

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ (Πάτρα)

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Τεχνικές Προβλέψεων στον τομέα της Υγειονομικής Περίθαλψης

Predictive Analytics in Healthcare

Παπαπούλου Αργυρώ

Τόγιας Παναγιώτης



Υπό την επίβλεψη του
Αντζουλάτου Γεράσιμου

ΠΑΤΡΑ, ΙΟΥΛΙΟΣ 2015

The only way to predict the future is to
have power to shape the future. - *Eric Hoffer*

Ευχαριστίες

Πρώτα, θέλουμε να ευχαριστήσουμε τον καθηγητή μας κ. Αντζουλάτο Γεράσιμο για την εμπιστοσύνη που μας έδειξε καθώς και για την καθοδήγησή του καθ' όλη την περίοδο της ετοιμασίας αυτής της πτυχιακής εργασίας.

Επίσης ευχαριστούμε το Τμήμα Περιφερειακού Χάρτη και Πρωτοβάθμιας Φροντίδας Υγείας της 6ης Υγειονομικής Περιφέρειας Πάτρας, Το Γενικό Νοσοκομείο Πατρών "Ο Άγιος Ανδρέας" και το Πανεπιστημιακό Γενικό Νοσοκομείο Ρίου "Παναγία η βοήθεια", για την συνεργασία και τα στοιχεία που μας παρείχαν.

Τέλος, ευχαριστούμε ιδιαίτερα τους φίλους και την οικογένειά μας για την κατανόηση, την συμπαράσταση και την υπομονή που έδειξαν όλο αυτό τον καιρό.

Περίληψη

Ο όγκος των συσσωρευμένων ηλεκτρονικών δεδομένων στον τομέα της Υγείας και Πρόνοιας συνεχώς αυξάνεται. Τεράστιες ποσότητες ιατρικών δεδομένων, όπως για παράδειγμα κλινικά δεδομένα, ιατρικές εικόνες και εγγραφές ασθενών, συμπεριφορά και προτιμήσεις ασθενών, οικονομικά δεδομένα, ζήτηση και χρησιμοποίηση των πόρων υγείας και του ανθρώπινου δυναμικού συλλέγονται στα πληροφοριακά συστήματα των δομών υγείας και πρόνοιας. Η επεξεργασία των δεδομένων αυτών απαιτεί ολοκληρωμένες και ευφυείς λύσεις ικανές για ανίχνευση έγκυρων και αξιόπιστων πληροφοριών χρήσιμων στη λήψη αποφάσεων και προβλέψεων. Οι τεχνικές προγνωστικής μοντελοποίησης (predictive analytics) αφορούν στην ανάλυση των ιστορικών δεδομένων με στόχο τις ασφαλείς προβλέψεις για μελλοντικά γεγονότα και τάσεις. Η χρησιμοποίηση από τους οργανισμούς τέτοιων τεχνικών βοηθά στο να καθορίσουν τις σχέσεις μεταξύ του κόστους, της ποιότητας και να προσδιορίσουν τις βέλτιστες κλινικές πρακτικές καθώς και να προβλέψουν και αξιολογήσουν ατομικά και πληθυσμιακά επίπεδα κινδύνων στον τομέα της υγείας.

Στην παρούσα πτυχιακή εργασία διεξήχθη έρευνα για το πώς οι τιμές δεικτών υγείας, χωρών εντός και εκτός Ευρωπαϊκής Ένωσης, θα κυμανθούν στο μέλλον με τη βοήθεια τεχνικών προβλέψεων και αναλύσεων. Στόχος είναι η ανάλυση των αποτελεσμάτων από κάθε «σενάριο» και η σύγκριση των αντίστοιχων τιμών των δεικτών ενδιαφέροντος μεταξύ των χωρών αυτών με απώτερο σκοπό την εξαγωγή χρήσιμων συμπερασμάτων. Ένας δεύτερος στόχος της πτυχιακής εργασίας ήταν η συλλογή δεδομένων υγειονομικής περίθαλψης από τα δύο κυριότερα νοσοκομεία του Ν. Αχαΐας και η ανάλυσή τους για εξαγωγή συμπερασμάτων. Η ανάλυση και παρουσίαση των αποτελεσμάτων έγινε με χρήση του λογισμικού πακέτου ανοιχτού κώδικα R.

Στο πρώτο κεφάλαιο αναλύονται εισαγωγικές έννοιες περί υγειονομικής περίθαλψης καθώς και οι δείκτες υγείας που χρησιμοποιήθηκαν. Στο δεύτερο κεφάλαιο γίνεται αναφορά στις τεχνικές προβλέψεων και τον τρόπο λειτουργίας τους και ακόμα αναφέρονται γνωστές μελέτες περιπτώσεων τεχνικών προβλέψεων

στην υγειονομική περίθαλψη. Τέλος, στο τρίτο κεφάλαιο αναλύονται με χρήση των προαναφερθέντων τεχνικών προβλέψεων και εξάγονται αποτελέσματα και συμπεράσματα από δεδομένα που αντλήθηκαν ύστερα από πρωτογενή και δευτερογενή έρευνα.

Λέξεις Κλειδιά: Υγειονομική Περίθαλψη, Τεχνικές Προβλέψεων, Εξόρυξη Δεδομένων

ABSTRACT

The volume of digital data in Healthcare keeps rising. Huge amount of medical data such as, clinical data, patients' scans, logs, behavior and preferences, economic data, demand and the use of medical and human resources are being collected in information systems of Healthcare. The elaboration of these data requires efficient and intelligent clarifications, capable of tracing valid and trustful information which will be useful for the decision making process and prediction. Predictive Analytics analyze past data and are able to give trustful outcomes (predictions) for future circumstances and trends. Organizations use these kind of techniques in order to determine the relationships between costs and quality and also to redefine better medical approaches as well as to predict and evaluate individual and general levels of danger in Healthcare.

The aspect of this thesis is the research that has been made about how European Union and non-European Union countries' health indicators, will hover in future with the use of Predictive Analytics. This aspect, eventually leads to results from each time different scenarios, as well to the comparison of their values with the purpose of creating conclusions. In order for this to happen, we collected data concerning healthcare from the two main public sector hospitals in the state Achaia. The presentation and analysis of the results was made with the use of the open-source software, R.

In the first chapter, we analyze introductory concepts about healthcare and we present its aspects co-existing with Predictive Analytics while health indicators that have been used are mentioned analytically. The techniques and the use of Predictive Analytics, in each case, are presented in the second chapter, with references to well-known Predictive Analytics related case studies. Finally, the third chapter hosts the presentation of the results and conclusions, with the elaboration made by the prediction techniques, after preliminary and secondary research.

Key Words: Healthcare, Predictive Analytics, Data Mining

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	III
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ ΕΙΚΟΝΩΝ.....	IV
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ ΠΙΝΑΚΩΝ	V
1° ΚΕΦΑΛΑΙΟ.....	6
ΥΓΕΙΟΝΟΜΙΚΗ ΠΕΡΙΘΑΛΨΗ	6
1.1 Η ΥΓΕΙΑ ΚΑΙ Η ΥΓΕΙΟΝΟΜΙΚΗ ΠΕΡΙΘΑΛΨΗ	6
1.2 ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΣΤΗΝ ΟΡΓΑΝΩΣΗ ΚΑΙ ΔΙΟΙΚΗΣΗ ΤΗΣ ΥΓΕΙΑΣ	7
1.3 Η ΠΟΙΟΤΗΤΑ ΤΩΝ ΥΠΗΡΕΣΙΩΝ ΥΓΕΙΑΣ ΚΑΙ Η ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ.....	9
1.4 Ο ΚΥΚΛΟΣ ΠΟΙΟΤΗΤΑΣ.....	9
1.5 ΟΙ ΔΕΙΚΤΕΣ ΑΞΙΟΛΟΓΗΣΗΣ ΤΩΝ ΝΟΣΟΚΟΜΕΙΩΝ.....	11
1.5.1 ΔΕΙΚΤΕΣ ΑΞΙΟΛΟΓΗΣΗΣ	11
1.5.2 ΔΕΙΚΤΕΣ ΠΟΙΟΤΗΤΑΣ.....	11
1.5.3 ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΟΙ ΔΕΙΚΤΕΣ	14
1.6 Η ΣΗΜΑΣΙΑ ΤΟΥ ΥΓΕΙΟΝΟΜΙΚΟΥ ΧΑΡΤΗ ΚΑΙ Ο ΣΤΟΧΟΣ ΤΟΥ.....	19
2° ΚΕΦΑΛΑΙΟ.....	20
ΤΕΧΝΙΚΕΣ ΠΡΟΒΛΕΨΕΩΝ.....	20
2.1 ΕΙΣΑΓΩΓΗ.....	20
2.2 ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ	22
2.3 ΕΞΟΥΡΥΞΗ ΔΕΔΟΜΕΝΩΝ (DATA MINING).....	23
2.4 ΣΥΣΤΗΜΑΤΑ ΥΠΟΣΤΗΡΙΞΗΣ ΚΛΙΝΙΚΩΝ ΑΠΟΦΑΣΕΩΝ	25
2.5 Ο ΡΟΛΟΣ ΤΗΣ ΠΡΟΒΛΕΨΗΣ ΣΤΟΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟ ΕΝΕΡΓΕΙΩΝ ΚΑΙ ΤΗΝ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ 26	
2.6 ΚΑΤΗΓΟΡΙΕΣ ΤΕΧΝΙΚΩΝ ΠΡΟΒΛΕΨΕΩΝ.....	29
2.7 ΠΟΣΟΤΙΚΑ ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΕΩΝ.....	30
2.7.1 ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	30
2.7.1.1 Απλή Γραμμική Παλινδρόμηση.....	31
2.7.1.2 Πολλαπλή γραμμική παλινδρόμηση	33
2.7.2 ΜΟΝΤΕΛΑ ΧΡΟΝΟΣΕΙΡΩΝ	34
2.7.2.1 Ποιοτικά Χαρακτηριστικά Χρονοσειρών	34
2.7.2.2 Διαχείριση κενών και μηδενικών τιμών	34
2.7.3 ΜΕΘΟΔΟΙ ΕΚΘΕΤΙΚΗΣ ΕΞΟΜΑΛΥΝΣΗΣ	35
2.7.3.1 Απλή εκθετική εξομάλυνση	36
2.7.3.2 Μοντέλο Γραμμικής Τάσης (Μέθοδος Holt).....	37
2.7.4 ΑΥΤΟΠΑΛΙΝΔΡΟΜΙΚΑ ΜΟΝΤΕΛΑ ΚΙΝΗΤΟΥ ΜΕΣΟΥ ΟΡΟΥ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ARIMA	37
2.7.4.1 Τιμές διαφόρισης (d)	39

2.8 ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	40
2.8.1 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ	40
2.8.2 ΠΡΟΤΥΠΟ ΕΝΟΣ ΤΕΧΝΗΤΟΥ ΝΕΥΡΩΝΑ	41
2.8.3 ΔΙΚΤΥΑ ΕΜΠΡΟΘΕΤΗΣ ΤΡΟΦΟΔΟΤΗΣΗΣ	43
2.8.4 ΑΛΓΟΡΙΘΜΟΙ ΕΚΠΑΙΔΕΥΣΗΣ ΤΕΧΝΙΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	45
2.9 ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ ΠΡΟΒΛΕΨΗΣ	46
2.10 ΣΦΑΛΜΑΤΑ	48
2.11 ΜΕΛΕΤΕΣ ΠΕΡΙΠΤΩΣΗΣ ΠΡΟΒΛΕΨΗΣ ΣΤΗΝ ΥΓΕΙΟΝΟΜΙΚΗ ΠΕΡΙΘΑΛΨΗ	48
2.11.1 Η ΔΗΜΙΟΥΡΓΙΑ ΜΟΝΤΕΛΟΥ ΠΡΟΒΛΕΨΗΣ ΑΠΟ ΤΗΝ GOOGLE	49
2.11.2 ΠΡΟΒΛΕΨΗ ΕΠΙΒΙΩΣΙΜΟΤΗΤΑΣ ΚΑΡΚΙΝΟΥ	50
<u>3° ΚΕΦΑΛΑΙΟ</u>	<u>52</u>
<u>ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΩΝ ΠΡΟΒΛΕΨΗΣ</u>	<u>52</u>
3.1 ΤΟ ΣΤΑΤΙΣΤΙΚΟ ΠΑΚΕΤΟ R	52
3.2 ΠΡΟΦΙΛ WORLD BANK	53
3.2.1 ΔΙΑΔΙΚΑΣΙΑ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΑΠΟ WORLD BANK	54
3.3 ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ	55
3.4 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	57
3.4.1 ΑΝΑΛΥΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ ΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	59
3.5 ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	63
3.5.1 ΑΝΑΛΥΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ ΠΟΛΛΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ.....	65
3.6 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	70
3.6.1 ΑΝΑΛΥΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	73
3.6.2 ΜΗΤΡΑ ΣΥΓΧΥΣΗΣ	75
3.7 ΧΡΟΝΟΣΕΙΡΕΣ (ΜΕΘΟΔΟΣ ARIMA)	76
3.7.1 ΑΝΑΛΥΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ ARIMA	78
3.6 ΠΡΟΦΙΛ ΕΛΛΗΝΙΚΩΝ ΝΟΣΟΚΟΜΕΙΩΝ	80
3.6.1 Γ.Π.Ν. ΡΙΟΥ	80
3.6.2 Γ.Ν. ΆΓΙΟΣ ΑΝΔΡΕΑΣ.....	81
3.7 ΔΗΜΙΟΥΡΓΙΑ ΕΦΑΡΜΟΓΗΣ ΠΑΡΟΥΣΙΑΣΗΣ ΠΕΡΙΓΡΑΦΙΚΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ ΠΑΚΕΤΟΥ SHINY ΤΗΣ R	82
<u>ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ</u>	<u>88</u>
<u>ΒΙΒΛΙΟΓΡΑΦΙΑ</u>	<u>90</u>
<u>ΠΑΡΑΡΤΗΜΑ</u>	<u>93</u>

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ ΕΙΚΟΝΩΝ

ΕΙΚΟΝΑ 1 ΤΡΟΠΟΣ ΛΕΙΤΟΥΡΓΕΙΑΣ ΕΝΟΣ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ	9
ΕΙΚΟΝΑ 2 ΚΥΚΛΟΣ ΠΟΙΟΤΗΤΑΣ	10
ΕΙΚΟΝΑ 3 ΒΗΜΑΤΑ ΑΝΑΛΥΤΙΚΗΣ ΠΡΟΒΛΕΨΗΣ	22
ΕΙΚΟΝΑ 4 ΚΑΤΗΓΟΡΙΕΣ ΑΝΑΛΥΤΙΚΗΣ ΠΡΟΒΛΕΨΗΣ	23
ΕΙΚΟΝΑ 5 ΒΗΜΑΤΑ ΔΙΑΔΙΚΑΣΙΑΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ	25
<i>ΕΙΚΟΝΑ 6 ΛΕΙΤΟΥΡΓΙΑ ΕΝΟΣ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ</i>	42
ΕΙΚΟΝΑ 7 ΛΕΙΤΟΥΡΓΙΑ ΔΙΚΤΥΟΥ ΕΜΠΡΟΘΕΤΗΣ ΤΡΟΦΟΔΟΤΗΣΗΣ	44
ΕΙΚΟΝΑ 8 ΣΥΓΚΡΙΣΗ ΠΡΑΓΜΑΤΙΚΗΣ ΚΑΙ ΠΡΟΒΛΕΠΟΜΕΝΗΣ ΤΙΜΗΣ ΑΝΑ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ ΧΩΡΕΣ (ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ)	57
ΕΙΚΟΝΑ 9 ΣΥΓΚΡΙΣΗ ΠΡΑΓΜΑΤΙΚΗΣ ΚΑΙ ΠΡΟΒΛΕΠΟΜΕΝΗΣ ΤΙΜΗΣ ΑΝΑ ΜΗ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ ΧΩΡΕΣ (ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ)	58
ΕΙΚΟΝΑ 10 ΣΥΓΚΡΙΣΗ ΠΡΑΓΜΑΤΙΚΗΣ ΚΑΙ ΠΡΟΒΛΕΠΟΜΕΝΗΣ ΤΙΜΗΣ ΑΝΑ ΕΥΡΩΠΑΪΚΗΣ ΕΝΩΣΗΣ ΧΩΡΑ (ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ)	64
ΕΙΚΟΝΑ 11 ΣΥΓΚΡΙΣΗ ΠΡΑΓΜΑΤΙΚΗΣ ΚΑΙ ΠΡΟΒΛΕΠΟΜΕΝΗΣ ΤΙΜΗΣ ΑΝΑ ΜΗ ΕΥΡΩΠΑΪΚΗΣ ΕΝΩΣΗΣ ΧΩΡΑ (ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ)	64
ΕΙΚΟΝΑ 12 ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΔΕΙΚΤΩΝ ΧΩΡΩΝ ΜΗ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ (ΈΞΟΔΟΣ ΑΡΙΘΜΗΤΙΚΗΣ ΤΙΜΗΣ)	71
ΕΙΚΟΝΑ 13 ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΔΕΙΚΤΩΝ ΧΩΡΩΝ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ (ΈΞΟΔΟΣ ΑΡΙΘΜΗΤΙΚΗΣ ΤΙΜΗΣ)	71
ΕΙΚΟΝΑ 14 ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΔΕΙΚΤΩΝ ΧΩΡΩΝ ΜΗ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ (ΈΞΟΔΟΣ ΚΑΤΗΓΟΡΙΚΩΝ ΤΙΜΩΝ)	72
ΕΙΚΟΝΑ 15 ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΔΕΙΚΤΩΝ ΧΩΡΩΝ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ (ΈΞΟΔΟΣ ΚΑΤΗΓΟΡΙΚΩΝ ΤΙΜΩΝ)	72
ΕΙΚΟΝΑ 16 ΠΡΟΒΛΕΨΗ ΓΙΑ ΤΑ ΕΤΗ 2014-2018 ΤΟΥ ΔΕΙΚΤΗ ΣΥΝΟΛΙΚΩΝ ΔΑΠΑΝΩΝ ΥΓΕΙΑΣ ΕΛΒΕΤΙΑΣ ΚΑΙ ΡΩΣΙΑΣ	76
ΕΙΚΟΝΑ 17 ΠΡΟΒΛΕΨΗ ΓΙΑ ΤΑ ΕΤΗ 2014-2018 ΤΟΥ ΔΕΙΚΤΗ ΣΥΝΟΛΙΚΩΝ ΔΑΠΑΝΩΝ ΥΓΕΙΑΣ ΓΑΛΛΙΑΣ ΚΑΙ ΣΟΥΗΔΙΑΣ	77
ΕΙΚΟΝΑ 18 Γ.Π.ΝΡΙΟ	80
ΕΙΚΟΝΑ 19 Γ.Ν. ΑΓΙΟΣ ΑΝΔΡΕΑΣ	81
ΕΙΚΟΝΑ 20 ΤΟ ΓΡΑΦΙΚΟ ΠΕΡΙΒΑΛΛΟΝ ΤΟΥ RSTUDIO	82
ΕΙΚΟΝΑ 21 ΑΡΧΙΚΟ ΠΑΡΑΘΥΡΟ ΕΦΑΡΜΟΓΗΣ	83
ΕΙΚΟΝΑ 22 ΠΑΝΕΛ ΕΠΕΞΕΡΓΑΣΙΑΣ ΠΑΡΑΜΕΤΡΩΝ	84
ΕΙΚΟΝΑ 23 ΠΑΡΟΥΣΙΑΣΗ ΠΕΡΙΛΙΠΤΙΚΩΝ ΣΤΟΙΧΕΙΩΝ	84
ΕΙΚΟΝΑ 24 ΓΡΑΦΗΜΑ ΓΡΑΜΜΗΣ (LINE CHART)	85
ΕΙΚΟΝΑ 25 ΓΡΑΦΗΜΑ ΠΙΤΑΣ (PIE CHART) 1	85
ΕΙΚΟΝΑ 26 ΓΡΑΦΗΜΑ ΠΙΤΑΣ (PIE CHART) 2	86
ΕΙΚΟΝΑ 27 ΡΑΒΔΟΓΡΑΜΜΑ ΒΑΣΗ ΑΠΟΛΥΤΩΝ ΤΙΜΩΝ ΚΑΙ ΧΡΟΝΟΛΟΓΙΑΣ (STACKED GROUPED BAR)	86
ΕΙΚΟΝΑ 28 ΡΑΒΔΟΓΡΑΜΜΑ ΒΑΣΗ ΠΟΣΟΣΤΙΑΙΩΝ ΤΙΜΩΝ ΚΑΙ ΧΡΟΝΟΛΟΓΙΑΣ (100% STACKED GROUPED BAR)	87

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ ΠΙΝΑΚΩΝ

ΠΙΝΑΚΑΣ 1 ΠΙΝΑΚΑΣ ΠΕΡΙΓΡΑΦΙΚΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΧΩΡΩΝ ΜΗ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ	56
ΠΙΝΑΚΑΣ 2 ΠΙΝΑΚΑΣ ΠΕΡΙΓΡΑΦΙΚΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΧΩΡΩΝ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ	56
ΠΙΝΑΚΑΣ 3 ΠΙΝΑΚΑ ΠΕΡΙΓΡΑΦΙΚΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ ΚΑΙ ΜΗ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ	59
ΠΙΝΑΚΑΣ 4 ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΟ 2013 ΧΩΡΩΝ ΕΕ ΚΑΙ ΜΗ ΕΕ	60
ΠΙΝΑΚΑΣ 5 ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΕΥ	60
ΠΙΝΑΚΑΣ 6 ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΝΟΕΥ	60
ΠΙΝΑΚΑΣ 7 ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ ΕΥ ΚΑΙ ΝΟΕΥ	62
ΠΙΝΑΚΑΣ 8 ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΟ 2013 ΧΩΡΩΝ ΕΕ ΚΑΙ ΜΗ ΕΕ	65
ΠΙΝΑΚΑΣ 9 ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΝΟΕΥ	66
ΠΙΝΑΚΑΣ 10 ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΩΝ ΕΥ	67
ΠΙΝΑΚΑΣ 11 ΠΙΝΑΚΑ ΠΕΡΙΓΡΑΦΙΚΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ ΚΑΙ ΜΗ ΕΥΡΩΠΑΪΚΗΣ ΈΝΩΣΗΣ	68
ΠΙΝΑΚΑΣ 12 ΠΙΝΑΚΑΣ ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ ΕΥ ΚΑΙ ΝΟΕΥ	68
ΠΙΝΑΚΑΣ 13 ΣΤΑΤΙΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΟ 2013 ΧΩΡΩΝ ΕΕ ΚΑΙ ΜΗ ΕΕ	73
ΠΙΝΑΚΑΣ 14 ΠΙΝΑΚΑΣ ΠΕΡΙΓΡΑΦΙΚΩΝ ΣΤΑΤΙΣΤΙΚΩΝ ΕΥΡΩΠΗΣ ΚΑΙ ΜΗ ΕΥΡΩΠΗΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	74
ΠΙΝΑΚΑΣ 15 ΥΠΟΛΕΙΜΜΑΤΙΚΟΣ ΠΙΝΑΚΑΣ ΓΑΛΛΙΑΣ ΚΑΙ ΣΟΥΗΔΙΑΣ	78
ΠΙΝΑΚΑΣ 16 ΥΠΟΛΕΙΜΜΑΤΙΚΟΣ ΠΙΝΑΚΑΣ ΕΛΒΕΤΙΑΣ ΚΑΙ ΡΩΣΙΑΣ	78
ΠΙΝΑΚΑΣ 17 ΠΙΝΑΚΑΣ ΣΦΑΛΜΑΤΩΝ ΧΡΟΝΟΣΕΙΡΑΣ ΓΑΛΛΙΑΣ	79
ΠΙΝΑΚΑΣ 18 ΠΙΝΑΚΑΣ ΣΦΑΛΜΑΤΩΝ ΧΡΟΝΟΣΕΙΡΑΣ ΣΟΥΗΔΙΑΣ	79
ΠΙΝΑΚΑΣ 19 ΠΙΝΑΚΑΣ ΣΦΑΛΜΑΤΩΝ ΧΡΟΝΟΣΕΙΡΑΣ ΕΛΒΕΤΙΑΣ	79
ΠΙΝΑΚΑΣ 20 ΠΙΝΑΚΑΣ ΣΦΑΛΜΑΤΩΝ ΧΡΟΝΟΣΕΙΡΑΣ ΡΩΣΙΑΣ	79

1^ο Κεφάλαιο

Υγειονομική Περίθαλψη

1.1 Η Υγεία και η Υγειονομική Περίθαλψη

Ο ορισμός της Υγείας περιλαμβάνεται στον πρόλογο του καταστατικού του Παγκόσμιου Οργανισμού Υγείας από το 1976 (WHO, 1976) :*"Υγεία είναι μια κατάσταση πλήρους σωματικής, νοητικής και κοινωνικής ευεξίας και όχι απλώς η απουσία νόσου ή αναπηρίας."* (*"Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity."*).

Το δίπολο «υγεία/αρρώστια» είναι ένα από τα θέματα που απασχολούν όλο και πιο συστηματικά τις σύγχρονες κοινωνίες. Οι επιστημονικές περιοχές, όπως η ιατρική, η νοσηλευτική, η βιολογία εμπλέκονται στη διερεύνηση και κατανόηση των παραπάνω όρων. Όπως η υγεία είναι κοινωνικό αγαθό έτσι και η αντιμετώπιση της αρρώστιας είναι καθήκον κοινωνικής πολιτικής, μέσα από μορφές υπηρεσιών υγείας. Όπως σε κάθε προηγμένη χώρα έτσι και στην Ελλάδα υπάρχει ένα σύστημα υγειονομικής περίθαλψης που περιλαμβάνει διάφορες υπηρεσίες, σε σταθμούς, κέντρα υγείας, ιδρύματα πρόληψης, θεραπείας ασθένειας και προστασία της δημόσιας υγείας.

Η υγειονομική περίθαλψη ουσιαστικά αποτελεί το σύνολο των διαδικασιών που απαιτούνται έτσι ώστε το άτομο να περάσει από την φάση της ασθένειας στη φάση της υγείας. Η υγειονομική περίθαλψη χωρίζεται σε πρωτοβάθμια, δευτεροβάθμια και τριτοβάθμια. Η πρωτοβάθμια περιλαμβάνει τις ιατρικές, νοσηλευτικές, οδοντιατρικές πράξεις και φροντίδες που έχουν ως σκοπό την πρόληψη υγείας που δεν απαιτούν νοσηλεία στο νοσοκομείο. Η δευτεροβάθμια είναι η περίθαλψη που

παρέχεται στο νοσοκομείο και τέλος η τριτοβάθμια απευθύνεται σε μεγάλες πληθυσμιακές ομάδες όπου απαιτεί εξειδικευμένες γνώσεις, τεχνολογία, ικανότητες, υψηλό κόστος δημιουργίας και λειτουργίας, όπως η Μονάδες Τεχνικού Νεφρού, Μονάδα Εντατικής Θεραπείας, Μονάδα Άσηπτης Νοσηλείας (για αιματολογικές παθήσεις). (Γκουργκούλη, και συν., 2007)

1.2 Πληροφοριακά Συστήματα στην Οργάνωση και Διοίκηση της Υγείας

Σύμφωνα με την βιβλιογραφία τα συστήματα κατηγοριοποιούνται με τρεις τρόπους. Ο πρώτος τρόπος κατηγοριοποίησής τους είναι τα φυσικά συστήματα (π.χ. άνθρωπος) και τα τεχνητά συστήματα, τα οποία αφορούν συστήματα επικοινωνίας (π.χ. τηλέφωνο, φαξ κ.α.).

Ο δεύτερος τρόπος αφορά την κατηγοριοποίησή τους σε ανοιχτά συστήματα όπου έχουν να κάνουν με εξωτερικές παρεμβάσεις, δηλαδή νέα δεδομένα και απαιτήσεις και κλειστά συστήματα που δεν μπορούν να δεχτούν εξωτερικές παρεμβάσεις ενώ υπάρχει συγκεκριμένη τυπική διαδικασία για την επαφή με το περιβάλλον.

Ο τελευταίος τρόπος κατηγοριοποίησης είναι με ένα αιτιοκρατικό σύστημα που ισχύει όταν μια είσοδος (input) καταλήγει σε μια έξοδο (output), δηλαδή από ένα συγκεκριμένο input σε ένα συγκεκριμένο output ή με ένα στοχαστικό σύστημα ένα συγκεκριμένο input πιθανώς θα έχουμε διαφορετικό output. Στην κατηγορία αυτή, για παράδειγμα, έχω την ασθένεια του ανθρώπινου οργανισμού που δεν μπορεί να προβλεφθεί. (Μυλωνά, 2013)

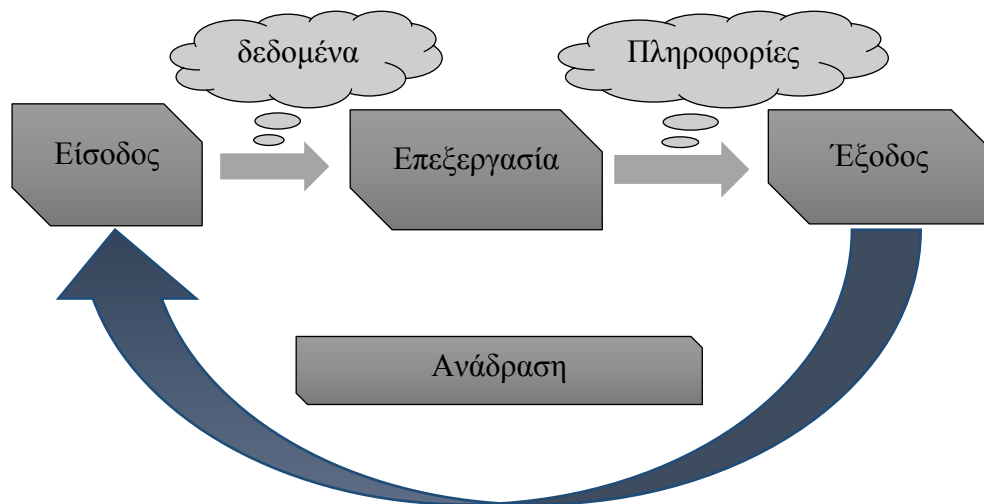
Σε πολλές περιπτώσεις η λήψη ορθολογικών αποφάσεων αποτελεί μια ανθρώπινη προσπάθεια και ιδιαίτερα όταν εμπλέκονται και άλλα πρόσωπα απαιτεί ικανότητες οργάνωσης και διοίκησης. Επί του παρόντος, η οργάνωση και η διοίκηση (management) ορίζεται «σαν μια δραστηριότητα κατά την οποία εργάζεται κανείς μαζί και μέσω άλλων ανθρώπων, για την ολοκλήρωση τόσο των στόχων της επιχείρησης όσο και των στόχων των ατόμων που εργάζονται σε αυτήν». (Πετρώφ, και συν., 2002). Βασικός στόχος στα επαγγέλματα της υγείας είναι η υψηλή ποιότητα υπηρεσιών υγείας αλλά και της αποδοτικότητας των υπηρεσιών υγείας.

Ο χώρος της υγείας μπορεί να επωφεληθεί πολύ από την εφαρμογή των πληροφοριακών συστημάτων. Στις μέρες μας η τεχνολογία της πληροφορίας έχει αποδειχθεί ως διευκόλυνση για τις επιχειρησιακές δραστηριότητες, χάρη στις δυνατότητες των πληροφοριακών συστημάτων, κάποιες από αυτές μπορεί να είναι:

- Εκτελούν ταχύτατα μεγάλο όγκο υπολογισμούς.
- Αποθηκεύουν τεράστιες ποσότητες πληροφοριών.
- Αυτοματοποιούν επιχειρησιακές διαδικασίες.
- Επιταχύνουν την επεξεργασία και καταχώρηση δεδομένων.
- Παρέχουν γρήγορη, ορθή και φθηνή επικοινωνία.
- Αυξάνουν την αποδοτικότητα και την αποτελεσματικότητα των ατόμων.

Η σπουδαιότητα των ηλεκτρονικών υπολογιστών και της μηχανογράφησης μπορεί κανείς να την αντιληφθεί αν λάβει υπόψη τον όγκο πληροφοριών και δεδομένων που αντιστοιχούν σε έναν μόνο ασθενή και την αδυναμία συλλογής, καταμέτρησης και επεξεργασίας τους από μεμονωμένα άτομα. (Φαράτζιαν, 2007)

Όσον αφορά την λειτουργία των πληροφοριακών συστημάτων περιβάλλεται αποτελείται από την είσοδο, την επεξεργασία, την έξοδο και την ανάδραση. Η είσοδος (input) αφορά την συλλογή των δεδομένων από το εξωτερικό περιβάλλον. Η επεξεργασία (processes) αφορά τον μετασχηματισμό των δεδομένων από ακατέργαστα σε μορφή που έχουν νόημα δηλαδή σε έξοδο. Η έξοδος (output) μεταφέρει την πληροφορία για να χρησιμοποιηθεί για την εξαγωγή συμπερασμάτων. Τέλος, η ανάδραση (feedback) είναι η πληροφορία που αφορά την απόδοση του συστήματος αλλά και την αξιολόγηση ώστε να προχωρήσουν σε βελτιώσεις από το στάδιο εισόδου. Ακολουθώντας μια τέτοια διαδικασία, όπου παράγονται όσο τον δυνατόν πιο ακριβείς προβλέψεις και αξιοποιώντας στο έπακρο όλη την διαθέσιμη πληροφορία, συμβάλουν στην σωστή λήψη αποφάσεων.



Εικόνα 1 Τρόπος Λειτουργίας Ενός Πληροφοριακών Συστημάτων

1.3 Η Ποιότητα των Υπηρεσιών Υγείας και η Αξιολόγηση της

Η αξιολόγηση και η βελτίωση της ποιότητας των υπηρεσιών απασχολούσαν τους ανθρώπους από την αρχαιότητα, με χαρακτηριστικό παράδειγμα ηθικής και πρακτικής τον Όρκο του Ιπποκράτη (5ος αιώνας π.Χ.). Η έννοια της ποιότητας στις υπηρεσίες υγείας περιλαμβάνει την αποτελεσματικότητα, την αποδοτικότητα, την ασφάλεια, την προσβασιμότητα και την ανάπτυξη των υπηρεσιών με επίκεντρο την φροντίδα του ασθενή. Η ποιότητα μπορεί να μετρηθεί. Αποτελείται από τρία χαρακτηριστικά τη δομή, τις διαδικασίες και τα αποτελέσματα. Σε κάθε κατηγορία υπάρχουν ποσοτικά χαρακτηριστικά που μπορούν να μετρηθούν όπως η εκπαίδευση, τα χρόνια εμπειρίας, η ακρίβεια των μηχανημάτων, ο βαθμός ικανοποίησης των ασθενών κ.α. αρκεί να βρεθούν τα κύρια σημεία αξιολόγησης. (Παπακωστίδη, et al., 2012)

1.4 Ο Κύκλος Ποιότητας

Ο κύκλος ποιότητας αποτελείται από μια ομάδα κατάλληλα εκπαιδευμένων με σκοπό να λύσουν συγκεκριμένα προβλήματα στον τομέα της εργασίας τους, με όσον το δυνατό μικρότερο κόστος για τον οργανισμό. Ο μηχανισμός που

περιγράφεται σχηματικά αρχίζει με την βελτίωση των ευκαιριών κατά σειρά προτεραιότητας.



Εικόνα 2 Κύκλος ποιότητας

Τα πρώτα τρία βήματα του κύκλου (ο σχεδιασμός, ο καθορισμός, τα πρότυπα ποιότητας, η ενημέρωση και η εκπαίδευση του προσωπικού) εντάσσονται στο πλαίσιο της διασφάλισης ποιότητας. Η έννοια της διασφάλισης ποιότητας αποτελείται από τον σχεδιασμό στόχων ποιότητας, τον καθορισμό προτύπων ποιότητας, την χρήση δεικτών και την συμμετοχή του προσωπικού. Έτσι είναι απαραίτητη για την βελτίωση της ποιότητας του στόχου. Ακολουθεί ο ποιοτικός έλεγχος, ενώ τα επόμενα έξι βήματα εντάσσονται στη διαδικασία βελτίωσης της ποιότητας, για την εξασφάλιση χαμηλού εύρους διακύμανσης και επομένως ελεγχόμενων αποτελεσμάτων. Αφού μελετηθεί η περίπτωση και επιλεγεί μια λύση τότε εφαρμόζεται, εφ' όσον παρέχει τα επιθυμητά αποτελέσματα, ενώ συνεχίζεται ο μηχανισμός βελτίωσης. Ο κύκλος ποιότητας λειτουργεί μέχρι την οριστική επίλυση των προβλημάτων. (Παπακωστίδη, et al., 2012)

1.5 Οι Δείκτες Αξιολόγησης των Νοσοκομείων

1.5.1 Δείκτες Αξιολόγησης

Τα βασικά κριτήρια για την επιλογή ενός δείκτη αξιολόγησης της λειτουργίας ενός υγειονομικού οργανισμού είναι τα παρακάτω (Παπακωστίδη, και συν., 2012):

- **Χρησιμότητα** (usefulness): Βοηθάει στην εξαγωγή των συμπερασμάτων και να συνδέεται με την δυνατότητα βελτίωσης των αποτελεσμάτων.
- **Εγκυρότητα** (validity): Φανερώνει κατά πόσο είναι αποδεκτή η μέτρηση του δείκτη, δυσκολία στην εγκυρότητα είναι ότι απαιτεί διαδοχικές μετρήσεις.
- **Αξιοπιστία** (reliability): Περιγράφει τον βαθμό συνέπειας μεταξύ των αποτελεσμάτων. Η αξιοπιστία μιας μέτρησης ελέγχεται με το συντελεστή συσχέτισης με επαναλαμβανόμενες μετρήσεις.
- **Δυνατότητα σύγκρισης** (comparability): Εξασφαλίζεται με την ύπαρξη πρότυπων δεικτών και την διαθεσιμότητα στοιχείων σε άλλες χώρες.
- **Ανταποκρισιμότητα** (responsiveness): Υποδηλώνει τον βαθμό ευαισθησίας ενός δείκτη να βρίσκει μεταβολές.
- **Ειδικότητα** (specificity): Πρόκειται για την ικανότητα εντόπισης διαφορών στις εξεταζόμενες ομάδες ασθενών.
- **Ευαισθησία** (sensitivity): Εξετάζει την ακρίβεια του δείκτη και καταγράφει τις αλλαγές που πραγματοποιούνται από την επαναλαμβανόμενη χρήση του.

1.5.2 Δείκτες Ποιότητας

Την εικόνα για τη λειτουργία ενός τμήματος στο νοσοκομείο την δίνουν οι δείκτες ποιότητας. Είναι χρήσιμοι για τη σύγκριση τόσο μεταξύ τμημάτων όσο και σε σχέση με καινούρια πρότυπα λειτουργίας. Με τους εν λόγω δείκτες εξάγεται το αποτέλεσμα των ενεργειών μιας χρονικής περιόδου και μπορούν να φανούν χρήσιμοι στην προσπάθεια για συνεχόμενη βελτίωση μέσα από την ανάδειξη των πλεονεκτημάτων και των αδυναμιών ενός τμήματος (Μωραΐτη, 2009).

Οι δείκτες που προσδιορίζουν τον τρόπο λειτουργίας και την αποδοτικότητα όλων των τύπων και μορφών νοσοκομείων είναι (Ζάβρας):

- **Κλίνες**

προσδιορίζουν το μέγεθος και τις δυνατότητες του νοσοκομείου στην παροχή υπηρεσιών

- **Αριθμός των Νοσηλευθέντων Ασθενών**

προσδιορίζεται από τις εξόδους των ασθενών ή τα «επεισόδια νοσηλείας». Το μεθοδολογικό πρόβλημα που ανακύπτει για να εξομοιωθεί με την ιατρική περίπτωση είναι η συνολική νοσηλεία του ίδιου ασθενούς (συμπεριλαμβανομένης της μεταφοράς σε άλλο τμήμα ή/και νοσοκομείο)

- **Ημέρες Νοσηλείας**

η αναφορά γίνεται στις ημέρες νοσηλείας μιας ιατρικής περίπτωσης ή συνολικά ενός νοσοκομείου

- **Μέση Διάρκεια Νοσηλείας**

υπολογίζεται με τις ημέρες νοσηλείας δια τον αριθμό των νοσηλευθέντων. Το αποτέλεσμα της αντανάκλα στην οργάνωση και στην ποιότητα των υπηρεσιών

- **Πληρότητα ή Κάλυψη**

μετριέται την δεδομένη στιγμή ή ημέρα και υπολογίζεται από το γινόμενο των νοσηλευθέντων επί εκατό (100) διά αριθμό κλινών ή μετριέται συνολικά ετησίως και υπολογίζεται με τις ημέρες νοσηλείας δια το γινόμενο 365 επί αριθμό κλινών. Αποτελεί δείκτη σύγκρισης της απόδοσης των νοσοκομείων.

- **Ρυθμός εισροής ασθενών ανά κλίνη**

υπολογίζεται από του γινομένου 365 επί ποσοστό πληρότητας διά το γινόμενο 100 επί Μέση Διάρκεια Νοσηλείας και μετρά τον ρυθμό χρησιμοποίησης μιας κλίνης και εκφράζεται σε αριθμό ασθενών ανά μονάδα χρόνου.

- **Διάστημα Εναλλαγής**

Εκφράζει τον ρυθμό εναλλαγής των ασθενών ή τον μέσο αριθμό ημερών που μένει κενή η κλίνη όπου υπολογίζεται από (365/Ρυθμός Εισροής Ασθενών)-Μέση Διάρκεια Νοσηλείας.

- **Προσδόκιμο ζωής κατά τη γέννηση (Συνολικά χρόνια) - *Life expectancy at birth, total (years)***

Δείχνει τον αριθμό των ετών που ένα νεογέννητο βρέφος θα ζήσει αν τα

πρότυπα θνησιμότητας κατά τη στιγμή της γέννησής του παραμένουν τα ίδια σε όλη τη ζωή του.

- **Αριθμός θανάτων βρεφών - *Number of infant deaths***

Αριθμός των βρεφών που πεθαίνουν πριν συμπληρώσουν ένα έτος της ηλικίας.

- **Ποσοστό βρεφικής θνησιμότητας (ανά 1000 γεννήσεις ζωντανών) - *Mortality rate, infant (per 1,000 live births)***

Είναι ο αριθμός των βρεφών που πεθαίνουν πριν συμπληρώσουν ένα έτος της ηλικίας τους, ανά 1.000 γεννήσεις σε ένα δεδομένο έτος.

- **Ποσοστό Νεογνικής θνησιμότητας (ανά 1000 γεννήσεις ζωντανών) - *Mortality rate, neonatal (per 1,000 live births)***

Είναι ο αριθμός των νεογνών που πεθαίνουν πριν συμπληρώσουν 28 ημέρες, ανά 1.000 γεννήσεις σε ένα δεδομένο έτος.

- **Αριθμός νεογνικών θανάτων - *Number of neonatal deaths***

Αριθμός θανάτων νεογνών πριν συμπληρώσουν 28 ημέρες.

- **Συνολικό ποσοστό γονιμότητας (γεννήσεις ανά γυναίκα) - *Fertility rate, total (births per woman)***

Αντιπροσωπεύει τον αριθμό των παιδιών που θα γεννηθούν από μια γυναίκα, αν αυτή επρόκειτο να ζήσει μέχρι το τέλος της αναπαραγωγικής ηλικίας της.

- **Ακαθάριστο ποσοστό γεννητικότητας (ανά 1000 άτομα) - *Birth rate, crude (per 1,000 people)***

Δείχνει τον αριθμό των γεννήσεων ζώντων που συμβαίνουν κατά τη διάρκεια του έτους, ανά 1.000 πληθυσμού και υπολογίζεται στα μέσα της χρονιάς.

- **Ακαθάριστο ποσοστό θνησιμότητας (ανά 1000 άτομα) - *Death rate, crude (per 1,000 people)***

Υποδεικνύει τον αριθμό των θανάτων που συμβαίνουν κατά τη διάρκεια του έτους, ανά 1.000 πληθυσμού υπολογίζεται στα μέσα της χρονιάς. Αφαιρώντας το ακαθάριστο ποσοστό θανάτου από το ακαθάριστο ποσοστό γεννητικότητας

παρέχει το ποσοστό φυσικής αύξησης, η οποία είναι ίση με το ρυθμό μεταβολής του πληθυσμού ελλείψει της μετανάστευσης.

- **Νοσηλευτικό προσωπικό και μαίες (ανά 1000 άτομα) - *Nurses and midwives (per 1,000 people)***

Η κατηγορία αυτή περιλαμβάνει επαγγελματίες νοσοκόμους, επαγγελματίες μαίες, βοηθητικό νοσηλευτικό προσωπικό, βοηθητικές μαίες, εγγεγραμμένους νοσηλευτές και μαίες και άλλο συναφή προσωπικό, όπως οδοντιατρικοί νοσηλευτές και πρωτοβάθμιοι νοσηλευτές.

- **Ιατρικό προσωπικό (ανά 1000 άτομα) - *Physicians (per 1,000 people)***

Περιλαμβάνουν όλους τους γενικό και ειδικό ιατρικό προσωπικό.

1.5.3 Χρηματοοικονομικοί Δείκτες

Στις δημόσιες μονάδες υγείας δεν υπάρχει το κίνητρο του κέρδους και για αυτό τον λόγο αναλύσεις χρηματοοικονομικής φύσεως δεν εφαρμόζονται πάνω σε αυτές. Η μόνη χρήση που μπορούν να έχουν είναι για την εξαγωγή συμπερασμάτων όπως: (Πονοζάχος, 2013)

- αν η χρηματοδότηση κατευθύνθηκε σε υποδομές που είναι απαραίτητες για την πραγματοποίηση του έργου για το οποίο υπάρχουν
- αν είναι οικονομικά βιώσιμοι
- αν έχουν την δυνατότητα να ανταποκριθούν στην πληρωμή των υποχρεώσεων τους
- αν διαθέτουν την ρευστότητα που είναι αναγκαία για την επίτευξη του κοινωνικούς τους έργου
- αν έχουν την δυνατότητα της αυτοχρηματοδότησης νέων δραστηριοτήτων ή και επέκταση των ήδη υλοποιούμενων

Οι αριθμοδείκτες που οδηγούν σε χρήσιμα συμπεράσματα για τους οι Μη Κερδοσκοπικοί Οργανισμοί είναι της: Ρευστότητας, Χρηματοδότησης και μέτρησης κεφαλαιακής δομής και Απόδοσης κεφαλαίων. Όπως περιγράφεται από τον Μη Κερδοσκοπικό Οργανισμό (ΜΚΟ Κατά του Καρκίνου Be Strong) οι πηγές των κερδών τους είναι:

- Εισφορές των εταίρων της
- Δωρεές τρίτων
- Χορηγίες
- Κληρονομίες
- Κληροδοσίες
- Έσοδα από διάφορες δημόσιες ή μη εκδηλώσεις και δραστηριότητες
- Έσοδα από την διάθεση εντύπων, περιοδικών, βιβλίων σε έντυπη ή ηλεκτρονική μορφή
- Νόμιμες επιχορηγήσεις από τους διάφορους οργανισμούς και ιδρύματα ημεδαπής και αλλοδαπής
- Κρατικές επιχορηγήσεις από τους οργανισμούς της τοπικής αυτοδιοίκησης επιχορηγήσεις της ευρωπαϊκής ένωσης, διεθνών οργανισμών και διεθνών ενώσεων
- Έσοδα από την μεταπώληση διαφόρων προϊόντων
- Έσοδα από μεταπώληση προϊόντων που η ίδια δημιουργεί και παράγει, έσοδα από τις υπηρεσίες και τα προϊόντα που είναι διαθέσιμα από τον διαδικτυακό τόπο της εταιρείας, που παρέχονται μέσω ηλεκτρονικού υπολογιστή
- Εισοδήματα από την αξιοποίηση της περιουσίας της εταιρείας και οποιοδήποτε άλλο έσοδο από νόμιμη αιτία

Αριθμοδείκτες Ρευστότητας

Η ρευστότητα έχει τρία βασικά συστατικά τα οποία είναι: Ο χρόνος, το ποσό και το κόστος. Ο χρόνος είναι αυτός που απαιτείται για την μετατροπή ενός περιουσιακού στοιχείου σε μετρητά. Το ποσό δείχνει αν ο φορέας μπορεί να καλύψει τις βραχυπρόθεσμες υποχρεώσεις του και κόστος είναι το τίμημα να μετατραπεί ένα περιουσιακό στοιχείο σε μετρητά (Γ. Λαζαρίδης - Δ. Παπαδόπουλος)

Οι αριθμοδείκτες ρευστότητας που χρησιμοποιούνται συχνότερα είναι :

- **Γενικής Ρευστότητας**

Ο δείκτης αυτός μετράει το περίσσειμα των ρευστών κεφαλαίων πάνω από τις τρέχουσες υποχρεώσεις. Το πλεόνασμα των κυκλοφοριακών περιουσιακών στοιχείων πάνω από τις τρέχουσες υποχρεώσεις, παρέχει ένα περιθώριο ασφαλείας.

$$\frac{\text{Κυκλοφορούν ενεργητικό Μ. Ο}}{\text{Βραχυπρόθεσμες υποχρεώσεις Μ. Ο}}$$

- **Άμεσης Ρευστότητας**

Ο δείκτης αυτός δείχνει την δυνατότητα της οικονομικής οντότητας να εξυπηρετήσει τις βραχυπρόθεσμες υποχρεώσεις της ρευστοποιώντας τα αμέσως ρευστοποιήσιμα στοιχεία. Αφαιρούνται τα αποθέματα διότι αφ' ενός παρουσιάζουν μεταξύ των κυκλοφορούντων περιουσιακών στοιχείων την μικρότερη ρευστότητα, αφ' ετέρου από την ρευστοποίηση των αποθεμάτων είναι πολύ πιθανό να προκύψει ζημιά.

$$\frac{(\text{Κυκλοφορούν ενεργητικό} - \text{Απόθεμα})\text{Μ. Ο.}}{\text{Βραχυπρόθεσμες Υποχρώσεις Μ. Ο.}}$$

- **Δείκτης Ανακύκλωσης Απαιτήσεων**

Ο δείκτης αυτός δείχνει πόσες φορές μέσα στην χρήση η οικονομική οντότητα εισπράττει τις απαιτήσεις της. Η τιμή του δείκτη για να είναι ικανοποιητικός πρέπει να είναι αρκετά μεγαλύτερος της μονάδας.

$$\frac{\text{Ετήσιες Πωλήσεις}}{\text{Απαιτήσεις Μ. Ο}}$$

- **Μέσος Χρόνος Είσπραξης Απαιτήσεων**

Ο δείκτης αυτός δείχνει σε πόσες ημέρες η οικονομική οντότητα εισπράξει τις απαιτήσεις της, από την στιγμή που πραγματοποιηθήκαν οι πωλήσεις. Όσο μικρότερος είναι ο χρόνος αυτός, τόσο μεγαλύτερη είναι η ταχύτητα είσπραξης, άρα μικρότερος ο χρόνος δέσμευσης των κεφαλαίων.

$$\frac{\text{Απαιτήσεις Μ.Ο.} * 365 \text{ ημέρες}}{\text{Ετήσιες Πωλήσεις}}$$

- **Χρόνος Εξόφλησης Βραχυπρόθεσμων Υποχρεώσεων**

Ο δείκτης αυτός μας δείχνει σε πόσες ημέρες η οικονομική οντότητα πληρώνει τους βραχυπρόθεσμους πιστωτές της. Σε συνδυασμό με τις ημέρες είσπραξης των απαιτήσεών της από τους πελάτες της γίνεται φανερή η διαμορφούμενη ρευστότητα της επιχείρησης αλλά και η πιστοληπτική της ικανότητα. Εάν ο χρόνος πληρωμής των βραχυπροθέσμων υποχρεώσεων είναι μεγαλύτερος από τον χρόνο είσπραξης των απαιτήσεων, τότε η οικονομική οντότητα έχει δημιουργήσει μια πηγή χρηματοδότησης που αποτελεί μια αποτελεσματική χρησιμοποίηση ξένων κεφαλαίων.

$$\frac{\text{Μ.Ο. Βραχυπρόθεσμων υποχρεώσεων} * 365 \text{ ημέρες}}{\text{Κόστος πωληθέντων} - \text{Αποσβέσεις}}$$

Αριθμοδείκτες Χρηματοδότησης και μέτρησης κεφαλαιακής δομής

- **Δείκτης Παγιοποίησης Περιουσίας.**

Ο δείκτης αυτός μας δείχνει αν η οικονομική οντότητα είναι εντάσεως παγίων περιουσιακών στοιχείων ή κυκλοφοριακών περιουσιακών στοιχείων.

$$\frac{\text{Σύνολο πάγιων περιουσιακών στοιχείων}}{\text{Σύνολο Ενεργητικού}}$$

Αριθμοδείκτες Απόδοσης Κεφαλαίων

- **Δείκτης περιθωρίου Κέρδους**

Στους δημόσιους οργανισμούς υγείας αναπτύσσεται η άποψη ότι πρέπει να είναι ελάχιστα άνω του μηδενός γιατί έτσι επιτυγχάνεται το καλύτερο οικονομικό αποτέλεσμα για το σύνολο της οικονομίας γιατί δεν επιβαρύνονται τα ασφαλιστικά ταμεία με επιπλέον κόστη.

$$\frac{\text{Πωλήσεις} - \text{Κόστος Πωληθέντων}}{\text{Πωλήσεις}}$$

- **Δείκτης Αποδοτικότητας Ενεργητικού**

Ο δείκτης αυτός αποτυπώνει την απόδοση του συνόλου των περιουσιακών στοιχείων. Δείχνει αν τελικά αξίζει να προσελκύσει κεφάλαια για επένδυση ή όχι.

$$\frac{\text{Καθαρά Κέρδη Χρήσης}}{\text{Σύνολο Ενεργητικού}}$$

- **Συνολικές δαπάνες υγείας (% του ΑΕΠ) - *Health expenditure, total (% of GDP)***

Οι συνολικές δαπάνες είναι το άθροισμα των δημοσίων και ιδιωτικών δαπανών. Καλύπτει την παροχή των υπηρεσιών υγείας (προληπτικές και θεραπευτικές), τις δραστηριότητες οικογενειακού προγραμματισμού, τις δραστηριότητες της διατροφής, και τις επείγουσες βοήθειες που ορίζονται για την υγεία.

- **Εξωτερικοί πόροι χρηματοδότησης για την υγεία (% των συνολικών δαπανών) - *External resources for health (% of total expenditure on health)***

Εξωτερικοί πόροι για την υγεία είναι κεφάλαια ή υπηρεσίες σε είδος που παρέχονται από φορείς που δεν μετέχουν στην εν λόγω χώρα. Οι πόροι μπορεί να προέρχονται από διεθνείς οργανισμούς, άλλες χώρες μέσω διμερών συμφωνιών, ή

ξένων μη κυβερνητικών οργανώσεων. Οι πόροι αυτοί αποτελούν μέρος των συνολικών δαπανών για την υγεία.

1.6 Η Σημασία του Υγειονομικού Χάρτη και ο Στόχος του

Σύμφωνα με το υπουργείο υγείας ο Υγειονομικός Χάρτης, αποτελεί ένα μηχανισμό διαρκούς συλλογής και επεξεργασίας στοιχείων και δεικτών υγείας σε διαχρονικό επίπεδο. (Υγειομονικός Χάρτης)

Πολλοί φορείς υγείας συλλέγουν στοιχεία σχετικά με τη δραστηριότητά τους, ωστόσο, στις περισσότερες περιπτώσεις αυτά τα στοιχεία δεν είναι επαρκή. Ο σχεδιασμός και η υλοποίηση ενός πληροφοριακού συστήματος για την έγκυρη πληροφόρηση θα έχει ως στόχο την αποτελεσματικότερη εξυπηρέτηση των πολιτών.

Ο στόχος του υγειονομικού συστήματος

1. Η **συλλογή δεδομένων** από πηγές δημοσίου και ιδιωτικού τομέα.
2. Η **επεξεργασία και ανάλυση των δεδομένων** και η **παρουσίαση τους υπό μορφή δεικτών**.
3. Η **αξιολόγηση** των παρεχόμενων υπηρεσιών και η πρόβλεψη των μελλοντικών αναγκών.
4. Η **ανάπτυξη** και σχεδιασμού νέων υπηρεσιών.

2^ο

Κεφάλαιο

Τεχνικές Προβλέψεων

2.1 Εισαγωγή

Σήμερα, το μέγεθος των δεδομένων των υπηρεσιών υγειονομικής περίθαλψης είναι απίστευτα μεγάλο. Σύμφωνα με την IBM (Schick, 2012), οι περισσότεροι οργανισμοί υγείας έχουν terabyte δεδομένων στα συστήματά τους. Περισσότερο από το 80% των δεδομένων των υπηρεσιών περίθαλψης δεν είναι αρχειοθετημένο, για παράδειγμα ιατρικές σημειώσεις, διάφορες φόρμες, αντίγραφα εξιτηρίων και έγγραφα ασθενών συναθροίζονται για να δημιουργήσουν ένα χάος από δεδομένα. Ως αποτέλεσμα, εκατομμύρια έγγραφα υγείας είναι μη διαθέσιμα και ένα ποσοστό από αυτά δεν βρίσκονται σε ηλεκτρονική μορφή. Αυτά τα έγγραφα περιέχουν πληροφορίες μεγάλης σημασίας που όμως δεν υπάρχει εύκολος τρόπος για να γίνει ανάλυσή τους. Οι τεχνικές προβλέψεων καταρρίπτουν ολοκληρωτικά αυτή την εικόνα.

Μόλις τα τελευταία χρόνια οι υπολογιστές χρησιμοποιούνται για την εξυπηρέτηση των ανθρώπων. Όσο η τεχνολογία αναπτύσσεται και κάνει τα δεδομένα πιο προσβάσιμα, πιο αξιόπιστα, πιο εύκολα στην χρήση, ενοποιητικές ενέργειες όπως η παγκοσμιοποίηση της ιατρικής συνδέουν τους τομείς της πρωτογενής και κλινικής έρευνας μαζί με την παρακλινία περίθαλψη (Pearson, 2012). Οι τεχνικές προβλέψεων βοηθούν στο να γίνουν αυτές οι συνδέσεις με τους ακόλουθους τρόπους:

- Διαχείριση ασθενειών

Μπορεί να χρησιμοποιηθεί για την πρόβλεψη κινδύνου του κόστους για κάθε

μέλος σε μία υγειονομική μονάδα απλά κάνοντας την ερώτηση “Πόσο πιθανό είναι ένα άτομο να είναι σε μεγάλο κίνδυνο ή να αποτελεί μεγάλο έξοδο;”

- Αναβάθμιση φροντίδας ασθενή

Οι υπηρεσίες υγειονομικής περίθαλψης μπορούν να προσεγγίσουν το θέμα της θεραπείας των ασθενών πιο ενεργά. Για παράδειγμα, προβλέποντας πιο αποτελεσματικά ποιοι ασθενείς θα αναπτύξουν χρόνιες παθήσεις, ή ποιοι θα ανταποκριθούν καλύτερα σε συγκεκριμένους τύπους φαρμάκων ή θεραπειών, οι υγειονομικές υπηρεσίες μπορούν να επικεντρωθούν όχι μόνο στην θεραπεία ήδη υπάρχοντων ασθενειών, αλλά ακόμα και να προλαβαίνουν εξάρσεις τους.

- Βελτιστοποίηση αξιοποίησης πόρων

Τάσεις στην εισαγωγή ασθενών στα νοσοκομεία, την κάλυψη κλινών, τη διάρκεια νοσηλείας και άλλοι δείκτες μπορούν να αναλυθούν και να προβλέψουν μελλοντικά πρότυπα. Τα νοσοκομεία μπορούν να είναι καλύτερα προετοιμασμένα και να διασφαλίζουν ότι υπάρχουν αρκετοί πόροι διαθέσιμοι για να προσφέρουν την καλύτερη δυνατή φροντίδα, και με τη σειρά τους καλύτερα εκπαιδευμένο προσωπικό, κλινικές, διαγνωστικά μηχανήματα κλπ.

- Ανίχνευση απάτης

Οι τεχνικές προβλέψεων μπορούν να βοηθήσουν του επαγγελματίες της υγειονομικής περίθαλψης να διακρίνουν περιπτώσεις οι οποίες χρειάζονται περαιτέρω ανάλυση για απάτη αυξάνοντας την πιθανότητα της ανακάλυψης δόλιας περίπτωσης.

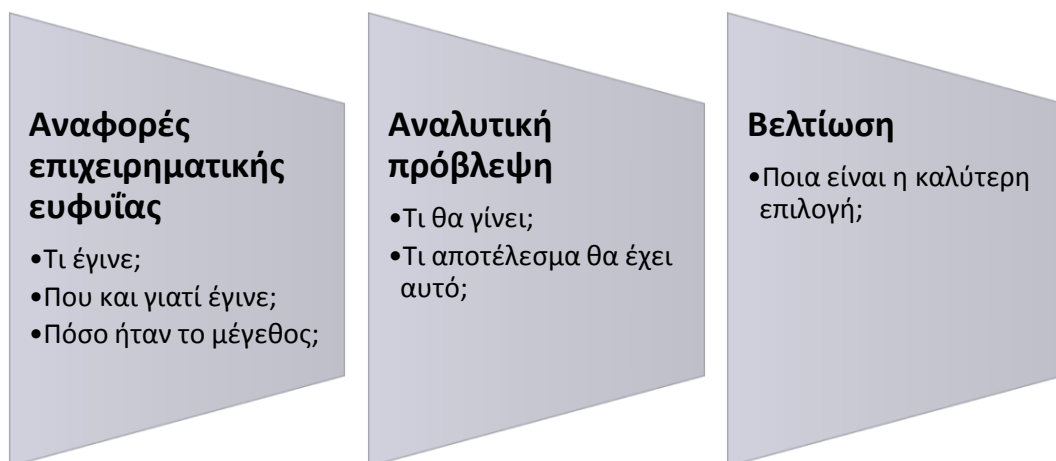
- Βελτίωση των ιατρικών αποτελεσμάτων και αύξηση εσόδων

Οι υπηρεσίες υγειονομικής περίθαλψης μπορούν να εξάγουν ιατρικά δεδομένα από μεγάλες ποσότητες πληροφοριών ασθενών έτσι ώστε να κατανοήσουν το ιστορικό των ασθενών και να προβλέψουν μελλοντικά αποτελέσματα. Αναλύοντας προσεκτικά ποια θεραπεία είναι η καλύτερη, τα νοσοκομεία και οι προμηθευτές αυτών μπορούν να πάρουν πιο έξυπνες αποφάσεις όσο αναφορά τις υπηρεσίες ή τα φάρμακα που προσφέρουν, ελαχιστοποιώντας τις επιπλοκές και τα έξοδα των ασθενών.

2.2 Μοντέλα πρόβλεψης

Ένα μοντέλο πρόβλεψης είναι μια μαθηματική αναπαράσταση της πραγματικότητας που προσφέρει μια άποψη βασισμένη σε επιστημονικές και στατιστικές αναλύσεις και η οποία χρησιμοποιεί δεδομένα που είναι διαθέσιμα ώστε να προβλεφθεί μια συγκεκριμένη κατάσταση σε κάποια στιγμή στο μέλλον. Τα μοντέλα αναζητούν στοιχεία στα δεδομένα που μπορεί να μην είναι πάντα ορατά και που φαίνεται να συνδέονται με την έκβαση κάποιου αποτελέσματος. Ανάλογα με τα δεδομένα το μοντέλο μπορεί να λειτουργήσει είτε με μάθηση με επίβλεψη (supervised learning), όπου το μοντέλο περιγράφει μια σχέση ανάμεσα σε ένα σύνολο ανεξάρτητων ιδιοτήτων και εξαρτώνται από ένα χαρακτηριστικό είτε με μάθηση χωρίς επίβλεψη (unsupervised learning) όπου το μοντέλο, το ίδιο, θα βρει τις σχέσεις των στοιχείων χωρίς αναφορά σε ανεξάρτητες ή εξαρτώμενες μεταβλητές. (Schick, 2012)

Μία υπηρεσία υγειονομικής περίθαλψης η οποία λαμβάνει αποφάσεις βάση αναλύσεων θα πρέπει να ακολουθεί τα εξής βήματα τα οποία απαντούν και σε αντίστοιχα ερωτήματα.



Εικόνα 3 Βήματα Αναλυτικής Πρόβλεψης

Τα παραπάνω βήματα σχετίζονται άμεσα με τις κατηγορίες που καταλήγουν στην αναλυτική πρόβλεψη, αυτές είναι:

Συναλλαγή εκθέσεων	<ul style="list-style-type: none"> • Βασικές Αναφορές • Διαγράμματα
Ενσωμάτωση δεδομένων – Αποθήκη δεδομένων	<ul style="list-style-type: none"> • Πίνακες • Κλινικά δεδομένα αποθηκών
Αναλύσεις υποστήριξης αποφάσεων	<ul style="list-style-type: none"> • Γενικές αναλύσεις υγειονομικής υπηρεσίας • Τεκμηριωμένη ιατρική
Αναλυτική πρόβλεψη	<ul style="list-style-type: none"> • Εξατομικευμένη υγειονομική περίθαλψη • Ανίχνευση απάτης • Συμπεριφορά ασθενούς

Εικόνα 4 Κατηγορίες Αναλυτικής Πρόβλεψης

2.3 Εξόρυξη δεδομένων (Data Mining)

Η εξόρυξη δεδομένων είναι η διαδικασία για την ανακάλυψη σημαντικής γνώσης από μεγάλο όγκο δεδομένων (Han, et al., 2006). Είναι ένας διεπιστημονικός τομέας ο οποίος προσφέρει σε πολλές κατηγορίες, όπως η στατιστική, η μηχανική μάθηση, ανάκτηση πληροφοριών, αναγνώριση προτύπων και βιοπληροφορική. Η εξόρυξη δεδομένων χρησιμοποιείται ευρέως σε πολλούς επαγγελματικούς κλάδους, όπως η οικονομία, τηλεπικοινωνίες και τα μέσα κοινωνικής δικτύωσης.

Οι κύριες τεχνικές για την εξόρυξη δεδομένων περιέχουν την ταξινόμηση και πρόβλεψη, συσταδοποίηση, εύρεση έκτοπων (outlier detection), τους κανόνες συσχέτισης, ανάλυση χρονοσειρών και εξόρυξη κειμένου, και ακόμα μερικές νέες τεχνικές όπως η ανάλυση των μέσων κοινωνικής δικτύωσης και ανάλυση συναισθήματος (Han, και συν., 2006) (Hand, et al., 2001) (Witten, et al., 2005).

Είναι σημαντικής σημασίας τομείς όπως συστήματα βάσεων δεδομένων, στατιστικής, μηχανικής μάθησης και αναγνώρισης πληροφοριών να συνεργαστούν με τέτοιο τρόπο έτσι ώστε η γνώση που θα προκύψει να είναι εύκολα επεξεργάσιμη και κατανοητή από τον αναλυτή.

Η εξόρυξη γνώσης από δεδομένα περιλαμβάνει δύο σημαντικά χαρακτηριστικά. Το πρώτο είναι η ανακάλυψη γνώσης από βάσεις δεδομένων και το δεύτερο είναι οι τεχνικές που χρησιμοποιούνται για την ανάλυση και εξαγωγή της γνώσης από σύνολα δεδομένων.

Τα στάδια της εξόρυξης δεδομένων είναι τα εξής (Παπαδόπουλος, 2011) (Mohammed, και συν., 2014):

1. Επιλογή του συνόλου των δεδομένων

Από την αποθήκη δεδομένων επιλέγονται εκείνα τα δεδομένα που θα χρειαστούν στη διαδικασία εξόρυξης δεδομένων που πρόκειται να εφαρμοστεί.

2. Καθαρισμός και προεπεξεργασία των δεδομένων

Σ' αυτό το βήμα γίνεται η συμπλήρωση των ελλιπών πεδίων δεδομένων, με βάση τις γνωστές τεχνικές διαχείρισης δεδομένων, και η αφαίρεση ακραίων τιμών δεδομένων (outliers).

3. Μετασχηματισμός των δεδομένων

Τα δεδομένα μετασχηματίζονται στην κατάλληλη μορφή για να επιτευχθεί η εξόρυξη. Χρησιμοποιούνται ποικίλες μέθοδοι και τεχνικές για να γίνει μείωση των διαστάσεων και των μεταβλητών. Επιπλέον κωδικοποιούνται τα δεδομένα για να αποφευχθούν, όπου είναι εφικτό, οι μεταβλητές.

4. Επιλογή των κατάλληλων αλγορίθμων εξόρυξης δεδομένων

Επιλέγονται οι μέθοδοι που θα χρησιμοποιηθούν, δεδομένων των στόχων που επιδιώκουμε. Αποτέλεσμα του βήματος αυτού είναι η παραγωγή προτύπων που συμβάλουν στην αποκόμιση νέων πληροφοριών και γνώσης.

5. Αξιολόγηση των προτύπων

Τα πρότυπα που έχουν εξαχθεί από το προηγούμενο βήμα αξιολογούνται με ορισμένα κριτήρια και αγνοούνται τελείως αυτά που δεν προσφέρουν καινούργια γνώση. Έτσι στο τέλος παραμένουν μόνο τα πρότυπα, τα οποία έχουν ενδιαφέρον και προσφέρουν κάτι το καινούργιο.

6. Παρουσίαση της ανακύπτουσας γνώσης

Στο βήμα αυτό, η νέα γνώση που έχει εξορυχθεί απεικονίζεται με χρήση τεχνικών στο χρήστη.

Η παραπάνω διαδικασία είναι επαναληπτική. Από οποιοδήποτε βήμα μπορεί κάποιος να μεταπηδήσει σε οποιοδήποτε προγενέστερο βήμα. Η ροή των βημάτων απεικονίζεται στην εικόνα 5. Παρόλο που το βήμα στο οποίο πραγματοποιείται η εξόρυξη δεδομένων αποτελεί την κύρια εργασία στη διαδικασία εξόρυξης γνώσης, όλα τα βήματα είναι εξίσου σημαντικά.



Εικόνα 5 Βήματα διαδικασίας εξόρυξης δεδομένων

2.4 Συστήματα Υποστήριξης Κλινικών Αποφάσεων

Ευφυή συστήματα σήμερα, έχουν μεγάλο ρόλο στην υποστήριξη συνταγογράφησης φαρμάκων, σε κλινικά εργαστήρια και διάφορες παρακολουθήσεις τους ή σε τομείς πλούσιους με δεδομένα όπως αυτός της

παρακολούθησης της εντατικής φροντίδας. Όταν η περίπτωση ενός ασθενή είναι πολύπλοκη, σπάνια ή το άτομο που κάνει τη διάγνωση είναι χωρίς την κατάλληλη εκπαίδευση, ένα έμπειρο σύστημα μπορεί να βοηθήσει στη σχηματοποίηση μιας πιθανής διάγνωσης βασισμένη στα δεδομένα του ασθενή που παρουσιάστηκαν και στην αντίληψη που έχει το σύστημα πάνω στην ασθένεια ως έχει αποθηκευτεί στις βάσεις δεδομένων. Μέσα από την επεξεργασία των δεδομένων λαμβάνουν μέρος ταυτόχρονα τεχνικές προβλέψεων οι οποίες θα αποθηκεύονται με σκοπό την μελλοντική τους χρήση σε αναλύσεις ασθενών ή ακόμα και προλήψεις κατά ασθενειών. Ένα τέτοιο έμπειρο σύστημα που μπορεί να εκπαιδευτεί έτσι ώστε να είναι ικανό να βοηθά το ιατρικό προσωπικό στην λήψη αποφάσεων ονομάζεται Σύστημα Υποστήριξης Κλινικών Αποφάσεων. (Bari, et al., 2014)

Οι λειτουργίες των συστημάτων αυτών, περιλαμβάνονται γύρω από τα ενδεχόμενα:

1. Να αποκτήσουν πεδίο γνώσεων
2. Να εξάγουν συμπεράσματα βάση της αποκτηθείσας γνώσης

Το πρώτο βήμα για την απόκτηση πεδίου γνώσεων μπορεί να συμβεί είτε μέσω οντολογικών κατασκευών που βρίσκονται μέσα στο σύστημα προγραμματιστικά κατά τη διάρκεια του χρόνου σχεδίασής τους, ή μέσω επιστημονικών μοντέλων που προέρχονται από τα δεδομένα που είχαν διατεθεί κατά τη διάρκεια του χρόνου λειτουργίας τους ή τον συνδυασμό και των δύο.

Η παραγωγή συμπερασμάτων από την άλλη έχει πολλές προσεγγίσεις οι οποίες αναπτύσσονται συνεχώς και επεκτείνονται από την κριτική βασισμένη στα δεδομένα, στη λογική κριτική στη λογική βασισμένη στην εργασία κοκ. Παρόλα αυτά, αυτές οι τεχνικές συμπερασμάτων συνδέονται στενά με τον μηχανισμό αναπαραγωγής γνώσεων που χρησιμοποιεί το σύστημα. (Bari, et al., 2014)

2.5 Ο Ρόλος της Πρόβλεψης στον Προγραμματισμό Ενεργειών και την Λήψη Αποφάσεων

Ο σκοπός της πρόβλεψης ήταν ανέκαθεν εύκολο να προσδιοριστεί. Παρόλα αυτά, κάποιος μπορεί να στρέψει την προσοχή του στο αντικείμενο που πρόκειται να

προβλεφθεί και ίσως ο καλύτερος τρόπος για να επιτευχθεί μία τέτοια προσέγγιση είναι ο υπολογισμός και του τομέα του προγραμματισμού ενεργειών και ο ρόλος αυτού στην συνολική οργάνωση που θα έχει μία πρόβλεψη. Η σωστή κατανόηση του προγραμματισμού ενεργειών μπορεί να βοηθήσει κάποιον να αποφασίσει ποιο είδος πρόβλεψης είναι πιο κατάλληλο και πιο χρήσιμο.

Η πρόβλεψη όσο και ο σχεδιασμός ενεργειών αφορούν συμβάντα του μέλλοντος και έτσι είναι σημαντικό να συνυπάρχουν μέσα σε μία υγειονομική υπηρεσία. Μόνο η γνώση μεθόδων πρόβλεψης δεν είναι αρκετή εάν δεν εφαρμόζονται σωστά έλεγχοι δραστηριοτήτων προγραμματισμού ενεργειών για μία σωστά εκτελεσμένη και έγκυρη πρόβλεψη.

Η λειτουργία του προγραμματισμού ενεργειών είναι χρήσιμη στις περιπτώσεις που: (Makridakis, et al., 1983)

α) Εμφανίζονται διάφορες ανάγκες ή ευκαιρίες, οι οποίες πρέπει να αντιμετωπίζονται χωρίς χάσιμο χρόνου

β) Το μέλλον επηρεάζεται εάν λάβουν μέρος συγκεκριμένες ενέργειες σε σύντομο χρονικό διάστημα

Να σημειωθεί ότι οι παραπάνω περιπτώσεις ισχύουν αφού έχουν πρώτα βασισθεί σε έγκυρες μελλοντικές προβλέψεις.

Ο προγραμματισμός ενεργειών δεν επιτρέπει μόνο τον επηρεασμό του μέλλοντος, αλλά παρέχει ακόμα χαρακτηριστικά που μπορούν να διαμορφώσουν ένα ακόμα πιο αποτελεσματικό περιβάλλον εργασίας. Ένα τέτοιο χαρακτηριστικό είναι τα αυξημένα κίνητρα των ατόμων μέσα σε μία υγειονομική υπηρεσία.

Πολλές επιχειρήσεις, μετά από έρευνες, έχουν καταλήξει στο συμπέρασμα ότι το ηθικό ενός εργαζομένου βελτιώνεται όταν αναλαμβάνει καθήκοντα πρόβλεψης του μέλλοντος της επιχείρησης και οι στόχοι αυτής είναι πιο πιθανό να επιτευχθούν όταν τους έχουν κατανοήσει και είναι αφοσιωμένοι σε αυτούς. Ο σχεδιασμός ενεργειών βοηθά ακόμα και στο συντονισμό διαφόρων ειδών δραστηριοτήτων και κάνει μία επιχείρηση ικανή στο να φτάσει τους στόχους της με τη χρήση των λιγότερο δυνατών πόρων.

Σύμφωνα με τον Steiner, τα χαρακτηριστικά του προγραμματισμού ενεργειών είναι τα ακόλουθα:

1. Επικεντρώνεται στην απάντηση ερωτημάτων μεγάλης σημασίας της επιχείρησης
2. Παρέχει δυνατότητες για την λήψη αποφάσεων όπως:
 - Προσομοιώνει το μέλλον
 - Αναγκάζει τον καθορισμό των στόχων
 - Αποκαλύπτει μελλοντικές ευκαιρίες και απειλές
 - Παρέχει αποφάσεις ενιαίου πλαισίου
 - Ενθαρρύνει την μέτρηση της αποδοτικότητας
 - Συνδέει μια ποικιλία λειτουργιών διαχείρισης
 - Φέρνει στο φως προβλήματα στρατηγικής σημασίας
3. Παρέχει βελτίωση της συμπεριφοράς εργαζομένων αποκτώντας επικοινωνία, εκπαίδευση και αίσθηση συμμετοχής.

Το επόμενο βήμα μετά τον προγραμματισμό ενεργειών στην πρόβλεψη είναι η λήψη αποφάσεων.

Επαγγελματίες χρησιμοποιούν ποιοτικές και ποσοτικές μεθόδους για την πρόβλεψη μελλοντικών θεμάτων όπως η ζήτηση κάποιας υπηρεσίας ή η έκβαση ασθενειών με σκοπό να παρθούν ανάλογες αποφάσεις. Ποιοτικές τεχνικές περιέχουν ερωτηματολόγια, γνώμες ειδικών και διάφοροι πειραματισμοί για μια συγκεκριμένη αγορά, ενώ οι ποσοτικές τεχνικές περιέχουν την ανάλυση χρονοσειρών, την βαρομετρική μέθοδο όπου εξετάζονται οι σχέσεις μεταξύ συνηθισμένων ή τυχαίων γεγονότων κ.α.

Οι οικονομικές προβλέψεις μπορούν να συμβάλουν στην καλύτερη κατανόηση των αλλαγών που λαμβάνουν μέρος στον αντίστοιχο κλάδο. Το ρίσκο και η αβεβαιότητα είναι δύο κύριοι παράγοντες της διαδικασίας λήψης αποφάσεων. Το ρίσκο είναι μία κατάσταση όπου ο επιχειρηματίας μπορεί να μετρήσει τα πιθανά έσοδα και έξοδα προερχόμενα από μία συγκεκριμένη απόφαση. Παρόλα αυτά, η αβεβαιότητα εμφανίζεται όταν το ρίσκο που εμπειρείχε μία λήψη απόφασης δεν μπορεί να υπολογιστεί.

Η πρόγνωση για την λήψη αποφάσεων γίνεται για να παρθούν σοβαρές αποφάσεις για το μέλλον μιας επιχείρησης και για αυτό το λόγο θα πρέπει να διαχειρίζεται σωστά από αναλυτές και στελέχη καθώς αποφάσεις τέτοιου είδους είναι μη αντιστρέψιμες. Έτσι, ο προϋπολογισμός των κεφαλαίων είναι απαραίτητος με σκοπό την αποτελεσματικότερη οργάνωση και διαχείριση μεγάλων αποφάσεων. (Makridakis, et al., 1983)

2.6 Κατηγορίες Τεχνικών Προβλέψεων

Οι μέθοδοι προβλέψεων χωρίζονται σε 3 κατηγορίες. Τα ποιοτικά μοντέλα, τα ποσοτικά μοντέλα και τα συστήματα τεχνητής νοημοσύνης. (Wikipedia)

Αναλυτικότερα τα **ποσοτικά μοντέλα**, γνωστά και ως στατιστικά, προσεγγίζουν μία πρόβλεψη αντικειμενικά. Παρέχουν μια σειρά από βήματα τα οποία μπορούν να αναπαραχθούν και να εφαρμοστούν σε επιχειρηματικές και οικονομικές περιπτώσεις. Στην κατηγορία των ποιοτικών μοντέλων πρόβλεψης ανήκουν οι χρονοσειρές και μέθοδοι παλινδρόμησης.

Οι χρονοσειρές ή αλλιώς τα αυτοπαλινδρομικά μοντέλα, προβλέπουν τιμές μιας μεταβλητής βασισμένα εξ 'ολοκλήρου σε ιστορικές παρατηρήσεις της. Οι αυτοπαλινδρομικές τεχνικές περιέχουν τις μεθόδους κινητού μέσου όρου, εκθετικής εξομάλυνσης, ανάλυση χρονοσειρών και τα μοντέλα ARIMA.

Οι μέθοδοι παλινδρόμησης είναι βασισμένοι στις σχέσεις που υπάρχουν σε θεωρητικό επίπεδο μεταξύ των μεταβλητών και στην οικονομική θεωρία. Λειτουργούν υποθέτοντας ότι μία τιμή μιας τυχαίας ανεξάρτητης μεταβλητής μπορεί να επηρεάσει την τιμή μίας άλλης, εξαρτημένης μεταβλητής. Ο τελικός στόχος είναι η εύρεση της ακριβής μορφής των σχέσεων που υπάρχουν μέσα στις αντίστοιχες μεταβλητές, κάτι το οποίο επιτυγχάνεται από την απορροή των υπολογισμών των συντελεστών παλινδρόμησης που περιέχουν ιστορικές σχέσεις μεταξύ ανεξάρτητων και εξαρτημένων τιμών. (Walter L. Perry, 2013)

Οι **ποιοτικές μέθοδοι** από την άλλη ονομάζονται μη στατιστικές ή κριτικές μέθοδοι. Αυτή η κατηγορία βασίζεται σημαντικά από τις γνώσεις, τις γνώμες και

την κριτική ικανότητα του ατόμου που κάνει την πρόβλεψη καθώς χρησιμοποιούνται σε περιπτώσεις όπου τα ιστορικά δεδομένα δεν είναι αρκετά.

Παρά την εξέλιξη των τεχνικών πρόβλεψης και την ολοένα μεγαλύτερη χρήση τους από αναλυτές, έρευνες έχουν δείξει ότι η χρήση απλών **συστημάτων τεχνητής νοημοσύνης** μπορεί να οδηγήσουν σε καλύτερα αποτελέσματα προβλέψεων από ότι μία απλή τεχνική (DeLurgio, 1998). Συνεπώς σήμερα χρησιμοποιούνται στη διαδικασία της πρόβλεψης συστήματα τεχνητής νοημοσύνης όπως είναι τα εξής:

- Τεχνητά Νευρωνικά Δίκτυα
- Μηχανές Διανυσμάτων Υποστήριξης
- Γενετικός Αλγόριθμος

2.7 Ποσοτικά Μοντέλα Προβλέψεων

2.7.1 Μοντέλα Παλινδρόμησης

Η ανάλυση της παλινδρόμησης (Regression analysis) αφορά την πρόβλεψη της τιμής μιας μεταβλητής με βάση τις τιμές άλλων μεταβλητών. Χρησιμοποιείται για την επαγωγική στατιστική αφού όλες οι μελέτες ασχολούνται με μελλοντικές προβλέψεις τιμών. Η κατασκευή για την ανάλυση της παλινδρόμησης γίνεται με μαθηματικές εξισώσεις όπου περιγράφουν την σχέση της μεταβλητής που θα προβλεφθεί, η οποία λέγεται εξαρτημένη μεταβλητή (Dependent Variable) και συμβολίζεται με το γράμμα Y , και μιας ή περισσότερων άλλων μεταβλητών όπου είναι οι ανεξάρτητες (Independent Variables) που θα συμβολίζονται ως X_1, X_2, \dots, X_k όπου το k είναι το πλήθος των ανεξάρτητων μεταβλητών. Με την βοήθεια των διαγραμμάτων διασποράς (Scatter Diagram), την ανάλυση της συσχέτισης (Correlation Analysis) και τον συντελεστή συσχέτισης (Coefficient or Correlation) μπορούμε να δούμε αν υπάρχει σχέση μεταξύ των μεταβλητών (Keller, 2010).

2.7.1.1 Απλή Γραμμική Παλινδρόμηση

Στην Απλή Γραμμική Παλινδρόμηση δημιουργείται μια μαθηματική εξίσωση ή μοντέλο περιγράφοντας την σχέση των μεταβλητών, ανεξάρτητα από το περιβάλλον όπου χρησιμοποιείται η παλινδρόμηση. Η κατασκευή της μαθηματικής εξίσωσης μπορεί να είναι πολύπλοκη δουλειά και προϋποθέτει την ύπαρξη γνώσης γιατί ο αριθμός των μαθηματικών μοντέλων είναι απεριόριστος. Στις εξισώσεις της φυσικής ($E=mc^2$ ενέργεια μάζας ηρεμίας και $S=g*t^2/2$ ελεύθερη πτώση) η εξαρτημένη μεταβλητή παίρνει την ίδια τιμή όσες φορές κι αν επαναληφθεί το πείραμα. Αυτό λέγεται αιτιοκρατικό ή ντετερμινιστικό μοντέλο. Για την κατασκευή ενός πιθανοθεωρητικού μοντέλου χρησιμοποιούμε ένα ντετερμινιστικό μοντέλο που προσεγγίζει την σχέση μεταξύ των μεταβλητών και προσθέτουμε έναν όρο που μετρά το τυχαίο σφάλμα. Για παράδειγμα στην πώληση ενός ακινήτου δεν μπορούμε να έχουμε πάντα την ίδια τιμή αφού επηρεάζεται από το μέγεθος, την τοποθεσία και πολλές άλλες μεταβλητές μη μετρήσιμες. Με άλλα λόγια το τυχαίο σφάλμα μετρά το σύνολο των μεταβλητών μετρήσιμων και μη που δεν υπάρχουν στο μοντέλο. Αυτό είναι το μοντέλο γραμμικής πρώτης τάξης (First-order Linear Model) ή μοντέλο απλής γραμμικής παλινδρόμησης (Simple Linear Regression Model) όπου δίνεται από την εξίσωση (Keller, 2010):

$$y = b_0 + b_1 * x + e$$

Όπου:

y εξαρτημένη μεταβλητή

x ανεξάρτητη μεταβλητή

b_0 ίχνος στον άξονα y

b_1 κλίση της ευθείας ανά μονάδα x

e μεταβλητή σφάλματος

Η απλή γραμμική παλινδρόμηση είναι η μέθοδος για την συσχέτιση δύο μεταβλητών που λαμβάνουν ποσοτικές συνεχείς τιμές με στόχο τη διερεύνηση των

δύο μεταβλητών x , y . Οι συντελεστές β_0 και β_1 δίνονται από την εξίσωση ευθείας γραμμής :

$$\hat{y} = b_0 + b_1 * x$$

Η ευθεία αυτή λέγεται γραμμή παλινδρόμησης (Regression Line) ή ευθεία ελάχιστων τετραγώνων (Least Squares Line) με \hat{y} η τιμή της πρόβλεψης για κάθε τιμή της ανεξάρτητης μεταβλητής.

Για να μπορεί να εφαρμοστεί η ανάλυση υπάρχουν 4 προϋποθέσεις που πρέπει να ελεγχθούν

- a. Η κατανομή του σφάλματος πρέπει να είναι κανονική.
- b. Ο μέσος της κατανομής πρέπει να είναι μηδέν
- c. Η τυπική απόκλιση του σφάλματος πρέπει να είναι σταθερή για όλα τα x .
- d. Η τιμή του σφάλματος για τις τιμές του y είναι ανεξάρτητη από την τιμή του σφάλματος για κάθε άλλη τιμή του y .

Η διαδικασία ανάλυση παλινδρόμησης αποτελείται από τα εξής βήματα:

1. Θεωρητική σύλληψη του μοντέλου: επιλογή ανεξάρτητης και εξαρτημένης μεταβλητής.
2. Συλλογή δεδομένων
3. Σχεδίαση του διαγράμματος (για παράδειγμα διασποράς)
4. Υπολογισμός της ευθείας γραμμής παλινδρόμησης
5. Υπολογισμός υπολοίπων και έλεγχος προϋποθέσεων:
 - αν είναι η διασπορά σταθερή,
 - τα σφάλματα εξαρτημένα,
 - αν υπάρχουν τιμές ακραίες,
 - αν έχει η μεταβλητή σφάλματος κανονική κατανομή.
6. Αξιολόγηση του μοντέλου:
 - Έλεγχος κλίσης
 - Εκτίμηση σφάλματος
 - Συντελεστής προσδιορισμού

7. Χρήση του μοντέλου για την πρόβλεψη και εκτίμηση τιμών.

2.7.1.2 Πολλαπλή γραμμική παλινδρόμηση

Στην πολλαπλή παλινδρόμηση χρησιμοποιούμε δύο ή περισσότερες ανεξάρτητες μεταβλητές. Έτσι θα ταιριάζουμε τα δεδομένα καλύτερα από ένα απλό γραμμικό μοντέλο. Η μορφή της εξίσωσης θα είναι :

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k + e$$

όπου b_0, b_1, \dots, b_k είναι οι συντελεστές και e η μεταβλητή σφάλματος. Οι μεταβλητές x_1, x_2, \dots, x_k μπορούν επίσης να πάρουν την μορφή άλλων μεταβλητών πιο σύνθετων. Όσον αφορά την απεικόνιση του διαγράμματος αφού έχουμε πολλές μεταβλητές ανεξάρτητες άρα θα έχουμε και μια πολυδιάστατη επιφάνεια που την ονομάζουν επιφάνεια απόκρισης (Response Surface).

Οι προϋποθέσεις είναι ίδιες με της απλής γραμμικής παλινδρόμησης για την μεταβλητή σφάλματος. Αν οι προϋποθέσεις δεν είναι έγκυρες τότε τα αποτελέσματα δεν έχουν νόημα και το μοντέλο θα πρέπει να διορθωθεί και επαναληφθεί η ανάλυση. Ύστερα τα βήματα της διαδικασίας είναι (Keller, 2010):

1. Χρήση υπολογιστή και λογισμικού για την δημιουργία συντελεστών
2. Έλεγχος των προϋποθέσεων
3. Αξιολόγηση των μοντέλων
4. Αν οι έλεγχοι δείξουν ότι τηρούνται οι προϋποθέσεις τότε μπορούμε να χρησιμοποιήσουμε το μοντέλο για τις προβλέψεις μας.

Συμπεραίνουμε λοιπόν ότι και οι δύο παλινδρομήσεις χρησιμοποιούνται είτε για πρόβλεψη μεμονωμένων τιμών, είτε για εκτίμηση της εξαρτημένης μεταβλητής. Η διαφορά τους είναι ότι επεκτείνει την απλή γραμμική παλινδρόμηση βάζοντας περισσότερες ανεξάρτητες μεταβλητές.

2.7.2 Μοντέλα Χρονοσειρών

2.7.2.1 Ποιοτικά Χαρακτηριστικά Χρονοσειρών

Υπάρχουν τέσσερα ποιοτικά χαρακτηριστικά μιας χρονοσειράς: η τάση, η κυκλικότητα, η εποχικότητα και η τυχαιότητα. (Makridakis, και συν., 1983)

- Τάση

Η τάση χρησιμοποιείται για την απεικόνιση μακροπρόθεσμων προβλέψεων οι οποίες μπορεί να είναι ανοδικές, καθοδικές ή στάσιμες.

- Κυκλικότητα

Αναφέρεται σε περιοδικές ανόδους ή καθόδους που παρατηρούνται πάνω από μια χρονική περίοδο και είναι γνωστή και ως επιχειρηματικός κύκλος. Καλύπτει ένα μεγαλύτερο χρονικό φάσμα και δεν υπόκειται σε πρότυπα που είναι εύκολα προβλέψιμα όπως είναι η περίοδος αιχμής και η περίοδος ύφεσης.

- Εποχικότητα

Σε μια χρονοσειρά η εποχικότητα αναφέρεται στη σταδιακή αλλαγή των τιμών των δεδομένων η οποία λαμβάνει μέρος την ίδια στιγμή κάθε χρόνο. Η εποχιακή επανάληψη μπορεί να είναι ακριβής ή κατά προσέγγιση.

- Τυχαιότητα

Αναφέρεται στις ακανόνιστες μεταβολές σε μια χρονοσειρά οι οποίες δεν είναι αποτέλεσμα των τριών παραπάνω χαρακτηριστικών. Είναι γνωστή και ως παράγοντας σφάλματος και ο οποίος δεν είναι προβλέψιμος και απομακρύνεται από την χρονοσειρά μέσω μεθόδων εξομάλυνσης.

2.7.2.2 Διαχείριση κενών και μηδενικών τιμών

Οι κενές τιμές αφορούν περιπτώσεις όπου η τιμή κάποιων περιόδων δεν είχε καταγραφεί και αποθηκευτεί στη βάση δεδομένων. Ο λόγος μπορεί να οφείλεται σε στοιχεία του πληροφοριακού συστήματος ή σε λάθος χειρισμό του υπεύθυνου

χρήστη. Ανεξαρτήτως αιτίας, ακολουθείται μία από τις παρακάτω διαδικασίες εκτίμησης της ελλείπουσας τιμής, ανάλογα, με την περίπτωση (Γεωργιάδης, 2013):

- Εύρεσης της κενής τιμής από άλλες πηγές ή απευθείας ορισμός αυτής, αν υπάρχει ασφαλής κριτική εκτίμηση για το ύψος στο οποίο κυμάνθηκε.
- Η κενή τιμή ορίζεται ως το ημιάθροισμα (μέσος όρος) της προηγούμενης και της επόμενης παρατήρησης, όταν η χρονοσειρά χαρακτηρίζεται από στασιμότητα και δεν παρατηρείται εποχιακή συμπεριφορά.
- Αν η χρονοσειρά παρουσιάζει σαφή εποχιακή συμπεριφορά, τότε η κενή τιμή ορίζεται ως ο μέσος όρος των τιμών των αντίστοιχων περιόδων. Για παράδειγμα, αν τα δεδομένα αποτελούνται από μηνιαίες παρατηρήσεις και παρατηρηθεί κενή τιμή στον Ιούνιο κάποιου έτους, τότε η κενή τιμή ορίζεται ως ο μέσος όρος των Ιουνίων.

2.7.3 Μέθοδοι Εκθετικής Εξομάλυνσης

Οι μέθοδοι εκθετικής εξομάλυνσης εφαρμόζουν το περισσότερο βάρος σε παρατηρήσεις οι οποίες είναι πιο πρόσφατες, «αμελώντας» αυτές που είναι πιο κοντά στο παρελθόν. Με αυτόν τον τρόπο οι μέθοδοι εξομάλυνσης εκμηδενίζουν την ανάγκη για αποθήκευση παρελθοντικών τιμών των μεταβλητών καθώς ακόμα διαχωρίζουν τις τυχαίες διακυμάνσεις από τα πραγματικά πρότυπα.

Η δημοτικότητα των μεθόδων αυτών οφείλεται στην απλότητα των μοντέλων που υιοθετούν, τις περιορισμένες απαιτήσεις τους σε αποθήκευση δεδομένων και τον μειωμένο υπολογιστικό φόρτο. Εμπειρικές μελέτες αποδεικνύουν ότι οι μέθοδοι εκθετικής εξομάλυνσης παρουσιάζουν ικανοποιητικά ποσοστά ακρίβειας σε σχέση με πιο πολύπλοκες μεθόδους πρόβλεψης. Το γεγονός αυτό οφείλεται στο ότι οι μέθοδοι εκθετικής εξομάλυνσης δεν επηρεάζονται από τις ιδιομορφίες των προτύπων των δεδομένων ή από περιστασιακά εμφανιζόμενες ακραίες τιμές, οι οποίες παρατηρούνται σε επιχειρησιακά δεδομένα (Γεωργιάδης, 2013).

2.7.3.1 Απλή εκθετική εξομάλυνση

Η απλή εκθετική εξομάλυνση χρησιμοποιείται για βραχυπρόθεσμες προβλέψεις, συνήθως μόνο για έναν μήνα στο μέλλον. Το μοντέλο αυτό λειτουργεί υποθέτοντας ότι τα δεδομένα κυμαίνονται γύρω από έναν σταθερό μέσο όρο.

Η συνάρτηση της απλής εκθετικής εξομάλυνσης είναι η εξής (Makridakis, και συν., 1983):

$$y_{t+1} = ay_t + (1 - a)y_{t-1}$$

όπου

y_{t+1} η νέα τιμή μετά την εξομάλυνση

a ο συντελεστής εξομάλυνσης ($0 < a < 1$)

y_t η πραγματική τιμή σε περίοδο t

y_{t-1} η παλιά τιμή μετά από εξομάλυνση σε περίοδο $t - 1$

Όταν εφαρμόζεται αναδρομικά σε κάθε διαδοχική παρατήρηση, κάθε πρόβλεψη λαμβάνεται ως ο σταθμισμένος μέσος όρος της παρούσας παρατήρησης και της προηγούμενης παρατήρησης που έχει γίνει εξομάλυνση. Κατά αυτόν τον τρόπο, κάθε τιμή με εξομάλυνση είναι ο σταθμισμένος μέσος των προηγούμενων παρατηρήσεων όπου το βάρος αυτών μειώνεται βάση της τιμής του συντελεστή εξομάλυνσης a .

Για κάθε τιμή του συντελεστή εξομάλυνσης (a) μεταξύ 0 και 1, τα επισυναπτόμενα βάρη στις παρατηρήσεις, μειώνονται εκθετικά όσο προχωράμε πίσω στον χρόνο, εξ ου και το όνομα «Εκθετική Εξομάλυνση». Αν η τιμή του a είναι μικρή (κοντά στο 0), περισσότερα βάρη αναθέτονται σε παρατηρήσεις οι οποίες είναι χρονικά πιο απομακρυσμένες. Αν η τιμή του a είναι μεγάλη (κοντά στο 1), περισσότερα βάρη ανατίθενται σε πιο πρόσφατες παρατηρήσεις. Στην ακραία περίπτωση όπου $a=1$, τότε οι παρατηρήσεις σε χρονική στιγμή $T+1$ και οι προβλέψεις, είναι ίσες με αυτές της Απλοϊκής μεθόδου πρόβλεψης. (Hyndman, et al.)

2.7.3.2 Μοντέλο Γραμμικής Τάσης (Μέθοδος Holt)

Ένα από τα πρότυπα των δεδομένων που υπάρχουν μέσα σε χρονοσειρές είναι η τάση. Το μοντέλο γραμμικής τάσης είναι η εξέλιξη της Απλής Εξομάλυνσης και έχει την δυνατότητα να μπορεί να την διαχειριστεί. Η μεθοδολογία του μοντέλου εφαρμόζει εξομάλυνση στην τάση που παρουσιάζεται στα δεδομένα χρησιμοποιώντας διαφορετικές σταθερές εξομάλυνσης για κάθε μία.

Για την εφαρμογή του μοντέλου, πρέπει πρώτα να υπολογιστούν (Makridakis, και συν., 1983)

- Οι τιμές των μεταβλητών μετά από εκθετική εξομάλυνση (1)
- Η τάση (2)
- Η πρόβλεψη της τιμής σε x περιόδους στο μέλλον (3)

$$(1) \quad A_t = \alpha y_t + (1 - \alpha)(A_{t-1} + T_{t-1})$$

$$(2) \quad T_t = b(A_t - A_{t-1}) + (1 - b)T_{t-1}$$

$$(3) \quad y_{t+x} = A_t + xT_t$$

όπου

A_t , η τιμή της εξομάλυνσης

α , ο συντελεστής εξομάλυνσης ($0 < \alpha < 1$)

b , ο συντελεστής εξομάλυνσης για τον υπολογισμό της τάσης ($0 < b < 1$)

T_t , το αποτέλεσμα της τάσης

x , οι περίοδοι για τους οποίους θα γίνει η πρόβλεψη

y_{t+x} , οι προβλέψεις για x περιόδους στο μέλλον

2.7.4 Αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου με την μέθοδο ARIMA


Η μέθοδος μοντελοποίησης ARIMA προέρχεται από τους Box και Jenkins (1976) και είναι μία σημαντική τεχνική μέθοδος πρόβλεψης εξαιτίας της δομημένης βάσης και της κοινώς αποδεκτής επίδοσής της πάνω στην πρόβλεψη καθώς, θεωρητικά και πρακτικά, είναι η πιο γενική κατηγορία μοντέλων πρόβλεψης μιας

χρονοσειράς. Ένα μοντέλο ARIMA προβλέπει μελλοντικές τιμές μίας χρονοσειράς με έναν γραμμικό συνδυασμό των παρελθοντικών της τιμών και μια σειρά από σφάλματα $\{a_t\}$. (Makridakis, και συν., 1983)

Υπάρχουν δύο μορφές μοντέλων, αυτά με εποχικές και μη εποχικές τιμές.

Οι περισσότερες χρονοσειρές είναι μη εποχιακές και για αυτό τον λόγο δεν μπορούμε να εφαρμόσουμε μοντέλα, AR, MA ή ARMA κατευθείαν. Μία πιθανή λύση για την διαχείριση μη εποχικών σειρών είναι να εφαρμοστεί διαφορίση έτσι ώστε να γίνουν εποχικά. (Chatfield, 2000)

Η μορφή με την οποία ένα μοντέλο μη εποχιακών όρων αναφέρεται είναι: ARIMA (p, d, q), όπου p (AR) είναι ο αριθμός των αυτοπαλινδρομικών όρων, d (I) είναι ο αριθμός των μη εποχιακών διαφορήσεων, και q (MA) είναι ο αριθμός των σφαλμάτων της πρόβλεψης στην αντίστοιχη συνάρτηση. Η γενική συνάρτηση του μοντέλου αυτού είναι (Hyndman, et al.):

$$(1 - \varphi B - \dots - \varphi_p B^p) (1 - B)^d y_t = c + (1 + \theta B + \dots + \theta_q B^q) e_t$$


Όπου

t, είναι ο χρονικός ορίζοντας

B, είναι ο τελεστής των προς τα πίσω διαφορών (τελεστής ολίσθησης)

c, ο μέσος όρος των μεταβολών μεταξύ διαδοχικών παρατηρήσεων

y_t , μια παρατήρηση σε μία χρονική στιγμή

φ , ο όρος για το μη εποχικό μοντέλο AR

θ , ο όρος για το μη εποχικό μοντέλο MA

Η μορφή ενός μοντέλου εποχιακών όρων είναι η εξής ARIMA (p, d, q)(P, D, Q)m όπου αντίστοιχα, m η εποχικότητα (αριθμός περιόδων ανά εποχή), P ο αριθμός των εποχιακών αυτοπαλινδρομικών όρων, D ο αριθμός των εποχιακών διαφορήσεων,

και Q ο αριθμός των εποχιακών αντίστοιχων σφαλμάτων της πρόβλεψης. Η γενική συνάρτηση είναι (Hyndman, et al.):

$$(1 - \varphi B) (1 - \Phi B^m) (1 - B) (1 - B^m) y_t = (1 + \theta B)(1 + \Theta B^m) e_t$$

Μη εποχικότητα AR(p)
Εποχικότητα AR(P)
Μη εποχική διαφόριση (d)
Εποχική διαφόριση (D)
Μη εποχικότητα MA(q)
Εποχικότητα MA(Q)

Όπου

t , είναι ο χρονικός ορίζοντας

B , είναι ο τελεστής των προς τα πίσω διαφορών (τελεστής ολίσθησης)

c , ο μέσος όρος των μεταβολών μεταξύ διαδοχικών παρατηρήσεων

y_t , μια παρατήρηση σε μία χρονική στιγμή

φ , ο όρος για το μη εποχικό μοντέλο AR

θ , ο όρος για το μη εποχικό μοντέλο MA

Φ , ο όρος για το εποχικό μοντέλο AR

Θ , ο όρος για το εποχικό μοντέλο MA

2.7.4.1 Τιμές διαφόρισης (d)

- Η μη διαφόριση ($d=0$) είναι δείγμα ύπαρξης σταθερότητας στην αρχική χρονοσειρά. Ωστόσο αυτή η σταθερότητα ενδέχεται εν τέλει να είναι ανεπαρκής. Η εισαγωγή λοιπόν μίας σταθεράς c μπορεί να βοηθήσει στον καλύτερο προσδιορισμό του επιπέδου της.
- Η διαφόριση πρώτης τάξης ($d=1$) συνεπάγεται ύπαρξη σταθερής τάσης στην αρχική χρονοσειρά. Αυτή θεωρητικά έχει εξαλειφθεί μετά τη διαφόριση, ωστόσο ενδέχεται να έχει υποπέσει στην παραπάνω περίπτωση. Έτσι, η εισαγωγή μίας σταθεράς μπορεί να γίνει, έχοντας πρώτα βεβαιωθεί για την

ανάγκη τόνωσης του επίπεδου της παραγόμενης χρονοσειράς. Στην πράξη αυτό αποφεύγεται.

- Η διαφόριση δεύτερης τάξης ($d=2$) συνεπάγεται χρονικά μεταβαλλόμενη τάση (τάση μέσα στην τάση) στην αρχική χρονοσειρά. Έτσι, η εισαγωγή σταθεράς σε αυτό την περίπτωση θεωρείται άστοχη επιλογή. (Πετρόπουλος Φώτιος, 2015)

2.8 Τεχνητά Νευρωνικά Δίκτυα

2.8.1 Ιστορική αναδρομή

Με τις εργασίες των McCulloch και Pitts το 1943 ξεκίνησαν τα τεχνητά νευρωνικά δίκτυα. Η πρώτη προσπάθεια έγινε από τον Hebb το 1949, οποίος πρότεινε ότι η μάθηση στα δίκτυα είναι αποτέλεσμα σχηματισμού πολλών ομάδων νευρωνικών. Θεμελιώδεις εργασίες στον τομέα των Τ.Ν.Δ. έγιναν από τους Rosenblatt το 1958, που εισήγαγε τα Perceptions καθώς και τον Widrow το 1959, που απέδειξε με πειραματικές μελέτες ότι τα Τ.Ν.Δ. κατέχουν δυνατότητες μάθησης. Η λειτουργία των πολυστρωματικών Τ.Ν.Δ. δεν είχε κατανοηθεί ακόμα γιατί η έρευνα εστίαζε κυρίως σε δομημένα πρότυπα. Ακολούθησαν οι Εκπαιδευόμενες Μηχανές (Learning Machines) το 1960. Το 1969 οι Minsky και Papert στο βιβλίο τους «Perceptions» υποστήριξαν ότι τα δίκτυα αυτά έχουν περιορισμένες δυνατότητες μάθησης. Με την έλλειψη αλγορίθμων είχαν ως αποτέλεσμα την καθοδική πορεία των Τ.Ν.Δ. για δεκαπέντε χρόνια (Κίνγκ, 2004).

Στις αρχές του 1980, επανέρχονται με νέους και ισχυρούς αλγόριθμους μάθησης, με την εργασία του Hopfield. Η εξέλιξη στη συνέχεια ήταν μεγάλη και σήμερα αποτελούν βασικό πεδίο εφαρμοσμένης έρευνας σε πολλούς τομείς όπως την επεξεργασία εικόνας και ομιλίας, τη μηχανική όραση, τη ρομποτική, την ιατρική, τα συστήματα ελέγχου και άλλα. Σε μια δημοσίευση του ο Hunt το 1992 συμπεραίνει τα εξής χαρακτηριστικά των Τ.Ν.Δ. (Κίνγκ, 2004) :

- Έχουν από την φύση τους εφαρμογή σε μη-γραμμικό έλεγχο
- Έχουν έμφυτες ικανότητες ανεκτικότητας σε βλάβες

- Έχουν την ικανότητα να γενικεύει τις άγνωστες εισόδους
- Έχουν την ικανότητα να λειτουργούν με αριθμητικά και λεκτικά δεδομένα
- Μπορούν να επεξεργαστούν πολλαπλά σήματα εισόδου και εξόδου

Τα Τ.Ν.Δ. αποτελούνται από πολλά γραμμικά στοιχεία ή νευρώνες που προσδιορίζονται από την δομή τους, τον τρόπο επικοινωνίας τους με το περιβάλλον, τη διαδικασία μάθησης κατά την επεξεργασία δεδομένων. Τα Τ.Ν.Δ. χωρίζονται σε δύο κατηγορίες:

- Τα *στατικά νευρωνικά δίκτυα* που δεν περιέχουν στοιχεία με μνήμη αλλά μπορούν να έχουν ως εισόδους προηγούμενες τιμές.
- Τα *δυναμικά νευρωνικά δίκτυα* με στοιχεία μνήμης κατάλληλα για την προτυποποίηση μη-γραμμικών συστημάτων.

2.8.2 Πρότυπο ενός Τεχνητού Νευρώνα

Τα τεχνητά νευρωνικά δίκτυα είναι εμπνευσμένα από την δομή του ανθρώπινου εγκεφάλου. Ένα νευρωνικό δίκτυο παρέχει ένα πιο απλό μοντέλο σε σχέση με αυτό του ανθρώπινου εγκεφάλου, αλλά ο σκοπός εκπληρώνεται χωρίς προβλήματα. Μέσα από την επεξεργασία παρελθοντικών και παροντικών δεδομένων υπολογίζονται οι μελλοντικές τιμές με τρόπο ανάλογο με αυτόν του ανθρώπινου εγκεφάλου (Κίνγκ, 2004).

Τα Τ.Ν.Δ. αποτελούνται από κυκλώματα που περιέχουν κόμβους που συνδέονται με διάφορες αρχιτεκτονικές. Οι συντελεστές διασύνδεσης, ή συναπτικά βάρη (synaptic weights) του δικτύου, είναι άγνωστα και στόχος της μάθησης είναι η εύρεση τους ελαχιστοποιώντας κάποιο μέτρο μεταξύ της επιθυμητής και πραγματικής εξόδου του δικτύου.

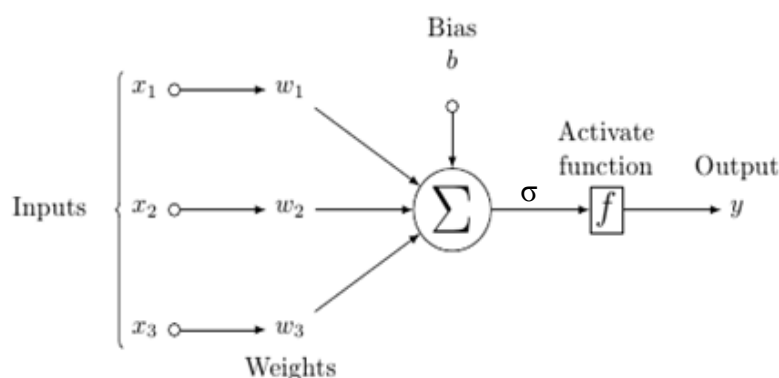
Η λειτουργία ενός νευρωνικού δικτύου, όπως βλέπουμε στην *Εικόνα 6*, δέχεται κάποιες τιμές εισόδου x_1, x_2, x_3 όπου αντιστοιχούν σε συνεχείς μεταβλητές. Αυτές οι τιμές μεταβάλλονται από μια τιμή βάρους w (weight) με μεταβλητές τις w_1, w_2, w_3 . Η τιμή του βάρους μπορεί να είναι θετική ή αρνητική. Το σώμα του τεχνητού

νευρώνα χωρίζεται σε δυο μέρη, τον *αθροιστή* (Σ) που προκύπτει από το άθροισμα του γινομένου της τιμής εισόδου με την τιμή του βάρους και την *συνάρτηση ενεργοποίησης* (activation function) η οποία διαμορφώνει την τελική τιμή του σήματος εξόδου y (output). Θεωρούμε επίσης ότι εκτός από τις τιμές εισόδου με τα αντίστοιχα βάρη, ο νευρώνας έχει και ένα βάρος w_0 το οποίο ονομάζεται *πόλωση* (Bias) ή *παράγοντας προδιάθεσης* του νευρώνα. Η διαφορά εδώ είναι ότι τα συγκεκριμένα βάρη είναι ότι επιδρά συνεχώς σε μια τιμή εισόδου ίση με 1, αυτά τα βάρη χρησιμοποιούνται για να καθορίσουν την δύναμη της θέσης της συνάρτησης ενεργοποίησης στο καρτεσιανό επίπεδο x - y (Βλαχάβας, και συν., 2006).

$$\sigma = \sum_{i=1}^3 xw + d * 1$$

Όπου x : τιμές εισόδου

w : συναπτικά βάρη



Εικόνα 6 Λειτουργία ενός νευρωνικού δικτύου
[Πηγή: <http://tex.stackexchange.com/>]

Οι κυριότερες συναρτήσεις ενεργοποίησης για τις τιμές εξόδου $y=f(\sigma)$ είναι (Κίνγκ, 2004) :

1. Γραμμικός νευρώνας (linear neuron) με την σχέση :

$$f(\sigma) = \sigma$$

2. Δυαδικός νευρώνας (binary) με τη σχέση κατωφλίου:

$$f(\sigma) = \begin{cases} 0 & \text{αν } \sigma < 0 \\ 1 & \text{αν } \sigma \geq 0 \end{cases} \text{ και}$$

3. Σιγμοειδής νευρώνας (sigmoidal) με τη λογιστική συνάρτηση:

$$f(\sigma) = \frac{1}{1+e^{-\sigma}} \in [0,1]$$

4. Νευρώνας υπερβολικής εφαπτομένης (hyperbolic tangent) με τη συνάρτηση:

$$f(\sigma) = \frac{1-e^{-\sigma}}{1+e^{-\sigma}} \in [-1,1]$$

2.8.3 Δίκτυα εμπρόθετης τροφοδότησης

Η δομή ενός νευρωνικού δικτύου εμπρόθετης τροφοδότησης αποτελείται από τουλάχιστον τρία επίπεδα (Günthe, et al., 2010) (Haykin, 1999), όπως φαίνεται στην Εικόνα 7:

1. Νευρώνες εισόδου

Τροφοδοτούν το δίκτυο με τιμές από το παρελθόν οι οποίες επεξεργάζονται από τους κρυφούς νευρώνες του επόμενου επιπέδου.

2. Κρυφοί Νευρώνες

Περιέχουν αρκετές σύνθετες λειτουργίες οι οποίες δημιουργούν προβλέψεις και είναι συχνά κρυφοί από τον χρήστη. Οι κόμβοι του επιπέδου αυτού αντιπροσωπεύουν μαθηματικές συναρτήσεις (συναρτήσεις ενεργοποίησης) οι οποίες τροποποιούν τα δεδομένα εισόδου. Το άθροισμα των γινομένων το οποίο έχει υπολογιστεί από το άθροισμα του συνόλου των τιμών του προηγούμενου επιπέδου επί την ισχύ του σημείου επικοινωνίας των αντίστοιχων νευρώνων (συναπτικό βάρος), τοποθετείται στην συνάρτηση ενεργοποίησης και αποτελεί την έξοδο του κρυφού νευρώνα και την είσοδο στον επόμενο.

Η συνάρτηση ενεργοποίησης είναι ως εξής:

$$y_k = f\left(\sum_{i=0}^N x_{ki} W_{ki}\right)$$

Όπου k , ο αντίστοιχος νευρώνας και y_k η έξοδός του.

Όπου x_{ki} , η i -οστή είσοδος του k νευρώνα.

Όπου w_{ki} , το i -οστό συναπτικό βάρος του k νευρώνα.

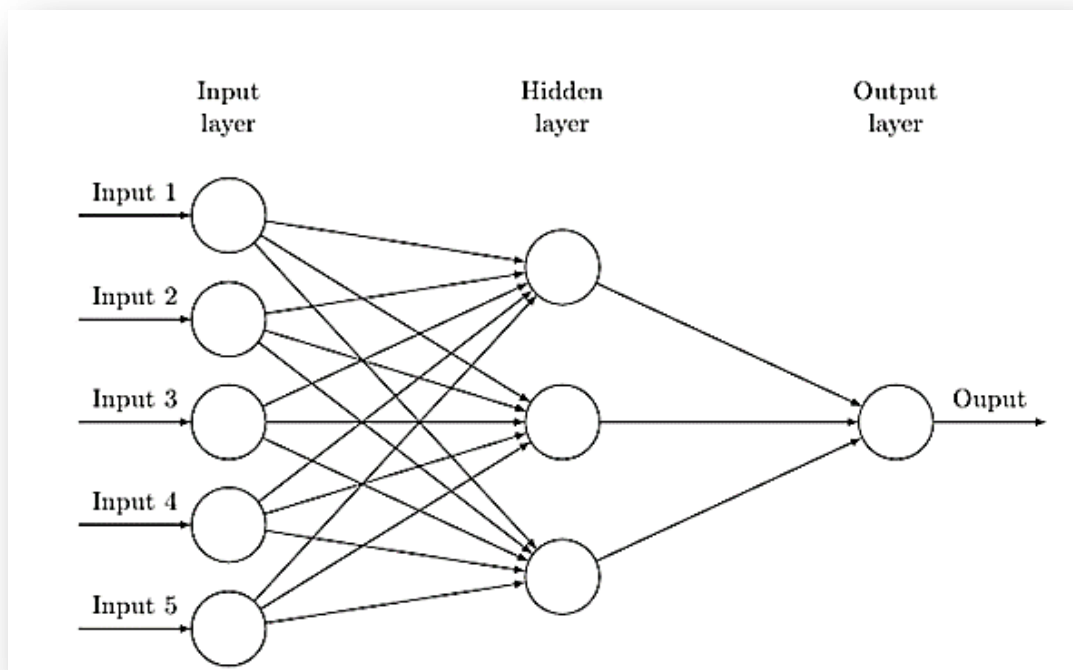
Όπου f η συνάρτηση ενεργοποίησης του k νευρώνα.

3. Νευρώνες εξόδου

Συλλέγει τις εξόδους που έγιναν στους κρυφούς νευρώνες και παράγει το τελικό αποτέλεσμα, δηλαδή την τελική πρόβλεψη.

Μία άλλη ιδέα η οποία εμπνεύστηκε από τον άνθρωπο είναι η εξής:

Ένας νευρώνας, μπορεί να μην χρησιμοποιηθεί ακόμα και καθόλου κατά τη διαδικασία της πρόβλεψης, δηλαδή να μην ενεργοποιηθεί ποτέ. Για να ισχύσει αυτή η υπόθεση, θα πρέπει το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα να έχει τιμή μικρότερη από το κατώφλι (bias), δηλαδή το συναπτικό βάρος w_{k0} του k -οστού νευρώνα, ($w_{k0} < 1$). Σε περίπτωση που η τιμή αυτή είναι μεγαλύτερη ή ίση της μονάδας, τότε ο νευρώνας ενεργοποιείται κανονικά. Να σημειωθεί ότι η τιμή εισόδου της πόλωσης (bias) είναι πάντα η μονάδα.



Εικόνα 7 Λειτουργία δικτύου εμπρόθετης τροφοδότησης

[Πηγή: tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network]

2.8.4 Αλγόριθμοι εκπαίδευσης τεχνητών νευρωνικών δικτύων

Ιδιαίτερη θέση στην εκμάθηση νευρωνικών δικτύων έχει ο αλγόριθμος μάθησης Ανάστροφης Διάδοσης (Back Propagation ή BP). Στόχος του είναι να μειώνεται το τετραγωνικό σφάλμα με την εξέλιξη του αλγορίθμου. Ένας τέτοιος αλγόριθμος αποτελείται από πολλά στρώματα νευρώνων. Έχει αποδειχθεί ότι τα τρία στρώματα, δηλαδή το στρώμα εισόδου, το ενδιάμεσο (κρυφό) και το στρώμα εξόδου επαρκούν για τις περισσότερες των περιπτώσεων. Όμως δεν είναι γνωστός ο αριθμός των νευρώνων που απαιτείται σε κάθε στρώμα άρα συνεπάγεται ο πειραματισμός σε αυτά. Για να γίνεται η μάθηση όσο πιο γρήγορα πρέπει να υπάρχει στον έλεγχο απλότητα και συνεπώς η ελαχιστοποίηση στρωμάτων και νευρώνων. Στην πράξη το κρυφό επίπεδο συχνά μπορεί να παραλειφθεί. Το πρόβλημα της μάθησης πολλαπλών στρωμάτων ουσιαστικά είναι το πρόβλημα συστηματικής εξεύρεσης των βαρών των νευρώνων στα διάφορα στρώματα. Επιπρόσθετα, είναι παραλλαγή του αλγορίθμου Widrow-Hoff με την διαφορά ότι τα διανύσματα στην επανάληψη γίνονται μήτρες, οι διαστάσεις των οποίων αλληλεξάρτιονται από τα στρώματα και τους νευρώνες σε κάθε στρώμα. Για παράδειγμα ένα τεχνητό νευρωνικό δίκτυο με χαρακτηρισμό 30-50-10, έχει στο στρώμα εισόδου 30 νευρώνες, στο κρυφό 50 νευρώνες και στην έξοδο 10 νευρώνες. Στην περίπτωση των T.N.Δ. το σφάλμα είναι η διαφορά μεταξύ της επιθυμητής και της πραγματικής εξόδου του δικτύου, δηλαδή

$$e = d - f(\sigma_3)$$

και συνεπώς πρέπει να οριστεί η παράγωγος του τετραγωνικού σφάλματος ως προς τα βάρη. (Κίνγκ, 2004)

Βήματα Backpropagation

Η συνολική αυτή διαδικασία αποτελεί έναν κύκλο (Haykin, 1999).

Βήμα 1^ο : Είσοδος Τα αρχικά βάρη και τα κατώφλια επιλέγονται χρησιμοποιώντας μικρές θετικές και τυχαίες αρχικές τιμές.

Βήμα 2^ο : Προς τα εμπρός πέρασμα: Το TND αρχικά λαμβάνει ως είσοδο τις τιμές των χαρακτηριστικών του πρώτου προτύπου εκπαίδευσης και με την διαδικασία που αναλύθηκε προηγουμένως παράγει την έξοδο.

Βήμα 3^ο : Προς τα πίσω πέρασμα: Γίνεται προσαρμογή των βαρών αρχίζοντας από τους νευρώνες εξόδου και προχωρώντας ανάστροφα προς το στρώμα εισόδου. Την τιμή εξόδου την συγκρίνει με την τιμή του στόχου. Εάν δεν υπάρχει διαφορά μεταξύ των δύο δεν συμβαίνει τίποτα και προχωράμε στο επόμενο πρότυπο. Εάν υπάρχει διαφορά τότε αλλάζουμε τις τιμές των w με τέτοιο τρόπο ώστε η διαφορά αυτή να ελαττωθεί.

Βήμα 4^ο : Μετά το τέλος ενός κύκλου επαναλαμβάνουμε την διαδικασία για πολλούς κύκλους από το βήμα 2^ο, έως ότου το σφάλμα φθάσει να είναι αρκετά μικρό.

Συμπερασματικά για τον αλγόριθμο εποπτευόμενης μάθησης (supervised learning) ανάστροφης διαδρομής αν και είναι διαδεδομένος η χρήση του για μεγάλα δίκτυα είναι προβληματική. Μια παραλλαγή είναι ο αλγόριθμος μάθησης με ορμή (momentum) όπου έχουμε έναν ακόμη όρο ανάλογος των βαρών στην περίπτωση μας με αποτέλεσμα την επιτάχυνση του αλγορίθμου (Haykin, 1999) (Βλαχάβας, και συν., 2006) (Hagan , et al., 1996).

2.9 Επιλογή μοντέλου πρόβλεψης

Για να γίνει σωστά η επιλογή μοντέλου πρόβλεψης θα πρέπει να τεθούν σε προτεραιότητα κριτήρια βάση της σημαντικότητάς τους στην πράξη. Τα κριτήρια αυτά, ταξινομημένα, είναι τα εξής (Chambers, et al., 1971):

1. Ακρίβεια

Στην πλειοψηφία των περιπτώσεων που απαιτείται πρόβλεψη, η ακρίβεια αντιμετωπίζεται ως κύριο κριτήριο για την επιλογή μοντέλου πρόβλεψης. Η ακρίβεια δεν είναι σημαντική μόνο εννοιολογικά, αλλά επειδή και άλλα κριτήρια εξαρτώνται από αυτή με σκοπό μια επιτυχημένη πρόβλεψη.

2. Συμπεριφορά δεδομένων

Διαφορετικά μοντέλα πρόβλεψης μπορούν να διαχειριστούν συγκεκριμένα πρότυπα δεδομένων, για τον λόγο αυτόν η επιλογή ενός κατάλληλου προτύπου συμπεριφοράς δεδομένων είναι σημαντική. Υπάρχουν, σαφώς, μοντέλα που έχουν την ικανότητα διαχείρισης μεγάλο φάσμα αλλά είναι συνήθως μεγάλες σε κόστος και πιο δύσκολες στην εφαρμογή τους. Το κριτήριο της συμπεριφοράς δεδομένων μπορεί να διαχωριστεί σε δύο υποκατηγορίες, οι οποίες είναι:

- a) Επαναλαμβανόμενα στον χρόνο πρότυπα (Περιοδικά)
- b) Πρότυπα που παρουσιάζουν σημεία καμψής και δεν επαναλαμβάνονται σε σταθερό διάστημα χρόνου

3. Χρονικός ορίζοντας

Κύριο χαρακτηριστικό επιλογής ενός μοντέλου πρόβλεψης είναι ο χρονικός ορίζοντας για τον οποίο θα παρουσιάζονται τα αποτελέσματα της πρόβλεψης. Είναι συνδεδεμένος με την συμπεριφορά των δεδομένων καθώς διαφορετικοί χρονικοί ορίζοντες περιλαμβάνουν διαφορετικά πρότυπα σε σχέση με τον αριθμό των δεδομένων που πρέπει να προβλεφθεί.

4. Κόστος

Το κόστος της πρόβλεψης εξαρτάται από το αντίστοιχο μοντέλο που υλοποιείται, την πολυπλοκότητά του, όσο ακόμα από τις απαιτήσεις που έχει σε θέμα δεδομένων και υπολογιστικής δύναμης και τον αριθμό των αντικειμένων που πρέπει να προβλεφθούν.

5. Ευκολία στην χρήση

Το κριτήριο αυτό εφαρμόζει όλα όσα χρειάζονται για την επιλογή μοντέλου πρόβλεψης τα οποία δεν έχουν καλυφθεί από τα προηγούμενα κριτήρια. Μερικά από αυτά είναι η πολυπλοκότητα του μοντέλου, οι χρονοσειρές των προβλέψεων, το επίπεδο γνώσης που απαιτείται για την εφαρμογή του μοντέλου και η ευκολία με την οποία ο τελικός χρήστης της πρόβλεψης χρησιμοποιεί τα δεδομένα.

2.10 Σφάλματα

Για την καλύτερη αξιολόγηση και σύγκριση των αποκλίσεων των αποτελεσμάτων ορίζονται τα παρακάτω σφάλματα:

- Σφάλμα

$$e_i = Y_i - F_i$$

- Μέσο σφάλμα (Mean Error)

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_t - F_t)$$

- Μέσο απόλυτο σφάλμα (Mean Absolut Error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_t - F_t|$$

- Μέσο τετραγωνικό σφάλμα (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_t - F_t)^2$$

Όπου

Y_i , η πραγματική τιμή την χρονική στιγμή t

F_i , η τιμή της πρόβλεψης την χρονική στιγμή t

2.11 Μελέτες περίπτωσης πρόβλεψης στην υγειονομική περίθαλψη¹

¹ Πηγή: (Bari, et al., 2014)

2.11.1 Η δημιουργία μοντέλου πρόβλεψης από την Google

Σήμερα, όλοι χρησιμοποιούν το Google για να σχεδόν οτιδήποτε, σχεδόν κάθε ώρα της ημέρας. Οι αναζητήσεις ποικίλουν ανάλογα με το ενδιαφέρον του χρήστη, όπως για παράδειγμα να μάθουν πληροφορίες για τον καλοκαιρινό τους προορισμό ή να αναζητήσουν πληροφορίες για κάποιο άρθρο που μόλις διάβασαν, ακόμα και να βρουν τα συμπτώματα μιας ασθένειας που ίσως απέκτησαν.

Μετά από πολλές αναζητήσεις, ερευνητές της Google, κατά την έρευνα του J. Ginsberg, κατέληξαν στο συμπέρασμα ότι συγκεκριμένοι όροι αναζήτησης δείχνουν ότι κάποια αρρώστια είναι σε έξαρση σε πραγματικό χρόνο. Στην συγκεκριμένη περίπτωση η κοινή γρίπη.

Από τα ιατρικά δεδομένα τα οποία είχε στη διάθεσή του ο Ginsberg και οι συνεργάτες του, προέκυψε μια δυνατή σύνδεση αυτών που έκαναν αναζητούσαν όρους που έχουν σχέση με τη γρίπη με αυτούς που όντως είχαν γρίπη εκείνο το χρονικό διάστημα. Αυτό δείχνει ότι υπάρχει μια συγκεκριμένη συμπεριφορά αυτών των δυο κατηγοριών η οποία είναι κοινή.

Μόλις τα αποτελέσματα της έρευνας δημοσιεύθηκαν η Google σχολίασε: *“Προφανώς, υπάρχουν περισσότερες αναζητήσεις που έχουν να κάνουν με αλλεργίες κατά τη διάρκεια έξαρσης αλλεργιών και περισσότερες αναζητήσεις που αφορούν εγκαύματα κατά τη διάρκεια της καλοκαιρινής περιόδου”*.

Επόμενο βήμα της Google είναι να πάρει τα συμπεράσματα της έρευνας και να δημιουργήσει τάσεις. Το αποτέλεσμα είναι η κατασκευή αξιόπιστων μοντέλων που μπορούν να προβλέψουν πραγματικά γεγονότα όπως κρούσματα σε συγκεκριμένες περιοχές του κόσμου σε πραγματικό χρόνο.

Βασισμένη σε εκπαιδευμένα πακέτα δεδομένων που περιέλαβε μετά από πέντε χρόνια αναζητήσεων και ιατρικά δεδομένα της γρίπης, η Google δημιούργησε ένα μοντέλο για την παγκόσμια παρακολούθηση και πρόβλεψη της κοινής γρίπης.

2.11.2 Πρόβλεψη επιβιωσιμότητας καρκίνου

Η ανάλυση της πρόβλεψης στην υγειονομική περίθαλψη περιλαμβάνει τη χρήση αλγορίθμων για την πρόβλεψη του ρυθμού επιβιωσιμότητας των ασθενών με καρκίνο του μαστού. Σύμφωνα με την Αμερικανική Έρευνα Παρακολούθησης για Άτομα με Καρκίνο, για την χρονιά 2013, περίπου 232.340 νέες περιπτώσεις ασθενών με καρκίνο αναμένονταν να εμφανιστούν και να διαγνωστούν στον γυναικείο πληθυσμό των ΗΠΑ με 39.620 θανάτους να προβλέπονται από καρκίνο του μαστού την ίδια χρονιά.

Στο πανεπιστήμιο George Washington των ΗΠΑ, ο καθηγητής Abdelghani Bellaachia και η ομάδα του, μέσα από έρευνα με τίτλο “Predicting Breast Cancer Survivability Using Data Mining Techniques”, συσχέτισαν την ανάλυση πρόβλεψης και το ανθρώπινο συμφέρον.

Για την δημιουργία ενός μοντέλου πρόβλεψης χρησιμοποιήθηκαν δεδομένα για την επιβίωση των ασθενών από τον καρκίνο του μαστού προερχόμενα από το Εθνικό Ινστιτούτο Παρακολούθησης Επιδημιών των ΗΠΑ. Τα δεδομένα ήταν κυρίως ιστορικά και εμπεριείχαν τον χρόνο επιβιωσιμότητας, σε ποιο στάδιο ήταν ο καρκίνος και η αιτία του θανάτου.

Το μοντέλο πρόβλεψης που χρησιμοποιήθηκε ήταν βασισμένο σε αυτό της ταξινόμησης (classification) και οι αλγόριθμοι που εφαρμόστηκαν, οι Naïve Bayes, Νευρωνικά Δίκτυα και Δέντρα Αποφάσεων. Αυτοί οι 3 συγκρίθηκαν μεταξύ τους, από τους ερευνητές, για την αποδοτικότητά τους και επέλεξαν τον καλύτερο για το συγκεκριμένο μοντέλο. Σε αυτή τη μελέτη περίπτωσης τα Δέντρα Αποφάσεων απόδωσαν καλύτερα, ταξινομώντας ακριβέστερα τα δεδομένα δοκιμής του μοντέλου (test data).

Από την στιγμή που η πρόβλεψη ήταν τύπου ταξινόμησης (classification-based), το μοντέλο έπρεπε να εκπαιδευθεί χρησιμοποιώντας συστάδες (κατηγορίες δεδομένων). Το βήμα πριν την ταξινόμηση απαιτεί την οργάνωση των δεδομένων σε συστάδες (κατηγορίες): *επιζώντες* και *μη επιζώντες*. Έτσι το σύνολο των δεδομένων αντιπροσωπεύει δύο κατηγορίες ασθενών με κάθε σειρά να αντιστοιχεί σε έναν ασθενή με καρκίνο του μαστού.

Τα δεδομένα για την εκπαίδευση (training set) περιέχουν τα δεδομένα των ασθενών που χρησιμοποιήθηκαν για τη δημιουργία αυτού του μοντέλου, και τα υπόλοιπα χρησιμοποιήθηκαν και να δοκιμαστεί το μοντέλο (test data).

3^ο Κεφάλαιο

Εφαρμογή Μεθόδων Πρόβλεψης²

Σε αυτό το κεφάλαιο γίνεται χρήση του στατιστικού πακέτου R, και με τη δημιουργία κατάλληλου κώδικα, για κάθε τεχνική πρόβλεψης, παρουσιάζουμε αποτελέσματα και συμπεράσματα για τα δεδομένα που έχουμε επιλέξει αντίστοιχα. Στόχος, ακόμα είναι η παρουσίαση στατιστικών περιγραφικών στοιχείων όπως για παράδειγμα τιμές σφαλμάτων έτσι ώστε να γίνει μια καλύτερη αξιολόγηση των αποτελεσμάτων την πρόβλεψης.

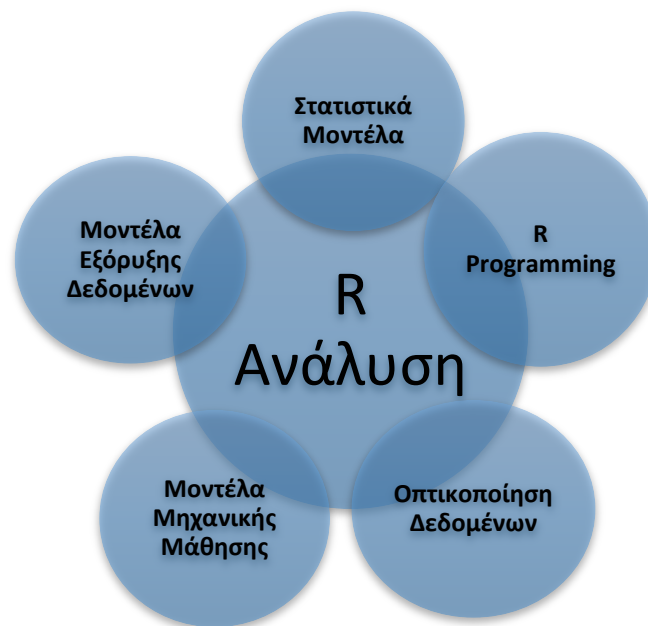
3.1 Το Στατιστικό Πακέτο R

Το λογισμικό πακέτο ανοιχτού κώδικα R παρέχει ευρύ φάσμα τεχνικών στατιστικής ανάλυσης (όπως γραμμικά και μη γραμμικά μοντέλα, στατιστικοί έλεγχοι, ανάλυση χρονοσειρών), μηχανικής μάθησης, εξόρυξης δεδομένων (όπως κατηγοριοποίησης, συσταδοποίησης κ.α.) καθώς και δημιουργίας γραφικών. Χρησιμεύει στην επεξηγηματική ανάλυση δεδομένων καθώς και στην εφαρμογή στατιστικών μοντέλων. Μπορεί να χρησιμοποιηθεί είτε με κατευθείαν εντολές είτε με προγράμματα τα οποία μπορούν να αναπτυχθούν και να δοθούν για εκτέλεση. Εφαρμόζει μια διάλεκτο της γλώσσας S η οποία είναι διερμηνέας γλώσσα προγραμματισμού και μεγάλο πλεονέκτημα αυτών είναι ότι επιτρέπουν σταδιακή ανάπτυξη του κώδικα. Με άλλα λόγια, μια συνάρτηση μπορεί να δημιουργηθεί, να εκτελεστεί και μετά να δημιουργηθεί μια καινούρια συνάρτηση η οποία καλεί την προηγούμενη κ.ο.κ. Η αλληλεπίδραση με την R επιτυγχάνεται πληκτρολογώντας

² Οι κώδικες που χρησιμοποιήθηκαν για κάθε μέθοδο, υπάρχουν ολοκληρωμένοι στο Παράρτημα.

εκφράσεις, τις οποίες ο διερμηνέας αξιολογεί και τις εκτελεί (Φωκιανός, και συν., 2010).

Για παράδειγμα όταν εκτελούμε την R, οι μεταβλητές, τα δεδομένα, οι λειτουργίες, τα αποτελέσματα κ.λπ. αποθηκεύονται στη μνήμη του υπολογιστή με τη μορφή των αντικειμένων και έχουν ένα όνομα. Ο χρήστης μπορεί να κάνει ενέργειες σε αυτά τα αντικείμενα όπως (αριθμητική, λογική, συγκριτική κτλ.) αλλά και στις συναρτήσεις που είναι και αυτές αντικείμενα (Paradis, 2005).



3.2 Προφίλ WorldBank

Η Παγκόσμια Τράπεζα είναι ένα διεθνές χρηματοπιστωτικό ίδρυμα και δημιουργήθηκε το 1944 από τους John Maynard Keynes και τον Harry Dexter White συμπεριλαμβανομένου και του Διεθνούς Νομισματικού Ταμείου (ΔΝΤ) και αποτελείται από 188 χώρες μέλη. Η Παγκόσμια Τράπεζα βρίσκεται στην Washington, D.C. Αποτελείται από δύο όργανα: τη Διεθνή Τράπεζα Ανασυγκρότησης και Ανάπτυξης (BIRD) και ο Διεθνής Οργανισμός Ανάπτυξης (IDA). Η Παγκόσμια τράπεζα παρέχει δάνεια σε αναπτυσσόμενες χώρες γιατί επίσημος στόχος της Παγκόσμιας Τράπεζας είναι η μείωση της φτώχειας. Ακόμα για αυτό το σκοπό παρέχει ελεύθερη και ανοικτή πρόσβαση σε δεδομένα που αφορούν την ανάπτυξη σε χώρες σε όλο τον κόσμο. Ο κατάλογος των δεδομένων

είναι μια λίστα διαθέσιμων συνόλων δεδομένων της Παγκόσμιας Τράπεζας, συμπεριλαμβανομένων των βάσεων δεδομένων.



3.2.1 Διαδικασία εξόρυξης δεδομένων από WorldBank

Η WorldBank διαθέτει στον ιστότοπό της, σύστημα³ το οποίο αποτελείται από βάση δεδομένων με υγειονομικά δεδομένα καθώς και στατιστικά στοιχεία για κάθε χώρα του κόσμου. Μέσα από αυτό το σύστημα, επιλέξαμε 17 δείκτες (Κεφ. 1.5.2, 1.5.3) οι οποίοι αντιπροσωπεύουν, επιλεκτικά, όλες τις χώρες της Ευρώπης καθώς και μερικές μη Ευρωπαϊκής Ένωσης για να υπάρξει μια περαιτέρω σύγκριση.

Οι μη Ευρωπαϊκής Ένωσης Χώρες που επιλέχθηκαν είναι :

- Βοσνία και Ερζεγοβίνη
- Ισλανδία
- Κοσσυφοπέδιο
- Λιχτενστάιν
- Σκόπια
- Μολδαβία
- Μονακό
- Μαυροβούνιο
- Νορβηγία
- Ρωσική Ομοσπονδία
- Σαν Μαρίνο
- Σερβία
- Ελβετία
- Τουρκία
- Ουκρανία

³ <http://data.worldbank.org/topic/health>

Σε αυτό το σημείο να επισημάνουμε ότι τα δεδομένα τα οποία παρέχει η WorldBank περιέχουν μεγάλο αριθμό ελλিপών (κενών) τιμών, με αποτέλεσμα να μην είναι δυνατή η διαχείρισή τους με τις γνωστές μεθόδους διαχείρισης. Για αυτό τον λόγο, τα διαγράμματα διασποράς των προβλέψεων Απλής και Πολλαπλής Γραμμικής παλινδρόμησης όσο και αυτά με ARIMA, για μελλοντικά χρόνια, παρουσιάζουν απότομες αυξομειώσεις, πράγμα που το οποίο, όμως, δεν αποτελεί εμπόδιο στην διαδικασία της πρόβλεψης.

3.3 Περιγραφική Στατιστική

Με την βοήθεια του πακέτου *pastecs* και με την βοήθεια της συνάρτησης `stat.desc()` μπορούμε να δούμε έναν πίνακα των περιγραφικών στατιστικών, έχοντας μετατρέψει το δείγμα μας σε πίνακα δεδομένων για έναν κατάλογο μεταβλητών όπως είναι οι χώρες της Ευρωπαϊκής Ένωσης και μη Ευρωπαϊκής Ένωσης. Ξεκινώντας με τα βασικά στατιστικά στοιχεία έχουμε τον αριθμό των τιμών (`nbr.val`), τον αριθμό των μηδενικών τιμών (`nbr.null`), τον αριθμό των τιμών που λείπουν (`nbr.na`), την ελάχιστη τιμή (`min`), τη μέγιστη τιμή (`max`), το εύρος και το άθροισμα όλων των τιμών.

Στην συνέχεια εμφανίζεται: η μέση τιμή ή μέσος όρος (`mean`), ο διάμεσος (`median`), το τυπικό σφάλμα επί του μέσου (`SE.mean`), το διάστημα εμπιστοσύνης της μέσης τιμής (`CI.mean`) στο p επίπεδο σημαντικότητας (p είναι η πιθανότητα που χρησιμοποιείται για τον υπολογισμό του διαστήματος εμπιστοσύνης (`CI.mean`) με τιμή $p = 0,95$), η διακύμανση (`var`), η τυπική απόκλιση (`std.dev`) και ο συντελεστής διακύμανσης (`coef.var`) όπου ορίζεται ως η τυπική απόκλιση διαιρούμενη με τη μέση. Τέλος μπορούμε να δούμε τα ίδια στατιστικά για έναν κατάλογο μεταβλητών όπως οι δείκτες που χρησιμοποιήσαμε.

row.names	Hospital beds, per 1,000 people	Analogia diathwn ana klinh	Analogia diathwn ana klinh	Analogia storktefen ana klinh	Analogia storktefen ana klinh	Nurses and midwives, per 1,000 people	Physicians, per 1,000 people	Health expenditure total, % of GDP	Health expenditure on health, % of total expenditure on health	External resources for health, % of total expenditure on health	Birth rate, crude, per 1,000 people	Death rate, crude, per 1,000 people	Life expectancy at birth, total, years	Number of infant deaths	Deaths, neonatal, this month	Mortality rate, infant, per 1,000 live births	Mortality rate, neonatal, per 1,000 live births	Number of neonatal deaths	Fertility rate, total, per woman
nbr.val	2.010.000.000	20.100.000.000	20.100.000.000	23.300.000.000	51.000.000	2.330.000.000	2.470.000.000	2.010.000.000	3,89E+09	3,89E+09	3,82E+08	3,83E+08	389.000.000	3,83E+09	3.120.000.000	312000000	3620000000		
nbr.null	0.0000000	2.200.000.000	16600000000	18300000000	0.000000	0.0000000	0.0000000	710000000	0.0000000	0.0000000	0.000000e+00	2,00E+06	129000000	0.0000000	0.0000000	19000000	0.00000000		
nbr.na	2.490.000.000	24.900.000.000	24.900.000.000	21.700.000.000	399.000.000	2.170.000.000	2.030.000.000	2.490.000.000	6,10E+08	6,10E+08	6,80E+07	6,70E+07	61.000.000	6,70E+08	1.380.000.000	138000000	8800000000		
min	209999999	0.00000000	0.00000000	0.00000000	0.612000	0.7000000	30.240.608	0.0000000	7,70E+07	57000000	61143200	0.000000e+00	0.000000e+00	16000000	0.9000000	0.000000e+00	108500000		
max	216.786.003	661.312.863	1.333.519.581	448.590.334	95.480.000	473.500.000	145.599.169	206.460.286	3,35E+08	1,66E+08	8,33E+07	1,11E+11	2.459.030.000.000	7,44E+08	312.000.000	4263600000	482000000		
range	195.786.004	661.312.863	1.333.519.581	448.590.334	94.868.000	466.500.000	115.358.562	206.460.286	2,58E+08	1,09E+08	2,22E+07	1,11E+11	2.459.030.000.000	7,28E+08	303.000.000	4263600000	373500000		
sum	14.784.878.981	8.441.322.524	7.667.122.342	12.848.422.124	530.634.100	7.181.476.012	18.774.534.092	2.333.387.126	5,54E+10	3,67E+10	2,80E+10	3,04E+12	82.727.810.000.000	5,42E+10	23.080.000.000	1.201.05E+12	68324100000		
median	55500002	0.36640211	0.00000000	0.00000000	7.599.000	29780000	76.894.521	0.2456623	1,28E+08	8,70E+07	7,36E+07	4,49E+08	11.879.700.000	1,04E+08	56.500.000	244500000	178400000		
mean	73556612	0.41996629	0.38144887	0.55143443	10.404.590	30821785	76.010.259	11.608.891	1,42E+08	9,43E+07	7,33E+07	7,94E+09	212.667.900.000	1,42E+08	73.974.359	3849516000	188740608		
SE.mean	0.3059666	0.03532025	0.09066359	0.07605818	1843907	0.2038319	0.1430152	0.1786413	0.2480179	0.1470770	2,75E+05	9,34E+08	2530388000	0.6300582	0.3216902	482263700	0.03091583		
CI.mean	0.6033343	0.06964787	0.17877920	0.14985302	3703596	0.4015982	0.2816905	0.3522622	0.4876273	0.2891676	5,40E+05	1,84E+09	4974988000	1,24E+07	0.6329645	948912300	0.06079774		
var	188167215	0.25075155	165.219.736	134.786.945	173.399.583	96.805.578	50.519.780	64.144.583	2,39E+08	8,41E+07	2,88E+07	3,34E+14	249.071.400.000.000.000	1,52E+09	322871960	7.25644E+13	0.34599544		
std.dev	43378245	0.50075098	128.537.829	116.097.780	13.168.128	31.113.595	22.476.606	25.326.781	4,89E+07	2,90E+07	5,37E+06	1,83E+10	499.070.500.000	1,23E+08	56821823	8518476000	0.58821377		
coef.var	0.5897260	119.235.993	336972629	210537780	1265608	10.094.676	0.2957049	21816710	0.3437018	0.3077577	7,32E+04	2,30E+06	2346713	0.8708900	0.7681286	2212869	0.31165194		

Πίνακας 1 Πίνακας Περιγραφικών Στατιστικών Χωρών Μη Ευρωπαϊκής Ένωσης

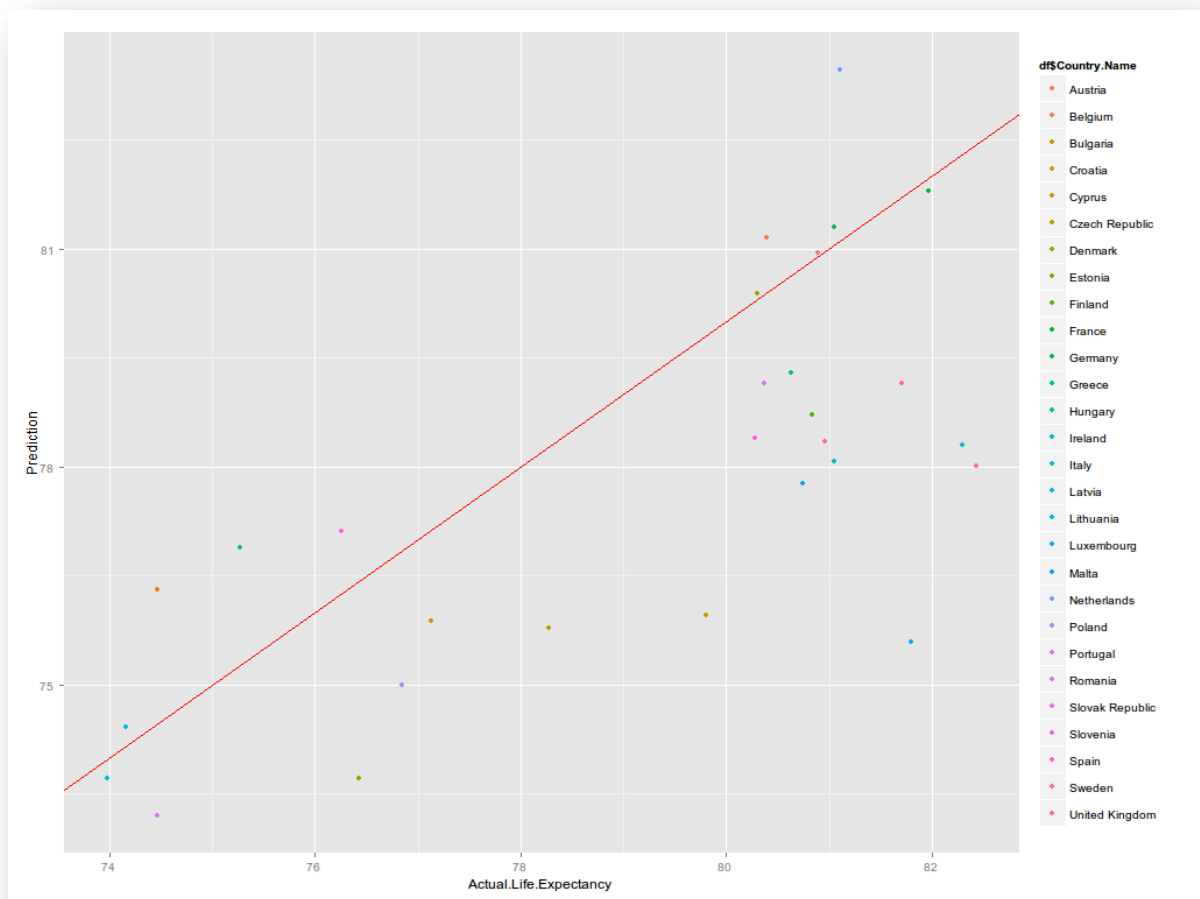
row.names	Hospital beds, per 1,000 people	Analogia diathwn ana klinh	Analogia storktefen ana klinh	Analogia storktefen ana klinh	Nurses and midwives, per 1,000 people	Physicians, per 1,000 people	Health expenditure total, % of GDP	External resources for health, % of total expenditure on health	Birth rate, crude, per 1,000 people	Death rate, crude, per 1,000 people	Life expectancy at birth, total, years	Number of infant deaths	Deaths, neonatal, this month	Mortality rate, infant, per 1,000 live births	Mortality rate, neonatal, per 1,000 live births	Number of neonatal deaths	Fertility rate, total, per woman
nbr.val	6,20E+08	62.000.000.000	62.000.000.000	61.100.000.000	1.360.000.000	6,11E+08	5,32E+08	3,14E+10	8,40E+08	8,40E+08	8,40E+08	8,40E+08	8,40E+08	8.400.000.000	6.720.000.000	6,72E+08	8,40E+08
nbr.null	0.000000e+00	12.400.000.000	51.500.000.000	48.800.000.000	0.0000000	0.000000e+00	0.000000e+00	1,67E+10	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1,68E+08	0.0000000	0.0000000	3.000000e+00	0.000000e+00
nbr.na	2,20E+08	22.000.000.000	22.000.000.000	22.900.000.000	7.040.000.000	2,29E+08	3,08E+08	5,26E+10	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0000000	1.680.000.000	1,68E+08	0.000000e+00
min	2,70E+06	0.00000000	0.00000000	0.00000000	0.1750000	1,38E+06	3,22E+06	0.00000000	7,60E+06	6,10E+06	6,57E+07	1,00E+07	0.000000e+00	16.000.000	0.9000000	6000000	1,09E+06
max	1,46E+07	145.407.407	497.492.064	824.647.160	239.560.000	6,17E+06	1,29E+07	4,99E+08	2,05E+07	1,66E+07	8,25E+07	1,40E+10	5,55E+11	326.000.000	168.000.000	7,47E+09	2,58E+06
range	1,19E+07	145.407.407	497.492.064	824.647.160	237.810.000	4,79E+06	9,67E+06	4,99E+08	1,29E+07	1,05E+07	1,68E+07	1,40E+10	5,55E+11	310.000.000	159.000.000	7,46E+09	1,49E+06
sum	4,29E+09	23.094.676.011	15.431.501.685	30.840.774.799	10.889.890.000	1,77E+09	4,24E+09	9,86E+09	9,70E+09	8,67E+09	6,36E+10	1,28E+12	5,12E+13	69.525.000.000	32.537.000.000	5,54E+11	1,35E+09
median	6,60E+06	0.36306342	0.00000000	0.00000000	71.300.000	2,90E+06	7,85E+06	0.00000000	1,12E+07	1,00E+07	7,61E+07	5,99E+08	2,21E+10	65.000.000	38.000.000	3,36E+08	1,56E+06
mean	6,92E+06	0.37249477	0.24889519	0.50475900	80.072.721	2,90E+06	7,97E+06	0.31393695	1,15E+07	1,03E+07	7,57E+07	7,40E+07	6,09E+10	82.767.857	48.418.155	8,24E+08	1,61E+06
SE.mean	9,62E+04	0.01083068	0.02761476	0.04743453	0.3428135	2,96E+04	7,69E+04	0.04398652	7,62E+04	6,45E+04	1,21E+05	1,40E+07	3,08E+09	0.1863587	0.1193553	4,06E+07	9,91E+03
CI.mean	1,89E+05	0.02126932	0.05422997	0.09315480	0.6779796	5,80E+04	1,51E+05	0.08654666	1,50E+05	1,27E+05	2,37E+05	1,45E+08	6,05E+09	0.3657841	0.2343549	7,98E+07	1,94E+04
var	5,73E+06	0.07272819	0.47279649	137.477.123	159.828.689	5,34E+05	3,14E+06	0.60753171	4,88E+06	3,49E+06	1,23E+07	4,60E+12	7,98E+15	291.728.454	95.731.073	1,11E+12	8,24E+04
std.dev	2,39E+06	0.26968165	0.68760198	117.250.639	39.978.580	7,31E+05	1,77E+06	0.77944321	2,21E+06	1,87E+06	3,50E+06	2,14E+09	8,93E+10	54.011.893	30.940.438	1,05E+09	2,87E+05
coef.var	3,46E+05	0.72398773	276.261.662	232.290.339	0.4992784	2,52E+05	2,22E+05	248.280.174	1,91E+05	1,81E+05	4,63E+04	1,41E+06	1,47E+06	0.6525709	0.6390256	1,28E+06	1,79E+05

Πίνακας 2 Πίνακας Περιγραφικών Στατιστικών Χωρών Ευρωπαϊκής Ένωσης

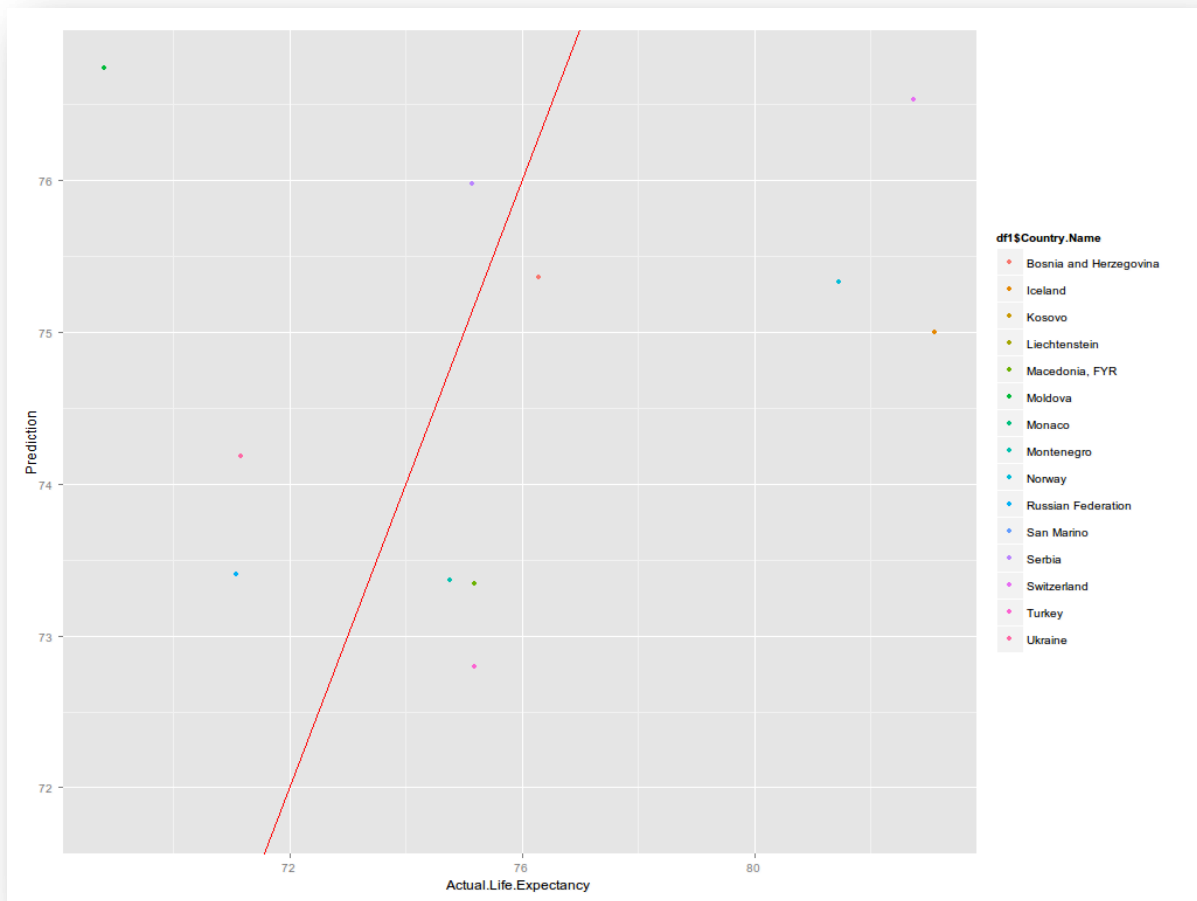
3.4 Απλή Γραμμική Παλινδρόμηση

Η απλούστερη μορφή της παλινδρόμησης είναι η Απλή Γραμμική Παλινδρόμηση και χρησιμοποιώντας την περιγραφική στατιστική μπορούμε να οργανώσουμε, να συνοψίσουμε και να παρουσιάσουμε τα στοιχεία μας με εύκολο και κατανοητό τρόπο. Τα στατιστικά στοιχεία είναι αυτά που χρησιμοποιούνται για να αντιπροσωπεύσουν κάθε σύνολο δεδομένων. Αποτελούν τη βάση για πολύ περίπλοκους υπολογισμούς και αναλύσεις με πολύ απλές μεθόδους.

Για την Απλή Γραμμική Παλινδρόμηση, χωρίσαμε τα δεδομένα σε Ευρωπαϊκής Ένωσης χώρες και μη Ευρωπαϊκής Ένωσης μαζί με τους δείκτες των συνολικών δαπανών και το προσδόκιμο ζωής στη γέννηση, με σκοπό να γίνει σύγκριση μεταξύ τους. Ως σύνολο εκπαίδευσης, χρησιμοποιήσαμε τα παραπάνω δεδομένα για τις χρονολογίες 1984 – 2012 και ως σύνολο ελέγχου τα αντίστοιχα για το έτος 2013.



Εικόνα 8 Σύγκριση πραγματικής και προβλεπόμενης τιμής ανά Ευρωπαϊκής Ένωσης χώρες (Απλή Γραμμική Παλινδρόμηση)



Εικόνα 9 Σύγκριση πραγματικής και προβλεπόμενης τιμής ανά μη Ευρωπαϊκής Ένωσης χώρες (Απλή Γραμμική Παλινδρόμηση)

Στις εικόνες 8 και 9 παρουσιάζονται, σε διαγράμματα διασποράς, η πραγματική τιμή με την προβλεπόμενη τιμή μαζί με την γραμμή παλινδρόμησης. Όπως βλέπουμε, οι τιμές απέχουν από την γραμμή παλινδρόμησης σε μεγάλο βαθμό, στην περίπτωση των ευρωπαϊκών χωρών αλλά και των μη, κάνοντας την μέθοδο στην συγκεκριμένη περίπτωση αναξιόπιστη. Το φαινόμενο αυτό οφείλεται στο ότι τα δεδομένα περιέχουν μεγάλο αριθμό κενών τιμών με αποτέλεσμα η μία ανεξάρτητη (Health Expenditure) μεταβλητή η οποία λαμβάνει μέρος, να μην είναι αρκετή για την πραγματοποίηση μιας αξιόπιστης πρόβλεψης. Στην επόμενη υποενότητα παρουσιάζονται, μαζί με άλλα στατιστικά στοιχεία, τα σφάλματα των αποτελεσμάτων της πρόβλεψης.

3.4.1 Ανάλυση του Μοντέλου Απλής Γραμμικής Παλινδρόμησης

Στη συγκεκριμένη υπόθεση ορίσαμε, στο βασικό μοντέλο, ως είσοδο τις συνολικές δαπάνες και ως έξοδο το προσδόκιμο ζωής.

```
fmla.eu <- as.formula(Life.Expectancy.Birth.Total ~ Health.Expenditure.Total)
s.regmodel.eu <- lm(fmla.eu, data = dset.eu.train)
```

```
fmla.noeu <- as.formula(Life.Expectancy.Birth.Total~Health.Expenditure.Total)
s.regmodel.noeu <- lm(fmla.noeu, data = dset.noeu.train)
```

Στον παρακάτω πίνακα βλέπουμε τις τιμές της με βάση το μοντέλο της απλής γραμμικής παλινδρόμησης για τις χώρες της Ευρωπαϊκής Ένωσης καθώς και τις μη Ευρωπαϊκής Ένωσης.

```
stat.desc(s.regmodel.eu$residuals) stat.desc(s.regmodel.noeu$residuals)
```

	Ευρωπαϊκής Ένωσης	Μη Ευρωπαϊκής Ένωσης
nbr.val	5.040000e+02	2.100000e+02
nbr.null	0.000000e+00	0.000000e+00
nbr.na	0.000000e+00	0.000000e+00
min	-7.384416e+00	-1.138780e+01
max	5.696650e+00	1.064894e+01
range	1.308107e+01	2.203674e+01
sum	2.099709e-14	-1.167122e-14
median	-6.515446e-02	1.136015e-01
mean	4.168843e-17	-5.561028e-17
SE.mean	1.052568e-01	3.536278e-01
CI.mean.0.95	2.067971e-01	6.971346e-01
var	5.583809e+00	2.626105e+01
std.dev	2.363008e+00	5.124554e+00
coef.var	5.668259e+16	-9.215120e+16

Πίνακας 3 Πίνακα Περιγραφικών Στατιστικών Ευρωπαϊκής Ένωσης και Μη Ευρωπαϊκής Ένωσης

Αυτό περιλαμβάνει την υπολειμματική αξία, την διάμεσο (Median ή το τεταρτημόριο Q2) ,τα τεταρτημόρια Q1,Q3 , μέγιστη και ελάχιστη τιμή.

```
summary.eu<-summary(s.regmodel.eu)
summary.noeu<-summary(s.regmodel.noeu)
```


Η έξοδος που θα απεικονίζεται από τον παραπάνω κώδικα είναι:

	<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
Χώρες Ευρωπαϊκής Ένωσης	-7.3844	-1.7165	-0.0652	1.9267	5.6967
Χώρες Μη Ευρωπαϊκής Ένωσης	-11.3878	-4.0904	0.1136	4.2954	10.6489

Πίνακας 4 Στατιστικά στοιχεία πρόβλεψης για το 2013 χωρών ΕΕ και μη ΕΕ

<i>Coefficients:</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr (> t)</i>
(Intercept)	65.92982	0.48792	135.12	<2e-16 ***
Health.Expenditure.Total	1.36113	0.06007	22.66	<2e-16 ***
Residual standard error: 2.365 on 502 degrees of freedom				
(308 observations deleted due to missingness)				

Πίνακας 5 Πίνακας Συντελεστών EU

Όταν η απόκλιση σφάλματος (Residual standard error) είναι ακριβώς μηδέν τότε το μοντέλο ταιριάζει τέλεια με τα δεδομένα. Για τις χώρες της Ευρωπαϊκής Ένωσης έχουμε ένα μοντέλο απλής παλινδρόμησης, μία από τις εξόδους είναι η απόκλιση σφάλματος των 2.365 στους 502 βαθμούς ελευθερίας ενώ αντίστοιχα για τις μη Ευρωπαϊκής Ένωσης χώρες έχουμε 5.137 στους 208 βαθμούς ελευθερίας.

<i>Coefficients:</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr (> t)</i>
(Intercept)	69.2517	1.4412	48.052	< 2e-16 ***
Health.Expenditure.Total	0.6344	0.1759	3.606	0.00039 ***
Residual standard error: 4.979 on 142 degrees of freedom				
(225 observations deleted due to missingness)				

Πίνακας 6 Πίνακας Συντελεστών NoEU

Στην έξοδο των παραπάνω πινάκων, διαπιστώνουμε ότι το σημείο τομής είναι 68.61371 του σταθερού όρου (intercept) και ο συντελεστής του ορίου ζωής είναι 1.00916 για τις χώρες της Ευρώπης εξίσωση (1) ενώ για τις μη Ευρωπαϊκής Ένωσης χώρες έχουμε αντίστοιχα την εξίσωση (2) :

$$y = 68.61371 + 1.00916 x \quad (1 \text{ EU})$$

$$y = 69.2517 + 0.6344x \quad (2 \text{ NoEU})$$

όπου η είσοδος $y = \text{Life.Expectancy.Birth.Total}$

και η έξοδος $x = \text{Health.Expenditure.Total}$

Το τυπικό σφάλμα πρόβλεψης (error) μετρά την ποιότητα του μοντέλου μας αλλά σε μονάδες μέτρησης της εισόδου μας και δηλώνεται σε απόλυτη τιμή επί την διαφορά της εξαρτημένης μεταβλητής μείον την προβλεθφείσα ώστε να μας δώσουν το μέγεθος του σφάλματος. Ενώ με την στήλη *t value* έχουμε τον έλεγχο σημαντικότητας των συντελεστών. Η σημαντικότητα της P-value φαίνεται και από τα 3 αστέρια ‘***’ στο τέλος κάθε αριθμού άρα όσο πιο πολλά αστέρια τόσο ο αριθμός είναι πιο κοντά στο μηδέν όπου βλέπουμε και από την έξοδο του κώδικα (Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1).

Root Mean Square Error

Η ρίζα του μέσου τετραγωνικού σφάλματος (sqrt) δίνει την τυπική απόκλιση του μοντέλου πρόβλεψης για το προσδόκιμο ζωής κατά τη γέννηση με τιμή 3.5002919 για τις χώρες της Ευρωπαϊκής Ένωσης και με τιμή 5.3131970 για τις χώρες εκτός Ευρωπαϊκής Ένωσης. Όσο μικρότερη είναι η τιμή του σφάλματος τόσο καλύτερη απόδοση θα έχει.

```
rmse.eu = sqrt(mean((pred.eu dset.eu$Life.expectancy.at.birth ..total..years.)  
^2, na.rm = TRUE))
```

```
rmse.noeu = sqrt(mean((pred.noeu dset.noeu$Life.expectancy.at.birth..total  
..years.)^2, na.rm = TRUE))
```

R-squared

Στους παρακάτω δύο πίνακες βλέπουμε τα αποτελέσματα για R-squared δηλαδή ο συντελεστής προσδιορισμού R^2 είναι το ποσοστό της μεταβλητότητας σε ένα σύνολο δεδομένων. Το $R^2 = 0.5056$ δείχνει τα χαρακτηριστικά της ανεξάρτητης μεταβλητής εξηγούν το 50.46% της μεταβλητότητας των δεδομένων για τις Ευρωπαϊκές Ένωσης χώρες ενώ το $R^2 = 0.05883$ δείχνει τα χαρακτηριστικά της ανεξάρτητης μεταβλητής για τις μη Ευρωπαϊκής Ένωσης χώρες όπου εξηγούν το 5,43% της μεταβλητότητας των δεδομένων.

Ευρωπαϊκής Ένωσης	
Multiple R-squared:	0.5056
Adjusted R-squared:	0.5046
F-statistic:	513.4 on 1 and 502 DF
p-value:	< 2.2e-16

Μη Ευρωπαϊκής Ένωσης	
Multiple R-squared:	0.05883
Adjusted R-squared:	0.0543
F-statistic:	13 on 1 and 208 DF
p-value:	0.0003897

Πίνακας 7 Πίνακας Συντελεστής Προσδιορισμού EU και NoEU

Με το **F-statistic** βλέπουμε την εγκυρότητα του μοντέλου 513.4 on 1 and 502 DF ενώ p-value όσο πιο κοντά στο 0 είναι τόσο καλύτερη είναι η σημαντικότητα της ανάλυσης για τις χώρες τις Ευρώπης. Ενώ βλέπουμε για τις χώρες τις μη Ευρωπαϊκής Ένωσης την εγκυρότητα του μοντέλου 13 on 1 and 208 DF και το p-value 0.0003897.

3.5 Πολλαπλή Γραμμική Παλινδρόμηση

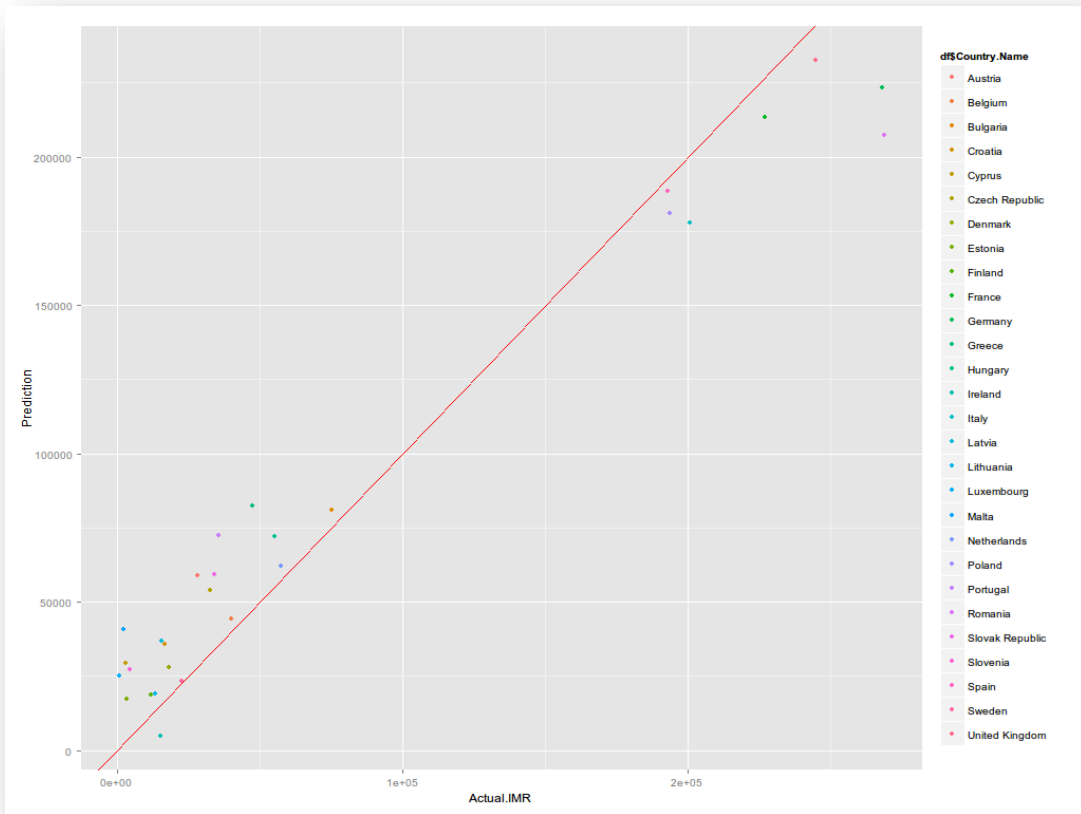
Για την Πολλαπλή Γραμμική Παλινδρόμηση, χωρίσαμε τα δεδομένα σε Ευρωπαϊκής Ένωσης χώρες και μη. Οι δείκτες που χρησιμοποιήσαμε είναι οι εξής:

- Ποσοστό βρεφικής θνησιμότητας (ανά 1000 γεννήσεις ζωντανών) - *Mortality rate, infant (per 1,000 live births)*
- Ακαθάριστο ποσοστό γεννητικότητας (ανά 1000 άτομα) - *Birth rate, crude (per 1,000 people)*
- Ακαθάριστο ποσοστό θνησιμότητας (ανά 1000 άτομα) - *Death rate, crude (per 1,000 people)*
- Προσδόκιμο ζωής κατά τη γέννηση (Συνολικά χρόνια) - *Life expectancy at birth, total (years)*
- Αριθμός θανάτων βρεφών - *Number of infant deaths*
- Συνολικό ποσοστό γονιμότητας (γεννήσεις ανά γυναίκα) - *Fertility rate, total (births per woman)*

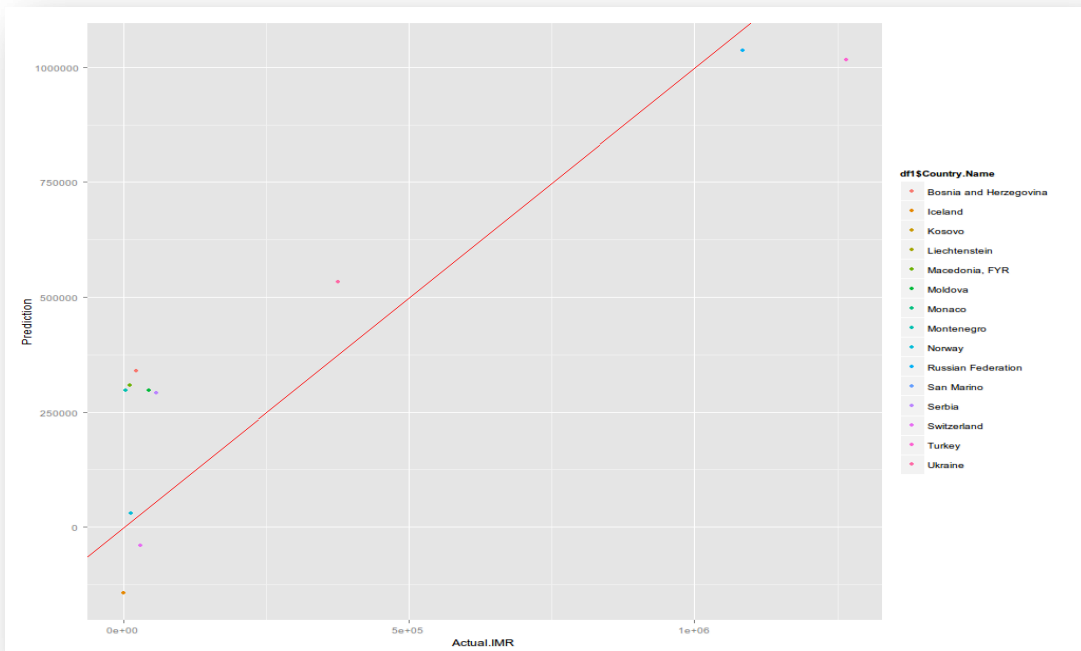
Ως σύνολο εκπαίδευσης, χρησιμοποιήσαμε τα παραπάνω δεδομένα για τις χρονολογίες 1984 – 2012 και ως σύνολο ελέγχου τα αντίστοιχα για το έτος 2013. Στη υπόθεση ορίσαμε, στο βασικό μοντέλο, ως εξαρτημένη μεταβλητή (έξοδο) τον Δείκτη Ποσοστού Βρεφικής Θνησιμότητας (MRI), και ως ανεξάρτητες μεταβλητές (είσοδο) τους υπόλοιπους από τους παραπάνω δείκτες.

```
fmla.eu <- as.formula(Mortality.rate.infant.per.1.000.live.births. ~  
Birth.rate.per.1.000.people. + Death.rate.per.1.000.people. +  
Life.expectancy.total. + Number.of.infant.deaths +  
Fertility.rate.total.births.per.woman.)  
s.regmodel.eu <- lm(fmla.eu, data = dset.eu.train)
```

```
fmla.noEU <- as.formula(Mortality.rate.infant.per.1.000.live.births. ~  
Birth.rate.per.1.000.people. + Death.rate.per.1.000.people.+  
Life.expectancy.total. + Number.of.infant.deaths + Fertility.rate.total.  
births.per.woman.)  
s.regmodel.noEU <- lm(fmla.noEU, data = dset.noEU.train)
```



Εικόνα 10 Σύγκριση πραγματικής και προβλεπόμενης τιμής ανά ευρωπαϊκής ένωσης χώρα (Πολλαπλή Γραμμική Παλινδρόμηση)



Εικόνα 11 Σύγκριση πραγματικής και προβλεπόμενης τιμής ανά μη ευρωπαϊκής ένωσης χώρα (Πολλαπλή Γραμμική Παλινδρόμηση)

Στην μέθοδο της Πολλαπλής Γραμμικής Παλινδρόμησης παρατηρούμε ότι οι τιμές βρίσκονται σε σημείο το οποίο είναι πιο κοντά στην γραμμή παλινδρόμησης με αποτέλεσμα οι προβλέψεις να είναι πιο έγκυρες. Εδώ, οι ανεξάρτητες μεταβλητές είναι πιο πολλές σε σχέση με την Απλή Γραμμική και έτσι ο αλγόριθμος μπορεί να αναλύσει καλύτερα τα δεδομένα και να εξάγει αποτελέσματα τα οποία συμβαδίζουν καλύτερα με τις πραγματικές τιμές. Στο επόμενο υποκεφάλαιο παρουσιάζονται τιμές στατιστικών εντολών της R, όσο ακόμα και οι τιμές των σφαλμάτων της πρόβλεψης.

3.5.1 Ανάλυση του Μοντέλου Πολλαπλής Γραμμικής Παλινδρόμησης

Σε πολλά πρακτικά προβλήματα είναι απαραίτητο να χρησιμοποιήσουμε δύο ή και περισσότερες ανεξάρτητες μεταβλητές προκειμένου να ερμηνεύσουμε με μεγαλύτερη ακρίβεια τα δεδομένα μας. Με την προϋπόθεση ότι ισχύουν οι υποθέσεις που προαναφέρθηκαν, για το μοντέλο βλέπουμε τα αποτελέσματα του ελέγχου μέσω της εντολής `summary.lm(s.regmodel.eu)` για τις χώρες της Ευρωπαϊκής Ένωσης και `summary.lm(s.regmodel.noeu)` για τις χώρες εκτός Ευρωπαϊκής Ένωσης. Η έξοδος αυτών εμφανίζονται στον παρακάτω πίνακα και περιλαμβάνει την υπολειμματική αξία, την διάμεσο (Median ή το τεταρτημόριο Q2), τα τεταρτημόρια Q1, Q3, μέγιστη και ελάχιστη τιμή.

	<i>Min</i>	<i>1Q</i>	<i>Median</i>	<i>3Q</i>	<i>Max</i>
<i>Χώρες Ευρωπαϊκής Ένωσης</i>	-5.8327	-1.8036	-0.1446	1.4022	10.4318
<i>Χώρες Μη Ευρωπαϊκής Ένωσης</i>	-10.3819	-3.0299	-0.1882	2.1810	16.7226

Πίνακας 8 Στατιστικά στοιχεία πρόβλεψης για το 2013 χωρών ΕΕ και μη ΕΕ

Για τις χώρες της Ευρωπαϊκής Ένωσης έχουμε ένα μοντέλο πολλαπλής παλινδρόμησης, μία από τις εξόδους είναι η απόκλιση σφάλματος (Residual standard error) των 2.658 στους 806 βαθμούς ελευθερίας ενώ αντίστοιχα για τις μη Ευρωπαϊκής Ένωσης χώρες έχουμε 4.267 στους 299 βαθμούς ελευθερίας.

Coefficients:	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	1.522e+02	9.226e+00	16.499	< 2e-16 ***
Birth.rate.per.1.000.people.	-5.823e-01	3.304e-01	-1.763	0.079
Death.rate.per.1.000.people.	-1.290e+00	1.764e-01	-7.316	2.37e-12 ***
Life.expectancy.total.	-1.951e+00	9.818e-02	-19.871	< 2e-16 ***
Number.of.infant.deaths	1.036e-04	1.811e-05	5.721	2.57e-08 ***
Fertility.rate.total.births.per.woman.	1.402e+01	2.701e+00	5.192	3.86e-07 ***
Residual standard error: 4.267 on 299 degrees of freedom				
(130 observations deleted due to missingness)				

Πίνακας 9 Πίνακας Συντελεστών NoEU

Διαπιστώνουμε ότι το σημείο τομής είναι $1.522e+02$ του σταθερού όρου (intercept) και ο συντελεστής του ακαθάριστου ποσοστού γεννητικότητας είναι $-5.823e-01$, το ακαθάριστο ποσοστό θνησιμότητας $-1.290e+00$, το προσδόκιμο ζωής κατά τη γέννηση $-1.951e+00$, ο αριθμός θανάτων των βρεφών $1.036e-04$, και το συνολικό ποσοστό γονιμότητας $1.402e+01$ για τις χώρες της Ευρωπαϊκής Ένωσης και η εξίσωση (3) και αντίστοιχα για τις χώρες εκτός Ευρωπαϊκής Ένωσης η εξίσωση (4) που δημιουργούνται είναι :

$$y = (1.522e + 02) + (-5.823e - 01)x_1 + (-1.290e + 00)x_2 + (-1.951e + 00)x_3 + (1.036e - 04)x_4 + (1.402e + 01)x_6 \quad (3)$$

$$y = (1.131e + 02) + (-3.925e - 01)x_1 + (-4.700e - 01)x_2 + (-1.364e + 00)x_3 + (7.359e - 04)x_4 + (4.159e + 00)x_5 \quad (4)$$

όπου η έξοδος $y =$ " Mortality.rate.infant.per.1.000.live.births."

και οι είσοδοι $x_1 =$ "Birth.rate.per.1.000.people.",

$x_2 =$ "Death.rate.per.1.000.people.",

$x_3 =$ "Life.expectancy.total.",

$x_4 =$ "Number.of.infant.deaths",

$x_5 =$ "Fertility.rate.total.births.per.woman."

Coefficients:	Estimate	Std. Error	t value	Pr (> t)
(Intercept)	1.131e+02	4.707e+00	24.027	< 2e-16 ***
Birth.rate.per.1.000.people.	-3.925e-01	1.339e-01	-2.930	0.00348 **
Death.rate.per.1.000.people.	-4.700e-01	9.274e-02	-5.068	4.99e-07 ***
Life.expectancy.total.	-1.364e+00	4.745e-02	-28.746	< 2e-16 ***
Number.of.infant.deaths	7.359e-04	4.400e-05	16.724	< 2e-16 ***
Fertility.rate.total.births.per.woman.	4.159e+00	9.102e-01	4.570	5.64e-06 ***
Residual standard error:	2.658 on 806 degrees of freedom			

Πίνακας 10 Πίνακας Συντελεστών EU

Στην στήλη *t value* αναγράφονται οι τιμές τις στατιστικής συνάρτησης βάση της οποίας γίνεται ο έλεγχος για την σημαντικότητα ύπαρξης των συντελεστών στην συνάρτηση. Ενώ η p-value όσο πιο κοντά στο 0 είναι τόσο καλύτερη είναι η σημαντικότητα της ανάλυσης για τις χώρες.

Ξεκινώντας με τα βασικά στατιστικά στοιχεία έχουμε τον αριθμό των τιμών (*nbr.val*), τον αριθμό των μηδενικών τιμών (*nbr.null*), τον αριθμό των τιμών που λείπουν (*nbr.na*), την ελάχιστη τιμή (*min*), τη μέγιστη τιμή (*max*), το εύρος και το άθροισμα όλων των τιμών, : η διάμεσος (*median*), η μέση τιμή ή μέσος όρος (*mean*), το τυπικό σφάλμα επί του μέσου (*SE.mean*), το διάστημα εμπιστοσύνης της μέσης τιμής (*CI.mean*) στο α επίπεδο (α είναι η πιθανότητα που χρησιμοποιείται για τον υπολογισμό του διαστήματος εμπιστοσύνης (*CI.mean*) με τιμή $\alpha = 0,95$), η διακύμανση (*var*), η τυπική απόκλιση (*std.dev*) και ο συντελεστής διακύμανσης (*coef.var*) όπου ορίζεται ως η τυπική απόκλιση διαιρούμενη με τη μέση.

`stat.desc(s.regmodel.eu$residuals)`

	Ευρωπαϊκής Ένωσης	Μη Ευρωπαϊκής Ένωσης
nbr.val	8.120000e+02	3.050000e+02
nbr.null	0.000000e+00	0.000000e+00
nbr.na	0.000000e+00	0.000000e+00
min	-5.832679e+00	-1.038188e+01
max	1.043176e+01	1.672255e+01
range	1.626444e+01	2.710443e+01
sum	3.377854e-14	3.406997e-14
median	-1.446284e-01	-1.882435e-01
mean	4.156504e-17	1.118264e-16
SE.mean	9.299634e-02	2.423269e-01

CI.mean.0.95	1.825419e-01	4.768504e-01
var	7.022435e+00	1.791031e+01
std.dev	2.649988e+00	4.232057e+00
coef.var	6.375520e+16	3.784489e+16

Πίνακας 11 Πίνακα Περιγραφικών Στατιστικών Ευρωπαϊκής Ένωσης και Μη Ευρωπαϊκής Ένωσης

Ο δείκτης R Square είναι ένδειξη του ποσοστού της διακύμανσης της εξαρτημένης μεταβλητής δηλαδή το ποσοστό βρεφικής θνησιμότητας που επεξηγεί η υπόθεση μας με τιμή 0.7611 που εξηγούν το 75.96% της μεταβλητότητας των δεδομένων για τις Ευρωπαϊκής Ένωσης χώρες ενώ για τις μη Ευρωπαϊκής Ένωσης χώρες είναι 0.8914 που εξηγούν το 88.95% της μεταβλητότητας των δεδομένων.

	Ευρωπαϊκής Ένωσης	Μη Ευρωπαϊκής Ένωσης
Multiple R-squared:	0.7611	0.8914
Adjusted R-squared:	0.7596	0.8895
F-statistic:	513.5 on 5 and 806 DF	490.7 on 5 and 299 DF
p-value:	< 2.2e-16	< 2.2e-16

Πίνα
κας
12
Πίνα
κας
Συντ
ελεσ
τής
Προ

σδιορισμού EU και NoEU

Με το F-statistic βλέπουμε την εγκυρότητα του μοντέλου 513.5 on 5 and 806 DF για τις χώρες της Ευρωπαϊκής Ένωσης ενώ 490.7 on 5 and 299 DF για τις χώρες εκτός Ευρωπαϊκής Ένωσης.

Root Mean Square Error

Η ρίζα του μέσου τετραγωνικού σφάλματος (sqrt) δίνει την τυπική απόκλιση του μοντέλου πρόβλεψης για το ποσοστό βρεφικής θνησιμότητας με τιμή 6.34043 για τις χώρες της Ευρωπαϊκής Ένωσης και με τιμή 12.8393 για τις χώρες εκτός Ευρωπαϊκής Ένωσης. Όσο μικρότερη είναι η τιμή του σφάλματος τόσο καλύτερη απόδοση θα έχει.

```
rmse.eu=sqrt(mean((pred.eu-
dset.eu$Mortality.rate..infant..per.1.000.live.births.)^2,na.rm=TRUE))
```

```
rmse.noeu=sqrt(mean((pred.noeu-  
dset.noeu$Mortality.rate..infant..per.1.000.live.births.)^2, na.rm=TRUE))
```

3.6 Νευρωνικά Δίκτυα

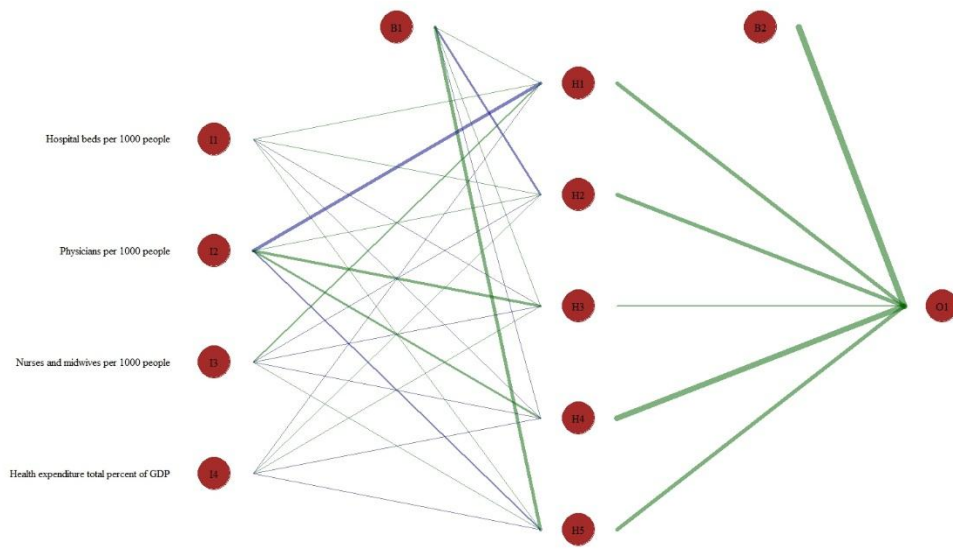
Στην περίπτωση των Νευρωνικών Δικτύων αφαιρέσαμε τις χώρες Λιχτενστάιν, Μονακό, Κόσσοβο καθώς δεν περιείχαν αρκετές τιμές και εφαρμόσαμε προεπεξεργασία στους δείκτες έτσι ώστε να αντικατασταθούν οι ελλιπείς τιμές με τον μέσο όρο του κάθε δείκτη ανά χώρα. Επίσης χωρίσαμε τα δεδομένα σε χώρες εντός και εκτός της Ευρωπαϊκής Ένωσης. Οι δείκτες που χρησιμοποιήσαμε για την συγκεκριμένη υπόθεση είναι οι ακόλουθοι:

- Κλίνες (ανά 1000 άτομα) - *Hospital beds (per 1,000 people)*
- Ιατρικό προσωπικό (ανά 1000 άτομα) - *Physicians (per 1,000 people)*
- Νοσοκόμες και μαίες (ανά 1000 άτομα) - *Nurses and midwives per 1000 people*
- Συνολικές δαπάνες υγείας (% του ΑΕΠ) - *Health expenditure, total (% of GDP)*
- Προσδόκιμο ζωής κατά τη γέννηση - *Life expectancy at birth total years*

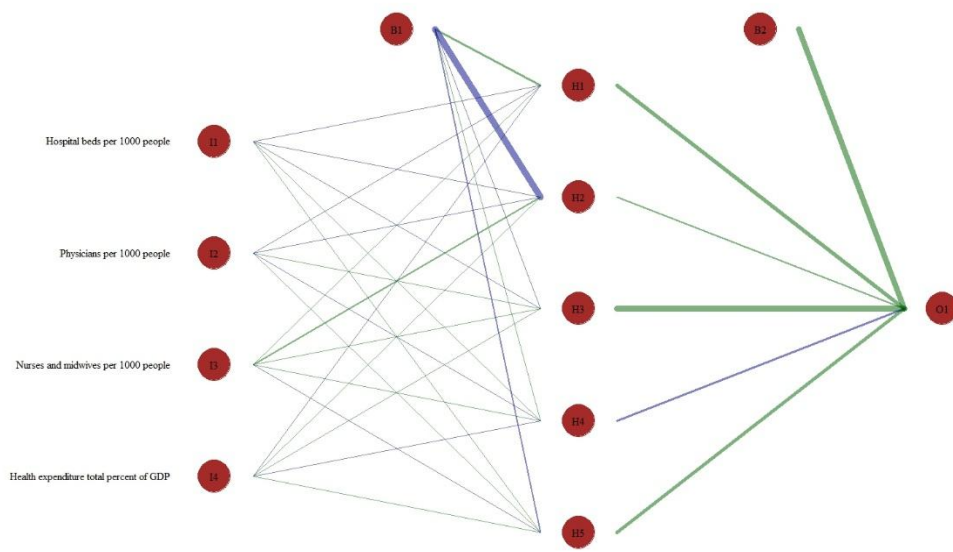
Ως σύνολο εκπαίδευσης, χρησιμοποιήσαμε τα παραπάνω δεδομένα για τις χρονολογίες 1984 – 2012 και ως σύνολο ελέγχου τα αντίστοιχα για το έτος 2013. Στη υπόθεση ορίσαμε, στο μοντέλο εκπαίδευσης, ως έξοδο το Προσδόκιμο ζωής κατά τη γέννηση, και ως είσοδο τους υπόλοιπους από τους παραπάνω δείκτες.

Εκτός από τις παραπάνω προεπεξεργασίες, στην συγκεκριμένη υπόθεση, έχουμε χωρίσει το Προσδόκιμο Ζωής σε 2 κατηγορικές τιμές ανάλογα την τιμή του δείκτη. Από 75 έτη και κάτω ως “Low”, ενώ άνω των 75 ετών ως “High”. Η επιλογή της τιμής αυτής έγινε καθώς ο μέσος όρος του δείκτη είναι 75.09 έτη.

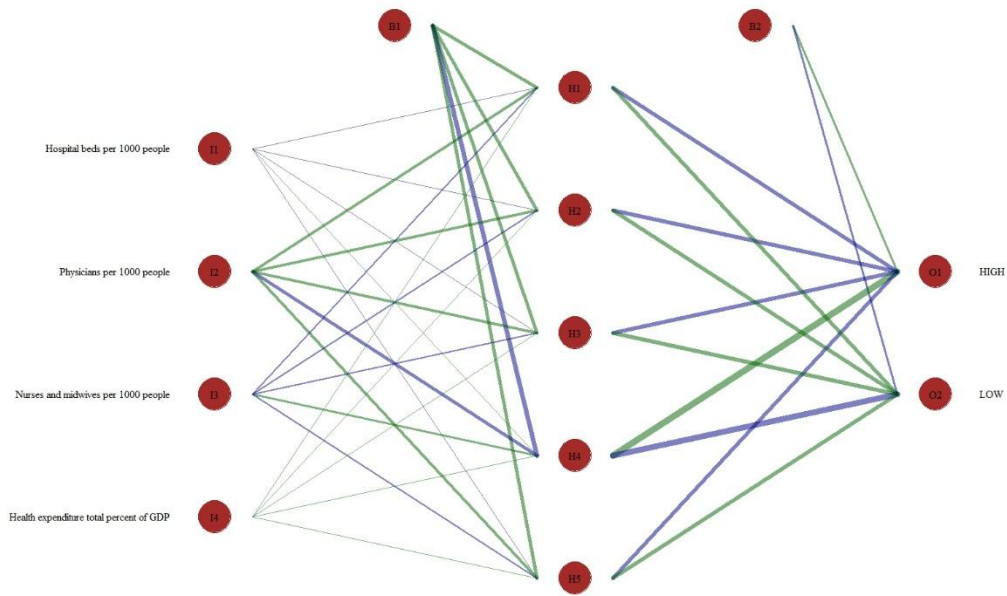
Στις εικόνες 12 και 13 παρουσιάζονται τα νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης 5 κρυφούς νευρώνες καθώς και η σημαντικότητα των βαρών κάθε εισόδου (δείκτη) και κρυφών νευρώνων για την δημιουργία εξόδου (πρόβλεψης), όπου η έξοδος είναι αριθμητική τιμή. Στις εικόνες 14 και 15 υπάρχουν τα αντίστοιχα νευρωνικά δίκτυα με εξόδους τις δύο κατηγορικές τιμές. Οι μπλε γραμμές φανερώνουν την αρνητικότητα των συναπτικών βαρών ενώ οι πράσινες την θετική τους τάση. Το πάχος αυτών προσδιορίζει την σημαντικότητά τους. (Beck)



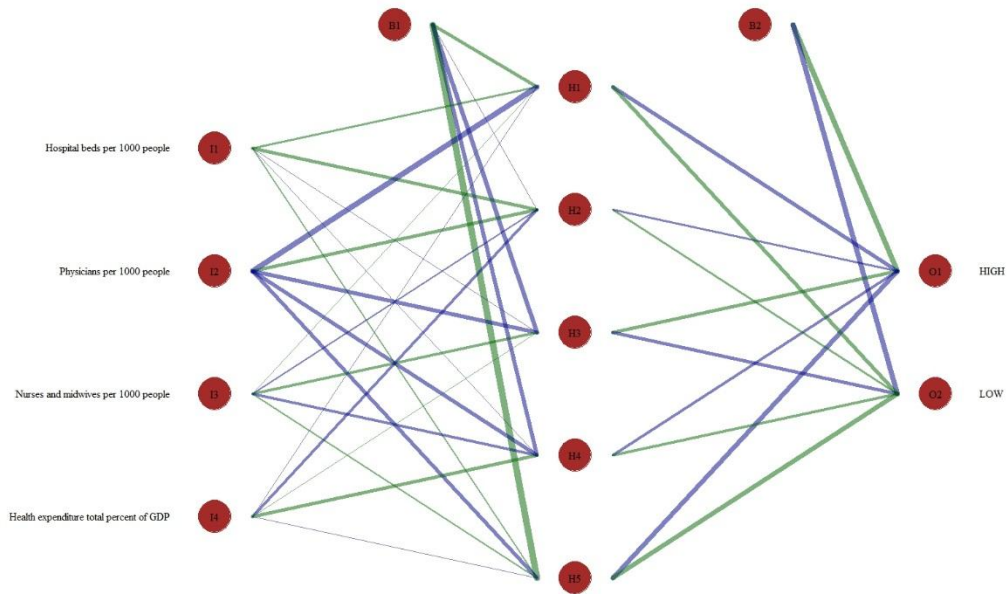
Εικόνα 12 Νευρωνικό δίκτυο δεικτών χωρών μη Ευρωπαϊκής Ένωσης (Εξόδος αριθμητικής τιμής)



Εικόνα 13 Νευρωνικό δίκτυο δεικτών χωρών Ευρωπαϊκής Ένωσης (Εξόδος αριθμητικής τιμής)



Εικόνα 14 Νευρωνικό δίκτυο δεικτών χωρών μη Ευρωπαϊκής Ένωσης (Εξοδος κατηγορικών τιμών)



Εικόνα 15 Νευρωνικό δίκτυο δεικτών χωρών Ευρωπαϊκής Ένωσης (Εξοδος κατηγορικών τιμών)

3.6.1 Ανάλυση του Μοντέλου Νευρωνικών Δικτύων

Στον παρακάτω πίνακα βλέπουμε τις τιμές της πρόβλεψης του δείκτη Προσδόκιμου Ζωής Κατά Τη Γέννηση για το 2013 με βάση το μοντέλο των νευρωνικών δικτύων για τις Ευρωπαϊκής Ένωσης χώρες καθώς και τις μη Ευρωπαϊκής Ένωσης.

Ο πίνακας 13 παρουσιάζει στατιστικά στοιχεία αυτής της πρόβλεψης οδηγώντας στο συμπέρασμα ότι το Προσδόκιμο Ζωής στις Ευρωπαϊκής Ένωσης χώρες είναι μεγαλύτερο από αυτό στις χώρες μη Ευρωπαϊκής Ένωσης. Συγκεκριμένα, η ελάχιστη και μέγιστη τιμή του δείκτη παρουσιάζει μια διαφορά της τάξεως του 3% περίπου προς όφελος των ΕΕ χωρών σε σχέση με αυτές των μη ΕΕ. Την ίδια περίπου διαφορά συναντάμε και σε επίπεδο τεταρτημόριων με αυτή να κυμαίνεται γύρω στο 2,5% με 3% περίπου, εκτός από το 3^ο τεταρτημόριο στο οποίο παρατηρείται μικρή κλίση της διαφοράς προς όφελος των χωρών μη ΕΕ (περίπου 0.5%).

Εντολές: `> summary(pred.eu)` και `> summary(pred.noeu)`

	Ευρωπαϊκής Ένωσης	Μη Ευρωπαϊκής Ένωσης
Min.	70.08	67.49
1st Qu.	74.22	67.68
Median	76.88	70.39
Mean	76.18	72.72
3rd Qu.	78.56	79.20
Max.	82.56	79.59

Πίνακας 13 Στατιστικά στοιχεία πρόβλεψης για το 2013 χωρών ΕΕ και μη ΕΕ

Σκοπός των στατιστικών μέτρων είναι η αντικατάσταση μιας μεγάλης μάζας αριθμητικών δεδομένων με λιγότερους αριθμούς, που μεταφέρουν το μεγαλύτερο ποσοστό της βασικής πληροφορίας που περιέχεται στα δεδομένα.

Με τις παρακάτω εντολές `> stat.desc(pred.eu)` και `> stat.desc(pred.noeu)` της υπόθεσης μας έχουμε την ακόλουθη έξοδο:

	Ευρωπαϊκής Ένωσης	Μη Ευρωπαϊκής Ένωσης
nbr.val	3.200000e+01	8.00000000
nbr.null	0.000000e+00	0.00000000
nbr.na	0.000000e+00	0.00000000
min	7.010019e+01	67.38291153
max	8.079198e+01	79.92015637
range	1.069179e+01	12.53724484
sum	2.429566e+03	586.13706399
median	7.683751e+01	72.30720210
mean	7.592393e+01	73.26713300
SE.mean	5.460975e-01	2.02057855
CI.mean.0.95	1.113773e+00	4.77790904
var	9.543120e+00	32.66190134
std.dev	3.089194e+00	5.71505917
coef.var	4.068802e-02	0.07800304
nbr.val	3.200000e+01	8.00000000

Πίνακας 14 Πίνακας Περιγραφικών Στατιστικών Ευρώπης και Μη Ευρώπης Νευρωνικών Δικτύων

Στον πίνακα 14 χρησιμοποιήσαμε διαφορετικό κώδικα έτσι ώστε να εμφανιστούν περισσότερα στατιστικά στοιχεία. Έτσι έχουμε, το εύρος (range) της πρόβλεψης του δείκτη (διαφορά μεγαλύτερης καταγεγραμμένης τιμής με την μικρότερη), περίπου 11 για τις χώρες της ΕΕ και περίπου 12 για αυτές των μη ΕΕ. Όσο πιο χαμηλή η τιμή του εύρους, τόσο πιο μικρή η απόσταση των τιμών των μεγαλύτερων και μικρότερων παρατηρήσεων. Το διάστημα εμπιστοσύνης του μέσου (CI.mean.95) είναι 1 για τις χώρες ΕΕ και πρέπει να προστεθεί στο μέσο για να δημιουργηθεί το διάστημα εμπιστοσύνης με επίπεδο σημαντικότητας 95%, ενώ η τιμή του διαστήματος για τις χώρες των μη ΕΕ είναι περίπου 5. Η τυπική απόκλιση (std.dev) είναι 3 για τις χώρες της ΕΕ και περίπου 6 για αυτές των μη ΕΕ. Ενδιαφέρον έχει η διαφορά της διακύμανσης (var) των χωρών ΕΕ από αυτές των μη ΕΕ (9.5 με 32 αντίστοιχα) καθώς έτσι καταλήγουμε στο ότι οι τιμές της πρόβλεψης του δείκτη για τις χώρες της ΕΕ απέχουν λιγότερο από τον μέσο όρο αυτής σε σχέση με τις χώρες των μη ΕΕ.

Η ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error) δίνει την τυπική απόκλιση του μοντέλου πρόβλεψης για το προσδόκιμο ζωής κατά τη γέννηση με τιμή 3.255060 για τις χώρες της Ευρωπαϊκής Ένωσης και με τιμή 3.048815 για τις χώρες εκτός Ευρωπαϊκής Ένωσης. Όσο μικρότερη είναι η τιμή του σφάλματος τόσο καλύτερη απόδοση θα έχει.

```
RMSE.noeu <- sqrt(sum((output.noeu.ts - pred.noeu)^2)/nrow(pred.noeu))
RMSE.eu <- sqrt(sum((output.eu.ts - pred.eu)^2)/nrow(pred.eu))
```

3.6.2 Μήτρα Σύγχυσης

Η μήτρα σύγχυσης είναι ένας πίνακας στον οποίον είναι αποθηκευμένο το πλήθος των εγγραφών ελέγχου που έχουν προβλεφθεί σωστά και λανθασμένα. Στην κύρια διαγώνια της μήτρας παρουσιάζεται το πλήθος των σωστών προβλέψεων, ενώ και τα μη διαγώνια στοιχεία φανερώνουν το πλήθος των λανθασμένων προβλέψεων.

Τα στατιστικά μέτρα που παρουσιάζουμε επιπλέον είναι:

- Ακρίβεια Πρόβλεψης (Accuracy): ορίζεται ως η αναλογία των δειγμάτων (θετικών και αρνητικών) που έχουν προβλεφθεί σωστά.
- Ευαισθησία (Sensitivity): ορίζεται ως η αναλογία των θετικών δειγμάτων που έχουν προβλεφθεί σωστά.
- Ιδιαιτερότητα (Specificity): ορίζεται ως η αναλογία των αρνητικών δειγμάτων που έχουν προβλεφθεί σωστά.

Μήτρα Σύγχυσης χωρών Ευρωπαϊκής Ένωσης και Στατιστικά Μέτρα

```
real.category      predictions
                  1  2
1                   20  8
2                    0  4

Accuracy: 0.75
Sensitivity: 0.7143
Specificity: 1.0000
```

Μήτρα Σύγχυσης χωρών μη Ευρωπαϊκής Ένωσης και Στατιστικά Μέτρα

```
real.category.noeu predictions.noeu
                  1  2
1                   3  1
2                    0  4

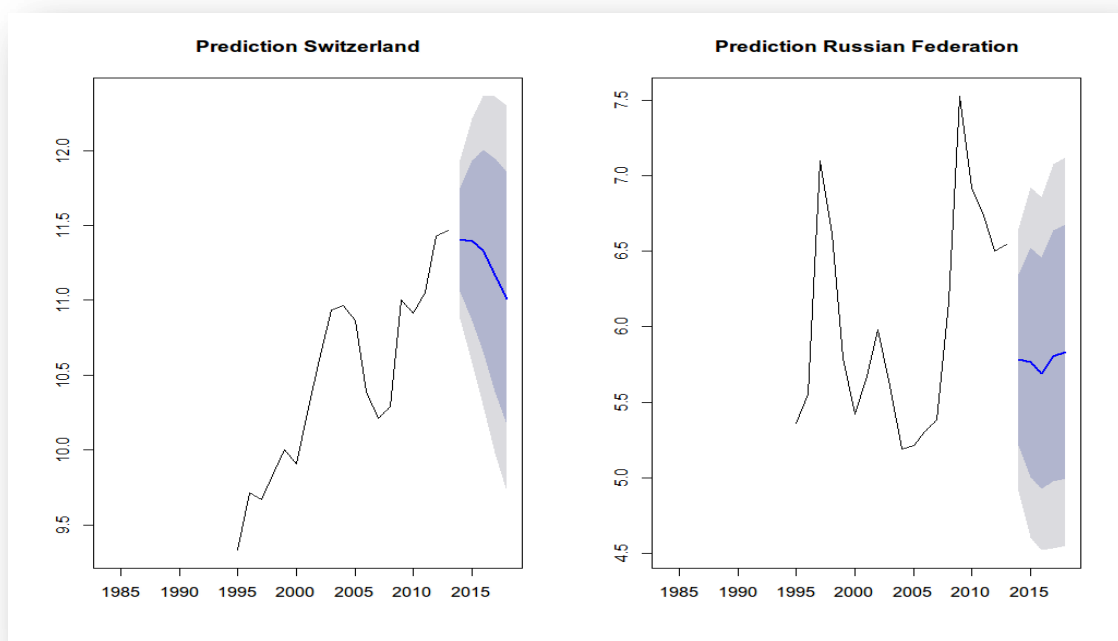
Accuracy: 0.875
Sensitivity: 0.750
Specificity: 1.000
```


Η ακρίβεια της πρόβλεψης παρατηρούμε ότι είναι σε μεγαλύτερες τιμές για τις χώρες μη ΕΕ. Τα ποσοστά της Ευαισθησίας και της Ιδιαιτερότητας είναι ψηλά και αυτό είναι ένα ακόμα καλό σημείο το οποίο δείχνει ότι τα αποτελέσματα της πρόβλεψης είναι αξιόπιστα καθώς, οι προβλέψεις στα δύο αυτά μέτρα παρουσιάζουν ποσοστά 71% και 100% αντίστοιχα και παράλληλα οι χώρες μη ΕΕ 75% και 100% αντίστοιχα.

3.7 Χρονοσειρές (Μέθοδος ARIMA)

Για την εφαρμογή της μεθόδου ARIMA, σε χρονοσειρά, χωρίσαμε το δείγμα μας και διαλέξαμε δύο Ευρωπαϊκής Ένωσης χώρες (Γαλλία, Σουηδία) και δύο μη (Ελβετία, Ρωσία), για να γίνει σύγκριση μεταξύ αυτών. Ο δείκτης που επιλέξαμε, για πρόβλεψη, στην συγκεκριμένη υπόθεση είναι αυτός των Συνολικών δαπάνες υγείας.

Το σύνολο εκπαίδευσης, εδώ, είναι όλα τα έτη τα οποία έχουμε στη διάθεσή μας και ως έλεγχο χρησιμοποιούμε μελλοντικά έτη τα οποία είναι του εύρους 2014-2018.

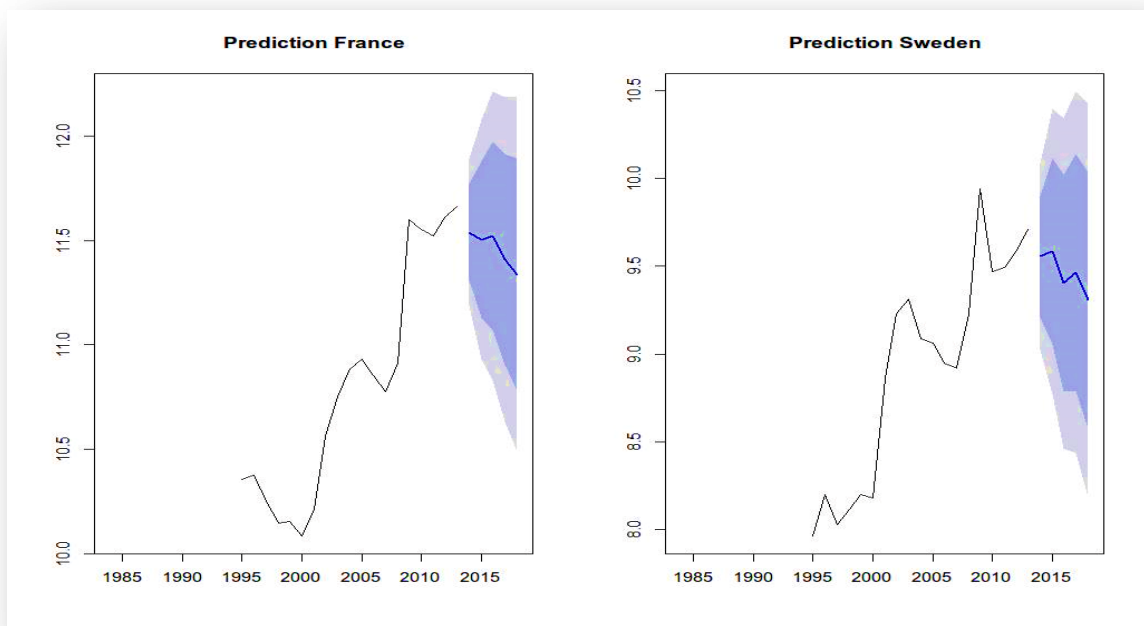


Εικόνα 16 Πρόβλεψη για τα έτη 2014-2018 του δείκτη Συνολικών Δαπανών Υγείας Ελβετίας και Ρωσίας

Παρατηρούμε ότι για το αμέσως επόμενο προβλεπόμενο έτος (2014) όλες οι χώρες παρουσιάζουν μείωση της τιμής του δείκτη.

Για τα επόμενα έτη η τιμή του δείκτη αυξομειώνεται, για να καταλήξει να παρουσιάζει καθοδική πορεία στο τελευταίο έτος (2018) σε όλες τις επιλεγμένες χώρες εκτός της Ρωσίας, η οποία παρουσιάζει άνοδο στις ποσοστιαίες της δαπάνες για την υγεία.

Από αυτή την σύγκριση, σε επίπεδο Ευρώπης, παρατηρούμε ότι οι προβλέψεις των χωρών όσο και τα διαστήματα εμπιστοσύνης αυτής βρίσκονται σε μεγαλύτερα επίπεδα από αυτά των μη ευρωπαϊκών χωρών. Αποτέλεσμα το οποίο εάν εφαρμοστεί στην περίπτωση της Ρωσίας βγάζουμε το συμπέρασμα ότι η τιμή των συνολικών δαπανών μπορεί να ανεβαίνει αλλά βρίσκεται ακόμα σε χαμηλά ποσοστιαία επίπεδα σε σχέση με αυτά των ευρωπαϊκών χωρών.



Εικόνα 17 Πρόβλεψη για τα έτη 2014-2018 του δείκτη Συνολικών Δαπανών Υγείας Γαλλίας και Σουηδίας

3.7.1 Ανάλυση του Μοντέλου ARIMA

Με την ανάλυση του μοντέλου ARIMA και τους παρακάτω κώδικες:

```
> summary(fore.france)           > summary(fore.sweden)
> summary(fore.switzerland)     > summary(fore.russia)
```

Ως έξοδο θα έχουμε τους υπολειμματικούς πίνακες της Γαλλίας, της Σουηδίας, της Ελβετίας και της Ρωσίας. Οι οποίοι θα δείχνουν τα $ar1$ = αυτοπαλίνδρομοι παράμετροι της τάξης 1 και αντίστοιχα $ar2$, $ar3$, $ma1$ = ο κινητός μέσος όρος των παραμέτρων της τάξης 1 όπου υπολογίζει τη μέση τιμή των τιμών της χρονοσειράς και αντίστοιχα τα $ma2$, $ma3$ και το $intercept$ τον σταθερό όρο.

Coefficients	Γαλλία	s.e.	Σουηδία	s.e
ar1	0.294	NaN	-0.4818	0.5365
ar2	-0.4208	NaN	0.7740	0.2173
ar3	0.906	NaN	0.4398	0.4457
ma1	1.0077	NaN	1.7280	0.5375
ma2	1.2190	0.2636	0.7522	0.9234
ma3	0.3246	NaN	-0.1331	0.4836
intercept	10.9054	0.4407	8.8507	0.5245

Πίνακας 15 Υπολειμματικός Πίνακας Γαλλίας και Σουηδίας

Coefficients	Ελβετία	s.e.	Ρωσία	s.e
ar1	1.5346	0.8474	0.4116	0.2575
ar2	-1.4126	0.9893	0.3067	0.3421
ar3	0.7385	0.3389	-0.0545	0.3221
ma1	-0.3751	0.8704	0.5030	0.2754
ma2	0.8466	0.4547	-0.5029	0.2406
ma3	0.2467	0.3941	-1.0000	0.3139
intercept	10.4488	0.5542	5.9845	0.1173

Πίνακας 16 Υπολειμματικός Πίνακας Ελβετίας και Ρωσίας

Εκτός από τους υπολειμματικούς πίνακες βλέπουμε και τους πίνακες των αντίστοιχων χωρών με τα μηνύματα σφαλμάτων.

Error measures:**Γαλλία**

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.03253865	0.174041	0.124017	0.2677031	1.136124	0.9425342	0.001132218

*Πίνακας 17 Πίνακας Σφαλμάτων Χρονοσειράς Γαλλίας***Error measures:****Σουηδία**

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.03253865	0.174041	0.124017	0.2677031	1.136124	0.9425342	0.001132218

*Πίνακας 18 Πίνακας Σφαλμάτων Χρονοσειράς Σουηδίας***Error measures:****Ελβετία**

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.06090888	0.2625044	0.2132404	0.5236454	2.038124	0.9384341	-0.1885759

*Πίνακας 19 Πίνακας Σφαλμάτων Χρονοσειράς Ελβετίας***Error measures:****Ρωσία**

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.02542542	0.4167167	0.346061	-0.9832702	5.680759	0.7615042	0.01159276

Πίνακας 20 Πίνακας Σφαλμάτων Χρονοσειράς Ρωσίας

Root Mean Square Error, RMSE: Η ρίζα του μέσου τετραγωνικού σφάλματος (sqrt) δίνει την τυπική απόκλιση του μοντέλου πρόβλεψης για τις συνολικές δαπάνες υγείας.

Mean Absolute Error, MAE: Το μέσο απόλυτο σφάλμα χρησιμοποιείται κατά τη σύγκριση των προβλέψεων, σε ένα ενιαίο σύνολο δεδομένων.

Mean Percent Error, MPE: Μέσο Ποσοστιαίο Σφάλμα. Το μέσο ποσοστιαίο σφάλμα το χρησιμοποιούμε όταν ενδιαφερόμαστε να προσδιορίσουμε τις προβλεπόμενες τιμές αν είναι συστηματικά μεγαλύτερες ή μικρότερες από τις αντίστοιχες πραγματικές.

Mean Absolute Percent Error, MAPE: Μέσο Απόλυτο Ποσοστιαίο Σφάλμα. Το μέσο απόλυτο ποσοστιαίο σφάλμα εξετάζει τη συμπεριφορά της απόλυτης τιμής του σφάλματος της πρόβλεψης σε σχέση με την πραγματική τιμή της χρονοσειράς.

Mean Absolute Scaled Error, MASE: Η μέση απόλυτη κλίμακα σφάλματος είναι ένα μέτρο πρόβλεψης της ακρίβειας.

Auto Correlation Function, ACFI: Όπου εμφανίζει όλες τις συσχετίσεις εντός των ορίων του δείχνοντας και πως τα κατάλοιπα συμπεριφέρονται σαν λευκός θόρυβος.

3.6 Προφίλ Ελληνικών Νοσοκομείων

3.6.1 Γ.Π.Ν. Ρίου

Το Πανεπιστημιακό Γενικό Νοσοκομείο Πατρών αποτελεί μια από τις μεγαλύτερες νοσηλευτικές μονάδες της Ελλάδας, η οποία, εδώ και χρόνια, προσφέρει ιατρική φροντίδα σύγχρονων προδιαγραφών και υπηρεσίες υγείας υψηλής



Εικόνα 18 Γ.Π.Ν.Ρίου

ποιότητας, γεγονός που πιστοποιείται από τους χιλιάδες πολίτες που το επισκέπτονται από όλα τα μέρη της χώρας μας. Ετησίως εξυπηρετούνται στο Νοσοκομείο άνω των 300.000 πολιτών. Είναι ένα σύγχρονο νοσοκομείο, που προσφέρει σημαντικότατο κοινωνικό έργο στους τομείς της πρωτοβάθμιας και δευτεροβάθμιας περίθαλψης. Αποτελεί μια από τις μεγαλύτερες και κορυφαίες νοσηλευτικές μονάδες της χώρας, δυναμικότητας 800 κλινών. Κτίστηκε το 1988 και στις εγκαταστάσεις του, που εκτείνονται σε 75 χιλιάδες τ.μ., αναπτύσσονται κλινικές και εξωτερικά ιατρεία όλων των ειδικοτήτων, πλήρως εξοπλισμένα με μηχανήματα τελευταίας τεχνολογίας.

3.6.2 Γ.Ν. Άγιος Ανδρέας

Το Νοσοκομείο Αγ. Ανδρέας ιδρύθηκε το 1871 με τίτλο «Δημοτικό Νοσοκομείο Πατρών» και διοικείται από αδελφάτο. Το 1953 με τον Νόμο 2592/1953 «Περί οργάνωσης της Ιατρικής Αντίληψης» μετετράπη σε Γενικό Νοσοκομείο, απετέλεσε Νομικό Πρόσωπο Δημοσίου Δικαίου και διοικείται από Διοικητικό Συμβούλιο,

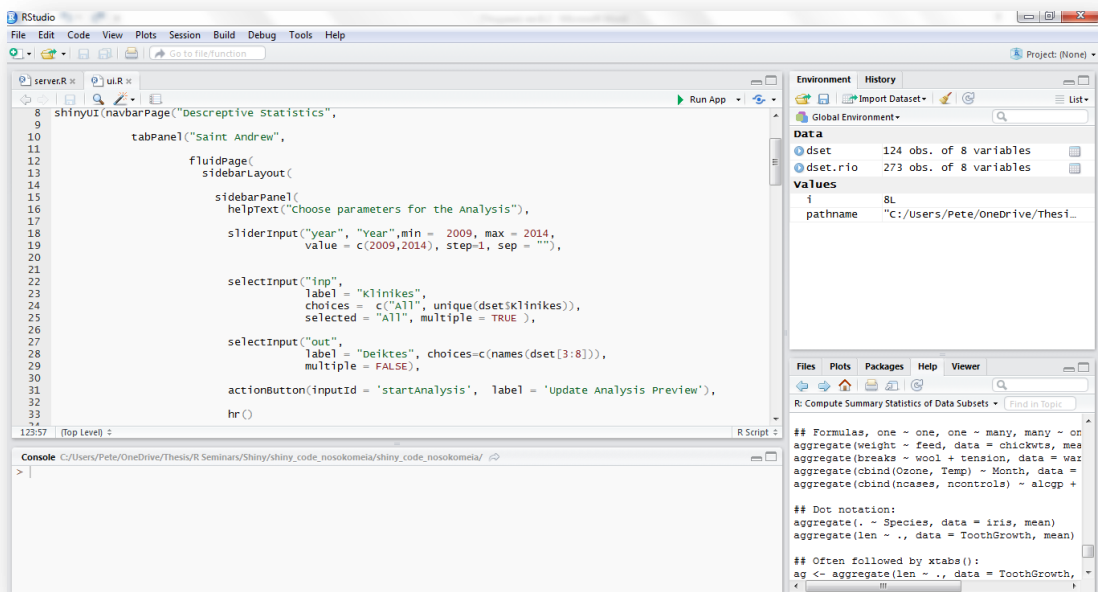


Εικόνα 19 Γ.Ν. Άγιος Ανδρέας

όμως η πραγματική υλοποίηση της μετατροπής του έγινε το 1959 με την διαρκή ανάπτυξη κλινικών και νέων τμημάτων καθώς και αύξηση προσωπικού. Από το 1981 εγκαταστάθηκαν σε αυτό Πανεπιστημιακές Κλινικές οι οποίες λειτουργούσαν παράλληλα με τις κρατικές κλινικές και είχαν παράλληλη ανάπτυξη. Το 1983 με τον νόμο 1397/83 του ΕΣΥ μετονομάστηκε «Περιφερειακό Γενικό Νοσοκομείο» με έδρα την Πάτρα και το 1986 συστάθηκαν τα Κέντρα Υγείας στα οποία συμπεριελήφθησαν τα αγροτικά Ιατρεία ανά Περιφέρεια και συγχωνεύτηκαν οι Υγειονομικοί Σταθμοί. Το 1988 οι Πανεπιστημιακές κλινικές μεταφέρθηκαν στο ΠΠΝΠ του Ρίου. Στο Γ.Ν.Π. “Ο Άγιος Ανδρέας” ανήκει και το πρώην Στρατιωτικό Νοσοκομείο (409) που λειτουργεί σαν δορυφορικό Τμήμα. Με τον Νόμο 2889/2001 μετονομάστηκε σε Γενικό Νοσοκομείο και υπάγεται στο Πε.Σ.Υ.Π. Δυτικής Ελλάδας (νυν Δ.Υ.ΠΕ).

3.7 Δημιουργία εφαρμογής παρουσίασης περιγραφικών στατιστικών με τη χρήση του πακέτου Shiny της R

Σκοπός σε αυτό το κεφάλαιο είναι η ανάπτυξη εφαρμογής με το πακέτο Shiny της R, με δεδομένα των δύο νοσοκομείων του Ν. Αχαΐας. Το Shiny είναι ένα πακέτο, με το οποίο, ο χρήστης μπορεί πολύ εύκολα να κατασκευάσει εφαρμογές με ένα εύχρηστο περιβάλλον εργασίας. Η δομή μιας τέτοιας εφαρμογής, εν συντομία, αποτελείται από δύο αρχεία, το ui το οποίο περιλαμβάνει τον κώδικα που έχει να κάνει με το περιβάλλον και «καλεί» δεδομένα από το παρασκήνιο και το server που έχει ως σκοπό την εκτέλεση των εντολών στο παρασκήνιο πριν γίνουν γραφικά από το αρχείο ui. Ο κώδικας που χρησιμοποιήσαμε, όπως και όλος ο υπόλοιπος κώδικας είναι γραμμένος με το βοηθητικό πρόγραμμα RStudio (Εικόνα 16).



Εικόνα 20 Το γραφικό περιβάλλον του RStudio

Η εφαρμογή αποτελείται από δύο καρτέλες οι οποίες είναι αντίστοιχες με τα νοσοκομεία τα οποία μελετάμε και κάθε μια από αυτή περιέχει τέσσερις διαφορετικές μεθόδους απεικόνισης της περιγραφικής στατιστικής.

Οι μέθοδοι που επιλέξαμε είναι:

- Περίληψη (Summary)

Παρουσιάζει συνοπτικά περιγραφικά στοιχεία για τα δεδομένα που επιλέχθηκαν (min, max, min κ.α.)

- Γράφημα Γραμμής (Line Chart)

Εμφανίζει την τάση των τιμών του δείκτη που επιλέγει ο χρήστης ανά χρονιά και κλινική. Ακόμα, υπάρχει η δυνατότητα επιλογής μέχρι τριών δεικτών ταυτόχρονα για να γίνει σύγκριση.

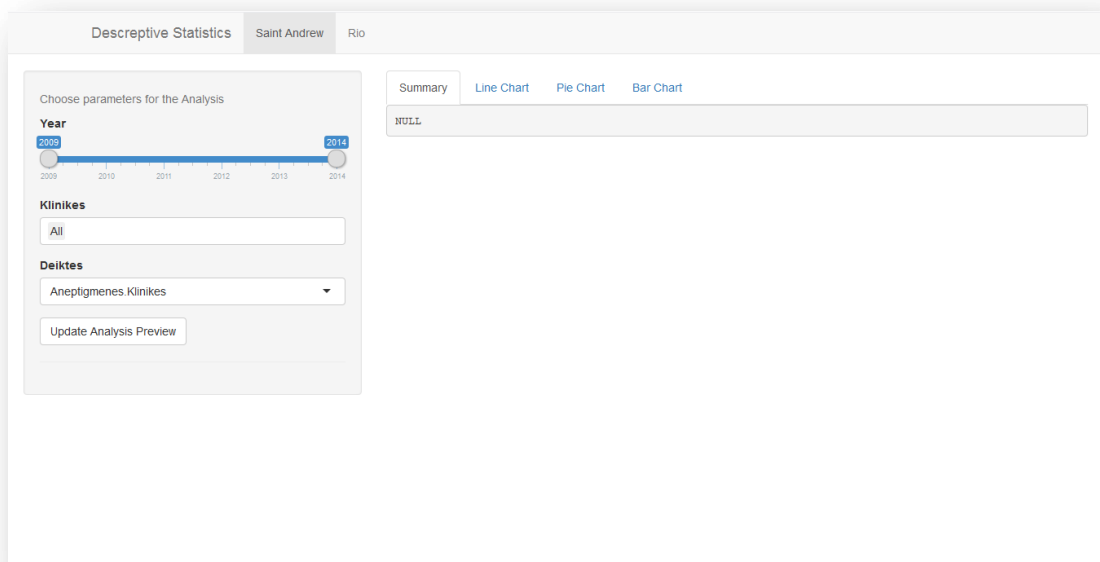
- Γράφημα Πίτας (Pie Chart)

Κατατάσσει τα επιλεγμένα δεδομένα σε μία πίτα καθώς ακόμα, ο χρήστης, μπορεί να δει περισσότερες πληροφορίες για αυτά μετακινώντας τον κέρσορα πάνω από αυτά.

- Ραβδόγραμμα (Bar Chart)

Ο σκοπός εδώ είναι ίδιος με το γράφημα πίτας με την διαφορά ότι τα επιλεγμένα δεδομένα παρουσιάζονται σε δύο κατηγορίες ραβδογραμμάτων, αυτές είναι:

- Ραβδόγραμμα βάση απόλυτων τιμών και χρονολογίας (Stacked Grouped Bar)
- Ραβδόγραμμα βάση ποσοστιαίων τιμών και χρονολογίας (100% Stacked Grouped Bar)



Εικόνα 21 Αρχικό παράθυρο εφαρμογής

Από την αριστερή μπάρα, προσφέρονται επιλογές για εμφάνιση των δεδομένων από τον χρήστη. Αναλυτικά, οι παράμετροι που προφέρονται προς αλλαγή είναι ο χρονικός ορίζοντας, οι κλινικές κάθε νοσοκομείου (δυνατή η πολλαπλή επιλογή αυτών) και οι δείκτες υγείας.

Choose parameters for the Analysis

Year

2009 2014

2009 2010 2011 2012 2013 2014

Klinikes

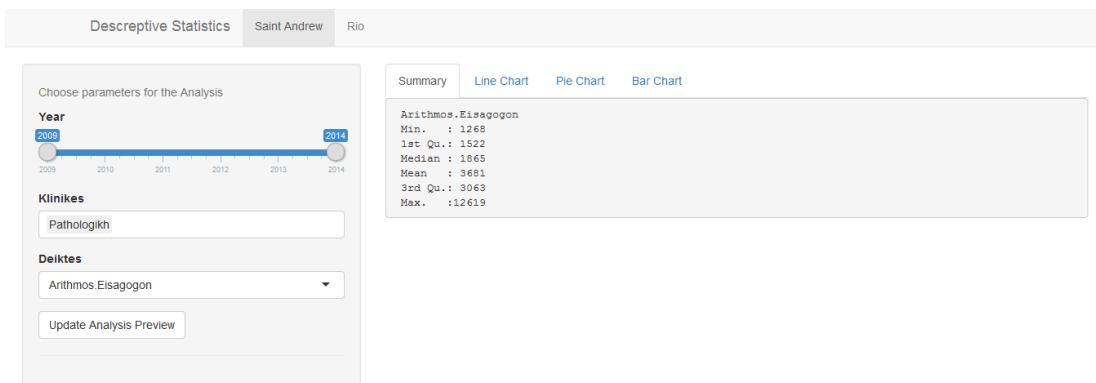
Pathologikh

Deiktes

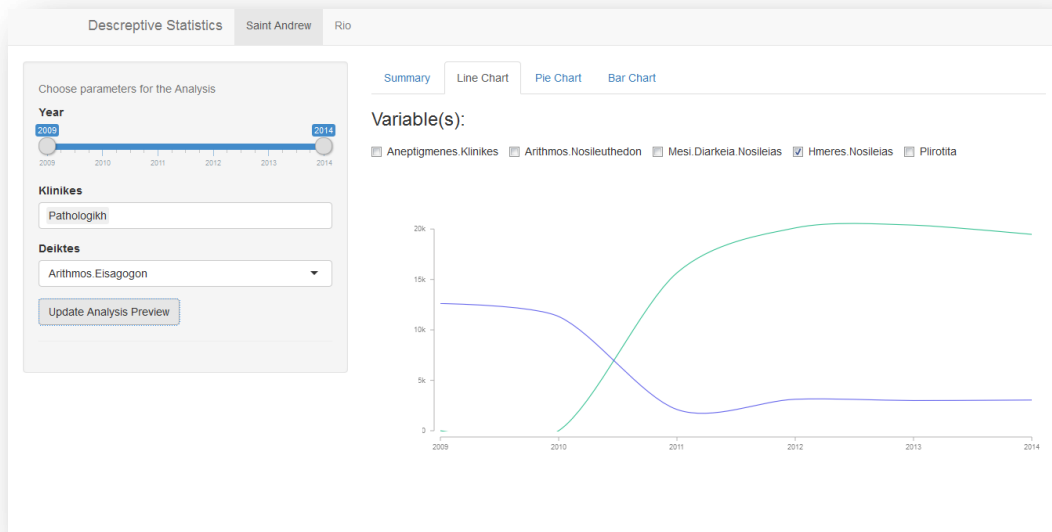
Mesi.Diarkeia.Nosileias

Update Analysis Preview

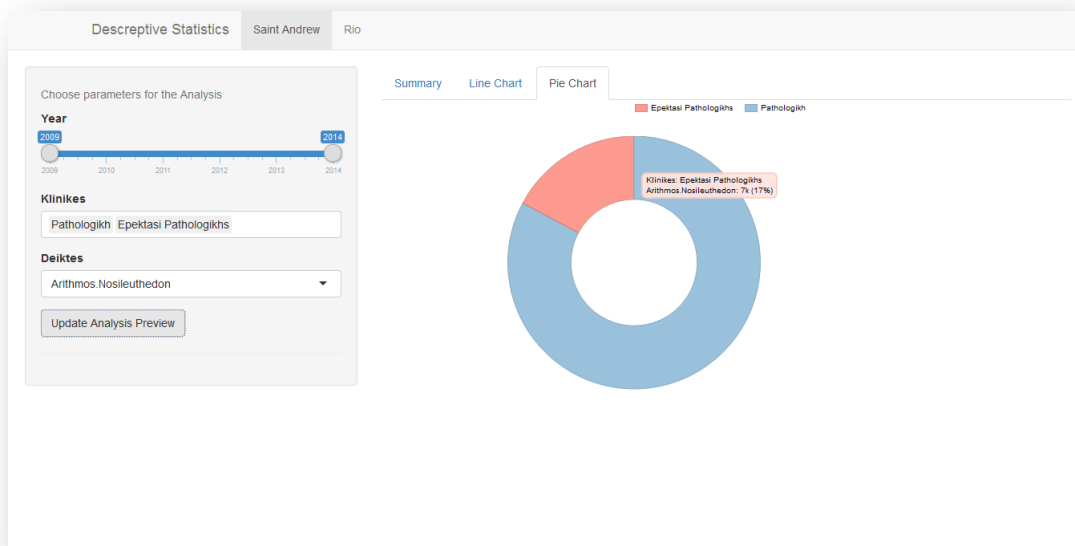
Εικόνα 22 Πάνελ επεξεργασίας παραμέτρων



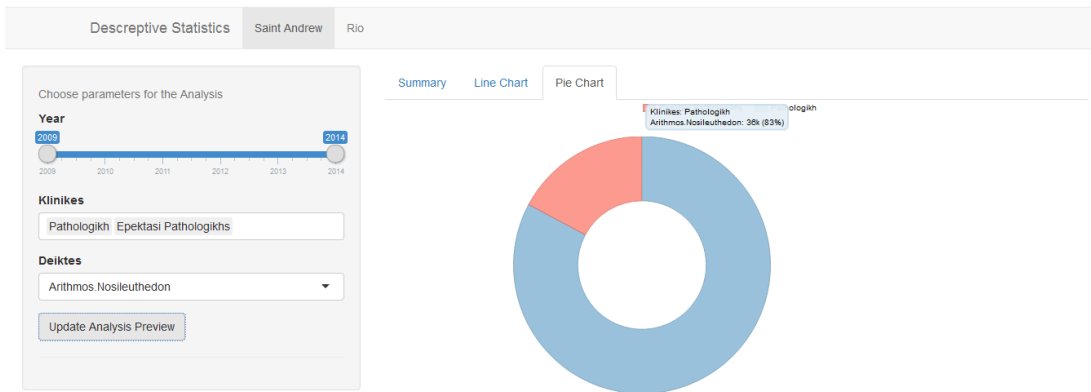
Εικόνα 23 Παρουσίαση περιληπτικών στοιχείων



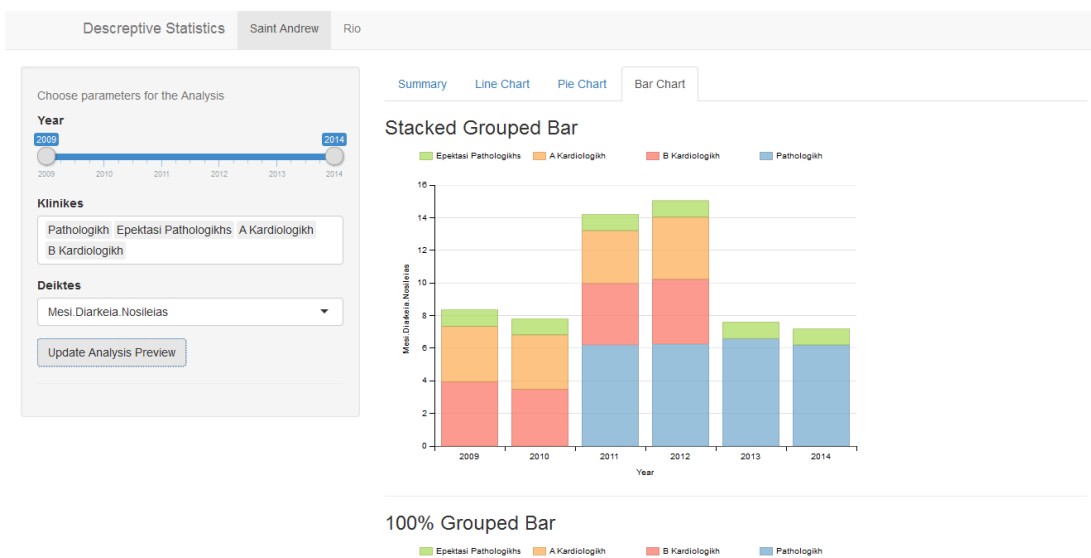
Εικόνα 24 Γράφημα Γραμμής (Line Chart)



Εικόνα 25 Γράφημα Πίτας (Pie Chart) 1

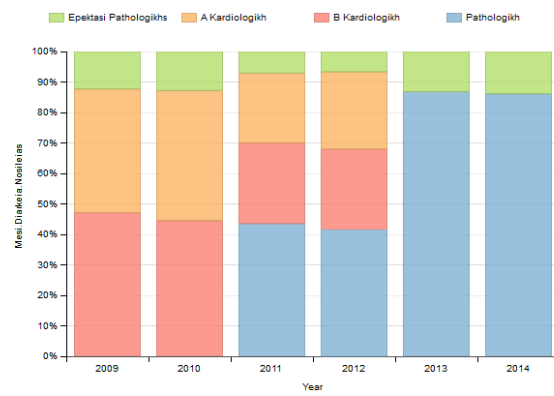


Εικόνα 26 Γράφημα Πίτας (Pie Chart) 2



Εικόνα 27 Ραβδόγραμμα βάση απόλυτων τιμών και χρονολογίας (Stacked Grouped Bar)

100% Grouped Bar



Εικόνα 28 Ραβδόγραμμα βάση ποσοσטיαίων τιμών και χρονολογίας (100% Stacked Grouped Bar)

Συμπεράσματα και Μελλοντικές Επεκτάσεις

Μέσα από τους δείκτες αξιολόγησης υγείας κάθε χώρα σχηματίζει την θέση της, ως προς τον συγκεκριμένο τομέα, δίπλα στις άλλες χώρες της Ευρωπαϊκής Ένωσης, ενώ ταυτόχρονα είναι σε κατάσταση να κάνει εσωτερική αξιολόγηση σε θέματα βελτιστοποίησης του τομέα της υγείας και σε συνδυασμό με μεθόδους προγνωστικής μοντελοποίησης, οι δείκτες, μπορούν να δώσουν μια ιδέα σε κάθε ενδιαφερόμενο ως προς την μελλοντική κατάσταση της υγείας οποιασδήποτε χώρας. Ενέργεια που δυστυχώς δεν εφαρμόζεται από το ελληνικό κράτος, αποτελεσματικά, για την αντιμετώπιση της κρίσης στον τομέα της υγείας.

Μεγάλη διαφορά παρουσιάζεται στην προσβασιμότητα και την ποιότητα των δεδομένων μεταξύ ελληνικών και παγκόσμιων βάσεων δεδομένων. Συγκεκριμένα, για τους δείκτες που χρησιμοποιήσαμε, η διαδικασία που χρειάστηκε για την συλλογή των ελληνικών δεικτών νοσοκομείων διήρκησε μήνες και τα δεδομένα δεν κάλυπταν πλήρως τις ανάγκες μας καθώς δεδομένα προ του 2009 είναι δύσκολο να ευρεθούν λόγω της μη ηλεκτρονικής καταγραφής τους. Η συλλογή παγκόσμιων δεικτών, από την άλλη, είναι άμεσα διαθέσιμη στο διαδίκτυο από οποιονδήποτε, έχοντας περισσότερες διαθέσιμες πληροφορίες για τον κάθε δείκτη, με τη χρήση μεταδεδομένων, και ο χρονικός διαθέσιμος ορίζοντας είναι σαφώς αρκετά μεγαλύτερος.

Έχοντας εξετάσει και εφαρμόσει τις τεχνικές της Απλής & Πολλαπλής Γραμμικής Παλινδρόμησης, Νευρωνικών Δικτύων και ανάλυσης χρονοσειρών με μοντέλα ARIMA στους δείκτες αξιολόγησης, καταλήγουμε ότι για πρόβλεψη άγνωστων μελλοντικών τιμών οι μέθοδοι των Νευρωνικών Δικτύων και των μοντέλων ARIMA είναι οι πιο χρήσιμες καθώς οι αλγόριθμοι που χρησιμοποιούν είναι σχεδιασμένοι ιδανικά για τέτοιες περιπτώσεις. Επιπλέον, απέδωσαν καλύτερα αποτελέσματα και παρουσίασαν ικανοποιητικές τιμές σφαλμάτων και διαστήματα εμπιστοσύνης.

Για την παρούσα πτυχιακή έχουμε χωρίσει τις προοπτικές εξέλιξής της σε τρία σημεία:

1. Συγκέντρωση της θεματολογίας αυστηρά μόνο στην περίπτωση των ελληνικών δημόσιων δεικτών αξιολόγησης υγείας νοσοκομείων, έτσι ώστε να υπάρξει μια καλύτερη και συγκεντρωμένη εικόνα για το σε ποια κατάσταση είναι και θα είναι η δημόσια υγεία στην Ελλάδα.
2. Εστίαση στους παγκόσμιους δείκτες αξιολόγησης υγείας ανά χώρα, ξεχωριστά με την διεξαγωγή έρευνας η οποία θα σχετίζεται με το πώς δείκτες υψηλής σημασίας επηρεάζουν την κατάταξη χωρών σε διάφορες λίστες διεθνών οργανισμών και πρόβλεψη των αυξομειώσεών τους.
3. Συνδυαστική χρήση μοντέλων πρόβλεψης και των δύο παραπάνω περιπτώσεων μεμονωμένα ή συνδυαστικά, καθώς είναι πιθανόν να δοθούν καλύτερης ποιότητας τιμές και συμπεράσματα.

Βιβλιογραφία

- Bari, Annase, Chaouchi, Mohamed και Jung, Tommy. 2014.** *Predictive Analytics For Dummies*. s.l. : John Wiley & Sons, Inc, 2014.
- Beck, Marcus W.** R is my friend. *Visualizing neural networks from the nnet package*. [Ηλεκτρονικό] <https://beckmw.wordpress.com/2013/03/04/visualizing-neural-networks-from-the-nnet-package/>.
- Bowerman, Bruce και O'Connell, Richard. 1987.** *Time series forecasting : unified concepts and computer implementation*. 2nd ed. Boston : Duxbury Press, 1987.
- Chambers, John C, Mullick, Satinder K και Smith, Donald D. 1971.** *How to Choose the Right Forecasting Technique*. s.l. : Harvard Business Review, 1971.
- Chatfield, Chris. 2000.** *Time-Series Forecasting*. s.l. : CHAPMAN & HALL/CRC, 2000.
- Crockett, David. 2013.** Using Predictive Analytics in Healthcare: Technology Hype vs. Reality. *Health Catalyst*. 2013.
- DeLurgio. 1998.** *Forecasting Principles and Applications*. s.l. : Irwin/McGraw-Hill, 1998.
- Gopalakrishma, Palem. 2013.** *The Practice of Predictive Analytics in Healthcare*. 2013.
- Guazzelli, Alex. 2011.** Predictive analytics in healthcare, The importance of open standards. *developerWorks*. 29 Νοεμβρίου 2011.
- Günthe, Frauke και Fritsch, Stefan. 2010.** neuralnet: Training of Neural Networks. *The R Journal*. 2010.
- Hagan , Martin T, και συν. 1996.** *Neural Network Design*. s.l. : PWS Publishing Company, 1996.
- Han, Jiawei και Kamber, Micheline. 2006.** *Data Mining Concepts and Techniques*. 2006.
- Hand, David, Heikki, Mannila και Padhraic, Smyth. 2001.** *Principles of Data Mining*. s.l. : The MIT Press, 2001.
- Haykin, Simon. 1999.** *Neural Networks and Learning*. s.l. : Prentice Hall International, Inc., 1999.
- Hoshmand, A. Reza. 2010.** *Business Forecasting: A practical Approach*. Second Edition. 270 Madison Avenue, New York : Routledge, 2010.
- Hyndman, Rob J και Athanasopoulos, George.** *Forecasting: Principles and Practice*.
- Keller, Gerald. 2010.** *Στατιστική για Οικονομικά και Διοίκηση Επιχειρήσεων*. Θεσσαλονίκη : Επίκεντρο Α.Ε., 2010.

Makridakis, Spyros, McGee, Victor E και Wheelwright, Steven C. 1983. *Forecasting : methods and applications.* 2nd ed. New York : Wiley Publishing, 1983.

Mohammed, Zaki J. και Wagner, Meira Jr. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms.* s.l. : Cambridge University Press, 2014.

Paradis, Emmanuel. 2005. *R For Beginners.* France : s.n., 2005.

Pearson, Andrew. 2012. *Predictive Analytics For The Healthcare Industry.* 2012.

Schick, Charlie. 2012. *The predictive power of big data in healthcare.* s.l. : IBM, 2012.

Tan, P.N., Steinbach, M. και Kumar, V. 2010. *Introduction to Data Mining.* [μεταφρ.] Β. Βερούκιος και Σ. Σουραβλάς. s.l. : Addison Wesley, 2010.

Walter L. Perry, Brian McInnis, Carter C. Price, Susan C. Smith, John S. Hollywood. 2013. *PREDICTIVE POLICING: The Role of Crime Forecasting in Law Enforcement Operations.* s.l. : RAND Corporation, 2013.

WHO, Παγκόσμιος Οργανισμός Υγείας. 1976. <http://www.who.int/about/en/>. *Παγκόσμιος Οργανισμός Υγείας.* [Ηλεκτρονικό] 1976.

Wikipedia. Forecasting.
https://en.wikipedia.org/wiki/Forecasting#Qualitative_vs._quantitative_methods.
[Ηλεκτρονικό]

Witten, Ian H και Frank, Eibe. 2005. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems).* s.l. : Morgan Kaufmann Publishers Inc., 2005.

Αντζουλάτος Γεράσιμος. 2012. *Τεχνητά Νευρωνικά Δίκτυα Επιβλεπόμενης Μάθησης Πολυστρωματικά Perceptron.* 2012.

Βλαχάβας, Ιωάννης, και συν. 2006. *Τεχνητή Νοημοσύνη.* Θεσσαλονίκη : Πανεπιστημίου Μακεδονίας, 2006.

Γεωργιάδης, Γεώργιος Ι. 2013. *Σύγκριση και αξιολόγηση μεθόδων πρόβλεψης για εκτίμηση κρίσεων στο ελληνικό χρηματοπιστωτικό σύστημα.* Αθήνα : Εθνικό Μετσόβιο Πολυτεχνείο, 2013. Διπλωματική Εργασία.

Γκουργκούλη, Ελένη, και συν. 2007. *Νοσηλευτική Θεωρία-Εργαστήριο.* Αθήνα : Οργανισμός Εκδόσεων Διδακτικών Βιβλίων, 2007, Κεφάλαιο 1.

Γραβάνης, Σεραφείμ Ε. 2012. *Πρόβλεψη χρηματοοικονομικών δεικτών με χρήση Τεχνητών Νευρωνικών Δικτύων.* s.l. : ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ, ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ, 2012. Διπλωματική Εργασία.

ΕΘΝΙΚΗ ΣΧΟΛΗ ΔΗΜΟΣΙΑΣ ΥΓΕΙΑΣ . 2012. *Ανάλυση δεικτών επίδοσης νοσοκομειακών μονάδων του ΕΣΥ.* s.l. : ΕΘΝΙΚΗ ΣΧΟΛΗ ΔΗΜΟΣΙΑΣ ΥΓΕΙΑΣ –

ΤΟΜΕΑΣ ΟΙΚΟΝΟΜΙΚΩΝ ΤΗΣ ΥΓΕΙΑΣ, 2012. Τελική Αναφορά Αποτελεσμάτων.

Ζάβρας. <http://www.stat-athens.aueb.gr/~jpan/diatrives/Zavras/chapter5.pdf>.

Καχριμάνη, Ελευθερία. 2012. *Πρόβλεψη Τεχνικών Χρονοσειρών με Οικονομετρικές Μεθόδους και Νευρωνικά Δίκτυα, Εφαρμογή μοντέλων σε εκπαιδευτικά Portal.* s.l. : Πανεπιστήμιο Πηραιώς, Τμήμα Ψηφιακών Συστημάτων, 2012. Διπλωματική Εργασία.

Κίνγκ, Ροβέρτος. 2004. *Ευφύης Έλεγχος.* Πάτρα : Τζιόλα, 2004.

ΜΚΟ Κατά του Καρκίνου Be Strong. <http://www.bestrong.org.gr/el/organization/aboutus/income/>.

Μυλωνά, Δόμνα. 2013. *Πληροφοριακά Συστήματα Υγείας.* Αθήνα : Πανεπιστήμιο Πειραιά, 2013.

Μωραΐτη, Ειρήνη. 2009. *Αξιολόγηση της λειτουργίας και της οικονομικής αποδοτικότητας των δημόσιων νοσοκομείων της περιφέρειας Κρήτης.* s.l. : Πανεπιστήμιο Κρήτης, 2009. ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ.

Παπαδόπουλος, Δημήτριος. 2011. *Τεχνικές Εξόρυξης Δεδομένων σε Σύστημα Υποστήριξης Ακαδημαϊκών Επαγγελματικών Αποφάσεων.* s.l. : ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ, Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων, 2011. Διπλωματική Εργασία.

Παπακωστίδη, Α και Τσουκαλάς, Ν. 2012. *Η Ποιότητα στις Υπηρεσίες Υγείας.* www.mednet.gr. [Ηλεκτρονικό] 2012.

Πετρόπουλος Φώτιος . 2015. *Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου Όρου (ARIMA). Τεχνικές Προβλέψεων.* s.l. : Εθνικό Μετσόβιο Πολυτεχνείο Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ, 2015.

Πετρόφ, Γιάννης, Τζωρτζάκης, Κώστας και Τζωρτζάκη, Αλεξία. 2002. *Μάρκετινγκ Μάνατζμεντ η Ελληνική Προσέγγιση.* Αθήνα : Rosili, 2002.

Πονοζάχος, Δημήτριος. 2013. *Χρηματοοικονομικές Καταστάσεις Επιχειρησιακών Μονάδων που δεν επιδιώκουν Κέρδος: Σύνταξη & Αξιολόγηση στο πλαίσιο των Ελληνικών Μονάδων Υγείας.* Θεσσαλονίκη : ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ, 2013. ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ.

Σταυρόπουλος, Νικόλαος. 2009. *Ποιότητα Υπηρεσιών Υγείας Μελέτη Ικανοποίησης Ασθενών.* <http://digilib.lib.unipi.gr/dspace/handle/unipi/3461>. [Ηλεκτρονικό] 2009.

Υγειομονικός Χάρτης. <http://ygeiamap.gov.gr/>. *Υγειομονικός Χάρτης.* [Ηλεκτρονικό]

Φαράτζιαν, Αλίσ. 2007. *Στρατηγική και Ποιότητα Μονάδων και Υπηρεσιών Υγείας.* s.l. : Πανεπιστήμιο Πειραιώς, Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων, 2007. Διπλωματική Εργασία.

Παράρτημα

Κώδικας Απλής Γραμμικής Παλινδρόμησης

```
library('car')
pathname = paste(getwd(), "worldbank_final.csv", sep="/")
dset <- read.csv2(pathname, dec=',',header = TRUE, sep=';',na='na')
dset.eu <- subset(dset, dset$Europe=="EU")
dset.noEU <- subset(dset, dset$Europe=="NoEU")
```

#1. Creation of training and testing data sets

```
dset.eu.train <- dset.eu[ dset.eu$Year <= 2012 ,
c("Health.expenditure..total....of.GDP.", "Life.expectancy.at.birth.
.total..years.",
"Year", "Country.Name")]
dset.eu.test <- dset.eu[ dset.eu$Year >= 2013,
c("Health.expenditure..total....of.GDP.", "Life.expectancy.at.birth.
.total..years.",
"Year", "Country.Name") ]
```

```
names(dset.eu.train) <- c("Health.Expenditure.Total",
"Life.Expectancy.Birth.Total", "Year", "Country.Name")
names(dset.eu.test) <-
c("Health.Expenditure.Total", "Life.Expectancy.Birth.Total", "Year",
"Country.Name")
```

```
dset.noEU.train <- dset.noEU[ as.vector(dset.noEU$Year) <= 2012,
c("Health.expenditure..total....of.GDP.",
"Life.expectancy.at.birth..total..years.", "Year", "Country.Name") ]
dset.noEU.test <- dset.noEU[ as.vector(dset.noEU$Year) >= 2013,
c("Health.expenditure..total....of.GDP.", "Life.expectancy.at.birth.
.total..years.", "Year", "Country.Name") ]
```

```
names(dset.noEU.train) <- c("Health.Expenditure.Total",
"Life.Expectancy.Birth.Total", "Year", "Country.Name")
names(dset.noEU.test) <- c("Health.Expenditure.Total",
"Life.Expectancy.Birth.Total", "Year", "Country.Name")
```

#2. Forecast formula and model creation

```
fmla.eu <- as.formula(Life.Expectancy.Birth.Total ~
Health.Expenditure.Total)
```

```

s.regmodel.eu <- lm(fmla.eu, data = dset.eu.train)

fmla.noeu <- as.formula(Life.Expectancy.Birth.Total ~
Health.Expenditure.Total)
s.regmodel.noeu <- lm(fmla.noeu, data = dset.noeu.train)

#3. Coefficient, Summary and Influence plotting

summary.eu <- summary(s.regmodel.eu)
coefs.eu <- s.regmodel.eu$coefficients

res.cor.eu <- matrix( c( summary.eu$r.squared*100,
summary.eu$adj.r.squared*100, AIC(s.regmodel.eu) ), c(1,3) )
colnames(res.cor.eu) <- c("R-Squared", "Adjusted R-Squared", "AIC" )
print(res.cor.eu[1,], row.names=FALSE, digits=5)

summary.noeu <- summary(s.regmodel.noeu)
coefs.noeu <- s.regmodel.noeu$coefficients

res.cor.noeu <- matrix( c( summary.noeu$r.squared*100,
summary.eu$adj.r.squared*100, AIC(s.regmodel.noeu) ), c(1,3) )
colnames(res.cor.noeu) <- c("R-Squared", "Adjusted R-Squared", "AIC")
print(res.cor.noeu[1,], row.names=FALSE, digits=5)

#3.1
png("LM.Life.Health.EU.png")
par(mfrow=c(2,2))
plot(s.regmodel.eu, which = c(1,3), id.n=10)
plot(s.regmodel.eu, which = c(2,4), id.n=10)
dev.off()

par(mfrow=c(1,1))
ci.eu <- confint(s.regmodel.eu)
print(ci.eu)

png("LM.Life.Health.noEU.png")
par(mfrow=c(2,2))
plot(s.regmodel.noeu, which = c(1,3), id.n=10)
plot(s.regmodel.noeu, which = c(2,4), id.n=10)
dev.off()

par(mfrow=c(1,1))
ci.noeu <- confint(s.regmodel.noeu)
print(ci.noeu)

#3.2 Influence Plot
p<-influencePlot(s.regmodel.eu, main="Influence Plot for EU
Countries", scale=5, id.n=10, id.cex=.75, id.col='blue',
sub="Circle size is proportional to Cook's distance")
results <- as.data.frame( cbind('Observation' = rownames(p), p) )

z<-influencePlot(s.regmodel.noeu, main="Influence Plot for non EU

```

```

Countries", scale=5, id.n=10, id.cex=.75, id.col='blue',
sub="Circle size is proportional to Cook's distance")
results <- as.data.frame( cbind('Observation' = rownames(z), z) )

# Outliers Detection
outliers.eu <- outlierTest(s.regmodel.eu)

outliers.noeu <- outlierTest(s.regmodel.noeu)
#4. The use of predict() function / Plotting the prediction

pred.eu <- predict(s.regmodel.eu, dset.eu.test)

plot( x=dset.eu.test$Life.Expectancy.Birth.Total, y=pred.eu, xlab =
'Life Expectancy', ylab='Predicted Values')

library('ggplot2')
df <- data.frame( dset.eu.test$Life.Expectancy.Birth.Total,
pred.eu, dset.eu.test$Country.Name )
names(df) <-c("Actual.Life.Expectancy", "Prediction", "Country.Name")

ggplot(df, aes_string(x="Actual.Life.Expectancy", y="Prediction"))
+ geom_point(aes(colour = df$Country.Name))+geom_abline(col="red")

pred.noeu <- predict(s.regmodel.noeu, dset.noeu.test)

plot( x=dset.noeu.test$Life.Expectancy.Birth.Total, y=pred.noeu,
xlab = 'Life Expectancy', ylab='Predicted Values')

library('ggplot2')
df1 <- data.frame( dset.noeu.test$Life.Expectancy.Birth.Total,
pred.noeu, dset.noeu.test$Country.Name )
names(df1)<-c("Actual.Life.Expectancy", "Prediction", "Country.Name")

ggplot(df1, aes_string(x="Actual.Life.Expectancy",y="Prediction"))
+ geom_point(aes(colour = df1$Country.Name))+geom_abline(col="red")

#Descriptive Statistics
library(pastecs)

summary.lm(s.regmodel.eu)
summary.lm(s.regmodel.noeu)

stat.desc(s.regmodel.eu$residuals)
stat.desc(s.regmodel.noeu$residuals)

#for indicators
stat.desc (dset.noeu)
stat.desc (dset.eu)

#Root Mean Square Error
rmse.eu = sqrt(mean((pred.eu -
dset.eu$Life.expectancy.at.birth..total..years.)^2, na.rm = TRUE))

```

```
rmse.noeu = sqrt(mean((pred.noeu -
dset.noeu$Life.expectancy.at.birth..total..years.)^2, na.rm =
TRUE))
```

Κώδικας Πολλαπλής Παλινδρόμησης

```
library('car')
pathname = paste(getwd(), "worldbank_final.csv", sep="/")
dset <- read.csv2(pathname, dec=',', header = TRUE, sep=';', na='na')

dset.eu <- subset(dset, dset$Europe=="EU")
dset.noeu <- subset(dset, dset$Europe=="NoEU")

#1. Creation of training and testing data sets

dset.eu.train <- dset.eu[ dset.eu$Year <= 2012 ,
c("Birth.rate..crude..per.1.000.people.",
"Death.rate..crude..per.1.000.people.",
"Life.expectancy.at.birth..total..years.", "Number.of.infant.deaths"
, "Mortality.rate..infant..per.1.000.live.births.",
"Fertility.rate..total..births.per.woman.", "Year", "Country.Name")]

dset.eu.test <- dset.eu[ dset.eu$Year >= 2013,
c("Birth.rate..crude..per.1.000.people.",
"Death.rate..crude..per.1.000.people.",
"Life.expectancy.at.birth..total..years.",
"Number.of.infant.deaths", "Mortality.rate..infant..per.1.000.live.b
irths.", "Fertility.rate..total..births.per.woman.",
"Year", "Country.Name")]

names(dset.eu.train) <- c("Birth.rate.per.1.000.people.",
"Death.rate.per.1.000.people.", "Life.expectancy.total.", "Number.of.
infant.deaths", "Mortality.rate.infant.per.1.000.live.births.",
"Fertility.rate.total.births.per.woman.", "Year", "Country.Name")

names(dset.eu.test) <- c("Birth.rate.per.1.000.people.",
"Death.rate.per.1.000.people.", "Life.expectancy.total.",
"Number.of.infant.deaths",
"Mortality.rate.infant.per.1.000.live.births.",
"Fertility.rate.total.births.per.woman.", "Year", "Country.Name")
```

```

dset.noeu.train <- dset.noeu[ as.vector(dset.noeu$Year) <= 2012, c(
  "Birth.rate..crude..per.1.000.people.",
  "Death.rate..crude..per.1.000.people.",
  "Life.expectancy.at.birth..total..years.",
  "Number.of.infant.deaths",
  "Mortality.rate..infant..per.1.000.live.births.",
  "Fertility.rate..total..births.per.woman.", "Year", "Country.Name")]

dset.noeu.test <- dset.noeu[ as.vector(dset.noeu$Year) >= 2013,

c("Birth.rate..crude..per.1.000.people.", "Death.rate..crude..per.1.
000.people.", "Life.expectancy.at.birth..total..years.",
"Number.of.infant.deaths",
"Mortality.rate..infant..per.1.000.live.births.",
"Fertility.rate..total..births.per.woman.", "Year", "Country.Name")]

names(dset.noeu.train) <- c("Birth.rate.per.1.000.people.",
"Death.rate.per.1.000.people.", "Life.expectancy.total.",
"Number.of.infant.deaths",
"Mortality.rate.infant.per.1.000.live.births.",
"Fertility.rate.total.births.per.woman.", "Year", "Country.Name")

names(dset.noeu.test) <- c("Birth.rate.per.1.000.people.",
"Death.rate.per.1.000.people.", "Life.expectancy.total.",
"Number.of.infant.deaths",
"Mortality.rate.infant.per.1.000.live.births.",
"Fertility.rate.total.births.per.woman.", "Year", "Country.Name")

#2. Forecast formula and model creation

fmla.eu <- as.formula(Mortality.rate.infant.per.1.000.live.births.
~ Birth.rate.per.1.000.people. +Death.rate.per.1.000.people. +
Life.expectancy.total. + Number.of.infant.deaths +
Fertility.rate.total.births.per.woman.)

s.regmodel.eu <- lm(fmla.eu, data = dset.eu.train)

fmla.noeu <-
as.formula(Mortality.rate.infant.per.1.000.live.births. ~
Birth.rate.per.1.000.people. +Death.rate.per.1.000.people. +
Life.expectancy.total. + Number.of.infant.deaths +
Fertility.rate.total.births.per.woman.)

s.regmodel.noeu <- lm(fmla.noeu, data = dset.noeu.train)

#3. Coefficient, Summary and Influence plotting

summary.eu <- summary(s.regmodel.eu)
coefs.eu <- s.regmodel.eu$coefficients

res.cor.eu <- matrix( c( summary.eu$r.squared*100,
summary.eu$adj.r.squared*100, AIC(s.regmodel.eu) ), c(1,3) )

```

```

colnames(res.cor.eu) <- c("R-Squared", "Adjusted R-Squared", "AIC" )
print(res.cor.eu[1,], row.names=FALSE, digits=5)

summary.noeu <- summary(s.regmodel.noeu)
coefs.noeu <- s.regmodel.noeu$coefficients

res.cor.noeu <- matrix( c( summary.noeu$r.squared*100,
summary.noeu$adj.r.squared*100, AIC(s.regmodel.noeu) ), c(1,3) )
colnames(res.cor.noeu) <-c("R-Squared", "Adjusted R-Squared", "AIC" )
print(res.cor.noeu[1,], row.names=FALSE, digits=5)

#3.1
png("LM.Life.Health.EU.MR.png")
par(mfrow=c(2,2))
plot(s.regmodel.eu, which = c(1,3), id.n=10)
plot(s.regmodel.eu, which = c(2,4), id.n=10)
dev.off()

par(mfrow=c(1,1))
ci.eu <- confint(s.regmodel.eu)
print(ci.eu)

png("LM.Life.Health.noEU.MR.png")
par(mfrow=c(2,2))
plot(s.regmodel.noeu, which = c(1,3), id.n=10)
plot(s.regmodel.noeu, which = c(2,4), id.n=10)
dev.off()
par(mfrow=c(1,1))
ci.noeu <- confint(s.regmodel.noeu)
print(ci.noeu)

#3.2 Influence Plot
p<-influencePlot(s.regmodel.eu, main="Influence Plot", scale=5,
id.n=10, id.cex=.75, id.col='blue', sub="Circle size is
proportional to Cook's distance")
results <- as.data.frame( cbind('Observation' = rownames(p), p))

z<-influencePlot(s.regmodel.noeu, main="Influence Plot", scale=5,
id.n=10, id.cex=.75, id.col='blue', sub="Circle size is
proportional to Cook's distance")
results <- as.data.frame( cbind('Observation' = rownames(z), z) )

# Outliers Detection
outliers.eu <- outlierTest(s.regmodel.eu)
outliers.noeu <- outlierTest(s.regmodel.noeu)

#4. The use of predict() function / Plotting the prediction
pred.eu <- predict(s.regmodel.eu, dset.eu.test)

library('ggplot2')
df <- data.frame(

```

```

dset.eu.test$Mortality.rate.infant.per.1.000.live.births., pred.eu,
dset.eu.test$Country.Name )
names(df) <- c( "Actual.IMR", "Prediction", "Country.Name")
ggplot(df, aes_string(x="Actual.IMR", y="Prediction") ) +
geom_point(aes(colour = df$Country.Name)) + geom_abline(col="red")

pred.noeu <- predict(s.regmodel.noeu, dset.noeu.test)

library('ggplot2')
df1 <- data.frame(
dset.noeu.test$Mortality.rate.infant.per.1.000.live.births.,
pred.noeu, dset.noeu.test$Country.Name )
names(df1) <- c( "Actual.IMR", "Prediction", "Country.Name")
ggplot(df1, aes_string(x="Actual.IMR", y="Prediction") ) +
geom_point(aes(colour = df1$Country.Name)) + geom_abline(col="red")

#Descriptive Statistics
library(pastecs)

summary.lm(s.regmodel.eu)
summary.lm(s.regmodel.noeu)
stat.desc(s.regmodel.eu$residuals)
stat.desc(s.regmodel.noeu$residuals)

#for indicators
stat.desc (dset.noeu)
stat.desc (dset.eu)

#Root Mean Square Error

rmse.eu = sqrt(mean((pred.eu -
dset.eu$Mortality.rate..infant..per.1.000.live.births)^2, na.rm =
TRUE))

rmse.noeu = sqrt(mean((pred.noeu -
dset.noeu$Mortality.rate..infant..per.1.000.live.births)^2, na.rm =
TRUE))

```

Κώδικας Νευρωνικών Δικτύων

```

library('nnet')
library('caret')
library('ggplot2')
library('readxl')

source('gar.fun.R')
source('plot.nnet.R')

# load dataset
pathname = paste(getwd(), "worldbank_final_summarised.xlsx",
sep="/")
dset <- read_excel(pathname, sheet = 1 ,col_names = TRUE, na='na')

```



```

#-----
# Preprocess dataset in order to use for create a Feed-forward
Neural Network

# Discretize the variable "Life expectancy at birth total years".
If an instance has value
# lower than the 75 years old (mean of Life expectancy at birth
total years) is considered as
# "LOW" else belong to "HIGH" class.
#
dset$Life.expectancy.class <- ifelse(dset$"Life expectancy at birth
total years"<=75,"LOW", "HIGH")

# Split dataset depending on EU countries and non EU countries
dset.eu <- subset(dset, dset$Europe=="EU")
dset.noeu <- subset(dset, dset$Europe=="NoEU")

#.....
....
# Training / Testing Sets (European Union countries)
dset.eu.train <- dset.eu[dset.eu$Year<= 2012, ]

dset.eu.test <- dset.eu[dset.eu$Year == 2013, ]

# Training / Testing Sets (Non European Union countries)
dset.noeu.train <- dset.noeu[dset.noeu$Year<= 2012, ]
dset.noeu.test <- dset.noeu[dset.noeu$Year == 2013, ]

#-----
#
# First Hypothesis: NN to solve regression problem. Response
variable is "Life expectancy at birth total years"
#
# Create and Train Neural Networks for Prediction - Europe

# Create input patterns and targets for NN
input.eu <- dset.eu.train[, c("Hospital beds per 1000 people",
                             "Physicians per 1000 people", "Nurses
and midwives per 1000 people",
                             "Health expenditure total percent of
GDP"
                             )]
input.eu$Year <- as.factor(input.eu$Year)
#
output.eu <- dset.eu.train[,c("Life expectancy at birth total
years")] # Life.expectancy.class

#
input.eu.ts <- dset.eu.test[, c("Hospital beds per 1000 people",
                              "Physicians per 1000 people",

```

```

"Nurses and midwives per 1000 people",
                                "Health expenditure total percent
of GDP"
                                )]
input.eu.ts$Year <- as.factor(input.eu.ts$Year)

output.eu.ts <- dset.eu.test[,c("Life expectancy at birth total
years")] # Life.expectancy.class

#
# # Train NN
trainingdata1 <- nnet(input.eu, output.eu, size=5, maxit=1e+6,
abstol=1e-06, reltol=1e-10,
                                softmax=FALSE, censored = FALSE, entropy =
FALSE,
                                linout=TRUE, trace = TRUE, rang=0.1, decay =
0.05)

par( mar=c(4,5,4,2), oma=c(0,0,0,0), family='serif')
plot.nnet(trainingdata1, pos.col='darkgreen', neg.col='darkblue',
alpha.val=0.5, rel.rsc=10,
                                circle.cex=7, cex.val=1, circle.col='brown', all.in = T,
all.out = T)

pred.eu <- predict(trainingdata1, input.eu.ts, type="raw")

RMSE.eu <- sqrt(mean( (output.eu.ts - pred.eu)^2 ))
plot(x=output.eu.ts, y=pred.eu, type='p', xlab='Target values',
ylab='Prediction values')

# Prediction - Non Europe

input.noeu <- dset.noeu.train[,
                                c("Hospital beds per 1000 people",
                                "Physicians per 1000 people",
"Nurses and midwives per 1000 people",
                                "Health expenditure total percent
of GDP")]

input.noeu$Year <- as.factor(input.eu$Year)

output.noeu <- dset.noeu.train[,c("Life expectancy at birth total
years")]

input.noeu.ts <- dset.noeu.test[,
                                c("Hospital beds per 1000 people",
                                "Physicians per 1000 people",
"Nurses and midwives per 1000 people",
                                "Health expenditure total percent
of GDP")]

input.noeu.ts$Year <- as.factor(input.noeu.ts$Year)

```

```

output.noeu.ts <- dset.noeu.test[,c("Life expectancy at birth total
years")]

trainingdata1.noeu <- nnet(input.noeu, output.noeu, size=5,
maxit=10000, abstol=1e-06, reltol=1e-10,
                        softmax=FALSE, censored = FALSE,
                        linout=TRUE, trace = TRUE, rang=0.1,
decay = 0.05)

par( mar=c(4,5,4,2), oma=c(0,0,0,0), family='serif')
plot.nnet(trainingdata1.noeu, pos.col='darkgreen',
neg.col='darkblue', alpha.val=0.5, rel.rsc=10,
          circle.cex=7, cex.val=1, circle.col='brown', all.in = T,
all.out = T)

pred.noeu <- predict(trainingdata1.noeu, input.noeu.ts, type="raw")

RMSE.noeu <- sqrt(mean( (output.noeu.ts - pred.noeu)^2 ))

plot(x=output.noeu.ts, y=pred.noeu, type='p', xlab='Target values',
ylab='Prediction values')

#-----
#
# Second Hypothesis: NN to solve classification problem. Response
variable is "Life.expectancy.class"
#
# Create and Train Neural Networks for Prediction - European Union

# Create input patterns and targets for NN
input.eu.cl <- dset.eu.train[, c("Hospital beds per 1000 people",
"Physicians per 1000 people", "Nurses
and midwives per 1000 people",
"Health expenditure total percent of
GDP"
)]
input.eu.cl$Year <- as.factor(input.eu$Year)
#
output.eu.cl <- dset.eu.train[,c("Life.expectancy.class")]
targets.output.eu.cl <-class.ind(output.eu.cl)

#
input.eu.cl.ts <- dset.eu.test[, c("Hospital beds per 1000 people",
"Physicians per 1000 people",
"Nurses and midwives per 1000 people",
"Health expenditure total percent
of GDP"
)]
input.eu.cl.ts$Year <- as.factor(input.eu.ts$Year)

```

```

output.eu.cl.ts <- dset.eu.test[,c("Life.expectancy.class")]
targets.output.eu.cl.ts <-class.ind(output.eu.cl.ts)

#
# # Train NN
trainingdata2 <- nnet(input.eu.cl, targets.output.eu.cl, size=5,
maxit=1e+6, abstol=1e-06, reltol=1e-10,
                    softmax=FALSE, censored = FALSE, entropy =
TRUE,
                    linout=FALSE, trace = TRUE, rang=0.1, decay =
0.05)

par( mar=c(4,5,4,2), oma=c(0,0,0,0), family='serif')
plot.nnet(trainingdata2, pos.col='darkgreen', neg.col='darkblue',
alpha.val=0.5, rel.rsc=10,
         circle.cex=7, cex.val=1, circle.col='brown', all.in = T,
all.out = T)

preds.eu.cl <- predict(trainingdata2, input.eu.cl.ts)

# Calculate the Confusion Matrix
real.category <- max.col(targets.output.eu.cl.ts)
predictions <- max.col(preds.eu.cl)
cf <- table(real.category, predictions)

Accuracy <- sum( diag(cf) )/sum(cf) * 100

# Create and Train Neural Networks for Prediction - Non European
Union

# Create input patterns and targets for NN
input.no.eu.cl <- dset.no.eu.train[, c("Hospital beds per 1000
people",
                                     "Physicians per 1000 people",
                                     "Nurses and midwives per 1000 people",
                                     "Health expenditure total
percent of GDP"
)]
#
output.no.eu.cl <- dset.no.eu.train[,c("Life.expectancy.class")]
targets.output.no.eu.cl <-class.ind(output.no.eu.cl)

#
input.no.eu.cl.ts <- dset.no.eu.test[, c("Hospital beds per 1000
people",
                                     "Physicians per 1000
people", "Nurses and midwives per 1000 people",
                                     "Health expenditure total
percent of GDP"
)]

```

```

output.noeu.cl.ts <- dset.noeu.test[,c("Life.expectancy.class")]
targets.output.noeu.cl.ts <-class.ind(output.noeu.cl.ts)

#
# # Train NN
trainingdata2.noeu <- nnet(input.noeu.cl, targets.output.noeu.cl,
size=5, maxit=1e+6, abstol=1e-06, reltol=1e-10,
                        softmax=FALSE, censored = FALSE, entropy
= TRUE,
                        linout=FALSE, trace = TRUE, rang=0.1,
decay = 0.05)

par( mar=c(4,5,4,2), oma=c(0,0,0,0), family='serif')
plot.nnet(trainingdata2.noeu, pos.col='darkgreen',
neg.col='darkblue', alpha.val=0.5, rel.rsc=10,
          circle.cex=7, cex.val=1, circle.col='brown', all.in = T,
all.out = T)

preds.noeu.cl <- predict(trainingdata2.noeu, input.noeu.cl.ts)

# Calculate the Confusion Matrix
real.category.noeu <- max.col(targets.output.noeu.cl.ts)
predictions.noeu <- max.col(preds.noeu.cl)
cf.noeu <- table(real.category.noeu, predictions.noeu)

Accuracy.noeu <- sum( diag(cf.noeu) )/sum(cf.noeu) * 100
confusionMatrix(cf)
confusionMatrix(cf.noeu)

```

Κώδικας μοντέλου ARIMA

```

library(forecast)

pathname = paste(getwd(), "worldbank_final.csv", sep="/")
dset <-read.csv2(pathname, dec=',', header = TRUE, sep=';',na='na')

ds.france <- dset[ dset$Country.Name=='France', ]
ds.sweden <- dset[ dset$Country.Name=='Sweden', ]

ds.switzerland <- dset[ dset$Country.Name=='Switzerland', ]
ds.russia <- dset[ dset$Country.Name=='Russian Federation', ]

ts.france <- ts(ds.france$Health.expenditure..total....of.GDP.,
start = 1984, end = 2013, frequency = 1)
ts.sweden <- ts(ds.sweden$Health.expenditure..total....of.GDP.,
start = 1984, end = 2013, frequency = 1)

ts.switzerland <-
ts(ds.switzerland$Health.expenditure..total....of.GDP., start =

```

```

1984, end = 2013, frequency = 1)
ts.russia <- ts(ds.russia$Health.expenditure..total....of.GDP.,
start = 1984, end = 2013, frequency = 1)

fit.france <- arima(ts.france, order=c(3,0,3))
fore.france <- forecast(fit.france, h=5)

fit.sweden <- arima(ts.sweden, order=c(3,0,3))
fore.sweden <- forecast(fit.sweden, h=5)

fit.switzerland <- arima(ts.switzerland, order=c(3,0,3))
fore.switzerland <- forecast(fit.switzerland, h=5)

fit.russia <- arima(ts.russia, order=c(3,0,3))
fore.russia <- forecast(fit.russia, h=5)

par(mfrow=c(1,2))
plot(fore.france, main = "Prediction France")
plot(fore.sweden, main = "Prediction Sweden")

par(mfrow=c(1,2))
plot(fore.switzerland, main = "Prediction Switzerland")
plot(fore.russia, main = "Prediction Russian Federation")
summary(fore.france)
summary(fore.sweden)
summary(fore.switzerland)
summary(fore.russia)

```

Κώδικας εφαρμογής με Shiny

ui.r

```

library("shiny")
library("car")

# load helper.R file
source("helper.R")

shinyUI(navbarPage("Descriptive Statistics",
  tabPanel("Saint Andrew",
    fluidPage(
      sidebarLayout(
        sidebarPanel(

```

```

        helpText("Choose parameters for the
Analysis"),
        sliderInput("year", "Year",min = 2009,
max = 2014,
        value = c(2009,2014),
step=1, sep = ""),

        selectInput("inp",
label = "Klinikes",
choices = c("All",
unique(dset$Klinikes)),
        selected = "All", multiple =
TRUE ),

        selectInput("out",
label = "Deiktes",
choices=c(names(dset[3:8])),
        multiple = FALSE),

        actionButton(inputId = 'startAnalysis',
label = 'Update Analysis Preview'),

        hr()

    ), #End SideBarPanel

    mainPanel(
        tabsetPanel(
            tabPanel("Summary",

verbatimTextOutput("summary1")

            ), #End tabPanel - Summary
            tabPanel("Line Chart",
                uiOutput("YZ"),

metricsgraphicsOutput("intlines", width = "100%", height = "400px")

            ), #End tabPanel - LineChart
            tabPanel("Pie Chart",

                dimpleOutput("pie")

            ), # End tabPanel - Pie Chart
            tabPanel("Bar Chart",

                h3("Stacked Grouped Bar"),
                dimpleOutput("barStacked"),
                hr(),
                h3("100% Grouped Bar"),

```

```

dimpleOutput("barStackedPerc")
    )
    ) #End tabsetPanel
  ) # End mainPanel
) # End sidebarLayout
) # End fluidPage
), # End tabPanel_St.Andrew

tabPanel("Rio",
  fluidPage(
    sidebarLayout(
      sidebarPanel(
        helpText("Choose parameters for the Analysis"),
sliderInput("yearrio", "Year",min = 2009, max = 2014,
  value = c(2009,2014), step=1, sep = ""),

        selectInput("inprio",
                    label = "Klinikes",
choices = c("All", unique(dset.rio$Klinikes)),
selected = "All", multiple = TRUE ),

        selectInput("outrio",
label = "Deiktes", choices=c(names(dset.rio[3:8])),
multiple = FALSE),

        actionButton(inputId =
'startAnalysisrio', label = 'Update Analysis Preview'),

        hr()

      ), #End SideBarPanel

      mainPanel(
        tabsetPanel(
          tabPanel("Summary",

verbatimTextOutput("summary1rio")

          ), #End tabPanel - Summary
        tabPanel("Line Chart",
          uiOutput("YZrio"),

metricsgraphicsOutput("intlinesrio", width = "100%", height =
"400px")

```



```

        ), #End tabPanel - LineChart
        tabPanel("Pie Chart",
                dimpleOutput("pierio")
        ), # End tabPanel - Pie Chart
        tabPanel("Bar Chart",
                h3("Stacked Grouped Bar"),
dimpleOutput("barStackedrio"),
                hr(),
                h3("100% Grouped Bar"),
dimpleOutput("barStackedPercrio")
        ) #End tabpanel - Bar Chart
    ) #End tabsetPanel
  ) # End mainPanel
) # End sidebarLayout
) # End fluidPage
) # End tabPanel_Rio

) # End navbarPage
) # End Shiny

```

server.r

```

library(shiny)
library(car)
library(datasets)

shinyServer(function(input, output, session){
  # subsetting dataset using user-defined criteria
  passData <- reactive({
    vars <- names(dset)
    new.dset <- subset(dset, select=vars %in% unlist(input$out) )
    new.dset <- as.data.frame( cbind( "Year"=dset$Year,
    "Klinikes"=dset$Klinikes, new.dset ) )
  })

```

```

new.dset <- new.dset[ new.dset$Year >= as.numeric(input$year[1])
&
                                new.dset$Year <=
as.numeric(input$year[2]) , ]
  if ( input$inp != "All" ){
    new.dset <- subset(new.dset, new.dset$Klinikes %in%
unlist(input$inp) )
  }
  return(new.dset)
})

output$summary1 <- renderPrint({

  # Re-run when button is clicked
  if (input$startAnalysis == 0){
    return() }
  else{
    isolate({

      dset.sa <- passData()
      ds <- subset(dset.sa, select=names(dset.sa) %in%
unlist(input$out) )

      s<-summary(ds)
      return( s )

    })
  }
})

output$YZ <- renderUI({
  if (is.null(input$out)) { return() }
  deiktes <- names(dset[3:8])
  deiktes <- deiktes[ deiktes != input$out ]
  checkboxGroupInput(inputId='YZ.indices',
label=h3('Variable(s):'),
                                choices=deiktes, selected = NULL, inline =
TRUE)
})

data.lines <- reactive({

  if ( !is.null(input$YZ.indices) ){
    vars1 <- list(input$out, input$YZ.indices)
  }else{
    vars1 <- list(input$out)
  }
  new.dset1 <- subset(dset, select=names(dset) %in% unlist(vars1) )
  new.dset1 <- as.data.frame( cbind( "Year"=dset$Year,
"Klinikes"=dset$Klinikes, new.dset1 ) )

  new.dset1 <- new.dset1[ new.dset1$Year >=
as.numeric(input$year[1]) &

```

```

                                new.dset1$Year <=
as.numeric(input$year[2]) , ]

  if ( input$inp != "All" ){
    new.dset1 <- subset(new.dset1, new.dset1$Klinikes %in%
unlist(input$inp) )
  }
  return(new.dset1)
})

output$intlines <- renderMetricsgraphics({
  if (input$startAnalysis == 0){
    return() }
  else{
    isolate({

      dset.lines <- data.lines()

      tid <- unlist(input$YZ.indices)
      if ( is.null(tid) ){
        inds <- list( input$out )
      }else if ( length(tid) ==1 ) {
        inds <- list( input$out, tid[1] )
      }else {
        inds <- list( input$out, tid[1], tid[2] )
      }

      if ( length(inds) == 1 )
      {
        tmp <- data.frame(year= dset.lines$Year,
d=as.numeric(dset.lines[,names(dset.lines) %in% inds[1] ] ))
        tmp %>%
          mjs_plot(x=year, y=d) %>%
          mjs_line() %>%
          mjs_add_baseline(mean(
dset.lines[,names(dset.lines) %in% input$out ]), "Mean")
      }
      else if ( length(inds) == 2 ) {
        tmp <- data.frame(year= dset.lines$Year,
d=as.numeric(dset.lines[, names(dset.lines) %in% inds[1] ] ),
                    d1=as.numeric(dset.lines[,
names(dset.lines) %in% inds[2] ] ) )
        tmp %>%
          mjs_plot(x=year, y=d) %>%
          mjs_line() %>%
          mjs_add_line(d1) %>%
          mjs_add_legend(legend = names(as.data.frame(inds)) )
      }
      else{
        tmp <- data.frame(year= dset.lines$Year,
d=as.numeric(dset.lines[, names(dset.lines) %in% inds[1] ] ),

```

```

d1=as.numeric(dset.lines[, names(dset.lines) %in% inds[2] ]),
d2=as.numeric(dset.lines[, names(dset.lines) %in% inds[3] ]) )
    tmp %>%
      mjs_plot(x=year, y=d) %>%
      mjs_line() %>%
      mjs_add_line(d1) %>%
      mjs_add_line(d2) %>%
      mjs_add_legend(legend = names(as.data.frame(inds)) )
  }
})
})
output$pie <- renderdimple({
  if (input$startAnalysis == 0){
    return() }
  else{
    isolate({
      dset.sa <- passData()
      legend_value = ifelse( length(input$inp) >=2 &
length(input$inp) <= 8, TRUE, FALSE )
      dimple(dset.sa,
        pMeasure=input$out,
        pCategory="Year",
        series="Klinikes",
        legend=legend_value,
        chartType="pie",
        ringInnerRadius="50%")
    })
  }
})
output$barStacked <- renderdimple({
  if (input$startAnalysis == 0){
    return() }
  else{
    isolate({
      dset.sa <- passData()
      legend_value = ifelse( length(input$inp) >=2 &
length(input$inp) <= 8, TRUE, FALSE )
      dimple(dset.sa,
        xCategory=c("Year"),
        yMeasure=input$out,
        series="Klinikes",
        legend=legend_value)
    })
  }
})

```

```

    })
  }
})

output$barStackedPerc <- renderdimple({

  if (input$startAnalysis == 0){
    return() }
  else{
    isolate({

      dset.sa <- passData()
      legend_value = ifelse( length(input$inp) >=2 &
length(input$inp) <= 8, TRUE, FALSE )

      dimple(dset.sa,
        xCategory=c("Year"),
        yPct=input$out,
        series="Klinikes",
        legend=legend_value)

    })
  }
})

passDatario <- reactive({

  varsrio <- names(dset.rio)
  new.dsetrio <- subset(dset.rio, select=varsrio %in%
unlist(input$outrio) )
  new.dsetrio <- as.data.frame( cbind( "Year"=dset.rio$Year,
"Klinikes"=dset.rio$Klinikes, new.dsetrio ) )
  new.dsetrio <- new.dsetrio[ new.dsetrio$Year >=
as.numeric(input$yearrio[1]) &
new.dsetrio$Year <=
as.numeric(input$yearrio[2]) , ]
  if ( input$inprio != "All" ){
    new.dsetrio <- subset(new.dsetrio, new.dsetrio$Klinikes %in%
unlist(input$inprio) )
  }
  return(new.dsetrio)
})

output$summary1rio <- renderPrint({

  # Re-run when button is clicked
  if (input$startAnalysisrio == 0){
    return() }
  else{
    isolate({

```

```

    dset.rio1 <- passDatario()
    dsrio <- subset(dset.rio1, select=names(dset.rio1) %in%
unlist(input$outrio) )
    srio<-summary(dsrio)
    return( srio )

  })
}
})

output$YZrio <- renderUI({
  if (is.null(input$outrio)) { return() }
  deiktesrio <- names(dset.rio[3:8])
  deiktesrio <- deiktesrio[ deiktesrio != input$outrio ]
  checkboxGroupInput(inputId='YZ.indicesrio',
label=h3('Variable(s):'),
                    choices=deiktesrio, selected = NULL, inline =
TRUE)
})

data.linesrio <- reactive({

  if ( !is.null(input$YZ.indicesrio) ){
    vars1rio <- list(input$outrio, input$YZ.indicesrio)
  }else{
    vars1rio <- list(input$outrio)
  }
  new.dset1rio <- subset(dset.rio, select=names(dset.rio) %in%
unlist(vars1rio) )
  new.dset1rio <- as.data.frame( cbind( "Year"=dset.rio$Year,
"Klinikes"=dset.rio$Klinikes, new.dset1rio ) )

  new.dset1rio <- new.dset1rio[ new.dset1rio$Year >=
as.numeric(input$yearrio[1]) &
                        new.dset1rio$Year <=
as.numeric(input$yearrio[2]) , ]

  if ( input$inprio != "All" ){
    new.dset1rio <- subset(new.dset1rio, new.dset1rio$Klinikes %in%
unlist(input$inprio) )
  }
  return(new.dset1rio)

})

output$intlinesrio <- renderMetricsgraphics({
  if (input$startAnalysisrio == 0){
    return() }
  else{
    isolate({

      dset.linesrio <- data.linesrio()

```

```

tidrio <- unlist(input$YZ.indicesrio)
if ( is.null(tidrio) ){
  indsrio <- list( input$outrio )
}else if ( length(tidrio) ==1 ) {
  indsrio <- list( input$outrio, tidrio[1] )
}else {
  indsrio <- list( input$outrio, tidrio[1], tidrio[2] )
}

if ( length(indsrio) == 1 )
{
  tmprio <- data.frame(yearrio= dset.linesrio$Year,
drio=as.numeric(dset.linesrio[,names(dset.linesrio) %in% indsrio[1]
] ))
  tmprio %>%
  mjs_plot(x=yearrio, y=drio) %>%
  mjs_line() %>%
  mjs_add_baseline(mean(
dset.linesrio[,names(dset.linesrio) %in% input$outrio ]), "Mean")
}
else if ( length(indsrio) == 2 ) {
  tmprio <- data.frame(yearrio= dset.linesrio$Year,
drio=as.numeric(dset.linesrio[, names(dset.linesrio) %in%
indsrio[1] ] )),
  d1rio=as.numeric(dset.linesrio[,
names(dset.linesrio) %in% indsrio[2] ])) )
  tmprio %>%
  mjs_plot(x=yearrio, y=drio) %>%
  mjs_line() %>%
  mjs_add_line(d1rio) %>%
  mjs_add_legend(legend = c("X", "Y") )
}
else{
  tmprio <- data.frame(yearrio= dset.linesrio$Year,
drio=as.numeric(dset.linesrio[, names(dset.linesrio) %in%
indsrio[1] ] )),
  d1rio=as.numeric(dset.linesrio[,
names(dset.linesrio) %in% indsrio[2] ]),
  d2rio=as.numeric(dset.linesrio[,
names(dset.linesrio) %in% indsrio[3] ])) )
  tmprio %>%
  mjs_plot(x=yearrio, y=drio) %>%
  mjs_line() %>%
  mjs_add_line(d1rio) %>%
  mjs_add_line(d2rio) %>%
  mjs_add_legend(legend = names(as.data.frame(indsrio)), )
}

})
}
})

output$pierio <- renderdimple({

```

```

if (input$startAnalysisrio == 0){
  return() }
else{
  isolate({

    dset.rio1 <- passDatario()

    dimple(dset.rio1,
           pMeasure=input$outrio,
           pCategory="Year",
           series="Klinikas",
           legend=TRUE,
           chartType="pie",
           ringInnerRadius="50%")

  })
}
})

output$barStackedrio <- renderdimple({

  if (input$startAnalysisrio == 0){
    return() }
  else{
    isolate({

      dset.rio1 <- passDatario()
      legend_value = ifelse( length(input$inprio) >=2 &
length(input$inprio) <= 8, TRUE, FALSE )

      dimple(dset.rio1,
             xCategory=c("Year"),
             yMeasure=input$outrio,
             series="Klinikas",
             legend=legend_value)

    })
  }
})

output$barStackedPercrio <- renderdimple({

  if (input$startAnalysisrio == 0){
    return() }
  else{
    isolate({

      dset.rio1 <- passDatario()
      legend_value = ifelse( length(input$inprio) >=2 &
length(input$inprio) <= 8, TRUE, FALSE )

      dimple(dset.rio1,
             xCategory=c("Year"),

```



```
        yPct=input$outrio,  
        series="Klinikes",  
        legend=legend_value)  
    })  
}  
})  
  
}) # end shinyServer
```