ΕΡΓΑΣΙΑ ΣΤΟ ΠΙΣΤΩΤΙΚΟ ΚΙΝΔΥΝΟ

Διδάσκων : Μ. Κούτρας

****

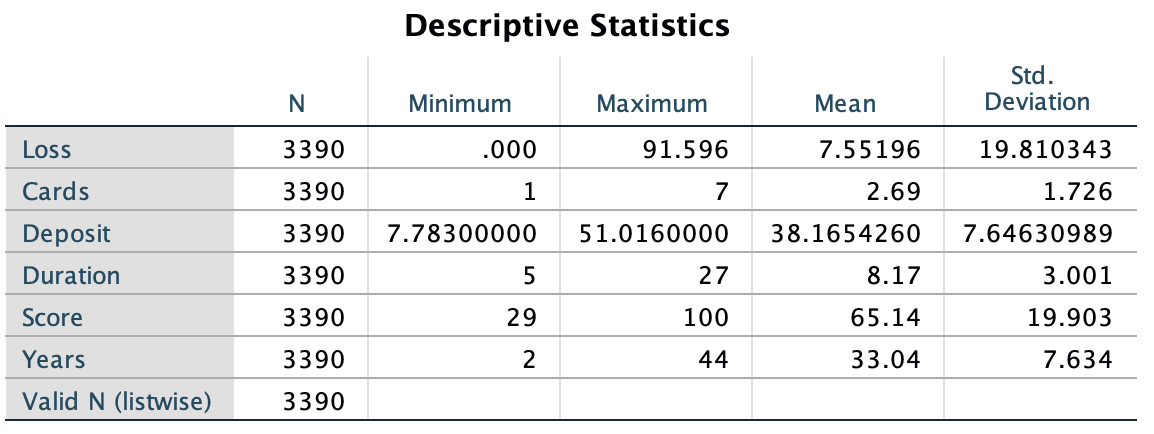
ΕΜΜΑΝΟΥΗΛ ΒΕΡΥΚΟΚΙΔΗΣ

ΜΑΕ19010

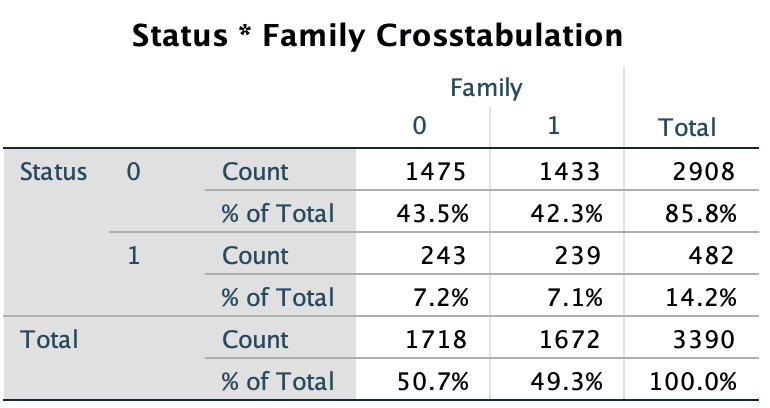
MSc: Risk Management & Actuarial Science

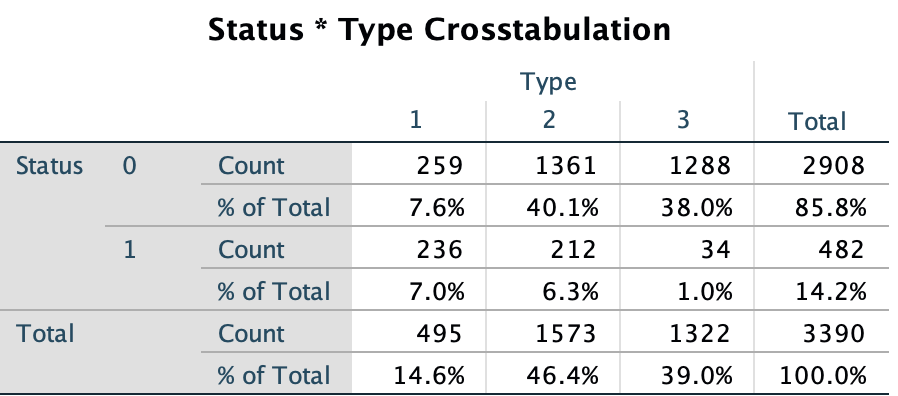
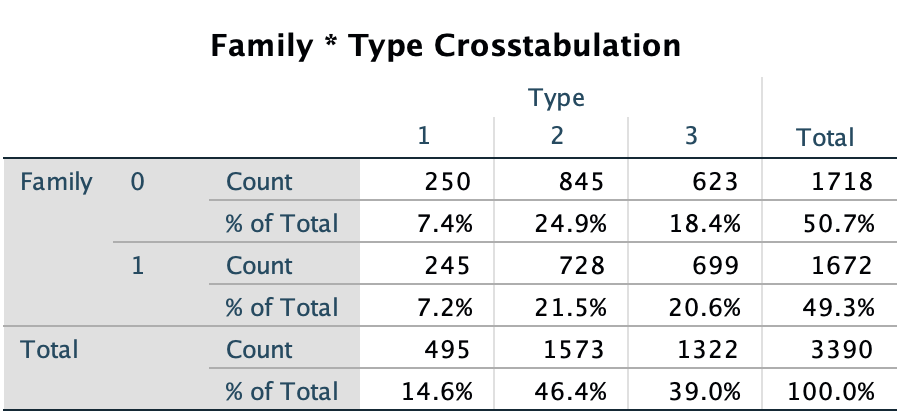
**QUESTION A.**

Για τα ποσοτικά χαρακτηριστικά έχουμε τα παρακάτω αποτελέσματα :



Για τις κατηγορικές μεταβλητές (status, family, type) έχουμε :





Οι παραπάνω πίνακες διπλής εισόδου εμφανίζουν τις συχνότητες ,που εμφανίζονται οι παρατηρήσεις, αλλά και τα ποσοστά τους ως προς το σύνολο όλων των παρατηρήσεων (3390).

**ΣΗΜΑΝΤΙΚΟ :**

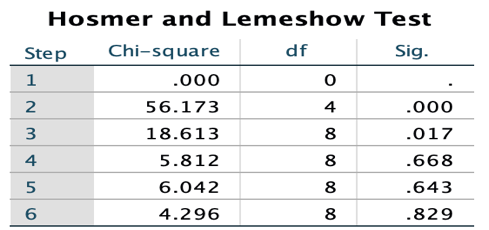
Πριν προχωρήσουμε στην ερώτηση Β θα πρέπει να χωρίσουμε την κατηγορική μεταβλητή Type, η οποία παίρνει 3 τιμές ανάλογα με το είδος εργασίας, σε 2 νέες δείκτριες τις οποίες εμείς θα ονομάσουμε type01 και type02.

Η type01 παίρνει την τιμή 1 αν έχουμε εποχική εργασία , η type02 παίρνει την τιμή 1 αν έχουμε σύμβαση αορίστου χρόνου και αν είναι και οι 2 0 έχουμε σύμβαση ορισμένου χρόνου.

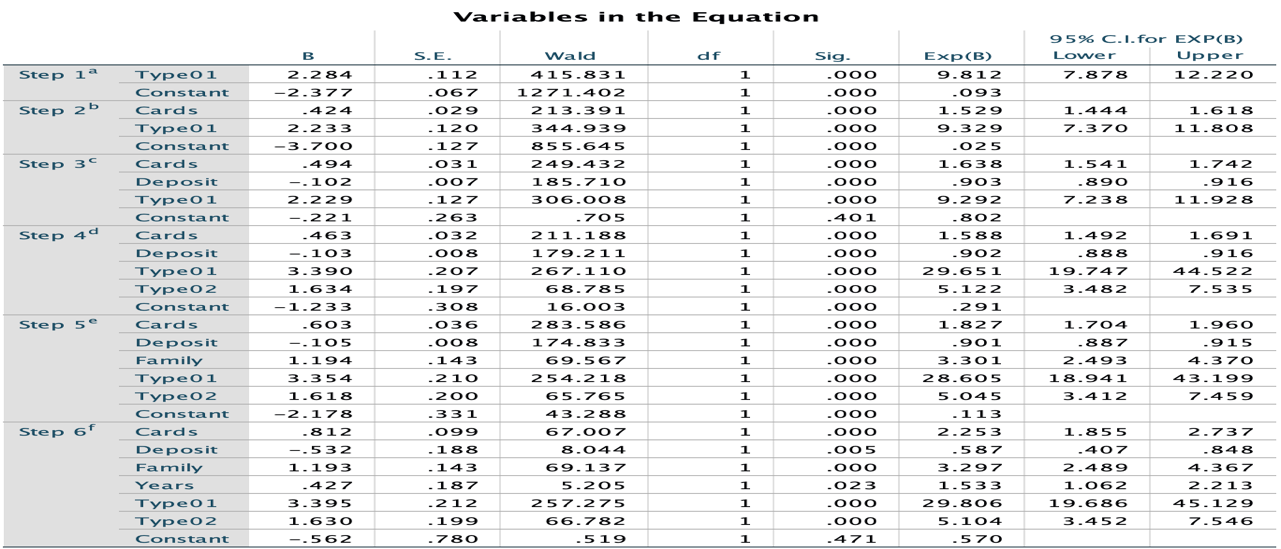
**QUESTION B**

**Υπό-ερώτημα 1.**

Αρχικά θα τρέξουμε το Hosmer-Lemeshow test ώστε να γίνει ο έλεγχος για την καταλληλότητα του μοντέλου.



Από το παραπάνω test συμπεραίνουμε ότι η μηδενική υπόθεση Ho δεν απορρίπτεται άρα το μοντέλο έχει καλή προσαρμογή στα δεδομένα.



Στον παραπάνω πίνακα παρατηρούμε στο 6ο βήμα τις μεταβλητές που είναι στατιστικά σημαντικές για το μοντέλο μας.

Έχουμε δηλαδή τις :

**1.Cards**

**2.Deposit**

**3.Family**

**4.Years**

**5.Type01**

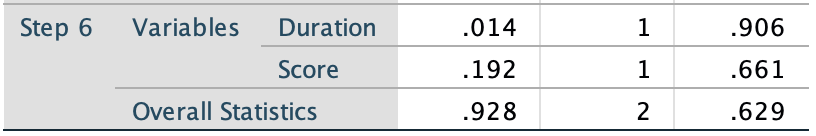
**6.Type02**

Από την άλλη οι μεταβλητές που δεν χρειάζονται είναι οι :

**1.Duration**

**2.Score**

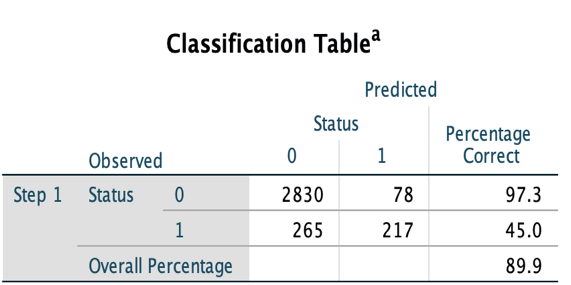
Όπως φαίνονται και στον παρακάτω πίνακα **(Variables not in the equation)** :

****

Από την πρώτη στήλη του πίνακα **Variables in the equation** παίρνουμε τους συντελεστές β για την κατασκευή του μοντέλου παλινδρόμησης , έτσι έχουμε :

**Ln[p/(1-p)] = - 0,562 + 0,812\*Cards – 0,532\*Deposit + 1,193\*Family + 0,427\*Years +3,395\*Type01 + 1,63\*Type02**

**Υπό-ερώτημα 2**

****

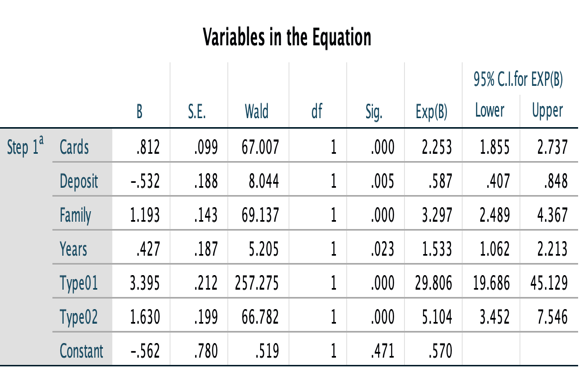
Για το μοντέλο μας ο πίνακας ορθών ταξινομήσεων είναι ο παραπάνω , ο οποίος μα δείχνει ότι οι καλοί πελάτες ταξινομούνται σωστά με ποσοστό 97,3% και οι κακοί με ποσοστό 45%.

Αν θέλουμε το ποσοστό ορθών ταξινομήσεων των κακών πελατών να αυξηθεί πάνω από 85% πρέπει να μειώσουμε το σημείο αποκοπής cut-off από 0,1 και κάτω.

Το μοντέλο μας δεν θα αλλάξει με την αλλαγή του cut-off έτσι θα έχουμε :

**1-p =**

**Υπό-ερώτημα 3**

****

Γνωρίζουμε ότι **odds’ = exp(βi)\*odds** άρα αν αυξηθούν οι κάρτες κατά 1 η σχετική πιθανότητα θα είναι 2,253 φορές η προηγούμενη σχετική πιθανότητα επομένως θα έχουμε μια αύξηση κατά 125,3%.

Ομοίως αν αυξηθούν τα δάνεια κατά 1000 ευρώ η σχετική πιθανότητα θα μειωθεί κατά 41,5%.

**Υπό-ερώτημα 4**

Δεδομένα:

**Years = 40**

**Cards = 5**

**Deposit = 20**

**Duration = 15**

**Family = 0**

**Type01=0**

**Type02=0**

**Score = 80**

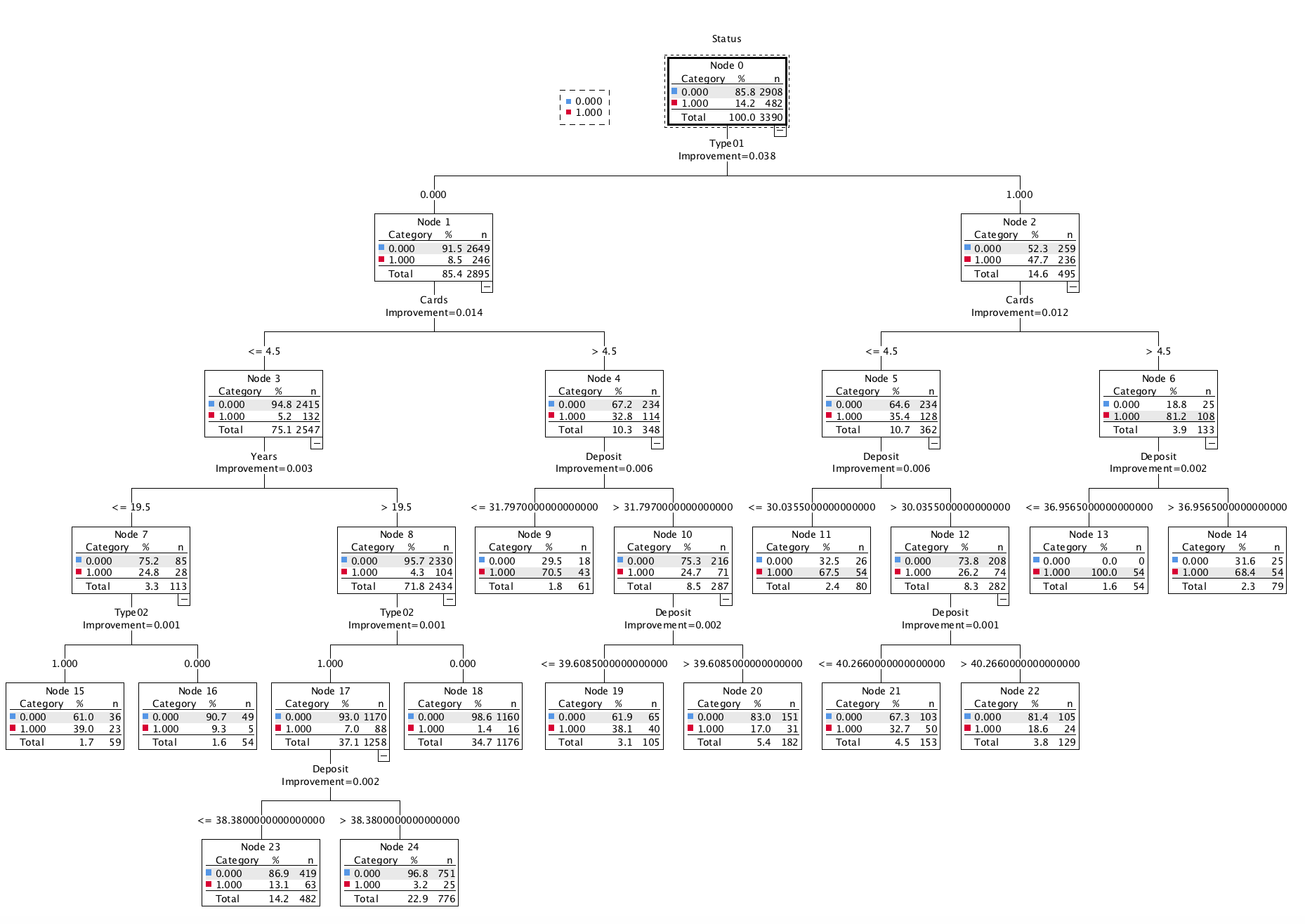
Χρησιμοποιόντας τον τύπο **p = exp(z)/(1+exp(z))**.Η πιθανότητα χρεοκοπίας του συγκεκριμένου ατόμου είναι 99,995% επομένως δεν θα του χορηγούσαμε δάνειο.

**Υπό-ερώτημα 5**

Έχουμε 0,1=exp(z)/(1+exp(z)) => exp(z)=0,11. Από το μοντέλο παλινδρόμησης έχουμε ότι αν κάποιος είναι έγγαμος το z αυξάνεται κατά 1,193 μονάδες άρα έχουμε :

p` = exp(z)\*exp(1,193)/(1+exp(z)\*exp(1,193)) => p’=0,266. Άρα η πιθανότητα αθέτησης είναι 26,6%.

**QUESTION Γ**

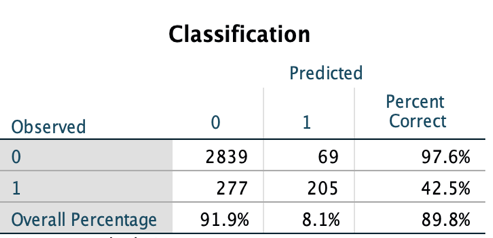


**Υπό-ερώτημα 1**

Σύμφωνα με το παραπάνω μοντέλο οι μεταβλητές που χρειάζονται είναι :

**Type01,Cards,Years,Deposit,Type02**

Από κάτω παρουσιάζεται ο πίνακας ορθών ταξινομήσεων :

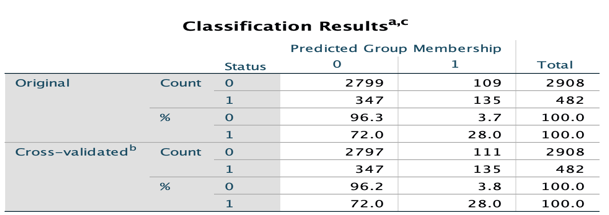
****

**Υπό-ερώτημα 2**

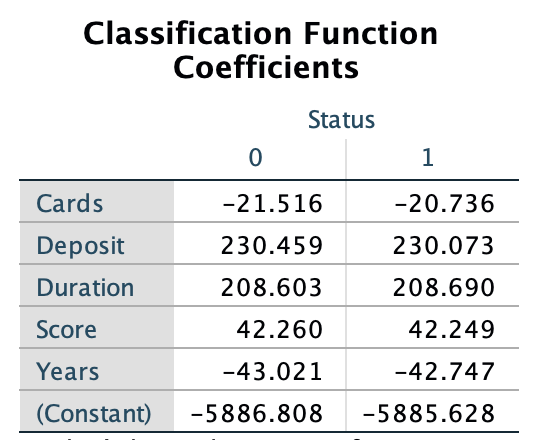
Από το δέντρο έχουμε ότι η πιθανότητα να αθετήσει ο συγκεκριμένος πελάτης ,του ερωτήματος Β4, είναι 70,5%. Επομένως δεν του χορηγούμε δάνειο.

**QUESTION Δ**

**Υπό-ερώτημα 1**

****Παρακάτω παρουσιάζεται ο πίνακας ορθών ταξινομήσεων για το μοντέλο διαχωριστικής ανάλυσης.

Επισημαίνεται ότι χρησιμοποιήσαμε την επιλογή **Compute from group sizes** ώστε οι πιθανότητες να υπολογιστούν βάση τα ποσοστά καλών και κακών πελατών από τα δεδομένα.

****

Σύμφωνα με τον πίνακα οι γραμμικές εξισώσεις είναι:

* Για τους καλούς πελάτες : **Yo = -5886,808 - 21,516\*Cards + 230,459\*Deposit + 208,603\*Duration + 42,260\*score - 43,021\*Years**
* Για τους κακούς πελάτες : **Υ1= -5885,628 - 20,736\*Cards + 230,073\*Deposit + 208,690 \*Duration + 42,249\*Score – 42,747\*Years**

**Υπό-ερώτημα 2**

Για τα δεδομένα του Β4 έχουμε :

Υο= -5886,808 – 21,516\*5 + 230,459\*20 + 208,603\*15 + 42,260\*80 – 43,021\*40 => **Υο= 3.403,797**

Υ1= -5885,628 – 20,736\*5 + 230,073\*20 + 208,690\*15 + 42,249\*80 – 42,747\*40 => **Υ1= 3.412,542**

Επειδή Υ1 > Υο δεν εγκρίνεται το δάνειο για τον συγκεκριμένο υποψήφιο δανειολήπτη.

**QUESTION E**

**Λογιστική Παλινδρόμηση :**

**Πλεονεκτήματα**

* Οι εκτιµήσεις της πιθανότητας βρίσκονται αυστηρά στο διάστηµα [0,1].
* Τόσο οι ανεξάρτητες µεταβλητές όσο και τα κατάλοιπα της παλινδρόµησης δεν είναι απαραίτητο να κατανέµονται κανονικά.
* Δεν κρίνεται ως απαραίτητη προϋπόθεση η ύπαρξη γραµµικής σχέσης ανάµεσα στην εξαρτηµένη και στις ανεξάρτητες µεταβλητές.
* Χρησιμοποιούνται τόσο ποσοτικές όσο και κατηγορικές μεταβλητές.

**Μειονεκτήματα**

* Προϋποθέτει την µη ύπαρξη πολυσυγκραµµικότητας στις ανεξάρτητες µεταβλητές.

**Δέντρο Ταξινόμησης :**

**Πλεονεκτήματα**

* Αν και η όλη διαδικασία για την κατασκευή του δέντρου είναι αλγοριθμικά πολύπλοκη, τα δέντρα που σχηματίζει είναι απλά και κατανοητά.
* Μπορεί να αντεπεξέλθει αποτελεσματικά σε μεγάλο αριθμό δεδομένων και πολυδιάστατα προβλήματα και μάλιστα χρησιμοποιώντας ένα μικρό αριθμό μεταβλητών.
* Οι μεταβλητές μπορούν να είναι κατηγορικές και ποσοτικές.

**Μειονεκτήματα**

* Για κάθε διαχωρισμό και κάθε υποδιαίρεση, χρησιμοποιείται μόνο ένα κριτήριο.
* Για να εφαρμοστεί επιτυχώς χρειάζεται μεγάλος αριθμός παρατηρήσεων.
* Σε σχέση με άλλες στατιστικές και αναλυτικές τεχνικές, αυτό είναι ένα «αμβλύ» όργανο.

**Διαχωριστική Ανάλυση:**

**Πλεονεκτήματα**

* Η διαχωριστική ανάλυση στηρίζεται σε πιο ρεαλιστικές μεθόδους και υπολογιστικά είναι πιο εύκολη.

**Μειονεκτήματα**

* Δεν χρησιμοποιεί κατηγορικές μεταβλητές
* Κάνει υπόθεση για τις ανεξάρτητες μεταβλητές.
* Γίνονται περίπλοκες υποθέσεις.

**QUESTION ΣΤ**

**Υπό-ερώτημα 1**

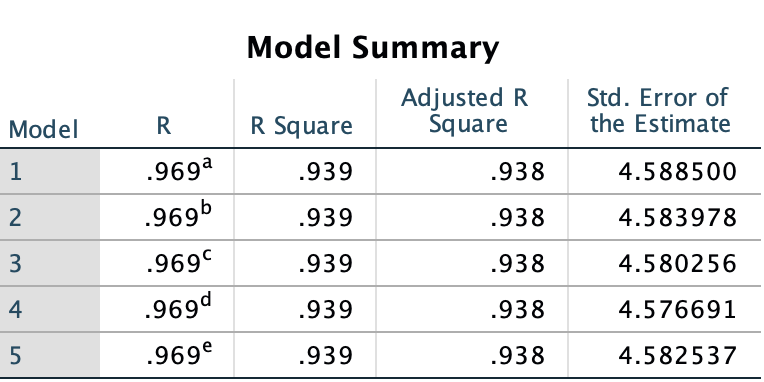
Τρέχουμε την Backward method και στον παρακάτω πίνακα (Βήμα 5ο) παίρνουμε τα εξής αποτελέσματα για τη σημαντικότητα των μεταβλητών :





Η μέθοδος backward αφού αφαιρέσει τις μεταβλητές οι οποίες δεν είναι χρήσιμες , στα βήματα 1-4 , καταλήγει στις παραπάνω χρήσιμες και το μοντέλο μας είναι :

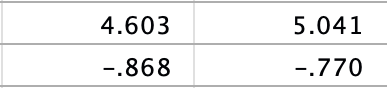
Loss = 69,878 + 4,822\*Cards – 0,819\*Deposit – 17,956\*Family + 0,913\*Type01



Στον πίνακα Model Summary στη τρίτη στήλη R square μπορούμε να δούμε την ποιότητα του μοντέλου η όποια από ότι φαίνεται και στον πίνακα είναι αρκετά καλή.

**Υπό-ερώτημα 2**

****

****

Στον παραπάνω πίνακα η πρώτη γραμμή αναφέρεται στη μεταβλητή cards ενώ η δεύτερη στη μεταβλητή deposit και δείχνουν τα διαστήματα εμπιστοσύνης ( 95%).

Από τα διαστήματα εμπιστοσύνης καταλαβαίνουμε τη μέγιστη και την ελάχιστη μεταβολή της ζημίας ,με 95% βεβαιότητα, αν μεταβληθούν οι κάρτες κατά 1 η οι καταθέσεις κατά 1000 ευρώ.

Μας παρουσιάζει τα δηλαδή τα ανώτατα και τα κατώτατα όρια.

Παράδειγμα αν αυξηθούν οι κάρτες κατά 1 η ζημία θα αυξηθεί σύμφωνα με το υπόδειγμά μας κατά 4,822 . Ο συγκεκριμένος συντελεστής μπορεί να μεταβάλλεται από 4,603 έως 5,041. Οι συγκεκριμένες τιμές αποτελούν ακρότατα για το συγκεκριμένο συντελεστή με πιθανότητα λάθους 5%.

Το ίδιο ισχύει και με τις καταθέσεις αν αυξηθούν κατά 1000 ευρώ η ζημιά θα μειωθεί κατά 0,819 με κατώτατο όριο το -0,868 και ανώτατο το -0,770 με πιθανότητα 95% όπως βλέπουμε και στον τελευταίο πίνακα.

**Υπό-ερώτημα 3**

**Αν ο εργαζόμενος είναι εποχικός η ζημιά θα είναι :**

Loss1 = 69,878 + 4,822\*Cards – 0,819\*Deposit – 17,956\*Family + 0,913\*1

**Αν ο εργαζόμενος είναι με σύμβαση ορισμένου χρόνου η ζημία θα είναι :**

Loss2 = 69,878 + 4,822\*Cards – 0,819\*Deposit – 17,956\*Family + 0,913\*0

Επομένως η ζημία θα **μειωθεί** κατά Loss1-Loss2 = 0,913

Αυτή η **μείωση** σύμφωνα με το διάστημα εμπιστοσύνης μπορεί να κυμανθεί από **0,091** έως **1,735** **με πιθανότητα 95%.** Δηλαδή υπάρχει 5% αβεβαιότητα .

**Υπό-ερώτημα 4**

**Αν ο υποψήφιος πελάτης είναι έγγαμος τότε family = 1 οπότε έχουμε :**

Loss1= 69,878 + 4,822\*Cards – 0,819\*Deposit – 17,956\*1 + 0,913\*Type01

**Αν ο υποψήφιος πελάτης είναι άγαμος τότε family = 0 οπότε έχουμε :**

Loss = 69,878 + 4,822\*Cards – 0,819\*Deposit – 17,956\*0 + 0,913\*Type01

Επομένως η ζημία **θα αυξηθεί** κατά Loss1-Loss2 = 17,956

Αυτή η αύξηση σύμφωνα με το διάστημα εμπιστοσύνης δεν μπορεί να είναι μεγαλύτερη από **18,922** και μικρότερη από **16,990** **με πιθανότητα σφάλματος 5%.**

**Υπό-ερώτημα 5**

Έχουμε τα δεδομένα :

**Years = 40**

**Cards = 5**

**Deposit = 20**

**Duration = 15**

**Family = 0**

**Type01=0**

**Score = 80**

Οπότε το μοντέλο μας διαμορφώνεται ως εξής :

Loss = 69,878 + 4,822\*5 – 0,819\*20 – 17,956\*0 + 0,913\*0 => **Loss = 77,608**

Άρα η ζημιά σε περίπτωση χρεοκοπίας θα είναι 77,608 με 95% βεβαιότητα και θα κυμαίνεται μεταξύ των διαστημάτων **68,55686 έως 86,66637**.

Η μέση ζημία για τα άτομα με τα παραπάνω χαρακτηριστικά θα είναι από **76,65851** έως **78,56472** με 95% βεβαιότητα άρα κατά μέσο όρο θα είναι 77,611.

**Υπό-ερώτημα 6**

Η αναμενόμενη ζημιά θα είναι η ζημία Loss που βγαίνει από την εξίσωση επί την πιθανότητα χρεοκοπίας που βρήκαμε στο ερώτημα Β4.

Επομένως έχουμε : 0,9995 \* (Loss) = 0,9995 \* (69,878 + 4,822\*5 – 0,819\*20 – 17,956\*0 + 0,913\*0) = 0,9995 \* 77,608 = **77,569196**

**Υπό-ερώτημα 7**

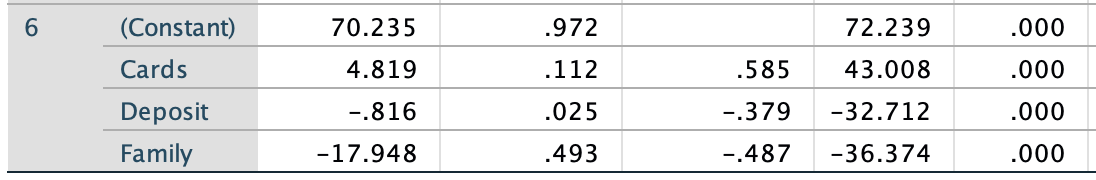
Αν θεωρήσουμε ότι το περιθώριο σφάλματος είναι 1% οι μεταβλητές οι οποίες παραμένουν στο μοντέλο μας είναι εκείνες οι οποίες έχουν p-value μικρότερο από το 0,01.

Οπότε έχουμε τις εξής υποθέσεις:

Ηο : η ζημία να μην επηρεάζεται αν ο δανειολήπτης είναι εργαζόμενος με σύμβαση ορισμένου η αορίστου χρόνου δηλαδή πρέπει ο συντελεστής του Type02 να είναι 0

Η1 : ο συντελεστής του Type02 να είναι διάφορος του 0

Αν ρυθμίσουμε την παλινδρόμηση έτσι ώστε να αφαιρεί τις μεταβλητές οι οποίες έχουν p-value > 0.01 θα καταλήξουμε στον παρακάτω πίνακα



Βλέπουμε λοιπόν ότι δεν υπάρχει η μεταβλητή type02 γιατί το sig. ήταν μεγαλύτερο από το 0,01. Επομένως το μοντέλο μας δεν εξαρτάται από την τιμή της .

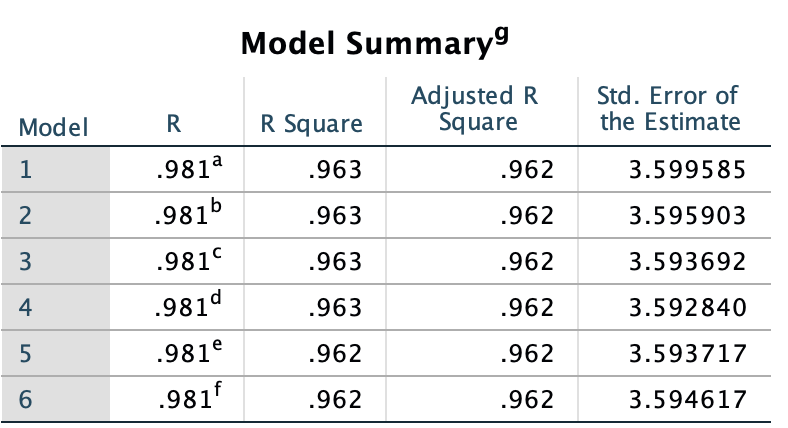
Άρα ισχύει η Ho δηλαδή το στέλεχος της τράπεζας έχει δίκιο.

**Υπό-ερώτημα 8**

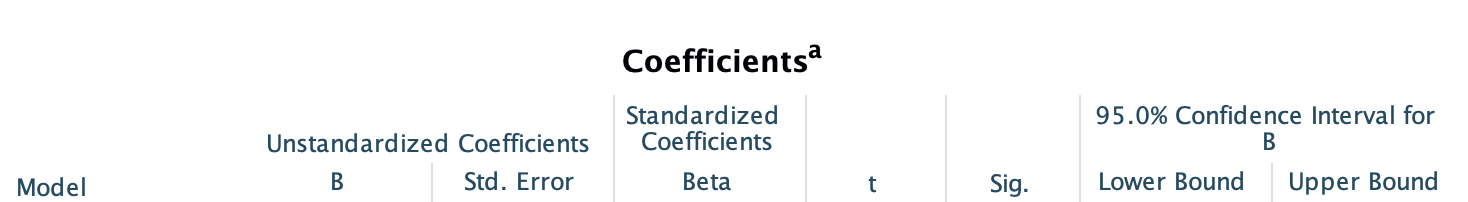
Για το συγκεκριμένο ερώτημα φτιάξαμε 2 νέες μεταβλητές ώστε να βρούμε τους όρους αλληλεπίδρασης μεταξύ του family και των deposit – cards οι οποίες ονομάστηκαν famdep & famcards.

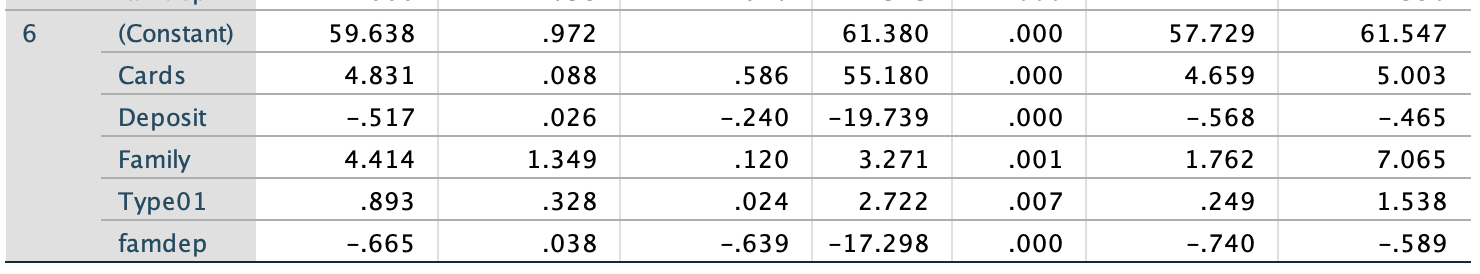
Έτσι τρέξαμε τη μέθοδο backward και είχαμε τα εξής αποτελέσματα :

Για την ποιότητα του δείγματος έγινε έλεγχος του R square



Ενώ για περιθώριο σφάλματος 5% το μοντέλο μας είναι :





Δηλαδή : Loss = 59,638 + 4,831\*Cards – 0,517\*Deposit + 4,414\*Family + 0,893\*Type01 – 0.665\*family\*deposit

Άρα για έναν έγγαμο έχουμε :

Loss = 59,638 + 4,831\*Cards – 0,517\*Deposit + 4,414\*1 + 0,893\*Type01 – 0.665\*1\*depοsit

Και για έναν άγαμο έχουμε :

Loss = 59,638 + 4,831\*Cards – 0,517\*Deposit + 4,414\*0 + 0,893\*Type01 – 0.665\*0\*depοsit

Αν αφαιρέσουμε κατά μέλη τις παραπάνω δύο σχέσεις έχουμε :

**ΔLoss = 4,414 – 0,665\*deposit.**

Άρα η διαφορά των ζημιών μεταξύ ενός έγγαμου και ενός άγαμου μειώνεται κατά 0,665 ανά 1000 ευρώ καταθέσεων.