

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ
ΚΙΝΔΥΝΟΥ

ΠΡΟΣΑΡΜΟΓΗ ΚΑΤΑΝΟΜΩΝ
ΠΙΘΑΝΟΤΗΤΑΣ ΣΕ ΧΑΡΤΟΦΥΛΑΚΙΟ
ΖΗΜΙΩΝ ΑΣΤΙΚΗΣ ΕΥΘΥΝΗΣ
ΑΥΤΟΚΙΝΗΤΩΝ

Αθανάσιος Παλαιοδημόπουλος

Διπλωματική Εργασία

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

Πειραιάς

Ιούλιος 2016

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

Τα μέλη της Επιτροπής ήταν:

- Πολίτης Κωνσταντίνος (Επιβλέπων)
- Μπούτσικας Μιχαήλ
- Μπερσίμης Σωτήριος

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

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS AND INSURANCE
SCIENCE

POSTGRADUATE PROGRAM IN ACTUARIAL
SCIENCE AND RISK MANAGEMENT

**FITTING PROBABILITY
DISTRIBUTIONS TO LOSSES IN A
CAR INSURANCE PORTFOLIO**

Athanasios Palaiodimopoulos

MSc Dissertation

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

Piraeus, Greece

July 2016

Στον παππού και στη γιαγιά μου

Ευχαριστίες

Θα ήθελα να εκφράσω τις ευχαριστίες μου στον επιβλέποντα καθηγητή κο Κωσταντίνο Πολίτη για την καθοδήγηση και βοήθειά του στην εκπόνηση της εργασίας.

Επίσης ευχαριστώ την ασφαλιστική εταιρεία που μου εμπιστεύτηκε τα δεδομένα και ιδιαίτερα τον υπεύθυνο διαχείρισης κινδύνων της εταιρείας, για τον χρόνο και τις πληροφορίες που τόσο πρόθυμα και γενναιόδωρα διέθεσε κατά τη διάρκεια της συνεργασίας μας.

Θα ήταν παράλειψη να μην ευχαριστήσω την οικογένεια μου, για την υπομονή και την υποστήριξή της.

Τέλος, ένα μεγάλο ευχαριστώ στην συμφοιτήτρια και φίλη μου Ειρήνη Σταματοπούλου για την ενθάρρυνση και συμπαράστασή της σε όλη την διάρκεια των σπουδών μας.

Περίληψη

Η παρούσα διπλωματική αφορά τη μελέτη χαρτοφυλακίου ζημιών αστικής ευθύνης αυτοκινήτου, ελληνικής ασφαλιστικής εταιρείας. Αρχικά, γίνεται ο διαχωρισμός χαρτοφυλακίου σε σωματικές βλάβες και υλικές ζημιές. Στη συνέχεια, παρουσιάζεται μια σειρά θεωρητικών κατανομών και οι βασικές ιδιότητες τους. Ακολουθεί εφαρμογή των θεωρητικών αυτών κατανομών στα εμπειρικά δεδομένα με τη χρήση του στατιστικού πακέτου R. Επιπρόσθετα, παρουσιάζεται ένας εναλλακτικός τρόπος προσέγγισης για την εύρεση της κατανομής που ακολουθεί το ύψος των αποζημιώσεων, χρησιμοποιώντας την έννοια του Μέσου Υπολειπόμενου Χρόνου Ζωής (Mean Residual Life) με τη βοήθεια του υπολογιστικού πακέτου Wolfram Mathematica. Τέλος, τα αποτελέσματα της μελέτης δίνουν σημαντικές πληροφορίες που μπορεί να χρησιμοποιήσει η εταιρεία για την διαχείριση και αντιμετώπιση των κινδύνων που απορρέουν από τον κλάδο αστικής ευθύνης αυτοκινήτων.

Abstract

In this thesis, we are concerned with the study of the motor vehicle liability losses in a portfolio of a Greek insurance company. We begin by separating the portfolio based on two criteria, the bodily injuries and the property damages. Afterwards, we briefly review a series of theoretical distributions and their basic properties. We apply these theoretical distributions to the empirical data by using the R statistics package. Furthermore, we present an alternative approach for finding the distribution that follows the amount of compensations, by using the concept of the Mean Residual Life, with the help of Wolfram Mathematica algebraic package. Our results provide important information that can be used by the company in order to manage and deal with the dangers that stem from the motor vehicle liability sector.

Περιεχόμενα

ΕΥΧΑΡΙΣΤΙΕΣ.....	III
ΠΕΡΙΛΗΨΗ.....	V
ABSTRACT.....	VII
ΚΕΦΑΛΑΙΟ 1.....	1
ΕΙΣΑΓΩΓΗ.....	1
ΚΕΦΑΛΑΙΟ 2.....	3
ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ.....	3
2.1 Ανάλυση του πλήθους των ζημιών.....	3
2.2 Ανάλυση του ύψους των ζημιών.....	5
2.3 Περιγραφική Στατιστική Ανάλυση Χαρτοφυλακίου.....	8
2.3.1 Εβδομαδιαία Ανάλυση.....	8
2.3.2 Μηνιαία Ανάλυση.....	11
ΚΕΦΑΛΑΙΟ 3.....	15
ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΑΣ ΓΙΑ ΤΟ ΠΛΗΘΟΣ ΚΑΙ ΤΟ ΥΨΟΣ ΤΩΝ ΖΗΜΙΩΝ.....	15
3.1 Διωνυμική κατανομή.....	15
3.2 Κατανομή Poisson.....	16
3.3 Αρνητική Διωνυμική κατανομή.....	17
3.4 Κατανομή Γάμμα.....	18
3.5 Κανονική Κατανομή.....	20
3.6 Λογαριθμοκανονική Κατανομή.....	21
3.7 Σχέση Λογαριθμοκανονικής και Τοποποιημένης κανονικής κατανομής.....	21
ΚΕΦΑΛΑΙΟ 4.....	23
ΕΥΡΕΣΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΤΟΥ ΥΨΟΥΣ ΚΑΙ ΤΟΥ ΠΛΗΘΟΥΣ ΤΩΝ ΑΠΟΖΗΜΙΩΣΕΩΝ.....	23
4.1 Εβδομαδιαία Ανάλυση.....	23
4.1.1 Ύψος Αποζημιώσεων.....	23
4.1.2 Πλήθος Ζημιών.....	28
4.2 Μηνιαία Ανάλυση.....	29
4.2.1 Ύψος Αποζημιώσεων.....	29
4.2.2 Πλήθος ζημιών.....	36
ΚΕΦΑΛΑΙΟ 5.....	37
ΑΝΑΛΥΣΗ ΜΕ ΒΑΣΗ ΤΟ ΜΕΣΟ ΥΠΟΛΕΙΠΟΜΕΝΟ ΧΡΟΝΟ ΖΩΗΣ (MEAN RESIDUAL LIFE).....	37
5.1 Προϋποθέσεις – Βασικές Ιδιότητες.....	38
5.2 Μελέτη Κατανομών Γάμμα.....	38
5.2.1 Κατανομή Γάμμα.....	38
5.2.2 Μίξη κατανομών Γάμμα.....	40
5.3 Μέση υπερβάλλονσα ζημιά – Εφαρμογή στα δεδομένα.....	44
5.3.1 Εβδομαδιαίο ύψος αποζημιώσεων.....	44
5.3.2 Μηνιαίο ύψος αποζημιώσεων.....	46
5.4 Ατομικά ύψη ζημιών.....	49
ΚΕΦΑΛΑΙΟ 6.....	50
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	50
ΠΑΡΑΡΤΗΜΑ Ι.....	51
ΠΑΡΑΡΤΗΜΑ ΙΙ.....	53
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	60

Πίνακας Γραφημάτων

Γράφημα 2-1: Θηκόγραμμα και Ιστόγραμμα των εβδομαδιαίων πληρωμών.....	9
Γράφημα 2-2: Διάγραμμα χρονολογικής ανάλυσης των εβδομαδιαίων πληρωμών	9
Γράφημα 2-3: Θηκόγραμμα και Ιστόγραμμα του εβδομαδιαίου πλήθους ζημιών.....	10
Γράφημα 2-4: Διάγραμμα χρονολογικής ανάλυσης εβδομαδιαίου πλήθους ζημιών.....	10
Γράφημα 2-5: Θηκόγραμμα και Ιστόγραμμα των μηνιαίων πληρωμών	12
Γράφημα 2-6: Διάγραμμα χρονολογικής ανάλυσης των μηνιαίων πληρωμών.....	12
Γράφημα 2-7: Θηκόγραμμα και Ιστόγραμμα του μηνιαίου πλήθους ζημιών	13
Γράφημα 2-8: Διάγραμμα χρονολογικής ανάλυσης του μηνιαίου πλήθους των ζημιών.....	13
Γράφημα 3-1: Συνάρτηση πυκνότητας πιθανότητας κατανομής Γάμμα για διάφορα β	19
Γράφημα 4-1: Προσαρμογή κατανομών στο ύψος των εβδομαδιαίων ζημιών	24
Γράφημα 4-2: Σύγκριση της συνάρτησης κατανομής των κατανομών Γάμμα και Κανονικής με αυτή των εμπειρικών δεδομένων	25
Γράφημα 4-3: Συγκριτικό ιστόγραμμα μηνιαίων ζημιών με την κανονική κατανομή	29
Γράφημα 4-4: Συγκριτικό διάγραμμα κατανομών.....	30
Γράφημα 4-5: Εφαρμογή δύο διαφορετικών κατανομών Γάμμα στα εμπειρικά δεδομένα.....	31
Γράφημα 4-6: Συνάρτηση πυκνότητας της μίξης Γάμμα κατανομών, όπου οι παράμετροι υπολογίστηκαν με τη μέθοδο της μεγίστης πιθανοφάνειας.....	33
Γράφημα 4-7: Συνάρτηση πυκνότητας της μίξης Γάμμα κατανομών, όπου οι παράμετροι υπολογίστηκαν με τη μέθοδο των ροπών.....	34
Γράφημα 4-8: Σύγκριση της εμπειρικής συνάρτησης κατανομής με των θεωρητικών.....	35
Γράφημα 4-9: Γραφικός έλεγχος καλής προσαρμογής της κανονικής κατανομής στο μηνιαίο πλήθος των ζημιών.....	36
Γράφημα 5-1: Γραφική απεικόνιση της συνάρτησης επιβίωσης, της έντασης κινδύνου και του μέσου υπολειπόμενου χρόνου ζωής για την κατανομή Γάμμα.....	39
Γράφημα 5-2: Μίξη κατανομών Γάμμα με παραμέτρους $a_1 = 3, a_2 = 9, \beta = 5$ και $w_1 = w_2 = 0.5$	41
Γράφημα 5-3: Μίξη κατανομών Γάμμα με παραμέτρους $a_1 = 0.5, a_2 = 0.2, \beta = 5$ και $w_1 = w_2 = 0.5$..	42
Γράφημα 5-4: Μίξη κατανομών Γάμμα με παραμέτρους $a_1 = 0.5, a_2 = 9, \beta = 5$ και $w_1 = w_2 = 0.5$	43
Γράφημα 5-5: Συνάρτηση μέσου υπολειπόμενου χρόνου ζωής εβδομαδιαίων αποζημιώσεων.....	44
Γράφημα 5-6: Σύγκριση εμπειρικού και θεωρητικού μέσου υπολειπόμενου χρόνου ζωής των εβδομαδιαίων αποζημιώσεων	45
Γράφημα 5-7: Συνάρτηση μέσου υπολειπόμενου χρόνου ζωής των μηνιαίων ζημιών	46
Γράφημα 5-8: Διαφορές μεταξύ του εμπειρικού mrl και του mrl που προκύπτει με τη μέθοδο μεγίστης πιθανοφάνειας και με τη μέθοδο των ροπών.....	47
Γράφημα 5-9: Σύγκριση της συνάρτησης mrl με τους δύο διαφορετικούς τρόπους εκτίμησης των παραμέτρων.....	47

Πίνακας Πινάκων

Πίνακας 2-1: Πλήθος ζημιών ανά έτος αναγγελίας.....	3
Πίνακας 2-2: Πλήθος υλικών ζημιών ανά έτος αναγγελίας	4
Πίνακας 2-3: Πλήθος ζημιών που αφορούν σωματικές βλάβες ανά έτος αναγγελίας	4
Πίνακας 2-4: Τρίγωνο εξέλιξης πληρωθεισών ζημιών	5
Πίνακας 2-5: Τρίγωνο εξέλιξης πληρωθεισών υλικών ζημιών	6
Πίνακας 2-6: Τρίγωνο εξέλιξης πληρωθεισών σωματικών βλαβών.....	6
Πίνακας 2-7: Τρίγωνο με το ποσοστό των κλειστών υλικών ζημιών επί του συνόλου των υλικών ζημιών ανά έτος αναγγελίας	7
Πίνακας 2-8: Περιγραφικά στατιστικά στοιχεία για το ύψος και το πλήθος των ζημιών που αναγγέλλονται ανά εβδομάδα	8
Πίνακας 2-9: Περιγραφικά στατιστικά στοιχεία για το ύψος και το πλήθος των ζημιών που αναγγέλλονται ανά εβδομάδα	11
Πίνακας 3-1: Μέτρα θέσης και διασποράς της κανονικής κατανομής	20
Πίνακας 4-1: Εκτίμηση των παραμέτρων της κανονικής και της Γάμμα κατανομής	24
Πίνακας 4-2: p-values του Kolmogorov –Smirnov test	25
Πίνακας 4-3: Πλήθος θεωρητικών και πραγματικών τιμών σε κάθε κλάση	26
Πίνακας 4-4: Πίνακας υπολογισμών για τον έλεγχο χ^2	26
Πίνακας 4-5: Πίνακας αποτελεσμάτων για τον έλεγχο χ^2	26
Πίνακας 4-6: Εκτιμώμενες παράμετροι της Κανονικής κατανομής.....	29
Πίνακας 4-7: Έλεγχος κανονικότητας	30
Πίνακας 4-8: Εκτίμηση παραμέτρων με τη μέθοδο μεγίστης πιθανοφάνειας.....	31
Πίνακας 4-9: Έλεγχος Kolmogorov – Smirnov.....	35
Πίνακας 4-10: Εκτίμηση παραμέτρων της κανονικής κατανομής	36
Πίνακας 4-11: Έλεγχος κανονικότητας για το μηνιαίο πλήθος των αποζημιώσεων	36

Κεφάλαιο 1

Εισαγωγή

Τα τελευταία χρόνια, στην Ελλάδα, οι ασφαλιστικές εταιρείες γενικών ασφαλίσεων απαιτείται να κάνουν χρήση σύνθετων μοντέλων και επιπρόσθετες μελέτες για την εκτίμηση των αποθεμάτων τους, πολύ μεγαλύτερης ανάλυσης από το παρελθόν. Η απαίτηση αυτή προκύπτει ως αποτέλεσμα των παρακάτω παραγόντων:

- Η τεχνολογική ανάπτυξη παρέχει τη δυνατότητα οργανωμένης αποθήκευσης και γρήγορης επεξεργασίας μεγάλου όγκου δεδομένων.
- Η οικονομική ύφεση και ο κλιμακούμενος ανταγωνισμός, επιβάλλει την όσο το δυνατόν πιο ακριβή εκτίμηση της επέλευσης του κινδύνου, με σκοπό την επίτευξη του ιδανικού συνδυασμού της ανταγωνιστικότητας των τιμών με την κερδοφορία της εταιρείας.
- Η συμμόρφωση της ελληνικής νομοθεσίας με την Οδηγία 2009/138/EK του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 25ης Νοεμβρίου 2009 (Φερεγγυότητα II).

Η φύση και ο ρόλος των ασφαλιστικών εταιρειών είναι να λαμβάνουν κινδύνους έναντι ασφαλιστρού. Για τον καλύτερο έλεγχο, τη μελέτη και τη διαχείριση των κινδύνων, οι εταιρείες κατηγοριοποιούν τους κινδύνους σε ομοιογενείς ομάδες. Τα κριτήρια βάσει των οποίων ομαδοποιούνται τα χαρτοφυλάκια κινδύνων επιλέγονται έτσι ώστε να εξασφαλίσουν τη μεγαλύτερη δυνατή ομοιογένεια.

Στις γενικές ασφαλίσεις η ομαδοποίηση γίνεται κατά κλάδο ασφάλισης και ειδικότερα για τον κλάδο αστικής ευθύνης αυτοκινήτων, ο αναλογιστής κατά κανόνα διαχωρίζει το χαρτοφυλάκιο του περαιτέρω σε επιμέρους ομάδες, προκειμένου να επιτευχθεί μεγαλύτερη ομοιογένεια. Το πιο σύνηθες είναι να γίνεται διαχωρισμός των ζημιών που αφορούν σωματικές βλάβες από εκείνες που αφορούν υλικές ζημιές. Αυτό γίνεται διότι η συμπεριφορά των δύο αυτών υποομάδων είναι εντελώς διαφορετική. Συγκεκριμένα, πέραν του ότι τα ποσά των τελικών αποζημιώσεων των ζημιών που αφορούν σωματικές βλάβες είναι τις περισσότερες φορές πολύ μεγαλύτερα από εκείνα που αφορούν υλικές ζημιές, ο χρόνος που απαιτείται για να 'κλείσει' μια ζημιά με σωματικές βλάβες είναι επίσης κατά πολύ μεγαλύτερος από τον αντίστοιχο των υλικών ζημιών. Ως εκ τούτου, ο αναλογιστής κατά τη διαδικασία σχηματισμού τεχνικών αποθεμάτων, τα οποία επηρεάζονται από το ποσό και την εξέλιξη των αποζημιώσεων, με την ομαδοποίηση αυτήν επιτυγχάνει καλύτερες εκτιμήσεις.

Στην εργασία αυτή, τα δεδομένα τα οποία χρησιμοποιούνται είναι πραγματικά δεδομένα ελληνικής ασφαλιστικής εταιρείας. Καταρχήν αποδεικνύεται η σημαντικότητα του διαχωρισμού των δεδομένων ανάλογα με το είδος κάλυψης και στη συνέχεια μελετάται η προσαρμογή των δεδομένων σε διάφορες κατανομές, με σκοπό την προσεκτική μελέτη του κινδύνου και το σχηματισμό των κατάλληλων αποθεμάτων προς διασφάλιση των συμφερόντων τόσο των ασφαλισμένων όσο και της εταιρείας.

Στη βιβλιογραφία αναφέρονται κατανομές ή οικογένειες κατανομών οι οποίες προσαρμόζονται ικανοποιητικά σε δεδομένα από ζημιές αυτοκινήτων. Κάποιες εξηγούν καλύτερα τη συμπεριφορά του πλήθους των ζημιών του χαρτοφυλακίου και κάποιες τα ποσά αυτών. Στο τρίτο κεφάλαιο της εργασίας, επιλέγονται κατανομές οι οποίες έχουν εφαρμογή στα δεδομένα και παρουσιάζονται οι βασικές ιδιότητες αυτών.

Στη συνέχεια, με τη χρήση του στατιστικού πακέτου R, γίνονται έλεγχοι καλής προσαρμογής των εμπειρικών δεδομένων στο σύνολο των θεωρητικών κατανομών που παρουσιάστηκαν στην προηγούμενη ενότητα, με σκοπό την εκτίμηση της κατανομής την οποία ακολουθούν αυτά.

Στο 5^ο κεφάλαιο, παρουσιάζεται ένας εναλλακτικός τρόπος προσέγγισης για την εύρεση της κατανομής που ακολουθεί το ύψος των αποζημιώσεων, χρησιμοποιώντας την έννοια του Μέσου Υπολειπόμενου Χρόνου Ζωής (MRL). Συγκεκριμένα, παρουσιάζεται η μορφή της συνάρτησης MRL για την κατανομή Γάμμα και τη μίξη κατανομών Γάμμα και κατασκευάζονται τα γραφήματά τους για διαφορετικές τιμές των παραμέτρων τους. Στη συνέχεια κατασκευάζεται ο αλγόριθμος για τον υπολογισμό της συνάρτησης του μέσου υπολειπόμενου χρόνου ζωής (MRL) των εμπειρικών δεδομένων, με τη βοήθεια του υπολογιστικού προγράμματος Wolfram Mathematica και συγκρίνονται γραφικά τα αποτελέσματα.

Τέλος αναλύονται τα αποτελέσματα της μελέτης και τα στοιχεία αυτά τα οποία μπορούν να συνεισφέρουν στις ανάγκες μιας ασφαλιστικής εταιρείας.

Στα παραρτήματα της εργασίας παρατίθενται αναλυτικά οι εντολές που εισήχθησαν στο στατιστικό πακέτο R και στο μαθηματικό πακέτο Wolfram Mathematica.

Κεφάλαιο 2

Περιγραφή Δεδομένων

Στο κεφάλαιο αυτό θα γίνει η περιγραφή των δεδομένων (αρχείο πληρωμών) που λάβαμε από την εταιρεία. Για το χαρτοφυλάκιο αστικής ευθύνης αυτοκινήτων που θα μελετήσουμε δίνονται τα παρακάτω στοιχεία:

- Ο αριθμός ζημιάς ή αλλιώς αριθμός φακέλου
- Η ημερομηνία αναγγελίας της ζημιάς
- Η ημερομηνία πληρωμής της ζημιάς
- Το ποσό πληρωμής
- Το είδος κάλυψης
- Ο χαρακτηρισμός της ζημιάς σε ανοιχτή ή κλειστή

Ειδικότερα σημειώνεται ότι ο αριθμός ζημιάς είναι μοναδικός, οι πληρωμές που γίνονται για ένα φάκελο μπορεί να είναι περισσότερες από μία και το είδος κάλυψης διαχωρίζει την πληρωμή, σε πληρωμή που έγινε για σωματικές βλάβες σε τρίτους και σε πληρωμή που έγινε για υλικές ζημιές σε τρίτους. Ο διαχωρισμός αυτός είναι πολύ σημαντικός διότι όπως θα δούμε και στη συνέχεια η συμπεριφορά της εξέλιξης των πληρωμών διαφέρει κατά πολύ ανάλογα με το είδος κάλυψης.

Τα έτη τα οποία συμπεριλαμβάνονται στη μελέτη είναι η πενταετία 2010 - 2014. Σαν μονάδα του χρόνου χρησιμοποιούνται η εβδομάδα, ο μήνας ή ακόμη και το έτος. Το δείγμα που θα αναλυθεί συμπεριλαμβάνει μόνο κλειστούς φακέλους, καθώς πληροφορίες για το απόθεμα των εκκρεμών ζημιών δεν έχουν δοθεί από την εταιρεία. Επομένως, το πρώτο υπολογιστικό βήμα της μελέτης αναφέρεται στο τι ποσό και τι ποσοστό των ζημιών αποκλείεται από το συνολικό χαρτοφυλάκιο και κατά πόσο η μελέτη μόνο των κλειστών φακέλων καταλήγει σε αξιόπιστα αποτελέσματα.

2.1 Ανάλυση του πλήθους των ζημιών

Ως προς το πλήθος των ζημιών, ανάλογα με το αν αυτές είναι κλειστές ή εκκρεμείς στις 31/12/2014 έχουμε τα παρακάτω στοιχεία:

Χαρτοφυλάκιο Αστικής Ευθύνης Αυτοκινήτων

Έτος Αναγγελίας	Πλήθος Ζημιών	Πλήθος Κλειστών Ζημιών	Πλήθος Εκκρεμών Ζημιών
2010	4.578	4.510	68
2011	4.877	4.747	130
2012	4.416	4.239	177
2013	4.020	3.799	221
2014	3.616	3.088	528

Πίνακας 2-1: Πλήθος ζημιών ανά έτος αναγγελίας

Όπως έχει προαναφερθεί, ο διαχωρισμός του χαρτοφυλακίου σε σωματικές βλάβες και υλικές ζημιές είναι απαραίτητος καθώς η συμπεριφορά ως προς την εξέλιξη του διακανονισμού της ζημιάς είναι απόλυτα εξαρτημένη από το είδος κάλυψης. Τα στοιχεία στους παρακάτω πίνακες παρουσιάζουν αυτή την αναγκαιότητα διαχωρισμού ανάλογα με το είδος κάλυψης.

Υλικές Ζημιές

Έτος Αναγγελίας	Πλήθος Ζημιών	Πλήθος Κλειστών Ζημιών	Πλήθος Εκκρεμών Ζημιών	% Εκκρεμείς Ζημιές
2010	4.530	4.481	49	1%
2011	4.847	4.728	119	2%
2012	4.388	4.227	161	4%
2013	3.999	3.794	205	5%
2014	3.587	3.080	507	14%

Πίνακας 2-2: Πλήθος υλικών ζημιών ανά έτος αναγγελίας

Σωματικές Βλάβες

Έτος Αναγγελίας	Πλήθος Ζημιών	Πλήθος Κλειστών Ζημιών	Πλήθος Εκκρεμών Ζημιών	% Εκκρεμείς Ζημιές
2010	105	74	31	30%
2011	88	52	36	41%
2012	91	48	43	47%
2013	50	27	23	46%
2014	29	8	21	72%

Πίνακας 2-3: Πλήθος ζημιών που αφορούν σωματικές βλάβες ανά έτος αναγγελίας

Παρατηρώντας τους παραπάνω πίνακες μπορεί κανείς να αναρωτηθεί γιατί το άθροισμα του πλήθους των υλικών ζημιών και των σωματικών βλαβών δεν ισούται με το πλήθος του συνόλου. Η απάντηση είναι απλή αλλά καθόλου προφανής. Μία ζημιά μπορεί να έχει πληρωμές που καλύπτουν τόσο σωματικές βλάβες όσο και υλικές ζημιές. Για παράδειγμα, ο παθών, για τον οποίον ανοίχτηκε ο φάκελος, μπορεί να έχει σπάσει το χέρι του αλλά παράλληλα να πρέπει να αποζημιωθεί και για το παρμπρίζ του αυτοκίνητου του. Σε αυτή την περίπτωση μιλάμε για μία ζημιά, αν μελετηθεί ως προς το σύνολο, και για δύο ζημιές, αν μελετηθεί ξεχωριστά σύμφωνα με το είδος κάλυψης.

Προτού ερμηνεύσουμε τη συμπεριφορά του πλήθους των ζημιών επισημαίνεται ότι τα δεδομένα μας είναι «στιγμιότυπο» με ημερομηνία 31/12/2014 οπότε οι ανοιχτές ζημιές του έτους 2010, για παράδειγμα, έχουν παραμείνει ανοιχτές παρόλο που έχουν 5 έτη «ωρίμανσης» και συνεπώς θα είναι αισθητά λιγότερες από μεταγενέστερα έτη που έχουν μικρότερη «ωρίμανση». Αντί του όρου *ωρίμανσης* από εδώ και στο εξής θα χρησιμοποιείται ο όρος *εξέλιξη*. Επομένως, για τα τελευταία έτη, για παράδειγμα το 2014, λαμβάνοντας υπόψη μόνο τις κλειστές ζημιές χάνουμε μεγάλη πληροφορία και το έτος ουσιαστικά δεν μπορεί να συγκριθεί με τα προηγούμενα.

Επίσης, άλλο ένα σημείο που μπορεί να κινήσει την περιέργεια του αναγνώστη είναι ότι οι εκκρεμείς φάκελοι των υλικών ζημιών είναι αρκετοί, ειδικά για 5 χρόνια μετά την αναγγελία τους.

Μελετώντας φάκελο προς φάκελο το αρχείο πληρωμών παρατηρούμε ότι για το έτος 2010, από τους 49 φακέλους που εκκρεμούν, οι 31 εμπεριέχουν και πληρωμές για σωματικές βλάβες. Αυτό σημαίνει ότι οι φάκελοι έχουν παραμείνει ανοιχτοί διότι ο διακανονιστής της εταιρείας

περιμένει να γίνουν και άλλες πληρωμές. Οι πληρωμές αυτές κατά πάσα πιθανότητα δεν θα αφορούν υλικές ζημιές αλλά σωματικές βλάβες.

Αρα, η εταιρεία κρατάει απόθεμα, συνεπώς ανοιχτούς τους φακέλους, για μόλις 18 φακέλους που αφορούν αποκλειστικά υλικές ζημιές από τους 4.481. Σε κάθε φάκελο το απόθεμα που εκτιμάται από την εταιρεία κάθε φορά έχει και χαρακτηρισμό για το είδος της κάλυψης που πρόκειται να χρησιμοποιηθεί. Όμως, η πληροφορία αυτή δεν μας έχει δοθεί από την εταιρεία, επομένως δεν μπορούμε να βρούμε το ακριβές πλήθος των εκκρεμών ζημιών για υλικές ζημιές και σωματικές βλάβες ξεχωριστά. Όλα τα ανωτέρω δεν δημιουργούν σφάλμα το οποίο να επηρεάζει την αξιοπιστία της ανάλυσης, αλλά είναι σημαντικό να αναφερθούν για την ορθή κατανόηση της δομής των δεδομένων.

Τέλος παρατηρούμε ότι για το έτος 2010 στην εταιρεία αναγγέλθηκαν 4.578 ζημιές εκ των οποίων, το 2014, δεν έχουν κλείσει οι 68 από αυτές. Το 2011 αναγγέλθηκαν 4.877 εκ των οποίων, το 2014, οι 130 δεν έχουν κλείσει. Αντίστοιχα και τα επόμενα έτη παρατηρούμε ότι ο αριθμός των εκκρεμών ζημιών μεγαλώνει, όπως είναι φυσιολογικό, καθώς μικραίνει η χρονική διάρκεια διακανονισμού της ζημιάς.

Επομένως, δουλεύοντας μόνο με κλειστούς φακέλους, ουσιαστικά δεν συμπεριλαμβάνουμε ένα πλήθος ζημιών της τάξης του 1%, 2%, 4%, 5% και 15% για τα έτη 2010, 2011, 2012, 2013, 2014 αντίστοιχα του συνολικού αριθμού ζημιών που έχουν έρθει στην εταιρεία. Το ποσοστό αυτό στην πραγματικότητα είναι ακόμα μικρότερο για τις υλικές ζημιές όπως αναλύθηκε στην προηγούμενη παράγραφο. Αντίθετα, για στις σωματικές βλάβες τα ποσοστά των εκκρεμών ζημιών είναι πολύ υψηλά για να τα αγνοήσουμε.

Ως εκ τούτου, αποφασίζουμε να δουλέψουμε με τις κλειστές ζημιές των οποίων οι πληρωμές καλύπτουν υλικές ζημιές αστικής ευθύνης προς τρίτους. Το έτος 2014 είναι υπό αμφισβήτηση αν θα ενσωματωθεί στη μελέτη μας.

2.2 Ανάλυση του ύψους των ζημιών

Ως προς το συνολικό ποσό πληρωμής των ζημιών σύμφωνα με το έτος αναγγελίας, τα στοιχεία που δίνονται από την εταιρεία επεξεργάζονται με τριγωνική ανάλυση, τα αποτελέσματα της οποίας παρατίθενται στον παρακάτω πίνακα:

<i>Χαρτοφυλάκιο Αστικής Ευθύνης Αυτοκινήτων (Σύνολο)</i>						
Πληρωθείσες Ζημιές						
Έτος Αναγγελίας	Σύνολο	E0	E1	E2	E3	E4
2010	5.118.941	2.331.748	1.127.277	563.569	220.244	876.103
2011	4.834.686	2.521.461	1.034.436	421.470	857.319	
2012	3.236.728	1.916.631	694.654	625.444		
2013	2.443.912	1.610.684	833.228			
2014	1.833.533	1.833.533				

Πίνακας 2-4: Τρίγωνο εξέλιξης πληρωθεισών ζημιών

Ο Πίνακας 2.4 μας δείχνει την εξέλιξη των πληρωμών μέχρι το 2014. Στην πρώτη στήλη είναι το έτος αναγγελίας το οποίο μπορούμε να το λάβουμε και ως έτος ατυχήματος θεωρώντας ότι η εταιρεία δεν έχει IBNR¹ ζημιές.

Στην δεύτερη στήλη υπολογίζεται το άθροισμα των πληρωμών που αφορούν ζημιές που αναγγέλθηκαν στο εκάστοτε έτος. Το άθροισμα αυτό αναλύεται στις επόμενες στήλες προκειμένου να διερευνηθεί η εξέλιξη διακανονισμού των ζημιών. Για παράδειγμα, το 2012 η εταιρεία πλήρωσε 1.916.631 ευρώ για τις ζημιές που ήρθαν το 2012. Τον επόμενο χρόνο, το

¹ Incurred But Not Reported

2013, οι ζημιές του 2012 έφεραν απαιτήσεις ύψους 694.654 ευρώ και το 2014 οι ζημιές του 2012 έφεραν απαιτήσεις ύψους 625.444 ευρώ.

Με την παραπάνω ανάλυση γίνεται αντιληπτό ότι για τις ζημιές του έτους 2014, στο οποίο δεν έχουμε εξέλιξη, η πληροφορία είναι ελλιπής ως προς την εξέλιξη του ύψους του ποσού. Το ίδιο παρατηρείται και για το έτος 2013, με αποτέλεσμα να δημιουργείται πρόβλημα ως προς την επιλογή του τελικού χαρτοφυλακίου.

Περαιτέρω ανάλυση και σχολιασμός των αποτελεσμάτων θα γίνει μετά τον διαχωρισμό του χαρτοφυλακίου σε Υλικές ζημιές και Σωματικές βλάβες.

<u>Υλικές Ζημιές</u>						
Πληρωθείσες Ζημιές						
Έτος Αναγγελίας	Σύνολο	E0	E1	E2	E3	E4
2010	3.496.303	2.201.041	845.269	141.651	30.684	277.659
2011	3.221.460	2.216.437	834.493	104.691	65.840	
2012	2.538.634	1.871.883	575.125	91.626		
2013	2.294.182	1.577.146	717.036			
2014	1.787.213	1.787.213				

Πίνακας 2-5: Τρίγωνο εξέλιξης πληρωθεισών υλικών ζημιών

<u>Σωματικές Βλάβες</u>						
Πληρωθείσες Ζημιές						
Έτος Αναγγελίας	Σύνολο	E0	E1	E2	E3	E4
2010	1.622.637	130.707	282.008	421.918	189.559	598.445
2011	1.613.226	305.024	199.944	316.779	791.479	
2012	698.094	44.748	119.528	533.818		
2013	149.730	33.538	116.193			
2014	46.320	46.320				

Πίνακας 2-6: Τρίγωνο εξέλιξης πληρωθεισών σωματικών βλαβών

Στον Πίνακα 2.5 φαίνεται για παράδειγμα ότι η εταιρεία το 2014 πλήρωσε 227.659 ευρώ για ζημιές που αναγγέλθηκαν το 2010. Για τις υλικές ζημιές, περιμένουμε ότι όσο περνούν τα χρόνια τα ποσά που πληρώνονται για ζημιές παλαιότερων ετών μειώνονται. Αυτό συμβαίνει σε όλα τα χρόνια εκτός από το 2014, όπου πληρώθηκαν 227.659 ζημιές που αναγγέλθηκαν το 2010.

Από το αρχείο πληρωμών που έχουμε λάβει, απομονώνοντας τις ζημιές του 2010 που έχουν πληρωμές το 2014 παρατηρούμε ότι, από τα 227.659 ευρώ, τα 165.991 ευρώ αφορούν μια μόνο ζημιά. Για αυτή τη ζημιά η εταιρεία μπορεί να ήταν σε δικαστική διαμάχη για την υπαιτιότητα και για αυτό το λόγο η πληρωμή να καθυστέρησε αλλά και να επιβαρύνθηκε και με άλλα δικαστικά έξοδα πέραν της αποζημίωσης για τις υλικές ζημιές.

Επίσης στο τέταρτο έτος εξέλιξης του 2010 παρουσιάζονται άλλες 6 πληρωμές πάνω από 9.500 ευρώ. Η αιτιολόγηση μπορεί να είναι παρόμοια με την αιτιολόγηση που δώσαμε παραπάνω αλλά επίσης, πιθανόν είναι η εταιρεία να θέλησε να πληρώσει για λόγους στρατηγικής την συγκεκριμένη χρονιά τις παραπάνω ζημιές.

Παρατηρούμε ότι οι υλικές ζημιές και οι σωματικές βλάβες έχουν ακριβώς αντίθετη συμπεριφορά ως προς τον χρόνο αποπληρωμής τους. Για τις υλικές ζημιές το μεγαλύτερο μέρος των πληρωμών γίνεται μέσα στην διετία ενώ για τις σωματικές βλάβες το μεγαλύτερο μέρος των πληρωμών γίνεται τρία ή και περισσότερα χρόνια μετά την αναγγελία τους.

Συνεπώς, με το δεδομένο αρχείο πληρωμών που διαθέτουμε, προκειμένου η ανάλυσή μας να έχει μεγαλύτερη δυνατή αξιοπιστία, εξαιρούμε από τη μελέτη μας τις σωματικές βλάβες και συνεχίζουμε να εργαζόμαστε με τις υλικές ζημιές.

Για τις υλικές ζημιές, υπολογίζεται το ποσοστό των φακέλων που έχουν διεκπεραιωθεί (κλειστοί φάκελοι) και παρατίθεται με το παρακάτω τρίγωνο.

<i>Κλειστές υλικές ζημιές ως ποσοστό επι του συνόλου των υλικών ζημιών</i>					
Πληρωθείσες Ζημιές					
Έτος Αναγγελίας	Ε0	Ε1	Ε2	Ε3	Ε4
2010	99%	99%	78%	82%	36%
2011	99%	93%	72%	83%	
2012	99%	86%	74%		
2013	97%	91%			
2014	94%				

Πίνακας 2-7: Τρίγωνο με το ποσοστό των κλειστών υλικών ζημιών επί του συνόλου των υλικών ζημιών ανά έτος αναγγελίας

Παρατηρούμε ότι το ποσοστό των κλειστών φακέλων τον πρώτο χρόνο εξέλιξης είναι αρκετά υψηλό. Αυτό πρακτικά σημαίνει, ότι το 94% και άνω των ζημιών που αναγγέλθηκαν για κάθε έτος, είναι κλειστές με ημερομηνία αποτίμησης το 2014. Αντιθέτως, ζημιές που φέρνουν πληρωμές και σε μεταγενέστερα έτη παρατηρείται ότι παραμένουν ανοιχτές. Δηλαδή, από το συνολικό ποσό που πληρώθηκε για υλικές ζημιές, που αναγγέλθηκαν το 2010, μέσα στο 2014 μόλις το 36% έκλεισαν. Αντίστοιχα, το 82% του ποσού που δόθηκε για αποζημιώσεις που αφορούν ζημιές με έτος αναγγελίας το 2010, αφορά κλειστές ζημιές. Σημειώνεται ιδιαίτερος ότι κλειστές ζημιές χαρακτηρίζονται το 2014. Επομένως, είναι φυσικό το ποσοστό αυτό να φθίνει όσο πλησιάζει την διαγώνιο του τριγώνου.

Οι λόγοι που μας επιτρέπουν να συνεχίσουμε την ανάλυσή μας μόνο με κλειστές υλικές ζημιές και χωρίς να έχουμε μεγάλο σφάλμα στα δεδομένα μας είναι οι ακόλουθοι:

- Το ποσό των πληρωμών για τις υλικές ζημιές μέσα στη διετία έχει ωριμάσει σε βαθμό μεγαλύτερο του 87% και
- Οι απαιτήσεις των κλειστών υλικών ζημιών στη διετία είναι κατά μέσο όρο, για όλα τα έτη αναγγελίας, μεγαλύτερο από 95% του συνολικού ποσού των υλικών ζημιών.

Επομένως, σε συνδυασμό και με την ανάλυση που κάναμε για το πλήθος των ζημιών, θα εργαστούμε από εδώ και στο εξής με τις κλειστές ζημιές που ήρθαν στην εταιρεία από το 2010 μέχρι το 2013 και αφορούν αποζημιώσεις για την κάλυψη υλικών ζημιών. Στη συνέχεια για αυτό το δείγμα, το οποίο θα αποκαλείται «Χαρτοφυλάκιο», γίνεται περαιτέρω ανάλυση.

2.3 Περιγραφική Στατιστική Ανάλυση Χαρτοφυλακίου

Σε αυτή την υπό-ενότητα, γίνεται περιγραφική στατιστική ανάλυση των ζημιών που έφτασαν στην εταιρεία για τα έτη 2010 – 2013. Η ανάλυση των δεδομένων γίνεται τόσο σε εβδομαδιαία όσο και σε μηνιαία βάση καθώς, όπως θα δούμε και σε επόμενα κεφάλαια, για τις ατομικές ζημιές δεν είναι εύκολο να βρεθεί κάποια κατανομή που να τις προσεγγίζει ικανοποιητικά.

2.3.1 Εβδομαδιαία Ανάλυση

Ξεκινώντας με την διερευνητική ανάλυση των δεδομένων, στον επόμενο πίνακα παρουσιάζονται οι τιμές από τα κυριότερα περιγραφικά στατιστικά χαρακτηριστικά σχετικά με το ύψος πληρωμής και το πλήθος ζημιών σε εβδομαδιαία βάση.

Περιγραφικά Στατιστικά Στοιχεία	Πληρωθείσες Ζημιές	Πλήθος Ζημιών
Μέσος	52.046,28	82,05
Τυπικό σφάλμα	1.188,36	1,24
Διάμεσος	48.781,93	83,50
Επικρατούσα τιμή	-	88,00
Μέση απόκλιση τετραγώνου	17.221,00	17,91
Διακύμανση	296.562.671,61	320,89
Κύρτωση	0,54	0,92
Ασυμμετρία	0,53	-0,50
Εύρος	106.091,28	115,00
Ελάχιστο	5.643,45	14,00
Μέγιστο	111.734,73	129,00
Άθροισμα	10.929.719,58	17.230,00
Πλήθος	210,00	210,00

Πίνακας 2-8: Περιγραφικά στατιστικά στοιχεία για το ύψος και το πλήθος των ζημιών που αναγγέλλονται ανά εβδομάδα

Στην εταιρεία από το 2010 μέχρι και το 2013 αναγγέλλονται κατά μέσο όρο 82 ζημιές εβδομαδιαία και το εβδομαδιαίο μέσο ποσό αποζημίωσης είναι 52.046,28€. Άρα το μέσο κόστος ανά ζημιό είναι 634,324 ευρώ.

Παρατηρούμε ότι η διάμεσος είναι αρκετά κοντά με τη μέση τιμή τόσο για τις πληρωμές όσο και για το πλήθος των ζημιών. Επομένως, περιμένουμε ότι η κατανομή και για τα δύο δείγματα θα είναι μάλλον συμμετρική.

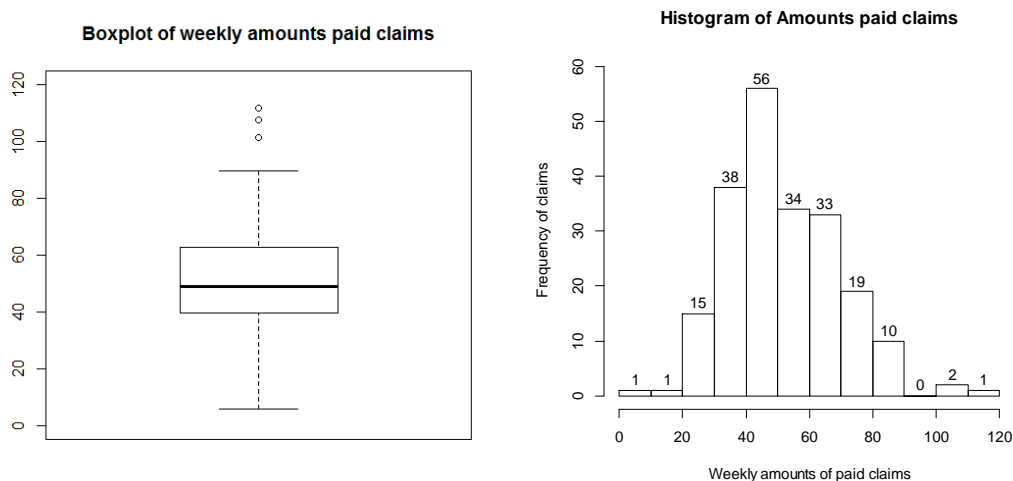
Ο συντελεστής ασυμμετρίας ενισχύει αυτή την ένδειξη καθώς, τόσο για τις πληρωμές όσο και για το πλήθος, είναι κοντά στο 0.

Για το πλήθος των ζημιών η διακύμανση είναι 320,89 και η τυπική απόκλιση του μέσου υπολογίζεται 1,24. Αυτό συνεπάγεται μεγάλη συγκέντρωση των τιμών γύρω από το μέσο.

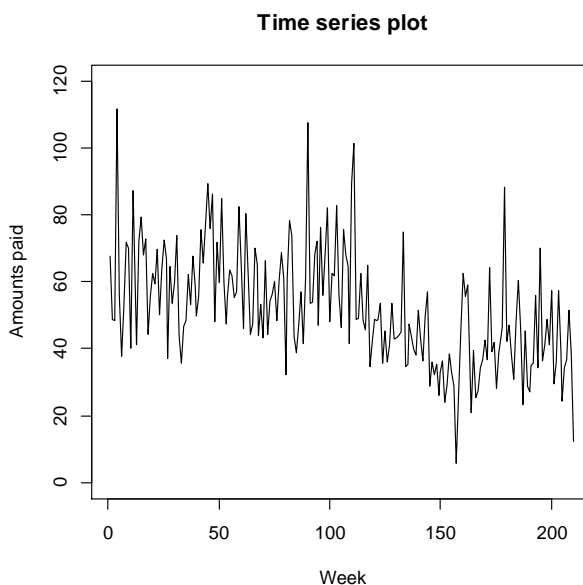
Αντιθέτως οι πληρωμές παρουσιάζουν μεγάλη απόκλιση από τον μέσο, δηλαδή μεγάλη διασπορά. Η οξύτητα της κορυφής της κατανομής των πληρωμών και του πλήθους των ζημιών, δίνεται από τον συντελεστή κύρτωσης, ο οποίος είναι 0,54 και 0,92 αντίστοιχα. Συνεπώς, οι κατανομές του πλήθους και των πληρωμών είναι πλατύκυρτες.

Όσα αναφέρθηκαν, ερμηνεύονται και κατανοούνται καλύτερα από τα γραφήματα που ακολουθούν.

Α) Γραφήματα για το ύψος των αποζημιώσεων



Γράφημα 2-1: Θηκόγραμμα και Ιστόγραμμα των εβδομαδιαίων πληρωμών²



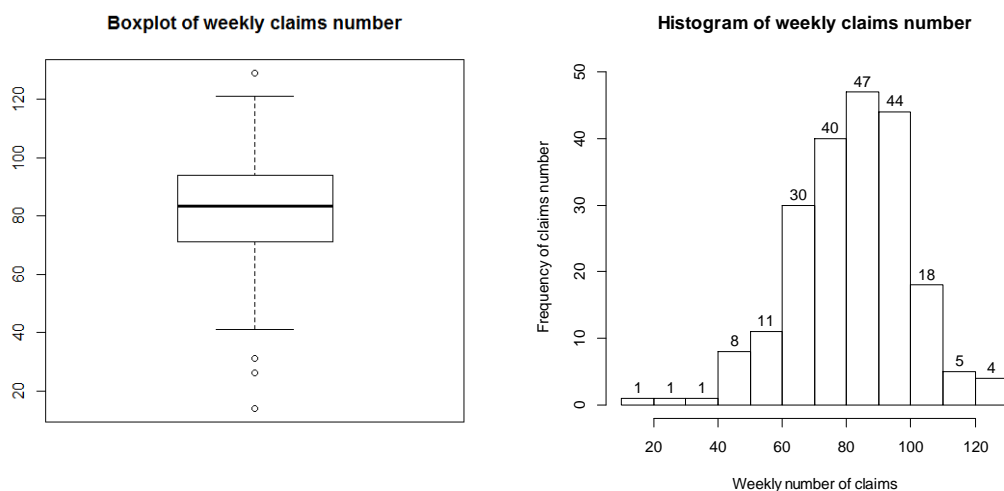
Γράφημα 2-2: Διάγραμμα χρονολογικής ανάλυσης των εβδομαδιαίων πληρωμών²

Η γραφική ένδειξη που λαμβάνουμε, ακολουθεί τα συμπεράσματα από τα περιγραφικά μέτρα που υπολογίσαμε και σχολιάσαμε παραπάνω. Από το θηκόγραμμα παρατηρείται ότι το ύψος των πληρωμών παρουσιάζει μικρό πλήθος ακραίων τιμών. Από το ιστόγραμμα έχουμε ενδείξεις ότι η κατανομή των πληρωμών είναι συμμετρική. Όσον αφορά το διάγραμμα

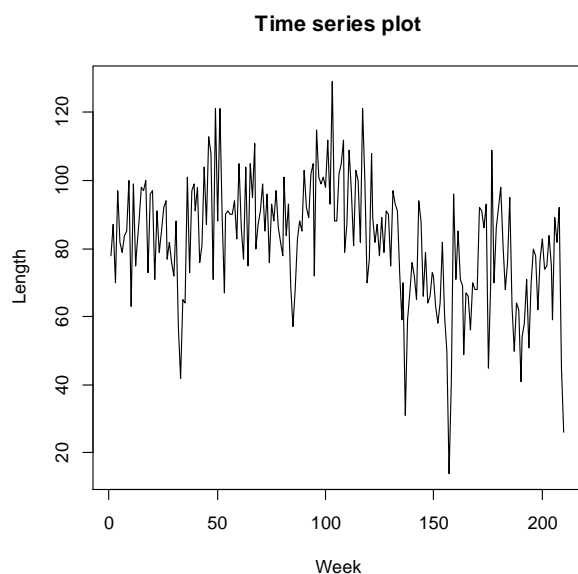
² Τα ποσά είναι σε χιλιάδες ευρώ

χρονολογικής ανάλυσης, παρατηρείται μια μετατόπιση του ύψους των ζημιών από το 2012 και μετά προς τα κάτω, περίπου κατά 20.000 ευρώ.

B) Γραφήματα για το πλήθος των ζημιών



Γράφημα 2-3: Θηκόγραμμα και Ιστόγραμμα του εβδομαδιαίου πλήθους ζημιών



Γράφημα 2-4: Διάγραμμα χρονολογικής ανάλυσης εβδομαδιαίου πλήθους ζημιών

Από το θηκόγραμμα του πλήθους των ζημιών βλέπουμε ότι παρουσιάζονται τέσσερις ακραίες τιμές, οι τρεις εκ των οποίων είναι πολύ μικρές και η μια εξ' αυτών μεγάλη. Η εμφάνιση μικρού πλήθους αναγγελιών μπορεί να οφείλεται σε εβδομάδες που η εταιρεία δεν λειτούργησε λόγω αργιών, για παράδειγμα τις εορταστικές ημέρες των Χριστουγέννων και του Πάσχα. Πιθανό όμως είναι ότι δεν έχουμε ακριβώς μετά (χρονικά) από αυτές τις τιμές μεγάλο αριθμό παρατηρήσεων. Επίσης το ελάχιστο πλήθος και το ελάχιστο ποσό παρατηρούνται την ίδια χρονική περίοδο (τελευταία εβδομάδα του 2012) και όπως αναμενόταν η εβδομάδα αυτή εμπεριέχει την χριστουγεννιάτικη αργία.

Από το ιστόγραμμα υπάρχει ένδειξη ότι η κατανομή του πλήθους των ζημιών είναι συμμετρική. Όσον αφορά το διάγραμμα χρονολογικής ανάλυσης, παρατηρούμε πως στα πρώτα δύο χρόνια

εμφανίζεται η θετική ακραία τιμή ενώ από το 2012 μέχρι το 2013 εμφανίζονται οι τρεις χαμηλότερες τιμές.

2.3.2 Μηνιαία Ανάλυση

Αντίστοιχα με την εβδομαδιαία διερευνητική ανάλυση των δεδομένων που έγινε και στην προηγούμενη ενότητα, παρουσιάζονται οι τιμές από τα κυριότερα περιγραφικά στατιστικά χαρακτηριστικά σχετικά με το ύψος πληρωμής και το πλήθος ζημιών σε μηνιαία βάση.

<i>Περιγραφικά Στατιστικά Στοιχεία</i>	<i>Πληρωθείσες Ζημιές</i>	<i>Πλήθος Ζημιών</i>
Μέσος	227.702,49	358,96
Τυπικό σφάλμα	7.488,57	7,93
Διάμεσος	239.107,97	362,00
Επικρατούσα τιμή	-	355,00
Μέση απόκλιση τετραγώνου	51.882,35	54,95
Διακύμανση	2.691.778.001,32	3.019,27
Κύρτωση	-0,21	-0,35
Ασυμμετρία	-0,04	-0,21
Εύρος	244.304,50	223,00
Ελάχιστο	119.778,16	239,00
Μέγιστο	364.082,66	462,00
Άθροισμα	10.929.719,58	17.230,00
Πλήθος	48,00	48,00

Πίνακας 2-9: Περιγραφικά στατιστικά στοιχεία για το ύψος και το πλήθος των ζημιών που αναγγέλλονται ανά εβδομάδα

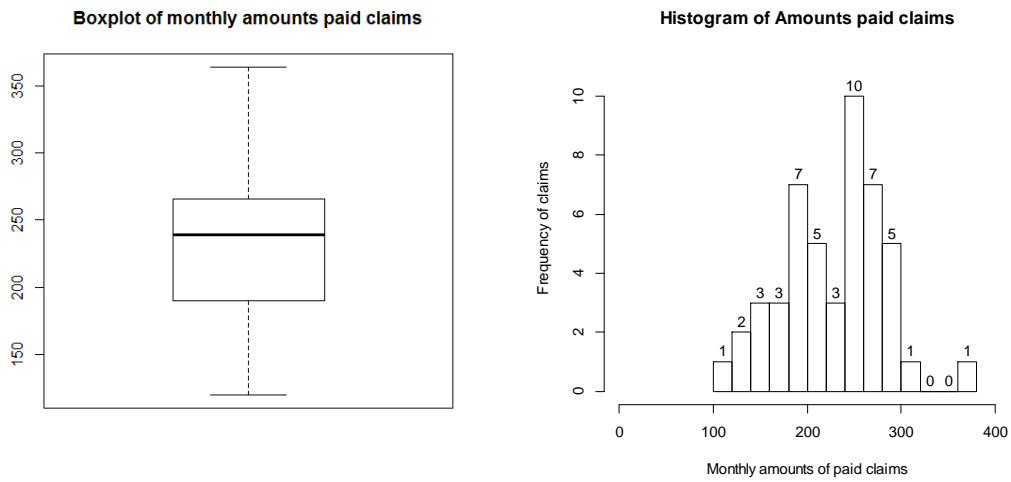
Στην εταιρεία από το 2010 μέχρι και το 2013 αναγγέλλονται κατά μέσο όρο 359 ζημιές μηνιαία και το μέσο μηνιαίο κόστος που καλείται να πληρώσει η εταιρεία είναι 227.702,49€. Παρατηρούμε ότι η διάμεσος είναι αρκετά κοντά με τη μέση τιμή τόσο για τις πληρωμές όσο και για το πλήθος των ζημιών. Επομένως, περιμένουμε ότι οι κατανομή και για τα δύο δείγματα θα είναι συμμετρική.

Ο συντελεστής ασυμμετρίας έρχεται και ενισχύει αυτή την ένδειξη καθώς, τόσο για τις πληρωμές όσο και για το πλήθος, είναι κοντά στο 0. Για το πλήθος των ζημιών η διακύμανση είναι 3.019,27 και η τυπική απόκλιση του μέσου υπολογίζεται 7,93. Αυτό συνεπάγεται μεγάλη συγκέντρωση των τιμών γύρο από το μέσο.

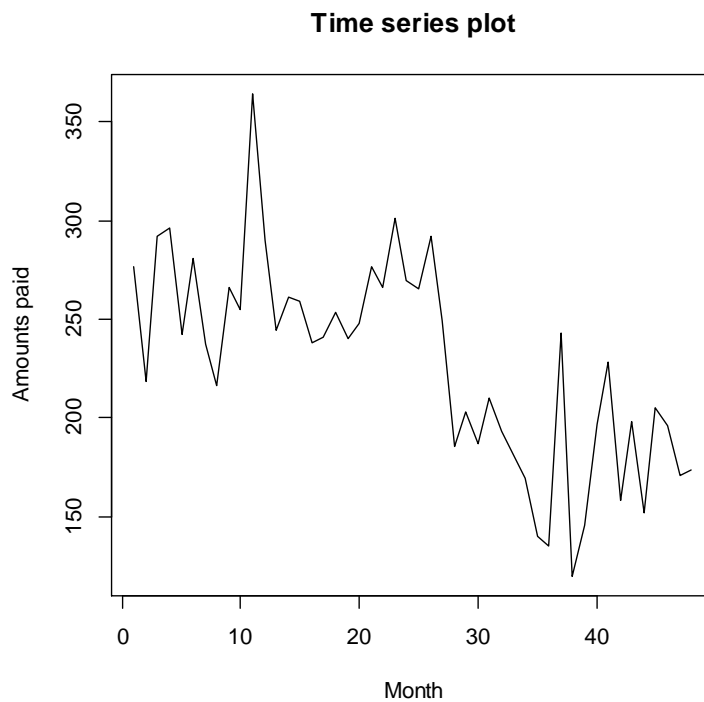
Αντιθέτως οι πληρωμές παρουσιάζουν μεγάλη απόκλιση από τον μέσο, δηλαδή μεγάλη διασπορά. Η οξύτητα της κορυφής της κατανομής των πληρωμών και του πλήθους των ζημιών, δίνεται από τον συντελεστή κύρτωσης, ο οποίος είναι -0,21 και -0,35 αντίστοιχα. Συνεπώς, οι κατανομές του πλήθους και των πληρωμών είναι πλατύκυρτες.

Όσα αναφέρθηκαν, παρουσιάζονται και κατανοούνται καλύτερα με τα παρακάτω γραφήματα:

Α) Γραφήματα για το ύψος των αποζημιώσεων



Γράφημα 2-5: Θηκόγραμμα και Ιστόγραμμα των μηνιαίων πληρωμών³



Γράφημα 2-6: Διάγραμμα χρονολογικής ανάλυσης των μηνιαίων πληρωμών³

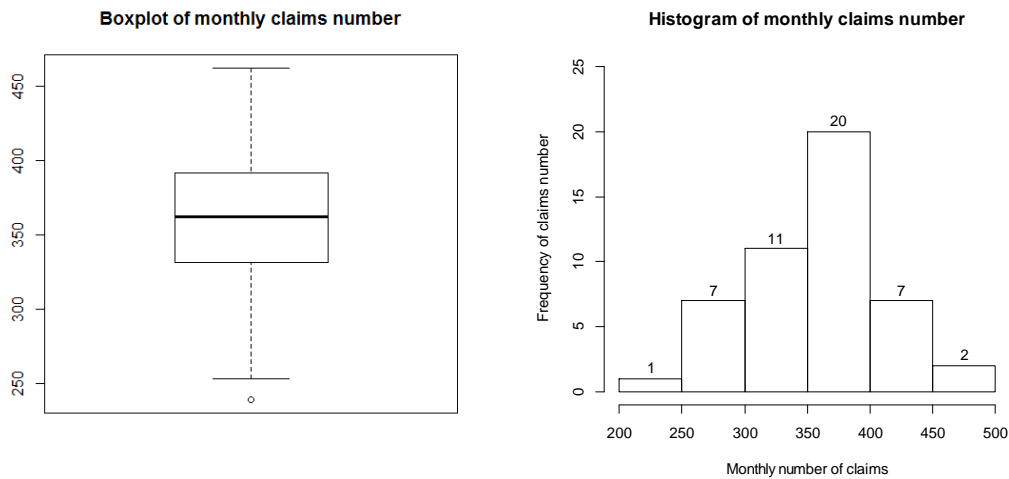
Από τα παραπάνω γραφήματα, είναι εμφανές ότι τα συνολικά ποσά που πληρώθηκαν για τις μηνιαίες αναγγελθείσες ζημιές δεν παρουσιάζουν ακραίες τιμές. Από το ιστόγραμμα, βλέπουμε ότι η κατανομή τους εμφανίζει δύο κορυφές.

³ Τα ποσά είναι σε χιλιάδες ευρώ

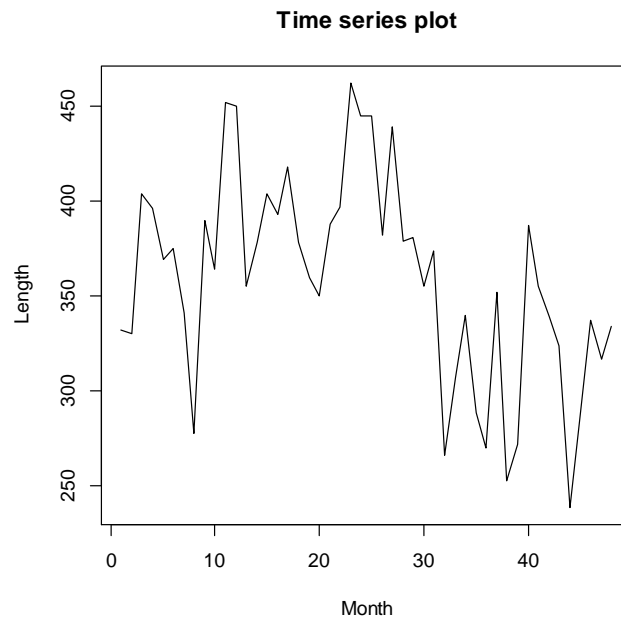
Αν και παρατηρούνται δύο κορυφές από το θηκόγραμμα έχουμε ενδείξεις για κανονική κατανομή, το οποίο είναι λογικό και σύμφωνο με τον νόμο των μεγάλων αριθμών.

Το διάγραμμα χρονολογικής ανάλυσης εμφανίζει ανάλογη εικόνα με εκείνη της εβδομαδιαίας ανάλυσης κάνοντας πιο εμφανή την μετατόπιση του ύψους των ζημιών από το 2012 και μετά προς τα κάτω. Επίσης, τα δεδομένα μας δεν φαίνεται να παρουσιάζουν κάποια εποχικότητα καθώς δεν σχηματίζεται κάποιος κανόνας είτε για μήνα είτε για τρίμηνο προκειμένου να διεξάγουμε το συμπέρασμα ότι συμπεριφέρονται παρόμοια τα ποσά των πληρωμών σε κάθε έτος.

B) Γραφήματα για το πλήθος των ζημιών



Γράφημα 2-7: Θηκόγραμμα και Ιστόγραμμα του μηνιαίου πλήθους ζημιών



Γράφημα 2-8: Διάγραμμα χρονολογικής ανάλυσης του μηνιαίου πλήθους των ζημιών

Από τα διαγράμματα υπάρχει ένδειξη ότι το πλήθος των ζημιών ακολουθεί την κανονική κατανομή. Αυτό προκύπτει τόσο από το ιστόγραμμα όσο και από το θηκόγραμμα. Στο

θηκόγραμμα παρουσιάζεται μόνο μία ακραία τιμή, γεγονός το οποίο δεν επηρεάζει το συμπέρασμά μας για συμμετρική κατανομή.

Όσον αφορά το διάγραμμα χρονολογικής ανάλυσης, παρατηρούμε ότι η συμπεριφορά του πλήθους είναι παρόμοια με τη συμπεριφορά του ύψους των ζημιών. Επίσης, τα πρώτα δύο χρόνια το μηνιαίο πλήθος αναγγελιών στην εταιρία ήταν αρκετά μεγαλύτερο σε σχέση με τα δύο επόμενα. Το γεγονός αυτό μπορεί να οφείλεται σε αλλαγή της πολιτικής ανάληψης κινδύνων της εταιρείας, για παράδειγμα μπορεί να η εταιρεία να είχε ασφαλίσει μια κοορτή με μεγαλύτερη έκθεση στον κίνδυνο τα παλαιότερα χρόνια. Επιπλέον, η μείωση αυτή μπορεί να οφείλεται σε μείωση της παραγωγής, δηλαδή στην μείωση του πλήθους των ασφαλισμένων οχημάτων. Τέλος δεν φαίνεται να παρουσιάζεται κάποια τάση ή περιοδικότητα στα στοιχεία.

Κεφάλαιο 3

Κατανομές πιθανότητας για το πλήθος και το ύψος των ζημιών

Στο κεφάλαιο αυτό θα αναφέρουμε κάποιες βασικές κατανομές της θεωρίας πιθανοτήτων τις οποίες και θα χρησιμοποιήσουμε για τη μελέτη της κατανομής που ακολουθούν τα εμπειρικά δεδομένα.

Συγκεκριμένα, γίνεται αναφορά σε ένα σύνολο από διακριτές και συνεχείς κατανομές. Οι διακριτές χρησιμοποιούνται συνήθως για να ερμηνεύσουν τη συμπεριφορά του πλήθους των αποζημιώσεων και οι συνεχείς το ύψος.

3.1 Διωνυμική κατανομή

Ορισμός

Έστω X ο αριθμός των επιτυχιών σε μια ακολουθία n (ανεξάρτητων) δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p και αποτυχίας $q = 1 - p$ σε όλες τις δοκιμές. Η κατανομή της τυχαίας μεταβλητής X καλείται Διωνυμική κατανομή με παραμέτρους n και p και συμβολίζεται $b(n, p)$ βλέπε π.χ. (Κούτρας Μ., 2004, σ. 259).

Η συνάρτηση πιθανότητας της Διωνυμικής κατανομής δίνεται από τον τύπο:

$$f(x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \dots, n.$$

Ροπές

Έστω η τυχαία μεταβλητή X με σύνολο τιμών $R_x = \{0, 1, 2, \dots\}$ ακολουθεί $b(n, p)$. Οι παρακάτω σχέσεις δίνουν τις ροπές γύρω από το μηδέν και γύρω από το μέσο.

Ροπές γύρω από το μηδέν: $\mu'_1 = E(X) = np$ "Μέσος"

$$\mu'_2 = E(X^2) = npq + n^2 p^2$$

$$\mu'_3 = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$$

Ροπές γύρω από το μέσο: $\mu_1 = E[X - E(X)] = 0$

$$\mu_2 = E[X - E(X)]^2 = npq$$
 "Διακύμανση"

$$\mu_3 = E[X - E(X)]^3 = npq(q-p)$$

Μια παρατήρηση που μπορούμε να κάνουμε από τα παραπάνω, είναι ότι η μέση τιμή είναι πάντα μεγαλύτερη από την διακύμανση. Οι αποδείξεις των παραπάνω σχέσεων δίνονται αναλυτικά στη βιβλιογραφία π.χ. (Kumar, A. and Chaudhry, A., 2008, p. 178).

3.2 Κατανομή Poisson

Η Διωνυμική κατανομή που αναφέραμε παραπάνω είναι εύχρηστη στους υπολογισμούς για μικρά δείγματα. Σε περίπτωση όμως, που το πλήθος n παίρνει μεγάλες τιμές και η πιθανότητα επιτυχίας p είναι πολύ μικρή, ο υπολογισμός των πιθανοτήτων μέσω του τύπου της Διωνυμικής κατανομής καθίσταται αρκετά δύσκολος.

Ο *Simeon – Denis Poisson* το 1837 δημοσίευσε μια μελέτη που είχε ως αποτέλεσμα την ανακάλυψη μιας έμμεσης μεθόδου υπολογισμού των πιθανοτήτων της Διωνυμικής κατανομής όταν το πλήθος είναι μεγάλο ($n \rightarrow \infty$) και η πιθανότητα επιτυχίας πολύ μικρή ($p \rightarrow 0$). Απέδειξε ότι, αν ο μέσος αριθμός επιτυχιών στις n δοκιμές παίρνει «μέτριες» τιμές, δηλαδή το γινόμενο np συγκλίνει προς κάποιο σταθερό αριθμό $\lambda > 0$ τότε ισχύει,

$$\lim_{n \rightarrow \infty} f_{binomial}(x) = \frac{e^{-\lambda} \lambda^x}{x!}.$$

Μετά από μισό αιώνα, το 1889, ο Ρώσο-Γερμανός μαθηματικός L. V. Bortriewicz παρατήρησε ότι η συνάρτηση – προσέγγιση στην οποία είχε φτάσει ο *Poisson* είχε όλες τις ιδιότητες μιας συνάρτησης πιθανότητας αφού είναι μη αρνητική για όλα τα $x = 0, 1, 2, \dots, n$, και επιπλέον ισχύει:

$$\sum_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_0^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Η συνάρτηση πιθανότητας της κατανομής Poisson δίδεται από στη σχέση:

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots$$

και η πιθανογεννήτρια από την σχέση : $P_X(z) = e^{\lambda(z-1)}$.

Μέσω της πιθανογεννήτριας υπολογίζεται η μέση τιμή και η διακύμανση, $E(X) = \lambda$ και $V(X) = \lambda$ αντίστοιχα.

Παρατηρούμε ότι η διακύμανση και η μέση τιμή είναι ίσες. Αυτό είναι μια βασική ιδιότητα της κατανομής Poisson. Άλλη μια βασική ιδιότητα για την κατανομή Poisson δίδεται παρακάτω σε μορφή θεωρήματος.

Θεώρημα: Έστω ότι οι τυχαίες μεταβλητές X_1, X_2, \dots, X_n είναι ανεξάρτητες και προέρχονται από την κατανομή Poisson με παραμέτρους $\lambda_1, \lambda_2, \dots, \lambda_n$, αντίστοιχα. Τότε, η τυχαία μεταβλητή $S_1 = X_1 + X_2 + \dots + X_n$ ακολουθεί Poisson με παράμετρο $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$.

3.3 Αρνητική Διωνυμική κατανομή

Ορισμός

Θεωρούμε μια ακολουθία (ανεξάρτητων) δοκιμών Bernoulli με πιθανότητα επιτυχίας p (και αποτυχίας $q = 1 - p$) σταθερή για όλες τις δοκιμές. Έστω X ο αριθμός των δοκιμών μέχρι την εμφάνιση της r επιτυχίας. Η κατανομή της τυχαίας μεταβλητής X καλείται Αρνητική Διωνυμική κατανομή με παραμέτρους r και p , και συμβολίζεται με $Nb(r, p)$ βλέπε π.χ. (Κούτρας Μ., 2004, σ. 277).

Η συνάρτηση πιθανότητας της Αρνητικής Διωνυμικής κατανομής δίνεται από την σχέση:

$$f(x) = \binom{x-1}{r-1} p^r q^{x-r}, x = r, r+1, r+2, \dots$$

Θέτοντας, $Y = X - r$ παίρνουμε την σχέση:

$$f(y) = \binom{y+r-1}{r-1} p^r q^y, y = 1, 2, \dots$$

η οποία αποτελεί συνάρτηση πιθανότητας, καθώς $\sum_{y=0}^{\infty} f(y) = 1$, και επομένως αποτελεί τη συνάρτηση πιθανότητας της Y , η οποία αντιπροσωπεύει τον αριθμό αποτυχιών πριν την r επιτυχία.

Μέτρα θέσης και διασποράς

Χρησιμοποιώντας την πιθανογεννήτρια της αρνητικής διωνυμικής κατανομής, η οποία είναι

$P_X(z) = \left(\frac{1-p}{1-pz}\right)^r$ υπολογίζουμε τα ακόλουθα:

$$E(X) = \frac{rq}{p}$$

$$V(X) = \frac{rq}{p^2}$$

$$Skewness = \frac{1+p}{\sqrt{pr}}$$

$$Kurtosis = \frac{6}{r} + \frac{(1-p)^2}{pr}$$

Παρατηρούμε ότι η διακύμανση, για την διωνυμική κατανομή, είναι πάντα μεγαλύτερη από την μέση τιμή σε αντίθεση με την κατανομή Poisson που η διακύμανση ισούται με την μέση τιμή.

Επομένως, όταν τα εμπειρικά δεδομένα παρουσιάζουν διακύμανση μεγαλύτερη από τη μέση τιμή, τότε είναι προτιμότερο να χρησιμοποιείται η αρνητική διωνυμική έναντι της Poisson.

3.4 Κατανομή Γάμμα

Στην θεωρία πιθανοτήτων και στην στατιστική, η κατανομή Γάμμα είναι από τις πιο γνωστές και συχνά χρησιμοποιούμενες κατανομές. Οι αναλογιστές την χρησιμοποιούν προκειμένου να περιγράψουν τα δεδομένα τους, τόσο σε χαρτοφυλάκια ζωής όσο και σε χαρτοφυλάκια γενικών ασφαλίσεων. Έχει παρατηρηθεί ότι προσαρμόζεται επίσης ικανοποιητικά σε χαρτοφυλάκια ζημιών αυτοκινήτων.

Αναφέρουμε αρχικά τη γενική περίπτωση μιας Γάμμα κατανομής με τρεις παραμέτρους και στη συνέχεια τη, συνηθέστερη, μορφή της κατανομής με δύο παραμέτρους, η οποία μας είναι γνωστή από τη θεωρία πιθανοτήτων.

Ορισμός

Η συνεχής τυχαία μεταβλητή $X \geq \gamma$ ακολουθεί την κατανομή Γάμμα(α, β, γ), με παραμέτρους $\alpha, \beta > 0$ και $x > \gamma$, αν η συνάρτηση πυκνότητας πιθανότητας είναι της μορφής:

$$f_X(x) = \frac{(x - \gamma)^{\alpha-1} e^{-\frac{x-\gamma}{\beta}}}{\beta^\alpha \Gamma(\alpha)}, \quad x > \gamma \text{ με } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \quad \alpha > 0$$

και συμβολικά γράφουμε $X \sim \text{Γάμμα}(\alpha, \beta, \gamma)$.

Η παραπάνω σχέση εμπεριέχει τρεις παραμέτρους α, β, γ . Εάν η παράμετρος $\gamma = 0$, τότε η κατανομή μπορεί να οριστεί ως:

$$f_X(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}, \quad x > 0 \quad (3.1)$$

Επομένως, όταν η τυχαία μεταβλητή X είναι μη αρνητική, τότε εργαζόμαστε με τη σχέση 3.1 και λέμε ότι $X \sim \text{Γάμμα}(\alpha, \beta)$ (βλέπε π.χ Johnson, 1995, p. 337).

Θεώρημα

Αν η τυχαία μεταβλητή $X \sim \text{Γάμμα}(\alpha, \beta)$ με $\alpha, \beta > 0$ τότε ισχύουν τα ακόλουθα:

1. $M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha, t < \beta.$
2. $F_X(x) = \Gamma\left(\alpha, \frac{x}{\beta}\right), x > 0.$
3. $E(X^k) = \frac{\Gamma(k+\alpha)}{\beta^k \Gamma(\alpha)}, k = 1, 2, 3, \dots$
4. $E(X) = \alpha\beta$ και $\text{Var}(X) = \alpha\beta^2.$

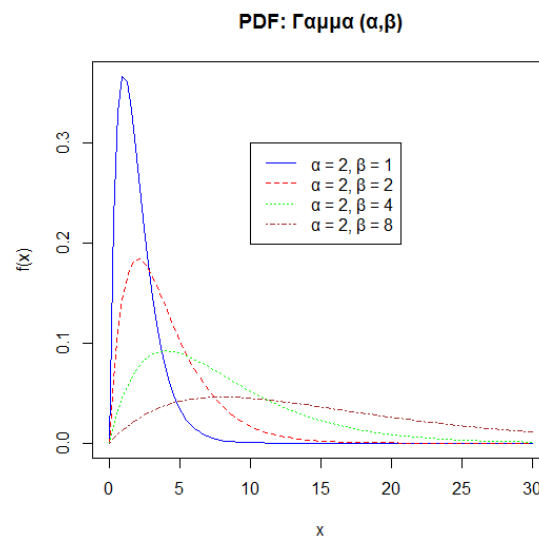
Για την απόδειξη του θεωρήματος βλέπε π.χ (Χατζηκωνσταντινίδης Ε., 2015)

Βασικές Ιδιότητες

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \alpha > 0$
- $\Gamma(n) = (n - 1)!, n = 1, 2, 3, \dots$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Γραφική απεικόνιση της συνάρτησης πυκνότητας της Γάμμα

Το παρακάτω γράφημα δείχνει την μορφή της συνάρτησης πυκνότητας πιθανότητας (PDF) της κατανομής Γάμμα για διαφορετικές τιμές της παραμέτρου β (παραμέτρος κλίμακας)



Γράφημα 3-1: Συνάρτηση πυκνότητας πιθανότητας κατανομής Γάμμα για διάφορα β

Όταν η παράμετρος α είναι σταθερή, παρατηρούμε ότι όσο το β παίρνει μεγαλύτερες τιμές, τόσο μεγαλώνει η ουρά της κατανομής, δηλαδή η κατανομή Γάμμα μπορεί και ερμηνεύει μεγαλύτερες ζημιές.

3.5 Κανονική Κατανομή

Ορισμός

Μια συνεχής τυχαία μεταβλητή X θα λέμε ότι ακολουθεί την κανονική κατανομή με παραμέτρους μ και σ^2 αν η συνάρτηση πυκνότητας f της X δίνεται από τον τύπο:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, \mu \in \mathbb{R}, \sigma > 0$$

Η παράμετρος μ είναι η μέση τιμή της τυχαίας μεταβλητής X και η παράμετρος σ^2 η διακύμανσή της και συμβολικά γράφεται $X \sim N(\mu, \sigma^2)$.

Θέτοντας, $z = \frac{x-\mu}{\sigma}$ παίρνουμε την τυποποιημένη κανονική κατανομή της οποίας η συνάρτηση πυκνότητας δίνεται από τον τύπο:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < \infty.$$

Η μέση τιμή της Z είναι μηδέν και η διασπορά της ισούται με τη μονάδα. Συμβολικά θα γράφουμε ότι $Z \sim N(0,1)$.

Βασικές Ιδιότητες

Εστω, $F(x)$ και $\Phi(z)$ η συνάρτηση κατανομής των τυχαίων μεταβλητών X και Z αντίστοιχα.

Οι τυχαίες μεταβλητές X και Z συνδέονται γραμμικά καθώς $Z = \frac{X-\mu}{\sigma}$. Επομένως έχουμε:

- $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- $\Phi(z) = F(\mu + \sigma z)$

Επίσης, οι $F(x)$ και $\Phi(z)$ είναι συμμετρικές ως προς το μ και το 0 αντίστοιχα. Μαθηματικά αυτό διατυπώνεται με τη σχέση: $f\left[\frac{(x-\mu)}{\sigma}\right] = f\left[\frac{(\mu-x)}{\sigma}\right]$.

Για την τυχαία μεταβλητή $X \sim N(\mu, \sigma^2)$ τα βασικά μέτρα θέσης και διασποράς είναι τα ακόλουθα:

Μέση τιμή	μ
Διάμεσος	μ
Διακύμανση	σ^2
Συντελεστής Μεταβλητότητας	σ/μ
Συντελεστής Κύρτωσης	3
Συντελεστής Ασυμμετρίας	0

Πίνακας 3-1: Μέτρα θέσης και διασποράς της κανονικής κατανομής

Οι παραπάνω σχέσεις παρουσιάζονται αναλυτικά στη βιβλιογραφία π.χ. (Patel and Read, 1982, p. 22).

3.6 Λογαριθμοκανονική Κατανομή

Ορισμός

Η συνεχής τυχαία μεταβλητή X λέμε ότι ακολουθεί τη Λογαριθμοκανονική κατανομή με παραμέτρους μ και σ^2 , και γράφουμε συμβολικά $X \sim L(\mu, \sigma^2)$, αν αυτή λαμβάνει τιμές στο διάστημα $(0, \infty)$ και ο λογάριθμός της $Y = \log X$, ακολουθεί την κατανομή $N(\mu, \sigma^2)$ βλέπε π.χ. (Αθανασόπουλος Δ., 1991, σ. 175). Όπου \log δηλώνει τον φυσικό λογάριθμο.

Η συνάρτηση πυκνότητας της λογαριθμοκανονικής κατανομής δίνεται από τη σχέση:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2}, \quad 0 < x < \infty$$

3.7 Σχέση Λογαριθμοκανονικής και Τυποποιημένης κανονικής κατανομής

Γνωρίζουμε ότι, $Y = \log X \Rightarrow X = e^Y$. Επομένως,

$$g(y) = f(e^y) \frac{dx}{dy} = \frac{1}{e^y \sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} e^y \Rightarrow g(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \Rightarrow Y \sim N(\mu, \sigma).$$

Είναι φανερό ότι οποιοδήποτε πρόβλημα αναφέρεται στη λογαριθμοκανονική κατανομή μπορούμε να το ανάγουμε (απλουστεύοντάς το) στην τυποποιημένη κανονική κατανομή. Για παράδειγμα έστω ότι η τυχαία μεταβλητή $X \sim L(\mu, \sigma^2)$ και θέλουμε να υπολογίσουμε την $P(\alpha < X < \beta)$. Εργαζόμαστε ως εξής:

$$P(\alpha < X < \beta) = P(\log \alpha < \log X < \log \beta) = P\left(\frac{\log \alpha - \mu}{\sigma} < \frac{\log X - \mu}{\sigma} < \frac{\log \beta - \mu}{\sigma}\right) \Rightarrow$$

$$P(\alpha < X < \beta) = \Phi\left(\frac{\log \beta - \mu}{\sigma}\right) - \Phi\left(\frac{\log \alpha - \mu}{\sigma}\right), \text{ όπου } \Phi(\cdot) \text{ η συνάρτηση κατανομής της } N(0,1) \text{ βλέπε π.χ. (Αθανασόπουλος Δ., 1991, σ. 177).}$$

Μέτρα θέσης και διασποράς

Στη συνέχεια, παρατίθενται οι ροπές μιας τυχαίας μεταβλητής X η οποία ακολουθεί τη λογαριθμοκανονική κατανομή και υπολογίζονται τα βασικά μέτρα θέσης και διασποράς. Οι αποδείξεις για τα παρακάτω δίδονται στη βιβλιογραφία π.χ. (Αθανασόπουλος Δ., 1991).

- $E(X^r) = \exp\left\{r\mu + \frac{1}{2}r^2\sigma^2\right\}$

Συνεπώς έχουμε:

1. $E(X) = e^{\mu + \frac{1}{2}\sigma^2}$
2. $V(X) = \exp\{2\mu + 2\sigma^2(e^{\sigma^2} - 1)\}$
3. $CV(X) = \sqrt{e^{\sigma^2} - 1}$

Από την παράμετρο σ^2 και μόνον εξαρτώνται επίσης, και αυξάνονται με αυτήν, οι συντελεστές ασυμμετρίας και κύρτωσης.

4. $skweness = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$
5. $kurtosis = e^{3\sigma^2}(e^{\sigma^2} + 2) + 3(e^{2\sigma^2} - 1)$

Από τις παραπάνω σχέσεις συμπεραίνεται ότι σε κάθε περίπτωση ο συντελεστής κύρτωσης είναι μεγαλύτερος του 3. Επομένως, η λογαριθμοκανονική κατανομή θα πρέπει να εφαρμόζεται σε εμπειρικές κατανομές, οι οποίες παρουσιάζουν όχι μόνο θετική ασυμμετρία αλλά και κύρτωση μεγαλύτερη της κανονικής κατανομής.

Επίσης, μια σχέση που ισχύει πάντα για την διάμεσο M , την μέση τιμή $E(X)$, και την επικρατούσα τιμή x_0 είναι η ακόλουθη:

$$x_0 < M < E(X)$$

Κεφάλαιο 4

Εύρεση της Κατανομής του ύψους και του πλήθους των αποζημιώσεων

Στο κεφάλαιο αυτό, θα προσπαθήσουμε να βρούμε ποια ή ποιες κατανομές προσαρμόζονται καλύτερα στα εμπειρικά δεδομένα της εταιρείας, δηλαδή στο ύψος και στο πλήθος των ζημιών. Στο δεύτερο κεφάλαιο παρουσιάστηκαν περιγραφικά στατιστικά στοιχεία τόσο για το πλήθος όσο και για το ύψος των αποζημιώσεων. Στο τρίτο κεφάλαιο μελετήσαμε ένα σύνολο από γνωστές κατανομές αναφέροντας τις βασικές ιδιότητές τους. Επίσης, υπολογίσαμε τα μέτρα θέσης και διασποράς για κάθε μια από αυτές.

Σε αυτό το κεφάλαιο καλούμαστε να επιλέξουμε ποια ή ποιες κατανομές έχουν κοινά χαρακτηριστικά με την κατανομή των εμπειρικών δεδομένων. Στη συνέχεια, για αυτές τις κατανομές, η διαδικασία που ακολουθείται είναι η εξής:

1. Κατασκευάζουμε ένα γράφημα με την συνάρτηση κατανομής των εμπειρικών δεδομένων και της συνάρτησης κατανομής της θεωρητικής κατανομής. Στη συνέχεια, παρατηρούμε γραφικά κατά πόσο τα δεδομένα μας προσαρμόζονται ή όχι σε αυτήν.
2. Αν η θεωρητική και η εμπειρική κατανομή δεν έχουν μεγάλες αποκλίσεις, εκτιμούμε τις παραμέτρους της θεωρητικής κατανομής. Για την εκτίμηση μπορεί να χρησιμοποιήσουμε διαφορετικές μεθόδους, όπως για παράδειγμα την μέθοδο των ροπών, της μέγιστης πιθανοφάνειας κ.α.
3. Τέλος, εκτελούμε στατιστικούς ελέγχους καλής προσαρμογής (X^2 τεστ, Kolmogorov-Smirnov test κ.α) προκειμένου να απορρίψουμε ή να μην απορρίψουμε την υπόθεση ότι τα δεδομένα μας προέρχονται από τη συγκεκριμένη θεωρητική κατανομή.

Η παραπάνω διαδικασία γίνεται τόσο για το ύψος όσο και για το πλήθος των εβδομαδιαίων και μηνιαίων ζημιών που έρχονται στην εταιρεία.

4.1 Εβδομαδιαία Ανάλυση

Για το σύνολο των κατανομών που αναφέρθηκαν στο Κεφάλαιο 3 καθώς και άλλων γνωστών κατανομών, που δεν αναφέρθηκαν όπως για παράδειγμα οι κατανομές Weibull, Pareto κ.α., εξετάζεται ποια ή ποιες από αυτές είναι πιο «κοντά» στα εμπειρικά δεδομένα.

Για μεγαλύτερη ευκολία από δω και στο εξής το εβδομαδιαίο ύψος των ζημιών που αναγγέλλονται στην εταιρεία θα συμβολίζεται με X^w και το πλήθος αντίστοιχα με N^w .

4.1.1 Ύψος Αποζημιώσεων

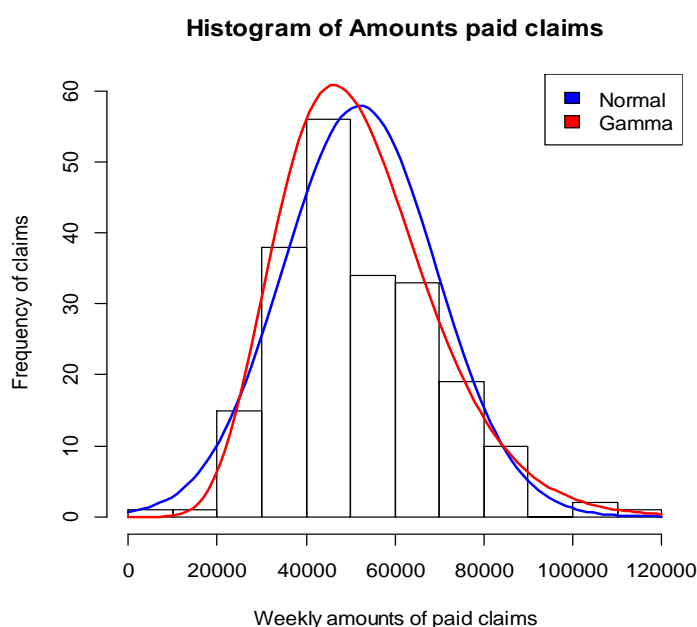
Παρατηρώντας το ιστόγραμμα στο Γράφημα 2.2 η μορφή του παραπέμπει στην κανονική κατανομή. Επομένως ελέγχουμε, αρχικά γραφικά, κατά πόσο η κανονική κατανομή είναι κοντά στα εμπειρικά δεδομένα. Επίσης, και η κατανομή Γάμμα φαίνεται ότι, με τις κατάλληλες παραμέτρους, μπορεί να ερμηνεύσει τα εμπειρικά δεδομένα.

Κατά συνέπεια, εκτιμούμε τις παραμέτρους⁴ της κατανομής Γάμμα και της κανονικής κατανομής και στη συνέχεια κατασκευάζουμε το ακόλουθο γράφημα

Distribution	Parameters		Method
Normal	mean = 52046.26	sd = 17179.97	Maximum Likelihood
Gamma	shape = 9.177712	rate = 0.000176	Matching moments

Πίνακας 4-1: Εκτίμηση των παραμέτρων της κανονικής και της Γάμμα κατανομής

Στην τρίτη στήλη του Πίνακα 4.1 αναφέρεται η μέθοδος με την οποία εκτιμήθηκαν οι παράμετροι. Το στατιστικό πακέτο R που χρησιμοποιήσαμε για τον υπολογισμό των παραμέτρων χρησιμοποιεί την μέθοδο της μέγιστης πιθανοφάνειας (maximum likelihood method) και την μέθοδο των ροπών (moments method). Η πρώτη δίνει πιο ακριβή αποτελέσματα αλλά σε περίπτωση που δεν μπορεί να εφαρμοστεί⁵ χρησιμοποιείται η μέθοδος των ροπών.



Γράφημα 4-1: Προσαρμογή κατανομών στο ύψος των εβδομαδιαίων ζημιών

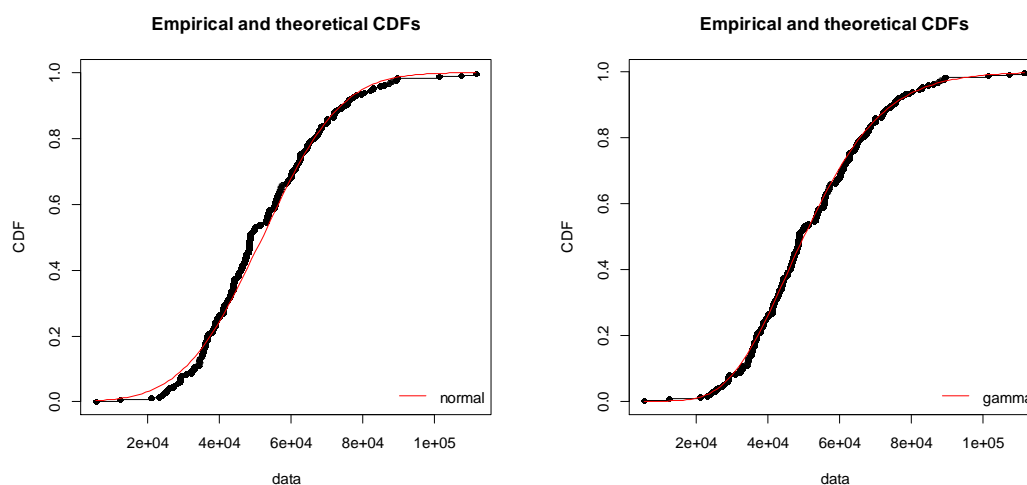
Από το παραπάνω γράφημα η κατανομή Γάμμα φαίνεται να προσαρμόζεται ικανοποιητικά, και καλύτερα από την Κανονική κατανομή, στα εμπειρικά δεδομένα καθώς το εμβαδόν που είναι εκτός της καμπύλης είναι μικρό. Αυτό θα ελεγχθεί στη συνέχεια με τη χρήση στατιστικών ελέγχων.

Επίσης, παρατηρούμε ότι το πηλίκο των παραμέτρων της κατανομής Γάμμα ισούται με τη μέση τιμή των εμπειρικών δεδομένων. Όπως έχει αναφερθεί στην παράγραφο 3.4, αυτό αποτελεί ένα βασικό στοιχείο ως προς την καταλληλότητα της κατανομής.

⁴ Για την εκτίμηση των παραμέτρων χρησιμοποιούμε το στατιστικό πακέτο R

⁵ Αυτό συμβαίνει σε περίπτωση που δεν υπάρχει γραμμική σχέση μεταξύ των παραμέτρων, και χρειάζεται εφαρμογή αριθμητικών μεθόδων για την λύση της εξίσωσης $\frac{\partial l(\theta;x)}{\partial \theta} = 0$.

Στη συνέχεια κατασκευάζουμε τα συγκριτικά γραφήματα της θεωρητικής και της εμπειρικής συνάρτησης κατανομής για την κανονική κατανομή και την κατανομή Γάμμα ξεχωριστά.



Γράφημα 4-2: Σύγκριση της συνάρτησης κατανομής των κατανομών Γάμμα και Κανονικής με αυτή των εμπειρικών δεδομένων

Από το Γράφημα 4.2 παρατηρούμε ότι η τυχαία μεταβλητή X^w προσαρμόζεται αρκετά καλύτερα στην κατανομή Γάμμα καθώς οι αποκλίσεις μεταξύ εμπειρικών και θεωρητικών δεδομένων είναι πολύ μικρότερες από τις αντίστοιχες αποκλίσεις της Κανονικής κατανομής.

Για να διαπιστώσουμε, και στατιστικά, σε ποια κατανομή προσαρμόζονται καλύτερα τα εμπειρικά δεδομένα υπολογίζουμε τα p-values με τον έλεγχο Kolmogorov-Smirnov, υπό την προϋπόθεση ότι τα δεδομένα μας ακολουθούν Γάμμα /Κανονική με τις παραμέτρους του Πίνακα 4.1.

Distribution	Kolmogorov-Smirnov test
Gamma	p-value = 0,7871
Normal	p-value = 0,081

Πίνακας 4-2: p-values του Kolmogorov-Smirnov test

Έστω, H_0 : Τα δεδομένα προέρχονται από την θεωρητική κατανομή

H_1 : Τα δεδομένα δεν προέρχονται από την θεωρητική κατανομή

Με κριτήριο απόρριψης της H_0 το p-value του ελέγχου Kolmogorov-Smirnov, σε επίπεδο σημαντικότητας 5%, δεν μπορούμε να απορρίψουμε την H_0 ούτε για την κατανομή Γάμμα, με παραμέτρους $shape = 9,17771$ και $rate = 0,00018$, ούτε για την Κανονική κατανομή, με παραμέτρους $mean = 52.046,26$ και $sd = 17.179,97$.

Ένα μειονέκτημα του Kolmogorov-Smirnov test, είναι ότι η κατανομή της στατιστικής συνάρτησης ελέγχου εξαρτάται από τις παραμέτρους του πληθυσμού. Έτσι αν δεν είναι εκ των προτέρων γνωστές οι παράμετροι και αναγκαστούμε να τις εκτιμήσουμε, όπως στην περίπτωσή μας, και να τις θέσουμε ως πραγματικές τιμές των παραμέτρων των αντίστοιχων ορισμάτων της συνάρτησης ks.test, τότε το αποτέλεσμα του ελέγχου δεν θα είναι αξιόπιστο (Αντζουλάκος Δ., 2011).

Προκειμένου να πάρουμε πιο αξιόπιστα αποτελέσματα, για την κατανομή Γάμμα θα χρησιμοποιήσουμε το χ^2 τεστ και για την κανονική κατανομή θα χρησιμοποιήσουμε τους ελέγχους Shapiro-Wilk και Lilliefors.

χ^2 τεστ (Chi Square test)

Κατασκευάζουμε τον ακόλουθο πίνακα:

Κλάσεις	O_i	E_i^G
[0 - 20.000)	2	1,802
[20.000 - 40.000)	53	52,248
[40.000 - 60.000)	90	94,825
[60.000 - 80.000)	52	47,333
[80.000 - άνω)	13	11,671

Πίνακας 4-3: Πλήθος θεωρητικών και πραγματικών τιμών σε κάθε κλάση

O_i : είναι ο παρατηρούμενος αριθμός απαιτήσεων για κάθε κλάση i , $i = 1, 2, \dots, 5$.

E_i^G : είναι ο αναμενόμενος αριθμός απαιτήσεων της Γάμμα κατανομής για την i κλάση.

Στη συνέχεια κάνουμε τους ακόλουθους υπολογισμούς:

Κλάσεις	$(O_i - E_i^G)^2$	$\frac{(O_i - E_i^G)^2}{E_i^G}$
[0 - 20.000)	0,039204	0,021756
[20.000 - 40.000)	0,565504	0,010823
[40.000 - 60.000)	23,28063	0,245511
[60.000 - 80.000)	21,78089	0,460163
[80.000 - άνω)	1,766241	0,151336
Σύνολο		0,88958

Πίνακας 4-4: Πίνακας υπολογισμών για τον έλεγχο χ^2

Έστω, $H_0 : F_o(x) = F_{Gamma}(x)$

$H_1 : F_o(x) \neq F_{Gamma}(x)$

Η απόρριψη ή μη απόρριψη της H_0 σε επίπεδο σημαντικότητας α , θα γίνει βάσει της τιμής

$\chi^2 = \sum_{i=1}^5 \frac{(O_i - E_i^G)^2}{E_i^G}$. Εάν $\chi^2 > \chi_{\kappa-m-1, \alpha}^2$ τότε απορρίπτεται η H_0 και συνεπώς τα δεδομένα

δεν προέρχονται από την θεωρητική κατανομή. Το κ δηλώνει το πλήθος των κλάσεων και το m το πλήθος των παραμέτρων της κατανομής.

Τα αποτελέσματα της εφαρμογής της παραπάνω διαδικασίας στα δεδομένα μας παρουσιάζονται στον πίνακα που ακολουθεί:

α	$\chi_{\kappa-m-1, \alpha}^2 = \chi_{2, \alpha}^2$	χ_G^2	Απόφαση
5%	5,991	0,88958	δεν απορρ. H_0
2,5%	7,378	0,88958	δεν απορρ. H_0
1%	9,210	0,88958	δεν απορρ. H_0

Πίνακας 4-5: Πίνακας αποτελεσμάτων για τον έλεγχο χ^2

Συγκρίνοντας τις τιμές της δεύτερης στήλης του Πίνακα 4.5, παρατηρούμε ότι η χ_G^2 είναι μικρότερη από την $\chi_{3, \alpha}^2$ για $\alpha = 5\%$.

Επομένως, δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση με αποτέλεσμα να δίνεται η δυνατότητα να θεωρήσουμε πως οι εβδομαδιαίες πληρωμές ακολουθούν την κατανομή Γάμμα με παραμέτρους $shape = 9,17771$ και $rate = 0,00018$.

Η ορθότητα της παραπάνω διαδικασίας εξασφαλίστηκε καθώς τηρούνται οι προϋπόθεσης του ελέγχου X^2 (βλέπε π.χ. Γναρδέλης Χ., 2009), δηλαδή το πλήθος των αναμενόμενων συχνοτήτων, βλέπε Πίνακα 4.3, είναι μεγαλύτερες της μονάδας και αυτές που είναι μικρότερες του 5 δεν υπερβαίνουν το 20% του συνόλου.

Τεστ Κανονικότητας (Test of Normality)

Στην παρούσα παράγραφο δίνονται έλεγχοι για να αξιολογήσουμε την υπόθεση της κανονικότητας του δείγματος, δηλαδή κατά πόσο η τυχαία μεταβλητή X^W προέρχεται από την κανονική κατανομή.

Πέραν από τον έλεγχο Kolmogorov-Smirnov, που προαναφέραμε και το βασικό του μειονέκτημα, θα χρησιμοποιήσουμε τον έλεγχο Shapiro-Wilk ο οποίος αν και προτείνεται γενικά για μικρά δείγματα ($n < 50$) συμπεριφέρεται καλά και για δείγματα με $n < 2000$. Η στατιστική συνάρτηση ελέγχου αποτιμά πόσο κοντά είναι τα εμπειρικά ποσοστιαία σημεία του δείγματος από τα αντίστοιχα θεωρητικά ποσοστιαία σημεία μιας κατάλληλης κανονικής κατανομής.

Για μεγαλύτερη αξιοπιστία θα πραγματοποιήσουμε και τον έλεγχο Lilliefors, ο οποίος χρησιμοποιεί προσομοίωση Monte Carlo και αποδίδει μια προσαρμοσμένη p – value.

Μετά από τις κατάλληλες εντολές που δόθηκαν⁶ στο στατιστικό πακέτο R παίρνουμε τα ακόλουθα αποτελέσματα:

Test of Normality

Shapiro-Wilk	p-value = 0.005107
Lilliefors	p-value = 0.0005324

Παρατηρούμε ότι, ενώ γραφικά είχαμε ενδείξεις ότι η τυχαία μεταβλητή X^W ακολουθεί την Κανονική κατανομή, οι τιμές των p-values είναι πολύ μικρές (μικρότερες από 1%) και συνεπώς απορρίπτεται η υπόθεση ότι τα εμπειρικά δεδομένα ακολουθούν την κανονική κατανομή.

⁶ Βλέπε παράρτημα με κώδικες

4.1.2 Πλήθος Ζημιών

Για το πλήθος των εβδομαδιαίων αναγγελιών που έρχονται στην εταιρεία, N^w , όπως είδαμε και από το ιστόγραμμα (Γράφημα 2.8), φαίνεται να κατανέμεται, σχετικά ομοιόμορφα, γύρω από το μέσο. Επομένως, αρχικά ελέγχουμε αν η τυχαία μεταβλητή N^w ακολουθεί την Κανονική κατανομή.

Για τον ισχυρισμό αυτό πραγματοποιούμε τους ελέγχους Shapiro-Wilk και Lilliefors.

Test of Normality

Shapiro-Wilk	p-value = 0.01732
Lilliefors	p-value = 0.185

Παρατηρούμε ότι, σε επίπεδο σημαντικότητας 5%, με τον έλεγχο Shapiro-Wilk απορρίπτουμε την υπόθεση ότι δεδομένα μας προέρχονται από την κανονική κατανομή. Ενώ με τον έλεγχο Lilliefors σε επίπεδο σημαντικότητας 5%, δεν μπορούμε να απορρίψουμε την υπόθεση ότι τα δεδομένα μας προέρχονται από την κανονική κατανομή.

Επομένως προχωράμε σε περαιτέρω στατιστικούς ελέγχους τα αποτελέσματα των οποίων παρουσιάζονται στον ακόλουθο πίνακα:

Test of Normality

Anderson-Darling	p-value = 0.07611
Cramer-von Mises	p-value = 0.1145
Pearson chi-square	p-value = 0.8377

Οι τιμές των p-value που προκύπτουν από τους παραπάνω ελέγχους είναι $\geq 5\%$. Συνεπώς σε επίπεδο σημαντικότητας 5%, δεν μπορούμε να απορρίψουμε την υπόθεση ότι τα δεδομένα μας προέρχονται από την κανονική κατανομή.

4.2 Μηνιαία Ανάλυση

4.2.1 Ύψος Αποζημιώσεων

Από τη μορφή του Γραφήματος 2.5 φαίνεται ότι τα εμπειρικά δεδομένα παρουσιάζουν δύο κορυφές. Μέχρι στιγμής, δεν έχει αναφερθεί κάποια θεωρητική κατανομή η οποία να παρουσιάζει δύο κορυφές προκειμένου να την συγκρίνουμε με το δείγμα μας. Συνεπώς, υπάρχουν δύο τρόποι να προχωρήσει η ανάλυση.

Ο πρώτος είναι, να προσαρμοστεί μια θεωρητική κατανομή γνωρίζοντας εκ των προτέρων ότι υπάρχουν αποκλίσεις, εμπειρικής και θεωρητικής, για τις τιμές που βρίσκονται ανάμεσα στις κορυφές. Στη συνέχεια να γίνει στατιστικός έλεγχος για να απορρίψουμε ή να μην απορρίψουμε ότι τα δεδομένα μας προέρχονται από αυτή.

Ο δεύτερος τρόπος, και πιο ορθός, είναι να χωρίσουμε το δείγμα σε δύο και στη συνέχεια να μελετήσουμε το κάθε δείγμα ξεχωριστά. Εάν επιτευχθεί καλή προσαρμογή, σε κάθε δείγμα, με μια από τις γνωστές κατανομές, τότε το επόμενο βήμα είναι να εκφραστεί η συνάρτηση πυκνότητας πιθανότητας του συνολικού δείγματος ως γραμμικός συνδυασμός των επιμέρους κατανομών.

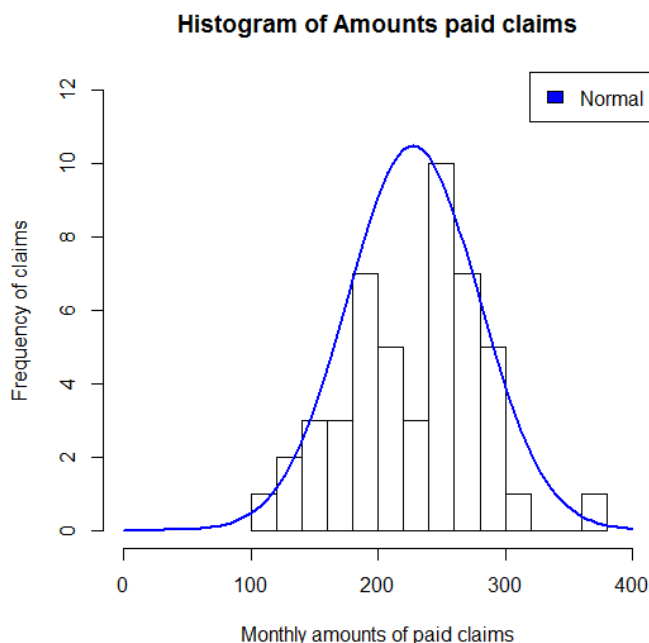
1^{ος} Τρόπος

Η κατανομή που φαίνεται να ταιριάζει στα δεδομένα είναι η Κανονική κατανομή. Για τον έλεγχο της ένδειξης αυτής, καταρχήν γίνεται εκτίμηση των παραμέτρων:

<i>Distribution</i>	<i>Parameter</i>
Normal	mean = 227.702,48 sd = 51.33314

Πίνακας 4-6: Εκτιμώμενες παράμετροι της Κανονικής κατανομής

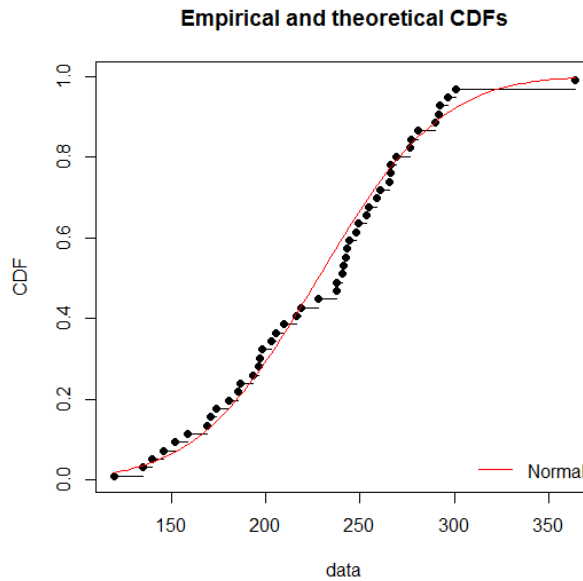
Μετά τον υπολογισμό των παραμέτρων ελέγχουμε τόσο γραφικά όσο και στατιστικά την ισχύ της υπόθεσής μας.



Γράφημα 4-3: Συγκριτικό ιστόγραμμα μηνιαίων ζημιών με την κανονική κατανομή

Από το παραπάνω γράφημα, παρατηρούμε ότι το ύψος των μηνιαίων αποζημιώσεων της εταιρείας φαίνεται να προσαρμόζεται ικανοποιητικά στην Κανονική κατανομή εκτός από δύο σημεία στα οποία δημιουργούνται μεγάλες αποκλίσεις.

Όπως αναμενόταν, οι αποκλίσεις παρουσιάζονται μεταξύ των δύο κορυφών. Επίσης, και οι ζημιές που ξεπερνούν τις 350 χιλιάδες ευρώ (μία τιμή) δεν μπορούν να περιγραφούν από την Κανονική κατανομή με $mean = 227.70248$ και $sd = 51.33314$. Παρόμοια εικόνα δίνεται και από το συγκριτικό διάγραμμα της εμπειρικής και της θεωρητικής (κανονική) κατανομής.



Γράφημα 4-4: Συγκριτικό διάγραμμα κατανομών

Στον παρακάτω πίνακα παρουσιάζονται οι τιμές των p-value για διαφορετικούς ελέγχους κανονικότητας των δεδομένων.

Test of Normality

Kolmogorov-Smirnov	p-value = 0.518
Shapiro-Wilk	p-value = 0.5608
Anderson-Darling	p-value = 0.4117
Pearson chi-square	p-value = 0.4735
Lilliefors	p-value = 0.09779

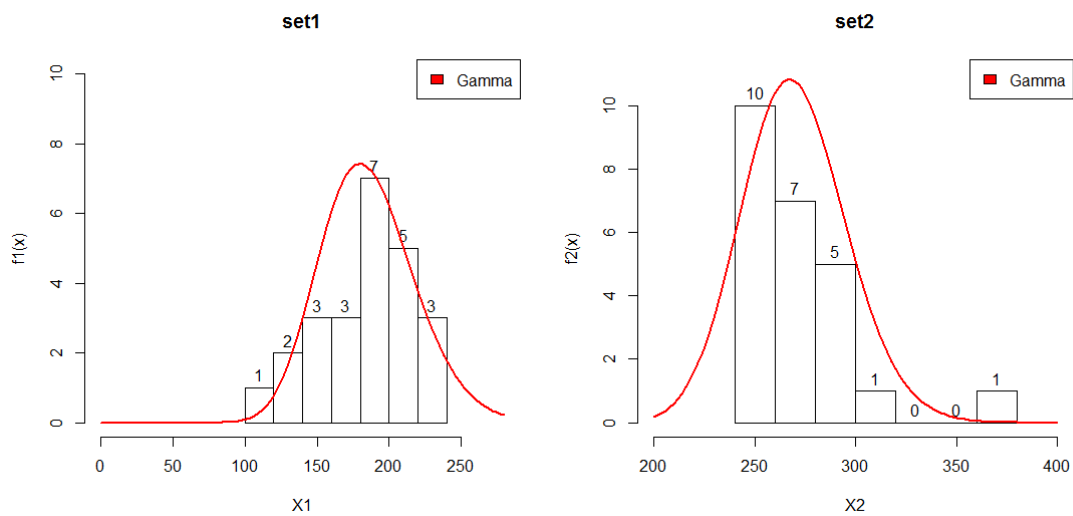
Πίνακας 4-7: Έλεγχος κανονικότητας

Είναι προφανές ότι δεν μπορούμε να απορρίψουμε, σε επίπεδο σημαντικότητας 5%, την υπόθεση ότι τα δεδομένα μας ακολουθούν την Κανονική κατανομή με μέση τιμή 227.70248 και τυπική απόκλιση ίση με 51.33914 καθώς όλες οι τιμές της δεύτερης στήλης είναι μεγαλύτερες από 5%.

Παρατηρούμε ότι τα τέσσερα πρώτα τεστ κανονικότητας δίνουν πολύ μεγαλύτερο p-value από τον έλεγχο του Lilliefors.

2^{ος} Τρόπος:

Ο δεύτερος τρόπος, είναι να χωρίσουμε το δείγμα στα δύο και να βρούμε την κατανομή που προσαρμόζεται στο κάθε δείγμα ξεχωριστά. Έστω $f_1(x)$ η συνάρτηση πυκνότητας του πρώτου δείγματος και $f_2(x)$ η συνάρτηση πυκνότητας του δεύτερου δείγματος. Μετά από διαδοχικές δοκιμές καταλήγουμε ότι το σημείο στο οποίο η πρώτη κατανομή αρχίζει να μην μπορεί να εξηγήσει την συμπεριφορά των ζημιών είναι το 240⁷. Επομένως, κατασκευάζουμε ένα ιστόγραμμα για τις ζημιές μέχρι το 240 και ένα ιστόγραμμα των ζημιών που είναι πάνω από 240 και προσαρμόζουμε σε αυτά γνωστές κατανομές. Για την επιλογή της κατάλληλης κατανομής εφαρμόζουμε μια σειρά από κατανομές που πιστεύουμε ότι είναι κατάλληλες. Στη συνέχεια γίνονται οι απαιτούμενοι στατιστικοί έλεγχοι μέσω των οποίων καταλήγουμε στην κατανομή που έχει την καλύτερη προσαρμογή.



Γράφημα 4-5: Εφαρμογή δύο διαφορετικών κατανομών Γάμμα στα εμπειρικά δεδομένα.

Όπως φαίνεται και στο παραπάνω Γράφημα, και για τα δύο δείγματα, καλύτερη προσαρμογή ενδέχεται να έχει η κατανομή Γάμμα.

Προκειμένου να βάλουμε στο γράφημα την συνάρτηση πυκνότητας πιθανότητας της κατανομής Γάμμα, και να παρατηρηθούν οι αποκλίσεις, έχει προηγηθεί η εκτίμηση των παραμέτρων της κατανομής για κάθε δείγμα (set 1 και set 2). Η εκτίμηση των παραμέτρων γίνεται με τη μέθοδο της μεγίστης πιθανοφάνειας και τα αποτελέσματα δίνονται στον παρακάτω πίνακα:

Distribution	Parameters	
	set 1	set 2
Gamma	shape = 32.299025 rate = 0.1739431	shape = 108.4179482 rate = 0.4019538

Πίνακας 4-8: Εκτίμηση παραμέτρων με τη μέθοδο μεγίστης πιθανοφάνειας

⁷σε χιλιάδες ευρώ

Στη συνέχεια εκτελείται ο στατιστικός έλεγχος Kolmogorov-Smirnov.

Distribution	p-value	
	set 1	set 2
Gamma	0.9118	0.8368

Παρατηρούμε ότι τα p-values του ελέγχου είναι αρκετά μεγάλα. Επομένως, δεν μπορούμε να απορρίψουμε ότι τα δεδομένα τόσο του set1 όσο και του set2 προέρχονται από την κατανομή Γάμμα με τις παραμέτρους του Πίνακα 4.9.

Άρα έχουμε:

- $f_1(x) = \frac{\beta_1^{\alpha_1} x^{\alpha_1-1} e^{-\beta_1 x}}{\Gamma(\alpha_1)}$ με μέση τιμή ίση με $E_1(X) = \int_{-\infty}^{\infty} x f_1(x) dx = \alpha_1 \beta_1$
- $f_2(x) = \frac{\beta_2^{\alpha_2} x^{\alpha_2-1} e^{-\beta_2 x}}{\Gamma(\alpha_2)}$ με μέση τιμή ίση με $E_2(X) = \int_{-\infty}^{\infty} x f_2(x) dx = \alpha_2 \beta_2$

Ισχύει,

$$f_{Total}(x) = w_1 f_1(x) + (1 - w_1) f_2(x) \text{ με } E_T(X) = \int_{-\infty}^{\infty} x f_{Total}(x) dx$$

$$\Rightarrow E_T(X) = \int_{-\infty}^{\infty} x [w_1 f_1(x) + (1 - w_1) f_2(x)] dx = w_1 \int_{-\infty}^{\infty} x f_1(x) dx + (1 - w_1) \int_{-\infty}^{\infty} x f_2(x) dx$$

$$\Rightarrow E_T(X) = w_1 E_1(X) + (1 - w_1) E_2(X) \Rightarrow E_T(X) = w_1 \alpha_1 \beta_1 + (1 - w_1) \alpha_2 \beta_2.$$

Οι παράμετροι $\alpha_2, \beta_2, \alpha_1, \beta_1$ είναι γνωστές. Συνεπώς, ο μόνος άγνωστος είναι το w_1 .

Για τον υπολογισμό του w_1 εξισώνουμε την θεωρητική μέση τιμή με την μέση τιμή του δείγματος.

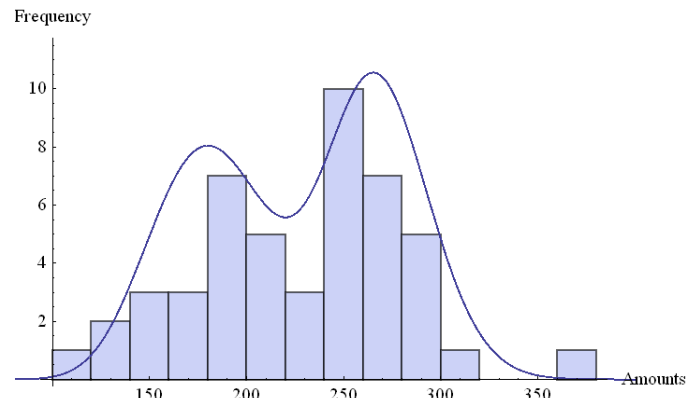
$$E_T(X) = E_{Ob}(X) \Rightarrow w_1 \alpha_1 \beta_1 + (1 - w_1) \alpha_2 \beta_2 = 227.70249 \Rightarrow w_1 = 0.500058 \cong 0.5$$

Επομένως, η συνάρτηση που προσεγγίζει τα εμπειρικά δεδομένα είναι μια μίξη δύο Γάμμα κατανομών και δίνεται από τον παρακάτω τύπο:

$$f_{Total}(x) = 6.32 \times 10^{-60} e^{-0.174x} x^{31.3} + 7.0228 \times 10^{-217} e^{-0.402x} x^{107.418}, x > 0 \quad (4.1)^8$$

⁸ Ο τρόπος υπολογισμού της σχέσης 4.1 έγινε με τη χρήση του Mathematica. Το αναλυτικό αποτέλεσμα παρουσιάζεται στο Παράρτημα II

Τέλος, δίνεται η γραφική παράσταση της $f_{Total}(x)$.



Γράφημα 4-6: Συνάρτηση πυκνότητας της μίξης Γάμμα κατανομών, όπου οι παράμετροι υπολογίστηκαν με τη μέθοδο της μέγιστης πιθανοφάνειας

Προκειμένου να εκτιμηθούν οι παράμετροι για το κάθε σετ χρησιμοποιήθηκε η μέθοδος μέγιστης πιθανοφάνειας. Άλλος ένας τρόπος να υπολογίσουμε τις παραμέτρους είναι η μέθοδος των ροπών.

Όμως, σε αυτή την περίπτωση θα πράξουμε διαφορετικά. Δεν θα χωρίσουμε το δείγμα μας στα δύο, αλλά θα εξισώσουμε τις πρώτες 5 ροπές των εμπειρικών δεδομένων με τις 5 πρώτες ροπές μιας τυχαίας μεταβλητής που ακολουθεί μίξη δύο κατανομών Γάμμα.

Επομένως, έχουμε:

Πέντε πρώτες θεωρητικές ροπές

- 1^η $\alpha_1 w_1 \theta_1 + (1 - w_1)\alpha_2\theta_2$
- 2^η $\alpha_1 (1 + \alpha_1)w_1\theta_1^2 + \alpha_2(1 + \alpha_2)(1 - w_1)\theta_2^2$
- 3^η $\alpha_1 (1 + \alpha_1)(2 + \alpha_1)w_1\theta_1^3 + \alpha_2 (1 + \alpha_2)(2 + \alpha_2)(1 - w_1)\theta_2^3$
- 4^η $\alpha_1 (1 + \alpha_1)(2 + \alpha_1)(3 + \alpha_1)w_1\theta_1^4 + \alpha_2 (1 + \alpha_2)(2 + \alpha_2)(3 + \alpha_2)(1 - w_1)\theta_2^4$
- 5^η $\alpha_1 (1 + \alpha_1)(2 + \alpha_1)(3 + \alpha_1)(4 + \alpha_1)w_1\theta_1^5 + \alpha_2 (1 + \alpha_2)(2 + \alpha_2)(3 + \alpha_2)(4 + \alpha_2)(1 - w_1)\theta_2^4$

Πέντε πρώτες ροπές δείγματος

- 1^η 227.702
- 2^η 54484.1
- 3^η $1.36017 * 10^7$
- 4^η $3.52256 * 10^9$
- 5^η $9.42491 * 10^{11}$

Από τη λύση του συστήματος⁹ παίρνουμε τα ακόλουθα αποτελέσματα:

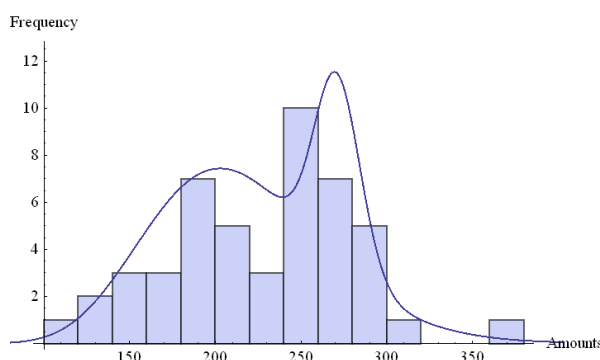
Παράμετρος	Τιμή
α_1	17.6079
θ_1	12.2152
α_2	417.128
θ_2	0.65129
w_1	0.777001

⁹ Η λύση του συστήματος έγινε με το Mathematica (βλέπε Παράρτημα II)

Άρα, η συνάρτηση πυκνότητας πιθανότητας δίνεται από τον τύπο:

$$f(x) = 4.86 \times 10^{-34} e^{-0.0818x} x^{16.6079} + 1.28 \times 10^{-834} e^{-1.53541x} x^{416.128}, \quad x > 0 \quad (4.2)$$

και η γραφική παράσταση είναι η εξής:



Γράφημα 4-7: Συνάρτηση πυκνότητας της μίξης Γάμμα κατανομών, όπου οι παράμετροι υπολογίστηκαν με τη μέθοδο των ροπών

Παρατηρούμε ότι, η σχέση 4.2 με τη σχέση 4.1 έχουν διαφορές. Οι διαφορές προκύπτουν αφενός, επειδή στη σχέση 4.1 έχει γίνει εκτίμηση των παραμέτρων με τη μέθοδο μεγίστης πιθανοφάνειας ενώ στην σχέση 4.2 έχει εφαρμοστεί η μέθοδος των ροπών, αφετέρου, όπως προαναφέραμε, για την εκτίμηση των παραμέτρων δεν έχει γίνει διαχωρισμός των δεδομένων και η εκτίμηση των παραμέτρων έχει γίνει με βάσει το σύνολο των παρατηρήσεων.

Στην συγκεκριμένη περίπτωση, δουλεύουμε με τα μηνιαία ύψη των αποζημιώσεων όπου το πλήθος των παρατηρήσεων είναι σχετικά μικρό (μόλις 48). Επομένως, χωρίζοντας το δείγμα τα αποτελέσματα (των ελέγχων) ενδέχεται να μην είναι αξιόπιστα.

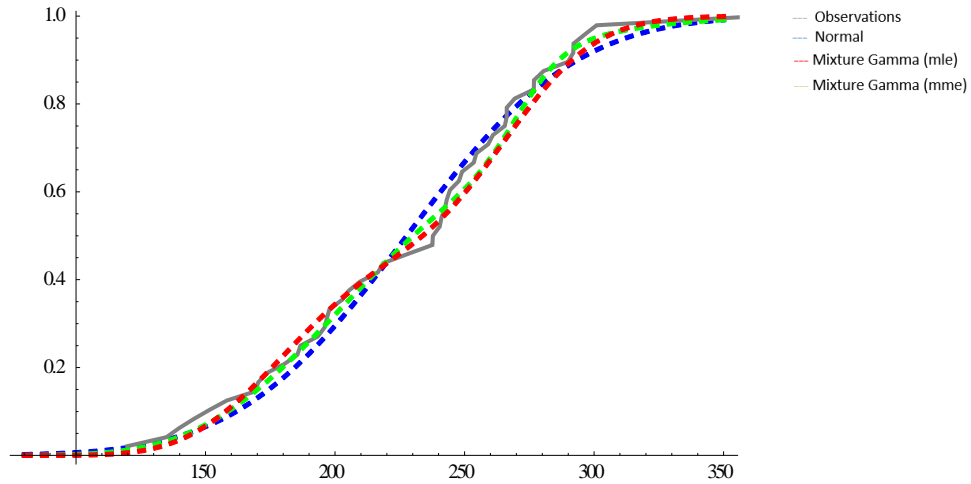
Είναι προφανές ότι, για την εκτίμηση ακραίων ζημιών η πυκνότητα που αντιστοιχεί στο Γράφημα 4.7 δίνει καλύτερες εκτιμήσεις, ενώ η πυκνότητα που προκύπτει με τη μέθοδο μεγίστης πιθανοφάνειας αποτυγχάνει να προβλέψει τις ζημιές αυτές.

Επίσης, ένας επιπλέον έλεγχος που μπορεί να γίνει είναι να υπολογιστούν οι παράμετροι με την μέθοδο των ροπών για κάθε δείγμα.

Στον παρακάτω πίνακα συνοψίζονται τα αποτελέσματα που διεξάχθηκαν για την κατανομή που ακολουθεί το μηνιαίο ύψος των αποζημιώσεων.

Κατανομή	Παράμετροι	Μέθοδος εκτίμησης παραμέτρων
Κανονική (F_1)	$mean = 227.702, sd = 51.33314$	Maximum Likelihood
Μίξη Γάμμα (F_2)	$\alpha_1 = 32.299025, \beta_1 = 1/0.1739,$ $a_2 = 108.41795, \beta_2 = 1/4019538,$ $w_1 = 0.5$	Maximum Likelihood
Μίξη Γάμμα (F_3)	$\alpha_1 = 17.6079, \beta_1 = 12.2152,$ $a_2 = 417.128, \beta_2 = 0.65129,$ $w_1 = 0.777001$	Matching Moments

Για κάθε μια από τις παραπάνω κατανομές, υπολογίζεται η συνάρτηση κατανομής και παρατίθενται με το παρακάτω γράφημα, μέσω των οποίων εφαρμόζεται το κριτήριο Kolmogorov-Smirnov.



Γράφημα 4-8: Σύγκριση της εμπειρικής συνάρτησης κατανομής με των θεωρητικών

Υπολογίζονται οι διαφορές μεταξύ της συνάρτησης κατανομής των δεδομένων και των θεωρητικών κατανομών, κατά απόλυτο τιμή, και παρουσιάζεται η μέγιστη διαφορά για κάθε μια κατανομή:

Kolmogorov-Smirnov		
$\sup F_{ob} - F_1 $	$\sup F_{ob} - F_2 $	$\sup F_{ob} - F_3 $
0.0969894	0.0681725	0.0597862

Πίνακας 4-9: Έλεγχος Kolmogorov – Smirnov

Παρατηρούμε ότι τη μικρότερη, κατά απόλυτη τιμή, μέγιστη διαφορά μεταξύ εμπειρικής και θεωρητικής κατανομής παρουσιάζει η F_3 . Επομένως, με το κριτήριο του Kolmogorov-Smirnov καταλήγουμε στο συμπέρασμα ότι οι μηνιαίες αποζημιώσεις ακολουθούν μια μίξη κατανομών Γάμμα με παραμέτρους $\alpha_1 = 17.6079$, $\beta_1 = 12.2152$, $\alpha_2 = 417.128$, $\beta_2 = 0.65129$, $w_1 = 0.777001$.

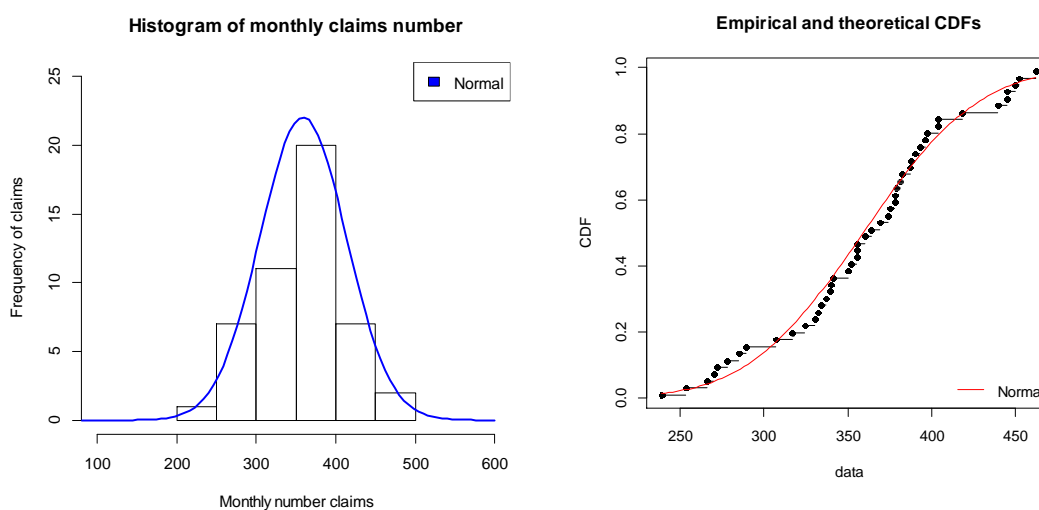
Τόσο το γράφημα όσο και ο παραπάνω πίνακας έγιναν με τη χρήση του υπολογιστικού προγράμματος Mathematica. Οι εντολές δίνονται αναλυτικά στο Παράρτημα II.

4.2.2 Πλήθος ζημιών

Σε αυτή την περίπτωση, η μορφή του ιστογράμματος του πλήθους των μηνιαίων ζημιών παραπέμπει στην Κανονική κατανομή. Αρχικά, υπολογίζονται οι παράμετροι και στη συνέχεια γίνεται γραφικός και στατιστικός έλεγχος καλής προσαρμογής των εμπειρικών δεδομένων στην Κανονική κατανομή.

<i>Distribution</i>	<i>Parameter</i>
Normal	mean = 358,96 sd = 54,37

Πίνακας 4-10: Εκτίμηση παραμέτρων της κανονικής κατανομής



Γράφημα 4-9: Γραφικός έλεγχος καλής προσαρμογής της κανονικής κατανομής στο μηνιαίο πλήθος των ζημιών

Test of Normality

Kolmogorov-Smirnov	p-value = 0.9686
Shapiro-Wilk	p-value = 0.3551
Anderson-Darling	p-value = 0.3773
Pearson chi-square	p-value = 0.1499
Lilliefors	p-value = 0.8064

Πίνακας 4-11: Έλεγχος κανονικότητας για το μηνιαίο πλήθος των αποζημιώσεων

Το πλήθος των μηνιαίων αναγγελθεισών ζημιών παρατηρούμε ότι ακολουθεί την κανονική κατανομή με μέση τιμή ίση με 358,96 και τυπική απόκλιση ίση με 54,37. Διαφορετικοί έλεγχοι πραγματοποιούνται προκειμένου να είμαστε σίγουροι ότι δεν απορρίπτεται η υπόθεση ότι τα δεδομένα μας ακολουθούν την κανονική κατανομή. Ενδιαφέρον παρουσιάζει το ότι το p-value του ελέγχου Lilliefors είναι αρκετά κοντά με το p-value του Kolmogorov-Smirnov. Αυτό πρακτικά σημαίνει ότι ο έλεγχος Kolmogorov-Smirnov δίνει αξιόπιστα αποτελέσματα. Οι υπόλοιποι τρεις έλεγχοι δίνουν μικρότερο p-value καθώς είναι πιο συντηρητικοί.

Κεφάλαιο 5

Ανάλυση με βάση το Μέσο Υπολειπόμενο Χρόνο Ζωής (Mean Residual Life)

Σε αυτή την ενότητα θα μελετήσουμε την συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής (*mean residual life function*) για τα μεγέθη των ζημιών. Παρόλο που η συνάρτηση αυτή χρησιμοποιείται συχνά στις ασφαλίσεις ζωής, και γενικότερα στην ανάλυση επιβίωσης, εμείς θα την χρησιμοποιήσουμε στις γενικές ασφαλίσεις.

Αρχικά ορίζουμε τα κάτωθι:

- X : Ορίζεται η μη-αρνητική συνεχής τ.μ. που αντιπροσωπεύει το ύψος της αποζημίωσης.
- $f(x)$: Ορίζουμε την συνάρτηση πυκνότητας πιθανότητας της τ.μ. X .
- $F(x)$: Ορίζουμε την συνάρτηση κατανομής της τ.μ. X .
- $S(x)$: Ορίζουμε την συνάρτηση επιβίωσης
- $h(x)$: Ορίζουμε την ένταση κινδύνου
- $m(x)$: Ορίζουμε την συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής (*MRL function*).

Η συνάρτηση κατανομής της τ.μ. X υπολογίζει την πιθανότητα, το ποσό της αποζημίωσης να πάρει τιμή μικρότερη ή ίση από την τιμή x και δίνεται από την επόμενη σχέση:

$$F(x) = P(X \leq x) = \int_0^x f(t)dt.$$

Η συνάρτηση επιβίωσης της τ.μ. X υπολογίζει την πιθανότητα η τ.μ. X να πάρει τιμή μεγαλύτερη από την τιμή x και δίνεται από την παρακάτω σχέση:

$$S(x) = P(X > x) = \int_x^\infty f(t)dt = 1 - F(x).$$

Η ένταση κινδύνου της τ.μ. X υπολογίζει την δεσμευμένη πιθανότητα το ύψος της αποζημίωσης να έχει μια τιμή περίπου ίση με x δοθέντος ότι η τ.μ. X είναι τουλάχιστον ίση με x και δίνεται από την παρακάτω σχέση:

$$h(x) = P(x < X < x + dx | X > x) = \frac{P\{(x < X < x + dx) \cap (X > x)\}}{P(X > x)}$$

$$\frac{P(x < X < x + dx)}{P(X > x)} = \frac{f(x)dx}{S(x)}, \quad x > 0$$

Η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής υπολογίζει την αναμενόμενη μέση αποζημίωση δοθέντος ότι η τιμή της αποζημίωσης είναι μεγαλύτερη από μια συγκεκριμένη τιμή x και δίνεται από την παρακάτω σχέση:

$$m(x) = E(X - x | X > x) = \frac{\int_x^\infty (t - x)f(t)dt}{S(x)} = \frac{\int_x^\infty S(t)dt}{S(x)} \text{ για κάθε } x > 0 \text{ και } E(X) < \infty.$$

5.1 Προϋποθέσεις – Βασικές Ιδιότητες

Στην ανάλυση επιβίωσης, η μελέτη του υπολειπόμενου χρόνου ζωής στοχεύει στην εκτίμηση του χρονικού περιθωρίου ζωής που έχει ένα μηχάνημα ή ένας οργανισμός. Στη δική μας περίπτωση, σε χαρτοφυλάκιο ζημιών, η συνάρτηση του μέσου υπολειπόμενου χρόνου ζωής χρησιμοποιείται προκειμένου να βρεθεί η κατανομή των αποζημιώσεων. Επομένως, από εδώ και στο εξής θα χρησιμοποιούμε τον όρο *μέση υπερβάλλουσα ζημιά*.

Σύμφωνα με τους (Shaked M. and Shanthikumar G., 2007, p. 82) μια μη-αρνητική συνάρτηση είναι συνάρτηση μέσης υπερβάλλουσας ζημιάς αν και μόνον αν τηρούνται οι ακόλουθες προϋποθέσεις.

- $0 \leq m(x) < \infty$ για όλα τα $x \geq 0$.
- $m(0) > 0$.
- $m(x)$ είναι συνεχής συνάρτηση.
- $m(x) + x$ είναι αύξουσα συνάρτηση στο $[0, \infty)$.
- Εάν υπάρχει x_0 τέτοιο ώστε $m(x_0) = 0$, τότε ισχύει $m(x) = 0$ για κάθε $x \geq x_0$.
Διαφορετικά, όταν δεν υπάρχει x_0 τέτοιο ώστε $m(x_0) = 0$, τότε $\int_0^{\infty} \frac{1}{m(x)} dx = \infty$.

5.2 Μελέτη Κατανομών Γάμμα

Στην παράγραφο αυτήν, παρουσιάζονται η συνάρτηση πυκνότητας πιθανότητας, η συνάρτηση επιβίωσης, η συνάρτηση κατανομής, η ένταση κινδύνου καθώς και η συνάρτηση της μέσης υπερβάλλουσας ζημιάς τόσο για την κατανομή Γάμμα όσο και για τη μίξη κατανομών Γάμμα. Επιπλέον, παρατίθενται τα γραφήματα των ανωτέρω συναρτήσεων για διαφορετικές τιμές των παραμέτρων κάθε κατανομής.

5.2.1 Κατανομή Γάμμα

Έστω μια μη - αρνητική τυχαία μεταβλητή που ακολουθεί την κατανομή Γάμμα, $X \sim \text{Γάμμα}(\alpha, \beta)$, όπου $\alpha > 0$ παράμετρος σχήματος και $\beta > 0$ παράμετρος κλίμακας. Η X έχει συνάρτηση πυκνότητας πιθανότητας την

$$f(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)}, \text{ όπως είδαμε στην ενότητα 3.4}$$

Η συνάρτηση επιβίωσης είναι

$$S(x) = Q\left(\alpha, \frac{x}{\beta}\right) = \frac{\Gamma\left(\alpha, \frac{x}{\beta}\right)}{\Gamma(\alpha)} \text{ με } \Gamma\left(\alpha, \frac{x}{\beta}\right) \text{ την συνάρτηση κατανομής της } X$$

και η ένταση κινδύνου δίνεται από τη σχέση,

$$h(x) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma\left(\alpha, \frac{x}{\beta}\right)}$$

Συνεπώς παρατηρούμε ότι ισχύουν τα ακόλουθα:

(α) η $h(x)$ είναι γνησίως φθίνουσα, όταν $0 < \alpha < 1$.

(β) η $h(x)$ είναι σταθερή, όταν $\alpha = 1$

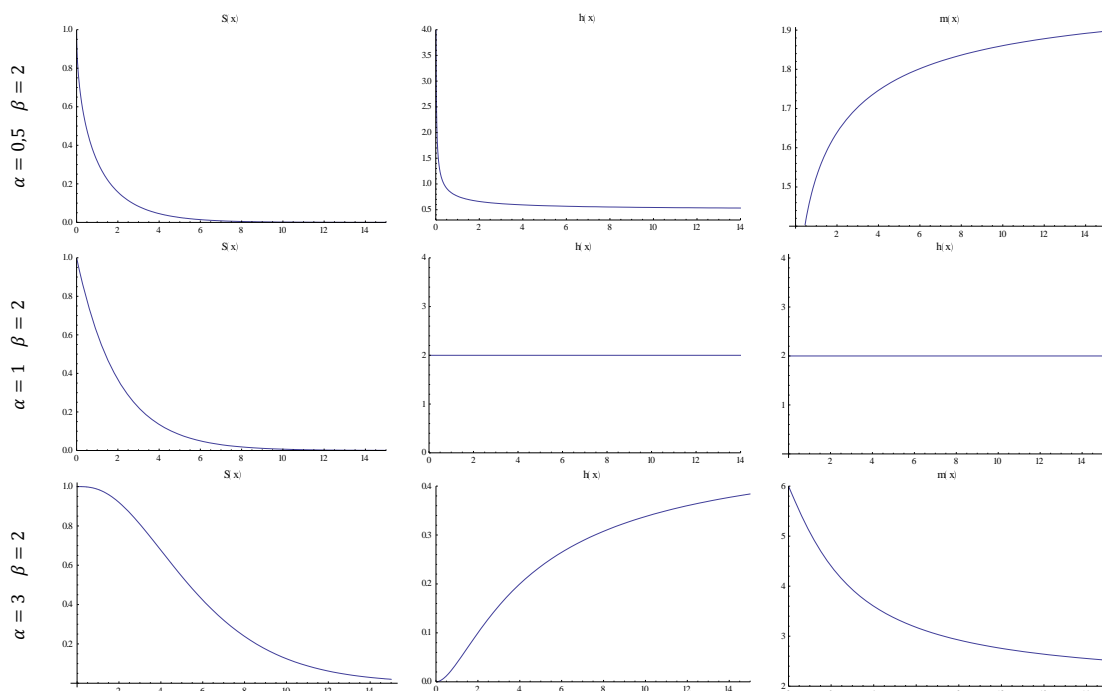
(γ) η $h(x)$ είναι γνησίως αύξουσα, όταν $\alpha > 1$

Η μέση υπερβάλλουσα ζημία υπολογίζεται από τη σχέση:

$$m(x) = \frac{x^\alpha e^{-\frac{x}{\beta}}}{\beta^{\alpha-1} \Gamma(\alpha) S(x)} + \beta\alpha - x$$

Η παραπάνω σχέση είναι αρκετά χρήσιμη καθώς, η συνάρτηση $m(x)$ σε πολλές περιπτώσεις δεν μπορεί να υπολογιστεί με τη σχέση $m(x) = \frac{\int_x^\infty S(t)dt}{S(x)}$. Η απόδειξη δίνεται αναλυτικά στην βιβλιογραφία (βλέπε π.χ. Rounor A., 2010).

Στη συνέχεια παρουσιάζονται οι γραφικές παραστάσεις των συναρτήσεων $S(x)$, $h(x)$ και $m(x)$ για την κατανομή Γάμμα.



Γράφημα 5-1: Γραφική απεικόνιση της συνάρτησης επιβίωσης, της έντασης κινδύνου και του μέσου υπολειπόμενου χρόνου ζωής για την κατανομή Γάμμα.

Το παραπάνω γράφημα, απεικονίζει την συνάρτηση επιβίωσης (αριστερά), την ένταση κινδύνου (κέντρο) και την μέση υπερβάλλουσα ζημία (δεξιά) της κατανομής Γάμμα για διαφορετικές τιμές της παραμέτρου α (σχήματος).

Όπως έχει προαναφερθεί, όταν το $0 < \alpha < 1$ η ένταση κινδύνου είναι γνησίως φθίνουσα. Παρατηρούμε, ότι η συνάρτηση της μέσης υπερβάλλουσας ζημίας είναι γνησίως αύξουσα. Το ακριβώς αντίθετο συμβαίνει για $\alpha > 1$. Ενώ για $\alpha = 1$ τόσο η μέση υπερβάλλουσα ζημία όσο και η συνάρτηση της έντασης κινδύνου, είναι σταθερές και ίσες με την παράμετρο κλίμακος, β . Η παράμετρος β δεν επηρεάζει την συμπεριφορά των συναρτήσεων $h(x)$ και $m(x)$.

Συνεπώς, παρατηρούμε ότι ισχύουν τα ακόλουθα:

- Όταν η $h(x)$ είναι γνησίως αύξουσα $\Rightarrow m(x)$ είναι γνησίως φθίνουσα (*DMRL*)
- Όταν η $h(x)$ είναι γνησίως φθίνουσα $\Rightarrow m(x)$ είναι γνησίως αύξουσα (*IMRL*)
- Όταν η $h(x)$ είναι σταθερή $\Rightarrow m(x)$ είναι σταθερή.

Επίσης, παρουσιάζονται δύο εναλλακτικοί τύποι υπολογισμού της μέσης υπερβάλλουσας ζημίας με την βοήθεια των οποίων κατασκευάστηκαν τα γραφήματα των συναρτήσεων $m(x)$ σε αυτό το κεφάλαιο.

- $$m(x) = \frac{\int_x^\infty tf(t)dt}{S(x)} - x$$
- $$m(x) = \frac{\mu - \int_0^x S(t)dt}{S(x)}$$

Επισημαίνεται ότι, όπου ήταν δυνατόν, η συνάρτηση $m(x)$ υπολογίστηκε και με τους τρεις τρόπους προς επαλήθευση των αποτελεσμάτων.

5.2.2 Μίξη κατανομών Γάμμα

Όπως είδαμε και στο 4^ο κεφάλαιο, οι αποζημιώσεις ενδέχεται να ακολουθούν μια μίξη κατανομών Γάμμα. Επίσης, ο (Lee C. K. S. and Lin, 2012) αναφέρουν ότι σε οποιοδήποτε χαρτοφυλάκιο ζημιών μπορεί να προσαρμοστεί μια μίξη κατανομών Erlang¹⁰. Επομένως, ενδιαφέρον παρουσιάζει η μελέτη της συνάρτησης κινδύνου και της συνάρτησης της μέσης υπερβάλλουσας ζημίας για μια μίξη κατανομών Γάμμα.

Στην υπο - ενότητα 5.2.1 αναφέρθηκαν κάποιοι γενικοί κανόνες που ισχύουν για την μορφή των συναρτήσεων $h(x)$ και $m(x)$ για διαφορετικές τιμές της παραμέτρου α . Στην περίπτωση της μίξης Γάμμα θα προσπαθήσουμε να δούμε αν μπορεί να γενικευτεί κάτι αντίστοιχο.

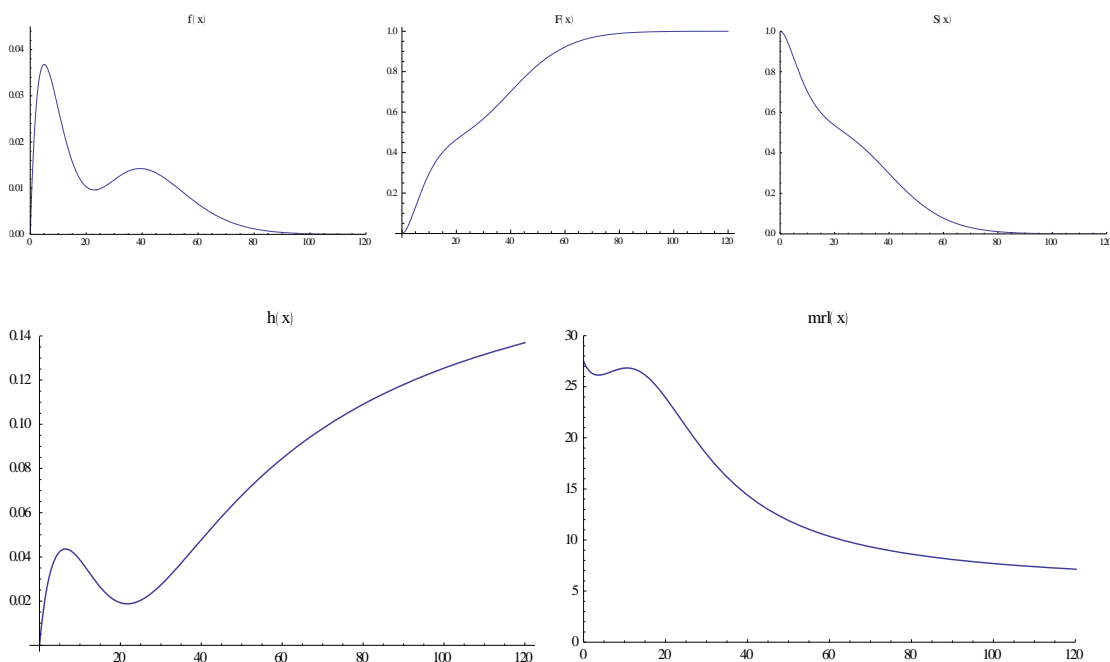
¹⁰ Ειδική περίπτωση της κατανομής Γάμμα όπου η παράμετρος α είναι ακέραιος.

Μίξη δύο κατανομών Γάμμα με παραμέτρους $\alpha_1, \alpha_2 > 1$ και β ίδιο

Έστω ότι X ακολουθεί μια μίξη κατανομών Γάμμα με συνάρτηση πυκνότητας πιθανότητας την,

$$f(x) = \frac{e^{-\frac{x}{\beta}} w_1 x^{-1+a_1} \beta^{-a_1}}{\Gamma(a_1)} + \frac{e^{-\frac{x}{\beta}} w_2 x^{-1+a_2} \beta^{-a_2}}{\Gamma(a_2)}, \quad x \geq 0$$

Για $a_1 = 3$, $a_2 = 9$, $\beta = 5$ και $w_1 = w_2 = 0.5$ υπολογίζονται οι συναρτήσεις $f(x), F(x), S(x), h(x), m(x)$ και παρατίθενται οι γραφικές παραστάσεις τους.

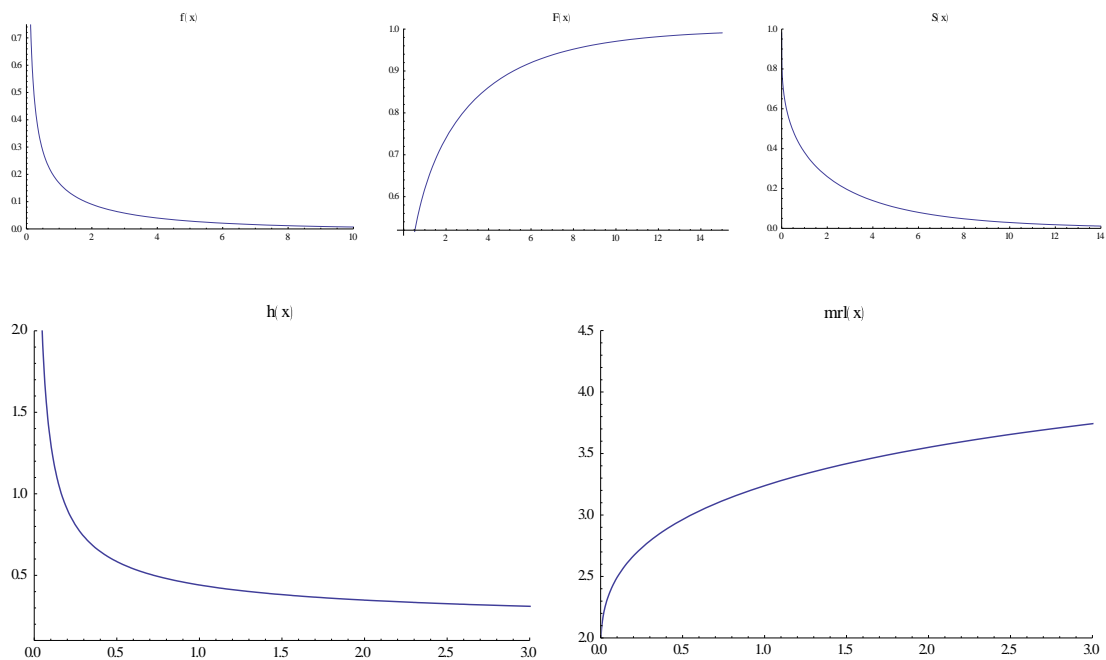


Γράφημα 5-2: Μίξη κατανομών Γάμμα με παραμέτρους $\alpha_1 = 3$, $\alpha_2 = 9$, $\beta = 5$ και $w_1 = w_2 = 0.5$

Στο παραπάνω γράφημα παρατηρείται ότι, η ένταση κινδύνου αλλάζει την μονοτονία της δύο φορές. Το ίδιο συμβαίνει, και στη συνάρτηση της μέσης υπερβάλλουσας ζημίας. Η μονοτονία των δύο αυτών συναρτήσεων έχει αντίστροφη συμπεριφορά. Δηλαδή, εκεί που η μια είναι αύξουσα, η άλλη είναι φθίνουσα και το αντίστροφο.

Μίξη δύο κατανομών Γάμμα με παραμέτρους $0 < \alpha_1, \alpha_2 < 1$ και β ίδιο

Για $a_1 = 0.5$, $a_2 = 0.2$, $\beta = 5$ και $w_1 = w_2 = 0.5$ υπολογίζονται οι συναρτήσεις $f(x), F(x), S(x), h(x), m(x)$ και παρατίθενται οι γραφικές παραστάσεις τους.



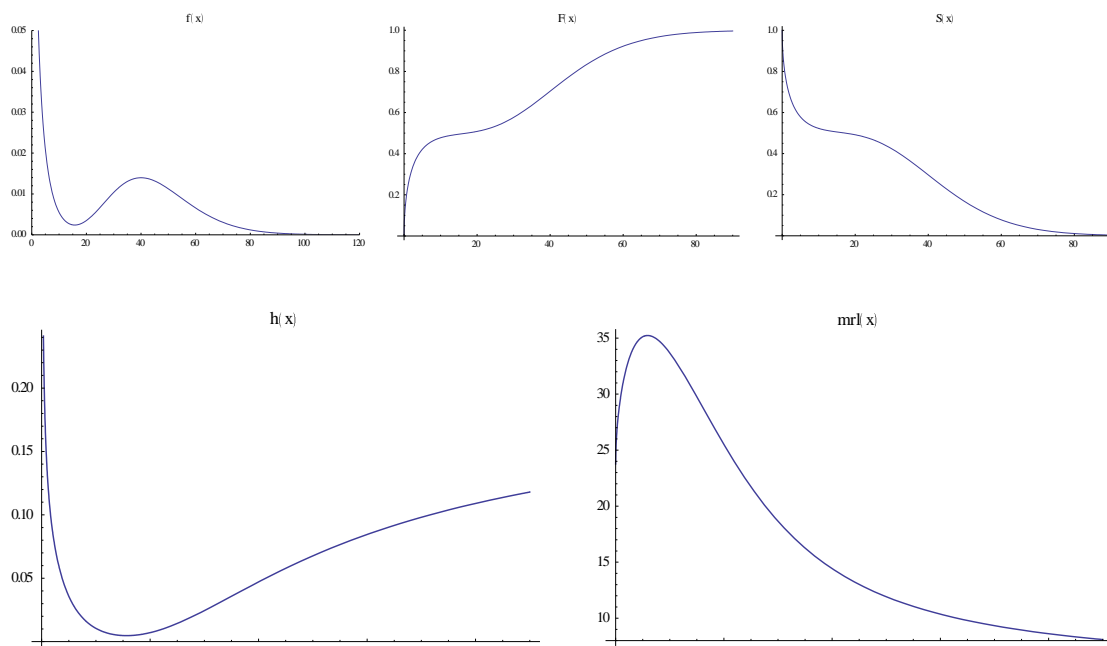
Γράφημα 5-3: Μίξη κατανομών Γάμμα με παραμέτρους $a_1 = 0.5$, $a_2 = 0.2$, $\beta = 5$ και $w_1 = w_2 = 0.5$

Από το παραπάνω γράφημα παρατηρούμε ότι τόσο η ένταση κινδύνου όσο και η συνάρτηση της μέσης υπερβάλλουσας ζημίας δεν αλλάζουν την μονοτονία τους. Συγκεκριμένα, η $h(x)$ είναι γνησίως φθίνουσα στο $(0, \infty)$ ενώ η συνάρτηση $m(x)$ είναι γνησίως αύξουσα.

Γενικά ισχύει ότι, για κάθε $0 < a_1, a_2 < 1$ η ένταση κινδύνου της μίξης δύο Γάμμα κατανομών είναι γνησίως φθίνουσα (βλέπε π.χ. Barlow R. and Proschan F.1996)

Μίξη δύο κατανομών Γάμμα με παραμέτρους $0 < \alpha_1 < 1$, $\alpha_2 > 1$ και β ίδιο

Για $\alpha_1 = 0.5$, $\alpha_2 = 9$, $\beta = 5$ και $w_1 = w_2 = 0.5$ υπολογίζονται οι συναρτήσεις $f(x)$, $F(x)$, $S(x)$, $h(x)$, $m(x)$ και παρατίθενται οι γραφικές παραστάσεις τους



Γράφημα 5-4: Μίξη κατανομών Γάμμα με παραμέτρους $\alpha_1 = 0.5$, $\alpha_2 = 9$, $\beta = 5$ και $w_1 = w_2 = 0.5$

Από το παραπάνω γράφημα, παρατηρείτε ότι η ένταση κινδύνου είναι γνησίως φθίνουσα μέχρι την τιμή 15 και στη συνέχεια γίνεται γνησίως αύξουσα. Αντίθετη συμπεριφορά παρουσιάζει η μονοτονία της συνάρτησης της μέσης υπερβάλλουσας ζημίας.

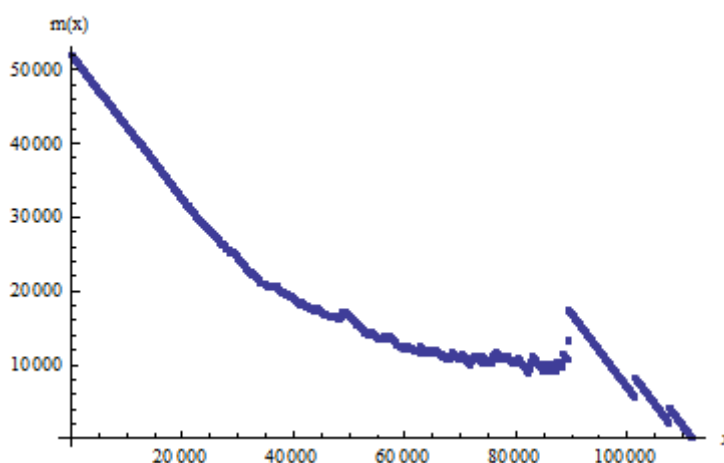
5.3 Μέση υπερβάλλουσα ζημιά – Εφαρμογή στα δεδομένα

Στην παράγραφο αυτή υπολογίζεται η μέση υπερβάλλουσα ζημιά στο εξεταζόμενο χαρτοφυλάκιο. Αναλυτικά, υπολογίζεται και συγκρίνεται με θεωρητικές κατανομές η μέση υπερβάλλουσα ζημιά τόσο για το ύψος των εβδομαδιαίων αποζημιώσεων όσο και για το ύψος των μηνιαίων και εξάγονται συμπεράσματα για τη μορφή της κατανομής των αποζημιώσεων.

Παράλληλα, ελέγχεται κατά πόσο οι εκτιμήσεις που έχουν γίνει για το ύψος των ζημιών επαληθεύονται γραφικά με τα αποτελέσματα που προκύπτουν από την ανάλυση της μέσης υπερβάλλουσα ζημιάς. Τέλος, παρουσιάζεται η μέση υπερβάλλουσα ζημιά του ατομικού ύψους των αποζημιώσεων.

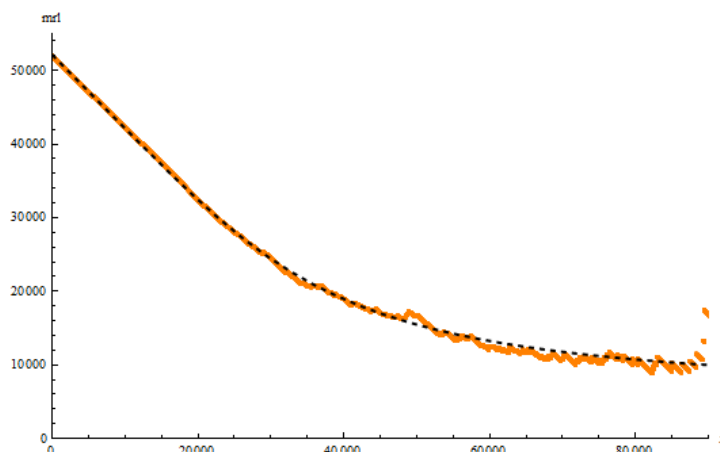
5.3.1 Εβδομαδιαίο ύψος αποζημιώσεων

Όπως είδαμε και στην ενότητα 4.2.1, η τυχαία μεταβλητή X^w που αντιπροσωπεύει τις εβδομαδιαίες αποζημιώσεις στο χαρτοφυλάκιο ακολουθεί *Γάμμα*(9.177, 05681.8). Συνεπώς, αναμένουμε ότι η μέση υπερβάλλουσα ζημιά της X^w θα είναι, σύμφωνα με όσα προαναφέραμε στην ενότητα 5.2.1, φθίνουσα.



Γράφημα 5-5: Συνάρτηση μέσου υπολειπόμενου χρόνου ζωής εβδομαδιαίων αποζημιώσεων

Παρατηρούμε ότι, η συνάρτηση της μέσης υπερβάλλουσας ζημιάς της X^w , μέχρι τις 90.000, έχει παρόμοια μορφή με αυτή που παρουσιάστηκε στο γράφημα 5.1 για $\alpha > 1$. Από τις 90.000 και πάνω το σχήμα χάνει την ομαλότητά του. Αυτό οφείλεται στις τρεις μεγαλύτερες ζημιές, οι οποίες έχουν μεγάλη απόκλιση από τις υπόλοιπες, και οι οποίες θα εξαιρεθούν στη συνέχεια. Πιο συγκεκριμένα, στα εμπειρικά δεδομένα, μετά την ζημιά με ποσό 90.000 ευρώ, η αμέσως επόμενη μεγαλύτερη είναι στα 101.000 ευρώ. Το ερώτημα που τίθεται, είναι αν η συνάρτηση της μέσης υπερβάλλουσας ζημιάς μιας κατανομής *Γάμμα* με παραμέτρους $shape = 9,17771$ και $rate = 0,00018$ θα είναι παρόμοια με αυτή της X^w . Δηλαδή, κατά πόσο η εκτίμηση που έγινε στην ενότητα 4.2.1 είναι καλή ή όχι.



Γράφημα 5-6: Σύγκριση εμπειρικού και θεωρητικού μέσου υπολειπόμενου χρόνου ζωής των εβδομαδιαίων αποζημιώσεων

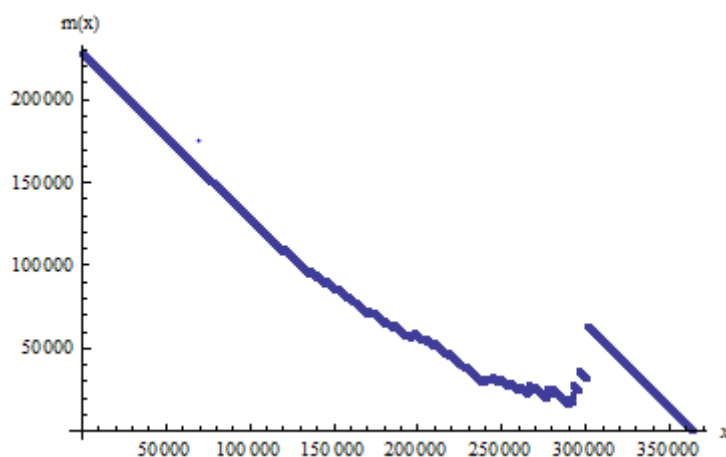
Το παραπάνω γράφημα απεικονίζει την συνάρτηση της μέσης υπερβάλλουσας ζημίας των εμπειρικών δεδομένων (πορτοκαλί γραμμή) και την συνάρτηση της μέσης υπερβάλλουσας ζημίας της κατανομής Γάμμα με παραμέτρους $shape = 9,17771$ και $rate = 0,00018$. Είναι φανερό ότι μέχρι τις 20.000 οι δύο καμπύλες σχεδόν ταυτίζονται. Από τις 40.000 μέχρι τις 60.000 οι δύο καμπύλες αποκλίνουν αισθητά σε ένα σημείο, αλλά γενικότερα είναι αρκετά κοντά. Από τις 60.000 μέχρι τις 80.000 οι δύο καμπύλες έχουν μια σταθερή διαφορά στα περισσότερα σημεία, με εξαίρεση στο τέλος. Από τις 80.000 και άνω οι δύο καμπύλες σχεδόν ταυτίζονται. Θα πρέπει να αναφέρουμε ότι, έχουν εξαιρεθεί οι τρεις μεγαλύτερες ζημιές.

Ο παραπάνω σχολιασμός για το Γράφημα 5-6 έχει ενδιαφέρον να συγκριθεί και με τα αποτελέσματα του Πίνακα υπολογισμών για τον έλεγχο X^2 (Πίνακας 4-4).

Συνεπώς, μέσω της συνάρτησης της μέσης υπερβάλλουσας ζημίας των εμπειρικών δεδομένων μπορούμε να εξάγουμε συμπεράσματα ως προς την μορφή της κατανομής που ακολουθούν οι εβδομαδιαίες ζημιές.

5.3.2 Μηνιαίο ύψος αποζημιώσεων

Στην ενότητα 4.2.1 είχαμε καταλήξει στο ότι το ύψος των αποζημιώσεων ακολουθεί μία μίξη κατανομών Γάμμα. Όμως, για την εκτίμηση των παραμέτρων καθώς και για τα αποτελέσματα των ελέγχων που έγιναν είμαστε αρκετά επιφυλακτικοί ως προς την αξιοπιστία τους, καθώς το δείγμα ήταν πολύ μικρό. Επομένως, κατασκευάζοντας τη συνάρτηση $m(x)$ για τα δεδομένα ελέγχεται κατά πόσο έχει παρόμοια μορφή με την $m(x)$ μιας μίξης Γάμμα (βλέπε Γράφημα 5-2).



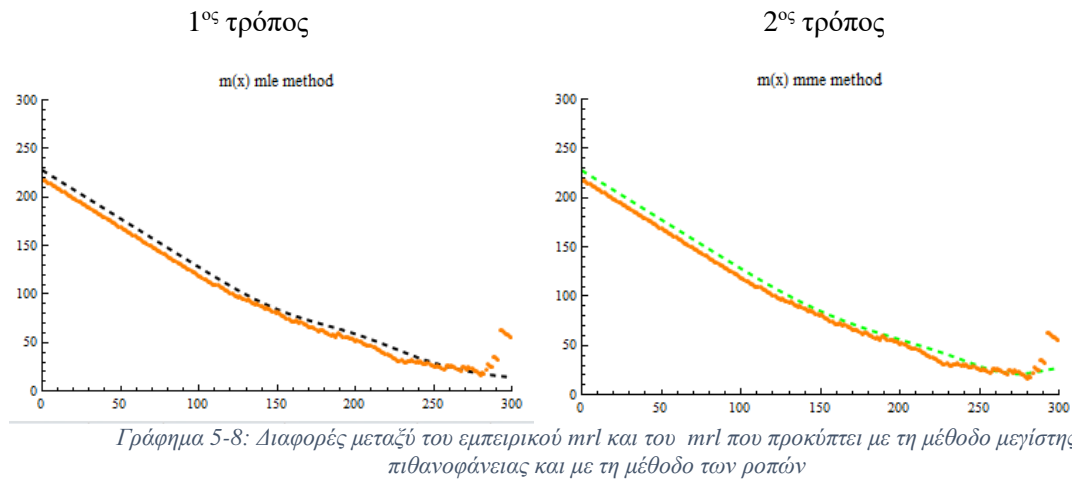
Γράφημα 5-7: Συνάρτηση μέσου υπολειπόμενου χρόνου ζωής των μηνιαίων ζημιών

Η μέση υπερβάλλουσα ζημία των μηνιαίων αποζημιώσεων έχει παρόμοια συμπεριφορά με τη μέση υπερβάλλουσα ζημία μιας μίξης Γάμμα με παραμέτρους $\alpha_1, \alpha_2 > 1$.

Για το μηνιαίο ύψος των αποζημιώσεων, έχουμε εκτιμήσει τις παραμέτρους $\alpha_1, \alpha_2, \beta_1, \beta_2$ και τα βάρη w_1, w_2 με δύο διαφορετικούς τρόπους. Στον πρώτο τρόπο, εκτιμήσαμε τις παραμέτρους με την μέθοδο μεγίστης πιθανοφάνειας και στη συνέχεια εξισώσαμε τις ροπές προκειμένου να βρούμε τα βάρη. Ο δεύτερος τρόπος είναι η μέθοδος των ροπών και για τις παραμέτρους και για τα βάρη (βλέπε παράγραφο 4.2.1).

Στη συνέχεια, παρατίθενται τα συγκριτικά διαγράμματα με τη συνάρτηση $m(x)$ των εμπειρικών δεδομένων και τη συνάρτηση $m(x)$ όπως υπολογίζεται για κάθε τρόπο ξεχωριστά.

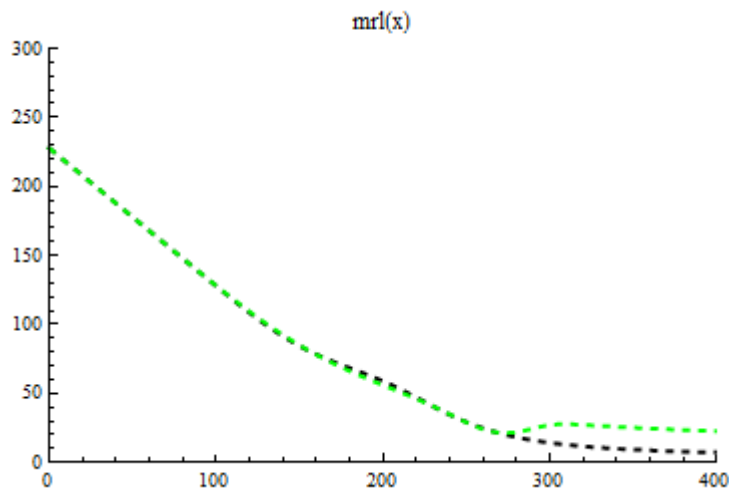
Σημειώνεται ότι, έχει αφαιρεθεί μία ζημία από τα εμπειρικά δεδομένα (η μέγιστη) η οποία επηρεάζει αρνητικά την ομαλότητα των αποτελεσμάτων.



Παρατηρούμε ότι, η συνάρτηση $m(x)$ που προκύπτει για τις παραμέτρους που εκτιμήθηκαν με τη μέθοδο των ροπών, είναι πιο κοντά στα εμπειρικά δεδομένα από την συνάρτηση $m(x)$ που προκύπτει για τις παραμέτρους που εκτιμήθηκαν με την μέθοδο της μεγίστης πιθανοφάνειας, αν και οι διαφορές είναι πολύ μικρές.

Συνεπώς, καταλήγουμε ότι το μηνιαίο ύψος των αποζημιώσεων ενδέχεται να ακολουθεί μια μίξη κατανομών Γάμμα με παραμέτρους $\alpha_1 = 17.6079, \alpha_2 = 417.128, \theta_1 = 12.2152, \theta_2 = 0.65129, w_1 = 0.77001$.

Τέλος, συγκρίνονται και οι συναρτήσεις των μέσων υπερβαλλουσών ζημιών, $m(x)$, για τις μίξεις Γάμμα με τις παραμέτρους που εκτιμήθηκαν σύμφωνα με τη μέθοδο των ροπών (πράσινο) και με τις παραμέτρους που εκτιμήθηκαν σύμφωνα με τη μέθοδο της μεγίστης πιθανοφάνειας (μαύρο).



Γράφημα 5-9: Σύγκριση της συνάρτησης mrl με τους δύο διαφορετικούς τρόπους εκτίμησης των παραμέτρων

Γραμμική μέση υπερβάλλουσα ζημιά

Στην παρακάτω υπο-ενότητα παρουσιάζεται ένας εναλλακτικός τρόπος μελέτης συνάρτησης της μέσης υπερβάλλουσας ζημιάς όταν υπάρχει υποψία εμφάνισης γραμμικότητας.

Στο Γράφημα 5-7 εμφανίζεται η σχέση μεταξύ της μέσης υπερβάλλουσας ζημιάς των μηνιαίων αποζημιώσεων και της τυχαίας μεταβλητής X που αναφέρεται ως ένα ποσό ζημιάς, δηλαδή βλέπουμε ότι για κάθε τιμή της X εμφανίζεται η αντίστοιχη τιμή του $m(x)$. Από το εν λόγω διάγραμμα, παρατηρούμε ότι τα δύο αυτά ποσά έχουν αντίστροφη σχέση.

Επιπρόσθετα από το Γράφημα 5-7, θα μπορούσε κανείς να πει ότι τα δεδομένα ενδέχεται να μπορούν να περιγραφούν μέσω μίας γραμμικής σχέσης, για παράδειγμα ως $m(x) = ax + \beta$ και μια πρώτη σκέψη είναι η εφαρμογή ενός μοντέλου απλής γραμμικής παλινδρόμησης. Μέσω αυτής της διαδικασίας, θα μπορούσε κανείς να περιγράψει τα δεδομένα δημιουργώντας ένα μοντέλο προβλεπτικής ικανότητας καθώς και να εκτιμήσει τις παραμέτρους a και β (ειδικότερα για τις τιμές από 0 έως 100.000).

Η ιδέα περιγραφής των δεδομένων από ένα γραμμικό μοντέλο δεν μπορεί να εφαρμοστεί καθώς δεν τηρούνται οι βασικές προϋποθέσεις της απλής γραμμικής παλινδρόμησης.

Εναλλακτικά, αναφέρεται στη βιβλιογραφία, ότι μία τέτοιου είδους γραμμική προσαρμογή μπορεί να γίνει χρησιμοποιώντας το μοντέλο των (Hall W. J. and Wellner J. A., 1981)

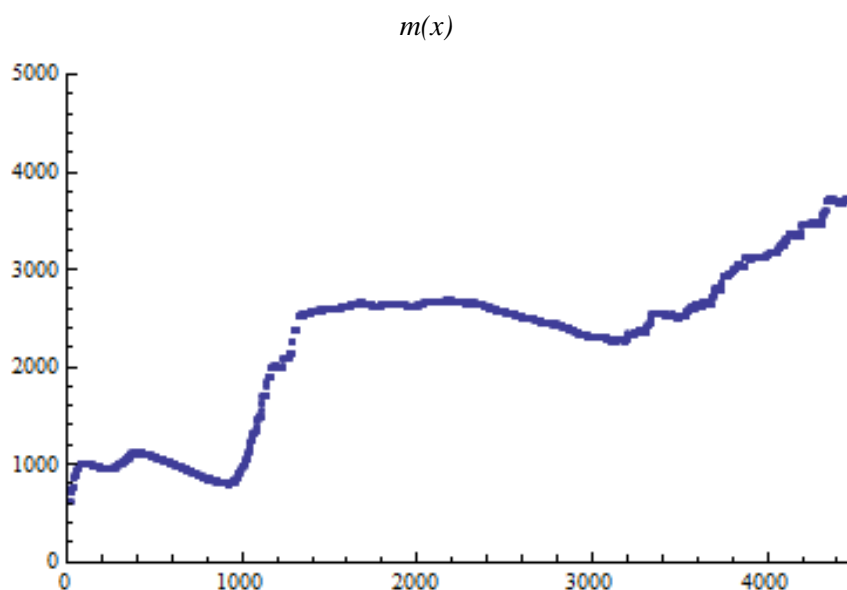
Σε αυτή την περίπτωση, χρησιμοποιώντας την σχέση (Jeong H., 2014, p. 21):

$$S(x) = \left(\frac{\beta}{ax + \beta} \right)_+^{1+\frac{1}{\alpha}}$$

μπορούμε να υπολογίσουμε την συνάρτηση επιβίωσης των μηνιαίων αποζημιώσεων και στην συνέχεια να βρεθεί και ένας τύπος για την συνάρτηση πυκνότητας.

5.4 Ατομικά ύψη ζημιών

Στο παρακάτω γράφημα παρουσιάζεται η μορφή της μέσης υπερβάλλουσας ζημιάς των ατομικών ζημιών.



Από το γράφημα φαίνεται, ότι δεν μπορούμε να προσεγγίσουμε την $m(x)$ με κάποια γνωστή κατανομή για το σύνολο των τιμών των ζημιών. Παρατηρούμε όμως ότι για υποσύνολα των τιμών του x , η $m(x)$ φαίνεται να έχει παρόμοια μορφή με τη $m(x)$ γνωστών κατανομών. Επομένως χωρίζοντας τα δεδομένα στα σημεία που αλλάζει η μονοτονία της $m(x)$ και δημιουργώντας επί μέρους σετ, θα μπορούσαμε να εφαρμόσουμε ελέγχους καλής προσαρμογής γνωστών κατανομών σε αυτά.

Μια τέτοια ανάλυση, ξεφεύγει από τα πλαίσια της παρούσας εργασίας αλλά θα μπορούσε να αποτελέσει το θέμα νέας εργασίας, ως συνέχεια της παρούσας, με στόχο την εύρεση μιας πολύκλαδης συνάρτησης πυκνότητας πιθανότητας των εμπειρικών δεδομένων μέσω της $m(x)$.

Κεφάλαιο 6

Συμπεράσματα

Στην παρούσα μελέτη αναπτύχθηκε μοντέλο πρόβλεψης ζημιών τόσο για το πλήθος όσο και για το ύψος των υλικών ζημιών του κλάδου αστικής ευθύνης αυτοκινήτου. Παρατηρήθηκε ότι ο διαχωρισμός σε σωματικές βλάβες και υλικές ζημιές για το χαρτοφυλάκιο ζημιών αυτοκινήτων είναι αναγκαίος καθώς η συμπεριφορά των αποζημιώσεων είναι πολύ διαφορετική ανά κάλυψη.

Όσον αφορά τις υλικές ζημιές της εταιρείας, που αναγγέλλονται σε εβδομαδιαία βάση, το ποσό της αποζημίωσης ακολουθεί την κατανομή Γάμμα με παραμέτρους $shape = 9,17771$ και $rate = 0,00018$ ενώ το πλήθος ακολουθεί κανονική κατανομή.

Όσον αφορά τις υλικές ζημιές της εταιρείας, που αναγγέλλονται στην εταιρεία σε μηνιαία βάση, το ποσό της αποζημίωσης μπορεί να περιγραφεί από τρεις διαφορετικές κατανομές όπως παρατίθενται στη συνέχεια:

Κατανομή	Παράμετροι	Μέθοδος εκτίμησης παραμέτρων
Κανονική (F_1)	$mean = 227.702, sd = 51.33314$	Maximum Likelihood
Μίξη Γάμμα (F_2)	$\alpha_1 = 32.299025, \beta_1 = 1/0.1739,$ $a_2 = 108.41795, \beta_2 = 1/0.4019538,$ $w_1 = 0.5$	Maximum Likelihood
Μίξη Γάμμα (F_3)	$\alpha_1 = 17.6079, \beta_1 = 12.2152,$ $a_2 = 417.128, \beta_2 = 0.65129,$ $w_1 = 0.777001$	Matching Moments

Επιπλέον, ο προτεινόμενος εναλλακτικός τρόπος παρακολούθησης της συμπεριφοράς των ζημιών, με τη μελέτη της μέσης υπερβάλλουσας ζημιάς, έδωσε παρόμοια συμπεράσματα ως προς το ποια είναι η κατάλληλη κατανομή για την περιγραφή της συμπεριφοράς του ύψους των μηνιαίων και εβδομαδιαίων αποζημιώσεων.

Η μελέτη του χαρτοφυλακίου με τη μεθοδολογία που εφαρμόστηκε στην εργασία αυτή, προσφέρει πολλές και χρήσιμες πληροφορίες για την συμπεριφορά των υλικών ζημιών και τη διαχείριση και αντιμετώπιση του κινδύνου που πηγάζει από αυτές. Για παράδειγμα μια εταιρεία μπορεί να εκτιμήσει το μέσο κόστος ζημιάς του επόμενου μήνα, την πιθανότητα να έρθει ζημιά μεγαλύτερη από ένα συγκεκριμένο ποσό καθώς και να διαφοροποιήσει τον τρόπο διακανονισμού και γενικότερης διαχείρισης των ζημιών και όλων των εμπλεκόμενων σε αυτές (συνεργεία, πραγματογνώμονες, διακανονιστές κλπ.). Τέλος, μπορεί να λάβει χρήσιμες πληροφορίες για την τιμολόγηση ενός νέου προϊόντος.

Παράρτημα I

Στο Παράρτημα I παρουσιάζονται οι εντολές που δόθηκαν στο στατιστικό πακέτο R προκειμένου να εξαχθούν τα αποτελέσματα της εργασίας.

Γραφήματα 2^ο κεφαλαίου

Γράφημα 2.1 και 2.2

```
➤ X<-X/1000
➤ par(mfrow = c(1,2))
➤ boxplot(X,ylim=c(0,120),main="Boxplot of weekly amounts paid claims")
➤ hist(X,xlim=c(0,120),ylim=c(0,60), xlab="Weekly amounts of paid claims",
ylab="Frequency of claims", labels=TRUE, main="Histogram of Amounts paid claims")
➤ par(mfrow = c(1,1))
➤ ts.plot(X,ylim=c(0,120),xlab="Week",ylab="Amounts paid",main="Time series plot")
```

Οι παραπάνω εντολές δίνουν τα γραφήματα 2.1 και 2.2. Αντίστοιχες είναι και οι εντολές που δόθηκαν για την κατασκευή των γραφημάτων 2.3 μέχρι 2.8 που αφορούν το πλήθος των εβδομαδιαίων αποζημιώσεων και της μηνιαίας ανάλυσης.

Γραφήματα 4^ο κεφαλαίου

Γράφημα 4.1

```
➤ hist(X,xlim=c(0,120000), ylim=c(0,60), xlab="Weekly amounts of paid claims",ylab="Frequency of claims", main = "Histogram of Amounts paid claims", labels=TRUE)
➤ curve(2500000*dnorm(x,mean=52046.28,sd=17211),from=0,to=120000,add=TRUE,col="blue",lwd=2)
➤ curve(2500000*dgamma(x,shape=9.1777122342,rate=0.0001763376),from=0,to=120000,add=TRUE,col="red",lwd=2)
➤ legend("topright",c("Normal","Gamma"),fill=c("blue","red",horiz=T))
```

Γράφημα 4.2

```
➤ fn<-fitdist(X, "name of distribution")
➤ cdfcomp(fn,legendtext="name of distribution")
```

Σημειώνεται ότι, με την εντολή `fitdist` έχουν εκτιμηθεί οι παράμετροι της κατανομής που έχει εισαχθεί στο «name distribution». Για την εμφάνιση των παραμέτρων εκτελείτε η εντολή:

```
➤ summary (fn)
```

Τα υπόλοιπα γραφήματα του 4^ο κεφαλαίου γίνονται με αντίστοιχες εντολές κατά περίπτωση.

Έλεγχοι 4ου κεφαλαίου

Οι εντολές που δόθηκαν για να πραγματοποιηθούν οι έλεγχοι καλής προσαρμογής είναι οι ακόλουθες:

- `ks.test(X,pnorm,mean=52046.26,sd=17179.97,exact=FALSE)`
- `shapiro.test(X)`
- `ad.test(X)`
- `cvm.test(X)`
- `pearson.test(X)`
- `lillie.test(X)`

Για τον έλεγχο του Kolmogorov Smirnov πρέπει να εισαχθούν και οι παράμετροι. Επίσης, αντί για `pnorm` μπορεί να μπει το όνομα κάποιας άλλης κατανομής, όπως για παράδειγμα `pgamma`, `rexp` και ούτω κάθε εξής. Οι υπόλοιποι έλεγχοι αφορούν αποκλειστικά έλεγχο κανονικότητας.

Για τον έλεγχο X^2 φτιάχτηκε η ακόλουθη ρουτίνα:

- `#Classes`
- `a=20000;`
- `b=40000;`
- `c=60000;`
- `d=80000;`
- `e=100000;`
- `#Observation (X)`
- `O1<-length(X[X<a])`
- `O2<-length(X[X>=a &X<b])`
- `O3<-length(X[X>=b &X<c])`
- `O4<-length(X[X>=c &X<d])`
- `O5<-length(X[X>d])`
- `#Distribution (Gamma)`
- `shape=9.1777122342;rate=0.0001763376`
- `eg1<-(pgamma(a,shape,rate))*length(X)`
- `eg2<-(pgamma(b,shape,rate)-pgamma(a,shape,rate))*length(X)`
- `eg3<-(pgamma(c,shape,rate)-pgamma(b,shape,rate))*length(X)`
- `eg4<-(pgamma(d,shape,rate)-pgamma(c,shape,rate))*length(X)`
- `eg5<-(1-pgamma(d,shape,rate))*length(X)`

Μετά την εκτέλεση των παραπάνω εντολών κατασκευάζουμε τον πίνακα 4.3 προκειμένου να εφαρμόσουμε τον έλεγχο X^2 . Προτού ξεκινήσει η κατασκευή γραφημάτων και ελέγχων στο στατιστικό πακέτο R, μελετήθηκε το Package 'fitdistrplus' July 2, 2016.

Παράρτημα II

Στο Παράρτημα II παρουσιάζονται οι εντολές που δόθηκαν στο αλγεβρικό πακέτο Wolfram Mathematica στην παρούσα εργασία.

Υπολογισμοί 4^{ου} κεφαλαίου

Για τον υπολογισμό της σχέσης 4.2 εκτελέστηκαν οι παρακάτω εντολές:

```
α1 = 32.299025; α2 = 108.4179482; β1 = 1/0.1739431; β2 = 1/0.4019538; w1 = 0.5; w2 = 1 - w1;  
Γ[α_] = Integrate[t^(α - 1) * Exp[-t], {t, 0, Infinity}, Assumptions -> Re[α] > 0];  
f1[x_] = (x^(α1 - 1) * Exp[-(x/β1)]) / (β1^α1 * Γ[α1]);  
f2[x_] = (x^(α2 - 1) * Exp[-(x/β2)]) / (β2^α2 * Γ[α2]);  
f[x_] = w1 * f1[x] + w2 * f2[x]  
6.32688 × 10-60 e-0.173943x x31.299 + 7.02276 × 10-217 e-0.401954x x107.418
```

Στη συνέχεια, για επαλήθευση υπολογίζεται η μέση τιμή και το ολοκλήρωμα από 0 μέχρι άπειρο για την μίξη Γάμμα.

```
Integrate[f[x], {x, 0, Infinity}, Assumptions -> x > 0]  
Integrate[x * f[x], {x, 0, Infinity}, Assumptions -> x > 0]  
1.  
227.707
```

Για να υπολογιστεί η συνάρτηση κατανομής, καθώς και για την κατασκευή των Πινάκων εκτελούνται οι ακόλουθες εντολές:

```
cdf[t_] = Integrate[f[x], {x, 0, t}, Assumptions -> x > 0]  
1.95884 × 10-61 t32.299 Hypergeometric1F1[32.299, 33.299, -0.173943 t] + 6.47749 × 10-219 t108.418 Hypergeometric1F1[108.418, 109.418, -0.401954 t]  
Plot[cdf[t], {t, 0, 350}]
```

Για τον υπολογισμό της σχέσης 4.3 εκτελούνται οι ακόλουθες εντολές:

```

m1 = MomentEvaluate[Moment[1], GammaDistribution[a1, θ1]] * w1 +
    (1 - w1) * MomentEvaluate[Moment[1], GammaDistribution[a2, θ2]]
m2 = MomentEvaluate[Moment[2], GammaDistribution[a1, θ1]] * w1 +
    (1 - w1) * MomentEvaluate[Moment[2], GammaDistribution[a2, θ2]]
m3 = MomentEvaluate[Moment[3], GammaDistribution[a1, θ1]] * w1 +
    (1 - w1) * MomentEvaluate[Moment[3], GammaDistribution[a2, θ2]]
m4 = MomentEvaluate[Moment[4], GammaDistribution[a1, θ1]] * w1 +
    (1 - w1) * MomentEvaluate[Moment[4], GammaDistribution[a2, θ2]]
m5 = MomentEvaluate[Moment[5], GammaDistribution[a1, θ1]] * w1 +
    (1 - w1) * MomentEvaluate[Moment[5], GammaDistribution[a2, θ2]]

a1 w1 θ1 + a2 (1 - w1) θ2

a1 (1 + a1) w1 θ12 + a2 (1 + a2) (1 - w1) θ22

a1 (1 + a1) (2 + a1) w1 θ13 + a2 (1 + a2) (2 + a2) (1 - w1) θ23

a1 (1 + a1) (2 + a1) (3 + a1) w1 θ14 + a2 (1 + a2) (2 + a2) (3 + a2) (1 - w1) θ24

a1 (1 + a1) (2 + a1) (3 + a1) (4 + a1) w1 θ15 + a2 (1 + a2) (2 + a2) (3 + a2) (4 + a2) (1 - w1) θ25

```

Αρχικά υπολογίζουμε τις 5 πρώτες ροπές της μίξης Γάμμα, συναρτήσει των παραμέτρων. Έπειτα, υπολογίζουμε τις 5 πρώτες ροπές των δεδομένων και τις εξισώνουμε με τις ροπές της μίξης Γάμμα.

```

Oe1 = N[MomentEvaluate[Moment[1], data1]]
Oe2 = N[MomentEvaluate[Moment[2], data1]]
Oe3 = N[MomentEvaluate[Moment[3], data1]]
Oe4 = N[MomentEvaluate[Moment[4], data1]]
Oe5 = N[MomentEvaluate[Moment[5], data1]]

227.702

54484.1

1.36017 × 107

3.52256 × 109

9.42491 × 1011

NSolve[m1 - Oe1 = 0 && m2 - Oe2 = 0 && m3 - Oe3 = 0 && m4 - Oe4 = 0 && m5 - Oe5 = 0,
{a1, a2, θ1, θ2, w1}]

{{a1 → 17.6079, a2 → 417.128, θ1 → 12.2152, θ2 → 0.65129, w1 → 0.777001},
{a1 → 417.128, a2 → 17.6079, θ1 → 0.65129, θ2 → 12.2152, w1 → 0.222999},
{a1 → 23.2675, a2 → 130.999, θ1 → 9.39333, θ2 → 1.52313, w1 → 1.48043},
{a1 → 130.999, a2 → 23.2675, θ1 → 1.52313, θ2 → 9.39333, w1 → -0.480429},
{a1 → 15.6678 + 8.66365 i, a2 → 79.8448 - 0.0491093 i,
θ1 → 8.39444 - 7.17995 i, θ2 → 3.36601 - 0.232063 i, w1 → 0.442221 - 0.373411 i},

```

Λύνοντας το σύστημα, εμφανίζονται και μιγαδικές ρίζες. Λαμβάνουμε υπόψη μας μόνο την πραγματική ρίζα.

Έχοντας εκτιμήσει τις παραμέτρους, εκτελούμε τις ίδιες εντολές που εκτελέσαμε και για τον υπολογισμό της σχέσης 4.2.

```

α1 = 17.607878878838232; α2 = 417.128160186056; β1 = 12.215176529778596;
β2 = 0.6512903872451707;
w1 = 0.7770007238924711;

Γ[α_] = Integrate[t^(α - 1) * Exp[-t], {t, 0, Infinity}, Assumptions -> Re[α] > 0];
f1[x_] = (x^(α1 - 1) * Exp[-(x/β1)]) / (β1^α1 * Γ[α1]);
f2[x_] = (x^(α2 - 1) * Exp[-(x/β2)]) / (β2^α2 * Γ[α2]);

f[x_] = w1 * f1[x] + (1 - w1) f2[x]
4.86269 × 10-34 e-0.0818654 x x16.6079 + 1.280412443026867 × 10-834 e-1.53541 x x416.128

Integrate[f[x], {x, 0, Infinity}, Assumptions -> x > 0]
1.

Integrate[x * f[x], {x, 0, Infinity}, Assumptions -> x > 0]
227.702

```

Τέλος, για την συνάρτηση κατανομής της κανονικής κατανομής έχουμε:

```

m = 227.70248; s = 51.33314;

F[x_] = CDF[NormalDistribution[m, s], x]
1/2 Erfc[0.0137749 (227.702 - x)]

F[150]
0.0650522

```

Έλεγχος Kolmogorov – Smirnov

Κατασκευή της συνάρτησης κατανομής των δεδομένων

```
data1 = Import["C:\\Users\\XXX\\Desktop\\Md.txt", "List"];
data1 = data1 / 1000;

data1;

data1 = N[data1];

data1 = Sort[data1];

list1 = {}; k = 0;

For[i = 1, i ≤ Length[data1], i = i + 1, k = k + 1 / Length[data1];
  AppendTo[list1, k]]

list1
{
 $\frac{1}{48}, \frac{1}{24}, \frac{1}{16}, \frac{1}{12}, \frac{5}{48}, \frac{1}{8}, \frac{7}{48}, \frac{1}{6}, \frac{3}{16}, \frac{5}{24}, \frac{11}{48}, \frac{1}{4}, \frac{13}{48}, \frac{7}{24}, \frac{5}{16},$ 
 $\frac{1}{3}, \frac{17}{48}, \frac{3}{8}, \frac{19}{48}, \frac{5}{12}, \frac{7}{16}, \frac{11}{24}, \frac{23}{48}, \frac{1}{2}, \frac{25}{48}, \frac{13}{24}, \frac{9}{16}, \frac{7}{12}, \frac{29}{48}, \frac{5}{8}, \frac{31}{48},$ 
 $\frac{2}{3}, \frac{11}{16}, \frac{17}{24}, \frac{35}{48}, \frac{3}{4}, \frac{37}{48}, \frac{19}{24}, \frac{13}{16}, \frac{5}{6}, \frac{41}{48}, \frac{7}{8}, \frac{43}{48}, \frac{11}{12}, \frac{15}{16}, \frac{23}{24}, \frac{47}{48}, 1$ 
}

T = Transpose[{data1, list1}];

Fob = ListLinePlot[T, PlotRange → {{80, 350}, {0, 1}}, PlotStyle → {Gray, Thick}]

Show[Fnorm, Fob, FGmle, FGmme]
```

Fnorm: η συνάρτηση κατανομής της κανονικής κατανομής

Fob: η συνάρτηση κατανομής των εμπειρικών δεδομένων

FGmle: η συνάρτηση κατανομής της μίξης Γάμμα κατανομής της οποίας οι παράμετροι εκτιμήθηκαν με την μέθοδο της μεγίστης πιθανοφάνειας.

FGmme: η συνάρτηση κατανομής της μίξης Γάμμα κατανομής της οποίας οι παράμετροι εκτιμήθηκαν με την μέθοδο των ροπών.

Με την εντολή Show προκύπτει το Γράφημα

Σύγκριση με Normal

```
In[40]= listF = {};  
      For[i = 1, i ≤ 48, i = i + 1,  
        AppendTo[listF, Abs[list1[[i]] - F[data1[[i]]]]]]  
  
In[80]= listF;  
  
In[43]= Max[listF]  
Out[43]= 0.0969894  
  
In[44]= Position[listF, Max[listF]]  
Out[44]= {{23}}  
  
In[45]= data1[[23]]  
Out[45]= 237.562
```

Σύγκριση με Gamma mme

```
listF = {};  
For[i = 1, i ≤ 48, i = i + 1,  
  AppendTo[listF, Abs[list1[[i]] - FGme[data1[[i]]]]]]  
  
listF  
{0.00873999, 0.00939982, 0.0201106, 0.0269325, 0.0293322, 0.026696, 0.00041245,  
 0.0152395, 0.0198002, 0.00393154, 0.00367106, 0.0106704, 0.00894644,  
 0.00426799, 0.0113927, 0.0260564, 0.0165582, 0.0218274, 0.0157073, 0.00523802,  
 0.00336981, 0.0298431, 0.0597862, 0.0401154, 0.0329839, 0.015541, 0.00177459,  
 0.015172, 0.0296383, 0.0314989, 0.0455774, 0.0382749, 0.051994, 0.0388477,  
 0.0447743, 0.0244981, 0.0384774, 0.0590183, 0.0507682, 0.00268108, 0.0230908,  
 0.0138892, 0.0214406, 0.00909166, 0.0113825, 0.0179208, 0.0278711, 0.00487366}  
  
Max[listF]  
0.0597862  
  
Position[listF, Max[listF]]  
{{23}}  
  
data1[[23]]  
237.562
```

Σύγκριση με Gamma mle

```
listF = {};  
For[i = 1, i ≤ 48, i = i + 1,  
  AppendTo[listF, Abs[list1[[i]] - FGme[data1[[i]]]]]]  
  
listF  
{0.0149013, 0.0175796, 0.0275747, 0.0323564, 0.0309597, 0.0226424, 0.0149873,  
 0.00145296, 0.000121275, 0.0215242, 0.0314932, 0.0174394, 0.0366708, 0.0307922,  
 0.0146238, 0.000722283, 0.00425797, 0.00402297, 0.00405026, 0.00583062,  
 0.00612795, 0.0141929, 0.0402461, 0.0206293, 0.0146278, 0.00240405, 0.0151687,  
 0.0314914, 0.0448462, 0.0430572, 0.0558422, 0.0441368, 0.0571043, 0.0426283,  
 0.0490508, 0.0326872, 0.0475909, 0.0681725, 0.0643836, 0.0274688, 0.0479373,  
 0.042092, 0.00422341, 0.014202, 0.0345573, 0.0348796, 0.0377986, 0.000229154}  
  
Max[listF]  
0.0681725  
  
Position[listF, Max[listF]]  
{{38}}  
  
data1[[38]]  
266.287
```


Γραφήματα 5^{ου} κεφαλαίου

Γράφημα 5.2

```
α1 = 1/2; α2 = 9; β = 5; w1 = 0.5; w2 = 1 - w1;

Γ[α_] = Integrate[t^(α - 1) * Exp[-t], {t, 0, Infinity}, Assumptions -> Re[α] > 0];

f1[x_] = (x^(α1 - 1) * Exp[-(x/β)]) / (β^α1 * Γ[α1]);
f2[x_] = (x^(α2 - 1) * Exp[-(x/β)]) / (β^α2 * Γ[α2]);

Integrate[f2[x], {x, 0, Infinity}, Assumptions -> x > 0];
mean = Integrate[x * f1[x], {x, 0, Infinity}, Assumptions -> x > 0];

f[x_] = w1 * f1[x] + w2 * f2[x];
Plot[f[x], {x, 0, 120}, PlotRange -> {{0, 120}, {0, 0.05}}, PlotLabel -> "f(x)"];

F[x_] = Integrate[f[t], {t, 0, x}];
Plot[F[x], {x, 0, 90}, PlotLabel -> "F(x)"];

S[x_] = Simplify[1 - F[x]];
Plot[S[x], {x, 0, 90}, PlotLabel -> "S(x)"];

h[x_] = Simplify[f[x] / S[x]];
Plot[h[x], {x, 0, 90}, PlotLabel -> "h(x)"];

mrl[x_] = Simplify[Integrate[S[t], {t, x, Infinity}] / S[x], Assumptions -> x > 0 && mrl[x] ≥ 0];
Plot[mrl[x], {x, 0, 90}, PlotLabel -> "mrl(x)"];
```

Οι παραπάνω εντολές εκτελούνται για διαφορετικές τιμές των παραμέτρων και δίνουν τα γραφήματα 5.2 με 5.4. Επίσης, προκειμένου να προκύψουν τα γραφήματα της εργασίας θα πρέπει να προσαρμόζονται τα plot ranges του κάθε άξονα αναλόγως.

Γράφημα 5.5

```
data1 = Import["C:\\Users\\xxx\\Desktop\\Wdata.txt", "Table"];
data1;
claim = data1[[2 ;; All, 2]];
Sort[claim];
mrl = {};
step = 500; i = 1; sum = 0; b = 0; j = 1;
For[k = 10, k ≤ Max[claim], k = k + step,
  For[i = 1, i ≤ Length[claim], i++,
    If[claim[[i]] >= k, sum = claim[[i]] - k + sum; b = b + 1]; If[b > 0, AppendTo[mrl, sum / b]];
    sum = 0; b = 0;]
ListPlot[mrl]
```

Ο παραπάνω αλγόριθμος υπολογίζει της μέση υπερβάλλουσα ζημιά των δεδομένων. Στην πρώτη γραμμή γίνεται η εισαγωγή των δεδομένων. Στη συνέχεια ελέγχουμε ότι τα δεδομένα έχουν περάσει σωστά στο Mathematica. Στην τρίτη γραμμή ορίζουμε claim τα ποσά των αποζημιώσεων που βρίσκονται στη δεύτερη στήλη των δεδομένων.

Το επόμενο βήμα είναι να υπολογίσουμε της μέση υπερβάλλουσα ζημιά των δεδομένων. Για να το επιτυγχάνουμε αυτό ορίζουμε αρχικά τα ακόλουθα:

*step=ότι βήμα θέλουμε (αν ορισθεί ίσο με 1 θα βγει το γράφημα της εργασίας), mrl={}, i=1,
sum=0, b=0, j=1*

και μετά εκτελούμε τον αλγόριθμο.

Για την κατασκευή των υπόλοιπων γραφημάτων, χρησιμοποιούμε την εντολή Show[graph1,graph2]. Με αυτόν τον τρόπο επιτυγχάνουμε την κατασκευή των συγκριτικών διαγραμμάτων αυτής της ενότητας.

Βιβλιογραφία

Ελληνική

Αθανασόπουλος Δ. (1991). Θεωρία Πιθανοτήτων Μέρος II Κατανομές Πιθανότητας Τυχαίων Μεταβλητών. Πειραιάς: Εκδόσεις Α. Σταμούλης.

Αντζουλάκος Δ. (2011). Ανάλυση Δεδομένων με τη Χρήση Στατιστικών Πακέτων - Εισαγωγή στην R. ΠΜΣ στην Εφαρμοσμένη Στατιστική. Πειραιάς.

Γναρδέλης Χ. (2009). ανάλυση δεδομένων με το PASW Statistics 17.0. Αθήνα: Εκδόσεις Παπαζήση.

Κούτρας Μ. (2004). Εισαγωγή στις Πιθανότητες - Θεωρία και Εφαρμογές. Αθήνα: Εκδόσεις Σταμούλη.

Χατζηκωνσταντινίδης Ε. (2015). Κατανομές Απώλειας. Σημειώσεις μαθήματος, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, Αθήνα.

Ξένη

Barlow R. and Proschan F. (1996). Mathematical Theory of Reliability. USA: John Wiley

Hall W. J. and Wellner J. A. (1981). Mean residual life. North - Holland, Amsterdam.

Jeong H. (2014). Statistics for Biology and Health. New York: Springer Science+Business Media.

Johnson K. S. (1995). Continuous Univariate Distributions. New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons.

Klugman S.A., Panjer H., Gordon E. Willmot. (2012). Loss Models From Data Decisions. John Wiley & Sons.

Kumar A. and Chaudhry A. (2008). Probability Theory. Krishna Prakashan Media (P) Ltd.

Lee C. K. S. and Lin. (2010). Modeling and Evaluating Insurance Losses Via Mixtures of Erlang Disributions. North American Actuarial Journal, 14(1), 107-130.

Patel and Read. (1982). Handbook of The Normal Distribution. New York and Basel: Marcel Dekker, Inc.

Poynor A., (2010, December). Bayesian inference for mean residual life functions in survival analysis. Msc thesis. California, USA: University of California Santa Cruz.

Shaked M. and Shanthikumar G. (2007). Stochastic Orders. New York: Springer Science+Business Media.

