

# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

## ΑΝΑΠΤΥΞΗ ΥΠΟΔΕΙΓΜΑΤΟΣ ΜΕΤΡΗΣΗΣ ΤΟΥ ΠΙΣΤΩΤΙΚΟΥ ΚΙΝΔΥΝΟΥ

Παναγιώτης – Νικόλαος Σ. Ροδουσάκης

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

Πειραιάς

Αύγουστος 2009

# РАНЕЕЗНАМО ТЕРПАА

# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

## ΑΝΑΠΤΥΞΗ ΥΠΟΔΕΙΓΜΑΤΟΣ ΜΕΤΡΗΣΗΣ ΤΟΥ ΠΙΣΤΩΤΙΚΟΥ ΚΙΝΔΥΝΟΥ

Παναγιώτης – Νικόλαος Σ. Ροδουσάκης

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

Πειραιάς

Αύγουστος 2009

# РАНЕЕЗНАМО ТЕРПАА

UNIVERSITY OF PIRAEUS



DEPARTMENT OF  
STATISTICS AND INSURANCE SCIENCE

POSTGRADUATE STUDY IN  
ACTUARIAL SCIENCE AND RISK MANAGEMENT

**DEVELOPMENT OF CREDIT RISK  
ESTIMATION MODEL**

Panagiotis – Nikolaos S. Rodousakis

**A dissertation submitted to the Department of  
Statistics and Insurance Science in accordance  
of the requirements of the degree of Msc in  
Actuarial Science and Risk Management.**

Piraeus

August 2009

# РАНЕЕЗНАМО ПЕРПАА

# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΔΑΛΗ

*Στην οικογένειά μου,*

# ТАНЕЦЪМО ТЕРПА



## Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Μιχάλη Γκλεζάκο για την πολύτιμη βοήθειά και καθοδήγησή του καθώς και τα μέλη της τριμελούς επιτροπής κα. Αικατερίνη Πανοπούλου και κ. Σπυρίδων Βρόντο. Επίσης, θα ήθελα να ευχαριστήσω τον κ. Παναγιώτη Καπόπουλο για τις πολύτιμες συμβουλές του.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και την Ασπασία Φασουλά για την συμπαράστασή τους κατά την εκπόνηση της πτυχιακής εργασίας.



## ΠΕΡΙΛΗΨΗ

Αντικείμενο της πτυχιακής εργασίας είναι η ανάπτυξη ενός υποδείγματος εκτίμησης του πιστωτικού κινδύνου.

Αφού εξετάστηκε ο πιστωτικός κίνδυνος και οι βασικές παράμετροί του σε θεωρητική βάση, καθώς στα πλαίσια του Συμφώνου της Βασιλείας II, αξιολογήθηκαν τρία βασικά υποδείγματα εκτίμησης της πιθανότητας αθέτησης (PD): Multivariate Discriminant Analysis, Logit και Probit.

Στη συνέχεια, αναπτύχθηκε ένα υπόδειγμα Logit, για την εκτίμηση της πιθανότητας αθέτησης σε χαρτοφυλάκιο καταναλωτικών δανείων ελληνικής τράπεζας.

Το υπόδειγμα αυτό έχει ικανοποιητική προβλεπτική ικανότητα, δεδομένου ότι κατατάσσει σωστά μέχρι και το 70% των πελατών του δείγματος. Όμως, με αυτά τα ποσοστά επιτυχίας δεν μπορεί να χαρακτηριστεί αποτελεσματικό. Ενδεχόμενα το υπόδειγμα αυτό επηρεάστηκε από την αυστηρή ταξινόμηση των πελατών της τράπεζας του δείγματος σε συνεπείς και ασυνεπείς.

# ТАНЕЦЫ И ТЕАТР

## **ABSTRACT**

The subject of this study is to develop a model which enables the estimation of credit risk.

At the beginning, the concept of credit risk, as well as the relevant subjects of Basel II, were presented and discussed. Next, three different methodologies aiming to measure credit risk were evaluated.

Finally, a Logit type model was developed, based on the relevant data of a Greek bank. This model exhibited a satisfactory predictive ability, classifying correctly up to 70% of the customers in the sample. However, higher percentages are required to be characterized as efficient. A possible explanation of its limited ability, in the above sense, may be due to the strict classification of the clients in the sample, from the part of the bank which provided the needed data.

# ТАНЕЦЫ И ТЕАТР

## Περιεχόμενα

Εισαγωγή/Στόχος της εργασίας.....	1
ΚΕΦΑΛΑΙΟ 1: Ο ΠΙΣΤΩΤΙΚΟΣ ΚΙΝΔΥΝΟΣ : ΕΝΝΟΙΑ, ΒΑΣΙΚΕΣ ΠΑΡΑΜΕΤΡΟΙ, ΒΑΣΙΛΕΙΑ Π.....	3
1.1 Ο Πιστωτικός Κίνδυνος από τη σκοπιά των τραπεζικών επιχειρήσεων.....	3
1.2 Ταξινόμηση των τραπεζικών απαιτήσεων .....	3
1.3 Παράμετροι του πιστωτικού κινδύνου .....	5
1.4 Η μέτρηση του πιστωτικού κινδύνου στα πλαίσια του Συμφώνου της Βασιλείας Π .....	8
1.4 Τυποποιημένη προσέγγιση.....	10
1.6 Η Προσέγγιση των εσωτερικών διαβαθμίσεων (IRB) .....	13
1.7 Προσέγγιση εσωτερικών διαβαθμίσεων για χαρτοφυλάκιο λιανικής τραπεζικής σύμφωνα με τη Βασιλεία Π.....	13
1.8 Απαιτήσεις Βασιλείας Π για την εκτίμηση της πιθανότητας αθέτησης.....	20
ΚΕΦΑΛΑΙΟ 2: ΥΠΟΔΕΙΓΜΑΤΑ ΕΚΤΙΜΗΣΗΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΑΘΕΤΗΣΗΣ.....	23
2.1 Multivariate discriminate analysis (MDA) .....	23
2.2 Υπόδειγμα Λογιστικής Παλινδρόμησης-Logit .....	41
2.3 Υπόδειγμα Probit.....	51
ΚΕΦΑΛΑΙΟ 3: ΘΕΩΡΗΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΣΤΗΝ ΚΑΤΑΣΚΕΥΗ ΥΠΟΔΕΙΓΜΑΤΩΝ ΕΚΤΙΜΗΣΗΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΑΘΕΤΗΣΗΣ.....	53
3.1 Επιλογή υποδείγματος.....	53
3.2 Αρχική επιλογή επεξηγηματικών μεταβλητών .....	54
3.3 Επιλογή και μορφοποίηση των δεδομένων .....	57
3.3.1 Αντιπροσωπευτικότητα των δεδομένων .....	57
3.3.2 Πληρότητα των δεδομένων .....	57
3.3.3 Μέθοδοι αντιμετώπισης των ελλειπόν δεδομένων.....	59
3.3.4 Κωδικοποίηση Ερμηνευτικών μεταβλητών .....	61
3.3.5 Διαχωρισμός του δείγματος σε δείγμα εκτίμησης και δείγμα επικύρωσης .....	62
3.4 Θεωρητική προσέγγιση στην ανάπτυξη υποδείγματος Logit .....	62
3.4.1 Κατασκευή της συνάρτησης διάκρισης .....	62
3.4.2 Στατιστική επικύρωση του υποδείγματος .....	63
3.4.3 Καθορισμός του σημείου διαχωρισμού (cut off point) .....	65
3.4.4 Αποτελεσματικότητα (προβλεπτική ικανότητα) του υποδείγματος.....	66
ΚΕΦΑΛΑΙΟ 4: ΑΝΑΠΤΥΞΗ ΥΠΟΔΕΙΓΜΑΤΟΣ ΕΚΤΙΜΗΣΗΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΑΘΕΤΗΣΗΣ, ΓΙΑ ΜΙΑ ΕΛΛΗΝΙΚΗ ΤΡΑΠΕΖΑ .....	67

4.1 Το Δείγμα .....	67
4.2 Αρχική επιλογή των ερμηνευτικών μεταβλητών .....	69
4.3 Ανάλυση και κωδικοποίηση των μεταβλητών .....	70
4.4 Εφαρμογή του υποδείγματος Λογιστικής Παλινδρόμησης.....	83
4.4.1 Κατασκευή της συνάρτησης διαχωρισμού.....	84
4.4.2 Στατιστική σημαντικότητα του υποδείγματος .....	85
4.4.3 Σχολιασμός των συντελεστών.....	86
4.4.4 Έλεγχος της αποτελεσματικότητας του υποδείγματος.....	89
4.4.5 Συμπληρωματικές παρατηρήσεις .....	96
ΚΕΦΑΛΑΙΟ 5: ΑΝΑΚΕΦΑΛΑΙΩΣΗ – ΣΥΜΠΕΡΑΣΜΑΤΑ.....	99
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	105
ΠΑΡΑΡΤΗΜΑΤΑ.....	109
ΠΑΡΑΡΤΗΜΑ 1 : Υπολογισμός των πινάκων E και H .....	109
ΠΑΡΑΡΤΗΜΑ 2 : Πίνακας συσχετίσεων των ερμηνευτικών μεταβλητών .....	110
ΠΑΡΑΡΤΗΜΑ 3 : Equation output .....	111
ΠΑΡΑΡΤΗΜΑ 4 : Correlogram of Standardized Squared Residuals.....	112
ΠΑΡΑΡΤΗΜΑ 5 : Hosmer-Lemeshow test.....	113
ΠΑΡΑΡΤΗΜΑ 6 : Σφάλμα τύπου I και II για Cut off point = 15,04 %.....	114



## Εισαγωγή/Στόχος της εργασίας

Το μεγαλύτερο μέρος της παγκόσμιας οικονομικής δραστηριότητας οφείλεται στον ετεροχρονισμό μεταξύ του χρόνου πραγματοποίησης των συναλλαγών και του χρόνου εκκαθάρισής τους. Ο ετεροχρονισμός αυτός αποτελεί την πιστωτική λειτουργία των αγορών (= παροχή πιστώσεων) και εφαρμόζεται τόσο στην πραγματική οικονομία (καταβολή του τιμήματος των αγαθών με χρονική υστέρηση) όσο και στις κεφαλαιαγορές/χρηματαγορές (απόκτηση χρηματιστηριακών τίτλων, δανεισμός κλπ).

Όπως είναι φυσικό, η παροχή πιστώσεων εμπεριέχει τον κίνδυνο μη ανταπόκρισης του οφειλέτη («πιστωτικός κίνδυνος»), είτε λόγω αδυναμίας είτε λόγω πρόθεσης αθέτησης. Η αποτίμηση, επομένως, του πιστωτικού κινδύνου είναι στο επίκεντρο της προσοχής κάθε επιχείρησης η οποία παρέχει πιστώσεις και ιδιαίτερα των τραπεζών, που έχουν ως κύρια δραστηριότητα την χορήγηση πιστώσεων.

Γι' αυτό, οι σχετικές προσπάθειες ξεκίνησαν κυρίως από αυτές, πριν αρκετά χρόνια, και συνεχίζονται με αμείωτη ένταση. Ειδικότερα, έχει ήδη διαμορφωθεί μια πλούσια βιβλιογραφία τόσο με θεωρητικές όσο και με εμπειρικές προσεγγίσεις στο θέμα αυτό, η οποία εξετάζει ένα πλήθος προβλημάτων που συνδέονται με τον πιστωτικό κίνδυνο, όπως πχ η ανάπτυξη εναλλακτικών μεθοδολογιών μέτρησής του, ο εντοπισμός των βασικών παραμέτρων που τον συνθέτουν, η προσαρμογή διαφόρων θεωρητικών υποδειγμάτων για την εφαρμογή τους σε συγκεκριμένες επιχειρήσεις ή κατηγορίες χρεωστών κλπ. Σε κάθε περίπτωση, αυτό που επιδιώκεται από τους ερευνητές είναι η ενίσχυση της ικανότητας των επιχειρήσεων που παρέχουν πιστώσεις να αποτιμούν τον πιστωτικό κίνδυνο πρό της παροχής της πίστωσης, ώστε να περιορίζονται οι αντίστοιχες απώλειες.

Κατά τα τελευταία 20 χρόνια, η Ευρωπαϊκή Ένωση, σε μια προσπάθεια διασφάλισης της βιωσιμότητας του τραπεζικού συστήματος (το οποίο αποτελεί καταλυτικό παράγοντα της οικονομικής δραστηριότητας και ανάπτυξης), έχει οδηγήσει στη διαμόρφωση συστήματος αρχών και κανόνων, η εφαρμογή των οποίων αποσκοπεί στον περιορισμό του πιστωτικού κινδύνου («Βασιλεία II»). Σύμφωνα με την Βασιλεία II, οι τράπεζες μπορούν να αναπτύξουν τα δικά τους συστήματα

αποτίμησης των πιθανών απωλειών από επισφαλείς απαιτήσεις. Το γεγονός αυτό έχει δώσει νέα ώθηση στις σχετικές ερευνητικές προσπάθειες, με έμφαση στη διαμόρφωση υποδειγμάτων που επιτρέπουν την πρόγνωση απωλειών σε επίπεδο τράπεζας.

Στα πλαίσια αυτών των προσπαθειών εντάσσεται και η παρούσα εργασία, η οποία αποσκοπεί αφενός μεν στην παρουσίαση του θεωρητικού πλαισίου αποτίμησης του πιστωτικού κινδύνου, αφετέρου δε στην ανάπτυξη υποδείγματος μέτρησής του με βάση τα δεδομένα συγκεκριμένης ελληνικής τράπεζας και σε συνάρτηση με τους περιορισμούς της Βασιλείας II.

Η εργασία διαρθρώνεται ως εξής :

Στο Πρώτο Κεφάλαιο, ορίζεται ο πιστωτικός κίνδυνος και επισημαίνονται η σημασία του και οι παράμετροι που τον συνθέτουν. Επίσης, παρουσιάζεται το πλαίσιο αρχών και κανόνων της Βασιλείας II, το οποίο οριοθετεί αρκετά περιοριστικά τον πιστωτικό κίνδυνο.

Στο Κεφάλαιο 2 γίνεται αναφορά στα υποδείγματα εκτίμησης της πιθανότητας αθέτησης (PD), με αναλυτική αναφορά στα κυριότερα από αυτά, ήτοι στο υπόδειγμα Multivariate Discriminant Analysis (MDA) και στα υποδείγματα Λογιστικής Παλινδρόμησης-Logit και Probit.

Η θεωρητική προσέγγιση στην κατασκευή υποδειγμάτων προσδιορισμού της πιθανότητας αθέτησης, αποτελεί το αντικείμενο του Κεφαλαίου 3.

Η ανάπτυξη συγκεκριμένου υποδείγματος εκτίμησης της πιθανότητας αθέτησης, για την τράπεζα του δείγματος (επιλογή υποδείγματος, επιλογή επεξηγηματικών μεταβλητών, επιλογή και μορφοποίηση δεδομένων, κατασκευή της συνάρτησης διάκρισης, στατιστική επικύρωση του υποδείγματος, καθορισμός του σημείου διαχωρισμού (cut off point), Μέτρηση της αποτελεσματικότητας του υποδείγματος) εμπεριέχεται στο Κεφάλαιο 4.

Τέλος, το Κεφάλαιο 5 περιλαμβάνει τα συμπεράσματα της εργασίας, ενώ ακολουθούν οι βιβλιογραφικές αναφορές και το Παράρτημα, στο οποίο παρουσιάζεται περαιτέρω ανάλυση των δεδομένων και των αποτελεσμάτων της ανάλυσης.

## ΚΕΦΑΛΑΙΟ 1

### Ο ΠΙΣΤΩΤΙΚΟΣ ΚΙΝΔΥΝΟΣ : ΕΝΝΟΙΑ, ΒΑΣΙΚΕΣ ΠΑΡΑΜΕΤΡΟΙ, ΒΑΣΙΛΕΙΑ II

#### 1.1 Ο Πιστωτικός Κίνδυνος από τη σκοπιά των τραπεζικών επιχειρήσεων

Η βασική λειτουργία των τραπεζών είναι η ανακατανομή κεφαλαιακών πόρων από πλεονάζουσες σε ελλειμματικές μονάδες. Δανείζονται χρήματα από ιδιώτες ή επιχειρήσεις που παρουσιάζουν πλεόνασμα ρευστότητας σε κάποια χρονική στιγμή και τα δανείζουν σε ιδιώτες ή επιχειρήσεις που παρουσιάζουν έλλειμμα την ίδια χρονική περίοδο. Κατά τη χορήγηση της πίστωσης, οι τράπεζες προσδοκούν ότι οι πελάτες τους θα είναι συνεπείς στους όρους δανεισμού και επομένως, θα εκπληρώσουν τις υποχρεώσεις τους σε αυτές. Έτσι, θα μπορούν και αυτές να ανταποκριθούν στους καταθέτες και λοιπούς δανειστές τους, δεδομένου ότι θα εξασφαλισθεί η αρμονία εισροών – εκροών τους. Αν αυτό δεν συμβεί, οι τράπεζες θα αντιμετωπίσουν αντίστοιχα προβλήματα, τα οποία θα εκτείνονται από την απλή μείωση των κερδών τους μέχρι τη χρεοκοπία τους, ανάλογα με το μέγεθος των απωλειών.

Οι τράπεζες, λοιπόν, αναλαμβάνουν εξ ολοκλήρου τον πιστωτικό κίνδυνο, γι' αυτό είναι ζωτικής σημασίας γι' αυτές ο έγκαιρος εντοπισμός και η μέτρησή του.

Με βάση τα πιο πάνω, καθίσταται αυτονόητο ότι ο πιστωτικός κίνδυνος υπάρχει σε κάθε χρηματοδότηση και συνιστά την αβεβαιότητα που υπάρχει όσον αφορά στην εκπλήρωση της υποχρέωσης από τον αντισυμβαλλόμενο σύμφωνα με τους όρους που έχουν συμφωνηθεί (*Bessis , 2002*).

#### 1.2 Ταξινόμηση των τραπεζικών απαιτήσεων

Οι απαιτήσεις που έχουν οι τράπεζες στον ενεργητικό τους είναι συνήθως ομαδοποιημένες σε πέντε βασικές κατηγορίες, ανάλογα το είδος της απαίτησης και τα χαρακτηριστικά του αντισυμβαλλομένου:

- Απαιτήσεις έναντι κρατών. Τέτοιες απαιτήσεις προκύπτουν από την δανειοδότηση κάποιου κράτους, πχ με την αγορά ομολόγων ή με παροχή χορηγήσεων. Στην κατηγορία αυτή συγκαταλέγονται και οι απαιτήσεις από τις κεντρικές τράπεζες των κρατών.
- Απαιτήσεις έναντι επιχειρήσεων. Σ' αυτήν την κατηγορία ανήκουν οι απαιτήσεις έναντι μεσαίων και μεγάλων επιχειρήσεων. Στις απαιτήσεις αυτές συναντώνται και τα ειδικές περιπτώσεις δανεισμού (Specialized Lendings ή SL) τα οποία χωρίζονται σε 5 ομάδες: Project finance, object finance, commodities finance, income-producing real estate, high volatility commercial real estate.
- Απαιτήσεις έναντι τραπεζών και ασφαλιστικών εταιρειών.
- Απαιτήσεις λιανικής τραπεζικής (Retail banking), οι οποίες προκύπτουν από τη διάθεση τυποποιημένων προϊόντων που απευθύνονται σε ιδιώτες ή μικρές επιχειρήσεις και τα οποία περιλαμβάνουν κυρίως :
  - Στεγαστικά δάνεια,
  - Καταναλωτικά/προσωπικά δάνεια, ανοιχτούς αλληλόχρεους λογαριασμούς, overdrafts, πιστωτικές κάρτες.
  - Επιχειρηματικά δάνεια από μικρές επιχειρήσεις.

Μικρές επιχειρήσεις θεωρούνται αυτές όπου ο κύκλος εργασιών τους δεν υπερβαίνει 2,5 εκ. ευρώ. Οι υπόλοιπες ανοίκουν στις μεσαίες και μεγάλες επιχειρήσεις.

Οι απαιτήσεις στην λιανική τραπεζική δεν πρέπει να υπερβαίνουν το 1.000.000 ευρώ. Επίσης, το χαρτοφυλάκιο λιανικής τραπεζικής πρέπει να έχει μικρό κίνδυνο συγκέντρωσης. Έτσι για παράδειγμα, προκειμένου να επιτευχθεί η απαιτούμενη διασπορά οι τράπεζες μπορούν να περιορίσουν την έκθεση σε ένα αντισυμβαλλόμενο σε ποσοστό μικρότερο ή ίσο του το 0.2% του συνολικού των συνολικών απαιτήσεων του χαρτοφυλακίου λιανικής. (Βασιλεία II παρ. 70)

- Μετοχές, συμμετοχές σε άλλες εταιρείες.

Είναι προφανές ότι, για να διατηρηθεί ο πιστωτικός κίνδυνος σε ένα ανεκτό επίπεδο πρέπει να περιοριστούν οι χορηγήσεις σε πελάτες χαμηλής φερεγγυότητας. Όμως, με αυτόν τον τρόπο περιορίζεται και η αναμενόμενη κερδοφορία της τράπεζας (Γκλεζάκος, 2008). Προφανώς, οι επενδύσεις υψηλού κινδύνου δεν είναι επιθυμητές,

αλλά πολλές φορές επιλέγονται λόγω της προσδοκώμενης υψηλής αποδοτικότητάς τους. Γενικά, ο κίνδυνος και η αναμενόμενη απόδοση συνδέονται θετικά, γι' αυτό δεν μπορεί οποιοσδήποτε επενδυτής να προσδοκά υψηλή απόδοση αναλαμβάνοντας χαμηλό επίπεδο κινδύνου. Έτσι, η τράπεζα πρέπει να επιλέξει τον συνδυασμό κινδύνου και απόδοσης που είναι το ιδανικότερο για εκείνη.

Όταν οι χορηγήσεις επιλέγονται με μοναδικό κριτήριο την απόδοση, οι συνέπειες είναι καταστροφικές. Δυστυχώς, κρίσεις όπως αυτή που περνά η παγκόσμια οικονομία σήμερα έχουν τις ρίζες τους σε στρατηγικές που στόχευαν μόνο στο κέρδος αγνηφώντας τον κίνδυνο.

### 1.3 Παράμετροι του πιστωτικού κινδύνου

Οι βασικές παράμετροι που παίζουν πολύ σημαντικό ρόλο στην μέτρηση του πιστωτικού κινδύνου και κατά συνέπεια στον υπολογισμό των ελάχιστων κεφαλαιακών απαιτήσεων, είναι οι εξής :

- *Πιθανότητα αθέτησης του αντισυμβαλλομένου (Probability of Default ή PD)*
- *Ζημία λόγω της αθέτησης (Loss Given Default ή LGD)*
- *Κεφάλαια που εκτίθενται σε κίνδυνο (Exposure at Default ή EAD).*

**Η «Πιθανότητα αθέτησης» (PD)** του αντισυμβαλλομένου εκφράζει τον βαθμό στον οποίο είναι ενδεχόμενο να αδυνατεί ο αντισυμβαλλόμενος να εκπληρώσει την υποχρέωσή του απέναντι στην τράπεζα. Η παράμετρος αυτή είναι κρίσιμη για την εκτίμηση των ελάχιστων κεφαλαιακών απαιτήσεων και γι' αυτό οι τράπεζες δίνουν μεγάλη βαρύτητα στην εκτίμηση της.

Η εκτίμηση τη πιθανότητας αθέτησης βασίζεται στην πιστοληπτική διαβάθμιση των δανειοληπτών με βάση τα επιμέρους χαρακτηριστικά τους. Η τράπεζα, μέσω της αξιολόγησης των χαρακτηριστικών αυτών, αποδίδει μια τιμή (score) σε κάθε

δανειολήπτη. Στη συνέχεια, με βάση αυτό το score, υπολογίζει την πιθανότητα αθέτησης. (Loffler and Posch., 2007).

Αρκετά υποδείγματα χρησιμοποιούνται για την εκτίμηση της πιθανότητας αθέτησης και τα κυριότερα από αυτά παρουσιάζονται στο Κεφάλαιο 2.

**Η «Ζημία λόγω της αθέτησης» (LGD)** του αντισυμβαλλομένου, εκφράζει τη ζημία στην περίπτωση που ο αντισυμβαλλόμενος αδυνατεί να εκπληρώσει την υποχρέωση του. Η LGD δίνεται από τον τύπο :

$$LGD = 1 - Recovery Rate \quad (1.1)$$

Το ποσοστό ανάκτησης (Recovery Rate ή RR) εκφράζει το ποσοστό της αξίας των απαιτήσεων της τράπεζας που θα ικανοποιηθεί σε περίπτωση που ο δανειολήπτης αθετήσει την υποχρέωση του. Το RR εξαρτάται από το είδος και την ποιότητα των εξασφαλίσεων καθώς και τη σειρά προτεραιότητας που έχει η τράπεζα ανάμεσα το σύνολο των δανειστών του δανειολήπτη (Bluhm et al., 2003).

**Τα Κεφάλαια που εκτίθενται σε κίνδυνο (EAD)**, αποτελούν ένα μέγεθος που εκφράζει την έκθεση που έχει η τράπεζα στον αντισυμβαλλόμενο. Γενικά, όταν η τράπεζα έχει εγκρίνει ένα δάνειο η συνολική έκθεση που έχει μπορεί να χωριστεί σε δύο μέρη, τα ήδη χορηγηθέντα κεφάλαια (outstandings) και τα αχρησιμοποίητα κεφάλαια για τα οποία υπάρχει δέσμευση της τράπεζας για χορήγησή τους (commitments) – βλ. Bluhm et al., (2003). Σε περίπτωση οριστικής αθέτησης από τον δανειολήπτη, η τράπεζα χάνει το σύνολο των outstandings, διότι τα έχει ήδη εκταμιεύσει. Τα commitments προσδιορίζονται από το πιστωτικό όριο που έχει εγκριθεί στον δανειολήπτη. Η έκθεση που έχει η τράπεζα αποτελεί ένα μέρος ( $\gamma$ ) των commitments, εκτός αν έχει ήδη εξαντληθεί το πιστωτικό όριο από τον δανειολήπτη. Έχει παρατηρηθεί ότι σε δυσχερείς οικονομικές καταστάσεις υπάρχει η τάση από τους δανειολήπτες να εξαντλούν αυτό το πιστωτικό όριο που τους παρέχει η τράπεζα.

Τα κεφάλαια σε κίνδυνο (EAD) δίνονται από τον επόμενο τύπο:

$$EAD = outstandings + \gamma \times commitments \quad (1.2)$$

Οι πιο πάνω παράμετροι είναι πολύ σημαντικές για την εκτίμηση της αναμενόμενης και της μη αναμενόμενης απώλειας για τις τράπεζες, δεδομένου ότι διευκολύνουν τον προσδιορισμό του πιστωτικού κινδύνου (Bluhm et al., 2003).



Οι παράμετροι PD, LGD είναι οι αναμενόμενες τιμές των τυχαίων μεταβλητών  $I_D$  και SEV(severity).

Η μεταβλητή  $I_D$  εκφράζει την αθέτηση από τον δανειολήπτη μέσα ένα συγκεκριμένο χρονικό διάστημα το οποίο συμβαίνει με πιθανότητα PD :

$$I_D = \begin{cases} 1, & PD \\ 0, & 1 - PD \end{cases}$$

Όπως γίνεται αντιληπτό, ισχύει :  $E(I_D) = PD$ .

Η τυχαία μεταβλητή SEV εκφράζει το μέγεθος της απώλειας σε περίπτωση αθέτησης. Όπως έχει αναφερθεί ισχύει  $E(SEV)=LGD$ .

Οι  $I_D$  και SEV μεταβλητές θεωρούνται ανεξάρτητες. Επίσης, γίνεται η παραδοχή ότι το EAD είναι ντετερμινιστικό.

Η απώλεια εκφράζεται με την τυχαία μεταβλητή L και δίνεται από τον τύπο:

$$L = I_D \times SEV \times EAD \quad (1.3)$$

Η αναμενόμενη απώλεια προκύπτει από την (1.3) ως εξής:

$$E(L) = PD \times LGD \times EAD \quad (1.4)$$

Πρέπει να σημειωθεί ότι, στην πραγματικότητα, το EAD εκφράζει την μέση τιμή μιας τυχαίας μεταβλητής. Επίσης, ότι η υπόθεση της ανεξαρτησίας των τυχαίων μεταβλητών είναι αρκετά απλοϊκή.

Έχοντας υπολογίσει την αναμενόμενη απώλεια, οι τράπεζες πρέπει να διατηρούν τα ίδια κεφάλαια σε τέτοιο επίπεδο ώστε να καλύπτονται από αυτά οι πιθανές απώλειες. Στις περισσότερες όμως των περιπτώσεων, η ζημία είναι δυνατόν να υπερβεί την αναμενόμενη απώλεια και γι' αυτό το λόγο οι τράπεζες υποχρεούνται να υπολογίσουν και την «Μη αναμενόμενη απώλεια» UL (Unexpected Loss) η οποία εκτιμάται ως η τυπική απόκλιση της τυχαίας μεταβλητής L (*Bluhm et al., 2003*).

$$UL = \sqrt{Var(I_D \times SEV \times EAD)} \quad (1.5)$$

Με βάση τις παραδοχές ότι  $I_D$  και SEV είναι ανεξάρτητες και ότι το EAD είναι ντετερμινιστικό το UL δίνεται από τον τύπο:

$$UL = EAD \times \sqrt{\text{Var}(SEV) \times PD + LGD^2 \times PD(1 - PD)} \quad (1.6)$$

#### 1.4 Η μέτρηση του πιστωτικού κινδύνου στα πλαίσια του Συμφώνου της Βασιλείας II

Κάθε τράπεζα, για να μπορεί να είναι συνεπής απέναντι στους δανειστές της πρέπει ανάλογα με το ενεργητικό της να διαθέτει ένα ποσό ιδίων κεφαλαίων, τέτοιο, που να επαρκεί για την κάλυψη ζημιών λόγω του πιστωτικού κινδύνου. Όσο μεγαλύτερο πιστωτικό κίνδυνο ενέχει μια δανειοδότηση τόσο περισσότερα ίδια κεφάλαια πρέπει να έχει η τράπεζα για να είναι καλυμμένη. Προσπάθειες εκτίμησης του πιστωτικού κινδύνου, είτε ποιοτικά, είτε ποσοτικά, γίνονταν σε όλο το φάσμα των χρηματοπιστωτικών ιδρυμάτων, κάποιες σε ατομικό επίπεδο, άλλες σε επίπεδο χώρας, ποτέ όμως σε διεθνές επίπεδο. Η μέτρηση του πιστωτικού κινδύνου και ο υπολογισμός των ελάχιστων κεφαλαιακών απαιτήσεων, που είναι ζωτικής σημασίας για την τράπεζα, παύουν να γίνονται αυθαίρετα μετά την 1<sup>η</sup> Συνθήκη της Βασιλείας το 1988.

Η επιτροπή της Βασιλείας ιδρύθηκε το τέλος του 1974 από τους διοικητές των κεντρικών τραπεζών της ομάδας 7 κρατών, μελών της ΕΕ, των ΗΠΑ, του Καναδά και της Ιαπωνίας (G 10). Στόχος της Επιτροπής της Βασιλείας είναι η διασφάλιση της σταθερότητας του διεθνούς χρηματοπιστωτικού συστήματος, η διαμόρφωση δίκαιης αγοράς, καθώς και η ενίσχυση της ανταγωνιστικότητας σε διεθνές επίπεδο, μέσα από ισοδύναμους όρους.

Το πρώτο σύμφωνο για την κεφαλαιακή επάρκεια των τραπεζών (Basel Capital Accord I, International Convergence of Capital Measurement and Capital Standards) υπογράφηκε το 1988 και προέβλεπε τον τρόπο υπολογισμού των κεφαλαιακών απαιτήσεων για την κάλυψη του πιστωτικού κινδύνου. Τροποποιήθηκε το 1996 όπου συμπεριλήφθηκε και ο κίνδυνος της αγοράς. Σήμερα είναι σε ισχύ το 2<sup>ο</sup> Σύμφωνο της Βασιλείας (Basel Capital Accord II ή Βασιλεία II) που εκδόθηκε το 2004 και έχει φτάσει στη σημερινή του μορφή μετά από αλληπάλληλες τροποποιήσεις.

Βασική διαφορά του Συμφώνου της Βασιλείας II από το I είναι ότι εκτιμά τον λειτουργικό κίνδυνο και τις κεφαλαιακές απαιτήσεις για την κάλυψή του ξεχωριστά.



Επίσης, υπάρχουν και άλλοι κίνδυνοι που έχουν συμπεριληφθεί όπως ο επιτοκιακός, ο κίνδυνος ρευστότητας, ο κίνδυνος συγκέντρωσης και ο πολιτικός.

Το Σύμφωνο της Βασιλείας II στηρίζεται σε τρεις πυλώνες :

**Πυλώνας 1** : Ορίζει τους κανόνες υπολογισμού των ελάχιστων κεφαλαιακών απαιτήσεων για τον πιστωτικό κίνδυνο, τον κίνδυνο αγοράς και τον λειτουργικό κίνδυνο.

Όπως έχει αναφερθεί ήδη, ο λειτουργικός κίνδυνος είναι η ουσιαστική διαφορά της Βασιλείας II σε σχέση με τη Βασιλείας I. Αντίθετα, ο υπολογισμός των ελάχιστων κεφαλαιακών απαιτήσεων για τον κίνδυνο αγοράς είναι, σε γενικές γραμμές, ο ίδιος με το 1<sup>ο</sup> Σύμφωνο. Όσον αφορά στον πιστωτικό κίνδυνο υπάρχουν κάποιες διαφοροποιήσεις που κάνουν την προσέγγιση πιο ευαίσθητη στον κίνδυνο. Στη συνέχεια, θα γίνει εκτενής αναφορά στον υπολογισμό των ελάχιστων κεφαλαιακών απαιτήσεων για τον πιστωτικό κίνδυνο.

**Πυλώνας 2** : Αναφέρεται στη διαδικασία του εποπτικού ελέγχου σύμφωνα με την οποία αξιολογείται από τις αντίστοιχες κεντρικές τράπεζες κατά πόσο οι τράπεζες είναι ικανές να εκτιμήσουν τους κινδύνους και να υπολογίσουν τις ελάχιστες κεφαλαιακές απαιτήσεις.

Στον 2<sup>ο</sup> πυλώνα υπάρχουν οδηγίες για την εκτίμηση των διαφόρων κατηγοριών κινδύνων που δεν αναφέρονται στον Πυλώνα 1.

**Πυλώνας 3** : Εστιάζει στην πειθαρχία της αγοράς επιβάλλοντας στις τράπεζες τη εκτενή δημοσιοποίηση στοιχείων. Έτσι, πρέπει να δημοσιοποιούνται οικονομικά στοιχεία όπως είναι η κεφαλαιακή επάρκεια, επίσης δε κατά πόσο οι τράπεζες έχουν συμμορφωθεί με τις οδηγίες της Βασιλείας II. Η διαφάνεια αποσκοπεί στην ενίσχυση της πειθαρχίας που επιβάλλει η αγορά.

Ειδικά για τον πιστωτικό κίνδυνο, ο πυλώνας 1 επιτρέπει τη χρησιμοποίηση δύο προσεγγίσεων για την εκτίμηση των ελάχιστων κεφαλαιακών απαιτήσεων:

- α) Την Τυποποιημένη Προσέγγιση (Standardized Approach).
- β) Την Προσέγγιση Εσωτερικών Διαβαθμίσεων (Internal Ratings Based Approach ή IRB).

## 1.4 Τυποποιημένη προσέγγιση

Σύμφωνα με την προσέγγιση αυτή οι τράπεζες καλούνται να μετρήσουν τον πιστωτικό κίνδυνο με μια τυποποιημένη μέθοδο που όλες οι παράμετροι του πιστωτικού κινδύνου PD, LGD, EAD είναι προκαθορισμένες από την εποπτική αρχή. Επίσης, γίνεται χρήση των πιστοληπτικών διαβαθμίσεων από καθορισμένους Εξωτερικούς Οργανισμούς Πιστοληπτικής Αξιολόγησης (Ε.Ο.Π.Α.).

Η κεφαλαιακή επάρκεια βασίζεται στον υπολογισμό των Risk Weights (RW) τα οποία είναι Συντελεστές Στάθμισης που εκφράζουν τον βαθμό επικινδυνότητας των απαιτήσεων. Τα Risk Weights παίρνουν συγκεκριμένες τιμές αναλόγως με την πιστοληπτική ικανότητα του αντισυμβαλλομένου και τα χαρακτηριστικά της απαίτησης. Κατά συνέπεια, για τον υπολογισμό τους γίνεται μια ομαδοποίηση των στοιχείων του ενεργητικού της τράπεζας με βάση την επικινδυνότητα τους.

Μετά τον υπολογισμό των Risk Weights υπολογίζονται τα Σταθμισμένα Κατά Κίνδυνο Ανοίγματα (Risk Weighted Assets ή RWA) καθώς και οι ελάχιστες κεφαλαιακές απαιτήσεις (Minimum Capital requirements ή MCR) για την κάλυψη από τον πιστωτικό κίνδυνο, σύμφωνα με τις επόμενες σχέσεις :

$$RWA = RW \times \text{Αξία Ανοίγματος} \quad (2.1)$$

$$MCR = RWA \times 8\% \quad (2.2)$$

Ο δείκτης κεφαλαιακής επάρκειας μιας τράπεζας ορίζεται ως εξής:

$$\text{Capital Ratio} = \frac{\text{Ιδία Κεφάλαια}}{\text{Risk Weighted Assets}} \quad (2.3)$$

Κατ' απαίτηση της Βασιλείας II, θα πρέπει να ισχύει : Capital Ratio (CR)  $\geq$  8%. Αν μια τράπεζα έχει CR  $>$  8%, αυτό σημαίνει ότι δεσμεύει περισσότερα ίδια κεφάλαια για τον πιστωτικό κίνδυνο με βάση τις χορηγήσεις που έχει στο ενεργητικό της.

Για τον υπολογισμό των Risk Weights γίνεται χρήση της πιστοληπτικής διαβάθμισης των αντισυμβαλλομένων η οποία προκύπτει από καθορισμένους Ε.Ο.Π.Α. οι οποίοι είναι οι εξής :

- Moody's

- Fitch
- S&P

Στους πίνακες που ακολουθούν, περιλαμβάνονται ενδεικτικά τα Risk Weights για ορισμένες κατηγορίες στοιχείων ενεργητικού.

Απαιτήσεις έναντι κυβερνήσεων κρατών και κεντρικών τραπεζών :

Credit Assessment	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Risk Weight	0%	20%	50%	100%	150%	100%

Απαιτήσεις έναντι χρηματοπιστωτικών ιδρυμάτων/ Τραπεζών :

Στην κατηγορία αυτή, κατόπιν έγκρισης της εκάστοτε εθνικής εποπτικής αρχής, δίνονται στις τράπεζες δύο επιλογές:

α) Να υπολογιστούν τα Risk Weights βάσει του πίνακα :

Credit assessment of Sovereign	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Risk weight under Option 1	20%	50%	100%	100%	150%	100%

β) Να χωρίσουν τις απαιτήσεις σε βραχυχρόνιες και μακροχρόνιες. Βραχυχρόνιες απαιτήσεις είναι αυτές που έχουν διάρκεια μικρότερη των τριών μηνών. Τα RW υπολογίζονται με βάση τον παρακάτω πίνακα :

Credit assessment of Banks	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Risk weight under Option 2	20%	50%	50%	100%	150%	50%
Risk weight for short-term claims <sup>26</sup> under Option 2	20%	20%	20%	50%	150%	20%

Ποιά από τις δύο επιλογές θα χρησιμοποιηθεί καθορίζεται από την εθνική εποπτική αρχή και επιβάλλεται σε όλες τις τράπεζες του κράτους.

Απαιτήσεις έναντι μεσαίων και μεγάλων επιχειρήσεων :

Credit assessment	AAA to AA-	A+ to A-	BBB+ to BB-	Below BB-	Unrated
Risk weight	20%	50%	100%	150%	100%

Απαιτήσεις Λιανικής τραπεζικής :

Οι απαιτήσεις λιανικής τραπεζικής έχουν συντελεστή στάθμισης  $RW = 75\%$ .

Στις απαιτήσεις όπου ως εξασφάλιση έχει προσημειωθεί κάποιο ακίνητο γίνεται χρήση του δείκτη LTV που με βάση την τιμή του διαφοροποιείται το Risk Weight.

$$LTV = \frac{LOAN}{VALUE} \quad (2.4)$$

Ο δείκτης LTV εκφράζει ως ποσοστό τη σχέση μεταξύ του χορηγηθέντος ποσού (Loan) και της αξίας του υποθηκευμένου ακινήτου (Value). Η κριτική τιμή του LTV είναι 75%. Οι χορηγήσεις με  $LTV < 75\%$  θεωρούνται πιο ασφαλείς και έχουν μικρότερο Risk Weight.

Απαιτήσεις που καλύπτονται από αστικά ακίνητα :

Residential Real Estate	Risk Weight
LTV < 75%	35%
LTV > 75%	75%

Απαιτήσεις που καλύπτονται από επιχειρηματικά ακίνητα :

Commercial Real Estate	Risk Weight
LTV < 75%	50%
LTV > 75%	100%

Απαιτήσεις σε καθυστέρηση :

Οι τιμές που παίρνει ο συντελεστής στάθμισης  $RW$  για τις απαιτήσεις που είναι σε καθυστέρηση είναι :

α) 150% εάν οι προβλέψεις απομείωσης του ανοίγματος είναι μικρότερες από το 20% του μη καλυμμένου με εξασφαλίσεις τμήματος της απαίτησης προ προβλέψεων.

β) 100% εάν οι προβλέψεις απομείωσης του ανοίγματος αντιπροσωπεύουν τουλάχιστον το 20% του μη καλυμμένου με εξασφαλίσεις τμήματος της απαίτησης προ προβλέψεων.

Προβλέψεις απομείωσης του ανοίγματος είναι το ποσό που προβλέπει η τράπεζα ότι δεν θα εισπράξει σίγουρα.

## **1.6 Η Προσέγγιση των εσωτερικών διαβαθμίσεων (IRB)**

Οι τράπεζες που πληρούν τις προϋποθέσεις και έχουν την σχετική άδεια από την εποπτική αρχή μπορούν να χρησιμοποιήσουν την προσέγγιση εσωτερικών διαβαθμίσεων για να υπολογίσουν τις κεφαλαιακές απαιτήσεις. Σύμφωνα με αυτήν την προσέγγιση οι τράπεζες πρέπει να εκτιμήσουν τις βασικές παραμέτρους του πιστωτικού κινδύνου (PD, LGD, EAD) αναπτύσσοντας δικά τους υποδείγματα, βάσει των οποίων θα υπολογίσουν την αναμενόμενη απώλεια (EL:Expected Loss) και τη μη αναμενόμενη απώλεια (UL:Unexpected Loss).

## **1.7 Προσέγγιση εσωτερικών διαβαθμίσεων για χαρτοφυλάκιο λιανικής τραπεζικής σύμφωνα με τη Βασιλεία II**

Το Σύμφωνο της Βασιλείας σε αρκετές περιπτώσεις δίνει την δυνατότητα στις τράπεζες να χρησιμοποιήσουν δύο εναλλακτικές μορφές της IRB, την Foundation και την Advanced.

Οι τράπεζες που υιοθετούν την Foundation IRB (FIRB) καλούνται να υπολογίσουν από τις παραμέτρους του πιστωτικού κινδύνου μόνο το PD με δικό τους τρόπο. Για τον υπολογισμό των άλλων παραμέτρων (LGD , EAD) βασίζονται στις εκτιμήσεις της εποπτικής αρχής. Αντίθετα, σύμφωνα με την Advanced IRB (AIRB) απαιτείται από τις τράπεζες να υπολογίσουν όλες τις παραμέτρους του πιστωτικού κινδύνου με δικό τους τρόπο, ικανοποιώντας βέβαια τις απαιτήσεις της εποπτικής αρχής.

Για τις απαιτήσεις που ανήκουν στην κατηγορία της λιανικής τραπεζικής δεν υπάρχει η δυνατότητα επιλογής ανάμεσα τις προσεγγίσεις FIRB και AIRB. Οι τράπεζες

πρέπει να υπολογίσουν όλες τις παραμέτρους του πιστωτικού κινδύνου με δικό τους τρόπο. Το PD και το LGD εκτιμώνται ως ποσοστά ενώ το EAD μετριέται σε μονάδες χρήματος (ευρώ).

Από την Βασιλεία II προβλέπεται μια περίοδος μετάβασης για τις τράπεζες που πρόκειται να υιοθετήσουν την προσέγγιση εσωτερικών διαβαθμίσεων (FIRB ή AIRB) η οποία λήγει τρία χρόνια από την ημερομηνία υιοθέτησης της. Κατά την περίοδο μετάβασης θα είναι υποχρεωμένες να εκτιμούν τις κεφαλαιακές απαιτήσεις με την IRB προσέγγιση αλλά παράλληλα και με βάση το Σύμφωνο την Βασιλείας I του 1988, όπως παρουσιάζονται στις παραγράφους 45 ως 47 της Βασιλείας II. Ο περιορισμός αυτός ουσιαστικά παρέχει μια κατώτατη τιμή στις ελάχιστες κεφαλαιακές απαιτήσεις. Η εποπτική αρχή μπορεί να κρίνει ότι κατά την περίοδο μετάβασης κάποιες (συγκεκριμένες) απαιτήσεις δύναται να χαλαρώσουν.

Εφόσον μια τράπεζα υιοθετήσει την προσέγγιση IRB θα την εφαρμόσει για το σύνολο των απαιτήσεων και των business units. Επίσης, θα πρέπει να συνεχίσουν να εφαρμόζουν την IRB προσέγγιση καθώς η επιστροφή στην τυποποιημένη προσέγγιση δεν επιτρέπεται παρά μόνο σε πολύ ειδικές περιπτώσεις.

### **Παράμετροι του πιστωτικού κινδύνου στην λιανική τραπεζική**

Η τράπεζα αφού χωρίσει σε κατηγορίες τις απαιτήσεις πρέπει να εκτιμήσει το PD και το LGD για κάθε ομάδα σύμφωνα με τις ελάχιστες απαιτήσεις του συμφώνου της Βασιλείας II.

Το PD υπολογίζεται ως η μεγαλύτερη τιμή της πιθανότητας αθέτησης σε χρονικό ορίζοντα ενός έτους (1 year PD) με βάση τον εσωτερικό βαθμό πιστοληπτικής ικανότητας που έχει αποδοθεί στην ομάδα, όμως δεν μπορεί να είναι μικρότερο του 0,03% έτσι :

$$PD = \max (1 \text{ year } PD, 0.03\%) \quad (2.5)$$

Παράγοντες που μειώνουν τον πιστωτικό κίνδυνο όπως εγγυήσεις και παράγωγα προϊόντα πιστωτικού κινδύνου πρέπει να λαμβάνονται υπόψη στην εκτίμηση των PD, LGD.

Όταν υπάρχει εγγυητής στην απαίτηση έναντι ενός πελάτη τότε το Risk Weight δεν μπορεί να είναι μικρότερο από αυτό που θα προέκυπτε στην περίπτωση που η απαίτηση ήταν έναντι του εγγυητή αποκλειστικά. Η ύπαρξη εγγυητή για τον αντισυμβαλλόμενο δεν επιτρέπει στην τράπεζα να κάνει χρήση της διπλής εξασφάλισης (double default treatment).

Για τον υπολογισμό του EAD λαμβάνονται υπόψη οι απαιτήσεις εντός και εκτός ισολογισμού. Το EAD για τα χρήματα που έχουν εκταμιευθεί για μια απαίτηση δεν μπορεί να είναι μικρότερο από το άθροισμα:

- α) Του ποσού κατά το οποίο θα μειώνονταν τα εποπτικά κεφάλαια αν η απαίτηση διαγραφόταν τελείως.
- β) Των ειδικών προβλέψεων και μερικών διαγραφών.

Η ειδική πρόβλεψη είναι το ποσό που εκτιμά η τράπεζα ότι δεν θα ανακτήσει.

Για απαιτήσεις λιανικής τραπεζικής εκτός ισολογισμού οι τράπεζες πρέπει να εκτιμήσουν έναν συντελεστή μετατροπής τους σε ισοδύναμες πιστώσεις (Credit Convertor Factor ή CCF).

### **Εκτίμηση της Μη Αναμενόμενης Απώλειας για απαιτήσεις Λιανικής Τραπεζικής**

Μετά τον υπολογισμό των παραμέτρων πιστωτικού κινδύνου PD, LGD, EAD το χαρτοφυλάκιο των απαιτήσεων χωρίζεται σε τρεις κατηγορίες :

- α) Στεγαστικά δάνεια (Residential retail exposures)
- β) Ανοιχτοί αλληλόχρεοι λογαριασμοί. (Qualifying revolving retail exposures)
- γ) Λοιπές απαιτήσεις λιανικής (Other retail exposures).

Για κάθε μια από τις τρεις κατηγορίες ορίζονται συγκεκριμένες Risk weight συναρτήσεις που παίρνουν ως δεδομένες τις παραμέτρους PD, LGD, EAD, που έχει εκτιμήσει η τράπεζα με δικό της τρόπο και παράγουν τις κεφαλαιακές απαιτήσεις για το UL.

### Απαιτήσεις που καλύπτονται με αστικά ακίνητα

Για απαιτήσεις που δεν είναι σε καθυστέρηση (default) και έχουν ως κάλυμμα αστικά ακίνητα, τα risk weights υπολογίζονται με τους ακόλουθους τύπους:

- $Correlation (R) = 0.15$
- $Capital\ requirements (K) = LGD \times N \left[ (1 - R)^{-0.5} \times N^{-1}(PD) + \left( \frac{R}{1-R} \right)^{0.5} \times N - 10.999 - PD \times LGD \right]$
- $RWA = K \times 12.5 \times EAD$

### Ανοιχτοί αλληλόχρεοι λογαριασμοί

Για τους ανοιχτούς αλληλόχρεους λογαριασμούς που δεν είναι σε καθυστέρηση (default) τα risk weights υπολογίζονται με τις ακόλουθους τύπους :

- $Correlation (R) = 0.04$
- $Capital\ requirements (K) = LGD \times N \left[ (1 - R)^{-0.5} \times N^{-1}(PD) + \left( \frac{R}{1-R} \right)^{0.5} \times N - 10.999 - PD \times LGD \right]$
- $Risk\ weighted\ assets = K \times 12.5 \times EAD$

### Λοιπές απαιτήσεις λιανικής

Για τις λοιπές απαιτήσεις λιανικής που δεν είναι σε καθυστέρηση (default) τα risk weights υπολογίζονται με τους ακόλουθους τύπους :

- $Correlation (R) = \frac{0.03 \times (1 - e^{(-35 \times PD)})}{1 - e^{-35}} + 0.16 \times \left[ 1 - \frac{1 - e^{(-35 \times PD)}}{1 - e^{-35}} \right]$
- $Capital\ requirements (K) = LGD \times N \left[ (1 - R)^{-0.5} \times N^{-1}(PD) + \left( \frac{R}{1-R} \right)^{0.5} \times N - 10.999 - PD \times LGD \right]$
- $Risk\ weighted\ assets = K \times 12.5 \times EAD$

Όπου, N: η συνάρτηση κατανομής της τυπικής κανονικής κατανομής.

Για απαιτήσεις λιανικής τραπεζικής που είναι σε καθυστέρηση η κεφαλαιακή απαίτηση δίνεται από τον τύπο  $K = \max(0, LGD - EL)$ , όπου EL είναι η καλύτερη εκτίμηση της τράπεζας για την αναμενόμενη απώλεια. Τα RWA υπολογίζονται με τον ίδιο τρόπο.



### **Υπολογισμός της αναμενόμενης απώλειας**

Εφόσον η τράπεζα έχει εκτιμήσει τις παραμέτρους του πιστωτικού κινδύνου μπορεί να εκτιμήσει την αναμενόμενη απώλεια ως το γινόμενο:

$$EL = PD \times LGD \quad (2.6)$$

Με τον τύπο (2.6) υπολογίζεται το EL ως ποσοστό. Αν πολλαπλασιαστεί και με το EAD τότε προκύπτει η αναμενόμενη απώλεια ως χρηματικό ποσό.

$$EL = PD \times LGD \times EAD \quad (2.7)$$

Οι σχέσεις (2.6) και (2.7) ισχύουν για απαιτήσεις που δεν είναι σε καθυστέρηση. Υπάρχει διαφοροποίηση στον υπολογισμό του EL των απαιτήσεων που είναι σε καθυστέρηση καθώς επίσης και για ειδικές περιπτώσεις δανεισμού (SL) και απαιτήσεις υψηλής διακύμανσης (HVCRE).

### **Ομαδοποίηση και εσωτερική διαβάθμιση των απαιτήσεων**

Η τράπεζα πρέπει να έχει αναπτύξει μεθόδους εσωτερικής διαβάθμισης των απαιτήσεων που βασίζεται σε χαρακτηριστικά επικινδυνότητας και του αντισυμβαλλομένου και του είδους της απαίτησης. Επίσης, πρέπει να ομαδοποιηθούν οι απαιτήσεις με βάση τα ομοιογενή χαρακτηριστικά τους. Στη συνέχεια, η τράπεζα πρέπει για κάθε ομάδα να εκτιμήσει τις παραμέτρους του πιστωτικού κινδύνου PD, LGD, EAD.

Για να τοποθετήσει η τράπεζα μια απαίτηση σε μια ομάδα πρέπει να λάβει υπόψη της τα χαρακτηριστικά επικινδυνότητας και του αντισυμβαλλομένου και του είδους της απαίτησης. Επίσης, η τράπεζα πρέπει να είναι σε θέση να αναγνωρίζει τις απαιτήσεις που είναι επισφαλείς (Delinquent).

Κατά την ομαδοποίηση ο αριθμός των απαιτήσεων σε μια ομάδα πρέπει να εξασφαλίζει ένα επίπεδο διαφοροποίησης τέτοιο ώστε η εκτίμηση των παραμέτρων του πιστωτικού κινδύνου που να έχει νόημα. Πρέπει να δοθεί ιδιαίτερη σημασία στην κατανομή των απαιτήσεων στις ομάδες για τη σωστή εκτίμηση των παραμέτρων PD, LGD, EAD.

Η τράπεζα πρέπει να έχει αναπτύξει ένα μοντέλο εσωτερικής διαβάθμισης με συγκεκριμένους ορισμούς, διαδικασίες και κριτήρια για την αντιστοίχιση ενός score σε κάθε απαίτηση. Οι ορισμοί και τα κριτήρια πρέπει να είναι εύλογα και διαισθητικά επίσης, πρέπει να συμβάλουν στην διαφοροποίηση των απαιτήσεων ως προς τον κίνδυνο.

- Πρέπει οι περιγραφές των διαβαθμίσεων και τα κριτήρια να είναι λεπτομερώς καταγεγραμμένα έτσι ώστε τα άτομα τα οποία είναι επιφορτισμένα να με την διαβάθμιση να απονέμουν με συνέπεια το ίδιο score στις απαιτήσεις με το ίδιο επίπεδο κινδύνου.
- Οι ορισμοί του συστήματος διαβάθμισης πρέπει να είναι ξεκάθαροι και αναλυτικοί, έτσι ώστε να επιτρέπει σε τρίτους να τους κατανοήσουν.
- Τα κριτήρια πρέπει να αποτελούνται από τα εσωτερικά στάνταρ δανειοδότησης και τις τακτικές της τράπεζας για το χειρισμό των «προβληματικών» απαιτήσεων.

Για την σωστή διαβάθμιση πρέπει να συμπεριλαμβάνεται στο μοντέλο η περισσότερη δυνατή πληροφορία για τους αντισυμβαλλομένους και τις απαιτήσεις. Όσο λιγότερη πληροφορία υπάρχει τόσο πιο συντηρητική πρέπει να είναι η τράπεζα ως προς τα αποτελέσματα της εσωτερικής διαβάθμισης των απαιτήσεων.

Ο χρονικός ορίζοντας για την εκτίμηση του PD είναι ένας χρόνος, όμως αναμένεται από τις τράπεζες να κάνουν την εσωτερική διαβάθμιση με βάση ένα μεγαλύτερο χρονικό ορίζοντα. Το score που αποδίδει η τράπεζα σε έναν δανειζόμενο πρέπει εκφράζει την συμπεριφορά του όσον αφορά στην υποχρέωσή του και σε δυσμενείς οικονομικές συνθήκες ή σε αναπάντεχα γεγονότα. Γι' αυτό πρέπει να χρησιμοποιούνται κατάλληλα σενάρια προσομοίωσης καταστάσεων κρίσης (stress scenarios) για την απόδοση του κατάλληλου score στον δανειολήπτη. Όμως, επειδή υπάρχει δυσκολία στην εκτίμηση των μελλοντικών γεγονότων και της επιρροής αυτών πρέπει η τράπεζα να ρίχνει μια συντηρητική ματιά στις εκτιμώμενες μελλοντικές πληροφορίες.

Τέλος, η τράπεζα πρέπει να ελέγχει τα χαρακτηριστικά επικινδυνότητας και την συμπεριφορά των αντισυμβαλλομένων της κάθε ομάδας απαιτήσεων τουλάχιστον σε ετήσια βάση (σε επίπεδο ομάδας). Ο έλεγχος θα πρέπει να γίνεται και σε ατομικό

επίπεδο έτσι ώστε να εξετάζεται αν οι δανειολήπτες εξακολουθούν να βρίσκονται στη σωστή ομάδα. Ένα αντιπροσωπευτικό δείγμα είναι ικανό να δώσει ικανοποιητικά αποτελέσματα για τους παραπάνω ελέγχους.

### **Χρήση υποδειγμάτων για την εκτίμηση των παραμέτρων του πιστωτικού κινδύνου**

Η εκτίμηση των παραμέτρων του πιστωτικού κινδύνου βασίζεται κυρίως σε στατιστικά υποδείγματα και άλλες μηχανικές μεθόδους. Τα υποδείγματα πιστοληπτικής διαβάθμισης (Credit Scoring Models) χρησιμοποιούν ένα υποσύνολο της διαθέσιμης πληροφορίας. Τα αποτελέσματα από τα στατιστικά υποδείγματα και τις μηχανικές μεθόδους πρέπει να ελέγχονται από την ανθρώπινη κρίση και λογική για να ελαχιστοποιούνται τυχόν λάθη.

- Η τράπεζα πρέπει να αποδείξει στην εποπτική αρχή ότι το υπόδειγμα ή η διαδικασία που χρησιμοποιεί έχει υψηλή προβλεπτική ικανότητα και ότι με τη χρήση του δεν διαστρεβλώνονται οι ελάχιστες κεφαλαιακές απαιτήσεις. Οι μεταβλητές που χρησιμοποιούνται στο υπόδειγμα πρέπει να αποτελούν ένα λογικό σύνολο ερμηνευτικών μεταβλητών. Επίσης, το υπόδειγμα πρέπει να είναι αποτελεσματικό κατά μέσο όρο στο σύνολο των απαιτήσεων που εφαρμόζεται.
- Η τράπεζα πρέπει να αποδείξει ότι τα δεδομένα που χρησιμοποιήθηκαν για την ανάπτυξη του υποδείματος είναι αντιπροσωπευτικά του πληθυσμού της.
- Συνδυάζοντας την ανθρώπινη κρίση με το υπόδειγμα πρέπει η ανθρώπινη κρίση να μπορεί να αντιληφθεί όλη τη σημαντική πληροφορία που δεν λαμβάνεται υπόψη από το υπόδειγμα. Η τράπεζα πρέπει να έχει οδηγίες για το πώς μπορεί η ανθρώπινη κρίση να συνδυαστεί με τα αποτελέσματα του υποδείματος.
- Η τράπεζα πρέπει να έχει αναπτύξει διαδικασίες για τον εντοπισμό των αδυναμιών του υποδείματος με σκοπό την βελτίωση του.

- Η επικύρωση του υποδείγματος πρέπει να βασίζεται σε έναν πλήρη κανονικό κύκλο επικύρωσης υποδειγμάτων που περιλαμβάνει την παρακολούθηση της αποτελεσματικότητας και της σταθερότητας, επισκόπηση των σχέσεων που συμπεριλαμβάνονται στο υπόδειγμα καθώς και έλεγχο των αποτελεσμάτων.

### **Τεκμηρίωση του υποδείγματος**

Η τράπεζα πρέπει να αποδείξει ότι το υπόδειγμα που χρησιμοποιεί είναι σύμφωνο με τις ελάχιστες απαιτήσεις της Βασιλείας II. Επίσης, πρέπει η θεωρία που βασίζεται το υπόδειγμα να παρουσιασθεί αναλυτικά. Για την επικύρωση του υποδείγματος πρέπει να εφαρμοστούν αυστηρές στατιστικές διαδικασίες όπως μέτρηση της αποτελεσματικότητας του σε διαφορετικό δείγμα (out of sample performance test). Τέλος, πρέπει να αναφερθεί κάτω από ποιές συνθήκες το υπόδειγμα δεν είναι αποτελεσματικό.

### **Ορισμός της αθέτησης**

Ο ορισμός της καθυστέρησης/αθέτησης της υποχρέωσης του αντισυμβαλλομένου έναντι στην τράπεζα συμπεριλαμβάνεται στις δύο ακόλουθες περιπτώσεις :

- Η τράπεζα θεωρεί ότι ο αντισυμβαλλόμενος είναι αδύνατο να εκπληρώσει τις υποχρεώσεις του χωρίς αναχρηματοδότηση.
- Ο αντισυμβαλλόμενος έχει καθυστερήσει τις πληρωμές του περισσότερο από 90 μέρες για οποιαδήποτε οφειλή του στην τράπεζα.

Στην λιανική τραπεζική η αθέτηση ορίζεται ανάλογα σε επίπεδο απαίτησης και όχι οφειλέτη. Έτσι αν κάποιος δανειολήπτης έχει παραπάνω από μια οφειλές στην τράπεζα η κάθε μια εξετάζεται ξεχωριστά.

## **1.8 Απαιτήσεις Βασιλείας II για την εκτίμηση της πιθανότητας αθέτησης**

α) Απαιτήσεις έναντι μεγάλων και μεσαίων επιχειρήσεων, κυβερνήσεων κρατών και πιστωτικών ιδρυμάτων:

Η τράπεζα πρέπει να χρησιμοποιεί πληροφορίες και τεχνικές που να λαμβάνουν κατάλληλα υπόψη τους την μακρόχρονη εμπειρία για την εκτίμηση του PD. Μπορεί

για παράδειγμα να χρησιμοποιήσει μια από τις τρεις τεχνικές: internal default experience, mapping to external data, and statistical default models.

Συνήθως χρησιμοποιείται μια βασική τεχνική για την εκτίμηση του PD όμως υπάρχει η δυνατότητα να χρησιμοποιεί τις άλλες τεχνικές για να κάνει σύγκριση των αποτελεσμάτων. Η τεχνική πρέπει να υποστηρίζεται από ανάλυση για να γίνει αποδεκτή από την εποπτική αρχή.

- Η τράπεζα μπορεί να χρησιμοποιήσει δικά της δεδομένα για την εκτίμηση του PD. Στην περίπτωση που τα εσωτερικά δεδομένα είναι περιορισμένα η τράπεζα πρέπει να προσθέσει ένα επιπλέον περιθώριο (margin of conservatism) στην εκτίμηση της πιθανότητας αθέτησης. Η τράπεζα μπορεί να χρησιμοποιήσει ομαδοποιημένα εξωτερικά δεδομένα αρκεί να αποδείξει ότι τα κριτήρια ομαδοποίησης και οι μέθοδοι εσωτερικής διαβάθμισης αυτών είναι παρόμοια με τα δικά της.
- Η πιστοληπτική διαβάθμιση και ο προσδιορισμός της πιθανότητας αθέτησης πρέπει να γίνει σε κλίμακα εξουσιοδοτημένων Εξωτερικών Οργανισμών Πιστοληπτικής Αξιολόγησης (Ε.Ο.Π.Α.). Ο σχεδιασμός του υποδείγματος πρέπει να γίνει συγκρίνοντας τα κριτήρια της τράπεζας με τα κριτήρια που χρησιμοποιεί η εταιρεία πιστοληπτικής αξιολόγησης καθώς και τα scores που έχουν αποδοθεί σε δανειολήπτες με παρόμοια χαρακτηριστικά εκατέρωθεν. Τα κριτήρια που χρησιμοποιούνται για την ποσοτικοποίηση του κινδύνου πρέπει να αναφέρονται στον δανειολήπτη και όχι στο είδος του δανείου. Επίσης, πρέπει να συγκριθεί και ο ορισμός της αθέτησης που χρησιμοποιεί η τράπεζα με αυτόν της εταιρείας πιστοληπτικής αξιολόγησης.
- Η τράπεζα μπορεί να χρησιμοποιήσει το μέσο όρο των PD για δανειολήπτες που έχουν το ίδιο score στην πιστοληπτική διαβάθμιση.

Τα ιστορικά δεδομένα στα οποία βασίζεται η εκτίμηση της πιθανότητας αθέτησης πρέπει να είναι τουλάχιστον πενταετίας. Σε περίπτωση που θεωρηθεί απαραίτητο δύναται να συμπεριληφθούν στο δείγμα και δεδομένα τα οποία εκτείνονται πέραν της πενταετίας. Επίσης, υπάρχει η δυνατότητα να μην δοθεί η ίδια βαρύτητα σε όλα τα δεδομένα εφόσον αποδειχθεί από την τράπεζα ότι τα πιο πρόσφατα προβλέπουν καλύτερα τον βαθμό απώλειας.

## β) Απαιτήσεις στη λιανική τραπεζική:

1. Οι τράπεζες πρέπει να αναπτύξουν συγκεκριμένη μεθοδολογία για να ομαδοποιούν τις απαιτήσεις. Κατ' αρχήν, πρέπει να δώσουν προτεραιότητα στα δικά τους δεδομένα για την εκτίμηση της πιθανότητας αθέτησης. Υπάρχει όμως και η δυνατότητα να χρησιμοποιηθούν εξωτερικά δεδομένα και στατιστικά υποδείγματα αρκεί να αποδειχθεί ότι υπάρχει δυνατή σχέση ανάμεσα α) στην εσωτερική και εξωτερική διαδικασία ομαδοποίησης των απαιτήσεων και β) στο προφίλ κινδύνου της τράπεζας και στη σύνθεση των εξωτερικών δεδομένων.

2. Τα ιστορικά δεδομένα στα οποία βασίζεται η εκτίμηση της πιθανότητας αθέτησης πρέπει να είναι τουλάχιστον πενταετίας. Σε περίπτωση που θεωρηθεί απαραίτητο δύναται να συμπεριληφθούν στο δείγμα και δεδομένα τα οποία εκτείνονται πέραν της πενταετίας. Επίσης, υπάρχει η δυνατότητα να μην δοθεί η ίδια βαρύτητα σε όλα τα δεδομένα εφόσον αποδειχθεί ότι τα πιο πρόσφατα προβλέπουν καλύτερα τον βαθμό απώλειας (loss rate).

3. Οι τράπεζες πρέπει να λάβουν υπόψη τους στην εκτίμηση της πιθανότητας αθέτησης και το seasoning effect καθώς η επιτροπή της Βασιλείας θεωρεί ότι αποτελεί ένα χρήσιμο εργαλείο. Συγκεκριμένα, για τις απαιτήσεις που χαρακτηρίζονται από seasoning effect, το οποίο κορυφώνεται μερικά χρόνια μετά την σύναψη (origination), η τράπεζα ύστερα από κάποια χρονική στιγμή πρέπει να περιμένει ραγδαία αύξηση της έκθεσης στον κίνδυνο. Η τράπεζα πρέπει να έχει πάρει τα απαραίτητα μέτρα απέναντι στο seasoning effect, έτσι ώστε να αποδείξει στην εποπτική αρχή ότι οι τεχνικές που έχει αναπτύξει είναι αποτελεσματικές και ότι δεν πρόκειται να αντιμετωπίσει πρόβλημα κεφαλαιακής επάρκειας στο μέλλον.

## ΚΕΦΑΛΑΙΟ 2

### ΥΠΟΔΕΙΓΜΑΤΑ ΕΚΤΙΜΗΣΗΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΑΘΕΤΗΣΗΣ

Σ' αυτό το κεφάλαιο παρουσιάζονται τρία από τα βασικότερα και ευρέως διαδεδομένα υποδείγματα εκτίμησης της πιθανότητας αθέτησης.

Τα υποδείγματα που θα παρουσιαστούν είναι τα εξής :

- α) η Πολυμεταβλητή Ανάλυση Διαχωρισμού (Multivariate Discriminate Analysis) που ανήκει στην οικογένεια των υποδειγμάτων διαχωρισμού,
- β) η Λογιστική Παλινδρόμηση (Logit) και
- γ) το υπόδειγμα Probit.

Τα υποδείγματα Probit και Logit ανήκουν στην οικογένεια των γενικευμένων υποδειγμάτων εκτίμησης πιθανοτήτων.

#### 2.1 Multivariate discriminate analysis (MDA)

Η MDA είναι μια δοκιμασμένη μεθοδολογία ταξινόμησης των μονάδων ενός πληθυσμού σε διακριτές ομάδες (*Hair et al. (1998) και Huberty και Olenjnik (2006)*).

Η ταξινόμηση αυτή επιτυγχάνει μέσω του εντοπισμού των παραμέτρων που χαρακτηρίζουν κάθε μονάδα του πληθυσμού, της συναρτησιακής σχέσης που τις συνδέει και της διαμόρφωσης ενός μεγέθους (score) το οποίο παίρνει διαφορετικές τιμές για διαφορετικές ομάδες.

Η MDA είναι κατάλληλη σε περιπτώσεις που η εξαρτημένη μεταβλητή  $Y$  δεν μπορεί να εκφρασθεί ποσοτικά, δηλαδή δεν έχει μονάδα μέτρησης. Όταν επιχειρείται ο διαχωρισμός του πληθυσμού σε δύο ομάδες, η MDA αναφέρεται ως «Two-group Discriminant Analysis», ενώ για τρεις και περισσότερες ομάδες ονομάζεται «Multivariate Discriminant Analysis».

Για παράδειγμα, κατά τον διαχωρισμό των απαιτήσεων μιας τράπεζας σε υγιείς και επισφαλείς, εφαρμογή έχει η Two-group Discriminant Analysis, της οποίας η εξαρτημένη τυχαία μεταβλητή  $Y_i$  εκφράζεται ως εξής :

$$Y_i = \begin{cases} 1, & \text{ασυνεπής} \\ 0, & \text{συνεπής} \end{cases}$$

Σε αυτή την περίπτωση, η εκτίμηση της πιθανότητας αθέτησης (PD) για τον  $i$  πιστούχο, εκφράζεται με την πιθανότητα η τυχαία μεταβλητή  $Y_i$  να πάρει την τιμή 1,  $Pr(Y_i = 1)$ .

Η MDA παράγει τις συναρτήσεις διάκρισης (Discriminate Function) οι οποίες αποτελούν συνδυασμό των ανεξάρτητων τυχαίων μεταβλητών που εκφράζουν τα χαρακτηριστικά στα οποία οφείλεται η διαφορετική συμπεριφορά των ατόμων του πληθυσμού.

Η συνάρτηση διάκρισης έχει την παρακάτω μορφή :

$$Z_i = b_{i0} + b_1X_{i1} + b_2X_{i2} + \dots + b_kX_{ik} \quad (3.1)$$

Για την ανάπτυξη υποδειγμάτων Discriminant Analysis, οι συναρτήσεις διάκρισης που απαιτούνται είναι κατά μία λιγότερες από τον αριθμό των κατηγοριών στις οποίες χωρίζεται ο πληθυσμός. Έτσι, στην περίπτωση του διαχωρισμού των δανειοληπτών σε συνεπείς και ασυνεπείς, χρειάζεται μόνο μια συνάρτηση διάκρισης.

Η διάκριση επιτυγχάνεται υπολογίζοντας τα βάρη των τυχαίων μεταβλητών έτσι ώστε να επιτευχθεί μεγιστοποίηση της διασποράς μεταξύ των κλάσεων και συνάμα να ελαχιστοποιηθεί η διασπορά μεταξύ των στοιχείων της κάθε ομάδας.

Μετά τον υπολογισμό των συντελεστών των τυχαίων μεταβλητών  $X_i$ , για κάθε μονάδα αντιστοιχεί ένα score («Discriminant Z score»), που προκύπτει από τις τιμές των  $X_i$  και αποτελεί την εξαρτημένη τυχαία μεταβλητή  $Z_i$ .

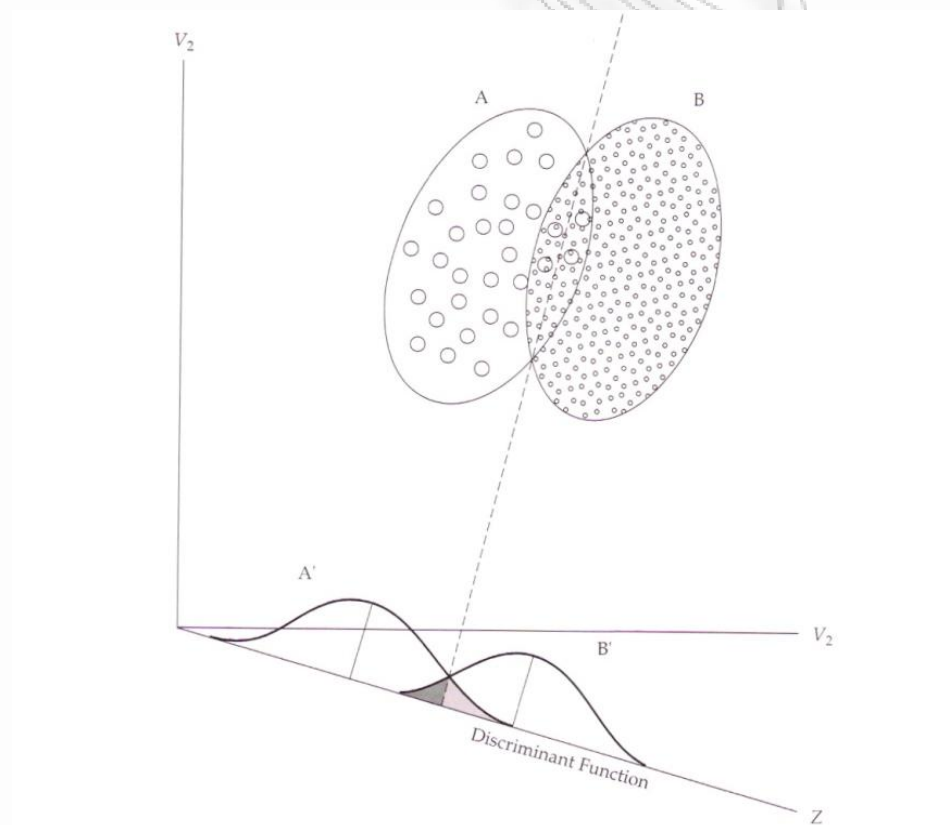
Ο μέσος όρος των  $Z$  ονομάζεται κεντροειδής (Centroid). Οι τιμές των  $Z$  κάθε ομάδας κυμαίνονται γύρω από τον κεντροειδή και έτσι η απόσταση μεταξύ των κεντροειδών των ομάδων δείχνει ως πρώτη προσέγγιση πόσο μακριά είναι οι ομάδες, δηλαδή πόσο αισθητά διαφέρουν οι παρατηρήσεις που περιέχουν.



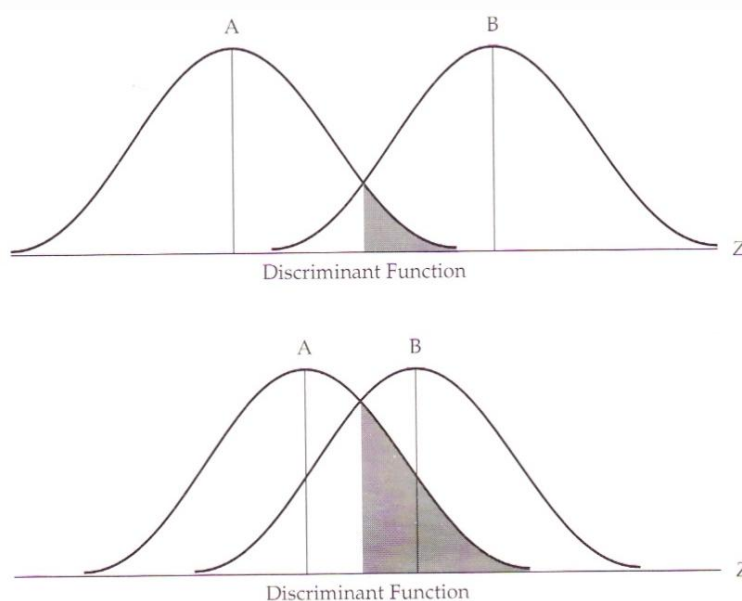
Πέρα όμως από τον κεντροειδή, σημαντικό ρόλο στη διαφοροποίηση των κλάσεων έχει και η διασπορά που υπάρχει σε κάθε μια από αυτές. Στόχος είναι οι κατανομές των παρατηρήσεων της κάθε ομάδας να μην επικαλύπτονται, δηλαδή να μην υπάρχουν παρατηρήσεις που ανήκουν σε άλλη ομάδα. Για να είναι εμφανής ο διαχωρισμός των κατηγοριών πρέπει οι παρατηρήσεις να είναι αρκετά κοντά στον κεντροειδή, δηλαδή πρέπει η διασπορά μέσα στις κλάσεις να είναι μικρή.

Η αποτελεσματικότητα των υποδειγμάτων MDA ενισχύεται όταν τα δεδομένα του δείγματος διαχωρίζονται σε κλάσεις που απέχουν μεταξύ τους όσο γίνεται περισσότερο, αλλά παρουσιάζουν την ελάχιστη δυνατή εσωτερική διασπορά.

Στο γράφημα 3.1 που ακολουθεί απεικονίζεται γραφικά το πως η συνάρτηση διάκρισης χωρίζει έναν πληθυσμό σε δύο κλάσεις.



Γράφημα 3.1 (Hair et al., 1998)



Γράφημα 3.2 (Hair et al., 1998)

Στο γράφημα 2 παρουσιάζονται δύο ομάδες με κεντροειδείς A και B που στη μία περίπτωση ο διαχωρισμός είναι ικανοποιητικός ενώ στην άλλη υπάρχει επικάλυψη των κατανομών των παρατηρήσεων, δεν διαχωρίζονται καλά οι ομάδες. Γίνεται αντιληπτό ότι στην δεύτερη περίπτωση θα υπήρχε καλύτερος διαχωρισμός αν οι κεντροειδείς A και B απείχαν περισσότερο ή αν η διασπορά μέσα στις ομάδες ήταν πιο μικρή, σαφώς το ιδανικό θα ήταν να ίσχυαν και τα δύο.

Για την κατάταξη των παρατηρήσεων σε κατηγορίες πρέπει να υπολογιστούν τα σημεία πρόκρισης-απόρριψης τα οποία ονομάζονται cut off points ή cutting scores. Τα Z scores που υπολογίζονται για κάθε μονάδα συγκρίνονται με τα cutting scores και ανάλογα τοποθετούνται στις ομάδες. Όταν υπάρχουν δύο μόνο κατηγορίες τότε πρέπει να εκτιμηθεί μόνο ένα cutting score. Έτσι, το Z score μιας μονάδος συγκρίνεται με το μοναδικό cutting score και αν είναι μεγαλύτερο κατατάσσεται στη μία ομάδα ενώ αν είναι μικρότερο κατατάσσεται στην άλλη. Παρακάτω παρουσιάζεται ο τρόπος εκτίμηση τους.

### Επιλογή των ερμηνευτικών μεταβλητών και κανόνες δειγματοληψίας

Η επιλογή των ανεξάρτητων τυχαίων μεταβλητών  $X_i$  γίνεται με δύο τρόπους :

- Βασίζεται σε παλαιότερη έρευνα ή σε κάποιο θεωρητικό μοντέλο που αποτελεί τη βάση της τωρινής έρευνας.
- Βασίζεται στη γνώση και τη διαίσθηση του ερευνητή που αποφασίζει για την χρησιμότητα των τυχαίων μεταβλητών που με βάση τη λογική έχουν σχέση με την κατηγοριοποίηση της εξαρτημένης τυχαίας μεταβλητής.

Η MDA είναι ευαίσθητη στην αναλογία του δείγματος με το σύνολο των ανεξάρτητων ερμηνευτικών μεταβλητών (όχι μόνο της Discriminant function). Πολλές έρευνες εισηγούνται ότι η χρήση 20πλάσιων παρατηρήσεων του αριθμού των μεταβλητών, είναι ικανοποιητική για την ανάπτυξη υποδειγμάτων MDA (Hair et al., 1998). Σε κάθε περίπτωση, η αναλογία αυτή θα πρέπει να είναι τουλάχιστον πενταπλάσια. Σαφώς, οι εκτιμήσεις που βασίζονται σε ένα μικρό δείγμα δεν είναι αποτελεσματικές. Επίσης, σημαντικό είναι και το δείγμα που απαρτίζει τις κατηγορίες. Το πλήθος των παρατηρήσεων που κατατάσσονται σε μια ομάδα δεν πρέπει να είναι μικρότερο από το πλήθος όλων των ερμηνευτικών τυχαίων μεταβλητών. Εάν το πλήθος των παρατηρήσεων της κάθε ομάδας υπερβαίνει τις 20 παρατηρήσεις, τότε πρέπει να ληφθεί υπόψη η διαφορά μεγέθους μεταξύ των κατηγοριών.

Πολλές φορές, όταν το δείγμα είναι αρκετά μεγάλο, χωρίζεται σε δύο επιμέρους δείγματα που το ένα (analysis sample) χρησιμοποιείται για την εκτίμηση των συναρτήσεων διάκρισης και το άλλο (hold out sample) για την επικύρωση των αποτελεσμάτων. Κατά το διαχωρισμό του δείγματος πρέπει να προσεχτούν οι αναλογίες μεταξύ των ομάδων στις οποίες επιδιώκεται η ταξινόμηση του πληθυσμού.

### Υποθέσεις της MDA

Η MDA βασίζεται στις πιο κάτω βασικές υποθέσεις:

- Οι ανεξάρτητες ερμηνευτικές τυχαίες μεταβλητές  $X_i$  ακολουθούν την κανονική κατανομή.
- Οι πίνακες Διακύμανσης- Συνδιακύμανσης των επιμέρους κατηγοριών είναι ίσοι.

- Οι ανεξάρτητες τυχαίες μεταβλητές έχουν μικρή συσχέτιση.

Ο ερευνητής πρέπει να εξετάσει τα δεδομένα και αν παραβιάζονται οι βασικές υποθέσεις πρέπει να συμπεριλάβει στην μελέτη του την επίδρασή τους και εφόσον κριθεί απαραίτητο θα πρέπει να επιλεγεί άλλη μέθοδος για την έρευνα. Όταν η υπόθεση της κανονικότητας των  $X_i$  παραβιάζεται τότε πιθανόν να υπάρχει πρόβλημα στην εκτίμηση των συναρτήσεων διάκρισης. Στην περίπτωση αυτή προτιμάται η μέθοδος της Λογιστικής Παλινδρόμησης. Όταν η υπόθεση της κανονικότητας ικανοποιείται αλλά οι πίνακες Διακύμανσης - Συνδιακύμανσης των ομάδων δεν είναι ίσοι τότε δεν πρέπει να χρησιμοποιηθούν γραμμικές συναρτήσεις διάκρισης αλλά μη γραμμικές συναρτήσεις 2<sup>ου</sup> βαθμού.

### **Εκτίμηση των συντελεστών $b_i$ της συνάρτησης διάκρισης**

Οι συντελεστές  $b_i$  των τυχαίων μεταβλητών  $X_i$  υπολογίζονται με την μεγιστοποίηση της διασποράς μεταξύ των ομάδων και έναντι της διασποράς μέσα στις ομάδες δηλαδή μεγιστοποιώντας την ποσότητα:

$$V = \frac{\text{Variance between groups}}{\text{Variance within groups}} \quad (3.2)$$

Ισοδύναμα, για την εκτίμηση των  $b_i$  μπορεί να μεγιστοποιηθεί ο δείκτης  $\xi$  που ορίζεται ως εξής (Huberty και Olenjnik, 2006):

$$\xi = \frac{SSH}{SSE} \quad (3.3)$$

όπου,

$$SSH = \sum_{j=1}^J \sum_{u=1}^{n_j} (X_{uj} - \bar{X}_{..})^2 \quad (3.4)$$

$$SSE = \sum_{j=1}^J \sum_{u=1}^{n_j} (X_{uj} - \bar{X}_{.j})^2 \quad (3.5)$$

όπου,

J = σύνολο των κλάσεων

$u$  = μονάδα  $u$

$n_j$  = πλήθος μονάδων στην κλάση  $j$

$X_{uj}$  = τιμή της μεταβλητής  $X$  για τη μονάδα  $u$  στην κλάση  $j$

$\bar{X}_{..}$  = μέση τιμή της μεταβλητής  $X$  για το σύνολο των μονάδων μέσα στο δείγμα

$\bar{X}_{.j}$  = μέση τιμή της μεταβλητής  $X$  στην κλάση  $j$

Σκοπός είναι να υπολογισθούν οι συντελεστές  $b_i$  των  $X_i$  της συνάρτησης διάκρισης

$$Z = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (3.7)$$

μεγιστοποιώντας την ποσότητα  $\xi$  έτσι ώστε ο διαχωρισμός μεταξύ των ομάδων να είναι ο βέλτιστος δυνατός. Για την προσέγγιση της ποσότητας  $\xi$  χρησιμοποιείται το γινόμενο των πινάκων  $E^{-1}H$

όπου,

$E$  = Error sum of squares and cross product matrix (βλ. Παράρτημα 1)

$H$  = sum of square and cross product matrix for the Hypothesis  $H_0$  (βλ. Παράρτημα 1)

Ο πίνακας  $E^{-1}H$  αποτελεί προσέγγιση του  $\xi$ . Για τον υπολογισμό των  $b_i$  πρέπει να υπολογιστεί η μέγιστη ιδιοτιμή  $\lambda_1$  του πίνακα  $E^{-1}H$  από τη σχέση :

$$|E^{-1}H - \lambda_1 I| = 0 \quad (3.8)$$

Μετά τον υπολογισμό της ιδιοτιμής  $\lambda_1$  υπολογίζεται το διάνυσμα  $b$  των συντελεστών  $b_i$  λύνοντας την εξίσωση :

$$(E^{-1}H - \lambda_1 I)b = 0 \quad (3.9)$$

### **Μέτρηση της απόστασης των ομάδων και έλεγχοι στατιστικής σημαντικότητας**

Όπως έχει προαναφερθεί για να είναι σημαντική η διάκριση που κάνουν οι συναρτήσεις διαχωρισμού πρέπει οι κεντροειδείς των κατηγοριών να έχουν αρκετή απόσταση μεταξύ τους. Πιο κάτω, παρουσιάζονται τόσο το μέτρο μέτρησης της απόστασης των κεντροειδών, όσο και οι σημαντικότεροι στατιστικοί έλεγχοι

σημαντικότητας της απόστασης αυτής, ώστε να θεωρούνται οι κεντροειδείς των ομάδων διαφορετικοί.

### Το μέτρο Mahalanobis Distance

Για την μέτρηση της απόστασης μεταξύ δύο ομάδων χρησιμοποιείται το μέτρο Mahalanobis Distance (*Huberty και Olenjnik, 2006*). Το μέτρο αυτό αποτελεί μια γενίκευση της Ευκλείδειας απόστασης και είναι κατάλληλο γιατί λαμβάνει υπόψη της τη συνδιακύμανση μεταξύ των μεταβλητών καθώς και την διαφορετική κλίμακα μέτρησής τους.

Έστω δύο ομάδες G1 και G2, στις οποίες υπάρχουν p τυχαίες μεταβλητές. Το μέτρο Mahalanobis Distance, για τις G1 και G2, δίνεται από την πιο κάτω σχέση :

$$D^2 = (CV_1 - CV_2)^T \times S_e^{-1} \times (CV_1 - CV_2) \quad (3.10)$$

όπου,

$CV_1$  και  $CV_2$  είναι τα διανύσματα των κεντροειδών των p μεταβλητών (Centroid Vectors) για κάθε κλάση,

$S_e^{-1}$  = ο αντίστροφος πίνακας του Error Covariance Matrix  $S_e$ , ο οποίος υπολογίζεται από τον άθροισμα των πινάκων sum of squares and cross products (SSCP) όλων των κλάσεων, ως εξής:

$$S_e = \frac{1}{N - J} \sum_{j=1}^J SSCP_j \quad (3.11)$$

όπου,  $SSCP_j = (n - 1) \times \text{Covariance Matrix}$  (βλ. Παράρτημα 1)

### Έλεγχοι στατιστικής σημαντικότητας για την Two Group Discriminant Analysis

#### *α) Μονομεταβλητή ανάλυση*

Όταν το υπόδειγμα είναι μονομεταβλητό και ο διαχωρισμός γίνεται σε δύο μόνο ομάδες ο έλεγχος για το αν η απόσταση μεταξύ των δύο ομάδων είναι σημαντική γίνεται με τον έλεγχο της μηδενικής υπόθεσης  $H_0: \mu_1 = \mu_2$ , όπου  $\mu_1, \mu_2$  είναι οι κεντροειδείς για την μοναδική μεταβλητή  $X_1$  για τις ομάδες 1 και 2 αντίστοιχα.

Ο έλεγχος γίνεται με τη χρήση δύο ισοδύναμων έλεγχων, του t-Student test και του F-test. Συγκεκριμένα το F-test παράγεται με ένα μετασχηματισμό από το t-Student test.

### To t-Student test

Το t-Student test γίνεται κάνοντας χρήση του στατιστικού t που ακολουθεί t-Student κατανομή με βαθμούς ελευθερίας  $df_e = n_1 + n_2 - 2 = N - 2$ .

$$t = \frac{\bar{X}_{11} - \bar{X}_{12}}{\sqrt{S_e^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

όπου  $\bar{X}_{11}, \bar{X}_{12}$  η μέση τιμή της μεταβλητής  $X_1$  για τις ομάδες 1 και 2 αντίστοιχα και όπου  $S_e^2$  είναι η error variance που δίνεται από τον τύπο:

$$S_e^2 = \frac{\sum_{j=1}^J (n_j - 1) S_j^2}{N - J} = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

όπου  $S_j^2$  είναι η διασπορά των τιμών της  $X_1$  στην ομάδα j.

Αν το στατιστικό αυτό μέγεθος είναι μεγαλύτερο από την τιμή της t-Student κατανομής απορρίπτεται η  $H_0$ .

### To F-test

Το τετράγωνο του στατιστικού t ακολουθεί την κατανομή F με βαθμούς ελευθερίας  $\nu_1 = 1$  και  $\nu_2 = N - 2$  :

$$t^2 = \frac{(\bar{X}_{11} - \bar{X}_{12})^2}{S_e^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = \frac{(\bar{X}_{11} - \bar{X}_{12})^2}{S_e^2 \left( \frac{n_1 + n_2}{n_1 n_2} \right)}$$

το οποίο μπορεί να γραφεί στην μορφή:

$$t^2 = \frac{n_1 n_2}{n_1 + n_2} \left[ (\bar{X}_{11} - \bar{X}_{12}) \times (S_e^2)^{-1} \times (\bar{X}_{11} - \bar{X}_{12}) \right]$$



Ο παράγοντας  $\left[ (\bar{X}_{11} - \bar{X}_{12}) \times (S_e^2)^{-1} \times (\bar{X}_{11} - \bar{X}_{12}) \right]$  είναι το μέτρο του Mahalanobis Distance στην μονομεταβλητή ανάλυση. Αν το στατιστικό  $t^2$  μέγεθος είναι μεγαλύτερο από την τιμή της F κατανομής απορρίπτεται η  $H_0$

### β) Πολυμεταβλητή ανάλυση

Το μέτρο του Mahalanobis βρίσκει επίσης εφαρμογή στην πολυμεταβλητή ανάλυση σε μια γενίκευση του τετραγωνισμένου Student-t στατιστικού από τον H.Hotelling για τον έλεγχο  $H_0 : \mu_1 = \mu_2$  για δύο ομάδες. Το  $\mu$  είναι ένα διάνυσμα που αποτελείται από τους κεντροειδείς όλων των  $p$  μεταβλητών για κάθε ομάδα. Για το πλήθος των μεταβλητών πρέπει να ισχύει  $p < N-1$ .

Το στατιστικό Hotelling δίνεται από τον τύπο :

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} [(CV_1 - CV_2)^T \times S_e^{-1} \times (CV_1 - CV_2)]$$

Ο έλεγχος  $H_0 : \mu_1 = \mu_2$  γίνεται με τη χρήση ενός στατιστικού F που παράγεται με έναν μετασχηματισμό από το  $T^2$ .

### Το F-test

Με τον παρακάτω μετασχηματισμό παράγεται το στατιστικό μέγεθος που ακολουθεί την κατανομή F με βαθμούς ελευθερίας  $\nu_1 = p$  και  $\nu_2 = df_e - p + 1 = N - p - 1$ .

$$F = \frac{df_e - p + 1}{p * df_e} T^2$$

όπου,

$df_e$  είναι το Error degrees of freedom = N-J

$p$  είναι το πλήθος των τυχαίων μεταβλητών  $X_i$  και

$T^2$  είναι το στατιστικό του Hotelling.

Αν το στατιστικό αυτό μέγεθος είναι μεγαλύτερο από την τιμή της F κατανομής απορρίπτεται η  $H_0$ , δηλαδή θεωρείται ότι τα διανύσματα των κεντροειδών δεν είναι ίσα.



## Έλεγχοι στατιστικής σημαντικότητας για την Multiple Group Discriminant Analysis

### α) Μονομεταβλητή ανάλυση

Όταν έχουμε μονομεταβλητό υπόδειγμα τότε ο έλεγχος  $H_0: \mu_1 = \mu_2 = \dots = \mu_j$  γίνεται με το στατιστικό μέγεθος  $F$ ,

$$F = \frac{MSH}{MSE}$$

το οποίο ακολουθεί την κατανομή  $F$  με βαθμούς ελευθερίας  $v_1 = df_h = N - 1$ ,  
 $v_2 = df_e = N - J$

$MSH$  είναι το Hypothesis Mean Square και δίνεται από τον τύπο

$$MSH = \frac{\sum_{j=1}^J n_j (\bar{X}_j - \bar{X})^2}{J - 1}$$

$MSE$  είναι το Mean Square Error

$$MSE = \frac{\sum_{j=1}^J \sum_{u=1}^{n_j} (X_{uj} - \bar{X}_j)^2}{N - J} = \frac{SSE}{df_e}$$

όπου,

$X_{uj}$  είναι η τιμή της τυχαίας μεταβλητής  $X$  της μονάδας  $u$  στην ομάδα  $j$

$\bar{X}_j$  είναι η μέση τιμή της  $X$  στην ομάδα  $j$  και δίνεται από τον τύπο:  $\bar{X}_j = \sum_{u=1}^{n_j} \frac{X_{uj}}{n_j}$

$\bar{X}$  είναι η μέση τιμή της  $X$  από όλες τις ομάδες και δίνεται από τον τύπο:  $\bar{X} =$

$$\sum_{j=1}^J \sum_{u=1}^{n_j} \frac{X_{uj}}{N}$$

$SSE$  είναι το Error Sum of Squares

$SSH$  είναι το Hypothesis Sum of Squares

$df_e$  είναι το Error degrees of freedom (=  $N - J$ )

### β) Πολυμεταβλητή ανάλυση

Στην πολυμεταβλητή ο έλεγχος  $H_0: \mu_1 = \mu_2 = \dots = \mu_j$  μπορεί να γίνει με τη χρήση αρκετών κριτηρίων τα περισσότερα παραθέτονται παρακάτω

### 1. Wilks Lambda F test

Το πιο παλιό και το πιο ευρέως διαδεδομένο κριτήριο για τον έλεγχο της υπόθεσης  $H_0: \mu_1 = \mu_2 = \dots = \mu_j$  είναι αυτό του *Wilks S.S. (1932)* το οποίο ονομάζεται και Wilks Lambda. Βασίζεται στο στατιστικό  $\Lambda$  :

$$\Lambda = \frac{|E|}{|H + E|}$$

όπου,

$E$  = Error sum of squares and cross product matrix

$H$  = Sum of square and cross product matrix for the Hypothesis  $H_0$

Το κριτήριο  $\Lambda$ , για  $p$  ανεξάρτητες τυχαίες μεταβλητές και για βαθμούς ελευθερίας  $df_h = J - 1$  (που καθορίζονται από την υπόθεση ελέγχου) μετασχηματίζεται σε ένα κριτήριο που ακολουθεί την κατανομή  $F$ .

Όταν έχουμε δύο κλάσεις, τότε  $df_h = 2 - 1 = 1$  και το στατιστικό αυτό παίρνει την ακόλουθη μορφή :

$$F = \frac{1 - \Lambda}{\Lambda} \times \frac{df_e - p + 1}{p}$$

Σε αυτή την περίπτωση, ακολουθεί την κατανομή  $F$  με βαθμούς ελευθερίας  $v_1 = p$  και  $v_2 = df_e - p + 1$ .

Το κριτήριο προϋποθέτει ότι οι μονάδες του δείγματος είναι ανεξάρτητες, ότι ισχύει η κανονικότητα και η ισότητα(ομοιογένεια) των Πινάκων Διακύμανσης.

Εναλλακτικά, το  $\Lambda$  μπορεί να υπολογισθεί με τη χρήση των ιδιοτιμών του πίνακα  $E^{-1}H$ . Έτσι λύνοντας την εξίσωση  $|E^{-1}H - \lambda I| = 0$  γίνεται ο υπολογισμός των ιδιοτιμών  $\lambda$  του πίνακα  $E^{-1}H$ , οπότε το  $\Lambda$  του Wilks ισούται με :

$$\Lambda = \prod_{u=1}^r \left( \frac{1}{1 + \lambda_u} \right)$$

όπου,  $r = \min(p, df_h)$

Ο έλεγχος αυτός δείχνει ποιές ανεξάρτητες μεταβλητές  $X_i$  συνεισφέρουν περισσότερο στον διαχωρισμό των κλάσεων. Το  $\Lambda$  παίρνει τιμές στο  $[0,1]$  όπου στην τιμή 1 οι κεντροειδείς ταυτίζονται. Αντίθετα, στην τιμή μηδέν υπάρχει πλήρης διαχωρισμός. Όσο μικρότερο λοιπόν είναι το  $\Lambda$  της τυχαίας μεταβλητής  $X_i$  τόσο πιο σημαντική προσφορά έχει στον διαχωρισμό.

## 2. Bartlett-Pillai κριτήριο

Ένα διαφορετικό κριτήριο για να ελέγξουμε την υπόθεση  $H_0: \mu_1 = \mu_2 = \dots = \mu_j$  είναι αυτό των M.S. Barlett και K.C.S. Pillai. Το κριτήριο αυτό όπως και του Wilks βασίζεται στις ιδιοτιμές του πίνακα  $E^{-1}H$ .

Το στατιστικό του Bartlett-Pillai ορίζεται ως εξής :

$$U = \sum_{u=1}^r \left( \frac{\lambda_u}{1 + \lambda_u} \right)$$

όπου,  $r = \min(p, df_h)$

Αντίστοιχα και στην περίπτωση αυτή υπάρχει ένας μετασχηματισμός που παράγει ένα στατιστικό που ακολουθεί κατανομή  $F$  με βαθμούς ελευθερίας  $\nu_1 = b \times r$  και  $\nu_2 = r(df_e - p + r)$  όπου  $b = \max(p, df_h)$  :

$$F = \frac{U}{r - U} \times \frac{df_e - p + r}{b}$$

## Αποτίμηση της αποτελεσματικότητας του υποδείγματος

Μετά από την κατασκευή των συναρτήσεων διάκρισης το επόμενο βήμα είναι ο υπολογισμός όλων των Z scores για κάθε μια μονάδα (παρατήρηση) μέσα στο δείγμα και για κάθε συνάρτηση διαχωρισμού. Έτσι, γίνεται μια κατάταξη των μονάδων του δείγματος με βάση κάποια επιμέρους χαρακτηριστικά τους που εκφράζονται από τις ανεξάρτητες τυχαίες μεταβλητές που έχουν επιλεγθεί. Οι παραπλήσιες τιμές του Z scores πρέπει να εκφράζουν παρατηρήσεις με παρόμοια χαρακτηριστικά.

Για την εφαρμογή της MDA υπολογίζονται τα Z scores όλων των δανειοληπτών για την μοναδική συνάρτηση διάκρισης, εφόσον ο διαχωρισμός θα γίνει σε δύο κλάσεις: τους συνεπείς και τους ασυνεπείς. Με αυτόν τον τρόπο γίνεται σύγκριση της συμπεριφοράς των δανειοληπτών. Αναμένεται δανειολήπτες με παρόμοιο Z score να έχουν παρόμοια συμπεριφορά.

Όπως έχει προαναφερθεί, στόχος της MDA είναι η κατασκευή συναρτήσεων διάκρισης που «αναγνωρίζουν» τα διαφορετικά χαρακτηριστικά των παρατηρήσεων του δείγματος τα οποία ουσιαστικά αποτελούν τη βάση για την σωστή εκτίμηση της συμπεριφοράς μιας νέας μονάδας και την κατάταξή της σε μια κατηγορία.

Πρέπει να αποδειχθεί ότι το υπόδειγμα είναι ικανό να κάνει σωστή διάκριση. Ένας τρόπος για να εκτιμηθεί η ικανότητα του μοντέλου είναι ο προσδιορισμός της συμπεριφοράς ανάμεσα στα μέλη διαφορετικών κλάσεων. Ο έλεγχος αυτός επιτυγχάνεται με την σύγκριση των κεντροειδών των κλάσεων μέσω του μεγέθους Mahalanobis Distance, που στη συνέχεια θα εκτιμηθεί αν η απόσταση είναι στατιστικά σημαντική. Μια καλή εικόνα θα έδινε και η γραφική απεικόνιση των κεντροειδών.

Η αποτελεσματικότητα του υποδείγματος σχετίζεται και με το κατά πόσο έχει κατατάξει τις μονάδες του δείγματος στις σωστές κλάσεις με βάση το Z score. Οπότε, πρέπει να ελεγχθεί κατά πόσο ο διαχωρισμός που επιτυγχάνεται από τις συναρτήσεις διάκρισης είναι καλός.

Γι' αυτό το λόγο κατασκευάζεται ο πίνακας ταξινόμησης (classification matrix), ο οποίος έχει και την εκτιμημένη κατάταξη και την πραγματική. Στην κύρια διαγώνιο του πίνακα υπάρχει ο αριθμός των σωστά τοποθετημένων παρατηρήσεων ενώ στην δευτερεύουσα διαγώνιο το πλήθος των μη σωστά τοποθετημένων. Από τον πίνακα αυτόν υπολογίζεται το Hit Ratio το οποίο είναι το ποσοστό των σωστά τοποθετημένων μονάδων. Το Hit Ratio δείχνει πόσο καλά διαχωρίζουν το δείγμα οι συναρτήσεις διάκρισης.

Παράδειγμα πίνακα ταξινόμησης για τον πιστωτικό κίνδυνο:

Πχ, μια τράπεζα χωρίζει δείγμα 50 δανειοληπτών σε συνεπείς (Ομάδα 0) και ασυνεπείς (Ομάδα 1), με 25 άτομα σε κάθε ομάδα. Έστω, ότι η κατάταξη που έκανε η

συνάρτηση διάκρισης διαφέρει από την πραγματικότητα, όπως φαίνεται στον παρακάτω πίνακα ταξινόμησης :

Actual group	Predicted group		Actual group size
	1	0	
1	22	3	25
0	5	20	25
Predicted group size	27	23	50

Για την ομάδα 1, οι σωστά τοποθετημένες παρατηρήσεις είναι 22 (88%) και για την ομάδα 0 είναι 20 (80%). Έτσι συνολικά από τους 50 πελάτες της T, η συνάρτηση διάκρισης τοποθέτησε σωστά τους 42. Σε αυτή την περίπτωση, η τιμή του Hit Ratio είναι 84% ( $42/50=84\%$ ).

Όταν το πλήθος των παρατηρήσεων το επιτρέπει, ο ερευνητής χωρίζει το δείγμα σε δύο μέρη. Το ένα από αυτά (analysis sample) το χρησιμοποιεί για την κατασκευή των συναρτήσεων διάκρισης και το άλλο (holdout sample) για τη επικύρωσή τους.

### Υπολογισμός του Cutting Score

Όπως έχει προαναφερθεί, τα cutting scores αποτελούν σημεία πρόκρισης ή απόρριψης τα οποία συγκρινόμενα με το Z score μίας μονάδος από το δείγμα κατατάσσουν τη μονάδα στην κατάλληλη κλάση. Τα cutting scores εξαρτώνται από το μέγεθος της κάθε κλάσης καθώς επίσης και από το πόσο διαφέρουν οι κλάσεις μεταξύ τους.

Οι σχέσεις με βάση τις οποίες υπολογίζονται για δύο ομάδες πχ A και B, είναι οι εξής:

(α) Όταν τα A και B έχουν ίσα μεγέθη:

$$Z_{cs} = \frac{Z_A + Z_B}{2} \quad (3.12)$$

όπου,

$Z_{cs}$  = Cutting score

$Z_A$  = κεντροειδής της ομάδας A

$Z_B$  = κεντροειδής της ομάδας B

Όταν τα A και B δεν έχουν ίσα μεγέθη και ο ερευνητής εκτιμά ότι αντικατοπτρίζουν την αναλογία του πληθυσμού :

$$Z_{cs} = \frac{N_A Z_A + N_B Z_B}{N_A + N_B} \quad (3.13)$$

όπου,

$Z_{cs}$  = Cutting score

$Z_A$  = κεντροειδής της ομάδας A

$Z_B$  = κεντροειδής της ομάδας B

$N_A$  = μέγεθος της ομάδας A

$N_B$  = μέγεθος της ομάδας B

Υπάρχει όμως και το ενδεχόμενο τα μεγέθη των ομάδων να μην είναι ίσα άλλα να κριθεί απαραίτητο ότι πρέπει να θεωρηθούν ίσα ώστε να αντιπροσωπεύουν την πραγματική αναλογία τους στον πληθυσμό. Ουσιαστικά, με την διαδικασία αυτή, καθορίζεται η a priori πιθανότητα να ανήκει κάποιος σε μια ομάδα.

### **Κόστος λανθασμένης κατάταξης (Cost of Misclassification)**

Η λανθασμένη ταξινόμηση των παρατηρήσεων σε κατηγορίες επιφέρει κάποιο κόστος το οποίο εκφράζεται από τα σφάλματα τύπου I και σφάλμα τύπου II. Στην περίπτωση του πιστωτικού κινδύνου που αντιμετωπίζουν οι τράπεζες, το σφάλμα τύπου I είναι η τοποθέτηση ενός μη συνεπούς δανειολήπτη στην κατηγορία των συνεπών. Το σφάλμα αυτό σημαίνει ότι μπορεί να δοθεί έγκριση για ένα επισφαλές δάνειο και επομένως να θέσει σε κίνδυνο το συνολικό ποσό της αντίστοιχης χορήγησης.

Από την άλλη πλευρά, το σφάλμα τύπου II σημαίνει την κατάταξη ενός συνεπούς πελάτη στην κατηγορία των μη συνεπών. Το σφάλμα αυτό πιθανόν να έχει ως αποτέλεσμα διαφυγόντα κέρδη για την τράπεζα.

Το σφάλμα τύπου I είναι πιο σημαντικό από αυτό του τύπου II αφού επιφέρει ζημία και πρέπει να δίνεται περισσότερη προσοχή για την αποφυγή του. Έτσι, για την εκτίμηση του Cutting Score λαμβάνεται υπόψη η ελαχιστοποίηση των δύο σφαλμάτων δίνοντας ιδιαίτερη βαρύτητα στην ελαχιστοποίηση του σφάλματος τύπου I.

### **Μέτρηση της ακρίβειας των προβλέψεων (Predictive accuracy)**

Ένας τρόπος όπως έχει προαναφερθεί για την μέτρηση της ικανότητας πρόβλεψης είναι το Hit Ratio το οποίο υπολογίζεται από τον πίνακα ταξινόμησης. Το ερώτημα που προκύπτει είναι από ποιά τιμή του Hit Ratio και πάνω είναι αποδεκτή η προβλεπτική ικανότητα του μοντέλου. Η απάντηση βασίζεται στην μελέτη των μεγεθών των κλάσεων καθώς και των πιθανοτήτων μια παρατήρηση  $i$  να ανήκει σε μια κατηγορία  $j$ . Όταν υπάρχουν  $n$  ομάδες, οι οποίες έχουν το ίδιο μέγεθος, αποδεκτά γίνονται τα Hit Ratio που είναι μεγαλύτερα από  $1/n$ , δηλαδή για 2 ομάδες πρέπει Hit Ratio  $>50\%$ , για 3 πρέπει Hit Ratio  $>33\%$  κλπ. Στην περίπτωση που οι ομάδες δεν έχουν ίσα μεγέθη γίνεται η χρήση των δύο κριτηρίων :

#### α) Maximum Chance criterion

Όταν οι ομάδες δεν είναι ισομεγέθεις υπολογίζεται με τί πιθανότητα κατατάσσεται σε κάθε ομάδα μια παρατήρηση και επιλέγεται η μεγαλύτερη πιθανότητα. Έτσι όταν υπάρχουν δύο ομάδες που η μία έχει 30 παρατηρήσεις και η άλλη 70, τότε το Hit Ratio πρέπει να είναι πάνω από 70% για να είναι αποδεκτό. Το κριτήριο αυτό χρησιμοποιείται όταν ο μοναδικός στόχος είναι η μεγιστοποίηση των ποσοστού των σωστών προβλέψεων.

#### β) Proportional Chance criterion

Αυτό το κριτήριο εφαρμόζεται όταν επιθυμείται η μεγιστοποίηση των σωστών ταξινομήσεων από όλες τις ομάδες. Έτσι για δύο ομάδες ο τύπος σύμφωνα με τον οποίο υπολογίζεται η ελάχιστη τιμή του Hit Ratio είναι :

$$C = p^2 + (1 - p)^2 \quad (3.14)$$

όπου,

$p$  = ποσοστιαία αναλογία του μεγέθους της πρώτης κλάσης ως προς τον πληθυσμό

$1 - p$  = ποσοστιαία αναλογία του μεγέθους της δεύτερης κλάσης ως προς τον πληθυσμό

Στο ίδιο παράδειγμα, για ομάδες με μεγέθη 30 και 70 το αποτέλεσμα είναι  $C = 0.7^2 + 0.3^2 = 58\%$  έναντι 70% με το Maximum Chance criterion. Hit Ratio με τιμή μικρότερη του 58% δεν γίνονται αποδεκτά.

Τα κριτήρια αυτά είναι χρήσιμα όταν εφαρμόζονται στο δείγμα επικύρωσης (holdout sample), όταν ο πληθυσμός χωρίζεται σε δύο δείγματα.

Για να είναι σημαντική η προβλεπτική ικανότητα του μοντέλου πρέπει το Hit Ratio να υπερβαίνει την κατώτατη αποδεκτή τιμή τουλάχιστον κατά το  $\frac{1}{4}$  αυτής. Δηλαδή, αν το κατώτατο όριο είναι 50% πρέπει για να είναι σημαντική η προβλεπτική ικανότητα το Hit Ratio να είναι τουλάχιστον  $0.5 \times 1.25 = 62.5\%$ . Το κριτήριο αυτό αποτελεί μια χοντρική προσέγγιση του επιπέδου της προβλεπτικής ικανότητας.

Από την άλλη πλευρά, υπάρχει ένας έλεγχος της ικανότητας διάκρισης του μοντέλου το οποίο ονομάζεται Press's Q statistic και βασίζεται στο στατιστικό :

$$Press's Q = \frac{[N - (nK)]^2}{N(K - 1)} \quad (3.15)$$

όπου,

$N$  = συνολικό μέγεθος του δείγματος

$n$  = πλήθος των σωστά τοποθετημένων παρατηρήσεων

$K$  = πλήθος των κατηγοριών

Η τιμή του πιο πάνω κριτηρίου συγκρίνεται με την κριτική τιμή από την  $X^2$  κατανομή με ένα βαθμό ελευθερίας και το επιθυμητό επίπεδο εμπιστοσύνης. Όταν  $Press's Q >$  κριτική τιμή της  $X^2$  το μοντέλο είναι σημαντικό ως προς την ικανότητα διάκρισης. Το κριτήριο αυτό λειτουργεί καλύτερα σε μεγάλο δείγμα.

Τέλος, ουσιαστικής σημασίας είναι η μελέτη των παρατηρήσεων μία προς μία και ο εντοπισμός των χαρακτηριστικών που συνεισφέρουν περισσότερο και λιγότερο στη διάκριση των ομάδων. Έτσι, δημιουργείται ένα προφίλ για κάθε ομάδα και αναμένεται κάποια χαρακτηριστικά να έχουν ένα συγκεκριμένο εύρος τιμών για τις



παρατηρήσεις που ανήκουν σε συγκεκριμένη ομάδα. Υπάρχουν βέβαια και εξαιρέσεις που η συνάρτηση διαχωρισμού δεν έχει τοποθετήσει κάποιες παρατηρήσεις στη σωστή ομάδα.

Σημαντικό είναι να εντοπισθούν τα χαρακτηριστικά που οδήγησαν σε λανθασμένη τοποθέτηση. Η ερμηνεία της μη αναμενόμενης συμπεριφοράς του μοντέλου οδηγεί στην βελτίωσή του.

## 2.2 Υπόδειγμα Λογιστικής Παλινδρόμησης-Logit

Το υπόδειγμα LOGIT ανήκει στην οικογένεια των γενικευμένων γραμμικών μοντέλων εκτίμησης πιθανοτήτων. Το υπόδειγμα αυτό κατηγοριοποιεί τις μονάδες ενός πληθυσμού σε κλάσεις με βάση τα ομοιογενή επιμέρους χαρακτηριστικά τους, υπό την προϋπόθεση ότι αυτά εκφράζονται με ανεξάρτητες τυχαίες μεταβλητές. Πλεονέκτημα του υποδείγματος είναι ότι προσεγγίζει αριθμητικά την πιθανότητα να ανήκει μια μονάδα σε μια συγκεκριμένη ομάδα.

Στην μέτρηση του πιστωτικού κινδύνου το υπόδειγμα βρίσκει εφαρμογή στην εκτίμηση της πιθανότητας αθέτησης (Probability of Default ή PD) που εκφράζει πχ το ενδεχόμενο μη ανταπόκρισης δανειοληπτών στις υποχρεώσεις τους έναντι των τραπεζών που τους δανειοδότησαν, καθώς και στον διαχωρισμό αυτών σε συνεπείς και ασυνεπείς.

Το υπόδειγμα LOGIT βασίζεται στην εξής λογική (*Bierens, 2008*):

Η τυχαία μεταβλητή  $Y_i$  εκφράζει την συμπεριφορά ενός τυχαίου δανειολήπτη ως προς την συνέπειά του στην αποπληρωμή του δανείου του και παίρνει δύο τιμές :

$$Y_i = \begin{cases} 1, & \text{συνεπής} \\ 0, & \text{ασυνεπής} \end{cases}$$

Η κατανομή της  $Y_i$  είναι *Bernouli* και παίρνει τις τιμές 1 και 0 με πιθανότητα  $P(Y_i = 1)$  και  $1 - P(Y_i = 1)$  αντίστοιχα. Η συνάρτηση πυκνότητας πιθανότητας μπορεί να γραφεί στη μορφή :

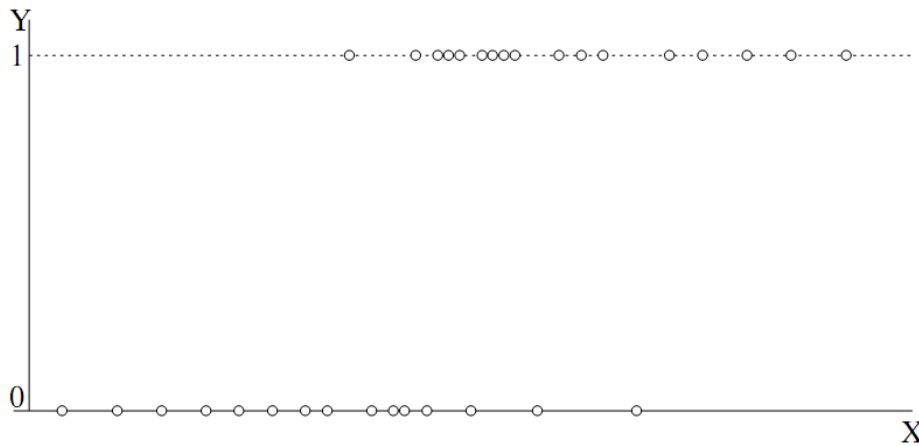
$$\Pr(Y_i = y_i) = P^{y_i}(1 - P)^{(1-y_i)} \quad (3.16)$$

με μέση τιμή και διασπορά :

$$E(Y_i) = \mu_i = P_i$$

$$Var(Y_i) = \sigma_i^2 = P_i(1 - P_i)$$

Παρατηρείται ότι και η μέση τιμή και η διασπορά εξαρτώνται από την πιθανότητα αθέτησης.



Γράφημα 3.3 (*J. S. Cramer, 2003*)

Στο γράφημα 3.3 παρουσιάζονται τιμές της μεταβλητής  $Y_i$  η οποία εξαρτάται από μία τυχαία μεταβλητή  $X$ .

Στις περισσότερες περιπτώσεις η ανάλυση είναι πολυμεταβλητή, δηλαδή η συμπεριφορά του δανειολήπτη ως προς την αποπληρωμή του δανείου επηρεάζεται από ένα πλήθος ανεξάρτητων τυχαίων μεταβλητών  $X_i$ . Το υπόδειγμα LOGIT βασίζεται κατ' αρχήν στην σωστή επιλογή των  $X_i$ , καθώς αποτελούν την διαθέσιμη πληροφορία βάση της οποίας θα γίνει ο διαχωρισμός του των δανειοληπτών σε συνεπείς και ασυνεπείς. Επίσης, η εκτίμηση της πιθανότητας αθέτησης βασίζεται στις ανεξάρτητες τυχαίες μεταβλητές  $X_i$ .

Ειδικότερα, έστω ότι  $N$  το πλήθος ανεξάρτητες τυχαίες μεταβλητές  $X_i$  επηρεάζουν την τιμή της  $Y_i$ . Στο υπόδειγμα LOGIT όπως προαναφέρθηκε η πιθανότητα αθέτησης υπολογίζεται με βάση τις τιμές των ερμηνευτικών μεταβλητών  $X_i$ .

Στην πράξη υπάρχει μια τυχαία μεταβλητή  $S_i$  η οποία δεν είναι παρατηρήσιμη και συνδέεται γραμμικά με τις ανεξάρτητες τυχαίες μεταβλητές  $X_i$  με την εξής σχέση (Liao, 1994):

$$S_i = \sum_{k=1}^N b_{ik}X_{ik} + \varepsilon_i \quad (3.17)$$

Η τυχαία μεταβλητή  $S_i$  αποτελεί μια ποσότητα (SCORE) που υπολογίζεται για τον κάθε δανειολήπτη. Το κατάλοιπο αυτής της γραμμικής σχέσης  $\varepsilon_i$  κατανέμεται συμμετρικά και έχει μέση τιμή μηδέν. Η κατανομή που ακολουθεί παίζει πολύ σημαντικό ρόλο, καθώς η Συνάρτηση Κατανομής του  $\varepsilon_i$ ,  $F(\varepsilon)$  είναι καταλυτικής σημασίας στην εκτίμηση της πιθανότητας αθέτησης.

Η μεταβλητή που παρατηρείται στην πράξη είναι η  $Y_i$ , η οποία είναι ψευδομεταβλητή (dummy variable). Οι δύο μεταβλητές  $Y_i$  και  $S_i$  συνδέονται με την παρακάτω σχέση :

$$Y_i = \begin{cases} 1, & \text{αν } S_i > 0 \text{ (Default)} \\ 0, & \text{διαφορετικά} \end{cases} \quad (3.18)$$

Υπάρχει μία τιμή  $c$  την οποία αν υπερβεί το SCORE του δανειολήπτη τότε θεωρείται ότι αυτός δεν θα ανταποκριθεί στις υποχρεώσεις του. Το  $c$  ονομάζεται cutting score και αποτελεί το σημείο πρόκρισης ή απόρριψης.

Η αναμενόμενη τιμή του score  $S_i$  για τον  $i$ -οστο δανειολήπτη είναι :

$$E(S_i|X_1 \dots X_N) = \sum_{k=1}^N b_{ik}X_{ik} \quad (3.19)$$

Ο υπολογισμός της πιθανότητα αθέτησης μέσω της Συνάρτησης Κατανομής (CDF) του  $\varepsilon_i$ ,  $F(\varepsilon)$  έχει ως εξής (Liao, 1994) :

$$\begin{aligned} PD &= \Pr(Y_i = 1) = \Pr\left(\sum_{k=1}^N b_{ik}X_{ik} + \varepsilon_i > 0\right) \\ &= \Pr\left(\varepsilon_i > -\sum_{k=1}^N b_{ik}X_{ik}\right) \Leftrightarrow \end{aligned}$$

$$PD = 1 - F\left(-\sum_{k=1}^N b_{ik}X_{ik}\right) \quad (3.20)$$

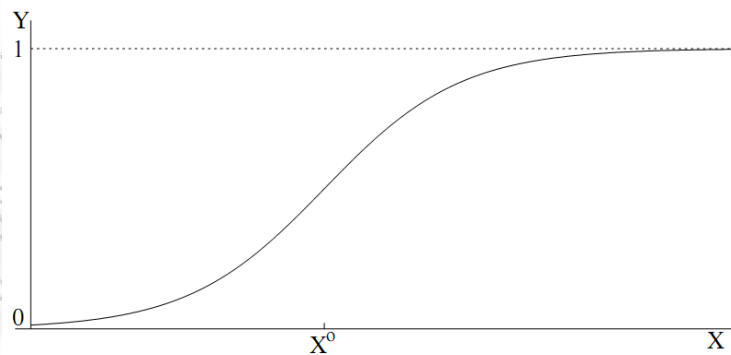
Παρατηρείται ότι το PD είναι συνάρτηση της αναμενόμενης τιμής  $E(S_i|X_1 \dots X_N)$ . Θα πρέπει να σημειωθεί στην περίπτωση που χρησιμοποιηθεί και σταθερά  $b_{i0}$  το άθροισμα ξεκινάει από το μηδέν με  $X_{i0} = 1$  και η πιθανότητα αθέτησης μπορεί να πάρει τη μορφή :

$$PD = \Pr(Y_i = 1) = 1 - F\left(-\sum_{k=0}^N b_{ik}X_{ik}\right) \quad (3.21)$$

Επειδή το PD είναι πιθανότητα, πρέπει να παίρνει τιμές αυστηρά μέσα στο διάστημα  $[0,1]$ . Μια συνάρτηση που έχει την ιδιότητα να παίρνει τιμές από το σύνολο των πραγματικών αριθμών και να επιστρέφει τιμές που ανήκουν στο  $[0,1]$  είναι η λογιστική συνάρτηση με τύπο:

$$\Phi(x) = \frac{1}{1 + e^{-x}} \quad (3.22)$$

Η συνάρτηση αυτή είναι μια σιγμοειδής καμπύλη όπως φαίνεται στο γράφημα 3.4 που για μικρές τιμές του  $x$  η τιμή της συνάρτησης  $\Phi(x)$  τείνει στο 0 ενώ για μεγάλες τείνει στο 1.



Γράφημα 3.4 (Cramer, 2003)

Το υπόδειγμα LOGIT χρησιμοποιεί την λογιστική συνάρτηση για να αντιστοιχίσει σε κάθε αναμενόμενο score μια πιθανότητα (Hair et al., 1998).

Υποθέτοντας ότι  $\varepsilon_i$  ακολουθεί τη λογιστική κατανομή (Liao, 1994) με συνάρτηση κατανομής  $F(x) = \frac{1}{1+e^{-x}}$  τότε :

$$PD = \Pr(Y_i = 1) = 1 - \frac{1}{1 + e^{\sum_{k=0}^N b_{ik}X_{ik}}} = \frac{e^{\sum_{k=0}^N b_{ik}X_{ik}}}{1 + e^{\sum_{k=0}^N b_{ik}X_{ik}}} \Leftrightarrow$$

$$PD = \frac{1}{1 + e^{-\sum_{k=0}^N b_{ik}X_{ik}}} \quad (3.23)$$

Επίσης,

$$\Pr(Y_i = 0) = 1 - \Pr(Y_i = 1) = \frac{1}{1 + e^{\sum_{k=0}^N b_{ik}X_{ik}}} \quad (3.24)$$

Όσο μεγαλύτερη είναι η τιμή του score ενός δανειολήπτη τόσο μικρότερη πιθανότητα αθέτησης έχει.

Η συμπεριφορά του PD με βάση τις τιμές του score  $= \sum_{k=0}^N \hat{b}_k X_{ik}$  είναι:

$$\begin{cases} PD \rightarrow 1 \text{ καθώς } score \rightarrow \infty \\ PD = 1/2 \text{ καθώς } score = 0 \\ PD \rightarrow 0 \text{ καθώς } score \rightarrow -\infty \end{cases}$$

Μια διαφορετική προσέγγιση είναι η έκφραση της πιθανότητας αθέτησης  $\Pr(Y_i = 1)$  ως συνάρτηση των ανεξάρτητων τυχαίας μεταβλητής  $X_i$  μέσω της γραμμικής σχέσης  $\Pr(Y_i = 1) = \sum_{k=0}^N b_{ik}X_{ik}$ . Στην περίπτωση αυτή υπάρχει πρόβλημα γιατί το αριστερό μέλος παίρνει τιμές στο διάστημα  $[0,1]$ , ενώ το δεξιό μπορεί να πάρει τιμές πέραν του διαστήματος  $[0,1]$  οπότε πρέπει να τεθούν κάποιοι περιορισμοί.

Για να αποφευχθούν οι περιορισμοί αυτοί γίνεται χρήση ενός μετασχηματισμού ο οποίος μετασχηματίζει την πιθανότητα  $P$  σε *odds*. Τα *odds* εκφράζουν την πιθανότητα αθέτησης σε σχέση με τη συμπληρωματική της και δίνονται από τον τύπο:

$$odds(P) = \frac{P}{1 - P} \quad (3.25)$$

Από τη λογαρίθμηση της σχέσης αυτής, προκύπτει :

$$logit(\Pr(Y_i = 1)) = \log(odds(P)) = \log \frac{P}{1-P} \quad (3.26)$$

Το κέρδος από αυτόν το μετασχηματισμό είναι ότι με τα *logits* έχει επιτευχθεί μια αντιστοιχία ανάμεσα στο σύνολο των πραγματικών αριθμών και στο διάστημα  $[0,1]$ , αφού  $P \in [0,1]$  και  $\text{logit} \in (-\infty, \infty)$ .

$$\text{logit}(\text{Pr}(Y_i = 1)) \begin{cases} < 0, & \text{οταν } P < 1/2 \\ = 0, & \text{οταν } P = 1/2 \\ > 0, & \text{οταν } P > 1/2 \end{cases}$$

Αντικαθιστώντας την πιθανότητα αθέτησης από τη σχέση (3.23) στη σχέση (3.26) προκύπτει η γραμμική σχέση (Cramer, 2003):

$$\text{logit}(P(Y_i = 1)) = \sum_{k=0}^N b_{ik} X_{ik} \quad (3.27)$$

Οι σχέσεις (3.23) και (3.27) είναι ισοδύναμες και ουσιαστικά αυτές έχουν δώσει στο υπόδειγμα τα δύο ονόματα Λογιστική Παλινδρόμηση και Logit. Από την πρώτη προκύπτει το όνομα Λογιστική Παλινδρόμηση και από τη δεύτερη το Logit. Η εκτίμηση των συντελεστών  $b_i$  πραγματοποιείται με την μέθοδο της Μεγίστης Πιθανοφάνειας, επειδή η τυχαία μεταβλητή  $Y_i$  παίρνει μόνο δυο τιμές, 0 και 1 (Liao, 1994).

### Συνάρτηση Μεγίστης Πιθανοφάνειας (Maximum Likelihood Estimation)

Έστω ένα χαρτοφυλάκιο δανείων με ένα τυχαίο αριθμό  $m$  δανειοληπτών για το οποίο έχει γίνει η επιλογή των ερμηνευτικών μεταβλητών  $X_i$  που επηρεάζουν τη συμπεριφορά των δανειοληπτών. Η συμπεριφορά αυτή είναι γνωστή και δίνεται από την τιμή της τυχαίας μεταβλητής  $Y_i$  η οποία παίρνει δύο τιμές, 0 ή 1. Έτσι, για κάθε δανειολήπτη υπάρχουν οι τιμές των τυχαίων μεταβλητών  $Y_i, X_{i,1}, \dots, X_{i,n}$ . Επίσης, έχει αποδειχθεί ότι για έναν δανειολήπτη ισχύει:

$$PD = \text{Pr}(Y_i = 1) = \frac{1}{1 + e^{-\sum_{k=0}^N b_{ik} X_{ik}}} = \Lambda(b^T X) \quad (3.28)$$

$$\text{Pr}(Y_i = 0) = 1 - \text{Pr}(Y_i = 1) = \frac{1}{1 + e^{\sum_{k=0}^N b_{ik} X_{ik}}} = 1 - \Lambda(b^T X) \quad (3.29)$$

Ανάλογα με την (3.16) προκύπτει η σχέση:

$$\text{Pr}(Y_i = y_i) = L_i = \Lambda(b^T X)^{y_i} (1 - \Lambda(b^T X))^{1-y_i} \quad (3.30)$$



όπου,

$b^T$  είναι το ανάστροφο διάνυσμα των συντελεστών  $b_i$

$X$  είναι το διάνυσμα των τιμών των τυχαίων μεταβλητών  $X_i$

Επομένως, για το σύνολο των δανειοληπτών θα ισχύει η πιο κάτω συνάρτηση πιθανοφάνειας, υπό την προϋπόθεση ότι τα Defaults είναι ανεξάρτητα:

$$L = \prod_{i=1}^m L_i = \prod_{i=1}^m \Lambda(b^T X)^{y_i} (1 - \Lambda(b^T X))^{1-y_i} \quad (3.31)$$

Η λογαρίθμηση της (3.31), οδηγεί στη σχέση :

$$\ln L = \sum_{i=1}^m y_i \ln(\Lambda(b^T X)) + (1 - y_i) \ln(1 - \Lambda(b^T X)) \quad (3.32)$$

Το ζητούμενο είναι να υπολογισθεί η τιμή των συντελεστών των  $X_i$ , δηλαδή το διάνυσμα  $b$  που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας (*Loffler and Posch, 2007*). Στη συνέχεια, θα πρέπει να ελεγχθεί κατά πόσο οι υπολογισθείσες τιμές  $\hat{b}$  του διανύσματος  $b$  είναι στατιστικά σημαντικές.

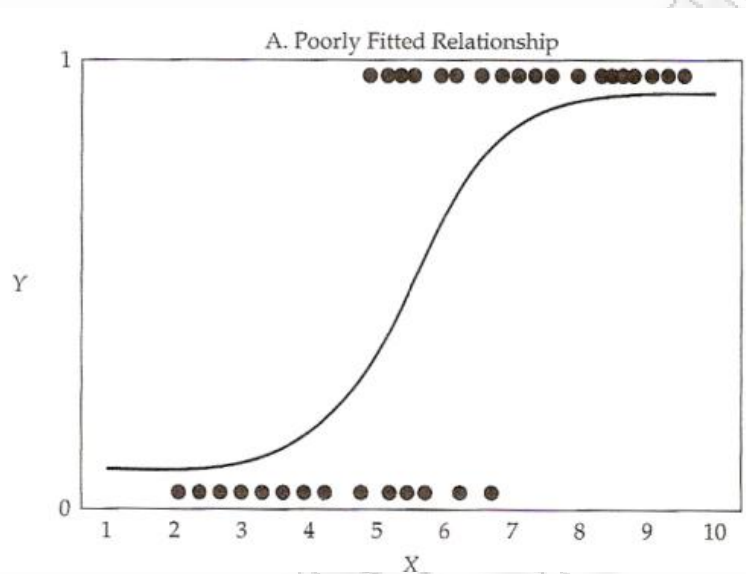
### Εκτίμηση της Πιθανότητας Αθέτησης

Αφού έχει εκτιμηθεί το διάνυσμα  $b$  για έναν οποιοδήποτε πελάτη της τράπεζας είναι δυνατόν να εκτιμηθεί η πιθανότητα αθέτησης εφόσον είναι γνωστές οι τιμές των παραμέτρων  $X_i$  για τον συγκεκριμένο πελάτη. Απλά αντικαθιστάται το διάνυσμα  $\hat{b}$  στον τύπο του PD για συγκεκριμένα πάντα  $X_i$

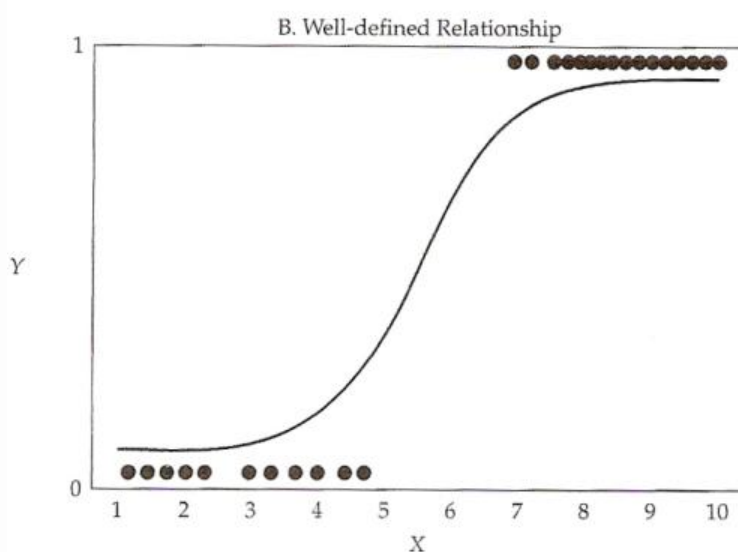
$$PD = \Pr(Y_i = 1) = \frac{1}{1 + e^{-\sum_{k=0}^N \hat{b}_k X_{ik}}} \quad (3.33)$$

Εκτός όμως από την εκτίμηση της πιθανότητας αθέτησης πολύ σημαντικό είναι να υπολογιστεί το cut off point δηλαδή από ποιο σημείο και μετά ένας δανειολήπτης κατατάσσεται στους ασυνεπείς. Συνηθίζεται να εκφράζεται σε τιμές του score  $S_i$  όπως έχει οριστεί προηγουμένως. Επειδή το υπόδειγμα Logit σε κάθε score αντιστοιχίζει μια πιθανότητα και το cut of point μπορεί να οριστεί και ως πιθανότητα.

Στην πραγματικότητα, για τιμές που είναι αρκετά κοντά στο cut off point δεν μπορούν να προκύψουν πολύ ασφαλή συμπεράσματα, ουσιαστικά υπάρχει μια γκριζα περιοχή στην οποία η διάκριση δεν είναι πολύ εμφανής όπως φαίνεται στο γράφημα 3.5. Το φαινόμενο αυτό πρέπει να λαμβάνεται υπόψη, διότι επηρεάζει αρνητικά την αποτελεσματικότητα του υποδείγματος. Επιθυμητό είναι ένα δείγμα με παρατηρήσεις που απέχουν αρκετά από το cut of point, διότι το υπόδειγμα παράγει πιο σημαντικά συμπεράσματα όταν υπάρχει ισχυρή πόλωση όπως φαίνεται στο γράφημα 3.6.



Γράφημα 3.5 (*Hair et al., 1998*)



Γράφημα 3.6 (*Hair et al., 1998*)



### Έλεγχος στατιστικής σημαντικότητας του υποδείγματος

#### Από κοινού σημαντικότητα (Joint Significance)

Για να ελεγχθεί κατά πόσο το μοντέλο είναι σημαντικό εξετάζεται κατά πόσο απορρίπτεται ή όχι η ακόλουθη μηδενική υπόθεση (Bierens, 2008):

$$H_0: b_1 = b_2 = \dots = b_m = 0 \text{ VS } H_1: \text{τουλάχιστον ένα } b_i \neq 0$$

Γίνεται χρήση του στατιστικού Likelihood Ratio (LR) που ακολουθεί  $\chi^2_{df}$  κατανομή

$$LR = -2 \left( \ln L(b_1 = 0, \dots, b_m = 0) - \ln L(\hat{b}_1, \dots, \hat{b}_m) \right) \sim \chi^2_{df}(\alpha)$$

Ο έλεγχος αυτός είναι ένα  $\chi^2$  τεστ επιλέγοντας το διάστημα εμπιστοσύνης (επίπεδο σημαντικότητας) και γνωρίζοντας τους βαθμούς ελευθερίας ( $df=m$ ) απορρίπτεται η  $H_0$  όταν  $LR > \chi^2_{df}(\alpha)$ , στην περίπτωση αυτή το μοντέλο είναι σημαντικό. Στην περίπτωση που  $LR \leq \chi^2_{df}(\alpha)$  δεν μπορεί να απορριφθεί η  $H_0$ , προσοχή αυτό δεν σημαίνει ότι γίνεται αποδεκτή, απλά δεν δύναται να χαρακτηριστεί σημαντικό το μοντέλο.

#### Έλεγχος σημαντικότητας συντελεστών

Για να ελεγχθεί κατά πόσο είναι σημαντικές οι μεταβλητές  $X_i$  εξετάζεται ο εξής αμφίπλευρος έλεγχος (Bierens, 2008):

$$H_0: b_i = 0 \text{ VS } H_1: b_i \neq 0$$

Γίνεται χρήση του στατιστικού :

$$Z = \frac{\sqrt{n}(\hat{b} - 0)}{\hat{\sigma}_b} \sim N(0,1)$$

Συγκρίνεται το στατιστικό  $Z$  με την κριτική τιμή από τον πίνακα της Τυπικής Κανονικής κατανομής για συγκεκριμένο επίπεδο εμπιστοσύνης. Έτσι πχ, για αμφίπλευρο έλεγχο με επίπεδο εμπιστοσύνης 95% η κριτική τιμή είναι περίπου 1,96. Όταν  $|Z| > 1,96$  απορρίπτεται η  $H_0$  δηλαδή, η  $X_i$  μεταβλητή είναι σημαντική. Ενώ, όταν  $|Z| \leq 1,96$  δεν μπορεί να απορριφθεί η  $H_0$ . Για να συμπεριληφθεί στο υπόδειγμα μια μεταβλητή πρέπει να είναι σημαντική.

Ο έλεγχος αυτός μπορεί να γίνει χρησιμοποιώντας και την p-value τιμή από τα αποτελέσματα. Όταν  $p - \text{value} < \alpha$  τότε απορρίπτεται η  $H_0$ .

### Έλεγχος προσαρμογής

Ο έλεγχος προσαρμογής του μοντέλου μπορεί να γίνει με το McFadden R-Squared το οποίο παίρνει τιμές στο (0,1) και όσο πιο κοντά είναι στο 1, τόσο καλύτερη είναι η προσαρμογή του υποδείγματος. Το McFadden R-Squared δίνεται από τον τύπο :

$$R^2 = 1 - \frac{\ln L(\hat{b}_1, \dots, \hat{b}_m)}{\ln L(b_1 = 0, \dots, b_m = 0)}$$

### Αποτελέσματα οριακών μεταβολών (Marginal effects)

Είναι πολύ σημαντικό να μελετηθεί το πώς επηρεάζουν την πιθανότητα αθέτησης πιθανές μεταβολές των ανεξάρτητων μεταβλητών  $X_i$  που έχουν συμπεριληφθεί στο υπόδειγμα (Liao, 1994). Γι' αυτό το λόγο εξετάζεται η επίδραση της οριακής μεταβολής κάθε μεταβλητής  $X_i$  πάνω στην πιθανότητα αθέτησης, κάνοντας χρήση της μερικής παραγώγου του PD ως προς τη μεταβλητή  $X_i$ :

$$\frac{\partial PD}{\partial X_k} = \frac{e^{\sum_{k=0}^N b_k X_{ik}}}{(1 + e^{\sum_{k=0}^N b_{ik} X_{ik}})^2} b_k = b_k PD(1 - PD)$$

Πρέπει να σημειωθεί ότι όταν η μεταβλητή  $X_i$  είναι διακριτή τότε η μερική παράγωγος αποτελεί μια πρόχειρη (rough) προσέγγιση του marginal effect καθώς για να ορίζεται η παράγωγος πρέπει η μεταβλητή  $X_i$  να είναι συνεχής, παρόλο που μπορεί να δώσει μια πρόχειρη προσέγγιση του marginal effect. Στην περίπτωση που έχουμε μια συνεχή μεταβλητή  $X_i$  η προσέγγιση του marginal effect είναι πολύ καλή.

### Odds and Odds Ratios

Πολύ σημαντικό ρόλο στην ανάλυση των αποτελεσμάτων έχουν τα Odds και τα Odds Ratios (Liao, 1994). Τα Odds δίνονται από τον τύπο:

$$Odds(X_i) = \frac{\Pr(Y = 1|X_i)}{1 - \Pr(Y = 1|X_i)}$$

Έστω ότι  $Odds(X_i) = 0.256$ , σημαίνει ότι 256 άτομα κάνουν Default για κάθε 1000 που δεν κάνουν Default.

Τα Odds Ratios δημιουργούνται από τα Odds διαιρώντας τα Odds διαφορετικών κλάσεων αφού βέβαια έχουμε χωρίσει σε κλάσεις τον πληθυσμό ή διαιρώντας Odds με διαφορετικά  $X_i$  δηλαδή,  $\frac{Odds(X+\Delta X)}{Odds(X)}$  (Bierens, 2008).

### **Κόστος λανθασμένης κατάταξης (Cost of Misclassification)**

Η λανθασμένη τοποθέτηση των δανειοληπτών στις δύο κατηγορίες επιφέρει κάποιο κόστος το οποίο εκφράζεται από τα σφάλματα τύπου I και σφάλμα τύπου II. Το σφάλμα τύπου I είναι η τοποθέτηση ενός μη συνεπούς δανειολήπτη στην κατηγορία των συνεπών. Το σφάλμα αυτό σημαίνει ότι μπορεί να δοθεί έγκριση για ένα επισφαλές δάνειο και επομένως, να θέσει σε κίνδυνο το συνολικό ποσό της αντίστοιχης χορήγησης.

Από την άλλη πλευρά, το σφάλμα τύπου II σημαίνει την κατάταξη ενός συνεπούς πελάτη στην κατηγορία των μη συνεπών. Το σφάλμα αυτό έχει ως αποτέλεσμα διαφυγόντα κέρδη για την τράπεζα.

Το σφάλμα τύπου I είναι πιο σημαντικό από αυτό του τύπου II αφού επιφέρει ζημιά και πρέπει να δίνεται περισσότερη προσοχή για την αποφυγή του. Έτσι, για την εκτίμηση της πιθανότητας αθέτησης λαμβάνεται υπόψη η ελαχιστοποίηση των δύο σφαλμάτων δίνοντας ιδιαίτερη βαρύτητα στην ελαχιστοποίηση του σφάλματος τύπου I.

### **2.3 Υπόδειγμα Probit**

Το υπόδειγμα Probit ανήκει επίσης στην οικογένεια των Γενικευμένων Γραμμικών Μοντέλων εκτίμησης πιθανοτήτων. Σύμφωνα με τον *Γκλεζάκο (2008)* το υπόδειγμα αυτό έχει πολλές ομοιότητες με το Logit καθώς κάνει χρήση της συσσωρευτικής συνάρτησης πιθανότητας και προσδιορίζεται με τη μέθοδο μεγίστης πιθανοφάνειας επίσης. Ουσιαστικά, η μοναδική διαφοροποίηση από το υπόδειγμα Logit είναι ότι στο Probit το κατάλοιπο  $\varepsilon_i$  ακολουθεί την Κανονική κατανομή  $N(0,1)$  αντί της Λογιστικής με συνάρτηση κατανομής :

$$\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x \exp(-y^2/2) dy$$

Έτσι ακολουθώντας την ίδια διαδικασία με αυτήν στο υπόδειγμα Logit το υπόδειγμα Probit παίρνει τη μορφή (Liao, 1994) :

$$PD = \Pr(Y_i = 1) = \Phi \left( \sum_{k=1}^N b_{ik} X_{ik} \right) \quad (3.34)$$

Και

$$1 - PD = \Pr(Y_i = 0) = 1 - \Phi \left( \sum_{k=1}^N b_{ik} X_{ik} \right) \quad (3.35)$$

Η σχέση (3.34) είναι αντίστοιχη με την σχέση (3.23) από το υπόδειγμα Logit. Το υπόδειγμα Probit έχει μόνο μια μορφή σε αντίθεση με το Logit που όπως έχει αποδειχθεί εκφράζεται με δυο ισοδύναμες σχέσεις.

Αξίζει να σημειωθεί ότι στην περίπτωση που χρησιμοποιηθεί και σταθερά  $b_{i0}$  το άθροισμα ξεκινάει από το μηδέν με  $X_{i0} = 1$  και η πιθανότητα αθέτησης μπορεί να πάρει τη μορφή :

$$PD = \Pr(Y_i = 1) = \Phi \left( \sum_{k=0}^N b_{ik} X_{ik} \right) \quad (3.36)$$

Οι συντελεστές εκτιμώνται μέσω της μεθόδου της μεγίστης πιθανοφάνειας και στη συνέχεια η πιθανότητα αθέτησης εκτιμάται από τον τύπο 3.37:

$$PD = \Pr(Y_i = 1) = \Phi \left( \sum_{k=0}^N \hat{b}_k X_{ik} \right) \quad (3.37)$$

Όλοι οι στατιστικοί έλεγχοι είναι οι ίδιοι που εφαρμόζονται στο υπόδειγμα Logit.

Σύμφωνα με τον Bessis (2002) τα αποτελέσματα των υποδειγμάτων Probit και Logit είναι παρόμοια, με μόνη διαφορά ότι το Logit είναι πιο εύκολο στη χρήση. Άλλωστε, η λογιστική κατανομή μοιάζει με την κανονική με τη διαφορά ότι οι ακραίες τιμές της παρουσιάζουν σχετικά μεγαλύτερη συχνότητα (πιο παχιές ουρές).

## ΚΕΦΑΛΑΙΟ 3

### ΘΕΩΡΗΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΣΤΗΝ ΚΑΤΑΣΚΕΥΗ ΥΠΟΔΕΙΓΜΑΤΩΝ ΕΚΤΙΜΗΣΗΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΑΘΕΤΗΣΗΣ

Η κατασκευή υποδειγμάτων για τον υπολογισμό της πιθανότητας αθέτησης προϋποθέτει την πραγματοποίηση μιας σειράς ενεργειών οι οποίες διασφαλίζουν κατ' ελάχιστον τα εξής :

- Η μορφή του υποδείγματος που θα επιλεγεί είναι κατάλληλη. Η καταλληλότητα αξιολογείται μέσω της εμπειρικής έρευνας, με κριτήριο την αποτελεσματικότητα που έχει επιδείξει το υπόδειγμα κατά την εφαρμογή του σε ανάλογες περιπτώσεις.
- Οι ανεξάρτητες μεταβλητές που επιλέγονται είναι δοκιμασμένες (με επιτυχία) σε αντίστοιχες περιπτώσεις.
- Τα δεδομένα είναι αντιπροσωπευτικά και έχουν την αναγκαία πληρότητα.

Τα βήματα υλοποίησης ενός υποδείγματος για τον υπολογισμό της πιθανότητας αθέτησης είναι τα εξής :

Επιλογή υποδείγματος

- Επιλογή των ερμηνευτικών μεταβλητών
- Επιλογή και μορφοποίηση δεδομένων
- Ανάπτυξη του υποδείγματος
- Έλεγχος του υποδείγματος
- Οριστικοποίηση της μορφής του υποδείγματος.

#### **3.1 Επιλογή υποδείγματος**

Σύμφωνα με τους *Hastie et al. (2001)* τα υποδείγματα διαχωρισμού αποτελούν βασική προσέγγιση για την εκτίμηση της πιθανότητας αθέτησης. Τα πιο δημοφιλή υποδείγματα διαχωρισμού είναι η Ανάλυση διαχωρισμού (Discriminant Analysis) και



τα υποδείγματα εκτίμησης πιθανότητας (Probability models) (*Duffie and Singleton, 2003*). Δύο από τα πιο δημοφιλή υποδείγματα που ανήκουν στην κατηγορία των Probability models είναι τα Probit και Logit (Λογιστική παλινδρόμηση).

Τα Probability models έχουν συγκριτικό πλεονέκτημα έναντι της Discriminant Analysis καθώς κάποιες από τις βασικές υποθέσεις της Discriminant Analysis δεν ικανοποιούνται στην πράξη (*Medema et al., 2007*). Για παράδειγμα, η βασική υπόθεση ότι οι ανεξάρτητες επεξηγηματικές μεταβλητές πρέπει να κατανέμονται σύμφωνα με την κανονική κατανομή σπάνια συναντάται. Επίσης, ένα ακόμη σημαντικό μειονέκτημα της Discriminant Analysis είναι ότι δεν υπάρχει άμεσος και προφανής τρόπος εκτίμησης της πιθανότητας αθέτησης σε αντίθεση με τα Probability models.

Με βάση τα ευρήματα των *Boyle et al. (1992)*, *Desai et al. (1996,1997)*, *Hentley (1995)*, *Srinivasan και Kim (1987)*, *Yobas et al. (2000)* και του *Thomas (2000)* επιλέγεται το υπόδειγμα Logit για την εκτίμηση της πιθανότητας αθέτησης.

Αξίζει να σημειωθεί ότι οι *Dinh και Kleimeier (2007)* εκτίμησαν την πιθανότητα αθέτησης στην λιανική τραπεζική για τράπεζα του Βιετνάμ. Ανάπτυξη scoring model για μικρά επιχειρηματικά δάνεια σε αναπτυσσόμενες χώρες, όπως το Βιετνάμ, έχουν κάνει επίσης ο *Schreier (2004)* για την Βολιβία και η *Vigano (1993)* για την Μπουρκίνα Φάσο, χώρα της δυτικής Αφρικής. Οι *Medema et al. (2007)* ανέπτυξαν ένα υπόδειγμα για την εκτίμηση της πιθανότητας αθέτησης σε χαρτοφυλάκιο στεγαστικών δανείων για την ολλανδική τράπεζα Friesland Bank. Οι *Yang et al. (2009)* εκτίμησαν την πιθανότητα αθέτησης σε χαρτοφυλάκιο στεγαστικών δανείων για τράπεζα στην Κίνα. Όλοι οι παραπάνω ερευνητές έχουν χρησιμοποιήσει το υπόδειγμα Logit εκτός από τη *Vigano (1993)* που έκανε χρήση της Discriminant Analysis.

### **3.2 Αρχική επιλογή επεξηγηματικών μεταβλητών**

Υπάρχουν πάρα πολλές παράμετροι που επηρεάζουν την συμπεριφορά των δανειοληπτών, οι οποίες εκφράζονται σε ένα υπόδειγμα με τις επεξηγηματικές τυχαίες μεταβλητές. Η μελέτη της συμπεριφοράς όλων αυτών των χαρακτηριστικών

είναι εξαιρετικά δύσκολη, γι' αυτό πρέπει να εξετασθεί μόνο ένα υποσύνολο των συνολικά διαθέσιμων επεξηγηματικών μεταβλητών. Το υποσύνολο αυτό, θα αποτελέσει τη βάση στην οποία θα αναπτυχθεί το τελικό υπόδειγμα. (*Dinh and Kleimeier, 2007*).

Σύμφωνα με τον *Siddiqi (2005)* μια απλή στατιστική μελέτη του δείγματος μπορεί να δώσει πολύ χρήσιμες πληροφορίες. Ειδικότερα, για κάθε τυχαία μεταβλητή εξετάζονται η κατανομή, η μέση τιμή, η διασπορά, το εύρος των τιμών, το ποσοστό των ελλιπών τιμών καθώς και οι ακραίες τιμές που παρατηρούνται. Με απλά στατιστικά μέτρα μπορεί να γίνει μια αρχική εκτίμηση για την σημαντικότητα των τυχαίων μεταβλητών.

Σημαντικό είναι να επιλεγούν οι μεταβλητές που παρέχουν ουσιώδεις πληροφορίες, δηλαδή αυτές που συνεισφέρουν περισσότερο στην διάκριση των δανειοληπτών σε συνεπείς και ασυνεπείς. Η επιλογή αυτή μπορεί να γίνει είτε με ποιοτικά κριτήρια, που βασίζονται στην ανθρώπινη κρίση, είτε με ποσοτικά κριτήρια.

Η αξιολόγηση των μεταβλητών με ποιοτικά κριτήρια μπορεί να πραγματοποιηθεί μέσω της αξιοποίησης των ευρημάτων των σχετικών εμπειρικών ερευνών. Ειδικότερα, η συμμετοχή κάποιων μεταβλητών σε προηγούμενα υποδείγματα εκτίμησης της πιθανότητας αθέτησης μπορεί να αποτελέσει σημαντικό κριτήριο για την αρχική επιλογή. Επίσης, η γνώση και η εμπειρία των ειδικών είναι αρκετά σημαντικές έτσι ώστε να λαμβάνονται υπόψη στην αρχική επιλογή των τυχαίων μεταβλητών.

Όσον αφορά την ποσοτική αξιολόγηση, υπάρχουν διάφορες προτάσεις, όπως πχ του *Siddiqi (2005)*, ο οποίος προτείνει ως ποσοτικό μέτρο για την εκτίμηση της σημαντικότητας των επεξηγηματικών μεταβλητών την «πληροφοριακή αξία» της μεταβλητής - Information Value (IV). Η «πληροφοριακή αξία» υπολογίζεται μόνο για διακριτές τυχαίες μεταβλητές, έτσι οι συνεχείς μεταβλητές πρέπει να διακριτοποιηθούν, δηλαδή να χωριστούν σε κλάσεις. Στη συνέχεια, για κάθε τυχαία μεταβλητή υπολογίζεται το μέτρο «Βαρύτητα των ενδείξεων» (Weight of Evidence - WOE ) για κάθε κλάση τιμών. Το μέγεθος αυτό εκτιμά την προβλεπτική ικανότητα κάθε κλάσης.

$$WOE = \ln \left( \frac{Distr. Good}{Distr. Bad} \right) \quad (4.1)$$

όπου,

Distr. Good είναι το ποσοστό των συνεπών δανειοληπτών μέσα στην κλάση

Distr. Bad είναι το ποσοστό των ασυνεπών δανειοληπτών μέσα στην κλάση

Η προβλεπτική ικανότητα μιας τυχαίας μεταβλητής εκτιμάται με βάση το IV το οποίο δίνεται από τον τύπο:

$$IV = \sum_{i=1}^N (Distr. Good_i - Distr. Bad_i) \times WOE_i \quad (4.2)$$

όπου,

N είναι το πλήθος μονάδων της κλάσης  $i$

Έτσι ανάλογα με την τιμή του IV για κάθε μεταβλητή εκτιμάται και η σημαντικότητα της με βάση τον παρακάτω πίνακα :

$IV < 0.015$	Μηδενική προβλεπτική ικανότητα
$0.015 < IV < 0.1$	Χαμηλή προβλεπτική ικανότητα
$0.1 < IV < 0.3$	Μέτρια προβλεπτική ικανότητα
$0.3 < IV < 0.5$	Δυνατή προβλεπτική ικανότητα
$0.5 < IV$	Πολύ δυνατή προβλεπτική ικανότητα

Το μέτρο IV αναφέρεται και από τους *Hand και Henley (1997)* ως κατάλληλο για την επιλογή των πιο σημαντικών χαρακτηριστικών των δανειοληπτών.

Αξίζει να σημειωθεί ότι η ανάπτυξη του υποδείγματος εκτίμησης της πιθανότητας αθέτησης μπορεί να γίνει κάνοντας χρήση και διακριτών και συνεχών τυχαίων μεταβλητών.



### **3.3 Επιλογή και μορφοποίηση των δεδομένων**

Στην φάση αυτή εξετάζεται η πληρότητα των δεδομένων και αξιολογείται η αντιπροσωπευτικότητά τους.

#### **3.3.1 Αντιπροσωπευτικότητα των δεδομένων**

Τα δεδομένα που χρησιμοποιούνται για την ανάπτυξη του υποδείγματος πρέπει να είναι αντιπροσωπευτικά του αντίστοιχου πληθυσμού. Πχ, στα πλαίσια μιας τράπεζας της ΕΕ, ο πληθυσμός μπορεί να αποτελείται από το δικό της χαρτοφυλάκιο καθώς και από σχετικά εξωτερικά δεδομένα (*Medema et al., 2007*).

Σε κάθε περίπτωση, τα δεδομένα είναι αντιπροσωπευτικά εφόσον χρησιμοποιείται ολόκληρος ο πληθυσμός. Υπάρχει όμως και η δυνατότητα χρησιμοποίησης ενός τυχαίου δείγματος, το οποίο μπορεί να είναι αντιπροσωπευτικό του πληθυσμού, εφόσον έχει κάποιες συγκεκριμένες στατιστικές ιδιότητες.

Στην περίπτωση όπου χρησιμοποιούνται δεδομένα από άλλη εξωτερική πηγή πρέπει να αποδειχθεί ότι είναι αντιπροσωπευτικά του πληθυσμού της.

#### **3.3.2 Πληρότητα των δεδομένων**

Σύμφωνα με τους *Medema et al. (2007)* η πληρότητα των δεδομένων του πληθυσμού ή του δείγματος στο οποίο εφαρμόζεται ένα υπόδειγμα για την εκτίμηση της πιθανότητας αθέτησης αποτελεί έναν από τους βασικούς συντελεστές για την εγκυρότητα και την αξιοπιστία των αποτελεσμάτων. Στην πράξη όμως, στις περισσότερες των περιπτώσεων οι τράπεζες έχουν δεδομένα τα οποία δεν είναι πλήρη. Οι υποψήφιοι δανειολήπτες πολλές φορές δεν παρέχουν τις ζητούμενες πληροφορίες για διάφορους λόγους, πχ γιατί δεν θέλουν να απαντήσουν ή γιατί δεν τους αφορά η ερώτηση ή γιατί μπορεί η ερώτηση να είναι προαιρετική κλπ. Σε κάθε περίπτωση η αντιμετώπιση του προβλήματος των ελλιπών στοιχείων αποτελεί προτεραιότητα, γι' αυτό και η σχετική βιβλιογραφία προτείνει αρκετές εναλλακτικές λύσεις. Σε κάποιες περιπτώσεις, η επιλογή της μεθόδου εξαρτάται από τον τρόπο που

δημιουργήθηκαν οι ελλείψεις. Οι *Little and Rubin (2002)* προτείνουν παρακάτω προσέγγιση:

Έστω ότι για κάθε δανειολήπτη  $i$  παρατηρείται η τιμή της εξαρτημένης μεταβλητής  $Y_i$ , ενώ στις ανεξάρτητες μεταβλητές  $X_{ik}$  παρατηρείται κάποιες από αυτές να μην έχουν τιμή. Έτσι, για κάθε τυχαία μεταβλητή χωρίζεται το διάνυσμα τιμών της,  $X_i$  σε δύο μέρη στις παρατηρούμενες τιμές και στις ελλείψεις,  $X_i = (X_{i\text{obs}}, X_{i\text{mis}})$ . Στη συνέχεια, για κάθε τυχαία μεταβλητή ορίζεται μία «μεταβλητή-δείκτης»  $M_{ik}$ :

$$M_{ik} = \begin{cases} 1 & \text{αν έχει παρατηρηθεί τιμή για το } X_{ik} \\ 0 & \text{αν δεν έχει παρατηρηθεί τιμή για το } X_{ik} \end{cases}$$

Οι περιπτώσεις που παρατηρείται ελλιπής πληροφορία χωρίζονται σε τρεις κατηγορίες ανάλογα με την μορφή της πιθανότητα της τυχαίας μεταβλητής  $M_{ik}$  δοθέντος  $Y, X$   $P(M|Y, X, \theta)$ , όπου  $\theta$  ένα διάνυσμα από άγνωστες παραμέτρους και  $X = (X_{\text{obs}}, X_{\text{mis}})$ .

Η περίπτωση κατά την οποία ισχύει  $P(M|Y, X, \theta) = P(M|\theta)$ , κατατάσσεται στην κατηγορία των «εντελώς τυχαίων ελλείψεων» - «Missing Completely at Random (MCAR)». Σε αυτήν την περίπτωση η πιθανότητα να μη δοθεί κάποια πληροφορία είναι ίδια για όλους και είναι σταθερή, εξαρτάται μόνο από την παράμετρο  $\theta$ .

Η περίπτωση κατά την οποία ισχύει  $P(M|Y, X, \theta) = P(M|Y, X_{\text{obs}}, \theta)$ , κατατάσσεται στην κατηγορία των «τυχαίων ελλείψεων» - Missing at Random (MAR). Η πιθανότητα να μη δοθεί κάποια πληροφορία δεν είναι σταθερή αλλά εξαρτάται από τα δεδομένα που έχουν παρατηρηθεί.

Η περίπτωση κατά την οποία ισχύει  $P(M|Y, X, \theta) = P(M|X_{\text{mis}}, \theta)$  κατατάσσεται στην κατηγορία «μη τυχαίων ελλείψεων» - Not Missing at Random (NMAR). Η πιθανότητα να μη δοθεί κάποια πληροφορία εξαρτάται από τα δεδομένα που λείπουν. Η περίπτωση αυτή είναι πιο δύσκολη στην αντιμετώπισή της από τις άλλες δύο.

### 3.3.3 Μέθοδοι αντιμετώπισης των ελλιπών δεδομένων.

#### 1.) Διαγραφή μεταβλητών (Variable deletion)

Σύμφωνα με τη μέθοδο, όποια τυχαία μεταβλητή παρατηρείται να της λείπουν τιμές αφαιρείται από το υπόδειγμα (Siddiqi, 2005). Πολλές φορές, η μέθοδος αυτή χρησιμοποιείται όταν το ποσοστό των ελλιπών πληροφοριών υπερβεί ένα προκαθορισμένο όριο πχ 5%. Στην περίπτωση που η ερμηνευτική μεταβλητή επηρεάζει σημαντικά την τιμή της ανεξάρτητης μεταβλητής δεν συνιστάται η αφαίρεση της. Με τη χρήση αυτής της μεθόδου μπορεί να μειωθεί σημαντικά το πλήθος των ερμηνευτικών μεταβλητών. Παρόλο που η μέθοδος είναι απλή στην εφαρμογή, σε πολλές περιπτώσεις δεν οδηγεί σε ικανοποιητικά αποτελέσματα.

#### 2.) Αφαίρεση δεδομένων από το δείγμα (listwise deletion)

Επίσης, απλή στην εφαρμογή είναι και η αφαίρεση των δανειοληπτών που έχουν δώσει ελλιπή στοιχεία σε μεταβλητές που χρησιμοποιούνται στην ανάλυση (Brown, 1983). Η μέθοδος αυτή ενδείκνυται όταν οι περιπτώσεις ελλιπών στοιχείων είναι λίγες σε σχέση με το σύνολο του δείγματος, διαφορετικά μπορεί να προκύψει σημαντική μείωση του δείγματος. Η μέθοδος προϋποθέτει τα δεδομένα που λείπουν να ανήκουν στην κατηγορία MCAR (Little and Rubin, 2002). Υπάρχει πάντα ο κίνδυνος το υπόλοιπο του δείγματος να μην είναι αντιπροσωπευτικό του πληθυσμού.

#### 3.) Αφαίρεση στοιχείων κατά ζεύγη (Pairwise deletion)

Η μέθοδος αυτή χρησιμοποιεί όλα τα διαθέσιμα δεδομένα, αφαιρώντας μόνο τις περιπτώσεις που λείπουν στοιχεία. Ειδικότερα με αυτήν τη μέθοδο υπολογίζεται η συσχέτιση  $r_{ij}$  όλων των ερμηνευτικών μεταβλητών ανά ζευγάρια ( $X_i, X_j$ ) για τις περιπτώσεις που υπάρχουν τιμές και για τις δύο τυχαίες μεταβλητές (Wothke, 1998), αφαιρούνται οι περιπτώσεις των δανειοληπτών που δεν έχουν δώσει στοιχεία έστω και για μία από τις δύο τυχαίες μεταβλητές (Kline, 1998).

Ένα πρόβλημα της παραπάνω διαδικασίας είναι ότι τα στοιχεία του πίνακα διακύμανσης - συνδιακύμανσης πιθανώς να είναι υπολογισμένα με βάση διαφορετικά δείγματα. Η μέθοδος προϋποθέτει τα δεδομένα που λείπουν να ανήκουν στην κατηγορία MCAR (Little and Rubin, 2002).

#### 4.) Δημιουργία νέας κατηγορίας

Στην περίπτωση που η τυχαία μεταβλητή είναι διακριτή δημιουργείται μια κατηγορία και σε αυτήν τοποθετούνται οι περιπτώσεις για τις οποίες υπάρχουν ελλιπή στοιχεία.

#### 5.) Υποκατάσταση δεδομένων (Single Imputation)

Η μέθοδος αυτή τοποθετεί στην θέση των τιμών που λείπουν μια τιμή η οποία θεωρείται κατάλληλη υπό την προϋπόθεση ότι τα δεδομένα που λείπουν ανήκουν στην κατηγορία MCAR.

Σύμφωνα με τον *Kline (1998)* υπάρχουν τρεις τύποι της μεθόδου Single Imputation:

1. Υποκατάσταση με την μέση τιμή (Mean imputation).

Στη θέση των τιμών που λείπουν τοποθετείται ο μέσος όλων των παρατηρούμενων τιμών της τυχαίας μεταβλητής. Σε κάποιες περιπτώσεις χρησιμοποιείται ο μέσος των τιμών ενός υποσυνόλου των παρατηρούμενων τιμών της τυχαίας μεταβλητής εφόσον κρίνεται καταλληλότερος.

2. Υποκατάσταση με εκτίμηση μέσω παλινδρόμησης (Regression imputation).

Στην θέση της τιμής που λείπει τοποθετείται η εκτίμηση από ένα υπόδειγμα πολλαπλής παλινδρόμησης που βασίζεται στις παρατηρούμενες τιμές άλλων ερμηνευτικών μεταβλητών.

3. Υποκατάσταση με συναφή τιμή της ίδιας μεταβλητής (Pattern matching).

Στην θέση της τιμής που λείπει τοποθετείται η τιμή της ίδιας τυχαίας μεταβλητής από μια περίπτωση που έχει παρόμοιο προφίλ. Έτσι, όταν ένας δανειολήπτης δεν έχει δώσει κάποια πληροφορία το κενό αυτό μπορεί να καλυφθεί από έναν δανειολήπτη με παρόμοια χαρακτηριστικά.

Ένα μειονέκτημα αυτής της μεθόδου είναι ότι μειώνει την διακύμανση των τιμών με αποτέλεσμα να υπο-εκτιμά τα standard errors και να υπερεκτιμά το συντελεστή προσαρμογής  $R^2$ .

Πολλές φορές η μέθοδος Single imputation συνδυάζεται με την προσθήκη μιας ψευδομεταβλητής  $X_{ik}missing$  στην αρχική επιλογή των ερμηνευτικών τυχαίων μεταβλητών του υποδείγματος όπου :

$$X_{ik}missing = \begin{cases} 1 & \text{αν δεν έχει παρατηρηθεί τιμή για το } X_{ik} \\ 0 & \text{αν έχει παρατηρηθεί τιμή για το } X_{ik} \end{cases}$$

Όταν χρησιμοποιείται η  $X_{ik}missing$  υπάρχει η δυνατότητα στην θέση της τιμής που λείπει να τοποθετηθεί η τιμή μηδέν.

#### 6.) Πολλαπλή υποκατάσταση με τιμές που προκύπτουν από τη χρήση αλγορίθμων (Multiple Imputations (MI) with Expectation Maximization (EM) Algorithm)

Κατά τον [Newman \(2003\)](#) η MI είναι μια διαδικασία σύμφωνα με την οποία στη θέση των τιμών που λείπουν τοποθετούνται διάφορες τιμές ούτως ώστε να παραχθούν διάφορα ολοκληρωμένα σετ δεδομένων, στα οποία εκτιμώνται οι βασικές στατιστικές παράμετροι (μέση τιμή, διασπορά κλπ). Στη συνέχεια, οι εκτιμημένες αυτές παράμετροι από κάθε ολοκληρωμένο σετ δεδομένων συνδυάζονται για να δώσουν μια τελική εκτίμηση των παραμέτρων καθώς και μια λογική εκτίμηση των τυπικών σφαλμάτων (standard errors).

Το ζητούμενο είναι η εκτίμηση ενός σετ δεδομένων το οποίο θα προσεγγίζει όσο το δυνατόν καλύτερα τα πραγματικά δεδομένα. Εφαρμόζονται προσεγγίσεις που χρησιμοποιούν τη μέθοδο μεγίστης πιθανοφάνειας και εκτιμούν τις παραμέτρους εκείνες που μεγιστοποιούν την πιθανότητα τα δεδομένα να είναι τα πραγματικά (αυτά που παρατηρήθηκαν).

### **3.3.4 Κωδικοποίηση Ερμηνευτικών μεταβλητών**

Η κωδικοποίηση των διακριτών ερμηνευτικών μεταβλητών θα πρέπει να γίνεται με τη χρήση των λογαρίθμων odds - [Crook et al. \(1992\)](#). Έτσι, σε μια από τις  $m$  κλάσεις μιας διακριτής τυχαίας μεταβλητής αντιστοιχίζεται η τιμή  $z_i$  σύμφωνα με τον τύπο:

$$z_i = \ln\left(\frac{g_i}{b_i}\right) + \ln\left(\frac{B}{G}\right) \quad (4.3)$$

όπου,

$g_i$  είναι ο αριθμός των συνεπών δανειοληπτών στην κλάση  $i$

$b_i$  είναι ο αριθμός των ασυνεπών δανειοληπτών στην κλάση  $i$



$B$  είναι ο συνολικός αριθμός των ασυνεπών δανειοληπτών στο δείγμα,  $B = \sum_{i=1}^m b_i$

$G$  είναι ο συνολικός αριθμός των συνεπών δανειοληπτών στο δείγμα,  $G = \sum_{i=1}^m g_i$

Την κωδικοποίηση αυτή έχουν εφαρμόσει και οι *Dinh και Kleimeier (2007)* που εκτίμησαν την πιθανότητα αθέτησης στην λιανική τραπεζική για τράπεζα του Βιετνάμ.

### 3.3.5 Διαχωρισμός του δείγματος σε δείγμα εκτίμησης και δείγμα επικύρωσης

Σύμφωνα με τη Βασιλεία II (παρ. 420) για την επικύρωση του υποδείγματος πρέπει να ελεγχθεί η αποτελεσματικότητά του σε δείγμα διαφορετικό του δείγματος εκτίμησης των συντελεστών της συνάρτησης διάκρισης (Out of sample validation).

Για να είναι εφικτή η επικύρωσή του, το δείγμα χωρίζεται τυχαία σε δύο μέρη :

- α) το δείγμα εκτίμησης των συντελεστών των ερμηνευτικών μεταβλητών και
- β) το δείγμα επικύρωσης του υποδείγματος.

Το δείγμα εκτίμησης χρησιμοποιείται για την ανάπτυξη του υποδείγματος και το δείγμα επικύρωσης στο οποίο επικυρώνεται η αποτελεσματικότητά του (*Dinh και Kleimeier (2007)* και οι *Medema et al. (2007)*).

## 3.4 Θεωρητική προσέγγιση στην ανάπτυξη υποδείγματος Logit

### 3.4.1 Κατασκευή της συνάρτησης διάκρισης

Η συνάρτηση διάκρισης (Discrimination function) κατασκευάζεται με δύο μεθόδους  
α) Την προσέγγιση bottom-up (stepwise) και β) Την προσέγγιση top-down (inverse stepwise) (*Dinh and Kleimeier, 2007*).

#### 1. Προσέγγιση Bottom-up

Σύμφωνα με την μέθοδο αυτή, η πρώτη μεταβλητή που συμπεριλαμβάνεται στην συνάρτηση είναι αυτή που παρουσιάζει την μεγαλύτερη στατιστική σημαντικότητα για ένα συγκεκριμένο διάστημα εμπιστοσύνης πχ 95%.

Στη συνέχεια, προσθέτονται διαδοχικά οι μεταβλητές που βελτιώνουν περισσότερο την στατιστική σημαντικότητα της συνάρτησης. Η διαδικασία αυτή ολοκληρώνεται όταν η προσθήκη επιπλέον μεταβλητών στη συνάρτηση δεν ενισχύει τη σημαντικότητα του υποδείγματος.

## 2. Προσέγγιση Top-Down

Κατά τη μέθοδο αυτή εφαρμόζεται αντίθετος τρόπος κατασκευής της συνάρτησης διάκρισης σε σχέση με την μέθοδο Bottom-up. Η κατασκευή ξεκινά με τη χρήση όλων των ερμηνευτικών μεταβλητών που έχουν επιλεγεί αρχικά. Σε κάθε βήμα αυτής της μεθόδου αφαιρείται η μεταβλητή που είναι στατιστικά πιο ασήμαντη για ένα συγκεκριμένο διάστημα εμπιστοσύνης. Η διαδικασία αυτή ολοκληρώνεται όταν όλες οι μεταβλητές που συμπεριλαμβάνονται στο υπόδειγμα είναι στατιστικά σημαντικές με συνέπεια η αφαίρεση κάποιας να κάνει πιο ασήμαντη την συνάρτηση διαχωρισμού.

Η συνάρτηση διάκρισης που παράγεται με τις δύο μεθόδους για το ίδιο διάστημα εμπιστοσύνης πρέπει να είναι ακριβώς η ίδια (*Dinh and Kleimeier, 2007*). Έτσι επικυρώνεται και η ορθότητα του αποτελέσματος.

### **3.4.2 Στατιστική επικύρωση του υποδείγματος**

Η σύγκριση της στατιστικής σημαντικότητας των μεταβλητών και των συναρτήσεων διάκρισης σε κάθε βήμα των μεθόδων bottom-up και top-down γίνεται με τα παρακάτω κριτήρια:

Έστω ότι επιλέγεται διάστημα εμπιστοσύνης 95% τότε μια μεταβλητή είναι σημαντική εφόσον  $|z\text{-statistic}| > 1,96$  (ή το αντίστοιχο  $p\text{-value} < 0,05$ ). Ανάμεσα σε δύο μεταβλητές σημαντικότερη είναι αυτή που έχει τη μεγαλύτερη τιμή στο  $|z\text{-statistic}|$  (ή τη μικρότερη στο αντίστοιχο  $p\text{-value}$ ). Επίσης, η συνάρτηση διαχωρισμού που έχει κατασκευαστεί από ένα σύνολο ερμηνευτικών μεταβλητών είναι σημαντική εφόσον  $\text{Likelihood Ratio (LR)} > \chi^2_{df}(\alpha)$ . Ανάμεσα σε δύο συναρτήσεις διαχωρισμού σημαντικότερη είναι αυτή που το LR έχει μεγαλύτερη τιμή. Ανάμεσα σε δυο υποδείγματα σημαντικότερο είναι εκείνο που έχει μικρότερες τιμές στα κριτήρια Akaiki και Swartz.

Σημαντικό κριτήριο είναι και το Macfadden  $R^2$  που όσο μεγαλύτερη τιμή έχει, τόσο καλύτερα ερμηνεύει το υπόδειγμα τις διακυμάνσεις του δείγματος δηλαδή έχει καλύτερη προσαρμογή.

Το υπόδειγμα για να γίνει αποδεκτό πρέπει τα κατάλοιπα να μην έχουν ετεροσκεδαστικότητα, έτσι εξετάζεται αν τα squared residuals έχουν μηδενική συσχέτιση.

### Έλεγχος προσαρμογής με το Hosmer-Lemeshow Statistic

Η προσαρμογή του υποδείγματος ελέγχεται με το κριτήριο Hosmer-Lemeshow το οποίο εξετάζει αν το εκτιμημένο υπόδειγμα είναι κατάλληλο (*Hair et al., 1998*).

Στο H-L εξετάζεται η μηδενική υπόθεση :

$H_0$  : Δεν υπάρχει έλλειψη προσαρμογής

$H_1$  : Υπάρχει έλλειψη προσαρμογής

Το H-L χωρίζει το δείγμα  $n$  απαιτήσεων σε  $k$  ισομεγέθεις (μέγεθος =  $n/k$ ) κλάσεις με βάση την πιθανότητα αθέτησης. Στην πρώτη κλάση τοποθετούνται οι  $n/k$  απαιτήσεις με την μικρότερη εκτιμημένη πιθανότητα αθέτησης και στην τελευταία οι  $n/k$  απαιτήσεις με την μεγαλύτερη.

Σε κάθε κλάση συγκρίνεται η παρατηρούμενη συχνότητα των «προβληματικών» απαιτήσεων με την αναμενόμενη τους συχνότητα. Η παρατηρούμενη συχνότητα δίνεται από τη σχέση  $\sum_{j=1}^{n_i} y_{ij}$  και η αναμενόμενη από την σχέση  $\sum_{j=1}^{n_i} \widehat{PD}_{ij}$ .

Το H-L κριτήριο κάνει χρήση του στατιστικού HL που ακολουθεί την κατανομή  $\chi^2_{k-2, \alpha}$  και δίνεται από τον τύπο :

$$HL = \sum_{i=1}^k \frac{\left(\sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \widehat{PD}_{ij}\right)^2}{\left(\sum_{j=1}^{n_i} \widehat{PD}_{ij}\right) \left(1 - \frac{\sum_{j=1}^{n_i} \widehat{PD}_{ij}}{n_i}\right)} \quad (4.4)$$

όπου,

$n_i$  = πλήθος των απαιτήσεων στην  $i$  κλάση



$\overline{PD}_{ij}$  = η εκτιμημένη πιθανότητα αθέτησης για τον δανειολήπτη j στην κλάση i

Σύμφωνα με το κριτήριο Hosmer-Lemeshow υπάρχει σίγουρα έλλειψη προσαρμογής στην περίπτωση που απορρίπτεται η μηδενική υπόθεση απορρίπτεται, δηλαδή όταν  $HL > X_{k-2,\alpha}^2$ .

Για να γίνει αποδεκτό το υπόδειγμα πρέπει να μην μπορεί να απορριφθεί η μηδενική υπόθεση στο H-L κριτήριο, δηλαδή να ισχύει  $HL < X_{k-2,\alpha}^2$ .

Το κριτήριο αυτό έχουν χρησιμοποιήσει οι *Yang et al. (2009)* στο υπόδειγμα για την εκτίμηση της πιθανότητας αθέτησης σε χαρτοφυλάκιο στεγαστικών δανείων Κινέζικης τράπεζας.

### 3.4.3 Καθορισμός του σημείου διαχωρισμού (cut off point)

Η τιμή cut off point έχει καθοριστική σημασία για τον υπολογισμό της αποτελεσματικότητας του υποδείγματος. Το cut off point προσδιορίζεται ανάλογα με το προφίλ και την πολιτική της τράπεζας γιατί ουσιαστικά με τον καθορισμό του γίνεται η επιλογή των σφαλμάτων τύπου I και II του υποδείγματος. Συνήθως το cut off point επιλέγεται έτσι ώστε να ελαχιστοποιηθούν τα σφάλματα δίνοντας βαρύτητα στο σφάλμα τύπου I.

Σύμφωνα με τους *Hair et al. (1998)* το cut off point επίσης μπορεί να προσδιοριστεί και από τον τύπο:

$$Z_{cut} = \frac{N_0 \overline{PD}_0 + N_1 \overline{PD}_1}{N_0 + N_1} \quad (4.5)$$

όπου,

$N_0$ : το πλήθος των συνεπών

$N_1$  : το πλήθος των ασυνεπών

$\overline{PD}_0$  : το μέσο PD των συνεπών

$\overline{PD}_1$  : το μέσο PD των ασυνεπών

Ο προσδιορισμός του cut off point δεν βασίζεται μόνο σε ποσοτικούς υπολογισμούς καθώς αυτοί πρέπει πάντα συμφωνεί και με την ανθρώπινη κρίση. Πχ μια τράπεζα

που επιθυμεί επιθετική πολιτική πιθανά θα αυξήσει το cut off point έτσι ώστε να εγκρίνει περισσότερα δάνεια αναλαμβάνοντας όμως και μεγαλύτερο κίνδυνο καθώς η πιθανότητα να εγκρίνει και επισφαλή δάνεια είναι αυξημένη.

#### **3.4.4 Αποτελεσματικότητα (προβλεπτική ικανότητα) του υποδείγματος**

Η αποτελεσματικότητα του υποδείγματος εκτιμάται με βάση το ποσοστό των σωστά τοποθετημένων δανειοληπτών. Έτσι, εφόσον έχει καθοριστεί το cut off point εκτιμώνται τα μεγέθη PCCgood, PCCbad και PCCtotal (PCC= Percentage of Correctly Classified) (*Dinh and Kleimeier, 2007*).

όπου,

PCCgood είναι το ποσοστό των σωστά τοποθετημένων συνεπών δανειοληπτών

PCCbad είναι το ποσοστό των σωστά τοποθετημένων ασυνεπών δανειοληπτών

PCCtotal είναι το ποσοστό των σωστά τοποθετημένων δανειοληπτών συνολικά

Επίσης, εξετάζονται τα σφάλματα τύπου I και II τα οποία προκύπτουν άμεσα από τα παραπάνω μεγέθη. Επιθυμείται η ελαχιστοποίηση τους δίνοντας έμφαση στο σφάλμα τύπου I που επιφέρει ζημιά στην τράπεζα. Τα δύο σφάλματα έχουν αρνητική συσχέτιση εξ' ορισμού. Για να είναι αποτελεσματικό το υπόδειγμα θα πρέπει η συσχέτιση των σφαλμάτων να είναι μικρή κατ' απόλυτη τιμή. Διαφορετικά οι μικρές τιμές για ένα σφάλμα θα συνεπάγονται αυτόματα μεγάλες τιμές για το άλλο.

Η αποτελεσματικότητα του υποδείγματος εκτιμάται αρχικά στο δείγμα εκτίμησης και στη συνέχεια στο δείγμα επικύρωσης διότι όπως έχει προαναφερθεί είναι απαραίτητο να ελεγχθεί η αποτελεσματικότητα του σε δείγμα διαφορετικό του δείγματος εκτίμησης.

## ΚΕΦΑΛΑΙΟ 4

### ΑΝΑΠΤΥΞΗ ΥΠΟΔΕΙΓΜΑΤΟΣ ΕΚΤΙΜΗΣΗΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΑΘΕΤΗΣΗΣ, ΓΙΑ ΜΙΑ ΕΛΛΗΝΙΚΗ ΤΡΑΠΕΖΑ

Η ανάπτυξη υποδείγματος εκτίμησης της πιθανότητας αθέτησης για τους πελάτες μιας τράπεζας, ακολουθεί τα βήματα που ήδη αναφέρθηκαν για την κατάρτιση υποδειγμάτων σε θεωρητική βάση.

Στα πλαίσια αυτά και με στόχο την ανάπτυξη του υποδείγματος της παρούσας εργασίας για την ελληνική τράπεζα, ακολουθήθηκαν τα εξής επιμέρους βήματα :

- Επιλογή του δείγματος,
- Επιλογή επεξηγηματικών μεταβλητών,
- Επιλογή και μορφοποίηση δεδομένων,
- Κατασκευή της συνάρτησης διάκρισης,
- Στατιστική επικύρωση του υποδείγματος,
- Καθορισμός του Σημείου Διαχωρισμού (cut off point),
- Μέτρηση της αποτελεσματικότητας του υποδείγματος).

#### 4.1. Το Δείγμα

Η ανάπτυξη του υποδείγματος εκτίμησης της πιθανότητας αθέτησης βασίζεται σε δείγμα 15.786 καταναλωτικών δανείων από χαρτοφυλάκιο ελληνικής τράπεζας. Οι ημερομηνίες αιτήσεων των δανειοληπτών είναι από 11/2006 έως 4/2007. Η βάση δεδομένων της τράπεζας περιέχει 26 επιμέρους προσωπικά στοιχεία των δανειοληπτών τα οποία χρησιμοποιούνται στην εργασία ως ερμηνευτικές μεταβλητές. Περιλαμβάνει επίσης τον χαρακτηρισμό κάθε δανειολήπτη ως φερέγγυο ή αφερέγγυο ανάλογα με την συμπεριφορά του, από την ημερομηνία έγκρισης του δανείου μέχρι τις 15/7/2009.

Οι μεταβλητές που περιέχονται στην βάση δεδομένων της τράπεζας είναι :

1. **Ημερομηνία αίτησης (ηη/μμ/εε)**
2. **Ημερομηνία γέννησης (ηη/μμ/εε)**
3. **Κάτοχος πιστωτικής κάρτας:** Εκφράζει κατά πόσο ο πελάτης είχε πιστωτική κάρτα ή όχι.
4. **Ιδιοκτησία:** Εκφράζει κατά πόσο το οίκημα που διαμένει ο δανειολήπτης είναι ιδιόκτητο ή όχι.
5. **Οικογενειακή κατάσταση**
6. **Τηλέφωνα:** Δείχνει αν δόθηκαν αριθμοί κλήσης για τηλεφωνική επικοινωνία. ή όχι.
7. **Επάγγελμα**
8. **Πελάτης της τράπεζας:** Εκφράζει κατά πόσο ο δανειολήπτης ήταν ήδη πελάτης της τράπεζας.
9. **Χειρότερη καταχώρηση Τειρεσία:** Εμφανίζει την χειρότερη καταχώρηση που έχει γίνει στην Τειρεσία για τον δανειολήπτη.
10. **Έτη από την πιο πρόσφατη καταχώρηση Τειρεσία:** Αναφέρει τα έτη που έχουν περάσει από την πιο πρόσφατη καταχώρηση στην Τειρεσία για τον δανειολήπτη.
11. **Διάρκεια δανείου**
12. **Μηνιαία δόση προς μηνιαίο εισόδημα:** Εκφράζει τι ποσοστό του μηνιαίου μισθού αποτελεί η δόση του δανείου.
13. **Επιτόκιο δανείου**
14. **Φύλο**
15. **Αριθμός παιδιών**
16. **Επίπεδο μόρφωσης**
17. **Είδος προσωπικού εγγράφου:** Εκφράζει το προσωπικό έγγραφο με το οποίο ο δανειολήπτης επιβεβαιώνει τα προσωπικά του στοιχεία (αστυνομική ταυτότητα διαβατήριο κλπ.).
18. **Έτη στην τωρινή διεύθυνση**
19. **Έτη στην τωρινή εργασία**
20. **Ταχυδρομικός κωδικός**
21. **Νέος κωδικός επαγγέλματος**
22. **Κωδικός επαγγέλματος Singlo**
23. **Χώρα κατοικίας**
24. **Υπηκοότητα**

## 25. Ετήσιο εισόδημα

## 26. Συμπεριφορά του δανειολήπτη, Υ.

Παρατηρείται ότι οι δανειολήπτες που έχουν χαρακτηριστεί ασυνεπείς είναι 2.389, δηλαδή το 15,13% του συνολικού δείγματος (ασυνεπείς χαρακτηρίζονται όσοι έχουν καθυστερήσει την αποπληρωμή του δανείου για τρεις συνεχόμενους μήνες τουλάχιστον μια φορά μέχρι τις 15/7/2009).

### 4.2. Αρχική επιλογή των ερμηνευτικών μεταβλητών

Ο αριθμός των ερμηνευτικών μεταβλητών είναι αρκετά μικρός και γι' αυτό δεν επιβάλλεται η μείωσή τους μέσω κάποιας διαδικασίας επιλογής για χρηστικούς λόγους, καθώς είναι εφικτή η χρησιμοποίηση και των 26 μεταβλητών.

Παρόλα αυτά, πρέπει να αποκλειστούν κάποιες μεταβλητές γιατί οι πληροφορίες που παρέχουν δεν δύναται να συνεισφέρουν σημαντικά στη διάκριση της συμπεριφοράς των δανειοληπτών. Οι μεταβλητές αυτές είναι οι εξής :

- *Χώρα κατοικίας*, διότι μόνο 25 στους 15786 δανειολήπτες κατοικούν εκτός Ελλάδος.
- *Υπηκοότητα*, διότι μόνο 155 από τους 15786 δανειολήπτες έχουν διαφορετική υπηκοότητα από την ελληνική.
- *Επίπεδο μόρφωσης*, διότι κανένας πελάτης δεν έχει απαντήσει αυτήν την ερώτηση.
- *Έτη στην τωρινή διεύθυνση*, διότι μόνο 81 άτομα έχουν δώσει απάντηση σε αυτήν την ερώτηση.
- *Έτη στην τωρινή εργασία*, διότι μόνο 79 απάντησαν σε αυτήν την ερώτηση.
- *Τηλέφωνα*, διότι μόνο 117 άτομα από τα 15786 δεν έδωσαν τηλέφωνο επικοινωνίας.

Επίσης, αποκλείονται και οι μεταβλητές *Επάγγελμα*, *Επιτόκιο δανείου*, *Ταχυδρομικός κωδικός*, *Νέος κωδικός επαγγέλματος*, *Κωδικός επαγγέλματος Singlo* διότι η κωδικοποίηση τους ήταν σε μορφή δύσκολα επεξεργάσιμη χωρίς περεταίρω πληροφόρηση από την τράπεζα.

Έτσι μετά την αφαίρεση των πιο πάνω μεταβλητών απέμειναν 14 μεταβλητές για την ανάπτυξη του υποδείγματος εκτίμησης της πιθανότητας αθέτησης οι οποίες θα αναλυθούν στην συνέχεια.

### **4.3. Ανάλυση και κωδικοποίηση των μεταβλητών**

#### Εξαρτημένη μεταβλητή Y (διακριτή) : Φερεγγυότητα του πελάτη

Η μεταβλητή Y παίρνει την τιμή 1 αν ο πελάτης έχει αθετήσει την υποχρέωση του, δηλαδή δεν έχει πληρώσει τις δόσεις για τρεις τουλάχιστον συνεχόμενους μήνες ανεξάρτητα με το αν μετά έγινε συνεπής στις υποχρεώσεις του.

Αν η συμπεριφορά του πελάτη είναι σωστή τότε η μεταβλητή Y έχει την τιμή μηδέν.

Αξίζει να σημειωθεί ότι μπορεί να υπάρχουν περιπτώσεις πελατών που ήταν ασυνεπείς στο παρελθόν για κάποιους λόγους αλλά στη συνέχεια έγιναν συνεπείς. Ο συγκεκριμένος τρόπος χαρακτηρισμού λοιπόν της συμπεριφοράς των πελατών έχει κάποιες αδυναμίες καθώς ένας πελάτης μπορεί να έχει χαρακτηριστεί ασυνεπής αλλά να είναι τελικά συνεπής, γι' αυτό η τράπεζα δεν κινεί διαδικασίες απόκτησης των αντίστοιχων εξασφαλίσεων.

Γίνεται αντιληπτό λοιπόν ότι ο χαρακτηρισμός της τράπεζας «ασυνεπής πελάτης» είναι αμετάκλητος και αυστηρός, γεγονός που σίγουρα θα έχει επίδραση στην κατασκευή και στην προβλεπτική ικανότητα του υποδείγματος εκτίμησης της πιθανότητας αθέτησης.

#### Ανεξάρτητη μεταβλητή X1: Ημερομηνία αίτησης (συνεχής)

Η μεταβλητή X1 εκφράζει την ημερομηνία αίτησης. Η περίοδος των αιτήσεων του δείγματος είναι από 11/2006 έως 4/2007. Εκτιμάται ότι η συμπεριφορά των δανειοληπτών δεν διαφοροποιείται σημαντικά σε επίπεδο ημέρας.

Αντίθετα, ο μήνας της αίτησης υπάρχει αυξημένη πιθανότητα να έχει συνεισφορά στην διάκριση της συμπεριφοράς των δανειοληπτών, καθώς στην Ελλάδα υπάρχουν κάποιοι μήνες που δημιουργούν συγκεκριμένες τάσεις στην ζήτηση δανείων.

Πχ, κατά τις γιορτινές περιόδους (Χριστούγεννα, Πάσχα) παρατηρείται υψηλή ζήτηση καθώς αρκετός κόσμος καταφεύγει στην ζήτηση δανεικών για να αντιμετωπίσει τις αυξημένες δαπάνες των ημερών αυτών. Επίσης, παρατηρείται αύξηση των δανείων για την αγορά αυτοκινήτου προς το τέλος του χρόνου καθώς ο κόσμος θέλει να αποφύγει τα Τέλη ταξινόμησης για την απερχόμενη χρονιά.

Αξίζει να σημειωθεί ότι αρκετά δάνεια που δίδονται σε συγκεκριμένους μήνες μπορεί να έχουν και συγκεκριμένο σκοπό, οπότε ενδέχεται να υπάρχει μια τάση στην συμπεριφορά των δανειοληπτών. Στην περίπτωση αυτή, η μεταβλητή X1 αναμένεται να έχει συνεισφορά στην εκτίμηση της πιθανότητας αθέτησης.

Η κωδικοποίηση της μεταβλητής X1 από την τράπεζα πραγματοποιείται από έναν οκταψήφιο αριθμό που σχηματίζεται από την αναγραφή του έτους, του μήνα και της ημέρας. Πχ η ημερομηνία 10-12-2006 εκφράζεται με τον αριθμό 20061210. Η κωδικοποίηση αυτή παρουσιάζει πρόβλημα κατά την αλλαγή του έτους, όταν δύο διαδοχικές ημέρες απέχουν 8870 μονάδες, ενώ κατά το υπόλοιπο έτος απέχουν μόνο μία μονάδα. Για παράδειγμα, οι ημερομηνίες 01-01-2007 και 31-12-2006 απέχουν :  $20070101-20061231=8870$  μονάδες. Ενώ πχ, οι ημερομηνίες 12-02-2007 και 13-02-2007 απέχουν :  $20070213-20070212=1$  μονάδα.

Για να αντιμετωπιστεί αυτό το πρόβλημα προτιμήθηκε η μεταβλητή X1 να εκφράζει τους μήνες που έχουν περάσει από την ημερομηνία αίτησης ως σήμερα 15-7-2009. Αξίζει να σημειωθεί ότι γίνεται παράλειψη των ημερών έτσι ώστε πχ η ημερομηνία 10-11-2006 παίρνει την τιμή 32 ( $(2009-2006)*12+(7-11)=32$ ), δηλαδή έχουν περάσει 32 μήνες από την ημερομηνία αίτησης.

#### Ανεξάρτητη μεταβλητή X2: Ημερομηνία γέννησης (συνεχής)

Η ημερομηνία γέννησης αναμένεται να έχει συνεισφορά στη διάκριση της συμπεριφοράς των δανειοληπτών καθώς κάθε ηλικία κατά γενικό κανόνα έχει συγκεκριμένα χαρακτηριστικά, ορισμένα από αυτά είναι η επαγγελματική σταθερότητα, η οικονομική κατάσταση, η οικογενειακή κατάσταση κλπ.

Η κωδικοποίηση που υπάρχει στη βάση δεδομένων είναι η ίδια με αυτήν της ημερομηνίας αίτησης.

Γίνεται αποκοπή των μηνών και των ημερών και στην μεταβλητή δίνεται η ηλικία του δανειολήπτη σε ακέραια έτη από την ημερομηνία γέννησής του μέχρι τις 15/7/09. Για παράδειγμα, η μεταβλητή 12520116 αντιστοιχίζεται στην τιμή 57 που εκφράζει πλέον την ηλικία σε ακέραια έτη.

#### Ανεξάρτητη μεταβλητή X3: Κάτοχος πιστωτικής κάρτας (διακριτή)

Ο πελάτης που έχει πιστωτική κάρτα ενδέχεται να έχει διαφορετική συμπεριφορά με αυτόν που δεν έχει καθώς για να του έχει εγκριθεί η κάρτα σημαίνει ότι έχει αξιολογηθεί θετικά από προηγούμενο έλεγχο, ενώ αντίθετα για τον πελάτη που δεν έχει κάρτα δεν υπάρχουν προηγούμενες «συστάσεις». Επίσης, η άρνηση του πελάτη να απαντήσει σε αυτό το ερώτημα είναι μία ένδειξη επικινδυνότητας.

Η μεταβλητή X3 σύμφωνα με τη βάση δεδομένων της τράπεζας παίρνει την τιμή 1 αν ο πελάτης είναι κάτοχος πιστωτικής και την τιμή 2 αν δεν είναι κάτοχος αλλά και στην περίπτωση που δεν έχει απαντήσει το ερώτημα.

Η συμπεριφορά των δανειοληπτών ως προς αυτό το χαρακτηριστικό φαίνεται στον παρακάτω πίνακα :

	ΚΑΤΟΧΟΣ ΠΙΣΤΩΤΙΚΗΣ 1	ΟΧΙ/ΔΕΝ ΑΠΑΝΤΗΣΕ 2
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	7426	8360
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	6,362	7,035
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	1,064	1,325
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	85.67%	84.15%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	14.33%	15.85%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	16.72%	18.83%

Παρατηρείται ότι η συμπεριφορά των δανειοληπτών διαφοροποιείται αρκετά με βάση το παραπάνω κριτήριο, γι' αυτό είναι ενδεχόμενο συνεισφορά στην διάκριση της εν γένει συμπεριφοράς των δανειοληπτών.

#### Ανεξάρτητη μεταβλητή X4: Ιδιοκτησία (οίκημα κατοικίας - διακριτή)

Η μεταβλητή αυτή εκφράζει κατά πόσο οι πελάτες διαμένουν σε οίκημα ιδιόκτητο, ή όχι. Ενδέχεται να είναι σημαντική στην εκτίμηση της πιθανότητας αθέτησης καθώς έμμεσα εκφράζει κατά ένα μέρος την οικονομική κατάσταση των πελατών. Επίσης, οι δανειολήπτες με ιδιόκτητη κατοικία αναμένεται να είναι πιο συνεπείς γιατί σε



περίπτωση αθέτησης υπάρχει το ενδεχόμενο η τράπεζα να την διεκδικήσει και να την χάσουν.

Η X4 αναμένεται να είναι περισσότερο ενδεικτική για την οικονομική κατάσταση αυτών που δεν διαμένουν σε ιδιόκτητο οίκημα καθώς στην Ελλάδα υπάρχει υψηλό ποσοστό ιδιοκατοίκησης, οπότε η ιδιοκτησία ενός οικήματος δεν συνεπάγεται και οικονομική ευρωστία. Αντίθετα, η μη ιδιοκτησία οικήματος ίσως να συνεπάγεται οικονομική δυσχέρεια.

Σύμφωνα με την βάση δεδομένων της τράπεζας, οι τιμές που παίρνει η μεταβλητή X4 είναι οι εξής: 3 αν το οίκημα είναι ιδιόκτητο, 4 αν ο πελάτης το ενοικιάζει και 5 στην περίπτωση που ο πελάτης είναι φιλοξενούμενος σε κάποιο οίκημα, πχ των γονιών του ή των παιδιών του.

Παρατηρείται ότι σε κάποιους πελάτες έχει δοθεί η τιμή 0 η οποία δεν αντιστοιχίζεται σε κάποια κατηγορία οπότε θα θεωρηθεί ότι είτε έχει τοποθετηθεί από λάθος, είτε δεν δόθηκε απάντηση. Οι περιπτώσεις αυτές διαγράφονται.

Η συμπεριφορά των δανειοληπτών φαίνεται στον παρακάτω πίνακα:

	0	ΙΔΙΟΚΤΗΤΗΣ 3	ΕΝΟΙΚΙΑΣΤΗΣ 4	ΦΙΛΟΞ/ΝΟΣ 5
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	36	12262	1830	1658
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	30	10,510	1,517	1,340
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	6	1,752	313	318
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	83.33%	85.71%	82.90%	80.82%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	16.67%	14.29%	17.10%	19.18%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	20.00%	16.67%	20.63%	23.73%

Από τον πίνακα παρατηρείται ότι υπάρχει αξιόλογη διαφορά στην συμπεριφορά των δανειοληπτών ανάλογα με την κλάση στην οποία ανήκουν, καθώς οι ιδιοκτήτες των οικημάτων που διαμένουν είναι συνεπέστεροι από όσους ενοικιάζουν και αυτοί με τη σειρά τους συνεπέστεροι από αυτούς που φιλοξενούνται. Έτσι, επιβεβαιώνεται ότι η μεταβλητή αυτή αποτελεί μια ένδειξη της οικονομικής κατάστασης των πελατών.

#### Ανεξάρτητη μεταβλητή X5: Οικογενειακή κατάσταση (διακριτή)

Η μεταβλητή αυτή αναμένεται να είναι πολύ σημαντική καθώς εκτιμάται ότι οι πελάτες έχουν διαφορετική συμπεριφορά έχουν ανάλογα με την οικογενειακή τους

κατάσταση. Ένας έγγαμος δανειολήπτης αναμένεται να είναι συνεπέστερος από έναν άγαμο ή διαζευγμένο καθώς ενδεχόμενη ασυνεπεία του θα έχει επίδραση και στην οικογένειά του. Αντίθετα, ένας άγαμος ή διαζευγμένος είναι πιθανό να μην έχει την ευθύνη τρίτων και έτσι μπορεί ευκολότερα να αθετήσει την υποχρέωση του προς την τράπεζα. Ειδικά ο διαζευγμένος, εκτός του ότι είναι ελεύθερος πιθανά θα πληρώνει και κάποια χρήματα ως διατροφή γεγονός που δυσχεραίνει την οικονομική του κατάσταση.

Οι τιμές που παίρνει η μεταβλητή αυτή σύμφωνα με την βάση δεδομένων της τράπεζας είναι: 1 αν ο πελάτης είναι έγγαμος, 2 αν είναι άγαμος, 3 στην περίπτωση που είναι διαζευγμένος και 4 για την κατηγορία ΑΛΛΟ.

Η συμπεριφορά των δανειοληπτών φαίνεται στον παρακάτω πίνακα:

	ΕΓΓΑΜΟΣ 1	ΑΓΑΜΟΣ 2	ΔΙΑΖ/ΝΟΣ 3	ΆΛΛΟ 4
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	11204	3395	632	515
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	9,710	2,706	506	440
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	1,494	689	126	75
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	86.67%	79.71%	80.06%	85.44%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	13.33%	20.29%	19.94%	14.56%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	15.39%	25.46%	24.90%	17.05%

Ο παραπάνω πίνακας επιβεβαιώνει την ex-ante εκτίμηση ότι θα υπάρχει σημαντική διαφοροποίηση στην συμπεριφορά των δανειοληπτών. Παρατηρείται ότι ο έγγαμος πελάτης είναι σημαντικά συνεπέστερος σε σχέση με τον άγαμο ή τον διαζευγμένο οι οποίοι έχουν παρόμοια συμπεριφορά.

Όσον αφορά την κατηγορία «ΑΛΛΟ» εκτιμάται ότι οι περισσότεροι των πελατών που ανήκουν σ' αυτήν είναι λίγο πριν την ένταξή τους στην κατηγορία των εγγάμων, δηλαδή είναι είτε αρραβωνιασμένοι, είτε πρόκειται να παντρευτούν στο εγγύς μέλλον και έτσι δικαιολογείται η παρόμοια συμπεριφορά τους με τους έγγαμους.

#### Ανεξάρτητη μεταβλητή X6: Πελάτης της τράπεζας (διακριτή)

Η μεταβλητή αυτή εκφράζει κατά πόσο ο δανειολήπτης ήταν ήδη πελάτης της τράπεζας και αναμένεται να έχει πολύ σημαντική συνεισφορά στην διάκριση της συμπεριφοράς των πιστούχων. Ειδικότερα εφόσον η τράπεζα έχει εγκρίνει ένα

καταναλωτικό δάνειο σε άτομο που είναι ήδη πελάτης σημαίνει ότι δεν έχει πέσει στην αντίληψη της κάτι αρνητικό στην συμπεριφορά του. Η γνώση της προηγούμενης συμπεριφοράς ενός πελάτη μπορεί να αποτρέψει μια πιστοδότηση υψηλού κινδύνου. Αντίθετα, στην περίπτωση που ο δανειολήπτης δεν είναι ήδη πελάτης της τράπεζας ο κίνδυνος έγκρισης ενός επισφαλούς δανείου αναμένεται να είναι υψηλότερος.

Η βάση δεδομένων της τράπεζας δίνει στην μεταβλητή αυτή την τιμή 1 εφόσον ο δανειολήπτης είναι ήδη πελάτης και την τιμή 2 στην περίπτωση που είναι νέος πελάτης.

Η συμπεριφορά των δανειοληπτών παρουσιάζεται στον παρακάτω πίνακα :

	ΝΑΙ ΠΕΛΑΤΗΣ 1	ΌΧΙ ΠΕΛΑΤΗΣ 2
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	14420	1366
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	12,327	1,070
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	2,093	296
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	85.49%	78.33%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	14.51%	21.67%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	16.98%	27.66%

Παρατηρείται ότι υπάρχει μεγάλη διαφορά στη συμπεριφορά των δανειοληπτών που ήταν ήδη πελάτες της τράπεζας έναντι των νέων πελατών. Οι ήδη πελάτες είναι σημαντικά πιο συνεπείς στην αποπληρωμή από τους νέους. Σύμφωνα με την προηγούμενη ανάλυση παρατήρηση αυτή ήταν αναμενόμενη.

#### Ανεξάρτητη μεταβλητή X7: Χειρότερη καταχώρηση Τειρεσία (διακριτή)

Η μεταβλητή αυτή υπό κανονικές συνθήκες έχει σημαντικό ρόλο στην έγκριση του δανείου, δεδομένου ότι δύσκολα εγκρίνεται δάνειο σε άτομο που έχει δυσμενή καταχώριση στον Τειρεσία. Έτσι, δεν αναμένεται να υπάρχει μεγάλο ποσοστό δανειοληπτών με δυσμενή καταχώριση στον Τειρεσία εφόσον η μελέτη γίνεται σε χαρτοφυλάκιο εγκεκριμένων δανείων.

Η συνεισφορά της μεταβλητής αυτής στην εκτίμηση της πιθανότητας αθέτησης εξαρτάται μόνο από το ποσοστό των δανείων που έχουν δοθεί σε πελάτες με δυσμενή

καταχώρηση στον Τειρεσία καθώς η σημαντικά διαφορετική συμπεριφορά των δανειοληπτών θεωρείται δεδομένη. Ειδικότερα το δάνειο που έχει εγκριθεί σε πελάτη ο οποίος έχει δυσμενή καταχώρηση στον Τειρεσία αναμένεται να είναι υψηλότερου κινδύνου καθώς είναι ήδη γνωστό ότι ο δανειολήπτης έχει ήδη υπάρξει ασυνεπής στο παρελθόν.

Σύμφωνα με την βάση δεδομένων της τράπεζας η μεταβλητή παίρνει την τιμή 1 εφόσον δεν υπάρχει δυσμενής καταχώρηση στον Τειρεσία για τον δανειολήπτη και παίρνει την τιμή 2 εφόσον υπάρχει δυσμενής καταχώρηση και τις τιμές 3,4...21,22 ανάλογα με το είδος της δυσμενούς καταχώρησης.

Επειδή το ποσοστό των δανείων που έχουν εγκριθεί έχοντας γνώση δυσμενούς καταχώρησης στον Τειρεσία είναι αρκετά μικρό θα περιοριστούν οι τιμές της μεταβλητής στις εξής : 1 εφόσον δεν υπάρχει δυσμενής καταχώρηση και 2 εφόσον υπάρχει. Έτσι το ποσοστό των δυσμενών καταχωρήσεων αποκτά άλλη βαρύτητα καθώς δεν διασπάται σε 22 κλάσεις.

Η συμπεριφορά των δανειοληπτών παρουσιάζεται στον παρακάτω πίνακα :

	ΟΧΙ ΔΥΣΜΕΝΗΣ ΚΑΤΑΧΩΡΗΣΗ 1	ΝΑΙ ΔΥΣΜΕΝΗΣ ΚΑΤΑΧΩΡΗΣΗ 2
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	15102	684
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	12,870	527
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	2,232	157
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	85.22%	77.05%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	14.78%	22.95%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	17.34%	29.79%

Επιβεβαιώνεται και στην περίπτωση αυτή, ότι η διαφορά στην συμπεριφορά των δανειοληπτών είναι σημαντική. Παρατηρείται επίσης ότι ο αριθμός των δανείων που έχουν εγκριθεί σε πελάτες που είχαν δυσμενή καταχώρηση είναι πολύ μικρός.

Αξίζει να σημειωθεί ότι αρκετές τράπεζες δεν εγκρίνουν δάνεια σε πελάτες που έχουν δυσμενή καταχώρηση στον Τειρεσία ανεξάρτητα από την πιθανότητα αθέτησης που θα εκτιμήσει τον υπόδειγμα της τράπεζας. Δηλαδή, η μη καταχώρηση στον Τειρεσία αποτελεί ένα κρίσιμο κριτήριο που πρέπει να ικανοποιείται προκειμένου να εγκριθεί το δάνειο.

### Ανεξάρτητη μεταβλητή X8 : Έτη από την χειρότερη καταχώρηση Τειρεσία (διακριτή)

Εκτός από το αν υπάρχει δυσμενής καταχώρηση στον Τειρεσία πολύ σημαντική είναι και η μεταβλητή που εκφράζει το χρονικό διάστημα από την τελευταία καταχώρηση. Ειδικότερα, όσο πιο παλιά είναι η καταχώρηση τόσο μεγαλύτερη πιθανότητα υπάρχει ο δανειολήπτης να έχει διαμορφώσει τέτοιες συνθήκες ώστε να είναι περισσότερο φερέγγυος και συνεπής σε σχέση με το παρελθόν. Αντίθετα, όταν η δυσμενής καταχώρηση είναι πρόσφατη είναι πολύ μεγάλη η πιθανότητα ο πελάτης να έχει παραμείνει ασυνεπής.

Αναμένεται επομένως ότι, ο αριθμός των δανείων σε άτομα με δυσμενή καταχώρηση στον Τειρεσία θα είναι μικρότερος στην περίπτωση που η καταχώρηση είναι πρόσφατη καθώς τέτοιες καταχωρήσεις θεωρούνται υψηλής επικινδυνότητας και δεν εγκρίνονται εύκολα.

Σύμφωνα με την βάση δεδομένων οι τιμές που παίρνει η μεταβλητή αυτή είναι: 1 στην περίπτωση που δεν υπάρχει δυσμενής καταχώρηση, 2 εφόσον υπάρχει και είναι πιο πρόσφατη των δύο ετών και 3 εφόσον η καταχώρηση είναι παλαιότερη των τριών ετών.

Όλα τα παραπάνω επιβεβαιώνονται και από τον παρακάτω πίνακα που δείχνει την συμπεριφορά των δανειοληπτών :

	ΟΧΙ ΔΥΣΜΕΝΗ 1	ΜΙΚΡΟΤΕΡΗ ΤΩΝ 2 ΕΤΩΝ 2	3 ΕΤΩΝ ΚΑΙ ΠΑΝΩ 3
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	15102	119	565
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	12,870	74	453
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	2,232	45	112
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	85.22%	62.18%	80.18%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	14.78%	37.82%	19.82%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	17.34%	60.81%	24.72%

Παρατηρείται μεγάλη διαφοροποίηση στη συμπεριφορά μεταξύ των δανειοληπτών με πρόσφατη δυσμενή καταχώρηση και αυτών με καταχώρηση μεγαλύτερη των τριών ετών. Αναμένεται η μεταβλητή αυτή να έχει σημαντική συνεισφορά στην εκτίμηση της πιθανότητας αθέτησης.

Αξίζει να σημειωθεί ότι η μεταβλητή αυτή έρχεται να συμπληρώσει την X7 καθώς δεν προσδίδει καμία νέα πληροφορία για τα άτομα που δεν έχουν δυσμενή

καταχώρηση. Αντίθετα, κάνει έναν ουσιαστικό διαχωρισμό στα άτομα με δυσμενή καταχώρηση ανάλογα με την παλαιότητα αυτής.

#### Ανεξάρτητη μεταβλητή X9: Διάρκεια δανείου (διακριτή)

Η διάρκεια του δανείου γενικά αποτελεί σημαντικό παράγοντα στον καθορισμό της συμπεριφοράς των δανειοληπτών και χρήζει ιδιαίτερης ανάλυσης. Κατά γενική ομολογία, ένα δάνειο όσο πιο μεγάλη διάρκεια έχει τόσο πιο επισφαλές είναι καθώς υπάρχει αυξημένη πιθανότητα να μεταβληθούν κάποιες παράμετροι που επηρεάζουν την συνέπεια του δανειολήπτη όσον αφορά την αποπληρωμή. Τέτοιες παράμετροι που ενδέχεται να μεταβληθούν μακροπρόθεσμα μπορεί να είναι η οικονομική κατάσταση του δανειολήπτη ή ο θάνατος του δανειολήπτη κλπ.

Επίσης, είναι σύνηθες το φαινόμενο κάποια άτομα να υπερεκτιμούν τις οικονομικές τους δυνατότητες, και βασιζόμενοι στην μικρή μηνιαία δόση που τους προσφέρουν τα μακροπρόθεσμα καταναλωτικά δάνεια, να αποκτούν με πίστωση καταναλωτικά αγαθά διαρκείας για να καλύψουν τις παρούσες ανάγκες. Ένα αντιπροσωπευτικό παράδειγμα είναι τα δάνεια για την αγορά αυτοκινήτου που υπάρχει η δυνατότητα για τον πελάτη να πάρει εύκολα ένα υψηλό χρηματικό ποσό με πίστωση, με αρκετά μεγάλο ορίζοντα αποπληρωμής, μεγαλύτερη των πενήντα μηνών. Όπως είναι αναμενόμενο σε αρκετές από αυτές τις περιπτώσεις υπάρχει αθέτηση των υποχρεώσεων των δανειοληπτών έναντι της τράπεζας.

Στη βάση δεδομένων της τράπεζας έχουν χωριστεί τα δάνεια σε τρεις κλάσεις ανάλογα με την διάρκεια και οι τιμές που μπορεί να πάρει η μεταβλητή είναι 11 εφόσον η διάρκεια του δανείου είναι από 0 έως 35 μήνες, 12 για διάρκεια από 36 έως 49 και 13 για δάνεια με ορίζοντα αποπληρωμής μεγαλύτερο των 50 μηνών.

Η συμπεριφορά των δανειοληπτών φαίνεται στον παρακάτω πίνακα :

	0-35 ΜΗΝΕΣ 11	36-49 ΜΗΝΕΣ 12	> 50 ΜΗΝΕΣ 13
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	1349	2290	12147
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	1,248	2,110	10,039
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	101	180	2,108
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	92.51%	92.14%	82.65%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	7.49%	7.86%	17.35%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	8.09%	8.53%	21.00%

Όπως αναμενόταν τα δάνεια με μεγαλύτερο ορίζοντα αποπληρωμής αποδεικνύονται πιο επισφαλής στο χαρτοφυλάκιο της τράπεζας. Η συμπεριφορά των πελατών διαφέρει σημαντικά για δάνεια με διάρκεια μεγαλύτερη των 50 μηνών, γεγονός που αποτελεί ένδειξη ότι η μεταβλητή μπορεί να είναι σημαντική στην εκτίμηση της πιθανότητας αθέτησης.

#### Ανεξάρτητη μεταβλητή X10: Μηνιαία δόση προς μηνιαίο εισόδημα (διακριτή).

Η μηνιαία δόση προς το μηνιαίο εισόδημα, υπό κανονικές συνθήκες, αποτελεί ένα σημαντικό κριτήριο για την έγκριση ενός δανείου. Συνήθως, ως ανώτατο επιτρεπόμενο όριο της μεταβλητής αυτής είναι το 40%. Η μικρή δόση φυσιολογικά επιτρέπει στον δανειολήπτη να είναι συνεπής στις υποχρεώσεις του.

Η μεταβλητή αυτή έχει δύο σημαντικά μειονεκτήματα. Το πρώτο είναι ότι το δηλωθέν εισόδημα στην Ελλάδα δεν είναι πάντα το πραγματικό οπότε αρκετοί δανειολήπτες δεν κατατάσσονται στην σωστή κατηγορία και έτσι προκύπτουν λάθος συμπεράσματα. Το δεύτερο μειονέκτημα είναι ότι η μικρή δόση είναι άρρηκτα συνδεδεμένη με τον μακρύ ορίζοντα αποπληρωμής χαρακτηριστικό το οποίο είναι ένδειξη επισφάλειας.

Οι τιμές που μπορεί να πάρει η μεταβλητή αυτή σύμφωνα με την βάση δεδομένων της τράπεζας είναι: 1 για > 9%, 2 για 10%-19%, 3 για 20%-29%, 4 για 30-39%, 5 για >40%

Ο παρακάτω πίνακας που δείχνει την συμπεριφορά των δανειοληπτών επιβεβαιώνει ότι η μεταβλητή αυτή δεν αναμένεται να έχει σημαντική συνεισφορά στην εκτίμηση της πιθανότητας αθέτησης καθώς η συμπεριφορά των δανειοληπτών δεν διαφοροποιείται σημαντικά.

	ΕΩΣ 9% 1	10% - 19% 2	20%-29% 3	30%-39% 4	> 40% 5
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	7586	5857	1341	419	583
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	6396	4,976	1,164	362	499
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	1190	881	177	57	84
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	84.31%	84.96%	86.80%	86.40%	85.59%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	15.69%	15.04%	13.20%	13.60%	14.41%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	18.61%	17.70%	15.21%	15.75%	16.83%

Συγκρίνοντας τους πίνακες των μεταβλητών X9 και X10 παρατηρείται ότι το μεγαλύτερο μέρος των δανείων είναι μακροπρόθεσμα με χαμηλή δόση. Ειδικότερα, το 85% των δανείων έχει μηνιαία δόση έως 19 % του μηνιαίου μισθού και το 77% των δανείων έχουν χρονικό ορίζοντα μεγαλύτερο των 50 μηνών. Επίσης, γίνεται αντιληπτό ότι τα περισσότερα επισφαλή καταναλωτικά δάνεια για το χαρτοφυλάκιο της τράπεζας είναι τα μακροπρόθεσμα με μικρή δόση σε σχέση με τον μηνιαίο μισθό.

#### Ανεξάρτητη μεταβλητή X11: Φύλο (διακριτή)

Το φύλο των δανειοληπτών ενδεχομένως να μην έχει σημαντική συνεισφορά στην διάκριση της συμπεριφοράς τους. Η μεταβλητή αυτή παλαιότερα θα ήταν σχεδόν σίγουρα ένας σημαντικός παράγοντας στην εκτίμηση της πιθανότητας αθέτησης γεγονός που οφείλεται στην κοινωνική και οικονομική διάκριση που υπήρχε μεταξύ των δύο φύλων. Τη σημερινή εποχή τα δύο φύλα δεν διαφοροποιούνται τόσο σημαντικά, καθώς τα κοινωνικά και οικονομικά χαρακτηριστικά τους έχουν εξομοιωθεί. Ως αποτέλεσμα, αναμένεται η συμπεριφορά των δανειοληπτών να μην διαφοροποιείται σημαντικά ανάλογα το φύλο τους.

Σύμφωνα με την βάση δεδομένων της τράπεζας, οι τιμές που παίρνει η μεταβλητή αυτή είναι 1 αν ο δανειολήπτης είναι άνδρας και 0 αν είναι γυναίκα.

Από τον παρακάτω πίνακα επιβεβαιώνεται ότι τα δύο φύλα δεν διαφέρουν σημαντικά ως προς την συμπεριφορά τους.



ΚΛΑΣΕΙΣ ΤΙΜΩΝ	ΓΥΝΑΙΚΑ 0	ΑΝΔΡΑΣ 1
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	4943	10843
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	4,167	9,230
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	776	1,613
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	84.30%	85.12%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	15.70%	14.88%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	18.62%	17.48%

#### Ανεξάρτητη μεταβλητή X12: Αριθμός παιδιών (διακριτή)

Ο αριθμός των παιδιών ενδέχεται να έχει συνεισφορά στη διάκριση της συμπεριφοράς των δανειοληπτών. Ο δανειολήπτης που έχει παιδιά ενδεχομένως να είναι πιο υπεύθυνος καθώς έχει την ευθύνη τρίτων προσώπων. Πιθανώς να θέλει να προστατεύσει τα παιδιά του από τα αποτελέσματα της δικής του ασυνέπειας.

Οι τιμές που παίρνει η μεταβλητή αυτή στη βάση δεδομένων είναι απλά ο αριθμός των παιδιών που έχει κάθε δανειολήπτης. Ειδικότερα, παίρνει την τιμή 1 εφόσον ο δανειολήπτης δεν έχει παιδιά, 2 εφόσον έχει μέχρι δύο παιδιά και 3 στην περίπτωση που ο δανειολήπτης είναι πολύτεκνος, δηλαδή να έχει από τρία παιδιά και πάνω.

	ΟΧΙ ΠΑΙΔΙΑ 1	1 - 2 ΠΑΙΔΙΑ 2	> 3 ΠΑΙΔΙΑ 3
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	7750	6488	1548
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	6460	5607	1330
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	1290	881	218
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	83.35%	86.42%	85.92%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	16.65%	13.58%	14.08%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	19.97%	15.71%	16.39%

#### Ανεξάρτητη μεταβλητή X13: Είδος προσωπικού εγγράφου (διακριτή)

Το έγγραφο με το οποίο ο δανειολήπτης αποδεικνύει την ταυτότητα του ενδέχεται να είναι σημαντικό χαρακτηριστικό για την εκτίμηση της πιθανότητας αθέτησης. Τα αποδεκτά προσωπικά έγγραφα από την τράπεζα είναι η αστυνομική ταυτότητα, το διαβατήριο, η ταυτότητα των ενόπλων δυνάμεων καθώς και των σωμάτων ασφαλείας και τέλος η ταυτότητα των αλλοδαπών. Ενδεχομένως να υπάρχει διάκριση στην συμπεριφορά αυτών των μονάδων.

Στους δανειολήπτες που έχουν προσκομίσει ταυτότητα η διαβατήριο που είναι τα συνηθέστερα προσωπικά έγγραφα θα δοθεί η τιμή 1 και στις υπόλοιπες κατηγορίες θα δοθεί η τιμή 2. Με αυτόν τον τρόπο θα ελεγχθεί κατά πόσο υπάρχει αυξημένη επικινδυνότητα στην περίπτωση που δεν προσκομίζεται αστυνομική ταυτότητα ή διαβατήριο.

Η συμπεριφορά των δανειοληπτών στις δύο κλάσεις φαίνεται στον παρακάτω πίνακα.

ΚΛΑΣΕΙΣ ΤΙΜΩΝ	ΤΑΥΤΟΤΗΤΑ /ΔΙΑΒΑΤΗΡΙΟ 1	ΛΟΙΠΟΙ 2
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	15322	464
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	12985	412
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	2337	52
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	84.75%	88.79%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	15.25%	11.21%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	18.00%	12.62%

#### Ανεξάρτητη μεταβλητή X14: Ετήσιο εισόδημα (συνεχής)

Το ετήσιο εισόδημα είναι μία μεταβλητή που δεν αναμένεται να είναι σημαντική στην εκτίμηση της πιθανότητας αθέτησης. Το παράδοξο αυτό συμβαίνει γιατί το δηλωθέν ετήσιο εισόδημα πολλές φορές δεν είναι το πραγματικό λόγω του μεγάλου ποσοστού φοροδιαφυγής στην Ελλάδα. Γενικά πάντως θα έπρεπε να ήταν η κύρια μεταβλητή για την έκφραση της οικονομικής κατάστασης του δανειολήπτη, όμως παραδόξως υπάρχουν μεταβλητές, όπως ο αριθμός των παιδιών, που εκφράζουν πιο έγκυρα την οικονομική κατάσταση των πελατών.

Χωρίζοντας το δείγμα σε κλάσεις ανάλογα με το δηλωθέν ετήσιο εισόδημα, όπως φαίνεται στον παρακάτω πίνακα, γίνεται αντιληπτό ότι δεν υπάρχει διαφοροποίηση στη συμπεριφορά των δανειοληπτών με βάση το εισόδημα τους.

	< 12000	12000 έως 30000	30000 έως 75000	> 75000
ΠΛΗΘΟΣ ΑΤΟΜΩΝ	4499	8968	2113	206
ΠΛΗΘΟΣ ΣΥΝΕΠΩΝ	3822	7611	1789	175
ΠΛΗΘΟΣ ΑΣΥΝΕΠΩΝ	677	1357	324	31
ΠΟΣΟΣΤΟ(%) ΣΥΝΕΠΩΝ	84.95%	84.87%	84.67%	84.95%
ΠΟΣΟΣΤΟ(%) ΑΣΥΝΕΠΩΝ	15.05%	15.13%	15.33%	15.05%
ODDS(ΑΣΥΝΕΠΩΝ/ΣΥΝΕΠΩΝ)	17.71%	17.83%	18.11%	17.71%

### **Κωδικοποίηση των ερμηνευτικών μεταβλητών**

Η κωδικοποίηση των διακριτών ερμηνευτικών μεταβλητών θα γίνει σύμφωνα με τους *Crook et al. (1992)* όπως έχει προαναφερθεί στην ενότητα 3. Όσον αφορά τις συνεχείς ερμηνευτικές μεταβλητές θα διατηρηθεί η κωδικοποίηση που έχει προαναφερθεί κατά την ανάλυση τους.

Αξίζει να σημειωθεί ότι κάποιες μεταβλητές που αναμένεται να είναι σημαντικές ενδέχεται να μην συμπεριληφθούν στην συνάρτηση διαχωρισμού καθώς μπορεί να έχουν υψηλή συσχέτιση με μεταβλητές που είναι πιο σημαντικές. Το φαινόμενο ονομάζεται *πολλυσυγγραμμικότητα (multicollinearity)*. Ο πίνακας συσχετίσεων παρουσιάζεται στο Παράρτημα 2.

#### **4.4 Εφαρμογή του υποδείγματος Λογιστικής Παλινδρόμησης**

Όπως έχει προαναφερθεί στην ανάλυση των μεταβλητών X5 και X6 υπάρχουν συνολικά 76 περιπτώσεις δανείων στις οποίες έχει δοθεί λανθασμένη τιμή. Έτσι από το δείγμα διαγράφονται 76 δάνεια. Η ανάπτυξη του υποδείγματος Logit θα γίνει σε δείγμα 15710 δανείων, που οι ασυνεπείς δανειολήπτες είναι 2378.

Το δείγμα των 15710 δανείων χωρίζεται τυχαία σε δύο μέρη, το *δείγμα εκτίμησης* των συντελεστών των ερμηνευτικών μεταβλητών και το *δείγμα επικύρωσης* του υποδείγματος. Η επικύρωση είναι σημαντικό να γίνεται σε δείγμα διαφορετικό του δείγματος εκτίμησης του υποδείγματος (out of sample validation).

Το δείγμα εκτίμησης αποτελείται από τα  $\frac{2}{3}$  του δείγματος που αντιστοιχούν σε 10522 δάνεια θα χρησιμοποιηθούν για την ανάπτυξη του υποδείγματος Logit. Το δείγμα επικύρωσης είναι το  $\frac{1}{3}$  του δείγματος που αντιστοιχεί σε 5188 δάνεια θα χρησιμοποιηθεί για την επικύρωση του. Στο δείγμα εκτίμησης οι ασυνεπείς δανειολήπτες είναι 1583 με ποσοστό 15,04% ενώ στο δείγμα επικύρωσης οι ασυνεπείς δανειολήπτες είναι 795 δηλαδή 15,32%.

#### 4.4.1 Κατασκευή της συνάρτησης διαχωρισμού

Η κατασκευή της συνάρτησης διαχωρισμού του υποδείγματος Logit γίνεται με τη βοήθεια του προγράμματος Eviews. Αναζητείται ο στατιστικά σημαντικότερος συνδυασμός από τις 14 ερμηνευτικές μεταβλητές για την κατασκευή της συνάρτησης:

$$\text{logit}(P(Y_i = 1)) = \sum_{k=0}^N b_{ik}X_{ik} \quad (5.1)$$

Η κατασκευή της συνάρτησης διαχωρισμού γίνεται με την εφαρμογή των μεθόδων bottom-up (stepwise) και top-down (inverse step wise) για διάστημα εμπιστοσύνης 95% δηλαδή  $\alpha = 0,05$ . Όπως αναμενόταν και οι δύο μέθοδοι δίνουν το ίδιο αποτέλεσμα:

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.097812	0.508985	-4.121564	0.0000
X1	0.043452	0.016727	2.597729	0.0094
X2	-0.020037	0.002587	-7.744944	0.0000
X3	-1.005937	0.479575	-2.097561	0.0359
X5	-0.556438	0.132397	-4.202809	0.0000
X6	-0.885579	0.179594	-4.930995	0.0000
X8	-1.118201	0.166914	-6.699271	0.0000
X9	-0.876596	0.088331	-9.923960	0.0000
X13	-1.664859	0.548184	-3.037047	0.0024
Mean dependent var	0.150447	S.D. dependent var	0.357526	
S.E. of regression	0.351629	Akaike info criterion	0.814827	
Sum squared resid	1299.858	Schwarz criterion	0.821038	
Log likelihood	-4277.804	Hannan-Quinn criter.	0.816924	
Restr. log likelihood	-4455.889	Avg. log likelihood	-0.406558	
LR statistic (8 df)	356.1705	McFadden R-squared	0.039966	
Probability(LR stat)	0.000000			

(βλ. Παράρτημα 3)

Το υπόδειγμα δίνεται από την εξίσωση:

$$\text{logit}(P(Y_i = 1)) = -2.097812 + 0.043452X_1 - 0.020037X_2 - 1.005937X_3 - 0.556438X_5 - 0.885579X_6 - 1.118201X_8 - 0.876596X_9 - 1.664859X_{13}$$

Οι ερμηνευτικές μεταβλητές που συμπεριλήφθησαν είναι:

X1: Ημερομηνία αίτησης (ηη/μμ/εε)

X2: Ημερομηνία γέννησης (ηη/μμ/εε)

X3: Κάτοχος πιστωτικής κάρτας

X5: Οικογενειακή κατάσταση

X6: Πελάτης της τράπεζας

X8: Έτη από την πιο πρόσφατη καταχώρηση Τειρεσία.

X9: Διάρκεια δανείου

X13: Είδος προσωπικού εγγράφου

Οι *Crook et al. (1992)* σε υπόδειγμα εκτίμησης της πιθανότητας αθέτησης έχουν συμπεριλάβει τις μεταβλητές X2, X3, X5 και X6.

Ο *Schreiner (2004)* έχει συμπεριλάβει στο υπόδειγμά του τη μεταβλητή X1.

Η *Vigano (1993)* έχει συμπεριλάβει στο υπόδειγμα της τις X2, X3, X5, X6, X9.

Οι *Dinh και Kleimeier (2007)* στο υπόδειγμά τους έχουν συμπεριλάβει τις μεταβλητές X5, X6, X9.

Οι *Yang et al. (2009)* στο υπόδειγμά τους έχουν συμπεριλάβει τις μεταβλητές X2, X9.

Παρατηρείται ότι οι περισσότερες από τις μεταβλητές που συμπεριλήφθησαν στο υπόδειγμα εκτίμησης της πιθανότητας αθέτησης έχουν συμπεριληφθεί και σε υποδείγματα αρκετών ερευνητών, γεγονός που ισχυροποιεί την σημαντικότητά τους.

#### 4.4.2 Στατιστική σημαντικότητα του υποδείγματος

Όσον αφορά την στατιστική σημαντικότητα το υπόδειγμα είναι επαρκές διότι :

1. Όλοι οι συντελεστές είναι στατιστικά σημαντικοί καθώς  $|Z_{statistic}| > 1,96$  ή ισοδύναμα όλα τα  $p_{value} < 0,05$ .

2. Το κριτήριο Likelihood Ratio δείχνει ότι το υπόδειγμα είναι στατιστικά σημαντικό καθώς:  $LR = 356,1705 > 15,51 = \chi_8^2(0,05)$  ή ισοδύναμα  $p_{value} = 0 < 0,05$ .
3. Όσον αφορά στην ετεροσκεδαστικότητα, τα p-values των squared residuals είναι όλα μεγαλύτερα του 0,05 οπότε δεν απορρίπτεται η υπόθεση ότι η συσχέτιση των squared residuals είναι μηδενική (βλ. Παράρτημα 4).
4. Το κριτήριο Hosmer-Lemeshow για  $k=10$  δίνει τα παρακάτω αποτελέσματα:

	Quantile of Risk		Dep=0		Dep=1		Total Obs	H-L Value
	Low	High	Actual	Expect	Actual	Expect		
1	0.0287	0.0672	999	995.106	53	56.8936	1052	0.28171
2	0.0672	0.0986	967	964.767	85	87.2333	1052	0.06235
3	0.0986	0.1159	955	938.367	97	113.633	1052	2.72939
4	0.1159	0.1298	914	922.895	138	129.105	1052	0.69865
5	0.1298	0.1428	906	909.641	147	143.359	1053	0.10702
6	0.1428	0.1559	902	895.029	150	156.971	1052	0.36391
7	0.1559	0.1728	861	879.376	191	172.624	1052	2.34023
8	0.1728	0.1970	846	858.515	206	193.485	1052	0.99198
9	0.1970	0.2377	831	824.610	221	227.390	1052	0.22910
10	0.2377	0.6675	758	750.694	295	302.306	1053	0.24770
Total			8939	8939.00	1583	1583.00	10522	8.05204
H-L Statistic:			8.0520		Prob. Chi-Sq(8)		0.4284	

Παρατηρείται ότι δεν μπορεί να υποτεθεί ότι το υπόδειγμα δεν προσαρμόζεται σωστά στα δεδομένα διότι  $HL = 8,0520 < 15,51 = \chi_8^2(0,05)$ . (βλ.Παράρτημα 5)

#### 4.4.3 Σχολιασμός των συντελεστών

Η εκτίμηση της πιθανότητας αθέτησης δίνεται από τον τύπο :

$$PD = \Pr(Y_i = 1) = \frac{1}{1 + e^{-\sum_{k=0}^N \hat{\beta}_k X_{ik}}} \quad (5.2)$$

Η συμπεριφορά του PD με βάση τις τιμές του  $score = \sum_{k=0}^N \hat{\beta}_k X_{ik}$  είναι:

$$\begin{cases} PD \rightarrow 1 \text{ καθώς } score \rightarrow \infty \\ PD = 1/2 \text{ καθώς } score = 0 \\ PD \rightarrow 0 \text{ καθώς } score \rightarrow -\infty \end{cases}$$

Από τη συμπεριφορά του PD προκύπτει ότι όσο μεγαλύτερο είναι το score για έναν δανειολήπτη τόσο μεγαλύτερη πιθανότητα αθέτησης του προσδίδεται. Έτσι για μια



μεταβλητή με θετικό συντελεστή οι μεγάλες τιμές είναι ένδειξη επικινδυνότητας. Το αντίθετο συμβαίνει στην περίπτωση που ο συντελεστής της τυχαίας μεταβλητής είναι αρνητικός. Όπως παρατηρείται στον παρακάτω πίνακα οι περισσότεροι συντελεστές είναι αρνητικοί:

Variable	Coefficient
C	-2.097812
X1	0.043452
X2	-0.020037
X3	-1.005937
X5	-0.556438
X6	-0.885579
X8	-1.118201
X9	-0.876596
X13	-1.664859

#### *Συνεχείς ερμηνευτικές μεταβλητές*

Η μεταβλητή X1 είναι η ημερομηνία αίτησης και εκφράζει τους μήνες που έχουν περάσει από την ημερομηνία αίτησης μέχρι τις 15/7/2009. Η X1 έχει θετικό συντελεστή.

Δηλαδή, όσο πιο παλιά είναι η αίτηση τόσο μεγαλύτερη και η πιθανότητα αθέτησης. Το φαινόμενο αυτό είναι λογικό διότι σε μεγαλύτερο χρονικό διάστημα υπάρχουν περισσότερες χρονικές στιγμές που πιθανώς να υπάρξει αθέτηση. Επίσης, η πιθανότητα αθέτησης είναι αύξουσα σε σχέση με το χρόνο για τα πρώτα χρόνια από την έγκριση του δανείου.

Η μεταβλητή X2 είναι η ηλικία σε ακέραια έτη μέχρι την 15/7/2009 του δανειολήπτη και έχει αρνητικό συντελεστή, δηλαδή όσο πιο μικρός είναι ο δανειολήπτης τόσο πιο επικίνδυνος θεωρείται.

#### *Διακριτές ερμηνευτικές μεταβλητές*

Η κωδικοποίηση που επιλέχθηκε για τις διακριτές ερμηνευτικές μεταβλητές (*Crook et al., 1992*) προσδίδει στις κλάσεις με τους πιο ασυνεπείς δανειολήπτες μικρότερη τιμή (συνήθως αρνητική) σε σχέση με τις τιμές των κλάσεων με τους συνεπέστερους

δανειολήπτες. Έτσι όπως αναμενόταν, λόγω της κωδικοποίησης, οι συντελεστές των διακριτών τυχαιών μεταβλητών είναι αρνητικοί.

Παρακάτω παρατίθενται οι πίνακες με τις τιμές που παίρνουν οι κλάσεις των διακριτών μεταβλητών που συμμετέχουν στο υπόδειγμα σύμφωνα με την κωδικοποίηση των *Crook et al. (1992)*. Οι κλάσεις στους πίνακες είναι με σειρά επικινδυνότητας ξεκινώντας με τους πιο ασυνεπείς δανειολήπτες. Παρατηρείται ότι όσο πιο ασυνεπείς είναι οι δανειολήπτες σε μία κλάση τόσο πιο μικρή τιμή τους προσδίδεται, γεγονός που δικαιολογεί τους αρνητικούς συντελεστές των μεταβλητών.

X3: Κάτοχος πιστωτικής κάρτας

ΚΛΑΣΕΙΣ ΤΙΜΩΝ	ΟΧΙ/ΔΕΝ ΑΠΑΝΤΗΣΕ	ΚΑΤΟΧΟΣ ΠΙΣΤΩΤΙΚΗΣ
ΚΩΔΙΚΟΠΟΙΗΣΗ	-0.054670702	0.064151469

X5: Οικογενειακή κατάσταση

ΚΛΑΣΕΙΣ ΤΙΜΩΝ	ΑΓΑΜΟΣ	ΔΙΑΖ/ΝΟΣ	ΆΛΛΟ	ΕΓΓΑΜΟΣ
ΚΩΔΙΚΟΠΟΙΗΣΗ	-0.35617	-0.333901	0.0451307	0.1475433

X6: Πελάτης της τράπεζας

ΚΛΑΣΕΙΣ ΤΙΜΩΝ	ΌΧΙ ΠΕΛΑΤΗΣ	ΝΑΙ ΠΕΛΑΤΗΣ
ΚΩΔΙΚΟΠΟΙΗΣΗ	-0.4391	0.0490

X8 : Έτη χειρότερης καταχώρηση Τειρεσία

ΚΛΑΣΕΙΣ ΤΙΜΩΝ	ΜΙΚΡΟΤΕΡΗ ΤΩΝ 2 ΕΤΩΝ	3 ΕΤΗ ΚΑΙ ΠΑΝΩ	ΌΧΙ ΔΥΣΜΕΝΗ
ΚΩΔΙΚΟΠΟΙΗΣΗ	-1.22675333	-0.326763	0.027845044

X9: Διάρκεια δανείου

ΚΛΑΣΕΙΣ ΤΙΜΩΝ	50 ΚΑΙ ΑΝΩ	36-49	0-35
ΚΩΔΙΚΟΠΟΙΗΣΗ	-0.16341806	0.737330443	0.790021099

X13: Είδος προσωπικού εγγράφου

ΚΛΑΣΕΙΣ ΤΙΜΩΝ	ΤΑΥΤΟΤΗΤΑ /ΔΙΑΒΑΤΗΡΙΟ	ΛΟΙΠΟΙ
ΚΩΔΙΚΟΠΟΙΗΣΗ	-0.009229143	0.345623698



#### 4.4.4 Έλεγχος της αποτελεσματικότητας του υποδείγματος

Μετά τον υπολογισμό των συντελεστών των ερμηνευτικών μεταβλητών υπολογίζεται η εκτίμηση της πιθανότητας αθέτησης για κάθε δανειολήπτη και στο δείγμα εκτίμησης και στο δείγμα επικύρωσης. Ειδικότερα, η πιθανότητα αυτή υπολογίζεται αντικαθιστώντας στην εξίσωση  $PD = \frac{1}{1 + e^{-\sum_{k=0}^N \hat{\beta}_k X_{ik}}}$  τις τιμές των μεταβλητών για κάθε δανειολήπτη χωριστά. Στην συνέχεια, γίνεται ταξινόμηση των δανειοληπτών με βάση την πιθανότητα αθέτησης, έτσι ώστε να υπολογιστεί το cut off point.

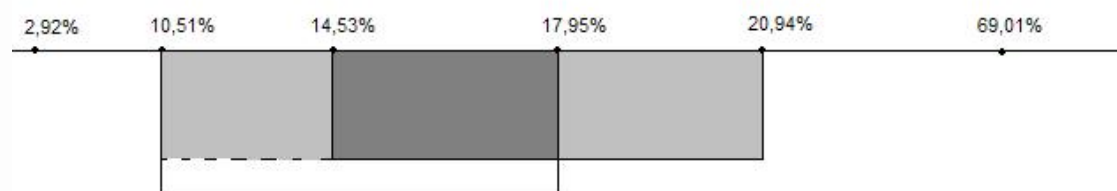
Ο προσδιορισμός του cut off point είναι πολύ σημαντικός για την αποτελεσματικότητα του υποδείγματος διότι από την τιμή του εξαρτώνται και τα σφάλματα τύπου I και II. Το cut off point προσδιορίζεται στο δείγμα εκτίμησης.

Με βάση την τιμή του cut off point η αποτελεσματικότητα του υποδείγματος ελέγχεται αρχικά στο δείγμα εκτίμησης και στη συνέχεια επικυρώνεται στο δείγμα επικύρωσης.

##### **Το Δείγμα Εκτίμησης του υποδείγματος**

Το δείγμα εκτίμησης αποτελείται από 10522 απαιτήσεις. Οι ασυνεπείς δανειολήπτες είναι 1583 δηλαδή το 15,04% του συνόλου. Το μέσο PD των ασυνεπών δανειοληπτών είναι 17,95% και η τυπική απόκλιση 7,44%. Από την άλλη πλευρά, οι συνεπείς δανειολήπτες έχουν μέσο PD 14,53% και τυπική απόκλιση 6,41%.

Προσθέτοντας μια τυπική απόκλιση στο μέσο των συνεπών και αφαιρώντας μια τυπική απόκλιση από τον μέσο των ασυνεπών ορίζεται η γκριζα περιοχή [10,51% , 20,94%] η οποία φαίνεται στο παρακάτω σχεδιάγραμμα:



Η περιοχή αυτή ονομάζεται γκριζα γιατί σε αυτήν όλοι οι δανειολήπτες έχουν παρόμοια χαρακτηριστικά ανεξάρτητα από το αν είναι συνεπείς ή ασυνεπείς. Στην γκριζα περιοχή υπάρχουν συνολικά 6363 απαιτήσεις. Οι ασυνεπείς δανειολήπτες

είναι 985 (15,48%) και οι συνεπείς είναι 5378 (84,52%). Το μέσο PD των ασυνεπών είναι 14,74% και η τυπική απόκλιση είναι 2,55%. Όσον αφορά τους συνεπείς έχουν μέσο PD 14,60% και τυπική απόκλιση 2,47%.

Παρατηρείται ότι οι συνεπείς και οι ασυνεπείς έχουν σχεδόν τα ίδια χαρακτηριστικά και έτσι σε αυτήν την περιοχή δεν μπορούν να εξαχθούν συμπεράσματα, με κριτήριο την πιθανότητα αθέτησης, για τη συμπεριφορά των δανειοληπτών ως προς την αποπληρωμή του δανείου.

Στον παρακάτω πίνακα παρουσιάζεται η κατανομή των ασυνεπών δανειοληπτών με βάση τις τιμές της πιθανότητας αθέτησης:

PD	Πλήθος απαιτήσεων	Ασυνεπείς	Ποσοστό ως προς τον πληθυσμό της κλάσης	Ποσοστό ως προς το σύνολο των ασυνεπών
> 20,94%	1730	436	25.20%	27.54%
17,95%-20,94%	1111	229	20.61%	14.47%
14,53%-17,95%	2206	368	16.68%	23.25%
10,51%-14,53%	3046	388	12.74%	24.51%
< 10,51%	2429	162	6.67%	10.23%

#### Προσδιορισμός του Cut off point

Επειδή οι κλάσεις των συνεπών και ασυνεπών δανειοληπτών δεν είναι ισομεγέθεις το cut off point προσδιορίζεται από την σχέση (4.5):

$$Z_{cut} = \frac{N_0 \overline{PD}_0 + N_1 \overline{PD}_1}{N_0 + N_1}$$

όπου,

$N_0$ : το πλήθος των συνεπών

$N_1$ : το πλήθος των ασυνεπών

$\overline{PD}_0$ : το μέσο PD των συνεπών

$\overline{PD}_1$ : το μέσο PD των ασυνεπών

Αντικαθιστώντας τις τιμές των παραμέτρων προκύπτει ότι  $Z_{cut} = 15,04\%$ . Οι αιτήσεις δανείων με πιθανότητα  $PD > 15,04\%$  απορρίπτονται, ενώ αυτές  $PD \leq 15,04\%$  εγκρίνονται.

### Ταξινόμηση των δανειοληπτών από άποψη φερεγγυότητας

Ο πίνακας ταξινόμησης που προκύπτει για cut off point 15,04% είναι ο εξής :

Cut off point=15,04%	Predicted Group		Actual Group Size
Actual Group	1	0	
1	976	607	1583
0	3638	5301	8939
Predicted Group Size	4614	5908	

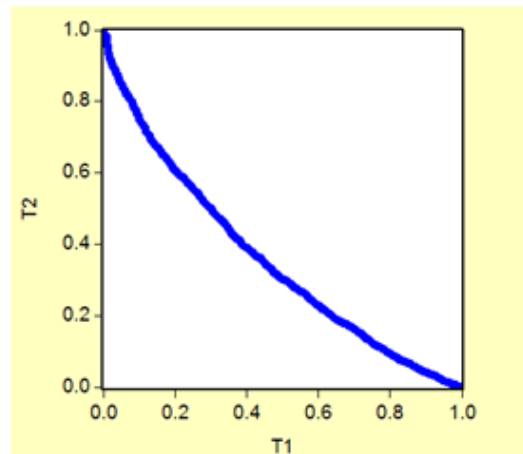
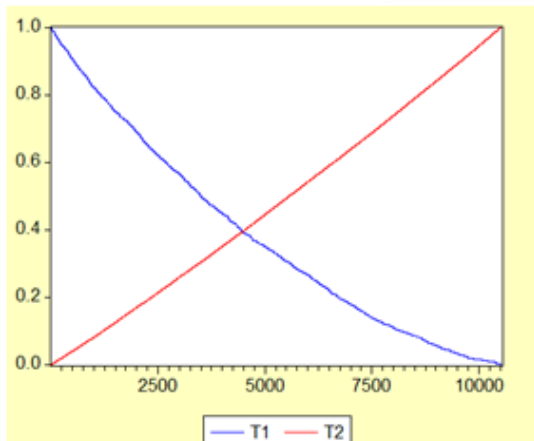
Οι σωστά τοποθετημένοι συνεπείς δανειολήπτες είναι 5301 δηλαδή  $PCC_{good} = 59,30\%$  ενώ το αντίστοιχο πλήθος των ασυνεπών είναι 976 δηλαδή  $PCC_{bad} = 61,66\%$ . Συνολικά οι δανειολήπτες που έχουν τοποθετηθεί στην σωστή κλάση είναι 6277 δηλαδή  $PCC_{total} = 59,66\%$ . (PCC= ποσοστό σωστά τοποθετημένων (Percentage of Correctly Classified) )

Από τον πίνακα διάκρισης προκύπτουν τα σφάλματα τύπου I και II (βλ. Παράρτημα 6):

ΣΦΑΛΜΑ ΤΥΠΟΥ I	38.34%
ΣΦΑΛΜΑ ΤΥΠΟΥ II	40.70%

Παρατηρείται ότι το σφάλμα τύπου I είναι μικρότερο από το τύπου II πράγμα επιθυμητό διότι το σφάλμα τύπου I οδηγεί σε ζημία ενώ το τύπου II σε διαφυγόντα κέρδη. Τα σφάλματα τύπου I και II είναι σχετικά υψηλά γεγονός που δείχνει ότι το υπόδειγμα δεν είναι πολύ αποτελεσματικό.

Η γραφική απεικόνιση των δύο σφαλμάτων δείχνει ότι έχουν πολύ υψηλή αρνητική συσχέτιση.



όπου,

T1 : Σφάλμα τύπου I

T2 : Σφάλμα τύπου II

Ο πίνακας συσχέτισης των σφαλμάτων τύπου I και II που ακολουθεί επιβεβαιώνει την υψηλή αρνητική συσχέτιση δηλαδή ότι οι χαμηλές τιμές του ενός σφάλματος συνεπάγονται υψηλές τιμές του άλλου.

	T1	T2
T1	1.000000	-0.979108
T2	-0.979108	1.000000

Έτσι, μεταβάλλοντας τις τιμές του cut off point προσδοκώντας καλύτερο αποτέλεσμα είναι αδύνατο να μειωθούν ικανοποιητικά και τα δύο σφάλματα.

### Δείγμα επικύρωσης

Για να επικυρωθεί το υπόδειγμα είναι σημαντικό να εξετασθεί η αποτελεσματικότητα και σε δείγμα διαφορετικό του δείγματος εκτίμησης, το δείγμα επικύρωσης (Out of sample validation).

Το δείγμα επικύρωσης αποτελείται από 5188 απαιτήσεις, από τους οποίους οι 795 είναι ασυνεπείς και ως ποσοστό ανέρχεται στο 15,32%. Το μέσο PD των ασυνεπών δανειοληπτών είναι 17,65% με τυπική απόκλιση 7,07%. Από την άλλη, οι συνεπείς δανειολήπτες έχουν μέσο PD 14,52% με τυπική απόκλιση 6,41%.

Παρατηρείται ότι υπάρχει μεγάλη ομοιότητα με το δείγμα εκτίμησης, έτσι η αποτελεσματικότητα του μοντέλου αναμένεται να είναι παρόμοια. Η εκτίμηση της

αποτελεσματικότητας του υποδείγματος θα γίνει με το cut off point που υπολογίστηκε στο δείγμα εκτίμησης,  $Z_{cut} = 15,04\%$ . Οι αιτήσεις δανείων με πιθανότητα  $PD > 15,04\%$  απορρίπτονται, ενώ αυτά με  $PD \leq 15,04\%$  εγκρίνονται.

Ο πίνακας ταξινόμησης που προκύπτει είναι:

Cut off point=15,04%	Predicted Group		Actual Group Size
Actual Group	1	0	
1	482	313	795
0	1810	2583	4,393
Predicted Group Size	2292	2896	

Η αποτελεσματικότητα του υποδείγματος γίνεται αντιληπτή από τα ποσοστά των σωστά τοποθετημένων δανειοληπτών. Από τον πίνακα παρατηρείται ότι οι σωστά τοποθετημένοι συνεπείς είναι 2583 ενώ οι ασυνεπείς είναι 482 οπότε  $PCC_{good} = 58,8\%$  και  $PCC_{bad} = 60,63\%$ . Συνολικά, οι σωστά τοποθετημένοι είναι 3065 δηλαδή,  $PCC_{total} = 59,08\%$ .

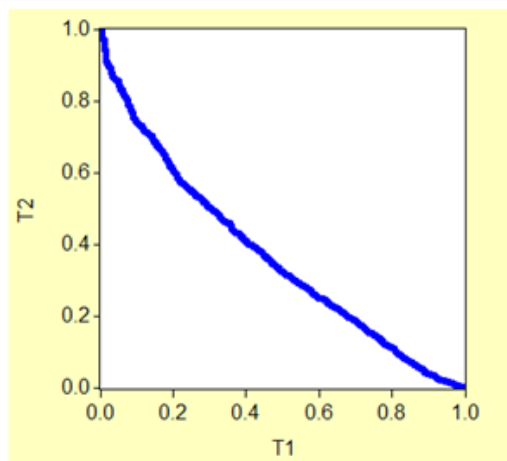
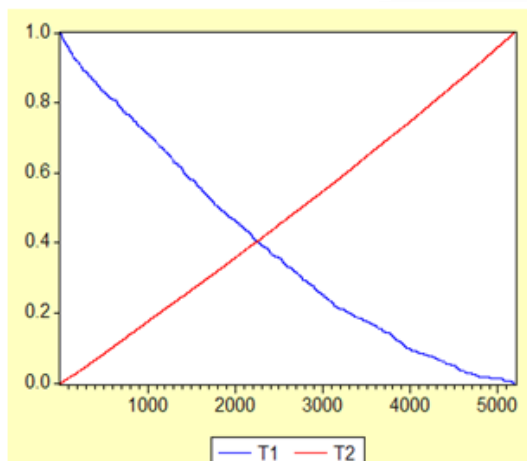
Αξίζει να σημειωθεί ότι οι *Desai et al. (1996)* στο υπόδειγμα τους έχουν  $PCC_{bad} = 41,38\%$  και  $PCC_{total} = 86,90\%$ .

Όπως αναμενόταν τα αποτελέσματα είναι παρόμοια με αυτά στο δείγμα εκτίμησης. Η αποτελεσματικότητα στο δείγμα επικύρωσης είναι ελαφρά χειρότερη. Το υπόδειγμα είναι λογικό να είναι πιο αποτελεσματικό στο δείγμα το οποίο κατασκευάστηκε *Hair et al (1998)*.

Από τον πίνακα ταξινόμησης προκύπτουν τα σφάλματα τύπου I και II:

ΣΦΑΛΜΑ ΤΥΠΟΥ I	39.37%
ΣΦΑΛΜΑ ΤΥΠΟΥ II	41.20%

Παρατηρείται ότι τα σφάλματα τύπου I και II είναι επίσης αρκετά υψηλά γεγονός που δείχνει ότι το υπόδειγμα δεν είναι πολύ αποτελεσματικό. Οι γραφικές παραστάσεις και ο πίνακας συσχέτισης των σφαλμάτων δείχνει ότι έχουν πολύ υψηλή αρνητική συσχέτιση όπως και στο δείγμα εκτίμησης.



όπου,

T1 : Σφάλμα τύπου I

T2 : Σφάλμα τύπου II

Πίνακας συσχέτισης:

	T1	T2
T1	1.000000	-0.982216
T2	-0.982216	1.000000

Όπως γίνεται αντιληπτό δεν είναι δυνατόν να επιτευχθεί ένα ζευγάρι σφαλμάτων με αρκετά μικρές τιμές και καθώς οι μικρές τιμές για το ένα συνεπάγονται μεγάλες για το άλλο.

Το υπόδειγμα είναι σχετικά αναποτελεσματικό διότι σύμφωνα με τα σφάλματα τύπου I και II το κόστος του να μην εγκριθεί το 60,63% των επισφαλών απαιτήσεων είναι η απόρριψη του 41,20% των υγιών απαιτήσεων γεγονός που σημαίνει ότι η τράπεζα παίρνει σωστές αποφάσεις μόνο για το 59,08% των αιτήσεων για δανειοδότηση. Προκειμένου δηλαδή, να προστατευθεί από τα επισφαλή δάνεια θα πρέπει να έχει σχετικά υψηλά διαφυγόντα κέρδη.

Μια εναλλακτική προσέγγιση είναι να ληφθεί υπόψη η γκρίζα περιοχή [10,51%, 20,94%] η οποία έχει οριστεί στο δείγμα εκτίμησης. Με βάση τα όρια της γκρίζας περιοχής το σημείο έγκρισης ή απόρριψης των δανείων δεν είναι πλέον το  $Z_{cut} = 15,04\%$ .



Σύμφωνα με αυτήν την προσέγγιση οι αιτήσεις δανείων με  $PD > 20,94\%$  απορρίπτονται, ενώ αυτές με  $PD \leq 10,51\%$  εγκρίνονται. Όσον αφορά τις αιτήσεις δανείων που με βάση την πιθανότητα αθέτησης ανήκουν στην γκριζα περιοχή δεν μπορεί να ληφθεί απόφαση για την έγκριση ή την απόρριψή της με βάση μόνο το PD, πρέπει να γίνει περαιτέρω ανάλυση.

Έτσι το δείγμα εκτίμησης χωρίζεται σε δύο κλάσεις στα δάνεια που δύναται να απορριφθούν ή να εγκριθούν με βάση την πιθανότητα αθέτησης που εκτιμάται από το υπόδειγμα και σε αυτά που ανήκουν στην γκριζα περιοχή.

#### Δάνεια εκτός γκριζας περιοχής

Τα δάνεια που δεν ανήκουν στην γκριζα περιοχή είναι 2076 από τα οποία σε 281 οι δανειολήπτες είναι ασυνεπείς. Ο πίνακας διάκρισης για αυτόν μόνο τον πληθυσμό είναι :

Cut off point=15,04%	Predicted Group		Actual Group Size
	(> 20,94%) 1	(< 10,51%) 0	
Actual Group			
1	199	82	281
0	635	1160	1,795
Predicted Group Size	834	1242	

Η αποτελεσματικότητα σε αυτό το δείγμα των 2076 απαιτήσεων φαίνεται από τους σωστά τοποθετημένους δανειολήπτες. Οι σωστά τοποθετημένοι συνεπείς είναι 1160 δηλαδή  $PCC_{good} = 64,62\%$ , οι ασυνεπείς είναι 199 δηλαδή  $PCC_{bad} = 70,82\%$ , και συνολικά 1359 δηλαδή  $PCC_{total} = 65,46\%$ .

Τα σφάλματα τύπου I και II είναι:

ΣΦΑΛΜΑ ΤΥΠΟΥ I	29.18%
ΣΦΑΛΜΑ ΤΥΠΟΥ II	35.38%

Γίνεται αντιληπτό ότι το υπόδειγμα είναι πολύ πιο αποτελεσματικό εκτός της γκριζας περιοχής. Οπότε η έγκριση ή η απόρριψη δανείων που δεν ανήκουν στην γκριζα περιοχή γίνεται με σημαντικά μικρότερη πιθανότητα λάθους.

Αξίζει να σημειωθεί ότι για την συνολική αποτελεσματικότητα του μοντέλου στις τιμές των σφαλμάτων αυτών πρέπει να συμπεριληφθούν και αυτές από τα σφάλματα τύπου I και II που προκύπτουν στην γκρίζα περιοχή.

#### Δάνεια που ανήκουν στην γκρίζα περιοχή

Στην γκρίζα περιοχή ανήκουν 3112 δάνεια και από αυτά σε 514 οι δανειολήπτες είναι ασυνεπείς. Δεν είναι δυνατόν να γίνει διαχωρισμός των συνεπών και των ασυνεπών με κριτήριο την εκτιμημένη πιθανότητα αθέτησης.

Σε αυτήν την περίπτωση πρέπει να γίνει η διάκριση με διαφορετική προσέγγιση η οποία μπορεί να βασίζεται και σε ποιοτικά κριτήρια. Σε αυτές τις περιπτώσεις οι εμπειρία και η γνώσεις των Risk Managers έχουν καταλυτικό ρόλο. Επίσης, ενδεχομένως να χρησιμοποιηθούν συγκεκριμένες ερμηνευτικές μεταβλητές, η επιλογή των οποίων γίνεται ποιοτικά βάσει των οποίων γίνεται η διάκριση των δανειοληπτών σε συνεπείς και ασυνεπείς.

Πχ μία αίτηση για δάνειο από πελάτη για τον οποίο υπάρχει δυσμενής καταχώρηση στον Τειρεσία απορρίπτεται ανεξαρτήτως της εκτίμησης της πιθανότητας αθέτησης που προκύπτει από το υπόδειγμα. Επίσης, παρόλο που το ετήσιο εισόδημα δεν συνεισφέρει στην διάκριση της συμπεριφοράς των δανειοληπτών, ενδεχομένως σε μεμονωμένες περιπτώσεις, που η τράπεζα θεωρεί ότι είναι σημαντική μπορεί να παίζει σημαντικό ρόλο για την έγκριση του δανείου.

#### **4.4.5 Συμπληρωματικές παρατηρήσεις**

Η σχετική αναποτελεσματικότητα του υποδείγματος πιθανώς να οφείλεται στον ορισμό που δίνει η Τράπεζα στην αθέτηση της αποπληρωμής των δανείων. Όπως έχει προαναφερθεί η τράπεζα θεωρεί έναν δανειολήπτη ασυνεπή εφόσον δεν έχει εξοφλήσει τις δόσεις τριών συνεχόμενων μηνών έστω και για μία φορά, παρόλο που στη συνέχεια μπορεί να είναι συνεπής στις υποχρεώσεις του. Ο ορισμός αυτός είναι σύμφωνος με την Βασιλεία II αλλά είναι αυστηρός για την Ελληνική πραγματικότητα.

Σύμφωνα με την Βασιλεία II ένα δάνειο είναι σε καθυστέρηση εφόσον :



1. Η τράπεζα θεωρεί ότι ο αντισυμβαλλόμενος είναι αδύνατο να εκπληρώσει τις υποχρεώσεις του χωρίς αναχρηματοδότηση.
2. Ο αντισυμβαλλόμενος έχει καθυστερήσει τις πληρωμές του περισσότερο από 90 μέρες για οποιαδήποτε οφειλή του στην τράπεζα.

Εφόσον ένα δάνειο είναι σε καθυστέρηση η τράπεζα κινεί τις διαδικασίες απόκτησης των εξασφαλίσεων.

Αντίθετα, στην Ελλάδα ο ορισμός της καθυστέρησης (ισοδύναμος του ασυνεπής δανειολήπτη) είναι πιο ελαστικός. Ένας δανειολήπτης κατατάσσεται στους ασυνεπείς με βάση το κριτήριο της καθυστέρησης των πληρωμών πέρα των 90 ημερών. Εφόσον όμως εξοφλήσει το ληξιπρόθεσμο ποσό χαρακτηρίζεται και πάλι συνεπής.

Ο υπολογισμός του ποσοστού αθέτησης, σε ένα χαρτοφυλάκιο δανείων, γίνεται «φωτογραφικά» με βάση τον αριθμό των πελατών που εκείνη τη στιγμή είναι σε προσωρινή καθυστέρηση, ανεξάρτητα με το ιστορικό του κάθε πελάτη. Έτσι λοιπόν, στην Ελλάδα μία τέτοιου είδους αθέτηση χαρακτηρίζεται προσωρινή.

Η ελαστικότητα αυτή εκτιμάται ότι επηρεάζει την συμπεριφορά των δανειοληπτών όσον αφορά στην αποπληρωμή του δανείου. Έτσι παρατηρείται άτομα με αρκετά διαφορετικά χαρακτηριστικά να έχουν παρόμοια συμπεριφορά.

Με το ακόλουθο παράδειγμα γίνεται πιο εύκολα αντιληπτό το πώς ο χαλαρός ορισμός της αθέτησης επηρεάζει αρνητικά την αποτελεσματικότητα του υποδείγματος.

Έστω ότι υπάρχουν δύο δανειολήπτες με πολύ διαφορετικά χαρακτηριστικά οι οποίοι έχουν ένα καταναλωτικό δάνειο. Και οι δύο δεν πληρώνουν για τρεις συνεχόμενους μήνες τις δόσεις, αλλά για εντελώς διαφορετικούς λόγους. Ο ένας, δεν μπορεί οικονομικά να ανταπεξέλθει στις υποχρεώσεις του και ο άλλος, απλά το έχει αμελήσει, εφόσον γνωρίζει ότι η τράπεζα δεν θα κινηθεί άμεσα για την απόκτηση των εξασφαλίσεων. Ως αποτέλεσμα και οι δύο πελάτες έχουν χαρακτηριστεί ασυνεπείς παρόλο που ο ένας συνέχισε την αποπληρωμή κανονικά σε αντίθεση με τον άλλο που η οικονομική του κατάσταση ανάγκασε την τράπεζα, μετά από κάποιο χρονικό διάστημα, να κινηθεί νομικά εναντίον του.

Στο υπόδειγμα εκτίμησης της πιθανότητας αθέτησης και οι δύο δανειολήπτες θα έχουν στην ανεξάρτητη τυχαία μεταβλητή  $Y$  την τιμή 1 παρόλο που οι ερμηνευτικές

μεταβλητές τους διαφοροποιούνται σημαντικά. Έτσι δημιουργείται πρόβλημα στην διάκριση της συμπεριφοράς των δανειοληπτών με βάση τα χαρακτηριστικά τους και κατά συνέπεια στην αποτελεσματικότητα του υποδείγματος.

Αξίζει να σημειωθεί ότι σύμφωνα με τις δημοσιεύσεις της Τράπεζας της Ελλάδος ο μέσος όρος των καθυστερήσεων των καταναλωτικών δανείων για την περίοδο 1/1/2007-15/7/2009 εκτιμήθηκε στο 10%. Το χαρτοφυλάκιο της τράπεζας που χρησιμοποιήθηκε για την εκτίμηση της πιθανότητας αθέτησης έχει ποσοστό καθυστερήσεων 15% γεγονός που δείχνει ότι ο χαρακτηρισμός από την τράπεζα ενός δανειολήπτη ως ασυνεπή ήταν μάλλον αυστηρός. Έτσι, πολλοί πελάτες χαρακτηρίστηκαν λανθασμένα ασυνεπείς γεγονός που επηρέασε την αποτελεσματικότητα του υποδείγματος.

## ΚΕΦΑΛΑΙΟ 5

### ΑΝΑΚΕΦΑΛΑΙΩΣΗ – ΣΥΜΠΕΡΑΣΜΑΤΑ

Σήμερα, με δεδομένη την πολυπλοκότητα των οικονομικών σχέσεων, την παγκοσμιοποίηση της οικονομίας και την αύξηση του συστημικού κινδύνου, ο πιστωτικός κίνδυνος είναι αναμφίβολα ο σημαντικότερος κίνδυνος που εκτίθενται όλες οι επιχειρήσεις και κυρίως τα χρηματοπιστωτικά ιδρύματα.

Οι προσπάθειες εκτίμησης του πιστωτικού κινδύνου είχαν ως αποτέλεσμα την ανάπτυξη αρκετών μεθοδολογιών, ορισμένες από τις οποίες παρουσιάζουν ικανοποιητική αποτελεσματικότητα. Όλες αυτές οι μεθοδολογίες αποσκοπούν στην προϋπολογιστική ταξινόμηση των χρεωστών σε δύο βασικές κατηγορίες : Σε εκείνους που αναμένεται να είναι συνεπείς και σε εκείνους που αναμένεται να είναι ασυνεπείς.

Όπως είναι φυσικό, η λανθασμένη ταξινόμηση των χρεωστών δημιουργεί κινδύνους, διότι οδηγεί σε λανθασμένες αποφάσεις. Για παράδειγμα, η a priori κατάταξη ενός ασυνεπούς χρεώστη στους συνεπείς, σημαίνει ότι αυτός θα πάρει τη ζητούμενη πίστωση, με αποτέλεσμα την μερική ή ολική απώλεια των αντίστοιχων κεφαλαίων.

Για την διασφάλιση της βιωσιμότητας του τραπεζικού συστήματος έχει αναπτυχθεί ένα πλαίσιο κανονισμών, γνωστό ως «Σύμφωνο Βασιλείας II», το οποίο υποχρεώνει τις τράπεζες να παρουσιάζουν τις απαιτήσεις τους κατά τρόπο που προσεγγίζει ικανοποιητικά την πραγματικότητα, όσον αφορά την πιθανότητα είσπραξής τους.

Ειδικότερα, για τον πιστωτικό κίνδυνο, το σύμφωνο της Βασιλείας II ορίζει αυστηρά τις αρχές και τους κανόνες που πρέπει να διέπουν τις προσπάθειες εκτίμησης του. Με βάση τον εκτιμώμενο κίνδυνο, η Βασιλεία II ορίζει τα ελάχιστα ίδια κεφάλαια που πρέπει να διαθέτουν οι τράπεζες για την αντιμετώπισή του.

Ο πιο απλός υπολογισμός των κεφαλαιακών απαιτήσεων μπορεί να γίνει με την τυποποιημένη μέθοδο που ορίζει η Βασιλεία II.

Εναλλακτικά, δίνεται η δυνατότητα στις τράπεζες να εκτιμήσουν τον πιστωτικό κίνδυνο κάνοντας χρήση της Προσέγγισης Εσωτερικών Διαβαθμίσεων (IRB)

σύμφωνα με την οποία θα πρέπει να αναπτύξουν δικά τους υποδείγματα εκτίμησης του πιστωτικού κινδύνου.

Στα πλαίσια της παρούσας εργασίας, αρχικά παρουσιάστηκαν τα βασικά υποδείγματα εκτίμησης της πιθανότητας αθέτησης (PD), που αποτελεί την πιο σημαντική παράμετρο του πιστωτικού κινδύνου. Τα υποδείγματα αυτά είναι:

α) Η Πολυμεταβλητή Ανάλυση διαχωρισμού (Multivariate discriminate analysis),

β) Η Λογιστική Παλινδρόμηση (Logit) και

γ) Το υπόδειγμα Probit.

Τα υποδείγματα Logit και Probit έχουν ένα συγκριτικό πλεονέκτημα έναντι της Πολυμεταβλητή Ανάλυσης Διαχωρισμού διότι κάποιες από τις υποθέσεις της ικανοποιούνται δύσκολα στην πράξη. Επίσης, τα υποδείγματα Logit και Probit εκτιμούν άμεσα την πιθανότητα αθέτησης σε αντίθεση με την Πολυμεταβλητή Ανάλυση διαχωρισμού. Τα υποδείγματα Probit και Logit είναι παρόμοια, αλλά το Logit είναι πιο εύκολο στη χρήση.

Στη συνέχεια, παρουσιάστηκε η θεωρητική προσέγγιση της κατασκευής υποδειγμάτων εκτίμησης της πιθανότητας αθέτησης, η οποία πραγματεύεται την λύση των εξής επιμέρους προβλημάτων :

- Επιλογή υποδείγματος,
- Αρχική επιλογή των ανεξάρτητων ερμηνευτικών μεταβλητών,
- Επιλογή και η μορφοποίηση των δεδομένων,
- Πληρότητα των δεδομένων,
- Κωδικοποίηση των ερμηνευτικών μεταβλητών,
- Διαχωρισμός του δείγματος σε δείγμα εκτίμησης και δείγμα επικύρωσης,
- Ανάπτυξη του υποδείγματος Logit:
  - Κατασκευή της συνάρτησης διαχωρισμού,
  - Στατιστική επικύρωση,
  - Καθορισμός του cut off point,
  - Αποτελεσματικότητα υποδείγματος.

Τέλος, αναπτύχθηκε ένα υπόδειγμα εκτίμησης της πιθανότητας αθέτησης σε χαρτοφυλάκιο 15.786 καταναλωτικών δανείων ελληνικής τράπεζας στο οποίο το ποσοστό των ασυνεπών δανειοληπτών ανέρχεται στο 15,13%.

Το υπόδειγμα που επιλέχθηκε για την εκτίμηση της πιθανότητας αθέτησης είναι το Logit και η επιλογή έγινε βάσει των αποτελεσμάτων προηγούμενων ερευνών.

Στα δεδομένα υπήρχαν 26 επιμέρους προσωπικά στοιχεία των δανειοληπτών τα οποία χρησιμοποιήθηκαν στην εργασία ως ανεξάρτητες ερμηνευτικές μεταβλητές. Μετά την εφαρμογή τεχνικών αρχικής επιλογής των ερμηνευτικών μεταβλητών απέμειναν 14 μεταβλητές για την ανάπτυξη του υποδείγματος εκτίμησης της πιθανότητας αθέτησης οι οποίες είναι :

1. *Ημερομηνία αίτησης (ηη/μμ/εε)*
2. *Ημερομηνία γέννησης (ηη/μμ/εε)*
3. *Κάτοχος πιστωτικής κάρτας*
4. *Ιδιοκτησία*
5. *Οικογενειακή κατάσταση*
6. *Πελάτης της τράπεζας*
7. *Χειρότερη καταχώρηση Τειρεσία*
8. *Έτη από την πιο πρόσφατη καταχώρηση Τειρεσία*
9. *Διάρκεια δανείου*
10. *Μηνιαία δόση προς μηνιαίο εισόδημα*
11. *Φύλο*
12. *Αριθμός παιδιών*
13. *Είδος προσωπικού εγγράφου*
14. *Ετήσιο εισόδημα*

Στη συνέχεια τα δάνεια χωρίστηκαν σε δύο δείγματα:

- α) Το δείγμα εκτίμησης των συντελεστών των ερμηνευτικών μεταβλητών και*
- β) Το δείγμα επικύρωσης του υποδείγματος.*

Το δείγμα εκτίμησης περιέχει τα 2/3 των δανείων ενώ το δείγμα επικύρωσης το 1/3 αυτών.

Το στατιστικά σημαντικό υπόδειγμα Logit που προέκυψε από το δείγμα εκτίμησης είναι το εξής<sup>1</sup>:

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.097812	0.508985	-4.121564	0.0000
X1	0.043452	0.016727	2.597729	0.0094
X2	-0.020037	0.002587	-7.744944	0.0000
X3	-1.005937	0.479575	-2.097561	0.0359
X5	-0.556438	0.132397	-4.202809	0.0000
X6	-0.885579	0.179594	-4.930995	0.0000
X8	-1.118201	0.166914	-6.699271	0.0000
X9	-0.876596	0.088331	-9.923960	0.0000
X13	-1.664859	0.548184	-3.037047	0.0024

Η τιμή του cut off point στο δείγμα εκτίμησης προσδιορίστηκε ίση με 15,04%.

Η αποτελεσματικότητα του υποδείματος μελετήθηκε και στο δείγμα εκτίμησης και στο δείγμα επικύρωσης. Τα αποτελέσματα στο δείγμα επικύρωσης παρουσιάζουν πιο αντιπροσωπευτικά την προβλεπτική ικανότητα του υποδείματος. Τα σφάλματα τύπου I και II φαίνονται στον πίνακα :

ΣΦΑΛΜΑ ΤΥΠΟΥ I	39.37%
ΣΦΑΛΜΑ ΤΥΠΟΥ II	41.20%

Δηλαδή, με πιθανότητα 39,37% η τράπεζα κατατάσσει έναν ασυνεπή δανειολήπτη στους συνεπείς, παράλληλα με πιθανότητα 41,20% κατατάσσει έναν συνεπή στους ασυνεπείς. Το υπόδειγμα έχει σχετικά μικρή προβλεπτική ικανότητα καθώς τα σφάλματα τύπου I και II είναι σχετικά μεγάλα.

Παρατηρήθηκε ότι υπάρχει μια γκριζα περιοχή όπου η φερεγγυότητα των δανειοληπτών δεν προσδιορίζεται εμφανώς από τα επιμέρους χαρακτηριστικά τους. Αφαιρώντας τους δανειολήπτες που ανήκουν στην γκριζα περιοχή η αποτελεσματικότητα του υποδείματος βελτιώθηκε σημαντικά:

<sup>1</sup> Οι ερμηνευτικές μεταβλητές του υποδείματος αντιστοιχούν στα χαρακτηριστικά των πελατών που αναφέρθηκαν πιο πάνω, με την ίδια αρίθμηση. Για παράδειγμα, η X2 αντιστοιχεί στο χαρακτηριστικό Νο 2 που είναι η ημερομηνία γέννησης.

ΣΦΑΛΜΑ ΤΥΠΟΥ Ι	29.18%
ΣΦΑΛΜΑ ΤΥΠΟΥ ΙΙ	35.38%

Συμπερασματικά, η εμπειρική ανάλυση στα πλαίσια της παρούσας εργασίας έδειξε ότι είναι εφικτή η ανάπτυξη υποδειγμάτων με αξιόλογη προβλεπτική ικανότητα, τα οποία επιτρέπουν στις τράπεζες να αξιολογούν τους υποψήφιους δανειολήπτες με αντικειμενικά δεδομένα και αποκλειστικά με ποιοτικά κριτήρια ή διαισθητικά.

Ειδικότερα, στα πλαίσια του αναπτυχθέντος υποδείγματος, έχοντας αφαιρέσει τους δανειολήπτες που ανήκουν στην γκρίζα περιοχή, το σφάλμα τύπου Ι ήταν 29,18% (δηλαδή σωστές ταξινομήσεις 70,71%) και το σφάλμα τύπου ΙΙ 35,38% (σωστές ταξινομήσεις 64,62%).

Το υπόδειγμα πιθανότατα θα ήταν αποτελεσματικότερο αν η τράπεζα του δείγματος δεν ήταν αυστηρή στον χαρακτηρισμό ενός δανειολήπτη ως ασυνεπή. Η αυστηρότητα αυτή επισημαίνεται και από το μέσο ποσοστό αθετήσεων στα καταναλωτικά δάνεια για την αντίστοιχη περίοδο και για όλες τις τράπεζες, το οποίο ήταν 10%, έναντι 15% της εν λόγω τράπεζας.

Τέλος, θα πρέπει να επισημανθεί ότι, πέρα από τη χρήση υποδειγμάτων ποσοτικής ανάλυσης, είναι αναγκαίο να αξιοποιείται και η εμπειρία των ειδικών, ώστε να οδηγούμαστε σε ασφαλέστερα συμπεράσματα.

# РАНЕЕЗНАМО ПЕРПАА



## ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική :

1. Μιχάλης Γκλεζάκος, (2008). *Αξιολόγηση των επιχειρήσεων*. Εκδότης: Μιχάλης Γκλεζάκος.

Ξένα :

2. *Basel Committee on Banking Supervision. International Convergence of Capital Measurement and Capital Standards. A Revised Framework Comprehensive Version. June 2006*
3. Boyle, M., Crook, J. N., Hamilton, R., & Thomas, L. C. (1992). *Methods for credit scoring applied to slow payers*. In L. C. Thomas, J. N. Crook, & D. B. Edelman (Eds.), *Credit scoring and credit control* (pp. 75–90). Oxford: Oxford University Press.
4. Brown, C. H. (1983). *Asymptotic comparison of missing data procedures for estimating factor loading*. *Psychometrika*, 48(2), 269-292.
5. Carl J. Huberty and Stephen Olenjnik (2006). *Applied MANOVA and Discriminant Analysis* (2<sup>nd</sup> ed.). John Willey and Sons. Inc.
6. Christian Bluhm, Ludger Overbeck and Cristoph Wagner (2003). *An introduction to Credit Risk Modeling*. Chapman and Hall/CRC.
7. Crook, J. N., Hamilton, R., & Thomas, L. C. (1992). *A comparison of discriminations under alternative definitions of credit default*. In L. C. Thomas, J. N. Crook, & D. B. Edelman (Eds.), *Credit scoring and credit control* (pp. 217–245). Oxford: Oxford University Press.
8. D. J. Hand and W. E. Henley (1997). *Statistical Classification Methods in Consumer Credit Scoring: a Review*. *Journal of the Royal Statistic Society* vol. 160, No 3, pages 523-541.
9. Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). *A comparison of neural networks and linear scoring models in the credit union environment*. *European Journal of Operational Research*, 95, 24–37.
10. Desai, V. S., Crook, J. N., & Overstreet, G. A. (1997). *Credit scoring models in the credit union environment using neural networks and genetic algorithms*. *IMA Journal of Mathematics Applied in Business and Industry*, 8, 323–346.
11. Duffie, D. and K.J. Singleton (2003). *Credit Risk*. Princeton: Princeton University Press.
12. Gunter Löffler and Peter N. Posch, (2007). *Credit risk modeling using Excel and VBA*. Willey Finance series. John Willey and Sons, Ltd.

13. Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag.
14. Henley, W. E. (1995). *Statistical aspects of credit scoring*. Open University, Dissertation.
15. Herman J. Bierens (2008). *The Logit Model: Estimation, Testing and Interpretation*.
16. J. S. Cramer (2003). *Logit Models from Economics and Other Fields*. Cambridge University Press 2003
17. Joel Bessis, (2002). *Risk Mangement in Banking 2<sup>nd</sup> ed.* John Willey and Sons, Ltd.
18. Joseph F. Hair, JR. Rolph E. Anderson. Ronald L. Tatham. William C. Black (1998). *Multivariate Data Analysis (5<sup>th</sup> ed.)*. Prentice Hall
19. Kline, R. B. (1998). *Principles and practices of structural equation modeling*. New York: Guilford.
20. Little, R.J.A. and D.B. Rubin (2002). *Statistical Analysis with Missing Data*. New York: Wiley Interscience.
21. Ludwing Fahrmeir, Gerhard Tutz (2001). *Multivariate statistical modeling Based on generalized linear models (second edition)*. Springer series in statistics. Springer
22. Lydian Medema, Ruud, H. Koning and Robert Lensik (2007). *A Practical Approach to Validating a PD Model*. *Journals of Banking and Finance* (33)4 April 2009, pages 701-708
23. Naeem Siddiqi (2005). *Credit Risk Scorecards, Developing and Implementing Intelligent Credit Scoring*. John Willey and Sons, Inc. pages 70- 90
24. Newman, Daniel A (2003). *Longitudinal Modeling with Randomly and Systematically Missing Data: A Simulation of Ad Hoc, Maximum Likelihood, and Multiple Imputation Techniques*. *Organization Research Methods*, Vol. 6 No3, July 2003 pp. 328-362.
25. Schreiner, M. (2004). *Scoring arrears at a microlender in Bolivia*. *Journal of Microfinance*, 6,65–88.
26. Srinivasan, V., & Kim, Y. H. (1987). *The Bierman–Hausman credit granting model: A note*. *Management Science*, 33, 1361–1362.
27. Thi Huyen Thanh Dinh and Stefanie Kleimeier (2007). *A credit scoring model for Vietnam's retail banking market*. *International Review of Financial Analysis* 16, 471–495
28. Thomas, L. C. (2000). *A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers*. *International Journal of Forecasting*, 16, 149–172.

29. Tim Futing Liao 1994. *Interpreting probability models Logit, Probit, and other generalized linear models. Series: Quantitative Applications in the Social Science. Sage university paper.*
30. Viganó, L. (1993). *A credit scoring model for development banks: An African case study. Savings and Development 4, 441–482.*
31. Wothke, W. (1998). *Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, and J. Baumert [Eds.]. Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples.. Mahwah, NJ: Lawrence Erlbaum Publishers.*
32. Wilks, S.S. (1932). *Certain generalizations of the analysis of variance. Biometrika, 39, 471–494 (48).*
33. Yihan Yang, Guangli Nie and Lingling Zhang (2009). *Retail Exposures Credit Scoring Models for Chinese Commercial Banks. Lecture Notes in Computer science (LNCS) vol. 5545 pp. 633-642. Springer-Verlag Berlin Heidelberg 2009.*
34. Yobas, M. B., Crook, J. N., & Ross, P. (2000). *Credit scoring using neural and evolutionary techniques. IMA Journal of Mathematics Applied in Business and Industry, 11, 111–125.*

# РАНЕЕЗНАМО ПЕРПАА

## ΠΑΡΑΡΤΗΜΑΤΑ

### ΠΑΡΑΡΤΗΜΑ 1 : Υπολογισμός των πινάκων Ε και Η

Ο πίνακας SSCP<sub>j</sub> προκύπτει από το γινόμενο:

$$C_j^T C_j = (Y_{uj} - \bar{Y}_{1,j})^T (Y_{uj} - \bar{Y}_{1,j})$$

όπου,

$Y_{uj}$  = το διάνυσμα των τυχαίων μεταβλητών  $Y_i$  για τη μονάδα  $u$  του γκρουπ  $j$ .

$\bar{Y}_{1,j}$  = το διάνυσμα των μέσων των τυχαίων μεταβλητών  $Y_i$  για τη μονάδα  $u$  του γκρουπ  $j$ .

Ενδεικτικά για δύο μεταβλητές  $Y_1$  και  $Y_2$  ο πίνακας SSCP<sub>j</sub> για το γκρουπ  $j$  είναι:

$$\begin{aligned} C_j^T C_j &= \mathbf{SSCP}_j \\ &= \begin{bmatrix} \sum_{u=1}^{n_j} (Y_{1uj} - \bar{Y}_{1,j})^2 & \sum_{u=1}^{n_j} (Y_{1uj} - \bar{Y}_{1,j})(Y_{2uj} - \bar{Y}_{2,j}) \\ \sum_{u=1}^{n_j} (Y_{1uj} - \bar{Y}_{1,j})(Y_{2uj} - \bar{Y}_{2,j}) & \sum_{u=1}^{n_j} (Y_{2uj} - \bar{Y}_{2,j})^2 \end{bmatrix} \end{aligned}$$

Υπολογισμός του πίνακα Ε γίνεται αθροίζοντας τους SSCP για όλες τις κλάσεις.

$$\mathbf{E} = \sum_{j=1}^J \mathbf{SSCP}_j$$

Ο πίνακας Η υπολογίζεται όπως ο Ε με την διαφορά ότι υπολογίζουμε τους SSCP πίνακες κάνοντας χρήση του διανύσματος του συνολικού μέσου και όχι του μέσου ανά κλάση. Αντί του  $Y_{1,j}, Y_{2,j}$  χρησιμοποιείται το  $Y_{..}$ .

Υπολογισμός του Error Covariance Matrix γίνεται διαιρώντας τον πίνακα Ε με το  $N-J$

$$\mathbf{S}_e = 1/(N - J)\mathbf{E}$$

**ΠΑΡΑΡΤΗΜΑ 2 : Πίνακας συσχετίσεων των ερμηνευτικών μεταβλητών**

	X1	X2	X3	X4	X5	X6	X7
Y	0.024019	-0.108182	0.027826	0.046072	0.055983	0.058730	0.035754
X1	1.000000	0.017510	-0.003001	0.000449	0.028111	0.024389	-0.001339
X2	0.017510	1.000000	-0.036031	-0.342283	-0.111980	-0.086431	0.049534
X3	-0.003001	-0.036031	1.000000	0.090790	0.032413	0.149033	0.005776
X4	0.000449	-0.342283	0.090790	1.000000	0.200760	0.122138	-0.009943
X5	0.028111	-0.111980	0.032413	0.200760	1.000000	0.029773	0.011347
X6	0.024389	-0.086431	0.149033	0.122138	0.029773	1.000000	0.008263
X7	-0.001339	0.049534	0.005776	-0.009943	0.011347	0.008263	1.000000
X8	-0.001983	0.056048	0.008955	-0.014315	0.011343	0.007452	0.975911
X9	-0.008215	-0.124761	-0.100432	-0.007908	0.015104	-0.003561	0.014630
X10	-0.029421	-0.038330	0.050656	0.020261	0.016795	-0.004704	-0.031258
X11	-0.020596	0.034074	0.007122	0.020408	-0.183925	-0.004958	0.045204
X12	-0.001372	0.105451	-0.063111	-0.175782	-0.285395	-0.043690	0.021660
X13	-0.012558	-0.102949	-0.009185	0.030441	0.002736	0.000115	-0.005070
X14	0.007922	0.076588	-0.138193	-0.096370	-0.061651	-0.061878	0.018064

	X8	X9	X10	X11	X12	X13	X14
Y	0.024856	0.102938	-0.012847	-0.006505	-0.030911	-0.018612	0.011446
X1	-0.001983	-0.008215	-0.029421	-0.020596	-0.001372	-0.012558	0.007922
X2	0.056048	-0.124761	-0.038330	0.034074	0.105451	-0.102949	0.076588
X3	0.008955	-0.100432	0.050656	0.007122	-0.063111	-0.009185	-0.138193
X4	-0.014315	-0.007908	0.020261	0.020408	-0.175782	0.030441	-0.096370
X5	0.011343	0.015104	0.016795	-0.183925	-0.285395	0.002736	-0.061651
X6	0.007452	-0.003561	-0.004704	-0.004958	-0.043690	0.000115	-0.061878
X7	0.975911	0.014630	-0.031258	0.045204	0.021660	-0.005070	0.018064
X8	1.000000	0.009605	-0.028450	0.044552	0.021105	-0.006590	0.017845
X9	0.009605	1.000000	-0.216322	-0.027200	-0.005572	-0.013616	0.071127
X10	-0.028450	-0.216322	1.000000	-0.030589	-0.051259	-0.045515	-0.235850
X11	0.044552	-0.027200	-0.030589	1.000000	0.005986	0.080555	0.034252
X12	0.021105	-0.005572	-0.051259	0.005986	1.000000	-0.006927	0.052060
X13	-0.006590	-0.013616	-0.045515	0.080555	-0.006927	1.000000	0.004196
X14	0.017845	0.071127	-0.235850	0.034252	0.052060	0.004196	1.000000



### IIAPATHMA 3 : Equation output

Dependent Variable: Y  
Method: ML - Binary Logit (Quadratic hill climbing)  
Date: 07/21/09 Time: 17:43  
Sample: 1 10522  
Included observations: 10522  
Convergence achieved after 6 iterations  
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.097812	0.508985	-4.121564	0.0000
X1	0.043452	0.016727	2.597729	0.0094
X2	-0.020037	0.002587	-7.744944	0.0000
X3	-1.005937	0.479575	-2.097561	0.0359
X5	-0.556438	0.132397	-4.202809	0.0000
X6	-0.885579	0.179594	-4.930995	0.0000
X8	-1.118201	0.166914	-6.699271	0.0000
X9	-0.876596	0.088331	-9.923960	0.0000
X13	-1.664859	0.548184	-3.037047	0.0024
Mean dependent var	0.150447	S.D. dependent var	0.357526	
S.E. of regression	0.351629	Akaike info criterion	0.814827	
Sum squared resid	1299.858	Schwarz criterion	0.821038	
Log likelihood	-4277.804	Hannan-Quinn criter.	0.816924	
Restr. log likelihood	-4455.889	Avg. log likelihood	-0.406558	
LR statistic (8 df)	356.1705	McFadden R-squared	0.039966	
Probability(LR stat)	0.000000			
Obs with Dep=0	8939	Total obs	10522	
Obs with Dep=1	1583			

## IIAPAPTHMA 4: Correlogram of Standardized Squared Residuals

Date: 09/15/09 Time: 10:13

Sample: 1 10522

Included observations: 10522

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.012	0.012	1.4895	0.222		
2	-0.000	-0.000	1.4901	0.475		
3	-0.011	-0.011	2.8646	0.413		
4	0.002	0.002	2.9031	0.574		
5	0.007	0.007	3.3761	0.642		
6	0.002	0.002	3.4183	0.755		
7	0.026	0.026	10.388	0.168		
8	0.018	0.017	13.617	0.092		
9	0.002	0.002	13.659	0.135		
10	0.008	0.008	14.307	0.159		
11	0.004	0.004	14.468	0.208		
12	-0.003	-0.003	14.541	0.267		
13	0.006	0.005	14.860	0.316		
14	0.005	0.005	15.170	0.367		
15	0.011	0.010	16.474	0.351		
16	0.007	0.006	16.934	0.390		
17	-0.002	-0.002	16.965	0.457		
18	-0.001	-0.002	16.982	0.524		
19	-0.003	-0.003	17.083	0.584		
20	-0.001	-0.001	17.088	0.647		
21	0.011	0.010	18.278	0.631		
22	-0.020	-0.021	22.595	0.425		
23	-0.005	-0.006	22.893	0.467		
24	-0.005	-0.005	23.202	0.508		
25	-0.014	-0.014	25.153	0.454		
26	0.012	0.013	26.792	0.420		
27	0.001	0.001	26.796	0.475		
28	0.004	0.003	26.968	0.520		
29	-0.010	-0.009	28.054	0.515		
30	0.020	0.021	32.087	0.364		
31	-0.010	-0.011	33.193	0.361		
32	0.007	0.008	33.738	0.383		
33	-0.013	-0.013	35.533	0.350		
34	0.006	0.006	35.945	0.377		
35	0.008	0.008	36.694	0.390		



## ΠΑΡΑΡΤΗΜΑ 5 : Hosmer-Lemeshow test

Dependent Variable: Y

Method: ML - Binary Logit (Quadratic hill climbing)

Date: 08/04/09 Time: 15:16

Sample: 1 10522

Included observations: 10522

Andrews and Hosmer-Lemeshow Goodness-of-Fit Tests

Grouping based upon predicted risk (randomize ties)

	Quantile of Risk		Dep=0		Dep=1		Total Obs	H-L Value
	Low	High	Actual	Expect	Actual	Expect		
1	0.0287	0.0672	999	995.106	53	56.8936	1052	0.28171
2	0.0672	0.0986	967	964.767	85	87.2333	1052	0.06235
3	0.0986	0.1159	955	938.367	97	113.633	1052	2.72939
4	0.1159	0.1298	914	922.895	138	129.105	1052	0.69865
5	0.1298	0.1428	906	909.641	147	143.359	1053	0.10702
6	0.1428	0.1559	902	895.029	150	156.971	1052	0.36391
7	0.1559	0.1728	861	879.376	191	172.624	1052	2.34023
8	0.1728	0.1970	846	858.515	206	193.485	1052	0.99198
9	0.1970	0.2377	831	824.610	221	227.390	1052	0.22910
10	0.2377	0.6675	758	750.694	295	302.306	1053	0.24770
Total			8939	8939.00	1583	1583.00	10522	8.05204
H-L Statistic:			8.0520		Prob. Chi-Sq(8)		0.4284	
Andrews Statistic:			11.4353		Prob. Chi-Sq(10)		0.3246	

**ΠΑΡΑΡΤΗΜΑ 6 : Σφάλμα τύπου I και II για Cut off point = 15,04 %**

Dependent Variable: Y  
 Method: ML - Binary Logit (Quadratic hill climbing)  
 Date: 07/21/09 Time: 17:43  
 Sample: 1 10522  
 Included observations: 10522  
 Prediction Evaluation (success cutoff C = 0.1504)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	5301	607	5908	0	0	0
P(Dep=1)>C	3638	976	4614	8939	1583	10522
Total	8939	1583	10522	8939	1583	10522
Correct	5301	976	6277	0	1583	1583
% Correct	59.30	61.66	59.66	0.00	100.00	15.04
% Incorrect	40.70	38.34	40.34	100.00	0.00	84.96
Total Gain*	59.30	-38.34	44.61			
Percent Ga...	59.30	NA	52.51			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	7640.19	1298.81	8939.00	7594.16	1344.84	8939.00
E(# of Dep=1)	1298.81	284.19	1583.00	1344.84	238.16	1583.00
Total	8939.00	1583.00	10522.00	8939.00	1583.00	10522.00
Correct	7640.19	284.19	7924.38	7594.16	238.16	7832.31
% Correct	85.47	17.95	75.31	84.96	15.04	74.44
% Incorrect	14.53	82.05	24.69	15.04	84.96	25.56
Total Gain*	0.51	2.91	0.87			
Percent Ga...	3.42	3.42	3.42			

\*Change in "% Correct" from default (constant probability) specification

\*\*Percent of incorrect (default) prediction corrected by equation

# РАНЕЕЗНАМО ПЕРПАА