

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ
ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Μεταπτυχιακό Πρόγραμμα Σπουδών στην
Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Εκτίμηση Ασφαλίσεων με Μοντέλα
Αξιοπιστίας Σταυρωτής Ταξινόμησης

Διπλωματική Εργασία
Αριστοτέλης Ντέκας

Επιβλέπων: Πιτσέλης Γεώργιος

Επίκουρος Καθηγητής, Πανεπιστήμιο Πειραιά

Πειραιάς, Ιούνιος 2018

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS AND
INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN ACTUARIAL
SCIENCE AND RISK MANAGEMENT

**Premium Estimation with Cross Classification
Credibility Models**

M. SC. Thesis
Aristotelis Ntekas

Advisor: Pitselis Georgios
Assistant Professor, University of Piraeus

Piraeus, June 2018

Περίληψη

Η εκτίμηση ασφαλιστρων είναι ένα μείζον ζήτημα για τις ασφαλιστικές εταιρείες. Για τον υπολογισμό τους υπάρχουν αρκετές μέθοδοι. Μια από αυτές, είναι η τεχνική της Αξιοπιστίας. Στην παρούσα εργασία, αρχικά, γίνεται εισαγωγή σε βασικές έννοιες της θεωρίας Αξιοπιστίας χαρτοφυλακίου (Credibility Theory) και της μεθόδου της Ανάλυσης Διασποράς.

Στη συνέχεια παρουσιάζεται η επιστημονική εξέλιξη των μοντέλων Αξιοπιστίας χαρτοφυλακίου, με κατάληξη στο Ιεραρχικό μοντέλο. Παράλληλα πραγματοποιείται προσέγγιση όλων των μοντέλων μέσω της μεθόδου της Ανάλυσης Διασποράς, με σκοπό τον προσδιορισμό των δεσμών μεταξύ αυτής και των μοντέλων Αξιοπιστίας.

Ως διερεύνηση του Ιεραρχικού μοντέλου Αξιοπιστίας αναπτύσσεται το μοντέλο Αξιοπιστίας Σταυρωτής Ταξινόμησης, με στόχο να καλυφθούν περιπτώσεις σύνθετων χαρτοφυλακίων για τις ασφαλιστικές εταιρείες, ενώ γίνεται αντίστοιχη προσέγγιση και με τη μέθοδο της Ανάλυσης Διασποράς.

Τέλος, παρουσιάζονται εφαρμογές των μοντέλων με χρήση του στατιστικού πακέτου R, και προσδιορίζεται η σχέση τους με τη μέθοδο της Ανάλυσης Διασποράς.

Abstract

Premium estimation is a major issue which insurance companies are faced with. There are lots of methods for premium estimation. One of them is the Credibility technique. In this thesis, firstly, the main concepts of the Credibility Theory and the Analysis of Variance Method are introduced.

Afterwards, we present the scientific development of Credibility Models, ending in Hierarchical Model. Simultaneously, all models are approached through the method of Analysis of Variance, in order links between that method and Credibility Theory to be identified.

As an inquiry of the Hierarchical Credibility Model, the Cross Classification Model is expounded, in order to cover cases of complex portfolios, while an approach with the use of method of Analysis of Variance is made similarly to the other models.

Finally, we present applications of Credibility models and their correlation with the technique of Analysis of Variance, using the statistical package R.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον Καθηγητή μου, κύριο Γεώργιο Πιτσέλη, για την δυνατότητα που μου έδωσε να ασχοληθώ με το θέμα της παρούσας εργασίας, καθώς και για τη βοήθεια και το χρόνο που διέθεσε κατά τη διάρκεια της συγγραφής της. Επιπλέον θα ήθελα να ευχαριστήσω τον κύριο Απόστολο Μποζίκα για την πολύτιμη βοήθεια του και τις συμβουλές του, και τους κ. Γεώργιο Τζαβελά και κ. Χρήστο Αγιακλόγλου, για την τιμή που μου έκαναν να είναι μέλη της συμβουλευτικής μου επιτροπής. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την συμπαράσταση και την υποστήριξη κατά τη διάρκεια των σπουδών μου.

Περιεχόμενα

1	Εισαγωγή	1
2	Το Μοντέλο Αξιοπιστίας του Buhlmann	7
2.1	Μοντέλο Buhlmann	7
2.2	Υποθέσεις	8
2.3	Συμβολισμοί και Παράμετροι Δόμησης	8
2.4	Σχέσεις Συνδιασπορών και Μέσες Τιμές	9
2.5	Αμερόληπτοι Εκτιμητές Παραμέτρων	10
2.6	Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς	10
2.7	Αριθμητικό Παράδειγμα στο Μοντέλο του Buhlmann	16
3	Το Μοντέλο Αξιοπιστίας των Buhlmann – Straub	19
3.1	Μοντέλο Buhlmann-Straub	19
3.2	Υποθέσεις	20
3.3	Συμβολισμοί και Παράμετροι Δόμησης	20
3.4	Σχέσεις Συνδιασπορών και Μέσες Τιμές	22
3.5	Αμερόληπτοι Εκτιμητές Παραμέτρων	22
3.6	Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς	23
3.7	Αριθμητικό Παράδειγμα στο Μοντέλο των Buhlmann-Straub	28
4	Το Μοντέλο Αξιοπιστίας των Hachemeister	31
4.1	Μοντέλο Hachemeister	31
4.2	Υποθέσεις	32
4.3	Συμβολισμοί και Παράμετροι Δόμησης	33
4.4	Σχέσεις Συνδιασπορών και Μέσες Τιμές	34
4.5	Αμερόληπτοι Εκτιμητές Παραμέτρων	34
4.6	Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς	34
4.7	Αριθμητικό Παράδειγμα στο Μοντέλο του Hachemeister	38

5	Το Μοντέλο Αξιοπιστίας του Jewell	40
5.1	Μοντέλο Jewell	41
5.2	Υποθέσεις	41
5.3	Παράμετροι Δόμησης και Συμβολισμοί	41
5.4	Σχέσεις Συνδιασπορών	43
5.5	Αμερόληπτοι Εκτιμητές Παραμέτρων	43
5.6	Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς	44
5.7	Αριθμητικό Παράδειγμα στο Μοντέλο του Jewell	53
6	Το Μοντέλο Αξιοπιστίας Σταυρωτής Ταξινόμησης	56
6.1	Μοντέλο Σταυρωτής Ταξινόμησης (Cross Classification)	60
6.2	Υποθέσεις	60
6.3	Παράμετροι Δόμησης και Συμβολισμοί	61
6.4	Σχέσεις Συνδιασπορών	62
6.5	Αμερόληπτοι Εκτιμητές Παραμέτρων	62
6.6	Προσέγγιση με τη μέθοδο της Ανάλυσης Διασποράς	65
6.7	Αριθμητικό Παράδειγμα στο Μοντέλο Σταυρωτής Ταξινόμησης (Cross Classification)	75
7	Επίλογος	79
A'	Κώδικας R	83
A'.1	Κώδικας R για Buhlmann	83
A'.2	Κώδικας R για Buhlmann-Straub	84
A'.3	Κώδικας R για Hachemeister	85
A'.4	Κώδικας R για Hierarchical	85
A'.5	Κώδικας R για Σταυρωτή Ταξινόμηση 2 Παραγόντων (Cross Classification) .	87

Κεφάλαιο 1

Εισαγωγή

Συχνά οι ασφαλιστικές εταιρείες καλούνται να καθορίσουν το ασφάλιστρο της επόμενης χρονιάς για μια ομάδα ασφαλιστικών συμβολαίων. Η διαδικασία αυτή τις φέρνει αντιμέτωπες με προβλήματα όταν η εμπειρία σχετικά με ένα είδος συμβάσεων είναι περιορισμένη σε σχέση με άλλα είδη που σχετίζονται είτε λιγότερο είτε περισσότερο μεταξύ τους. Πιο συγκεκριμένα πρέπει να δημιουργηθεί ένα πρόγραμμα βαθμολόγησης της εμπειρίας αυτής ώστε η τιμολόγηση κάθε ομάδας συμβολαίων να γίνει λαμβάνοντας υπόψη τόσο την ατομική (individual experience) όσο και τη συλλογική εμπειρία (collective experience) εξέλιξης ζημιών ενός χαρτοφυλακίου.

Οι επιλογές που υπάρχουν είναι οι εξής δύο: Η πρώτη είναι να ορίζεται το ίδιο ασφάλιστρο με βάση το γενικό μέσο όρο \bar{X} των δεδομένων από όλες τις ομάδες συμβολαίων του χαρτοφυλακίου. Η δεύτερη επιλογή είναι να χρεώνουμε σε κάθε ομάδα j συμβολαίων τις αντίστοιχες μέσες αξιώσεις \bar{X}_j ως ατομικό ασφάλιστρο.

Και οι δύο επιλογές ωστόσο παρουσιάζουν «αδυναμίες». Για να μπορεί να γίνει χρήση της πρώτης επιλογής θα πρέπει το χαρτοφυλάκιο να είναι ομοιογενές. Γενικά όμως τα χαρτοφυλάκια είναι ετερογενή, δηλαδή, ασφαλιστήρια συμβόλαια με σχετικά όμοιες ιδιότητες ως προς ορισμένα χαρακτηριστικά κινδύνου, έχουν διαφορετικές κατανομές και αναμενόμενες τιμές ζημιών εξαιτίας άλλων χαρακτηριστικών κινδύνου που είναι δύσκολο να παρατηρηθούν. Αυτά τα χαρακτηριστικά είναι που φέρουν ως αποτέλεσμα την ετερογένεια του χαρτοφυλακίου. Αντίστοιχα η δεύτερη μέθοδος επιτρέπει την ετερογένεια στο χαρτοφυλάκιο, αλλά η χρήση της είναι εφικτή μόνο εάν η εμπειρία σε κάθε ομάδα συμβολαίων είναι αρκετά μεγάλη. Έτσι, από τις αρχές του 1900, συχνά επιλέγεται το ασφάλιστρο να εξάγεται ως σταθμισμένος μέσος όρος των ασφαλίσεων από τις δύο προαναφερόμενες τεχνικές [βλέπε Dannenburg et al. (1996)].

Αυτή η τεχνική τιμολόγησης επομένως, που ονομάζεται αξιοπιστία (Credibility), χρησιμοποιείται για τον καθορισμό ασφαλίσεων σε ομάδες ασφαλιστηρίων συμβολαίων ετερογενών χαρτοφυλακίων, στην περίπτωση που υπάρχει περιορισμένη εμπειρία εξέλιξης ζημιών για κάθε ομάδα, αλλά απεριόριστη εμπειρία εξέλιξης ζημιών σε μεγαλύτερη ομάδα ασφαλιστηρίων

συμβολαίων (χαρτοφυλάκιο) που έχουν όμοια χαρακτηριστικά.

Ο γενικός τύπος της θεωρίας αξιοπιστίας είναι:

$$C = Z_j \bar{X}_j + (1 - Z_j) \bar{X}, \quad 0 < Z < 1 \quad (1.1)$$

όπου ο παράγοντας Z_j εκφράζει πόσο αξιόπιστη είναι η ατομική εξέλιξη ζημιών, και καλείται συντελεστής αξιοπιστίας, ενώ όπως έχει ήδη αναφερθεί το \bar{X}_j είναι η μέση τιμή από τα δεδομένα της ομάδας j συμβολαίων και \bar{X} ο γενικός μέσος όρος των δεδομένων από όλες τις ομάδες συμβολαίων του χαρτοφυλακίου. Η τιμολόγηση βάσει ατομικής και συλλογικής εμπειρίας είναι δικαιολογημένη, δεδομένου ότι το χαρτοφυλάκιο δεν είναι ούτε εντελώς ετερογενές, ούτε εντελώς ομοιογενές. Οι κίνδυνοι στην ομάδα j έχουν κοινά χαρακτηριστικά με τους αντίστοιχους κινδύνους που υπάρχουν σε άλλες ομάδες από την j , καθώς και τα δικά τους χαρακτηριστικά [βλέπε Kaas et al. (2008)].

Κάποιος θα επέλεγε την τιμή του Z_j κοντά στο ένα υπό τις παρακάτω συνθήκες:

1. Στην ομάδα j η εμπειρία κινδύνου είναι τεράστια
2. Παρουσιάζεται μόνο μικρή διακύμανση εντός της ομάδας j
3. Η διαφορά μεταξύ των ομάδων είναι σημαντική

Για τον καθορισμό της τιμής του Z_j υπάρχουν δύο μέθοδοι. Η περιορισμένης διακύμανσης θεωρία αξιοπιστίας, όπου σε μια ομάδα συμβολαίων δίνεται η τιμή $Z_j = 1$ (πλήρης αξιοπιστία) εάν η εμπειρία είναι επαρκώς μεγάλη. Αυτό σημαίνει πως η πιθανότητα ύπαρξης τουλάχιστον ενός σχετικού σφάλματος στον μεμονωμένο μέσο δεν υπερβαίνει ένα συγκεκριμένο όριο. Διαφορετικά, ο συντελεστής αξιοπιστίας λαμβάνεται ως αναλογία της πραγματικής εμπειρίας και της εμπειρίας που απαιτείται για την πλήρη αξιοπιστία. Μεγαλύτερο ενδιαφέρον παρουσιάζει η δεύτερη, η μέγιστης ακρίβειας θεωρία αξιοπιστίας, όπου οι συντελεστές αξιοπιστίας συμπεριφέρονται ως βέλτιστοι εκτιμητές γραμμικής μορφής σε ένα Μπευζιανό μοντέλο. Σε αυτή την περίπτωση λέμε ότι ο συντελεστής αξιοπιστίας είναι ακριβής Μπευζιανός. Το μοντέλο αυτό αναπτύχθηκε το 1960 από τον Buhlmann [κεφ. (2)] [βλέπε Whitney (1918)]. Σχετικά με την ομοιογένεια-ετερογένεια σε ένα χαρτοφυλάκιο εισάγεται η παράμετρος κινδύνου Θ . Έτσι, δοθείσης της τιμής της παραμέτρου Θ , θεωρούμε την δεσμευμένη κατανομή της τυχαίας μεταβλητής X . Η διασπορά κινδύνου εκφράζεται ως η διασπορά της δεσμευμένης μέσης τιμής $V[E(X|\Theta)]$ και ως η αναμενόμενη τιμή της ενδο-διασποράς, $E[V(X|\Theta)]$. Θεωρούμε ένα ετερογενές ασφαλιστικό χαρτοφυλάκιο το οποίο αποτελείται από ομάδες ομογενών κινδύνων. Κάθε ομάδα ομογενούς κινδύνου για την οποία έχουμε εμπειρία εξέλιξης ζημιών σε μια χρονική περίοδο t , με $t = 1, 2, \dots, T$, αναφέρεται ως ασφαλιστήριο συμβόλαιο (contract). Κάθε ασφαλιστήριο συμβόλαιο j χαρακτηρίζεται από μια παράμετρο κινδύνου Θ_j . Η παράμετρος Θ_{ij} περιγράφει τα χαρακτηριστικά κινδύνου του ασφαλιστηρίου συμβολαίου j ,

με $j = 1, 2, \dots, J$ σε μια χρονική περίοδο t . Υποθέτουμε δηλαδή πως όλες οι διαφορές εμπειρίας εξέλιξης ζημιών μεταξύ των ασφαλιστήριων συμβολαίων και μεταξύ χρονικών περιόδων προκαλούνται από τις παράμετρους κινδύνου.

$$\Theta_{11}, \Theta_{21}, \dots, \Theta_{T1}, \dots, \Theta_{1J}, \Theta_{2J}, \dots, \Theta_{TJ}.$$

Αν όλες οι παράμετροι κινδύνου είναι ίσες, τότε δεν υπάρχει λόγος να εφαρμόσουμε τεχνικές αξιοπιστίας. Στα περισσότερα μοντέλα αξιοπιστίας υποθέτουμε ότι οι Θ_{ij} είναι χρονικά ομογενείς, δηλαδή, δοθέντος του ασφαλιστήριου συμβολαίου η παράμετρος κινδύνου παραμένει αμετάβλητη στο χρόνο. Στην περίπτωση αυτή παραλείπουμε το δείκτη i και έχουμε

$$\Theta_1, \Theta_2, \dots, \Theta_J.$$

Συνήθως οι παράμετροι κινδύνου είναι άγνωστοι ή δεν μπορούν να παρατηρηθούν και στο εξής θα χρησιμοποιούνται ως τυχαίες μεταβλητές. Συνάρτηση κατανομής των τυχαίων μεταβλητών αυτών ορίζουμε την $U(\theta)$, η οποία ονομάζεται συνάρτηση δόμησης (structure function). Έτσι, τα ασφαλιστήρια συμβόλαια είναι όμοια μεταξύ τους με την έννοια ότι οι παράμετροι κινδύνου Θ_j για κάθε ασφαλιστήριο συμβόλαιο έχουν την ίδια συνάρτηση κατανομής $U(\theta)$, αλλά διαφέρουν στο ότι οι τιμές θ_j των Θ_j , αν και μη παρατηρούμενες, είναι διαφορετικές σε κάθε ασφαλιστήριο συμβόλαιο. Όπως οι Θ_j , έτσι και η συνάρτηση δόμησης $U(\theta)$ είναι δύσκολο να προσδιοριστεί. Αν και αυτό είναι ένα μειονέκτημα για τις περισσότερες στατιστικές μεθόδους, δεν συμβαίνει το ίδιο και με τις διαδικασίες αξιοπιστίας. Οι διαδικασίες στις οποίες δεν απαιτείται η κατανομή κινδύνου ονομάζονται διαδικασίες απαλλαγμένες κατανομών (distribution free) [βλέπε Bühlmann (1967)].

Ωστόσο η ομοιογένεια-ετερογένεια σε ένα σύνολο δεδομένων, όπως ένα χαρτοφυλάκιο με πλήθος ασφαλιστήριων συμβολαίων, σε όρους στατιστικής μπορεί να μεταφραστεί ως τη μεταβλητότητα που παρουσιάζεται στα δεδομένα αυτά. Μια παραδοσιακή και χρήσιμη μέθοδος που συνοψίζει τη μεταβλητότητα είναι γνωστή ως ανάλυση διασποράς (Analysis of Variance). Η μέθοδος αυτή δεν παρουσιάζει μόνο μια διαμέριση της παρατηρούμενης μεταβλητότητας, αλλά συνοψίζει επίσης τους υπολογισμούς που μας επιτρέπουν να εξετάσουμε, υπό ορισμένες προϋποθέσεις (κανονικότητας), τις σημαντικές διαφορές μεταξύ των μέσων των υποσυνόλων των δεδομένων. Ο Ronald Fisher, ο δημιουργός της ανάλυσης διασποράς, αναφέρει σε μια επιστολή του προς τον George Snedecor: "Η ανάλυση διασποράς είναι, όχι ένα μαθηματικό θεώρημα, αλλά μια απλή μέθοδος οργάνωσης αριθμητικών γεγονότων έτσι ώστε να απομονώνονται και να εμφανίζονται τα βασικά χαρακτηριστικά ενός συνόλου δεδομένων με την μεγαλύτερη δυνατή απλότητα".

Αρχικά η τεχνική της ανάλυσης διασποράς αναπτύχθηκε για την εξέταση των διαφορών μεταξύ των μέσων, ωστόσο αργότερα προσαρμόστηκε στην εκτίμηση των μεταβλητών διακύμανσης. Αν και οι περιγραφές του Fisher σχετίζονταν με τα αθροίσματα τετραγώνων των διαφορών

μεταξύ των παρατηρηθέντων μέσων, η τάση των τελευταίων δεκαετιών ήταν να παρουσιαστούν πολλές ιδέες της ανάλυσης διασποράς ως κομμάτι των γραμμικών μοντέλων.

Το σημείο εκίνησης του πίνακα ανάλυσης διασποράς του Fisher ήταν οι διαφορετικοί μέσοι όροι και οι διαφορετικές μέσες τιμές που προκύπτουν από ένα σύνολο δεδομένων. Για παράδειγμα εάν έχουμε τις τιμές X_{jt} που είναι το ύψος αποζημιώσεων ασφαλιστηρίων συμβολαίων j , την χρονική στιγμή t , θα έχουμε πλήθος σε μέσους \bar{X}_j , αλλά και τον ολικό μέσο $\bar{X}_{..}$. Διευκρινίζεται ότι:

$$\bar{X}_j = \frac{\sum_{t=1}^T X_{jt}}{T},$$

και

$$\bar{X}_{..} = \frac{\sum_{j=1}^J \sum_{t=1}^T X_{jt}}{JT},$$

όπου $j = 1, \dots, J$ και $t = 1, \dots, T$. Έτσι έχουμε ότι J είναι το πλήθος των συμβολαίων, και T οι παρατηρηθείσες τιμές (περίοδοι) σε κάθε J . Ξεκινώντας από την (1.2) έχουμε:

$$X_{jt} - \bar{X}_{..} = (X_{jt} - \bar{X}_j) + (\bar{X}_j - \bar{X}_{..}) \quad (1.2)$$

$$\sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{..})^2 = \sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_j)^2 + \sum_{j=1}^J \sum_{t=1}^T (\bar{X}_j - \bar{X}_{..})^2, \quad (1.3)$$

το οποίο προκύπτει καθώς το

$$\sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_j)(\bar{X}_j - \bar{X}_{..}) = 0. \quad (1.4)$$

Κάθε όρος της (1.3) είναι ανάλογος του αθροίσματος τετραγώνων που χρησιμοποιείται στη συνηθισμένη εκτίμηση διασποράς ενός τυχαίου δείγματος T παρατηρήσεων (x_1, x_2, \dots, x_T) , όπου:

$$s^2 = \frac{1}{T-1} \sum_{t=1}^T (x_t - \bar{x})^2.$$

Έτσι ο όρος στο αριστερό μέρος της (1.3) είναι το συνολικό άθροισμα των τετραγώνων της διαφοράς όλων των παρατηρήσεων από τον μέσο τους, και οι δύο όροι στο δεξί κομμάτι της (1.3) είναι το άθροισμα των τετραγώνων της διαφοράς των παρατηρήσεων από τον αντίστοιχο μέσο της ομάδας που ανήκουν, \bar{X}_j , και αυτών των μέσων των ομάδων από τον μέσο $\bar{X}_{..}$. Έτσι η (1.3) μπορούμε να πούμε πως αποτελεί μια διαμέριση, του συνολικού αθροίσματος

τετραγώνων σε δύο άλλα αθροίσματα τετραγώνων, η οποία δίνεται στον παρακάτω Πίνακα 1.1:

Πηγή Διασποράς	Άθροισμα Τετραγώνων
Μεταξύ των Ομάδων	$SSB = \sum_{j=1}^J \sum_{t=1}^T (\bar{X}_j - \bar{X}_{..})^2$
Εντός των Ομάδων	$SSW = \sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_j)^2$
Συνολικά	$SST = \sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{..})^2$

Πίνακας 1.1: Αθροισμάτων Τετραγώνων.

Υπό τις συνθήκες ανεξαρτησίας και κανονικότητας, ότι τα X_{jt} είναι τιμές των ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν κανονική κατανομή με $E(X_{jt}) = \mu_j$ και $Var(X_{jt}) = s^2$, ο Fisher απέδειξε ότι τα SSB και SSW είναι το κάθε ένα κατανομημένο ως πολλαπλάσιο μιας χ^2 κατανομής, στοχαστικά ανεξάρτητα και πως

$$F = \frac{SSB/(J-1)}{SSW/J(T-1)} \sim F_{J(T-1)}^{J-1},$$

που σημαίνει πως το F είναι κατανομημένο σύμφωνα με την κατανομή του Fisher με $(J-1)$ και $J(T-1)$ βαθμούς ελευθερίας για τον αριθμητή και παρανομαστή αντίστοιχα. Όλα τα παραπάνω συνοψίζονται στον Πίνακα 1.2. Η πιο απλή χρήση για την οποία ο Fisher σχεδίασε

Πηγή Διασποράς	Βαθμοί Ελ.	Αθρ. Τετραγ.	Μέσος Αθρ. Τετρ.	$F : H_0$
Μεταξύ των Ομάδων	$J-1$	SSB	$MSB = SSB/(J-1)$	$F = \frac{MSB}{MSW}$
Εντός των Ομάδων	$J(T-1)$	SSW	$MSW = SSW/J(T-1)$	
Συνολικά	$JT-1$	SST		

Πίνακας 1.2: Ανάλυση Διασποράς.

τον πίνακα ανάλυσης διασποράς είναι το στατιστικό F , το οποίο είναι ένα τεστ της υπόθεσης [βλέπε Searle et al. (2009)]:

- $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$,
- $H_1 : \Delta$ ύο, τουλάχιστον, από τους μέσους διαφέρουν.

με την H_0 να απορρίπτεται όταν $F > F_{J-1, J(T-1), 1-\alpha}$. Οι κρίσιμες τιμές του F , $F_{J-1, J(T-1), 1-\alpha}$, μπορούν να βρεθούν σε έναν F -πίνακα.

Η συσχέτιση μεταξύ της τεχνικής της Ανάλυσης Διασποράς και της μεθόδου Αξιοπιστίας έχει εξεταστεί και αναλυθεί σε αρκετές δημοσιεύσεις. Για παράδειγμα οι Morirs-Van Slyke (1978), καθώς και ο Venter (1990) αναλύουν ένα μοντέλο όμοιο με αυτό της ανάλυσης διασποράς για έναν παράγοντα.

Στο Κεφάλαιο 2 της εργασίας γίνεται παρουσίαση του μοντέλου αξιοπιστίας του Buhlmann (1967), ένα μοντέλο που εξετάζει συμβόλαια με ίδιο πλήθος παρεχόμενων πληροφοριών σε ολόκληρο τον πληθυσμό. Στο 3ο Κεφάλαιο έχουμε το μοντέλο των Buhlmann-Sraub (1970) όπου είναι δυνατή η διαφοροποίηση του πλήθους παρατηρήσεων ανά συμβόλαιο. Επιπλέον δίνεται η επιλογή η ακρίβεια (βαρύτητα) της κάθε παρατήρησης να είναι διαφορετική. Το βασικό χαρακτηριστικό του μοντέλου που εξετάζεται στο επόμενο κεφάλαιο, το Τέταρτο, είναι πως τα συμβόλαια σε ένα χαρτοφυλάκιο πρέπει να ομαδοποιούνται σύμφωνα με ποσοτικούς παράγοντες κινδύνου, που υποδηλώνουν την κάθε κατηγορία συμβάσεων. Πρόκειται για το μοντέλο αξιοπιστίας παλινδρόμησης, το οποίο εισήγαγε στην αναλογιστική βιβλιογραφία ο Hachemeister (1975). Τα τελευταία δύο κεφάλαια παρουσιάζουν μοντέλα στα οποία τα συμβόλαια των χαρτοφυλακίων ομαδοποιούνται με βάση ποιοτικούς παράγοντες κινδύνου. Αρχικά στο Κεφάλαιο 5 παρουσιάζεται το ιεραρχικό μοντέλο αξιοπιστίας του Jewell (1975). Σε μερικές περιπτώσεις τα χαρτοφυλάκια μπορούν να υποδιαιρεθούν σε κλάδους-τομείς οι οποίοι παρουσιάζουν μεγαλύτερη ομοιογένεια στο εσωτερικό τους. Το μοντέλο αυτό είναι μια άμεση επέκταση του μοντέλου των Buhlmann-Sraub (1970) που αναφέρεται σε τέτοιες περιπτώσεις. Αυτό όμως που δεν εξετάζει το προαναφερόμενο μοντέλο αξιοπιστίας είναι την πιθανότητα ύπαρξης αλληλεπίδρασης μεταξύ διαφόρων ποιοτικών παραγόντων κινδύνου με αδυναμία ιεράρχησης αυτών. Το κατάλληλο μοντέλο για μια τέτοια περίπτωση, είναι το μοντέλο αξιοπιστίας σταυρωτής ταξινόμησης που εξετάζεται στο 6ο Κεφάλαιο, και εισήχθη στη βιβλιογραφία από τον Dannenburg (1996).

Στο σύνολο των μοντέλων της παρούσας εργασίας γίνεται προσέγγιση με τη χρήση της μεθόδου της ανάλυσης διασποράς (Ανονα) με σκοπό την εξέταση της μεταβλητότητας που παρουσιάζουν τα δεδομένα και ερευνάται η συσχέτιση της μεθόδου αυτής με τα προαναφερόμενα μοντέλα αξιοπιστίας. Επιπλέον σε κάθε κεφάλαιο γίνεται επίλυση αριθμητικού παραδειγμάτος και με τις δύο μεθόδους στο στατιστικό πακέτο R.

Κεφάλαιο 2

Το Μοντέλο Αξιοπιστίας του Buhlmann

2.1 Μοντέλο Buhlmann

Για να γίνουν κατανοητές κύριες έννοιες στις οποίες βασίζεται η θεωρία αξιοπιστίας, στο κεφάλαιο αυτό θα εξετάσουμε ένα βασικό μοντέλο αξιοπιστίας. Πιο συγκεκριμένα θα παρουσιαστεί το ισορροπημένο μοντέλο του Buhlmann. Ισορροπημένο (Balanced) καθώς τα δεδομένα μπορούν να χαρακτηριστούν με διάφορους τρόπους που θα εξαρτώνται από το εάν κάθε κελί περιέχει τον ίδιο αριθμό παρατηρήσεων ή όχι. Όταν αυτοί οι αριθμοί είναι ίδιοι, τα δεδομένα πρέπει να περιγράφονται ως ισορροπημένα δεδομένα. Όσα δεδομένα δεν είναι ισορροπημένα (Balanced Data), περιγράφονται ως μη ισορροπημένα (Unbalanced Data) [βλέπε Searle et al. (2009)].

Για ευκολία παίρνουμε ως δεδομένο πως η κάθε ομάδα συμβολαίων περιέχει μόνο ένα συμβολαίο, και πως κάθε ομάδα έχει παρατηρηθεί κατά τη διάρκεια T περιόδων.

Θ_1	Θ_2	...	Θ_J
X_{11}	X_{21}	...	X_{J1}
X_{12}	X_{22}	...	X_{J2}
X_{13}	X_{23}	...	X_{J3}
.
.
.
X_{1T}	X_{2T}	...	X_{JT}

Πίνακας 2.1: Αξιώσεις των συμβολαίων J σε T χρονικές στιγμές.

Επομένως για κάθε j η μεταβλητή t παίρνει τιμές $t = 1, 2, \dots, T$. Έτσι λοιπόν θεωρούμε

την τυχαία μεταβλητή X_{jt} , η οποία αντιπροσωπεύει την αξίωση που έχουμε από μια ομάδα συμβολαίων j , με $j = 1, 2, \dots, J$, για το έτος t , καθώς και την παράμετρο κινδύνου Θ_j . Από την παράμετρο κινδύνου Θ_j εξαρτάται η κατανομή της X_{jt} . Εάν για παράδειγμα η τυχαία μεταβλητή X_{1t} ακολουθεί την κατανομή Poisson με παράμετρο λ , η θ_j , πραγματική τιμή της Θ_j , θα μπορούσε να είναι η παράμετρος λ της κατανομής Poisson. Ακόμα έχουμε το τυχαίο διάνυσμα $(\Theta_j, X_{j1}, X_{j2}, \dots, X_{jT})$ και το υπό δέσμευση διάνυσμα $(X_{j1}, X_{j2}, \dots, X_{jT}) | \Theta_j = \theta_j$. Οι υπό δέσμευση αναμενόμενες τιμές των ασφαλιστηρίων συμβολαίων είναι

$$E(X_{jt} | \Theta_j)$$

και η

$$E(X_{jt}) = E[E(X_{jt} | \Theta_j)]$$

είναι η μέση τιμή των αναμενόμενων τιμών των ασφαλιστηρίων συμβολαίων.

2.2 Υποθέσεις

1. Τα ασφαλιστήρια συμβόλαια είναι ανεξάρτητα και ισόνομα καταναμεμένα και οι τυχαίες μεταβλητές Θ_j είναι ανεξάρτητες και ισόνομα καταναμεμένες.
2. Δοθέντος Θ_j , οι τυχαίες μεταβλητές X_{jt} είναι υπό δέσμευση ανεξάρτητες και ισόνομα καταναμεμένα.
3. $E(X_{jt} | \Theta_j) = \mu(\Theta_j)$ για όλα τα $j = 1, \dots, J$ και $t = 1, \dots, T$.
4. $Cov(X_{jr}, X_{jt} | \Theta_j) = \delta_{rt} \sigma^2(\Theta_j)$, με $t, r = 1, 2, \dots, T$ όπου δ_{rt} είναι το σύμβολο του Kronecker, το οποίο ισούται με ένα στην περίπτωση που $r = t$ και με μηδέν σε διαφορετική περίπτωση [βλέπε Bühlmann (1970)].

2.3 Συμβολισμοί και Παράμετροι Δόμησης

Στο επόμενο θεώρημα, ψάχνουμε για μια πρόβλεψη της τυχαίας, μέχρι στιγμής, μη παρατηρούμενης μεταβλητής $X_{j,T+1}$. Απαιτείται αυτή η πρόβλεψη να είναι ένας γραμμικός συνδυασμός των παρατηρήσιμων δεδομένων X_{11}, \dots, X_{JT} και με τον ίδιο μέσο όρο του $X_{j,T+1}$. Αποδεικνύεται [βλέπε Kaas et al. (2008)], πως ο εκτιμητής αυτός έχει τη μορφή του γενικού τύπου της θεωρίας αξιοπιστίας (1.1), οπότε είναι ένας σταθμισμένος μέσος

όρος της εμπειρίας των μεμονωμένων απαιτήσεων και της συνολικής μέσης απαίτησης. Το θεώρημα, μας παρέχει επιπλέον τη βέλτιστη τιμή του συντελεστή αξιοπιστίας Z . Ο λόγος που θέλουμε να γνωρίζουμε τον βέλτιστο εκτιμητή του ποσού που θα πληρωθεί στην επόμενη περίοδο $T + 1$, είναι πως αυτό είναι το ασφάλιστρο που θα πρέπει να χρεωθεί για την παρούσα σύμβαση. Σύμφωνα με τις υποθέσεις και τους συμβολισμούς, ο ζητούμενος εκτιμητής ταυτίζεται με το εκτιμητή αξιοπιστίας της $\mu(\Theta_j)$.

Θεώρημα 1. Έστω ότι στο χαρτοφυλάκιο μας ισχύουν οι υποθέσεις 3 και 4 του μοντέλου. Τότε οι βέλτιστες μη ομογενείς εκτιμήτριες αξιοπιστίας $\mu(\Theta_j)^{cred}$, $j = 1, \dots, J$ της $\mu(\Theta_j)$:

$$\mu(\Theta_j)^{cred} = Z\bar{X}_{\cdot j} + (1 - Z)\mu \quad (2.1)$$

όπου:

$$Z = \frac{\sigma_\alpha^2 T}{\sigma_\alpha^2 T + s^2} = \frac{T}{T + \frac{s^2}{\sigma_\alpha^2}}, \quad (2.2)$$

είναι ο προκύπτων καλύτερος συντελεστής αξιοπιστίας (ο οποίος στην περίπτωση αυτή είναι ίσος για όλα τα j), $\bar{X}_{\cdot j} = \frac{1}{JT} \sum_{j=1}^J \sum_{t=1}^T X_{jt}$ είναι ο συλλογικός εκτιμητής του μ , και $\bar{X}_{\cdot j}$ είναι ο ατομικός εκτιμητής του μ .

Οι μ , s^2 , σ_α^2 , είναι παράμετροι στο υπό θεώρηση μοντέλο μας και ονομάζονται παράμετροι δόμησης (structural parameters). Για αυτούς ισχύουν τα παρακάτω:

1. $\mu = E(X_{jt}) = E[\mu(\Theta_j)]$
2. $s^2 = E[\sigma^2(\Theta_j)]$
3. $\sigma_\alpha^2 = V[\mu(\Theta_j)]$

[βλέπε Bühlmann (1967)].

2.4 Σχέσεις Συνδιασπορών και Μέσες Τιμές

Θεώρημα 2. Ισχύουν:

- $E(X_{jt}) = E(\bar{X}_{\cdot t}) = E(\bar{X}_{\cdot j}) = E[\mu(\Theta_j)] = \mu$
- $Cov(X_{jr}, X_{jt}) = \delta_{rt}s^2 + \sigma_\alpha^2$
- $Cov(X_{jt}, \bar{X}_{\cdot j}) = Cov(\bar{X}_{\cdot j}, \bar{X}_{\cdot j}) = \frac{s^2}{T} + \sigma_\alpha^2$

- $Cov(X_{jt}, \bar{X}_{..}) = Cov(\bar{X}_{j.}, \bar{X}_{..}) = Cov(\bar{X}_{..}, \bar{X}_{..}) = (\frac{s^2}{T} + \sigma_\alpha^2)/J$
- $Cov(X_{jt}, \mu(\Theta_j)) = Cov(\bar{X}_{j.}, \mu(\Theta_j)) = \sigma_\alpha^2$
- $Cov(\bar{X}_{..}, \mu(\Theta_j)) = \frac{\sigma_\alpha^2}{J}$

Απόδειξη: [βλέπε Pitselis (2018)].

2.5 Αμερόληπτοι Εκτιμητές Παραμέτρων

Αμερόληπτοι εκτιμητές των παραμέτρων μ , s^2 , και α είναι αντίστοιχα οι κάτωθι:

$$\hat{\mu} = \bar{X}_{..},$$

$$\hat{s}^2 = \frac{1}{J} \sum_{j=1}^J \hat{s}_j^2,$$

όπου

$$\hat{s}_j^2 = \frac{1}{(T-1)} \sum_{t=1}^T (X_{jt} - \bar{X}_{.t})^2$$

και

$$\hat{\sigma}_\alpha^2 = \frac{1}{J-1} \sum_{j=1}^J (\bar{X}_{j.} - \bar{X})^2 - \frac{1}{T} \hat{s}^2.$$

[βλέπε Goovaerts et al. (1990)].

2.6 Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς

Όπως αναφέρθηκε και στην εισαγωγή, η τεχνική της ανάλυσης της διασποράς αναπτύχθηκε αρχικά για την εξέταση των διαφορών μεταξύ των μέσων, ωστόσο στη συνέχεια χρησιμοποιήθηκε για την εκτίμηση των παραμέτρων διασποράς.

Αρχικά θεωρούμε πως από στατιστικής πλευράς η αξίωση αυτή (ασφάλιστρο) είναι το άθροισμα του μέσου μ_j για κάθε ομάδα συμβολαίων προσθέτοντας τον όρο που εκφράζει το σφάλμα, καθώς και πως όλα τα X_{jt} είναι ανεξάρτητα και $X_{jt} \sim N(\mu_j, s^2)$ (κανονικά κατανομημένα).

Πιθανόν οι τιμές μ_j , μέσοι, για κάθε ομάδα j να είναι άνισες, αλλά έχουν την ίδια διασπορά s^2 . Υπάρχει η δυνατότητα ελέγχου μεταξύ των μέσων της κάθε ομάδας συμβολαίων με τη χρήση της μεθόδου της ανάλυσης διασποράς (Analysis of Variance) που αναφέρθηκε στην εισαγωγή. Στην περίπτωση που η αρχική υπόθεση, ότι όλα τα μ_j είναι ίσα, δεν ισχύει οδηγούμαστε στο συμπέρασμα πως υπάρχει μεγαλύτερη διασπορά μεταξύ των μέσων τιμών \bar{X}_j και του γενικού μέσου όρου $\bar{X}_{..}$ από αυτή που αναμέναμε. Για το λόγο αυτό εξετάζουμε την τυχαία μεταβλητή SSB (sum of squares between):

$$SSB = \sum_{j=1}^J \sum_{t=1}^T (\bar{X}_{j.} - \bar{X}_{..})^2 = T \sum_{j=1}^J (\bar{X}_{j.} - \bar{X}_{..})^2. \quad (2.3)$$

Κάτω από τη μηδενική υπόθεση πως όλες οι ομάδες συμβολαίων έχουν τον ίδιο μέσο μ_j , μπορεί να αποδειχθεί ότι η τυχαία μεταβλητή SSB έχει μέση τιμή $(J-1)s^2$. Δεδομένου ότι το s^2 είναι άγνωστο, πρέπει να εκτιμήσουμε τη παράμετρο αυτή ξεχωριστά. Η εκτίμηση αυτή γίνεται από το παρακάτω άθροισμα τετραγώνων που συμβολίζεται SSW (sum of squares within):

$$SSW = \sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{j.})^2. \quad (2.4)$$

Είναι εύκολο να αποδειχθεί πως η τυχαία μεταβλητή SSW έχει μέσο $J(T-1)s^2$. Διαιρώντας το SSB με $J-1$ και το SSW με $J(T-1)$ παίρνουμε δύο τυχαίες μεταβλητές, κάθε μία με μέση τιμή s^2 , που αντίστοιχα τις ονομάζουμε MSB (mean square between) και MSW (mean square within). Στη συνέχεια μπορούμε να πραγματοποιήσουμε ένα F-test, όπου συγκρίνοντας τις μεγάλες τιμές του MSB με του MSW να καταλήξουμε σχετικά με τη μηδενική υπόθεση, ότι όλοι οι μέσοι των ομάδων συμβολαίων είναι ίσοι, εάν θα απορριφθεί ή όχι. Ο στατιστικός έλεγχος που πρέπει να χρησιμοποιηθεί είναι ο λεγόμενος F-ratio ή έλεγχος αναλογίας διακύμανσης:

$$F = \frac{MSB}{MSW} = \frac{\frac{1}{J-1} \sum_{j=1}^J \sum_{t=1}^T (\bar{X}_{j.} - \bar{X}_{..})^2}{\frac{1}{J(T-1)} \sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{j.})^2}. \quad (2.5)$$

Επίσης, δεδομένης της μηδενικής υπόθεσης, το SSB διαιρούμενο από s^2 έχει κατανομή $\chi^2(J-1)$, ενώ το SSW διαιρούμενο από s^2 έχει κατανομή $\chi^2(J(T-1))$. Επιπροσθέτως, είναι εφικτό ναδειχθεί πως οι τυχαίες μεταβλητές αυτές είναι ανεξάρτητες. Επομένως, η αναλογία (ratio) F έχει μια $F(J-1, J(T-1))$ κατανομή.

Εάν η μηδενική υπόθεση, πως όλοι οι μέσοι είναι ίσοι, αποτύχει να απορριφθεί δεν θα υπάρχει κανένα στατιστικό στοιχείο ότι το χαρτοφυλάκιο είναι ανομοιογενές. Επομένως δεν θα υπάρχει λόγος για να μην θεωρήσουμε το ίδιο ασφάλιστρο για κάθε σύμβαση. Εάν ωστόσο απορριφθεί, προφανώς θα υπάρχει διαφοροποίηση μεταξύ των μέσων μ_j . Σε μια τέτοια περίπτωση οι αριθμοί αυτοί θα μπορούσαν να θεωρηθούν σταθεροί άγνωστοι και να γίνει προσπάθεια να υπολογιστούν μέσω ενός συστήματος μεταξύ αυτών των αριθμών, για παράδειγμα κάνοντας μια παλινδρόμηση στα συσχετισμένα δεδομένα. Μια άλλη προσέγγιση είναι να θεωρηθεί πως οι αριθμοί αυτοί (μ_j) έχουν παραχθεί από ένα τυχαίο μηχανισμό, με σφάλμα παρόμοιο με αυτό που είναι υπεύθυνο για τις αποκλίσεις από τον μέσο όρο μέσα σε κάθε ομάδα συμβολαίων.

Έτσι, υποθέτουμε ότι η αξίωση X_{jt} για την ομάδα j στην περίοδο t μπορεί να γραφεί ως το άθροισμα των στοχαστικά ανεξάρτητων στοιχείων, ως εξής:

$$X_{jt} = \mu + \alpha_j + \varepsilon_{jt}, \quad j = 1, \dots, J \quad t = 1, \dots, T,$$

όπου μ είναι ο γενικός μέσος των μ_j και οι τυχαίες μεταβλητές α_j είναι ανεξάρτητες και ισόνομα κατανομημένες, με $E[\alpha_j] = 0$ και $Var[\alpha_j] = \sigma_a^2$. Επίσης οι τυχαίες μεταβλητές ε_{jt} είναι ανεξάρτητα και ισόνομα κατανομημένες με $E[\varepsilon_{jt}] = 0$ και $Var[\varepsilon_{jt}] = \sigma_\varepsilon^2 = s^2$ για κάθε j και t . Επιπλέον το α_j δηλώνει την απόκλιση για το συμβόλαιο j από τον μέσο μ , και το ε_{jt} είναι η απόκλιση της τιμής του συμβολαίου j για t από τον μέσο του συμβολαίου μ_j [βλέπε Kaas et al. (2008)].

Από τα παραπάνω συνεπάγονται τα εξής:

$$s^2 = Var[\varepsilon_{jt}] = E[\varepsilon_{jt}^2] - E[\varepsilon_{jt}]^2 = E[\varepsilon_{jt}^2] - 0 = E[\varepsilon_{jt}^2].$$

Ομοίως προκύπτει πως $\sigma_a^2 = Var[\alpha_j] = E[\alpha_j^2]$ [βλέπε Dannenburg et al. (1996)].

Θεώρημα 3. Σύμφωνα με τις σχέσεις (2.3), (2.4), καθώς και το Θεώρημα (2), για το ισορροπημένο μοντέλο οι αμερόληπτοι εκτιμητές των s^2 και σ_α^2 είναι:

$$\widehat{s}^2 = MSW,$$

$$\widehat{\sigma}_\alpha^2 = \frac{MSB - MSW}{T}.$$

[βλέπε *Kaas et al. (2008)*]

Απόδειξη: Στην περίπτωση του ισορροπημένου μοντέλου η ανάλυση διασποράς δίνεται από τον παρακάτω πίνακα:

Πηγή Διασποράς	Βαθμοί Ελευθ.	Αθρ. Τετρ.	Μέσος Αθρ. Τετρ.
Μεταξύ των Ομάδων	$J - 1$	$SSB = \sum_{j=1}^J \sum_{t=1}^T (\bar{X}_{j.} - \bar{X}_{..})^2$	MSB
Εντός των Ομάδων	$J(T - 1)$	$SSW = \sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{j.})^2$	MSW
Συνολικά	$JT - 1$	$SST = \sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{..})^2$	-

Πίνακας 2.2: Ανάλυση διασποράς ισορροπημένων δεδομένων στο μοντέλο του Buhlmann.

Παίρνοντας τις μέσες τιμές των μέσων αθροισμάτων τετραγώνων έχουμε,

$$\begin{aligned}
 E(\widehat{s}^2) &= E(MSW) \\
 &= \frac{E(SSW)}{J(T-1)} \\
 &= \frac{E(\sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{j.})^2)}{J(T-1)} \\
 &= \frac{E[\sum_{j=1}^J \sum_{t=1}^T (\mu + \alpha_j + \varepsilon_{jt} - \mu - \alpha_j - \varepsilon_{j.})^2]}{J(T-1)} \\
 &= \frac{E[\sum_{j=1}^J \sum_{t=1}^T (\varepsilon_{jt} - \varepsilon_{j.})^2]}{J(T-1)}
 \end{aligned} \tag{2.11}$$

$$\begin{aligned}
&= \frac{\sum_{j=1}^J E[\sum_{t=1}^T (\varepsilon_{jt} - \varepsilon_j)^2]}{J(T-1)} \\
&= \frac{\sum_{j=1}^J (\sum_{t=1}^T E(\varepsilon_{jt})^2 - TE(\varepsilon_j)^2)}{J(T-1)} \\
&= \frac{\sum_{j=1}^J T(\mu^2 + s^2) - \frac{T(T\mu^2 + s^2)}{T}}{J(T-1)} \\
&= \frac{\sum_{j=1}^J (T-1)s^2}{J(T-1)} \\
&= s^2.
\end{aligned}$$

Άρα από τη μέση τιμή της σχέσης (2.11) προκύπτει ότι

$$\begin{aligned}
\hat{s}^2 &= \frac{\sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_j)^2}{J(T-1)} \\
&= \frac{SSW}{J(T-1)} \\
&= MSW.
\end{aligned}$$

Με τον ίδιο τρόπο έχουμε και την απόδειξη του αμέροληπτου εκτιμητή για το σ_α^2 .

$$\begin{aligned}
E(MSB) &= \frac{E(SSB)}{J-1} \\
&= \frac{T(E[\sum_{j=1}^J (\bar{X}_j - \bar{X}_{..})^2])}{J-1} \\
&= \frac{T(E[\sum_{j=1}^J (\mu + \alpha_j + \bar{\varepsilon}_j - \mu - \bar{\alpha}_{..} - \bar{\varepsilon}_{..})^2])}{J-1} \\
&= \frac{T(E[\sum_{j=1}^J (\alpha_j + \bar{\varepsilon}_j - \bar{\alpha}_{..} - \bar{\varepsilon}_{..})^2])}{J-1} \\
&= \frac{T}{J-1} \sum_{j=1}^J [E(\alpha_j - \bar{\alpha}_{..})^2 + E(\bar{\varepsilon}_j - \bar{\varepsilon}_{..})^2]
\end{aligned} \tag{2.24}$$

$$\begin{aligned}
&= \frac{T}{J-1} \sum_{j=1}^J [\text{Var}(\alpha_j - \bar{\alpha}_.) + \text{Var}(\bar{\varepsilon}_j - \bar{\varepsilon}_.)] \\
&= \frac{T}{J-1} \left[\sum_{j=1}^J (\sigma_a^2 - 2\frac{\sigma_a^2}{J} + \frac{\sigma_a^2}{J}) + \sum_{j=1}^J (\frac{s^2}{T} - 2\frac{Ts^2}{JT^2} + \frac{s^2}{JT}) \right] \\
&= \frac{T}{J-1} \left[\sum_{j=1}^J (1 - \frac{1}{J})\sigma_a^2 + \sum_{j=1}^J (\frac{1}{T} - \frac{1}{JT})s^2 \right] \\
&= \frac{T}{J-1} \left(\frac{J(J-1)}{J}\sigma_a^2 + \frac{J}{T} \frac{J-1}{J} s^2 \right) \\
&= T\sigma_a^2 + s^2.
\end{aligned}$$

Επομένως, από την μέση τιμή της (2.24) έχουμε:

$$\begin{aligned}
T \frac{\sum_{j=1}^J (\bar{X}_j - \bar{X}_{..})^2}{J-1} &= T\hat{\sigma}_a^2 + \hat{s}^2 \\
\frac{\sum_{j=1}^J (\bar{X}_j - \bar{X}_{..})^2}{J-1} &= \hat{\sigma}_a^2 + \frac{\hat{s}^2}{T}.
\end{aligned}$$

Άρα

$$\begin{aligned}
\hat{\sigma}_a^2 &= \frac{\sum_{j=1}^J (\bar{X}_j - \bar{X}_{..})^2}{J-1} - \frac{\hat{s}^2}{T} \\
&= \frac{\frac{SSB}{J-1} - \hat{s}^2}{T} \\
&= \frac{MSB - MSW}{T}.
\end{aligned}$$

Επιπλέον, κάνοντας χρήση των σχέσεων (2.2) και (2.5) έχουμε τον συντελεστή αξιοπιστίας

$$\begin{aligned}
Z &= \frac{T}{T + \frac{s^2}{\sigma_a^2}} \\
&= \frac{T}{T + \frac{MSW}{\frac{MSB - MSW}{T}}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{T}{T + T \frac{MSW}{MSB - MSW}} \\
&= \frac{1}{1 + \frac{MSW}{MSB - MSW}} \\
&= \frac{1}{1 + \frac{\frac{MSW}{MSW}}{\frac{MSB - MSW}{MSW}}} \\
&= \frac{1}{1 + \frac{1}{F - 1}} \\
&= \frac{F - 1}{F}.
\end{aligned}$$

Άρα

$$Z = 1 - \frac{1}{F}. \quad (2.45)$$

Παρατήρηση: Σημειώνεται πως δεν εξετάστηκε η ανάλυση διασποράς για την περίπτωση μη ισορροπημένων δεδομένων, καθώς όπως αναφέρθηκε στην παρουσίαση, στην παράγραφο (2.1), στο μοντέλο του Buhlmann τα δεδομένα σε κάθε κελί (συμβολαίο) είναι ίδιου πλήθους.

2.7 Αριθμητικό Παράδειγμα στο Μοντέλο του Buhlmann

Στην παράγραφο αυτή παρουσιάζεται ένα αριθμητικό παράδειγμα, το οποίο επιλύεται τόσο με τη μεθοδολογία της ανάλυσης διασποράς καθώς και σύμφωνα με το μοντέλο του Buhlmann.

Έστω ότι έχουμε τις παρακάτω παρατηρήσεις για 3 ομάδες συμβολαίων ($j = 3$) σε περίοδο 5 ετών ($t = 5$) [βλέπε Dannenburg et al. (1996)]. Ζητούμενο, ο υπολογισμός του ποσού που θα πρέπει να καταβληθεί για το κάθε συμβολαίο j την χρονική περίοδο $T + 1$.

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	\bar{X}_j
$j = 1$	99.3	93.7	103.9	92.5	110.6	100.0
$j = 2$	112.5	108.3	118.0	99.4	111.8	110.0
$j = 3$	129.2	140.9	108.3	105.0	116.6	120.0

Πίνακας 2.3: Δεδομένα αριθμητικού παραδείγματος στο μοντέλο του Buhlmann.

Ανάλυση Διασποράς: Από τους υπολογισμούς σύμφωνα με τον πίνακα της ανάλυσης διασποράς, τους τύπους (2.3), (2.4) και (2.5) [βλέπε Κώδικα (Α'.1)] έχουμε:

$$MSB = \frac{\sum_{i=1}^I \sum_{j=1}^J (\bar{X}_{i.} - \bar{X}_{..})^2}{I - 1} = 500$$

$$MSW = \frac{\sum_{j=1}^J \sum_{t=1}^T (X_{jt} - \bar{X}_{j.})^2}{I(J - 1)} = 108.97$$

και

$$F = \frac{MSB}{MSW} = 4.588418831$$

Παρατήρηση 1.: Έχουμε λοιπόν πως το MSB ισούται με 500 με 2 βαθμούς ελευθερίας, ενώ το MSW είναι ίσο με 108.97 και 12 βαθμούς ελευθερίας. Η κρίσιμη τιμή $F_{2,12,1-0.95}$ είναι ίση με 3.89. Έτσι καταλήγουμε στο συμπέρασμα πως οι μέσες τιμές αξιώσεων σε κάθε ομάδα συμβολαίων δεν είναι όλες ίσες. Σε περίπτωση που η μηδενική υπόθεση, πως όλοι οι μέσοι είναι ίσοι, δεν απορριπτόταν, τότε το χαρτοφυλάκιο μας θα ήταν ομοιογενές και ως ασφάλιστρο για όλες τις ομάδες συμβολαίων θα οριζόταν ο γενικός μέσος των δεδομένων.

Σύμφωνα με το Θεώρημα (3), που αναφέρεται στο ισορροπημένο μοντέλο, και από τα αποτελέσματα της ανάλυσης διασποράς, οι εκτιμητές των παραμέτρων θα είναι:

$$\hat{\mu} = 110$$

$$\hat{s}^2 = MSW = 108.97$$

$$\hat{\sigma}_\alpha^2 = \frac{MSB - MSW}{T} = 78.206$$

Μοντέλο Buhlmann: Σύμφωνα με το μοντέλο του Buhlmann παίρνουμε τα εξής αποτελέσματα [βλέπε Κώδικα (Α'.1)]:

$$\hat{\mu} = 110 \quad \hat{s}_1^2 = 56$$

$$\hat{s}_2^2 = 47.185 \quad \hat{s}_3^2 = 223.725$$

$$\hat{s}^2 = 108.97 \quad \hat{\sigma}_\alpha^2 = 78.206$$

Παρατήρηση 2.: Από τα παραπάνω αποτελέσματα γίνεται αντιληπτό πως ο υπολογισμός των παραμέτρων δόμησης και με τις δύο μεθόδους οδήγησε σε ίδια αποτελέσματα. Τέλος κάνοντας χρήση των σχέσεων (2.45) και (2.1) προκύπτουν ο συντελεστής αξιοπιστίας Z , καθώς και οι εκτιμήσεις ζημιών του επόμενου έτους για κάθε συμβόλαιο j , δηλαδή τα ασφάλιστρα.

$$\hat{Z} = 1 - \frac{1}{F} = 0.78206$$

και

$$\hat{X}_{1,T+1}^{Cred} = 102.1794$$

$$\hat{X}_{2,T+1}^{Cred} = 110$$

$$\hat{X}_{3,T+1}^{Cred} = 117.8206$$

Κεφάλαιο 3

Το Μοντέλο Αξιοπιστίας των Buhlmann – Straub

Στο προηγούμενο κεφάλαιο παρουσιάστηκε το μοντέλο ισορροπημένων δεδομένων (Balanced) του Buhlmann. Εκείνο το μοντέλο είναι ισορροπημένο, όπως περιγράφεται και στην παράγραφο (2.1), καθώς το πλήθος των παρατηρήσεων για κάθε συμβόλαιο ήταν T (χρόνια) και η διασπορά κάθε παρατήρησης ήταν ίδια. Στο κεφάλαιο αυτό θα παρουσιαστεί το μοντέλο Buhlmann-Straub [βλέπε Buhlmann and Straub (1970)], που είναι μια γενίκευση του κλασικού μοντέλου αξιοπιστίας του Buhlmann (1969) που εξετάσαμε ήδη.

3.1 Μοντέλο Buhlmann-Straub

Δημιουργήθηκε από τους Buhlmann και Straub ως μέσο κοστολόγησης των αντασφαλιστικών συμβάσεων. Από τότε, το μοντέλο έχει χρησιμοποιηθεί ευρέως στην αντασφάλιση και στις ασφάλειες αυτοκινήτων [βλέπε Goulet (1998)].

Στο μοντέλο αυτό, τα συμβόλαια παύουν να είναι στη λογική του μοντέλου του Buhlmann. Πιο συγκεκριμένα υπάρχει η δυνατότητα το πλήθος παρατηρήσεων για κάθε συμβόλαιο να διαφέρει, επομένως θα έχουμε πλέον T_j . Μια ακόμα αλλαγή είναι πως συσχετίζονται διαφορετικές διακυμάνσεις στα συμβόλαια, π.χ. επιτρέπουμε η ακρίβεια κάθε παρατήρησης να είναι διαφορετική. Αυτές οι διαφορετικές διακυμάνσεις, που μπορεί να διαφέρουν στο χρόνο, εξελίσσονται φυσιολογικά όταν οι κίνδυνοι στα συμβόλαια αντικαθίστανται από μέσους αναγνωρισμένων κινδύνων. Σε αυτή την περίπτωση η διασπορά είναι αντιστρόφως ανάλογη του πλήθους αυτών των κινδύνων (φυσικά βάρη). Εάν η πληροφορία (φυσικά βάρη) δεν είναι διαθέσιμη, τότε είναι δυνατή η χρήση ως δεδομένου μιας προσέγγισης της ολικής έκθεσης. Παραδείγματα είναι το ύψος ασφαλιστρού ενός γκρουπ συμβολαίων τη δεδομένη χρονική περίοδο ή ακόμα και το ύψος σε μια προηγούμενη περίοδο. Όταν οι κίνδυνοι αντιπροσωπεύουν ποσοστά ζημιών, και έτσι οι απαιτήσεις διαιρούνται από τα ασφάλιστρα, είναι αναμενόμενο οι διακυμάνσεις των αξιώσεων να είναι ανάλογες των μέσων τους καθώς επίσης είναι λογικό

να λάβουμε τα ασφάλιστρα ως βάρη. Επιπλέον πιθανό είναι ένας αναλογιστής να θέλει να εισάγει τα δικά του βάρη (w) σε ορισμένες παρατηρήσεις, εάν θεωρεί κάποιες από τις παρατηρήσεις πιο αξιόπιστες [βλέπε Dannenburg et al. (1996)].

Θ_1	Θ_2	...	Θ_J
X_{11}, w_{11}	X_{21}, w_{21}	...	X_{J1}, w_{J1}
X_{12}, w_{12}	X_{22}, w_{22}	...	X_{J2}, w_{J2}
X_{13}, w_{13}	X_{23}, w_{23}	...	X_{J3}, w_{J3}
.
.
.
X_{1T_1}, w_{1T_1}	X_{2T_2}, w_{2T_2}	...	X_{JT_J}, w_{JT_J}

Πίνακας 3.1: Σταθμισμένες παρατηρήσεις των J συμβολαίων σε T χρονικές στιγμές.

3.2 Υποθέσεις

1. Τα ασφαλιστήρια συμβόλαια είναι ανεξάρτητα και ισόνομα και οι τυχαίες μεταβλητές Θ_j είναι ανεξάρτητες και ισόνομα κατανεμημένες.
2. Δοθέντος Θ_j , οι τυχαίες μεταβλητές X_{jt} είναι υπό δέσμευση ανεξάρτητες και ισόνομες.
3. $E(X_{jt}|\Theta_j) = \mu(\Theta_j)$, για όλα τα $j = 1, \dots, J$ και $t = 1, \dots, T_j$
4. $Cov(X_{rt}, X_{jt}|\Theta_j) = \delta_{jt} \frac{1}{w_{jt}} \sigma^2(\Theta_j)$, με $j, r = 1, 2, \dots, J$, όπου δ_{jr} είναι το σύμβολο του *Kroneker*.

[βλέπε Bühlmann and Gisler (2006)].

3.3 Συμβολισμοί και Παράμετροι Δόμησης

Όπως στην παράγραφο (2.3), έτσι και εδώ δίνεται το θεώρημα για τον βέλτιστο εκτιμητή του ποσού που θα πληρωθεί στην επόμενη περίοδο $T + 1$, καθώς και τη βέλτιστη τιμή του συντελεστή αξιοπιστίας Z_j . Σύμφωνα με τις υποθέσεις και τους συμβολισμούς, ο ζητούμενος εκτιμητής του ποσού που θα πληρωθεί στην επόμενη περίοδο ταυτίζεται με το εκτιμητή αξιοπιστίας της $\mu(\Theta_j)$.

Θεώρημα 4. Έστω ότι στο χαρτοφυλάκιο μας ισχύουν οι υποθέσεις 3 και 4. Τότε οι βέλτιστες μη ομογενείς εκτιμήτριες αξιοπιστίας $\mu(\Theta_j)^{Cred}$, $j = 1, \dots, J$ της $\mu(\Theta_j)$ για σταθερό j , είναι

$$\mu(\Theta_j)^{Cred} = Z_j X_{jw} + (1 - Z_j)\mu,$$

όπου $X_{jw} = \sum_{t=1}^{T_j} \frac{w_{jt}}{w_j} X_{jt}$ είναι ο ατομικός εκτιμητής της $\mu(\Theta_j)$ και Z_j είναι ο συντελεστής αξιοπιστίας για το ασφαλιστήριο συμβόλαιο j που δίνεται από τη σχέση:

$$\begin{aligned} Z_j &= \frac{\sigma_\alpha^2 w_j}{\sigma_\alpha^2 w_j + s^2} \\ &= \frac{w_j}{w_j + \frac{s^2}{\sigma_\alpha^2}}, \end{aligned}$$

όπου

$$w_j = \sum_{t=1}^{T_j} w_{jt}.$$

Επιπλέον συμβολισμοί είναι

$$\begin{aligned} w_{..} &= \sum_{j=1}^J w_j, & X_{ww} &= \sum_{j=1}^J \frac{w_j}{w_{..}} X_{jw}, \\ X_{zw} &= \sum_{j=1}^J \frac{Z_j}{Z_{..}} X_{jw}, & Z_{..} &= \sum_{j=1}^J Z_j. \end{aligned}$$

Ακόμα, μ , σ_α^2 και s^2 είναι οι παράμετροι δόμησης για τις οποίες ισχύουν τα παρακάτω:

1. $\mu = E(X_{jt})$
2. $s^2 = E[\sigma^2(\Theta_j)]$
3. $\sigma_\alpha^2 = V[\mu(\Theta_j)]$

Απόδειξη: [Goovaerts et al. (1990)].

3.4 Σχέσεις Συνδιασπορών και Μέσες Τιμές

Θεώρημα 5. Ισχύουν:

- $E(X_{jt}) = E(X_{jw}) = E(X_{ww}) = E(X_{wz}) = E[\mu(\Theta_j)] = \mu$
- $Cov(X_{jt}, X_{jr}) = \delta_{rt} \frac{1}{w_{jt}} s^2 + \sigma_\alpha^2$
- $Cov(X_{jt}, X_{jw}) = Cov(X_{jw}, X_{jw}) = \frac{s^2}{w_j} + \sigma_\alpha^2$
- $Cov(X_{jw}, X_{zw}) = Cov(X_{zw}, X_{zw}) = \frac{\sigma_\alpha^2}{Z}$
- $Cov(X_{jw}, X_{ww}) = \frac{s^2}{w_{..}} + \sigma_\alpha^2 \left(\frac{w_j}{w_{..}}\right)$
- $Cov(X_{ww}, X_{ww}) = \frac{s^2}{w_{..}} + \sigma_\alpha^2 \sum_{j=1}^J \left(\frac{w_j}{w_{..}}\right)^2$
- $Cov(X_{j't}, \mu(\Theta_j)) = \sigma_\alpha^2 \delta_{jj'}$
- $Var(X_{ww}) = \frac{s^2}{w_{..}} + \sigma_\alpha^2 \sum_{j=1}^J \left(\frac{w_j}{w_{..}}\right)^2$.

[[βλέπε Pitselis (2018)].

3.5 Αμερόληπτοι Εκτιμητές Παραμέτρων

Οι αμερόληπτοι εκτιμητές των παραμέτρων μ , s^2 και σ_α^2 είναι αντίστοιχα οι κάτωθι:

$$\hat{\mu} = X_{ww}$$

$$\hat{s}^2 = \frac{1}{J} \sum_{j=1}^J \hat{s}_j^2$$

,
όπου

$$\hat{s}_j^2 = \frac{1}{T-1} \sum_{t=1}^T w_{jt} (X_{jt} - X_{jw})^2.$$

Δηλαδή,

$$\hat{s}^2 = \frac{1}{J(T-1)} \sum_{j=1}^J \sum_{t=1}^T w_{jt} (X_{jt} - X_{jw})^2.$$

Τέλος,

$$\hat{\sigma}_\alpha^2 = \left[\frac{w_{..}}{w_{..}^2} - \sum_{j=1}^J w_j^2 \right] \sum_{j=1}^J (X_{wj} - X_{ww})^2 - (J-1)\hat{s}^2.$$

Παρατήρηση: Σημειώνεται πως οι παραπάνω εκτιμητές ισχύουν στην περίπτωση όπου ο αριθμός των περιόδων είναι ίσος με T σε όλα τα ασφαλιστήρια συμβόλαια (balanced model). Σε περίπτωση που ο αριθμός είναι διαφορετικός (unbalanced model) σε ένα τουλάχιστον ασφαλιστήριο συμβόλαιο, η \hat{s}^2 γίνεται ως εξής:

$$\hat{s}^2 = \sum_{j=1}^J f_j \hat{s}_j^2 = \frac{1}{\sum_{j=1}^J (T_j - 1)} \sum_{j=1}^J \sum_{t=1}^{T_j} w_{jt} (X_{jt} - X_{jw})^2,$$

όπου $f_j = \frac{T_j - 1}{\sum_{j=1}^J (T_j - 1)}$, T_j ο αριθμός παρατηρήσεων για το ασφαλιστήριο συμβόλαιο j και

$$\hat{s}_j^2 = \frac{1}{t_j - 1} \sum_{t=1}^{T_j} w_{jt} (X_{jt} - X_{jw})^2.$$

[βλέπε Dannenburg et al. (1996)].

3.6 Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς

Όπως ακριβώς αναλύθηκε στην ενότητα 2.6, έτσι και στο μοντέλο *Buhlmann – Straub* οι παρατηρήσεις X_{jt} μπορούν να γραφούν ως εξής:

$$X_{jt} = \mu + \alpha_j + \varepsilon_{jt}, \quad j = 1, \dots, J, \quad t = 1, \dots, T_j$$

όπου οι μη παρατηρήσιμες μεταβλητές α_j , $j = 1, \dots, J$ είναι ανεξάρτητες και γνωστά κατανομημένες, με μέση τιμή μηδέν και διακύμανση σ_α^2 . Επιπλέον τα ε_{jt} θεωρούνται ανεξάρτητα με μέσο μηδέν. Η διακύμανση των συνιστωσών ε_{jt} είναι $\frac{s^2}{w_{jt}}$, όπου το w_{jt} είναι το βάρος

που συνδέεται με την παρατήρηση X_{jt} . Παρατηρήσεις με διακυμάνσεις όπως αυτή, ανακύπτουν όταν το X_{jt} είναι ένας μέσος όρος που προκύπτει ως $X_{jt} = \sum_k X_{jtk}/w_{jt}$, όπου $X_{jtk} = \mu + \alpha_j + \varepsilon_{jtk}$, με ε_{jtk} ανεξάρτητο, γνωστά κατανομημένο, μέση τιμή μηδέν και διασπορά s^2 . Οι τυχαίες μεταβλητές ε_{jtk} τότε, δηλώνουν αποκλίσεις από το ασφάλιστρο κινδύνου $\mu + \alpha_j$ για την k ατομική σύμβαση σε χρονική περίοδο t και ομάδα j [βλέπε Kaas et al. (2008)].

Όπως και στην παράγραφο (2.5), έτσι και σε αυτό το κεφάλαιο, οι εκτιμητές των παραμέτρων σ_α^2 και s^2 μπορούν να υπολογιστούν από αθροίσματα τετραγώνων SSB και SSW . Η διαδικασία είναι ακριβώς η ίδια με μόνη διαφορά πως εδώ είναι σταθμισμένα (weighted) [βλέπε Kaas et al. (2008)].

Θεώρημα 6. Οι αμερόληπτοι εκτιμητές των σ_α^2 και s^2 για το ισορροπημένο μοντέλο είναι:

$$\hat{s}^2 = MSW$$

$$\hat{\sigma}_\alpha^2 = \frac{(J-1)(MSB - MSW)}{(w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}})},$$

[βλέπε Kaas et al. (2008)].

Απόδειξη: Η περίπτωση του ισορροπημένου μοντέλου περιγράφεται από τον πίνακα (2.2) με μόνη διαφορά την ύπαρξη των w_{jt} (βάρη). Δηλαδή

$$SSW = \sum_{j=1}^J \sum_{t=1}^T w_{jt} (X_{jt} - X_{jw})^2, \quad (3.6)$$

και

$$SSB = \sum_{j=1}^J w_{j.} (X_{jw} - X_{ww})^2. \quad (3.7)$$

Όπως και στην ενότητα 2.6, υπολογίζοντας το $E(MSW)$ και το $E(MSB)$, σύμφωνα με τις σχέσεις (3.32), (3.33), καθώς και το θεώρημα (5), έχουμε:

$$\begin{aligned} E(MSW) &= \frac{E(SSW)}{J(T-1)} \\ &= \frac{E(\sum_{j=1}^J \sum_{t=1}^T w_{jt} (X_{jt} - X_{jw})^2)}{J(T-1)} \\ &= \frac{\sum_{j=1}^J \sum_{t=1}^T w_{jt} [Var(X_{jt}) + Var(X_{jw}) - 2Cov(X_{jt}, X_{jw})]}{J(T-1)} \end{aligned}$$

$$\begin{aligned}
&= \frac{\sum_{j=1}^J \sum_{t=1}^T w_{jt} [\sigma_\alpha^2 + \frac{s^2}{w_{jt}} + \sigma_\alpha^2 + \frac{s^2}{w_j} - 2(\sigma_\alpha^2 + \frac{s^2}{w_j})]}{J(T-1)} \\
&= \frac{J(T-1)s^2}{J(T-1)} \\
&= s^2.
\end{aligned}$$

Επομένως,

$$\hat{s}^2 = MSW.$$

Αντίστοιχα υπολογίζουμε

$$\begin{aligned}
E(MSB) &= \frac{E(SSB)}{J-1} \\
&= \frac{E(\sum_{j=1}^J w_j (X_{jw} - X_{ww})^2)}{J-1} & (3.18) \\
&= \frac{\sum_{j=1}^J w_j [Var(X_{jw}) + Var(X_{ww}) - 2Cov(X_{jw}, X_{ww})]}{J-1} \\
&= \frac{\sum_{j=1}^J w_j [\sigma_\alpha^2 + \frac{s^2}{w_j} + \sigma_\alpha^2 \sum_{j'} \frac{w_{j'}^2}{w_{..}^2} + \frac{s^2}{w_{..}} - 2(\frac{s^2}{w_{..}} + \frac{\sigma_\alpha^2 w_{j.}}{w_{..}})]}{J-1} \\
&= \frac{\sigma_\alpha^2 \sum_{j=1}^J w_j (1 + \sum_{j'} \frac{w_{j'}^2}{w_{..}^2} - 2\frac{w_{j.}}{w_{..}}) + s^2 \sum_{j=1}^J w_j (\frac{1}{w_j} - \frac{1}{w_{..}})}{J-1} \\
&= \frac{\sigma_\alpha^2 (w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}}) + (J-1)s^2}{J-1}.
\end{aligned}$$

Από τη μέση τιμή της σχέσης (3.18) θα έχουμε ότι

$$\frac{\sum_{j=1}^J w_j (X_{jw} - X_{ww})^2}{J-1} = \frac{\hat{\sigma}_\alpha^2 (w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}}) + (J-1)\hat{s}^2}{J-1}$$

$$\sum_{j=1}^J w_j (X_{jw} - X_{ww})^2 = \hat{\sigma}_\alpha^2 (w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}}) + (J-1)\hat{s}^2$$

$$\sum_{j=1}^J w_j (X_{jw} - X_{ww})^2 - (J-1)\hat{s}^2 = \hat{\sigma}_\alpha^2 (w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}}).$$

Λύνοντας ως προς σ_a^2 έχουμε

$$\begin{aligned} \hat{\sigma}_a^2 &= \frac{\sum_{j=1}^J w_j (X_{jw} - X_{ww})^2 - (J-1)\hat{s}^2}{(w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}})} \\ &= \frac{[SSB - (J-1)MSW]}{(w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}})} \\ &= \frac{(J-1)(MSB - MSW)}{(w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}})}. \end{aligned}$$

Θεώρημα 7. Οι αμερόληπτοι εκτιμητές των σ_α^2 και s^2 για το μη ισορροπημένο μοντέλο είναι:

$$\hat{s}^2 = MSW$$

$$\hat{\sigma}_\alpha^2 = \frac{(J-1)(MSB - MSW)}{(w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}})}.$$

Απόδειξη: Στο μη ισορροπημένο μοντέλο η ανάλυση διασποράς δίνεται από:

Πηγή Διασποράς	Βαθμοί Ελευθ.	Αθρ. Τετρ.	Μέσος Αθρ. Τετρ.
Μεταξύ των Ομάδων	$J - 1$	SSB	MSB
Εντός των Ομάδων	$T. - J$	SSW	MSW

Πίνακας 3.2: Ανάλυση διασποράς μη ισορροπημένων δεδομένων για Buhlmann-Straub.

όπου $T. = \sum_{j=1}^J T_j$ και $T. - J = \sum_{j=1}^J (T_j - 1)$.

Ισχύει

$$SSW = \sum_{j=1}^J \sum_{t=1}^{T_j} w_{jt} (X_{jt} - X_{jw})^2, \quad (3.32)$$

και

$$SSB = \sum_{j=1}^J w_j (X_{jw} - X_{ww})^2. \quad (3.33)$$

Υπολογίζοντας τις μέσες τιμές έχουμε

$$\begin{aligned} E(MSW) &= \frac{E(SSW)}{T - J} \\ &= \frac{E(\sum_{j=1}^J \sum_{t=1}^{T_j} w_{jt} (X_{jt} - X_{jw})^2)}{T - J} \\ &= \frac{\sum_{j=1}^J \sum_{t=1}^{T_j} w_{jt} [Var(X_{jt}) + Var(X_{jw}) - 2Cov(X_{jt}, X_{jw})]}{T - J} \\ &= \frac{\sum_{j=1}^J \sum_{t=1}^{T_j} w_{jt} [\sigma_\alpha^2 + \frac{s^2}{w_{jt}} + \sigma_\alpha^2 + \frac{s^2}{w_j} - 2(\sigma_\alpha^2 + \frac{s^2}{w_j})]}{T - J} \\ &= \frac{(T - J)s^2}{(T - J)} \\ &= s^2. \end{aligned}$$

Με ίδιο τρόπο εργαζόμαστε για την εκτίμηση του $\hat{\sigma}_\alpha^2$, η οποία ωστόσο θα οδηγήσει και για το μη ισορροπημένο μοντέλο στον ίδιο εκτιμητή με του θεωρήματος για το ισορροπημένο μοντέλο, καθώς το SSB είναι ανεξάρτητο του t .

Επιπλέον, σύμφωνα με την ενότητα 3.3 και όμοια με την ενότητα 2.6 παίρνουμε το Z_j [βλέπε Bühlmann and Gisler (2006)].

Παρατήρηση: Στην περίπτωση που όλα τα βάρη (w_{jt}) είναι ίσα με ένα, οδηγούμαστε στο μοντέλο του Bühlmann που αναλύθηκε στο προηγούμενο κεφάλαιο.

3.7 Αριθμητικό Παράδειγμα στο Μοντέλο των Buhlmann-Straub

Στην παράγραφο αυτή, θα παρουσιαστούν τα αποτελέσματα αυτού του κεφαλαίου με ένα αριθμητικό παράδειγμα. Το παράδειγμα πραγματεύεται συλλογικά δεδομένα ασφάλισης αυτοκινήτων [βλέπε Dannenburg et al. (1996)], και όπως και στην ενότητα 2.7, η προσέγγιση της λύσης του προβλήματος γίνεται και με τις δύο μεθόδους.

Θεωρούμε ένα χαρτοφυλάκιο αποτελούμενο από εννέα στόλους αυτοκινήτων ($j = 1, \dots, 9$) που παρατηρήθηκε για δέκα έτη ($t = 1, \dots, 10$). Ο κίνδυνος αντιπροσωπεύεται από την τυχαία μεταβλητή X_{jt} , η οποία είναι ένας μέσος που λαμβάνεται από w_{jt} αυτοκίνητα. Υποθέτουμε ότι οι κίνδυνοι X_{jt} με (φυσικά) βάρη w_{jt} πληρούν τις παραδοχές του Buhlmann-Straub.

Έτος/Στόλος	1	2	3	4	5	6	7	8	9
1	540	99	0	275	543	0	333	494	1667
(w)	(44)	(20)	(8)	(22)	(26)	(6)	(18)	(16)	(6)
2	514	103	400	278	984	0	404	133	313
(w)	(50)	(20)	(6)	(22)	(24)	(8)	(20)	(16)	(6)
3	576	163	1042	430	727	0	400	735	556
(w)	(56)	(24)	(10)	(18)	(22)	(6)	(20)	(14)	(4)
4	483	126	313	196	562	645	361	519	769
(w)	(58)	(32)	(6)	(20)	(18)	(6)	(16)	(16)	(2)
5	481	0	0	667	722	833	588	1000	1818
(w)	(58)	(28)	(8)	(12)	(20)	(2)	(18)	(14)	(4)
6	493	219	833	185	610	0	349	641	0
(w)	(56)	(28)	(4)	(10)	(16)	(4)	(18)	(16)	(2)
7	438	370	0	517	794	0	435	339	1429
(w)	(54)	(28)	(6)	(12)	(12)	(2)	(14)	(12)	(4)
8	588	273	0	204	299	769	476	513	0
(w)	(52)	(22)	(4)	(10)	(14)	(2)	(12)	(8)	(2)
9	541	155	0	323	580	0	635	227	0
(w)	(52)	(26)	(4)	(6)	(14)	(2)	(12)	(8)	(4)
10	441	275	0	968	488	0	556	244	0
(w)	(46)	(22)	(4)	(6)	(8)	(2)	(10)	(8)	(2)

Πίνακας 3.3: Παρατηρούμενες αξιώσεις X_{jt} χαρτοφυλακίου εννέα συμβολαίων.

Ο παραπάνω Πίνακας 3.3 περιέχει τις παρατηρηθείσες τιμές X_{jt} , και το πλήθος των οχημάτων (w_{jt}). Παρατηρούμε πως το συνολικό πλήθος των οχημάτων είναι 1510 και πως για τους στόλους 3, 6 και 9 η εμπειρία που έχουμε είναι περιορισμένη εξαιτίας του μικρού αριθμού

οχημάτων.

Για τον υπολογισμό του ποσού που θα πρέπει να καταβληθεί για το κάθε συμβόλαιο j την χρονική περίοδο $T + 1$ έχουμε:

Ανάλυση Διασποράς: Στο αριθμητικό παράδειγμα αυτό για την ανάλυση διασποράς γίνεται χρήση της εντολής «anoona» στον κώδικα [βλέπε κώδικα (A'.2)], που δίνει τα αποτελέσματα που προκύπτουν από τη διαδικασία της ανάλυσης αιασποράς χωρίς να χρειάζεται να υπολογιστούν αναλυτικά οι τύποι (3.32) και (3.33). Τα αποτελέσματα που εξάγονται είναι τα παρακάτω:

$$MSB = 4.693.376$$

$$MSW = 695.107$$

$$F = 6.752$$

με $F_{8,81,0.95} = 2.06$. Επομένως η αρχική υπόθεση απορρίπτεται, δηλαδή το χαροφυλάκιο είναι ανομοιογενές.

Έτσι, όπως αποδείχθηκε στην ενότητα 3.6 έχουμε:

$$\hat{s}^2 = MSW = (833.73)^2,$$

και

$$\hat{\sigma}_a^2 = \frac{(J - 1)(MSB - MSW)}{(w_{..} - \sum_{j=1}^J \frac{w_{j.}^2}{w_{..}})} = 26195.97 = (161.8517)^2.$$

Μοντέλο Buhlmann – Straub: Σύμφωνα με το μοντέλο του Buhlmann-Straub και των εκτιμητών $\hat{\mu}$, \hat{s}^2 , $\hat{\alpha}$, που προαναφέρθηκαν στην ενότητα 3.5 πραγματοποιούνται οι υπολογισμοί όπως παρατίθενται στο Παράρτημα [βλέπε Κώδικα (A'.2)], παίρνοντας τα εξής αποτελέσματα:

$$\hat{\mu} = 439.8344,$$

$$\hat{s}^2 = (833.7308)^2,$$

$$\hat{\sigma}_\alpha^2 = (161.8517)^2.$$

καθώς και

$$\hat{Z}_j = \begin{pmatrix} 0.9519761 \\ 0.9040451 \\ 0.6933620 \\ 0.8387279 \\ 0.8676795 \\ 0.6011884 \\ 0.8562067 \\ 0.8282920 \\ 0.5756787 \end{pmatrix} .$$

Τέλος,

$$\hat{X}_{j,T+1}^{Cred} = \begin{pmatrix} 505.9463 \\ 203.3485 \\ 343.2252 \\ 372.8143 \\ 625.5917 \\ 281.7312 \\ 440.9408 \\ 494.9883 \\ 644.4556 \end{pmatrix} .$$

Παρατήρηση: Από τα αποτελέσματα γίνεται αντιληπτό πως ο υπολογισμός των παραμέτρων δόμησης και με τις δύο μεθόδους οδήγησε σε ίδια αποτελέσματα.

Κεφάλαιο 4

Το Μοντέλο Αξιοπιστίας των Hachemeister

Όπως θα δούμε στα επόμενα κεφάλαια, για να εφαρμόσουμε τα μοντέλα που θα εξετάσουμε, θα πρέπει οι συμβάσεις σε ένα χαρτοφυλάκιο ασφαλιστηρίων συμβολαίων να ομαδοποιηθούν σύμφωνα με ποιοτικούς παράγοντες κινδύνου, που υποδηλώνουν την κάθε κατηγορία. Για παράδειγμα, τα μέρη μιας χώρας στα οποία κατοικούν οι ασφαλισμένοι μπορούν να χαρακτηρίζονται ως βόρειες, κεντρικές και νότιες περιοχές. Ωστόσο, σε πολλές περιπτώσεις, είναι διαθέσιμα ποσοτικά δεδομένα σχετικά με τους κινδύνους, με άμεσο αντίκτυπο στο επίπεδο των ασφαλιστρών κινδύνου, όπως το βάρος ενός αυτοκινήτου στην περίπτωση που αναφερόμαστε σε μια τέτοιου τύπου ασφάλιση. Το ασφάλιστρο κινδύνου στη συνέχεια διαμορφώνεται ως συνάρτηση αυτού του βάρους, για παράδειγμα ανάλογο αυτού. Το βάρος επίσης μπορεί να χρησιμοποιηθεί και με ποιοτικό τρόπο: έτσι τα ασφαλιστικά συμβόλαια ομαδοποιούνται σε κατηγορίες βάρους, χαμηλά, μεσαία και μεγάλα βάρη [βλέπε Dannenburg et al. (1996)].

4.1 Μοντέλο Hachemeister

Στο κεφάλαιο αυτό θα παρουσιάσουμε το μοντέλο αξιοπιστίας παλινδρόμησης που περιέχει ποσοτικούς παράγοντες κινδύνου. Εισήχθη στην αναλογιστική βιβλιογραφία από τον Hachemeister (1975), ο οποίος το ανέπτυξε για να προβλέψει τα μεγέθη των απαιτήσεων από σωματικές βλάβες χρησιμοποιώντας δεδομένα που παρατηρήθηκαν κατά τη διάρκεια των ετών 1970-1973 στις ΗΠΑ.

Στο μοντέλο του Bulhmann-Straub είχαμε την υπόθεση πως $E(X_{jt}|\Theta_j) = \mu(\Theta_j)$ για $j = 1, \dots, J$. Στην πραγματικότητα η υπόθεση ότι δοθέντος του Θ_j , τα X_{jt} είναι ανεξάρτητα και ισόνομα δεν ισχύει. Για να αποφευχθεί αυτός ο περιορισμός ο Hachemeister (1975) επέκτεινε το μοντέλο του Bulhmann Straub, παρουσιάζοντας το μοντέλο αξιοπιστίας μέσω παλινδρόμησης. Για να είμαστε συμβατοί με την βιβλιογραφία που αφορά την παλινδρόμηση

(το Y είναι η εξαρτημένη μεταβλητή και X είναι ο πίνακας σχεδιασμού που περιέχει τις ανεξάρτητες) στη θέση του X του Bullmann Straub θέτουμε το Y (βλέπε Pitselis (2018)).

4.2 Υποθέσεις

Το μοντέλο αυτό, όπως αναφέρθηκε ήδη, αποκλίνει από την υπόθεση του ανεξάρτητα και ισόνομα κατανομημένα επιτρέποντας διαφορετικούς μέσους και διαφορετικές διακυμάνσεις, τα οποία είναι και πάλι συναρτήσσει γνωστών βαρών, όπως και στην περίπτωση του Bullmann Straub. Επιπροσθέτως, επιτρέπει την πιθανή αλληλεπίδραση μεταξύ των παρατηρήσεων και των συμβολαίων [βλέπε Chi H. L. et al. (2007)].

1. Για $j = 1, \dots, K$ τα ασφαλιστήρια συμβόλαια, δηλαδή τα ζεύγη (Θ_j, \mathbf{Y}_j) , είναι ανεξάρτητα και οι μεταβητές $\Theta_1, \dots, \Theta_K$ είναι ανεξάρτητες και ισόνομες.
2. $\mu(\Theta_j) = E(\mathbf{Y}_j|\Theta_j) = \mathbf{X}_j\boldsymbol{\beta}(\Theta_j)$, όπου \mathbf{Y}_j είναι ένα διάνυσμα διάστασης $t \times 1$, \mathbf{X}_j είναι ένας $t \times p$ πίνακας και $\boldsymbol{\beta}(\Theta_j)$ είναι ένα άγνωστο $p \times 1$ διάνυσμα παραμέτρων παλινδρόμησης.
3. $Cov(\mathbf{Y}_j|\Theta_j) = \sigma^2(\Theta_j)\mathbf{W}_j^{-1}$, όπου \mathbf{W}_j είναι ένας $t_j \times t_j$ πίνακας από βάρη.

Γενικά ο πίνακας σχεδιασμού \mathbf{X}_j μπορεί να αποτελείται από διάφορες ανεξάρτητες μεταβλητές που επηρεάζουν το μέγεθος της ζημιάς, για παράδειγμα σε μια ασφάλιση αυτοκινήτου θα μπορούσαμε να έχουμε ανεξάρτητη μεταβλητή τα κυβικά του αυτοκινήτου ή την ηλικία του οδηγού μαζί με άλλες. Στο μοντέλο του Hachemeister, ο πίνακας σχεδιασμού \mathbf{X}_j είναι ο ίδιος για όλα τα ασφαλιστήρια συμβόλαια, με μόνη ανεξάρτητη μεταβλητή την περίοδο ασφάλισης, και για αυτό στη θέση του \mathbf{X}_j θέτουμε το \mathbf{X} . Επιπλέον ο πίνακας \mathbf{W}_j δεν είναι απαραίτητο να είναι διαγώνιος. Το μοντέλο του Bullmann Straub είναι μια ειδική περίπτωση του μοντέλου παλινδρόμησης που εξετάζουμε, όταν ο πίνακας \mathbf{W}_j είναι ένας διαγώνιος πίνακας με στοιχεία της διαγωνίου $\mathbf{W}_j = \text{diag}(w_{1j}, w_{2j}, \dots, w_{Tj})$ και ο πίνακας σχεδιασμού \mathbf{X} είναι ο μοναδιαίος πίνακας $\mathbf{X} = (1, 1, \dots, 1)'$, το οποίο συνεπάγεται ότι $\mu_t(\Theta_j) = \beta_1(\Theta_j)$ για όλα τα t [βλέπε Pitselis (2018)].

4.3 Συμβολισμοί και Παράμετροι Δόμησης

Θεώρημα 8. Ο βέλτιστος γραμμικός εκτιμητής του $E[\boldsymbol{\beta}(\Theta_j)|\mathbf{X}_j]$ δίνεται από την σχέση

$$\mathbf{B}_j^{Cred} = \mathbf{Z}_j \widehat{\boldsymbol{\beta}}_j + (\mathbf{I} - \mathbf{Z}_j) \mathbf{b},$$

όπου ο $\widehat{\boldsymbol{\beta}}_j$ είναι ο εκτιμητής ελαχίστων τετραγώνων του $\boldsymbol{\beta}(\Theta_j)$, όπως παρουσιάστηκε στη σχέση (4.6) και $E[\widehat{\boldsymbol{\beta}}_j] = \mathbf{b}$, με

$$\mathbf{Z}_j = \mathbf{A}[\mathbf{A} + s^2(\mathbf{X}'_j \mathbf{W}_j \mathbf{X}_j)^{-1}]^{-1}. \quad (4.1)$$

Σημειώνεται πως ο \mathbf{B}_j^{Cred} αναφέρεται και ως εκτιμητής αξιοπιστίας του $\boldsymbol{\beta}(\Theta_j)$. Ο συλλογικός εκτιμητής του $E[\boldsymbol{\beta}(\Theta_j)]$ είναι:

$$\widehat{\mathbf{b}} = \left(\sum_{j=1}^K \mathbf{Z}_j \right)^{-1} \sum_{j=1}^K \mathbf{Z}_j \boldsymbol{\beta}_j,$$

με τον συντελεστή αξιοπιστίας \mathbf{Z}_j , όπως ορίζεται στη σχέση (4.1).

Οι παράμετροι \mathbf{A} , s^2 και \mathbf{b} ονομάζονται παράμετροι δόμησης και ισχύει ότι:

$$\begin{aligned} \mathbf{b} &= E[\boldsymbol{\beta}(\Theta_j)], \\ s^2 &= E[\sigma(\Theta_j)], \\ \mathbf{A} &= Cov[\boldsymbol{\beta}(\Theta_j)]. \end{aligned}$$

Ο εκτιμητής ελαχίστων τετραγώνων του $\boldsymbol{\beta}(\Theta_j)$ είναι:

$$\widehat{\boldsymbol{\beta}}_j = (\mathbf{X}'_j \mathbf{W}_j \mathbf{X}_j)^{-1} \mathbf{X}'_j \mathbf{W}_j \mathbf{Y}_j \quad (4.6)$$

[βλέπε Hachemeister (1975)].

4.4 Σχέσεις Συνδιασπορών και Μέσες Τιμές

Θεώρημα 9. Ισχύουν:

$$\begin{aligned}
 E[\mathbf{Y}_j] &= \mathbf{X}_j \mathbf{b} \\
 E[\widehat{\boldsymbol{\beta}}_j | \Theta_j] &= \boldsymbol{\beta}(\Theta_j) \\
 E[\widehat{\boldsymbol{\beta}}_j] &= \mathbf{b} \\
 Cov[\mathbf{Y}_j] &= s^2 \mathbf{W}_j^{-1} + \mathbf{X}_j \mathbf{A} \mathbf{X}_j' \\
 Cov[\boldsymbol{\beta}(\Theta_j), \mathbf{Y}_j] &= \mathbf{A} \mathbf{X}_j' \\
 Cov[\widehat{\boldsymbol{\beta}}_j, \boldsymbol{\beta}(\Theta_j)] &= \mathbf{A} \\
 Cov[\widehat{\boldsymbol{\beta}}_j, \mathbf{Y}_j] &= (\mathbf{A} + s^2 (\mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j)^{-1}) \mathbf{X}_j' \\
 Cov[\widehat{\boldsymbol{\beta}}_j] &= \mathbf{A} + s^2 (\mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j)^{-1}
 \end{aligned}$$

Απόδειξη: [βλέπε Goovaerts et al. (1990)].

4.5 Αμερόληπτοι Εκτιμητές Παραμέτρων

Οι αμερόληπτοι εκτιμητές για το s^2 και \mathbf{A} δίνονται από τις σχέσεις:

$$\begin{aligned}
 \widehat{s}^2 &= \frac{1}{K(n-p)} \sum_{j=1}^K (\mathbf{Y}_j - \mathbf{X}_j \widehat{\boldsymbol{\beta}}_j)' \mathbf{W}_j (\mathbf{Y}_j - \mathbf{X}_j \widehat{\boldsymbol{\beta}}_j) \\
 \widehat{\mathbf{A}} &= \frac{1}{K-1} \sum_{j=1}^K \mathbf{Z}_j (\widehat{\boldsymbol{\beta}}_j - \widehat{\mathbf{b}}) (\widehat{\boldsymbol{\beta}}_j - \widehat{\mathbf{b}})'
 \end{aligned}$$

[βλέπε Goovaerts et al. (1990)].

4.6 Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς

Θεωρούμε το ακόλουθο μοντέλο:

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \quad j = 1, \dots, K. \quad (4.16)$$

Κάθε στοιχείο Y_{jt} στο $t \times 1$ διάνυσμα \mathbf{Y}_j είναι η παρατηρηθείσα τιμή ενός μετρήσιμου χαρακτηριστικού για το συμβόλαιο j στην t περίοδο παρατήρησης. \mathbf{X}_j είναι ένας $t \times p$

πίνακας γνωστών σταθερών. Η διάσταση του διανύσματος των συντελεστών παλινδρόμησης, που την ορίσαμε p , είναι ουσιαστικά ο αριθμός των παραμέτρων παλινδρόμησης που υπάρχουν στο μοντέλο μας, συμπεριλαμβανομένου του σταθερού όρου [βλέπε Pitselis (2018)].

Δηλαδή,

$$\begin{pmatrix} Y_{j1} \\ Y_{j2} \\ \cdot \\ \cdot \\ \cdot \\ Y_{jt} \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2p-1} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & X_{t1} & X_{t2} & \dots & X_{tp-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_t \end{pmatrix}$$

Σε αντίθεση με τα κλασικά μοντέλα παλινδρόμησης, τα β_j είναι τυχαίας φύσης, ανεξάρτητα και ισόνομα κατανομημένα με κοινό μέσο β και πίνακα συνδιασπορών \mathbf{A} για κάθε j , δηλαδή $E(\beta_j) = \mathbf{b}$ και $V(\beta_j) = \mathbf{A}$. Επιπλέον, για το διάνυσμα των σφαλμάτων ε_j θεωρούνται ως ανεξάρτητα κατανομημένα με μέσο μηδέν, $E(\varepsilon_j | \Theta_j) = 0$.

Το μοντέλο που παρουσίασε ο Hachemeister μπορεί να υλοποιηθεί βάσει του μοντέλου (4.16) εάν ρυθμιστεί ως εξής: $p = 2$ και $X_{jt} = (1, t)$, με το μοντέλο που προκύπτει πλέον να είναι:

$$Y_{jt} = \beta_{j1} + \beta_{j2}t + \varepsilon_{jt}.$$

Θεώρημα 10. Ο αμερόληπτος εκτιμητής του s^2 είναι:

$$\hat{s}^2 = MSE$$

[βλέπε Guo (2012)].

Απόδειξη: Ο πίνακας ανάλυσης διασποράς για το μοντέλο που εξετάζουμε είναι ως εξής:

Πηγή Διασποράς	Βαθμοί Ελευθ.	Αθρ. Τετρ.	Μέσος Αθρ. Τετρ.
Παλινδρόμηση	$K * 1$	$SSR = \sum_{j=1}^K (\hat{\mathbf{Y}}_j - \bar{\mathbf{Y}})^2$	$MSR = SSR / K * 1$
Σφάλμα	$K(t - 2)$	$SSE = \sum_{j=1}^K (\mathbf{Y}_j - \hat{\mathbf{Y}}_j)^2$	$MSE = SSE / K(t - 2)$
Συνολικά	$K(t - 1)$	$SST = \sum_{j=1}^K (\mathbf{Y}_j - \bar{\mathbf{Y}})^2$	-

Πίνακας 4.1: Ανάλυση διασποράς στο μοντέλο Hachemeister.

όπου $\hat{\mathbf{Y}}_j = \mathbf{X}_j \hat{\beta}_j$ η εκτίμηση της συνάρτησης παλινδρόμησης.

Η απόκλιση κάθε \mathbf{Y}_j από τον μέσο $\bar{\mathbf{Y}}$ είναι $\mathbf{Y}_j - \bar{\mathbf{Y}}$. Το μέτρο της συνολικής διασποράς επομένως υπολογίζεται από

$$SST = \sum_{j=1}^K (\mathbf{Y}_j - \bar{\mathbf{Y}})^2,$$

για το οποίο αναμένουμε πως εάν όλα τα \mathbf{Y}_j είναι ίσα, τότε $SST = 0$ καθώς και πως όσο μεγαλύτερη είναι η διασπορά τόσο μεγαλύτερη τιμή αναμένεται για το SST . Ακόμα $\varepsilon_i = \mathbf{Y}_i - \hat{\mathbf{Y}}_i$. Το μέτρο της συνολικής διασποράς μπορεί να γραφεί και ως εξής:

$$\mathbf{Y}_j - \bar{\mathbf{Y}} = \hat{\mathbf{Y}}_j - \bar{\mathbf{Y}} + \mathbf{Y}_j - \hat{\mathbf{Y}}_j,$$

από όπου συνεπάγονται τα εξής:

$$\begin{aligned} \sum_{j=1}^K (\mathbf{Y}_j - \bar{\mathbf{Y}})^2 &= \sum_{j=1}^K [(\hat{\mathbf{Y}}_j - \bar{\mathbf{Y}}) + (\mathbf{Y}_j - \hat{\mathbf{Y}}_j)]^2 \\ &= \sum_{j=1}^K [(\hat{\mathbf{Y}}_j - \bar{\mathbf{Y}})^2 + (\mathbf{Y}_j - \hat{\mathbf{Y}}_j)^2 + 2(\hat{\mathbf{Y}}_j - \bar{\mathbf{Y}})(\mathbf{Y}_j - \hat{\mathbf{Y}}_j)]. \end{aligned}$$

Αλλά όπως είδαμε στο (1.4) συνεπάγεται:

$$\sum_{j=1}^K (\mathbf{Y}_j - \bar{\mathbf{Y}})^2 = \sum_{j=1}^K (\hat{\mathbf{Y}}_j - \bar{\mathbf{Y}})^2 + \sum_{j=1}^K (\mathbf{Y}_j - \hat{\mathbf{Y}}_j)^2.$$

Υπολογίζοντας, όπως στα προηγούμενα κεφάλαια, το $E(MSE)$ έχουμε:

$$\begin{aligned} E(MSE) &= \frac{E(SSE)}{t - p - 1} \\ &= \frac{E[\sum_{j=1}^K (\mathbf{Y}_j - \hat{\mathbf{Y}}_j)^2]}{t - 2} \\ &= \frac{\sum_{j=1}^K E(\mathbf{Y}_j - \hat{\mathbf{Y}}_j)^2)}{t - 2} \end{aligned}$$

$$\begin{aligned}
&= \frac{\sum_{j=1}^K \text{Var}(\mathbf{Y}_j - \widehat{\mathbf{Y}}_j) + [E(\mathbf{Y}_j - \widehat{\mathbf{Y}}_j)]^2}{t-2} \\
&= \frac{\sum_{j=1}^K \text{Var}[(\mathbf{Y}_j - \bar{\mathbf{Y}} - \widehat{\beta}_1(\mathbf{X}_j - \bar{\mathbf{X}})]^2}{t-2} \\
&= \frac{\sum_{j=1}^K [\text{Var}(\mathbf{Y}_j - \bar{\mathbf{Y}}) - 2\text{Cov}(\mathbf{Y}_j - \bar{\mathbf{Y}}, \widehat{\beta}_1(\mathbf{X}_j - \bar{\mathbf{X}})) + \text{Var}(\widehat{\beta}_1)(\mathbf{X}_j - \bar{\mathbf{X}})^2]}{t-2} \\
&= \frac{\sum_{j=1}^K [\text{Var}(\mathbf{Y}_j - \bar{\mathbf{Y}}) - 2\text{Cov}((\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{X}_j - \bar{\mathbf{X}}), \widehat{\beta}_1) + \text{Var}(\widehat{\beta}_1)(\mathbf{X}_j - \bar{\mathbf{X}})^2]}{t-2} \\
&= \frac{\sum_{j=1}^K [\text{Var}(\varepsilon_j - \bar{\varepsilon}) - 2\text{Cov}((\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{X}_j - \bar{\mathbf{X}}), \widehat{\beta}_1) + \text{Var}(\widehat{\beta}_1)(\mathbf{X}_j - \bar{\mathbf{X}})^2]}{t-2} \\
&= \frac{(t-1)s^2 - 2\text{Cov}(\sum_{j=1}^K (\mathbf{Y}_j - \bar{\mathbf{Y}})(\mathbf{X}_j - \bar{\mathbf{X}}), \widehat{\beta}_1) + \text{Var}(\widehat{\beta}_1)(\mathbf{X}_j - \bar{\mathbf{X}})^2}{t-2} \\
&= \frac{(t-1)s^2 - 2\text{Cov}(\widehat{\beta}_1 \sum_{j=1}^K (\mathbf{X}_j - \bar{\mathbf{X}})^2, \widehat{\beta}_1) + \text{Var}(\widehat{\beta}_1)(\mathbf{X}_j - \bar{\mathbf{X}})^2}{t-2} \\
&= \frac{(t-1)s^2 - \text{Var}(\widehat{\beta}_1) \sum_{j=1}^K (\mathbf{X}_j - \bar{\mathbf{X}})^2}{t-2} \\
&= \frac{(t-2)s^2}{t-2} \\
&= s^2.
\end{aligned}$$

Έτσι προκύπτει πως:

$$\widehat{s}^2 = MSE.$$

Ο υπολογισμός του $E(MSR)$ παραλείπεται καθώς δεν οδηγεί στον υπολογισμό κάποιας παραμέτρου δόμησης.

Οι συντελεστές αξιοπιστίας μπορούν να ληφθούν με τη χρήση του παρακάτω τύπου [βλέπε Frees et al. (1999)]:

$$Z_j = \frac{\det(\mathbf{D}\mathbf{W}_j)\mathbf{I}_2 + s^2\mathbf{D}\mathbf{W}_j}{\det(\mathbf{D}\mathbf{W}_j) + s^2\text{trace}(\mathbf{D}\mathbf{W}_j) + s^4},$$

όπου

$$\mathbf{W}_j = \begin{pmatrix} \sum_{t=1}^T w_{jt} & \sum_{t=1}^T tw_{jt} \\ \sum_{t=1}^T tw_{jt} & \sum_{t=1}^T t^2w_{jt} \end{pmatrix},$$

και

$$\mathbf{D} = \text{Var}(\alpha_j) = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}.$$

4.7 Αριθμητικό Παράδειγμα στο Μοντέλο του Hachemeister

Το σύνολο των δεδομένων του Hachemeister (1975) [βλέπε Hachemeister (1975)] αποτελείται από το μέσο όρο των απαιτήσεων που προέκυψαν από ασφαλιστήρια συμβόλαια για σωματικές βλάβες ιδιωτών επιβατών, καθώς και τον αντίστοιχο αριθμό απαιτήσεων για πέντε πολιτείες των ΗΠΑ σε δώδεκα τρίμηνα στο χρονικό διάστημα μεταξύ Ιουλίου 1970 και Ιουνίου 1973. Τα δεδομένα παρουσιάζονται στον παρακάτω πίνακα όπου η πρώτη στήλη περιέχει τους δείκτες για κάθε πολιτεία, οι επόμενες 12 περιέχουν τους μέσους όρους των απαιτήσεων ανά τρίμηνο και οι τελευταίες 12 το πλήθος των αξιώσεων (βάρη) ανά τρίμηνο [βλέπε Kahn (1975)]. Ζητούμενο του αριθμητικού παραδείγματος αυτού, όπως και στα προηγούμενα κεφάλαια, είναι η εκτίμηση του ποσού που θα πληρωθεί σε κάθε συμβόλαιο την περίοδο $T + 1$.

Πολ.	1ο T.	2ο T.	3ο T.	4ο T.	5ο T.	6ο T.	7ο T.	8ο T.	9ο T.	10ο T.	11ο T.
1	1738	1642	1794	2051	2079	2234	2032	2035	2115	2262	2267
2	1364	1408	1597	1444	1342	1675	1470	1448	1464	1831	1612
3	1759	1685	1479	1763	1674	2103	1502	1622	1828	2155	2233
4	1223	1146	1010	1257	1426	1532	1953	1123	1343	1243	1762
5	1456	1499	1609	1741	1482	1572	1606	1735	1607	1573	1613

Πίνακας 4.2: Δεδομένα αριθμητικού παραδείγματος *Hachemeister*

12ο T.	B.1	B.2	B.3	B.4	B.5	B.6	B.7	B.8	B.9	B.10	B.11	B.12
2517	7861	9251	8706	8575	7917	8263	9456	8003	7365	7832	7849	9077
1471	1622	1742	1523	1515	1622	1602	1964	1515	1527	1748	1654	1861
2059	1147	1357	1329	1204	998	1077	1277	1218	896	1003	1108	1121
1306	407	396	348	341	315	328	352	331	287	384	321	342
1690	2902	3172	3046	3068	2693	2910	3275	2697	2663	3017	3242	3425

Μοντέλο Hachemeister: Με χρήση των τύπων του μοντέλου του Hachemeister όπως αυτοί περιγράφησαν στο Θεώρημα (8) και στην ενότητα 4.5 παίρνουμε τους αμερόληπτους εκτιμητές των παραμέτρων. Τα αποτελέσματα από τους υπολογισμούς [βλέπε Κώδικα (Α'3)] είναι:

$$\hat{b}_1 = 2469.57440, \quad \hat{b}_2 = 1621.11925,$$

$$\hat{b}_3 = 2095.99391, \quad \hat{b}_4 = 1538.19530,$$

$$\hat{b}_5 = 1676.26757,$$

$$\hat{s}^2 = 49870187.$$

Τέλος, από θεώρημα (8)

$$\hat{Y}_1^{Cred} = 2436.752,$$

$$\hat{Y}_2^{Cred} = 1650.533,$$

$$\hat{Y}_3^{Cred} = 2073.296,$$

$$\hat{Y}_4^{Cred} = 1507.07,$$

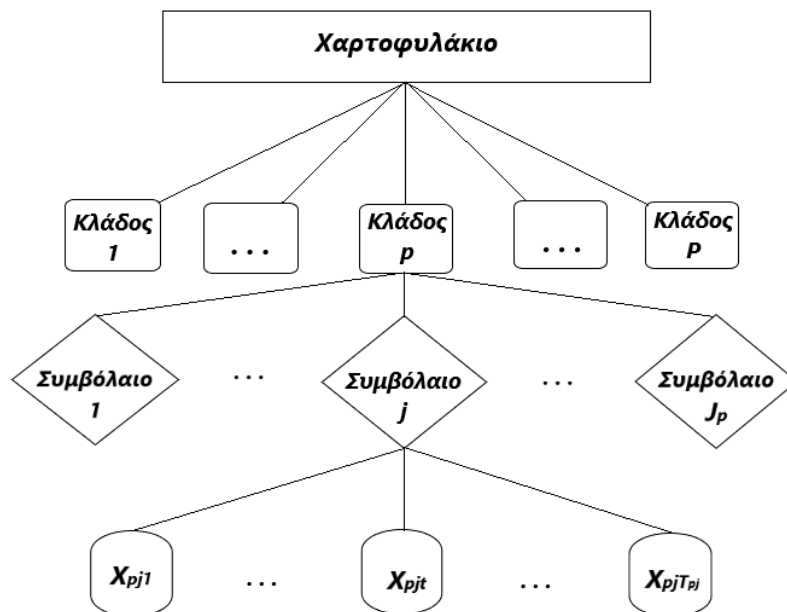
$$\hat{Y}_5^{Cred} = 1759.403.$$

Παρατήρηση: Η μέθοδος της ανάλυσης διασποράς δεν αναπτύχθηκε στο αριθμητικό παράδειγμα, καθώς όπως αναφέρθηκε στην παράγραφο (4.6) δεν είναι δυνατός ο υπολογισμός όλων των παραμέτρων δόμησης μέσω αυτής.

Κεφάλαιο 5

Το Μοντέλο Αξιοπιστίας του Jewell

Όπως είδαμε το μοντέλο των Buhlmann-Straub, ένα χαρτοφυλάκιο υποδιαιρείται σε πολλές ομάδες, περισσότερο ή λιγότερο ομογενείς. Τα τμήματα αυτά μπορούν να αποτελούνται από έναν αριθμό όμοιων συμβολαίων, είτε να είναι ένα μεμονωμένο συμβόλαιο. Ωστόσο είναι πιθανό τα συμβόλαια να ομαδοποιηθούν σε διάφορους τομείς με κοινά χαρακτηριστικά, επιπλέον των δικών τους ατομικών χαρακτηριστικών. Για παράδειγμα σε ένα χαρτοφυλάκιο ασφάλισης πυρός μπορεί να γίνει μια πρώτη υποδιαίρεση σε κλάδο ασφάλισης πυρός ιδιωτικής κατοικίας και βιομηχανικών ακινήτων. Η υποδιαίρεση μπορεί να συνεχιστεί και σε ένα δεύτερο επίπεδο σχετικά με το μέγεθος των ακινήτων αυτών.



Σχήμα 5.1: Γραφική αναπαράσταση Ιεραρχικού Μοντέλου 3 επιπέδων.

5.1 Μοντέλο Jewell

Πρόκειται ουσιαστικά για μια επέκταση του μοντέλου Buhlmann-Straub που έγινε από τον Jewell (1975) . Θεωρούμε ένα χαρτοφυλάκιο ασφαλιστηρίων συμβολαίων το οποίο σε πρώτη φάση υποδιαιρείται σε p κλάδους (πρώτος παράγοντας) , με $p = 1, \dots, P$. Κάθε κλάδος χαρακτηρίζεται από μια παράμετρο κινδύνου Θ_p , που όπως έχει προαναφερθεί περιγράφει την ετερογένεια μεταξύ των κλάδων. Στο επόμενο επίπεδο, κάθε κλάδος υποδιαιρείται σε ασφαλιστήρια συμβόλαια j (δεύτερος παράγοντας). Δοθέντος του κλάδου, κάθε ασφαλιστήριο συμβόλαιο j χαρακτηρίζεται από μια άλλη παράμετρο κινδύνου Θ_{pj} με $j = 1, \dots, J$. Έτσι κάθε ασφαλιστήριο συμβόλαιο αποτελείται από τις μεταβλητές X_{pjt}, Θ_{pj} με $t = 1, \dots, t_{pj}$.

5.2 Υποθέσεις

1. Οι κλάδοι είναι μεταξύ τους ανεξάρτητοι.
2. Για κάθε κλάδο p , τα συμβόλαια είναι κατά συνθήκη ανεξάρτητα δοθέντος Θ_p .
3. Για κλαυθε $p = 1, \dots, P$ και κάθε $j = 1, \dots, J_p$ οι αξιώσεις $X_{pj1}, \dots, X_{pj t_{pj}}$ είναι κατά συνθήκη ανεξάρτητες δοθέντος (Θ_p, Θ_{pj}) .
4. Όλα τα ζεύγη των μεταβλητών (Θ_p, Θ_{pj}) για $p = 1, \dots, P$ και $j = 1, \dots, j_p$ είναι ισόνομα κατανομημένα.
5. Για κάθε p, j, t ισχύει:

$$E(X_{pjt}|\Theta_p, \Theta_{pj}) = \mu(\Theta_p, \Theta_{pj}), \quad t = 1, \dots, t_{pj},$$

και

$$Var(X_{pjt}|\Theta_p, \Theta_{pj}) = \frac{1}{w_{pjt}} \sigma^2(\Theta_p, \Theta_{pj}), \quad t = 1, \dots, t_{pj}$$

όπου w_{pjt} είναι γνωστά βάρη.

[βλέπε Goovaerts et al. (1990)].

5.3 Παράμετροι Δόμησης και Συμβολισμοί

Για τις παραμέτρους δόμησης ισχύουν τα εξής:

1. $\mu = E[\mu(\Theta_p, \Theta_{pj})] = E(X_{pjt}|\Theta_p)$

2. $s^2 = E[\sigma^2(\Theta_p, \Theta_{pj})]$
3. $\sigma_\alpha^2 = E[Var[\mu(\Theta_p, \Theta_{pj})|\Theta_p]]$
4. $\sigma_\beta^2 = Var[\mu(\Theta_p)]$

Επιπλέον συμβολισμοί είναι,

$$\begin{aligned}
 w_{p..} &= \sum_{j=1}^{J_p} w_{pj.} = \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} w_{pjt}, \\
 X_{pjw} &= \sum_{t=1}^{t_{pj}} \frac{w_{pjt}}{w_{pj.}} X_{pjt}, \\
 X_{pzw} &= \sum_{j=1}^{J_p} \frac{Z_{pj}}{w_{p.}} X_{pjw}, \\
 X_{zzw} &= \sum_{p=1}^P \frac{Z_p}{Z_{..}} X_{pzw}.
 \end{aligned}$$

Θεώρημα 11. *Ο συντελεστής αξιοπιστίας στο επίπεδο των συμβολαίων είναι:*

$$Z_{pj} = \frac{\sigma_\alpha^2 w_{pj.}}{s^2 + \sigma_\alpha^2 w_{pj.}},$$

και

$$Z_p = \sum_{j=1}^{J_p} Z_{pj}.$$

Ο συντελεστής αξιοπιστίας στο επίπεδο των κλάδων είναι:

$$Z_p = \frac{\sigma_\beta^2 Z_p}{\sigma_\alpha^2 + \sigma_\beta^2 Z_p},$$

και

$$Z_{..} = \sum_{p=1}^P Z_p$$

Προς αποφυγή σύγχυσης $Z_p \neq Z_{p.} = \sum_{j=1}^{J_p} Z_{pj}$

Με βάση τις υποθέσεις μπορούμε να λάβουμε την κάτωθι γραμμικοποιημένη μη ομογενή εκτιμήτρια για το καθαρό ασφάλιστρο του κλάδου p

$$\mu(\Theta_p)^{Cred} = Z_p X_{pzw} + (1 - Z_p)\mu,$$

καθώς και την μη ομογενή εκτιμήτρια για το καθαρό ασφάλιστρο των ασφαλιστηρίων συμβολαίων που δίνεται από:

$$\mu(\Theta_p, \Theta_{pj})^{Cred} = Z_{pj} X_{pjw} + (1 - Z_{pj}) \mu(\Theta_p)^{Cred}.$$

[βλέπε *Goovaerts et al. (1990)*].

5.4 Σχέσεις Συνδιασπορών

Θεώρημα 12. *Ισχύουν:*

- $Cov(X_{pjt}, X_{pj't'}) = \delta_{tt'} \frac{s^2}{w_{pjw}} + \sigma_\alpha^2 + \sigma_\beta^2$
- $Cov(X_{pj't}, X_{pj't'}) = \delta_{jj'} (\delta_{tt'} \frac{s^2}{w_{pjw}} + \sigma_\alpha^2 + \sigma_\beta^2)$
- $Cov(X_{pj't}, X_{p'j't'}) = 0, p \neq p'$
- $Cov(X_{pj't}, X_{pj't'}) = Cov(X_{pjw}, X_{pj'w}) = \delta_{jj'} \frac{\sigma_\alpha^2}{Z_{pj}} + \sigma_\beta^2$
- $Cov(X_{pjw}, X_{pzw}) = Cov(X_{pzw}, X_{pzw}) = \frac{\sigma_\alpha^2}{Z_p} + \sigma_\beta^2 = \frac{\sigma_\beta^2}{Z_p}$
- $Cov(X_{pzw}, X_{zzw}) = Cov(X_{zzw}, X_{zzw}) = \frac{\sigma_\beta^2}{Z}$
- $Cov[\mu(\Theta_p), X_{p'jt}] = Cov[\mu(\Theta_p), X_{p'jw}] = \delta_{pp'} \sigma_\beta^2$
- $Cov[\mu(\Theta_{pj}, \Theta_p), X_{p'jt}] = \delta_{pp'} (\delta_{jj'} \sigma_\alpha^2 + \sigma_\beta^2)$

[βλέπε *Goovaerts et al. (1990)*].

5.5 Αμερόληπτοι Εκτιμητές Παραμέτρων

Οι κάτωθι είναι αμερόληπτοι εκτιμητές των παραμέτρων μ_p , μ , s^2 , σ_α^2 και σ_β^2 αντίστοιχα.

$$\begin{aligned}
\hat{\mu}_p &= X_{pzw}, \\
\hat{\mu} &= X_{zzw}, \\
\hat{s}^2 &= \frac{\sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} w_{pjt} (X_{pjt} - X_{pjw})^2}{\sum_{p=1}^P \sum_{j=1}^{J_p} (t_{pj} - 1)_-}, \\
\hat{\sigma}_\alpha^2 &= \frac{\sum_{p=1}^P \sum_{j=1}^{J_p} Z_{pj} (X_{pjw} - X_{pzw})^2}{\sum_{p=1}^P (J_p - 1)_+}, \\
\hat{\sigma}_\beta^2 &= \frac{\sum_{p=1}^P Z_p (X_{pzw} - X_{zzw})^2}{P - 1}.
\end{aligned}$$

[βλέπε Goovaerts et al. (1990)].

5.6 Προσέγγιση με τη Μέθοδο της Ανάλυσης Διασποράς

Αν μοντελοποιήσουμε όπως και στα προηγούμενα κεφάλαια, ο παρατηρούμενος κίνδυνος του τομέα p για το συμβόλαιο j στη μονάδα έκθεσης t μπορεί να αποσυντεθεί ως εξής:

$$X_{pjt} = \mu + \alpha_p + \beta_{pj} + \varepsilon_{pjt},$$

όπου μ είναι οι συνολικές αναμενόμενες αξιώσεις, α_p είναι η απόκλιση από τις κατά συνθήκη αναμενόμενες αξιώσεις του κλάδου p από το μ , β_{pj} είναι η απόκλιση των κατα συνθήκη αναμενόμενων αξιώσεων του συμβολαίου j στον κλάδο p από το προσδοκώμενο του κλάδου, δηλαδή από το $\mu + \alpha_p$ και ε_{pjt} είναι η απόκλιση της παρατήρησης X_{pjt} από το $\mu + \alpha_p + \beta_{pj}$. Οι αναμενόμενες απαιτήσεις από τον κλάδο p είναι $E(X_{pjt}|\alpha_p) = \mu + \alpha_p$ και οι αναμενόμενες απαιτήσεις του συμβολαίου j εντός του κλάδου p είναι $E(X_{pjt}|\alpha_p, \beta_{pj}) = \mu + \alpha_p + \beta_{pj}$. Θεωρούμε ότι οι μεταβλητές α_p , β_{pj} και ε_{pjt} είναι πλήρως ασυσχέτιστες με μέσες τιμές μηδέν και διακυμάνσεις σ_α^2 , σ_β^2 , σ_ε^2 αντίστοιχα. Οι παράμετροι αυτοί, σ_α^2 , σ_β^2 , σ_ε^2 , είναι γνωστές ως συστατικά διακύμανσης. Επιπλέον σε κάθε συμβόλαιο j του κλάδου p στον χρόνο t δίνεται ένα βάρος w_{pjt} , για παράδειγμα ίσο με τον αριθμό των συμβολαίων στον κλάδο τον δεδομένο χρόνο t . Ακόμα είναι κατανοητό πως δεν είναι διαθέσιμα όλα τα δεδομένα X_{pjt} , επειδή η τάξη pj είναι κενή σε κάποιες περιόδους. Επομένως κάποιες πληροφορίες ίσως δεν είναι διαθέσιμες. Έτσι το πλήθος των παρατηρήσεων για το κάθε ασφαλιστήριο συμβόλαιο j του κλάδου p το συμβολίζουμε t_{pj} [βλέπε Frees et al. (1999)].

Για την ανάλυση διασποράς των δεδομένων που περιγράφονται από το ιεραρχικό μοντέλο

υπολογίζουμε τα αθροίσματα τετραγώνων των αντίστοιχων όρων, ώστε να γίνει ο έλεγχος μεταξύ των μέσων των κλάδων p με τον γενικό μέσο, καθώς και των μέσων των συμβολαίων του κάθε κλάδου p με τον μέσο όλων των κλάδων. Τέλος για την ανάλυση της διασποράς υπολογίζεται το άθροισμα τετραγώνων της διαφοράς των μέσων των συμβολαίων του κάθε κλάδου p , \bar{X}_{pj} , από την κάθε τιμή X_{pjt} για τα αντίστοιχα p και j . Έτσι για το ισορροπημένο μοντέλο έχουμε:

Πηγή Διασποράς	Βαθμοί Ελευθ.	Αθρ. Τετρ.	Μέσος Τετρ.
Παράγοντας Α	$P - 1$	SSP	MSP
Παρ. Β εντός Α	$P(J - 1)$	SSJ	MSJ
Σφάλμα	$PJ(T - 1)$	SSE	MSE

Πίνακας 5.1: Ανάλυση διασποράς ισορροπημένων δεδομένων.

με τα αθροίσματα τετραγώνων να είναι:

$$SSP = JT \sum_{p=1}^P (\bar{X}_{p..} - \bar{X}_{...})^2,$$

$$SSJ = T \sum_{p=1}^P \sum_{j=1}^{J_p} (\bar{X}_{pj.} - \bar{X}_{p..})^2,$$

$$SSE = \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^T (X_{pjt} - \bar{X}_{pj.})^2.$$

Θεώρημα 13. Οι υπολογισμοί των $E(MSE)$, $E(MSP)$ και $E(MSJ)$ καταλήγουν στους προκύπτοντες αμερόληπτους εκτιμητές των συνιστωσών διακύμανσης s^2 , σ_α^2 και σ_β^2 [βλέπε *Sahai and Ojeda (2013)*]. Για το ισορροπημένο μοντέλο αυτοί είναι:

$$\begin{aligned} \hat{s}^2 &= MSE, \\ \hat{\sigma}_\beta^2 &= \frac{1}{T}(MSJ - MSE), \\ \hat{\sigma}_\alpha^2 &= \frac{1}{TJ}(MSP - MSJ). \end{aligned}$$

Απόδειξη: Η απόδειξη του θεωρήματος είναι όμοια με εκείνη του Θεωρήματος (3) και παραλείπεται.

Θεώρημα 14. Στο μη ισορροπημένο μοντέλο, οι αμερόληπτοι εκτιμητές των παραμέτρων s^2 , σ_α^2 και σ_β^2 είναι:

$$\begin{aligned}\widehat{s}^2 &= MSE, \\ \widehat{\sigma}_\alpha^2 &= \frac{1}{r_3} \left[MSP - r_2 \frac{\frac{SSJ}{(J. - P)} - \left(\frac{SSE}{T - J.} \right)}{r_1} - MSE \right], \\ \widehat{\sigma}_\beta^2 &= \frac{\frac{SSJ}{(J. - P)} - \left(\frac{SSE}{T - J.} \right)}{r_1}.\end{aligned}$$

[βλέπε Sahai and Ojeda (2007)]

Απόδειξη: Η ανάλυση διασποράς για το μη ισορροπημένο μοντέλο δίνεται από τον παρακάτω πίνακα:

Πηγή Διασποράς	Βαθμοί Ελευθ.	Αθρ. Τετρ.	Μέσος Τετρ.
Παράγοντας Α	$P - 1$	SSP	MSP
Παρ. Β εντός Α	$J. - P$	SSJ	MSJ
Σφάλμα	$T. - J.$	SSE	MSE

Πίνακας 5.2: Ανάλυση διασποράς μη ισορροπημένων δεδομένων.

όπου $J. = \sum_{p=1}^P J_p$ με J_p να είναι το πλήθος συμβολαίων του ανά κλάδο.

Έτσι τα αθροίσματα των τετραγώνων ορίζονται ως εξής [βλέπε Sahai and Ojeda (2007)]:

$$SSP = \sum_{p=1}^P t_{p.} (\bar{X}_{p.} - \bar{X}_{...})^2 = \sum_{p=1}^P \frac{X_{p.}^2}{t_{p.}} - \frac{X_{...}^2}{T}, \quad (5.20)$$

$$SSJ = \sum_{p=1}^P \sum_{j=1}^{J_p} t_{pj} (\bar{X}_{pj.} - \bar{X}_{p.})^2 = \sum_{p=1}^P \sum_{j=1}^{J_p} \frac{X_{pj.}^2}{t_{pj}} - \sum_{p=1}^P \frac{X_{p.}^2}{t_{p.}}, \quad (5.21)$$

$$SSE = \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} (X_{pjt} - \bar{X}_{pj.})^2 = \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} (X_{pjt}^2) - \sum_{p=1}^P \sum_{j=1}^{J_p} \frac{X_{pj.}^2}{t_{pj}}, \quad (5.22)$$

όπου

$$X_{pj.} = \sum_{t=1}^{t_{pj}} X_{pjt}, \quad \bar{X}_{pj.} = \frac{X_{pj.}}{t_{pj}},$$

$$X_{p..} = \sum_{j=1}^{J_p} X_{pj.}, \quad \bar{X}_{p..} = \frac{X_{p..}}{t_p},$$

$$X_{...} = \sum_{p=1}^P X_{p..}, \quad \bar{X}_{...} = \frac{X_{...}}{T},$$

$$t_p = \sum_{j=1}^{J_p} t_{pj}, \quad T = \sum_{p=1}^P t_p.$$

Προς διευκόλυνση των πράξεων για τη συνέχεια θεωρούμε:

$$D_p = \sum_{p=1}^P \frac{X_{p..}^2}{t_p},$$

$$D_j = \sum_{p=1}^P \sum_{j=1}^{J_p} \frac{X_{pj.}^2}{t_{pj}},$$

$$D_o = \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} (X_{pjt}^2),$$

$$D_\mu = \frac{X_{...}^2}{T}.$$

Έτσι λοιπόν οι σχέσεις (5.20), (5.21) και (5.22) συνοψίζονται ως εξής:

$$SSP = D_p - D_\mu,$$

$$SSJ = D_j - D_p,$$

$$SSE = D_o - D_j.$$

Όπως αναφέρθηκε στην παρουσίαση του μοντέλου ισχύει ότι $E(\alpha_p) = E(\beta_{pj}) = E(\varepsilon_{pjt}) = 0$

και επομένως έχουμε ότι:

$$\begin{aligned} s^2 &= E(\varepsilon_{pjt}^2), \\ \sigma_\alpha^2 &= E(\alpha_p^2), \\ \sigma_\beta^2 &= E(\beta_{pj}^2). \end{aligned}$$

Επιπλέον έχουμε:

$$\begin{aligned} E(D_o) &= \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} E(X_{pjt}^2) \\ &= \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} E(\mu + \alpha_p + \beta_{pj} + \varepsilon_{pjt})^2 \\ &= \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} (\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2 + s^2) \\ &= T(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2 + s^2). \end{aligned}$$

$$\begin{aligned} E(D_\mu) &= E\left(\frac{X^2}{T}\right) \\ &= \frac{E(T\mu + \sum_{p=1}^P t_p \alpha_p + \sum_{p=1}^P \sum_{j=1}^{J_p} t_{pj} \beta_{pj} + \sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} \varepsilon_{pjt})^2}{T} \\ &= \frac{[T^2\mu^2 + \sum_{p=1}^P t_p^2 \sigma_\alpha^2 + \sum_{p=1}^P \sum_{j=1}^{J_p} t_{pj}^2 \sigma_\beta^2 + Ts^2]}{T} \\ &= T\mu^2 + K_1\sigma_\alpha^2 + K_2\sigma_\beta^2 + s^2, \end{aligned}$$

όπου,

$$\begin{aligned} K_1 &= \frac{\sum_{p=1}^P t_p^2}{T}, \\ K_2 &= \frac{\sum_{p=1}^P \sum_{j=1}^{J_p} t_{pj}^2}{T}. \end{aligned}$$

Ακόμα

$$\begin{aligned}
E(D_p) &= \sum_{p=1}^P E\left(\frac{X_{p..}^2}{t_p}\right) \\
&= \sum_{p=1}^P E\left[\frac{t_p(\mu + \alpha_p) + \sum_{j=1}^{J_p} t_{pj}\beta_{pj} + \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} \varepsilon_{pj}t}{t_p}\right]^2 \\
&= \sum_{p=1}^P \left[\frac{t_p^2(\mu^2 + \sigma_\alpha^2) + \sum_{j=1}^{J_p} t_{pj}^2\sigma_\beta^2 + t_p \cdot s^2}{t_p}\right] \\
&= \sum_{p=1}^P [t_p(\mu^2 + \sigma_\alpha^2) + \sum_{j=1}^{J_p} \left(\frac{t_{pj}^2}{t_p}\right)\sigma_\beta^2 + s^2] \\
&= T(\mu^2 + \sigma_\alpha^2) + K_{12}\sigma_\beta^2 + P s^2,
\end{aligned}$$

με

$$K_{12} = \sum_{p=1}^P \sum_{j=1}^{J_p} \frac{t_{pj}^2}{t_p}.$$

Ακόμα,

$$\begin{aligned}
E(D_j) &= \sum_{p=1}^P \sum_{j=1}^{J_p} E\left(\frac{X_{pj.}^2}{t_{pj}}\right) \\
&= \sum_{p=1}^P \sum_{j=1}^{J_p} \frac{E[t_{pj}(\mu + \alpha_p + \beta_{pj}) + \sum_{t=1}^{t_{pj}} \varepsilon_{pj}t]^2}{t_{pj}} \\
&= \sum_{p=1}^P \sum_{j=1}^{J_p} \frac{[t_{pj}^2(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2 + s^2)]}{t_{pj}} \\
&= \sum_{p=1}^P \sum_{j=1}^{J_p} [t_{pj}(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2) + s^2] \\
&= T(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2) + J \cdot s^2,
\end{aligned}$$

με J να συμβολίζει το συνολικό πλήθος των συμβολαίων στο χαρτοφυλάκιο ανεξαρτήτως κλάδου.

Έτσι οι αναμενόμενες τιμές των αθροισμάτων των τετραγώνων και των μέσων τετραγώνων δίνονται ως εξής:

$$\begin{aligned}
 E(SSE) &= E(D_o - D_j) \\
 &= E(D_o) - E(D_j) \\
 &= T(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2 + s^2) - T(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2) - J \cdot s^2 \\
 &= T s^2 - J \cdot s^2 \\
 &= (T - J) s^2.
 \end{aligned}$$

Επομένως προκύπτει

$$\begin{aligned}
 E(MSE) &= \frac{E(SSE)}{T - J} \\
 &= \frac{(T - J) s^2}{T - J},
 \end{aligned}$$

και άρα,

$$\begin{aligned}
 \hat{s}^2 &= MSE = \frac{SSE}{T - J} \\
 &= \frac{\sum_{p=1}^P \sum_{j=1}^{J_p} \sum_{t=1}^{t_{pj}} (X_{pjt} - \bar{X}_{pj.})^2}{T - J}.
 \end{aligned}$$

Με όμοιο τρόπο υπολογίζουμε,

$$\begin{aligned}
 E(SSJ) &= E(D_J - D_P) \\
 &= E(D_J) - E(D_P) \\
 &= T(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2) + J \cdot s^2 - T(\mu^2 + \sigma_\alpha^2) - K_{12} \sigma_\beta^2 - P s^2 \\
 &= T \sigma_\beta^2 + J \cdot s^2 - K_{12} \sigma_\beta^2 - P s^2 \\
 &= (T - K_{12}) \sigma_\beta^2 + (J - P) s^2,
 \end{aligned}$$

άρα

$$\begin{aligned} E(MSJ) &= \frac{E(SSJ)}{J_i - P} \\ &= \frac{(T - K_{12})\sigma_\beta^2 + (J_i - P)s^2}{J_i - P} \\ &= \frac{T - K_{12}}{J_i - P} \sigma_\beta^2 + s^2. \end{aligned}$$

Θέτοντας,

$$r_1 = \frac{T - K_{12}}{J_i - P},$$

καταλήγουμε ότι

$$E(MSJ) = r_1 \sigma_\beta^2 + s^2.$$

Δηλαδή,

$$MSJ = r_1 \hat{\sigma}_\beta^2 + \hat{s}^2,$$

και άρα,

$$\begin{aligned} \hat{\sigma}_\beta^2 &= \frac{MSJ - MSE}{r_1} \\ &= \frac{\left(\frac{SSJ}{J_i - P}\right) - \left(\frac{SSE}{T - J_i}\right)}{r_1}. \end{aligned}$$

Τέλος,

$$\begin{aligned} E(SSP) &= E(D_p - D_\mu) \\ &= E(D_p) - E(D_\mu) \\ &= T(\mu^2 + \sigma_\alpha^2) + K_{12}\sigma_\beta^2 + Ps^2 - T\mu^2 - K_1\sigma_\alpha^2 - K_2\sigma_\beta^2 - s^2 \\ &= (T - K_1)\sigma_\alpha^2 + (K_{12} - K_2)\sigma_\beta^2 + (P - 1)s^2. \end{aligned}$$

Συνεπώς,

$$\begin{aligned} E(MSP) &= \frac{E(SSP)}{P-1} \\ &= s^2 + \frac{T-K_1}{P-1}\sigma_\alpha^2 + \frac{K_{12}-K_2}{P-1}\sigma_\beta^2. \end{aligned}$$

Θέτοντας

$$r_2 = \frac{K_{12}-K_2}{P-1}$$

και

$$r_3 = \frac{T-K_1}{P-1},$$

καταλήγουμε ότι,

$$E(MSP) = s^2 + r_2\sigma_\beta^2 + r_3\sigma_\alpha^2.$$

Από τη μέση τιμή αυτής έχουμε

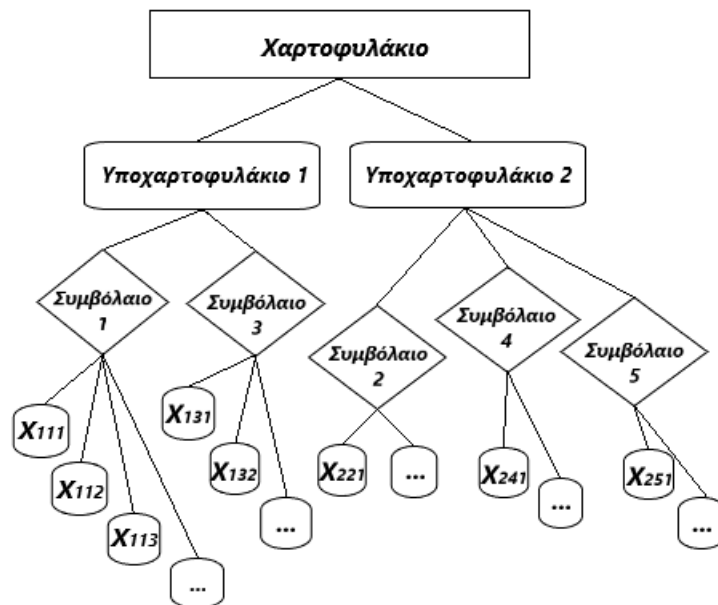
$$MSP = \hat{s}^2 + r_2\hat{\sigma}_\beta^2 + r_3\hat{\sigma}_\alpha^2,$$

από όπου προκύπτει πως

$$\hat{\sigma}_\alpha^2 = \frac{1}{r_3} \left[MSP - r_2 \frac{\left(\frac{SSJ}{J-P}\right) - \left(\frac{SSE}{T-J}\right)}{r_1} - MSE \right].$$

5.7 Αριθμητικό Παράδειγμα στο Μοντέλο του Jewell

Για να παρουσιαστεί και στην πράξη η χρήση του μοντέλου του Jewell, στην παράγραφο αυτή θα εφαρμόσουμε μια προσομοίωση του μοντέλου μας στα δεδομένα του Hachemeister, όπως αυτά δίνονται στην παράγραφο (4.7), με κάποιες διαφοροποιήσεις. Συγκεκριμένα θα θεωρήσουμε ένα τυχαίο χαρακτηριστικό σύμφωνα με το οποίο το χαρτοφυλάκιο θα υποδιαιρεθεί σε δύο κλάδους, 2 υποχαρτοφυλάκια ($p = 1, 2$), όπου το πρώτο θα αποτελείται από τα συμβόλαια 1 και 3 ($j = 1, j = 3$) και το δεύτερο από τα υπόλοιπα ($j = 2, j = 4, j = 5$), καθώς και πως όλα τα δεδομένα είναι ίσης βαρύτητας ($w_{pjt} = 1$).



Σχήμα 5.2: Γραφική αναπαράσταση Ιεραρχικού Μοντέλου αριθμητικού παραδείγματος.

Ανάλυση Διασποράς: Μέσω της ανάλυσης διασποράς για τα δεδομένα μας [βλέπε κώδικα (A.4)] και τη χρήση των σχέσεων του Θεωρήματος (14), λαμβάνονται οι τιμές των αμερόληπτων εκτιμητών των παραμέτρων δόμησης. Αυτές είναι:

$$SSP = 2955372$$

$$SSJ = 699671.1$$

$$SSE = 2532226$$

$$\hat{s}^2 = 46040.47,$$

$$\hat{\sigma}_\alpha^2 = 15598.6,$$

$$\hat{\sigma}_\beta^2 = 94519.04.$$

Μοντέλο Hachemeister: Πραγματοποιώντας τους υπολογισμούς, σύμφωνα με το μοντέλο του *Jewell*, [βλέπε Κώδικα (Α'.4)] προκύπτουν τα εξής αποτελέσματα:

$$\hat{s}^2 = 46040.47,$$

$$\hat{\sigma}_\alpha^2 = 15598.6,$$

$$\hat{\sigma}_\beta^2 = 94519.04.$$

Οι συντελεστές αξιοπιστίας στο επίπεδο των συμβολαίων για τους κλάδους $p=1,2$, όπως αναμενόταν αφού τα $w_{pj} = 1$, είναι μεταξύ τους ίσοι. Συγκεκριμένα:

$$\hat{Z}_{pj} = 0.8025909, \quad p = 1, 2 \quad j = 1, 2, 3, 4, 5$$

Επιπλέον οι συντελεστές αξιοπιστίας στο επίπεδο των κλάδων είναι:

$$\hat{Z}_p = (0.9067732 \quad 0.9358555), \quad p = 1, 2.$$

Παρατήρηση: Από τα αποτελέσματα γίνεται αντιληπτό πως ο υπολογισμός των παραμέτρων δόμησης και με τις δύο μεθόδους οδήγησε σε ίδια αποτελέσματα.

Τέλος, κάνοντας χρήση των εκτιμητών των παραμέτρων δόμησης, καθώς και των συντελεστών αξιοπιστίας, υπολογίζονται οι εκτιμήτριες των καθαρών ασφαλίσεων τόσο των κλάδων

p , όσο και των συμβολαίων j σύμφωνα με τις σχέσεις του θεωρήματος (11)[βλέπε κώδικα (A.4)].

$$\widehat{X}_p^{Cred} = (1921.383 \quad 1504.106), \quad p = 1, 2.$$

και

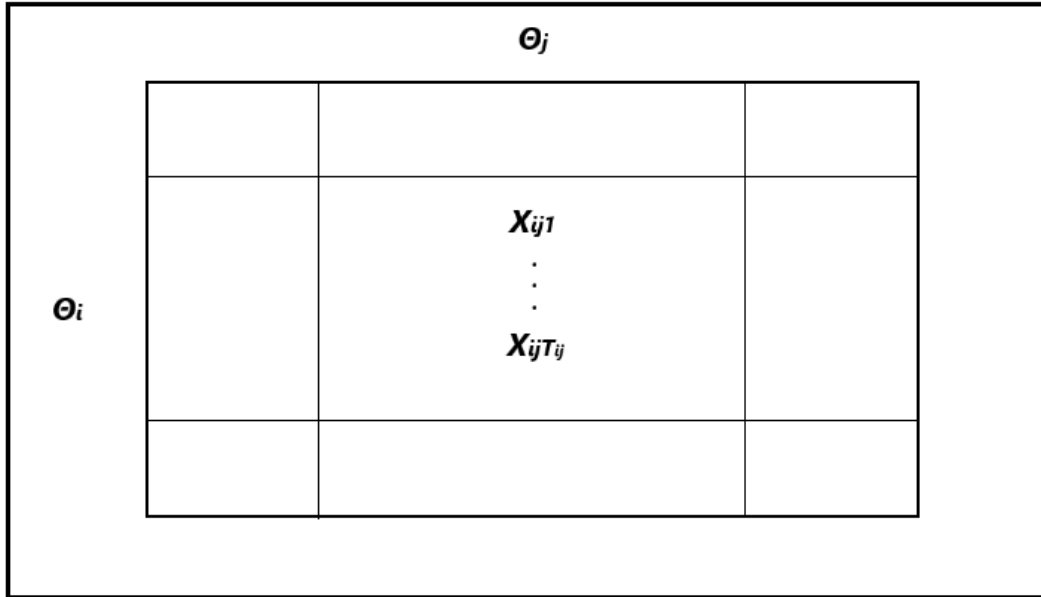
$$\widehat{X}_{1j}^{Cred} = (2035.712 \quad 1841.485), \quad j = 1, 3.$$

$$\widehat{X}_{2j}^{Cred} = (1509.238 \quad 1388.715 \quad 1579.933), \quad j = 2, 4, 5.$$

Κεφάλαιο 6

Το Μοντέλο Αξιοπιστίας Σταυρωτής Ταξινόμησης

Στο προηγούμενο κεφάλαιο κατασκευάστηκαν και εξετάστηκαν συντελεστές αξιοπιστίας για το μοντέλο του Jewell. Το μοντέλο που θα εξετάσουμε στο κεφάλαιο αυτό, μοντέλο παραγόντων σταυρωτής ταξινόμησης, υποθέτει πως τα δεδομένα μπορούν να ταξινομηθούν με ιεραρχικό τρόπο, αλλά αυτό δεν μπορεί να επιτευχθεί πάντα. Πιο συγκεκριμένα, σε ένα πιθανό χαρτοφυλάκιο ασφάλισης αυτοκινήτων με κινδύνους κατηγοριοποιημένους ανάλογα με το φύλο και την ηλικία του ασφαλισμένου δεν μπορεί να υπάρξει αιτιολόγηση ιεράρχησης των κινδύνων αυτών. Δεν υπάρχει λόγος για τον οποίο το φύλο θα πρέπει στο μοντέλο να εισάγεται σε υψηλότερο επίπεδο στην ιεράρχηση από την ηλικία, ή αντίστροφα. Έτσι, ένα μοντέλο σαν του Jewell δεν είναι κατάλληλο σε μια τέτοια περίπτωση χαρτοφυλακίου. Τέτοιες καταστάσεις, με μη ιεραρχικούς παράγοντες κινδύνων, παρατηρούνται συχνά στην πράξη. Για να χειριστούμε δομές δεδομένων σαν αυτές, χρησιμοποιούμε το μοντέλο αξιοπιστίας σταυρωτής ταξινόμησης n -παραγόντων, όπου n το πλήθος των παραγόντων. Στο μοντέλο μας, 2 παραγόντων, ένα χαρτοφυλάκιο ασφαλίσεων υποδιαιρείται από δύο ποιοτικούς παράγοντες κινδύνου με συμμετρικό τρόπο, καθώς όπως ήδη αναφέρθηκε, οι παράγοντες είναι ίσης βαρύτητας σε επίπεδο ιεράρχησης. Αντί της δομής δέντρου στο μοντέλο του Jewell, που είδαμε στο προηγούμενο κεφάλαιο, ένας πιο φυσικός τρόπος για να αντιληφθούμε τη δομή ενός χαρτοφυλακίου, σαν αυτού που περιγράψαμε, είναι ο παρακάτω:



Σχήμα 6.1: Γενική μορφή μοντέλου 2 παραγόντων Σταυρωτής Ταξινόμησης.

Σε αυτό οι τιμές των ασφαλιστηρίων συμβολαίων συσχετίζονται με τις αντίστοιχες τιμές των παραγόντων που βρίσκονται στην ίδια σειρά και αντίστοιχα στην ίδια στήλη [βλέπε Dannenburg et al. (1996)].

Θεωρούμε το μοντέλο του *Jewell*,

$$X_{ijt} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijt}. \quad (6.1)$$

Η επίδραση στο ύψος των απαιτήσεων συγκεκριμένα για την κατηγορία i του πρώτου παράγοντα κινδύνου, τον οποίο θα ονομάσουμε παράγοντα γραμμών, δηλώνεται με α_i . Η αντίστοιχη επίδραση για την κατηγορία j του δεύτερου παράγοντα, του παράγοντα στηλών, για τον κίνδυνο στην γραμμή i περιγράφεται από την τυχαία μεταβλητή β_{ij} . Ωστόσο, δεν υπάρχει τυχαία μεταβλητή που να αντιπροσωπεύει μια κοινή επίδραση για όλους τους κινδύνους της στήλης j , ανεξαρτήτως από τη σειρά που αντιστοιχεί. Εάν τέτοιες πληροφορίες προστεθούν στο μοντέλο μέσω της τυχαίας μεταβλητής β_j τότε παίρνουμε την ακόλουθη επέκταση του μοντέλου του *Jewell*:

$$X_{ijt} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijt}. \quad (6.2)$$

Στην (6.2), σε σύγκριση με την (6.1), και οι δύο παράγοντες στο μοντέλο αυτό είναι μοντελοποιημένοι συμμετρικά. Οι τυχαίες μεταβλητές α_i και β_j υποδηλώνουν τις κύριες

επιπτώσεις της σειράς και της στήλης αντίστοιχα, και οι μεταβλητές $\alpha\beta_{ij}$ αντιπροσωπεύουν την αλληλεπίδραση μεταξύ των δύο παραγόντων (γραμμής-στήλης). Αυτό σημαίνει πως για κάθε τιμή του i και j το στοιχείο $\alpha\beta_{ij}$ υποδηλώνει τη συγκεκριμένη επίδραση των συμβολαίων που ανήκουν στη σειρά i και στη στήλη j . Ο εναπομείνας παράγοντας ε_{ijt} αντιπροσωπεύει την απόκλιση στο συμβόλαιο για την περίοδο παρατήρησης t από το μέγεθος της αναμενόμενης αξίωσης, δοθέντος α_i , β_j , και $\alpha\beta_{ij}$.

Εάν για παράδειγμα υποδιαιρέσουμε το χαρτοφυλάκιο σε κελιά σύμφωνα με το φύλο του ασφαλισμένου ($i = 1, 2$, θηλυκό - αρσενικό) και σύμφωνα με τις τρεις ηλικιακές κλάσεις ($j = 1, 2, 3$, νέοι - μέσης ηλικίας - ηλικιωμένοι), τότε αυτό μπορεί να απεικονιστεί ως εξής:

		Νέος	Μέσης Ηλικίας	Ηλικιωμένος	
Θηλ.		$\alpha\beta_{11}$	$\alpha\beta_{12}$	$\alpha\beta_{13}$	α_1
Αρσ.		$\alpha\beta_{21}$	$\alpha\beta_{22}$	$\alpha\beta_{23}$	α_2
		β_1	β_2	β_3	

Σχήμα 6.2: Αναπαράσταση του μοντέλου 2 παραγόντων (Ηλικία-Γένος) Σταυρωτής Ταξινόμησης.

Όπως φαίνεται στην εικόνα, η επίδραση στο μέγεθος της αξίωσης του να είναι νέος, είτε αρσενικού είτε θηλυκού γένους, αντιπροσωπεύεται από την τυχαία μεταβλητή β_1 . Ομοίως, α_1 είναι η επίδραση του να είναι γένους θηλυκού στο χαρτοφυλάκιο. Επιπλέον, η μεταβλητή $\alpha\beta_{11}$ εκφράζει την επιπρόσθετη επίδραση στο μέγεθος της αξίωσης του να είναι νέος και ταυτόχρονα θηλυκού γένους.

Στο κεφάλαιο αυτό θεωρούμε μόνο μοντέλο με δύο παράγοντες κινδύνου. Ωστόσο είναι δυνατό να κατασκευαστούν μοντέλα σταυρωτής ταξινόμησης με περισσότερους παράγοντες. Επίσης, είναι αντιληπτό πως τα δεδομένα δεν μπορούν να μοντελοποιηθούν

ούτε από ένα καθαρά σταυρωτής ταξινόμησης μοντέλο, ούτε από ένα ιεραρχικό, αλλά μόνο από ένα συνδιασμό αυτών. Ας θεωρήσουμε για παράδειγμα ένα χαρτοφυλάκιο ασφαλειών αυτοκινήτων στο οποίο οι ασφαλισμένοι κατηγοριοποιούνται σύμφωνα με την ηλικία τους και τα χιλιόμετρα που διένυσαν. Κάθε κελί που αντιστοιχεί σε συγκεκριμένο εύρος χιλιομέτρων και ηλικιών έχει τη δική του επίδραση και περιλαμβάνει έναν αριθμό συμβολαίων για τα οποία υπάρχει ένα πλήθος παρατηρήσεων κατά τη διάρκεια πολλών περιόδων. Εάν X_{ijkl} είναι ο κίνδυνος στην περίοδο t της παρατήρησης του συμβολαίου k που ανήκει στην ομάδα συμβολαίων ηλικιών j και κατηγορία χιλιομέτρων i , τότε έχουμε:

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_{ijk} + \varepsilon_{ijkl}. \quad (6.3)$$

Τα στοιχεία α_i και β_j είναι οι κύριες επιπτώσεις στην i κλάση της κατηγορίας χιλιομέτρων, και στην j των ηλικιών. Επιπλέον το $\alpha\beta_{ij}$ αντιπροσωπεύει το αποτέλεσμα της αλληλεπίδρασης μεταξύ χιλιομέτρων και ηλικίας. Μια επιπρόσθετη επίδραση ειδικά για το συμβόλαιο k στο κελί (i,j) περιγράφεται από το γ_{ijk} . Τυχαίες μεταβλητές, όπως $\gamma_{i,k}$, που δηλώνουν την επίδραση για κάθε k συμβόλαιο στην σειρά i δεν παρουσιάζονται στο (6.3). Η χρήση αυτού του στοιχείου θα δήλωνε ότι εντός των κλάσεων χιλιομέτρων i θα υπήρχε μια κοινή επίδραση σε όλα τα k συμβόλαια σε κάθε κλάση ηλικίας j . Σε πολλές περιπτώσεις αυτό είναι μη ρεαλιστικό, και ως εκ τούτου η (6.3) έχει εν μέρει ιεραρχική δομή όπως στο μοντέλο του Jewell.

Στη συνέχεια του κεφαλαίου, θα περιγραφεί η μεθοδολογία υπολογισμού του καθαρού ασφαλίστρου για το μοντέλο σταυρωτής ταξινόμησης δύο παραγόντων. Θα αποδειχθεί πως η εμπειρία κινδύνου εντός των παραγόντων (είτε της σειράς είτε της στήλης) συνοψίζεται από ένα σταθμισμένο μέσο όρο παρατηρούμενων κινδύνων στην κλάση του παράγοντα αυτού, προσαρμοσμένο για την επίδραση των χαρακτηριστικών του άλλου παράγοντα. Μια ειδική περίπτωση αποτελείται από τα ισορροπημένα μοντέλα (balanced), όπου όπως αναφέρθηκε και στο (2.1), σε κάθε κελί υπάρχει ίδιο πλήθος παρατηρήσεων και επιπλέον όλα τα βάρη w_{ijt} θα είναι ίσα με ένα. Στην πράξη ο αριθμός των παρατηρήσεων T_{ij} διαφέρει από το ένα κελί στο άλλο, ενώ κάποια μπορεί να μην έχουν καμία παρατήρηση. Εάν το X_{ijt} αντιπροσωπεύει μέσους όρους, τα βάρη συνήθως είναι διαφορετικά. Τέλος, όπως και στα προηγούμενα κεφάλαια έτσι και εδώ θα παρουσιαστεί η δυνατότητα υπολογισμού των συντελεστών του μοντέλου μας με τη χρήση της μεθόδου ανάλυσης διασποράς και θα χρησιμοποιηθούν όλα για την επίλυση ενός αριθμητικού προβλήματος στο στατιστικό πακέτο R [βλέπε Dannenburg et al. (1996)].

6.1 Μοντέλο Σταυρωτής Ταξινόμησης (Cross Classification)

Στο μοντέλο δύο παραγόντων σταυρωτής ταξινόμησης, οι κίνδυνοι σε ένα χαρτοφυλάκιο ασφαλίσεων ομαδοποιούνται κατά δυο παράγοντες κινδύνου. Το πλήθος των κλάσεων για τον πρώτο παράγοντα κινδύνου ορίζεται ως I (γραμμές) και του δεύτερου παράγοντα ως J (στήλες). Εντός των κελιών (i, j) υπάρχουν T_{ij} διαθέσιμες παρατηρήσεις των κινδύνων,

$$X_{ijt} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijt}, \quad t = 1, \dots, T_{ij}. \quad (6.4)$$

Τα στοιχεία, α , β , $\alpha\beta$, ε είναι ανεξάρτητα και με μέση τιμή μηδέν. Επίσης οι μεταβλητές α_i , β_j , $\alpha\beta_{ij}$, και ε_{ijt} έχουν διασπορές σ_α^2 , σ_β^2 , $\sigma_{\alpha\beta}^2$ και $\sigma_\varepsilon^2/w_{ijt}$ αντίστοιχα. Το σ_ε^2 , που όπως και στα προηγούμενα κεφάλαια θα συμβολίζεται και ως s^2 , είναι ένα μέτρο της μεταβλητότητας των κινδύνων μέσα στο κελί (i, j) , και το $\sigma_{\alpha\beta}^2$ αντανακλά τη διασπορά μεταξύ των κελιών που δεν σχετίζεται με τους δύο ατομικούς παράγοντες κινδύνου. Επιλέον οι παράμετροι σ_α^2 και σ_β^2 μπορούν να ερμηνευτούν ως η διασπορά μεταξύ των κελιών σε διαφορετικές γραμμές και στήλες αντίστοιχα [βλέπε Dannenburg et al. (1996)].

6.2 Υποθέσεις

1. Τα στοιχεία σε κάθε ένα από τα παρακάτω τυχαία διανύσματα είναι ανεξάρτητα και γνωστά κατανομημένα:
 - $\Theta_i = (\Theta_{i1}, \dots, \Theta_{iI})$
 - $\Theta_j = (\Theta_{j1}, \dots, \Theta_{jJ})$
2. Τα $(I + J)$ στοιχεία των τυχαίων διανυσμάτων Θ_i, Θ_j είναι ομοίως ανεξάρτητα μεταξύ των διανυσμάτων.
3. Οι κατά συνθήκη μέσοι, μπορούν να εκφραστούν ως εξής:

$$E[X_{ijt}|\Theta_i, \Theta_j] = \mu(\Theta_i, \Theta_j)$$

$$E[X_{ijt}|\Theta_i] = \mu(\Theta_i)$$

4. $E[Cov(X_{ijt}|\Theta_i, \Theta_j)] = \delta_{ijt,pqu}s^2/w_{ijt}$.

Όπως έχουμε δει και σε προηγούμενο κεφάλαιο, το $\delta_{ijt,pqu}$ είναι το σύμβολο Kronecker που ισούται με 1 εάν το $i = p$, $j = q$, και το $t = u$, διαφορετικά μηδέν [βλέπε Goulet (1998)].

6.3 Παράμετροι Δόμησης και Συμβολισμοί

Όπως και στα προηγούμενα κεφάλαια, έτσι και σε αυτό ο υπολογισμός των ασφαλιστρών γίνεται με χρήση του συντελεστή αξιοπιστίας.

Θεώρημα 15. Για το κελί (i, j) , ο συντελεστής αξιοπιστίας z_{ij} και ο σταθμισμένος μέσος όρος X_{ijw} που συνοψίζει την ειδική εμπειρία κινδύνου είναι:

$$Z_{ij} = \frac{\sigma_{\alpha\beta}^2}{\sigma_{\alpha\beta}^2 + \sigma_{\varepsilon}^2/w_{ij}}, \quad (6.8)$$

και

$$X_{ijw} = \sum_{t=1}^{T_{ij}} \frac{w_{ijt}}{w_{ij}} X_{ijt}.$$

Αυτός ο συντελεστής αξιοπιστίας έχει μια παρόμοια ερμηνεία με τον παράγοντα αξιοπιστίας του μοντέλου Jewell στο επίπεδο συμβολαίων. Αν η διακύμανση $\sigma_{\alpha\beta}^2$ μεταξύ των κελιών είναι σχετικά μεγάλη, τότε δίνεται μεγαλύτερη αξιοπιστία στο κελί (i, j) . Για την εμπειρία κινδύνου της γραμμής i , χρησιμοποιείται ο ακόλουθος συντελεστής αξιοπιστίας ως το βάρος με τον σταθμισμένο μέσο όρο αξιοπιστίας X_{izw} :

$$Z_i = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\alpha\beta}^2/z_i}, \quad (6.9)$$

και

$$X_{izw} = \sum_{j=1}^J \frac{z_{ij}}{z_i} X_{ijw}.$$

Αυτό σημαίνει πως εάν η διασπορά σ_{α}^2 μεταξύ των κλάσεων είναι υψηλή σε σχέση με τη διασπορά $\sigma_{\alpha\beta}^2$ μεταξύ των κελιών στην γραμμή i , περισσότερη αξιοπιστία δίνεται στην εμπειρία κινδύνου εντός της γραμμής i . Όμοια, ο συντελεστής αξιοπιστίας που αντιστοιχεί στην στήλη j είναι:

$$Z_j = \frac{\sigma_{\beta}^2}{\sigma_{\beta}^2 + \sigma_{\alpha\beta}^2/z_{.j}}, \quad (6.10)$$

και

$$X_{zjw} = \sum_{i=1}^I \frac{z_{ij}}{z_{.j}} X_{ijw}.$$

Η εκτιμήτρια για το ασφάλιστρο στο μοντέλο σταυρωτής ταξινόμησης δύο παραγόντων δίνεται από τον παρακάτω τύπο:

$$\mu(\Theta_{ij})^{Cred} = \mu + (1 - Z_{ij})(\hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\hat{\beta}_{ij}) + Z_{ij}(X_{ijw} - \mu),$$

όπου,

$$\begin{aligned}\hat{\alpha}_i &= Z_i(X_{izw} - \mu) - Z_i \sum_{j=1}^J \frac{Z_{ij}}{Z_i} \hat{\beta}_j, \\ \hat{\beta}_j &= Z_j(X_{zjw} - \mu) - Z_j \sum_{i=1}^I \frac{Z_{ij}}{Z_{\cdot j}} \hat{\alpha}_i, \\ \hat{\alpha}\hat{\beta}_{ij} &= Z_{ij}(X_{ijw} - \mu) - Z_{ij}(\hat{\alpha}_i + \hat{\beta}_j).\end{aligned}\tag{6.11}$$

και από την επίλυση του συστήματος των δύο πρώτων σχέσεων υπολογίζουμε τους όρους $\hat{\alpha}_i$, $\hat{\beta}_j$, τους οποίους χρησιμοποιούμε στην τρίτη σχέση για τον υπολογισμό του $\hat{\alpha}\hat{\beta}_{ij}$.

Απόδειξη: [βλέπε Goulet (1998)].

6.4 Σχέσεις Συνδιασπορών

Στο ισορροπημένο (balanced) μοντέλο σταυρωτής ταξινόμησης, ισχύουν οι κάτωθι τύποι:

$$Cov[X_{ijt}, X_{klu}] = \delta_{ik}\sigma_\alpha^2 + \delta_{jl}\sigma_\beta^2 + \delta_{ik}\delta_{jl}\sigma_{\alpha\beta}^2 + \delta_{ik}\delta_{jl}\delta_{tu} \frac{s^2}{w_{ijt}},$$

$$Cov[\bar{X}_{ij}, X_{klu}] = \delta_{ik}\sigma_\alpha^2 + \delta_{jl}\sigma_\beta^2 + \delta_{ik}\delta_{jl}(\sigma_{\alpha\beta}^2 + \frac{s^2}{T}),$$

[βλέπε Dannenburg et al. (1996)].

6.5 Αμερόληπτοι Εκτιμητές Παραμέτρων

Όπως είδαμε, και στο κεφάλαιο αυτό, ο υπολογισμός του ασφάλιστρου εξαρτάται από τις άγνωστες παραμέτρους δόμησης. Οι ακόλουθοι εκτιμητές των παραμέτρων αυτών προέρχονται από εκείνους του Dannenburg (1995).

Ένας απλός αμερόληπτος εκτιμητής του συνολικού μέσου όρου μ είναι ο σταθμισμένος μέσος όρος των παρατηρήσεων του χαροφυλακίου, χρησιμοποιώντας τα φυσικά βάρη:

$$\hat{\mu} = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^{t_{ij}} \frac{w_{ijt}}{w_{\dots}} X_{ijt},$$

με

$$w_{...} = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^{t_{ij}} w_{ijt}.$$

Ο ακόλουθος εκτιμητής για την παράμετρο s^2 είναι στενά συνδεδεμένος με τον εκτιμητή του s^2 στον μοντέλο του Jewell:

$$\hat{s}^2 = \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^{t_{ij}} w_{ijt} (X_{ijt} - X_{ijw})^2}{\sum_{i=1}^I \sum_{j=1}^J (t_{ij} - 1)_+}.$$

Για τον εκτιμητή του s^2 στην περίπτωση του ισορροπημένου μοντέλου (balanced), έχουμε τον ίδιο τύπο με μόνη διαφορά στον παρονομαστή όπου πλέον για κάθε i, j ισχύει $T_{ij} = T$. Άρα ισχύει ότι, $\sum_{i=1}^I \sum_{j=1}^J (t_{ij} - 1)_+ = IJ(T - 1)$.

Οι εκτιμητές των παραμέτρων σ_α^2 , σ_β^2 , $\sigma_{\alpha\beta}^2$, μπορούν να δοθούν ως λύσεις ενός γραμμικού συστήματος εξισώσεων. Έχουμε,

$$\sum_{i=1}^I g_i \left(\sum_{j=1}^J \frac{w_{ij.}}{w_{i..}} (X_{ijw} - X_{iww})^2 - \hat{s}^2 \frac{J-1}{w_{i..}} \right) = (\sigma_\beta^2 + \sigma_{\alpha\beta}^2) \left(1 - \sum_{i=1}^I g_i \sum_{j=1}^J \left(\frac{w_{ij.}}{w_{i..}} \right)^2 \right), \quad (6.17)$$

$$\sum_{j=1}^J g_j \left(\sum_{i=1}^I \frac{w_{ij.}}{w_{.j.}} (X_{ijw} - X_{wjwt})^2 - \hat{s}^2 \frac{I-1}{w_{.j.}} \right) = (\sigma_\beta^2 + \sigma_{\alpha\beta}^2) \left(1 - \sum_{j=1}^J g_j \sum_{i=1}^I \left(\frac{w_{ij.}}{w_{.j.}} \right)^2 \right), \quad (6.18)$$

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \frac{w_{ij.}}{w_{...}} (X_{ijw} - X_{www})^2 - \hat{s}^2 \frac{IJ-1}{w_{...}} &= \sigma_\alpha^2 \left(1 - \sum_{i=1}^I \left(\frac{w_{i..}}{w_{...}} \right)^2 \right) + \sigma_\beta^2 \left(1 - \sum_{j=1}^J \left(\frac{w_{.j.}}{w_{...}} \right)^2 \right) \\ &+ \sigma_{\alpha\beta}^2 \left(1 - \sum_{i=1}^I \sum_{j=1}^J \left(\frac{w_{ij.}}{w_{...}} \right)^2 \right). \end{aligned} \quad (6.19)$$

Οι τρεις γραμμικές εξισώσεις, (6.17), η ανάλογη αυτής (6.18) με τον ρόλο των i και j να εναλλάσσονται, και η (6.19) μπορούν να λυθούν και να δώσουν της εκτιμήτριες των σ_α^2 , σ_β^2 , και $\sigma_{\alpha\beta}^2$. Επιπλέον τα g_i , g_j , είναι αυθαίρετα βάρη που αθροίζουν σε ένα. Σημειώνεται ότι οι τιμές τους δεν είναι αναγκαστικά μη αρνητικές. Εάν εκ των προτέρων κανένας παράγοντας κινδύνου δεν φαίνεται να είναι πιο σημαντικός από τους άλλους για την εκτίμηση παραμέτρων, τότε πρέπει να χρησιμοποιηθούν ίσα βάρη.

Η διαδικασία εκτίμησης είναι απλούστερη από την αρχική εμφάνιση. Ωστόσο απαιτούνται αρκετοί υπολογισμοί όσο ο αριθμός των παραγόντων και των κατηγοριών κινδύνου αυξάνεται.

Το πλήρες σύστημα εξισώσεων που χρειάζεται να λυθεί για τον προσδιορισμό των παραμέτρων σ_α^2 , σ_β^2 , και $\sigma_{\alpha\beta}^2$ δίνεται παρακάτω, θεωρώντας ίσα βάρη. Αρχικά ορίζονται οι ακόλουθοι όροι:

$$\begin{aligned} x_1 &= \frac{1}{I} \sum_{i=1}^I \left(\sum_{j=1}^J \frac{w_{ij.}}{w_{i..}} (X_{ijw} - X_{iww})^2 - \hat{s}^2 \frac{J-1}{w_{i..}} \right), \\ x_2 &= \frac{1}{J} \sum_{j=1}^J \left(\sum_{i=1}^I \frac{w_{ij.}}{w_{.j}} (X_{ijw} - X_{wjw})^2 - \hat{s}^2 \frac{I-1}{w_{.j}} \right), \\ x_3 &= \sum_{i=1}^I \sum_{j=1}^J \frac{w_{ij.}}{w_{...}} (X_{ijw} - X_{www})^2 - \hat{s}^2 \frac{IJ-1}{w_{...}}, \end{aligned}$$

και

$$\begin{aligned} y_1 &= 1 - \frac{1}{I} \sum_{i=1}^I \sum_{j=1}^J \left(\frac{w_{ij.}}{w_{i..}} \right)^2, \\ y_2 &= 1 - \frac{1}{J} \sum_{j=1}^J \sum_{i=1}^I \left(\frac{w_{ij.}}{w_{.j}} \right)^2, \\ y_3 &= 1 - \sum_{i=1}^I \left(\frac{w_{i..}}{w_{...}} \right)^2, \\ y_4 &= 1 - \sum_{j=1}^J \left(\frac{w_{.j}}{w_{...}} \right)^2, \\ y_5 &= 1 - \sum_{i=1}^I \sum_{j=1}^J \left(\frac{w_{ij.}}{w_{...}} \right)^2. \end{aligned}$$

Οι εκτιμητές, $\hat{\sigma}_\alpha^2$, $\hat{\sigma}_\beta^2$, $\hat{\sigma}_{\alpha\beta}^2$, είναι στη συνέχεια οι λύσεις του συστήματος των γραμμικών εξισώσεων που προκύπτουν από:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 & y_1 & y_1 \\ y_2 & 0 & y_2 \\ y_3 & y_4 & y_5 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\beta^2 \\ \hat{\sigma}_{\alpha\beta}^2 \end{pmatrix}.$$

Συνεπώς

$$\begin{aligned}x_1 &= y_1\hat{\sigma}_\beta^2 + y_1\hat{\sigma}_{\alpha\beta}^2, \\x_2 &= y_2\hat{\sigma}_\alpha^2 + y_2\hat{\sigma}_{\alpha\beta}^2, \\x_3 &= y_3\hat{\sigma}_\alpha^2 + y_4\hat{\sigma}_\beta^2 + y_5\hat{\sigma}_{\alpha\beta}^2.\end{aligned}$$

[βλέπε Goulet (1998)]

6.6 Προσέγγιση με τη μέθοδο της Ανάλυσης Διασποράς

Για το μοντέλο δύο παραγόντων σταυρωτής ταξινόμησης (6.4) δεν υπάρχει μοναδική ανάλυση διασποράς.

Θεώρημα 16. Στο ισορροπημένο μοντέλο, οι αμερόληπτοι εκτιμητές των παραμέτρων s^2 , σ_α^2 , σ_β^2 , $\sigma_{\alpha\beta}^2$ είναι:

$$\begin{aligned}\hat{s}^2 &= MSE \\ \hat{\sigma}_{\alpha\beta}^2 &= \frac{MSAB - MSE}{T}, \\ \hat{\sigma}_\alpha^2 &= \frac{MSA - MSAB}{JT}, \\ \hat{\sigma}_\beta^2 &= \frac{MSB - MSAB}{IT}.\end{aligned}$$

[βλέπε Sahai and Ojeda (2013)]

Απόδειξη: Για το ισορροπημένο μοντέλο ο πίνακας ανάλυσης διασποράς είναι:

Πηγή Διασποράς	Βαθμοί Ελευθ.	Αθρ. Τετρ.	Μέσος Τετρ.
Παράγοντας A	$I - 1$	SSA	MSA
Παράγοντας B	$J - 1$	SSB	MSB
Αλληλεπίδραση AB	$(I - 1)(J - 1)$	$SSAB$	$MSAB$
Σφάλμα	$IJ(T - 1)$	SSE	MSE

Πίνακας 6.1: Ανάλυση διασποράς ισορροπημένων δεδομένων.

Από τα αθροίσματα των τετραγώνων του πίνακα προκύπτουν τα εξής:

$$\begin{aligned}
E(MSE) &= \frac{1}{IJ(T-1)} E\left[\sum_{i=1}^I \sum_{j=1}^J (T-1)s^2\right] \\
&= \frac{IJ(T-1)s^2}{IJ(T-1)} \\
&= s^2.
\end{aligned}$$

Έτσι έχουμε,

$$\hat{s}^2 = MSE.$$

Ομοίως έχουμε:

$$\begin{aligned}
E(MSAB) &= \frac{T}{(I-1)(J-1)} E\left[\sum_{i=1}^I \sum_{j=1}^J (\bar{X}_{ij} - \bar{X}_{i..} - \bar{X}_{.j} + \bar{X}_{...})^2\right] \\
&= \frac{T}{(I-1)(J-1)} \sum_{i=1}^I [E\left[\sum_{j=1}^J (\alpha\beta_{ij} - \bar{\alpha}\beta_{i.} - \bar{\alpha}\beta_{.j} + \bar{\alpha}\beta_{..})^2\right] \\
&\quad + E\left[\sum_{j=1}^J (\bar{\varepsilon}_{ij} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j} + \bar{\varepsilon}_{...})^2\right]],
\end{aligned}$$

όπου θέτοντας $q_j = \alpha\beta_{ij} - \bar{\alpha}\beta_{.j}$, με $j = 1, 2, \dots, J$, $\bar{q} = \bar{\alpha}\beta_{i.} - \bar{\alpha}\beta_{..}$, $Var(q_j) = (I-1)\sigma_{\alpha\beta}^2/I$ και τα q_j ασυσχέτιστα με μέσο μ έχουμε εξ ορισμού:

$$\begin{aligned}
E\left[\sum_{j=1}^J (q_j - \bar{q})^2\right] &= E(q_j^2) - JE(\bar{q})^2 \\
&= J(\mu^2 + (I-1)\sigma_{\alpha\beta}^2/I) - (J\mu^2 + (I-1)\sigma_{\alpha\beta}^2/I) \\
&= (J-1)[(I-1)\sigma_{\alpha\beta}^2/I] \tag{6.46}
\end{aligned}$$

Όμοια με τη σχέση (6.46) έχουμε πως:

$$E\left[\sum_{j=1}^J (\bar{\varepsilon}_{ij} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j} + \bar{\varepsilon}_{...})^2\right] = (J-1)[(I-1)s^2/IT].$$

Από τα δύο τελευταία καταλήγουμε πως:

$$\begin{aligned} E(MSAB) &= \frac{T}{(I-1)(J-1)} \sum_{i=1}^I [(J-1)(I-1)\sigma_{\alpha\beta}^2/I + (J-1)(I-1)s^2/IT] \\ &= s^2 + T\sigma_{\alpha\beta}^2. \end{aligned}$$

Έτσι,

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{MSAB - MSE}{T}.$$

Επιπλέον,

$$\begin{aligned} E(MSB) &= \frac{IT}{J-1} E\left[\sum_{j=1}^J (\bar{X}_{.j} - \bar{X}_{...})^2\right] \\ &= \frac{IT}{J-1} [E\left[\sum_{j=1}^J (\beta_j - \bar{\beta})^2\right] + E\left[\sum_{j=1}^J (\bar{\alpha}\beta_{.j} - \bar{\alpha}\bar{\beta})^2\right] \\ &\quad + E\left[\sum_{j=1}^J (\bar{\varepsilon}_{.j} - \bar{\varepsilon}_{...})^2\right], \end{aligned} \quad (6.57)$$

και αντίστοιχα με τη σχέση (6.46) προκύπτουν:

$$E\left[\sum_{j=1}^J (\beta_j - \bar{\beta})^2\right] = (J-1)\sigma_{\beta}^2, \quad (6.58)$$

$$E\left[\sum_{j=1}^J (\bar{\alpha}\beta_{.j} - \bar{\alpha}\bar{\beta})^2\right] = (J-1)\sigma_{\alpha\beta}^2/I, \quad (6.59)$$

και

$$E\left[\sum_{j=1}^J (\bar{\varepsilon}_{.j} - \bar{\varepsilon}_{...})^2\right] = (J-1)s^2/IT. \quad (6.60)$$

Αντικαθιστώντας στην σχέση (6.57) τις σχέσεις (6.58), (6.59), (6.60) καταλήγουμε πως

$$E(MSB) = \frac{IT}{J-1} [(J-1)\sigma_{\beta}^2 + (J-1)\sigma_{\alpha\beta}^2/I + (J-1)s^2/IT]$$

$$= s^2 + T\sigma_{\alpha\beta}^2 + IT\sigma_{\beta}^2. \quad (6.61)$$

Τέλος από τη συμμετρία μεταξύ MSA και MSB είναι επακόλουθο πως

$$E(MSA) = s^2 + T\sigma_{\alpha\beta}^2 + JT\sigma_{\alpha}^2. \quad (6.62)$$

Από τις (6.61), (6.62) προκύπτουν οι εκτιμητές $\hat{\sigma}_{\alpha}^2$ και $\hat{\sigma}_{\beta}^2$.

Θεώρημα 17. Οι αμερόληπτοι εκτιμητές των παραμέτρων s^2 , σ_{α}^2 , σ_{β}^2 , $\sigma_{\alpha\beta}^2$ για το μη ισορροπημένο μοντέλο είναι:

$$\begin{aligned} \hat{s}^2 &= \frac{SSE}{T - k}, \\ \hat{\sigma}_{\alpha}^2 &= \frac{1}{T - r_4} [SSAB + SSA - (k - J)\hat{s}^2] - \hat{\sigma}_{\alpha\beta}^2, \\ \hat{\sigma}_{\beta}^2 &= \frac{1}{T - r_3} [SSAB + SSB - (k - I)\hat{s}^2] - \hat{\sigma}_{\alpha\beta}^2, \\ \hat{\sigma}_{\alpha\beta}^2 &= \left\{ \frac{T - r_1}{T - r_4} [SSAB + SSA - (k - J)\hat{s}^2] + \frac{r_3 - r_2}{T - r_3} [SSAB \right. \\ &\quad \left. + SSB - (k - I)\hat{s}^2] - [SSA - (I - 1)\hat{s}^2] \right\} / (T - r_1 - r_2 + r_{12}). \end{aligned}$$

[βλέπε *Sahai and Ojeda (2007)*]

Απόδειξη: Η συμβατική ανάλυση διασποράς για το μη ισορροπημένο μοντέλο, λαμβάνεται από μια αναλογία με την αντίστοιχη ισορροπημένη ανάλυση και δίνεται στον παρακάτω πίνακα:

Πηγή Διασποράς	Βαθμοί Ελευθ.	Αθρ. Τετρ.	Μέσος Τετρ.
Παράγοντας Α	$I - 1$	SSA	MSA
Παράγοντας Β	$I(J - 1)$	SSB	MSB
Αλληλεπίδραση ΑΒ	$k - I - J + 1$	$SSAB$	$MSAB$
Σφάλμα	$T - k$	SSE	MSE

Πίνακας 6.2: Ανάλυση διασποράς μη ισορροπημένων δεδομένων.

όπου k είναι το πλήθος των μη κενών κελιών.

Τα αθροίσματα των τετραγώνων του πίνακα ορίζονται ως εξής:

$$\begin{aligned}
 SSA &= \sum_{i=1}^I t_i (\bar{X}_{i..} - \bar{X}_{...})^2 \\
 &= \sum_{i=1}^I \frac{X_{i..}^2}{t_i} - \frac{X_{...}^2}{T}, \\
 SSB &= \sum_{j=1}^J t_j (\bar{X}_{.j} - \bar{X}_{...})^2 \\
 &= \sum_{j=1}^J \frac{X_{.j}^2}{t_j} - \frac{X_{...}^2}{T}, \\
 SSAB &= \sum_{i=1}^I \sum_{j=1}^J t_{ij} \bar{X}_{ij}^2 - \sum_{i=1}^I t_i \bar{X}_{i..}^2 - \sum_{j=1}^J t_j \bar{X}_{.j}^2 + T \bar{X}_{...}^2 \\
 &= \sum_{i=1}^I \sum_{j=1}^J \frac{X_{ij}^2}{t_{ij}} - \sum_{i=1}^I \frac{X_{i..}^2}{t_i} - \sum_{j=1}^J \frac{X_{.j}^2}{t_j} + \frac{X_{...}^2}{T},
 \end{aligned}$$

και

$$\begin{aligned}
 SSE &= \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^{t_{ij}} (X_{ijt} - \bar{X}_{ij.})^2 \\
 &= \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^{t_{ij}} X_{ijt}^2 - \sum_{i=1}^I \sum_{j=1}^J \frac{X_{ij.}^2}{t_{ij}},
 \end{aligned}$$

όπου

$$\begin{aligned}
 X_{ij.} &= \sum_{t=1}^{t_{ij}} X_{ijt}, & \bar{X}_{ij.} &= \frac{X_{ij.}}{t_{ij}} \\
 X_{i..} &= \sum_{j=1}^J X_{ij.}, & \bar{X}_{i..} &= \frac{X_{i..}}{t_i} \\
 X_{.j} &= \sum_{i=1}^I X_{ij.}, & \bar{X}_{.j} &= \frac{X_{.j}}{t_j}
 \end{aligned}$$

$$X_{...} = \sum_{i=1}^I X_{i..} = \sum_{j=1}^J X_{.j.}, \quad \bar{X}_{...} = \frac{X_{...}}{T},$$

με

$$t_{i.} = \sum_{j=1}^J t_{ij}, \quad t_{.j} = \sum_{i=1}^I t_{ij},$$

$$T = \sum_{i=1}^I t_{i.} = \sum_{j=1}^J t_{.j} = \sum_{i=1}^I \sum_{j=1}^J t_{ij}.$$

Τα SSA , SSB , $SSAB$, και SSE έχουν οριστεί καθιερώνοντας μια αναλογία με τους αντίστοιχους όρους στην περίπτωση των ισορροπημένων δεδομένων.

Εάν θέσουμε τα εξής:

$$D_A = \sum_{i=1}^I \frac{X_{i..}^2}{t_{i.}},$$

$$D_B = \sum_{j=1}^J \frac{X_{.j.}^2}{t_{.j}},$$

$$D_{AB} = \sum_{i=1}^I \sum_{j=1}^J \frac{X_{ij.}^2}{t_{ij}},$$

$$D_0 = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^{t_{ij}} X_{ijt}^2,$$

$$D_{\mu} = \frac{X_{...}^2}{T},$$

τότε τα αθροίσματα των τετραγώνων, μπορούν να γραφούν ως εξής:

$$SSA = D_A - D_{\mu},$$

$$SSB = D_B - D_{\mu},$$

$$SSAB = D_{AB} - D_A - D_B + D_{\mu},$$

$$SSE = D_0 - D_{AB}.$$

Γενικά, οι παραπάνω εκφράσεις δεν είναι όλες αθροίσματα τετραγώνων. Το $SSAB$ για παράδειγμα, μπορεί να είναι αρνητικό και έτσι δεν μπορεί να είναι άθροισμα τετραγώνων. Τα μέσα τετράγωνα ως συνήθως λαμβάνονται αν διαιρέσουμε τα αθροίσματα των τετραγώνων με τους αντίστοιχους βαθμούς ελευθερίας. Στη συνέχεια περιγράφονται τα αποτελέσματα των αναμενόμενων μέσων τετραγώνων.

Τα αναμενόμενα ποσά τετραγώνων ή μέσων τετραγώνων λαμβάνονται εύκολα, αν υπολογίσουμε αρχικά τις αναμενόμενες τιμές των ποσοτήτων D_0 , D_A , D_B , και D_μ . Αρχικά να σημειωθεί ότι όπως προέκυψε και στο (2.6) έτσι και εδώ ισχύει ότι $E(\alpha_i) = 0$, και $E(\alpha_i^2) = \sigma_\alpha^2$. Αντίστοιχα αποτελέσματα έχουμε για τα β_j , $\alpha\beta_{ij}$ και ε_{ij} . Επιπλέον, όλες οι συνδιασπορές μεταξύ ζευγών των τυχαίων μεταβλητών είναι μηδέν. Σύμφωνα με αυτά προκύπτουν τα εξής:

$$\begin{aligned}
E(D_0) &= \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^{t_{ij}} E[\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijt}]^2 \\
&= T(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + s^2), \\
E(D_{AB}) &= \sum_{i=1}^I \sum_{j=1}^J \frac{1}{t_{ij}} E[\mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij}]^2 \\
&= T(\mu^2 + \sigma_\alpha^2 + \sigma_\beta^2) + ks^2, \\
E(D_A) &= \sum_{i=1}^I \frac{1}{t_i} E[t_i\mu + t_i\alpha_i + \sum_{j=1}^J t_{ij}\beta_j + \sum_{j=1}^J t_{ij}(\alpha\beta)_{ij} + \varepsilon_{i..}]^2 \\
&= T(\mu^2 + \sigma_\alpha^2) + r_3(\sigma_\beta^2 + \sigma_{\alpha\beta}^2) + Is^2, \\
E(D_B) &= \sum_{j=1}^J \frac{1}{t_{.j}} E[t_{.j}\mu + t_{.j}\beta_j + \sum_{i=1}^I t_{ij}\alpha\beta_{ij} + \varepsilon_{.j}]^2 \\
&= T(\mu^2 + \sigma_\beta^2) + r_4(\sigma_{\alpha\beta}^2) + Js^2, \\
E(D_\mu) &= \frac{1}{T} E[T\mu + \sum_{i=1}^I t_i\alpha_i + \sum_{j=1}^J t_{.j}\beta_j + \sum_{i=1}^I \sum_{j=1}^J t_{ij}\alpha\beta_{ij} + \varepsilon_{...}]^2 \\
&= T\mu^2 + r_1\sigma_\alpha^2 + r_2\sigma_\beta^2 + r_{12}\sigma_{\alpha\beta}^2 + s^2,
\end{aligned}$$

όπου

$$\begin{aligned}
r_1 &= \frac{1}{T} \sum_{i=1}^I t_i^2, & r_2 &= \frac{1}{T} \sum_{j=1}^J t_{.j}^2 \\
r_3 &= \sum_{i=1}^I \frac{\sum_{j=1}^J t_{ij}^2}{t_i}, & r_4 &= \sum_{j=1}^J \frac{\sum_{i=1}^I t_{ij}^2}{t_{.j}} \\
r_{12} &= \frac{1}{T} \sum_{i=1}^I \sum_{j=1}^J t_{ij}^2
\end{aligned}$$

Έτσι, οι αναμενόμενες τιμές των αθροισμάτων τετραγώνων και των μέσων τετραγώνων δίνονται παρακάτω:

$$\begin{aligned}
E(SSE) &= E[D_0 - D_{AB}] \\
&= (T - k)s^2, \\
E(MSE) &= \frac{1}{T - k}E(SSE) \\
&= s^2, \\
E(SSAB) &= E[D_{AB} - D_A - D_B + D_\mu] \\
&= (k - I - J + 1)s^2 + (T - r_3 - r_4 + r_{12})\sigma_{\alpha\beta}^2 + (r_2 - r_3)\sigma_\beta^2 + (r_1 + r_4)\sigma_\alpha^2, \\
E(MSAB) &= \frac{1}{k - I - J + 1}E(SSAB) \\
&= s^2 + \frac{(T - r_3 - r_4 + r_{12})\sigma_{\alpha\beta}^2 + (r_2 - r_3)\sigma_\beta^2 + (r_1 + r_4)\sigma_\alpha^2}{k - I - J + 1}, \\
E(SSB) &= E[D_B - D_\mu] \\
&= (J - 1)s^2 + (r_4 - r_{12})\sigma_{\alpha\beta}^2 + (T - r_2)\sigma_\beta^2 + (r_4 - r_1)\sigma_\alpha^2, \\
E(MSB) &= \frac{1}{J - 1}E(SSB) \\
&= s^2 + \frac{(r_4 - r_{12})\sigma_{\alpha\beta}^2 + (T - r_2)\sigma_\beta^2 + (r_4 - r_1)\sigma_\alpha^2}{J - 1}, \\
E(SSA) &= E[D_A - D_\mu] \\
&= (I - 1)s^2 + (r_3 - r_{12})\sigma_{\alpha\beta}^2 + (r_3 - r_2)\sigma_\beta^2 + (T - r_1)\sigma_\alpha^2, \\
E(MSA) &= \frac{1}{I - 1}E(SSA) \\
&= s^2 + \frac{(r_3 - r_{12})\sigma_{\alpha\beta}^2 + (r_3 - r_2)\sigma_\beta^2 + (T - r_1)\sigma_\alpha^2}{I - 1}.
\end{aligned}$$

Μια σημαντική όψη του αποτελέσματος $E(MSA)$ είναι πως έχει ένα μη μηδενικό συντελεστή για κάθε μεταβλητή διασποράς στο μοντέλο, ενώ με ισορροπημένα δεδομένα η αντίστοιχη αναμενόμενη τιμή δεν περιέχει όρους σ_β^2 . Ο όρος αυτός μειώνεται στο μηδέν στην περίπτωση του ισορροπημένου μοντέλου. Έτσι όταν έχουμε, $t_{ij} = t$, $t_{i.} = Jt$, $t_{.j} = It$ και $T = IJt$, ο συντελεστής του σ_β^2 στο $E(MSA)$ είναι:

$$\frac{r_3 - r_2}{I - 1} = \frac{\sum_{i=1}^I \frac{\sum_{j=1}^J t_{ij}^2}{t_{i.}} - \frac{1}{T} \sum_{j=1}^J t_{.j}^2}{I - 1}$$

$$\begin{aligned}
&= \frac{I\left(\frac{Jt^2}{Jt}\right) - J\left(\frac{I^2t^2}{IJt}\right)}{I-1} \\
&= 0.
\end{aligned}$$

Αντίστοιχα, ο συντελεστής του όρου σ_α^2 στο $E(MSA)$ γίνεται:

$$\begin{aligned}
\frac{T-r_1}{I-1} &= \frac{T - \frac{1}{T} \sum_{i=1}^I t_i^2}{I-1} \\
&= \frac{IJt - \frac{IJ^2t^2}{IJt}}{I-1} \\
&= Jt.
\end{aligned}$$

Τέλος ο συντελεστής του $\sigma_{\alpha\beta}^2$ ισούται με:

$$\begin{aligned}
\frac{r_3 - r_{12}}{I-1} &= \frac{\sum_{i=1}^I \frac{\sum_{j=1}^J t_{ij}^2}{t_i} - \frac{1}{T} \sum_{i=1}^I \sum_{j=1}^J t_{ij}^2}{I-1} \\
&= \frac{I\left(\frac{Jt^2}{Jt}\right) - J\left(\frac{Jt^2}{IJt}\right)}{I-1} \\
&= t.
\end{aligned}$$

Έτσι, για ισορροπημένα δεδομένα έχουμε:

$$E(MSA) = \sigma_\varepsilon^2 + t\sigma_{\alpha\beta}^2 + Jt\sigma_\alpha^2.$$

Παρόμοιες παρατηρήσεις και απλουστεύσεις ισχύουν και για τους όρους MSB και $MSAB$.

Για την εκτίμηση των παραμέτρων διασποράς αρκεί να εξισώσουμε τα αθροίσματα τετραγώνων ή των μέσων τετραγώνων με τις αντίστοιχες αναμενόμενες τιμές. Έτσι έχουμε:

$$SSE = (T-k)\hat{s}^2,$$

$$SSA = (T-r_1)\hat{\sigma}_\alpha^2 + (r_3-r_2)\hat{\sigma}_\beta^2 + (r_3-r_{12})\hat{\sigma}_{\alpha\beta}^2 + (I-1)\hat{s}^2,$$

$$SSB = (r_4 - r_1)\hat{\sigma}_\alpha^2 + (T - r_2)\hat{\sigma}_\beta^2 + (r_4 - r_{12})\hat{\sigma}_{\alpha\beta}^2 + (J - 1)\hat{s}^2,$$

και

$$SSAB = (r_1 - r_4)\hat{\sigma}_\alpha^2 + (r_2 - r_3)\hat{\sigma}_\beta^2 + (T - r_3 - r_4 + r_{12})\hat{\sigma}_{\alpha\beta}^2 + (k - I - J + 1)\hat{s}^2.$$

Οι εκτιμητές των παραμέτρων λαμβάνονται λύνοντας τις παραπάνω εξισώσεις ως προς $\hat{\sigma}_\alpha^2$, $\hat{\sigma}_\beta^2$, $\hat{\sigma}_{\alpha\beta}^2$ και \hat{s}^2 . Συγκεκριμένα:

$$\hat{s}^2 = \frac{SSE}{T - k}, \quad (6.133)$$

και

$$\begin{pmatrix} \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\beta^2 \\ \hat{\sigma}_{\alpha\beta}^2 \end{pmatrix} = P^{-1} \begin{pmatrix} SSA - (I - 1)\hat{s}^2 \\ SSB - (J - 1)\hat{s}^2 \\ SSAB - (k - I - J + 1)\hat{s}^2 \end{pmatrix},$$

όπου

$$P = \begin{pmatrix} T - r_1 & r_3 - r_2 & r_3 - r_{12} \\ r_4 - r_1 & T - r_2 & r_4 - r_{12} \\ r_1 - r_4 & r_2 - r_3 & T - r_3 - r_4 + r_{12} \end{pmatrix}.$$

Έτσι τελικά λαμβάνουμε τα παρακάτω:

$$\hat{\sigma}_\alpha^2 = \frac{1}{T - r_4} [SSAB + SSA - (k - J)\hat{s}^2] - \hat{\sigma}_{\alpha\beta}^2, \quad (6.134)$$

$$\hat{\sigma}_\beta^2 = \frac{1}{T - r_3} [SSAB + SSB - (k - I)\hat{s}^2] - \hat{\sigma}_{\alpha\beta}^2, \quad (6.135)$$

$$\begin{aligned} \hat{\sigma}_{\alpha\beta}^2 = & \left\{ \frac{T - r_1}{T - r_4} [SSAB + SSA - (k - J)\hat{s}^2] + \frac{r_3 - r_2}{T - r_3} [SSAB + SSB \right. \\ & \left. - (k - I)\hat{s}^2] - [SSA - (I - 1)\hat{s}^2] \right\} / (T - r_1 - r_2 + r_{12}). \end{aligned} \quad (6.136)$$

6.7 Αριθμητικό Παράδειγμα στο Μοντέλο Σταυρωτής Ταξινόμησης (Cross Classification)

Ως παράδειγμα θα εφαρμόσουμε το μοντέλο δύο παραγόντων σταυρωτής ταξινόμησης σε δεδομένα που αναφέρονται σε ιδιωτικά δάνεια [βλέπε Dannenburg et al. (1996)]. Ο Πίνακας που ακολουθεί περιέχει πληρωμές του ασφαλιστή πιστώσεων σε διάφορες τράπεζες για την κάλυψη

$i=1, j=1$	75.30	7.72	133.23	202.63	223.68	199.95	49.62	63.97	293.02	218.24	242.69
	113.68	196.30	380.65	388.53	326.63	91.41	326.63	106.38	266.48	156.06	43.80
	484.22	90.07	27.06	122.02	80.97	61.24	133.68	445.67	265.84	231.10	10.29
	325.23	68.63	87.61	84.79	27.48	314.31	248.60				
$i=2, j=1$	54.92	164.22	396.50	554.12	163.77	52.67	77.99	65.40	102.35	107.72	79.17
	208.78	106.38	268.77	209.69	80.80	166.02	148.62	71.84	84.18	254.47	97.79
	80.97	665.58	73.22	109.34	154.87	184.61	55.57	173.10	7.06	486.38	176.24
	62.66	67.74	26.78	305.27	213.05	47.18	87.88	321.16	86.70	452.24	300.52
$i=3, j=1$	69.13	141.89	134.91	118.57	180.37	106.38	491.71	5.05	321.15	93.98	
	363.10	267.58	66.82	93.05	295.78	59.36	313.61	395.56	118.71	150.96	371.92
	166.47	312.18	103.37	452.58	293.79	286.09	96.27	403.13	540.88	523.66	103.35
	140.82	196.93	298.89	96.28	187.40	136.81	24.92	94.36	261.02	109.34	81.67
$i=1, j=2$	150.11	87.44	137.64	113.57	332.67	51.60					
	158.72	616.01	405.37	53.58	123.04	34.98	316.44	323.07	364.56	221.55	165.78
	477.92	107.48	253.04	406.20	58.31	77.93	103.37	400.06	550.82	101.68	358.21
	56.18	513.20	82.79	151.52	342.59	402.08	413.70	257.69	47.99	638.10	97.31
$i=2, j=2$	133.48	113.57	177.84	406.86	225.10	234.76	406.86	56.09	25.05	147.45	
	513.14	40.35	529.67	260.59	425.27	163.77	640.47	53.00	272.63	81.67	378.23
	494.40	264.02	95.01	316.37	133.37	532.93	145.99	42.67	157.20	202.68	331.85
	27.32	563.61	273.65	97.04	150.13	332.02	184.14	102.29	398.88	77.34	47.94
$i=3, j=2$	319.02	97.72	416.72	456.57	107.96	150.90	483.78	134.27	147.06	192.16	149.56
	309.78	98.96	111.04	166.00	56.23	134.27	46.66	228.23	194.81		
	173.69	63.47	414.06	205.19	295.89	536.37	8.29	74.51	100.01	20.97	403.24
	221.20	196.30	400.06	405.70	203.34	398.01	90.92	609.78	12.82	142.28	608.41
$i=1, j=3$	311.53	496.28	268.45	111.55	168.30	33.02	84.40	350.82	497.90	12.11	122.62
	457.48	168.02	309.32	381.98	748.68	366.05					
	293.78	185.62	488.89	314.31	46.47	49.86	483.03	73.09	527.52	214.25	37.12
	295.07	447.74	242.96	298.26	222.98	219.79	539.43	22.68	420.87	155.79	34.23
$i=2, j=3$	170.73	252.84	84.31	358.32	287.12	411.58	434.72	502.44	247.09	157.63	145.27
	314.85	400.06	56.56	210.21	251.92	561.44	117.82	145.92			
	215.06	440.24	48.16	10.29	382.19	374.19	563.36	271.07	37.10	36.43	324.44
	133.16	392.61	135.26	238.92	313.01	388.53	139.42	77.37	314.01	60.60	134.11
$i=3, j=3$	163.87	290.97	361.60	414.72	529.82	471.13	272.92	520.18	258.75	232.75	94.52
	421.51	487.45	380.65	416.76	277.07	214.37	157.63	89.62	36.85	99.73	290.97
	78.43	209.32	209.51	143.69							
	48.56	57.06	518.60	287.28	480.38	423.87	711.86	259.00	505.18	177.51	394.34
$i=3, j=3$	145.99	95.04	110.22	463.98	129.01	469.67	516.97	203.34	386.37	103.11	331.60
	105.88	643.38	724.22	648.69	783.54	814.95	251.56	685.96	162.40	602.69	203.00
	591.94	239.66	221.54	182.72	297.93	556.78	208.54	133.16	509.60	274.23	469.48

των ζημιών που προκλήθηκαν από πελάτες που δεν μπόρεσαν να εξοφλήσουν τα προσωπικά τους δάνεια.

Οι παρατηρήσεις έχουν κατηγοριοποιηθεί σύμφωνα με την οικογενειακή κατάσταση του δανειολήπτη (I) καθώς και το χρονικό διάστημα που βρίσκεται στον εργασιακό χώρο που απασχολείται (J). Κάθε ένας από τους δύο αυτούς παράγοντες κινδύνου αποτελείται από τρεις κατηγορίες και επομένως για το μοντέλο μας θα ισχύει $I = J = 3$. Ο πελάτης μπορεί να είναι μόνος, διαζευγμένος ή άλλο (το οποίο περιέχει: παντρεμένος, συζεί ή να είναι χήρος). Ο χρόνος εργασίας με τον συγκεκριμένο εργοδότη διακρίνεται επίσης σε τρεις κατηγορίες, με την πρώτη να είναι λιγότερο από δύο έτη, η επομένη από δύο έως δέκα έτη και η τρίτη πάνω από δέκα χρόνια.

Το συνολικό πλήθος των παρατηρήσεων είναι 401, με το πλήθος να κυμαίνονται από 39 ως 54 ανά κελί. Στον παρακάτω Πίνακα 6.3, συνοψίζονται τα μέσα ποσά \bar{X}_{ijw} . Τα εννεά αυτά ποσά είναι οι μέσοι των κελιών, δηλαδή οι μέσοι των τιμών του πίνακα για κάθε ζευγάρι (i, j) . Στα περιθώρια του πίνακα δίνονται οι μέσοι της κάθε σειράς \bar{X}_{iww} και στήλης αντίστοιχα \bar{X}_{wjw} , ενώ η τιμή κάτω δεξιά είναι η τιμή \bar{X}_{www} που θα μπορούσε να περιγραφεί ως γενικός μέσος.

	< 2 ($j = 1$)	2 – 10 ($j = 2$)	≥ 10 ($j = 3$)	
Μόνος ($i = 1$)	180,385	246,705	261,574	230,228
Διαζευγ. ($i = 2$)	172,045	232,652	253,215	217,905
Άλλο ($i = 3$)	212,299	269,564	366,608	286,258
-	186,357	247,792	293,305	242,51

Πίνακας 6.3: Σταθμισμένοι μέσοι όροι των δεδομένων.

Ανάλυση Διασποράς: Από την ανάλυση διασπορά για τα δεδομένα μας [βλέπε κώδικα (A'.5)] παίρνουμε τα αποτελέσματα των SSA , SSB , $SSAB$ και SSE , και στη συνέχεια με χρήση των σχέσεων του θεωρήματος (17), λαμβάνονται οι τιμές των αμερόληπτων εκτιμητών των παραμέτρων δόμησης. Συγκεκριμένα έχουμε:

$$\begin{aligned}
 SSA &= 346036.3 \\
 SSB &= 766300.4 \\
 SSAB &= 112341 \\
 SSE &= 10578862
 \end{aligned}$$

και έτσι από θεώρημα (17) προκύπτουν

$$\begin{aligned}\widehat{s}^2 &= (164.2769)^2, \\ \widehat{\sigma}_\alpha^2 &= (32.95554)^2, \\ \widehat{\sigma}_\beta^2 &= (51.50535)^2, \\ \widehat{\sigma}_{\alpha\beta}^2 &= (5.593303)^2.\end{aligned}$$

Μοντέλο Σταυρωτής Ταξινόμησης: Πραγματοποιώντας τους υπολογισμούς, σύμφωνα με το μοντέλο σταυρωτής ταξινόμησης, κάνοντας χρήση των σχέσεων (6.133), (6.134), (6.135) και (6.136) [βλέπε Κώδικα (Α'.5)] προκύπτουν τα εξής αποτελέσματα:

$$\widehat{s}^2 = 26986.89 = (164.2769)^2,$$

$$\widehat{\sigma}_\alpha^2 = 26986.89 = (30.99753)^2,$$

$$\widehat{\sigma}_\beta^2 = 2645.731 = (51.43667)^2,$$

$$\widehat{\sigma}_{\alpha\beta}^2 = 34.24742 = (5.852129)^2.$$

Παρατήρηση 1.: Είναι φανερό, εξετάζοντας τα αποτελέσματα των $\widehat{\sigma}_\alpha^2$, $\widehat{\sigma}_\beta^2$, $\widehat{\sigma}_{\alpha\beta}^2$, πως περισσότερη αξιοπιστία δίνεται στην εμπειρία κινδύνου εντός των στηλών j , δηλαδή στα έτη συνεχούς εργασίας στον τελευταίο εργοδότη.

Επιπλέον, οι συντελεστές αξιοπιστίας, μέσω των σχέσεων, (6.8), (6.9) και (6.10) είναι:

$$\widehat{Z}_{ij} = \begin{pmatrix} 0.04830930 & 0.05174501 & 0.04945730 \\ 0.06413317 & 0.06302037 & 0.05741641 \\ 0.04715852 & 0.04715852 & 0.05288475 \end{pmatrix}$$

$$\widehat{Z}_i = \begin{pmatrix} 0.8074962 \\ 0.8381432 \\ 0.8050644 \end{pmatrix}$$

και

$$\widehat{Z}_j = (0.9249797 \quad 0.9259762 \quad 0.9250481).$$

Τέλος από τη λύση του συστήματος (6.11), και με χρήση των συντελεστών αξιοπιστίας, παίρνουμε τις εκτιμήσεις των ποσών που θα πληρωθούν την επόμενη περίοδο, $T + 1$. Αυτές είναι:

$$\widehat{X}_{ij,T+1}^{Cred} = \begin{pmatrix} 180.8747 & 238.7310 & 276.4352 \\ 172.4621 & 229.8255 & 267.7878 \\ 223.7261 & 280.6700 & 326.5767 \end{pmatrix}.$$

Κεφάλαιο 7

Επίλογος

Ο βασικός στόχος της παρούσας εργασίας ήταν η παρουσίαση της επιστημονικής εξέλιξης των μοντέλων αξιοπιστίας χαρτοφυλακίων, με τη βοήθεια της τεχνικής την ανάλυσης διασποράς. Στο πρώτο μέρος της εργασίας πραγματοποιήθηκε εισαγωγή σε βασικές έννοιες της θεωρίας αξιοπιστίας, ενώ παράλληλα αναλύθηκε και η μέθοδος της ανάλυσης διασποράς.

Στο δεύτερο Κεφάλαιο παρουσιάστηκε το μοντέλο αξιοπιστίας του Bühlmann (1967), και παρουσιάστηκε η δυνατότητα υπολογισμού των παραμέτρων δόμησης μέσω της ανάλυσης διασποράς κατά ένα παράγοντα. Επιπλέον μέσω της ανάλυσης διασποράς δόθηκε και ένας άμεσος τρόπος υπολογισμού του συντελεστή αξιοπιστίας Z , που δεν απαιτεί τον εκ των προτέρων υπολογισμό των παραμέτρων δόμησης του μοντέλου.

Η εξέλιξη του μοντέλου αυτού ήταν η δυνατότητα εισαγωγής βαρών στα δεδομένα, εάν κάποιες παρατηρήσεις θεωρηθούν πιο έμπιστες, καθώς και η διαφορά στο πλήθος των παρατηρήσεων ανά συμβόλαιο. Προκειται για το μοντέλο των Bühlmann-Straub (1970). Όπως και στο προηγούμενο μοντέλο, έτσι και σε αυτό αναλύθηκε η δυνατότητα συσχέτισης με τη θεωρία της ανάλυσης διασποράς.

Στο Κεφάλαιο 4 παρουσιάστηκε το μοντέλο του Hachemeister (1975), μια επέκταση του μοντέλου των Bühlmann-Straub (1970) μέσω παλινδρόμησης.

Στο Ιεραρχικό μοντέλο, το μοντέλο του Jewell (1975), που παρουσιάστηκε στο επόμενο μέρος, εξετάστηκε η περίπτωση συμβολαίων που ομαδοποιούνται σε τομείς με κοινά χαρακτηριστικά επιπλέον των δικών τους ατομικών χαρακτηριστικών και ιεραρχούνται. Και εδώ αποδείχθηκε η δυνατότητα υπολογισμού των παραμέτρων δόμησης μέσω της ανάλυσης διασποράς 2 παραγόντων με ιεράρχιση.

Μέσω διερεύνησης του προαναφερομένου μοντέλου παρουσιάστηκε και μελετήθηκε στο 6ο κεφάλαιο ένα ακόμα μοντέλο. Το μοντέλο αξιοπιστίας Σταυρωτής Ταξινόμησης παραγόντων, όπως αναλύθηκε, καλύπτει τις περιπτώσεις όπου οι κίνδυνοι δεν μπορούν να ιεραρχηθούν και παρουσιάζουν αλληλεπίδραση. Στο μοντέλο που παρουσιάστηκε, εξετάστηκε η περίπτωση δύο παραγόντων και μελετήθηκε η συσχέτιση του μοντέλου αξιοπιστίας με την τεχνική της ανάλυσης διασποράς για δύο παράγοντες με αλληλεπίδραση, και διαπιστώθηκε η δυνατότη-

τα υπολογισμού των παραμέτρων δόμησης και με αυτή. Συνολικά από την ανάπτυξη των μοντέλων μέσω της τεχνικής της ανάλυσης διασποράς μπορούμε να πούμε πως η μέθοδος αυτή προσφέρει τον άμεσο έλεγχο ομοιογένειας-ετερογένειας σε ένα χαρτοφυλάκιο, επιπλέον δίνει την δυνατότητα να γίνεται άμεσα και εύκολα αντιληπτό το επίπεδο του σφάλματος, και φυσικά να πραγματοποιηθούν οι υπολογισμοί των παραμέτρων δόμησης.

Βιβλιογραφία

- Chi H. L., Wing K.F. and Zhong Y.Z. (2007), Structural Parameter Estimation using Generalized Estimating Equations for Regression Credibility Models, A. S. *ASTIN Bulletin*. Number τ . 37. 323-341.
- Bühlmann, H. (1967). *Experience Rating and Credibility*. *ASTIN Bulletin*, 4, 199-207.
- Bühlmann, H. (1970). *Mathematical Models in Risk Theory*. Springer.
- Bühlmann, H. and Gisler, A. (2006). *A Course in Credibility Theory and its Applications*. Universitext. Springer Berlin Heidelberg.
- Bühlmann, H. and Straub, E. (1970). "*Glaubwürdigkeit für Schadensätze*". *Mitt. Ver. Schweiz. Ver.*, 70, 111-133.
- Dannenburg, D., Kaas, R., and Goovaerts, M. (1996). *Practical Actuarial Credibility Models, PhD Thesis*. Institute of Actuarial Science and Econometrics.
- Frees, E. W., Young, V. R., and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24(3): 229–247.
- Goovaerts, Kaas, M., Heerwaarden, R. V., and T., B. (1990). *Effective Actuarial Methods*. Elsevier Science Publishers B.V.
- Goulet, V. (1998). *Journal of Actuarial Practice, Principles and Application of Credibility Theory*. Number τ . 6. Absalom Press, 40-59.
- Guo, W. (2012). *Simple Linear Regression, Regression Analysis Method*.
- Hachemeister, C. A. (1975). *Credibility for Regression Models with Application to Trend, from Credibility: Theory and Applications*, edited by P.M. Kahn. Academic Press, Inc.
- Jewell, W. S. (1975). *The Use of Collateral Data in Credibility Theory: A Hierarchical Model*. IIASA Research Memorandum. IIASA, Laxenburg, Austria.
- Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern Actuarial Risk Theory*. Springer.

- Kahn, P. (1975). *Credibility: Theory and Applications : Proceedings of the Berkeley Actuarial Research Conference on Credibility, September 19-21, 1974, The University of California, Berkley [sic]*. Academic Press rapid manuscript reproductions. Academic Press.
- Pitselis, G. (2018). *Credibility Theory Notes*. University of Piraeus.
- Sahai, H. and Ojeda, M. (2007). *Analysis of Variance for Random Models, Volume 2: Unbalanced Data: Theory, Methods, Applications, and Data Analysis*. Birkhäuser Boston.
- Sahai, H. and Ojeda, M. (2013). *Analysis of Variance for Random Models, Volume 1: Balanced Data: Theory, Methods, Applications, and Data Analysis*. Birkhäuser Boston.
- Searle, S., Casella, G., and McCulloch, C. (2009). *Variance Components*. Wiley Series in Probability and Statistics. Wiley.
- Ward, C. (1997). *Robust Theory Applied to Jewell's Heirarchical Credibility Model, Master Thesis*. Concordia University Montreal.
- Whitney, A.W. (1918). *The theory of experience rating. Proceedings of the Casualty Actuarial Society 4, 274-292.*

Παράρτημα Α'

Κώδικας R

Α'.1 Κώδικας R για Buhlmann

Κώδικας ANOVA

```
$A=read.table("bproblem.txt",header=F,dec=".")
attach(A)
A
t=5
Xbari. <- matrix(0,3,1)
for (i in 1:3)
{
Xbari.[i] <- rowMeans(A[i,])
}
m <- mean(Xbari.);m
MSB <- sum(t*(Xbari.-m)^2)/(nrow(A)-1)
MSB

MSW <- 0
for(i in 1:3)
for(j in 1:5)
{
MSW <- MSW+(A[i,j]-Xbari.[i])^2/(nrow(A)*(ncol(A)-1))
}
MSW
F = MSW/MSB
F
```

Κώδικας Buhlmann

```
install.packages("actuar")
library(actuar)
bprob<-rbind(c(99.3, 93.7, 103.9, 92.5, 110.6),
c(112.5, 108.3, 118.0, 99.4, 111.8),c(129.2, 140.9, 108.3, 105.0, 116.6))
bprob<-cbind(c(1,2,3),bprob)
colnames(bprob)<-c("num","ratio.1","ratio.2","ratio.3","ratio.4","ratio.5")
bprob
fit <- cm(~num, bprob,
ratios = ratio.1:ratio.5)
```

```
fit predict(fit)
summary(fit)
```

A'.2 Κώδικας R για Buhlmann-Straub

Κώδικας ANOVA

```
J <- 9; K <- 10; X <- scan(n=J*K)
540 514 576 483 481 493 438 588 541 441
99 103 163 126 0 219 370 273 155 275
0 400 1042 313 0 833 0 0 0 0
275 278 430 196 667 185 517 204 323 968
543 984 727 562 722 610 794 299 580 488
0 0 0 645 833 0 0 769 0 0
333 404 400 361 588 349 435 476 635 556
494 133 735 519 1000 641 339 513 227 244
1667 313 556 769 1818 0 1429 0 0 0

j <- rep(1:J, each=K); j <- as.factor(j)

w <-c(44, 50, 56, 58, 58, 56, 54, 52, 52, 46, 20, 20, 24,
32, 28, 28, 28, 22, 26, 22, 8, 6, 10, 6, 8, 4, 6, 4, 4, 4,
22, 22, 18, 20, 12, 10, 12, 10, 6, 6, 26, 24, 22, 18, 20,
16, 12, 14, 14, 8, 6, 8, 6, 6, 2, 4, 2, 2, 2, 2, 18, 20, 20,
16, 18, 18, 14, 12, 12, 10, 16, 16, 14, 16, 14, 16, 12,
8, 8, 8, 6, 6, 4, 2, 4, 2, 4, 2, 4, 2)

anova(lm(X~j, weight=w))

w1 <- sum(w)

w2 <- sum(w[1:10])^2+sum(w[11:20])^2+sum(w[21:30])^2
+sum(w[31:40])^2+sum(w[41:50])^2+sum(w[51:60])^2
+sum(w[61:70])^2+sum(w[71:80])^2+sum(w[81:90])^2
w2

s2a <- (J-1)*(4693376-695107)/(w1-w2/w1); s2a; sqrt(s2a)
```

Κώδικας Buhlmann-Straub

```
bsprob=read.table("BS.txt", header=F, dec=",")
attach(bsprob)
bsprob
colnames(bsprob)<-c("num", "ratio.1", "ratio.2", "ratio.3", "ratio.4", "ratio.5",
"ratio.6", "ratio.7", "ratio.8", "ratio.9", "ratio.10", "weight.1", "weight.2",
"weight.3", "weight.4", "weight.5", "weight.6", "weight.7", "weight.8", "weight.9", "weight.10")
bsprob
fit <- cm(~num, bsprob,
ratios = ratio.1:ratio.10, weights = weight.1:weight.10)
fit
predict(fit)
summary(fit)
```

A.3 Κώδικας R για Hachemeister

Κώδικας Hachemeister

```
fit <- cm(~state, hachemeister,
regformula = ~time, regdata = data.frame(time = 12:1),
ratios = ratio.1:ratio.12, weights = weight.1:weight.12)
fit
predict(fit, newdata = data.frame(time = 0))
summary(fit, newdata = data.frame(time = 0))
```

A.4 Κώδικας R για Hierarchical

Κώδικας ANOVA

```
P<-2
J<-5
T<-12
T.<-J*T
X<-rbind(c(1738, 1364, 1759, 1223, 1456),
c(1642, 1408, 1685, 1146, 1499),
c(1794, 1597, 1479, 1010, 1609),
c(2051, 1444, 1763, 1257, 1741),
c(2079, 1342, 1674, 1426, 1482),
c(2234, 1675, 2103, 1532, 1572),
c(2032, 1470, 1502, 1953, 1606),
c(2035, 1448, 1622, 1123, 1735),
c(2115, 1464, 1828, 1343, 1607),
c(2262, 1831, 2155, 1243, 1573),
c(2267, 1612, 2233, 1762, 1613),
c(2517, 1471, 2059, 1306, 1690))
X11t<-X[,1]
X12t<-X[,3]
X1jt<-cbind(X11t, X12t)
X21t<-X[,2]
X22t<-X[,4]
X23t<-X[,5]
X2jt<-cbind(X21t, X22t, X23t)
X1jt;X2jt
Xpjt<-cbind(X1jt, X2jt)

Xbarp..<-matrix(0,1,2)
Xbarpj.<-matrix(0,1,5)

for(i in 1:J){
Xbarpj.[i]<-sum(Xpjt[,i])/T
}
colnames(Xbarpj.)<-c("Xbar11.", "Xbar13.", "Xbar22.",
"Xbar24.", "Xbar25.")
Xbarpj.

Xbarp..[1]<-(Xbarpj.[1]+Xbarpj.[2])/2
Xbarp..[2]<-(Xbarpj.[3]+Xbarpj.[4]+Xbarpj.[5])/3
colnames(Xbarp..) <-c("Xbar1..", "Xbar2..")
```



```

Xbarp..
Xbar...<-sum(Xpjt)/(J*T)
Xbar...

tp.<-matrix(0,P)
tp.[1]<-2*nrow(Xpjt)
tp.[2]<-3*nrow(Xpjt)

tpj<-matrix(12,5)

k1<-sum(tp.[]^2)/T.;k1
k2<-sum(tpj[]^2)/T.;k2
k12<-(tpj[1]^2+tpj[2]^2)/tp.[1]+
(tpj[3]^2+tpj[4]^2+tpj[5]^2)/tp.[2];k12

r1<-(T.-k12)/(J-P);r1
r2<-(k12-k2)/(P-1);r2
r3<-(T.-k1)/(P-1);r3

SSE<-0
for(i in 1:5){
for(j in 1:12){
SSE<-SSE+(Xpjt[j,i]-Xbarpj.[i])^2
}
}
SSE
MSE<-SSE/(T.-J);MSE

s2<-MSE

SSP<-(tp.[1]*(Xbarp..[1]-Xbar...)^2)+
tp.[2]*(Xbarp..[2]-Xbar...)^2)
MSP<-SSP/(P-1);MSP

SSJ<-T*(
(Xbarpj.[1]-Xbarp..[1])^2)+
(Xbarpj.[2]-Xbarp..[1])^2)+
(Xbarpj.[3]-Xbarp..[2])^2)+
(Xbarpj.[4]-Xbarp..[2])^2)+
(Xbarpj.[5]-Xbarp..[2])^2)
)
SSJ
MSJ<-SSJ/(J-P);MSJ

s2
S2a<-(MSP-r2/r1*(SSJ/(J-P)-SSE/(T.-J))-MSE)/r3;S2a
S2b<-(SSJ/(J-P)-SSE/(T.-J))/r1;S2b

```

Κώδικας Hierarchical

```

X <- data.frame(unit = c("A", "B", "A", "B", "B"), hachemeister)
Jewell_iter <- cm(~unit + unit:state, X, ratio.1:ratio.12,method="iterative")
summary(Jewell_iter)

```

Α'.5 Κώδικας R για Σταυρωτή Ταξινόμηση 2 Πα- ραγόντων (Cross Classification)

Κώδικας ANOVA

```
C=read.table("2way.dplm.txt", header=T)
attach(C)
C
T<-nrow(C);T
I<-3
J<-3
K<-I*J;K

Xij.<-matrix(0,3,3)

t11<-0
t12<-0
t13<-0
t21<-0
t22<-0
t23<-0
t31<-0
t32<-0
t33<-0

for(i in 1:401){
if (C[i,1]==1 & C[i,2]==1 )
t11<-t11+1
else if(C[i,1]==1 & C[i,2]==2 )
t12<-t12+1
else if(C[i,1]==1 & C[i,2]==3 )
t13<-t13+1
else if(C[i,1]==2 & C[i,2]==1 )
t21<-t21+1
else if(C[i,1]==2 & C[i,2]==2 )
t22<-t22+1
else if(C[i,1]==2 & C[i,2]==3 )
t23<-t23+1
else if(C[i,1]==3 & C[i,2]==1 )
t31<-t31+1
else if(C[i,1]==3 & C[i,2]==2 )
t32<-t32+1
else if(C[i,1]==3 & C[i,2]==3 )
t33<-t33+1
}

t11;t12;t13;t21;t22;t23;t31;t32;t33
tij<-matrix(c(t11,t21,t31,t12,t22,t32,t13,t23,t33),
nrow=3,ncol=3)
tij

for(i in 1:401){

if (C[i,1]==1 & C[i,2]==1)
Xij.[1,1]<-Xij.[1,1]+C[i,3]

else if(C[i,1]==1 & C[i,2]==2 )
```

```

Xij.[1,2]<-Xij.[1,2]+C[i,3]

else if(C[i,1]==1 & C[i,2]==3 )
Xij.[1,3]<-Xij.[1,3]+C[i,3]

else if(C[i,1]==2 & C[i,2]==1 )
Xij.[2,1]<-Xij.[2,1]+C[i,3]

else if(C[i,1]==2 & C[i,2]==2 )
Xij.[2,2]<-Xij.[2,2]+C[i,3]

else if(C[i,1]==2 & C[i,2]==3 )
Xij.[2,3]<-Xij.[2,3]+C[i,3]

else if(C[i,1]==3 & C[i,2]==1 )
Xij.[3,1]<-Xij.[3,1]+C[i,3]

else if(C[i,1]==3 & C[i,2]==2 )
Xij.[3,2]<-Xij.[3,2]+C[i,3]

else if(C[i,1]==3 & C[i,2]==3 )
Xij.[3,3]<-Xij.[3,3]+C[i,3]
}
Xij.

Xijw<-matrix(0,3,3)

for (i in 1:3){
for (j in 1:3){
Xijw[i,j]<-Xij.[i,j]/tij[i,j]
}
}
Xijw

sum11<-0
sum21<-0
sum31<-0
sum12<-0
sum22<-0
sum32<-0
sum13<-0
sum23<-0
sum33<-0

for ( i in 1:401) {

if (C[i,1]==1 & C[i,2]==1)
sum11<-sum11+(C[i,3] - Xijw[1,1])^2

if (C[i,1]==2 & C[i,2]==1)
sum21<-sum21+(C[i,3] - Xijw[2,1])^2

if (C[i,1]==3 & C[i,2]==1)
sum31<-sum31+(C[i,3] - Xijw[3,1])^2

if (C[i,1]==1 & C[i,2]==2)
sum12<-sum12+(C[i,3] - Xijw[1,2])^2

```

```

if (C[i,1]==2 & C[i,2]==2)
sum22<-sum22+(C[i,3] - Xijw[2,2])^2

if (C[i,1]==3 & C[i,2]==2)
sum32<-sum32+(C[i,3] - Xijw[3,2])^2

if (C[i,1]==1 & C[i,2]==3)
sum13<-sum13+(C[i,3] - Xijw[1,3])^2

if (C[i,1]==2 & C[i,2]==3)
sum23<-sum23+(C[i,3] - Xijw[2,3])^2

if (C[i,1]==3 & C[i,2]==3)
sum33<-sum33+(C[i,3] - Xijw[3,3])^2

}

sum11;sum21;sum31;sum12;sum22;sum32;sum13;sum23;sum33

SSE<-sum (sum11,sum21,sum31,sum12,sum22,sum32,sum13,
sum23,sum33);SSE

s2<-SSE/(nrow(C)-I*J)

sqrt(s2)

ti.<-matrix(0,3,1)

for(i in 1:3){
ti.[i]<-sum(tij[i,])
}
ti.

Xi..<-matrix(0,3,1)

for(i in 1:401){
if (C[i,1]==1)
Xi..[1]<-Xi..[1]+C[i,3]
else if (C[i,1]==2)
Xi..[2]<-Xi..[2]+C[i,3]
else if (C[i,1]==3)
Xi..[3]<-Xi..[3]+C[i,3]
}

t.j<-matrix(0,1,3)
t.j
for(j in 1:3){
t.j[j]<-sum(tij[,j])
}
t.j

X.j.<-matrix(0,1,3)

for(i in 1:401){
if (C[i,2]==1)
X.j.[1]<-X.j.[1]+C[i,3]
else if (C[i,2]==2)
X.j.[2]<-X.j.[2]+C[i,3]
}

```

```

else if (C[i,2]==3)
X.j.[3]<-X.j.[3]+C[i,3]
}

SSA<-(Xi..[1]^2)/ti.[1]+Xi..[2]^2)/ti.[2]+
Xi..[3]^2)/ti.[3] )-sum(Amounts)^2)/T
SSA

SSB<-(X.j.[1]^2)/t.j[1]+X.j.[2]^2)/t.j[2]+
X.j.[3]^2)/t.j[3] )-sum(Amounts)^2)/T
SSB

SSAB<-(Xij.[1,1]^2)/tij[1,1]+Xij.[1,2]^2)/tij[1,2]+
Xij.[1,3]^2)/tij[1,3]+Xij.[2,1]^2)/tij[2,1]+
Xij.[2,2]^2)/tij[2,2]+Xij.[2,3]^2)/tij[2,3]+
Xij.[3,1]^2)/tij[3,1]+Xij.[3,2]^2)/tij[3,2]+
Xij.[3,3]^2)/tij[3,3])-(Xi..[1]^2)/ti.[1]+
Xi..[2]^2)/ti.[2]+Xi..[3]^2)/ti.[3]) -
(X.j.[1]^2)/t.j[1]+X.j.[2]^2)/t.j[2]+
X.j.[3]^2)/t.j[3]) +sum(Amounts)^2)/T
SSAB

r1<-(1/T)*(ti.[1]^2+ti.[2]^2+ti.[3]^2);r1
r2<-(1/T)*(t.j[1]^2+t.j[2]^2+t.j[3]^2);r2

r12<-(1/T)*(tij[1,1]^2+tij[1,2]^2+tij[1,3]^2+
tij[2,1]^2+tij[2,2]^2+tij[2,3]^2+
tij[3,1]^2+tij[3,2]^2+tij[3,3]^2);r12

r3<-((tij[1,1]^2+tij[1,2]^2+tij[1,3]^2)/ti.[1])+
((tij[2,1]^2+tij[2,2]^2+tij[2,3]^2)/ti.[2])+
((tij[3,1]^2+tij[3,2]^2+tij[3,3]^2)/ti.[3]));r3

r4<-((tij[1,1]^2+tij[2,1]^2+tij[3,1]^2)/t.j[1])+
((tij[1,2]^2+tij[2,2]^2+tij[3,2]^2)/t.j[2])+
((tij[1,3]^2+tij[2,3]^2+tij[3,3]^2)/t.j[3]));r4

s2ab<-(((T-r1)/(T-r4))*(SSAB+SSA-(K-J)*s2)+
((r3-r2)/(T-r3))*(SSAB+SSB-(K-I)*s2)-
(SSA-(I-1)*s2))/(T-r1-r2+r12);s2ab
sqrt(s2ab)

s2a<-(1/(T-r4))*(SSAB+SSA-(K-J)*s2)-s2ab;s2a
sqrt(s2a)

s2b<-(1/(T-r3))*(SSAB+SSB-(K-I)*s2)-s2ab;s2b
sqrt(s2b)

```

Κώδικας Σταυρωτής Ταξινόμησης 2 Παραγόντων

```

C=read.table("2way.dplm.txt", header=T)
attach(C)
C
I<-3
J<-3

```

```

T<-nrow(dataTwoWay);T
Xijw<-matrix(0,3,3)
rownames(Xijw)<-c("single(i=1)", "divorced(i=2)",
"otherwise(i=3)")
colnames(Xijw)<-c("num.of.years<2", "num.of.years 2-10", "num.of.years>=10")

M<-sum(Amounts)/length(Amounts);M

t11<-0
t12<-0
t13<-0
t21<-0
t22<-0
t23<-0
t31<-0
t32<-0
t33<-0

for(i in 1:401){
if (C[i,1]==1 & C[i,2]==1 )
t11<-t11+1
else if (C[i,1]==1 & C[i,2]==2 )
t12<-t12+1
else if (C[i,1]==1 & C[i,2]==3 )
t13<-t13+1
else if (C[i,1]==2 & C[i,2]==1 )
t21<-t21+1
else if (C[i,1]==2 & C[i,2]==2 )
t22<-t22+1
else if (C[i,1]==2 & C[i,2]==3 )
t23<-t23+1
else if (C[i,1]==3 & C[i,2]==1 )
t31<-t31+1
else if (C[i,1]==3 & C[i,2]==2 )
t32<-t32+1
else if (C[i,1]==3 & C[i,2]==3 )
t33<-t33+1
}
t11;t12;t13;t21;t22;t23;t31;t32;t33
tij<-matrix(c(t11,t21,t31,t12,t22,t32,t13,t23,t33),
nrow=3,ncol=3)
tij

for(i in 1:401){

if (C[i,1]==1 & C[i,2]==1)
Xijw[1,1]<-Xijw[1,1]+C[i,3]

else if (C[i,1]==1 & C[i,2]==2 )
Xijw[1,2]<-Xijw[1,2]+C[i,3]

else if (C[i,1]==1 & C[i,2]==3 )
Xijw[1,3]<-Xijw[1,3]+C[i,3]

else if (C[i,1]==2 & C[i,2]==1 )
Xijw[2,1]<-Xijw[2,1]+C[i,3]

else if (C[i,1]==2 & C[i,2]==2 )
Xijw[2,2]<-Xijw[2,2]+C[i,3]

```

```

else if(C[i,1]==2 & C[i,2]==3 )
Xijw[2,3]<-Xijw[2,3]+C[i,3]

else if(C[i,1]==3 & C[i,2]==1 )
Xijw[3,1]<-Xijw[3,1]+C[i,3]

else if(C[i,1]==3 & C[i,2]==2 )
Xijw[3,2]<-Xijw[3,2]+C[i,3]

else if(C[i,1]==3 & C[i,2]==3 )
Xijw[3,3]<-Xijw[3,3]+C[i,3]
}
Xijw

for (i in 1:3){
for (j in 1:3){
Xijw[i,j]<-Xijw[i,j]/tij[i,j]
}
}
Xijw

sum11<-0
sum21<-0
sum31<-0
sum12<-0
sum22<-0
sum32<-0
sum13<-0
sum23<-0
sum33<-0

for ( i in 1:401) {

if (C[i,1]==1 & C[i,2]==1)
sum11<-sum11+(C[i,3] - Xijw[1,1])^2

if (C[i,1]==2 & C[i,2]==1)
sum21<-sum21+(C[i,3] - Xijw[2,1])^2

if (C[i,1]==3 & C[i,2]==1)
sum31<-sum31+(C[i,3] - Xijw[3,1])^2

if (C[i,1]==1 & C[i,2]==2)
sum12<-sum12+(C[i,3] - Xijw[1,2])^2

if (C[i,1]==2 & C[i,2]==2)
sum22<-sum22+(C[i,3] - Xijw[2,2])^2

if (C[i,1]==3 & C[i,2]==2)
sum32<-sum32+(C[i,3] - Xijw[3,2])^2

if (C[i,1]==1 & C[i,2]==3)
sum13<-sum13+(C[i,3] - Xijw[1,3])^2

if (C[i,1]==2 & C[i,2]==3)
sum23<-sum23+(C[i,3] - Xijw[2,3])^2

```

```

if (C[i,1]==3 & C[i,2]==3)
sum33<-sum33+(C[i,3] - Xijw[3,3])^2

}

sum11;sum21;sum31;sum12;sum22;sum32;sum13;sum23;sum33

s2<-sum (sum11, sum21, sum31, sum12, sum22, sum32, sum13,
sum23, sum33)/(nrow(C)-I*J)
s2
sqrt(s2)

ti.<-matrix(0,3,1)

for(i in 1:3){
ti.[i]<-sum(tij[i,])
}
ti.

Xiww<-matrix(0,3,1)

for(i in 1:401){
if (C[i,1]==1)
Xiww[1]<-Xiww[1]+C[i,3]
else if (C[i,1]==2)
Xiww[2]<-Xiww[2]+C[i,3]
else if (C[i,1]==3)
Xiww[3]<-Xiww[3]+C[i,3]
}
Xiww<-Xiww/ti.[]
Xiww

t.j<-matrix(0,1,3)
t.j
for(j in 1:3){
t.j[j]<-sum(tij[,j])
}
t.j

Xwjw<-matrix(0,1,3)
Xwjw

for(i in 1:401){
if (C[i,2]==1)
Xwjw[1]<-Xwjw[1]+C[i,3]
else if (C[i,2]==2)
Xwjw[2]<-Xwjw[2]+C[i,3]
else if (C[i,2]==3)
Xwjw[3]<-Xwjw[3]+C[i,3]
}
Xwjw<-Xwjw/t.j[]
Xwjw

X1<-0
for(i in 1:3){
for(j in 1:3){
X1<-X1+(tij[i,j]/ti.[i])*(Xijw[i,j]-Xiww[i])^2
}
}

```



```

}
Xx1<-0
for(i in 1:3){
Xx1<-Xx1+s2*((J-1)/ti.[i])
}
X1<-(1/I)*(X1-Xx1)

X2<-0
for(j in 1:3){
for(i in 1:3){
X2<-X2+(tij[i,j]/t.j[j])*(Xijw[i,j]-Xwjw[j])^2
}
}
Xx2<-0
for(j in 1:3){
Xx2<-Xx2+s2*((I-1)/t.j[j])
}
X2<-(1/J)*(X2-Xx2)

X3<-0
for(i in 1:3){
for(j in 1:3){
X3<-X3+(tij[i,j]/T)*(Xijw[i,j]-M)^2
}
}

X3<-X3-s2*((I*J-1)/T)

X1;X2;X3

Y1<-0
Y2<-0
Y3<-0
Y4<-0
Y5<-0

for(i in 1:3){
for(j in 1:3){
Y1<-Y1+(tij[i,j]/ti.[i])^2
}
}
Y1<-1-(1/I)*Y1

for(j in 1:3){
for(i in 1:3){
Y2<-Y2+(tij[i,j]/t.j[j])^2
}
}
Y2<-1-(1/J)*Y2

for(i in 1:3){
Y3<-Y3+(ti.[i]/T)^2
}
Y3<-1-Y3

for(j in 1:3){
Y4<-Y4+(t.j[j]/T)^2
}

```

```

Y4<-1-Y4

for(i in 1:3){
for(j in 1:3){
Y5<-Y5+(tij[i,j]/T)^2
}
}
Y5<-1-Y5

Y1;Y2;Y3;Y4;Y5

A<-matrix(c(0, Y1, Y1,
            Y2, 0, Y2,
            Y3, Y4, Y5),3 ,3, byrow=TRUE)
b<-c(X1,X2,X3)
solve(A,b)
s2_<-matrix(0,1,3)
s2_<-solve(A,b)
s2a<-s2_[1]
s2b<-s2_[2]
s2ab<-s2_[3]
sqrt(s2a);sqrt(s2b);sqrt(s2ab)

Zij<-matrix(0,3,3)
for(i in 1:3){
for(j in 1:3){
Zij[i,j]<-s2ab/(s2ab+s2/tij[i,j])
}
}
Zij

Zi.<-matrix(0,3,1)
for(i in 1:3){
for(j in 1:3){
Zi.[i]<-Zi.[i]+Zij[i,j]
}
}
Zi.

Z.j<-matrix(0,1,3)
for(j in 1:3){
for(i in 1:3){
Z.j[j]<-Z.j[j]+Zij[i,j]
}
}
Z.j

Zi<-matrix(0,3,1)
for(i in 1:3){
Zi[i]<-s2a/(s2a+s2ab/Zi.[i])
}
Zi

Zj<-matrix(0,1,3)
for(j in 1:3){
Zj[j]<-s2b/(s2b+s2ab/Z.j[j])
}
Zj

```

```
Xizw<-matrix(0,3,1)
for (i in 1:3){
for (j in 1:3){
Xizw[i]<-Xizw[i]+(Zij[i, j]/Zi.[i])*Xijw[i, j]
}
}
Xizw
```

```
Xzjw<-matrix(0,1,3)
for(j in 1:3){
for(i in 1:3){
Xzjw[j]<-Xzjw[j]+(Zij[i, j]/Z.j[j])*Xijw[i, j]
}
}
Xzjw
```