



**ΕΛΛΗΝΙΚΟ ΑΝΟΙΚΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ**

**ΣΧΟΛΗ**

**ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ**

**ΕΦΑΡΜΟΣΜΕΝΑ ΜΑΘΗΜΑΤΙΚΑ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ**

**ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ**

**ΟΝΟΜΑ ΦΟΙΤΗΤΡΙΑΣ**

**ΤΣΙΑΡΑ ΘΕΟΧΑΡΙΑ**

**ΟΝΟΜΑ ΕΠΙΒΛΕΠΟΝΤΑ ΚΑΘΗΓΗΤΗ**

**ΠΟΛΙΤΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ**

**ΠΑΤΡΑ**

**ΜΗΝΑΣ ΙΟΥΝΙΟΣ, ΕΤΟΣ 2022**

## ΕΥΧΑΡΙΣΤΙΕΣ

Με την περάτωση του μεταπτυχιακού κύκλου σπουδών μου θα ήθελα να ευχαριστήσω τον θείο μου Θεοχαρίδη Γιώργο , Χημικό , που πριν από 40 χρόνια μου είπε το πρώτο ψάξε στα βιβλία να βρεις τις απαντήσεις στα ερωτήματα σου και τον εκλειπόντα καθηγητή μου στα Μαθηματικά , Σιτζόγλου Αθανάσιο , από τον τρόπο διδασκαλίας του οποίου διδάχθηκα τη δυναμική της θεωρίας στην επίλυση ασκήσεων.

Επίσης , θα ήθελα να ευχαριστήσω τον Επιβλέποντα Καθηγητή μου κ. Πολίτη Κωνσταντίνο για τα πραγματικά δεδομένα του SHARE που μου έδωσε ώστε να τα επεξεργαστώ μέσω του Λογισμικού Προγράμματος της R με την μέθοδο της Λογιστικής Παλινδρόμησης και για τις απαραίτητες σχολαστικές του παρατηρήσεις που συντέλεσαν στην ολοκλήρωση της Διπλωματικής μου.

Περισσότερο από όλους θα ήθελα να ευχαριστήσω το συμφοιτητή μου Τσιάρα Βασίλη που μου έδωσε τις σωστές εντολές του πακέτου R για την υλοποίηση της Λογιστικής Παλινδρόμησης.

## ΠΕΡΙΛΗΨΗ

Η Δίτιμη Λογιστική Παλινδρόμηση είναι μία προέκταση της Γραμμικής Παλινδρόμησης , όπου η επεξηγηματική μεταβλητή παίρνει μόνο δύο τιμές, το 0 και το 1 . Η εφαρμογή της σε πραγματικά δεδομένα εκτιμά την πιθανότητα να συμβεί ή όχι ένα γεγονός που έχει δύο εκβάσεις. (Για παράδειγμα την πιθανότητα να εμφανίσει κάποιος Ευρωπαίος ή Ισραηλινός , ηλικίας άνω των 50 ετών, κατάθλιψη.)

Αυτή είναι η πρώτη μου εφαρμογή , η οποία γίνεται για να διαπιστώσουμε αρχικά πόσο καλή προσαρμογή έχουν τα πραγματικά δεδομένα στα αποτελέσματα της Λογιστικής Παλινδρόμησης και εκ των υστέρων πόσο καλή πρόβλεψη της πιθανότητας του ενδεχομένου που αναζητούμε , δίνει η Λογιστική Παλινδρόμηση.

Τα δεδομένα είναι πραγματικά ,παρμένα ,από την Ευρωπαϊκή Πηγή Δεδομένων SHARE για την οποία γίνεται εκτενής αναφορά κυρίως στα ευρήματά της που η χρησιμότητα τους είναι αναμφισβήτητη.

Η, δε, επεξεργασία των δεδομένων γίνεται με το Στατιστικό Λογισμικό της Γλώσσας Προγραμματισμού R για την οποία γίνεται επίσης μνεία.

Επιπλέον στην αρχή της διπλωματικής μου διατριβής γίνεται αναφορά στο θεωρητικό Σκέλος της Λογιστικής Παλινδρόμησης το οποίο αποδεικνύει την εγκυρότητα της .

Το Ερευνητικό μου Ερώτημα , πόσο ωφέλιμη είναι η Λογιστική Παλινδρόμηση , απαντιέται από τα αποτελέσματα των Εφαρμογών σε πραγματικά δεδομένα

## **Abstract**

The two chance's Logistic Regression is an expansion of Linear Regression where the depended variable takes only two rates 0 and 1 .Its application to real data estimates the probability of happening one of the two chances of our research.(For instance , the probability, an over 50 years old European or Israeli has to be illed of depression.)

This is my first application which becomes in order to verify , initially , how good is the adjustment of real data to the results of the Logistic Regression and then how well is the prediction of probability we search, according to results of the analysis of Logistic Regression,

The data that we use are real ,taken from the European Data Bank of SHARE. A wide reference is become about , mainly, the research findings of SHARE which are undoubtedly useful.

Moreover the analysis of data is become by the statistic Computer Program of R for which a shortly reference takes places.

Initially, in my diploma , I refer the theory that Logistic Regression is based that it proves it's valid.

My Search Question ,how useful is the Logistic Regression , is proved by the results of the applications in real data.

## Περιεχόμενα

Κεφάλαιο 1:.....σελ 1
Κεφάλαιο2.....σελ 9
Κεφάλαιο3.....σελ 12
Κεφάλαιο4.....σελ.17
Κεφάλαιο 5.....σελ 22
Κεφάλαιο 6.....σελ 31
Κεφάλαιο 7.....σελ48
Κεφάλαιο 8.....σελ61
Εντολές R.....σελ 62
Βιβλιογραφία.....σελ 64
ΠΑΡΑΡΤΗΜΑ.....σελ65

## ΕΙΣΑΓΩΓΗ

Στην παρούσα εργασία θα ασχοληθούμε με το στατιστικό πρόβλημα ανάλυσης δεδομένων χρησιμοποιώντας τη μέθοδο της Λογιστικής Παλινδρόμησης. Στην παλινδρόμηση γενικά μελετάμε τη (γραμμική) σχέση που υπάρχει ανάμεσα σε μία ή περισσότερες ανεξάρτητες, ή επεξηγηματικές, μεταβλητές, με μία εξαρτημένη μεταβλητή (μεταβλητή απόκρισης, response variable). Στη λογιστική παλινδρόμηση η εξαρτημένη μεταβλητή παίρνει δύο δυνατές τιμές (οι οποίες συμβολίζονται με τις τιμές 0 και 1). Το 0 παριστάνει την κατάσταση αποτυχίας ενός ενδεχομένου ενώ το 1 παριστάνει την κατάσταση επιτυχίας του επίμαχου ενδεχομένου.

**Στο 1ο Κεφάλαιο** εξηγείται αρχικά, τι είναι η Δίτιμη Λογιστική Παλινδρόμηση, που βρίσκει εφαρμογές, ποια είναι η αναγκαιότητα της εφαρμογής της μεθόδου αυτής συγκρίνοντάς την με την μέθοδο της Γραμμικής Παλινδρόμησης και στη συνέχεια αναφέρονται τα βασικά χαρακτηριστικά της.

**Στο 2ο Κεφάλαιο** γίνεται έλεγχος των μέτρων προσαρμογής του μοντέλου της Λογιστικής Παλινδρόμησης.

**Στο 3ο Κεφάλαιο** γίνεται αναφορά στη Λογιστική Συνάρτηση και θα περιγράψουμε τις ιδιότητες της καθώς επίσης και τη σχέση της με τη Λογιστική Παλινδρόμηση.

**Στο 4ο Κεφάλαιο** γίνεται αναφορά στη σχέση της Συνάρτησης Σίγμα με τη Λογιστική Συνάρτηση.

**Στο 5ο Κεφάλαιο** γίνεται αναφορά στο υπολογιστικό στατιστικό λογισμικό της γλώσσας προγραμματισμού της R, με το οποίο θα αναλύσουμε τα στατιστικά δεδομένα στα παραδείγματα στα οποία θα εφαρμόσουμε τη μέθοδο της Λογιστικής Παλινδρόμησης για να εκτιμήσουμε την πιθανότητα να συμβεί το ένα από τα δύο ενδεχόμενα της δύο τιμών εξαρτημένης μεταβλητής.

Επιπλέον γίνεται αναφορά στην έγκυρη Ευρωπαϊκή Πηγή δεδομένων SHARE από την οποία θα αντλήσουμε στατιστικά δεδομένα που έχουν συλλεχθεί σύμφωνα με τις βασικές αρχές δειγματοληψίας ώστε να αποτελούν τυχαίο αντιπροσωπευτικό δείγμα του πληθυσμού με όφελος την εξασφάλιση της αμεροληψίας και της εγκυρότητας, κατά συνέπεια, των εκτιμώμενων πιθανοτήτων, σε οποιοδήποτε επίπεδο σημαντικότητας.

**Στο 6ο Κεφάλαιο** εφαρμόζεται η μέθοδος της **ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ** σε 2 εφαρμογές πραγματικών δεδομένων.

1ο Εφαρμογή:

Ποια η εκτιμώμενη πιθανότητα να αποκτήσει κάποιος κατάθλιψη;

Εδώ ως επεξηγηματικές μεταβλητές θα χρησιμοποιηθούν μία σειρά από δημογραφικά και ιατρικά χαρακτηριστικά κάθε ατόμου, π.χ. φύλο, ηλικία, οικογενειακή κατάσταση, αριθμός χρόνιων παθήσεων κλπ.

2ο Εφαρμογή:

Ποια η εκτιμώμενη πιθανότητα να έχει καλή υγεία ένας Δανός, Έλληνας, Αυστριακός ή Γερμανός πολίτης;

**Στο 7ο Κεφάλαιο** αναφέρονται οι διαφορετικές προσεγγίσεις (εκτιμήσεις) των συντελεστών των διαφορετικών μοντέλων της Λογιστικής Παλινδρόμησης στα ίδια δεδομένα.

**Στο 8ο Κεφάλαιο** αναφέρονται τα συμπεράσματα αυτής της Διπλωματικής Έρευνας

**Η μέθοδος** της Λογιστικής Παλινδρόμησης είναι η Γλώσσα Προγραμματισμού R και τα **προσδοκώμενα αποτελέσματα** είναι :

1ο ) η καλή προσαρμογή των πραγματικών δεδομένων στο Λογιστικό Μοντέλο

2ο ) η καλή προσαρμογή της Λογιστικής Συνάρτησης στο μοντέλο της Λογιστικής Παλινδρόμησης

3ο ) η καλή προσαρμογή της Λογιστικής Παλινδρόμησης σε εφαρμογές πραγματικών δεδομένων που έχουν παρθεί από την Ευρωπαϊκή Πηγή Δεδομένων SHARE

4ο ) το μεγάλο πλήθος συμπερασμάτων που δύναται να δώσει η Λογιστική Παλινδρόμηση το οποίο αποδεικνύει την τεράστια συνεισφορά της στην επιστήμη , την τεχνική και την κοινωνία γενικότερα.

5ο) Οι συντελεστές που βρίσκει η Λογιστική Παλινδρόμηση είναι εκτιμήσεις (προσεγγίσεις) , που κυμαίνονται σε διαστήματα εμπιστοσύνης , με μικρή πιθανότητα να βρίσκονται έξω από αυτά.

## ΚΕΦΑΛΑΙΟ 1ο

### 1.1 Εισαγωγή στη Λογιστική Παλινδρόμηση

Η Λογιστική Παλινδρόμηση είναι μία Στατιστική Τεχνική η οποία ερμηνεύει τη σχέση που έχει μία εξαρτημένη μεταβλητή που παίρνει δύο τιμές μόνο (π. χ. υγιής/ασθενής, επιτυχία/αποτυχία, κανονικό/ελαττωματικό, ζωή/θάνατος) με μία ή περισσότερες ανεξάρτητες μεταβλητές οι οποίες μπορεί να είναι διακριτές ή συνεχείς. Χρησιμοποιεί τη Λογιστική Συνάρτηση στην οποία θα αναφερθούμε διεξοδικά στο 3ο Κεφάλαιο για να μοντελοποιήσει τις τιμές της δίτιμης εξαρτημένης μεταβλητής.

Μία αξιοσημείωτη διαφορά που έχει από τη Γραμμική Παλινδρόμηση είναι ότι στη δεύτερη η μεταβλητή απόκρισης είναι συνεχής. Επιπλέον στο λογιστικό μοντέλο τα σφάλματα δεν ακολουθούν την κανονική κατανομή και οι υποθέσεις της Γραμμικής Παλινδρόμησης όπως η ομοσκεδαστικότητα και κανονικότητα παραβιάζονται. Τα φαινόμενα αυτά εμφανίζονται ιδιαίτερα όταν τα δείγματα είναι μικρά ( $n < 200$ ), (ελάχιστο μέγεθος δείγματος 100 ή 200 ανάλογα με το πλήθος των εξηγηματικών μεταβλητών Kerling & Pedhazur 1973) ή όταν τα δεδομένα είναι σπάνια.

Βρίσκει εφαρμογές σε όλες τις επιστήμες όπως στην βιοϊατρική (π.χ. εκτιμώντας την πιθανότητα ένα αντιβιοτικό να είναι δραστικό ή όχι), στη μηχανική (π.χ. όταν ερευνούμε αν μία μηχανή αυτοκινήτου έχει μεγάλη διάρκεια ζωής ή όχι). Βέβαια σε αυτό το σημείο υπεισέρχεται μία νέα μαθηματική έννοια, αυτή της Ασάφειας, το μετά από πόσα χρόνια ζωής, δηλαδή, μία μηχανή θεωρείται ότι έχει μεγάλη διάρκεια ζωής), στην κοινωνιολογία (π.χ. εκτιμώντας την πιθανότητα εμφύλιου πολέμου) και σε πολλές άλλες.

Βέβαια, εδώ δράττουμε την ευκαιρία να τονίσουμε ότι τα σπάνια δίτιμα ενδεχόμενα, όπως η πιθανότητα ενός μεγάλου σεισμού ή εμφάνισης μιας σπάνιας ασθένειας, μελετώνται με περισσότερη ακρίβεια και λιγότερα σφάλματα με την τεχνική της Λογιστικής Παλινδρόμησης και όχι με αυτήν της Γραμμικής Παλινδρόμησης.

#### 1.2.1 Ορισμός Λογιστικού Μοντέλου

Αν  $Y$  είναι ανεξάρτητη τυχαία μεταβλητή Bernoulli το λογιστικό μοντέλο ορίζεται ως εξής:

$$Y_i = E(Y_i) + \varepsilon_i \quad \text{με} \quad E(Y_i) = \frac{e^{\beta_0 + \sum_i x_i \cdot \beta_i}}{1 + e^{\beta_0 + \sum_i x_i \cdot \beta_i}}$$

για κάθε  $i$  άτομο του δείγματος.



2

$$\text{και } \text{Var}(Y_i) = E((Y_i - E(Y_i))^2)$$

όπου  $X_i$  οι ανεξάρτητες τυχαίες μεταβλητές,  $\beta_0$  ο σταθερός όρος και  $\beta_i$  είναι οι συντελεστές των ανεξάρτητων μεταβλητών  $X_i$  οι οποίοι είναι άγνωστοι.

Αυτό αποδεικνύεται ως εξής:

$$E(Y_i) = 1 * \pi_i + 0 * (1 - \pi_i) = \pi_i$$

για τις πιθανότητες να συμβεί και να μην συμβεί ένα γεγονός αντίστοιχα. Οπότε η αναμενόμενη τιμή της  $Y_i$  είναι

$$E(Y_i) = 1 * \pi_i + 0 * (1 - \pi) = \pi_i$$

που σημαίνει ότι η αναμενόμενη τιμή της  $Y_i$  είναι η πιθανότητα να συμβεί το γεγονός. Επιθυμώντας, τώρα, να συνδέσουμε την πιθανότητα  $\pi_i$  με την γραμμική έκφραση

$$\beta_0 + \sum_{i=1}^k x_i * \beta_i$$

θα χρησιμοποιήσουμε μία συνάρτηση  $g(\pi)$  γνησίως αύξουσα με πεδίο ορισμού το διάστημα  $[0,1]$  και σύνολο τιμών  $(-\infty, +\infty)$ . Υπάρχουν αρκετές τέτοιες συναρτήσεις, όμως οι τρεις παρακάτω είναι οι πιο βασικές:

### Λογιστική Συνάρτηση

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \quad (\text{Reed \& Pedhazur 1973})$$

### Probit η αντίστροφη Κανονική Συνάρτηση

$$g(\pi) = \Phi^{-1}(\pi) \quad (\text{Gauss 1809})$$

### Συμπληρωματική log-log Συνάρτηση

$$g(\pi) = \log(-\log(1-\pi)) \quad (\text{Chester Ittner Bliss 1934})$$

Τις περισσότερες φορές βέβαια χρησιμοποιείται η Λογιστική Συνάρτηση γιατί είναι εύκολο να ερμηνευτεί καθώς είναι ίση με τον λογάριθμο της σχετικής πιθανότητας (odds) να συμβεί το γεγονός.

Επομένως στο Λογιστικό Μοντέλο έχουμε ότι:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{i=1}^k \beta_i * X_i \Leftrightarrow \frac{\pi_i}{1-\pi_i} = e^{\beta_0 + \sum_{i=1}^k \beta_i * X_i}$$

όπου  $\pi_i$  είναι η πιθανότητα ενός ενδεχομένου  $i$  που αποτελείται από ένα διάνυσμα  $k$  τιμών οι οποίες αντιστοιχούν στις  $X_k$  ανεξάρτητες μεταβλητές.

$$\Leftrightarrow \pi_i = \frac{\exp(\beta_0 + \sum_{i=1}^k \beta_i * X_i)}{1 + \exp(\beta_0 + \sum_{i=1}^k \beta_i * X_i)}$$

Σε αυτό το σημείο θα γίνει αναφορά για την έννοια της ομοσκεδαστικότητας η οποία δεν εμφανίζεται απαραίτητα ως προϋπόθεση για την εφαρμογή του μοντέλου της Λογιστικής Παλινδρόμησης ενώ είναι απαραίτητη προϋπόθεση για την εφαρμογή του μοντέλου της Γραμμικής Παλινδρόμησης.

Αυτό σημαίνει ότι στη Λογιστική Παλινδρόμηση τα σφάλματα  $\varepsilon_i$  δεν έχουν απαραίτητα ίσες διασπορές.

Μία απλή απόδειξη αυτού του ισχυρισμού είναι η εξής:

$$\text{Var}(\varepsilon_i) = \text{Var}(Y_i - \pi_i) = \text{Var}(Y_i) - \text{Var}(-\pi_i) = \text{Var}(Y_i) + 0 = \text{Var}(Y_i)$$

Μία δεύτερη αναλυτικότερη απόδειξη του παραπάνω ισχυρισμού είναι :

$$\begin{aligned} \text{Var}(Y_i) &= E((Y_i - EY_i)^2) = E(Y_i - \pi_i)^2 = E(Y_i - \pi_i - \pi_i + \pi_i)^2 = E((Y_i - \pi_i) - EY_i + \pi_i)^2 \\ &= E((Y_i - EY_i) - E[Y_i - EY_i])^2 = \text{Var}(Y_i - EY_i) = \text{Var}(\varepsilon_i) \end{aligned}$$

(Τσιάρα Θ.)

Άρα η διασπορά του σφάλματος είναι διαφορετική για κάθε  $i$

### 1.2.2 Εκτίμηση Παραμέτρων με τη μέθοδο της Μέγιστης Πιθανοφάνειας

Είναι φανερό πως στο Λογιστικό Μοντέλο δεν μπορεί να εφαρμοστεί η μέθοδος των Ελαχίστων Τετραγώνων για να εκτιμηθούν οι συντελεστές των ανεξάρτητων μεταβλητών  $X_i$  για τον λόγο ότι η εξαρτημένη μεταβλητή  $Y$  δεν ακολουθεί την Κανονική Κατανομή, οπότε δεν είναι συνεχής αφού παίρνει δύο μόνο τιμές, τις 0 και 1. Κατά συνέπεια, τα αθροίσματα τετραγώνων που αφορούν την μεταβλητή  $Y$  δεν ακολουθούν κάποια γνωστή κατανομή, όπως στην συνήθη γραμμική παλινδρόμηση. Επιπλέον λόγοι είναι ότι οι υποθέσεις ομοσκεδαστικότητας, κανονικότητας και γραμμικότητας παραβιάζονται.

Αναζητώντας, επομένως, κάποια άλλη μέθοδο για να υπολογιστούν οι παράμετροι (συντελεστές των Επεξηγηματικών Μεταβλητών  $X_i$ ) οι επιστήμονες ΜΑΘΗΜΑΤΙΚΟΙ-ΣΤΑΤΙΣΤΙΚΟΙ

κατέληξαν στην εφαρμογή της Μέγιστης Πιθανοφάνειας που είναι ανεξάρτητη από τις παραπάνω παραβιάσεις.

Η μέθοδος Μέγιστης Πιθανοφάνειας οφείλεται στον Fisher και είναι μία τεχνική κατασκευής εκτιμητών παραμέτρων  $\vec{b}$ . Στην Λογιστική Παλινδρόμηση η παράμετρος  $\vec{b}$  είναι το διάνυσμα  $\vec{b}=(\beta_0, \beta_1, \dots, \beta_k)$  με συνιστώσες τους συντελεστές των ανεξάρτητων μεταβλητών  $X_i$ .

Η υπολογιστική διαδικασία της παραμέτρου  $\vec{b}$  με τη μέθοδο της Μέγιστης Πιθανοφάνειας πραγματοποιείται σε δύο βήματα. Στο πρώτο φτιάχνουμε τη συνάρτηση Πιθανοφάνειας αποδίδοντας στην εξαρτημένη μεταβλητή  $Y$  τη διωνυμική με μία δοκιμή (Bernoulli) και παράμετρο  $\pi_i$  η οποία θεωρείται ότι εξαρτάται από τις επεξηγηματικές μεταβλητές  $X_i$ . Οπότε την  $Y_i$  την εξισώνουμε με μία παράσταση που περιέχει τις ανεξάρτητες ή επεξηγηματικές μεταβλητές. Έτσι δημιουργείται η συνάρτηση Πιθανοφάνειας. Προφανώς δεχόμαστε ότι οι παρατηρήσεις είναι ανεξάρτητες μεταξύ των ατόμων.

Στο δεύτερο βήμα μεγιστοποιούμε τη συνάρτηση Πιθανοφάνειας ως προς τις άγνωστες παραμέτρους. Αυτό πραγματοποιείται εφαρμόζοντας μία επαναληπτική μέθοδο προσέγγισης της λύσης η οποία συντελείται με τη βοήθεια των ηλεκτρονικών υπολογιστών τα τελευταία χρόνια γρήγορα.

Αναλυτικότερα:

Για κάθε παρατήρηση  $i$  έχουμε το  $Y_i$  και το  $X_i$ , όπου  $Y_i$  είναι μία τυχαία μεταβλητή με δυνατές τιμές 0 ή 1 και  $\vec{X}_i=[1, X_{i,1}, X_{i,2}, X_{i,3}, \dots, X_{i,k}]^T$  είναι ένας πίνακας στήλη που περιέχει τις επεξηγηματικές μεταβλητές με τη μονάδα που αντιπροσωπεύει το σταθερό όρο. Για λόγους ευκολίας θα τον γράφουμε  $X_i$ . Υποθέτουμε ότι η πιθανότητα  $\pi_i$  είναι η πιθανότητα το  $Y_i$  να είναι 1.

$$Pr(Y_i=1)=\pi_i.$$

Αν τα δεδομένα προέρχονται από το Λογιστικό μοντέλο έχουμε αποδείξει ότι

$$\pi_i = \frac{1}{1 + e^{-\beta_0 - \sum_{i=1}^k \beta_i * X_i}} \quad (1)$$

ΤΟΤΕ η συνάρτηση Πιθανοφάνειας, λόγω του γεγονότος της ανεξαρτησίας των  $Y_i$ , ορίζεται:

$$L = Pr(Y_1, Y_2, \dots, Y_k) = Pr(Y_1) * Pr(Y_2) * \dots * Pr(Y_n) = \prod_{i=1}^n Pr(Y_i)$$

και εφόσον αποδείχθηκε παραπάνω ότι  $Pr(Y_i) = \pi_i^{Y_i} * (1 - \pi_i)^{1 - Y_i}$  θα ισχύσει τελικά

$$L = \prod_{i=1}^n \left( \frac{\pi_i}{1 - \pi_i} \right)^{Y_i} * (1 - \pi_i)$$

Αν, στη συνέχεια, λογαριθμίσουμε και τα δύο μέρη της παραπάνω ισότητας, διότι είναι πιο εύκολο να μεγιστοποιήσουμε τον λογάριθμο  $\log(L)$  αντί για την ίδια την  $L$ . Αυτό είναι ενδιαφέρον γιατί ο λογάριθμος είναι γνησίως αύξουσα συνάρτηση και οι δύο συναρτήσεις ( $\log L$ ,  $L$ ) μεγιστοποιούνται στο ίδιο σημείο ακριβώς. Έτσι καταλήγουμε στην ακόλουθη Συνάρτηση

$$\log(L) = \sum_{i=1}^n Y_i * \log\left(\frac{\pi_i}{1-\pi_i}\right) + \sum_{i=1}^n \log(1-\pi_i) \quad (2)$$

τέλος, αντικαθιστώντας, στην παραπάνω σχέση την σχέση (1) καταλήγουμε στη συνάρτηση Πιθανοφάνειας:

$$\log L = \sum_{i=1}^n Y_i * [\beta_0 + \sum_{i=1}^n \beta_i * X_i] - k * \log [1 + e^{\beta_0 + \sum_{i=1}^n \beta_i * X_i}] \quad (3)$$

Έχοντας βρει, λοιπόν, την συνάρτηση Πιθανοφάνειας θα προχωρήσουμε στο επόμενο βήμα που είναι να βρούμε τα  $\vec{\beta} = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]$  που μεγιστοποιούν την (3). Βρίσκουμε, επομένως, την παράγωγο της (3) ως προς  $\beta$  και τη θέτουμε ίση με 0. Οπότε καταλήγουμε, μετά την εφαρμογή αυτής της διαδικασίας στο παρακάτω αποτέλεσμα:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^n X_i * Y_i - \sum_{i=1}^n X_i * (1 + e^{-\beta_0 - \sum_{i=1}^n \beta_i * X_i})^{-1}$$

$$\text{και} \quad \frac{\partial \log L}{\partial \beta} = 0 \Leftrightarrow \sum_{i=1}^n X_i * Y_i - \sum_{i=1}^n X_i * \hat{Y}_i = 0$$

$$\text{και έτσι ισοδύναμα προκύπτει ότι} \quad \hat{Y}_i = \frac{1}{1 + e^{-\beta_0 - \sum_{i=1}^n \beta_i * X_i}} = \frac{e^{\beta_0 + \sum_{i=1}^n X_i * \beta_i}}{1 + e^{\beta_0 + \sum_{i=1}^n X_i * \beta_i}}$$

Από την (1) θεωρείται ότι η  $\Pi_i$  είναι η εκτιμώμενη τιμή του  $Y$  για δοσμένο  $X_i$ . Επειδή το  $\vec{\beta}$  είναι πίνακας στήλη, η παραπάνω σχέση στην πραγματικότητα είναι ένα σύστημα  $k+1$  εξισώσεων, ένα για κάθε  $\beta$ .

Εκτός από σπάνιες περιπτώσεις που η λύση είναι εύκολο να βρεθεί χρησιμοποιούνται επαναληπτικές μέθοδοι που έχουν διαδοχικές προσεγγίσεις μέχρι η λύση να συγκλίνει στην πραγματική τιμή.

Υπάρχουν και εδώ αρκετές μέθοδοι υπολογισμού που δίνουν την ίδια λύση, διαφέρουν όμως στην ταχύτητα σύγκλισης η οποία επηρεάζεται από τις αρχικές τιμές.

Οι πιο διαδεδομένες μέθοδοι είναι η Newton-Raphson και η Fisher Scoring.

Η Newton-Raphson μέθοδος χρησιμοποιεί τον ακόλουθο τύπο

$$\beta_{j+1} = \beta_j - H^{-1}(\beta_j) * U(\beta_j)$$

όπου η  $U(\beta)$  είναι ένα διάνυσμα με τις πρώτες παραγώγους της

$\log L$  ως προς  $\beta$ , και πιο συγκεκριμένα

$$U(\beta) = \frac{\partial \log L}{\partial \beta} = \sum_{i=1}^n X_i * Y_i - \sum_{i=1}^n X_i * \hat{Y}_i$$

Επίσης ο  $H(\beta)$  είναι ένας πίνακας με τις δεύτερες παραγώγους της  $\log L$  ως προς  $\beta$ , και ονομάζεται Εσσιανός πίνακας έχοντας τον ακόλουθο τύπο:

$$H(\beta) = \frac{\partial^2 \log L}{\partial \beta * \partial \beta} = \sum_{i=1}^n X_i * X_i^T * \hat{Y}_i * (1 - \hat{Y}_i)^T$$

Η μέθοδος Fisher Scoring που πρότεινε ο Fisher ήρθε για να ξεπεράσει το πρόβλημα της σύγκλισης που είχε κάποιες φορές η μέθοδος Newton-Raphson. Ουσιαστικά αντικατέστησε τον Εσσιανό πίνακα- $H(\beta)$  με τον πίνακα πληροφορίας  $I(\beta)$ . Ο πίνακας αυτός ισούται με την αναμενόμενη αρνητική τιμή των δεύτερων παραγώγων του  $\log L$ , δηλαδή ισχύει ότι

$$I(\beta) = -E(\nabla^2 \log L)$$

Έτσι, λοιπόν, ο τύπος της μεθόδου Fisher Scoring είναι ο ακόλουθος:

$$\beta_{j+1} = \beta_j - I^{-1}(\beta_j) * U(\beta_j) \quad (1925 \text{ Statistical method of Research Workers, βιβλίο του}$$

Fisher)

Σχετικά με αυτές τις δύο μεθόδους θα λέγαμε ότι:

- Η μέθοδος Newton-Raphson συγκλίνει πιο γρήγορα στην λύση, όμως υπάρχουν και περιπτώσεις που αποτυγχάνει να βρεί λύση. Τέτοιες περιπτώσεις συναντάμε όταν η συνάρτηση δεν είναι ομαλή ή όταν περιέχει περισσότερα από ένα τοπικά ακρότατα.
- Η μέθοδος Fisher Scoring αντίθετα είναι πιο αξιόπιστη και συγκλίνει σε περιπτώσεις που η Newton-Raphson δεν μπορεί, κατά συνέπεια προτείνετε στις περισσότερες των περιπτώσεων.
- Επιπλέον η Fisher Scoring μέθοδος δεν εξαρτάται από τα δεδομένα μας και έτσι υπολογιστικά είναι πιο εύκολη.

### 1.2.3 Έλεγχος Σημαντικότητας των Συντελεστών

Αφού εκτιμήσαμε τους συντελεστές με την μέθοδο μέγιστης Πιθανοφάνειας στην συνέχεια ελέγχουμε αν αυτοί οι συντελεστές είναι σημαντικοί. Υπάρχει ο έλεγχος λόγου πιθανοφανειών ( Agresti 2007) που χρησιμοποιεί την στατιστική συνάρτηση:

$$G^2 = \text{Deviance}(\text{without variables}) - \text{Deviance}(\text{with variables})$$

και συγκρίνει το μοντέλο με τις μεταβλητές και αυτό με το σταθερό όρο μόνο. Βέβαια μπορούμε να συγκρίνουμε και ένα μοντέλο που περιέχει κάποιες μεταβλητές με ένα μοντέλο με περισσότερες για να καταλήξουμε στο αν οι επιπρόσθετες είναι σημαντικές και πρέπει να συμπεριληφθούν στο μοντέλο. Η απόκλιση (**Deviance**) που πρώτοι ανέφεραν οι Nelder και Wedderburn (1972) είναι η ακόλουθη ποσότητα

$$D = -2 \log \left[ \frac{\text{Πιθανοφάνεια του Εξεταζόμενου Μοντέλου}}{\text{Πιθανοφάνεια του Κορεσμένου Μοντέλου}} \right]$$

Όπου κορεσμένο (saturated) μοντέλο είναι αυτό που περιέχει τόσες παραμέτρους όσα είναι τα δεδομένα μας. Στην Λογιστική Παλινδρόμηση όπου η ανεξάρτητη μεταβλητή είναι δίτιμη {1 ή 0} η πιθανοφάνεια του κορεσμένου μοντέλου ισούται με 1. Πράγματι αυτό συμβαίνει γιατί στην επίλυση του συστήματος (3) ο αριθμός των εξισώσεων είναι όσες και ο αριθμός των αγνώστων και έτσι έχουμε την τέλεια προσαρμογή.

Συνεπώς η απόκλιση στο Λογιστικό μοντέλο είναι:

$$D = -2 \log(\text{likelihood fitted model})$$

Συνεπώς η στατιστική Συνάρτηση  $G^2$  στην λογιστική παλινδρόμηση παίρνει την ακόλουθη μορφή :

$$G^2 = 2(\log L_{\text{with variables}} - \log L_{\text{without variables}})$$

Τέλος για να αποφανθούμε για την σημαντικότητα των μεταβλητών συγκρίνουμε τη στατιστική συνάρτηση  $G^2$  με το κατάλληλο ποσοστιαίο σημείο της

$$X^2_{DF} = [df(\text{without variables}) - df(\text{with variables})]$$

και αν  $G^2 > X^2_{DF}$

τότε λέμε ότι τουλάχιστον μία μεταβλητή είναι στατιστικά σημαντική.

Κάνοντας έλεγχο ενός μοντέλου με το μοντέλο με τον σταθερό όρο η στατιστική συνάρτηση  $G^2$  παίζει τον ίδιο ρόλο που παίζει και η  $F$  στην Γραμμική παλινδρόμηση. Γενικότερα, θα λέγαμε ότι η απόκλιση γενικεύει την έννοια του αθροίσματος τετραγώνων που οφείλονται στο σφάλμα (SSE) στη συνήθη γραμμική παλινδρόμηση. Διαισθητικά, μπορούμε να πούμε ότι η απόκλιση αντιπροσωπεύει το τμήμα της μεταβλητότητας της  $Y$  που δεν ερμηνεύεται από το υπόδειγμά μας.

Υπάρχει βέβαια και ο έλεγχος Wald που χρησιμοποιεί την στατιστική συνάρτηση

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

Από τις ασυμπτωτικές ιδιότητες των ΕΜΠ έχουμε ότι:

- $\hat{\beta}_i \sim N(\bar{\beta}_i, \sigma^2(\beta_i))$
- Ο  $\hat{\beta}_i$  συνεπής εκτιμητής του  $\beta_i$

Από το Θεώρημα του Slutsky (1925), έχουμε ότι η στατιστική συνάρτηση  $W$  είναι ασυμπτωτικά κανονική. Επομένως το αντίστοιχο  $p$ -value υπολογίζεται ως εξής:

$$P(|Z| > W) .$$

Ο έλεγχος του Wald μας απαντάει στο ερώτημα εάν η παράμετρος  $\beta_i$  είναι στατιστικά σημαντική, δηλαδή αν διαφέρει σημαντικά από την τιμή μηδέν.

Βέβαια όταν η μεταβλητή είναι κατηγορική δεν μας δίνει αξιόπιστα αποτελέσματα για αυτό θα πρέπει να εκτελούμε έλεγχο λόγου πιθανοφανειών ανάμεσα στο μοντέλο με την συγκεκριμένη μεταβλητή και σε αυτό χωρίς αυτήν.

## ΚΕΦΑΛΑΙΟ 2

### 2.1 ΜΕΤΡΑ ΠΡΟΣΑΡΜΟΓΗΣ ΜΟΝΤΕΛΟΥ

Στη συνήθη γραμμική παλινδρόμηση ως δείκτης ποιότητας του μοντέλου έχουμε το ευρέως γνωστό  $R^2$ , όπου παίρνει τιμές στο διάστημα 0 έως 1

Τιμές του  $R^2$  κοντά στην μονάδα, μας φανερώνουν καλή προσαρμογή του μοντέλου στα δεδομένα ενώ αντίθετα μικρές τιμές του  $R^2$  κοντά στο 0 μας δείχνουν κακή προσαρμογή. Το  $R^2$  ερμηνεύεται ως το ποσοστό της συνολικής διακύμανσης που εξηγείται από το μοντέλο.

Για το Λογιστικό Μοντέλο υπάρχουν πολλές παραλλαγές του  $R^2$ , από τις οποίες η καλύτερη θεωρείται αυτή των Cox and Snell και αν οι τιμές του κυμαίνονται μεταξύ των 0,2 και 0,4 δείχνουν καλή προσαρμογή.

Σύμφωνα με αυτούς είναι 
$$R^2 = 1 - \left( \frac{\log L_{null}}{\log L_M} \right)^{2/n}$$

όπου  $n$  είναι το πλήθος των παρατηρήσεων.

$\log L_M$  είναι ο λογάριθμος της μέγιστης Πιθανοφάνειας του μοντέλου  $M$  και  $\log L_{null}$  είναι ο λογάριθμος της μέγιστης Πιθανοφάνειας του μοντέλου που δεν χρησιμοποιεί επεξηγηματικές μεταβλητές.

Η παραπάνω εξίσωση μπορεί να χρησιμοποιηθεί για κάθε μοντέλο παλινδρόμησης που εκτιμάται με την μέθοδο της μέγιστης Πιθανοφάνειας.

### 2.2 ΠΡΟΒΛΕΠΤΙΚΗ ΙΣΧΥΣ ΕΝΟΣ ΜΟΝΤΕΛΟΥ

#### ΚΑΜΠΥΛΕΣ ROC

Σε αυτό το σημείο αξίζει να αναφέρουμε τις καμπύλες ROC (David W. Hosmer, Jr. Stanley Lemeshow 2001) που μας περιγράφουν το πόσο καλά ταξινομεί το μοντέλο μας την εξαρτημένη μεταβλητή  $Y$  (0 ή 1) για δοθέντες τιμές ανεξάρτητων μεταβλητών. Πριν αναλύσουμε τις καμπύλες αυτές είναι χρήσιμο να αναφέρουμε δύο μέτρα όπου βοηθούν στην κατανόηση και στην δημιουργία τους. Αυτά είναι:

1. *Ευαισθησία* =  $P(\hat{Y}=1/Y=1)$

2. *Ειδικότητα* =  $P(\hat{Y}=0/Y=0)$

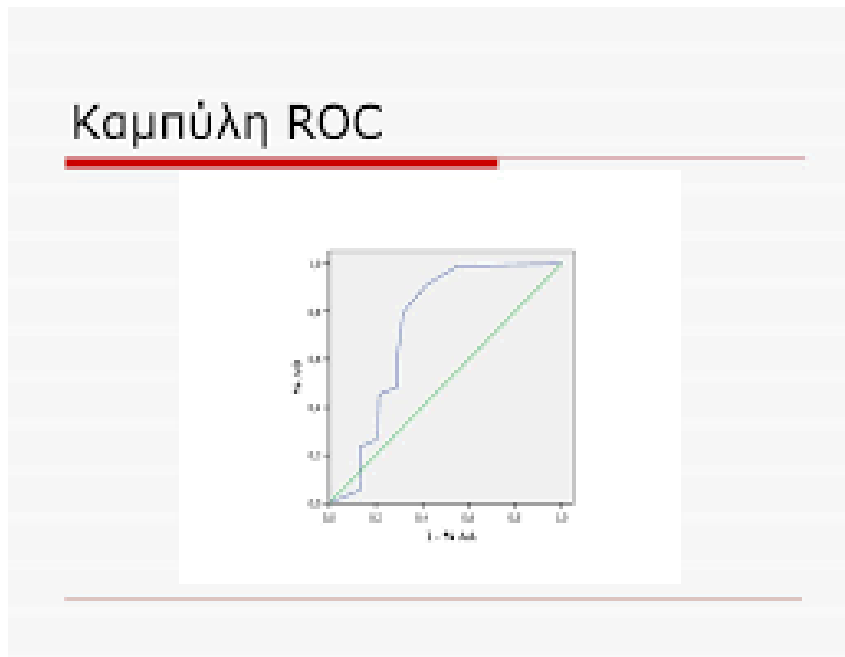


Οι ποσότητες αυτές όμως εξαρτώνται από την τιμή μιας διαδρομής. Η δημοφιλέστερη μέθοδος για την επιλογή της διαδρομής είναι η μέθοδος Youden (1950) που χρησιμοποιεί τον παρακάτω δείκτη:

$$J = \text{Ευαισθησία} - (1 - \text{Ειδικότητα})$$

Αφού λοιπόν έχουμε υπολογίσει τις τιμές αυτές (ευαισθησία, ειδικότητα) για τις διάφορες διαδρομές επιλέγουμε αυτό που μεγιστοποιεί την ποσότητα J.

Μία τέτοια καμπύλη ROC έχει την παρακάτω μορφή:



Η περιοχή κάτω από την καμπύλη που ονομάζεται AUC (Area Under the Curve) παίρνει τιμές από 0 έως 1 και μας δείχνει την ικανότητα του μοντέλου να διαχωρίζει τα δεδομένα και να δίνει τιμή  $Y=1$  ή  $Y=0$  για την εξαρτημένη μεταβλητή. Ένας κανόνας αξιολόγησης της ταξινόμησης είναι ο ακόλουθος:

- Αν  $AUC = 0.5$  έχουμε κακή ταξινόμηση (ταξινομεί με πιθανότητα 50%)
- Αν  $0.5 < AUC < 0.7$ , έχουμε αρκετά φτωχή ταξινόμησης
- Αν  $0.7 \leq AUC < 0.8$ , έχουμε μια αποδεκτή ταξινόμηση
- Αν  $0.8 \leq AUC < 0.9$ , έχουμε πολύ καλή ταξινόμηση
- Αν  $AUC \geq 0.9$ , έχουμε τέλεια ταξινόμηση

(Αλεξάνδρα Πουλοπούλου 2017)

Η διαγώνιος είναι η καμπύλη για το μοντέλο που ταξινομεί με πιθανότητα 50% , ανεξάρτητα από τις τιμές των ανεξάρτητων μεταβλητών. Επίσης στην πράξη είναι εξαιρετικά ασυνήθιστο να πετύχουμε  $AUC \geq 0.9$

Εδώ μπορούμε να αναφέρουμε ότι ένα μοντέλο μπορεί να μην έχει καλή προσαρμογή αλλά να έχει κακή ταξινόμηση για αυτό προτείνεται να ελέγχουμε και τα δύο χαρακτηριστικά.

Προφανώς όσο πιο ψηλά είναι η καμπύλη τόσο καλύτερη προβλεπτική ικανότητα έχει το μοντέλο καθώς και η περιοχή κάτω από την καμπύλη (Area Under the Curve) είναι μεγαλύτερη.

Σύμφωνα με τον Bob Horton διευθυντή δεδομένων της Microsoft η καμπύλη ROC είναι ένας δημοφιλής τρόπος στο να παρουσιάσουμε τις αντιστοιχίες μεταξύ της Ευαισθησίας και της Ειδικότητας σε ένα διωνυμικό ταξινομητή. Ο ταξινομητής χρησιμοποιείται για να διαχωρίσει περιπτώσεις ώστε ,από το περισσότερο στο λιγότερο πιθανό να είναι θετικές και η Logo χελώνα βαδίζει κατά μήκος αυτής της αλυσίδας περιπτώσεων. Η χελώνα θεωρεί όλες τις περιπτώσεις που έχει περάσει ως θετικές. Αυτές θεωρούνται ότι είναι θετικές ψευδώς (FP) ή θετικές αληθώς (TP). Αυτό είναι ισοδύναμο με το να προσδιοριστεί το σκορ του βήματος. Όταν η χελώνα περνά ένα TP βαδίζει ένα βήμα επάνω παράλληλη με τον άξονα  $y'g$  και όταν περνά ένα FP βαδίζει ένα βήμα προς τα δεξιά , παράλληλη με τον άξονα  $x'x$  . Το μονοπάτι της χελώνας πάντα ξεκινά από το σημείο (0,0) και τελειώνει στο σημείο (1,1). Το αποτέλεσμα είναι ένα σχεδιάγραμμα του ρυθμού των θετικών σωστά (sensitivity) ως προς των θετικών λανθασμένα (1-specificity) και ονομάζεται καμπύλη ROC.

Η ερμηνεία της καμπύλης ROC από την πλευρά των πιθανοτήτων είναι η εξής: Εάν κάποιος τυχαία διαλέξει μία θετική και μία αρνητική υπόθεση, η πιθανότητα η θετική υπόθεση να ξεπερνά την αρνητική σύμφωνα με τον ταξινομητή δίνεται από το εμβαδό της περιοχής που βρίσκεται κάτω από την καμπύλη ROC και πάνω από τη διαγώνιο του τετραγώνου πλευράς 1 με κορυφές της μιας διαγωνίου του τις (0,0) και (1,1).

#### **ΕΡΕΥΝΗΤΙΚΑ ΕΡΩΤΗΜΑΤΑ:**

- 1ο ) Γιατί οι εφαρμογές με 2 εκβάσεις και πολλές ανεξάρτητες μεταβλητές (συνεχείς ή διακριτές ή κατηγορικές) θα πρέπει να έχουν καλή προσαρμογή στο θεωρητικό κομμάτι με τη μέθοδο της Λογιστικής Παλινδρόμησης;
- 2ο ) Πόσο καλή είναι η προσαρμογή της Λογιστικής Παλινδρόμησης στις παραπάνω εφαρμογές;
- 3ο ) Πόσο καλή είναι η προβλεψιμότητα της Λογιστικής Παλινδρόμησης;
- 4ο ) Πόσα συμπεράσματα μπορεί κάποιος να εξαγάγει από την εύρεση της Λογιστικής Συνάρτησης;
- 5ο ) Είναι οι συντελεστές  $b_i$  με  $i=0,1,\dots,k$  που εξάγονται από τη διαδικασία της Λογιστικής Παλινδρόμησης σταθεροί αριθμοί ή όχι;

## ΚΕΦΑΛΑΙΟ 3

### 3.1 ΙΣΤΟΡΙΑ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΣΥΝΑΡΤΗΣΗΣ

Η Λογιστική Συνάρτηση εισείχθηκε στα Μαθηματικά με μία σειρά από 3 συγγράμματα από τον Pierre Francois Verhulst μεταξύ 1838 και 1847 ο οποίος την πρότεινε σαν ένα μοντέλο ανάπτυξης πληθυσμού από την προσαρμογή του μοντέλου της εκθετικής ανάπτυξης υπό τις οδηγίες του Adolphe Quetelet.

Ο Verhulst πρώτος, ανακάλυψε τη συνάρτηση στα μέσα του 1830 , δημοσίευσε μία σύντομη σημείωση το 1838 , μετά παρουσίασε μία εκτεταμένη ανάλυση και ονόμασε τη συνάρτηση το 1844 (εκδόθηκε το 1845). Στο τρίτο σύγγραμμα προσαρμόσε την ορθότητα των όρων στο Μοντέλο της ανάπτυξης του Βελγικού πληθυσμού.

Το αρχικό τμήμα της ανάπτυξης του πληθυσμού είναι προσεγγιστικά εκθετικό . Μετά ,όταν ο κορεσμός έχει επιτευχθεί , η ανάπτυξη μετατρέπεται σε γραμμική και στο τέλος η ανάπτυξη σταματά ( επέρχεται δηλαδή η ωρίμανση της ανάπτυξης του πληθυσμού.)

Ο Verhulst δεν εξήγησε την επιλογή του όρου Λογιστικής (Γαλλική *logistique*) όμως είναι πιθανόν ως αντίθεση της λογαριθμικής συνάρτησης. Το αναπτυξιακό του έργο παρουσιάζεται ως μία σύζευξη αριθμητικής και γεωμετρικής ανάπτυξης ( την καμπύλη των οποίων , ονόμασε λογαριθμική σε αντίθεση του μοντέρνου όρου εκθετική).Οπότε η “Λογιστική Ανάπτυξη” είναι πολύ πιθανόν να προέρχεται από την αναλογία και Λογιστική που πηγάζει από την Αρχαία Ελλάδα.

### 3.2 ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΣΥΝΑΡΤΗΣΗΣ

Ο τύπος της Λογιστικής Συνάρτησης είναι :

$$f(x) = \frac{L}{1 + e^{-K*(x-x_0)^2}} \quad \text{είναι παράμετροι.}$$

Η ακριβής Λογιστική Συνάρτηση είναι η Συνάρτηση Σίγμα με παραμέτρους  $K=1$  ,  $x_0 = 0$  ,  $L=1$  , η οποία παράγει την

$$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x} = \frac{1}{2} + \frac{1}{2} * \tanh\left(\frac{x}{2}\right)$$

Πρακτικά εξαιτίας της φύσης του εκθετικού όρου  $e^{-x}$  , είναι συχνά επαρκές να υπολογίζουμε την τυπική Λογιστική Συνάρτηση του  $x$  ένα περιορισμένο εύρος πραγματικών τιμών , όπως  $[-6,6]$  , καθώς γρήγορα συγκλίνει στις φυσικές της τιμές 0 και 1

Η Λογιστική Συνάρτηση έχει την Συμμετρική ιδιότητα  $f(-x)+f(x)=1$

η οποία δείχνει ότι έχει κέντρο συμμετρίας το σημείο  $(0, \frac{1}{2})$ . Επιπλέον η παραπάνω ισότητα δείχνει ότι η συνάρτηση  $f(x) - \frac{1}{2}$  είναι περιττή.

Η Λογιστική Συνάρτηση είναι απόρροια της τμηματικής Συνάρτησης υπερβολικής εφαπτομένης

$$f(x) = \frac{1}{2} + \frac{1}{2} * \tanh\left(\frac{x}{2}\right) \Leftrightarrow \tanh(x) = 2 * f(2x) - 1$$

Αυτό εξάγεται από

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^x * (1 - e^{-2x})}{e^x * (1 + e^{-2x})} = f(2x) - \frac{e^{-2x} + 1 - 1}{1 + e^{-2x}} = f(2x) - 1 + \frac{1}{1 + e^{-2x}}$$

$$= f(2x) - 1 + f(2x) = 2 * f(2x) - 1$$

### Παράγωγος της Λογιστικής Συνάρτησης

$$\frac{d}{dx}(f(x)) = f(x) * (1 - f(x))$$

### Ολοκλήρωμα της Λογιστικής Συνάρτησης

είναι η συνάρτηση  $F(x) = \ln(1 + e^x)$  με αρχική Συνθήκη  $f(0) = 1/2$

Στα τεχνητά νευρωνικά Δίκτυα αυτή είναι γνωστή ως τμηματικά λεία (ομαλή) συνάρτηση που είναι προσέγγιση μιας ασυνεχούς συνάρτησης όπως π.χ. η λογιστική Συνάρτηση είναι λεία προσέγγιση της Συνάρτησης Heaviside.

### Διαφορική Εξίσωση Λογιστικής Συνάρτησης

Η Τυπική Λογιστική Συνάρτηση είναι λύση της 1ης τάξης μη γραμμικής διαφορικής εξίσωσης

$$\frac{d}{dx}(f(x)) = f(x) * (1 - f(x))$$

με αρχική συνθήκη  $f(0) = 1/2$

Αυτή η εξίσωση είναι η Συνεχής εκδοχή της Λογιστικής Συνάρτησης.

Η ποιοτική συμπεριφορά αυτής είναι εύκολα κατανοητή σε όρους γραμμικής φάσης. Η παράγωγος είναι 0 όταν η συνάρτηση είναι 1 και η παράγωγος είναι θετική όταν η συνάρτηση είναι μεταξύ 0 και 1 ενώ αλλού η παράγωγος είναι αρνητική. Σημειώστε όμως ότι αρνητικοί πληθυσμοί δε συμφωνούν με το φυσικό περιβάλλον. Αυτό δηλώνει μία μη σταθερή ισορροπία στο 0 και μία σταθερή ισορροπία στο 1. Επομένως για κάθε τιμή της συνάρτησης μεγαλύτερης του 0 και μικρότερης του 1, συγκλίνει στο 1.

Η Λογιστική Διαφορική Εξίσωση είναι Ειδική περίπτωση της Διαφορικής Εξίσωσης Bernoulli η οποία έχει την ακόλουθη λύση

$$f(x) = \frac{e^x}{c + e^x}$$

Για  $c=1$  παίρνουμε τη γνωστή Σιγμοειδή Συνάρτηση

$$f(x) = \frac{e^x}{1 + e^x}$$

Ποσοτικά η γραφική παράσταση της Σιγμοειδής Συνάρτησης δείχνει εκθετική αύξηση για αρνητικά  $x$  όπου η κλίση της αυξάνει μέχρι την τιμή  $\frac{1}{4}$  κοντά στο 0 και έπειτα δείχνει εκθετική μείωση πλησιάζοντας το 1 όπου η κλίση της προσεγγίζει το 0.

Η παραπάνω Διαφορική Εξίσωση είναι η ειδική περίπτωση Γενικής Διαφορικής Εξίσωσης που μόνο μοντελοποιεί τη Σίγμα Συνάρτηση για  $x$  θετικά. Σε πολλές εφαρμογές μοντελοποίησης η πιο Γενική Μορφή είναι:

$$\frac{d}{dx}(f(x)) = \frac{K}{a} * f(x) * (a - f(x)) \quad , \quad f(0) = \frac{a}{1 + e^{K*r}}$$

Η λύση της είναι η παραμετρική σιγμοειδής συνάρτηση

$$a * S * K(x - r)$$

Η Συνάρτηση Υπερβολικής Συνεφαπτομένης προσομοιάζει μία άλλη μορφή Παραγώγου Λογιστικής Συνάρτησης:

$$\frac{d}{dx}(f(x)) = \frac{1}{4} * \text{Sech}^2\left(\frac{x}{2}\right)$$

η οποία αναδεικνύει τη Λογιστική Κατανομή.

### 3.3 Εφαρμογές Λογιστικής Συνάρτησης

Η πρώτη απόδειξη ότι η Λογιστική Συνάρτηση μπορεί να κατέχει στη βάση της τη Στοχαστική Διαδικασία είναι ότι παράγει την πιθανότητα της πρώτης ισότητας ή υπέρβασης ενός φράγματος (φράγμα θεωρείται το σημείο καμπής της γραφικής της παράστασης όπου από εκεί και μετά η κυρτότητα της αλλάζει και η τιμή της βρίσκεται πλέον γρηγορότερα κοντά στη μονάδα ενώ αριστερά από το σημείο καμπής η τιμή της βρίσκεται γρηγορότερα πιο κοντά στο μηδέν), οπότε έχει πολλές εφαρμογές σε όλες τις επιστήμες.

**Στην Οικολογία:** Το μοντέλο Πληθυσμιακής Ανάπτυξης είναι:

$$\frac{dP}{dt} = r * P * \left(1 - \frac{P}{K}\right)$$

όπου  $r$  είναι ο ρυθμός ανάπτυξης,  $P$  είναι ο πληθυσμός και  $K$  είναι ο μέγιστος αριθμός πληθυσμού που μπορεί να υπάρξει για να είναι το σύστημα βιώσιμο.

Είναι δεδομένο επομένως ότι:

$$\lim_{t \rightarrow \infty} P(t) = K$$

Εάν το  $K$  περιοδικά εξαρτάται από το χρόνο  $t$ , όπου το  $t$  μετριέται σε έτη, τότε μπορούμε να επισημάνουμε την πιθανότητα  $P(t)$  του μέγιστου πληθυσμού  $K$  για τον οποίο το σύστημα είναι βιώσιμο.

### Στο Μοντέλο Ανάπτυξης Καρκίνου

Εάν  $X(t)$  είναι το μέγεθος του καρκίνου τη χρονική στιγμή  $t$  η εξέλιξή του δίνεται από τον τύπο

$$x' = r * \left(1 - \frac{x}{k}\right) * x$$

Εάν ο ασθενής ξεκινήσει έγκαιρα τη χημειοθεραπεία τότε ο ρυθμός μείωσης των καρκινικών κυττάρων αν είναι  $c(t)$  η εξίσωση μετατρέπεται σε

$$x' = r * \left(1 - \frac{x}{k}\right) * x - c(t) * x$$

όπου, όταν σε κάποιο ασθενή ισχύσει η σχέση

$$\frac{1}{T} \int_0^T c(t) dt > r \Leftrightarrow \lim_{t \rightarrow \infty} x(t) = 0$$

που σημαίνει ότι ο ασθενής (θεωρητικά) σώζεται. Λέμε θεωρητικά γιατί δεν συνυπολογίστηκε η αντοχή του οργανισμού του ασθενή στην ακτινοβολία.

Εδώ βέβαια γεννιέται το ελπιδοφόρο ερωτηματικό εάν ο ρυθμός μείωσης των καρκινικών κυττάρων μπορούσε να επέλθει με άλλου είδους θεραπεία τότε το παραπάνω αποτέλεσμα θα μπορούσε να φαίνεται πραγματικό.

### Στη Χημεία:

Στα Μοντέλα Χημικών Αντιδράσεων. Η συγκέντρωση των Αντιδρώντων και Προϊόντων σε μία Χημική Αντίδραση χωρίς καταλύτες ακολουθεί τη Λογιστική Συνάρτηση.

### Στη Φυσική:

#### Η Fermion Dirac Κατανομή:

Είναι αυτή που σε οποιοδήποτε ενεργειακό επίπεδο συστήματος μετρά την πιθανότητα το σύστημα να κατέχεται από μία τουλάχιστον Fermion

### Στη Γεωργία: Η Μοντελοποίηση της Συγκομιδής

### Στην Κοινωνιολογία:

Το Μοντέλο Αποδοχής Καινοτομίας μέσα στο Χρόνο.

Ο Gabriel Trade περιγράφει την διάδοση(αποδοχή) της καινοτομίας σε 3 στάδια.

- Στο 1ο αντανακλάται η δυσκολία που βρίσκει αρχικά να διαδοθεί μία καινούργια εφαρμογή , εξαιτίας του εχθρικού περιβάλλοντος (αντίθετες συνήθειες , κάθε τι καινούργιο και άγνωστο τρομάζει)
- Στο 2ο απεικονίζεται η εκθετική διάδοση(αποδοχή) της καινοτομίας. ΚΑΙ
- Στο 3ο απεικονίζεται η παγίωση της αποδοχής πλέον της καινοτομίας γιατί καθώς ο χρόνος περνά μία άλλη καινοτομία εμφανίζεται η οποία απασχολεί τον πληθυσμό.

### Στις Επιχειρήσεις:

Η Charlota Perez (2003) απεικονίζει τον κύκλο της επιχειρηματικότητας του Kondratiev (1920) με τις επόμενες ετικέτες:

1ο) Την αρχή της Τεχνολογικής Εποχής ως ΚΑΤΑΡΡΕΥΣΗ

2ο) Την άνοδό της ως ΤΡΕΛΛΑ

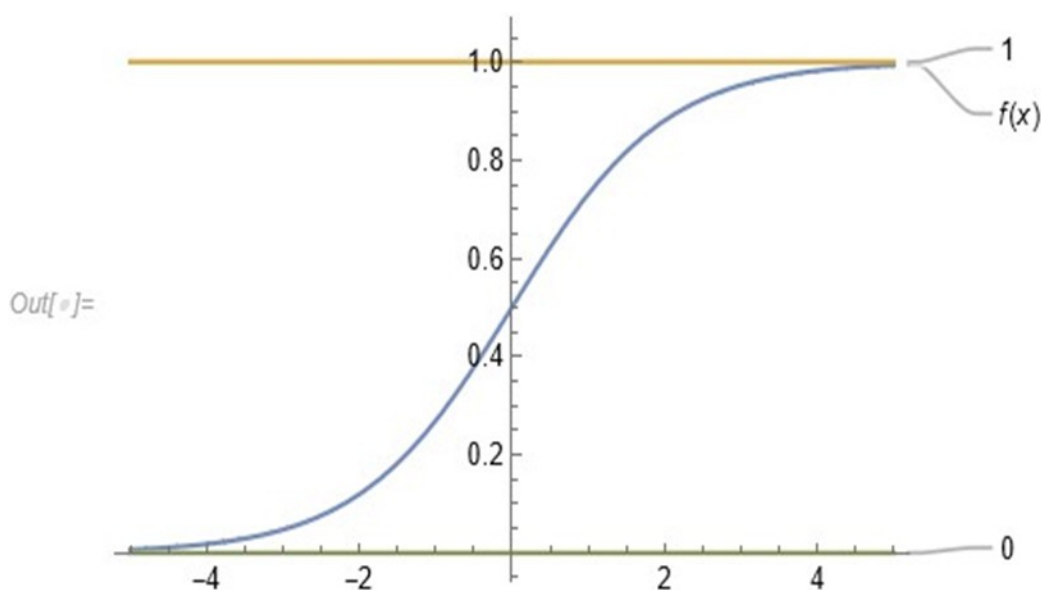
3ο) Το γρήγορο ξέσπασμά τους ως ΣΥΝΕΡΓΕΙΑ

4ο) Την Ολοκλήρωσή της ως ΩΡΙΜΑΝΣΗ.

## ΚΕΦΑΛΑΙΟ 4

### Η ΛΟΓΙΣΤΙΚΗ ΣΥΝΑΡΤΗΣΗ ΚΑΙ Η ΣΧΕΣΗ ΤΗΣ ΜΕ ΤΗΝ ΕΞΑΡΤΗΜΕΝΗ ΜΕΤΑΒΛΗΤΗ Υ ΤΗΣ ΛΟΓΙΣΤΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

#### 4.1 Συνάρτηση Σίγμα- Λογιστική Συνάρτηση και η Σχέση τους.



Η παραπάνω γραφική παράσταση είναι της συνάρτησης σίγμα και ονομάζεται σιγμοειδής καμπύλη. Ο τύπος της συνάρτησης σίγμα

$$S(x) = \frac{1}{1 + e^{-x}}$$

Παρατηρούμε ότι έχει Ασυμπτωτική Συμπεριφορά καθώς  $x \rightarrow -\infty$  και  $x \rightarrow +\infty$

Ειδικότερα στην πρώτη περίπτωση όταν  $x \rightarrow -\infty$  το  $S(x) \rightarrow 0$

και όταν  $x \rightarrow +\infty$  το  $S(x) \rightarrow 1$ . Άρα οι ευθείες  $\psi=0$  και  $\psi=1$  είναι Οριζόντιες Ασύμπτωτες της γραφικής παράστασης της σίγμα συνάρτησης στο  $-\infty$  και στο  $+\infty$  αντίστοιχα. Επομένως το σύνολο τιμών αυτής είναι το  $(0,1)$

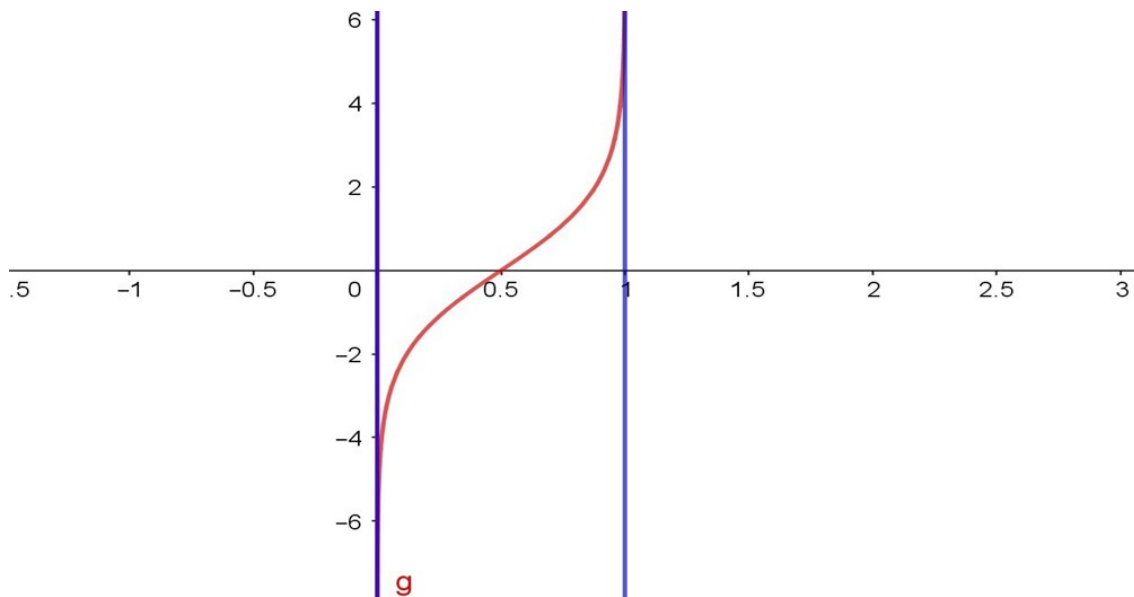


Γνωρίζοντας όμως ότι η πιθανότητα ενός ενδεχομένου παίρνει τιμές στο  $[0,1]$  και με την παραδοχή ότι το  $(0,1)$  σύνολο τιμών προσεγγίζει το  $[0,1]$  καθώς το  $x$  παίρνει ακραίες τιμές στο  $-\infty$  και στο  $+\infty$ , εκτιμήσαμε ότι αν υπήρχε μία συνάρτηση με πεδίο ορισμού το  $\mathbb{R}$  και σύνολο τιμών το  $(0,1)$  αυτή θα προσέγγιζε τα πραγματικά μας δεδομένα.

Αυτή λοιπόν, δεν είναι καμία άλλη παρά η Λογιστική Συνάρτηση, η οποία έχει τύπο:

$$G(x) = \ln\left(\frac{x}{1-x}\right)$$

Η Λογιστική Συνάρτηση είναι η αντίστροφη της φυσικής (logit) Λογαριθμικής Συνάρτησης και επομένως μπορεί να χρησιμοποιηθεί για να μετατρέψει το λογάριθμο της σχετικής συχνότητας (odds) σε πιθανότητα. Η μετατροπή από Λογαριθμική αναλογία σε 2 εναλλακτικές, συμπληρωματικές περιπτώσεις προσεγγίζει τη καμπύλη της Λογιστικής Συνάρτησης η οποία είναι η κόκκινη.



Αναδεικνύεται στο σχήμα η ασυμπτωτική συμπεριφορά της Λογιστικής συνάρτησης.

Δηλαδή

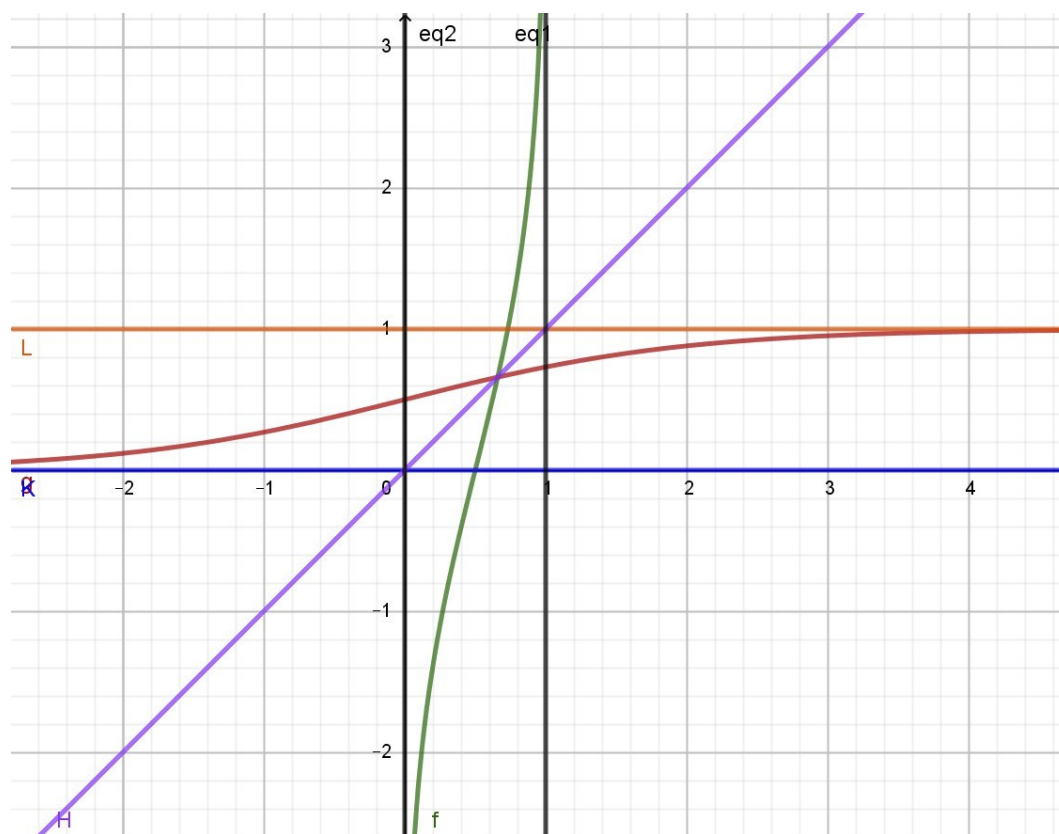
$$\lim_{x \rightarrow 0} G(x) = -\infty$$

Άρα ο άξονας  $y'g$  είναι Κατακόρυφη Ασύμπτωτη αυτής .

και 
$$\lim_{x \rightarrow 1} G(x) = +\infty$$

που σημαίνει ότι η ευθεία  $x=1$  είναι Κατακόρυφη Ασύμπτωτη αυτής.

Οι δύο γραφικές παραστάσεις είναι αντίστροφες μεταξύ τους αφού είναι συμμετρικές ως προς την ευθεία  $y=x$  που είναι η διχοτόμος της 1ης και 3ης γωνίας όπως φαίνεται στο επόμενο σχήμα. Η σίγμα συνάρτηση είναι η κόκκινη ενώ η λογιστική συνάρτηση είναι η πράσινη, τέμνονται πάνω στη διχοτόμο  $y=x$  και είναι συμμετρικές ως προς αυτή (μωβ χρώματος).



#### 4.2 Σχέση της Δίτιμης Εξαρτημένης μεταβλητής $Y_i$ με τη Συνάρτηση Σίγμα και τη Λογιστική Συνάρτηση.

Στο Κεφάλαιο 2 εκτιμήθηκαν τα  $\beta_i$

Οπότε

$$Y_i = b_0 + b_1 * X_{1,i} + b_2 * X_{2,i} + \dots + b_k * X_{k,i} \quad (1)$$

Η εξαρτημένη μεταβλητή  $Y_i$  παίρνει τιμές στο  $R$  για κάθε διατεταγμένη  $k$ -άδα των ανεξάρτητων μεταβλητών

$(X_{1,i}, X_{2,i}, \dots, X_{k,i})$ .

Ακριβέστερα για κάθε διατεταγμένη  $k$ -άδα  $(X_{1,i}, X_{2,i}, \dots, X_{k,i})$ , υπάρχει ένας μοναδικός πραγματικός αριθμός τέτοιος ώστε να ισχύει η σχέση (1).

Αν αυτόν τον αριθμό  $Y_i$  τον εξισώσουμε με το  $G(\pi_i)$

της Λογιστικής Συνάρτησης, επειδή η  $G$  είναι 1-1 θα υπάρχει μοναδικός αριθμός  $\pi_i \in (0,1)$  τέτοιος ώστε  $G(\pi_i) = Y_i$

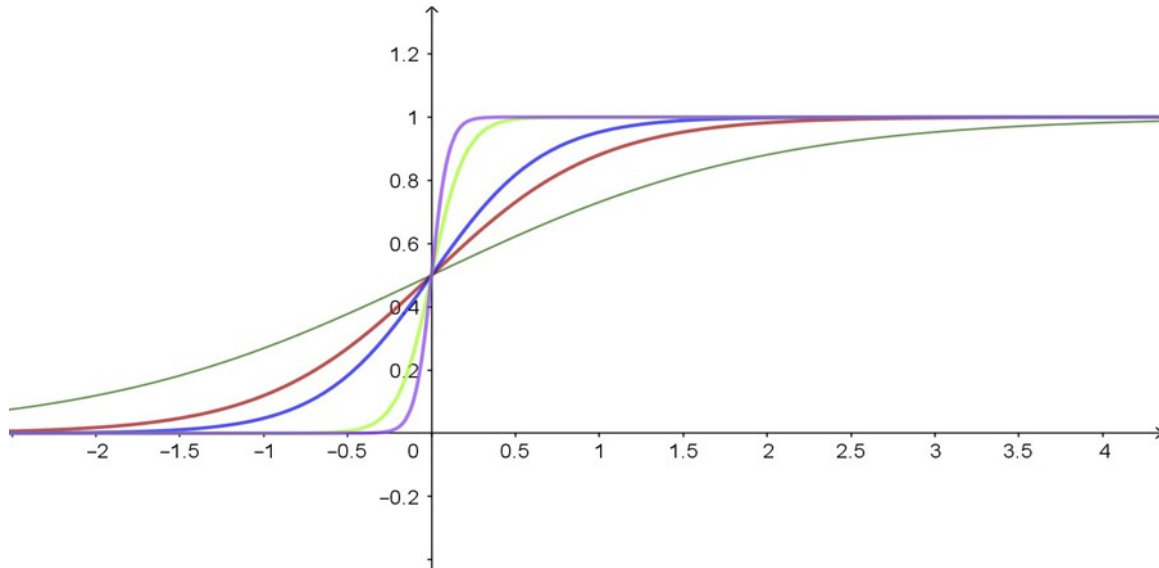
Ο αριθμός  $\pi_i$  είναι η πιθανότητα η εξαρτημένη μεταβλητή  $Y_i$  να πάρει την τιμή  $Y_i = G(\pi_i)$ .

Οπότε ισχύουν οι ισότητες:

$$P(Y=Y_i) = \pi_i = S(X_i) = G^{-1}(Y) \quad (4)$$

Αν τώρα στην  $S$  προσδώσουμε και κάποια βαρύτητα η γραφική παράσταση αυτής θα προσεγγίσει πιο γρήγορα, (απότομα) τις τιμές 0 και 1 που ουσιαστικά είναι οι μοναδικές τιμές που μπορεί να πάρει η δίτιμη μεταβλητή Bernoulli με αποτέλεσμα τα δεδομένα μας να έχουν καλύτερη προσαρμογή. Αυτή η βαρύτητα, προφανώς, έχει δοθεί, και είναι το  $\tilde{\beta}$ , το οποίο εκτιμήθηκε μετά τον έλεγχο προσαρμογής και προβλεψιμότητας του μοντέλου.

Οι παρακάτω γραφικές παραστάσεις είναι της Σίγμα Συνάρτησης για διαφορετικές τιμές του συντελεστή της ανεξάρτητης μεταβλητής  $t$ . Παρατηρούμε ότι όσο αυτός μεγαλώνει τόσο πιο γρήγορα η εξαρτημένη μεταβλητή  $Y$  παίρνει τις τιμές 0 και 1.



Είμαστε υποχρεωμένοι όμως να διαλέξουμε εκείνη τη Σίγμα Συνάρτηση ώστε η προσαρμογή και η προβλεψιμότητα του μοντέλου να είναι καλύτερη.

## ΚΕΦΑΛΑΙΟ 5

### ΕΥΡΩΠΑΪΚΗ ΠΗΓΗ ΔΕΔΟΜΕΝΩΝ SHARE

#### **5.1 Τι είναι η Ευρωπαϊκή Πηγή Δεδομένων SHARE;**

Η Έρευνα SHARE είναι μία βάση δεδομένων, στην οποία συγκεντρώνονται στοιχεία από Ευρωπαίους ηλικίας άνω των 50 και 50 ετών. Τα δεδομένα αυτά αντλούνται από διαφορετικές χώρες της Ευρώπης μέσω ενός δειγματοληπτικά εγκεκριμένου δείγματος ατόμων και αναφέρονται σε διαφορετικούς επιστημονικούς τομείς.

Στην πρώτη προσπάθεια ανάκτησης δεδομένων η έρευνα SHARE, που έγινε το 2004, 11 Ευρωπαϊκές χώρες συνεισέφεραν στοιχεία. Από τη Σκανδιναβία ήταν η Δανία και η Σουηδία, από την Κεντρική Ευρώπη ήταν η Γαλλία, η Ολλανδία, η Ελβετία, η Αυστρία, το Βέλγιο και η Γερμανία και από τη Μεσόγειο ήταν η Ελλάδα, η Ισπανία και η Ιταλία. Στο τέλος του 2004 συνεισέφερε δεδομένα και το Ισραήλ.

Στην δεύτερη προσπάθεια ανάκτησης δεδομένων της έρευνας SHARE, που έλαβε χώρα το 2006-2007, η Τσεχία, η Πολωνία και η Ιρλανδία εισχώρησαν σε αυτήν.

Στην τρίτη προσπάθεια ανάκτησης δεδομένων έλαβαν χώρα συνολικά 16 Ευρωπαϊκές Χώρες με την προσθήκη της Σλοβενίας ως νέο μέλος, το 2008-2009. Τα στοιχεία της Έρευνας συλλέχθηκαν από αναλυτικές διαδρομές ζωής και από αυτό πήρε τον τίτλο SHARELIFE.

Η Έρευνα SHARE αυτή τη στιγμή συμπεριλαμβάνει 27 Ευρωπαϊκές Χώρες καθώς επίσης και το Ισραήλ.

Στο 6ο κύμα συλλογής δεδομένων συμμετείχε και η Ελλάδα η οποία δεν συμμετείχε στην 4η και 5η προσπάθεια. Συνολικά η τράπεζα δεδομένων SHARE συμπεριλαμβάνει στοιχεία που αφορούν 140.000 άτομα (380.000 συνεντεύξεις) ηλικίας 50 ετών και άνω, διαφόρων κοινωνικών τάξεων, οικονομικής και οικογενειακής κατάστασης.

Η μελέτη SHARE αποδίδει την πλήρη εικόνα της διαδικασίας γήρανσης στην Ευρώπη. Τα στοιχεία που έχουν συλλεχθεί περιέχουν μεταβλητές υγείας, βιολογικούς δείκτες, ψυχολογικούς δείκτες και μεταβλητές οικονομικών και κοινωνικών καταστάσεων.

Για παράδειγμα, οι αναφορές για την κατάσταση υγείας του ερωτώμενου είναι οι :

- δείκτης μάζας σώματος,
- παθήσεις,
- ψυχική υγεία,
- ποιότητα ζωής,
- εργασιακή δραστηριότητα,
- χαρακτηριστικά απασχόλησης,
- περιουσιακά στοιχεία,
- στέγαση,
- εκπαίδευση,
- παροχή και
- λήψη οικογενειακής βοήθειας.

Επιπλέον, συμπεριλαμβάνονται μεταβλητές από περιγραφές ερωτώμενων των Ερευνών COMPARE και AMANDA.

## 5.2 Ότι θα Θέλατε να Γνωρίζετε Σχετικά με τα Μεγαλύτερα σε Ηλικία Άτομα στην Ευρώπη

Η αναλογία των ηλικιωμένων στο συνολικό πληθυσμό στην Ευρώπη είναι υψηλότερη από κάθε άλλη ήπειρο – και το φαινόμενο της γήρανσης του πληθυσμού θα συνεχίζεται με αμείωτη ένταση στη διάρκεια αυτού του αιώνα. Η Έρευνα για την Υγεία, τη Γήρανση και τη Συνταξιοδότηση στην Ευρώπη (SHARE), της οποίας τα δεδομένα του πρώτου ερευνητικού κύματος συλλέχθηκαν κατά το 2004-05, είναι η πρώτη επιστημονική μελέτη που εξετάζει τους διαφορετικούς τρόπους ζωής ατόμων ηλικίας 50 ετών και άνω σε έντεκα Ευρωπαϊκές χώρες, από τη Σκανδιναβία έως τη Μεσόγειο.

Η έρευνα βασίστηκε σε ένα αντιπροσωπευτικό δείγμα που ξεπέρασε τα 30.000 άτομα, εστιάζοντας σε θέματα όπως το οικογενειακό περιβάλλον, οι οικογενειακές σχέσεις, η εργασία, οι συντάξεις, η οικονομική ασφάλεια και η υγεία. Τα αποτελέσματα παρέχουν χρήσιμα σημεία αναφοράς για την άσκηση εναλλακτικών πολιτικών στο μέλλον και θα συνεχίσουν να ενισχύονται καθώς θα συμπληρώνονται από τα επόμενα ερευνητικά κύματα της SHARE.

Η πηγή SHARE περιέχει αναλυτικότερα:

- Ότι θα θέλαμε να γνωρίζουμε σχετικά με τα οικονομικά των μεγαλύτερων σε ηλικία ατόμων στην Ευρώπη.
- Ότι θα θέλαμε να γνωρίζουμε σχετικά με την απασχόληση των μεγαλύτερων σε ηλικία ατόμων στην Ευρώπη.
- Ότι θα θέλαμε να γνωρίζουμε σχετικά με την οικογενειακή δικτύωση των μεγαλύτερων σε ηλικία ατόμων στην Ευρώπη.
- Ότι θα θέλαμε να γνωρίζουμε σχετικά με τη υγεία των μεγαλύτερων σε ηλικία ατόμων στην Ευρώπη.

## 5.3 Χρήσιμα Συμπεράσματα από τη διεξαγωγή της Έρευνας SHARE

- *Ο κίνδυνος φτώχειας μετριάζεται από μη-χρηματικές πηγές:*

Τα επίπεδα διαφέρουν σημαντικά μεταξύ των χωρών της Ευρώπης. Ενώ η **φτώχεια και ο κοινωνικός αποκλεισμός** συνεχίζουν να αποτελούν σοβαρά προβλήματα σε ορισμένες χώρες που συμμετείχαν στην έρευνα, ο κίνδυνος της οικονομικής επισφάλειας περιορίζεται εφόσον ληφθεί υπόψη η ιδιοκατοίκηση. Αυτό ισχύει κυρίως για τις χώρες της Νότιας Ευρώπης. Επιπλέον, η εγγύτητα με τα παιδιά (συγκατοίκηση ή διαμονή σε κοντινή απόσταση) παραμένει ένας σοβαρός μηχανισμός που ενδέχεται να μειώνει τη φτώχεια, όχι μόνο στις χώρες της Μεσογείου, αλλά και στη Γερμανία .

- *Η ανισότητα στην κατανάλωση είναι πολύ μικρότερη από την χρηματική ανισότητα.*

- Τα κίνητρα για πρόωρη συνταξιοδότηση επιφέρουν την έξοδο από την αγορά εργασίας προς την πρόωρη συνταξιοδότηση: Τα διαφορετικά συστήματα κοινωνικής προστασίας επηρεάζουν ξεκάθαρα την ηλικιακή κατανομή και τα διαφορετικά ηλικιακά πρότυπα συμμετοχής στην αγορά εργασίας και συνταξιοδότησης. Η πρόωρη συνταξιοδότηση συναντάται περισσότερο συχνά στις

χώρες της Νότιας Ευρώπης καθώς και την Αυστρία και την Γαλλία, στις οποίες επιτρέπεται η πρόωρη συνταξιοδότηση και/ή συχνά προσφέρονται υψηλές παροχές.

- Σαν αποτέλεσμα, εντοπίζεται πως υπάρχει ένα σημαντικό αναξιοποίητο εργατικό δυναμικό σε χώρες όπως η Αυστρία, η Ιταλία και η Γαλλία, όπου πολλά υγιή άτομα βρίσκονται εκτός εργατικού δυναμικού. Παράλληλα, οι θεσμικές διαφορές που ισχύουν ανάμεσα στις χώρες επηρεάζουν τις ατομικές διαδρομές προς τη συνταξιοδότηση. Σχετικά χαμηλά επίπεδα ιδιωτικών συντάξεων στην Ευρώπη: Σχεδόν όλοι οι ηλικιωμένοι πολίτες στην Ευρώπη λαμβάνουν με μια δημόσια σύνταξη. Περισσότεροι από τους μισούς συμπληρώνουν τη σύνταξη αυτή με πρόσθετες πηγές εισοδήματος όπως είναι οι ιδιωτικές συντάξεις. Σε χώρες όπως η Σουηδία ή η Ολλανδία σχεδόν τα δύο τρίτα των συνταξιούχων έχουν μια συμπληρωματική επαγγελματική σύνταξη. Ωστόσο, στις περισσότερες χώρες οι επαγγελματικές συντάξεις αντιστοιχούν σε ένα πολύ μικρότερο μερίδιο του εισοδήματος από ότι η δημόσια σύνταξη. Ακόμη λιγότερη μέριμνα λαμβάνει χώρα ιδιωτικά, για παράδειγμα, διαμέσου της ιδιωτικής ασφάλισης ζωής. Η Σουηδία ιεραρχείται υψηλότερα σε αυτό το ζήτημα με σχεδόν το 20 τοις εκατό των συνταξιούχων να εισπράττουν τέτοιου είδους εισοδήματα. Σε αντίθεση, ο Ευρωπαϊκός μέσος όρος είναι 10 τοις εκατό. Οι παροχές συντάξεων είναι μικρότερες για τα άτομα με χαμηλή ειδίκευση – ειδικά αν λάβουν σύνταξη πρόωρα.
- *Οι καλές συνθήκες εργασίας υποστηρίζουν την παράταση του εργασιακού βίου:* Η ποιότητα της κατά το διάστημα πριν τη συνταξιοδότηση - για παράδειγμα πόσο έλεγχο ασκούμε πάνω στη δουλειά μας καθώς και πόσο ανταμείβονται οι προσπάθειές μας - διαφέρει σημαντικά μεταξύ των χωρών της Ευρώπης για τις οποίες έχει γίνει διερεύνηση, με την ποιότητα της εργασίας να είναι καλύτερη στην Βόρεια Ευρώπη σε σχέση με τη Νότια. Επίσης, η ποιότητα της απασχόλησης σχετίζεται έντονα με την ευημερία: Η χαμηλή ποιότητα απασχόλησης συνοδεύεται από κακή κατάσταση υγείας και κατάθλιψη. Η SHARE φανερώνει, επίσης, ότι υπάρχουν σημαντικές διαφορές ανάμεσα στις χώρες όσον αφορά την ποιότητα ζωής των ατόμων.
- Οι οικογενειακοί δεσμοί έχουν παραμείνει ισχυροί ανάμεσα στις γενεές: Η οικογένεια διαδραματίζει σημαντικό ρόλο στις ζωές των ηλικιωμένων ατόμων σε όλη την Ευρώπη. Σήμερα, περισσότεροι άνθρωποι παρά ποτέ έχουν παιδιά όταν προσεγγίζουν τη μεγαλύτερη ηλικία. Παρόλα αυτά, τα δεδομένα της SHARE δείχνουν πως υπάρχουν αξιοσημείωτες διαφορές μεταξύ νότιας και βόρειας Ευρώπης σχετικά με τους οικογενειακούς διακανονισμούς, την οικογενειακή δομή, εγγύτητα και επαφές. Στη Δανία, για παράδειγμα, μονάχα 13 τοις εκατό των ερευνώμενων ζουν με τα παιδιά τους. Στην Ισπανία η αντιστοιχία είναι 52 τοις εκατό, ενώ πάνω από 80 τοις εκατό ζουν πολύ κοντά στα παιδιά τους. Μια παρόμοια εικόνα είναι εμφανής σχετικά με το πόσο συχνά έχουν επαφή οι γονείς με τα παιδιά τους: ενώ 42 τοις εκατό των Δανών βλέπουν τα παιδιά τους κάθε μέρα το αντίστοιχο στοιχείο για τις Μεσογειακές χώρες φθάνει το 86 τοις εκατό. Το δυναμικό για καθημερινή υποστήριξη εντός της οικογένειας είναι υψηλό σε όλη την Ευρώπη. Ειδικά ο χρόνος που διατίθεται για την παροχή βοήθειας προς άλλα μέλη ή για τη φροντίδα των εγγονιών είναι ιδιαίτερα σημαντικός σε μέγεθος. Το ένα τρίτο των ερευνώμενων ηλικίας άνω των 65 ετών δήλωσαν πως παρέχουν καθημερινή βοήθεια προς άλλα μέλη ή φροντίδα προς τα εγγόνια, και πως διέθεσαν κατά μέσο όρο 4,6 ώρες την ημέρα σε τέτοιου είδους δραστηριότητες.

## ➤ 5.6 ΕΥΡΗΜΑΤΑ ΤΗΣ SHARE

- **Hank, K. and Erlinghagen, M. (2010)** Volunteering in “Old” Europe. *Journal of Applied Gerontology*, 29, 3-20.

### Περίληψη:

Αυτό το άρθρο στοχεύει να παρέχει μία συνολική εικόνα σχεδίων ,δυνατοτήτων και περιορισμών του επίσημου εθελοντισμού ανάμεσα σε ηλικιωμένους Ευρωπαίους.. Βασισμένη σε δεδομένα από μελέτη του 2004 για την υγεία , γηρατειά και συνταξιοδότηση στην Ευρώπη, αρχικά περιγράφουμε τις διεθνείς αποκλίσεις της επίσημης εθελοντικής ενασχόλησης και αναφέρουμε πρόσφατη ανάλυση που ερευνά τους κοινωνικούς , καθοριστικούς παράγοντες εθελοντισμού. Θα συζητήσουμε γενικά θέματα σχετικά με τακτικές και προγράμματα που προωθούν την εθελοντική ενασχόληση των ηλικιωμένων . Αυτό συμπληρώνεται από κάποιες κριτικές παρατηρήσεις των πρόσφατων “παραγωγικά γηρατειά” συζητήσεων στην Ευρώπη. Αναδεικνύοντας τα εμπόδια του εθελοντισμού που υπάρχουν στον ηλικιωμένο πληθυσμό, καταλήγουμε προτείνοντας μια ευρύ αντίληψη γήρανσης η οποία συμπεριλαμβάνει και παραγωγικά και καταναλωτικά στοιχεία σαν ένα κατάλληλο μοντέλο για την κοινωνία των ηλικιωμένων.

- **Igel, C., Brandt, M., Haberken, K., et al (2009)**

Ιδιαιτερότητες μεταξύ Οικογένειας και Πολιτείας. Διεθνείς Χρονικοί Ανταποκριτές της Δυτικής Ευρώπης. Δημοσιογραφικές Μελέτες Συγκρινόμενων Οικογενειών 40, 203-227.

### Περίληψη:

Χρονικοί Ανταποκριτές της Δυτικής Ευρώπης διαφοροποιούνται και χωρίζονται σε δύο υποστηρικτικές κατηγορίες: Δραστηριότητες Βοήθειας και Φροντίδας. Οι Ενήλικες υποστηρίζουν τους ηλικιωμένους γονείς τους με περισσότερη ή λιγότερη θέληση και ευρέως διαφορετική προσφορά που κυμαίνεται από βοήθεια σε εσωτερικές και εξωτερικές δουλειές σπιτιού μέχρι προσωπική φροντίδα. Ωστόσο , ηλικιωμένοι άνθρωποι είναι επίσης μία σημαντική πηγή , διεθνώς , υποστήριξης , καθώς βοηθούν τα παιδιά τους, φροντίζοντας τα εγγόνια τους για παράδειγμα. Γενικά διεθνή σχέδια αλληλεγγύης είναι επηρεασμένα από την ευκαιρία , την ανάγκη , την οικογένεια, και σχετικά με τα έθιμα υποστάσεις , τα οποία έχουν διαφορετικές επιπτώσεις στη βοήθεια και φροντίδα. Κατά κύριο λόγο η φροντίδα εξαρτάται από την ανάγκη που έχει αυτός που τη λαμβάνει , ενώ οι δραστηριότητες βοήθειας στους γονείς κυρίως εξαρτώνται από την ευκαιρία (ελεύθερο χρόνο) που έχει αυτός που τη δίνει. Επιπρόσθετα , χρησιμοποιώντας τα δεδομένα της SHARE

- η λογιστική πολυεπίπεδη ανάλυση επιτρέπει την εθνική βοήθεια και φροντίδα να αναζητώνται στις δημόσιες παροχές.
- Τα εμπειρικά ευρήματα υποστηρίζουν την “εξειδικευμένη υπόθεση” .



- Το υψηλότερο εθνικό επίπεδο κοινωνικών δημοσίων παροχών συμπίπτει με τη λιγότερη εντατική βοήθεια και περισσότερη απαιτητική φροντίδα.
- Οι πλούσιες πολιτείες επομένως μειώνουν το ρίσκο της εξασθένησης μιας οικογένειας και θωρακίζουν την παντελή υποστήριξη των γηραιότερων και νέων οικογενειών διαμέσου αποτελεσματικής συνεργασίας μεταξύ οικογένειας και πολιτείας.
- **Kalbarczyk, M. and Nicinska, A.** Αλτρουισμός στη μοντέρνα οικογένεια :Προσωπικές δωρεές μεταξύ γονιών και των βιολογικών και μη-βιολογικών παιδιών. CenEAS; Εργατικό Σύγγραμμα 09/09/2010.

#### **Απόσπασμα:**

- Είναι αναγνωρισμένο εδώ και πολύ καιρό στη λογοτεχνία ότι υπάρχουν σχέσεις αλτρουισμού μέσα σε μία οικογένεια από γονείς και βιολογικά παιδιά .Η διένεξη στις πηγές αλτρουισμού αναφέρεται κυρίως σε κοινωνικο-βιολογικές και γενετικές βελτιώσεις παρ'όλο που υπάρχουν συζητήσεις για άλλο μηχανισμό λειτουργίας βασιζόμενο περισσότερο στην κοινωνικές παρά στη βιολογικές σχέσεις. Οι αλλαγές στο σχήμα της οικογένειας , η μονιμότητα και και το οικοδόμημα που έλαβε χώρα στον αιώνα μας ίσως έχει επηρεάσει σημαντικά τις σχέσεις μέσα στην οικογένεια. Επιπρόσθετα , η Ευρώπη αντιμετωπίζει ευκαιρίες και απειλές προερχόμενες από τα γηρατεία τα οποία δημιουργούν καθοριστικές αντιδράσεις , ιδιαίτερα σχετικές. Για αυτούς τους λόγους η υπόθεση της συγγενικής αλτρουιστικής παρουσίας μέσα στις οικογένειες θα διερευνηθεί με περισσότερη λεπτομέρεια. Βρίσκουμε μαρτυρία σε δείγμα πληθυσμού 50 ετών και πάνω δείχνοντας ότι δεν υπάρχει λόγος να πιστεύουμε ότι οι γονείς πάνω από 50 ετών συμπεριφέρονται στα βιολογικά τους παιδιά διαφορετικά από ότι στα μη-βιολογικά τους παιδιά μέχρι που οι ζωντανόι συγγενείς ενδιαφέρονται. Ωστόσο παιδιά μεταξύ των οποίων υπάρχει ένα μη-βιολογικό παιδί είναι λιγότερο πιθανόν να παρέχουν οικονομική και μη-οικονομική βοήθεια στους γονείς τους.
- **Tsimbos, C.(2009)** Μία Εκτίμηση των Κοινωνικο-Οικονομικών Ανισοτήτων στην Υγεία ανάμεσα στους Ηλικιωμένους στην Ελλάδα , Ιταλία ,και Ισπανία. Διεθνή Δημοσιογραφία Δημόσιας Υγείας , 55, 5-15.

**Σκοπός:** Η μελέτη αυτή ερευνά τις κοινωνικο -οικονομικές ανισότητες στη υγεία μεταξύ των κατοίκων της Μεσογείου που έχουν ηλικία 50 και πάνω έτη.

**Μέθοδος:** Τα δεδομένα που χρησιμοποιήθηκαν στην ανάλυση προέρχονται από τη Μελέτη της Υγείας , Γηρατεία και Συνταξιοδότηση στην Ευρώπη , κύμα 1, κύμα 2, το δείγμα συμπεριλαμβάνει 2671 Έλληνες , 2502 Ιταλούς και 2343 Ισπανούς. Επτά δείκτες Υγείας εξετάζονται χρησιμοποιώντας ηλικία-φύλλο σταθεροποιημένες συχνότητες και μοντέλα λογιστικής παλινδρόμησης. Πίνακες Συγκέντρωσης υπολογίζονται για αυτοαξιολογική Υγεία.

**Αποτελέσματα:** Η Κοινωνικο-οικονομική θέση των ατόμων αποκλίνει με την ηλικία. Άτομα χαμηλής κοινωνικο-οικονομικής επιφάνειας έχουν τη χειρότερη υγεία σε όλες τις περιστάσεις.

Ανεξάρτητα από την εκπαίδευση και το φύλο οι Έλληνες εκθέτουν τη χαμηλότερη συχνότητα της αυτοαξιολογημένης υγείας και σωματικά και καταθλιπτικά συμπτώματα, οι Ισπανοί εκθέτουν την υψηλότερη κλίμακα για χρόνιες συνθήκες και οι Ιταλοί παρουσιάζουν καλύτερη θεωρώντας λειτουργικούς περιορισμούς. Η ανάλυση μεταξύ των 3 κρατών δείχνει ότι η σημασία των κοινωνικο-οικονομικών ανισοτήτων στη αυτοαξιολογική υγεία είναι μεγαλύτερη στην Ελλάδα, ακολουθούμενη από την Ισπανία και τελευταία από την Ιταλία.

### **Συμπεράσματα:**

Η ανάλυση επιβεβαιώνει το πλεονέκτημα της υψηλής έναντι της χαμηλής κοινωνικο-οικονομικής θέσης για όλες τις χώρες στους δείκτες υγείας και αποδεικνύει ότι η εκπαίδευση είναι ένας σημαντικός συσχετισμός συγκρινόμενη με τον πλούτο και το εισόδημα ανάμεσα στους γηραιότερους.

([info@share-project.org](mailto:info@share-project.org))

### ΠΡΟΣΦΑΤΗ ΕΡΕΥΝΑ ΣΕ ΕΞΕΛΙΞΗ

- Από το 8 και 9 κύμα της τράπεζας δεδομένων SHARE έχουμε το ερωτηματολόγιο και τα δεδομένα που έγινε σε Ευρωπαίους και Ισραηλινούς άνω των 50 ετών μέσω τηλεφωνικής κλίσης διάρκειας 30 λεπτών κατά την εποχή της πανδημίας του COVID19.
- Αυτό αφορά την έρευνα γύρω από :
  - 1) Την πρόσβαση που είχε κάποιος σε είδη περίθαλψης και φροντίδας αφού είτε οικειοθελώς είτε από τους παρόχους φροντίδας γίνονταν ακυρώσεις στα ραντεβού του λόγω έλλειψης προσοχής ή μεγάλου χρόνου αναμονής .
  - 2) Την απώλεια της εργασίας κάποιου
  - 3) Την απώλεια του εισοδήματος κάποιου
  - 4) Τον αντίκτυπο της πανδημίας στις κοινωνικές σχέσεις κάποιου
  - 5) Την δυσκολία που είχε κάποιος να παρέχει ή να λαμβάνει φροντίδα κατά την διάρκεια της πανδημίας.

(Αναφορές: [Gerret, N. Boerma, W., Kroneman, M., Hutchinson, A. & Saltman, R. B. \(2012\). Φροντίδα στο σπίτι σε όλη την Ευρώπη : Τρέχουσες υποδομές και μελλοντικές προκλήσεις. Διαθέσιμο στην ιστοσελίδα \[https://intranet.euro.who.int/\\\_data/assets/pdf\\\_file/0008/181799/e96757.pdf\]\(https://intranet.euro.who.int/\_data/assets/pdf\_file/0008/181799/e96757.pdf\)](#)

### 5.4 Δημοσιεύσεις Ελλήνων Ερευνητών μέσω δεδομένων της Τράπεζας SHARE

- Λυμπεράκη, Α., Τήνιος, Πλ. και Φιλαλήθης, Τ. επιμ., 2009, *‘Ζωή 50+: Υγεία, γήρανση και σύνταξη στην Ελλάδα και στην Ευρώπη’*, εκδ. Κριτική.

- **Λυμπεράκη, Α., 2009.** Πολυδιάστατες ανισότητες και το χάσμα των φύλων μετά την ηλικία των 50: συγκριτικές επιδόσεις της Ελλάδας. Εισήγηση στο Συνέδριο «Κοινωνικές Ανισότητες στη Σύγχρονη Ελλάδα», Πάντειο Πανεπιστήμιο, Αθήνα, Νοέμβριος 2009.
- **Γεωργιάδης Θ., 2010.** Μοναχικοί ηλικιωμένοι στην Ευρώπη: Πού και Πόσο φαίνεται η απουσία της «οικογενειακής αλληλεγγύης»; Εισήγηση στο 11ο Πανελλήνιο Συνέδριο Γεροντολογίας Γηριατρικής, Αθήνα 4-6 Φεβρουαρίου.
- **Λαμπρέλλη, Δ., 2010.** Πόσο κοστίζει η υγεία στους ηλικιωμένους; Εισήγηση στο 11ο Πανελλήνιο Συνέδριο Γεροντολογίας Γηριατρικής, Αθήνα 4-6 Φεβρουαρίου.
- **Παπαδούδης, Γ., 2010.** Εργασιακό παρελθόν και κατάσταση υγείας των ατόμων ηλικίας 80+ στην Ελλάδα και στην Ευρώπη. Εισήγηση στο 11ο Πανελλήνιο Συνέδριο Γεροντολογίας Γηριατρικής, Αθήνα 4-6 Φεβρουαρίου.

## 5.5ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ R

Εισαγωγή στο R

- Το R είναι μία γλώσσα και ένα περιβάλλον για στατιστικούς υπολογισμούς και γραφικά. Είναι ένα έργο GNU παρόμοιο με τη γλώσσα και το περιβάλλον S που αναπτύχθηκε στα εργαστήρια Bell (πρώην AT & T, τώρα Lucent Technologies) από τον John Chambers και τους συνεργάτες του. Το R μπορεί να θεωρηθεί ως μία διαφορετική υλοποίηση του S. Υπάρχουν μερικές σημαντικές διαφορές, αλλά μεγάλο μέρος του κώδικα που γράφτηκε για το S, εκτελείται αμετάβλητο στο R.
- Το R παρέχει μια μεγάλη ποικιλία στατιστικών (γραμμική και μη γραμμική μοντελοποίηση, κλασικές στατιστικές δοκιμές, ανάλυση χρονοσειρών, ταξινόμηση, ομαδοποίηση ...) και γραφικές τεχνικές και είναι εξαιρετικά επεκτάσιμη.
- Η γλώσσα S είναι συχνά το όχημα επιλογής για έρευνα στη στατιστική μεθοδολογία και η R παρέχει μία διαδρομή ανοικτού κώδικα για τη συμμετοχή σε αυτή τη δραστηριότητα.
- Έχει δοθεί μεγάλη προσοχή στις προεπιλογές για τις μικρές σχεδιαστικές επιλογές στα γραφικά, αλλά ο χρήστης διατηρεί τον πλήρη έλεγχο.
- Το R είναι διαθέσιμο ως Ελεύθερο Λογισμικό υπό τους όρους της Γενικής Δημόσιας Άδειας GNU του Ιδρύματος Ελεύθερου Λογισμικού σε μορφή πηγαίου κώδικα. Μεταγλωττίζει και εκτελείται σε μια μεγάλη ποικιλία πλατφορμών UNIX και παρόμοιων συστημάτων (συμπεριλαμβανομένων των FreeBSD και Linux), Windows και MacOS.

### Το περιβάλλον R

Το R είναι μια ολοκληρωμένη σουίτα εγκαταστάσεων λογισμικού για χειρισμό δεδομένων, υπολογισμούς και γραφική απεικόνιση. Περιλαμβάνει

- μια αποτελεσματική εγκατάσταση διαχείρισης και αποθήκευσης δεδομένων,
- μια σουίτα τελεστών για υπολογισμούς σε πίνακες, ιδίως πίνακες,
- μια μεγάλη, συνεκτική, ολοκληρωμένη συλλογή ενδιάμεσων εργαλείων για την ανάλυση δεδομένων,
- γραφικές εγκαταστάσεις για ανάλυση και εμφάνιση δεδομένων είτε στην οθόνη είτε σε έντυπη μορφή και
- μια καλά ανεπτυγμένη, απλή και αποτελεσματική γλώσσα προγραμματισμού που περιλαμβάνει όρους, βρόχους, αναδρομικές συναρτήσεις που καθορίζονται από τον χρήστη και εγκαταστάσεις εισόδου και εξόδου.

Ο όρος «περιβάλλον» προορίζεται να το χαρακτηρίσει ως ένα πλήρως σχεδιασμένο και συνεκτικό σύστημα, παρά ως μια σταδιακή αύξηση πολύ συγκεκριμένων και ανελαστικών εργαλείων, όπως συμβαίνει συχνά με άλλα λογισμικά ανάλυσης δεδομένων.

Το R, όπως και το S, έχει σχεδιαστεί γύρω από μια αληθινή γλώσσα υπολογιστή και επιτρέπει στους χρήστες να προσθέτουν επιπλέον λειτουργίες ορίζοντας νέες λειτουργίες. Μεγάλο μέρος του συστήματος είναι το ίδιο γραμμένο στη διάλεκτο R του S, γεγονός που διευκολύνει τους χρήστες να ακολουθήσουν τις αλγοριθμικές επιλογές που έγιναν. Για εργασίες με υπολογιστική ένταση, οι κώδικες C, C++ και Fortran μπορούν να συνδεθούν και να κληθούν κατά την εκτέλεση. Οι προχωρημένοι χρήστες μπορούν να γράψουν κώδικα C για να χειριστούν απευθείας αντικείμενα R.

Πολλοί χρήστες σκέφτονται το R ως ένα στατιστικό σύστημα. Προτιμούμε να το σκεφτόμαστε ως ένα περιβάλλον μέσα στο οποίο εφαρμόζονται στατιστικές τεχνικές. Το R μπορεί να επεκταθεί (εύκολα) μέσω πακέτων. Υπάρχουν περίπου οκτώ πακέτα που παρέχονται με τη διανομή R και πολλά άλλα είναι διαθέσιμα μέσω της οικογένειας δικτυακών τόπων CRAN που καλύπτουν ένα πολύ ευρύ φάσμα σύγχρονων στατιστικών.

- Η R έχει τη δική της μορφή τεκμηρίωσης που μοιάζει με LaTeX, η οποία χρησιμοποιείται για την παροχή ολοκληρωμένης τεκμηρίωσης, τόσο on-line σε διάφορες μορφές όσο και σε έντυπη μορφή.
- Η R είναι μια γλώσσα προγραμματισμού ανοικτού κώδικα και περιβάλλον που παρέχει στο χρήστη τη δυνατότητα να κάνει υπολογιστική στατιστική και γραφήματα. Παρέχει τα απαραίτητα εργαλεία προκειμένου να υλοποιηθεί μία στατιστική ανάλυση. Κάποια από τα στατιστικά της εργαλεία είναι:

1. δημιουργία τυχαίων δειγμάτων
2. διακριτές και συνεχείς μεταβλητές (Poisson ,Gamma , Exponential,κτλ)
3. έλεγχοι υποθέσεων
4. στατιστικά τεστ(Kolmogorov-Smirnoff)
5. δημιουργία γραφημάτων(ιστόγραμμα , qq ,plot ,pie chart , bar chart κτλ)

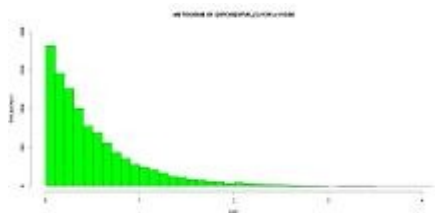
Είναι GNU λογισμικό , δηλαδή ο καθένας μπορεί να κάνει βελτιώσεις στον πηγαίο κώδικα της R και να τις δημοσιεύει και επίσης διανέμεται δωρεάν. Λόγω του ότι όλοι έχουν πρόσβαση στον πηγαίο κώδικά της , έχουν γίνει πολλές βελτιώσεις από τότε που δημιουργήθηκε. Οι αρχικοί δημιουργοί ήταν οι Robert Gentleman και Ross Ihaka, αργότερα όμως συνέβαλλαν και άλλοι στην ανάπτυξή της.

### Πλεονεκτήματα και Μειονεκτήματα

Η R προέρχεται από την S, αλλά η R μπορεί να κάνει τα ίδια πράγματα με την S και με λιγότερο κώδικα. Είναι open source περιβάλλον άρα όλοι μπορούν να έχουν πρόσβαση στον κώδικά της και να κάνουν διορθώσεις. Επιπλέον επιτρέπει στο χρήστη να αλληλεπιδρά και με άλλες γλώσσες(C/C++, Java , Python), με αρχεία δεδομένων (Excel, Access) και με άλλα στατιστικά πακέτα (SASS, SPSS , MINITAB, STATA). Τέλος το περιβάλλον της R παρέχεται δωρεάν. Η R δε συνιστάται για ανάλυση μεγάλων δεδομένων. Το βασικό μειονέκτημα της R είναι ότι καταναλώνει πολύ μνήμη. Είναι "αργή" γλώσσα και επομένως όχι τόσο αποδοτική ως προς το χρόνο εκτέλεσης των εντολών της.

### Εφαρμογές

#### Ιστόγραμμα Εκθετικής κατανομής



Λόγω ότι η R έχει πάνω από 15000 πακέτα , χρησιμοποιείται σε πολλούς επιστημονικούς τομείς. Αρχικά χρησιμοποιείται σε εταιρίες όπως η Google , LinkendIn , Facebook που κάνουν ανάλυση δεδομένων . Επιπλέον χρησιμοποιείται και σε άλλους τομείς , όπως στα οικονομικά , στην αστρονομία , στην χημεία , στην φαρμακευτική , στην ιατρική και στο μάρκετινγκ κ.τ.λ.

R. Chapman & Hall/CRC, Boca Raton, FL, 2004.

## ΚΕΦΑΛΑΙΟ 6

Το πρώτο παράδειγμα της Λογιστικής Παλινδρόμησης, όπως αναφέραμε νωρίτερα στην περίληψη είναι να εκτιμηθεί η πιθανότητα ενός Ευρωπαίου ή Ισραηλινού, ηλικίας άνω των 50 ετών να αποκτήσει κατάθλιψη.

Η εξαρτημένη μεταβλητή παίρνει την τιμή 1 δηλαδή  $Y=1$ , όταν το άτομο έχει κατάθλιψη και παίρνει την τιμή 0,  $Y=0$ , όταν το άτομο ΔΕΝ έχει κατάθλιψη.

Με τη βοήθεια της Λογιστικής Παλινδρόμησης θα εξετάσουμε κατά πόσο αυτή εξαρτάται από τις ανεξάρτητες μεταβλητές  $X_i$  όπου  $i=1, \dots, n$ .

Αναλυτικότερα

η  $X_1$  εκφράζει αν κάποιος ζει μόνος του ή όχι με τιμές 1, 0 αντίστοιχα.

η  $X_2$  εκφράζει το αν το άτομο είναι άνδρας ή Γυναίκα με τιμές 1 αν είναι άνδρας και 2 αν είναι γυναίκα.

η  $X_3$  εκφράζει την ηλικία του ατόμου με ακέραιες θετικές τιμές πάνω από 50

η  $X_4$  εκφράζει κατά πόσο το άτομο αξιολογεί την υγεία του με τιμές 1,2,3,4,5, (1=άριστη,2=πολύ καλή,3=καλή,4=κακή,5=πολύ κακή).

η  $X_5$  εκφράζει αν έχει περιορισμό στις δραστηριότητες ή όχι με τιμές 1,0 αντίστοιχα.

η  $X_6$  εκφράζει αν τον αριθμό των χρόνων παθήσεων που έχει το άτομο με τιμές 1,2,3,4,5,6,7,8,9,10,11

η  $X_7$  εκφράζει τη σωματική του ανικανότητα ή όχι με τιμές 1,0 αντίστοιχα

η  $X_8$  εκφράζει το κατά πόσο το άτομο συντηρείται οικονομικά, με τιμές 1=άριστη οικονομική συντήρηση, 2= καλή, 3=μέτρια, 4= κακή.

Αρχικά θα εισάγουμε τα δεδομένα στην R και θα ελέγξουμε πόσο σημαντικές είναι οι ανεξάρτητες μεταβλητές ως προς την εξαρτημένη. Αν παρατηρηθεί ότι κάποια είναι σχετικά μη σημαντική θα την αφαιρέσουμε από τα δεδομένα για να εκτιμήσουμε την πιθανότητα ένα άτομο να πάθει κατάθλιψη μετά την ηλικία των 50 ετών.

Στη συνέχεια θα εκτιμήσουμε την προβλεψιμότητα της Λογιστικής Συνάρτησης με την σχεδίαση της καμπύλης ROC.

```
> anova(model2)
```

```
Analysis of Deviance Table
```

```
Model: binomial, link: logit
```

```
Response: Y
```

```
Terms added sequentially (first to last)
```

	Df	Deviance Resid.	Df Resid.	Dev
	NULL		1879	1994.8
X1	1	2.312	1878	1992.4
X2	1	0.037	1877	1992.4
X3	1	7.963	1876	1984.4
X4	4	301.831	1872	1682.6
X5	1	22.447	1871	1660.2
X6	1	16.568	1870	1643.6
X7	1	19.334	1869	1624.3
X8	3	38.507	1866	1585.8

Ισχύουν τα στατιστικά  $\chi^2_{1,0.95}=3.84$  ,  $\chi^2_{3,0.95}=7.81$  και  $\chi^2_{4,0.95}=9.49$

➤ Επειδή  $3.84 > 2.312$  και  $3.84 > 0.037$  οι ανεξάρτητες μεταβλητές X1 , X2 θεωρούνται **Μη Σημαντικές**

➤ Όλες οι υπόλοιπες ανεξάρτητες μεταβλητές με 1 βαθμό ελευθερίας έχουν μεγαλύτερο Deviance Resid. από το 3,84 οπότε θεωρούνται σημαντικές.

➤ Επειδή  $7.81 < 38.507$  η μεταβλητή X8 θεωρείται σημαντική

➤ Και επειδή  $9.49 < 301.831$  η μεταβλητή X4 θεωρείται σημαντική

Άρα θα αφαιρέσουμε τις **Μη σημαντικές** X1 ,X2 μεταβλητές και θα ξανατρέξουμε τη Λογιστική Παλινδρόμηση  
anova(model2)

Analysis of Deviance Table

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance Resid.	Df Resid.	Dev
	NULL		1879	1994.8
X3	1	9.299	1878	1985.5
X4	4	300.062	1874	1685.4
X5	1	22.936	1873	1662.5
X6	1	16.040	1872	1646.4
X7	1	19.207	1871	1627.2
X8	3	38.758	1868	1588.5

summary(model2)

Call:

glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0911	-0.6265	-0.4251	-0.2361	2.6132

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.557994	1.167811	-2.190	0.028494 *
X3	-0.003018	0.007721	-0.391	0.695894
X42	0.283073	1.056291	0.268	0.788709
X43	0.864497	1.036797	0.834	0.404385
X44	1.516475	1.043650	1.453	0.146210
X45	2.086441	1.063444	1.962	0.049766 *
X51	0.496601	0.154924	3.205	0.001348 **
X6	0.179472	0.050701	3.540	0.000400 ***
X71	1.051468	0.237631	4.425	9.65e-06 ***
X82	-0.646991	0.134918	-4.795	1.62e-06 ***
X83	-1.104316	0.298796	-3.696	0.000219 ***
X84	-1.766155	0.656192	-2.692	0.007113 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1994.8 on 1879 degrees of freedom  
 Residual deviance: 1588.5 on 1868 degrees of freedom  
 AIC: 1612.5

Number of Fisher Scoring iterations: 5

>Area under the curve: 0.7789

Number of Fisher Scoring iterations: 5

Και η προσαρμογή και η προβλεψιμότητα είναι καλή.

Το Λογιστικό Μοντέλο που προκύπτει είναι:



$$\log\left(\frac{p}{1-p}\right) = -2,557994 - 0,003018 * X_3 + 0,283073 * X_{42} + 0,864497 * X_{43} + 1,516475 * X_{44} \\ + 2,086441 * X_{45} + 0,496601 * X_5 + 0,179472 * X_6 + 1,051468 * X_7 \\ - 0,646991 * X_{82} - 1,104316 * X_{83} - 1,766155 * X_{84}$$

Οι Ανεξάρτητες μεταβλητές  $X_{42}$ ,  $X_{43}$ ,  $X_{44}$ ,  $X_{45}$  παίρνουν την τιμή 1 μόνο όταν είναι αντίστοιχα ίσες με 2,3,4,5 ενώ όταν όλες παίρνουν 0 τότε με βεβαιότητα 100% η  $X_4=1$

Επίσης οι Ανεξάρτητες μεταβλητές  $X_{82}$ ,  $X_{83}$ ,  $X_{84}$  παίρνουν την τιμή 1 μόνο όταν είναι αντίστοιχα 2,3,4 ενώ όταν όλες είναι 0, τότε με βεβαιότητα 100% η  $X_8=1$

ΕΡΜΗΝΕΙΣ :

- i. Οι θετικοί συντελεστές των ανεξάρτητων μεταβλητών αυξάνουν την πιθανότητα εμφάνισης του ενδεχομένου της επιτυχίας  $\{Y=1\}$ , ενώ στον αντίποδα οι αρνητικοί συντελεστές αυτών τη μειώνουν.
- ii. Αν η μεταβλητή  $X_3$  αυξηθεί κατά 1 μονάδα, από την πρόβλεψη

$$\log\left(\frac{\pi_3}{1-\pi_3}\right) = b_0 + b_3 * X_3 + b_4 * X_4 + \dots + b_7 * X_7 + b_8 * X_8 \quad (1)$$

θα μεταβούμε στην πρόβλεψη

$$\log\left(\frac{p_3}{1-p_3}\right) = b_0 + b_3 * X_3 + b_3 + b_4 * X_4 + \dots + b_7 * X_7 + b_8 * X_8 \quad (2)$$

Αφαιρώντας την πρώτη από τη δεύτερη κατά μέλη θα βρούμε ότι

$$b_3 = \log\left(\frac{\pi_3}{1-\pi_3}\right) - \log\left(\frac{p_3}{1-p_3}\right) = \log\left(\frac{p_3 * (1-\pi_3)}{\pi_3 * (1-p_3)}\right)$$

Δηλαδή το  $b_3$  είναι ίσο με τη διαφορά των δύο λογαρίθμων του λόγου των σχετικών συχνοτήτων των δύο ενδεχομένων, οι οποίοι προκύπτουν όταν η ανεξάρτητη μεταβλητή  $X_3$  αυξηθεί κατά 1 μονάδα.

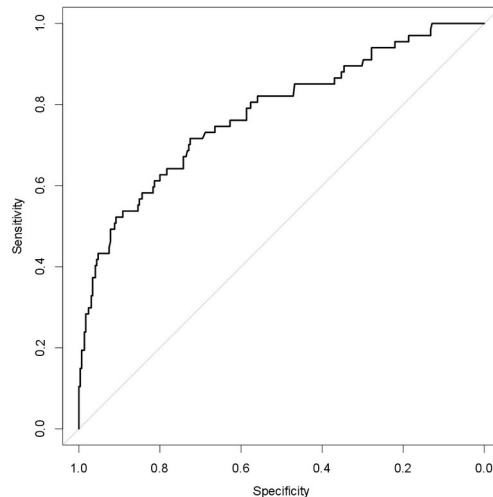
Ομοίως και για τους υπόλοιπους συντελεστές  $b_i$  βρίσκουμε με τι είναι ίσοι.

Επιπλέον όσο οι συντελεστές μεγαλώνουν τόσο σημαντικές γίνονται οι αντίστοιχες ανεξάρτητες μεταβλητές, με την προϋπόθεση ότι το τυπικό σφάλμα της εκτίμησης είναι μικρό.

**AUC:**

Κατά πόσο το μοντέλο έχει καλή προβλεψιμότητα φαίνεται από την καμπύλη ROC που δείχνει το εμβαδόν της περιοχής που βρίσκεται κάτω από την καμπύλη και πάνω από τη διαγώνιο. Όσο

μεγαλύτερο είναι το χωρίο , τόσο καλύτερη είναι η προβλεψιμότητα του. Αυτό φαίνεται στο παρακάτω σχήμα.



Από ότι φαίνεται είναι καλή η προβλεψιμότητα του αφού το εμβαδόν του χωρίου μεταξύ της καμπύλης ROC και της διαγωνίου φαίνεται μεγαλύτερο από το μισό του του εμβαδού του ορθογωνίου τριγώνου.

Πράγματι :

Είναι: Area under the curve: 0.7789

Άρα επειδή η προσεγγιστική τιμή της AUC είναι 0,7789 θα διαπιστώσουμε ότι η προβλεψιμότητα του είναι καλή , έχουμε μία καλή ταξινόμηση των δεδομένων.

### Παράδειγμα

Άν ένα άτομο, ανεξάρτητα από το αν είναι Άνδρας ή Γυναίκα, ζεί μόνος του ή με συντροφιά, αφού αυτές οι μεταβλητές έχουν αφαιρεθεί ως μη σημαντικές είναι 54 χρονών , νοιώθει ότι η υγεία του είναι καλή , δεν έχει περιορισμό στις δραστηριότητες, δεν έχει χρόνιες παθήσεις , έχει κάποια σωματική ανικανότητα και τα οικονομικά του είναι πολύ καλά τότε η εκτιμώμενη πιθανότητα να πάθει κατάθλιψη είναι  $p$  όπου

$$\log\left(\frac{p}{1-p}\right) = -2,557994 - 0,003018 * 54 + 0,283073 * 0 + 0,864497 * 1 \\ + 1,516475 * 0 + 2,086441 * 0 + 0,496601 * 0 \\ + 1,051468 * 1 - 0,646991 * 0 - 1,104316 * 1 - 1,766155 * 0$$

$$\log\left(\frac{p}{1-p}\right) = -1,909342$$

ή

Κάνοντας τις πράξεις παίρνουμε ότι:

$$\frac{p}{1-p} = e^{-1,909342} \Leftrightarrow p = \frac{e^{-1,909342}}{1+e^{-1,909342}} \Leftrightarrow p \approx 0,129055$$

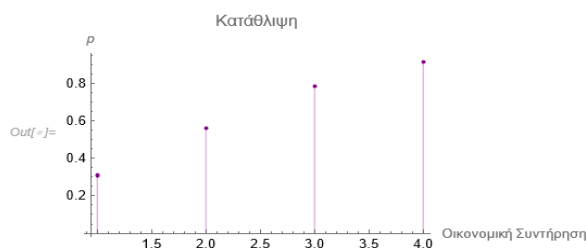
**Σχόλιο:** Είναι πολύ μικρή η πιθανότητα κάποιου ατόμου άνδρα ή γυναίκα να πάθει κατάθλιψη ,με τα παραπάνω δεδομένα.

#### **ΕΠΕΚΤΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ:**

1) Διατηρώντας σταθερές τις τιμές των ανεξάρτητων μεταβλητών  $X_3$  ,  $X_4$  ,  $X_5$  ,  $X_6$  ,  $X_7$  και μεταβάλλοντας την τιμή της ανεξάρτητης μεταβλητής  $X_8$  παίρνουμε 4 πιθανότητες οι οποίες αποτυπώνονται στον παρακάτω πίνακα:

$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$p$
54	3	0	0	1	1	0,308957
54	3	0	0	1	2	0,561307
54	3	0	0	1	3	0,785488
54	3	0	0	1	4	0,9122887

Το αντίστοιχο ραβδόγραμμα του παραπάνω πίνακα της Οικονομικής Συντήρησης και της  $p$  ,πιθανότητας εμφάνισης κατάθλιψης είναι το παρακάτω:

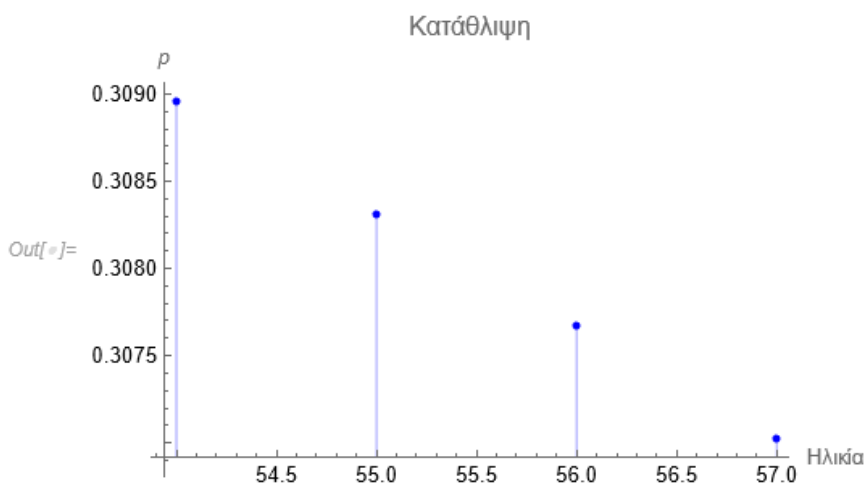


Παρατηρούμε ότι όταν κάποιος Ευρωπαίος έχει αίσθηση υγείας 3(καλή) , είναι 54 ετών , έχει κάποια σωματική ανικανότητα και δεν έχει περιορισμό στις δραστηριότητες και δεν έχει κάποια χρόνια πάθηση ,η πιθανότητα να πάθει κατάθλιψη αυξάνεται όσο τα Οικονομικά του μειώνονται.

2) Διατηρώντας τις ανεξάρτητες μεταβλητές Χ4 , Χ5 , Χ6 , Χ7 , Χ8 σταθερές και αυξάνοντας την ανεξάρτητη μεταβλητή Χ3 της ηλικίας παρατηρούμε ότι η πιθανότητα να πάθει κάποιος κατάθλιψη ,καθώς μεγαλώνει η ηλικία του, φθίνει αλλά με πολύ βραδύ ρυθμό.

Χ3	Χ4	Χ5	Χ6	Χ7	Χ8	ρ
54	3	0	0	1	1	0,308957
55	3	0	0	1	1	0,308313
56	3	0	0	1	1	0,307670
57	3	0	0	1	1	0,307027

Το αντίστοιχο ραβδόγραμμα του παραπάνω πίνακα της ηλικίας και της ρ ,πιθανότητας εμφάνισης κατάθλιψης είναι το παρακάτω:

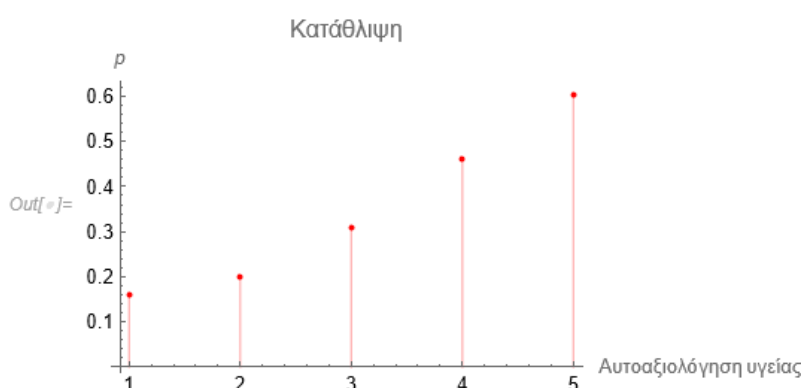


Παρατηρούμε ότι όταν κάποιος Ευρωπαίος έχει αίσθηση καλής υγείας (3), κανέναν περιορισμό στις δραστηριότητες, καμία χρόνια πάθηση , έχει κάποια σωματική ανικανότητα και έχει άριστα οικονομικά, η πιθανότητα να πάθει κατάθλιψη μειώνεται ελάχιστα όσο η ηλικία του αυξάνεται.

3) Διατηρώντας σταθερή τις τιμές της Χ3 (ηλικίας), Χ5=0(δεν έχει περιορισμό στις Δραστηριότητες) ,Χ6=0(0 Πλήθος χρόνων παθήσεων), Χ7=1( Έχει Φυσική Ανικανότητα) , Χ8(Οικονομικά) και αυξάνοντας την Χ4(Αυτοαξιολόγηση Υγείας) παίρνουμε τις πιθανότητες του παραπάνω πίνακα οι οποίες αυξάνονται .

X3	X4	X5	X6	X7	X8	p
54	1	0	0	1	1	0,158491
54	2	0	0	1	1	0,19998
54	3	0	0	1	1	0,308957
54	4	0	0	1	1	0,461819
54	5	0	0	1	1	0,602752

Το αντίστοιχο ραβδόγραμμα του παραπάνω πίνακα της ηλικίας και της  $p$  ,πιθανότητας εμφάνισης κατάθλιψη είναι το παρακάτω:



Παρατηρούμε ότι όταν κάποιος Ευρωπαίος είναι 54 ετών , έχει κάποια σωματική ανικανότητα , οι δραστηριότητες του δεν είναι περιορισμένες και δεν έχει κάποια χρόνια πάθηση , η πιθανότητα να πάθει κατάθλιψη αυξάνεται όσο η X4 αυξάνεται, δηλαδή όσο χειρότερη εκτίμηση έχει για την υγεία του.

4) Διατηρώντας σταθερή τις τιμές της X3 (ηλικίας),X4( Αυτοεκτίμηση της Υγείας του) , X6(Πλήθος χρόνων παθήσεων), X7(Φυσική Ανικανότητα) , X8(Οικονομικά) και αυξάνοντας την X5( Περιορισμός στις Δραστηριότητες) παίρνουμε τις παρακάτω πιθανότητες:

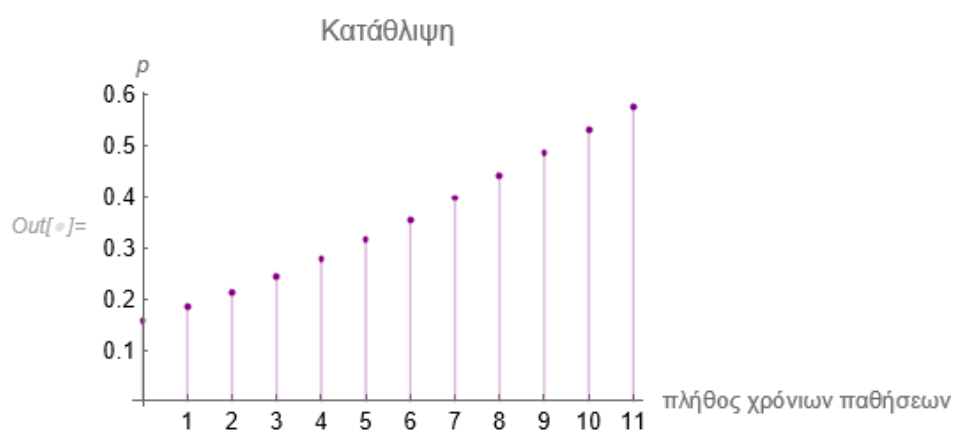
X3	X4	X5	X6	X7	X8	p
54	1	0	0	1	1	0,129057
54	1	1	0	1	1	0,195805

Παρατηρούμε ότι η πιθανότητα να έχει κατάθλιψη, κάποιος που είναι 54 ετών και έχει άριστη εκτίμηση της υγείας του ,καμία χρόνια πάθηση ,με κάποια σωματική ανικανότητα , άριστα οικονομικά και δεν έχει περιορισμό στις δραστηριότητες είναι μικρότερη από αυτόν που έχει περιορισμό στις δραστηριότητες.

5) Διατηρώντας σταθερή τις τιμές της Χ3 (ηλικίας),Χ4( Αυτοεκτίμηση Υγείας) ,Χ5( Περιορισμός στις Δραστηριότητες) ,Χ7(Φυσική Ανικανότητα) , Χ8(Οικονομικά) και αυξάνοντας τις τιμές της Χ6(Πλήθος χρόνιων παθήσεων) παίρνουμε τις παρακάτω πιθανότητες:

Χ3	Χ4	Χ5	Χ6	Χ7	Χ8	ρ
54	1	0	0	1	1	0,158491
54	1	0	1	1	1	0,183917
54	1	0	2	1	1	0,212395
54	1	0	3	1	1	0,243962
54	1	0	4	1	1	0,278561
54	1	0	5	1	1	0,316017
54	1	0	6	1	1	0,356023
54	1	0	7	1	1	0,398147
54	1	0	8	1	1	0,441834
54	1	0	9	1	1	0,486441
54	1	0	10	1	1	0,531265
54	1	0	11	1	1	0,57559

Το αντίστοιχο ραβδόγραμμα του παραπάνω πίνακα του πλήθους των χρόνιων παθήσεων και της ρ ,πιθανότητας εμφάνισης κατάθλιψης είναι το παρακάτω:



Παρατηρούμε ότι καθώς αυξάνεται το πλήθος των χρόνιων παθήσεων η πιθανότητα να πάθει κάποιος κατάθλιψη αυξάνει, με δεδομένο ότι είναι 54 ετών, εκτιμά ότι η υγεία του είναι άριστη, δεν περιορισμό στις δραστηριότητες, έχει άριστη φυσική κατάσταση και τα οικονομικά του είναι άριστα.

**ΕΦΑΡΜΟΓΗ 2:**

Στο επόμενο παράδειγμα θα εξετάσουμε κατά πόσο στην υγεία ενός ατόμου που ζει στις 4 χώρες (Αυστρία, Γερμανία, Δανία, Ελλάδα) παίζουν ρόλο η χώρα διαμονής του, το φύλο του, η κοινωνική θέση του, η ηλικία του, το πλήθος των παιδιών που έχει, το πλήθος των χρόνιων παθήσεων που έχει, το αν καπνίζει ή όχι και το ύψος του εισοδήματος που έχει.

Τα δεδομένα είναι παρμένα από το SHARE.

Οπότε οι μεταβλητές είναι οι εξής:

Εξαρτημένη μεταβλητή Y: η κατάσταση της υγείας που έχει η οποία εκτιμάται σε υγιής ή μη .

Οι τιμές της είναι 1 ή 0

Το 0 σημαίνει ότι είναι υγιής (καλή υγεία) και το 1 (κακή υγεία) σημαίνει ότι δεν είναι υγιής.

Η πιθανότητα  $p$  της Λογιστικής Παλινδρόμησης είναι του ενδεχομένου 1 (κακή υγεία).

Ανεξάρτητες μεταβλητές

➤ **Country** η χώρα προέλευσης του ατόμου. Η **Country** παίρνει τις τιμές 11,12,13,14 που αντιστοιχούν στις χώρες Αυστρία, Γερμανία, Δανία, Ελλάδα.

Αυστρία: 11, Γερμανία :12, Δανία :13, Ελλάδα:14

- **Gender**, το φύλο με τιμές 1:άνδρας, 2: γυναίκα
- **Status**, η κοινωνική θέση του με τιμές: 1:σε γάμο, 2:σε σχέση, 3: μόνος
- **Age** η ηλικία με τιμές ακέραιες θετικές
- **Child** το πλήθος παιδιών
- **Chronic** το πλήθος χρόνιων παθήσεων
- **Smoke**, κάπνισμα: με τιμές 0:όχι καπνίζων και 1:καπνίζων
- **Income**, εισόδημα από εργασία και συντάξεις

Αρχικά θα εισάγουμε τα δεδομένα στην R και θα ελέγξουμε πόσο σημαντικές είναι οι ανεξάρτητες μεταβλητές ως προς την εξαρτημένη. Αν παρατηρηθεί ότι κάποια είναι σχετικά ασήμαντη θα την αφαιρέσουμε από τα δεδομένα για να εκτιμήσουμε την πιθανότητα ένα άτομο να έχει καλή υγεία ή όχι ζώντας στην Αυστρία, στη Γερμανία, στη Δανία ή στην Ελλάδα.

```
> anova(model3)
```

```
Analysis of Deviance Table
```

```
Model: binomial, link: logit
```

```
Response: health
```

```
Terms added sequentially (first to last)
```

	Df	Deviance Resid.	Df	Resid. Dev
NULL			602	658.26
age	1	0.0182	601	658.25
child	1	0.0283	600	658.22
chronic	1	28.3279	599	629.89
income	1	12.6896	598	617.20
status	2	0.4587	596	616.74
smoke	1	1.4171	595	615.32
gender	1	5.4418	594	609.88
country	3	15.1912	591	594.69

Όταν τα στατιστικά  $\chi^2_{1,0.95}=3.84$  και  $\chi^2_{2,0.95}=5.99$  και

$\chi^2_{3,0.95}=7.81$  συγκριθούν με τα DevianceResid. των ανεξάρτητων μεταβλητών με τους ίδιους βαθμούς ελευθερίας είναι μεγαλύτερα από αυτά, τότε οι ανεξάρτητες μεταβλητές age, child, smoke, status είναι στατιστικά μη σημαντικές.

Άρα αν ξανατρέξουμε το μοντέλο με μεταβλητές τις health, chronic, income, gender, και country θα πάρουμε τον ακόλουθο anova

```
> anova(model6)
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: health

Terms added sequentially (first to last)

	Df	Deviance Resid.	Df	Resid. Dev
NULL			602	658.26
chronic	1	26.3434	601	631.92
income	1	8.3255	600	623.60
gender	1	7.3229	599	616.27
country	3	17.6826	596	598.59

```
> summary(model6)
```

Call:

```
glm(formula = health ~ chronic + income + gender + country, family = binomial(link = "logit"))
```



Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6055	-0.7742	-0.5852	-0.2920	2.7590

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.237e+00	2.687e-01	-8.326	< 2e-16 ***
chronic	3.258e-01	6.075e-02	5.364	8.14e-08 ***
income	-3.867e-05	1.617e-05	-2.392	0.01676 *
gender2	5.961e-01	2.110e-01	2.826	0.00472 **
country12	2.633e-01	2.594e-01	1.015	0.31008
country13	-1.223e+00	5.212e-01	-2.346	0.01898 *
country14	5.964e-01	2.506e-01	2.380	0.01732 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

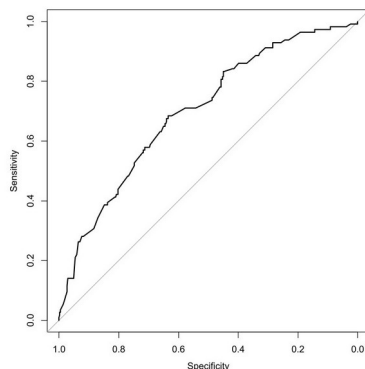
Null deviance: 658.26 on 602 degrees of freedom  
 Residual deviance: 598.59 on 596 degrees of freedom  
 AIC: 612.59

Number of Fisher Scoring iterations: 5

Οπότε το εκτιμώμενο Λογιστικό Μοντέλο είναι:

$$\log\left(\frac{p}{1-p}\right) = -2,237 + 0,3258 * chronic - 3,867 / 10^5 * income + 0,5961 * gender 2 + 0,2633 * country 12 - 1,223 * country 13 + 0,5964 * country 14$$

Area under the curve: 0.6991



Η προβλεψιμότητα του μοντέλου είναι αρκετά καλή.

Οι μεταβλητές country14 ,country13, country12 παίρνουν την τιμή 1 μόνο όταν το άτομο είναι από την Ελλάδα, από την Δανία ,από την Γερμανία ,διαφορετικά παίρνουν την τιμή 0. Άν επιπλέον όλες είναι μηδέν , τότε το άτομο είναι από την Αυστρία.

Επίσης η μεταβλητή gender2 παίρνει την τιμή 1 μόνο όταν το άτομο είναι γυναίκα και παίρνει την τιμή 0 μόνο όταν το άτομο είναι άνδρας.

### Παράδειγμα

Η προσεγγιστική πιθανότητα ενός άνδρα, που μένει στην Ελλάδα , έχοντας εισόδημα 24,000 ευρώ και 1 χρόνια πάθηση ,να είναι η υγεία του καλή είναι  $1-p$  όπου

$$\log\left(\frac{p}{1-p}\right) = -2,237 + 0,3258 * 1 - \frac{3,867}{10^5} * 24000 + 0,5961 * 0 + 2,633 * 0 - 1,223 * 0 + 0,5964 * 1$$

$$\Leftrightarrow \log\left(\frac{p}{1-p}\right) = -2,24288 \Leftrightarrow \frac{p}{1-p} = 2,718^{-2,24288} \Leftrightarrow p = \frac{2,718^{-2,24288}}{1+2,718^{-2,24288}} \Leftrightarrow p = 0,0959654$$

Άρα  $p \approx 0,1$

Είναι σχεδόν βέβαιο ότι η υγεία αυτού του άνδρα είναι καλή, αφού η πιθανότητα του ενδεχομένου της επιτυχίας {1}, αφορά την κακή υγεία και είναι η  $p$  ,όπου  $p \approx 0,1$

### 1η Επέκταση παραδείγματος ως προς τις χώρες διαμονής:

1) Άν ο άνδρας έμενε στην Αυστρία τότε θα ίσχυε:

$$\log\left(\frac{p}{1-p}\right) = -2,237 + 0,3258 * 1 - \frac{3,867}{10^5} * 24000 + 0,5961 * 0 + 2,633 * 0 - 1,223 * 0 + 5,964 * 0$$

$$\Leftrightarrow \log\left(\frac{p}{1-p}\right) = -2,83928 \Leftrightarrow \frac{p}{1-p} = 2,718^{-2,83928} \Leftrightarrow p = \frac{2,718^{-2,83928}}{1+2,718^{-2,83928}} \Leftrightarrow p = 0,0552381$$

Οπότε είναι φανερό ότι στην Αυστρία ο άνδρας με 1 χρόνια πάθηση και με εισόδημα 24000 ευρώ έχει μικρότερη πιθανότητα να έχει κακή υγεία από κάποιον άνδρα της Ελλάδας, δηλαδή μεγαλύτερη πιθανότητα να έχει καλή υγεία.

2) Άν ο άνδρας έμενε στην Γερμανία τότε θα ίσχυε:

$$\log\left(\frac{p}{1-p}\right) = -2,237 + 0,2633 * 1 - \frac{3,867}{10^5} * 24000 + 0,5961 * 0 + 0,3258 * 1$$

$$\Leftrightarrow \log\left(\frac{p}{1-p}\right) = -2,57598 \Leftrightarrow \frac{p}{1-p} = 2,718^{-2,57598} \Leftrightarrow p = \frac{2,718^{-2,57598}}{1+2,718^{-2,57598}} \Leftrightarrow p = 0,0707004$$

3) Αν ο άνδρας έμεινε στην Δανία θα ίσχυε:

$$\log\left(\frac{p}{1-p}\right) = -2,237 + 0,3258 * 1 - \frac{3,867}{10^5} * 24000 + 5,961 * 0 + 0,2633 * 0 - 1,223 * 1 + 0,5964 * 0$$

$$\Leftrightarrow \log\left(\frac{p}{1-p}\right) = -4,06228 \Leftrightarrow \frac{p}{1-p} = 2,718^{-4,06228} \Leftrightarrow p = \frac{2,718^{-4,06228}}{1+2,718^{-4,06228}} \Leftrightarrow p = 0,169186$$

Στον παρακάτω πίνακα φαίνονται συγκεντρωτικά οι πιθανότητες ενός άνδρα με εισόδημα 24000 ευρώ , 1 χρόνια πάθηση να έχει κακή υγεία ζώντας στις χώρες αντίστοιχα Αυστρία , Γερμανία , Δανία και Ελλάδα

#### ΠΙΘΑΝΟΤΗΤΕΣ

Αυστρία	Γερμανία	Δανία	Ελλάδα
0,0552381	0,0707004	0,169186	0,0959654

Οπότε προκύπτει ότι Π.Α.<Π.Γ.<Π.Ε.<Π.Δ.

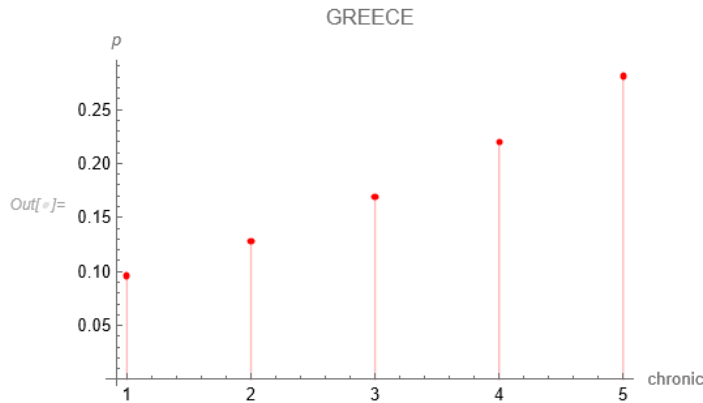
Δηλαδή στην Αυστρία έχει μικρότερη πιθανότητα να έχει κακή υγεία ενώ στην Δανία έχει μεγαλύτερη πιθανότητα να έχει κακή υγεία.

#### 2η Επέκταση παραδείγματος ως προς το πλήθος των χρόνιων παθήσεων:

Διατηρώντας σταθερές τις τιμές των ανεξάρτητων (επεξηγηματικών) μεταβλητών income , country , gender και μεταβάλλοντας τις τιμές της ανεξάρτητης μεταβλητής chronic συμπληρώνουμε τον παρακάτω πίνακα.

Chronic	income	gender	country	p
1	24000	1	Greece	0,0959654
2	24000	1	Greece	0,128188
3	24000	1	Greece	0,169204
4	24000	1	Greece	0,220032
5	24000	1	Greece	0,280965

Το αντίστοιχο ραβδόγραμμα του παραπάνω πίνακα του chronic και της πιθανότητας είναι το επόμενο



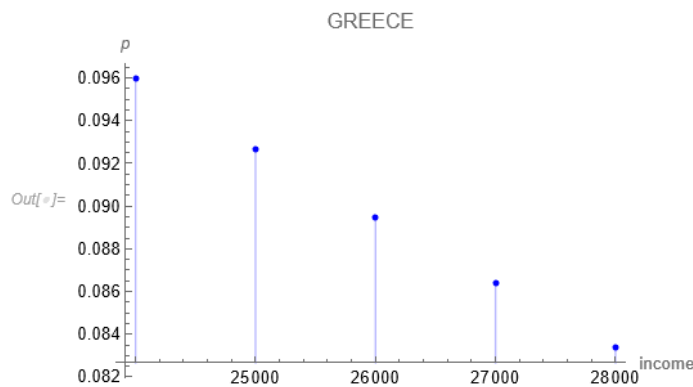
Παρατηρούμε ότι η πιθανότητα κακής υγείας κάποιου άνδρα που ζει στην Ελλάδα αυξάνεται καθώς αυξάνεται το πλήθος των χρόνιων παθήσεων του.

**3η Επέκταση παραδείγματος ως προς το εισόδημα:**

Διατηρώντας σταθερές τις τιμές των ανεξάρτητων (επεξηγηματικών) μεταβλητών, chronic, country, gender και μεταβάλλοντας τις τιμές της ανεξάρτητης μεταβλητής income συμπληρώνουμε τον παρακάτω πίνακα.

Chronic	income	gender	country	p
1	24000	1	Greece	0,0959654
1	25000	1	Greece	0,0926626
1	26000	1	Greece	0,0894621
1	27000	1	Greece	0,0863687
1	28000	1	Greece	0,0833589

Το αντίστοιχο ραβδόγραμμα του παραπάνω πίνακα του income και της πιθανότητας είναι το επόμενο



Παρατηρούμε ότι η πιθανότητα κακής υγείας κάποιου άνδρα μειώνεται καθώς αυξάνεται το εισόδημα του στην Ελλάδα.

### **1839174η Επέκταση παραδείγματος ως προς το φύλο:**

\_Διατηρώντας σταθερές τις τιμές των ανεξάρτητων (επεξηγηματικών) μεταβλητών, chronic, country, income και μεταβάλλοντας τις τιμές της ανεξάρτητης μεταβλητής gender συμπληρώνουμε τον παρακάτω πίνακα.

-  
**α)**

Chronic	income	gender	country	p
1	24000	1	Greece	0,0960168
1	24000	2	Greece	0,161545

Παρατηρούμε ότι στην Ελλάδα ο άνδρας έχει μικρότερη πιθανότητα να έχει κακή υγεία από τη γυναίκα.

**β)**

Chronic	income	gender	country	p
1	24000	1	Austria	0,0552381
1	24000	2	Austria	0,0959394

Παρατηρούμε ότι στην Αυστρία ο άνδρας έχει μικρότερη πιθανότητα να έχει κακή υγεία από τη γυναίκα.

**γ)**

Chronic	income	gender	country	p
1	24000	1	Germany	0,0707007
1	24000	2	Germany	0,121332

Παρατηρούμε ότι στην Γερμανία ο άνδρας έχει μικρότερη πιθανότητα να έχει κακή υγεία από τη γυναίκα.

**δ)**

Chronic	income	gender	country	p
1	24000	1	Denmark	0,165719
1	24000	2	Denmark	0,264992

Παρατηρούμε ότι στην Δανία ο άνδρας έχει μικρότερη πιθανότητα να έχει κακή υγεία από τη γυναίκα.

**ΓΕΝΙΚΟ ΣΥΜΠΕΡΑΣΜΑ:** Και στις 4 χώρες ο άνδρας έχει μεγαλύτερη πιθανότητα να έχει καλή υγεία από τη γυναίκα.

Θα μπορούσαμε επομένως να προβλέψουμε ότι η υγεία του άνδρα στις χώρες της Ευρώπης και του Ισραήλ είναι καλύτερη από την υγεία της γυναίκας.

## ΚΕΦΑΛΑΙΟ 7

### ΔΙΑΦΟΡΕΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΣΥΝΤΕΛΕΣΤΩΝ ΤΗΣ GLM

Σε αυτό το κεφάλαιο κρίνεται σκόπιμο να εξεταστεί πόσο καλές είναι οι προσεγγίσεις των συντελεστών ίδιων δεδομένων με διαφορετικές εντολές Λογιστικής Παλινδρόμησης. Ο λόγος αυτής της εξέτασης απορρέει από το γεγονός ότι όλες οι εκτιμήσεις στη Στατιστική δίνουν τιμές που κυμαίνονται σε διαστήματα εμπιστοσύνης με κάποια πιθανότητα σφάλματος να βρίσκονται έξω από αυτά.

**1η προσέγγιση:** Στην εντολή `glm` ΔΕΝ αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με `cut-off` (διαχωρισμό) , `c(0.8,0.2)`), και έχει συνάρτηση σύνδεσης την `logit`.

```
> Y<-as.numeric(my$Y)
> X3<-as.numeric(my$X3)
> X4<-as.factor(my$X4)
> X5<-as.factor(my$X5)
> X6<-as.numeric(my$X6)
> X7<-as.factor(my$X7)
> X8<-as.factor(my$X8)
> set.seed(1234)
> ind<-sample(2,nrow(my),replace=T,prob=c(0.8,0.2))
> trainData<-my[ind==1,]
> testData<-my[ind==2,]
> model1=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="logit"))
> anova(model1)
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1879	1994.8
X3	1	9.299	1878	1985.5
X4	4	300.062	1874	1685.4
X5	1	22.936	1873	1662.5
X6	1	16.040	1872	1646.4
X7	1	19.207	1871	1627.2
X8	3	38.758	1868	1588.5

```
> summary(model1)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0911	-0.6265	-0.4251	-0.2361	2.6132

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.557994	1.167811	-2.190	0.028494 *
X3	-0.003018	0.007721	-0.391	0.695894
X42	0.283073	1.056291	0.268	0.788709
X43	0.864497	1.036797	0.834	0.404385
X44	1.516475	1.043650	1.453	0.146210
X45	2.086441	1.063444	1.962	0.049766 *
X51	0.496601	0.154924	3.205	0.001348 **
X6	0.179472	0.050701	3.540	0.000400 ***
X71	1.051468	0.237631	4.425	9.65e-06 ***
X82	-0.646991	0.134918	-4.795	1.62e-06 ***
X83	-1.104316	0.298796	-3.696	0.000219 ***
X84	-1.766155	0.656192	-2.692	0.007113 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1994.8 on 1879 degrees of freedom  
 Residual deviance: 1588.5 on 1868 degrees of freedom  
 AIC: 1612.5

Number of Fisher Scoring iterations: 5

**2η προσέγγιση:** Στην εντολή `glm` αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με `cut-off c(0.8,0.2)`, και έχει συνάρτηση σύνδεσης τη `logit`.

```
model2=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="logit"),data=trainData)
> anova(model2)
Analysis of Deviance Table
```

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1517	1644.1
X3	1	9.574	1516	1634.5
X4	1	246.976	1515	1387.6
X5	1	22.266	1514	1365.3
X6	1	11.299	1513	1354.0



50

```
X7 1 11.754 1512 1342.3  
X8 1 37.608 1511 1304.7
```

```
> summary(model2)
```

```
Call:
```

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"),  
data = trainData)
```

```
Deviance Residuals:
```

```
Min 1Q Median 3Q Max  
-2.0963 -0.6489 -0.4310 -0.2141 2.5487
```

```
Coefficients:
```

```
Estimate Std. Error z value Pr(>|z|)  
(Intercept) -3.087610 0.734411 -4.204 2.62e-05 ***  
X3 0.002604 0.008445 0.308 0.757823  
X4 0.587715 0.116883 5.028 4.95e-07 ***  
X5 0.568888 0.168817 3.370 0.000752 ***  
X6 0.171246 0.056404 3.036 0.002397 **  
X7 0.906526 0.256883 3.529 0.000417 ***  
X8 -0.674798 0.115810 -5.827 5.65e-09 ***
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 1644.1 on 1517 degrees of freedom  
Residual deviance: 1304.6 on 1511 degrees of freedom  
AIC: 1318.6
```

```
Number of Fisher Scoring iterations: 5
```

Παρατηρούμε ό,τι το AIC της 1ης προσέγγισης είναι 1612.5 ενώ της 2ης προσέγγισης είναι μικρότερο 1318.5. Προφανώς η 2η προσέγγιση είναι καλύτερη

**3η προσέγγιση:** Στην εντολή `glm` ΔΕΝ αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με `cut-off c(0.7,0.3)`, και έχει συνάρτηση σύνδεσης τη `logit`.

```
> my=read.csv2("data3.csv",header=T)  
> Y<-as.numeric(my$Y)  
> X3<-as.numeric(my$X3)  
> X4<-as.factor(my$X4)  
> X5<-as.factor(my$X5)  
> X6<-as.numeric(my$X6)  
> X7<-as.factor(my$X7)  
> X8<-as.factor(my$X8)  
> set.seed(1234)  
> ind<-sample(2,nrow(my),replace=T,prob=c(0.7,0.3))
```

```
> trainData<-my[ind==1,]
> testData<-my[ind==2,]
> model3=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="logit"))
> anova(model3)
Analysis of Deviance Table
```

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1879	1994.8
X3	1	9.299	1878	1985.5
X4	4	300.062	1874	1685.4
X5	1	22.936	1873	1662.5
X6	1	16.040	1872	1646.4
X7	1	19.207	1871	1627.2
X8	3	38.758	1868	1588.5

```
> summary(model3)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0911	-0.6265	-0.4251	-0.2361	2.6132

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.557994	1.167811	-2.190	0.028494 *
X3	-0.003018	0.007721	-0.391	0.695894
X42	0.283073	1.056291	0.268	0.788709
X43	0.864497	1.036797	0.834	0.404385
X44	1.516475	1.043650	1.453	0.146210
X45	2.086441	1.063444	1.962	0.049766 *
X51	0.496601	0.154924	3.205	0.001348 **
X6	0.179472	0.050701	3.540	0.000400 ***
X71	1.051468	0.237631	4.425	9.65e-06 ***
X82	-0.646991	0.134918	-4.795	1.62e-06 ***
X83	-1.104316	0.298796	-3.696	0.000219 ***
X84	-1.766155	0.656192	-2.692	0.007113 **

---

Signif. Codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1994.8 on 1879 degrees of freedom  
Residual deviance: 1588.5 on 1868 degrees of freedom  
AIC: 1612.5

Number of Fisher Scoring iterations: 5

**Παρατηρούμε ό,τι:** οι προσεγγίσεις 1 και 3 είναι αρκετά όμοιες με το ίδιο AIC και ίδιους συντελεστές.

**4η προσέγγιση:** Στην εντολή `glm` αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με cut-off  $c(0.7,0.3)$ , και έχει συνάρτηση σύνδεσης τη `logit`.

```
model4=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="logit"),data=trainData)
> anova(model4)
Analysis of Deviance Table
```

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1344	1440.1
X3	1	9.584	1343	1430.5
X4	1	226.103	1342	1204.4
X5	1	19.133	1341	1185.2
X6	1	10.352	1340	1174.9
X7	1	16.594	1339	1158.3
X8	1	30.691	1338	1127.6

```
> summary(model4)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"),
    data = trainData)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1807	-0.6268	-0.4209	-0.2062	2.5936

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.931080	0.798257	-3.672	0.000241 ***
X3	-0.002528	0.009227	-0.274	0.784120
X4	0.626330	0.125479	4.992	5.99e-07 ***

53

```
X5      0.556386  0.181031  3.073 0.002116 **
X6      0.174756  0.060648  2.881 0.003958 **
X7      1.187632  0.290128  4.093 4.25e-05 ***
X8     -0.665104  0.125963  -5.280 1.29e-07 ***
```

---

Signif. Codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1440.1 on 1344 degrees of freedom  
Residual deviance: 1127.6 on 1338 degrees of freedom  
AIC: 1141.6  
Number of Fisher Scoring iterations: 5

**Παρατηρούμε ό,τι:** η 4η προσέγγιση έχει μικρότερο AIC από όλες, άρα είναι καλύτερη από όλες.

**ΓΕΝΙΚΑ:** Οι προσεγγίσεις 2 και 4 που αναφέρουν τον πίνακα δεδομένων της glm θεωρούν τις ανεξάρτητες μεταβλητές συνεχείς. Στο γεγονός αυτό οφείλετε ό,τι το 1 είναι ο βαθμός ελευθερίας όλων ανεξάρτητων μεταβλητών. Και με αυτή τη θεώρηση παίρνουμε το μικρότερο AIC το οποίο δηλώνει την καλύτερη προσέγγιση.

**5η προσέγγιση:** Στην εντολή `glm ΔΕΝ` αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με `cut-off`, και έχει συνάρτηση σύνδεσης τη `probit`.

```
> model5=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="probit"))
> anova(model5)
Analysis of Deviance Table
Model: binomial, link: probit
Response: Y
Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1879	1994.8
X3 1	9.089		1878	1985.7
X4 4	300.213		1874	1685.5
X5 1	22.816		1873	1662.6
X6 1	16.844		1872	1645.8
X7 1	19.279		1871	1626.5
X8 3	38.536		1868	1588.0

```
> summary(model5)
```

Call:

54

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "probit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0950	-0.6311	-0.4168	-0.2173	2.6454

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.455133	0.568291	-2.561	0.010451 *
X3	-0.001163	0.004377	-0.266	0.790517
X42	0.125960	0.487033	0.259	0.795923
X43	0.408082	0.478583	0.853	0.393833
X44	0.776972	0.484285	1.604	0.108633
X45	1.130900	0.499944	2.262	0.023694 *
X51	0.298167	0.091461	3.260	0.001114 **
X6	0.103559	0.029129	3.555	0.000378 ***
X71	0.622496	0.141012	4.414	1.01e-05 ***
X82	-0.364300	0.076302	-4.774	1.80e-06 ***
X83	-0.582045	0.152156	-3.825	0.000131 ***
X84	-1.066709	0.353775	-3.015	0.002568 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1994.8 on 1879 degrees of freedom  
Residual deviance: 1588.0 on 1868 degrees of freedom  
AIC: 1612  
Number of Fisher Scoring iterations: 6

**6η προσέγγιση:** Στην εντολή `glm` αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με `cut-off c(0.8,0.2)`, και έχει συνάρτηση σύνδεσης τη `probit`.

```
model6=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="probit"),data=trainData)
```

```
> anova(model6)
```

Analysis of Deviance Table

Model: binomial, link: probit

Response: Y

Terms added sequentially (first to last)

55

```
      Df  Deviance Resid.  Df Resid. Dev
NULL                                1517  1644.1
X3  1      9.370          1516  1634.8
X4  1    244.920          1515  1389.8
X5  1    23.517          1514  1366.3
X6  1    12.098          1513  1354.2
X7  1    12.591          1512  1341.6
X8  1    37.947          1511  1303.7
```

```
> summary(model6)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "probit"),
     data = trainData)
```

Deviance Residuals:

```
  Min    1Q  Median    3Q   Max
-2.0876 -0.6620 -0.4279 -0.1754  2.6320
```

Coefficients:

```
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.813597  0.416345 -4.356 1.32e-05 ***
X3           0.002057  0.004822  0.427 0.669608
X4           0.325794  0.066305  4.914 8.94e-07 ***
X5           0.346268  0.099706  3.473 0.000515 ***
X6           0.099688  0.032549  3.063 0.002194 **
X7           0.548297  0.152229  3.602 0.000316 ***
X8          -0.378914  0.063605 -5.957 2.56e-09 ***
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1644.1 on 1517 degrees of freedom
Residual deviance: 1303.7 on 1511 degrees of freedom
AIC: 1317.7
```

Number of Fisher Scoring iterations: 5

**7η προσέγγιση:** Στην εντολή `glm` αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με `cut-off c(0.7,0.3)`, και έχει συνάρτηση σύνδεσης τη `probit`.

```
model7=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="probit"),data=trainData)
> anova(model7)
```

## Analysis of Deviance Table

Model: binomial, link: probit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid.	Dev
NULL				1344		1440.1
X3	1	9.369		1343		1430.7
X4	1	222.297		1342		1208.4
X5	1	20.190		1341		1188.2
X6	1	11.623		1340		1176.6
X7	1	17.416		1339		1159.2
X8	1	31.848		1338		1127.3

&gt; summary(model7)

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "probit"),
     data = trainData)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1792	-0.6387	-0.4164	-0.1664	2.6368

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.7002141	0.4496957	-3.781	0.000156 ***
X3	-0.0006578	0.0052370	-0.126	0.900043
X4	0.3376820	0.0707722	4.771	1.83e-06 ***
X5	0.3357734	0.1065482	3.151	0.001625 **
X6	0.1044742	0.0348778	2.995	0.002741 **
X7	0.7034642	0.1698092	4.143	3.43e-05 ***
X8	-0.3787331	0.0691018	-5.481	4.23e-08 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1440.1 on 1344 degrees of freedom  
 Residual deviance: 1127.3 on 1338 degrees of freedom  
 AIC: 1141.3

Number of Fisher Scoring iterations: 5

8η προσέγγιση: Στην εντολή `glm` ΔΕΝ αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει με cut-off  $c(0.7, 0.3)$ , και έχει συνάρτηση σύνδεσης τη `probit`.

```
model8=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="probit"))
```

```
> anova(model8)
```

Analysis of Deviance Table

Model: binomial, link: probit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1879	1994.8
X3	1	9.089	1878	1985.7
X4	4	300.213	1874	1685.5
X5	1	22.816	1873	1662.6
X6	1	16.844	1872	1645.8
X7	1	19.279	1871	1626.5
X8	3	38.536	1868	1588.0

```
> summary(model8)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "probit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0950	-0.6311	-0.4168	-0.2173	2.6454

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.455133	0.568291	-2.561	0.010451 *
X3	-0.001163	0.004377	-0.266	0.790517
X42	0.125960	0.487033	0.259	0.795923
X43	0.408082	0.478583	0.853	0.393833
X44	0.776972	0.484285	1.604	0.108633
X45	1.130900	0.499944	2.262	0.023694 *
X51	0.298167	0.091461	3.260	0.001114 **
X6	0.103559	0.029129	3.555	0.000378 ***
X71	0.622496	0.141012	4.414	1.01e-05 ***
X82	-0.364300	0.076302	-4.774	1.80e-06 ***
X83	-0.582045	0.152156	-3.825	0.000131 ***
X84	-1.066709	0.353775	-3.015	0.002568 **



58

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Dispersion parameter for binomial family taken to be 1)  
Null deviance: 1994.8 on 1879 degrees of freedom  
Residual deviance: 1588.0 on 1868 degrees of freedom  
AIC: 1612  
Number of Fisher Scoring iterations: 6

9η προσέγγιση: Στην εντολή `glm` ΔΕΝ αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει, και έχει συνάρτηση σύνδεσης τη συμπληρωματική της λογιστικής `log-log`.

```
model9=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="cloglog"))
> anova(model9)
Analysis of Deviance Table
Model: binomial, link: cloglog
Response: Y
Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1879	1994.8
X3	1	9.529	1878	1985.2
X4	4	300.119	1874	1685.1
X5	1	22.956	1873	1662.2
X6	1	15.514	1872	1646.6
X7	1	16.140	1871	1630.5
X8	3	36.269	1868	1594.2

```
> summary(model9)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "cloglog"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2957	-0.6235	-0.4442	-0.2440	2.6074

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.7575711	1.0891450	-2.532	0.011346 *
X3	-0.0009558	0.0060019	-0.159	0.873469
X42	0.2606370	1.0257121	0.254	0.799416
X43	0.8841673	1.0058417	0.879	0.379384
X44	1.4958573	1.0097557	1.481	0.138499

```

X45    1.8281384  1.0199825  1.792 0.073081 .
X51    0.4247837  0.1270241  3.344 0.000825 ***
X6     0.1317469  0.0373057  3.532 0.000413 ***
X71    0.6354946  0.1546235  4.110 3.96e-05 ***
X82   -0.5059039  0.1103732 -4.584 4.57e-06 ***
X83   -0.9597797  0.2723381 -3.524 0.000425 ***
X84   -1.3434644  0.5543784 -2.423 0.015377 *

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1994.8 on 1879 degrees of freedom  
Residual deviance: 1594.2 on 1868 degrees of freedom  
AIC: 1618.2  
Number of Fisher Scoring iterations: 6

10η προσέγγιση: Στην εντολή `glm` αναφέρεται ο πίνακας των δεδομένων που αυτή διαβάζει, και έχει συνάρτηση σύνδεσης τη συμπληρωματική της λογιστικής `log-log`.

```

> set.seed(1234)
> ind<-sample(2,nrow(my),replace=T,prob=c(0.8,0.2))
> trainData<-my[ind==1,]
> testData<-my[ind==2,]
> model10=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="cloglog"),data=trainData)
> anova(model10)

```

Analysis of Deviance Table

Model: binomial, link: cloglog

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1517	1644.1
X3	1	9.820	1516	1634.3
X4	1	246.113	1515	1388.2
X5	1	21.996	1514	1366.2
X6	1	11.126	1513	1355.1
X7	1	7.292	1512	1347.8
X8	1	36.048	1511	1311.7

```
> summary(model10)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "cloglog"),
    data = trainData)
```

Deviance Residuals:

```
   Min      1Q   Median      3Q      Max
-2.3569 -0.6441 -0.4512 -0.2518  2.4613
```

Coefficients:

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.038362  0.580631  -5.233 1.67e-07 ***
X3           0.003790  0.006513   0.582 0.560575
X4           0.492772  0.092386   5.334 9.62e-08 ***
X5           0.497934  0.138934   3.584 0.000338 ***
X6           0.119677  0.041118   2.911 0.003608 **
X7           0.476478  0.167374   2.847 0.004416 **
X8          -0.549659  0.096733  -5.682 1.33e-08 ***
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1644.1 on 1517 degrees of freedom  
Residual deviance: 1311.7 on 1511 degrees of freedom  
AIC: 1325.7

Number of Fisher Scoring iterations: 6

**Παρατήρηση:** Μεταξύ όλων των μοντέλων καλύτερη εκτίμηση(προσαρμογή) των συντελεστών γίνεται από αυτό που έχει το μικρότερο AIC. (Πολίτης Κ.)

Παρακάτω παραθέτουμε πίνακα των AIC από κάθε μοντέλο.

model1	model2	model3	model4	model5	model6	model7	model8	model9	model10
1612,5	1318,6	1612,5	1141,6	1612	1317,7	1141,3	1612	1618,2	1325,7

Άρα το μοντέλο 7 εκτιμά καλύτερα τους συντελεστές με μικρή διαφορά από το μοντέλο 4.

## ΚΕΦΑΛΑΙΟ 8 ΣΥΜΠΕΡΑΣΜΑΤΑ

1. Η Δίτιμη Λογιστική Παλινδρόμηση είναι ένα θεωρητικά έγκυρο μοντέλο ανάλυσης Εφαρμογών με αποτέλεσμα δύο τιμές , εξαρτώμενο από πολλές ανεξάρτητες μεταβλητές (ή αλλιώς επεξηγηματικές) οι οποίες μπορεί να είναι συνεχείς ή διακριτές ή κατηγορικές.
2. Η χρησιμότητά της οφείλεται στο γεγονός ότι είναι επέκταση του μοντέλου της Γραμμικής Παλινδρόμησης επειδή εφαρμόζεται και εκεί που παραβιάζεται η ομοσκεδαστικότητα , η συνέχεια και η κανονικότητα της εξαρτημένης μεταβλητής.
3. Η σύλληψη της ιδέας της ανάλυσης Δίτιμων Εφαρμογών , εξαρτώμενων από ανεξάρτητες (επεξηγηματικές) μεταβλητές κάθε είδους (συνεχείς, διακριτές, κατηγορικές) με το μοντέλο της Λογιστική Παλινδρόμησης , βασίζεται στη Σιγμοειδή Συνάρτηση με σύνολο τιμών (0,1) της οποίας η αντίστροφη είναι η Λογιστική Συνάρτηση με πεδίο ορισμού το (0,1).
4. Με τη βοήθεια της Γλώσσας Προγραμματισμού R , εκτελέστηκε η Λογιστική Παλινδρόμηση σε 2 εφαρμογές με πραγματικά δεδομένα , παρμένα από την Ευρωπαϊκή Πηγή Δεδομένων SHARE , όπου η προσαρμογή των αποτελεσμάτων κρίνεται ικανοποιητική.
5. Τα πραγματικά δεδομένα της SHARE είναι παγκοσμίως αποδεκτά και συντελούν στην εξαγωγή πολλών συμπερασμάτων που αφορούν ηλικιωμένα άτομα άνω των 50 ετών που ζουν στην Ευρώπη και το Ισραήλ.
6. Τα παραδείγματα των 2 εφαρμογών που χρησιμοποιήθηκαν μπορούν να επεκταθούν πολύ περισσότερο , ώστε ο κάθε αναγνώστης αυτής της Διπλωματικής , να βγάλει χρήσιμα και έγκυρα και πολλά συμπεράσματα. Βάζοντας τις δικές του τιμές στις επεξηγηματικές μεταβλητές μπορεί να εκτιμήσει τις επιθυμητές πιθανότητες που αναζητά. Ακόμη θα μπορούσε να παρακολουθήσει την εξέλιξη της πιθανότητας ενός ενδεχομένου , σταθεροποιώντας τις κ-1 τιμές των ανεξάρτητων μεταβλητών και τρέχοντας την τιμή της κ κατά σειρά προτίμησης ανεξάρτητης μεταβλητής με διάφορες τιμές που μπορεί να πάρει αυτή.
7. Όσο περισσότερα δεδομένα χρησιμοποιούνται σε κάθε εφαρμογή , τόσο πιο γρήγορα παγιώνονται οι εκτιμήσεις των συντελεστών των Ανεξάρτητων Μεταβλητών της Λογιστικής Παλινδρόμησης.
8. Η Γλώσσα Προγραμματισμού της R έχει τη δυνατότητα να πολλαπλασιάζει τις κατηγορικές ανεξάρτητες μεταβλητές και να δίνει σε κάθε μία έναν συντελεστή που πολλαπλασιάζεται με 1 αν συμβαίνει ή με 0 αν δεν συμβαίνει.
9. Όλα τα αποτελέσματα στη στατιστική δεν είναι απόλυτα , αλλά προσεγγίζουν τα δεδομένα με κάποιο σφάλμα. Υπάρχουν κεντρικές(σημειακές) τιμές που κυμαίνονται σε διαστήματα εμπιστοσύνης.  
π.χ. ο b3 ανήκει στο διάστημα  $(-0.003018-1.96*0.007721, -0.003018+1.96*0.007721)=$   
 $(-0.0181512, 0.0121152)$

Εντολές της R για τη λειτουργία της Λογιστικής Παλινδρόμησης

1. `install.packages("pROC")`
2. `library(pROC)`
3. File→Change Directory→Folder.....
4. `mydata=read.csv2("data.csv",header=T)`
5. `set.seed(1234)`
6. `ind←sample(2,nrow(mydata),replace=T, prob=c(0.8,0.2))`
7. `trainData←mydata[ind==1,]`
8. `testData←mydata[ind==2,]`
9. `model=glm(Y~X1+X2+X3+.....+X10 , family=binomial(link="logit"),data=trainData)`
10. `anova(model)`
11. `summary(model)`
12. `pred_Y=predict(model, trainData, type="response")`
13. `roc1=roc(response=trainData$Y,predictor=pred_Y)`
14. `jpeg("rplot.jpg")` (προαιρετικό)
15. `pdf("rplot.pdf")`
16. `plot.roc(roc1)`
17. `dev.off()`
18. `auc(roc1)`

**Παρατηρήσεις:**

1) Η εντολή `auc(roc1)` ενεργοποιείται και στην Rstudio. Εκεί όμως οι εντολές 3 , 4 αντικαθίστανται από την `File → ImportDataSet → FromText(base) → μεταβαίνουμε στην επιφάνεια εργασίας και βρίσκουμε το αρχείο (όχι φάκελο) το οποίο έχει επέκταση csv`

2) Επιπρόσθετα αν δε δουλεύουμε στην Rstudio και δουλεύουμε στην R.4.2.1 ή στην R.4.2.0 θα πρέπει να μετατρέψουμε τις ανεξάρτητες μεταβλητές σε αριθμητικές, όπως για παράδειγμα

```
health←as.numeric(mydata$health) ή σε κατηγορικές , όπως για παράδειγμα
gender←as.factor(mydata$gender)
```

3) Ο θεωρητικά σωστός πίνακας απονα με βαθμό ελευθερίας 1 σε κάθε συνεχή μεταβλητή και k-1 σε κάθε διακριτή και κατηγορική μεταβλητή με k διαφορετικές τιμές δίνεται μόνο όταν η εντολή 9 γίνεται `model=glm(Y~X1+X2+X3+.....+X10 , family=binomial(link="logit"))` αφού προηγουμένως δηλώσουμε τις μεταβλητές ως κατηγορικές ή αριθμητικές.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

### Ελληνική

1. Φωκιανός Σημειώσεις για την R. Πανεπιστήμιο Κύπρου.
2. Αντζουλάκος, Δ.(2013) Ανάλυση δεδομένων με τη Χρήση Στατιστικών Πακέτων- Εισαγωγή στην R. Πανεπιστήμιο Πειραιώς.
3. Ηλιόπουλος Γ.(2017) ,Γενικευμένα Γραμμικά Μοντέλα , πανεπιστημιακές σημειώσεις. Πανεπιστήμιο Πειραιώς.
4. Πολίτης Κ.(2018), Βιοστατιστική και Στατιστικές Μέθοδοι στην Επιδημιολογία , πανεπιστημιακές σημειώσεις. Πανεπιστήμιο Πειραιώς.
5. Πουλοπούλου Αλεξάνδρα. Λογιστική Παλινδρόμηση Διπλωματική στο ΕΜΠ.

### Ξένα

6. Agresti A.(2007), An Introduction to Categorical Data Analysis ,Wiley , New York
7. Cox. D.R. and Snell, E.J.,(1989).Analysis of binary data (2<sup>nd</sup> ed.) Analysis of Binary Data (2<sup>nd</sup> ed.) Chapman and Hall
8. Efron , B. ( 1978) , Regression and ANOVA with zero-one data: measures of residual variation , Joynral of the American statistical Association 73 ,113-121
9. Hosmer David W., Jr., Lemeshow S. (2001), Applied Logistic Regression
10. McFadden, D. (1974) , Conditional logit analysis of qualitative choice behaviour , in: P.Zarembeka (ed.) , Frontiers in Econometrics , Academic Press , New York, 105-142
11. Youden, W.J. (1950) , Index for rating diagnostic tests-Cancer

### Διαδυκτιακές Ιστοσελίδες

- [https://en.wikipedia.org/wiki/Main\\_Page](https://en.wikipedia.org/wiki/Main_Page)
- [https://www.r\\_bloggers.com/example-8-15-firth-logistic-regression/](https://www.r_bloggers.com/example-8-15-firth-logistic-regression/)
- <https://www.quara.com/>

**Κάποια βιβλιογραφία ικανεί να διαφωτίσει οποιονδήποτε επιθυμεί να ασχοληθεί με τη γλώσσα προγραμματισμού R είναι η παρακάτω:**

1. Richard A. Becker, John M. Chambers, and Allan R. Wilks. The New S
2. John M. Chambers and Trevor J. Hastie. Statistical Models in S. Chapman & Hall, London, 1992.
3. William N. Venables and Brian D. Ripley. Modern Applied Statistics with S. Fourth Edition. Springer, New York, 2002.
4. William N. Venables and Brian D. Ripley. S Programming. Springer, New York, 2000.
5. Frank E. Harrell. Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression.
6. John Fox. An R and S-Plus Companion to Applied Regression. Sage Publications, Thousand Oaks, CA, USA, 2002.

ΠΑΡΑΡΤΗΜΑ ΕΝΤΟΛΩΝ ΕΦΑΡΜΟΣΜΕΝΩΝ ΣΤΗΝ R  
ΓΙΑ ΤΗ ΛΗΨΗ ΤΟΥ ΠΙΝΑΚΑ (ΑΝΟΝΑ) ΚΑΙ ΤΟΥ ΠΙΝΑΚΑ  
ΤΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΤΩΝ ΑΝΕΞΑΡΤΗΤΩΝ ΜΕΤΑΒΛΗΤΩΝ (SUMMARY)

```
> summary(model1)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"))
```

Deviance Residuals:

```
  Min    1Q  Median    3Q   Max
-1.5970 -0.6842 -0.6371 -0.5281  2.0834
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.254471	0.689537	-1.819	0.0689 .
X3	0.002192	0.007084	0.309	0.7570
X42	-0.688585	0.487132	-1.414	0.1575
X43	-0.383052	0.474917	-0.807	0.4199
X44	-0.295662	0.495075	-0.597	0.5504
X45	-0.442704	0.544857	-0.813	0.4165
X51	0.087844	0.161274	0.545	0.5860
X6	0.048985	0.047988	1.021	0.3074
X71	1.891622	0.229875	8.229	<2e-16 ***
X82	-0.085661	0.126976	-0.675	0.4999
X83	-0.131577	0.217341	-0.605	0.5449
X84	-0.248444	0.397841	-0.624	0.5323

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1994.2 on 1878 degrees of freedom  
Residual deviance: 1860.1 on 1867 degrees of freedom  
AIC: 1884.1

Number of Fisher Scoring iterations: 4



Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

```
> my=read.csv2("d.csv",header=T)
```

```
> my
```

```
  X3 X4 X5 X6 X7 X8 Y
1  65 3 0 2 0 1 0
2  74 4 1 3 0 3 0
3  74 2 0 1 0 3 0
4  74 3 0 1 0 2 0
5  66 4 0 3 0 1 0
6  75 4 1 6 0 1 0
7  86 3 0 2 0 3 0
8  69 2 0 1 0 3 0
9  65 4 0 1 0 2 0
10 71 2 0 0 0 2 0
11 75 2 0 1 0 2 0
12 65 5 1 5 0 1 0
13 61 3 0 2 0 1 0
14 79 3 0 1 0 1 0
15 71 4 0 4 0 3 0
16 67 2 0 1 0 3 0
17 70 2 0 1 0 3 0
18 64 3 0 1 0 1 0
19 73 3 0 2 0 2 0
20 73 2 0 2 0 2 1
21 79 3 0 2 0 2 1
22 64 3 0 2 0 1 0
23 92 3 0 0 0 2 0
24 68 4 0 2 0 2 1
25 57 3 1 4 0 2 0
26 62 3 1 3 0 2 0
```

27 66 4 0 2 0 3 0  
28 82 3 0 1 0 3 0  
29 68 3 0 1 0 1 1  
30 72 4 1 4 0 1 0  
31 79 3 0 2 0 1 0  
32 65 3 0 1 0 2 0  
33 61 3 0 3 0 1 0  
34 61 2 0 0 0 1 0  
35 85 3 1 2 0 1 0  
36 75 4 1 2 0 1 0  
37 81 2 0 1 0 2 0  
38 71 3 0 1 0 2 1  
39 63 5 1 4 0 2 1  
40 49 5 0 1 0 2 0  
41 77 3 0 2 0 1 0  
42 89 2 0 0 0 1 0  
43 67 3 0 1 0 2 0  
44 62 3 0 1 0 3 0  
45 69 3 0 2 0 3 0  
46 63 5 1 4 0 2 0  
47 56 2 0 0 0 2 0  
48 64 2 0 1 0 2 0  
49 76 3 0 4 0 2 1  
50 61 4 0 4 0 1 0  
51 70 3 1 5 0 1 1  
52 81 5 1 5 0 1 0  
53 72 4 0 3 0 2 0  
54 83 3 0 3 0 1 0  
55 73 3 1 2 0 1 1  
56 80 5 1 3 0 1 1  
57 78 3 0 2 0 1 0  
58 74 4 0 4 0 2 0  
59 76 3 0 1 0 2 0  
60 70 4 0 0 0 1 0  
61 67 2 0 1 0 3 0  
62 59 3 0 1 0 1 0  
63 77 3 0 4 0 2 0  
64 76 3 0 0 0 1 0  
65 74 3 0 1 0 1 0  
66 70 3 1 2 0 1 0  
67 71 3 0 1 0 2 0  
68 71 2 0 0 0 2 0  
69 70 3 0 2 0 2 0

70 77 3 0 2 0 3 1  
71 66 4 0 3 0 2 0  
72 66 3 0 3 0 3 0  
73 78 3 0 1 0 3 0  
74 66 3 0 1 0 2 0  
75 66 5 1 3 0 2 0  
76 71 2 0 1 0 3 0  
77 71 4 0 2 0 1 0  
78 63 3 0 0 0 1 0  
79 59 4 1 3 0 1 0  
80 68 4 0 2 0 1 0  
81 77 4 0 6 0 2 0  
82 64 3 0 1 0 1 0  
83 65 4 1 7 0 2 0  
84 91 3 0 3 0 1 0  
85 75 3 0 3 0 1 0  
86 85 3 0 0 0 2 1  
87 71 5 1 7 0 1 1  
88 65 5 1 6 0 1 0  
89 62 2 0 0 0 3 0  
90 73 2 0 0 0 3 0  
91 66 2 0 0 0 3 1  
92 74 5 1 5 0 1 0  
93 74 5 0 2 0 2 1  
94 60 5 1 3 0 2 0  
95 69 4 0 4 0 1 1  
96 64 4 0 5 0 1 0  
97 64 4 0 1 0 2 1  
98 61 4 0 3 0 2 0  
99 66 3 0 3 0 2 0  
100 65 2 0 1 0 3 0  
101 82 3 1 1 0 1 0  
102 62 3 0 1 0 2 0  
103 59 2 0 0 0 2 0  
104 65 3 0 0 0 3 0  
105 81 3 0 1 0 3 0  
106 67 4 0 2 0 1 0  
107 70 3 0 2 0 1 0  
108 59 3 0 3 0 2 0  
109 86 4 1 5 0 1 0  
110 73 3 0 1 0 1 0  
111 85 3 1 1 0 1 0  
112 69 5 1 3 0 1 1

113 73 3 0 2 0 3 1  
114 71 4 0 5 0 2 0  
115 64 2 0 0 0 2 0  
116 63 4 0 3 0 4 0  
117 79 2 0 1 0 2 1  
118 79 3 0 1 0 2 0  
119 71 2 0 0 0 2 0  
120 63 3 0 1 0 2 0  
121 58 2 0 1 0 2 1  
122 81 4 1 2 0 1 0  
123 81 3 0 1 0 2 0  
124 73 3 0 1 0 2 1  
125 78 3 0 2 0 2 1  
126 84 5 1 4 0 1 0  
127 68 3 1 6 0 1 0  
128 77 3 0 2 0 2 0  
129 73 2 0 2 0 2 0  
130 73 3 0 0 0 2 0  
131 76 3 0 1 0 1 1  
132 67 2 0 0 0 1 0  
133 58 4 0 2 0 2 0  
134 61 3 0 2 0 2 0  
135 67 3 0 2 0 3 0  
136 66 2 0 2 0 3 1  
137 71 3 0 3 0 2 1  
138 77 3 0 1 0 2 0  
139 66 4 1 3 0 1 0  
140 72 2 0 2 0 1 1  
141 63 4 1 3 0 2 0  
142 78 3 0 3 0 2 0  
143 66 2 0 1 0 1 1  
144 72 4 1 1 0 1 0  
145 63 3 0 2 0 2 0  
146 65 2 0 3 0 2 0  
147 77 4 0 3 0 2 0  
148 67 4 0 4 0 2 0  
149 60 2 0 0 0 2 0  
150 64 3 0 0 0 2 0  
151 66 3 0 2 0 2 0  
152 79 2 0 1 0 1 0  
153 62 2 0 2 0 1 0  
154 72 2 0 0 0 2 0  
155 68 3 0 1 0 2 0

156 83 3 0 2 0 1 0  
157 79 3 0 0 0 3 1  
158 66 4 0 3 0 1 1  
159 65 4 1 3 0 1 0  
160 71 3 1 1 0 3 0  
161 74 3 0 0 0 3 0  
162 61 4 0 3 0 2 0  
163 63 1 0 0 0 2 1  
164 58 4 1 3 0 2 0  
165 71 4 0 4 0 2 0  
166 66 3 0 1 0 2 0  
167 73 3 0 4 0 2 0  
168 66 3 0 4 0 2 0  
169 72 5 1 5 0 2 0  
170 69 3 0 2 0 2 1  
171 86 5 1 3 0 2 1  
172 62 5 1 4 0 2 0  
173 64 3 0 3 0 2 1  
174 57 3 0 2 0 1 0  
175 66 3 0 3 0 1 0  
176 63 4 1 5 0 2 1  
177 65 5 1 1 0 1 0  
178 62 2 0 1 0 2 1  
179 53 4 0 4 0 1 0  
180 65 2 0 2 0 1 1  
181 75 5 1 5 0 1 0  
182 68 4 0 4 0 1 1  
183 66 5 1 6 0 1 0  
184 68 4 0 4 0 1 0  
185 61 3 0 2 0 1 1  
186 76 4 1 4 0 2 1  
187 64 3 0 2 0 2 0  
188 69 4 0 4 0 1 0  
189 64 3 0 2 0 1 1  
190 70 4 0 3 0 1 0  
191 60 3 0 2 0 2 0  
192 76 3 0 2 0 2 0  
193 74 2 0 0 0 4 0  
194 69 1 0 0 0 4 1  
195 70 4 0 3 0 1 0  
196 62 3 0 1 0 1 0  
197 73 4 0 2 0 1 0  
198 78 3 0 1 0 1 0

199 71 4 0 3 0 1 0  
200 72 4 1 2 0 2 0  
201 79 4 1 3 0 2 0  
202 68 3 0 1 0 4 0  
203 72 3 0 0 0 2 1  
204 82 3 0 1 0 2 1  
205 82 2 0 0 0 1 0  
206 73 2 0 0 0 2 0  
207 83 2 0 1 0 2 0  
208 67 3 0 2 0 1 1  
209 67 4 1 4 0 1 0  
210 76 4 0 1 0 1 0  
211 87 4 1 2 0 1 0  
212 78 3 0 0 0 2 0  
213 86 3 0 2 0 1 0  
214 79 2 0 1 0 3 0  
215 70 3 0 1 0 1 1  
216 82 4 1 2 0 1 1  
217 67 4 1 1 0 1 1  
218 72 3 0 2 0 1 1  
219 61 4 1 3 0 1 0  
220 87 2 0 0 0 2 0  
221 64 3 0 3 0 2 0  
222 65 3 0 2 0 2 0  
223 62 3 0 0 0 2 1  
224 55 3 0 3 0 1 1  
225 76 3 0 2 0 1 0  
226 72 2 0 1 0 2 0  
227 69 2 0 0 0 2 0  
228 62 2 0 2 0 2 0  
229 83 3 0 0 0 2 0  
230 74 3 0 1 0 2 1  
231 81 4 1 4 0 1 1  
232 69 4 0 3 0 1 0  
233 62 5 1 2 0 1 0  
234 61 3 0 2 0 2 0  
235 66 4 1 3 0 2 0  
236 68 4 1 3 0 1 1  
237 81 4 1 4 0 1 0  
238 62 4 0 4 0 2 1  
239 71 4 0 3 0 2 0  
240 77 4 0 2 0 1 0  
241 83 4 0 4 0 1 0

242 66 4 0 3 0 2 0  
243 80 3 0 2 0 2 0  
244 74 2 0 0 0 2 0  
245 69 3 0 0 0 2 0  
246 85 3 1 4 0 1 0  
247 75 2 0 2 0 2 0  
248 67 3 0 2 0 2 0  
249 77 3 0 3 0 2 1  
250 64 3 0 0 0 2 0  
251 76 3 1 4 0 1 0  
252 79 3 1 2 0 1 1  
253 71 4 1 3 0 1 0  
254 70 4 0 2 0 1 0  
255 66 3 0 1 0 1 0  
256 67 3 0 3 0 3 0  
257 61 3 0 2 0 3 0  
258 71 4 0 6 0 1 0  
259 66 2 0 1 0 2 0  
260 68 3 0 1 0 2 1  
261 85 4 1 5 0 1 0  
262 67 4 0 4 0 1 0  
263 58 4 1 3 0 1 1  
264 82 4 1 2 0 1 0  
265 86 3 0 0 0 2 0  
266 82 4 0 2 0 2 0  
267 83 2 0 1 0 1 0  
268 73 3 0 1 0 2 0  
269 84 3 0 2 0 1 0  
270 67 3 0 2 0 1 0  
271 64 3 0 0 0 2 0  
272 71 2 0 0 0 2 0  
273 69 4 1 2 0 1 0  
274 65 4 1 4 0 1 0  
275 62 3 0 1 0 2 0  
276 82 4 1 3 0 1 1  
277 66 3 1 2 0 1 0  
278 71 3 0 2 0 1 1  
279 67 3 0 1 0 1 0  
280 59 2 0 0 0 2 0  
281 76 2 0 0 0 2 0  
282 80 3 0 1 0 2 0  
283 79 4 1 5 0 2 0  
284 63 2 0 0 0 2 0

285 66 3 0 2 0 2 0  
286 83 4 1 4 0 2 0  
287 73 4 1 4 0 2 0  
288 82 4 0 4 0 2 0  
289 58 3 0 2 0 1 0  
290 65 2 0 2 0 1 0  
291 73 3 0 3 0 1 0  
292 77 1 0 1 0 3 0  
293 87 2 0 1 0 3 0  
294 88 4 1 2 0 1 0  
295 75 2 0 0 0 1 0  
296 75 3 0 2 0 1 0  
297 63 1 0 0 0 2 1  
298 68 5 1 4 0 1 1  
299 63 4 0 3 0 1 0  
300 56 2 0 0 0 2 0  
301 70 2 0 0 0 2 1  
302 66 3 0 3 0 1 0  
303 83 5 1 4 0 2 0  
304 78 4 1 7 0 2 0  
305 64 3 0 2 0 1 0  
306 79 3 0 2 0 2 0  
307 65 3 0 1 0 2 0  
308 88 3 0 1 0 1 0  
309 83 5 0 4 0 1 0  
310 64 2 0 0 0 1 0  
311 68 4 1 2 0 1 1  
312 65 4 1 3 0 1 1  
313 75 3 0 1 0 1 0  
314 63 4 1 3 0 1 0  
315 69 3 0 2 0 1 0  
316 72 2 0 1 0 2 0  
317 81 3 0 3 0 2 1  
318 83 5 1 6 0 2 0  
319 66 3 0 2 0 2 0  
320 81 4 1 3 0 1 0  
321 80 3 0 1 0 1 0  
322 63 3 0 1 0 2 0  
323 66 2 0 1 0 3 1  
324 82 4 0 3 0 2 0  
325 65 3 0 2 0 2 0  
326 74 4 0 1 0 1 0  
327 69 4 0 1 0 1 1



328 79 4 1 5 0 1 0  
329 76 3 0 1 0 1 0  
330 60 2 0 1 0 3 0  
331 65 2 0 1 0 3 0  
332 74 3 0 1 0 1 0  
333 82 4 0 0 0 1 0  
334 69 3 0 2 0 1 0  
335 67 2 0 0 0 1 0  
336 61 3 0 3 0 1 0  
337 72 3 1 2 0 1 0  
338 64 4 0 4 0 2 0  
339 82 2 0 0 0 2 0  
340 62 2 0 1 0 3 0  
341 69 3 1 0 0 2 0  
342 82 3 0 1 0 2 0  
343 76 2 0 0 0 2 0  
344 77 2 0 0 0 2 0  
345 77 4 1 3 0 2 0  
346 66 5 1 4 0 2 1  
347 67 3 1 3 0 2 0  
348 74 3 0 3 0 4 0  
349 66 2 0 1 0 4 0  
350 63 3 0 2 0 2 0  
351 70 3 0 0 0 2 0  
352 77 5 1 6 0 1 0  
353 77 3 0 0 0 2 0  
354 75 4 0 3 0 1 0  
355 71 2 0 0 0 2 1  
356 66 2 0 0 0 2 1  
357 61 5 1 5 0 2 0  
358 77 2 0 0 0 1 1  
359 71 3 0 2 0 1 0  
360 67 3 0 1 0 1 0  
361 60 3 1 1 0 1 0  
362 79 3 0 2 0 1 1  
363 70 2 0 2 0 2 0  
364 61 3 1 3 0 2 0  
365 59 3 0 0 0 2 0  
366 61 2 0 0 0 2 0  
367 56 2 0 0 0 1 0  
368 79 3 0 2 0 1 0  
369 65 2 0 0 0 4 1  
370 68 4 1 4 0 1 0

371 63 3 0 0 0 2 0  
372 74 2 0 1 0 2 0  
373 64 4 1 3 0 2 0  
374 65 2 1 0 0 1 0  
375 72 4 1 1 0 1 0  
376 77 2 0 1 0 2 0  
377 77 2 0 0 0 2 0  
378 68 2 0 2 0 2 1  
379 58 5 1 3 0 1 0  
380 59 3 0 3 0 1 1  
381 64 4 0 3 0 1 1  
382 84 2 0 2 0 2 0  
383 80 3 0 1 0 1 0  
384 81 4 0 1 0 1 0  
385 65 4 0 3 0 2 0  
386 61 3 0 1 0 2 1  
387 87 4 0 2 0 1 1  
388 64 4 0 3 0 1 1  
389 67 4 0 4 0 1 0  
390 64 3 0 0 0 1 0  
391 68 4 0 4 0 1 0  
392 83 4 1 4 0 1 0  
393 66 3 0 3 0 2 0  
394 73 3 0 2 0 2 0  
395 66 2 0 2 0 1 0  
396 59 3 0 2 0 1 0  
397 65 3 1 2 0 2 0  
398 80 2 0 0 0 2 1  
399 63 4 1 2 0 3 0  
400 62 2 0 2 0 2 0  
401 77 3 0 3 0 2 0  
402 72 3 0 1 0 3 0  
403 83 3 0 1 0 3 0  
404 61 2 0 0 0 3 0  
405 57 3 0 1 0 1 0  
406 69 4 0 3 0 1 1  
407 54 3 0 0 0 3 1  
408 64 3 0 2 0 3 1  
409 79 4 1 2 0 1 0  
410 66 3 0 1 0 2 0  
411 82 4 0 2 0 1 0  
412 71 5 1 5 0 1 0  
413 83 3 0 0 0 2 0

414 80 4 1 5 0 1 0  
415 72 3 0 3 0 2 0  
416 78 4 0 6 0 1 0  
417 83 3 1 3 0 2 0  
418 70 2 0 0 0 2 0  
419 87 3 0 2 0 3 0  
420 58 2 0 0 0 3 0  
421 66 4 1 4 0 1 1  
422 85 3 1 2 0 2 0  
423 68 3 0 3 0 1 1  
424 68 3 0 2 0 1 0  
425 63 3 0 0 0 3 0  
426 66 3 0 0 0 3 1  
427 70 3 0 2 0 1 0  
428 73 2 0 0 0 1 0  
429 76 2 0 0 0 1 0  
430 70 3 0 0 0 1 0  
431 69 2 0 1 0 1 1  
432 65 3 0 2 0 1 0  
433 58 2 0 1 0 2 0  
434 65 4 0 4 0 2 1  
435 65 4 1 4 0 1 0  
436 76 4 1 4 0 2 0  
437 67 3 1 3 0 2 0  
438 69 3 0 3 0 4 0  
439 73 2 0 0 0 2 0  
440 70 4 1 3 0 1 0  
441 82 3 0 2 0 1 1  
442 80 4 0 3 0 1 0  
443 79 3 0 2 0 1 1  
444 80 4 0 1 0 1 0  
445 66 3 0 4 0 1 0  
446 72 3 0 5 0 3 0  
447 68 2 0 0 0 2 0  
448 64 2 0 0 0 2 0  
449 68 3 0 2 0 2 0  
450 61 2 0 0 0 2 0  
451 82 3 0 2 0 1 1  
452 76 3 0 3 0 1 0  
453 77 4 1 3 0 1 0  
454 71 4 0 3 0 4 0  
455 64 4 1 2 0 4 0  
456 70 3 0 1 0 3 0

457 77 2 0 0 0 3 0  
458 65 3 0 1 0 2 0  
459 57 4 1 3 0 1 0  
460 63 3 0 1 0 1 0  
461 69 1 0 0 0 2 0  
462 58 2 0 1 0 2 0  
463 59 3 0 1 0 1 1  
464 63 4 0 2 0 1 0  
465 77 4 1 3 0 2 1  
466 78 4 1 4 0 2 1  
467 83 4 1 4 0 1 1  
468 66 4 0 1 0 1 0  
469 63 4 0 3 0 2 0  
470 76 3 0 1 0 2 0  
471 78 2 0 1 0 2 0  
472 63 3 0 4 0 2 0  
473 66 3 0 2 0 2 0  
474 90 3 0 2 0 3 0  
475 65 2 0 1 0 1 0  
476 81 2 0 2 0 1 0  
477 65 3 1 2 0 2 0  
478 67 4 1 2 0 2 0  
479 65 3 0 1 0 2 0  
480 65 2 0 0 0 2 0  
481 65 4 1 4 0 1 0  
482 59 4 1 4 0 1 0  
483 78 3 0 2 0 1 0  
484 70 3 0 2 0 1 0  
485 82 2 0 2 0 3 0  
486 65 3 0 3 0 1 0  
487 57 3 0 0 0 1 1  
488 65 3 0 3 0 1 1  
489 73 2 1 2 0 3 1  
490 75 5 1 7 0 1 0  
491 64 2 0 0 0 2 0  
492 51 1 0 0 0 2 0  
493 71 3 0 1 0 1 1  
494 67 4 0 3 0 1 0  
495 75 3 1 2 0 1 0  
496 88 4 1 2 0 2 0  
497 74 2 0 1 0 2 0  
498 82 4 0 4 0 2 0  
499 71 4 0 3 0 3 0

500 63 2 0 0 0 3 0  
501 79 2 0 2 0 3 0  
502 72 2 0 4 0 3 1  
503 73 4 1 4 0 1 1  
504 70 4 0 3 0 1 0  
505 64 3 0 1 0 1 0  
506 77 3 0 1 0 2 0  
507 71 3 0 1 0 1 0  
508 73 3 0 1 0 1 0  
509 64 3 0 3 0 3 0  
510 67 2 0 0 0 3 0  
511 79 3 0 3 0 1 0  
512 78 3 0 1 0 1 1  
513 61 5 1 4 0 3 1  
514 71 3 0 2 0 2 0  
515 68 3 0 3 0 1 0  
516 62 5 0 6 0 1 0  
517 54 4 0 5 0 1 0  
518 77 3 0 2 0 2 0  
519 72 3 0 1 0 1 0  
520 79 3 0 2 0 1 0  
521 64 3 0 1 0 1 1  
522 59 4 1 3 0 1 0  
523 73 4 1 2 0 1 1  
524 67 3 0 4 0 1 0  
525 65 1 0 0 0 2 0  
526 73 3 0 0 0 2 0  
527 73 2 0 0 0 2 0  
528 61 3 0 0 0 2 0  
529 78 3 0 3 0 2 0  
530 69 4 0 5 0 2 0  
531 74 2 0 1 0 2 0  
532 69 2 0 0 0 2 0  
533 68 4 1 4 0 2 0  
534 87 4 0 3 0 2 0  
535 84 3 0 2 0 1 1  
536 65 4 1 2 0 1 0  
537 70 4 0 2 0 2 0  
538 57 3 0 1 0 2 0  
539 71 3 0 2 0 1 0  
540 71 3 0 2 0 2 0  
541 67 2 0 1 0 2 1  
542 77 5 0 4 0 1 0

543 88 1 0 0 0 2 0  
544 78 2 0 0 0 2 0  
545 64 3 0 2 0 2 0  
546 85 3 0 2 0 2 0  
547 75 2 0 1 0 2 1  
548 64 3 0 5 0 2 0  
549 71 3 0 2 0 2 0  
550 66 3 0 2 0 2 0  
551 58 4 1 3 0 1 0  
552 83 3 1 3 0 2 1  
553 87 4 0 5 0 2 0  
554 78 3 1 2 0 1 0  
555 63 4 0 2 0 1 0  
556 51 3 0 1 0 2 0  
557 63 3 0 3 0 1 0  
558 76 4 0 3 0 1 0  
559 64 3 0 2 0 2 0  
560 70 3 0 2 0 2 0  
561 65 4 1 3 0 2 0  
562 58 2 0 0 0 2 0  
563 75 5 1 4 0 2 0  
564 71 2 0 1 0 2 0  
565 67 3 0 2 0 2 0  
566 60 2 0 0 0 2 0  
567 73 3 0 0 0 1 0  
568 71 3 0 1 0 1 1  
569 69 3 0 2 0 3 0  
570 67 4 1 5 0 2 0  
571 52 4 1 2 0 2 0  
572 83 5 1 5 0 1 0  
573 63 4 0 3 0 2 0  
574 52 4 1 3 0 2 0  
575 71 3 0 1 0 3 0  
576 64 2 0 0 0 3 1  
577 74 3 0 3 0 1 0  
578 71 2 0 0 0 2 0  
579 74 3 0 2 0 1 0  
580 68 3 0 3 0 1 0  
581 70 2 0 0 0 1 0  
582 74 3 0 2 0 1 1  
583 93 5 1 9 0 1 1  
584 84 5 1 5 0 1 0  
585 85 3 0 0 0 3 0

586 66 2 0 0 0 3 0  
587 80 2 0 2 0 2 0  
588 73 2 0 1 0 2 0  
589 68 2 0 1 0 2 0  
590 71 3 1 3 0 2 0  
591 65 4 0 2 0 2 0  
592 62 3 1 1 0 2 0  
593 82 3 0 4 0 2 0  
594 64 3 0 3 0 2 1  
595 59 2 0 0 0 2 0  
596 65 3 0 2 0 3 0  
597 88 3 0 2 0 2 0  
598 83 3 0 1 0 2 1  
599 64 4 1 3 0 2 1  
600 82 4 0 5 0 1 1  
601 65 4 0 3 0 1 1  
602 78 4 1 3 0 1 0  
603 77 4 0 3 0 1 0  
604 69 2 0 0 0 2 0  
605 63 2 0 1 0 2 1  
606 61 4 0 0 0 2 0  
607 71 3 0 3 0 1 0  
608 63 3 0 2 0 2 0  
609 67 4 0 4 0 2 0  
610 61 4 1 3 0 1 0  
611 57 2 0 1 0 2 0  
612 66 4 0 0 0 2 0  
613 81 4 0 2 0 1 0  
614 80 3 0 1 0 2 0  
615 64 4 1 3 0 2 0  
616 63 2 0 3 0 3 0  
617 61 3 0 1 0 3 1  
618 74 5 0 4 0 2 1  
619 67 4 1 3 0 1 0  
620 67 3 0 2 0 2 0  
621 63 1 0 1 0 4 0  
622 64 3 0 2 0 3 0  
623 80 2 0 0 0 1 0  
624 73 3 0 2 0 1 1  
625 72 4 0 4 0 1 0  
626 64 3 0 2 0 1 0  
627 63 3 0 2 0 1 0  
628 68 3 0 1 0 1 0

629 77 4 0 4 0 1 0  
630 68 4 1 2 0 2 0  
631 61 2 0 1 0 2 0  
632 80 5 1 2 0 1 0  
633 78 2 0 2 0 2 0  
634 74 2 0 1 0 2 1  
635 63 3 0 2 0 3 1  
636 70 3 1 1 0 3 0  
637 73 3 0 1 0 2 1  
638 61 3 0 1 0 2 0  
639 58 2 0 1 0 1 0  
640 84 3 1 1 0 1 1  
641 64 5 0 3 0 1 0  
642 61 2 0 1 0 3 0  
643 80 2 0 0 0 3 0  
644 62 2 0 0 0 1 0  
645 78 3 0 1 0 1 1  
646 72 3 0 2 0 1 1  
647 73 4 1 2 0 1 1  
648 82 4 1 4 0 1 0  
649 86 4 1 2 0 1 0  
650 81 1 0 1 0 2 1  
651 86 1 0 3 0 2 1  
652 66 4 1 3 0 1 0  
653 77 3 0 2 0 2 0  
654 71 3 0 0 0 2 0  
655 81 4 0 2 0 1 0  
656 70 3 0 0 0 2 0  
657 69 3 0 1 0 2 0  
658 74 3 0 2 0 2 0  
659 74 3 0 2 0 2 0  
660 66 3 0 0 0 3 0  
661 59 2 0 0 0 3 0  
662 75 3 0 4 0 1 0  
663 61 4 0 3 0 1 0  
664 59 3 0 1 0 1 0  
665 68 3 0 0 0 1 0  
666 67 3 1 5 0 2 0  
667 71 2 0 2 0 2 0  
668 64 2 0 3 0 2 0  
669 65 3 0 2 0 1 0  
670 77 3 0 2 0 1 0  
671 83 2 0 1 0 1 0



672 65 3 0 1 0 1 1  
673 64 5 1 5 0 2 0  
674 64 4 0 2 0 1 0  
675 63 3 0 1 0 2 0  
676 69 4 0 4 0 2 0  
677 72 4 1 3 0 2 0  
678 66 3 1 1 0 1 0  
679 72 3 0 2 0 1 0  
680 65 4 1 2 0 2 0  
681 49 3 0 1 0 2 1  
682 78 4 1 3 0 1 0  
683 70 3 0 2 0 1 0  
684 61 3 0 0 0 1 0  
685 80 2 0 0 0 1 1  
686 66 4 0 2 0 1 0  
687 83 2 0 0 0 1 1  
688 73 3 0 1 0 2 1  
689 75 4 1 2 0 2 0  
690 74 3 0 0 0 4 0  
691 67 2 0 0 0 2 0  
692 84 3 0 0 0 2 0  
693 84 2 0 0 0 1 0  
694 73 2 0 3 0 1 0  
695 75 3 1 6 0 1 0  
696 82 2 0 2 0 2 0  
697 66 3 0 3 0 2 0  
698 65 3 0 1 0 2 0  
699 73 2 0 0 0 1 0  
700 77 3 0 1 0 1 0  
701 65 3 0 2 0 1 0  
702 57 4 1 2 0 1 0  
703 63 3 0 4 0 3 0  
704 70 3 0 2 0 3 1  
705 79 5 1 4 0 1 1  
706 62 4 0 4 0 1 1  
707 72 4 1 5 0 1 1  
708 78 4 0 1 0 1 1  
709 81 4 1 6 0 2 0  
710 72 5 1 4 0 1 0  
711 75 4 1 4 0 1 1  
712 83 3 1 2 0 1 0  
713 68 2 0 2 0 1 0  
714 68 3 0 1 0 2 1

715 60 3 0 3 0 1 0  
716 63 2 0 1 0 1 0  
717 65 4 0 3 0 1 1  
718 65 4 1 4 0 1 0  
719 70 3 0 1 0 2 1  
720 68 4 0 3 0 2 0  
721 61 4 0 1 0 2 0  
722 58 3 0 0 0 2 0  
723 81 4 1 4 0 1 0  
724 81 2 0 1 0 3 0  
725 80 2 0 1 0 3 0  
726 67 2 0 0 0 1 0  
727 87 4 0 3 0 1 0  
728 64 3 0 2 0 2 0  
729 61 5 1 5 0 1 0  
730 75 3 1 2 0 1 0  
731 61 3 0 2 0 1 0  
732 57 3 0 1 0 2 0  
733 72 2 0 0 0 2 0  
734 63 4 1 5 0 2 0  
735 78 2 0 1 0 2 0  
736 72 3 0 1 0 1 0  
737 67 2 0 0 0 1 0  
738 69 3 0 1 0 2 1  
739 66 3 0 0 0 2 0  
740 77 4 1 6 0 1 1  
741 77 5 1 6 0 1 0  
742 67 3 1 3 0 2 0  
743 67 4 1 5 0 3 0  
744 67 2 0 2 0 1 0  
745 72 2 0 1 0 1 1  
746 65 4 0 0 0 2 0  
747 61 3 0 3 0 1 0  
748 87 2 0 2 0 1 1  
749 65 3 1 3 0 1 1  
750 77 3 0 1 0 1 0  
751 61 3 1 3 0 2 0  
752 70 3 0 0 0 2 0  
753 85 4 0 2 0 2 0  
754 68 2 0 1 0 3 0  
755 78 3 0 2 0 2 0  
756 69 3 0 2 0 2 1  
757 77 4 0 5 0 1 1

758 82 4 0 6 0 2 0  
759 82 4 1 6 0 2 0  
760 84 4 0 4 0 3 0  
761 79 3 0 1 0 3 0  
762 65 3 0 3 0 1 0  
763 69 4 0 3 0 1 1  
764 73 4 1 3 0 1 1  
765 64 4 1 4 0 2 0  
766 63 3 0 3 0 2 0  
767 55 2 0 3 0 3 1  
768 71 4 0 3 0 1 1  
769 63 4 1 4 0 2 0  
770 65 2 0 1 0 3 0  
771 85 3 0 1 0 3 0  
772 73 3 1 1 0 2 0  
773 87 2 0 0 0 2 0  
774 65 4 1 3 0 1 0  
775 61 4 1 3 0 1 0  
776 73 3 0 2 0 2 0  
777 82 3 0 3 0 2 0  
778 73 3 0 2 0 2 0  
779 68 3 0 1 0 2 0  
780 79 4 0 3 0 1 1  
781 80 3 0 1 0 3 1  
782 68 3 1 2 0 1 0  
783 75 3 0 2 0 2 0  
784 73 3 0 3 0 2 1  
785 71 4 1 3 0 1 0  
786 69 2 0 0 0 2 0  
787 73 2 0 2 0 2 0  
788 73 3 1 1 0 1 0  
789 65 4 1 3 0 2 0  
790 70 4 1 5 0 2 1  
791 64 3 1 1 0 1 0  
792 69 1 0 0 0 1 0  
793 71 3 0 1 0 1 0  
794 75 4 0 2 0 1 0  
795 68 2 0 0 0 2 0  
796 66 3 0 0 0 2 0  
797 79 3 0 2 0 2 0  
798 85 2 0 0 0 1 0  
799 83 3 0 1 0 2 0  
800 65 4 1 2 0 2 0

801 70 3 0 1 0 2 0  
802 87 2 0 0 0 2 0  
803 78 3 0 2 0 1 0  
804 79 4 1 3 0 3 0  
805 79 3 0 3 0 2 1  
806 73 4 0 3 0 1 0  
807 77 3 0 1 0 1 1  
808 69 4 0 5 0 1 0  
809 63 3 0 0 0 2 0  
810 77 3 0 0 0 2 1  
811 66 4 0 4 0 1 1  
812 74 3 0 3 0 1 0  
813 73 3 0 2 0 1 0  
814 70 3 1 1 0 2 0  
815 84 2 0 0 0 2 0  
816 64 3 1 1 0 2 0  
817 61 3 1 2 0 1 0  
818 69 3 0 0 0 1 0  
819 83 3 1 2 0 3 0  
820 81 3 0 0 0 3 0  
821 67 5 1 4 0 1 0  
822 55 3 1 4 0 2 0  
823 68 3 0 2 0 1 0  
824 63 2 0 0 0 2 0  
825 63 2 0 1 0 2 1  
826 64 3 0 4 0 2 1  
827 71 4 0 2 0 1 0  
828 76 3 0 0 0 1 1  
829 60 3 0 2 0 1 0  
830 65 2 0 3 0 2 0  
831 62 2 0 0 0 4 0  
832 69 4 0 1 0 3 0  
833 78 2 0 1 0 3 0  
834 82 3 0 2 0 2 0  
835 86 3 0 1 0 2 1  
836 62 3 0 2 0 1 0  
837 77 2 0 0 0 4 0  
838 73 2 0 0 0 4 1  
839 67 5 1 5 0 2 0  
840 60 4 0 3 0 2 1  
841 84 3 0 1 0 2 0  
842 80 4 0 3 0 1 0  
843 71 2 0 2 0 2 0

844 62 3 0 2 0 2 0  
845 56 5 0 2 0 3 0  
846 74 4 0 3 0 1 0  
847 72 4 0 2 0 1 0  
848 74 4 1 5 0 2 1  
849 75 4 1 5 0 2 0  
850 72 3 0 2 0 2 0  
851 69 3 0 2 0 2 1  
852 80 4 0 1 0 1 0  
853 88 3 0 2 0 1 0  
854 67 4 1 5 0 3 0  
855 58 5 0 4 0 3 0  
856 61 4 1 3 0 1 1  
857 84 2 0 2 0 1 0  
858 85 2 0 2 0 2 0  
859 73 4 0 4 0 2 0  
860 59 3 0 3 0 1 0  
861 54 4 1 3 0 1 0  
862 70 2 1 1 0 4 0  
863 59 2 1 1 0 4 0  
864 67 3 0 4 0 1 1  
865 65 5 1 3 0 2 0  
866 67 3 0 2 0 2 0  
867 64 4 0 3 0 2 0  
868 52 2 0 1 0 3 0  
869 77 3 0 3 0 3 0  
870 62 2 0 0 0 4 0  
871 60 4 1 4 0 4 0  
872 64 3 0 1 0 2 0  
873 74 4 0 3 0 2 1  
874 79 3 0 5 0 1 0  
875 80 2 0 1 0 3 0  
876 65 3 0 2 0 3 0  
877 63 4 0 3 0 1 0  
878 75 3 0 2 0 2 0  
879 75 3 0 2 0 1 0  
880 70 3 0 1 0 1 0  
881 76 4 1 3 0 2 0  
882 77 4 0 2 0 1 0  
883 78 3 0 2 0 2 0  
884 86 3 0 2 0 3 0  
885 87 3 0 2 0 3 0  
886 81 4 1 3 0 1 0

887 74 4 1 2 0 1 0  
888 76 3 0 2 0 2 0  
889 83 2 0 2 0 2 0  
890 78 4 1 3 0 2 0  
891 72 4 1 4 0 2 0  
892 62 2 0 1 0 2 0  
893 57 2 0 0 0 2 0  
894 74 2 0 1 0 2 1  
895 82 4 0 2 0 1 1  
896 74 3 0 2 0 1 0  
897 80 4 0 4 0 1 1  
898 76 3 1 4 0 1 0  
899 75 3 0 1 0 2 0  
900 66 3 0 2 0 2 0  
901 61 3 0 2 0 2 0  
902 69 3 0 1 0 2 0  
903 60 3 0 1 0 1 1  
904 87 4 0 3 0 2 0  
905 64 4 0 2 0 1 0  
906 54 3 0 2 0 1 1  
907 81 4 0 4 0 1 0  
908 75 4 1 3 0 1 0  
909 84 3 0 3 0 2 0  
910 74 4 0 1 0 2 0  
911 67 4 0 2 0 2 0  
912 80 3 0 4 0 2 0  
913 75 3 0 1 0 2 0  
914 66 1 0 0 0 2 0  
915 72 2 0 0 0 2 0  
916 64 2 0 0 0 2 0  
917 72 3 0 1 0 2 0  
918 74 3 0 1 0 1 0  
919 60 3 1 3 0 1 0  
920 86 2 0 0 0 3 0  
921 79 2 0 0 0 3 0  
922 59 3 0 2 0 2 0  
923 68 4 0 3 0 1 1  
924 67 4 1 7 0 1 0  
925 72 3 0 1 0 2 0  
926 71 3 0 2 0 2 1  
927 78 4 1 4 0 1 1  
928 89 5 1 2 0 1 0  
929 63 3 1 4 0 2 0

930 71 2 0 0 0 2 0  
931 65 4 0 0 0 1 1  
932 59 4 0 3 0 1 0  
933 83 3 0 0 0 2 0  
934 79 3 0 3 0 2 0  
935 68 3 0 1 0 2 0  
936 67 4 1 3 0 2 0  
937 81 3 0 2 0 1 0  
938 74 3 0 4 0 3 0  
939 68 3 0 2 0 3 0  
940 74 3 0 1 0 2 0  
941 76 3 1 1 0 2 0  
942 70 4 1 4 0 3 0  
943 56 4 1 3 0 3 0  
944 66 4 0 3 0 2 0  
945 70 3 0 3 0 2 0  
946 71 3 0 1 0 1 0  
947 67 3 0 2 0 1 0  
948 61 3 0 1 0 1 0  
949 80 3 0 1 0 3 0  
950 62 5 1 4 0 1 0  
951 54 4 0 3 0 3 0  
952 66 5 0 4 0 3 1  
953 77 5 1 5 0 1 0  
954 79 5 1 4 0 2 0  
955 61 2 0 1 0 3 0  
956 71 2 0 1 0 3 0  
957 63 3 0 1 0 1 0  
958 53 3 0 1 0 2 0  
959 73 2 0 1 0 2 0  
960 68 4 0 4 0 1 0  
961 61 2 0 0 0 3 0  
962 56 4 0 1 0 2 0  
963 63 3 0 3 0 1 0  
964 73 3 0 2 0 1 0  
965 75 3 0 1 0 2 0  
966 64 3 0 1 0 2 0  
967 72 4 1 3 0 1 0  
968 69 2 0 1 0 1 0  
969 72 3 1 1 0 1 0  
970 66 3 0 1 0 1 0  
971 63 4 0 2 0 1 0  
972 71 4 1 2 0 1 0

973 62 3 0 0 0 1 0  
974 75 2 0 2 0 1 0  
975 87 4 1 5 0 2 0  
976 72 3 0 1 0 1 1  
977 73 4 0 3 0 2 0  
978 65 2 0 0 0 2 0  
979 77 3 0 1 0 2 0  
980 69 3 0 3 0 2 0  
981 76 2 0 0 0 2 0  
982 66 3 0 0 0 2 0  
983 83 3 0 1 0 2 1  
984 66 3 0 2 0 2 0  
985 68 3 0 2 0 3 0  
986 61 3 0 2 0 2 0  
987 85 2 0 0 0 2 0  
988 72 4 0 1 0 2 0  
989 74 3 0 1 0 2 0  
990 76 3 0 3 0 2 1  
991 65 3 0 4 0 2 0  
992 80 2 0 2 0 2 0  
993 84 3 0 3 0 2 0  
994 86 4 1 2 0 2 0  
995 59 4 1 3 0 2 0  
996 68 3 0 3 0 2 0  
997 77 5 1 4 0 2 0  
998 69 4 1 3 0 1 0  
999 61 3 0 1 0 1 0  
1000 74 4 0 3 0 1 1  
1001 72 4 1 2 0 1 0  
1002 83 4 0 4 0 2 0  
1003 77 4 1 5 0 1 1  
1004 65 4 1 0 0 1 0  
1005 86 3 0 2 0 1 0  
1006 64 4 0 2 0 1 1  
1007 87 5 1 4 0 1 0  
1008 57 2 0 0 0 2 0  
1009 67 2 0 1 0 2 0  
1010 69 3 0 2 0 1 0  
1011 65 3 0 0 0 2 0  
1012 59 4 0 0 0 2 1  
1013 60 4 1 4 0 2 0  
1014 66 3 0 1 0 2 0  
1015 71 3 0 1 0 2 0



1016 64 4 1 1 0 2 0  
1017 66 3 0 2 0 2 0  
1018 74 4 0 5 0 1 0  
1019 71 3 0 2 0 3 0  
1020 74 4 0 4 0 2 0  
1021 71 4 0 4 0 2 0  
1022 63 5 1 8 0 1 0  
1023 56 4 0 3 0 1 0  
1024 78 4 0 5 0 1 0  
1025 78 4 1 4 0 2 1  
1026 80 4 0 4 0 1 0  
1027 78 3 0 3 0 2 1  
1028 83 5 1 7 0 1 1  
1029 66 4 1 5 0 1 0  
1030 62 3 0 2 0 1 0  
1031 73 4 1 1 0 2 1  
1032 73 4 1 3 0 2 0  
1033 90 3 0 1 0 2 0  
1034 65 4 0 2 0 2 1  
1035 71 2 0 1 0 2 0  
1036 71 3 0 1 0 1 0  
1037 79 3 0 3 0 1 0  
1038 68 4 0 3 0 2 0  
1039 69 3 0 1 0 2 1  
1040 77 3 0 3 0 1 0  
1041 65 3 1 1 0 2 1  
1042 59 3 0 1 0 2 1  
1043 66 3 0 3 0 2 0  
1044 65 3 0 1 0 1 0  
1045 66 3 0 1 0 1 0  
1046 70 4 0 4 0 1 0  
1047 79 3 0 2 0 1 0  
1048 73 3 0 5 0 4 0  
1049 86 4 1 3 0 1 0  
1050 81 4 1 2 0 1 0  
1051 83 4 1 4 0 1 0  
1052 84 4 1 3 0 1 0  
1053 77 2 0 0 0 3 0  
1054 81 3 0 3 0 1 0  
1055 79 3 0 2 0 1 0  
1056 87 3 0 1 0 1 1  
1057 64 4 0 0 0 1 0  
1058 84 3 0 1 0 2 1

1059 66 4 1 4 0 2 0  
1060 68 2 1 1 0 2 0  
1061 64 3 0 2 0 1 0  
1062 57 3 1 1 0 1 0  
1063 65 2 0 2 0 2 0  
1064 63 3 0 1 0 2 0  
1065 57 3 0 2 0 2 0  
1066 63 3 0 1 0 2 0  
1067 65 2 0 0 0 3 0  
1068 65 3 0 0 0 2 0  
1069 64 4 1 2 0 2 0  
1070 53 2 0 0 0 2 0  
1071 69 3 0 1 0 1 0  
1072 65 3 0 3 0 1 0  
1073 69 4 0 3 0 1 1  
1074 67 4 1 3 0 1 0  
1075 71 2 0 0 0 2 0  
1076 73 3 0 1 0 1 1  
1077 69 4 0 2 0 1 0  
1078 82 4 0 3 0 2 0  
1079 82 2 0 0 0 2 0  
1080 77 3 0 2 0 2 0  
1081 85 4 0 3 0 1 0  
1082 85 4 1 4 0 1 0  
1083 52 3 0 1 0 2 1  
1084 67 5 1 3 0 2 0  
1085 69 3 0 3 0 1 0  
1086 62 2 0 0 0 2 0  
1087 84 4 1 2 0 2 0  
1088 79 3 0 2 0 2 0  
1089 64 4 1 5 0 1 1  
1090 58 5 1 5 0 1 0  
1091 64 3 0 3 0 2 0  
1092 62 3 0 1 0 2 0  
1093 65 3 1 2 0 2 0  
1094 65 4 1 2 0 2 0  
1095 67 4 0 2 0 2 0  
1096 59 3 0 0 0 2 0  
1097 77 1 0 0 0 2 0  
1098 69 3 0 3 0 2 0  
1099 57 3 0 1 0 2 0  
1100 63 4 0 2 0 4 0  
1101 65 3 0 1 0 4 1

1102 61 5 1 6 0 1 0  
1103 73 3 0 2 0 1 0  
1104 82 4 0 1 0 3 0  
1105 63 4 1 3 0 2 0  
1106 60 4 1 4 0 2 1  
1107 65 4 0 4 0 1 0  
1108 74 3 0 3 0 1 0  
1109 60 3 0 1 0 1 0  
1110 72 2 0 0 0 1 0  
1111 89 2 0 0 0 2 0  
1112 86 2 0 1 0 2 0  
1113 75 3 0 3 0 1 0  
1114 92 4 0 2 0 1 0  
1115 81 2 0 1 0 1 0  
1116 65 4 0 2 0 1 0  
1117 68 3 0 1 0 3 0  
1118 69 3 0 1 0 3 0  
1119 82 2 0 3 0 3 0  
1120 79 3 0 2 0 1 0  
1121 61 2 0 3 0 1 0  
1122 67 3 1 1 0 1 0  
1123 80 2 0 1 0 1 0  
1124 77 2 0 0 0 2 0  
1125 83 3 0 0 0 2 0  
1126 66 4 0 3 0 1 1  
1127 59 5 0 5 0 1 0  
1128 79 4 1 4 0 2 1  
1129 66 4 1 4 0 1 0  
1130 67 4 0 3 0 1 0  
1131 64 3 0 3 0 1 0  
1132 88 3 0 0 0 1 0  
1133 75 3 0 2 0 3 0  
1134 70 2 0 0 0 3 0  
1135 74 3 0 1 0 1 1  
1136 67 3 0 1 0 1 0  
1137 78 2 0 0 0 2 0  
1138 77 3 1 1 0 1 0  
1139 80 4 0 2 0 2 0  
1140 64 4 1 3 0 4 0  
1141 65 3 0 2 0 2 0  
1142 61 4 0 4 0 1 0  
1143 78 5 1 4 0 1 0  
1144 72 3 0 0 0 1 0

1145 72 3 0 1 0 1 1  
1146 62 2 0 0 0 1 0  
1147 65 3 0 1 0 1 0  
1148 69 4 1 2 0 2 0  
1149 69 3 0 0 0 2 0  
1150 73 4 0 3 0 1 0  
1151 64 4 0 3 0 1 1  
1152 66 2 0 2 0 2 0  
1153 63 3 0 3 0 2 0  
1154 66 3 0 2 0 2 0  
1155 66 4 0 2 0 1 0  
1156 70 2 0 0 0 2 0  
1157 79 3 0 3 0 1 0  
1158 76 3 0 2 0 1 0  
1159 64 3 0 2 0 2 0  
1160 67 3 0 0 0 1 0  
1161 81 3 0 2 0 1 0  
1162 74 2 0 0 0 2 0  
1163 71 4 0 3 0 2 0  
1164 62 4 0 3 0 1 0  
1165 65 4 0 4 0 1 0  
1166 62 2 0 0 0 3 0  
1167 68 3 0 0 0 2 0  
1168 66 3 0 1 0 2 0  
1169 70 3 0 1 0 3 1  
1170 67 3 0 0 0 3 1  
1171 67 4 0 2 0 1 1  
1172 75 3 1 1 0 2 0  
1173 75 2 0 0 0 2 0  
1174 70 2 0 1 0 2 0  
1175 72 3 0 2 0 2 1  
1176 67 4 1 7 0 1 0  
1177 63 4 0 2 0 1 0  
1178 67 3 1 2 0 2 0  
1179 74 3 0 3 0 2 1  
1180 79 5 1 6 0 1 0  
1181 59 4 0 4 0 1 0  
1182 68 4 0 4 0 2 1  
1183 64 3 1 0 0 1 0  
1184 69 3 0 4 0 3 1  
1185 77 4 0 2 0 1 0  
1186 64 4 1 3 0 1 0  
1187 78 3 0 2 0 2 0

1188 68 4 0 2 0 2 1  
1189 77 4 1 4 0 2 0  
1190 71 5 1 4 0 1 0  
1191 71 2 0 0 0 1 1  
1192 58 2 0 1 0 1 0  
1193 83 3 0 2 0 4 0  
1194 71 2 0 2 0 4 0  
1195 67 5 1 4 0 2 0  
1196 67 3 0 2 0 2 0  
1197 64 4 0 1 0 2 0  
1198 62 3 0 0 0 2 0  
1199 63 4 1 4 0 2 1  
1200 59 5 1 3 0 2 0  
1201 70 3 0 3 0 2 1  
1202 81 3 0 2 0 1 0  
1203 73 3 0 2 0 1 0  
1204 74 2 0 0 0 3 0  
1205 72 2 0 1 0 3 0  
1206 73 3 0 2 0 2 0  
1207 69 3 0 1 0 2 1  
1208 77 4 0 4 0 1 0  
1209 74 2 0 1 0 2 1  
1210 88 3 0 1 0 2 0  
1211 80 3 0 2 0 3 0  
1212 68 3 0 2 0 1 1  
1213 68 3 0 3 0 1 1  
1214 83 5 1 7 0 1 0  
1215 73 2 0 2 0 4 0  
1216 78 3 0 1 0 4 0  
1217 82 3 0 1 0 2 1  
1218 82 3 0 0 0 2 0  
1219 89 4 0 0 0 2 0  
1220 91 3 0 3 0 2 0  
1221 86 3 0 1 0 2 0  
1222 71 3 0 0 0 2 0  
1223 61 3 1 3 0 2 0  
1224 55 3 1 3 0 2 0  
1225 73 4 0 3 0 1 1  
1226 70 4 1 3 0 1 0  
1227 64 3 0 2 0 1 0  
1228 70 2 1 0 0 2 1  
1229 73 3 1 0 0 2 0  
1230 73 3 0 0 0 2 1

1231 75 3 0 3 0 2 0  
1232 69 4 1 4 0 2 1  
1233 61 3 0 1 0 2 0  
1234 68 3 0 2 0 3 0  
1235 57 3 0 1 0 3 0  
1236 78 3 0 2 0 2 0  
1237 73 3 0 0 0 2 0  
1238 71 4 0 4 0 2 1  
1239 68 4 1 3 0 1 0  
1240 75 2 0 0 0 1 0  
1241 70 1 0 0 0 2 0  
1242 66 1 0 0 0 2 0  
1243 62 3 0 0 0 2 0  
1244 79 2 0 0 0 2 1  
1245 66 5 1 4 0 1 0  
1246 60 3 0 3 0 2 0  
1247 72 3 0 1 0 2 1  
1248 63 5 1 6 0 1 1  
1249 64 4 1 4 0 2 0  
1250 84 2 0 3 0 1 0  
1251 70 2 0 3 0 2 1  
1252 68 4 1 3 0 1 0  
1253 86 3 0 2 0 1 0  
1254 81 3 0 4 0 1 1  
1255 69 4 0 3 0 1 0  
1256 84 4 1 5 0 1 0  
1257 73 3 1 2 0 1 0  
1258 64 2 0 1 0 2 0  
1259 57 3 0 3 0 2 1  
1260 67 3 0 1 0 1 0  
1261 64 4 1 4 0 2 0  
1262 66 1 0 0 0 2 0  
1263 62 3 0 2 0 2 0  
1264 66 3 0 2 0 2 0  
1265 66 2 0 1 0 3 0  
1266 68 2 0 1 0 3 0  
1267 60 3 0 1 0 2 0  
1268 86 2 0 0 0 2 0  
1269 85 5 1 3 0 2 0  
1270 77 2 0 0 0 3 0  
1271 61 2 0 0 0 3 1  
1272 57 3 0 1 0 1 0  
1273 66 3 0 3 0 3 0

1274 61 4 0 2 0 3 0  
1275 82 3 0 4 0 3 0  
1276 73 3 0 1 0 3 1  
1277 64 4 1 4 0 1 1  
1278 90 4 0 4 0 1 0  
1279 83 4 1 5 0 2 0  
1280 73 3 0 1 0 2 0  
1281 69 3 0 3 0 2 0  
1282 63 3 1 3 0 2 0  
1283 60 3 0 1 0 1 0  
1284 80 3 0 1 0 2 0  
1285 74 2 0 0 0 2 0  
1286 81 3 0 1 0 2 0  
1287 72 2 0 1 0 2 0  
1288 75 4 1 5 0 1 0  
1289 73 3 0 1 0 4 0  
1290 67 4 0 4 0 4 0  
1291 65 3 0 2 0 2 0  
1292 66 2 0 0 0 2 1  
1293 80 5 1 6 0 1 0  
1294 81 3 0 2 0 1 0  
1295 73 4 1 6 0 2 0  
1296 67 4 0 2 0 2 0  
1297 62 2 0 0 0 2 0  
1298 65 3 0 1 0 2 1  
1299 83 2 0 2 0 1 0  
1300 70 3 0 1 0 2 0  
1301 76 3 0 2 0 2 0  
1302 68 4 1 3 0 3 1  
1303 63 5 1 6 0 3 1  
1304 69 3 0 3 0 1 0  
1305 63 3 0 3 0 2 0  
1306 68 3 0 1 0 4 0  
1307 87 3 0 2 0 1 0  
1308 75 3 0 2 0 1 0  
1309 74 4 0 1 0 1 0  
1310 64 1 0 0 0 2 0  
1311 79 2 0 2 0 2 0  
1312 72 2 0 0 0 2 0  
1313 66 3 0 2 0 1 0  
1314 63 2 0 1 0 1 0  
1315 58 3 0 0 0 1 0  
1316 68 4 0 2 0 1 1

1317 59 3 0 4 0 1 0  
1318 61 3 0 1 0 2 0  
1319 62 3 0 1 0 1 0  
1320 68 3 0 1 0 1 0  
1321 63 3 0 1 0 2 0  
1322 62 4 1 3 0 2 1  
1323 85 3 0 0 0 2 1  
1324 63 4 1 5 0 1 1  
1325 76 2 0 1 0 1 0  
1326 67 3 0 2 0 2 1  
1327 71 5 1 3 0 2 0  
1328 57 3 0 1 0 1 0  
1329 81 4 0 2 0 3 0  
1330 66 2 0 2 0 1 0  
1331 78 3 0 1 0 1 0  
1332 77 3 0 1 0 3 0  
1333 67 3 0 2 0 3 0  
1334 68 3 0 1 0 1 0  
1335 79 4 0 1 0 1 0  
1336 69 2 0 1 0 3 0  
1337 64 2 0 1 0 1 0  
1338 64 3 0 2 0 2 0  
1339 72 4 1 3 0 1 0  
1340 76 4 1 2 0 1 0  
1341 71 4 0 2 0 2 0  
1342 67 2 0 0 0 1 0  
1343 63 3 0 3 0 2 0  
1344 59 3 0 2 0 2 0  
1345 66 3 0 1 0 2 1  
1346 89 3 0 1 0 2 0  
1347 63 4 0 3 0 2 0  
1348 77 3 0 0 0 1 0  
1349 66 3 0 1 0 1 0  
1350 63 2 0 3 0 1 0  
1351 76 3 0 3 0 1 0  
1352 80 3 0 3 0 2 0  
1353 65 2 0 2 0 2 0  
1354 69 3 0 2 0 1 0  
1355 58 3 0 3 0 1 0  
1356 68 3 0 2 0 2 0  
1357 75 3 1 3 0 2 1  
1358 69 4 1 5 0 1 0  
1359 66 3 0 2 0 1 0



1360 64 2 0 2 0 3 0  
1361 59 3 0 2 0 3 0  
1362 65 4 0 4 0 2 0  
1363 68 4 0 4 0 2 0  
1364 63 4 0 1 0 4 0  
1365 62 3 0 0 0 4 0  
1366 66 2 0 0 0 3 0  
1367 68 2 0 0 0 3 0  
1368 79 3 0 1 0 2 0  
1369 75 2 0 0 0 2 0  
1370 68 2 0 0 0 2 0  
1371 73 3 0 0 0 2 0  
1372 64 3 0 1 0 1 0  
1373 72 3 0 2 0 1 0  
1374 73 4 0 2 0 1 0  
1375 67 3 0 1 0 2 0  
1376 58 3 0 1 0 1 0  
1377 63 3 0 2 0 1 0  
1378 69 3 0 2 0 1 0  
1379 72 3 0 2 0 1 0  
1380 67 3 0 3 0 1 0  
1381 72 2 0 0 0 3 1  
1382 64 3 0 2 0 1 0  
1383 62 3 0 1 0 3 0  
1384 70 2 0 1 0 3 0  
1385 64 3 0 2 0 3 0  
1386 70 2 0 1 0 1 0  
1387 71 3 0 0 0 1 0  
1388 79 3 0 2 0 2 0  
1389 75 3 0 2 0 2 0  
1390 67 4 1 3 0 2 0  
1391 64 3 0 2 0 2 0  
1392 70 4 0 3 0 1 0  
1393 78 3 1 2 0 4 0  
1394 63 3 0 2 0 4 1  
1395 62 3 1 4 0 1 1  
1396 67 3 0 2 0 2 1  
1397 71 5 1 1 0 1 0  
1398 71 2 0 1 0 2 0  
1399 65 3 0 2 0 2 0  
1400 64 4 1 5 0 1 0  
1401 73 4 0 4 0 1 0  
1402 68 4 0 1 0 1 0

1403 79 4 0 2 0 1 0  
1404 62 3 0 1 0 2 0  
1405 63 3 0 1 0 2 0  
1406 66 3 0 4 0 1 0  
1407 71 3 0 1 0 1 0  
1408 68 3 0 1 0 1 1  
1409 70 4 1 2 0 1 1  
1410 77 3 0 4 0 1 0  
1411 63 2 1 2 0 2 0  
1412 89 4 1 1 0 2 0  
1413 71 4 0 7 0 2 1  
1414 65 4 1 3 0 2 1  
1415 65 5 1 7 0 1 1  
1416 71 3 0 1 0 2 1  
1417 66 4 1 3 0 1 0  
1418 66 2 0 0 0 3 0  
1419 62 4 0 4 0 3 0  
1420 83 2 0 1 0 2 0  
1421 81 3 1 2 0 1 1  
1422 71 4 1 3 0 1 1  
1423 60 4 1 5 0 2 1  
1424 76 2 0 1 0 2 0  
1425 76 3 1 2 0 4 1  
1426 72 3 0 1 0 1 0  
1427 84 3 0 2 0 2 0  
1428 65 3 0 3 0 2 0  
1429 67 5 1 6 0 1 0  
1430 69 3 1 2 0 1 1  
1431 65 3 1 3 0 1 0  
1432 82 2 0 3 0 1 1  
1433 83 5 1 7 0 1 0  
1434 64 3 0 1 0 1 0  
1435 66 2 0 0 0 1 0  
1436 79 3 0 1 0 3 0  
1437 65 3 0 0 0 3 0  
1438 62 3 0 1 0 2 0  
1439 80 3 0 4 0 3 0  
1440 71 2 0 0 0 3 0  
1441 70 3 0 3 0 2 0  
1442 66 4 0 2 0 2 1  
1443 74 4 1 2 0 2 0  
1444 67 3 0 1 0 2 0  
1445 74 4 0 4 0 1 0

1446 76 3 1 1 0 2 0  
1447 75 3 0 0 0 2 0  
1448 73 2 0 2 0 2 0  
1449 84 2 0 1 0 2 0  
1450 81 2 0 0 0 2 0  
1451 75 2 0 1 0 2 0  
1452 72 2 0 0 0 3 0  
1453 77 2 0 0 0 3 0  
1454 68 4 0 0 0 1 0  
1455 76 3 0 1 0 1 0  
1456 75 4 0 2 0 1 0  
1457 67 5 0 4 0 1 0  
1458 71 3 0 0 0 1 1  
1459 81 4 0 2 0 2 1  
1460 76 4 1 3 0 1 0  
1461 79 4 0 2 0 2 0  
1462 66 4 0 2 0 2 0  
1463 85 4 0 2 0 1 0  
1464 69 2 0 2 0 1 0  
1465 66 3 0 1 0 3 0  
1466 71 3 0 2 0 1 1  
1467 59 4 0 2 0 1 0  
1468 68 4 1 2 0 1 0  
1469 75 3 0 2 0 1 1  
1470 72 3 0 2 0 1 0  
1471 73 5 1 4 0 1 1  
1472 63 3 1 4 0 2 1  
1473 62 4 0 2 0 1 0  
1474 66 3 0 2 0 1 0  
1475 73 2 0 1 0 1 1  
1476 66 4 1 2 0 1 1  
1477 60 4 1 5 0 1 1  
1478 71 3 1 4 0 2 1  
1479 72 3 0 2 0 2 0  
1480 70 3 0 1 0 1 0  
1481 68 2 0 0 0 1 0  
1482 61 4 1 5 0 1 0  
1483 59 4 1 2 0 1 0  
1484 70 2 0 0 0 2 0  
1485 86 2 0 1 0 2 0  
1486 77 3 0 2 0 3 0  
1487 68 2 0 0 0 3 0  
1488 65 2 0 1 0 2 0

1489 76 2 0 0 0 2 0  
1490 85 4 1 4 0 3 0  
1491 74 4 0 3 0 3 0  
1492 73 3 0 1 0 1 0  
1493 85 3 0 3 0 1 0  
1494 73 4 1 2 0 2 1  
1495 85 4 1 4 0 2 1  
1496 73 4 0 2 0 1 0  
1497 67 2 0 1 0 2 0  
1498 71 2 0 0 0 2 1  
1499 66 4 0 2 0 1 0  
1500 69 4 0 3 0 2 0  
1501 72 3 0 1 0 2 1  
1502 81 4 1 6 0 1 1  
1503 65 4 1 3 0 1 0  
1504 70 3 0 2 0 1 1  
1505 68 3 0 2 0 1 0  
1506 75 2 0 0 0 1 0  
1507 79 2 0 0 0 1 0  
1508 77 5 1 2 0 2 0  
1509 64 2 0 2 0 2 0  
1510 66 4 0 4 0 1 1  
1511 64 5 1 3 0 1 0  
1512 77 2 0 0 0 2 0  
1513 79 1 0 0 0 2 1  
1514 64 4 0 4 0 1 0  
1515 80 2 0 0 0 3 0  
1516 61 3 0 0 0 3 0  
1517 66 3 0 2 0 1 1  
1518 64 4 0 4 0 1 0  
1519 62 3 0 1 0 1 1  
1520 84 2 0 0 0 1 0  
1521 87 2 0 0 0 2 0  
1522 71 3 1 0 0 2 1  
1523 62 4 1 1 0 1 0  
1524 87 3 1 2 0 2 1  
1525 88 5 1 5 0 1 0  
1526 83 3 0 2 0 2 1  
1527 63 5 1 3 0 1 0  
1528 51 2 0 1 0 4 0  
1529 78 3 1 2 0 4 1  
1530 71 2 0 0 0 1 0  
1531 71 3 0 3 0 1 0

1532 69 3 0 1 0 3 0  
1533 67 3 0 2 0 2 0  
1534 64 3 0 3 0 1 0  
1535 71 1 0 0 0 1 0  
1536 63 4 1 2 0 1 0  
1537 59 3 0 0 0 1 1  
1538 73 4 1 8 0 1 0  
1539 73 3 0 1 0 1 0  
1540 67 3 0 2 0 3 1  
1541 60 2 0 1 0 3 0  
1542 64 2 0 0 0 1 0  
1543 67 3 1 1 0 1 0  
1544 85 4 1 4 0 2 0  
1545 62 5 1 4 0 1 0  
1546 56 3 0 1 0 1 0  
1547 69 3 0 3 0 2 0  
1548 64 3 0 2 0 2 0  
1549 64 4 1 2 0 3 0  
1550 56 3 0 1 0 3 0  
1551 66 2 0 2 0 2 0  
1552 60 2 0 3 0 2 0  
1553 63 3 0 2 0 2 0  
1554 76 2 0 1 0 2 0  
1555 83 2 0 0 0 2 0  
1556 64 4 1 8 0 1 0  
1557 85 4 0 2 0 1 1  
1558 63 3 0 1 0 2 0  
1559 59 3 0 1 0 2 0  
1560 62 4 1 3 0 1 1  
1561 90 4 1 0 0 1 0  
1562 67 3 0 3 0 1 0  
1563 85 3 0 0 0 2 0  
1564 77 3 0 0 0 2 0  
1565 75 4 0 3 0 2 0  
1566 85 1 0 1 0 2 0  
1567 63 2 0 0 0 2 0  
1568 63 2 0 1 0 3 0  
1569 70 2 0 0 0 3 0  
1570 66 2 0 0 0 2 0  
1571 55 3 0 2 0 2 0  
1572 63 3 0 2 0 2 0  
1573 57 2 0 2 0 3 0  
1574 65 2 0 1 0 3 0

1575 70 3 0 3 0 1 0  
1576 66 1 0 2 0 2 0  
1577 62 3 1 3 0 2 0  
1578 75 3 0 2 0 2 0  
1579 69 2 0 0 0 1 1  
1580 59 4 1 6 0 2 0  
1581 70 3 0 1 0 2 0  
1582 65 3 0 2 0 2 0  
1583 65 3 0 3 0 2 0  
1584 65 3 0 1 0 2 0  
1585 67 2 0 0 0 2 0  
1586 80 3 1 2 0 3 0  
1587 80 3 1 3 0 2 0  
1588 74 2 0 1 0 4 0  
1589 67 1 0 0 0 4 0  
1590 72 3 1 2 0 1 0  
1591 84 3 0 2 0 1 0  
1592 74 3 0 2 0 1 0  
1593 57 2 1 4 0 2 0  
1594 61 3 1 4 0 2 1  
1595 85 4 1 1 0 1 0  
1596 67 3 0 1 0 4 0  
1597 58 3 0 2 0 4 0  
1598 68 3 0 3 0 2 1  
1599 63 4 0 2 0 1 0  
1600 77 4 1 2 0 1 0  
1601 79 3 0 2 0 2 0  
1602 66 2 0 1 0 2 0  
1603 78 2 0 1 0 2 0  
1604 85 2 0 2 0 2 0  
1605 63 2 0 2 0 2 0  
1606 71 3 0 1 0 1 0  
1607 63 2 0 2 0 2 0  
1608 78 3 0 1 0 2 0  
1609 61 3 0 2 0 2 0  
1610 74 3 0 3 0 1 0  
1611 84 3 0 1 0 1 1  
1612 66 3 0 6 0 2 0  
1613 78 4 0 2 0 2 1  
1614 63 4 1 4 0 1 0  
1615 57 2 0 2 0 1 0  
1616 88 3 0 3 0 2 0  
1617 65 3 0 4 0 4 0

1618 66 2 0 2 0 4 0  
1619 64 4 0 5 0 1 0  
1620 67 3 0 1 0 2 0  
1621 70 2 0 1 0 2 0  
1622 83 3 1 2 0 1 0  
1623 67 4 1 3 0 1 0  
1624 71 4 1 4 0 1 0  
1625 71 3 0 2 0 1 0  
1626 61 3 0 1 0 1 0  
1627 61 2 0 0 0 2 1  
1628 70 3 1 3 0 2 1  
1629 71 5 0 2 0 1 0  
1630 62 4 0 3 0 2 0  
1631 73 3 0 1 0 2 0  
1632 65 4 0 1 0 1 0  
1633 70 4 0 2 0 1 1  
1634 68 4 0 3 0 2 0  
1635 62 3 0 2 0 2 0  
1636 63 3 0 3 0 1 0  
1637 57 3 1 3 0 1 0  
1638 68 3 0 1 0 2 0  
1639 76 2 0 0 0 2 0  
1640 69 3 0 0 0 1 1  
1641 58 4 0 2 0 1 1  
1642 67 3 0 1 0 2 0  
1643 78 4 1 5 0 2 0  
1644 83 4 1 3 0 1 0  
1645 67 4 0 3 0 1 1  
1646 71 4 1 7 0 1 0  
1647 65 2 0 1 0 2 0  
1648 63 2 0 1 0 2 0  
1649 87 4 1 4 0 1 1  
1650 64 4 1 6 0 1 0  
1651 59 3 1 3 0 2 0  
1652 90 2 0 0 0 2 0  
1653 71 3 0 2 0 1 1  
1654 67 4 1 3 0 2 1  
1655 73 4 1 4 0 2 1  
1656 68 5 1 4 0 2 0  
1657 66 1 0 0 0 3 0  
1658 62 2 0 0 0 3 1  
1659 86 4 1 4 0 3 1  
1660 55 4 0 2 0 3 0

1661 73 4 0 2 0 2 0  
1662 80 4 0 2 0 2 0  
1663 74 2 0 1 0 1 0  
1664 89 2 0 0 0 1 0  
1665 63 3 0 2 0 1 1  
1666 83 3 0 2 0 2 0  
1667 89 3 0 1 0 4 0  
1668 75 3 0 2 0 2 0  
1669 70 3 0 0 0 4 0  
1670 71 3 0 0 0 4 0  
1671 60 4 1 2 0 1 0  
1672 66 3 0 1 0 1 0  
1673 71 2 0 0 0 3 0  
1674 74 2 0 0 0 3 0  
1675 80 2 0 1 0 2 0  
1676 85 3 0 2 0 2 0  
1677 63 2 0 0 0 1 0  
1678 67 2 0 0 0 1 0  
1679 69 3 0 3 0 1 0  
1680 83 3 0 4 0 1 0  
1681 62 2 0 1 0 3 0  
1682 51 3 0 2 0 1 0  
1683 76 3 0 2 0 1 0  
1684 69 2 0 0 0 2 0  
1685 79 3 0 4 0 2 1  
1686 82 4 1 4 0 1 0  
1687 64 4 0 4 0 1 0  
1688 58 4 0 5 0 1 0  
1689 76 4 0 4 0 1 0  
1690 72 2 0 1 0 1 0  
1691 65 3 0 1 0 1 0  
1692 71 3 0 3 0 1 0  
1693 74 3 0 2 0 1 1  
1694 64 3 1 5 0 1 0  
1695 71 2 0 0 0 1 0  
1696 71 3 0 1 0 1 0  
1697 66 3 0 1 0 1 0  
1698 69 4 0 3 0 1 1  
1699 71 5 1 6 0 1 1  
1700 65 4 1 4 0 1 1  
1701 81 4 1 2 0 1 0  
1702 77 2 0 0 0 3 0  
1703 64 3 0 0 0 3 0



1704 49 3 0 1 0 1 0  
1705 67 3 1 3 0 1 0  
1706 75 3 0 2 0 2 0  
1707 68 4 0 2 0 2 1  
1708 59 3 0 1 0 2 0  
1709 73 3 0 1 0 2 0  
1710 70 4 1 4 0 2 0  
1711 63 4 0 4 0 2 0  
1712 65 4 0 2 0 2 0  
1713 55 2 0 1 0 2 0  
1714 65 3 0 2 0 2 0  
1715 60 3 0 1 0 2 0  
1716 72 3 1 3 0 2 0  
1717 80 5 1 4 0 2 1  
1718 80 3 0 2 0 2 1  
1719 66 5 1 4 0 2 0  
1720 66 4 1 2 0 2 0  
1721 65 3 0 2 0 3 0  
1722 68 3 0 3 0 3 0  
1723 60 2 0 0 0 3 1  
1724 79 3 0 0 0 2 1  
1725 60 4 0 3 0 1 0  
1726 63 3 0 2 0 1 0  
1727 69 4 0 2 0 2 0  
1728 70 2 0 0 0 2 0  
1729 74 3 0 2 0 2 0  
1730 77 4 1 3 0 2 0  
1731 80 5 1 2 0 2 0  
1732 54 2 0 0 0 2 0  
1733 64 3 0 1 0 2 0  
1734 70 2 0 1 0 1 0  
1735 85 2 0 2 0 1 0  
1736 82 3 0 1 0 1 0  
1737 77 2 0 2 0 3 0  
1738 56 2 0 0 0 3 0  
1739 65 3 0 1 0 2 0  
1740 70 3 0 1 0 3 0  
1741 74 5 0 4 0 1 1  
1742 77 4 1 3 0 1 1  
1743 67 4 1 3 0 1 1  
1744 63 2 0 1 0 1 1  
1745 68 3 0 2 0 1 0  
1746 79 4 0 4 0 1 0

1747 74 4 0 4 0 1 0  
1748 79 4 1 4 0 2 0  
1749 65 4 0 3 0 2 1  
1750 65 4 1 4 0 1 0  
1751 75 3 0 1 0 2 1  
1752 74 5 1 2 1 2 1  
1753 65 5 1 5 1 1 1  
1754 63 4 1 5 1 1 0  
1755 68 3 0 1 1 3 0  
1756 77 5 1 4 1 3 1  
1757 61 5 1 6 1 2 1  
1758 57 5 1 3 1 1 1  
1759 74 4 1 3 1 1 0  
1760 79 4 1 3 1 2 1  
1761 82 5 1 10 1 2 1  
1762 61 4 1 3 1 4 1  
1763 65 5 1 4 1 4 0  
1764 61 4 1 2 1 1 1  
1765 49 5 1 5 1 2 1  
1766 70 4 1 5 1 1 1  
1767 77 4 1 3 1 2 0  
1768 62 5 1 8 1 2 0  
1769 65 3 0 2 1 2 1  
1770 81 4 1 3 1 1 1  
1771 70 5 1 6 1 1 1  
1772 73 5 1 4 1 1 1  
1773 71 3 0 3 1 2 0  
1774 68 1 0 1 1 2 1  
1775 59 5 1 3 1 1 0  
1776 79 5 1 5 1 1 0  
1777 75 1 0 2 1 2 0  
1778 66 5 0 5 1 1 0  
1779 73 3 1 1 1 1 1  
1780 76 4 0 4 1 1 0  
1781 69 4 0 1 1 2 1  
1782 76 5 1 5 1 1 0  
1783 88 2 1 1 1 3 1  
1784 84 4 0 3 1 1 1  
1785 66 5 1 4 1 1 1  
1786 75 5 1 5 1 1 1  
1787 91 3 0 0 1 1 1  
1788 64 4 1 3 1 2 1  
1789 64 4 1 2 1 2 1

1790 71 5 1 3 1 1 1  
1791 89 4 1 3 1 1 1  
1792 92 5 1 1 1 2 0  
1793 89 4 0 1 1 2 1  
1794 85 4 1 9 1 1 1  
1795 78 3 0 2 1 1 0  
1796 77 5 1 4 1 1 1  
1797 83 5 1 6 1 1 1  
1798 87 5 1 5 1 1 0  
1799 69 4 1 2 1 1 1  
1800 89 4 1 3 1 2 1  
1801 87 4 1 4 1 1 1  
1802 85 4 1 3 1 1 1  
1803 80 4 1 5 1 1 0  
1804 63 5 1 3 1 2 0  
1805 92 4 1 3 1 2 0  
1806 81 5 1 2 1 2 1  
1807 67 5 1 4 1 2 1  
1808 66 5 1 4 1 1 1  
1809 80 3 0 2 1 2 1  
1810 83 3 0 5 1 2 1  
1811 68 5 1 5 1 1 1  
1812 80 4 1 4 1 1 0  
1813 81 3 1 3 1 1 1  
1814 80 5 1 10 1 1 0  
1815 86 4 1 4 1 1 1  
1816 91 5 1 5 1 1 0  
1817 65 5 1 4 1 1 1  
1818 79 5 1 5 1 1 1  
1819 69 5 1 5 1 1 1  
1820 67 5 1 4 1 4 1  
1821 93 4 1 5 1 1 1  
1822 86 4 1 2 1 1 0  
1823 75 5 1 5 1 1 1  
1824 93 4 1 2 1 2 0  
1825 57 5 1 1 1 1 0  
1826 89 4 1 4 1 1 0  
1827 88 4 1 4 1 1 1  
1828 78 5 1 5 1 2 1  
1829 84 5 1 6 1 2 0  
1830 76 4 0 3 1 1 1  
1831 85 5 1 5 1 1 0  
1832 87 4 1 4 1 1 1

1833 79 5 1 3 1 1 0  
1834 78 5 1 4 1 2 1  
1835 83 4 1 3 1 2 1  
1836 78 4 1 7 1 1 1  
1837 78 5 1 4 1 2 0  
1838 92 3 1 3 1 2 0  
1839 75 4 1 3 1 2 1  
1840 87 5 1 5 1 1 0  
1841 88 5 1 4 1 1 0  
1842 83 4 1 1 1 1 1  
1843 75 5 1 3 1 1 0  
1844 89 3 0 1 1 2 1  
1845 83 5 1 4 1 2 1  
1846 92 4 1 4 1 1 1  
1847 88 3 0 3 1 1 0  
1848 92 4 0 2 1 1 1  
1849 85 4 1 3 1 1 0  
1850 63 4 1 2 1 2 1  
1851 87 4 1 4 1 1 0  
1852 88 5 1 7 1 1 1  
1853 89 4 1 1 1 2 0  
1854 81 3 1 3 1 2 0  
1855 74 5 1 4 1 3 1  
1856 92 5 1 4 1 2 1  
1857 86 5 1 2 1 2 1  
1858 87 5 1 2 1 1 1  
1859 92 4 1 3 1 2 1  
1860 87 4 1 11 1 1 1  
1861 90 5 1 4 1 1 1  
1862 84 5 1 4 1 1 0  
1863 61 5 1 3 1 1 1  
1864 84 4 1 5 1 1 1  
1865 79 3 1 3 1 1 1  
1866 81 5 1 3 1 1 1  
1867 85 4 1 4 1 2 0  
1868 82 5 1 3 1 2 0  
1869 65 5 1 5 1 1 0  
1870 81 5 1 4 1 2 1  
1871 80 5 1 8 1 1 1  
1872 90 5 1 7 1 2 1  
1873 76 5 1 4 1 2 1  
1874 82 3 0 1 1 1 1  
1875 67 5 0 4 1 2 1

```

1876 76 5 1 6 1 1 1
1877 69 5 1 5 1 1 1
1878 82 5 1 4 1 1 0
1879 88 4 1 4 1 1 1
> Y<-as.numeric(my$Y)
> X3<-as.numeric(my$X3)
> X4<-as.factor(my$X4)
> X5<-as.factor(my$X5)
> X6<-as.numeric(my$X6)
> X7<-as.factor(my$X7)
> X8<-as.factor(my$X8)
> model1=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="logit"))
> anova(model1)

```

Analysis of Deviance Table

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1878	1994.2
X3 1	10.114		1877	1984.1
X4 4	43.919		1873	1940.2
X5 1	4.874		1872	1935.3
X6 1	1.589		1871	1933.8
X7 1	72.821		1870	1860.9
X8 3	0.875		1867	1860.1

>

```
> summary(model1)
```

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5970	-0.6842	-0.6371	-0.5281	2.0834

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.254471	0.689537	-1.819	0.0689 .

X3	0.002192	0.007084	0.309	0.7570
X42	-0.688585	0.487132	-1.414	0.1575
X43	-0.383052	0.474917	-0.807	0.4199
X44	-0.295662	0.495075	-0.597	0.5504
X45	-0.442704	0.544857	-0.813	0.4165
X51	0.087844	0.161274	0.545	0.5860
X6	0.048985	0.047988	1.021	0.3074
X71	1.891622	0.229875	8.229	<2e-16 ***
X82	-0.085661	0.126976	-0.675	0.4999
X83	-0.131577	0.217341	-0.605	0.5449
X84	-0.248444	0.397841	-0.624	0.5323

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1994.2 on 1878 degrees of freedom  
 Residual deviance: 1860.1 on 1867 degrees of freedom  
 AIC: 1884.1

Number of Fisher Scoring iterations: 4

```

> X3<-as.numeric(my$X3)
> X4<-as.factor(my$X4)
> X5<-as.factor(my$X5)
> X6<-as.numeric(my$X6)
> X7<-as.factor(my$X7)
> X8<-as.factor(my$X8)
> set.seed(1234)
> ind<-sample(2,nrow(my),replace=T,prob=c(0.8,0.2))
> trainData<-my[ind==1,]
> testData<-my[ind==2,]
> model2=glm(formula=Y~X3+X4+X5+X6+X7+X8,family=binomial(link="logit"),data=trainData)
> anova(model2)

```

Analysis of Deviance Table

Model: binomial, link: logit

Response: Y

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid.	Dev
NULL				1516		1601.8
X3	1	8.423		1515		1593.4
X4	1	37.464		1514		1555.9
X5	1	4.678		1513		1551.3
X6	1	0.142		1512		1551.1
X7	1	52.980		1511		1498.1
X8	1	2.468		1510		1495.7

> summary(model2)

Call:

```
glm(formula = Y ~ X3 + X4 + X5 + X6 + X7 + X8, family = binomial(link = "logit"),
    data = trainData)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5716	-0.6695	-0.6148	-0.5322	2.0766

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.883939	0.673686	-2.796	0.00517 **
X3	0.002113	0.007880	0.268	0.78860
X4	0.145434	0.106456	1.366	0.17190
X5	0.101746	0.176744	0.576	0.56484
X6	0.018368	0.053914	0.341	0.73333
X7	1.742186	0.244243	7.133	9.82e-13 ***
X8	-0.148538	0.095394	-1.557	0.11945

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1601.8 on 1516 degrees of freedom  
 Residual deviance: 1495.7 on 1510 degrees of freedom  
 AIC: 1509.7

Number of Fisher Scoring iterations: 4

```
> pred_Y=predict(model2,trainData,type="response")
> roc1=roc(response=trainData$Y,predictor=pred_Y)
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> pdf("rplot.pdf")
```

```
> plot.roc(roc1)
> dev.off()
null device
  1
> auc(roc1)
Area under the curve: 0.6441
```

Ανοίγω το αρχείο pdf το οποίο βρίσκεται στο φάκελο που όρισα το αρχείο csv των δεδομένων και το οποίο έκανα εισαγωγή στην R , και αντιγράφω και επικολλώ τη γραφική παράσταση.

#### **ΔΙΕΥΚΡΙΝΗΣΗ**

Οι γραφικές παραστάσεις και οι πιθανότητες εκτιμήθηκαν με τη βοήθεια του λογισμικού mathematica.

<<Δηλώνω ρητά ότι , σύμφωνα με το άρθρο 8 του Ν.1599/1968 η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας ,προσωπικότητας και προσωπικών δεδομένων τρίτων,δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών /δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής ,οι πηγές δε που χρησιμοποιούνται περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.>>

ΤΣΙΑΡΑ ΘΕΟΧΑΡΙΑ



