

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΕΦΑΡΜΟΓΗ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΠΟΛΥΜΕΤΑΒΛΗΤΗΣ ΑΝΑΛΥΣΗΣ ΣΕ ΦΟΡΟΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ ΕΠΙΧΕΙΡΗΣΕΩΝ

Ευδοκία Θ. Λόκκα

Διατριβή

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Μάιος 2016

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**APPLICATION OF MULTIVARIATE STATISTICAL
METHODS IN ENTERPRISE TAXATION DATA**

By

Evdokia Th. Lokka

Thesis

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
Applied Statistics

Piraeus, Greece
May 2016

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών.

Τα μέλη της Επιτροπής ήσαν:

- Καθηγητής Μάρκος Κούτρας (Επιβλέπων)
- Καθηγητής Κλέων Τσίμπος
- Αναπλ. Καθηγήτρια Γεωργία Βεροπούλου

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

Στους γονείς μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω βαθύτατα τον επιβλέποντα καθηγητή μου κ. Μάρκο Κούτρα για την αμέριστη καθοδήγηση που μου προσέφερε καθ' όλη τη διάρκεια εκπόνησης της διπλωματικής αυτής εργασίας.

Περίληψη

Στα πλαίσια της διπλωματικής αυτής εργασίας εφαρμόζονται μέθοδοι Πολυμεταβλητής Ανάλυσης σε οικονομικά δεδομένα Ελληνικών βιομηχανικών επιχειρήσεων. Τα οικονομικά αυτά δεδομένα προέρχονται από τα δεδομένα που χρησιμοποιούνται από τις φορολογικές αρχές για την εκτίμηση της οικονομικής θέσης των επιχειρήσεων και την αξιολόγηση της βιωσιμότητάς τους. Οι εκτιμήσεις αυτές μπορούν να καταστούν χρήσιμες τόσο για τις επιχειρήσεις όσο και για το ίδιο το κράτος καθώς παρέχουν ενδείξεις για πιθανή αναγκαιότητα λήψης προληπτικών μέτρων που θα ευνοήσουν την αποφυγή δυσμενών μελλοντικών οικονομικών καταστάσεων όπως η πτώχευση.

Ως κριτήρια πρόβλεψης της διακοπής των λειτουργιών μιας επιχείρησης χρησιμοποιούνται βασικοί αριθμοδείκτες αποδοτικότητας, κεφαλαιακής διάρθρωσης, δραστηριότητας και ρευστότητας. Η προβλεπτική ικανότητα των δεικτών αυτών εξετάζεται μέσω μεθόδων Πολυμεταβλητής Ανάλυσης όπως η Λογιστική Παλινδρόμηση, τα μοντέλα Probit, η Ανάλυση Κύριων Συνιστωσών και οι μέθοδοι Συσταδοποίησης. Όλες οι μέθοδοι, πέραν αυτής της Συσταδοποίησης, έχουν καλή προσαρμογή στα δεδομένα, ενώ από την εφαρμογή της Λογιστικής Παλινδρόμησης προκύπτει πως ο δείκτης «Ίδια Κεφάλαια προς Ενεργητικό» έχει την καλύτερη προβλεπτική ικανότητα για τη διακοπή λειτουργίας μιας επιχείρησης.

Abstract

In the present MSc. thesis, methods of Multivariate Analysis are applied on Greek industrial business and financial data. These financial data are used by tax authorities to assess companies' financial position and viability. These estimates may become useful for both companies and the state itself since they provide clues for a possible need of preventive measures that will favour the avoidance of adverse future economic situations like (e.g. bankruptcy).

Key financial ratios of efficiency, capital structure, activity and liquidity are used for predicting business' default. The predictive power of these indicators is examined by the use of methods of Multivariate Analysis such as Logistic Regression, Probit Analysis, Principal Component Analysis and Clustering methods. All methods, apart from Clustering, exhibit a good fit to the data. Logistic Regression revealed that the index "Equity to Assets" owns the best power for predicting a business' default.

Περιεχόμενα

Κατάλογος Πινάκων.....	xv
Κατάλογος Διαγραμμάτων.....	xvii
1 Εισαγωγή.....	19
1.1 Αντικειμενικός Σκοπός της Διπλωματικής Εργασίας	19
2 Πηγές Δεδομένων και Παρουσίαση Μεταβλητών.....	21
2.1 Πηγές Δεδομένων	21
2.2 Παρουσίαση Μεταβλητών.....	23
2.2.1 Παρουσίαση Ποιοτικών Μεταβλητών	24
Περιγραφικά Στατιστικά των σημαντικότερων Ποιοτικών Μεταβλητών	24
2.2.2 Παρουσίαση Ποσοτικών Μεταβλητών: Μεγέθη	29
2.2.3 Παρουσίαση Ποσοτικών Μεταβλητών: Αριθμοδείκτες	36
Αριθμοδείκτες Αποδοτικότητας.....	36
Αριθμοδείκτες Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας	44
Αριθμοδείκτες Δραστηριότητας	48
Αριθμοδείκτες Δαπανών Λειτουργίας	51
Αριθμοδείκτες Ρευστότητας	53
3 Βιβλιογραφική Επισκόπηση	57
4 Θεωρία Στατιστικών Μεθόδων που χρησιμοποιούνται στην Ανάλυση	65
4.1 <i>Logit και Probit – Γενικευμένα Γραμμικά Μοντέλα</i>	65
4.2 Ανάλυση Κυρίων Συνιστωσών.....	68
4.3 Ανάλυση κατά Συστάδες	74
4.3.1 Μέτρα Απόστασης	74
4.3.2 Μέτρα Ομοιότητας.....	76
4.3.3 Μέθοδοι Ομαδοποίησης.....	76
5 Μεθοδολογία και Αποτελέσματα.....	83
5.1 Καθαρισμός και προετοιμασία των δεδομένων για τη στατιστική ανάλυση	83
5.2 Στατιστική Ανάλυση – Εφαρμογή μοντέλων	91
5.2.1 Εφαρμογή Λογιστικής Παλινδρόμησης	91
5.2.2 Εφαρμογή μοντέλου Probit	112
5.2.3 Εφαρμογή Ανάλυσης Κυρίων Συνιστωσών.....	115
5.2.4 Εφαρμογή Ανάλυσης κατά Συστάδες	123

6	Συμπεράσματα και προτάσεις για περαιτέρω έρευνα.....	132
	Βιβλιογραφία.....	134

Κατάλογος Πινάκων

Πίνακας 2.2-1 Ονομασία και Περιγραφή Ποιοτικών Μεταβλητών	24
Πίνακας 2.2-2 Περιγραφικά Στατιστικά μεταβλητής "Νομική Μορφή Επιχείρησης"	25
Πίνακας 2.2-3 Περιγραφικά Στατιστικά μεταβλητής "Εξαγωγέας"	25
Πίνακας 2.2-4 Περιγραφικά Στατιστικά μεταβλητής "Κλάδος Δραστηριότητας"	26
Πίνακας 2.2-5 Περιγραφικά Στατιστικά μεταβλητής "Περιοχή Δραστηριοποίησης"	27
Πίνακας 2.2-6 Περιγραφικά Στατιστικά μεταβλητής "Κατάσταση Επιχείρησης"	27
Πίνακας 2.2-7 Διαθεσιμότητα Δεδομένων ως προς τα Έτη Αναφοράς.....	28
Πίνακας 2.2-8 Σύνολο Μεταβλητών που θα χρησιμοποιηθούν στην Ανάλυση	54
Πίνακας 5.2-1 Σημαντικές Μεταβλητές μετά τη Λογιστική Παλινδρόμηση.....	92
Πίνακας 5.2-2 Αποτελέσματα Confusion Matrix της Λογιστικής Παλινδρόμησης	106
Πίνακας 5.2-3 Εκτιμήσεις συντελεστών που προέκυψαν από τα μοντέλα Logit και Probit .	113
Πίνακας 5.2-4 Εκτιμήσεις οριακών αποτελεσμάτων στον μέσο και μέσων οριακών αποτελεσμάτων που προέκυψαν από τα μοντέλα Logit και Probit.....	114

Κατάλογος Διαγραμμάτων

Διάγραμμα 5.2-1 Sensitivity / Specifity vs Probability Cutoff	107
Διάγραμμα 5.2-2 Καμπύλη ROC του μοντέλου της Λογιστικής Παλινδρόμησης	108
Διάγραμμα 5.2-3 Screeplot της Ανάλυσης Κυρίων Συνιστωσών	117
Διάγραμμα 5.2-4: Διάγραμμα διασποράς 1ης και 2ης συνιστώσας.....	120
Διάγραμμα 5.2-5: Διάγραμμα διασποράς 1ης και 3ης συνιστώσας.....	121
Διάγραμμα 5.2-6: Διάγραμμα διασποράς 1ης και 4ης συνιστώσας.....	121

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Αντικειμενικός Σκοπός της Διπλωματικής Εργασίας

Αντικείμενο της διπλωματικής αυτής αποτελεί η εφαρμογή στατιστικών μεθόδων Πολυμεταβλητής Ανάλυσης σε οικονομικά – φορολογικά δεδομένα ελληνικών βιομηχανικών επιχειρήσεων. Βάσει ενός συνόλου οικονομικών δεικτών, επιδιώκεται η ανίχνευση εκείνων των δεικτών που μπορούν να χρησιμοποιηθούν για την κατηγοριοποίηση των επιχειρήσεων σε βιώσιμες και μη βιώσιμες. Κριτήριο της κατηγοριοποίησης αποτελεί η ομοιότητα που θα εμφανίσουν οι εξεταζόμενες επιχειρήσεις ως προς τους δείκτες αυτούς. Η εν λόγω κατηγοριοποίηση αναφέρεται ουσιαστικά στο ευρύτατα αναλυμένο στη βιβλιογραφία θέμα της διαχείρισης και πρόβλεψης του πιστωτικού κινδύνου των επιχειρήσεων.

Οι στατιστικές μέθοδοι που χρησιμοποιούνται αντλούνται από τα πεδία τεχνικών διαχωρισμού δεδομένων και μείωσης των διαστάσεων αυτών. Στα πλαίσια της εργασίας αυτής, οι μέθοδοι που επιλέχθηκαν να εφαρμοστούν, λαμβάνοντας υπόψη και τη σχετική βιβλιογραφία για την αντιμετώπιση του πιστωτικού κινδύνου των επιχειρήσεων, είναι οι ακόλουθες: τα μοντέλα Λογιστικής Παλινδρόμησης, η Ανάλυση Κύριων Συνιστωσών και τέλος οι αλγόριθμοι Συσταδοποίησης. Η δομή της ανάλυσης που θα επακολουθήσει έχει ως εξής:

1. Αρχικά, θα πραγματοποιηθεί αναφορά των πηγών που χρησιμοποιήθηκαν για την ανεύρεση των δεδομένων καθώς και παράθεση γενικών πληροφοριών σχετικά με τις πηγές και τα δεδομένα αυτά. Επίσης, θα παρουσιασθεί το σύνολο των μεταβλητών που θα χρησιμοποιηθούν στην εργασία. Η παρουσίαση αυτή περιλαμβάνει περιγραφή των ποιοτικών μεταβλητών και χρηματοοικονομική και λογιστική ερμηνεία των οικονομικών ποσοτικών μεταβλητών.
2. Στη συνέχεια, θα πραγματοποιηθεί, με βάση τη βιβλιογραφία, σύντομη επισκόπηση προηγούμενων μελετών που εκπονήθηκαν στο συγκεκριμένο θέμα και παράθεση στατιστικών μεθόδων που εφαρμόστηκαν για σκοπό αντίστοιχο με αυτόν της διπλωματικής.

3. Θα παρουσιασθεί το θεωρητικό σκέλος των προαναφερθέντων βασικών στατιστικών τεχνικών Πολυμεταβλητής Ανάλυσης που θα εφαρμοσθούν στα δεδομένων.
4. Τέλος, θα παρουσιασθούν και θα αναλυθούν τα αποτελέσματα της στατιστικής επεξεργασίας των δεδομένων από τα οποία θα προκύψουν και τα τελικά συμπεράσματα της διπλωματικής.

ΚΕΦΑΛΑΙΟ 2

Πηγές Δεδομένων και Παρουσίαση Μεταβλητών

2.1 Πηγές Δεδομένων

Τα δεδομένα που χρησιμοποιούνται στη διπλωματική αυτή προέρχονται σχεδόν στο σύνολό τους από την ιστοσελίδα www.inr.gr και αφορούν πληροφοριακά (ποιοτικά) και οικονομικά (ποσοτικά) δεδομένα βιομηχανικών επιχειρήσεων που δραστηριοποιούνται στην ελληνική επικράτεια. Όπως αναφέρεται στην ίδια την ιστοσελίδα, «συμπεριλαμβάνονται εταιρείες ιχθυοκαλλιεργειών, διαλογής και τυποποίησης οπωροκηπευτικών, κτηνοτροφικές, επεξεργασίας φύλλων καπνού, εκκόκκισης βάμβακος, εκδοτικές, παραγωγής λογισμικού και άλλων προϊόντων πληροφορικής, επισκευής και μετασκευής μεταφορικών μέσων και μηχανών, καθώς και ανακύκλωσης. Αρκετές εταιρείες είναι συγχρόνως βιομηχανικές και εμπορικές, καθώς συνδυάζουν τη βιομηχανική με την εμπορική δραστηριότητα».

Η ιστοσελίδα αυτή παρουσιάζει, για κάθε μία από τις περίπου 6.540 επιχειρήσεις που περιλαμβάνει, επιλεγμένα οικονομικά δεδομένα που προέρχονται από τις δημοσιευμένες οικονομικές τους καταστάσεις, και συγκεκριμένα από τον Ισολογισμό και την Κατάσταση Αποτελεσμάτων Χρήσεως, για τα έτη 2011 έως 2014. Τα οικονομικά δεδομένα της ιστοσελίδας δεν καλύπτουν για όλες τις επιχειρήσεις το διάστημα αυτό. Παρακάτω γίνεται λεπτομερής αναφορά του πλήθους των επιχειρήσεων που καλύπτουν το εκάστοτε διάστημα. Η ιστοσελίδα αναφέρει ότι για τις περιπτώσεις επιχειρήσεων που πραγματοποίησαν υπερδωδεκάμηνη οικονομική χρήση (αυτό συμβαίνει συνήθως στον πρώτο χρόνο λειτουργίας των επιχειρήσεων καθώς, για να δημοσιεύσει η επιχείρηση μια πιο ουσιαστική και εμπειριστατωμένη εικόνα των αποτελεσμάτων της επιλέγει να δημοσιεύσει τις οικονομικές της καταστάσεις στο τέλος της χρονιάς που έπεται της ενάρξεως της αντί στην πρώτη διαχειριστική λήξη) έχει γίνει αναγωγή των αποτελεσμάτων σε δωδεκάμηνη βάση ή και καταμερισμός των αποτελεσμάτων σε δύο χρήσεις. Επίσης η ιστοσελίδα αναφέρει πως «δεν εγγυάται την ακρίβεια των συνοπτικών στοιχείων που παρουσιάζει». Πραγματοποιήθηκε δειγματοληπτική επιβεβαίωση των οικονομικών στοιχείων κάποιων εταιριών, με βάση τις δημοσιευμένες καταστάσεις στις

αντίστοιχες εταιρικές ιστοσελίδες και στο εθνικό τυπογραφείο, και τα οικονομικά στοιχεία βρέθηκαν ακριβή. Συνεπώς, με την επιφύλαξη της ακρίβειας αυτής, θεωρήθηκε ότι τα δεδομένα της ιστοσελίδας μπορούν να χρησιμοποιηθούν για τους ερευνητικούς σκοπούς της διπλωματικής.

Για όλες τις επιχειρήσεις, εκτός των οικονομικών δεδομένων, παρέχεται από την ιστοσελίδα ο Αριθμός Φορολογικού Μητρώου τους (ΑΦΜ) και ο Κλάδος στον οποίο δραστηριοποιούνται, ενώ για τις περισσότερες επιχειρήσεις παρέχονται επιπλέον πληροφοριακά στοιχεία όπως η περιοχή δραστηριοποίησης τους και ο Κωδικός Αριθμός Δραστηριότητάς (ΚΑΔ) τους με την περιγραφή του. Ο αριθμός ΚΑΔ είναι ένας 8-ψήφιος αριθμός που προσδιορίζει μια συγκεκριμένη επαγγελματική δραστηριότητα.

Η ιστοσελίδα δεν περιέχει τα δεδομένα των επιχειρήσεων σε μια ενοποιημένη βάση δεδομένων. Για το λόγο αυτό, για να καταστεί δυνατή η στατιστική επεξεργασία τους και να εκπονηθεί η παρούσα μελέτη, πραγματοποιήθηκε άντληση των οικονομικών και των πληροφοριακών δεδομένων για μία προς μία όλων των περίπου 6.540 επιχειρήσεων που περιλαμβάνονται στην ιστοσελίδα. Περαιτέρω πραγματοποιήθηκε εκτεταμένη επεξεργασία για την ενοποίηση των δεδομένων αυτών έως ότου συγκεντρωθούν σε μία ενιαία βάση δεδομένων. Όπου δεν κατέστη δυνατό να αντληθούν τα απαραίτητα πληροφοριακά δεδομένα για την κάθε επιχείρηση, πραγματοποιήθηκε αναζήτηση με βάση τον ΑΦΜ είτε στην κύρια ιστοσελίδα της επιχείρησης είτε σε άλλες ιστοσελίδες που παρέχουν πληροφοριακά δεδομένα κυρίως μητρώου όπως: η επίσημη ιστοσελίδα του εθνικού τυπογραφείου (www.et.gr), η επίσημη ιστοσελίδα της Γενικής Γραμματείας Εμπορίου του Υπουργείου Ανάπτυξης της Ελλάδας (ΓΕΜΗ) (<https://www.businessregistry.gr/publicity.aspx>), και οι επίσημες ιστοσελίδες των εμποροβιομηχανικών επιμελητηρίων της επικράτειας.

2.2 Παρουσίαση Μεταβλητών

Η ενότητα αυτή αποτελείται από τρεις υποενότητες. Η κάθε υποενότητα αφορά σε μία ειδική κατηγορία μεταβλητών.

Στην πρώτη υποενότητα θα παρουσιαστεί πίνακας όπου θα δίνεται το περιεχόμενο των ποιοτικών μεταβλητών. Οι ποιοτικές μεταβλητές αποτελούν τα πληροφοριακά στοιχεία των επιχειρήσεων. Στην ενότητα θα παρουσιαστούν περιγραφικοί στατιστικοί πίνακες για τα μοναδικά πεδία κάποιων ποιοτικών μεταβλητών που χρήζουν περαιτέρω ανάλυσης για την κατανόησή τους. Οι συγκεντρωτικοί πίνακες θα περιλαμβάνουν τη συχνότητα και τη σχετική συχνότητα των επιχειρήσεων που αντιστοιχούν στο κάθε πεδίο. Επίσης θα παρουσιαστεί αναλυτικά για κάθε έτος από το 2011 έως το 2014 το πλήθος των επιχειρήσεων για τις οποίες υπήρξε διαθεσιμότητα οικονομικών στοιχείων στην ιστοσελίδα του inf.gr.

Στη δεύτερη και τρίτη υποενότητα θα παρουσιαστεί λεπτομερώς το περιεχόμενο του συνόλου των ποσοτικών μεταβλητών. Οι ποσοτικές μεταβλητές αποτελούν τα οικονομικά στοιχεία των επιχειρήσεων και θα διακριθούν σε δύο κατηγορίες: στις μεταβλητές που αποτελούν *μεγέθη*, είτε του ισολογισμού είτε της κατάστασης αποτελεσμάτων χρήσεως, και στις μεταβλητές που αποτελούν *αριθμοδείκτες*, δηλαδή κλάσματα των προηγούμενων μεγεθών με εξαιρετικά χρήσιμη οικονομική σημασία.

Στη δεύτερη υποενότητα, συγκεκριμένα, θα παρουσιαστεί το σύνολο των οικονομικών στοιχείων που αφορά τα μεγέθη. Αρχικά θα παρουσιασθούν τα *πρωτογενή μεγέθη*, δηλαδή τα μεγέθη που λήφθηκαν αυτούσια από την ιστοσελίδα του inf.gr, ενώ στη συνέχεια θα παρουσιασθούν τα *δευτερογενή μεγέθη*, δηλαδή τα μεγέθη που παράχθηκαν από τα πρωτογενή με κατάλληλους μαθηματικούς τύπους. Και για τα δύο είδη μεγεθών θα δίνεται ο μαθηματικός τύπος και η οικονομική και λογιστική τους ερμηνεία. Τα μεγέθη αποτελούν μεταβλητές ιδιαίτερης χρησιμότητας, καθώς αποτελούν τα συστατικά στοιχεία των αριθμοδεικτών (αριθμητής και παρονομαστής).

Στην τρίτη υποενότητα θα ακολουθήσει λεπτομερής επεξήγηση του συνόλου των αριθμοδεικτών. Οι αριθμοδείκτες κατέχουν θεμελιώδη θέση στην στατιστική αυτή ανάλυση, για αυτό και θα δοθεί αναλυτική περιγραφή της λογιστικής και χρηματοοικονομικής τους έννοιας. Η παρουσίασή τους θα γίνει κατά είδος ανάλογα με το σκοπό που εξυπηρετεί ο υπολογισμός τους.

2.2.1 Παρουσίαση Ποιοτικών Μεταβλητών

Παρακάτω δίνεται πίνακας με την ονομασία και την περιγραφή των ποιοτικών μεταβλητών που περιλαμβάνονται στο σύνολο των δεδομένων . Η πλειοψηφία των ποιοτικών μεταβλητών δεν θα χρησιμοποιηθεί στην ανάλυση, παρατίθενται όμως στατιστικά στοιχεία για κάποιες από αυτές τις μεταβλητές, διότι αυτό θα βοηθήσει στην καλύτερη κατανόηση του συνόλου των δεδομένων που αντλήθηκαν από την ιστοσελίδα.

Πίνακας 2.2-1 Ονομασία και Περιγραφή Ποιοτικών Μεταβλητών

a/a	Όνομα Μεταβλητής	Περιγραφή
1	ΑΦΜ	Αναγνωριστικό της κάθε εγγραφής (id).
2	Επωνυμία	Επωνυμία της Επιχείρησης
3	Νομική μορφή	Νομική μορφή της επιχείρησης (Α.Ε, Ε.Π.Ε κ.α.).
4	Εξαγωγέας	Υποδηλώνει εάν η επιχείρηση πραγματοποιεί εξαγωγές σύμφωνα με τις πληροφορίες της ιστοσελίδας inr.gr.
5	Κλάδος Δραστηριοποίησης	Επαγγελματικός κλάδος δραστηριοποίησης της επιχείρησης σύμφωνα με την κατάταξη της στην ιστοσελίδα inr.gr.
6	Περιοχή Δραστηριοποίησης	Περιοχή δραστηριοποίησης της επιχείρησης σύμφωνα με τον ταχυδρομικό κώδικα της περιοχής όπου ανήκει η επιχείρηση.
7	Έτος Διακοπής	Έτος διακοπής λειτουργίας της επιχείρησης.
8	Κατάσταση Επιχείρησης	Ορίζεται αν η επιχείρηση δεν είναι σε ενεργή δραστηριοποίηση και βρίσκεται σε κατάσταση που δεν της επιτρέπει την παραγωγική της λειτουργία (πτώχευση, εκκαθάριση, διακοπή). Αν η επιχείρηση είναι σε ενεργή κατάσταση η μεταβλητή δεν παίρνει τιμές. Η μεταβλητή αυτή συμβαδίζει με τη μεταβλητή "Έτος Διακοπής" καθώς όταν υπάρχει κάποια ημερομηνία στο Έτος Διακοπής η επιχείρηση δεν είναι ενεργή.

Η 2^η μεταβλητή με την περιγραφή «Επωνυμία της Επιχείρησης» έχει επεξηγηματικό ρόλο καθώς το αναγνωριστικό της κάθε εγγραφής είναι ο ΑΦΜ της επιχείρησης.

Περιγραφικά Στατιστικά των σημαντικότερων Ποιοτικών Μεταβλητών

Παρακάτω θα παρουσιαστούν συγκεντρωτικοί πίνακες με τα μοναδικά πεδία των σημαντικότερων ποιοτικών μεταβλητών και θα αφορούν το πλήθος και το ποσοστό επί του συνολικού πλήθους των επιχειρήσεων που αναφέρονται σε αυτά τα μοναδικά πεδία. Τα μοναδικά πεδία των μεταβλητών εμφανίζονται ταξινομημένα ως προς τις συχνότητές τους από το μεγαλύτερο προς το μικρότερο.

i. Νομική Μορφή Επιχείρησης

Πίνακας 2.2-2 Περιγραφικά Στατιστικά μεταβλητής "Νομική Μορφή Επιχείρησης"

Νομική Μορφή	Πλήθος Εταιριών	Ποσοστό επί του Πλήθους
ΑΕ (Ανώνυμη Εταιρία)	5386	82,32%
ΕΠΕ (Εταιρία Περιορισμένης Ευθύνης)	856	13,08%
ΜΕΠΕ (Μονοπρόσωπη Εταιρία Περιορισμένης Ευθύνης)	279	4,26%
ΑΓΡΟΤΙΚΟΣ ΣΥΝΕΤΑΙΡΙΣΜΟΣ	19	0,29%
ΙΚΕ (Ιδιωτική Κεφαλαιουχική Εταιρία)	3	0,05%
Γενικό Άθροισμα Εταιριών	6543	100%

Παρατηρούμε ότι το μεγαλύτερο πλήθος των εταιριών ανήκει στην κατηγορία των ανώνυμων εταιριών πράγμα αναμενόμενο καθώς πρόκειται για βιομηχανικές εταιρίες. Η μεταβλητή αυτή δεν θα χρησιμοποιηθεί στη στατιστική ανάλυση καθώς η νομική μορφή έχει να κάνει κυρίως με τα αρχικά επενδυμένα κεφάλαια και δεν επηρεάζει γενικότερα τη βιωσιμότητα της επιχείρησης.

ii. Εξαγωγέας

Πίνακας 2.2-3 Περιγραφικά Στατιστικά μεταβλητής "Εξαγωγέας"

Εξαγωγέας	Πλήθος Εταιριών	Ποσοστό επί του Πλήθους
ΝΑΙ	885	13,53%
(κενό)	5658	86,47%
Γενικό Άθροισμα Εταιριών	6543	100%

Η πληροφορία για τη μεταβλητή αυτή προέρχεται αποκλειστικά από την ιστοσελίδα του inr.gr. Αυτό σημαίνει ότι μόνο για το 13,5% των επιχειρήσεων γνωρίζουμε με βεβαιότητα ότι πραγματοποιούν εξαγωγές. Για το υπόλοιπο ποσοστό δεν μπορούμε να ισχυριστούμε με σιγουριά ότι δεν είναι εξαγωγικές καθώς υπάρχει πιθανότητα, και μάλιστα πολύ μεγάλη επειδή πρόκειται για βιομηχανίες, οι περισσότερες από αυτές να πραγματοποιούν εξαγωγές. Ως επακόλουθο, η μεταβλητή δεν θα χρησιμοποιηθεί στην ανάλυση.

iii. Κλάδος Δραστηριοποίησης

Πίνακας 2.2-4 Περιγραφικά Στατιστικά μεταβλητής "Κλάδος Δραστηριότητας"

Κλάδος Δραστηριοποίησης	Πλήθος Εταιριών	Ποσοστό επί του Πλήθους
ΤΡΟΦΙΜΑ - ΑΓΡΟΤΙΚΑ ΠΡΟΪΟΝΤΑ	1659	25,36%
ΜΗ ΜΕΤΑΛΛΙΚΑ ΟΡΥΚΤΑ	610	9,32%
ΕΚΔΟΣΕΙΣ ΕΚΤΥΠΩΣΕΙΣ	609	9,31%
ΣΙΔΗΡΟΥΧΑ ΜΕΤΑΛΛΑ	452	6,91%
ΔΙΑΦΟΡΑ ΠΡΟΪΟΝΤΑ	324	4,95%
ΠΛΑΣΤΙΚΑ	292	4,46%
ΕΝΔΥΜΑΤΑ	281	4,29%
ΜΕΤΑΦΟΡΙΚΑ ΜΕΣΑ	279	4,26%
ΜΗΧΑΝΗΜΑΤΑ	226	3,45%
ΠΟΤΑ	220	3,36%
ΕΠΙΠΛΑ	215	3,29%
ΧΗΜΙΚΑ	199	3,04%
ΚΛΩΣΤΟΪΦΑΝΤΟΥΡΓΙΚΕΣ ΥΛΕΣ	181	2,77%
ΧΑΡΤΙ	179	2,74%
ΜΗ ΣΙΔΗΡΟΥΧΑ ΜΕΤΑΛΛΑ	172	2,63%
ΗΛΕΚΤΡΟΛΟΓΙΚΟ - ΗΛΕΚΤΡΟΝΙΚΟ ΥΛΙΚΟ	138	2,11%
ΞΥΛΟ	123	1,88%
ΦΑΡΜΑΚΑ - ΚΑΛΛΥΝΤΙΚΑ - ΑΠΟΡΡΥΠΑΝΤΙΚΑ	99	1,51%
ΗΛΕΚΤΡΙΚΕΣ ΣΥΣΚΕΥΕΣ ΦΩΤΙΣΤΙΚΑ	89	1,36%
ΠΡΟΪΟΝΤΑ ΠΛΗΡΟΦΟΡΙΚΗΣ	59	0,90%
ΟΡΥΚΤΑ - ΜΕΤΑΛΛΕΥΜΑΤΑ	41	0,63%
ΠΑΡΑΓΩΓΑ ΠΕΤΡΕΛΑΙΟΥ	36	0,55%
ΥΠΟΔΗΜΑΤΑ ΔΕΡΜΑΤΙΝΑ ΕΙΔΗ	34	0,52%
ΔΕΡΜΑ	13	0,20%
ΚΑΠΝΑ ΣΕ ΦΥΛΛΑ	9	0,14%
ΤΣΙΓΑΡΑ	4	0,06%
Γενικό Αθροισμα Εταιριών	6543	100%

Παρατηρούμε πως το μεγαλύτερο ποσοστό των διαθέσιμων εταιριών αφορά τον κλάδο των τροφίμων ενώ ακολουθούν οι κλάδοι των μη μεταλλικών ορυκτών και αυτός των εκτυπώσεων. Οι υπόλοιποι κλάδοι εκπροσωπούνται από πολύ μικρότερο πλήθος εταιριών.

iv. Περιοχή Δραστηριοποίησης

Πίνακας 2.2-5 Περιγραφικά Στατιστικά μεταβλητής "Περιοχή Δραστηριοποίησης"

Περιοχή Δραστηριοποίησης	Πλήθος Εταιριών	Ποσοστό επί του Πλήθους
ΑΤΤΙΚΗ	2685	41,04%
ΜΑΚΕΔΟΝΙΑ	1742	26,62%
ΣΤΕΡΕΑ ΕΛΛΑΔΑ	598	9,14%
ΠΕΛΟΠΟΝΝΗΣΟΣ	544	8,31%
ΚΕΝΤΡΙΚΗ ΕΛΛΑΔΑ	464	7,09%
ΚΡΗΤΗ	316	4,83%
ΝΗΣΙΑ	195	2,97%
Γενικό Άθροισμα Εταιριών	6543	100%

Παρατηρούμε πως σχεδόν το 40% των εταιριών που εξετάζονται δραστηριοποιούνται στην Αττική ενώ το 26% στην Μακεδονία. Τα ποσοστά είναι αναμενόμενα από τη στιγμή που οι δύο αυτές περιοχές διαθέτουν τη μεγαλύτερη βιομηχανική ζώνη της επικράτειας. Για τη διευκόλυνση της ανάλυσης και για την καλύτερη ερμηνεία των αποτελεσμάτων, ως προς αυτό το ποιοτικό χαρακτηριστικό, θα δημιουργηθεί νέα μεταβλητή στα δεδομένα με την ονομασία «Γεωγραφικός Τομέας». Η μεταβλητή αυτή θα έχει 3 επίπεδα: στο πρώτο επίπεδο θα περιλαμβάνονται οι περιοχές που βρίσκονται εντός Αττικής, στο δεύτερο επίπεδο οι περιοχές που βρίσκονται στο χώρο της Μακεδονίας και στο τρίτο επίπεδο οι περιοχές που βρίσκονται στην υπόλοιπη Ελλάδα.

v. Κατάσταση Επιχείρησης

Πίνακας 2.2-6 Περιγραφικά Στατιστικά μεταβλητής "Κατάσταση Επιχείρησης"

Κατάσταση Επιχείρησης	Πλήθος Εταιριών	Ποσοστό επί του Πλήθους
ΕΚΚΑΘΑΡΙΣΗ	3	0,05%
ΠΤΩΧΟΣ	43	0,66%
ΔΙΑΚΟΠΗ	103	1,57%
ΕΝΕΡΓΗ	6394	97,72%
Γενικό Άθροισμα Εταιριών	6543	100%

Παρατηρούμε πως οι επιχειρήσεις που δεν είναι ενεργές, δηλαδή η κατάστασή τους είναι είτε σε «Διακοπή», είτε σε «Εκκαθάριση» είτε σε «Πτώχευση», είναι συνολικά 149 δηλαδή το

2,27% του συνολικού πλήθους. Η μεταβλητή «Κατάσταση Επιχείρησης» αποτελεί βασικό στοιχείο της Πολυμεταβλητής Ανάλυσης που θα επακολουθήσει καθώς αποτελεί την «απόκριση» για την κατάσταση βιωσιμότητας της επιχείρησης. Ειδικότερα, οι 149 επιχειρήσεις που δεν είναι ενεργές θα αποτελέσουν σημείο αναφοράς των αλγορίθμων που θα χρησιμοποιηθούν στην ανάλυση.

Παρακάτω δίνεται πίνακας που παρουσιάζει τη διαθεσιμότητα των οικονομικών στοιχείων.

Πίνακας 2.2-7 Διαθεσιμότητα Δεδομένων ως προς τα Έτη Αναφοράς

Διαθεσιμότητα Δεδομένων για τα παρακάτω Συνεχόμενα Έτη	Πλήθος Εταιριών	Ποσοστό επί του Πλήθους
2011	546	8,34%
2011, 2012	788	12,04%
2011, 2012, 2013	1005	15,36%
2011, 2012, 2013, 2014	4204	64,25%

Παρατηρούμε ότι σχεδόν για το 65% των εταιριών διατίθεται οικονομικές πληροφορίες και για τα 4 έτη αναφοράς.

2.2.2 Παρουσίαση Ποσοτικών Μεταβλητών: Μεγέθη

Στη συγκεκριμένη ενότητα θα πραγματοποιηθεί χρηματοοικονομική ανάλυση των πρωτογενών και των δευτερογενών οικονομικών μεγεθών που χρησιμοποιούνται στην εργασία.

Πρωτογενή Μεγέθη

1. Πάγιο Ενεργητικό

Το «Πάγιο Ενεργητικό» αποτελεί στοιχείο του Ισολογισμού και ανήκει στους λογαριασμούς του Ενεργητικού. Το μέγεθος αυτό αναφέρεται στα εξής: στο σύνολο των ασώματων ακινητοποιήσεων (εμπορικά και βιομηχανικά σήματα, ευρεσιτεχνίες, δικαιώματα πνευματικής ιδιοκτησίας, πελατεία, φήμη), των ενσώματων ακινητοποιήσεων (οικόπεδα, κτίρια, μηχανήματα, μεταφορικά μέσα, υλικοτεχνικός εξοπλισμός κτλ.), των εξόδων πολυετούς απόσβεσης (δαπάνες που πραγματοποιούνται για την ίδρυση και την οργάνωση της επιχείρησης), και των συμμετοχών της επιχείρησης σε άλλες επιχειρήσεις. Όλα τα στοιχεία του ενσώματου παγίου ενεργητικού, εκτός της γης, υπόκεινται σε αποσβέσεις για τη διάρκεια της ωφέλιμης ζωής τους. Αποτελεί εξαιρετικά σημαντικό μέγεθος για την παραγωγική δυναμικότητα της επιχείρησης. Ο τύπος του συνοπτικά είναι ο ακόλουθος:

Πάγιο Ενεργητικό

$$= \text{Ασώματες Ακινήτοποιήσεις} + \text{Ενσώματες Ακινήτοποιήσεις} \\ + \text{Έξοδα Πολυετούς Απόσβεσης} + \text{Συμμετοχές}$$

2. Λοιπό Ενεργητικό

Το «Λοιπό Ενεργητικό» αποτελεί στοιχείο του Ισολογισμού και ανήκει στους λογαριασμούς του Ενεργητικού. Το μέγεθος αυτό αναφέρεται στο άθροισμα του ποσού του κυκλοφορούντος ενεργητικού (αποθέματα υλικών αξιών, βραχυπρόθεσμες απαιτήσεις και χρηματικά διαθέσιμα) και των μεταβατικών λογαριασμών του ενεργητικού (έξοδα επόμενων χρήσεων κ.α.). Ο τύπος του συνοπτικά είναι ο ακόλουθος:

$$\text{Λοιπό Ενεργητικό} = \text{Κυκλοφορούν Ενεργητικό} + \text{Μεταβατικοί Λογαριασμοί Ενεργητικού}$$

3. Σύνολο Ενεργητικού

Το «Σύνολο του Ενεργητικού» αποτελεί στοιχείο του Ισολογισμού και ανήκει στους λογαριασμούς του Ενεργητικού. Το μέγεθος αυτό αναφέρεται στο άθροισμα του παγίου και του λοιπού ενεργητικού. Ο τύπος του συνοπτικά είναι ο ακόλουθος:

$$\text{Σύνολο Ενεργητικού} = \text{Πάγιο Ενεργητικό} + \text{Λοιπό Ενεργητικό}$$

4. Ίδια Κεφάλαια

Τα «Ίδια Κεφάλαια» αποτελούν στοιχείο του Ισολογισμού και ανήκουν στους λογαριασμούς του Παθητικού. Το μέγεθος αυτό αναφέρεται στην καθαρή περιουσία της επιχείρησης και αποτελείται από το άθροισμα των μετοχικών κεφαλαίων, των αποθεματικών (μη διανεμηθέντα κέρδη στους μετόχους) και των καθαρών αποτελεσμάτων της χρήσης (είτε αυτά είναι κέρδη είτε ζημιές). Για το λόγο αυτό μπορεί η μεταβλητή να λάβει και αρνητικές τιμές. Ο τύπος του συνοπτικά είναι ο ακόλουθος:

$$\text{Ίδια Κεφάλαια} = \text{Μετοχικό Κεφάλαιο} + \text{Αποθεματικά} + \text{Αποτελέσματα χρήσης}$$

5. Μ / Μ Χρέος (Μακροπρόθεσμο Χρέος)

Το «Μακροπρόθεσμο Χρέος» αποτελεί στοιχείο του Ισολογισμού και ανήκει στους λογαριασμούς του Παθητικού. Το μέγεθος αυτό αναφέρεται στο σύνολο των δανείων μακροπρόθεσμης διάρκειας (τραπεζικά δάνεια, ομολογίες) που έχει συνάψει η επιχείρηση με τους πιστωτές της. Οι υποχρεώσεις αυτές λήγουν μετά το πέρας της επόμενης χρήσης.

6. Βραχυπρόθεσμο Χρέος

Το «Βραχυπρόθεσμο Χρέος» αποτελεί στοιχείο του Ισολογισμού και ανήκει στους λογαριασμούς του Παθητικού. Το μέγεθος αυτό αναφέρεται στο σύνολο των βραχυπρόθεσμων υποχρεώσεων της επιχείρησης και αφορά τις πιστώσεις προμηθευτών, τους πληρωτέους φόρους και τέλη, τα βραχυπρόθεσμα δάνεια, τα πληρωτέα μερίσματα, τις προκαταβολές πελατών, τα πληρωτέα γραμμάτια, τις εισφορές σε ασφαλιστικούς οργανισμούς κ.α. Οι υποχρεώσεις αυτές λήγουν το αργότερο μέχρι το πέρας της επόμενης χρήσης.

7. Σύνολο Παθητικού

Το «Σύνολο του Παθητικού» αποτελεί στοιχείο του Ισολογισμού και ανήκει στους λογαριασμούς του Παθητικού. Το μέγεθος αυτό αναφέρεται στο σύνολο των υποχρεώσεων της επιχείρησης, δηλαδή αυτές που αφορούν τους μετόχους και τους πιστωτές της, και ισούται με το άθροισμα των ιδίων κεφαλαίων και των βραχυπρόθεσμων και μακροπρόθεσμων δανείων. Το συνολικό ποσό του Παθητικού είναι πάντα ίσο με το συνολικό ποσό του Ενεργητικού σε κάθε ισολογισμό. Για το λόγο αυτό, η μεταβλητή «Συνολικό Παθητικό», που λαμβάνεται από το *inp.gr*, είναι πλεονάζουσα και θα *διαγραφεί* από το σύνολο των δεδομένων. Σε οποιονδήποτε υπολογισμό της εργασίας απαιτηθεί η χρησιμοποίηση του μεγέθους του Παθητικού (ή αλλιώς του Συνόλου των Υποχρεώσεων της επιχείρησης) θα χρησιμοποιείται αντί αυτού το μέγεθος «Συνολικό Ενεργητικό». Ο τύπος του συνοπτικά είναι ο ακόλουθος:

$$\begin{aligned} \text{Συνολικό Παθητικό} &= \text{Ίδια Κεφάλαια} + \text{Μακροπρόθεσμα Χρέη} + \text{Βραχυπρόθεσμα Χρέη} \\ \text{Σύνολο Υποχρεώσεων} &= \text{Συνολικό Παθητικό} = \text{Συνολικό Ενεργητικό} \end{aligned}$$

8. Πωλήσεις

Οι «Πωλήσεις» αποτελούν στοιχείο της Κατάστασης Αποτελεσμάτων Χρήσης. Το μέγεθος αυτό αναφέρεται στον κύκλο εργασιών της επιχείρησης ή αλλιώς στον τζίρο της.

9. Μεικτό Κέρδος

Το «Μεικτό Κέρδος» αποτελεί στοιχείο της Κατάστασης Αποτελεσμάτων Χρήσης. Το μέγεθος αυτό αποτελεί το ποσό που απομένει αν από τις πωλήσεις αφαιρεθεί το κόστος των πωληθέντων προϊόντων. Το κόστος των πωληθέντων είναι σημαντικό μέγεθος και θα αναλυθεί ως έννοια παρακάτω καθώς αποτελεί δευτερογενές μέγεθος. Το μεικτό κέρδος αποτελεί το οργανικό αποτέλεσμα της επιχείρησης και δεν υπολογίζει λειτουργικά κόστη, έκτακτα κέρδη, άλλα έσοδα, ζημιές και άλλα έξοδα. Ο τύπος του συνοπτικά είναι ο ακόλουθος:

$$\text{Μεικτό Κέρδος} = \text{Πωλήσεις} - \text{Κόστος Πωληθέντων}$$

10. EBITDA (*Earnings Before Interest, Taxes, Depreciation and Amortization*)

Ο «EBITDA» αποτελεί στοιχείο της Κατάστασης Αποτελεσμάτων Χρήσης. Το μέγεθος αυτό αποτελεί τα **λειτουργικά κέρδη** της επιχείρησης και προκύπτει αν αφαιρεθούν από τα μεικτά κέρδη τα λειτουργικά έξοδα της επιχείρησης δηλαδή τα έξοδα λειτουργίας της διοίκησης, της

διάθεσης (διαφημίσεις), της έρευνας και ανάπτυξης και της χρηματοοικονομικής λειτουργίας. Ο τύπος του συνοπτικά είναι ο ακόλουθος:

$$EBITDA = \text{Μεικτό Κέρδος} - \text{Λειτουργικά Έξοδα}$$

11. EBIT (*Earnings Before Interest, Taxes*)

Ο «EBIT» αποτελεί στοιχείο της Κατάστασης Αποτελεσμάτων Χρήσης. Το μέγεθος αυτό αποτελεί τα **μερικά αποτελέσματα εκμετάλλευσης** της επιχείρησης δηλαδή τα κέρδη που προκύπτουν αν αφαιρεθούν από τον EBITDA οι αποσβέσεις των παγίων. Ο τύπος του συνοπτικά είναι ο ακόλουθος:

$$EBIT = EBITDA - \text{Αποσβέσεις Παγίων}$$

12. Κέρδη προ Φόρων

Τα «Κέρδη προ Φόρων» αποτελούν στοιχείο της Κατάστασης Αποτελεσμάτων Χρήσης. Το μέγεθος αυτό αποτελεί το καθαρό αποτέλεσμα της χρήσης πριν αφαιρεθούν οι φόροι. Ο τύπος του συνοπτικά είναι ο ακόλουθος:

Κέρδη προ Φόρων

$$= EBIT - \text{Χρεωστικοί Τόκοι} - \text{Έκτακτα Έξοδα} + \text{Πιστωτικοί Τόκοι} \\ + \text{Έκτακτα Έσοδα}$$

13. Φόροι

Οι «Φόροι» αποτελούν στοιχείο της Κατάστασης Αποτελεσμάτων Χρήσης. Το μέγεθος αυτό εκφράζει τον φόρο που αναλογεί στα κέρδη προ φόρων όταν αυτά είναι θετικά. Υπάρχουν περιπτώσεις στις οποίες τα κέρδη προ φόρων είναι μεν θετικά όμως ο φόρος που αναλογεί σε αυτά εμφανίζεται ως μηδέν ή μικρότερος από τον φόρο που θα αντιστοιχούσε σύμφωνα με τη νομοθεσία. Στις περιπτώσεις αυτές, η επιχείρηση, είτε μεταφέρει στα αποτελέσματά της ζημιές παρελθουσών χρήσεων (έως και 5 προηγούμενες χρήσεις) οι οποίες συμψηφίζονται με το τωρινό φορολογητέο αποτέλεσμα και δεν της επιβάλλεται φόρος, είτε έχει υπάρξει προκαταβολή φόρου του προηγούμενου έτους που συμψηφίζεται με το φόρο που αντιστοιχεί στην τρέχουσα χρήση.

14. Καθαρό Κέρδος

Το «Καθαρό Κέρδος» αποτελεί στοιχείο της Κατάστασης Αποτελεσμάτων Χρήσης. Το μέγεθος εκφράζει το αποτέλεσμα που προκύπτει αν από τα κέρδη προ φόρων αφαιρεθούν οι φόροι. Αποτελεί ουσιαστικά τα κέρδη προς διάθεση στους μετόχους και στα αποθεματικά της επιχείρησης.

$$\text{Καθαρό Κερδος} = \text{Κέρδη προ Φόρων} - \text{Φόροι}$$

Στη συνέχεια παρουσιάζονται τα δευτερογενή μεγέθη όπως προκύπτουν με τη βοήθεια των πρωτογενών μεγεθών.

Δευτερογενή Μεγέθη

15. Κόστος των Πωληθέντων Προϊόντων

Το «Κόστος των Πωληθέντων Προϊόντων» αντιπροσωπεύει το κόστος των αγαθών που πουλήθηκαν από την επιχείρηση κατά τη διάρκεια της επιχειρηματικής της δραστηριότητας. Αποτελεί εξαιρετικά σημαντικό μέγεθος καθώς προσδιορίζει το χρηματικό ποσό που κατέβαλε η επιχείρηση για να αποκτήσει τα εμπορεύματά της. Το κόστος αγοράς αποτελείται από την τιμή αγοράς της μετρητοίς (+), τα άμεσα ή ειδικά έξοδα αγορών (+), τις εκπτώσεις των αγορών (-) και τις επιστροφές των αγορών (-) (Ευρετήριο, 2015). Αν από τις πωλήσεις αφαιρεθεί το κόστος των πωληθέντων προκύπτει το μεικτό κέρδος. Συνεπώς, από τα δεδομένα που διαθέτουμε το κόστος των πωληθέντων ισούται με:

$$\text{Κόστος Πωληθέντων} = \text{Πωλήσεις} - \text{Μεικτό Κέρδος}$$

16. Λειτουργικά Έξοδα

Τα «Λειτουργικά Έξοδα» αποτελούν τα έξοδα που οφείλονται στις συνήθεις δραστηριότητες της επιχείρησης. Στο σημείο αυτό θα υπολογιστούν τα λειτουργικά έξοδα που οφείλονται στις λειτουργίες της διοίκησης, της διάθεσης, της έρευνας και ανάπτυξης και των επενδύσεων. Παραδείγματα τέτοιων εξόδων αποτελούν οι αμοιβές προσωπικού, τα ενοίκια, τα έξοδα ηλεκτρισμού – ύδρευσης, των διαφημίσεων, των αναλωσίμων κ.α. Από τα δεδομένα που διαθέτουμε τα λειτουργικά έξοδα ισούται με:

$$\text{Λειτουργικά Έξοδα} = \text{Μεικτό Κέρδος} - \text{EBITDA}$$

17. Καθαρό Κεφάλαιο Κίνησης

Το «Καθαρό Κεφάλαιο Κίνησης» αποτελεί το ποσό των κυκλοφοριακών στοιχείων της επιχείρησης που απομένει, εάν υποθεθεί ότι εξοφλούνται όλες οι βραχυχρόνιες υποχρεώσεις. Το μέγεθος αυτό απεικονίζει το βαθμό που η επιχείρηση ανταποκρίνεται: α) στην πληρωμή των τρεχουσών υποχρεώσεών της, των σταθερών δαπανών, των απαιτητών τόκων και μερισμάτων και β) στην απορρόφηση τυχόν τρεχουσών ζημιών και καθυστερήσεων εισπράξεων μεγάλων ποσών απαιτήσεων (Γκούμας & Λεκαράκου, 2013). Προκύπτει αν από το σύνολο των στοιχείων του κυκλοφορούντος ενεργητικού αφαιρεθούν οι βραχυπρόθεσμες υποχρεώσεις. Το κυκλοφορούν ενεργητικό προκύπτει αν από το σύνολο του ενεργητικού αφαιρεθεί το πάγιο ενεργητικό και οι μεταβατικοί λογαριασμοί του ενεργητικού. Οι μεταβατικοί λογαριασμοί αποτελούν το μέσο μετάβασης στην επόμενη χρήση των εσόδων ή εξόδων που έχουν προεισπραχθεί ή προκαταβληθεί αντίστοιχα στην τρέχουσα χρήση. Συνήθως οι μεταβατικοί λογαριασμοί του ενεργητικού περιέχουν ποσά μικρής αξίας και δημιουργούνται με σκοπό την χρονική τακτοποίηση των εσόδων και των εξόδων. Από τα δεδομένα που διαθέτουμε δεν μπορεί να απομονωθεί το ποσό των μεταβατικών λογαριασμών για αυτό και για τον υπολογισμό του καθαρού κεφαλαίου κίνησης θα θεωρηθεί ως κυκλοφορούν ενεργητικό το «Λοιπό Ενεργητικό», το οποίο περιλαμβάνει και τους μεταβατικούς λογαριασμούς. Συνεπώς, από τα δεδομένα που διαθέτουμε το καθαρό κεφάλαιο κίνησης ισούται με:

$$\text{Καθαρό Κεφάλαιο Κίνησης} = \text{Λοιπό Ενεργητικό} - \text{Βραχυπρόθεσμο Χρέος}$$

18. Αποσβέσεις

Οι «Αποσβέσεις» αφορούν το ποσό των αποσβέσεων των παγίων στοιχείων του ενεργητικού. Οι αποσβέσεις υπολογίζονται από τον τύπο:

$$\text{Αποσβέσεις} = \text{EBITDA} - \text{EBIT}$$

19. Ξένα Κεφάλαια

Τα «Ξένα Κεφάλαια» αποτελούν το σύνολο των κεφαλαίων που έχουν δανείσει στην επιχείρηση οι πιστωτές της. Συνεπώς, αποτελούνται από το σύνολο των μακροπρόθεσμων και των βραχυπρόθεσμων δανείων. Το μέγεθος υπολογίζεται από τον τύπο:

$$\text{Ξένα Κεφάλαια} = \text{Μακροπρόθεσμο Χρέος} + \text{Βραχυπρόθεσμο Χρέος}$$

20. Τόκοι Χρεωστικοί

Οι «Τόκοι Χρεωστικοί» αφορούν το σύνολο των τόκων που η επιχείρηση κατέβαλε στους πιστωτές της την τρέχουσα χρήση. Το μέγεθος αυτό υπολογίζεται από τον τύπο:

$$\text{Τόκοι Χρεωστικοί} = \text{Κέρδη προ φόρων και τόκων (EBIT)} - \text{Κέρδη προ φόρων}$$

Στη συνέχεια ακολουθεί η παρουσίαση όλων των αριθμοδεικτών που θα χρησιμοποιηθούν στην εργασία.

2.2.3 Παρουσίαση Ποσοτικών Μεταβλητών: Αριθμοδείκτες

Στη συγκεκριμένη ενότητα θα παρουσιασθεί ο τύπος και η χρηματοοικονομική ανάλυση των αριθμοδεικτών που χρησιμοποιούνται στην ανάλυση. Τα είδη των αριθμοδεικτών που χρησιμοποιούνται είναι τα παρακάτω:

- Αριθμοδείκτες αποδοτικότητας
- Αριθμοδείκτες κεφαλαιακής διάρθρωσης
- Αριθμοδείκτες δραστηριότητας
- Αριθμοδείκτες δαπανών λειτουργίας
- Αριθμοδείκτες ρευστότητας.

Αριθμοδείκτες Αποδοτικότητας

Οι αριθμοδείκτες αποδοτικότητας (*Profitability ratios*) εκφράζουν την ικανότητα της επιχείρησης να πραγματοποιεί κέρδη. Οι αριθμοδείκτες αυτοί απεικονίζουν τις σχέσεις κερδών και απασχολουμένων κεφαλαίων στην επιχείρηση καθώς επίσης και τις σχέσεις κερδών και πωλήσεων (Βασιλείου & Ηρειώτης, 2008). Οι αριθμοδείκτες που θα χρησιμοποιηθούν από την κατηγορία αυτή είναι οι ακόλουθοι:

1. Μεικτό Περιθώριο Κέρδους (*Gross Profit Margin*)

Ο αριθμοδείκτης μεικτού περιθωρίου κέρδους δίνεται από τον τύπο:

$$\frac{\text{Μεικτό Κέρδος}}{\text{Πωλήσεις}} = 1 - \frac{\text{Κόστος Πωληθέντων}}{\text{Πωλήσεις}}$$

Ο δείκτης απεικονίζει το μεικτό κέρδος της επιχείρησης για κάθε ένα ευρώ καθαρών πωλήσεων που πραγματοποιεί. Εκφράζει τη λειτουργική αποτελεσματικότητα μιας οικονομικής μονάδας καθώς και την τιμολογιακή πολιτική της, δηλαδή την πρόσθετη αξία που έχουν οι πωλήσεις έναντι του κόστους τους (Δασίλας, 2012). Ένας υψηλός δείκτης μεικτού περιθωρίου κέρδους παρέχει ενδείξεις ικανότητας της διοίκησης να πραγματοποιεί φθηνές αγορές και να πωλεί σε υψηλές τιμές. Αντίθετα ένας χαμηλός δείκτης δεν υποδεικνύει καλή πολιτική της διοίκησης στον τομέα αγορών και πωλήσεων (Παπαναστασόπουλος, 2014).

2. Περιθώριο EBITDA (*EBITDA Margin*)

Ο αριθμοδείκτης περιθωρίου EBITDA δίνεται από τον τύπο:

$$\frac{EBITDA}{Πωλήσεις}$$

Ο δείκτης παρουσιάζει τη λειτουργική κερδοφορία της επιχείρησης υποδεικνύοντας την ανάπτυξη και το πραγματικό κέρδος των επιχειρήσεων πριν όμως αφαιρεθούν οι αποσβέσεις, οι τόκοι και οι φόροι (Ependysopedia, 2014).

3. Περιθώριο EBIT (*EBIT Margin*)

Ο αριθμοδείκτης περιθωρίου EBIT δίνεται από τον τύπο:

$$\frac{EBIT}{Πωλήσεις}$$

Ο δείκτης είναι ανάλογος του προηγούμενου, ενδείκνυται κυρίως για συγκρίσεις μεταξύ ομοειδών επιχειρήσεων και παρέχει ενδείξεις για την πορεία ανάπτυξης της επιχείρησης. Ο EBIT αποτελεί μέγεθος που παρακολουθείται στενά από τους πιστωτές της επιχείρησης καθώς αντιπροσωπεύει το ποσό των μετρητών που η επιχείρηση θα είναι σε θέση να χρησιμοποιήσει για να εξοφλήσει τις οφειλές της προς αυτούς (InvestorWords, 2015).

4. Καθαρό Περιθώριο Κέρδους (*Net Profit Margin*)

Ο αριθμοδείκτης καθαρού περιθωρίου κέρδους δίνεται από τον τύπο:

$$\frac{Καθαρό Κέρδος}{Πωλήσεις}$$

Ο δείκτης εκφράζει το καθαρό κέρδος της επιχείρησης για κάθε ένα ευρώ πωλήσεων που αυτή πραγματοποιεί. Όσο μεγαλύτερος είναι ο συγκεκριμένος αριθμοδείκτης τόσο πιο επικερδής είναι η οικονομική μονάδα. Σε συνδυασμό με το μεικτό περιθώριο κέρδους η ανάλυση του καθαρού κέρδους δίνει πολλά χρήσιμα συμπεράσματα. Σε ένα συγκεκριμένο διάστημα ετών, αν το καθαρό περιθώριο κέρδους μειώνεται ενώ το μεικτό περιθώριο παραμένει σταθερό τότε ο ρυθμός αύξησης των εξόδων που δεν συμπεριλαμβάνονται στο κόστος

πωληθέντων είναι μεγαλύτερος από την αύξηση των πωλήσεων. Αντίθετα, αν το καθαρό περιθώριο κέρδους παραμένει σταθερό και το μεικτό περιθώριο κέρδους μειώνεται τότε το κόστος παραγωγής αυξήθηκε περισσότερο από τις πωλήσεις (Βασιλείου & Ηρειώτης, 2008).

5. Περιθώριο Αποσβέσεων (*Ratio of Depreciation to Net Sales*)

Ο αριθμοδείκτης περιθωρίου αποσβέσεων δίνεται από τον τύπο:

$$\frac{\text{Αποσβέσεις}}{\text{Πωλήσεις}}$$

Ο δείκτης είναι σημαντικός όταν οι αποσβέσεις είναι ανάλογες με το βαθμό απασχόλησης των παγίων. Καθώς τα μεγέθη αφορούν βιομηχανίες, που οι πωλήσεις τους στηρίζονται κατά κόρον στη λειτουργία των παγίων, ο δείκτης αυτός κρίνεται σημαντικός. Οι αποσβέσεις αποτελούν στοιχείο κόστους λειτουργίας τα επιχείρησης και δεν απαιτούν τη χρήση κεφαλαίων κίνησης. Το μέγεθος των αποσβέσεων ενδέχεται να διαφέρει από έτος σε έτος, τόσο για την ίδια οικονομική μονάδα όσο και ανάμεσα στις οικονομικές μονάδες του κλάδου, εξαιτίας ποικίλων παραγόντων όπως: η σύνθεση και η ηλικία των παγίων, οι δαπάνες συντήρησης και επισκευών κ.α. (Βασιλείου & Ηρειώτης, 2008).

6. Λοιπό Ενεργητικό προς Πωλήσεις

Ο αριθμοδείκτης λοιπού ενεργητικού προς πωλήσεις δίνεται από τον τύπο:

$$\frac{\text{Λοιπό Ενεργητικό}}{\text{Πωλήσεις}}$$

Ο δείκτης εκφράζει τα κυκλοφορούντα στοιχεία που αντιστοιχούν σε κάθε ένα ευρώ πωλήσεων που πραγματοποιεί. Η διαχρονική μεταβολή του είναι σημαντική στην αξιολόγηση της πιστοληπτικής ικανότητας της επιχείρησης (ICAP GROUP A.E, Αύγουστος 2013).

7. Ενεργητικό προς Πωλήσεις

Ο αριθμοδείκτης ενεργητικού προς πωλήσεις δίνεται από τον τύπο:

$$\frac{\text{Ενεργητικό}}{\text{Πωλήσεις}}$$

Ο δείκτης εκφράζει το ποσό των συνολικών επενδυμένων κεφαλαίων που αντιστοιχούν σε κάθε ένα ευρώ πωλήσεων που πραγματοποιεί η επιχείρηση. Η διαχρονική μεταβολή του είναι σημαντική στην αξιολόγηση της πιστοληπτικής ικανότητας της επιχείρησης (ICAP GROUP A.E, Αύγουστος 2013).

8. EBITDA προς Ενεργητικό

Ο αριθμοδείκτης EBITDA προς ενεργητικό δίνεται από τον τύπο:

$$\frac{EBITDA}{\text{Ενεργητικό}}$$

Ο δείκτης μετρά πόσο αποτελεσματικά μια εταιρεία παράγει τα κέρδη EBITDA. Αυτό σημαίνει ότι η διάρθρωση του κεφαλαίου, οι διαφορετικοί φορολογικοί συντελεστές, και το διαφορετικό κόστος των κεφαλαιακών δαπανών δεν επηρεάζει τις συγκρίσεις μεταξύ των εταιρειών. Ο δείκτης αυτός έχει περιορισμένη χρησιμότητα για τους περισσότερους επενδυτές (Faulkenberry, 2015).

9. Αριθμοδείκτης Βασικής Ικανότητας Κερδών (*Basic Earnings Power*)

Ο αριθμοδείκτης βασικής ικανότητας κερδών δίνεται από τον τύπο:

$$\frac{EBIT}{\text{Ενεργητικό}}$$

Η αριθμοδείκτης αυτός μετρά τα κέρδη της επιχείρησης ως προς το σύνολο της περιουσίας της πριν όμως πληρωθούν οι φορολογικές της υποχρεώσεις και πριν ικανοποιηθούν οι πιστωτές της. Όσο μεγαλώνει η τιμή που λαμβάνει ο δείκτης τόσο πιο επικερδής είναι η στρατηγική και η πορεία της επιχείρησης.

10. Απόδοση Ενεργητικού προ Φόρων

Ο αριθμοδείκτης απόδοσης ενεργητικού προ φόρων δίνεται από τον τύπο:

$$\frac{\text{Κέρδη προ Φόρων}}{\text{Ενεργητικό}}$$

Ο δείκτης μετρά πόσο αποτελεσματικά μια εταιρεία παράγει κέρδη προ φόρων σύμφωνα με τα περιουσιακά της στοιχεία. Για τον υπολογισμό των κερδών προ φόρων έχει ληφθεί υπόψη το κόστος των κεφαλαιακών δαπανών.

11. Απόδοση Ενεργητικού ή Λειτουργίας μετά Φόρων (*Return on Asset*)

Ο αριθμοδείκτης απόδοσης ενεργητικού δίνεται από τον τύπο:

$$\frac{\text{Καθαρό Κέρδος}}{\text{Ενεργητικό}}$$

Ο δείκτης υποδεικνύει την καθαρή απόδοση του ενεργητικού της επιχείρησης και αντανακλά την ικανότητα της διοίκησης να αξιοποιεί αποτελεσματικά τους οικονομικούς της πόρους, δηλαδή τα ίδια και τα ξένα επενδυμένα κεφάλαια, για τη δημιουργία καθαρών κερδών. Η απόδοση ενεργητικού μιας επιχείρησης εξαρτάται από δύο κυρίως παράγοντες, το περιθώριο καθαρού κέρδους (Καθαρά Κέρδη / Πωλήσεις) και την κυκλοφοριακή ταχύτητα ενεργητικού (Καθαρές Πωλήσεις / Σύνολο Ενεργητικού) (Δασίλας, 2012).

12. Αποδοτικότητα Απασχολούμενων Κεφαλαίων (*Return to Total Capital Employed*)

Ο αριθμοδείκτης αποδοτικότητας απασχολούμενων κεφαλαίων δίνεται από τον τύπο:

$$\frac{\text{Καθαρά Κέρδη} + \text{Χρεωστικοί Τόκοι}}{\text{Ενεργητικό}}$$

Ο δείκτης εκφράζει την κερδοφόρα δυναμικότητα των συνολικών απασχολούμενων κεφαλαίων της επιχείρησης, δηλαδή κατά πόσο τα κεφάλαιά της, ανεξάρτητα από την προέλευσή τους, αξιοποιούνται αποδοτικά. Ο δείκτης αυτός είναι πολύ σημαντικός για την επιχείρηση για τους εξής λόγους: α) δίνει κατευθυντήρια γραμμή σε περίπτωση που η επιχείρηση επιθυμεί να εξαγοράσει άλλη ή άλλες επιχειρήσεις και γενικά να αναλάβει νέες δραστηριότητες, β) μηδενίζεται στην περίπτωση που η επιχείρηση βρίσκεται σε κρίση, γ) μια μόνιμα χαμηλή τιμή του ενδεχομένως να αποτελεί ένδειξη διακοπής της δραστηριότητας της επιχείρησης, δ) τέλος αν είναι χαμηλότερος από το κόστος των δανειακών κεφαλαίων (δηλαδή το επιτόκιο δανεισμού) θα έχει ως αποτέλεσμα μία αύξηση των δανείων να προκαλέσει μείωση των κερδών ανά μετοχή της επιχείρησης (Βασιλείου & Ηρειώτης, 2008).

13. Αποδοτικότητα Ιδίων Κεφαλαίων

Ο αριθμοδείκτης αποδοτικότητας ιδίων κεφαλαίων δίνεται από τον τύπο:

$$\frac{EBIT}{\text{Ίδια Κεφάλαια}}$$

Ο δείκτης εκφράζει το κέρδος EBIT που μπορεί να παράγει η επιχείρηση χρησιμοποιώντας τα κεφάλαια των φορέων της, δηλαδή τα μετοχικά κεφάλαια και τα αποθεματικά. Πολλές φορές, μέσα στη διάρκεια της χρήσης, μπορεί να πραγματοποιηθεί αύξηση των ιδίων κεφαλαίων της επιχείρησης, όπως για παράδειγμα με αύξηση του μετοχικού κεφαλαίου με μετρητά ή με αναπροσαρμογή της αξίας των περιουσιακών στοιχείων. Αυτή η πληροφορία δεν υπάρχει στα δεδομένα της εργασίας αυτής οπότε θα λαμβάνεται υπόψη η εικόνα των ιδίων κεφαλαίων όπως δηλώθηκαν στον ισολογισμό (Βασιλείου & Ηρειώτης, 2008).

14. Απόδοση Ιδίων Κεφαλαίων προ Φόρων

Ο αριθμοδείκτης απόδοσης ιδίων κεφαλαίων προ φόρων δίνεται από τον τύπο:

$$\frac{\text{Κέρδη προ Φόρων}}{\text{Ίδια Κεφάλαια}}$$

Ο δείκτης είναι ανάλογος του προηγούμενου. Χαμηλές τιμές του δείκτη μπορεί να είναι ενδεικτικές κάποιου προβλήματος που αντιμετωπίζει η επιχείρηση χωρίς όμως να μπορεί να γίνει άμεσα αντιληπτή η πηγή του προβλήματος. Τέτοια προβλήματα θα μπορούσαν να είναι η ανεπαρκής διοίκηση, η χαμηλή παραγωγικότητα και οι δυσμενείς οικονομικές συνθήκες. Αντίθετα, υψηλές τιμές υποδεικνύουν οικονομική ευημερία και εύστοχη χρησιμοποίηση των κεφαλαίων. Αυτό σημαίνει ότι το κόστος δανεισμού είναι μικρότερο από την απόδοση των κεφαλαίων (Γκούμας & Λεκαράκου, 2013).

15. Καθαρή Απόδοση Ιδίων Κεφαλαίων (*Return on net worth*)

Ο αριθμοδείκτης καθαρής απόδοσης ιδίων κεφαλαίων δίνεται από τον τύπο:

$$\frac{\text{Καθαρά Κέρδη}}{\text{Ίδια Κεφάλαια}}$$

Ο δείκτης αποτελεί μέτρο της κερδοφόρας δυναμικότητας των ιδίων κεφαλαίων της επιχείρησης. Μια υψηλή τιμή του παρέχει ενδείξεις για επάρκεια ποσού των ιδίων κεφαλαίων, για αποτελεσματική χρησιμοποίηση τους στην πραγματοποίηση κερδών, και για ύπαρξη ικανοποιητικού περιθωρίου ασφαλείας σε περίπτωση ανεπιθύμητων εξελίξεων στα κέρδη. Οι επενδυτές προτιμούν εταιρίες με υψηλή και αυξανόμενη απόδοση ιδίων κεφαλαίων. «Μια εταιρεία μπορεί να δημιουργήσει αξία για τους μετόχους της, εάν η απόδοση ιδίων κεφαλαίων είναι μεγαλύτερη από το κόστος ευκαιρίας των κεφαλαίων αυτών. Αν δηλαδή η αναμενόμενη απόδοση που απαιτούν οι μέτοχοι για να επενδύσουν στην εταιρία είναι υψηλότερη από την απόδοση εναλλακτικών μορφών επενδύσεων με παρόμοιο επίπεδο κεφαλαιακού κινδύνου». (Ορων, 2015) Γενικά, επιχειρήσεις με υψηλό δείκτη χρέους προς ίδια κεφάλαια θα έχουν συνήθως καλύτερη απόδοση ιδίων κεφαλαίων αν και υψηλά χρέη αυξάνουν τις πιθανότητες πτώχευσής τους (Παπαναστασόπουλος, 2014).

16. Αριθμοδείκτης Οικονομικής Μόχλευσης (*Financial Leverage Ratio*)

Ο αριθμοδείκτης οικονομικής μόχλευσης δίνεται από τον τύπο:

$$\frac{\text{Καθαρή Απόδοση Ιδίων Κεφαλαίων (Αριθμοδείκτης 13)}}{\text{Αποδοτικότητα Απασχολούμενων Κεφαλαίων (Αριθμοδείκτης 10)}}$$

Ο δείκτης εκφράζει εμμέσως τον τρόπο που επιδρούν τα ξένα κεφάλαια στα κέρδη της επιχείρησης. Η επίδραση των δανειακών κεφαλαίων στα κέρδη είναι θετική και επωφελής, όταν η αποδοτικότητα των ιδίων κεφαλαίων είναι μεγαλύτερη από την αποδοτικότητα του συνόλου των απασχολούμενων κεφαλαίων. Συνεπώς, όταν ο αριθμοδείκτης είναι μεγαλύτερος της μονάδος η χρησιμοποίηση των ξένων κεφαλαίων είναι θετική για την επιχείρηση. Αντίθετα, η επίδραση των ξένων κεφαλαίων στα κέρδη της επιχείρησης είναι αρνητική όταν ο δείκτης είναι μικρότερος της μονάδος. Αυτό σημαίνει ότι η επιχείρηση δανείζεται με επαχθείς όρους. Τέλος, όταν ο δείκτης ισούται με τη μονάδα τα ξένα κεφάλαια δεν επιδρούν στα κέρδη της επιχείρησης (Βασιλείου & Ηρειώτης, 2008).

17. Αριθμοδείκτης Απόσβεσης Παγίων (*Ratio of depreciation to fixed assets*)

Ο αριθμοδείκτης απόσβεσης παγίων δίνεται από τον τύπο:

$$\frac{\text{Αποσβέσεις}}{\text{Πάγιο Ενεργητικό + Αποσβέσεις}}$$

Ο δείκτης απεικονίζει το ποσοστό των παγίων που αποσβαίνει η επιχείρηση από κάθε μονάδα παγίων μέσα σε μία χρονική περίοδο. Καθορίζεται έτσι αν η ετήσια απόσβεση που πραγματοποιήθηκε είναι επαρκής ή όχι, και αν η επιχείρηση ακολουθεί παρόμοια πολιτική αποσβέσεων διαχρονικά ή αν επηρεάζεται από το ύψος των πραγματοποιούμενων κερδών στην εκάστοτε χρονιά. Μία διαχρονική αύξηση του αριθμοδείκτη υποδεικνύει ότι η επιχείρηση ακολουθεί μία πιο φιλελεύθερη πολιτική αποσβέσεων μειώνοντας τον πραγματικό χρόνο αποσβέσεως των παγίων στοιχείων της. Αντιθέτως, μία διαχρονική μείωση του αριθμοδείκτη υποδεικνύει ανεπάρκεια αποσβέσεων ή αλλαγή της πολιτικής αποσβέσεων που ακολουθεί η επιχείρηση (Βασιλείου & Ηρειώτης, 2008).

18. Αριθμοδείκτης Φορολογικής Επιβάρυνσης

Ο αριθμοδείκτης φορολογικής επιβάρυνσης δίνεται από τον τύπο:

$$\frac{\text{Καθαρά Κέρδη}}{\text{Κέρδη προ Φόρων}}$$

Ο δείκτης υποδεικνύει τη φορολογική μεταχείριση των επιχειρήσεων από πλευράς φορολογικής νομοθεσίας από έτος σε έτος. Τα κέρδη των επιχειρήσεων φορολογούνται κλιμακωτά και το ποσοστό φορολόγησης της κάθε κλίμακας είναι συνήθως μεταβλητό τα τελευταία χρόνια. Το ποσοστό αυτό διαφοροποιείται από επιχείρηση σε επιχείρηση εξαιτίας του διαφορετικού ποσού του προκαταβεβλημένου φόρου που μπορεί να κατέβαλε η επιχείρηση κατά τα προηγούμενα έτη και το οποίο συμψηφίζεται με το ποσό του φόρου που η επιχείρηση οφείλει να καταβάλλει στο τρέχον έτος.

19. Απόδοση Μακροπρόθεσμων Κεφαλαίων

Ο αριθμοδείκτης απόδοσης μακροπρόθεσμων κεφαλαίων δίνεται από τον τύπο:

$$\frac{\text{Καθαρά Κέρδη}}{\text{Μακροπρόθεσμο Χρέος + Ίδια Κεφάλαια}}$$

Ο δείκτης αποτελεί μέτρο της κερδοφόρας δυναμικότητας των μακροπρόθεσμων κεφαλαίων της επιχείρησης, δηλαδή και των ίδιων κεφαλαίων της αλλά και των μακροπρόθεσμων δανείων που αυτή έχει συνάψει. Μια υψηλή τιμή του δείκτη παρέχει ενδείξεις για αποτελεσματική

χρησιμοποίηση των μακροπρόθεσμων κεφαλαίων για πραγματοποίηση κερδών και για ικανοποιητικό περιθώριο ασφαλείας σε περίπτωση ανεπιθύμητων εξελίξεων στα κέρδη.

Αριθμοδείκτες Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας

Οι δείκτες Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας (*Financial structure and viability ratios*) χρησιμοποιούνται κυρίως για την ανίχνευση της δυνατότητας της επιχείρησης να ανταποκρίνεται στις μακροχρόνιες υποχρεώσεις της αλλά και για την απεικόνιση του βαθμού διασφάλισης που απολαμβάνουν οι πιστωτές της σε περιπτώσεις ανεπιθύμητων εξελίξεων. Ο δείκτης κεφαλαιακής διάρθρωσης ή μόχλευσης ή αλλιώς δείκτης χρέους υποδεικνύουν το πώς η επιχείρηση χρηματοδοτεί τις συνολικές της επενδύσεις, δηλαδή το ενεργητικό της. Η χρηματοδότηση αυτή μπορεί να συνίσταται είτε σε δανειακά κεφάλαια είτε σε ίδια κεφάλαια. Οι αριθμοδείκτες που θα χρησιμοποιηθούν από την κατηγορία αυτή είναι οι ακόλουθοι:

20. Αριθμοδείκτης Σχέσεως Κεφαλαίων (αυτονομίας) ή Αριθμοδείκτης Ιδίων προς Συνολικά Κεφάλαια (*Ratio of Owner's Equity to Total Assets*)

Ο αριθμοδείκτης σχέσεως κεφαλαίων δίνεται από τον τύπο:

$$\frac{\text{Ίδια Κεφάλαια}}{\text{Συνολικά Κεφάλαια}} = \frac{\text{Ίδια Κεφάλαια}}{\text{Ενεργητικό}}$$

Ο δείκτης εκφράζει το ποσοστό του συνόλου των περιουσιακών στοιχείων της επιχείρησης (ενεργητικό) που χρηματοδοτείται από τους φορείς της επιχείρησης (ίδια κεφάλαια). Θεωρείται πολύ σημαντικός δείκτης καθώς υποδηλώνει την οικονομική αυτάρκεια της επιχείρησης, την οικονομική δυναμικότητα της και τη μακροχρόνια ρευστότητα της. Μια υψηλή τιμή του μπορεί να υπονοεί την ύπαρξη χαμηλού δανεισμού, θετικές συνθήκες μακροχρόνιας εξέλιξης, ικανοποιητικά περιθώρια ασφαλείας για την αντιμετώπιση πιθανών μελλοντικών ζημιών, χαμηλή πιθανότητα αθέτησης της αποπληρωμής τοκοχρεολυσίων και χαμηλή πιθανότητα χρεοκοπίας (Δασίλας, 2012) (Παπαναστασόπουλος, 2014).

21. Αριθμοδείκτης Δανειακής Ικανότητας ή Φερεγγυότητας ή Δανειακής Επιβάρυνσης (Ratio of Total Liabilities to Total Assets)

Ο αριθμοδείκτης δανειακής ικανότητας δίνεται από τον τύπο:

$$\frac{\text{Ξένα Κεφάλαια}}{\text{Συνολικά Κεφάλαια}} = \frac{\text{Ξένα Κεφάλαια}}{\text{Ενεργητικό}}$$

Ο δείκτης εκφράζει το ποσοστό των ξένων κεφαλαίων που χρησιμοποιήθηκαν για τη χρηματοδότηση των περιουσιακών στοιχείων της επιχείρησης. Μια υψηλή τιμή του μπορεί να υπονοεί την ύπαρξη υψηλού δανεισμού, αρνητικές προοπτικές μακροχρόνιας εξέλιξης, μικρά περιθώρια ασφάλειας για την αντιμετώπιση πιθανών μελλοντικών ζημιών, υψηλή πιθανότητα αθέτησης της αποπληρωμής τοκοχρεολυσίων και πιθανόν ενδείξεις χρεοκοπίας. Οι πιστωτές της επιχείρησης προτιμούν χαμηλά επίπεδα συνολικής δανειακής επιβάρυνσης, διότι όσο μικρότερος είναι ο δείκτης, τόσο περισσότερα ίδια κεφάλαια υπάρχουν για να ικανοποιηθούν οι πιστωτές σε περίπτωση χρεοκοπίας της επιχείρησης. Αντίθετα οι μέτοχοι της επιχείρησης προτιμούν υψηλά σχετικά επίπεδα του δείκτη αυτού, καθώς αυτό σημαίνει ότι οι επενδύσεις της επιχείρησης δεν χρηματοδοτούνται με ίδια κεφάλαια και άρα απομένουν για τους μετόχους περισσότερα κέρδη προς διανομή (Δασίλας, 2012; Παπαναστασόπουλος, 2014).

22. Αριθμοδείκτης Ιδίων Κεφαλαίων προς Πάγια (Ratio of Owner's Equity to Fixed Assets)

Ο αριθμοδείκτης ιδίων κεφαλαίων προς πάγια δίνεται από τον τύπο:

$$\frac{\text{Ίδια Κεφάλαια}}{\text{Πάγιο Ενεργητικό}}$$

Ο δείκτης υποδεικνύει τον τρόπο που χρηματοδοτούνται οι πάγιες επενδύσεις της επιχείρησης. Όταν τα ίδια κεφάλαια είναι μεγαλύτερα από το σύνολο των επενδύσεων σε πάγια, τότε ένα μέρος των κεφαλαίων κίνησης προέρχεται από τους μετόχους της. Όταν συμβαίνει το αντίθετο, η χρηματοδότηση των παγίων στοιχείων πραγματοποιήθηκε με δανειακά κεφάλαια (Βασιλείου & Ηρειώτης, 2008).

23. Αριθμοδείκτης Ιδίων Κεφαλαίων προς Δανειακά Κεφάλαια (μόχλευσης) – Δανειακής Επιβάρυνσης (*Ratio of Owner's Equity to Total Liabilities*)

Ο αριθμοδείκτης δανειακής επιβάρυνσης δίνεται από τον τύπο:

$$\frac{\text{Ίδια Κεφάλαια}}{\text{Ξένα Κεφάλαια}}$$

Ο δείκτης χρησιμοποιείται για την εξακρίβωση ύπαρξης υπερδανεισμού στην επιχείρηση και προσδιορίζει κατά πόσο τα ίδια κεφάλαια καλύπτουν τη χρηματοδότησή της από εξωτερικό δανεισμό. Για το λόγο αυτό θεωρείται δείκτης οικονομικής αυτοτέλειας καθώς όταν είναι μεγαλύτερος της μονάδος οι φορείς της επιχείρησης συμμετέχουν σε αυτή με περισσότερα κεφάλαια από ότι οι πιστωτές της που σημαίνει ότι τόσο μεγαλύτερη είναι και η ασφάλεια που απολαμβάνουν οι πιστωτές της (Βασιλείου & Ηρειώτης, 2008; Δασίλας, 2012).

24. Αριθμοδείκτης Κυκλοφορούντος Ενεργητικού προς Συνολικές Υποχρεώσεις (*Ratio of Current Assets to Total Liabilities*)

Ο αριθμοδείκτης κυκλοφορούντος ενεργητικού προς συνολικές υποχρεώσεις δίνεται από τον τύπο:

$$\frac{\text{Λοιπό Ενεργητικό}}{\text{Ξένα Κεφάλαια}}$$

Ο δείκτης υποδεικνύει τη ρευστότητα των μακροχρόνιων υποχρεώσεων της επιχείρησης. Η σημαντικότητα του δείκτη εναπόκειται στο γεγονός ότι ορισμένα κυκλοφοριακά στοιχεία του ενεργητικού έχουν σχετικά χαμηλό βαθμό ρευστοποίησης. Επηρεάζεται αρνητικά από τις λειτουργικές και έκτακτες ζημιές. Υψηλή τιμή του ίσως να αποτελεί ένδειξη ότι οι μακροπρόθεσμες υποχρεώσεις θα μπορέσουν να εξοφληθούν από τα κεφάλαια κίνησης, στην περίπτωση που δεν υπάρξουν υπερβολικές ζημιές από τη ρευστοποίηση των κυκλοφοριακών στοιχείων (Βασιλείου & Ηρειώτης, 2008).

25. Αριθμοδείκτης Βραχυπρόθεσμων Υποχρεώσεων προς Συνολικές Υποχρεώσεις

Ο αριθμοδείκτης βραχυπρόθεσμων υποχρεώσεων προς συνολικές υποχρεώσεις δίνεται από τον τύπο:

$$\frac{\text{Βραχυπρόθεσμο Χρέος}}{\text{Ξένα Κεφάλαια}}$$

Ο δείκτης αυτός αποτελεί ουσιαστικά το ποσοστό των βραχυπρόθεσμου δανεισμού της επιχείρησης ως προς το συνολικό δανεισμό. Ένας μικρός σχετικά δείκτης υποδεικνύει ότι η επιχείρηση καλείται να εξοφλήσει τις υποχρεώσεις της μακροπρόθεσμα γεγονός που της επιτρέπει μια σχετική ευελιξία στην τακτοποίηση των πληρωμών της.

26. Αριθμοδείκτης Κάλυψης Χρηματοοικονομικών Εξόδων (*Number of Times Interest Earned*)

Ο αριθμοδείκτης καθαρής απόδοσης ιδίων κεφαλαίων δίνεται από τον τύπο:

$$\frac{\text{Χρεωστικοί Τόκοι}}{\text{EBIT}}$$

Ο δείκτης υποδεικνύει την ικανότητα που έχει η επιχείρηση να ανταποκριθεί στις υποχρεώσεις της απέναντι στους πιστωτές της και αποτυπώνει την δανειακή της κατάσταση σε σχέση με την κερδοφόρα δυναμικότητά της. Στη βιβλιογραφία ο δείκτης εμφανίζεται με την αντίστροφη μορφή του, στην ανάλυση όμως αυτή θα χρησιμοποιηθεί ο τύπος ως έχει παραπάνω για να αποφευχθούν τα κενά πεδία τιμών στις περιπτώσεις που οι επιχειρήσεις δεν έχουν δανειακά κεφάλαια. Ουσιαστικά εκφράζει την ικανότητα κάλυψης του κόστους χρησιμοποίησης δανειακών κεφαλαίων μέσω των κερδών που η επιχείρηση έχει αποκτήσει από τις δραστηριότητες της πριν την πληρωμή των χρεωστικών τόκων και του φόρου εισοδήματος. Είναι φανερό πως όσο μικρότερος είναι ο δείκτης αυτός τόσο πιο βιώσιμη είναι οικονομικά η επιχείρηση, τόσο μεγαλύτερο το περιθώριο ασφαλείας των πιστωτών της, τόσο μικρότερος ο κίνδυνος αθέτησης εξόφλησης των υποχρεώσεων της και τόσο πιο αυξημένη η πιστοληπτική της ικανότητα (Παπαναστασόπουλος, 2014; Βασιλείου & Ηρειώτης, 2008; Δασίλας, 2012).

27. Αριθμοδείκτης Συνόλου Μακροπρόθεσμων Υποχρεώσεων προς Ίδια Κεφάλαια (*Ratio of Long Term Liabilities to Owners's Equity*)

Ο αριθμοδείκτης συνόλου μακροπρόθεσμων υποχρεώσεων προς ίδια κεφάλαια δίνεται από τον τύπο:

$$\frac{\text{Μακροπρόθεσμο Χρέος}}{\text{Ίδια Κεφάλαια}}$$

Ο δείκτης εκφράζει το ποσοστό των μακροπρόθεσμων υποχρεώσεων που αποτελούν τα ίδια κεφάλαια της επιχείρησης. Στη βιβλιογραφία ο δείκτης αυτός εμφανίζεται, όπως και ο παραπάνω, με την αντίστροφη μορφή του, στην ανάλυση όμως αυτή θα χρησιμοποιηθεί όπως παρουσιάζεται για να αποφευχθούν πολλά κενά πεδία τιμών στις περιπτώσεις που οι επιχειρήσεις δεν έχουν δανειακά κεφάλαια. Είναι αντιληπτό πως όταν ο δείκτης είναι μικρότερος της μονάδος η επιχείρηση δύναται τη συγκεκριμένη στιγμή που γίνεται η μέτρηση να εξοφλήσει τα δάνεια που οφείλονται σε μακροπρόθεσμο χρονικό ορίζοντα. Το αντίθετο συμβαίνει όταν ο δείκτης είναι μεγαλύτερος της μονάδος γεγονός που υποδεικνύει χαμηλό περιθώριο ασφαλείας και πιθανόν οικονομική δυσχέρεια της επιχείρησης. Ο δείκτης αποτελεί δείκτη φερεγγυότητας της επιχείρησης και παρακολουθείται στενά από τους ενεργούς επενδυτές της επιχείρησης αλλά και αυτούς που επιθυμούν να επενδύσουν μελλοντικά σε αυτή.

28. Αριθμοδείκτης Μακροπρόθεσμων Υποχρεώσεων προς Πάγια (*Ratio of Long Term Liabilities to Fixed Assets*)

Ο αριθμοδείκτης μακροπρόθεσμων υποχρεώσεων προς πάγια δίνεται από τον τύπο:

$$\frac{\text{Μακροπρόθεσμο Χρέος}}{\text{Πάγιο Ενεργητικό}}$$

Ο δείκτης υποδεικνύει την ασφάλεια που η επιχείρηση μπορεί να παρέχει στους μακροχρόνιους πιστωτές της μέσω της περιουσίας της. Μέσω διαχρονικής παρακολούθησης του αριθμοδείκτη αυτού μπορεί να εκτιμηθεί η πολιτική της επιχείρησης ως προς τη χρηματοδότηση των πάγιων περιουσιακών της στοιχείων. Μία διαχρονική μείωση του δείκτη μπορεί να υπονοεί την αύξηση του περιθωρίου ασφαλείας για τους μακροχρόνιους πιστωτές και επιπλέον ότι μία πιθανή επέκταση των πάγιων περιουσιακών στοιχείων χρηματοδοτήθηκε από τους ίδιους φορείς. Αντίθετα, μία διαχρονική αύξηση μπορεί να υπονοεί ότι μειώθηκε το περιθώριο ασφαλείας για τους μακροχρόνιους πιστωτές και επιπλέον ότι μία πιθανή επέκταση των πάγιων περιουσιακών στοιχείων χρηματοδοτήθηκε από προσφυγή σε εξωτερική χρηματοδότηση (Βασιλείου & Ηρειώτης, 2008; Γκούμας & Λεκαράκου, 2013).

Αριθμοδείκτες Δραστηριότητας

Οι αριθμοδείκτες δραστηριότητας (*Activity Ratios*) χρησιμοποιούνται για να μετρήσουν το βαθμό στον οποίο μία οικονομική μονάδα είναι αποτελεσματική στη διαχείριση των

περιουσιακών της στοιχείων, δηλαδή κατά πόσο ικανοποιητικά ή όχι τα χρησιμοποιεί (Βασιλείου & Ηρειώτης, 2008). Οι αριθμοδείκτες που θα χρησιμοποιηθούν από την κατηγορία αυτή είναι οι ακόλουθοι:

29. Αριθμοδείκτης Ανακύκλωσης ή Κυκλοφορίας Καθαρού Κεφαλαίου Κίνησης (*Net Working Capital Turnover Ratio*)

Ο αριθμοδείκτης κυκλοφορίας καθαρού κεφαλαίου κίνησης δίνεται από τον τύπο:

$$\frac{\text{Πωλήσεις}}{\text{Καθαρό Κεφάλαιο Κίνησης}}$$

Ο δείκτης εκφράζει πόσες φορές κατά μέσο όρο χρησιμοποιήθηκαν τα κυκλοφοριακά στοιχεία μιας επιχείρησης στην τρέχουσα χρήση προκειμένου να πραγματοποιηθούν έσοδα από πωλήσεις. Όσο μεγαλύτερη είναι η τιμή του δείκτη τόσο καλύτερη μπορεί να χαρακτηριστεί η θέση της επιχείρησης από πλευράς διαχείρισεως του κεφαλαίου κινήσεως. Μπορεί όμως να σημαίνει και χαμηλές επενδύσεις σε κεφάλαιο κινήσεως, με άμεση συνέπεια την αδυναμία απορρόφησης ζημιών από ενδεχόμενη μείωση των πωλήσεων (Παπαναστασόπουλος, 2014). Η σχέση μεταξύ των πωλήσεων και του κεφαλαίου κίνησης είναι στενή, διότι όσο αυξάνουν οι πωλήσεις τόσο περισσότερα κεφάλαια κίνησης απαιτούνται για αποθέματα και για αυξημένες ενδεχομένως πιστώσεις προς τους πελάτες της (Νιάρχος, 2004).

30. Αριθμοδείκτης Ταχύτητας Κυκλοφορίας Ενεργητικού ή Εκμετάλλευσης Περιουσιακών Στοιχείων (*Asset Turnover Ratio*)

Ο αριθμοδείκτης ταχύτητας κυκλοφορίας ενεργητικού δίνεται από τον τύπο:

$$\frac{\text{Πωλήσεις}}{\text{Ενεργητικό}}$$

Ο δείκτης υποδεικνύει το βαθμό που χρησιμοποιούνται τα στοιχεία του ενεργητικού σε σχέση με τις πωλήσεις της και αν υπάρχει υπερεπένδυση κεφαλαίων στην οικονομική μονάδα σε σχέση με το ύψος των πωλήσεων που πραγματοποιεί (Βασιλείου & Ηρειώτης, 2008).

31. Αριθμοδείκτης Ανακύκλωσης ή Ταχύτητας Κυκλοφορίας Ιδίων Κεφαλαίων (Owner's Equity Turnover Ratio)

Ο αριθμοδείκτης ταχύτητας κυκλοφορίας ιδίων κεφαλαίων δίνεται από τον τύπο:

$$\frac{\text{Πωλήσεις}}{\text{Ίδια Κεφάλαια}}$$

Ο δείκτης υποδεικνύει το βαθμό που χρησιμοποιούνται τα ίδια κεφάλαια σε σχέση με τις πωλήσεις της επιχείρησης. Δείχνει δηλαδή το ύψος των πωλήσεων που πραγματοποίησε η επιχείρηση με κάθε μονάδα ιδίων κεφαλαίων Ένας υψηλός αριθμοδείκτης υποδεικνύει μία ευνοϊκή κατάσταση για την επιχείρηση, διότι η επιχείρηση φαίνεται να πραγματοποιεί μεγάλες πωλήσεις με σχετικά μικρό ύψος ιδίων κεφαλαίων, πράγμα που ενδέχεται να την οδηγήσει σε υψηλά κέρδη. Παρόλα αυτά όταν ο αριθμοδείκτης αυξάνεται πολύ, τόσο πιο δυσμενής είναι η θέση της οικονομικής μονάδας, αφού λειτουργεί βασιζόμενη στον δανεισμό ξένων κεφαλαίων (Βασιλείου & Ηρειώτης, 2008).

32. Αριθμοδείκτης Ανακύκλωσης Παγίου Ενεργητικού ή Δείκτης Ταχύτητας Κυκλοφορίας Παγίων (Fixed Asset Turnover Ratio)

Ο αριθμοδείκτης ταχύτητας κυκλοφορίας παγίων δίνεται από τον τύπο:

$$\frac{\text{Πωλήσεις}}{\text{Πάγιο Ενεργητικό}}$$

Ο δείκτης εκφράζει την κεφαλαιακή ένταση της επιχείρησης δηλαδή το βαθμό στον οποίο χρησιμοποιούνται τα πάγια περιουσιακά στοιχεία της οικονομικής μονάδας σε σχέση με τις πωλήσεις (Δασίλας, 2012). Όσο μεγαλύτερος είναι ο δείκτης τόσο καλύτερη μπορεί να χαρακτηριστεί η θέση της επιχείρησης από πλευράς διαχείρισης του παγίου ενεργητικού της, μπορεί όμως να υποδεικνύει και χαμηλές επενδύσεις σε πάγια παραγωγικά στοιχεία (μηχανήματα, νέες τεχνολογίες), με άμεση συνέπεια την μείωση της παραγωγικότητας, της ανταγωνιστικότητας και της κερδοφορίας της (Παπαναστασόπουλος, 2014). Όσο μικρότερος είναι ο δείκτης, τόσο περισσότερα πάγια χρησιμοποιεί μια επιχείρηση για τη δημιουργία του συγκεκριμένου ύψους πωλήσεων, κάτι που μπορεί να υπονοεί υπερεπένδυση σε πάγια στοιχεία (Βασιλείου & Ηρειώτης, 2008). Διαχρονική μείωση της τιμής του δείκτη υποδηλώνει μη αποδοτική χρησιμοποίηση των παγίων για τη δημιουργία πωλήσεων.

«Στις περισσότερες περιπτώσεις παρατηρείται ότι μια σχετικά μεγάλη ταχύτητα κυκλοφορίας περιουσιακών στοιχείων συνδέεται με σχετικά χαμηλό περιθώριο κέρδους. Αν οι πωλήσεις μιας επιχείρησης βρίσκονται σε υψηλά επίπεδα και το περιθώριο κέρδους είναι χαμηλό, ο καλύτερος τρόπος για να αυξηθεί το περιθώριο αυτό είναι να μειωθεί το κόστος πωληθέντων. Αυτό, διότι, μια αύξηση του όγκου των πωλήσεων θα οδηγήσει σε υπερβολική ταχύτητα κυκλοφορίας περιουσιακών στοιχείων, μεγαλύτερη ίσως από την επιτρεπτή, ενώ μια αύξηση της τιμής πωλήσεως μπορεί να οδηγήσει σε ανεπιθύμητη μείωσή τους. Έτσι αν υπάρξει κάποια ύφεση, αυτή θα οδηγήσει σε μείωση των πωλήσεων, με συνέπεια να είναι δύσκολη η ρευστοποίηση των αποθεμάτων και η μη έγκαιρη είσπραξη των πωλήσεων με ή χωρίς μείωση της τιμής πωλήσεως» (Νιάρχος 2004).

33. Αριθμοδείκτης Κυκλοφορίας Υποχρεώσεων (*Total Liabilities Turnover Ratio*)

Ο αριθμοδείκτης κυκλοφορίας υποχρεώσεων δίνεται από τον τύπο:

$$\frac{\text{Πωλήσεις}}{\text{Ξένα Κεφάλαια}}$$

Ο δείκτης εκφράζει πόσες φορές οι πωλήσεις της επιχείρησης καλύπτουν το σύνολο των υποχρεώσεων προς τους δανειστές της. Και ο δείκτης αυτός αποτελεί δείκτη φερεγγυότητας της επιχείρησης, συνεπώς όσο μεγαλύτερος είναι τόσο πιο πιθανή είναι η συνεπής εκπλήρωση των υποχρεώσεων της επιχείρησης.

Αριθμοδείκτες Δαπανών Λειτουργίας

Οι αριθμοδείκτες δαπανών λειτουργίας (*operating expense ratios*) αυτοί χρησιμοποιούνται για να απεικονίσουν την πολιτική που ακολουθεί η επιχείρηση ως προς τις διάφορες δαπάνες λειτουργίας που πραγματοποιεί αλλά και την αποτελεσματικότητά της ως προς τις δαπάνες αυτές (Βασιλείου & Ηρειώτης, 2008). Οι αριθμοδείκτες που θα χρησιμοποιηθούν από την κατηγορία αυτή είναι οι ακόλουθοι:

34. Αριθμοδείκτης Λειτουργικών Εξόδων προς Πωλήσεις (*Operating Expenses to Net Sales Ratio*)

Ο αριθμοδείκτης λειτουργικών εξόδων προς πωλήσεις δίνεται από τον τύπο:

$$\frac{\text{Λειτουργικά Έξοδα}}{\text{Πωλήσεις}}$$

Ο δείκτης εκφράζει τη σχέση των λειτουργικών εξόδων της επιχείρησης ως προς τις πωλήσεις που πραγματοποίησε στη χρήση. Ο δείκτης απεικονίζει την τάση των εξόδων λειτουργίας και χρησιμοποιείται για τη διάγνωση ή και την πρόληψη τυχόν προβλημάτων στη λειτουργία της επιχείρησης. Τα λειτουργικά έξοδα των επιχειρήσεων διαφέρουν μεταξύ τους, ανάλογα με το είδος των εργασιών τους, των υπηρεσιών τους και των λειτουργικών τους αναγκών. Όσο μεγαλύτερος είναι ο δείκτης τόσο δυσχερέστερη είναι η θέση της επιχείρησης, καθώς το υπόλοιπο των λειτουργικών κερδών πιθανόν να μην επαρκεί για την εξυπηρέτηση των χρηματοοικονομικών και των άλλων εξόδων της επιχείρησης (Γκούμας & Λεκαράκου, 2013; Βασιλείου & Ηρειώτης, 2008).

35. Αριθμοδείκτης λειτουργικών εξόδων (*Operating ratio*)

Ο αριθμοδείκτης λειτουργικών εξόδων δίνεται από τον τύπο:

$$\frac{\text{Κόστος Πωληθέντων} + \text{Λειτουργικά Έξοδα}}{\text{Πωλήσεις}}$$

Ο δείκτης εκφράζει το ποσοστό των καθαρών πωλήσεων που απορροφάται από τα λειτουργικά έξοδα και το κόστος πωληθέντων. Όπως και με τον προηγούμενο δείκτη, όσο μεγαλύτερος είναι ο δείκτης τόσο δυσχερέστερη είναι η θέση της επιχείρησης, διότι ενδέχεται το υπόλοιπο των λειτουργικών κερδών να μην επαρκεί για την εξυπηρέτηση των χρηματοοικονομικών και των άλλων εξόδων της επιχείρησης. Ο δείκτης δεν μπορεί να χρησιμοποιηθεί σαν κριτήριο οικονομικής κατάστασης της επιχείρησης αν δεν συμπεριληφθούν σε αυτόν και τα διάφορα τακτικά και έκτακτα έσοδα και έξοδα που λαμβάνουν μέρος στο οικονομικό αποτέλεσμα (Βασιλείου & Ηρειώτης, 2008).

Αριθμοδείκτες Ρευστότητας

Οι αριθμοδείκτες ρευστότητας (*Liquidity Ratios*) χρησιμοποιούνται για να προσδιορίσουν την βραχυχρόνια οικονομική θέση της επιχείρησης, καθώς επίσης και την ικανότητά της να ανταποκρίνεται στις βραχυπρόθεσμες υποχρεώσεις της. Οι αριθμοδείκτες που θα χρησιμοποιηθούν από την κατηγορία αυτή είναι οι ακόλουθοι:

36. Αριθμοδείκτης Γενικής Ρευστότητας ή Κεφαλαίου Κίνησης (*Current Ratio*)

Ο αριθμοδείκτης γενικής ρευστότητας δίνεται από τον τύπο:

$$\frac{\text{Λοιπό Ενεργητικό}}{\text{Βραχυπρόθεσμο Χρέος}}$$

Ο δείκτης αποτελεί δείκτη πιστοληπτικής ικανότητας της επιχείρησης. Υποδεικνύει αν και κατά πόσο οι βραχυπρόθεσμες υποχρεώσεις της καλύπτονται στη δεδομένη στιγμή από τα άμεσα ρευστοποιήσιμα στοιχεία του ενεργητικού της, δηλαδή το κυκλοφορούν ενεργητικό. Ο δείκτης αποτελεί μέτρο ρευστότητας και περιθωρίου ασφάλειας για την αντιμετώπιση μίας τυχόν ανεπιθύμητης εξέλιξης στη ροή κεφαλαίων κίνησης και στη βραχυπρόθεσμη δανειοδότηση της επιχείρησης. Όπως αναφέρθηκε και παραπάνω, το σύνολο των κυκλοφοριακών στοιχείων αποτελεί το άθροισμα των διαθεσίμων, των απαιτήσεων και των αποθεμάτων μίας οικονομικής μονάδας, ενώ το σύνολο των βραχυχρόνιων υποχρεώσεων αποτελούν οι πιστώσεις προμηθευτών, οι πληρωτέοι φόροι, τα βραχυπρόθεσμα δάνεια, τα μερίσματα πληρωτέα και οι προκαταβολές πελατών. Από τα δεδομένα που διαθέτουμε δεν μπορούμε δυστυχώς να προσδιορίσουμε το ποσοστό συμμετοχής του κάθε παράγοντα στα συνολικά μεγέθη γεγονός που δυσχεραίνει τον προσδιορισμό της έντασης του δείκτη. Και αυτό γιατί για παράδειγμα μία επιχείρηση που διαθέτει μεγαλύτερο ποσοστό σε μετρητά είναι σε ευνοϊκότερη θέση, από άποψη ρευστότητας, από μία άλλη που διαθέτει μεγαλύτερο ποσοστό σε αποθέματα, ακόμη και αν έχουν τον ίδιο αριθμοδείκτη γενικής ρευστότητας. Διαχρονική μείωση του δείκτη μπορεί να υποδεικνύει οικονομικές δυσμένειες τουλάχιστον σε ότι αφορά την πλήρωση των τρεχουσών υποχρεώσεων της επιχείρησης ωστόσο, υπάρχει και η πιθανότητα μίας πιο ορθολογικής χρησιμοποίησης του υπάρχοντος κεφαλαίου κίνησης (Βασιλείου & Ηρειώτης, 2008; Δασίλας, 2012; Παπαναστασόπουλος, 2014).

37. Αριθμοδείκτης Βραχυπρόθεσμων Υποχρεώσεων προς Καθαρό Κεφάλαιο Κίνησης (Ratio of Long Term Liabilities to Owners's Equity)

Ο αριθμοδείκτης βραχυπρόθεσμων υποχρεώσεων προς καθαρό κεφάλαιο κίνησης δίνεται από τον τύπο:

$$\frac{\text{Βραχυπρόθεσμο Χρέος}}{\text{Καθαρό Κεφάλαιο Κίνησης}}$$

Ο δείκτης αποτελεί δείκτη πιστοληπτικής ικανότητας της επιχείρησης. Υποδεικνύει το ποσοστό του καθαρού κεφαλαίου κίνησης που αποτελούν οι βραχυπρόθεσμες υποχρεώσεις της επιχείρησης. Όσο μικρότερος είναι ο δείκτης τόσο μεγαλύτερο είναι το κεφάλαιο κίνησης και άρα τόσο μεγαλύτερη η ασφάλεια ρευστότητας που απολαμβάνει η επιχείρηση.

Συνολικά οι μεταβλητές που θα χρησιμοποιηθούν στην ανάλυση θα είναι οι ακόλουθες:

Πίνακας 2.2-8 Σύνολο Μεταβλητών που θα χρησιμοποιηθούν στην Ανάλυση

a/a	Αριθμοδείκτης	Αριθμητής	Παρονομαστής	Τύπος
1	Μεικτό Περιθώριο Κέρδους	Μεικτό Κέρδος	Πωλήσεις	Αποδοτικότητα
2	Περιθώριο EBITDA	EBITDA	Πωλήσεις	Αποδοτικότητα
3	Περιθώριο EBIT (Περιθώριο Λειτουργικού Κέρδους)	EBIT (Καθαρά Λειτουργικά Κέρδη)	Πωλήσεις	Αποδοτικότητα
4	Καθαρό Περιθώριο Κέρδους	Καθαρά Κέρδη	Πωλήσεις	Αποδοτικότητα
5	Περιθώριο Αποσβέσεων	Αποσβέσεις	Πωλήσεις	Αποδοτικότητα
6	Λοιπό Ενεργητικό προς Πωλήσεις	Λοιπό Ενεργητικό	Πωλήσεις	Αποδοτικότητα
7	Ενεργητικό προς Πωλήσεις	Ενεργητικό	Πωλήσεις	Αποδοτικότητα
8	EBITDA προς Ενεργητικό	EBITDA	Ενεργητικό	Αποδοτικότητα
9	Αριθμοδείκτης Βασικής Ικανότητας Κερδών	EBIT	Ενεργητικό	Αποδοτικότητα
10	Απόδοση Ενεργητικού προ Φόρων	Κέρδη προ Φόρων	Ενεργητικό	Αποδοτικότητα
11	Απόδοση Ενεργητικού ή Λειτουργίας μετά Φόρων	Καθαρά Κέρδη	Ενεργητικό	Αποδοτικότητα
12	Αποδοτικότητα Απασχολούμενων Κεφαλαίων	Καθαρά Κέρδη + Χρεωστικοί Τόκοι	Ενεργητικό	Αποδοτικότητα
13	Αποδοτικότητα Ιδίων Κεφαλαίων	EBIT	Ίδια Κεφάλαια	Αποδοτικότητα
14	Απόδοση Ιδίων Κεφαλαίων προ Φόρων	Κέρδη προ Φόρων	Ίδια Κεφάλαια	Αποδοτικότητα
15	Καθαρή Απόδοση Ιδίων Κεφαλαίων	Καθαρά Κέρδη	Ίδια Κεφάλαια	Αποδοτικότητα
16	Αριθμοδείκτης Οικονομικής	Καθαρή Απόδοση	Αποδοτικότητα	Αποδοτικότητα

	Μόχλευσης	Ιδίων Κεφαλαίων	Απασχολούμενων Κεφαλαίων	
17	Αριθμοδείκτης Απόσβεσης Παγίων	Αποσβέσεις	Πάγιο Ενεργητικό + Αποσβέσεις	Αποδοτικότητα
18	Αριθμοδείκτης Φορολογικής Επιβάρυνσης	Καθαρά Κέρδη	Κέρδη προ Φόρων	Αποδοτικότητα
19	Απόδοση Μακροπρόθεσμων Κεφαλαίων	Καθαρά Κέρδη	Μακροπρόθεσμο Χρέος + Ίδια Κεφάλαια	Αποδοτικότητα
20	Αριθμοδείκτης Ιδίων Κεφαλαίων προς Συνολικά Κεφάλαια	Ίδια Κεφάλαια	Ενεργητικό	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
21	Αριθμοδείκτης Δανειακής Επιβάρυνσης	Ξένα Κεφάλαια	Ενεργητικό	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
22	Αριθμοδείκτης Ιδίων Κεφαλαίων προς Πάγια	Ίδια Κεφάλαια	Πάγιο Ενεργητικό	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
23	Αριθμοδείκτης Ιδίων Κεφαλαίων προς Δανειακά Κεφάλαια	Ίδια Κεφάλαια	Ξένα Κεφάλαια	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
24	Αριθμοδείκτης Κυκλοφορούντος Ενεργητικού προς Συνολικές Υποχρεώσεις	Λοιπό Ενεργητικό	Ξένα Κεφάλαια	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
25	Αριθμοδείκτης Βραχυπρόθεσμων Υποχρεώσεων προς Συνολικές Υποχρεώσεις	Βραχυπρόθεσμο Χρέος	Ξένα Κεφάλαια	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
26	Αριθμοδείκτης Κάλυψης Χρηματοοικονομικών Εξόδων	Χρεωστικοί Τόκοι	EBIT	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
27	Αριθμοδείκτης Ίδια Κεφάλαια προς Μακροπρόθεσμες Υποχρεώσεις	Μακροπρόθεσμο Χρέος	Ίδια Κεφάλαια	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
28	Αριθμοδείκτης Μακροπρόθεσμων Υποχρεώσεων προς Πάγια	Μακροπρόθεσμο Χρέος	Πάγιο Ενεργητικό	Κεφαλαιακής Διάρθρωσης και Βιωσιμότητας
29	Αριθμοδείκτης Κυκλοφορίας Καθαρού Κεφαλαίου Κίνησης	Πωλήσεις	Καθαρό Κεφάλαιο Κίνησης	Δραστηριότητας
30	Αριθμοδείκτης Ταχύτητας Κυκλοφορίας Ενεργητικού	Πωλήσεις	Ενεργητικό	Δραστηριότητας
31	Αριθμοδείκτης Ταχύτητας Κυκλοφορίας Ιδίων Κεφαλαίων	Πωλήσεις	Ίδια Κεφάλαια	Δραστηριότητας
32	Αριθμοδείκτης Ταχύτητας Κυκλοφορίας Παγίων	Πωλήσεις	Πάγιο Ενεργητικό	Δραστηριότητας
33	Αριθμοδείκτης Ταχύτητας Κυκλοφορίας Υποχρεώσεων	Πωλήσεις	Ξένα Κεφάλαια	Δραστηριότητας
34	Αριθμοδείκτης Λειτουργικών Εξόδων προς Πωλήσεις	Λειτουργικά Έξοδα	Πωλήσεις	Δαπανών Λειτουργίας
35	Αριθμοδείκτης Λειτουργικών Εξόδων	Κόστος Πωληθέντων + Λειτουργικά Έξοδα	Πωλήσεις	Δαπανών Λειτουργίας

36	Αριθμοδείκτης Γενικής Ρευστότητας	Λοιπό Ενεργητικό	Βραχυπρόθεσμο Χρέος	Ρευστότητας
37	Αριθμοδείκτης Βραχυπρόθεσμων Υποχρεώσεων προς Καθαρό Κεφάλαιο Κίνησης	Βραχυπρόθεσμο Χρέος	Καθαρό Κεφάλαιο Κίνησης	Ρευστότητας
38	Γεωγραφικός Τομέας			

ΚΕΦΑΛΑΙΟ 3

Βιβλιογραφική Επισκόπηση

Το κεφάλαιο αυτό περιλαμβάνει μια επισκόπηση των στατιστικών μεθόδων και μοντέλων Πολυμεταβλητής Ανάλυσης που αναφέρεται στη βιβλιογραφία ότι χρησιμοποιούνται για τη διαχείριση του πιστωτικού κινδύνου των επιχειρήσεων.

Τα μοντέλα διαχείρισης πιστωτικού κινδύνου διακρίνονται στα παραδοσιακά υποδείγματα, τα λεγόμενα μοντέλα πιστωτικής βαθμονόμησης (*credit scoring models*), και τα δομικά υποδείγματα (*structural models*).

Στην πρώτη κατηγορία, η διαχείριση του πιστωτικού κινδύνου αποτελεί διαδικασία κατά την οποία μοντελοποιείται η πιστοληπτική ικανότητα μιας επιχείρησης ή ενός οργανισμού, μέσω συλλογής, ανάλυσης και στατιστικής επεξεργασίας οικονομικών δεδομένων που προέρχονται από τις λογιστικές τους καταστάσεις. Κατά τη διαδικασία επεξεργασίας μελετώνται τα χαρακτηριστικά και οι ιδιαιτερότητες των επιχειρήσεων που αθέτησαν στο παρελθόν και συγκρίνονται με τα χαρακτηριστικά των επιχειρήσεων που δεν αθέτησαν. Κατόπιν, για κάθε νέα επιχείρηση για την οποία εξετάζεται ο κίνδυνος αθέτησης των υποχρεώσεών της, πραγματοποιείται σύγκριση των χαρακτηριστικών της εξεταζόμενης επιχείρησης με τα χαρακτηριστικά των επιχειρήσεων που αθέτησαν και που δεν αθέτησαν, εκτιμάται η πιθανότητα αθέτησης, και ανάλογα με το όριο κινδύνου που έχει τεθεί από τον ερευνητή, η εξεταζόμενη επιχείρηση κατανέμεται είτε στην κατηγορία των επιχειρήσεων που «πιθανόν να αθετήσουν» είτε στην κατηγορία εκείνων που «πιθανόν να μην αθετήσουν» (Abdou & Pointon, 2011).

Πέραν της βασικής αξιολόγησης της αθέτησης των υποχρεώσεων πληρωμών των εταιριών, τα μοντέλα πιστωτικού κινδύνου χρησιμοποιήθηκαν και χρησιμοποιούνται ευρέως στην αξιολόγηση των αιτήσεων εγκρίσεως και χορήγησης δανείων από τα χρηματοπιστωτικά ιδρύματα στους εταιρικούς και ιδιώτες πελάτες τους. Τα δεδομένα που χρησιμοποιούν τα ιδρύματα αυτά αφορούν κυρίως την ιστορικότητα των συναλλαγών των αιτούντων των δανείων, καθώς και άλλα δημογραφικά, επαγγελματικά και προσωπικά χαρακτηριστικά (*micro – data*) τους. Η ανάγκη διάκρισης των πελατών των τραπεζών σε «καλούς» και σε «κακούς» καθώς και η εφαρμογή στατιστικών μεθόδων για την εν λόγω διαδικασία χρονολογείται από πολύ παλιά

και συνεχίζεται ευρύτατα έως και σήμερα (Durand, 1941; Sustersic, et al., 2009; Orgler, 1971; Steenackers & Goovaerts, 1989).

Η δεύτερη κατηγορία μοντέλων διαχείρισης πιστωτικού κινδύνου αφορά τα λεγόμενα δομικά υποδείγματα και βασίζεται στη θεωρία χρηματοοικονομικών δικαιωμάτων (*option theory*), συγκεκριμένα στη χρήση τεχνικών τιμολόγησης των δικαιωμάτων αυτών. Με βάση αυτές, και σε συνδυασμό με τις πληροφορίες που προέρχονται από τη χρηματιστηριακή αγορά, πραγματοποιείται συνδυαστική εκτίμηση του πιστωτικού κινδύνου μιας επιχείρησης. Τα μοντέλα που χρησιμοποιούνται σήμερα είναι πιο εξελιγμένα και βασίζονται στη στοχαστική εκτίμηση του πιστωτικού κινδύνου (Παπαναστασόπουλος, 2007). Το πρώτο μοντέλο που υιοθετήθηκε για την εφαρμογή αυτής της διαδικασίας ήταν το μοντέλο τιμολόγησης δικαιωμάτων των Black – Scholes (1973) και Merton (1973, 1974).

Στα πλαίσια της εργασίας αυτής θα γίνει αναφορά και χρήση των μοντέλων της πρώτης κατηγορίας μοντέλων, δηλαδή των παραδοσιακών υποδειγμάτων, καθώς τα δεδομένα που χρησιμοποιούνται δεν αφορούν εισηγμένες βιομηχανικές εταιρίες και συνεπώς δεν περιλαμβάνονται στο υπόδειγμα πληροφορίες της αγοράς.

Τα παραδοσιακά υποδείγματα πιστωτικού σκορ, όπως προαναφέρθηκε, χρησιμοποιούν ως δεδομένα συγκεκριμένα λογιστικά μεγέθη καθώς και χρηματοοικονομικούς δείκτες που λαμβάνονται από τις λογιστικές καταστάσεις των επιχειρήσεων. Το σύνολο των δεδομένων αυτών αποτελούν τα δεδομένα εισόδου της στατιστικής ανάλυσης που πραγματοποιείται από τα υποδείγματα. Οι πληροφορίες που φέρουν τα δεδομένα εισόδου αφορούν κυρίως τη ρευστότητα (*liquidity*) της επιχείρησης, την αποδοτικότητά της (*profitability*), τη δραστηριότητά της (*efficiency*), τη διάρθρωση των κεφαλαίων και τη βιωσιμότητά της (*capital structure and viability*), το μέγεθός της και τις ταμειακές της ροές (Παπαναστασόπουλος, 2007).

Τα παραδοσιακά μοντέλα πιστωτικού κινδύνου, αν και είναι ευρέως διαδεδομένα λόγω της καλής προσαρμογής τους και της καλής προβλεπτικής τους ικανότητας, είναι στατικά και βασίζονται σε ιστορικά λογιστικά δεδομένα τα οποία περιγράφουν στιγμιότυπα της τωρινής και της παρελθοντικής συμπεριφοράς της επιχείρησης (π.χ. η απεικόνιση μπορεί να αναφέρεται στην τελευταία ημέρα του χρόνου, ή την τελευταία ημέρα του τριμήνου) (Παπαναστασόπουλος, 2007). Δεν επιτυγχάνουν να ενσωματώσουν και να αποτυπώσουν επαρκώς τη μελλοντική συμπεριφορά των εταιριών και τις τάσεις του επιχειρηματικού τους περιβάλλοντος (Agarwal & Taffler, 2008). Τέλος στηρίζονται σε μια γραμμική απεικόνιση της κατάστασης που περιγράφουν κάτι το οποίο δεν ισχύει στην πραγματικότητα (Altman & Saunders, 1998).

Υπάρχουν το λιγότερο τρεις παραδοσιακές μεθοδολογικές προσεγγίσεις για την ανάπτυξη πολυμεταβλητών μοντέλων διαχείρισης πιστωτικού κινδύνου. Οι μέθοδοι αυτές βασίζονται

στην κατηγοριοποίηση των παρατηρήσεων ενός πληθυσμού σε ομάδες – επίπεδα πιστωτικού κινδύνου και αφορούν τα εξής μοντέλα: α) τα γραμμικά μοντέλα (*linear models*), β) τα μοντέλα διακριτής ανάλυσης (*discriminant analysis*) και γ) τα μοντέλα logit (λογιστική παλινδρόμηση) και probit. Στα πλαίσια των επιστημονικών εκδόσεων του περιοδικού “Journal of Banking and Finance”, η λογιστική παλινδρόμηση αναφέρεται ως το ευρύτερα χρησιμοποιούμενο μοντέλο εκ των προαναφερθέντων τριών για τη διαχείριση του πιστωτικού κινδύνου (Altman & Saunders, 1998).

Τα μοντέλα που προαναφέρθηκαν βασίζονται σε τεχνικές κατηγοριοποίησης που αποσκοπούν στην πρόβλεψη της κλάσης όπου θα ανήκει το σύνολο των παρατηρήσεων ενός πληθυσμού. Για τη δημιουργία του μοντέλου πρόβλεψης απαιτείται ένα σύνολο δεδομένων «εκπαίδευσης» (*training set*) το οποίο συνιστά ένα μεγάλο υποσύνολο του αρχικού συνόλου των δεδομένων. Τα δεδομένα εκπαίδευσης κατέχουν καθοριστικό ρόλο στην διαδικασία της κατηγοριοποίησης καθώς αποτελούν την πηγή από όπου ο αλγόριθμος θα αποκομίσει τη γνώση που χρειάζεται για να χτίσει το μοντέλο πρόβλεψης (*data mining*). Η απόδοση του μοντέλου κρίνεται βάσει της προβλεπτικής ικανότητας των μεταβλητών που ο αλγόριθμος θα ξεχωρίσει ως πιο αποτελεσματικές για τη διενέργεια της πρόβλεψης, και εν τέλει της ορθής ταξινόμησης των παρατηρήσεων στις εκάστοτε ομάδες. Για να επιβεβαιωθεί η αποδοτικότητα του αλγορίθμου χρησιμοποιείται το υπόλοιπο σύνολο των αρχικών δεδομένων που δεν χρησιμοποιήθηκαν κατά την εκπαίδευση. Το υπολειπόμενο αυτό σύνολο αποτελεί τα δεδομένα δοκιμής (*test set*) και με αυτά εξετάζεται αν η γνώση που ανέπτυξε ο αλγόριθμος και το μοντέλο που δημιούργησε κατά τη διαδικασία εκμάθησης είναι επαρκή και ικανά για να ταξινομήσουν σωστά τα δεδομένα δοκιμής στις κατάλληλες ομάδες. Ακολούθως, κάθε νέα παρατήρηση (*validation data*) που θα προσδιορίζεται από τα ίδια χαρακτηριστικά με τα αρχικά δεδομένα θα μπορεί να κατηγοριοποιηθεί σε κάποια ομάδα βάσει της γνώσης που δημιουργήθηκε από το μοντέλο. Η διαδικασία της μάθησης αυτής ονομάζεται στη βιβλιογραφία «εποπτευόμενη μάθηση» διότι είναι απαραίτητη η εκ των προτέρων γνώση της κλάσης – κατάταξης που ανήκει το αρχικό σύνολο των δεδομένων, δηλαδή τα δεδομένα εκπαίδευσης και τα δεδομένα δοκιμής. (Πελέκης, 2014).

Ο τρόπος με τον οποίο εφαρμόζεται η κατηγοριοποίηση σε κάθε μια από τις τρεις προαναφερθείσες μεθοδολογικές προσεγγίσεις καθώς και η εφαρμογή των προσεγγίσεων αυτών, σύμφωνα με τη βιβλιογραφία, στη διαχείριση του πιστωτικού κινδύνου αναλύεται ακολούθως.

Τα γραμμικά μοντέλα που χρησιμοποιούνται στην διαχείριση του πιστωτικού κινδύνου είναι κυρίως η γραμμική παλινδρόμηση, στην οποία χρησιμοποιείται κατηγορική μεταβλητή που

υποδηλώνει τις κλάσεις του προβλήματος, και η παλινδρόμηση Poisson. Τέτοιου είδους παλινδρομήσεις χρησιμοποίησε ο Orgler (1971) για να αξιολογήσει, ως προς την πιθανότητα αθέτησης πληρωμής, ένα σύνολο δανείων πελατών τραπεζών. Η χρήση της παλινδρόμησης επεκτείνεται και σε άλλου είδους χρηματοοικονομικά ζητήματα, που αφορούν γενικότερα τον προσδιορισμό και την αξιολόγηση της οικονομικής κατάστασης των επιχειρήσεων, (Hand & Henley, 1997; Hand, et al., 1998), τα αποτελέσματα όμως που παράγει δεν είναι τόσο ικανοποιητικά όσο τα αποτελέσματα των άλλων μεθοδολογικών προσεγγίσεων.

Η *διακριτή ανάλυση* είναι μια απλή παραμετρική στατιστική τεχνική που χρησιμοποιείται για το διαχωρισμό ενός συνόλου παρατηρήσεων σε δύο ή περισσότερες ομάδες. Η διακριτή ανάλυση προτάθηκε αρχικά από τον Fisher (1936) σαν τεχνική διαφοροποίησης και κατηγοριοποίησης. Στον τομέα της χρηματοοικονομικής, ο Beaver (1968) εφάρμοσε για πρώτη φορά μονομεταβλητή διακριτή ανάλυση (*univariate discriminant analysis*) και συγκεκριμένα εφάρμοσε ελέγχους διχοτομικής ταξινόμησης (*dichotomous classification tests*) για να διακρίνει ποιοι χρηματοοικονομικοί αριθμοδείκτες παρουσιάζουν καλύτερη προβλεπτική ικανότητα στην εταιρική βιωσιμότητα και την επιχειρηματική αποτυχία (*solvency position and firm failure*). Το 1968 ο Altman δημιούργησε το μοντέλο “Z – score” που αποτελεί το πρώτο υπόδειγμα πιστωτικού σκορ. Το υπόδειγμα αυτό, που βασίζεται σε πέντε αριθμοδείκτες – προβλέπουσες μεταβλητές, αποτέλεσε την αρχή για την εφαρμογή της πολυμεταβλητής ανάλυσης στη διαχείριση του πιστωτικού κινδύνου.

Μετά το 1970, η πολυμεταβλητή διακριτή ανάλυση (*multivariate discriminant analysis*) χρησιμοποιήθηκε αρκετά ως συγκριτική μέθοδος στα μοντέλα πιστωτικού κινδύνου. Στην πιο συνηθισμένη της μορφή η μέθοδος επιδιώκει να αναπτύξει μια γραμμική συνάρτηση μεταξύ των οικονομικών μεταβλητών που θα αποφασίσει να χρησιμοποιήσει ο ερευνητής στο μοντέλο, και βάσει της συνάρτησης αυτής να επιτευχθεί ο διαχωρισμός των οντοτήτων που πιθανόν να αθετήσουν και εκείνων που πιθανόν να μην αθετήσουν. Αυτό απαιτεί στατιστική ανάλυση κατά την οποία, μέσω του συνόλου των μεταβλητών που θα τεθεί προς επεξεργασία στο μοντέλο, θα διαμορφωθούν ομάδες μεταβλητών που θα πληρούν δύο προϋποθέσεις: θα επιτυγχάνεται μεγιστοποίηση της διακύμανσης μεταξύ διαφορετικών ομάδων μεταβλητών και θα ελαχιστοποιείται η διακύμανση μεταξύ των μεταβλητών που αποτελούν μια ομάδα (Altman & Saunders, 1998). Η μεθοδολογία αυτή αποτελεί προάγγελο της εφαρμογής της λογιστικής παλινδρόμησης και της ανάλυσης συστάδων. Αργότερα, οι Altman et al. (1977) ,με βάση τη διακριτή ανάλυση, διερεύνησαν την προβλεπτική ικανότητα επτά οικονομικών μεταβλητών και βελτίωσαν την προηγούμενη έκδοση του μοντέλου “Zeta”.

Ένα από τα μειονεκτήματα που παρουσιάζει η διακριτή ανάλυση αποτελεί το γεγονός ότι το εύρος των εκτιμήσεων των πιθανοτήτων αθέτησης δεν περιορίζεται αυστηρά στο διάστημα 0 και 1. Επίσης ζητήματα εγείρονται ως προς το ποια κατανομή θα πρέπει να ακολουθούν οι μεταβλητές που χρησιμοποιούνται στο μοντέλο για να έχει στατιστική βαρύτητα η ανάλυση, ποια είναι η επιλογή των κατάλληλων *a priori* πιθανοτήτων και / ή τα κόστη της εσφαλμένης κατάταξης, πώς πραγματοποιείται η εκτίμηση των ρυθμών λάθους ταξινόμησης κ.α. (Eisenbeis, 1977). Παρ' όλα αυτά η διακριτή ανάλυση, ακόμη και σήμερα, συνεχίζει να χρησιμοποιείται αρκετά σε εφαρμογές διαχείρισης πιστωτικού κινδύνου (Abdou & Pointon, 2009; Caouette, et al., 1998; Hand, et al., 1998; Hand & Henley, 1997; Desai, et al., 1996).

Εξαιτίας των παραπάνω μειονεκτημάτων, αργότερα εμφανίστηκαν πιο εξελιγμένα μοντέλα, τα λεγόμενα γενικευμένα γραμμικά μοντέλα, με ευρύτερα χρησιμοποιούμενα τα μοντέλα *logit* (λογιστική παλινδρόμηση) και *probit*. Τα μοντέλα αυτά δίνουν αρκετά παρόμοιες εκτιμήσεις των παραμέτρων τους γι' αυτό και το μοντέλο *probit* χρησιμοποιείται συνήθως εναλλακτικά του μοντέλου *logit* (Abdou & Pointon, 2011). Τα δύο μοντέλα διαφέρουν ως προς την υπόθεση κατανομής των καταλοίπων. Συγκεκριμένα, στο μοντέλο *logit* γίνεται η υπόθεση λογιστικής κατανομής των καταλοίπων ενώ στο μοντέλο *probit* γίνεται η υπόθεση κανονικής κατανομής (Παπαναστασόπουλος, 2007). Όσον αφορά τη διαχείριση κινδύνου, τα μοντέλα αυτά, λαμβάνοντας υπόψη ποικίλες μεταβλητές που αφορούν χρηματοοικονομικά μεγέθη καθώς και την αλληλεπίδρασή αυτών των μεγεθών, είναι σε θέση να διαχωρίσουν τις επιχειρήσεις σε αυτές που έχουν πιθανότητα να αθετήσουν και που δεν θα αθετήσουν. Η πιθανότητα αυτή εκτιμάται με τη μέθοδο της μέγιστης πιθανοφάνειας (*maximum likelihood method*) και είναι δεδομένο ότι θα βρίσκεται μεταξύ του 0 και 1. Τα μοντέλα αυτά είναι ευαίσθητα στην πολυσυγγραμμικότητα των μεταβλητών, για αυτό και πριν την χρησιμοποίησή τους πολλές φορές εφαρμόζονται στατιστικές τεχνικές μείωσης της διάστασης των δεδομένων όπως είναι η Ανάλυση Κυρίων Συνιστωσών (*Principal Component Analysis*) και η Ανάλυση Παραγόντων (*Factor Analysis*).

Ιδιαίτερα το μοντέλο της λογιστικής παλινδρόμησης χρησιμοποιείται κατά κόρον σε εφαρμογές διαχείρισης πιστωτικού κινδύνου. Ο Martin (1977) χρησιμοποίησε το μοντέλο *logit* και το μοντέλο της διακριτής ανάλυσης για να προβλέψει τραπεζικές χρεοκοπίες κατά την περίοδο 1975 – 1976, όταν 23 τράπεζες χρεοκόπησαν. Και τα δύο μοντέλα έδωσαν παρόμοιες κατηγοριοποιήσεις στην αναγνώρισή των χρεοκοπημένων και των μη χρεοκοπημένων τραπεζών. Το υπόδειγμα *logit* χρησιμοποιήθηκε στη διαχείριση πιστωτικού κινδύνου και από τον Ohlson το 1980. Ο West (1985) χρησιμοποίησε το μοντέλο *logit* σε συνδυασμό με Ανάλυση Παραγόντων (*factor analysis*) για να εκτιμήσει τη χρηματοοικονομική κατάσταση χρηματοπιστωτικών ιδρυμάτων και να τους αποδώσει μια πιθανότητα χρεοκοπίας. Πράγματι οι

παράγοντες που αναγνώρισε το μοντέλο logit ήταν παρόμοιοι με τις CAMEL συνιστώσες που χρησιμοποιούσαν οι ελεγκτές των τραπεζών για την αναγνώριση κινδύνου στο κάθε ίδρυμα. Οι Crook et al (2007) χρησιμοποίησαν τη λογιστική παλινδρόμηση για να εκτιμήσουν τον κίνδυνο που απορρέει από το τραπεζικό δανεισμό πελατών – νοικοκυριών. Οι Baesens et al (2003) συνέκριναν τη λογιστική παλινδρόμηση και τη γραμμική διακριτή ανάλυση με πιο εξελιγμένες τεχνικές κατηγοριοποίησης σε οχτώ σύνολα δεδομένων που προέρχονται από χρηματοπιστωτικά ιδρύματα της Αγγλίας, του Βελγίου, της Ολλανδίας και του Λουξεμβούργου. Οι ερευνητές διαπίστωσαν πως οι δύο παραδοσιακές τεχνικές συμπεριφέρονται πολύ καλά στην πρόβλεψη του πιστωτικού κινδύνου. Σε παρόμοιο αποτέλεσμα κατέληξαν και οι Desai et al (2000) στην έρευνά τους για το ίδιο θέμα.

Συμπερασματικά προκύπτει ότι, στη βιβλιογραφία η χρήση των παραπάνω τεσσάρων τεχνικών αποτελεί βασικό εργαλείο στη διαμόρφωση των μοντέλων πιστωτικού κινδύνου (Abdou & Pointon, 2011; Sarlija, et al., 2009; Banasik, et al., 2001; Greene, 1998; Leonard, 1995; Steenackers & Goovaerts, 1989; Boyes, et al., 1989; Orgler, 1971).

Σήμερα, εκτός των παραπάνω μοντέλων, οι τεχνικές ομαδοποίησης και κατηγοριοποίησης δεδομένων χρησιμοποιούν πολλές μεθόδους από τα πεδία της τεχνητής νοημοσύνης και ειδικότερα από τις τεχνικές εξόρυξης δεδομένων. Οι τεχνικές αυτές είναι ιδιαίτερα εξελιγμένες καθώς έχουν τη δυνατότητα να μοντελοποιούν εξαιρετικά πολύπλοκες συναρτήσεις. (Abdou & Pointon, 2011). Τέτοιες τεχνικές αποτελούν τα νευρωνικά δίκτυα (*neural networks*), οι γενετικοί αλγόριθμοι (*genetic algorithms*), τα δέντρα αποφάσεων (*decision trees*) και οι τεχνικές συσταδοποίησης (*cluster analysis*).

Στην εργασία αυτή χρησιμοποιούνται τεχνικές συσταδοποίησης ως περαιτέρω τεχνικές για την ομαδοποίηση των δεδομένων και την εξαγωγή αποτελεσμάτων.

Οι τεχνικές συσταδοποίησης δε δίνουν εκτιμήσεις πιθανοτήτων μέσω υπολογισμού βαρών – σκορ των παρατηρήσεων. Αντίθετα, διαφέρουν από τις τεχνικές κατηγοριοποίησης ως προς την απουσία γνώσης της κλάσης στην οποία ανήκουν τα δεδομένα για αυτό το λόγο και η διαδικασία της μάθησης θεωρείται μη εποπτευόμενη. Σκοπός της εφαρμογής των τεχνικών συσταδοποίησης στη διαχείριση του πιστωτικού κινδύνου αποτελεί ο διαχωρισμός των παρατηρήσεων σε ομάδες (συστάδες), επιδιώκοντας η κάθε ομάδα να είναι κατά το δυνατόν πιο ομογενής όσον αφορά το επίπεδο ρίσκου αθέτησης αλλά και η κάθε ομάδα να διαφέρει σημαντικά και όσο γίνεται περισσότερο από τις υπόλοιπες ομάδες ρίσκου (Anderberg, 2014). Οι Henley & Hand (1996) εφάρμοσαν τη μέθοδο του κοντινότερου γείτονα και την ευκλείδεια απόσταση για να αναγνωρίσουν πρότυπα και να εκτιμήσουν την πιστοληπτική ικανότητα των δανειοληπτών. Ο Tsai (2014) εφάρμοσε τεχνικές συσταδοποίησης σε συνδυασμό με λογιστική

παλινδρόμηση, δέντρα αποφάσεων και νευρωνικά δίκτυα για να προβλέψει την εταιρική αποτυχία. Οι τεχνικές συσταδοποίησης σε συνδυασμό με τα νευρωνικά δίκτυα είχαν την καλύτερη προβλεπτική ικανότητα και τα χαμηλότερα σφάλματα τύπου I και II. Γενικότερα οι τεχνικές συσταδοποίησης έχουν και αυτές ευρεία εφαρμογή στη διαχείριση του πιστωτικού κινδύνου (Lim & Sohn, 2007; Hsieh & Hung, 2010).

Συμπερασματικά, η εφαρμογή όλων των παραπάνω μοντέλων πιστωτικού σκορ σε διάφορους τομείς της λογιστικής και της χρηματοοικονομικής αναπτύσσεται ολοένα και περισσότερο (Abdou & Pointon, 2011; Ben-David & Frank, 2009; Laha, 2007) και συνοψίζεται στα ακόλουθα πεδία εφαρμογής:

α) Στην πρόβλεψη και την κατηγοριοποίηση των χρεοκοπιών των τραπεζών (Min & Lee, 2008; Nanni & Lumini, 2009),

β) Στα θέματα κατηγοριοποίησης ρίσκου πελατών επιχειρήσεων και ειδικότερα αξιολόγησης δανείων πελατών τραπεζών (Crook, et al., 2007; Leonard, 1995; Steenackers & Goovaerts, 1989),

γ) Στα ζητήματα χορήγησης εταιρικών δανείων κυρίως σε μικρές επιχειρήσεις με τη χρήση αριθμοδεικτών (Bensic, et al., 2005)

δ) Στα ζητήματα συσταδοποίησης πελατών τραπεζών (Lim & Sohn, 2007)

ε) Στα προβλήματα ανίχνευσης απάτης μέσω χρήσης πιστωτικών καρτών (Leonard, 1993; Quah & Sriganesh, 2008),

στ) Στα ζητήματα διαχείρισης ενυπόθηκων δανείων (Haughwout, et al., 2008; Heuson, et al., 2001),

ζ) Στα προβλήματα αποτελεσματικής λήψης χρηματοοικονομικών αποφάσεων μέσω ανίχνευσης χρηματοοικονομικών δυσκολιών (*financial distress*) (Beynon, 2005; Glen, 2001; Hu & Ansell, 2007).

Σε πολλές περιπτώσεις στη βιβλιογραφία η επιλογή των αρχικών μεταβλητών, δηλαδή των μεταβλητών που εισέρχονται αρχικά το μοντέλο, πραγματοποιείται με στατιστική ανάλυση και κυρίως με τη χρήση της *stepwise* λογιστικής παλινδρόμησης, της γραμμικής παλινδρόμησης κ.α. (Orgler, 1971; Steenackers & Goovaerts, 1989; Lee & Chen, 2005). Παρά τη χρήση των τεχνικών αυτών, σε ένα ευρύτατο φάσμα της βιβλιογραφίας δεν τεκμηριώνεται από τους ερευνητές για ποιο λόγο επιλέγουν το συγκεκριμένο δείγμα των μεταβλητών που χρησιμοποιούν στα μοντέλα. Στις περισσότερες περιπτώσεις, οι ερευνητές επισημαίνουν πως η επιλογή των μεταβλητών εξαρτάται από τη διαθεσιμότητα των δεδομένων, τον πάροχο, τη χώρα, την οικονομία, την κοινωνία την οποία αφορά το μοντέλο, και γενικότερα τη φύση των δεδομένων και του μοντέλου που επιθυμείται να δομηθεί. Για αυτό και η χρήση του συνόλου

των μεταβλητών που χρησιμοποιούν θεωρείται εκ των πραγμάτων σημαντική (Abdou & Pointon, 2011). Επίσης θεωρείται ότι δεν υπάρχει βέλτιστο πλήθος μεταβλητών που θα χρησιμοποιηθούν σε ένα scoring μοντέλο. Σε αρκετές χρηματοοικονομικές εφαρμογές των μοντέλων ο αριθμός των μεταβλητών ποικίλει. Έτσι υπάρχουν ερευνητές που χρησιμοποίησαν μόνο τρεις μεταβλητές (Pendharkar, 2005; Fletcher & Goss, 1993), άλλοι έως και είκοσι περίπου μεταβλητές (Jo, et al., 1997; Desai, et al., 1996), οι Salchenberger et al. (1992) χρησιμοποίησαν είκοσι εννέα μεταβλητές ενώ οι Leshno & Spector (1996) σαράντα μία. Τα τελευταία χρόνια τα μοντέλα που εφαρμόζονται μπορεί να περιέχουν πολύ μεγαλύτερο πλήθος μεταβλητών καθώς η εξέλιξη της ισχύος των υπολογιστών επιτρέπει τη διενέργεια πολύπλοκων υπολογισμών σε ελάχιστο χρόνο.

Επίσης στη βιβλιογραφία γίνεται λόγος για το μέγεθος του δείγματος που χρησιμοποιείται στα μοντέλα που διαχειρίζονται τον πιστωτικό κίνδυνο. Πιστεύεται ότι όσο μεγαλύτερο είναι το δείγμα που χρησιμοποιείται τόσο καλύτερη είναι και η προβλεπτική ικανότητα και η ακρίβεια του μοντέλου. Στον τομέα της χρηματοοικονομικής, στις πρώτες έρευνες που πραγματοποιήθηκαν ο αριθμός των παρατηρήσεων που περιλήφθηκαν στο δείγμα ήταν περίπου πενήντα (Fletcher & Goss, 1993), ενώ στις περισσότερες τελευταίες αναλύσεις χρησιμοποιούνται κάποιες χιλιάδες παρατηρήσεων (Bensic, et al., 2005; Hsieh, 2004). Υπάρχουν βέβαια και έρευνες που βασίστηκαν σε δείγμα λιγότερο των χιλίων εκατό παρατηρήσεων (Sustersic, et al., 2009; Lee & Chen, 2005).

ΚΕΦΑΛΑΙΟ 4

Θεωρία Στατιστικών Μεθόδων που χρησιμοποιούνται στην Ανάλυση

Το κεφάλαιο αυτό περιλαμβάνει τη θεωρητική προσέγγιση των μοντέλων που χρησιμοποιήθηκαν στη στατιστική ανάλυση των δεδομένων για την εξαγωγή των τελικών αποτελεσμάτων και συμπερασμάτων. Συνοπτικά οι μέθοδοι που χρησιμοποιούνται κατά σειρά εφαρμογής είναι οι ακόλουθες:

1. Γενικευμένα Γραμμικά Μοντέλα (Λογιστική Παλινδρόμηση και μοντέλο probit)
2. Ανάλυση Κυρίων Συνιστωσών
3. Ανάλυση κατά Συστάδες

Για την κάθε μία από τις παραπάνω στατιστικές μεθόδους ακολουθεί η περιγραφή του μοντέλου της και το μαθηματικό της υπόδειγμα.

4.1 *Logit και Probit – Γενικευμένα Γραμμικά Μοντέλα*

Στα πλαίσια των γενικευμένων γραμμικών μοντέλων εξετάζονται εξαρτημένες μεταβλητές που μπορεί να αφορούν είτε σε δεδομένα ομαδοποιημένα, οπότε μιλάμε για διωνυμικά δεδομένα (*binomial data*), είτε σε δεδομένα μη ομαδοποιημένα, οπότε μιλάμε για δίτιμα δεδομένα (*binary data*) όπου για κάθε άτομο στο δείγμα γνωρίζουμε την τιμή απόκρισής του (0 = επιτυχία, 1 = αποτυχία). Τα δεδομένα που αναλύονται στη διπλωματική αυτή είναι δίτιμα δεδομένα αφού για κάθε εταιρία στο δείγμα υπάρχει η τιμή απόκρισης (0 = ενεργή, 1 = σε διακοπή).

Στα δίτιμα και τα διωνυμικά δεδομένα δεν μπορεί να εφαρμοστεί το μοντέλο γραμμικής παλινδρόμησης για τους εξής λόγους: α) δεν μπορούμε να χρησιμοποιήσουμε κανονική προσέγγιση στη διωνυμική κατανομή $Bi(n, p_i)$ όταν $n = 1$, β) η διακύμανση της εξαρτημένης μεταβλητής Y_i ισούται με $p_i * (1 - p_i)$ που δεν είναι σταθερή, γ) η εκτιμώμενη τιμή $\hat{p}_i = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j X_{ij}$ είναι πιθανόν να μην ανήκει στο διάστημα $[0, 1]$ όπως θα έπρεπε.

Για τους λόγους αυτούς, για την επεξεργασία των δίτιμων και διωνυμικών δεδομένων χρησιμοποιούνται άλλα γραμμικά μοντέλα που διαφοροποιούνται ανάλογα με τη συνάρτηση σύνδεσης (συνάρτηση η οποία συνδέει το στοχαστικό τμήμα του μοντέλου δηλαδή τη μέση τιμή της τυχαίας μεταβλητής Y με το μη στοχαστικό τμήμα δηλαδή το γραμμικό συνδυασμό των ερμηνευτικών μεταβλητών X_j) που χρησιμοποιούν.

Τα πιο γνωστά μοντέλα επεξεργασίας αυτών των δεδομένων είναι το μοντέλο **logit**, δηλαδή η **λογιστική παλινδρόμηση**, και το μοντέλο **probit**. Τα δύο μοντέλα διαφέρουν ως προς τη συνάρτηση σύνδεσης στην οποία βασίζονται ενώ δίνουν σχεδόν τις ίδιες εκτιμήσεις για τις παραμέτρους των μοντέλων τους.

Η συνάρτηση σύνδεσης για το μοντέλο **probit** είναι η ακόλουθη:

$$\eta_i = Probit(p_i) = \Phi^{-1}(p_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

όπου Φ είναι η αθροιστική συνάρτηση κατανομής της τυποποιημένης κανονικής.

Στο μοντέλο **logit** χρησιμοποιείται ως συνάρτηση σύνδεσης ο λογάριθμος του λόγου της σχετικής πιθανότητας (*odds ratio*) του ενδεχομένου που μας ενδιαφέρει («επιτυχία»). Η βασική εξίσωση του μοντέλου είναι της μορφής:

$$\eta_i = logit(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

όπου p_i είναι η πιθανότητα επιτυχίας στην τιμή απόκρισης, $1 - p_i$ είναι η πιθανότητα αποτυχίας στην τιμή απόκρισης και X_1, X_2, \dots, X_k οι ανεξάρτητες ποσοτικές ή ποιοτικές μεταβλητές.

Από τα παραπάνω προκύπτει ότι η εκτίμηση της σχετικής πιθανότητας επιτυχίας ισούται με:

$$p_i = \frac{e^{\sum_{j=0}^k \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^k \beta_j x_{ij}}}$$

Η εκτίμηση των παραμέτρων $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ και στα δύο μοντέλα γίνεται με τη μέθοδο της μέγιστης πιθανοφάνειας (*maximum likelihood*). Για δίτιμα δεδομένα, όπως είναι τα δεδομένα της εργασίας αυτής, με τη μέθοδο της μέγιστης πιθανοφάνειας το διάνυσμα των παραμέτρων:

$$\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)$$

εκτιμάται μεγιστοποιώντας τη συνάρτηση πιθανοφάνειας, ή ισοδύναμα το λογάριθμό της, επιλύοντας δηλαδή το σύστημα:

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_i} = 0 \text{ για } i = 1, 2, \dots, k$$

Το $100(1 - \alpha)100\%$ διάστημα εμπιστοσύνης για τις παραμέτρους είναι το:

$$\hat{\beta} \pm z_{\alpha/2} * s.e.(\hat{\beta})$$

όπου $s.e.(\hat{\beta})$ το τυπικό σφάλμα του εκτιμητή της μέγιστης πιθανοφάνειας. Το διάστημα αυτό είναι προσεγγιστικό (ασυμπτωτικό για μεγάλα διαστήματα).

Η έννοια της σχετικής πιθανότητας είναι πολύ σημαντική για την ερμηνεία των παραμέτρων του μοντέλου καθώς καθιστά την ερμηνεία των αποτελεσμάτων εύκολη και διαισθητική. Ανάμεσα στην πιθανότητα επιτυχίας και το λογάριθμο της σχετικής πιθανότητας υπάρχει 1 – 1 αντιστοιχία γιατί ισχύει ότι:

$$p = \frac{odds}{1 + odds}$$

Για παράδειγμα σε ένα μοντέλο λογιστικής παλινδρόμησης με δύο ανεξάρτητες μεταβλητές η σχέση:

$$\log\left(\frac{\hat{p}}{1 - \hat{p}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

μπορεί να γραφτεί ως:

$$\hat{p} = \frac{e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2)}}{1 + e^{(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2)}}$$

Από την παραπάνω σχέση διακρίνεται ότι η μοναδιαία αύξηση της τιμής της μεταβλητής x_1 προκαλεί πολλαπλασιαστική αύξηση της σχετικής πιθανότητας επιτυχίας κατά $\exp(\hat{\beta}_1)$ όταν η μεταβλητή x_2 είναι σταθερή.

4.2 Ανάλυση Κυρίων Συνιστωσών

Η Ανάλυση Κυρίων Συνιστωσών (*Principal Component Analysis*) αποτελεί τεχνική των μοντέλων της Πολυμεταβλητής Στατιστικής Ανάλυσης που βασικό στόχο έχει τη μείωση των διαστάσεων των πολυμεταβλητών δεδομένων. Αυτό πρακτικά σημαίνει πως με τη μέθοδο επιδιώκεται να αντικατασταθεί ένα σύνολο μεταβλητών X_1, X_2, \dots, X_p με ένα σημαντικά μικρότερο στο πλήθος σύνολο μεταβλητών F_1, F_2, \dots το οποίο θα αποτελείται από γραμμικούς συνδυασμούς των αρχικών μεταβλητών και θα διατηρεί ένα σημαντικό μέρος των πληροφοριών που περιέχει το αρχικό σύνολο. Ως «πληροφορία» νοείται το ποσοστό της συνολικής διασποράς που μεταφέρεται στο πλήθος των μεταβλητών F_1, F_2, \dots . Η πρώτη μεταβλητή F_1 ονομάζεται πρώτη κύρια συνιστώσα και περιέχει τη μέγιστη δυνατή πληροφορία που μπορεί να μεταφερθεί από το σύνολο των συνιστωσών που θα χρησιμοποιηθούν. Η επόμενη μεταβλητή στη σειρά η F_2 ονομάζεται δεύτερη κύρια συνιστώσα, περιέχει λιγότερη πληροφορία από την πρώτη, καλύπτει μέρος της εναπομένουσας διασποράς που δεν μεταφέρθηκε στην πρώτη συνιστώσα και δεν συσχετίζεται με την πρώτη συνιστώσα. Ανάλογη ερμηνεία αποδίδεται και στις υπόλοιπες στη σειρά συνιστώσες.

Η Ανάλυση Κυρίων Συνιστωσών χρησιμοποιείται στις περιπτώσεις που υπάρχει επαρκής συσχέτιση μεταξύ των αρχικών μεταβλητών για να εξασφαλιστεί η αναπαράστασή τους στις κύριες συνιστώσες.

Στόχος της Ανάλυσης Κυρίων Συνιστωσών είναι να βρεθούν οι συνιστώσες $z = [z_1, z_2, \dots, z_p]$ που αποτελούν γραμμικούς συνδυασμούς των αρχικών μεταβλητών $x = [x_1, x_2, \dots, x_p]$ και επιτυγχάνουν να αποτυπώσουν την μεγαλύτερη δυνατή διακύμανση. Η χαρακτηριστική εξίσωση η οποία μεγιστοποιεί τη διακύμανση είναι η $\mathbf{z} = \mathbf{x}\alpha$ τέτοια ώστε $\alpha'\alpha = \mathbf{1}$ όπου $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_p]'$. Αναλυτικότερα το μαθηματικό υπόδειγμα του προβλήματος της μεγιστοποίησης έχει ως εξής:

Αν θεωρήσουμε δείγμα n ατόμων από έναν πληθυσμό και για το κάθε άτομο παρατηρούμε ($p \geq 2$) χαρακτηριστικά (τυχαίες μεταβλητές) X_1, X_2, \dots, X_p τότε ο πίνακας $X = (x_{ij})$:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

θα περιέχει στη θέση (i, j) την τιμή x_{ij} που καταγράφηκε στο i άτομο για το χαρακτηριστικό j . Θεωρώντας τη μεταβλητή Y ως τη συνιστώσα που θα αποτελεί γραμμικό συνδυασμό των αρχικών μεταβλητών X_1, X_2, \dots, X_p αυτή θα έχει τη μορφή:

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p$$

Η ανάλυση εστιάζει στον προσδιορισμό των $\alpha_1, \alpha_2, \dots, \alpha_p \in \Re$ ώστε οι τιμές (scores) των n ατόμων για τη μεταβλητή Y , δηλαδή τα:

$$y_i = \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_p x_{ip} \quad i = 1, 2, \dots, n$$

να «διατηρούν» όσο το δυνατόν περισσότερο τις αποστάσεις που έχουν τα άτομα από τις αρχικές μεταβλητές. Αυτό θα επιτευχθεί αν μεγιστοποιηθεί η ποσότητα:

$$SS_Y = \sum_{r=1}^n \sum_{i=r+1}^n (y_r - y_i)^2$$

Η μεταβλητότητα του συνόλου των τιμών y_1, y_2, \dots, y_n συμβολίζεται με $Dis(Y)$ και ισούται με:

$$Dis(Y) = \sum_{i=1}^n (y_i - \bar{y})^2$$

Κατόπιν υπολογισμών προκύπτει ότι:

$$SS_Y = nDIS(Y)$$

Το πρόβλημα της μεγιστοποίησης της SS_Y είναι ισοδύναμο με την εύρεση των αριθμών $\alpha_1, \alpha_2, \dots, \alpha_p$ για τους οποίους μεγιστοποιείται η ποσότητα $Dis(Y)$. Η ποσότητα αυτή μπορεί να γραφτεί αλλιώς με τη βοήθεια του παραπάνω πίνακα τιμών των δεδομένων, δηλαδή του πίνακα $X = (x_{ij})$ των p μεταβλητών X_1, X_2, \dots, X_p ($p \geq 2$) για n άτομα. Έστω ότι συμβολίζουμε με:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad j = 1, 2, \dots, p$$

τους δειγματικούς μέσους για τις τιμές του κάθε χαρακτηριστικού X_j . Επίσης από τον παραπάνω ορισμό των y_i θα ισχύει ότι:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (a_1 x_{i1} + a_2 x_{i2} + \dots + a_p x_{ip})$$

Από όπου προκύπτει ότι:

$$\bar{y} = a_1 \bar{x}_1 + a_2 \bar{x}_2 + \dots + a_p \bar{x}_p$$

Αφαιρώντας κατά μέλη την παραπάνω ισότητα από την αρχική ισότητα των y_i προκύπτει ότι:

$$y_i - \bar{y} = a_1(x_{i1} - \bar{x}_1) + a_2(x_{i2} - \bar{x}_2) + \dots + a_p(x_{ip} - \bar{x}_p) \quad i = 1, 2, \dots, n$$

Η σχέση αυτή μπορεί να γραφτεί ισοδύναμα ως:

$$f_i = y_i - \bar{y} = \mathbf{z}'_i \mathbf{a} \quad i = 1, 2, \dots, n$$

όπου z_i, a συμβολίζονται τα διανύσματα στήλης:

$$z_i = \begin{bmatrix} x_{i1} - \bar{x}_1 \\ x_{i2} - \bar{x}_2 \\ \vdots \\ x_{ip} - \bar{x}_p \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

Επίσης:

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}, \quad \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}, \quad \bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

Όπου:

- Οι αριθμοί $f_i = y_i - \bar{y}, i = 1, 2, \dots, n$ συμβολίζουν την απόκλιση του score y_i του i ατόμου από το μέσο score \bar{y} . Το διάνυσμα f ονομάζεται διάνυσμα κεντρικοποιημένων score.
- Το διάνυσμα γραμμή $\mathbf{x}'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ περιέχει τις παρατηρήσεις που αφορούν το i άτομο για κάθε μια από τις p μεταβλητές.
- Το διάνυσμα $\bar{\mathbf{x}}$ αποτελεί τους δειγματικούς μέσους ανά μεταβλητή.

Βάσει των παραπάνω μπορεί να γραφτεί ότι:

$$z_i = x_i - \bar{x}, \quad y_i = \mathbf{x}'_i \mathbf{a}, \quad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} = \begin{bmatrix} z'_1 \mathbf{a} \\ z'_2 \mathbf{a} \\ \vdots \\ z'_n \mathbf{a} \end{bmatrix} = \begin{bmatrix} z'_1 \\ z'_2 \\ \vdots \\ z'_n \end{bmatrix} \mathbf{a} = \mathbf{Z} \mathbf{a}$$

όπου με \mathbf{Z} συμβολίζεται ο πίνακας:

$$\mathbf{Z} = \begin{bmatrix} z'_1 \\ z'_2 \\ \vdots \\ z'_n \end{bmatrix} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p \end{bmatrix}$$

Συνεπώς η ποσότητα $Dis(Y)$ που θα μεγιστοποιηθεί για να δώσει το σύνολο των συνιστωσών ισούται με:

$$Dis(Y) = \sum_{i=1}^n f_i^2 = \mathbf{f}' \mathbf{f} = (\mathbf{Z} \mathbf{a})' (\mathbf{Z} \mathbf{a}) = \mathbf{a}' \mathbf{Z}' \mathbf{Z} \mathbf{a}$$

Η ποσότητα αυτή αποτελεί τη διασπορά του συνόλου N των παρατηρήσεων κατά μήκος του διανύσματος \mathbf{a} και συμβολίζεται με $Dis_{\mathbf{a}}(N)$.

Η διασπορά αυτή, που επιχειρείται να μεγιστοποιηθεί, μπορεί να αυξάνει απεριόριστα. Για να περιοριστεί η δυνατότητα αυτή ο συνήθης περιορισμός που τίθεται για το διάνυσμα \mathbf{a} είναι να έχει μέτρο ίσο με τη μονάδα (μοναδιαίο διάνυσμα). Αυτό γιατί η προσέγγιση του προβλήματος που εξετάζουμε γίνεται με τη χρήση Ευκλείδειων αποστάσεων. Συνεπώς θα ισχύει ότι:

$$\|\mathbf{a}\| = 1 \Leftrightarrow \mathbf{a}' \mathbf{a} = 1 \Leftrightarrow \sum_{i=1}^p a_i^2 = 1$$

Το πρόβλημα το οποίο τελικά θα επιλυθεί είναι το ακόλουθο:

Για δεδομένο πίνακα Z διάστασης $n \times p$ να βρεθεί μοναδιαίο διάνυσμα $\mathbf{a} = (\alpha_1, \alpha_2, \dots, \alpha_p)'$ το οποίο μεγιστοποιεί την τετραγωνική μορφή $Dis_a(Y) = Dis(Y) = \mathbf{a}'Z'Z\mathbf{a}$.

Η λύση του προβλήματος αυτού προκύπτει με τη χρήση των ιδιοτιμών και των ιδιοδιανυσμάτων του πίνακα $Z'Z$. Ο πίνακας $Z'Z$ έχει μη αρνητικές ιδιοτιμές, έστω τις $\lambda_1, \lambda_2, \dots, \lambda_p$. Αν θεωρήσουμε ότι $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ και αν συμβολίσουμε u_1, u_2, \dots, u_p τα αντίστοιχα μοναδιαία ιδιοδιανύσματα τότε:

1. Το διάνυσμα \mathbf{a} που μεγιστοποιεί την $\mathbf{a}'Z'Z\mathbf{a}$ είναι το μοναδιαίο διάνυσμα u_1 που αντιστοιχεί στη μεγαλύτερη ιδιοτιμή λ_1 .
2. Η μέγιστη τιμή της τετραγωνικής μορφής είναι ίση με λ_1 και ισχύει ότι:

$$\max_{\|\mathbf{a}\|=1} Dis_a(N) = Dis_{u_1}(N) = \lambda_1$$

Η μεταβλητή Y για την οποία επιτυγχάνεται η παραπάνω μεγιστοποίηση αποτελεί την **πρώτη κύρια συνιστώσα (first principal component)**

Οι ιδιοτιμές λ αποτελούν τη διακύμανση της κάθε συνιστώσας y που αντιπροσωπεύουν. Όσο περισσότερες συνιστώσες χρησιμοποιούνται τόσο μεγαλύτερο μέρος της συνολικής διακύμανσης ερμηνεύεται από τις συνιστώσες. Το άθροισμα του συνόλου των ιδιοτιμών λ που θα προκύψουν, το πλήθος των οποίων ισούται με το πλήθος p των αρχικών μεταβλητών, αποτελεί τη συνολική διακύμανση των αρχικών δεδομένων. Η διακύμανση αυτή ισούται με το ίχνος των στοιχείων του πίνακα $Z'Z$ (άθροισμα των στοιχείων της κύριας διαγωνίου). Ισχύει δηλαδή ότι:

$$Dis(N) = tr(Z'Z) = \sum_{j=1}^p \lambda_j = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Καθώς στόχο της Ανάλυσης Κυρίων Συνιστωσών αποτελεί η μείωση των διαστάσεων των δεδομένων, θα πρέπει να επιτευχθεί μια «χρυσή τομή» μεταξύ του πλήθους των συνιστωσών που τελικά θα χρησιμοποιηθούν και του πλήθους των συνιστωσών που θεωρείται επαρκές για την ερμηνεία του μεγαλύτερου ποσού της συνολικής μεταβλητότητας. Τα πιο συνηθισμένα κριτήρια διατήρησης κυρίων συνιστωσών είναι τα ακόλουθα:

i. Ποσοστό συνολικής διακύμανσης που ερμηνεύουν οι συνιστώσες.

Συνήθως επιλέγεται το πλήθος των συνιστωσών που ερμηνεύει αθροιστικά μεγαλύτερο ποσοστό διακύμανσης από το όριο που τίθεται. Είναι εύκολο στην πράξη αλλά δεν δίνει πάντα τα επιθυμητά αποτελέσματα ειδικά στην περίπτωση τίθεται αρκετά υψηλό όριο και άρα απαιτείται η διατήρηση μεγάλου αριθμού συνιστωσών.

ii. Κριτήριο του Kaiser

Σύμφωνα με το κριτήριο αυτό θα πρέπει να διατηρηθούν οι συνιστώσες που η ιδιοτιμή τους είναι μεγαλύτερη από 1. Διαισθητικά, ο κανόνας αυτός υποδεικνύει ότι η συνιστώσα που θα διατηρηθεί θα πρέπει να ενσωματώνει τουλάχιστον τόση διακύμανση όση και η διακύμανση που περιέχεται σε μια οποιαδήποτε αρχική μεταβλητή.

iii. Ποσοστό της διακύμανσης που ερμηνεύεται για κάθε αρχική μεταβλητή.

Σύμφωνα με το κριτήριο αυτό τίθεται όριο στο ποσό της διακύμανσης της αρχικής μεταβλητής που θα πρέπει να ερμηνεύει η κάθε συνιστώσα. Το πλήθος των συνιστωσών που θα επιλεγεί θα εξασφαλίζει ότι για κάθε μεταβλητή ερμηνεύεται ποσοστό τουλάχιστον ίσο με το προκαθορισμένο όριο που τέθηκε. Και αυτή η μέθοδος μπορεί να μη δίνει καλά αποτελέσματα καθώς το όριο που τίθεται είναι υποκειμενικό και αν κάποια μεταβλητή δεν ερμηνεύεται από τις πρώτες συνιστώσες θα απαιτηθεί διατήρηση μεγάλου αριθμού συνιστωσών.

iv. Scree plot

Το scree plot είναι γράφημα το οποίο αναπαριστά τις ιδιοτιμές με βάση τη σειρά μεγέθους τους. Η επιλογή του πλήθους των συνιστωσών γίνεται στο σημείο όπου το γράφημα γίνεται περίπου οριζόντιο διατηρώντας τόσες συνιστώσες όσες υποδεικνύονται από το σημείο αυτό.

4.3 Ανάλυση κατά Συστάδες

Η Ανάλυση κατά Συστάδες (*Cluster Analysis*) αποτελεί και αυτή κομμάτι της Πολυμεταβλητής Ανάλυσης. Βασικός της στόχος είναι να εξετάσει κατά πόσο είναι όμοιες κάποιες παρατηρήσεις ως προς κάποιον ή κάποιους παράγοντες. Απώτερος σκοπός είναι να δημιουργηθούν συστάδες (ομάδες) από εκείνες τις παρατηρήσεις που ομοιάζουν μεταξύ τους, δηλαδή οι παρατηρήσεις μέσα σε κάθε ομάδα να είναι όσο γίνεται πιο ομοιογενείς, ενώ οι παρατηρήσεις διαφορετικών ομάδων να διαφέρουν όσο γίνεται περισσότερο. Με αυτόν τον τρόπο επιτυγχάνεται η ευκολότερη και αποδοτικότερη επεξεργασία των δεδομένων που διατίθενται.

4.3.1 Μέτρα Απόστασης

Για να μπορέσει να πραγματοποιηθεί η διαδικασία ομαδοποίησης θα πρέπει να υπάρχουν οι κατάλληλες ποσότητες που θα μπορούν να χρησιμοποιηθούν για να υποδείξουν ότι δύο άτομα (παρατηρήσεις) είναι όμοια ή ανόμοια μεταξύ τους. Τέτοια μέτρα είναι οι αποστάσεις (*distances*). Παρατηρήσεις που μοιάζουν πολύ μεταξύ τους θα πρέπει να δίνουν πολύ μικρή τιμή στην **απόσταση** και επομένως θα τοποθετούνται στην ίδια ομάδα. Συνεπώς στόχο της ανάλυσης κατά συστάδες αποτελεί η δημιουργία ομάδων μέσα στις οποίες οι παρατηρήσεις απέχουν λίγο ενώ οι παρατηρήσεις διαφορετικών ομάδων να απέχουν μεταξύ τους αρκετά. Η παρακάτω ανάλυση αφορά τον υπολογισμό των αποστάσεων μεταξύ ατόμων στα οποία οι παρατηρήσεις τους περιγράφονται με ποσοτικές μεταβλητές. Στη διπλωματική αυτή η Ανάλυση κατά Συστάδες θα χρησιμοποιηθεί μόνο για τις ποσοτικές μεταβλητές του συνόλου των δεδομένων.

Όπως και στην Ανάλυση των Κύριων Συνιστωσών, αν θεωρήσουμε δείγμα n ατόμων από έναν πληθυσμό και για το κάθε άτομο παρατηρούμε ($p \geq 2$) χαρακτηριστικά (τυχαίες μεταβλητές) X_1, X_2, \dots, X_p τότε ο πίνακας $X = (x_{ij})$:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

θα περιέχει στη θέση (i, j) την τιμή x_{ij} που καταγράφηκε στο i άτομο για το χαρακτηριστικό j . Αν $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ είναι το διάνυσμα των παρατηρήσεων που αφορούν το i άτομο ($i = 1, 2, \dots, n$) για τις p μεταβλητές τότε το πιο γνωστό μέτρο απόστασης μεταξύ δύο παρατηρήσεων $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ και $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})$ είναι η Ευκλείδεια Απόσταση και ορίζεται από τον τύπο:

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

Όταν η κλίμακα μέτρησης είναι διαφορετική μεταξύ των μεταβλητών τότε η ευκλείδεια απόσταση δεν ενδείκνυται ως καλό μέτρο απόστασης. Αντί αυτής χρησιμοποιείται η απόσταση του Pearson. Η απόσταση αυτή ισούται με:

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

όπου s_r είναι η διακύμανση της r μεταβλητής και ισούται με:

$$s_r = \left[\frac{1}{n-1} \sum_{i=1}^n (x_{ir} - \bar{x}_r)^2 \right]^{1/2}, \quad \bar{x}_r = \frac{1}{n} \sum_{i=1}^n x_{ir}$$

Υπάρχουν και άλλα είδη αποστάσεων ενδεικτικότερες από τις οποίες είναι οι ακόλουθες:

- i. **Απόσταση Mahalanobis:** $d_{ij} = (x_i - x_j)\Sigma^{-1}(x_i - x_j)'$ όπου Σ^{-1} είναι ο δειγματικός πίνακας διακύμανσης – συνδιακύμανσης.
- ii. **Απόσταση Manhattan ή City – block metric:** $d_{ij} = \sum_{r=1}^p |x_{ir} - x_{jr}|$
- iii. **Απόσταση Minkowski:** $d_{ij} = \left(\sum_{r=1}^p |x_{ir} - x_{jr}|^\lambda \right)^{1/\lambda}$ όπου $\lambda \geq 1$ δεδομένη παράμετρος.
- iv. **Απόσταση max ή Chebyshev:** $d_{ij} = \max_{r=1,2,\dots,p} |x_{ir} - x_{jr}|$
- v. **Απόσταση του Gower:** $d_{ij} = \sum_{r=1}^p \frac{|x_{ir} - x_{jr}|}{R_r}$ όπου $R_r = \max_{i=1,2,\dots,n} x_{ir} - \min_{i=1,2,\dots,n} x_{ir}$

- vi. **Απόσταση Bhattacharyya:** $d_{ij} = \left(\sum_{r=1}^p (\sqrt{x_{ir}} - \sqrt{x_{jr}})^2 \right)^{1/2}$
- vii. **Canberra metric:** $d_{ij} = \sum_{r=1}^p \frac{|x_{ir} - x_{jr}|}{|x_{ir}| + |x_{jr}|}$

4.3.2 Μέτρα Ομοιότητας

Τα μέτρα ομοιότητας (*similarity measures* ή *affinity measures*) υποδεικνύουν και αυτά αν δύο άτομα (παρατηρήσεις) είναι όμοια ή ανόμοια μεταξύ τους. Όταν οι παρατηρήσεις ομοιάζουν πολύ μεταξύ τους τα μέτρα ομοιότητας λαμβάνουν μεγάλες τιμές, ενώ για ανόμοιες παρατηρήσεις τα μέτρα λαμβάνουν πολύ μικρές τιμές. Συνεπώς τα ζεύγη παρατηρήσεων τοποθετούνται στην ίδια ή σε διαφορετική ομάδα ανάλογα με το αν η τιμή του μέτρου είναι μεγάλη ή μικρή αντίστοιχα. Το πιο γνωστό μέτρο ομοιότητας για τις ποσοτικές παρατηρήσεις είναι ο (δειγματικός) συντελεστής συσχέτισης που δίνεται από τον τύπο:

$$s_{ij} = \frac{\sum_{r=1}^p (x_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j)}{\left(\sum_{r=1}^p (x_{ir} - \bar{x}_i)^2 \sum_{k=1}^p (x_{jr} - \bar{x}_j)^2 \right)^{1/2}}$$

όπου $\bar{x}_i = \frac{1}{p} \sum_{r=1}^p x_{ir}$, $\bar{x}_j = \frac{1}{p} \sum_{r=1}^p x_{jr}$

4.3.3 Μέθοδοι Ομαδοποίησης

Στην παράγραφο αυτή θα γίνει αναφορά στις μεθόδους ομαδοποίησης που έχουν μαθηματική και στατιστική βάση. Οι μέθοδοι αυτοί, ανάλογα με τον τρόπο που προχωρούν στη διαμόρφωση των ομάδων διακρίνονται σε μη ιεραρχικές και ιεραρχικές μεθόδους.

Μη ιεραρχικές μέθοδοι

Στις μη ιεραρχικές μεθόδους ο αριθμός των ομάδων που θα δημιουργηθούν θεωρείται γνωστός εκ των προτέρων γι αυτό και καθορίζεται πριν την επαναληπτική διαδικασία του αλγορίθμου. Οι μέθοδοι αυτοί έχουν καλή εφαρμογή σε μεγάλα δείγματα καθώς σχηματίζουν ομάδες παρεμφερούς μεγέθους, επηρεάζονται όμως αρκετά από το αρχικό πλήθος που θέτει ο αναλυτής. Οι αλγόριθμοι υλοποίησης των μη ιεραρχικών μεθόδων βασίζονται στην έννοια του κέντρου μιας ομάδας (κέντρο βάρους, centroid) το οποίο αποτελεί τη μέση τιμή για κάθε

μεταβλητή όλων των παρατηρήσεων της ομάδας. Κατά την επαναληπτική διαδικασία, οι παρατηρήσεις κατατάσσονται σε ομάδες ανάλογα με την απόστασή τους από τα κέντρα όλων των ομάδων. Αυτό σημαίνει ότι ο αλγόριθμος υπολογίζει την απόσταση της κάθε παρατήρησης από τα κέντρα των ομάδων που έχουν ήδη δημιουργηθεί και κατόπιν κατατάσσει την παρατήρηση στην ομάδα που βρίσκεται πιο κοντά, δηλαδή στην ομάδα με κέντρο πιο κοντά στην παρατήρηση. Οι μέθοδοι διαφοροποιούνται ανάλογα με το σημείο όπου γίνεται η ανανέωση των κέντρων των ομάδων και η ταξινόμηση των υπόλοιπων παρατηρήσεων σε αυτές. Ο υπολογισμός των αποστάσεων συνήθως βασίζεται στην ευκλείδεια απόσταση, χωρίς να αποκλείεται η χρήση άλλων αποστάσεων όπως αναφέρθηκαν παραπάνω.

Η πιο γνωστή και πιο αντιπροσωπευτική τεχνική ομαδοποίησης αυτής της κατηγορίας είναι η μέθοδος **MacQueen** ή **k-means method**. Στη μέθοδο αυτή το κέντρο βάρους των ομάδων υπολογίζεται σύμφωνα με τη σύνθεση που προκύπτει αφού ταξινομηθεί ένα άτομο και όχι αφού ταξινομηθούν όλα τα άτομα (μέθοδος *Forgy*). Η ολοκλήρωση της διαδικασίας ομαδοποίησης απαιτεί σχετικά λίγες επαναλήψεις γεγονός που καθιστά τον αλγόριθμο ιδιαίτερα χρήσιμο στη διαχείριση μεγάλων συνόλων δεδομένων. Συνήθως οι ομάδες που δημιουργούνται τελικά περιέχουν ίσο περίπου αριθμό παρατηρήσεων. Τα μειονεκτήματα που παρουσιάζει ο αλγόριθμος είναι τα ακόλουθα:

- Ο αλγόριθμος επηρεάζεται σημαντικά από την επιλογή των αρχικών σημείων κατά την εκκίνηση της διαδικασίας γεγονός που μπορεί να οδηγήσει σε εντελώς διαφορετική ομαδοποίηση από την υπάρχουσα φυσική ομαδοποίηση των δεδομένων. Για το λόγο αυτό ενδείκνυται ο αλγόριθμος να τρέχει με διάφορες επιλογές για να αποφευχθεί κατά το δυνατόν η παγίδευση σε μη βέλτιστη λύση.
- Λόγω ύπαρξης έκτροπων παρατηρήσεων ο αλγόριθμος μπορεί να οδηγήσει σε ομάδες με πολύ διασπαρμένα άτομα.
- Υπάρχει πιθανότητα να γνωρίζουμε εκ των προτέρων ότι η φυσική ομαδοποίηση του πληθυσμού που μελετάμε να αποτελείται από k ομάδες και κατά την εφαρμογή της διαδικασίας της ομαδοποίησης κάποια από αυτές τις ομάδες να μην αντιπροσωπεύεται τελικά. Για την αποφυγή αυτού ο αλγόριθμος εφαρμόζεται πολλές φορές για διαφορετικά k και πραγματοποιείται σύγκριση των αποτελεσμάτων λαμβάνοντας υπόψη και τη διαίσθηση του αναλυτή.

Ιεραρχικές μέθοδοι

Σε αντίθεση με τις μη ιεραρχικές μεθόδους, στις ιεραρχικές μεθόδους οι ομάδες σχηματίζονται σταδιακά παράγοντας μια ιεραρχία δενδροειδούς μορφής όπου στα διάφορα

στάδια της διαδικασίας το πλήθος k των ομάδων παίρνει όλες τις δυνατές τιμές από το 1 έως το n . Οι ιεραρχικές μέθοδοι δεν ενδείκνυνται για μεγάλο πλήθος δεδομένων καθώς απαιτούν πολύ χρόνο, μνήμη και ισχύ.

Ανάλογα με τη διαδικασία που ακολουθείται οι μέθοδοι διακρίνονται **διαιρετικές** και σε **συσσωρευτικές**.

Στις **διαιρετικές** μεθόδους οι ομάδες σχηματίζονται με διαίρεση των ομάδων σε μικρότερες μέχρις ότου κάθε παρατήρηση θα αποτελεί από μόνη της μια ομάδα. Ουσιαστικά οι αλγόριθμοι αυτοί βρίσκουν υποομάδες των ήδη διαμορφωμένων ομάδων που είναι περισσότερο απομακρυσμένες και τις διαχωρίζουν. Η πρακτική χρησιμότητά τους είναι σχετικά μικρή καθώς απαιτούν πολύ περισσότερους υπολογισμούς από ότι οι συσσωρευτικές μέθοδοι και στα πλαίσια αυτής της διπλωματικής δεν θα χρησιμοποιηθούν.

Στις **συσσωρευτικές** μεθόδους οι ομάδες σχηματίζονται με συνένωση μικρότερων ομάδων σχηματίζοντας συνεχώς μεγαλύτερες ομάδες μέχρις ότου όλα τα δεδομένα να υπάρχουν σε μια ομάδα. Οι ιεραρχικές μέθοδοι είναι σχεδόν συνυφασμένες με τις συσσωρευτικές μεθόδους. Στην εφαρμογή των μεθόδων αυτών χρησιμοποιείται όχι η απόσταση μεταξύ στοιχείων αλλά η απόσταση μεταξύ ομάδων που έχουν προέλθει είτε από συγχώνευση άλλων ομάδων είτε από συγχώνευση παρατηρήσεων.

Παρακάτω αναφέρονται συνοπτικά οι αντιπροσωπευτικότερες μέθοδοι προσδιορισμού των παραπάνω αποστάσεων, ο γενικός τύπος ορισμού απόστασης των ομάδων που χρησιμοποιούν, και τέλος ο γενικός τύπος ορισμού απόστασης που εφαρμόζει ο αλγόριθμος μετά τη συγχώνευση δύο ομάδων για να ανανεώσει τον πίνακα αποστάσεων D (όταν δηλαδή διαγράφει τις γραμμές και στήλες που αντιστοιχούσαν στις ομάδες που συγχωνεύτηκαν, και προσθέτει μια γραμμή και μια στήλη που περιέχει τις αποστάσεις της νέας ομάδας από τις υπόλοιπες):

i. Μέθοδος της απλής συνένωσης (Single Linkage Method)

Η μέθοδος αυτή είναι η απλούστερη και πιο παλιά όλων των ιεραρχικών μεθόδων. Στη μέθοδο αυτή η απόσταση μεταξύ δύο ομάδων ορίζεται ως η μικρότερη απόσταση από μια παρατήρηση μέσα στη μια ομάδα με κάποια παρατήρηση στην άλλη ομάδα (μέθοδος του πλησιέστερου γείτονα – *nearest neighbor method*). Έτσι αν έχουμε τις ομάδες R και Q τότε η απόστασή τους $d(R, Q)$ ορίζεται από τον τύπο:

$$d(R, Q) = \min_{i \in R, j \in Q} d_{ij}$$

Η μέθοδος δεν έχει τη δυνατότητα να κρατήσει ξεχωριστά δύο εμφανώς διαφορετικές ομάδες που όμως έχουν κάποιο σημείο ή ένα σύνολο σημείων που τις «συνδέει», και τις συγχωνεύει.

Όταν δύο ομάδες A και B ενωθούν για να δημιουργήσουν μια νέα ομάδα, έστω $R=(A,B)$, τότε όλες οι αποστάσεις μεταξύ της R και των άλλων ομάδων, έστω Q , προκύπτουν από την παρακάτω απόσταση που χρησιμοποιείται στην ανανέωση του πίνακα αποστάσεων:

$$d(R, Q) = \min\{d(A, Q), d(B, Q)\} = \frac{1}{2}(d(A, Q) + d(B, Q)) - \frac{1}{2}|d(A, Q) - d(B, Q)|$$

Από δεδομένα προσομοίωσης έχει διαπιστωθεί ότι η μέθοδος αυτή παρουσιάζει τη χειρότερη επίδοση ως προς τη δημιουργία κατάλληλης ομαδοποίησης των δεδομένων.

ii. Μέθοδος της πλήρους συνένωσης (*Complete Linkage Method*)

Η μέθοδος αυτή εφαρμόζει την αντίθετη έκφραση της απόστασης δύο ομάδων σε σχέση με τη μέθοδο του πλησιέστερου γείτονα. Έτσι η απόσταση ανάμεσα σε δύο ομάδες ορίζεται ως η μεγαλύτερη απόσταση από μια παρατήρηση μέσα στην μια ομάδα με κάποια παρατήρηση στην άλλη ομάδα. Ο τύπος της απόστασης των ομάδων είναι:

$$d(R, Q) = \max_{i \in R, j \in Q} d_{ij}$$

Ο τύπος ανανέωσης των αποστάσεων για τη μέθοδο αυτή έχει τη μορφή:

$$d(R, Q) = \max\{d(A, Q), d(B, Q)\} = \frac{1}{2}(d(A, Q) + d(B, Q)) + \frac{1}{2}|d(A, Q) - d(B, Q)|$$

Η μέθοδος αυτή δημιουργεί συνήθως μεγάλες και συμπαγείς ομάδες, ενώ αρκετά συχνά δεν δύναται να ξεχωρίσει κάποιες πολύ συμπαγείς μικρές ομάδες.

iii. Μέθοδος των σταθμισμένων μέσων (*Weighted Average Linkage Method*)

Στη μέθοδο αυτή η απόσταση ανάμεσα στις ομάδες εκφράζεται ως ο μέσος των αποστάσεων όλων των στοιχείων της μιας ομάδας με τα στοιχεία της άλλης. Ο τύπος ανανέωσης των αποστάσεων για τη μέθοδο αυτή είναι της μορφής:

$$d(R, Q) = \frac{|A|d(A, Q) + |B|d(B, Q)}{|A| + |B|}$$

Από δεδομένα προσομοίωσης έχει διαπιστωθεί ότι η μέθοδος αυτή επιτυγχάνει πολύ καλές ομαδοποιήσεις των δεδομένων.

iv. Μέθοδος των κέντρων βάρους (*Centroid Method*)

Η μέθοδος αυτή θεωρεί ως απόσταση των ομάδων την απόσταση των κέντρων των ομάδων. Αν x_{ir} είναι η r μέτρηση του i αντικειμένου x_i , όπου το i παίρνει τιμές από 1 έως n και το r από 1 έως p , το κέντρο βάρους μιας ομάδας R θα είναι το σημείο:

$$\bar{x}(R) = (\bar{x}_1(R), \bar{x}_2(R), \dots, \bar{x}_p(R))$$

το οποίο έχει ως r συντεταγμένη την:

$$\bar{x}_r(R) = \frac{1}{|R|} \sum_{i \in R} x_{ir} \quad \text{για } r = 1, 2, \dots, p$$

Η απόσταση μεταξύ των ομάδων R και Q ορίζεται ως η ευκλείδεια απόσταση μεταξύ των κέντρων βάρους τους, δηλαδή:

$$d(R, Q) = d(\bar{x}(R), \bar{x}(Q)) = \sqrt{\sum_{r=1}^p (\bar{x}_r(R) - \bar{x}_r(Q))^2}$$

Ο τύπος ανανέωσης του πίνακα αποστάσεων είναι ο ακόλουθος:

$$d^2(R, Q) = \frac{|A|}{|R|} d^2(A, Q) + \frac{|B|}{|R|} d^2(B, Q) - \frac{|A||B|}{|R|^2} d^2(A, B)$$

Η μέθοδος αυτή έχει αρκετές καλές ιδιότητες και οι ομάδες που σχηματίζει είναι συνήθως συμπαγείς και ελλειπτικές. Ένα μειονέκτημα είναι ότι μπορεί να εφαρμοστεί μόνο σε ποιοτικά δεδομένα.

v. Μέθοδος του Ward (*Ward's method*)

Η συγκεκριμένη μέθοδος προσδοκεί να ελαχιστοποιήσει τη διακύμανση μέσα στις ομάδες. Για κάθε παρατήρηση υπολογίζεται η ευκλείδεια απόστασή της από το κέντρο της ομάδας. Στη συνέχεια αθροίζονται οι αποστάσεις αυτές για όλα τα άτομα μιας ομάδας C και λαμβάνεται η παρακάτω ποσότητα που αποτελεί τη συνεκτικότητα των στοιχείων της ομάδας:

$$ESS(C) = \sum_{i \in C} (d(x_i, \bar{x}(C)))^2$$

Η ποσότητα αυτή ονομάζεται άθροισμα των τετραγωνικών αποκλίσεων (τετραγώνων των αποστάσεων) της ομάδας C . Αν υπάρχουν k ομάδες το συνολικό άθροισμα των τετραγωνικών αποκλίσεων θα είναι ίσο:

$$ESS = ESS_1 + ESS_2 + \dots + ESS_k$$

Καθώς στην αρχή κάθε παρατήρηση αποτελεί και μια ομάδα το συνολικό άθροισμα είναι ίσο με το 0. Σε κάθε βήμα συγχωνεύονται εκείνες οι δύο ομάδες που η ομαδοποίησή τους οδηγεί στη μικρότερη αύξηση του ESS δηλαδή επιτυγχάνεται η ελάχιστη απώλεια πληροφορίας. Στο τελικό βήμα όταν και τα n άτομα θα έχουν συγχωνευτεί σε μια ομάδα το ESS θα ισούται με:

$$ESS = \sum_{j=1}^N (x_j - \bar{x})' (x_j - \bar{x})$$

όπου x_j είναι η πολυδιάστατη μέτρηση του j -οστού ατόμου και \bar{x} είναι η μέση τιμή όλων των ατόμων.

Η απόσταση μεταξύ των ομάδων είναι η ευκλείδεια απόσταση των κέντρων βαρών τους:

$$d^2(R, Q) = \frac{2|R||Q|}{|R| + |Q|} d^2(\bar{x}(R), \bar{x}(Q))$$

Ενώ ο τύπος ανανέωσης του πίνακα αποστάσεων είναι:

$$d^2(R, Q) = \frac{|A| + |Q|}{|R| + |Q|} d^2(A, Q) + \frac{|B| + |Q|}{|R| + |Q|} d^2(B, Q) - \frac{|Q|}{|R| + |Q|} d^2(A, B)$$

Η μέθοδος χρησιμοποιείται στην πράξη πολύ συχνά γιατί έχει μερικές πολύ καλές ιδιότητες και συνήθως δημιουργεί ομάδες με παρόμοιο αριθμό παρατηρήσεων.

ΚΕΦΑΛΑΙΟ 5

Μεθοδολογία και Αποτελέσματα

Στην ενότητα αυτή θα παρουσιαστεί η μεθοδολογία που ακολουθήθηκε για τον καθαρισμό των δεδομένων, η στατιστική ανάλυση που πραγματοποιήθηκε στα τελικά δεδομένα μετά τον καθαρισμό και τα αποτελέσματα που προέκυψαν σε κάθε βήμα της ανάλυσης.

5.1 Καθαρισμός και προετοιμασία των δεδομένων για τη στατιστική ανάλυση

Στην υποενότητα αυτή θα παρουσιασθεί η βήμα προς βήμα διαδικασία που ακολουθήθηκε προκειμένου τα δεδομένα να μην περιέχουν ανακρίβειες και λογικά σφάλματα και να αποτελέσουν ακολούθως επαρκή και ικανή βάση για τη στατιστική ανάλυση.

A. Διαχείριση δεδομένων απόκρισης

Ένα από τα σημαντικότερα ζητήματα για την ανάλυση των δεδομένων και την εξαγωγή ορθών αποτελεσμάτων αποτέλεσε η εύρεση των εταιριών που διέκοψαν τη λειτουργία τους, εταιρίες που ουσιαστικά αποτέλεσαν την «απόκριση» στην στατιστική επεξεργασία της ανάλυσης. Οι εταιρίες που συνιστούν το σύνολο των δεδομένων ελέγχθηκαν μία προς μία. Για κάθε εταιρία που εντοπίστηκε πως έχει διακόψει τη λειτουργία της χρησιμοποιήθηκε ως έτος διακοπής το τελευταίο έτος για το οποίο υπάρχουν διαθέσιμα δεδομένα στην ιστοσελίδα από την οποία αντλήθηκαν τα δεδομένα, ανεξάρτητα από το πραγματικό έτος διακοπής της εταιρίας. Αυτό έγινε για να καταστεί δυνατή η εφαρμογή των στατιστικών μοντέλων. Η παραμετροποίηση αυτή δεν εμπόδισε την εξαγωγή ορθών αποτελεσμάτων καθώς το πραγματικό έτος διακοπής, στις περιπτώσεις που αυτό δεν συνέπιπτε με το τελευταίο έτος διαθέσιμων στοιχείων, δεν απέκλινε παραπάνω του ενός έτους.

B. Διαχείριση αρνητικών τιμών των μη αρνητικών πρωτογενών δεδομένων

Τα πρωτογενή δεδομένα περιέχουν μεγέθη τα οποία από τη φύση τους δεν δύνανται να λαμβάνουν αρνητικές τιμές. Τέτοια μεγέθη αποτελούν το πάγιο και το λοιπό ενεργητικό, τα μακροπρόθεσμα και τα βραχυπρόθεσμα χρέη, οι πωλήσεις (τζίρος) και το μεικτό κέρδος. Για τα μεγέθη αυτά πραγματοποιήθηκαν έλεγχοι αρνητικών τιμών, και στις περιπτώσεις όπου διαπιστώθηκε ύπαρξη αρνητικού ποσού, αυτό αντικαταστάθηκε με το μηδέν. Κατόπιν πραγματοποιήθηκε ο υπολογισμός των δευτερογενών μεγεθών.

C. Διαχείριση μηδενικών τιμών των παρονομαστών κλασμάτων

Αμέσως μετά τον υπολογισμό των δευτερογενών μεγεθών ακολούθησε ο υπολογισμός των αριθμοδεικτών. Στην περίπτωση που ο παρονομαστής του αριθμοδείκτη ήταν ίσος με το μηδέν, ο αριθμοδείκτης τέθηκε ίσος με το μηδέν. Αυτό γιατί, ανεξάρτητα του πώς εκφράζεται ο αριθμοδείκτης, είτε α/β είτε β/α , η φύση του και το οικονομικό του περιεχόμενο δεν αλλάζει παρά μόνο η ερμηνεία του. Συνεπώς αν στον αριθμοδείκτη α/β ο παρονομαστής β ήταν ίσος με το μηδέν αυτός θα ισούταν με το άπειρο. Αν ο αριθμοδείκτης εκφραζόταν στην αντίστροφη μορφή του τότε θα ισούταν με το μηδέν. Για το λόγο αυτό πραγματοποιήθηκε η παραπάνω παραμετροποίηση.

D. Διαχείριση ελλιπών τιμών

Πραγματοποιήθηκε έλεγχος των ελλιπών τιμών (*missing values*) των παρατηρήσεων που δεν αποτελούν παρατήρηση «απόκρισης». Για την καλύτερη εφαρμογή των μοντέλων και για την εξασφάλιση της μέγιστης δυνατής χρονοσειράς που διατίθεται, διαγράφηκαν οι παρατηρήσεις που αφενός δεν αποτελούσαν παρατήρηση απόκρισης και αφετέρου δεν είχαν ολοκληρωμένη εικόνα δεδομένων στο “*panel*” τους, δηλαδή δεν είχαν δεδομένα για όλα τα έτη από το 2011 έως 2014. Κριτήριο ολοκληρωμένης εικόνας “*panel*” αποτέλεσε η μεταβλητή «Πωλήσεις (Τζίρος)» καθώς μια εταιρία όταν έχει δραστηριότητα αυτομάτως έχει και ποσό τζίρου. Συνεπώς, οι παρατηρήσεις που απέμειναν προς επεξεργασία σε αυτό το στάδιο αποτελούνταν από τις παρατηρήσεις απόκρισης και τις παρατηρήσεις που διέθεταν δεδομένα τζίρου, και άρα και τα υπόλοιπα μεγέθη, για όλα τα έτη από το 2011 έως και το 2014.

E. Διαχείριση υψηλά συσχετισμένων αριθμοδεικτών

Καθώς η στατιστική ανάλυση στηρίζεται στους αριθμοδείκτες, πραγματοποιήθηκε ανάλυση συσχέτισης για το σύνολο των αριθμοδεικτών με σκοπό την εύρεση εκείνων των ζευγών των

αριθμοδεικτών που σχετίζονται υψηλά. Οι αριθμοδείκτες που θεωρήθηκαν ως υψηλά συσχετισμένοι μεταξύ τους ήταν αυτοί που ανά ζεύγη είχαν ποσοστό συσχέτισης πάνω από 90%. Από τους αριθμοδείκτες αυτούς, προκειμένου να αποφευχθούν προβλήματα πολυσυγγραμμικότητας στην ανάλυση, διατηρήθηκε ο ένας εκ των δύο με βάση τη σημαντικότητά του ως προς το οικονομικό του περιεχόμενο.

Εν τέλει, από το σύνολο των αριθμοδεικτών διαγράφηκαν αυτοί που αναφέρονται στους 9 ακόλουθους δείκτες: Περιθώριο EBIT, Περιθώριο Λοιπού Ενεργητικού, EBIT προς Ενεργητικό, Κέρδη προ Φόρων προς Ενεργητικό, Καθαρά Κέρδη προς Ενεργητικό, Καθαρά Κέρδη και Τόκοι προς Ενεργητικό, Κέρδη προ Φόρων προς Ίδια, Καθαρά Κέρδη προς Ίδια, Κόστος Πωληθέντων και Λειτουργικά Έξοδα προς Πωλήσεις.

Η επιλογή του δείκτη που διατηρήθηκε βασίστηκε κατά περίπτωση ως ακολούθως:

1. Βαθμός συσχέτισης **«Περιθωρίου EBITDA» και «Περιθωρίου EBIT»: 0,9515.**

Μεταξύ των δύο, διατηρήθηκε ο δείκτης «Περιθώριο EBITDA». Γενικά, το μέγεθος του EBITDA χρησιμοποιείται εκτενέστερα στη βιβλιογραφία για την εκτίμηση της κερδοφορίας και τις αποτιμήσεις των εταιριών. Η διαφορά του με τον EBIT εναπόκειται στην αφαίρεση των αποσβέσεων. Η μέθοδος των αποσβέσεων που χρησιμοποιεί κάθε επιχείρηση μπορεί να διαφέρει από κλάδο σε κλάδο και από επιχείρηση σε επιχείρηση. Το ποσό των αποσβέσεων δεν επηρεάζεται από τη λειτουργία της επιχείρησης ενώ επηρεάζει την συνολική της κερδοφορία καθώς αποτελεί έξοδο. Για τους λόγους αυτούς, το μέγεθος των κερδών πριν την αφαίρεση των αποσβέσεων είναι προτιμότερο του μεγέθους των κερδών μετά την αφαίρεση αυτών, καθώς αποτελεί αντιπροσωπευτικότερο μέγεθος των λειτουργικών κερδών της επιχείρησης.

2. Βαθμός συσχέτισης **«Περιθωρίου Ενεργητικού» και «Περιθωρίου Λοιπού Ενεργητικού»: 0,9214.**

Μεταξύ των δύο, διατηρήθηκε ο δείκτης «Περιθώριο Ενεργητικού». Αυτό γιατί περικλείει μεγαλύτερο μέρος οικονομικής πληροφορίας αφού περιλαμβάνει το σύνολο του Ενεργητικού.

3. Βαθμός συσχέτισης **«EBITDA προς Ενεργητικό» και «EBIT προς Ενεργητικό»: 0,9994.**

Μεταξύ των δύο, διατηρήθηκε ο δείκτης «EBITDA προς Ενεργητικό» για τους ίδιους λόγους με την 1^η περίπτωση.

4. Βαθμός συσχέτισης **«EBITDA προς Ενεργητικό»** και **«Κέρδη προ Φόρων προς Ενεργητικό»**: **0,9660**.

Μεταξύ των δύο, διατηρήθηκε ο δείκτης **«EBITDA προς Ενεργητικό»**. Αυτό γιατί ο EBITDA διαφέρει με τα Κέρδη προ Φόρων ως προς το ποσό των αποσβέσεων και των πληρωτέων τόκων. Όπως ακριβώς και οι αποσβέσεις, οι πληρωτέοι τόκοι αποτελούν έξοδο για την επιχείρηση και επηρεάζουν την κερδοφορία της. Επίσης, δεν εξαρτώνται από τη λειτουργία της επιχείρησης καθώς ορίζονται από τους πιστωτές και τους όρους της δανειακής σύμβασης.

5. Βαθμός συσχέτισης **«EBITDA προς Ενεργητικό»** και **«Καθαρά Κέρδη προς Ενεργητικό»**: **0,9648**.

Μεταξύ των δύο, διατηρείται ο δείκτης **«EBITDA προς Ενεργητικό»**. Αυτό γιατί, όπως συμβαίνει με τις αποσβέσεις και τους πληρωτέους τόκους, το μέγεθος του φόρου επηρεάζει την κερδοφορία, ο υπολογισμός του δεν εξαρτάται από την επιχείρηση, και διαφέρει από έτος σε έτος ανάλογα με τους φορολογικούς συντελεστές που ορίζει η νομοθεσία.

6. Βαθμός συσχέτισης **«EBITDA προς Ενεργητικό»** και **«Καθαρά Κέρδη και Τόκοι προς Ενεργητικό»**: **0,9981**

Μεταξύ των δύο, διατηρείται ο δείκτης **«EBITDA προς Ενεργητικό»** για τους ίδιους λόγους που αναφέρθηκαν και παραπάνω.

7. Βαθμός συσχέτισης **«EBIT προς Ίδια Κεφάλαια»** και **«Κέρδη προ Φόρων προς Ίδια Κεφάλαια»**: **0,9518**

Μεταξύ των δύο, διατηρείται ο δείκτης **«EBIT προς Ίδια Κεφάλαια»** για τους ίδιους λόγους που αναφέρθηκαν και παραπάνω.

8. Βαθμός συσχέτισης **«EBIT προς Ίδια Κεφάλαια»** και **«Καθαρά Κέρδη προς Ίδια Κεφάλαια»**: **0,9482**

Μεταξύ των δύο, διατηρείται ο δείκτης **«EBIT προς Ίδια Κεφάλαια»** για τους ίδιους λόγους που αναφέρθηκαν και παραπάνω.

9. Βαθμός συσχέτισης «Λειτουργικά Έξοδα προς Πωλήσεις» και «Κόστος Πωληθέντων και Λειτουργικά Έξοδα προς Πωλήσεις»: **0,9992**

Μεταξύ των δύο, διατηρείται ο δείκτης «Κόστος Πωληθέντων και Λειτουργικά Έξοδα προς Πωλήσεις» γιατί ο δείκτης περιλαμβάνει περισσότερη οικονομική πληροφορία. Βέβαια, οποιοσδήποτε εκ των δύο και να διατηρούνταν θα ήταν ακριβώς το ίδιο καθώς η συσχέτιση των δύο δεικτών είναι σχεδόν 100%.

Ε. Έλεγχος πολυσυγγραμμικότητας με διαγνωστικό κριτήριο τον Παράγοντα

Διόγκωσης Διακύμανσης

Ο Παράγοντας Διόγκωσης Διακύμανσης (*Variance Inflation Factor – VIF*) αποτελεί έναν δείκτη που χρησιμοποιείται ευρέως ως διαγνωστικό κριτήριο για την ύπαρξη πολυσυγγραμμικότητας. Ο δείκτης αυτός ορίζεται για κάθε ανεξάρτητη μεταβλητή $X_k, k = 1, 2, \dots, p - 1$ ενός μοντέλου πολλαπλής παλινδρόμησης με p παραμέτρους και ισούται με:

$$VIF_k = \frac{1}{1 - R_k^2}, \quad k = 1, 2, \dots, p - 1$$

όπου R_k^2 είναι ο συντελεστής προσδιορισμού του μοντέλου που χρησιμοποιεί ως εξαρτημένη μεταβλητή την X_k και ως ανεξάρτητες τις υπόλοιπες $p - 2$ ανεξάρτητες μεταβλητές $X_j, j \neq k$.

Αν συμβεί η μεταβλητή X_k να εκφράζεται με ικανοποιητική ακρίβεια ως γραμμικός συνδυασμός των υπόλοιπων ανεξάρτητων μεταβλητών, ο συντελεστής προσδιορισμού R_k^2 θα λάβει τιμή κοντά στο 1 με αποτέλεσμα ο δείκτης VIF_k να γίνεται υπερβολικά μεγάλος. Αντίστροφα, αν συμβεί να ισχύει $VIF_k \cong 1$ θα ισχύει ότι $R_k^2 = 0$ οπότε η X_k δεν μπορεί να εκφρασθεί γραμμικά μέσω των $X_j, j \neq k$. Εξαιτίας αυτών χρησιμοποιούνται συνήθως οι εξής κανόνες:

- Αν $VIF_k \cong 1$, η αντίστοιχη ανεξάρτητη μεταβλητή X_k δεν έχει πρόβλημα πολυσυγγραμμικότητας (σε σχέση με τις υπόλοιπες).
- Αν $VIF_k > 10$ τότε η X_k εμφανίζει πρόβλημα πολυσυγγραμμικότητας με τις υπόλοιπες ανεξάρτητες μεταβλητές.

Για την εύρεση πολυσυγγραμμικότητας στους αριθμοδείκτες πραγματοποιήθηκε η ακόλουθη ανάλυση:

- i. Πραγματοποιήθηκε γραμμική παλινδρόμηση με εξαρτημένη μεταβλητή τον αριθμοδείκτη «Μεικτό Περιθώριο» και ανεξάρτητες του υπόλοιπους αριθμοδείκτες που απέμειναν μέχρι και αυτό το βήμα καθαρισμού των δεδομένων.
- ii. Μετά την γραμμική παλινδρόμηση υπολογίστηκε ο δείκτης VIF για τις ανεξάρτητες μεταβλητές. Οι μεταβλητές που παρουσίασαν τιμή VIF άνω του 10 είναι οι ακόλουθες:

Μεταβλητή	VIF	1/VIF
Περιθώριο EBITDA	1842.64	0.000543
Κόστος Πωληθέντων και Λειτουργικά προς Πωλήσεις	1840.96	0.000543
EBITDA προς Ενεργητικό	27.1	0.036904
Ίδια Κεφάλαια προς Ενεργητικό	23.59	0.042395
Ξένα Κεφάλαια προς Ενεργητικό	22.85	0.04377

Από τα παραπάνω αποτελέσματα διαπιστώνεται πάρα πολύ μεγάλη, και μάλιστα σχεδόν ίδια, τιμή του VIF για τους δύο πρώτους αριθμοδείκτες. Στο επόμενο βήμα θα πραγματοποιηθεί εκ νέου γραμμική παλινδρόμηση με τις ίδιες μεταβλητές του βήματος 1 πλην της μιας εκ των δύο μεταβλητών με το πολύ υψηλό VIF. Η μεταβλητή που επιλέγεται να αποκλειστεί είναι το «Κόστος Πωληθέντων και Λειτουργικά προς Πωλήσεις» και αυτό γιατί το «Περιθώριο EBITDA» χρησιμοποιείται ευρύτερα στη βιβλιογραφία και αποτελεί πολύ σημαντικότερο μέγεθος στην αξιολόγηση των επιχειρήσεων.

- iii. Μετά τη γραμμική παλινδρόμηση που πραγματοποιήθηκε με ανεξάρτητη μεταβλητή τον αριθμοδείκτη «Μεικτό Περιθώριο» και ανεξάρτητες του υπόλοιπους αριθμοδείκτες πλην του «Κόστος Πωληθέντων και τα Λειτουργικά προς Πωλήσεις» υπολογίστηκε ο δείκτης VIF για τις ανεξάρτητες μεταβλητές. Οι μεταβλητές που παρουσίασαν τιμή VIF άνω του 10 είναι οι ακόλουθες:

Μεταβλητή	VIF	1/VIF
EBITDA προς Ενεργητικό	27.05	0.036969
Ίδια Κεφάλαια προς Ενεργητικό	23.56	0.042443
Ξένα Κεφάλαια προς Ενεργητικό	22.84	0.043775

Παρατηρείται πως ο αριθμοδείκτης «Περιθώριο EBIDTA», που προηγουμένως παρουσίαζε πολύ μεγάλο VIF, τώρα δεν σημειώνει τιμή του δείκτη άνω του 10 (για την ακρίβεια η τιμή του είναι 1,96). Στο σημείο αυτό, ο αριθμοδείκτης που θα αποκλειστεί από την επόμενη παλινδρόμηση θα είναι ο «EBIDTA προς Ενεργητικό», καθώς έχει το μεγαλύτερο VIF από τους 3 παραπάνω αριθμοδείκτες.

- iv. Μετά την παλινδρόμηση που πραγματοποιήθηκε με ανεξάρτητη μεταβλητή τον αριθμοδείκτη «Μεικτό Περιθώριο» και ανεξάρτητες του υπόλοιπους αριθμοδείκτες του προηγούμενου βήματος πλην του «EBIDTA προς Ενεργητικό» καμία μεταβλητή δεν παρουσίασε VIF άνω του 10.

Συγκεντρωτικά από την παραπάνω διαδικασία αφαιρούνται οι δείκτες «Κόστος Πωληθέντων και τα Λειτουργικά προς Πωλήσεις» και «EBIDTA προς Ενεργητικό».

G. Διαχείριση ακραίων τιμών

Στις μεταβλητές των αριθμοδεικτών, ως ακραίες τιμές θεωρήθηκαν οι τιμές που περιέχονταν κάτω από το όριο του 0,5% και πάνω από το όριο του 99,5% της κατανομής της εκάστοτε μεταβλητής. Ουσιαστικά, εξαιρέθηκε ισόποσα το 0,5% της αριστερής ουράς και το 0,5% της δεξιάς ουράς της κατανομής του κάθε αριθμοδείκτη, και το εναπομείναν ποσοστό 99% των τιμών του θεωρήθηκε ως μη ακραίο.

Όσον αφορά στις παρατηρήσεις – εταιρίες που δεν αποτελούν παρατήρηση απόκρισης, στην περίπτωση που έστω και μια μεταβλητή τους περιείχε τιμή που ενέπιπτε εκτός του επιτρεπόμενου περιθωρίου, δηλαδή η τιμή της θεωρούταν ακραία, τότε για την παρατήρηση αυτή διαγράφηκε ολόκληρο το panel των τεσσάρων ετών με όλες του τις μεταβλητές.

Όσον αφορά στις παρατηρήσεις – εταιρίες που αποτελούν παρατήρηση απόκρισης, επειδή θεωρήθηκε πιο πιθανό οι παρατηρήσεις αυτές να περιέχουν στις μεταβλητές τους ακραίες τιμές, ακολουθήθηκε διαφορετική διαδικασία. Αρχικά δημιουργήθηκαν βοηθητικές μεταβλητές οι οποίες λάμβαναν την τιμή 1 κάθε φορά που η τιμή μιας μεταβλητής θεωρούνταν ακραία. Στη συνέχεια, για την κάθε παρατήρηση αθροίστηκαν οι τιμές των βοηθητικών μεταβλητών για κάθε έτος αλλά και για όλα τα έτη συνολικά. Εν τέλει, οι μεταβλητές απόκρισης που παρέμειναν προς ανάλυση περιείχαν στο panel τους το πολύ 6 ακραίες παρατηρήσεις που μπορεί να βρίσκονταν σε ένα οποιοδήποτε έτος και να αφορούσαν μια οποιαδήποτε μεταβλητή.

Το σύνολο των δεδομένων που προέκυψε μετά από τη διαδικασία αυτή ανήλθε στο ποσό των 3.076 εταιριών που δεν περιείχαν σε κανένα έτος και καμία μεταβλητή καμία ακραία παρατήρηση και σε 134 εταιρίες απόκρισης εκ των οποίων: 82 εταιρίες δεν περιείχαν καμία ακραία τιμή, 15 εταιρίες που περιείχαν 1 ακραία τιμή κάπου στο panel τους, 16 εταιρίες που περιείχαν 2 ακραίες τιμές κάπου στο panel τους, 9 εταιρίες που περιείχαν 3 ακραίες τιμές, 5 εταιρίες που περιείχαν 4 ακραίες τιμές, 3 εταιρίες που περιείχαν 5 ακραίες τιμές και 4 εταιρίες που περιείχαν 6 ακραίες τιμές.

5.2 Στατιστική Ανάλυση – Εφαρμογή μοντέλων

Στην ενότητα αυτή θα παρουσιαστούν τα αποτελέσματα εφαρμογής των στατιστικών μεθόδων Πολυμεταβλητής Ανάλυσης που αναφέρθηκαν στην προηγούμενη ενότητα, δηλαδή της Λογιστικής Παλινδρόμησης, της ανάλυσης Probit, της Ανάλυσης Κύριων Συνιστωσών και της Ανάλυσης Συσταδοποίησης.

5.2.1 Εφαρμογή Λογιστικής Παλινδρόμησης

Παρακάτω ακολουθούν η παρουσίαση και η επεξήγηση όλων των αποτελεσμάτων και των εκτιμήσεων που σχετίζονται με την εφαρμογή της Λογιστικής Παλινδρόμησης.

A. Εκτίμηση και επεξήγηση του μοντέλου logit

Η φιλοσοφία της ανάλυσης βασίζεται στο να διερευνηθεί αν υπάρχουν αριθμοδείκτες των οποίων η τιμή κατά την προηγούμενη χρονική περίοδο ($t - 1$) μπορεί να χρησιμοποιηθεί ως προβλέπουσα του γεγονότος διακοπής – πτώχευσης της επιχείρησης την εξεταζόμενη χρονική περίοδο (t). Δηλαδή ουσιαστικά εξετάζεται αν η τιμή ενός αριθμοδείκτη μπορεί να δώσει «σινιάλο» για την εμφάνιση οικονομικού προβλήματος στην επιχείρηση ένα χρόνο πριν την πραγματοποίηση του προβλήματος.

Αναγκαστικά, εφόσον οι τιμές των δεδομένων αφορούν τα έτη 2011 έως 2014, τα μοντέλα που θα εφαρμοστούν στο σημείο αυτό θα περιλαμβάνουν τις επιχειρήσεις που πτώχευσαν από το 2012 και έπειτα. Οι τιμές των παρατηρήσεων που θα χρησιμοποιηθούν ως προβλέπουσες αφορούν τα έτη 2011 έως και 2013 και αποτελούν την υστέρηση (*lag*) ενός έτους του πραγματικού δείκτη. Παράδειγμα της μορφής των προβλεπουσών μεταβλητών αποτελεί η ακόλουθη:

Εταιρία	Έτος	Δείκτης	Lag Δείκτη
1	2011	50	.
1	2012	10	50
1	2013	15	10
1	2014	40	15
2	2011	20	.
2	2012	25	20
2	2013	30	25
2	2014	10	30

Μετά τον μετασχηματισμό αυτό, οι επιχειρήσεις που αποτελούν την απόκριση είναι πλέον στο σύνολό τους 84 ($y=1$) ενώ οι υπόλοιπες παρατηρήσεις είναι στο σύνολό τους 9.270 ($y=0$). Η αναλογία μεταξύ των τιμών αποκρίσεων προς το συνολικό δείγμα είναι 0,9%, ποσοστό πολύ μικρό για τη διενέργεια προβλέψεων και την εξαγωγή ασφαλών συμπερασμάτων. Στη βιβλιογραφία προτείνεται ως αποδεκτή αναλογία τιμών αποκρίσεων προς το συνολικό δείγμα των παρατηρήσεων ποσοστό περίπου ίσο με 50% (Abdou & Pointon, 2011; ICAP GROUP A.E, 2013). Για το λόγο αυτό ακολουθήθηκε η εξής διαδικασία:

Για την επίτευξη της αναλογίας του 50%, λήφθηκε δείγμα από το σύνολο των παρατηρήσεων το οποίο αποτελούταν σε ποσοστό 50% από το 100% των τιμών των αποκρίσεων ($n=84$) και σε ποσοστό 50% από τυχαίες παρατηρήσεις του συνόλου των 9.270 παρατηρήσεων που δεν αποτελούν απόκριση. Αυτή η τυχαία δειγματοληψία είναι στατιστικώς αποδεκτή, καθώς η λογιστική παλινδρόμηση, παρόλο που οι παρατηρήσεις αποτελούν panel data, διαχειρίζεται την εκάστοτε παρατήρηση ως ανεξάρτητη εγγραφή και όχι ως συσχετισμένη χρονικά με την προηγούμενη ή επόμενη εγγραφή της. Η διαδικασία αυτή πραγματοποιήθηκε 100 φορές, ενώ από την εφαρμογή της λογιστικής παλινδρόμησης σε κάθε δείγμα, προέκυψε η συχνότητα εμφάνισης των μεταβλητών που προέκυψαν στατιστικά σημαντικές σε επίπεδο σημαντικότητας περίπου 12%. Τα αποτελέσματα της διαδικασίας είχαν ως ακολούθως:

Πίνακας 5.2-1 Σημαντικές Μεταβλητές μετά τη Λογιστική Παλινδρόμηση

Όνομασία Μεταβλητής	Φορές Εμφάνισης ως Σημαντική στις 100 δοκιμές	Μέσος Όρος P-value στις φορές που εμφανίστηκε σημαντική
Ίδια Κεφάλαια προς Ενεργητικό	93	0.0305
Γεωγραφικός Τομέας – Επίπεδο 3	32	0.0953
Βραχυπρόθεσμο Χρέος προς Ξένα	20	0.0984
Απόσβεση Παγίων	4	0.1010
Σταθερά	31	0.1013
Πωλήσεις προς Ενεργητικό	37	0.1088
Μεικτό Περιθώριο Κέρδους	14	0.1094
Βραχυπρόθεσμο Χρέος προς Κεφάλαιο Κίνησης	31	0.1135
Ίδια προς Ξένα	18	0.1162
Γεωγραφικός Τομέας – Επίπεδο 2	11	0.1193
Λοιπό Ενεργητικό προς Ξένα	17	0.1199
Περιθώριο EBITDA	12	0.1252
Λοιπό Ενεργητικό προς Βραχυπρόθεσμο Χρέος	16	0.1259
Πωλήσεις προς Κεφάλαιο Κίνησης	17	0.1274
Μακροπρόθεσμο Χρέος προς Πάγιο	14	0.1282
Πωλήσεις προς Πάγιο Ενεργητικό	22	0.1287

Πωλήσεις προς Ίδια Κεφάλαια	17	0.1306
Ίδια προς Πάγια	9	0.1336
Περιθώριο Αποσβέσεων	11	0.1341
Καθαρά κέρδη προς Μακροπρόθεσμα και Ίδια	22	0.1343
Περιθώριο Ενεργητικού	9	0.1350
Πωλήσεις προς Ξένα	37	0.1356
Τόκοι προς EBIT	16	0.1443
Φορολογική Επιβάρυνση	28	0.1484
Καθαρό Περιθώριο Κέρδους	3	0.1497
EBIT προς Ίδια	4	0.1628
Μακροπρόθεσμο Χρέος προς Ίδια	3	0.1647

Από τα παραπάνω συμπεραίνεται ότι ο πιο σημαντικός αριθμοδείκτης, του οποίου η τιμή δύναται να δώσει σινιάλο διακοπής της επιχείρησης μια χρονική περίοδο πριν την πραγματοποίηση της διακοπής, είναι τα «**Ίδια Κεφάλαια προς Ενεργητικό**». Ο αριθμοδείκτης αυτός βρέθηκε σημαντικός 93 φορές στις 100 δοκιμές με μέσο *p-value* 0.0305 που σημαίνει ότι ο δείκτης είναι σημαντικός σε κατά μέσο όρο επίπεδο σημαντικότητας 3%. Επίσης ο «**Γεωγραφικός Τομέας**», ως κατηγορική μεταβλητή, θεωρείται σημαντικός, καθώς το επίπεδο 3 που εμφανίζεται περισσότερες φορές σημαντικό (32 φορές) συσχετίζεται με το επίπεδο 2 που εμφανίζεται λιγότερες φορές σημαντικό (11 φορές). Όσον αφορά τις υπόλοιπες μεταβλητές, επιλέχθηκαν 9 σημαντικές μεταβλητές, αυτές που βρίσκονται στο σκιασμένο κομμάτι του πίνακα. Κριτήριο επιλογής αποτέλεσε ο μέσος όρος της *p-value* που προέκυψε από τις τιμές *p-value* που σημειώθηκαν κατά τις φορές που οι μεταβλητές εμφανίστηκαν σημαντικές. Αποδεκτές έγιναν οι μεταβλητές που σημείωσαν μέσο όρο *p-value* κάτω από 12%.

Παρακάτω λαμβάνεται ενδεικτικό δείγμα παρατηρήσεων το οποίο θα χρησιμοποιηθεί και καθ' όλη την ανάλυση που θα επακολουθήσει. Θα πρέπει να σημειωθεί ότι στο δείγμα παρατηρήσεων που αναλύεται ακολούθως δεν εμφανίζονται σημαντικές όλες οι μεταβλητές που επιλέχθηκαν να χρησιμοποιηθούν στην ανάλυση. Τα αποτελέσματα της λογιστικής παλινδρόμησης για το επιλεγμένο δείγμα είναι τα ακόλουθα:

Iteration 0: log likelihood = -118.50185	
Iteration 1: log likelihood = -100.23496	
Iteration 2: log likelihood = -99.611029	
Iteration 3: log likelihood = -99.608894	
Iteration 4: log likelihood = -99.608894	
Logistic regression	Number of obs = 171
	LR chi2(10) = 37.79
	Prob > chi2 = 0.0000
	Pseudo R2 = 0.1594

Απόκριση	Coef. Std. Err. z P> z [95% Conf. Interval]
Γεωγραφ. Τομέας	

2	-.698835	.4191986	-1.67	0.095	-1.520449	.1227791
3	-1.092131	.4473632	-2.44	0.015	-1.968947	-.2153152
ΜΕΙΚΤΟ Περιθώριο	-.0000519	.0096867	-0.01	0.996	-.0190375	.0189338
Απόσβεση Παγίων	.0251079	.0163544	1.54	0.125	-.0069461	.057162
Ιδια/Ενεργητικό	-.0412251	.0094174	-4.38	0.000	-.059683	-.0227673
Ιδια/Ξένα	.0053022	.0019686	2.69	0.007	.0014438	.0091606
Λοιπό Ενεργ/Ξένα	-.0036861	.0021469	-1.72	0.086	-.0078938	.0005217
Βραχ Χρέος/Ξένα	.0052698	.0084244	0.63	0.532	-.0112417	.0217814
Πωλήσεις/Ενεργητικό	.0000362	.0026812	0.01	0.989	-.0052188	.0052913
Βραχ Χρέος/Κεφ.Κίνησης	1.52e-06	.0001565	0.01	0.992	-.0003053	.0003083
Σταθερά	1.285939	.8179877	1.57	0.116	-.3172878	2.889165

Εν συνεχεία ακολουθεί η επεξήγηση του αποτελέσματος της Λογιστικής Παλινδρόμησης.

a) Iteration log: (0, 1, 2, 3, 4)

Τα μεγέθη που βρίσκονται στο πάνω μέρος του αποτελέσματος, υπό τον τίτλο αυτό, αποτελούν τις εκτιμήσεις της μέγιστης πιθανοφάνειας σε κάθε επανάληψη του αλγορίθμου της λογιστικής παλινδρόμησης, καθώς η εκτίμηση της μέγιστης πιθανοφάνειας αποτελεί μια επαναληπτική διαδικασία. Η πρώτη επανάληψη (iteration 0) είναι η εκτίμηση της μέγιστης πιθανοφάνειας για το «μηδενικό» μοντέλο (“null” or “empty” model). Το «μηδενικό» μοντέλο είναι το εξεταζόμενο μοντέλο χωρίς τις προβλέπουσες μεταβλητές. Στην επόμενη επανάληψη, στο μοντέλο περιλαμβάνονται και οι προβλέπουσες μεταβλητές. Σε κάθε επανάληψη η μέγιστη πιθανοφάνεια αυξάνεται καθώς στόχος του αλγορίθμου είναι η μεγιστοποίηση του λογαρίθμου της πιθανοφάνειας. Όταν η διαφορά μεταξύ δύο συνεχόμενων επαναλήψεων είναι πολύ μικρή τότε θεωρείται ότι το μοντέλο έχει «συγκλίνει», η επαναληπτική διαδικασία σταματά, και προκύπτει το αποτέλεσμα.

b) Log Likelihood = - 99.608894

Το μέγεθος αυτό εκφράζει το μέγεθος της πιθανοφάνειας του προσαρμοσμένου μοντέλου (*fitted model*) που προέκυψε. Χρησιμοποιείται στη διενέργεια του τεστ Likelihood Ratio Chi – Square, στο οποίο διερευνάται αν οι συντελεστές των προβλεπουσών μεταβλητών που χρησιμοποιήθηκαν στην παλινδρόμηση είναι ταυτόχρονα όλοι ίσοι με το μηδέν. Χρησιμοποιείται επίσης και στα τεστ των εμφωλευμένων μοντέλων.

c) Number of obs = 171

Το μέγεθος αυτό αποτελεί το πλήθος των παρατηρήσεων που χρησιμοποιήθηκαν στο μοντέλο της λογιστικής παλινδρόμησης. Ο αριθμός αυτός μπορεί να είναι μικρότερος από το αρχικό πλήθος των εγγραφών στην περίπτωση που στο αρχικό σετ δεδομένων υπάρχουν κενές τιμές σε οποιαδήποτε μεταβλητή. Οι εγγραφές που περιέχουν κενές τιμές σε κάποια μεταβλητή διαγράφονται από το σετ δεδομένων της ανάλυσης. Στην περίπτωση των δεδομένων της

εργασίας αυτής, αυτό δεν συμβαίνει, διότι έχει προηγηθεί έλεγχος κενών τιμών και απαλοιφή των κενών δεδομένων.

d) LR ch2 (10) = 39.79

Το μέγεθος αυτό αποτελεί τη στατιστική του τεστ Likelihood Ratio (LR) Chi – Square στο οποίο γίνεται έλεγχος αν στο μοντέλο υπάρχει τουλάχιστον ένας συντελεστής κάποιας προβλέπουσας μεταβλητής που είναι διάφορος του μηδενός. Ο αριθμός εντός της παρένθεσης υποδεικνύει τον αριθμό των βαθμών ελευθερίας που χρησιμοποιήθηκαν από την κατανομή Chi – Square για να πραγματοποιηθεί ο έλεγχος της στατιστικής LR Chi – Square. Οι βαθμοί ελευθερίας υπολογίζονται από τον αριθμό των μοντέλων που υπολογίστηκαν, στην περίπτωση μας 1, επί τον αριθμό των προβλεπουσών μεταβλητών στο τελικό μοντέλο, στην περίπτωση μας 10 ($1 * 10 = 10$). Η στατιστική LR Chi – Square ή αλλιώς “*Deviance*” υπολογίζεται από τον ακόλουθο τύπο:

$$LR\ Chi - Square = -2 * (L(null\ model) - L(fitted\ model))$$

Όπου $L(null\ model)$ είναι ο λογάριθμος της πιθανοφάνειας του μοντέλου που περιέχει μόνο τη μεταβλητή απόκρισης (iteration 0), και $L(fitted\ model)$ είναι ο λογάριθμος της πιθανοφάνειας του μοντέλου της τελικής επανάληψης που περιέχει όλες τις μεταβλητές (για την περίπτωση που στο μοντέλο έχει επιτευχθεί η σύγκλιση). Στην περίπτωση μας η στατιστική ισούται με:

$$LR - Square = -2 * ((-118.50) - 99.60) = 37.79$$

e) Prob > ch2 = 0.000

Το μέγεθος αυτό εκτιμά την πιθανότητα να ληφθεί μια στατιστική LR Chi – Square τόσο ακραία ή περισσότερο ακραία από τη στατιστική που λάβαμε υπό τη μηδενική υπόθεση. Η μηδενική υπόθεση θεωρεί ότι όλοι οι συντελεστές της παλινδρόμησης στο μοντέλο είναι ταυτόχρονα όλοι ίσοι με το μηδέν. Με άλλα λόγια το παραπάνω μέγεθος εκτιμά την πιθανότητα να λάβουμε τη στατιστική LR Chi – Square που υπολογίστηκε (δηλαδή την 37.79) ή μια στατιστική ακόμα πιο ακραία, αν στην πραγματικότητα οι προβλέπουσες μεταβλητές δεν είχαν καμία επίδραση στη μεταβλητή απόκρισης. Αυτή η $p - value$ συγκρίνεται με την κριτική τιμή άλφα α , που υποδεικνύει την ανοχή αποδοχής ενός σφάλματος τύπου I, η οποία συνήθως λαμβάνεται ίση με 0.05 ή 0.01. Η μηδενική $p - value$ που προέκυψε από το τεστ LR σημαίνει

ότι το μοντέλο είναι στατιστικά σημαντικό και τουλάχιστον ένας συντελεστής μιας προβλέπουσας μεταβλητής της παλινδρόμησης είναι διάφορος του μηδενός. Η παράμετρος της κατανομής χ^2 που χρησιμοποιήθηκε για τον έλεγχο της μηδενικής υπόθεσης είναι ίση με τους βαθμούς ελευθερίας που ορίστηκαν στο προηγούμενο μέγεθος LR $\chi^2(10)$.

f) Pseudo R² = 0.1594

Το μέγεθος αυτό αποτελεί την McFadden ψευδό R^2 . Η λογιστική παλινδρόμηση δεν έχει ένα ισοδύναμο R^2 με αυτό της γραμμικής παλινδρόμησης. Υπάρχει μια ευρεία ποικιλία στατιστικών ψευδό R^2 . Ο τύπος υπολογισμού της στατιστικής McFadden's pseudo R^2 είναι ο ακόλουθος¹:

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})}$$

Όπου:

M_{Full} : Το μοντέλο με τις προβλέπουσες μεταβλητές

$M_{intercept}$: Το μοντέλο χωρίς τις προβλέπουσες μεταβλητές

\hat{L} : η εκτιμώμενη πιθανοφάνεια

Ο λόγος των πιθανοφανειών υποδεικνύει το επίπεδο βελτίωσης που επιτυγχάνεται στο μοντέλο χωρίς τις προβλέπουσες μεταβλητές μέσω του μοντέλου με τις προβλέπουσες μεταβλητές. Αφού η πιθανοφάνεια βρίσκεται μεταξύ του 0 και 1, τότε και ο λογάριθμος της πιθανοφάνειας θα είναι μικρότερος ή ίσος του μηδενός. Αν το μοντέλο έχει πολύ μικρή πιθανοφάνεια, τότε ο λογάριθμος της πιθανοφάνειας θα έχει μεγαλύτερο μέγεθος από ότι ο λογάριθμος ενός πιο πιθανού μοντέλου. Έτσι ένας μικρός λόγος των log των πιθανοφανειών υποδεικνύει ότι το μοντέλο με τις προβλέπουσες μεταβλητές έχει μακράν καλύτερη προσαρμογή από ότι το μοντέλο χωρίς αυτές.

Στην περίπτωση του μοντέλου της εργασίας, η στατιστική McFadden ισούται με 0.1594 που σημαίνει ότι ο λόγος των log πιθανοφανειών είναι 0.8406 (1 – 0.1594) που είναι αρκετά μεγάλος γεγονός που σημαίνει ότι το μοντέλο με όλες τις προβλέπουσες ίσως να μην έχει τόσο καλή προσαρμογή.

¹ Οι πληροφορίες σχετικά με τη στατιστική McFadden's pseudo R^2 λήφθηκαν από το link: http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm

g) Απόκριση

Αποτελεί τη μεταβλητή απόκρισης που υποδεικνύει αν η επιχείρηση είναι ενεργή ($y=0$) ή είναι μη ενεργή ($y=1$).

h) Συντελεστές των προβλεπουσών μεταβλητών

Οι συντελεστές αποτελούν τις τιμές που εκτιμήθηκαν από την λογιστική παλινδρόμηση για την πρόβλεψη της τιμής απόκρισης βάσει των προβλεπουσών μεταβλητών. Οι τιμές αυτές είναι σε μονάδες log – odds. Το μοντέλο της λογιστικής παλινδρόμησης που εφαρμόστηκε είναι το ακόλουθο:

Αν p η πιθανότητα επιτυχίας ($y = 1$), δηλαδή η πιθανότητα διακοπής της λειτουργίας της επιχείρησης, και $1 - p$ η πιθανότητα αποτυχίας ($y = 0$), δηλαδή η πιθανότητα μη διακοπής της και ομαλής συνέχισης της λειτουργίας της τότε:

$$\log\left(\frac{p}{1-p}\right) = c + \beta_1 * \text{Γεωγραφικός Τομέας}_2 + \beta_2 * \text{Γεωγραφικός Τομέας}_3 + \beta_3$$

* Μικτό Περιθώριο Κέρδους + β_4 * Απόσβεση Παγίων + β_5

* Ίδια προς Ενεργητικό + β_6 * Ίδια προς Ξένα + β_7 * Λοιπό Ενεργητικό προς Ξένα

+ β_8 * Βραχυπρόθεσμο Χρέος προς Ξένα + β_9 * Πωλήσεις προς Ενεργητικό + β_{10}

* Βραχυπρόθεσμο Χρέος προς Κεφάλαιο Κίνησης

Θεωρώντας ότι στατιστικά σημαντικές είναι οι μεταβλητές που η $p - value$ τους είναι μικρότερη από 0.12 (επίπεδο σημαντικότητας 88%) η εξίσωση της λογιστικής παλινδρόμησης σύμφωνα με τους εκτιμένους συντελεστές διαμορφώνεται ως ακολούθως:

$$\log\left(\frac{p}{1-p}\right) = 1.2859 - 0.6988 * \text{Γεωγραφικός Τομέας}_2 - 1.0921 * \text{Γεωγραφικός Τομέας}_3$$
$$+ 0.0251 * \text{Απόσβεση Παγίων} - 0.0412 * \text{Ίδια προς Ενεργητικό} + 0.0053$$
$$* \text{Ίδια προς Ξένα} - 0.0036 * \text{Λοιπό Ενεργητικό προς Ξένα}$$

Οι εκτιμήσεις των παραπάνω συντελεστών εκφράζουν τη σχέση των προβλεπουσών μεταβλητών με τη μεταβλητή απόκρισης, όπου η μεταβλητή απόκρισης είναι εκφρασμένη στην κλίμακα του logit. Οι εκτιμήσεις των συντελεστών εκφράζουν το ποσό που θα αυξηθεί ή θα μειωθεί, ανάλογα με το πρόσημο του συντελεστή, το προβλεπόμενο log odds για τιμή απόκρισης ίση με 1 (διακοπή επιχείρησης), όταν η τιμή μιας προβλέπουσας μεταβλητής αυξηθεί κατά 1 μονάδα, κρατώντας τις τιμές των υπόλοιπων προβλεπουσών μεταβλητών σταθερές.

Εξαιτίας του γεγονότος ότι οι τιμές των εκτιμημένων συντελεστών είναι σε μονάδες log – odds, δεν ερμηνεύεται κυρίως το μέγεθός τους αλλά το *πρόσημό* τους.

Στο παραπάνω μοντέλο η ερμηνεία των συντελεστών έχει ως ακολούθως:

Σταθερά: Ο συντελεστής της σταθεράς είναι ίσος με +1.2859. Η τιμή αυτή εκφράζει την αναμενόμενη τιμή του log – odds της απόκρισης όταν όλες οι προβλέπουσες μεταβλητές είναι ίσες με το μηδέν. Στην περίπτωση του μοντέλου αυτού δεν έχει ενδιαφέρον να ερμηνευθεί η σταθερά γιατί οι προβλέπουσες μεταβλητές αν λάβουν ταυτόχρονα την τιμή 0 θα σημαίνει ότι η επιχείρηση δεν βρίσκεται σε λειτουργία.

Γεωγραφικός Τομέας (επίπεδο 2): Ο συντελεστής της μεταβλητής αυτής είναι ίσος με – 0.6988. Η μεταβλητή αυτή είναι κατηγορική και το αποτέλεσμά της ερμηνεύεται ως προς το επίπεδο βάσης που έχει ορισθεί, δηλαδή το επίπεδο 1. Το επίπεδο 2 αποτελεί τον γεωγραφικό τομέα «Μακεδονία» ενώ το επίπεδο 1 τον γεωγραφικό τομέα «Αττική». Η ερμηνεία του συντελεστή έχει ως εξής: κρατώντας σταθερές όλες τις υπόλοιπες μεταβλητές, μια εταιρία που βρίσκεται στην περιοχή της Μακεδονίας έχει λιγότερες πιθανότητες να πτωχεύσει από μια εταιρία που βρίσκεται στην περιοχή της Αττικής.

Γεωγραφικός Τομέας (επίπεδο 3): Ο συντελεστής της μεταβλητής αυτής είναι ίσος με – 1.0921. Το επίπεδο 3 αποτελεί τον γεωγραφικό τομέα «Υπόλοιπη Ελλάδα». Η ερμηνεία του συντελεστή έχει ως εξής: κρατώντας σταθερές όλες τις υπόλοιπες μεταβλητές, μια εταιρία που βρίσκεται σε μια περιοχή εκτός της Αττικής και της Μακεδονίας έχει λιγότερες πιθανότητες να πτωχεύσει από μια εταιρία που βρίσκεται στην περιοχή της Αττικής.

Απόσβεση Παγίων: Ο συντελεστής της μεταβλητής αυτής είναι ίσος με +0.0251. Η μεταβλητή αυτή είναι ποσοτική και η ερμηνεία του συντελεστή της έχει ως εξής: για κάθε αύξηση μιας μονάδας του δείκτη απόσβεσης παγίων αναμένεται μια 0.0251 αύξηση του log – odds της απόκρισης, κρατώντας όλες τις υπόλοιπες μεταβλητές σταθερές. Γενικότερα μια αύξηση του δείκτη αυξάνει τις πιθανότητες διακοπής της επιχείρησης.

Ίδια Κεφάλαια προς Ενεργητικό: Ο συντελεστής της μεταβλητής αυτής είναι ίσος με – 0.0412. Η μεταβλητή αυτή είναι ποσοτική, είναι η περισσότερο στατιστικά σημαντική από όλες τις μεταβλητές που χρησιμοποιούνται στο μοντέλο, και η ερμηνεία του συντελεστή της έχει ως

εξής: για κάθε αύξηση μιας μονάδας του δείκτη Ίδια Κεφάλαια προς Ενεργητικό αναμένεται μια 0.0412 μείωση του log – odds της απόκρισης, κρατώντας όλες τις υπόλοιπες μεταβλητές σταθερές. Γενικότερα μια αύξηση του δείκτη μειώνει τις πιθανότητες διακοπής της επιχείρησης.

Ανάλογη ερμηνεία έχουν και οι μεταβλητές «Ίδια Κεφάλαια προς Ξένα Κεφάλαια» και «Λοιπό Ενεργητικό προς Ξένα Κεφάλαια».

i) Standard Errors

Τα μεγέθη αυτά αποτελούν τα τυπικά σφάλματα των εκτιμημένων συντελεστών των μεταβλητών. Το τυπικό σφάλμα χρησιμοποιείται στον έλεγχο στατιστικής σημαντικότητας των συντελεστών, αν δηλαδή οι συντελεστές είναι στατιστικά σημαντικά διάφοροι του μηδενός. Η διαίρεση των τιμών των συντελεστών με τα τυπικά τους σφάλματα δίνει την τιμή z (επόμενη στήλη). Τα τυπικά σφάλματα χρησιμοποιούνται επίσης στον υπολογισμό των διαστημάτων εμπιστοσύνης των συντελεστών (τελευταία στήλη)

j) z – value και p – value

Τα μεγέθη αυτά αποτελούν τη στατιστική τιμή z και την τιμή p – value του δίπλευρου ελέγχου που πραγματοποιήθηκε για την εξέταση της μηδενικής υπόθεσης ότι ο εξεταζόμενος συντελεστής της μεταβλητής απόκρισης ισούται με το μηδέν. Στο δίπλευρο έλεγχο η p – value συγκρίνεται με την τιμή $\alpha = 0.12$ που έχει επιλεγεί ως αποδεκτό όριο σφάλματος. Οι συντελεστές που έχουν p – value μικρότερη από α θεωρούνται στατιστικά σημαντικοί. Η τιμή z ακολουθεί την τυπική κανονική κατανομή. Η πιο στατιστικά σημαντική μεταβλητή, όπως αναφέρθηκε και προηγουμένως αλλά και στις αρχικές δοκιμές, είναι ο δείκτης «Ίδια Κεφάλαια προς Ενεργητικό».

k) [95% Confidence Interval]

Το μέγεθος αυτό αποτελεί το 95% διάστημα εμπιστοσύνης του συντελεστή της προβλέπουσας μεταβλητής. Πρόκειται για ιδιαίτερα αξιόλογο μέγεθος καθώς υποδεικνύει την κατώτατη και την ανώτατη τιμή που μπορεί να λάβει ο συντελεστής στον πραγματικό πληθυσμό με πιθανότητα 95%. Όταν το διάστημα εμπιστοσύνης δεν περιλαμβάνει το 0 σημαίνει πως η μεταβλητή είναι ιδιαίτερα στατιστικά σημαντική με τιμή p – value σχεδόν μηδενική.

B. Εκτιμήσεις οριακών αποτελεσμάτων

Καθώς τα αποτελέσματα του μοντέλου της λογιστικής παλινδρόμησης είναι εκφρασμένα σε μονάδες log – odds, για την φυσική ερμηνεία των συντελεστών είναι απαραίτητη η εκτίμηση του οριακού αποτελέσματος της κάθε μεταβλητής. Το οριακό αποτέλεσμα εκφράζει ουσιαστικά πόσο θα μεταβληθεί η πιθανότητα να λάβει η μεταβλητή απόκριση την τιμή $y=1$ όταν μεταβληθεί η τιμή μιας προβλέπουσας μεταβλητής κατά 1 μονάδα. Παρακάτω αναλύονται οι εκτιμήσεις των πιθανοτήτων στη μέση τιμή των μεταβλητών και οι εκτιμήσεις των μέσων οριακών αποτελεσμάτων.

Εκτίμηση πιθανοτήτων στη μέση τιμή των μεταβλητών (Marginal Effect at the Mean)

Η εκτίμηση αυτή αφορά τον υπολογισμό της πιθανότητας διακοπής μιας επιχείρησης για την οποία οι αριθμοδείκτες που εξετάζονται λαμβάνουν τη μέση τιμή τους στο δείγμα. Βασίζεται στον τύπο:

$$\partial p / \partial x_j = F'(\bar{x}'\beta)\beta_j = \Lambda(x'\beta)[1 - \Lambda(x'\beta)]\beta_j = \frac{e^{x'\beta}}{1 + e^{x'\beta}}\beta_j$$

Οι εκτιμήσεις των οριακών αποτελεσμάτων έχουν το ίδιο πρόσημο με τους συντελεστές καθώς $F'(\bar{x}'\beta) > 0$.

Τα περισσότερα άρθρα στη βιβλιογραφία δίνουν εκτιμήσεις για αυτή την πιθανότητα. Ένα πρόβλημα σχετικά με αυτές τις εκτιμήσεις είναι ότι πιθανόν να μην υπάρχει τέτοια εγγραφή στο δείγμα. (Katchova, 2013)

Καθώς στην ανάλυση περιέχεται και κατηγορική μεταβλητή θα υπολογιστεί για κάθε επίπεδο γεωγραφικού τομέα η τιμή της πιθανότητας η μεταβλητή απόκρισης να λάβει την τιμή 1 όταν οι τιμές των προβλεπουσών μεταβλητών λαμβάνουν τη μέση τιμή τους. Τα αποτελέσματα των υπολογισμών των πιθανοτήτων έχουν ως ακολούθως:

Adjusted predictions	Number of obs	=	171
Model VCE	: OIM		
Expression	: Pr(apo), predict()		
1._at	: Γεωγραφ. Τομέας	=	1
	Μεικτό Περιθώριο	=	22.92134 (mean)
	Απόσβεση Παγίων	=	10.58113 (mean)
	Ίδια/Ενεργητικό	=	40.00259 (mean)
	Ίδια/Ξένα	=	191.4451 (mean)
	Λοιπό Ενεργ/Ξένα	=	197.0819 (mean)
	Βραχ Χρέος/Ξένα	=	83.35787 (mean)

	Πωλήσεις/Ενεργητικό	=	76.11614	(mean)		
	Βραχ Χρέος/Κεφ.Κίνησης	=	102.8201	(mean)		
2._at	: Γεωγραφ. Τομέας	=	2			
	Μικτό Περιθώριο	=	22.92134	(mean)		
	Απόσβεση Παγίων	=	10.58113	(mean)		
	Ίδια/Ενεργητικό	=	40.00259	(mean)		
	Ίδια/Ξένα	=	191.4451	(mean)		
	Λοιπό Ενεργ/Ξένα	=	197.0819	(mean)		
	Βραχ Χρέος/Ξένα	=	83.35787	(mean)		
	Πωλήσεις/Ενεργητικό	=	76.11614	(mean)		
	Βραχ Χρέος/Κεφ.Κίνησης	=	102.8201	(mean)		
3._at	: Γεωγραφ. Τομέας	=	3			
	Μικτό Περιθώριο	=	22.92134	(mean)		
	Απόσβεση Παγίων	=	10.58113	(mean)		
	Ίδια/Ενεργητικό	=	40.00259	(mean)		
	Ίδια/Ξένα	=	191.4451	(mean)		
	Λοιπό Ενεργ/Ξένα	=	197.0819	(mean)		
	Βραχ Χρέος/Ξένα	=	83.35787	(mean)		
	Πωλήσεις/Ενεργητικό	=	76.11614	(mean)		
	Βραχ Χρέος/Κεφ.Κίνησης	=	102.8201	(mean)		

			Delta-method			
		Margin	Std. Err.	z	P> z	[95% Conf. Interval]

_at						
1		.6529714	.0672376	9.71	0.000	.5211882 .7847547
2		.483329	.0747904	6.46	0.000	.3367426 .6299154
3		.3869847	.0779587	4.96	0.000	.2341885 .539781

Το πάνω μέρος του πίνακα περιλαμβάνει τη μέση τιμή της κάθε ποσοτικής προβλέπουσας μεταβλητής ενώ οι εκτιμήσεις για κάθε επίπεδο γεωγραφικού τομέα βρίσκονται στο κάτω μέρος του πίνακα (Delta – Method). Τα αποτελέσματα έχουν ως εξής:

1. Η πιθανότητα μια επιχείρηση να διακόψει τη λειτουργία της ($y=1$), όταν όλες οι προβλέπουσες μεταβλητές λαμβάνουν τη μέση τιμή τους και η επιχείρηση βρίσκεται στην Αττική, είναι 65.29%.
2. Η πιθανότητα μια επιχείρηση να διακόψει τη λειτουργία της ($y=1$), όταν όλες οι προβλέπουσες μεταβλητές λαμβάνουν τη μέση τιμή τους και η επιχείρηση βρίσκεται στην περιοχή της Μακεδονίας, είναι 48.33%.
3. Η πιθανότητα μια επιχείρηση να διακόψει τη λειτουργία της ($y=1$), όταν όλες οι προβλέπουσες μεταβλητές λαμβάνουν τη μέση τιμή τους και η επιχείρηση βρίσκεται στην υπόλοιπη Ελλάδα πέραν της Μακεδονίας και της Αττικής, είναι 38.69%.

Τα αποτελέσματα είναι όλα στατιστικά σημαντικά, αν λάβουμε υπόψη την εκτιμηθείσα p – *value* σε κάθε περίπτωση, δείχνουν όμως επιπρόσθετα πως όταν οι δείκτες λαμβάνουν τη μέση τιμή τους τότε η πιθανότητα διακοπής είναι πάρα πολύ αυξημένη γεγονός που υποδεικνύει ότι

μια εύρωστη οικονομικά επιχείρηση θα πρέπει να λαμβάνει τιμές πολύ πιο πέρα (και από τις δύο πλευρές ανάλογα με τη φυσική σημασία του δείκτη) από το μέσο όρο.

Εκτίμηση μέσου οριακού αποτελέσματος (Average Marginal Effect)

Στην εκτίμηση αυτή τα οριακά αποτελέσματα υπολογίζονται ως ο μέσος όρος των μεμονωμένων οριακών αποτελεσμάτων. Το μέσο οριακό αποτέλεσμα βασίζεται στον τύπο:

$$\partial p / \partial x_j = \frac{\sum F'(x'\beta)}{n} \beta_j$$

Θα μπορούσε να θεωρηθεί καλή προσέγγιση εκτίμησης των πιθανοτήτων αν και στη βιβλιογραφία δεν χρησιμοποιείται ευρέως (Katchova, 2013). Στην πραγματικότητα και η εκτίμηση του “marginal effect at mean” και η εκτίμηση του “average marginal effect” παράγουν σχεδόν ίδια αποτελέσματα όπως φαίνεται και στον παρακάτω πίνακα με τα εκτιμημένα μέσα οριακά τα αποτελέσματα:

Predictive margins		Number of obs =		171		
Model VCE : OIM						
Expression : Pr(apo), predict()						
1._at	: Γεωγρ. Τομέας =			1		
2._at	: Γεωγρ. Τομέας =			2		
3._at	: Γεωγρ. Τομέας =			3		

		Delta-method				
		Margin	Std. Err.	z	P> z	[95% Conf. Interval]

_at						
1		.6053505	.0581742	10.41	0.000	.4913311 .7193699
2		.4612112	.0600059	7.69	0.000	.3436018 .5788206
3		.3837974	.0618488	6.21	0.000	.2625761 .5050188

C. Εκτιμήσεις odds ratio / σχετικού κινδύνου

Το odds ratio ή αλλιώς ο σχετικός κίνδυνος (*relative risk ratio*) αποτελεί το μέγεθος:

$$\text{πιθανότητα επιτυχίας} / \text{πιθανότητα απότυχίας}$$

Το μέγεθος αυτό εκτιμά την πιθανότητα πραγματοποίησης του ενδεχομένου επιτυχίας (y=1) σε σχέση με την πιθανότητα απότυχίας (y=0). Το odds ratio του αριστερού μέλους της εξίσωσης

αποτελεί την εκθετική τιμή του δεξιού μέλους της εξίσωσης της λογιστικής παλινδρόμησης και συγκεκριμένα:

$$\frac{p}{1-p} = \exp(x'\beta)$$

Το odds ratio ενός συντελεστή μιας προβλέπουσας μεταβλητής είναι η εκθετική τιμή του συντελεστή. Το odds ratio ή αλλιώς ο σχετικός κίνδυνος εκφράζει το πώς, το ενδεχόμενο που αφορά το εξεταζόμενο επίπεδο, μεταβάλλεται σε σχέση με το επίπεδο βάσης, σε κάθε αλλαγή της μεταβλητής του συντελεστή. Ένα odds ratio > 1 εκφράζει ότι η πιθανότητα του ενδεχομένου που αφορά το εξεταζόμενο γκρουπ σε σχέση με την πιθανότητα που αφορά το αναφερόμενο γκρουπ (επίπεδο βάσης) αυξάνεται όταν η τιμή της προβλέπουσας μεταβλητής αυξάνεται. Ένα odds ratio < 1 εκφράζει ότι η πιθανότητα του ενδεχομένου που αφορά το εξεταζόμενο γκρουπ σε σχέση με την πιθανότητα που αφορά το αναφερόμενο γκρουπ (επίπεδο βάσης) μειώνεται όταν η τιμή της προβλέπουσας μεταβλητής αυξάνεται. (UCLA, 2016)

Τα αποτελέσματα εκτίμησης των odds ratio στο μοντέλο λογιστικής παλινδρόμησης έχουν ως ακολούθως:

Logistic regression		Number of obs	=	171		
		LR chi2(10)	=	37.79		
		Prob > chi2	=	0.0000		
Log likelihood = -99.608894		Pseudo R2	=	0.1594		

	απόκριση	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]

	Γεωργ. Τομέας					
	2	.4971642	.2084105	-1.67	0.095	.2186137 1.130635
	3	.3355008	.1500907	-2.44	0.015	.1396038 .8062873

	Μεικτό Περιθώριο	.9999481	.0096862	-0.01	0.996	.9811426 1.019114
	Απόσβεση Παγίων	1.025426	.0167702	1.54	0.125	.9930779 1.058827
	Ίδια/Ενεργητικό	.9596131	.0090371	-4.38	0.000	.9420631 .9774899
	Ίδια/Ξένα	1.005316	.0019791	2.69	0.007	1.001445 1.009203
	Λοιπό Ενεργ/Ξένα	.9963207	.002139	-1.72	0.086	.9921372 1.000522
	Βραχ Χρέος/Ξένα	1.005284	.0084689	0.63	0.532	.9888213 1.02202
	Πωλήσεις/Ενεργητικό	1.000036	.0026813	0.01	0.989	.9947948 1.005305
	Βραχ Χρέος/Κεφ.Κίνησης	1.000002	.0001565	0.01	0.992	.9996948 1.000308
	Σταθερά	3.618062	2.95953	1.57	0.116	.7281212 17.97829

Παρακάτω θα επεξηγηθούν τα odds ratio των μεταβλητών «Γεωγραφικός Τομέας (επίπεδο 2)» και της πιο σημαντικής μεταβλητής «Ίδια Κεφάλαια / Ενεργητικό». Αντίστοιχη ερμηνεία με αυτές έχουν και οι υπόλοιπες μεταβλητές.

Γεωγραφικός Τομέας (επίπεδο 2): Το odds ratio της μεταβλητής αυτής ισούται με $0.497 < 1$. Αυτό σημαίνει ότι, για τις επιχειρήσεις που βρίσκονται στη Μακεδονία εν συγκρίσει με τις επιχειρήσεις που βρίσκονται στην Αττική, ο σχετικός κίνδυνος διακοπής σε σχέση με τη μη διακοπή αναμένεται να μειωθεί κατά 0.497 δεδομένου ότι όλες οι άλλες μεταβλητές παραμένουν σταθερές. Με άλλα λόγια το odds των εταιριών που βρίσκονται στη Μακεδονία είναι 64% ($\exp(0.479 - 1)$) χαμηλότερο από το odds των εταιριών που βρίσκονται στην Αττική, άρα οι εταιρίες που βρίσκονται στη Μακεδονία έχουν λιγότερες πιθανότητες να πτωχεύσουν από ότι οι εταιρίες που βρίσκονται στην Αττική.

Ίδια Κεφάλαια / Ενεργητικό: Το odds ratio της μεταβλητής αυτής ισούται με $0.959 < 1$. Αυτός είναι ο σχετικός κίνδυνος διακοπής της επιχείρησης για κάθε αύξηση κατά μια μονάδα της μεταβλητής «Ίδια Κεφάλαια / Ενεργητικό», δεδομένου ότι όλες οι υπόλοιπες μεταβλητές παραμένουν σταθερές. Αν μια επιχείρηση αύξανε το συγκεκριμένο δείκτη κατά μια μονάδα, τότε το odds της διακοπής σε σχέση με τη μη διακοπή θα μειωνόταν κατά 0.959 μονάδες ή αλλιώς θα παρατηρούταν μείωση του σχετικού κινδύνου διακοπής κατά 160% ($\exp(0.959 - 1)$), με σταθερούς τους υπόλοιπους δείκτες. Με άλλα λόγια, οι εταιρίες που αυξάνουν το δείκτη «Ίδια Κεφάλαια / Ενεργητικό» μειώνουν αρκετά τις πιθανότητες τους να πτωχεύσουν.

D. Πρόβλεψη πιθανοτήτων και μέτρα καλής προσαρμογής

Στην ενότητα αυτή θα υπολογισθεί για κάθε επιχείρηση στο δείγμα η πιθανότητα διακοπής της ($y=1$) (predicted probabilities) σύμφωνα με το μοντέλο που εκτιμήθηκε. Θα εφαρμοστεί ουσιαστικά ο τύπος:

$$\hat{p} = pr[y = 1|x] = F(x'\hat{\beta})$$

Η προβλεπόμενη πιθανότητα, που εκτιμάται από το στατιστικό πακέτο, υποδεικνύει την πιθανότητα πραγματοποίησης του ενδεχομένου $y=1$. Η by default “cutoff” πιθανότητα που χρησιμοποιεί το στατιστικό πακέτο είναι η 0.5 και αφήνεται ως έχει. Έτσι, αν η εκτιμώμενη πιθανότητα είναι πάνω από 0.5 τότε εκτιμάται ότι θα πραγματοποιηθεί το ενδεχόμενο διακοπής

της λειτουργίας της επιχείρησης ($\hat{y} = 1$) και η εκτίμηση αυτή καλείται θετική – positive (+), διαφορετικά, αν η εκτιμώμενη πιθανότητα είναι κάτω από 0.5, η επιχείρηση θεωρείται ότι δεν θα διακόψει τη λειτουργία της ($\hat{y} = 0$) και η εκτίμηση αυτή καλείται αρνητική – negative (-).

Ο πίνακας (confusion matrix) που θα προκύψει από την ακόλουθη ανάλυση θα έχει το ακόλουθο εννοιολογικό περιεχόμενο:

	Actual $y = 1$	Actual $y = 0$
(+) Predicted $\hat{y} = 1$	True Positive	False Positive
(-) Predicted $\hat{y} = 0$	False Negative	True Negative

Το ποσοστό των σωστών προβλεφθέντων τιμών είναι το άθροισμα των συχνοτήτων της κύριας διαγωνίου, δηλαδή όλες οι αληθείς τιμές, προς το σύνολο των παρατηρήσεων του δείγματος. Τα αποτελέσματα κατηγοριοποίησης των προβλέψεων έχουν ως ακολούθως:

Logistic model for απόκριση			
Classified	----- True -----		Total
	D	~D	
+	50	20	70
-	34	67	101
Total	84	87	171
Classified + if predicted $\Pr(D) \geq .5$ True D defined as απόκριση != 0			
Sensitivity	$\Pr(+ D)$		59.52%
Specificity	$\Pr(- \sim D)$		77.01%
Positive predictive value	$\Pr(D +)$		71.43%
Negative predictive value	$\Pr(\sim D -)$		66.34%
False + rate for true ~D	$\Pr(+ \sim D)$		22.99%
False - rate for true D	$\Pr(- D)$		40.48%
False + rate for classified +	$\Pr(\sim D +)$		28.57%
False - rate for classified -	$\Pr(D -)$		33.66%
Correctly classified			68.42%

Τα δεδομένα του αποτελέσματος, από τα οποία θα προκύψουν τα μέτρα καλής προσαρμογής (*goodness of fit measures*) συνοψίζονται στον παρακάτω πίνακα:

Πίνακας 5.2-2 Αποτελέσματα Confusion Matrix της Λογιστικής Παλινδρόμησης

	Actual $y = 1$	Actual $y = 0$	Total
(+) Predicted $\hat{y} = 1$	50 (True Positive)	20 (False Positive)	70
(-) Predicted $\hat{y} = 0$	34 (False Negative)	67 (True Negative)	101
Total	84	87	171

ενώ η επεξήγηση του αποτελέσματος της κατηγοριοποίησης είναι η ακόλουθη:

a. Correctly Classified: 68.42%

Παρατηρούμε ότι το ποσοστό των συνολικών σωστών κατηγοριοποιήσεων είναι 68.42%. Το ποσοστό αυτό προκύπτει από το άθροισμα των παρατηρήσεων που το αποτέλεσμα της ταξινόμησής τους ($\hat{y} = 1$ ή $\hat{y} = 0$) ήταν αντίστοιχα ίδιο με το την κατάσταση τους στην πραγματικότητα ($y = 1$ ή $y = 0$). Ουσιαστικά προκύπτει από το άθροισμα των στοιχείων της κύριας διαγωνίου του πίνακα “Confusion Matrix”, δηλαδή:

$$correctly\ classified = 68.42\% = \frac{50 + 67}{171} = \frac{117}{171}$$

b. Sensitivity Pr (+|D): 59.52%

Το ποσοστό “sensitivity” αποτελεί το ποσοστό των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει ότι $y = 1$ και ταξινομήθηκαν σωστά από την εκτίμηση της πρόβλεψης ως $\hat{y} = 1$ (true positive) ως προς το σύνολο των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 1$. Το ποσοστό αυτό λέγεται αλλιώς “**True Positive Rate**” (TPR). Από τα δεδομένα το ποσοστό sensitivity προκύπτει ως εξής:

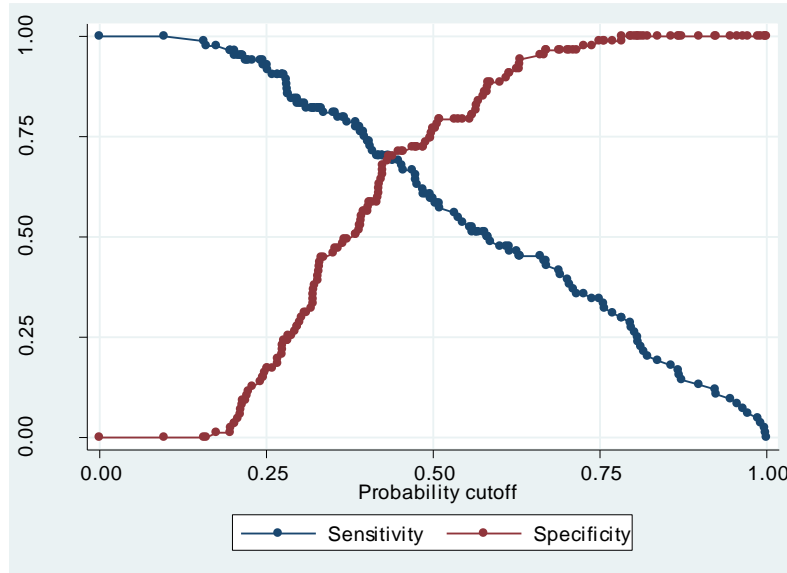
$$sensitivity\ (True\ Positive\ Rate - TPR) = 59.52\% = \frac{50}{84}$$

c. Specificity Pr(-|~D): 77.01%

Το ποσοστό “specificity” αποτελεί το ποσοστό των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει ότι $y = 0$ και ταξινομήθηκαν σωστά από την εκτίμηση της πρόβλεψης ως $\hat{y} = 0$ (true negative) ως προς το σύνολο των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 0$. Το ποσοστό αυτό λέγεται αλλιώς “**True Negative Rate**” (TNR). Από τα δεδομένα το specificity προκύπτει ως εξής:

$$\text{specificity (True Negative Rate - TNR)} = 77.01\% = \frac{67}{87}$$

Ακολούθως απεικονίζεται το διάγραμμα “sensitivity and specificity vs. Probability cutoff”.



Διάγραμμα 5.2-1 Sensitivity / Specificity vs Probability Cutoff

d. Positive predictive value $\Pr(D|+)$: 71.43%

Το ποσοστό αυτό ονομάζεται αλλιώς “precision” και εκφράζει το ποσοστό των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 1$ και ταξινομήθηκαν σωστά από την εκτίμηση της πρόβλεψης ως $\hat{y} = 1$ (true positive) ως προς το σύνολο των παρατηρήσεων που ταξινομήθηκαν από την εκτίμηση της πρόβλεψης ως $\hat{y} = 1$. Από τα δεδομένα αυτό προκύπτει ως εξής:

$$\text{Positive Predictive Value} = 71.43\% = \frac{50}{70}$$

e. Negative predictive value $\Pr(\sim D|-)$: 66.34%

Το ποσοστό αυτό εκφράζει το ποσοστό των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 0$ και ταξινομήθηκαν σωστά από την εκτίμηση της πρόβλεψης ως $\hat{y} = 0$ (true negative) ως προς το σύνολο των παρατηρήσεων που ταξινομήθηκαν από την εκτίμηση της πρόβλεψης ως $\hat{y} = 0$. Από τα δεδομένα αυτό προκύπτει ως εξής:

$$\text{Negative Predictive Value} = 66.34\% = \frac{67}{101}$$

f. False + rate for true ~D Pr(+|~D): 22.99%

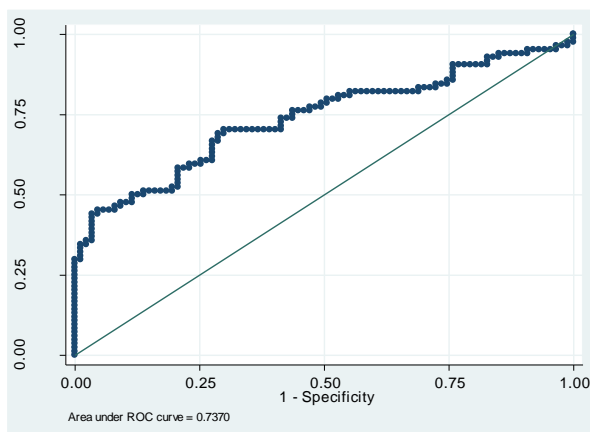
Το ποσοστό αυτό εκφράζει το ποσοστό των παρατηρήσεων που ταξινομήθηκαν από την εκτίμηση της πρόβλεψης ως $\hat{y} = 1$ ενώ στην πραγματικότητα για αυτές ισχύει ότι $y = 0$ (false positive) ως προς το σύνολο των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 0$. Το ποσοστό αυτό λέγεται αλλιώς “**Fall out or False Positive Rate**” (FTR). Για το ποσοστό αυτό ισχύει ότι ισούται με $1 - \text{Specificity}$. Από τα δεδομένα αυτό προκύπτει ως εξής:

$$\text{Fall out or False Positive Rate} = 22.99\% = \frac{20}{87} = 1 - \text{specificity} = 1 - 77.01\%$$

Παρακάτω ακολουθεί το διάγραμμα της καμπύλης ROC (Receiver Operating Characteristic) που αποτελεί μέτρο μέτρησης καλής προσαρμογής του μοντέλου.

Η καμπύλη απεικονίζει τη συνάρτηση του True Positive Rate (Sensitivity) ως προς το “Fall out or False Positive Rate ($1 - \text{Specificity}$)” σε κάθε αλλαγή του σημείου cutoff, ενώ παράλληλα υπολογίζεται για το μοντέλο που εκτιμήθηκε το εμβαδόν κάτω από την καμπύλη που αντιστοιχεί στο cutoff που χρησιμοποιήθηκε. Η καμπύλη ξεκινάει από το σημείο (0,0) που αντιστοιχεί σε cutoff = 1 και συνεχίζει ως το σημείο (1,1) που αντιστοιχεί σε cutoff = 0. Η καμπύλη ROC ενός μοντέλου με καμία προβλεπτική ικανότητα θα συμπέσει με τη γραμμή των 45 μοιρών. Όσο πιο μεγάλη είναι η προβλεπτική ικανότητα του μοντέλου τόσο πιο κυρτή θα είναι η καμπύλη ROC γι αυτό και το εμβαδόν κάτω από αυτή χρησιμοποιείται ως μονάδα μέτρησης της ικανότητας αυτής. Ένα μοντέλο με καμία προβλεπτική ικανότητα έχει εμβαδό 0.5, ενώ ένα μοντέλο με τέλεια προβλεπτική ικανότητα έχει εμβαδό 1 (Green & Swets, 1966).

Στο διάγραμμα που ακολουθεί το εμβαδόν κάτω από την καμπύλη ισούται με 0.7370 σύμφωνα με το οποίο συμπεραίνεται πως το μοντέλο έχει κάποια προβλεπτική ικανότητα.



Διάγραμμα 5.2-2 Καμπύλη ROC του μοντέλου της Λογιστικής Παλινδρόμησης

g. False - rate for true D Pr(-| D): 40.48%

Το ποσοστό αυτό εκφράζει το ποσοστό των παρατηρήσεων που ταξινομήθηκαν από την εκτίμηση της πρόβλεψης ως $\hat{y} = 0$ ενώ στην πραγματικότητα για αυτές ισχύει ότι $y = 1$ (false negative) ως προς το σύνολο των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 1$. Το ποσοστό αυτό λέγεται αλλιώς “**False Negative Rate**” (FTR). Από τα δεδομένα αυτό προκύπτει ως εξής:

$$\text{False Negative Rate} = 40.48\% = \frac{34}{84}$$

h. False + rate for classified + Pr(~D| +): 28.57%

Το ποσοστό αυτό εκφράζει το ποσοστό των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 0$ ενώ ταξινομήθηκαν εσφαλμένα ως $\hat{y} = 1$ (false positive) ως προς το σύνολο των παρατηρήσεων που ταξινομήθηκαν από την εκτίμηση της πρόβλεψης ως $\hat{y} = 1$. Το ποσοστό αυτό λέγεται αλλιώς “**False Discovery Rate**” (FDR). Για το ποσοστό αυτό ισχύει ότι ισούται με $1 - \text{Positive Predictive Value}$. Από τα δεδομένα αυτό προκύπτει ως εξής:

$$\text{False Discovery Rate} = 28.57\% = \frac{20}{70} = 1 - \text{Positive predictive Value} = 1 - 71.43\%$$

i. False - rate for classified - Pr(D| -): 33.66%

Το ποσοστό αυτό εκφράζει το ποσοστό των παρατηρήσεων για τις οποίες στην πραγματικότητα ισχύει $y = 1$ ενώ ταξινομήθηκαν εσφαλμένα ως $\hat{y} = 0$ (false negative) ως προς το σύνολο των παρατηρήσεων που ταξινομήθηκαν από την εκτίμηση της πρόβλεψης ως $\hat{y} = 0$. Για το ποσοστό αυτό ισχύει ότι ισούται με $1 - \text{Negative Predictive Value}$. Από τα δεδομένα αυτό προκύπτει ως εξής:

$$33.66\% = \frac{34}{101} = 1 - \text{Negative predictive Value} = 1 - 66.34\%$$

Hosmer–Lemeshow goodness-of-fit test

Το τελευταίο μέτρο καλής προσαρμογής που θα εξετασθεί για το μοντέλο είναι το “**Hosmer–Lemeshow goodness-of-fit test**” που ουσιαστικά διενεργεί το “Pearson χ^2 goodness-of-fit test”. Το τεστ αυτό αφορά τη σχέση μεταξύ των παρατηρούμενων τιμών εν συγκρίσει με τον

αναμενόμενο αριθμό αποκρίσεων που προκύπτουν από τις παρατηρήσεις που ορίζουν τα διαφορετικά “covariate patterns”.² Το αποτέλεσμα της εκτίμησης του τεστ έχει ως ακολούθως:

Logistic model for απόκριση, goodness-of-fit test

number of observations =	171
number of covariate patterns =	171
Pearson chi2(160) =	166.36
Prob > chi2 =	0.3490

Από το τεστ συμπεραίνεται πως το μοντέλο έχει κάποια προβλεπτική ικανότητα καθώς η p -value = 0.34 δεν απορρίπτει τη μηδενική υπόθεση που θεωρεί ότι το μοντέλο έχει κάποια καλή προσαρμογή. Παρόλα αυτά παρατηρείται ότι το πλήθος των “covariate patterns” συμπίπτει με το πλήθος των παρατηρήσεων κάνοντας την εφαρμογή του τεστ αμφισβητήσιμη αλλά όχι απαραίτητα ακατάλληλη.

Οι Hosmer, Lemeshow, and Sturdivant (2013), για την αντιμετώπιση αυτού του θέματος, προτείνουν, την επαναομαδοποίηση των παρατηρήσεων κατά σειρά κατάταξης των προβλεφθέντων πιθανοτήτων, κατόπιν το σχηματισμό για παράδειγμα 10 ισοπληθών ομάδων, και με τη ολοκλήρωση αυτού τη διενέργεια του τεστ. (Hosmer, et al., 2013)

Τα αποτελέσματα της διενέργειας του τεστ μετά την επαναομαδοποίηση έχουν ως εξής:

Logistic model for apo, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

number of observations =	171
number of groups =	10
Hosmer-Lemeshow chi2(8) =	11.94
Prob > chi2 =	0.1538

Από το τεστ συμπεραίνεται πως το μοντέλο και μετά την προσαρμογή έχει κάποια προβλεπτική ικανότητα καθώς η p -value = 0.1538 δεν απορρίπτει τη μηδενική υπόθεση που θεωρεί ότι το μοντέλο έχει κάποια προσαρμογή.

² <http://www.stata.com/manuals13/restatgof.pdf#restatgof>

Συμπεράσματα από την εφαρμογή της Λογιστικής Παλινδρόμησης

Από τη Λογιστική Παλινδρόμηση προέκυψε ότι η σημαντικότερη μεταβλητή που μπορεί να δώσει προειδοποιητικό σήμα διακοπής της επιχείρησης μια χρονική περίοδο πριν την πραγματοποίηση της διακοπής, είναι τα «Ίδια Κεφάλαια προς Ενεργητικό». Η μεταβλητή αυτή είναι σημαντική σε επίπεδο εμπιστοσύνης 97%. Αν μια επιχείρηση αύξανε το συγκεκριμένο δείκτη κατά μια μονάδα, τότε το odds της διακοπής σε σχέση με τη μη διακοπή θα μειωνόταν κατά 0.959 μονάδες ή αλλιώς θα παρατηρούταν μείωση του σχετικού κινδύνου διακοπής κατά 160% ($\exp(0.959 - 1)$), με σταθερούς τους υπόλοιπους δείκτες. Με άλλα λόγια, οι εταιρίες που αυξάνουν το δείκτη «Ίδια Κεφάλαια / Ενεργητικό» μειώνουν αρκετά τις πιθανότητες τους να πτωχεύσουν. Άλλες μεταβλητές που προέκυψαν σημαντικές σε επίπεδο σημαντικότητας 12% είναι οι ακόλουθες: Βραχυπρόθεσμο Χρέος προς Ξένα Κεφάλαια, Απόσβεση Παγίων, Πωλήσεις προς Ενεργητικό, Μεικτό Περιθώριο Κέρδους, Βραχυπρόθεσμο Χρέος προς Κεφάλαιο Κίνησης, Ίδια Κεφάλαια προς Ξένα Κεφάλαια, Λοιπό Ενεργητικό προς Ξένα Κεφάλαια ενώ η τελευταία αφορά την ποιοτική μεταβλητή του Γεωγραφικού Τομέα που ανήκει η επιχείρηση. Από την εφαρμογή του μοντέλου, το ποσοστό των συνολικών σωστών κατηγοριοποιήσεων είναι 68.42% ενώ από τον έλεγχο καλής προσαρμογής των “Hosmer–Lemeshow” συμπεραίνεται πως το μοντέλο έχει κάποια προβλεπτική ικανότητα.

5.2.2 Εφαρμογή μοντέλου Probit

Παρακάτω θα παρουσιαστούν τα αποτελέσματα της ανάλυσης που αφορούν το μοντέλο probit. Οι ερμηνείες των συντελεστών είναι ίδιες με αυτές του μοντέλου της λογιστικής παλινδρόμησης γι' αυτό και θα παραληφθούν.

Στο τέλος της παρουσίασης των αποτελεσμάτων probit παρατίθενται δύο συγκεντρωτικοί πίνακες με τα σημαντικότερα αποτελέσματα που προέκυψαν από τις δύο αναλύσεις logit και probit. Στον πρώτο πίνακα απεικονίζονται οι εκτιμήσεις των συντελεστών που προέκυψαν από τα δύο μοντέλα. Στον δεύτερο πίνακα απεικονίζονται οι εκτιμήσεις των οριακών αποτελεσμάτων στον μέσο και των μέσων οριακών αποτελεσμάτων που προέκυψαν από τα δύο μοντέλα.

A. Εκτίμηση του μοντέλου probit

Τα αποτελέσματα εκτίμησης του μοντέλου probit είναι τα ακόλουθα:

Iteration 0:	log likelihood = -118.50185					
Iteration 1:	log likelihood = -100.24604					
Iteration 2:	log likelihood = -99.589577					
Iteration 3:	log likelihood = -99.586005					
Iteration 4:	log likelihood = -99.586005					
Probit regression						
Number of obs	= 171					
LR chi2(10)	= 37.83					
Prob > chi2	= 0.0000					
Pseudo R2	= 0.1596					
Log likelihood = -99.586005						

	Απόκριση	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

Γεωγραφ. Τομέας						
2		-.4205332	.2527183	-1.66	0.096	-.9158519 .0747855
3		-.6522601	.2665104	-2.45	0.014	-1.174611 -.1299094
Μεικτό Περιθώριο		.0002111	.0057092	0.04	0.971	-.0109786 .0114009
Απόσβεση Παγίων		.015143	.0096771	1.56	0.118	-.0038237 .0341097
Τόξα/Ενεργητικό		-.0249887	.00553	-4.52	0.000	-.0358273 -.0141502
Τόξα/Ξένα		.0031714	.001152	2.75	0.006	.0009136 .0054293
Λοιπό Ενεργ/Ξένα		-.0021736	.0012466	-1.74	0.081	-.0046168 .0002696
Βραχ Χρέος/Ξένα		.003439	.0050647	0.68	0.497	-.0064877 .0133657
Πωλήσεις/Ενεργητικό		-.0000335	.0016288	-0.02	0.984	-.003226 .0031589
Βραχ Χρέος/Κεφ.Κίνησης		2.86e-06	.0000969	0.03	0.976	-.0001872 .0001929
Σταθερά		.7592496	.4833521	1.57	0.116	-.188103 1.706602

B. Εκτιμήσεις οριακών αποτελεσμάτων

Οι εκτιμήσεις των οριακών αποτελεσμάτων στο μέσο έχουν ως ακολούθως:

Adjusted predictions		Number of obs = 171				
Model VCE : OIM						
		Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]	
_at						
1	.6509515	.0660706	9.85	0.000	.5214555	.7804474
2	.4869798	.0728486	6.68	0.000	.3441991	.6297605
3	.3957476	.0757219	5.23	0.000	.2473354	.5441598

Οι εκτιμήσεις των μέσων οριακών αποτελεσμάτων έχουν ως ακολούθως:

Predictive margins		Number of obs = 171				
Model VCE : OIM						
		Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]	
_at						
1	.607763	.0576692	10.54	0.000	.4947335	.7207925
2	.4653831	.0603996	7.71	0.000	.3470022	.5837641
3	.389401	.0621278	6.27	0.000	.2676327	.5111693

C. Πίνακες συγκρίσεων αποτελεσμάτων μεταξύ των μοντέλων Logit και Probit

Πίνακας 5.2-3 Εκτιμήσεις συντελεστών που προέκυψαν από τα μοντέλα Logit και Probit

Προβλέπουσες Μεταβλητές	Συντελεστές Logit	P - Value Συντέλ. Logit	Συντελεστές Probit	P - Value Συντέλ. Probit
Pseudo - R2	0.15940		0.15960	
Γεωγραφ. Τομέας_ Επίπεδο 2	-0.69884	0.095	-0.42053	0.096
Γεωγραφ. Τομέας_ Επίπεδο 3	-1.09213	0.015	-0.65226	0.014
Μεικτό Περιθώριο	-0.00005	0.996	0.00021	0.971
Απόσβεση Παγίων	0.02511	0.125	0.01514	0.118
Ίδια / Ενεργητικό	-0.04123	0.000	-0.02499	0.000
Ίδια / Ξένα	0.00530	0.007	0.00317	0.006
Λοιπό Ενεργητικό / Ξένα	-0.00369	0.086	-0.00217	0.081
Βραχ. Χρέος / Ξένα	0.00527	0.532	0.00344	0.497
Πωλήσεις / Ενεργητικό	0.00004	0.989	-0.00003	0.984
Βραχ. Χρέος/ Κεφ. Κίνησης	0.00000	0.992	0.00000	0.976
Σταθερά	1.28594	0.116	0.75925	0.116

Από τον πίνακα παρατηρούνται κάποιες μικρές διαφορές στις εκτιμήσεις των συντελεστών. Αυτό συμβαίνει διότι το δείγμα παρατηρήσεων δεν είναι ιδιαίτερα μεγάλο. Σε περίπτωση που το δείγμα ήταν μεγαλύτερο οι τιμές των συντελεστών θα έτειναν να συμπέσουν στα δύο μοντέλα.

Πίνακας 5.2-4 Εκτιμήσεις οριακών αποτελεσμάτων στον μέσο και μέσω οριακών αποτελεσμάτων που προέκυψαν από τα μοντέλα Logit και Probit

Επίπεδα Γεωγραφικού Τομέα	Logit marginal effect at mean	Logit average marginal effect	Probit marginal effect at mean	Probit average marginal effect
1	0.6530	0.6054	0.6510	0.6078
2	0.4833	0.4612	0.4870	0.4654
3	0.3870	0.3838	0.3957	0.3894

Στην περίπτωση των οριακών αποτελεσμάτων που εκτιμήθηκαν από τα δύο μοντέλα οι διαφορές που παρατηρούνται είναι απειροελάχιστες.

5.2.3 Εφαρμογή Ανάλυσης Κυρίων Συνιστωσών

Όπως έχει αναφερθεί, σκοπός της Ανάλυσης Κυρίων Συνιστωσών (*Principal Component Analysis – PCA*) είναι η μείωση της διάστασης πολυμεταβλητών δεδομένων. Στην ανάλυση δεν θα χρησιμοποιηθεί η μεταβλητή «Γεωγραφικός Τομέας» καθώς αποτελεί κατηγορική μεταβλητή και η Ανάλυση Κυρίων Συνιστωσών δεν εφαρμόζεται σε σύνολο δεδομένων που περιέχει κατηγορικές μεταβλητές.

Η ανάλυση που θα πραγματοποιηθεί βασίζεται στο μήτρα συσχετίσεων των προβλεπουσών μεταβλητών. Αυτό γιατί οι τιμές τους είναι αρκετά μεγάλες και σε μια μήτρα συνδιακυμάνσεων θα επικρατούσαν τα πολύ μεγάλα μεγέθη έναντι των μικρότερων. Το αποτελέσματα της ανάλυσης για το σύνολο των στατιστικά σημαντικών ποσοτικών προβλεπουσών μεταβλητών έχει ως ακολούθως:

Principal components/correlation				
	Number of obs	=	171	
	Number of comp.	=	8	
	Trace	=	8	
Rotation: (unrotated = principal)	Rho	=	1.0000	

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	2.45883	.925285	0.3074	0.3074
Comp2	1.53355	.423248	0.1917	0.4990
Comp3	1.1103	.160081	0.1388	0.6378
Comp4	.950218	.215189	0.1188	0.7566
Comp5	.735029	.0796588	0.0919	0.8485
Comp6	.65537	.147454	0.0819	0.9304
Comp7	.507916	.459128	0.0635	0.9939
Comp8	.0487879	.	0.0061	1.0000

Η επεξήγηση του αποτελέσματος της εκτίμησης του μοντέλου είναι η ακόλουθη:

a. Eigenvalue

Η στήλη αυτή αποτελεί την εκτίμηση της ιδιοτιμής λ της κάθε συνιστώσας. Η τιμή λ είναι η διακύμανση της συνιστώσας. Έτσι, η διακύμανση της 1^{ης} συνιστώσας είναι 2.45883, η διακύμανση της 2^{ης} είναι 1.53355 κ.ο.κ.

b. Difference

Η στήλη αυτή αποτελεί τη διαφορά της ιδιοτιμής λ της κάθε συνιστώσας με την ιδιοτιμή της επόμενης της. Έτσι, για παράδειγμα, η διαφορά της ιδιοτιμής της 1^{ης} συνιστώσας από τη 2^η συνιστώσα είναι $2.45883 - 1.53355 = 0.925285$. κ.ο.κ

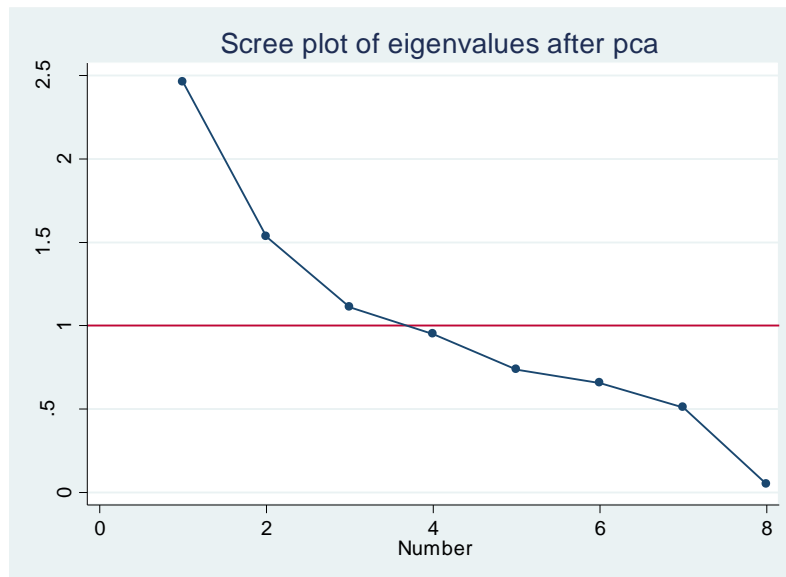
c. Proportion

Η στήλη αυτή αποτελεί το ποσό της συνολικής μεταβλητότητας που εξηγεί η εκάστοτε συνιστώσα. Έτσι, για παράδειγμα, η 1^η συνιστώσα εξηγεί το 30.74% της συνολικής μεταβλητότητας των δεδομένων, η 2^η συνιστώσα εξηγεί επιπλέον 19.17% της συνολικής μεταβλητότητας κ.ο.κ

d. Cumulative

Η στήλη αυτή αποτελεί το ποσό της συνολικής μεταβλητότητας του μοντέλου που ερμηνεύεται με την προσθήκη κάθε επιπλέον συνιστώσας. Όταν χρησιμοποιηθούν όλες οι συνιστώσες το ποσοστό που ερμηνεύεται είναι προφανώς 100% για αυτό και στο επίπεδο της τελευταίας συνιστώσας η τιμή της στήλης αυτής είναι 1. Έτσι, αν χρησιμοποιηθεί στην ανάλυση μόνο η 1^η συνιστώσα θα έχει επιτευχθεί η εξήγηση του 30.74% της συνολικής μεταβλητότητας, αν προστεθεί και η 2^η συνιστώσα στην ανάλυση θα έχει επιτευχθεί η εξήγηση του 49.90% (30.74% + 19.17%) της συνολικής μεταβλητότητας κ.ο.κ.

Παρατηρούμε ότι οι 3 πρώτες συνιστώσες έχουν ιδιοτιμή μεγαλύτερη του 1. Αυτό σημαίνει ότι η κάθε μία από αυτές τις συνιστώσες αυτές ερμηνεύει τουλάχιστον τόση μεταβλητότητα όση και οι αρχικές μεταβλητές. Η μεταβλητότητα που ερμηνεύουν αθροιστικά οι 3 αυτές μεταβλητές ισούται με το 63.78% της συνολικής μεταβλητότητας. Αν προστεθεί και η 4^η συνιστώσα στην ανάλυση, που έχει ιδιοτιμή πολύ κοντά στο 1 ($\lambda_4=0.95$), τότε επιτυγχάνεται ερμηνεία του 75.66% της συνολικής μεταβλητότητας. Το ποσοστό αυτό είναι αρκετά καλό για αυτό και στην ανάλυση θα χρησιμοποιηθούν οι πρώτες 4 συνιστώσες. Το διάγραμμα screeplot και για τις 8 συνιστώσες είναι το ακόλουθο:



Διάγραμμα 5.2-3 Screeplot της Ανάλυσης Κυρίων Συνιστωσών

Από το διάγραμμα παρατηρείται «σπάσιμο» στην καμπύλη ή αλλιώς σχηματισμός «αγκώνα» στην 3^η συνιστώσα. Αυτό σημαίνει ότι θα μπορούσαν να χρησιμοποιηθούν και 3 συνιστώσες αντί για 4.

Συνεπώς με την Ανάλυση Κύριων Συνιστωσών που εφαρμόστηκε πραγματοποιήθηκε μείωση της διάστασης των δεδομένων, αφού αντί για 8 προβλέπουσες μεταβλητές μπορούν να χρησιμοποιηθούν 3 συνιστώσες ή και 4 συνιστώσες.

Παρακάτω παρουσιάζονται τα αποτελέσματα των μοναδιαίων ιδιοδιανυσμάτων που αντιστοιχούν στην ιδιοτιμή της κάθε συνιστώσας. Το άθροισμα των τετραγώνων των φορτίων της κάθε συνιστώσας ισούται με 1.

Principal components (eigenvectors)									
Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Unexplained
Μεικτό Περιθώριο	0.0766	0.1826	0.7764	0.2977	0.4347	0.1028	-0.2630	0.0265	0
Απόσβεση Παγίων	0.0002	0.5744	0.2538	-0.0843	-0.5589	-0.5289	-0.0747	0.0301	0
Ίδια/Ενεργητικό	0.4771	-0.0701	0.2302	0.1734	-0.0760	-0.0334	0.8182	-0.0899	0
Ίδια/Ξένα	0.6058	-0.0929	-0.0716	-0.0495	-0.1270	0.0850	-0.2592	0.7255	0
Λοιπό Ενεργ/Ξένα	0.5964	-0.0189	-0.0890	-0.0700	-0.1025	0.0729	-0.3908	-0.6801	0
Βραχ Χρέος/Ξένα	0.2091	0.4372	-0.3447	-0.1851	0.6727	-0.3846	0.1023	0.0340	0
Πωλήσεις/Ενεργητικ	-0.0087	0.6160	-0.0690	-0.1988	-0.0871	0.7402	0.1432	0.0149	0
Βραχ Χρέος/Κεφ.Κίν	0.0035	0.2287	-0.3783	0.8906	-0.0781	0.0179	-0.0703	0.0088	0

Τα φορτία που παρουσιάζονται στον πίνακα για την κάθε συνιστώσα αποτελούν τη συσχέτιση μεταξύ της συνιστώσας και της αρχικής μεταβλητής. Έτσι για παράδειγμα, αντί να

χρησιμοποιηθεί το μοναδιαίο διάνυσμα, θεωρείται το ιδιοδιάνυσμα που αντιστοιχεί στο λ_1 και έχει μέτρο $\sqrt{\lambda_1}$ δηλαδή το:

$$\begin{aligned}\sqrt{\lambda_1}u_1 &= \sqrt{2.45883} (0.04885, \\ &0.00013, 0.30426, 0.38634, 0.38034, 0.13335, -0.00555, 0.00223)' \\ &= (0.0766, 0.0002, 0.4771, 0.6058, 0.5964, 0.2091, -0.0087, 0.0035)'\end{aligned}$$

Έτσι οι συντεταγμένες του πρώτου ιδιοδιανύσματος είναι ίσες με $\sqrt{\lambda_1}u_1$ και δίνουν απευθείας τους συντελεστές συσχέτισης της πρώτης κύριας συνιστώσας με τις μεταβλητές X_1, X_2, \dots, X_8 .

Επίσης για την 1^η συνιστώσα, όπως και για τις υπόλοιπες, ισχύει ότι:

$$0.0766^2 + 0.0002^2 + 0.4771^2 + 0.6058^2 + 0.5964^2 + 0.2091^2 + (-0.0087)^2 + 0.0035^2 = 1$$

Για τη χρήση των τεσσάρων πρώτων συνιστωσών ο παραπάνω πίνακας έχει ως εξής:

Principal components (eigenvectors)						
Variable	Comp1	Comp2	Comp3	Comp4	Unexplained	
Μεικτό Περιθώριο	0.0766	0.1826	0.7764	0.2977	.181	
Απόσβεση Παγίων	0.0002	0.5744	0.2538	-0.0843	.4158	
Ιδια/Ενεργητικό	0.4771	-0.0701	0.2302	0.1734	.3454	
Ιδια/Ξένα	0.6058	-0.0929	-0.0716	-0.0495	.07637	
Λοιπό Ενεργ/Ξένα	0.5964	-0.0189	-0.0890	-0.0700	.1113	
Βραχ χρέος/Ξένα	0.2091	0.4372	-0.3447	-0.1851	.4349	
Πωλήσεις/Ενεργητικ	-0.0087	0.6160	-0.0690	-0.1988	.3751	
Βραχ χρέος/Κεφ.κίν	0.0035	0.2287	-0.3783	0.8906	.007214	

Παρατηρούμε πως η ανερμήνευτη μεταβλητότητα διαφέρει σε μέγεθος από μεταβλητή σε μεταβλητή. Ο μέσος όρος της ανερμήνευτης μεταβλητότητας είναι 24.33% που προκύπτει αν αφαιρέσουμε από το 100 το ποσοστό της μεταβλητότητας που ερμηνεύεται με τις τέσσερις συνιστώσες δηλαδή το 75.66%.

Αν θεωρηθεί ότι μια συνιστώσα σχετίζεται ισχυρά με μια μεταβλητή όταν η μεταξύ τους συσχέτιση είναι πάνω από 30% ή κάτω από -30%, τότε ο παρακάτω πίνακας, που είναι ο ίδιος ουσιαστικά με τον προηγούμενο, απεικονίζει πόσες φορές μια συνιστώσα σχετίζεται ισχυρά με μια μεταβλητή και σε τι μέγεθος.

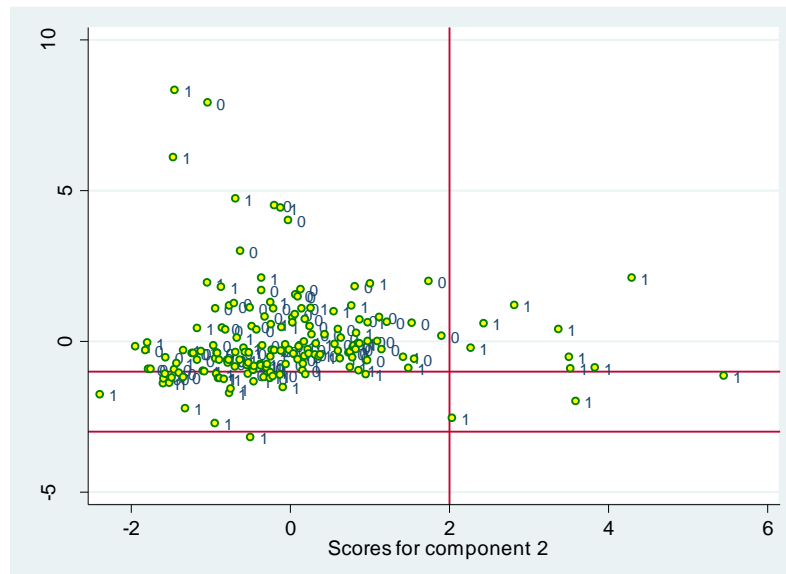
Principal components (eigenvectors) (blanks are abs(loading)<.3)					
Variable	Comp1	Comp2	Comp3	Comp4	Unexplained
Μεικτό Περιθώριο			0.7764		.181
Απόσβεση Παγίων		0.5744			.4158
Ίδια/Ενεργητικό	0.4771				.3454
Ίδια/Ξένα	0.6058				.07637
Λοιπό Ενεργ/Ξένα	0.5964				.1113
Βραχ Χρέος/Ξένα		0.4372	-0.3447		.4349
Πωλήσεις/Ενεργητικ		0.6160			.3751
Βραχ Χρέος/Κεφ.Κίν			-0.3783	0.8906	.007214

Από τον πίνακα προκύπτει ότι δύο συνιστώσες, η 2^η και η 3^η, σχετίζονται ισχυρά με τη μεταβλητή «Βραχυπρόθεσμο Χρέος / Ξένα κεφάλαια», και δύο συνιστώσες, η 3^η και η 4^η, σχετίζονται ισχυρά με τη μεταβλητή «Βραχυπρόθεσμο Χρέος / Κεφάλαιο Κίνησης» με την 4^η να σχετίζεται πολύ ισχυρά με τη μεταβλητή (0.8906).

Στη συνέχεια για την κάθε μία παρατήρηση του δείγματος υπολογίζεται το score της στις 4 συνιστώσες. Έτσι δημιουργούνται 4 νέες μεταβλητές, οι pc1, pc2, pc3 και pc4, που περιέχουν την εκτιμώμενη τιμή της κάθε μία συνιστώσας για την κάθε μία παρατήρηση. Προκειμένου να καθοριστεί ένα “cutoff” το οποίο θα προσδιορίζει το αν και το πότε οι συνιστώσες θα μπορούσαν να δίνουν σινιάλο για τη διακοπή μιας επιχείρησης θα δοθούν τα διαγράμματα διασποράς μεταξύ 1^{ης} και 2^{ης} συνιστώσας, μεταξύ 1^{ης} και 3^{ης} συνιστώσας και μεταξύ 1^{ης} και 4^{ης} συνιστώσας. Οι συνδυασμοί 2^{ης} και 3^{ης} συνιστώσας, 2^{ης} και 4^{ης} συνιστώσας και 3^{ης} και 4^{ης} συνιστώσας δεν θα παρατεθούν γιατί η 2^η, η 3^η και η 4^η συνιστώσα εμπεριέχουν πολύ μικρό ποσοστό διακύμανσης και δεν ενδείκνυνται για την εξαγωγή συμπερασμάτων εν τη απουσία της 1^{ης} συνιστώσας.

Διαγράμματα Διασποράς μεταξύ της 1^{ης} συνιστώσας και των υπολοίπων συνιστωσών

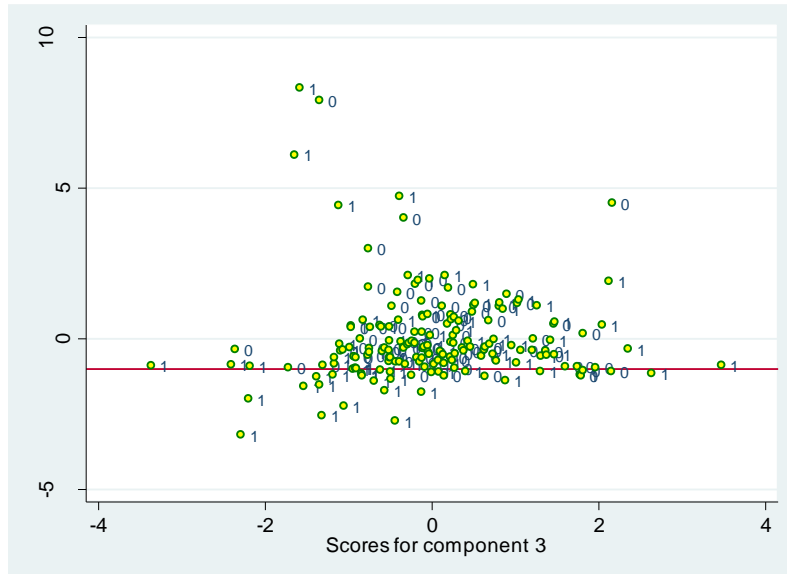
Παρακάτω απεικονίζεται το διάγραμμα διασποράς μεταξύ της 1^{ης} και της 2^{ης} συνιστώσας. Οι κουκίδες με την ένδειξη 1 αποτελούν τις παρατηρήσεις απόκρισης, δηλαδή τις επιχειρήσεις που διέκοψαν τη λειτουργία τους, ενώ οι κουκίδες με την ένδειξη 0 είναι οι επιχειρήσεις που συνεχίζουν τη λειτουργία τους.



Διάγραμμα 5.2-4: Διάγραμμα διασποράς 1ης και 2ης συνιστώσας

Από το διάγραμμα παρατηρείται πως οι παρατηρήσεις που τα σκορ τους είναι πάνω από -3 στην 1^η συνιστώσα και πάνω από $+2$ στη 2^η συνιστώσα είναι όλες παρατηρήσεις απόκρισης. Επίσης οι παρατηρήσεις που έχουν σκορ κάτω από -1 στην 1^η συνιστώσα είναι επίσης παρατηρήσεις απόκρισης. Συνεπώς το σημείο $(2, -3)$ και η ευθεία $y = -1$ μπορούν να αποτελέσουν σημεία “cutoff” για την πρόβλεψη της διακοπής μιας επιχείρησης.

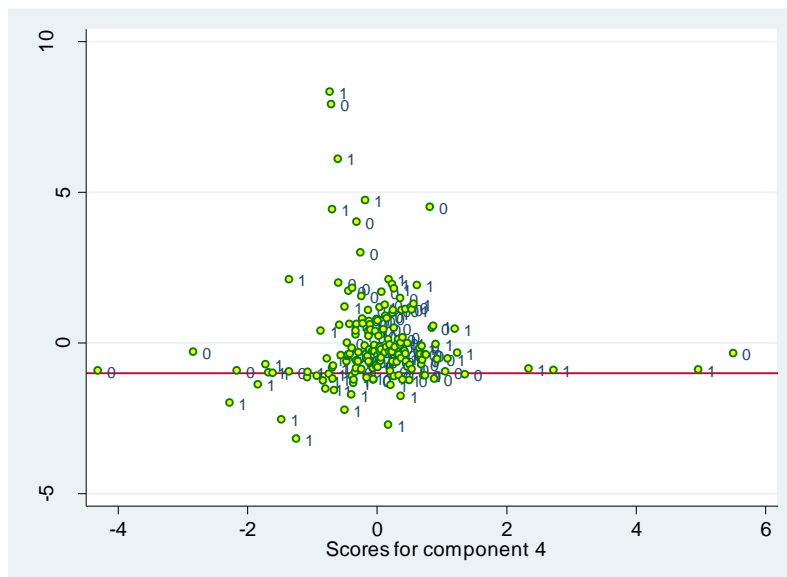
Παρακάτω απεικονίζεται το διάγραμμα διασποράς μεταξύ της 1^{ης} και της 3^{ης} συνιστώσας. Οι κουκίδες στα επόμενα δύο διαγράμματα ακολουθούν την ίδια σημειολογία με παραπάνω.



Διάγραμμα 5.2-5: Διάγραμμα διασποράς 1ης και 3ης συνιστώσας

Από το διάγραμμα παρατηρείται πως οι παρατηρήσεις που έχουν σκορ κάτω από -1 στην 1^η συνιστώσα είναι παρατηρήσεις απόκρισης. Δεν φαίνεται να υπάρχει κάποιο σημείο που να συνδυάζει την 1^η και την 3^η συνιστώσα για να αποτελέσει σημείο cutoff. Συνεπώς σε αυτή την περίπτωση μόνο η ευθεία $y = -1$ μπορεί να θεωρηθεί “cutoff” για την πρόβλεψη της διακοπής μιας επιχείρησης.

Παρακάτω απεικονίζεται το διάγραμμα διασποράς μεταξύ της 1^{ης} και της 4^{ης} συνιστώσας.



Διάγραμμα 5.2-6: Διάγραμμα διασποράς 1ης και 4ης συνιστώσας

Ομοίως με παραπάνω από το διάγραμμα παρατηρείται πως οι παρατηρήσεις που έχουν σκορ κάτω από -1 στην 1^η συνιστώσα είναι παρατηρήσεις απόκρισης. Δεν φαίνεται να υπάρχει κάποιο σημείο που να συνδυάζει την 1^η και την 4^η συνιστώσα για να αποτελέσει σημείο cutoff.

Εκτίμηση κριτηρίου Kaiser – Meyer – Olkin για την επάρκεια του δείγματος

Τέλος ακολουθεί η εκτίμηση του δείκτη Kaiser-Meyer-Olkin (KMO) που σχετίζεται με τη μέτρηση της επάρκειας του δείγματος.

Kaiser-Meyer-Olkin measure of sampling adequacy	
variable	kmo
Μεικτό Περιθώριο	0.3994
Απόσβεση Παγίων	0.5512
Ίδια/Ενεργητικό	0.7347
Ίδια/Ξένα	0.5496
Λοιπό Ενεργ/Ξένα	0.5635
Βραχ Χρέος/Ξένα	0.6748
Πωλήσεις/Ενεργητικ	0.5832
Βραχ Χρέος/Κεφ.κίν	0.5902
overall	0.5840

Οι τιμές του κριτηρίου για όλες τις μεταβλητές, πλην της πρώτης «Μεικτό Περιθώριο», είναι πάνω από το όριο 0.5, που σημαίνει ότι οι μεταβλητές έχουν συσχέτιση και άρα ενδείκνυται για ανάλυση PCA.

Συμπεράσματα από την εφαρμογή της Ανάλυσης Κυρίων Συνιστωσών

Συμπερασματικά, από την Ανάλυση Κυρίων Συνιστωσών προέκυψε ότι αντί για 8 ποσοτικές μεταβλητές στην ανάλυση μπορούν να χρησιμοποιηθούν 4 συνιστώσες που ερμηνεύουν ποσοστό 75.66% της συνολικής μεταβλητότητας. Από τα διαγράμματα διασποράς της 1^{ης} συνιστώσας με τις υπόλοιπες 3 παρατηρήθηκαν δύο σημεία cutoff που μπορούν να χρησιμοποιηθούν για την πρόβλεψη της διακοπής μιας επιχείρησης. Το πρώτο σημείο ορίζεται εκεί όπου οι παρατηρήσεις έχουν σκορ στην 1^η συνιστώσα πάνω από -3 και στην 2^η συνιστώσα πάνω από $+2$. Το δεύτερο σημείο ορίζεται εκεί όπου οι παρατηρήσεις έχουν σκορ στην 1^η συνιστώσα κάτω από -1 . Οι τιμές του κριτηρίου Kaiser – Meyer – Olkin για όλες τις μεταβλητές, πλην αυτής του «Μεικτού Περιθωρίου», είναι πάνω από το όριο 0.5, πράγμα που υπέδειξε ότι οι μεταβλητές έχουν συσχέτιση και άρα ενδείκνυται για ανάλυση PCA.

5.2.4 Εφαρμογή Ανάλυσης κατά Συστάδες

Στην ενότητα αυτή θα παρουσιασθούν τα αποτελέσματα των μεθόδων ανάλυσης κατά συστάδες που παρουσιάστηκαν στο 4^ο κεφάλαιο.

Αρχικά πραγματοποιήθηκε εφαρμογή των μη ιεραρχικών αλγορίθμων. Κατά την εφαρμογή τους ζητήθηκε ο σχηματισμός 2 ομάδων, 3 ομάδων και 4 ομάδων κατηγοριοποίησης των δεδομένων. Αυτό έγινε για να διαπιστωθεί αν ο αλγόριθμος είναι σε θέση να ξεχωρίσει ομάδες επιχειρήσεων που παρουσιάζουν κάποια διαβάθμιση κινδύνου. Αν πράγματι υπήρχε αυτή η δυνατότητα, τότε από την κατηγοριοποίηση θα προέκυπταν ομάδες που η κάθε μια τους θα περιείχε είτε μεγάλο αριθμό επιχειρήσεων που έχουν διακόψει εν συγκρίσει με αυτές που δεν έχουν διακόψει, είτε μεγάλο αριθμό επιχειρήσεων που δεν έχουν διακόψει εν συγκρίσει με αυτές που έχουν διακόψει. Τα αποτελέσματα των κατηγοριοποιήσεων και στις τρεις περιπτώσεις ήταν γενικά παρόμοια, δηλαδή στο σύνολό τους οι αλγόριθμοι σχημάτισαν ομάδες που περιείχαν ίσο πλήθος επιχειρήσεων που έχουν διακόψει και που δεν έχουν διακόψει τη λειτουργία τους. Για τον κάθε αλγόριθμο παρατίθενται τα αποτελέσματα σχηματισμού 2 ομάδων και 4 ομάδων.

A. Μέθοδος ανάλυσης $k - means$

Η μέθοδος αυτή αποτελεί μη ιεραρχική μέθοδο ανάλυσης συσταδοποίησης. Συνεπώς, θα πρέπει να οριστεί εκ των προτέρων ο αριθμός των συστάδων που θα δημιουργήσει ο αλγόριθμος. Ως επιθυμητός αριθμός ομάδων ορίζεται το 2, ούτως ώστε να διαπιστωθεί αν ο αλγόριθμος είναι σε θέση να διακρίνει τις επιχειρήσεις που έχουν κάνει διακοπή και αυτές που είναι σε λειτουργία.

Τα αποτελέσματα της συσταδοποίησης ως προς το πλήθος των ομάδων που σχηματίστηκαν μετά τη συσταδοποίηση είναι τα ακόλουθα:

Omadess k-means	Freq.	Percent	Cum.
1	163	95.32	95.32
2	8	4.68	100.00
Total	171	100.00	

Παρατηρείται ότι οι 2 ομάδες που ζητήθηκε να σχηματισθούν διαφέρουν πάρα πολύ στο πλήθος. Η κατανομή των παρατηρήσεων απόκρισης στις 2 ομάδες φαίνεται ως ακολούθως:

k-means	απόκριση		Total
	0	1	
1	0	1	.49079755
	0	0	.50145588
	83	80	163
2	0	1	.5
	0	0	.53452248
	4	4	8
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Παρατηρείται ότι οι 80 από τις 84 παρατηρήσεις που έχουν διακοπή βρίσκονται στην 1^η ομάδα και οι υπόλοιπες 4 στη 2^η. Συνεπώς σχεδόν όλες οι επιχειρήσεις βρίσκονται στην 1^η ομάδα και ελάχιστες στη 2^η. Συμπεραίνεται ότι ο αλγόριθμος αυτός δεν ενδείκνυται για το σχηματισμό 2 ομάδων που επιχειρείται για τα δεδομένα.

Παρακάτω πραγματοποιείται εφαρμογή του αλγορίθμου με επιθυμητό αριθμό ομάδων το 4. Η κατανομή των παρατηρήσεων απόκρισης στις 4 ομάδες φαίνεται ως ακολούθως:

k-means	απόκριση		Total
	0	1	
1	0	1	.5
	0	0	.53452248
	4	4	8
2	0	1	.75
	0	0	.5
	1	3	4
3	0	1	.48344371
	0	0	.5013888
	78	73	151
4	0	1	.5
	0	0	.53452248
	4	4	8
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Και σε αυτή την περίπτωση παρατηρούνται ομάδες που περιέχουν ίσο πλήθος επιχειρήσεων μεταξύ αυτών που έχουν διακόψει και που δεν έχουν διακόψει τη λειτουργία τους.

B. Μέθοδος ανάλυσης k – median

Η μέθοδος αυτή αποτελεί επίσης μη ιεραρχική μέθοδο ανάλυσης συσταδοποίησης. Η φιλοσοφία του αλγορίθμου είναι ίδια με αυτή του προηγούμενου απλά αντί για τον μέσο όρο ως σημείο αναφοράς χρησιμοποιείται η διάμεσος. Τα αποτελέσματα της συσταδοποίησης ως προς το πλήθος των ομάδων που σχηματίστηκαν μετά τη συσταδοποίηση με επιθυμητό αριθμό ομάδων το 2 είναι τα ακόλουθα:

Omadēs k-median	Freq.	Percent	Cum.
1	130	76.02	76.02
2	41	23.98	100.00
Total	171	100.00	

Τα αποτελέσματα του αλγορίθμου αυτού διαφέρουν από αυτά του προηγούμενου ως προς το πλήθος των παρατηρήσεων εντός των ομάδων. Η κατανομή των παρατηρήσεων απόκρισης στις 2 ομάδες φαίνεται ως ακολούθως:

k-median	απόκριση		Total
	0	1	
1	0	1	.41538462
	0	0	.49469457
	76	54	130
2	0	1	.73170732
	0	0	.44857498
	11	30	41
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Ο αριθμός των επιχειρήσεων που διέκοψαν τη λειτουργία τους και κατηγοριοποιήθηκαν στην ομάδα 1 είναι 54 ενώ οι υπόλοιπες 30 στην ομάδα 2.

Παρακάτω πραγματοποιείται εφαρμογή του αλγορίθμου με επιθυμητό αριθμό ομάδων το 4. Η κατανομή των παρατηρήσεων απόκρισης στις 4 ομάδες φαίνεται ως ακολούθως:

k-median	απόκριση		Total
	0	1	
1	0	1	.38666667
	0	0	.4902656
	46	29	75

2	0	1	.44680851
	0	0	.5025375
	26	21	47
3	0	1	.73170732
	0	0	.44857498
	11	30	41
4	0	1	.5
	0	0	.53452248
	4	4	8
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Και στις δύο κατηγοριοποιήσεις παρατηρείται μια ομάδα, συγκεκριμένα η 3^η που περιέχει μεγαλύτερο αριθμό παρατηρήσεων απόκρισης (30) σε σχέση με τη μη απόκριση (11). Συνεπώς θα μπορούσε να προκύψει ως συμπέρασμα ότι η 3^η ομάδα αυτής της κατηγοριοποίησης είναι πιθανό να υποδεικνύει ύπαρξη κινδύνου διακοπής και ο αλγόριθμος αυτός να είναι σε θέση να δώσει κάποια εικόνα για την πρόβλεψη διακοπής της επιχείρησης.

Παρακάτω ακολουθεί η εφαρμογή των ιεραρχικών αλγορίθμων. Κατά την εφαρμογή τους ζητήθηκε επίσης ο σχηματισμός 2 ομάδων, 3 ομάδων και 4 ομάδων κατηγοριοποίησης των δεδομένων. Τα αποτελέσματα των κατηγοριοποιήσεων και στις τρεις περιπτώσεις ήταν παρόμοια, δηλαδή στο σύνολό τους οι αλγόριθμοι σχημάτισαν ομάδες που περιείχαν ίσο πλήθος επιχειρήσεων που έχουν διακόψει και που δεν έχουν διακόψει τη λειτουργία τους, και κατά συνέπεια κρίνεται ότι οι αλγόριθμοι δεν λειτουργούν ικανοποιητικά στα δεδομένα. Για τον κάθε αλγόριθμο παρατίθενται τα αποτελέσματα σχηματισμού 2 ομάδων. Συγκεκριμένα για τον αλγόριθμο Ward, επειδή χρησιμοποιείται στην πράξη πολύ συχνά, παρατίθενται τα αποτελέσματα σχηματισμού 2, 3 και 4 ομάδων για να δοθεί μια εικόνα των κατηγοριοποιήσεων που προέκυψαν.

C. Μέθοδος Απλής Συνένωσης (Simple Linkage Method)

Η μέθοδος αυτή αποτελεί ιεραρχική μέθοδο ανάλυσης συσταδοποίησης. Μετά τον υπολογισμό των αποστάσεων από τον αλγόριθμο, όταν ζητηθεί ο σχηματισμός 2 τελικών ομάδων, η κατανομή των παρατηρήσεων στις 2 ομάδες έχει ως ακολούθως:

Clus	Freq.	Percent	Cum.
1	2	1.17	1.17

2	169	98.83	100.00
-----+-----			
Total	171	100.00	

Η κατανομή των παρατηρήσεων απόκρισης στις 2 ομάδες φαίνεται ως ακολούθως:

Clus	απόκριση		Total
	0	1	
1	0	1	.5
	0	0	.70710678
	1	1	2
-----+-----			
2	0	1	.49112426
	0	0	.50140687
	86	83	169
-----+-----			
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Ο αριθμός των επιχειρήσεων που διέκοψαν τη λειτουργία τους και κατηγοριοποιήθηκαν στην ομάδα 1 είναι 1 ενώ οι υπόλοιπες 83 στην ομάδα 2 μαζί με όλες επιχειρήσεις που δεν διέκοψαν τη λειτουργία τους.

D. Μέθοδος Πλήρους Συνένωσης (Complete Linkage Method)

Η μέθοδος αυτή αποτελεί ιεραρχική μέθοδο ανάλυσης συσταδοποίησης. Μετά τον υπολογισμό των αποστάσεων από τον αλγόριθμο, όταν ζητηθεί ο σχηματισμός 2 τελικών ομάδων, η κατανομή των παρατηρήσεων στις 2 ομάδες έχει ως ακολούθως:

Clus	Freq.	Percent	Cum.
1	167	97.66	97.66
2	4	2.34	100.00
-----+-----			
Total	171	100.00	

Η κατανομή των παρατηρήσεων απόκρισης στις 2 ομάδες φαίνεται ως ακολούθως:

Clus	απόκριση		Total
	0	1	
1	0	1	.48502994
	0	0	.50127894
	86	81	167
-----+-----			
2	0	1	.75
	0	0	.5

	1	3	4
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Ο αριθμός των επιχειρήσεων που διέκοψαν τη λειτουργία τους και κατηγοριοποιήθηκαν στην ομάδα 1 είναι 81 μαζί με 86 επιχειρήσεις που δεν διέκοψαν τη λειτουργία τους.

E. Μέθοδος Σταθμισμένων Μέσων (Weighted Average Linkage Method)

Η μέθοδος αυτή αποτελεί ιεραρχική μέθοδο ανάλυσης συσταδοποίησης. Μετά τον υπολογισμό των αποστάσεων από τον αλγόριθμο, όταν ζητηθεί ο σχηματισμός 2 τελικών ομάδων, η κατανομή των παρατηρήσεων στις 2 ομάδες έχει ως ακολούθως:

Clus	Freq.	Percent	Cum.
1	167	97.66	97.66
2	4	2.34	100.00
Total	171	100.00	

Η κατανομή των παρατηρήσεων απόκρισης στις 2 ομάδες φαίνεται ως ακολούθως:

Clus	απόκριση		Total
	0	1	
1	0	1	.48502994
	0	0	.50127894
	86	81	167
2	0	1	.75
	0	0	.5
	1	3	4
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Η μέθοδος δίνει τα ίδια αποτελέσματα με την προηγούμενή της. Ο αριθμός των επιχειρήσεων που διέκοψαν τη λειτουργία τους και κατηγοριοποιήθηκαν στην ομάδα 1 είναι 81 μαζί με 86 επιχειρήσεις που δεν διέκοψαν τη λειτουργία τους.

F. Μέθοδος Κέντρων Βάρους (Centroid Linkage Method)

Η μέθοδος αυτή αποτελεί ιεραρχική μέθοδο ανάλυσης συσταδοποίησης. Μετά τον υπολογισμό των αποστάσεων από τον αλγόριθμο, όταν ζητηθεί ο σχηματισμός 2 τελικών ομάδων, η κατανομή των παρατηρήσεων στις 2 ομάδες έχει ως ακολούθως:

Clus	Freq.	Percent	Cum.
1	170	99.42	99.42
2	1	0.58	100.00
Total	171	100.00	

Η κατανομή των παρατηρήσεων απόκρισης στις 2 ομάδες φαίνεται ως ακολούθως:

Clus	απόκριση		Total
	0	1	
1	0	1	.48502994
	0	0	.50127894
	86	81	167
2	0	1	.75
	0	0	.5
	1	3	4
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Ο αριθμός των επιχειρήσεων που διέκοψαν τη λειτουργία τους και κατηγοριοποιήθηκαν στην ομάδα 1 είναι 81 μαζί με 86 επιχειρήσεις που δεν διέκοψαν τη λειτουργία τους.

G. Μέθοδος Ward

Η μέθοδος αυτή αποτελεί ιεραρχική μέθοδο ανάλυσης συσταδοποίησης. Μετά τον υπολογισμό των αποστάσεων από τον αλγόριθμο, όταν ζητηθεί ο σχηματισμός 2 τελικών ομάδων, η κατανομή των παρατηρήσεων στις 2 ομάδες έχει ως ακολούθως:

Clus	Freq.	Percent	Cum.
1	167	97.66	97.66
2	4	2.34	100.00
Total	171	100.00	

Η κατανομή των παρατηρήσεων απόκρισης στις 2 ομάδες φαίνεται ως ακολούθως:

Clus	απόκριση		Total
	0	1	

	0	1	Total
1	0	1	.48502994
	0	0	.50127894
	86	81	167
2	0	1	.75
	0	0	.5
	1	3	4
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Η μέθοδος δίνει τα ίδια αποτελέσματα με την προηγούμενή της. Ο αριθμός των επιχειρήσεων που διέκοψαν τη λειτουργία τους και κατηγοριοποιήθηκαν στην ομάδα 1 είναι 81 μαζί με 86 επιχειρήσεις που δεν διέκοψαν τη λειτουργία τους.

Η κατανομή των παρατηρήσεων απόκρισης στις 3 ομάδες φαίνεται ως ακολούθως:

Clus	απόκριση		Total
	0	1	
1	0	1	.48427673
	0	0	.50133172
	82	77	159
2	0	1	.5
	0	0	.53452248
	4	4	8
3	0	1	.75
	0	0	.5
	1	3	4
Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Η κατανομή των παρατηρήσεων απόκρισης στις 4 ομάδες φαίνεται ως ακολούθως:

Clus	απόκριση		Total
	0	1	
1	0	1	.48026316
	0	0	.50126191
	79	73	152
2	0	1	.57142857
	0	0	.53452248
	3	4	7
3	0	1	.5

	0	0	.53452248
	4	4	8

4	0	1	.75
	0	0	.5
	1	3	4

Total	0	1	.49122807
	0	0	.50139125
	87	84	171

Όπως και στην κατανομή των παρατηρήσεων στις 2 ομάδες, έτσι και στην κατανομή των παρατηρήσεων στις 3 και τις 4 ομάδες σχηματίζονται ομάδες ίσου πλήθους επιχειρήσεων που διέκοψαν και που δεν διέκοψαν τη λειτουργία τους με ιδιαίτερα μεγάλη συγκέντρωση πλήθους επιχειρήσεων στην πρώτη σχηματισθείσα ομάδα σε κάθε περίπτωση.

Συμπεράσματα από την εφαρμογή της Ανάλυσης κατά Συστάδες

Συμπερασματικά, η συσταδοποίηση δεν ενδείκνυται για το συγκεκριμένο σετ δεδομένων. Αυτό μπορεί να οφείλεται στο ότι το δείγμα είναι αρκετά μικρό για να διενεργηθεί μια ευρύτερη ανάλυση. Θα μπορούσε να θεωρηθεί ότι ο αλγόριθμος k – median έδωσε εν μέρει ένα ικανοποιητικό αποτέλεσμα κατηγοριοποίησης.

ΚΕΦΑΛΑΙΟ 6

Συμπεράσματα και προτάσεις για περαιτέρω έρευνα

Στη διπλωματική αυτή εφαρμόστηκαν μέθοδοι Πολυμεταβλητής Ανάλυσης στα οικονομικά δεδομένα 6.500 περίπου ελληνικών βιομηχανικών επιχειρήσεων με σκοπό την εκτίμηση της οικονομικής τους βιωσιμότητας και την ανίχνευση ενδείξεων που μπορεί να υποδεικνύουν κίνδυνο μελλοντικής διακοπής των επιχειρηματικών τους εργασιών. Το σύνολο των μεταβλητών που χρησιμοποιήθηκε αφορούσε χρηματοοικονομικούς αριθμοδείκτες που σχηματίστηκαν με βάση τα διαθέσιμα οικονομικά στοιχεία που αντλήθηκαν από την ιστοσελίδα inf.gr για τις εξετασθείσες βιομηχανίες.

Τέσσερις στατιστικές μέθοδοι πολυμεταβλητής ανάλυσης εφαρμόστηκαν στα δεδομένα προκειμένου να διαπιστωθεί αν τελικά μπορεί να διαμορφωθεί αξιόπιστο μοντέλο πρόβλεψης της διακοπής της λειτουργίας μιας επιχείρησης από το σύνολο των συγκεκριμένων δεδομένων.

Οι δύο μέθοδοι προέρχονται από την κατηγορία των γενικευμένων γραμμικών μοντέλων και αφορούσαν τα μοντέλα logit, δηλαδή τη λογιστική παλινδρόμηση, και probit. Τα δύο μοντέλα έδωσαν παρόμοιες εκτιμήσεις των συντελεστών τους και είχαν αρκετά καλή προβλεπτική ικανότητα για την επικείμενη διακοπή των εργασιών της επιχείρησης. Από τα μοντέλα προέκυψαν 9 σημαντικές μεταβλητές από τις οποίες οι 8 αναφέρονται στα εξής ποσοτικά μεγέθη: Ίδια Κεφάλαια προς Ενεργητικό, Βραχυπρόθεσμο Χρέος προς Ξένα Κεφάλαια, Απόσβεση Παγίων, Πωλήσεις προς Ενεργητικό, Μεικτό Περιθώριο Κέρδους, Βραχυπρόθεσμο Χρέος προς Κεφάλαιο Κίνησης, Ίδια Κεφάλαια προς Ξένα Κεφάλαια, Λοιπό Ενεργητικό προς Ξένα Κεφάλαια ενώ η τελευταία αφορά την ποιοτική μεταβλητή του Γεωγραφικού Τομέα που ανήκει η επιχείρηση. Η μεταβλητή η οποία κρίθηκε περισσότερο σημαντική στο μοντέλο πρόβλεψης ήταν αυτή που αφορούσε τον αριθμοδείκτη «Ίδια Κεφάλαια προς Ενεργητικό».

Η τρίτη μέθοδος που εφαρμόστηκε στα δεδομένα ήταν η Ανάλυση Κυρίων Συνιστωσών. Η μέθοδος εφαρμόστηκε στις παραπάνω 8 ποσοτικές μεταβλητές που προέκυψαν από την εκτίμηση του μοντέλου της Λογιστικής Παλινδρόμησης. Από την εφαρμογή του μοντέλου της Ανάλυσης Κύριων Συνιστωσών προέκυψε ότι μπορεί να επιτευχθεί μείωση της διάστασης των δεδομένων από το σύνολο των 8 μεταβλητών στο σύνολο των 4 μεταβλητών – κυρίων

συνιστώσών που ερμήνευσαν ποσοστό ίσο με το 75.66% της συνολικής μεταβλητότητας. Από τα διαγράμματα διασποράς της 1^{ης} συνιστώσας με τις υπόλοιπες 3 παρατηρήθηκαν δύο σημεία cutoff που μπορούν να χρησιμοποιηθούν για την πρόβλεψη της διακοπής μιας επιχείρησης. Το πρώτο σημείο ορίστηκε εκεί όπου οι παρατηρήσεις είχαν σκορ στην 1^η συνιστώσα πάνω από -3 και στην 2^η συνιστώσα πάνω από +2. Το δεύτερο σημείο ορίστηκε εκεί όπου οι παρατηρήσεις είχαν σκορ στην 1^η συνιστώσα κάτω από -1.

Τέλος, στα δεδομένα εφαρμόστηκαν τεχνικές συσταδοποίησης προκειμένου να διαπιστωθεί αν οι γνωστοί αλγόριθμοι συσταδοποίησης είναι σε θέση να ανιχνεύσουν και να ομαδοποιήσουν κατάλληλα τις οικονομικά βιώσιμες και μη βιώσιμες επιχειρήσεις. Οι αλγόριθμοι αυτοί δεν έδωσαν ικανοποιητικά αποτελέσματα πέραν του αλγορίθμου k - median που θα μπορούσε να χρησιμοποιηθεί και να δώσει σε ένα μικρό βαθμό ικανοποιητικές κατηγοριοποιήσεις.

Το βασικότερο πρόβλημα που υπήρξε στη διαδικασία της ανάλυσης, και που στάθηκε καθοριστικό στην εξαγωγή των συμπερασμάτων, ήταν ο μικρός αριθμός των παρατηρήσεων που αποτέλεσαν το σύνολο των επιχειρήσεων που διέκοψαν τη λειτουργία τους, δηλαδή τις παρατηρήσεις απόκρισης. Ενώ στην αρχή οι παρατηρήσεις της απόκρισης ήταν 149, στη συνέχεια, μετά τον καθαρισμό των δεδομένων και τη κατάλληλη διαμόρφωσή των μεταβλητών για την εξέταση των χρονικών τους υστερήσεων, οι παρατηρήσεις που απέμειναν να αποτελούν την απόκριση ήταν μόλις 87. Αυτό επηρέασε και το συνολικό δείγμα που χρησιμοποιήθηκε για να μπορέσει να επιτευχθεί επαρκής αναλογία δεδομένων απόκρισης και βάσης για την εφαρμογή των μοντέλων.

Σε ότι αφορά τις προτάσεις για μελλοντική έρευνα, τα δεδομένα που αντλήθηκαν αποτελούν εξαιρετικό σύνολο οικονομικών δεδομένων που με τα χρόνια θα εμπλουτίζεται και θα μπορεί να αποτελέσει πολύ σημαντικό εργαλείο σε κάθε ερευνητή που θα ήθελε να μελετήσει περαιτέρω τη βιωσιμότητα των ελληνικών βιομηχανικών επιχειρήσεων. Μειονέκτημα αποτελεί το γεγονός ότι στη συγκεκριμένη ιστοσελίδα δεν περιλαμβάνονται δεδομένα πολλών βιομηχανικών επιχειρήσεων που διέκοψαν τη λειτουργία τους με αποτέλεσμα η εφαρμογή μοντέλων κατηγοριοποίησης να μην παρουσιάζει τη μέγιστη αποτελεσματικότητά της. Επιπλέον μοντέλα κατηγοριοποίησης, πιο σύγχρονα και πιο εξελιγμένα, όπως τα νευρωνικά δίκτυα, οι γενετικοί αλγόριθμοι, τα δέντρα αποφάσεων κ.α. μπορούν να εφαρμοστούν και να διερευνηθεί περαιτέρω η προβλεπτική τους ικανότητα στην ανίχνευση ενδείξεων πιστωτικού κινδύνου που μπορεί να ελλοχεύει στην οικονομική δραστηριότητα των επιχειρήσεων είτε συγκεκριμένων κλάδων είτε γενικότερα στο σύνολο των ελληνικών βιομηχανιών.

Βιβλιογραφία

- Βασιλείου, Δ. & Ηρειώτης, Ν., 2008. *Χρηματοοικονομική Διοίκηση - Θεωρία και Πρακτική*. 1η επιμ. Αθήνα: Rosili.
- Γκούμας, Σ. & Λεκαράκου, Α., 2013. Σημειώσεις στο "Διπλογραφικό Λογιστικό Σύστημα" Γενική και Αναλυτική Λογιστική Εκμετάλλευσης . Στο: Αθήνα: Εθνικό Κέντρο Δημόσιας Διοίκησης και Αυτοδιοίκησης Ινστιτούτο Επιμόρφωσης.
- Κούτρας, Μ., 2012. *Εφαρμοσμένη Πολυμεταβλητή Ανάλυση*. Πειραιάς: Πανεπιστήμιο Πειραιώς.
- Νιάρχος, Ν., 2004. *Χρηματοοικονομική Ανάλυση Λογιστικών Καταστάσεων*. 7η επιμ. Αθήνα: Σταμούλης.
- Παπαναστασόπουλος, Γ., 2007. *Θεμελιώδη Λογιστικά Μεγέθη ως Παράγοντες Πρόβλεψης Εταιρικών Αποδόσεων και Διαχείρισης Πιστωτικού Κινδύνου*, Πειραιάς: Διδακτορική Διατριβή.
- Παπαναστασόπουλος, Γ., 2014. *Ειδικά Θέματα Λογιστικής* . Στο: Αθήνα: Πανεπιστήμιο Πειραιώς, ΠΜΣ στη Λογιστική και Έλεγχο Δημοσίων Οργανισμών και Επιχειρήσεων.
- Πελέκης, Ν., 2014. *Στατιστικές Μέθοδοι Εξόρυξης Δεδομένων*. Πειραιάς: Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης.
- Πολίτης, Κ., 2012. *Γενικευμένα Γραμμικά Μοντέλα*. Πειραιάς: Πανεπιστήμιο Πειραιώς
- Abdou, H. A. & Pointon, J., 2011. Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature. *Intelligent Systems in Accounting, Finance & Management*, pp. 18 (2-3), pp. 59-88.
- Abdou, H. A. & Pointon, J., 2009. Credit scoring and decision making in Egyptian public sector banks. *International Journal of Managerial Finance*, 5(4), pp. 391-406.
- Agarwal, V. & Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, pp. 32(8), 1541-1551.
- Altman, E., Hadelmann, R. & Narayanan, P., 1977. ZETA analysis: a new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance*, June, pp. pp470-492.
- Altman, E. I. & Saunders, A., 1998. Credit risk measurement: Developments over the last 20 years. *Journal of banking and finance*, p. 21.11/12.
- Anderberg, M. R., 2014. *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks*. 19 επιμ. s.l.:Academic press.

- Arminger, G., Enache, D. & Bonne, T., 1997. Analyzing credit risk data: A comparison of logistic discrimination, classification tree analysis, and feedforward networks. *Computational Statistics*, p. 12(2).
- Baesens, B. και συν., 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, 54(6), pp. 627-635.
- Banasik, J., Crook, J. & Thomas, L., 2001. Scoring by usage. *Journal of the Operational Research Society*, pp. 997-1006.
- Beaver, W., 1968. Market prices, financial ratios, and the prediction of failure. *Journal of accounting research*, pp. 179-192.
- Ben-David, A. & Frank, E., 2009. Accuracy of machine learning models versus “hand crafted” expert systems—a credit scoring case study. *Expert Systems with Applications*, 36(3), pp. 5264-5271.
- Bensic, M., Sarlija, N. & Zekic-Susac, M., 2005. Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance and Management* , 13(3), pp. 133-150.
- Beynon, M. J., 2005. Optimizing object classification under ambiguity/ignorance: application to the credit rating problem. *Intelligent Systems in Accounting, Finance and Management*, 13(2), pp. 113-130.
- Black, F. & Scholes, M., 1973. The pricing of options and corporate liabilities. *The journal of political economy*, pp. 637-654.
- Boyes, W. J., Hoffman, D. L. & Low, S. A., 1989. An econometric analysis of the bank credit scoring problem. *Journal of Econometrics*, 40(1), pp. 3-14.
- Breiman, L., Friedman, J., Stone, C. J. & Olshen, A. R., 1984. *Classification and regression trees*. s.l.:CRC press.
- Caouette, J. B., Altman, E. I. & Narayanan, P., 1998. Managing credit risk: the next great financial challenge. *Vol. 2. John Wiley & Sons*.
- Coffman, J. Y., 1986. The proper role of tree analysis in forecasting the risk behavior of borrowers. *Management Decision Systems, Atlanta, MDS Reports*, pp. 3(4), 7.
- Crook, J. N., Edelman, D. B. & Thomas, L. C., 2007. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), pp. 1447-1465.
- Desai, V. S., Crook, J. N. & Overstreet, G. A., 1996. A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95(1), pp. 24-37.

- Durand, D., 1941. Risk elements in consumer instalment financing. *NBER Books*.
- Eisenbeis, R. A., 1977. Pitfalls in the application of discriminant analysis in business, finance, and economics. *The Journal of Finance*, 32(3), pp. 875-900.
- Fisher, R. A., 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), pp. 179-188.
- Fletcher, D. & Goss, E., 1993. Forecasting with neural networks: an application using bankruptcy data. *Information & Management*, 24(3), pp. 159-167.
- Glen, J. J., 2001. Classification accuracy in discriminate analysis: a mixed integer programming approach. *Journal of the Operational Research Society*, 52(3), pp. 328-339.
- Greene, W., 1998. Sample selection in credit-scoring models. *Japan and the world economy*, 10(3), pp. 299-316.
- Green, M. D. & Swets, J. A., 1966. *Signal Detection Theory and Psychophysics*. Ney York: Wiley.
- Hand, D. J. & Henley, W. E., 1997. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), pp. 523-541.
- Hand, D. J. & Jacka, S. D., 1998. *Statistics in finance*. London: Arnold.
- Haughwout, A., Peach, R. & Tracy, J., 2008. Juvenile delinquent mortgages: Bad credit or bad economy?. *Journal of Urban Economics*, 64(2), pp. 246-257.
- Henley, E. W. & Hand, D. J., 1996. A k-nearest-neighbour classifier for assessing consumer credit risk. *The Statistician*, 45(1), pp. 77-95.
- Heuson, A., Passmore, W. & Sparks, R., 2001. Credit scoring and mortgage securitization: Implications for mortgage rates and credit availability. *The Journal of Real Estate Finance and Economic*, 23(3), pp. 337-363.
- Hosmer, W. D., Lameshow, S. A. & Sturdivant, R. X., 2013. *Applied Logistic Regression*. 3rd ed. Hoboken επιμ. NJ: Wiley.
- Hsieh, N.-C., 2004. An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert systems with applications*, 27(4), pp. 623-633.
- Hsieh, N.-C. & Hung, L.-P., 2010. A data driven ensemble classifier for credit scoring analysis. *Expert Systems with Applications*, pp. 37(1), 534-545.
- Hu, Y.-C. & Ansell, J., 2007. Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research*, 183(3), pp. 1595-1606.

ICAP GROUP A.E, Αύγουστος 2013. *Μεθοδολογία Απόδοσης Αξιολογήσεων Πιστοληπτικής Ικανότητας*, Αθήνα: ICAP GROUP.

Jo, H., Han, I. & Lee, H., 1997. Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Systems with Applications*, 13(2), pp. 97-108.

Laha, A., 2007. Building contextual classifiers by integrating fuzzy rule based classification technique and k-nn method for credit scoring. *Advanced Engineering Informatics*, 21(3), pp. 281 - 291.

Laha, A., 2007. Building contextual classifiers by integrating fuzzy rule based classification technique and k-nn method for credit scoring. *Advanced Engineering Informatics*, pp. 21(3), 281-291.

Lee, T.-S. & Chen, I.-F., 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4), pp. 743-752.

Leonard, K. J., 1993. Detecting credit card fraud using expert systems. *Computers & industrial engineering*, 25(1-4), pp. 103-106.

Leonard, K. J., 1995. The development of credit scoring quality measures for consumer credit applications. *International Journal of Quality & Reliability Management*, 12(4), pp. 79-85.

Leshno, M. & Spector, Y., 1996. Neural network prediction analysis: The bankruptcy case. *Neurocomputing*, 10(2), pp. 125-147.

Lim, M. K. & Sohn, S. Y., 2007. Cluster-based dynamic scoring model. *Expert Systems with Applications*, 32(2), pp. 427-431.

Martin, D., 1977. Early warning of bank failure: A logit regression approach. *Journal of banking and finance*, pp. 1.3: 249-276.

Merton, R. C., 1973. Theory of rational option pricing. *The Bell Journal of economics and management science*, pp. 141 - 183.

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, pp. 29, 449-470.

Min, J. H. & Lee, Y.-C., 2008. A practical approach to credit scoring. *Expert Systems with Applications*, pp. 35(4), 1762-1770.

Nanni, L. & Lumini, A., 2009. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, pp. 3028-3033.

Ohlson, J. A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, pp. 109-131.

- Orgler, Y., 1971. Evaluation of bank consumer loans with credit scoring models. *el-Aviv University, Department of Environmental Sciences*.
- Pendharkar, P. C., 2005. A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem. *Computers & Operations Research*, 32(10), p. 2561–2582.
- Quah, J. T. & Sriganesh, M., 2008. Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications*, 35(4), pp. 1721-1732.
- Salchenberger, L. M., Cinar, E. & Lash, N. A., 1992. Neural networks: A new tool for predicting thrift failures. *Decision Sciences*, 23(4), pp. 899-916.
- Sarlija, N., Bensic, M. & Zekic-Susac, M., 2009. Comparison procedure of predicting the time to default in behavioural scoring. *Expert Systems with Applications*, 36(5), pp. 8778-8788.
- Steenackers, A. & Goovaerts, M. J., 1989. A credit scoring model for personal loans. *Insurance: Mathematics and Economics*, 8(8), pp. 31-34.
- Sustersic, M., Mramor, D. & Zupan, J., 2009. Consumer credit scoring models with limited data. *Expert Systems with Applications*, 36(3), pp. 4736-4744.
- Tsai, C.-F., 2014. Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion*, pp. 16, 46-58.
- West, R. C., 1985. A factor-analytic approach to bank condition. *Journal of Banking & Finance*, pp. 9(2), 253-266.
- Yap, B. W., Ong, S. H. & Husain, N. H. M., 2011. Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications*, pp. 38(10), 13274-13283.
- Zmijewski, M. E., 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, pp. 59-82.

Ηλεκτρονική Βιβλιογραφία

- Δασίλας, Α., 2012. *Πρόγραμμα μεταπτυχιακών σπουδών στη Διοίκηση Επιχειρήσεων, ΤΕΙ Κεντρικής Μακεδονίας*. [Ηλεκτρονικό]
Available at: <http://business.teicm.gr/pms/sites/default/files/pdfs/Analysi%20katastasevn.pdf>
[Πρόσβαση 20 July 2015].
- Όρων, Ε. Ο., 2015. *Ευρετήριο Οικονομικών Όρων*. [Ηλεκτρονικό]
Available at: <http://www.euretirio.com/apodosi-idion-kefalaion-roe/>
[Πρόσβαση 20 July 2015].

Όρων, Ε. Ο., 2015. *Ευρετήριο Οικονομικών Όρων*. [Ηλεκτρονικό]

Available at: <http://www.euretirio.com/kostos-polithenton/>

[Πρόσβαση 22 May 2015].

Ευρετήριο, Ο. Ο., 2015. *Ευρετήριο Οικονομικών Όρων*. [Ηλεκτρονικό]

Available at: <http://www.euretirio.com/kostos-polithenton/>

[Πρόσβαση 22 May 2015].

Ependysopedia, 2014. *Ependysopedia*. [Ηλεκτρονικό]

Available at: <http://www.ependysopedia.gr/ebitda>

[Πρόσβαση 22 July 2015].

Faulkenberry, K., 2015. *Arbor Investment Planner*. [Ηλεκτρονικό]

Available at: <http://arborinvestmentplanner.com/return-on-total-assets-ratios-calculations/>

[Πρόσβαση 20 July 2015].

InvestorWords, 2015. *Investor Words*. [Ηλεκτρονικό]

Available at: http://www.investorwords.com/6797/EBIT_margin.html

[Πρόσβαση 22 July 2015].

Katchova, A., 2013. <https://sites.google.com/site/econometricsacademy/>. [Ηλεκτρονικό]

Available at:

<https://docs.google.com/file/d/0BwogTI8d6EEidjg2OGl4b3dxVmc/edit?pref=2&pli=1>

[Πρόσβαση 15 03 2016].

UCLA, 2016. *IDRE (Intitute for Digital Reserach and Education) - UCLA*. [Ηλεκτρονικό]

Available at: http://www.ats.ucla.edu/stat/stata/output/stata_logistic.htm

[Πρόσβαση 16 04 2016].

