ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ
& ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ &
ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

με θέμα

«Μοντέλα πιθανότητας για τα μεγέθη των ζημιών και το χρόνο αποπληρωμής τους στον κλάδο ασφάλισης αυτοκινήτου»

Σκόνδρα Αικατερίνη

ΜΑΕ/13018

Επιβλέπων Καθηγητής:

Πολίτης Κωνσταντίνος (Αναπληρωτής Καθηγητής)

Πειραιάς, 2016

“Probability models for the claim amounts and the time until the last payment per claim in the car insurance sector”

By

Aikaterini Skondra

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science in Actuarial Science and Risk Management

UNIVERSITY OF PIRAEUS

Department of Statistics and Insurance Science

Στους γονείς μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον κ. Πολίτη Κωνσταντίνο, επιβλέποντα καθηγητή μου, για τη βοήθεια και τις συμβουλές που μου παρείχε καθ' όλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας.

Θα ήθελα επίσης να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς συμβουλευτικής επιτροπής, τον κ. Μπούτσικα Μιχαήλ και τον κ. Τζαβελά Γεώργιο.

Περίληψη

Η ασφάλιση αποτελεί μια μορφή διαχείρισης κινδύνου που χρησιμοποιείται για να προστατεύσει ενάντια στον κίνδυνο πιθανών οικονομικών απωλειών. Το βασικό χαρακτηριστικό της έγκειται στο γεγονός της αποζημίωσης του ασφαλισμένου από την ασφαλιστική εταιρία σε ένα τυχαίο χρόνο και κατά ένα τυχαίο ποσό, ποσότητες οι οποίες μπορούν να αποδοθούν με τη μορφή πιθανοθεωρητικών κατανομών. Σκοπός αυτής της εργασίας είναι η εύρεση μοντέλων πιθανότητας που θα καταφέρουν να περιγράψουν τα μεγέθη των αποζημιώσεων και το χρόνο αποπληρωμής τους στον κλάδο ασφάλισης αστικής ευθύνης αυτοκινήτου. Θα δοκιμαστεί η προσαρμογή των μεταβλητών των αποζημιώσεων σε θεωρητικές κατανομές και θα κατασκευαστεί ένα νέο "πολυγραμμικό" μοντέλο για να περιγράψει τη συμπεριφορά του ύψους των αποζημιώσεων.

Abstract

Insurance is a form of risk management used to protect against the risk of potential financial losses. The main feature lies in the fact of the insured compensation from the insurance company at a random time during a random amount, amounts of which can be attributed to the form of probabilistic distributions. The aim of this thesis is to find the probability models that will be able to describe total payments and the settlement period on motor vehicle liability insurance. The adjustment of the variables will be tested on theoretical distributions and a new “multilinear” model will be constructed to describe the behavior of total payments.

Πίνακας περιεχομένων

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ 10

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ 12

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ 14

ΚΕΦΑΛΑΙΟ 2: ΔΙΕΡΕΥΝΗΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΈΛΕΓΧΟΙ ΚΑΝΟΝΙΚΟΤΗΤΑΣ 16

2.1 Περιγραφική ανάλυση δεδομένων 16

2.1.1 Περιγραφική ανάλυση δεδομένων υλικών ζημιών 18

2.1.2 Περιγραφική ανάλυση δεδομένων σωματικών βλαβών 25

2.2 Σύγκριση των περιγραφικών χαρακτηριστικών των κατηγοριών ζημιών 29

2.3 Συσχέτιση μεταβλητών των κατηγοριών ζημιών 33

2.3.1 Συσχέτιση μεταξύ των μεταβλητών υλικών ζημιών 34

2.3.2 Συσχέτιση μεταξύ των μεταβλητών σωματικών βλαβών 36

ΚΕΦΑΛΑΙΟ 3:ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ 39

3.1 Διερεύνηση της κατανομής του ύψους των αποζημιώσεων 39

3.1.1 Προσαρμογή στα αρχικά δεδομένα 39

3.1.2 Προσαρμογή σε μετασχηματισμένα δεδομένα 40

3.1.2.1 Μετασχηματισμός λογάριθμος 41

3.1.2.2 Μετασχηματισμός τετραγωνική ρίζα 43

3.1.2.3 Μετασχηματισμός κυβική ρίζα 45

3.1.2.4 Μετασχηματισμός διπλός λογάριθμος 46

3.2 Διερεύνηση της κατανομής του χρόνου αποπληρωμής 54

3.2.1 Προσαρμογή στα αρχικά δεδομένα 54

3.2.2 Προσαρμογή σε μετασχηματισμένα δεδομένα 54

3.2.2.1 Μετασχηματισμός λογάριθμος 54

3.2.2.2 Μετασχηματισμός τετραγωνική ρίζα 57

3.2.2.3 Μετασχηματισμός κυβική ρίζα 59

3.2.2.4 Μετασχηματισμός διπλός λογάριθμος 60

3.3 Διερεύνηση της κατανομής του χρόνου καθυστέρησης αναγγελίας της ζημιάς στην ασφαλιστική εταιρία 65

3.3.1 Προσαρμογή στα αρχικά δεδομένα 65

3.3.2 Προσαρμογή σε μετασχηματισμένα δεδομένα 66

3.3.2.1 Μετασχηματισμός λογάριθμος 66

3.3.2.2 Μετασχηματισμός τετραγωνική ρίζα 68

3.3.2.3 Μετασχηματισμός κυβική ρίζα 70

3.3.2.4 Μετασχηματισμός διπλός λογάριθμος 71

3.4 Box-Cox Power Transformation 74

3.4.1 Εφαρμογή του Box-Cox Power Transformation στα δεδομένα υλικών ζημιών 76

3.4.2 Εφαρμογή του Box-Cox Power Transformation στα δεδομένα σωματικών βλαβών 79

ΚΕΦΑΛΑΙΟ 4: ΔΙΕΡΕΥΝΗΣΗ ΚΑΤΑΝΟΜΗΣ ΜΕ ΜΟΝΤΕΛΑ ΜΙΚΤΟΥ ΤΥΠΟΥ 85

4.1 Μοντέλα μείξης κατανομών 85

4.1.1 Μοντέλο μείξης κατανομών – Εφαρμογή στο ύψος αποζημιώσεων 88

4.1.2 Μοντέλο μείξης κατανομών – Εφαρμογή στο χρόνο αποπληρωμής 90

4.2 Έλεγχος καλής προσαρμογής εμπειρικών κατανομών σε θεωρητικές 91

ΚΕΦΑΛΑΙΟ 5: ΠΟΛΥΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ 96

5.1 Συμφωνία άμεσου διακανονισμού ζημιών από τροχαίο ατύχημα – Σύστημα άμεσης πληρωμής (Σ.Α.Π.) 96

5.2 Εύρεση μίας νέας κατανομής για το ύψος αποζημιώσεων 99

5.2.1 Κατασκευή της “Πολυγραμμικής κατανομής” 100

5.2.2 Πολυγραμμική κατανομή του ύψους των αποζημιώσεων υλικών ζημιών 104

5.2.3 Πολυγραμμική κατανομή του ύψους των αποζημιώσεων σωματικών βλαβών 110

Παράρτημα Α 115

Παράρτημα Β 116

Παράρτημα Γ 118

ΒΙΒΛΙΟΓΡΑΦΙΑ 120

Κατάλογος Πινάκων

Πίνακας 2.1: Περιγραφικά χαρακτηριστικά μεταβλητών υλικών ζημιών. 18
Πίνακας 2.2: Αποτελέσματα των τεστ κανονικότητας μεταβλητών υλικών ζημιών. . 24
Πίνακας 2.3: Περιγραφικά χαρακτηριστικά μεταβλητών σωματικών βλαβών. 25
Πίνακας 2.4: Αποτελέσματα των τεστ κανονικότητας μεταβλητών σωματικών βλαβών. 28
Πίνακας 2.5:Περιγραφικά χαρακτηριστικά "Total payment". 29
Πίνακας 2.6:Περιγραφικά χαρακτηριστικά "Total time". 31
Πίνακας 2.7:Περιγραφικά χαρακτηριστικά "Delay". 32
Πίνακας 3.1.1: Περιγραφικά χαρακτηριστικά – μετασχηματισμός λογάριθμος. 42
Πίνακας 3.1.2: Περιγραφικά χαρακτηριστικά – μετασχηματισμός τετραγωνική ρίζα. 43
Πίνακας 3.1.3: Περιγραφικά χαρακτηριστικά – μετασχηματισμός κυβική ρίζα. 45
Πίνακας 3.1.4: Περιγραφικά χαρακτηριστικά – μετασχηματισμός διπλός λογάριθμος. 46
Πίνακας 3.1.5: Αποτελέσματα των τεστ κανονικότητας του "Total payment". 48
Πίνακας 3.1.6: Αποτελέσματα των τεστ προσαρμογής του "EasyFit" για τη GEV. ... 50
Πίνακας 3.1.7: Αποτελέσματα των τεστ προσαρμογής του "EasyFit" για τη Burr (4P). 52
Πίνακας 3.2.1: Περιγραφικά χαρακτηριστικά – μετασχηματισμός λογάριθμος. 55
Πίνακας 3.2.2: Περιγραφικά χαρακτηριστικά – μετασχηματισμός τετραγωνική ρίζα. 57
Πίνακας 3.2.3: Περιγραφικά χαρακτηριστικά – μετασχηματισμός κυβική ρίζα. 59
Πίνακας 3.2.4: Περιγραφικά χαρακτηριστικά – μετασχηματισμός διπλός λογάριθμος. 60
Πίνακας 3.2.5: Αποτελέσματα των τεστ κανονικότητας του "Total time". 62
Πίνακας 3.2.6: Αποτελέσματα του τεστ προσαρμογής του "EasyFit" για τη Wakeby. 64
Πίνακας 3.3.1: Περιγραφικά χαρακτηριστικά – μετασχηματισμός λογάριθμος. 66
Πίνακας 3.3.2: Περιγραφικά χαρακτηριστικά – μετασχηματισμός τετραγωνική ρίζα. 68
Πίνακας 3.3.3: Περιγραφικά χαρακτηριστικά – μετασχηματισμός κυβική ρίζα. 70
Πίνακας 3.3.4: Περιγραφικά χαρακτηριστικά – μετασχηματισμός διπλός λογάριθμος. 71
Πίνακας 3.3.5: Αποτελέσματα των τεστ κανονικότητας του "Delay". 73
Πίνακας 3.4.1: Μετασχηματισμοί Box – Cox Power Transformation. 75
Πίνακας 3.4.2: Αποτελέσματα για τους συντελεστές λ των μεταβλητών. 75
Πίνακας 3.4.3: Περιγραφικά χαρακτηριστικά μετασχηματισμών των μεταβλητών υλικών ζημιών. 76
Πίνακας 3.4.4: Αποτελέσματα των τεστ κανονικότητας των μεταβλητών υλικών ζημιών. 78

Πίνακας 3.4.5: Περιγραφικά χαρακτηριστικά μετασχηματισμών των μεταβλητών σωματικών βλαβών. 79
Πίνακας 3.4.6: Αποτελέσματα των τεστ κανονικότητας των μεταβλητών σωματικών βλαβών. 81
Πίνακας 3.4.7: Αποτελέσματα των τεστ προσαρμογής του “EasyFit” για τη GEV.... 83
Πίνακας 4.1: Ροπές της μεταβλητής “Total payment”. 88
Πίνακας 4.2: Εκτιμήσεις παραμέτρων μικτού μοντέλου του “Total payment”..... 89
Πίνακας 4.3: Ροπές της μεταβλητής “Total time”. 90
Πίνακας 4.4: Εκτιμήσεις παραμέτρων μικτού μοντέλου του “Total time”. 90
Πίνακας 5.1: Συντελεστές ευθειών που συνθέτουν το πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων υλικών ζημιών. 105
Πίνακας 5.2: Συντελεστές ευθειών που συνθέτουν το πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων σωματικών βλαβών. 111

Κατάλογος Σχημάτων

Σχήμα 2.1: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot “Total payment AD”.....19
Σχήμα 2.2: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot “Total time AD”. 20
Σχήμα 2.3: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot “Delay AD”. 20
Σχήμα 2.4: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot “Total payment BI”. 26
Σχήμα 2.5: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot “Total time BI”. 27
Σχήμα 2.6: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot “Delay BI”. 28
Σχήμα 2.7: Θηκογράμματα της μεταβλητής “Total payment”. 30
Σχήμα 2.8: Θηκογράμματα της μεταβλητής “Total time”.....32
Σχήμα 2.9: Θηκογράμματα της μεταβλητής “Delay”.....33
Σχήμα 2.10: Διάγραμμα διασποράς “Total payment” – “Total time”. 35
Σχήμα 2.11: Διάγραμμα διασποράς “Total payment” – “Delay”. 35
Σχήμα 2.12: Διάγραμμα διασποράς “Total time” – “Delay”.....36
Σχήμα 2.13: Διάγραμμα διασποράς “Total payment” – “Total time”. 37
Σχήμα 2.14: Διάγραμμα διασποράς “Total payment” – “Delay”. 37
Σχήμα 2.15: Διάγραμμα διασποράς “Total time” – “Delay”.....38
Σχήμα 3.1.1: Ιστογράμματα και Normal Q-Q Plots ln (Total payment). 42
Σχήμα 3.1.2: Ιστογράμματα και Normal Q-Q Plots sqrt (Total payment).....44
Σχήμα 3.1.3: Ιστογράμματα και Normal Q-Q Plots cube root (Total payment).....45
Σχήμα 3.1.4: Ιστογράμματα και Normal Q-Q Plots ln (ln (Total payment)).....46
Σχήμα 3.1.5: Ιστόγραμμα λογαρίθμου “Total payment” σωματικών βλαβών. 49
Σχήμα 3.1.6: Ιστόγραμμα κυβικής ρίζας “Total payment” σωματικών βλαβών. 51
Σχήμα 3.2.1: Ιστογράμματα και Normal Q-Q Plots ln (Total time). 56
Σχήμα 3.2.2: Ιστογράμματα και Normal Q-Q Plots sqrt (Total time). 58
Σχήμα 3.2.3: Ιστογράμματα και Normal Q-Q Plots cube root (Total time). 59
Σχήμα 3.2.4: Ιστογράμματα και Normal Q-Q Plots ln (ln (Total time)). 61
Σχήμα 3.2.5: Ιστόγραμμα κυβικής ρίζας “Total time” σωματικών βλαβών.....63
Σχήμα 3.3.1: Ιστογράμματα και Normal Q-Q Plots ln (Delay). 67
Σχήμα 3.3.2: Ιστογράμματα και Normal Q-Q Plots sqrt (Delay). 69
Σχήμα 3.3.3: Ιστογράμματα και Normal Q-Q Plots cube root (Delay). 70
Σχήμα 3.3.4: Ιστογράμματα και Normal Q-Q Plots ln (ln (Delay)). 72
Σχήμα 3.4.1: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του “Total payment”. 77
Σχήμα 3.4.2: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του “Total time”.....77
Σχήμα 3.4.3: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του “Delay”.....78
Σχήμα 3.4.4: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του “Total payment”. 80
Σχήμα 3.4.5: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του “Total time”.....80

Σχήμα 3.4.6: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του “Delay”.....81
Σχήμα 3.4.7: Ιστόγραμμα μετασχηματισμού Box – Cox “Total payment” σωματικών βλαβών. 82
Σχήμα 5.1: Πολυγραμμικό μοντέλο. 101
Σχήμα 5.2: Πολυγωνική γραμμή και πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων υλικών ζημιών. 108
Σχήμα 5.3:Ιστόγραμμα συχνοτήτων της σ.π.π. και της πολυγραμμικής κατανομής του ύψους των αποζημιώσεων υλικών ζημιών. 109
Σχήμα 5.4: Πολυγωνική γραμμή και πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων σωματικών βλαβών. 112
Σχήμα 5.5:Ιστόγραμμα συχνοτήτων της σ.π.π. και της πολυγραμμικής κατανομής του ύψους των αποζημιώσεων σωματικών βλαβών. 113

Κεφάλαιο 1: Εισαγωγή

Ο κλάδος της αστικής ευθύνης χερσαίων οχημάτων είναι ο μεγαλύτερος κλάδος των Γενικών Ασφαλίσεων με παραγωγή 806 εκατ. € και 44% επί του συνόλου (ΕΑΕΕ, Παραγωγή ασφαλίστρων δωδεκαμήνου 2015). Η ασφάλιση της αστικής ευθύνης αυτοκινήτου αποτελεί μορφή υποχρεωτικής ασφάλισης. Σύμφωνα με το Νόμο, η ασφάλιση αστικής ευθύνης από ατύχημα αυτοκινήτου καλύπτει την έναντι τρίτων αστική ευθύνη του κυρίου, του κατόχου και κάθε οδηγού, σε περίπτωση πρόκλησης τροχαίου ατυχήματος με υπαιτιότητά τους.

Ο κλάδος αστικής ευθύνης αυτοκινήτου ανήκει στην κατηγορία των Γενικών ασφαλίσεων, επίσης γνωστές ως Ασφαλίσεις κατά Ζημιών. Ο όρος χρησιμοποιείται καθώς εν αντιθέσει προς τις Ασφαλίσεις Ζωής που σχεδόν πάντα το κεφάλαιο του ασφαλιζομένου επιστρέφεται και είναι γνωστό κατά τη χρονική στιγμή σύναψης της ασφάλισης, εδώ ο ασφαλιζόμενος θα εισπράξει κάποιο ποσό, ανάλογο προς τη σοβαρότητα της ζημιάς, από την ασφαλιστική εταιρία μόνο αν επέλθει ζημιά στο χρονικό διάστημα της κάλυψης. Λόγω αυτής της ιδιαιτερότητας, στους κλάδους των Γενικών ασφαλίσεων μεγέθη όπως το αναμενόμενο πλήθος ζημιών και μέσο αναμενόμενο κόστος αποζημιώσεων από ασφαλιστικούς κινδύνους είναι απαραίτητα για την διεκπεραίωση σημαντικών εργασιών. Η σπουδαιότερη από αυτές είναι η τιμολόγηση ασφαλίστρου. Ο καθορισμός, δηλαδή, του επιπέδου του ασφαλίστρου το οποίο αναμένεται να είναι επαρκές για την κάλυψη του κόστους αποζημίωσης και των εξόδων κάθε ομοιογενούς κατηγορίας κινδύνων, ώστε να μειωθούν στο ελάχιστο οι οικονομικές απώλειες της ασφαλιστικής εταιρίας από την επέλευση ασφαλιστικών γεγονότων.

Και στον τομέα της αντασφάλισης, η γνώση της κατανομής που ακολουθούν οι αποζημιώσεις κρίνεται πολύ σημαντική. Ειδικότερα, για την κάλυψη του κλάδου αστικής ευθύνης αυτοκινήτου η συνήθης αντασφαλιστική σύμβαση είναι η αντασφάλιση υπερβάλλουσας ζημίας (excess of loss). Σε αυτή την κατηγορία αντασφάλισης, ο αντασφαλιστής καλείται να καλύψει τις ζημιές, οι αποζημιώσεις των οποίων υπερβαίνουν ένα προσυμφωνηθέν με την εκάστοτε εταιρία όριο και για ποσά αποζημιώσεων πάνω από αυτό. Επομένως, η ασφαλιστική εταιρία θα μπορεί να επιλέξει το όριο στο οποίο θα ενεργοποιείται η αντασφαλιστική κάλυψη βάσει του κινδύνου που επιθυμεί να αναλάβει. Η πληροφόρηση για την κατανομή που ακολουθούν οι αποζημιώσεις είναι πολύ σημαντική και για την αντασφαλιστική εταιρία. Θα είναι σε θέση να εκτιμήσει σε τι ποσοστό οι αποζημιώσεις θα υπερβαίνουν το όριο ενεργοποίησης της κάλυψης μέσω των ποσοστημορίων της κατανομής, καθώς επίσης και πόσο βαριά είναι η δεξιά ουρά της κατανομής των αποζημιώσεων.

Πέραν του ύψους των ποσών πληρωμών μία ασφαλιστική εταιρία θα ήθελε να γνωρίζει και άλλα δύο μεγέθη που συνδέονται με τις αποζημιώσεις της. Το πρώτο είναι ο χρόνος που απαιτείται για τον οριστικό διακανονισμό των αποζημιώσεων.

Έτσι κατά το άνοιγμα ενός φακέλου ζημιάς θα είναι σε θέση να γνωρίζει κατά μέσο όρο τον χρόνο που χρειάζεται για την αποπληρωμή των υποχρεώσεων της. Επίσης, θα μπορεί να εκτιμήσει πόσες ζημιές θα διακανονιστούν πλήρως σε διάστημα ενός έτους ή ενός εξαμήνου. Τέλος, θα έχει μία εικόνα των ακραίων περιπτώσεων χρόνων αποπληρωμής, που συνήθως συνδέονται με δικαστικές διαμάχες. Το δεύτερο είναι ο χρόνος καθυστέρησης αναγγελίας της ζημιάς από τον εμπλεκόμενο σε ατύχημα ασφαλισμένο. Γνωρίζοντας η εταιρία την κατανομή του χρόνου καθυστέρησης αναγγελίας της ζημιάς θα μπορεί να εκτιμήσει πόσες IBNR (Incurred But Not Reported, δηλ. προκληθείσες μα μη δηλωθείσες) ζημιές να αναμένει για παράδειγμα τον επόμενο μήνα.

Κατανοώντας τη σημαντικότητα των παραπάνω, στην παρούσα διπλωματική εργασία θα ασχοληθούμε με την εύρεση μοντέλων πιθανότητας που περιγράφουν τα μεγέθη αποζημιώσεων που καταβάλει στους εμπλεκόμενους ατυχημάτων η ασφαλιστική εταιρία καθώς και το χρόνο αποπληρωμής τους για τον κλάδο αστικής ευθύνης αυτοκινήτου. Για τον λόγο αυτό θα αναλύσουμε πραγματικά δεδομένα αποζημιώσεων κλειστών φακέλων ζημιάς του κλάδου αστικής ευθύνης αυτοκινήτου ασφαλιστικής εταιρίας. Πιο αναλυτικά η δομή της εργασίας θα είναι η εξής:

Στο δεύτερο κεφάλαιο θα παρουσιαστούν αναλυτικά τα δεδομένα αποζημιώσεων του κλάδου αστικής ευθύνης αυτοκινήτου της ασφαλιστικής εταιρίας. Για κάθε μία από τις τρεις μεταβλητές που περιγράφουν τις αποζημιώσεις θα δοθούν περιγραφικά χαρακτηριστικά και θα πραγματοποιηθούν έλεγχοι κανονικότητας για να διαπιστωθεί αν τα δεδομένα κατανέμονται κανονικά. Επίσης, θα προσδιοριστεί ο βαθμός συσχέτισης μεταξύ των μεταβλητών.

Στο τρίτο κεφάλαιο θα πραγματοποιηθεί διερεύνηση της κατανομής που ακολουθεί η κάθε μεταβλητή που περιγράφει τις αποζημιώσεις. Πέραν της προσπάθειας προσαρμογής κατανομών στα αρχικά δεδομένα, θα δοκιμαστεί η προσαρμογή των μετασχηματισμών αυτών σε διάφορες θεωρητικές κατανομές.

Στο τέταρτο κεφάλαιο θα προσεγγίσουμε τις συναρτήσεις πυκνότητας πιθανότητας του ύψους των αποζημιώσεων και του χρόνου αποπληρωμής τους από μία μείξη δύο γάμμα κατανομών. Για να το επιτύχουμε αυτό θα χρησιμοποιήσουμε τη μέθοδο των ροπών. Στη συνέχεια θα αξιολογήσουμε την προσαρμογή στα μοντέλα μείξης γάμμα με το στατιστικό τεστ καλής προσαρμογής χ^2, ώστε να διαπιστώσουμε αν η προσαρμογή τους είναι ικανοποιητική.

Στο τελευταίο κεφάλαιο θα δοθεί ερμηνεία για τη συμπεριφορά των αποζημιώσεων της εταιρίας μέσω της πληροφόρησης για τον τρόπο λειτουργίας του Συστήματος Άμεσης Πληρωμής (Σ.Α.Π.). Θα κατασκευαστεί ένα νέο “πολυγραμμικό” μοντέλο για την περιγραφή και τον προσδιορισμό του ύψους των αποζημιώσεων. Το μοντέλο αυτό θα προσεγγίζει την καμπύλη της κατανομής που ακολουθούν τα δεδομένα αποζημιώσεων με διαδοχικές ευθείες γραμμές.

Κεφάλαιο 2: Διερευνητική ανάλυση και Έλεγχοι κανονικότητας

Στο παρόν κεφάλαιο γίνεται παρουσίαση των στοιχείων των αποζημιώσεων του κλάδου ασφάλισης της αστικής ευθύνης αυτοκινήτου που θα χρησιμοποιήσουμε για τη μελέτη μας. Στη διάθεσή μας έχουμε στοιχεία αποζημιώσεων που προκύπτουν από την εμπλοκή οχημάτων του στόλου ασφαλισμένων οχημάτων της εταιρίας σε τροχαία ατυχήματα περιόδου οκταετίας. Θα πραγματοποιηθεί παρουσίαση των βασικών μεταβλητών των αποζημιώσεων, δηλαδή του ύψους των πληρωμών, του χρόνου έως τον οριστικό διακανονισμό των ζημιών και του χρόνου καθυστέρησης αναγγελίας αυτών στην ασφαλιστική εταιρία και των ιδιαίτερων χαρακτηριστικών τους. Τέλος, θα προσδιοριστεί ο βαθμός στον οποίο οι μεταβλητές αποζημιώσεων συσχετίζονται μεταξύ τους.

2.1 Περιγραφική ανάλυση δεδομένων

Τα στοιχεία της ασφαλιστικής εταιρίας αφορούν σε καταβολές αποζημιώσεων του κλάδου Αστικής Ευθύνης Αυτοκινήτου. Πιο συγκεκριμένα τα στοιχεία προέρχονται από χαρτοφυλάκιο ζημιών ασφαλιστικής επιχείρησης και αφορούν σε καταβολές αποζημιώσεων για ζημιές με ημερομηνία συμβάντος από 1/1/2003 μέχρι και 31/12/2010. Για τις αποζημιώσεις της ασφαλιστική επιχείρηση δίνεται το σύνολο των κλειστών φακέλων ζημιάς και οι μεμονωμένες πληρωμές που διεκπεραιώθηκαν, ανεξάρτητα αν αφορούν καταβολή αποζημίωσης, έξοδα αποζημίωσης ή επανείσπραξη (ανάκτηση). Επιπλέον για κάθε φάκελο ζημιάς δίνονται η ημερομηνία συμβάντος της ζημιάς (accident date), η ημερομηνία αναγγελίας (δήλωσης) της ζημιάς (announcement date), η ημερομηνία τελευταίας πληρωμής (last payment date) και τα ποσά πληρωμής ανά έτος σε ευρώ.

Τα παραπάνω στοιχεία χρησιμοποιούνται για την πραγματοποίηση στατιστικής ανάλυσης που αφορά στο χαρτοφυλάκιο της ασφαλιστικής εταιρίας. Η μελέτη μας θα εστιαστεί σε τρεις μεταβλητές που αφορούν στις αποζημιώσεις που πληρώθηκαν για τον οριστικό διακανονισμό των ζημιών της οκταετίας 2003-2010.

Η πρώτη μεταβλητή που θα εξετάσουμε είναι το συνολικό ύψος των αποζημιώσεων που καταβλήθηκαν από την εταιρία (“Total payment”). Η δεύτερη μεταβλητή, ο συνολικός χρόνος που απαιτείται για τον οριστικό διακανονισμό των ζημιών και επομένως το “κλείσιμο” των φακέλων ζημίας. Τέλος, θα εξετάσουμε το χρόνο καθυστέρησης αναγγελίας από πλευράς ασφαλισμένου της προκληθείσας ζημιάς στην ασφαλιστική εταιρία.

Για τη διερεύνηση των παραπάνω μεταβλητών, υπολογίστηκαν από τα στοιχεία για κάθε ζημιά τα εξής:

- Total payment (συνολικό ύψος αποζημίωσης) = άθροισμα του συνόλου των μεμονωμένων πληρωμών για κάθε φάκελο ζημιάς

- Total time (συνολικός χρόνος αποπληρωμής – συνολικός χρόνος έως το "κλείσιμο" του φακέλου ζημιάς) = last payment date – announcement date

- Delay (καθυστέρηση αναγγελίας της ζημιάς στην ασφαλιστική εταιρία) = announcement date – accident date

Όσον αφορά στα δεδομένα ζημιών πραγματοποιήθηκε διαχωρισμός αυτών σε δύο κατηγορίες. Η πρώτη κατηγορία είναι οι υλικές ζημιές (Accidental Damage) και η δεύτερη οι σωματικές βλάβες (Third Party Bodily Injury). Ο διαχωρισμός αυτός κρίνεται απαραίτητος για την ανάλυση των δεδομένων αποζημιώσεων, καθώς σε κάθε περίπτωση συναντούμε μεγάλες αποκλίσεις στο ύψος των αποζημιώσεων και στη διάρκεια αποπληρωμής τους. Σημειώνεται ότι για τις περιπτώσεις φακέλων ζημιών με αποζημιώσεις για σωματικές βλάβες και υλικές ζημιές γίνεται επιλογή της ένταξής τους στην κατηγορία των σωματικών βλαβών. Ο λόγος που η επιλογή αυτή κρίνεται απαραίτητη είναι αρχικά πως ο φάκελος ζημιάς "κλείνει" μία φορά και πως δεν μπορούμε να θεωρήσουμε πως ο χρόνος αποπληρωμής είναι ανεξάρτητος του είδους αποζημίωσης. Ζημιές που αφορούν σε σωματικές βλάβες, κατά κανόνα έχουν μεγαλύτερο χρόνο αποπληρωμής από αυτές που αφορούν σε υλικές ζημίες (δικαστικές διαμάχες, χρόνιες σωματικές βλάβες, ιατροφαρμακευτική περίθαλψη κτλ).

Μετά τον τελικό διαχωρισμό των ζημιών του χαρτοφυλακίου της ασφαλιστικής επιχείρησης σε υλικές ζημιές και σωματικές βλάβες, έχουμε:

- 102.454 κλειστούς φακέλους για υλικές ζημιές
- 3.629 κλειστούς φακέλους για σωματικές βλάβες

Σε προκαταρκτική εξέταση των δεδομένων αλλά και για καλύτερη κατανόηση, δίνονται παρακάτω περιγραφικά χαρακτηριστικά που αφορούν στο συνολικό ύψος ζημιάς (total payment), το συνολικό χρόνο αποπληρωμής ζημιάς (total time) και την καθυστέρηση αναγγελίας ζημιάς (delay) για την ασφαλιστική εταιρία. Επιπλέον με τη χρήση διαγραμμάτων αλλά και ελέγχων κανονικότητας των δεδομένων εξετάζεται αν τα δεδομένα προέρχονται από την κανονική κατανομή. Για τη διενέργεια των υπολογισμών και την εξαγωγή των γραφημάτων γίνεται χρήση του στατιστικού πακέτου R.

2.1.1 Περιγραφική ανάλυση δεδομένων υλικών ζημιών

Τα αποτελέσματα στον παρακάτω πίνακα αφορούν την πρώτη κατηγορία φακέλων ζημιών, δηλαδή αποζημιώσεις για υλικές ζημίες.

Πίνακας 2.1: Περιγραφικά χαρακτηριστικά μεταβλητών υλικών ζημιών.

Accidental Damage	Total payment	Total time	Delay
min	93	0	0
25% quantile	412	32	5
median (50% quantile)	952	61	8
mean	1589	125.1	16.79
75% quantile	1277	85	14
max	594000	4195	2642
variance	25327776	62432.2	2638.1
standard deviation	5032.7	249.9	51.4
99% quantile	12924.5	1358.0	155.0
skewness	47.6	5.2	19.1
kurtosis	3791.8	39.4	551.5

Ξεκινώντας από τα αποτελέσματα του πίνακα για το ύψος αποζημιώσεων υλικών ζημιών, η μέση ισούται με 1589, δηλαδή κατά μέσο όρο μία αποζημίωση για υλικές ζημιές θα "κοστίσει" στην ασφαλιστική εταιρία περίπου €1.590. Η μέγιστη αποζημίωση που κλήθηκε να καταβάλει είναι ίση με €594.000 και η μικρότερη μόλις €93. Παρατηρούμε μεγάλη μεταβλητότητα στο ύψος των καταβληθέντων αποζημιώσεων, με τυπική απόκλιση ίση με 5032,7. Η κατανομή που ακολουθούν οι συνολικές πληρωμές υλικών ζημιών είναι λεπτόκυρτη με συντελεστή κύρτωσης ίσο με 3791,8. Η υψηλή τιμή του συντελεστή κύρτωσης μας δίνει την πληροφορία ότι μεγάλος όγκος των τιμών των αποζημιώσεων θα είναι συγκεντρωμένος σε κάποιες τιμές.

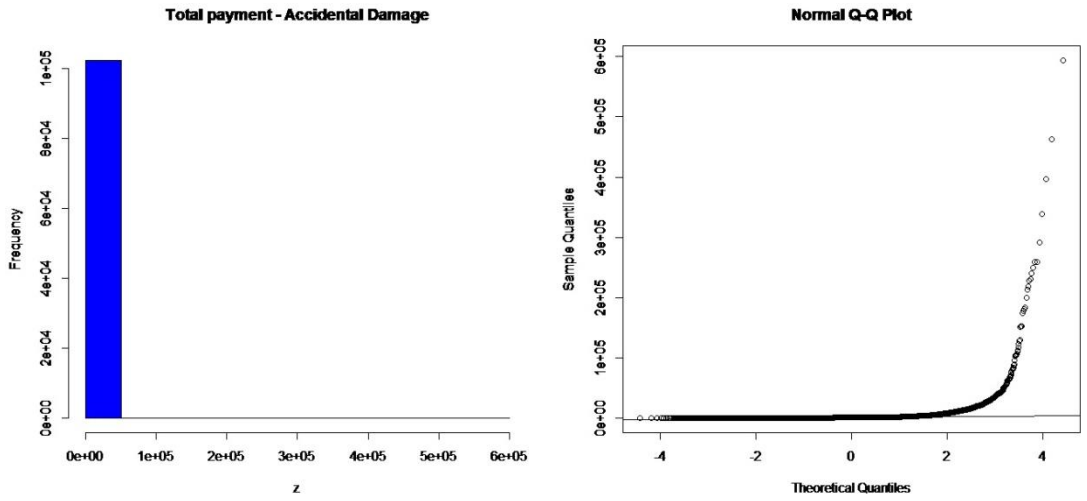
Αντίστοιχα, για το χρόνο αποπληρωμής υλικών ζημιών, η μέση τιμή ισούται με 125,1 και επομένως κατά μέσο όρο η υλική ζημιά διακανονίζεται σε περίπου 125 ημέρες. Ο μέγιστος χρόνος που απαιτήθηκε για αποπληρωμή από την εταιρία είναι 4195 ημέρες και ο ελάχιστος 0 ημέρες, δηλαδή την ημέρα αναγγελίας της ζημίας πραγματοποιήθηκε και η πληρωμή. Η τυπική απόκλιση είναι ίση με 249,9. Στην περίπτωση του "Total time" αναμένουμε μία κατανομή θετικά ασύμμετρη και κυρτή σύμφωνα με τις τιμές των συντελεστών όπως φαίνονται στον παραπάνω πίνακα.

Τέλος, κατά μέσο όρο ο χρόνος καθυστέρησης για να δηλωθεί η υλική ζημιά στην εταιρία είναι 16,8 ημέρες. Η μικρότερη τιμή για τη μεταβλητή "Delay" είναι 0, αναμενόμενο αν σκεφτούμε πως κανονικά οι ζημιές δηλώνονται στην εταιρία όταν συμβούν κυρίως στις περιπτώσεις που απαιτείται πραγματογνώμονας. Η μέγιστη τιμή ισούται με 2642 ημέρες, δηλαδή στα στοιχεία αποζημιώσεων υπάρχει ζημιά που

καθυστέρησε να δηλωθεί στην ασφαλιστική επτά περίπου χρόνια! Η τυπική απόκλιση των χρόνων είναι 51,4. Η τιμή του συντελεστή κύρτωσης, μαρτυρά λεπτόκυρτη κατανομή με συγκέντρωση σε κάποια τιμή.

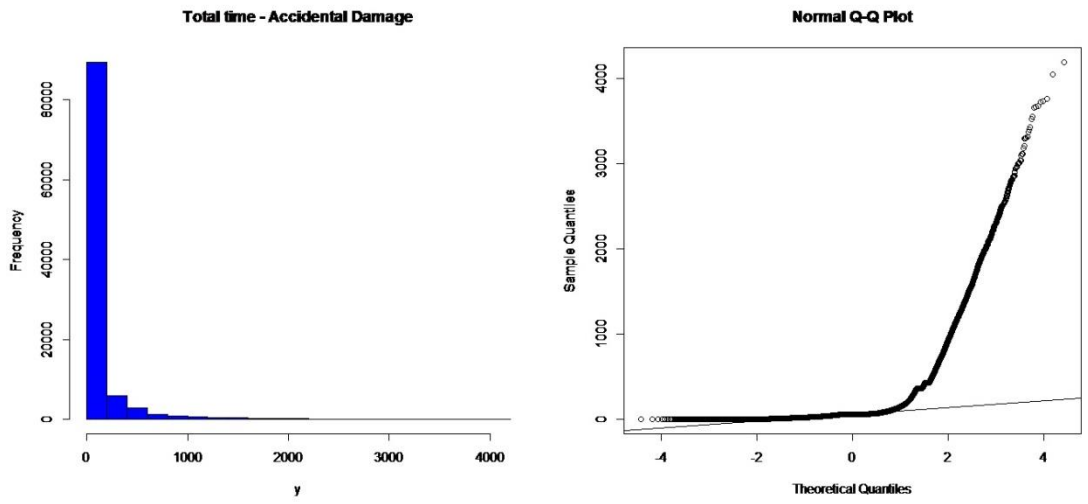
Με σκοπό να έχουμε καλύτερη οπτική της συμπεριφοράς των τιμών των μεταβλητών υλικών ζημιών, θα επιχειρήσουμε να εξάγουμε συμπεράσματα μέσω γραφικών μεθόδων.

Γραφική απεικόνιση για τη μορφή των δεδομένων δίνεται παρακάτω με ιστογράμματα συχνοτήτων. Πρόκειται για τρόπο περιγραφής της κατανομής των δεδομένων, όπου το ύψος κάθε ράβδου ισούται με την παρατηρούμενη συχνότητα στην κατηγορία αυτή. Κατασκευάζονται, επίσης, Q-Q Plots για την κανονική κατανομή για να διαπιστώσουμε αν υπάρχουν γραφικές ενδείξεις για την καταλληλότητα της επιλογής της κανονικής κατανομής. Σε ένα Q-Q Plot διακρίνουμε την ευθεία γραμμή της κανονικής κατανομής. Όσο πιο πολύ προσεγγίζουν τα δεδομένα την ευθεία αυτή, τόσο πιο ισχυρή γραφική ένδειξη υπάρχει πως τα δεδομένα του δείγματος που εξετάζεται προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή. Αντίθετα αποκλίσεις από τη γραμμή, μας δίνουν ένδειξη απόρριψης της υπόθεσης ότι τα δεδομένα είναι κανονικά κατανεμημένα.



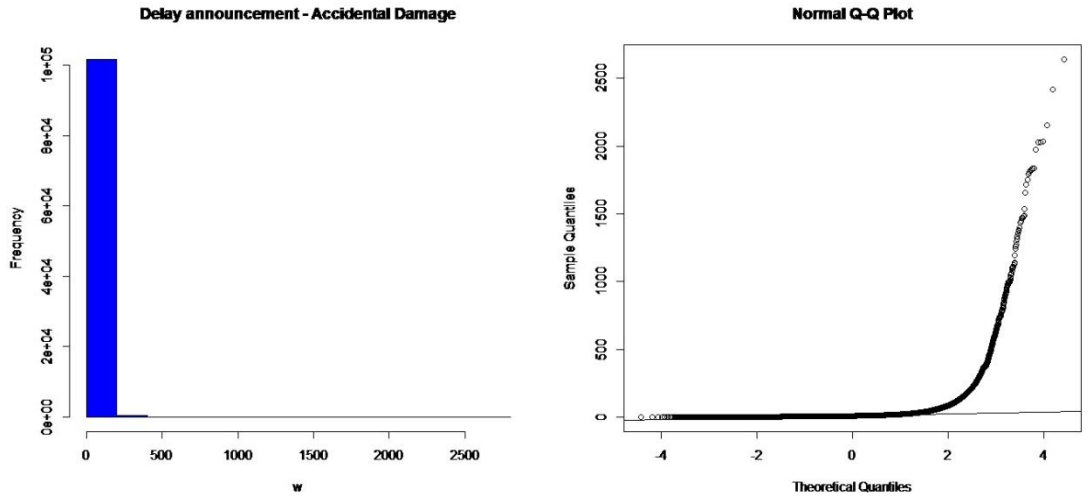
Σχήμα 2.1: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot "Total payment AD".

Στο παραπάνω ιστόγραμμα συχνοτήτων παρουσιάζεται η συχνότητα του ύψους των συνολικών πληρωμών ("Total payment") για φακέλους που αφορούν μόνο σε υλικές ζημιές. Από την μορφή του ιστογράμματος δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα για την κατανομή που ακολουθούν τα δεδομένα. Το Q-Q Plot δεν επιδεικνύει καλή προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά μία κατανομή με βαριά δεξιά ουρά.



Σχήμα 2.2: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot "Total time AD".

Στο παραπάνω ιστόγραμμα συχνοτήτων παρουσιάζεται η συχνότητα του συνολικού χρόνου κλεισίματος των φακέλων υλικών ζημιών ("Total time"). Από την μορφή του ιστογράμματος συμπεραίνουμε ότι εμφανίζεται μεγάλος βαθμός συγκέντρωσης στην πρώτη τάξη, δηλαδή για χρόνο κλεισίματος φακέλου από 0 έως 250 ημέρες. Παρατηρούμε επίσης θετική ασυμμετρία, γεγονός που επιβεβαιώνεται και από τον παραπάνω πίνακα. Όσον αφορά στην κατανομή που ακολουθούν τα δεδομένα, δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα. Το Q-Q Plot δεν επιδεικνύει προσαρμογή των δεδομένων στη κανονική κατανομή, αλλά κάποια κατανομή με πολύ βαριά δεξιά ουρά.



Σχήμα 2.3: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot "Delay AD".

Τέλος, στο παραπάνω ιστόγραμμα συχνοτήτων παρουσιάζεται η συχνότητα του χρόνου καθυστέρησης αναγγελίας της ζημιάς από τον ασφαλισμένο στην ασφαλιστική εταιρία ("Delay"). Και εδώ, από τη μορφή του ιστογράμματος συμπεραίνουμε ότι εμφανίζεται μεγάλος βαθμός συγκέντρωσης στην πρώτη τάξη, δηλαδή για καθυστέρηση αναγγελίας ζημιάς από 0 έως 250 ημέρες. Από την μορφή του ιστογράμματος δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα για την κατανομή που ακολουθούν τα δεδομένα. Το Q-Q Plot δεν επιδεικνύει προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά κάποια κατανομή με πολύ βαριά δεξιά ουρά.

Για να εξάγουμε τελικό συμπέρασμα για το αν τα δεδομένα αποζημιώσεων των τριών μεταβλητών υλικών ζημιών προέρχονται από την κανονική κατανομή, προχωρούμε στην επιλογή στατιστικών τεστ που δεν στηρίζονται σε γραφικές μεθόδους, όπως τα Lilliefors (Kolmogorov-Smirnov) normality test, Anderson-Darling normality test, Cramer-von Mises normality test και Pearson chi-square normality test.

Τα ιστογράμματα και οι συντελεστές ασυμμετρίας και κύρτωσης χρησιμοποιούνται για μια αρχική επισκόπηση της κατανομής του δείγματος. Από την ανάγνωση όμως των αποτελεσμάτων των συντελεστών, δεν προκύπτει καθαρό αποτέλεσμα. Για να είμαστε περισσότερο ακριβείς, θα χρησιμοποιήσουμε κριτήρια ελέγχου κανονικότητας για να κατανοήσουμε αν τα δεδομένα μας ακολουθούν κανονική κατανομή. Η εφαρμογή των κριτηρίων μας δίνει σαν αποτέλεσμα την απόρριψη ή μη της υπόθεσης ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

Ο έλεγχος κανονικότητας για τις μεταβλητές "Total payment", "Total time" και "Delay" πραγματοποιήθηκε με τη βοήθεια του στατιστικού πακέτου R, την εγκατάσταση και χρήση του πακέτου nortest, που δίνει τη δυνατότητα πραγματοποίησης των τεσσάρων παρακάτω τεστ κανονικότητας.

- ***Anderson-darling Goodness-of-fit test***

Το Anderson-Darling normality test χρησιμοποιείται για να ελέγξουμε αν ένα δείγμα προέρχεται από πληθυσμό που ακολουθεί την κανονική κατανομή. Αποτελεί τροποποίηση του Kolmogorov-Smirnov τεστ και δίνει μεγαλύτερη βαρύτητα στις ουρές της κατανομής του δείγματος απ΄ ότι το Kolmogorov-Smirnov τεστ.

Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: *Τα δεδομένα ακολουθούν την κανονική κατανομή.*

H_1: *Τα δεδομένα δεν ακολουθούν την κανονική κατανομή.*

Η στατιστική ελέγχου του τεστ είναι:

$$A = -n - \frac{1}{n}\sum_{i=1}^{n}[2i - n]\left[\ln(p_{(i)}) + \ln\left(1 - p_{(n-i+1)}\right)\right],$$

όπου $p_{(i)} = \Phi([x_{(i)} - \bar{x}]/s)$ και $x_{(i)}$ οι διατεταγμένες παρατηρήσεις του δείγματος. Η Φ είναι η αθροιστική συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής και $\bar{x}$ και s είναι η μέση τιμή και η τυπική απόκλιση των δεδομένων του δείγματος. Το p-value υπολογίζεται από την τροποποιημένη στατιστική (modified statistic) $Z = A(1.0 + 0.75/n + 2.25/n^2)$ σύμφωνα με τον πίνακα 4.9 του Stephens (1986).

Η υπόθεση H_0 απορρίπτεται για μεγάλες τιμές της A.

- ***Cramer-von Mises Goodness-of-fit test***

Οι υποθέσεις που γίνονται για τη διεξαγωγή του ελέγχου κανονικότητας είναι οι εξής:

H_0: Τα δεδομένα ακολουθούν την κανονική κατανομή.

H_1: Τα δεδομένα δεν ακολουθούν την κανονική κατανομή.

Η στατιστική ελέγχου του τεστ είναι:

$$W = \frac{1}{12n} + \sum_{i=1}^{n}\left(p_{(i)} - \frac{2i-1}{2n}\right)^2$$

όπου $p_{(i)} = \Phi([x_{(i)} - \bar{x}]/s)$. Η Φ είναι η αθροιστική συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής και $\bar{x}$ και s είναι η μέση τιμή και η τυπική απόκλιση των δεδομένων του δείγματος. Το p-value υπολογίζεται από την τροποποιημένη στατιστική (modified statistic) $Z = W(1.0 + 0.5/n)$ σύμφωνα με τον πίνακα 4.9 του Stephens (1986).

Για μεγάλες τιμές της W απορρίπτουμε τη μηδενική υπόθεση H_0.

- ***Lilliefors Goodness-of-fit test***

Ένα μειονέκτημα του ελέγχου Komogorov-Smirnov, 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 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. Σε αυτή την περίπτωση είναι προτιμότερο να χρησιμοποιήσουμε μια παραλλαγή του ελέγχου Komogorov-Smirnov, τον έλεγχο Lilliefors, που χρησιμοποιεί προσομοίωση Monte Carlo και αποδίδει μια προσαρμοσμένη p-value. Ο συγκεκριμένος έλεγχος εκτελείται με τη συνάρτηση lillie.test του πακέτου nortest.

Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: Τα δεδομένα ακολουθούν την κανονική κατανομή.

H_1: Τα δεδομένα δεν ακολουθούν την κανονική κατανομή.

Η στατιστική ελέγχου του τεστ είναι η μέγιστη απόλυτη διαφορά ανάμεσα στην εμπειρική και την αθροιστική συνάρτηση της κανονικής κατανομής. Υπολογίζεται ως $D = max\{D^+, D^-\}$, όπου

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

όπου $p_{(i)} = \Phi([x_{(i)} - \bar{x}]/s)$. Η Φ είναι η αθροιστική συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής και $\bar{x}$ και s είναι η μέση τιμή και η τυπική απόκλιση των δεδομένων του δείγματος. Το p-value υπολογίζεται από την Dallal-Wilkinson (1986) formula, η οποία είναι η μόνη αξιόπιστη όταν το p-value είναι μικρότερο από 0,1. Αν το Dallal-Wilkinson p-value είναι μεγαλύτερο από 0,1 τότε το p-value υπολογίζεται από την κατανομή της τροποποιημένης στατιστικής (modified statistic) $Z = D(\sqrt{n} - 0.01 + 0.85/\sqrt{n})$, Stephens (1974), η πραγματική p-value formula λαμβάνεται από μια διαδικασία προσομοίωσης και προσέγγισης.

Προφανώς και στην περίπτωση αυτού του ελέγχου, είναι οι μεγάλες τιμές της στατιστικής συνάρτησης D, οι οποίες συνηγορούν υπέρ της απόρριψης της μηδενικής υπόθεσης, αφού τέτοιες τιμές αντανακλούν χαμηλό βαθμό εγγύτητας της εμπειρικής συνάρτησης κατανομής του δείγματος προς την συνάρτηση κατανομής της κανονικής κατανομής.

- ***Pearson chi-squared Goodness-of-fit test***

Οι υποθέσεις που γίνονται για τον έλεγχο της κανονικότητας των δεδομένων του δείγματος είναι οι εξής:

H_0: Τα δεδομένα ακολουθούν την κανονική κατανομή.

H_1: Τα δεδομένα δεν ακολουθούν την κανονική κατανομή.

Η στατιστική ελέγχου του τεστ είναι:

$$P = \sum \frac{(C_i - E_i)^2}{E_i}$$

όπου C_i είναι το πλήθος των παρατηρήσεων στην i κατηγορία και E_i είναι το αναμενόμενο πλήθος παρατηρήσεων στην i κατηγορία κάτω από τη μηδενική υπόθεση. Το p-value του ελέγχου υπολογίζεται από την κατανομή χ^2 με i-1 βαθμούς ελευθερίας.

Μεγάλες τιμές της στατιστικής συνάρτησης P συνηγορούν υπέρ της απόρριψης της μηδενικής υπόθεσης ότι τα δεδομένα προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή.

Έχοντας πραγματοποιήσει μία σύντομη περιγραφή των τεσσάρων τεστ κανονικότητας, προχωρούμε στην παρουσίαση των αποτελεσμάτων που εξήχθησαν με το στατιστικό πακέτο R.

Οι τιμές των στατιστικών ελέγχου και οι αντίστοιχες τιμές των p-value από την εφαρμογή των τεσσάρων τεστ κανονικότητας στα δεδομένα υλικών ζημιών εμφανίζονται στον παρακάτω πίνακα.

Πίνακας 2.2: Αποτελέσματα των τεστ κανονικότητας μεταβλητών υλικών ζημιών.

Accidental Damage	**Total payment**		**Total time**		**Delay**	
	D/A/W/P	p-value	D/A/W/P	p-value	D/A/W/P	p-value
Lilliefors	0.3832	<2,2 e-16	0.3244	<2,2 e-16	0.3768	<2,2 e-16
Anderson-Darling	23667.71	<2,2 e-16	19565.61	<2,2 e-16	23930.94	<2,2 e-16
Cramer-von Mises	4844.033	7.37E-10	3989.01	7.37E-10	4893.92	7.37E-10
Pearson	714761.8	<2,2 e-16	1276707	<2,2 e-16	914601.9	<2,2 e-16

Τα αποτελέσματα που προκύπτουν από τους ελέγχους κανονικότητας επιβεβαιώνουν την αρχική μας πρόβλεψη ότι τα δεδομένα δεν προέρχονται από την κανονική κατανομή, αφού τα p-value των ελέγχων εμφανίζουν την ελάχιστη τιμή που δίνεται από το στατιστικό πακέτο R. Επομένως, απορρίπτουμε τη μηδενική υπόθεση, δηλαδή την υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή σε επίπεδο σημαντικότητας 5%.

2.1.2 Περιγραφική ανάλυση δεδομένων σωματικών βλαβών

Κατά τον ίδιο τρόπο που εργαστήκαμε για τις αποζημιώσεις υλικών ζημιών, στην υποενότητα αυτή, προχωρούμε στην ανάλυση της δεύτερης κατηγορίας αποζημιώσεων του χαρτοφυλακίου της εταιρίας. Τα αποτελέσματα στον παρακάτω πίνακα αφορούν στα περιγραφικά χαρακτηριστικά αποζημιώσεων για σωματικές βλάβες.

Πίνακας 2.3: Περιγραφικά χαρακτηριστικά μεταβλητών σωματικών βλαβών.

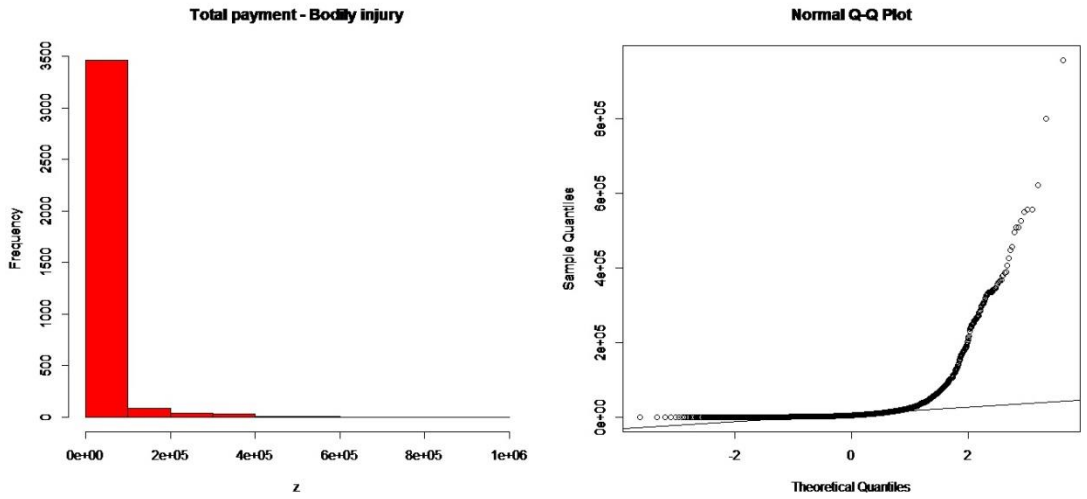
Bodily Injury	Total payment	Total time	Delay
min	235	0	0
25% quantile	2000	105	5
median (50% quantile)	5500	515	9
mean	21650	768.4	32.9
75% quantile	15000	1273	17
max	958000	3944	2241
variance	3363830675	591076.1	14895.9
standard deviation	57998.54	768.81	122.05
99% quantile	326778.3	3087.36	603.16
skewness	6.26	1.09	9.49
kurtosis	56.93	3.61	119.24

Για το ύψος αποζημιώσεων σωματικών βλαβών, η μέση ισούται με 21650, δηλαδή κατά μέσο όρο η εταιρία καλείται να καταβάλει €21.650 ως αποζημίωση σε τρίτο πρόσωπο που ενεπλάκει σε συμβάν (τροχαίο ατύχημα) με τον ασφαλισμένο της για σωματικές βλάβες. Η μέγιστη αποζημίωση που δόθηκε για ατύχημα της οκταετίας που εξετάζουμε, είναι ίση με €958.000 και η μικρότερη €235. Υπάρχει πολύ μεγάλη απόκλιση στις τιμές του ύψους των καταβληθέντων αποζημιώσεων, με τυπική απόκλιση ίση με 57998,54. Η κατανομή που ακολουθούν οι συνολικές πληρωμές σωματικών βλαβών θα εμφανίζει βαριά δεξιά ουρά, αφού το 75% ποσοστημόριο ισούται με €15.000 και η μέγιστη τιμή με €958.000.

Για το χρόνο αποπληρωμής σωματικών βλαβών, η μέση τιμή ισούται με 768,4 και επομένως κατά μέσο όρο για να διακανονιστεί μια ζημιά αυτής της κατηγορίας απαιτούνται περίπου 2,1 χρόνια. Ο μέγιστος χρόνος που απαιτήθηκε για αποπληρωμή από την εταιρία είναι 3944 ημέρες και ο ελάχιστος 0 ημέρες, δηλαδή την ημέρα αναγγελίας της ζημίας πραγματοποιήθηκε και η πληρωμή. Η τυπική απόκλιση είναι ίση με 768,81. Στην περίπτωση του “Total time” αναμένουμε μία κατανομή θετικά ασύμμετρη, καθώς ο συντελεστής ασυμμετρίας ισούται με 1,09. Τέλος, η κατανομή είναι κυρτή με συντελεστή κύρτωσης 3,61 πολύ κοντά στην τιμή 3 μιας μεσόκυρτης κατανομής.

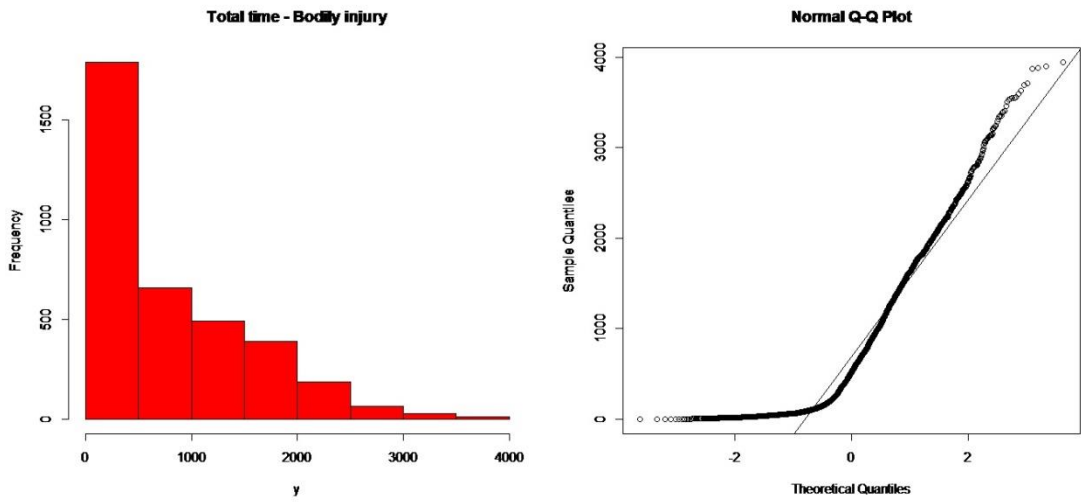
Αντίστοιχα, κατά μέσο όρο ο χρόνος καθυστέρησης για να δηλωθεί ατύχημα με σωματική βλάβη τρίτου προσώπου στην εταιρία είναι 33 περίπου ημέρες. Η μικρότερη τιμή για τη μεταβλητή "Delay" είναι και σε αυτή την κατηγορία ζημιών ίση με 0. Η μέγιστη τιμή ισούται με 2241 ημέρες, δηλαδή στα στοιχεία αποζημιώσεων υπάρχει ζημιά που καθυστέρησε να δηλωθεί στην ασφαλιστική εταιρία έξι περίπου χρόνια! Η τυπική απόκλιση των χρόνων καθυστέρησης δήλωσης του συμβάντος είναι 122,05. Η τιμή του συντελεστή κύρτωσης, μαρτυρά λεπτόκυρτη κατανομή με συγκέντρωση σε κάποια τιμή.

Γραφική απεικόνιση για τη μορφή των δεδομένων δίνεται παρακάτω με ιστογράμματα συχνοτήτων, ώστε να σχηματίσουμε μία αρχική εικόνα για την κατανομή από την οποία προέρχονται τα δεδομένα μας. Κατασκευάζουμε, επίσης, Q-Q Plots για την κανονική κατανομή για να διαπιστώσουμε αν υπάρχουν γραφικές ενδείξεις για την καταλληλότητα της επιλογής της κανονικής κατανομής ως κατανομή που περιγράφει τα δεδομένα μας.



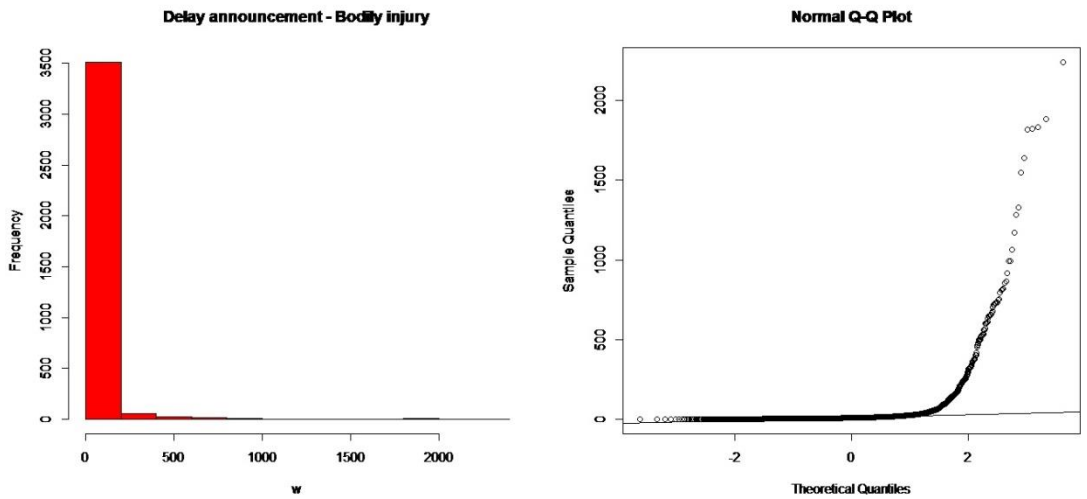
Σχήμα 2.4: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot "Total payment BI".

Στο παραπάνω ιστόγραμμα συχνοτήτων παρουσιάζεται η συχνότητα του ύψους των συνολικών πληρωμών - αποζημιώσεων ("Total payment") για φακέλους της κατηγορίας σωματικές βλάβες. Από την μορφή του ιστογράμματος δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα για την κατανομή που ακολουθούν τα δεδομένα. Το Q-Q Plot δεν επιδεικνύει καλή προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά μαρτυρά ότι τα δεδομένα ακολουθούν μία κατανομή με πολύ βαριά δεξιά ουρά, όπως αναφέρουμε και παραπάνω.



Σχήμα 2.5: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot "Total time BI".

Στο παραπάνω ιστόγραμμα συχνοτήτων παρουσιάζεται η συχνότητα του συνολικού χρόνου "κλεισίματος" των φακέλων ζημιών ("Total time") για σωματικές βλάβες. Από την μορφή του ιστογράμματος συμπεραίνουμε ότι για την πλειονότητα των ζημιών, οι χρόνοι αποπληρωμής τους ανήκουν στην πρώτη κλάση, δηλαδή 0 έως 500 ημέρες. Σε αντίθεση όμως με τα δεδομένα των υλικών ζημιών, εδώ παρατηρούμε συγκέντρωση χρόνων αποπληρωμής και στις επόμενες κλάσεις. Πιο συγκεκριμένα, περίπου 650 ζημιές αποπληρώθηκαν σε διάστημα 500 έως 1000 ημερών, 500 ζημιές σε διάστημα 1000 έως 1500 ημερών κτλ. Παρατηρούμε επίσης θετική ασυμμετρία, γεγονός που επιβεβαιώνεται και από τον παραπάνω πίνακα. Όσον αφορά στην κατανομή που ακολουθούν τα δεδομένα, δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα. Το Q-Q Plot δεν επιδεικνύει προσαρμογή των δεδομένων στη κανονική κατανομή. Επίσης μπορούμε να συμπεράνουμε πως δεν υπάρχει προσαρμογή στην κανονική κατανομή κυρίως για μικρές τιμές και για τις ακραίες τιμές. Όπως φαίνεται και στο Q-Q Plot οι ουρές της κατανομής των δεδομένων δεν φαίνεται να προσαρμόζονται στην κανονική κατανομή.



Σχήμα 2.6: Ιστόγραμμα συχνοτήτων και Normal Q-Q Plot "Delay BI".

Τέλος, στο παραπάνω ιστόγραμμα συχνοτήτων παρουσιάζεται η συχνότητα του χρόνου καθυστέρησης αναγγελίας της ζημιάς ("Delay") στην ασφαλιστική εταιρία. Και εδώ, από τη μορφή του ιστογράμματος συμπεραίνουμε ότι εμφανίζεται μεγάλος βαθμός συγκέντρωσης στην πρώτη τάξη, δηλαδή για καθυστέρηση αναγγελίας ζημιάς από 0 έως 250 ημέρες. Από την μορφή του ιστογράμματος δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα για την κατανομή που ακολουθούν τα δεδομένα. Το Q-Q Plot δεν επιδεικνύει προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά κάποια κατανομή με πολύ βαριά δεξιά ουρά.

Για να εξάγουμε τελικό συμπέρασμα για το αν τα δεδομένα αποζημιώσεων για σωματικές βλάβες προέρχονται από την κανονική κατανομή, προχωρούμε στη διενέργεια των στατιστικών τεστ του πακέτου nortest της R (Lilliefors (Kolmogorov-Smirnov) normality test, Anderson-Darling normality test, Cramer-von Mises normality test και Pearson chi-square normality test).

Οι τιμές των στατιστικών ελέγχου και οι αντίστοιχες τιμές των p-value από την εφαρμογή των τεσσάρων τεστ κανονικότητας στα δεδομένα εμφανίζονται στον παρακάτω πίνακα.

Πίνακας 2.4: Αποτελέσματα των τεστ κανονικότητας μεταβλητών σωματικών βλαβών.

Bodily Injury	**Total payment**		**Total time**		**Delay**	
	D/A/W/P	p-value	D/A/W/P	p-value	D/A/W/P	p-value
Lilliefors	0.356	<2,2 e-16	0.1588	<2,2 e-16	0.3948	<2,2 e-16
Anderson-Darling	780.936	<2,2 e-16	158.0419	<2,2 e-16	986.6415	<2,2 e-16
Cramer-von Mises	159.0949	7.37E-10	26.2316	7.37E-10	207.2938	7.37E-10
Pearson	26194.14	<2,2 e-16	6327.43	<2,2 e-16	38489.37	<2,2 e-16

Τα αποτελέσματα που προκύπτουν από τους ελέγχους κανονικότητας επιβεβαιώνουν την αρχική μας πρόβλεψη ότι τα δεδομένα δεν προέρχονται από την κανονική κατανομή, αφού τα p-value των ελέγχων εμφανίζουν την ελάχιστη τιμή που δίνεται από το στατιστικό πακέτο R. Επομένως, απορρίπτουμε τη μηδενική υπόθεση, δηλαδή την υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή σε επίπεδο σημαντικότητας 5%.

2.2 Σύγκριση των περιγραφικών χαρακτηριστικών των κατηγοριών ζημιών

Όπως αναφέραμε και παραπάνω, τα στοιχεία φακέλων ζημιών χωρίζονται σε δύο κατηγορίες, δηλαδή σε αποζημιώσεις υλικών ζημιών και σωματικών βλαβών. Σε αυτό το σημείο θεωρούμε ότι μία σύγκριση μεταξύ των περιγραφικών χαρακτηριστικών των τιμών των μεταβλητών "Total payment", "Total time" και "Delay" για τις δύο κατηγορίες που χωρίσαμε τις ζημιές του χαρτοφυλακίου της εταιρίας, θα μας δώσει σημαντικές πληροφορίες για την ανάλυσή μας.

Ο παρακάτω πίνακας αναφέρεται στο ύψος αποζημιώσεων των δύο κατηγοριών, δηλαδή των υλικών ζημιών και σωματικών βλαβών.

Πίνακας 2.5:Περιγραφικά χαρακτηριστικά "Total payment".

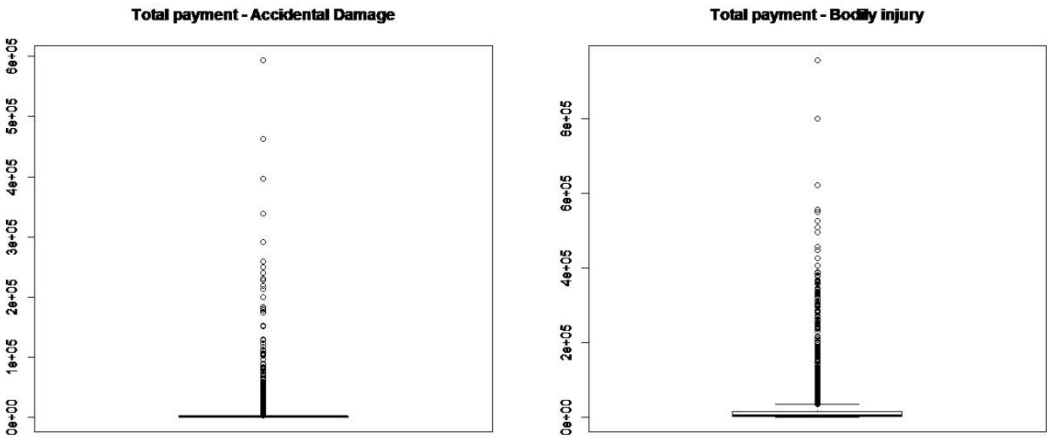
Total payment	Accidental Damage	Bodily Injury
min	93	235
mean	1589	21650
max	594000	958000
median (50% quantile)	952	5500
standard deviation	5032.7	57998.5
Q3-Q1	865	13000
skewness	47.6	6.3
kurtosis	3791.8	56.9

Όπως είναι αναμενόμενο στην περίπτωση των σωματικών βλαβών οι τιμές που παρατηρούμε για το ύψος των αποζημιώσεων είναι πολύ μεγαλύτερες. Η μέση τιμή των αποζημιώσεων για υλικές ζημιές ισούται με €1.589 ενώ για τις σωματικές βλάβες με €21.650. Επίσης, στην περίπτωση των σωματικών βλαβών παρατηρούμε πολύ μεγαλύτερη διακύμανση των τιμών του ύψους αποζημιώσεων σε σχέση με τις αποζημιώσεις για υλικές ζημιές. Οι πληρωμές της ασφαλιστικής εταιρίας για υλικές ζημίες κυμαίνονται στο διάστημα €93 έως €594.000. Οι πληρωμές για σωματικές βλάβες παίρνουν τιμές στο διάστημα €235 έως και €958.000. Η διαπίστωση αυτή μπορεί να εξηγηθεί εύκολα από το γεγονός ότι στις περιπτώσεις αποζημιώσεων ατυχημάτων που περιλαμβάνουν σωματικές βλάβες οι αποζημιώσεις που δίνονται

από τις εταιρίες είναι κατά κανόνα υψηλότερες από αυτές των υλικών ζημιών. Ειδικότερα στις περιπτώσεις δικαστικών διαμαχών μεταξύ των εμπλεκόμενων μερών, τα ποσά αποζημιώσεων παίρνουν πολύ μεγάλες τιμές που ίσως εξηγεί και την τεράστια μεταβλητότητα στα δεδομένα μας.

Δίνονται επίσης, οι τιμές του ενδοτεταρτημοριακού εύρους, δηλαδή της διαφοράς ανάμεσα στην τιμή του τρίτου και του πρώτου τεταρτημορίου των διατεταγμένων παρατηρήσεων της μεταβλητής. Όπως είναι αναμενόμενο από τα παραπάνω, το μεγαλύτερο εύρος παρατηρείται στις αποζημιώσεις σωματικών βλαβών, που εμφανίζουν και μεγαλύτερη μεταβλητότητα. Παρατηρούμε, επίσης, ότι και οι δύο μεταβλητές παρουσιάζουν θετική ασυμμετρία. Τέλος, από τις τιμές που παίρνουν οι συντελεστές κύρτωσης αντιλαμβανόμαστε πως στην περίπτωση των αποζημιώσεων υλικών ζημιών υπάρχει πολύ μεγαλύτερη συγκέντρωση σε κάποιες τιμές απ' ότι στις σωματικές βλάβες.

Δίνονται παρακάτω τα boxplot για τη μεταβλητή του ύψους των αποζημιώσεων υλικών ζημιών και σωματικών βλαβών.



Σχήμα 2.7: Θηκογράμματα της μεταβλητής "Total payment".

Η μορφή των boxplot για τις αποζημιώσεις και των δύο κατηγοριών ζημιών δεν μας δίνει τη δυνατότητα να εξάγουμε συμπεράσματα πέραν της ύπαρξης πολλών ακραίων τιμών στα δεδομένα.

Ο παρακάτω πίνακας αναφέρεται στο συνολικό χρόνο που απαιτείται για την αποπληρωμή των αποζημιώσεων υλικών ζημιών και σωματικών βλαβών.

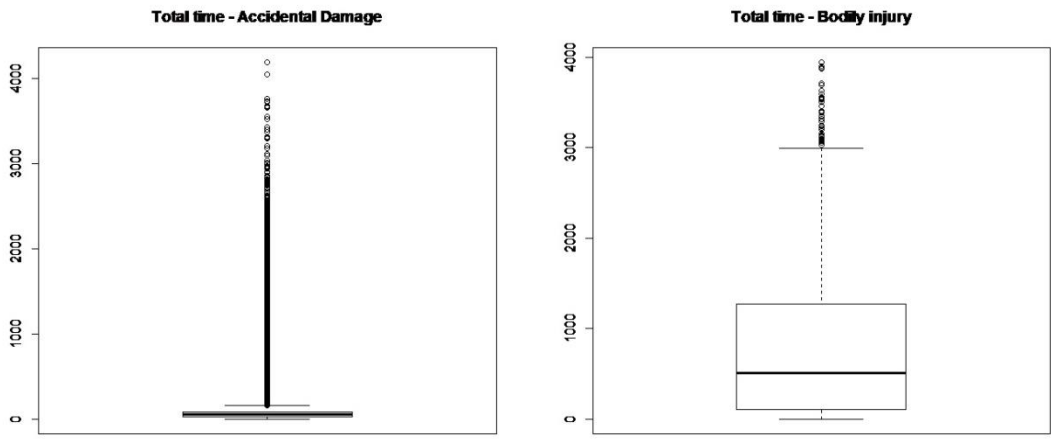
Πίνακας 2.6:Περιγραφικά χαρακτηριστικά "Total time".

Total time	Accidental Damage	Bodily Injury
min	0	0
mean	125.1	768.4
max	4195	3944
median (50% quantile)	61	515
standard deviation	249.9	768.8
Q3-Q1	53	1168
skewness	5.2	1.1
kurtosis	39.4	3.6

Όπως και για το ύψος αποζημιώσεων, έτσι και για τους χρόνους αποπληρωμής αναμένουμε οι τιμές για σωματικές βλάβες να είναι μεγαλύτερες. Αναφέραμε παραπάνω την περίπτωση δικαστικών διαμαχών, οι οποίες στην πλειονότητα τους αποδεικνύονται χρονοβόρες.

Από τη σύγκριση των τιμών του παραπάνω πίνακα επιβεβαιώνονται οι σκέψεις μας, ότι οι χρόνοι για την κατηγορία σωματικές βλάβες είναι σαφώς μεγαλύτεροι από αυτούς των υλικών ζημιών. Πιο συγκεκριμένα, η μέση τιμή του "Total time" για υλικές ζημίες είναι 125 ημέρες και αντίστοιχα για σωματικές βλάβες 768 ημέρες. Και εδώ η διασπορά των παρατηρούμενων τιμών είναι πολύ μεγαλύτερη για τις σωματικές βλάβες (768,8 έναντι 249,9). Όπως είναι αναμενόμενο από τα παραπάνω, το μεγαλύτερο ενδοτεταρτημοριακό εύρος παρατηρείται στις αποζημιώσεις σωματικών βλαβών, που εμφανίζουν και μεγαλύτερη μεταβλητότητα, ίσο με 1168 ημέρες. Αντίστοιχα, η τιμή για χρόνο αποπληρωμής υλικών ζημιών είναι 53 ημέρες. Εντύπωση μας προκαλεί η μέγιστη τιμή για το χρόνο αποπληρωμής υλικής ζημίας ίση με 4165 ημέρες, ενώ αντίστοιχα η μέγιστη για σωματικές βλάβες είναι 3944 ημέρες. Προφανώς πρόκειται για κάποια ακραία περίπτωση, καθώς κατά μέσο όρο όπως αναφέραμε ένας φάκελος ζημιάς που αφορά σε σωματική βλάβη τρίτου, καθυστερεί περισσότερο να "κλείσει" σε σχέση με έναν για υλική ζημιά. Τέλος, παρατηρούμε ότι και οι δύο μεταβλητές παρουσιάζουν θετική ασυμμετρία με τους χρόνους αποπληρωμής αποζημιώσεων σωματικών βλαβών να κατανέμονται πιο συμμετρικά από τους χρόνους για σωματικές βλάβες.

Τα παραπάνω συμπεράσματα μπορούν να γίνουν εύκολα αντιληπτά με τη βοήθεια του παρακάτω γραφήματος που περιλαμβάνει τα boxplot του χρόνου αποπληρωμής των απαιτήσεων των ασφαλισμένων από την ασφαλιστική εταιρία και για τις δύο κατηγορίες.



Σχήμα 2.8: Θηκογράμματα της μεταβλητής "Total time".

Η μορφή του boxplot για υλικές ζημιές δεν μας δίνει τη δυνατότητα να οδηγηθούμε σε συμπεράσματα πέραν της ύπαρξης πολλών ακραίων τιμών. Αντίθετα, η μορφή του boxplot των σωματικών βλαβών μας επιτρέπει να επιβεβαιώσουμε τα αποτελέσματα του παραπάνω πίνακα περιγραφικών χαρακτηριστικών. Όπως αναφέραμε οι τιμές χρόνου αποπληρωμής για σωματικές βλάβες κατανέμονται πιο συμμετρικά και παρόλο που η διακύμανση των τιμών είναι πολύ μεγάλη εδώ δεν παρατηρούμε τόσες ακραίες τιμές.

Τέλος, ο παρακάτω πίνακας αναφέρεται στην καθυστέρηση δήλωσης της ζημιάς στην εταιρία ("Delay") για τις δύο κατηγορίες αποζημιώσεων.

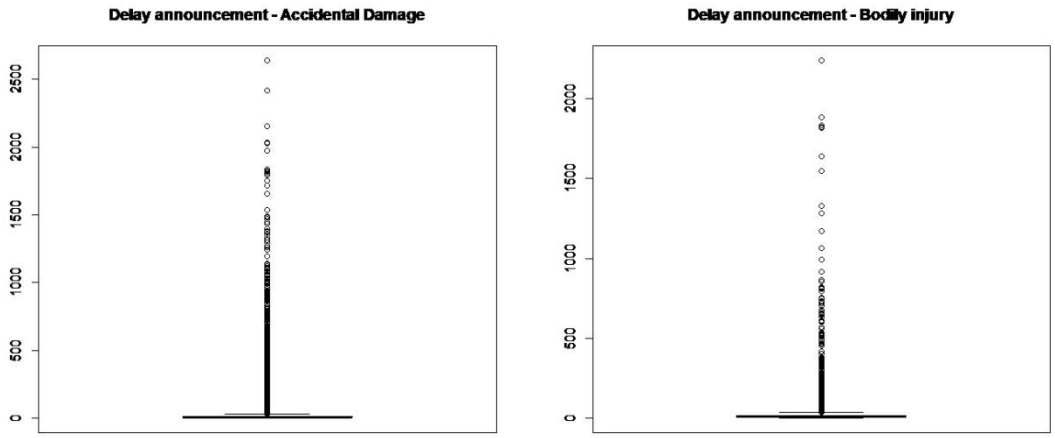
Πίνακας 2.7:Περιγραφικά χαρακτηριστικά "Delay".

Delay	Accidental Damage	Bodily Injury
min	0	0
mean	16.79	32.86
max	2642	2241
median (50% quantile)	8	9
standard deviation	51.4	122
Q3-Q1	9	12
skewness	19.1	9.5
kurtosis	551.5	119.2

Σε αντίθεση με τις προηγούμενες δύο μεταβλητές, σε αυτή την περίπτωση δεν παρατηρούνται τόσο μεγάλες διαφορές μεταξύ των τιμών. Η μέση τιμή της καθυστέρησης αναγγελίας της υλικής ζημιάς είναι 17 ημέρες και 33 ημέρες για σωματικές βλάβες. Η τυπική απόκλιση των τιμών της μεταβλητής είναι και εδώ μεγαλύτερη στην περίπτωση των αποζημιώσεων σωματικών βλαβών. Όπως είναι αναμενόμενο, το μεγαλύτερο ενδοτεταρτημοριακό εύρος παρατηρείται στις

αποζημιώσεις σωματικών βλαβών, που εμφανίζουν τη μεγαλύτερη μεταβλητότητα, και είναι ίσο με 12 ημέρες. Τέλος, και σε αυτή την περίπτωση παρατηρούμε ότι οι δύο μεταβλητές παρουσιάζουν έντονη θετική ασυμμετρία.

Δίνονται τα boxplot για τη μεταβλητή καθυστέρηση αναγγελίας συμβάντος που δημιουργεί απαίτηση για την καταβολή αποζημίωσης από την ασφαλιστική εταιρία για τις δύο κατηγορίες ζημιών, δηλαδή για υλικές ζημιές και σωματικές βλάβες.



Σχήμα 2.9: Θηκογράμματα της μεταβλητής "Delay".

Δυστυχώς στην περίπτωση αυτή, η μορφή των θηκογραμμάτων δεν μας δίνει τη δυνατότητα να εξάγουμε συμπεράσματα πέραν της ύπαρξης ακραίων τιμών για τους χρόνους καθυστέρησης αναγγελίας ζημιάς και στις δύο κατηγορίες αποζημιώσεων.

2.3 Συσχέτιση μεταβλητών των κατηγοριών ζημιών

Σε συνέχεια της ανάλυσης που πραγματοποιήσαμε στις προηγούμενες υποενότητες για τις τρεις μεταβλητές σε κάθε κατηγορία ζημιάς, αναζητούμε στην παρούσα υποενότητα το βαθμό στον οποίο οι τιμές της μίας μεταβλητής συσχετίζονται με τις τιμές που λαμβάνει μία από τις άλλες δύο μεταβλητές.

Συσχέτιση μεταξύ δύο τυχαίων μεταβλητών υπάρχει όταν η μία επηρεάζει την άλλη, δηλαδή όταν η τιμή που λαμβάνει η μία μεταβλητή επηρεάζει την τιμή που λαμβάνει η άλλη μεταβλητή.

Ένας απλός τρόπος για να αποκτήσουμε μία πρώτη ιδέα για το αν και πως δύο μεταβλητές συσχετίζονται είναι να κατασκευάσουμε το διάγραμμα διασποράς (Scatter Plot). Με το διάγραμμα διασποράς αναπαριστούμε τα ζεύγη των

παρατηρήσεων σε ένα διάγραμμα, ελέγχοντας γραφικά την ύπαρξη μιας ανάλογης ή αντιστρόφως ανάλογης σχέσης μεταξύ των δύο μεταβλητών. Στην περίπτωση που επιθυμούμε να ποσοτικοποιήσουμε τη σχέση αυτή, δηλαδή να μετρήσουμε το βαθμό στον οποίο οι τιμές της μιας μεταβλητής συσχετίζονται με της άλλης, χρησιμοποιούμε το συντελεστή γραμμικής συσχέτισης του Pearson.

Στη συνέχεια μελετούμε την ύπαρξη και το βαθμό συσχέτισης των μεταβλητών "Total payment", "Total time" και "Delay" ανά δύο για τις αποζημιώσεις υλικών ζημιών και σωματικών βλαβών.

2.3.1 Συσχέτιση μεταξύ των μεταβλητών υλικών ζημιών

Ξεκινάμε την ανάλυση μας με τη μελέτη της ύπαρξης συσχέτισης μεταξύ των μεταβλητών που αναφέρονται στις αποζημιώσεις υλικών ζημιών. Πιο συγκεκριμένα δίνουμε τα διαγράμματα διασποράς και τους συντελεστές συσχέτισης του Pearson για τα ζεύγη "Total payment" – "Total time" , "Total payment" – "Delay", "Total time" – "Delay" υλικών ζημιών.

Ο δειγματικός συντελεστής γραμμικής συσχέτισης του Pearson συμβολίζεται με r και δίνει ένα μέτρο του μεγέθους της γραμμικής συσχέτισης μεταξύ δύο τ.μ. X και Y με δειγματική διασπορά s_X^2 και s_Y^2 αντίστοιχα και συνδιασπορά $s_{XY} = COV(X,Y)$. Ορίζεται από τον τύπο:

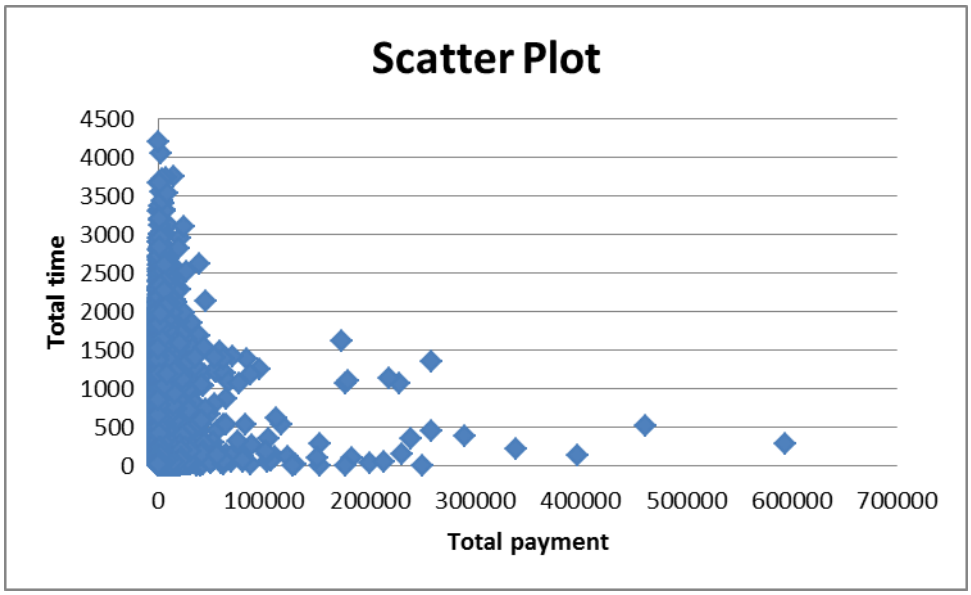
$$r = \frac{s_{XY}}{s_X s_Y}$$

Ο δειγματικός συντελεστής συσχέτισης r, όπως και η συνδιασπορά s_{XY} , εκφράζει το βαθμό και τον τρόπο που οι δύο μεταβλητές συσχετίζονται, δηλαδή πως η μία τ.μ. μεταβάλλεται ως προς την άλλη. Η s_{XY} παίρνει τιμές που εξαρτώνται από το πεδίο τιμών των X και Y, ενώ ο συντελεστής r παίρνει τιμές στο διάστημα [-1, 1]. Οι χαρακτηριστικές τιμές του r ερμηνεύονται ως εξής:

- Αν r = ±1 υπάρχει τέλεια γραμμική συσχέτιση.
- Αν $-0{,}3 \leq r < 0{,}3$ δεν υπάρχει γραμμική συσχέτιση. Αυτό, όμως, δεν σημαίνει ότι δεν υπάρχει άλλου είδους συσχέτιση μεταξύ των δύο μεταβλητών.
- Αν $-0{,}5 < r \leq -0{,}3$ ή $0{,}3 \leq r < 0{,}5$ υπάρχει ασθενής γραμμική συσχέτιση.
- Αν $-0{,}7 < r \leq -0{,}5$ ή $0{,}5 \leq r < 0{,}7$ υπάρχει μέση γραμμική συσχέτιση.
- Αν $-0{,}8 < r \leq -0{,}7$ ή $0{,}7 \leq r < 0{,}8$ υπάρχει ισχυρή γραμμική συσχέτιση.
- Αν $-1 < r \leq -0{,}8$ ή $0{,}8 \leq r < 1$ υπάρχει πολύ ισχυρή γραμμική συσχέτιση.

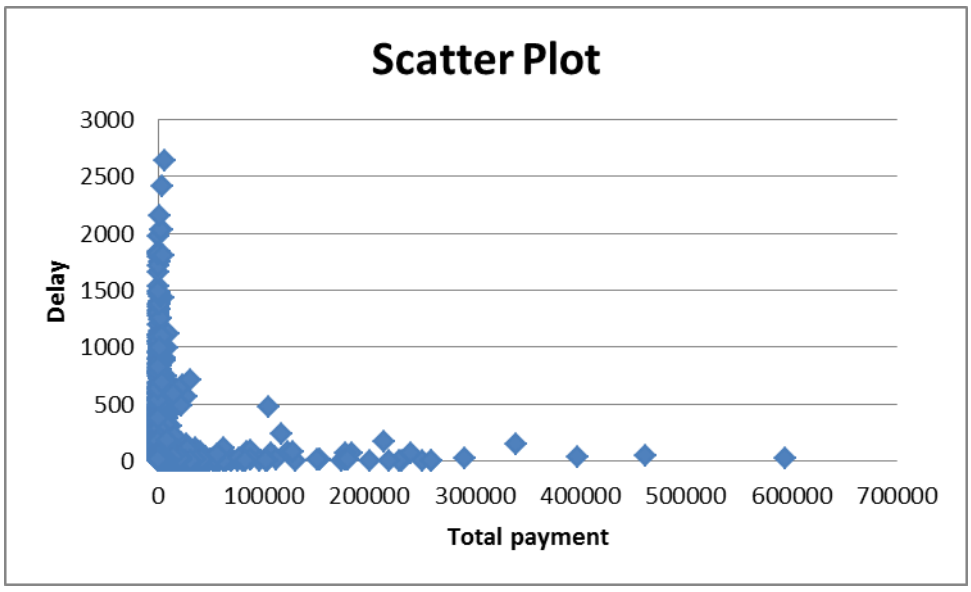
Ο βαθμός γραμμικής συσχέτισης καθορίζεται από την απόλυτη τιμή του r και όχι από το πρόσημο του. Το πρόσημο του r καθορίζει το είδος μόνο της συσχέτισης (θετική ή αρνητική). Μας πληροφορεί δηλαδή για το αν αύξηση της μιας μεταβλητής αντιστοιχεί σε αύξηση ή σε μείωση της άλλης μεταβλητής.

Οι συντελεστές συσχέτισης υπολογίστηκαν με την εντολή CORREL που είναι διαθέσιμη στο Excel. Ακολουθούν τα διαγράμματα διασποράς των μεταβλητών που αφορούν στις αποζημιώσεις υλικών ζημιών ανά ζεύγη.



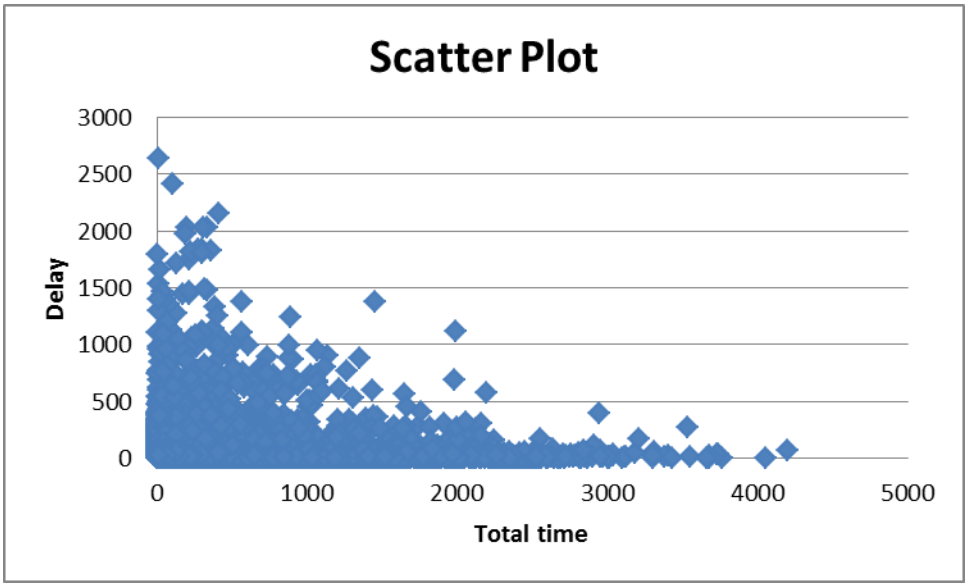
Σχήμα 2.10: Διάγραμμα διασποράς "Total payment" – "Total time".

Το διάγραμμα διασποράς, καθώς και ο συντελεστής συσχέτισης των μεταβλητών, που ισούται με 0,1, μαρτυρούν πως δεν υπάρχει γραμμική συσχέτιση μεταξύ του "Total payment" και "Total time" υλικών ζημιών.



Σχήμα 2.11: Διάγραμμα διασποράς "Total payment" – "Delay".

Από τη μορφή του διαγράμματος 2.10 δεν φαίνεται να υπάρχει συσχέτιση μεταξύ των τιμών που λαμβάνουν οι μεταβλητές “Total payment” και “Delay”. Ο συντελεστής συσχέτισης ισούται με 0,01 και επομένως δεν υπάρχει γραμμική συσχέτιση μεταξύ των δύο μεταβλητών.



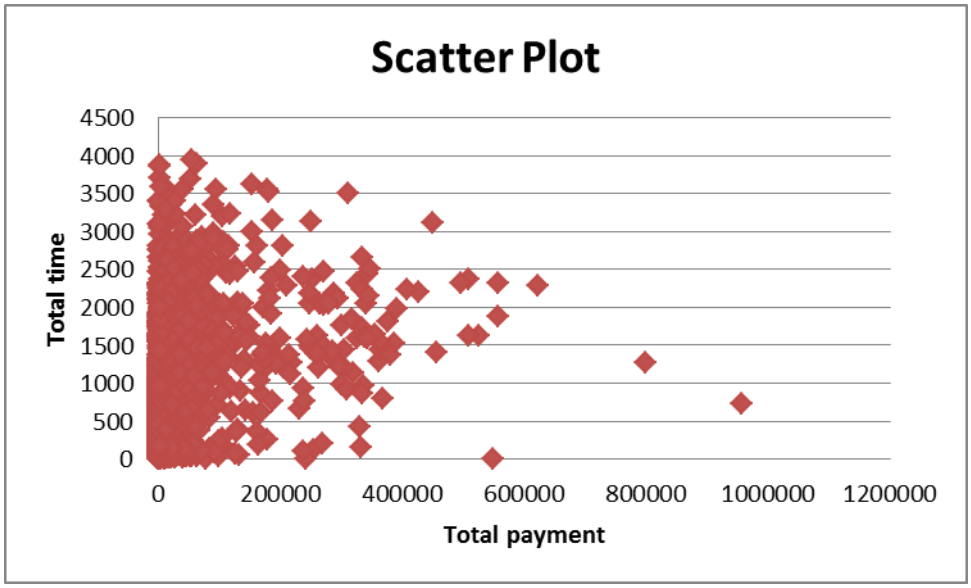
Σχήμα 2.12: Διάγραμμα διασποράς “Total time” – “Delay”.

Τέλος, και από το διάγραμμα 2.11 δεν βλέπουμε την ύπαρξη κάποιας γραμμικής σχέσης μεταξύ του “Total time” και του “Delay” υλικών ζημιών. Η τιμή του r που είναι ίση με 0,07, επιβεβαιώνει τη μη ύπαρξη συσχέτισης μεταξύ των τιμών των δύο μεταβλητών.

Συμπερασματικά, μπορούμε να πούμε πως δεν υπάρχει συσχέτιση μεταξύ των τιμών των μεταβλητών που περιγράφουν τις αποζημιώσεις υλικών ζημιών. Οι τιμές που λαμβάνει ο συντελεστής συσχέτισης είναι τέτοιες που μας επιτρέπουν να πούμε με σιγουριά πως οι τιμές που λαμβάνει η μία μεταβλητή δεν επηρεάζουν τις τιμές που θα λάβουν οι άλλες δύο, καθώς και στις τρεις περιπτώσεις οι τιμές του r είναι πολύ κοντά στο μηδέν.

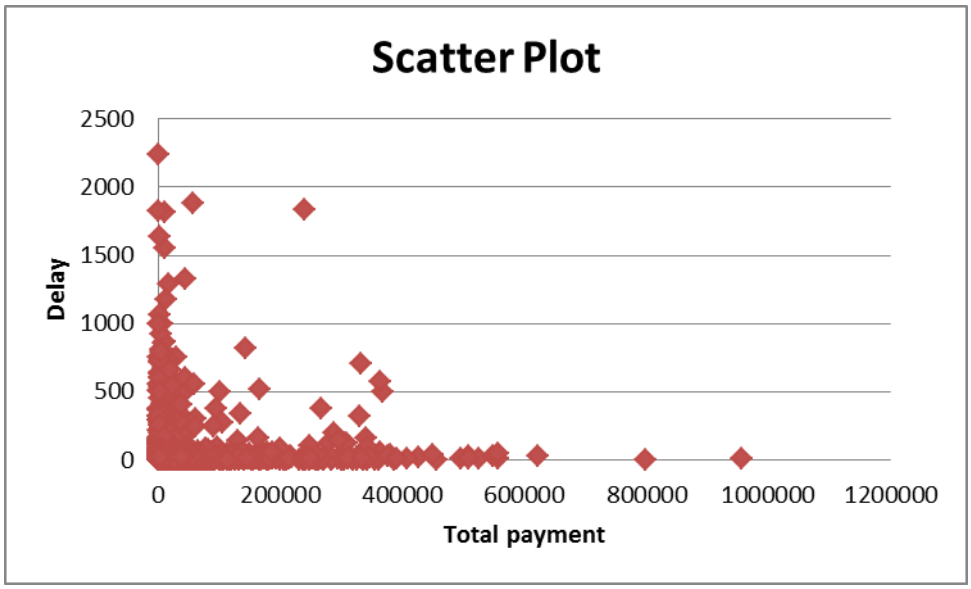
2.3.2 Συσχέτιση μεταξύ των μεταβλητών σωματικών βλαβών

Ακριβώς όπως εργαστήκαμε στην προηγούμενη υποενότητα για τις υλικές ζημιές, μελετούμε στην παρούσα την ύπαρξη συσχέτισης μεταξύ των μεταβλητών που αναφέρονται στις αποζημιώσεις σωματικών βλαβών. Κατασκευάζουμε τα διαγράμματα διασποράς και υπολογίζουμε τους συντελεστές συσχέτισης του Pearson για τα ζεύγη “Total payment” – “Total time” , “Total payment” – “Delay”, “Total time” – “Delay” ώστε να ελέγξουμε γραφικά αλλά και μέσω ενός συντελεστή την ύπαρξη γραμμικής συσχέτισης.



Σχήμα 2.13: Διάγραμμα διασποράς "Total payment" – "Total time".

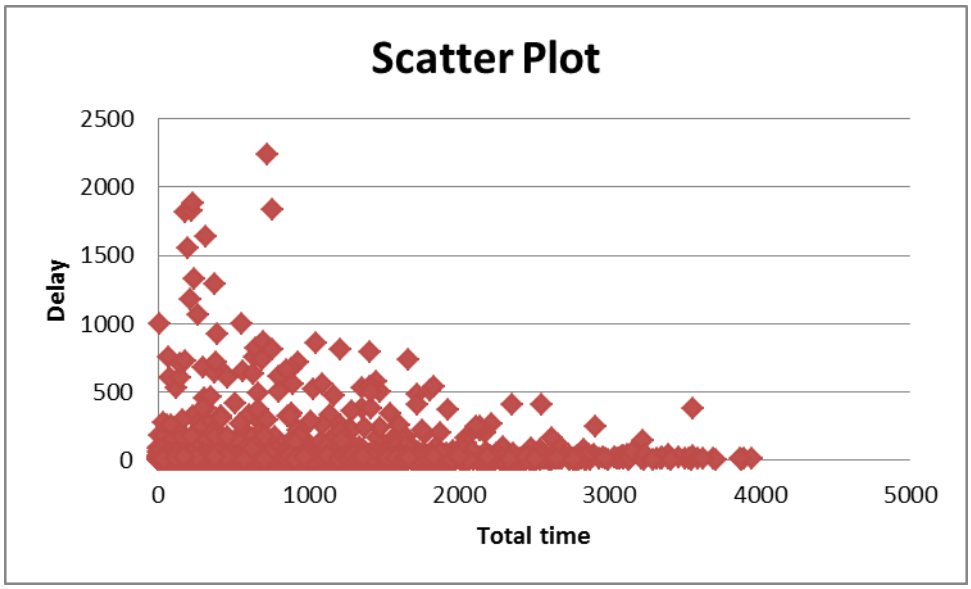
Από τη μορφή του διαγράμματος 2.12 δεν φαίνεται να υπάρχει συσχέτιση μεταξύ των τιμών που λαμβάνουν οι μεταβλητές "Total payment" και "Total time" σωματικών βλαβών. Ο συντελεστής συσχέτισης ισούται με 0,29 και επομένως σύμφωνα με τις χαρακτηριστικές τιμές που δώσαμε για το r δεν υπάρχει γραμμική συσχέτιση μεταξύ των δύο μεταβλητών.



Σχήμα 2.14: Διάγραμμα διασποράς "Total payment" – "Delay".

Στο ίδιο συμπέρασμα, δηλαδή τη μη ύπαρξη συσχέτισης, οδηγούμαστε παρατηρώντας το παραπάνω διάγραμμα διασποράς για το ύψος αποζημιώσεων και το

χρόνο καθυστέρησης αναγγελίας της ζημιάς στην ασφαλιστική. Η τιμή του r ίση με 0,05 επιβεβαιώνει ότι οι δύο μεταβλητές δεν παρουσιάζουν γραμμική συσχέτιση.



Σχήμα 2.15: Διάγραμμα διασποράς "Total time" – "Delay".

Το διάγραμμα διασποράς 2.14, καθώς και ο συντελεστής συσχέτισης των μεταβλητών, που ισούται με 0,001, μαρτυρούν πως δεν υπάρχει γραμμική συσχέτιση μεταξύ του "Total time" και "Delay " σωματικών βλαβών.

Όπως στις μεταβλητές των αποζημιώσεων υλικών ζημιών, έτσι και για τις μεταβλητές που περιγράφουν τις αποζημιώσεις σωματικών βλαβών δεν υπάρχει συσχέτιση μεταξύ των τιμών που λαμβάνουν.

Κεφάλαιο 3:Διερεύνηση της κατανομής των μεταβλητών

Σε συνέχεια της μελέτης των αποζημιώσεων κλάδου αστικής ευθύνης αυτοκινήτου, στο κεφάλαιο αυτό αναζητούμε τις κατανομές που θα περιγράψουν τη συμπεριφορά των μεταβλητών ύψος αποζημίωσης, χρόνος αποπληρωμής και χρόνος καθυστέρησης δήλωσης της ζημιάς στην ασφαλιστική εταιρία. Για το σκοπό αυτό, θα δοκιμάσουμε την προσαρμογή κατανομών πιθανότητας στις μεταβλητές αποζημιώσεων για τα αρχικά αλλά και για μετασχηματισμένα δεδομένα. Σε κάθε περίπτωση θα εκτελούνται στατιστικά τεστ καλής προσαρμογής για να διευκρινιστεί κατά πόσον η προσαρμογή είναι ικανοποιητική.

3.1 Διερεύνηση της κατανομής του ύψους των αποζημιώσεων

3.1.1 Προσαρμογή στα αρχικά δεδομένα

Μια πρώτη προσπάθεια διερεύνησης της κατανομής της μεταβλητής “Total payment” επιχειρήθηκε στην προηγούμενη ενότητα. Με τη βοήθεια διαγραμμάτων (ιστογράμματα συχνοτήτων, Normal Q-Q Plot διαγράμματα) και με τη διενέργεια των τεσσάρων τεστ κανονικότητας (Lilliefors, Anderson-Darling, Cramer-von Mises, Pearson normality test) ελέγχτηκε αν τα δεδομένα προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή. Τα αποτελέσματα των ελέγχων ήταν πως απορρίπτουμε την υπόθεση ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

Στην παρούσα υποενότητα με τη βοήθεια του προγράμματος “EasyFit” προχωρούμε στη διερεύνηση της κατανομής της μεταβλητής “Total payment” για τις δύο κατηγορίες ζημιών της ασφαλιστικής εταιρίας, δηλαδή για αποζημιώσεις υλικών ζημιών και σωματικών βλαβών, δοκιμάζοντας την προσαρμογή των δεδομένων σε διάφορες θεωρητικές κατανομές.

Το “EasyFit” αποτελεί ένα πρόγραμμα ανάλυσης και προσομοίωσης δεδομένων που επιτρέπει την προσαρμογή κατανομών πιθανότητας σε δείγμα δεδομένων, επιλέγει την καλύτερη κατανομή σύμφωνα με τα δεδομένα που εισάγονται προς ανάλυση δίνοντας έτσι τη δυνατότητα στο χρήστη να εφαρμόζει τα αποτελέσματα της ανάλυσης ώστε να λαμβάνει καλύτερες αποφάσεις. Το πρόγραμμα επιτρέπει την προσαρμογή δεδομένων σε θεωρητικές κατανομές σε ελάχιστο χρόνο εκτιμώντας με τη βοήθεια ισχυρού αλγόριθμου τις παραμέτρους των κατανομών αυτών.

Το “EasyFit” υποστηρίζει όλες τις συνήθως χρησιμοποιούμενες συνεχείς και διακριτές κατανομές (βλ. αναλυτικά Πίνακας 1 – Παράρτημα Α). Μάλιστα, πολλές κατανομές είναι διαθέσιμες σε δύο εκδοχές. Για παράδειγμα υποστηρίζονται η κατανομή Weibull με 2 και με 3 παραμέτρους. Επτά επιπλέον κατανομές είναι διαθέσιμες (βλ. αναλυτικά Πίνακας 2 – Παράρτημα Α).

Με την προσαρμογή των κατανομών στα δεδομένα, παράλληλα δίνονται από το πρόγραμμα οι εκτιμήσεις των παραμέτρων της κάθε κατανομής και πραγματοποιούνται τεστ καλής προσαρμογής. Τα τεστ καλής προσαρμογής (Goodness of fit (gof) tests) μετρούν την συμβατότητα του τυχαίου δείγματος με τη θεωρητική συνάρτηση κατανομής πιθανότητας, δηλαδή δείχνουν στον χρήστη πόσο καλά ταιριάζει η επιλεγμένη κατανομή στα προς εξέταση δεδομένα. Οι έλεγχοι καλής προσαρμογής που εκτελεί το "EasyFit" σε διάφορα επίπεδα σημαντικότητας είναι οι ακόλουθοι:

- Kolmogorov-Smirnov Goodness-of-fit test
- Anderson-Darling Goodness-of-fit test
- Chi-Squared Goodness-of-fit test

Τα αποτελέσματα των ελέγχων καλής προσαρμογής παραθέτονται συνολικά δίνοντας στο χρήστη τη δυνατότητα να επιλέξει ένα ή περισσότερα μοντέλα τα οποία ταιριάζουν καλύτερα στα δεδομένα. Επιπλέον, δίνεται η δυνατότητα σύγκρισης μεταξύ κατανομών και η τελική επιλογή λαμβάνοντας υπόψη το p-value του στατιστικού ελέγχου για την προσαρμογή των δεδομένων σε μια θεωρητική κατανομή.

Οι δοκιμές που πραγματοποιήθηκαν αφορούν στα αρχικά δεδομένα. Πιο αναλυτικά ελέγχθηκε η προσαρμογή των αρχικών τιμών της μεταβλητής "Total payment" στις συνεχείς κατανομές που περιλαμβάνει το πακέτο "EasyFit". Σύμφωνα με τα τρία τεστ καλής προσαρμογής που χρησιμοποιεί το πακέτο (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test) απορρίπτουμε σε επίπεδο σημαντικότητας 1% την υπόθεση ότι τα δεδομένα προέρχονται από κάποια από τις παραπάνω 57 θεωρητικές κατανομές.

3.1.2 Προσαρμογή σε μετασχηματισμένα δεδομένα

Η ανάλυση στην προηγούμενη ενότητα των ανεξάρτητων μεταβλητών "Total payment", "Total time" και "Delay" και για τις δύο κατηγορίες ζημιών, παρόλο που μας έδωσε μία πρώτη εικόνα για τα δεδομένα, δεν οδήγησε στην εξαγωγή συμπερασμάτων για την κατανομή που αυτά ακολουθούν. Κυρίως οι απεικονίσεις των ιστογραμμάτων είναι τέτοιες που δεν μας επέτρεψαν να έχουμε εικόνα για την κατανομή των δεδομένων. Εφόσον η υπόθεση της κανονικότητας των δεδομένων απορρίπτεται και από τα τεστ κανονικότητας που διενεργήσαμε και με τη βοήθεια του προγράμματος "EasyFit" δεν μπορέσαμε να δεχτούμε την υπόθεση πως τα δεδομένα προέρχονται από κάποια γνωστή κατανομή, θα τροποποιήσουμε τα δεδομένα έτσι ώστε να πάρουμε πιο σαφή εικόνα για τη μορφή τους καθώς και για την κατανομή που αυτά ακολουθούν. Στην παρούσα ενότητα θα επανεξετάσουμε τα δεδομένα στην περίπτωση που αυτά μετασχηματιστούν.

Ο μετασχηματισμός των τιμών μιας μεταβλητής είναι μία πολύ δημοφιλής πρακτική στη στατιστική. Οι μετασχηματισμοί στοχεύουν στον υπολογισμό βασικών μέτρων θέσης και διασποράς όπως και στην περιγραφή συνολικά της κατανομής των τιμών. Οι μετασχηματισμοί εφαρμόζονται για να βελτιώσουν την εμφάνιση των γραφημάτων ώστε ο παρατηρητής να μπορεί να έχει καλύτερη εικόνα της συμπεριφοράς των τιμών μιας μεταβλητής. Στη στατιστική ανάλυση ο κυριότερος λόγος για τον οποίο πραγματοποιούνται μετασχηματισμοί των αρχικών δεδομένων είναι η εξέταση του αν τα δεδομένα προέρχονται από την κανονική κατανομή, κάτι που επιτρέπει τη χρήση στατιστικών μεθόδων που στηρίζονται στην υπόθεση της κανονικότητας.

Στην υποενότητα αυτή θα εφαρμόσουμε στα αρχικά δεδομένα των “Total payment”, “Total time” και “Delay” και για τις δύο κατηγορίες ζημιών τους παρακάτω τέσσερις κοινώς χρησιμοποιούμενους μετασχηματισμούς:

- Λογάριθμος των αρχικών δεδομένων
- Τετραγωνική ρίζα των αρχικών δεδομένων
- Κυβική ρίζα των αρχικών δεδομένων
- Διπλός λογάριθμος των αρχικών δεδομένων

Σημειώνεται ότι για να εφαρμοστεί ο λογάριθμος στα αρχικά δεδομένα για τις μεταβλητές “Total time” και “Delay” προστέθηκε στις μηδενικές τιμές η μονάδα ώστε να μην απειρίζεται το αποτέλεσμα του μετασχηματισμού.
Κατά τον ίδιο τρόπο για να εφαρμοστεί ο διπλός λογάριθμος στα αρχικά δεδομένα για τις μεταβλητές “Total time” και “Delay” προστέθηκε στις μηδενικές τιμές το δύο και στις αρχικές τιμές που ισούνται με μονάδα, η μονάδα ώστε η μικρότερη τιμή που θα εμφανίζουν τα δεδομένα πριν μετασχηματιστούν να είναι ίση με δύο.

Για κάθε μετασχηματισμό των δεδομένων που θα πραγματοποιηθεί θα δίνονται περιγραφικά χαρακτηριστικά, καθώς και τα ιστογράμματα συχνοτήτων και τα Normal Q-Q Plot. Τέλος, για κάθε μετασχηματισμό θα πραγματοποιούνται τα τέσσερα τεστ κανονικότητας που χρησιμοποιήσαμε και στην προηγούμενη ενότητα για να ελεγχθεί η υπόθεση ότι τα μετασχηματισμένα δεδομένα προέρχονται από την κανονική κατανομή. Οι υπολογισμοί πραγματοποιούνται με τη βοήθεια του στατιστικού πακέτου R.

3.1.2.1 Μετασχηματισμός λογάριθμος

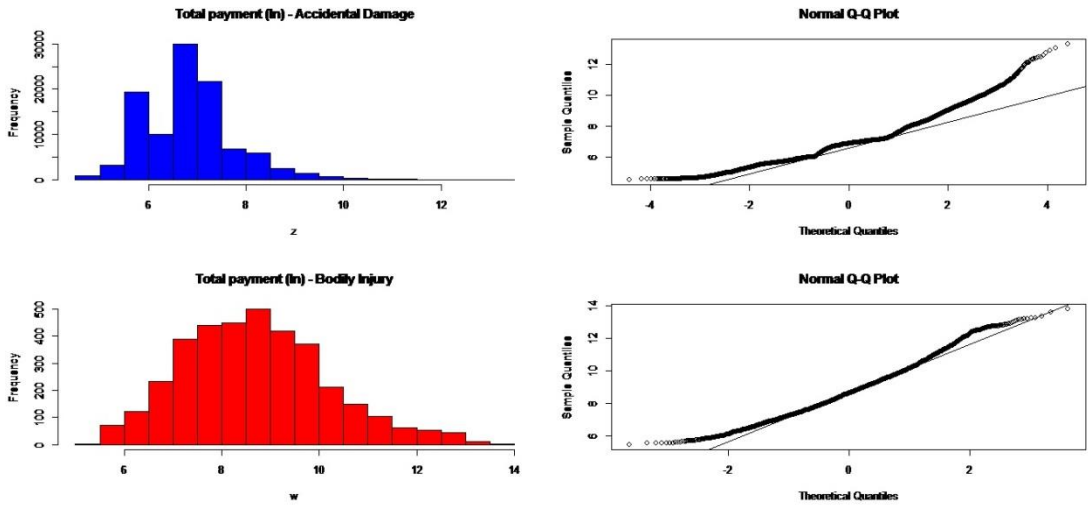
Για το ύψος αποζημιώσεων “Total payment” παρουσιάζονται στον παρακάτω πίνακα τα περιγραφικά χαρακτηριστικά του λογαρίθμου των αρχικών δεδομένων. Η δεύτερη στήλη αφορά στις αποζημιώσεις υλικών ζημιών και η τρίτη στις αποζημιώσεις σωματικών βλαβών.

Πίνακας 3.1.1: Περιγραφικά χαρακτηριστικά – μετασχηματισμός λογάριθμος.

Total payment	Accidental Damage	Bodily Injury
min	4.53	5.46
25% quantile	6.02	7.60
median (50% quantile)	6.86	8.61
mean	6.82	8.7
75% quantile	7.15	9.62
max	13.29	13.77
variance	0.8146	2.2022
standard deviation	0.9026	1.484
99% quantile	9.4669	12.697
skewness	0.6929	0.4711
kurtosis	4.3376	3.0411

Τα περιγραφικά χαρακτηριστικά, όπως και στα αρχικά δεδομένα, παίρνουν μεγαλύτερες τιμές για τις αποζημιώσεις σωματικών βλαβών. Η διαπίστωση αυτή είναι αναμενόμενη, καθώς οι μετασχηματισμοί εφαρμόζονται στα αρχικά δεδομένα για τα οποία έχουμε ήδη μία πρώτη εικόνα από την σύγκριση που πραγματοποιήσαμε στην ενότητα 2.2.

Η διαγραμματική απεικόνιση θα μας δώσει σημαντικές πληροφορίες για τη μορφή των δεδομένων και για το αν τουλάχιστον διαγραμματικά μπορούμε να υποθέσουμε ότι τα μετασχηματισμένα δεδομένα προέρχονται από την κανονική κατανομή.



Σχήμα 3.1.1: Ιστογράμματα και Normal Q-Q Plots ln (Total payment).

Η μορφή των ιστογραμμάτων στην περίπτωση του λογαρίθμου των αρχικών δεδομένων είναι σαφώς βελτιωμένη και μας δίνει εικόνα για την συχνότητα του ύψους των αποζημιώσεων. Σε αντίθεση με τα ιστογράμματα των αρχικών δεδομένων που εμφάνιζαν συγκέντρωση τιμών στην πρώτη κλάση και επομένως δεν μας έδιναν τη δυνατότητα παρατήρησης της συχνότητας σε διάφορα ύψη αποζημιώσεων, τώρα μπορούμε να πάρουμε σημαντικές πληροφορίες για τη μελέτη μας.

Όσον αφορά στις αποζημιώσεις υλικών ζημιών, το ιστόγραμμα μαρτυρά ότι τα δεδομένα προέρχονται από κατανομή που παρουσιάζει τουλάχιστον 2 κορυφές. Αν παρατηρήσουμε τη μορφή του ιστογράμματος θα δούμε την πρώτη κορυφή για λογάριθμο αποζημιώσεων ίσο με 6 και τη δεύτερη κορυφή στην τιμή 7. Το Q-Q plot δεν επιδεικνύει καλή προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά κάποια κατανομή με βαριά δεξιά ουρά.

Όσον αφορά στις αποζημιώσεις σωματικών βλαβών, η μορφή του ιστογράμματος δείχνει ότι τα δεδομένα κατανέμονται πιο συμμετρικά και η κατανομή είναι μονοκόρυφη με μέγιστη παρατηρούμενη συχνότητα στην τιμή 9 περίπου. Επίσης τα δεδομένα προέρχονται από κατανομή που παρουσιάζει θετική ασυμμετρία, όπως φαίνεται και στον πίνακα περιγραφικών χαρακτηριστικών, ίση με 0,4711. Η εικόνα του Q-Q plot είναι καλύτερη στις σωματικές βλάβες. Εδώ, δεν παρατηρούμε τόσο μεγάλη απόκλιση των σημείων από την ευθεία της κανονικής κατανομής.

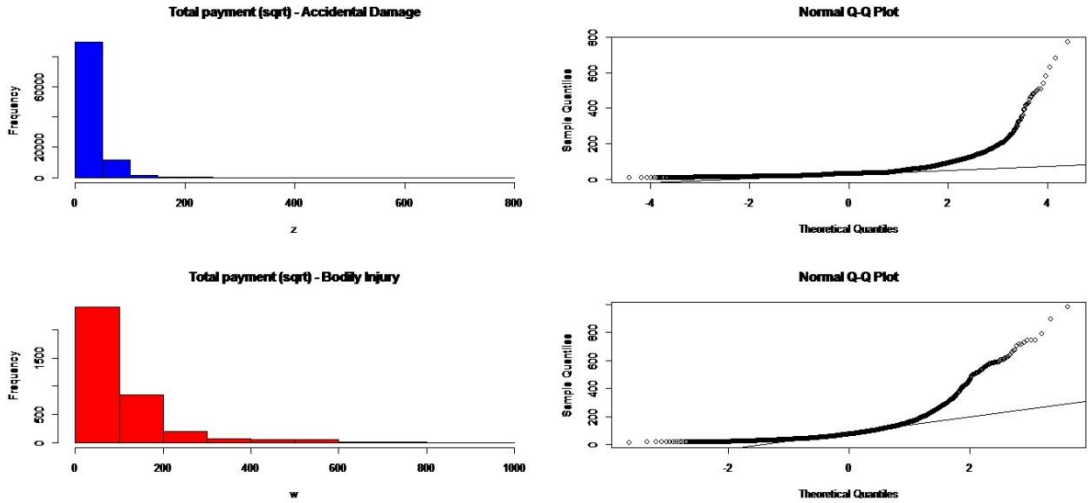
3.1.2.2 Μετασχηματισμός τετραγωνική ρίζα

Στον παρακάτω πίνακα δίνονται τα περιγραφικά χαρακτηριστικά για το μετασχηματισμό τετραγωνική ρίζα των τιμών των αποζημιώσεων.

Πίνακας 3.1.2: Περιγραφικά χαρακτηριστικά – μετασχηματισμός τετραγωνική ρίζα.

Total payment	Accidental Damage	Bodily Injury
min	9.64	15.33
25% quantile	20.30	44.72
median (50% quantile)	30.85	74.16
mean	33.95	105.2
75% quantile	35.74	122.5
max	770.7	978.8
variance	436.777	10598.54
standard deviation	20.8992	102.949
99% quantile	113.6858	571.645
skewness	5.5727	2.9948
kurtosis	86.9978	14.5423

Από τα αποτελέσματα του πίνακα παρατηρούμε ότι οι αποζημιώσεις σωματικών βλαβών παρουσιάζουν μεγάλη μεταβλητότητα σε σχέση με αυτή των υλικών ζημιών. Η διακύμανση για υλικές ζημιές ισούται με 436,777 ενώ για σωματικές βλάβες είναι σχεδόν τριακονταπλάσια. Επιπλέον, η τιμή του συντελεστή κύρτωσης για αποζημιώσεις υλικών ζημιών μαρτυρά πολύ μεγάλη συγκέντρωση γύρω από κάποια τιμή. Τα ιστογράμματα και τα Q-Q plots θα μας δώσουν πιο καθαρή εικόνα.



Σχήμα 3.1.2: Ιστογράμματα και Normal Q-Q Plots sqrt (Total payment).

Και για τις δύο κατηγορίες αποζημιώσεων, για το μετασχηματισμό τετραγωνική ρίζα των αποζημιώσεων, παρατηρούμε συγκέντρωση τιμών στην πρώτη κλάση του κάθε ιστογράμματος αντίστοιχα. Από τα Q-Q plots συμπεραίνουμε πως τα δεδομένα δεν προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή, αλλά από κάποια κατανομή με βαριά δεξιά ουρά.

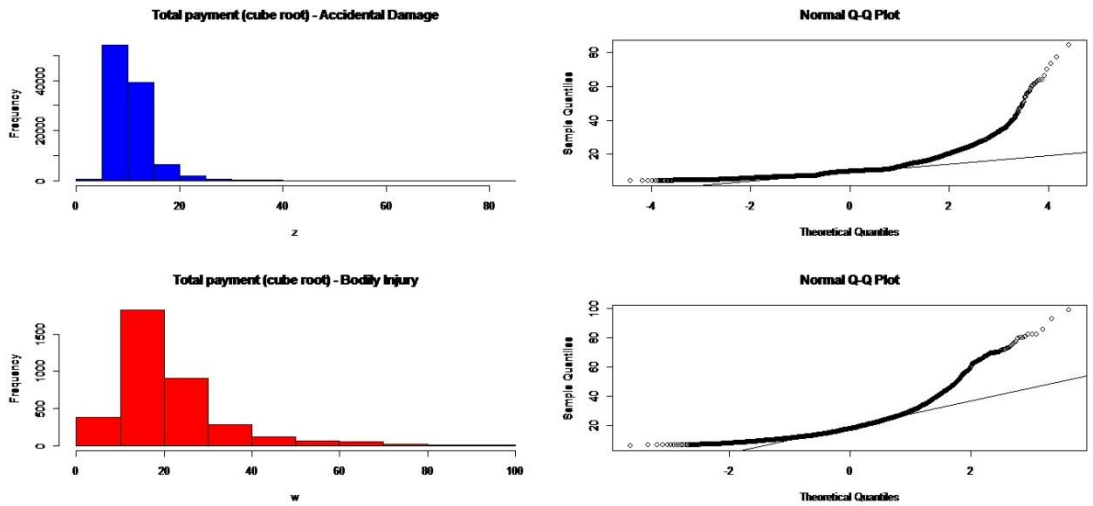
3.1.2.3 Μετασχηματισμός κυβική ρίζα

Στη συνέχεια παρουσιάζουμε στον πίνακα τα περιγραφικά χαρακτηριστικά για τον μετασχηματισμό κυβική ρίζα των αποζημιώσεων για τις δύο κατηγορίες ζημιών.

Πίνακας 3.1.3: Περιγραφικά χαρακτηριστικά – μετασχηματισμός κυβική ρίζα.

Total payment	Accidental Damage	Bodily Injury
min	4.53	6.17
25% quantile	7.44	12.60
median (50% quantile)	9.84	17.65
mean	10.20	20.73
75% quantile	10.85	24.7
max	84.1	98.6
variance	13.051	146.54
standard deviation	3.6126	12.105
99% quantile	23.4677	68.879
skewness	2.7645	2.0493
kurtosis	21.8818	8.413

Τα αποτελέσματα επαληθεύουν για ακόμα μία φορά τους αρχικούς μας ισχυρισμούς όπως είναι αναμενόμενο. Οι αποζημιώσεις υλικών ζημιών γενικά παίρνουν τιμές μικρότερες από τις αποζημιώσεις σωματικών βλαβών. Η διακύμανση στις αποζημιώσεις σωματικών βλαβών είναι πολύ μεγαλύτερη, ενώ αντίθετα στις υλικές ζημίες έχουμε μεγαλύτερη συγκέντρωση σε κάποιες τιμές.



Σχήμα 3.1.3: Ιστογράμματα και Normal Q-Q Plots cube root (Total payment).

Και στην περίπτωση του μετασχηματισμού κυβική ρίζα των αποζημιώσεων τα παραπάνω διαγράμματα επιβεβαιώνουν πως τα δεδομένα δεν προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή.

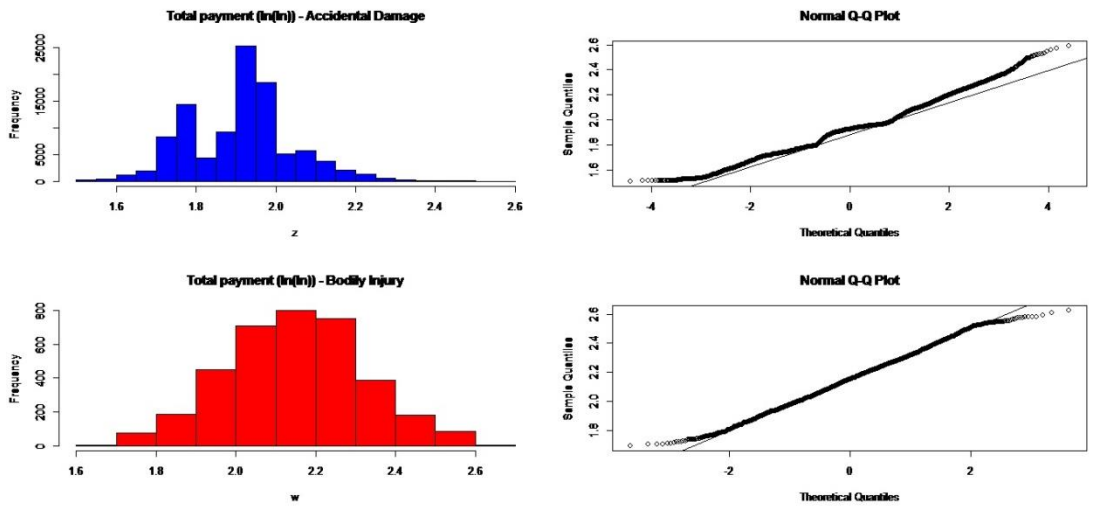
3.1.2.4 Μετασχηματισμός διπλός λογάριθμος

Τέλος, εξετάζουμε την περίπτωση του μετασχηματισμού του διπλού λογαρίθμου του ύψους των αποζημιώσεων. Δίνουμε και εδώ, συγκεντρωτικά, σε ένα πίνακα τα περιγραφικά χαρακτηριστικά για τις δύο κατηγορίες ζημιών.

Πίνακας 3.1.4: Περιγραφικά χαρακτηριστικά – μετασχηματισμός διπλός λογάριθμος.

Total payment	Accidental Damage	Bodily Injury
min	1.51	1.70
25% quantile	1.80	2.03
median (50% quantile)	1.93	2.15
mean	1.91	2.149
75% quantile	1.97	2.3
max	2.587	2.623
variance	0.017	0.03
standard deviation	0.1298	0.169
99% quantile	2.2478	2.541
skewness	0.2031	0.0285
kurtosis	3.3953	2.728

Τα αποτελέσματα του πίνακα δεν έχουν να μας δώσουν κάποια επιπλέον πληροφορία σε σχέση με τα περιγραφικά χαρακτηριστικά των τριών προηγούμενων μετασχηματισμών.



Σχήμα 3.1.4: Ιστογράμματα και Normal Q-Q Plots ln (ln (Total payment)).

Η μορφή των ιστογραμμάτων στην περίπτωση του διπλού λογαρίθμου των αρχικών δεδομένων μας δίνει πληροφορίες για την συχνότητα του ύψους των αποζημιώσεων. Σε αντίθεση με τα ιστογράμματα των μετασχηματισμών τετραγωνική και κυβική ρίζα, η εικόνα που παρουσιάζουν εδώ τα δεδομένα είναι η ίδια περίπου με την περίπτωση του λογάριθμου.

Όσον αφορά στις αποζημιώσεις υλικών ζημιών, το ιστόγραμμα μαρτυρά ότι τα δεδομένα προέρχονται από κατανομή που παρουσιάζει 3 κορυφές. Η πρώτη κορυφή παρατηρείται στην τιμή 1,75 η δεύτερη στην τιμή 1,9 και η τρίτη στην τιμή 2,1. Το Q-Q plot δεν επιδεικνύει καλή προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά κάποια κατανομή με βαριά δεξιά ουρά. Το συμπέρασμα αυτό είναι δεδομένο από το ιστόγραμμα καθώς μία κατανομή με 3 κορυφές δεν μπορεί να είναι η κανονική κατανομή.

Όσον αφορά στις αποζημιώσεις σωματικών βλαβών, η μορφή του ιστογράμματος δείχνει ότι τα δεδομένα κατανέμονται πιο συμμετρικά και εμφανίζουν μία κορυφή. Μάλιστα, από τις τιμές των περιγραφικών χαρακτηριστικών του παραπάνω πίνακα βλέπουμε πως τα δεδομένα κατανέμονται σχεδόν συμμετρικά, καθώς ο συντελεστής ασυμμετρίας παίρνει τιμή πολύ κοντά στο μηδέν ίση με 0,0285. Επιπλέον, ο συντελεστής κύρτωσης ισούται με 2,728 δηλαδή πολύ κοντά στο 3, την τιμή της μεσόκυρτης – κανονικής κατανομής. Το Q-Q plot των σωματικών βλαβών μπορούμε να πούμε πως ίσως να υποδεικνύει πως ο διπλός λογάριθμος του ύψους των αποζημιώσεων προέρχεται από πληθυσμό που ακολουθεί κανονική κατανομή. Δεν παρατηρούμε απόκλιση των σημείων από τη γραμμή της κανονικής κατανομής, πέραν κάποιων ακραίων παρατηρήσεων στη δεξιά ουρά.

Θα επαληθεύσουμε τα παραπάνω συμπεράσματα από τα ιστογράμματα και τα Q-Q plots με τα τέσσερα τεστ κανονικότητας που χρησιμοποιήσαμε και στην προηγούμενη ενότητα. Τα αποτελέσματα παρουσιάζονται συνοπτικά στον παρακάτω πίνακα.

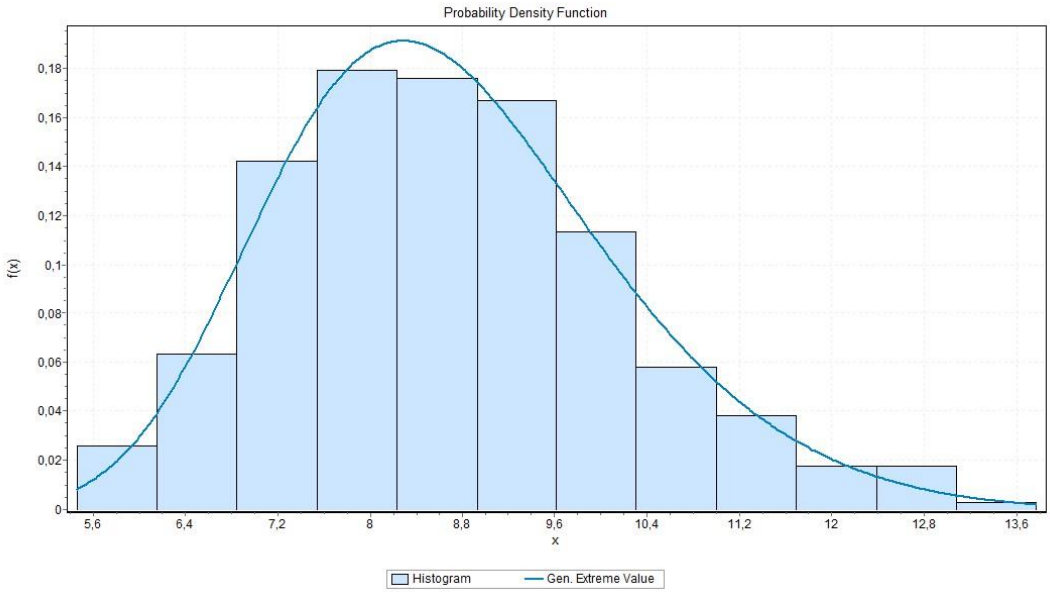
Πίνακας 3.1.5: Αποτελέσματα των τεστ κανονικότητας του "Total payment".

Normality Testing	**Ln(Total payment)**			
	Accidental Damage		**Bodily Injury**	
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.1198	<2,2 e-16	0.0355	<2,2 e-16
Anderson-Darling	1283.843	<2,2 e-16	11.5567	<2,2 e-16
Cramer-von Mises	234.6349	7.37E-10	1.5622	7.37E-10
Pearson	88979.27	<2,2 e-16	268.8655	<2,2 e-16
	Sqrt(Total payment)			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.2323	<2,2 e-16	0.1927	<2,2 e-16
Anderson-Darling	7090.255	<2,2 e-16	311.8053	<2,2 e-16
Cramer-von Mises	1305.356	7.37E-10	55.9903	7.37E-10
Pearson	147897.6	<2,2 e-16	3587.781	<2,2 e-16
	Cube root (Total payment)			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.1926	<2,2 e-16	0.1356	<2,2 e-16
Anderson-Darling	4090.305	<2,2 e-16	167.9746	<2,2 e-16
Cramer-von Mises	729.7132	7.37E-10	28.0099	7.37E-10
Pearson	119561	<2,2 e-16	1738.725	<2,2 e-16
	Ln(ln(Total payment))			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.0972	<2,2 e-16	0.0173	0.01419
Anderson-Darling	1044.019	<2,2 e-16	0.9817	0.01365
Cramer-von Mises	213.6937	7.37E-10	0.1338	0.03908
Pearson	75715.11	<2,2 e-16	112.4158	1.618 e-06

Τα αποτελέσματα που προκύπτουν από τους ελέγχους κανονικότητας επιβεβαιώνουν την αρχική μας πρόβλεψη ότι τα δεδομένα δεν προέρχονται από την κανονική κατανομή. Με εξαίρεση τα p-value των ελέγχων για το μετασχηματισμό του διπλού λογαρίθμου του ύψους αποζημιώσεων σωματικών βλαβών, τα p-value των ελέγχων εμφανίζουν την ελάχιστη τιμή που δίνεται από το στατιστικό πακέτο R. Και στην περίπτωση του διπλού λογαρίθμου για σωματικές βλάβες απορρίπτουμε τη μηδενική υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή, καθώς για κάθε ένα από τα τεστ το p-value είναι μικρότερο από το επίπεδο σημαντικότητας 5%. Επομένως, απορρίπτουμε για κάθε μετασχηματισμό τη μηδενική υπόθεση, δηλαδή την υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή σε επίπεδο σημαντικότητας 5%.

Για τους τέσσερις παραπάνω μετασχηματισμούς της μεταβλητής “Total payment”, δοκιμάστηκε η προσαρμογή τους και στις θεωρητικές συνεχείς κατανομές που είναι διαθέσιμες από το πρόγραμμα “EasyFit”. Με εξαίρεση τους μετασχηματισμούς λογάριθμος και κυβική ρίζα του “Total payment” σωματικών βλαβών, σύμφωνα με τα τρία τεστ καλής προσαρμογής που εκτελούνται αυτόματα από το “EasyFit” (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test), απορρίπτουμε την υπόθεση ότι τα μετασχηματισμένα δεδομένα προέρχονται από κάποια από τις κατανομές αυτές.

Για το μετασχηματισμό λογάριθμος του ύψους αποζημιώσεων σωματικών βλαβών, δίνεται το ιστόγραμμα συχνοτήτων όπως εξήχθη από το “EasyFit”.



Σχήμα 3.1.5: Ιστόγραμμα λογαρίθμου “Total payment” σωματικών βλαβών.

Σύμφωνα με τα τεστ καλής προσαρμογής του “EasyFit” δεν απορρίπτουμε την υπόθεση ότι ο λογάριθμος του “Total payment” σωματικών βλαβών ακολουθεί την κατανομή Generalized Extreme Value (GEV). Η μπλε γραμμή που εμφανίζεται στο ιστόγραμμα είναι η καμπύλη της συνάρτησης πυκνότητας πιθανότητας της κατανομής GEV.

Τα αποτελέσματα των ελέγχων καλής προσαρμογής του προγράμματος εμφανίζονται στον παρακάτω πίνακα.

Πίνακας 3.1.6: Αποτελέσματα των τεστ προσαρμογής του "EasyFit" για τη GEV.

Kolmogorov-Smirnov					
Sample Size	3629				
Statistic	0,01437				
P-Value	0,43773				
Rank	2				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,01781	0,0203	0,02254	0,0252	0,02704
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	3629				
Statistic	0,86327				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	13,749	19,286	25,018	32,892	39,074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom	11				
Statistic	19,932				
P-Value	0,04628				
Rank	11				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	14,631	17,275	19,675	22,618	24,725
Reject?	Yes	Yes	Yes	No	No

Σύμφωνα με τα αποτελέσματα του πίνακα, η υπόθεση ότι τα δεδομένα ακολουθούν την κατανομή GEV δεν απορρίπτεται από τα τεστ Kolmogorov-Smirnov και Anderson-Darling καθώς το p-value των ελέγχων είναι μεγαλύτερο από τα χρησιμοποιούμενα επίπεδα σημαντικότητας. Αντίθετα, σύμφωνα με το Chi-squared test σε επίπεδα σημαντικότητας μεγαλύτερο του 5% η υπόθεση ότι τα δεδομένα προέρχονται από την κατανομή GEV απορρίπτεται. Ο λόγος που επιλέγεται η κατανομή GEV είναι ότι αξιολογείται ως η προτιμότερη για την περιγραφή των δεδομένων του δείγματος σύμφωνα με το Anderson-Darling gof test. Υπενθυμίζεται ότι το Anderson-Darling αποτελεί τροποποίηση του Kolmogorov-Smirnov τεστ και δίνει μεγαλύτερη βαρύτητα στις ουρές της κατανομής του δείγματος απ΄ ότι το Kolmogorov-Smirnov τεστ.

Δίνονται, τέλος, οι εκτιμήσεις των τριών παραμέτρων της κατανομής, όπως υπολογίζονται από το πρόγραμμα.

21	Gen. Extreme Value	k=-0,13889 σ=1,3468 μ=8,087

Η κατανομή Generalized Extreme Value (GEV) είναι ένα ευέλικτο μοντέλο τριών παραμέτρων που συνδυάζει τις κατανομές Gumbel, Frechet και Weibull.

Η συνάρτηση πυκνότητας πιθανότητας της GEV είναι:

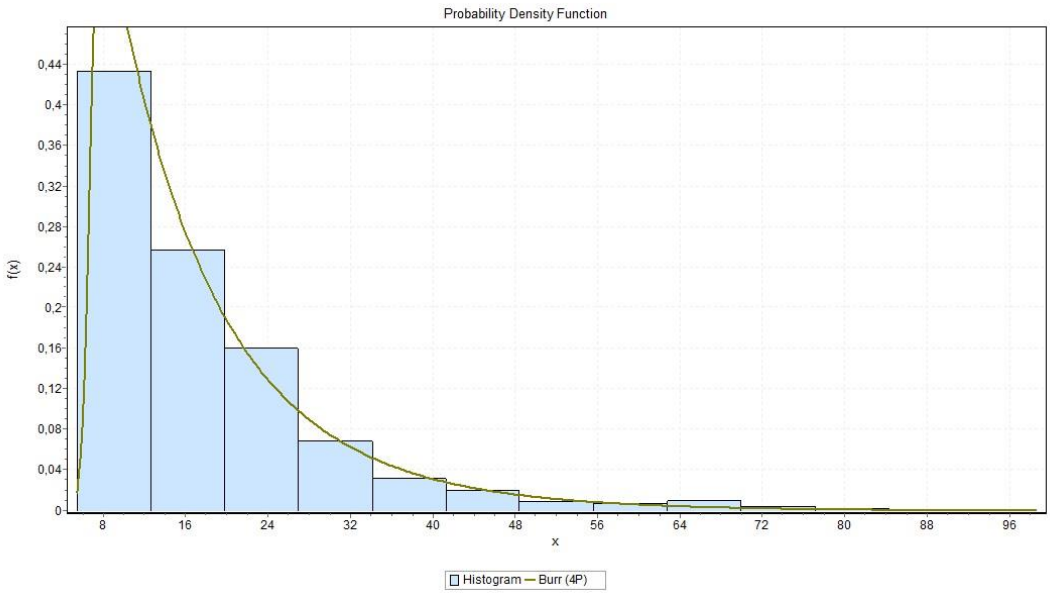
$$f(x) = \begin{cases} \frac{1}{\sigma}\, e^{-(1+kz)^{\frac{1}{k}}}\, (1+kz)^{-1-\frac{1}{k}} & k \neq 0 \\ \frac{1}{\sigma}\, e^{-z-e^{-z}} & k = 0 \end{cases}$$

όπου $z = \frac{x-\mu}{\sigma}$ και k, σ, μ οι παράμετροι σχήματος (shape), κλίμακας (scale) και θέσης (location) αντίστοιχα. Η παράμετρος κλίμακας πρέπει να είναι θετική (σ >0) και οι παράμετροι σχήματος και θέσης παίρνουν πραγματικές τιμές. Το πεδίο ορισμού της κατανομής GEV εξαρτάται από την τιμή της παραμέτρου σχήματος k.

- Για $k \neq 0$, το πεδίο ορισμού της GEV είναι: $1 + k\frac{x-\mu}{\sigma} > 0$
- Για $k = 0$, το πεδίο ορισμού της GEV είναι: $-\infty < x < +\infty$

Για κατάλληλες τιμές της παραμέτρου σχήματος k προκύπτουν οι τρεις τύποι κατανομών ακροτάτων. Πιο συγκεκριμένα για k=0 έχουμε κατανομή Gumbel, για k >0 κατανομή Frechét και για k <0 την αντεστραμμένη κατανομή Weibull. Επομένως, όταν προσαρμόζεις την GEV σε δείγμα τιμών, η τιμή της παραμέτρου k υποδεικνύει συνήθως ποιο από τα τρία μοντέλα περιγράφει καλύτερα τα δεδομένα.

Τέλος, δίνεται το ιστόγραμμα συχνοτήτων όπως εξήχθη από το "EasyFit", για το μετασχηματισμό κυβική ρίζα του ύψους αποζημιώσεων σωματικών βλαβών.



Σχήμα 3.1.6: Ιστόγραμμα κυβικής ρίζας "Total payment" σωματικών βλαβών.

Σύμφωνα με τα τεστ καλής προσαρμογής του “EasyFit” δεν απορρίπτουμε την υπόθεση ότι η κυβική ρίζα του “Total payment” σωματικών βλαβών ακολουθεί την κατανομή Burr με τέσσερις παραμέτρους (Burr (4P)). Η πράσινη γραμμή που εμφανίζεται στο ιστόγραμμα είναι η καμπύλη της συνάρτησης πυκνότητας πιθανότητας της κατανομής Burr (4P).

Τα αποτελέσματα των ελέγχων καλής προσαρμογής που πραγματοποιήθηκαν αυτόματα μέσω του προγράμματος εμφανίζονται στον παρακάτω πίνακα.

Πίνακας 3.1.7: Αποτελέσματα των τεστ προσαρμογής του “EasyFit” για τη Burr (4P).

Kolmogorov-Smirnov					
Sample Size	4784				
Statistic	0,01969				
P-Value	0,04841				
Rank	3				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,01551	0,01768	0,01963	0,02195	0,02355
Reject?	Yes	Yes	Yes	No	No
Anderson-Darling					
Sample Size	4784				
Statistic	18,433				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	13,749	19,286	25,018	32,892	39,074
Reject?	Yes	No	No	No	No
Chi-Squared					
Deg. of freedom	12				
Statistic	44,715				
P-Value	1.15E-01				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	15,812	18,549	21,026	24,054	26,217
Reject?	Yes	Yes	Yes	Yes	Yes

Σύμφωνα με τα αποτελέσματα του πίνακα, η υπόθεση ότι τα δεδομένα ακολουθούν την κατανομή Burr (4P) δεν απορρίπτεται από τα τεστ Kolmogorov-Smirnov και Anderson-Darling σε κάποια επίπεδα σημαντικότητας, καθώς το p-value των ελέγχων είναι μεγαλύτερο από αυτά τα επίπεδα σημαντικότητας. Αντίθετα, σύμφωνα με το Chi-squared test στα χρησιμοποιούμενα επίπεδα σημαντικότητας από το πρόγραμμα απορρίπτεται η υπόθεση ότι τα δεδομένα προέρχονται από πληθυσμό που ακολουθεί την κατανομή Burr (4P). Η επιλογή της κατανομής Burr (4P) ως η προτιμότερη για να περιγράψει τα δεδομένα αποζημιώσεων γίνεται και πάλι με βάση το Anderson-Darling gof test.

Δίνονται, τέλος, οι εκτιμήσεις των τεσσάρων παραμέτρων της κατανομής, όπως υπολογίζονται από το πρόγραμμα.

3	Burr (4P)	k=0,0333 α=526,12 β=183,28 γ=-176,56

Η κατανομή Burr (4P) προσδιορίζεται από τις 4 παραμέτρους:

- k – συνεχής, παράμετρος σχήματος (k >0)
- α – συνεχής, παράμετρος σχήματος (α >0)
- β – συνεχής, παράμετρος κλίμακας (β >0)
- γ – συνεχής, παράμετρος θέσης

Το πεδίο ορισμού της κατανομής είναι το διάστημα: $\gamma \leq x \leq +\infty$

Η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της κατανομής Burr (4P) είναι:

$$f(x) = \frac{ak\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}}{\beta\left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{k+1}}$$

Η αθροιστική συνάρτηση πυκνότητας πιθανότητας (α.σ.π.π.) είναι:

$$F(x) = 1 - \left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{-k}$$

3.2 Διερεύνηση της κατανομής του χρόνου αποπληρωμής

3.2.1 Προσαρμογή στα αρχικά δεδομένα

Μια πρώτη προσπάθεια διερεύνησης της κατανομής της μεταβλητής “Total time” επιχειρήθηκε στην προηγούμενη ενότητα. Με τη βοήθεια διαγραμμάτων (ιστογράμματα συχνοτήτων, Normal Q-Q Plot διαγράμματα) και με τη διενέργεια των τεσσάρων τεστ κανονικότητας (Lilliefors, Anderson-Darling, Cramer-von Mises, Pearson normality test) ελέγχτηκε αν τα δεδομένα προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή. Τα αποτελέσματα των ελέγχων ήταν πως απορρίπτουμε την υπόθεση ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

Στην παρούσα υποενότητα με τη βοήθεια του προγράμματος “EasyFit” προχωρούμε στη διερεύνηση της κατανομής της μεταβλητής “Total time” για τις δύο κατηγορίες ζημιών της ασφαλιστικής εταιρίας, δηλαδή για αποζημιώσεις υλικών ζημιών και σωματικών βλαβών, δοκιμάζοντας την προσαρμογή των δεδομένων σε διάφορες θεωρητικές κατανομές.

Οι δοκιμές που έγιναν αφορούν στα αρχικά δεδομένα. Πιο αναλυτικά ελέγχθηκε η προσαρμογή των αρχικών τιμών της μεταβλητής “Total time” στις συνεχείς και διακριτές κατανομές που περιλαμβάνει το πακέτο “EasyFit”. Σύμφωνα με τα τρία τεστ καλής προσαρμογής που χρησιμοποιεί το πακέτο (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test) απορρίπτουμε σε κάθε ένα από τα συνήθη επίπεδα σημαντικότητας την υπόθεση ότι τα δεδομένα προέρχονται από κάποια από τις παραπάνω κατανομές.

3.2.2 Προσαρμογή σε μετασχηματισμένα δεδομένα

Όπως εργαστήκαμε στην προηγούμενη ενότητα για το ύψος ζημιών, έτσι και εδώ θα μετασχηματίσουμε τα αρχικά δεδομένα του “Total time” και θα ελέγξουμε διαγραμματικά αλλά και μέσω στατιστικών τεστ την υπόθεση ότι οι μετασχηματισμοί αυτών προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή. Τα περιγραφικά χαρακτηριστικά που δίνονται παρακάτω καθώς και τα ιστογράμματα και τα Q-Q plots εξήχθησαν με το στατιστικό πακέτο R.

3.2.2.1 Μετασχηματισμός λογάριθμος

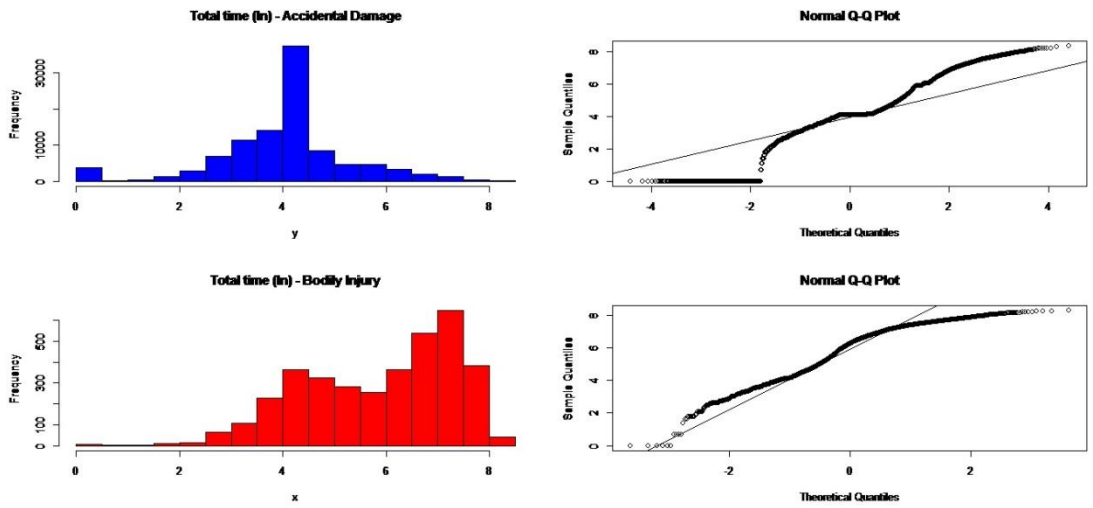
Για το χρόνο αποπληρωμής απαιτήσεων από την ασφαλιστική εταιρία “Total time” παρουσιάζονται στον παρακάτω πίνακα τα περιγραφικά χαρακτηριστικά του λογαρίθμου των αρχικών δεδομένων. Η δεύτερη στήλη αφορά στο χρόνο αποπληρωμής αποζημιώσεων για υλικές ζημιές και η τρίτη για σωματικές βλάβες.

Πίνακας 3.2.1: Περιγραφικά χαρακτηριστικά – μετασχηματισμός λογάριθμος.

Total time	Accidental Damage	Bodily Injury
min	0.00	0.00
25% quantile	3.47	4.65
median (50% quantile)	4.11	6.24
mean	4.00	5.88
75% quantile	4.44	7.15
max	8.34	8.28
variance	1.6913	2.1527
standard deviation	1.3005	1.467
99% quantile	7.2138	8.035
skewness	-0.5104	-0.5759
kurtosis	5.2474	2.5625

Για τους χρόνους αποπληρωμής τα αποτελέσματα του πίνακα δείχνουν πως οι τιμές για το χρόνο αποπληρωμής αποζημιώσεων σωματικών βλαβών είναι γενικά μεγαλύτερες. Όπως είχαμε αναφέρει και στην πρώτη ενότητα, εντύπωση προκαλεί η μέγιστη τιμή που παίρνει η μεταβλητή "Total time" για υλικές ζημίες, η οποία είναι μεγαλύτερη από την αντίστοιχη για σωματικές βλάβες. Προφανώς πρόκειται για κάποια ή κάποιες εξαιρετικές περιπτώσεις. Παρατηρούμε στις τιμές του πίνακα ότι και για τις δύο κατηγορίες ζημιών ο λογάριθμος των αρχικών δεδομένων ακολουθεί κάποια κατανομή με αρνητική ασυμμετρία. Επιπλέον παρατηρούμε πως η κατανομή του χρόνου αποπληρωμής υλικών ζημιών είναι πολύ πιο κυρτή από αυτή των σωματικών βλαβών, μάλιστα ο συντελεστής κύρτωσης είναι ίσος με 5,24 σχεδόν διπλάσιος από των σωματικών βλαβών. Αυτό σημαίνει πως υπάρχει συγκέντρωση παρατηρούμενων χρόνων σε κάποιες τιμές πολύ μεγαλύτερη από των σωματικών βλαβών. Βέβαια, αυτό το επαληθεύουμε και από τη διακύμανση των τιμών όπως παρουσιάζεται παραπάνω που για τους χρόνους αποπληρωμής αποζημιώσεων σωματικών βλαβών είναι σαφώς μεγαλύτερη.

Δίνονται τα ιστογράμματα συχνοτήτων και τα Q-Q plots ώστε να ελέγξουμε οπτικά την προσαρμογή του λογάριθμου των τιμών της μεταβλητής "Total time" για τις δύο κατηγορίες αποζημιώσεων και να εξάγουμε χρήσιμα συμπεράσματα.



Σχήμα 3.2.1: Ιστογράμματα και Normal Q-Q Plots ln (Total time).

Τα γραφήματα που προκύπτουν από τα μετασχηματισμένα δεδομένα μας προσφέρουν, τουλάχιστον οπτικά πολλές πληροφορίες σε σχέση με τα αντίστοιχα για τα αρχικά δεδομένα.

Όσον αφορά στο λογάριθμο του χρόνου αποπληρωμής υλικών ζημιών, το ιστόγραμμα μαρτυρά ότι τα δεδομένα προέρχονται από κατανομή που παρουσιάζει κορυφή γύρω από την τιμή 4 και συγκέντρωση στις πολύ μικρές τιμές και ίσες με το μηδέν. Από το ύψος της κλάσης στην επικρατούσα τιμή μπορούμε να πούμε πως το μεγαλύτερο ποσοστό του λογαρίθμου του χρόνου αποπληρωμής παίρνει τιμές στο διάστημα 4 έως 4,5 δηλαδή εξήντα με ενενήντα ημέρες (2-3 μήνες). Ο μέσος της κατανομής ισούται με 4 και η διάμεσος με 4,11. Παρουσιάζει αρνητική ασυμμετρία και όπως είναι αναμενόμενο πρόκειται για μία λεπτόκυρτη κατανομή με συντελεστή κύρτωσης ίσο με 5,25. Το Q-Q plot δεν επιδεικνύει προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά κάποια κατανομή με πολύ βαριά δεξιά ουρά και γενικά πρόβλημα προσαρμογής στις ακραίες τιμές. Ειδικότερα μπορούμε να δούμε πόσο πολύ αποκλίνουν τα σημεία της κατανομής του λογαρίθμου του χρόνου αποπληρωμής από την ευθεία της κανονικής κατανομής για τιμές κοντά στο μηδέν.

Όσον αφορά στο λογάριθμο του χρόνου αποπληρωμής σωματικών βλαβών, το πρώτο συμπέρασμα στο οποίο μπορούμε να προβούμε παρατηρώντας το ιστόγραμμα συχνοτήτων είναι ότι τα δεδομένα προέρχονται από κατανομή με δύο κορυφές. Μάλιστα, το πρώτο τοπικό μέγιστο είναι στην κλάση 4–4,5 δηλαδή για χρόνους αποπληρωμής 2-3 μήνες και το δεύτερο στην κλάση 7-7,5 δηλαδή για 3-5 έτη. Και εδώ παρατηρείται αρνητική ασυμμετρία όπως φαίνεται και στον πίνακα 3.2.1. Σε αντίθεση με το λογάριθμο του χρόνου αποπληρωμής υλικών ζημιών οι τιμές δεν παρουσιάζουν συγκέντρωση σε ένα σημείο, αλλά η κατανομή είναι πλατύκυρτη με συντελεστή κύρτωσης να ισούται με 2,56. Το Q-Q plot δεν επιδεικνύει προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά σε κάποια κατανομή με βαριά δεξιά ουρά.

3.2.2.2 Μετασχηματισμός τετραγωνική ρίζα

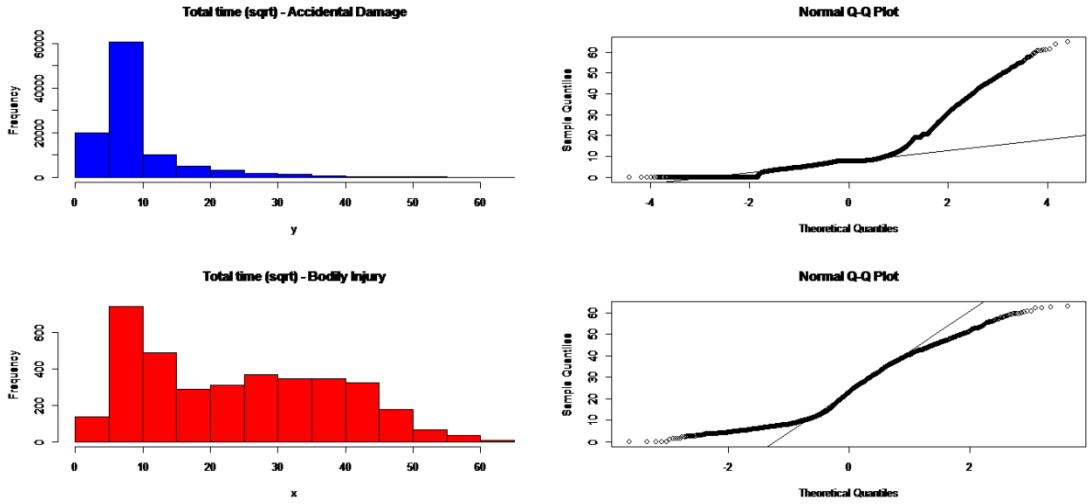
Στον παρακάτω πίνακα δίνονται τα περιγραφικά χαρακτηριστικά για το μετασχηματισμό τετραγωνική ρίζα των τιμών των χρόνων αποπληρωμής υποχρεώσεων της ασφαλιστικής εταιρίας στους ασφαλισμένους.

Πίνακας 3.2.2: Περιγραφικά χαρακτηριστικά – μετασχηματισμός τετραγωνική ρίζα.

Total time	Accidental Damage	Bodily Injury
min	0.00	0.00
25% quantile	5.66	10.25
median (50% quantile)	7.81	22.69
mean	9.00	23.80
75% quantile	9.22	35.68
max	64.77	62.80
variance	44.0258	202.2181
standard deviation	6.6352	14.22
99% quantile	36.8511	55.564
skewness	2.5758	0.3404
kurtosis	11.8639	1.9758

Στα ίδια συμπεράσματα που οδηγηθήκαμε στην περίπτωση του λογαρίθμου, μπορούμε να καταλήξουμε παρατηρώντας τις τιμές του παραπάνω πίνακα. Η διακύμανση των τιμών της τετραγωνικής ρίζας του χρόνου αποπληρωμής για σωματικές βλάβες είναι τετραπλάσια της αντίστοιχης για υλικές ζημιές. Επιπλέον, για τις σωματικές βλάβες παρατηρούμε μια πλατύκυρτη κατανομή με μικρή θετική ασυμμετρία, ενώ αντίθετα για τις σωματικές βλάβες μία λεπτόκυρτη καθώς ο συντελεστής ασυμμετρίας ισούται με 11,86.

Ακολουθούν τα ιστογράμματα και τα Q-Q plots για την τετραγωνική ρίζα του χρόνου αποπληρωμής των αποζημιώσεων για τις δύο κατηγορίες ζημιών.



Σχήμα 3.2.2: Ιστογράμματα και Normal Q-Q Plots sqrt (Total time).

Το ιστόγραμμα για τις υλικές ζημιές επιβεβαιώνει τα συμπεράσματα από τη μελέτη των περιγραφικών χαρακτηριστικών. Παρατηρούμε μεγάλη συγκέντρωση τιμών στη δεύτερη κλάση, δηλαδή για χρόνους αποπληρωμής 50-100 ημέρες. Το Q-Q plot δεν επιδεικνύει προσαρμογή των δεδομένων στην κανονική κατανομή αλλά σε μία κατανομή με πολύ βαριά δεξιά ουρά.

Το ιστόγραμμα για τις σωματικές βλάβες υποδεικνύει ότι τα μετασχηματισμένα δεδομένα του χρόνου αποπληρωμής ακολουθούν κάποια δικόρυφη κατανομή. Ως αποτέλεσμα, που επιβεβαιώνεται και από το Q-Q plot, τα δεδομένα αποκλίνουν από τη γραμμή της κανονικής κατανομής.

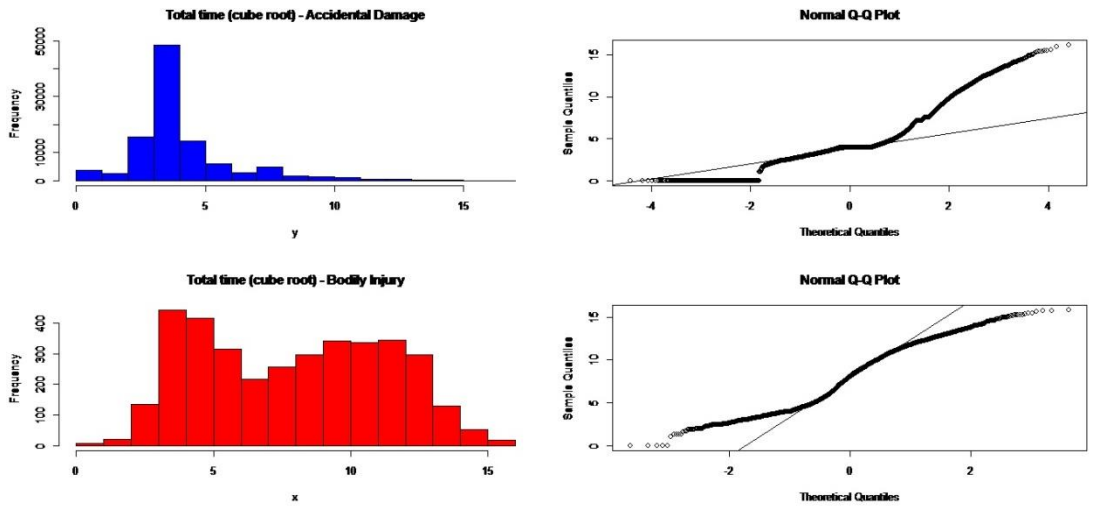
3.2.2.3 Μετασχηματισμός κυβική ρίζα

Στη συνέχεια παρουσιάζουμε στον πίνακα 3.2.3 τα περιγραφικά χαρακτηριστικά για τον μετασχηματισμό κυβική ρίζα των χρόνων αποπληρωμής αποζημιώσεων για τις δύο κατηγορίες ζημιών.

Πίνακας 3.2.3: Περιγραφικά χαρακτηριστικά – μετασχηματισμός κυβική ρίζα.

Total time	Accidental Damage	Bodily Injury
min	0.00	0.00
25% quantile	3.18	4.72
median (50% quantile)	3.94	8.02
mean	4.12	7.91
75% quantile	4.40	10.84
max	16.13	15.80
variance	3.6741	11.3904
standard deviation	1.9168	3.37
99% quantile	11.0739	14.561
skewness	1.429	0.0783
kurtosis	7.2065	1.8152

Τα συμπεράσματα από τα περιγραφικά χαρακτηριστικά του παραπάνω πίνακα είναι κοινά με αυτά που οδηγηθήκαμε από τους μετασχηματισμούς λογάριθμος και τετραγωνική ρίζα του χρόνου αποπληρωμής απαιτήσεων “Total time” στις υποενότητες 3.2.2.1 και 3.2.2.2 αντίστοιχα.



Σχήμα 3.2.3: Ιστογράμματα και Normal Q-Q Plots cube root (Total time).

Όπως αναμέναμε μπορούμε και πάλι να δούμε καθαρά πως οι χρόνοι για αποπληρωμή σωματικών βλαβών προέρχονται από κατανομή με δύο κορυφές και επομένως όχι από την κανονική.

Αντίθετα, από το ιστόγραμμα της κυβικής ρίζας για τις τιμές του "Total time" μπορούμε να πάρουμε μία σημαντική πληροφορία που κανένα ιστόγραμμα από τους προηγούμενους μετασχηματισμούς δεν μας την παρείχε. Αν παρατηρήσουμε προσεκτικά τις κλάσεις του ιστογράμματος θα δούμε ότι η κατανομή παρουσιάζει τουλάχιστον 2 κορυφές. Πιο συγκεκριμένα παρατηρούμε συγκέντρωση τιμών στην πρώτη κλάση, δηλαδή για πολύ μικρές και μηδενικές τιμές. Στην τέταρτη κλάση έχουμε πολύ μεγάλη συγκέντρωση τιμών και επομένως μία ακόμα κορυφή. Τέλος η τρίτη κορυφή φαίνεται στην όγδοη κλάση. Η πληροφορία αυτή είναι πολύ σημαντική για την προσπάθεια διερεύνησης της κατανομής που ακολουθούν οι χρόνοι αποπληρωμής υλικών ζημιών. Η μορφή των Q-Q plots επιβεβαιώνει τα παραπάνω συμπεράσματα καθώς υπάρχει πολύ μεγάλη απόκλιση από την ευθεία της κανονικής κατανομής.

3.2.2.4 Μετασχηματισμός διπλός λογάριθμος

Τέλος, δίνονται τα περιγραφικά χαρακτηριστικά για το μετασχηματισμό του διπλού λογαρίθμου των αρχικών τιμών του χρόνου αποπληρωμής για υλικές ζημιές και σωματικές βλάβες.

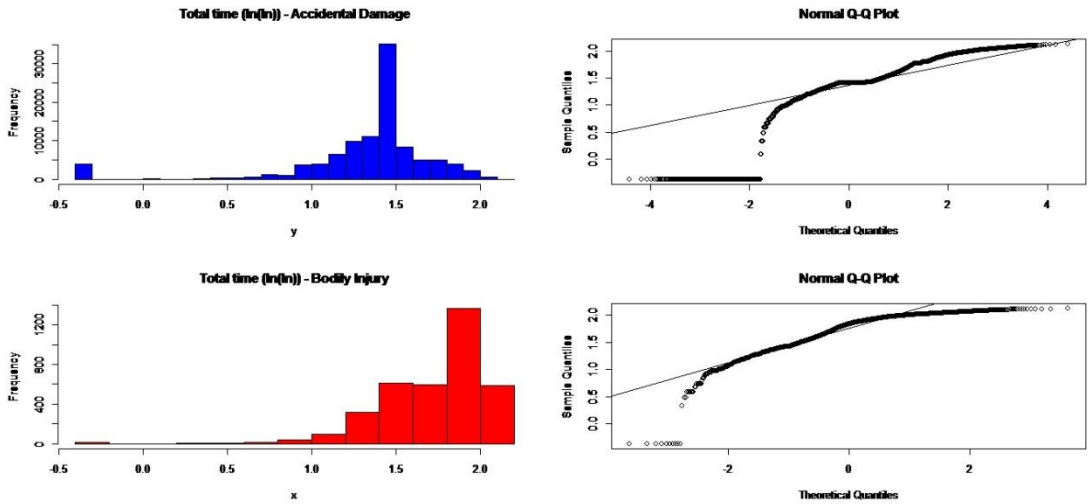
Πίνακας 3.2.4: Περιγραφικά χαρακτηριστικά – μετασχηματισμός διπλός λογάριθμος.

Total time	Accidental Damage	Bodily Injury
min	-0.37	-0.37
25% quantile	1.24	1.54
median (50% quantile)	1.41	1.83
mean	1.33	1.73
75% quantile	1.49	1.97
max	2.12	2.11
variance	0.1781	0.0886
standard deviation	0.4220	0.30
99% quantile	1.9760	2.084
skewness	-2.384	-1.7193
kurtosis	10.4437	9.5302

Παρατηρώντας τα περιγραφικά χαρακτηριστικά του τελευταίου μετασχηματισμού που εφαρμόσαμε στα αρχικά δεδομένα για το χρόνο αποπληρωμής καταλήγουμε για ακόμα μία φορά όπως ήταν αναμενόμενο στα ίδια συμπεράσματα με πριν. Βέβαια, στην περίπτωση του διπλού λογαρίθμου των τιμών παρατηρούμε αρνητικούς συντελεστές ασυμμετρίας και για τις δύο κατηγορίες τιμών.

Η μόνη διαφορά που παρατηρούμε στα περιγραφικά χαρακτηριστικά σε σχέση με τους μετασχηματισμούς που πραγματοποιήθηκαν νωρίτερα είναι ότι για το διπλό λογάριθμο των τιμών της μεταβλητής “Total time” η διακύμανση των τιμών είναι μεγαλύτερη από αυτή των σωματικών βλαβών. Αυτό οφείλεται στην πρόσθεση του δύο για τις αρχικές τιμές που ισούνται με μηδέν και στην πρόσθεση της μονάδας σε αυτές που ισούνται με μονάδα ώστε το αποτέλεσμα της διπλής λογαρίθμισης να μην απειρίζεται. Ως αποτέλεσμα για τους χρόνους αποπληρωμής υλικών ζημιών μετά το μετασχηματισμό παρουσιάζεται μεγάλη συγκέντρωση σε πολύ μικρές τιμές, μηδενικές και αρνητικές κοντά στο μηδέν. Η συγκέντρωση αυτή έχει ως αποτέλεσμα τα μετασχηματισμένα δεδομένα να εμφανίζουν μεγαλύτερη διακύμανση. Η διαπίστωση αυτή έγινε σε σύγκριση με την αντίστοιχη τιμή της διακύμανσης για τις σωματικές βλάβες. Σε κάθε έναν από τους μετασχηματισμούς που επιχειρήσαμε προηγουμένως οι τιμές του χρόνου αποπληρωμής σωματικών βλαβών εμφάνιζαν μεγαλύτερη διασπορά.

Τα ιστογράμματα συχνοτήτων και τα Q-Q plots για το διπλό λογάριθμο των τιμών της μεταβλητής “Total time” για τις δύο κατηγορίες αποζημιώσεων θα μας βοηθήσουν να επαληθεύσουμε ή να απορρίψουμε τους παραπάνω ισχυρισμούς.



Σχήμα 3.2.4: Ιστογράμματα και Normal Q-Q Plots ln (ln (Total time)).

Το ιστόγραμμα συχνοτήτων του μετασχηματισμού του χρόνου αποπληρωμής υλικών ζημιών όπως αναμέναμε παρουσιάζει συγκέντρωση σε αρνητικές και μηδενικές τιμές ως αποτέλεσμα της διπλής λογαρίθμισης. Στο Q-Q plot φαίνεται ξεκάθαρα το πρόβλημα προσαρμογής των δεδομένων της κάτω αριστερής ουράς. Τα δεδομένα δεν προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή.

Στο ιστόγραμμα συχνοτήτων του μετασχηματισμού του χρόνου αποπληρωμής σωματικών βλαβών παρατηρούμε επίσης μία μικρή συγκέντρωση στην πρώτη κλάση. Σε καμία περίπτωση η πρώτη κλάση των σωματικών βλαβών δεν παρουσιάζει τη

συχνότητα που είδαμε στις υλικές ζημιές. Το Q-Q plot δεν επιδεικνύει καλή προσαρμογή των δεδομένων στην κανονική κατανομή, αλλά κάποια κατανομή με βαριά δεξιά ουρά. Και εδώ βλέπουμε πρόβλημα στην προσαρμογή της κάτω αριστερής ουράς, λόγω των αρνητικών τιμών που προέκυψαν από τη διπλή λογαρίθμιση των αρχικών τιμών της μεταβλητής.

Θα επαληθεύσουμε τα παραπάνω συμπεράσματα από τα ιστογράμματα και τα Q-Q plots με τα τέσσερα τεστ κανονικότητας που χρησιμοποιήσαμε για τη μελέτη του ύψους αποζημιώσεων. Τα αποτελέσματα παρουσιάζονται συνοπτικά στον παρακάτω πίνακα.

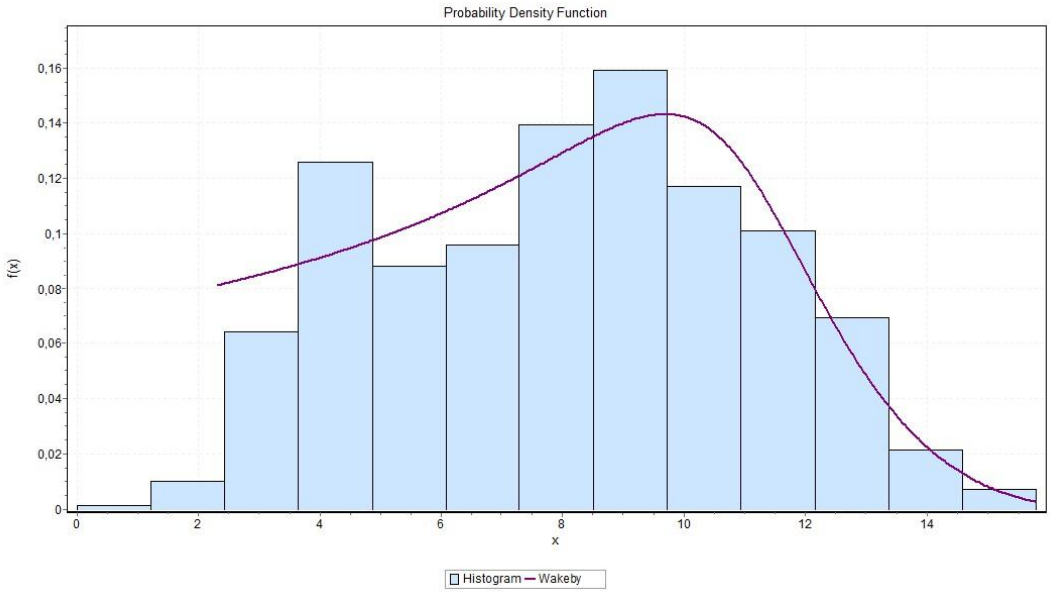
Πίνακας 3.2.5: Αποτελέσματα των τεστ κανονικότητας του "Total time".

Normality Testing	**Ln(Total time)**			
	Accidental Damage		**Bodily Injury**	
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.1333	<2,2 e-16	0.1099	<2,2 e-16
Anderson-Darling	3189.604	<2,2 e-16	72.7313	<2,2 e-16
Cramer-von Mises	615.1138	7.37E-10	12.5524	7.37E-10
Pearson	672956.4	<2,2 e-16	1176.018	<2,2 e-16
	Sqrt(Total time)			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.2425	<2,2 e-16	0.1095	<2,2 e-16
Anderson-Darling	8485.006	<2,2 e-16	67.6243	<2,2 e-16
Cramer-von Mises	1673.545	7.37E-10	10.4176	7.37E-10
Pearson	722107.2	<2,2 e-16	1520.82	<2,2 e-16
	Cube root (Total time)			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.2061	<2,2 e-16	0.0925	<2,2 e-16
Anderson-Darling	5394.714	<2,2 e-16	54.7923	<2,2 e-16
Cramer-von Mises	1076.49	7.37E-10	8.7744	7.37E-10
Pearson	685125.5	<2,2 e-16	1044.418	<2,2 e-16
	Ln(ln(Total time))			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.1807	<2,2 e-16	0.1323	<2,2 e-16
Anderson-Darling	7225.167	<2,2 e-16	109.501	<2,2 e-16
Cramer-von Mises	1287.055	7.37E-10	18.7165	7.37E-10
Pearson	853406	<2,2 e-16	1909.222	<2,2 e-16

Τα αποτελέσματα που προκύπτουν από τους ελέγχους κανονικότητας επιβεβαιώνουν την αρχική μας πρόβλεψη ότι τα δεδομένα δεν προέρχονται από την κανονική κατανομή. Τα p-value των ελέγχων εμφανίζουν την ελάχιστη τιμή που δίνεται από το στατιστικό πακέτο R. Επομένως, απορρίπτουμε για κάθε μετασχηματισμό τη μηδενική υπόθεση, δηλαδή την υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή σε επίπεδο σημαντικότητας 5%.

Για τους τέσσερις παραπάνω μετασχηματισμούς της μεταβλητής “Total time”, δοκιμάστηκε η προσαρμογή τους και στις θεωρητικές συνεχείς κατανομές που είναι διαθέσιμες από το πρόγραμμα “EasyFit”. Με εξαίρεση το μετασχηματισμό κυβική ρίζα του “Total time” σωματικών βλαβών, σύμφωνα με τα τρία τεστ καλής προσαρμογής που εκτελούνται αυτόματα από το “EasyFit” (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test), απορρίπτουμε την υπόθεση ότι τα μετασχηματισμένα δεδομένα προέρχονται από κάποια από τις κατανομές αυτές.

Για το μετασχηματισμό κυβική ρίζα του χρόνου αποπληρωμής των αποζημιώσεων σωματικών βλαβών, δίνεται το ιστόγραμμα συχνοτήτων όπως εξήχθη από το “EasyFit”.



Σχήμα 3.2.5: Ιστόγραμμα κυβικής ρίζας “Total time” σωματικών βλαβών.

Σύμφωνα με το τεστ καλής προσαρμογής Kolmogorov-Smirnov δεν απορρίπτουμε την υπόθεση ότι τα μετασχηματισμένα δεδομένα του δείγματος προέρχονται από την κατανομή Wakeby σε κάποια από τα επίπεδα σημαντικότητας που χρησιμοποιεί το “EasyFit”. Τα άλλα δύο τεστ καλής προσαρμογής του πακέτου απορρίπτουν την παραπάνω υπόθεση. Η μωβ γραμμή που εμφανίζεται στο ιστόγραμμα είναι η καμπύλη της συνάρτησης πυκνότητας πιθανότητας της κατανομής Wakeby.

Τα αποτελέσματα του ελέγχου Kolmogorov-Smirnov όπως εξήχθησαν από το πρόγραμμα εμφανίζονται στον παρακάτω πίνακα.

Πίνακας 3.2.6: Αποτελέσματα του τεστ προσαρμογής του "EasyFit" για τη Wakeby.

Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	4784 0,02151 0,02354 1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,01551	0,01768	0,01963	0,02195	0,02355
Reject?	Yes	Yes	Yes	No	No

Όπως αναφέρουμε και παραπάνω σύμφωνα με το Kolmogorov-Smirnov gof test δεν απορρίπτουμε την υπόθεση ότι η κυβική ρίζα του "Total time" σωματικών βλαβών ακολουθεί κατανομή Wakeby σε επίπεδο σημαντικότητας 2% καθώς το p-value του ελέγχου ισούται με 2,354% και είναι μεγαλύτερο από το επιλεγμένο επίπεδο σημαντικότητας. Για κάθε επίπεδο σημαντικότητας μεγαλύτερο του 2% η υπόθεση ότι τα δεδομένα προέρχονται από την κατανομή Wakeby απορρίπτεται.

Δίνονται, τέλος, οι εκτιμήσεις των πέντε παραμέτρων της κατανομής, όπως υπολογίζονται από το πρόγραμμα.

58	Wakeby	α=12,699 β=2,2405 γ=2,237 δ=-0,22517 ξ=2,3337

Η κατανομή Wakeby είναι μία κατανομή πιθανότητας πέντε παραμέτρων που ορίζεται από τη συνάρτηση:

$$x(U) = \xi + \frac{\alpha}{\beta}\left(1-(1-U)^{\beta}\right) - \frac{\gamma}{\delta}\left(1-(1-U)^{-\delta}\right)$$

όπου U είναι μία ομοιόμορφη τυχαία μεταβλητή στο μοναδιαίο διάστημα. Η παραπάνω συνάρτηση ορίζει τη συνάρτηση ποσοστιαίων σημείων (ποσοστημορίων – quantile function) για την κατανομή Wakeby. Οι παράμετροι β, γ και δ είναι παράμετροι σχήματος. Οι παράμετροι ξ και α είναι παράμετροι θέσης.

Για τις παραμέτρους της κατανομής ισχύουν οι παρακάτω περιορισμοί:

- $\beta + \delta \geq 0$
- Είτε $\beta + \delta > 0$ είτε $\beta = \gamma = \delta = 0$
- Αν $\gamma > 0$, τότε $\delta > 0$
- $\gamma \geq 0$

- $\alpha + \gamma \geq 0$

Το πεδίο ορισμού της κατανομής είναι:

- $\xi \leq x < +\infty$, αν $\delta \geq 0$ και $\gamma > 0$
- $\xi \leq x \leq \alpha/\beta - \gamma/\delta$, αν $\delta < 0$ ή $\gamma = 0$

Η αθροιστική συνάρτηση κατανομής υπολογίζεται με την αριθμητική αντιστροφή της συνάρτησης των ποσοστημορίων όπως δόθηκε παραπάνω. Η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της κατανομής δίνεται μέσω της επόμενης σχέσης (βλ. σελ. 46 Johnson, Kotz, & Balakrishnan, 1994):

$$f(x) = \frac{\left(1 - F(x)\right)^{\delta+1}}{\alpha t + \gamma}$$

όπου F η αθροιστική συνάρτηση πυκνότητας πιθανότητας (α.σ.π.π.) της κατανομής και $t = \left(1 - F(x)\right)^{\beta+\delta}$.

3.3 Διερεύνηση της κατανομής του χρόνου καθυστέρησης αναγγελίας της ζημιάς στην ασφαλιστική εταιρία

3.3.1 Προσαρμογή στα αρχικά δεδομένα

Μια πρώτη προσπάθεια διερεύνησης της κατανομής της μεταβλητής “Delay” επιχειρήθηκε στην προηγούμενη ενότητα. Με τη βοήθεια διαγραμμάτων (ιστογράμματα συχνοτήτων, Normal Q-Q Plot διαγράμματα) και με τη διενέργεια των τεσσάρων τεστ κανονικότητας (Lilliefors, Anderson-Darling, Cramer-von Mises, Pearson normality test) ελέγχθηκε αν τα δεδομένα προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή. Τα αποτελέσματα των ελέγχων ήταν πως απορρίπτουμε την υπόθεση ότι τα δεδομένα προέρχονται από την κανονική κατανομή.

Στην παρούσα υποενότητα με τη βοήθεια του προγράμματος “EasyFit” προχωρούμε στη διερεύνηση της κατανομής της μεταβλητής “Delay” για τις δύο κατηγορίες ζημιών της ασφαλιστικής εταιρίας, δηλαδή για αποζημιώσεις υλικών ζημιών και σωματικών βλαβών, δοκιμάζοντας την προσαρμογή των δεδομένων στις θεωρητικές κατανομές που προσφέρει το πακέτο.

Οι δοκιμές που έγιναν αφορούν στα αρχικά δεδομένα. Πιο αναλυτικά ελέγχθηκε η προσαρμογή των αρχικών τιμών της μεταβλητής “Delay” στις συνεχείς και διακριτές κατανομές που περιλαμβάνει το πακέτο “EasyFit”. Σύμφωνα με τα τρία τεστ καλής

προσαρμογής που χρησιμοποιεί το πακέτο (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test) απορρίπτουμε σε κάθε σύνηθες επίπεδο σημαντικότητας την υπόθεση ότι τα δεδομένα προέρχονται από κάποια από τις παραπάνω κατανομές.

3.3.2 Προσαρμογή σε μετασχηματισμένα δεδομένα

Όπως εργαστήκαμε στις προηγούμενες υποενότητες για το ύψος ζημιών και για το χρόνο αποπληρωμής αποζημιώσεων, έτσι και εδώ θα μετασχηματίσουμε τα αρχικά δεδομένα του "Delay" και θα ελέγξουμε διαγραμματικά αλλά και μέσω στατιστικών τεστ την υπόθεση ότι οι μετασχηματισμοί αυτών προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή. Τα περιγραφικά χαρακτηριστικά που δίνονται παρακάτω καθώς και τα ιστογράμματα και τα Q-Q plots εξήχθησαν με το στατιστικό πακέτο R.

3.3.2.1 Μετασχηματισμός λογάριθμος

Για το χρόνο καθυστέρησης αναγγελίας της ζημιάς στην ασφαλιστική εταιρία "Delay" παρουσιάζονται στον παρακάτω πίνακα τα περιγραφικά χαρακτηριστικά του λογαρίθμου των αρχικών δεδομένων. Η δεύτερη στήλη αφορά στο χρόνο καθυστέρησης αναγγελίας υλικών ζημιών και η τρίτη σωματικών βλαβών.

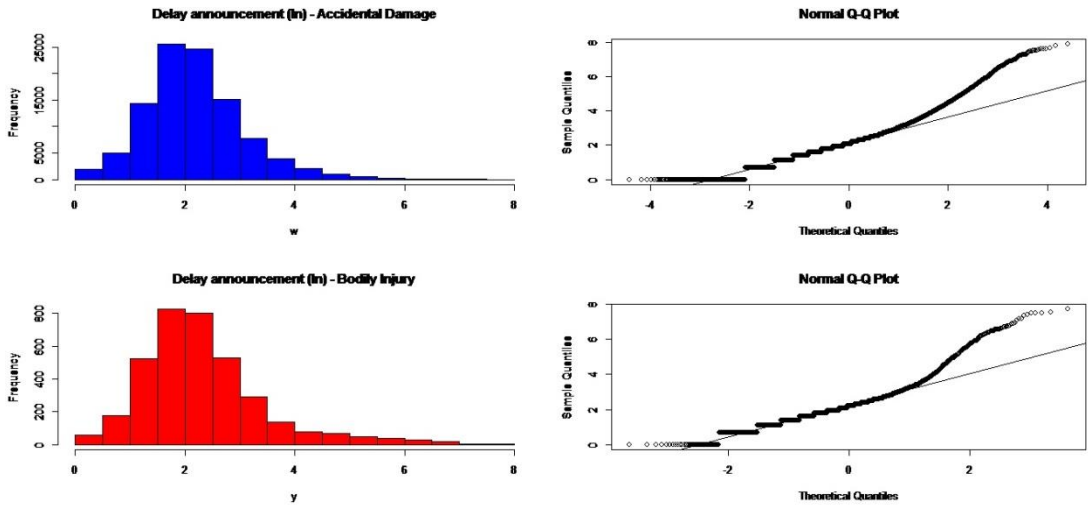
Πίνακας 3.3.1: Περιγραφικά χαρακτηριστικά – μετασχηματισμός λογάριθμος.

Delay	Accidental Damage	Bodily Injury
min	0.00	0.00
25% quantile	1.61	1.61
median (50% quantile)	2.08	2.20
mean	2.19	2.34
75% quantile	2.64	2.83
max	7.88	7.72
variance	0.8871	1.354
standard deviation	0.9419	1.164
99% quantile	5.0434	6.402
skewness	0.7378	1.2603
kurtosis	4.9063	5.6497

Όπως είχαμε αναφέρει και στην πρώτη προσπάθεια σύγκρισης των περιγραφικών χαρακτηριστικών του χρόνου καθυστέρησης αναγγελίας της ζημιάς στην υποενότητα 2.2, τα αποτελέσματα μεταξύ των δύο κατηγοριών ζημιών δεν διαφέρουν τόσο όσο των μεταβλητών "Total payment" και "Total time". Παρατηρώντας τον παραπάνω πίνακα, θα διαπιστώσουμε ότι η μέση τιμή της κατανομής για υλικές ζημιές δεν διαφέρει σημαντικά από αυτή για σωματικές βλάβες. Επίσης, για το λογάριθμο των δεδομένων του "Delay" για υλικές ζημιές, η διακύμανση των τιμών, οι συντελεστές

ασυμμετρίας και κύρτωσης παίρνουν τιμές που δεν διαφοροποιούνται σημαντικά από τις αντίστοιχες για σωματικές βλάβες. Γενικά αναμένουμε οι κατανομές που ακολουθούν τα δεδομένα για τις δύο κατηγορίες ζημιών να "μοιάζουν".

Τα ιστογράμματα συχνοτήτων θα μας βοηθήσουν να διαπιστώσουμε τουλάχιστον οπτικά αν μπορούμε να θεωρήσουμε ότι είναι πιθανόν ο λογάριθμος της καθυστέρησης αναγγελίας για υλικές ζημιές να ακολουθεί την ίδια κατανομή με αυτή των σωματικών βλαβών.



Σχήμα 3.3.1: Ιστογράμματα και Normal Q-Q Plots ln (Delay).

Με μία πρώτη ματιά ο ισχυρισμός μας φαίνεται να ευσταθεί, δηλαδή οι χρόνοι καθυστέρησης να προέρχονται από την ίδια κατανομή. Μία όμως πιο προσεκτική ματιά, ιδίως στα Q-Q plots μας δείχνει ότι στην περίπτωση των χρόνων καθυστέρησης αναγγελίας για υλικές ζημίες έχουμε μεγαλύτερη συγκέντρωση στην πρώτη κλάση, δηλαδή για τιμές πολύ κοντά στο μηδέν ή μηδενικές. Επίσης, για τις τιμές του λογαρίθμου του "Delay" των σωματικών βλαβών παρατηρούμε πιο βαριά δεξιά ουρά. Όπως φαίνεται στα ιστογράμματα στην τέταρτη κλάση εμφανίζεται η μεγαλύτερη συγκέντρωση και για τις δύο κατηγορίες ζημιών, δηλαδή για χρόνους καθυστέρησης ίσους με 8 ημέρες περίπου.

Μία παρατήρηση που πρέπει να κάνουμε σε αυτό το σημείο είναι ότι, ενώ στη μελέτη των περιγραφικών χαρακτηριστικών για τις αρχικές τιμές του "Delay" η κατανομή των δεδομένων για υλικές ζημιές εμφανίζει μεγαλύτερη θετική ασυμμετρία και κυρτότητα σε σχέση με την κατανομή των σωματικών βλαβών, εδώ συμβαίνει το αντίθετο. Σύμφωνα με τις τιμές του παραπάνω πίνακα, ο λογάριθμος του "Delay" σωματικών βλαβών ακολουθεί πιο ασύμμετρη και κυρτή κατανομή από το λογάριθμο υλικών ζημιών. Αυτό οφείλεται στη λογαρίθμιση των χρόνων καθυστέρησης.

Όσον αφορά την υπόθεση ότι τα δεδομένα για τους χρόνους καθυστέρησης προέρχονται από την κανονική κατανομή, οι αποκλίσεις των σημείων από την ευθεία της κανονικής κατανομής και στα δύο Q-Q plots συνηγορούν στην απόρριψη της.

3.3.2.2 Μετασχηματισμός τετραγωνική ρίζα

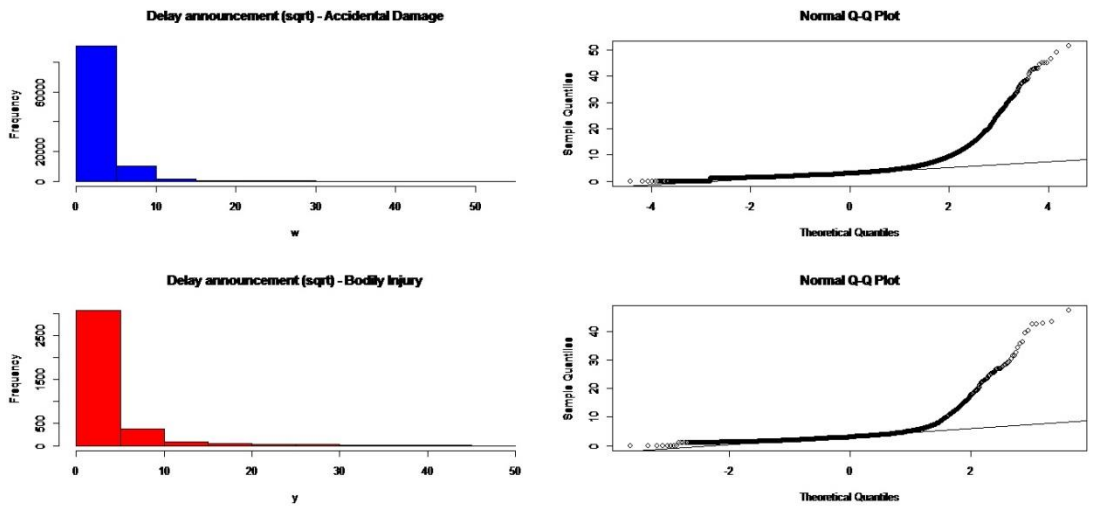
Συνεχίζουμε την ανάλυση μας με το δεύτερο μετασχηματισμό της μεταβλητής "Delay", δηλαδή την τετραγωνική ρίζα των αρχικών τιμών. Παραθέτουμε τον πίνακα με τα περιγραφικά χαρακτηριστικά.

Πίνακας 3.3.2: Περιγραφικά χαρακτηριστικά – μετασχηματισμός τετραγωνική ρίζα.

Delay	Accidental Damage	Bodily Injury
min	0.00	0.00
25% quantile	2.24	2.24
median (50% quantile)	2.83	3.00
mean	3.39	4.02
75% quantile	3.74	4.12
max	51.40	47.34
variance	5.3115	16.717
standard deviation	2.3047	4.089
99% quantile	12.4499	24.559
skewness	5.2873	4.7072
kurtosis	54.7653	31.939

Τα αποτελέσματα για το μετασχηματισμό της τετραγωνικής ρίζας των αρχικών δεδομένων, μας μαρτυρούν και πάλι πως κατά μέσο όρο μία ζημιά που αφορά σωματική βλάβη "αργεί" να ανακοινωθεί στην ασφαλιστική εταιρία λίγο περισσότερο απ' ότι στην περίπτωση μιας υλικής ζημίας. Επιπλέον τα αποτελέσματα του πίνακα του μετασχηματισμού τετραγωνική ρίζα επιβεβαιώνουν τα αποτελέσματα που προέκυψαν από την ανάλυση των αρχικών τιμών, δηλαδή ότι η κατανομή για υλικές ζημιές είναι πιο κυρτή και ασύμμετρη από αυτή για σωματικές βλάβες.

Δίνονται τα ιστογράμματα και τα Q-Q plots των μετασχηματισμένων τιμών για υλικές ζημιές και σωματικές βλάβες.



Σχήμα 3.3.2: Ιστογράμματα και Normal Q-Q Plots sqrt (Delay).

Η μορφή των παραπάνω ιστογραμμάτων δεν μπορεί να μας δώσει πρόσθετη πληροφορία για την κατανομή που ακολουθούν τα μετασχηματισμένα δεδομένα. Και στην περίπτωση της τετραγωνικής ρίζας βλέπουμε να μοιάζουν οπτικά οι κατανομές των χρόνων αναγγελίας των δύο κατηγοριών ζημιών. Η πρώτη κλάση συγκεντρώνει το μεγαλύτερο μέρος των παρατηρούμενων τιμών. Τα Q-Q plots δεν επιδεικνύουν προσαρμογή στην κανονική κατανομή, αλλά κάποια κατανομή με πολύ βαριά δεξιά ουρά.

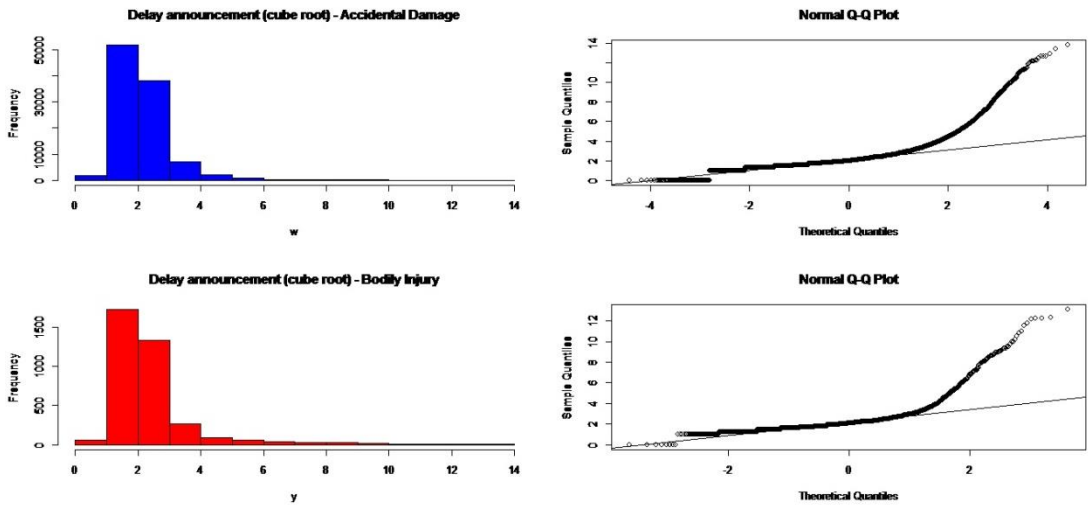
3.3.2.3 Μετασχηματισμός κυβική ρίζα

Προχωρούμε στο μετασχηματισμό κυβική ρίζα των τιμών της μεταβλητής “Delay”. Όπως και προηγουμένως δίνουμε τον πίνακα με τα περιγραφικά χαρακτηριστικά που προκύπτουν από το μετασχηματισμό των αρχικών τιμών.

Πίνακας 3.3.3: Περιγραφικά χαρακτηριστικά – μετασχηματισμός κυβική ρίζα.

Delay	Accidental Damage	Bodily Injury
min	0.00	0.00
25% quantile	1.71	1.71
median (50% quantile)	2.00	2.08
mean	2.18	2.38
75% quantile	2.41	2.57
max	13.82	13.09
variance	0.7037	1.684
standard deviation	0.8389	1.298
99% quantile	5.3717	8.449
skewness	2.9853	3.3587
kurtosis	21.1294	18.121

Συγκρίνοντας με τα αποτελέσματα που προέκυψαν από τον προηγούμενο μετασχηματισμό, δηλαδή την τετραγωνική ρίζα, δεν μπορούμε να πούμε πως υπάρχει διαφοροποίηση στα συμπεράσματά μας εξετάζοντας τον παραπάνω πίνακα.



Σχήμα 3.3.3: Ιστογράμματα και Normal Q-Q Plots cube root (Delay).

Το ίδιο ισχύει και για τα ιστογράμματα και τα Q-Q plots. Η μορφή τους δεν μας δίνει κάποια επιπλέον πληροφορία για την κατανομή που ακολουθούν τα δεδομένα. Τα

δεδομένα δεν προέρχονται από κανονική κατανομή, αλλά από κάποια κατανομή με βαριά δεξιά ουρά.

3.3.2.4 Μετασχηματισμός διπλός λογάριθμος

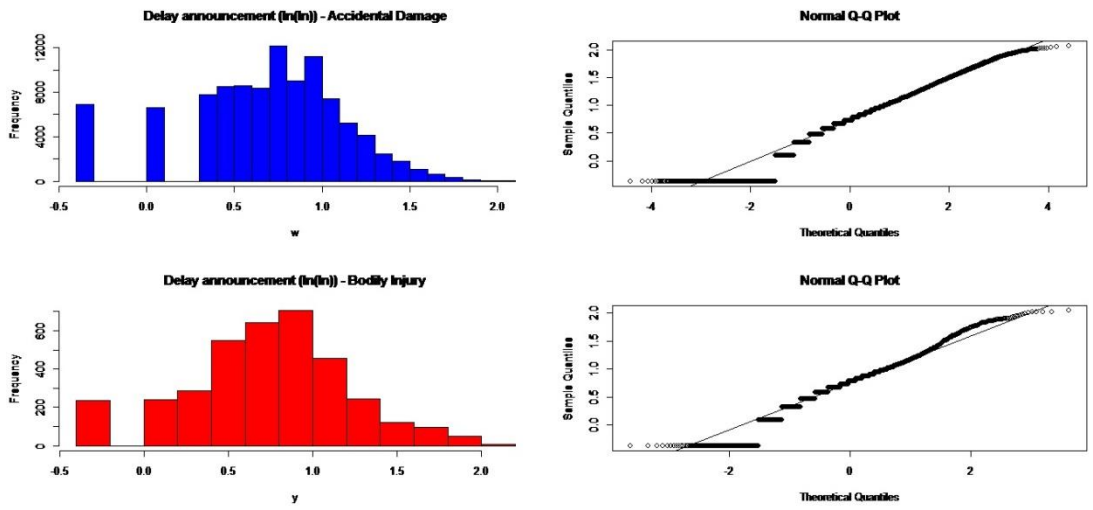
Τέλος, δίνονται τα περιγραφικά χαρακτηριστικά για το μετασχηματισμό λογάριθμος του λογαρίθμου των αρχικών τιμών του χρόνου καθυστέρησης αναγγελίας της ζημίας στην ασφαλιστική εταιρία για υλικές ζημιές και σωματικές βλάβες.

Πίνακας 3.3.4: Περιγραφικά χαρακτηριστικά – μετασχηματισμός διπλός λογάριθμος.

Delay	Accidental Damage	Bodily Injury
min	-0.37	-0.37
25% quantile	0.48	0.48
median (50% quantile)	0.73	0.79
mean	0.70	0.74
75% quantile	0.97	1.04
max	2.06	2.04
variance	0.1971	0.234
standard deviation	0.4439	0.484
99% quantile	1.6181	1.857
skewness	-0.5685	-0.2819
kurtosis	3.4259	3.304

Οι αρνητικές τιμές που παίρνει η ελάχιστη τιμή των μετασχηματισμένων δεδομένων οφείλεται στην διπλή λογαρίθμιση. Όσον αφορά στα άλλα χαρακτηριστικά δεν μας παρέχουν κάποια πρόσθετη πληροφορία.

Προχωρούμε στα ιστογράμματα συχνοτήτων και τα normal Q-Q plots για τις δύο κατηγορίες ζημιών με σκοπό την εύρεση κάποιας πληροφορίας που θα μας βοηθήσει στον προσδιορισμό της κατανομής που ακολουθούν τα δεδομένα.



Σχήμα 3.3.4: Ιστογράμματα και Normal Q-Q Plots ln (ln (Delay)).

Παρατηρώντας τα δύο ιστογράμματα διακρίνουμε αμέσως το αποτέλεσμα της διπλής λογαρίθμισης, δηλαδή συγκέντρωση σε μικρές τιμές. Τόσο στην περίπτωση των υλικών ζημιών αλλά και των σωματικών βλαβών παρατηρούμε κλάσεις που εμφανίζουν συγκέντρωση και είναι αποκομμένες από τις υπόλοιπες. Ειδικότερα για τις υλικές ζημιές η πρώτη κλάση εμφανίζεται με κέντρο στην τιμή -0,37 και είναι αποτέλεσμα της διπλής λογαρίθμισης των αρχικών τιμών που ισούνται με 2. Αντίστοιχα για άλλες μικρές τιμές στα αρχικά δεδομένα πριν το μετασχηματισμό εμφανίζεται η δεύτερη κλάση. Μία πληροφορία που παίρνουμε από το ιστόγραμμα είναι οι 2 κορυφές που εμφανίζει πλέον με το μετασχηματισμό του διπλού λογαρίθμου η κατανομή του “Delay” για υλικές ζημίες.

Όσον αφορά στο ιστόγραμμα των σωματικών βλαβών παρατηρούμε και εδώ την πρώτη κλάση με αρνητικές τιμές ως αποτέλεσμα της διπλής λογαρίθμισης. Στην περίπτωση των σωματικών βλαβών η κατανομή παρουσιάζει μία κορυφή στην τιμή 0,8 δηλαδή για χρόνους καθυστέρησης ίσους με 8-10 ημέρες.

Τα Q-Q plots δεν επιδεικνύουν προσαρμογή των τιμών στην κανονική κατανομή. Για τις δύο κατηγορίες ζημιών εμφανίζεται έντονο πρόβλημα προσαρμογής στην κάτω αριστερή ουρά, δηλαδή για αρνητικές και μηδενικές τιμές.

Θα επαληθεύσουμε τα παραπάνω συμπεράσματα από τα ιστογράμματα και τα Q-Q plots με τα τέσσερα τεστ κανονικότητας που χρησιμοποιήσαμε στις προηγούμενες υποενότητες. Τα αποτελέσματα παρουσιάζονται συνοπτικά στον παρακάτω πίνακα.

Πίνακας 3.3.5: Αποτελέσματα των τεστ κανονικότητας του "Delay".

Normality Testing	**Ln(Delay)**			
	Accidental Damage		**Bodily Injury**	
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.0742	<2,2 e-16	0.1124	<2,2 e-16
Anderson-Darling	818.4106	<2,2 e-16	74.9099	<2,2 e-16
Cramer-von Mises	138.5614	7.37E-10	12.2688	7.37E-10
Pearson	924813.3	<2,2 e-16	5678.419	<2,2 e-16
	Sqrt(Delay)			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.192	<2,2 e-16	0.2585	<2,2 e-16
Anderson-Darling	7853.328	<2,2 e-16	501.2107	<2,2 e-16
Cramer-von Mises	1421.557	7.37E-10	96.0875	7.37E-10
Pearson	922392.1	<2,2 e-16	6060.508	<2,2 e-16
	Cube root (Delay)			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.148	<2,2 e-16	0.2073	<2,2 e-16
Anderson-Darling	4282.976	<2,2 e-16	317.5708	<2,2 e-16
Cramer-von Mises	748.6976	7.37E-10	57.7296	7.37E-10
Pearson	925465.9	<2,2 e-16	5630.326	<2,2 e-16
	Ln(ln(Delay))			
	D/A/W/P	**p-value**	**D/A/W/P**	**p-value**
Lilliefors	0.1085	<2,2 e-16	0.0885	<2,2 e-16
Anderson-Darling	1226.904	<2,2 e-16	27.967	<2,2 e-16
Cramer-von Mises	169.339	7.37E-10	3.7783	7.37E-10
Pearson	956735.6	<2,2 e-16	5913.108	<2,2 e-16

Τα αποτελέσματα που προκύπτουν από τους ελέγχους κανονικότητας επιβεβαιώνουν την αρχική μας πρόβλεψη ότι τα δεδομένα δεν προέρχονται από την κανονική κατανομή. Τα p-value των ελέγχων εμφανίζουν την ελάχιστη τιμή που δίνεται από το στατιστικό πακέτο R.

Για τους τέσσερις παραπάνω μετασχηματισμούς της μεταβλητής "Delay", δοκιμάστηκε η προσαρμογή τους και στις θεωρητικές συνεχείς κατανομές που είναι διαθέσιμες από το πρόγραμμα "EasyFit". Σύμφωνα με τα τρία τεστ καλής προσαρμογής που εκτελούνται αυτόματα από το "EasyFit" (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test), απορρίπτουμε την υπόθεση ότι τα μετασχηματισμένα δεδομένα του "Delay" υλικών ζημιών και σωματικών βλαβών προέρχονται από κάποια από τις κατανομές αυτές.

3.4 Box-Cox Power Transformation

Για την στατιστική ανάλυση των δεδομένων και την εξαγωγή σωστών συμπερασμάτων είναι συνήθως πολύ χρήσιμη η κανονικότητα στα δεδομένα (τα προς εξέταση δεδομένα να ακολουθούν την κανονική κατανομή). Όταν τα δεδομένα δεν είναι κανονικά κατανεμημένα, η αιτία της μη κανονικότητας θα πρέπει να προσδιορίζεται και να ληφθούν τα κατάλληλα διορθωτικά μέτρα πριν ξεκινήσει η στατιστική ανάλυση, η διενέργεια υπολογισμών και η εξαγωγή τελικών συμπερασμάτων.

Ο μετασχηματισμός δεδομένων και ιδιαίτερα ο Box-Cox Power Transformation (Box & Cox, 1964) είναι ένας από αυτούς τους μετασχηματισμούς που μπορούν να οδηγήσουν στο επιθυμητό αποτέλεσμα, δηλαδή στο να μετασχηματίσουμε τα δεδομένα σε κανονικά.

Το να μετασχηματίζεις δεδομένα πρακτικά σημαίνει να εφαρμόζεις την ίδια μαθηματική πράξη σε κάθε μία τιμή των αρχικών δεδομένων. Οι πιο συχνά εφαρμόσιμοι μετασχηματισμοί είναι ο λογάριθμος, η τετραγωνική ρίζα, η κυβική ρίζα, ο διπλός λογάριθμος. Η πρακτική αυτή ακολουθήθηκε ήδη στις προηγούμενες υποενότητες για τις μεταβλητές του δείγματος. Στην υποενότητα αυτή μετασχηματίζουμε τα δεδομένα σύμφωνα με τη γενικότερη οικογένεια μετασχηματισμών, Box-Cox Power Transformation.

Οι στατιστικοί George Box και David Cox ανέπτυξαν μία διαδικασία για τον προσδιορισμό του κατάλληλου δείκτη-συντελεστή λ (lambda) που μπορεί να χρησιμοποιηθεί για την μετατροπή των δεδομένων σε “κανονικά”. Η τιμή του συντελεστή λ μπορεί να είναι είτε θετική, είτε αρνητική, είτε μηδέν στο διάστημα [-5,5]. Η τιμή του λ δείχνει τη δύναμη στην οποία θα πρέπει να υψωθούν όλα τα δεδομένα για να μετασχηματιστούν.

Έστω ότι y είναι τα αρχικά δεδομένα που θέλουν μετασχηματισμό. Η συνάρτηση του Box-Cox μετασχηματισμού είναι:

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

Παρατηρούμε ότι για λ=0, ο μετασχηματισμός δεν είναι y^0 (γιατί αυτό θα ήταν 1 για κάθε τιμή-αξία), αλλά αντίθετα είναι ο λογάριθμος του y.

Η διαδικασία Box-Cox αναζητεί το καλύτερο λ στο πεδίο τιμών [-5,5]. Η τιμή του λ που επιλέγεται είναι αυτή που μεγιστοποιεί το λογάριθμο της συνάρτησης πιθανοφάνειας:

$$f(y,\lambda) = -\frac{n}{2}\ln\left[\sum_{i=1}^{n}\frac{\left(\left(y_i(\lambda)-\bar{y}(\lambda)\right)\right)^2}{n}\right] + (\lambda-1)\sum_{i=1}^{n}\ln(y_i),$$

όπου $\bar{y}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} y_i(\lambda)$ ο αριθμητικός μέσος των μετασχηματισμένων δεδομένων.

Ο παρακάτω πίνακας περιέχει τους πιο κοινούς μετασχηματισμούς, όπου y(λ) είναι ο μετασχηματισμός των αρχικών δεδομένων y.

Πίνακας 3.4.1: Μετασχηματισμοί Box – Cox Power Transformation.

Box-Cox Transformations	
λ	**Y(λ)**
-2	$Y^{-2}=1/Y^2$
-1	$Y^{-1}=1/Y^1$
-0.5	$Y^{-0,5}$=1/(Sqrt(Y))
0	Log(Y)
0.5	$Y^{0,5}$=Sqrt(Y)
1	$Y^1=Y$
2	Y^2

Πρέπει να σημειωθεί ότι η εφαρμογή του μετασχηματισμού Box-Cox δεν εγγυάται την κανονικότητα στα μετασχηματισμένα δεδομένα. Αυτό οφείλεται στο γεγονός ότι στην πραγματικότητα η μέθοδος "ψάχνει" για την τιμή του λ που θα έχει ως αποτέλεσμα τη μικρότερη δυνατή τυπική απόκλιση στα μετασχηματισμένα δεδομένα.

Ο μετασχηματισμός Box-Cox λειτουργεί μόνο εάν όλα τα δεδομένα έχουν θετικές τιμές. Αυτό πάντως επιτυγχάνεται εύκολα με την πρόσθεση μιας σταθεράς (c) σε όλα τα αρχικά δεδομένα πριν μετασχηματιστούν.

Με τη βοήθεια της R και την εγκατάσταση του πακέτου forecast, προσδιορίστηκαν οι κατάλληλοι συντελεστές λ (lambda) για τον μετασχηματισμό Box-Cox των δεδομένων για τις δύο κατηγορίες αποζημιώσεων. Τα αποτελέσματα παραθέτονται συγκεντρωτικά στον παρακάτω πίνακα.

Πίνακας 3.4.2: Αποτελέσματα για τους συντελεστές λ των μεταβλητών.

lambda (λ)	**Accidental Damage**	**Bodily injury**
Total payment	-0.21	-0.12
Total time	0.07	0.21
Delay	-0.3	-0.39

3.4.1 Εφαρμογή του Box-Cox Power Transformation στα δεδομένα υλικών ζημιών

Για την κατηγορία υλικές ζημιές προσδιορίστηκαν οι συντελεστές λ για τις μεταβλητές “Total payment”, “Total time” και “Delay”. Από τον υπολογισμό προέκυψε τιμή του συντελεστή ίση με -0,21 για το ύψος αποζημιώσεων, 0,07 για το χρόνο αποπληρωμής και -0,3 για το χρόνο καθυστέρησης αναγγελίας της ζημιάς.

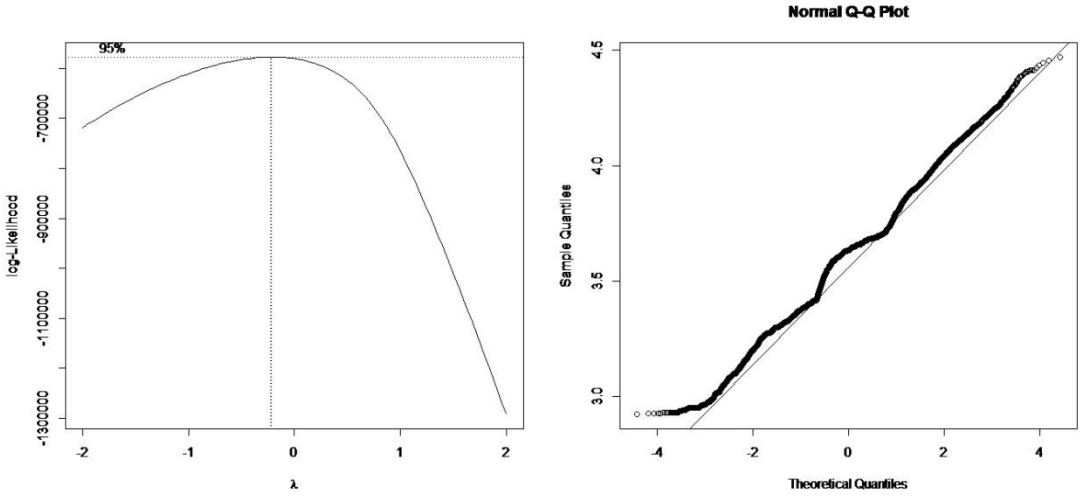
Τα δεδομένα μετασχηματίστηκαν σύμφωνα με τις παραπάνω τιμές του συντελεστή λ. Συνοπτικά δίνονται στον πίνακα που ακολουθεί περιγραφικά χαρακτηριστικά για τα μετασχηματισμένα δεδομένα.

Πίνακας 3.4.3: Περιγραφικά χαρακτηριστικά μετασχηματισμών των μεταβλητών υλικών ζημιών.

Accidental damage	Total payment	Total time	Delay
min	2.92	0	0
25% quantile	3.42	3.96	1.39
median (50% quantile)	3.63	4.79	1.61
mean	3.61	4.72	1.63
75% quantile	3.70	5.23	1.85
max	4.47	11.33	3.02
variance	0.0436	2.8165	0.1560
standard deviation	0.2088	1.6782	0.3949
99% quantile	4.1097	9.3859	2.6006
skewness	-0.0027	0.0303	-0.0456
kurtosis	0.1322	1.9310	0.7215

Επιπλέον, δίνονται τα διαγράμματα για την παράμετρο λ, καθώς και τα Normal Q-Q Plot διαγράμματα για τα μετασχηματισμένα δεδομένα.

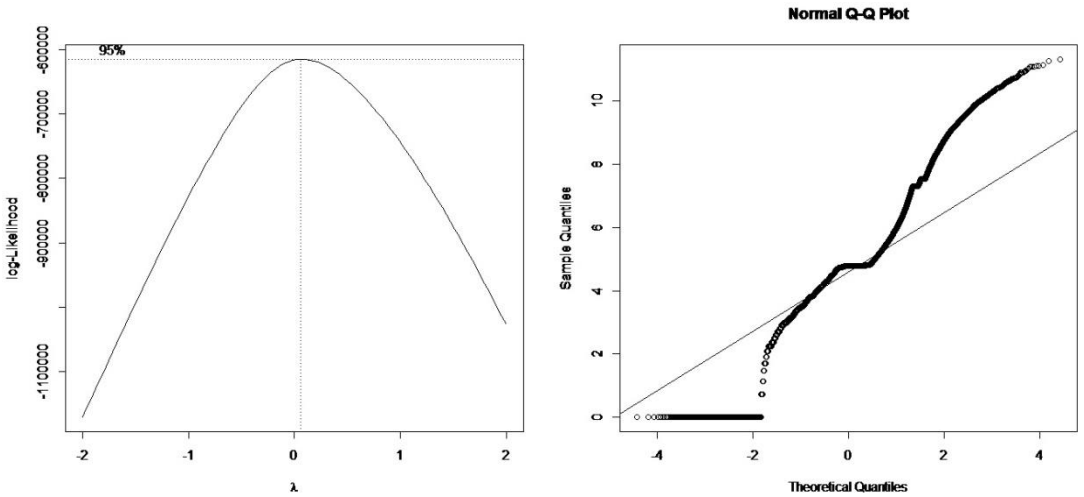
Για τα ύψη αποζημιώσεων υλικών ζημιών:



Σχήμα 3.4.1: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του "Total payment".

Η τιμή του λ που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας είναι -0,21 όπως προαναφέραμε. Η προσαρμογή των δεδομένων στην κανονική κατανομή είναι βελτιωμένη σε σχέση με τα αρχικά δεδομένα σύμφωνα με το Q-Q Plot διάγραμμα, αλλά και πάλι βλέπουμε σοβαρό πρόβλημα, ιδιαίτερα στην κάτω αριστερή ουρά της κατανομής. Διαγραμματικά τουλάχιστον δεν μπορούμε να εξάγουμε το συμπέρασμα ότι τα μετασχηματισμένα δεδομένα προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή.

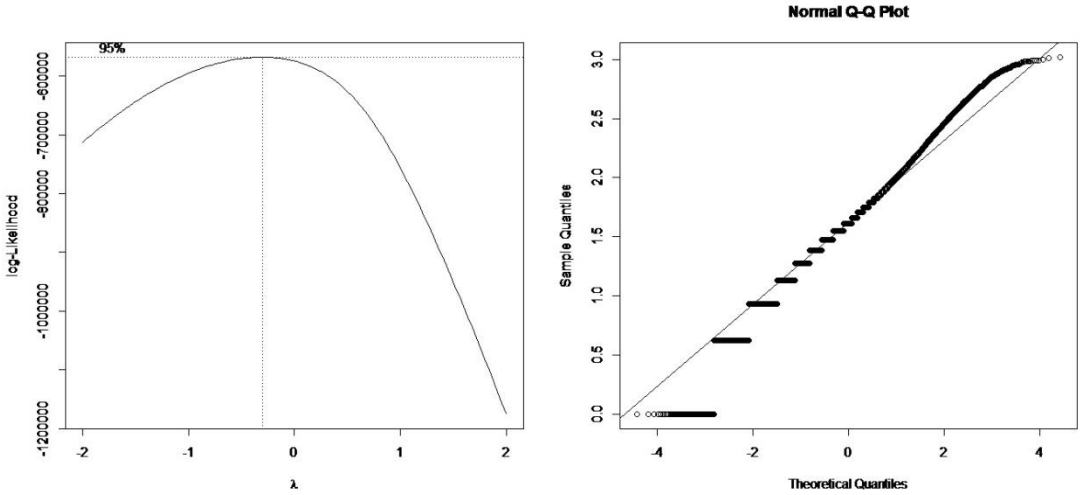
Για το χρόνο αποπληρωμής υλικών ζημιών:



Σχήμα 3.4.2: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του "Total time".

Η τιμή του λ που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας είναι 0,07. Σύμφωνα με το Q-Q Plot διάγραμμα, τα μετασχηματισμένα δεδομένα δεν εμφανίζουν καλή προσαρμογή και επομένως δεν προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή.

Για το χρόνο καθυστέρησης αναγγελίας της ζημιάς στην εταιρία:



Σχήμα 3.4.3: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του "Delay".

Η τιμή του λ που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας είναι -0,3. Από τη μορφή του Q-Q Plot διαγράμματος, τα μετασχηματισμένα δεδομένα δεν εμφανίζουν καλή προσαρμογή και επομένως δεν προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή.

Για να επαληθεύσουμε τους παραπάνω ισχυρισμούς μας ότι τα μετασχηματισμένα δεδομένα και για τις τρεις μεταβλητές δεν είναι κανονικά κατανεμημένα, διεξάγουμε τα τέσσερα τεστ κανονικότητας με τη βοήθεια του πακέτου R, όπως και στην προηγούμενη ενότητα. Τα αποτελέσματα εμφανίζονται στον πίνακα που ακολουθεί.

Πίνακας 3.4.4: Αποτελέσματα των τεστ κανονικότητας των μεταβλητών υλικών ζημιών.

Accidental damage	**Total Payment**		**Total time**		**Delay**	
	D/A/W/P	p-value	D/A/W/P	p-value	D/A/W/P	p-value
Lilliefors	0.0954	<2,2 e-16	0.1528	<2,2 e-16	0.0641	<2,2 e-16
Anderson-Darling	1063.974	<2,2 e-16	3139.101	<2,2 e-16	299.4115	<2,2 e-16
Cramer-von Mises	226.517	7.37E-10	619.2932	7.37E-10	48.8231	7.37E-10
Pearson	78339.26	<2,2 e-16	672269.2	<2,2 e-16	919258.5	<2,2 e-16

Τα αποτελέσματα που προκύπτουν από τους ελέγχους κανονικότητας επιβεβαιώνουν τις αρχικές μας προβλέψεις ότι τα δεδομένα δεν προέρχονται από την κανονική κατανομή, αφού τα p-value των ελέγχων εμφανίζουν την ελάχιστη τιμή που δίνεται

από το στατιστικό πακέτο R. Επομένως, απορρίπτουμε τη μηδενική υπόθεση, δηλαδή την υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή σε επίπεδο σημαντικότητας 5%.

Για τις μετασχηματισμένες τιμές των μεταβλητών που αφορούν σε υλικές ζημιές, δοκιμάστηκε η προσαρμογή τους και στις θεωρητικές συνεχείς κατανομές που είναι διαθέσιμες από το πρόγραμμα "EasyFit". Σύμφωνα με τα τρία τεστ καλής προσαρμογής που εκτελούνται αυτόματα από το "EasyFit" (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test), απορρίπτουμε την υπόθεση ότι τα μετασχηματισμένα δεδομένα του "Total payment", "Total time" και "Delay" υλικών ζημιών προέρχονται από κάποια από τις κατανομές αυτές.

3.4.2 Εφαρμογή του Box-Cox Power Transformation στα δεδομένα σωματικών βλαβών

Για την κατηγορία σωματικές βλάβες προσδιορίστηκαν οι συντελεστές λ για τις μεταβλητές "Total payment", "Total time" και "Delay". Από τον υπολογισμό προέκυψε τιμή του συντελεστή ίση με -0,12 για το ύψος αποζημιώσεων, 0,21 για το χρόνο αποπληρωμής και -0,39 για το χρόνο καθυστέρησης αναγγελίας της ζημιάς.

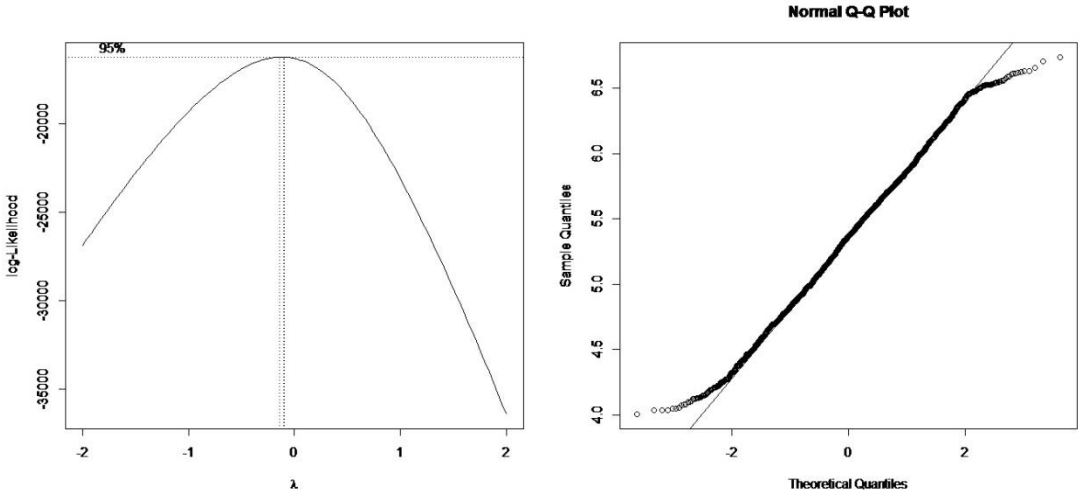
Τα δεδομένα μετασχηματίστηκαν σύμφωνα με τις παραπάνω τιμές του συντελεστή λ Συνοπτικά δίνονται στον πίνακα που ακολουθεί περιγραφικά χαρακτηριστικά για τα μετασχηματισμένα δεδομένα.

Πίνακας 3.4.5: Περιγραφικά χαρακτηριστικά μετασχηματισμών των μεταβλητών σωματικών βλαβών.

Bodily injury	Total payment	Total time	Delay
min	4.01	0	0
25% quantile	4.99	7.92	1.29
median (50% quantile)	5.37	12.92	1.52
mean	5.35	12.38	1.51
75% quantile	5.71	16.61	1.73
max	6.74	22.34	2.44
variance	0.2622	23.1870	0.1248
standard deviation	0.5120	4.8153	0.3533
99% quantile	6.5174	20.9790	2.3531
skewness	0.0096	-0.1160	-0.0379
kurtosis	-0.3377	-1.1766	0.5904

Επιπλέον, δίνονται τα διαγράμματα για την παράμετρο λ, καθώς και τα Normal Q-Q Plot διαγράμματα για τα μετασχηματισμένα δεδομένα.

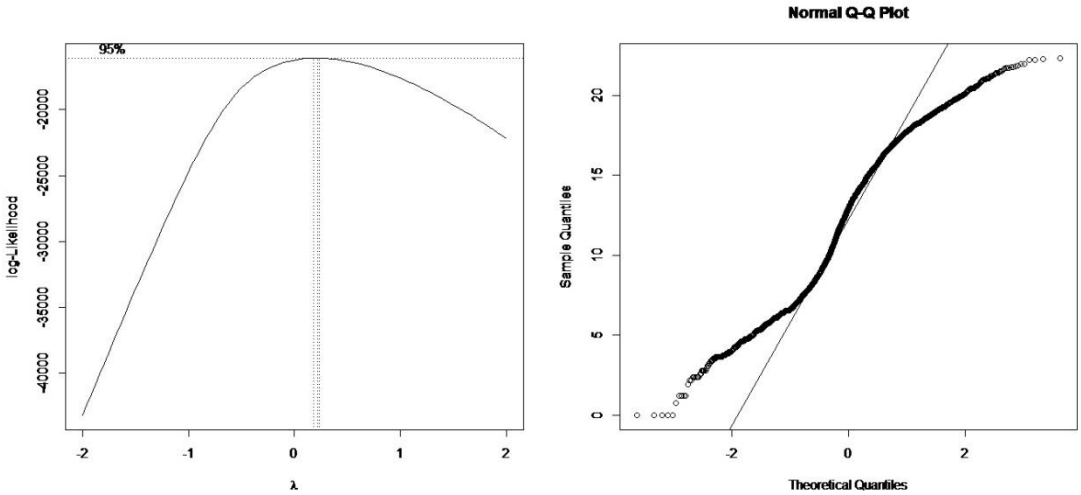
Για τα ύψη αποζημιώσεων σωματικών βλαβών:



Σχήμα 3.4.4: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του "Total payment".

Η τιμή του λ που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας είναι -0,12 όπως προαναφέραμε. Η προσαρμογή των δεδομένων στην κανονική κατανομή είναι βελτιωμένη σε σχέση με τα αρχικά δεδομένα σύμφωνα με το Q-Q Plot διάγραμμα, αλλά και πάλι βλέπουμε πρόβλημα στην προσαρμογή των ουρών της κατανομής. Διαγραμματικά τουλάχιστον δεν μπορούμε να εξάγουμε το συμπέρασμα ότι τα δεδομένα προέρχονται από πληθυσμό που ακολουθεί την κανονική κατανομή.

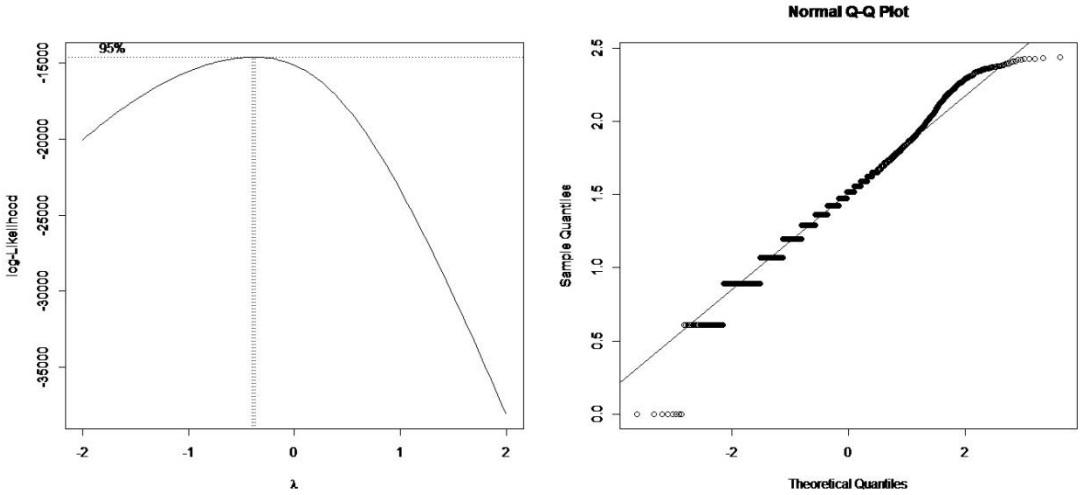
Για το χρόνο αποπληρωμής σωματικών βλαβών:



Σχήμα 3.4.5: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του "Total time".

Η τιμή του λ που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας είναι 0,21. Σύμφωνα με το Q-Q Plot διάγραμμα, τα μετασχηματισμένα δεδομένα δεν εμφανίζουν καλή προσαρμογή και επομένως είναι εμφανές ότι δεν προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή.

Για το χρόνο καθυστέρησης αναγγελίας της ζημιάς στην εταιρία:



Σχήμα 3.4.6: Διάγραμμα για το συντελεστή λ και Normal Q-Q Plot των μετασχηματισμένων δεδομένων του "Delay".

Η τιμή του λ που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας είναι -0,39. Από τη μορφή του Q-Q Plot διαγράμματος, τα μετασχηματισμένα δεδομένα δεν εμφανίζουν καλή προσαρμογή και επομένως δεν προέρχονται από πληθυσμό που ακολουθεί κανονική κατανομή.

Για να επαληθεύσουμε τους παραπάνω ισχυρισμούς μας ότι τα μετασχηματισμένα δεδομένα και για τις τρεις μεταβλητές δεν είναι κανονικά κατανεμημένα, διεξάγουμε τα τέσσερα τεστ κανονικότητας με τη βοήθεια του πακέτου R, όπως στην περίπτωση των υλικών ζημιών. Τα αποτελέσματα εμφανίζονται στον πίνακα που ακολουθεί.

Πίνακας 3.4.6: Αποτελέσματα των τεστ κανονικότητας των μεταβλητών σωματικών βλαβών.

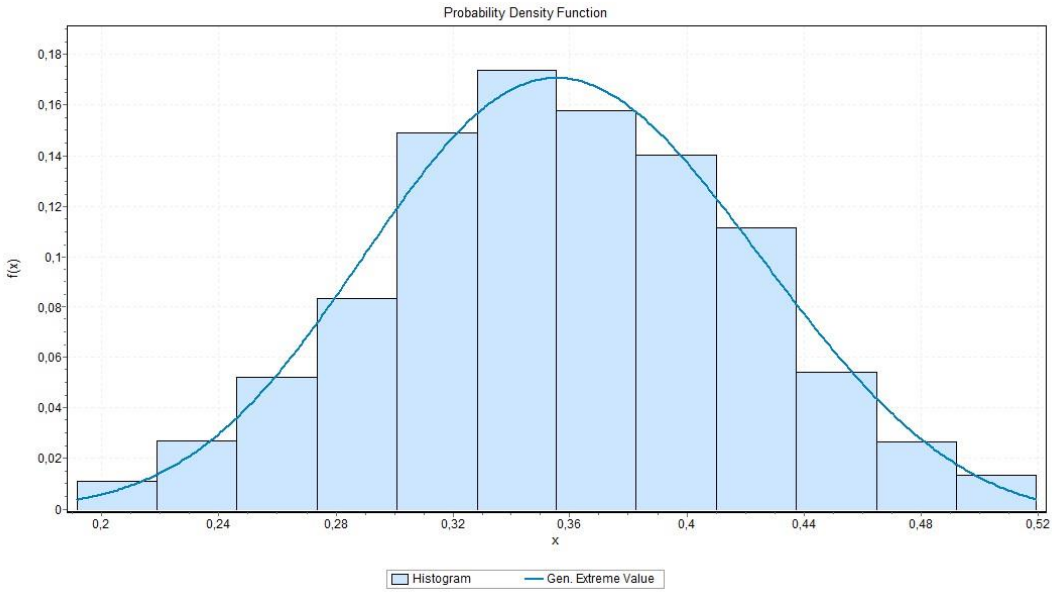
Bodily injury	**Total Payment**		**Total time**		**Delay**	
	D/A/W/P	p-value	D/A/W/P	p-value	D/A/W/P	p-value
Lilliefors	0.0192	0.003811	0.0791	<2,2 e-16	0.0542	<2,2 e-16
Anderson-Darling	1.2432	0.0031	55.0044	<2,2 e-16	10.6787	<2,2 e-16
Cramer-von Mises	0.1865	0.00798	9.1513	7.37E-10	1.5654	7.37E-10
Pearson	106.2257	9,29 e-06	928.472	<2,2 e-16	5632.558	<2,2 e-16

Τα αποτελέσματα που προκύπτουν από τους ελέγχους κανονικότητας επιβεβαιώνουν τις αρχικές μας προβλέψεις ότι τα δεδομένα δεν προέρχονται από την κανονική

κατανομή, αφού τα p-value των ελέγχων για τις μεταβλητές “Total time” και “Delay” εμφανίζουν την ελάχιστη τιμή που δίνεται από το στατιστικό πακέτο R. Για τη μεταβλητή “Total payment” τα p-value των ελέγχων δεν εμφανίζουν την ελάχιστη τιμή. Όμως και πάλι απορρίπτουμε τη μηδενική υπόθεση, καθώς για κάθε ένα από τα τεστ το p-value είναι μικρότερο από το επίπεδο σημαντικότητας 5%. Επομένως, απορρίπτουμε για κάθε μετασχηματισμό τη μηδενική υπόθεση, δηλαδή την υπόθεση ότι τα δεδομένα ακολουθούν κανονική κατανομή σε επίπεδο σημαντικότητας 5%.

Για τις μετασχηματισμένες τιμές των μεταβλητών που αφορούν σε σωματικές βλάβες, δοκιμάστηκε η προσαρμογή τους και στις θεωρητικές συνεχείς κατανομές που είναι διαθέσιμες από το πρόγραμμα “EasyFit”. Με εξαίρεση το μετασχηματισμό “Total payment”, σύμφωνα με τα τρία τεστ καλής προσαρμογής που εκτελούνται αυτόματα από το “EasyFit” (Kolmogorov-Smirnov, Anderson-Darling, Chi-Squared gof test), απορρίπτουμε την υπόθεση ότι τα μετασχηματισμένα δεδομένα του “Total time” και “Delay” προέρχονται από κάποια από τις κατανομές αυτές.

Για το μετασχηματισμό Box-Cox του ύψους των αποζημιώσεων σωματικών βλαβών, δίνεται το ιστόγραμμα συχνοτήτων.



Σχήμα 3.4.7: Ιστόγραμμα μετασχηματισμού Box – Cox “Total payment” σωματικών βλαβών.

Σύμφωνα με τα τεστ καλής προσαρμογής του “EasyFit” δεν απορρίπτουμε την υπόθεση ότι τα μετασχηματισμένα δεδομένα του δείγματος προέρχονται από την κατανομή Generalized Extreme Value. Η μπλε γραμμή που εμφανίζεται στο ιστόγραμμα είναι η καμπύλη της συνάρτησης πυκνότητας πιθανότητας της κατανομής Generalized Extreme Value.

Τα αποτελέσματα των ελέγχων καλής προσαρμογής του προγράμματος εμφανίζονται στον παρακάτω πίνακα.

Πίνακας 3.4.7: Αποτελέσματα των τεστ προσαρμογής του "EasyFit" για τη GEV.

Kolmogorov-Smirnov					
Sample Size	3629				
Statistic	0,01218				
P-Value	0,65017				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	0,01781	0,0203	0,02254	0,0252	0,02704
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size	3629				
Statistic	0,73186				
Rank	1				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	13,749	19,286	25,018	32,892	39,074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom	11				
Statistic	19,063				
P-Value	0,05998				
Rank	2				
a	0,2	0,1	0,05	0,02	0,01
Critical Value	14,631	17,275	19,675	22,618	24,725
Reject?	Yes	Yes	No	No	No

Σύμφωνα με τα αποτελέσματα του πίνακα, η υπόθεση ότι τα δεδομένα ακολουθούν την κατανομή Generalized Extreme Value δεν απορρίπτεται από τα τεστ Kolmogorov-Smirnov και Anderson-Darling καθώς το p-value των ελέγχων είναι μεγαλύτερο από τα χρησιμοποιούμενα επίπεδα σημαντικότητας. Αντίθετα, σύμφωνα με το Chi-squared test σε επίπεδα σημαντικότητας μεγαλύτερο του 5% η υπόθεση ότι τα δεδομένα προέρχονται από την κατανομή Generalized Extreme Value απορρίπτεται.

Δίνονται, τέλος, οι εκτιμήσεις των τριών παραμέτρων της κατανομής, όπως υπολογίζονται από το πρόγραμμα.

19	Gen. Extreme Value	k=-0,2839 σ=0,06162 μ=0,33582

Από την παραπάνω ανάλυση που πραγματοποιήσαμε για τις μεταβλητές που προσδιορίζουν το χαρτοφυλάκιο ζημιών, μπορούμε να συμπεράνουμε πως τα δεδομένα φαίνεται να προέρχονται από κατανομές που εμφανίζουν 2 ή και 3 κορυφές. Παρόλο που σε κάποιες περιπτώσεις μετασχηματισμών για τις μεταβλητές του ύψους αποζημιώσεων και χρόνου αποπληρωμής σωματικών βλαβών, τα στατιστικά τεστ καλής προσαρμογής οδήγησαν στη μη απόρριψη των επιλεγμένων κατανομών, εντούτοις δεν μας παρέχουν πληροφόρηση για τις κατανομές που ακολουθούν τα αρχικά δεδομένα του χαρτοφυλακίου ζημίων. Ωστόσο, από τις διαγραμματικές απεικονίσεις των μετασχηματισμών καταφέραμε να αντλήσουμε πρόσθετη γνώση για τις ιδιομορφίες των μεταβλητών υλικών ζημιών και σωματικών βλαβών. Ειδικότερα, στην περίπτωση του ύψους αποζημιώσεων αναμένουμε ότι τα δεδομένα προέρχονται από δικόρυφες κατανομές. Το ίδιο αναμένουμε και για τους χρόνους αποπληρωμής. Αντίθετα, από τη μελέτη μας για το χρόνο καθυστέρησης δήλωσης της ζημιάς στην ασφαλιστική εταιρία, τα δεδομένα περιμένουμε να προέρχονται από κάποια μονοκόρυφη κατανομή.

Κεφάλαιο 4: Διερεύνηση κατανομής με μοντέλα μικτού τύπου

4.1 Μοντέλα μείξης κατανομών

Στα κεφάλαια 2 και 3 διαπιστώσαμε με γραφικές μεθόδους, αλλά και με τη χρήση του "EasyFit" πως τα αρχικά δεδομένα του δείγματος δεν προέρχονται από κάποια γνωστή κατανομή. Ο μετασχηματισμός των δεδομένων των υπό εξέταση μεταβλητών, οδήγησε στην αποδοχή κατανομής για τους μετασχηματισμούς των "Total payment" και "Total time" σωματικών βλαβών. Από την ανάλυση που πραγματοποιήσαμε οδηγούμαστε στο συμπέρασμα ότι τα δεδομένα φαίνεται να προέρχονται από κατανομή ή κατανομές που παρουσιάζουν τουλάχιστον δύο κορυφές.

Στη στατιστική, πολυκόρυφη κατανομή (multimodal distribution) είναι μία συνεχής κατανομή πιθανότητας η οποία εμφανίζει δύο ή περισσότερες κορυφές. Αυτές εμφανίζονται ως διακριτές κορυφές (τοπικά μέγιστα) στη συνάρτηση πυκνότητας πιθανότητας. Οι ευρέως χρησιμοποιούμενες κατανομές, οι οποίες είναι διαθέσιμες και από το "EasyFit", είναι μονοκόρυφες κατανομές. Ένας τρόπος για να κατασκευαστεί μία κατανομή που παρουσιάζει περισσότερες από μία κορυφές είναι η μείξη δύο ή περισσοτέρων γνωστών κατανομών.

Έστω τ.μ. X_i έτσι ώστε $f_i(x) = P(X_i = x)$ για i=1,2,….κ και επίσης σταθερά w_i τέτοια ώστε:

$$0<w_i<1 \text{ και } \sum_{i=1}^{\kappa} w_i = 1$$

Τότε, η μορφή της συνάρτησης πυκνότητας πιθανότητας του μικτού μοντέλου θα είναι:

$$f(x) = \sum_{i=1}^{\kappa} w_i f_i(x)$$

Στο κεφάλαιο αυτό γίνεται προσπάθεια προσέγγισης της κατανομής που ακολουθούν τα δεδομένα για τα μεγέθη των αποζημιώσεων και για τους χρόνους αποπληρωμής από ένα μοντέλο μείξης δύο γάμμα κατανομών, δηλαδή για κ=2.

Έστω δηλαδή ότι, $f_1(x) \sim Ga(a_1, \beta_1)$ με $(a_1, \beta_1) > 0$ και $f_2(x) \sim Ga(a_2, \beta_2)$ με $(a_2, \beta_2) > 0$, τότε:

$$f(x) = wf_1(x) + (1 - w)f_2(x), \quad x > 0$$

Για να εκτιμήσουμε τις παραμέτρους του μοντέλου θα χρησιμοποιήσουμε τη μέθοδο των ροπών. Η μέθοδος αυτή οφείλεται στον K. Pearson και είναι η παλαιότερη όλων των άλλων μεθόδων σημειακής εκτίμησης.

Η μέθοδος αυτή συνίσταται στην εξίσωση των πρώτων ροπών της κατανομής με τις αντίστοιχες δειγματικές ροπές. Προκύπτει έτσι ένα σύστημα εξισώσεων, με άγνωστες τις προς εκτίμηση παραμέτρους της κατανομής. Η λύση του συστήματος αυτού μας δίνει τις ζητούμενες εκτιμήσεις των άγνωστων παραμέτρων της κατανομής.

Εφόσον το μοντέλο μας αποτελεί μία μείξη γάμμα κατανομών η ροπή κ τάξης θα είναι:

$$m_k = E(X^k) = \int_0^\infty x^k f(x)\ dx$$

και η ροπή κ τάξεως των δεδομένων θα είναι:

$$m'_k = \frac{1}{n}\sum x_i^k\ ,\ \ \text{όπου } n \text{ το πλήθος των παρατηρήσεων}$$

Η μέθοδος των ροπών συνίσταται στην εξίσωση των ροπών της κατανομής με τις αντίστοιχες ροπές που προκύπτουν από τα δεδομένα, δηλαδή

$m_1 = m'_1$

$m_2 = m'_2$

.

.

$m_k = m'_k$

Η λύση του παραπάνω συστήματος θα μας δώσει τις εκτιμήτριες των άγνωστων παραμέτρων της κατανομής.

Για το μοντέλο της μείξης δύο γάμμα κατανομών πρέπει να γίνει εκτίμηση πέντε παραμέτρων, δηλαδή των $a_1, \beta_1, a_2, \beta_2, w$. Επομένως θα εξισώσουμε τις πέντε πρώτες ροπές της νέας κατανομής με τις πρώτες πέντε ροπές των δεδομένων μας για κάθε υπό εξέταση μεταβλητή. Οι υπολογισμοί θα γίνουν με τη βοήθεια του Mathematica.

Η συνάρτηση πυκνότητας πιθανότητας της γάμμα (α, β) είναι:

$$f(x) = \frac{e^{-x/\beta} x^{-1+a} \beta^{-\alpha}}{\Gamma(\alpha)},\ \ x > 0$$

και επομένως η συνάρτηση πυκνότητας πιθανότητας του μοντέλου της μείξης των δύο γάμμα κατανομών θα είναι:

$$f(x) = w\frac{e^{-x/\beta_1} x^{-1+\alpha_1} {\beta_1}^{-\alpha_1}}{\Gamma(\alpha_1)} + (1-w)\frac{e^{-\frac{x}{\beta_2}} x^{-1+\alpha_2} {\beta_2}^{-\alpha_2}}{\Gamma(\alpha_2)}, \qquad x > 0$$

Η μέση τιμή μιας μεταβλητής X με πυκνότητα f(x) όπως αυτή ορίζεται πιο πάνω, θα ισούται με:

$$E(X) = \int_0^\infty x\, f(x)dx = \int_0^\infty x\left(wf_1(x) + (1-w)f_2(x)\right) dx =$$

$$= \int_0^\infty x\,(w\frac{e^{-x/\beta_1}x^{-1+\alpha_1}{\beta_1}^{-\alpha_1}}{\Gamma(\alpha_1)} + (1-w)\frac{e^{-\frac{x}{\beta_2}}x^{-1+\alpha_2}{\beta_2}^{-\alpha_2}}{\Gamma(\alpha_2)})\, dx =$$

$$= \int_0^\infty w\, x\, \frac{e^{-x/\beta_1}x^{-1+\alpha_1}{\beta_1}^{-\alpha_1}}{\Gamma(\alpha_1)}\, dx + \int_0^\infty (1-w)x\, \frac{e^{-\frac{x}{\beta_2}}x^{-1+\alpha_2}{\beta_2}^{-\alpha_2}}{\Gamma(\alpha_2)}\, dx =$$

$$= w\int_0^\infty x\, \frac{e^{-x/\beta_1}x^{-1+\alpha_1}{\beta_1}^{-\alpha_1}}{\Gamma(\alpha_1)}\, dx + \int_0^\infty x\, \frac{e^{-\frac{x}{\beta_2}}x^{-1+\alpha_2}{\beta_2}^{-\alpha_2}}{\Gamma(\alpha_2)}\, dx - w\int_0^\infty x\, \frac{e^{-\frac{x}{\beta_2}}x^{-1+\alpha_2}{\beta_2}^{-\alpha_2}}{\Gamma(\alpha_2)}\, dx =$$

$$= w\, E(x, a_1, \beta_1) + E(x, a_2, \beta_2) - w\, E(x, a_2, \beta_2) =$$

$$= w\, E(x, a_1, \beta_1) + (1-w)\, E(x, a_2, \beta_2)$$

Η μέση τιμή της κατανομής γάμμα με παραμέτρους (α, β) >0 που ισούται με την ροπή 1^{ης} τάξης είναι:

$$E(X) = m_1 = a\beta$$

και επομένως, η μέση τιμή του μοντέλου των δύο γάμμα θα είναι:

$$E(X) = w\, a_1\beta_1 + (1-w)a_2\beta_2$$

Με τη βοήθεια του Mathematica ορίζουμε ως και τη ροπή 5^{ης} τάξης μιας γάμμα κατανομής με παραμέτρους (α, β)>0.

- $m_2 = E(X^2) = a\,(1+a)\,\beta^2$
- $m_3 = E(X^3) = \alpha\,(1+\alpha)(2+\alpha)\,\beta^3$
- $m_4 = E(X^4) = \alpha\,(1+\alpha)\,(2+\alpha)\,(3+\alpha)\,\beta^4$
- $m_5 = E(X^5) = \alpha\,(1+\alpha)\,(2+\alpha)\,(3+\alpha)\,(4+\alpha)\,\beta^5$

Οι πρώτες πέντε ροπές του μοντέλου της μείξης των δύο Γάμμα κατανομών, όπως αυτές ορίστηκαν παραπάνω θα είναι:

- $m_1 = w(a_1\beta_1) + (1-w)(a_2\beta_2)$
- $m_2 = w(a_1(1+a_1)\,\beta_1^2) + (1-w)(a_2(1+a_2)\,\beta_2^2)$
- $m_3 = w(a_1(1+a_1)\,(2+a_1)\,\beta_1^3) + (1-w)(a_2(1+a_2)\,(2+a_2)\,\beta_2^3)$

- $m_4 = w(a_1(1+a_1)\,(2+a_1)\,(3+a_1)\,\beta_1^4) + (1-w\,)(a_2(1+a_2)\,(2+a_2)\,(3+a_2)\,\beta_2^4)$

- $m_5 = w\big(a_1(1+a_1)\,(2+a_1)\,(3+a_1)\,(4+a_1)\,\beta_1^5\big) + (1-w\,)\big(a_2(1+a_2)\,(2+a_2)\,(3+a_2)\,(4+a_2)\,\beta_2^5\big)$

Θα εξισώσουμε τις ροπές του μοντέλου με τις πρώτες πέντε ροπές από τα δεδομένα για το ύψος αποζημιώσεων "Total payment" και για το χρόνο αποπληρωμής από την ασφαλιστική εταιρία απαιτήσεων "Total time" για τις δύο κατηγορίες ζημιών, δηλαδή για υλικές ζημιές και σωματικές βλάβες.

4.1.1 Μοντέλο μείξης κατανομών – Εφαρμογή στο ύψος αποζημιώσεων

Στην παρούσα υποενότητα, εφαρμόζουμε το μοντέλο της μείξης των γάμμα κατανομών στο ύψος αποζημιώσεων υλικών ζημιών και σωματικών βλαβών. Δίνονται στον παρακάτω πίνακα οι τιμές των πέντε πρώτων ροπών όπως υπολογίστηκαν από τα αρχικά δεδομένα.

Πίνακας 4.1: Ροπές της μεταβλητής "Total payment".

Total payment	Accidental damage	Bodily injury
m1	1.59	21.65
m2	27.85	3831.78
m3	6196.43	1449325.13
m4	2471400.41	759274056.00
m5	1167881548.83	493534896015.94

Σημειώνεται ότι πριν την εξαγωγή των τιμών των ροπών, όπως φαίνονται στον παραπάνω πίνακα, οι τιμές των αρχικών αποζημιώσεων διαιρέθηκαν με το 1000.

Με τη βοήθεια του Mathematica, για το "Total payment" υλικών ζημιών λύνουμε το παρακάτω σύστημα εξισώσεων:

```
eq1=w*(a1*b1)+(1-w)*(a2*b2)==1.59

eq2=w*(a1*(1+a1)*b1^2)+(1-w)*(a2*(1+a2)*b2^2)==27.85

eq3=w*(a1*(1+a1)*(2+a1)*b1^3)+(1-w)*(a2*(1+a2)*(2+a2)*b2^3)==6196.43

eq4=w*(a1*(1+a1)*(2+a1)*(3+a1)*b1^4)+(1-
w)*(a2*(1+a2)*(2+a2)*(3+a2)*b2^4)==2471400.41

eq5=w*(a1*(1+a1)*(2+a1)*(3+a1)*(4+a1)*b1^5)+(1-w)*(a2*(1+a2)*(2+a2)

*(3+a2)*(4+a2)*b2^5)==1167881548.83

Solve[{eq1,eq2,eq3,eq4,eq5},{w,a1,b1,a2,b2}]
```

Αντίστοιχα εργαζόμαστε για το “Total payment” σωματικών βλαβών, εξισώνοντας δηλαδή τις πέντε πρώτες ροπές της μείξης των γάμμα κατανομών με τις ροπές όπως υπολογίστηκαν από τα δεδομένα και φαίνονται στον πίνακα 4.1.

Η λύση του συστήματος που δεχόμαστε είναι αυτή που πληροί τη συνθήκη $\alpha_1, \beta_1, \alpha_2, \beta_2 > 0$, δηλαδή οι παράμετροι των δύο γάμμα κατανομών είναι θετικές και $0 \leq w \leq 1$.

Οι εκτιμήτριες των παραμέτρων που προκύπτουν από τη μέθοδο των ροπών για το “Total payment” υλικών ζημιών και σωματικών βλαβών (βλ. αναλυτικά αποτελέσματα - Παράρτημα Γ), οι οποίες πληρούν τις συνθήκες που ορίσαμε είναι:

Πίνακας 4.2: Εκτιμήσεις παραμέτρων μικτού μοντέλου του “Total payment”.

Total payment	Accidental damage	Bodily injury
α1	3.06845	1.38293
β1	66.9031	120.995
α2	0.3062	0.402624
β2	4.97906	25.5616
w	0.000321018	0.0723293

Έχοντας εκτιμήσει τις παραμέτρους των μικτών μοντέλων για το ύψος αποζημιώσεων μπορούμε να κατασκευάσουμε τις συναρτήσεις πυκνότητας πιθανότητας με τη βοήθεια του Mathematica.

Η σ.π.π. του ύψους των αποζημιώσεων υλικών ζημιών:

$$f(x) = \begin{cases} 0.000321\,(1.1744896 * 10^{-6}\ e^{-0.01494699\,x}\ x^{2.06845}) + 0.99968\ \dfrac{0.2089125\ e^{-0.20084\,x}}{x^{0.6938}} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

Η σ.π.π. του ύψους των αποζημιώσεων σωματικών βλαβών:

$$f(x) = \begin{cases} 0.0723293\,(0.00148287\ e^{-0.0082648\,x}\ x^{0.38293}) + 0.9276707\ \dfrac{0.123079744\ e^{-0.0391211817\ x}}{x^{0.597376}} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

4.1.2 Μοντέλο μείξης κατανομών – Εφαρμογή στο χρόνο αποπληρωμής

Με τον ίδιο τρόπο που εργαστήκαμε για τη μεταβλητή ύψος αποζημιώσεων, θα εργαστούμε και για την εκτίμηση των παραμέτρων του μοντέλου της μείξης των δύο γάμμα κατανομών για το χρόνο αποπληρωμής "Total time" των κατηγοριών ζημιών.

Στον πίνακα που ακολουθεί δίνονται οι τιμές των πέντε πρώτων ροπών όπως υπολογίστηκαν από τα δεδομένα.

Πίνακας 4.3: Ροπές της μεταβλητής "Total time".

Total time	Accidental damage	Bodily injury
m1	4.17	25.61
m2	86.76	1312.62
m3	3967.37	85544.10
m4	247487.82	6444108.48
m5	18501200.19	540652504.05

Σημειώνεται ότι πριν την εξαγωγή των τιμών των ροπών, όπως φαίνονται στον παραπάνω πίνακα, οι τιμές των αρχικών χρόνων αποπληρωμής διαιρέθηκαν με το 30, διαφορετικά θα προέκυπταν πάρα πολύ μεγάλες τιμές για τις ροπές $4^{ης}$ και $5^{ης}$ τάξης.

Όπως στην προηγούμενη ενότητα θα επιλύσουμε δύο συστήματα εξισώσεων. Στην περίπτωση του χρόνου αποπληρωμής σωματικών βλαβών δεν μπορούμε να δεχτούμε κανένα από τα αποτελέσματα που εξάγονται από το Mathematica, καθώς δεν πληρούνται οι συνθήκες που θέσαμε για τις παραμέτρους.

Τα αποτελέσματα για το χρόνο αποπληρωμής υλικών ζημιών φαίνονται στον πίνακα που ακολουθεί.

Πίνακας 4.4: Εκτιμήσεις παραμέτρων μικτού μοντέλου του "Total time".

Total time	Accidental damage
α1	2.82873
β1	11.0086
α2	0.77106
β2	3.37669
w	0.0548898

Έχοντας εκτιμήσει τις παραμέτρους του μικτού μοντέλου για το χρόνο αποπληρωμής υλικών ζημιών μπορούμε να κατασκευάσουμε τη συνάρτηση πυκνότητας πιθανότητας.

Η σ.π.π. του χρόνου αποπληρωμής υλικών ζημιών:

$$f(x) = \begin{cases} 0.05489\ (0.000658\ e^{-0.09083807\ x}\ x^{1.82873}) + 0.94511\ \dfrac{0.32651869\ \ e^{-0.296148\ x}}{x^{0.22894}} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

Έχοντας παράγει με τη μέθοδο των ροπών τα μοντέλα των μεικτών γάμμα κατανομών, θα πρέπει να διαπιστώσουμε κατά πόσον η προσέγγιση αυτή είναι σωστή και το βαθμό στον οποίο τελικά τα δεδομένα μας για το "Total payment" και για το "Total time" υλικών ζημιών και σωματικών βλαβών προέρχονται από μοντέλα αυτής της μορφής. Για να εξετάσουμε την προσαρμογή των πραγματικών παρατηρήσεων στα παραπάνω μοντέλα μπορούμε όπως και σε προηγούμενες ενότητες να εργαστούμε είτε διαγραμματικά, είτε μέσω κάπου τεστ καλής προσαρμογής.

Στο κεφάλαιο 2 επιχειρήσαμε να εξάγουμε συμπεράσματα με τη βοήθεια του στατιστικού πακέτου R για τη μορφή της κατανομής των δεδομένων μας. Η φύση των δεδομένων και για τις δύο μεταβλητές δεν μας επέτρεψε να καταλήξουμε σε κάποιο συμπέρασμα μέσω των διαγραμμάτων που παράγαμε, καθώς η μορφή των ιστογραμμάτων δεν μας έδινε πληροφορία για τη κατανομή που αυτά μπορεί να ακολουθούν. Επομένως, στο σημείο αυτό θα εργαστούμε με το τεστ chi-squared καλής προσαρμογής για να καταλήξουμε στο αν τελικά μπορούμε να δεχτούμε πως τα δεδομένα προέρχονται από τη μείξη δύο γάμμα κατανομών ή όχι.

4.2 Έλεγχος καλής προσαρμογής εμπειρικών κατανομών σε θεωρητικές

Στη στατιστική ανάλυση πολλές φορές θέλουμε να ελέγξουμε κατά πόσο μία εμπειρική κατανομή, που προέρχεται από ένα τυχαίο δείγμα, ακολουθεί μία "γνωστή" κατανομή, δηλαδή κατά πόσο το δείγμα προέρχεται από μία ορισμένη θεωρητική κατανομή. Ο έλεγχος αυτός ορθής προσαρμογής γίνεται συνήθως με το χ^2.

Ο έλεγχος καλής ή όχι προσαρμογής μιας θεωρητικής κατανομής σε μία εμπειρική κατανομή συχνοτήτων στηρίζεται στη σύγκριση μεταξύ θεωρητικών και εμπειρικών συχνοτήτων. Υποθέτουμε ότι η εμπειρική κατανομή ακολουθεί μία "γνωστή" θεωρητική κατανομή και στη συνέχεια την υπόθεση την ελέγχουμε με το κριτήριο:

$$\chi^2 = \sum \frac{(f_i - np_i)^2}{np_i}$$

όπου f_i είναι οι εμπειρικές συχνότητες που αντιστοιχούν σε κάθε τιμή της μεταβλητής X, np_i οι αντίστοιχες θεωρητικές συχνότητες και n ο συνολικός αριθμός συχνοτήτων της κατανομής του δείγματος. Την τιμή χ^2 τη συγκρίνουμε με την τιμή $\chi^2_{\nu,\alpha}$ με $\nu =$

$n - \kappa - 1$ βαθμούς ελευθερίας, όπου κ είναι ο αριθμός των παραμέτρων που χρησιμοποιήσαμε για να κάνουμε την προσαρμογή.

Στην περίπτωση που ο έλεγχος πραγματοποιείται κατατάσσοντας τις n παρατηρήσεις της μεταβλητής X σε k κατηγορίες τιμών, τότε αντίστοιχα f_i είναι οι εμπειρικές συχνότητες που αντιστοιχούν σε κάθε κατηγορία και np_i οι θεωρητικές συχνότητες της κατηγορίας αυτής. Την τιμή χ^2 τη συγκρίνουμε με την τιμή $\chi^2_{\nu,\alpha}$ με $\nu = k - 1$ βαθμούς ελευθερίας, όπου k το πλήθος των κατηγοριών που εντάσσονται οι παρατηρήσεις.

Για την εφαρμογή του παραπάνω κριτηρίου, θα πρέπει οι θεωρητικές συχνότητες που αντιστοιχούν σε κάθε κατηγορία ή τιμή της μεταβλητής X να είναι μεγαλύτερες ή ίσες με 5. Αν έχουμε συχνότητες μικρότερες από 5, τότε τις προσθέτουμε στις προηγούμενες ή τις επόμενες συχνότητες, ώστε να έχουμε άθροισμα μεγαλύτερο ή ίσο του 5.

Για τη διερεύνηση της προσαρμογής των τιμών των μεταβλητών ύψος αποζημιώσεων και χρόνος αποπληρωμής απαιτήσεων από την ασφαλιστική εταιρία, πραγματοποιούμε τεστ καλής προσαρμογής χ^2 με χρήση του Excel.

Οι υποθέσεις που γίνονται είναι οι εξής:

H_0: Τα δεδομένα προέρχονται από το μοντέλο μείξης 2 γάμμα κατανομών

H_1: Τα δεδομένα δεν προέρχονται από το μοντέλο μείξης 2 γάμμα κατανομών

Για τον έλεγχο της μηδενικής υπόθεσης ακολουθούμε την παρακάτω διαδικασία.

Το σύνολο των παρατηρήσεων μας για τη μεταβλητή ύψος αποζημιώσεων υλικών ζημιών είναι $n = 102.454$. Κατατάσσουμε τις τιμές αυτές σε $k = 38$ κατηγορίες. Σύμφωνα με την κατηγοριοποίηση που πραγματοποιούμε το άθροισμα των αναμενόμενων συχνοτήτων σε κάθε κατηγορία είναι μεγαλύτερο του 5 για να είναι δυνατή η εφαρμογή του κριτηρίου.

Οι παρατηρούμενες συχνότητες σε κάθε κατηγορία είναι $f_1, f_2, \ldots\ldots, f_k$ έτσι ώστε $\sum_{i=1}^{k} f_i = n = 102.454$.

Θα υπολογίσουμε το αναμενόμενο πλήθος παρατηρήσεων np_i που θα βρεθούν σε κάθε κατηγορία, κάτω από τη μηδενική υπόθεση, δηλαδή υπό την προϋπόθεση ότι τα δεδομένα προέρχονται από το μικτό μοντέλο που κατασκευάσαμε.

Έστω η πρώτη κατηγορία περιέχει τιμές του “Total payment” υλικών ζημιών που ανήκουν στο διάστημα $(0,\ z_1)$, η δεύτερη $(z_1,\ z_2)$, η τρίτη $(z_2,\ z_3)$ κ.τ.λ. όπου z_i το άνω άκρο της κάθε κατηγορίας. Οι πιθανότητες $p_1, p_2,\ \ldots,\ p_k$ υπολογίζονται μέσω της αθροιστικής συνάρτησης πυκνότητας πιθανότητας του μοντέλου μείξης γάμμα κατανομών $F(z_i)$ ως εξής:

$$
\begin{aligned}
&p_1 = F(z_1) = P(0 \le x \le z_1)\\
&p_2 = F(z_2) - F(z_1) = P(0 \le x \le z_2) - P(0 \le x \le z_1)\\
&p_3 = F(z_3) - F(z_2) = P(0 \le x \le z_3) - P(0 \le x \le z_2)\\
&.\\
&.\\
&p_k = F(z_k) - F(z_{k-1}) = P(0 \le x \le z_k) - P(0 \le x \le z_{k-1})
\end{aligned}
$$

όπου $F(z_i) = P(0 \le x \le z_i) = \int_0^{z_i} f(x)\ dx$, έτσι ώστε να ισχύει:

$$np_1 + np_2 + np_3 + \ldots + np_k = n\sum_{i=1}^{k} p_i = n$$

Υπολογίσουμε τις πιθανότητες $F(z_i)$ κάνοντας χρήση του Mathematica και της εντολής Integrate.

Πιο συγκεκριμένα, η σ.π.π. του ύψους των αποζημιώσεων υλικών ζημιών όπως έχει ήδη οριστεί είναι:

$$f(x) = \begin{cases} 0.000321\,(1.1744896 * 10^{-6}\ e^{-0.01494699\,x}\ x^{2.06845}) + 0.99968\ \dfrac{0.2089125\ e^{-0.20084\,x}}{x^{0.6938}} & , x > 0 \\ 0 & , x \le 0 \end{cases}$$

Η εντολή που εκτελείται στο Mathematica είναι:

```
Integrate[0.000321018*1.1744895822360309*(10^-6) *E^(-
0.014946990498198141* x)*(x^2.06845)+0.999678982*0.20891253947208602*
E^(-0.20084112262153903*x)/x^0.6938,{x,0,z}]
```

Επαναλαμβάνουμε τόσες φορές όσες και οι κατηγορίες που έχουμε χωρίσει τις παρατηρήσεις μας, δηλαδή στην περίπτωση του “Total payment” υλικών ζημιών 38 φορές, αντικαθιστώντας κάθε φορά το z_i με το άνω άκρο της αντίστοιχης κατηγορίας.

Τα αποτελέσματα που δίνονται από το Mathematica τα εισάγουμε στο υπολογιστικό φύλλο του Excel. Με μία απλή αφαίρεση διαδοχικών κελιών έχουμε ως αποτέλεσμα τις πιθανότητες $p_1, p_2,\ \ldots,\ p_k$ οι τιμές του “Total payment” υλικών ζημιών να ανήκουν αντίστοιχα στα διάστημα $(0,\ z_1)$, $(z_1,\ z_2)$, $(z_2,\ z_3)$ κ.τ.λ.

Πολλαπλασιάζουμε τις πιθανότητες p_i με το συνολικό αριθμό παρατηρήσεων του δείγματος $n = 102.454$. Ως αποτέλεσμα έχουμε πλέον το αναμενόμενο πλήθος παρατηρήσεων np_i που θα βρεθούν σε κάθε κατηγορία, κάτω από τη μηδενική

υπόθεση. Στη συνέχεια υπολογίζουμε τις διαφορές μεταξύ των παρατηρούμενων συχνοτήτων f_i και των αναμενόμενων συχνοτήτων np_i σε κάθε μια από τις 38 κατηγορίες που έχουμε χωρίσει τα δεδομένα του δείγματος. Το άθροισμα των τετραγώνων των διαφορών διαιρεμένο με την αντίστοιχη για κάθε κατηγορία αναμενόμενη συχνότητα, μας δίνει την τιμή της στατιστικής ελέγχου χ^2.

Τέλος, η τιμή της στατιστικής συνάρτησης χ^2 συγκρίνεται με την τιμή $\chi^2_{\nu,\alpha}$ με $\nu = k - 1$. Μεγάλες τιμές της στατιστικής συνάρτησης ελέγχου χ^2 συνηγορούν υπέρ της απόρριψης της μηδενικής υπόθεσης ότι τα δεδομένα προέρχονται από πληθυσμό που ακολουθεί τη θεωρητική κατανομή. Για συγκεκριμένο επίπεδο σημαντικότητας α απορρίπτουμε τη μηδενική υπόθεση όταν $\chi^2 > \chi^2_{\nu,\alpha}$.

Στην περίπτωση των αποζημιώσεων υλικών ζημιών η τιμή της στατιστικής συνάρτησης χ^2 ισούται με 180.757,5. Την τιμή αυτή τη συγκρίνουμε με την τιμή της $\chi^2_{\nu,\alpha}$ για $\nu = 37$ βαθμούς ελευθερίας και σε επίπεδο σημαντικότητας α=5%, που ισούται με 18,59.

Απορρίπτουμε την μηδενική υπόθεση ότι τα δεδομένα αποζημιώσεων υλικών ζημιών ακολουθούν το μοντέλο μείξης γάμμα κατανομών σε επίπεδο σημαντικότητας 5%, καθώς χ^2=180.757,5 είναι μεγαλύτερο από το $\chi^2_{\nu,\alpha}$=18,59.

Η διαδικασία που αναλύθηκε για τον έλεγχο της προσαρμογής των τιμών του ύψους αποζημιώσεων υλικών ζημιών στο θεωρητικό μοντέλο της μείξης δύο γάμμα κατανομών που κατασκευάσαμε, ακολουθήθηκε και στην περίπτωση των αποζημιώσεων σωματικών βλαβών. Το σύνολο των παρατηρήσεων μας για το "Total payment" σωματικών βλαβών είναι $n = 3.629$. Οι τιμές του "Total payment" σωματικών βλαβών χωρίστηκαν σε $k = 23$ κατηγορίες.

Στην περίπτωση των αποζημιώσεων σωματικών βλαβών η τιμή της στατιστικής συνάρτησης χ^2 ισούται με 1.009,4. Την τιμή αυτή τη συγκρίνουμε με την τιμή της $\chi^2_{\nu,\alpha}$ για $\nu = 22$ βαθμούς ελευθερίας και σε επίπεδο σημαντικότητας α=5%, που ισούται με 8,64.

Και σε αυτή την περίπτωση αποζημιώσεων, απορρίπτουμε την μηδενική υπόθεση ότι τα δεδομένα ακολουθούν το μοντέλο μείξης γάμμα κατανομών σε επίπεδο σημαντικότητας 5%, καθώς χ^2=1.009,4 είναι μεγαλύτερο από το $\chi^2_{\nu,\alpha}$=8,64.

Τέλος, με τον ίδιο τρόπο εργαστήκαμε για να ελέγξουμε τη συμβατότητα του μοντέλου μείξης γάμμα κατανομών με τα δεδομένα που περιγράφουν το χρόνο αποπληρωμής των αποζημιώσεων υλικών ζημιών. Το σύνολο των παρατηρήσεων μας για το "Total time" υλικών ζημιών είναι $n = 102.454$. Οι τιμές του "Total time" υλικών ζημιών χωρίστηκαν σε $k = 29$ κατηγορίες.

Η τιμή της στατιστικής συνάρτησης χ^2 υπολογίζεται ίση με 71.071,05. Την τιμή αυτή τη συγκρίνουμε με την τιμή της $\chi^2_{\nu,\alpha}$ για $\nu = 28$ βαθμούς ελευθερίας και σε επίπεδο σημαντικότητας α=5%, που ισούται με 12,46.

Για ακόμα μία φορά, απορρίπτουμε την μηδενική υπόθεση ότι τα δεδομένα ακολουθούν το μοντέλο μείξης γάμμα κατανομών σε επίπεδο σημαντικότητας 5%, καθώς χ^2=71.071,05 είναι μεγαλύτερο από το $\chi^2_{\nu,\alpha}$= 12,46.

Συμπερασματικά, από την ανάλυση που πραγματοποιήθηκε στο παρόν κεφάλαιο δεν καταφέραμε να εντοπίσουμε την κατανομή που περιγράφει τα δεδομένα των ζημιών της ασφαλιστικής εταιρίας. Οι αποκλίσεις των παρατηρούμενων συχνοτήτων από τις αντίστοιχες των θεωρητικών μοντέλων μείξης γάμμα κατανομών παίρνουν μεγάλες τιμές και οδηγούν στην απόρριψη του μικτού μοντέλου για την περιγραφή των μεταβλητών. Οι προσπάθειες διερεύνησης της κατανομής των αρχικών δεδομένων των τριών μεταβλητών (“Total payment”, “Total time”, “Delay”) στο παρόν αλλά και στα κεφάλαια 2 και 3 δεν οδήγησαν στο επιθυμητό αποτέλεσμα. Οι διαδικασίες που ακολουθήθηκαν τόσο για τα αρχικά αλλά και για τα μετασχηματισμένα δεδομένα ζημιών μας έχουν δώσει την πληροφορία ότι οι κατανομές των μεταβλητών παρουσιάζουν ιδιομορφίες (π.χ. 2-3 κορυφές) και επομένως πρέπει να αναζητηθεί ένας άλλος τρόπος περιγραφής τους. Εφόσον και το μικτό μοντέλο απέτυχε να ερμηνεύσει τη συμπεριφορά του ύψους των ζημιών και του χρόνου αποπληρωμής θα αναζητήσουμε έναν άλλο τρόπο περιγραφής των τιμών των μεταβλητών.

Κεφάλαιο 5: Πολυγραμμικό μοντέλο

Στην υποενότητα 2.1 δόθηκε μία σύντομη παρουσίαση του αρχείου των ζημιών που χρησιμοποιήθηκε για την πραγματοποίηση της παρούσας διπλωματικής εργασίας. Όπως έχουμε ήδη αναφέρει τα στοιχεία προέρχονται από χαρτοφυλάκιο ασφαλιστικής εταιρίας και αφορούν σε καταβολές αποζημιώσεων για ζημιές με ημερομηνία συμβάντος από 1/1/2003 μέχρι και 31/12/2010. Πέραν των μεμονωμένων πληρωμών που διεκπεραιώθηκαν με τα χαρακτηριστικά τους (ποσό πληρωμής, ημερομηνία πληρωμής, αριθμό φακέλου ζημιάς, ένδειξη για το αν αφορούν σε υλικές ζημιές ή σωματικές βλάβες) δεν υπάρχει πληροφόρηση για τον τρόπο με τον οποίο διακανονίστηκαν οι ζημιές της εταιρίας. Στο παρόν κεφάλαιο θα γίνει παρουσίαση του συστήματος άμεσης πληρωμής (Σ.Α.Π.) που θεσπίστηκε για το διακανονισμό αποζημιώσεων του κλάδου αστικής ευθύνης αυτοκινήτου. Βάσει αυτού θα δοθεί ερμηνεία της συμπεριφοράς της μεταβλητής του ύψους των αποζημιώσεων, καθώς υπάρχει η πληροφορία ότι η εταιρία το χαρτοφυλάκιο ζημιών της οποίας μελετούμε συμμετέχει σε αυτό. Τέλος, θα κατασκευαστεί ένα νέο πολυγραμμικό μοντέλο για την προσέγγιση της κατανομής που ακολουθούν τα δεδομένα αποζημιώσεων.

5.1 Συμφωνία άμεσου διακανονισμού ζημιών από τροχαίο ατύχημα – Σύστημα άμεσης πληρωμής (Σ.Α.Π.)

Από την 1η Μαίου 2000 οι ασφαλιστικές επιχειρήσεις που ασκούν στην Ελλάδα τον κλάδο ασφάλισης αστικής ευθύνης αυτοκινήτων καθιέρωσαν ένα σύστημα αποζημίωσης των αναίτιων ασφαλιζομένων πελατών τους που εμπλέκονται σε τροχαίο ατύχημα, απαλλάσσοντάς τους από την ανάγκη να στραφούν κατά της ασφαλιστικής επιχείρησης που καλύπτει την ευθύνη του υπαίτιου οδηγού. Πρόκειται για μια Συμφωνία Κυρίων ανοικτή σε όλες τις ασφαλιστικές επιχειρήσεις, εφόσον ασκούν τον κλάδο ασφάλισης αστικής ευθύνης αυτοκινήτων.

Κατ' εφαρμογή της Συμφωνίας που καθιερώνει το Σύστημα Άμεσης Πληρωμής (Σ.Α.Π.) ή Φιλικός Διακανονισμός σε περίπτωση τροχαίου ατυχήματος, εφόσον και οι δύο εμπλεκόμενες ασφαλιστικές εταιρίες έχουν προσχωρήσει σ' αυτήν, επιτρέπεται στην ασφαλιστική εταιρία του αναίτιου οδηγού να τον αποζημιώσει για λογαριασμό της ασφαλιστικής εταιρίας που καλύπτει την ευθύνη του υπαίτιου οδηγού.

Η εφαρμογή του Συστήματος Άμεσης Πληρωμής ξεκίνησε με την αποζημίωση ατυχήματος μόνο με υλικές ζημιές. Από 1ης Σεπτεμβρίου 2005 καλύπτονται και ατυχήματα που προκαλούν μικρές σωματικές βλάβες.

Το Σ.Α.Π. ή Φιλικός Διακανονισμός εφαρμόζεται μόνο σε ατυχήματα που συμβαίνουν εντός Ελλάδος και εφόσον πληρούνται οι εξής προϋποθέσεις:

- Το τροχαίο ατύχημα έχει αποκλειστικά και μόνο υλικές ζημιές μέχρι 6.500 € ή σωματικές βλάβες μέχρι 12.000 € ανά ζημιωθέν πρόσωπο και μέχρι 30.000 € ανά συμβάν (όρια για την εφαρμογή Σ.Α.Π.).
- Σε περίπτωση ατυχήματος με σύγκρουση περισσοτέρων των δύο οχημάτων, το εκάστοτε ισχύον όριο των υλικών ζημιών εφαρμόζεται ανά όχημα.
- Δεν μπορούν να αποζημιωθούν ατυχήματα άνευ επαφής.

Με τη Συμφωνία, χωρίς να αλλάξει σε τίποτε η υφιστάμενη νομοθεσία που διέπει την υποχρεωτική ασφάλιση της αστικής ευθύνης από την κυκλοφορία αυτοκινήτων, καθιερώνεται σύστημα εξυπηρέτησης των πελατών μιας ασφαλιστικής επιχείρησης, οι οποίοι προφανώς προτιμούν και σ' αυτή την περίπτωση να κάνουν χρήση των υπηρεσιών της δικής τους εταιρίας, της εταιρίας δηλαδή που επέλεξαν για να ασφαλιστούν έχοντας λάβει υπ' όψιν τους τα ιδιαίτερα χαρακτηριστικά της, άρα και την ποιότητα των υπηρεσιών της. Κατά συνέπεια η ύπαρξη του Συστήματος Άμεσης Πληρωμής διευκολύνει και επιταχύνει τις διαδικασίες του διακανονισμού και της αποζημίωσης.

Αναπόσπαστο μέρος της Συμφωνίας αποτελούν οι Πίνακες Υπαιτιότητας. Πρόκειται για συμφωνημένες περιπτώσεις ατυχημάτων, που όταν συμβούν οδηγούν με απόλυτη ασφάλεια στο συμπέρασμα για το ποιος από τους δύο εμπλεκόμενους οδηγούς ευθύνεται για το ατύχημα εξ' ολοκλήρου ή έστω κατά 50%. Αποζημιώνοντας οι ασφαλιστικές εταιρίες τους πελάτες τους για ατύχημα για το οποίο δε φέρουν ευθύνη ενεργούν, όπως ελέχθη για λογαριασμό της ασφαλιστικής εταιρίας που καλύπτει την ευθύνη του υπαίτιου οδηγού.

Η μεταξύ των δύο εμπλεκομένων ασφαλιστικών εταιριών χρηματική συναλλαγή τακτοποιείται μέσω ενός μηχανισμού συμψηφισμού που λειτουργεί μέσα στην Ένωση Ασφαλιστικών Εταιριών Ελλάδος. Οι εκκαθαρίσεις των πληρωμών γίνονται σε μηνιαία βάση και οι επανεισπράξεις υπολογίζονται με μέσους όρους. Το όργανο παρακολούθησης της καθημερινής εφαρμογής της Συμφωνίας είναι η Διαχειριστική Επιτροπή, ένα 10μελές όργανο που συνεδριάζει πάντα με την παρουσία νομικού συμβούλου. Οι εκκαθαρίσεις των πληρωμών, ο υπολογισμός των μέσων όρων και η στατιστική παρακολούθηση της Συμφωνίας γίνεται από το Γραφείο Συμψηφισμού, ενώ το ανώτατο όργανο της Συμφωνίας είναι η Γενική Συνέλευση των συμβαλλομένων εταιριών. Τα μέλη της Συμφωνίας καλύπτουν τη συντριπτική πλειοψηφία του ασφαλιζόμενου στόλου αυτοκινήτων οχημάτων στην Ελλάδα.

Η σύντομη περιγραφή του συστήματος άμεσης πληρωμής μας δίνει τη δυνατότητα εξαγωγής σημαντικών συμπερασμάτων για τον τρόπο επηρεασμού τόσο της κατανομής του ύψους αποζημίωσης όσο και της κατανομής του χρόνου αποπληρωμής της ζημιάς. Ως αποτέλεσμα της συμμετοχής μιας ασφαλιστικής εταιρίας στο Σ.Α.Π., στις αποζημιώσεις που καταβάλλονται στους μη υπαίτιους ασφαλισμένους της για λογαριασμό της ασφαλιστικής εταιρίας του υπαίτιου οδηγού,

δεν εμπεριέχεται η πληροφόρηση που εμπεριέχεται στις αποζημιώσεις που δεν διακανονίζονται μέσω Σ.Α.Π. Θα δώσουμε ένα σύντομο παράδειγμα για να γίνουν κατανοητά τα παραπάνω.

Έστω ο οδηγός Α, που είναι ασφαλισμένος στην εταιρία Α και ο οδηγός Β ασφαλισμένος αντίστοιχα στην εταιρία Β. Οι εταιρίες Α και Β δεν συμμετέχουν στο σύστημα άμεσης πληρωμής. Λαμβάνει χώρα ένα ατύχημα με υπαίτιο τον οδηγό Α. Η ασφαλιστική εταιρία Α, του υπαίτιου οδηγού, καλείται να αποζημιώσει τον αναίτιο οδηγό Β. Αφού κινηθούν οι απαραίτητες διαδικασίες από την εταιρία Α, εκτιμάται το ποσό της αποζημίωσης, έστω €1000 και ξεκινά η αποπληρωμή της ζημιάς στον Β. Προφανώς και οι διαδικασίες που λαμβάνουν χώρα απαιτούν χρόνο.

Έστω τώρα ότι οι εμπλεκόμενες εταιρίες Α και Β του παραδείγματος συμμετέχουν στο Σ.Α.Π. Εφόσον, έχει εξαχθεί με απόλυτη ασφάλεια ότι είναι υπεύθυνος ο οδηγός Α για το ατύχημα, ο αναίτιος οδηγός Β δεν αποζημιώνεται από την εταιρία Α, αλλά από την εταιρία στην οποία είναι ο ίδιος ασφαλισμένος, δηλαδή την εταιρία Β. Εφόσον ενεργοποιείται η αποζημίωση της ζημιάς μέσω Σ.Α.Π. η εταιρία Β γνωστοποιεί στην εταιρία Α ότι κινεί τις απαραίτητες διαδικασίες. Η εταιρία Β, λοιπόν, καλείται να καλύψει εξ' ολοκλήρου τον αναίτιο οδηγό της Β, με το ποσό των €1000. Ως αποτέλεσμα επιταχύνεται η διαδικασία διακανονισμού και αποπληρωμής της ζημιάς στον οδηγό Β.

Στην παραπάνω ζημιά που διακανονίζεται μέσω Σ.Α.Π. υπάρχει και ένα ακόμα πολύ σημαντικό αποτέλεσμα. Όπως είπαμε η εταιρία του αναίτιου οδηγού ειδοποιεί την εταιρία του υπαίτιου οδηγού για την αποπληρωμή της ζημιάς. Παρόλο που η εταιρία καταβάλει ποσό €1000 στον ασφαλισμένο, η άλλη εταιρία (του υπαίτιου οδηγού) καταγράφει στο σύστημα της τη ζημιά αυτή, με ποσό ίσο με το μέσο όρο των ζημιών ίδιας κατηγορίας (Πίνακες Υπαιτιότητας – προσυμφωνημένες περιπτώσεις ατυχημάτων), έστω €850 όπως έχει υπολογιστεί από το Γραφείο Συμψηφισμού. Με άλλα λόγια οι ζημιές που διακανονίζονται μέσω Σ.Α.Π. καταγράφονται στο σύστημα της εταιρίας των υπαίτιων οδηγών όχι με το ποσό που πραγματικά "κοστίζουν", αλλά με το ποσό που ισχύει την ημερομηνία που έλαβε χώρα το συμβάν για ατυχήματα της ίδιας σοβαρότητας. Τα ποσά αποζημιώσεων μέσω Σ.Α.Π. υπολογίζονται ως μέσοι όροι αποζημιώσεων ατυχημάτων ίδιας κατηγορίας και αναπροσαρμόζονται σε τακτά χρονικά διαστήματα από το αρμόδιο όργανο.

Η διαπίστωση που έγινε για τις επιδράσεις του Σ.Α.Π. στα στοιχεία αποζημιώσεων μιας εταιρίας είναι πολύ σημαντική για την ανάλυσή μας. Αν υποθέσουμε ότι η εταιρία Α είναι η εταιρία το χαρτοφυλάκιο της οποίας χρησιμοποιούμε στην παρούσα εργασία, τότε για τις ζημιές των ασφαλισμένων της που διακανονίστηκαν μέσω Σ.Α.Π. τα ποσά στα στοιχεία που μελετούμε δεν αντανακλούν το πραγματικό κόστος της κάθε ζημιάς. Σύμφωνα με το παράδειγμα που δόθηκε παραπάνω, η ζημιά "κόστισε" στην εταιρία Β (αναίτιου οδηγού) €1000 και στα στοιχεία της εταιρίας Α (υπαίτιου οδηγού) εμφανίζεται με κόστος €850.

Ο τρόπος με τον οποίο διακανονίζονται οι ζημιές μέσω Σ.Α.Π. οδηγεί σε αυτό το αποτέλεσμα. Σκοπός του Σ.Α.Π. και της καθιέρωσης του είναι η αποζημίωση για το 100% της ζημιάς (στα όρια που ορίζονται από το νόμο) του αναίτιου οδηγού από την ασφαλιστική του εταιρία. Η εταιρία του υπαίτιου οδηγού ενημερώνεται είτε από τον ασφαλισμένο της είτε από την άλλη εταιρία για τον διακανονισμό μέσω Σ.Α.Π. και περνά στο σύστημα το ποσό που ορίζεται από το Γραφείο Συμψηφισμού. Η διαδικασία βέβαια δεν σταματά εκεί. Όλες οι εταιρίες που συμμετέχουν στο σύστημα, υποχρεούνται να υποβάλλουν στο Γραφείο Συμψηφισμού όλα τα απαραίτητα αποδεικτικά για τις ζημιές που διακανονίστηκαν μέσω Σ.Α.Π. Έτσι είναι εφικτό να γίνουν οι εκκαθαρίσεις των πληρωμών μεταξύ των ασφαλιστικών εταιριών που συμμετέχουν. Στο παράδειγμά μας το Γραφείο Συμψηφισμού ενημερώνεται για το ατύχημα μεταξύ των Α και Β και για το ποσό των €1000 που καταβλήθηκαν από την εταιρία Β στον οδηγό Β. Ο μέσος όρος για ατύχημα της ίδιας κατηγορίας έχει οριστεί στα €850. Η διαφορά που προκύπτει μεταξύ των δύο ποσών αποζημιώσεων, ήτοι €150, είναι το ποσό εκκαθάρισης μεταξύ των δύο εταιριών αν υποθέσουμε ότι αυτό είναι το μοναδικό ατύχημα που διακανονίστηκε μέσω Σ.Α.Π. για τις δύο εμπλεκόμενες εταιρίες. Το ποσό των €150 θα χρεωθεί στο λογαριασμό που διατηρεί το Γραφείο Συμψηφισμού για την εταιρία Β και αντίστοιχα θα πιστωθεί στο λογαριασμό της εταιρίας Α. Τέλος, η ενημέρωση του Γραφείο Συμψηφισμού για τα πραγματικά κόστη των ατυχημάτων είναι απαραίτητη για τον επαναπροσδιορισμό των μέσων όρων των κατηγοριών ζημιών σύμφωνα με την κείμενη νομοθεσία.

5.2 Εύρεση μίας νέας κατανομής για το ύψος αποζημιώσεων

Στην παρούσα διπλωματική εργασία έγινε προσπάθεια διερεύνησης της κατανομής που ακολουθούν οι αποζημιώσεις υλικών και σωματικών βλαβών. Από την ανάλυση που πραγματοποιήσαμε στα κεφάλαια 2 και 3 διαπιστώσαμε πως τα ύψη αποζημιώσεων προέρχονται από κατανομή που εμφανίζει τουλάχιστον 2 ή και περισσότερες κορυφές. Λαμβάνοντας υπόψη μας την πληροφορία ότι η εταιρία συμμετέχει στο Σ.Α.Π. μπορούμε ίσως να δώσουμε μία εξήγηση για τη μεγάλη συγκέντρωση σε κάποιες τιμές αποζημιώσεων και κατ' επέκταση την εμφάνιση περισσότερων της μίας κορυφής στην κατανομή του ύψους αποζημιώσεων.

Όπως αναφέρουμε παραπάνω για τις πληρωμές μέσω Σ.Α.Π. η εταιρία του υπαίτιου οδηγού παρόλο που δεν αποζημιώνει άμεσα τον αναίτιο οδηγό, πρέπει να καταγράφει στο σύστημα της τη ζημιά. Προφανώς και η καταγραφή των ζημιών που δεν διακανονίζονται άμεσα από την εταιρία είναι απαραίτητη για διαφόρους λόγους, όπως λογιστικούς, οικονομικές καταστάσεις, εκτίμηση μελλοντικών ζημιών, εκτίμηση των τεχνικών προβλέψεων κ.τ.λ. Η απεικόνιση αυτή του ποσού αποζημίωσης, όμως, ισούται με το μέσο όρο των αποζημιώσεων για ζημιές της ίδιας κατηγορίας και δεν αντικατοπτρίζει το πραγματικό ποσό πληρωμής και κατ' επέκταση κόστος της ζημιάς. Αν λάβουμε υπόψη μας πως μέσω Σ.Α.Π. διακανονίζονται χιλιάδες ζημιές, αυτές θα

εμφανίζονται στα δεδομένα αποζημιώσεων του χαρτοφυλακίου της εταιρίας των υπαίτιων οδηγών με ίδιο κόστος και θα δημιουργούν συγκεντρώσεις σε σημεία που αντιστοιχούν στους μέσους όρους των ζημιών όπως υπολογίζονται από το Γραφείο Συμψηφισμού.

Δεδομένου ότι δεν υπάρχει πληροφόρηση για το ποια είναι αυτά τα ποσά αποζημιώσεων, προχωρούμε στην εύρεση μιας κατανομής που θα απεικονίζει τη συμπεριφορά του ύψους αποζημιώσεων υλικών ζημιών και σωματικών βλαβών όπως αυτή εξάγεται από τα δεδομένα πληρωμών των κλειστών φακέλων της εταιρίας. Σκοπός μας είναι η εύρεση μιας κατανομής με την οποία ο υπολογισμός της συχνότητας εμφάνισης διαφόρων ποσών αποζημιώσεων θα μπορεί να προσδιοριστεί εύκολα και γρήγορα.

5.2.1 Κατασκευή της "Πολυγραμμικής κατανομής"

Στην παρούσα υποενότητα επιχειρείται η εύρεση και κατασκευή μίας νέας κατανομής που θα καταφέρει να περιγράψει τη συμπεριφορά της μεταβλητής του ύψους των αποζημιώσεων για τις δύο κατηγορίες ζημιών. Ο προσδιορισμός της νέας κατανομής θα γίνει με τη μέθοδο της προσέγγισης της καμπύλης της σ.π.π. των δεδομένων από διαδοχικές ευθείες γραμμές. Σκοπός είναι να δημιουργήσουμε μία σειρά από ευθείες, της μορφής $y = ax + b$, που θα προσεγγίζουν όσο το δυνατόν την καμπύλη της σ.π.π. της κατανομής που ακολουθούν οι μεταβλητές του ύψους των αποζημιώσεων.

Η ιδέα αφορά στην κατασκευή μιας "πολυγραμμικής" κατανομής (Multilinear distibution) που θα αναπαριστά όσο το δυνατόν την κατανομή που ακολουθούν τα δεδομένα. Το εγχείρημα ξεκινά από κάτι που είναι γνωστό ως "Piecewise distributions" και έχει εξεταστεί από πολλούς ερευνητές (βλέπε για παράδειγμα Klugman *et al.*, 2004) και αφορά σε κατανομές που αναπαρίστανται από πολλές υποκατανομές (τμηματικές κατανομές). Η καθεμία από αυτές τις υποκατανομές ερμηνεύει τη συμπεριφορά της μεταβλητής σε συγκεκριμένο τμήμα των τιμών που λαμβάνει.

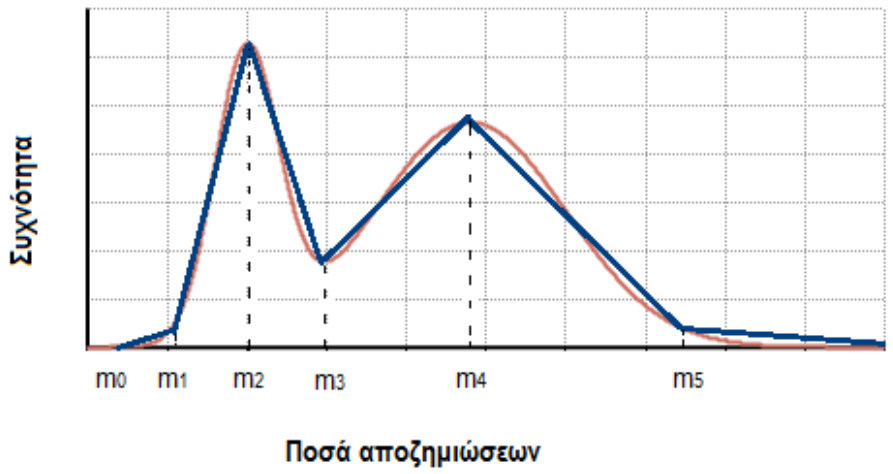
Η γενική μορφή μιας "πολυγραμμικής" σ.π.π. θα είναι:

$$f(x) = \begin{cases} \alpha_1 x + b_1\,, & x \in (m_0, m_1) \\ & \dots \\ \alpha_k x + b_k\,, & x \in (m_{k-1}, m_k) \end{cases}$$

όπου α_i οι συντελεστές διεύθυνσης των ευθειών, b_i οι σταθερές για $x = 0$ και m_i τα σημεία αλλαγής διεύθυνσης της ευθείας.

Για να γίνει κατανοητός ο τρόπος με τον οποίο θα κατασκευάσουμε την πολυγραμμική συνάρτηση κατανομής, δίνεται το παρακάτω σχήμα, όπου με κόκκινο χρώμα εμφανίζεται η καμπύλη της κατανομής των δεδομένων και με μπλε χρώμα οι διαδοχικές ευθείες που προσεγγίζουν την καμπύλη αυτή. Ο οριζόντιος άξονας αφορά στις τιμές των δεδομένων αποζημιώσεων και ο κάθετος άξονας στη συχνότητα

εμφάνισης των τιμών αυτών. Τα σημεία m_i που εμφανίζονται στον οριζόντιο άξονα είναι αυτά στα οποία γίνεται αλλαγή κλίσης της πολυγωνικής γραμμής.



Σχήμα 5.1: Πολυγραμμικό μοντέλο.

Όπως φαίνεται στο σχήμα 5.1, η σ.π.π. του ύψους των αποζημιώσεων προσεγγίζεται από 6 ευθείες γραμμές. Η πρώτη ευθεία αναπαριστά τη σ.π.π. για τιμές αποζημιώσεων στο διάστημα (m_0, m_1), η δεύτερη στο διάστημα (m_1, m_2) κ.τ.λ. Σε κάθε διάστημα (m_{i-1}, m_i) του πεδίου ορισμού, που αφορά στο σύνολο των παρατηρούμενων τιμών αποζημιώσεων, θα έχουμε την αντίστοιχη σ.π.π. $f_i(x) = y_i = a_i x + b_i$ του πολυγραμμικού μοντέλου.

Έχοντας κατανοήσει τη μορφή του πολυγραμμικού μοντέλου και τον τρόπο εφαρμογής του, το επόμενο βήμα είναι να μπορέσουμε να υπολογίσουμε τις τιμές των ζευγών $(a_i\,, b_i)$ που ορίζουν τις διαδοχικές ευθείες σε κάθε διάστημα $(m_{i-1}\,, m_i)$.

Για να κατασκευάσουμε τη συνάρτηση κατανομής του μοντέλου σε κάθε περίπτωση χρησιμοποιήσουμε τις παρατηρούμενες αθροιστικές συχνότητες C_i στα σημεία αλλαγής διεύθυνσης m_i, έτσι ώστε να ισχύει:

$$C_i - C_{i-1} = \int_{m_{i-1}}^{m_i} (a_i x + b_i)\ dx$$

Χρησιμοποιώντας την παραπάνω σχέση, υπολογίζουμε τις τιμές των $(a_i\,, b_i)$, ως εξής:

$$C_i - C_{i-1} = \int_{m_{i-1}}^{m_i} (a_i x + b_i)\ dx \ => \ C_i = \int_{m_{i-1}}^{m_i} (a_i x + b_i)\ dx + C_{i-1} \ =>$$

$$C_i = \int_{m_{i-1}}^{m_i} \alpha_i x \ dx + \int_{m_{i-1}}^{m_i} b_i \ dx + C_{i-1} \ => \ C_i = \alpha_i \int_{m_{i-1}}^{m_i} x \ dx + \int_{m_{i-1}}^{m_i} b_i \ dx + C_{i-1}$$

$$=> C_i = C_{i-1} + \ \alpha_i \frac{m_i^2 - m_{i-1}^2}{2} + b_i(m_i - m_{i-1}) =>$$

$$=> \ C_i = C_{i-1} + \ 0.5 \ \alpha_i \left(m_i^2 - m_{i-1}^2\right) + b_i(m_i - m_{i-1}) \quad (1)$$

$$(1) => \ \alpha_i = \frac{2 \ (C_i - C_{i-1}) - b_i(m_i - m_{i-1})}{\left(m_i^2 - m_{i-1}^2\right)} \quad (2)$$

Στα σημεία m_i μπορεί να αλλάζει το α, δηλαδή η κλίση της ευθείας, όμως το b δηλαδή το σημείο στο οποίο η προέκταση της ευθείας τέμνει τον κάθετο άξονα παραμένει σταθερό. Επομένως θα έχουμε:

$$\alpha_i \, m_{i-1} + \ b_i = \ \alpha_{i-1} \, m_{i-1} + \ b_{i-1} \ => \ b_i = \ \alpha_{i-1} \, m_{i-1} - \ \alpha_i \, m_{i-1} + b_{i-1} \ =>$$

$$=> \ b_i = \ (\alpha_{i-1} - \alpha_i) \, m_{i-1} + b_{i-1} \quad (3)$$

$$(2) => \ \alpha_i = \frac{2 \ (C_i - C_{i-1})}{(m_i^2 - m_{i-1}^2)} - \frac{2 \ b_i(m_i - m_{i-1})}{(m_i^2 - m_{i-1}^2)} = \frac{2 \ (C_i - C_{i-1})}{(m_i^2 - m_{i-1}^2)} - \frac{2 \ b_i(m_i - m_{i-1})}{(m_i - m_{i-1}) \ (m_i + m_{i-1})}$$

$$=> \ \alpha_i = \frac{2 \ (C_i - C_{i-1})}{\left(m_i^2 - m_{i-1}^2\right)} - \frac{2 \ b_i}{(m_i + m_{i-1})} \quad (4)$$

$$(3), (4) => \alpha_i = \frac{2 \ (C_i - C_{i-1})}{\left(m_i^2 - m_{i-1}^2\right)} - \frac{2 \ ((\alpha_{i-1} - \alpha_i) \, m_{i-1} + b_{i-1})}{(m_i + m_{i-1})} =$$

$$= \frac{2 \ (C_i - C_{i-1})}{\left(m_i^2 - m_{i-1}^2\right)} - \frac{2 \ (\alpha_{i-1} - \alpha_i) \, m_{i-1}}{(m_i + m_{i-1})} - \frac{2 \ b_{i-1}}{(m_i + m_{i-1})} =$$

$$= \frac{2 \ (C_i - C_{i-1})}{\left(m_i^2 - m_{i-1}^2\right)} - \frac{2 \ a_{i-1} m_{i-1}}{(m_i + m_{i-1})} + \frac{2 \ a_i m_{i-1}}{(m_i + m_{i-1})} - \frac{2 \ b_{i-1}}{(m_i + m_{i-1})} =>$$

$$=> \ \alpha_i - \frac{2 \ a_i m_{i-1}}{(m_i + m_{i-1})} = \frac{2 \ (C_i - C_{i-1})}{\left(m_i^2 - m_{i-1}^2\right)} - \frac{2 \ a_{i-1} m_{i-1}}{(m_i + m_{i-1})} - \frac{2 \ b_{i-1}}{(m_i + m_{i-1})} =>$$

$$=> \alpha_i \ (m_i + m_{i-1}) - 2 \ a_i m_{i-1} = \frac{2 \ (C_i - C_{i-1})}{(m_i - m_{i-1})} - 2 \ a_{i-1} m_{i-1} - 2 \ b_{i-1} =>$$

$$=> \alpha_i \ (m_i + m_{i-1} - 2m_{i-1}) = \frac{2\,(C_i - C_{i-1})}{(m_i - m_{i-1})} - 2\,a_{i-1}m_{i-1} - 2\,b_{i-1} =>$$

$$=> \ \alpha_i \ (m_i - m_{i-1}) = \frac{2\,(C_i - C_{i-1})}{(m_i - m_{i-1})} - \ 2\,(a_{i-1}m_{i-1} + b_{i-1}) =>$$

$$=> \alpha_i \ \frac{(m_i - m_{i-1})}{(m_i - m_{i-1})} = \frac{2\,(C_i - C_{i-1})}{(m_i - m_{i-1})}\frac{1}{(m_i - m_{i-1})} - 2\,(a_{i-1}m_{i-1} + b_{i-1})\frac{1}{(m_i - m_{i-1})}$$

$$=> \ \alpha_i = \frac{2\,(C_i - C_{i-1})}{(m_i - m_{i-1})^2} - \ 2\ \frac{(a_{i-1}m_{i-1} + b_{i-1})}{(m_i - m_{i-1})} =>$$

$$=> \alpha_i = 2\left(\frac{C_i - C_{i-1}}{(m_i - m_{i-1})^2} - \frac{a_{i-1}m_{i-1} + b_{i-1}}{(m_i - m_{i-1})}\right) \ (5)$$

Οι τιμές των α_i, b_i για κάθε διάστημα $(m_{i-1}\ ,\ m_i)$ υπολογίζονται μέσω των σχέσεων (3),(5) δηλαδή:

$$\alpha_i = 2\left(\frac{C_i - C_{i-1}}{(m_i - m_{i-1})^2} - \frac{a_{i-1}m_{i-1} + b_{i-1}}{(m_i - m_{i-1})}\right)$$

$$b_i = \ (\alpha_{i-1} - \alpha_i)\, m_{i-1} + b_{i-1}$$

5.2.2 Πολυγραμμική κατανομή του ύψους των αποζημιώσεων υλικών ζημιών

Στην υποενότητα 5.2.2 παρουσιάζουμε αναλυτικά, τον τρόπο με τον οποίο κατασκευάζουμε την πολυγραμμική κατανομή για το ύψος αποζημιώσεων υλικών ζημιών με τη χρήση των δεδομένων του χαρτοφυλακίου ζημιών της εταιρίας.

Το πρώτο βήμα αποτελεί ο διαχωρισμός των ποσών αποζημιώσεων υλικών ζημιών σε διαστήματα. Το πλάτος του κάθε διαστήματος επιλέχθηκε να είναι ίσο με €100. Ο λόγος είναι ότι σύμφωνα με τα αρχικά δεδομένα που αναλύθηκαν στο κεφάλαιο 2, το 95% των τιμών των αποζημιώσεων αυτής της κατηγορίας παίρνει τιμές μικρότερες ή ίσες των €4.700. Επομένως, κρίνεται σκόπιμο να εστιάσουμε σε τόσο μικρά διαστήματα τιμών ώστε να εκμεταλλευτούμε όλη τη διαθέσιμη πληροφόρηση. Μέχρι την τιμή των €20.000 (τιμή που αντανακλά το 99,56% περίπου των παρατηρούμενων τιμών αποζημιώσεων) κατηγοριοποιήθηκαν οι αποζημιώσεις σε διαστήματα πλάτους €100. Οι υπόλοιπες παρατηρήσεις του δείγματος περιλαμβάνονται στο τελευταίο διάστημα για ποσά αποζημιώσεων μεγαλύτερα των €20.000.

Ο συνολικός αριθμός αποζημιώσεων υλικών ζημιών είναι $n = 102.454$. Ως αποτέλεσμα του παραπάνω διαχωρισμού οι αρχικές τιμές αποζημιώσεων ομαδοποιούνται σε k=201 ομάδες. Για τις ομάδες αυτές, υπολογίζουμε την παρατηρούμενη συχνότητα αποζημιώσεων ν_i σε κάθε διάστημα τιμών, έτσι ώστε $\sum_{i=1}^{k} \nu_i = n = 102.454$. Επιπλέον, αθροίζοντας διαδοχικά τις συχνότητες κάθε ομάδας έχουμε την αθροιστική συχνότητα N_i των αποζημιώσεων στο άνω άκρο κάθε διαστήματος. Διαιρώντας τις παρατηρούμενες συχνότητες σε κάθε διάστημα τιμών με το σύνολο $n = 102.454$, υπολογίζουμε τη σχετική συχνότητα f_i και αντίστοιχα την αθροιστική σχετική συχνότητα F_i στο άνω άκρο κάθε διαστήματος.

Έχοντας ολοκληρώσει τους παραπάνω υπολογισμούς και κάνοντας χρήση των σχέσεων (3) και (5) της υποενότητας 5.2.1, μπορούμε να προσδιορίσουμε τις τιμές των συντελεστών α_i και b_i των ευθειών που θα προσεγγίζουν την καμπύλη της σ.π.π. της κατανομής των αποζημιώσεων υλικών ζημιών.

Στο σημείο αυτό σημειώνουμε ότι πρέπει να αποφασιστούν τα βέλτιστα σημεία m_i, στα οποία θα γίνεται αλλαγή του συντελεστή διεύθυνσης α, που ορίζει τις διαδοχικές ευθείες που συνθέτουν το πολυγραμμικό μοντέλο. Σκοπός μας δεν είναι να προσεγγίσουμε την καμπύλη της σ.π.π. των αποζημιώσεων ανά €100, δηλαδή για κάθε ένα από τα 201 διαστήματα τιμών να κατασκευάζουμε και μία διαφορετική ευθεία. Αντίθετα, βρίσκουμε σκόπιμο να προσεγγίσουμε την κατανομή των αποζημιώσεων με περίπου 10-12 διαδοχικές ευθείες, ώστε το μοντέλο να είναι χρήσιμο και εύκολο προς χρήση.

Η επιλογή των σημείων m_i αλλαγής κλίσης της πολυγωνικής γραμμής, πραγματοποιήθηκε λαμβάνοντας υπόψη τα παρακάτω:

- Σε κανένα διάστημα τιμών η πολυγραμμική κατανομή δεν μπορεί να παίρνει αρνητικές τιμές.
- Οι ευθείες θέλουμε να αντανακλούν όσο το δυνατόν τη συμπεριφορά του ύψους των αποζημιώσεων, αλλά ταυτόχρονα να είναι όσο το δυνατόν λιγότερες σε αριθμό.
- Η βέλτιστη επιλογή μεταξύ των σημείων που θα μπορούσαν να επιλεγούν ως σημεία αλλαγής διεύθυνσης, έγινε λαμβάνοντας υπόψη τις απόλυτες διαφορές που προέκυπταν ανάμεσα στις παρατηρούμενες σε κάθε ένα από τα k διαστήματα σχετικές συχνότητες f_i και στις σχετικές συχνότητες που υπολογίζονται από το πολυγραμμικό μοντέλο.

Στον πίνακα που ακολουθεί δίνονται οι συντελεστές των ευθειών που συνθέτουν το πολυγραμμικό μοντέλο για το ύψος αποζημιώσεων υλικών ζημιών.

Πίνακας 5.1: Συντελεστές ευθειών που συνθέτουν το πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων υλικών ζημιών.

Multilinear model - "Total payment Accidental Damage"			
mi	**CDF (mi)**	**α**	**b**
0	0.00%	0.00E+00	0.00E+00
200	2.24%	1.12E-06	0.00E+00
300	9.99%	1.10E-05	-1.98E-03
500	27.83%	-4.35E-06	2.63E-03
700	34.17%	-1.40E-06	1.16E-03
900	45.27%	3.79E-06	-2.47E-03
1600	81.43%	-1.19E-06	2.01E-03
2200	85.70%	-9.39E-08	2.50E-04
2400	86.67%	5.35E-08	-7.46E-05
6000	96.44%	-1.48E-08	8.93E-05
143500	100.00%	-3.76E-12	5.40E-07

Η πρώτη στήλη του πίνακα αφορά στα σημεία αλλαγής διεύθυνσης m_i, η δεύτερη στην αθροιστική συνάρτηση κατανομής (Cumulative Distribution Function) του πολυγραμμικού μοντέλου στα σημεία m_i και οι δύο τελευταίες στους συντελεστές (α_i, b_i) που προσδιορίζουν τις ευθείες στα διαστήματα τιμών (m_{i-1}, m_i).

Για να γίνει κατανοητός ο τρόπος με τον οποίο εργαστήκαμε για να εξάγουμε τα αποτελέσματα του πίνακα 5.1, περιγράφουμε αναλυτικά τη διαδικασία που ακολουθήσαμε για τον υπολογισμό των συντελεστών των πρώτων δύο ευθειών του πολυγραμμικού μοντέλου.

Για να είναι δυνατή η προσέγγιση της σ.π.π. των αποζημιώσεων από τις διαδοχικές ευθείες, επιλέγουμε ως σημεία m_i άνω άκρα από τα k διαστήματα που χωρίσαμε τα δεδομένα αποζημιώσεων. Στα σημεία αυτά θέτουμε την αθροιστική συχνότητα του πολυγραμμικού μοντέλου ίση με την α.σ.π.π. όπως υπολογίστηκε από τα δεδομένα.

Παρατηρούμε από τον πίνακα 5.1 ότι για ποσά αποζημιώσεων ίσα με το 0, έχουμε συχνότητα ίση με το 0, αφού στο αρχείο αποζημιώσεων της εταιρίας δεν εμφανίζονται μηδενικές αποζημιώσεις. Επομένως η α.σ.π.π. στο 0 ισούται με 0.

Προχωρούμε να υπολογίσουμε τους συντελεστές της πρώτης ευθείας α_1και b_1. Το σημείο αλλαγής διεύθυνσης m_1 ορίζεται στα €200. Έχουμε υπολογίσει από τα δεδομένα ότι η F(200) ισούται με 2,24% και F(0) με 0. Επομένως:

$$C(200) - C(0) = \int_0^{200} (a_1 x + b_1)\ dx$$

Κάνοντας χρήση της σχέσης (3) της υποενότητας 5.2.1 έχουμε:

$$\alpha_1 = 2\left(\frac{C_1 - C_0}{(m_1 - m_0)^2} - \frac{a_0 m_0 + b_0}{(m_1 - m_0)}\right)$$

Αντικαθιστώντας στην παραπάνω σχέση όπου $C_1 = C(200) = F(200) = 2{,}24\%$, $C_0 = C(0) = F(0) = 0$, $m_1 = 200$, $m_0 = 0$, $a_0 = b_0 = 0$ λαμβάνουμε την τιμή του συντελεστή διεύθυνσης α_1 ίσο με 1,12 E-06.

Επιπλέον, από τη σχέση (5) της υποενότητας 5.2.1 έχουμε:

$$b_1 = (\alpha_0 - \alpha_1)\, m_0 + b_0$$

Αντικαθιστώντας στη σχέση όπου $m_0 = 0$, $a_0 = b_0 = 0$ λαμβάνουμε την τιμή του συντελεστή b_1 ίσο με 0.

Αντίστοιχα, εργαζόμαστε για τον προσδιορισμό της δεύτερης ευθείας για το διάστημα αποζημιώσεων €200 - €300.

$$C(300) - C(200) = \int_{200}^{300} (a_2 x + b_2)\ dx$$

Χρησιμοποιούμε και πάλι τις σχέσεις (3) και (5) της υποενότητας 5.2.1 και έχουμε να επιλύσουμε το παρακάτω σύστημα εξισώσεων:

$$\alpha_2 = 2\left(\frac{C_2 - C_1}{(m_2 - m_1)^2} - \frac{a_1 m_1 + b_1}{(m_2 - m_1)}\right)$$

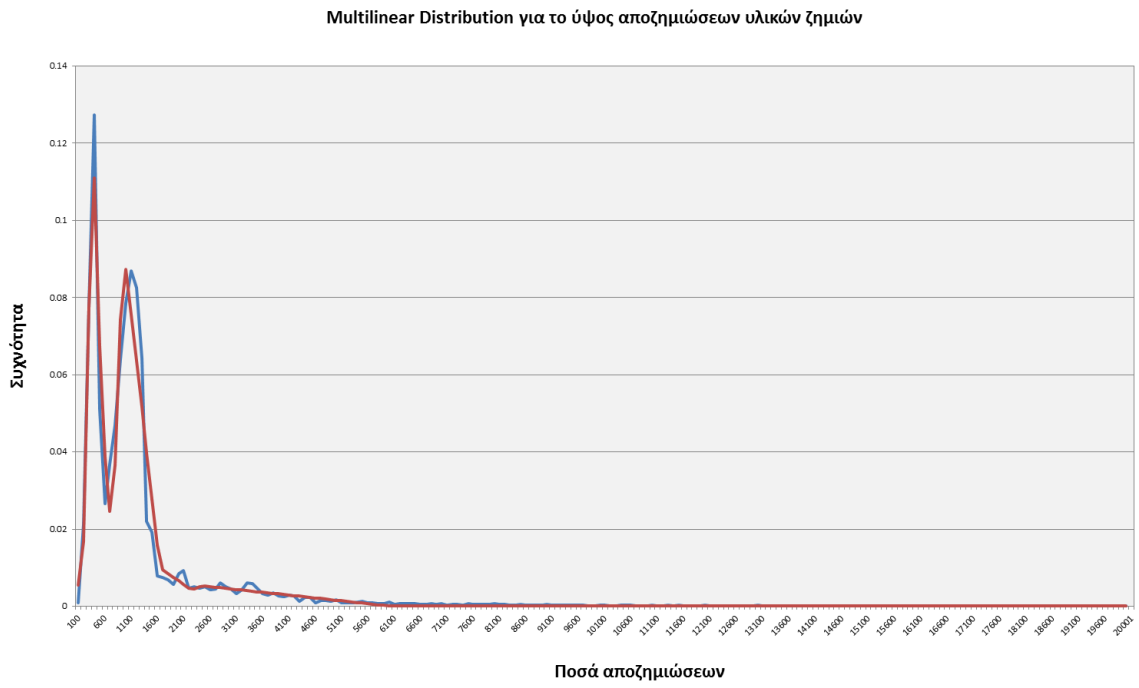
$$b_2 = (\alpha_1 - \alpha_2)\, m_1 + b_1$$

Αντικαθιστώντας στις παραπάνω σχέσεις όπου $C_2 = C(300) = F(300) = 9{,}99\%$, $C_1 = C(200) = F(200) = 2{,}24\%$, $m_2 = 300$, $m_1 = 200$, $a_1 = 1{,}12\ \mathrm{E} - 06$, $b_1 = 0$, λαμβάνουμε τις τιμές των συντελεστών της δεύτερης ευθείας που ισούνται με $a_2 = 1.10\ \mathrm{E} - 05$ και $b_2 = -1.98\ \mathrm{E} - 03$, όπως φαίνονται στον πίνακα 5.1.

Ως αποτέλεσμα της χρήσης των δύο σχέσεων της υποενότητας 5.2.1, υπολογίσαμε τους συντελεστές της 1ης και 2ης ευθείας που συνθέτουν το πολυγραμμικό μοντέλο. Ακολουθώντας την ίδια διαδικασία για τα υπόλοιπα σημεία m_i, οδηγούμαστε στη δημιουργία της πολυγραμμικής κατανομής. Ο τρόπος κατασκευής της κατανομής αυτής, όπως περιγράφεται, αντανακλά την πληροφόρηση που υπάρχει στις τιμές των αποζημιώσεων.

Στο σημείο αυτό σημειώνουμε ότι για την επιλογή των "κατάλληλων" σημείων αλλαγής κλίσης χρησιμοποιήθηκαν και δύο διαγράμματα, που τουλάχιστον οπτικά, έδιναν μία αίσθηση του κατά πόσο το νέο μοντέλο προσεγγίζει την σ.π.π. που ακολουθούν τα δεδομένα αποζημιώσεων και κατ' επέκταση βοήθησαν στην επιλογή των σημείων στα οποία είναι προτιμότερη η αλλαγή της κλίσης της πολυγωνικής γραμμής.

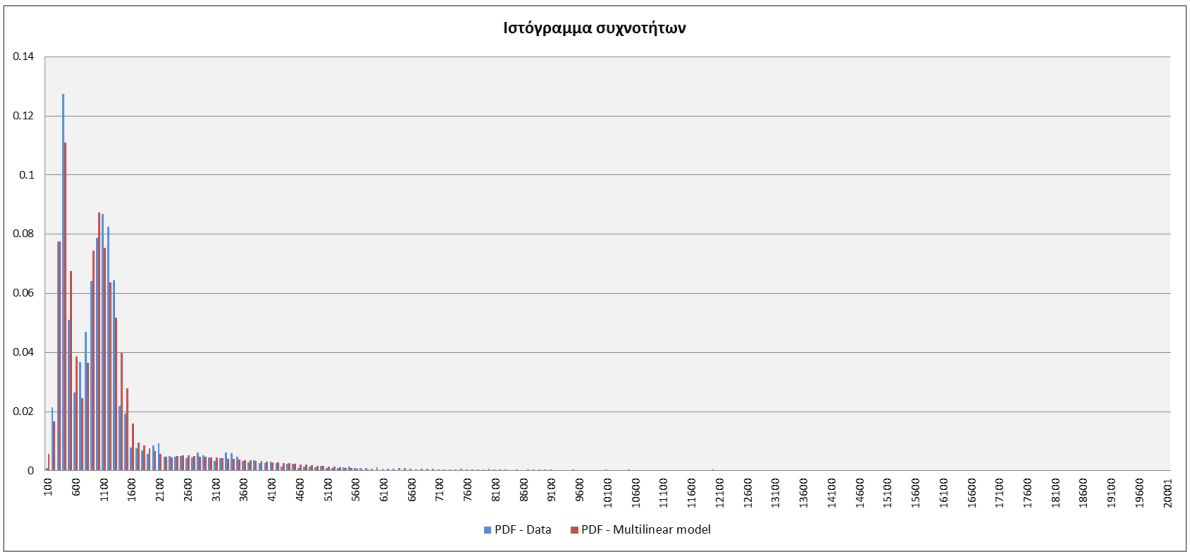
Στο σχήμα 5.2 που δίνεται παρακάτω, παρατηρούμε δύο γραμμές διαφορετικού χρώματος. Η μπλε γραμμή εξήχθη ως αποτέλεσμα της ένωσης των άνω άκρων των k=201 ομάδων, πλάτους €100, που χωρίσαμε τα αρχικά δεδομένα αποζημιώσεων υλικών ζημιών. Πρόκειται ουσιαστικά για μία πολυγωνική γραμμή, η οποία αναπαριστά ακριβώς τη μορφή της συνάρτησης κατανομής που θα είχε το πολυγραμμικό μοντέλο αν επιλεγόταν κάθε φορά μία διαφορετική ευθεία για να περιγράψει τη συμπεριφορά των δεδομένων αποζημιώσεων σε κάθε μία από τις k ομάδες. Αποτέλεσμα αυτής της πρακτικής θα ήταν η δημιουργία 201 διαφορετικών ευθειών για να προσεγγίσουμε την καμπύλη της σ.π.π. του ύψους των αποζημιώσεων. Όπως αναφέρουμε και παραπάνω δεν θεωρούμε σκόπιμη τη δημιουργία πολυγραμμικού μοντέλου αυτής της μορφής.



Σχήμα 5.2: Πολυγωνική γραμμή και πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων υλικών ζημιών.

Η κόκκινη γραμμή αναπαριστά το πολυγραμμικό μοντέλο που επιλέξαμε για να περιγράψει τη συμπεριφορά της μεταβλητής "Total payment" υλικών ζημιών. Οι συνιστώσες του μοντέλου, δηλαδή οι διαδοχικές ευθείες που το συνθέτουν σε κάθε διάστημα (m_{i-1}, m_i), προσδιορίζονται από τους συντελεστές του πίνακα 5.1. Παρατηρούμε πως υπάρχουν κάποιες αποκλίσεις μεταξύ των δύο γραμμών, που αντικατοπτρίζουν διαφορές ανάμεσα στις παρατηρούμενες συχνότητες και σε αυτές που υπολογίζονται από το μοντέλο. Παρά τις αποκλίσεις αυτές, με τον τρόπο που κατασκευάσαμε την πολυγραμμική κατανομή κατορθώσαμε να προσεγγίσουμε αρκετά καλά τις δύο πρώτες κορυφές που εμφανίζουν τα δεδομένα αποζημιώσεων υλικών ζημιών. Τα αποτελέσματα της μελέτης που πραγματοποιήσαμε στο κεφάλαιο 3 επιβεβαιώνονται, καθώς παρατηρούμε για τις αποζημιώσεις υλικών ζημιών περισσότερες των τριών κορυφές στην κατανομή του ύψους τους. Επιπλέον, από το διάγραμμα παρατηρούμε τη μεγάλη ουρά στα δεδομένα αποζημιώσεων. Μάλιστα, οι δύο γραμμές φαίνονται να εφάπτονται με τον οριζόντιο άξονα στην τιμή αποζημίωσης ίση με €7.500 και για τιμές μεγαλύτερες από αυτή.

Το δεύτερο διάγραμμα που χρησιμοποιήσαμε για να ελέγξουμε την προσέγγιση της κατανομής που ακολουθούν τα δεδομένα αποζημιώσεων υλικών ζημιών από το νέο μοντέλο, είναι ένα συνδυαστικό ιστόγραμμα συχνοτήτων. Συνδυαστικό υπό την έννοια ότι μας επιτρέπει να βλέπουμε ταυτόχρονα δύο ιστογράμματα μαζί σε ένα γράφημα. Πιο συγκεκριμένα για κάθε μία από τις k=201 ομάδες, αντιστοιχίζεται μία κλάση για τη σχετική συχνότητα όπως υπολογίστηκε από τα αρχικά δεδομένα (μπλε στήλες) και μία δεύτερη κλάση για τη σχετική συχνότητα όπως υπολογίστηκε από το πολυγραμμικό μοντέλο (κόκκινες στήλες).



Σχήμα 5.3:Ιστόγραμμα συχνοτήτων της σ.π.π. και της πολυγραμμικής κατανομής του ύψους των αποζημιώσεων υλικών ζημιών.

Με τις κλάσεις να παρουσιάζονται με διαφορετικό χρόνο η μία δίπλα στην άλλη, μπορούμε εύκολα να ελέγχουμε κατά πόσο απέχουν οι εκτιμώμενες σύμφωνα με το πολυγραμμικό μοντέλο σχετικές συχνότητες, με τις αντίστοιχες που υπολογίστηκαν από τα δεδομένα. Παρατηρώντας και το ιστόγραμμα συχνοτήτων επιβεβαιώνουμε την ύπαρξη αποκλίσεων μεταξύ των συχνοτήτων. Αν παρατηρήσουμε προσεκτικά τα γραφήματα 5.2 και 5.3 θα διαπιστώσουμε πως οι μεγαλύτερες αποκλίσεις παρουσιάζονται στις κλάσεις που αφορούν σε τιμές αποζημιώσεων €300 έως €400, €700 έως €800 και €1100 έως €1200. Το πολυγραμμικό μοντέλο μοιάζει να υποεκτιμά τη συχνότητα εμφάνισης αποζημιώσεων σε αυτά τα διαστήματα τιμών, όπως και γύρω από την τιμή των €1800 που εμφανίζεται μία ακόμα κορυφή. Αντίθετα, υπερεκτιμά την εμφάνιση αποζημιώσεων στα διαστήματα €500 έως €600, €900 έως €1000 και €1400 έως €1500. Οι διαφορές αυτές είναι της τάξης του 1% και είναι αναμενόμενες από τη στιγμή που το μοντέλο κατασκευάστηκε με σκοπό να προσομοιώνει τη συμπεριφορά των αποζημιώσεων υλικών ζημιών με 10 ευθείες περίπου. Καλύτερη προσέγγιση θα μπορούσε να επιτευχθεί με την αύξηση του αριθμού των ευθειών, αυξάνοντας ταυτόχρονα την πολυπλοκότητα στην εφαρμογή του. Δεν μπορούμε να μην παρατηρήσουμε και από τη μορφή του ιστογράμματος την έντονη δεξιά ουρά της κατανομής των αποζημιώσεων.

Στο σημείο αυτό πρέπει να γίνει μία ακόμα σημαντική παρατήρηση, που αφορά στα αποτελέσματα του πίνακα 5.1. Αν κοιτάξουμε την τελευταία γραμμή του πίνακα, παρατηρούμε πως το τελευταίο σημείο αλλαγής διεύθυνσης της πολυγωνικής γραμμής του μοντέλου έχει υπολογιστεί στην τιμή €143.500. Ο προσδιορισμός αυτού του σημείου πραγματοποιήθηκε λαμβάνοντας υπόψη την ιδιότητα μιας συνεχούς τυχαίας μεταβλητής, ότι η αθροιστική σχετική συχνότητα για το σύνολο των τιμών που λαμβάνει η τ.μ. ισούται με τη μονάδα (100%). Αφού το πολυγραμμικό μοντέλο

αναπαρίσταται από μία πολυγωνική γραμμή, η τελευταία ευθεία που το συνθέτει θα δημιουργεί με τον οριζόντιο άξονα ένα τρίγωνο. Το εμβαδό του τριγώνου αυτού θα ισούται με τη συχνότητα αποζημιώσεων που λαμβάνουν τιμές στο διάστημα (m_{i-1}, m_i).

Επομένως, έχοντας προσδιορίσει το προτελευταίο σημείο αλλαγής κλίσης, υπολογίζουμε το σημείο m_i στο οποίο η πολυγωνική γραμμή συναντά τον οριζόντιο άξονα, λύνοντας την παρακάτω σχέση:

$$1 - C_{m_{i-1}} = \frac{(m_i - m_{i-1})\,(a_{i-1}m_{i-1} + b_{i-1})}{2}$$

Αν αντικαταστήσουμε στη σχέση αυτή όπου $C_{m_{i-1}} = F(6000) = 96{,}4\%$, $m_{i-1} = 6000$, $a_{i-1} = -1.479E-08$ και $b_{i-1} = 8.926E-05$ λαμβάνουμε ως αποτέλεσμα την τιμή του $m_i = 143.500$.

Σύμφωνα, λοιπόν, με το μοντέλο που κατασκευάσαμε για τις αποζημιώσεις υλικών ζημιών στο σημείο €143.500 η πολυγωνική γραμμή τέμνει τον οριζόντιο άξονα και επομένως εγκλείεται όλη η πληροφορία για τα δεδομένα αποζημιώσεων. Στο σημείο $m_i = 143.500$ η αθροιστική συνάρτηση πιθανότητας παίρνει την τιμή 1. Γνωρίζοντας τις πραγματικές τιμές που λαμβάνουν τα δεδομένα υλικών ζημιών καταλήγουμε στο συμπέρασμα ότι το πολυγραμμικό μοντέλο αδυνατεί να ερμηνεύσει το 0,02% των ακραίων παρατηρήσεων για αποζημιώσεις υλικών ζημιών.

5.2.3 Πολυγραμμική κατανομή του ύψους των αποζημιώσεων σωματικών βλαβών

Στην προηγούμενη υποενότητα παρουσιάστηκε αναλυτικά ο τρόπος κατασκευής του πολυγραμμικού μοντέλου που επιλέξαμε για την ερμηνεία της συμπεριφοράς της μεταβλητής “Total payment” υλικών ζημιών. Την ίδια διαδικασία θα ακολουθούμε στην παρούσα υποενότητα για τον υπολογισμό των συντελεστών (α_i, b_i) των ευθειών που θα συνθέσουν το πολυγραμμικό μοντέλο για το ύψος αποζημιώσεων σωματικών βλαβών.

Αρχικά, χωρίζουμε τα ποσά αποζημιώσεων σωματικών βλαβών σε διαστήματα πλάτους €500. Σύμφωνα με τις τιμές των δεδομένων όπως προκύπτουν από τους κλειστούς φακέλους ζημιάς της εταιρίας, το 95,4% των παρατηρούμενων αποζημιώσεων παίρνει τιμές έως τις €100.000. Επομένως, κατηγοριοποιούμε τις αποζημιώσεις σε διαστήματα πλάτους €500 έως την τιμή των €100.000. Οι υπόλοιπες παρατηρήσεις του δείγματος περιλαμβάνονται στο τελευταίο διάστημα, δηλαδή για ποσά αποζημιώσεων μεγαλύτερα των €100.000. Ως αποτέλεσμα του διαχωρισμού αυτού, οι $n = 3.629$ παρατηρήσεις ομαδοποιούνται σε k=201 ομάδες. Για κάθε

διάστημα υπολογίζουμε τη συχνότητα και την αθροιστική συχνότητα των αποζημιώσεων για το άνω άκρο, καθώς και τις αντίστοιχες σχετικές συχνότητες.

Πριν προχωρήσουμε στον προσδιορισμό των συντελεστών των ευθειών που συνθέτουν την πολυγραμμική κατανομή, υπενθυμίζουμε τη σημαντικότητα της επιλογής των σημείων αλλαγής διεύθυνσης m_i. Όπως αναφέρουμε και στην υποενότητα 5.2.2 σκοπός μας είναι να προσεγγίσουμε την καμπύλη της κατανομής των αποζημιώσεων με όσο το δυνατό λιγότερες ευθείες, υπό τον περιορισμό ότι η πολυγραμμική κατανομή δεν μπορεί να παίρνει αρνητικές τιμές. Επιλέγουμε και εδώ η προσέγγιση να πραγματοποιηθεί με περίπου 10-12 διαδοχικές ευθείες γραμμές, ώστε το μοντέλο να είναι εύκολο προς χρήση.

Ακολουθεί ο πίνακας με τα ζεύγη των συντελεστών (α_i, b_i) των ευθειών που συνθέτουν το πολυγραμμικό μοντέλο, καθώς και με σημεία αλλαγής διεύθυνσης m_i για το ύψος αποζημιώσεων σωματικών βλαβών.

Πίνακας 5.2: Συντελεστές ευθειών που συνθέτουν το πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων σωματικών βλαβών.

Multilinear model - "Total payment Bodily Injury"			
mi	**CDF (mi)**	**α**	**b**
0	0.00%	0.00E+00	0.00E+00
500	3.09%	2.47E-07	0.00E+00
1000	10.33%	8.60E-08	8.05E-05
1500	18.21%	-3.53E-08	2.02E-04
3500	39.18%	-4.40E-08	2.15E-04
5500	50.34%	-5.10E-09	7.87E-05
8000	60.46%	-8.20E-09	9.58E-05
11000	68.23%	-2.87E-09	5.31E-05
18000	78.86%	-1.83E-09	4.17E-05
25000	84.16%	-3.51E-10	1.51E-05
43500	90.22%	-3.30E-10	1.46E-05
917500	100.00%	-2.56E-13	2.35E-07

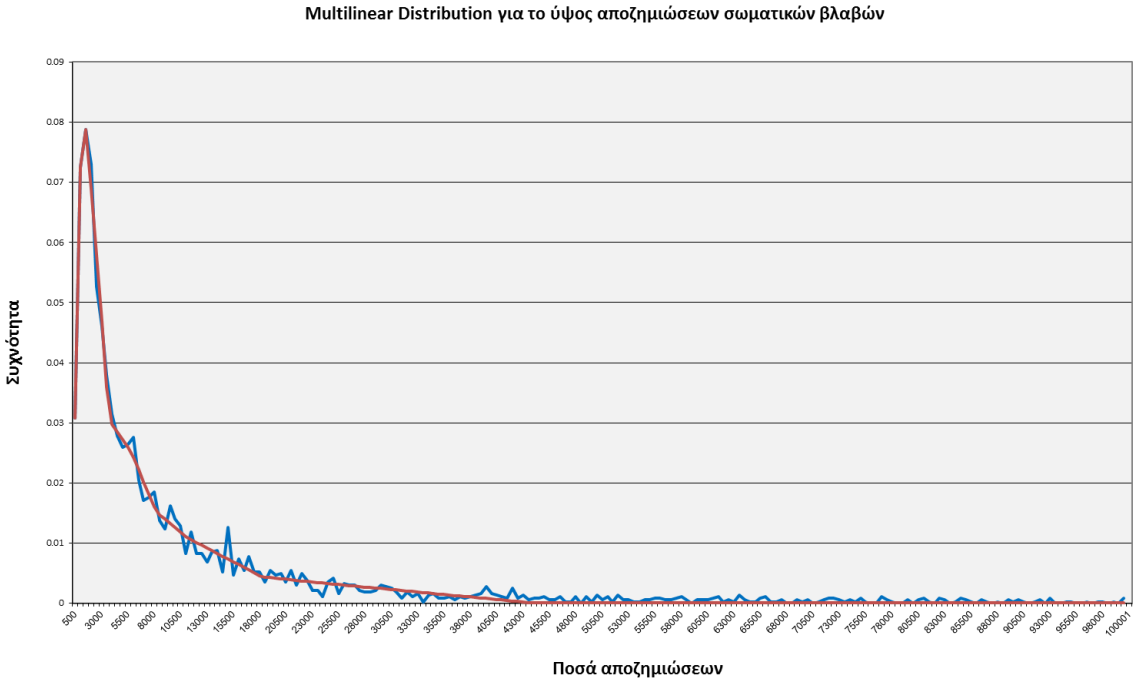
Τα αποτελέσματα του παραπάνω πίνακα προέκυψαν ακολουθώντας τη διαδικασία που αναλύθηκε στην υποενότητα 5.2.2 για τις αποζημιώσεις υλικών ζημιών. Τα ζεύγη των συντελεστών των ευθειών (α_i, b_i) υπολογίστηκαν βάσει των σχέσεων (3) και (5) της υποενότητας 5.2.1. όπως φαίνονται παρακάτω:

$$\alpha_2 = 2\left(\frac{C_2 - C_1}{(m_2 - m_1)^2} - \frac{a_1 m_1 + b_1}{(m_2 - m_1)}\right)$$

$$b_2 = (\alpha_1 - \alpha_2)\, m_1 + b_1$$

Για την επιλογή των “κατάλληλων” σημείων αλλαγής διεύθυνσης της πολυγωνικής γραμμής του μοντέλου κατασκευάσαμε για τις αποζημιώσεις σωματικών βλαβών, τα δύο γραφήματα που παρουσιάστηκαν στην προηγούμενη υποενότητα.

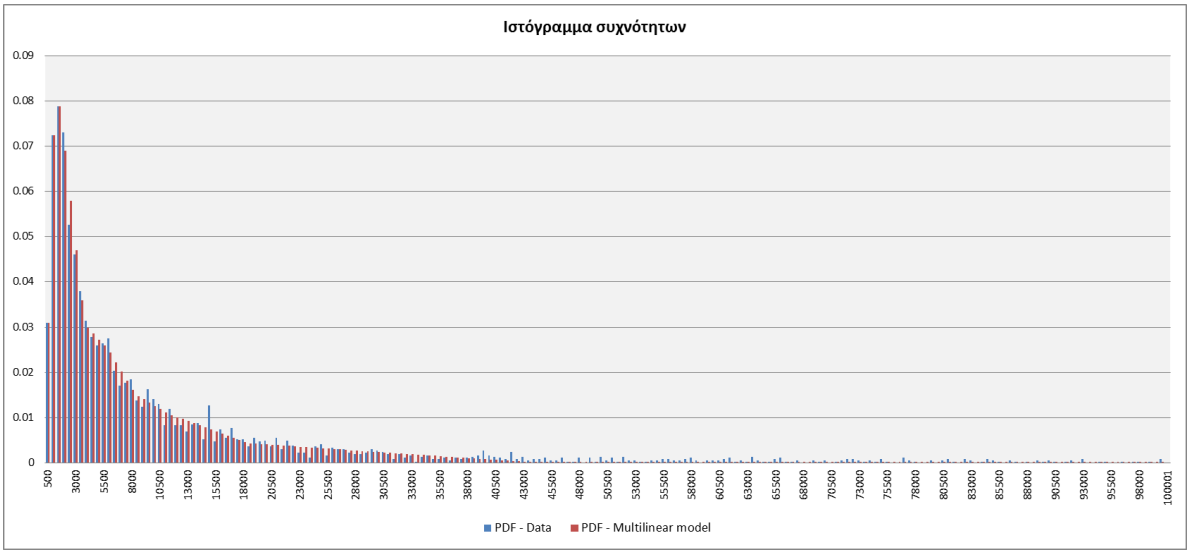
Δίνεται αρχικά το σχήμα 5.4, στο οποίο αναπαρίστανται με μπλε χρώμα η πολυγωνική γραμμή που προκύπτει από την ένωση των άκρων των k=201 διαστημάτων και με κόκκινο χρώμα το πολυγραμμικό μοντέλο του πίνακα 5.2 για τη μεταβλητή “Total payment” σωματικών βλαβών.



Σχήμα 5.4: Πολυγωνική γραμμή και πολυγραμμικό μοντέλο του ύψους των αποζημιώσεων σωματικών βλαβών.

Το σχήμα 5.4, μας παρέχει την πληροφόρηση πως και στην περίπτωση των σωματικών βλαβών η κατανομή που ακολουθούν οι αποζημιώσεις εμφανίζει περισσότερες από μία κορυφές. Είναι εύκολο να διακρίνουμε τρεις από αυτές στα διαστήματα για τιμές αποζημιώσεων περί τις €6.000, €9.500 και €15.000. Παρατηρούμε πως υπάρχουν κάποιες αποκλίσεις μεταξύ των δύο γραμμών, δηλαδή διαφορές στις παρατηρούμενες συχνότητες και σε αυτές που υπολογίζονται από το μοντέλο. Παρόλα αυτά, από το παραπάνω διάγραμμα βλέπουμε το μοντέλο να προσεγγίζει αρκετά καλά την πολυγωνική γραμμή και ιδιαίτερα την πρώτη κορυφή της κατανομής των δεδομένων αποζημιώσεων. Επιπλέον, από τη μορφή του διαγράμματος επιβεβαιώνουμε την ύπαρξη βαριάς δεξιάς ουράς στα δεδομένα αποζημιώσεων σωματικών βλαβών όπως έχουμε ήδη συμπεράνει και από την ανάλυση που πραγματοποιήσαμε στα κεφάλαια 2 και 3.

Το δεύτερο γράφημα που χρησιμοποιήσαμε για να ελέγξουμε την προσέγγιση της κατανομής που ακολουθούν τα δεδομένα αποζημιώσεων από το νέο μοντέλο, είναι το συνδυαστικό ιστόγραμμα συχνοτήτων.



Σχήμα 5.5:Ιστόγραμμα συχνοτήτων της σ.π.π. και της πολυγραμμικής κατανομής του ύψους των αποζημιώσεων σωματικών βλαβών.

Οι στήλες με μπλε χρώμα του ιστογράμματος 5.5, αντιστοιχούν στις σχετικές συχνότητες που υπολογίστηκαν από τα δεδομένα αποζημιώσεων για κάθε ένα από τα k διαστήματα των €500. Αντίστοιχα, οι κόκκινες στήλες προέρχονται από την εφαρμογή του πολυγραμμικού μοντέλου στα διαστήματα αυτά. Συγκρίνοντας μεταξύ τους τις κλάσεις του ιστογράμματος για ίδια διαστήματα τιμών, καταλήγουμε στο συμπέρασμα ότι εμφανίζονται πολύ μικρές αποκλίσεις μεταξύ των συχνοτήτων. Οι μεγαλύτερες από αυτές τις αποκλίσεις εμφανίζονται στα διαστήματα από €2.000 έως €2.500 και €14.500 έως €15.000. Στην πρώτη περίπτωση το μοντέλο υπερεκτιμά τη συχνότητα εμφάνισης αποζημιώσεων αυτού του ύψους, ενώ στη δεύτερη την υποεκτιμά. Πρόκειται για την κορυφή που εμφανίζεται στην τιμή των €15.000 όπως βλέπουμε και στο σχήμα 5.4. Βέβαια, να σημειώσουμε ότι οι διαφορές αυτές δεν ξεπερνούν το 0,5%. Το μοντέλο που κατασκευάσαμε επιλέξαμε να συντίθεται από 10 περίπου ευθείες γραμμές, επομένως όπως έχουμε προαναφέρει τέτοιες αποκλίσεις είναι αναμενόμενες.

Επιστρέφοντας στον πίνακα 5.2, το τελευταίο σημείο αλλαγής κλίσης της γραμμής του μοντέλου υπολογίστηκε με τη λογική που παρουσιάστηκε στην υποενότητα 5.2.2, δηλαδή με το εμβαδόν τριγώνου της τελευταίας ευθείας που συνθέτει το πολυγραμμικό μοντέλο και του οριζόντιου άξονα. Σύμφωνα με τον υπολογισμό του εμβαδού, το σημείο m_i που το μοντέλο τέμνει τον άξονα των αποζημιώσεων και η αθροιστική κατανομή παίρνει την τιμή 1, είναι στις €917.500. Γνωρίζοντας τις πραγματικές τιμές που λαμβάνουν τα δεδομένα αποζημιώσεων σωματικών βλαβών της εταιρίας καταλήγουμε στο συμπέρασμα ότι το πολυγραμμικό μοντέλο αδυνατεί

να ερμηνεύσει το 0,03% των ακραίων παρατηρήσεων για αποζημιώσεις σωματικών βλαβών.

Οι υπολογισμοί που έγιναν για τον προσδιορισμό των συντελεστών των ευθειών των πολυγραμμικών μοντέλων των υποενοτήτων 5.2.2 και 5.2.3, καθώς και οι διαγραμματικές απεικονίσεις πραγματοποιήθηκαν με το πρόγραμμα Excel. Μας δόθηκε έτσι η δυνατότητα να πραγματοποιούμε δοκιμές για την επιλογή των σημείων αλλαγής διεύθυνσης m_i των δύο μοντέλων με γρήγορο και εύκολο τρόπο, μέχρι να καταλήξουμε σε αυτά που περιγράφουν καλύτερα τη συμπεριφορά των δεδομένων αποζημιώσεων του κλάδου αστικής ευθύνης αυτοκινήτου της εταιρίας.

Παράρτημα Α

Παρακάτω παρατίθεται συνοπτικά ο κώδικας στην R που χρησιμοποιήθηκε για τον υπολογισμό των περιγραφικών χαρακτηριστικών των μεταβλητών, τη δημιουργία των γραφημάτων και την πραγματοποίηση τεστ κανονικότητας τα οποία παρουσιάστηκαν αναλυτικά στις ενότητες 2.1 και 2.2 της παρούσας εργασίας. Σημειώνεται ότι δίνουμε αναλυτικά τις εντολές για τη μεταβλητή "Total payment" καθώς οι ίδιες χρησιμοποιήθηκαν και για τις μεταβλητές "Total time" και "Delay".

```
TP191<-read.table("TP191.txt",header=TRUE)
z<-TP191[[1]]
z
length(z)
summary(z)
var(z)
sqrt(var(z))
quantile(z,c(0.99))
skewness(z)
kurtosis(z)
par(mfrow=c(1,2))
hist(z,main= paste("Total payment - Accidental Damage"),col="blue")
qqnorm(z);qqline(z)

TP191<-read.table("TP191.txt",header=TRUE)
z<-TP191[[1]]
z
lillie.test(z)
ad.test(z)
cvm.test(z)
pearson.test(z)

TP191<-read.table("TP191.txt",header=TRUE)
x<-TP191[[1]]
TP192<-read.table("TP192.txt",header=TRUE)
y<-TP192[[1]]
par(mfrow=c(1,2))
boxplot(x,main= paste("Total payment - Accidental Damage"))
boxplot(y,main= paste("Total payment - Bodily injury"))

TP191<-read.table("TP191.txt",header=TRUE)
x<-TP191[[1]]
x
TP192<-read.table("TP192.txt",header=TRUE)
y<-TP192[[1]]
y
boxplot(x,y,names=c("Accidental Damage","Bodily injury"),
col=c("blue","red"),xlab="Insurance Company",
main="Boxplot Total payment")
```

Παράρτημα Β

Συνεχείς και διακριτές κατανομές διαθέσιμες στο πρόγραμμα “EasyFit”.

Πίνακας 1

Continuous Distributions	
Beta	Levy
Burr (Burr Type 12, or Singh-Maddala)	Laplace (Double Exponential)
Cauchy (Lorentz)	Logistic
Chi-Squared	Log-Gamma
Dagum (Burr Type 3, or Inverse Burr)	Log-Logistic (Fisk)
Erlang	Lognormal
Error (Exponential Power, or Generalized Error)	Nakagami (Nakagami-m)
Error Function	Normal (Gaussian)
Exponential	Pareto - first kind
F Distribution	Pareto - second kind (Lomax)
Fatigue Life (Birnbaum-Saunders)	Pearson Type 5 (Inverse Gamma)
Frechet (Maximum Extreme Value Type 2)	Pearson Type 6 (Beta dist. of the second kind)
Gamma	Pert
Generalized Gamma	Power Function
Gumbel Max (Maximum Extreme Value Type 1)	Rayleigh
Gumbel Min (Minimum Extreme Value Type 1)	Reciprocal
Hyperbolic Secant	Rice (Ricean, or Nakagami-n)
Inverse Gaussian	Student's
Johnson SB	Triangular
Johnson SU	Uniform
Kumaraswamy	Weibull
Discrete Distributions	
Bernoulli	Hypergeometric
Binomial	Logarithmic
Discrete Uniform	Negative Binomial
Geometric	Poisson

Πίνακας 2

Advanced Distributions	
Generalized Extreme Value	Phased Bi-Exponential
Generalized Logistic	Phased Bi-Weibull
Generalized Pareto	Wakeby
Log-Pearson 3 (LP3)	

Παρακάτω παρατίθεται συνοπτικά ο κώδικας στην R που χρησιμοποιήθηκε για το μετασχηματισμό Box-Cox των δεδομένων, που παρουσιάστηκε στην ενότητα 3.4 της παρούσας εργασίας. Σημειώνεται ότι δίνουμε αναλυτικά τις εντολές για τη μεταβλητή "Total payment" καθώς οι ίδιες χρησιμοποιήθηκαν και για τις μεταβλητές "Total time" και "Delay".

```
TP191<-read.table("TP191.txt",header=TRUE)
y<-TP191[[1]]
y

x<-(y+0)
x
res=boxcox(x~1,lambda=seq(-2,2,by=0.01),plotit=TRUE)
res

lambda<-with(res,x[which.max(y)])
lambda
trans.x=BoxCox(x,lambda)
trans.x

summary(trans.x)
var(trans.x)
sqrt(var(trans.x))
quantile(trans.x,c(0.99))
skewness(trans.x)
kurtosis(trans.x)

par(mfrow=c(1,2))
boxcox(x~1)
qqnorm(trans.x);qqline(trans.x)

lillie.test(trans.x)
ad.test(trans.x)
cvm.test(trans.x)
pearson.test(trans.x)
```

Παράρτημα Γ

Παρακάτω παρατίθενται αναλυτικά τα αποτελέσματα του Mathematica από τη λύση των συστημάτων της μεθόδου των ροπών για την εκτίμηση των παραμέτρων των μοντέλων μείξης γάμμα κατανομών που παρουσιάστηκε στην ενότητα 4.1 της παρούσας εργασίας.

<u>“Total payment” υλικών ζημιών</u>

{w->0.989782,a1->0.00169648,b1->124.192,a2->-1.01032,b2->-133.821},{w->0.000352098 -0.000348458 I,a1->-2.04306-0.737051 I,b1->187.763 +100.675 I,a2->-0.000877168+0.0576817 I,b2->-0.259398-31.5396 I},{w->0.000352098 +0.000348458 I,a1->-2.04306+0.737051 I,b1->187.763 -100.675 I,a2->-0.000877168-0.0576817 I,b2->-0.259398+31.5396 I},{w->0.000321018,a1->3.06845,b1->66.9031,a2->0.3062,b2->4.97906},{w->0.999679,a1->0.3062,b1->4.97906,a2->3.06845,b2->66.9031},{w->-0.085305,a1->-0.515333,b1->97.2116,a2->-0.0368345,b2->67.1259},{w->1.0853,a1->-0.0368345,b1->67.1259,a2->-0.515333,b2->97.2116},{w->0.0102178,a1->-1.01032,b1->-133.821,a2->0.00169648,b2->124.192},{w->0.999648 -0.000348458 I,a1->-0.000877168-0.0576817 I,b1->-0.259398+31.5396 I,a2->-2.04306+0.737051 I,b2->187.763 -100.675 I},{w->0.999648 +0.000348458 I,a1->-0.000877168+0.0576817 I,b1->-0.259398-31.5396 I,a2->-2.04306-0.737051 I,b2->187.763 +100.675 I},{w->1.00811,a1->-2.24896*10^-6,b1->424.899,a2->-0.916927,b2->213.817},{w->-0.00558438-0.00134751 I,a1->-0.915766+0.000417854 I,b1->302.947 -79.7665 I,a2->-0.000228205+0.0000939014 I,b2->268.334 -41.7435 I},{w->-0.00558438+0.00134751 I,a1->-0.915766-0.000417854 I,b1->302.947 +79.7665 I,a2->-0.000228205-0.0000939014 I,b2->268.334 +41.7435 I},{w->0.5 +3.37724 I,a1->-1.00006+0.0000605081 I,b1->280.728 +41.8144 I,a2->-1.00006-0.0000605081 I,b2->280.728 -41.8144 I},{w->0.5 -3.37724 I,a1->-1.00006-0.0000605081 I,b1->280.728 -41.8144 I,a2->-1.00006+0.0000605081 I,b2->280.728 +41.8144 I},{w->1.00558 +0.00134751 I,a1->-0.000228205+0.0000939014 I,b1->268.334 -41.7435 I,a2->-0.915766+0.000417854 I,b2->302.947 -79.7665 I},{w->1.00558 -0.00134751 I,a1->-0.000228205-0.0000939014 I,b1->268.334 +41.7435 I,a2->-0.915766-0.000417854 I,b2->302.947 +79.7665 I},{w->-0.00811491,a1->-0.916927,b1->213.817,a2->-2.24896*10^-6,b2->424.899}

<u>“Total payment” σωματικών βλαβών</u>

{w->1.00001,a1->0.130387,b1->166.531,a2->-2.0008,b2->-4009.75},{w->1.03286,a1->0.180274,b1->150.856,a2->-13.6598,b2->-14.3469},{w->0.0723293,a1->1.38293,b1->120.995,a2->0.402624,b2->25.5616},{w->-0.0115612+0.0201274 I,a1->0.523542 -4.44642 I,b1->72.8194 +54.3575 I,a2->0.16916 +0.095272 I,b2->63.1452 -86.2083 I},{w->-0.0115612-0.0201274 I,a1->0.523542 +4.44642 I,b1->72.8194 -54.3575 I,a2->0.16916 -0.095272 I,b2->63.1452 +86.2083 I},{w->1.01156 +0.0201274 I,a1->0.16916 -0.095272 I,b1->63.1452 +86.2083 I,a2->0.523542 +4.44642 I,b2->72.8194 -54.3575 I},{w->1.01156 -0.0201274 I,a1->0.16916 +0.095272 I,b1->63.1452 -86.2083 I,a2->0.523542 -4.44642 I,b2->72.8194 +54.3575 I},{w->0.927671,a1->0.402624,b1->25.5616,a2->1.38293,b2->120.995},{w->0.5 +1.05843 I,a1->-0.189649-0.0966446 I,b1->141.583 +39.22 I,a2->-0.189649+0.0966446 I,b2->141.583 -39.22 I},{w->0.5 -1.05843 I,a1->-0.189649+0.0966446 I,b1->141.583 -39.22 I,a2->-0.189649-0.0966446 I,b2->141.583 +39.22 I},{w->-0.0328562,a1->-13.6598,b1->-14.3469,a2->0.180274,b2->150.856},{w->-7.94285*10^-6,a1->-2.0008,b1->-4009.75,a2->0.130387,b2->166.531},{w->-0.000323797,a1->-1.77221,b1->773.221,a2->0.123264,b2->171.984},{w->0.5 +1.45855 I,a1->-0.898055-0.00229828 I,b1->176.697 +68.3249 I,a2->-0.898055+0.00229828 I,b2->176.697 -68.3249 I},{w->0.5 -1.45855 I,a1->-0.898055+0.00229828 I,b1->176.697 -68.3249 I,a2->-0.898055-0.00229828 I,b2->176.697 +68.3249 I},{w->0.5 +0.172372 I,a1->-0.0207334-0.18537 I,b1->433.604 +15.2225 I,a2->-0.0207334+0.18537 I,b2->433.604 -15.2225 I},{w->0.5 -0.172372 I,a1->-0.0207334+0.18537

I,b1->433.604 -15.2225 I,a2->-0.0207334-0.18537 I,b2->433.604 +15.2225 I},{w->1.00032,a1->0.123264,b1->171.984,a2->-1.77221,b2->773.221}

“Total time” υλικών ζημιών

{w->1.00314,a1->0.243654,b1->18.2089,a2->-2.27262,b2->-39.2779},{w->1.04722,a1->0.320038,b1->16.6173,a2->-6.83975,b2->-4.33274},{w->0.0548898,a1->2.82873,b1->11.0086,a2->0.77106,b2->3.37669},{w->-0.0205602+0.0240759 I,a1->-1.22409-4.43718 I,b1->7.26898 +9.71841 I,a2->0.27366 +0.243776 I,b2->4.76677 -10.1216 I},{w->-0.0205602-0.0240759 I,a1->-1.22409+4.43718 I,b1->7.26898 -9.71841 I,a2->0.27366 -0.243776 I,b2->4.76677 +10.1216 I},{w->1.02056 +0.0240759 I,a1->0.27366 -0.243776 I,b1->4.76677 +10.1216 I,a2->-1.22409+4.43718 I,b2->7.26898 -9.71841 I},{w->1.02056 -0.0240759 I,a1->0.27366 +0.243776 I,b1->4.76677 -10.1216 I,a2->-1.22409-4.43718 I,b2->7.26898 +9.71841 I},{w->0.94511,a1->0.77106,b1->3.37669,a2->2.82873,b2->11.0086},{w->-0.047217,a1->-6.83975,b1->-4.33274,a2->0.320038,b2->16.6173},{w->0.5 +0.75665 I,a1->-0.206539-0.162898 I,b1->17.6023 +7.29065 I,a2->-0.206539+0.162898 I,b2->17.6023 -7.29065 I},{w->0.5 -0.75665 I,a1->-0.206539+0.162898 I,b1->17.6023 -7.29065 I,a2->-0.206539-0.162898 I,b2->17.6023 +7.29065 I},{w->-0.00314354,a1->-2.27262,b1->-39.2779,a2->0.243654,b2->18.2089},{w->-0.00265932,a1->-1.46826,b1->68.0046,a2->0.190029,b2->20.4922},{w->0.5 +1.31307 I,a1->-0.867335-0.0019375 I,b1->21.659 +10.0213 I,a2->-0.867335+0.0019375 I,b2->21.659 -10.0213 I},{w->0.5 -1.31307 I,a1->-0.867335+0.0019375 I,b1->21.659 -10.0213 I,a2->-0.867335-0.0019375 I,b2->21.659 +10.0213 I},{w->0.5 +0.365178 I,a1->-0.0803833-0.198289 I,b1->56.0363 +2.17561 I,a2->-0.0803833+0.198289 I,b2->56.0363 -2.17561 I},{w->0.5 -0.365178 I,a1->-0.0803833+0.198289 I,b1->56.0363 -2.17561 I,a2->-0.0803833-0.198289 I,b2->56.0363 +2.17561 I},{w->1.00266,a1->0.190029,b1->20.4922,a2->-1.46826,b2->68.0046}

Βιβλιογραφία

Ελληνική βιβλιογραφία

1. Αντζουλάκος Δ. (2011), *Ανάλυση Δεδομένων με τη Χρήση Στατιστικών Πακέτων Εισαγωγή στο R.* Σημειώσεις παραδόσεων. Πανεπιστήμιο Πειραιώς ΠΜΣ στην Εφαρμοσμένη Στατιστική.
2. Κιόχος Π. (1993), *Στατιστική.* Εκδόσεις INTERBOOKS.
3. Μαυραγάνης Π. (1979), *Κριτήρια Κανονικότητας, Γενικεύσεις, Εφαρμογές*. Εκδόσεις Λαλιώτη.
4. Μπούτσικας Μ. (2008), *Σημειώσεις διαλέξεων στη «Θεωρία Ακραίων τιμών»*. ΠΜΣ στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου, Τμ. Στατιστικής & Ασφ. Επιστ., Πανεπ. Πειραιώς.
5. Ξεκαλάκη Ε. (2001), *Μη Παραμετρική Στατιστική.* Οικονομικό Πανεπιστήμιο Αθηνών.
6. Πολίτης Κ. (2012), *Εισαγωγή στη Θεωρία Συλλογικού Κινδύνου.* Εκδόσεις Σταμούλης.
7. Τσίμπος Κ. & Γεωργιακώδης Φ. (1999), *Περιγραφική & Διερευνητική Στατιστική Ανάλυση Δεδομένων, τόμος Α΄, Μονοδιάστατη ανάλυση.* Εκδόσεις Σταμούλης.
8. Φωκιανός Κ. & Χαραλάμπους Χ. (2010), *Εισαγωγή στην R Πρόχειρες Σημειώσεις*. Τμ .Μαθηματικών & Στατιστικής, Πανεπιστήμιο Κύπρου.

Ξενόγλωσση βιβλιογραφία

1. Cooray K. & Ananda M.M.A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal* 5, 321-334.
2. Johnson N., Kotz S. & Balakrishnan N. (1994), *Continuous Univariate Distributions, Volume 1, Second Edition.* John Wiley & Sons, Inc.
3. Klugman S., Panjer H. & Wilmot G.E. (2004), *Loss Models: From Data to Deci*sions, *Second edition*. New York: John Wiley and Sons.
4. Li P. (2005), *Box-Cox Transformations: An Overview.* University of Connecticut, Department of Statistics.

5. Nadarajah S. & Bakar S. (2014), New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal* 2, 180-187.
6. Osborne J. (2010), Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment Research & Evaluation, Volume 15, No 12.*
7. Pakacova V. & Zapletal D. (2013), Mixture Distributions in Modelling of Insurance Losses. Proceedings of the International Conference on Applied Mathematics and Computation Methods in Engineering, 122 – 125.

Διαδικτυακοί Τόποι

1. Box G. & Cox D. (1964), An Analysis of Transformations. *Journal of the Royal Statistical Society,. Series B, Vol. 26, No. 2.*, 211-252. Διαθέσιμο στο διαδικτυακό τόπο: http://www.ime.usp.br/~abe/lista/pdfQWaCMboK68.pdf
2. Buthmann A. (n.d.), Making Data Normal Using Box-Cox Power Transformation. Διαθέσιμο στον διαδικτυακό τόπο: https://www.isixsigma.com/tools-templates/normality/making-data-normal-using-box-cox-power-transformation/
3. Chhay L. & Hyndman (n.d.), Automatic selection of Box Cox transformation parameter, Package 'forecast'. http://www.inside-r.org/packages/cran/forecast/docs/BoxCox.lambda
4. Gross J. & Ligges U. (2015), Test for Normality, Package 'nortest'. https://cran.r-project.org/web/packages/nortest/nortest.pdf
5. Mathwave Technologies, EasyFit. http://www.mathwave.com/products/easyfit_desc.html
6. Sherman P. (n.d.), Tips for Recognizing and Transforming Non-normal Data. Διαθέσιμο στον διαδικτυακό τόπο: https://www.isixsigma.com/tools-templates/normality/tips-recognizing-and-transforming-non-normal-data/
7. Wikipedia (2014), Multimodal distribution. Διαθέσιμο στο διαδικτυακό τόπο: https://en.wikipedia.org/wiki/Multimodal_distribution
8. Wikipedia (2012), Piecewise. Διαθέσιμο στο διαδικτυακό τόπο: https://en.wikipedia.org/wiki/Piecewise
9. Wikipedia (2012), Piecewise linear function. Διαθέσιμο στον διαδικτυακό

τόπο: https://en.wikipedia.org/wiki/Piecewise_linear_function

10. ΕΑΕΕ (2016), Παραγωγή ασφαλίστρων δωδεκαμήνου 2015. Υπηρεσία Μελετών & Στατιστικής, Οικονομικές Μελέτες 78. Διαθέσιμο στον διαδικτυακό τόπο: http://www.eaee.gr/cms/paragogi-asfalistron
11. ΕΑΕΕ (2016), Συμφωνία Άμεσου Διακανονισμού Ζημιών από τροχαίο ατύχημα Σύστημα Άμεσης Πληρωμής. Διαθέσιμο στον διαδικτυακό τόπο: http://www.eaee.gr/cms/basic-page/77/symfonia-amesoy-diakanonismoy-zimion-apo-trohaio-atyhima-systima-amesis-pliromis