ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΠΜΣ ΑΝΑΛΟΓΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ & ΔΙΟΙΚΗΤΙΚΗΣ ΚΙΝΔΥΝΟΥ

ΤΙΜΟΛΟΓΗΣΗ ΑΣΦΑΛΙΣΤΗΡΙΩΝ ΣΥΜΒΟΛΑΙΩΝ ΑΣΤΙΚΗΣ ΕΥΘΥΝΗΣ ΕΝΑΝΤΙ ΤΡΙΤΩΝ ΣΤΟΝ ΚΛΑΔΟ ΑΥΤΟΚΙΝΗΤΟΥ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΟΜΑΤΟΣ ΑΘΑΝΑΣΙΟΣ
ΜΑΕ 08026

Τρίτη, 31 Ιανουαρίου 2012

ΠΕΡΙΛΗΨΗ

Στην ασφαλιστική αγορά η τιμολόγηση ασφαλιστηρίων συμβολαίων αστικής ευθύνης έναντι τρίτων όσον αφορά τον κλάδο αυτοκινήτου αποτελεί ζήτημα μείζονος σημασίας, διότι μόνο με την σωστή τιμολόγηση η ασφαλιστική εταιρία θα μπορέσει να ανταπεξέλθει τόσο στις υποχρεώσεις της προς τους ασφαλισμένους της και το κοινωνικό σύνολο, όσο και απέναντι στις συνθήκες ανταγωνισμού που επικρατούν στην ασφαλιστική αγορά. Αποτελεί επιπλέον, ζήτημα αξιοπιστίας και φερεγγυότητας αυτής.

Αρχικά, παρουσιάζονται στην εισαγωγή της εργασίας ορισμένα συστήματα τιμολόγησης ασφαλίστρων που επικρατούν στην διεθνή ασφαλιστική αγορά. Έπειτα, ορισμένα μεμειγμένα πρότυπα της κατανομής Poisson για τους αριθμούς των ατυχημάτων, μια εισαγωγή στα γενικευμένα γραμμικά μοντέλα, τις κυριότερες ζημιοκατανομές, την ταξινόμηση των κινδύνων με εκ των προτέρων κριτήρια και τέλος, παρουσιάζονται τα αποτελέσματα και τα συμπεράσματα των εμπειρικών δεδομένων που χρησιμοποιήσαμε για την εργασία μας.

Θα ήταν παράλειψή μου να μην ευχαριστήσω τον επιβλέποντα Λέκτορα του Πανεπιστημίου Πειραιώς κ. Σπυρίδων Βρόντο για την καθοδήγηση, το χρόνο του αλλά και την πολύτιμη βοήθειά του στη συγγραφή αυτής της εργασίας. Θα ήθελα να ευχαριστήσω επίσης την οικογένειά μου για την αμέριστη συμπαράσταση που έδειξαν στο πρόσωπό μου καθ’ όλη τη διάρκεια των σπουδών μου.

Τρίτη, 31 Ιανουαρίου 2012

Γομάτος Αθανάσιος

ABSTRACT

In the actuarial market the pricing of insurance contracts against third party liability with regard to motor insurance, constitutes a question of major importance, because only with the correct pricing the actuarial company might be able to cope with her obligations towards to her insurers and to the social total, but also towards to the conditions of competition that prevail in the actuarial market. It constitutes moreover, a question of the company's solvency and economic strength.

Initially, in the introduction of our thesis we presented certain systems of premium pricing that prevail in the international actuarial market, then some certain mixed models of Poisson distribution for the numbers of accidents, an introduction in the generalized linear models, the main loss distributions, the classification of risks with a priori criteria, and finally, the results and the conclusions of empirical data that we used for our thesis.

It would be a great omission if i didn't thank Dr. S. Vrontos for his guidance, time and valuable help while writing this thesis. I would also like to express my thanks to my family who has been very supportive throughout my studies.

Τρίτη, Ιανουάριος 31, 2012

Gomatos Athanassios

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 1
ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή

Τον τελευταίο καιρό είναι πάρα πολύ δύσκολο για τις ασφαλιστικές εταιρίες να μελετήσουν και να μοντελοποιήσουν τις διαφορετικές κατηγορίες κινδύνων μέσα σε μια ανταγωνιστική αγορά. Αν, για παράδειγμα, αποδειχτεί ότι οι γυναίκες οδηγοί προκαλούν σημαντικά λιγότερα ατυχήματα από τους άντρες και αν η ασφαλιστική εταιρία αγνοήσει αυτή την περίπτωση και χρεώσει ένα μέσο ασφάλιστρο για όλα τα ασφαλιστήρια συμβόλαια ανεξαρτήτως φύλου, τότε οι περισσότερες γυναίκες οδηγοί θα τείνουν να απευθυνθούν σε άλλη ασφαλιστική εταιρία που θα τους προσφέρει καλύτερο ασφάλιστρο. Σε αυτήν την περίπτωση, η πρώτη ασφαλιστική εταιρία θα μείνει με ένα δυσανάλογο αριθμό ασφαλιστηρίων συμβολαίων με άντρες οδηγούς και με ένα ανεπαρκές ασφάλιστρο για ανταποκριθεί στις ενδεχόμενες απαιτήσεις που θα προκύψουν.

Για να αποφευχθούν τέτοιου είδους καταστάσεις σε ένα ανταγωνιστικό περιβάλλον όπως είναι αυτό της ασφάλισης, οι αναλογιστές θα πρέπει να τιμολογήσουν με τέτοιο τρόπο το ασφάλιστρο ώστε να κατανέμεται ομαλά και να επιμερίζεται το ύψος των απαιτήσεων μεταξύ του χαρτοφυλακίου των ασφαλιστήριων συμβολαίων της εταιρίας. Έτσι, οδηγούνται στην κατηγοριοποίηση των κινδύνων σε διάφορα επίπεδα με τους ασφαλισμένους να πληρώνουν ασφάλιστρο ανάλογα με το επίπεδο που ανήκουν. Κάθε φορά βέβαια που μια ανταγωνιστική εταιρία αλλάξει την τιμολογιακή της πολιτική, ο αναλογιστής θα πρέπει να προσαρμόζει το ασφάλιστρο στα νέα δεδομένα της αγοράς για να μην κινδυνέψει να «χάσει» ασφαλιστήρια συμβόλαια.

Σε ένα περιβάλλον ελεύθερης αγοράς, οι ασφαλιστικές εταιρίες πρέπει να χρησιμοποιούν μια τιμολογιακή πολιτική η οποία να προσεγγίζει ικανοποιητικά τους ενδεχόμενους κινδύνους στους οποίους απευθύνονται όσο το δυνατόν καλύτερα γίνεται, ενώ να είναι ταυτόχρονα και πιο κοντά στις τιμολογιακές πολιτικές των ανταγωνιστών τους. Αυτό σημαίνει ότι θα πρέπει να χρησιμοποιήσουν όσο το δυνατόν περισσότερες μεταβλητές που συσχετίζονται με τους κινδύνους που αναλαμβάνουν (και την κατηγοριοποίηση αυτών), έτσι ώστε να μπορούν να έχουν το δικαίωμα επιλογής ασφαλισμένων και να μην καταλήγουν σε ένα χαρτοφυλάκιο το οποίο θα αποφέρει ζημίες. Συνεπώς, ο ανταγωνισμός μεταξύ των ασφαλιστικών εταιριών και η κατηγοριοποίηση των κινδύνων μέσω της αναλογιστικής επιστήμης καθορίζει την ποιότητα του χαρτοφυλακίου τους. Η πολιτική αυτή της κατηγοριοποίησης κινδύνων μπορεί να οδηγήσει στην περίπτωση: οι οδηγοί που έχουν προκαλέσει αρκετά ατυχήματα μέσα σε μια περίοδο να μη βρίσκουν ασφαλιστική κάλυψη για το αυτοκίνητό τους σε μια λογική τιμή – στα πλαίσια της πολιτικής underwriting των εταιριών - και προτιμούν να οδηγούν χωρίς ασφάλιση. Το ζήτημα αυτό είναι πολύ σημαντικό και αποτελεί θέμα προς συζήτηση.

Οι ασφαλιστές αστικής ευθύνης αυτοκινήτου και πυρός χρησιμοποιούν ένα σχέδιο κατηγοριοποίησης για να δημιουργήσουν κλάσεις κινδύνων. Οι μεταβλητές που χρησιμοποιούν για να καταλήξουν στις επιθυμητές κλάσεις ονομάζονται a priori μεταβλητές (και αυτό γιατί οι τιμές τους μπορούν να καθοριστούν πριν ξεκινήσει να οδηγεί ο κάτοχος του ασφαλιστηρίου συμβολαίου). Τα ασφάλιστρα για την αστική ευθύνη αυτοκινήτου διαφέρουν ανάλογα με την περιοχή στην οποία κινείται το αυτοκίνητο, την χρήση του (επαγγελματική ή ιδιωτική χρήση) και τα ατομικά χαρακτηριστικά του οδηγού (όπως είναι η ηλικία, το φύλο, το επάγγελμα, η οικογενειακή κατάσταση κ.α.). Αν οι ασφαλισμένοι δεν είναι ειλικρινείς στη δήλωση των στοιχείων τους στο ασφαλιστήριο συμβόλαιο, τότε εκείνο μπορεί να θεωρηθεί άκυρο και να μην είναι δυνατή η καταβολή οποιασδήποτε αποζημίωσης. Γι’ αυτό τον λόγο είναι πολύ σημαντικό να δίνονται τα αληθινά στοιχεία στη δήλωση για οποιαδήποτε ασφαλιστική κάλυψη.

Στην εργασία αυτή θα χρησιμοποιήσουμε τα γενικευμένα γραμμικά μοντέλα όπως αυτά εισήχθησαν από τους Nelder and Wedderburn (1972). Οι συγγραφείς αυτοί ανακάλυψαν ότι ορισμένα μοντέλα παλινδρόμησης που η μεταβλητή απόκρισής τους ακολουθεί στην εκθετική οικογένεια κατανομών έχουν τα ίδια χαρακτηριστικά. Μέλη αυτής της οικογένειας κατανομών είναι η Κανονική, η Διωνυμική, η Poisson, η Γάμμα και η Inverse Gaussian. Πιο συγκεκριμένα: η κατανομή Poisson και η Αρνητική διωνυμική χρησιμοποιούνται για να μοντελοποιήσουν τον αριθμό των ατυχημάτων και οι κατανομές Gamma, Lognormal, Pareto, Weibull χρησιμοποιούνται για να μοντελοποιήσουν το ύψος ζημιάς, που θα απαιτηθούν σε μια περίοδο. Χρησιμοποιώντας τις παραπάνω κατανομές οδηγούμαστε σε μοντέλα τα οποία καταλήγουν σε χρήσιμα αποτελέσματα για τις ασφαλιστικές εταιρίες (M. Denuit et al (2007)).

1.2 Σύστημα Pay-As-You-Drive

Όσα περισσότερα χιλιόμετρα διανύει το αυτοκίνητο, τόσος περισσότερος κίνδυνος μεταφέρεται από τον ασφαλισμένο στον ασφαλιστή. Συνεπώς, το συνολικό κόστος της κάλυψης αυξάνει από χιλιόμετρο σε χιλιόμετρο. Για τον λόγο αυτό, αρκετοί ερευνητές πρότειναν την χρέωση ενός λεπτού ανά χιλιόμετρο ως βασικό ασφάλιστρο. Ορισμένες ασφαλιστικές εταιρίες αστικής ευθύνης αυτοκινήτων τα τελευταία χρόνια υιοθετούν το νέο αυτό σύστημα κάλυψης το οποίο καλείται Pay-As-You-Drive (συντομογραφία PAYD). Κατ' αυτό το σύστημα, ο οδηγός πληρώνει ασφάλιστρο ανάλογα με τα χιλιόμετρα που πραγματοποιεί. Επιπλέον, θα πληρώσει μεγαλύτερο ασφάλιστρο αν το όχημα του το χρησιμοποιεί σε πολυσύχναστους δρόμους (σε δρόμους δηλαδή με κίνηση πολλών οχημάτων) και χαμηλότερο σε δρόμους αραιής κυκλοφορίας.

Αρκετές εταιρίες τον τελευταίο καιρό προσφέρουν ασφαλιστική κάλυψη βασισμένη στο σύστημα Pay-As-You-Drive (όπως είναι η εταιρία Norwich Union) έπειτα από ένα πετυχημένο πιλοτικό πρόγραμμα που συμπεριέλαβε χιλιάδες οχήματα. Με τα συστήματα Pay-As-You-Drive, οι οδηγοί προμηθεύονται από τις ασφαλιστικές τους εταιρίες συστήματα GPS (Global Positioning System) μαζί με κατάλληλους οδικούς χάρτες, έτσι ώστε να μπορεί η ασφαλιστική εταιρία να υπολογίσει το ασφάλιστρο για κάθε ταξίδι που θα πραγματοποιηθεί από τον ασφαλισμένο (συνυπολογίζεται βέβαια και αν το ταξίδι πραγματοποιήθηκε μέρα ή νύχτα καθώς επίσης και ο τύπος του οδοστρώματος). Μια συγκεκριμένη συσκευή τοποθετείται στο όχημα του ασφαλισμένου το οποίο λαμβάνει σήματα από το GPS έτσι ώστε οποιαδήποτε στιγμή να είναι γνωστή η τοποθεσία του οχήματος, η ταχύτητα του, η ώρα και την απόσταση που έχει διανύσει. Η συσκευή αυτή λειτουργεί ως ένα ασύρματο modem το οποίο εκπέμπει τις παραπάνω πληροφορίες μέσω δικτύων κινητής τηλεφωνίας και τις μεταφέρει στον ασφαλιστή. Έπειτα, ο ασφαλιστής στέλνει τον μηνιαίο λογαριασμό στον πελάτη του βασισμένο στην χρήση που έκανε, τα χιλιόμετρα που διένυσε και τη χρονική στιγμή της ημέρας που πραγματοποιήθηκε η χρήση. Με αυτό τον τρόπο λοιπόν, δημιουργούνται ιστορικά δεδομένα για το πώς, πότε και που τα αμάξια χρησιμοποιούνται πραγματικά, και το κατά πόσο τα ατυχήματα και οι απαιτήσεις που θα ζητηθούν μπορούν να εκτιμηθούν με τη χρήση συγκεκριμένων παραγόντων. Επίσης, η συσκευή αυτή εντοπίζει τις διακυμάνσεις στην ταχύτητα του οχήματος και γενικά, το κατά πόσο ήρεμος ή νευρικός είναι ο οδηγός (δηλαδή τη συμπεριφορά πίσω από το τιμόνι). Υψηλότερα ασφάλιστρα χρεώνονται σε εκείνους που οδηγούν ατίθασα και νευρικά έτσι ώστε να συνετιστούν και να οδηγούν ασφαλέστερα. Για να επιτευχθεί λοιπόν η καλύτερη εκτίμηση του κινδύνου από τον αναλογιστή, πρέπει να ληφθεί υπόψη η ηλικία του οδηγού, η ταχύτητα με την οποία κινεί το όχημά του, η τοποθεσία και η ώρα της ημέρας.

Η γενίκευση του παραπάνω συστήματος αναμένεται να αλλάξει την συμπεριφορά των οδηγών επειδή εκείνοι με τη σειρά τους για να αποφύγουν τα υψηλά ασφάλιστρα θα σκεφτούν πάρα πολύ την οδηγική τους συμπεριφορά. Αρκετές έρευνες που πραγματοποιήθηκαν στην Βόρειο Αμερική έδειξαν ότι τα συστήματα Pay-As-You-Drive μπορούν να ελαττώσουν τη χρήση αυτοκινήτου περισσότερο από 10%. Επίσης, το σύστημα αυτό αναμένεται να μειώσει την κίνηση στους δρόμους και την περιβαλλοντική μόλυνση λόγω του καυσαερίου (για τον απλό λόγο ότι οι δρόμοι που έχουν την περισσότερη κίνηση και χρησιμοποιούνται από τους ασφαλισμένους, επιβαρύνονται με υψηλά ασφάλιστρα) (M. Denuit et al (2007)).

1.3 Σύστημα Τιμολόγησης Ασφαλίστρων βασισμένο στην Οδηγική Εμπειρία

Ορισμένες ασφαλιστικές εταιρίες αποκλείουν ορισμένους παράγοντες στην κατηγοριοποίηση των κινδύνων, όπως είναι η ηλικία και το φύλο. Τα αποτελέσματα αυτών των αποκλεισμών, ενός δηλαδή εκ των προτέρων (a priori) συστήματος τιμολόγησης, μπορεί να διορθωθεί χρησιμοποιώντας τον παρελθοντικό αριθμό των απαιτήσεων για να υπολογιστούν εκ νέου τα μελλοντικά ασφάλιστρα. Η αξιοπιστία λοιπόν ενός αναλογιστικού μοντέλου τιμολόγησης βρίσκεται στη διατήρηση μιας ισορροπίας μεταξύ της πιθανότητας ύπαρξης ενός άτυχου αλλά καλού οδηγού (ο οποίος υπέστη ζημία και αξίωσε αποζημίωση) και της πιθανότητας ύπαρξης ενός πραγματικά κακού οδηγού (ο οποίος θα υποστεί αύξηση στο ασφάλιστρό του λόγω των ατυχημάτων που προκαλεί). Είναι λοιπόν πιο δίκαιο, να διορθωθούν οι ατέλειες ενός εκ των προτέρων συστήματος τιμολόγησης με τη χρήση ενός επαρκούς συστήματος τιμολόγησης ασφαλίστρων βασισμένο στην οδηγική εμπειρία, όπως είναι αυτό του συστήματος Bonus - Malus, που μπορεί να είναι περισσότερο αποδεκτό από τους ασφαλισμένους από εκείνο του a priori συστήματος τιμολόγησης.

Επίσης, ορισμένοι σημαντικοί παράγοντες κινδύνου δεν λαμβάνονται υπόψη στην κατηγοριοποίηση τους χρησιμοποιώντας εκ των προτέρων κριτήρια. Για παράδειγμα, τα αντανακλαστικά, η τάση για κατανάλωση αλκοόλ ή ο σεβασμός που επιδεικνύει ο οδηγός στον κώδικα οδικής κυκλοφορίας. Η τυχαία μεταβλητή που περιγράφει τα κρυφά χαρακτηριστικά ενός οδηγού αποκαλύπτεται μερικώς από τον αριθμό των απαιτήσεων που θα απαιτηθούν από τους ασφαλισμένους. Αρκετές εμπειρικές μελέτες έχουν δείξει ότι αν οι ασφαλιστικές εταιρίες επιτρέψουν τη χρήση μίας μόνο μεταβλητής για την αξιολόγηση των κινδύνων, τότε θα καταλήξουν σε ένα μοντέλο τιμολόγησης το οποίο δεν θα είναι αντιπροσωπευτικό, και για να δώσουμε ένα παράδειγμα, ο καλύτερος εκτιμητής του αριθμού των απαιτήσεων που απαιτούνται από τον ασφαλισμένο στο μέλλον δεν είναι η ηλικία ή ο τύπος του αυτοκινήτου, αλλά τα ιστορικά δεδομένα των απαιτήσεων.

Ένα τροχαίο ατύχημα μπορεί να τροποποιήσει ή να μεταβάλλει το ποσοστό κινδύνου κατά τέτοιο τρόπο ώστε να μειωθεί ο κίνδυνος εμφάνισης άλλης αποζημίωσης στο μέλλον. Τα συστήματα τιμολόγησης ασφαλίστρων που βασίζονται στην οδηγική εμπειρία του ασφαλισμένου παρέχουν κίνητρα για ασφαλή οδήγηση. Παρόλα αυτά, είναι εξωγενείς οι λόγοι που ερμηνεύουν την ασφάλιση αυτοκινήτου, αφού για τον αριθμό των απαιτήσεων συνυπολογίζεται το γεγονός ότι οι ασφαλισμένοι που απαίτησαν κάποια αποζημίωση στο παρελθόν είναι πιθανό να απαιτήσουν μια απαίτηση και στο μέλλον ((M. Denuit et al (2007)).

1.4 Συστήματα Bones – Malus

Σε πολλές Ευρωπαϊκές και Ασιατικές χώρες, καθώς και αρκετές πολιτείες της Βορείου Αμερικής, οι ασφαλιστές χρησιμοποιούν την τιμολόγηση ασφαλίστρων βασισμένοι στην οδηγική εμπειρία των ασφαλισμένων για να συσχετίσουν το ποσό του ασφαλίστρου με τις παρελθοντικές απαιτήσεις που έχουν απαιτηθεί για την κάλυψη αστικής ευθύνης αυτοκινήτου. Τέτοιου είδους συστήματα επιβαρύνουν τους ασφαλισμένους οδηγούς που είναι υπεύθυνοι για ένα ή περισσότερα ατυχήματα μέσα στην ασφαλιστική περίοδο με επασφάλιστρο (malus) και επιβραβεύουν τους οδηγούς που δεν απαίτησαν κάποια αποζημίωση κατά τη διάρκεια της ασφαλιστικής περιόδου με κάποια έκπτωση (bonus). Τέτοια συστήματα καλούνται Bonus – Malus συστήματα και με αυτά θα ασχοληθούμε σε αυτή τη παράγραφο.

Εκπτώσεις στα ασφάλιστρα ενός οδηγού στην περίπτωση που δεν προκαλέσει κάποιο ατύχημα κατά τη διάρκεια της ασφαλιστικής του κάλυψης εφαρμόστηκαν για πρώτη φορά στο Ηνωμένο Βασίλειο στις αρχές του 1910. Εκείνη την περίοδο η έκπτωση αυτή δινόταν περισσότερο ως κίνητρο για την ανανέωση της ασφαλιστικής κάλυψης παρά ως επιβράβευση για ασφαλή οδήγηση. Τα πρώτα συστήματα bonus-malus εισάχθηκαν για πρώτη φορά από τον Grenander (1957). Σήμερα, υπάρχουν διάφορα τέτοια συστήματα που χρησιμοποιούνται από τις ασφαλιστικές εταιρίες ανά τον κόσμο. Μια τυπική μορφή των συστημάτων αυτών που εφαρμόζεται στο Ηνωμένο Βασίλειο, έχει την ακόλουθη μορφή:

- Για ένα χρόνο χωρίς απαίτηση, έκπτωση 25%
- Δύο χρόνια χωρίς απαίτηση, έκπτωση 40%

- Τρία χρόνια χωρίς απαίτηση, έκπτωση 50%, και
- Τέσσερα χρόνια χωρίς απαίτηση, έκπτωση 60%.

Οι οδηγοί μπορούν να κερδίσουν έναν επιπλέον χρόνο έκπτωσης στο ασφάλιστρό τους για κάθε χρόνο που δεν προκαλούν κάποιο ατύχημα με μέγιστη χρονική διάρκεια τα τέσσερα χρόνια, αλλά χάνουν δύο χρόνια έκπτωσης κάθε φορά που προκαλούν ατύχημα και απαιτήσουν αποζημίωση. Με ένα τέτοιο σύστημα, μέσα σε λίγα μόνο χρόνια ένας ώριμος και ικανός οδηγός μπορεί να επιτύχει την μέγιστη έκπτωση στο ασφάλιστρό του.

Τα συστήματα bonus-malus που χρησιμοποιούνται στην υπόλοιπη Ευρώπη είναι συνήθως πιο ευέλικτα. Οι βαθμίδες που χωρίζονται περιέχουν έναν πεπερασμένο αριθμό επιπέδων, κάθε μια έχει τη δική της σχετικότητα (ή αλλιώς, σχετικό ασφάλιστρο). Το ποσό του ασφαλίστρου που καταβάλλεται από τον ασφαλισμένο προέρχεται από την πρόσθεση ενός βασικού ασφαλίστρου με το ποσό που αντιστοιχεί στην βαθμίδα και το επίπεδο που ανήκει ο οδηγός την δεδομένη στιγμή που συνάπτεται η ασφάλιση. Οι νέοι ασφαλισμένοι ανήκουν σε ένα συγκεκριμένο επίπεδο και πληρώνουν συγκεκριμένο ασφάλιστρο. Μετά από κάθε χρόνο, το ασφάλιστρο κινείται προς τα πάνω ή προς τα κάτω, ανάλογα με τη συμπεριφορά του οδηγού κατά την ασφαλιστική περίοδο (ανάλογα δηλαδή αν θα προκαλέσει ή όχι ατύχημα). Μπορούμε να πούμε συνεπώς ότι ο μηχανισμός bonus-malus μπορεί να θεωρηθεί ως μία βελτιωμένη εκ των προτέρων εκτίμηση του κινδύνου, αφού καταφέρνει να διαχωρίσει την κάθε κλάση κινδύνου σε έναν αριθμό υποκατηγοριών, σύμφωνα πάντα με τις απαιτήσεις που απαιτήθηκαν στο παρελθόν για κάθε κατηγορία οδηγού.

Κατά τη διάρκεια του 20^{ου} αιώνα, οι περισσότερες Ευρωπαϊκές χώρες εφάρμοσαν ένα ομοιόμορφο bonus-malus σύστημα σε όλες τις ασφαλιστικές εταιρίες που λειτουργούσαν στην περιοχή τους. Το 1994 η Ευρωπαϊκή Ένωση αποφάσισε ότι όλα τα μέλη της θα πρέπει να αποσύρουν τα υποχρεωτικά συστήματα bonus-malus, με τον ισχυρισμό ότι τέτοιου είδους συστήματα μειώνουν τον ανταγωνισμό μεταξύ των ασφαλιστικών εταιριών και έρχονται σε αντίθεση με την ελεύθερη τιμολόγηση των ασφαλιστικών προϊόντων. Από εκείνη την ημερομηνία λοιπόν, στο Βέλγιο για παράδειγμα, καταργήθηκε το υποχρεωτικό αυτό σύστημα, αλλά παρόλα αυτά οι ασφαλιστικές εταιρίες που λειτουργούν στη χώρα αυτή ακόμη εφαρμόζουν το παλιό ομοιόμορφο σύστημα (με μικρές τροποποιήσεις για τους ασφαλισμένους που ανήκουν στο χαμηλότερο επίπεδο της διαβάθμισης). Σε άλλες Ευρωπαϊκές χώρες, ωστόσο, οι ασφαλιστικές εταιρίες ανταγωνίζονται στη βάση των συστημάτων bonus-malus (αυτό συμβαίνει στις χώρες Ισπανία και Πορτογαλία).

Το Γαλλικό σύστημα δεν βασίζεται σε κάποιες βαθμίδες αλλά χρησιμοποιεί έναν συντελεστή αύξησης – μείωσης του ασφαλίστρου (coefficient de reduction – majoration). Πιο συγκεκριμένα, το Γαλλικό αυτό σύστημα προτείνει μια επιβάρυνση της τάξεως του 25% ανά αποζημίωση και μια έκπτωση της τάξεως του 5% για τους οδηγούς που δεν προκαλούν ατύχημα κατά τη διάρκεια της ασφαλιστικής περιόδου. Γι' αυτό κάθε ασφαλισμένος χρεώνεται με ένα βασικό ασφάλιστρο και αυτό το βασικό ασφάλιστρο προσαρμόζεται ανάλογα με τον αριθμό των απαιτήσεων που απαιτούνται από την ασφαλιστική εταιρία, πολλαπλασιασμένο με 1.25 κάθε φορά που απαιτείται αποζημίωση και με 0.95 για καμία αποζημίωση (για τον κάθε ασφαλισμένο ξεχωριστά) (M. Denuit et al (2007)).

1.5 Αναλογιστικές και Οικονομικές Πιστοποιήσεις για το σύστημα Bonus – Malus

Τα συστήματα αυτά επιτρέπουν στα ασφάλιστρα να προσαρμόζονται ανάλογα με τους κρυμμένους ατομικούς παράγοντες κινδύνου και δίνουν κίνητρα για ασφαλή οδήγηση, λαμβάνοντας υπόψη το ιστορικό απαιτήσεων. Το γεγονός αυτό μπορεί να πιστοποιηθεί από την πληροφόρηση που αντλείται μεταξύ της εκάστοτε ασφαλιστικής εταιρίας και του ασφαλισμένου οδηγού της. Η μη συμμετρική αυτή πληροφορία εμφανίζεται στην ασφαλιστική αγορά όταν οι ασφαλιστικές εταιρίες αντιμετωπίζουν δυσκολίες στη κρίση της επικινδυνότητας που απορρέει από εκείνους που αποκτούν ασφαλιστική κάλυψη. Υπάρχουν κυρίως δύο όψεις γι' αυτό το φαινόμενο: δυσμενής επιλογή ασφαλισμένων και ηθικός κίνδυνος (adverse selection and moral hazard). Η δυσμενής επιλογή εμφανίζεται όταν οι ασφαλισμένοι γνωρίζουν καλύτερα από την ασφαλιστική τους εταιρία την οδική τους συμπεριφορά και το εκμεταλλεύονται αποκτώντας ασφαλιστική κάλυψη (και αναφερόμαστε ειδικά στους κακούς οδηγούς που προκαλούν ατυχήματα και αναγκάζουν τις ασφαλιστικές εταιρίες να καταβάλλουν απαιτήσεις). Στα πλαίσια όμως της υποχρεωτικής αστικής ευθύνης αυτοκινήτου προς τρίτους, η δυσμενής επιλογή δεν καθιστά τόσο σημαντικό πρόβλημα όσο ο ηθικός κίνδυνος, όπου οι ασφαλιστικές εταιρίες χρεώνουν παρόμοιο ασφάλιστρο για όλους τους ασφαλισμένους. Εφόσον πολλές ετερογενείς οδηγικές συμπεριφορές παρατηρούνται ανάμεσα στο χαρτοφυλάκιο των ασφαλισμένων μιας εταιρίας, χρησιμοποιώντας τις ίδιες εκ των προτέρων μεταβλητές, τότε η δυσμενής επιλογή ασφαλισμένων δεν θα μπορέσει αποφευχθεί.

Όσον αφορά την δυσμενή επιλογή, οι ασφαλισμένοι αποκαλύπτουν μερικώς τους κινδύνους στους οποίους εκτίθενται στη δήλωσή τους για ασφάλιση, γεγονός το οποίο θα πρέπει να ληφθεί σοβαρά υπόψη στην δημιουργία ενός επαρκούς ασφαλιστικού πλαισίου. Στην ύπαρξη μιας μη παρατηρήσιμης ετερογένειας, οι πιο ριψοκίνδυνοι ασφαλιστές θα επιλέξουν μια πιο εξειδικευμένη κάλυψη και οι κινδυνόφοβοι θα επιλέξουν υψηλά αφαιρετέα ποσά για να μειωθεί και το ασφάλιστρό τους.

Είναι πολύ ενδιαφέρον να μελετήσουμε τις προσεγγίσεις των οικονομολόγων και των αναλογιστών στην τιμολόγηση των ασφαλίστρων που βασίζεται στην οδηγική εμπειρία. Στην βιβλιογραφία για την οικονομική επιστήμη, οι εκπτώσεις και οι επιβαρύνσεις στα ασφάλιστρα λειτουργούν κυρίως ως αντίμετρα για τον ηθικό κίνδυνο. Αντίθετα, στην βιβλιογραφία για την αναλογιστική επιστήμη, ο κύριος στόχος είναι η καλύτερη εκτίμηση του ατομικού κινδύνου έτσι ώστε να μπορούν να πληρώνουν όλοι το ασφάλιστρο που τους αναλογεί, βραχυχρόνια, και που αντιστοιχεί στην συχνότητα που προκαλούν τα ατυχήματα οι εκάστοτε οδηγοί. Συνεπώς, οι αναλογιστές ενδιαφέρονται περισσότερο για την περίπτωση της δυσμενούς επιλογής ασφαλισμένων παρά για τον ηθικό κίνδυνο (M. Denuit et al (2007)).

1.6 Κόστος των Απαιτήσεων

Τα περισσότερα συστήματα bonus-malus τα οποία βρίσκονται σε ισχύ ανά τον κόσμο επιβαρύνουν με επιπλέον ασφάλιστρο τον ασφαλισμένο που θα προκαλέσει περισσότερα από ένα ατυχήματα κατά τη διάρκεια της ασφαλιστικής του κάλυψης, αλλά δεν δίνουν σημασία στο ποσό της αποζημίωσης. Για παράδειγμα ένα σοβαρό ατύχημα που θα προκαλέσει σοβαρές σωματικές βλάβες και υλικές ζημιές θα επιβαρύνει το ασφάλιστρο το ίδιο με ένα «επιπόλαιο» ατύχημα. Ο λόγος που η κατηγοριοποίηση του κινδύνου βασίζεται στη συχνότητα των ατυχημάτων είναι η δυσκολία που υπάρχει για να εκτιμηθεί το κόστος των σωματικών βλαβών και άλλων σοβαρών ατυχημάτων. Επίσης, ο λόγος που δεν συμπεριλαμβάνονται τα μεγέθη των απαιτήσεων στα συστήματα αυτά και η εκ των προτέρων κατηγοριοποίηση του κινδύνου είναι διότι απαιτεί μια υπόθεση ανεξαρτησίας ανάμεσα στις τυχαίες μεταβλητές «αριθμός των απαιτήσεων» και «μέγεθος των απαιτήσεων», καθώς επίσης και η πεποίθηση ότι η δεύτερη τυχαία μεταβλητή δεν εξαρτάται από τα χαρακτηριστικά του οδηγού. Αυτό σημαίνει ότι οι αναλογιστές υποθέτουν ότι το κόστος του ατυχήματος είναι, το μεγαλύτερο του μέρος, πέραν της ικανότητας του οδηγού και πιο συγκεκριμένα: ένας προσεκτικός οδηγός μειώνει τον αριθμό των ατυχημάτων, αλλά δεν μπορεί να ελεγχθεί το κόστος αυτών των ενδεχομένων ατυχημάτων.

Οι επιβαρύνσεις των ασφαλίστρων που εφαρμόζονται από τα συστήματα bonus-malus που είναι ανεξάρτητα από τα ποσά της αποζημίωσης, αποτρέπουν τους ασφαλισμένους να δηλώνουν ατυχήματα με μικρές υλικές ζημιές για να αποφύγουν την επιβάρυνση του ασφαλίστρου τους. Οι μικρές ζημιές συνεπώς απορροφούνται από τους ίδιους τους ασφαλισμένους παρά από τις ασφαλιστικές εταιρίες. Σε μερικά συστήματα bonus-malus όμως, η δήλωση ενός σοβαρού ατυχήματος (που περιλαμβάνει σωματικές βλάβες), επιφέρει στον ασφαλισμένο μια πολύ σημαντική επιβάρυνση στο ασφάλιστρό του σε σύγκριση με τη δήλωση ενός επιπόλαιου ατυχήματος με υλικές ζημιές. Στο σύστημα που λειτουργούσε στην Ιαπωνία πριν το 1993, οι απαιτήσεις που απευθυνόντουσαν σε ατυχήματα με σωματικές βλάβες κατέβαζαν το επίπεδο του οδηγού κατά τέσσερα επίπεδα (με κατάλληλη επιβάρυνση ασφαλίστρου), ενώ οι απαιτήσεις για υλικές ζημιές το κατέβαζαν κατά δύο επίπεδα μόνο (M. Denuit et al (2007)).

ΚΕΦΑΛΑΙΟ 2
ΜΕΜΕΙΓΜΕΝΑ ΠΡΟΤΥΠΑ POISSON ΓΙΑ ΤΟΥΣ ΑΡΙΘΜΟΥΣ ΑΤΥΧΗΜΑΤΩΝ

2 Μοντελοποίηση της κατανομής Poisson για τον αριθμό ατυχημάτων

Λαμβάνοντας υπόψη την οικονομική σημασία της ασφάλισης για την αστική ευθύνη έναντι τρίτων στον κλάδο αυτοκινήτου στις βιομηχανικές χώρες, πολλές προσπάθειες έχουν γίνει στην ασφαλιστική βιβλιογραφία έτσι ώστε να βρεθεί ένα πιθανοτικό μοντέλο για την κατανομή του αριθμού ατυχημάτων (αξιώσεων) που αναφέρονται από τους ασφαλισμένους οδηγούς. Αυτό το κεφάλαιο στοχεύει να εισάγει τα βασικά μοντέλα πιθανοτήτων για τα αριθμητικά δεδομένα που θα εφαρμοστούν στην ασφάλιση μηχανοκίνητων οχημάτων.

Η διωνυμική, είναι η διακριτή κατανομή πιθανότητας του αριθμού επιτυχιών σε μια ακολουθία ανεξάρτητων n πειραμάτων, κάθε ένα από τα οποία παράγει την επιτυχία με την πιθανότητα q. Ένα τέτοιο πείραμα επιτυχίας/αποτυχίας καλείται επίσης πείραμα Bernoulli ή δοκιμή Bernoulli. Δύο σημαντικές κατανομές προκύπτουν ως προσεγγίσεις των διωνυμικών κατανομών. Εάν το n είναι αρκετά μεγάλο και η εκτροπή της κατανομής δεν είναι πάρα πολύ μεγάλη (δηλαδή το q δεν είναι πάρα πολύ κοντά στο 0 ή 1), τότε η διωνυμική κατανομή προσεγγίζεται καλά από την κανονική κατανομή. Όταν ο αριθμός των παρατηρήσεων n είναι μεγάλος και η πιθανότητα επιτυχίας q είναι μικρή, τότε η διωνυμική κατανομή προσεγγίζεται πολύ καλά από την κατανομή Poisson με μέση τιμή $\lambda = nq$.

Η κατανομή Poisson ανακαλύφθηκε από τον Poisson (1838). Χαρακτηριστικά, μία τυχαία μεταβλητή η οποία ακολουθεί την κατανομή Poisson είναι μια μέτρηση του αριθμού γεγονότων που εμφανίζονται σε ένα ορισμένο χρονικό διάστημα ή περιοχή. Παραδείγματος χάριν, ο αριθμός αυτοκινήτων που περνούν ένα σταθερό σημείο σε ένα διάστημα πέντε λεπτών, ή ο αριθμός ζημιών που αναγγέλλονται σε μια ασφαλιστική εταιρεία από έναν ασφαλισμένο οδηγό σε μια δεδομένη περίοδο. Ένα τυπικό χαρακτηριστικό που συνδέεται με τη κατανομή Poisson είναι ότι η διακύμανση της τυχαίας μεταβλητής είναι ίση με τη μέση τιμή της. (βλ, Lemaire, 1985).

2.1 Ετερογένεια και Μεμιγμένο πρότυπο Poisson

Η κατανομή Poisson έχει έναν προεξέχοντα ρόλο στη μοντελοποίηση διακριτών δεδομένων που αφορούν αριθμό γεγονότων, κυρίως λόγω της περιγραφικής επάρκειάς της ως πρότυπο όταν υπάρχει μόνο τυχαιότητα και ο πληθυσμός είναι ομοιογενής. Δυστυχώς, αυτό δεν είναι μια ρεαλιστική υπόθεση έτσι ώστε να μοντελοποιήσουμε ασφαλιστικά στοιχεία. Οι μίξεις της Poisson, χρησιμοποιούνται για την περιγραφή των μη - ομογενών πληθυσμών. Ειδικού ενδιαφέροντος είναι οι πληθυσμοί που αποτελούνται από έναν πεπερασμένο αριθμό ομοιογενών υποσυνόλων πληθυσμού. Σε αυτές τις περιπτώσεις η κατανομή πιθανότητας του πληθυσμού μπορεί θεωρηθεί ως πεπερασμένη μείξη κατανομών Poisson.

Το πρόβλημα της απαρατήρητης ετερογένειας προκύπτει επειδή οι διαφορές στη συμπεριφορά οδήγησης μεταξύ των ατόμων δεν μπορούν να παρατηρηθούν από τον αναλογιστή. Μια από τις γνωστές συνέπειες της απαρατήρητης ετερογένειας στην ανάλυση των δεδομένων είναι η υπερδιασπορά (overdispersion): η διακύμανση της μεταβλητής που εξετάζεται είναι μεγαλύτερη από το μέσο. Εκτός από τις επιπτώσεις της στη δομή της μεταβλητής, η απαρατήρητη ετερογένεια έχει τις σημαντικές επιπτώσεις στη δομή πιθανότητας του προτύπου που υποθέτουμε. Τα φαινόμενα της μεγάλης συχνότητας μηδενικού αριθμού γεγονότων καθώς επίσης και της βαριάς δεξιάς ουράς στα περισσότερα ασφαλιστικά στοιχεία μπορούν να ληφθούν ως επίπτωση της απαρατήρητης ετερογένειας. Είναι σύνηθες όταν υπάρχει απαρατήρητη ετερογένεια, να θέσουμε μια τυχαία μεταβλητή (αποκαλούμενη τυχαία επίδραση) στο μέσο της παραμέτρου της κατανομής Poisson κι έτσι να δημιουργήσουμε ένα μικτό πρότυπο Poisson. Σε μια μεμιγμένη διαδικασία Poisson, η ίδια η ετήσια αναμενόμενη συχνότητα ατυχημάτων γίνεται τυχαία.

2.2 Κατανομή Poisson για ομοιογενή χαρτοφυλάκια

Η κατανομή Poisson χρησιμοποιείται για να περιγράψει τη συμπεριφορά των ασφαλισμένων· θεωρείται όμως ανεπαρκής για την περιγραφή του αριθμού των ζημιών σε ένα τέτοιο χαρτοφυλάκιο. Σε αυτό το μοντέλο, υποθέτουμε ότι όλοι οι ασφαλισμένοι υπόκεινται στον ίδιο κίνδυνο. Ο κίνδυνος ο οποίος έχει επέλθει προέρχεται από τυχαίο γεγονός και δεν υπάρχει λόγος να «τιμωρήσουμε» τους υπαίτιους για τη ζημία ασφαλισμένους.

Διατυπώνοντας τις τρεις ακόλουθες αρχικές υποθέσεις ορίζουμε την $N(t, t+ \Delta t)$ η οποία απεικονίζει τον αριθμό των ατυχημάτων στο χρονικό διάστημα $(t, t+ \Delta t)$:

1. $P[N(t,t+\Delta t)=1]=\lambda\Delta t+o(\Delta t)$, (η πιθανότητα να έχουμε ένα γεγονός στο [t, t+Δt) είναι ανάλογη του λ και του μήκους του διαστήματος)

2. $P[N(t,t+\Delta t)>1]=o(\Delta t)$, (η πιθανότητα να έχουμε περισσότερα από ένα γεγονότα στο [t, t+Δt))

3. Ας θέσουμε τα t και t΄ να είναι δύο διαφορετικά χρονικά διαστήματα, τότε:

$$P[\{N(t)=k\} \cap \{N(t')=k'\}]=P[N(t)=k]\cdot P[N(t')=k']$$

Η ερμηνεία αυτών των τριών υποθέσεων είναι εμφανής. Η πρώτη υποθέτει ότι η πιθανότητα ενός ατυχήματος σε ένα χρονικό διάστημα (t, t+ Δt) είναι – αγνοώντας υψηλότερους όρους διάταξης – ανάλογη του μήκους του διαστήματος αυτού και του λ. Η δεύτερη υποθέτει ότι η πιθανότητα να συμβούν δύο ή περισσότερα ατυχήματα σε αυτό το διάστημα είναι αμελητέα ενώ η δε τρίτη ότι ο αριθμός των ατυχημάτων που σχετίζονται με ξεχωριστά χρονικά διαστήματα είναι ανεξάρτητος. Είναι γνωστό ότι τα παραπάνω υποθέτουν ότι ο αριθμός των ατυχημάτων ακολουθεί την κατανομή Poisson. Δηλαδή έχουμε: Αν $p_k(t)=P[N(0,t)=k]$έχουμε

$$p_k(t+\Delta t)=p_k(t)\cdot P[N(t,t+\Delta t)=0]+p_{k-1}(t)\cdot P[N(t,t+\Delta t)=1]+\sum_{i=2}^{k} p_{k-i}(t)\cdot P[N(t,t+\Delta t)=i]$$

$$=p_k(t)[1-\lambda\Delta t+o(t)]+p_{k-1}(t)[\lambda\Delta t+o(\Delta t)]+\sum_{i=2}^{k} p_{k-i}(t)\cdot o(\Delta t)$$

$$=p_k(t)(1-\lambda\Delta t)+p_{k-1}(t)\lambda\Delta t+o(\Delta t).$$

k=0, 1,…(θέτοντας $p_{-1}(t)=0$)

Έτσι

$$\frac{p_k(t+\Delta t)-p_k(t)}{\Delta t}=-\lambda p_k(t)+\lambda p_{k-1}(t)+\frac{o(\Delta t)}{\Delta t}.$$

Παίρνοντας το όριο για Δt → 0,

$$p_k'(t)=-\lambda p_k(t)+\lambda p_{k-1}(t) \quad k=1, 2\ldots$$

$$p_0'(t)=-\lambda p_0(t) \quad k=0.$$

Λύνοντας τις ομάδες των εξισώσεων αυτών παίρνουμε με τις αρχικές υποθέσεις :

$$p_0(0)=1 \text{ και } p_\kappa(0)=0 \text{ αν } k>0,$$

Παίρνουμε :

$$p_k(t)=\frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

Για μία χρονική περίοδο

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Θυμίζουμε ότι η μέση τιμή και η διακύμανση αυτής της κατανομής είναι ίση με λ.

Οι μίξεις των κατανομών χρησιμοποιούνται συνήθως από τις ασφαλιστικές εταιρίες για να μοντελοποιήσουν τον αριθμό των ζημιών και των αποζημιώσεων που καταγράφονται σε μια ασφαλιστική περίοδο. Για παράδειγμα, η χρήση της κατανομής Poisson για τη μοντελοποίηση και περιγραφή του αριθμού των αποζημιώσεων που προκαλούνται από τους ασφαλισμένους κατά τη διάρκεια μια συγκεκριμένης ασφαλιστικής περιόδου απορρίπτεται, αφού η οδηγική συμπεριφορά των ασφαλισμένων ολόκληρου του χαρτοφυλακίου της ασφαλιστικής εταιρίας είναι ετερογενής. Συνεπώς, η παράμετρος της κατανομής Poisson, λ, ποικίλει μεταξύ των ασφαλισμένων οδηγών λόγω των διαφορετικών κινδύνων που ο καθένας υπόκεινται και δεν μπορεί να θεωρηθεί ως σταθερά. Για αυτόν ακριβώς τον λόγο πρέπει να υιοθετηθεί ένα μοντέλο που να ανταποκρίνεται στην αβεβαιότητα της εμφάνισης της συχνότητας των αποζημιώσεων. Είναι λογικό λοιπόν να θεωρηθεί ότι μια διακριτή κατανομή περιγράφει τον αριθμό των αποζημιώσεων ή των ζημιών που προκαλούνται από τους ασφαλισμένους και ότι η οδηγική συμπεριφορά του κάθε ασφαλισμένου ακολουθεί μια κατανομή Poisson με παράμετρο λ. Θεωρείται ένα λ για κάθε ασφαλισμένο ως το αποτέλεσμα μιας τυχαίας μεταβλητής Λ με μια γνωστή συνάρτηση κατανομής U κάτω από ένα δεδομένο επίπεδο κινδύνου του αριθμού των αποζημιώσεων που ακολουθεί την κατανομή Poisson, με παράμετρο τη δεδομένη τιμή του επιπέδου κινδύνου. Άρα, ο αριθμός των αποζημιώσεων ή των ζημιών Ν (για μια δεδομένη ασφαλιστική περίοδο, πχ., ένα έτος) που προκαλείται από έναν οδηγό που επιλέγεται τυχαία από το χαρτοφυλάκιο μιας ασφαλιστικής εταιρίας, ακολουθεί μια μεμιγμένη Poisson κατανομή (D.L. Antzoulakos & S. Chadjiconstantinidis, (2004)).

Άρα, η συνάρτηση πιθανότητας της τυχαίας μεταβλητής Ν δίνεται από τη σχέση:

$$p(n) = P(N = n) = \int_{\lambda_1}^{\lambda_2} p_\lambda(n)\,dU(\lambda),\ \text{n} = 0, 1, 2, \ldots$$

όπου $p_\lambda(.)$ αποτελεί την συνάρτηση πιθανότητας μιας τυχαίας μεταβλητής N_λ που ακολουθεί την κατανομή Poisson με παράμετρο λ > 0, δηλαδή:

$$p_\lambda(n) = P(N_\lambda = n) = e^{-\lambda}\frac{\lambda^n}{n!},\ \text{n} = 0, 1, 2, \ldots$$

Η συνάρτηση κατανομής U μιας μικτικής τυχαίας μεταβλητής L, η οποία ορίζεται σε ένα κλειστό διάστημα $[\lambda_1, \lambda_2]$, $0 \le \lambda_1 < \lambda_2 \le \infty$, αναπαριστά ένα μέτρο ετερογένειας ενός ετερογενούς χαρτοφυλακίου κάποιας ασφαλιστικής εταιρίας. Έστω λοιπόν, $u(.)$ να αποτελεί μια συνάρτηση πυκνότητας πιθανότητας μιας μικτικής τυχαίας μεταβλητής L που θα συμβολίζεται με MP(u) ή MP ή οποία ουσιαστικά αποτελεί την μεμιγμένη Poisson κατανομή της τυχαίας μεταβλητής N με μικτική τυχαία μεταβλητή την L με συνάρτηση πυκνότητας πιθανότητας $u(.)$ (D.L. Antzoulakos & S. Chadjiconstantinidis, (2004)).

Η κλάση των MP κατανομών είναι από τις πιο σημαντικές κλάσεις κατανομών που μοντελοποιούν τον αριθμό των αποζημιώσεων στα χαρτοφυλάκια των ασφαλιστικών εταιριών. Για παράδειγμα, η Αρνητική Διωνυμική κατανομή και η μίξη Poisson – Gamma κατανομή είναι κατάλληλες για να μοντελοποιήσουν δεδομένα με μακριές ουρές. Εφόσον οι κατανομές MP έχουν πιο λεπτές ουρές από την κατανομή Poisson, είναι καταλληλότερες για να μοντελοποιήσουν μετρίσιμα δεδομένα. Ένα άλλο πλεονέκτημα των MP κατανομών είναι ότι μπορεί να ερμηνευτούν και να προέλθουν από μια στοχαστική διαδικασία. Το κύριο μειονέκτημά τους είναι ότι η συνάρτηση πιθανότητάς τους είναι δύσκολο να υπολογιστεί (D.L. Antzoulakos & S. Chadjiconstantinidis, (2004)).

Ας υποθέσουμε ότι η συνάρτηση πυκνότητας πιθανότητας $u(.)$ ικανοποιεί την σχέση:

$$\frac{d}{d\lambda}\log u(\lambda) = \frac{\alpha(\lambda)}{b(\lambda)} = \frac{\sum_{i=1}^{r} a_i \lambda^i}{\sum_{i=0}^{r} b_i \lambda^i}, \quad \lambda \in [\lambda_1, \lambda_2]$$

Μια τυχαία μεταβλητή L υποτίθεται ότι ανήκει στην κλάση κατανομών W(r) αν υπάρχει ένας θετικός αριθμός r και κάποιες σταθερές a_i και b_i $(1 \le i \le r)$ όπως στην περίπτωση της συνάρτησης πυκνότητας πιθανότητας $u(.)$ της τυχαίας μεταβλητής L ικανοποιεί την παραπάνω εξίσωση $(L \in W(r) \quad ή \quad \mathrm{u}(.) \in \mathrm{W(r)})$.

Έχει προταθεί μια φόρμουλα για την αποτίμηση της p(.) στην περίπτωση όπου $L \in W(r)$. Πιο συγκεκριμένα για $n \ge 0$:

$$\sum_{i=1}^{r} \left(a_i - b_i + (n+i+1)b_{i+1}\right)(n+i)^{(i)}\, p(n+i) = C_{\lambda_2}(n) - C_{\lambda_1}(n),$$

Όπου: $a^{(b)} = \prod_{i=1}^{b}(a+1-i), \quad p(-1) = 0, \quad και$

$$\mathrm{C}_\lambda(n) = \begin{cases} b_0 u(0), & \lambda = 0, \quad \mathrm{n} = 0 \\ 0, & \lambda = 0, \quad \mathrm{n} > 0 \\ b(\lambda)u(\lambda)p_\lambda(\mathrm{n}), & 0 < \lambda < \infty \\ 0, & \lambda = \infty \end{cases}$$

Σε αυτό το σημείο θα υποθέσουμε ότι οι $X_1, X_2,...$ είναι μια ακολουθία ανεξάρτητων και ισόνομων διακριτών τυχαίων μεταβλητών με μια κοινή συνάρτηση πιθανότητας f(x), x = 0, 1, 2, … και έστω $S = X_1 + X_2 + ... + X_N$, με το S = 0 αν το N = 0, όπου N και $\{X_i\}_{i \geq 1}$ είναι ανεξάρτητες τυχαίες μεταβλητές. Στη θεωρία κινδύνων η τυχαία μεταβλητή X υποδηλώνει το ποσό μιας αποζημίωσης (ή ζημιάς) σε μια συγκεκριμένη ασφαλιστική περίοδο και η τυχαία μεταβλητή S υποδηλώνει το συνολικό ποσό των αποζημιώσεων (ή ζημιών) ενός χαρτοφυλακίου μιας ασφαλιστικής εταιρίας σε μια συγκεκριμένη ασφαλιστική περίοδο. Η τυχαία μεταβλητή S έχει ακολουθεί μια σύνθετη μεμιγμένη Poisson κατανομή (η οποία συμβολίζεται με CMP(f, u)) και η συνάρτηση πιθανότητάς της δίνεται από τη σχέση:

$$g(x) = P(S = x) = \sum_{n=0}^{\infty} p(n) f^{*n}(x) = \int_{\lambda_1}^{\lambda_2} g_\lambda(x) u(\lambda) \mathrm{d}\lambda, \ \mathrm{x} = 0, 1, 2, \ldots$$

Όπου ο όρος $g_\lambda(.)$ υποδηλώνει την συνάρτηση πιθανότητας της τυχαίας μεταβλητής:

$$S_\lambda = X_1 + X_2 + ... + X_{N_\lambda},$$

η οποία είναι μια σύνθετη Poisson τυχαία μεταβλητή και δίνεται από τον τύπο:

$$g_\lambda(x) = P(S_\lambda = x) = \sum_{n=0}^{\infty} p_\lambda(n) f^{*n}(x), \ \mathrm{x} = 0, 1, 2, \ldots$$

(D.L. Antzoulakos & S. Chadjiconstantinidis, (2004)).

2.3 Αρνητική Διωνυμική Κατανομή για ετερογενή χαρτοφυλάκια

Σε αυτή την περίπτωση υποθέτουμε ότι οι ασφαλισμένοι δεν υπόκεινται όλοι στον ίδιο κίνδυνο. Η συμπεριφορά των ασφαλισμένων είναι ετερογενής και δικαιολογεί για παράδειγμα την εισαγωγή ενός συστήματος Bonus Malus. Πιο αναλυτικά, υποθέτουμε ότι η κατανομή του αριθμού των απαιτήσεων για κάθε ασφαλισμένο ακολουθεί κατανομή Poisson,

$$p_k(\lambda) = \frac{e^{-\lambda} \lambda^k}{k!}, k = 0,1,...$$

της οποίας η παράμετρος ποικίλει από άτομο σε άτομο. Κάθε ασφαλισμένος χαρακτηρίζεται από την τιμή της παραμέτρου του λ, όπου το λ λογίζεται ως τυχαία μεταβλητή. Ας διαλέξουμε τώρα ως κατανομή του λ μία κατανομή Γάμμα με συνάρτηση πυκνότητας πιθανότητας, δηλαδή:

$$\mathrm{u}(\lambda) = \frac{\mathrm{dU}(\lambda)}{\mathrm{d}\lambda} = \frac{\tau^\alpha \mathrm{e}^{-\mathrm{x}\lambda} \lambda^{\alpha-1}}{\Gamma(\alpha)}, (\alpha, \tau) > 0, \text{ με μέση τιμή } \frac{\alpha}{\tau} \text{ και διακύμανση } \frac{\alpha}{\tau^2}.$$

Κάποιες Ιδιότητες για τη Γάμμα Συνάρτηση:

$$\Gamma(\alpha)=\int_0^\infty \mathrm{t}^{\alpha-1}\mathrm{e}^{-\mathrm{t}}dt,\ \Gamma(\alpha+1)=\alpha\Gamma(\alpha),\ \Gamma(\alpha+1)=\alpha!$$

Έστω p_k (*k=0, 1,…*) είναι η κατανομή του αριθμού των απαιτήσεων στο χαρτοφυλάκιο. Τότε έχουμε

$$p_k=\int_0^\infty p_k(\lambda)dU(\lambda)=\int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!}dU(\lambda)=\int_0^\infty \frac{e^{-\lambda(1+\tau)}\lambda^{k+a-1}\tau^a}{k!\Gamma(a)}d\lambda$$

$$=\frac{\tau^a}{k!\Gamma(a)(1+\tau)^{k+a}}\int e^{-\lambda(1+\tau)}[\lambda(1+\tau)]^{k+a-1}d[\lambda(1+\tau)]$$

$$=\frac{\Gamma(k+a)}{\Gamma(k+1)\Gamma(a)}\frac{\tau^a}{(1+\tau)^{k+a}}=$$

$$=\binom{k+a-1}{k}\left(\frac{\tau}{1+\tau}\right)^a\left(\frac{1}{1+\tau}\right)^k,$$

Ορίζοντας ως γενικευμένο συνδυαστικό συντελεστή,

$$\binom{k+a-1}{k}=\frac{\Gamma(k+a)}{\Gamma(k+1)\Gamma(a)}$$

Παίρνουμε μία Αρνητική Διωνυμική Κατανομή με μέση τιμή $m=\dfrac{a}{\tau}$ και διακύμανση $\sigma^2=\dfrac{a}{\tau}\left(1+\frac{1}{\tau}\right)$.

Προσεγγίζοντας την παρατηρούμενη κατανομή από μία αρνητική διωνυμική κατανομή συνεπάγεται, για τη μέθοδο των ροπών, ότι η μέση τιμή *m* και η σ^2 διακύμανση αυτής της κατανομής προσεγγίζεται από τις αντίστοιχες της παρατηρούμενης. (βλ, Lemaire, 1985). Αυτό οδηγεί στους κάτωθι εκτιμητές:

$$\hat{\tau}=\frac{\bar{x}}{s^2-\bar{x}},\qquad \hat{\alpha}=\frac{\bar{x}^2}{s^2-\bar{x}}$$

2.4 Ιδιότητες των Μεικτών Κατανομών Poisson

Πριν προχωρήσουμε στην ανάλυση των ιδιοτήτων των Μεικτών Κατανομών Poisson, θα ήταν πολύ χρήσιμο να αναφερθούμε στις μεμειγμένες κατανομές γενικότερα (για περισσότερα μπορεί κανείς να δει, Κουτσόπουλος, 1999). Αρχικά παραδεχόμαστε ότι μια μεμιγμένη κατανομή περιγράφεται από την περιθώρια κατανομή $F_X(x)$ που προκύπτει από την δεσμευμένη κατανομής $F_{X|Y}(x,y)$ και την συνάρτηση κατανομής $F_Y(y)$, της τυχαίας μεταβλητής Y. Συνεπώς, προκύπτει:

$$f_X(x) = \int_0^{\infty} f_{X|Y}(x,y)dF_Y(y)$$

Για την παραπάνω εξίσωση μπορούμε να πούμε ότι αποτελεί μια συγκέντρωση κινδύνων X, δοθέντος του παραμετρικού χώρου Θ, όπου το είδος της κατανομής είναι η ίδια για όλους τους κινδύνους, με διαφορετική όμως τιμή της παραμέτρου Θ για κάθε κίνδυνο. Επιπλέον, οι τιμές της παραμέτρου Θ «προέρχονται» από κάποια συγκεκριμένη κατανομή $U_\Theta(\theta)$. Ακολουθεί ο παρακάτω ορισμός που μας δίνει κάποιες πληροφορίες για την κατανομή $U_\Theta(\theta)$.

Ορισμός 2.4.1. Αν η αδέσμευτη κατανομή μιας τυχαίας μεταβλητής X δίδεται από τη σχέση $f_X(x) = \int_0^{\infty} f_{X|\Theta}(x,\theta)dU_\Theta(\theta)$, τότε η κατανομή καλείται μεμιγμένη με μεικτική συνάρτηση κατανομής $U_\Theta(\theta)$.

Οι μεμιγμένες κατανομές μας βοηθούν να εντοπίσουμε τις διαφορές στην συχνότητα εμφάνισης των ζημιών ανάμεσα στους κινδύνους ενός χαρτοφυλακίου σε μια δεδομένη χρονική περίοδο. Επίσης, μας βοηθούν να εντοπίσουμε τις διαφορές στις συχνότητες ζημιών μεταξύ δυο διαδοχικών χρονικών περιόδων. Παρόλα αυτά, οι μαθηματικοί χειρισμοί είναι οι ίδιοι και στις δύο περιπτώσεις, αν και πρόκειται για δυο διαφορετικά προβλήματα (Κουτσόπουλος, 1999).

2.5 Μεμειγμένες Κατανομές για το Πλήθος των Ζημιών

Οι ασφαλιστικές εταιρίες ταξινομούν τους κινδύνους με τέτοιο τρόπο έτσι ώστε κάθε κατηγορία κινδύνων που έχουν κοινό ασφάλιστρο να είναι όσο το δυνατόν πιο ομοιογενείς ως προς το επίπεδο κινδύνου. Όμως, είναι εξαιρετικά δύσκολο να υπολογιστούν και να εκτιμηθούν όλα τα χαρακτηριστικά ενός κινδύνου (στην ασφάλιση αυτοκινήτων, π.χ. μπορεί να λάβει κανείς υπόψη το είδος του οχήματος και ενδεχομένως το γεωγραφικό διαμέρισμα, το φύλο ή και την ηλικία του οδηγού, κλπ, δεν είναι όμως εφικτό να λάβει υπόψη τα διανυόμενα χιλιόμετρα κατ' έτος (δηλαδή την έκταση της έκθεσης στον κίνδυνο)). Οι διαφορές στο επίπεδο του ρίσκου μπορούν να λάβουν ποσοτικό χαρακτήρα με τη χρήση μιας ή περισσοτέρων παραμέτρων. Περιοριζόμαστε λοιπόν, στις μεμιγμένες κατανομές για τη συχνότητα των κινδύνων και στις προκύπτουσες απ' αυτές σύνθετες μεμιγμένες κατανομές συνολικών απαιτήσεων.

Αν το πλήθος των ζημιών N είναι μια μεμιγμένη τυχαία μεταβλητή με μεικτική τυχαία μεταβλητή Θ, τότε, ακολουθεί το παρακάτω θεώρημα που μας δίνει τα αποτελέσματα για το μέσο, τη διασπορά και τη ροπογεννήτρια (της συχνότητας) ενός κινδύνου επιλεγμένου τυχαία από τη σχετική συλλογή κινδύνων, έχουμε:

Θεώρημα 2.5.1. Αν η N είναι μεμιγμένη τυχαία μεταβλητή με μεικτική τυχαία μεταβλητή Θ, τότε:

I. $\mathrm{E}(\mathrm{N}) = E(E(N \mid \Theta))$,

II. $\mathrm{Var}(\mathrm{N}) = E(Var(N \mid \Theta)) + Var(E(N \mid \Theta))$,

III. $\mathrm{M}_{\mathrm{N}}(t) = E(M_{N|\Theta}(\mathrm{t}))$

Στη δεύτερη σχέση του παραπάνω Θεωρήματος, ο όρος Var(E(N|Θ)), που εκφράζει τη συνολική διασπορά, αντιπροσωπεύει την πρόσθετη διασπορά που προκύπτει από την ετερογένεια που διακρίνει τη συχνότητα των ζημιών της υπό θεώρηση συλλογής κινδύνων. Επίσης, όσον αφορά την τυχαία μεταβλητή N μπορούν να υπολογιστούν και κάποια άλλα χαρακτηριστικά της (π.χ. πιθανογεννήτρια, ροπές τάξης $k \geq 3$,κλπ) εφόσον γενικότερα ισχύει, $E(\phi(N)) = E(E(\phi(N) \mid \Theta))$.

2.5.1 Μεμειγμένες και Σύνθετες Μεμειγμένες Poisson

Στην περίπτωση μιας μεμιγμένης Poisson(Λ), για την μέση τιμή, τη διακύμανση και την ροπογεννήτρια της, παραθέτουμε το παρακάτω Θεώρημα:

Θεώρημα 2.5.1.1. Αν η S είναι σύνθετη μεμιγμένη Poisson με παράμετρο Λ, τότε:

I. E(N) = E(Λ), Var(N) = Var(Λ) + E(Λ), $M_S(t) = M_\Lambda(e^t - 1)$,

Σύμφωνα με το Θ.2.5.1.1., προκύπτει και το παρακάτω Θεώρημα:

Θεώρημα 2.5.1.2. Αν $\Pr(N = n \mid \Lambda = \lambda) = e^{-\lambda}\dfrac{\lambda^n}{n!}$, n = 0,1,2,… και $f_\Lambda(\lambda) = \dfrac{\beta^a}{\Gamma(a)}\lambda^{a-1}e^{-\beta\lambda}$, λ>0 τότε $\Pr(N = n) = \dfrac{\Gamma(n+a)}{n!\Gamma(a)}\left(\dfrac{\beta}{\beta+1}\right)^a\left(\dfrac{1}{\beta+1}\right)^n$ n=0,1,2,…

όπου φαίνεται ότι η μείξη τυχαίων μεταβλητών Poisson με συνάρτηση πυκνότητας πιθανότητας Γάμμα (με παραμέτρους α και β) οδηγεί σε αρνητική διωνυμική τυχαία μεταβλητή με παραμέτρους r = a, $p = \dfrac{\beta}{\beta+1}$, $q = \dfrac{1}{\beta+1}$.

Οι μεμιγμένες Poisson έχουν μια σειρά από σημαντικές και χρήσιμες ιδιότητες. Πέρα από το μέσο Ε(Ν) λοιπόν, μπορούν να υπολογιστούν όλες οι παραγοντικές ροπές μιας μεμειγμένης Poisson N με μεικτική Λ, αρκεί να ικανοποιούν τη σχέση:

$$E((N)_k) = E(\Lambda^k), \qquad k = 1,2,3,...$$

Προκύπτει ότι η πιθανογεννήτρια της τυχαίας μεταβλητής Ν είναι $P_N(t) = M_\Lambda(\ln(t)) = M_\Lambda(t-1)$ και η παραπάνω σχέση, προέρχεται από παραγώγιση της $P_N(t) = M_\Lambda(t-1)$ και αντικατάσταση t = 1. Επίσης, κάθε μεμιγμένη Poisson έχει διασπορά μεγαλύτερη από τη μέση τιμή της, εφόσον ισχύει η σχέση Var(N) = Var(Λ) + E(Λ) > E(Λ) = E(N). Κατά συνέπεια, οι μεμιγμένες Poisson (όπως, π.χ. η αρνητική διωνυμική) θεωρούνται ότι παρέχουν καλύτερες προβλέψεις από την αρχική Poisson, εφόσον οδηγούν σε πιο δυσμενείς προβλέψεις για την συχνότητα εμφάνισης των ζημιών και, επομένως, σε μια περισσότερο συντηρητική ασφαλιστική πολιτική από τις εταιρίες ανάληψης κινδύνων.

Στην αναλογιστική επιστήμη είναι γνωστό ότι το άθροισμα ανεξαρτήτων μεμιγμένων Poisson κατανομών είναι επίσης μεμιγμένη Poisson. Εξίσου σημαντικό είναι το γεγονός ότι κάθε μεμιγμένη Poisson προέρχεται από μια μοναδική μεικτική τυχαία μεταβλητή. Έτσι, π.χ. η Pascal $\binom{n+r-1}{r-1} p^r q^n$ είναι μεμιγμένη Poisson με μεικτική Γάμμα $\dfrac{\left(\frac{p}{q}\right)^r}{(r-1)!} \lambda^{r-1} e^{\frac{p}{q}\lambda}$ και μόνο αυτή. Τέλος, υπάρχει στενή σχέση μεταξύ μεμιγμένων Poisson και σύνθετων Poisson.

Η πιο συνήθης μεικτική κατανομή για μια Poisson είναι η Γάμμα. Αν η αρνητική διωνυμική που προκύπτει δεν κρίνεται ικανοποιητική (κυρίως στη δεξιά ουρά), μπορεί να αποδειχθεί πιο κατάλληλη μεικτική η αντίστροφη κανονική (Inverse Gaussian), με $f(x) = \left(\frac{\mu}{x}\right)^{\frac{3}{2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mu}{x}\frac{(x-\mu)^2}{2\sigma^2}}$ x>0, της οποίας η λοξότητα υπερβαίνει τη λοξότητα της αρνητικής Διωνυμικής με τον ίδιο μέσο μ και την ίδια διασπορά σ^2. Η μεμιγμένη Poisson με μεικτική αντίστροφη κανονική έχει μέσο μ και διασπορά $\mu + \sigma^2$. Ανεξάρτητες μεμιγμένες Poisson με μεικτική αντίστροφη Gaussian αθροίζονται σε μεμιγμένη Poisson με μεικτική αντίστροφη Gaussian υπό τον όρο ότι $\sigma_i^2 = \beta\mu_i, \forall i$. Η μεμιγμένη Poisson – αντίστροφη Gaussian ανάγεται στην Poisson με λ = μ όταν ο λόγος $\frac{\sigma^2}{\mu}$ τείνει στο μηδέν (Denuit et al, (1999)).

Άλλες μικτικές κατανομές για την Poisson μπορούν να θεωρηθούν οι: γενικευμένη αντίστροφη κανονική και η γενικευμένη αρνητική διωνυμική. Επίσης, η κατανομή Βήτα έχει χρησιμοποιηθεί ως μεικτική τόσο με την αρνητική διωνυμική όσο και με τη διωνυμική. Ο ορισμός των κατανομών αυτών (των λεγόμενων ζημιοκατανομών (loss distributions)) αναλύεται στο 3ο Κεφάλαιο της παρούσας εργασίας.

2.5.2 Τυποποίηση της Μεικτικής Τυχαίας Μεταβλητής

Πολύ συχνά χρησιμοποιείται ο μετασχηματισμός μια μεικτικής τυχαίας μεταβλητής Θ σε Θ΄ = Θ/ Ε(Θ), έτσι ώστε η μετασχηματισμένη τυχαία μεταβλητή να έχει μέση τιμή ίση με τη μονάδα. Στην περίπτωση, π.χ. της κατανομής Poisson, θέτουμε την παράμετρο της Poisson ίση με λQ, όπου Q τυχαία μεταβλητή με μέση τιμή E(Q) = 1. Η δεσμευμένη συνάρτηση πυκνότητας της Poisson όταν Q = q είναι τώρα $e^{-\lambda q}\frac{(\lambda q)^n}{n!}$ n = 0,1,2,… η αδέσμευτη πιθανότητα Pr(N = n) είναι ίση με $\int_0^\infty e^{-\lambda q}\frac{(\lambda q)^n}{n!}dU_Q(q)$, οπότε έχουμε ότι:

Θεώρημα 2.5.2.1

I. E(N) = λ, $Var(N)=\lambda+\lambda^2\mathrm{Var(Q)}$, $M_N(t)=M_Q\left(\lambda\left(\mathrm{e}^t-1\right)\right)$

Στην περίπτωση μεμιγμένης Poisson N με τυποποιημένη μεικτική τυχαία μεταβλητή Q, παραθέτουμε το παρακάτω Θεώρημα που περιγράφει την 3η και 4η ροπή της τυχαίας μεταβλητής N.

Θεώρημα 2.5.2.2. Αν η N είναι Poisson με μεικτική τυχαία μεταβλητή Q, τότε $k_3(N)=\lambda+3\lambda^2Var(Q)+\lambda^3k_3(Q),\ k_4(N)=\lambda+7\lambda^2Var(Q)+6\lambda^3k_3(Q)+\lambda^4k_4(Q).$

Στην ειδική περίπτωση όπου η μεικτική τυχαία μεταβλητή Q είναι Γάμμα με μέση τιμή E(Q) = 1, δηλαδή, αν $u(q)=\frac{h}{\Gamma(h)}q^{k-1}e^{-kq}$, q>0, τότε θα οδηγηθούμε στην αρνητική διωνυμική με μεμιγμένη Poisson που περιγράφεται από τη σχέση για n = 0,1,2,…:

$$\Pr(N=n)=\binom{n+h-1}{h-1}\left(\frac{h}{\lambda+h}\right)^h\left(\frac{\lambda}{\lambda+h}\right)^n$$

Σύμφωνα με τις παραπάνω σχέσεις προκύπτει και το ακόλουθο Θεώρημα:

Θεώρημα 2.5.2.3. Για την Poisson N με παράμετρο λQ και μεικτική συνάρτηση πυκνότητας πιθανότητας $u(q)=\frac{h}{\Gamma(h)}q^{h-1}e^{-kq}$, q>0, ισχύουν οι σχέσεις:

i. $\mathrm{E(N)} = \lambda,\ Var(N) = \lambda + \frac{\lambda^2}{\mathrm{h}}, M_N(t) = \left[\frac{h}{(\lambda + \mathrm{h}) - \lambda e^{t}}\right]^k$

ii. $k_3(N) = \lambda + 3\frac{\lambda^2}{h} + 2\frac{\lambda^3}{h^2}$, $k_4(N) = \lambda + 7\frac{\lambda^2}{h} + 12\frac{\lambda^3}{h^2} + 6\frac{\lambda^4}{h^3}$.

Οι δείκτες λοξότητας και κυρτότητας είναι, για μεμιγμένη Poisson γενικά, $\frac{k_3(N)}{\left[\lambda + \lambda^2 Var(Q)\right]^{\frac{3}{2}}}$, $\frac{k_4(N)}{\left[\lambda + \lambda^2 Var(Q)\right]^2}$ και για την αρνητική διωνυμική ειδικά, $\frac{2\lambda + \mathrm{h}}{\sqrt{\lambda \mathrm{h}(\lambda + \mathrm{h})}}$ και $\frac{6\lambda^2 + 6\lambda h + h^2}{\lambda h(\lambda + h)}$.

2.6 Σύνθετες κατανομές για το πλήθος ζημιών

Οι σύνθετες κατανομές για το πλήθος των ζημιών αποτελούν χρήσιμα εργαλεία για να αντιμετωπιστούν από τις ασφαλιστικές εταιρίες περιπτώσεις όπου πολλαπλές απαιτήσεις για αποζημίωση μπορεί να προκύψουν από το ίδιο ατύχημα ή επανειλημμένες απαιτήσεις για αποζημίωση μπορεί να προκύψουν (μέσα σε μία περίοδο) από τον ίδιο ασφαλισμένο κίνδυνο, κλπ. Αν για παράδειγμα, ο αριθμός των περιπτώσεων που αξιώθηκε αποζημίωση κατά ατύχημα είναι τυχαία μεταβλητή Μ και το πλήθος των ατυχημάτων (σε μια χρονική περίοδο) είναι τυχαία μεταβλητή Κ, τότε το συνολικό πλήθος περιπτώσεων προς αποζημίωση (κατά την περίοδο) είναι $N = M_1 + M_2 + ... + M_K$.

Στο σημείο αυτό θα δώσουμε μερικά παραδείγματα (τρία στο σύνολο) σύνθετων κατανομών για το πλήθος των απαιτήσεων Ν. Η τυχαία μεταβλητή Κ υποθέτουμε ότι ακολουθεί κατανομή Poisson, οπότε το πλήθος Ν των απαιτήσεων έχει ροπογεννήτρια $e^{\lambda(M_M(t)-1)}$, όπου $M_M(t)$ είναι η ροπογεννήτρια της δευτερεύουσας τυχαίας μεταβλητής Μ (αριθμού απαιτήσεων κατά ατύχημα). Στην περίπτωση τώρα που η Μ έχει λογαριθμική συνάρτηση πυκνότητας $f_M(m) = -\frac{1}{m \ln p} q^m$ m = 1, 2, 3, …, τότε η Ν έχει αρνητική διωνυμική συνάρτηση πυκνότητας $f_N(n) = \binom{n+r-1}{r-1} p^r q^n$ n = 0, 1, 2, …, με $r = \frac{\lambda}{\ln p}$.

Αν η τυχαία μεταβλητή Μ ακολουθεί τη διωνυμική κατανομή με αριθμό δοκιμών m, η ροπογεννήτρια της Ν είναι $e^{\lambda\left[\left(p+qe^t\right)^m - 1\right]}$. Όταν m = 1 (που σημαίνει το πολύ μια αποζημίωση κατά ατύχημα), η N ακολουθεί την κατανομή Poisson με παράμετρο λq. Ο μέσος της N είναι mλq και η διασπορά mλq[1 + (m-1)q] (Κουτσόπουλος, 1999).

Σε περίπτωση που (πέραν της τυχαίας μεταβλητής K) ακολουθεί κατανομή Poisson και η τυχαία μεταβλητή M, τότε η ροπογεννήτρια της N είναι $e^{\lambda_\kappa \left\{\exp\left[\lambda_{\mathrm{M}}(e^t - 1)\right] - 1\right\}}$. Η λοξότητα αυτής της κατανομής είναι μικρότερη από την λοξότητα της αρνητικής Διωνυμικής (που έχει τον ίδιο μέσο και την ίδια διασπορά) (Κουτσόπουλος, 1999).

Στην περίπτωση τώρα που η τυχαία μεταβλητή K ακολουθεί την γεωμετρική κατανομή, η ροπογεννήτρια της N είναι $\dfrac{p}{1 - qM_M(t)}$. Αν και η M είναι γεωμετρική με παράμετρο p', η ροπογεννήτρια της N είναι $\dfrac{p - pq'e^t}{1 - qp' - q'e^t} = p + q\dfrac{\dfrac{pp'}{q' + pp'}}{1 - \dfrac{q'}{q' + pp'}e^t}$. Άρα η N είναι απλά «τροποποιημένη γεωμετρική» με p(0) = p και $p(n) = q\left(\dfrac{pp'}{q' + pp'}\right)\left(\dfrac{q'}{q' + pp'}\right)^n$, n = 1, 2, 3, … . Η ίδια ακριβώς γραφή της ροπογεννήτριας δείχνει ότι η N μπορεί να θεωρηθεί, εναλλακτικά, ως σύνθετη Bernoulli με Pr(K = 1) = q και M γεωμετρική με παράμετρο $p'' = \dfrac{pp'}{q' + pp'}$. Αν, αντί για γεωμετρική, η M είναι λογαριθμική (με αυτό το μέσο και αυτή τη διασπορά), η προκύπτουσα N έχει μεγαλύτερη λοξότητα (βλ., Κουτσόπουλος, 1999).

2.7 Τροποποιήσεις Κατανομών

Μια τροποποίηση κατανομών μπορεί να προκύψει όταν η απαίτηση για αποζημίωση είναι μηδενική. Τέτοιου είδους παραδείγματα ζημιογόνων γεγονότων για τα οποία η αποζημίωση που θα αξιωθεί είναι μηδέν, υπάρχουν πολλά. Ένα ασφαλιστήριο μπορεί να έχει εκτιμήσει για παράδειγμα τη λεγόμενη «απαλλαγή» (μη καταβολή μικρών απαιτήσεων λόγω ύπαρξης αφαιρετέου ποσού (deductible)) ή ότι δεν θα καταβληθεί αποζημίωση αν αυτή έχει ήδη καταβληθεί από φορέα κοινωνικής ασφάλισης. Ένας αντασφαλιστής που παρέχει κάλυψη υπερβάλλοντος ζημιάς θα πρέπει να υπολογίσει ότι στις απαιτήσεις που θα καταβάλει, θα αφαιρέσει εκείνες τις ζημιές που εμπίπτουν στην ίδια κράτηση του πρωτασφαλιστή. Υπάρχουν επίσης απαιτήσεις όπου η καταβολή τους δεν γίνονται από τον ασφαλιστή, διότι δεν αποτελούν λόγω ασφάλισης βάση του ασφαλιστηρίου συμβολαίου, άλλες τις οποίες επανεισπράττει (στρεφόμενος κατά τρίτων υπευθύνων), κ.ο.κ. Για όλους αυτούς τους λόγους, είναι συχνά αναγκαία η τροποποίηση του ύψους της αποζημίωσης ώστε, να ληφθεί υπόψη το ενδεχόμενο Χ = 0.

Αν για παράδειγμα η πιθανότητα Pr(X = 0) = π και Y = X| X > 0, η ροπογεννήτρια της τυχαίας μεταβλητής Χ είναι $M_X(t) = \pi + (1-\pi)M_Y(t)$ και η ροπογεννήτρια των συνολικών απαιτήσεων είναι $M_S(t) = M_N(\ln[\pi + (1-\pi)M_Y(t)])$. Η ροπογεννήτρια αυτή μπορεί να απλοποιηθεί σημαντικά στην περίπτωση ορισμένων τυχαίων μεταβλητών για το πλήθος των ζημιών Ν.

Θεώρημα 2.7.1. Αν η ροπογεννήτρια του πλήθους των ζημιών είναι της μορφής $\phi(\theta(e^t - 1))$ και η ροπογεννήτρια του ύψους της αποζημίωσης της μορφής $\pi + (1-\pi)M_Y(t), Y = X \mid X > 0$, τότε η ροπογεννήτρια των συνολικών απαιτήσεων είναι της μορφής $M_S(t) = M_N(\ln[M_Y(t)])$ με παράμετρο (1 – π)θ (όπου θ είναι η παράμετρος της αριθμήτριας τυχαίας μεταβλητής Ν).

Το παραπάνω Θεώρημα δείχνει, ότι ο αριθμός των ζημιών με θετική αποζημίωση έχει κατανομή της ίδιας μορφής με το συνολικό πλήθος των ζημιών Ν, αλλά με παράμετρο (1 – π)θ αντί θ, και η ροπογεννήτρια $M_X(t)$ μπορεί να αντικατασταθεί με τη ροπογεννήτρια $M_Y(t)$ του δεσμευμένου ύψους αποζημίωσης Χ | Χ>0. Τυχαίες μεταβλητές για τις οποίες ισχύει η συνθήκη του Θεωρήματος είναι η Poisson $(\phi(x) = e^x \quad \kappa\alpha\iota \quad \theta = \lambda)$, η αρνητική διωνυμική $\left(\phi(x) = \frac{1}{1-x} \quad \kappa\alpha\iota \quad \theta = \frac{q}{p}\right)$ και η διωνυμική $(\phi(x) = (1+x)^m \quad \kappa\alpha\iota \quad \theta = q)$.

ΚΕΦΑΛΑΙΟ 3
ΖΗΜΙΟΚΑΤΑΝΟΜΕΣ

3 Εισαγωγή

Το πρόβλημα αναζήτησης της κατάλληλης κατανομής για την περιγραφή της κατανομής των συνολικών απαιτήσεων προϋποθέτει τον προσδιορισμό της κατανομής του πλήθους των ζημιών και τον προσδιορισμό της κατανομής του ύψους μιας αποζημίωσης. Η επιλογή κατάλληλης κατανομής για το ύψος της αποζημίωσης αποτελεί ένα από τα πιο δύσκολα προβλήματα που αντιμετωπίζονται από τις ασφαλιστικές και αντασφαλιστικές εταιρίες. Οι σχετικές κατανομές φέρουν το όνομα «κατανομές απώλειας» (loss distributions) ή «ζημιοκατανομές» και αποτελούν το αντικείμενο του παρόντος κεφαλαίου. Οι ζημιοκατανομές που έχουν χρησιμοποιηθεί περισσότερο σε αναλογιστικές εφαρμογές δίδονται (μαζί με τα κυριότερα χαρακτηριστικά τους) στο τέλος του κεφαλαίου αυτού.

3.1 Γενικές Παρατηρήσεις

Ίσως το πρώτο πρόβλημα που μπορεί να συναντήσει κανείς μελετώντας ένα χαρτοφυλάκιο με ασφαλιστήρια συμβόλαια (ή αλλιώς ένα χαρτοφυλάκιο κινδύνων), είναι η επιλογή της κατάλληλης κατανομής που περιγράφουν τα στοιχεία του χαρτοφυλακίου αυτού. Ο λόγος που συμβαίνει αυτό είναι διότι τα στοιχεία που συλλέγονται από τις ασφαλιστικές και αντασφαλιστικές εταιρίες συχνά εμφανίζουν συστηματικές αποκλίσεις από το τυχαίο καθεστώς που διέπει τις απαιτήσεις και πρέπει, κατά συνέπεια, να αντιμετωπίζονται με προσοχή και να υποβάλλονται σε αναγκαίες τροποποιήσεις. Η κατανομή των απαιτήσεων μπορεί να εμφανίζει έξαρση του πλήθους των απαιτήσεων που ανήκουν σε μια συγκεκριμένη ομάδα ασφαλισμένων ή είναι κοντά σε κάποιο συγκεκριμένο ύψος. Η συγκέντρωση πολλών απαιτήσεων γύρω από κάποιο ποσό μπορεί να οφείλεται, π.χ., σε όρο απαλλαγής στην κάλυψη ή σε μέγιστο όριο κάλυψης. Επίσης, μπορεί να εμφανιστούν ασυνήθιστες απαιτήσεις συγκρίνοντας τα ποσά ιδίας κράτησης και εκχώρησης.

Ένα εξίσου σημαντικό πρόβλημα είναι η ανίχνευση τάσεων στα δεδομένα. Υπενθυμίζουμε σχετικά ότι ο πληθωρισμός δεν επηρεάζει ομοιόμορφα όλα τα επίπεδα της κάλυψης και η σχετική ανάλυση πρέπει να λαμβάνει υπόψη το γεγονός ότι άλλη περίπτωση είναι η τιμολόγηση κάλυψης με απαλλαγή (κάτω άκρο της κατανομής) και άλλο η κοστολόγηση κάλυψης υπερβάλλοντος ζημιάς (άνω άκρο της κατανομής). Επίσης, η ίδια παρατήρηση ισχύει για τις δοκιμασίες καλής εφαρμογής της αναλυτικής συνάρτησης κατανομής πάνω στα εμπειρικά δεδομένα χαρτοφυλακίου όπου τον κυριότερο ρόλο παίζει η καλή προσαρμογή της θεωρητικής συνάρτησης κατανομής στο φάσμα εκείνο των απαιτήσεων που ενδιαφέρει. Γι' αυτό τον λόγο, δεν είναι ασυνήθιστο δυο διαδοχικά excess loss για την ίδια κάλυψη να τιμολογούνται βάσει διαφορετικών συναρτήσεων κατανομής, έτσι ώστε να επιτυγχάνεται η καλύτερη δυνατή εφαρμογή συνάρτησης κατανομής σε κάθε κάλυψη.

Πέραν όμως όλων αυτών που αναφέραμε στην προηγούμενη παράγραφο, μπορεί από μόνα τους τα στοιχεία να είναι ελλιπή. Αν για παράδειγμα το μόνο στοιχείο που έχουμε είναι ένα περιγραφικό μέτρο όπως αυτό της μέσης τιμής, είμαστε υποχρεωμένοι να περιοριστούμε σε μονοπαραμετρικά πρότυπα και δεν μπορεί να αναμένεται καλή εφαρμογή του προτύπου στα δεδομένα. Γενικά, αν γνωρίζουμε k διαφορετικά χαρακτηριστικά της εμπειρικής κατανομής, μπορούμε να χρησιμοποιήσουμε ένα μοντέλο με k παραμέτρους. Στην περίπτωση όμως που καταλήγουμε σε ένα μοντέλο με πάρα πολλούς παραμέτρους, η χρήση ενός τέτοιου προτύπου κρίνεται ως μη επιθυμητή δεδομένου ότι ένα απλούστερο πρότυπο διευκολύνει και τη διαδικασία της εφαρμογής και την ερμηνεία των αποτελεσμάτων. Εξάλλου, η πιστότητα της εφαρμογής δεν πρέπει να συγχέεται με τις προβλεπτικές δυνατότητες ενός προτύπου. Ο πραγματικός στόχος δεν είναι η τέλεια εφαρμογή στα δεδομένα, αλλά η εύρεση όσο το δυνατόν πιο απλού μοντέλου που περιγράφει τα δεδομένα αυτά. Δεν υπάρχει άλλωστε σωστό πρότυπο και η επιδίωξη είναι ένα πρότυπο που αντικατοπτρίζει σε ικανοποιητικό βαθμό τα ουσιώδη χαρακτηριστικά κάθε περίστασης. Στην εκτίμηση του βαθμού ανταπόκρισης στα εμπειρικά δεδομένα μπορεί να βοηθήσει και η εφαρμογή δύο ή περισσότερων προτύπων.

3.2 Η Επιλογή Προτύπου

Για να προχωρήσουμε στην επιλογή του κατάλληλου προτύπου για τα δεδομένα μας, πρέπει πρώτα να λάβουμε υπόψη και τις τροποποιήσεις στην εφαρμογή της αναλυτικής συνάρτησης κατανομής στα δεδομένα όπως είναι οι απαλλαγές, οι αντασφαλιστικές καλύψεις, κλπ. Πρέπει συνεπώς να βρεθεί μια κατάλληλη κατανομή που περιγράφει το ύψος της αποζημίωσης (για τη συνολική απώλεια) X, ενώ τα διαθέσιμα στοιχεία συνήθως αφορούν στην καταβλητέα αποζημίωση $Y = I(X)$ (όπως αυτή προκύπτει από τους όρους του ασφαλιστηρίου). Στις περισσότερες εφαρμογές, η τυχαία μεταβλητή Y προκύπτει από μετασχηματισμό της τυχαίας μεταβλητής X. Αν, π.χ., η κάλυψη περιλαμβάνει απαλλαγή, έχουμε μετασχηματισμό – μετατροπής της κατανομής στο κάτω άκρο, ενώ η ιδία κράτηση υπό ασφαλιστική κάλυψη excess loss συνεπάγεται αλλαγή στο άνω άκρο της κατανομής. Όταν όμως υπάρχει κάλυψη με απαλλαγή και ταυτόχρονα μέγιστο όριο ευθύνης του ασφαλιστή ή, ισοδύναμα, η ιδία κράτηση σε μια ασφάλιση που έχει απαλλαγή και ταυτόχρονα αντασφαλιστική κάλυψη excess loss ή, ισοδύναμα, μια αντασφαλιστική κάλυψη excess loss με ανώτατο όριο ευθύνης του αντασφαλιστή συνεπάγεται μετατροπή της τυχαίας μεταβλητής X και στα δύο άκρα.

Στην περίπτωση που ένα άκρο της κατανομής της τυχαίας μεταβλητής που περιγράφει το ύψος των απαιτήσεων αποκόπτεται, για οποιοδήποτε λόγο, το τμήμα της κατανομής που απομένει συνήθως τροποποιείται ώστε να αποτελέσει συνάρτηση κατανομής. Σε της σχεδόν της εφαρμογές, η τροποποίηση παίρνει μια από δύο μορφές. Στην πρώτη περίπτωση, η πιθανότητα του τμήματος που αποκόπτεται τοποθετείται, ως διακριτή πιθανότητα, στο σημείο αποκοπής. Είναι η περίπτωση, π.χ., της ιδίας κράτησης που αντιστοιχεί σε κάλυψη υπερβάλλοντος ζημιάς I_d με συνάρτηση πιθανότητας $I_d = \begin{Bmatrix} f_X(x), & x \le d \\ 1 - F_X(d), & x = d \end{Bmatrix}$ που ουσιαστικά για μέγεθος ζημιάς που είναι μικρότερο ή ίσο με το αφαιρετέο ποσό d η δείκτρια συνάρτηση ισούται με την συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X, ενώ για μέγεθος ζημιάς ίση με το αφαιρετέο ποσό, η δείκτρια συνάρτηση ισούται με την συνάρτηση επιβίωσης της τυχαίας μεταβλητής X. Στην δεύτερη περίπτωση, η πιθανότητα του τμήματος που αποκόπτεται αγνοείται και η πιθανότητα που είναι κατανεμημένη πάνω στο μη αποκομμένο τμήμα ανάγεται σε συνάρτηση κατανομής μέσω πολλαπλασιασμού με κατάλληλη σταθερά. Είναι η περίπτωση, π.χ., της ιδίας κράτησης που αντιστοιχεί σε αντασφαλιστική κάλυψη $I_d = \begin{Bmatrix} 0, & X \le d \\ X, & X > d \end{Bmatrix}$ και έχει συνάρτηση κατανομής $\dfrac{F_X(x)}{F(d)}$ (η σταθερά εδώ είναι $\dfrac{1}{F(d)}$). Σε όλες τις περιπτώσεις υφίσταται πρόβλημα αναγωγής των εμπειρικών δεδομένων για την τυχαία μεταβλητή που εκφράζει την καταβλητέα αποζημίωση, σε αναλυτική συνάρτηση κατανομής για την τυχαία μεταβλητή που περιγράφει το ύψος της αποζημίωσης.

Παρόλο που η τελική επιλογή προτύπου γίνεται χρησιμοποιώντας τις στατιστικές δοκιμασίες καλής εφαρμογής, στην όλη διαδικασία κατάλληλων προτύπων συχνά γίνεται χρήση και ορισμένων ad hoc δοκιμασιών βασισμένων σε μεγέθη όπως είναι εκείνα της «μέσης υπόλοιπης ζωής», «περιορισμένης μαθηματικής ελπίδας» και του «συντελεστή απαλοιφής ζημιάς». Η περιορισμένη μαθηματική ελπίδα δίνεται από τον τύπο,

$$E(K_d) = \int_0^d x f_X(x)dx + d[1 - F_X(d)] \tag{1}$$

Ο συντελεστής απαλοιφής ζημιάς (LER) ασφαλιστή είναι σε περίπτωση απαλλαγής

$$LER(d) = \frac{E(K_d)}{E(X)} \tag{1α}$$

σε περίπτωση αντασφαλιστικής κάλυψης υπερβάλλοντος ζημιάς

$$LER(d) = \frac{E(I_d)}{E(X)} \tag{1β}$$

και σε περίπτωση απαλλαγής d και excess loss με ιδία κράτηση d'

$$LER(d,d') = \frac{E(K_d) + E(I_{d'})}{E(X)} \tag{1γ}$$

Για τον λόγο (1β) χρησιμοποιείται και ο όρος EPR (excess premium ratio), όρος που δικαιολογείται από το γεγονός ότι πρόκειται για ποσοστό του ασφαλίστρου που δεν χρειάζεται να εισπραχθεί αν η ασφάλιση έχει ανώτατο όριο ευθύνης d.

3.3 Προκριματικές Δοκιμασίες

Για να μπορέσουμε να εκτιμήσουμε τους παραμέτρους που προκύπτουν από την μελέτη εμπειρικών δεδομένων ενός χαρτοφυλακίου, και πιο συγκεκριμένα της άγνωστης παραμέτρου θ, θα χρησιμοποιήσουμε την συνάρτηση κατανομής $F_X(x,\theta)$, όπου $\theta = (\theta_1, \theta_2, ..., \theta_k)$. Συνεπώς, σύμφωνα με τη μέθοδο των ροπών, π.χ., τα $\theta_i, i = 1,2,...,k$, προκύπτουν από τις εξισώσεις

$$E(X^i) = \frac{1}{n}\sum_{j=1}^{n} x_j^i, \quad i = 1,2,...,k \tag{1}$$

όπου $x_1, x_2, ..., x_n$ είναι οι απαιτήσεις του χαρτοφυλακίου, ενώ οι ροπές $E(X^i)$ είναι βέβαια συναρτήσεις του ζητούμενου θ. Στην απλούστερη εφαρμογή της συγκεκριμένης μεθόδου, επιλέγεται μια συγκεκριμένη συνάρτηση κατανομής $\theta = (\theta_1, \theta_2)$ της οποίας ο μέσος και η διασπορά συμπίπτουν με το μέσο και την διασπορά των εμπειρικών δεδομένων. Όμως, ο αριθμός των συναρτήσεων κατανομής που ικανοποιούν αυτή τη συνθήκη είναι συνήθως μεγάλο. Επιπλέον, η ύπαρξη μιας τέτοιας κατανομής δεν οδηγεί απαραίτητα στην καλή εφαρμογή της συνάρτησης κατανομής σε όλα τα επίπεδα του κινδύνου. Υπάρχει συνεπώς ανάγκη, εύρεσης ενός ελέγχου για την εφαρμογή της συνάρτησης κατανομής σε όλο το φάσμα των τιμών του κινδύνου (Κουτσόπουλος, (1999)).

3.4 Προσδιορισμός Κατάλληλης Ζημιοκατανομής

Για τον προσδιορισμό της κατάλληλης κατανομής που να περιγράφει το ύψος της αποζημίωσης X έχουν καταγραφεί αρκετές μέθοδοι. Οι μέθοδοι αυτοί περιλαμβάνουν διαδικασίες με ορισμένες στατιστικές ιδιότητες, αλλά και πιο πρόχειρες διαδικασίες χωρίς σπουδαίο στατιστικό υπόβαθρο. Μια από αυτές τις μεθόδους, που χρησιμοποιείται όταν ο προσδιορισμός της συνάρτησης κατανομής της τυχαίας μεταβλητής X σε κλειστή μορφή δεν είναι δυνατός ή είναι πολύ δύσκολος, είναι η προσομοίωση. Επειδή όμως δεν θα εξετάσουμε τη μέθοδο αυτή στην παρούσα εργασία, θα αναφέρουμε παραμετρικές μεθόδους με επαρκή στατιστική βάση. Οποιοσδήποτε θέλει να ασχοληθεί ουσιαστικά με την εύρεση κατάλληλης ζημιοκατανομής στα εμπειρικά δεδομένα που εξετάζει, θα πρέπει να εξοικειωθεί με τα διάφορα είδη εκτιμητριών καθώς και με τις ιδιότητές τους (εκτιμήτριες μέγιστης πιθανοφάνειας, Bayes, ελάχιστης απόστασης, ελαχίστου χι-τετράγωνο, ελαχίστων τετραγώνων, συναρτήσεις απώλειας (ζημιάς), συναρτήσεις και κριτήριο minimax, εκτιμήτριες αμερόληπτες, αμερόληπτες ελάχιστης διακύμανσης (αποτελεσματικές, ανισότητα Cramer – Rao, πληροφοριακός αριθμός του Fisher), επαρκείς (θεώρημα Blackwell – Rao), συνεπείς και ασθενώς συνεπείς, ασυμπτωτικά κανονικές και άριστες ασυμπτωτικά κανονικές, πλήρεις (θεώρημα Lehmann – Scheffe)). Εκτός από την γνώση όλων των παραπάνω σε βάθος, απαιτείται και η γνώση των αντίστοιχων μεθόδων για τον προσδιορισμό διαστημάτων εμπιστοσύνης για τις τιμές των παραμέτρων καθώς και γνώση των διαδικασιών ελέγχου των στατιστικών υποθέσεων (τουλάχιστον του λήμματος Neyman – Pearson και του γενικευμένου λόγου πιθανοφάνειας). Τέλος, είναι αναγκαία η γνώση της χρήσης των κυριότερων στατιστικών διαδικασιών καλής εφαρμογής των προτύπων (Κουτσόπουλος, (1999)).

3.5 Κυριότερες Ζημιοκατανομές

Σε αυτή την παράγραφο, θα παραθέσουμε τη συνάρτηση πυκνότητας πιθανότητας, τη συνάρτηση κατανομής, τις ροπές, την επικρατούσα τιμή και την περιορισμένη μαθηματική ελπίδα για ορισμένες ζημιοκατανομές. Ορισμένα από τα αποτελέσματα διατυπώνονται με τη βοήθεια των συναρτήσεων Γ(α), ψ(α), Γ(x;α), Β(x;α,β) που ορίζονται ως:

$$\Gamma(\alpha)=\int_0^\infty x^{a-1}e^{-x}dx \qquad (1),$$

$$\psi(a)=\frac{d}{da}\ln\Gamma(a)=\frac{\Gamma'(a)}{\Gamma(a)} \qquad (2),$$

$$\Gamma(x;a)=\frac{1}{\Gamma(a)}\int_0^x y^{a-1}e^{-y}dy \qquad (3),$$

$$B(x;a,\beta)=\frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)}\int_0^x y^{a-1}(1-y)^{\beta-1}dy \qquad (4).$$

Δίνουμε τώρα τα χαρακτηριστικά οκτώ οικογενειών κατανομών που όλες ορίζονται στο x>0, με εξαίρεση τη λογαριθμική Γάμμα με πεδίο ορισμού x>1.

Κατανομή Burr

$f(x)=a\lambda^a \tau x^{\tau-1}\left(\lambda+x^\tau\right)^{-(a+1)}, a>0, \lambda>0, \tau>0$

$F(x)=1-\lambda^a\left(\lambda+x^\tau\right)^{-a}$.

$p_n : \lambda^{\frac{n}{\tau}}\dfrac{\Gamma\left(1+\frac{n}{\tau}\right)\Gamma\left(a-\frac{n}{\tau}\right)}{\Gamma(a)}, n<a\tau$

$m: 0 \quad \gamma\iota\alpha \quad \tau\le 1, \qquad \left(\dfrac{\lambda(\tau-1)}{\alpha\tau+1}\right)^{\frac{1}{\tau}}, \gamma\iota\alpha \quad \tau>1$

Περιορισμένη Μαθηματική Ελπίδα:

$$\alpha\lambda^{\frac{1}{\tau}}\frac{\Gamma\left(1+\frac{1}{\tau}\right)\Gamma\left(\alpha-\frac{n}{\tau}\right)}{\Gamma(a+1)}B\left(\frac{x^\tau}{\lambda+x^\tau};1+\frac{1}{\tau};a-\frac{1}{\tau}\right)+x\lambda^a\left(\lambda+x^\tau\right)^{-a}$$

Κατανομή Γάμμα

$f(x)=\dfrac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}, a>0, \lambda>0$.

$F(x)=\Gamma(\lambda x;a)$.

$p_n : \dfrac{\prod_{i=1}^{n-1}(a+i)}{\lambda^n}$

$m: 0 \quad \gamma\iota\alpha \quad a\le 1, \qquad \dfrac{a-1}{\lambda}, \gamma\iota\alpha \quad a>1$

Περιορισμένη Μαθηματική Ελπίδα:

$$\frac{a}{\lambda}\Gamma(\lambda x; a+1) + x[1-\Gamma(\lambda x; a)]$$

Κατανομή Pareto

$f(x) = a\lambda^{a}\left(\lambda + x^{\tau}\right)^{-(a+1)}, a > 0, \lambda > 0$.

$F(x) = 1 - \lambda^{a}\left(\lambda + x\right)^{-a}$.

$p_n : \lambda^{n} \dfrac{n!}{\prod_{i=1}^{n}(a-i)}, n < a$

$m : 0$

Περιορισμένη Μαθηματική Ελπίδα:

$$\frac{\lambda}{a-1}\left[1 - a\lambda^{\alpha-1}\left(\lambda + x\right)^{-(a-1)} + (a-1)\lambda^{\alpha}(\lambda + x)^{-a}\right] + x\lambda^{a}\left(\lambda + x\right)^{-a}$$

Κατανομή Weibull

$f(x) = \beta\tau x^{\tau-1}e^{-\beta x^{2}}, \beta > 0, \tau > 0$

$F(x) = 1 - e^{-\beta x^{2}}$

$p_n : \dfrac{\Gamma\left(1+\frac{n}{\tau}\right)}{\beta^{\frac{n}{\tau}}}$

$m : 0 \quad \gamma\iota\alpha \quad \tau \leq 1, \qquad \left(\dfrac{\tau-1}{\beta\tau}\right)^{\frac{1}{\tau}}, \gamma\iota\alpha \quad \tau > 1$

Περιορισμένη Μαθηματική Ελπίδα:

$$\frac{\Gamma\left(1+\frac{1}{\tau}\right)}{\beta^{\frac{1}{\tau}}}\Gamma\left(\beta x^{\tau}; 1+\frac{1}{\tau}\right) + xe^{-\beta x^{\tau}}$$

Κατανομή Log – Normal

$f(x) = \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^{2}}{2\sigma^{2}}}, x > 0$

$$F(x) = \frac{1}{2} erfc\left[-\frac{\ln x - \mu}{\sigma\sqrt{2}} \right] = \phi\left(\frac{\ln x - \mu}{\sigma} \right)$$

Ροπή τάξης s:

$$E\left[X^s\right] = e^{s\mu + \frac{1}{2}s^2\sigma^2}$$

ΚΕΦΑΛΑΙΟ 4
ΓΕΝΙΚΕΥΜΕΝΑ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

4 Εισαγωγή στα Γενικευμένα Γραμμικά Μοντέλα

Και στις δυο περιπτώσεις, γραμμικών και μη γραμμικών μοντέλων, κυρίαρχο ρόλο έχει η Κανονική Κατανομή $\mathrm{N}(\mu, \sigma^2)$ την οποία υποθέσαμε ότι ακολουθεί η απόκριση. Αυτό όμως δε συμβαίνει πάντοτε, π.χ., η απόκριση μπορεί να ακολουθεί την Κατανομή Poisson, δηλαδή, οι αποκρίσεις να είναι της μορφής 0,1,2,3...τα οποία αναπαριστούν τον αριθμό γεγονότων σε συγκεκριμένο χρονικό διάστημα. Στην περίπτωση αυτή, έχουμε τα Γενικευμένα Γραμμικά Μοντέλα (Generalized Linear Models ή GLM) των οποίων τα δεδομένα ακολουθούν κατανομές της Εκθετικής Οικογένειας.

Οι κατανομές που θα εξεταστούν στην παρούσα εργασία είναι οι εξής: Poisson, Αρνητική Διωνυμική και Γάμμα. Για παράδειγμα, αν $\mathrm{y_i}$ οι αποκρίσεις, τότε το Γενικευμένο Γραμμικό Μοντέλο δίνεται από τον τύπο:

$$g(\mu_i) = g[E(y_i)] = x_i^{'}\beta$$

Όπου, $\mathrm{x_i}$ το διάνυσμα των μεταβλητών παλινδρόμησης και β: το διάνυσμα των άγνωστων παραμέτρων.

Στο Κεφάλαιο αυτό θα παραθέσουμε μια περιληπτική περιγραφή του γενικού μοντέλου GLM. Στην κλασσική ανάλυση παλινδρόμησης λοιπόν, θεωρούμε $Y_1, Y_2, ..., Y_\mathrm{n}$ μεταβλητές απόκρισης (dependent variables) και $X_1, X_2, ..., X_\mathrm{n}$ ερμηνευτικές μεταβλητές $\left(X_i = (1, X_{i1}, X_{i2}, ..., X_{i,p-1})^T\right)$ με την υπόθεση ότι υπάρχουν παράμετροι $\beta_0, \beta_1, ..., \beta_{\mathrm{p-1}}$ ώστε:

$Y_i \sim N(\mu_\mathrm{i}, \sigma^2)$ όπου $\mu_\mathrm{i} = \beta_0 + \beta_1 X_{i1} + ... + \beta_1 X_{i,p-1}$

Σκοπός μας είναι η μελέτη του παραπάνω μοντέλου (εκτίμηση των β_i, έλεγχος καλής προσαρμογής, κλπ).

Γενικότερα τώρα, θεωρούμε ότι οι μεταβλητές απόκρισης Y_i είναι δυνατό να προέρχονται από μία γενικότερη οικογένεια κατανομών. Και σε αυτή την περίπτωση, θεωρούμε ότι η μέση τιμή μ_i του Y_i εξαρτάται από ένα σύνολο παραμέτρων $\beta_0, \beta_1, ..., \beta_{\mathrm{p-1}}$. Συγκεκριμένα θεωρούμε ότι:

$$g(\mu_i) = \eta_i = X_i^T \beta$$

για κάποια συνάρτηση g (στο κλασσικό μοντέλο είναι $g(\mu_i) = \mu_i$) την οποία ονομάζουμε συνάρτηση σύνδεσης (link function) και η οποία ουσιαστικά παριστάνει ένα μετασχηματισμό του μέσου g(μ), ο οποίος είναι γραμμικά συσχετισμένος με τις επεξηγηματικές μεταβλητές x. Μια αρκετά ευρεία κλάση κατανομών η οποία περιλαμβάνει την κανονική, την διωνυμική, την Poisson κ.α. είναι η εκθετική οικογένεια κατανομών. Θεωρούμε λοιπόν ότι οι μεταβλητές απόκρισης $Y_1, Y_2, ..., Y_{\mathrm{n}}$ προέρχονται από την εκθετική οικογένεια κατανομών, δηλαδή η Y_{i} έχει συνάρτηση πυκνότητας πιθανότητας (ή συνάρτηση πυκνότητας) της μορφής:

$$f(y_i;\theta_i,\phi) = \exp\left\{\frac{\mathrm{y_i}\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right\}, \qquad \mu_i = E(Y_i) = b'(\theta_i)$$

Σύμφωνα με τα παραπάνω, τα θ_{i}, i = 1,2,…,n εξαρτώνται από ένα μικρότερο σύνολο παραμέτρων $\beta_0, \beta_1, ..., \beta_{\mathrm{p-1}}$, δηλαδή $\theta_{\mathrm{i}} = \theta_{\mathrm{i}}(\beta)$ διότι $g(b'(\theta_i)) = X_i^T \beta$. Μπορούμε να πούμε ότι η επιλογή του b(θ) καθορίζει την κατανομής της μεταβλητής απόκρισης και η επιλογή της g(μ), η οποία καλείται συνάρτηση σύνδεσης, καθορίζει πως ο μέσος σχετίζεται με τις επεξηγηματικές μεταβλητές. Σκοπός μας και πάλι είναι η μελέτη του συγκεκριμένου γενικότερου μοντέλου (εκτίμηση των β_{i}, έλεγχος καλής προσαρμογής, κλπ) (De Jong & Heller, (2008)).

4.1 Εκτίμηση των β_{i}

Η εκτίμηση των β_{i} γίνεται χρησιμοποιώντας τις εκτιμήτριες μέγιστης πιθανοφάνειας (MLE's) οι οποίες ως γνωστό προκύπτουν από τη λύση του συστήματος:

$$\frac{\vartheta}{\vartheta\beta} \mathrm{l}(\beta) = 0$$

ως προς β_{i} όπου $l(\beta) = \mathrm{lnL}(\beta)$και $\mathrm{L}(\beta) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{f}(\mathrm{y_i};\theta_{\mathrm{i}})$ είναι η συνάρτηση πιθανοφάνειας. Συγκεκριμένα, έχουμε το σύστημα:

$$\frac{\vartheta}{\vartheta\beta} \mathrm{l}(\beta) = \frac{\vartheta}{\vartheta\beta} \ln \prod_{i=1}^{n} f(y_i;\theta_{\mathrm{i}}) = \frac{\vartheta}{\vartheta\beta} \sum_{i=1}^{n} \left(\frac{y_i\theta_{\mathrm{i}} - b(\theta_{\mathrm{i}})}{a_i(\phi)} + c(y_i,\phi) \right) = 0$$

$$\Leftrightarrow \left(\frac{\vartheta}{\vartheta\beta_0} \sum_{i=1}^{n} \left(\frac{y_i\theta_{\mathrm{i}} - b(\theta_{\mathrm{i}})}{a_i(\phi)} + c(y_i,\phi) \right), ..., \frac{\vartheta}{\vartheta\beta_{p-1}} \sum_{i=1}^{n} \left(\frac{y_i\theta_{\mathrm{i}} - b(\theta_{\mathrm{i}})}{a_i(\phi)} + c(y_i,\phi) \right) \right) = (0,0,...,0)$$

Λόγω του ότι το τελευταίο σύστημα (p εξισώσεων με p αγνώστους $\beta_0, \beta_1, ..., \beta_{p-1}$) βρίσκεται γενικά σε πεπλεγμένη μορφή (είναι μη γραμμικό), ο υπολογισμός των λύσεων γίνεται προσεγγιστικά μέσα από μία επαναληπτική διαδικασία (αρχικά υπολογίζουμε $\beta^{(0)}$, με βάση αυτό υπολογίζουμε ένα $\beta^{(1)}$, με βάση το $\beta^{(1)}$ υπολογίζουμε ένα $\beta^{(2)}$ κοκ ώστε $\beta^{(k)} \to \beta$ για $k \to \infty$).

Συγκεκριμένα συχνά χρησιμοποιείται μια παραλλαγή της προσεγγιστικής (p-διάστατης) μεθόδου $Newton - Raphson$[1] όπου:

$$\beta^{(m)} = (X^T W^{(m-1)} X)^{-1} X^T W^{(m-1)} z^{(m-1)}$$

Όπου τα $W^{(m-1)}, z^{(m-1)}$ υπολογίζονται με βάση τα $\beta^{(m-1)}$. Συγκεκριμένα,

$$\mathrm{W}^{(\mathrm{m-1})} = diag(1/(V(\mu_i^{(m-1)}))g'(\mu_i^{(m-1)})^2))_i, \quad z_i^{(m-1)} = \eta_i^{(m-1)} + (y_i - \mu_i^{(m-1)})g'(\mu_i^{(m-1)})$$

$$\mu_i^{(m-1)} = g^{-1}(x_i^T \beta^{(m-1)}), \quad \eta_i^{(m-1)} = x_i^T \beta^{(m-1)}$$

Στο αρχικό βήμα συνήθως επιλέγουμε: $\mu_i^{(0)} = Y_i, n_i^{(0)} = g(Y_i), \beta^{(0)} = \left(X^T X\right)^{-1} X^T g(Y)$. Η επαναληπτική διαδικασία σταματάει όταν $\sum_{i=1}^{n} \left(\beta_1^{(m)} - \beta_i^{(m-1)}\right)^2 < \varepsilon$ για δεδομένο ε, ή μετά από προκαθορισμένο αριθμό βημάτων (συνήθως όλη η παραπάνω διαδικασία για την εκτίμηση του **β** γίνεται αυτόματα από το στατιστικό πακέτο που χρησιμοποιείται από τον χρήστη).

([1] Γενικά η n-διάστατη N-R μέθοδος χρησιμοποιείται για την εύρεση των ριζών ενός συστήματος εξισώσεων

$$f(x) = 0 \Leftrightarrow (f_1(x_1, x_2, ..., x_p), ..., f_p(x_1, x_2, ..., x_p)) = (0,0,...,0)$$

Ο γενικός αναγωγικός τύπος που χρησιμοποιείται είναι ο

$$x^{(m)} = x^{(m-1)} - J^{-1}_{x^{(m-1)}} \cdot f(x^{(m-1)}), \quad J = \left[\frac{\vartheta f_i}{\theta x_j}\right]_{i,j},$$

Στη μονοδιάστατη περίπτωση όμως έχουμε: $\mathrm{x}^{(m)} = x^{(m-1)} - \dfrac{f\left(x^{(m-1)}\right)}{f'\left(x^{(m-1)}\right)}$.

Στη συγκεκριμένη περίπτωση είναι $f_i = \dfrac{\vartheta l(\beta)}{\vartheta \beta_i}, x_i = \beta_i$ και άρα $J = \left[\dfrac{\vartheta^2 l(\beta)}{\vartheta \beta_i \vartheta \beta_j}\right]_{i,j}$).

4.2 Έλεγχος καλής προσαρμογής του μοντέλου

Ο έλεγχος για την καλή προσαρμογή του μοντέλου γίνεται με την χρήση του γενικευμένου λόγου πιθανοφανειών. Συγκεκριμένα θέλουμε να ελέγξουμε την μηδενική υπόθεση $H_0 : g(\mu) = \mathrm{X}\beta$, ότι δηλαδή ο μέσος σχετίζεται με τις επεξηγηματικές μεταβλητές. Σύμφωνα με το κριτήριο του γενικευμένου λόγου πιθανοφανειών απορρίπτουμε την H_0 όταν:

$$\frac{\sup\limits_{\mu\in\Theta_0} L(\mu, y)}{\sup\limits_{\mu\in R^n} L(\mu, y)} < c \Leftrightarrow -2\ln\frac{L_c(\hat{\mu}, y)}{L_f(y, y)} = 2(l_f(y, y) - l_c(\hat{\mu}, y)) = S_{c,f}(y, \hat{\mu}) > c'$$

Όπου $\Theta_0 = \{\mu : \mathrm{g}(\mu) = \mathrm{X}\cdot\beta, \beta \in \Re^{\mathrm{p}}\}$. Η στατιστική συνάρτηση $S_{c,f}\left(y, \hat{\mu}\right) = D_{c,f}\left(y, \hat{\mu}\right)/\phi$ καλείται Scaled Deviance και ασυμπτωτικά ακολουθεί την κανονική χ^2 κατανομή με $\dim\left(R^n\right) - \dim\left(\Theta_0\right) = n - p$ βαθμούς ελευθερίας. Άρα απορρίπτουμε την H_0 (ότι δηλαδή το μοντέλο είναι σωστό) όταν:

$$S_{c,f}(y, \hat{\mu}) > \chi^2_{n-p;a}$$

Ένα άλλο κριτήριο καλής προσαρμογής είναι το χ^2 του Pearson (ασυμπτωτικά ισοδύναμο του $\mathrm{S}_{\mathrm{c,f}}\left(y, \hat{\mu}\right)$):

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \sim \chi^2_{n-p}$$

Σύμφωνα με αυτό το κριτήριο απορρίπτουμε την H_0 όταν $X^2 > \chi^2_{\mathrm{n\text{-}p,x}}$ (τα στατιστικά πακέτα συνήθως δίνουν την τιμή της $\mathrm{S}_{\mathrm{c,f}}$ ή του συντελεστή χ^2 του Pearson) (Denuit et al., (2007)).

4.3 Σύγκριση μοντέλων

Αν τώρα θέλουμε να ελέγξουμε την υπόθεση της μορφής $H_0 : g(\mu) = \mathrm{X}\beta$ έναντι του μεγαλύτερου μοντέλου $H_1 : g'(\mu) = \mathrm{X}'\beta'$, τότε πάλι σύμφωνα με το κριτήριο του γενικού λόγου πιθανοφανειών, απορρίπτουμε την H_0 όταν:

$$\frac{\sup\limits_{\mu\in\Theta_0} L(\mu,y)}{\sup\limits_{\mu\in\Theta_1} L(\mu,y)} < c \Leftrightarrow -2\ln\frac{L_c(\hat{\mu},y)}{L_f(y,\hat{\mu}')} = 2(l(y,\hat{\mu}') - l(y,\hat{\mu}')) = S_{c,f}(y,\hat{\mu}) - S_{c,f}(y,\hat{\mu}) > c'$$

και όμοια με τα παραπάνω, ακολουθεί ασυμπτωτικά τη χ^2 κατανομή με $\dim(\Theta_1) - \dim(\Theta_0) = p' - p$ βαθμούς ελευθερίας. Άρα, απορρίπτουμε την H_0 (ότι δηλαδή το μικρότερο μοντέλο είναι το σωστό) όταν:

$$S_{c,f}(y,\hat{\mu}) - S_{c,f}(y,\hat{\mu}') > \chi^2_{p'-p}$$

(Denuit et al., (2007))

4.4 Έλεγχοι υποθέσεων για το β

Οι έλεγχοι για το **β** βασίζονται στο γεγονός ότι αν ισχύει η $H_0 : \beta = \beta_0$, τότε ασυμπτωτικά (βλ. παρατήρηση παρακάτω) ισχύουν τα εξής:

I. $U \sim N(0, Inf(\beta_0)) \Rightarrow U^T (Inf(\beta_0))^{-1} U \sim \chi^2_p,\ Inf(\beta) = X^T W X,\ W = diag\left(1/V(y_i)(g'(\mu_i))^2\right)$

II. $\hat{\beta} \sim N(\beta_0, Inf(\beta_0)^{-1}) \Rightarrow \left(\hat{\beta} - \beta_0\right)^T Inf(\beta_0)\left(\hat{\beta} - \beta_0\right) \sim \chi^2_p$

III. $-2\ln\dfrac{\sup\limits_{\theta\in\Theta_0} L(\beta_0)}{\sup\limits_{\theta\in\Theta_1} L\left(\hat{\beta}\right)} = 2\left(l\left(\hat{\beta}\right) - l(\beta_0)\right) \sim \chi^2_p$

4.5 Κατάλοιπα

Οι υποθέσεις του γενικευμένου γραμμικού μοντέλου για τα κατάλοιπα είναι διαφορετικές από εκείνες του κανονικού γραμμικού προτύπου, που οδηγεί σε διαφορετικές διαγνωστικές δοκιμές. Οι στόχοι, εντούτοις, είναι οι ίδιοι, όπως για παράδειγμα για να ανιχνεύσουν τα στοιχεία ενάντια στις αρχικές υποθέσεις. Τα κατάλοιπα χρησιμοποιούνται για να ελέγξουν την καταλληλότητα μιας επιλεγμένης κατανομής απόκρισης και για τις απομακρυσμένες (ακραίες) τιμές. Τα κατάλοιπα και η τυποποιημένη τους μορφή είναι απαραίτητα για να διαμορφώσουν τον έλεγχο για το κανονικό γραμμικό πρότυπο. Για τα γενικευμένα γραμμικά μοντέλα, αυτά τα κατάλοιπα δεν κατανέμονται ούτε κανονικά, ούτε έχουν σταθερή διακύμανση, για οποιαδήποτε κατανομή απόκρισης εκτός από εκείνη της κανονικής κατανομής.

Όσον αφορά τώρα τα κατάλοιπα των μετρήσεων, αυτά υπολογίζονται από τον τύπο (του Pearson):

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{V}(y_i)}}$$

Τα τυποποιημένα κατάλοιπα κατά τον Pearson υπολογίζονται από τον τύπο:

$$r_i^{PS} = \frac{r_i^P}{\sqrt{1-h_{ij}}} \quad \left((h_{ij}) = W^{-\frac{1}{2}} X \left(X^T W X \right)^{-1} X^T W^{-\frac{1}{2}} \right)$$

Τέλος, τα κατάλοιπα απόκρισης υπολογίζονται:

$$r_i^D = \operatorname{sgn}(y_i - \hat{\mu}_i)\sqrt{2\phi(l_f(y_i, y_i) - l_c(\hat{\mu}_i, y_i))}, \quad \left(D_{c,f} = \sum_{i=1}^{n} \left(r_i^D \right)^2 \right)$$

Παρατήρηση. (Ασυμπτωτικά αποτελέσματα για τις εκτιμήτριες μέγιστης πιθανοφάνειας (MLE's) και την Score function U). Έστω $\mathrm{Y}_1, \mathrm{Y}_2,.,, \mathrm{Y}_\mathrm{n}$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με συνάρτηση πιθανοφάνειας $L(\theta) = \mathrm{L}(\theta_1, \theta_2, ..., \theta_\mathrm{p})$. Ως Score function του δείγματος καλείται η τυχαία μεταβλητή:

$$U = \frac{\vartheta}{\vartheta\theta} l(\theta) = \frac{\vartheta}{\vartheta\theta} \ln L(\theta) = \sum_{i=1}^{n} \frac{\vartheta}{\vartheta\theta} \ln f(Y_i; \theta)$$

Κάτω από τις κατάλληλες συνθήκες ομαλότητας της L(θ) ισχύει ότι:

I. $E(U) = 0, \quad V(U) = -E\left(\frac{\vartheta^2 l}{\vartheta\theta^2} \right) \equiv Inf(\theta)$ (πίνακας πληροφορίας κατά Fisher)

II. $U \underset{n\to\infty}{\rightarrow} N(0, Inf(\theta))$

III. $\hat{\theta} \underset{n\to\infty}{\rightarrow} N\left(\theta, Inf(\theta)^{-1}\right)$, όπου $\hat{\theta}$ η εκτιμήτρια μέγιστης πιθανοφάνειας του θ.

IV. $-2\ln \frac{\sup_{\theta\in\Theta_0} \mathrm{L}(\theta)}{\sup_{\theta\in\Theta_1} \mathrm{L}(\theta)} \underset{n\to\infty}{\rightarrow} \chi^2_{\dim(\Theta_0)-\dim(\Theta_1)}$, όπου dim(Θ) είναι η διάσταση του χώρου Θ.

(De Jong & Heller, (2008)).

4.6 Poisson Παλινδρόμηση

Στην παλινδρόμηση με κατανομή Poisson οι αποκρίσεις είναι Poisson μετρήσεις που ακολουθούν ανεξάρτητες Poisson κατανομές με $E(y_i) = \mu_i$ και $Var(y_i) = \mu_i$. Μια σειρά από μεταβλητές $y_1, y_2, ..., y_n$ επηρεάζουν το μ μέσω του μοντέλου:

$$\mu_i = \exp\left(x_i^{'}\beta\right) \quad i = 1,2,3,...,n \tag{4.7.1}$$

με $x_i^{'} = \left[1, x_{i1}, x_{i2}, ..., x_{ik}\right]$. Ωστόσο, πρέπει να αναφέρουμε ότι όπως διάφορα μοντέλα αντί της λογιστικής είναι κατάλληλα για διωνυμικές αποκρίσεις, το ίδιο ισχύει και για την Poisson παλινδρόμηση. Για τη μεμειγμένη κατανομή Poisson παραθέτουμε επίσης τον παρακάτω τύπο:

$$N_i \sim Poi\left(d_i \exp\left(\beta_o + \sum_{j=1}^{p} \beta_j x_{ij}\right)\right), i = 1,2,...,n$$

όπου η τυχαία μεταβλητή ε_i αντιπροσωπεύει την επίδραση των καταλοίπων των κρυμμένων χαρακτηριστικών. Η συχνότητα των απαιτήσεων περιγράφεται από την τυχαία μεταβλητή $\lambda_i \Theta_i$ όπου $\Theta_i = \exp(\varepsilon_i)$.

Η μέθοδος της Μέγιστης Πιθανοφάνειας για τις αποκρίσεις $y_1, y_2, ..., y_n$ είναι η εξής:

$$\ln L(\beta; y) = \log \prod_{i=1}^{n} \left[\frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}\right]$$

$$= \sum_{i=1}^{n} \left[-\exp(x_i^{'}\beta) + y_i x_i^{'}\beta - \ln y_i!\right]$$

Παραγωγίζοντας την παραπάνω σχέση και θέτοντάς τη ίση με το μηδέν, έχουμε:

$$\frac{\vartheta L(\beta; y)}{\vartheta \beta} = 0$$

$$= \sum_{i=1}^{n} \left[y_i x_i - \exp(x_i^{'}\beta) x_i\right] = 0$$

ή όπως αλλιώς μπορεί να γραφτεί η παραπάνω εξίσωση: $\sum_{i=1}^{n} (y_i - \mu_i) x_i = 0$

Η σχέση αυτή με μορφή πινάκων παίρνει τη μορφή $X^{'}(y - \mu) = 0$. Το ίδιο ισχύει και για τη σταθμισμένη Poisson Παλινδρόμηση η οποία συσχετίζεται με αυτή της Λογιστικής. Για το μοντέλο της σχέσης (4.7.1), το τετράγωνο του αθροίσματος των σταθμισμένων υπολοίπων είναι:

$$S = \sum_{i=1}^{n}\left[\frac{(y_i - \mu_i)^2}{\sigma_i^2}\right]$$

ενώ η ελαχιστοποίηση του παραπάνω μας δίνει

$$\vartheta S / \vartheta \beta = 0 \quad 2\sum_{i=1}^{n}\left[y_i - \exp(x_i^{'}\beta)\right]x_i \quad \rightarrow X'(y-\mu) = 0$$

Οι εφαρμογές της Poisson παλινδρόμησης συμπεριλαμβάνουν προβλήματα όπου η απόκριση είναι θετική τιμή: 0,1,2,3,… . Οι ακέραιες τιμές περιγράφουν την επίδοση ενός βιολογικού ή κοινωνικού συστήματος ή την παραγωγική διαδικασία μιας βιομηχανικής εφαρμογής. Στην περίπτωση του μοντέλου αυτού η μέθοδος της Μέγιστης Πιθανοφάνειας δίνει ως εκτίμηση του μ_{i} στο κορεσμένο μοντέλο το y_{i} επειδή στο κορεσμένο μοντέλο υποθέτουμε ότι οι n παρατηρήσεις είναι ανεξάρτητες και ανεπηρέαστες από τις μεταβλητές παλινδρόμησης x_{i}. Έτσι στο κορεσμένο μοντέλο έχουμε:

$$\ln L(\mu) = -\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i} + \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i} \ln \mathrm{y_i} - \sum_{i=1}^{n} \mathrm{y_i}!$$

Για το Poisson μοντέλο, έχουμε:

$$\ln L(\beta) = -\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{\mu_i} + \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i} \ln \mathrm{\mu_i} - \sum_{i=1}^{n} \mathrm{y_i}!$$

Όπου το $\hat{\mu}_{\mathrm{i}}$ προέρχεται από τη Μέγιστη Πιθανοφάνεια. Η απόκλιση:

$$\begin{aligned} D(\beta) &= -2\ln\left[\mathrm{L}(\beta)/\mathrm{L}(\lambda)\right] \\ &= 2\left[-\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} - \mathrm{\mu_i}) + \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i}(\ln \mathrm{y_i} - \ln \mathrm{\mu_i})\right] \\ &= 2\left[-\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} - \mathrm{\mu_i}) + \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i}\left[\ln(\mathrm{y_i}/\mathrm{\mu_i})\right]\right] \end{aligned}$$

η οποία αν y_{i} είναι κοντά στο $\hat{\mu}_{\mathrm{i}}$ πλησιάζει το μηδέν. Επίσης, αν θέλουμε να απλοποιήσουμε τη σχέση της απόκλισης πρέπει να λάβουμε υπόψη τη σχέση αποτελέσματος X’(y – μ) = 0, η οποία μας δίνει ότι $\sum_{i=1}^{n}(y_i - \mu_{\mathrm{i}}) = 0$ και η τελική μορφή της απόκλισης είναι:

$$D(\beta) = 2\sum_{i=1}^{\mathrm{n}} y_i \ln\frac{y_i}{\mu_i}$$

(Denuit et al., (2007)).

4.7 Μοντέλο Παλινδρόμησης Αρνητικής Διωνυμικής

Έστω ότι $\Theta_1,...,\Theta_n$ είναι ανεξάρτητες Γάμμα κατανομές (Gamma(a,a)) όπου η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Θ_i δίνεται από τη σχέση:

$$f_\Theta(\theta) = \frac{1}{\Gamma(a)} a^a \theta^{a-1} \exp(-a\theta), \qquad \theta > 0$$

Σε αυτή την περίπτωση έχουμε $E[\Theta_i] = 1$ και $Var[\Theta_i] = 1/a$.

Δοθέντος των παρατηρήσιμων χαρακτηριστικών που συνοψίζονται και παρουσιάζονται στο διάνυσμα x_i και στις επιδράσεις των τυχαίων μεταβλητών Θ_i, ο ετήσιος αριθμός των απαιτήσεων που αξιώνονται από το i ασφαλιστήριο συμβόλαιο ακολουθεί την κατανομή Poisson. Με άλλα λόγια, το λ_i είναι η αναμενόμενη συχνότητα απαιτήσεων του ασφαλιστηρίου συμβολαίου i (που βασίζεται στο διάνυσμα x_i) και το θ αντιπροσωπεύει το επίπεδο κινδύνου για το συγκεκριμένο ασφαλιστήριο (αν θ<1 ο ασφαλισμένος είναι λιγότερο επικίνδυνος στο να προκαλέσει ατύχημα από τους άλλους οδηγούς που ανήκουν στο ίδιο επίπεδο κινδύνου).

Για την Αρνητική Διωνυμική κατανομή έχουμε:

$$\begin{aligned}\Pr[N_i = k_i \mid x_i] &= \int_0^\infty \Pr[N_i = k_i \mid x_i, \Theta_i = \theta] f_\Theta(\theta) d\theta \\ &= \binom{a + k_i + 1}{k_i} \left(\frac{\lambda_i}{a + \lambda_i}\right)^{k_i} \left(\frac{a}{a + \lambda_i}\right)^a\end{aligned}$$

Για την πιθανοφάνεια της, έχουμε:

$$L(\beta, a) = \prod_{i=1}^n \frac{\lambda_i^{k_i}}{k_i!} \left(\frac{a}{a + \lambda_i}\right)^a (a + \lambda_i)^{-k_i} \frac{\Gamma(a + k_i)}{\Gamma(a)}$$

Για να υπολογίσουμε τον εκτιμητή μέγιστης πιθανοφάνειας για τις παραμέτρους β και α, αρκεί να λύσουμε την:

$$\frac{\vartheta}{\vartheta\beta} L(\beta, a) = \sum_{i=1}^n \tilde{x}_i \left(k_i - \lambda_i \frac{a + k_i}{a + \lambda_i}\right) = 0$$

Σημειώνουμε ότι οι παραπάνω εξισώσεις είναι παρόμοιες με εκείνες στην περίπτωση της κατανομής Poisson, μόνο που τώρα αντί του λ_i έχουμε το $\lambda_i \dfrac{a + k_i}{a + \lambda_i}$.

Παρατήρηση 4.8.1. Στο σημείο αυτό θα δώσουμε μια εξήγηση για τον λόγο $\frac{a+k_i}{a+\lambda_i}$ που συμπεριλαμβάνεται στις εξισώσεις πιθανοφάνειας της Αρνητικής Διωνυμικής κατανομής. Η από κοινού συνάρτηση πυκνότητας πιθανότητας των τυχαίων μεταβλητών (N_i, Θ_i), ισούται με:

$$\exp(-\theta_i\lambda_i)\frac{(\theta_i\lambda_i)^{k_i}}{k_i!}\frac{1}{\Gamma(a)}a^a\theta^{a-1}\exp(-a\theta_i)$$
$$\propto \exp(-\theta_i\lambda_i)\theta_i^{k_i+a-1}\exp(-a\theta_i)$$

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Θ_i δοθέντος ότι $N_i = k_i$, είναι:

$$\frac{\exp(-\theta_i(a+\lambda_i))\theta_i^{a+k_i-1}}{\int_0^{+\infty}\exp(-\xi(a+\lambda_i))\xi^{a+k_i-1}d\xi}$$
$$= \exp(-\theta_i(a+\lambda_i))\theta_i^{a+k_i-1}\frac{(a+\lambda_i)^{a+k_i}}{\Gamma(a+k_i)}$$

έτσι ώστε το Θ_i δοθέντος ότι $N_i = k_i$ ακολουθεί κατανομή $Gam(a+k_i, a+\lambda_i)$. Συνεπώς, έχουμε:

$$E[\Theta_i \mid N_i = k_i] = \frac{a+k_i}{a+\lambda_i}$$

και οι εκτιμητές μέγιστης πιθανοφάνειας για το μοντέλο παλινδρόμησης της Αρνητικής Διωνυμικής κατανομής προκύπτουν από τη λύση της εξίσωσης:

$$\sum_{i=1}^{n} x_i(k_i - \lambda_i E[\Theta_i \mid N_i = k_i]) = 0$$

Συγκρίνοντας τα παραπάνω αποτελέσματα με τις εξισώσεις πιθανοφάνειας της κατανομής Poisson, καταλαβαίνουμε ότι ο εκτιμώμενος αναμενόμενος αριθμός απαιτήσεων λ_i αντικαθίσταται από τον $\lambda_i E[\Theta_i \mid N_i = k_i]$ που βασίζεται στην πληροφόρηση που περιέχεται στον αριθμό k_i των απαιτήσεων που καταβλήθηκαν για το ασφαλιστήριο συμβόλαιο i.

Μπορεί κανείς να επιλύσει τις εξισώσεις πιθανοφάνειας της Αρνητικής Διωνυμικής κατανομής με τη βοήθεια της διαδικασίας Newton – Raphson. Οι αρχικές τιμές για την διαδικασία Newton – Raphson συνήθως ανακτώνται ως εξής: Ο εκτιμητής μέγιστης πιθανοφάνειας της κατανομής Poisson $\hat{\beta}$ είναι γνωστό ότι είναι συνεπής εκτιμητής, οπότε το χρησιμοποιούμε αυτόν ως αρχική τιμή. Πρέπει όμως να βρούμε την αρχική εκτίμηση για το $\tau = \mathrm{V}[\Theta_i] = 1/a$, γι’ αυτό τον λόγο υπολογίζουμε πρώτα την διακύμανση των N_i η οποία δίνεται από τη σχέση:

$$V[N_i] = E[N_i] + (d_i \exp(score_i))^2 \tau$$

Έπειτα, γράφοντας την εμπειρική αναλογία της τελευταίας σχέσης, θα πάρουμε:

$$\sum_{i=1}^{n} \left((k_i - d_i \exp(score_i))^2 - k_i - (d_i \exp(score_i))^2 \tau \right) = 0$$

Συνεπώς, ο εκτιμητής του τ δίνεται από τη σχέση:

$$\frac{1}{\hat{a}} = \hat{\tau} = \frac{\sum_{i=1}^{n} \left(\left(k_i - d_i \exp\left(\hat{score}_i \right) \right)^2 - k_i \right)}{\sum_{i=1}^{n} \left(d_i \exp\left(\hat{score}_i \right) \right)^2}$$

όπου $\hat{score}_i = \tilde{x}_i^T \hat{\beta}$ και $\hat{\beta}$ ο εκτιμητής μέγιστης πιθανοφάνειας του β για την κατανομή Poisson. Οι εκτιμητές $\hat{\beta}$ και $\hat{\tau}$ είναι συνεπής για το μεμιγμένο μοντέλο Poisson και άρα αποτελούν κατάλληλες αρχικές τιμές για να βρεθούν εκτιμητές μέγιστης πιθανοφάνειας (Denuit et al., (2007)).

4.8 Μοντέλο Παλινδρόμησης για την κατανομή Poisson – LogNormal

Έστω ότι οι τυχαίες μεταβλητές $\Theta_1, \Theta_2.,,, \Theta_n$ είναι ανεξάρτητες και ακολουθούν την Λογαριθμοκανονική κατανομή. Στο σημείο αυτό και πριν προχωρήσουμε στην ανάλυση του παλινδρομικού μοντέλου της κατανομής αυτής, θα ήταν χρήσιμο να πούμε λίγα λόγια για την ίδια την Λογαριθμοκανονική κατανομή.

Γνωρίζουμε ότι αν μια τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή με μέσο μ και διακύμανση $\sigma^2, X \sim N(\mu, \sigma^2)$, τότε η συνάρτηση κατανομής της είναι η:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma} \right), \quad \acute{o}\pi o \upsilon \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-y^2/2) dy$$

Συνεπώς, η τυχαία μεταβλητή X ακολουθεί την λογαριθμοκανονική κατανομή με παραμέτρους μ και σ, , αν η τυχαία μεταβλητή lnX ακολουθεί την κανονική κατανομή με μέσο μ και διασπορά σ^2, με συνάρτηση πυκνότητας πιθανότητας:

$$f(x)=\frac{1}{\sqrt{2\pi}x\sigma}\exp\left(-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right), x>0$$

Αν $X \sim LN(\mu,\sigma^2)$ τότε ο μέσος και η διακύμανση δίνονται από τις σχέσεις:

$$E[X]=\exp\left(\mu+\frac{\sigma^2}{2}\right),\quad V[X]=\exp(2\mu+\sigma^2)(\exp(\sigma^2)-1)$$

Επανερχόμαστε τώρα στην περίπτωση της Poisson Λογαριθμοκανονικής κατανομής, όπου η συνάρτηση πυκνότητας πιθανότητας των Θ_i δίνεται από τη σχέση:

$$f_\Theta(\theta)=\frac{1}{\theta\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln\theta+\sigma^2/2)^2}{2\sigma^2}\right), \theta>0$$

Ενώ η συνάρτηση κατανομής δίνεται από τη σχέση:

$$\Pr[N=k]=\frac{1}{\sigma\sqrt{2\pi}}\frac{(\lambda d)^k}{k!}\int_0^\infty \exp(-\lambda d\theta)\theta^{k-1}\exp\left(-\frac{(\ln\theta+\sigma^2/2)^2}{2\sigma^2}\right)d\theta$$

Συνυπολογίζοντας τα παραπάνω καταλήγουμε στο ότι $\mathrm{E}[\Theta_i]=1$ και $V[\Theta_i]=\exp(\sigma^2)-1$ $\left(E[N]=\lambda, V[N]=\lambda+\lambda^2(\exp(\sigma^2)-1)\right)$. Επίσης μπορεί να δειχθεί ότι η ασυμμετρία αυτής της κατανομής προκύπτει πάλι από τον τύπο:

$$\gamma[N]=\frac{1}{(V[N])^{3/2}}\left(3V[N]-2E[N]+\frac{\gamma[\Theta]}{\sqrt{V[\Theta]}}\frac{(V[N]-E[N])^2}{E[N]}\right)$$

όπου $\gamma[\Theta]/\sqrt{\mathrm{V}[\Theta]}=2+\exp(\sigma^2)$. Συνεπώς, η ασυμμετρία της Poisson Λογαριθμοκανονικής κατανομής υπερβαίνει αυτής της Poisson – Inverse Gaussian κατανομής έχοντας τον ίδιο μέσο και την ίδια διασπορά.

Η πιθανότητα το ασφαλιστήριο συμβόλαιο i να απαιτήσει k_i απαιτήσεις υπολογίζεται από τον τύπο:

$$\Pr[N_i=k_i \mid x_i]=\int_0^\infty \exp(-\theta\lambda_i)\frac{(\theta\lambda_i)^{k_i}}{k_i!}\frac{1}{\theta\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln\theta+\sigma^2/2)^2}{2\sigma^2}\right)d\theta$$

όπου μπορούμε να αντλήσουμε τους εκτιμητές μέγιστης πιθανοφάνειας που θα μας βοηθήσουν στο μοντέλο παλινδρόμησης (Denuit et al., (2007)).

4.9 Μοντέλα Παλινδρόμησης για Ζημιοκατανομές

Το μοναδικό μοντέλο παλινδρόμησης που θα περιγραφεί στην παρούσα εργασία είναι εκείνο της κατανομής Γάμμα.

4.10 Μοντέλο Παλινδρόμησης για την Γάμμα Κατανομή

Η κατανομή αυτή έχει εφαρμογή σε προβλήματα παλινδρόμησης όπου η απόκριση είναι συνεχής και η διακύμανση δεν είναι σταθερή αλλά ανάλογη του τετράγωνου του μέσου μ. Μια λύση στο πρόβλημα αυτό είναι να χρησιμοποιήσουμε λογαριθμικό μετασχηματισμό για να σταθεροποιηθεί η διακύμανση όπου όλοι οι συντελεστές είναι αμερόληπτοι και η απόκριση είναι λογαριθμοκανονική. Η τομή είναι μεροληπτική κατά $(\sigma/\mu)^2/2$ αφού από την ανάλυση κατά Taylor έχουμε (Denuit, M. (2002)):

$$E[\ln y] = \ln\mu - (\sigma/\mu)^2/2$$

Η συνάρτηση πυκνότητας πιθανότητας της κατανομής είναι:

$$f(y) = \frac{1}{\Gamma(r)}\left(\frac{1}{\lambda}\right)^r \exp(-y/\lambda)y^{r-1}, r>0, \lambda>0$$

η οποία βάσει της συνάρτησης της εκθετικής οικογένειας κατανομών, δίνει:

$$\theta = -\frac{1}{\lambda r} = -\frac{1}{\mu},\ \mu = r\lambda,\ \operatorname{var} y = \frac{\mu^2}{r} \to \frac{\operatorname{var} y}{\mu^2} = r\lambda^2,\ \alpha(\phi) = r^{-1},\ b(\theta) = -\ln(-\theta)$$

$$c(\phi) = r\ln r - \ln\Gamma(r) + (r-1)\ln y$$

όπου r είναι η παράμετρος κλίμακας. Όταν έχουμε r= 1 τότε καταλήγουμε στην Εκθετική κατανομή. Στην περίπτωση όπου το r δεν ποικίλει αλλά είναι σταθερό κατά την ανάλυση των δεδομένων, τότε μέσω της παραμέτρου λ έχουμε αλλαγή του μέσου.

Το Γενικευμένο Γραμμικό Μοντέλο με μεταβλητή απόκρισης να ακολουθεί την Γάμμα κατανομή, περιγράφεται από τη σχέση:

$$y \sim G(\mu,\nu), \quad g(\mu) = x'\beta$$

Έχει αποδειχθεί ότι η σχέση μεταξύ του μεγέθους των λογαριθμισμένων απαιτήσεων και της χρονικής ισχύς του ασφαλιστηρίου συμβολαίου, είναι γραμμική (De Jong & Heller, (2008)). Συνεπώς, το γεγονός αυτό μας οδηγεί στο συμπέρασμα ότι πρέπει να χρησιμοποιηθεί ένα μοντέλο που αποτιμά το μέγεθος των απαιτήσεων με λογαριθμική συνάρτηση σύνδεσης και ανεξάρτητες μεταβλητές την χρονική διάρκεια του ασφαλιστηρίου συμβολαίου και την νομική προστασία. Καθώς το μέγεθος των απαιτήσεων είναι συνεχής τυχαία μεταβλητή και η κατανομή τους έχει αρνητική ασυμμετρία, η Γάμμα κατανομή θεωρείται ως μια πολύ καλή λύση για την μεταβλητή απόκρισης.

Σύμφωνα με τα παραπάνω, το μοντέλο γράφεται τώρα ως εξής:

$$y \sim G(\mu, \nu), \quad \ln\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

όπου η μεταβλητή x_1 περιγράφει τον χρόνο της ασφαλιστικής κάλυψης, η μεταβλητή x_2 είναι μια δείκτρια τυχαία μεταβλητή για το αν έχει νομική προστασία ή όχι το ασφαλιστήριο συμβόλαιο και τέλος, η μεταβλητή x_3 παριστάνει την αλληλεπίδραση των δυο τυχαίων μεταβλητών.

Στην μελέτη των De Jong & Heller αποδείχθηκε ότι οι απαιτήσεις που προέρχονται από ασφαλιστήρια συμβόλαια με μικρή χρονική διάρκεια, έχουν μέση τιμή απαιτήσεων χαμηλότερη από εκείνη που προβλέπεται από το μοντέλο. Για τα ασφαλιστήρια συμβόλαια που δεν έχουν την κάλυψη της νομικής προστασίας, παρουσιάζεται μια συστηματική μείωση του μεγέθους των καταλοίπων για την χρονική ισχύ του συμβολαίου η οποία πλησιάζει το μηδέν.

Το προτεινόμενο μοντέλο, είναι το:

$$y_{ij} \sim G(\mu_{ij}, \nu), \quad \ln\mu_{ij} = a + \gamma_i + \delta_i, i = 1,...,n, \quad j = 0,...,n-1$$

Ο χρόνος του ατυχήματος είναι ο i = 1 και ο χρόνος αξιοποίησης της αποζημίωσης είναι ο j = 0. Ο συνδυασμός των δύο παραπάνω αποτελεί τη βασική υπόθεση για το μοντέλο.

ΚΕΦΑΛΑΙΟ 5

ΤΑΞΙΝΟΜΗΣΗ ΑΣΦΑΛΙΣΜΕΝΩΝ ΜΕ ΕΚ ΤΩΝ ΠΡΟΤΕΡΩΝ ΚΡΙΤΗΡΙΑ

5 Ταξινόμηση Κινδύνων και Μοντέλα Παλινδρόμησης

Η τιμολόγηση για το ασφάλιστρο του αυτοκινήτου είναι μια πολύ σημαντική διαδικασία και γι αυτό το λόγο θα πρέπει να ταξινομηθούν και να εφαρμοστούν διάφορες πολιτικές σύμφωνα με τα χαρακτηριστικά του κινδύνου που υπόκεινται το αυτοκίνητο. Οι μεταβλητές που θα χρησιμοποιηθούν για την ταξινόμηση αυτή καλούνται ως εκ των προτέρων μεταβλητές (διότι οι τιμές τους μπορούν να καθοριστούν πριν την ημερομηνία απόκτησης διπλώματος του ασφαλισμένου). Στην ασφάλιση του αυτοκινήτου, οι μεταβλητές αυτές περιλαμβάνουν την ηλικία, το φύλο και το επάγγελμα του ασφαλισμένου, τον τύπο και την χρήση του αυτοκινήτου, τον τόπο που διαμένει και μερικές φορές τον αριθμό των αυτοκινήτων του νοικοκυριού του ασφαλισμένου, την οικογενειακή κατάσταση ή το χρώμα του οχήματος.

Όλα αυτά τα παρατηρήσιμα χαρακτηριστικά κινδύνων μπορούν να θεωρηθούν τυπικά ως μη τυχαίοι συντελεστές κινδύνου. Άλλα χαρακτηριστικά κινδύνων είναι μη παρατηρήσημα και θα πρέπει να θεωρηθούν ως άγνωστοι παράμετροι ή, στα πλαίσια της πιστοληπτικής θεωρίας, latent μεταβλητές με μια κοινή κατανομή. Μέχρι στιγμής η βιβλιογραφία που αναφέρεται σε τιμολόγηση ασφαλίστρων για το αυτοκίνητο επικεντρώνεται σε δυο βασικές προσεγγίσεις: α) η πρώτη, αγνοεί τους παρατηρήσημους συντελεστές και μετατρέπει όλα τα ατομικά χαρακτηριστικά κινδύνου μετατρέποντάς τα σε τυχαίες μεταβλητές, β) η δεύτερη, αγνοεί τελείως τα τυχαία ατομικά χαρακτηριστικά κινδύνου και προσπαθεί αντίθετα να συλλέξει όλες τις ατομικές διακυμάνσεις του κινδύνου από τους ίδιους τους συντελεστές κινδύνου.

Αν τα δεδομένα που εξετάζονται έχουν υποδιαιρεθεί σε κλάσεις κινδύνου από εκ των προτέρων μεταβλητές, τότε οι αναλογιστές εργάζονται με μεγέθη που είναι σχετικά μικρά σε σύγκριση με τα μεγέθη των απαιτήσεων και έκθεσης στον κίνδυνο. Για αυτό τον λόγο χρησιμοποιούνται περιγραφικά στατιστικά και μοντέλα παλινδρόμησης. Η παλινδρόμηση, αναλύει τη σχέση μεταξύ μιας μεταβλητής (εξαρτημένη μεταβλητή ή μεταβλητή απόκρισης) με ένα σύνολο άλλων μεταβλητών (ανεξάρτητες μεταβλητές). Αυτή η σχέση εκφράζεται με μια εξίσωση η οποία προβλέπει μια μεταβλητή απόκρισης (π.χ., τον αναμενόμενο αριθμό των απαιτήσεων που δόθηκαν σε έναν συγκεκριμένο ασφαλισμένο) από μια συνάρτηση επεξηγηματικών μεταβλητών και παραμέτρων (συμπεριλαμβάνοντας έναν γραμμικό συνδυασμό των επεξηγηματικών μεταβλητών αυτών και παραμέτρων, ο οποίος καλείται ως γραμμικός εκτιμητής). Οι παράμετροι εκτιμούνται με τέτοιο τρόπο ώστε να βελτιστοποιείται ένα μέτρο καλής προσαρμογής (όπως είναι για παράδειγμα η λογιστική πιθανοφάνεια στις περισσότερες των περιπτώσεων). Οι αναλογιστές χρησιμοποιούν τεχνικές παλινδρόμησης για να προβλέψουν τον αναμενόμενο αριθμό των απαιτήσεων, έχοντας κάποιες ενδείξεις - πληροφορίες για την οδική συμπεριφορά των ασφαλισμένων, των οχημάτων και των διάφορων τύπων ασφαλιστηρίων συμβολαίων. Αξίζει να σημειωθεί όμως ότι ακόμα και αν όλες οι μεταβλητές έχουν συμπεριληφθεί στην αποτίμηση των συμβολαίων αυτών, υπάρχουν ακόμη και άλλες παράμετροι κινδύνων που ενδεχομένως να μην μπορούν να προβλεφθούν (όπως είναι η νευρικότητα του οδηγού, η ικανότητα του, η γνώση του κώδικα οδικής κυκλοφορίας κα).

Στην ασφάλιση αστικής ευθύνης, το καθαρό ασφάλιστρο υπολογίζεται ως το αναμενόμενο κόστος όλων των απαιτήσεων που θα απαιτήσουν οι ασφαλισμένοι κατά τη διάρκεια την ασφαλιστικής κάλυψης (υπό την υπόθεση του Νόμου των Μεγάλων Αριθμών). Ο υπολογισμός αυτού του ασφαλίστρου βασίζεται σε ένα στατιστικό μοντέλο το οποίο συμπεριλαμβάνει όλες τις διαθέσιμες πληροφορίες για τον ασφαλιστικό κίνδυνο. Η τεχνική αυτή του υπολογισμού στοχεύει στο να εκτιμήσει όσο πιο καλά μπορεί (με περισσότερη ακρίβεια) το καθαρό ασφάλιστρο για κάθε ασφαλισμένο μέσω μεθόδων παλινδρόμησης. Είναι γνωστό στην ασφαλιστική αγορά ότι τα ασφάλιστρα που υπολογίζονται ως ενδεικτικά για κάθε ασφαλιστικό κίνδυνο, διαφέρουν με τα ασφάλιστρα που υπολογίζονται από τους αναλογιστές για την εταιρία που εργάζονται για τους ίδιους ασφαλιστικούς κινδύνους.

Ορισμένες φορές, η τιμολόγηση των ασφαλίστρων για το αυτοκίνητο βασίζεται σε δεδομένα πολλών ετών. Με τον τρόπο αυτό, συνυπολογίζοντας δηλαδή δεδομένα αρκετών ασφαλιστικών περιόδων, προκύπτουν κάποια πλεονεκτήματα: αυξάνεται το μέγεθος του δείγματος και αποφεύγεται να δίνεται περισσότερη σημασία σε μια και μόνο ασφαλιστική περίοδο (κατά την οποία ενδεχομένως λόγω των καιρικών συνθηκών να υπήρχε αύξηση ή μείωση των ατυχημάτων). Ωστόσο, το γεγονός αυτό μπορεί να επιφέρει κάποια εξάρτηση στα δεδομένα, αφού ο αριθμός των ζημιών που σχετίζονται με έναν ασφαλισμένο κατά τη διάρκεια του χρόνου αναμένεται να συσχετίζεται. Η ανάλυση αυτή των συσχετισμένων δεδομένων με την χρήση της διαδικασίας Poisson που προκύπτουν από επαναλαμβανόμενες μετρήσεις, μπορεί να γίνει με τη βοήθεια των Γενικευμένων Συναρτήσεων Εκτίμησης (Generalized Estimating Equations (GEEs)). Οι συναρτήσεις αυτές παρέχουν μια πρακτική μέθοδο με στατιστικά σημαντική αποδοτικότητα για να αναλύσουν αυτά τα διαχρονικά δεδομένα.

Επίσης, οι συναρτήσεις αυτές δίνουν τη δυνατότητα να χρησιμοποιηθούν διαδικασίες μέγιστης πιθανοφάνειας πάνω σε μοντέλα για μακροχρόνια δεδομένα (Denuit et al, (2007)).

5.1 Επιμερισμός Κινδύνων σε Τμήματα

Έστω λοιπόν ότι έχουμε ένα χαρτοφυλάκιο με n ασφαλιστήρια συμβόλαια που αφορούν την κάλυψη αστικής ευθύνης αυτοκινήτου έναντι τρίτων. Η τυχαία μεταβλητή Y_i (που είναι και η εξαρτημένη μεταβλητή) περιγράφει το σύνολο των απαιτήσεων που θα απαιτηθούν σε μια περίοδο για το ασφαλιστήριο συμβόλαιο i. Για να ερμηνευτούν λοιπόν τα αποτελέσματα της μεταβλητής Y_i, ο αναλογιστής θα χρησιμοποιήσει ένα διάνυσμα παρατηρήσιμων μεταβλητών $X_i^T = (X_{i1}, X_{i2},...)$ που έχει στη διάθεσή του (π.χ., την ηλικία, το φύλο και το επάγγελμα του κατόχου του ασφαλιστηρίου συμβολαίου i, την περιοχή κατοικίας του, τον τύπο και την χρήση του αυτοκινήτου του). Ωστόσο, η μεταβλητή Y_i εξαρτάται επίσης από μια ακολουθία αγνώστων χαρακτηριστικών $Z_i^T = (Z_{i1}, Z_{i2},...)$ όπως είναι τα αντανακλαστικά του οδηγού, η επιθετικότητα πίσω από το τιμόνι, η τάση προς την κατανάλωση αλκοόλ κλπ. Ορισμένες από αυτές τις ποσότητες δεν είναι μετρήσιμες και κάποιες άλλες δεν μπορούν να τιμολογηθούν διότι η αποτίμηση τους θεωρείται δαπανηρή.

Το αληθινό ασφάλιστρο για το ασφαλιστήριο συμβόλαιο i δίδεται από τη σχέση $E[Y_i \mid X_i, Z_i]$. Είναι μια σύνθετη συνάρτηση των X_i και Z_i που βρίσκονται πιο κοντά στο Y_i, τέτοια ώστε το $E\left[(Y_i - g(X_i, Z_i))^2\right]$ να είναι ελάχιστο για $g(X_i, Z_i) = E[Y_i \mid X_i, Z_i]$. Αν ο ασφαλιστής χρεώνει με ασφάλιστρο $E[Y_i \mid X_i, Z_i]$ το ασφαλιστήριο συμβόλαιο i, τότε ο ασφαλισμένος πληρώνει ουσιαστικά το ασφάλιστρο για να απορροφηθούν οι διακυμάνσεις στις τιμές των ασφαλίστρων που επηρεάζονται από τα προσωπικά χαρακτηριστικά του οδηγού, όπως εκείνα οροθετήθηκαν στις μεταβλητές X_i και Z_i, η σημαντικότητα των οποίων περιγράφεται από την σχέση $V[E[Y_i \mid X_i, Z_i]]$. Ο επιμερισμός του κινδύνου μπορεί να συνοψισθεί ως εξής: χρησιμοποιώντας τον τύπο

$$V[Y_i] = \underbrace{E[V[Y_i \mid X_i, Z_i]]}_{\to ασφαλιστής} + \underbrace{V[E[Y_i \mid X_i, Z_i]]}_{\to ασφαλισμένος}$$

Επίσης, εφόσον τα στοιχεία της μεταβλητής Z_i είναι άγνωστα για τον ασφαλιστή, τότε ο παραπάνω τύπος είναι καθαρά θεωρητικός. Εφόσον λοιπόν, η ασφαλιστική εταιρία γνωρίζει μόνο τα στοιχεία της μεταβλητής X_i, τότε ο ασφαλιστής μπορεί να χρεώνει μόνο ασφάλιστρο το οποίο προκύπτει από την $E[Y_i \mid X_i]$. Ο επιμερισμός του κινδύνου σε αυτή την περίπτωση περιγράφεται από τον τύπο:

$$V[Y_i] = \underbrace{E[V[Y_i \mid X_i]]}_{\rightarrow ασφαλιστής} + \underbrace{V[E[Y_i \mid X_i]]}_{\rightarrow ασφαλισμένος}$$

Στον παραπάνω τύπο η διακύμανση του ασφαλιστή είναι μεγαλύτερη, γεγονός που οφείλεται στην ετερογένεια των καταλοίπων που απορρέουν από τις απαιτήσεις της εταιρίας. Για να γίνουμε πιο συγκεκριμένοι, αρκεί να γράψουμε:

$$E[V[Y_i \mid X_i]] = E[V[Y_i \mid X_i, Z_i]] + E[V[E[Y_i \mid X_i, Z_i] \mid X_i]]$$

Ο όρος $E[V[Y_i \mid X_i, Z_i]]$ αντιπροσωπεύει τις καθαρές τυχαίες διακυμάνσεις του κινδύνου και βασίζεται στην βασική αρχή την ασφαλιστικής κάλυψης που ακολουθούν όλες οι ασφαλιστικές εταιρίες. Αντίθετα, ο δεύτερος όρος $E[V[E[Y_i \mid X_i, Z_i] \mid X_i]]$, αντιπροσωπεύει την διακύμανση των αναμενόμενων απαιτήσεων που οφείλονται στα άγνωστα χαρακτηριστικά την τυχαίας μεταβλητής Z_i. Γι' αυτό τον λόγο, η παραπάνω ποσότητα θα πρέπει να διορθωθεί από έναν εμπειρικό μηχανισμό αξιολόγησης των κινδύνων .

Σε αυτό το σημείο πρέπει να δούμε την σύνδεση που υπάρχει ανάμεσα στην a priori και την a posteriori εκτίμηση και αξιολόγηση των κινδύνων. Οι παρελθοντικές απαιτήσεις αποκαλύπτουν τα κριμένα χαρακτηριστικά της τυχαίας μεταβλητής Z_i. Έστω λοιπόν ότι η τυχαία μεταβλητή $Y_i^{\leftarrow}$ υποδηλώνει τις απαιτήσεις που αξιώθηκαν στο παρελθόν για την τυχαία μεταβλητή Y_i . Η πληροφορίες που περιέχονται στο διάστημα $(X_i, Y_i^{\leftarrow})$ είναι συγκρίσιμες με τις πληροφορίες που περιέχονται στο διάστημα (X_i, Z_i) καθώς περνάει ο χρόνος. Γι' αυτό το λόγο το a posteriori ασφάλιστρο είναι το $E[Y_i \mid X_i, Y_i^{\leftarrow}]$. Ο μηχανισμός τιμολόγησης των ασφαλίστρων που βασίζεται ανάλογα με το πόσες φορές παρουσιαστεί αξίωση για αποζημίωση (δηλαδή αν ο ασφαλισμένος δεν υποστεί ή προκαλέσει κάποιο ατύχημα μέσα σε μια δεδομένη χρονική περίοδο τότε θα έχει μια ευνοϊκή τιμολόγηση για το ασφάλιστρο του, τα ονομαζόμενα bonuses, ενώ στην αντίθετη περίπτωση θα έχει επιβαρυμένο ασφάλιστρο, τα ονομαζόμενα maluses).

Στην a priori τιμολόγηση, ο αναλογιστής στοχεύει στο να αναγνωρίσει τα χαρακτηριστικά της τυχαίας μεταβλητής X_i και να υπολογίσει το ασφάλιστρο $E[Y_i \mid X_i]$. Στην a posteriori τιμολόγηση, ο αναλογιστής υπολογίζει το ασφάλιστρο σύμφωνα με τα στοιχεία την τυχαίας μεταβλητής $Y_i^{\leftarrow}$ για να μπορέσει να καλύψει τις μη διαθέσιμες πληροφορίες που περιέχονται στην τυχαία μεταβλητή Z_i (Denuit et al., (2007)).

ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΓΕΝΙΚΕΥΜΕΝΑ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

5.1.1 Εισαγωγή

Εάν επιθυμούμε να μελετήσουμε πίνακες συνάφειας με περισσότερες από δύο διαστάσεις, τότε ο υπολογισμός των εκτιμητών μέγιστης πιθανοφάνειας δεν είναι τόσο απλός και πρέπει να βασιστεί στην γενική θεωρία των γενικευμένων γραμμικών μοντέλων. Επιπλέον η θεωρία των γενικευμένων γραμμικών μοντέλων προσφέρει ένα θεωρητικό πλαίσιο για την ενιαία μελέτη των πινάκων συνάφειας όλων των διαστάσεων. Παρακάτω θα δούμε πως μπορεί να ενταχθεί η μελέτη των πινάκων συνάφειας στα γενικευμένα γραμμικά μοντέλα ξεκινώντας από την απλούστερη περίπτωση των διδιάστατων πινάκων συνάφειας.

5.1.2 Μελέτη διδιάστατων πινάκων χρησιμοποιώντας γενικευμένα γραμμικά μοντέλα

Έστω *X, Y* δύο κατηγορικές μεταβλητές απόκρισης με *I* και *J* στάθμες αντίστοιχα. Οι αποκρίσεις (*X,Y*) ενός τυχαία επιλεγμένου ατόμου από έναν πληθυσμό θεωρούνται ως τυχαίες μεταβλητές που ακολουθούν κάποια κατανομή. Συμβολίζουμε με *πij* την πιθανότητα

$$\pi_{ij} = \Pr((X,Y) = (i,j)) = \Pr(X = i, Y = J), i = 1,\dots, I \quad j = 1,\dots, J$$

Επίσης, θα συμβολίσουμε με $\pi_{i\bullet} = P(X = i)$, $\pi_{\bullet j} = P(Y = j)$, τις περιθώριες κατανομές των X, Y. Είναι προφανές ότι

$$\pi_{i\bullet} = \sum_{j=1}^{J} \pi_{ij}, i = 1,2,\dots, I \qquad \kappa\alpha\iota \quad \pi_{\bullet j} = \sum_{i=1}^{I} \pi_{ij}, \quad j = 1,2,\dots, J$$

Μπορούμε να παραστήσουμε τη διδιάστατη κατανομή των τυχαίων μεταβλητών (X,Y) χρησιμοποιώντας ένα πίνακα με I γραμμές και J στήλες. Πιο συγκεκριμένα, έχουμε τον ακόλουθο πίνακα:

Y

		1	2	…	J	Σύνολο
	1	π_{11}	π_{12}	…	π_{1J}	$\pi_{1\bullet}$
	2	π_{21}	π_{22}	…	π_{2J}	$\pi_{2\bullet}$
X	…	…	…	…	…	…
	I	π_{I1}	π_{I2}	…	π_{IJ}	$\pi_{I\bullet}$
	Σύνολο	$\pi_{\bullet 1}$	$\pi_{\bullet 2}$		$\pi_{\bullet J}$	$\pi_{\bullet\bullet}=1$

Έστω τώρα ότι έχουμε *N* τυχαία επιλεγμένα άτομα (*N* είναι είτε σταθερά είτε τυχαία μεταβλητή) από τον πληθυσμό και N_{ij} το πλήθος των ατόμων (από τα *N*) τα οποία ταξινομούνται στο κελί (*i*, *j*) (κάθε άτομο ταξινομείται στην θέση (*i*, *j*) του πίνακα συνάφειας με πιθανότητα π_{ij}). Ο *I*×*J* πίνακας ο οποίος περιέχει τις παρατηρούμενες συχνότητες N_{ij} καλείται πίνακας συνάφειας. Όμοια με παραπάνω συμβολίζουμε με $N_{i\bullet} N_{\bullet j}$, τα αθροίσματα των γραμμών και των στηλών αντίστοιχα, δηλαδή,

$$N_{i\bullet} = \sum_{j=1}^{J} N_{ij}, i = 1,2,..., I \qquad \kappa\alpha\iota \quad N_{\bullet j} = \sum_{i=1}^{I} N_{ij}, \quad j = 1,2,..., J$$

και θα πάρουμε τον παρακάτω πίνακα:

Y

		1	2	…	J	Σύνολο
	1	N_{11}	N_{12}	…	N_{1J}	$N_{1\bullet}$
	2	N_{21}	N_{22}	…	N_{2J}	$N_{2\bullet}$
X	…	…	…	…	…	…
	I	N_{I1}	N_{I2}	…	N_{IJ}	$N_{I\bullet}$
	Σύνολο	$N_{\bullet 1}$	$N_{\bullet 2}$		$N_{\bullet J}$	$N_{\bullet\bullet} = N$

Σκοπός μας είναι η μελέτη διαφόρων χαρακτηριστικών της διδιάστατης κατανομής π_{ij} με βάση τις παρατηρούμενες συχνότητες N_{ij}. Η δειγματική διδιάστατη κατανομή θα συμβολίζεται με p_{ij} και θα ισχύει ότι $p_{ij} = N_{ij} / N$.

Στους διδιάστατους πίνακες μπορούμε να θεωρήσουμε το εξής μοντέλο:

1) *Μοντέλο Poisson:* Ισχύει ότι $E(N_{ij}) = m_{ij} = m_{\bullet\bullet}\pi_{\mathrm{ij}}$ και η υπόθεση της ανεξαρτησίας (γραμμών και στηλών) μεταφράζεται ως εξής: $E(N_{ij}) = m_{\bullet\bullet}\pi_{\mathrm{i}\bullet}\pi_{\bullet j}$ (βλ, Cox, (1970)).

5.1.3 Παραμετροποίηση του μοντέλου

Παρατηρούμε ότι οι συνήθεις υποθέσεις (ανεξαρτησία, ίδια κατανομή στις γραμμές κλπ) μπορούν να εκφραστούν ως πολλαπλασιαστικά μοντέλα στα οποία οι αναμενόμενες συχνότητες είναι ίσες με γινόμενα περιθωρίων πιθανοτήτων και προκαθορισμένων αθροισμάτων συχνοτήτων. Συνεπώς, κάτω από αυτές τις υποθέσεις ο λογάριθμος των αναμενόμενων συχνοτήτων μπορεί να γραφεί στη μορφή:

$$\eta_{ij} = \log m_{ij} = \log E(N_{ij}) = x_i^T \beta$$

για κατάλληλες παραμέτρους β. Μια αρκετά βολική παραμετροποίηση του μοντέλου είναι η ακόλουθη:

$$\mu = \bar{\eta}_{\bullet\bullet}, \quad \lambda_{\mathrm{i}}^{\mathrm{X}} = \bar{\eta}_{i\bullet} - \bar{\eta}_{\bullet\bullet}, \quad \lambda_{\mathrm{j}}^{\mathrm{Y}} = \bar{\eta}_{\bullet j} - \bar{\eta}_{\bullet\bullet}, \quad \lambda_{\mathrm{ij}}^{\mathrm{XY}} = \eta_{ij} - \bar{\eta}_{i\bullet} - \bar{\eta}_{\bullet j} + \bar{\eta}_{\bullet\bullet}$$

όπου $\bar{\eta}_{i\bullet} = \frac{1}{J}\sum_{j=1}^{J}\eta_{ij}, \quad \bar{\eta}_{\bullet j} = \frac{1}{I}\sum_{i=1}^{I}\eta_{ij}, \quad \bar{\eta}_{\bullet\bullet} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\eta_{ij},$ και επομένως

$$\eta_{\mathrm{ij}} = \log m_{ij} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Y}} + \lambda_{\mathrm{ij}}^{\mathrm{XY}}$$

Το μοντέλο αυτό περιγράφει επακριβώς οποιοδήποτε σύνολο από αναμενόμενες συχνότητες (πλήρες ή κορεσμένο μοντέλο, saturated model). Οι παράμετροι $\lambda_{\mathrm{i}}^{\mathrm{X}}, \lambda_{\mathrm{j}}^{\mathrm{Y}}$ θεωρούνται αποκλίσεις από το γενικό μέσο μ, και είναι εύκολο να δούμε ότι αθροίζουν στο 0. Επομένως υπάρχουν I – 1 ανεξάρτητες παράμετροι γραμμών και J – 1 παράμετροι στηλών. Αντίστοιχα, οι αλληλεπιδράσεις $\lambda_{\mathrm{ij}}^{\mathrm{XY}}$ ικανοποιούν την

$$\sum_i \lambda_{\mathrm{ij}}^{\mathrm{XY}} = \sum_j \lambda_{\mathrm{ij}}^{\mathrm{XY}} = 0$$

Επομένως (I – 1)(J – 1) από αυτές τις παραμέτρους είναι γραμμικά ανεξάρτητες (βλ, Cox, (1970)).

5.1.4 Έλεγχοι ανεξαρτησίας

Η παραπάνω παραμετροποίηση είναι αρκετά βολική στην περίπτωση που θέλουμε να κάνουμε έλεγχο της ανεξαρτησίας των X, Y. Για παράδειγμα, στο πολυωνυμικό μοντέλο η υπόθεση της ανεξαρτησίας μπορεί να εκφραστεί ως εξής:

$$\eta_{ij} = \log E(N_{ij}) = \log n + \log \pi_{i\bullet} + \log \pi_{\bullet j} = \mu + \lambda_i^X + \lambda_j^Y$$

διότι αν $\eta_{ij} = \log n + \log \pi_{i\bullet} + \log \pi_{\bullet j}$ τότε είναι εύκολο να δούμε ότι

$$\lambda_i^X = \log \pi_{i\bullet} - \frac{1}{I}\sum_h \log \pi_{h\bullet}, \quad \lambda_j^Y = \log \pi_{\bullet j} - \frac{1}{J}\sum_h \log \pi_{\bullet h}, \quad \mu = \log n + \frac{1}{I}\sum_h \log \pi_{h\bullet} + \frac{1}{J}\sum_h \log \pi_{\bullet h},$$

Επομένως, το μοντέλο ανεξαρτησίας των X, Y είναι ειδική περίπτωση του παραπάνω πλήρους μοντέλου θεωρώντας ότι $\lambda_{ij}^{XY} = 0$ (οι παράμετροι λ_{ij}^{XY} αντικατοπτρίζουν απόκλιση από την ανεξαρτησία). Το πλήρες μοντέλο έχει 1+(I-1)+(J-1)+(I-1)(J-1) = I J ανεξάρτητες παραμέτρους, ενώ το μοντέλο ανεξαρτησίας έχει 1+(I-1)+(J-1) = I+J-1 ανεξάρτητες παραμέτρους. Ο έλεγχος ανεξαρτησίας γίνεται εξετάζοντας αν το μοντέλο $\eta_{ij} = \mu + \lambda_i^X + \lambda_j^Y$ είναι αποδεκτό (βλ, Cox, (1970)).

5.1.5 Μοντέλο κατανομής συχνοτήτων κελιών

Ας θεωρήσουμε τώρα το απλούστερο μοντέλο για την κατανομή των συχνοτήτων των κελιών που είναι το μοντέλο Poisson (οι συχνότητες είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με μέση τιμή m_{ij}). Η από κοινού συνάρτηση πιθανότητας των συχνοτήτων N_{ij} των κελιών, είναι της μορφής:

$$f(n;m) = \prod_{i=1}^{I}\prod_{j=1}^{J} e^{-m_{ij}} \frac{m_{ij}^{n_{ij}}}{n_{ij}!}$$

όπου οι μέσες τιμές m_{ij} εξαρτώνται από κάποιες παραμέτρους **β** μέσω μιας σχέσης της μορφής

$$\ln m_{ij} = \ln E(N_{ij}) = x_i^T \beta$$

ανάλογα με το μοντέλο. Παρατηρούμε ότι η παραπάνω κατανομή f(**n;m**) ανήκει στην εκθετική οικογένεια κατανομών και μάλιστα τα m_{ij} εξαρτώνται από ένα μικρότερο σύνολο παραμέτρων (π.χ. κάποιες από τις $\mu, \lambda_i^X, \lambda_j^Y, \lambda_{ij}^{XY}$) με λογαριθμική συνάρτηση σύνδεσης (δηλαδή η σχέση των m_{ij} με τα $\mu, \lambda_i^X, \lambda_j^Y, \lambda_{ij}^{XY}$ είναι λογαριθμοκανονική, (loglinear)). Επομένως έχουμε το κλασσικό γενικευμένο γραμμικό μοντέλο και οι εκτιμήτριες μέγιστης πιθανοφάνειας μπορούν να βρεθούν χρησιμοποιώντας τις γνωστές επαναληπτικές μεθόδους (Fisher's scoring method) (βλ, Cox, (1970)).

5.1.6 Έλεγχος καλής προσαρμογής

Για διδιάστατους πίνακες συνάφειας, η στατιστική συνάρτηση που προκύπτει από το γενικευμένο λόγο πιθανοφανειών (likelihood ratio statistic ή Deviance) και το χ^2 του Pearson αντίστοιχα είναι:

$$G^2 = 2\sum_{i,j} N_{ij} \log \frac{N_{ij}}{\hat{m}_{ij}}, \qquad X^2 = \sum_{i,j} \frac{\left(N_{ij} - \hat{m}_{ij}\right)}{\hat{m}_{ij}}$$

Όταν το μοντέλο είναι σωστό, οι παραπάνω στατιστικές συναρτήσεις ακολουθούν ασυμπτωτικά χι-τετράγωνο κατανομή. Οι βαθμοί ελευθερίας είναι ίσοι με το πλήθος των παραμέτρων στο πλήρες μοντέλο μείον το πλήθος των (ανεξάρτητων) παραμέτρων στο μοντέλο που εξετάζουμε. Άρα απορρίπτουμε το εκάστοτε μοντέλο σε επίπεδο σημαντικότητας α, αν

$$G^2 > \chi^2_{\mathrm{df}, \alpha} \quad \text{ή} \quad \mathrm{X}^2 > \chi^2_{\mathrm{df}, \alpha}$$

Τα διαστήματα εμπιστοσύνης και οι έλεγχοι για τις παραμέτρους του μοντέλου γίνονται σύμφωνα με τη γενική θεωρία των γενικευμένων γραμμικών μοντέλων (βλ, Cox, (1970)).

5.1.7 Μελέτη τρισδιάστατων πινάκων χρησιμοποιώντας γενικευμένα γραμμικά μοντέλα

Σε πίνακες συνάφειας με τρεις διαστάσεις (X, Y, Z κατηγορικές μεταβλητές με I, J, K στάθμες αντίστοιχα) μπορούμε και πάλι να θεωρήσουμε το μοντέλο:

1) *Poisson:* Οι συχνότητες N_{ijk} σε κάθε κελί είναι ανεξάρτητες και ακολουθούν την κατανομή Poisson με μέση τιμή m_{ijk} (βλ, Dobson, (1990)).

5.1.8 Παραμετροποίηση του μοντέλου

Και σε πίνακες συνάφειας με περισσότερες από δύο διαστάσεις οι συνθήκες υπόθεσης (π.χ. ανεξαρτησίας μεταξύ X, Y, Z) μπορούν να εκφραστούν ως πολλαπλασιαστικά μοντέλα στα οποία οι αναμενόμενες συχνότητες είναι ίσες με γινόμενα περιθωρίων πιθανοτήτων και προκαθορισμένων αθροισμάτων συχνοτήτων (βλ, Dobson, (1990)). Συνεπώς, και στις τρεις διαστάσεις υπό τις συνήθεις υποθέσεις, ο λογάριθμος των αναμενόμενων συχνοτήτων γράφεται στη μορφή

$$\eta_{ijk} = \log m_{ijk} = \log E(N_{ijk}) = x_i^T \beta$$

για κατάλληλες παραμέτρους β. Η ανάλογη με τη διδιάστατη περίπτωση παραμετροποίηση εδώ είναι:

$$\mu = \bar{\eta}_{\bullet\bullet\bullet}, \quad \lambda_{\mathrm{i}}^{\mathrm{X}} = \bar{\eta}_{i\bullet\bullet} - \bar{\eta}_{\bullet\bullet\bullet}, \quad \lambda_{\mathrm{j}}^{\mathrm{Y}} = \bar{\eta}_{\bullet j\bullet} - \bar{\eta}_{\bullet\bullet\bullet}, \quad \lambda_{\mathrm{k}}^{\mathrm{Z}} = \bar{\eta}_{\bullet\bullet k} - \bar{\eta}_{\bullet\bullet\bullet}$$

$$\lambda_{\mathrm{ij}}^{\mathrm{XY}} = \bar{\eta}_{\mathrm{ij}\bullet} - \bar{\eta}_{\mathrm{i}\bullet\bullet} - \bar{\eta}_{\bullet\mathrm{j}\bullet} + \bar{\eta}_{\bullet\bullet\bullet}, \quad \lambda_{\mathrm{ik}}^{\mathrm{XZ}} = \bar{\eta}_{\mathrm{i}\bullet k} - \bar{\eta}_{\mathrm{i}\bullet\bullet} - \bar{\eta}_{\bullet\bullet k} + \bar{\eta}_{\bullet\bullet\bullet}, \quad \lambda_{\mathrm{jk}}^{\mathrm{YZ}} = \bar{\eta}_{\bullet jk} - \bar{\eta}_{\bullet j\bullet} - \bar{\eta}_{\bullet\bullet k} + \bar{\eta}_{\bullet\bullet\bullet}$$

$$\lambda_{\mathrm{ijk}}^{\mathrm{XYZ}} = \eta_{\mathrm{ijk}} - \bar{\eta}_{\mathrm{ij}\bullet} - \bar{\eta}_{\mathrm{i}\bullet k} - \bar{\eta}_{\bullet jk} + \bar{\eta}_{\mathrm{i}\bullet\bullet} + \bar{\eta}_{\bullet j\bullet} + \bar{\eta}_{\bullet\bullet k} + \bar{\eta}_{\bullet\bullet\bullet}$$

και αντικαθιστώντας προκύπτει το πλήρες μοντέλο

$$\eta_{\mathrm{ij}} k = \log m_{ijk} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Y}} + \lambda_{\mathrm{k}}^{\mathrm{Z}} + \lambda_{\mathrm{ij}}^{\mathrm{XY}} + \lambda_{\mathrm{ik}}^{\mathrm{XZ}} + \lambda_{\mathrm{jk}}^{\mathrm{YZ}} + \lambda_{\mathrm{ijk}}^{\mathrm{XYZ}}$$

Τα αθροίσματα των παραμέτρων ως προς οποιοδήποτε δείκτη είναι ίσο με 0:

$$\sum_i \lambda_{\mathrm{i}}^{\mathrm{X}} = \sum_j \lambda_{\mathrm{j}}^{\mathrm{Y}} = \sum_k \lambda_{\mathrm{k}}^{\mathrm{Z}} = \sum_i \lambda_{\mathrm{ij}}^{\mathrm{XY}} = \sum_j \lambda_{\mathrm{ij}}^{\mathrm{XY}} = \ldots \sum_k \lambda_{\mathrm{ijk}}^{\mathrm{XYZ}} = 0$$

Το πλήρες μοντέλο προφανώς έχει IJK ανεξάρτητες παραμέτρους (ίσες με το πλήθος κελιών του πίνακα).

Στο εξής θα θεωρήσουμε ιεραρχικά μοντέλα. Στα μοντέλα αυτά όταν υπάρχει η αλληλεπίδραση μεταξύ κάποιων μεταβλητών τότε στο μοντέλο θα υπάρχουν και οι μικρότερες αλληλεπιδράσεις καθώς και οι κύριες επιδράσεις μεταξύ των συγκεκριμένων μεταβλητών. Για παράδειγμα, αν στο μοντέλο υπάρχει η $\lambda_{\mathrm{ik}}^{\mathrm{XZ}}$ τότε θα υπάρχουν και οι $\lambda_{\mathrm{i}}^{\mathrm{X}}, \lambda_{\mathrm{i}}^{\mathrm{Z}}$ ενώ π.χ. αν υπάρχει η $\lambda_{\mathrm{ijk}}^{\mathrm{XYZ}}$ τότε θα υπάρχουν όλες οι ανά οι δύο αλληλεπιδράσεις των X, Y, Z καθώς και οι κύριες επιδράσεις τους, δηλαδή θα έχουμε το πλήρες μοντέλο. Μερικά από τα ιεραρχικά μοντέλα που μπορούν να θεωρηθούν σε ένα τριδιάστατο πίνακα συνάφειας δίνονται στον παρακάτω πίνακα:

Log Linear Model	*Συμβολισμός*
$\log m_{ijk} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Y}} + \lambda_{\mathrm{k}}^{\mathrm{Z}}$	(X, Y, Z)
$\log m_{ijk} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Y}} + \lambda_{\mathrm{k}}^{\mathrm{Z}} + \lambda_{\mathrm{ij}}^{\mathrm{XY}}$	(XY, Z)
$\log m_{ijk} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Y}} + \lambda_{\mathrm{k}}^{\mathrm{Z}} + \lambda_{\mathrm{ij}}^{\mathrm{XY}} + \lambda_{\mathrm{jk}}^{\mathrm{YZ}}$	(XY, YZ)
$\log m_{ijk} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Y}} + \lambda_{\mathrm{k}}^{\mathrm{Z}} + \lambda_{\mathrm{ij}}^{\mathrm{XY}} + \lambda_{\mathrm{jk}}^{\mathrm{YZ}} + \lambda_{\mathrm{ik}}^{\mathrm{XZ}}$	(XY, YZ, XZ)
$\log m_{ijk} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Y}} + \lambda_{\mathrm{k}}^{\mathrm{Z}} + \lambda_{\mathrm{ij}}^{\mathrm{XY}} + \lambda_{\mathrm{jk}}^{\mathrm{YZ}} + \lambda_{\mathrm{ik}}^{\mathrm{XZ}} + \lambda_{\mathrm{ijk}}^{\mathrm{XYZ}}$	(XYZ)

5.1.9 Μοντέλο κατανομής συχνοτήτων κελιών

Ας θεωρήσουμε αρχικά και εδώ το μοντέλο Poisson (οι συχνότητες σε κάθε κελί είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με μέση τιμή m_{ijk}). Η από κοινού συνάρτηση πιθανότητας των συχνοτήτων N_{ijk} των κελιών σε τρισδιάστατους πίνακες συνάφειας της μορφής:

$$f(n;m) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} e^{-m_{ijk}} \frac{m_{ijk}^{n_{ijk}}}{n_{ijk}!}$$

όπου οι μέσες τιμές m_{ijk} εξαρτώνται από κάποιες από τις παραμέτρους:

$$\mu, \lambda_{\mathrm{i}}^{\mathrm{X}}, \lambda_{\mathrm{j}}^{\mathrm{Y}}, \lambda_{\mathrm{k}}^{\mathrm{Z}}, \lambda_{\mathrm{ij}}^{\mathrm{XY}}, \lambda_{\mathrm{ik}}^{\mathrm{XZ}}, \lambda_{\mathrm{jk}}^{\mathrm{YZ}}, \lambda_{\mathrm{ijk}}^{\mathrm{XYZ}}$$

ανάλογα με το μοντέλο (π.χ. $\log m_{ijk} = \mu + \lambda_{\mathrm{i}}^{\mathrm{X}} + \lambda_{\mathrm{j}}^{\mathrm{Z}} + \lambda_{\mathrm{k}}^{\mathrm{Z}}$ κάτω από την υπόθεση της αμοιβαίας ανεξαρτησίας). Αν και σε ορισμένες περιπτώσεις (π.χ. της αμοιβαίας ανεξαρτησίας, από κοινού ανεξαρτησίας, δεσμευμένης ανεξαρτησίας) μπορούμε να βρούμε κλειστές εκφράσεις για τις εκτιμήτριες μέγιστης πιθανοφάνειας όσων από τις παραπάνω παραμέτρους περιέχονται στο μοντέλο μεγιστοποιώντας την συνάρτηση πιθανοφάνειας, είναι γενικά προτιμότερο (και αυτό γίνεται και από τα περισσότερα στατιστικά πακέτα) να χρησιμοποιήσουμε τη θεωρία των γενικευμένων γραμμικών μοντέλων.

Παρατηρούμε ότι η παραπάνω κατανομή f(**n;m**) ανήκει στην εκθετική οικογένεια κατανομών και μάλιστα τα m_{ijk}, εξαρτώνται από ένα μικρότερο σύνολο παραμέτρων (π.χ. κάποιες από τις $\mu, \lambda_{\mathrm{i}}^{\mathrm{X}}, \lambda_{\mathrm{j}}^{\mathrm{Z}}, \lambda_{\mathrm{k}}^{\mathrm{Z}}, \lambda_{\mathrm{ij}}^{\mathrm{XY}}, \lambda_{\mathrm{ik}}^{\mathrm{XZ}}, \lambda_{\mathrm{jk}}^{\mathrm{YZ}}, \lambda_{\mathrm{ijk}}^{\mathrm{XYZ}}$), με λογαριθμική συνάρτηση σύνδεσης (loglinear model). Επομένως και εδώ έχουμε το κλασσικό γενικευμένο γραμμικό μοντέλο και οι εκτιμήτριες μέγιστης πιθανοφάνειας μπορούν να βρεθούν χρησιμοποιώντας τις γνωστές επαναληπτικές μεθόδους (βλ, Dobson, (1990)).

5.1.10 Έλεγχος καλής προσαρμογής

Για τρισδιάστατους πίνακες συνάφειας, η στατιστική συνάρτηση που προκύπτει από το γενικευμένο λόγο πιθανοφανειών (likelihood ratio statistic ή Deviance) και το χι-τετράγωνο του Pearson αντίστοιχα είναι:

$$G^2 = 2\sum_{i,j,k} N_{ijk} \log \frac{N_{ijk}}{\hat{m}_{ijk}}, \qquad X^2 = \sum_{i,j,k} \frac{\left(N_{ijk} - \hat{m}_{ijk}\right)}{\hat{m}_{ijk}}$$

Όταν το μοντέλο είναι σωστό, οι παραπάνω στατιστικές συναρτήσεις ακολουθούν ασυμπτωτικά χ^2 κατανομή. Οι βαθμοί ελευθερίας της κατανομής αυτής είναι ίσοι με το πλήθος των παραμέτρων στο πλήρες μοντέλο μείον το πλήθος των (ανεξάρτητων) παραμέτρων στο μοντέλο που εξετάζουμε. Άρα απορρίπτουμε το εκάστοτε μοντέλο σε επίπεδο σημαντικότητας α, αν

$$G^2 > \chi^2_{\mathrm{df}, \alpha} \quad \text{ή} \quad \mathrm{X}^2 > \chi^2_{\mathrm{df}, \alpha}$$

Τα διαστήματα εμπιστοσύνης και οι έλεγχοι για τις παραμέτρους του μοντέλου γίνονται και εδώ σύμφωνα με τη γενική θεωρία των γενικευμένων γραμμικών μοντέλων (βλ, Dobson, (1990)).

5.1.11 Σύγκριση μοντέλων

Έστω δυο παραμετρικά μοντέλα M_1 και M_2 τέτοια ώστε το M_2 να περιέχεται στο M_1. Αν επιθυμούμε να ελέγξουμε την υπόθεση της μορφής: H_0: «το M_2 είναι γνωστό» έναντι του μεγαλύτερου μοντέλου H_1: «το M_1 είναι γνωστό», τότε σύμφωνα με το κριτήριο του γενικευμένου λόγου πιθανοφανειών, απορρίπτουμε σε επίπεδο σημαντικότητας α την H_0 όταν:

$$G^2(M_2 \mid M_1) = G^2(M_2) - G^2(M_1) > \chi^2_{\nu_2 - \nu_1; a}$$

Στους διδιάστατους και τριδιάστατους πίνακες συνάφειας έχουμε τα ακόλουθα μοντέλα:

Ι. Μοντέλα σε διδιάστατους πίνακες συνάφειας

α) X, Y: *μεταβλητές απόκρισης*

Ερμηνεία		**Μοντέλο**	**β.ε**
i) X, Y ανεξάρτητες	$m_{ij} = n\pi_{i\bullet}\pi_{\bullet j}$	(X, Y)	IJ-I-J+1
ii) Πλήρες μοντέλο	$m_{ij} = n\pi_{ij}$	(XY)	0

β) *X: μεταβλητή απόκρισης, Y ερμηνευτική μεταβλητή*. Στα μοντέλα πρέπει να περιλαμβάνεται ο όρος $\mu + \lambda_j^Y$.

Ερμηνεία		**Μοντέλο**	**β.ε**
i) Η X έχει την ίδια κατανομή σε όλες τις στάθμες της Y	$m_{ij} = n_{\bullet j}\pi_{i\bullet}$	(X, Y)	IJ-I-J+1
ii) Πλήρες μοντέλο	$m_{ij} = n_{\bullet j}\pi_{ij}$	(XY)	0

ΙΙ. Μοντέλα σε τριδιάστατους πίνακες συνάφειας

α) X, Y, Z: *μεταβλητές απόκρισης*

Ερμηνεία		**Μοντέλο**	**β.ε**
i) X,Y,Z αμοιβαία ανεξάρτητες	$m_{ijk} = n\pi_{i\bullet\bullet}\pi_{\bullet j\bullet}\pi_{\bullet\bullet\bullet}$	(X, Y, Z)	IJK-I-J-K+2
ii) Z ανεξάρτητη από τις X,Y	$m_{ijk} = n\pi_{ij\bullet}\pi_{\bullet\bullet k}$	(XY, Y)	(K-1)(IJ-1)
iii) X,Z ανεξάρτητες δεδομένης της Y	$m_{ijk} = n\pi_{ij\bullet}\pi_{\bullet jk}\pi_{\bullet j\bullet}$	(XY, YZ)	J(I-1)(K-1)
iv) Ανά δύο εξαρτημένες	$m_{ijk} = n\pi_{ij\bullet}\pi_{\bullet jk}\pi_{i\bullet k}$	(XY, XZ, YZ)	(I-1)(J-1)(K-1)
v) Πλήρες μοντέλο	$m_{ij} = n\pi_{ijk}$	(XYZ)	0

β) *X,Y μεταβλητές απόκρισης, Z: ερμηνευτική μεταβλητή*. Στα μοντέλα πρέπει να περιλαμβάνεται ο όρος $\mu + \lambda_k^Z$.

Ερμηνεία		**Μοντέλο**	**β.ε**
i) Οι X,Y έχουν την ίδια κατανομή σε όλες τις στάθμες της Z	$m_{ijk} = n_{\bullet\bullet k}\pi_{ij\bullet}$	(XY, Z)	(K-1)(IJ-1)
ii) Οι X,Y ανεξάρτητες σε κάθε στάθμη της ερμηνευτικής μεταβλητής Z	$m_{ijk} = n_{\bullet\bullet k}\pi_{i\bullet k}\pi_{\bullet jk}$	(XY, YZ)	K(I-1)(J-1)
iii) Πλήρες μοντέλο	$m_{ijk} = n_{\bullet\bullet k}\pi_{ijk}$	(XYZ)	0

γ) *X: μεταβλητή απόκρισης, Y,Z: ερμηνευτικές μεταβλητές.* Στα μοντέλα πρέπει να περιλαμβάνεται ο όρος $\mu + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$.

Ερμηνεία		**Μοντέλο**	**β.ε**
i) Η X έχει ίδια κατανομή σε όλες τις στάθμες της Y (για δεδομένη στάθμη της Z)	$m_{ijk} = n_{\bullet jk}\pi_{i\bullet k}$	(XY, YZ)	K(I-1)(J-1)
ii) Η X έχει ίδια κατανομή σε όλες τις στάθμες των Y,Z	$m_{ijk} = n_{\bullet jk}\pi_{i\bullet\bullet}$	(X, YZ)	(I-1)(JK-1)
iii) Πλήρες μοντέλο	$m_{ijk} = n_{\bullet jk}\pi_{ijk}$	(XYZ)	0

(βλ, Dobson, (1990))

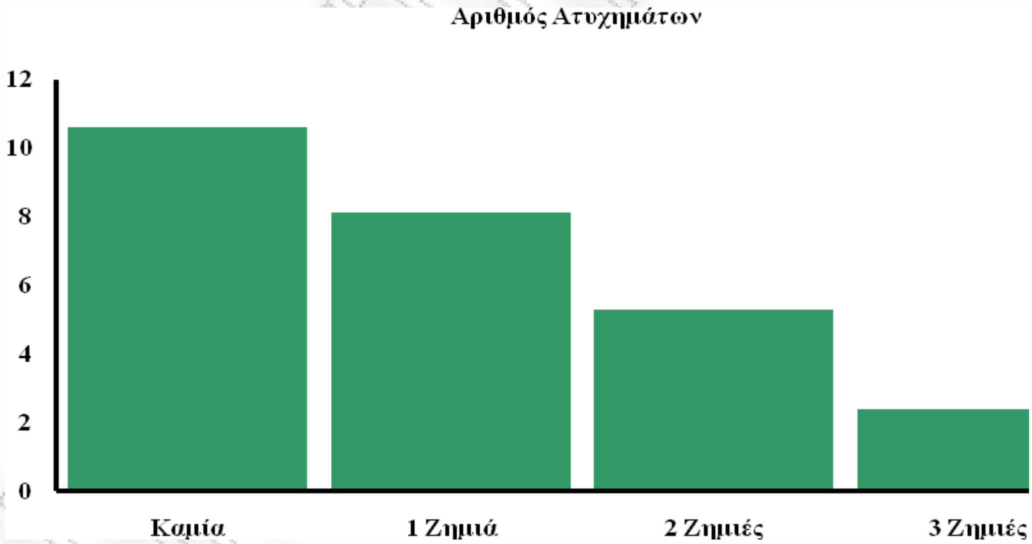
ΚΕΦΑΛΑΙΟ 6
ΕΜΠΕΙΡΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

6 Περιγραφικά Στατιστικά για τις Απαιτήσεις

Οι παρακάτω πίνακες – γραφήματα μας δίνουν κάποια αποτελέσματα για τις συχνότητες των μεταβλητών μας τα οποία θα μας βοηθήσουν να κατανοήσουμε τη δομή του χαρτοφυλακίου το οποίο εξετάζεται και να βγάλουμε κάποια συμπεράσματα. Θα πρέπει βέβαια να σημειωθεί, ότι έχουν γίνει κάποιες ομαδοποιήσεις στις συγγενείς μεταβλητές, έτσι ώστε να μπορέσουμε να τις εξετάσουμε καλύτερα.

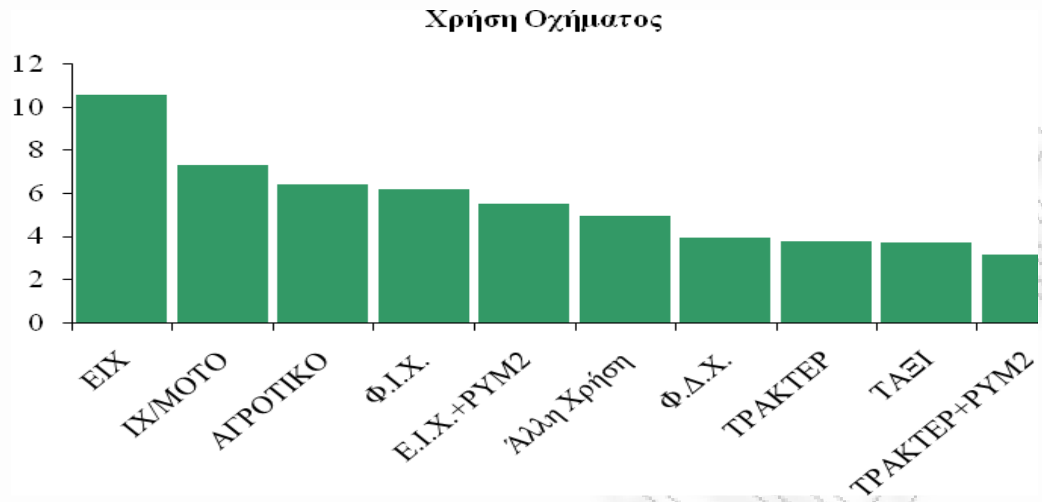
Πιο συγκεκριμένα σύμφωνα με το παρακάτω γράφημα: το 91,7% (40.183 περιπτώσεις από το σύνολο του δείγματος που είναι ίσο με 43.819) των ασφαλισμένων δεν έχει εμπλακεί σε ατύχημα (Claim), το 7,8% έχει προκαλέσει 1 ζημία (3.429 περιπτώσεις), το 0,4% 2 ζημίες (196 περιπτώσεις) ενώ οι οδηγοί που έχουν προξενήσει 3 ατυχήματα αποτελούν το 0.03% (11 περιπτώσεις), ποσότητα αμελητέα για το μέγεθος του χαρτοφυλακίου.

Γράφημα 6.1. Συχνότητες Αριθμού Ατυχημάτων



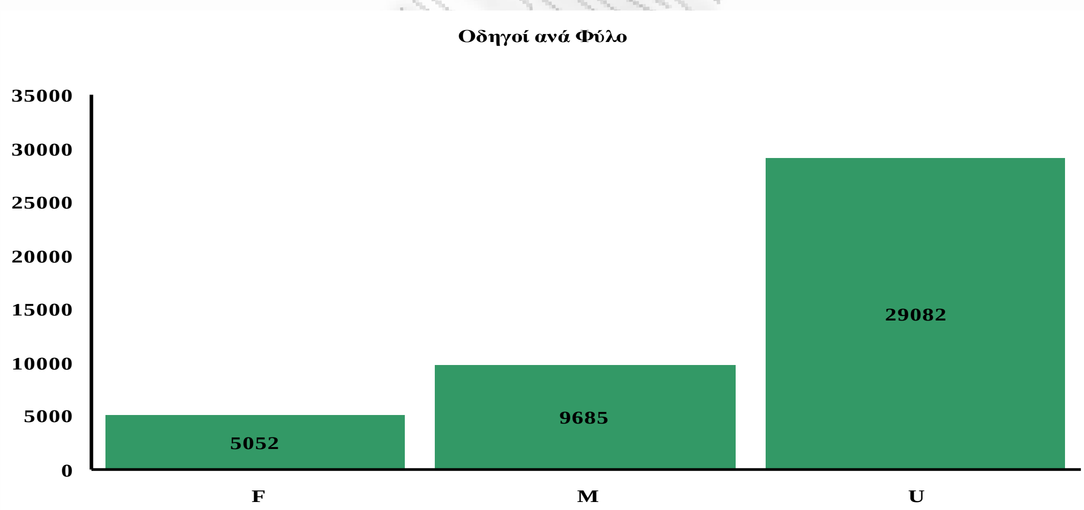
Επιπλέον, σύμφωνα με το παρακάτω γράφημα (Coded Use) το 93,4% των ατυχημάτων αφορούν Επιβατικά Ιδιωτικής Χρήσης, το 3,4% Μηχανάκια και Μοτοποδήλατα το 1,4% Φορτηγά Ιδιωτικής Χρήσης Αγροτικά, το 1,1% Φορτηγά Ιδιωτικής Χρήσης ενώ οι υπόλοιπες συχνότητες είναι πολύ μικρές όπως φαίνεται και παρακάτω.

Γράφημα 6.2. Συχνότητες για τον τρόπο χρήση του Οχήματος



Διπλάσιο είναι το ποσοστό των ανδρών οδηγών σε σχέση με αυτό των γυναικών, ενώ το υπόλοιπο 66,4% των οδηγών του χαρτοφυλακίου αποτελείται από άνδρες και γυναίκες οδηγούς για το ίδιο ασφαλιζόμενο όχημα ή απλά δεν έχουμε την πληροφορία για το φύλο του καθενός ξεχωριστά.

Γράφημα 6.3. Συχνότητες για τους Οδηγούς οχημάτων ανά Φύλο



Παρακάτω, ο πίνακας με το φύλο του οδηγού, παρουσιάζει τη μέση συχνότητα της ζημίας κατά φθίνουσα σειρά που κάνει ο ασφαλιζόμενος, ανά νομό δήλωσης της συνήθους περιοχής κίνησης του οχήματος. Η συνολική συχνότητα ανέρχεται στο 8,7% (δηλαδή, από τα 100 οχήματα τα 8.7 προκαλούν ατύχημα). Επιπροσθέτως, είναι άξιο παρατήρησης ότι η Αττική αν και κατέχει το μεγαλύτερο πλήθος των ασφαλιζόμενων οχημάτων έχει μέση συχνότητα ατυχήματος 9,6%, δηλαδή μία (1) ποσοστιαία μονάδα πάνω από το μέσο όρο, ενώ οι υπόλοιπες με υψηλή συχνότητα αποκλίνουν από τη μέση ζημία έως και 70%, όπως συμβαίνει πχ., στην ακραία περίπτωση της Θεσπρωτίας.

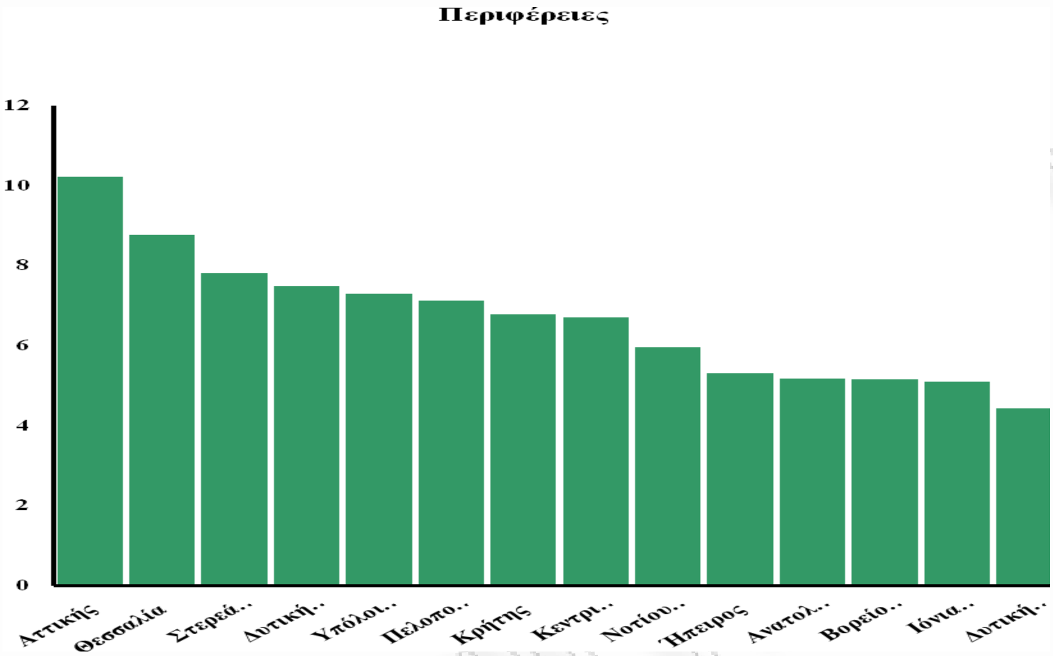
Πίνακας 6.1. Μέση συχνότητα απαιτήσεων ανά Νομό

Ονομασία Έδρας	Μέσος	Ν	τ.α
ΗΜΑΘΙΑ	0,263158	38	0,446258
ΘΕΣΠΡΩΤΙΑ	0,25	8	0,707107
ΚΑΒΑΛΑ	0,210526	19	0,535303
ΚΕΦΑΛΛΗΝΙΑ	0,173913	23	0,387553
ΛΕΥΚΑΔΑ	0,166667	12	0,389249
ΠΙΕΡΙΑ	0,157895	19	0,374634
ΦΛΩΡΙΝΑ	0,153846	13	0,375534
ΑΧΑΙΑ	0,134694	735	0,382992
ΞΑΝΘΗ	0,125	16	0,341565
ΕΛΛΑΔΑ	0,119241	1476	0,369162
ΡΕΘΥΜΝΟ	0,116279	43	0,390927
ΜΕΣΣΗΝΙΑ	0,104839	124	0,37867
ΘΕΣΣΑΛΟΝΙΚΗ	0,101587	630	0,322693
ΑΤΤΙΚΗ	0,096682	27337	0,315177
ΗΡΑΚΛΕΙΟ	0,096154	364	0,313316
ΑΡΤΑ	0,09375	32	0,296145
ΔΡΑΜΑ	0,09375	32	0,296145
ΚΥΚΛΑΔΕΣ	0,092369	249	0,316709
ΚΑΣΤΟΡΙΑ	0,090909	22	0,294245
ΧΙΟΣ	0,088235	34	0,287902
ΒΟΙΩΤΙΑ	0,08519	763	0,293102
ΗΛΕΙΑ	0,08062	645	0,272462
ΙΩΑΝΝΙΝΑ	0,076087	92	0,26659
ΛΑΡΙΣΑ	0,075384	2149	0,272744
ΛΑΚΩΝΙΑ	0,072607	303	0,284259
ΝΗΣΙΑ ΣΑΡΩΝΙΚΟΥ ΕΚΤΟΣ ΣΑΛΑΜΙΝΑΣ	0,070175	57	0,319578
ΡΟΔΟΠΗ	0,070175	57	0,257713
ΑΙΤΩΛΟΑΚΑΡΝΑΝΙΑ	0,068966	406	0,253708
ΑΡΓΟΛΙΔΑ	0,067568	74	0,252716
ΑΡΚΑΔΙΑ	0,065217	92	0,248262
ΚΙΛΚΙΣ	0,064516	31	0,249731
ΚΕΡΚΥΡΑ	0,06422	109	0,281373
ΚΟΡΙΝΘΙΑ	0,061747	664	0,247059

ΕΥΒΟΙΑ	0,059801	602	0,237314
ΣΑΛΑΜΙΝΑ	0,058252	103	0,235365
ΧΑΝΙΑ	0,056054	446	0,230284
ΠΕΛΛΑ	0,055556	18	0,235702
ΚΑΡΔΙΤΣΑ	0,054916	1129	0,231774
ΦΘΙΩΤΙΔΑ	0,053926	1057	0,225979
ΖΑΚΥΝΘΟΣ	0,052632	19	0,229416
ΚΟΖΑΝΗ	0,047619	21	0,218218
ΤΡΙΚΑΛΑ	0,047155	1951	0,219161
ΜΑΓΝΗΣΙΑ	0,044861	1226	0,207084
ΠΡΕΒΕΖΑ	0,043478	69	0,205425
ΦΩΚΙΔΑ	0,04	25	0,2
ΣΑΜΟΣ	0,039216	51	0,196039
ΣΕΡΡΕΣ	0,038462	26	0,196116
ΕΒΡΟΣ	0,038462	52	0,194184
ΓΡΕΒΕΝΑ	0,035714	28	0,188982
ΔΩΔΕΚΑΝΗΣΑ	0,029197	137	0,168976
ΧΑΛΚΙΔΙΚΗ	0,017544	57	0,132453
ΕΥΡΥΤΑΝΙΑ	0	13	0
ΚΥΘΗΡΑ	0	2	0
ΛΑΣΙΘΙ	0	31	0
ΛΕΣΒΟΣ	0	88	0
Σύνολο	**0,087953**	**43819**	**0,301117**

Για λόγους ευκολίας προσαρμογής του μοντέλου χωρίζουμε τους Νομούς στις 13 περιφέρειες σύμφωνα με το νόμο του Καλλικράτη. Ομαδοποιούμε σύμφωνα με το παρακάτω γράφημα (με κωδικό 1 χαρακτηρίσαμε τα οχήματα που κινούνται στην περιοχή της Ανατολικής Μακεδονίας και Θράκης, με 2 της Κεντρικής Μακεδονίας, με 3 της Δυτικής Μακεδονίας, με 4 της Ηπείρου, με 5 της Θεσσαλίας, με 6 των Ιονίων Νήσων, με 7 της Δυτικής Ελλάδας και Αιτωλοακαρνανίας, με 8 της Στερεάς Ελλάδας, με 9 της ευρύτερης περιοχής της Αττικής, με 10 της Πελοποννήσου, με 11 του Βορείου Αιγαίου, με 12 του Νοτίου Αιγαίου, με 13 της Κρήτης και με 14 της Υπόλοιπης Ελλάδας):

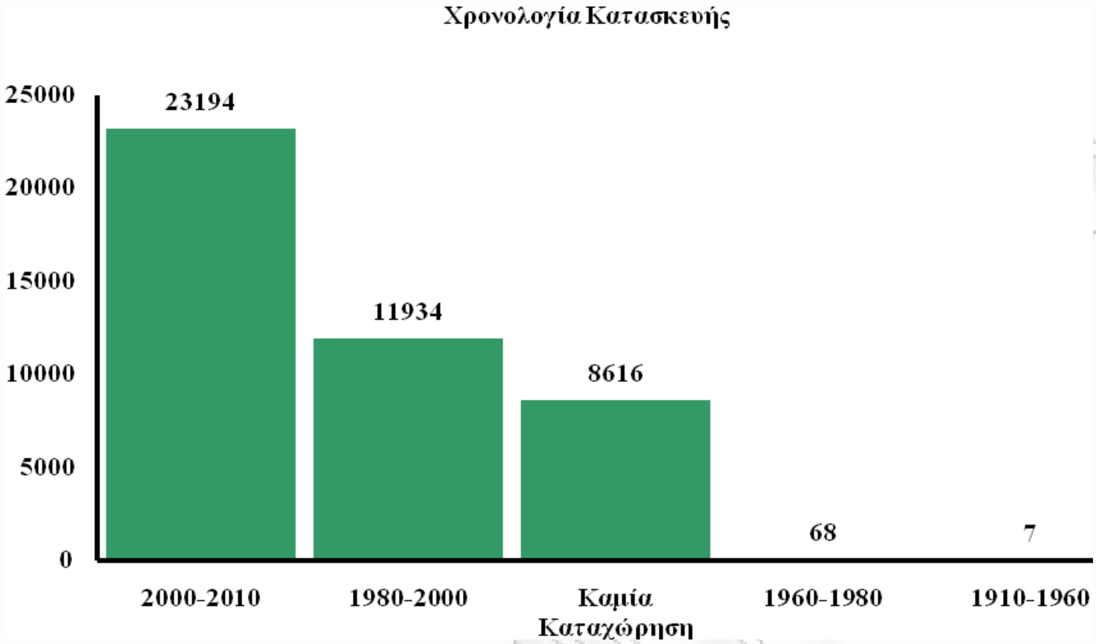
Γράφημα 6.4. Συχνότητες για τον αριθμό οχημάτων ανά Περιφέρεια



Παρατηρούμε από το παραπάνω γράφημα ότι 27.499 οχήματα από το σύνολο των 43.819 ολόκληρου του χαρτοφυλακίου (δηλαδή ποσοστό 63% περίπου), προέρχεται από την ευρύτερη περιοχή της Αττικής (με κωδικό περιφέρειας 9). Επίσης, 6.455 οχήματα (ποσοστό 15% περίπου του συνόλου) προέρχονται από την περιοχή της Θεσσαλίας (κωδικός 5) κοκ, μέχρι την περιφέρεια της Δυτικής Μακεδονίας (κωδικός 3) που παρατηρήθηκαν συνολικά 84 οχήματα (ποσοστό 0,2% του συνόλου).

Για το έτος κατασκευής των αυτοκινήτων του χαρτοφυλακίου μας, ενδεικτικό είναι το ακόλουθο γράφημα που περιέχει τις συχνότητες καθώς και τις σχετικές συχνότητες για κάθε περίοδο κατασκευής του οχήματος. Πιο συγκεκριμένα, το 53% των οχημάτων του χαρτοφυλακίου έχει έτος κατασκευής από το 2000 μέχρι σήμερα (σύνολο 23.194 οχήματα από τα 43.819), το 27% κατασκευάστηκε από το 1980 μέχρι το 2000 (σύνολο 11.934 οχήματα) κοκ.

Γράφημα 6.5. Συχνότητες για τη Χρονολογία Κατασκευής των οχημάτων



Όσον αφορά τις απαιτήσεις, θα ακολουθηθούν ορισμένοι πίνακες που θα περιέχουν περιγραφικά μέτρα (μέσος όρος και τυπική απόκλιση) που μας βοηθούν να κατανοήσουμε, κατά προσέγγιση, το πώς κινήθηκαν οι ζημιές κατά την ασφαλιστική περίοδο που εξετάστηκαν.

Πίνακας 6.2. Περιγραφικά Μέτρα των Απαιτήσεων ανάλογα με την Χρήση του οχήματος.

Κωδικός Χρήσης	Μέσος	N	Τυπική Απόκλιση
Άλλη Χρήση	0,2781	187	0,56578
ΕΙΧ	0,0887	40910	0,30124
ΦΙΧ	0,1521	480	0,40841
ΦΙΧ Αγροτικά	0,0835	623	0,29914
Μοτοσικλέτες	0,0293	1500	0,17652
Φ.Δ.Χ.	0,0769	52	0,26907
ΤΡΑΚΤΕΡ	0,0149	67	0,12217
Σύνολο	**0,088**	**43819**	**0,30112**

Ο παραπάνω πίνακας ουσιαστικά μας περιγράφει ότι στο σύνολο των 43.819 οχημάτων, τα 187 ανήκουν στην κατηγορία «Άλλη Χρήση», όπου κατά μέσο όρο το 27% των απαιτήσεων τον 187 οχημάτων αξιώνει απαίτηση, με μία απόκλιση της τάξεως του 0,56 ή 56%. Επίσης, στο σύνολο των οχημάτων τα 40.910 ανήκουν στην κατηγορία «Ε.Ι.Χ.» όπου κατά μέσο όρο το 8% των οχημάτων αυτών αξιώνει κάποια απαίτηση με μία απόκλιση της τάξεως του 0,30 ή

30% περίπου. Ανάλογα ερμηνεύονται και τα υπόλοιπα δεδομένα του πίνακα. Όσον αφορά τώρα το φύλο του οδηγού, παρατηρήθηκε ότι ένα ποσοστό κατά μέσο όρο περίπου 10% των 5.052 αντρών οδηγών (από τα 43.819 συμβόλαια) αξιώνουν κάποια απαίτηση, με μια απόκλιση της τάξεως του 33% περίπου. Επίσης, κατά μέσο όρο το 8% των 9.685 οδηγών που είναι γυναίκες αξιώνουν κάποια απαίτηση με απόκλιση της τάξεως του 29% περίπου, και τέλος, και για τα δυο φύλα (unisex) κατά μέσο όρο το 8% των 29.082 αξιώνουν κάποια απαίτηση με μια απόκλιση της τάξεως του 29% περίπου. Πολύ κατατοπιστικά είναι και τα αποτελέσματα του παρακάτω πίνακα.

Πίνακας 6.3. Περιγραφικά Μέτρα των Απαιτήσεων ανάλογα με το Φύλο του οδηγού.

Φύλο	**Μέσος**	**N**	**Τυπική Απόκλιση**
Άντρες	0.1047	5052	0.33226
Γυναίκες	0.0881	9685	0.29279
Unisex	0.0850	29082	0.29658
Σύνολο	**0.0880**	**43819**	**0.30112**

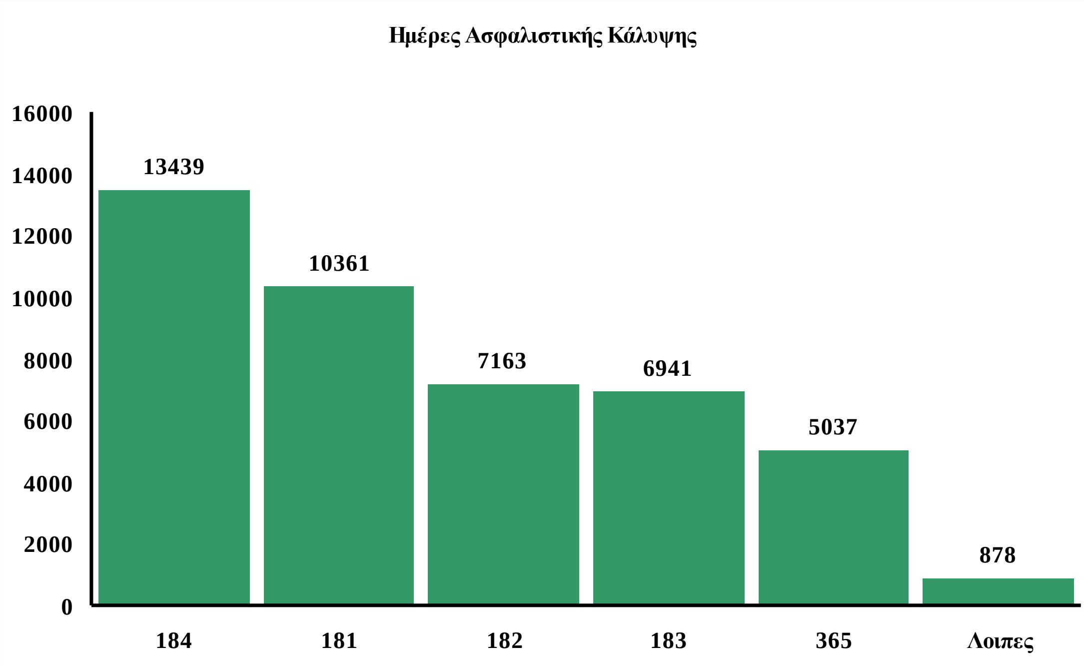
Στη συνέχεια, θα περιγράψουμε τον μέσο όρο των απαιτήσεων ανά περίοδο ασφαλιστικής κάλυψης. Σύμφωνα με τον παρακάτω πίνακα παρατηρούμε ότι, για εξαμηνιαία ασφαλιστική κάλυψη το 8% των 37.904 συμβολαίων κατά μέσο όρο παρουσίασε αξίωση για απαίτηση με τυπική απόκλιση της τάξεως του 27% περίπου. Τα ετήσια συμβόλαια παρουσίασαν αξίωση κατά μέσο όρο περίπου 10% των 5.037 με τυπική απόκλιση της τάξεως του 38% περίπου και για διάφορες ασφαλιστικές περιόδους, ο μέσος όρος απαίτησης άγγιξε το 23% των 878 συμβολαίων περίπου.

Πίνακας 6.4. Περιγραφικά Μέτρα των Απαιτήσεων ανάλογα με την Περίοδο Ασφαλιστικής Κάλυψης.

Ασφ. Περίοδος	**Μέσος**	**N**	**Τυπική Απόκλιση**
Εξάμηνο	,0814	37904	,27352
Ετήσιο	,1114	5037	,38398
Άλλο	,2346	878	,65099
Σύνολο	,0880	43819	,30112

Ακολουθεί το γράφημα για τις ημέρες ασφαλιστικής κάλυψης.

Γράφημα 6.6. Συχνότητες για την αξίωση απαίτησης σε σχέση με τις ημέρες ασφαλιστικής κάλυψης



6.1 Γενικευμένο Γραμμικό Μοντέλο Κατανομής Poisson για τις Απαιτήσεις

Στο σημείο αυτό θα εξετάσουμε το μοντέλο πολλαπλής γραμμικής παλινδρόμησης με τη βοήθεια της κατανομής Poisson (χρησιμοποιώντας τα γενικευμένα γραμμικά μοντέλα). Πιο συγκεκριμένα, στο χαρτοφυλάκιο των απαιτήσεων θεωρούμε ως εξαρτημένη μεταβλητή τις απαιτήσεις και ως ανεξάρτητες το Φύλο, τον τρόπο χρήσης του οχήματος και τις ημέρες της ασφαλιστική κάλυψης. Από τα αποτελέσματα που προέκυψαν παρατηρήσαμε ότι επηρεάζουν το μοντέλο οι γυναίκες, τα οχήματα Άλλης Χρήσης (με κωδικό χρήσης = 0), τα Φορτηγά Ιδιωτικής Χρήσης (με κωδικό χρήσης = 2), τα εξαμηνιαία και τα ετήσια συμβόλαιο με p-values μικρότερα από το επίπεδο σημαντικότητας που έχουμε ορίσει και είναι το α = 0,05 ή 5%. Ενδεικτικοί είναι οι πίνακες που ακολουθούν. Ο πίνακας 4 μας παρουσιάζει τα αποτελέσματα της ανάλυσης παλινδρόμησης θεωρώντας ότι η εξαρτημένη μεταβλητή ακολουθεί την κατανομή Poisson, και πιο συγκεκριμένα παρουσιάζει την τιμή των συντελεστών της πολλαπλής παλινδρόμησης των ανεξάρτητων μεταβλητών καθώς και το 95% διάστημα εμπιστοσύνης για τις τιμές που μπορούν να πάρουν οι συντελεστές αυτοί. Μια πρώτη εκτίμηση για την σημαντικότητα που έχουν οι ανεξάρτητες μεταβλητές στο μοντέλο (κατά πόσο δηλαδή επηρεάζουν το μοντέλο), μπορούμε να αποκτήσουμε από τα διαστήματα εμπιστοσύνης. Στις μεταβλητές όπου το διάστημα εμπιστοσύνης δεν περιέχει την τιμή μηδέν (0), περιμένουμε να έχουν επιρροή στο μοντέλο.

Πίνακας 6.5. 95% Διαστήματα Εμπιστοσύνης για το Μοντέλο Παλινδρόμησης Poisson

Παράμετροι	B	Τυπικό Σφάλμα	95% Wald Διάστημα Εμπιστοσύνης	
			Κάτω Φράγμα	Άνω Φράγμα
(Σταθερά)	-3,357	1,0025	-5,322	-1,392
Γυναίκες	,164	,0483	,069	,258
Άντρες	-,012	,0404	-,092	,067
Άλλη Χρήση	2,889	1,0097	,910	4,868
Ε.Ι.Χ.	1,885	1,0003	-,075	3,845
Φ.Ι.Χ.	2,341	1,0069	,368	4,315
Φ.Ι.Χ. Αγροτικά	1,806	1,0096	-,173	3,785
Μοτοσικλέτες	,644	1,0113	-1,338	2,626
Φ.Δ.Χ.	1,839	1,1183	-,353	4,031
Εξαμηνιαία	-1,047	,0732	-1,190	-,904
Ετήσια	-,673	,0820	-,834	-,513

Τα δεδομένα του πίνακα 6.5 επιβεβαιώνουν τα δεδομένα του πίνακα 6.4, όσον αφορά βέβαια την επιρροή των ανεξάρτητων στον μοντέλο παλινδρόμησης με την κατανομή Poisson. Πράγματι, οι μεταβλητές Γυναίκες, Άλλη Χρήση, Φορτηγά Ιδιωτικής Χρήσης, Εξαμηνιαία και Ετήσια επηρεάζουν την εξαρτημένη μεταβλητή των Απαιτήσεων με αποτέλεσμα τα p-values να είναι μικρότερο από το επίπεδο σημαντικότητας που έχει οριστεί για την εργασία μας και είναι ίσο με α = 0,05 ή 5%.

Πίνακας 6.6. Αποτελέσματα p-values

Παράμετροι	B	Τεστ Υποθέσεων		
		Wald χ^2	β.ε.	p-value
(Σταθερά)	-3,357	11,214	1	,001
Γυναίκες	,164	11,445	1	**,001**
Άντρες	-,012	,095	1	,758
Άλλη Χρήση	2,889	8,186	1	**,004**
Ε.Ι.Χ.	1,885	3,551	1	,059
Φ.Ι.Χ.	2,341	5,406	1	**,020**
Φ.Ι.Χ. Αγροτικά	1,806	3,200	1	,074
Μοτοσικλέτες	,644	,405	1	,524
Φ.Δ.Χ.	1,839	2,705	1	,100
Εξαμηνιαία	-1,047	204,749	1	**,000**
Ετήσια	-,673	67,419	1	**,000**

6.2 Γενικευμένο Γραμμικό Μοντέλο Αρνητικής Διωνυμικής Κατανομής για τις Απαιτήσεις

Στο σημείο αυτό θα προχωρήσουμε στην ανάλυση των δεδομένων μας με την χρήση των Γενικευμένων Γραμμικών Μοντέλων και πιο συγκεκριμένα θα θεωρήσουμε στο μοντέλο παλινδρόμησης ως εξαρτημένη μεταβλητή εκείνη των Απαιτήσεων και ανεξάρτητες το Φύλο, τον Κωδικό Χρήσης και τις Ημέρες Ασφαλιστικής Κάλυψης. Το σύνολο του δείγματος αποτέλεσαν 43.819 καταγεγραμμένες περιπτώσεις ασφαλιστηρίων συμβολαίων. Για να έχουμε μια πιο καλή εικόνα των δεδομένων μας, παρατίθεται ο πίνακας 6 που περιέχει βασικές περιγραφικές πληροφορίες ανά ανεξάρτητη μεταβλητή. Πιο συγκεκριμένα, παρατηρούμε ότι 5.052 ασφαλιστήρια συμβόλαια προήλθαν από γυναίκες οδηγούς (11,5% του συνόλου), 9.685 από άντρες οδηγούς (περίπου 22,1% του συνόλου) και τα υπόλοιπα 29.082 από οχήματα που τα οδηγούσαν άντρες και γυναίκες οδηγοί (το υπόλοιπο 66,4%).

Πίνακας 6.7. Βασικές πληροφορίες ανά Ανεξάρτητη Μεταβλητή

Ανεξάρτητες Μεταβλητές		N	Ποσοστό
Φύλο	Γυναίκες	5.052	11,5%
	Άντρες	9.685	22,1%
	Unisex	29.082	66,4%
	Σύνολο	43.819	100,0%
Κωδικός Χρήσης	Άλλη Χρήση	187	,4%
	ΕΙΧ	40.910	93,4%
	ΦΙΧ	480	1,1%
	ΦΙΧ Αγροτικά	623	1,4%
	Μοτοσικλέτα	1.500	3,4%
	Φ.Δ.Χ	52	,1%
	ΤΡΑΚΤΕΡ	67	,2%
	Σύνολο	43.819	100,0%
Ημέρες Κάλυψης	Εξάμηνο	37.904	86,5%
	Έτος	5.037	11,5%
	Άλλο	878	2,0%
	Σύνολο	43.819	100,0%

Αφού λοιπόν πραγματοποιήσαμε μια περιγραφική ανάλυση των δεδομένων μας, θα προχωρήσουμε στην πολλαπλή ανάλυση παλινδρόμησης με την υπόθεση ότι η εξαρτημένη μεταβλητή των απαιτήσεων ακολουθεί την Negative Binomial κατανομή. Το αποτέλεσμα του Omnibus τεστ για το αν υπάρχουν ανεξάρτητες μεταβλητές που επηρεάζουν το μοντέλο, έδωσε p-value = 0.00 που είναι μικρότερο από το επίπεδο σημαντικότητας α = 0.05. Το γεγονός αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχουν, τουλάχιστον μια, ανεξάρτητες μεταβλητές που επηρεάζουν την εξαρτημένη μεταβλητή των απαιτήσεων. Το αποτέλεσμα αυτό θα επιβεβαιωθεί και από τους πίνακες που ακολουθούν και παραθέτουν τα 95% διαστήματα εμπιστοσύνης για τις τιμές των ανεξάρτητων μεταβλητών στο μοντέλο καθώς επίσης και τα p – values τους. Συμπληρωματικό των παραπάνω συμπερασμάτων είναι και το τεστ Τύπου ΙΙΙ που πραγματοποιήθηκε για την επίδραση που ασκούν οι ανεξάρτητες μεταβλητές στο μοντέλο, που έβγαλε αποτελέσματα p-value μικρότερα του 0.05 για κάθε ανεξάρτητη μεταβλητή αντίστοιχα.

Πίνακας 6.8. Τεστ Τύπου ΙΙΙ για την επίδραση στο Μοντέλο

Πηγή	Τύπου ΙΙΙ		
	Wald χ^2	β.ε.	p-value
(Σταθερά)	175,818	1	,000
Φύλο	11,293	2	**,004**
Κωδικός Χρήσης	123,205	6	**,000**
Ημέρες Ασφαλιστικής Κάλυψης	206,772	2	**,000**

Εξαρτημένη Μεταβλητή: Απαιτήσεις

Ανεξάρτητες: (Σταθερά), Φύλο, Κωδικός Χρήσης, Ημέρες Ασφ. Κάλυψης

Πίνακας 6.9. Εκτιμήσεις Παραμέτρων Παλινδρόμησης

Παράμετροι	B	τ.σ.	95% Wald δ.ε.		Τεστ Υποθέσεων		
			Κάτω Φράγμα	Άνω Φράγμα	Wald χ^2	β.ε.	p-value
(Σταθερά)	-3,360	1,0111	-5,341	-1,378	11,043	1	,001
Γυναίκες	,166	,0509	,066	,265	10,614	1	**,001**
Άντρες	-,003	,0421	-,085	,079	,005	1	,943
Άλλη Χρήση	2,903	1,0206	,903	4,904	8,094	1	**,004**
Ε.Ι.Χ.	1,886	1,0083	-,090	3,862	3,498	1	,061
Φ.Ι.Χ.	2,348	1,0160	,357	4,339	5,340	1	**,021**
Φ.Ι.Χ. Αγροτικά	1,801	1,0185	-,195	3,797	3,127	1	,077
Μοτοσικλέτες	,646	1,0196	-1,352	2,644	,401	1	,526
Φ.Δ.Χ.	1,842	1,1340	-,380	4,065	2,640	1	,104
Εξαμηνιαία	-1,048	,0811	-1,206	-,889	167,038	1	**,000**
Ετήσια	-,675	,0902	-,852	-,498	55,956	1	**,000**

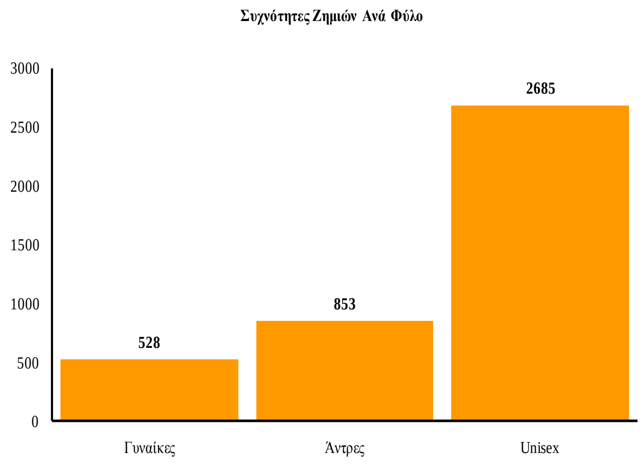
Εξαρτημένη Μεταβλητή: Απαιτήσεις
Ανεξάρτητες: (Σταθερά), Φύλο, Κωδικός Χρήσης, Ημέρες Ασφαλιστικής Κάλυψης
τ.σ: Τυπική Σφάλμα, δ.ε.: Διαστήματα Εμπιστοσύνης, β.ε.: Βαθμοί Ελευθερίας

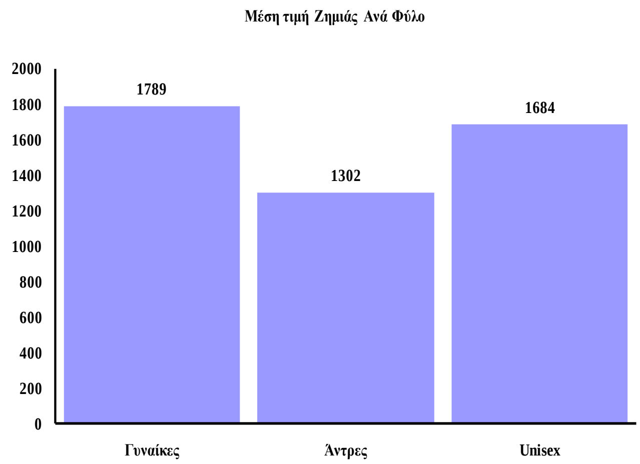
Τα ίδια αποτελέσματα με εκείνα της κατανομής Poisson προέκυψαν και για την κατανομή της Αρνητικής Διωνυμικής. Οι μεταβλητές που φαίνεται να επηρεάζουν την εξαρτημένη μεταβλητή των απαιτήσεων, είναι πάλι οι Γυναίκες, η Άλλη Χρήση του οχήματος, τα Φορτηγά Ιδιωτικής Χρήσης, τα Εξαμηνιαία και τα Ετήσια συμβόλαια με p-values μικρότερα από το επίπεδο σημαντικότητας α = 0.05 ή 5%. Καλύτερο μοντέλο όμως θεωρείται εκείνο που επιτυγχάνει μεγαλύτερη τιμή όσον αφορά το αποτέλεσμα του κριτηρίου Akaike (Akaike's Information Criterion (AIC)). Συνεπώς, το γενικευμένο γραμμικό μοντέλο με την κατανομή Poisson παρέχει πιο «φτωχά» αποτελέσματα από εκείνο της Αρνητικής Διωνυμικής, με αποτέλεσμα 26.406 έναντι του 26.461 αντίστοιχα.

6.3 Περιγραφικά Στατιστικά για τις Ζημιές

Σε αυτή την παράγραφο θα δώσουμε ορισμένα περιγραφικά μέτρα των εμπειρικών δεδομένων του χαρτοφυλακίου που εξετάσαμε σε σχέση με το μέγεθος των ζημιών. Πιο συγκεκριμένα, παρατηρήσαμε ότι παρόλο που οι άντρες οδηγοί προκαλούν περισσότερα στον αριθμό ατυχήματα, η μέση τιμή της ζημιάς τους είναι μικρότερη από εκείνη των γυναικών. Με άλλα λόγια, μπορεί οι γυναίκες να προκαλούν λιγότερα ατυχήματα, αλλά όταν συμβαίνει αυτό, το ύψος της ζημιάς παίρνει μεγάλες τιμές. Μπορούμε να πούμε δηλαδή, κατά προσέγγιση πάντα, βάση των περιγραφικών μέτρων που εξάγαμε, η μέση τιμή των ζημιών που προκαλούν οι άντρες είναι ίσο με € 1.302 ενώ η μέση τιμή ζημιών των γυναικών είναι ίση με €1.789. Ενδεικτικά είναι τα παρακάτω γραφήματα (βλ., Γραφήματα 6.7). Παρατηρούμε δηλαδή ότι οι γυναίκες προκαλούν ζημιές κατά μέσο όρο του ποσού των € 1.789 περίπου με συχνότητα 528 από το σύνολο των 4.066 ζημιών που προκλήθηκαν στο χαρτοφυλάκιο (ποσοστό περίπου 13%). Οι άντρες από την άλλη σημείωσαν τις 853 από τις 4.066 ζημιές του χαρτοφυλακίου (ποσοστό περίπου 21%) και κατά μέσο όρο αυτή η ζημιά δεν ξεπερνούσε τα € 1.302 περίπου. Τέλος, τις μεγαλύτερες ζημιές σημείωσαν τα οχήματα που οι οδηγοί τους ήταν και των δυο φύλων. Ουσιαστικά το 66% των ζημιών περίπου προκλήθηκαν από τους οδηγούς αυτούς (2.685 ζημίες από τις 4.066) και το μέσο ποσό της προκληθείσας ζημιάς άγγιξε το ποσό των € 1.684 περίπου.

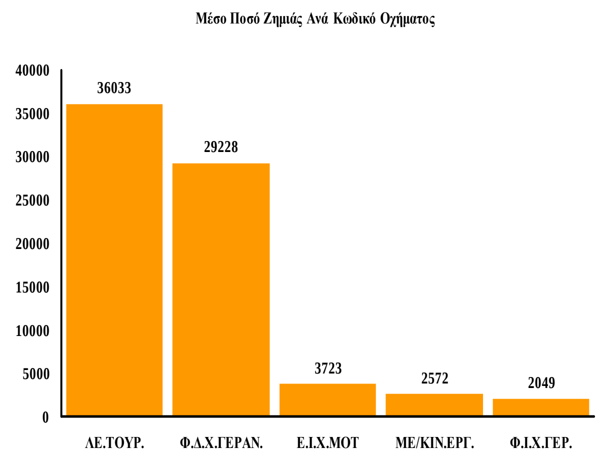
Γραφήματα 6.7. Περιγραφικά Μέτρα για τις Ζημιές Ανά Φύλο Οδηγού

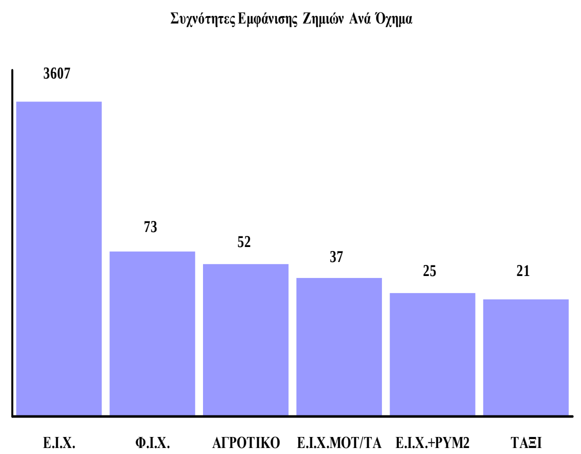




Ας δούμε τώρα τι συνέβη όταν κατηγοριοποιήσαμε τις ζημιές ανά κωδικό χρήσης του οχήματος. Παρατηρήσαμε λοιπόν, ότι την μεγαλύτερη μέση ζημιά προκάλεσαν τα τουριστικά λεωφορεία με 9 καταγεγραμμένα ατυχήματα (στο σύνολο των 4.066 ατυχημάτων) με ύψος μέσης ζημιάς που άγγιξαν το ποσό των € 36.033. Ακολούθησε 1 ατύχημα ενός φορτηγού ιδιωτικής χρήσης με γερανό που έφτασε το ποσό των € 29.228 περίπου. Παρόλα αυτά, τα περισσότερα ατυχήματα προκλήθηκαν από τα επιβατικά ιδιωτικής χρήσης που προκάλεσαν 3607 ζημιές, ακολούθησαν τα φορτηγά ιδιωτικής χρήσης με 73 ζημιές, τα αγροτικά με 52 κοκ. Ενδεικτικά είναι και τα παρακάτω γραφήματα για την καλύτερη κατανόηση συμπερασμάτων αυτής της παραγράφου.

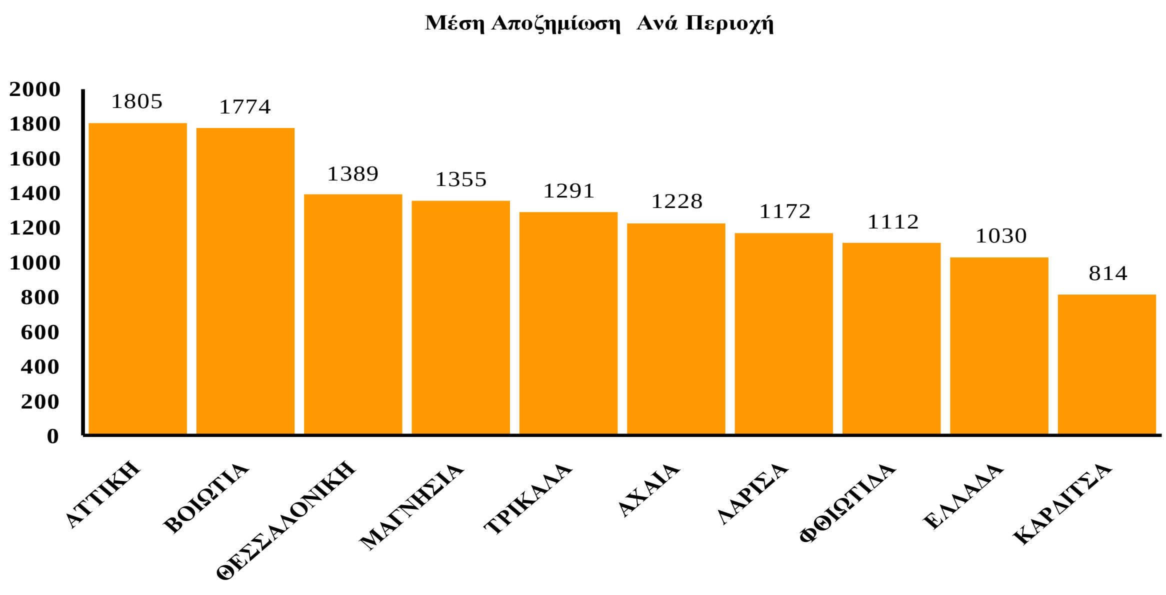
Γραφήματα 6.8. Περιγραφικά Μέτρα για τις Ζημιές Ανά Κωδικό Χρήσης του Οχήματος





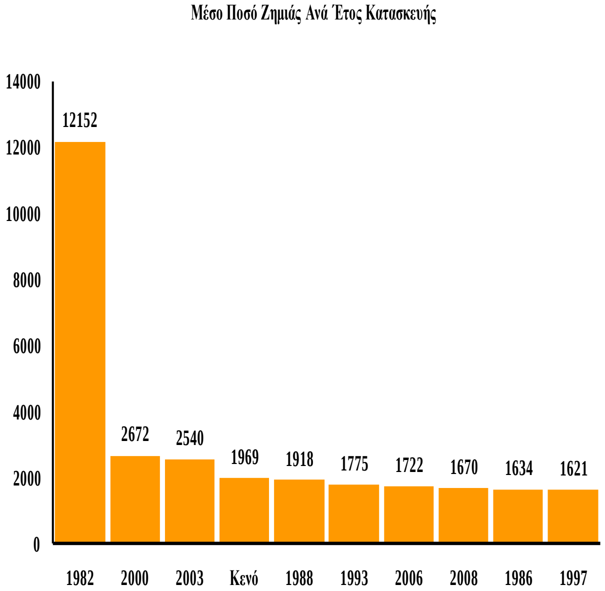
Στην περίπτωση της ταξινόμησης των ζημιών που πραγματοποιήσαμε ανά περιοχή παρατηρήσαμε ότι η περιοχή που παρουσιάζει την μεγαλύτερη μέση ζημιά, είναι εκείνη της ευρύτερης περιοχής της Αττικής με το ποσό να αγγίζει εκείνο των € 1.805 περίπου. Ακολουθεί η περιοχή της Βοιωτίας με μέση τιμή του ποσού των € 1.774 και η περιοχή που παρουσιάζει την πιο χαμηλή μέση ζημιά, σε σύγκριση με τις 10 μεγαλύτερες μέσες ζημιές, είναι εκείνη της Καρδίτσας με ποσό που αγγίζει το ποσό των € 814 περίπου. Ακολουθεί ενδεικτικό διάγραμμα:

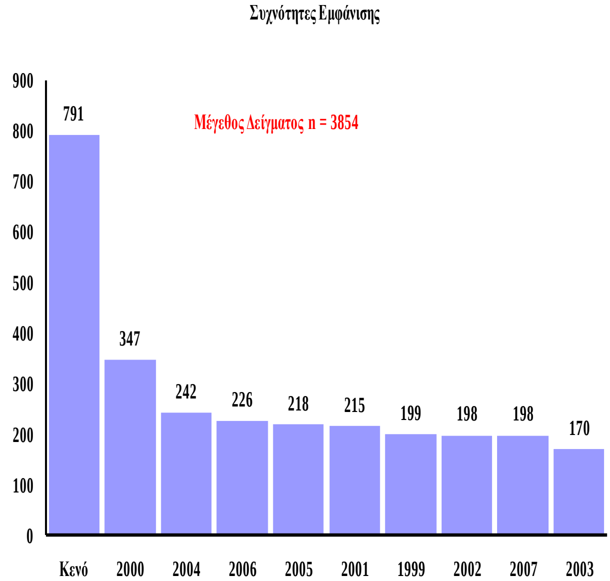
Γράφημα 6.9. Μέσα ποσά ζημιών ανά περιοχή



Πολύ σημαντικά αποτελέσματα προέκυψαν επίσης και από την ταξινόμηση των ζημιών ανάλογα με έτος κατασκευής του οχήματος. Παρατηρήσαμε λοιπόν, ότι στις 10 πιο ζημιογόνες χρονολογίες, τα οχήματα με έτος κατασκευής το 1982 προκάλεσαν ζημιά που άγγιξε περίπου το ποσό των € 12.152 (αποτελεί όμως 1 περίπτωση γι αυτό δεν την ορίσαμε ως μέση ζημιά). Στην αμέσως επόμενη ζημιογόνα χρονολογία οχήματος συμπεριλαμβάνεται το έτος 2000 (με 347 ατυχήματα) με μέσο ποσό ζημιάς που άγγιξε το ποσό των € 2.672 και καταλήγουμε στη δεκάδα των πιο ζημιογόνων χρονολογιών, στο έτος 1997 με μέση ζημιά του ποσού των € 1.621 (με 121 ατυχήματα).

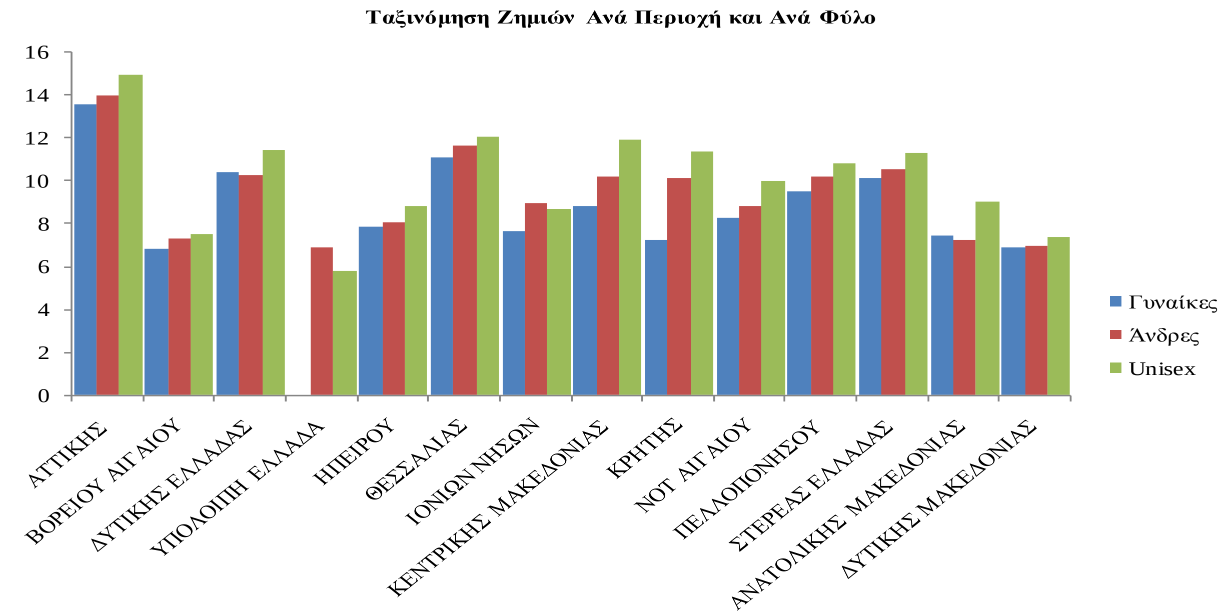
Γράφημα 6.10. Ζημιές ανά χρονολογία κατασκευής του οχήματος





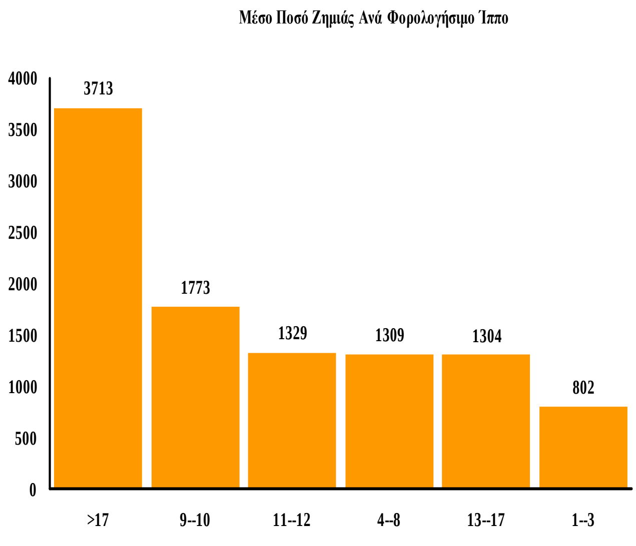
Στην ταξινόμηση που πραγματοποιήθηκε ανά περιοχή και ανά φύλο παρατηρήσαμε ότι πάλι η ευρύτερη περιοχή της Αττικής παρουσίασε τα μεγαλύτερα ποσά ζημιών (σύνολο € 5.063.000 περίπου) με τους άντρες να σημειώνουν συνολικά ζημιές ύψους € 1.201.000 περίπου, τις γυναίκες € 777.000 περίπου και οι δυο οδηγοί ανά όχημα € 3.086.000 περίπου. Αμέσως μετά ακολουθεί η περιοχή της Θεσσαλίας με συνολικές ζημιές € 360.510 περίπου, με τους άντρες να προκαλούν τις € 114.292 και οι γυναίκες τις € 68.000 περίπου. Η περιοχή με τις μικρότερες ζημιές είναι εκείνη της υπόλοιπης Ελλάδας με ζημιές συνολικού ύψους και από τους δυο οδηγούς (άντρες και γυναίκες) του ποσού των € 1.322 περίπου. Ακολουθεί το γράφημα που περιγράφει τα αποτελέσματα της παρούσας παραγράφου (σημείωση: για την καλύτερη απεικόνιση του γραφήματος, λογαριθμίσαμε τα ποσά των απαιτήσεων ανά περιοχή και ανά φύλο διότι υπήρχαν πολύ μεγάλες διαφορές ανάμεσα στις περιοχές που εξετάσαμε).

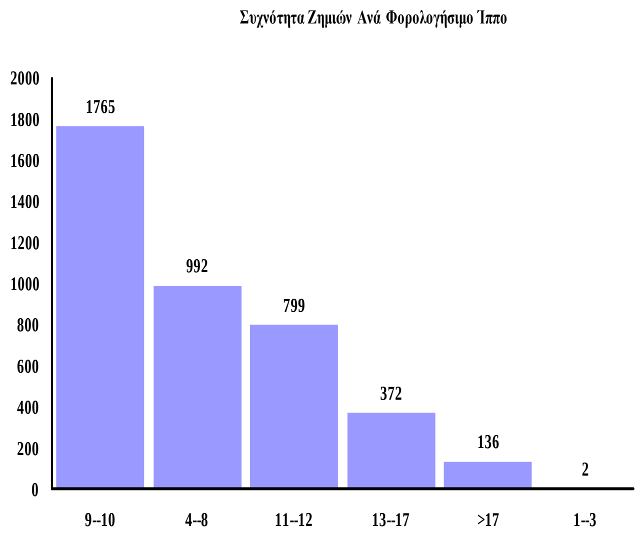
Γράφημα 6.11 Ζημιές ανά περιοχή και ανά φύλο



Κλείνοντας την παράγραφο των περιγραφικών στατιστικών για το μέγεθος της ζημιάς, ασχοληθήκαμε με τις ζημιές που προκλήθηκαν ανά φορολογήσιμο ίππο. Κωδικοποιήσαμε τους φορολογήσιμους ίππους χωρίζοντας τους σε έξι κατηγορίες ως εξής: από 1 έως 3 φορολογήσιμούς ίππους, από 4 έως 8, από 9 έως 10, από 11 έως 12, από 13 έως 17 και από 17 και πάνω. Τα αποτελέσματα που προέκυψαν παραθέτονται στα παρακάτω γραφήματα.

Γράφημα 6.11 Ζημιές ανά φορολογήσιμο ίππο





Παρατηρούμε συνεπώς ότι η ομάδα των φορολογήσιμων ίππων που παρουσίασε το μεγαλύτερο μέσο ποσό ζημιάς ήταν εκείνη των 17 ίππων και άνω με ποσό που άγγιξε τα € 3.713 σε 372 προκληθέντα ατυχήματα. Αμέσως μετά ακολούθησε η ομάδα των 9 έως 10 ίππων με μέση ζημιά ύψους € 1.773 περίπου σε 1.765 ατυχήματα και τέλος, η ομάδα με το μικρότερο μέσο ποσό ζημιάς ήταν εκείνη από 1 έως 3 ίππους με μέσο ποσό ζημιάς τα € 802 σε σύνολο 2 ατυχημάτων.

6.4 Γενικευμένο Γραμμικό Μοντέλο της Γάμμα Κατανομής για τις Ζημιές

Χρησιμοποιώντας το Γενικευμένο Γραμμικό Μοντέλο της κατανομής Γάμμα για να περιγράψουμε την συμπεριφορά των ζημιών στο χαρτοφυλάκιο των εμπειρικών δεδομένων μας, επιλέξαμε το σύνολο των 4.066 περιπτώσεων που κατέγραψαν ζημιογόνο γεγονός. Σε αυτό το σύνολο περιπτώσεων οι 528 περιπτώσεις ζημιάς προκλήθηκαν από γυναίκες οδηγούς, οι 853 από άντρες, οι 2.473 από οχήματα που οδηγήθηκαν από άντρες και γυναίκες οδηγούς ενώ 212 περιπτώσεις δεν κατέγραψαν το φύλο του οδηγού. Η μέγιστη ζημιά που προκλήθηκε από όχημα του εξεταζόμενου χαρτοφυλακίου άγγιξε το ποσό των € 418.650,00 € ενώ η μέση ζημιά άγγιξε το ποσό των € 1.593 περίπου με μία απόκλιση της τάξεως του ποσού των € 10.119 περίπου. Όσον αφορά τώρα τους φορολογήσιμους ίππους οι μέγιστοι ίπποι του χαρτοφυλακίου έφτασαν τους 140 με μέση τιμή των 10 ίππων και απόκλιση της τάξεως των 5 ίππων περίπου.

Επιπλέον, για το μοντέλο, υποθέσαμε ως εξαρτημένη μεταβλητή εκείνη των ζημιών και ανεξάρτητες τις μεταβλητές του Φύλου και των Φορολογήσιμων Ίππων. Ο έλεγχος καλής προσαρμογής του μοντέλου μας έδωσε τιμή του κριτηρίου Akaike εκείνη των 67.710 με 4.061 βαθμούς ελευθερίας. Επίσης το Omnibus τεστ έδωσε αποτέλεσμα p – value σχεδόν ίση με το μηδέν, γεγονός που σημαίνει ότι τουλάχιστον μια από τις ανεξάρτητες μεταβλητές επηρεάζουν την εξαρτημένη μεταβλητή των ζημιών. Τα αποτελέσματα αυτά επιβεβαιώνονται και από το τεστ για την επίδραση του μοντέλου (το λεγόμενο τεστ Τύπου ΙΙΙ) όπου τα p-value των ανεξάρτητων μεταβλητών είναι περίπου ίσα με το μηδέν (βλ., σχετικό πίνακα).

Πίνακας 6.10. Τεστ Τύπου ΙΙΙ για την επίδραση στο Μοντέλο

Πηγή	Τύπου ΙΙΙ		
	Wald χ^2	β.ε.	p-value
(Σταθερά)	46.238,39	1	0.000
Φύλο	44,115	3	**0.000**
Φορολογήσιμοι Ίπποι	49,568	1	**0.000**

Εξαρτημένη Μεταβλητή: Ζημιές

Ανεξάρτητες: (Σταθερά), Φύλο, Φορολογήσιμοι Ίπποι

Πίνακας 6.11. Εκτιμήσεις Παραμέτρων Παλινδρόμησης Γάμμα Κατανομής

Παράμετροι	B	τ.σ.	95% Wald δ.ε.		Τεστ Υποθέσεων		
			Κάτω Φράγμα	Άνω Φράγμα	Wald χ^2	β.ε.	p-value
(Σταθερά)	7,586	0,0333	7,521	7,652	51790,31	1	0.000
Καμία Καταχώρηση	-0,308	0,0836	-0,472	-0,145	13,625	1	**0.000**
Γυναίκες	0,048	0,0561	-0,062	0,158	0,727	1	0,394
Άντρες	-0,254	0,0464	-0,345	-0,163	29,972	1	**0.000**
Φορολογήσιμοι Ίπποι	-0,016	0,0023	-0,02	-0,012	49,568	1	**0.000**

Εξαρτημένη Μεταβλητή: Ζημιές

Ανεξάρτητες: (Σταθερά), Φύλο, Φορολογήσιμοι Ίπποι

τ.σ: Τυπική Σφάλμα, δ.ε.: Διαστήματα Εμπιστοσύνης, β.ε.: Βαθμοί Ελευθερίας

Σύμφωνα λοιπόν με τον Πίνακα 6.11 παρατηρούμε ότι εκείνοι οι οδηγοί οι οποίοι δεν δήλωσαν το φύλο τους (unisex), μαζί με τους άντρες οδηγούς και τους φορολογήσιμους ίππους φαίνονται να επηρεάζουν την εξαρτημένη μεταβλητή των ζημιών (με p-values ίσα με το μηδέν). Τελείως διαφορετικά αποτελέσματα όμως θα προκύψουν στο μοντέλο παλινδρόμησης χρησιμοποιώντας την Γάμμα κατανομή, όταν αφαιρεθούν από τον υπολογισμό των αποτελεσμάτων τέσσερις μεγάλες ζημίες (το σύνολο των ζημιών πέφτει τώρα στις 4.062). Πρόκειται για τις ζημιές με ποσά € 416.000, € 302.149, € 301.500 και € 201.500 αντίστοιχα που αφορούσαν καλύψεις για σωματικές βλάβες. Πιο συγκεκριμένα, ενδεικτικοί είναι οι παρακάτω πίνακες

Πίνακας 6.12. Τεστ Τύπου ΙΙΙ για την επίδραση στο Μοντέλο

Πηγή	Τύπου ΙΙΙ		
	Wald χ^2	β.ε.	p-value
(Σταθερά)	38857,97	1	0.000
Φύλο	4,149	3	0,246
Φορολογήσιμοι Ίπποι	2,446	1	0,118

Εξαρτημένη Μεταβλητή: Ζημιές

Ανεξάρτητες: (Σταθερά), Φύλο, Φορολογήσιμοι Ίπποι

Πίνακας 6.13. Εκτιμήσεις Παραμέτρων Παλινδρόμησης Γάμμα Κατανομής

Παράμετροι	Β	τ.σ.	95% Wald δ.ε.		Τεστ Υποθέσεων		
			Κάτω Φράγμα	Άνω Φράγμα	Wald χ^2	β.ε.	p-value
(Σταθερά)	7,194	0,0366	7,123	7,266	38573,31	1	0.000
Καμία Καταχώρηση	-0,031	0,0744	-0,177	0,114	0,177	1	0,674
Γυναίκες	0,096	0,0501	-0,002	0,194	3,678	1	0,055
Άντρες	0,022	0,0413	-0,059	0,103	0,285	1	0,593
Φορολογήσιμοι Ίπποι	-0,004	0,0029	-0,01	0,001	2,446	1	0,118

Εξαρτημένη Μεταβλητή: Ζημιές

Ανεξάρτητες: (Σταθερά), Φύλο, Φορολογήσιμοι Ίπποι

τ.σ: Τυπική Σφάλμα, δ.ε.: Διαστήματα Εμπιστοσύνης, β.ε.: Βαθμοί Ελευθερίας

Το Omnibus τεστ έδωσε αποτέλεσμα p-value ίσο με 0.131 γεγονός που σημαίνει πως καμία από τις ανεξάρτητες μεταβλητές του συγκεκριμένου μοντέλου δεν φαίνεται να επηρεάζει την εξαρτημένη μεταβλητή των ζημιών. Πράγματι, τα αποτελέσματα των πινάκων 6.12 και 6.13 επιβεβαιώνουν το παραπάνω τεστ με αποτελέσματα p-values μεγαλύτερα από το επίπεδο σημαντικότητας που έχουμε ορίσει για την έρευνά μας και δεν είναι άλλο από το α = 0.05 ή 5%.

6.5 Συμπεράσματα

Στην περίπτωση του αριθμού των απαιτήσεων το γενικευμένο μοντέλο της κατανομής Poisson μας έδωσε τα εξής αποτελέσματα: οι εκτιμήσεις των $b_0, b_1, b_2, ..., b_{10}$ δόθηκαν από τον Πίνακα 6.6 (-3.357, 0.164, …,-0.67 αντίστοιχα). Παρατηρούμε συνεπώς, ότι για τις γυναίκες ο συντελεστής της παλινδρόμησης έδωσε την τιμή $b_1 = 0.164$. Στην περίπτωση μιας γυναίκας οδηγού λοιπόν, ο μέσος αριθμός ατυχημάτων θα είναι ίσος με¨

$$e^{b_0+b_1} = e^{-3.357+0.164} = 0.041 = 4{,}1\%$$

δηλαδή, αν έχει δηλωθεί το φύλο του οδηγού και είναι γυναίκα, θα καταλήξουμε στο συμπέρασμα ότι θα έχουμε μέσο αριθμό ατυχημάτων ίσο με 4,1% περίπου. Στην περίπτωση που δεν έχει δηλωθεί το φύλο του οδηγού (unisex), ο μέσος αριθμός ατυχημάτων θα υπολογιστεί από την σχέση:

$$e^{b_0} = e^{-3.357} = 0.034 = 3{,}4\%$$

Αντίστοιχα παράγεται και η ερμηνεία των υπόλοιπων συντελεστών. Το μοντέλο παλινδρόμησης που προκύπτει και που μπορεί να χρησιμοποιηθεί και για προβλέψεις είναι το ακόλουθο:

$$\log(\text{Α}\textit{ριθμός} \quad \text{Α}\textit{τυχημάτων}) = b_0 + b_1(\Phi\textit{ύλο} = \Gamma) + b_2(\Phi\textit{ύλο} = \text{Α}) + ... + b_{10}(\Delta\textit{ιάρκεια} = 2)$$

$$\log(\text{Α}\textit{ριθμός} \quad \text{Α}\textit{τυχημάτων}) = -3.36 + 0.17(\Phi\textit{ύλο} = \Gamma) - 0.03(\Phi\textit{ύλο} = \text{Α}) + ... - 0.67(\Delta\textit{ιάρκεια} = 2)$$

Ακολούθησε το γενικευμένο γραμμικό μοντέλο της Αρνητικής Διωνυμικής κατανομής, που όπως παρατηρήσαμε παραπάνω από τον Πίνακα 6.9 έδωσε τα εξής αποτελέσματα: οι εκτιμήσεις των συντελεστών $b_0, b_1, b_2, ..., b_{10}$ είναι οι -3.36, 0.16, …, -0.65 αντίστοιχα. Παρατηρούμε συνεπώς πως παρόμοια αποτελέσματα με το μοντέλο της κατανομής Poisson προκύπτουν στην περίπτωση της αρνητικής διωνυμικής, με μέσο αριθμό ατυχημάτων για μια γυναίκα οδηγό να αγγίζει το 4,1% περίπου. Το μοντέλο παλινδρόμησης που προκύπτει είναι πάλι το:

$$\log(\text{Α}\textit{ριθμός} \quad \text{Α}\textit{τυχημάτων}) = b_0 + b_1(\Phi\textit{ύλο} = \Gamma) + b_2(\Phi\textit{ύλο} = \text{Α}) + ... + b_{10}(\Delta\textit{ιάρκεια} = 2)$$

Τέλος, όσον αφορά της ζημιές, χρησιμοποιήσαμε το γενικευμένο γραμμικό μοντέλο της Γάμμα κατανομής και καταλήξαμε σε ένα μοντέλο όπου στατιστικά σημαντικές θεωρήθηκαν οι ανεξάρτητες μεταβλητές των οδηγών χωρίς διαχωρισμό, των αντρών οδηγών και των φορολογήσιμων ίππων. Πιο συγκεκριμένα, οι τιμές των συντελεστών $b_0, b_1, ..., b_4$ που δίνονται από τον Πίνακα 6.11, είναι οι 7.586, -0.308, 0.048, -0.254, 0 και -0.016 αντίστοιχα. Παρατηρούμε ότι μόνο όταν μεταβάλλεται ο αριθμός των γυναικών οδηγών μεταβάλλονται τα ποσά των ζημιών. Συνεπώς, παρατηρούμε ότι οι γυναίκες οδηγοί έχουν μέσο ύψος ζημιάς ίσο με:

$$e^{b_0+b_2} = e^{7.586+0.048} = 2067€$$

ενώ, στην περίπτωση που δεν καταχωρηθεί το φύλο του οδηγού (unisex), θα έχουμε το αποτέλεσμα:

$$e^{b_0} = e^{7.586} = 1970€$$

Τέλος, οι άντρες θα έχουν μέσο ύψος ζημιάς ίσο με:

$$e^{b_0+b_2} = e^{7.586-0.254} = 1528€$$

Καταλήγουμε συνεπώς στο συμπέρασμα ότι οι άντρες οδηγοί έχουν μικρότερο αναμενόμενο ύψος ζημιάς από τις γυναίκες οδηγούς και τους unisex.

Στην συνέχεια, χρησιμοποιήσαμε το γενικευμένο γραμμικό μοντέλο της Γάμμα κατανομής στο χαρτοφυλάκιο μας, εξαιρώντας όμως τις τέσσερις μεγάλες ζημιές που αφορούν θανατηφόρα ατυχήματα. Στην περίπτωση αυτή το μοντέλο δεν κρίθηκε αξιόπιστο, διότι καμία από τις ανεξάρτητες μεταβλητές δεν φαίνεται να επηρεάζει την εξαρτημένη μεταβλητή του ύψους των ζημιών. Συνεπώς, δεν πραγματοποιήσαμε κάποια πρόβλεψη γι’ αυτό το μοντέλο και θα αρκεστούμε στα αποτελέσματα του πλήρους μοντέλου ζημιών.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Antzoulakos D. & Chadjiconstantinidis S., *On Mixed and Compound Poisson Distributions, Scandinavian Actuarial Journal, 2004:3, 161 – 188.*

Cox, D.R. (1970), *The Analysis of Binary Data. Chapman and Hall.*

Dobson, A. (1990), *An Introduction to Generalized Linear Models. Chapman and Hall.*

Denuit M., X. Maréchal, S. Pitrebois and J.-F. Walhin, (2007)*. Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems.*

De Jong P. & Heller G.Z (2008)*, Generalized Linear Models for Insurance Data.*

Denuit, M. (1997)*. A new distribution of Poisson-type for the number of claims. ASTIN Bulletin* ***27****, 229–242.*

Denuit, M., & Dhaene, J. (2001). *Bonus-Malus scales using exponential loss functions. German Actuarial Bulletin* ***25****, 13–27.*

Denuit, M., Dhaene, J., Goovaerts, M.J., & Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models. John Wiley & Sons, Inc., New York.*

Lord, D. (2006). *Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention 38, 751–766.*

Lemaire Jean,(1985). *Automobile Insurance Actuarial Models, Poisson Model – Homogeneous Portfolio.*

Lemaire Jean,(1985), *Automobile Insurance Actuarial Models, Negative Binominal Model – Heterogeneous Portfolio.*

Tucker, H.G. (1963). *An estimate of the compounding distribution of a compound Poisson distribution. Theory of Probability and Its Applications 8, 195–200.*

Viswanathan, K.S., & Lemaire, J. (2005). *Bonus-malus systems in a deregulated environment: Forecasting market shares using diffusion models.*

Κουτσόπουλος Κ.Ι.,(1999)*, Αναλογιστικά Μαθηματικά, Μέρος Ι: Θεωρία των Κινδύνων*