

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ**

**ΕΚΤΙΜΗΣΗ ΤΟΥ ΤΕΧΝΙΚΟΥ ΑΣΦΑΛΙΣΤΡΟΥ
ΣΤΟΝ ΚΛΑΔΟ ΟΧΗΜΑΤΩΝ ΚΑΙ
ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΟΥ ΠΛΗΘΟΥΣ ΚΑΙ ΥΨΟΥΣ
ΤΩΝ ΖΗΜΙΩΝ**

ΜΠΕΛΕΓΡΙΝΟΥ ΚΩΝΣΤΑΝΤΙΝΑ

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Πειραιάς, Ιούλιος 2016

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Τα μέλη της Επιτροπής ήταν:

- Πολίτης Κωνσταντίνος (Επιβλέπων)
- Μ. Μπούτσικας
- Γ. Ψαρράκος

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

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS AND
INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN ACTUARIAL
SCIENCE AND RISK MANAGEMENT**

**PREMIUM CALCULATION AND MODELING OF
THE NUMBER AND SIZE OF CLAIMS IN THE
VEHICLE INSURANCE SECTOR**

BELEGRINOU KONSTANTINA

MSc Dissertation

Submitted to the Department of Statistics and Insurance Science of the
University of Piraeus in partial fulfillment of the requirements for the degree of
Master of Science in Actuarial Science and Risk Management

Piraeus, July 2016

Πίνακας περιεχομένων

ΕΥΧΑΡΙΣΤΙΕΣ	6
ΠΕΡΙΛΗΨΗ	7
ABSTRACT	8
ΚΕΦΑΛΑΙΟ 1.....	9
Εισαγωγικά.....	9
1.1 Εισαγωγή.....	9
1.2 Κίνδυνος.....	9
1.3 Ασφάλιση	10
1.3.1 Πώς λειτουργεί η ασφάλιση.....	11
1.3.2 Κόστος ασφάλισης και δίκαια ασφάλιστρα.....	11
1.3.3 Τεχνικό Ασφάλιστρο.....	12
1.4 Διαχείριση των κινδύνων	13
1.5 Αναλογιστική ανάλυση	13
1.6 Εποπτεία ασφαλιστικών εταιριών	14
1.7 Η ελληνική ασφαλιστική αγορά σήμερα.....	14
1.6.1 Solvency I και II.....	15
1.8 Διάρθρωση διπλωματικής εργασίας.....	18
ΚΕΦΑΛΑΙΟ 2.....	21
Περιγραφική Ανάλυση.....	21
2.1 Εισαγωγή.....	21
2.2 Αριθμητικά περιγραφικά μέτρα των ζημιών	21
2.3 Κόστος ζημιών κάθε ανά μήνα για κάθε έτος.....	25
2.4 Μέσο κόστος ζημιών ανά περιφέρεια και ανά νομό	26
2.5 Μέσο κόστος και πλήθος ζημιών ανά χρήση οχήματος.....	29
2.6 Σχετική συχνότητα πλήθους ζημιών ανά χρήση και ανά περιφέρεια.....	31
2.7 Ποσοστά σωματικής βλάβης και θανατηφόρων ατυχημάτων ανά χρήση.....	34
2.8 Αναλυτικά στοιχεία τριετίας θανατηφόρων σύμφωνα με την Ε.Α.	36
2.9 Εξέλιξη των ζημιών για κάθε έτος ανά μήνα	41
2.10 Ακραίες ζημιές	43
ΚΕΦΑΛΑΙΟ 3.....	46
Προσαρμογή κατανομών πιθανότητας για το πλήθος και το μέγεθος των ζημιών.....	46
3.1 Εισαγωγή.....	46
3.2 Μοντέλα για το πλήθος των ζημιών.....	46

3.2.1 Προσαρμογή πλήθους ζημιών ανά ημέρα.....	46
3.2.2 Προσαρμογή πλήθους ζημιών ανά μήνα	57
3.3 Ανάλυση ανά ζημιά	61
3.4 Συνολικές αποζημιώσεις ανά μήνα	62
ΚΕΦΑΛΑΙΟ 4.....	67
BOOTSTRAP.....	67
4.1 Εισαγωγή.....	67
4.2 Η παραδοσιακή προσέγγιση.....	67
4.3 Η μέθοδος bootstrap	69
4.3.1 Μέθοδος bootstrap και τυπικό σφάλμα.....	71
4.3.2 Το bootstrap διάστημα εμπιστοσύνης.....	74
4.3.3 Αδυναμίες της μεθόδου bootstrap	76
4.4 Εφαρμογή της μεθόδου bootstrap στις μηνιαίες αποζημιώσεις του χαρτοφυλακίου	79
ΚΕΦΑΛΑΙΟ 5.....	84
Γενικευμένα Γραμμικά Μοντέλα	84
5.1 Εισαγωγή.....	84
5.2 Παλινδρόμηση.....	84
5.3 Γενικά Χαρακτηριστικά ενός Γενικευμένου Γραμμικού Μοντέλου.....	85
5.4 Εφαρμογή στα δεδομένα	90
ΠΑΡΑΡΤΗΜΑ	98
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	108

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή της διπλωματικής μου, κύριο Πολίτη Κωνσταντίνο για την πολύ σημαντική καθοδήγηση, υποστήριξη, βοήθεια και υπομονή σε όλη την διάρκεια υλοποίησής της. Τέλος θέλω να ευχαριστήσω τους γονείς μου και την γιαγιά μου, που πάντα με στηρίζουν στις σπουδές μου, την εταιρεία μου, τους φίλους μου, τόσο για την συμπαράσταση, όσο και για τις χρήσιμες υποδείξεις τους.

ΠΕΡΙΛΗΨΗ

Για τον υπολογισμό του τεχνικού ασφαλίστρου στον κλάδο ασφάλισης αυτοκινήτων, λαμβάνονται υπόψη διάφοροι παράγοντες που αφορούν το όχημα και τον ασφαλιζόμενο οδηγό. Στην παρούσα διπλωματική εργασία γίνεται επισκόπηση των μεθόδων που χρησιμοποιούνται από τις ασφαλιστικές εταιρείες για τον υπολογισμό του τεχνικού ασφαλίστρου. Επίσης γίνεται μοντελοποίηση (σε πραγματικά δεδομένα αποζημιώσεων) του πλήθους και του μεγέθους των ζημιών με την χρήση γνωστών κατανομών πιθανότητας. Τέλος χρησιμοποιήθηκαν μέθοδοι γενικευμένων γραμμικών μοντέλων για να εξεταστεί η συμβολή αυτών των παραγόντων στον καθορισμό του ασφαλίστρου.

ABSTRACT

For the calculation of technical premium in motor department different factors relating to the vehicle and the driver insured are considered. In this thesis we get an overview of the methods used by insurance companies for calculating technical premium. We study the number of losses and the size of them using known probability distributions. Finally, we used methods of generalized linear models to see the contribution of these factors in calculating the technical premium.

ΚΕΦΑΛΑΙΟ 1

Εισαγωγικά

1.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζουμε γενικές αρχές και έννοιες της ασφαλιστικής επιστήμης. Συγκεκριμένα δίνονται βασικές έννοιες και γενικοί ορισμοί της ασφαλιστικής επιστήμης σε ένα ευρύ φάσμα. Επιπρόσθετα γίνεται αναφορά για την κατάσταση που επικρατεί σήμερα στην ασφαλιστική αγορά. Στη τελευταία ενότητα του κεφαλαίου γίνεται αναφορά στην διάρθρωση της διπλωματικής εργασίας.

Στα πλαίσια της νέας οικονομικής πραγματικότητας που έχουν δημιουργήσει οι δυνάμεις της παγκοσμιοποίησης από την μια πλευρά και η τεχνολογική πρόοδος από την άλλη, διαπιστώνεται ότι η διεύρυνση της ασφαλισιμότητας των κινδύνων καθώς και ο εκσυγχρονισμός των μηχανισμών της ιδιωτικής ασφάλισης και κυρίως οι συμπράξεις δημοσίου- ιδιωτικού τομέα στη διαχείριση των κινδύνων μπορούν να συμβάλλουν καθοριστικά στην επιτυχή αντιμετώπιση καίριων κοινωνικών και οικονομικών ζητημάτων στις σύγχρονες κοινωνίες.

1.2 Κίνδυνος

Ο όρος κίνδυνος (risk) έχει πολλές έννοιες στην εμπορική και καθημερινή ζωή. Γενικά, ο όρος κίνδυνος χρησιμοποιείται για να περιγράψει κάθε κατάσταση στην οποία υπάρχει αβεβαιότητα σχετικά με το αναμενόμενο αποτέλεσμα. Στις πιθανότητες και στη στατιστική, στην οικονομική διαχείριση και στην διαχείριση των επενδύσεων ο όρος κίνδυνος συχνά χρησιμοποιείται με μια συγκεκριμένη έννοια, που σκοπό έχει να καταδείξει πιθανή απόκλιση μιας αναμενόμενης τιμής. Σε άλλες καταστάσεις, ο όρος κίνδυνος μπορεί να αναφέρεται στις ζημιές που συνδέονται με την έλευση ή μη κάποιου ζημιογόνου γεγονότος. Στην ασφαλιστική αγορά η έννοια του κινδύνου χρησιμοποιείται για να περιγράψει την ενδεχόμενη υψηλή τιμή των ζημιών που θα πληρωθούν από τον ασφαλιστή.

Η μεταφορά του κινδύνου στις ασφαλιστικές εταιρείες περιορίζει το επίπεδο του κινδύνου για την κοινωνία ως σύνολο. Κατά την μεταφορά του κινδύνου στις ασφαλιστικές εταιρείες ο κίνδυνος της ζημίας ή της μη ζημίας που αντιμετωπίζουμε αλλάζει. Πληρώνουμε ασφάλιστρα για να έχουμε την ασφάλιση της μη ζημίας. Όταν μεταφέρουμε τον κίνδυνο, οι ασφαλιστικές εταιρείες αναλαμβάνουν κίνδυνο ο οποίος είναι σε αρκετές περιπτώσεις πολύ μικρότερος : είναι ο κίνδυνος της κακής πρόβλεψης των ζημιών. Όσο μεγαλύτερος είναι ο αριθμός των ζημιογόνων ενδεχομένων, τόσο μικρότερος είναι ο κίνδυνος της κακής πρόβλεψης των

μελλοντικών ζημιών. Έτσι η μεταφορά του κινδύνου στις ασφαλιστικές εταιρείες μειώνει τον κίνδυνο για την κοινωνία ως σύνολο, χάρη στο νόμο των μεγάλων αριθμών.

Ο κίνδυνος και η κάλυψή του είναι ένα θέμα που έχει απασχολήσει πολλούς επιστήμονες. Για αιώνες η κάλυψη κινδύνου ήταν αντικείμενο μόνο των ασφαλιστικών εταιρειών. Τις τελευταίες δεκαετίες όμως τόσο οι τράπεζες όσο και τα χρηματοπιστωτικά ιδρύματα πωλούν κάλυψη κινδύνου σε εταιρείες ή ιδιώτες. Πλέον και οι τρεις παραπάνω φορείς συλλέγουν και διαχειρίζονται κινδύνους. Στις δραστηριότητες της διοικητικής κινδύνου (risk management) είναι ο διαχωρισμός και η αντιστάθμιση των κινδύνων (hedging). Κάποιοι κίνδυνοι δεν είναι δυνατόν να αντισταθμιστούν λόγω μη ύπαρξης αντιθέτου κινδύνου στην αγορά. Όταν λοιπόν δεν είναι δυνατή η αντιστάθμιση του κινδύνου πρέπει να μετρηθεί προσεκτικά ο κίνδυνος και να δημιουργηθούν τα κατάλληλα αποθέματα για την αντιμετώπιση του. Οι εταιρείες στις μέρες μας δίνουν μεγάλη σημασία και βελτιώνουν συνεχώς τις τεχνικές μέτρησης των κινδύνων και αξιολόγησης των κερδοφόρων τομέων της επιχείρησης. Η διοικητική των κινδύνων πρέπει να βρει την χρυσή τομή μεταξύ των αντικρουόμενων συμφερόντων πελατών και ασφαλισμένων με αυτά των μετόχων. Οι μεν πρώτοι επιθυμούν φυσικά την μεγιστοποίηση του πλούτου τους ενώ οι μέτοχοι από την πλευρά τους απαιτούν απόδοση ιδίων κεφαλαίων σύμφωνα με τον κίνδυνο της επένδυσή τους. (Μ. Νεκτάριος, 2014)

1.3 Ασφάλιση

Η ασφάλιση θεωρείται ως ένα κοινωνικό εργαλείο, σύμφωνα με το οποίο μια ομάδα ατόμων μεταφέρει τον κίνδυνο σε ένα άλλο μέρος με τέτοιο τρόπο που το μέρος αυτό (ασφαλιστική εταιρεία) συνθέτει ή ομαδοποιεί όλα τα ζημιόγωνα ενδεχόμενα μαζί. Η ομαδοποίηση των ζημιόγων ενδεχομένων επιτρέπει περισσότερο ακριβείς στατιστικές προβλέψεις των μελλοντικών ζημιών. Τα άτομα που μεταφέρουν τον κίνδυνο σε έναν εξειδικευμένο διαχειριστή ονομάζονται ασφαλισμένοι. Ο εξειδικευμένος διαχειριστής που αποδέχεται τους κινδύνους οι οποίοι μεταφέρονται από τους ασφαλισμένους ονομάζεται ασφαλιστής.

Η ασφάλιση είναι ένα από τα εργαλεία που στοχεύουν στην διατήρηση τόσο του πλούτου των πολιτών όσο και της αξίας των επιχειρήσεων. Άνθρωποι και επιχειρήσεις εργάζονται για να μεγιστοποιήσουν αυτή την αξία. Ένα ουσιαστικό στοιχείο στην μεγιστοποίηση της αξίας της περιουσίας μας είναι η διατήρηση και η σταθερότητα. Όταν η ασφάλιση αγοράζεται από μια αξιόπιστη και με υψηλή πιστοποίηση ασφαλιστική εταιρεία, εγγυάται την διατήρηση της περιουσίας και της οικονομικής αξίας. Τα αρνητικά αποτελέσματα των περισσότερων κινδύνων μπορούν να μετριαστούν μεταφέροντάς τα σε ασφαλιστικές εταιρείες. Οι

ασφαλιστικές εταιρείες πωλούν ξεχωριστά ασφαλιστήρια συμβόλαια, τα οποία καλύπτουν τους ξεχωριστούς κινδύνους.(M.Νεκτάριος,2014)

1.3.1 Πώς λειτουργεί η ασφάλιση

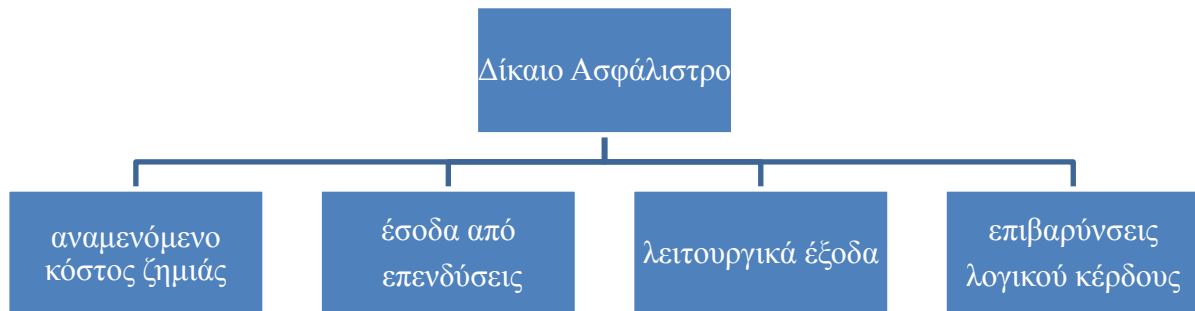
Η ασφάλιση λειτουργεί με τα παρακάτω βήματα

- ◆ Ο κίνδυνος μεταφέρεται από ένα άτομο ή μονάδα (ασφαλισμένος) σε ένα άλλο μέρος (ασφαλιστής)
- ◆ Ο ασφαλιστής συγκεντρώνει όλα τα ζημιόγωνα ενδεχόμενα μαζί, προκειμένου να υπολογίσει τις μελλοντικές πιθανές ζημιές με κάποιο επίπεδο ακρίβειας. Ο ασφαλιστής είναι σε θέση να χρησιμοποιήσει διάφορες τεχνικές πρόβλεψης που εξαρτώνται από την κατανομή των ζημιών.
- ◆ Η ομαδοποίηση των κινδύνων οδηγεί σε μείωση του συνολικού κινδύνου μέσα στην κοινωνία, αυτό οφείλεται στο γεγονός πως όσο μεγαλύτερος είναι ο αριθμός των ασφαλισμένων κινδύνων τόσο η πρόβλεψη του ασφαλιστή βελτιώνεται.
- ◆ Οι ασφαλιστές ομαδοποιούν ζημιόγωνα ενδεχόμενα που πρέπει να είναι ομοιογενή, προκειμένου να υπολογίσουν τον δικό τους κίνδυνο λανθασμένης πρόβλεψης.
- ◆ Οι ασφαλιστές κατηγοριοποιούν τους κινδύνους σε ομοιογενείς κατηγορίες μέσω του *underwriting*, δηλαδή της διαδικασίας εκτίμησης και ταξινόμησης του κινδύνου. (M.Νεκτάριος,2014)

1.3.2 Κόστος ασφάλισης και δίκαια ασφάλιστρα

Μια βασική αρχή της τιμολόγησης των ασφαλιστρών είναι ότι οι ασφαλιστικές εταιρείες, προκειμένου να προσφέρουν κάλυψη θα πρέπει να λαμβάνουν ασφάλιστρα που 1) είναι για αυτές επαρκή, ώστε να χρηματοδοτήσουν τις αναμενόμενες απαιτήσεις για αποζημιώσεις καθώς και τα λειτουργικά τους έξοδα και 2) παρέχουν ένα αναμενόμενο κέρδος, το οποίο αντισταθμίζει το κόστος απόκτησης των απαραίτητων κεφαλαίων για την υποστήριξη της πώλησης των ασφαλιστικών καλύψεων. Το επίπεδο του ασφαλίστρου που είναι απλώς επαρκές για την χρηματοδότηση του αναμενόμενου κόστους του ασφαλιστή και για την παροχή στους ιδιοκτήτες της ασφαλιστικής εταιρείας μιας δίκαιης απόδοσης των επενδυμένων κεφαλαίων τους είναι γνωστό ως δίκαιο ασφάλιστρο(*fair premium*). Το δίκαιο

ασφάλιστρο είναι το ασφάλιστρο που θα εφαρμόζονταν σε μια τέλεια ανταγωνιστική αγορά.



Το παραπάνω διάγραμμα παρουσιάζει τους σημαντικότερους παράγοντες που καθορίζουν το δίκαιο ασφάλιστρο. (Μ. Νεκτάριος, 2010)

1.3.3 Τεχνικό Ασφάλιστρο

Ένας από τους κύριους σκοπούς κάθε ασφαλιστικής εταιρείας είναι η ορθή τιμολόγηση του ασφαλιστρού. Το ολικό ασφάλιστρο είναι το ποσό που ο λήπτης της ασφάλισης οφείλει να καταβάλλει στην ασφαλιστική εταιρεία για την παροχή της ασφαλιστικής κάλυψης, ώστε να έχει ισχύ η ασφαλιστική σύμβαση. Το ολικό ασφάλιστρο προκύπτει ως ο πολλαπλασιασμός του τεχνικού ασφαλιστρού με κάποιες επιμέρους παραμέτρους (loadings). Οι παράμετροι αυτοί περιλαμβάνουν το κόστος προκλήσεως, τα διαχειριστικά έξοδα, το κέρδος και το ποσοστό ασφαλείας. Το τεχνικό ασφάλιστρο είναι το ασφάλιστρο που θα επαρκεί για την κάλυψη μόνο εκείνων των ζημιών που θα προκύψουν από κάποιο συγκεκριμένο ασφαλιστικό κίνδυνο (π.χ. Αστική Ευθύνη).

Συγκεκριμένα η συχνότητα ζημιών επί το μέσο κόστος ζημιάς του κινδύνου που εξετάζεται προσδιορίζουν το τεχνικό ασφάλιστρο του κινδύνου αυτού.

Στη συνέχεια παρουσιάζουμε πως προκύπτει η παραπάνω σχέση

$$T.A. = \frac{\text{Συνολικό Κόστος Ζημιών}}{\text{Πλήθος οχημάτων}}$$

$$= \frac{\text{Μέσο Κόστος Ζημιών} \times \text{Πλήθος Ζημιών}}{\text{Πλήθος Οχημάτων}} =$$

$$\frac{\text{Μέσο Κόστος Ζημιών} \times \text{Σχετική Συχνότητα} \times \text{Πλήθος Οχημάτων}}{\text{Πλήθος Οχημάτων}}$$

=Μέσο Κόστος Ζημιών x Σχετική Συχνότητα

Με τον όρο σχετική συχνότητα εννοούμε το ποσοστό των οχημάτων από τον συνολικό στόλο που έχουν ζημιά.

Στη παρούσα διπλωματική μελετάμε μοντέλα τόσο για το μέσο κόστος ζημιάς όσο και για τις συχνότητες των ζημιών σε ένα χαρτοφυλάκιο με πραγματικά δεδομένα από ζημιές αστικής ευθύνης αυτοκινήτων.

1.4 Διαχείριση των κινδύνων

Η πληροφορική, η παγκοσμιοποίηση και η καινοτομία στις τεχνολογίες της χρηματοοικονομικής οδήγησαν στον όρο διαχείριση επιχειρηματικών κινδύνων (enterprise risk management – ERM). Η διαχείριση επιχειρηματικών κινδύνων περιλαμβάνει τη διαχείριση τόσο των καθαρών όσο και των κερδοσκοπικών κινδύνων.

Η τυπική λειτουργία της διαχείρισης των κινδύνων περιλαμβάνει τα εξής στάδια

- ◆ Αναγνώριση των κινδύνων
- ◆ Εκτίμηση των κινδύνων
- ◆ Πρόβλεψη μελλοντικής συχνότητας και έντασης ζημιών
- ◆ Περιορισμό των κινδύνων
- ◆ Εξεύρεση λύσεων για τον περιορισμό των κινδύνων
- ◆ Δημιουργία σχεδίων για την εκπόνηση αναλύσεων κόστους – ωφέλειας για τις εφαρμογές προγραμμάτων ελέγχου των ζημιών

1.5 Αναλογιστική ανάλυση

Η αναλογιστική ανάλυση (actuarial analysis) είναι μια εξαιρετικά εξειδικευμένη μαθηματική ανάλυση που σχετίζεται τόσο με την οικονομική πλευρά της ασφάλισης όσο και με την εκτίμηση του κινδύνου. Η αναλογιστική ανάλυση λαμβάνει τα δεδομένα των ζημιών του παρελθόντος και τα προβάλλει στο μέλλον, προκειμένου να καθορίσει τα αποθέματα που

χρειάζεται να κρατήσει η ασφαλιστική εταιρεία και τα ασφάλιστρα που πρόκειται να χρεώσει. Ένας αναλογιστής καθορίζει τα κατάλληλα ασφάλιστρα και αποθέματα, πιστοποιεί τις οικονομικές καταστάσεις, συμμετέχει στην δημιουργία των προϊόντων και βοηθά στον προγραμματισμό της διοίκησης.

1.6 Εποπτεία ασφαλιστικών εταιριών

Οι ασφαλιστικές εργασίες εποπτεύονται αυστηρά στις περισσότερες αναπτυγμένες χώρες. Οι νομοθετικές ρυθμίσεις έχουν ως κύριο στόχο την εξασφάλιση της φερεγγυότητας, καθώς τα κίνητρα και οι συνθήκες μιας αγοράς χωρίς ρυθμίσεις ενδεχομένως να μην οδηγούν στο επιθυμητό επίπεδο φερεγγυότητας. Κάθε χώρα διαθέτει μια εποπτική αρχή, η οποία εφαρμόζει την κρατική νομοθεσία που διέπει την αγορά και πώληση ασφάλισης. Συμπληρωματικά με την ασφαλιστική νομοθεσία, η εποπτική αρχή συνήθως είναι εξουσιοδοτημένη να καθορίζει κανόνες και διαδικασίες, που αφορούν κυρίως την εφαρμογή των νομοθετικών ρυθμίσεων.

Οι ρυθμίσεις σχετικά με την φερεγγυότητα των ασφαλιστικών εταιρειών έχουν δυο κύρια μέρη:

- ✚ Το ρυθμιστικό πλαίσιο εποπτείας για τον έλεγχο του κινδύνου αφερεγγυότητας του ασφαλιστή καθώς και τους περιορισμούς στα κεφάλαια και στα στοιχεία του ενεργητικού των ασφαλιστών, και
- ✚ Τα κρατικά εγγυητικά κεφάλαια, τα οποία καλύπτουν ορισμένες από τις υποχρεώσεις των αφερέγγυων ασφαλιστικών εταιρειών προς τους ασφαλισμένους.

1.7 Η ελληνική ασφαλιστική αγορά σήμερα

Στην Ελλάδα η ιδιωτική ασφάλιση έχει κατά κύριο λόγο συμπληρωματικό ρόλο, καθώς το Ελληνικό κράτος παρέχει υποχρεωτική δημόσια ιατροφαρμακευτική ασφάλιση καθώς και (μέχρι πρόσφατα) πλήθος συνταξιοδοτικών ταμείων. Η μόνη εξαίρεση στον κανόνα είναι η ασφάλιση αστικής ευθύνης αυτοκινήτου η οποία είναι υποχρεωτική για όλους τους κατόχους μεταφορικών μέσων.

Εξαιτίας αυτού και σε συνδυασμό με την χαμηλό συντελεστή ασφαλιστικής συνείδησης των Ελλήνων, η ασφαλιστική αγορά στην Ελλάδα είναι συγκριτικά στην τελευταία θέση στην Ευρώπη όσον αφορά τον δείκτη Ασφάλιστρα προς Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ) (Ελλάδα 2,2% με στοιχεία 2013, ευρωπαϊκός μέσος όρος 7,69%).(ιστοσελίδα nextdeal.gr)

Στην Ελλάδα σήμερα (2016) δραστηριοποιούνται συνολικά 41 ιδιωτικές ασφαλιστικές εταιρείες, εκ των οποίων:

- ◆ 4 είναι Ασφαλιστικές εταιρείες Ζωής
- ◆ 19 είναι Ασφαλιστικές εταιρείες Ζημιών
- ◆ 17 είναι Ασφαλιστικές εταιρείες μικτής δραστηριότητας (πηγή ιστοσελίδα Τράπεζα Ελλάδος)

Αξίζει επίσης να σημειωθεί ότι υπάρχει μεγάλη διείσδυση ξένων πολυεθνικών ασφαλιστικών, οι οποίες δραστηριοποιούνται στην Ελλάδα μέσω υποκαταστημάτων τους εδώ. Από τις εταιρείες αυτές, οι περισσότερες λαμβάνουν τεχνογνωσία, όσον αφορά τα συστήματα και τις διαδικασίες που χρησιμοποιούν, από τις μητρικές εταιρείες τους, με αποτέλεσμα να απολαμβάνουν συγκριτικά πλεονεκτήματα σε σχέση με τις αμιγώς ελληνικές εταιρείες, οι οποίες είναι συνήθως μικρότερες όσον αφορά την παραγωγή ασφαλιστρών και των διαθεσίμων κεφαλαίων.

Τα τελευταία χρόνια στην Ελλάδα παρατηρείται μια μείωση της ιδιωτικής δαπάνης για ασφαλιστικά προϊόντα, τάση που επιβεβαιώθηκε και το 2015 και μάλιστα διευρύνθηκε εν μέρει λόγω της επιβολής των capital controls.

Επιπρόσθετα ο αριθμός των ασφαλιστικών εταιρειών στην Ελλάδα βαίνει μειούμενος τα τελευταία χρόνια κυρίως, αλλά όχι μόνο, λόγω της εφαρμογής ενός αυστηρότερου ρυθμιστικού πλαισίου (Solvency II), βάσει του οποίου οι εταιρείες χρειάζονται περισσότερα κεφάλαια προκειμένου να επιτρέπεται και να διασφαλίζεται η ομαλή λειτουργία τους. Στη συνέχεια του κεφαλαίου θα δώσουμε τα βασικά χαρακτηριστικά του νέου πλαισίου(Solvency II) επιχειρώντας μια σύγκριση με την παλαιότερη νομοθεσία (Solvency I).

1.6.1 Solvency I και II

Στην ελληνική ασφαλιστική αγορά, το πλαίσιο εποπτείας προέρχεται από μια σειρά κοινοτικών οδηγιών που τέθηκαν σε εφαρμογή το 1985 για πρώτη φορά και στη συνέχεια το πλαίσιο αυτό εναρμονίστηκε με τις διατάξεις νεότερων Οδηγιών, με πιο πρόσφατη αυτή του 2009 που επέβαλε νέο εποπτικό πλαίσιο που ισχύει από το 2016. Ο στόχος των οδηγιών της Ε.Ε. ήταν να δημιουργήσει τις προϋποθέσεις για την λειτουργία της Ενιαίας Ευρωπαϊκής

ασφαλιστικής Αγοράς. Αρχικά δημιουργήθηκε το Solvency I το οποίο παρέχει σχετική ομοιομορφία στον υπολογισμό της φερεγγυότητας των ασφαλιστικών εταιρειών που βασίζεται κυρίως σε χρηματοοικονομικούς παράγοντες, χωρίς κάποια ιδιαίτερη πολυπλοκότητα και χωρίς να δίνει έμφαση στους επιμέρους κινδύνους που πιθανό να διαταράξουν και να επηρεάσουν σημαντικά τη φερεγγυότητα μιας ασφαλιστικής επιχείρησης, λαμβάνοντας υπόψη μόνο τους ασφαλιστικούς κινδύνους (ύψος ασφαλίσεων και ύψος αποζημιώσεων). Με την πάροδο των ετών διαπιστώθηκε ότι το Solvency I δεν εξυπηρετεί επαρκώς τις ανάγκες εναρμόνισης της ευρωπαϊκής ασφαλιστικής αγοράς. Χαρακτηρίστηκε ως η εύκολη λύση (one size fits all) και κρίθηκε αναγκαία η αντικατάσταση του. (Μ.Νεκτάριος,2010)

Νέα σελίδα για την ασφαλιστική αγορά ήρθε με την οδηγία Solvency II που τέθηκε σε εφαρμογή από το 2016 ενώ οι προετοιμασίες για τον σχεδιασμό του άρχισαν το 2001.

Βασικοί στόχοι του Solvency II είναι οι ακόλουθοι :

- ◆ Προστασία των ευρωπαϊκών πολιτών- κατόχων ασφαλιστηρίων συμβολαίων
- ◆ Θέσπιση μεθόδου υπολογισμού του περιθωρίου φερεγγυότητας η οποία θα είναι προσαρμοσμένη στους πραγματικούς κινδύνους που αντιμετωπίζουν οι ασφαλιστικές εταιρείες
- ◆ Προσαρμογή του περιθωρίου φερεγγυότητας σύμφωνα με τις αυξανόμενες ανάγκες και εξελίξεις της ασφαλιστικής αγοράς
- ◆ Αποφυγή της υπερβολικής κεφαλαιοποίησης των ασφαλιστικών εταιρειών.

Η φερεγγυότητα των ασφαλιστικών επιχειρήσεων στο εξής θα βασίζεται σε 3 πυλώνες. Ο πρώτος πυλώνας καθορίζει τις ποιοτικές απαιτήσεις της φερεγγυότητας. Ο δεύτερος πυλώνας τις ποιοτικές απαιτήσεις, ενώ παράλληλα θέτει τις νέες εποπτικές δυνατότητες και λειτουργίες. Και ο τρίτος πυλώνας αφορά την διαφάνεια και δημοσιοποίηση των στοιχείων.(ιστοσελίδα Solvency II)

Αναλυτικά ο πρώτος πυλώνας περιλαμβάνει τον καθορισμό των κανόνων πάνω στους οποίους θα υπολογιστούν τα τεχνικά αποθέματα, οι επενδύσεις σε κεφάλαια, η διαχείριση των περιουσιακών στοιχείων καθώς και το κεφάλαιο που θα εξασφαλίζει την επιθυμητή φερεγγυότητα στις ασφαλιστικές εταιρείες. Στο στάδιο αυτό αναλύονται οι έννοιες του MCR(Minimum Capital Requirement) και SCR(Solvency Capital Requirement) όπου :

- MCR: είναι το ελάχιστο επιθυμητό κεφάλαιο που πρέπει να διατηρεί μια ασφαλιστική εταιρεία για να διασφαλίζει την φερεγγυότητά της. Οποιοδήποτε κεφάλαιο κάτω του ελάχιστου δίνει την δυνατότητα στην Εποπτική Αρχή να

επέμβει και να ανακαλέσει την άδεια της ασφαλιστικής εταιρείας μεταφέροντας το σύνολο των συμβολαίων της σε μια άλλη ασφαλιστική εταιρεία.

- SCR: Το κεφάλαιο φερεγγυότητας είναι το επιθυμητό κεφάλαιο το οποίο θα πρέπει να κατέχει μια ασφαλιστική εταιρεία λαμβάνοντας υπόψη όλους τους πιθανούς κινδύνους για την φερεγγυότητα της σύμφωνα με την οδηγία Solvency II. Το απαιτούμενο κεφάλαιο φερεγγυότητας SCR αντιστοιχεί σε ένα επίπεδο κεφαλαίου το οποίο επιτρέπει στην επιχείρηση να απορροφήσει σημαντικές απρόβλεπτες ζημιές και να παρέχει εύλογη κάλυψη στους ασφαλισμένους και τους δικαιούχους. Όταν μια επιχείρηση δεν πληροί το SCR θα πρέπει να αποκαταστήσει σε εύλογο χρόνο το απαραίτητο κεφάλαιο για την κάλυψη της απαίτησης αυτής, με βάση ένα συγκεκριμένο και εφικτό σχέδιο που θα υποβάλλεται προς έγκριση στην Εποπτική Αρχή.

Οι βασικοί κίνδυνοι στους οποίους εκτίθεται μια ασφαλιστική εταιρεία και στους οποίους καλείται να ανταπεξέλθει ο υπολογισμός του SCR είναι οι ακόλουθοι :

- i. Κίνδυνος underwriting
- ii. Κίνδυνος ρευστότητας
- iii. Κίνδυνος αγοράς
- iv. Πιστωτικός κίνδυνος
- v. Συστηματικός κίνδυνος
- vi. Επιχειρηματικός κίνδυνος

Ο δεύτερος πυλώνας καθορίζει τις ποιοτικές προδιαγραφές της φερεγγυότητας δηλαδή τις αρχές εσωτερικού ελέγχου πάνω στις οποίες θα βασίζεται η αξιολόγηση του ασφαλιστικού κινδύνου.

Ο τρίτος πυλώνας αφορά αποκλειστικά την ενίσχυση της πειθαρχίας της αγοράς και την οριοθέτηση των κανονισμών δημοσίευσης οικονομικών στοιχείων τα οποία θα είναι ομοιόμορφα σε όλη την ενιαία ευρωπαϊκή ασφαλιστική αγορά.

Η αρχή στην οποία βασίζεται το Solvency II σε αντίθεση με την προσέγγιση που βασίζεται στον κανόνα Solvency I είναι ότι δίνει στην ασφαλιστική εταιρεία την ευελιξία να εφαρμόσει είτε το πρότυπο μοντέλο είτε τη χρήση εσωτερικού υποδείγματος το οποίο πρέπει να εγκριθεί από την εποπτική αρχή όσον αφορά τον υπολογισμό των δυο κεφαλαιακών απαιτήσεων SCR, MCR.(Μ. Νεκτάριος,2010)

Στη συνέχεια παρατίθενται σχηματικά οι τρεις πυλώνες.



Το Solvency II μέσα από τους πυλώνες, Οικονομικό Παν. Αθηνών, τμ. Στατιστικής

Το Solvency II είναι μια αποστολή που έχει ως στόχο την εισαγωγή και την ανάπτυξη του κινδύνου προσανατολισμένη στο εποπτικό πλαίσιο για τις ασφαλιστικές εταιρείες με την έννοια ότι οι επιχειρήσεις θα πρέπει να έχουν ύψος αποθεματικού με βάση τους κινδύνους που έχουν να αντιμετωπίσουν. Σε ένα τέτοιο πλαίσιο η κατάλληλη διαχείριση του κινδύνου γίνεται πυρήνας για την ορθότητα και την αποτελεσματικότητα του όλου συστήματος. Το προφίλ κινδύνου μιας επιχείρησης θα πρέπει να λαμβάνει υπόψη τόσο τον εσωτερικό όσο και τον εξωτερικό κίνδυνο που αντιμετωπίζει καθώς και τους μετρήσιμους και μη μετρήσιμους κινδύνους. Για να το πράξει αυτό πρέπει να γίνει ο συνδυασμός του πυλώνα I και του πυλώνα II. Το Solvency II προσπαθεί να ενσταλάξει στην επιχείρηση την εστίαση και διαχείριση κινδύνου εντός της δομής της επιχείρησης.

1.8 Διάρθρωση διπλωματικής εργασίας

Τα δεδομένα που έχουμε συλλέξει προέρχονται από ασφαλιστική εταιρεία Ζημιών. Πρόκειται για εταιρεία γενικών ασφαλίσεων που περιλαμβάνει τους εξής κλάδους

- ◆ Αστική ευθύνη αυτοκινήτων
- ◆ Ατυχημάτων
- ◆ Γενικής αστικής ευθύνης
- ◆ Ζημιών περιούσιας
- ◆ Λοιπών ζημιών σε πράγματα
- ◆ Μεταφερόμενων εμπορευμάτων

Συγκεκριμένα επεξεργαστήκαμε τις ζημιές που προέρχονται από την αστική ευθύνη αυτοκινήτου για τη τριετία 2011, 2012, 2013. Πρόκειται για την αστική ευθύνη για ζημιές σε τρίτους από τροχαία ατυχήματα καθώς και αστική ευθύνη για υλικές (ίδιες) ζημιές στο αυτοκίνητο. Για τα τρία αυτά έτη του χαρτοφυλακίου της ασφαλιστικής εταιρείας είχαμε πληροφορίες που αφορούσαν το ύψος της ζημίας, την ακριβή ημερομηνία του ατυχήματος, την ηλικία του οδηγού, τον νομό, την χρήση του οχήματος, το αν υπήρχε σωματική βλάβη ή όχι και τέλος εάν το ατύχημα ήταν θανατηφόρο ή όχι. Τα προγράμματα που θα χρησιμοποιηθούν για την στατιστική ανάλυση και επεξεργασία των δεδομένων είναι κατά κύριο λόγο η R (γλώσσα προγραμματισμού), το SAS Enterprise Guide, το EasyFit 5.6 Professional και το Excel.

Στην παρούσα εργασία όπως αναφερθήκαμε και παραπάνω δίνεται μια γενική εικόνα της ασφαλιστικής επιστήμης (βασικές έννοιες-γενικοί ορισμοί) καθώς και η κατάσταση που επικρατεί στην ασφαλιστική αγορά σήμερα. Έπειτα μελετήθηκαν τα περιγραφικά στοιχεία του χαρτοφυλακίου της ασφαλιστικής εταιρείας (κεφάλαιο 2). Στη συνέχεια (κεφάλαιο 3) γίνεται η ανάλυση των δεδομένων της ασφαλιστικής εταιρείας, έγινε προσαρμογή κατανομών πιθανότητας για το πλήθος και το μέγεθος των ζημιών της εταιρείας. Επιπρόσθετα εξετάσαμε και εφαρμόσαμε τη μέθοδο επαναδειγματοληψίας Bootstrap (κεφάλαιο 4) και προβλέψαμε που θα κυμαίνονται κατά μέσο όρο τον μήνα οι αποζημιώσεις που η εταιρεία θα κληθεί να αποδώσει για τις πιθανές ζημιές. Τέλος χρησιμοποιήθηκαν μέθοδοι γενικευμένων γραμμικών μοντέλων (κεφάλαιο 5) για να εξεταστεί η συμβολή διαφόρων παραγόντων στον καθορισμό του ασφαλιστήριου.

Στο σημείο αυτό θα εξηγήσουμε τις συντομογραφίες που έχουμε χρησιμοποιήσει στα προαναφερθέντα κεφάλαια

ΕΠΕΞΗΓΗΣΗ ΟΡΩΝ

Για τις χρήσεις των οχημάτων είχαμε τα εξής :

EIX: επιβατικά οχήματα

EIX. ENOIK. : ενοικιαζόμενα οχήματα

ΦΙΧ.ΑΓΡ. : φορτηγά αγροτικά οχήματα

ΦΙΧ: φορτηγά οχήματα

ΦΔΧ: φορτηγά δημοσίας χρήσης

ΜΟΤΟ: μοτοσυκλέτες και μοτοποδήλατα

ΛΟΙΠΑ: στην κατηγορία αυτή ανήκουν διάφορες χρήσεις όπως ταξί, μηχανήματα έργων, γεωργικοί ελκυστήρες, θεριζ/κη μηχανή οχήματα σχολής οδηγών, λεωφορεία τουριστικά.

ΚΕΦΑΛΑΙΟ 2

Περιγραφική Ανάλυση

2.1 Εισαγωγή

Στο παρόν κεφάλαιο θα μελετηθούν τα περιγραφικά στοιχεία του χαρτοφυλακίου της ασφαλιστικής εταιρείας που εξετάζουμε. Τα δεδομένα που θα χρησιμοποιηθούν είναι στοιχεία πραγματικών ζημιών της εταιρείας. Τα διαθέσιμα στοιχεία που έχουμε αφορούν το πλήθος των ζημιών, το μέγεθος της αποζημίωσης, την ημερομηνία ατυχήματος, την ημερομηνία αναγγελίας του ατυχήματος, την ομάδα χρήσης του οχήματος, τον νομό, την περιοχή και τέλος την ένδειξη σωματικής βλάβης και θανατηφόρου ατυχήματος.

Τα προγράμματα που θα χρησιμοποιηθούν για την στατιστική ανάλυση και επεξεργασία των δεδομένων είναι κατά κύριο λόγο το SAS Enterprise Guide ,η R (γλώσσα προγραμματισμού) καθώς και το Excel.

Πιο συγκεκριμένα μελετήθηκαν τα περιγραφικά μέτρα όπως και το μέσο κόστος ζημιών για κάθε έτος .Μελετήθηκαν οι ανά περιοχή και ανά τύπο οχήματος συχνότητες των ζημιών, τα ποσοστά σωματικών βλαβών, θανατηφόρων ατυχημάτων καθώς και οι ακραίες τιμές του χαρτοφυλακίου.

2.2 Αριθμητικά περιγραφικά μέτρα των ζημιών

Βασικός σκοπός είναι η παρουσίαση των τιμών του χαρτοφυλακίου ώστε να γίνει μια πρώτη ερμηνεία των αποτελεσμάτων και να ανιχνευτούν κάποια ιδιαίτερα χαρακτηριστικά τους. Τα κυριότερα μέτρα που θα αναλύσουμε είναι :

- Το μέγιστο και ελάχιστο του κόστους των ζημιών
- Το πρώτο και τρίτο τεταρτημόριο
- Η διάμεσος
- Ο μέσος
- Η διακύμανση
- Η τυπική απόκλιση
- Η κύρτωση
- Η λοξότητα

Οι πρώτοι πίνακες αναφέρονται σε κάθε έτος ξεχωριστά ενώ ο τελευταίος εξετάζει σφαιρικά και τα τρία έτη του χαρτοφυλακίου.

Για κάθε έτος δίνονται τα παρακάτω στοιχεία.

Περιγραφικά για το πλήθος των ζημιών ανά έτος

	2011	2012	2013
Ελάχιστο	9,23	9,23	9,23
Μέγιστο	685.043,42	704.736,65	684.661
1 ^ο τεταρτημόριο	342,3	321,8	327,8
3 ^ο τεταρτημόριο	1120,77	1094	1019,25
Διάμεσος	975,9	943,1	852
Μέσος	2051,96	2065,46	1886,48
Διακύμανση	200765150	267843830	216369801
Τυπική Απόκλιση	14169,15	16365,94	14709,51
Κύρτωση	831,3817	832,7311	881,7723
Λοξότητα	25,53989	26,27532	26,56911
Συνολικό Κόστος	48.329.901,55	38.760.347,17	37.867.327,62

Σύμφωνα με τον παραπάνω πίνακα προκύπτουν τα παρακάτω συμπεράσματα για τις ζημιές της ασφαλιστικής εταιρείας:

- Η μικρότερη ζημία ύψους 9,23€ είναι ίδια και για τα 3 έτη ενώ η μεγαλύτερη ύψους 704.700€ σημειώθηκε το έτος 2012. Το μέσο κόστος ζημιών είναι μικρότερο το έτος 2013 και είναι 1.887€ ενώ το μεγαλύτερο το 2012 με 2.065€. Παρατηρείται μικρή διαφορά της μέσης τιμής του ύψους του κόστους ζημιάς για τα 3 έτη.
- Και στα 3 έτη η διασπορά είναι μεγάλη .Ειδικότερα το έτος 2012 οι τιμές απέχουν περισσότερο από τη μέση τιμή συγκριτικά με τα έτη 2011 και 2013.

- Το έτος 2011 το 1^ο τεταρτημόριο δηλαδή το σημείο κάτω από το οποίο βρίσκεται το 25% των αποζημιώσεων είναι 342 €, ενώ το 3^ο δηλαδή το σημείο κάτω από το οποίο βρίσκεται το 75% των αποζημιώσεων είναι 1.121€. Για το 2012 το 25% των αποζημιώσεων είναι κάτω από 322€ ενώ το 75% κάτω από 1.094€. Τέλος για το 2013 έχουμε 328€ και 1.019 € αντίστοιχα για κάθε τεταρτημόριο.
- Σχετικά με την κύρτωση αξίζει να σημειωθεί πως και στα 3 έτη του χαρτοφυλακίου παρατηρούμε παρόμοια αριθμητικά μεγέθη (831,3817 για το 2011 ,832,7311 για το 2012 και 881,7723 για το 2013). Η κύρτωση είναι μια έννοια που προσπαθεί να περιγράψει το σχήμα που έχει η κορυφή μια κατανομής πιθανοτήτων. Αν συγκρίνουμε την κύρτωση των αποζημιώσεων μας με αυτή της κανονικής κατανομής ,τότε παρατηρούμε πως η κατανομή των δεδομένων μας είναι λεπτόκυρτη. Για να εξάγουμε συμπεράσματα θα συγκρίνουμε την τιμή της κύρτωσης που υπολογίσαμε για κάθε έτος με την τιμή 3. Η τιμή αυτή είναι ο αντίστοιχος συντελεστής κύρτωσης για την κανονική κατανομή. Έστω β ο συντελεστής κύρτωσης, όταν $\beta > 3$ η κατανομή είναι λεπτόκυρτη και περιγράφει κορυφές που είναι αιχμηρές ,συνήθως έχει θετική υπερβάλλουσα κύρτωση.
- Το ίδιο παρατηρούμε και στα 3 έτη του χαρτοφυλακίου για την λοξότητα. Η λοξότητα εκφράζει την ασυμμετρία μιας κατανομής συχνοτήτων. Μέτρο της λοξότητας είναι ο συντελεστής της. Όπως διαπιστώνουμε και για τα 3 έτη αλλά και για το χαρτοφυλάκιο συνολικά έχει θετική τιμή. Αυτό δηλώνει πως οι περιπτώσεις ομαδοποιούνται περισσότερο στα αριστερά του μέσου με τις περισσότερες από τις ακραίες τιμές στα δεξιά.
- Στην τελευταία γραμμή του πίνακα παρουσιάζονται τα συνολικά κόστη που η εταιρεία απέδωσε σε κάθε έτος ξεχωριστά. Το 2011 οι αποζημιώσεις ήταν υψηλότερες σε σχέση με τα δυο άλλα έτη. Για το 2012, 2013 οι διαφορές στο ύψος των αποζημιώσεων καθώς κινούνται στα ίδια επίπεδα.

Συμπερασματικά μπορούμε να πούμε ότι παρατηρούμε παρόμοια χαρακτηριστικά και για τα 3 έτη. Οι ζημιές λοιπόν στην εταιρεία μας και στα 3 χρόνια δεν παρουσίασαν έντονες διαφορές τόσο στο πλήθος όσο και στο ύψος τους. Πρόκειται για μια σταθερή και αναμενόμενη εξέλιξη των ζημιών κατά τα 3 έτη μελέτη

Ο πίνακας που ακολουθεί περιλαμβάνει περιγραφικά μεγέθη για όλο το χαρτοφυλάκιο.

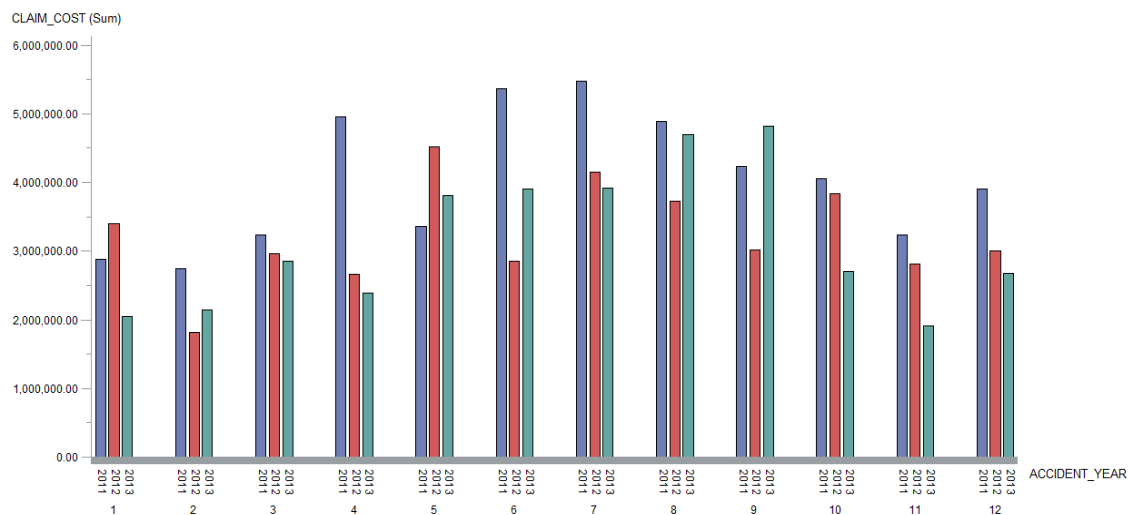
Ελάχιστο	9,23
Μέγιστο	704736,65
1^ο τεταρτημόριο	332,54
3^ο τεταρτημόριο	1089,52
Διάμεσος	896,13
Μέσος	2002,78
Διακόμανση	225960364,1
Τυπική Απόκλιση	15031,98
Κύρτωση	859,789
Λοξότητα	26,28198

Σχολιασμός ελάχιστης ζημιάς (9,23€)

Η εταιρεία έχει εξωτερικούς συνεργάτες οι οποίοι παίρνουν εντολή από τον κλάδο ζημιών της εταιρείας ώστε να προσκομίσουν στον κλάδο αντίγραφα βιβλίων συμβάντων (ΑΒΣ), δικογραφίες, έρευνες ζημιών και άλλα έγγραφα τα οποία χρειάζονται για τον διακανονισμό και τον έλεγχο μιας ζημιάς. Ύστερα από την παράδοση των εγγράφων αυτών προσκομίζουν στην εταιρεία τιμολόγια παροχής υπηρεσιών ώστε να πληρωθούν το αντίτιμο. Συνεπώς το 9,23€ είναι η πληρωμή ενός τέτοιου τιμολογίου παροχής υπηρεσιών.

2.3 Κόστος ζημιών κάθε ανά μήνα για κάθε έτος

Στο παρακάτω διάγραμμα απεικονίζεται το συνολικό ύψος του κόστους ζημιών ανά μήνα ατυχήματος για τα έτη 2011,2012,2013.



Διάγραμμα 2.1

Επιχειρώντας μια πρώτη ανάλυση του διαγράμματος είναι φανερό ότι το 2011 είναι η χρονιά με το υψηλότερο συνολικό κόστος αποζημιώσεων (περίπου 48 εκ.). Παρατηρείται ραγδαία πτώση το έτος 2012 ύψους 20% (περίπου 38 εκ.), ακολουθεί μικρή πλέον το έτος 2013 της τάξεως του 2% (περίπου 37 εκ.). Πιο συγκεκριμένα για το έτος 2011 ο μήνας που εμφανίζει το μεγαλύτερο ύψος αποζημιώσεων είναι ο Ιούλιος με τον Ιούνιο να ακολουθεί. Αντιθέτως το μικρότερο ύψος εντοπίζεται τον Φεβρουάριο. Αναφορικά με το έτος 2012 διαπιστώνουμε το υψηλότερο κόστος στις αποζημιώσεις να είναι τον μήνα Μάιο ενώ ο Φεβρουάριος είναι και πάλι ο μήνας με το χαμηλότερο. Σχετικά με το έτος 2013 ο Σεπτέμβριος είναι ο μήνας που συγκεντρώνει τις περισσότερες αποζημιώσεις ενώ ο Νοέμβριος τις χαμηλότερες. Ένα στοιχείο που παρουσιάζει ενδιαφέρον είναι πως το υψηλότερο συνολικό κόστος αποζημιώσεων παρατηρείται το διάστημα από τον μήνα Μάιο έως και το Σεπτέμβριο και για τα τρία έτη. Είναι άξιο αναφοράς ο μήνας Απρίλιος και για τα τρία έτη καθώς το 2011 οι αποζημιώσεις είναι σχεδόν 5 εκ. ενώ τα έτη 2012 και 2013 πέφτουν σχεδόν στο μισό. Ένα επιπλέον στοιχείο που παρουσιάζει ενδιαφέρον είναι το αυξημένο συνολικό κόστος τους μήνες Δεκέμβριο και Ιανουάριο σε σχέση με τον Νοέμβριο και Φεβρουάριο. Λαμβάνοντας υπόψη όλα τα παραπάνω στοιχεία καταλήγουμε πως υπάρχει υψηλό κόστος ζημιών κατά τους θερινούς μήνες ενώ χαμηλότερο κατά τους χειμερινούς και δικαιολογημένα συμβαίνει. Την καλοκαιρινή περίοδο η κίνηση στους δρόμους είναι πάντα

αυξημένη και αυτό έχει σαν επίπτωση τα υψηλά ποσοστά τροχαίων ατυχημάτων. Σχετικά με την αύξηση του κόστους των αποζημιώσεων τους μήνες Δεκέμβριο και Ιανουάριο αξίζει να επισημάνουμε πως είναι περίοδος χειμερινών διακοπών και κατ' επέκταση έχουμε αυξημένη κυκλοφορία οχημάτων στους αυτοκινητόδρομους της χώρας. Αξιοσημείωτη είναι και η μεγάλη μείωση του συνολικού ύψους των αποζημιώσεων κατά τα έτη 2012 και 2013 σε σύγκριση με το 2011. Το έτος 2011 τα τροχαία στην χώρα είναι περισσότερα και κατά συνέπεια και οι αποζημιώσεις καθόσον πρόκειται για την χρονική περίοδο που αρχίζει στην Ελλάδα η οικονομική κρίση. Σχετικά με τα έτη 2012 και 2013 τα τροχαία και κατ' επέκταση οι αποζημιώσεις μειώθηκαν καθώς η χώρα βιώνει την δημοσιονομική ανισορροπία, αρχίζουν μέτρα λιτότητας και περικοπές με άμεσο αντίκτυπο και στα τροχαία καθώς ελαττώνεται η κίνηση στους δρόμους.

2.4 Μέσο κόστος ζημιών ανά περιφέρεια και ανά νομό

Από τα δεδομένα της ασφαλιστικής εταιρείας υπολογίσαμε το μέσο κόστος ζημιών ανά νομό και στη συνέχεια εντάξαμε τον κάθε νομό στην αντίστοιχη περιφέρεια όπου ανήκει. Αποτέλεσμα των παραπάνω ήταν το Διάγραμμα 2.2 για τις περιφέρειες και ο πίνακας που ακολουθεί για τον κάθε νομό. Παρατηρούμε στο διάγραμμα πως το μεγαλύτερο μέσο κόστος αποζημίωσης για την εταιρεία είναι στο νότιο και βόρειο Αιγαίο. Σύμφωνα με την Γενική Αστυνομική Διεύθυνση Νοτίου Αιγαίου σε δημοσίευμα αναφέρει πως στον νομό Κυκλάδων και Δωδεκανήσου σημειώνονται τα περισσότερα τροχαία ατυχήματα τους θερινούς μήνες. Στο δημοσίευμα τους αυτό αναφέρουν τις κύριες αιτίες των τροχαίων καθώς και τους παράγοντες που επιτείνουν τις συνέπειες τους. Συγκεκριμένα αναφέρονται οι κάτωθι αιτίες

- Υπερβολική ταχύτητα
- Απόσπαση προσοχής του οδηγού
- Παραβίαση της προτεραιότητας
- Οδήγηση υπό την επήρεια αλκοόλ
- Κίνηση στο αντίθετο ρεύμα
- Οδήγηση χωρίς χρήση κράνους και προστατευτικής ζώνης

Στο διάγραμμα φαίνεται και η περιφέρεια της Κρήτης να έχει υψηλό κόστος ζημιών. Είναι απόλυτα λογικό να θεωρείται ζημιόγωνα περιφέρεια για αρκετούς λόγους. Αρχικά θεωρείται από τους πιο δημοφιλείς προορισμούς διακοπών. Άνθρωποι από όλη την Ελλάδα, αλλά και από τον υπόλοιπο κόσμο, επιλέγουν την Κρήτη την Κρήτη καθ' όλη την διάρκεια του

χρόνου. Όπως θα δούμε παρακάτω ο νομός Λασιθίου έχει το μεγαλύτερο ποσοστό τροχαίων ατυχημάτων.

Από τους πιο σημαντικούς λόγους είναι το κακό οδικό δίκτυο . Πρόκειται για μια περιοχή με πολλά οροπέδια, γεγονός που δυσκολεύει την οδική κυκλοφορία. Αξίζει επίσης να σχολιασθεί πως το μικρότερο μέσο κόστος αποζημίωσης είναι στην περιφέρεια της Μακεδονίας. Ο σημαντικότερος παράγοντας που συμβάλει σε αυτό είναι η ολοκλήρωση του έργου της 'Εγνατίας οδού ' που καλύπτει τις περιφέρειες κεντρικής ,δυτικής και ανατολικής Μακεδονίας και Θράκης.



Διάγραμμα 2.2

Για να υπολογίσουμε τις τιμές του παραπάνω πίνακα χρησιμοποιήσαμε το σύνολο των ασφαλισμένων οχημάτων ανά περιφέρεια καθώς και το πλήθος των ζημιών που είχαμε στην εκάστοτε περιοχή.

	ΠΟΣΟΣΤΟ ΟΧΗΜΑΤΩΝ ΣΤΟΛΟΥ	ΠΟΣΟΣΤΟ ΖΗΜΙΩΝ
ΑΝΑΤΟΛΙΚΗ ΜΑΚΕΔΟΝΙΑ ΘΡΑΚΗ	9%	8%
ΑΤΤΙΚΗ	13%	17%
ΒΟΡΕΙΟ ΑΙΓΑΙΟ	7%	5%
ΔΥΤΙΚΗ ΕΛΛΑΔΑ	8%	9%
ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	3%	3%
ΗΠΕΙΡΟΣ	4%	4%
ΘΕΣΣΑΛΙΑ	8%	7%
ΙΟΝΙΟΙ ΝΗΣΟΙ	4%	3%
ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	12%	14%
ΚΡΗΤΗ	4%	4%
ΝΟΤΙΟ ΑΙΓΑΙΟ	10%	8%
ΠΕΛΟΠΟΝΗΣΟΣ	12%	11%
ΣΤΕΡΕΑ ΕΛΛΑΔΑ	7%	7%

Στον παραπάνω πίνακα παρατηρούμε το πώς απεικονίζεται το ποσοστό του στόλου των οχημάτων της ασφαλιστικής εταιρείας σε σχέση με το ποσοστό ζημιών αυτών για κάθε περιφέρεια. Συγκεκριμένα το ποσοστό του στόλου οχημάτων που αναγράφεται στον πίνακα σε σχέση με το ποσοστό ζημιών βρίσκονται στην ίδια κλίμακα. Εξαιρέση αποτελεί η περιφέρεια της Αττικής και της κεντρικής Μακεδονίας όπου το ποσοστό ζημιών είναι κατά 2% με 4% μεγαλύτερο σε σχέση με το στόλο οχημάτων. Αυτό συμβαίνει διότι το πλήθος οχημάτων που είναι ασφαλισμένα σε αυτή την περιφέρεια είναι μεγαλύτερο συγκριτικά με τις υπόλοιπες περιφέρειες της χώρας. Αυτό έχει ως αποτέλεσμα να έχουμε και τις περισσότερες ζημιές στην περιοχή αυτή.

Στη συνέχεια παρατίθενται μετά από υπολογισμούς στοιχεία για τον μέσο όρο του κόστους ζημιάς ανά νομό. Ο πίνακας με τα αποτελέσματα είναι διαθέσιμος στο παράρτημα 1.

Από τον πίνακα προκύπτει πως το μεγαλύτερο μέσο κόστος ζημιάς είναι στους νομούς Λασιθίου(3.790,95), Αιτωλοακαρνανίας(3.416,54), Αρκαδίας(3.321,42), Ξάνθης(3.074,76) και Τρικάλων(3.658,58) ενώ το χαμηλότερο στην Καρδίτσα(1.052,62). Για τον νομό Λασιθίου, όπως αναφέρθηκε και πιο πάνω, η μορφολογική κατανομή του νομού ευνοεί την αύξηση των τροχαίων ατυχημάτων. Αναφορικά με τον νομό Αιτωλοακαρνανίας το μέσο κόστος αποζημίωσης είναι αρκετά μεγάλο καθώς τα τροχαία λόγω του κακού οδικού δικτύου της περιοχής είναι πάρα πολλά. Το οδικό δίκτυο της περιοχής είναι αρκετά απαρχαιωμένο και σε κάποια σημεία πολύ κακώς συντηρημένο. Είναι μια από τις περιοχές στην οποία έχει συμβεί θανατηφόρο ατύχημα ,γεγονός που οδήγησε την ασφαλιστική εταιρεία να δώσει πολύ μεγάλη αποζημίωση.

Σχετικά με τις περιοχές της Αρκαδίας, Ξάνθης και Τρικάλων μπορούμε να πούμε ότι τα τροχαία οφείλονται στην ορεινή μορφολογία των νομών αυτών. Πρόκειται για περιοχές με υψηλή επισκεψιμότητα. Είναι τουριστικά θέρετρα που επιλέγουν πολλά άτομα κατά τους χειμερινούς μήνες. Η ολισθηρότητα του δρόμου από τους πάγους αλλά και οι πολλές στροφές ευνοούν την ανάπτυξη ατυχημάτων.

2.5 Μέσο κόστος και πλήθος ζημιών ανά χρήση οχήματος

Σημαντικό για την ασφαλιστική εταιρεία είναι να μελετηθεί σε ποια χρήση αποδίδεται το μεγαλύτερο και το μικρότερο μέσο κόστος αποζημίωσης καθώς και από ποια χρήση προέρχονται οι περισσότερες ζημιές.

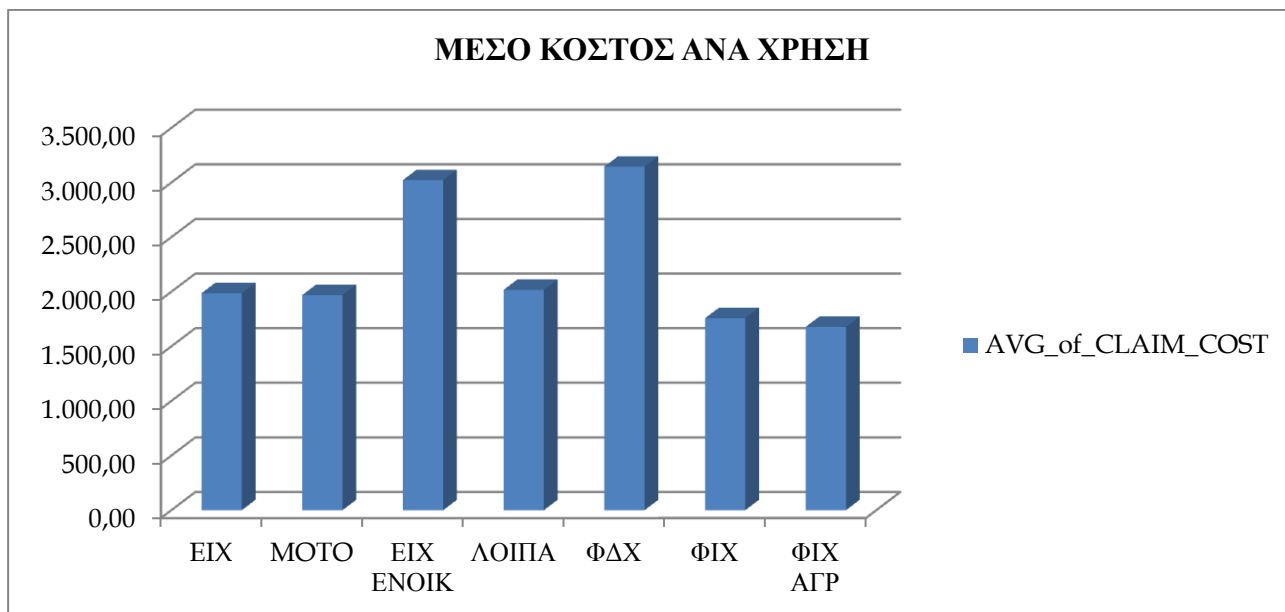
Το παρακάτω Διάγραμμα 2.3 απεικονίζει το μέσο κόστος ζημίας για κάθε χρήση οχήματος. Όπως παρατηρούμε τις υψηλότερες ζημιές κατά μέσο όρο τις έχουν τα φορτηγά δημοσίας χρήσεως και τα ενοικιαζόμενα οχήματα. Πρόκειται για δυο χρήσεις που ελλοχεύουν τον μεγαλύτερο κίνδυνο για την εταιρεία λόγω της μεγάλης έκθεσης τους στο οδικό δίκτυο της χώρας. Για τις υπόλοιπες χρήσεις παρατηρούνται μικρές διακυμάνσεις. Τα παραπάνω στοιχεία είναι χρήσιμα για την τιμολογιακή πολιτική μιας ασφαλιστικής επιχείρησης καθώς αναλογικά με το ύψος τους κινδύνου διαμορφώνεται το ασφάλιστρο.

Ενδεικτικά παρατίθεται στον επόμενο πίνακα το μέσο ασφάλιστρο που έχει το ίδιο όχημα με τα ίδια χαρακτηριστικά τιμολογώντας το είτε ως απλό επιβατικό είτε ως ενοικιαζόμενο.

Στη τέταρτη στήλη (Δ) του πίνακα δίνεται το ποσοστό αύξησης στην τιμή ενός οχήματος όταν τιμολογείται ως ενοικιαζόμενο.

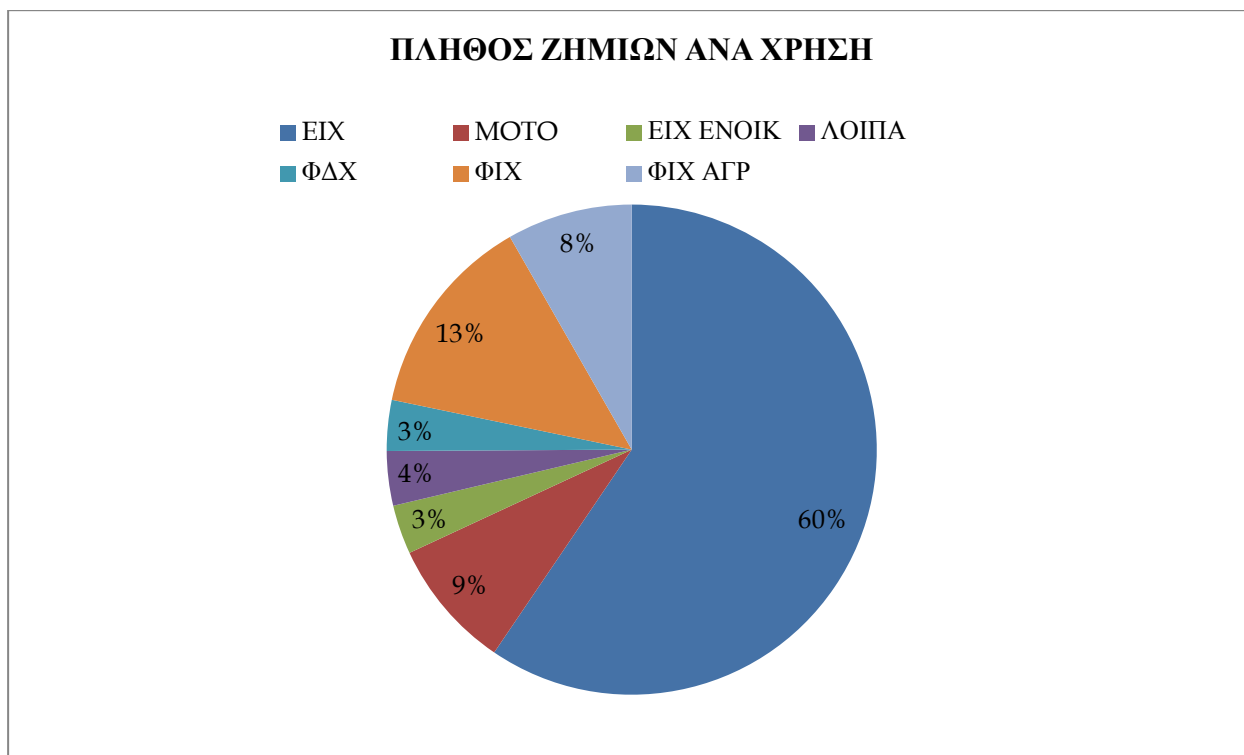
Μέσο ασφάλιστρο

ΕΤΟΣ	ΕΙΧ	ΕΙΧ ΕΝΟΙΚ	Δ.
2011	230	329	43,2%
2012	227	309	36,1%
2013	214	336	57,3%



Διάγραμμα 2.3

Στη συνέχεια παρατίθεται σε διάγραμμα πίτας το ποσοστό των ζημιών που αντιπροσωπεύει κάθε χρήση στο χαρτοφυλάκιο μας.



Διάγραμμα 2.4

Παρατηρούμε ότι το 60% του πλήθους των ζημιών του χαρτοφυλακίου προέρχεται από E.I.X., όπου είναι και το υψηλότερο ποσοστό. Το χαμηλότερο προέρχεται από EIX ENOIKIAZOMENA και είναι μόλις 3%.

	ΠΟΣΟΣΤΟ ΟΧΗΜΑΤΩΝ ΣΤΟΛΟΥ	ΠΟΣΟΣΤΟ ΖΗΜΙΩΝ
EIX	53%	60%
MOTO	21%	9%
EIX ENOIK.	3%	3%
ΛΟΙΠΑ	6%	4%
ΦΔΧ	1%	3%
ΦΙΧ	8%	13%
ΦΙΧ ΑΓΡ.	8%	8%

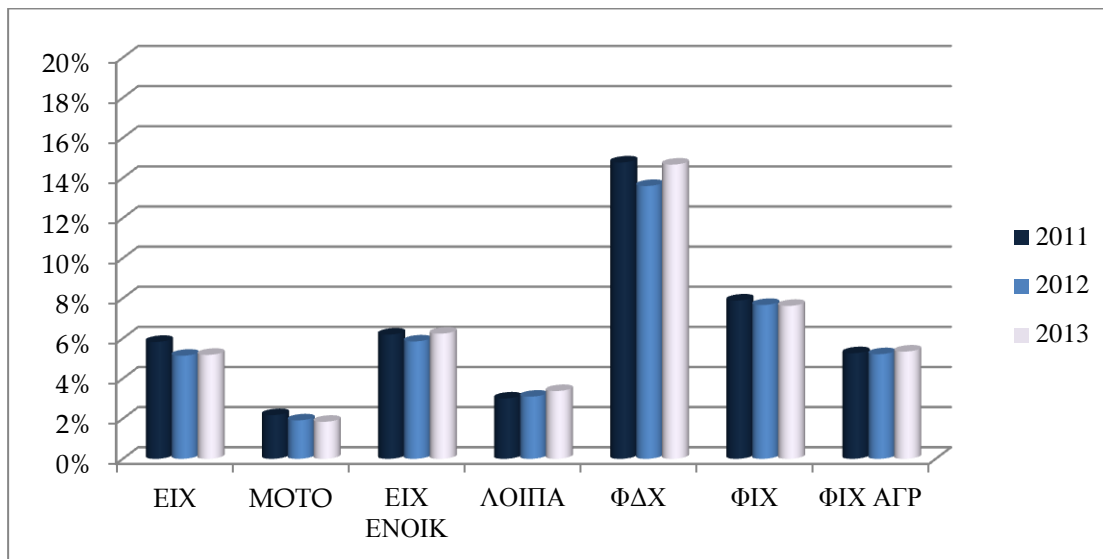
Ο παραπάνω πίνακας αποτελείται από δύο στήλες με ποσοστά. Στην πρώτη στήλη παρατηρούμε τι ποσοστό καταλαμβάνει κάθε χρήση σε όλο το χαρτοφυλάκιο. Συγκεκριμένα τα αυτοκίνητα καταλαμβάνουν το 53% όλου του στόλου που έχει ασφαλίσει η εταιρεία. Σημαντικό ποσοστό κατέχουν οι μοτοσυκλέτες σε όλο το χαρτοφυλάκιο (21%). Πολύ μικρό είναι το ποσοστό των φορτηγών δημοσίας χρήσης (1%). Στη δεύτερη στήλη έχουμε τα ποσοστά των ζημιών που κάθε χρήση οχήματος προκαλεί στην ασφαλιστική εταιρεία. Αξίζει να επισημάνουμε ότι τα EIX είναι τα περισσότερα ζημιογόνα οχήματα ,ποσοστό 60% ,καταλαμβάνει δηλαδή το μεγαλύτερο ποσοστό σε σχέση με τον μέσο όρο των ζημιών του χαρτοφυλακίου. Ακολουθούν τα ΦΙΧ ,ενώ πολύ μικρό ποσοστό ζημιών σε σχέση με το πλήθος τους εμφανίζουν οι μοτοσυκλέτες (MOTO). Στο σημείο αυτό θα μπορούσαμε να κάνουμε την εξής παρατήρηση ,πως ένας σημαντικός λόγος όπου τα ασφάλιστρα των MOTO είναι χαμηλά είναι ότι δεν είναι ζημιογόνα οχήματα για την εταιρεία .

2.6 Σχετική συχνότητα πλήθους ζημιών ανά χρήση και ανά περιφέρεια

Μείζονος σημασίας για την ασφαλιστική εταιρεία είναι το πλήθος των ζημιών που παρατηρούνται σε κάθε χρήση οχήματος καθώς και σε κάθε περιφέρεια .Υπολογίσαμε τις συχνότητες για κάθε έτος ξεχωριστά. Η σχετική συχνότητα αυτή είναι το πηλίκο της διαίρεσης του πλήθους των ζημιών προς τον αριθμό όλων των ασφαλισμένων κινδύνων.

Στο Διάγραμμα 2.5 βλέπουμε τις σχετικές συχνότητες των ζημιών που προέκυψαν στην ασφαλιστική εταιρεία για τα έτη 2011,2012,2013 ανά χρήση. Ως σχετικές συχνότητες ζημιών ορίζεται το ποσοστό των οχημάτων που παρουσίασαν μια τουλάχιστον απαίτηση μέσα σε ένα ημερολογιακό έτος. Παρατηρούμε ότι οι περισσότερες ζημιές ως ποσοστό επί του συνόλου

των ασφαλισμένων αυτού του τύπου προέρχονται από τα φορτηγά δημοσίας χρήσεως (ΦΔΧ) για όλα τα έτη ενώ οι μοτοσυκλέτες(MOTO) παρουσιάζουν το μικρότερο πλήθος ζημιών. Όπως φαίνεται από χρονιά σε χρονιά δεν παρατηρούμε σημαντικές διαφορές στη συχνότητα των ζημιών έτσι εύκολα θα μπορούσε κανείς να προβλέψει τις ζημιές της ακόλουθης χρονιάς για την κάθε χρήση. Η ασφαλιστική εταιρεία μπορεί να χρησιμοποιήσει όλες αυτές τις πληροφορίες και για την τιμολογιακή της πολιτική όπως είναι προφανές τα φορτηγά θα πρέπει να έχουν και υψηλότερο ασφάλιστρο συγκριτικά με τα επιβατικά και τις μοτοσυκλέτες.



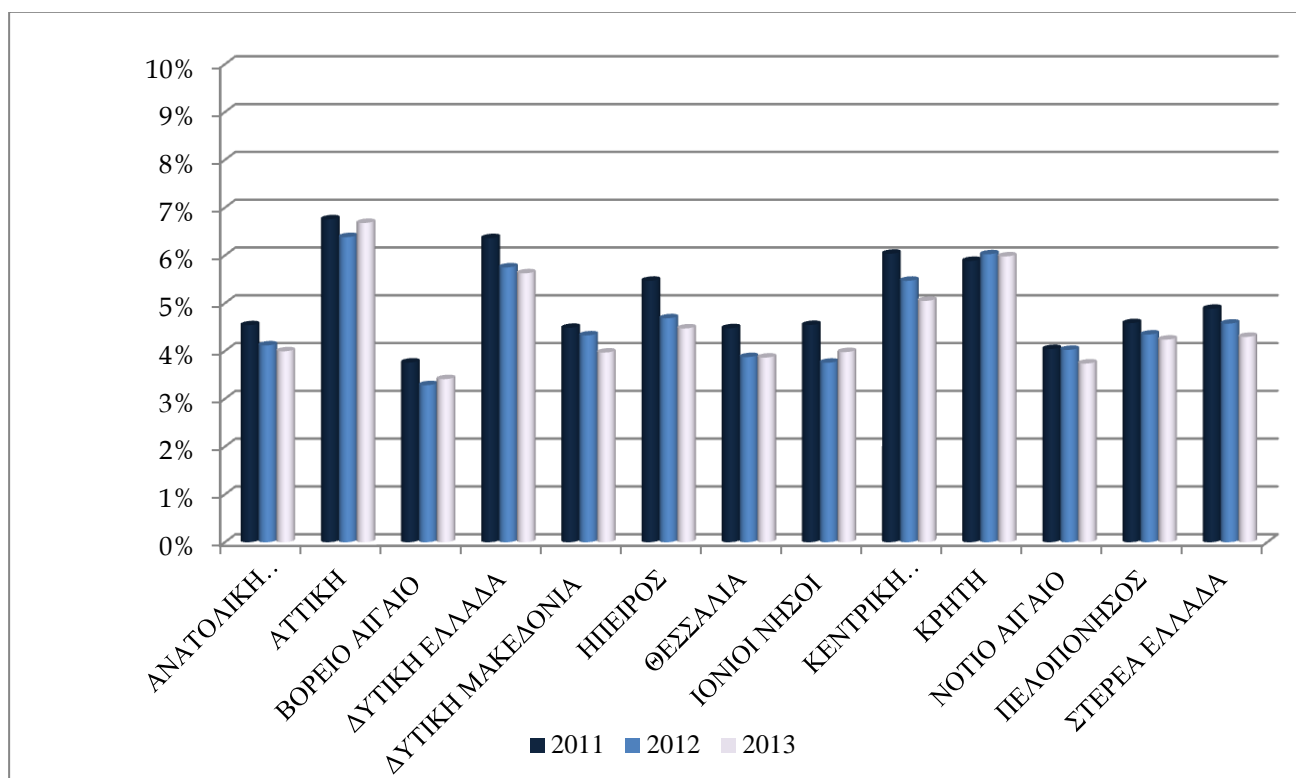
Διάγραμμα 2.5

Στο Διάγραμμα 2.6 παρουσιάζονται οι σχετικές συχνότητες του πλήθους των ζημιών ανά περιφέρεια. Συγκεκριμένα η μεγαλύτερη συγκέντρωση ζημιών βρίσκεται στις εξής περιφέρειες

- Αττική
- Δυτική Ελλάδα
- Κρήτη
- Κεντρική Μακεδονία

Για την Αττική είναι αρκετά λογικό να συγκεντρώνει πολλές ζημιές. Είναι ο νομός με τον περισσότερο πληθυσμό σε σχέση με την υπόλοιπη Ελλάδα . Συγκεκριμένα στην Αθήνα κινούνται τα πιο πολλά οχήματα, οι αποστάσεις που διανύουν είναι αρκετά μεγάλες οπότε είναι λογικό να έχουμε πολλά τροχαία ατυχήματα. Η κεντρική Μακεδονία είναι εξίσου ζημιογόνα περιφέρεια. Περιλαμβάνει την περιοχή της Θεσσαλονίκης η οποία είναι η δεύτερη μεγαλύτερη πόλη στην Ελλάδα και κατά συνέπεια παρουσιάζει μεγάλη συγκέντρωση οχημάτων. Η δυτική Ελλάδα ,όπως είχαμε αναφέρει προηγουμένως διαθέτει κακό οδικό δίκτυο, γεγονός που την χαρακτηρίζει ζημιογόνα περιοχή.

Αξιοσημείωτο είναι να θυμηθούμε το Διάγραμμα 2.2 του μέσου κόστους ζημιών ανά περιφέρεια. Κατά την συγκεκριμένη ανάλυση το βόρειο και νότιο Αιγαίο φαινόταν το πιο ζημιογόνο όμως συνδυαστικά με το παρακάτω διάγραμμα συμπεραίνουμε ότι το πλήθος των ζημιών είναι το μικρότερο σε σχέση με τις υπόλοιπες περιφέρειες. Συνεπώς σε συνδυαστική ανάλυση (ύψος –πλήθος ζημιών) το βόρειο και νότιο Αιγαίο δεν θα ήταν το πιο ζημιογόνο.

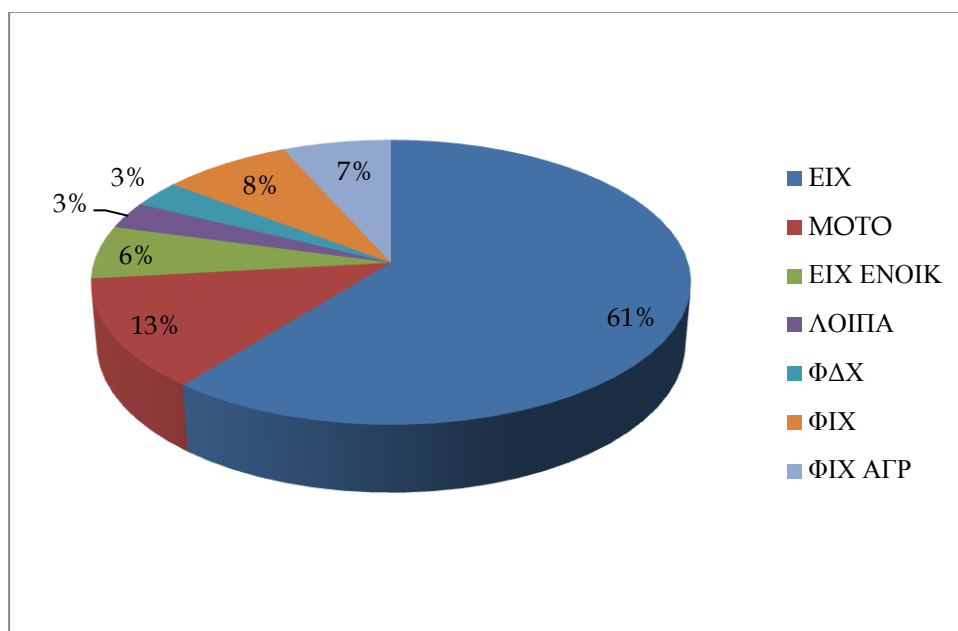


Διάγραμμα 2.6

2.7 Ποσοστά σωματικής βλάβης και θανατηφόρων ατυχημάτων ανά χρήση

Στο σημείο αυτό θα εξετάσουμε και θα κατηγοριοποιήσουμε τις ζημιές ανάλογα με τις επιπτώσεις τους. Τα τροχαία ατυχήματα δεν έχουν μόνο υλικές ζημιές. Συγκεκριμένα υπάρχουν ζημιές οι οποίες έχουν σαν αποτέλεσμα κάποια σωματική βλάβη στον οδηγό είτε στους επιβάτες. Στις σωματικές βλάβες ανήκουν οι ακρωτηριασμοί, ολική ή μερική ανικανότητα, τετραπληγία, διπληγία, εγκεφαλικές κακώσεις καθώς και άλλα. Επιπρόσθετα, η εταιρεία έχει να αντιμετωπίσει ατυχήματα που είναι θανατηφόρα. Οι αποζημιώσεις που καλείται να δώσει η εταιρεία σε αυτές τις περιπτώσεις είναι πολύ μεγάλες.

ΠΟΣΟΣΤΟ ΣΩΜΑΤΙΚΗΣ ΒΛΑΒΗΣ ΑΝΑ ΧΡΗΣΗ



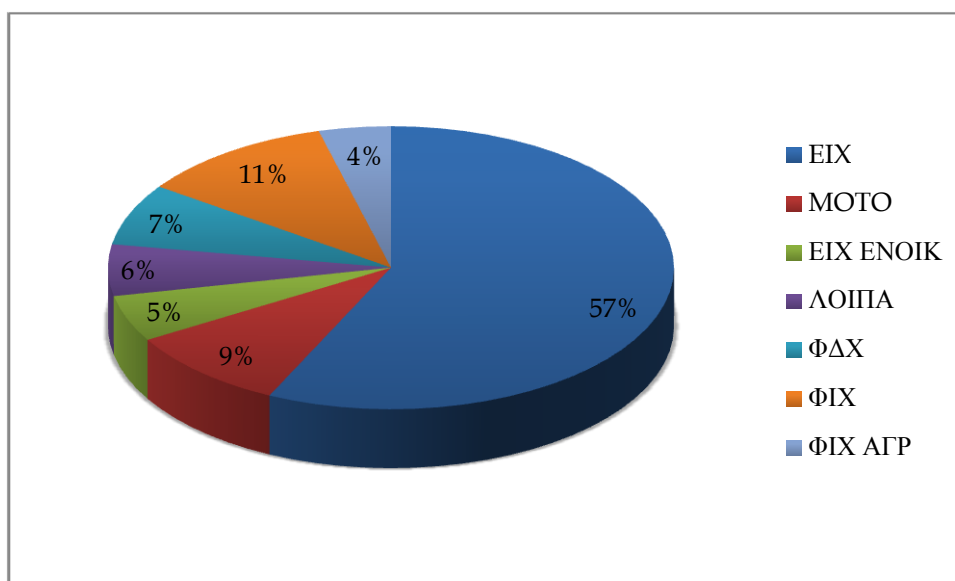
Διάγραμμα 2.7

Στην Διάγραμμα 2.7 παρουσιάζονται τα ποσοστά των ατυχημάτων που είχαν σωματική βλάβη καθώς και η χρήση από την οποία προέρχονται. Το μεγαλύτερο ποσοστό παρατηρείται στα ΕΙΧ όπου το 61% των ατυχημάτων με σωματική βλάβη προέρχεται από αυτή την χρήση αντίθετα τα ΦΔΧ έχουν το μικρότερο ποσοστό. Το μεγαλύτερο ποσοστό ατυχημάτων με σωματικές βλάβες είναι στην περιφέρεια της Αττικής (20%) ακολουθεί η κεντρική Μακεδονία και η Πελοπόννησος όπου τα ποσοστά κυμαίνονται στα ίδια επίπεδα (δηλαδή 10-11%). Οι περιοχές με το χαμηλότερο ποσοστό ατυχημάτων είναι η δυτική Μακεδονία, Ήπειρος και οι Ιόνιοι Νήσοι (2-3%).

	ΠΟΣΟΣΤΟ ΖΗΜΙΩΝ	ΠΟΣΟΣΤΟ ΖΗΜΙΩΝ(με Σ.Β.)
EIX	60%	61%
MOTO	9%	13%
EIX ΕΝΟΙΚ.	3%	6%
ΛΟΙΠΑ	4%	3%
ΦΔΧ	3%	3%
ΦΙΧ	13%	8%
ΦΙΧ ΑΓΡ.	8%	7%

Σύμφωνα με το παραπάνω πίνακα παρατηρούμε ότι στα EIX το ποσοστό ζημιών με σωματικές βλάβες σε σχέση με το συνολικό ποσοστό ζημιών με σωματικές βλάβες των υπόλοιπων χρήσεων είναι αισθητά μεγαλύτερο. Αυτό οφείλεται στο ότι το 60% των ζημιών είναι από EIX. Ακολουθούν τα MOTO ενώ στα ίδια επίπεδα εμφανίζονται τα ΦΔΧ με αντίστοιχα ποσοστά.

ΠΟΣΟΣΤΟ ΘΑΝΑΤΗΦΟΡΩΝ ΑΤΥΧΗΜΑΤΩΝ ΑΝΑ ΧΡΗΣΗ



Διάγραμμα 2.8

Στο Διάγραμμα 2.8 παρουσιάζονται τα ποσοστά των θανατηφόρων ατυχημάτων καθώς και η χρήση από την οποία προέρχονται. Τα θανατηφόρα ατυχήματα για τα τρία έτη που εξετάζουμε είναι στο σύνολο τους 170.

Όπως βλέπουμε και από το διάγραμμα της πίτας τα περισσότερα προέρχονται από την χρήση ΕΙΧ (57%) ενώ τα λιγότερα από τα ΦΙΧ ΑΓΡ.(4%). Οι περιοχές στις οποίες έχουν σημειωθεί τα περισσότερα είναι η κεντρική Μακεδονία η Αττική και το νότιο Αιγαίο .Τα λιγότερα έχουν γίνει στην περιφέρεια της Ηπείρου

	ΠΟΣΟΣΤΟ ΖΗΜΙΩΝ	ΠΟΣΟΣΤΟ ΖΗΜΙΩΝ (ΘΑΝΑΤΗΦΟΡΕΣ)
ΕΙΧ	60%	57%
ΜΟΤΟ	9%	9%
ΕΙΧ ΕΝΟΙΚ.	3%	5%
ΛΟΙΠΑ	4%	6%
ΦΔΧ	3%	7%
ΦΙΧ	13%	11%
ΦΙΧ ΑΓΡ.	8%	4%

Σύμφωνα με τον παραπάνω πίνακα παρατηρούμε ότι στα ΕΙΧ το ποσοστό ζημιών που αφορά θανατηφόρα ατυχήματα είναι πολύ μεγαλύτερο σε σχέση με το συνολικό ποσοστό θανατηφόρων ατυχημάτων στις υπόλοιπες χρήσεις. Ακολουθούν τα ΦΙΧ. Παρατηρούμε ότι στα ΦΔΧ το ποσοστό των θανατηφόρων είναι μεγαλύτερο σε σχέση με το ποσοστό των ζημιών. Αυτό σημαίνει πως στα ΦΔΧ συμβαίνουν λίγες ζημιές, αλλά αυτές που συμβαίνουν είναι συχνά θανατηφόρες.

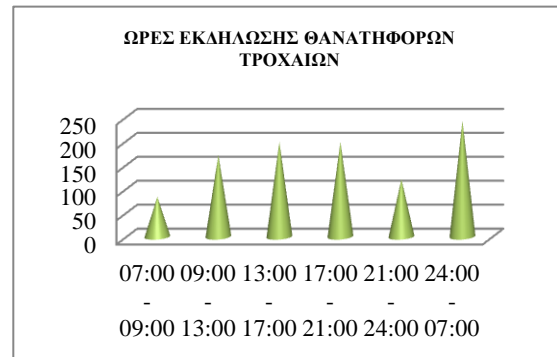
2.8 Αναλυτικά στοιχεία τριετίας θανατηφόρων σύμφωνα με την Ε.Α.

Στη συνέχεια του κεφαλαίου θα αναπτύξουμε κάποιες πληροφορίες γενικά για τα θανατηφόρα τροχαία ατυχήματα στην Ελλάδα όπως προέκυψαν από στοιχεία που συλλέξαμε από δημοσιεύσεις στο διαδίκτυο της ελληνικής αστυνομίας. Συγκεκριμένα για τα τρία υπό μελέτη έτη 2011,2012,2013 έχουμε στοιχεία για τα αίτια των θανατηφόρων ,για τα είδη, τις ώρες εκδήλωσης και την ηλικία. Ακολουθούν κάποια διαγράμματα για κάθε έτος ξεχωριστά.

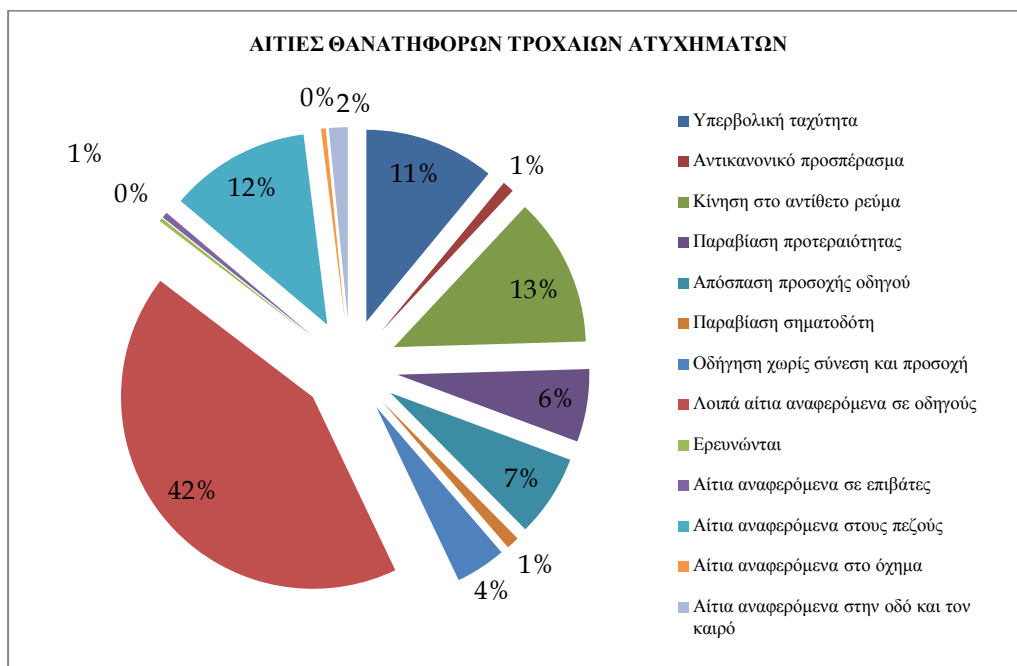
ΕΤΟΣ 2011



Διάγραμμα 2.9

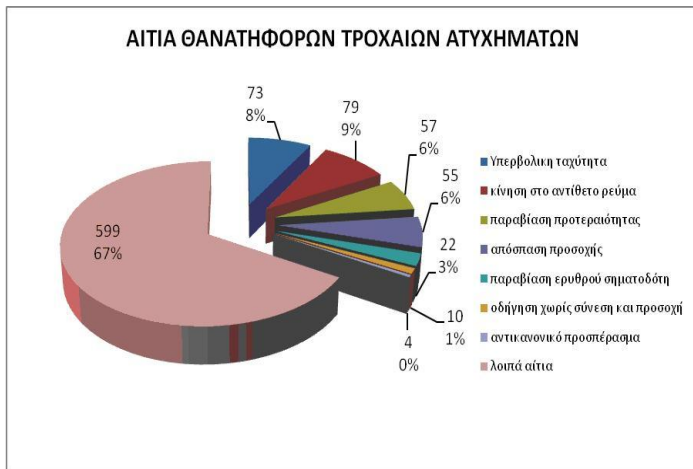


Διάγραμμα 2.10



Διάγραμμα 2.11

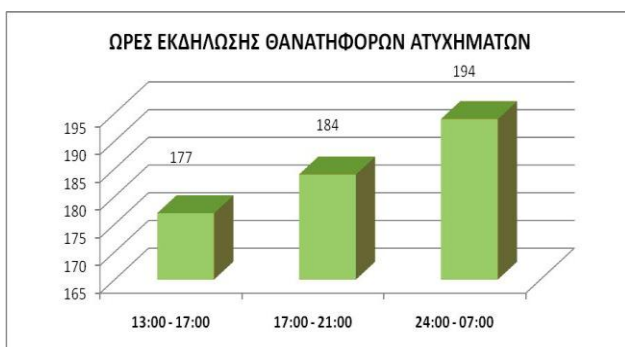
ΕΤΟΣ 2012



Διάγραμμα 2.12

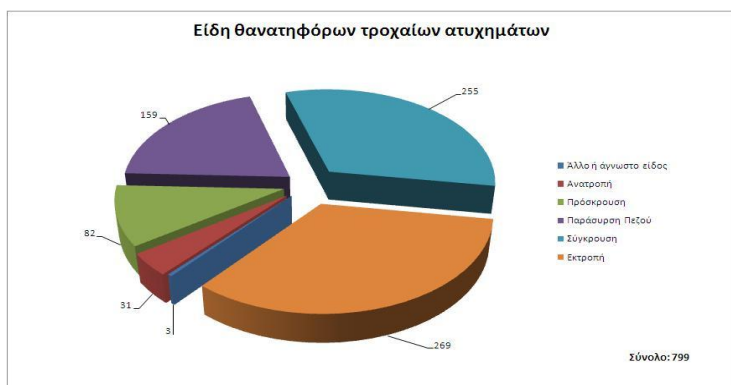


Διάγραμμα 2.13

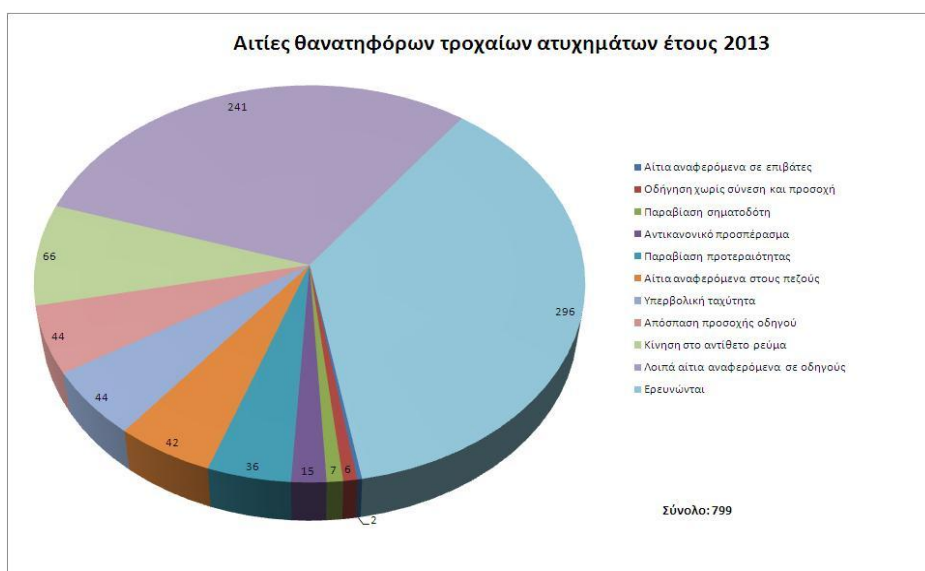


Διάγραμμα 2.14

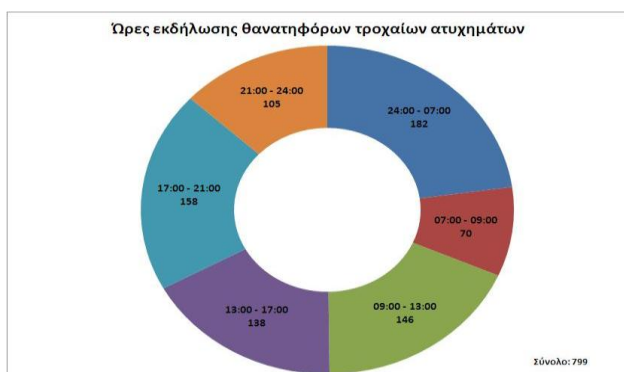
ΕΤΟΣ 2013



Διάγραμμα 2.15



Διάγραμμα 2.16



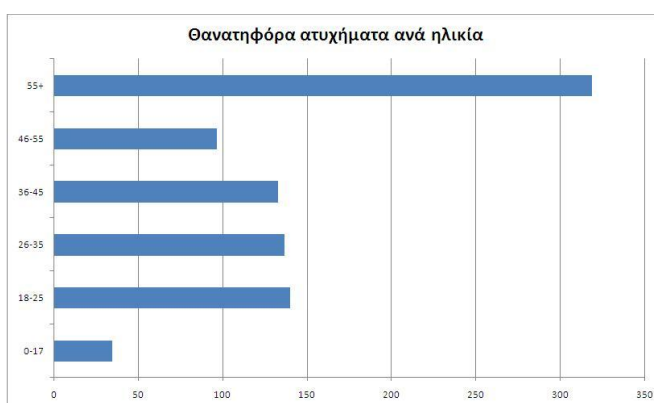
Διάγραμμα 2.17

Τα τροχαία ατυχήματα σήμερα είναι υπεύθυνα για το 40% των θανάτων στις περισσότερες ανεπτυγμένες χώρες του κόσμου. Θεωρείται η κυριότερη αιτία πρόωρων θανάτων, ανικανότητας και χαμένων χρόνων υγιούς ζωής. Ως μέτρο σύγκρισης θα ήταν χρήσιμο να αναφερθεί ότι στις ίδιες χώρες αποτελούν την τρίτη αιτία θνησιμότητας και επηρεάζουν τους δείκτες υγείας περισσότερο από τα καρδιαγγειακά νοσήματα και τις κακοήθεις νεοπλασίες. Παρατηρείται ότι το 50% των θανάτων μετά από οδικό ατύχημα λαμβάνει χώρα στον τόπο της σύγκρουσης και το υπόλοιπο κατά τη μετάβαση στο νοσοκομείο ή στο χώρο αυτού.

Συγκρίνοντας όλους τους πίνακες το σύνολο των συγκρούσεων οφείλεται σε λάθη ή αμέλεια των οδηγών. Η υπερβολική ταχύτητα και η κίνηση στο αντίθετο ρεύμα συνδέονται μεταξύ τους και είναι οι κυριότερες καταγεγραμμένες αιτίες των ατυχημάτων σε ποσοστό τουλάχιστον 36% το 2011, 17% το 2012, το οποίο το 2013 είναι σημαντικά μειωμένο (13,7%). Παρόμοια ποσοστά παρουσιάζει η παραβίαση του Κ.Ο.Κ. στο σύνολό του (παραβίαση προτεραιότητας, παραβίαση ερυθρού σηματοδότη, απόσπαση προσοχής), με ποσοστό 14% το 2011, 15% το 2012 και 10% το 2013 (σημαντικά μειωμένο).

Επί του συνόλου των θανατηφόρων ατυχημάτων η εκτροπή οχημάτων επιφέρει τα περισσότερα θύματα και ακολουθεί η απλή σύγκρουση μεταξύ τους και η παράσυρση πεζών, που οφείλεται σε απόσπαση της προσοχής των οδηγών, σε μικρότερα ποσοστά. Οι θάνατοι μεταξύ των ετών '11, '12 και '13 έχουν μειωθεί σημαντικά.

Όσον αφορά τις ώρες πραγματοποίησης των ατυχημάτων, αν και το βράδυ φαίνεται ο αριθμός αυξημένος, συμβαίνουν καθ' όλη τη διάρκεια της ημέρας και μάλιστα με αυξημένη ένταση διότι οι ώρες πραγματοποίησης είναι λιγότερες.

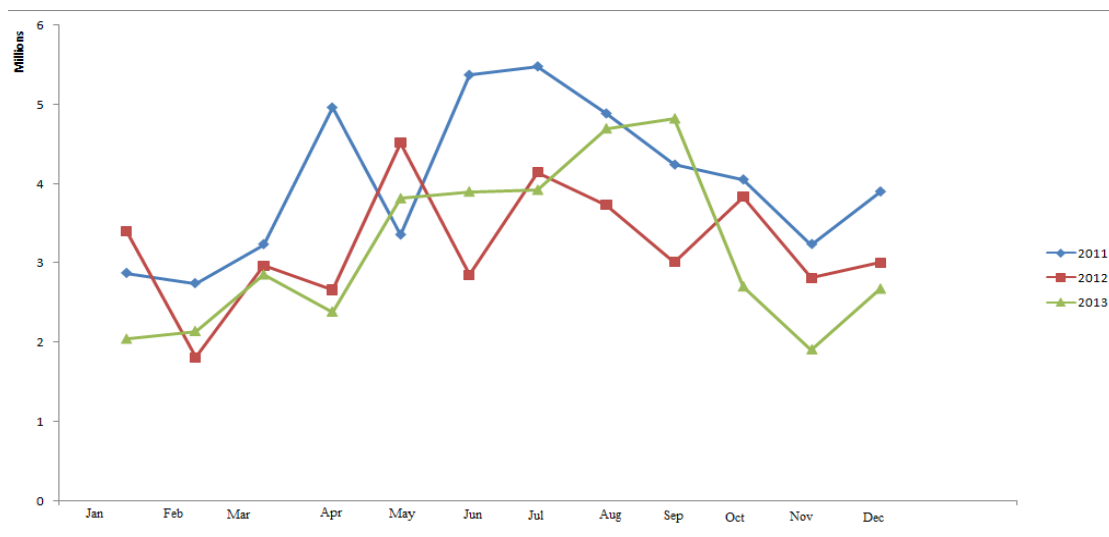


Διάγραμμα 2.18

Από το διάγραμμα 2.18 παρατηρούμε ότι οι ηλικίες που διατρέχουν μεγαλύτερο κίνδυνο να εμπλακούν σε θανατηφόρο ατύχημα είναι 18-35 και 55+ ετών. Σε αντίθεση με ό,τι πιστεύεται οι γυναίκες οδηγοί και κυρίως σε μεγάλες ηλικίες είναι πιο επιρρεπείς σε πρόκληση ατυχήματος.

Εν κατακλείδι τα τροχαία ατυχήματα αποτελούν σε όλο τον κόσμο σοβαρότατο πρόβλημα δημόσιας υγείας. Κάθε μέρα στην Ευρωπαϊκή Ένωση τα οδικά τροχαία ατυχήματα γίνονται η αιτία για το θάνατο σχεδόν 120 Ευρωπαίων πολιτών. Δυστυχώς η χώρα μας παρουσιάζει έναν από τους υψηλότερους δείκτες τροχαίων ατυχημάτων και θανάτων στην Ευρώπη που δικαιολογείται κατά κύριο λόγο από την έλλειψη συστηματικών προσπαθειών πρόληψης.

2.9 Εξέλιξη των ζημιών για κάθε έτος ανά μήνα



Διάγραμμα 2.19

Στο Διάγραμμα 2.19 παρατηρούμε το μηνιαίο κόστος συνολικών ζημιών για κάθε έτος ξεχωριστά. Βλέπουμε πώς εξελίσσονται οι ζημιές κάθε μήνα. Και για τα 3 έτη παρατηρούμε πώς υπάρχει η ίδια κινητικότητα σε σχέση με το μέσο μέγεθος ανά μήνα. Η πορεία του κόστους κινείται στα ίδια επίπεδα. Το κόστος φαίνεται να αυξάνεται στα μέσα του έτους,

και συγκεκριμένα τους καλοκαιρινούς μήνες. Τους τελευταίους μήνες του έτους μειώνεται και πάλι σταδιακά.

Είναι γεγονός πως ο δείκτης επικινδυνότητας το καλοκαίρι βρίσκεται σε πολύ υψηλό επίπεδο. Αυτό οφείλεται στο μεγάλο αριθμό ενοικιαζόμενων οχημάτων κατά την θερινή περίοδο.

Στα νησιά της χώρας και κυρίως του Αιγαίου διαπιστώσαμε πως ο αριθμός ατυχημάτων είναι αυξημένος. Το κακώς συντηρημένο οδικό δίκτυο καθώς και η οδήγηση υπό την επήρεια αλκοόλ είναι δυο από τους κυριότερους λόγους πρόκλησης ατυχημάτων. Επιπρόσθετα η οδήγηση σε υψηλές ταχύτητες χωρίς την απαραίτητη προστασία (ζώνη, κράνος) καθώς και ο μεγάλος αριθμός μοτοσικλετιστών οδηγούν στην αύξηση των ατυχημάτων κατά τους θερινούς μήνες.

Τέλος οδηγοί από διάφορες εθνικότητες με διαφορετική οδική κουλτούρα από αυτή του δυτικού κόσμου (οδήγηση από αριστερό ρεύμα κυκλοφορίας, σήμανση κυρίως οριζόντια και κάθετη με άλλο τύπο χρωματισμού) είναι ένα ακόμα αίτιο πρόκλησης οδικών ατυχημάτων. (ιστοσελίδα Ατλαντίς «νέα Σαντορίνης», 22/7/2014)

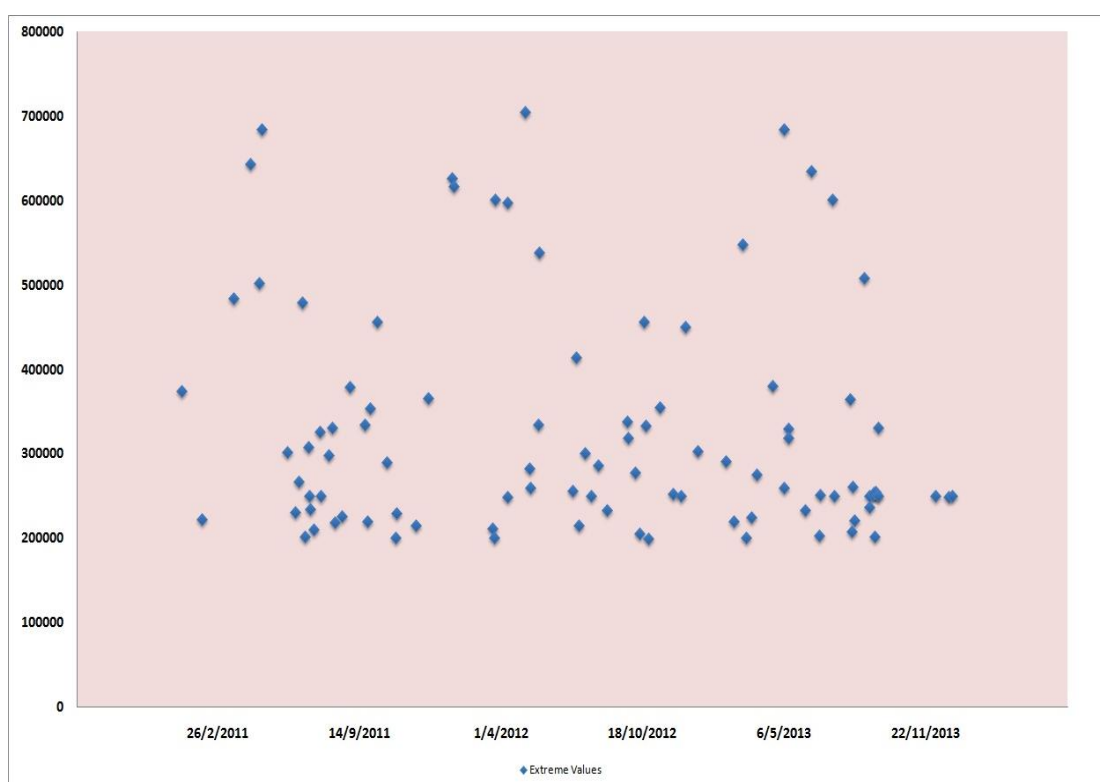
Παρατηρώντας στο διάγραμμα την εξέλιξη του κόστους των ζημιών διαπιστώνουμε στο 2013 μια σχετική μείωση σε σχέση με τα δυο προηγούμενα έτη (2011, 2012). Η μείωση αυτή μπορεί να δικαιολογηθεί σημαντικά από την οικονομική κρίση την οποία βιώνει η Ελλάδα τα τελευταία έτη.

Παράγοντες όπως οι μειωμένες διαδρομές με αυτοκίνητο, οι χαμηλότερες ταχύτητες ώστε να εξοικονομηθούν χρήματα από καύσιμα και να αποφευχθούν τυχόν ζημιόγωνα ατυχήματα, αλλά και οι λιγότερες μετακινήσεις εκτιμάται ότι έχουν επιδράσει θετικά στην μείωση του αριθμού των τροχαίων στη χώρα.

2.10 Ακραίες ζημιές

Εισαγωγή

Στην ενότητα που ακολουθεί θα γίνει μια ανάλυση για τις ακραίες ζημιές του χαρτοφυλακίου μας. Ως ακραίες ζημιές θεωρούνται οι αποζημιώσεις που καλείται να δώσει η εταιρεία στους ασφαλισμένους όταν υπερβαίνουν κάποιο συγκεκριμένο ποσό(200.000€). Οι ζημιές αυτές συνήθως εμπεριέχουν είτε θάνατο είτε σωματικές βλάβες. Εξετάσαμε τις ζημιές αυτές ανάλογα με το χρόνο που εμφανίστηκαν, τον τύπο του οχήματος και την περιοχή. Στη συνέχεια ακολουθούν κάποια γραφήματα για τις ακραίες τιμές.



Διάγραμμα 2.20

Στο παραπάνω Διάγραμμα 2.20 παρατηρούμε πώς κατανέμονται οι ακραίες ζημιές στο χαρτοφυλάκιο μας. Στη συγκεκριμένη εταιρεία από την οποία προήλθαν τα δεδομένα θεωρούνται ακραίες τιμές τις ζημιές που είναι πάνω από 200.000€. Παρατηρούμε ότι τουλάχιστον το 75% των ακραίων τιμών είναι από 200.000€ έως 400.000€. Όπως φαίνεται και από το γράφημα η μεγαλύτερη ζημία είναι στις 700.000€ περίπου.

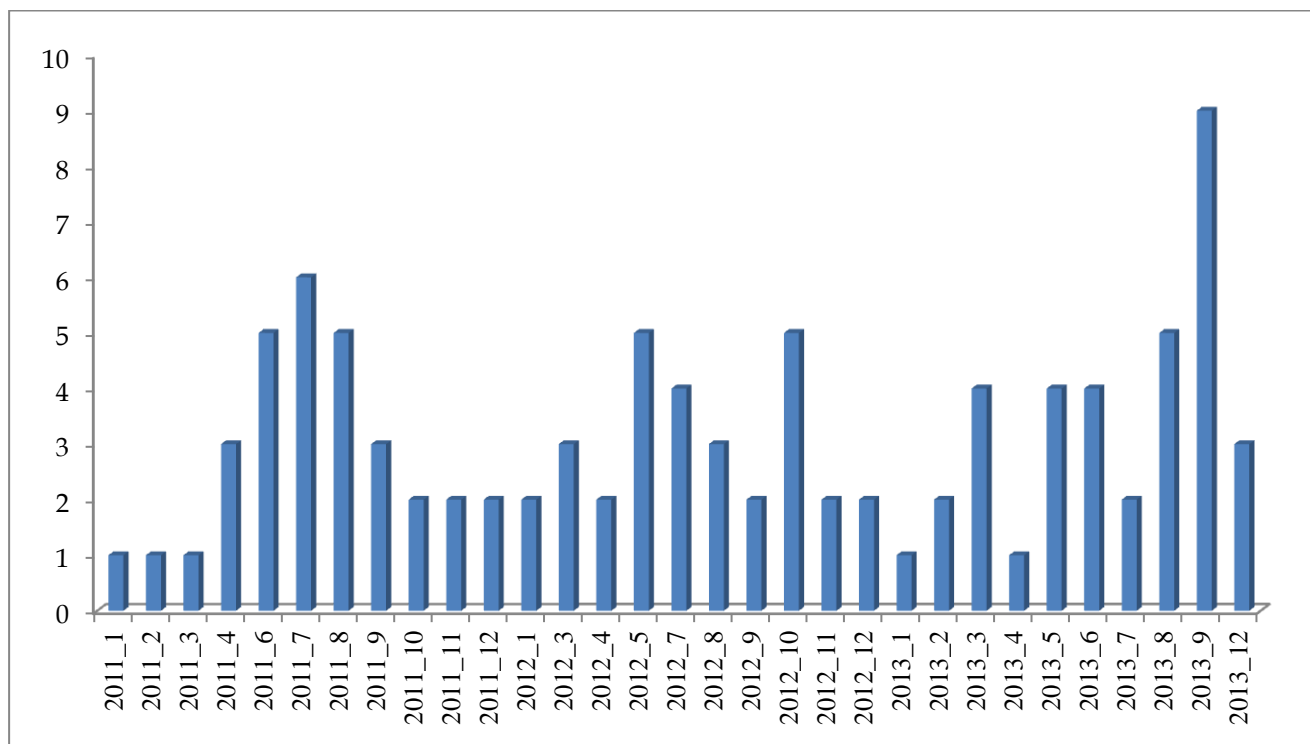
Οι ακραίες ζημιές του χαρτοφυλακίου είναι συνολικά 96 από τις οποίες οι 78 ήταν θανατηφόρες ενώ οι υπόλοιπες στην πλειοψηφία τους εμπεριέχουν σωματική βλάβη.

Σύμφωνα με τα στοιχεία που επεξεργαζόμαστε το 50% των ακραίων ζημιών προέρχονται από ΕΠΙΒΑΤΙΚΑ ενώ ακολουθούν οι ΜΟΤΟΣΥΚΛΕΤΕΣ και τα ΦΟΡΤΗΓΑ.

Αξίζει να αναφερθεί πως οι περισσότερες από τις υψηλές αποζημιώσεις που η εταιρεία έχει να αποδώσει εμφανίζονται στην Αττική, στο νότιο Αιγαίο και τέλος στην δυτική Ελλάδα.

Αντίθετα, στα Ιόνια Νησιά και στην Ήπειρο το ποσοστό είναι ελάχιστο. Ιδιαίτερο ενδιαφέρον όμως έχει να επικεντρωθεί κανείς στις ζημιές πάνω από 500.000€. Αυτές στο σύνολο τους είναι 14 και είναι διασκορπισμένες στις εξής περιφέρειες : δυτική Ελλάδα, Αττική, ανατολική και κεντρική Μακεδονία ,Θεσσαλία, Κρήτη, Πελοπόννησος , Στερεά Ελλάδα και Βόρειο Αιγαίο.Η μεγαλύτερη αποζημίωση που καλείται να δώσει η ασφαλιστική εταιρεία είναι ύψους 704.736€. Το ατύχημα έγινε το 2012 στην δυτική Ελλάδα και το όχημα το οποίο προκάλεσε το ατύχημα ήταν χρήση ΤΑΞΙ(ανήκει στην κατηγορία των ΛΟΙΠΩΝ ΧΡΗΣΕΩΝ).

Ιστόγραμμα πλήθους ακραίων ζημιών



Διάγραμμα 2.21

Στο παραπάνω ιστόγραμμα παρατηρούμε το πλήθος των ακραίων ζημιών που υπάρχουν στο χαρτοφυλάκιο μας για τα τρία έτη που εξετάζουμε (2011,2012,2013). Αρχικά μπορούμε να πούμε πως υπάρχουν αυξομειώσεις στο πλήθος των ακραίων ζημιών. Ενώ ο Ιανουάριος του 2011 ξεκινά με μικρό πλήθος βλέπουμε στη συνέχεια να αυξάνεται και έπειτα να φθίνει πάλι με κορύφωση τα τέλη του 2013. Παρατηρούμε πως τον Σεπτέμβριο του 2013 είχαμε τις περισσότερες ακραίες ζημιές συνεπώς και τις μεγαλύτερες αποζημιώσεις . Το γεγονός αυτό είναι εξαιρετικά ζημιογόνο για την ασφαλιστική εταιρεία εφόσον το κόστος καθεμίας ανέρχεται κατά μέσο όρο στα 300.000ευρω. Για το λόγο αυτό θα πρέπει η εταιρεία να έχει αποθέματα επαρκή για τέτοιες ακραίες περιπτώσεις.

Τέλος αξίζει να αναφέρουμε πως με βάση τα γραφήματα που κατασκευάσαμε για τις ακραίες ζημιές δεν παρατηρούμε να υπάρχει περιοδικότητα. Συγκεκριμένα δεν εμφανίζονται οι ίδιες ζημιές την ίδια χρονική περίοδο για κάθε έτος.

ΚΕΦΑΛΑΙΟ 3

Προσαρμογή κατανομών πιθανότητας για το πλήθος και το μέγεθος των ζημιών

3.1 Εισαγωγή

Στο παρόν κεφάλαιο θα εφαρμοστούν στο χαρτοφυλάκιο ασφάλισης των οχημάτων της ασφαλιστικής εταιρείας οι μέθοδοι εκτίμησης των κατανομών πιθανότητας. Το πρόγραμμα που πρόκειται να χρησιμοποιηθεί για την στατιστική επεξεργασία των δεδομένων είναι η R (γλώσσα προγραμματισμού) και το EasyFit 5.6 Professional.

Στη μελέτη που θα ακολουθήσει θα γίνει προσπάθεια να διερευνηθούν τα δεδομένα και να προκύψουν σημαντικά συμπεράσματα για τις πιθανοτικές κατανομές που ακολουθούν τόσο το μέγεθος όσο και το πλήθος των αποζημιώσεων.

Συγκεκριμένα για το πλήθος των ζημιών θα γίνει προσαρμογή ανά μήνα και ανά ημέρα για να διερευνηθεί ποια κατανομή προσεγγίζει καλύτερα αυτή την ποσότητα.

Στη συνέχεια θα εξετασθεί για το μέγεθος κάθε αποζημίωσης, τόσο μεμονωμένα όσο και ανά μήνα, ποια δηλαδή από τις γνωστές κατανομές το προσεγγίζει.

3.2 Μοντέλα για το πλήθος των ζημιών

Στην ενότητα αυτή θα γίνει ανάλυση για το πώς μπορεί να προσεγγιστεί από μια πιθανοθεωρητική κατανομή ο αριθμός των ζημιών.

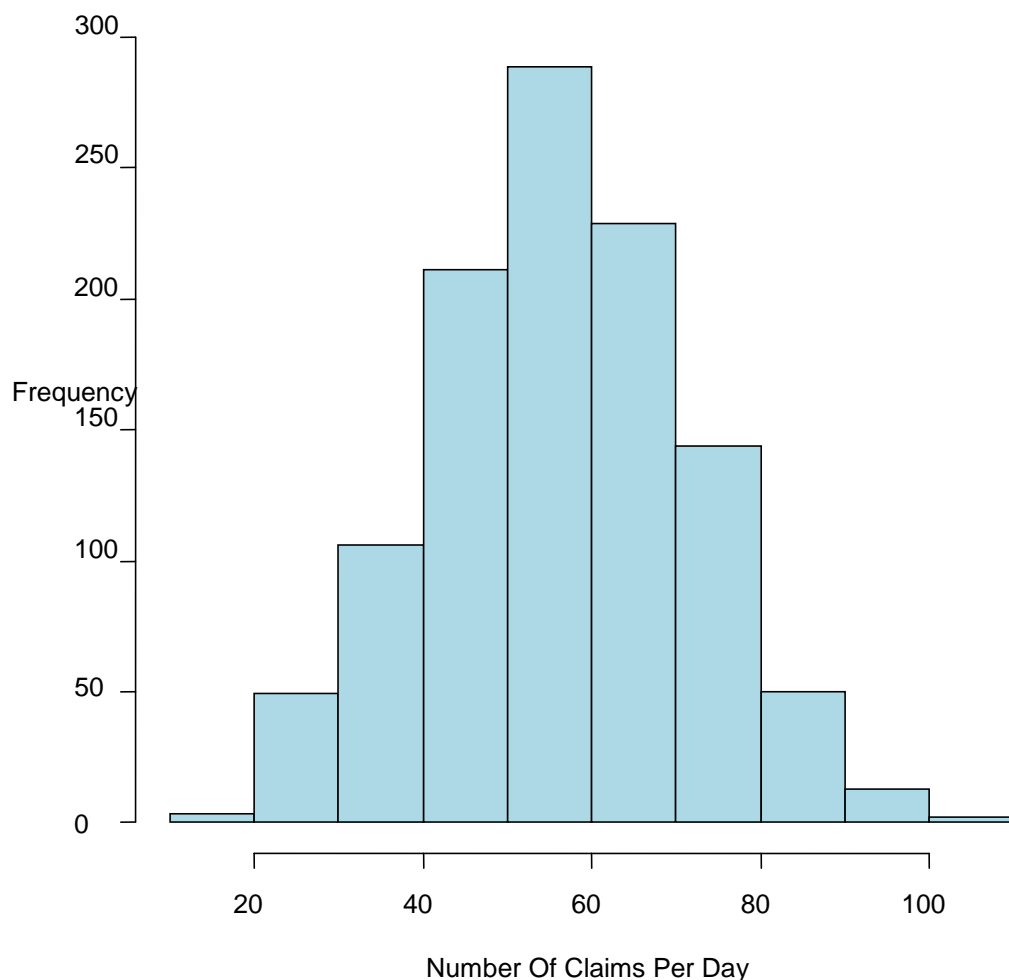
Για την μελέτη αυτή αρχικά κατασκευάσαμε μια μεταβλητή η οποία δίνει πόσες ζημιές έχουμε ανά ημέρα και πόσες ανά μήνα.

A) ANA ΗΜΕΡΑ

3.2.1 Προσαρμογή πλήθους ζημιών ανά ημέρα

Αρχικά θα ήταν χρήσιμο ένα ιστόγραμμα συχνοτήτων για να φανεί γραφικά αν η κατανομή του πλήθους των ζημιών ανά ημέρα θυμίζει κάποια γνωστή κατανομή.

ΙΣΤΟΓΡΑΜΜΑ ΠΛΗΘΟΥΣ ΑΠΟΖΗΜΙΩΣΕΩΝ ΑΝΑ ΗΜΕΡΑ



Στο παραπάνω ιστόγραμμα παρουσιάζεται η συχνότητα του αριθμού των αποζημιώσεων ανά ημέρα κατά την τριετία για την οποία εξετάζεται το χαρτοφυλάκιο.

Εκ πρώτης όψεως η κατανομή των δεδομένων φαίνεται να είναι αρκετά κοντά σ την κανονική κατανομή.

Κάθε όχημα για κάθε ημέρα είναι σαν μια δοκιμή Bernoulli δηλαδή εάν θα προκαλέσει ζημία ή όχι. Για το χαρτοφυλάκιο μας, εφόσον το εύρος των τιμών είναι πολύ μεγάλο η διωνομική κατανομή συγκλίνει στην κανονική .

Για μεγάλες τιμές του του n (μέγεθος δείγματος) η σύμπτωση των δυο κατανομών είναι σχεδόν τέλεια.

Αρκεί να εξετάσουμε την εγκυρότητα των παραπάνω ισχυρισμών.

Με τη χρήση του στατιστικού πακέτου R κάνουμε προσαρμογή στα δεδομένα μας την κανονική κατανομή . Υπολογίσαμε τις παραμέτρους της κατανομής και προέκυψαν τα ακόλουθα

ΠΑΡΑΜΕΤΡΟΙ ΚΑΝΟΝΙΚΗΣ ΚΑΤΑΝΟΜΗΣ		
		Τυπικό σφάλμα (st.error)
Μέσος (mean)	56,92701	0,4518811
Τυπική απόκλιση(standard deviation)	14,95993	0,3195282

P-P plot Q-Q plot

Τα P-P Plot και Q-Q plot (probability-probability plot και Quantile-Quantile plot) είναι δύο γραφήματα τα οποία μας βοηθούν να ελέγξουμε αν κάποια δεδομένα προέρχονται από κάποια συγκεκριμένη κατανομή (π.χ. κανονική). Τα γραφήματα αυτά βασίζονται στην ακόλουθη παρατήρηση:

Αν X_1, X_2, \dots, X_n είναι ένα τυχαίο δείγμα (ανεξ. τ.μ.) από μια (συνεχή) κατανομή με σ.κ. F τότε οι νέες τ.μ. $Y_1 = F(X_1), Y_2 = F(X_2), \dots, Y_n = F(X_n)$ είναι και αυτές ανεξάρτητες και ακολουθούν την ομοιόμορφη $U(0,1)$ κατανομή διότι

$$P(F(X) \leq x) = P(X \leq F^{-1}(x)) = F(F^{-1}(x)) = x, x \in [0,1].$$

Είναι εύκολο να αποδειχθεί ότι αν $Y_1, Y_2, \dots, Y_n \sim U(0,1)$ τότε κάθε μια από τις διατεταγμένες τ.μ. $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ ακολουθεί την κατανομή βήτα και συγκεκριμένα

$$Y_{(i)} \sim \text{Beta}(i, n-i+1) \quad \text{με} \quad E(Y_{(i)}) = i/(n+1).$$

Επομένως, για μεγάλο n θα ισχύει προσεγγιστικά ότι για $i = 1, 2, \dots, n$

$$Y_{(i)} = F(X_{(i)}) \approx \frac{i}{n+1} \quad \text{ή ισοδύναμα} \quad X_{(i)} \approx F^{-1}\left(\frac{i}{n+1}\right)$$

(διότι και $V(Y_{(i)}) \rightarrow_{n \rightarrow \infty} 0$)

Με άλλα λόγια, αν $X_i \sim F_0$ περιμένουμε ότι τα n σημεία του επιπέδου

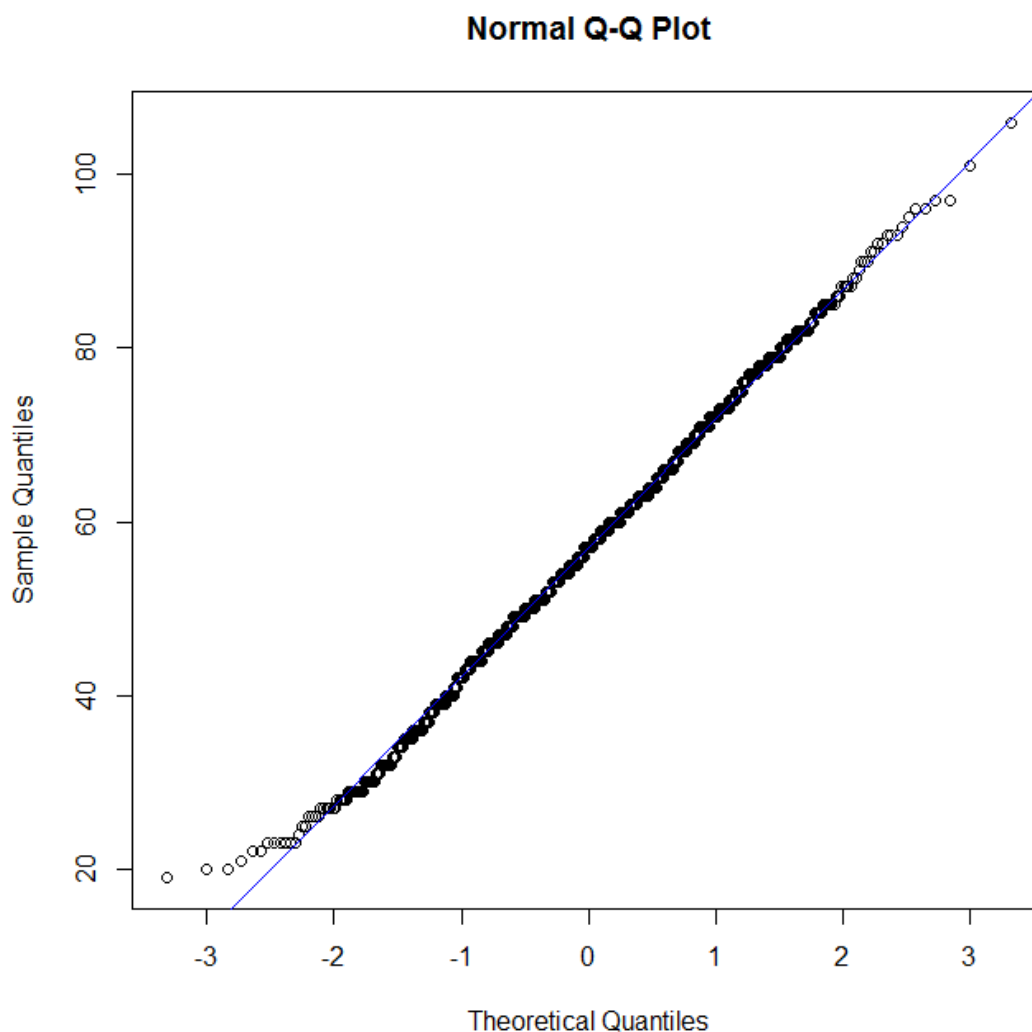
$$(F_0(X_{(i)}), \frac{i}{n+1}) , \quad i=1,2,\dots,n \quad (3.1)$$

ή ισοδύναμα

$$(X_{(i)}, F_0^{-1}(\frac{i}{n+1})) , \quad i=1,2,\dots,n \quad (3.2)$$

θα βρίσκονται «κοντά» στην διαγώνιο ($x = y$) που περνά από την αρχή των αξόνων. Το P-P plot ακριβώς είναι το γράφημα των πρώτων n σημείων, σχέση (3.1) (μαζί με τη διαγώνιο) ενώ το Q-Q Plot είναι το γράφημα των δεύτερων n σημείων, σχέση (3.2) (μαζί με τη διαγώνιο). Και στα δύο γραφήματα, αν τα σημεία βρίσκονται «κοντά» στη διαγώνιο (και «τυχαία» γύρω από αυτήν) τότε μπορεί να θεωρηθεί ότι τα δεδομένα προέρχονται από την F_0 . Ενδέχεται να μην είναι γνωστές όλες οι παράμετροι της κατανομής F_0 (π.χ. μπορεί να είναι κανονική με άγνωστο μ , σ^2). Σε αυτή την περίπτωση οι άγνωστες παράμετροι εκτιμώνται από τα δεδομένα. Ο έλεγχος μέσω αυτών των γραφημάτων είναι υποκειμενικός διότι δεν βασίζεται σε κάποιο στατιστικό κριτήριο που μας οδηγεί σε σωστή απόφαση π.χ. στο $1-\alpha$ % των περιπτώσεων. Συνήθως γίνεται για να πάρουμε μια πρώτη εποπτική εικόνα και για να δούμε αν υπάρχουν κάποιες έκτροπες, σε σχέση με τις αναμενόμενες υπό την F_0 , παρατηρήσεις. (Μ. Μπούτσικας, 2008)

Στη συνέχεια κατασκευάσαμε για τα δεδομένα μας ένα διάγραμμα Q-Q plot του πλήθους των αποζημιώσεων ανά ημέρα για την κανονική κατανομή

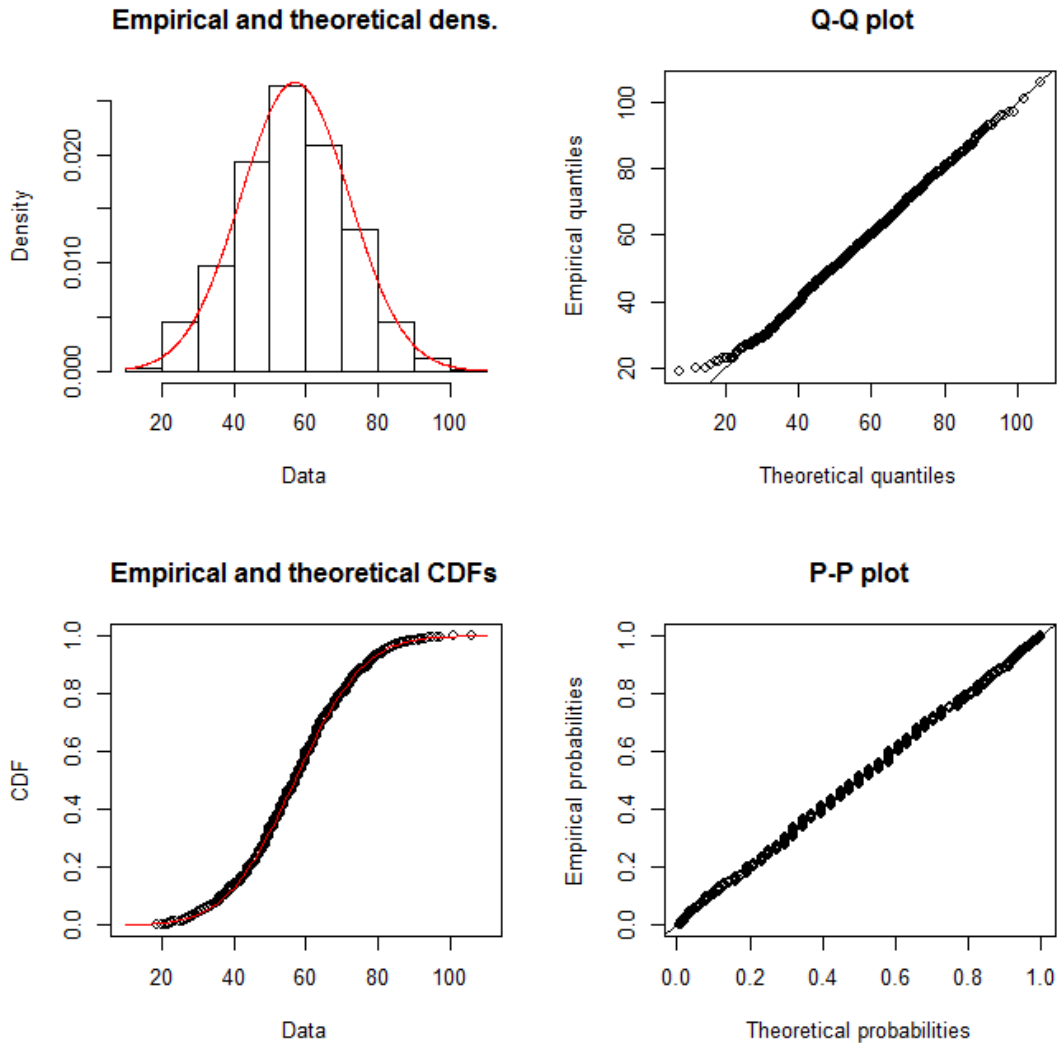


Σε μια ιδανική κατάσταση για την κανονικότητα όλα τα κυκλάκια θα βρίσκονται πάνω στην διχοτόμο της γωνίας των αξόνων.

Σε μια τέτοια περίπτωση μπορούμε να θεωρήσουμε ότι το δείγμα ακολουθεί προσεγγιστικά την κανονική κατανομή.

Στο συγκεκριμένο διάγραμμα δεν παρατηρούμε αποκλίσεις οπότε μπορούμε να πούμε ότι το δείγμα προέρχεται από την κανονική κατανομή. Όπως μπορεί και να διαπιστώσει κανείς το παραπάνω Q-Q plot επιδεικνύει άριστη προσαρμογή των δεδομένων μας στην κανονική κατανομή, εκτός από την αριστερή ουρά που αποκλίνει ελαφρώς από την διαγώνιο.

Στη συνέχεια παρατίθεται γραφήματα που προέκυψαν κατά την ανάλυση των δεδομένων στο στατιστικό πρόγραμμα R.



Συγκεκριμένα παρατηρούμε την προσαρμογή στα δεδομένα μας της εμπειρικής κατανομής το Q-Q plot και το P-P plot. Για τα διαγράμματα αυτά καταλήγουμε στο ίδιο συμπέρασμα ότι δηλαδή έχουμε άριστη προσαρμογή στη κανονική κατανομή.

Παρόλα αυτά υπάρχουν και στατιστικά τεστ που δεν στηρίζονται σε γραφικές μεθόδους όπως είναι το Kolmogorov-Smirnov και το Anderson –Darling GoF Test.

KOLMOGOROV-SMIRNOV

Το κριτήριο K-S χρησιμοποιείται για τον έλεγχο καλής προσαρμογής ενός δείγματος σε μία δεδομένη συνεχή κατανομή ($H_0: X_i \sim F_0$).

Το κριτήριο K-S βασίζεται στην διαφορά της εμπειρικής συνάρτησης κατανομής (που προέρχεται από το δείγμα) και της αναμενόμενης F_0 (υπό την H_0).

Πιο συγκεκριμένα, αν X_1, X_2, \dots, X_n είναι ένα τ.δ., η εμπειρική συνάρτηση κατανομής (ΕΣΚ) του δείγματος αυτού είναι

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) = \frac{\#\{X_i \leq x\}}{n},$$

(όπου $I(X_i \leq x) = 1$ ή 0 ανάλογα με το αν $X_i \leq x$ ή όχι) η οποία ως γνωστό αποτελεί εκτίμηση της συνάρτησης κατανομής των X_i διότι (από το νόμο των μεγάλων αριθμών,

θέτοντας $Y_i = I(X_i \leq x)$)

$$\begin{aligned} \hat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) = \frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow[n \rightarrow \infty]{} E(Y_1) = 0P(Y_1=0) + 1P(Y_1=1) = P(Y_1=1) \\ &= P(X_1 \leq x) = F(x) \text{ για κάθε } x \end{aligned}$$

Επομένως, υπό την H_0 , η ΕΣΚ θα πρέπει να είναι «κοντά» στην F_0 . Αντίθετα, αν δεν ισχύει η H_0 αναμένουμε σημαντική απόκλιση της ΕΣΚ από την F_0 .

Για να κατασκευάσουμε έναν έλεγχο με βάση αυτόν τον συλλογισμό, θα πρέπει να ορίσουμε μία «απόσταση» μεταξύ των δύο κατανομών (της ΕΣΚ και της F_0) και να απορρίπτουμε την H_0 όταν αυτή η απόσταση γίνεται «μεγάλη». Σχετικά έχουμε τον επόμενο ορισμό.

Ορισμός

Αν F, G είναι δύο συναρτήσεις κατανομής στον R , τότε η ποσότητα

$$D_K(F, G) = \sup_{x \in R} \{|F(x) - G(x)|\}$$

καλείται απόσταση Kolmogorov μεταξύ της F και της G .

Σύμφωνα με τα παραπάνω, θα απορρίπτουμε την $H_0: X_i \sim F_0$

όταν η στατιστική συνάρτηση

$$D_n = d_K(\hat{F}_n, F_0) = \sup_{x \in R} \{|\hat{F}_n(x) - F_0(x)|\},$$

λαμβάνει «ασυνήθιστα» μεγάλες τιμές, δηλαδή όταν $D_n > c$, για κάποια ‘κρίσιμη τιμή’ c .

Το κριτήριο αυτό είναι γνωστό ως κριτήριο Kolmogorov – Smirnov (και η στατιστική συνάρτηση D_n καλείται ελεγχοσυνάρτηση Kolmogorov – Smirnov). Προκειμένου να χρησιμοποιήσουμε το συγκεκριμένο κριτήριο θα πρέπει να προσδιορίσουμε την κατανομή της τ.μ. D_n κάτω από την H_0 έτσι ώστε να υπολογίσουμε το c (για δεδομένο επίπεδο οσφραντικότητας α) και το p -value ενός δείγματος.

Παραπάνω εξετάσαμε τον έλεγχο της υπόθεσης $H_0: X_i \sim F_0$ όπου η F_0 ήταν πλήρως καθορισμένη.

Συνηθέστερη όμως περίπτωση είναι να γνωρίζουμε την οικογένεια στην οποία ανήκει η F_0 με άγνωστες όμως παραμέτρους θ (π.χ. κανονική με άγνωστα μ, σ). Στην περίπτωση αυτή συνήθως εκτιμούμε τις παραμέτρους θ από τα δεδομένα και χρησιμοποιούμε την ίδια στατιστική συνάρτηση. (M. Μπούτσικας, 2008)

LILLIEFORS TEST

Ωστόσο, στην πράξη η μέση τιμή του πληθυσμού ή/και η διακύμανσή του δεν είναι συνήθως γνωστή. Μια καλή τακτική τότε θα ήταν να χρησιμοποιήσουμε στη θέση τους τις δειγματικές τιμές. Όμως σε αυτή την περίπτωση έχει παρατηρηθεί ότι ο έλεγχος γίνεται συντηρητικός (δηλαδή η ισχύς του ελέγχου είναι μικρότερη από την ισχύ του ελέγχου με γνωστές παραμέτρους). Μια καλή εναλλακτική λύση σε αυτή την περίπτωση είναι να εκτελεστεί ο έλεγχος του Lilliefors για τον έλεγχο της κανονικότητας των δεδομένων που αποτελεί μια παραλλαγή του ελέγχου Kolmogorov-Smirnov.

Ο έλεγχος Lilliefors (χρησιμοποιεί προσομοίωση Monte Carlo) που αποδίδει μια προσαρμοσμένη p -value.

Το Lilliefors είναι ένα γενικό test της εμπειρικής συνάρτησης κατανομής για την σύνθετη υπόθεση της κανονικότητας.

Η στατιστική συνάρτηση του test είναι η απόλυτη διαφορά μεταξύ της εμπειρικής και της υποθετικής συνάρτησης κατανομής.

Ορίζεται ως εξής

$$D = \max \{D^+, D^-\},$$

$$\text{Όπου } D^+ = \max_{i=1, \dots, n} \{i/n - p_{(i)}\} = \max_{i=1, \dots, n} \{p_{(i)} - (i-1)/n\}$$

$$p_{(i)} = \Phi([x_{(i)} - \bar{x}] / s)$$

Φ = αθροιστική συνάρτηση της τυπικής κανονικής κατανομής

\bar{x} = μέσος

s = τυπική απόκλιση

το p-value υπολογίζεται από τον τύπο των Dallal-Wilkinson (1986) ,το οποίο θεωρείται αξιόπιστο μόνο όταν είναι μικρότερο από 0,1.

Όταν το p-value είναι μεγαλύτερο από 0,1 τότε υπολογίζεται μέσω της τροποποιημένης στατιστικής συνάρτησης

$$Z = D (\sqrt{n} - 0.01 + 0.85/\sqrt{n}), \text{ [την οποία πρότεινε ο Stephens (1974)].}$$

(Δ. Αντζουλάκος, 2011)

P-VALUE

Για να αποφασίσουμε αν θα δεχθούμε ή όχι την μηδενική υπόθεση H_0 στηρίζομαστε στο p-value.

Το p-value μπορεί να ορισθεί ως η ελάχιστη τιμή του επιπέδου σημαντικότητας για την οποία απορρίπτεται η H_0 .

Το p-value είναι ένα μέτρο το οποίο εκφράζει πόσο ισχυρές είναι οι ενδείξεις που προκύπτουν από το δείγμα , εναντίον της H_0 . Έτσι , υπολογίζοντας το p-value ενός δείγματος, για συγκεκριμένο έλεγχο υποθέσεων ,γνωρίζουμε πόσο πιθανή ήταν η εμφάνιση του δείγματος που πήραμε αν η μηδενική υπόθεση ήταν αληθής. Επομένως ,όσο πιο μικρό είναι το p-value τόσο ισχυρότερες ενδείξεις εναντίον της H_0 προκύπτουν από το συγκεκριμένο τυχαίο δείγμα ή αλλιώς τόσο πιο σημαντική είναι η τιμή που προκύπτει από το δείγμα για την στατιστική συνάρτηση ελέγχου. (Wikipedia)

Μπορούμε να αποφασίσουμε με βάση τον ακόλουθο κανόνα

Για κάποια προκαθορισμένη τιμή α

- αν το $p\text{-value} < \alpha$: απορρίπτουμε την H_0

- αν το $p\text{-value} \geq \alpha$: δεν απορρίπτουμε την H_0

Αξίζει να σημειωθεί πως οι πιο συχνές τιμές για το α είναι το 1%, 5% και 10%

Στη συνέχεια θα εφαρμόσουμε στο χαρτοφυλάκιο μας το Lilliefors Test το οποίο πραγματοποιήσαμε με τη χρήση του στατιστικού πακέτου R.

Το αποτέλεσμα είναι το ακόλουθο

$D = 0.013$, $p\text{-value} = 0.956$

Σαν μηδενική υπόθεση θεωρούμε πως τα δεδομένα μας προέρχονται από κανονική κατανομή με παραμέτρους αυτές που έχουμε υπολογίσει για το χαρτοφυλάκιο μας παραπάνω.

$N(56.92701, 14.95993)$

Θεωρούμε επίπεδο σημαντικότητας $\alpha=5\%$

$P\text{-value} \geq 0.05$

Σύμφωνα λοιπόν με τα παραπάνω δεν μπορούμε να απορρίψουμε ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

ANDERSON –DARLING

Το Anderson –Darling είναι ένα στατιστικό έλεγχο καλής προσαρμογής που χρησιμοποιείται όταν τα δεδομένα μας προέρχονται από κάποια συγκεκριμένη κατανομή. Πρόκειται για μια παραλλαγή του Kolmogorov-Smirnov test, η διαφορά του όμως έγκειται στο γεγονός ότι το A-D test δίνει περισσότερη έμφαση στις ουρές από ότι το K-S test.

Το Anderson-Darling test χρησιμοποιεί πάντα κάποια συγκεκριμένη κατανομή για να εκτιμήσει τις παραμέτρους. Αυτό έχει σαν πλεονέκτημα ότι έτσι επιτρέπει να είναι πιο

ευαίσθητο, αλλά έχει και μειονέκτημα το γεγονός πως πρέπει να εκτιμώνται οι παράμετροι για κάθε κατανομή.

Η δοκιμή του test είναι μονόπλευρη και η κατανομή του συγκεκριμένου τύπου απορρίπτεται εφόσον το A-D test είναι μεγαλύτερο από την κρίσιμη τιμή.

Ο έλεγχος του A-D είναι καταλληλότερος για δεδομένα υψηλής επίδρασης γιατί είναι περισσότερο ευαίσθητος στην ουρά των δεδομένων.

Το Anderson-Darling test ορίζεται ως εξής

H_0 =Τα δεδομένα ακολουθούν κάποια συγκεκριμένη κατανομή F

H_1 =Τα δεδομένα δεν ακολουθούν την κατανομή F

Το A-D Test ορίζεται ως εξής

$$A^2 = -N \cdot S$$

$$\text{Όπου } S = \sum_{i=1}^N \frac{(2i-1)}{N} [\ln F(Y_i) + \ln(1-F(Y_{N+1-i}))]$$

F=Αθροιστική συνάρτηση συγκεκριμένης κατανομής

(Δ. Αντζουλάκος, 2011)

Στη συνέχεια θα εφαρμόσουμε στο χαρτοφυλάκιο μας το Anderson-Darling test το οποίο πραγματοποιήσαμε με τη χρήση του στατιστικού πακέτου R.

Το αποτέλεσμα είναι το ακόλουθο

$$AD = 0.5007, \text{ p-value} = 0.7461$$

Σαν μηδενική υπόθεση θεωρούμε πως τα δεδομένα μας προέρχονται από κανονική κατανομή με παραμέτρους αυτές που έχουμε υπολογίσει για το χαρτοφυλάκιο μας παραπάνω.

$$N(56.92701, 14.95993)$$

Θεωρούμε επίπεδο σημαντικότητας $\alpha=5\%$

$$P\text{-value} \geq 0.05$$

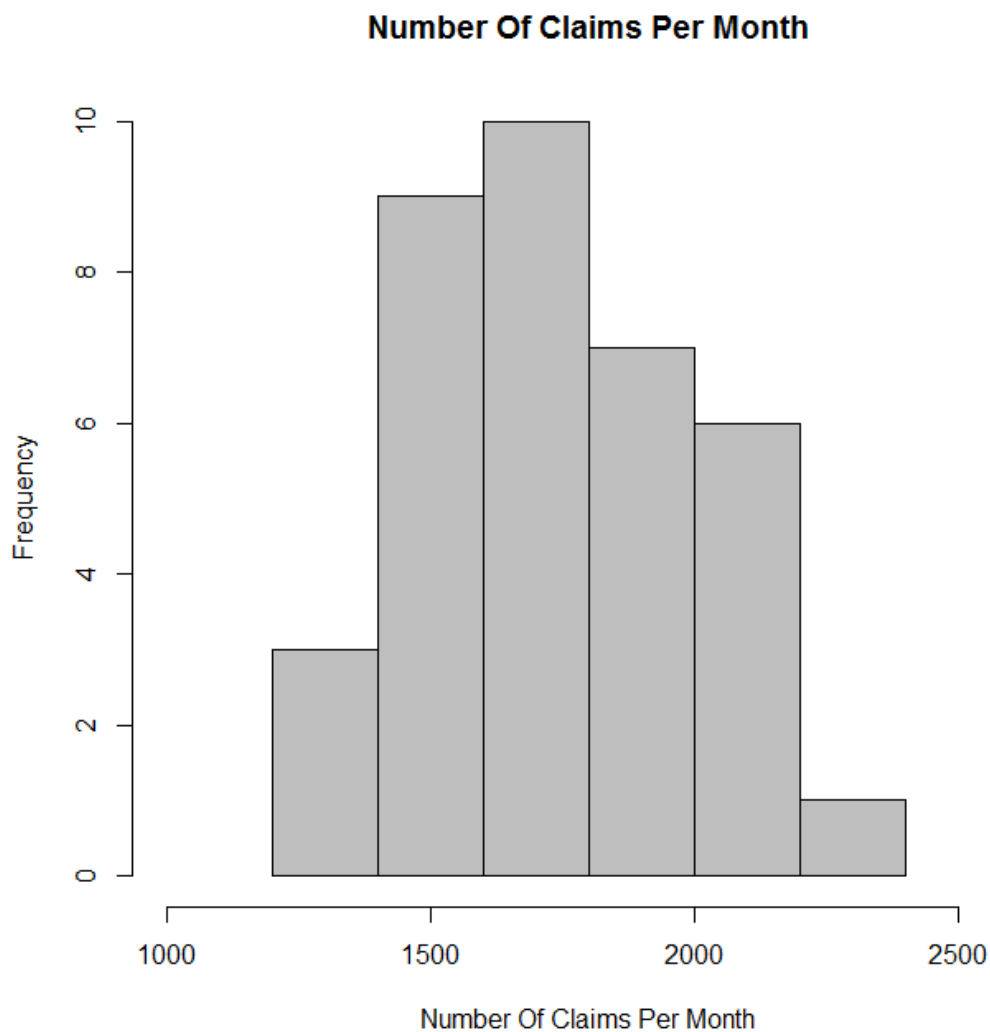
Σύμφωνα λοιπόν με τα παραπάνω δεν μπορούμε να απορρίψουμε ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

B)ANA ΜΗΝΑ

3.2.2 Προσαρμογή πλήθους ζημιών ανά μήνα

Στην ενότητα αυτή θα ακολουθήσουμε την ίδια διαδικασία όπως και στην προσαρμογή του πλήθους ζημιών ανά ημέρα (βλ.ενότητα 3.2.1)

Αρχικά θα ήταν χρήσιμο ένα ιστόγραμμα συχνοτήτων για να φανεί γραφικά αν η κατανομή του πλήθους των ζημιών ανά ημέρα θυμίζει κάποια γνωστή κατανομή.



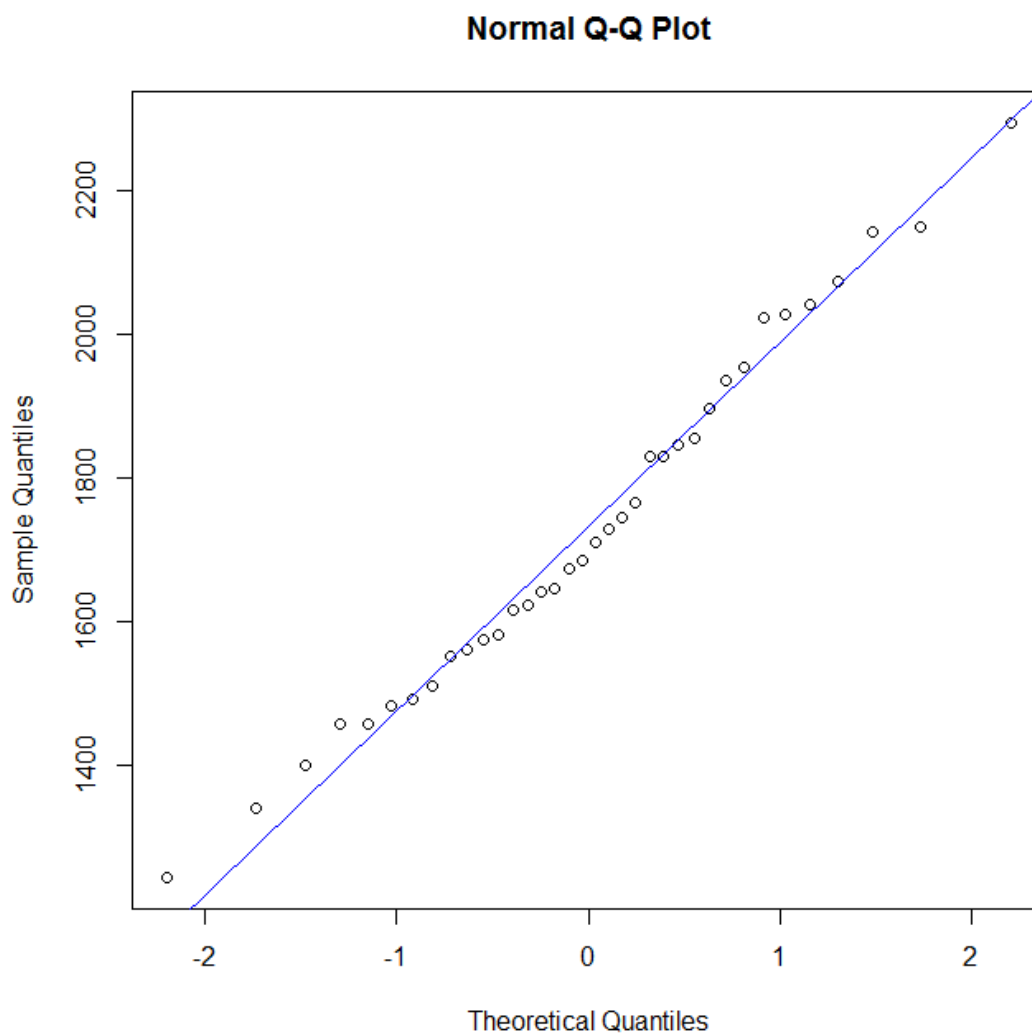
Στο παραπάνω ιστόγραμμα παρουσιάζεται η συχνότητα του αριθμού των αποζημιώσεων ανά μήνα κατά την τριετία για την οποία εξετάζεται το χαρτοφυλάκιο.

Εκ πρώτης όψεως η κατανομή των δεδομένων φαίνεται να προσομοιάζει την κανονική κατανομή.

Με τη χρήση του στατιστικού πακέτου R κάνουμε προσαρμογή στα δεδομένα μας την κανονική κατανομή . Υπολογίσαμε τις παραμέτρους της κατανομής και προέκυψαν τα ακόλουθα

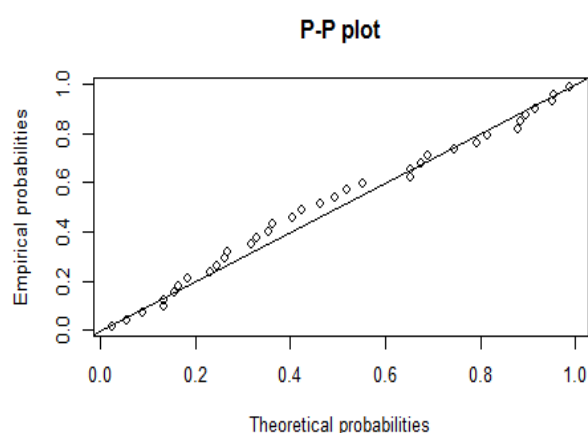
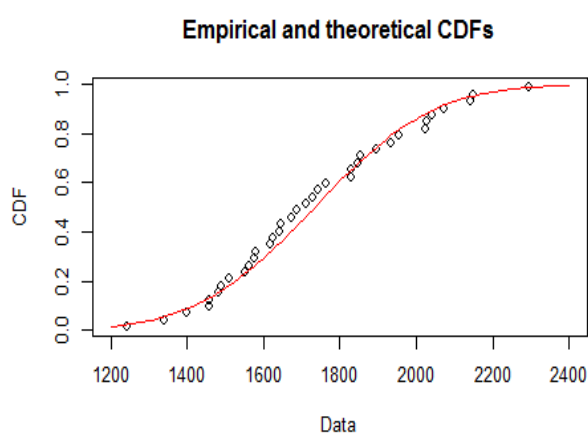
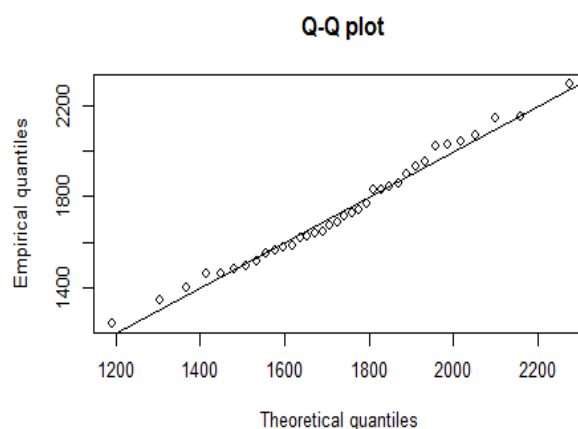
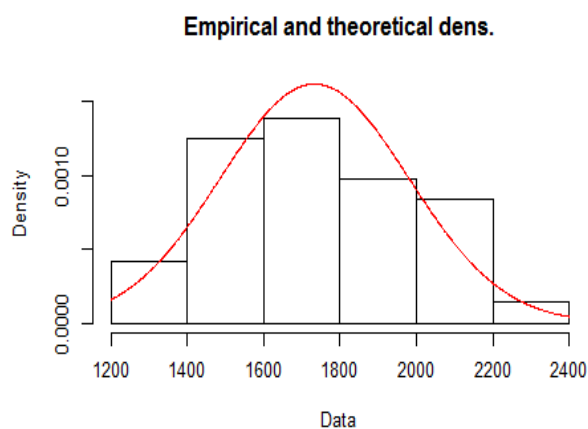
ΠΑΡΑΜΕΤΡΟΙ ΚΑΝΟΝΙΚΗΣ ΚΑΤΑΝΟΜΗΣ		
		Τυπικό σφάλμα (st.error)
Μέσος (mean)	1733,1111	41,12063
Τυπική απόκλιση(standard deviation)	246,7235	29,07685

Στη συνέχεια κατασκευάσαμε για τα δεδομένα μας ένα διάγραμμα Q-Q plot του πλήθους των αποζημιώσεων ανά μήνα για την κανονική κατανομή



Στο συγκεκριμένο διάγραμμα δεν παρατηρούμε σημαντικές αποκλίσεις από την διχοτόμο των αξόνων οπότε μπορούμε να πούμε ότι το δείγμα προέρχεται από την κανονική κατανομή. Όπως μπορεί και να διαπιστώσει κανείς το παραπάνω Q-Q plot επιδεικνύει πολύ καλή προσαρμογή των δεδομένων μας στην κανονική κατανομή.

Στη συνέχεια παρατίθεται γραφήματα που προέκυψαν κατά την ανάλυση των δεδομένων στο στατιστικό πρόγραμμα R.



Συγκεκριμένα παρατηρούμε την προσαρμογή στα δεδομένα μας της εμπειρικής κατανομής το Q-Q plot και το P-P plot. Για τα διαγράμματα αυτά καταλήγουμε στο ίδιο συμπέρασμα ότι δηλαδή έχουμε πολύ καλή προσαρμογή στη κανονική κατανομή.

Έπειτα ακολουθούν τα στατιστικά τεστ Lilliefors (Kolmogorov-Smirnov) και το Anderson-Darling GoF Test.

LILLIEFORS TEST

Σαν μηδενική υπόθεση θεωρούμε πως τα δεδομένα μας προέρχονται από κανονική κατανομή με παραμέτρους αυτές που έχουμε υπολογίσει για το χαρτοφυλάκιο μας παραπάνω.

$$N(1733.1111, 246.7235)$$

Το αποτέλεσμα από το στατιστικό τεστ είναι

$D = 0.0734$, $p\text{-value} = 0.8934$

Θεωρούμε επίπεδο σημαντικότητας $\alpha=5\%$

$P\text{-value} \geq 0.05$

Σύμφωνα λοιπόν με τα παραπάνω δεν μπορούμε να απορρίψουμε ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

ANDERSON –DARLING

Σαν μηδενική υπόθεση θεωρούμε πως τα δεδομένα μας προέρχονται από κανονική κατανομή με παραμέτρους αυτές που έχουμε υπολογίσει για το χαρτοφυλάκιο μας παραπάνω.

$N(1733.1111, 246.7235)$

Το αποτέλεσμα από το στατιστικό τεστ είναι

$AD = 0.2607$ $p\text{-value} = 0.9643$

Θεωρούμε επίπεδο σημαντικότητας $\alpha=5\%$

$P\text{-value} \geq 0.05$

Σύμφωνα λοιπόν με τα παραπάνω δεν μπορούμε να απορρίψουμε ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

Για το χαρτοφυλάκιο μας εκτός από την κανονική κατανομή εξετάσαμε και άλλες κατανομές όμως όλες απορροφήθηκαν.

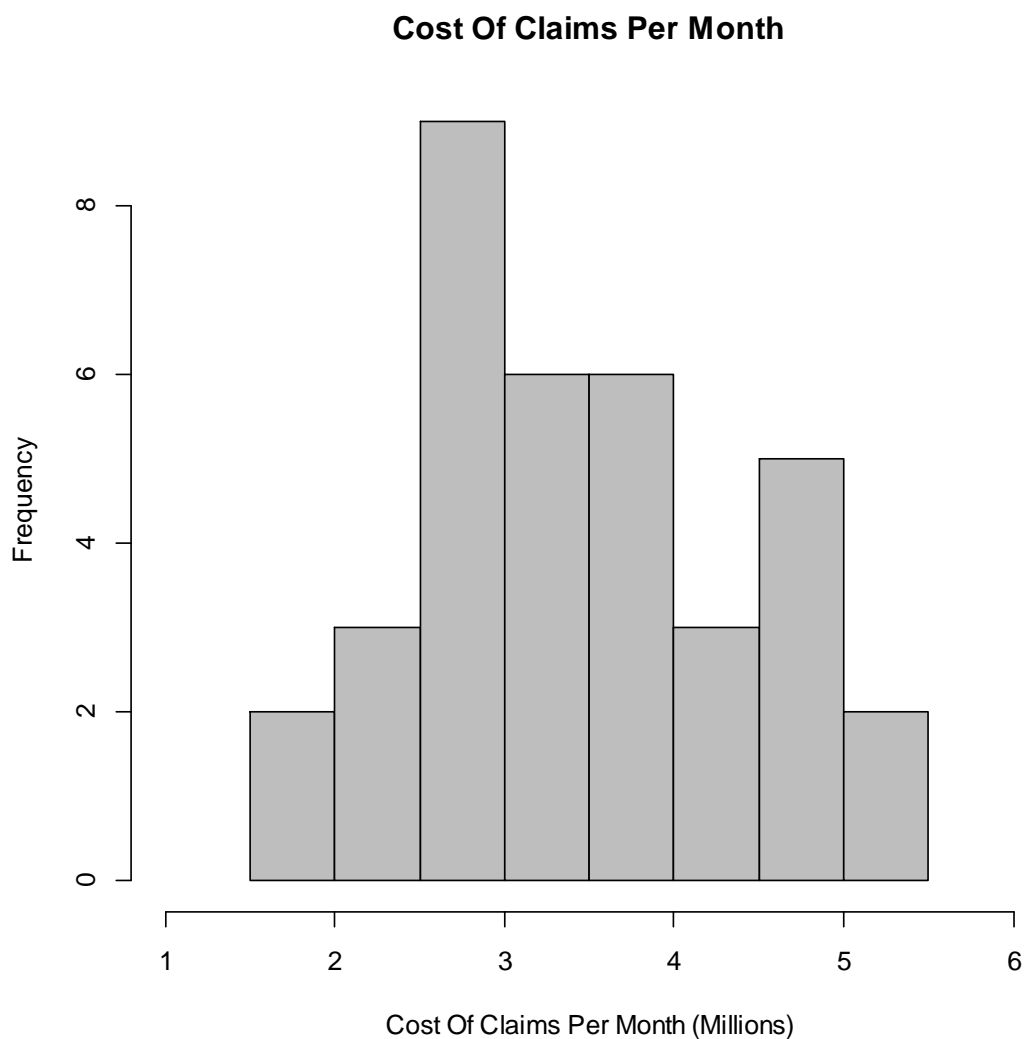
3.3 Ανάλυση ανά ζημιά

Στην ενότητα αυτή προσπαθήσαμε να κάνουμε ανάλυση ανά ζημιά. Για τα μεγέθη των ζημιών επιχειρήσαμε να κάνουμε προσαρμογή κάποιας συνεχούς κατανομής. Με τη βοήθεια του στατιστικού πακέτου EasyFit δοκιμάσαμε όλες τις κατανομές αν προσαρμόζονται στα δεδομένα μας. Μετά από όλους τους ελέγχους που έγιναν όμως δε βρέθηκε κάποια κατανομή που να προσαρμόζεται. Στο παράρτημα 2 παρατίθεται ο πίνακας με τα αποτελέσματα όλων των στατιστικών τεστ που έγιναν και απορρίπτουν όλες τις κατανομές.

Στη συνέχεια διαχωρίσαμε τις ζημιές σε μικρές και μεγάλες και δοκιμάσαμε αν κάποια κατανομή προσαρμόζεται όμως και πάλι απορρίφθηκαν όλες. Τέλος ούτε με τον διαχωρισμό ανάλογα με το είδος της ζημιάς ,σωματική βλάβη ή υλική ζημιά , είχαμε θετικά αποτελέσματα στα στατιστικά τεστ.

3.4 Συνολικές αποζημιώσεις ανά μήνα

Στην ενότητα αυτή θα μελετήσουμε τις αποζημιώσεις ανά μήνα για τα τρία έτη του χαρτοφυλακίου που εξετάζουμε. Αρχικά κατασκευάσαμε μια καινούργια μεταβλητή η οποία περιλαμβάνει το συνολικό μέγεθος ζημιών για κάθε μήνα ,έχουμε δηλαδή να εξετάσουμε 36 παρατηρήσεις (3 έτη x 12 μήνες). Αρχικά θα ήταν χρήσιμο ένα ιστόγραμμα συχνοτήτων για να φανεί γραφικά αν η κατανομή του πλήθους των ζημιών ανά ημέρα θυμίζει κάποια γνωστή κατανομή.



Στη συνέχεια δοκιμάσαμε κάποιες κατανομές και καταλήξαμε πως τα δεδομένα προσαρμόζονται ικανοποιητικά από την Weibull κατανομή.

Κατανομή WEIBULL

Η κατανομή Weibull πήρε το όνομα της από τον Σουηδό καθηγητή Wallodi Weibull ο οποίος την πρότεινε ως μια κατανομή με διάφορες εφαρμογές. Η ομοιότητα που παρατήρησε ανάμεσα στις παρατηρήσεις και στις προβλέψεις μέσω της κατανομής ήταν εντυπωσιακές.

Η κατανομή χρησιμοποιήθηκε για πρώτη φορά το 1933 από τον Ροσίν (Rosin) και τον Ράμλερ (Rammler) στην περιγραφή των «νόμων που διέπουν την λεπτότητα του κονιοροποιημένου άνθρακα» ("Laws governing the fineness of powdered coal"). Η κατανομή Weibull συμπεριλαμβάνει και την εκθετική κατανομή, ως μια ειδική περίπτωση και μερικές φορές θεωρείται γενίκευσή της. Το εκθετικό μοντέλο αναπτύχθηκε και χρησιμοποιήθηκε εκτενώς στην δεκαετία του 1950 και η κατανομή Weibull άρχισε να θεωρείται ανταγωνιστικό μοντέλο την επόμενη δεκαετία, ειδικά σε προβλήματα στα οποία ο χρόνος ζωής ήταν η υπό μελέτη μεταβλητή.

Συνάρτηση πυκνότητας πιθανότητας

$$f(x) = (a/b) (x/b)^{a-1} \exp(- (x/b)^a) \text{ για } x > 0$$

Συνάρτηση κατανομής

$$F(x) = 1 - \exp(- (x/b)^a) \text{ για } x > 0$$

Μέση Τιμή

$$E(X) = b \Gamma(1 + 1/a),$$

Διακύμανση

$$\text{Var}(X) = b^2 * (\Gamma(1 + 2/a) - (\Gamma(1 + 1/a))^2)$$

$$F(t) = 1 - \exp(-(\lambda t)^a) \quad \lambda > 0, \alpha \geq 0, t > 0$$

Η παράμετρος a καλείται παράμετρος μορφής (shape parameter)

Η παράμετρος b καλείται παράμετρος κλίμακας (scale parameter)

Η εκτίμηση των παραμέτρων γίνεται με την εκτίμηση μέγιστης πιθανοφάνειας ή με την μέθοδο των ελαχίστων τετραγώνων.

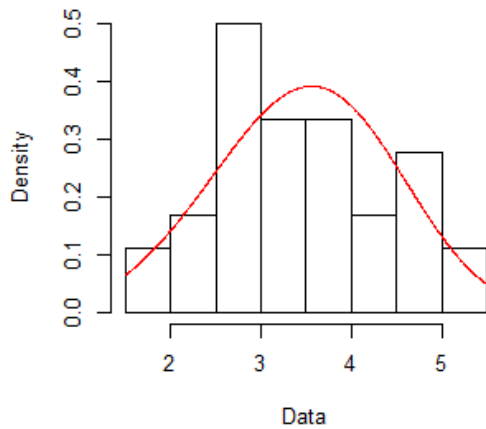
Στα δεδομένα που έχουμε για το χαρτοφυλάκιο μας υπολογίσαμε με την μέθοδο μέγιστης πιθανοφάνειας τις παραμέτρους.

Πρόέκυψαν τα ακόλουθα

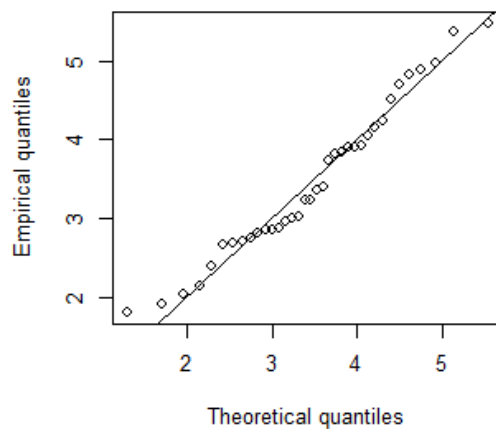
ΠΑΡΑΜΕΤΡΟΙ WEIBULL ΚΑΤΑΝΟΜΗΣ	
SHAPE a	0,687
SCALE b	1208,35

Με τη χρήση του στατιστικού πακέτου R κατασκευάσαμε τα κάτωθι

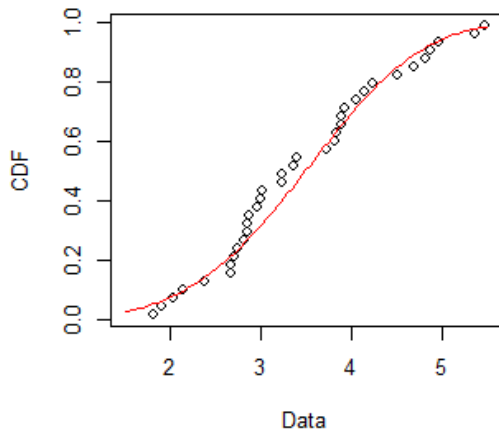
Empirical and theoretical dens.



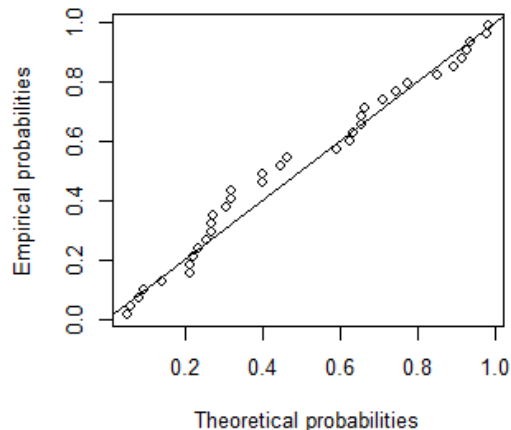
Q-Q plot



Empirical and theoretical CDFs



P-P plot



Παρατηρούμε από τα Q-Q plot και P-P plot, τα δεδομένα να προσαρμόζονται στην διχοτόμο της γωνίας των αξόνων πολύ καλά.

Έπειτα θα εφαρμόσουμε στο χαρτοφυλάκιο μας το Kolmogorov-Smirnov Test το οποίο πραγματοποιήσαμε με τη χρήση του στατιστικού πακέτου R.

Θα χρησιμοποιήσουμε τις παραμέτρους που έχουμε υπολογίσει για τα δεδομένα μας.

Ως μηδενική υπόθεση θεωρούμε πως τα δεδομένα μας προέρχονται από Weibull κατανομή

Θεωρούμε επίπεδο σημαντικότητας $\alpha=5\%$

Το αποτέλεσμα από το στατιστικό τεστ είναι

$D = 0.1235$, $p\text{-value} = 0.5992$

$P\text{-value} \geq 0.05$

Σύμφωνα λοιπόν με τα παραπάνω δεν μπορούμε να απορρίψουμε ότι τα δεδομένα προέρχονται από την Weibull κατανομή.

Στο επόμενο κεφάλαιο σύμφωνα με αυτή την παραδοχή για την κατανομή Weibull θα δοκιμάσουμε να κατασκευάσουμε διαστήματα εμπιστοσύνης.

ΚΕΦΑΛΑΙΟ 4

BOOTSTRAP

4.1 Εισαγωγή

Στο κεφάλαιο 4 παρουσιάζεται η μέθοδος bootstrap, η οποία είναι μια μέθοδος επαναδειγματοληψίας και χρησιμοποιείται για την εκτίμηση άγνωστων παραμέτρων σε ένα πρόβλημα που εξετάζουμε. Η εφαρμογή της μεθόδου εξαρτάται από την φύση του προβλήματος. Θα πρέπει σε κάθε περίπτωση η φύση του προβλήματος να είναι τέτοια που να αξίζει να προτιμήσουμε τη bootstrap από κάποια άλλη παραδοσιακή προσέγγιση. Αρχικά έγινε κάποια αναφορά στην παραδοσιακή προσέγγιση με επισημάνσεις στα μειονεκτήματά της. Στη συνέχεια προσεγγίσαμε βασικά σημεία της μεθόδου καθώς επίσης αναφερθήκαμε και σε περιπτώσεις όπου η μέθοδος αποτυγχάνει. Τέλος έγινε εφαρμογή της μεθόδου στο χαρτοφυλάκιο της εταιρείας που εξετάζουμε. Στην τελευταία ενότητα του κεφαλαίου 3 επεξεργαστήκαμε τις αποζημιώσεις που η εταιρεία κλήθηκε να αποδώσει για κάθε μήνα και για τα 3 έτη. Για τις 36 αυτές παρατηρήσεις θα υπολογίσουμε ένα διάστημα εμπιστοσύνης μέσα στο οποίο θα βρίσκεται με μεγάλη πιθανότητα το μηνιαίο κόστος της αποζημίωσης. Θα χρησιμοποιήσουμε την μέθοδο επαναδειγματοληψίας bootstrap επειδή τα δεδομένα μας δεν προέρχονται από την κανονική κατανομή αλλά όπως είδαμε μια πιθανή κατανομή είναι η Weibull.

4.2 Η παραδοσιακή προσέγγιση

Η αναγνώριση της αβεβαιότητας είναι μια κατηγορηματική παραδοχή της στατιστικής επιστήμης. Ιδέες και αρχές όπως η εκ των προτέρων πληροφορία, πιθανοθεωρητικά μοντέλα, πιθανοφάνεια, πληρότητα, τυπικά σφάλματα και διαστήματα εμπιστοσύνης έχουν κατασκευαστεί με σκοπό να επισημοποιήσουν την αβεβαιότητα και να την παρουσιάζουν σε στατιστικές αναλύσεις. Ο Rao(1989) συνοψίζοντας τα παραπάνω ανέφερε ότι “ ο βασικός στόχος των στατιστικών αναλύσεων είναι να εξάγουμε όσο το δυνατόν περισσότερες πληροφορίες από τα δεδομένα και να συμπεράνουμε τις ιδιότητες του πληθυσμού από τον οποίο προέρχονται τα δεδομένα ”.

Πριν από την συλλογή των δεδομένων, μια στατιστική συνάρτηση είναι μια τυχαία ποσότητα που έχει μια κατανομή πιθανότητας, η οποία ονομάζεται δειγματική κατανομή του στατιστικού και χρησιμοποιείται για την ανάλυση των δεδομένων. Ανάλογα με τον στόχο της ανάλυσης απαιτείται η γνώση της δειγματικής κατανομής ή κάποιου χαρακτηριστικού της. Για παράδειγμα, η κατασκευή διαστημάτων εμπιστοσύνης απαιτεί την γνώση της δειγματικής κατανομής ενώ η εκτίμηση μιας ποσότητας απαιτεί την γνώση συγκεκριμένων χαρακτηριστικών της δειγματικής κατανομής. Επειδή οποιοσδήποτε εκτιμητής μπορεί να έχει

σφάλμα στην εκτίμηση, μας ενδιαφέρει να έχουμε μια ένδειξη ενός μέτρου ακρίβειας της εκτίμησης. Τα μέτρα ακρίβειας, όπως η διασπορά, η μεροληψία και το μέσο τετραγωνικό σφάλμα χρησιμοποιούνται συνήθως για να επιλέξουμε τον καλύτερο εκτιμητή από μια κλάση εκτιμητών.

Η δειγματική κατανομή ενός στατιστικού και τα χαρακτηριστικά του συνήθως εξαρτώνται από τον πληθυσμό και είναι άγνωστα. Αυτά πρέπει να εκτιμηθούν ή να προσεγγιστούν από τα παρατηρούμενα δεδομένα στα περισσότερα προβλήματα εκτίμησης ή προβλήματα συμπερασματολογίας. Για να διαλέξουμε έναν εκτιμητή από μια κλάση εκτιμητών, μερικές φορές χρειάζεται να εκτιμήσουμε ή να προσεγγίσουμε τα μέτρα ακρίβειας. Στις περισσότερες περιπτώσεις η σχετική ακρίβεια των εκτιμητών εξαρτάται από το πληθυσμό που χρησιμοποιήσαμε τα δεδομένα, για να εκτιμήσουμε την σχετική ακρίβεια ώστε να επιλέξουμε τον εκτιμητή. Σύμφωνα με την παραδοσιακή προσέγγιση, ένα μέτρο ακρίβειας εκτιμάται από την εμπειρική αναλογία μιας θεωρητικής εξίσωσης του μέτρου ακρίβειας ή της προσέγγισης του, το οποίο λαμβάνεται από κάποιο μοντέλο.

Οι Shao and Tu (1995) και οι Efron and Tibshirani (1993) υποδεικνύουν μερικές αδυναμίες και μειονεκτήματα της παραδοσιακής προσέγγισης.

- ❖ Το μεγάλο μέγεθος του δείγματος (n) είναι αναγκαίο ώστε να έχουμε ακριβή διασπορά των εκτιμητών. Όταν χρησιμοποιούμε προσεγγίσεις για την διασπορά και επομένως όταν η τάξη του (n) δεν είναι η επιθυμητή τότε η ακρίβεια των αποτελεσμάτων είναι αμφισβητήσιμη. Κλασική είναι η περίπτωση όπου χρησιμοποιούμε προσεγγίσεις της κανονικής κατανομής και το (n) είναι μικρότερο του 30. Τότε συνήθως οδηγούμαστε σε παραπλανητικά αποτελέσματα.
- ❖ Η θεωρητική εξίσωση η οποία χρησιμοποιείται ή η προσέγγισή της βασίζεται σε ένα πιθανοθεωρητικό μοντέλο. Όταν το μοντέλο είναι ελαφρώς λανθασμένο τότε οι εκτιμητές της ακρίβειας μπορεί να μην ισχύουν. Αξίζει να σημειωθεί ότι λανθασμένα αποτελέσματα βασιζόμενα σε ασυμπτωτικές προσεγγίσεις μπορούν να δοθούν και όταν το μέγεθος του δείγματος είναι μεγαλύτερο του 30 αλλά οι προϋποθέσεις της χρήσης των ασυμπτωτικών αποτελεσμάτων δεν ισχύουν.
- ❖ Για να εφαρμόσουμε την παραδοσιακή προσέγγιση σε πολλά διαφορετικά προβλήματα όπως η πιθανοφάνεια, πρέπει να παραγωγίσουμε τους όρους μιας θεωρητικής εξίσωσης για κάθε πρόβλημα. Αυτές οι παραγωγίσεις μπορεί να είναι δύσκολες. Επιπλέον, η παραγωγή της θεωρητικής εξίσωσης απαιτεί από τους αναλυτές των δεδομένων να έχουν καλή γνώση της στατιστικής και μαθηματικής επιστήμης.
- ❖ Ίσως όμως το σημαντικότερο πρόβλημα που έχει να αντιμετωπίσει ο στατιστικός είναι να εξετάσει κατά πόσο πληρούνται οι προϋποθέσεις ενός παραμετρικού μοντέλου. Αυτό σημαίνει συνήθως ότι πρέπει να πραγματοποιηθεί ένας μεγάλος αριθμός διαγνωστικών ελέγχων προκειμένου να εξασφαλιστούν οι υποθέσεις αυτές. Ακόμα και όταν το αποτέλεσμα των διαγνωστικών τεστ είναι θετικό, υπάρχουν περιπτώσεις όπου δημιουργούνται απορίες για την εγκυρότητα τους. Ένα παράδειγμα μπορεί να προκύψει στην απλή γραμμική παλινδρόμηση, με την εξέταση της κανονικότητας των υπολοίπων.

Μπορεί μέσω του στατιστικού Kolmogorov Smirnov να είναι αποδεκτή αλλά μέσω γραφικών ελέγχων , για παράδειγμα του qq-plot να παρουσιάζονται αποκλίσεις από την διχοτόμο και επομένως εύλογες απορίες για την εγκυρότητα του αποτελέσματος.

- ❖ Η θεωρητική εξίσωση μπορεί να είναι πολύπλοκη για να είναι χρήσιμη στην εκτίμηση του μέτρου ακρίβειας.

Για την αντιμετώπιση των παραπάνω ο Efron(1979) παρουσίασε την μέθοδο bootstrap ως μια μη παραμετρική μέθοδο εκτίμησης του τυπικού σφάλματος εκτιμητών . Η απλότητα της ιδέας της συντέλεσε στη ραγδαία ανάπτυξη της και στην εφαρμογή της σε διάφορους τομείς τη στατιστικής συμπερασματολογίας, όπως για παράδειγμα στην εκτίμηση παραμέτρων, στην κατασκευή διαστημάτων εμπιστοσύνης ,τον έλεγχο υποθέσεων και αλλού. Η μέθοδος bootstrap χρησιμοποιείται ευρέως σε συνεχείς τυχαίες μεταβλητές ενώ η χρήση της είναι πιο περιορισμένη για διακριτές τυχαίες μεταβλητές.(Efron and Tibshirani,1993)

4.3 Η μέθοδος bootstrap

Η μέθοδος Bootstrap είναι μια από τις τεχνικές επαναδειγματοληψίας που εμφανίστηκε λίγο πριν το 1980. Χρησιμοποιείται για την εκτίμηση διαφόρων παραμέτρων. Η μέθοδος πήρε το όνομα της από τον μυθικό βαρόνο Munchausen,ο οποίος για να αποφύγει τον πνιγμό σε μια λίμνη, προσπάθησε να τραβήξει τον εαυτό του στην επιφάνεια τραβώντας προς τα πάνω τα λουριά από τις μπότες του.

Βασική ιδέα της μεθόδου είναι ότι από ένα αρχικό δείγμα κατασκευάζουμε νέα δείγματα κάνοντας απλή τυχαία δειγματοληψία με επανάθεση.

Το 1979 εισήχθη από τον Bradley Efron η μέθοδος bootstrap. Έκτοτε δεν έχει σταματήσει να αναπτύσσεται καθώς με την ανάπτυξη της τεχνολογίας είναι εφικτό να κάνουμε προσομοίωση και να κατασκευάζουμε νέα δείγματα μέσω της επαναδειγματοληψίας πολύ ταχύτερα από ότι στο παρελθόν.

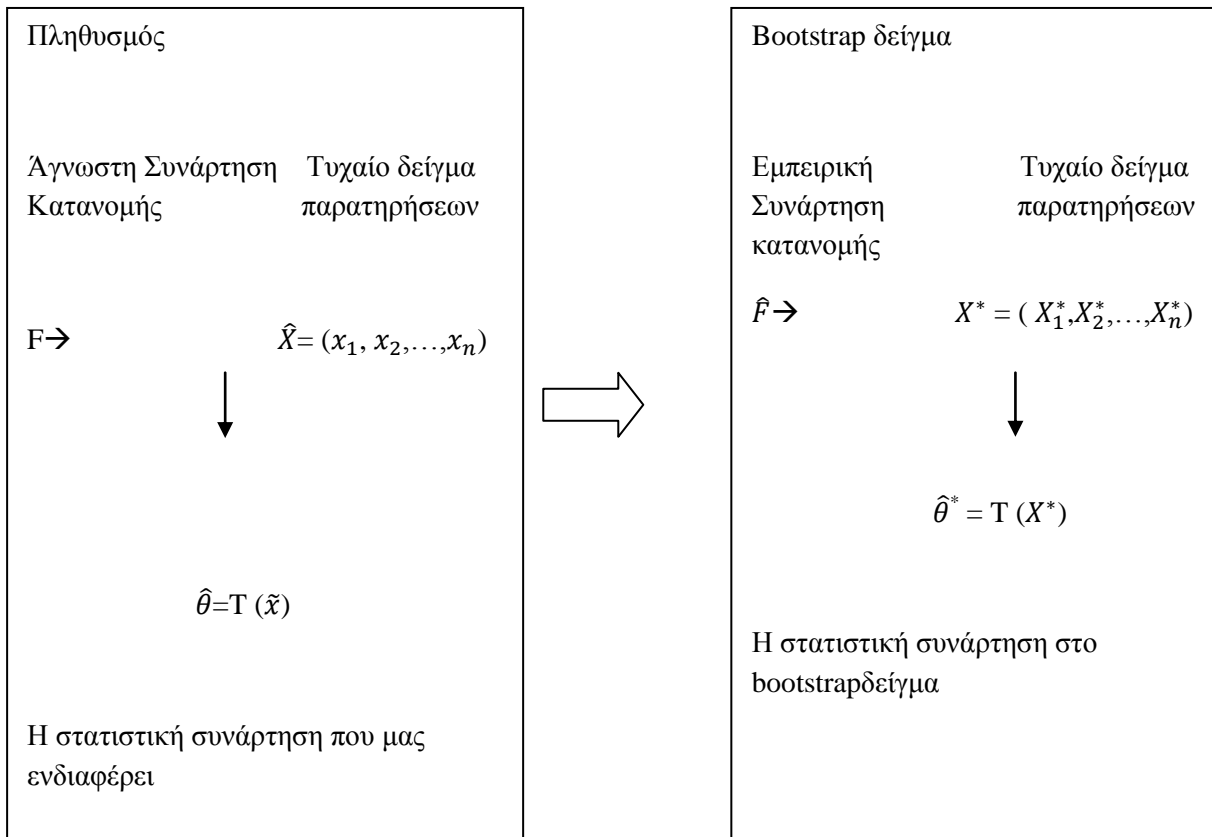
Πλέον επεξεργαζόμαστε ζητήματα χωρίς να εξαρτόμαστε από πολύπλοκους μαθηματικούς τύπους.

Παρόμοιες τεχνικές επαναδειγματοληψίας είχαν εμφανιστεί ήδη από το 1949, όπως η μέθοδος jackknife από τον Quenouille , εκείνη την εποχή άρχισαν να χρησιμοποιούνται και οι υπολογιστές για την προσομοίωση. Από τότε και μετά, με την ταυτόχρονη πρόοδο των υπολογιστών, η bootstrap αποτελεί μια τεχνική που αναπτύσσεται, ενώ σταδιακά προστέθηκαν και άλλες παραλλαγές της μεθόδου, όπως η ABC bootstrap (1992, DiCiccio-Efron) και η Bca [Bias Accelerated Bias –Corrected Percentile] (1996, DiCiccio-Efron).

Η μέθοδος Bootstrap καλείται να υπολογίσει γνώριμα στοιχεία στατιστικής, όπως είναι το τυπικό σφάλμα, η μεροληψία, τα διαστήματα εμπιστοσύνης με έναν τρόπο που διαφέρει από την παραδοσιακή προσέγγιση.

Εισαγωγή στη μέθοδο bootstrap

Από ένα αρχικό δείγμα παρατηρήσεων $\tilde{x} = x_1, x_2, \dots, x_n$ μεγέθους n , που προέρχεται από την κατανομή F , κατασκευάζουμε ένα bootstrap δείγμα του ίδιου μεγέθους πραγματοποιώντας δειγματοληψία με επανάθεση. Αυτό σημαίνει ότι κάποια από τις παρατηρήσεις του αρχικού μας δείγματος μπορεί να εμφανίζεται 0,1,2, ακόμα και n φορές στο bootstrap δείγμα. Αν συμβολίσουμε το bootstrap δείγμα ως $X_1^*, X_2^*, \dots, X_n^*$, τότε αυτό θα προέρχεται από την εμπειρική συνάρτηση κατανομής \hat{F}_n , κατανομή που εκτιμά την κατανομή F . Υπολογίζουμε για το $X_1^*, X_2^*, \dots, X_n^*$ την στατιστική συνάρτηση $T^* = T(X_1^*, X_2^*, \dots, X_n^*)$ και επαναλαμβάνοντας την παραπάνω διαδικασία B φορές, θα έχουμε συνολικά B bootstrap δείγματα, τα οποία θα συμβολίσουμε ως $X^{*1}, X^{*2}, \dots, X^{*B}$ και αντίστοιχα $T_1^*, T_2^*, \dots, T_B^*$. Σημειώνεται ότι μετά τις B επαναλήψεις της διαδικασίας αυτό που λαμβάνουμε είναι μια Monte Carlo εκτίμηση της στατιστικής συνάρτησης που μας ενδιαφέρει. Θα παραθέσουμε μία σχηματική αναπαράσταση που περιγράφει την παραπάνω διαδικασία ώστε να γίνει πιο κατανοητή η μέθοδος:



Εικόνα 4.1 Πηγή: Efron & Tibshirani(1993)

Όπως έγινε κατανοητό τα $X_1^*, X_2^*, \dots, X_n^*$ δεν αποτελούν το πραγματικό δείγμα, αλλά μια εκδοχή του πραγματικού δείγματος στο οποίο έχει γίνει δειγματοληψία με επανάθεση. Για αυτό άλλωστε και η μέθοδος bootstrap είναι μια μέθοδος επαναδειγματοληψίας, διότι κάνουμε δειγματοληψία μέσα από το δείγμα. Στη συγκεκριμένη περίπτωση μάλιστα, επειδή δεν κάναμε καμία υπόθεση για την κατανομή της F η μέθοδος καλείται *απαραμετρική μέθοδος bootstrap*. Η εκτιμήτρια \hat{F} που χρησιμοποιείται για την παραγωγή των bootstrap δειγμάτων μπορεί να είναι οποιαδήποτε εκτιμήτρια της F , όχι απαραίτητα η εμπειρική συνάρτηση κατανομής (Efron-Tibshirani 1993). Ωστόσο βέβαια μια μη παραμετρική εκτιμήτρια της F είναι η εμπειρική συνάρτηση κατανομής η οποία κατανέμει πιθανότητα σε καθένα από τα x_1, x_2, \dots, x_n ίση με $1/n$ και στην περίπτωση που έχουμε επαναλαμβανόμενες τιμές ή κατηγορικά δεδομένα, η εμπειρική συνάρτηση κατανομής δίνει βάρος σε κάθε διακριτή τιμή του δείγματος ίση με το ποσοστό της συχνότητας στο δείγμα. (Efron & Tibshirani, 1993)

4.3.1 Μέθοδος bootstrap και τυπικό σφάλμα

Όπως σε κάθε εκτίμηση μας ενδιαφέρει η ακρίβειά της, έτσι και στις εκτιμήσεις που λαμβάνουμε με την μέθοδο bootstrap μας ενδιαφέρει να υπολογίσουμε το τυπικό σφάλμα. Ας υποθέσουμε λοιπόν ότι έχουμε ένα δείγμα x_1, x_2, \dots, x_n το οποίο προέρχεται από τη κανονική συνάρτηση κατανομής με μέση τιμή και διακύμανση σ^2 , δηλαδή $N(\mu, \sigma^2)$. Όπως γνωρίζουμε ο δειγματικός μέσος εκτιμά αμερόληπτα την μέση τιμή του πληθυσμού. Η εκτίμηση του τυπικού σφάλματος για τον δειγματικό μέσο $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$ δίνεται από τον τύπο:

$$\widehat{se}(\bar{x}) = \sqrt{\frac{s^2}{n}}, \text{ όπου } s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1)$$

Αν τα x_i λοιπόν, προέρχονται από την κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2 , τότε ο δειγματικός μέσος ακολουθεί και αυτός την κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2/n , δηλαδή

$$\bar{x} \sim N(\mu, \sigma^2/n).$$

Όπως γίνεται αντιληπτό ο δειγματικός μέσος έχει μικρότερη διακύμανση από τα x_i και μάλιστα όσο μεγαλύτερο είναι το μέγεθος του δείγματος τόσο μικρότερη θα είναι η διακύμανση του δειγματικού μέσου και τόσο καλύτερη θα είναι η εκτίμηση. Στην περίπτωση όμως που το δείγμα μας προέρχεται από κάποια άγνωστη κατανομή F , δεν είναι εφικτό να κάνουμε μια εκτίμηση της διακύμανσης ώστε να εκτιμήσουμε το τυπικό σφάλμα. Δηλαδή αν τα x_1, x_2, \dots, x_n προέρχονται από την F , τότε ο δειγματικός μέσος θα έχει τις παρακάτω παραμέτρους

$$\bar{x} \sim \left(\mu_F, \frac{\sigma_F^2}{n} \right)$$

και το τυπικό σφάλμα θα δίνεται από τον τύπο:

$$se_F(\bar{x}) = [Var(\bar{x})]^{1/2} = \sigma_F / \sqrt{n}$$

Με βάση το κεντρικό οριακό θεώρημα, για $n \rightarrow \infty$ ο δειγματικός μέσος ακολουθεί ασυμπτωτικά την κανονική κατανομή $N(\mu_F, \frac{\sigma_F^2}{n})$ και έτσι μπορούμε να πούμε ότι ο μέσος για τον πληθυσμό αναμένεται με πιθανότητα 68,3% να απέχει από την εκτίμηση του το πολύ ένα τυπικό σφάλμα, ενώ με πιθανότητα 95,4% θα απέχει λιγότερο από δυο τυπικά σφάλματα από την εκτίμηση του. Για τις περισσότερες στατιστικές συναρτήσεις ωστόσο, δεν υπάρχει κάποιος τύπος ώστε να υπολογίζεται άμεσα το τυπικό σφάλμα και εκτός από αυτό δεν είναι πάντα εφικτό να εφαρμόσουμε το κεντρικό οριακό θεώρημα. Με τη βοήθεια της μεθόδου bootstrap απαλλασσόμαστε από τις υποθέσεις του κεντρικού οριακού θεωρήματος και λαμβάνουμε μια εκτίμηση για το τυπικό σφάλμα μιας οποιασδήποτε στατιστικής συνάρτησης $T(x)$ ακόμα και αν δεν γνωρίζουμε την κατανομή F από την οποία προέρχεται το δείγμα.

Ας υποθέσουμε λοιπόν πως από ένα τυχαίο δείγμα X_1, X_2, \dots, X_n που προέρχεται από την κατανομή F , έχουμε τις παρατηρήσεις x_1, x_2, \dots, x_n , και μας ενδιαφέρει η παράμετρος $\theta = t(F)$, η οποία εκτιμάται από την $\hat{\theta} = T(x)$.

Ουσιαστικά αν το τυπικό σφάλμα της εκτίμησης $\hat{\theta} = T(x)$ είναι το $se_F(\hat{\theta})$, εμείς θα το εκτιμήσουμε με το $se_F(\hat{\theta}^*)$. Το $se_F(\hat{\theta}^*)$ είναι μια plug-in εκτίμηση του $se_F(\hat{\theta})$, αφού αντί του αρχικού δείγματος που ακολουθεί την κατανομή F , θα χρησιμοποιήσουμε τα δείγματα μεγέθους n που προέρχονται από την συνάρτηση κατανομής \hat{F}_n , η οποία εκτιμά την F .

Ας χρησιμοποιήσουμε την εμπειρική συνάρτηση κατανομής \hat{F}_n στην θέση της F . Τότε η εκτίμηση $se_F(\hat{\theta}^*)$ καλείται ιδανική εκτίμηση bootstrap για το τυπικό σφάλμα (ideal bootstrap estimate of standard error σύμφωνα με τους Efron-Tibshirani 1993) της $\hat{\theta} = T(x)$ και είναι μη παραμετρική, διότι βασίζεται στην \hat{F}_n χωρίς να λάβουμε υπ' όψιν την F . Επειδή όπως αναφέρθηκε παραπάνω δεν υπάρχει πάντα κάποιος τύπος που να μας οδηγεί στην εκτίμηση του τυπικού σφάλματος προτείνεται στη συνέχεια ένας υπολογιστικός τρόπος για την προσέγγιση του $se_F(\hat{\theta}^*)$, ο οποίος και μας οδηγεί στην bootstrap εκτίμηση του τυπικού σφάλματος την οποία και θα συμβολίσουμε με \widehat{se}_B .

Αναλυτικότερα, όπως περιγράφηκε παραπάνω από τα αρχικά δεδομένα x_1, x_2, \dots, x_n λαμβάνουμε ένα δείγμα μεγέθους n με επανάθεση, έστω $X^* = (X_1^*, X_2^*, \dots, X_n^*)$. Το δείγμα των X^* αποτελείται από τα στοιχεία του αρχικού δείγματος (x_1, x_2, \dots, x_n) , κάποια από τα οποία εμφανίζονται μηδέν φορές, δύο φορές κ.ο.κ., αλλά φυσικά δεν αποτελεί το αυθεντικό σύνολο των δεδομένων. Επαναλαμβάνουμε τη διαδικασία B φορές και έτσι κατασκευάζουμε B ανεξάρτητα bootstrap δείγματα $X^{*1}, X^{*2}, \dots, X^{*B}$.

Η μέθοδος bootstrap θα μας βοηθήσει να εκτιμήσουμε το τυπικό σφάλμα της $T(\hat{x})$ χωρίς περιττούς υπολογισμούς, ανεξάρτητα από το πόσο πολύπλοκος είναι ο τύπος της. Για κάθε ένα από τα bootstrap δείγματα $X^{*1}, X^{*2}, \dots, X^{*B}$ υπολογίζουμε την τιμή $\hat{\theta}^* = T(X^{*b})$, όπου $b=1, \dots, B$. Η τυπική απόκλιση των $T(X^{*1}), T(X^{*2}), \dots, T(X^{*B})$ είναι η εκτίμηση του τυπικού σφάλματος της $T(\hat{x})$. Δηλαδή,

$$\widehat{se}_{boot} = \frac{\sum_{b=1}^B [T(X^{*b}) - T(\cdot)]^2}{B - 1}$$

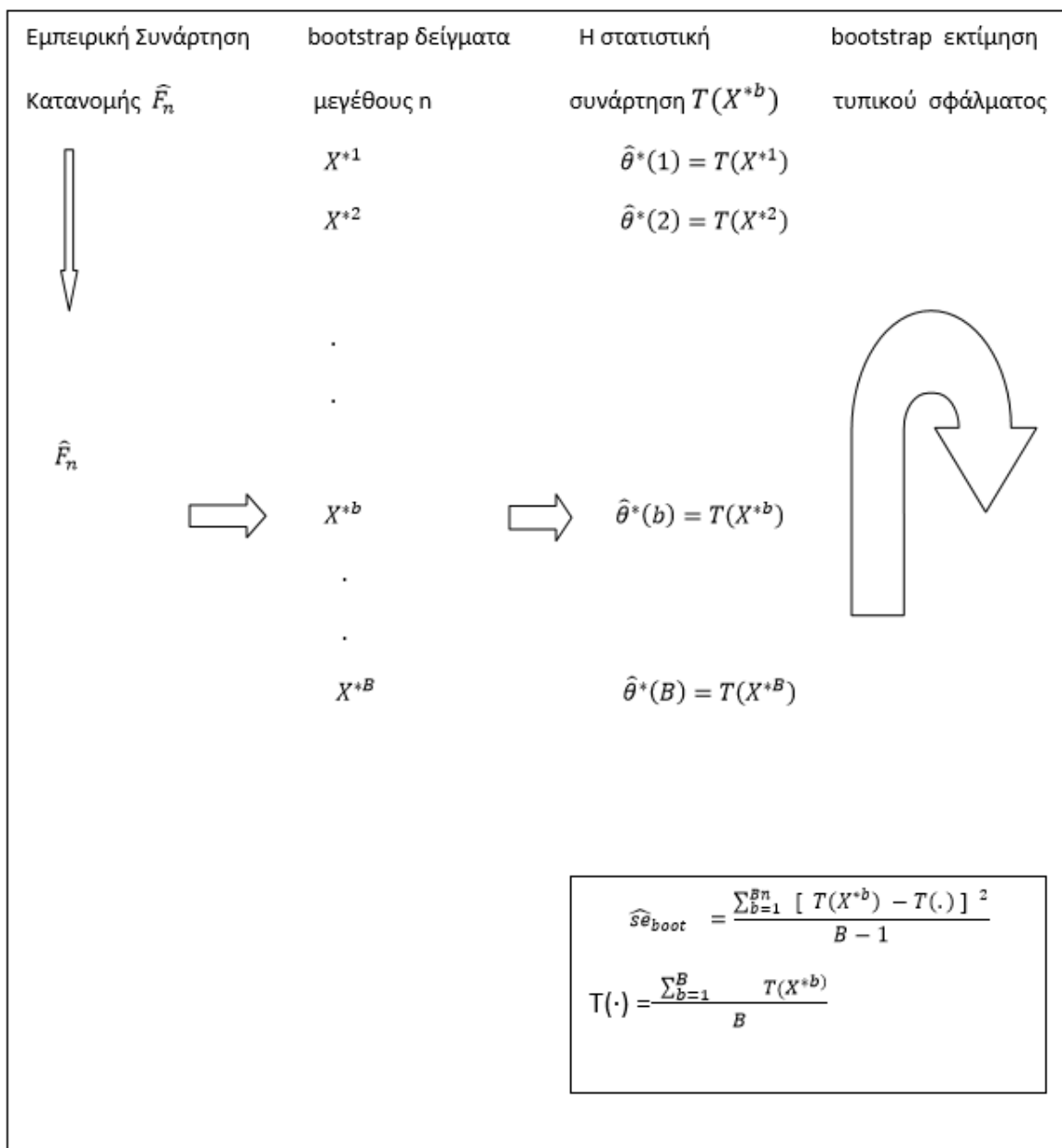
όπου το δίνεται από την σχέση:

$$T(\cdot) = \frac{\sum_{b=1}^B T(X^{*b})}{B}.$$

Σημειώνουμε ότι η ποσότητα $T(X^{*b})$ προκύπτει επειδή εφαρμόσαμε την στατιστική συνάρτηση σε κάθε ένα από τα bootstrap δείγματα ($b=1,2,\dots,B$). Για παράδειγμα αν η $T(X)=\bar{X}$, τότε για την συνάρτηση $T(X^{*b})$ θα έχω:

$$T(X^{*b}) = \bar{X}^* = \sum_{i=1}^n \frac{X_i^*}{n}$$

Στη συνέχεια παρουσιάζεται μια σχηματική αναπαράσταση της διαδικασίας για την bootstrap εκτίμηση του τυπικού σφάλματος μιας στατιστικής συνάρτησης.



Εικόνα 4.2 Πηγή (B.Efron, R.J. Tibshirani 1993)

4.3.2 Το bootstrap διάστημα εμπιστοσύνης

Υποθέτουμε ότι έχουμε τις παρατηρήσεις x_1, x_2, \dots, x_n οι οποίες προήλθαν κατόπιν τυχαίας δειγματοληψίας από την άγνωστη κατανομή F . Έστω $\theta = t(F)$ η παράμετρος που μας ενδιαφέρει, $\hat{\theta}$ η plug-in εκτίμησή της, $T = T(X_1, X_2, \dots, X_n)$ στατιστική συνάρτηση που χρησιμοποιήσαμε για την εκτίμηση της παραμέτρου και \widehat{se} μια εκτίμηση για το τυπικό σφάλμα της $\hat{\theta}$, η οποία προήλθε με τη μέθοδο bootstrap.

Σε αρκετές περιπτώσεις και ενώ το μέγεθος n του δείγματος μεγαλώνει τότε η κατανομή της $\hat{\theta}$ προσεγγίζει όλο και πιο πολύ την κανονική κατανομή με μέση τιμή κοντά στο θ και διακύμανση κοντά στο \widehat{se}^2 .

Δηλαδή:

$$\hat{\theta} \sim N(\theta, \widehat{se}^2) \quad \text{ή ισοδύναμα} \quad \frac{\hat{\theta} - \theta}{\widehat{se}} \sim N(0,1)$$

Η ποσότητα $\frac{\hat{\theta} - \theta}{\widehat{se}}$ είναι η ποσότητα οδηγός και το διάστημα εμπιστοσύνης με $1-\alpha$ πιθανότητα κάλυψης ή $100(1-\alpha)\%$ συντελεστή εμπιστοσύνης προκύπτει:

$$[\hat{\theta} - z_{1-\frac{\alpha}{2}} * \widehat{se}, \hat{\theta} + z_{\frac{\alpha}{2}} * \widehat{se}]$$

Ουσιαστικά το παραπάνω διάστημα, το οποίο χρησιμοποιεί την προσέγγιση της τυποποιημένης κανονικής κατανομής καλείται κανονικοποιημένο (standardized) και είναι ένα προσεγγιστικό διάστημα διότι η πιθανότητα κάλυψης στις περισσότερες περιπτώσεις δεν είναι ακριβώς $1-\alpha$.

Ωστόσο η παραπάνω διαδικασία που περιγράψαμε δεν είναι πάντα εφικτή διότι συχνά δεν είναι εύκολο να βρούμε μια ποσότητα οδηγό, της οποίας η κατανομή να μην εξαρτάται από την παράμετρο αλλά να είναι γνωστή και αρκετές φορές δεν μπορούμε να βασιστούμε σε υποθέσεις κανονικότητας. Θα χρησιμοποιήσουμε την ιδέα της μεθόδου bootstrap για να κατασκευάσουμε τα διαστήματα εμπιστοσύνης.

Αναλυτικότερα, θα πρέπει να χρησιμοποιήσουμε τα ποσοστιαία σημεία της κατανομής της $\theta - \hat{\theta}$, όπου $\hat{\theta} = T(X_1, X_2, \dots, X_n)$.

Παραθέτουμε την διαδικασία όπως αναφέρεται στην βιβλιογραφία (Μ. Μπούτσικας, 2008):

Αρκεί να εκτιμήσουμε τα σημεία $c_{\alpha/2}$ και $c_{1-\alpha/2}$, έτσι ώστε να ισχύει:

$$P(\hat{\theta} - \theta < c_{\alpha/2}) = \alpha/2 \quad \text{και} \quad P(\hat{\theta} - \theta < c_{1-\alpha/2}) = 1-\alpha/2$$

Τότε θα έχουμε ότι

$$P(c_{1-\alpha/2} < \hat{\theta} - \theta < c_{\alpha/2}) = 1-\alpha$$

και το διάστημα $(\hat{\theta} - \theta < c_{\alpha/2}, \hat{\theta} - \theta < c_{1-\alpha/2})$ θα είναι ένα διάστημα εμπιστοσύνης για την παράμετρο θ , με συντελεστή εμπιστοσύνης $1-\alpha$. Θεωρούμε ότι η παράμετρος θ είναι συνάρτηση μιας κατανομής F και ότι η εκτίμηση $\hat{\theta} = T$, ακολουθεί μια κατανομή την F_T .

Οπότε έχουμε:

$$P(\hat{\theta} - \theta_F > c_{\alpha/2}) = \alpha/2 \Leftrightarrow F_T(c_{\alpha/2} + \theta_F) = 1 - \frac{\alpha}{2}$$

$$\Leftrightarrow c_{\frac{\alpha}{2}} = F_T^{-1}(1 - \frac{\alpha}{2}) - \theta_F$$

Αντικαθιστούμε την F με την \hat{F} και επομένως την T με την T^* , οπότε θα είναι:

$$\hat{c}_{\frac{\alpha}{2}} = F_{T^*}^{-1}(1 - \frac{\alpha}{2}) - \theta_F$$

όπου $T^* = T(X_1^*, \dots, X_n^*)$. Επειδή είναι δύσκολος ο υπολογισμός της $F_{T^*}^{-1}(1 - \frac{\alpha}{2})$ χρησιμοποιούμε και πάλι προσομοίωση.

Παράγουμε $T_1^*, T_2^*, \dots, T_B^*$, τυχαίους αριθμούς και τότε η F_{T^*} προσεγγίζεται από την εμπειρική συνάρτηση κατανομής:

$$\hat{F}_{T^*} = \frac{1}{B} \sum_{i=1}^B I(T_i^* < x)$$

Διατάσσοντας τα T_i^* , από το μικρότερο στο μεγαλύτερο, έχουμε το σύνολο των $T_{(i)}^*$, οι εκτιμήσεις που παίρνουμε για το άνω και το κάτω άκρο του διαστήματος είναι :

$$\hat{c}_{\alpha/2} = T_{(B(1-\frac{\alpha}{2}))}^* - \theta_{\hat{F}} \text{ και } \hat{c}_{1-\alpha/2} = T_{(B(\frac{\alpha}{2}))}^* - \theta_{\hat{F}}$$

Το αντίστοιχο δ.ε. θα είναι:

$$[\hat{\theta} - T_{(B(1-\frac{\alpha}{2}))}^* - \theta_{\hat{F}}, \hat{\theta} - T_{(B(\frac{\alpha}{2}))}^* - \theta_{\hat{F}}]$$

Αν μάλιστα έχουμε θέσει ότι $\theta_{\hat{F}} = \hat{\theta}$, τότε το διάστημα εμπιστοσύνης θα είναι ίσο με:

$$[2\hat{\theta} - T_{(B(1-\frac{\alpha}{2}))}^* - \theta_{\hat{F}}, 2\hat{\theta} - T_{(B(\frac{\alpha}{2}))}^* - \theta_{\hat{F}}]$$

Το παραπάνω δ.ε. καλείται βασικό bootstrap διάστημα εμπιστοσύνης.

4.3.3 Αδυναμίες της μεθόδου bootstrap

Η μέθοδος δεν εφαρμόζεται το ίδιο αποτελεσματικά σε όλες τις περιπτώσεις. Ωστόσο θα πρέπει να λάβουμε υπ' όψη ότι ακόμα και για αυτές τις περιπτώσεις έχουν προταθεί αλγόριθμοι ή παραλλαγές της κλασικής μεθόδου που βελτιώνουν τα αποτελέσματα (Chernick 1999 & Davison and Hinkley 1997). Η εφαρμογή της μεθόδου αποτυγχάνει όταν έχουμε εξαρτημένα δεδομένα και αυτό διότι η ανεξαρτησία προϋποθέτει ότι η από κοινού συνάρτηση κατανομής των x_i μπορεί να γραφεί ως γινόμενο των περιθωρίων κατανομών τους, κάτι που προφανώς δεν ισχύει αν έχουμε εξαρτημένα δεδομένα. Στην περίπτωση που υπάρχει ασθενής εξάρτηση η μη παραμετρική μέθοδος δίνει ικανοποιητικά αποτελέσματα. Γενικότερα αν έχουμε κάποια πληροφορία για την κατανομή F του αρχικού δείγματος, καλό είναι να ακολουθείται η παραμετρική μέθοδος και όχι η μη παραμετρική. Αν γνωρίζουμε ότι τα δεδομένα προέρχονται από μια συνεχή κατανομή δεν είναι ωφέλιμο να προσπαθούμε να την προσεγγίσουμε με την εμπειρική συνάρτηση κατανομής η οποία είναι διακριτή.

Δυσκολία στην εφαρμογή της μεθόδου εντοπίζεται και στην περίπτωση που έχουμε ελλιπή δεδομένα ή έκτροπες παρατηρήσεις. Σε κάθε περίπτωση υποθέτουμε ότι έχουμε ένα αρχικό δείγμα παρατηρήσεων x_1, x_2, \dots, x_n και πάνω σε αυτό βασίζουμε την μέθοδο και τα συμπεράσματά μας. Υπάρχει ωστόσο το ενδεχόμενο το αρχικό μας δείγμα να μην είναι ολοκληρωμένο, όπως για παράδειγμα αν έχουμε επαναλαμβανόμενες μετρήσεις και κάποια μέτρηση απουσιάζει.

Από την άλλη όταν έχουμε έκτροπες παρατηρήσεις και εφαρμόζουμε την bootstrap είναι χρήσιμο να ελέγξουμε το αποτέλεσμα της προσομοίωσης και να δούμε αν τα αποτελέσματα εξαρτώνται από συγκεκριμένες παρατηρήσεις.

Κλασική περίπτωση στην οποία η bootstrap δεν ανταποκρίνεται ικανοποιητικά είναι στην περίπτωση ύπαρξης ακραίων τιμών (K.B. Athreya 1987). Ένα άλλο παράδειγμα αποτυχίας της μεθόδου είναι εάν θέλουμε να εκτιμήσουμε την παράμετρο θ από ένα δείγμα που ακολουθεί την ομοιόμορφη κατανομή $U(0,\theta)$, τότε η μέθοδος αποτυγχάνει διότι η κατανομή bootstrap δεν προσεγγίζει ικανοποιητικά την δειγματική κατανομή. Επιπλέον δεδομένα που προέρχονται από κατανομές με βαριές ουρές, όπου η πιθανότητα ακραίων παρατηρήσεων είναι μεγάλη, επηρεάζουν την ποιότητα της μη παραμετρικής bootstrap, ειδικά όταν η παράμετρος που μας ενδιαφέρει είναι η μέση τιμή.

Τέλος όπως έχει αναφερθεί ήδη, έχει προταθεί μια διαφορετική προσέγγιση όταν το μέγεθος του δείγματος είναι πολύ μικρό από τον Hall (1992). Το πρόβλημα έγκειται στο ότι με μικρό μέγεθος δείγματος, λόγω της επανάθεσης πολλές παρατηρήσεις θα επαναλαμβάνονται και πολλά bootstrap δείγματα θα ταυτίζονται, ακόμα και αν επιδιώξουμε να κάνουμε πολλές επαναλήψεις για να λάβουμε την Monte Carlo εκτίμηση. Επιπλέον στη περίπτωση του μικρού δείγματος η εμπειρική συνάρτηση συνήθως δεν είναι μια καλή εκτίμηση της κατανομής του πληθυσμού. (Chernick M. ,2008)

ΠΑΡΑΔΕΙΓΜΑ

Στη συνέχεια του κεφαλαίου θα παραθέσουμε ένα παράδειγμα, στο οποίο θα εφαρμόσουμε τη μέθοδο Bootstrap με τη βοήθεια του στατιστικού πακέτου R. Θα επιχειρήσουμε να εκτιμήσουμε το τυπικό σφάλμα και την μεροληψία για τον δειγματικό μέσο.

Έστω ότι έχουμε ένα δείγμα από υποψηφίους υπαλλήλους μιας πολυεθνικής εταιρείας για τους οποίους γνωρίζουμε τον βαθμό του ψυχομετρικού τεστ στο οποίο υποβλήθηκαν πριν την πρόσληψή τους. Στόχος είναι να προσληφθούν όσοι συγκεντρώσουν τους υψηλότερους βαθμούς. Έστω ότι έχουμε την βαθμολογία από οχτώ υποψήφιους

$$x_1=5.5, x_2=5.9, x_3=6.1, x_4=6.3, x_5=6.4, x_6=6.9, x_7=7.2, x_8=8.6$$

επιλέγω $B=8$ τυχαία δείγματα μεγέθους 8 από τα x_1, \dots, x_8 με επανάθεση.

Για το δείγμα $X_1^*, X_2^*, \dots, X_8^*$ κάθε ένα από τα X_i^* παίρνει τις τιμές x_i με πιθανότητα $1/8$.

Στη συνέχεια θα κατασκευάσουμε bootstrap δείγματα με τη χρήση του στατιστικού πακέτου R. Οι εντολές που θα χρησιμοποιήσουμε είναι οι κάτωθι

```

B<-8
x<-c(5.5,5.9,6.1,6.3,6.4,6.9,7.2,8.6)
output<-matrix(nrow=B,ncol=length(x))
for(i in 1:10){
boot.deigma<-sample(x,size=length(x),replace=TRUE)
output[i,]<-boot.deigma
print(output[i,])}

```

Πίνακας 4.1

Δώσαμε εντολή να 8 επαναλήψεις και προέκυψαν τα ακόλουθα bootstrap δείγματα

Επανάληψη	Δείγμα	$X_1^*, X_2^*, \dots, X_8^*$
1	X^{*1}	6.3, 6.1, 6.3, 8.6, 7.2, 6.4, 6.9, 6.9
2	X^{*2}	6.4, 8.6, 6.3, 6.3, 7.2, 6.4, 6.4, 7.2
3	X^{*3}	5.5, 6.4, 6.4, 5.5, 8.6, 5.9, 6.3, 6.1
4	X^{*4}	5.5, 7.2, 6.3, 5.5, 6.1, 5.9, 7.2, 6.3
5	X^{*5}	7.2, 5.5, 6.3, 6.9, 5.5, 6.3, 6.9, 8.6
6	X^{*6}	5.9, 6.1, 5.5, 6.1, 6.9, 6.3, 6.9, 5.5
7	X^{*7}	6.1, 5.5, 6.1, 7.2, 6.1, 6.9, 6.9, 7.2
8	X^{*8}	8.6, 6.1, 6.4, 6.3, 5.9, 6.9, 6.1, 8.6

Πίνακας 4.2

Με τις ακόλουθες εντολές στον πίνακα 3 θα εκτιμήσουμε την μέση τιμή με τη μέθοδο bootstrap μέσα από 100 δείγματα bootstrap.

```

> x<-c(5.5,5.9,6.1,6.3,6.4,6.9,7.2,8.6)
> B<-numeric(100)
> for(i in 1:100){
+ B[i]<-mean(sample(x,8,replace=TRUE))+ }
> mean(B)
[1] 6.7

```

Πίνακας 4.3

Ο μέσος που προέκυψε είναι 6.7

Στη συνέχεια μέσω της διαδικασίας κατασκευής bootstrap δειγμάτων για B=200 και B=1000 επαναλήψεις θα επιχειρήσουμε να εκτιμήσουμε το τυπικό σφάλμα και την μεροληψία εκτίμησης της μέσης τιμής έτσι ώστε να συγκρίνουμε τα αποτελέσματα .

Οι εντολές που δώσαμε για κάθε περίπτωση επαναλήψεων(200 και 1000) καθώς και τα αποτελέσματα τους δίνονται στους ακόλουθους πίνακες

```
x<-c(5.5,5.9,6.1,6.3,6.4,6.9,7.2,8.6)
B<-200
boot.mean<-function(x,i){mean(x[i])}
b1<-boot(x,statistic=boot.mean,B)
b1
```

Πίνακας 4.4

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:

```
boot(data = x, statistic = boot.mean, R = B)
```

Bootstrap Statistics :

	original	bias	std. error
t1*	6.6125	-0.02625	0.3002407

Πίνακας 4.5

```
x<-c(5.5,5.9,6.1,6.3,6.4,6.9,7.2,8.6)
B<-1000
boot.mean<-function(x,i){mean(x[i])}
b1<-boot(x,statistic=boot.mean,B)
b1
```

Πίνακας 4.6

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:

```
boot(data = x, statistic = boot.mean, R = B)
```

Bootstrap Statistics :

	original	bias	std. error
t1*	6.6125	-0.0003375	0.3189674

Πίνακας 4.7

Από τους πίνακες παρατηρούμε πως με τη μέθοδο bootstrap ο πληθυσμιακός μέσος υποεκτιμάται κατά 0.02625 όταν $B=200$ και κατά 0.0003375 όταν το $B=1000$.

4.4 Εφαρμογή της μεθόδου bootstrap στις μηνιαίες αποζημιώσεις του χαρτοφυλακίου

Στην ενότητα αυτή θα εφαρμόσουμε την μέθοδο επαναδειγματοληψίας bootstrap στα δεδομένα της ασφαλιστικής εταιρείας που εξετάζουμε. Συγκεκριμένα στο κεφάλαιο 3 στη υποενότητα 3.4 μελετήσαμε τις αποζημιώσεις ανά μήνα για τα τρία έτη του χαρτοφυλακίου. Σύμφωνα με αυτές τις 36 παρατηρήσεις (κάθε μια παρατήρηση αντιστοιχεί σε ένα μήνα) θα κατασκευάσουμε ένα διάστημα εμπιστοσύνης μέσα στο οποίο εκτιμάται με μεγάλη

πιθανότητα ότι θα βρίσκεται το μηνιαίο κόστος για την ασφαλιστική εταιρεία. Χρησιμοποιούμε τη μέθοδο bootstrap εφόσον τα δεδομένα δεν προέρχονται από κανονική κατανομή αλλά προσαρμόζονται ικανοποιητικά από την κατανομή Weibull.

Για τις 36 παρατηρήσεις υπολογίσαμε τα περιγραφικά τους στατιστικά και παρατίθενται στον επόμενο πίνακα

Ελάχιστο	1.810.000
1^ο τεταρτημόριο	2.795.000
Διάμεσος	3.298.000
Μέσος	3.471.000
3^ο τεταρτημόριο	4.076.000
Μέγιστο	5.476.000
Διακύμανση	956342834361
Τυπική απόκλιση	977927,8

Οι εντολές που δώσαμε στο στατιστικό πακέτο R είναι οι ακόλουθες

```
CLAIMSPM<-read.table("CLAIMSPM.txt",header=TRUE)
raw.data<-CLAIMSPM[[1]]
raw.data
boot.mean.sampling.distribution = function(x1, B=1000){
boot.dist = vector(length=B)
N= length(x1)
for(i in 1:B){
boot.sample = sample(x=x1, size=N, replace=TRUE)
boot.dist[i] = mean(boot.sample) }
boot.dist = sort(boot.dist)
return(boot.dist) }
boot.mean.sampling.distribution(raw.data)
y<-boot.mean.sampling.distribution(raw.data)
quantile(y,c(0.975))
quantile(y,c(0.025))
```

Με την παραπάνω διαδικασία εξήχθησαν 1000 τιμές (bootstrap δείγματα) που αντιπροσωπεύουν το μηνιαίο κόστος των αποζημιώσεων. Το κάτω και άνω άκρο του 95% διαστήματος εμπιστοσύνης για το μηνιαίο κόστος είναι αντίστοιχα το 2,5% και το 97,5% ποσοστημόριο των 1000 τιμών του δείγματος, δηλαδή (3.149.597 , 3.801.713). Η μέθοδος που χρησιμοποιήσαμε για την εκτίμηση ονομάζεται percentile bootstrap και θεωρείται ως η πιο απλή. Στη συνέχεια θα κατασκευάσουμε πάλι ένα διάστημα εμπιστοσύνης αλλά αυτή τη φορά θα χρησιμοποιήσουμε μια άλλη μέθοδο που ονομάζεται basic bootstrap ή empirical bootstrap.[MIT Open Course Ware]

Ο τύπος που θα χρησιμοποιήσουμε για να υπολογίσουμε το 95% δ.ε. μέσω της basic bootstrap είναι ο κάτωθι

$$[2\bar{x} - \theta_{0.975}, 2\bar{x} - \theta_{0.025}]$$

Όπου \bar{x} = η μέση τιμή των 36 παρατηρήσεων

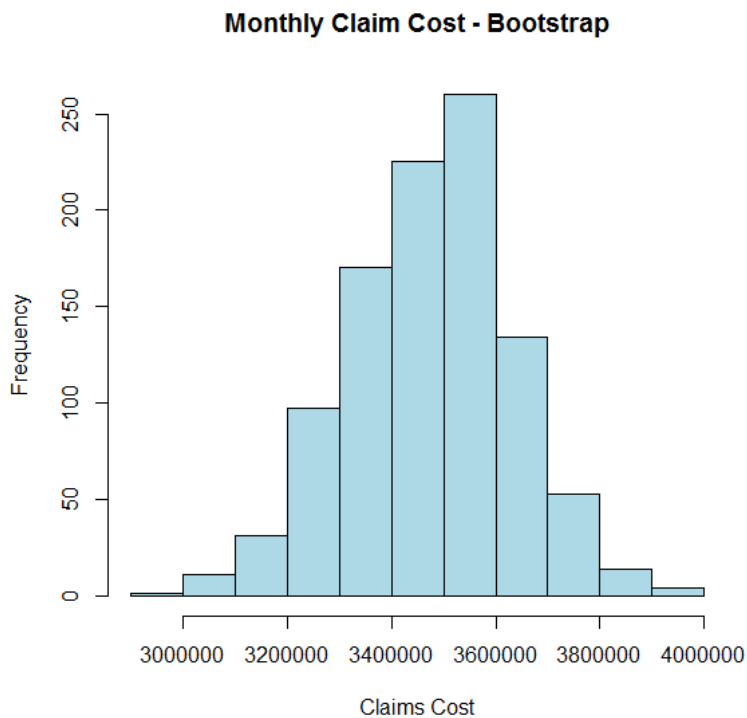
$\theta_{0.975}$ = το 97,5% ποσοστημόριο

$\theta_{0.025}$ = το 2,5% ποσοστημόριο

Με αντικατάσταση προκύπτει το δ.ε. [3.140.287 , 3.792.403]

Και οι δυο μέθοδοι έχουν το ίδιο εύρος διαστήματος εμπιστοσύνης.

Στη συνέχεια παρατίθεται το ιστόγραμμα των 1000 παρατηρήσεων που προέκυψαν από την μέθοδο bootstrap. Παρατηρούμε πως θα οι μηνιαίες αποζημιώσεις φαίνονται να ακολουθούν την κανονική κατανομή.



Στη συνέχεια θα εκτιμήσουμε το 95% διάστημα εμπιστοσύνης θεωρώντας πως τα δεδομένα μας (1000 παρατηρήσεις) προέρχονται από κανονική κατανομή και θα κάνουμε σύγκριση με το αποτέλεσμα που βρήκαμε μέσω της μεθόδου bootstrap.

Αφού θεωρούμε πως τα δεδομένα μας προέρχονται από κανονική κατανομή η δειγματική κατανομή του μέσου είναι κατά προσέγγιση κανονική με μέσο το μ και τυπική απόκλιση

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, η τιμή $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ προέρχεται από τυχαία μεταβλητή που ακολουθεί την τυπική κανονική κατανομή. Αφού μια τιμή της τυπικής κανονικής κατανομής βρίσκεται στο διάστημα $(-Z_{\alpha/2}, Z_{\alpha/2})$ ή $-Z_{\alpha/2} < Z < Z_{\alpha/2}$, με πιθανότητα $1-\alpha$, αντικαθιστώντας το Z με την τιμή του θα έχουμε:

$$-Z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}$$

από την οποία επιλύοντας και τις δυο ανισότητες ως προς μ παίρνουμε:

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (4.1)$$

Μπορούμε δηλαδή να ισχυριστούμε με εμπιστοσύνη $(1-\alpha)100\%$ ότι το διάστημα

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

που καθορίζεται από τα δεδομένα ενός μεγάλου δείγματος περιλαμβάνει τον μέσο του πληθυσμού που θέλουμε να εκτιμήσουμε. Αν η τυπική απόκλιση σ του πληθυσμού είναι άγνωστη αντικαθίσταται από τη δειγματική τυπική απόκλιση. Στην περίπτωση αυτή αντί για την κανονική κατανομή χρησιμοποιούμε την κατανομή t του student. Ωστόσο εδώ επειδή το δείγμα μας είναι μεγάλο, οι διαφορές ανάμεσα στις δυο κατανομές είναι πολύ μικρές. Το παραπάνω διάστημα καλείται διάστημα εμπιστοσύνης ενώ τα άκρα του λέγονται όρια εμπιστοσύνης και η τιμή $1-\alpha$ βαθμός εμπιστοσύνης. Οι συνήθεις τιμές του βαθμού εμπιστοσύνης είναι 0,95 και 0,99. Οι εκτιμήσεις που δίνονται με τη μορφή διαστήματος τιμών λέγονται εκτιμήσεις διαστήματος.

Για τις 1000 παρατηρήσεις θα εφαρμόσουμε τα παραπάνω. Ο έλεγχος που θα πραγματοποιήσουμε θα γίνει σε επίπεδο σημαντικότητας $\alpha=5\%$. Η τιμή του $Z_{0,025}$ την υπολογίσαμε από τους πίνακες της τυπικής κανονικής κατανομής.

Έχουμε υπολογίσει τα περιγραφικά στατιστικά των παρατηρήσεων αυτών και παρατίθενται στον ακόλουθο πίνακα

Ελάχιστο	2.951.000
1^ο τεταρτημόριο	3.369.000
Διάμεσος	3.484.000
Μέσος	3.475.000
3^ο τεταρτημόριο	3.576.000
Μέγιστο	3.977.000
Διακύμανση	24821815765
Τυπική Απόκλιση	157549,4

Με αντικατάσταση στην σχέση 4.1 προκύπτουν τα ακόλουθα

$$3475000 - 1,96 \frac{157549,4}{\sqrt{1000}} < \mu < 3475000 + 1,96 \frac{157549,4}{\sqrt{1000}}$$
$$3.465.235 < \mu < 3.484.765$$

Οπώς παρατηρούμε το παραπάνω διάστημα εμπιστοσύνης είναι πολύ πιο ‘στενό’ από το προηγούμενο, εφόσον εδώ χρησιμοποιείται παραμετρική μέθοδος.

Εν κατακλείδι μπορούμε να αναφέρουμε πως τα διαστήματα εμπιστοσύνης που κατασκευάσαμε με τη μέθοδο bootstrap για το μηνιαίο κόστος των αποζημιώσεων μπορούν να συμβάλλουν στις προβλέψεις των ζημιών.

ΚΕΦΑΛΑΙΟ 5

Γενικευμένα Γραμμικά Μοντέλα

5.1 Εισαγωγή

Στο κεφάλαιο αυτό αρχικά γίνεται αναφορά στην παλινδρόμηση απλή και πολλαπλή. Στη συνέχεια γίνεται εισαγωγή στα γενικευμένα γραμμικά μοντέλα και αναφέρονται τα βασικά χαρακτηριστικά τους καθώς και η συμβολή τους στην επεξεργασία ασφαλιστικών δεδομένων. Στο τέλος του κεφαλαίου γίνεται εφαρμογή των γενικευμένων γραμμικών μοντέλων στο χαρτοφυλάκιο της εταιρείας. Δημιουργήσαμε τρία διαφορετικά μοντέλα όπου για το καθένα είδαμε ποιες μεταβλητές είναι σημαντικές. Η κατασκευή των μοντέλων έγινε με τη χρήση του στατιστικού πακέτου R.

5.2 Παλινδρόμηση

Σε διάφορα προβλήματα της Στατιστικής το ενδιαφέρον μας εστιάζεται στην ταυτόχρονη μελέτη δύο ή περισσότερων μεταβλητών, για να προσδιορίσουμε με ποιο τρόπο οι μεταβλητές αυτές σχετίζονται μεταξύ τους. Για παράδειγμα

- το ύψος του ασφάλιστρου που καλείται να πληρώσει ένας ασφαλισμένος σχετίζεται με την ηλικία του, το βαθμό επικινδυνότητας του επαγγέλματός του κτλ.
- το ύψος των απαιτήσεων που εγείρονται σε μια ασφαλιστική εταιρεία από ζημιές που προκαλούνται από τους ασφαλισμένους σε αυτή οδηγούς συνδέεται με το χρόνο εμπειρίας του οδηγού

Ο κλάδος της Στατιστικής που εξετάζει τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών με απώτερο στόχο την πρόβλεψη μιας απ' αυτές μέσω των άλλων λέγεται ανάλυση παλινδρόμησης (regression analysis). Ο όρος "regression" χρησιμοποιήθηκε για πρώτη φορά από τον Άγγλο ανθρωπολόγο Galton (1822- 1911) το 1885.

Απλή και πολλαπλή Ανάλυση παλινδρόμησης

Η παλινδρόμηση στην οποία υπάρχει μόνο μια ανεξάρτητη μεταβλητή καλείται απλή παλινδρόμηση ενώ αν υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές λέγεται πολλαπλή παλινδρόμηση.

Ορολογία μεταβλητών

Χρησιμοποιούμε τον όρο εξαρτημένη μεταβλητή για τις μετρήσεις που θεωρούμε σαν τυχαίες μεταβλητές. Οι μεταβλητές αυτές θεωρούνται ότι μεταβάλλονται ελεύθερα εν αντιθέσει με τις ανεξάρτητες μεταβλητές οι οποίες θεωρούνται σαν μη τυχαίες μεταβλητές (δηλαδή παίρνουν συγκεκριμένες τιμές ανάλογα με τον σχεδιασμό του πειράματος).

Οι μεταβλητές μπορεί να ταξινομηθούν σαν

- ✓ κατηγορικές ή ποιοτικές

- ✓ διάταξης για τις οποίες υπάρχει κάποια φυσική διάταξη μεταξύ των κατηγοριών
- ✓ Συνεχείς μεταβλητές όπου οι παρατηρήσεις μπορούν να πάρουν οποιαδήποτε τιμή σε κάποιο διάστημα.

Μια ποιοτική ανεξάρτητη μεταβλητή καλείται παράγοντας (factor) και οι κατηγορίες λέγονται επίπεδα (level) του παράγοντα. Οι συνεχείς ανεξάρτητες μεταβλητές λέγονται covariates.

Απλή γραμμική παλινδρόμηση

Η απλούστερη περίπτωση παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση (simple linear regression), κατά την οποία χρησιμοποιούμε μόνο μια μεταβλητή X , και μια δεύτερη μεταβλητή Y η οποία μπορεί να προσεγγιστεί ικανοποιητικά από μία γραμμική συνάρτηση του X .

X : ανεξάρτητη μεταβλητή (independent or input variable)

Y : εξαρτημένη μεταβλητή (dependent or response variable)

5.3 Γενικά Χαρακτηριστικά ενός Γενικευμένου Γραμμικού Μοντέλου

Γενικευμένα Γραμμικά μοντέλα

Τα γενικευμένα γραμμικά μοντέλα είναι μια μεθοδολογία για την μοντελοποίηση των σχέσεων μεταξύ των μεταβλητών. Τα Γ.Γ.Μ. γενικεύουν το κλασικό κανονικό μοντέλο, χαλαρώνοντας ορισμένες από τις περιοριστικές υποθέσεις του και παρέχουν μεθόδους για την ανάλυση μη κανονικών δεδομένων. Τα εργαλεία χρονολογούνται από το αρχικό άρθρο των Nelder και Wedderburn (1972) και από τότε έχουν γίνει μέρος των βασικών στατιστικών στοιχείων που χρησιμοποιούνται σε πολλούς διαφορετικούς τομείς εφαρμογών. Με τα Γ.Γ.Μ η μεταβλητότητα μιας μεταβλητής ερμηνεύεται από τις μεταβολές των άλλων μεταβλητών του μοντέλου. Η μεταβλητή που ερμηνεύεται ονομάζεται "εξαρτημένη" ή "απόκρισης" μεταβλητή ενώ οι υπόλοιπες "επεξηγηματικές". Αναφέρονται επίσης και ως "παράγοντες κινδύνου". Το μοντέλο επεξηγεί την σύνδεση μεταξύ της μεταβλητής "απόκρισης" και των "επεξηγηματικών" μεταβλητών. Η στατιστική μοντελοποίηση στα γενικά αλλά και στα γενικευμένα γραμμικά μοντέλα ιδίως είναι η τέχνη ή η επιστήμη του σχεδιασμού της προσαρμογής και της ερμηνείας ενός μοντέλου. Ένα στατιστικό μοντέλο είναι τόσο καλό όσο και τα δεδομένα από τα οποία προέρχεται. Συνεπώς η καλή ερμηνεία και κατανόηση των δεδομένων είναι σημαντικό σημείο εκκίνησης για την μοντελοποίηση. Απαιτείται σημαντικός χρόνος για την εκκαθάριση των δεδομένων και την διερεύνησή τους. (de Jong, P and Heller, G.,2008)

Τα ασφαλιστικά δεδομένα συνήθως οργανώνονται σε μια σειρά αμφίδρομη ανάλογα με τις περιπτώσεις και τις μεταβλητές. Οι περιπτώσεις μπορεί να είναι τα ασφαλιστήρια συμβόλαια, οι απαιτήσεις ή τα ατυχήματα.

Οι μεταβλητές μπορεί να είναι το επίπεδο της ζημιάς , το φύλλο , το κόστος του νομίσματος , αν υπάρχει νομική εκπροσώπηση και άλλα ακόμα .Οι μεταβλητές μπορεί να είναι ποσοτικές ή ποιοτικές . Στη συνέχεια θα προσδιορίσουμε κάποιους τύπους μεταβλητών .

ΥΨΟΣ ΖΗΜΙΑΣ : Είναι ένα παράδειγμα του τι συνήθως μπορεί να θεωρηθεί ως συνεχής μεταβλητή. Οι συνεχείς μεταβλητές ονομάζονται και "διαστήματα" εξαιτίας του ότι μπορούν να πάρουν τιμές σε ένα διάστημα της πραγματικής γραμμής .

NΟΜΙΚΗ ΕΚΠΡΟΣΩΠΗΣΗ : Πρόκειται για μια κατηγορική μεταβλητή με δυο επίπεδα "όχι" ή "ναι". Οι μεταβλητές που είναι δίτιμες καλούνται και "Bernoulli" μεταβλητές. Οι δίτιμες μεταβλητές υποδεικνύουν την παρουσία ή την απουσία ενός χαρακτηριστικού, την εμφάνιση ή μη μιας εκδήλωσης ενδιαφέροντος όπως μια απαίτηση ή ένα θάνατο .

ΕΠΙΠΕΔΟ ΖΗΜΙΑΣ : Είναι μια κατηγορική ή ποιοτική μεταβλητή που λαμβάνει τιμές ανάλογα με το επίπεδο της σωματικής βλάβης . Ξεκινά από χαμηλό επίπεδο και καταλήγει σε θάνατο . Οι κατηγορικές μεταβλητές προσδιορίζονται από ένα διακριτό σύνολο τιμών ονομαστικού χαρακτήρα .

Τα δεδομένα συχνά μετατρέπονται σε μετρήσεις ή συχνότητες. Παράδειγμα μεταβλητών μέτρησης είναι : ο αριθμός των απαιτήσεων για τα ασφαλιστικά συμβόλαια ενός έτους , ο αριθμός των ατυχημάτων σε μια εβδομάδα , ο αριθμός των παιδιών σε μια οικογένεια , ο αριθμός θανάτων ενός πληθυσμού.

Τα Γ.Γ.Μ. χρησιμοποιούνται για την αξιολόγηση και την ποσοτικοποίηση της σχέσης μεταξύ της μεταβλητής απόκρισης και των επεξηγηματικών μεταβλητών.

Σε σχέση με τα συνήθη μοντέλα παλινδρόμησης, τα ΓΓΜ διαφέρουν κατά δυο έννοιες

- Η κατανομή της μεταβλητής απόκρισης δεν είναι κανονική αλλά επιλέγεται από την εκθετική οικογένεια κατανομών
- Ο μετασχηματισμός του μέσου της μεταβλητής απόκρισης σχετίζεται γραμμικά με τις επεξηγηματικές μεταβλητές

Τα γενικευμένα γραμμικά μοντέλα είναι σημαντικά και για την ανάλυση των ασφαλιστικών δεδομένων. Στα ασφαλιστικά δεδομένα οι παραδοχές του κανονικού μοντέλου δεν είναι συχνά εφαρμόσιμες. Για παράδειγμα το ύψος της ζημιάς, η συχνότητα της ζημιάς, η πιθανότητα να συμβεί ζημιά για ένα ασφαλιστήριο συμβόλαιο δεν μπορούν να προσδιοριστούν από την κανονική κατανομή.

Το κλασικό μοντέλο παλινδρόμησης με k ερμηνευτικές μεταβλητές:

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon_i$$

Η ισοδύναμα,

$$E(Y_i) = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

- Γραμμική σχέση ανάμεσα στη μέση απόκριση και τις ερμηνευτικές μεταβλητές.
- Βασική υπόθεση στο μοντέλο είναι ότι $\varepsilon_i \sim N(0, \sigma^2)$ και τα ε_i (σφάλματα) ανεξάρτητα ανά δύο.
- Αυτό σημαίνει ότι και τα Y_i ακολουθούν κανονική κατανομή με σταθερή διακύμανση.

Υπάρχουν όμως περιπτώσεις όπου οι παραπάνω υποθέσεις δεν ισχύουν. Λύση για αυτό επιχειρείται να δοθεί μέσω των γενικευμένων γραμμικών μοντέλων.

Το 1972 οι Nelder & Wedderburn παρουσίασαν μια ενοποιημένη θεωρία για γραμμικά μοντέλα, που δεν απαιτεί την υπόθεση της κανονικότητας για τη μεταβλητή απόκρισης. Σύμφωνα με αυτήν, τα γραμμικά μοντέλα μπορούν να μελετηθούν ενιαία κάτω από την υπόθεση ότι η κατανομή της μεταβλητής απόκρισης ανήκει στην εκθετική οικογένεια κατανομών.

Εκθετική οικογένεια κατανομών

Μία κατανομή πιθανότητας λέμε ότι ανήκει στην εκθετική οικογένεια κατανομών όταν η συνάρτηση πιθανότητας (ή πυκνότητας αν η κατανομή είναι συνεχής) της κατανομής μπορεί να γραφεί στη μορφή

$$f_Y(y; \theta; \varphi) = \exp \left[\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right]$$

όπου a, b, c είναι τρεις γνωστές συναρτήσεις, ενώ οι θ, φ είναι παράμετροι.

- Αν το φ είναι γνωστό, τότε έχουμε την εκθετική οικογένεια με μία παράμετρο και το θ αναφέρεται ως η κανονική παράμετρος (canonical parameter) της κατανομής.

• Αν το φ δεν είναι γνωστό, τότε μπορούμε σε πολλές περιπτώσεις να το θεωρήσουμε σαν μία παράμετρο κλίμακας για την κατανομή, οπότε αποκαλείται παράγοντας όχλησης (nuisance factor) της κατανομής.

Συναρτήσεις σύνδεσης (link functions)

Σε ένα ΓΓΜ, μία συνάρτηση σύνδεσης g είναι μια συνάρτηση η οποία συνδέει

- ❖ το στοχαστικό τμήμα του μοντέλου (μέση τιμή της τ.μ. Y)
- ❖ με το μη στοχαστικό τμήμα (γραμμικός συνδυασμός των ερμηνευτικών μεταβλητών X_j).

Συγκεκριμένα, έστω $\mu_i = E(Y_i)$ η μέση τιμή της μεταβλητής απόκρισης. Υποθέτουμε ότι αυτή εξαρτάται από τις τιμές των X_j για $j=1,2,\dots,k$

Θεωρούμε τη γραμμική συνάρτηση πρόβλεψης

$$n_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

όπου X_{ij} είναι η τιμή της μεταβλητής X_i για την παρατήρηση i .

Τότε η συνάρτηση σύνδεσης συνδέει τη μέση τιμή της μεταβλητής απόκρισης με την παραπάνω συνάρτηση πρόβλεψης,

$$n_i = g(\mu_i) = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

Η g θεωρούμε ότι είναι πάντα μία συνάρτηση

- μονότονη, και
- διαφορίσιμη.

(Κ. Πολίτης, 2015)

Πλεονεκτήματα των ΓΓΜ έναντι της συνήθους παλινδρόμησης

- ✓ Έχουν πολύ μεγαλύτερο φάσμα εφαρμογών. Χρησιμοποιούνται και σε περιπτώσεις όπου δεν μπορεί να υποτεθεί ότι η κατανομή της Y είναι κανονική, ούτε καν προσεγγιστικά – π.χ. όταν η Y παίρνει μόνο δύο τιμές (Bernoulli)

- ✓ Οι εκτιμητές των παραμέτρων προκύπτουν με τη μέθοδο μέγιστης πιθανοφάνειας – άρα έχουν μια σειρά από επιθυμητές ιδιότητες.
- ✓ Στις περισσότερες περιπτώσεις, δε χρειάζεται να υποθέσουμε σταθερή διακύμανση για τις τιμές της Y .
- ✓ Με την ενοποιημένη θεωρία των ΓΓΜ, δε χρειάζεται να χρησιμοποιήσουμε διαφορετικό μοντέλο ανάλογα με το αν οι ερμηνευτικές μεταβλητές είναι ποσοτικές ή ποιοτικές (ή μείξη των δύο).
- ✓ Στην περίπτωση που όλες οι μεταβλητές είναι κατηγορικές, τα ΓΓΜ αποτελούν ένα βασικό τρόπο ανάλυσης σε πίνακες συνάφειας. (de Jong, P. and Heller, G.,2008)

Σε πολλές περιπτώσεις η μεταβλητή απόκρισης σε ένα μοντέλο που μας ενδιαφέρει είναι διακριτή.

Απλούστερο παράδειγμα διακριτής κατανομής από την εκθετική οικογένεια είναι η διωνυμική κατανομή.

Στο πλαίσιο ενός ΓΓΜ, μία διάκριση που μας ενδιαφέρει είναι όταν

- Τα δεδομένα μας είναι ομαδοποιημένα (οπότε μιλάμε συνήθως για διωνυμικά δεδομένα, binomial data) και
- Όταν δεν είναι ομαδοποιημένα, οπότε γνωρίζουμε για κάθε άτομο στο δείγμα την τιμή της απόκρισης (0=αποτυχία, 1= επιτυχία). Τότε μιλάμε για δίτιμα δεδομένα (binary data).

Τρεις συναρτήσεις σύνδεσης που χρησιμοποιούμε για δίτιμα (ή διωνυμικά) δεδομένα:

1. Logit $n_i = \text{logit}(p_i) = \log \left[\frac{p_i}{1-p_i} \right]$

2. Probit $n_i = \text{probit}(p_i) = \Phi^{-1}(p_i)$, όπου Φ είναι η αθροιστική συνάρτηση κατανομής της τυποποιημένης κανονικής.

3. Complementary log –log, $n_i = \log[-\log(1-p_i)]$

Όταν χρησιμοποιούμε τη συνάρτηση σύνδεσης τη συνάρτηση logit, τότε πρόκειται για μοντέλο λογιστικής παλινδρόμησης. Τα μοντέλα με τις συναρτήσεις σύνδεσης logit και probit δίνουν αρκετά παρόμοια αποτελέσματα, τα οποία (με μια αλλαγή κλίμακας) μπορούν να γίνουν σχεδόν ταυτόσημα. (Κ. Πολίτης, 2015)

Εκτίμηση παραμέτρων

Όταν η μεταβλητή απόκρισης σε ένα ΓΓΜ δεν είναι κανονική, οι εκτιμητές των παραμέτρων βί που προκύπτουν με τις μεθόδους

- Ελαχίστων τετραγώνων
- Μέγιστης πιθανοφάνειας (εμπ) δεν είναι γενικά οι ίδιοι.

Αυτό έχει σημαντικές επιπτώσεις όσον αφορά στις ιδιότητες των εκτιμητών, π.χ. οι εκτιμητές ελαχίστων τετραγώνων μπορεί να μην είναι γενικά αμερόληπτοι ή συνεπείς.

Το στατιστικό πακέτο R βρίσκει τους εκτιμητές μέγιστης πιθανοφάνειας με μία επαναληπτική (αριθμητική) μέθοδο (iterative reweighted least squares, IRLS). Ο αριθμός των επαναλήψεων και το βήμα που χρησιμοποιείται για τον τερματισμό του αλγορίθμου μπορούν να καθοριστούν από το χρήστη (οι προεπιλογές είναι 50 και 0.01 αντίστοιχα). (Κ. Πολίτης, 2015)

5.4 Εφαρμογή στα δεδομένα

Για την κατασκευή των γενικευμένων γραμμικών μοντέλων θα χρησιμοποιήσουμε κάποια από τα δεδομένα της ασφαλιστικής εταιρείας. Συγκεκριμένα θέλουμε να εξάγουμε συμπεράσματα για τις Σωματικές Βλάβες, τα Θανατηφόρα ατυχήματα καθώς και τις Ακραίες Ζημιές. Τα παραπάνω έχουν τον ρόλο της μεταβλητής Y στα μοντέλα μας ενώ οι ερμηνευτικές μεταβλητές X είναι η χρήση του οχήματος, η ηλικία και η περιφέρεια. Η μεταβλητή της Ηλικίας (age) κατασκευάστηκε από την ηλικία γέννησης του οδηγού του οχήματος, όπου αυτή η πληροφορία ήταν διαθέσιμη. Στα αρχικά δεδομένα είχαμε 13 περιφέρειες όπου τις συμπύξαμε σε 6 επειδή διαπιστώσαμε από την αρχική μας ανάλυση πως δεν παρουσιάζουν σημαντικές διαφοροποιήσεις επειδή κάποιες περιφέρειες είναι γειτονικές μεταξύ τους. Τέλος όμως για την καλύτερη προσαρμογή στο μοντέλο χρειάστηκε να τις κατηγοριοποιήσουμε σε 3 .

Ακολουθούν οι πίνακες με τις περιφέρειες

ΒΟΡΕΙΑ ΕΛΛΑΔΑ (NORTH_GR)	ΚΕΝΤΡΙΚΗ ΕΛΛΑΔΑ (CENTRAL_GR)	ΑΤΤΙΚΗ	ΑΙΓΑΙΟ (AIGAION)	ΙΟΝΙΟ	ΝΟΤΙΑ ΕΛΛΑΔΑ (SOUTH_GR)
ΑΝΑΤΟΛΙΚΗ ΜΑΚΕΔΟΝΙΑ ΚΑΙ ΘΡΑΚΗ	ΔΥΤΙΚΗ ΕΛΛΑΔΑ		ΒΟΡΕΙΟ		ΠΕΛΛΟΠΟΝΗΣΟΣ
ΔΥΤΙΚΗ ΜΑΚΕΔΟΝΙΑ	ΗΠΕΙΡΟΣ		ΝΟΤΙΟ		ΚΡΗΤΗ
ΚΕΝΤΡΙΚΗ ΜΑΚΕΔΟΝΙΑ	ΘΕΣΣΑΛΙΑ				
	ΣΤΕΡΕΑ ΕΛΛΑΔΑ				

Πίνακας 5.1

ΑΤΤΙΚΗ(ΑΤΤΙΚΗ)	ΒΟΡΕΙΑ ΕΛΛΑΔΑ(NORTH_GR)	ΝΟΤΙΑ ΕΛΛΑΔΑ(SOUTH_GR)
	ΚΕΝΤΡΙΚΗ ΕΛΛΑΔΑ(CENTRAL.GR)	ΑΙΓΑΙΟ(AIGAION)
		ΙΟΝΙΟ(IONIAN)

Πίνακας 5.2

Μετά από την επεξεργασία των δεδομένων μας καταλήξαμε να δημιουργήσουμε τρία μοντέλα. Στο Α' μοντέλο η μεταβλητή απόκρισης είναι το εάν είχαμε σωματική βλάβη ή όχι στις ζημιές μας (modelSV) ενώ οι επεξηγηματικές μεταβλητές είναι η χρήση του οχήματος, η ηλικία του οδηγού και η περιφέρεια.

Στο Β' μοντέλο η μεταβλητή απόκρισης είναι τα θανατηφόρα ατυχήματα δηλαδή το εάν είχαμε θάνατο ή όχι (modelDEATH) ενώ οι επεξηγηματικές μεταβλητές είναι η χρήση του οχήματος, η ηλικία του οδηγού και η περιφέρεια.

Στο Γ' μοντέλο η μεταβλητή απόκρισης είναι το εάν είχαμε ακραία ζημιά ή όχι (modelCOST). Ως ακραία ζημιά ορίζουμε το ποσό της αποζημίωσης να είναι πάνω από 200.000€. Ενώ οι επεξηγηματικές μεταβλητές είναι η χρήση του οχήματος, η ηλικία του οδηγού και η περιφέρεια.

Στη συνέχεια ακολουθεί η ανάλυση για το κάθε μοντέλο.

Πίνακας ανάλυσης απόκλισης για την ένδειξη Σωματικής Βλάβης

Μοντέλο	Απόκλιση	Βαθμοί Ελευθερίας	Μεταβολή στην απόκλιση	Μεταβολή στους β.ε.	Στατιστικά σημαντικό σε ε.σ. 5 %
1 (constant)	19690	48563			
ΧΡΗΣΗ	19523	48557	167,125	6	ΝΑΙ
AGE	19521	48556	1,926	1	ΟΧΙ
PERIFEREIA	19508	48551	13,374	5	ΝΑΙ

Πίνακας 5.3

Πίνακας ανάλυσης απόκλισης για την ένδειξη Θανατηφόρου Ατυχήματος

Μοντέλο	Απόκλιση	Βαθμοί Ελευθερίας	Μεταβολή στην απόκλιση	Μεταβολή στους β.ε.	Στατιστικά σημαντικό σε ε.σ. 5 %
1 (constant)	1905,6	48563			
ΧΡΗΣΗ	1888,8	48557	16,8368	6	ΝΑΙ
AGE	1888,1	48556	0,6852	1	ΟΧΙ
PERIFEREIA	1882,7	48551	5,4262	5	ΟΧΙ

Πίνακας 5.4

Πίνακας ανάλυσης απόκλισης για την ένδειξη Ακραίας Ζημιάς

Μοντέλο	Απόκλιση	Βαθμοί Ελευθερίας	Μεταβολή στην απόκλιση	Μεταβολή στους β.ε.	Στατιστικά σημαντικό σε ε.σ. 5 %
1 (constant)	1146,7	48563			
ΧΡΗΣΗ	1139,9	48557	6,7861	6	ΟΧΙ
AGE	1138,4	48556	1,4691	1	ΟΧΙ
PERIFEREIA	1134,2	48551	4,2748	5	ΟΧΙ

Πίνακας 5.5

Από τους παραπάνω πίνακες ανάλυσης απόκλισης (apova) για κάθε μοντέλο προκύπτουν συμπεράσματα για το ποια μεταβλητή είναι σημαντική στην κάθε περίπτωση. Συγκεκριμένα εξετάζοντας το πρώτο μοντέλο για την σωματική βλάβη ,δηλαδή θεωρώντας επιτυχία το εάν ένα τροχαίο ατύχημα έχει σωματική βλάβη παρατηρούμε πως στατιστικά σημαντικές είναι οι μεταβλητές χρήση του οχήματος και η περιφέρεια, αντιθέτως η ηλικία του κάθε οδηγού δεν είναι σημαντική.

Σχετικά με το μοντέλο των θανατηφόρων ατυχημάτων, δηλαδή θεωρώντας επιτυχία να συμβεί θάνατος σε ένα τροχαίο ατύχημα καταλήγουμε πως στατιστικά σημαντική μεταβλητή είναι η χρήση ενώ η περιφέρεια και η ηλικία δεν είναι.

Τέλος από τον πίνακα 5.3 που παρουσιάζει το μοντέλο της ακραίας ζημιάς δηλαδή θεωρούμε επιτυχία το αν κάποια αποζημίωση είναι πάνω από 200.000 ευρώ προκύπτει πως καμία μεταβλητή δεν είναι στατιστικά σημαντική άρα δεν προβλέπεται για το μοντέλο αυτό παραπάνω ανάλυση.

Οι παραπάνω έλεγχοι που πραγματοποιήθηκαν παρουσιάζονται στο πίνακα 1 στο παράρτημα 3.

Αναφορικά με τους πίνακες (2 και 3) των συντελεστών των μοντέλων που βρίσκονται στο παράρτημα 3, μπορούμε να σχολιάσουμε αναλυτικά κάθε μεταβλητή του μοντέλου.

Μοντέλο Σωματικής βλάβης (SV)

Στο μοντέλο αυτό για την μεταβλητή χρήση έχουμε σαν επίπεδο αναφοράς τα αγροτικά οχήματα (επιλέγεται αυτόματα από την R το επίπεδο αναφοράς). Προκύπτει πως όλες οι υπόλοιπες κατηγορίες (ΕΙΧ, ΕΝΟΙΚ., ΦΔΧ,ΛΟΙΠΑ,ΜΟΤΟ) έχουν περισσότερες σωματικές βλάβες από τα αγροτικά ενώ τα φορτηγά (ΦΙΧ) έχουν λιγότερες. Από τον πίνακα προκύπτει πως οι χρήσεις που είναι στατιστικά σημαντικές για το μοντέλο είναι τα επιβατικά (ΕΙΧ), τα ενοικιαζόμενα (ΕΝΟΙΚ.) ,τα φορτηγά (ΦΙΧ) και οι μοτοσυκλέτες (ΜΟΤΟ). Όσον αφορά την μεταβλητή περιφέρειας εδώ το επίπεδο αναφοράς είναι το Αιγαίο. Όπως προκύπτει από τον πίνακα η Αττική έχει τις περισσότερες σωματικές βλάβες από το Αιγαίο ενώ ο Βορράς σε σύγκριση με το Αιγαίο τις λιγότερες. Επειδή όμως βλέπουμε ότι όλες οι άλλες περιφέρειες εκτός από τον Βορρά δεν διαφέρουν σημαντικά από το Αιγαίο εξετάσαμε μια ακόμα διαφοροποίηση σε 3 περιφέρειες.

Μοντέλο θανατηφόρων ατυχημάτων

Στο μοντέλο των θανατηφόρων με επίπεδο αναφοράς πάλι τα αγροτικά οχήματα προκύπτει πως όλες οι άλλες κατηγορίες έχουν περισσότερα θανατηφόρα ατυχήματα. Στατιστικά σημαντικές είναι οι χρήσεις ΦΔΧ και ΛΟΙΠΑ.

Ανάλυση μοντέλων με 3 περιφέρειες

Πίνακας ανάλυσης απόκλισης για την ένδειξη Σωματικής Βλάβης

Μοντέλο	Απόκλιση	Βαθμοί Ελευθερίας	Μεταβολή στην απόκλιση	Μεταβολή στους β.ε.	Στατιστικά σημαντικό σε ε.σ. 5 %
1 (constant)	19690	48563			
ΧΡΗΣΗ	19523	48557	167,125	6	ΝΑΙ
AGE	19521	48556	1,926	1	ΟΧΙ
PERIFEREIA	19515	48554	6,719	2	ΝΑΙ

Πίνακας 5.6

Πίνακας ανάλυσης απόκλισης για την ένδειξη Θανατηφόρου Ατυχήματος

Μοντέλο	Απόκλιση	Βαθμοί Ελευθερίας	Μεταβολή στην απόκλιση	Μεταβολή στους β.ε.	Στατιστικά σημαντικό σε ε.σ. 5 %
1 (constant)	1905,6	48563			
ΧΡΗΣΗ	1888,8	48557	16,8368	6	ΝΑΙ
AGE	1888,1	48556	0,6852	1	ΟΧΙ
PERIFEREIA	1884,3	48554	3,7719	2	ΟΧΙ

Πίνακας 5.7

Αφού δημιουργήσαμε 3 νέες κατηγορίες για την μεταβλητή της περιφέρειας από τους παραπάνω πίνακες ανάλυσης απόκλισης για το μοντέλο σωματικής βλάβης στατιστικά σημαντικές είναι πάλι η χρήση και η περιφέρεια. Για το μοντέλο των θανατηφόρων ατυχημάτων προκύπτει σημαντική μόνο η χρήση.

Από τους πίνακες ανάλυσης των συντελεστών των μοντέλων (παράρτημα 3) για την Σωματική βλάβη προκύπτει πως στην Αττική, που είναι τώρα το επίπεδο αναφοράς, έχουμε περισσότερες σωματικές βλάβες από ότι στην Βόρεια Ελλάδα και στην Νότια. Επίσης προκύπτει πως στατιστικά σημαντική είναι η Βόρεια Ελλάδα.

Όσον αφορά το μοντέλο των θανατηφόρων ατυχημάτων για την μεταβλητή χρήση που είναι η μόνη στατιστικά σημαντική όλες οι κατηγορίες έχουν περισσότερα θανατηφόρα από τα αγροτικά οχήματα(που είναι το επίπεδο αναφοράς. Στατιστικά σημαντικές όμως είναι τα ΦΔΧ και τα ΛΟΙΠΑ οχήματα.

Τελικά μοντέλα

Λαμβάνοντας υπόψη όλα τα παραπάνω καταλήξαμε στα τελικά μοντέλα τα οποία περιέχουν μόνο τις σημαντικές μεταβλητές. Ακολουθούν οι πίνακες με την ανάλυσή τους .

Πίνακας ανάλυσης απόκλισης για την ένδειξη Σωματικής Βλάβης

Μοντέλο	Απόκλιση	Βαθμοί Ελευθερίας	Μεταβολή στην απόκλιση	Μεταβολή στους β.ε.	Στατιστικά σημαντικό σε ε.σ. 5 %
1 (constant)	24538	62391			
ΧΡΗΣΗ	24324	62385	214.033	6	ΝΑΙ
PERIFEREIA	24293	62383	31.304	2	ΝΑΙ

Πίνακας 5.8

Πίνακας ανάλυσης συντελεστών για την ένδειξη Σωματικής Βλάβης

	Estimate	Std.Error	Z value	Pr(> z)
(Intercept)	-3,03795	0,08415	-36,103	<2e-16***
XRHSHEIX	0,23347	0,07543	3,095	0,00197**
XRHSHENOIK	0,79613	0,10675	7,458	8,77e14***
XRHSHFDX	0,11401	0,12752	0,894	0,37131
XRHSHFIX	-0,33839	0,09644	-3,509	0,00045***
XRHSHLOIPA	0,05847	0,12719	0,460	0,64572
XRHSHMOTO	0,56551	0,08939	6,327	2,51e-10***
PERIFEREIA NORTH_GR	-0,23338	0,05074	-4,599	4,24e-06***
PERIFEREIA SOUTH_GR	-0,03319	0,05388	-0,616	0,53784

Πίνακας 5.9

Το μοντέλο όπως προκύπτει από παραπάνω είναι το εξής

$$\log \left[\frac{p_i}{1-p_i} \right] = -3,03795 + 0,23EIX + 0,79613ENOIK. + 0,11401FDX - 0,33839FIX + 0,05847LOIPA + 0,5655MOTO - 0,23338NORTH_GR - 0,03319 SOUTH_GR$$

Η πρώτη παρατήρηση που θα πρέπει να κάνουμε για τον πίνακα ανάλυσης απόκλισης (apona) είναι η σημαντική διαφορά στις τιμές της απόκλισης και στους βαθμούς ελευθερίας σε σχέση με τα προηγούμενα μοντέλα. Αυτό οφείλεται στο γεγονός πως στο τελικό μοντέλο πλέον δεν υπάρχει η μεταβλητή της ηλικίας (age). Από τους παραπάνω πίνακες (5.8, 5.9) για την μεταβλητή χρήση έχοντας πάλι ως επίπεδο αναφοράς τα αγροτικά οχήματα παρατηρούμε πως όλες οι άλλες κατηγορίες έχουν περισσότερες σωματικές βλάβες ενώ τα ΦΙΧ λιγότερες. Στατιστικά σημαντικές στο τελικό μοντέλο είναι οι χρήσεις EIX, ENOIK., ΦΙΧ, ΜΟΤΟ. Σχετικά με την άλλη σημαντική μεταβλητή του μοντέλου την περιφέρεια προκύπτει πως η Βόρεια Ελλάδα και η Νότια Ελλάδα έχουν περισσότερες σωματικές βλάβες από την περιφέρεια της Αττικής που είναι το επίπεδο αναφοράς. Τέλος, στατιστικά σημαντική είναι η περιφέρεια της Βόρειας Ελλάδας.

Από τον πίνακα 5.9 μπορούμε επίσης να εξάγουμε κάποια συμπεράσματα για τον λογάριθμο της σχετικής πιθανότητας. Με δεδομένο πως το επίπεδο αναφοράς είναι η χρήση των αγροτικών οχημάτων ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας επιτυχίας (δηλαδή να έχουμε σωματική βλάβη) στα επιβατικά οχήματα (EIX) είναι κατά 0,23 μονάδες μεγαλύτερος σε σχέση με τον λογάριθμο της αντίστοιχης σχετικής πιθανότητας επιτυχίας για την χρήση αγροτικού οχήματος, Αντίστοιχα για την χρήση των ενοικιαζόμενων ο λογάριθμος είναι κατά 0,7961 μονάδες μεγαλύτερος από τον λογάριθμο την αγροτικών οχημάτων. Επίσης για τα φορτηγά δημοσίας χρήσεως είναι 0,11401 μονάδες μεγαλύτερος, για τα λοιπά οχήματα 0,0584 μεγαλύτερος και για τις μοτοσυκλέτες 0,5655 μεγαλύτερος. Αντιθέτως αξίζει να αναφέρουμε πως για την χρήση των φορτηγών ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας επιτυχίας είναι κατά 0,33839 μονάδες μικρότερος σε σχέση με τον λογάριθμο της αντίστοιχης σχετικής πιθανότητας επιτυχίας για τα αγροτικά οχήματα. Για την τιμή -0,039795 (intercept) πρέπει να αναφέρουμε πως είναι ο λογάριθμος της σχετικής πιθανότητας όταν οι υπόλοιπες μεταβλητές είναι 0.

Πίνακας ανάλυσης απόκλισης για την ένδειξη Θανατηφόρου ατυχήματος

Μοντέλο	Απόκλιση	Βαθμοί Ελευθερίας	Μεταβολή στην απόκλιση	Μεταβολή στους β.ε.	Στατιστικά σημαντικό σε ε.σ. 5 %
1 (constant)	2347,4	62391			
ΧΡΗΣΗ	2332,7	62385	14,631	6	ΝΑΙ

Πίνακας 5.10

Πίνακας ανάλυσης συντελεστών για την ένδειξη Σωματικής Βλάβης

	Estimate	Std. Error	Z value	Pr (> z)
(Intercept)	-6,6022	0,3782	-17,456	<2e-16***
XRHSHEIX	0,6578	0,3916	1,680	0,09302
XRHSHENOIK	1,1925	0,5046	2,363	0,01812*
XRHSHFDX	1,4484	0,4763	3,041	0,00236 **
XRHSHFIX	0,5127	0,4425	1,159	0,24661
XRHSHLOIPA	1,1946	0,4935	2,421	0,01548 *
XRHSHMOTO	0,7905	0,4536	1,743	0,08137

Πίνακας 5.11

$$\log \left[\frac{p_i}{1-p_i} \right] = -6,6022 + 0,6578EIX + 1,1925ENOIK. + 1,4484FDX + 0,5046FIX + 1,1946LOIPA + 0,7905MOTO$$

Από τους παραπάνω πίνακες (5.10 , 5.11) για την μεταβλητή χρήση έχοντας πάλι ως επίπεδο αναφοράς τα αγροτικά οχήματα παρατηρούμε πως όλες οι άλλες κατηγορίες έχουν περισσότερα θανατηφόρα ατυχήματα. Στατιστικά σημαντικές στο τελικό μοντέλο είναι οι χρήσεις ENOIK., ΦΔΧ., ΛΟΙΠΑ.

Από το πίνακα 5.11 μπορούμε να σχολιάσουμε για τον λογάριθμο της σχετικής πιθανότητας. Συγκεκριμένα παρατηρούμε πως ο λογάριθμος της εκτιμώμενης σχετικής πιθανότητας επιτυχίας είναι 0,657 μονάδες για τα επιβατικά, 1,1925 μονάδες για τα ενοικιαζόμενα, 1,4484 για τα φορτηγά δημοσίας χρήσεως, 0,5046 για τα φορτηγά, 1,1946 για τις λοιπές χρήσεις και 0,7905 για τις μοτοσυκλέτες, μεγαλύτερος σε σχέση με τον λογάριθμο της αντίστοιχης σχετικής πιθανότητας επιτυχίας για το επίπεδο αναφοράς τα αγροτικά οχήματα . Για την τιμή -6,6022 (intercept) πρέπει να αναφέρουμε και εδώ πως είναι ο λογάριθμος της σχετικής πιθανότητας όταν οι υπόλοιπες μεταβλητές είναι 0.

Εν κατακλείδι λαμβάνοντας υπόψη όλα τα παραπάνω στοιχεία που προέκυψαν από την εφαρμογή των γενικευμένων γραμμικών μοντέλων στα δεδομένα μας η ασφαλιστική εταιρεία θα εξετάσει ποιες είναι οι μεταβλητές που επιδρούν περισσότερο και πως θα πρέπει να διαμορφωθεί το τεχνικό ασφάλιστρο σύμφωνα με αυτές.

ΠΑΡΑΡΤΗΜΑ

Παράρτημα 1

NOMOS	MEAN_of_CLAIM_COST
ΑΙΤΩΛ/ΝΙΑ	3.416,54
ΑΡΓΟΛΙΔΑ	1.709,16
ΑΡΚΑΔΙΑ	3.321,42
ΑΡΤΑ	2.152,26
ΑΤΤΙΚΗ	2.059,34
ΑΧΑΙΑ	1.938,43
ΒΟΙΩΤΙΑ	1.365,96
ΓΡΕΒΕΝΑ	1.282,79
ΔΡΑΜΑ	1.704,72
ΔΩΔΕΚΑΝΗΣΑ	2.839,01
ΕΒΡΟΣ	1.671,14
ΕΥΒΟΙΑ	2.404,30
ΕΥΡΥΤΑΝΙΑ	1.389,28
ΖΑΚΥΝΘΟΣ	2.009,16
ΗΛΕΙΑ	2.389,02
ΗΜΑΘΙΑ	2.272,25
ΗΡΑΚΛΕΙΟ	2.333,70
ΘΕΣ/ΝΙΚΗ	1.620,43
ΘΕΣΠΡΩΤΙΑ	1.648,72
ΙΩΑΝΝΙΝΑ	1.690,40
ΚΑΒΑΛΑ	1.965,96
ΚΑΡΔΙΤΣΑ	1.052,62
ΚΑΣΤΟΡΙΑ	2.475,94
ΚΕΡΚΥΡΑ	2.330,71
ΚΕΦΑΛΛΟΝΙΑ	1.273,68
ΚΙΛΚΙΣ	1.959,97
ΚΟΖΑΝΗ	1.327,23
ΚΟΡΙΝΘΟΣ	1.469,88
ΚΥΚΛΑΔΕΣ	1.926,89
ΛΑΚΩΝΙΑ	1.332,56
ΛΑΡΙΣΑ	1.476,32
ΛΑΣΙΘΙ	3.790,95
ΛΕΣΒΟΣ	2.622,56
ΛΕΥΚΑΔΑ	1.277,27
ΜΑΓΝΗΣΙΑ	2.497,75
ΜΕΣΣΗΝΙΑ	2.062,47
ΞΑΝΘΗ	3.074,76

ΠΕΛΛΑ	2.014,56
ΠΙΕΡΙΑ	2.025,62
ΠΡΕΒΕΖΑ	1.785,85
ΡΕΘΥΜΝΟ	1.623,01
ΡΟΔΟΠΗ	2.005,14
ΣΑΜΟΣ	2.278,92
ΣΕΡΡΕΣ	1.414,93
ΤΡΙΚΑΛΑ	3.658,58
ΦΘΙΩΤΙΔΑ	1.682,10
ΦΛΩΡΙΝΑ	1.286,92
ΦΩΚΙΔΑ	1.667,03
ΧΑΛΚΙΔΙΚΗ	1.131,60
ΧΑΝΙΑ	1.612,05
ΧΙΟΣ	2.199,96

Παράρτημα 2

Easy Fit

Πρόκειται για ένα πρόγραμμα ανάλυσης και προσομοίωσης δεδομένων που επιτρέπει την προσαρμογή κατανομών πιθανότητας σε ένα δείγμα δεδομένων. Γίνεται εφαρμογή γνωστών διακριτών και συνεχών μεταβλητών. Εκτιμά τις παραμέτρους των κατανομών. Το easyFit πραγματοποιεί ταυτόχρονα τρία τεστ καλής προσαρμογής για τα δεδομένα

- Kolmogorov Smirnov Goodness-of-Fit -Test
- Anderson -Darling Goodness-of-Fit -Test
- Chi-Squared Goodness-of-Fit -Test

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Beta	0,66579	51	41193,0	50	N/A	
2	Burr	0,11772	4	1066,9	6	50737,0	12
3	Burr (4P)	0,12458	9	1034,3	3	44633,0	5
4	Cauchy	0,15045	17	1869,3	16	61215,0	24
5	Chi-Squared	0,88097	56	1,3010E+7	59	7,3216E+5	42
6	Chi-Squared (2P)	0,87062	55	1,0071E+7	58	7,1743E+5	41
7	Dagum	0,1145	2	1019,3	1	50418,0	11
8	Dagum (4P)	0,12444	8	1040,1	4	46252,0	7
9	Error	0,41448	38	18825,0	38	2,0739E+5	39
10	Error Function	0,50024	47	20633,0	45	33435,0	1
11	Exponential	0,38212	33	11237,0	31	85675,0	31
12	Exponential (2P)	0,38319	34	11398,0	32	85452,0	30
13	Fatigue Life	0,2427	29	4501,0	25	78994,0	29
14	Fatigue Life (3P)	0,28248	32	5273,9	27	61684,0	25
15	Frechet	0,17843	22	2991,3	21	47638,0	8
16	Frechet (3P)	0,12087	7	1044,7	5	49704,0	10
17	Gamma	0,86314	54	94990,0	54	1,3167E+6	45
18	Gamma (3P)	0,39966	35	17196,0	37	N/A	
19	Gen. Extreme Value	0,15721	19	3351,2	22	40095,0	2
20	Gen. Gamma	0,41586	39	16871,0	35	2,0081E+5	36
21	Gen. Gamma (4P)	0,18285	23	2092,5	17	61031,0	23
22	Gen. Logistic	0,15633	18	17112,0	36	N/A	

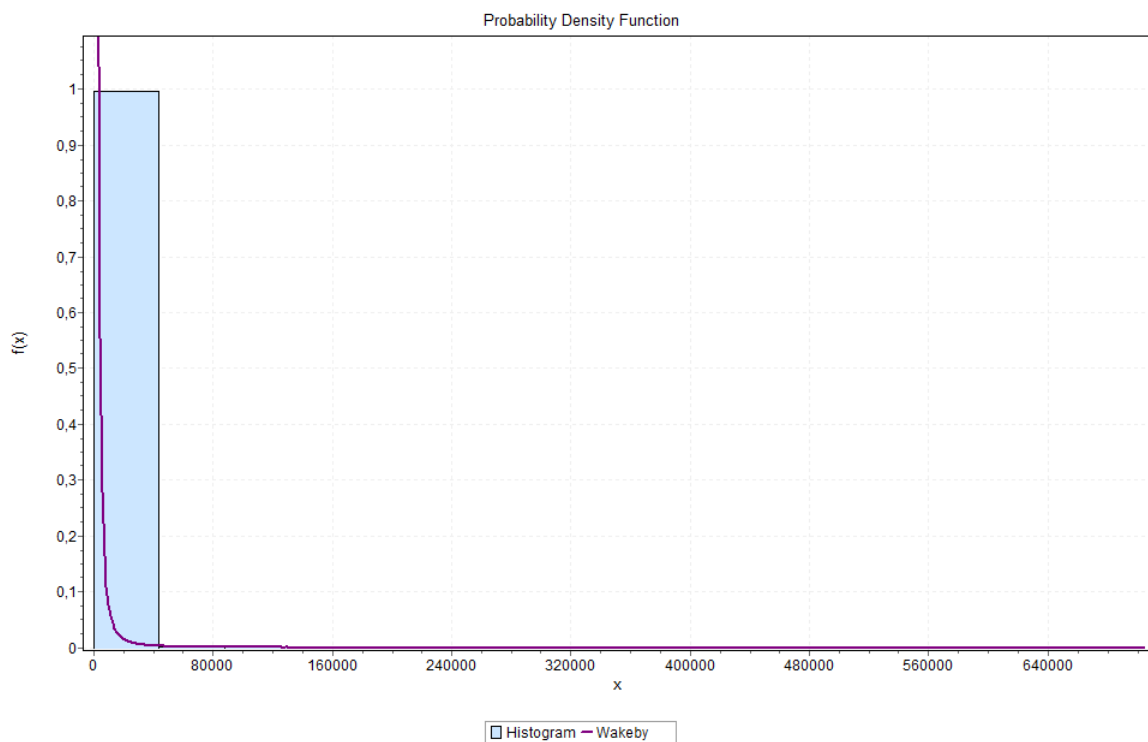
23	Gen. Pareto	0,1654	20	20140,0	42	N/A	
24	Gumbel Max	0,51398	48	20539,0	44	57526,0	18
25	Gumbel Min	0,46454	45	22637,0	48	N/A	
26	Hypersecant	0,43416	41	19558,0	40	54514,0	14
27	Inv. Gaussian	0,6116	50	39640,0	49	2,0130E+5	37
28	Inv. Gaussian (3P)	0,20654	26	2721,3	19	65450,0	27
29	Kumaraswamy	0,24866	30	7183,9	30	91037,0	32
30	Laplace	0,41448	37	18825,0	39	2,0739E+5	38
31	Levy	0,21704	28	4897,8	26	1,1073E+5	34
32	Levy (2P)	0,27254	31	5793,6	29	1,0932E+5	33
33	Log-Gamma	0,1355	12	1741,9	15	60559,0	21
34	Log-Logistic	0,12491	10	1209,8	9	58787,0	20
35	Log-Logistic (3P)	0,11754	3	1032,7	2	48393,0	9
36	Log-Pearson 3	0,13777	13	1527,4	13	60674,0	22
37	Logistic	0,44014	42	19853,0	41	41097,0	4
38	Lognormal	0,14116	14	1500,2	12	57915,0	19
39	Lognormal (3P)	0,14373	15	1318,5	11	45583,0	6
40	Normal	0,44724	43	20453,0	43	41066,0	3
41	Pareto	0,43363	40	16467,0	34	2,2031E+5	40
42	Pareto 2	0,14636	16	2147,9	18	53210,0	13
43	Pearson 5	0,17745	21	4122,7	24	65339,0	26
44	Pearson 5 (3P)	0,21069	27	5357,8	28	72983,0	28
45	Pearson 6	0,12848	11	1198,5	8	55692,0	17
46	Pearson 6 (4P)	0,12049	6	1071,5	7	54795,0	15

47	Pert	0,92862	57	2,3500E+5	55	8,9576E+5	43
48	Phased Bi-Exponential	0,11934	5	1659,1	14	55234,0	16
49	Power Function	0,47723	46	21058,0	46	N/A	
50	Rayleigh	0,59008	49	51604,0	51	N/A	
51	Rayleigh (2P)	0,68562	53	77897,0	53	N/A	
52	Reciprocal	0,41146	36	15929,0	33	1,8658E+5	35
53	Rice	0,68467	52	77129,0	52	N/A	
54	Student's t	0,99674	59	7,1778E+5	57	1,5705E+8	46
55	Triangular	0,95471	58	2,9159E+5	56	9,2062E+5	44
56	Uniform	0,46171	44	22253,0	47	N/A	
57	Wakeby	0,10721	1	1227,1	10	N/A	
58	Weibull	0,19216	24	2754,6	20	N/A	
59	Weibull (3P)	0,19721	25	4028,2	23	N/A	
60	Erlang	No fit					
61	Erlang (3P)	No fit					
62	Johnson SB	No fit					
63	Johnson SU	No fit					
64	Nakagami	No fit					
65	Phased Bi-Weibull	No fit					

#	Distribution	Parameters
1	Beta	$\alpha_1=0,04048$ $\alpha_2=1,6961$ $a=9,0$ $b=7,0478E+5$
2	Burr	$k=1,0607$ $\alpha=1,6333$ $\beta=722,94$
3	Burr (4P)	$k=0,64512$ $\alpha=2,4251$ $\beta=601,51$ $\gamma=-101,89$
4	Cauchy	$\sigma=383,12$ $\mu=758,28$
5	Chi-Squared	$v=2002$
6	Chi-Squared (2P)	$v=1800$ $\gamma=9,0$
7	Dagum	$k=0,74164$ $\alpha=1,8824$ $\beta=878,78$
8	Dagum (4P)	$k=5,8419$ $\alpha=1,7581$ $\beta=282,25$ $\gamma=-238,03$
9	Error	$k=1,0$ $\sigma=15032,0$ $\mu=2002,8$
10	Error Function	$h=4,7040E-5$
11	Exponential	$\lambda=4,9929E-4$
12	Exponential (2P)	$\lambda=5,0155E-4$ $\gamma=9,0$
13	Fatigue Life	$\alpha=1,7763$ $\beta=814,61$
14	Fatigue Life (3P)	$\alpha=1,5101$ $\beta=1033,2$ $\gamma=-39,269$
15	Frechet	$\alpha=1,0399$ $\beta=385,57$
16	Frechet (3P)	$\alpha=1,8084$ $\beta=832,79$ $\gamma=-336,57$
17	Gamma	$\alpha=0,01775$ $\beta=1,1282E+5$
18	Gamma (3P)	$\alpha=0,1431$ $\beta=4,1636E+5$ $\gamma=9,0$
19	Gen. Extreme Value	$k=0,78429$ $\sigma=372,52$ $\mu=464,91$

20	Gen. Gamma	$k=0,87893$ $\alpha=0,23487$ $\beta=1,1282E+5$
21	Gen. Gamma (4P)	$k=0,34665$ $\alpha=6,6862$ $\beta=3,9706$ $\gamma=-63,982$
22	Gen. Logistic	$k=0,78547$ $\sigma=367,69$ $\mu=620,07$
23	Gen. Pareto	$k=0,75969$ $\sigma=433,32$ $\mu=199,66$
24	Gumbel Max	$\sigma=11720,0$ $\mu=-4762,4$
25	Gumbel Min	$\sigma=11720,0$ $\mu=8768,0$
26	Hypersecant	$\sigma=15032,0$ $\mu=2002,8$
27	Inv. Gaussian	$\lambda=35,555$ $\mu=2002,8$
28	Inv. Gaussian (3P)	$\lambda=635,42$ $\mu=2056,9$ $\gamma=-54,091$
29	Kumaraswamy	$\alpha_1=0,46427$ $\alpha_2=16,612$ $a=9,0$ $b=7,1070E+5$
30	Laplace	$\lambda=9,4080E-5$ $\mu=2002,8$
31	Levy	$\sigma=302,09$
32	Levy (2P)	$\sigma=419,47$ $\gamma=-28,996$
33	Log-Gamma	$\alpha=31,282$ $\beta=0,2081$
34	Log-Logistic	$\alpha=1,5232$ $\beta=671,62$
35	Log-Logistic (3P)	$\alpha=1,7425$ $\beta=707,78$ $\gamma=-16,613$
36	Log-Pearson 3	$\alpha=748,34$ $\beta=0,04255$ $\gamma=-25,33$
37	Logistic	$\sigma=8287,6$ $\mu=2002,8$
38	Lognormal	$\sigma=1,1639$ $\mu=6,5098$
39	Lognormal (3P)	$\sigma=1,0697$ $\mu=6,5848$ $\gamma=-27,127$
40	Normal	$\sigma=15032,0$ $\mu=2002,8$

40	Normal	$\sigma=15032,0$ $\mu=2002,8$
41	Pareto	$\alpha=0,23188$ $\beta=9$
42	Pareto 2	$\alpha=2,7004$ $\beta=2243,9$
43	Pearson 5	$\alpha=0,74927$ $\beta=226,35$
44	Pearson 5 (3P)	$\alpha=0,64225$ $\beta=162,08$ $\gamma=8,7729$
45	Pearson 6	$\alpha_1=1,9049$ $\alpha_2=2,1817$ $\beta=799,51$
46	Pearson 6 (4P)	$\alpha_1=7,4907$ $\alpha_2=1,9996$ $\beta=187,68$ $\gamma=-115,94$
47	Pert	$m=9,0$ $a=9,0$ $b=7,0474E+5$
48	Phased Bi-Exponential	$\lambda_1=0,00105$ $\gamma_1=9$ $\lambda_2=3,2073E-5$ $\gamma_2=3253$
49	Power Function	$\alpha=0,10343$ $a=9,0$ $b=5,3719E+6$
50	Rayleigh	$\sigma=1598,0$
51	Rayleigh (2P)	$\sigma=2129,2$ $\gamma=7,7854$
52	Reciprocal	$a=9,0$ $b=7,0474E+5$
53	Rice	$v=263,95$ $\sigma=2127,5$
54	Student's t	$v=2$
55	Triangular	$m=9,0$ $a=9,0$ $b=7,0474E+5$
56	Uniform	$a=-24033,0$ $b=28039,0$
57	Wakeby	$\alpha=1389,6$ $\beta=1,9317$ $\gamma=204,21$ $\delta=0,86504$ $\xi=15,663$
58	Weibull	$\alpha=1,0433$ $\beta=1167,8$
59	Weibull (3P)	$\alpha=0,68295$ $\beta=1185,9$ $\gamma=9,0$



Παράρτημα 3

Η κρίσιμη τιμή της κατανομής χ^2 με τον αντίστοιχο βαθμό ελευθερίας κάθε φορά δίνεται στο ακόλουθο πίνακα

ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ P-VALUE

(167.125, 6) \rightarrow 0	<0,05
(1.926,1) \rightarrow 0,16	>0,05
(13.374,5) \rightarrow 0,02	<0,05
(16.8368,6) \rightarrow 0,009	<0,05
(0.6852,1) \rightarrow 0,40	>0,05
(5.4262,5) \rightarrow 0,36	>0,05
(6.7861,6) \rightarrow 0,34	>0,05
(1.4691,1) \rightarrow 0,22	>0,05
(4.2748,5) \rightarrow 0,51	>0,05
(6.719,2) \rightarrow 0,03	<0,05
(3.7719,2) \rightarrow 0,15	>0,05
(214.033,6) \rightarrow 0	<0,05
(31.304,2) \rightarrow 1.59375e-0.7	>0,05

(14.631,6)→0,02

<0,05

Πίνακας ανάλυσης συντελεστών μοντέλου Σωματικής βλάβης

	Estimate	Std.Error	Z value	Pr(> z)
(Intercept)	-3,219767	0,118455	-27,181	<2e-16***
XRHSHEIX	0,290891	0,083445	3,486	0,00049***
XRHSHENOIK	0,831447	0,124169	6,696	2,14e-11***
XRHSHFDX	0,260720	0,135915	1,918	0,05508
XRHSHFIX	-0,344950	0,107509	-3,209	0,00133**
XRHSHLOIPA	0,149949	0,138649	1,082	0,27947
XRHSHMOTO	0,617480	0,101349	6,093	1,11e-09***
AGE	0,002083	0,001414	1,473	0,14065
PERIFEREIA ATTIKH	0,040350	0,071873	0,561	0,57453
PERIFEREIACENTRAL_GR	0,027811	0,067548	-0,412	0,68055
PERIFEREIAIONIAN	0,056320	0,124231	-0,453	0,65030
PERIFEREIA NORTH_GR	0,187373	0,074407	-2,518	0,01180*
PERIFEREIA SOUTH_GR	0,007713	0,073920	0,104	0,91690

Πίνακας ανάλυσης συντελεστών μοντέλου θανατηφόρων οχημάτων

	Estimate	Std.Error	Z value	Pr(> z)
(Intercept)	-6,304689	0,577749	-10,912	<2e-16***
XRHSHEIX	0,830929	0,466169	1,782	0,07467
XRHSHENOIK	0,923631	0,650323	1,420	0,15553
XRHSHFDX	1,705962	0,542959	3,142	0,00168**
XRHSHFIX	0,624067	0,520019	1,200	0,23011
XRHSHLOIPA	1,559690	0,550024	2,836	0,00457**
XRHSHMOTO	1,001918	0,533633	1,878	0,06044
AGE	-0,005032	0,006144	-0,819	0,41278
PERIFEREIA ATTIKH	-0,590318	0,321520	-1,836	0,06635
PERIFEREIACENTRAL_GR	0,163282	0,264870	-0,616	0,53759
PERIFEREIAIONIAN	0,311753	0,539492	-0,578	0,56336
PERIFEREIA NORTH_GR	0,006929	0,274769	-0,025	0,97988
PERIFEREIA SOUTH_GR	-0,325011	0,310363	-1,047	0,29501

Πίνακας ανάλυσης συντελεστών μοντέλου Σωματικής βλάβης (3περιφέρειες)

	Estimate	Std.Error	Z value	Pr(> z)
(Intercept)	-3,164791	0,118169	-26,782	<2e-16***
XRHSHEIX	0,277234	0,083155	3,334	0,0008568***
XRHSHENOIK	0,814633	0,121640	6,697	2,13e-11***
XRHSHFDX	0,245667	0,135732	1,810	0,070304
XRHSHFIX	-0,354131	0,107354	-3,299	0,000971***
XRHSHLOIPA	0,142552	0,138516	1,029	0,303417
XRHSHMOTO	0,611261	0,100706	6,070	1,28e-09***

AGE	0,002018	0,001414	1,428	0,153351
PERIFEREIA NORTH_GR	-0,132599	0,056541	-2,345	0,019018*
PERIFEREIA SOUTH_GR	-0,042660	0,059536	-0,717	0,473656

Πίνακας ανάλυσης συντελεστών μοντέλου θανατηφόρων οχημάτων(3 περιφέρειες)

	Estimate	Std.Error	Z value	Pr(> z)
(Intercept)	-6,925516	0,600608	-11,531	<2e-16***
XRHSHEIX	0,861858	0,465257	1,852	0,06396
XRHSHENOIK	1,027947	0,642596	1,600	0,10967
XRHSHFDX	1,736995	0,542161	3,204	0,00136**
XRHSHFIX	0,648760	0,519466	1,249	0,21170
XRHSHLOIPA	1,580591	0,549520	2,876	0,00402
XRHSHMOTO	1,050055	0,531322	1,976	0,04812*
AGE	-0,005086	0,006138	-0,829	0,40735
PERIFEREIA NORTH_GR	0,498907	0,270894	1,842	0,06552
PERIFEREIA SOUTH_GR	0,418320	0,286440	1,460	0,14418

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική Βιβλιογραφία

1. Νεκτάριος Μ.(2014), *Ιδιωτική ασφάλιση και Διαχείριση Κινδύνων*, Εκδόσεις Παπαζήση
2. Νεκτάριος Μ.(2010), *Διοικητική Κινδύνων και Ασφαλίσεις Επιχειρήσεων*, Εκδόσεις Σταμούλη
3. Δαμιανός Χ. & Κούτρας Μάρκος(2003), *Εισαγωγή στη Στατιστική* , Εκδόσεις Συμμετρία
4. Αντζουλάκος Δ.(2009), *Ανάλυση Δεδομένων με τη Χρήση Στατιστικών Πακέτων Εισαγωγή στο R*, Σημειώσεις παραδόσεων. Πανεπιστήμιο Πειραιώς ΠΜΣ Αναλογιστή Επιστήμη και Διοικητική Κινδύνου
5. Φωκιανός Κ. & Χαραλάμπους Χ. (2010) , *Εισαγωγή στην R Πρόχειρες Σημειώσεις* , Τμ.. Μαθηματικών & Στατιστικής, Πανεπιστήμιο Κύπρου
6. Μπούτσικας Μ.(2008), *Σημειώσεις Διαλέξεων στα Στατιστικά Προγράμματα*, Τμ. Στατιστικής και Ασφαλιστικής επιστήμης Πανεπιστήμιο Πειραιά
7. Κ.Πολίτης(2015), *Σημειώσεις Διαλέξεων στα Γενικευμένα Γραμμικά Μοντέλα*, Πανεπιστήμιο Πειραιώς ΠΜΣ Εφαρμοσμένη Στατιστική

Ξενόγλωσση Βιβλιογραφία

1. Klugman S., Panjer H.Wilmot (1998), *Loss Models : From Data to DeCisions*.3rd edition, Willey
2. Athreya K.B (1987), *Bootstrap estimation of the mean in the infinite variance case*, Annals of Statistic
3. Chernick M. (2008), *Bootstrap methods, a guide for practitioners and researchers*, John Wiley and Sons Davison
4. Efron B.- Tibshirani R, (1993),*An Introduction to the bootstrap* , Chapman and Hall

5. Hall P. (1992a) , *Efficient bootstrap simulations, Exploring the limits of bootstrap*, LePage and L.Billard eds, Wiley
6. Shao J.- Tu D. (1995) ,*The jackknife and bootstrap*, Springer
7. de Jong, P. and Heller, G. (2008) *Generalized Linear Models for Insurance Data*. Cambridge University Press, Cambridge UK
8. Jeremy Orloff and Jonathan Bloom ,spring 2014,*Bootstrap confidence intervals*, , MIT Open Course Ware

Ηλεκτρονική Βιβλιογραφία

1. <http://www.eaee.gr/cms/>
2. <http://www.bankofgreece.gr/Pages/default.aspx>
3. <http://www.nextdeal.gr/>
4. <https://el.wikipedia.org/>