

# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



## ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

### ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΥΝΔΥΝΟΥ

## ΜΟΝΤΕΛΑ ΑΞΙΟΛΟΓΗΣΗΣ ΠΙΣΤΟΛΗΠΤΙΚΗΣ ΙΚΑΝΟΤΗΤΑΣ

Ιωάννης Αντωνόπουλος

Διπλωματική Εργασία

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

Πειραιάς  
Σεπτέμβριος 2015



# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



## ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

### ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΥΝΔΥΝΟΥ

## ΜΟΝΤΕΛΑ ΑΞΙΟΛΟΓΗΣΗΣ ΠΙΣΤΟΛΗΠΤΙΚΗΣ ΙΚΑΝΟΤΗΤΑΣ

Ιωάννης Αντωνόπουλος

Διπλωματική Εργασία

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

Πειραιάς  
Σεπτέμβριος 2015

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

Τα μέλη της Επιτροπής ήταν:

- Κούτρας Μάρκος (Επιβλέπων)
- Μιχαήλ Γκλεζάκος
- Γεώργιος Τζαβελάς

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

**UNIVERSITY OF PIRAEUS**



**DEPARTMENT OF STATISTICS  
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN  
ACTUARIAL SCIENCE AND RISK  
MANAGEMENT**

# **CREDIT SCORING MODELS**

By

**Ioannis Antonopoulos**

MSc Dissertation

submitted to the Department of Statistics and Insurance  
Science of the University of Piraeus in partial fulfilment of the  
requirements for the degree of Master of Science in Actuarial  
Science and Risk Management

Piraeus, Greece

September 2015









## **Ευχαριστίες**

Ευχαριστώ, τους Γονείς μου και τους φίλους μου που μου συμπαραστάθηκαν σε αυτήν την πορεία καθώς και τους καθηγητές μου για τον επαγγελματισμό τους και την στήριξή τους σε 'μένα και τους συμφοιτητές μου καθ' όλη τη διάρκεια του μεταπτυχιακού μας προγράμματος.



## Περίληψη

Η παρούσα εργασία αναφέρεται στα μοντέλα αξιολόγησης πιστοληπτικής ικανότητας. Αρχικά παρουσιάζεται το απαραίτητο θεωρητικό υπόβαθρο και γίνεται μια ιστορική αναδρομή επί του θέματος. Γίνεται αναφορά στο κανονιστικό πλαίσιο της Βασιλείας II για το πώς επιβάλλεται τα τραπεζικά ιδρύματα να διαχειρίζονται τον πιστωτικό κίνδυνο. Στη συνέχεια αναπτύσσεται εκτενέστερα η έννοια του πιστωτικού κινδύνου και παρατίθενται οι κατηγορίες των μοντέλων βαθμονόμησης πιστωτικού κινδύνου. Ακόμη περιγράφονται οι τρεις κύριες στατιστικές μέθοδοι που χρησιμοποιούνται για την κατασκευή μοντέλων βαθμονόμησης πιστωτικού κινδύνου. Τέλος, η εργασία ολοκληρώνεται με πρακτική εφαρμογή σε αληθινά δεδομένα από τη βάση δεδομένων δύο ελληνικών τραπεζών, χρησιμοποιώντας μία από τις τρεις μεθοδολογίες που έχουν αναπτυχθεί σε αυτή την εργασία.



## **Abstract**

The topic of the present MSc thesis is the Credit Scoring Models. Initially all necessary theoretical background is presented, along with a historical review on the topic. Reference is made to the regulatory framework of Basel II and the need all financial institutions to comply with that in order to effectively manage credit risk. Subsequently, several aspects of Credit Risk and the subcategories of Credit Scoring Models are discussed in some detail. Furthermore, three of the most important statistical methods widely used for the creation of Credit Scoring Models are described. Finally an application on real data, extracted from the databases of two Greek banks, is presented where one of those three methods is applied and several useful conclusions are drawn.



# Περιεχόμενα

<b>ΚΕΦΑΛΑΙΟ 1</b>	<b>1</b>
<b>ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ</b>	<b>1</b>
1.1 Μοντέλα Αξιολόγησης Πιστοληπτικής Ικανότητας	1
1.2 Το Σύμφωνο της Βασιλείας	3
1.3 Περιεχόμενα της διπλωματικής	5
<b>ΚΕΦΑΛΑΙΟ 2</b>	<b>7</b>
<b>ΒΑΣΙΛΕΙΑ II ΚΑΙ ΠΙΣΤΩΤΙΚΟΣ ΚΙΝΔΥΝΟΣ</b>	<b>7</b>
2.1 Βασιλεία I	7
2.2 Βασιλεία II	8
2.3 Πιστωτικός Κίνδυνος	10
2.3.1 Τυποποιημένη Μέθοδος	11
2.3.2 Μέθοδος Εσωτερικών Διαβαθμίσεων	16
<b>ΚΕΦΑΛΑΙΟ 3</b>	<b>19</b>
<b>ΜΟΝΤΕΛΑ ΒΑΘΜΟΝΟΜΗΣΗΣ ΠΙΣΤΩΤΙΚΟΥ ΚΙΝΔΥΝΟΥ</b>	<b>19</b>
3.1 Εισαγωγή	19
3.2 Κατηγορίες Μοντέλων Βαθμονόμησης Πιστοληπτικής Αξιολόγησης	20
3.3 Είδη Μοντέλων Ανάλογα με τον Τύπο Δεδομένων	26
3.4 Διαδικασία Ανάπτυξης Μοντέλων Πιστωτικού Κινδύνου	27
<b>ΚΕΦΑΛΑΙΟ 4</b>	<b>29</b>
<b>ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΙ ΚΑΤΑΣΚΕΥΗΣ ΣΚΟΡΟΧΑΡΤΩΝ</b>	<b>29</b>
4.1 Λογιστική Παλινδρόμηση	29
4.2 Διαχωριστική Ανάλυση	33
4.3 Μέθοδος του κοντινότερου Γείτονα	37
<b>ΚΕΦΑΛΑΙΟ 5</b>	<b>39</b>
<b>ΕΦΑΡΜΟΓΗ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΣΕ ΔΕΔΟΜΕΝΑ ΚΑΘΥΣΤΕΡΗΣΕΩΝ</b>	<b>39</b>
5.1 Γενικά Χαρακτηριστικά Δειγμάτων	39
5.2 Ανάλυση των δεδομένων	43
5.3 Ανάπτυξη Μοντέλων Λογιστικής Παλινδρόμησης	55
<b>ΚΕΦΑΛΑΙΟ 6</b>	<b>63</b>
<b>ΑΞΙΟΛΟΓΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ</b>	<b>63</b>
6.1 Εξαγωγή συμπερασμάτων με τη βοήθεια καμπύλων ROC	63

<b>6.2 Συμπεράσματα – Συζήτηση</b>	<b>65</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>	<b>67</b>



# Κατάλογος Πινάκων

Πίνακας 5.2.1 Περιγραφικά στατιστικά μέτρα για εξυπηρετούμενα δάνεια.....	43
Πίνακας 5.2.2 Περιγραφικά στατιστικά μέτρα για μη εξυπηρετούμενα δάνεια .	<b>Error!</b>
<b>Bookmark not defined.</b>	
Πίνακας 5.2.3 Δείκτης καθυστέρησης 90 ημερών .....	45
Πίνακας 5.2.4 Πελάτες .....	46
Πίνακας 5.2.5 Φύλο.....	47
Πίνακας 5.2.6 Οικογενειακή κατάσταση.....	48
Πίνακας 5.2.7 Επάγγελμα.....	48
Πίνακας 5.2.8 Ιδιοκτησία .....	49
Πίνακας 5.2.9 Κάτοχος πιστωτικής κάρτας .....	49
Πίνακας 5.2.10 Χειρότερη κατάσταση Τειρεσία .....	50
Πίνακας 5.2.11 Περιγραφικά στατιστικά .....	51
Πίνακας 5.2.12 Ένδειξη Αναδιάρθρωσης.....	52
Πίνακας 5.2.13 Κατάσταση Επιχείρησης.....	52
Πίνακας 5.2.14 Μέγιστο Επίπεδο Καθυστέρησης Δανείου .....	53
Πίνακας 5.2.15 Μέγιστο Επίπεδο Καθυστέρησης του Πελάτη.....	54
Πίνακας 5.3.1 Μεταβλητές που περιλαμβάνονται στο 1 <sup>ο</sup> μοντέλο.....	56
Πίνακας 5.3.2 Μεταβλητές που δεν περιλαμβάνονται στο 1 <sup>ο</sup> μοντέλο .....	57
Πίνακας 5.3.3 Αξιολόγηση του 1 <sup>ου</sup> μοντέλου.....	57
Πίνακας 5.3.4 Ερμηνευσιμότητα του 1 <sup>ου</sup> μοντέλου .....	57
Πίνακας 5.3.5 Πίνακας ταξινόμησης για το 1 <sup>ο</sup> μοντέλο.....	58
Πίνακας 5.3.6 Μεταβλητές που περιλαμβάνονται στο 2 <sup>ο</sup> μοντέλο.....	58
Πίνακας 5.3.7 Μεταβλητές που δεν περιλαμβάνονται στο 2 <sup>ο</sup> μοντέλο .....	61
Πίνακας 5.3.8 Αξιολόγηση του 2 <sup>ου</sup> μοντέλου.....	61
Πίνακας 5.3.9 Ερμηνευσιμότητα του 2 <sup>ου</sup> μοντέλου .....	62
Πίνακας 5.3.10 Πίνακας ταξινόμησης για το 2 <sup>ο</sup> μοντέλο.....	62
Πίνακας 6.1.1 Καμπύλη ROC για το 1 <sup>ο</sup> μοντέλο .....	65
Πίνακας 6.1.2 Καμπύλη ROC για το 2 <sup>ο</sup> μοντέλο .....	65



## Κατάλογος Σχημάτων

Σχήμα 5.2.1 Συγκριτικό θηκόγραμμα ηλικιών .....	44
Σχήμα 5.2.2 Ιστόγραμμα ετήσιου εισοδήματος.....	45
Σχήμα 5.2.3 Διάγραμμα πίτας του δείκτη καθυστέρησης 90 ημερών .....	46
Σχήμα 5.2.4 Διάγραμμα πίτας της μεταβλητής πελάτης ή μη της τράπεζας.....	47
Σχήμα 6.1.1 Καμπύλη ROC 1 <sup>ου</sup> μοντέλου .....	64
Σχήμα 6.1.2 Καμπύλη ROC 2 <sup>ου</sup> μοντέλου .....	64



# ΚΕΦΑΛΑΙΟ 1

## ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

### 1.1 Μοντέλα Αξιολόγησης Πιστοληπτικής Ικανότητας

Η ιστορία της πιστοδότησης ξεκίνησε όταν ο άνθρωπος άρχισε να οργανώνεται σε κοινωνίες. Τα κοινά συμφέροντα και η αμοιβαία συνεργασία για επιβίωση δημιούργησε την πρώτη πίστη. Η πιο απλή έκφραση της πιστοδότησης εμφανίστηκε σε προσωπικό επίπεδο, με την έννοια ότι μια πλεονασματική μονάδα χορηγούσε κεφάλαια οποιασδήποτε μορφής σε μια άλλη μονάδα, με σκοπό την αμοιβαία εξυπηρέτηση και το αμοιβαίο όφελος.

Παράλληλα με το σχηματισμό των πρώτων τοπικών κοινωνιών, εμφανίστηκε και η ανάγκη για τη διενέργεια πιο περίπλοκων οικονομικών συναλλαγών και έτσι εισήχθη το πρώτο νόμισμα. Ακολούθησε ο δανεισμός σε νομίσματα πλέον, αφού άνθρωποι που για κάποια χρονική περίοδο είχαν ένα πλεόνασμα νομισμάτων δέχονταν προτάσεις από άτομα που είχαν έλλειμμα, ώστε να τους δανείσουν τα κεφάλαια τους με αντάλλαγμα να τους τα επιστρέψουν με ένα παραπάνω ποσό, που αντιστοιχεί στην απόδοση του δανειστή. Εισάγεται έτσι η έννοια “πίστης” και εκφράζεται από τον δανειστή ως προς το ότι ο δανειζόμενος που θα λάβει τα χρήματά του, θα καταφέρει μετά το πέρας της συμφωνίας να επιστρέψει το κεφάλαιο και την απόδοση. Η μη δυνατή είσπραξη της απόδοσης ή του κεφαλαίου και της απόδοσης οδήγησε με την πάροδο του χρόνου πολλούς δανειστές να σκέφτονται τι δεν πήγε καλά στον πιστούχο και δεν μπόρεσε να ανταπεξέλθει στη συμφωνία ώστε να επιστρέψει τα χρήματα. Συνεπώς σύντομα η οξυδέρκεια, ο προβληματισμός, και για πολλούς η ικανότητα και η τύχη χαρακτήριζαν τον “επιτυχημένο δανειστή”. Ταυτόχρονα η φήμη, η κινητή και ακίνητη περιουσία, η ρητορική και η πειθώ, αλλά και ο σωστός προγραμματισμός χαρακτήριζαν τον “προτιμητέο” πιστούχο ώστε να εξασφαλιστεί η επιστροφή του κεφαλαίου και της απόδοσης στον δανειστή του.

Από τα παραπάνω αντιλαμβανόμαστε ότι η ιστορία της χορήγησης πιστώσεως ξεκινάει από παλιά. Παρ’ όλ’ αυτά η ανάπτυξη μοντέλων

αξιολόγησης της πιστοληπτικής ικανότητας του δανειολήπτη, δηλαδή η ποσοτικοποίηση των παραγόντων που επηρεάζουν την ικανότητά του να εκπληρώσει την υποχρέωση που έχει απέναντι στο δανειστή του, είναι αρκετά πρόσφατη. Συγκεκριμένα χρονολογείται το 1941, όποτε ο Davis Durand δημοσίευσε έρευνα που του είχε ανατεθεί το 1937 από τον Αμερικάνικο, μη-κερδοσκοπικό οργανισμό “National Bureau of Economic Research” (Durand, D.. (1941) Risk Elements in Consumer Instalment Financing, National Bureau of Economic Research, New York). Κατά την έρευνα αυτή, παρατηρήθηκε πτώση στην χρηματοδότηση καταναλωτικής πίστης σε σχέση με άλλες κατηγορίες χρηματοδότησης, ενώ με τη χρήση της στατιστικής αναλύθηκε δείγμα 7200 καταναλωτικών δανείων από 37 διαφορετικά πιστωτικά ιδρύματα και μελετήθηκαν οι παράγοντες του πιστωτικού κινδύνου στη χρηματοδότηση των καταναλωτών. Η μελέτη του Durand θεωρείται πως έθεσε τους πυλώνες του πλαισίου των τεχνικών αξιολόγησης πιστοληπτικής ικανότητας και άνοιξε το δρόμο για περαιτέρω έρευνα.

Μετά το τέλος του Β' παγκοσμίου πολέμου άρχισαν να αναπτύσσονται στατιστικές και αριθμητικές μέθοδοι για την αυτοματοποιημένη λήψη αποφάσεων για πιστοδότηση. Η πρώτη συμβουλευτική προσπάθεια προς τραπεζικά ιδρύματα, λιανοπωλητές και εταιρίες ταχυδρομικών πωλήσεων άρχισε στο San Francisco από τους Bill Fair and Earl Isaac. Την δεκαετία του 1960 η εμφάνιση των πιστωτικών καρτών και η μαζική πιστοδότηση από έναν εκδότη έκανε επιτακτική την ανάγκη για αξιοποίηση και ανάπτυξη μοντέλων πιστοληπτικής αξιολόγησης. Οι αιτήσεις για χορήγηση πιστωτικών καρτών αυξάνονταν μέρα με τη μέρα με γεωμετρικούς ρυθμούς με αποτέλεσμα το ανθρώπινο δυναμικό να μην επαρκεί για να ανταπεξέλθει στις απαιτήσεις της αγοράς. Ο έλεγχος της πιστοδότησης απομακρυνόταν από τη διοίκηση και η ανάγκη για κεντρικό έλεγχο και αυτοματοποιημένο σύστημα αποφάσεων ήταν άμεση.

Η παράλληλη ανάπτυξη της υπολογιστικής ισχύος των υπολογιστών βοήθησε σημαντικά τη διαδικασία πιστοληπτικής αξιολόγησης και έκανε πιο αποτελεσματικές τις αποφάσεις οδηγώντας σε 50% μείωση τις πτωχεύσεις των δανειοληπτών που λάμβαναν χρηματοδότηση με βάση τις νέες διαδικασίες. Από τότε άρχισε να αναπτύσσεται η ιδέα ότι πρέπει να βρεθούν οι παράγοντες

και τα χαρακτηριστικά ενός δανειολήπτη που επηρεάζουν και συσχετίζονται με την πιστοληπτική ικανότητα και θα έπρεπε συνεπώς να περιέχονται σε ένα έγκυρο σύστημα βαθμολόγησης. Η επίσημη πια αναγνώριση της αξίας των μεθόδων πιστοληπτικής αξιολόγησης έγινε με την θέσπιση από τις Η.Π.Α του νομοθετικού πλαισίου Equal Credit Opportunity Act (ECOA) που εφαρμόστηκε το 1974. Το ECOA υποχρέωνε τους πιστοδότες να μην περιλαμβάνουν παράγοντες φυλετικών ή άλλων διακρίσεων στο σύστημα πιστοληπτικής τους αξιολόγησης όπως η φυλή, το χρώμα του δέρματος, το φύλο κ.ά. Τα τελευταία 25 έτη η έρευνα στο συγκεκριμένο πεδίο έχει περάσει σε ανώτερο επιστημονικό επίπεδο και συνεχίζεται ακόμη και σήμερα. Η μεγάλη επιτυχία των μεθόδων για την αξιολόγηση των υποψήφιων δανειοληπτών στα δάνεια καταναλωτών έχει οδηγήσει τα τραπεζικά ιδρύματα στο να χρησιμοποιούν ευρέως τις μεθόδους αυτές σε αποφάσεις χρηματοδότησης όλων των αγορών.

## **1.2 Το Σύμφωνο της Βασιλείας**

Μετά το τέλος του Β΄ παγκοσμίου πολέμου η εκτενής εξάπλωση των τραπεζικών ιδρυμάτων με σκοπό την διαμεσολάβηση στην δανειοδότηση από τα πλεονασματικά νοικοκυριά, επιχειρήσεις ακόμη και κράτη προς άλλες ομοειδείς μονάδες έκανε επιτακτική την ανάγκη θέσπισης κανόνων για την προστασία αυτού του θεσμού. Στο δυτικό κόσμο η εισαγωγή του θεσμού αυτού είναι από τους παράγοντες που έχουν φέρει ανάπτυξη και έχουν βελτιώσει το βιοτικό επίπεδο των πολιτών.

Στην προηγούμενη ενότητα είδαμε την εξέλιξη στο χρόνο των μεθόδων πιστοληπτικής αξιολόγησης οι οποίες έχουν ως στόχο τη διασφάλιση της αποπληρωμής της υποχρέωσης από το δανειολήπτη. Ένα άλλο σκέλος για την ομαλή λειτουργία του τραπεζικού συστήματος και τη θωράκισή του από αρνητικά χρηματοοικονομικά συμβάντα, είναι η ενίσχυση της εμπιστοσύνης των καταθετών και η προστασία των κεφαλαίων τους. Αυτό επετεύχθη με το σύμφωνο της Βασιλείας.

Το 1930 ιδρύθηκε στην πόλη Βασιλεία της Ελβετίας η Τράπεζα Διεθνών Διακανονισμών (Bank of International Settlements) με σκοπό την επιτήρηση των πληρωμών για την αποκατάσταση των ζημιών που προκλήθηκαν από τον

πόλεμο, αλλά γρήγορα εξελίχθηκε σε μέσο για τη συνεργασία μεταξύ των κεντρικών τραπεζών με σκοπό τη νομισματική και οικονομική σταθερότητα. Έπειτα από τη σημαντική παγκόσμια οικονομική δυσπραγία του 1987, και συγκεκριμένα στις 15 Ιουλίου του 1988 δημοσιεύθηκε το Σύμφωνο της Βασιλείας (Basel Capital Accord) για την κεφαλαιακή Επάρκεια με τίτλο «Διεθνής σύγκλιση της Κεφαλαιακής Μέτρησης και των Κεφαλαιακών Προτύπων» (International Convergence of Capital Measurement and Capital Standards). Στόχος της συνθήκης, που χαρακτηρίζεται από πολλούς ως ορόσημο για την λειτουργία των τραπεζικών ιδρυμάτων, ήταν όπως προαναφέρθηκε η θωράκιση του τραπεζικού συστήματος από αρνητικά χρηματοοικονομικά συμβάντα, με απώτερο σκοπό την ενίσχυση της εμπιστοσύνης των καταθετών, που είναι και οι βασικοί χρηματοδότες του, προς το παγκόσμιο χρηματοπιστωτικό σύστημα.

Το σύμφωνο της Βασιλείας αποτέλεσε τη βάση για την εποπτεία των τραπεζικών ιδρυμάτων σε πάνω από 100 χώρες σε όλο την κόσμο. Η κάθε κρατική εποπτική αρχή προσάρμοσε το πλαίσιο στις ανάγκες της ανάλογα με την διάρθρωση της οικονομίας της και το επίπεδο ανάπτυξης του τραπεζικού της συστήματος. Ο πρώτος κίνδυνος που έλαβαν υπόψη τους οι συντάκτες του συμφώνου ήταν ο πιστωτικός κίνδυνος (Credit Risk) που είναι ο κίνδυνος μη εκπλήρωσης των συμβατικών υποχρεώσεων του δανειζόμενου προς το τραπεζικό ίδρυμα που είχε πιστοδοτήσει τον αντισυμβαλλόμενο, δηλαδή ο κίνδυνος μη αποπληρωμής του κεφαλαίου και των τόκων έπειτα από την χορήγηση κάποιου κεφαλαίου σε έναν δανειολήπτη. Η σκέψη του συμβουλίου ήταν ότι, ένα πιστωτικό ίδρυμα που έχει κύριους χρηματοδότες τους καταθέτες, θα έπρεπε να προστατευτεί επιβάλλοντας στους μετόχους του τραπεζικού ιδρύματος, που έχουν τον έλεγχο του μέσω του ορισμού του διοικητικού συμβουλίου, να προσαρμόζουν τα ίδια κεφάλαιά τους ανάλογα με την έκθεση της συνολικής περιουσίας του ιδρύματος στον πιστωτικό κίνδυνο.

Η Βασιλεία I εισήγαγε ένα απλό σύστημα υπολογισμού των ελαχίστων ιδίων κεφαλαίων που έπρεπε να διατηρούν οι μέτοχοι ανάλογα με τις τοποθετήσεις που επέλεγε να κάνει η τράπεζα. Το σύστημα αυτό όριζε βάρη ανάλογα με το είδος της τοποθέτησης που είχε η κάθε τράπεζα στο ενεργητικό της. Με αυτόν τον τρόπο υπολογιζόταν το σταθμισμένο στον κίνδυνο



ενεργητικό RWA (Risk Weighted Assets) και ένας απλός τύπος υπολόγιζε μέσω αυτού τα ελάχιστα απαιτούμενα ίδια κεφάλαια που έπρεπε να διατηρούν οι μέτοχοι στην τράπεζα.

Το 2001 δημοσιεύθηκε το πρώτο συμβουλευτικό έγγραφο που αφορούσε στη Βασιλεία II ενώ τον Ιούνιο του 2004 εκδόθηκε το τελικό έγγραφο με μια δεύτερη αναθεώρηση το Νοέμβριο του 2005. Η εφαρμογή της συμφωνίας αναμενόταν έως το 2008 και στόχευε στο να επιβάλει στους ιδιοκτήτες και στους μετόχους των τραπεζικών ιδρυμάτων, να διατηρούν τα απαιτούμενα κεφάλαια για την προστασία των καταθετών τους (δανειστών τους) ώστε τα χρήματά τους να μην καταλήξουν σε ελλειμματικές μονάδες που δεν θα μπορούσαν να αποπληρώσουν τις υποχρεώσεις τους.

Στο εποπτικό πλαίσιο της Βασιλείας II ο πρώτος και σημαντικότερος κίνδυνος που αναγνωρίζεται είναι ο πιστωτικός και είναι υπεύθυνος, όπως έχει δείξει η εφαρμογή του, για το 80% περίπου των απαιτούμενων εποπτικών κεφαλαίων.

Κλείνοντας την ιστορική αναδρομή, αναφέρεται ότι το 2011 και ως απόρροια της οικονομικής κρίσης του 2007-2008, δημοσιεύθηκε το σύμφωνο της Βασιλείας III με στόχο την εφαρμογή του έως το 2015, ενώ κάποιες αναθεωρήσεις το 2013 μετέθεσαν την εφαρμογή του πλαισίου έως το 2019.

### **1.3 Περιεχόμενα της διπλωματικής**

Στην παρούσα εργασία αφού παρατεθεί το απαραίτητο θεωρητικό υπόβαθρο για τα μοντέλα βαθμονόμησης πιστωτικού κινδύνου θα προχωρήσουμε στην παρουσίαση του τρόπου εφαρμογής μιας μεθόδου αξιολόγησης σε πραγματικά δεδομένα.

Στο πρώτο κεφάλαιο έχει γίνει μια σύντομη εισαγωγή στην έννοια του πιστωτικού κινδύνου και μια σύντομη ιστορική αναδρομή. Στο δεύτερο κεφάλαιο αναπτύσσεται το σύμφωνο της Βασιλείας II και αναπτύσσεται εκτενέστερα η έννοια του πιστωτικού κινδύνου. Στη συνέχεια, στο τρίτο κεφάλαιο παρουσιάζονται οι κατηγορίες των μοντέλων βαθμονόμησης πιστωτικού κινδύνου. Στο τέταρτο κεφάλαιο αναλύονται τρεις τεχνικές κατασκευής σκορόχαρτων με έμφαση τη μαθηματική περιγραφή των μεθόδων.

Πιο συγκεκριμένα παρουσιάζεται η τεχνική της Λογιστικής Παλινδρόμησης, η Διαχωριστική Ανάλυση και η μέθοδος του k-κοντινότερου γείτονα. Η εφαρμογή της Λογιστικής Παλινδρόμησης σε πραγματικά δεδομένα παρουσιάζεται στο πέμπτο κεφάλαιο ενώ η εργασία ολοκληρώνεται με τα συμπεράσματα της πρακτικής εφαρμογής στο έκτο και τελευταίο κεφάλαιο.

## ΚΕΦΑΛΑΙΟ 2

### ΒΑΣΙΛΕΙΑ Ι ΚΑΙ ΠΙΣΤΩΤΙΚΟΣ ΚΙΝΔΥΝΟΣ

#### 2.1 Βασιλεία Ι

Η Επιτροπή της Βασιλείας Ι για την Τραπεζική Εποπτεία (Basel I Committee on Banking Supervision) ιδρύθηκε το 1974 υπό την αιγίδα της Τράπεζας Διεθνών Διακανονισμών και αποβλέπει κυρίως στη διασφάλιση της σταθερότητας του διεθνούς χρηματοπιστωτικού συστήματος. Έτσι στοχεύει στην καθιέρωση αποδεκτών κανόνων με σκοπό την ενίσχυση της ικανότητας απορρόφησης ζημιών σε περίπτωση που επέλθουν μη προβλέψιμοι κίνδυνοι στους οποίους εκτίθενται οι τράπεζες με τη λειτουργία τους. Η επιτροπή της Βασιλείας δεν έχει νομική προσωπικότητα ούτε μορφή υπερεθνικής εποπτικής αρχής και επομένως οι προτάσεις της δεν έχουν νομική δεσμευτική ισχύ, αλλά οι εποπτικές αρχές κάθε χώρας έχουν την ευθύνη για την ενσωμάτωση και την εφαρμογή των κανόνων σε εθνικό επίπεδο.

Το πρώτο Σύμφωνο της Βασιλείας που δημοσιεύθηκε τον Ιούλιο του 1988 αφορούσε στον πιστωτικό κίνδυνο. Ο πιστωτικός κίνδυνος είναι ο σημαντικότερος σε ένα τραπεζικό οργανισμό, καθώς είναι ο κίνδυνος που συνδέεται με την κύρια δραστηριότητα του ιδρύματος, την πιστοδότηση. Έτσι το κείμενο του συμφώνου εστίασε σε κανόνες προκειμένου τα χρήματα των καταθετών και των άλλων πιστωτών των τραπεζικών ιδρυμάτων, να μην καταλήξουν σε πιστούχους που δεν θα έχουν την δυνατότητα να ανταποκριθούν στις υποχρεώσεις τους. Είναι φανερό ότι αν δεν είχαν θεσπιστεί κανόνες για να μη χάνονται τα χρήματα των πιστωτών η εμπιστοσύνη στο τραπεζικό σύστημα θα ήταν περιορισμένη.

Το κύριο σημείο ενδιαφέροντος είναι η κεφαλαιακή επάρκεια, δηλαδή οι εποπτικές αρχές υποχρεώνουν τα τραπεζικά ιδρύματα σε παρακράτηση κεφαλαίων ώστε να περιορίσουν τον κίνδυνο αθέτησης υποχρεώσεων. Το 8% ορίστηκε ως δείκτης ελάχιστης κεφαλαιακής επάρκειας, πιο συγκεκριμένα τα ιδρύματα θα πρέπει να φροντίσουν ο λόγος των ιδίων κεφαλαίων προς το

σταθμισμένο στον κίνδυνο ενεργητικό να μην υπολείπεται σε καμία περίπτωση του 8%. Πιο συγκεκριμένα θα πρέπει να ισχύει

$$\text{Minimum Capital Requirements} \geq 8\% \times RWA.$$

## 2.2 Βασιλεία II

Το 1996 η επιτροπή της Βασιλείας με μια σημαντική τροποποίηση του κειμένου, εισήγαγε έναν επιπλέον κίνδυνο που θα έπρεπε να προσμετρούν τα τραπεζικά ιδρύματα για τον υπολογισμό των ελάχιστων κεφαλαιακών απαιτήσεων, τον κίνδυνο αγοράς (Market Risk). Ο κίνδυνος αγοράς είναι ο κίνδυνος που αντιμετωπίζουν τα πιστωτικά ιδρύματα από τις διακυμάνσεις στις αξίες των στοιχείων του ενεργητικού τους. Η αξία των στοιχείων του ενεργητικού έχει διακύμανση που το ύψος της σχετίζεται άμεσα με το είδος του κάθε στοιχείου. Για παράδειγμα οι αξίες των μετοχών που μπορεί να διατηρεί ένα τραπεζικό ίδρυμα στο ενεργητικό, έχουν μεγαλύτερη διακύμανση στο χρόνο σε σχέση με την διακύμανση στις αξίες κρατικών ομολόγων.

Μετά το 1996 άρχισε η αναθεώρηση του πλαισίου της Βασιλείας I. Σε συνεργασία με την τραπεζική βιομηχανία, η επιτροπή της Βασιλείας έφτασε σε συμφωνία για την νέα μορφή του πλαισίου. Έπειτα από αρκετές μελέτες επίπτωσης στην τραπεζική φερεγγυότητα (Quantitative Impact Studies – QIS) συμφωνήθηκε το 2004 το κύριο σχέδιο της Βασιλείας II και το τελικό κείμενο δημοσιεύθηκε τον Ιούνιο του 2008 με τίτλο «International Convergence of Capital Measurement and Capital Standards A Revised Framework Comprehensive Version» (*Basel Committee on Banking Supervision (BCBS) (2006), International Convergence of Capital Measurement and Capital Standards – A Revised Framework (Comprehensive Version), Bank for International Settlements, Basel*). Το εποπτικό σχέδιο αυτό, ενσωματώθηκε στην νομοθεσία όλων των ενδιαφερόμενων κρατών.

Στο σύμφωνο της Βασιλείας II υπήρξε σημαντική αναθεώρηση για τον πιστωτικό κίνδυνο. Ο νέος κανονισμός είναι πιο ευαίσθητος στον πιστωτικό κίνδυνο σε σύγκριση με το παλαιότερο πλαίσιο ενώ αναγνωρίστηκαν τεχνικές μετρίωσης του πιστωτικού κινδύνου όπως είναι η εμπράγματη εξασφάλιση, οι εγγυήσεις, τα παράγωγα πιστωτικού κινδύνου, η διαφοροποίηση-διασπορά του

κινδύνου γεωγραφικά και μεταξύ των κλάδων μιας οικονομίας. Ακόμη δόθηκε η δυνατότητα οι ελάχιστες κεφαλαιακές απαιτήσεις να υπολογίζονται με διαφορετικές μεθοδολογίες.

Το νέο πλαίσιο προτείνει την τυποποιημένη μέθοδο υπολογισμού ελάχιστων κεφαλαιακών απαιτήσεων η οποία έχει ως πρότυπο μια τυπική τράπεζα. Ακόμη επιτρέπει τη δυνατότητα στα τραπεζικά ιδρύματα να υπολογίσουν τα απαιτούμενα κεφάλαια και με δυο επιπλέον πιο εξελιγμένες μεθόδους οι οποίες είναι πιο ευαίσθητες στον εκάστοτε κίνδυνο.

Επίσης, πέρα από τον κίνδυνο αγοράς που είχε εισαχθεί με την τροποποίηση του πλαισίου της Βασιλείας I το 1996, η Βασιλεία II εισάγει έναν ακόμη κίνδυνο προς υπολογισμό. Ο νέος κίνδυνος είναι ο λειτουργικός κίνδυνος (Operational Risk) και είναι αυτός που πηγάζει από τη δυσλειτουργία των λειτουργικών διαδικασιών της τράπεζας σε κάθε επιχειρηματικό τομέα που δραστηριοποιείται η κάθε τράπεζα (π.χ. κακή λειτουργία μηχανογραφικών συστημάτων, εσωτερικές και εξωτερικές απάτες κ.α.).

Ο νέος τύπος υπολογισμού των κεφαλαιακών απαιτήσεων λαμβάνει την παρακάτω μορφή:

$$MCR = 0.08 \cdot (RWA_{Credit Risk} + RWA_{Oper. Risk} + RWA_{Market Risk})$$

Πολλές εμπειρικές μελέτες αναφέρουν ότι σε μια μέση τράπεζα το 80% των απαιτούμενων ιδίων κεφαλαίων είναι απαραίτητα για να καλυφθούν απαιτήσεις που οφείλονται στον πιστωτικό κίνδυνο, το 15% για τις απαιτήσεις σε κεφάλαια από τον λειτουργικό κίνδυνο και το 5% για απαιτήσεις που προκύπτουν από τον κίνδυνο αγοράς. Η παραπάνω αναλογία είναι και αναμενόμενη καθώς ένα τραπεζικό ίδρυμα είναι ο συνδετικός κρίκος μεταξύ ελλειμματικών και πλεονασματικών μονάδων μιας οικονομίας που επιθυμούν να συνδιαλλαγούν και η κύρια δραστηριότητά του είναι η διαχείριση πιστώσεων. Κατά συνέπεια ο σημαντικότερος κίνδυνος είναι ο πιστωτικός.

Το πλαίσιο της Βασιλείας II προβλέπει μία προσέγγιση αρτιότερη και περισσότερο προσαρμοσμένη στη φύση των αναλαμβανόμενων κινδύνων, προάγοντας την ενισχυμένη διαχείριση του κινδύνου από τα χρηματοπιστωτικά ιδρύματα, πράγμα που αναμενόταν να συμβάλει στη σταθερότητα του χρηματοπιστωτικού συστήματος, να εμπνεύσει εμπιστοσύνη στα

χρηματοπιστωτικά ιδρύματα και να ενισχύσει την προστασία του καταναλωτή. Συνοψίζοντας, το σύμφωνο της Βασιλείας II καθιέρωσε τους εξής τρεις θεμελιώδεις άξονες - Πυλώνες (Pillars) εποπτείας :

- Μεθόδους προσδιορισμού των κεφαλαιακών απαιτήσεων έναντι των κινδύνων τους οποίους τα πιστωτικά ιδρύματα κατά κανόνα αντιμετωπίζουν στο πλαίσιο της άσκησης των δραστηριοτήτων τους. Καθιερώνονται επίσης κεφαλαιακές απαιτήσεις και για το λειτουργικό κίνδυνο (Πυλώνας 1).
- Αρχές, κριτήρια και διαδικασία με την οποία καταρχάς τα ίδια τα πιστωτικά ιδρύματα και κατ' ακολουθίαν η εποπτική αρχή (π.χ η Τράπεζα της Ελλάδος) αξιολογούν την επάρκεια των κεφαλαίων και των συστημάτων διαχείρισης κινδύνων κάθε πιστωτικού ιδρύματος χωριστά, σε σχέση με τους πάσης φύσεως κινδύνους στους οποίους αυτό εκτίθεται ή ενδέχεται να εκτεθεί, πέραν από εκείνους που αντιμετωπίζονται στον Πυλώνα 1 (Πυλώνας 2).
- Υποχρέωση δημοσιοποίησης στοιχείων για την ενίσχυση της διαφάνειας και της πειθαρχίας της αγοράς με την παροχή στους ενδιαφερόμενους της δυνατότητας σύγκρισης τόσο της πολιτικής για τη διαχείριση κινδύνων, της κεφαλαιακής και οργανωτικής επάρκειας των πιστωτικών ιδρυμάτων (παρέχοντας έτσι κίνητρο για την βελτίωσή τους) όσο και των μεθόδων και πρακτικών που εφαρμόζουν οι εποπτικές αρχές (Πυλώνας 3).

### 2.3 Πιστωτικός Κίνδυνος

Ο πιστωτικός κίνδυνος (Credit Risk) αποτελεί τον κίνδυνο που μπορεί να αναλάβει ένας χρηματοοικονομικός οργανισμός και σχετίζεται άμεσα με την αξιολόγηση της δυνατότητας ενός δανειολήπτη να αποπληρώσει τις υποχρεώσεις του βάσει των αμοιβαίως συμφωνηθέντων όρων που καλύπτουν την πίστωση αυτή. Μπορεί να επηρεάσει την Κεφαλαιακή Επάρκεια και παρ' όλες τις εγγυήσεις που λαμβάνονται για κάθε παρεχόμενο προϊόν χρηματοδότησης, είναι αυξημένος τόσο λόγω του πλήθους των παρεχόμενων υπηρεσιών και προϊόντων, όσο και της διαφορετικότητας των πελατών στους

οποίους απευθύνονται τα χρηματοπιστωτικά ιδρύματα. Για τους λόγους αυτούς θα πρέπει να μετριέται με την μεγαλύτερη δυνατή ακρίβεια, έτσι ώστε η Τράπεζα να δεσμεύει τα λιγότερα δυνατά κεφάλαια για την κάλυψη τυχόν επισφαλειών.

Ο πιστωτικός κίνδυνος μπορεί να υπολογιστεί με τις εξής τρεις διαφορετικές μεθόδους:

1. την τυποποιημένη μέθοδο (Standardized Approach)
2. τη θεμελιώδη μέθοδο εσωτερικών διαβαθμίσεων (Foundation internal-rating-based approach, IRB)
3. την προχωρημένη μέθοδο εσωτερικών διαβαθμίσεων (Advanced internal-rating-based approach)

Η τυποποιημένη μέθοδος είναι η λιγότερη ευαίσθητη στον κίνδυνο και καθορίζει μια συντηρητική οπτική για το μέσο επίπεδο του κινδύνου. Οι μέθοδοι των εσωτερικών διαβαθμίσεων (IRB) επιτρέπουν στα πιστωτικά ιδρύματα να υπολογίζουν μόνα τους το επίπεδο κινδύνου που έχουν αναλάβει μέσω του αντίστοιχου εποπτικού πλαισίου. Οι εσωτερικές μέθοδοι είναι πιο ευαίσθητες στον κίνδυνο και τον προσεγγίζουν με μεγαλύτερη ακρίβεια. Το αποτέλεσμα είναι ότι αυτές οι μέθοδοι είναι λιγότερο συντηρητικές από την προσέγγιση της τράπεζας-πρότυπο στην τυποποιημένη μέθοδο.

### **2.3.1 Τυποποιημένη Μέθοδος**

Στην τυποποιημένη μέθοδο (Standardized Approach) η ποσοτικοποίηση του πιστωτικού κινδύνου γίνεται με τη βοήθεια εξωτερικών ιδρυμάτων αξιολόγησης πιστούχων. Ο κανονισμός της Βασιλείας II επιβάλλει τον υπολογισμό του σταθμισμένου στον κίνδυνο ενεργητικού (Risk Weighted Assets). Χρησιμοποιεί βάρη (ποσοστά) επί του ποσού έκθεσης για το κάθε στοιχείο του ενεργητικού της τράπεζας και με αυτόν τον τρόπο υπολογίζεται το σταθμισμένο στον κίνδυνο ενεργητικό. Τα βάρη για κάθε επίπεδο κινδύνου και για την κάθε κατηγορία των στοιχείων του ενεργητικού δίνονται παρακάτω.

*α. Κυβερνητικές πιστώσεις και Πιστώσεις Κεντρικών Τραπεζών*

Μια τράπεζα, εφόσον έχει στο ενεργητικό της επενδύσεις σε ομόλογα ή δάνεια προς κράτη (Claims on Sovereign and Central Banks), τα βάρη κάθε επιπέδου κινδύνου για τον υπολογισμό του σταθμισμένου στον κίνδυνο ενεργητικού παρουσιάζονται στον παρακάτω πίνακα.

Sovereigns						
Επίπεδο Κινδύνου	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Βάρη	0%	20%	50%	100%	150%	100%

Το πλαίσιο της βασιλείας παρέχει κάποιες εξαιρέσεις για Κεντρικές Τράπεζες και Διεθνείς Οργανισμούς όπως το Διεθνές Νομισματικό Ταμείο, την Τράπεζα Διεθνών Διακανονισμών, την Παγκόσμια Τράπεζα και την Ευρωπαϊκή Επιτροπή έτσι ώστε οι τράπεζες που έχουν επενδύσει κεφάλαια σε αυτούς τους οργανισμούς να μην επιβαρύνονται με κεφάλαια (δηλαδή αυτά τα στοιχεία του ενεργητικού θεωρείται ότι έχουν βάρος 0). Η ίδια διευκόλυνση (μηδενικής κεφαλαιακής επιβάρυνσης) παρέχεται και στις κατά τόπους Κεντρικές Τράπεζες που ελέγχουν την εφαρμογή του πλαισίου εντός νομισματικής εμβέλειας.

*β. Απαιτήσεις από Τράπεζες*

Το σταθμισμένο στον κίνδυνο ενεργητικό αυτής της κατηγορίας υπολογίζεται βάσει δυο προσεγγίσεων. Με την πρώτη προσέγγιση οι επενδύσεις σε τραπεζικά ιδρύματα λαμβάνουν το βάρος μιας κατηγορίας μικρότερης από αυτήν που αντιστοιχεί στη χώρα στην οποία εδρεύει το πιστοδοτούμενο τραπεζικό ίδρυμα. Ωστόσο οι απαιτήσεις από τράπεζες που εδρεύουν σε χώρα που εντάσσεται στην κατηγορία με διαβάθμιση από BB+ έως B- λαμβάνουν βάρος 100%. Ο παρακάτω πίνακας αφορά τα βάρη για την πρώτη επιλογή.



Banks (Option 1)						
Κατηγορία Κινδύνου Κράτους	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Βάρη	20%	50%	100%	100%	150%	100%

Η δεύτερη προσέγγιση επιτρέπει τον υπολογισμό των βαρών μέσω της διαβάθμισης από εξωτερικό οίκο αξιολόγησης. Γίνεται επιπλέον και μια κατηγοριοποίηση σε βραχυπρόθεσμα και μακροπρόθεσμα στοιχεία ενεργητικού. Οι υποχρεώσεις χωρίζονται σε εκείνες που έχουν διάρκεια ως την λήξη τους έως 3 μήνες και υποχρεώσεις που λήγουν σε περισσότερους από 3 μήνες. Τα βάρη για τη δεύτερη προσέγγιση δίνονται από τους παρακάτω πίνακες. Ο πρώτος πίνακας αφορά τις μακροπρόθεσμες υποχρεώσεις ενώ ο δεύτερος αφορά τις βραχυπρόθεσμες υποχρεώσεις.

Banks Long-term claims (Option 2)						
Κατηγορία Κινδύνου Τράπεζας	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Βάρη	20%	50%	50%	100%	150%	50%

Banks Short-term claims (Option 2)						
Κατηγορία Κινδύνου Τράπεζας	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Βάρη	20%	20%	20%	50%	150%	20%

*γ. Απαιτήσεις από επιχειρήσεις*

Οι απαιτήσεις από μεγάλες επιχειρήσεις (Claims on Corporates) είναι η επόμενη κατηγορία πιστωτικού κινδύνου η οποία μεταξύ άλλων αναφέρεται και σε κινδύνους που προέρχονται από ασφαλιστικές εταιρείες. Ο παρακάτω πίνακας παρουσιάζει τους συντελεστές ανάλογα με τη διαβάθμιση της κάθε εταιρείας από εξωτερικό οίκο αξιολόγησης. Οι μη διαβαθμισμένες εταιρείες δεν μπορούν να έχουν συντελεστή μικρότερο από τον αντίστοιχο που έχει η χώρα που εδρεύουν και δραστηριοποιούνται.

Επιχειρήσεις					
Κατηγορία κινδύνου επιχείρησης	AAA to AA-	A+ to A-	BBB+ to BB-	Below BB-	Unrated
Βάρη	20%	50%	100%	150%	100%

*δ. Απαιτήσεις Χαρτοφυλακίου Λιανικής Τραπεζικής*

Ο γενικός κανόνας στο χαρτοφυλάκιο λιανικής είναι ότι ο συντελεστής βάρους είναι στο 75%.

Το χαρτοφυλάκιο λιανικής αφορά σε απαιτήσεις από φυσικά πρόσωπα και μικρές επιχειρήσεις. Οι χορηγήσεις σε κάθε φυσικό πρόσωπο δεν πρέπει να ξεπερνούν το 0,2% του συνολικού χαρτοφυλακίου ή το 1 εκ. ευρώ. Ο κανόνας αυτός εξυπηρετεί το σκοπό της διαφοροποίησης του χαρτοφυλακίου της λιανικής, που απαιτείται από το πλαίσιο της Βασιλείας II.

Οι παραπάνω χορηγήσεις αφορούν πιστωτικές κάρτες, όρια υπερανάληψης, καταναλωτικά δάνεια, χρηματοδοτικές μισθώσεις (leasing), δάνεια σπουδών, δάνεια εγκαταστάσεων κ.α.

Απαιτήσεις που είναι πλήρως εξασφαλισμένες με υποθήκη κατοικίας (residential mortgages) έχουν βάρος 35%.

Απαιτήσεις που έχουν ως εξασφάλιση εμπορικό ακίνητο έχουν βάρος 50% για ανεπτυγμένες και μακροβιότερες αγορές. Διαφορετικά σε οποιαδήποτε άλλη περίπτωση το βάρος είναι 100%.

*ε. Ληξιπρόθεσμα δάνεια & κατηγορίες υψηλού κινδύνου*

Στη Βασιλεία II εφαρμόζονται ειδικά βάρη για δάνεια τα οποία είναι ληξιπρόθεσμα (Past due loans) για πάνω από 90 ημέρες. Ο συντελεστής-βάρους για αυτά τα δάνεια είναι 100% ή 150% όταν οι προβλέψεις (Provisions) είναι πάνω ή κάτω από 20% του υπολοίπου του δανείου αντίστοιχα. Επίσης όταν οι προβλέψεις είναι πάνω από 50% με την σύμφωνη γνώμη των κατά τόπους εποπτικών αρχών το βάρος μπορεί να μειωθεί και στο 50%.

*στ. Στοιχεία Ενεργητικού Υψηλού Κινδύνου*

Οι κατηγορίες υψηλού κινδύνου (Higher-Risk categories) λαμβάνουν βάρος από 150% και πάνω και είναι οι παρακάτω:

- Πιστώσεις σε κράτη, δημοσίους οργανισμούς, τράπεζες με διαβάθμιση μικρότερη από B-.
- Πιστώσεις σε επιχειρήσεις με διαβάθμιση μικρότερη από BB-.
- Ληξιπρόθεσμα δάνεια όπως ορίστηκαν παραπάνω.
- Τιτλοποιημένα προϊόντα με διαβάθμιση μέχρι BB-.

Επίπεδο κινδύνου υλοποιημένου προϊόντος	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to BB-
Βάρος	20%	50%	100%	350%

*ζ. Λοιπά στοιχεία ενεργητικού*

Τα λοιπά στοιχεία του ενεργητικού (Other Assets) ενός τραπεζικού ιδρύματος όπως οι μετοχές έχουν ως γενικό κανόνα βάρος ίσο με 100% ή 150%.

*η. Στοιχεία εκτός Ισολογισμού*

Τα στοιχεία του ενεργητικού που δεν αναγράφονται στον ισολογισμό (Off-balance sheet items) μετατρέπονται σε πιστωτικά στοιχεία μέσω του

πιστωτικού παράγοντα μετατροπής (Credit Conversion Factor - CCF) μέσω της εξίσωσης:

$$\text{Ισοδύναμο πιστωτικό άνοιγμα} = CCF \times \text{off Balance sheet exposures}$$

Στον παραπάνω τύπο ο παράγοντας CCF λαμβάνει τις εξής τιμές ανά περίπτωση:

- CCF= 0% για δεσμεύσεις χωρίς όρους.
- CCF= 20% για δεσμεύσεις με ωρίμανση μέχρι και έτος, βραχείας διάρκειας παράγωγα κ.α.
- CCF= 50% για δεσμεύσεις με ωρίμανση πάνω από ένα έτος, μακροχρόνιας διάρκειας δεσμεύσεις όπως παράγωγα με διάρκεια μεγαλύτερη του έτους.
- CCF= 100% για δανεισμό χρεογράφων ή παροχή χρεογράφων ως εξασφάλιση.

### 2.3.2 Μέθοδος Εσωτερικών Διαβαθμίσεων

Ο γενικός κανόνας για αυτή τη μέθοδο των εσωτερικών διαβαθμίσεων (Internal rating-based approach IRBA) είναι ότι το σταθμισμένο στον κίνδυνο ενεργητικό είναι συνάρτηση των παρακάτω παραμέτρων :

- *PD*: Πιθανότητα Χρεοκοπίας (Probability of default)
- *EAD*: Έκθεση στον Κίνδυνο Χρεοκοπίας (Exposure at default)
- *LGD*: Ζημία Δεδομένης της Χρεοκοπίας (Loss Given Default)
- *M*: Ωρίμανση Υποχρέωσης (Maturity)

Η κάθε ομάδα ενεργητικού έχει διαφορετικό τύπο για τον υπολογισμό του σταθμισμένου στον κίνδυνο ενεργητικού. Για κάθε ομάδα ενεργητικού η επιτροπή της Βασιλείας επιτρέπει τον υπολογισμό με δυο διαφορετικές προσεγγίσεις, την βασική προσέγγιση των εσωτερικών διαβαθμίσεων (foundation internal rating based approach) και την προχωρημένη προσέγγιση (advanced internal rating based approach).

Στην βασική προσέγγιση το κάθε τραπεζικό ίδρυμα εκτιμά με δικά του μέσα την παράμετρο *PD* ενώ όλες οι υπόλοιπες παράμετροι προσμετρούνται με βάση τις οδηγίες της επιτροπής της Βασιλείας II. Αντίθετα στην προχωρημένη

μέθοδο τα τραπεζικά ιδρύματα εκτιμούν όλες τις παραπάνω παραμέτρους με δικά τους μέσα. Παρ' όλες τις διαφορές, με όποια από τις δύο παραπάνω μεθόδους και να γίνει ο υπολογισμός, θα πρέπει οπωσδήποτε να χρησιμοποιηθεί για τον υπολογισμό του σταθμισμένου στον κίνδυνο ενεργητικού ο τύπος που επιβάλλει το σύμφωνο της Βασιλείας II και είναι ο εξής

$$RWA = K \times 12.5 \times EAD$$

όπου η έκθεση στον κίνδυνο χρεοκοπίας (EAD) εκφράζεται σε μία χρηματική μονάδα μέτρησης π.χ. ευρώ.

Από όλες τις ομάδες ενεργητικού που εξετάζει το κείμενο της Βασιλείας II, εμείς θα αναφερθούμε μόνο στην κατηγορία της λιανικής τραπεζικής. Οι υπολογισμοί γίνονται με βάση τρεις διαφορετικές κατηγορίες στοιχείων ενεργητικού λιανικής αλλά υπάρχει ένας τύπος για τον υπολογισμό των κεφαλαίων ο οποίος είναι:

$$K = LGD \times \Phi \left[ \frac{1}{\sqrt{1-R}} \times \Phi^{-1}(PD) + \sqrt{\frac{R}{1-R}} \times \Phi^{-1}(0.999) \right] - PD \times LGD$$

Η ποσότητα  $PD$ , παίρνει τιμές μεταξύ 0 και 1 ενώ το σύμβολο  $\Phi$  παριστάνει την αθροιστική συνάρτηση πιθανότητας της κανονικής κατανομής.

Αυτό που διαφοροποιεί τις διαφορετικές κατηγορίες ενεργητικού λιανικής είναι η μεταβλητή  $R$  που ορίζει ο κανόνας για κάθε μία από τις κατηγορίες: έκθεση σε στεγαστικά δάνεια (Residential mortgage exposures), έκθεση σε ανακυκλούμενα δάνεια λιανικής (Qualifying revolving retail exposures), λοιπές εκθέσεις στο χαρτοφυλάκιο λιανικής (Other retail exposure).

Η συγκεκριμένη μεταβλητή είναι ένας συντελεστής συσχέτισης ο οποίος παίρνει τις ακόλουθες τιμές

- 0.15 όταν ο  $R$  εφαρμόζεται σε υποθέσεις στεγαστικών δανείων που δεν είναι πτωχευμένες και είναι εξασφαλισμένες μερικώς ή ολόκληρες.
- 0.04 όταν ο  $R$  εφαρμόζεται σε υποθέσεις δανείων χωρίς εξασφάλιση σε φυσικά πρόσωπα που δεν ξεπερνούν τις 100.000 € όπως είναι κάποια καταναλωτικά δάνεια.

- Για όλες τις υπόλοιπες απαιτήσεις λιανικής που δεν περιλαμβάνονται στις δυο παραπάνω κατηγορίες ο συντελεστής συσχέτισης  $R$  είναι συνάρτηση της πιθανότητας αθέτησης  $PD$  και δίνεται από τον τύπο

$$R = 0.03 \cdot \left( \frac{1 - e^{-35 \cdot PD}}{1 - e^{-35}} \right) + 0.16 \cdot \left( 1 - \frac{1 - e^{-35 \cdot PD}}{1 - e^{-35}} \right)$$

Από τα παραπάνω μπορεί κανείς να συμπεράνει τη σημαντικότητα των παραμέτρων  $PD$ ,  $LGD$  και  $EAD$  καθώς είναι καθοριστικής σημασίας παράμετροι για τις κεφαλαιακές απαιτήσεις ενός τραπεζικού ιδρύματος και κατά συνέπεια των αποδόσεων των δεικτών κερδοφορίας. Στα πλαίσια αυτής της εργασίας θα ασχοληθούμε με προτάσεις για τη μεθοδολογία του υπολογισμού της πιθανότητας αθέτησης ( $PD$ ) με τη χρήση μοντέλων πιστοληπτικής διαβάθμισης (Credit Scoring Models) όπως προτείνει και το κείμενο της Βασιλείας II.

# ΚΕΦΑΛΑΙΟ 3

## ΜΟΝΤΕΛΑ ΒΑΘΜΟΝΟΜΗΣΗΣ ΠΙΣΤΩΤΙΚΟΥ ΚΙΝΔΥΝΟΥ

### 3.1 Εισαγωγή

Ο όρος “Μοντέλα Βαθμονόμησης Πιστωτικού Κινδύνου” (Credit Scoring Models) αναφέρεται στην ανάπτυξη στατιστικών μοντέλων και τεχνικών για την πρόβλεψη, τη διαχείριση και την ανάλυση του κινδύνου που αναλαμβάνει ο δανειοδότης μέσω της αξιολόγησης του δανειολήπτη. Δεδομένα που αφορούν σε διάφορα χαρακτηριστικά των πελατών, συλλέγονται και αναλύονται με στόχο να βρεθούν εκείνα που ερμηνεύουν και προβλέπουν καλύτερα τον κίνδυνο, έτσι ώστε βάσει αυτών να αποδοθεί βαθμολογία (score) στους πελάτες.

Η βαθμολογία ενός μοντέλου βαθμονόμησης πιστωτικού κινδύνου μπορεί είτε να εκφράζει την πιθανότητα αθέτησης των υποχρεώσεων του πελάτη (probability of default) είτε να σχετίζεται με την κατάταξή του σε κατηγορίες πιστοληπτικής ικανότητας ως προς τον κίνδυνο που αναλαμβάνει ο πιστωτής. Έτσι, απόδοση υψηλής βαθμολογίας (score) σε ένα πιστούχο θα μπορούσε να συνεπάγεται χαμηλό έως και μηδενικό κίνδυνο δανειοδότησης και χαμηλή βαθμολογία (score) υψηλό κίνδυνο (ή και αντίστροφα, ανάλογα με την ανάπτυξη του μοντέλου).

Η διαδικασία υπολογισμού της βαθμολογίας κάθε πελάτη μέσω της χρήσης ενός στατιστικού μοντέλου θα πρέπει να είναι αμερόληπτη και δίκαιη έτσι ώστε να εξασφαλίζεται ότι το αποτέλεσμα που θα προκύπτει θα είναι αντικειμενικό και ακέραιο. Το αποτέλεσμα αυτό λέγεται αλλιώς και σκορόχαρτο (scorecard).

Σε κάθε μοντέλο υπολογίζεται ένα σημείο αποκοπής (cut off score) πάνω από το οποίο κάθε αίτηση για πίστωση γίνεται δεκτή ενώ κάτω από αυτό κάθε αίτηση απορρίπτεται. Όσο πιο μεγάλο είναι το εύρος βαθμολόγησης του μοντέλου, τόσο πιο καλό θεωρείται, λόγω της διαβάθμισης του πιστωτικού κινδύνου βάσει λεπτομερούς αξιολόγησης των πιστούχων, ενώ η αποτελεσματικότητά και η ποιότητά του κρίνεται και από την ακρίβεια του σημείου αποκοπής έτσι ώστε να επιτυγχάνεται το μικρότερο ποσοστό

απόρριψης πελατών με ταυτόχρονη ελαχιστοποίηση των «κακών» πελατών που θα γίνουν αποδεκτά.

Ένα μοντέλο βαθμονόμησης πιστωτικού κινδύνου δε θα πρέπει να είναι στατικό αλλά αναπτυσσόμενο. Θα πρέπει να επαναπροσδιορίζεται και να εξελίσσεται με την εισαγωγή νέων στοιχείων, λαμβάνοντας υπόψη τα νέα δεδομένα που προκύπτουν κατά την πάροδο του χρόνου, με τη συστηματική παρακολούθηση των πελατών.

### **3.2 Κατηγορίες Μοντέλων Βαθμονόμησης Πιστοληπτικής Αξιολόγησης**

Στο παρελθόν, η πιστοληπτική βαθμονόμηση χρησιμοποιούνταν για την αρχική αξιολόγηση ενός υποψήφιου πιστούχου και την κατάταξή του σε δυνητικά καλό ή κακό υποψήφιο, ώστε να ληφθεί απόφαση για το αν θα λάβει χρηματοδότηση. Σήμερα, οι διαδικασίες πιστοληπτικής αξιολόγησης λαμβάνουν χώρα σε όλη τη διάρκεια ζωής ενός χρηματοδοτικού προϊόντος - από το σχεδιασμό και την κατασκευή του προϊόντος, στην τιμολόγησή και τη διάθεσή του, έως την παρακολούθηση της αποπληρωμής του ή τη λήψη αποφάσεων σε περίπτωση μη αποπληρωμής του. Έτσι για παράδειγμα ένα προϊόν που απευθύνεται σε πελάτες υψηλού πιστωτικού κινδύνου θα έχει μεγαλύτερο επιτόκιο από ένα άλλο που απευθύνεται σε πελάτες χαμηλού κινδύνου, ενώ στην περίπτωση ενός ληξιπρόθεσμου δανείου η πρόβλεψη που θα σχηματιστεί για ένα πελάτη υψηλού κινδύνου θα είναι μεγαλύτερη από την αντίστοιχη ενός πελάτη που έχει χαμηλότερο score.

Από τα παραπάνω αντιλαμβανόμαστε ότι μπορούν να αναπτυχθούν πολλά είδη στατιστικών μοντέλων για τη βαθμολόγηση της πιστοληπτικής ικανότητας κατά τα διάφορα στάδια της πελατειακής σχέσης πιστωτή και πιστούχου ώστε να υπάρχει συνεχής παρακολούθηση του κινδύνου που διατρέχει το πιστωτικό ίδρυμα. Στη συνέχεια δίνονται πληροφορίες για τα σημαντικότερα από αυτά.



#### *α. Μοντέλα βαθμολόγησης προώθησης προϊόντων*

Το σκληρά ανταγωνιστικό περιβάλλον στην αγορά της λιανικής πιστοδότησης την έχει μετατρέψει από αγορά δανειστών σε αγορά δανειοληπτών. Σήμερα ο καταναλωτής μπορεί πλέον να επιλέγει από που θα δανειοδοτηθεί. Τα μοντέλα βαθμολόγησης για την προώθηση προϊόντων (marketing scoring models) είναι πλέον απαραίτητα για μια καλά σχεδιασμένη διαδικασία ανάπτυξης χρηματοδοτικών προϊόντων που θα είναι παραγωγικά και για τις δύο πλευρές.

Τα προσφερόμενα χρηματοδοτικά προϊόντα θα πρέπει να στοχεύουν στο κοινό το οποίο είναι ικανό να ανταπεξέλθει στις απαιτήσεις και ταυτόχρονα θα πρέπει να είναι ανταγωνιστικά ώστε να επιλεγθούν από αυτό. Η διαφήμιση και η προώθηση του προϊόντος προς το κατάλληλο κοινό είναι επίσης σημαντική ώστε να επιτευχθεί το μέγιστο επιθυμητό αποτέλεσμα για την κερδοφορία του ιδρύματος.

#### *β. Μοντέλα βαθμολόγησης αιτήσεων*

Ένα ευρύ πεδίο μελέτης στο οποίο βρίσκουν εφαρμογή τα μοντέλα διαβάθμισης είναι αυτό της πιστοληπτικής αξιολόγησης αιτούμενων μονάδων για χορηγήσεις ποσού πίστωσης. Κάθε τραπεζικό ίδρυμα έχει σύστημα αξιολόγησης αιτήσεων νέων δανειοληπτών έτσι ώστε με βάση τη διαθέσιμη πληροφορία τη στιγμή της αξιολόγησης, να κατατάσσονται οι υποψήφιοι δανειολήπτες σε μια κλίμακα, ανάλογα με το score που λαμβάνουν.

Η χορήγηση ή μη ποσού στον αιτούμενο αποφασίζεται με βάση το score του κάθε υποψηφίου και του σημείου αποκοπής που έχει ορίσει από την στρατηγική του ο πιστοδοτικός οργανισμός. Κάποιες φορές καθορίζεται ένα διάστημα γύρω από το σημείο αποκοπής (cut off) και ο υποψήφιος του οποίου το score θα βρεθεί εντός αυτού του διαστήματος παραπέμπεται σε περαιτέρω εξέταση για τη χορήγηση ή μη του ποσού που αιτήθηκε.

Η διαδικασία της πιστοληπτικής αξιολόγησης αιτήσεων (application scoring models) είναι πρωτίστως σημαντική για τον χαρακτηρισμό ενός υποψηφίου σε καλό ή κακό δανειολήπτη. Επίσης στην περίπτωση της θετικής γνωμοδότησης, το μοντέλο προσφέρει τη δυνατότητα ανάλογα με το ύψος του

score, να καθορίζεται το ανώτατο ποσό χορήγησης και να γίνεται τιμολόγηση της χορήγησης προσφέροντας ένα επιτόκιο δανεισμού ανάλογα με το score.

Αξίζει να σημειωθεί ότι στη λιανική τραπεζική η οποία αποτελεί μια αγορά πιστοδοτήσεων με μεγάλο ανταγωνισμό και γρήγορους ρυθμούς, τα αυτοματοποιημένα συστήματα πιστοληπτικής αξιολόγησης που χρησιμοποιούνται σήμερα μπορούν να διεξάγουν την αξιολόγηση σε πολύ σύντομο χρονικό διάστημα.

Ένα επιπλέον πλεονέκτημα χρήσης των μοντέλων βαθμονόμησης αιτήσεων είναι η δυνατότητα χάραξης στρατηγικής, η οποία μέσω των συστημάτων μπορεί να σχεδιαστεί και να ακολουθηθεί πιστά και σταθερά. Στη διαδικασία βαθμολόγησης της πιστοληπτικής ικανότητας πολλές φορές υπεισέρχονται παράγοντες που επιβάλουν περιορισμό στην ελευθερία επιλογής μεταβλητών αξιολόγησης. Για παράδειγμα, στις ΗΠΑ ο συνταγματικός νόμος περί ίσων πιστοδοτικών δικαιωμάτων του 1976 απαγορεύει τη χρήση σε μεθόδους πιστοληπτικής αξιολόγησης χαρακτηριστικών όπως το φύλο, το χρώμα του δέρματος, την οικογενειακή κατάσταση, την εθνικότητα, το αν λαμβάνει κάποιο κοινωνικό επίδομα, τη θρησκεία και άλλα χαρακτηριστικά, η επιλογή των οποίων μπορεί να μειώσει δικαιώματα κοινωνικών ομάδων. Ακόμη και όταν ένα άτομο λάβει αρνητική απάντηση από ένα πιστωτικό ίδρυμα, η απάντηση πρέπει να είναι λεπτομερώς τεκμηριωμένη.

#### *γ. Μοντέλα βαθμολόγησης απάτης*

Η ανίχνευση δόλιων πελατών είναι ένα ακόμα πεδίο εφαρμογής των μοντέλων βαθμολόγησης (fraud scoring models). Ένα τέτοιο μοντέλο υπολογίζει την πιθανότητα ένας υποψήφιος πιστούχος να μην έχει χορηγήσει αληθινές πληροφορίες προς τον πιστωτή.

Ένα παράδειγμα που βρίσκει εφαρμογή στον ασφαλιστικό κλάδο είναι το υψηλό score που συγκεντρώνει η αίτηση για αποζημίωση από έναν πελάτη, όπου μεγάλες τιμές στο score υποδεικνύουν μεγάλη πιθανότητα ο πελάτης να επιθυμεί να λάβει αποζημίωση χωρίς να τη δικαιούται, παρέχοντας στοιχεία που δεν αντικατοπτρίζουν την πραγματικότητα.

#### *δ. Μοντέλα βαθμολόγησης απόδοσης*

Η εκτίμηση της μελλοντικής εκπλήρωσης των συμβατικών υποχρεώσεων των ήδη δανειοδοτημένων πελατών, γίνεται με την εφαρμογή ενός μοντέλου βαθμολόγησης απόδοσης (performance scoring model). Είναι πολύ σημαντικό για ένα τραπεζικό ίδρυμα να γνωρίζει και να παρακολουθεί την ποιότητα, την αξία και τις προοπτικές αποπληρωμών ενός χαρτοφυλακίου δανειοληπτών.

Η κατασκευή ενός τέτοιου μοντέλου είναι πιο απλή σε σχέση με αυτή του μοντέλου βαθμολόγησης αιτήσεων. Αυτό συμβαίνει διότι για έναν υφιστάμενο πελάτη, υπάρχουν αρκετά διαθέσιμα στοιχεία και είναι πιο αξιόπιστα σε σχέση με έναν υποψήφιο πελάτη που τα στοιχεία του δεν είναι διαθέσιμα και ζητούνται από τον ίδιο.

Μια ακόμα διαφορά μεταξύ των δύο μοντέλων είναι η περίοδος πρόβλεψης, η οποία στα μοντέλα βαθμολόγησης απόδοσης κυμαίνεται μεταξύ 12 & 24 μηνών ενώ για τα μοντέλα βαθμολόγησης αιτήσεων είναι έως τη λήξη της δανειακής σύμβασης. Τα αποτελέσματα αυτής της διαδικασίας είναι χρήσιμα για το τραπεζικό ίδρυμα καθώς με βάση αυτά μπορεί να κάνει εκτιμήσεις για προβλέψεις και εποπτικές κεφαλαιακές απαιτήσεις. Επίσης, μπορεί να παρακολουθεί πιο στενά τους πελάτες που έχουν αυξημένη αβεβαιότητα αποπληρωμής και να δρα προτού κάποιος δηλώσει ολική αδυναμία αποπληρωμής. Οι δράσεις της τράπεζας μπορούν να αφορούν αναδιάρθρωση δανείων, μείωση πιστωτικών ορίων και απαιτήσεις για περαιτέρω εξασφαλίσεις.

Το σύμφωνο της Βασιλείας II δίνει μεγάλη έμφαση στην παρακολούθηση της εκτίμησης της αποπληρωμής για την ανίχνευση κινδύνων με ορίζοντα το ένα έτος.

#### *ε. Μοντέλα βαθμολόγησης συμπεριφοράς*

Η κατηγορία αυτών των μοντέλων έχει ως βάση την βαθμονόμηση της συμπεριφοράς των υπαρχόντων πελατών (behavioural scoring models). Η τράπεζα θέλει να γνωρίζει την επίπτωση στον πιστωτικό κίνδυνο από την αλλαγή της συμπεριφοράς των πελατών της, που είναι αποτέλεσμα δικών της ή εξωγενών παραγόντων.

Η κατασκευή του μοντέλου βασίζεται σε πρόσφατες παρατηρήσεις στην συναλλακτική συμπεριφορά του δανειολήπτη με την τράπεζα. Για παράδειγμα, ο αριθμός των πιστωτικών ανοιγμάτων που έχει με την τράπεζα ο πελάτης, οι κινήσεις των λογαριασμών του, η ενημέρωση των προσωπικών του στοιχείων είναι παράγοντες που μπορεί να χρησιμοποιήσει η τράπεζα για να εκτιμήσει τη μελλοντική συμπεριφορά των πελατών της, σε περίπτωση που παραμείνουν με τα υπάρχοντα προϊόντα ή ζητήσουν νέα ή τους προταθούν νέα προϊόντα από την τράπεζα.

#### *στ. Μοντέλα βαθμολόγησης έγκαιρης προειδοποίησης*

Το σύστημα έγκαιρης προειδοποίησης (early warning scoring models) εντοπίζει πιθανά προβλήματα με τους αντισυμβαλλόμενους. Αυτοί που εντοπίζονται ως προβληματικοί αντισυμβαλλόμενοι εισάγονται σε μια λίστα εντατικής παρακολούθησης. Τα μοντέλα που χρησιμοποιούνται σε αυτά τα συστήματα κάνουν χρήση μακροοικονομικών παραγόντων, λογιστικών στοιχείων καθώς και στοιχείων της αγοράς.

Τα μοντέλα που χρησιμοποιούν στοιχεία της αγοράς φαίνεται να έχουν καλύτερη προβλεπτική ικανότητα για βραχυπρόθεσμες προβλέψεις ενώ όσα στηρίζονται σε λογιστικά στοιχεία έχουν καλύτερη προβλεπτική ικανότητα για πιο μακροπρόθεσμες προβλέψεις.

Μοντέλα έγκαιρης προειδοποίησης κρατικών χρεογράφων έχουν ως σκοπό να αντιληφθούν προβλήματα στην ομαλή αποπληρωμή των υποχρεώσεων των κρατών παρά να εντοπίσουν πιστωτικά γεγονότα. Σε αυτήν την κατηγορία μπορούμε να εντάξουμε και τα μοντέλα το οποίο θα περιγράψουμε σε αυτήν την πτυχιακή εργασία, το οποίο έχει ως σκοπό να αντιληφθεί προβλήματα ομαλής αποπληρωμής μιας δανειακής υποχρέωσης και να προτείνει την βέλτιστη ενέργεια για το τραπεζικό ίδρυμα είτε μέσω κάποιας αναδιάρθρωσης είτε μέσω μεγαλύτερης προσπάθειας είσπραξης για κάποιον ληξιπρόθεσμο πελάτη.

#### *ζ. Μοντέλα βαθμολόγησης είσπραξης*

Τα μοντέλα βαθμονόμησης διαχείρισης ληξιπρόθεσμων απαιτήσεων είναι πολύ χρήσιμο εργαλείο για την αποτελεσματικότερη λειτουργία των

μεθόδων είσπραξης. Η λογική αυτών των μοντέλων είναι ότι το πιστωτικό ίδρυμα κατατάσσει τους πελάτες του ανάλογα με την πιθανότητα της δυνατότητας ανταπόκρισης στην ομαλή αποπληρωμή του κεφαλαίου που έχει δανειστεί από την τράπεζα.

Ανάλογα με τα χαρακτηριστικά του κάθε πελάτη, το τραπεζικό ίδρυμα εφαρμόζει κατάλληλη μέθοδο και προσπάθεια είσπραξης, με σκοπό την μεγιστοποίηση της απόδοσης του κόστους διαχείρισης ληξιπρόθεσμων απαιτήσεων. Για παράδειγμα δανειολήπτες με τον ίδιο χρόνο καθυστέρησης αποπληρωμής και ίδια έκθεση σε δανεισμό, δεν σημαίνει ότι πρέπει να αντιμετωπιστούν με τον ίδιο τρόπο από μια υπηρεσία είσπραξης ληξιπρόθεσμών απαιτήσεων. Κατ' ουσία το κόστος είσπραξης θα είναι διαφορετικό με το ίδιο αποτέλεσμα είσπραξης.

Είναι αρκετά τα χαρακτηριστικά τα οποία πρέπει να λάβει υπόψη του το τραπεζικό ίδρυμα ώστε να ορίσει μια στρατηγική είσπραξης. Η πρακτική που ακολουθούνταν να δαπανάται περισσότερο χρόνο και χρήμα για είσπραξη από πελάτες με μεγαλύτερη καθυστέρηση είναι ξεπερασμένη.

Τα μοντέλα βαθμονόμησης ληξιπρόθεσμων πελατών είναι σε στάδιο ανάπτυξης διεθνώς. Πολλά πιστωτικά ιδρύματα πειραματίζονται για να μπορέσουν να καθορίσουν ποιοι ληξιπρόθεσμοι πελάτες έχουν καλές προοπτικές είσπραξης και ποιοι όχι. Οι επιστημονικές δημοσιεύσεις με αυτό το αντικείμενο είναι σχετικά περιορισμένες, και ως τώρα το μεγαλύτερο ενδιαφέρον εντοπίζεται σε μοντέλα διαχείρισης απαιτήσεων επιχειρηματικών πιστοδοτήσεων. Τα τραπεζικά ιδρύματα καλούνται να διαχειριστούν ληξιπρόθεσμες χορηγήσεις ώστε οι απώλειες τους από τα ληξιπρόθεσμα χαρτοφυλάκια να είναι οι λιγότερες το δυνατό. Για να πετύχουν κάτι τέτοιο εφαρμόζουν μια μεγάλη γκάμα στρατηγικών ενεργειών όπως οι ρυθμίσεις δανείων, αναδιαρθρώσεις, προσπάθειες είσπραξης και προσφυγή σε νομικές ενέργειες. Κάθε απόφαση της τράπεζας για κάθε περίπτωση ξεχωριστά έχει κάποιο κόστος και κάποιο αποτέλεσμα. Ο σκοπός αυτών των ενεργειών είναι να μειωθεί η προσδοκώμενη απώλεια είτε μέσω της προσπάθειας μείωσης της πιθανότητας χρεωκοπίας του πελάτη βοηθώντας τον με ένα ευνοϊκότερο πρόγραμμα πληρωμών για κάποιο χρονικό διάστημα, είτε μέσω της ελαχιστοποίησης της απώλειας μέσω μεθόδων είσπραξης ή νομικών ενεργειών.

Πέραν των κλασικών μοντέλων που έχει στην διάθεσή του ένα τραπεζικό ίδρυμα για να μπορεί να αποφασίζει για τις πιστώσεις που χορηγεί στους πελάτες, απαραίτητη είναι και μια μέθοδος ώστε να μπορεί να αποφασίσει ποια από τις διαθέσιμες ενέργειες θα ακολουθήσει από την στιγμή που ένας ήδη πιστοδοτημένος πελάτης, αρχίσει να αντιμετωπίζει πρόβλημα με την ομαλή αποπληρωμή της υποχρέωσής του.

### **3.3 Είδη Μοντέλων Ανάλογα με τον Τύπο Δεδομένων**

Για την ανάπτυξη ενός μοντέλου βαθμολόγησης πιστωτικού κινδύνου, θα πρέπει να συλλεχθούν τα δεδομένα εκείνα που θα μελετηθούν και θα αναλυθούν. Ο τρόπος συλλογής των δεδομένων διαχωρίζει τα διάφορα μοντέλα στις παρακάτω δύο κατηγορίες .

#### *α. Πελατειακά μοντέλα βαθμολόγησης*

Στην κατηγορία αυτή, κάθε εταιρεία που θέλει να αναπτύξει ένα μοντέλο χρησιμοποιεί τα δεδομένα που έχει στις βάσεις της και είτε το αναπτύσσει η ίδια είτε παρέχει την πληροφορία σε μία άλλη εταιρεία η οποία θα κάνει την υλοποίηση. Για παράδειγμα μία τράπεζα, χρησιμοποιεί τη βάση δεδομένων της και τα στοιχεία των πελατών της για την ανάπτυξη ενός μοντέλου, που θα χρησιμοποιήσει στους ίδιους και σε μελλοντικούς πελάτες της. Τα πελατειακά σκορόχαρτα δεν είναι πάντα εφικτά σε περιπτώσεις που η εταιρεία ενδιαφέρεται να επεκταθεί σε ένα νέο τομέα για τον οποίο δεν έχει ιστορικότητα στοιχείων, ενώ άλλες φορές μπορεί η διαδικασία να απαιτεί μεγάλο κόστος και πολύ χρόνο.

#### *β. Γενικά μοντέλα βαθμολόγησης*

Στην κατηγορία αυτή, για την ανάπτυξη ενός μοντέλου τα δεδομένα αντλούνται εκτός εταιρείας, είτε από μια βάση δεδομένων με στοιχεία από διάφορες εταιρείες που μπορεί να είχε σχηματιστεί στο παρελθόν για παρόμοια ή και άλλη έρευνα ή χρήση, είτε από ένα γραφείο πίστης (credit bureau). Πλεονεκτήματα της άντλησης δεδομένων από έναν εξωτερικό προμηθευτή είναι ότι η εταιρεία έχει πρόσβαση στα στοιχεία του πελάτη της και σε άλλες εταιρείες καθώς και η ύπαρξη μεγάλου όγκου και συνεχούς ενημέρωση των

δεδομένων. Για παράδειγμα αν μία τράπεζα ελέγχοντας έναν πελάτη της που αιτήθηκε δανείου, διαπιστώσει ότι κατά το ίδιο χρονικό διάστημα αιτήθηκε ταυτόχρονα δάνειο και σε άλλα ιδρύματα ή ότι έχει ανοίγματα και σε άλλα ιδρύματα θα μπορούσε να είναι σημάδι υψηλού πιστωτικού κινδύνου.

### **3.4 Διαδικασία Ανάπτυξης Μοντέλων Πιστωτικού Κινδύνου**

Η διαδικασία για την ανάπτυξη ενός μοντέλου βαθμολόγησης του πιστωτικού κινδύνου περιλαμβάνει τα εξής τρία στάδια σύμφωνα με τους Liu και Schumann (2002) :

- Προσδιορισμός του προβλήματος
- Ανάπτυξη του μοντέλου
- Εφαρμογή και αξιολόγηση του μοντέλου

Κατά το πρώτο στάδιο καθορίζεται το προς επίλυση πρόβλημα, ο λόγος για τον οποίο θέλουμε να αναπτύξουμε ένα μοντέλο και ποια θα είναι η χρήση του. Ορίζονται εν συνεχεία τα δεδομένα που θα πρέπει να συλλεχθούν, το μέγεθος του δείγματος και το σύνολο των χαρακτηριστικών που θα πρέπει να έχει το κάθε στοιχείο του δείγματος. Εν συνεχεία λαμβάνεται το δείγμα με τέτοιο τρόπο ώστε να είναι τυχαίο, όσο πιο αντιπροσωπευτικό του πληθυσμού και όσο πιο αξιόπιστο γίνεται.

Κατά το δεύτερο στάδιο, γίνεται η ανάλυση των δεδομένων. Τα στοιχεία του δείγματος που συλλέχθηκαν θα διαχωριστούν αρχικά σε ομάδες «καλών», «κακών» και «απροσδιόριστων» πελατών και θα μελετηθούν τα χαρακτηριστικά τους ανά ομάδα. Ο σωστός διαχωρισμός έχει μεγάλη σημασία και θα επηρεάσει σε σημαντικό βαθμό το μοντέλο για αυτό και πρέπει να γίνει προσεκτικά και με σαφήνεια. Από τα αποτελέσματα της ανάλυσης θα προκύψει το ζητούμενο μοντέλο.

Τέλος στο τρίτο στάδιο, το μοντέλο που έχει αναπτυχθεί θα πρέπει να εφαρμοστεί και να ελεγχθεί η αποτελεσματικότητά του, προτού εφαρμοσθεί και στην πράξη.





# ΚΕΦΑΛΑΙΟ 4

## ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΙ ΚΑΤΑΣΚΕΥΗΣ

### ΣΚΟΡΟΧΑΡΤΩΝ

#### 4.1 Λογιστική Παλινδρόμηση

Η Λογιστική Παλινδρόμηση είναι μία τεχνική σχεδιασμένη για την πραγματοποίηση ανάλυσης δεδομένων, που αφορούν τη μελέτη και την πρόβλεψη των τιμών κάποιας κατηγορικής εξαρτημένης μεταβλητής μέσω κάποιων ποσοτικών και ποιοτικών ανεξάρτητων μεταβλητών. Για την κατάταξη των υποψήφιων πελατών σε «καλούς» ή «κακούς» η χρήση της μεθόδου της γραμμικής παλινδρόμησης δεν είναι εφικτή, εφόσον όταν προβλέπουμε τις τιμές μία κατηγορικής εξαρτημένης μεταβλητής, στην ουσία επιδιώκουμε να υπολογίσουμε την πιθανότητα με την οποία η εξαρτημένη μεταβλητή θα λάβει κάποια συγκεκριμένη τιμή. Η τιμή της πιθανότητας αυτής θα πρέπει, εξ ορισμού, να παίρνει τιμές μεταξύ του 0 και του 1 ενώ με τη χρήση της Γραμμικής Πολλαπλής Παλινδρόμησης μπορεί να προκύψουν τιμές μεγαλύτερες του 1 ή και μικρότερες του 0.

Έστω  $Y_i$  η δίτιμη μεταβλητή απόκρισης του  $i$  πελάτη του δείγματος, δηλαδή

$$Y_i = \begin{cases} 1, & \text{εάν ο πελάτης } i \text{ είναι καλός} \\ 0, & \text{διαφορετικά} \end{cases} \text{ για κάθε } i = 1, 2, \dots, n$$

$p_i = P(Y_i = 1) = P(G|x_i)$  η αντίστοιχη πιθανότητα επιτυχίας δηλαδή η πιθανότητα ο πελάτης  $i$  με τα χαρακτηριστικά  $x_p$  να είναι καλός, με τιμές από -1 έως +1. Η συνάρτηση του  $p_i$ ,  $\ln\left(\frac{p_i}{1-p_i}\right)$  ονομάζεται λογάριθμος της σχετικής πιθανότητας (odds) και παίρνει τιμές από  $-\infty$  έως  $+\infty$ . Η συνάρτηση σύνδεσης στο μοντέλο λογιστικής παλινδρόμησης είναι η

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \ln(\text{odds}(G|x_i))$$

δηλαδή το μοντέλο της λογιστικής παλινδρόμησης συνδέει γραμμικά το λογάριθμο της σχετικής πιθανότητας της δίτιμης απόκρισης που ονομάζεται logit με τις επεξηγηματικές μεταβλητές. Έτσι προκύπτει η έκφραση

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p, \quad i = 1, 2, \dots, n \quad (3.1).$$

Αφού η ποσότητα  $\frac{p_i}{1-p_i}$  παίρνει τιμές από 0 έως  $\infty$  η ποσότητα  $\ln\left(\frac{p_i}{1-p_i}\right)$

θα παίρνει τιμές από  $-\infty$  έως  $+\infty$ .

Το αριστερό μέρος της εξίσωσης (3.1) περιέχει τις τιμές της εξαρτημένης μεταβλητής με την μορφή του λογαρίθμου των odd δηλαδή, του λογαρίθμου της σχέσης:  $\text{odds} = \text{prob}/(1-\text{prob})$ . Το odds εναλλακτικά ονομάζεται logit και ο όρος prob εκφράζει την πιθανότητα να συμβεί το γεγονός που έχει ορισθεί ως επιτυχία του πειράματος. Το δεξί μέρος της εξίσωσης περιέχει ένα γραμμικό συνδυασμό των ανεξάρτητων μεταβλητών που συμμετέχουν στο μοντέλο της παλινδρόμησης. Οι συντελεστές των ανεξάρτητων αυτών μεταβλητών στην εξίσωση της παλινδρόμησης, εκτιμώνται με τη μέθοδο της Μέγιστης Πιθανοφάνειας. Βάσει της μεθόδου αυτής η τιμή των συντελεστών των ανεξάρτητων μεταβλητών είναι αυτή που κάνει τις παρατηρηθείς τιμές της εξαρτημένης μεταβλητής πιο πιθανές, βάσει των παρατηρηθεισών τιμών των ανεξάρτητων μεταβλητών.

Τα βήματα κατασκευής του μοντέλου της Λογιστικής Παλινδρόμησης είναι τα εξής

- Προσδιορίζουμε την ποσότητα που μας ενδιαφέρει (εξαρτημένη μεταβλητή) και το σύνολο των ανεξάρτητων μεταβλητών που θα συμμετέχουν στην παλινδρόμηση.
- Διερευνούμε τα δεδομένα για τυχόν ύπαρξη ασυνήθιστων κινήσεων όπως, ακραίες τιμές, ελλείπουσες τιμές κτλ.
- Εκτιμούμε τις παραμέτρους της εξίσωσης της Λογιστικής Παλινδρόμησης.
- Ελέγχουμε την ικανοποίηση των υποθέσεων για τη σωστή εφαρμογή της Λογιστικής Παλινδρόμησης.
- Μελετάμε την επίδραση κάθε ανεξάρτητης μεταβλητής στο μοντέλο.

Η αξιολόγηση της συνολικής προσαρμογής ενός μοντέλου στα δειγματικά δεδομένα γίνεται με το λόγο των τιμών των συναρτήσεων πιθανοφάνειας (likelihood ratio test) του μοντέλου που περιλαμβάνει μόνο τη σταθερά  $b_0$  ( $L_0$ ) προς το πλήρες μοντέλο ( $L_F$ ). Πιο συγκεκριμένα υπολογίζεται η ποσότητα  $-2 \ln\left(\frac{L_0}{L_F}\right)$  η οποία ακολουθεί την κατανομή  $\chi^2$  με βαθμούς ελευθερίας όσες είναι και οι μεταβλητές. Η μηδενική υπόθεση που ελέγχεται

είναι ότι όλοι οι συντελεστές είναι μηδέν με εναλλακτική τουλάχιστον ένας συντελεστής να είναι διάφορος του μηδέν.

Για την αξιολόγηση της προσαρμογής του λογιστικού μοντέλου, χρησιμοποιείται επιπλέον ο συντελεστής προσδιορισμού  $R^2$  των Cox και Snell ο οποίος ισούται με

$$R^2 = 1 - \left[ \frac{L_0}{L_f} \right]^{2/n}.$$

Το πρόβλημα με τον συγκεκριμένο συντελεστή είναι ότι ποτέ δεν καταλήγει να πάρει μέγιστη τιμή το 1. Έτσι χρησιμοποιείται εναλλακτικά και ο συντελεστής Nagelkerke ο οποίος υπολογίζεται από τον τύπο

$$\tilde{R}^2 = \frac{R^2}{R_{max}^2} \in (0,1)$$

όπου

$$R^2 = 1 - (L_0)^{\frac{2}{n}}.$$

Στο σημείο αυτό θα πρέπει να αναφερθεί ότι η Λογιστική Παλινδρόμηση, για τη σωστή εφαρμογή της απαιτεί μεγάλο δείγμα, προκειμένου να παράγει αξιόπιστο αποτέλεσμα. Ένας εμπειρικός κανόνας αναφέρει ότι το δείγμα θα πρέπει να είναι 30 φορές μεγαλύτερο από τον αριθμό των παραμέτρων που εκτιμά το μοντέλο. Επιπλέον, σε περίπτωση που ενδιαφερόμαστε να χρησιμοποιήσουμε το μοντέλο για πρόβλεψη, θα πρέπει να αξιολογήσουμε την αποτελεσματικότητά του. Αυτό σημαίνει ότι δημιουργούμε την εξίσωση βασιζόμενοι σε ένα μέρος των δεδομένων και σε επόμενο βήμα ελέγχουμε την αποτελεσματικότητά της, χρησιμοποιώντας στο υπόλοιπο δείγμα. Στο τελικό μοντέλο που προκύπτει υπολογίζουμε το ποσοστό των πελατών που προβλέφθηκαν καλοί και ήταν καλοί, το ποσοστό των πελατών που προβλέφθηκαν ως κακοί και ήταν κακοί αλλά και το συνολικό ποσοστό ακρίβειας του μοντέλου.

Μετά τον προσδιορισμό των μεταβλητών που θα συμμετέχουν, τελικά, στη διαδικασία κατασκευής του μοντέλου, θα πρέπει να ελέγξουμε την ικανοποίηση των υποθέσεων εφαρμογής του συγκεκριμένου είδους παλινδρόμησης. Υπάρχουν κάποιες υποθέσεις που πρέπει να διερευνηθούν πριν την κατασκευή του μοντέλου και κάποιες που θα πρέπει να διερευνηθούν μετά

τη δημιουργία της εξίσωσης. Πριν τη διαδικασία δημιουργίας του μοντέλου της Λογιστικής παλινδρόμησης θα πρέπει να ελέγχουμε ότι:

- Οι ανεξάρτητες μεταβλητές είναι ή ποσοτικές ή ποιοτικές ( οι οποίες σε ένα επόμενο στάδιο θα μετατραπούν σε ψευδομεταβλητές ).
- Οι ανεξάρτητες μεταβλητές θα πρέπει να συσχετίζονται με την εξαρτημένη μεταβλητή.

Μετά τη διαδικασία δημιουργίας του μοντέλου της παλινδρόμησης θα πρέπει να ελέγξουμε ότι:

- Δεν υπάρχει αυτοσυσχέτιση.
- Δεν υπάρχει συσχέτιση των καταλοίπων με τις ανεξάρτητες μεταβλητές.
- Δεν υπάρχει ένδειξη συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών.

Στην πραγματικότητα η Λογιστική Παλινδρόμηση ακολουθεί τη λογική της Γραμμικής Παλινδρόμησης. Για να επιτευχθούν τα βέλτιστα αποτελέσματα, θα πρέπει οι ανεξάρτητες μεταβλητές να συνδέονται ισχυρά με την εξαρτημένη μεταβλητή και οι ανεξάρτητες μεταβλητές να είναι ασυσχέτιστες μεταξύ τους. Επίσης, θα πρέπει, τα κατάλοιπα να μην κρύβουν κάποιο ανερμήνευτο συστηματικό παράγοντα ο οποίος δεν περιγράφεται από το μοντέλο. Στο σημείο αυτό θα πρέπει να επισημανθεί ότι, λόγω του σχεδιασμού της τεχνικής, δεν είναι δυνατόν να ικανοποιείται η υπόθεση της κανονικότητας των καταλοίπων. Στη Λογιστική Παλινδρόμηση τα κατάλοιπα ακολουθούν την διωνυμική κατανομή, η οποία σε μεγάλα δείγματα τείνει να προσομοιάζει ως προς τη μορφή με την κανονική κατανομή. Επίσης, στην τεχνική αυτή, δεν είναι δυνατόν να ισχύει η υπόθεση της ισότητας της διακύμανσης, λόγω της συναρτησιακής σχέσης της μέσης τιμής με τη διακύμανση. Πριν την εφαρμογή της τεχνικής θα πρέπει να μελετηθεί και η ένταση της συσχέτισης της εξαρτημένης μεταβλητής με κάθε μία ανεξάρτητη. Η μελέτη αυτή θα πρέπει να πραγματοποιηθεί μέσω του ελέγχου  $\chi^2$

Όταν κατασκευάζεται ένα μοντέλο λογιστικής παλινδρόμησης, είναι πολύ σημαντικό να εντοπιστεί το βέλτιστο υποσύνολο εξηγηματικών μεταβλητών χρησιμοποιώντας κάποια βηματική διαδικασία. Έτσι έχουμε να επιλέξουμε μεταξύ Forward και Backward αλγορίθμων εισαγωγής των

ανεξάρτητων μεταβλητών στο μοντέλο. Οι αλγόριθμοι Forward εισάγουν σε διαδοχικά βήμα μία ανεξάρτητη μεταβλητή στην εξίσωση, ξεκινώντας από εκείνη τη μεταβλητή που συνδέεται ισχυρότερα με το εξαρτημένο μέγεθος. Οι αλγόριθμοι Backward εισάγουν όλες τις προτεινόμενες ανεξάρτητες μεταβλητές στο μοντέλο και εξάγουν σε διαδοχικά βήματα τις μεταβλητές που δεν συνδέονται στατιστικά σημαντικά με την εξαρτημένη μεταβλητή. Τέλος η μέθοδος Enter παράγει την εξίσωση της παλινδρόμησης βάση όλων των προτεινόμενων από τον ερευνητή μεταβλητών. Στους αλγόριθμους Forward και στους Backward υπάρχουν τρεις εναλλακτικές επιλογές κάθε μία από τις οποίες χρησιμοποιεί διαφορετικό μηχανισμό για την έλεγχο της στατιστικής σημαντικότητας των συντελεστών των ανεξάρτητων μεταβλητών που συμμετέχουν στην παλινδρόμηση. Στα πλαίσια της παρούσας εργασίας, θα χρησιμοποιήσουμε το κριτήριο του Wald το οποίο εξετάζει την μηδενική υπόθεση  $H_0 : b_i = 0$  έναντι της  $H_1 : b_i \neq 0$  για συγκεκριμένο  $i=1,2,\dots,n$ .

## 4.2 Διαχωριστική Ανάλυση

Η διαχωριστική ανάλυση είναι μια στατιστική τεχνική που στόχο έχει να χωρίσει ένα πληθυσμό σε υποσύνολα (υποομάδες) ώστε να καταταχθούν σωστά όσο το δυνατό περισσότερες παρατηρήσεις. Έτσι για παράδειγμα στη χρηματοοικονομική επιστήμη, αναζητούμε τους κανόνες αυτούς με τους οποίους οι πελάτες διαχωρίζονται σε «καλούς» και «κακούς» ώστε να μπορεί στη συνέχεια να καταταχθεί ένας καινούριος πελάτης και να αποφασιστεί εάν θα του χορηγηθεί ή όχι κάποιο προϊόν.

Έστω λοιπόν ότι θέλουμε να βρούμε ένα κανόνα σύμφωνα με τον οποίο ο δειγματικός χώρος  $A$ , όλων πιθανών συνδυασμών των υπό εξέταση χαρακτηριστικών των πελατών, θα χωριστεί σε δύο περιοχές  $A_G$  και  $A_B$  για τους «καλούς» και τους «κακούς» πελάτες αντίστοιχα. Όπως σε κάθε μοντέλο πρόβλεψης έτσι και στη μέθοδο διαχωρισμού υπάρχει η πιθανότητα λάθους στην πρόβλεψη η οποία πρέπει να ελαχιστοποιηθεί. Ο πρώτος τύπος λάθους είναι ενώ ένας πελάτης είναι καλός, να χαρακτηριστεί ως κακός και να χάσουμε το κέρδος που θα είχαμε από αυτόν. Υποθέτουμε ότι το κέρδος είναι ίδιο για

όλους τους πελάτες και το συμβολίζουμε με  $L$ . Η αναμενόμενη απώλεια στην περίπτωση αυτού του τύπου λάθους θα είναι  $L \times P(\mathbf{x}|G)p_G$ . Ο δεύτερος τύπος λάθους είναι να κατατάξουμε έναν πελάτη στην υποομάδα με τους καλούς πελάτες ενώ είναι κακός και να επωμιστούμε το κόστος. Υποθέτουμε ότι το κόστος είναι ίδιο για όλους τους πελάτες και το συμβολίζουμε με  $D$ . Στην περίπτωση αυτή το κόστος θα είναι  $D \times P(\mathbf{x}|B)p_B$ . Επομένως για να ελαχιστοποιηθεί το αναμενόμενο κόστος, ο πελάτης που έχει τις ιδιότητες  $\mathbf{x}$  θα ταξινομείται στο σύνολο  $A_G$  εάν ισχύει η σχέση

$$D \times P(\mathbf{x}|B)p_B \leq L \times P(\mathbf{x}|G)p_G$$

Κατά συνέπεια θα ικανοποιούνται οι αιτήσεις των πελατών των οποίων τα χαρακτηριστικά ανήκουν στο σύνολο

$$A_G = \{\mathbf{x} | D \times P(\mathbf{x}|B)p_B \leq L \times P(\mathbf{x}|G)p_G\} = \left\{ \mathbf{x} \mid \frac{D}{L} \leq \frac{P(\mathbf{x}|G)p_G}{P(\mathbf{x}|B)p_B} \right\}.$$

Στην περίπτωση που τα χαρακτηριστικά δεν προέρχονται από διακριτές τυχαίες μεταβλητές αλλά από συνεχείς, τότε οι δεσμευμένες συναρτήσεις πιθανότητας αντικαθίστανται από τις δεσμευμένες συναρτήσεις πυκνότητας και η σχέση γράφεται

$$A_G = \{\mathbf{x} | Df(\mathbf{x}|B)p_B \leq Lf(\mathbf{x}|G)p_G\} = \left\{ \mathbf{x} \mid \frac{Dp_B}{Lp_G} \leq \frac{f(\mathbf{x}|G)}{f(\mathbf{x}|B)} \right\}$$

Από τις διάφορες συναρτήσεις πυκνότητας πιθανότητας που έχουν ερευνηθεί, το μεγαλύτερο ενδιαφέρον έχει εστιασθεί στην πολυδιάστατη κανονική κατανομή. Και αυτό, γιατί οι διαδικασίες ταξινόμησης, που είναι βασισμένες στους κανονικούς πληθυσμούς, υπερισχύουν στην πράξη λόγω της απλότητας και της υψηλής αποδοτικότητάς τους.

Στη συνέχεια παρουσιάζονται οι προϋποθέσεις για την κατάταξη ενός πελάτη με ιδιότητες  $\mathbf{x}$ , δεδομένων τυχαίων δειγμάτων «καλών» και «κακών» πελατών που ακολουθούν την κανονική κατανομή με κοινή και μη κοινή διακύμανση.

Αν  $\mathbf{x}_{G1}, \mathbf{x}_{G2}, \dots, \mathbf{x}_{Gn_G}$  είναι ένα τυχαίο δείγμα  $n_G$  «καλών» πελατών από την κατανομή  $N_p(\mu_G, \Sigma)$  και  $\mathbf{x}_{B1}, \mathbf{x}_{B2}, \dots, \mathbf{x}_{Bn_B}$  ένα τυχαίο δείγμα  $n_B$  «κακών» πελατών από την κατανομή  $N_p(\mu_B, \Sigma)$ , τότε ο πελάτης με τις ιδιότητες  $\mathbf{x}$  ταξινομείται στους «καλούς» αν ισχύει

$$xS_p^{-1}(m_G - m_B)^T \geq \ln\left(\frac{Dp_B}{Lp_G}\right) + \frac{m_G S_p^{-1} m_G^T - m_B S_p^{-1} m_B^T}{2}$$

και στους «κακούς» αν ισχύει

$$xS_p^{-1}(m_G - m_B)^T < \ln\left(\frac{Dp_B}{Lp_G}\right) + \frac{m_G S_p^{-1} m_G^T - m_B S_p^{-1} m_B^T}{2}$$

Για την εκτίμηση των παραμέτρων  $\mu_G$  και  $\mu_B$  χρησιμοποιούμε τους δειγματικούς μέσους

$$\mu_G = m_G = (m_{G1}, m_{G2}, \dots, m_{Gp})$$

όπου

$$m_{Gj} = \frac{1}{n_G} \sum_{i=1}^{n_G} x_{Gi}$$

και

$$\mu_B = m_B = (m_{B1}, m_{B2}, \dots, m_{Bp})$$

όπου

$$m_{Bj} = \frac{1}{n_B} \sum_{i=1}^{n_B} x_{Bi}$$

για κάθε  $j=1, 2, \dots, p$

Για την εκτίμηση του πίνακα συνδιακύμανσης χρησιμοποιούμε το δειγματικό πίνακα συνδιακύμανσης, δηλαδή

$$\Sigma = S_p = \frac{1}{n_G + n_B - 2} \left[ \sum_{i=1}^{n_G} (x_{Gi} - m_G)^T (x_{Gi} - m_G) + \sum_{i=1}^{n_B} (x_{Bi} - m_B)^T (x_{Bi} - m_B) \right]$$

Στην περίπτωση μη ίσων διακυμάνσεων, δηλαδή αν  $x_{G1}, x_{G2}, \dots, x_{Gn_G}$  είναι ένα τυχαίο δείγμα  $n_G$  «καλών» πελατών από την κατανομή  $N_p(\mu_G, \Sigma_G)$  και  $x_{B1}, x_{B2}, \dots, x_{Bn_B}$  ένα τυχαίο δείγμα  $n_B$  «κακών» πελατών από την κατανομή  $N_p(\mu_B, \Sigma_B)$ , τότε ο πελάτης με τις ιδιότητες  $\mathbf{x}$  ταξινομείται στους «καλούς» αν ισχύει

$$x(S_B^{-1} - S_G^{-1})x^T + 2x(S_G^{-1}m_G - S_B^{-1}m_B) \geq 2\ln\left(\frac{Dp_B}{Lp_G}\right) - 2\ln\left(\frac{\sqrt{|S_B|}}{\sqrt{|S_G|}}\right) + m_G S_G^{-1} m_G^T - m_B S_B^{-1} m_B^T$$

και στους «κακούς» αν ισχύει

$$x(S_B^{-1} - S_G^{-1})x^T + 2x(S_G^{-1}m_G - S_B^{-1}m_B) < 2\ln\left(\frac{Dp_B}{Lp_G}\right) - 2\ln\left(\frac{\sqrt{|S_B|}}{\sqrt{|S_G|}}\right) + m_G S_G^{-1} m_G^T - m_B S_B^{-1} m_B^T$$

Για την εκτίμηση των πινάκων συνδιακύμανσης  $\Sigma_G$  και  $\Sigma_B$  παίρνουμε το δειγματικό πίνακα συνδιακύμανσης, δηλαδή

$$\Sigma_G = S_G = \frac{1}{n_G - 1} \sum_{i=1}^{n_G} (\mathbf{x}_{Gi} - m_G)^T (\mathbf{x}_{Gi} - m_G)$$

και

$$\Sigma_B = S_B = \frac{1}{n_B - 1} \sum_{i=1}^{n_B} (\mathbf{x}_{Bi} - m_B)^T (\mathbf{x}_{Bi} - m_B)$$

Σε περίπτωση που δεν μπορούμε να υποθέσουμε κανονικότητα της κατανομής, τότε μπορεί να γίνει χρήση του διαχωριστικού κανόνα του Fisher (Fisher R.A. (1936), *The use of multiple measurements in taxonomic problems*, Annals of Eugenics, 7,179-188) ο οποίος σκέφτηκε να μετασχηματίσει τις πολυμεταβλητές παρατηρήσεις  $\mathbf{x}$  σε μονομεταβλητές παρατηρήσεις  $\mathbf{y}$  μέσω μιας συνάρτησης, η οποία λέγεται διαχωριστική συνάρτηση (Βασιλάκη Γ. Μαρία (2010), Διπλωματική Εργασία στα Στατιστικά Μοντέλα Βαθμολόγησης Πιστοληπτικής Ικανότητας). Έστω

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_p X_p = \mathbf{b} \mathbf{X}^T$$

ένας γραμμικός συνδυασμός των χαρακτηριστικών  $\mathbf{X} = (x_1, x_2, \dots, x_p)$  με αντίστοιχο διάνυσμα συντελεστών  $\mathbf{b}$ . Ο Fisher πρότεινε ως μέτρο διαχωρισμού  $M$  των ομάδων, την απόσταση των δειγματικών μέσων των ομάδων των «καλών» και των «κακών» πελατών προς την τετραγωνική ρίζα της δειγματικής διασποράς, υποθέτοντας πως οι δύο ομάδες έχουν κοινή διασπορά. Συμβολίζοντας με  $m_G$  και  $m_B$  τους δειγματικούς μέσους των ομάδων και με  $S_p$  την κοινή δειγματική διασπορά προκύπτει

$$M = \mathbf{b} \frac{(\mathbf{m}_G - \mathbf{m}_B)^T}{(b S_p b^T)^{\frac{1}{2}}}$$

Σκοπός της μεθόδου είναι να μεγιστοποιηθεί η ποσότητα  $M$  ώστε οι δύο ομάδες να είναι όσο γίνεται πιο διαφορετικές μεταξύ τους. Παραγωγίζουμε την ποσότητα  $M$  ως προς  $\mathbf{b}$ , εξισώνουμε την παράγωγο με μηδέν και στη συνέχεια υπολογίζουμε την μέγιστη τιμή της ποσότητας  $M^2$  καταλήγοντας στην παρακάτω σχέση



$$M^2 = \frac{[\mathbf{b}(\mathbf{m}_G - \mathbf{m}_B)^T]^2}{\mathbf{b}\mathcal{S}_p\mathbf{b}^T} \leq \mathcal{S}_p^{-1}(\mathbf{m}_G - \mathbf{m}_B)^T \mathcal{S}_p^{-1}(\mathbf{m}_G - \mathbf{m}_B)$$

Ο κανόνας διαχωρισμού του Fisher υπερτερεί στο ότι δε μας ενδιαφέρει τι κατανομή ακολουθεί το δείγμα μας. Το αποτέλεσμα ισχύει για όλες τις κατανομές αφού το μέτρο  $M$  της απόστασης περιλαμβάνει μόνο το μέσο όρο και τη διακύμανση των κατανομών.

### 4.3 Μέθοδος του κοντινότερου Γείτονα

Η μέθοδος του κοντινότερου γείτονα είναι μια γενική μέθοδος με εφαρμογές στην κατασκευή μοντέλων πρόβλεψης νέων τιμών που μπορεί να χρησιμοποιηθούν και για την κατάταξη παρατηρήσεων. Τόσο οι ανεξάρτητες όσο και οι εξαρτημένες μεταβλητές μπορούν να είναι είτε συνεχείς είτε κατηγορικές.

Έστω ότι έχουμε ένα δείγμα με παρατηρήσεις και θέλουμε να προβλέψουμε την τιμή της μεταβλητής  $y$  για μία νέα παρατήρηση  $\mathbf{x}$  χρησιμοποιώντας για την πρόβλεψη αυτή πληροφορίες από το ήδη υπάρχον δείγμα. Οι τιμές που θα χρησιμοποιήσουμε θα είναι από παρατηρήσεις που μοιάζουν περισσότερο με τη νέα παρατήρηση για την οποία θέλουμε να κάνουμε πρόβλεψη. Η εκτίμηση αυτή δίνεται ως

$$\hat{y} = \frac{1}{K} \sum_{x_i \in N_k(\mathbf{x})} y_i$$

όπου  $N_k(\mathbf{x})$  είναι το σύνολο που περιέχει τις  $k$  πιο κοντινές παρατηρήσεις στο  $x_i$  για το οποίο θέλουμε να κάνουμε πρόβλεψη. Δηλαδή βρίσκουμε ποιες παρατηρήσεις είναι πιο κοντά στην τιμή που θέλουμε να κάνουμε πρόβλεψη και προβλέπουμε παίρνοντας το μέσο όρο των τιμών των κοντινότερων παρατηρήσεων.

Οι βασικές παράμετροι της μεθόδου είναι η τιμή του  $k$ , δηλαδή του αριθμού των πιο κοντινών με το  $\mathbf{x}$  παρατηρήσεων και η απόσταση την οποία χρησιμοποιούμε για να βρούμε τις κοντινότερες παρατηρήσεις. Και για τα δύο αυτά χαρακτηριστικά τίθεται το θέμα της επιλογής τους. Σε κάθε πρόβλημα μικρή τιμή για το  $k$  οδηγεί σε πολύ μεγάλη διακύμανση όσον αφορά τις

προβλέψεις. Αντίθετα αν δώσουμε στο  $k$  μεγάλη τιμή τότε οδηγούμαστε σε ένα μοντέλο με μεγάλη μεροληψία. Από τα παραπάνω προκύπτει πως το  $k$  θα πρέπει να είναι αρκετά μεγάλο ώστε να ελαχιστοποιήσει την πιθανότητα λάθους κατάταξης αλλά και αρκετά μικρό (με σεβασμό πάντα στον αριθμό των παρατηρήσεων που περιλαμβάνει το δείγμα) ώστε οι  $k$  κοντινές παρατηρήσεις να είναι αρκετά κοντά στο άγνωστο σημείο. Έτσι λοιπόν υπάρχει μια βέλτιστη τιμή για το  $k$  η οποία καταφέρνει να φέρει την ισορροπία μεταξύ μεροληψίας και διακύμανσης στο μοντέλο. Αναζητούμε επομένως το βέλτιστο  $k$ , χρησιμοποιώντας το κατάλληλο μέτρο απόστασης με βάση τη φύση του προβλήματος, ενώ δίνεται σε όλες τις κοντινές παρατηρήσεις το ίδιο βάρος.

Μια παραλλαγή της μεθόδου είναι να σταθμίσουμε την απόσταση, δηλαδή πιο όμοιες παρατηρήσεις με μικρή απόσταση να λαμβάνονται περισσότερο υπόψη. Στις περιπτώσεις που χρησιμοποιούμε τη μέθοδο του κοντινότερου γείτονα για να κατατάξουμε κάποιες παρατηρήσεις, δεν μπορούμε να χρησιμοποιήσουμε το μέσο όρο. Συνήθως αυτό που χρησιμοποιούμε ως πρόβλεψη είναι η τιμή με τη μεγαλύτερη συχνότητα στις κοντινότερες παρατηρήσεις. Δηλαδή βρίσκουμε σε ποιες ομάδες ανήκουν οι κοντινότερες παρατηρήσεις και κατατάσσουμε την παρατήρησή μας στην ομάδα στην οποία ανήκουν οι περισσότερες. Εναλλακτικά καθώς έχουμε συχνότητες για κάθε ομάδα, μπορούμε να κατατάξουμε με πιθανότητα ανάλογη της σχετικής συχνότητας που βρήκαμε από τις κοντινότερες παρατηρήσεις.

Έστω για παράδειγμα ότι έχουμε ένα δείγμα 10 ανθρώπων τους οποίους θέλουμε να κατατάξουμε ως προς το γένος (άνδρας ή γυναίκα) έχοντας ως δεδομένο το ύψος σε εκατοστά. Οι τιμές του δείγματος είναι οι εξής : (Θ,163), (Θ,171), (Θ,158), (Θ,175), (Θ,169), (Α,172), (Α,174), (Α,169), (Α,185), (Α,190). Θέλοντας να κατατάξουμε άτομο ύψους 170 εκατοστών σύμφωνα με τη μέθοδο του κοντινότερου γείτονα για  $k=3$ , τότε θα αποφανθούμε ότι το άτομο είναι Θήλυ δεδομένου ότι οι τρεις κοντινότερες παρατηρήσεις είναι οι (Θ,171), (Θ,169), (Α,169). Εάν στο ίδιο παράδειγμα θελήσουμε να κατατάξουμε το άτομο για  $k=5$ , τότε θα το κατατάξουμε στους άνδρες δεδομένου ότι οι κοντινότερες παρατηρήσεις θα είναι οι (Θ,171), (Θ,169), (Α,169), (Α,172), (Α,174).

Οι κανόνες κατάταξης δεν έχουν κάποια συγκεκριμένη συναρτησιακή μορφή (για παράδειγμα γραμμική, τετραγωνική κλπ.). Τέλος η μέθοδος έχει πολλές ομοιότητες με τη μέθοδο της μη παραμετρικής παλινδρόμησης με τη χρήση των Kernels και απαιτεί σημαντικό υπολογιστικό φόρτο, για αυτό και έχουν σχεδιαστεί αλγόριθμοι που εκτελούν τους υπολογισμούς με έξυπνο τρόπο και κατά συνέπεια γρηγορότερα.

Όπως έχουμε προαναφέρει, δοθέντος ενός αγνώστου σημείου η μέθοδος των  $k$ -κοντινότερων γειτόνων βασίζεται στις  $k$  πιο κοντινές παρατηρήσεις. Για αυτό το λόγο για να κάνουμε πρόβλεψη θα πρέπει καθορίσουμε ένα μέτρο που να υπολογίζει την απόσταση ανάμεσα στο άγνωστο σημείο και στις παρατηρήσεις του δείγματος. Το πιο σύνηθες μέτρο που χρησιμοποιούμε για να μετρήσουμε αυτή την απόσταση είναι η ευκλείδεια απόσταση. Υπάρχουν όμως και άλλα μέτρα για να μετρήσουμε αυτή την απόσταση όπως το τετράγωνο της ευκλείδειας απόστασης, City block (Manhattan) distance και Chebychev distance.

Η ευκλείδεια απόσταση αποτελεί την πιο απλή και την πιο γνωστή απόσταση και δίνεται από τον τύπο:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{r=1}^p (x_{ir} - x_{jr})^2}$$

Στην περίπτωση που το σύνολο δεδομένων περιλαμβάνει και διακριτές και συνεχείς μεταβλητές χρησιμοποιείται ο τύπος

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{r=1}^p f(x_{ir}, x_{jr})}$$

όπου

$$f(x_i, x_j) = (x_i - x_j)^2$$

για τις συνεχείς μεταβλητές και

$$f(x_i, x_j) = \begin{cases} 0, & x_i = x_j \\ 1, & x_i \neq x_j \end{cases}$$

για τις διακριτές μεταβλητές.

Μερικές ιδιότητες είναι πως εξαρτάται από την κλίμακα μέτρησης κι επομένως αλλάζοντας την κλίμακα μπορούμε να πάρουμε ολότελα

διαφορετικές αποστάσεις. Επίσης μεταβλητές με μεγάλες απόλυτες τιμές έχουν πολύ μεγαλύτερο βάρος και σχεδόν καθορίζουν την απόσταση ανάμεσα σε παρατηρήσεις. Η ερμηνεία της απόστασης είναι πολύ εύκολο να αποδοθεί γεωμετρικά. Στην πραγματικότητα η απόσταση αγνοεί τις στατιστικές ιδιότητες των παρατηρήσεων όπως για παράδειγμα τη μεταβλητότητα κάθε μεταβλητής. Δεδομένου ότι παίρνουμε τετραγωνικές αποκλίσεις outliers έχουν μεγάλη επίδραση στον υπολογισμό της απόστασης.

Όλες οι παραπάνω αποστάσεις έχουν το μειονέκτημά ότι δεν λαμβάνουν υπόψη τους τις όποιες διαφορές στην κλίμακα των μεταβλητών όπως επίσης και τις διαφορές στις διακυμάνσεις τους. Επίσης τυχόν συσχετίσεις ανάμεσα στις μεταβλητές δεν λαμβάνονται υπόψη και έτσι κατά κάποιον τρόπο αν υπάρχουν συσχετισμένες μεταβλητές η απόσταση ανάμεσα σε δύο παρατηρήσεις μπορεί να είναι πλασματική.

Ένας τρόπος για να φέρουμε όλες τις μεταβλητές σε συγκρίσιμη κλίμακα είναι να διαιρέσουμε κάθε μία μεταβλητή με την τυπική της απόκλιση. Έτσι, αν συμβολίσουμε με  $s_r$  τη διακύμανση της  $r$  μεταβλητής προκύπτει ο παρακάτω τύπος γνωστός ως απόσταση του Pearson :

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^p \frac{(x_{ir} - x_{jr})^2}{s_r^2}}$$

όπου

$$s_r = \frac{1}{n-1} \sum_{i=1}^n (x_{ir} - \bar{x}_r)$$

και

$$\bar{x}_r = \frac{1}{n} \sum_{i=1}^n x_{ir}.$$

Η μέθοδος των  $k$ -κοντινότερων γειτονιών έχει βρει αρκετές εφαρμογές τα τελευταία χρόνια λόγω της μεγάλης ανάπτυξης των υπολογιστών, ενώ στην πράξη τα αποτελέσματα στα οποία καταλήγει η μέθοδος είναι αρκετά ικανοποιητικά.

## ΚΕΦΑΛΑΙΟ 5

### ΕΦΑΡΜΟΓΗ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΣΕ ΔΕΔΟΜΕΝΑ ΚΑΘΥΣΤΕΡΗΣΕΩΝ

#### 5.1 Γενικά Χαρακτηριστικά Δειγμάτων

Στα πλαίσια της παρούσας εργασίας θα αναλύσουμε δεδομένα που προέρχονται από δύο διαφορετικά δείγματα και με τη χρήση της λογιστικής παλινδρόμησης θα κατασκευάσουμε δύο μοντέλα με στόχο την πρόβλεψη και την ορθή κατάταξη ενός πελάτη σε «καλό» ή «κακό».

Το πρώτο δείγμα αποτελείται από 12949 πιστωτές (όχι απαραίτητα πελάτες της τράπεζας, χωρίς καταθετικά ή άλλα προϊόντα πέραν της χορήγησης πίστωσης) μίας ελληνικής συστημικής τράπεζας. Από αυτούς 9210 έχουν αξιολογηθεί ως «καλοί» και οι υπόλοιποι ως «κακοί» βάση του δείκτη 90 ημερών. Εάν δηλαδή ένας πελάτης έχει καθυστερήσει να πληρώσει τη δόση από 90 μέρες και πάνω θεωρείται «κακός». Οι μεταβλητές που έχουν καταγραφεί από το συγκεκριμένο δείγμα είναι οι ακόλουθες :

- ΚΩΔΙΚΟΣΠΕΛΑΤΗ: Είναι μία μεταβλητή που περιέχει τους κωδικούς του κάθε πελάτη όπως αυτοί έχουν αποδοθεί από την τράπεζα από την οποία άρθηκε το δείγμα
- ΑΑ: Είναι ο αύξων αριθμός κάθε εγγραφής/στοιχείου του δείγματος από το 1 έως το 12949
- ΗΜΕΡΟΜΗΝΙΑΓΕΝΝΗΣΗΣ: Η ημερομηνία γέννησης του κάθε πελάτη
- ΕΤΟΣΓΕΝΝΗΣΗΣ: Το έτος γέννησης του κάθε πελάτη
- ΗΛΙΚΙΑ: Η ηλικία του κάθε πελάτη το 2015, η οποία υπολογίστηκε με βάση την ημερομηνία γέννησης
- ΚΑΤΟΧΟΣΠΙΣΤΚΑΡΤΑΣ: Κατηγορική μεταβλητή με τιμή
  - 1 αν ο πελάτης έχει πιστωτική κάρτα
  - 2 αν δεν έχει
- ΙΔΙΟΚΤΗΣΙΑ: Κατηγορική μεταβλητή με τιμές
  - 1 εάν είναι ιδιοκτήτης

- 2 εάν είναι ενοικιαστής
- 3 εάν είναι φιλοξενούμενος
- ΟΙΚΟΓΚΑΤΑΣΤΑΣΗ: Κατηγορική μεταβλητή με τιμές
  - 1 εάν είναι έγγαμος
  - 2 εάν είναι άγαμος
  - 3 διαζευγμένος
  - 4 άλλο
  - 5 χήρος
  - 6 σε διάσταση
- ΕΠΑΓΓΕΛΜΑ: Κατηγορική μεταβλητή με τιμές
  - 1 ελεύθερος επαγγελματίας
  - 2 αυτοαπασχολούμενος
  - 3 ιδιωτικός υπάλληλος
  - 4 στρατιωτικός/σώματα ασφαλείας
  - 5 δημόσιος υπάλληλος
  - 6 συνταξιούχος
  - 7 άνεργος
  - 8 υπάλληλος εμπορικής τράπεζας
  - 9 λοιπά επαγγέλματα
- ΠΕΛΑΤΗΣΤΡΑΠΕΖΑΣ: Κατηγορική μεταβλητή με 1 εάν είναι πελάτης και 2 εάν όχι
- ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕΣΙΑ: Κατηγορική μεταβλητή που εκφράζει τη βαθμολογία του κάθε πελάτη στο διατραπεζικό σύστημα Τειρεσίας και στο δείγμα παίρνει μία από τις παρακάτω τιμές
  - 1 Όχι δυσμενή
  - 4 Διαταγές πληρωμής ποσό άνω των 1000€
  - 7 Πλειστηριασμοί ακινήτων ποσό άνω των 1000€
  - 9 Κατασχέσεις επιταγές ποσό άνω των 1000€
  - 10 Ακάλυπτες επιταγές ποσό κάτω ή ίσο των 1000€
  - 11 Απλήρωτες συναλλαγματικές ποσό κάτω ή ίσο των 1000€

- 12 Καταναλωτικά προσωπικά δάνεια
  - 13 Πιστωτικές κάρτες
  - 17 Πλειστηριασμοί ακινήτων ποσό κάτω ή ίσο των 1000€
  - 20 Διαταγές πληρωμής ποσό κάτω ή ίσο των 1000€
  - 22 Απλήρωτες συναλλαγματικές ποσό άνω των 1000€
- ΦΥΛΟ: Κατηγορική μεταβλητή με 1 για τις γυναίκες και 0 για τους άνδρες
  - ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ: Διακριτή ποσοτική μεταβλητή
  - ΕΤΗΣΙΟΕΙΣΟΔΗΜΑχιλ€: Συνεχής ποσοτική μεταβλητή που εκφράζει το ετήσιο εισόδημα του πελάτη σε χιλιάδες ευρώ
  - ΔΕΙΚΤΗΣΚΑΘΥΣΤΕΡΗΣΗΣ90ΗΜΕΡΩΝ: Κατηγορική μεταβλητή με 0 εάν το δάνειο είναι εξυπηρετούμενο και δεν υπάρχει καθυστέρηση σε πληρωμή μεγαλύτερη των 90 ημερών και 1 εάν υπάρχει καθυστερημένη πληρωμή με καθυστέρηση πέραν των 90 ημερών.

Το δεύτερο δείγμα αποτελείται από 3432 πελάτες μίας επίσης ελληνικής τράπεζας, οι οποίοι έχουν υποστεί ρύθμιση και των οποίων παρατηρούμε τη συμπεριφορά 6 μήνες μετά τη ρύθμιση αυτή. Για το δείγμα αυτό έχουν καταγραφεί οι εξής μεταβλητές:

- ΚωδικόςΠελάτη: Αύξων αριθμός από το 1 έως το 3432 για το διαχωρισμό των πελατών του δείγματος
- ΜήνεςΚαθυστέρησης: Ποσοτική μεταβλητή που εκφράζει τον αριθμό μηνών που βρίσκεται σε καθυστέρηση η τελευταία πληρωμή του τρέχοντος προϊόντος
- ΤρέχονΥπόλοιποΔανείου: Ποσοτική μεταβλητή για το οφειλόμενο υπόλοιπο
- ΠοσόΚαθυστέρησης: Ποσοτική μεταβλητή που εκφράζει το οφειλόμενο ποσό που εκκρεμεί για πληρωμή είτε είναι σε καθυστέρηση είτε όχι (δηλαδή συμπεριλαμβάνονται και οι οφειλές του τρέχοντος μήνα)
- Διαφορά: Ποσοτική μεταβλητή που εκφράζει τη διαφορά του ποσού καθυστέρησης από το τρέχον υπόλοιπο δανείου
- Περιοχή: Ποιοτική μεταβλητή για την περιοχή κατοικίας του πελάτη
- ΕπάγγελμαΠελάτη: Ποιοτική μεταβλητή που δηλώνει την κύρια απασχόληση του πελάτη

- Ένδειξη Αναδιάρθρωσης: Κατηγορική μεταβλητή με τιμές «yes» για τα δάνεια που έχουν αναδιρθρωθεί πάλι και «no» για αυτά που δεν έχουν αναδιρθρωθεί πλέον της αρχικής ρύθμισης
- CustomerLTV: Ποσοτική μεταβλητή που εκφράζει το επίπεδο του πελάτη με το δείκτη Σύνολο Οφειλών Πελάτη προς Αξία Υποθήκης
- Case LTV: Ποσοτική μεταβλητή που εκφράζει το επίπεδο του δανείου με το δείκτη Υπόλοιπο Δανείου προς Αξία Υποθήκης
- Έτος Εκταμίευσης: Μεταβλητή που εκφράζει την πληροφορία της χρονολογίας εκταμίευσης
- Στάτους Επιχείρησης: Κατηγορική μεταβλητή με τιμές Λειτουργεί η επιχείρηση, Δεν λειτουργεί η επιχείρηση και η επιχείρηση λειτουργεί αλλά δραστηριοποιείται με άλλη μορφή
- Τζίρος Επιχείρησης: Ποσοτική μεταβλητή του τζίρου της επιχείρησης το έτος εκχώρησης της πίστωσης
- Αρχικό Ποσό Εκταμίευσης: Ποσοτική μεταβλητή που εκφράζει το ποσό που εγκρίθηκε για εκταμίευση και πιστώθηκε στο λογαριασμό του πελάτη ή δόθηκε το αντίστοιχο πιστωτικό όριο
- Ποσό που Χρησ. Από Αρχικό Ποσό Εκταμίευσης: Ποσοτική μεταβλητή για το ποσό που τελικά εκταμιεύθηκε. Διαφέρει της προηγούμενης μεταβλητής στο προϊόν «Ανοιχτές Πιστώσεις» εφόσον υπάρχουν πελάτες που δεν εκταμίευσαν το σύνολο του πιστωτικού ορίου που τους εγκρίθηκε
- Έτος Λήξης Δανείου: Χρονολογία κατά την οποία θα πρέπει να έχει αποπληρωθεί το δάνειο
- Προϊόν: Κατηγορική μεταβλητή που μας πληροφορεί για τα προϊόντα τα οποία είναι overdrafts (υπεραναλήψεις), Ανοιχτές πιστώσεις και Τοκοχρεωλυτικό δάνειο
- Μέγιστο Επίπεδο Καθυστέρησης Δανείου: Κατηγορική μεταβλητή για τη μεγαλύτερη καθυστέρηση κάποιας πληρωμής που αφορά στο παρόν προϊόν
- Μέγιστο Επίπεδο Καθυστέρησης του Πελάτη: Κατηγορική μεταβλητή για τη μεγαλύτερη καθυστέρηση κάποιας πληρωμής που αφορά στον πελάτη και στο σύνολο των οφειλών του



- ΕνεργοποίησηΡύθμισης: Κατηγορική μεταβλητή με τιμές «Ενεργοποίηση ρύθμισης» και «No requests related» για το εάν το δάνειο βρίσκεται σε ρύθμιση ή όχι
- ΣυνολικήΈκθεσητουΠελάτη: Ποσοτική μεταβλητή που εκφράζει το άθροισμα των οικονομικών υποχρεώσεων του πελάτη απέναντι στο πιστωτικό ίδρυμα
- ΑριθμόςΣυμμετεχόντωνστοΔάνειο: Ποσοτική μεταβλητή για το αριθμό των ατόμων που συμμετέχουν στο δάνειο και επομένως είναι υπόχρεοι απέναντι στο πιστωτικό ίδρυμα
- ΣτάτουςΔανείου: Κατηγορική μεταβλητή με τιμές «normal» και «denounced» για το αν το δάνειο δεν έχει ή έχει καταγγελθεί αντίστοιχα
- ΠοσόΚαταγγελίας: Ποσοτική μεταβλητή που εκφράζει το ύψος του δανείου που έχει καταγγελθεί
- ΔείκτηςΚαθυστέρησης: Κατηγορική μεταβλητή με τιμή 0 για τα εξυπηρετούμενα δάνεια και 1 και τα δάνεια σε καθυστέρηση

## 5.2 Ανάλυση των δεδομένων

Τα δεδομένα των δύο data set, των οποίων οι μεταβλητές παρουσιάστηκαν στην προηγούμενη παράγραφο, αναλύθηκαν με τη χρήση του στατιστικού πακέτου SPSS 15.0. Τα αποτελέσματα παρουσιάζονται στη συνέχεια αυτής της παραγράφου.

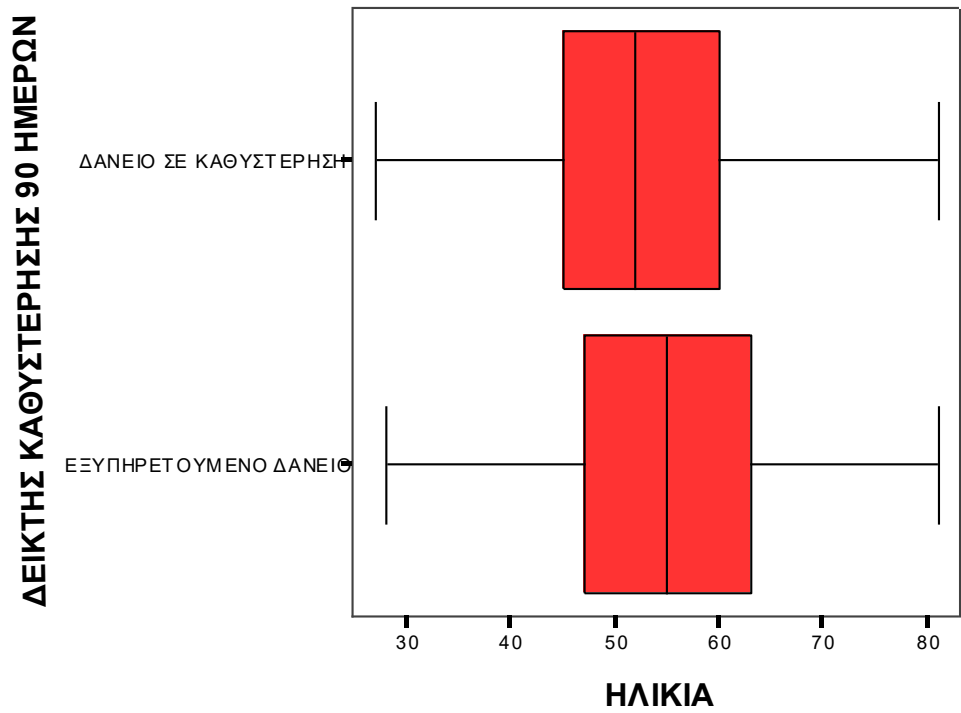
Στους Πίνακες 5.2.1 και 5.2.2. και τα σχήματα 5.2.1 και 5.2.2 δίνονται τα περιγραφικά στοιχεία των ποσοτικών μεταβλητών του πρώτου data set για τα εξυπηρετούμενα και για τα μη εξυπηρετούμενα δάνεια, δηλαδή για τους καλούς και τους κακούς πελάτες αντίστοιχα.

	N	Minimum	Maximum	Mean	Std. Deviation
ΗΛΙΚΙΑ	9210	28	81	55,05	10,802
ΕΤΗΣΙΟ ΕΙΣΟΔΗΜΑ (χιλ €)	9210	10	70	19,53	9,152
Valid N (listwise)	9210				

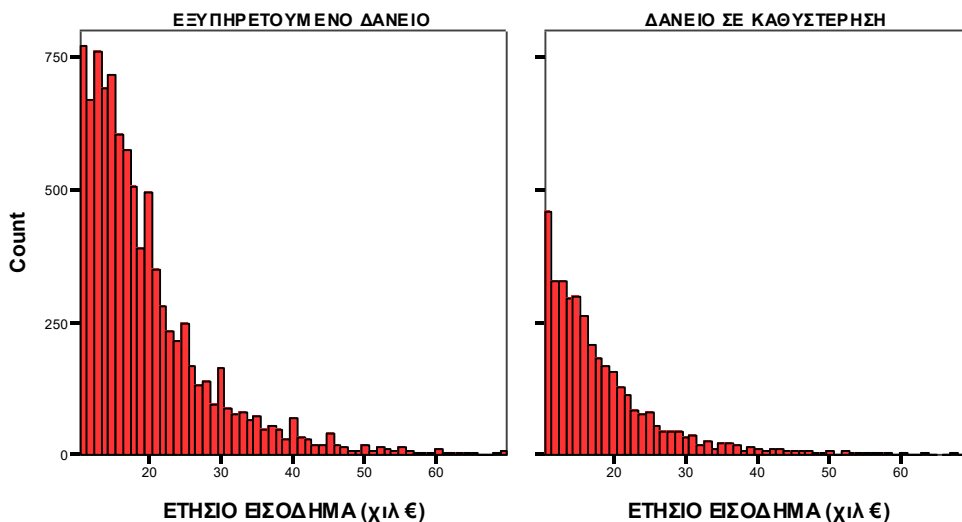
**Πίνακας 5.2.1** Περιγραφικά στατιστικά μέτρα για εξυπηρετούμενα δάνεια

	N	Minimum	Maximum	Mean	Std. Deviation
ΗΛΙΚΙΑ	3739	27	81	52,99	10,771
ΕΤΗΣΙΟ ΕΙΣΟΔΗΜΑ (χιλ €)	3739	10	70	18,56	9,235
Valid N (listwise)	3739				

**Πίνακας 5.2.2** Περιγραφικά στατιστικά μέτρα για μη εξυπηρετούμενα δάνεια



**Σχήμα 5.2.1** Συγκριτικό θηκόγραμμα ηλικιών



Σχήμα 5.2.2 Ιστόγραμμα ετήσιου εισοδήματος

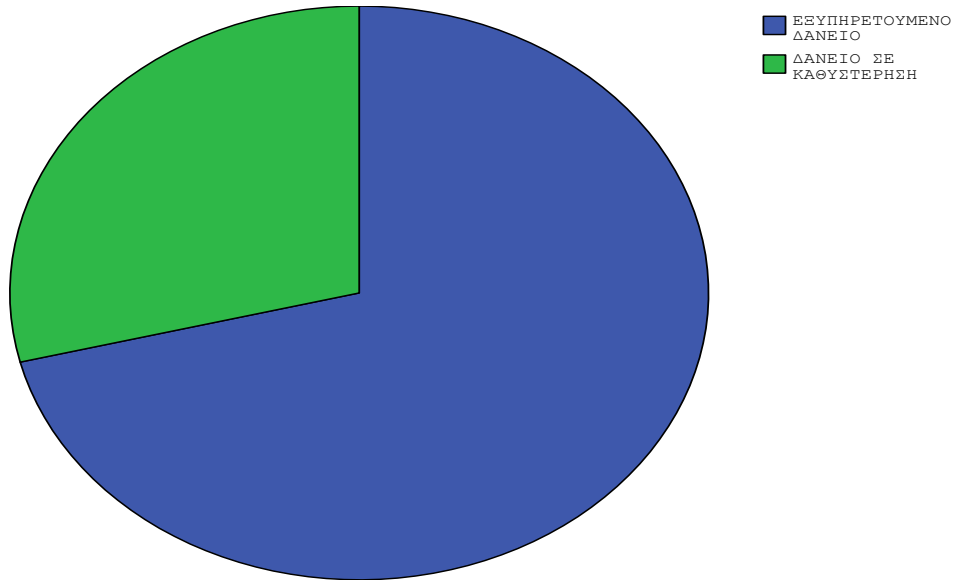
Παρατηρούμε ότι φαίνεται να υπάρχει διαφοροποίηση στην ηλικία των πελατών των δύο ομάδων, ενώ παρατηρείται και διαφορά στο μέσο όρο του ετήσιου εισοδήματος με τους πελάτες των εξυπηρετούμενων δανείων να έχουν μεγαλύτερο μέσο όρο από αυτό των «κακών» πελατών.

Εν συνεχεία, δίνονται οι πίνακες συχνοτήτων για τις ποιοτικές μεταβλητές των δεδομένων. Έτσι παρατηρείται στον πίνακα 5.2.3 ότι το δείγμα μας αποτελείται από 9210 πιστωτές με εξυπηρετούμενα δάνεια και 3739 με δάνεια σε καθυστέρηση πέρα των 90 ημερών και στον πίνακα 5.2.4 ότι 12226 είναι πελάτες του πιστωτικού ιδρύματος από το οποίο έχουν δανειοδοτηθεί ενώ 723 δεν είναι πελάτες.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ΕΞΥΠΗΡΕΤΟΥΜΕΝΟ ΔΑΝΕΙΟ	9210	71,1	71,1	71,1
	ΔΑΝΕΙΟ ΣΕ ΚΑΘΥΣΤΕΡΗΣΗ	3739	28,9	28,9	100,0
	Total	12949	100,0	100,0	

Πίνακας 5.2.3 Δείκτης καθυστέρησης 90 ημερών

**ΔΕΙΚΤΗΣ ΚΑΘΥΣΤΕΡΗΣΗΣ 90 ΗΜΕΡΩΝ**

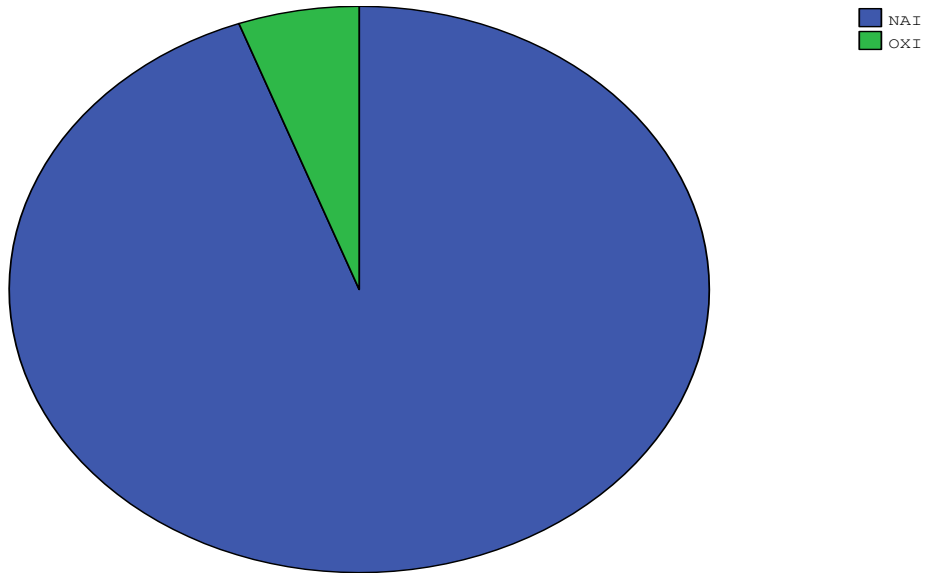


**Σχήμα 5.2.3** Διάγραμμα πίτας του δείκτη καθυστέρησης 90 ημερών

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NAI	12226	94,4	94,4	94,4
	OXI	723	5,6	5,6	100,0
	Total	12949	100,0	100,0	

**Πίνακας 5.2.4** Πελάτες

**ΠΕΛΑΤΗΣ ΤΡΑΠΕΖΑΣ**



**Σχήμα 5.2.4** Διάγραμμα πίτας της μεταβλητής πελάτης ή μη της τράπεζας

Στους επόμενους τρεις πίνακες πληροφορούμαστε για τα προσωπικά στοιχεία των πελατών όπως το φύλο, την οικογενειακή κατάστασή τους, τον αριθμό των παιδιών αλλά και το επάγγελμα .

	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός Άνδρας	3173	34,5	34,5	34,5
Καλός Γυναίκα	6037	65,5	65,5	65,5
Κακός Άνδρας	1150	12,5	12,5	30,8
Κακός Γυναίκα	2589	28,4	28,4	69,2

**Πίνακας 5.2.5** Φύλο

	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός ΕΓΓΑΜΟΣ	6791	73,7	73,7	73,7
ΑΓΑΜΟΣ	1851	20,1	20,1	93,8
ΔΙΑΖΕΥΓΜΕΝΟΣ	350	3,8	3,8	97,6
ΑΛΛΟ	195	2,1	2,1	99,8
ΧΗΡΟΣ	20	0,2	0,2	100,00
ΣΕ ΔΙΑΣΤΑΣΗ	3	,0	,0	100,00
Κακός ΕΓΓΑΜΟΣ	2457	65,7	65,7	65,7
ΑΓΑΜΟΣ	939	25,1	25,1	90,8
ΔΙΑΖΕΥΓΜΕΝΟΣ	222	5,9	5,9	96,8
ΑΛΛΟ	119	3,2	3,2	99,9
ΧΗΡΟΣ	2	0,1	0,1	100,00
ΣΕ ΔΙΑΣΤΑΣΗ	0	,0	,0	100,00

Πίνακας 5.2.6 Οικογενειακή κατάσταση

	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός ΕΛ.ΕΠΑΓΓΕΛΜΑΤΙΑΣ	947	10,3	10,3	10,3
ΑΥΤΟΑΠΑΣΧΟΛΟΥΜΕΝΟΣ	1250	13,6	13,6	23,9
ΙΔΙΟΤΙΚΟΣ ΥΠΑΛΛΗΛΟΣ	2250	24,4	24,4	48,3
ΣΤΡΑΤΙΩΤΙΚΟΣ / ΣΩΜΑΤΑ ΑΣΦΑΛΕΙΑΣ	430	4,7	4,7	53,00
ΔΗΜΟΣΙΟΣ ΥΠΑΛΛΗΛΟΣ	2661	28,9	28,9	81,8
ΣΥΝΤΑΞΙΟΥΧΟΣ	767	8,3	8,3	90,2
ΑΝΕΡΓΟΣ	56	0,6	0,6	90,8
ΥΠΑΛΛΗΛΟΣ ΕΜΠΟΡΙΚΗΣ ΤΡΑΠΕΖΑΣ	382	4,1	4,1	94,9
ΛΟΙΠΑ ΕΠΑΓΓΕΛΜΑΤΑ	467	5,1	5,1	100,00
Κακός ΕΛ.ΕΠΑΓΓΕΛΜΑΤΙΑΣ	347	9,3	9,3	9,3
ΑΥΤΟΑΠΑΣΧΟΛΟΥΜΕΝΟΣ	911	24,4	24,4	33,6
ΙΔΙΟΤΙΚΟΣ ΥΠΑΛΛΗΛΟΣ	1187	31,7	31,7	65,4
ΣΤΡΑΤΙΩΤΙΚΟΣ / ΣΩΜΑΤΑ ΑΣΦΑΛΕΙΑΣ	148	4,0	4,0	69,4
ΔΗΜΟΣΙΟΣ ΥΠΑΛΛΗΛΟΣ	598	16,0	16,0	85,3
ΣΥΝΤΑΞΙΟΥΧΟΣ	241	6,4	6,4	91,8
ΑΝΕΡΓΟΣ	12	0,3	0,3	92,1
ΥΠΑΛΛΗΛΟΣ ΕΜΠΟΡΙΚΗΣ ΤΡΑΠΕΖΑΣ	23	0,6	0,6	92,7
ΛΟΙΠΑ ΕΠΑΓΓΕΛΜΑΤΑ	272	7,3	7,3	100,00

Πίνακας 5.2.7 Επάγγελμα

Ακολουθούν οι Πίνακες 5.2.9, 5.2.10 και 5.2.11 όπου πληροφορούμαστε για στοιχεία που συνδέονται με την οικονομική κατάσταση των πιστούχων. Έτσι παρατηρείται ότι το 79% των καλών πελατών είναι ιδιοκτήτες έναντι του 71,4% των πελατών με δάνεια σε καθυστέρηση. Οι κάτοχοι πιστωτικής κάρτας με εξυπηρετούμενα δάνεια είναι περισσότεροι από τους κατόχους πιστωτικών καρτών με δάνεια σε καθυστέρηση με ποσοστά 56% και 48,5% αντίστοιχα. Τέλος, το ποσοστό με δυσμενή κατάσταση στο διατραπεζικό σύστημα Τειρεσία είναι 1,4% για τους «καλούς» έναντι 3,9% για τους «κακούς» πελάτες.

Αξιολόγηση Πελάτη	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός ΙΔΙΟΚΤΗΤΗΣ	7280	79,0	79,0	79,0
ΕΝΟΙΚΙΑΣΤΗΣ	1191	12,9	12,9	92,0
ΦΙΛΟΞΕΝΟΥΜΕΝΟΣ	739	8,0	8,0	100,0
Total	9210	100,0	100,0	
Κακός ΙΔΙΟΚΤΗΤΗΣ	2669	71,4	71,4	71,4
ΕΝΟΙΚΙΑΣΤΗΣ	563	15,1	15,1	86,4
ΦΙΛΟΞΕΝΟΥΜΕΝΟΣ	507	13,6	13,6	100,0
Total	3739	100,0	100,0	

**Πίνακας 5.2.8** Ιδιοκτησία

Αξιολόγηση Πελάτη	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός ΝΑΙ	5153	56,0	56,0	56,0
ΟΧΙ / ΝΑ	4057	44,0	44,0	100,0
Total	9210	100,0	100,0	
Κακός ΝΑΙ	1815	48,5	48,5	48,5
ΟΧΙ / ΝΑ	1924	51,5	51,5	100,0
Total	3739	100,0	100,0	

**Πίνακας 5.2.9** Κάτοχος πιστωτικής κάρτας

Αξιολόγηση Πελάτη	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός όχι δυσμενή	9080	98,6	98,6	98,6
Διαταγές Πληρωμής ποσό άνω1000€	13	,1	,1	98,7
Πλσμοί ακινήτων ποσό ανω 1000 ευρώ	2	,0	,0	98,8
Κατασχέσεις επιταγές ποσό άνω 1000 ευρώ	3	,0	,0	98,8
Ακάλυπτες επιταγές ποσό κάτω ή ίσο των1000€	13	,1	,1	98,9
πλήρωτες συναλλαγματικές ποσό κάτω ή ίσο των1000€	83	,9	,9	99,8
Καταναλωτικά προσωπικά δάνεια	7	,1	,1	99,9
Πιστωτικές κάρτες	1	,0	,0	99,9
Πλσμοί ακινήτων ποσό κάτω ή ίσο 1000 ευρώ	1	,0	,0	99,9
Διαταγές Πληρωμής ποσό κάτω ή ίσο των1000€	1	,0	,0	99,9
Απλήρωτες συναλλαγματικές ποσό άνω των1000€	6	,1	,1	100,0
Total	9210	100,0	100,0	
Κακός όχι δυσμενή	3624	96,9	96,9	96,9
Διαταγές Πληρωμής ποσό άνω1000€	8	,2	,2	97,1
Κατασχέσεις επιταγές ποσό άνω 1000 ευρώ	5	,1	,1	97,3
Ακάλυπτες επιταγές ποσό κάτω ή ίσο των1000€	13	,3	,3	97,6
Απλήρωτες συναλλαγματικές ποσό κάτω ή ίσο των1000€	75	2,0	2,0	99,6
Καταναλωτικά προσωπικά δάνεια	5	,1	,1	99,8
Πιστωτικές κάρτες	2	,1	,1	99,8
Διαταγές Πληρωμής ποσό κάτω ή ίσο των1000€	4	,1	,1	99,9
Απλήρωτες συναλλαγματικές ποσό άνω των1000€	3	,1	,1	100,0
Total	3739	100,0	100,0	

Πίνακας 5.2.10 Χειρότερη κατάσταση Τειρεσία



Στη συνέχεια δίνονται τα περιγραφικά στοιχεία του δεύτερου data set το οποίο αποτελείται από 3432 πελάτες εκ των οποίων οι 473 είναι «καλοί» δεδομένου ότι δεν έχουν καθυστερήσει σε πληρωμή ενώ οι 2959 είναι «κακοί» αφού οι οφειλές τους βρίσκονται σε καθυστέρηση με μέσο όρο τους 5,3 μήνες

Αξιολόγηση Πελάτη		N	Minimum	Maximum	Mean	Std. Deviation
Καλός	Μήνες Καθυστερήσης	473	0	0	,00	,000
	Τρέχον Υπόλοιπο Δανείου	473	36,90	1212849,40	69881,3113	112146,23308
	Ποσό Καθυστερήσης	473	,00	8047,54	280,4474	539,97068
	Υπόλοιπο Δανείου (επίπεδο πελάτη)/Αξία υποθήκης (Customer LTV)	383	,05	999,99	16,4806	124,24205
	Υπόλοιπο Δανείου (επίπεδο δανείου) /Αξία Υποθήκης (Case LTV)	296	,05	999,99	17,6292	128,98713
	Τζίρος Επιχείρησης	470	0	2516000	169198,08	327053,367
	Αρχικό Ποσό Εκταμίευσης	473	1,00	1300000,00	76550,4879	120142,00426
	Ποσό που Χρησ. Από Αρχικό Ποσό Εκταμίευσης	473	1,00	1300000,00	76550,4879	120142,00426
	Συνολική Έκθεση του Πελάτη	473	637,96	1443255,25	126392,1428	181170,20165
	Valid N (listwise)	296				
Κακός	Μήνες Καθυστερήσης	2959	3	12	5,33	2,580
	Τρέχον Υπόλοιπο Δανείου	2959	14,67	1540582,20	74663,8719	117527,04544
	Ποσό Καθυστερήσης	2959	60,00	88713,03	3887,9552	5830,60544
	Υπόλοιπο Δανείου (επίπεδο πελάτη)/Αξία υποθήκης (Customer LTV)	2045	,01	999,99	15,0460	114,65119
	Υπόλοιπο Δανείου (επίπεδο δανείου) / Αξία Υποθήκης (Case LTV)	1823	,02	999,99	14,5013	114,94722
	Τζίρος Επιχείρησης	2959	0	407357000	675880,47	13039846,878
	Αρχικό Ποσό Εκταμίευσης	2959	,00	1506046,91	87050,7878	123508,90699
	Ποσό που Χρησ. Από Αρχικό Ποσό Εκταμίευσης	2959	,00	1506046,91	86854,7443	123096,80176
	Συνολική Έκθεση του Πελάτη	2959	202,48	2742652,05	151850,4821	227560,90396
	Valid N (listwise)	1823				

Πίνακας 5.2.11 Περιγραφικά στατιστικά

Παρατηρείται μεταξύ άλλων στον πίνακα των περιγραφικών μέτρων ότι ο μέσος όρος της συνολικής έκθεσης των πελατών που βρίσκονται σε καθυστέρηση είναι μεγαλύτερος από αυτόν των πελατών που δεν είναι σε καθυστέρηση (151,850 έναντι 126,392) αλλά και η μεγάλη διαφορά στο τζίρο των επιχειρήσεων με μέσο όρο 169,198 για τους καλούς πελάτες έναντι 675,880 για τους κακούς.

Εν συνεχεία παρατίθενται οι πίνακες συχνότητας. Σύμφωνα με αυτούς έχει αναδιαρθρωθεί το 81% των εξυπηρετούμενων δανείων και το 69,1% των δανείων σε καθυστέρηση.

Αξιολόγηση Πελάτη	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός no	90	19,0	19,0	19,0
yes	383	81,0	81,0	100,0
Total	473	100,0	100,0	
Κακός no	914	30,9	30,9	30,9
yes	2045	69,1	69,1	100,0
Total	2959	100,0	100,0	

Πίνακας 5.2.12 Ένδειξη Αναδιάρθωσης

Δεν παρατηρούνται διαφορές όσον αφορά στο στάτους των επιχειρήσεων με ποσοστό άνω του 95% να είναι επιχειρήσεις που συνεχίζουν να δραστηριοποιούνται στον ίδιο τομέα ασχέτως αν είναι «καλοί» ή «κακοί» πελάτες σύμφωνα με τον πίνακα συχνότητας 5.2.14.

Αξιολόγηση Πελάτη	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός				
Δεν Λειτουργεί η επιχ	3	,6	,6	,6
Δραστηριοποίηση άλλης Μορφής	8	1,7	1,7	2,3
Λειτουργεί η επιχείρηση	1	,2	,2	2,5
Total	461	97,5	97,5	100,0
Total	473	100,0	100,0	
Κακός				
Δεν Λειτουργεί η επιχ	68	2,3	2,3	2,3
Δραστηριοποίηση άλλης Μορφής	3	,1	,1	2,4
Λειτουργεί η επιχείρηση	2888	97,6	97,6	100,0
Total	2959	100,0	100,0	

Πίνακας 5.2.13 Κατάσταση Επιχείρησης

Στον πίνακα συχνοτήτων 5.2.15 για το μέγιστο επίπεδο καθυστέρησης που έχει φτάσει ένα δάνειο, παρατηρείται ότι το αθροιστικό ποσοστό επί τοις εκατό για τους καλούς πελάτες φτάνει τα 90% μόλις στο επίπεδο 4 ενώ για τους κακούς πελάτες στο επίπεδο 10. Δηλαδή οι πελάτες που είναι σε καθυστέρηση έχουν υπάρξει «κακοπληρωτές» και στο παρελθόν με μεγαλύτερα διαστήματα καθυστέρησης σε σχέση με τους «καλοπληρωτές» εκ των οποίων όσοι καθυστέρησαν πληρωμές στο παρελθόν ήταν για μικρότερο χρονικό διάστημα (χαμηλότερο επίπεδο καθυστέρησης).

Αξιολόγηση Πελάτη	Frequency	Percent	Valid Percent	Cumulative Percent	
Καλός	0	63	13,3	13,3	
	1	242	51,2	64,5	
	2	48	10,1	74,6	
	3	41	8,7	83,3	
	4	24	5,1	88,4	
	5	15	3,2	91,5	
	6	8	1,7	93,2	
	7	7	1,5	94,7	
	8	6	1,3	96,0	
	9	8	1,7	97,7	
	10	2	,4	98,1	
	11	2	,4	98,5	
	12	3	,6	99,2	
	13	4	,8	100,0	
Total	473	100,0	100,0		
Κακός	2	476	16,1	16,1	
	3	372	12,6	28,7	
	4	292	9,9	38,5	
	5	591	20,0	58,5	
	6	658	22,2	80,7	
	7	148	5,0	85,7	
	8	75	2,5	88,3	
	9	33	1,1	89,4	
	10	21	,7	90,1	
	11	223	7,5	97,6	
	12	24	,8	98,4	
	13	46	1,6	100,0	
	Total	2959	100,0	100,0	

Πίνακας 5.2.14 Μέγιστο Επίπεδο Καθυστέρησης Δανείου

Αντίστοιχα με τον πίνακα συχνοτήτων 5.2.15, στον 5.2.16 βλέπουμε το μέγιστο επίπεδο καθυστέρησης σε επίπεδο πελάτη συνυπολογίζοντας δηλαδή και ιστορικά στοιχεία πληρωμών . Και εδώ οι καλοί πελάτες έχουν μεγαλύτερα ποσοστά σε μικρότερα επίπεδα καθυστέρησης σε σχέση με τους κακούς πελάτες. Έτσι το 50% των καλών πελατών βρίσκεται στο επίπεδο 5 και κάτω ενώ το 50% των κακών πελατών βρίσκεται στο επίπεδο 7 και κάτω.

Αξιολόγηση Πελάτη	Frequency	Percent	Valid Percent	Cumulative Percent
Καλός 1	23	4,9	4,9	4,9
2	30	6,3	6,3	11,2
3	49	10,4	10,4	21,6
4	74	15,6	15,6	37,2
5	61	12,9	12,9	50,1
6	55	11,6	11,6	61,7
7	32	6,8	6,8	68,5
8	23	4,9	4,9	73,4
9	23	4,9	4,9	78,2
10	11	2,3	2,3	80,5
11	7	1,5	1,5	82,0
12	54	11,4	11,4	93,4
13	31	6,6	6,6	100,0
Total	473	100,0	100,0	
Κακός 2	18	,6	,6	,6
3	130	4,4	4,4	5,0
4	239	8,1	8,1	13,1
5	436	14,7	14,7	27,8
6	555	18,8	18,8	46,6
7	281	9,5	9,5	56,1
8	217	7,3	7,3	63,4
9	157	5,3	5,3	68,7
10	116	3,9	3,9	72,6
11	237	8,0	8,0	80,6
12	281	9,5	9,5	90,1
13	282	9,5	9,5	99,7
50	10	,3	,3	100,0
Total	2959	100,0	100,0	

Πίνακας 5.2.15 Μέγιστο Επίπεδο Καθυστέρησης του Πελάτη

Ένα γενικό συμπέρασμα από τα παραπάνω στοιχεία είναι ότι δεν παρατηρούνται κάποιες διαφορές ανάμεσα στους πελάτες με πιστώσεις σε καθυστέρηση και σε αυτούς που αποπληρώνουν εγκαίρως τις οφειλές τους στο πρώτο data set όσο στο δεύτερο . Αυτό θα μπορούσε να συμβαίνει διότι το πρώτο data set δεν περιέχει αρκετές μεταβλητές και μεταβλητές με πληροφορίες ως προς την οικονομική κατάσταση των πελατών σε αντίθεση με το δεύτερο data set που είναι πιο εμπειριστατωμένο.

### 5.3 Ανάπτυξη Μοντέλων Λογιστικής Παλινδρόμησης

Στην παρούσα παράγραφο θα αναπτύξουμε τα δύο μοντέλα λογιστικής παλινδρόμησης για τα δύο data sets. Η διαδικασία επιλογής των σημαντικότερων μεταβλητών γίνεται και στα δύο με τη χρήση της βηματικής διαδικασίας (forward LR) στην οποία κάθε φορά προστίθεται και μία μεταβλητή που είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%. Τα τελικά μοντέλα θα έχουν τη μορφή:

$$\text{logit } p_i = \log\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_1 + \dots + b_nx_n.$$

Για τη δόμηση του μοντέλου της λογιστικής παλινδρόμησης του πρώτου data set χρησιμοποιούνται 9064 παρατηρήσεις δηλαδή το 70%, ενώ οι υπόλοιπες χρησιμοποιούνται για την επικύρωση του μοντέλου. Στον πίνακα 5.3.1 δίνονται όλες οι μεταβλητές μαζί με τους συντελεστές του τελικού μοντέλου (B) καθώς και τα p-values (Sig.) των ελέγχων ως προς τη σημαντικότητα των αντίστοιχων μεταβλητών.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 8(h)	ΗΛΙΚΙΑ	-,015	,003	25,232	1	,000	,985
	ΚΑΤΟΧΟΣΠΙΣΤΚΑΡΤΑΣ(1)	-,253	,050	25,188	1	,000	,776
	ΟΙΚΟΓΚΑΤΑΣΤΑΣΗ			71,643	5	,000	
	ΟΙΚΟΓΚΑΤΑΣΤΑΣΗ(1)	20,449	28389,080	,000	1	,999	76019436
	ΟΙΚΟΓΚΑΤΑΣΤΑΣΗ(2)	20,584	28389,080	,000	1	,999	3,072
	ΟΙΚΟΓΚΑΤΑΣΤΑΣΗ(3)	21,206	28389,080	,000	1	,999	86972426
	ΟΙΚΟΓΚΑΤΑΣΤΑΣΗ(4)	21,232	28389,080	,000	1	,999	3,292
	ΟΙΚΟΓΚΑΤΑΣΤΑΣΗ(5)	19,586	28389,080	,000	1	,999	16211302
	ΕΠΑΓΓΕΛΜΑ			328,884	8	,000	10,481
							16637583
							84,283
							32073944
							3,134

ΕΠΑΓΓΕΛΜΑ(1)	-427	,122	12,244	1	,000	,653
ΕΠΑΓΓΕΛΜΑ(2)	,218	,113	3,710	1	,054	1,243
ΕΠΑΓΓΕΛΜΑ(3)	-249	,110	5,158	1	,023	,779
ΕΠΑΓΓΕΛΜΑ(4)	-542	,148	13,440	1	,000	,581
ΕΠΑΓΓΕΛΜΑ(5)	-991	,113	77,576	1	,000	,371
ΕΠΑΓΓΕΛΜΑ(6)	-585	,144	16,475	1	,000	,557
ΕΠΑΓΓΕΛΜΑ(7)	-901	,386	5,445	1	,020	,406
ΕΠΑΓΓΕΛΜΑ(8)	-2,203	,257	73,233	1	,000	,110
ΠΕΛΑΤΗΣΤΡΑΠΕΖΑΣ(1)	-543	,104	27,195	1	,000	,581
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ			20,036	10	,029	
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(1)	,416	1,189	,123	1	,726	1,516
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(2)	,359	1,287	,078	1	,780	1,432
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(3)	-20,019	28419,359	,000	1	,999	,000
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(4)	1,646	1,410	1,363	1	,243	5,189
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(5)	1,605	1,289	1,549	1	,213	4,975
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(6)	,970	1,206	,647	1	,421	2,639
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(7)	,610	1,364	,200	1	,655	1,841
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(8)	1,146	1,861	,379	1	,538	3,146
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(9)	-20,137	40192,970	,000	1	1,000	,000
ΧΕΙΡΟΤΕΡΗΚΑΤΑΧΤΕΙΡΕ ΣΙΑ(10)	2,817	1,663	2,871	1	,090	16,730
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ			22,137	9	,008	
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(1)	-21,814	39833,014	,000	1	1,000	,000
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(2)	-21,811	39833,014	,000	1	1,000	,000
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(3)	-21,783	39833,014	,000	1	1,000	,000
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(4)	-21,649	39833,014	,000	1	1,000	,000
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(5)	-21,247	39833,014	,000	1	1,000	,000
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(6)	-20,625	39833,014	,000	1	1,000	,000
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(7)	-18,878	39833,014	,000	1	1,000	,000
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(8)	,464	56587,489	,000	1	1,000	1,591
ΑΡΙΘΜΟΣΠΑΙΔΙΩΝ(9)	-41,374	48784,329	,000	1	,999	,000
ΙΔΙΟΚΤΗΣΙΑ			36,092	2	,000	
ΙΔΙΟΚΤΗΣΙΑ(1)	-510	,091	31,570	1	,000	,600
ΙΔΙΟΚΤΗΣΙΑ(2)	-289	,107	7,344	1	,007	,749
Constant	2,151	49055,637	,000	1	1,000	8,593

Πίνακας 5.3.1 Μεταβλητές που περιλαμβάνονται στο 1<sup>ο</sup> μοντέλο

Ο έλεγχος της μηδενικής υπόθεσης ότι ο συντελεστής της εκάστοτε μεταβλητής ισούται με μηδέν, έναντι της εναλλακτικής ότι είναι διάφορος του μηδέν εξετάζεται με το κριτήριο του Wald. Από τον πίνακα 5.3.1 και σύμφωνα με το κριτήριο του Wald

περιλαμβάνονται στο μοντέλο όλες οι μεταβλητές που είναι στατιστικά σημαντικές δηλαδή όπου τα p-values είναι μικρότερα του 0.05. Το ετήσιο εισόδημα σε χιλιάδες και το φύλο είναι οι μεταβλητές που δεν περιλαμβάνονται στο μοντέλο και παρουσιάζονται στον πίνακα 5.3.2

			Score	df	Sig.
Step 8	Variables	ΕΤΗΣΙΟΕΙΣΟΔΗΜΑ χιλ€	1,961	1	,161
		ΦΥΛΟ(1)	3,204	1	,073
	Overall Statistics		5,528	2	,063

**Πίνακας 5.3.2** Μεταβλητές που δεν περιλαμβάνονται στο 1<sup>ο</sup> μοντέλο

Η αξιολόγηση της προσαρμογής του μοντέλου στα δειγματικά δεδομένα δίνεται στον πίνακα 5.3.3 και γίνεται με το λόγο των μέγιστων τιμών της συνάρτησης πιθανοφάνειας (likelihood ratio statistics) για το εξεταζόμενο μοντέλο και το μοντέλο που περιλαμβάνει μόνο τον σταθερό όρο. Το p-value του μοντέλου είναι 0,000 και επομένως το μοντέλο είναι στατιστικά σημαντικό, δηλαδή απορρίπτεται η μηδενική υπόθεση ότι οι συντελεστές των μεταβλητών είναι όλοι ίσοι με μηδέν.

		Chi-square	df	Sig.
Step 8	Step	23,603	10	,009
	Block	712,395	37	,000
	Model	712,395	37	,000

**Πίνακας 5.3.3** Αξιολόγηση του 1<sup>ου</sup> μοντέλου

Στον πίνακα 5.3.4 παρατηρούμε ότι ο συντελεστής προσδιορισμού του Nagelkerke στο 8ο βήμα της διαδικασίας, είναι 0,11 που σημαίνει ότι μόνο το 11% της μεταβλητότητας της μεταβλητής απόκρισης ερμηνεύεται από τις 9 μεταβλητές του τελικού μοντέλου.

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
8	9844,412(b)	,076	,110

**Πίνακας 5.3.4** Ερμηνευσιμότητα του 1<sup>ου</sup> μοντέλου

Τέλος, υπολογίζοντας τα ποσοστά σωστής και λανθασμένης ταξινόμησης με βάση τις 3884 παρατηρήσεις του δείγματος επικύρωσης για να μετρήσουμε την απόδοση του μοντέλου μας, προκύπτει από τον πίνακα 5.3.5 ότι το συνολικό ποσοστό ακρίβειας του τελικού μοντέλου είναι 67%

Observed			Predicted					
			Selected Cases(a)			Unselected Cases(b)		
			ΔΕΙΚΤΗΣ ΚΑΘΥΣΤΕΡΗΣΗΣ 90 ΗΜΕΡΩΝ		%	ΔΕΙΚΤΗΣ ΚΑΘΥΣΤΕΡΗΣΗΣ 90 ΗΜΕΡΩΝ		%
			ΕΞΥΠ/ΝΟ ΔΑΝΕΙΟ	ΔΑΝΕΙΟ ΣΕ ΚΑΘ/ΣΗ		ΕΞΥΠ/ΝΟ ΔΑΝΕΙΟ	ΔΑΝΕΙΟ ΣΕ ΚΑΘ/ΣΗ	
			Correct		Correct		Correct	
Step 9	ΔΕΙΚΤΗΣ ΚΑΘ/ΣΗΣ 90 ΗΜΕΡΩΝ	ΕΞΥΠ/ΝΟ ΔΑΝΕΙΟ	6465	160	97,6	2456	129	95,0
		ΔΑΝΕΙΟ ΣΕ ΚΑΘ/ΣΗ	2216	223	9,1	1154	126	11,2
		Overall Percentage			73,8			67,0

Πίνακας 5.3.5 Πίνακας ταξινόμησης για το 1<sup>ο</sup> μοντέλο

Για τη δόμηση του μοντέλου της λογιστικής παλινδρόμησης του δεύτερου data set χρησιμοποιούνται 2402 παρατηρήσεις δηλαδή το 70% και εδώ, ενώ οι υπόλοιπες χρησιμοποιούνται για την επικύρωση του μοντέλου. Στον πίνακα 5.3.6 δίνονται όλες οι μεταβλητές μαζί με τους συντελεστές του τελικού μοντέλου (B) καθώς και τα p-values (Sig.) των ελέγχων ως προς τη σημαντικότητα των αντίστοιχων μεταβλητών.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 3(c)	Μέγιστο Επίπεδο Καθυστέρησης Δανείου			27,791	13	,010	
	Μέγιστο Επίπεδο Καθυστέρησης Δανείου(1)	-46,360	18761,103	,000	1	,998	,000
	Μέγιστο Επίπεδο Καθυστέρησης Δανείου(2)	-54,308	14335,428	,000	1	,997	,000
	Μέγιστο Επίπεδο Καθυστέρησης Δανείου(3)	-21,047	13403,856	,000	1	,999	,000



ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(4)	-20,077	13403,856	,000	1	,999	,000
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(5)	-18,517	13403,856	,000	1	,999	,000
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(6)	-18,971	13403,856	,000	1	,999	,000
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(7)	-18,090	13403,856	,000	1	,999	,000
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(8)	-18,522	13403,856	,000	1	,999	,000
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(9)	,981	16656,595	,000	1	1,000	2,667
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(10)	-,104	18585,481	,000	1	1,000	,901
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(11)	-1,125	25466,013	,000	1	1,000	,325
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(12)	-,836	15048,624	,000	1	1,000	,433
ΜέγιστοΕπίπεδοΚαθυστέρησηςΔανείου(13)	-,644	19260,494	,000	1	1,000	,525
ΈτοςΕκταμίευσης			46,498	11	,000	
ΈτοςΕκταμίευσης(1)	-6,213	1,513	16,857	1	,000	,002
ΈτοςΕκταμίευσης(2)	14,323	12126,813	,000	1	,999	1661118,608
ΈτοςΕκταμίευσης(3)	-4,408	1,149	14,716	1	,000	,012
ΈτοςΕκταμίευσης(4)	12,771	3564,972	,000	1	,997	351912,110
ΈτοςΕκταμίευσης(5)	-4,294	1,330	10,427	1	,001	,014
ΈτοςΕκταμίευσης(6)	-4,219	,847	24,799	1	,000	,015
ΈτοςΕκταμίευσης(7)	-2,661	,942	7,987	1	,005	,070
ΈτοςΕκταμίευσης(8)	-3,391	1,008	11,312	1	,001	,034
ΈτοςΕκταμίευσης(9)	-1,969	1,191	2,734	1	,098	,140
ΈτοςΕκταμίευσης(10)	-3,412	,654	27,202	1	,000	,033
ΈτοςΕκταμίευσης(11)	-2,145	,561	14,635	1	,000	,117
ΑριθμόςΣυμμετεχόντωνστο Δάνειο			8,989	5	,110	
ΑριθμόςΣυμμετεχόντωνστο Δάνειο(1)	3,856	1,520	6,436	1	,011	47,293
ΑριθμόςΣυμμετεχόντωνστο Δάνειο(2)	4,258	1,563	7,417	1	,006	70,659
ΑριθμόςΣυμμετεχόντωνστο Δάνειο(3)	5,240	1,844	8,078	1	,004	188,747
ΑριθμόςΣυμμετεχόντωνστο Δάνειο(4)	3,807	1,743	4,772	1	,029	45,023

Αριθμός Συμμετεχόντων στο Δάνειο(5)	20,980	27047,039	,000	1	,999	12921665 18,326
Constant	20,597	13403,856	,000	1	,999	88097664 0,925

Πίνακας 5.3.6 Μεταβλητές που περιλαμβάνονται στο 2<sup>ο</sup> μοντέλο

Όπως αναφέραμε και στο πρώτο μοντέλο χρησιμοποιούμε το κριτήριο του Wald για να ελέγξουμε αν είναι στατιστικά σημαντικές οι μεταβλητές μας. Παρατηρούμε ότι στο μοντέλο συμμετέχουν οι μεταβλητές οι οποίες έχουν p-value μικρότερο του 0,05 και επομένως είναι στατιστικά σημαντικές ενώ οι μεταβλητές που δεν περιλαμβάνονται στο τελικό μοντέλο, φαίνονται στον πίνακα 5.3.7 .

		Score	df	Sig.
Step 3	Variables			
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη	8,411	13	,816
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(1)	,000	1	1,000
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(2)	5,081	1	,024
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(3)	,303	1	,582
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(4)	,414	1	,520
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(5)	,066	1	,798
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(6)	,013	1	,908
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(7)	,984	1	,321
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(8)	1,355	1	,244
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(9)	,136	1	,712
	Μέγιστο Επίπεδο Καθυστερήσεων του Πελάτη(10)	,084	1	,772

ΜέγιστοΕπίπεδοΚαθυστέρησηςτουΠελάτη(11)	,063	1	,802
ΜέγιστοΕπίπεδοΚαθυστέρησηςτουΠελάτη(12)	,020	1	,886
ΜέγιστοΕπίπεδοΚαθυστέρησηςτουΠελάτη(13)	,107	1	,744
ΤζίροςΕπιχείρησης	,513	1	,474
ΑρχικόΠοσόΕκταμίευσης	1,370	1	,242
ΠοσόπουΧρησ.ΑπόΑρχικόΠοσόΕκταμίευσης	1,364	1	,243
CustomerLTV	,165	1	,684
CaseLTV	,162	1	,688
ΣυνολικήΈκθεσητουΠελάτη	3,528	1	,060

**Πίνακας 5.3.7** Μεταβλητές που δεν περιλαμβάνονται στο 2<sup>ο</sup> μοντέλο

Η αξιολόγηση της προσαρμογής του μοντέλου στα δειγματικά δεδομένα δίνεται στον πίνακα 5.3.8 όπου το p-value του μοντέλου είναι 0,000 και επομένως το μοντέλο είναι στατιστικά σημαντικό, δηλαδή απορρίπτεται η μηδενική υπόθεση ότι οι συντελεστές των μεταβλητών είναι όλοι ίσοι με μηδέν.

	Chi-square	df	Sig.
Step 6 Step	7,922	5	,000
Block	500,615	29	,000
Model	500,615	29	,000

**Πίνακας 5.3.8** Αξιολόγηση του 2<sup>ου</sup> μοντέλου

Στον πίνακα 5.3.9 παρατηρούμε ότι ο συντελεστής προσδιορισμού του Nagelkerke στο 3<sup>ο</sup> βήμα της διαδικασίας, είναι 0,756 που σημαίνει ότι το 75,6% της μεταβλητότητας της μεταβλητής απόκρισης ερμηνεύεται από τις μεταβλητές του τελικού μοντέλου.

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
3	249,559(c)	,426	,756

**Πίνακας 5.3.9** Ερμηνευσιμότητα του 2<sup>ου</sup> μοντέλου

Τέλος, υπολογίζοντας τα ποσοστά σωστής και λανθασμένης ταξινόμησης με βάση τις παρατηρήσεις του δείγματος επικύρωσης για να μετρήσουμε την απόδοση του μοντέλου μας, προκύπτει από τον πίνακα 5.3.5 ότι το συνολικό ποσοστό ακρίβειας του τελικού μοντέλου είναι 95,6%

Observed	Predicted							
	Selected Cases(a)				Unselected Cases(b)			
	ΔΕΙΚΤΗΣ ΚΑΘΥΣΤΕΡΗΣΗΣ			%	ΔΕΙΚΤΗΣ ΚΑΘΥΣΤΕΡΗΣΗΣ			
	ΕΞΥΠ/ΝΟ ΔΑΝΕΙΟ	ΔΑΝΕΙΟ ΣΕ ΚΑΘ/ΣΗ	% Correct		ΕΞΥΠ /ΝΟ ΔΑΝΕΙΟ	ΔΑΝΕΙΟ ΣΕ ΚΑΘ/ΣΗ	% Correct	
Step 8	ΔΕΙΚΤΗΣ ΚΑΘ/ΣΗΣ	ΕΞΥΠ/ΝΟ ΔΑΝΕΙΟ	96	35	73,3	49	19	72,1
		ΔΑΝΕΙΟ ΣΕ ΚΑΘ/ΣΗ	1	769	99,9	2	403	99,5
	Overall Percentage				96,0			95,6

**Πίνακας 5.3.10** Πίνακας ταξινόμησης για το 2<sup>ο</sup> μοντέλο

## ΚΕΦΑΛΑΙΟ 6

### ΑΞΙΟΛΟΓΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

#### 6.1 Εξαγωγή συμπερασμάτων με τη βοήθεια καμπύλων ROC

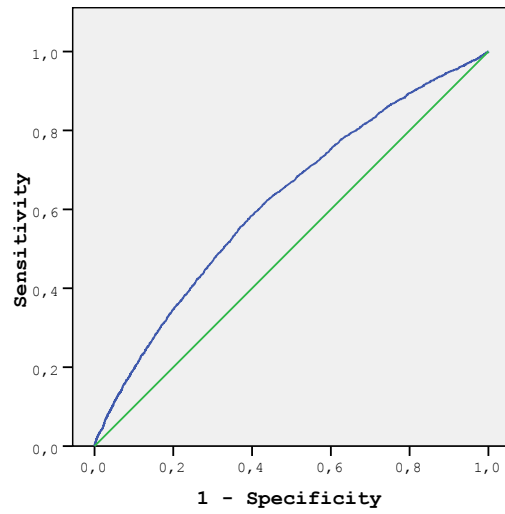
Στο προηγούμενο κεφάλαιο, επεξεργαστήκαμε δύο data sets για την αξιολόγηση πελατών ως προς την πιστοληπτική τους ικανότητα με τη δημιουργία και τη χρήση μοντέλων που αναπτύχθηκαν με τη μέθοδο της λογιστικής παλινδρόμησης. Ένας τρόπος για τη μέτρηση της απόδοσης κάθε μοντέλου είναι ο σχηματισμός της καμπύλης ROC και η μέτρηση του εμβαδού της περιοχής κάτω από τη καμπύλη.

Οι καμπύλες ROC ( Receiver Operating Characteristic curves) είναι η διαγραμματική απεικόνιση της αποτελεσματικότητας ενός μοντέλου οι οποίες μας βοηθάνε να εξετάσουμε την απόδοση του μοντέλου και να επιλέξουμε το σημείο διαχωρισμού (cut-off point) πέρα από το οποίο θεωρούμε ότι το δάνειο είναι σε καθυστέρηση. Στον άξονα X απεικονίζεται η αναλογία των ψευδών θετικών περιπτώσεων και στον άξονα Y η ευαισθησία για όλες τις τιμές του ελέγχου που παρατηρούμε στο δείγμα. Για ένα μοντέλο άριστης πρόβλεψης το οποίο θα ταξινομούσε με απόλυτη ακρίβεια ένα πιστωτή σε καλό ή κακό η καμπύλη έχει τη μορφή μιας ευθείας παράλληλης προς τον οριζόντιο άξονα η οποία τέμνει τον κάθετο άξονα στο σημείο (0,1). Στην αντίθετη περίπτωση, για ένα μη ικανοποιητικό μοντέλο η καμπύλη έχει τη μορφή μιας ευθείας γραμμής η οποία διέρχεται από την αρχή των αξόνων και έχει κλίση  $45^\circ$ .

Σημαντική ποσότητα στις καμπύλες ROC είναι η περιοχή κάτω από την καμπύλη (AUC - area under curve) και η μέτρηση του εμβαδού της περιοχής αυτής. Όσο μεγαλύτερο είναι το εμβαδό, τόσο καλύτερη θα είναι και η απόδοση του μοντέλου. Ο έλεγχος της υπόθεσης είναι  $H_0 : AUC = 0,05$  έναντι της εναλλακτικής  $H_1 : AUC > 0$ , όπου η τιμή 0,05 αντιστοιχεί σε ένα μοντέλο που μαντεύει τυχαία αν ένας πελάτης θα είναι καλός ή κακός και ουσιαστικά δεν υπάρχει καμία προγνωστική ικανότητα.

Στα σχήματα 6.1.1 και 6.1.2 παρουσιάζονται αντίστοιχα οι καμπύλες ROC για το πρώτο και το δεύτερο μοντέλο που της λογιστικής παλινδρόμησης του προηγούμενου κεφαλαίου.

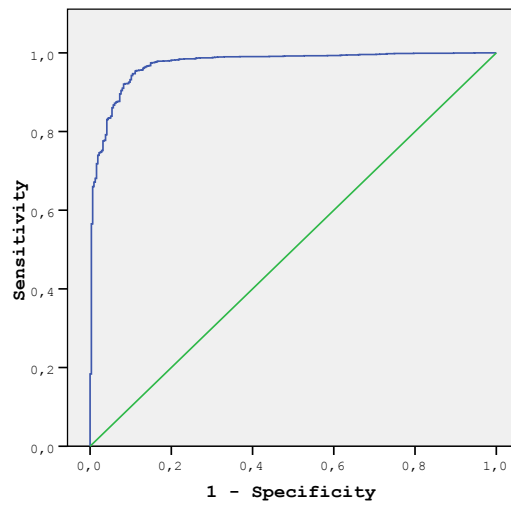
ROC Curve



Diagonal segments are produced by ties.

Σχήμα 6.1.1 Καμπύλη ROC 1<sup>ου</sup> μοντέλου

ROC Curve



Σχήμα 6.1.2 Καμπύλη ROC 2<sup>ου</sup> μοντέλου

Στους πίνακες 6.1.1 και 6.1.2 δίνονται τα εμβαδά της περιοχής κάτω από την καθεμία καμπύλη ROC τα οποία είναι 0,620 και 0,972 αντίστοιχα ενώ είναι σαφές ότι το 2ο μοντέλο που σχηματίσαμε με τη μέθοδο της λογιστικής παλινδρόμησης είναι σαφώς καλύτερο από το πρώτο .

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,620	,005	,000	,610	,631

**Πίνακας 6.1.1** Καμπύλη ROC για το 1<sup>ο</sup> μοντέλο

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,972	,005	,000	,963	,980

**Πίνακας 6.1.2** Καμπύλη ROC για το 2<sup>ο</sup> μοντέλο

## 6.2 Συμπεράσματα – Συζήτηση

Στην παρούσα εργασία, έπειτα από την θεωρητική παρουσίαση τριών τεχνικών ανάπτυξης μοντέλων για την ανάλυση δεδομένων, με στόχο την αξιολόγηση των πελατών ως προς τον πιστωτικό κίνδυνο προχωρήσαμε στην πρακτική εφαρμογή της ανάπτυξης δύο μοντέλων από δύο διαφορετικά δείγματα. Η ανάπτυξη των μοντέλων έγινε με τη χρήση του στατιστικού πακέτου SPSS 15.0 με τη μέθοδο της λογιστικής παλινδρόμησης ενώ και στα δύο δείγματα χρησιμοποιήθηκε το 30% των δεδομένων ως δείγμα επικύρωσης.

Από τα αποτελέσματα της ανάλυσης όπως αυτά παρουσιάστηκαν στο κεφάλαιο 5, προέκυψε ότι ενώ και τα δύο μοντέλα αξιολογήθηκαν ως στατιστικά σημαντικά το πρώτο από αυτά είχε χαμηλό ποσοστό πρόβλεψης (66%) καθώς και χαμηλό ποσοστό ερμηνείας της μεταβλητότητας της μεταβλητής απόκρισης (μόλις 6,1%). Στο δεύτερο μοντέλο, το συνολικό ποσοστό ακρίβειας του τελικού μοντέλου ήταν 97,3%. Τα αποτελέσματα αυτά και η υπεροχή του δεύτερου μοντέλου επικυρώθηκαν στην πρώτη παράγραφο του παρόντος κεφαλαίου και με τις αντίστοιχες καμπύλες ROC.

Δεδομένου ότι τα δύο data set τα οποία χρησιμοποιήθηκαν ήταν διαφορετικά και ως προς το μέγεθος και ως προς τη φύση των δεδομένων καταλήγουμε τονίζοντας τη σημασία που έχει η επιλογή των μεταβλητών που θα χρησιμοποιηθούν για τη σωστή και έγκυρη παραγωγή ενός μοντέλου πρόβλεψης και αξιολόγησης του πιστωτικού κινδύνου.



## ΒΙΒΛΙΟΓΡΑΦΙΑ

- Βασιλάκη Γ. Μαρία (2010), *Στατιστικά Μοντέλα Βαθμολόγησης Πιστοληπτικής Ικανότητας*, Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα σπουδών στην Εφαρμοσμένη Στατιστική του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιά
- Γναρδέλλης Χαράλαμπος (2006) *Ανάλυση Δεδομένων με το SPSS 14.0 for Windows*, Εκδόσεις Παπαζήση, Αθήνα
- Ζαπουνίδης Κωνσταντίνος και Λεμονάκης Χρήστος (2009) *Διαχείριση Πιστωτικού Κινδύνου*, Εκδόσεις Κλειδάριθμος, Αθήνα
- Καρλής Δημήτρης (2005) *Πολυμεταβλητή Στατιστική Ανάλυση*, Εκδόσεις Αθ.Σταμούλη, Αθήνα
- Ντζούφρας Ιωάννης & Περπέρογλου Άρης (2009) , *Εισαγωγή στην Βιοστατιστική και την Επιδημιολογία*
- Ξενή Μαρία (2012), *Λογιστική Παλινδρόμηση & Διαχωριστική Ανάλυση*, Διπλωματική εργασία, Μεταπτυχιακό Πρόγραμμα σπουδών «Μαθηματικά των υπολογιστών και των αποφάσεων» του τμήματος Μαθηματικών, Πανεπιστήμιο Πατρών
- Τράπεζα της Ελλάδος, Έγγραφο Διαβούλευσης II (2004), *Μέθοδος Εσωτερικών Διαβαθμίσεων*, έκδοση ΤτΕ, Αθήνα
- Τράπεζα της Ελλάδος, Έγγραφο Διαβούλευσης V (2004), *Βασικές Προϋποθέσεις για την Ανάπτυξη της Μεθόδου των Εσωτερικών Διαβαθμίσεων*, έκδοση ΤτΕ, Αθήνα
- Basel Committee on Banking Supervision (BCBS) (2006), *International Convergence of Capital Measurement and Capital Standards – A Revised Framework (Comprehensive Version)*, Bank for International Settlements, Basel.
- Bryan F.J. Manly (1993) *Multivariate Statistical Methods, A Primer*, second edition
- David W. Hosmer, Jr., Stanley Lemeshow, Rodney X. Sturdivant (2013) *Applied Logistic Regression* Third Edition, Wiley
- Durand D. (1941) *Risk Elements in Consumer Installment Financing*, National Bureau of Economic Research, New York
- Fisher R.A. (1936), *The use of multiple measurements in taxonomic problems*, Annals of Eugenics, 7,179-188
- Tony Van Gestel & Bart Baesens (2009), *Credit Risk Management* , Oxford University Press