



ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΩΝ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ
&
ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΕΠΙΒΙΩΣΗΣ

-

COX REGRESSION

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ: Δρ ΜΗΤΡΟΠΟΥΛΟΣ ΠΑΝΑΓΙΩΤΗΣ

ΕΠΙΜΕΛΕΙΑ ΕΡΓΑΣΙΑΣ:
ΑΝΑΣΤΟΠΟΥΛΟΥ ΕΥΓΕΝΙΑ
ΓΙΑΝΝΟΠΟΥΛΟΥ ΚΥΡΙΑΚΗ
ΜΗΤΡΟΠΟΥΛΟΥ ΑΝΑΣΤΑΣΙΑ

ΠΑΤΡΑ 2011

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ.....	4
1. ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ	5
1.1 Εισαγωγικοί ορισμοί.....	5
1.2 Λογοκριμένα δεδομένα.....	6
1.3 Προϋποθέσεις της ανάλυσης επιβίωσης.....	11
1.4 Συναρτήσεις του χρόνου επιβίωσης.....	11
1.4.1 Συνάρτηση επιβίωσης (Survival function).....	12
1.4.2 Συνάρτηση πυκνότητας πιθανότητας ή συνάρτηση πυκνότητας.....	13
1.4.3 Συνάρτηση κινδύνου (Hazard function).....	14
1.5 Μοντέλα επιβίωσης.....	16
2. ΜΟΝΤΕΛΟ ΕΠΙΒΙΩΣΗΣ COX REGRESSION.....	18
2.1 Επεκτάσεις του μοντέλου αναλογικού κινδύνου του Cox.....	19
2.1.1 Στρωματοποίηση (Strata).....	19
2.1.2 Μεταβλητές εξαρτώμενες από το χρόνο.....	20
3. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ.....	21
3.1 Μεθοδολογία.....	21
3.2 Παρουσίαση των δημογραφικών στοιχείων του δείγματος.....	22
3.3 COX Regression μονού παράγοντα.....	33
3.3.1 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΦΥΛΟ.....	33
3.3.2 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΥΠΗΚΟΟΤΗΤΑ.....	40
3.3.3 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΕΓΓΡΑΦΗ ΜΕ.....	48
3.3.4 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ.....	55
3.3.5 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ.....	62
3.3.6 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΠΤΥΧΙΑΚΗ.....	69
3.3.7 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ.....	72
3.3.8 Σύγκριση αποτελεσμάτων.....	75
3.4 COX Regression πολυπαραγοντικό.....	76
3.4.1 Επιλεγμένες μεταβλητές.....	76
3.4.2 Επιλογή μεταβλητών με τη μέθοδο Forward : LR.....	84
3.5 Έλεγχος υποθέσεων του μοντέλου επιβίωσης COX.....	88
3.5.1 Έλεγχος για τις ποιοτικές μεταβλητές.....	88
3.5.2 Έλεγχος για τις ποσοτικές μεταβλητές.....	93
3.6 Μοντέλο επιβίωσης COX με χρονικά εξαρτημένες.....	95
3.6.1 Ποιοτικές μεταβλητές.....	95

3.6.2. Ποσοτικές μεταβλητές.....	95
Συμπεράσματα.....	98
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	99

ΕΙΣΑΓΩΓΗ

Σε αυτή τη πτυχιακή θα παρουσιαστούν θέματα που σχετίζονται με τις μεθόδους ανάλυσης επιβίωσης και συγκεκριμένα με την Παλινδρόμηση του COX (Cox Regression).

Η ανάλυση επιβίωσης είναι επίσης γνωστή ως μελέτη του χρόνου μέχρι να συμβεί ένα γεγονός και κατ' ουσία είναι μια μέθοδος για να διαπιστωθεί αν θα συμβεί ένα γεγονός ή όχι.

Η ανάλυση κατά Cox είναι μία στατιστική τεχνική που χρησιμοποιείται για να καθοριστεί η σχέση μεταξύ επιβίωσης και πολλών ανεξάρτητων διερευνητικών μεταβλητών. Μας είναι χρήσιμη για την μοντελοποίηση του χρόνου μέχρι να συμβεί ένα συγκεκριμένο τελικό γεγονός με βάση την τιμή μιας δεδομένης συμμεταβλητής (covariate).

Η Cox Regression παρέχει μία εκτίμηση του ποσοστού επιβίωσης, μετά την προσαρμογή της διερευνητικής μεταβλητής. Το βασικό μοντέλο παράγει την αναλογική συνάρτηση κινδύνου, η οποία μπορεί να επεκταθεί με την χρήση μιας μεταβλητής στρωματοποίησης (strata variable) ή χρονικά εξαρτημένες συμμεταβλητές.

Στη συνέχεια θα παρουσιαστούν όλα τα συστατικά στοιχεία του μοντέλου αλλά και ορισμοί που θα βοηθήσουν στην κατανόηση της χρησιμότητας της ανάλυσης επιβίωσης και ιδιαίτερα του COX. Πέραν της παρουσίασης του θεωρητικού υπόβαθρου του μοντέλου, παρέχεται η εμπειρική ανάλυση επί πραγματικών δεδομένων. Χρησιμοποιούνται πραγματικά στοιχεία από φοιτητές του τμήματος Επιχειρηματικού Σχεδιασμού και Πληροφορικών Συστημάτων του ΤΕΙ Πάτρας για τα έτη 2000 έως 2009. Στην πτυχιακή αναλύονται η σχέση της διάρκειας φοίτησής (επιβίωση) με διάφορες επεξηγητικές μεταβλητές πρόβλεψης με βάση τα δημογραφικά χαρακτηριστικά των φοιτητών. Τέλος, γίνεται έλεγχος της υπόθεσης για τη σταθερότητα των διαφόρων συμμεταβλητών στο χρόνο και προσαρμόζονται εφόσον χρειάζεται μοντέλα με χρονικά μεταβαλλόμενες συμμεταβλητές.

1. ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ

1.1 Εισαγωγικοί ορισμοί

Ανάλυση επιβίωσης (survival analysis) είναι μια περιοχή έρευνας στη στατιστική, η οποία δημιουργήθηκε για την ανάλυση δεδομένων τα οποία δε μπορούν να επεξεργαστούν από τις συνηθισμένες στατιστικές μεθόδους. Τα δεδομένα αυτά δίνουν τη χρονική διάρκεια μέχρι να γίνει ένα συγκεκριμένο γεγονός.

Ο χρόνος επιβίωσης ή χρόνος αποτυχίας (survival time ή failure time), αναφέρεται σε μια μεταβλητή που μετράει το χρόνο (ημέρες, εβδομάδες, μήνες, κλπ) που μεσολαβεί από τη στιγμή της έναρξης της παρακολούθησης ενός ατόμου (άνθρωπος, αντικείμενο, φαινόμενο κτλ.), μέχρι τη στιγμή που το άτομο θα αντιμετωπίσει το ενδεχόμενο.

Ο χρόνος επιβίωσης είναι το βασικό σημείο ενδιαφέροντος σε πολλές ιατρικές εφαρμογές (π.χ. ο χρόνος μέχρι την κατάληξη ενός ασθενή), σε κοινωνικές (π.χ. ο χρόνος μέχρι την αποφοίτηση σπουδαστών), οικονομικές επιστήμες (π.χ. ο χρόνος μέχρι ένας χρηματιστηριακός δείκτης να ξεπεράσει ένα όριο) καθώς και στη βιομηχανία (π.χ. ο χρόνος μέχρι να χαλάσει ένα εξάρτημα μιας μηχανής). Ενδεχόμενα ερωτήματα που μπορεί να προκύψουν είναι ο χαρακτηρισμός της κατανομής του χρόνου επιβίωσης, καθώς και η σύγκριση αυτού του χρόνου μεταξύ διαφορετικών ομάδων ή ακόμη η μοντελοποίηση της σχέσης του χρόνου επιβίωσης σε σχέση με άλλες μεταβλητές.

Οι χρόνοι επιβίωσης συνήθως έχουν μια κατανομή η οποία διαφέρει πολύ από την κανονική. Πολλές από τις συνηθισμένες στατιστικές μεθόδους προϋποθέτουν ότι η κατανομή της μεταβλητής που εξετάζουμε είναι κανονική και έτσι δεν είναι κατάλληλες τέτοιες μέθοδοι για την ανάλυση δεδομένων επιβίωσης. Το ιδιαίτερο χαρακτηριστικό που εμφανίζεται στα δεδομένα επιβίωσης και δεν επιτρέπει τη χρήση των συνηθισμένων στατιστικών

τεχνικών (όπως αναφέρθηκε στην αρχή), είναι ότι οι χρόνοι επιβίωσης ορισμένων παρατηρήσεων είναι λογοκριμένοι (censored). Αυτό συνήθως συμβαίνει επειδή τα άτομα μπορεί να εισέρχονται στη μελέτη σε διαφορετικούς χρόνους, με συνέπεια ο χρόνος παρακολούθησης μερικών ατόμων να μην είναι επαρκής ώστε να καταγραφεί ο χρόνος μέχρι την πραγματοποίηση του υπό μελέτη γεγονότος (Κ.Φωκιανός , Χ.Χαραλάμπους, 2001).

1.2 Λογοκριμένα δεδομένα

Λογοκριμένα δεδομένα, έχουμε όταν υπάρχουν παρατηρήσεις, των οποίων οι χρόνοι επιβίωσης δεν είναι ακριβείς. Η πληροφορία για το χρόνο επιβίωσης ενός ατόμου στην περίπτωση αυτή είναι μερική (αφού γνωρίζουμε μόνο ένα κάτω φράγμα του χρόνου επιβίωσης). Ο όρος censoring χρησιμοποιήθηκε για πρώτη φορά από τον Hald (1949). Τα δεδομένα που δεν είναι λογοκριμένα ονομάζονται μη-λογοκριμένα ή πλήρη δεδομένα. Για να γίνει κατανοητή η έννοια των λογοκριμένων παρατηρήσεων, δίνεται ένα παράδειγμα στο οποίο υπάρχουν λογοκριμένες παρατηρήσεις.

Σε μια έρευνα για τη μελέτη του παραγόντων που επηρεάζουν το χρόνο αποφοίτησης σπουδαστών ΑΤΕΙ, η μεταβλητή που μας ενδιαφέρει είναι ο αριθμός των ετών που θα φοιτήσει ο κάθε σπουδαστής, δηλαδή ο χρόνος «επιβίωσης» κάθε ατόμου. Οι σπουδαστές μπορεί να εισέρχονται στη μελέτη σε διαφορετικούς χρόνους, ενώ η διάρκεια της μελέτης είναι προεπιλεγμένη. Έτσι, για κάθε σπουδαστή, καταγράφεται ο χρόνος, από την είσοδό του στη μελέτη, μέχρι την αποφοίτησή του. Στο τέλος της μελέτης, το πιθανότερο είναι να υπάρχουν σπουδαστές που δεν αποφοίτησαν σε ολόκληρη τη διάρκεια της μελέτης (συνήθως οι σπουδαστές που εισήλθαν αργά στη μελέτη), ενώ θα υπάρχουν σπουδαστές με τους οποίους χάθηκε η επαφή. Ο χρόνος επιβίωσης (αποφοίτησης) των ατόμων αυτών, θα είναι τουλάχιστον όσο ο χρόνος από την είσοδο τους στη μελέτη, μέχρι τη στιγμή που ολοκληρώθηκε η μελέτη (για την πρώτη περίπτωση), και μέχρι τη στιγμή που χάθηκε η επαφή (στη δεύτερη περίπτωση). Αυτές οι παρατηρήσεις, είναι λογοκριμένες (censored) παρατηρήσεις.

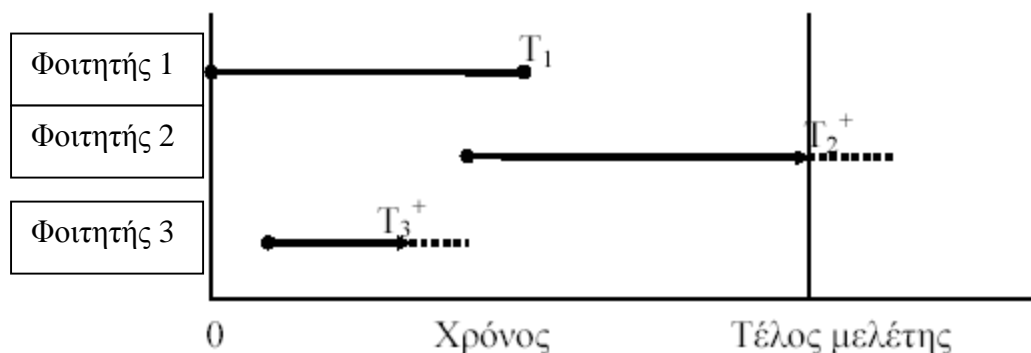
Σίγουρα, δε θα θέλαμε να αποκλείσουμε τα άτομα αυτά από τη μελέτη θεωρώντας τα ως ελλιπή δεδομένα. Κάτι τέτοιο, θα επηρέαζε πολύ την ανάλυση και τα αποτελέσματα που θα παίρναμε δε θα ήταν σωστά, αφού οι περισσότεροι από τους σπουδαστές αυτούς σπουδάζουν ακόμη και επομένως αντανακλούν την αδυναμία να αποφοιτήσουν έγκαιρα.

Οι λογοκριμένες παρατηρήσεις δεν προκύπτουν μόνο λόγω του χρόνου λήξης της έρευνας, μπορεί να προκύψουν και σε άλλες περιπτώσεις:

§ **Όταν ο σπουδαστής χάνεται από την παρακολούθηση:** Ο σπουδαστής μπορεί να πήρε μεταγραφή για άλλο εκπαιδευτικό ίδρυμα.

§ **Ο σπουδαστής αποσύρεται από την παρακολούθηση:** Ο σπουδαστής μπορεί να εγκατέλειψε τις σπουδές του.

Για να γίνουν κατανοητά τα πιο πάνω, θεωρούμε το σχήμα 1 όπου φαίνονται οι χρόνοι μέχρι την αποφοίτηση τριών φοιτητών.



Σχήμα 1: Χρόνοι φοίτησης τριών φοιτητών

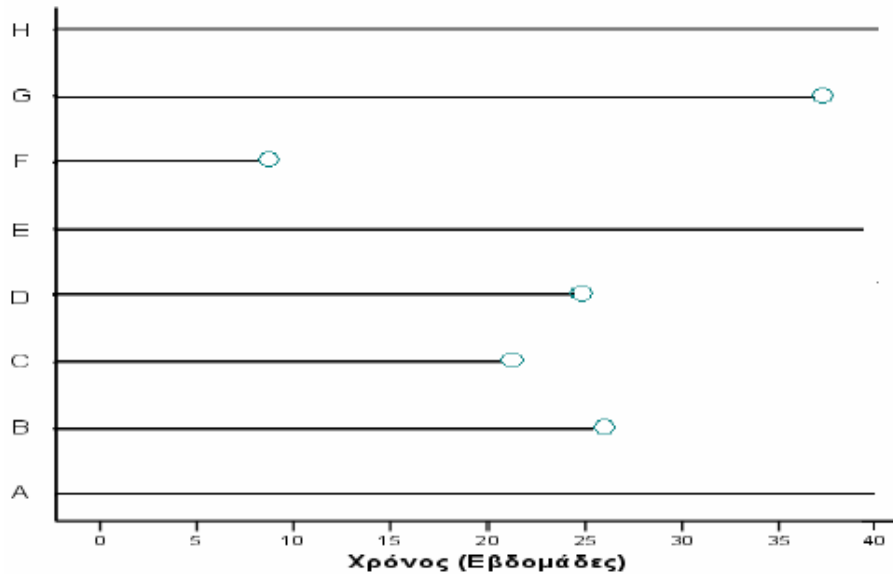
Υπάρχουν 3 είδη λογοκρισίας. Η **δεξιά λογοκρισία** (right censoring), η **αριστερή λογοκρισία** (left censoring) και η **λογοκρισία διαστήματος** (interval censoring) (Ε.Παύλου,2006).

Επιπλέον, η δεξιά λογοκρισία χωρίζεται σε 3 κατηγορίες, τη λογοκρισία τύπου I (Type I censoring), τη λογοκρισία τύπου II (Type II censoring) και την τυχαία λογοκρισία (random censoring). Θεωρούμε ότι T_i είναι ο χρόνος

επιβίωσης ή χρόνος αποτυχίας του ατόμου i και u ο χρόνος στον οποίο σταματά η μελέτη.

§ **Δεξιά λογοκρισία (right censoring):** Στην περίπτωση αυτή, ο χρόνος επιβίωσης T_i , είναι μεγαλύτερος από το χρόνο u . Δηλαδή, δε γνωρίζουμε τον ακριβή χρόνο επιβίωσης του i -στού ατόμου, γνωρίζουμε μόνο ότι ο χρόνος επιβίωσης του είναι στο διάστημα (u, ∞) . Η δεξιά λογοκρισία, είναι η πιο συνηθισμένη μορφή λογοκρισίας. Παρατηρείται σε περιπτώσεις όπου ένα άτομο χάνεται ή αποσύρεται από την παρακολούθηση, ή ακόμη όταν η μελέτη τερματίζεται σε ένα προκαθορισμένο χρόνο.

ü Λογοκρισία τύπου I (Type I censoring): Όταν από την αρχή της έρευνας προκαθορίζεται ο χρόνος διάρκειάς της, έστω u , τότε έχουμε λογοκρισία τύπου I. Ο χρόνος u ονομάζεται χρόνος λογοκρισίας (censoring time). Έτσι, ο ερευνητής καταγράφει τους χρόνους αποτυχίας ή επιβίωσης των ατόμων που απέτυχαν κατά τη διάρκεια της έρευνας, ενώ για τα υπόλοιπα άτομα το μόνο που είναι γνωστό είναι ότι οι χρόνοι επιβίωσης τους είναι μεγαλύτεροι από u . Στη λογοκρισία τύπου I, όταν δεν υπάρχουν απώλειες από ατυχήματα, όλες οι λογοκριμένες παρατηρήσεις ισούνται με το μήκος της περιόδου της μελέτης. Στο σχήμα 2 δίνεται ανάλογο παράδειγμα με τα H, E, A άτομα να έχουν χρόνους επιβίωσης μεγαλύτερους από το χρόνο μελέτης και να είναι λογοκριμένες παρατηρήσεις, ενώ τα F, D, C και B άτομα έχουν χρόνους επιβίωσης μικρότερους από το διάστημα μελέτης και καταγράφονται οι χρόνοι επιβίωσής τους κανονικά.



Σχήμα 2: Ένα παράδειγμα λογοκριμένων δεδομένων τύπου I

ü Λογοκρισία τύπου II (Type II censoring): Στην λογοκρισία τύπου II, η μελέτη συνεχίζεται μέχρι να 'αποτύχουν' r άτομα. Ο αριθμός r καθορίζεται πριν την έναρξη της μελέτης. Έτσι, αν έχουμε n άτομα υπό μελέτη, τότε στο τέλος της μελέτης, γνωρίζουμε τους χρόνους αποτυχίας r ατόμων, ενώ για τα υπόλοιπα $n-r$ άτομα, γνωρίζουμε μόνο ότι ο χρόνος επιβίωσης τους είναι μεγαλύτερος από το χρόνο επιβίωσης των r ατόμων που απέτυχαν. Δηλαδή, στη λογοκρισία τύπου II, οι λογοκριμένες παρατηρήσεις ισούνται με τη μεγαλύτερη μη-λογοκριμένη παρατήρηση.

ü Τυχαία Λογοκρισία (Random censoring): Στην περίπτωση αυτή, ο χρόνος λογοκρισίας που αντιστοιχεί σε κάθε υπό παρακολούθηση άτομο δεν είναι σταθερός, αλλά είναι τυχαίος. Για παράδειγμα, σε κλινικές μελέτες, ενώ οι χρονικές στιγμές έναρξης και λήξης της έρευνας είναι προκαθορισμένες, οι ασθενείς εισέρχονται σε αυτή σε διαφορετικές (τυχαίες) χρονικές στιγμές, με αποτέλεσμα οι χρόνοι λογοκρισίας τους να είναι τυχαίοι.

§ **Αριστερή Λογοκρισία (left censoring):** Το μόνο που είναι γνωστό στην περίπτωση αυτή, είναι ότι ο χρόνος επιβίωσης T , είναι μικρότερος από ένα χρονικό διάστημα. Ο ακριβής χρόνος επιβίωσης δεν είναι γνωστός.

§ **Λογοκρισία σε διάστημα (Interval censoring):** Στη λογοκρισία σε διάστημα, γνωρίζουμε μόνο ότι ο χρόνος επιβίωσης, T , βρίσκεται σε ένα διάστημα (U_1, U_2) . Αυτού του είδους η λογοκρισία, παρατηρείται συνήθως όταν έχουμε περιοδική παρακολούθηση.

1.3 Προϋποθέσεις της ανάλυσης επιβίωσης.

Οι μέθοδοι για την ανάλυση επιβίωσης, προϋποθέτουν ότι οι παρατηρήσεις είναι ανεξάρτητες. Επιπλέον, πρέπει να ισχύει ακόμη μία προϋπόθεση για τις μεθόδους που υπάρχουν για τα λογοκριμένα δεδομένα επιβίωσης, αυτή της ανεξάρτητης λογοκρισίας. Δηλαδή, ένα άτομο που είναι λογοκριμένο (censored) και είναι ζωντανό στο χρόνο t , πρέπει να έχει τον ίδιο κίνδυνο επακόλουθης αποτυχίας (ή να έχει την ίδια πιθανότητα να επιβιώσει) με ένα άτομο που είναι μη-λογοκριμένο (uncensored), στο χρόνο t . Στατιστικά, η προϋπόθεση αυτή, είναι ισοδύναμη με την ανεξαρτησία της διαδικασίας της λογοκρισίας με το χρόνο επιβίωσης.

Έτσι, αν υπάρχουν λογοκριμένες παρατηρήσεις που προκύπτουν μόνο λόγω διαφορετικών εισόδων στη μελέτη, τότε η υπόθεση της ανεξαρτησίας φαίνεται να ισχύει. Όταν όμως οι λογοκριμένες παρατηρήσεις προκύπτουν επειδή χάθηκε το άτομο από την παρακολούθηση, ή όταν προκύπτουν λόγω αποχώρησης του σπουδαστή από τη μελέτη εξαιτίας μεταγραφής σε άλλο ΑΤΕΙ, τότε είναι πιθανόν η υπόθεση της ανεξαρτησίας να μην ισχύει.

1.4 Συναρτήσεις του χρόνου επιβίωσης.

Η κατανομή των χρόνων επιβίωσης, χαρακτηρίζεται συνήθως από τρεις συναρτήσεις, τη συνάντηση επιβίωσης (survivorship function), τη συνάρτηση πυκνότητας πιθανότητας (probability density function) και τη συνάρτηση κινδύνου (hazard function). Στην πράξη, οι τρεις αυτές συναρτήσεις μπορούν να χρησιμοποιηθούν για να επεξηγήσουν διαφορετικές όψεις των δεδομένων. Έστω T ο χρόνος επιβίωσης.

1. 4. 1 Συνάρτηση επιβίωσης (Survival function).

Η συνάρτηση αυτή, που συμβολίζεται με $S(t)$, ορίζεται ως η πιθανότητα το άτομο να επιβιώνει για χρόνο μεγαλύτερο του t :

$$S(t)=P(\text{ένα άτομο επιβιώνει για χρόνο μεγαλύτερο του } t)=P(T>t)$$

Επειδή $F(t)=P(T < t)$ ή

$$S(t)=1-F(t)$$

Η $S(t)$ είναι μια φθίνουσα συνάρτηση του t με τις ιδιότητες:

- $S(t)=1$ για $t=0$ και
- $S(t)=0$ για $t=\infty$

Δηλαδή, η πιθανότητα το άτομο να επιβιώνει τουλάχιστον στο χρόνο 0 είναι 1 και η πιθανότητα επιβίωσης σε ένα άπειρο χρόνο είναι 0. Η γραφική παράσταση του $S(t)$ συναρτήσεως του t ονομάζεται καμπύλη επιβίωσης (survival curve) και προτάθηκε από τον Berkson το 1942 (Ε.Παύλου,2006).

Στην πράξη, όταν δεν υπάρχουν λογοκριμένες παρατηρήσεις, η συνάρτηση επιβίωσης εκτιμάται ως η αναλογία των ασθενών που επιβιώνουν για χρόνο μεγαλύτερο του t .

$$\hat{S}(t) = \frac{\text{αρ. ατόμων που επιβιώνουν για χρόνο μεγαλύτερο του } t}{\text{συνολικός αριθμός ατόμων}}$$

Όταν υπάρχουν λογοκριμένες παρατηρήσεις, τότε ο αριθμητής δε μπορεί πάντα να οριστεί. Έτσι, όταν έχουμε λογοκριμένα δεδομένα καταφεύγουμε σε μη παραμετρικές μεθόδους για την εκτίμηση του $S(t)$.

1. 4. 2 Συνάρτηση πυκνότητας πιθανότητας ή συνάρτηση πυκνότητας (Probability density function).

Όπως και με οποιαδήποτε άλλη συνεχή τυχαία μεταβλητή, η συνάρτηση πυκνότητας πιθανότητας του χρόνου επιβίωσης T , ορίζεται ως το όριο της πιθανότητας ότι ένα άτομο αποτυγχάνει σε ένα μικρό διάστημα $(t, t+\Delta t)$ ανά μονάδα πλάτους Δt . Μπορεί να εκφραστεί ως εξής:

$$f(t) = \frac{\text{αρ. ατόμων που αποτυγχάνουν στο διάστημα } (t, t+\Delta t)}{\Delta t}$$

ή

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

Η καμπύλη της $f(t)$ ονομάζεται καμπύλη πυκνότητας (density curve). Η συνάρτηση πυκνότητας έχει τις ακόλουθες δύο ιδιότητες:

- § $f(t)=0$ για $t < 0$
- § Το εμβαδόν μεταξύ της καμπύλης πυκνότητας και του άξονα των t ισούται με 1.

Στην πράξη, όταν δεν υπάρχουν λογοκριμένες παρατηρήσεις, η συνάρτηση πυκνότητας πιθανότητας $f(t)$, εκτιμάται ως την αναλογία των ατόμων που αποτυγχάνουν σε ένα διάστημα ανά μονάδα πλάτους:

$$\hat{f}(t) = \frac{\text{αρ. ατόμων που αποτυγχάνουν στο διάστημα που ξεκινά στο χρόνο } t}{(\text{συνολικός αριθμός ατόμων}) * (\text{πλάτος διαστήματος})}$$

Όπως και με την εκτίμηση της συνάρτησης επιβίωσης, η σχέση αυτή εφαρμόζεται μόνο όταν δεν υπάρχουν λογοκριμένα δεδομένα.

1. 4. 3 Συνάρτηση κινδύνου (Hazard function).

Ο όρος αυτός χρησιμοποιήθηκε για πρώτη φορά από τον Barlow το 1963. Η συνάρτηση κινδύνου ($h(t)$ ή $\lambda(t)$) καθορίζει το στιγμιαίο λόγο αποτυχίας στο $T=t$, δεδομένου της επιβίωσης στο χρόνο t και ορίζεται ως:

$$h(t) = \lim_{Dt \rightarrow 0} \frac{P(t \leq T < t + Dt | T \geq t)}{Dt}$$

και σε συνδυασμό με τις σχέσεις της προηγούμενης ενότητας παίρνουμε:

$$h(t) = \lim_{Dt \rightarrow 0} \frac{P(t \leq T < t + Dt)}{P(T \geq t)Dt} = \frac{1}{P(T \geq t)} f(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}$$

Η συνάρτηση κινδύνου είναι γνωστή και ως στιγμιαίος λόγος αποτυχίας (instantaneous failure rate) ή ως δεσμευμένος λόγος θνησιμότητας (conditional mortality). Η ποσότητα $h(t)\Delta t$ (Δt μικρό) είναι προσεγγιστικά η πιθανότητα θανάτου ενός ατόμου στο διάστημα $[t, t+\Delta t)$, γνωρίζοντας ότι το άτομο έχει επιβιώσει μέχρι τη χρονική στιγμή t .

Στην πράξη, όταν δεν υπάρχουν λογοκριμένες παρατηρήσεις, η συνάρτηση κινδύνου εκτιμάται ως η αναλογία των ατόμων που αποτυγχάνουν σε ένα διάστημα ανά μονάδα χρόνου, δεδομένου ότι επιβίωσαν μέχρι την αρχή του διαστήματος:

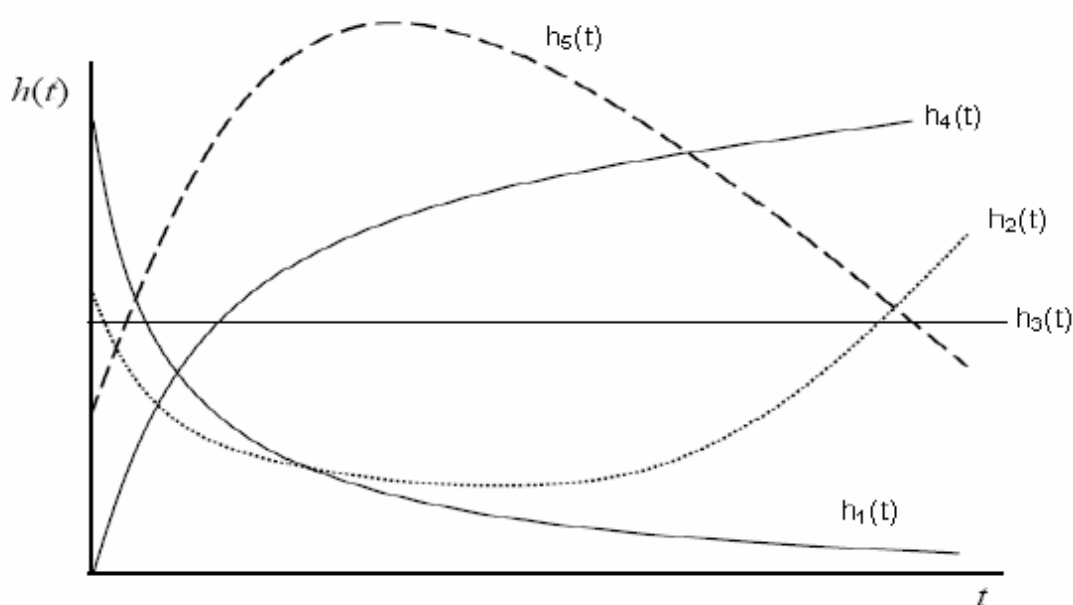
$$\hat{h}(t) = \frac{\text{αρ. ατόμων που αποτυγχάνουν στο διάστημα που ξεκινά στο χρόνο } t}{(\text{αρ. ατόμων που είναι ζωντανά μέχρι το χρόνο } t) * (\text{πλάτος χρονικού διαστήματος})}$$
$$= \frac{\text{αρ. ατόμων που αποτυγχάνουν ανά μονάδα χρόνου στο διάστημα που ξεκινά στο } t}{\text{αρ. ατόμων που είναι ζωντανά σ το χρόνο } t}$$

Οι ασφαλιστές χρησιμοποιούν συνήθως μία διαφορετική εκτίμηση της συνάρτησης κινδύνου, το μέσο λόγο κινδύνου του διαστήματος, στον οποίο οι αριθμοί των ασθενών που πεθαίνουν ανά μονάδα χρόνου στο διάστημα,

διαίρεται με το μέσο αριθμό των επιζώντων στο μέσο σημείο του διαστήματος.

$$\hat{h}^*(t) = \frac{\text{αρ. ατόμων που αποτυγχάνουν ανά μονάδα χρόνου στο διάστημα}}{\text{αρ. ατόμων που επιβιώνουν μέχρι το } t - 1/2 \text{ (αρ. αποτυχιών στο διάστημα)}}$$

Ο λόγος κινδύνου μπορεί να αυξάνει, να μειώνεται, να μένει σταθερός ή να δηλώνει μια πιο περίπλοκη διαδικασία. Στο παρακάτω σχήμα 3 φαίνονται διάφορες μορφές της συνάρτησης κινδύνου.



Σχήμα 3: Παραδείγματα συναρτήσεων κινδύνου (Ε.Παύλου,2006)

Η συνάρτηση $h_1(t)$ είναι μια φθίνουσα συνάρτηση κινδύνου και δεν εμφανίζεται συχνά στην πράξη. Δείχνει ότι σε αρχικούς χρόνους ο κίνδυνος είναι μεγάλος, ενώ όσο περνάει ο χρόνος ο κίνδυνος μειώνεται. Η $h_2(t)$ είναι η λεκανοειδής καμπύλη (bathtub curve) όπως ονομάζεται, και περιγράφει την εξέλιξη της ανθρώπινης ζωής. Στη διάρκεια μιας αρχικής περιόδου, ο κίνδυνος είναι μεγάλος (υψηλή βρεφική θνησιμότητα), στη συνέχεια, μέχρι μια συγκεκριμένη ηλικία ο κίνδυνος παραμένει σταθερός ενώ σε μεγαλύτερες ηλικίες αυξάνεται ακόμη περισσότερο. Η $h_3(t)$ είναι μια σταθερή συνάρτηση κινδύνου, δηλαδή ο κίνδυνος παραμένει σταθερός. Αυτό συμβαίνει όταν π.χ. εξετάζουμε τον κίνδυνο θανάτου υγιών ατόμων ηλικίας 18-40, των οποίων οι

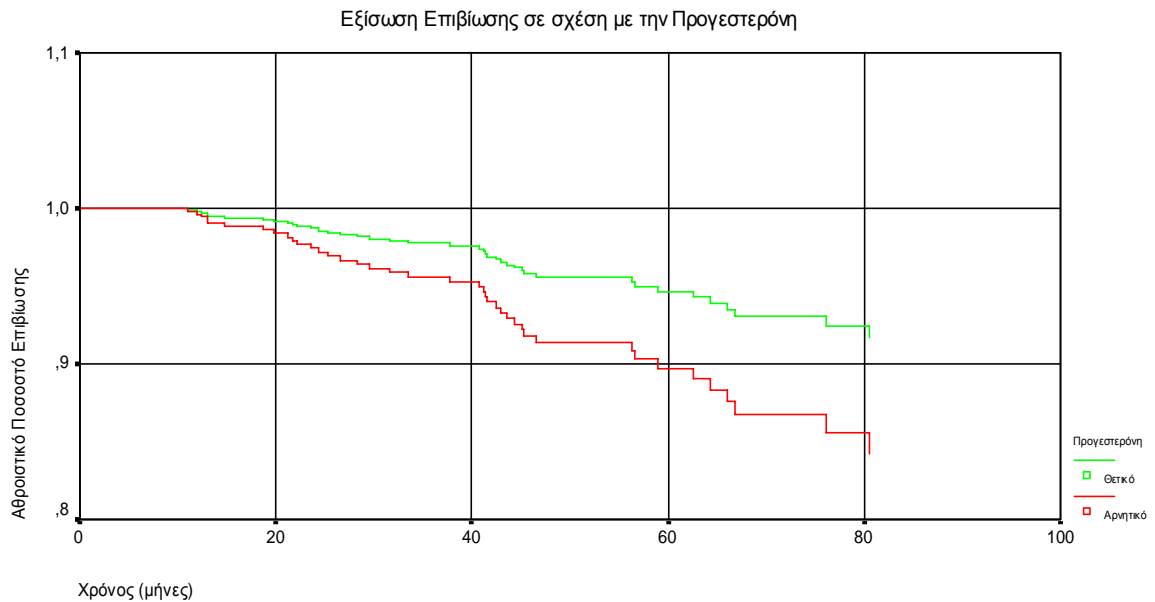
κύριες αιτίες θανάτου είναι τα ατυχήματα. Η $h_4(t)$ είναι μια αύξουσα συνάρτηση, η οποία συναντάται συχνά. Με την πάροδο του χρόνου το ρίσκο κινδύνου αυξάνεται. Για παράδειγμα ασθενείς με οξεία λευχαιμία, έχουν ένα αυξανόμενο λόγο διακινδύνευσης. Τέλος, ασθενείς με φυματίωση έχουν ρίσκο που αυξάνεται αρχικά, ενώ μετά από θεραπεία μειώνεται και η συνάρτηση κινδύνου στην περίπτωση αυτή έχει τη μορφή της $h_5(t)$.

1.5 Μοντέλα επιβίωσης

Στα μοντέλα επιβίωσης κύριο χαρακτηριστικό μέτρησης είναι η διάρκεια του χρόνου ώσπου να συντελεστεί το κρίσιμο γεγονός. Η ανάλυση μπορεί να αφορά μια απλή ομάδα (group) ή να εξετάζει το χρόνο επιβίωσης σαν μια συνάρτηση από κατηγορικές και συνεχείς ανεξάρτητες μεταβλητές. Συνήθως, τα μοντέλα επιβίωσης χρησιμοποιούνται σε ιατρικές έρευνες για την μελέτη των χαρακτηριστικών των ασθενών που επιβίωσαν σε σχέση με την ασθένεια που είχαν. Όπως αναφέραμε τα μοντέλα επιβίωσης εξετάζουν την διάρκεια του χρόνου ώσπου να συντελεστεί το κρίσιμο γεγονός, πιθανότητα σαν μια συνάρτηση από ανεξάρτητες μεταβλητές. Εάν το κρίσιμο γεγονός είναι θάνατος, τότε υπάρχει ιδιαίτερο ενδιαφέρον στο χρόνο επιβίωσης από άτομα που ανήκουν σε διαφορετικούς πληθυσμούς. Ιδιαίτερα στις ιατρικές έρευνες τους ενδιαφέρει ο χρόνος επιβίωσης σε σχέση με την θεραπεία που ακολουθήθηκε κάθε φορά. Ο λόγος που δεν χρησιμοποιούμε μοντέλα παλινδρόμησης ή ανάλυση της διακύμανσης (ANOVA) οφείλεται στο γεγονός ότι τα δεδομένα επιβίωσης περιέχουν λογοκριμένα δομένα: παρατηρήσεις για τις οποίες δεν γνωρίζουμε το τελικό αποτέλεσμα, παρόλα αυτά κάποια πληροφορία είναι διαθέσιμη προς εκμετάλλευση και μπορεί να χρησιμοποιηθεί από τα μοντέλα επιβίωσης.

Μια χρήσιμη πληροφορία είναι η συνάρτηση επιβίωσης κατά την διάρκεια του χρόνου. Το γράφημα (ή πίνακας) της αθροιστικής συνάρτησης επιβίωσης παρουσιάζει την εκτιμώμενη πιθανότητα στο να επιβιώσει ο ασθενής περισσότερο από το χρόνο που παρουσιάζεται στο γράφημα (ή πίνακα). Στο παρακάτω διάγραμμα 1 παρουσιάζεται ένα παράδειγμα

γραφήματος της αθροιστικής συνάρτησης επιβίωσης σε δύο ομάδες σε σχέση με την ύπαρξη προγεστερόνης ή μη στους ασθενείς (SPSSBI, 2008).



Διάγραμμα 1: Παράδειγμα αθροιστικής συνάρτησης επιβίωσης.

Παρατηρείται ότι στην αρχή της έρευνας η πιθανότητα της αθροιστικής επιβίωσης ξεκινάει από την μονάδα (100%) ενώ κατά την διάρκεια του χρόνου η πιθανότητα μειώνεται. Η πιθανότητα είναι άμεσα συνδεδεμένη με το χρόνο και μειώνεται κάθε φορά όταν συντελείται το κρίσιμο γεγονός(συγκεκριμένα στη περίπτωση μας είναι ο θάνατος). Τέλος παρατηρείται ότι η πιθανότητα επιβίωσης μειώνεται σε μεγαλύτερο ρυθμό στους ασθενείς που η ύπαρξη προγεστερόνης είναι αρνητικό σε σχέση με τους ασθενείς που η ύπαρξη προγεστερόνης είναι θετικό.

2. ΜΟΝΤΕΛΟ ΕΠΙΒΙΩΣΗΣ COX REGRESSION

Cox Regression είναι ένα μοντέλο επιβίωσης όπου αναπαριστά το ρίσκο επικινδυνότητας (hazard) στο να συντελεστεί το κρίσιμο γεγονός κατά τη διάρκεια του χρόνου που γίνεται η έρευνα. Το μοντέλο επιβίωσης Cox Regression χρησιμοποιεί ποσοτικές και ποιοτικές ανεξάρτητες μεταβλητές για την διερεύνηση της κίνησης μίας κατηγορικής εξαρτημένης μεταβλητής. Μπορεί να θεωρηθεί ως προσέγγιση μη παραμετρικής στατιστικής ή για να ακριβολογούμε μπορεί να θεωρηθεί ως ημι-παραμετρικό μοντέλο επειδή δεν απαιτείται μια συγκεκριμένη μορφή της καμπύλης επιβίωσης (survival) ή της καμπύλης επικινδυνότητας (hazard). Το μοντέλο προϋποθέτει ότι η αναλογία του ρυθμού επικινδυνότητας μεταξύ δύο ατόμων(individual) ή ομάδων(group) παραμένει σταθερή στο χρόνο.

Στο μοντέλο επιβίωσης Cox Regression, η συνάρτηση επικινδυνότητας στο χρόνο t σε συνάρτηση με τις ανεξάρτητες μεταβλητές X_1 έως X_p μπορεί να διατυπωθεί ως εξής :

$$H(t|X_1, X_2, \dots, X_p) = h_0(t) * e^{X_1 * B_1 + X_2 * B_2 + \dots + X_p * B_p}$$

Η συνάρτηση επικινδυνότητας (hazard function) μπορεί να διατυπωθεί ως αποτέλεσμα δύο συνιστωσών : στη βασική συνάρτηση επιβίωσης ($h_0(t)$), που αλλάζει στη διάρκεια του χρόνου και δεν εξαρτάται από τις ανεξάρτητες μεταβλητές και στις επιδράσεις των κατηγορικών και συνεχών μεταβλητών

$$e^{X_1 * B_1 + X_2 * B_2 + \dots + X_p * B_p}$$

οι οποίες είναι ανεξάρτητες στη διάρκεια του χρόνου και προσαρμόζουν τη βασική συνάρτηση επιβίωσης. Η μορφή της καμπύλης επιβίωσης δεν είναι καθορισμένη και εμπειρικά καθορίζεται από τα δεδομένα (εδώ φαίνεται και η μη παραμετρική πλευρά του μοντέλου). Εφόσον το μοντέλο επιβίωσης συνδέεται άμεσα με την εκθετική συνάρτηση μπορούμε να πούμε ότι η τιμή του συντελεστή του μοντέλου (π.χ. e^{B_1}) μπορεί να ερμηνευθεί ως η αλλαγή που συντελείται στη συνάρτηση επιβίωσης σε σχέση με μια μονάδα αλλαγής της ανεξάρτητης μεταβλητής (X_1) όταν οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές.

Ο διαχωρισμός του μοντέλου σε δύο συνιστώσες, στη συνάρτηση που εξαρτάται από το χρόνο και στις ανεξάρτητες μεταβλητές υπονοεί ότι η αναλογία των συναρτήσεων επιβίωσης μεταξύ δύο ατόμων (individual) ή ομάδων (group) παραμένει σταθερή στη διάρκεια του χρόνου. Αν δεν ισχύει αυτή η προϋπόθεση, ότι οι επιδράσεις των ανεξάρτητων μεταβλητών είναι σταθερές στο χρόνο, τότε το μοντέλο επιβίωσης Cox Regression δεν θα παρέχει καλή προσαρμογή των δεδομένων.

Επειδή η συνάρτηση επικινδυνότητας έχει άμεση σχέση με την αθροιστική συνάρτηση επιβίωσης, το μοντέλο μπορεί εναλλακτικά να εκφραστεί σε σχέση με τους όρους της αθροιστικής επιβίωσης. Διαγράμματα επιβίωσης που βασίζονται στο μοντέλο μπορούν να παραχθούν (SPSSBI, 2008).

2.1 Επεκτάσεις του μοντέλου αναλογικού κινδύνου του Cox

Το μοντέλο αναλογικού κινδύνου του Cox, αφού τροποποιηθεί κατάλληλα, μπορεί να χρησιμοποιηθεί και σε περιπτώσεις όπου οι μεταβλητές παρουσιάζουν κάποια χαρακτηριστικά, διαφορετικά από αυτά που ισχύουν στο μοντέλο του Cox. Για παράδειγμα, στις εξαρτώμενες από το χρόνο μεταβλητές, δεν είναι όλες οι μεταβλητές σταθερές, αλλά η τιμή κάποιων μεταβλητών μεταβάλλεται με το χρόνο. Επίσης στην περίπτωση των στρωματοποιημένων μεταβλητών η υπόθεση της αναλογικότητας των κινδύνων δεν ισχύει απαραίτητα. Έτσι, όταν χρειάζεται να εξεταστούν τέτοιες μεταβλητές, το μοντέλο του Cox γενικεύεται και τροποποιείται κατάλληλα ώστε να μπορεί να αντιμετωπίσει τις περιπτώσεις αυτές (Garson, 2010).

2.1.1 Στρωματοποίηση (Strata)

Όταν μια μεταβλητή έχει επίπεδα που δημιουργούν συναρτήσεις κινδύνου οι οποίες δεν ικανοποιούν την υπόθεση της αναλογικότητας, τότε στρωματοποιούμε ως προς τη μεταβλητή αυτή. Το μοντέλο που προκύπτει ονομάζεται στρωματοποιημένο μοντέλο του Cox (strata Cox model) και εφαρμόζεται αφού θεωρήσουμε τη στρωματοποίηση των δεδομένων σε

υποομάδες. Το στρωματοποιημένο μοντέλο επιτρέπει στη μορφή της συνάρτησης κινδύνου να αλλάζει ανάμεσα στα επίπεδα της στρωματοποιημένης μεταβλητής. Δηλαδή, άτομα που ανήκουν στο ίδιο στρώμα έχουν συναρτήσεις κινδύνου ανάλογες μεταξύ τους. Αντίθετα, άτομα που ανήκουν σε διαφορετικά στρώματα δεν έχουν ανάλογες συναρτήσεις κινδύνου, αφού οι αναφορικές συναρτήσεις κινδύνου κάθε στρώματος είναι αυθαίρετες συναρτήσεις του χρόνου και αφήνονται ασυσχέτιστες.

Η στρωματοποίηση γίνεται για να μειωθεί μία ανομοιόμορφη δειγματοληψία ή/και για να διασφαλισθεί η κατάλληλη εκπροσώπηση των μικρών ομάδων σε ένα δείγμα.

Η μεταβλητή στρωματοποίησης όταν εισέρχεται στο μοντέλο επιδρά σε αυτό αλλά δεν αντιμετωπίζεται ως μεταβλητή πρόγνωσης οπότε δεν υπολογίζονται συντελεστές για αυτήν στο μοντέλο.

Ένα πλεονέκτημα αυτής της προσέγγισης είναι ότι δεν αναγκαίο να βρούμε μία ιδιαίτερη μορφή αλληλεπίδρασης μεταξύ της μεταβλητής στρωματοποίησης και του χρόνου. Ένα μειονέκτημα είναι η αδυναμία να προβληθούν οι επιπτώσεις της μεταβλητής στρωματοποίησης. Ως εκ τούτου η στρωματοποίηση χρησιμοποιείται όταν η μεταβλητή δεν έχει άμεσο ενδιαφέρον ως προς την επίδραση της (Garson, 2010).

2.1.2 Μεταβλητές εξαρτώμενες από το χρόνο

Μια δεύτερη επέκταση του μοντέλου αναλογικού κινδύνου του Cox προκύπτει όταν έχουμε μεταβλητές εξαρτώμενες από το χρόνο. Μέχρι τώρα θεωρούσαμε ότι οι μεταβλητές x_i ήταν σταθερές στο χρόνο. Υπάρχουν όμως περιπτώσεις στις οποίες οι τιμές κάποιων μεταβλητών μεταβάλλονται με το χρόνο. Μια μεταβλητή εξαρτώμενη από το χρόνο, είναι μια επεξηγηματική μεταβλητή της οποίας η τιμή μπορεί να αλλάζει με το χρόνο. Έτσι, το άτομο i θα έχει τιμή $x_i(t)$ στο χρόνο t . Το μοντέλο αναλογικού κινδύνου του Cox μπορεί να επεκταθεί έτσι ώστε να ενσωματώνει τέτοιες μεταβλητές. Ο πιο συνηθισμένος τύπος εξαρτώμενης από το χρόνο μεταβλητής είναι μια επαναλαμβανόμενη μέτρηση σε ένα άτομο ή μια αλλαγή στη θεραπεία ενός ατόμου (Garson, 2010).

3. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

3.1 Μεθοδολογία

Μετά το θεωρητικό υπόβαθρο που αναλύθηκε στα προηγούμενα τμήματα, χρησιμοποιούνται πραγματικά δεδομένα προκειμένου να παρουσιαστούν στην πράξη τα μοντέλα ανάλυσης επιβίωσης.

Κατ' αρχήν, τα δεδομένα που χρησιμοποιούνται προέρχονται από το χώρο της εκπαίδευσης και αφορούν τους αποφοίτους του τμήματος Επιχειρηματικού Σχεδιασμού και Πληροφορικών Συστημάτων του ΤΕΙ Πάτρας για τα έτη 2000 έως 2009. Τα δεδομένα περιλαμβάνουν δημογραφικά χαρακτηριστικά αλλά και στοιχεία σε σχέση με τις επιδόσεις των σπουδαστών και τη διάρκεια φοίτησής τους. Πιο συγκεκριμένα, οι μεταβλητές που μετέχουν στην ανάλυση είναι:

§ Κατάσταση (status)

§ Χρόνος φοίτησης (time)

Ανεξάρτητες μεταβλητές (covariates)

§ Φύλο (Ανδρας, Γυναίκα)

§ Υπηκοότητα (Ελληνική, Κυπριακή, Αλβανική)

§ Εγγραφή με (Εισιτήριες εξετάσεις, Τ.Ε.Ε., Άλλο)

§ Πτυχιακή (Βαθμός πτυχιακής)

§ Βαθμός πτυχίου

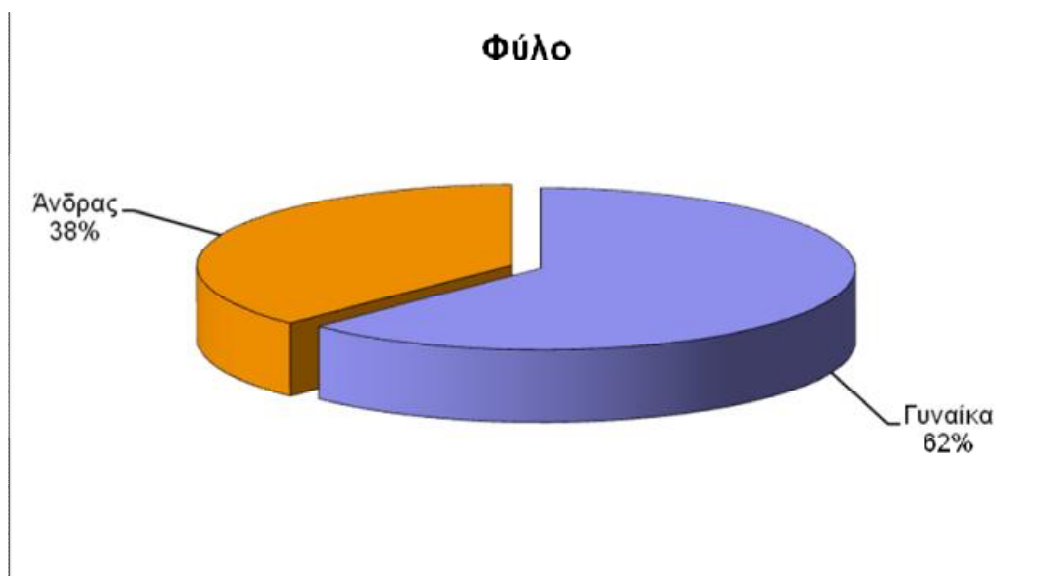
§ Έτος εισαγωγής (1999, 2000, 2001, 2002, 2003, 2004, 2005)

§ Εξάμηνο εισαγωγής (Χειμερινό εξάμηνο, Εαρινό εξάμηνο)

Η ανάλυση των δεδομένων έγινε με το στατιστικό πακέτο SPSS. Για την παρουσίαση των δημογραφικών χαρακτηριστικών χρησιμοποιήθηκαν μέτρα της περιγραφικής στατιστικής (συχνότητες, μέσοι όροι, crosstabs κλπ.) και διαγράμματα (ραβδογράμματα, ιστογράμματα κλπ.). Για την ανάλυση επιβίωσης χρησιμοποιήθηκε το μοντέλο COX Regression. Τονίζεται ότι σε όλα τα μοντέλα που ακολουθούν, το επίπεδο σημαντικότητας της στατιστικής ανάλυσης καθορίστηκε σε 0,05.

3.2 Παρουσίαση των δημογραφικών στοιχείων του δείγματος

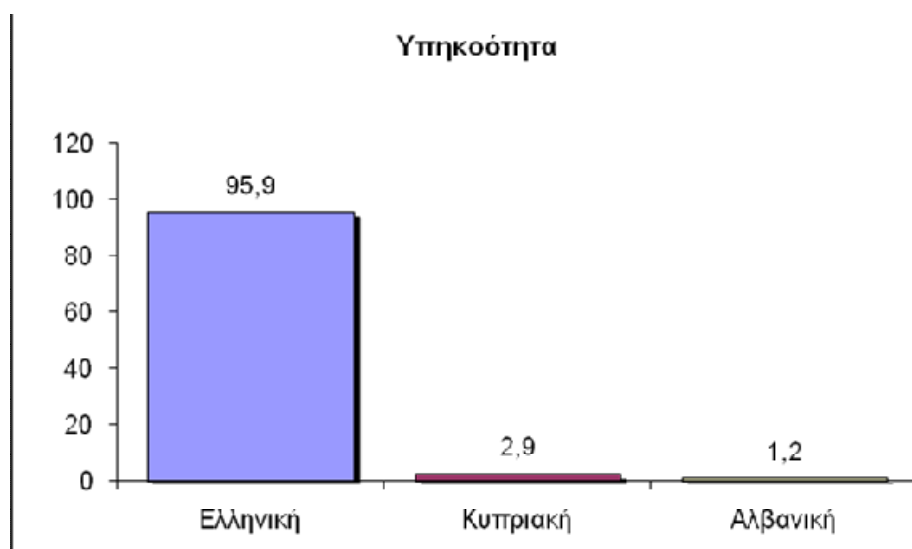
Ξεκινώντας την ανάλυση των δεδομένων, παραθέτουμε την ταυτότητα του δείγματος που θα χρησιμοποιήσουμε. Κατ' αρχήν, το δείγμα αποτελείται περισσότερο από γυναίκες και λιγότερο από άνδρες (διάγραμμα 2).



Διάγραμμα 2: Μεταβλητή «φύλο»

ΠΙΝΑΚΑΣ 1: ΦΥΛΟ	
	%
Άνδρας	37,6
Γυναίκα	62,4
Σύνολο	100,0

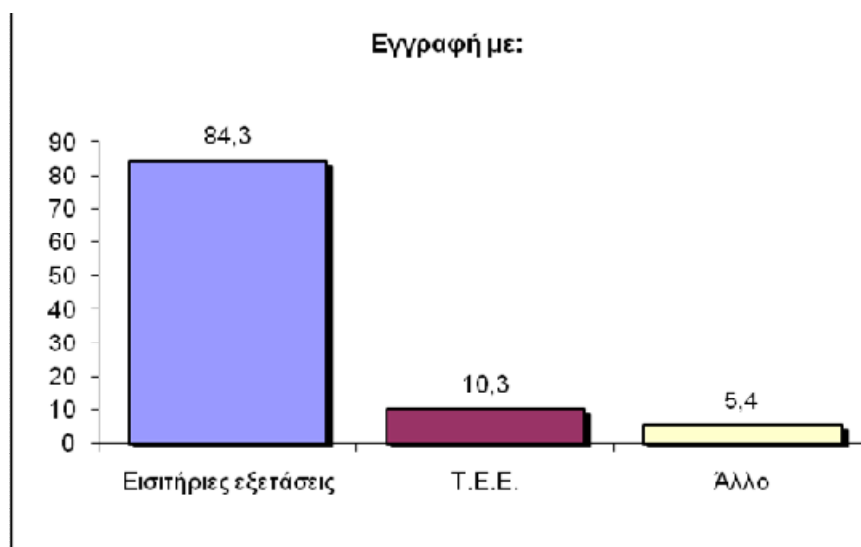
Η συντριπτική πλειοψηφία των φοιτητών του δείγματος έχουν Ελληνική υπηκοότητα. Υπάρχουν όμως και 3% φοιτητές από την Κύπρο αλλά και 1,2% φοιτητές από την Αλβανία (διάγραμμα 3).



Διάγραμμα 3: Μεταβλητή «υπηκοότητα»

ΠΙΝΑΚΑΣ 2: ΥΠΗΚΟΟΤΗΤΑ	
	%
Ελληνική	95,9
Κυπριακή	2,9
Αλβανική	1,2
Σύνολο	100,0

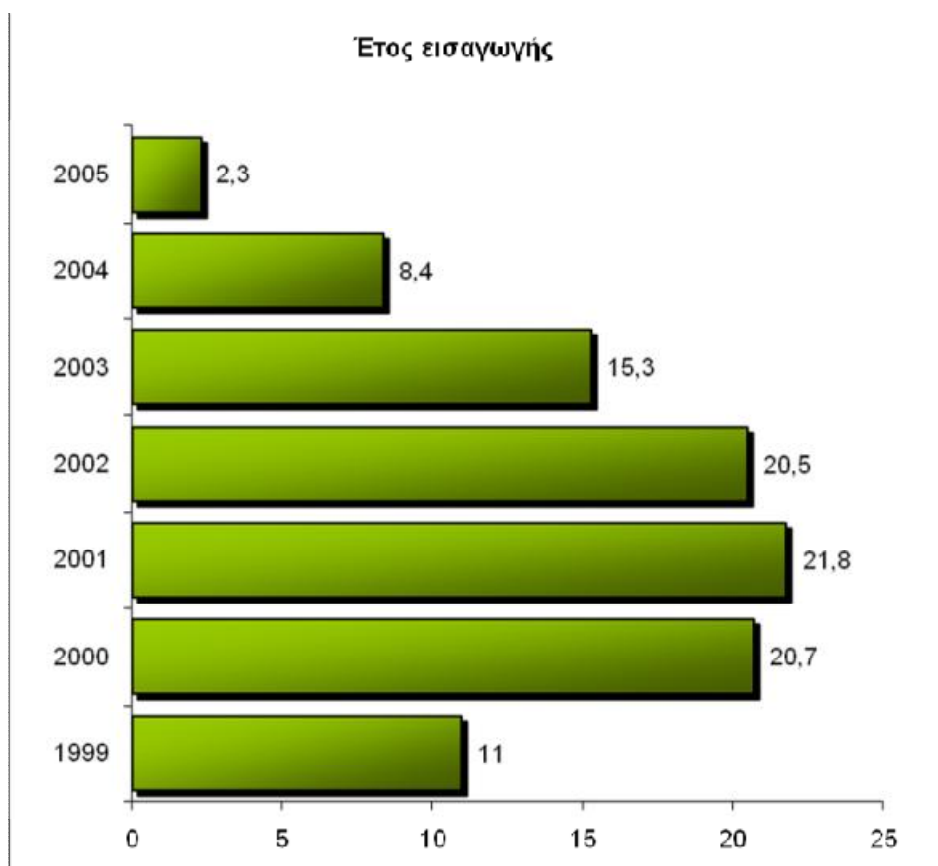
Οι περισσότεροι από τους φοιτητές ενεγράφησαν στο τμήμα δια μέσου εισιτηρίων εξετάσεων. Επίσης, σημαντικό είναι ότι 1 στους 10 εισακτέους προέρχονται από τα Τ.Ε.Ε. και περίπου 5% εισήχθησαν στο τμήμα με κάποιο άλλο τρόπο (κατάταξη, ομογενής, κύπριος ,κλπ.). Τα αποτελέσματα παρουσιάζονται στο διάγραμμα 4.



Διάγραμμα 4: Μεταβλητή «Εγγραφή»

ΠΙΝΑΚΑΣ 3: ΕΓΓΡΑΦΗ	
	%
Εισιτήριες εξετάσεις	84,3
Τ.Ε.Ε.	10,3
Άλλο	5,4
Σύνολο	100,0

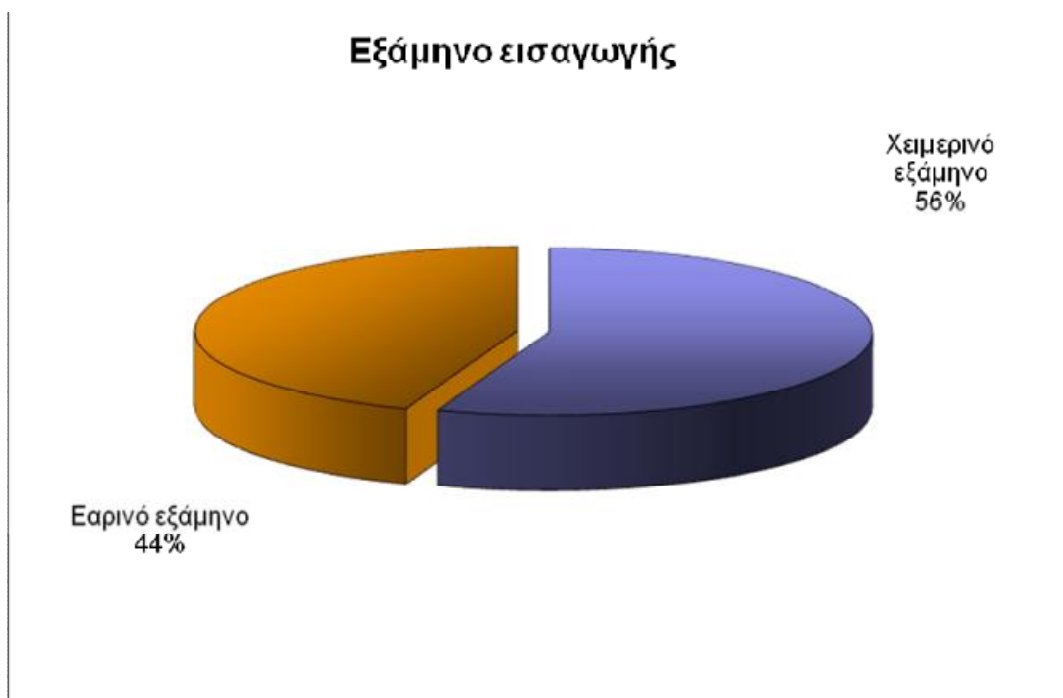
Τα περισσότερα μέλη του δείγματος εισήχθησαν στη σχολή τα έτη 2000, 2001 και 2002. Μάλιστα, οι εισακτέοι αυτών των ετών είναι ισομοιρασμένοι και καταλαμβάνουν περίπου το 60% του δείγματος. Οι εισακτέοι του 2005 είναι και οι λιγότεροι με μόλις 2,3% συμμετοχή (διάγραμμα 5).



Διάγραμμα 5: Μεταβλητή «Έτος Εισαγωγής»

ΠΙΝΑΚΑΣ 4: ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ	
	%
1999	11,0
2000	20,7
2001	21,8
2002	20,5
2003	15,3
2004	8,4
2005	2,3
Σύνολο	100,0

Αν και οι φοιτητές που ενεγράφησαν στο χειμερινό εξάμηνο είναι πλειοψηφία, οι φοιτητές του εαρινού εξαμήνου συμμετέχουν εξίσου σημαντικά στο δείγμα (διάγραμμα 6).

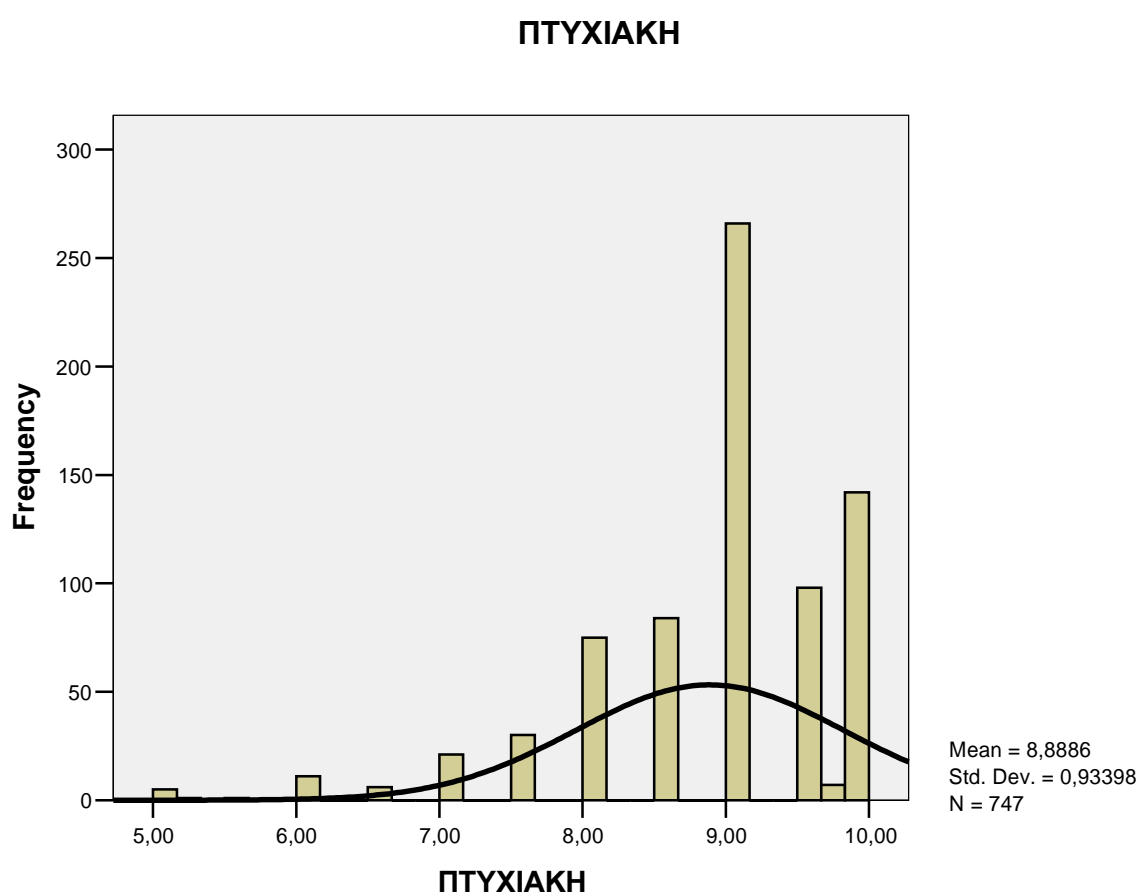


Διάγραμμα 6: Μεταβλητή «Εξάμηνο Εισαγωγής»

ΠΙΝΑΚΑΣ 5: ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ	
	%
Χειμερινό εξάμηνο	55,6
Εαρινό εξάμηνο	44,4
Σύνολο	100,0

Ο μεγαλύτερος βαθμός που έχει δοθεί σε πτυχιακή φοιτητή που συμμετέχει στο δείγμα είναι 10 και ο μικρότερος βαθμός είναι 5. Η μέση βαθμολογία των πτυχιακών είναι 8,89 (πίνακας 6). Από το διάγραμμα 7 που ακολουθεί είναι φανερό ότι οι περισσότερες βαθμολογίες συγκεντρώνονται στην περιοχή 8 έως 10.

ΠΙΝΑΚΑΣ 6: ΒΑΘΜΟΣ ΠΤΥΧΙΑΚΗΣ				
Μέσος	Διάμεσος	Τυπική απόκλιση	Min	Max
8,89	9,00	0,93	5,00	10,00



Διάγραμμα 7: Ιστόγραμμα μεταβλητής «Βαθμός Πτυχιακής»

Αναλύοντας περαιτέρω βλέπουμε ότι οι γυναίκες έχουν ελαφρώς μεγαλύτερο βαθμό πτυχιακής από τους άνδρες (πίνακας 7).

ΠΙΝΑΚΑΣ 7: ΜΕΣΟΣ ΒΑΘΜΟΣ ΠΤΥΧΙΑΚΗΣ ΑΝΑ ΦΥΛΟ	
Άνδρας	Γυναίκα
8,84	8,92

Επίσης, βλέπουμε στον πίνακα 8 ότι τους μεγαλύτερους μέσους βαθμούς πτυχιακών έχουν οι εισακτέοι των ετών 2004 και 2002 (με ελάχιστη, πάντως, διαφορά).

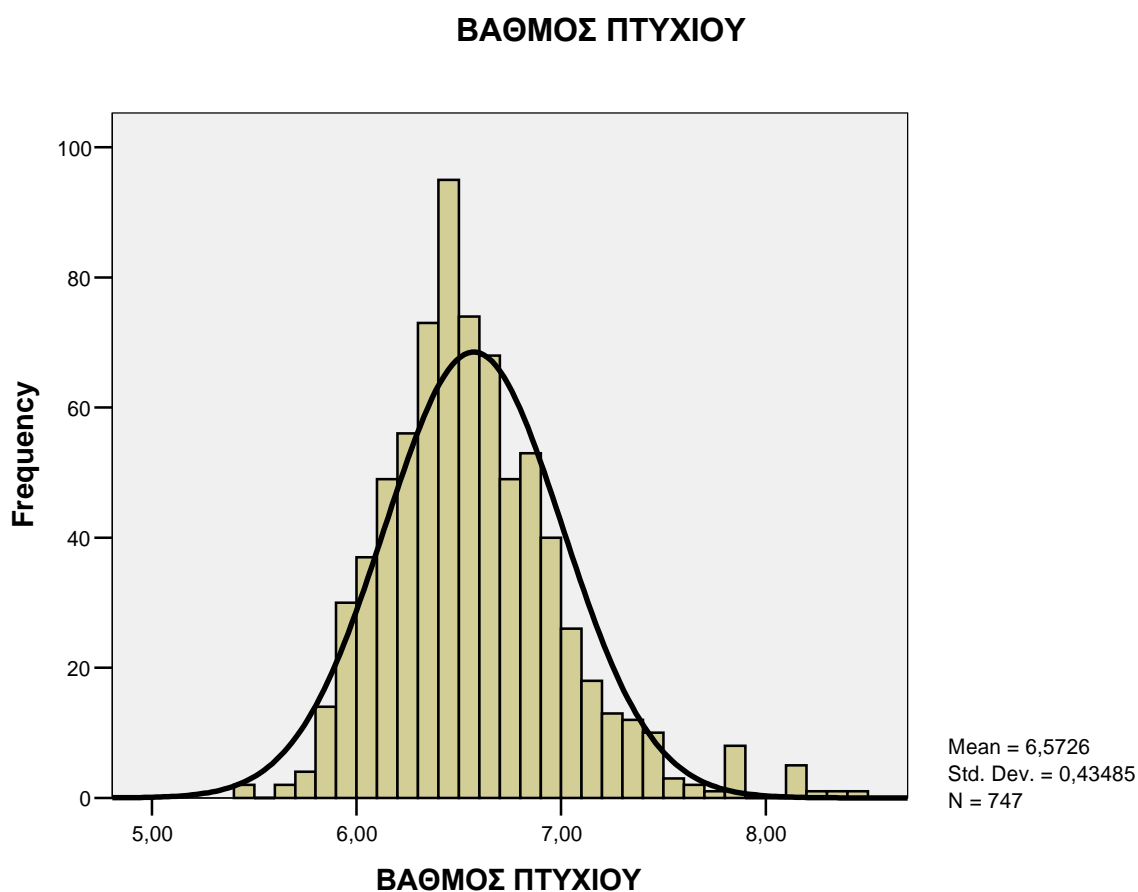
ΠΙΝΑΚΑΣ 8: ΜΕΣΟΣ ΒΑΘΜΟΣ ΠΤΥΧΙΑΚΗΣ ΑΝΑ ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ						
1999	2000	2001	2002	2003	2004	2005
8,77	8,91	8,76	9,00	8,85	9,15	8,85

Τέλος, οι εισακτέοι του εαρινού εξαμήνου φαίνεται να έχουν ελάχιστα μεγαλύτερες βαθμολογίες στις πτυχιακές τους από τους εισακτέους του χειμερινού εξαμήνου (πίνακας 9).

ΠΙΝΑΚΑΣ 9: ΜΕΣΟΣ ΒΑΘΜΟΣ ΠΤΥΧΙΑΚΗΣ ΑΝΑ ΕΞΑΜΗΝΟ ΕΓΓΡΑΦΗΣ	
Χειμερινό εξάμηνο	Εαρινό εξάμηνο
8,87	8,92

Από την άλλη μεριά, η μέση βαθμολογία πτυχίου για τους συμμετέχοντες στο δείγμα είναι 6,57. Ο μεγαλύτερος βαθμός που έχει επιτευχθεί είναι 8,43 και ο μικρότερος βαθμός 5,46 (πίνακας 10). Από το διάγραμμα 8 που ακολουθεί, βλέπουμε ότι οι βαθμολογίες ακολουθούν κανονική κατανομή και είναι συγκεντρωμένες κυρίως μεταξύ 6 και 7.

ΠΙΝΑΚΑΣ 10: ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ				
Μέσος	Διάμεσος	Τυπική απόκλιση	Min	Max
6,57	6,52	0,43	5,46	8,43



Διάγραμμα 8: Ιστόγραμμα μεταβλητής «Βαθμός Πτυχίου»

Αναλύοντας περαιτέρω βλέπουμε ότι οι γυναίκες έχουν ελαφρώς μεγαλύτερο μέσο βαθμό πτυχίου από τους άνδρες (πίνακας 11).

ΠΙΝΑΚΑΣ 11: ΜΕΣΟΣ ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ ΑΝΑ ΦΥΛΟ	
Άνδρας	Γυναίκα
6,52	6,61

Επίσης, στον πίνακα 12 βλέπουμε ότι τους μεγαλύτερους μέσους βαθμούς πτυχίου έχουν οι εισακτέοι του έτους 2005 (με μικρή, πάντως, διαφορά).

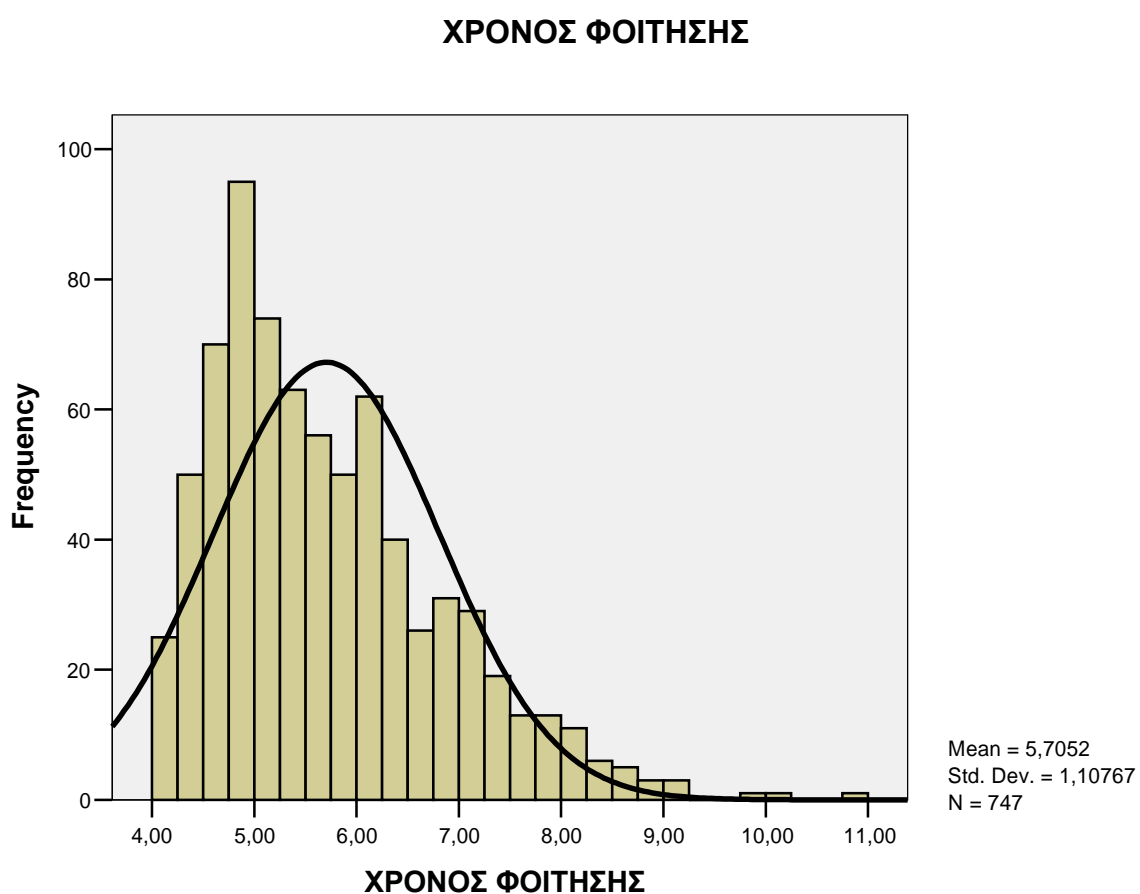
ΠΙΝΑΚΑΣ 12: ΜΕΣΟΣ ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ ΑΝΑ ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ						
1999	2000	2001	2002	2003	2004	2005
6,64	6,60	6,58	6,58	6,42	6,56	6,92

Τέλος, οι εισακτέοι του χειμερινού εξαμήνου φαίνεται να έχουν ελάχιστα μεγαλύτερο μέσο βαθμό πτυχίου από τους εισακτέους του εαρινού εξαμήνου (πίνακας 13).

ΠΙΝΑΚΑΣ 13: ΜΕΣΟΣ ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ ΑΝΑ ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ	
Χειμερινό εξάμηνο	Εαρινό εξάμηνο
6,60	6,54

Υπάρχουν φοιτητές στο δείγμα μας που έχουν τελειώσει τη σχολή ακριβώς σε 4 χρόνια, αλλά και φοιτητές που έχουν ξεπεράσει και τα 10 χρόνια. Η μέση διάρκεια φοίτησης, πάντως, είναι 5,7 χρόνια (πίνακας 14). Το ακόλουθο διάγραμμα 9 μας δείχνει ότι οι περισσότεροι φοιτητές τελειώνουν μέσα σε 4 με 7 χρόνια.

ΠΙΝΑΚΑΣ 14: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ				
Μέσος	Διάμεσος	Τυπική απόκλιση	Min	Max
5,70	5,40	1,11	4,00	10,80



Διάγραμμα 9: Ιστόγραμμα μεταβλητής «Χρόνος Φοίτησης»

Αναλύοντας περαιτέρω βλέπουμε ότι οι γυναίκες υπερτερούν και στο χρόνο φοίτησης αφού είναι ελαφρώς μικρότερος από το χρόνο φοίτησης των ανδρών (πίνακας 15).

ΠΙΝΑΚΑΣ 15: ΜΕΣΟΣ ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ ΑΝΑ ΦΥΛΟ	
Άνδρας	Γυναίκα
5,97	5,55

Επίσης, βλέπουμε ότι οι εισακτέοι των τελευταίων ετών έχουν μικρότερο μέσο χρόνο φοίτησης από του εισακτέους των αρχικών ετών. Μάλιστα, όσο προχωράμε από το 1999 προς το 2005 μειώνεται συνεχώς και ο χρόνος φοίτησης (πίνακας 16).

ΠΙΝΑΚΑΣ 16: ΜΕΣΟΣ ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ ΑΝΑ ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ						
1999	2000	2001	2002	2003	2004	2005
6,00	5,79	5,83	5,78	5,64	5,05	4,55

Τέλος, οι εισακτέοι του χειμερινού εξαμήνου φαίνεται να έχουν ελάχιστα μικρότερο μέσο χρόνο φοίτησης από τους εισακτέους του εαρινού εξαμήνου (πίνακας 17).

ΠΙΝΑΚΑΣ 17: ΜΕΣΟΣ ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ ΑΝΑ ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ	
Χειμερινό εξάμηνο	Εαρινό εξάμηνο
5,65	5,78

3.3 COX Regression μονού παράγοντα

3.3.1 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΦΥΛΟ

Επιλέγουμε την ποιοτική μεταβλητή «ΦΥΛΟ» ως πρώτο παράγοντα για να μας βοηθήσει να ερμηνεύσουμε τη συμπεριφορά της εξαρτημένης μεταβλητής «ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ». Προκειμένου να εκτελέσουμε τη στατιστική διαδικασία COX REGRESSION επιλέξαμε τις ακόλουθες μεταβλητές στο SPSS (SPSS- PASW statistics 18, 2009):

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΦΥΛΟ (κατηγορική μεταβλητή: Άνδρας-Γυναίκα)

Από την εκτέλεση της διαδικασίας προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

		N	Percent
Cases available in analysis	Event ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
	Total	747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 18: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 18 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα

μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Ο επόμενος πίνακας 19 απεικονίζει την ψευδομεταβλητή που δημιουργήθηκε για την εισαγωγή της κατηγορικής μεταβλητής στο μοντέλο. Σύμφωνα με αυτόν, η κατηγορία αναφοράς είναι οι άνδρες.

Categorical Variable Codings^b

		Frequency	(1)
sex ^a	1=ΑΝΔΡΑΣ	281	0
	2=ΓΥΝΑΙΚΑ	466	1

a. Indicator Parameter Coding

b. Category variable: sex (ΦΥΛΟ)

Πίνακας 19: Κωδικοποίηση κατηγορικής μεταβλητής

Αμέσως παρακάτω απεικονίζεται ο πίνακας 20 με τον έλεγχο στατιστικής σημαντικότητας της ανεξάρτητης μεταβλητής. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο. Βλέπουμε ότι αφού το $\text{sig.} = 0,000 < 0,05$, έστω και ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0. Βέβαια, στο μοντέλο μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, η οποία φαίνεται ότι είναι στατιστικά σημαντική.

Omnibus Tests of Model Coefficients^b

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8445,254	21,596	1	,000	21,914	1	,000	21,914	1	,000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168

b. Beginning Block Number 1. Method = Enter

Πίνακας 20: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 21 που απεικονίζει, εκτός των άλλων, τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Η μεταβλητή «ΦΥΛΟ» φαίνεται ότι είναι στατιστικά σημαντική, σε επίπεδο 5%, αφού το sig. = 0,000 < 0.05.

Άλλο στοιχείο του πίνακα είναι η τιμή της στατιστικής συνάρτησης WALD = 21,386 σύμφωνα με την οποία γίνεται ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή B βάσει της Χ-τετράγωνο κατανομής. Η τιμή του συντελεστή B είναι 0,354, ενώ η τιμή του Exp(B) = 1,425. Τέλος, το Exp(B) κυμαίνεται μεταξύ 1,226 και 1,656 με 95% εμπιστοσύνη.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
sex	,354	,077	21,386	1	,000	1,425	1,226	1,656

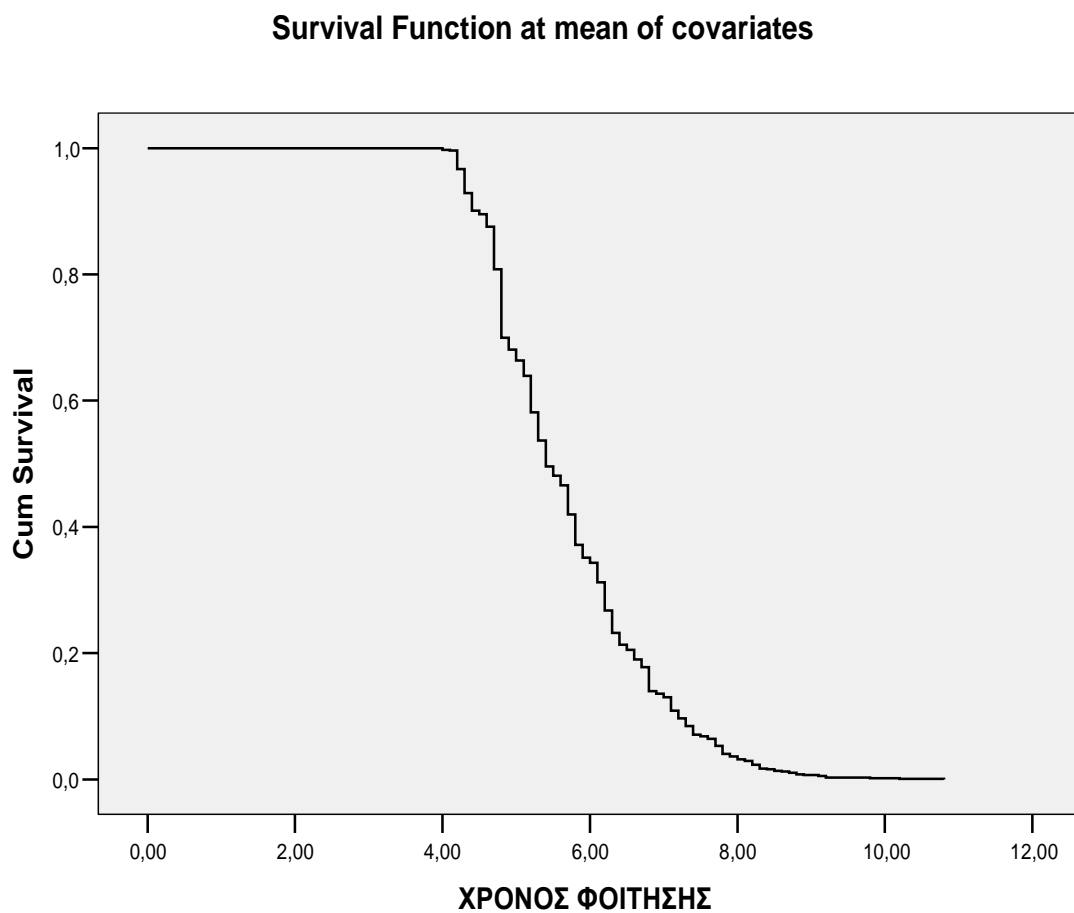
Πίνακας 21: Έλεγχος σημαντικότητας για την ανεξάρτητη μεταβλητή

Η στήλη Exp(B) εκφράζει την εκτιμώμενη αλλαγή του «κινδύνου» σε μια μονάδα αλλαγής της ανεξάρτητης μεταβλητής όταν οι υπόλοιπες παραμένουν σταθερές.

Στην προκειμένη περίπτωση έχουμε μόνο μια ανεξάρτητη μεταβλητή που είναι και ποιοτική. Οπότε η στήλη Exp(B) εκφράζει την αλλαγή επικινδυνότητας από την κατηγορία αναφοράς στην επόμενη κατηγορία. Εδώ βλέπουμε ότι ο συντελεστής (Hazard Ratio, HR) είναι HR=1,425, άρα η επικινδυνότητα να συντελεστεί το γεγονός στις γυναίκες είναι 1,425 φορές περισσότερη από ότι στους άνδρες. Επομένως, οι γυναίκες παρουσιάζουν μεγαλύτερο ρίσκο και χαμηλότερο δείκτη επιβίωσης.

Όλα τα προηγούμενα σημαίνουν ότι οι άνδρες έχουν μεγαλύτερους χρόνους φοίτησης από τις γυναίκες.

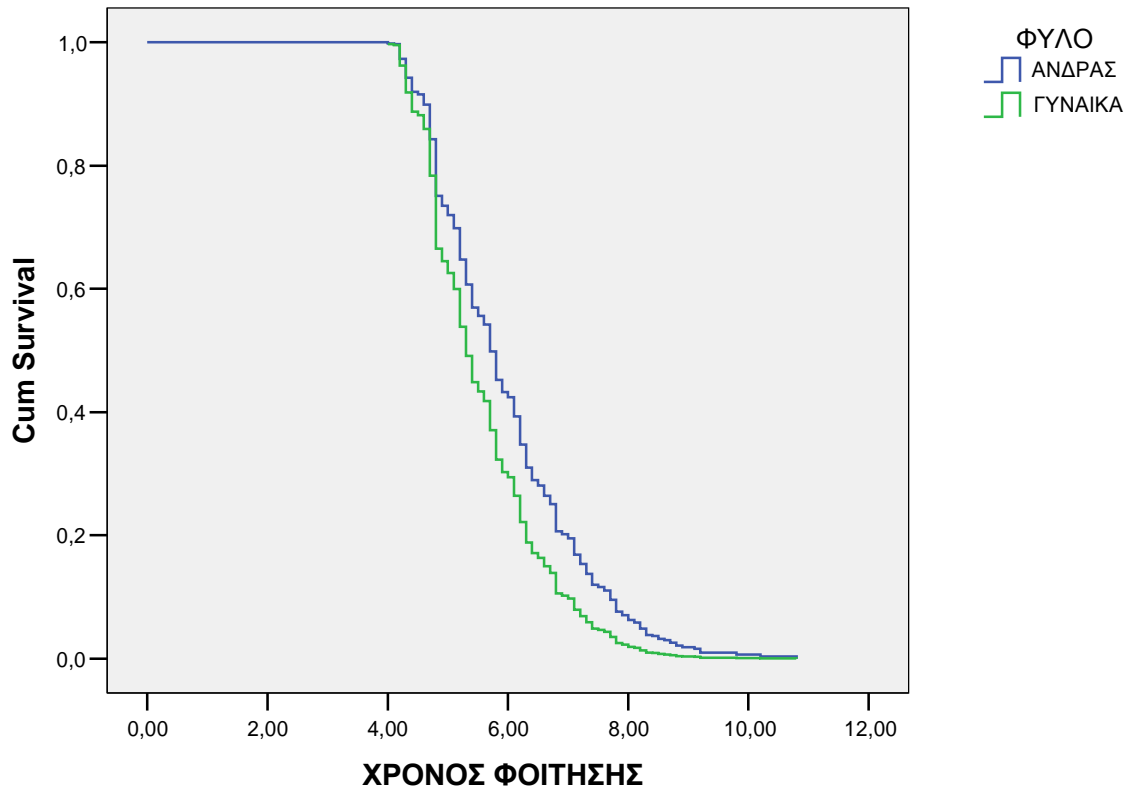
Η συνολική συνάρτηση επιβίωσης για το μοντέλο μας παρουσιάζεται στο παρακάτω διάγραμμα 10. Βλέπουμε ότι μετά τα 4 χρόνια, που είναι και ο ελάχιστος χρόνος φοίτησης, η πιθανότητα να μην έχει αποφοιτήσει κάποιος φοιτητής μειώνεται σχεδόν κατακόρυφα. Μάλιστα, λίγο μετά τα 8 χρόνια η πιθανότητα αυτή σχεδόν μηδενίζεται.



Διάγραμμα 10: Συνολική συνάρτηση επιβίωσης

Στο διάγραμμα 11 παρουσιάζονται, σε δύο διαφορετικές καμπύλες, οι συναρτήσεις επιβίωσης για τις γυναίκες και τους άνδρες. Οι γυναίκες παρουσιάζουν μεγαλύτερη «επικινδυνότητα» να αποφοιτήσουν και η γραμμή τους έχει χαμηλότερες τιμές για την αθροιστική συνάρτηση επιβίωσης.

Survival Function for patterns 1 - 2

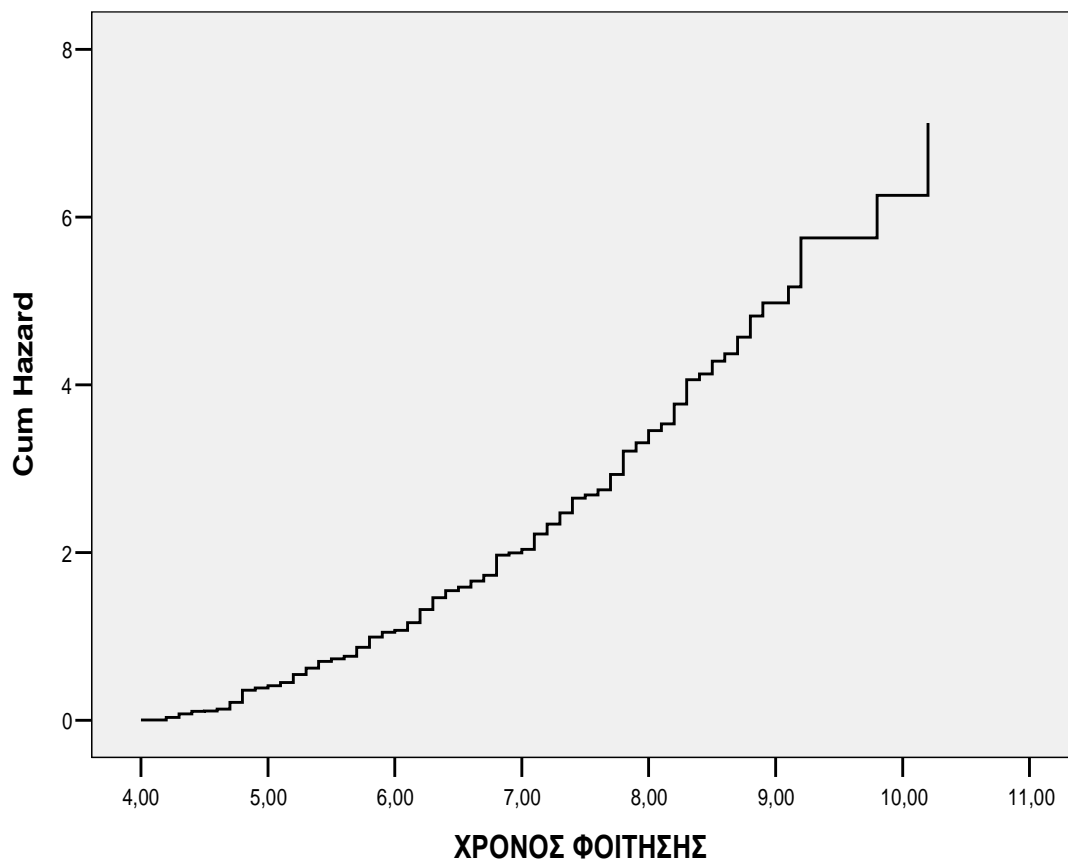


Διάγραμμα 11: Συνάρτηση επιβίωσης ανά φύλο

Οι άνδρες, από την άλλη μεριά, φαίνεται ότι παραμένουν περισσότερο χρόνο στη σχολή μέχρι να καταφέρουν να αποφοιτήσουν.

Ακριβώς ίδια συμπεράσματα βγαίνουν και από το γράφημα της συνάρτησης επικινδυνότητας (διάγραμμα 12). Μετά από τα 4 χρόνια φοίτησης, αυξάνεται και η πιθανότητα να αποφοιτήσει κάποιος.

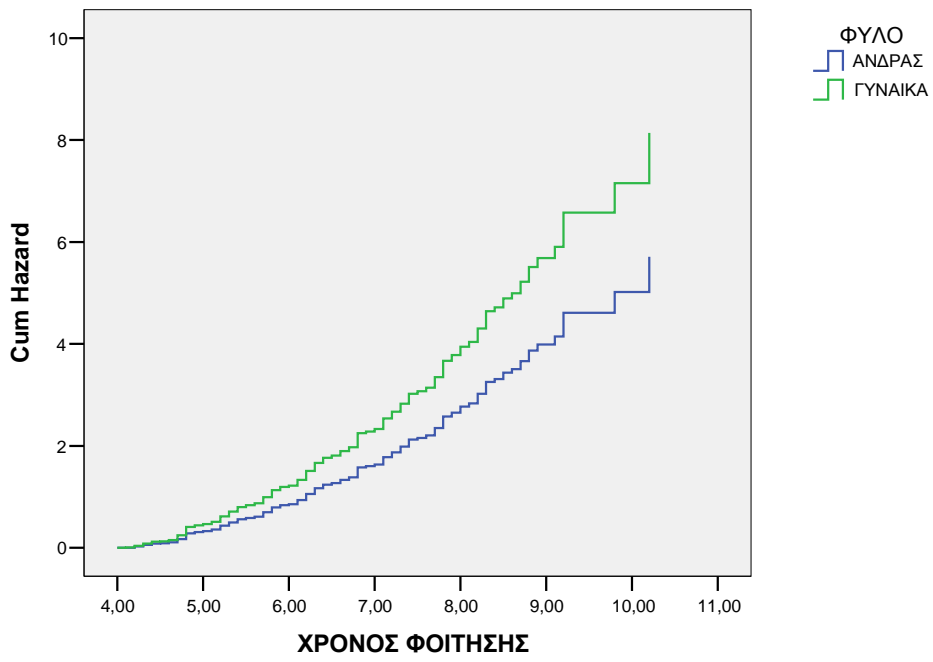
Hazard Function at mean of covariates



Διάγραμμα 12: Συνολική συνάρτηση επικινδυνότητας

Τέλος, παρουσιάζεται η συνάρτηση επικινδυνότητας για κάθε φύλο ξεχωριστά (διάγραμμα 13). Είναι φανερό ότι για τις γυναίκες υπάρχει μεγαλύτερο «κίνδυνο» να αποφοιτήσουν αφού η συνάρτησή τους έχει μεγαλύτερες τιμές.

Hazard Function for patterns 1 - 2



Διάγραμμα 13: Συνάρτηση επικινδυνότητας ανά φύλο

3.3.2 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΥΠΗΚΟΟΤΗΤΑ

Επιλέγουμε την ποιοτική μεταβλητή «ΥΠΗΚΟΟΤΗΤΑ» ως δεύτερο παράγοντα για να μας βοηθήσει να ερμηνεύσουμε τη συμπεριφορά της εξαρτημένης μεταβλητής «ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ». Προκειμένου να εκτελέσουμε τη στατιστική διαδικασία COX REGRESSION επιλέξαμε τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΥΠΗΚΟΟΤΗΤΑ (κατηγορική μεταβλητή: Ελλάδα-Κύπρος-Αλβανία)

Από την εκτέλεση της διαδικασίας προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

		N	Percent
Cases available in analysis	Event ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
Total		747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 22: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 22 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα

μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Ο πίνακας 23 απεικονίζει την ψευδομεταβλητή που δημιουργήθηκε για την εισαγωγή της κατηγορικής μεταβλητής στο μοντέλο. Σύμφωνα με αυτόν, η κατηγορία αναφοράς είναι η Αλβανία.

Categorical Variable Coding^b

		Frequency	(1)	(2)
origin ^a	1=ΕΛΛΑΔΑ	716	1	0
	2=ΚΥΠΡΟΣ	22	0	1
	3=ΑΛΒΑΝΙΑ	9	0	0

a. Indicator Parameter Coding

b. Category variable: origin (ΥΠΗΚΟΟΤΗΤΑ)

Πίνακας 23: Κωδικοποίηση κατηγορικής μεταβλητής

Αμέσως παρακάτω απεικονίζεται ο πίνακας 24 με τον έλεγχο για το αν έστω και μια ανεξάρτητη μεταβλητή του μοντέλου είναι στατιστικά σημαντική. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο. Βλέπουμε ότι αφού το $\text{sig.} = 0,725 > 0.05$, οι συντελεστές όλων των ανεξάρτητων μεταβλητών είναι ίσοι με το 0. Βέβαια, στο μοντέλο μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, η οποία φαίνεται ότι δεν είναι στατιστικά σημαντική.

Omnibus Tests of Model Coefficients

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8466,492	,642	2	,725	,676	2	,713	,676	2	,713

- a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168
 b. Beginning Block Number 1. Method = Enter

Πίνακας 24: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 25 που απεικονίζει, εκτός των άλλων, τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Η μεταβλητή «ΥΠΗΚΟΟΤΗΤΑ» φαίνεται ότι δεν είναι στατιστικά σημαντική, σε επίπεδο 5%, αφού το sig. = 0,726 > 0,05.

Άλλο στοιχείο του πίνακα είναι η τιμή της στατιστικής συνάρτησης WALD = 0,641 σύμφωνα με την οποία γίνεται ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή Β βάσει της Χ-τετράγωνο κατανομής.

Η τιμή του συντελεστή Β είναι 0,037 για τους Έλληνες και -0,136 για τους Κύπριους, ενώ η τιμή του Exp(B) = 1,037 για τους Έλληνες και 0,873 για τους Κύπριους. Τέλος, το Exp(B) κυμαίνεται μεταξύ 0,537 και 2,006 για τους Έλληνες και μεταξύ 0,401 και 1,901 για τους Κύπριους, με 95% εμπιστοσύνη.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
origin			,641	2	,726			
origin(1)	,037	,336	,012	1	,913	1,037	,537	2,006
origin(2)	-,136	,397	,117	1	,733	,873	,401	1,901

Πίνακας 25: Έλεγχος σημαντικότητας για την ανεξάρτητη μεταβλητή

Covariate Means and Pattern Values

	Mean	Pattern		
		1	2	3
origin(1)	,959	1,000	,000	,000
origin(2)	,029	,000	1,000	,000

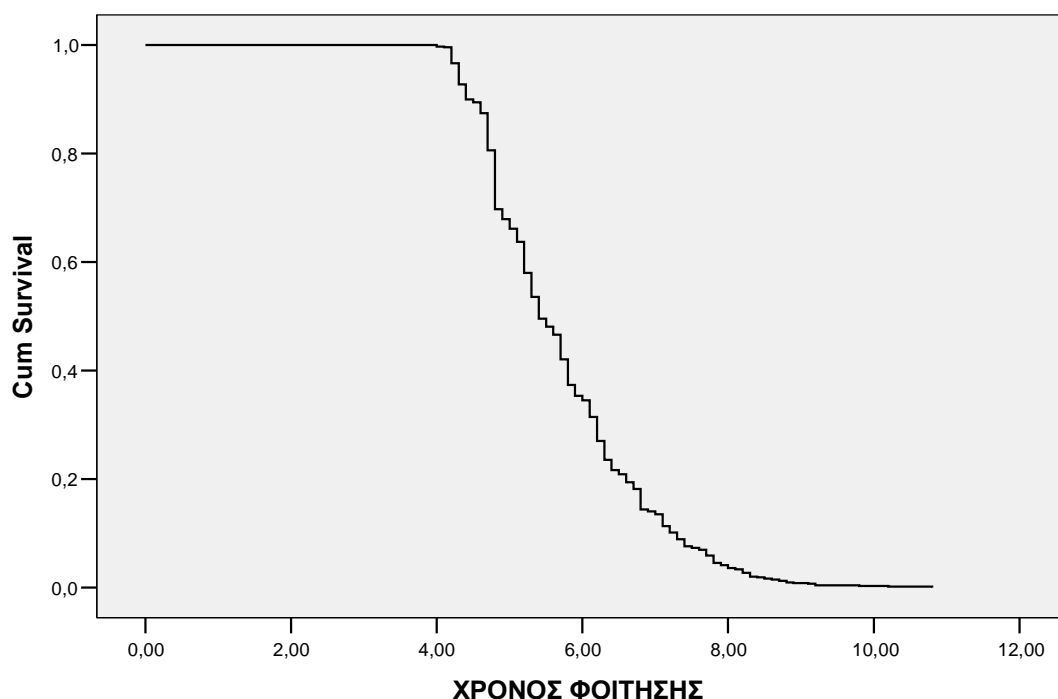
Πίνακας 26: Μέσοι όροι των ανεξάρτητων μεταβλητών

Στην περίπτωση μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, που είναι ποιοτική με τρεις κατηγορίες. Οπότε η στήλη Exp(B) εκφράζει την αλλαγή επικινδυνότητας από την κατηγορία αναφοράς (Αλβανοί φοιτητές) στις άλλες δύο κατηγορίες (Έλληνες και Κύπριοι). Εδώ βλέπουμε ότι ο συντελεστής είναι $HR=1,037$ για τους Έλληνες φοιτητές, άρα η επικινδυνότητα να συντελεστεί το γεγονός σε αυτούς είναι 1,037 φορές περισσότερη από ότι στους Αλβανούς. Επίσης, ο συντελεστής είναι 0,873 για τους Κύπριους φοιτητές, άρα η επικινδυνότητα να συντελεστεί το γεγονός σε αυτούς είναι λιγότερη από ότι στους Αλβανούς.

Όλα τα προηγούμενα δείχνουν ότι οι Έλληνες έχουν μικρότερους χρόνους φοίτησης από τους Αλβανούς, ενώ οι Κύπριοι μεγαλύτερους. Όμως οι διαφορές δεν είναι στατιστικά σημαντικές.

Η συνολική συνάρτηση επιβίωσης για το μοντέλο μας παρουσιάζεται στο παρακάτω διάγραμμα 14. Βλέπουμε ότι μετά τα 4 χρόνια, που είναι και ο ελάχιστος χρόνος φοίτησης, μειώνεται σχεδόν κατακόρυφα η πιθανότητα να μην έχει αποφοιτήσει κάποιος φοιτητής. Μάλιστα, λίγο μετά τα 8 χρόνια η πιθανότητα αυτή σχεδόν μηδενίζεται.

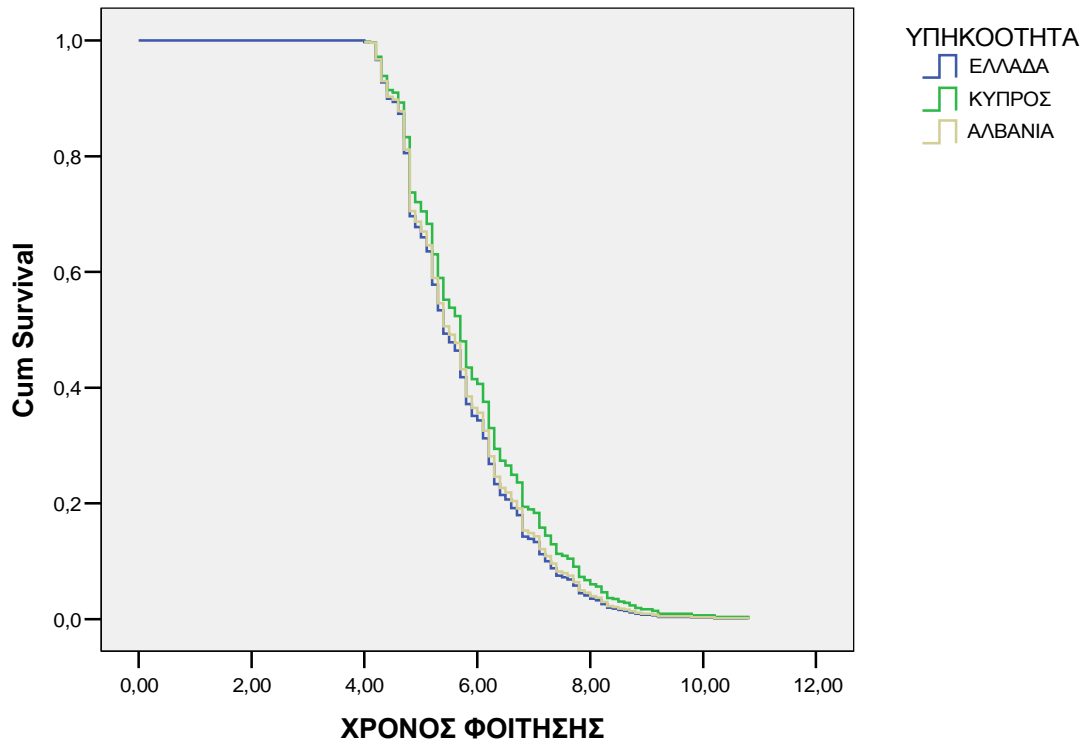
Survival Function at mean of covariates



Διάγραμμα 14: Συνολική συνάρτηση επιβίωσης

Στο διάγραμμα 15 παρουσιάζονται, σε τρεις διαφορετικές γραμμές, οι συναρτήσεις επιβίωσης για τους Έλληνες, τους Κύπριους και τους Αλβανούς φοιτητές. Οι Έλληνες παρουσιάζουν μεγαλύτερη «επικινδυνότητα» να αποφοιτήσουν και η γραμμή τους έχει χαμηλότερες τιμές για την αθροιστική συνάρτηση επιβίωσης. Ακολουθούν οι Αλβανοί φοιτητές και τέλος οι Κύπριοι που παρουσιάζουν και τις υψηλότερες τιμές επιβίωσης (φοίτησης).

Survival Function for patterns 1 - 3

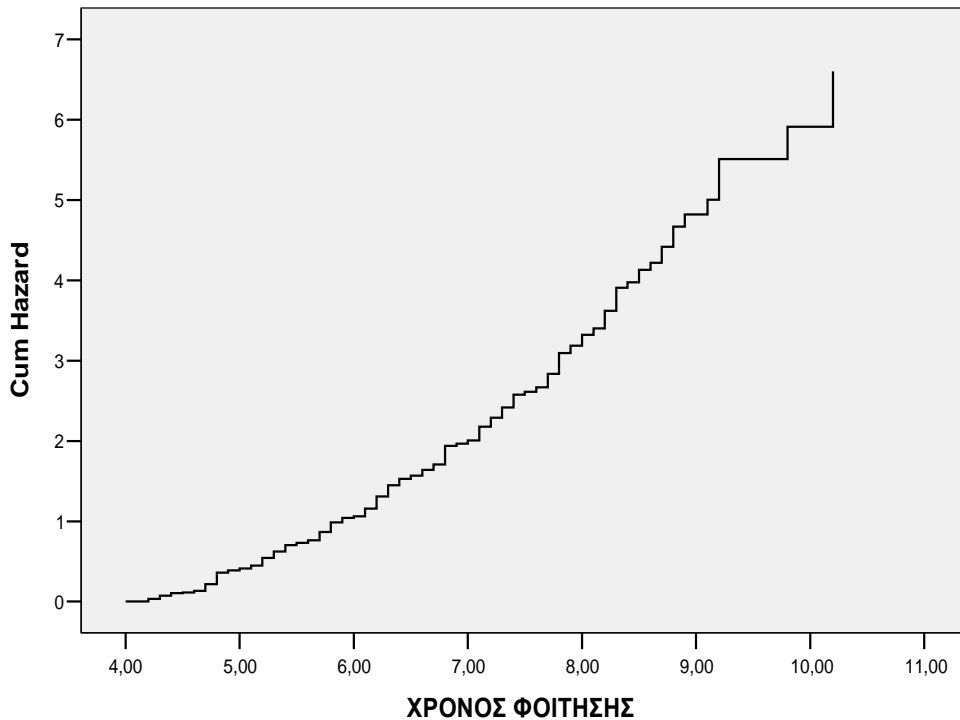


Διάγραμμα 15: Συνάρτηση επιβίωσης ανά υπηκοότητα

Το διάγραμμα επιβεβαιώνει τα προηγούμενα συμπεράσματα, δηλαδή δεν φαίνεται να υπάρχουν ουσιαστικές διαφορές για τις τρεις κατηγορίες φοιτητών.

Ακριβώς ίδια συμπεράσματα βγαίνουν και από το γράφημα της συνάρτησης επικινδυνότητας (διάγραμμα 16). Μετά από τα 4 χρόνια φοίτησης, αυξάνεται και η πιθανότητα να αποφοιτήσει κάποιος.

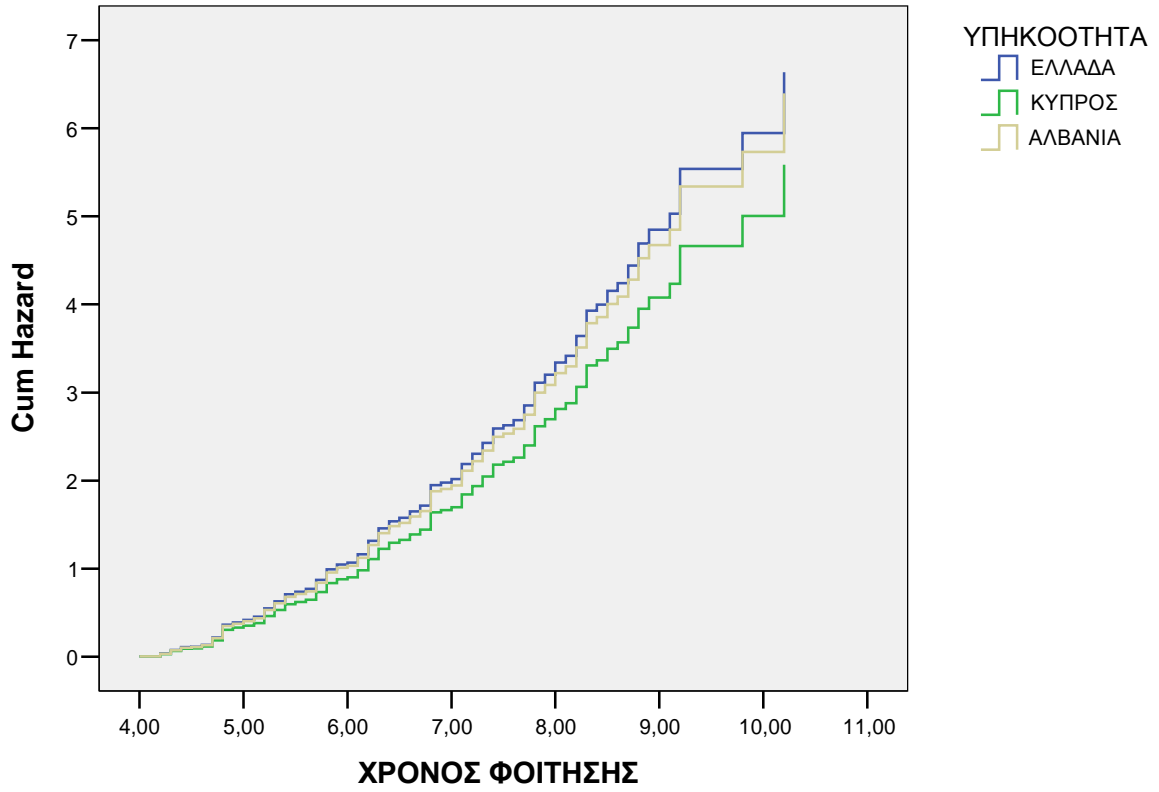
Hazard Function at mean of covariates



Διάγραμμα 16: Συνολική συνάρτηση επικινδυνότητας

Τέλος, παρουσιάζεται η συνάρτηση επικινδυνότητας για κάθε Υπηκοότητα ξεχωριστά (διάγραμμα 17). Είναι φανερό ότι για τους Έλληνες υπάρχει μεγαλύτερο «κίνδυνο» να αποφοιτήσουν αφού η συνάρτησή τους έχει μεγαλύτερες τιμές. Οι διαφορές όμως είναι αμελητέες για τα τρία γκρουπ.

Hazard Function for patterns 1 - 3



Διάγραμμα 17: Συνάρτηση επικινδυνότητας ανά υπηκοότητα

3.3.3 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΕΓΓΡΑΦΗ ΜΕ

Επιλέγουμε την ποιοτική μεταβλητή «ΕΓΓΡΑΦΗ ΜΕ» ως τρίτο παράγοντα για να μας βοηθήσει να ερμηνεύσουμε τη συμπεριφορά της εξαρτημένης μεταβλητής «ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ». Προκειμένου να εκτελέσουμε τη στατιστική διαδικασία COX REGRESSION επιλέξαμε τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΕΓΓΡΑΦΗ ΜΕ (κατηγορική μεταβλητή: Εισιτήριες εξετάσεις-
ΤΕΕ - Άλλο)

Από την εκτέλεση της διαδικασίας στο SPSS προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

Case Processing Summary		N	Percent
Cases available in analysis	Event ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
Total		747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 27: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 27 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα

μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Ο επόμενος πίνακας 28 απεικονίζει την ψευδομεταβλητή που δημιουργήθηκε για την εισαγωγή της κατηγορικής μεταβλητής στο μοντέλο. Σύμφωνα με αυτόν, η κατηγορία αναφοράς είναι η «Εισιτήριες εξετάσεις».

Categorical Variable Coding^b

	Frequency	(1)	(2)
regist ^a 1=ΕΙΣΙΤΗΡΙΕΣ ΕΞΕΤΑΣΕΙΣ	630	0	0
2=Τ.Ε.Ε.	77	1	0
3=ΑΛΛΟ	40	0	1

a. Indicator Parameter Coding

b. Category variable: regist (ΕΓΓΡΑΦΗ ΜΕ)

Πίνακας 28: Κωδικοποίηση κατηγορικής μεταβλητής

Αμέσως παρακάτω απεικονίζεται ο πίνακας 29 με τον έλεγχο για το αν έστω και μια ανεξάρτητη μεταβλητή του μοντέλου είναι στατιστικά σημαντική. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο. Βλέπουμε ότι αφού το $\text{sig.} = 0,431 > 0,05$, οι συντελεστές όλων των ανεξάρτητων μεταβλητών είναι ίσοι με το 0. Βέβαια, στο μοντέλο μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, η οποία φαίνεται ότι δεν είναι στατιστικά σημαντική.

Omnibus Tests of Model Coefficients^b

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8465,412	1,682	2	,431	1,756	2	,416	1,756	2	,416

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168

b. Beginning Block Number 1. Method = Enter

Πίνακας 29: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 30 που απεικονίζει τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Η μεταβλητή «ΕΓΓΡΑΦΗ ΜΕ» φαίνεται ότι είναι δεν είναι στατιστικά σημαντική, σε επίπεδο 5%, αφού το sig. = 0,432 > 0,05.

Άλλο στοιχείο του πίνακα είναι η τιμή της στατιστικής συνάρτησης WALD = 1,677 σύμφωνα με την οποία γίνεται ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή B βάσει της Χ-τετράγωνο κατανομής.

Η τιμή του συντελεστή B είναι -0,072 για τους εγγεγραμμένους μέσω ΤΕΕ και -0,196 για τους εγγεγραμμένους με άλλο τρόπο. Οι αντίστοιχες τιμές του Exp(B) είναι HR=0,931 και HR=0,822. Τέλος, το Exp(B) κυμαίνεται μεταξύ 0,734 και 1,180 για τους εγγεγραμμένους που προέρχονται από τα ΤΕΕ και μεταξύ 0,597 και 1,133 για τους εγγεγραμμένους με άλλο τρόπο, με 95% εμπιστοσύνη.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
regist			1,677	2	,432			
regist(1)	-,072	,121	,351	1	,554	,931	,734	1,180
regist(2)	-,196	,163	1,434	1	,231	,822	,597	1,133

Πίνακας 30: Έλεγχος σημαντικότητας για την ανεξάρτητη μεταβλητή

Covariate Means and Pattern Values

	Mean	Pattern		
		1	2	3
regist(1)	,843	1,000	,000	,000
regist(2)	,103	,000	1,000	,000

Πίνακας 31: Μέσοι όροι των ανεξάρτητων μεταβλητών

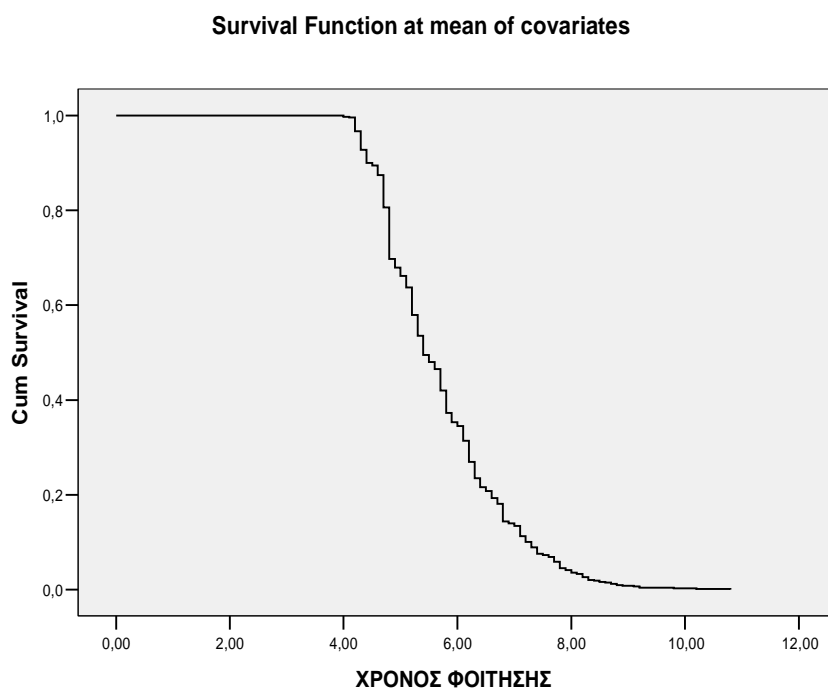
Στην περίπτωση μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, που είναι ποιοτική με τρεις κατηγορίες. Οπότε η στήλη Exp(B) εκφράζει την αλλαγή επικινδυνότητας από την κατηγορία αναφοράς (Εισιτήριες εξετάσεις) στις

άλλες δύο κατηγορίες (εγγεγραμμένοι από ΤΕΕ και εγγεγραμμένοι με άλλο τρόπο). Εδώ βλέπουμε ότι ο συντελεστής είναι $HR=0,931$ για τους εγγεγραμμένους από ΤΕΕ, άρα η «επικινδυνότητα» να αποφοιτήσουν είναι 0,931 φορές λιγότερη από ότι στους εγγεγραμμένους με εισιτήριες εξετάσεις.

Επίσης, ο συντελεστής είναι $HR=0,822$ για τους εγγεγραμμένους με άλλο τρόπο, άρα η «επικινδυνότητα» να αποφοιτήσουν είναι 0,822 φορές μικρότερη από ότι στους εγγεγραμμένους με εισιτήριες εξετάσεις.

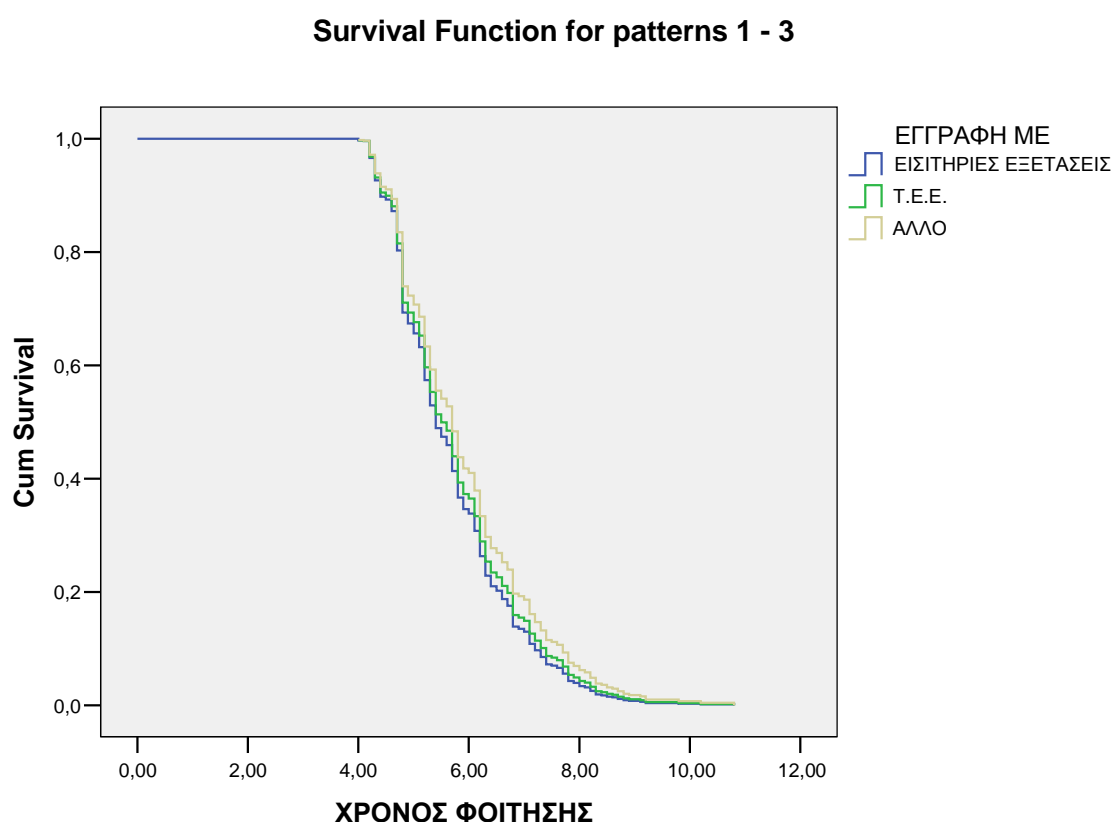
Όλα τα προηγούμενα δείχνουν ότι και οι εγγεγραμμένοι από ΤΕΕ και οι εγγεγραμμένοι με άλλο τρόπο έχουν μεγαλύτερους χρόνους φοίτησης από τους εγγεγραμμένους με εισιτήριες εξετάσεις. Όμως οι διαφορές αυτές δεν είναι στατιστικά σημαντικές.

Η συνολική συνάρτηση επιβίωσης για το μοντέλο μας παρουσιάζεται στο παρακάτω διάγραμμα 18. Βλέπουμε ότι μετά τα 4 χρόνια, που είναι και ο ελάχιστος χρόνος φοίτησης, η πιθανότητα να μην έχει αποφοιτήσει κάποιος φοιτητής μειώνεται. Μάλιστα, φαίνεται ότι σχεδόν όλοι οι φοιτητές αποφοιτούν στο διάστημα μεταξύ 4 και 8 χρόνων. Μετά τα 8 χρόνια η πιθανότητα επιβίωσης σχεδόν μηδενίζεται.



Διάγραμμα 18: Συνολική συνάρτηση επιβίωσης

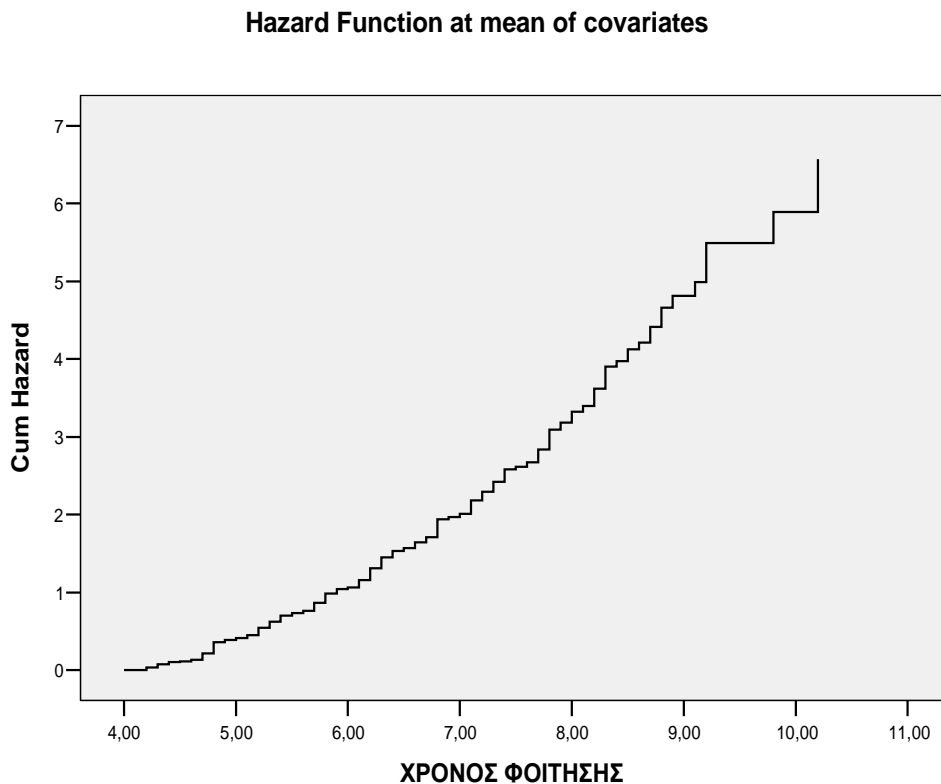
Στο διάγραμμα 19 παρουσιάζονται, σε τρεις διαφορετικές γραμμές, οι συναρτήσεις επιβίωσης για τους εγγεγραμμένους από εισιτήριες εξετάσεις, τους εγγεγραμμένους από ΤΕΕ και τους εγγεγραμμένους με άλλο τρόπο. Οι εγγεγραμμένοι από εισιτήριες εξετάσεις παρουσιάζουν μεγαλύτερη «επικινδυνότητα» να αποφοιτήσουν και η γραμμή τους έχει χαμηλότερες τιμές για την αθροιστική συνάρτηση επιβίωσης. Ακολουθούν οι εγγεγραμμένοι από ΤΕΕ και οι τέλος εγγεγραμμένοι με άλλο τρόπο που παρουσιάζουν και τις υψηλότερες τιμές επιβίωσης (φοίτησης).



Διάγραμμα 19: Συνάρτηση επιβίωσης ανά είδος εγγραφής

Βέβαια, το διάγραμμα επιβεβαιώνει τα προηγούμενα συμπεράσματα, δηλαδή δεν φαίνεται να υπάρχουν ουσιαστικές διαφορές για τις τρεις κατηγορίες φοιτητών.

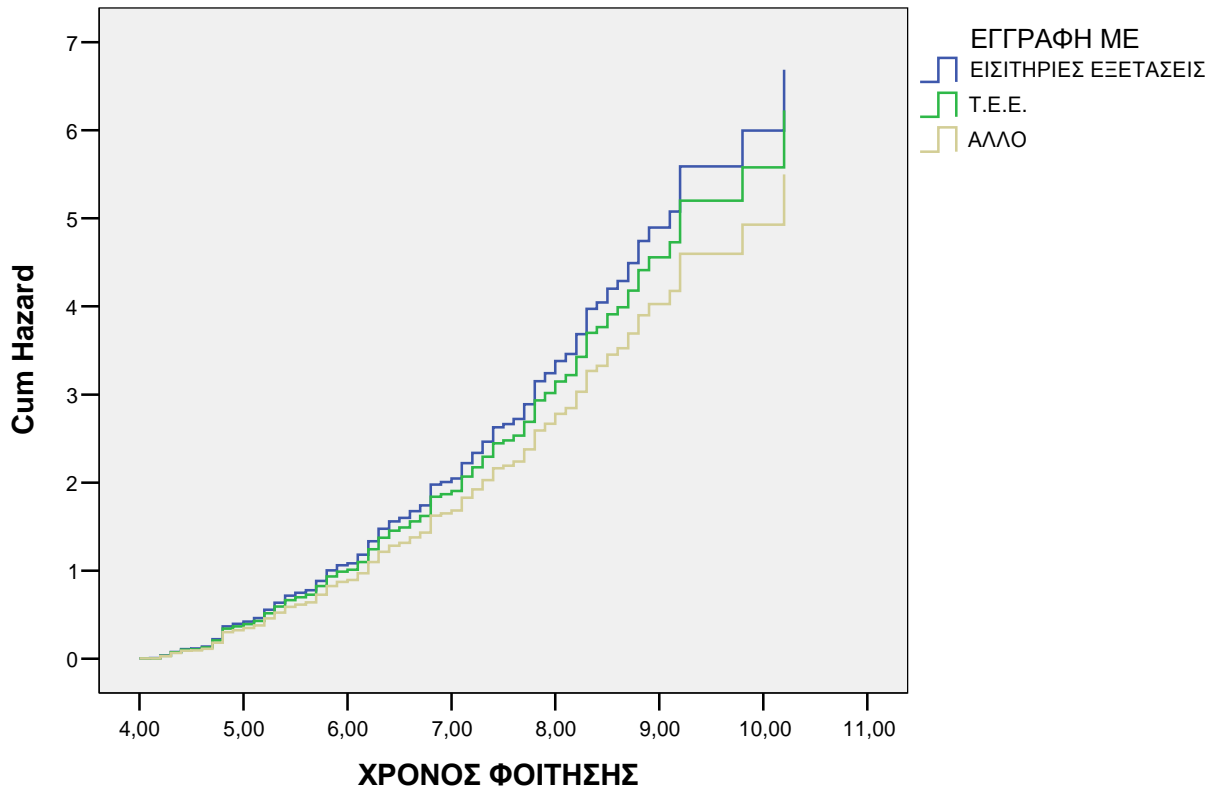
Ακριβώς ίδια συμπεράσματα βγαίνουν και από το γράφημα της συνάρτησης επικινδυνότητας (διάγραμμα 20). Μετά από τα 4 χρόνια φοίτησης, αυξάνεται και η πιθανότητα να αποφοιτήσει κάποιος.



Διάγραμμα 20: Συνολική συνάρτηση επικινδυνότητας

Τέλος, παρουσιάζεται η συνάρτηση επικινδυνότητας για κάθε Είδος Εγγραφής ξεχωριστά (διάγραμμα 21). Είναι φανερό ότι για τους εγγεγραμμένους με εισιτήριες εξετάσεις υπάρχει μεγαλύτερο «ρίσκο» να αποφοιτήσουν αφού η συνάρτησή τους έχει μεγαλύτερες τιμές. Οι διαφορές όμως είναι αμελητέες για τα τρία γκρουπ.

Hazard Function for patterns 1 - 3



Διάγραμμα 21: Συνάρτηση επικινδυνότητας ανά είδος εγγραφής

3.3.4 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ

Μια ανεξάρτητη μεταβλητή που μπορεί να χρησιμοποιηθεί είναι το «ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ». Θα ελέγξουμε αν οι φοιτητές που εισήχθησαν σε διαφορετικές χρονιές είχαν διαφορετικές διάρκειες φοίτησης στη σχολή. Προκειμένου να εκτελέσουμε τη στατιστική διαδικασία COX REGRESSION επιλέξαμε τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: 1999, 2000, 2001, 2002, 2003, 2004, 2005)

Από την εφαρμογή της διαδικασίας στο SPSS προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

		N	Percent
Cases available in analysis	Even ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
	Total	747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 32: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 32 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Ο πίνακας 33 απεικονίζει την ψευδομεταβλητή που δημιουργήθηκε για την εισαγωγή της κατηγορικής μεταβλητής στο μοντέλο. Σύμφωνα με αυτόν, η κατηγορία αναφοράς είναι το «2005».

Categorical Variable Codings

	Frequency	(1)	(2)	(3)	(4)	(5)	(6)
etos ^a 1=1999	82	1	0	0	0	0	0
2=2000	155	0	1	0	0	0	0
3=2001	163	0	0	1	0	0	0
4=2002	153	0	0	0	1	0	0
5=2003	114	0	0	0	0	1	0
6=2004	63	0	0	0	0	0	1
7=2005	17	0	0	0	0	0	0

a. Indicator Parameter Coding

b. Category variable: etos (ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ)

Πίνακας 33: Κωδικοποίηση κατηγορικής μεταβλητής

Αμέσως παρακάτω απεικονίζεται ο πίνακας 34 με τον έλεγχο για το αν έστω και μια ανεξάρτητη μεταβλητή του μοντέλου είναι στατιστικά σημαντική. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο. Βλέπουμε ότι αφού το $\text{sig.} = 0,000 < 0,05$, έστω και ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0. Βέβαια, στο μοντέλο μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, η οποία φαίνεται ότι είναι στατιστικά σημαντική.

Omnibus Tests of Model Coefficients

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8384,760	128,323	6	,000	82,408	6	,000	82,408	6	,000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168

b. Beginning Block Number 1. Method = Enter

Πίνακας 34: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 35 που απεικονίζει τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Η μεταβλητή «ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ» φαίνεται ότι είναι στατιστικά σημαντική, σε επίπεδο 5%, αφού το sig. = 0,000 < 0,05.

Άλλο στοιχείο του πίνακα είναι η τιμή της στατιστικής συνάρτησης WALD = 106,492 σύμφωνα με την οποία γίνεται ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή Β βάσει της Χ-τετράγωνο κατανομής.

Οι τιμές του Β για τα έτη 1999 έως και 2004 είναι -2.255, -2.066, -2.020, -1.903, -1.741, -1.052 ενώ οι τιμές του Exp(B) είναι HR={0.105, 0.127, 0.133, 0.149, 0.175 και 0.349} αντίστοιχα. Τέλος, για το Exp(B) παρουσιάζονται τα αντίστοιχα διαστήματα εμπιστοσύνης σε επίπεδο 95%.

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
etos			106,492	6	,000			
etos(1)	-2,255	,278	65,654	1	,000	,105	,061	,181
etos(2)	-2,066	,266	60,424	1	,000	,127	,075	,213
etos(3)	-2,020	,264	58,383	1	,000	,133	,079	,223
etos(4)	-1,903	,264	51,903	1	,000	,149	,089	,250
etos(5)	-1,741	,267	42,402	1	,000	,175	,104	,296
etos(6)	-1,052	,277	14,407	1	,000	,349	,203	,601

Πίνακας 35: Έλεγχος σημαντικότητας για την ανεξάρτητη μεταβλητή

	Mean	Pattern						
		1	2	3	4	5	6	7
etos(1)	,110	1,000	,000	,000	,000	,000	,000	,000
etos(2)	,207	,000	1,000	,000	,000	,000	,000	,000
etos(3)	,218	,000	,000	1,000	,000	,000	,000	,000
etos(4)	,205	,000	,000	,000	1,000	,000	,000	,000
etos(5)	,153	,000	,000	,000	,000	1,000	,000	,000
etos(6)	,084	,000	,000	,000	,000	,000	1,000	,000

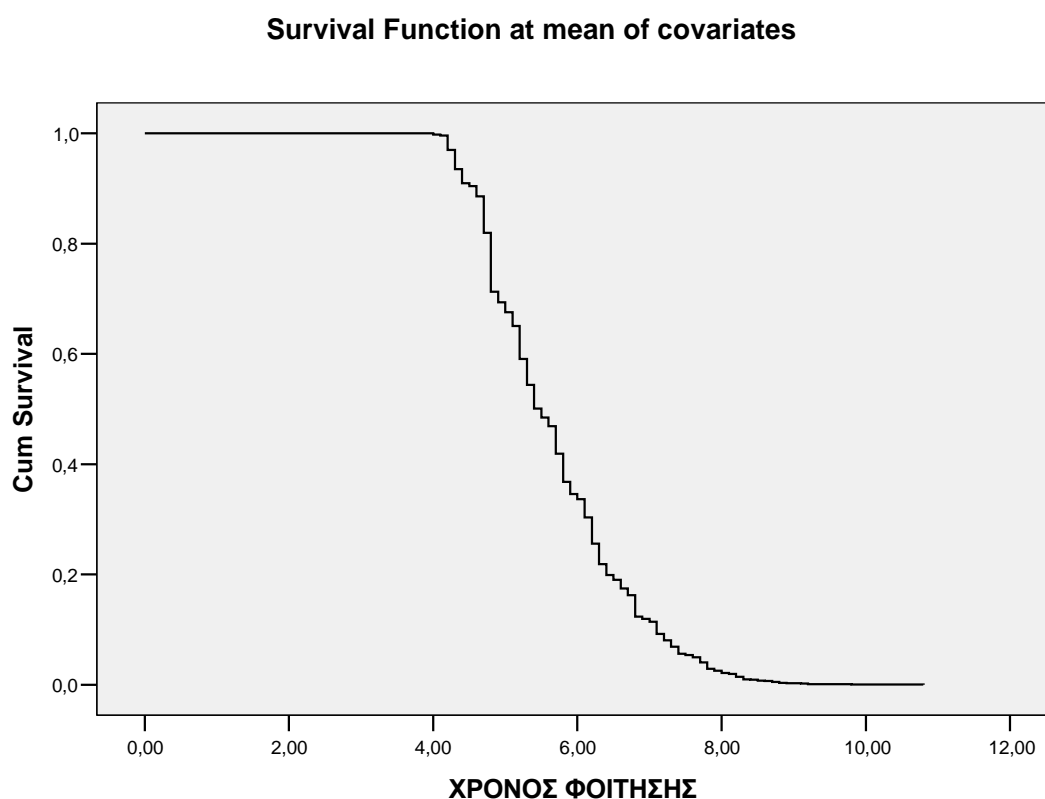
Πίνακας 36: Μέσοι όροι των ανεξάρτητων μεταβλητών

Στην περίπτωση μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, που είναι ποιοτική με επτά κατηγορίες. Οπότε η στήλη Exp(B) εκφράζει την αλλαγή επικινδυνότητας από την κατηγορία αναφοράς (έτος 2005) στις άλλες έξι

κατηγορίες (1999-2004). Εδώ βλέπουμε ότι ο συντελεστής είναι 0,105 για τους φοιτητές που εισήχθησαν το 1999, άρα η «επικινδυνότητα» να αποφοιτήσουν είναι πολύ μικρότερη από ότι στους εισαχθέντες του 2005. Το ίδιο συμπέρασμα προκύπτει και για τους εισαχθέντες των άλλων ετών. Μάλιστα φαίνεται ότι η «επικινδυνότητα» αποφοίτησης μεγαλώνει όσο πηγαίνουμε από το 1999 προς το 2005.

Όλα τα προηγούμενα δείχνουν ότι όσο περνάνε τα χρόνια, οι εισαχθέντες φοιτητές επιτυγχάνουν όλο και μικρότερους χρόνους φοίτησης στη σχολή. Η τάση αυτή είναι στατιστικά σημαντική.

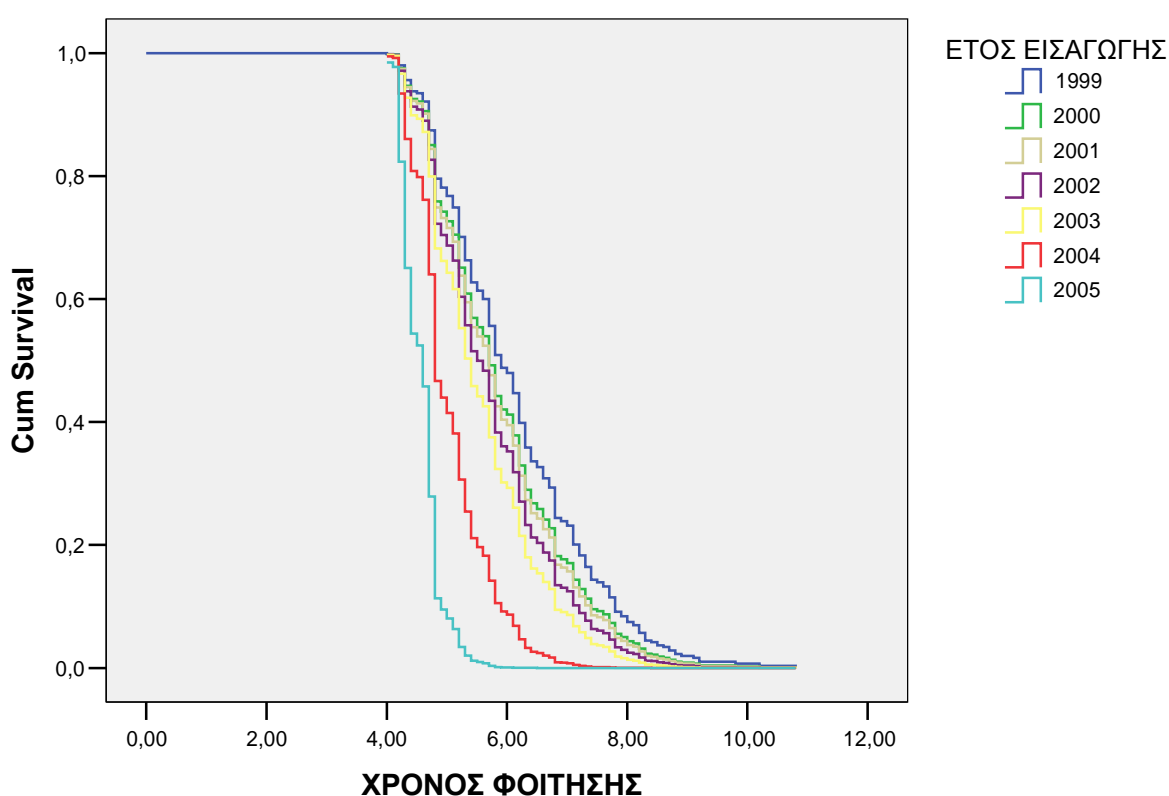
Η συνολική συνάρτηση επιβίωσης για το μοντέλο μας παρουσιάζεται στο παρακάτω διάγραμμα 22. Βλέπουμε ότι μετά τα 4 χρόνια, που είναι και ο ελάχιστος χρόνος φοίτησης, η πιθανότητα να μην έχει αποφοιτήσει κάποιος φοιτητής μειώνεται. Μάλιστα, φαίνεται ότι σχεδόν όλοι οι φοιτητές αποφοιτούν στο διάστημα μεταξύ 4 και 8 χρόνων. Μετά τα 8 χρόνια η πιθανότητα επιβίωσης σχεδόν μηδενίζεται.



Διάγραμμα 22: Συνολική συνάρτηση επιβίωσης

Στο επόμενο διάγραμμα 23 παρουσιάζονται, σε επτά διαφορετικές γραμμές, οι συναρτήσεις επιβίωσης για τους εισαχθέντες κατά τα έτη 1999 - 2005. Οι εισαχθέντες του 2005 παρουσιάζουν μεγαλύτερη «επικινδυνότητα» να αποφοιτήσουν και η γραμμή τους έχει χαμηλότερες τιμές για την αθροιστική συνάρτηση επιβίωσης. Μάλιστα, η πιθανότητα επιβίωσης μετά τα 5 χρόνια είναι σχεδόν μηδενική για αυτούς.

Survival Function for patterns 1 - 7

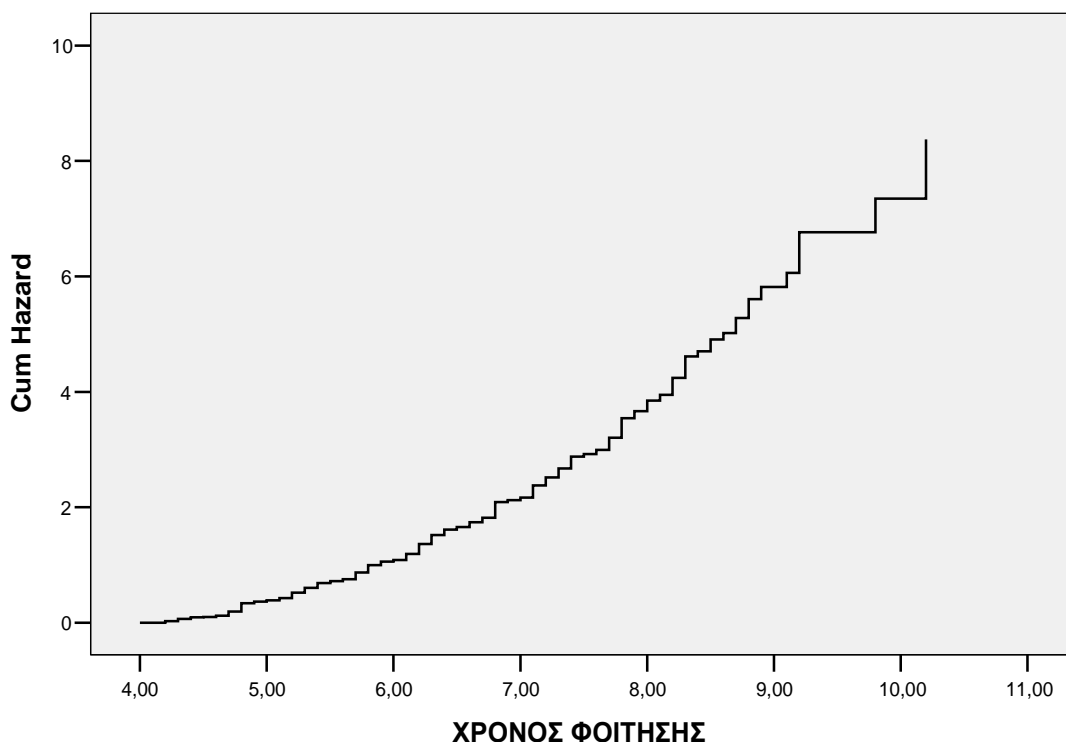


Διάγραμμα 23: Συνάρτηση επιβίωσης ανά έτος εισαγωγής

Επίσης, επιβεβαιώνεται και γραφικά το συμπέρασμα της προηγούμενης παραγράφου, δηλαδή ότι όσο περνάνε τα χρόνια μειώνεται η πιθανότητα επιβίωσης (μικραίνει ο χρόνος φοίτησης).

Ακριβώς ίδια συμπεράσματα βγαίνουν και από το γράφημα της συνάρτησης επικινδυνότητας (διάγραμμα 24). Μετά από τα 4 χρόνια φοίτησης, αυξάνεται και η πιθανότητα να αποφοιτήσει κάποιος.

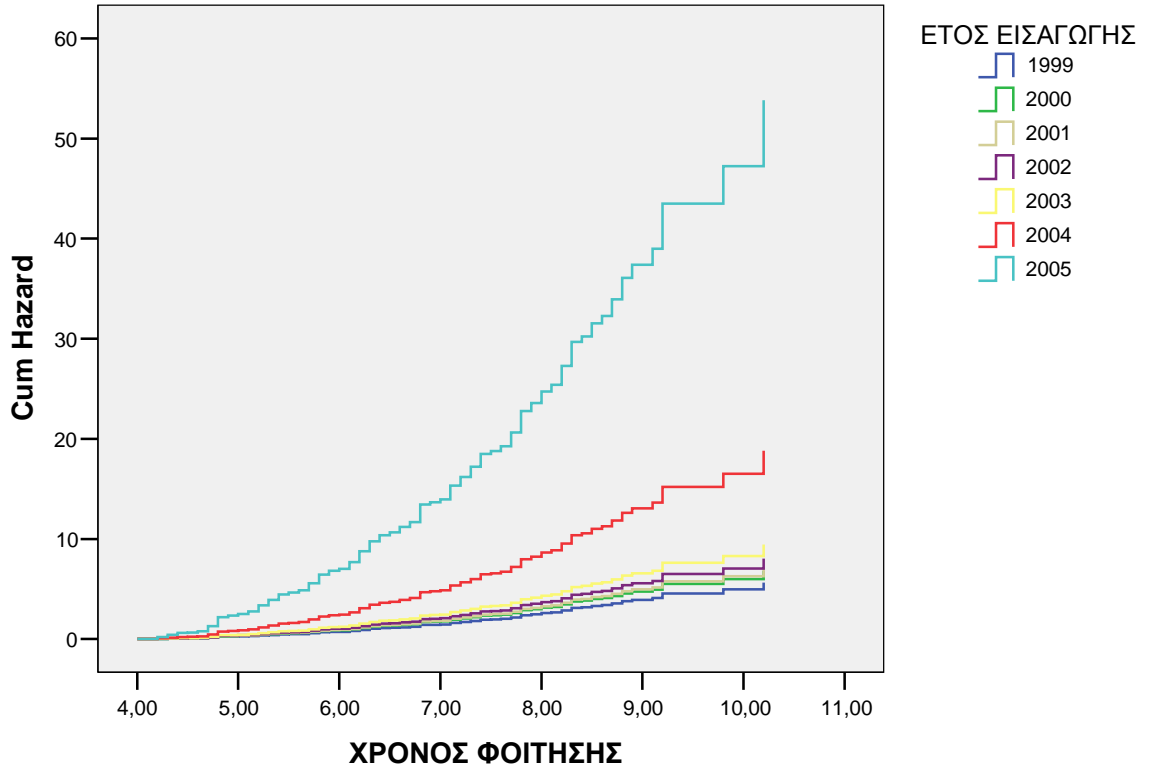
Hazard Function at mean of covariates



Διάγραμμα 24: Συνολική συνάρτηση επικινδυνότητας

Τέλος, παρουσιάζεται η συνάρτηση επικινδυνότητας για κάθε έτος εισαγωγής ξεχωριστά (διάγραμμα 25). Είναι φανερό ότι για τους εισαχθέντες του 2005 υπάρχει μεγαλύτερο «ρίσκο» να αποφοιτήσουν αφού η συνάρτησή τους έχει μεγαλύτερες τιμές. Οι διαφορές ανάμεσα στα έτη είναι εμφανείς και παρουσιάζεται η τάση αύξησης της «επικινδυνότητας» αποφοίτησης όσο κινούμαστε από το έτος εισαγωγής 1999 προς το έτος εισαγωγής 2005.

Hazard Function for patterns 1 - 7



Διάγραμμα 25: Συνάρτηση επικινδυνότητας ανά έτος εισαγωγής

3.3.5 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ

Στη συνέχεια ελέγχουμε αν οι φοιτητές που εισήχθησαν στο χειμερινό εξάμηνο έχουν διαφορετικές διάρκειες φοίτησης από τους φοιτητές που εισήχθησαν στο εαρινό εξάμηνο. Προκειμένου να εκτελέσουμε τη στατιστική διαδικασία COX REGRESSION επιλέξαμε τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: Χειμερινό εξάμηνο, Εαρινό εξάμηνο)

Από την εφαρμογή της διαδικασίας στο SPSS προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

Case Processing Summary		N	Percent
Cases available in analysis	Even ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
Total		747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 37: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 37 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Ο επόμενος πίνακας 38 απεικονίζει την ψευδομεταβλητή που δημιουργήθηκε για την εισαγωγή της κατηγορικής μεταβλητής στο μοντέλο. Σύμφωνα με αυτόν, η κατηγορία αναφοράς είναι το «Εαρινό εξάμηνο».

Categorical Variable Codings^b

	Frequency	(1)
semester ^a 1=X.E.	415	1
2=E.E.	332	0

a. Indicator Parameter Coding

b. Category variable: semester (ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ)

Πίνακας 38: Κωδικοποίηση κατηγορικής μεταβλητής

Αμέσως παρακάτω απεικονίζεται ο πίνακας 39 με τον έλεγχο για το αν έστω και μια ανεξάρτητη μεταβλητή του μοντέλου είναι στατιστικά σημαντική. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο. Βλέπουμε ότι αφού το $\text{sig.} = 0,310 > 0.05$, όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι ίσοι με το 0. Βέβαια, στο μοντέλο μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, η οποία φαίνεται ότι δεν είναι στατιστικά σημαντική.

Omnibus Tests of Model Coefficients^a

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8466,135	1,030	1	,310	1,032	1	,310	1,032	1	,310

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168

b. Beginning Block Number 1. Method = Enter

Πίνακας 39: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 40 που απεικονίζει τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Η μεταβλητή «ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ» φαίνεται ότι είναι δεν είναι στατιστικά σημαντική, σε επίπεδο 5%, αφού το sig. = 0,310 > 0,05.

Άλλο στοιχείο του πίνακα είναι η τιμή της στατιστικής συνάρτησης WALD = 1,030 σύμφωνα με την οποία γίνεται ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή B βάσει της Χ-τετράγωνο κατανομής.

Η τιμή του συντελεστή B είναι 0,075 για τους εγγεγραμμένους στο χειμερινό εξάμηνο ενώ η τιμή του Exp(B) είναι 1,078. Τέλος, το Exp(B) κυμαίνεται μεταξύ 0,932 και 1,246 για τους εγγεγραμμένους στο χειμερινό εξάμηνο, με 95% εμπιστοσύνη.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
semester	,075	,074	1,030	1	,310	1,078	,932	1,246

Πίνακας 40: Έλεγχος σημαντικότητας για την ανεξάρτητη μεταβλητή

Covariate Means and Pattern Values

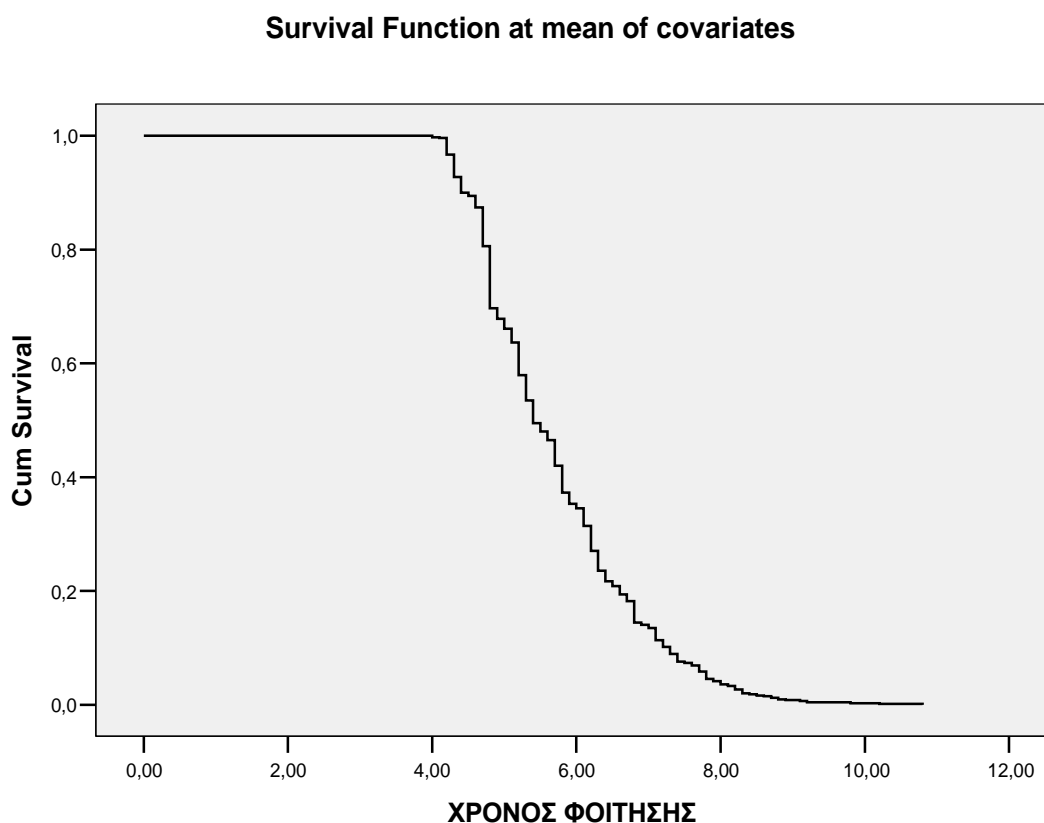
	Mean	Pattern	
		1	2
semester	,556	1,000	,000

Πίνακας 41: Μέσοι όροι των ανεξάρτητων μεταβλητών

Στην περίπτωση μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, που είναι ποιοτική με δύο κατηγορίες. Οπότε η στήλη Exp(B) εκφράζει την αλλαγή επικινδυνότητας από την κατηγορία αναφοράς (Εαρινό εξάμηνο) στην άλλη κατηγορία (Χειμερινό εξάμηνο). Εδώ βλέπουμε ότι ο συντελεστής είναι 1,078 για τους εγγεγραμμένους στο Χειμερινό εξάμηνο, άρα η «επικινδυνότητα» να αποφοιτήσουν είναι 1,078 φορές περισσότερη από ότι στους εγγεγραμμένους στο Εαρινό εξάμηνο.

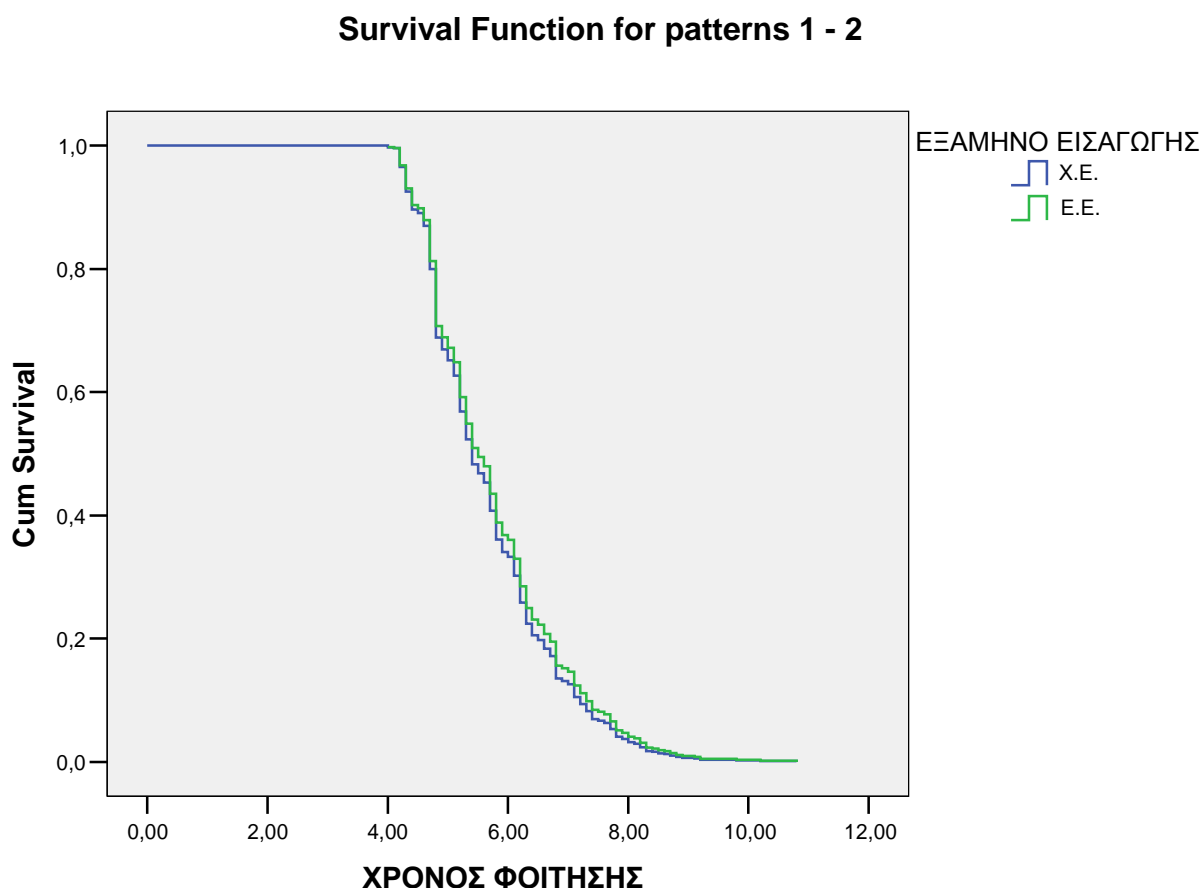
Όλα τα προηγούμενα δείχνουν ότι οι εγγεγραμμένοι στο Χειμερινό εξάμηνο έχουν μικρότερους χρόνους φοίτησης από τους εγγεγραμμένους στο Εαρινό εξάμηνο. Η διαφορά, όμως, αυτή δεν είναι στατιστικά σημαντική.

Η συνολική συνάρτηση επιβίωσης για το μοντέλο μας παρουσιάζεται στο παρακάτω διάγραμμα 26. Βλέπουμε ότι μετά τα 4 χρόνια, που είναι και ο ελάχιστος χρόνος φοίτησης, η πιθανότητα να μην έχει αποφοιτήσει κάποιος φοιτητής μειώνεται. Μάλιστα, φαίνεται ότι σχεδόν όλοι οι φοιτητές αποφοιτούν στο διάστημα μεταξύ 4 και 8 χρόνων. Μετά τα 8 χρόνια η πιθανότητα επιβίωσης σχεδόν μηδενίζεται.



Διάγραμμα 26: Συνολική συνάρτηση επιβίωσης

Στο επόμενο διάγραμμα 27 παρουσιάζονται, σε δύο διαφορετικές γραμμές, οι συναρτήσεις επιβίωσης για τους εγγεγραμμένους στο Χειμερινό και στο Εαρινό εξάμηνο. Οι εγγεγραμμένοι στο Χειμερινό εξάμηνο παρουσιάζουν μια ανεπαίσθητη μεγαλύτερη «επικινδυνότητα» να αποφοιτήσουν και η γραμμή τους έχει ελάχιστα χαμηλότερες τιμές για την αθροιστική συνάρτηση επιβίωσης.

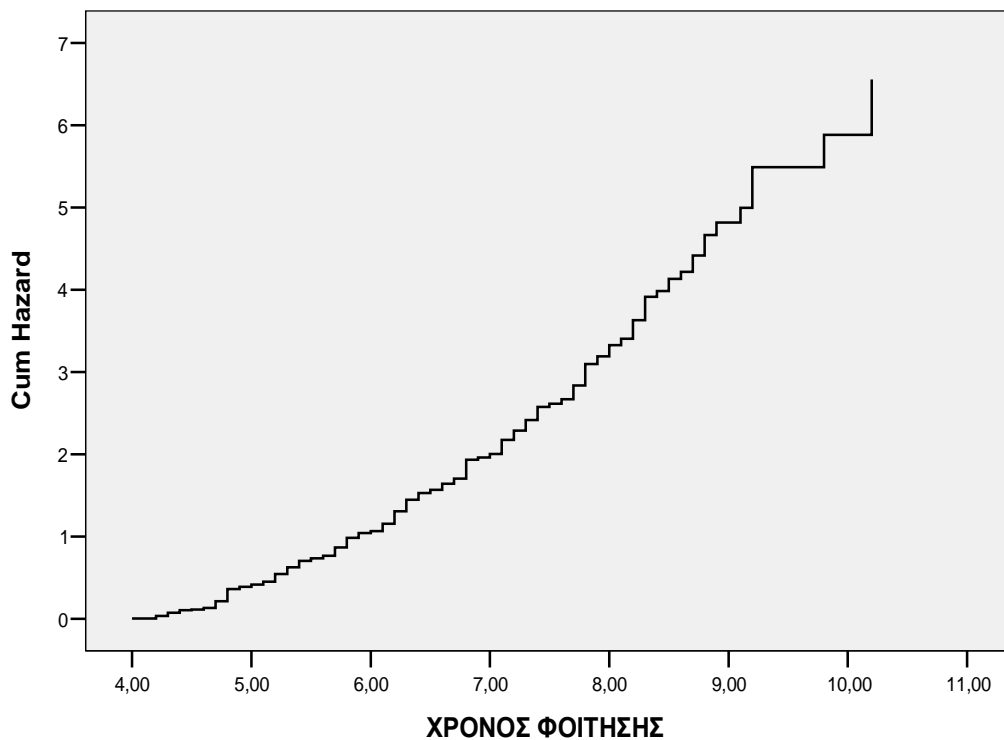


Διάγραμμα 27: Συνάρτηση επιβίωσης ανά εξάμηνο εισαγωγής

Το διάγραμμα επιβεβαιώνει τα προηγούμενα συμπεράσματα, δηλαδή δεν φαίνεται να υπάρχουν ουσιαστικές διαφορές για τις δύο κατηγορίες φοιτητών.

Ακριβώς ίδια συμπεράσματα βγαίνουν και από το γράφημα της συνάρτησης επικινδυνότητας (διάγραμμα 27). Μετά από τα 4 χρόνια φοίτησης, αυξάνεται και η πιθανότητα να αποφοιτήσει κάποιος.

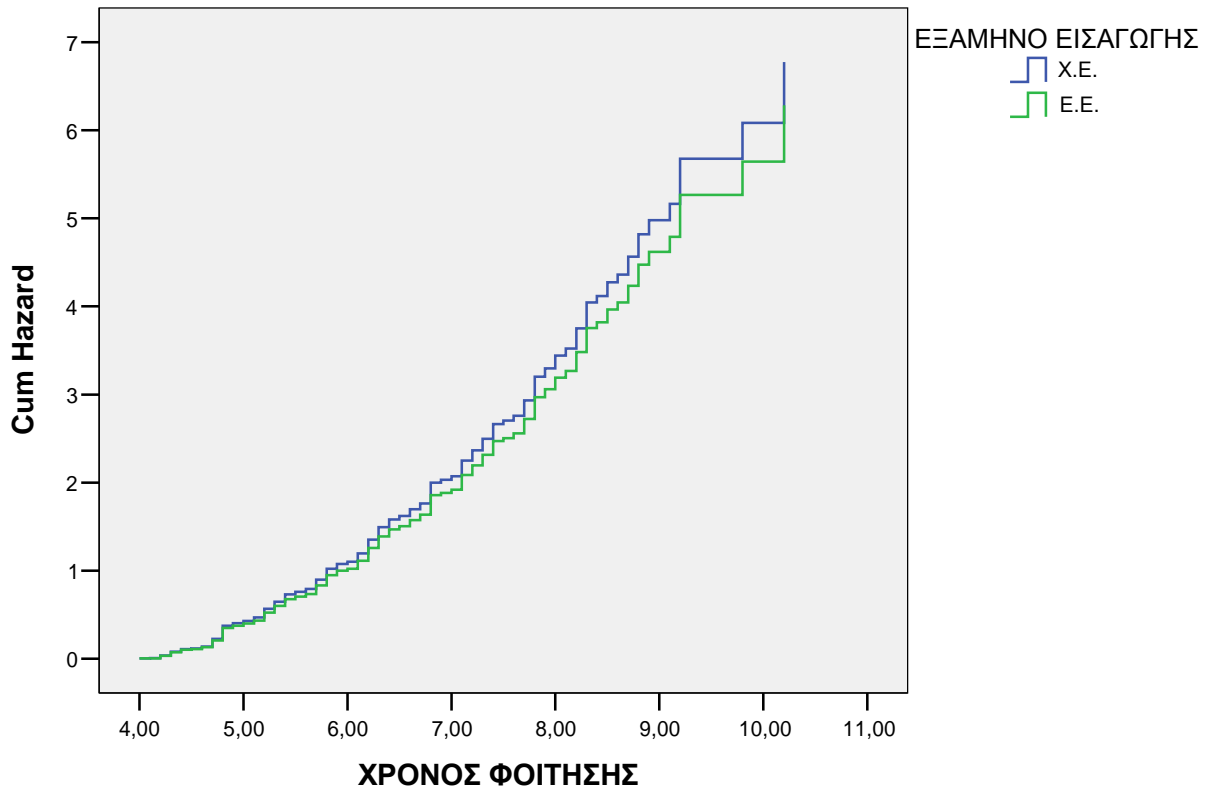
Hazard Function at mean of covariates



Διάγραμμα 27: Συνολική συνάρτηση επικινδυνότητας

Τέλος, παρουσιάζεται η συνάρτηση επικινδυνότητας για κάθε Εξάμηνο Εισαγωγής ξεχωριστά (διάγραμμα 28). Είναι φανερό ότι για τους εγγεγραμμένους στο Χειμερινό εξάμηνο υπάρχει μεγαλύτερο «κίνδυνος» να αποφοιτήσουν αφού η συνάρτησή τους έχει μεγαλύτερες τιμές. Οι διαφορές δεν είναι στατιστικά σημαντικές για τα δύο γκρουπ.

Hazard Function for patterns 1 - 2



Διάγραμμα 28: Συνάρτηση επικινδυνότητας ανά εξάμηνο εισαγωγής

3.3.6 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΠΤΥΧΙΑΚΗ

Έχοντας ολοκληρώσει την εξέταση όλων των ποιοτικών ανεξάρτητων μεταβλητών, θα χρησιμοποιήσουμε και δύο ποσοτικές μεταβλητές για να ελέγξουμε κατά πόσο επηρεάζουν το χρόνο φοίτησης. Κατ' αρχήν θα χρησιμοποιήσουμε το βαθμό που έχουν πάει οι φοιτητές στην πτυχιακή τους. Προκειμένου να εκτελέσουμε τη στατιστική διαδικασία COX REGRESSION επιλέξαμε τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΠΤΥΧΙΑΚΗ (ποσοτική μεταβλητή)

Από την εφαρμογή της διαδικασίας στο SPSS προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

		N	Percent
Cases available in analysis	Event ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
	Total	747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 42: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 42 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Αμέσως παρακάτω απεικονίζεται ο πίνακας 43 με τον έλεγχο για το αν έστω και μια ανεξάρτητη μεταβλητή του μοντέλου είναι στατιστικά σημαντική. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο. Βλέπουμε ότι αφού το $\text{sig.} = 0,000 < 0.05$, έστω και ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0. Βέβαια, στο μοντέλο μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, η οποία φαίνεται ότι είναι στατιστικά σημαντική.

Omnibus Tests of Model Coefficients

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8437,536	27,301	1	,000	29,632	1	,000	29,632	1	,000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168

b. Beginning Block Number 1. Method = Enter

Πίνακας 43: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 44 που απεικονίζει τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Η μεταβλητή «ΠΤΥΧΙΑΚΗ» φαίνεται ότι είναι στατιστικά σημαντική, σε επίπεδο 5%, αφού το $\text{sig.} = 0,000 < 0,05$.

Άλλο στοιχείο του πίνακα είναι η τιμή της στατιστικής συνάρτησης WALD = 27,284 σύμφωνα με την οποία γίνεται ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή Β βάσει της Χ-τετράγωνο κατανομής.

Η τιμή του συντελεστή Β είναι 0,209 ενώ η τιμή του $\text{Exp}(B)$ είναι 1,233. Τέλος, το $\text{Exp}(B)$ κυμαίνεται μεταξύ 1,140 και 1,334, με 95% εμπιστοσύνη.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
ptyxiaki	,209	,040	27,284	1	,000	1,233	1,140	1,334

Πίνακας 44: Έλεγχος σημαντικότητας για την ανεξάρτητη μεταβλητή

Στην περίπτωση μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, που είναι ποσοτική. Οπότε η στήλη $\text{Exp}(B)$ εκφράζει την εκτιμώμενη αλλαγή της επικινδυνότητας σε μια μονάδα αλλαγής της ανεξάρτητης μεταβλητής. Εδώ βλέπουμε ότι ο συντελεστής είναι $HR=1,233$, που υποδηλώνει ότι όταν ο βαθμός της πτυχιακής αυξάνεται κατά μια μονάδα τότε η «επικινδυνότητα» αποφοίτησης αυξάνει κατά 1,233 φορές.

Όλα τα προηγούμενα δείχνουν ότι όσο μεγαλύτερος ήταν ο βαθμός της πτυχιακής (που δείχνει φοιτητές με καλύτερες επιδόσεις) τόσο μικραίνει η διάρκεια φοίτησης.

3.3.7 ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ: ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ

Τέλος, ως τελευταία ερμηνευτική μεταβλητή θα χρησιμοποιήσουμε το βαθμό που έχουν πάρει οι φοιτητές στην πτυχίο τους. Έχοντας την ίδια λογική με την προηγούμενη μεταβλητή που χρησιμοποιήσαμε, αναμένουμε και παρόμοια αποτελέσματα. Προκειμένου να εκτελέσουμε τη στατιστική διαδικασία COX REGRESSION επιλέξαμε τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ (ποσοτική μεταβλητή)

Από την εκτέλεση της διαδικασίας προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

		N	Percent
Cases available in analysis	Even ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
	Total	747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 45: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 45 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Αμέσως παρακάτω απεικονίζεται ο πίνακας 46 με τον έλεγχο για το αν έστω και μια ανεξάρτητη μεταβλητή του μοντέλου είναι στατιστικά σημαντική. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο. Βλέπουμε ότι αφού το $\text{sig.} = 0,000 < 0,05$, έστω και ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0. Βέβαια, στο μοντέλο μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, η οποία φαίνεται ότι είναι στατιστικά σημαντική.

Omnibus Tests of Model Coefficients

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8300,500	189,486	1	,000	166,667	1	,000	166,667	1	,000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168

b. Beginning Block Number 1. Method = Enter

Πίνακας 46: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 47 που απεικονίζει τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Η μεταβλητή «ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ» φαίνεται ότι είναι στατιστικά σημαντική, σε επίπεδο 5%, αφού το $\text{sig.} = 0,000 < 0,05$.

Άλλο στοιχείο του πίνακα είναι η τιμή της στατιστικής συνάρτησης $WALD = 192,065$ σύμφωνα με την οποία γίνεται ο έλεγχος της στατιστικής σημαντικότητας του συντελεστή Β βάσει της Χ-τετράγωνο κατανομής.

Η τιμή του συντελεστή Β είναι 1,142 ενώ η τιμή του $\text{Exp}(B)$ είναι 3,133. Τέλος, το $\text{Exp}(B)$ κυμαίνεται μεταξύ 2,666 και 3,682, με 95% εμπιστοσύνη.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
ptyxio	1,142	,082	192,065	1	,000	3,133	2,666	3,682

Πίνακας 47: Έλεγχος σημαντικότητας για την ανεξάρτητη μεταβλητή

Στην περίπτωση μας έχουμε μόνο μια ανεξάρτητη μεταβλητή, που είναι ποσοτική. Οπότε η στήλη $\text{Exp}(B)$ εκφράζει την εκτιμώμενη αλλαγή της επικινδυνότητας σε μια μονάδα αλλαγής της ανεξάρτητης μεταβλητής. Εδώ βλέπουμε ότι ο συντελεστής είναι $HR=3,133$, που υποδηλώνει ότι όταν ο βαθμός του πτυχίου αυξάνεται κατά μια μονάδα τότε η «επικινδυνότητα» αποφοίτησης αυξάνει κατά 3,133 φορές.

Όλα τα προηγούμενα δείχνουν ότι όσο μεγαλύτερος ήταν ο βαθμός του πτυχίου (που δείχνει φοιτητές με καλύτερες επιδόσεις) τόσο μικραίνει η διάρκεια φοίτησης. Τα αποτελέσματα δείχνουν εδώ μια ισχυρότερη σχέση μεταξύ του βαθμού πτυχίου και της διάρκειας φοίτησης από ότι με τον βαθμό της πτυχιακής.

3.3.8 Σύγκριση αποτελεσμάτων

Έχοντας ολοκληρώσει την εξέταση όλων των ερμηνευτικών μεταβλητών συμπεραίνουμε ότι ορισμένες από αυτές ήταν πραγματικά στατιστικά σημαντικές και συνεισφέρουν στην ερμηνεία της εξαρτημένης μεταβλητής αφού είχαν sig. μικρότερο από 0,05. Οι μεταβλητές αυτές ήταν το φύλο, το έτος εισαγωγής, ο βαθμός της πτυχιακής και ο βαθμός του πτυχίου. Στον παρακάτω πίνακα 48 παρουσιάζονται όλες οι μεταβλητές που χρησιμοποιήθηκαν με το επίπεδο σημαντικότητας τους:

A/A	ΜΕΤΑΒΛΗΤΗ	ΕΙΔΟΣ	Sig.	ΣΤΑΤΙΣΤΙΚΑ ΣΗΜΑΝΤΙΚΗ (σε επίπεδο 5%)
1	ΦΥΛΟ	Ποιοτική	0,000	ΝΑΙ
2	ΥΠΗΚΟΟΤΗΤΑ	Ποιοτική	0,726	ΟΧΙ
3	ΕΓΓΡΑΦΗ ΜΕ	Ποιοτική	0,432	ΟΧΙ
4	ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ	Ποιοτική	0,000	ΝΑΙ
5	ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ	Ποιοτική	0,310	ΟΧΙ
6	ΠΤΥΧΙΑΚΗ	Ποσοτική	0,000	ΝΑΙ
7	ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ	Ποσοτική	0,000	ΝΑΙ

Πίνακας 48: Στατιστικά σημαντικές μεταβλητές

3.4 COX Regression πολυπαραγοντικό

3.4.1 Επιλεγμένες μεταβλητές

Προκειμένου να μπορέσουμε να πάρουμε ένα πλήρες μοντέλο που να ερμηνεύει καλύτερα τα δεδομένα και να μας δίνει καλύτερα αποτελέσματα, θα χρησιμοποιήσουμε ταυτόχρονα μερικές από τις ερμηνευτικές μεταβλητές που εξετάσαμε στις προηγούμενες παραγράφους. Εκτελούμε τη στατιστική διαδικασία COX REGRESSION τοποθετώντας τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΦΥΛΟ (κατηγορική μεταβλητή: Άνδρας, Γυναίκα)

-> ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: 1999-2005)

-> ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ(ποσοτική μεταβλητή)

Είναι φανερό ότι χρησιμοποιήσαμε τις δύο ποιοτικές μεταβλητές που βρήκαμε προηγουμένως ως στατιστικά σημαντικές και την ποσοτική μεταβλητή «βαθμός πτυχίου» που είχε μεγαλύτερη σημαντικότητα. Από την εκτέλεση της διαδικασίας προκύπτει, κατ' αρχήν, ο κάτωθι πίνακας:

Case Processing Summary

		N	Percent
Cases available in analysis	Event ^a	747	100,0%
	Censored	0	,0%
	Total	747	100,0%
Cases dropped	Cases with missing values	0	,0%
	Cases with negative time	0	,0%
	Censored cases before the earliest event in a stratum	0	,0%
	Total	0	,0%
Total		747	100,0%

a. Dependent Variable: ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Πίνακας 49: Συγκεντρωτική αναφορά του μοντέλου

Όπως φαίνεται στον πίνακα 49 το μοντέλο επιβίωσης δημιουργήθηκε από 747 παρατηρήσεις. Όλες αυτές αφορούν φοιτητές για τους οποίους συντελέστηκε το γεγονός, δηλαδή έχουν αποφοιτήσει. Επομένως, στο δείγμα μας δεν υπάρχουν καθόλου censored παρατηρήσεις. Επίσης, δεν έχει απορριφθεί καμία παρατήρηση για οποιοδήποτε λόγο.

Ο επόμενος πίνακας 50 απεικονίζει τις ψευδομεταβλητές που δημιουργήθηκαν για την εισαγωγή της κατηγορικών μεταβλητών στο μοντέλο. Σύμφωνα με αυτόν, η κατηγορία αναφοράς για το ΦΥΛΟ είναι οι «Γυναίκες» και για το ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ είναι το «2005».

Categorical Variable Codings

	Frequency	(1)	(2)	(3)	(4)	(5)	(6)
sex ^a 1=ΑΝΔΡΑΣ	281	1					
2=ΓΥΝΑΙΚΑ	466	0					
etos ^b 1=1999	82	1	0	0	0	0	0
2=2000	155	0	1	0	0	0	0
3=2001	163	0	0	1	0	0	0
4=2002	153	0	0	0	1	0	0
5=2003	114	0	0	0	0	1	0
6=2004	63	0	0	0	0	0	1
7=2005	17	0	0	0	0	0	0

a. Indicator Parameter Coding

b. Category variable: sex (ΦΥΛΟ)

c. Category variable: etos (ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ)

Πίνακας 50: Κωδικοποίηση κατηγορικής μεταβλητής

Αμέσως παρακάτω απεικονίζεται ο πίνακας 51 με τον έλεγχο για το αν έστω και μια ανεξάρτητη μεταβλητή του μοντέλου είναι στατιστικά σημαντική. Δηλαδή πραγματοποιούμε τον έλεγχο υποθέσεων:

H_0 : Όλοι οι συντελεστές των ανεξάρτητων μεταβλητών είναι 0.

H_1 : Έστω ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Ο έλεγχος πραγματοποιείται με το τεστ Χ-τετράγωνο και είναι ανάλογος με τον έλεγχο F για τη γραμμική παλινδρόμηση. Βλέπουμε ότι αφού το sig. = 0,000 < 0.05, έστω και ένας συντελεστής ανεξάρτητης μεταβλητής είναι διάφορος του 0.

Omnibus Tests of Model Coefficients

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
8206,291	325,158	8	,000	260,877	8	,000	260,877	8	,000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168

b. Beginning Block Number 1. Method = Enter

Πίνακας 51: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Το ανωτέρω συμπέρασμα επιβεβαιώνεται από τον ακόλουθο πίνακα 52 που απεικονίζει τον αντίστοιχο έλεγχο σημαντικότητας, για την κάθε μια ανεξάρτητη μεταβλητή ξεχωριστά. Μάλιστα, φαίνεται ότι όλες οι ερμηνευτικές μεταβλητές του μοντέλου είναι στατιστικά σημαντικές, σε επίπεδο 5%, αφού το sig. για κάθε μια είναι < 0,05.

Στον ίδιο πίνακα παρουσιάζονται οι τιμές του συντελεστή B και του Exp(B) , όπως και τα διαστήματα εμπιστοσύνης για το Exp(B), σε επίπεδο 95%.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
sex	-,189	,079	5,723	1	,017	,828	,709	,966
etos			98,154	6	,000			
etos(1)	-1,890	,281	45,329	1	,000	,151	,087	,262
etos(2)	-1,811	,268	45,563	1	,000	,163	,097	,277
etos(3)	-1,655	,267	38,348	1	,000	,191	,113	,323
etos(4)	-1,660	,266	39,042	1	,000	,190	,113	,320
etos(5)	-1,278	,271	22,271	1	,000	,278	,164	,474
etos(6)	-,694	,280	6,132	1	,013	,499	,288	,865
ptychio	1,168	,085	187,819	1	,000	3,216	2,721	3,800

Πίνακας 52: Έλεγχος σημαντικότητας για τις ανεξάρτητες μεταβλητές

Covariate Means and Pattern Values

	Mean	Pattern						
		1	2	3	4	5	6	7
sex	,376	,376	,376	,376	,376	,376	,376	,376
etos(1)	,110	1,000	,000	,000	,000	,000	,000	,000
etos(2)	,207	,000	1,000	,000	,000	,000	,000	,000
etos(3)	,218	,000	,000	1,000	,000	,000	,000	,000
etos(4)	,205	,000	,000	,000	1,000	,000	,000	,000
etos(5)	,153	,000	,000	,000	,000	1,000	,000	,000
etos(6)	,084	,000	,000	,000	,000	,000	1,000	,000
ptyxio	6,573	6,573	6,573	6,573	6,573	6,573	6,573	6,573

Πίνακας 53: Μέσοι όροι των ανεξάρτητων μεταβλητών

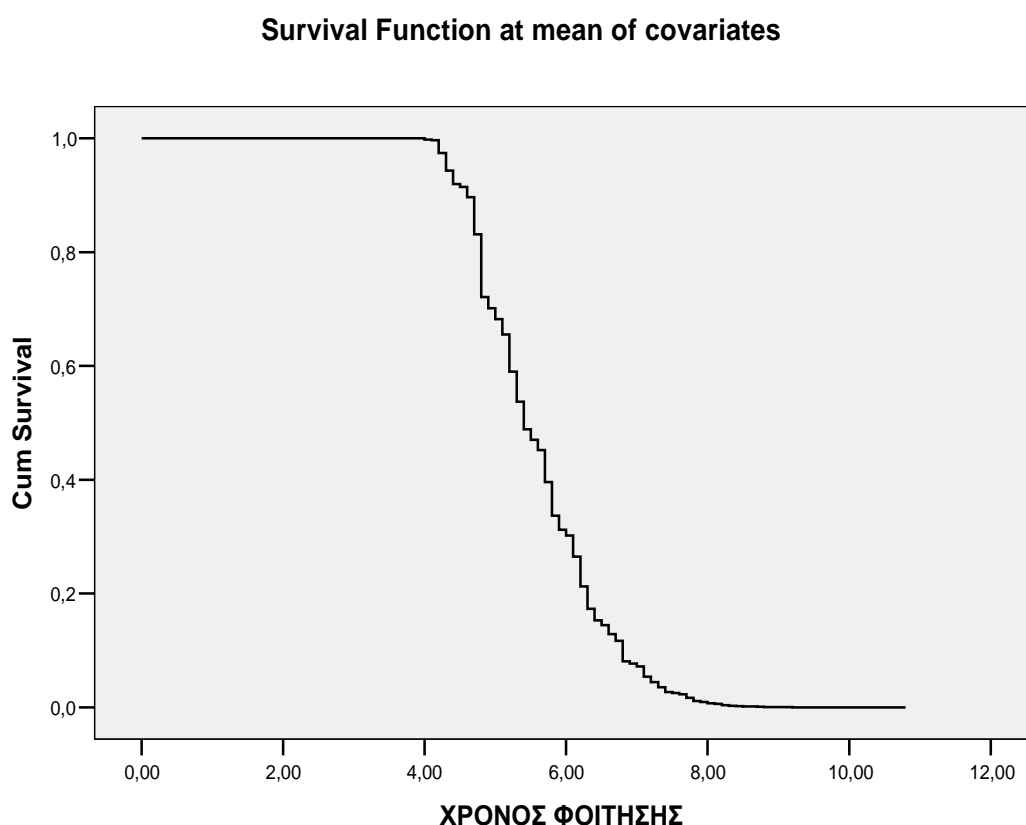
Στην περίπτωση μας έχουμε τρεις ανεξάρτητες μεταβλητές, δύο ποιοτικές και μια ποσοτική.

Όσον αφορά τις ποιοτικές μεταβλητές, η στήλη Expr(B) εκφράζει την αλλαγή επικινδυνότητας από την κατηγορία αναφοράς στις άλλες κατηγορίες. Ξεκινώντας από τη μεταβλητή ΦΥΛΟ, βλέπουμε ότι ο συντελεστής είναι $HR=0,828$ για τους άνδρες, άρα η «επικινδυνότητα» να αποφοιτήσουν είναι μικρότερη από ότι στις γυναίκες. Για τη μεταβλητή ΕΤΟΣ ΦΟΙΤΗΣΗΣ, βλέπουμε ότι ο συντελεστής είναι $HR=\{0.151, 0.163, 0.191, 0.190, 0.278\}$ και 0.499 για τα έτη 1999, 2000, 2001, 2002, 2003 και 2004 αντίστοιχα. Συνεπώς, η «επικινδυνότητα» να αποφοιτήσουν οι εισαχθέντες αυτών των ετών είναι μικρότερη από τους εισαχθέντες του 2005. Ενδιαφέρον παρουσιάζει το γεγονός ότι όσο περνάνε τα χρόνια από το 1999 προς το 2005, η «επικινδυνότητα» αποφοίτησης αυξάνει.

Όσον αφορά την ποσοτική μεταβλητή, η στήλη Expr(B) εκφράζει την εκτιμώμενη αλλαγή της επικινδυνότητας σε μια μονάδα αλλαγής της ανεξάρτητης μεταβλητής. Εδώ βλέπουμε ότι ο συντελεστής είναι $HR=3,216$, που υποδηλώνει ότι όταν ο βαθμός του πτυχίου αυξάνεται κατά μια μονάδα τότε η «επικινδυνότητα» αποφοίτησης αυξάνει κατά 3,216 φορές.

Με βάση το γεγονός ότι όλες οι προηγούμενες ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές, συμπεραίνουμε ότι οι γυναίκες παρουσιάζουν χαμηλότερο δείκτη επιβίωσης (μικρότερους χρόνους φοίτησης), όπως και οι εισαχθέντες σε νεότερα χρόνια. Τέλος, όσο μεγαλύτερος είναι ο βαθμός του πτυχίου τόσο μικρότερος χρόνος φοίτησης έχει επιτευχθεί από το φοιτητή.

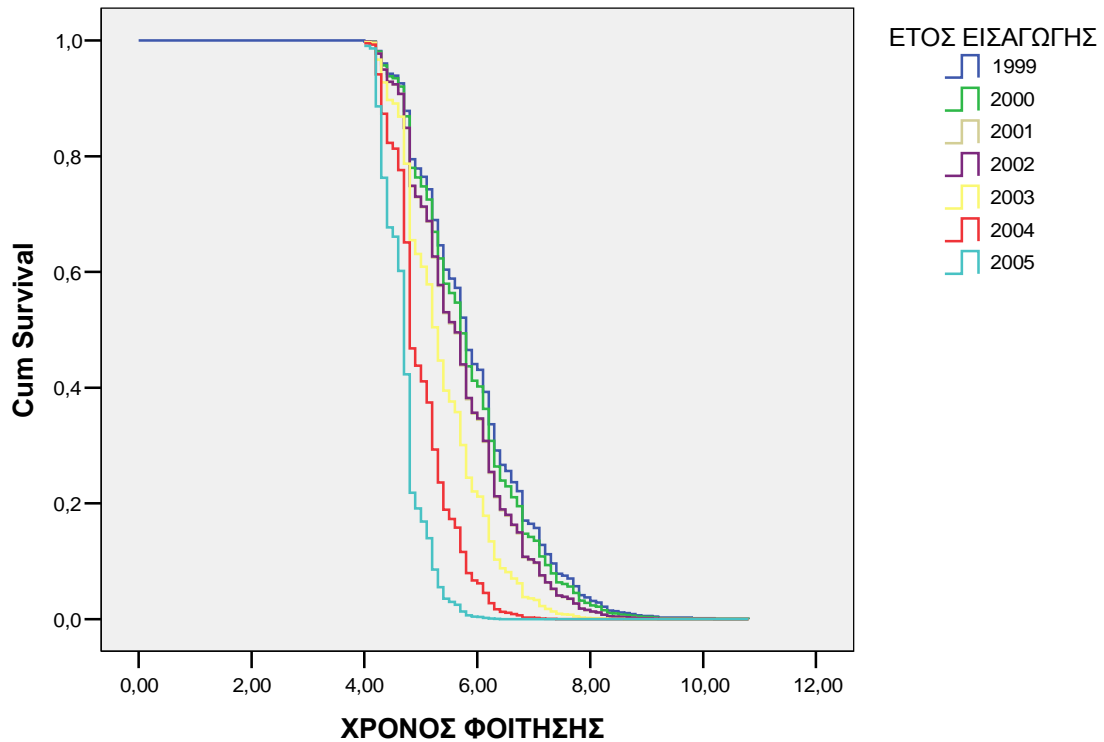
Η συνολική συνάρτηση επιβίωσης για το μοντέλο μας παρουσιάζεται στο παρακάτω διάγραμμα 29. Βλέπουμε ότι μετά τα 4 χρόνια, που είναι και ο ελάχιστος χρόνος φοίτησης, η πιθανότητα να μην έχει αποφοιτήσει κάποιος φοιτητής μειώνεται. Μάλιστα, φαίνεται ότι σχεδόν όλοι οι φοιτητές αποφοιτούν στο διάστημα μεταξύ 4 και 8 χρόνων. Μετά τα 8 χρόνια η πιθανότητα επιβίωσης σχεδόν μηδενίζεται.



Διάγραμμα 29: Συνολική συνάρτηση επιβίωσης

Στο επόμενο διάγραμμα 30 παρουσιάζονται, σε επτά διαφορετικές γραμμές, οι συναρτήσεις επιβίωσης για τους εισαχθέντες κατά τα έτη 1999 – 2005, οι οποίες είναι προσαρμοσμένες στο επίπεδο του μέσου όρου του «φύλου» και του «βαθμού πτυχίου». Οι εισαχθέντες του 2005 παρουσιάζουν μεγαλύτερη «επικινδυνότητα» να αποφοιτήσουν και η γραμμή τους έχει χαμηλότερες τιμές για την αθροιστική συνάρτηση επιβίωσης. Μάλιστα, η πιθανότητα επιβίωσης μετά τα 5 χρόνια είναι σχεδόν μηδενική για αυτούς.

Survival Function for patterns 1 - 7

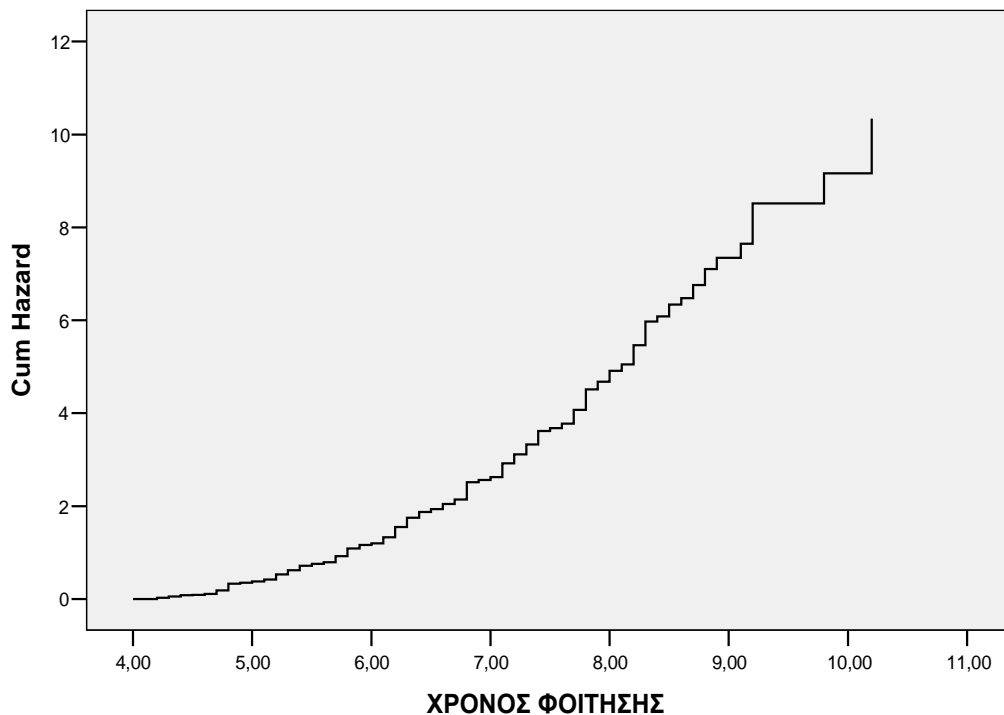


Διάγραμμα 30: Συνάρτηση επιβίωσης ανά έτος εισαγωγής

Επίσης, επιβεβαιώνεται και γραφικά το συμπέρασμα ότι όσο περνάνε τα χρόνια μειώνεται η πιθανότητα επιβίωσης (μικραίνει ο χρόνος φοίτησης).

Ακριβώς ίδια συμπεράσματα βγαίνουν και από το γράφημα της συνάρτησης επικινδυνότητας (διάγραμμα 31). Μετά από τα 4 χρόνια φοίτησης, αυξάνεται και η πιθανότητα να αποφοιτήσει κάποιος.

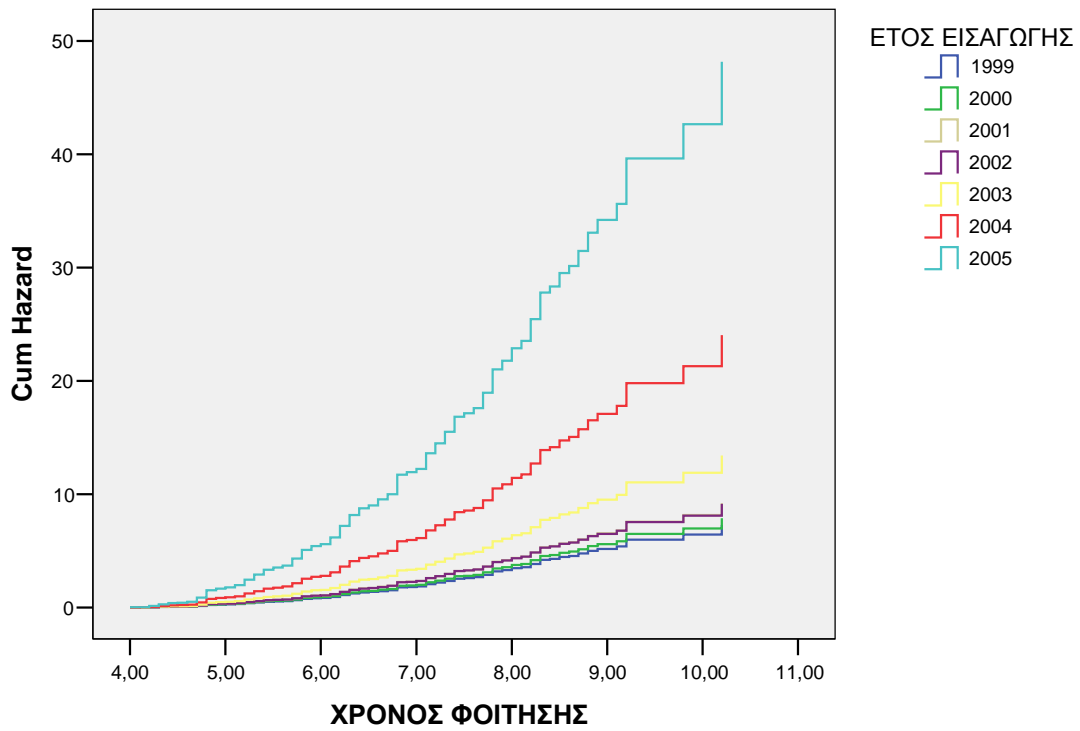
Hazard Function at mean of covariates



Διάγραμμα 31: Συνολική συνάρτηση επικινδυνότητας

Τέλος, παρουσιάζεται η συνάρτηση επικινδυνότητας για κάθε έτος εισαγωγής ξεχωριστά (διάγραμμα 32). Είναι φανερό ότι για τους εισαχθέντες του 2005 υπάρχει μεγαλύτερο «ρίσκο» να αποφοιτήσουν αφού η συνάρτησή τους έχει μεγαλύτερες τιμές. Οι διαφορές ανάμεσα στα έτη είναι εμφανείς και παρουσιάζεται η τάση αύξησης της «επικινδυνότητας» αποφοίτησης όσο κινούμαστε από το έτος εισαγωγής 1999 προς το έτος εισαγωγής 2005.

Hazard Function for patterns 1 - 7



Διάγραμμα 32: Συνάρτηση επικινδυνότητας ανά έτος εισαγωγής

3.4.2 Επιλογή μεταβλητών με τη μέθοδο Forward : LR

Για να επιβεβαιώσουμε την ορθότητα της επιλογής των μεταβλητών που κάναμε στην προηγούμενη ενότητα εκτελούμε ξανά την διαδικασία COX Regression βάζοντας αρχικά όλες τις μεταβλητές στο μοντέλο. Αυτή τη φορά όμως δεν χρησιμοποιούμε τη μέθοδο Enter η οποία χρησιμοποιεί όλες τις ανεξάρτητες μεταβλητές δείχνοντας τις συσχετίσεις τους αλλά τη μέθοδο Forward: LR η οποία θα δώσει και το τελικό μοντέλο πρόβλεψης απομακρύνοντας τις μεταβλητές που δεν είναι στατιστικά σημαντικές ή δεν προσθέτουν κάποια επιπλέον πληροφορία στο μοντέλο πρόβλεψης. Η μέθοδος αυτή επιλέγει προοδευτικά τις μεταβλητές που είναι στατιστικά σημαντικές και τις υπόλοιπες τις απορρίπτει. Αναλυτικότερα η μέθοδος, βασίζεται στη στατιστική του λόγου πιθανοτήτων (likelihood ratio statistics), όπου η παρατήρηση του τρέχοντος μοντέλου συγκρίνεται με το μοντέλο όπου η ανεξάρτητη μεταβλητή απομακρύνεται. Εάν η μεταβλητή που απομακρύνεται, δημιουργεί σημαντική διαφορά στο πόσο καλά το μοντέλο προσαρμόζεται στις παρατηρούμενες τιμές, τότε η μεταβλητή συγκρατείται στο μοντέλο. Εάν όμως, η μεταβλητή που απομακρύνεται, δημιουργεί μικρή διαφορά στο μοντέλο, τότε αυτή απορρίπτεται.

Εκτελούμε λοιπόν τη στατιστική διαδικασία COX REGRESSION τοποθετώντας τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΦΥΛΟ (κατηγορική μεταβλητή: Άνδρας – Γυναίκα)

-> ΥΠΗΚΟΟΤΗΤΑ (κατηγορική μεταβλητή: Ελληνική-Κυπριακή-Αλβανική)

-> ΕΓΓΡΑΦΗ ΜΕ (κατηγορική μεταβλητή: Εισιτήριες εξετάσεις-ΤΕΕ - Άλλο)

-> ΠΤΥΧΙΑΚΗ (ποσοτική μεταβλητή: Βαθμός πτυχιακής)

-> ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ (ποσοτική μεταβλητή)

- > ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: 1999-2005)
- > ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: Χειμερινό
εξάμηνο – Εαρινό εξάμηνο)

Αρχικά, όλες οι μεταβλητές είναι εκτός μοντέλου. Στο πρώτο βήμα εισέρχεται η ερμηνευτική μεταβλητή ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ, στο δεύτερο βήμα η μεταβλητή ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ και στο τρίτο βήμα η μεταβλητή ΦΥΛΟ (πίνακας 54).

Omnibus Tests of Model Coefficients

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1 ^a	8300,500	189,486	1	,000	166,667	1	,000	166,667	1	,000
2 ^b	8212,086	318,342	7	,000	88,415	6	,000	255,082	7	,000
3 ^c	8206,291	325,158	8	,000	5,795	1	,016	260,877	8	,000

- a. Variable(s) Entered at Step Number 1: ptyxio
- b. Variable(s) Entered at Step Number 2: etos
- c. Variable(s) Entered at Step Number 3: sex
- d. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 8467,168
- e. Beginning Block Number 1. Method = Forward Stepwise (Likelihood Ratio)

Πίνακας 54: Έλεγχος Χ-τετράγωνο για το συνολικό μοντέλο

Συνεπώς, και με αυτή τη μέθοδο, επιλέγονται να συμμετέχουν στο μοντέλο οι μεταβλητές που χρησιμοποιήσαμε στην προηγούμενη ενότητα. Παρακάτω ακολουθεί ο πίνακας 55 με τους συντελεστές Β και EXP(B), τις στατιστικές σημαντικότητες και τα διαστήματα εμπιστοσύνης για τον κάθε συντελεστή. Η παρουσίαση των αποτελεσμάτων γίνεται ανά βήμα και διαφοροποιούνται ελάχιστα καθώς εισέρχονται νέες μεταβλητές στο μοντέλο.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
Step 1 ptyxio	1,142	,082	192,065	1	,000	3,133	2,666	3,682
Step 2 ptyxio	1,193	,085	198,053	1	,000	3,297	2,792	3,893
etos			106,660	6	,000			
etos(1)	-1,948	,280	48,536	1	,000	,143	,082	,247
etos(2)	-1,834	,268	46,877	1	,000	,160	,095	,270
etos(3)	-1,696	,267	40,470	1	,000	,183	,109	,309
etos(4)	-1,669	,265	39,551	1	,000	,188	,112	,317
etos(5)	-1,288	,271	22,686	1	,000	,276	,162	,468
etos(6)	-,678	,280	5,872	1	,015	,507	,293	,878
Step 3 sex	-,189	,079	5,723	1	,017	,828	,709	,966
ptyxio	1,168	,085	187,819	1	,000	3,216	2,721	3,800
etos			98,154	6	,000			
etos(1)	-1,890	,281	45,329	1	,000	,151	,087	,262
etos(2)	-1,811	,268	45,563	1	,000	,163	,097	,277
etos(3)	-1,655	,267	38,348	1	,000	,191	,113	,323
etos(4)	-1,660	,266	39,042	1	,000	,190	,113	,320
etos(5)	-1,278	,271	22,271	1	,000	,278	,164	,474
etos(6)	-,694	,280	6,132	1	,013	,499	,288	,865

Πίνακας 55: Έλεγχος σημαντικότητας για τις ανεξάρτητες μεταβλητές

Στο βήμα 3 έχουμε τρεις ανεξάρτητες μεταβλητές, δύο ποιοτικές και μια ποσοτική.

Όσον αφορά τις ποιοτικές μεταβλητές, η στήλη Exp(B) εκφράζει την αλλαγή επικινδυνότητας (HR) από την κατηγορία αναφοράς στις άλλες κατηγορίες. Ξεκινώντας από τη μεταβλητή ΦΥΛΟ, βλέπουμε ότι ο συντελεστής είναι 0,828 για τους άνδρες, άρα η «επικινδυνότητα» να αποφοιτήσουν είναι μικρότερη από ότι στις γυναίκες. Για τη μεταβλητή ΕΤΟΣ ΦΟΙΤΗΣΗΣ, βλέπουμε ότι ο συντελεστής HR είναι 0.151, 0.163, 0.191, 0.190, 0.278 και 0.499 για τα έτη 1999, 2000, 2001, 2002, 2003 και 2004 αντίστοιχα. Συνεπώς, η «επικινδυνότητα» να αποφοιτήσουν οι εισαχθέντες αυτών των ετών είναι μικρότερη από τους εισαχθέντες του 2005. Ενδιαφέρον παρουσιάζει το γεγονός ότι όσο περνάνε τα χρόνια από το 1999 προς το 2005, η «επικινδυνότητα» αποφοίτησης αυξάνει.

Όσον αφορά την ποσοτική μεταβλητή, η στήλη Exp(B) εκφράζει την εκτιμώμενη αλλαγή της επικινδυνότητας σε μια μονάδα αλλαγής της ανεξάρτητης μεταβλητής. Εδώ βλέπουμε ότι ο συντελεστής είναι HR=3,216, που υποδηλώνει ότι όταν ο βαθμός του πτυχίου αυξάνεται κατά μια μονάδα τότε η «επικινδυνότητα» αποφοίτησης αυξάνει κατά 3,216 φορές.

Με βάση το γεγονός ότι όλες οι προηγούμενες ανεξάρτητες μεταβλητές είναι στατιστικά σημαντικές, συμπεραίνουμε ότι οι γυναίκες παρουσιάζουν χαμηλότερο δείκτη επιβίωσης (μικρότερους χρόνους φοίτησης), όπως και οι εισαχθέντες σε νεότερα χρόνια. Τέλος, όσο μεγαλύτερος είναι ο βαθμός του πτυχίου τόσο μικρότερος χρόνος φοίτησης έχει επιτευχθεί από το φοιτητή.

3.5 Έλεγχος υποθέσεων του μοντέλου επιβίωσης COX Regression

Τα συμπεράσματα που εξήχθησαν στην ενότητα 3.4 έχουν αξία μόνο υπό την προϋπόθεση ότι οι επιδράσεις των ανεξάρτητων μεταβλητών είναι σταθερές στο χρόνο. Αν δεν ισχύει κάτι τέτοιο, τότε το μοντέλο επιβίωσης Cox Regression δεν θα παρέχει καλή προσαρμογή των δεδομένων. Επομένως, πρέπει να ελέγξουμε την υπόθεση ότι η αναλογία των συναρτήσεων επιβίωσης μεταξύ δύο ομάδων παραμένει σταθερή στη διάρκεια του χρόνου. Υπάρχουν διαθέσιμες διαφορετικές προσεγγίσεις για τον έλεγχο αυτή της υπόθεσης:

- § Να εξεταστεί το γράφημα επιβίωσης ή το γράφημα Log-minus-log του μοντέλου επιβίωσης Cox Regression σε σχέση με την κατηγορική μεταβλητή που ορίζεται ως στρωματοποιημένη (strata).
- § Αποθηκεύουμε τα μερικά κατάλοιπα (partial residual) και δημιουργούμε το γράφημα τους σε σχέση με το χρόνο.
- § Προσαρμογή του μοντέλου επιβίωσης Cox Regression σε σχέση με την επίδραση του χρόνου σε διαφορετικά διαστήματα (time-varying covariate) και να ελεγχθεί η σημαντικότητα και η συνεισφορά.

3.5.1 Έλεγχος για τις ποιοτικές μεταβλητές

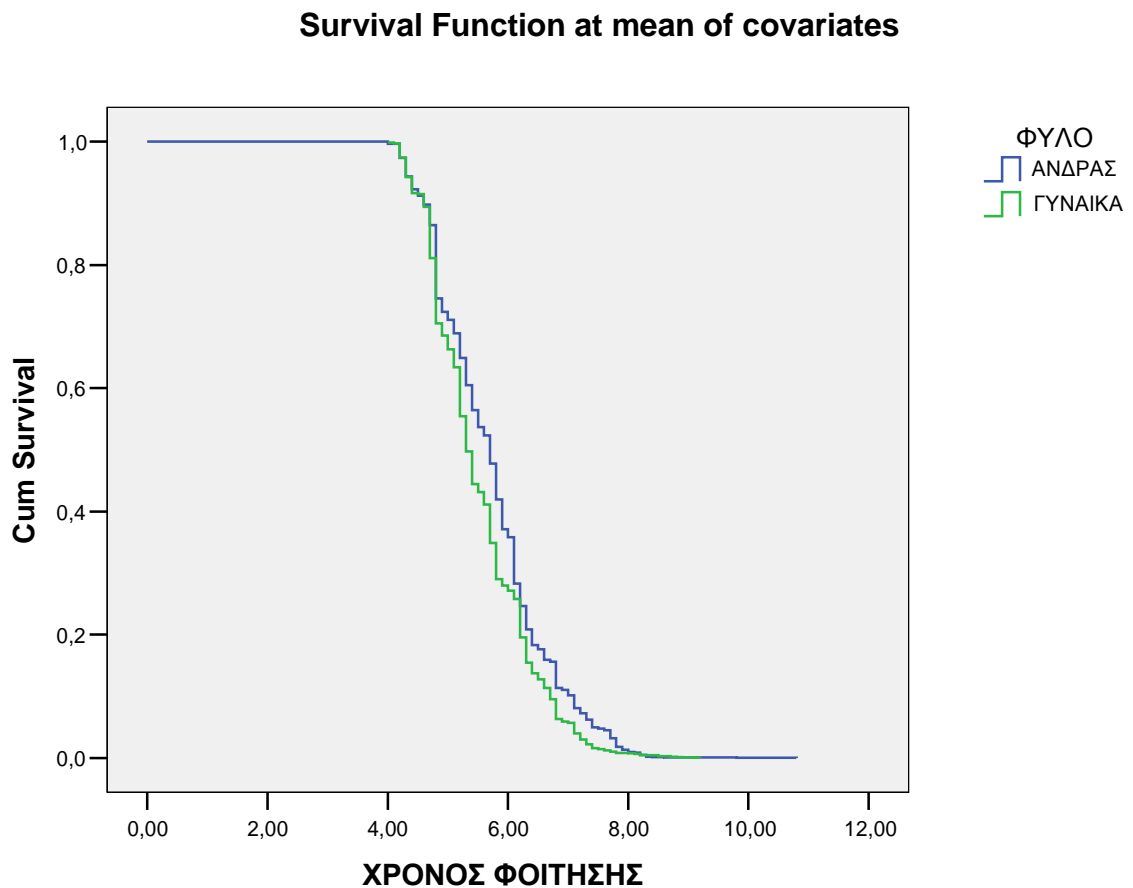
Εδώ, ο έλεγχος θα γίνει χρησιμοποιώντας την πρώτη προσέγγιση. Αρχικά, θα τοποθετήσουμε ως strata την μεταβλητή ΦΥΛΟ (πίνακας 56).

Stratum	Strata label	Event	Censored	Censored Percent
1	ΑΝΔΡΑΣ	281	0	,0%
2	ΓΥΝΑΙΚΑ	466	0	,0%
Total		747	0	,0%

a. The strata variable is : ΦΥΛΟ

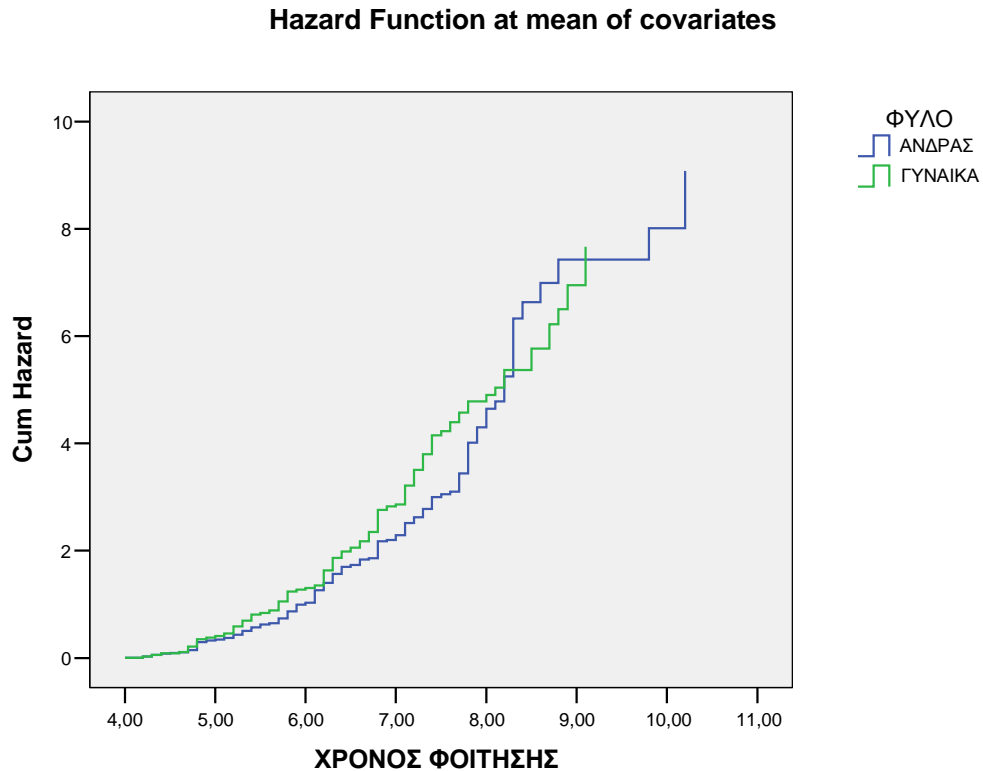
Πίνακας 56: Μεταβλητή strata το «φύλο»

Όσον αφορά το διάγραμμα επιβίωσης 33, παρατηρούμε ότι δεν υπάρχει απόκλιση μεταξύ των ανδρών και των γυναικών και οι συναρτήσεις τους σχεδόν ταυτίζονται. Δηλαδή, η αναλογία επικινδυνότητας μεταξύ των δύο ομάδων είναι σταθερή στο χρόνο και ικανοποιείται η υπόθεση του μοντέλου επιβίωσης Cox Regression. Άρα, τα συμπεράσματα που προέκυψαν στην προηγούμενη ενότητα έχουν ισχύ για τη μεταβλητή ΦΥΛΟ.



Διάγραμμα 33: Συνάρτηση επιβίωσης ανά φύλο

Στο ίδιο συμπέρασμα καταλήγουμε και με τη χρήση του διαγράμματος 34 (log minus log). Παρατηρούμε ότι οι γραμμές των ανδρών και των γυναικών είναι παράλληλες κατά τη διάρκεια του χρόνου που καλύπτουν τα δεδομένα μας. Άρα δεν υπάρχει πρόβλημα στον έλεγχο της υπόθεσης.



Διάγραμμα 34: *Log minus Log* ανά φύλο

Τον ίδιο έλεγχο διεξάγουμε και για την ανεξάρτητη ποιοτική μεταβλητή ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ που την τοποθετούμε τώρα ως strata (πίνακας 57).

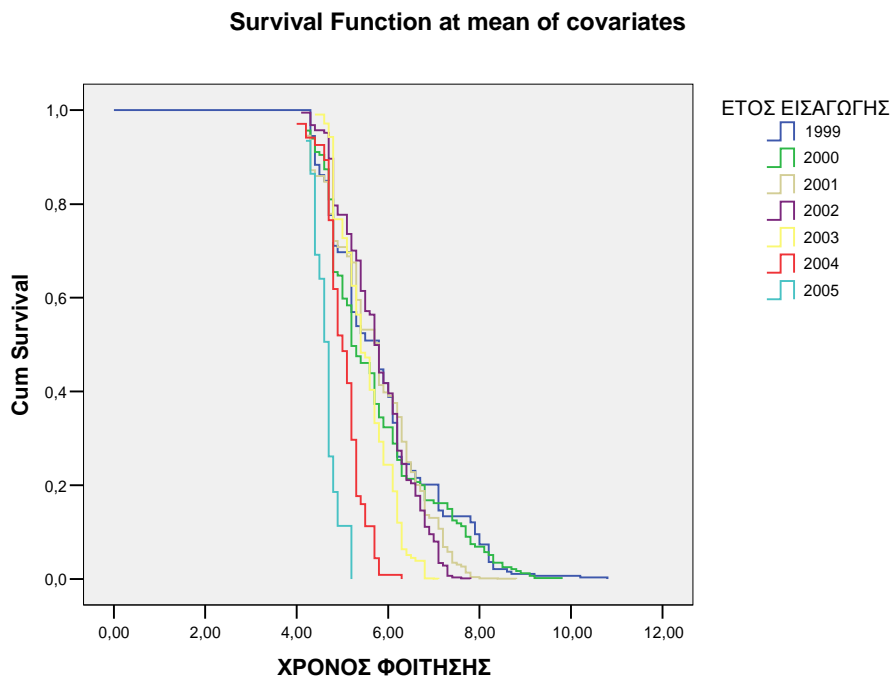
Stratum Status^a

Stratum	Strata label	Event	Censored	Censored Percent
1	1999	82	0	,0%
2	2000	155	0	,0%
3	2001	163	0	,0%
4	2002	153	0	,0%
5	2003	114	0	,0%
6	2004	63	0	,0%
7	2005	17	0	,0%
Total		747	0	,0%

a. The strata variable is : ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ

Πίνακας 57: Μεταβλητή strata το «έτος εισαγωγής»

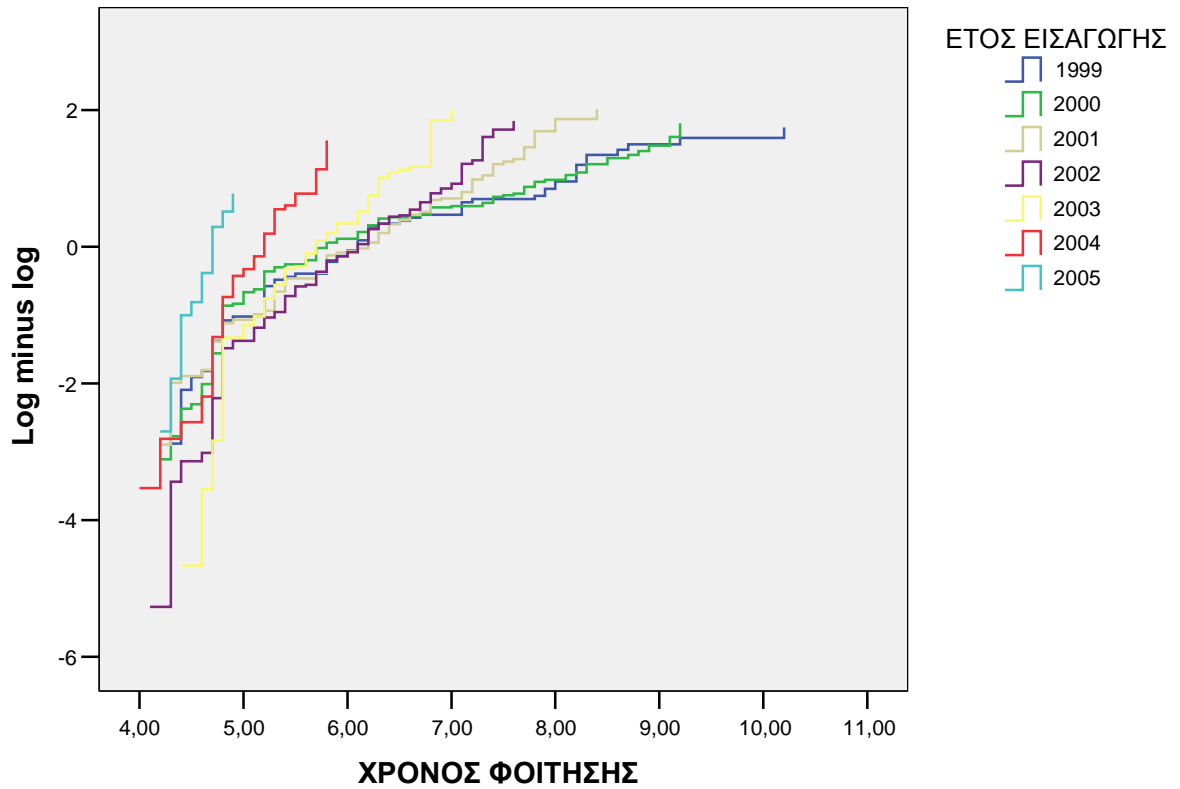
Όσον αφορά το διάγραμμα επιβίωσης 35, παρατηρούμε ότι υπάρχει απόκλιση κατά τη διάρκεια του χρόνου μεταξύ των διαφορετικών ετών εισαγωγής στη σχολή. Δηλαδή, η αναλογία επικινδυνότητας μεταξύ των διαφορετικών ομάδων δεν είναι σταθερή στο χρόνο και δεν ικανοποιείται η υπόθεση του μοντέλου επιβίωσης Cox Regression. Άρα, το μοντέλο δεν έχει καλή προσαρμογή στα δεδομένα για τη μεταβλητή αυτή.



Διάγραμμα 35: Συνάρτηση επιβίωσης ανά έτος εισαγωγής

Στο ίδιο συμπέρασμα καταλήγουμε και με τη χρήση του διαγράμματος 36 (log minus log). Παρατηρούμε ότι οι γραμμές των διαφόρων ετών εισαγωγής δεν είναι παράλληλες μεταξύ τους κατά τη διάρκεια του χρόνου που καλύπτουν τα δεδομένα μας, ιδιαίτερα μετά το 7^ο έτος. Άρα, υπάρχει πρόβλημα στον έλεγχο της υπόθεσης.

LML Function at mean of covariates



Διάγραμμα 36: *Log minus Log* ανά έτος εισαγωγής

Σε αυτή την περίπτωση μπορεί να δημιουργηθεί ένας όρος αλληλεπίδρασης μεταξύ της μεταβλητής «Έτος εισαγωγής» και μιας συνάρτησης του χρόνου, στη συνάρτηση επικινδυνότητας, που δεν είναι σταθερή στο χρόνο.

3.5.2 Έλεγχος για τις ποσοτικές μεταβλητές

Εδώ, ο έλεγχος θα γίνει χρησιμοποιώντας την δεύτερη προσέγγιση. Αρχικά, εκτελούμε τη στατιστική διαδικασία COX REGRESSION τοποθετώντας τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

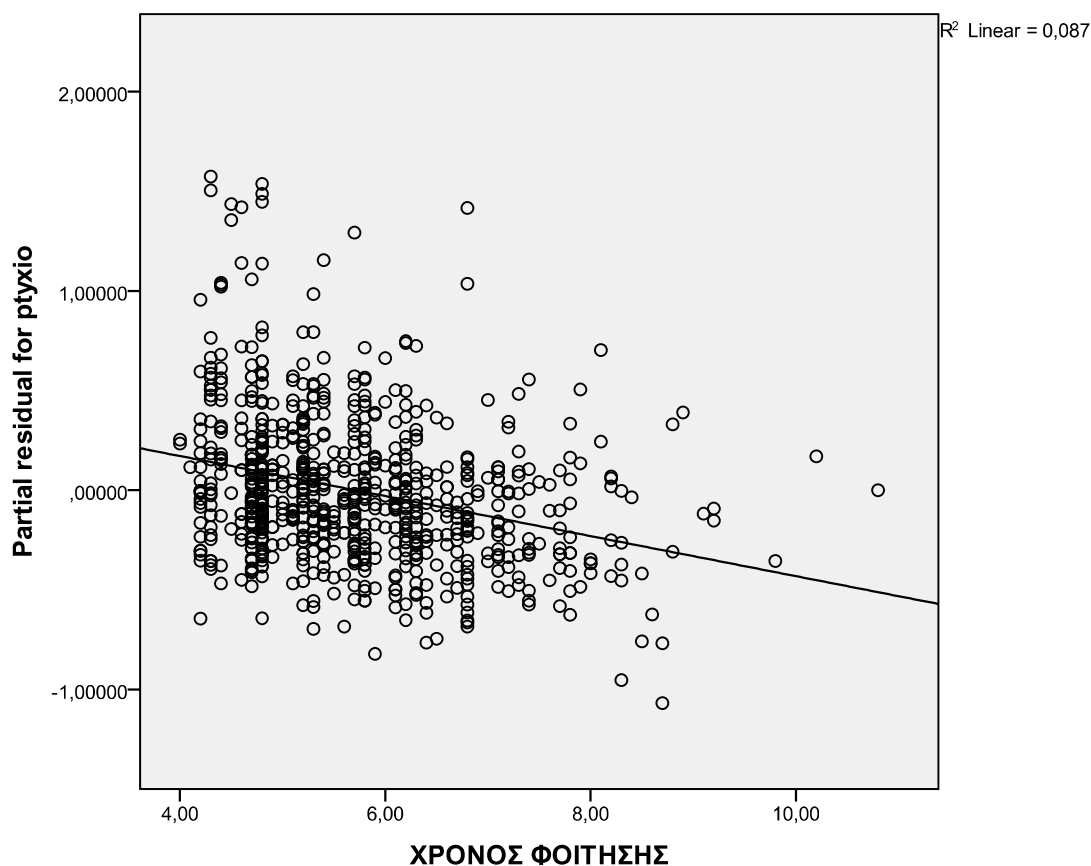
Status -> ΚΑΤΑΣΤΑΣΗ

Covariates -> ΦΥΛΟ (κατηγορική μεταβλητή: Άνδρας - Γυναίκα)

-> ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: 1999-2005)

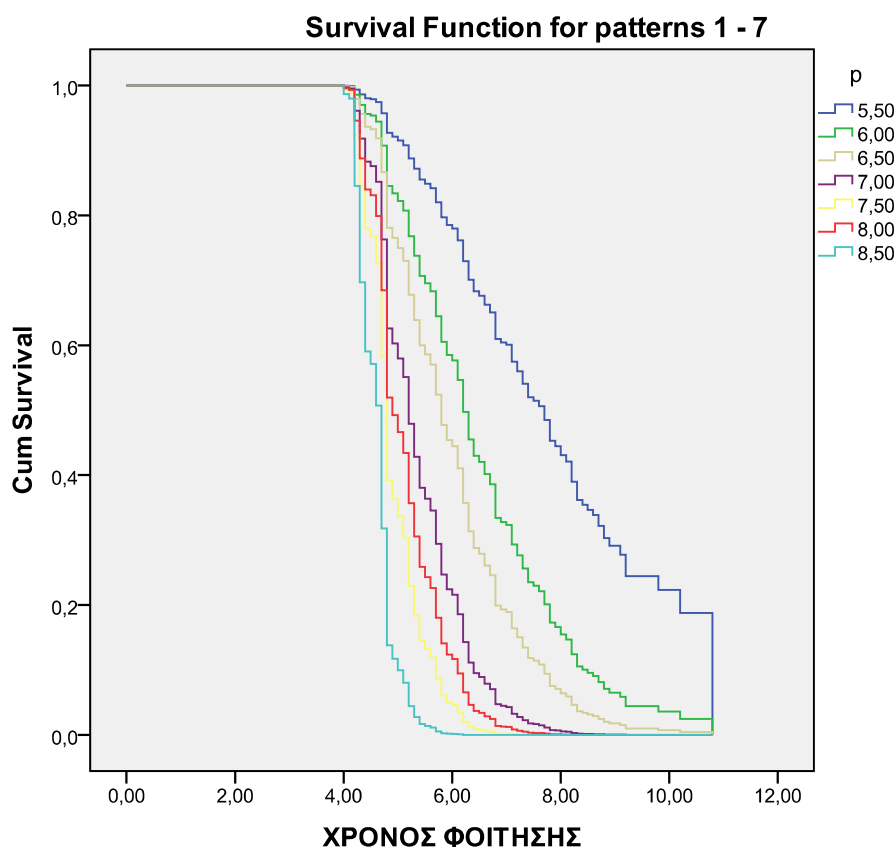
-> ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ (ποσοτική μεταβλητή)

Για το μοντέλο αυτό σώζουμε τα μερικά κατάλοιπα για την κάθε ανεξάρτητη μεταβλητή. Έπειτα, αντιπαραβάλλουμε σε ένα γράφημα τα μερικά κατάλοιπα της ανεξάρτητης μεταβλητής ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ σε σχέση με το χρόνο φοίτησης.



Διάγραμμα 37: Κατάλοιπα του «βαθμού πτυχίου» σε σχέση με το χρόνο

Για την μεταβλητή «βαθμός πτυχίου» από το διάγραμμα διασποράς 37 με τα μερικά κατάλοιπα της ανεξάρτητης μεταβλητής σε σχέση με το χρόνο φοίτησης προκύπτει ότι η μεταβλητή συστηματικά μεταβάλλεται σε σχέση με τον χρόνο (αρνητική συσχέτιση). Επιπροσθέτως, η μη ύπαρξη αναλογικότητας στον κίνδυνο λόγω της χρονικής εξάρτησης συνάγεται και από το διάγραμμα 38 το οποίο δημιουργήθηκε με την ομαδοποίηση της συνεχόμενης μεταβλητής «βαθμός πτυχίου» ανά 0,5 μονάδες δηλαδή: [5 έως 5,5), [5,5 έως 6) κλπ. Αναλυτικότερα, στο διάγραμμα 38 παρατηρείται μια σαφής διαφοροποίηση της αναλογικότητας (δηλαδή οι γραμμές δεν είναι παράλληλες μεταξύ τους στο χρόνο) στις πολύ μεγάλες βαθμολογίες (>8,5) όπως και στις πολύ μικρές βαθμολογίες (<5,5).



Διάγραμμα 38: Συνάρτηση επιβίωσης ανά βαθμό πτυχίου

3.6 Μοντέλο επιβίωσης COX με χρονικά εξαρτημένες ανεξάρτητες μεταβλητές (time-varying covariate)

3.6.1. Ποιοτικές μεταβλητές

Εφόσον υπάρχει πρόβλημα χρονικής εξάρτησης για την μεταβλητή «Έτος εισαγωγής» και οι γραμμές των διαφόρων ετών εισαγωγής δεν είναι παράλληλες μεταξύ τους στο διάγραμμα LML δεν μπορεί να εισέρθει με αυτήν την μορφή στο μοντέλο. Με δεδομένο ότι δεν παρουσιάζει ιδιαίτερο ενδιαφέρον από πλευράς πρόβλεψης αλλά είναι στατιστικά σημαντική και επηρεάζει το μοντέλο θα χρησιμοποιηθεί ως σωματοποιημένη μεταβλητή.

3.6.2. Ποσοτικές μεταβλητές

Εφόσον υπάρχει πρόβλημα χρονικής εξάρτησης με τη μεταβλητή «βαθμός πτυχίου» δημιουργήθηκε και εδώ μία νέα συνάρτηση χρόνου. Αυτή η μεταβλητή προκύπτει από τον πολλαπλασιασμό της μεταβλητής χρόνου T με τη μεταβλητή βαθμός πτυχίου δηλαδή «T * βαθμός πτυχίου», επομένως δημιουργείται μια νέα μεταβλητή αλληλεπίδρασης χρόνου και βαθμό πτυχίου. Ως εκ τούτου στο νέο εκτεταμένο μοντέλο επιβίωσης Cox Regression ο συντελεστής της μεταβλητής «βαθμός πτυχίου» θα αναπαριστά την επίδραση της μεταβλητής αυτής όταν ο χρόνος της έρευνας μεταβάλλεται ενώ ο συντελεστής για την αλλαγή του χρόνου επίδρασης (time-varying covariate) θα αναπαριστά την αλλαγή της επίδρασης της μεταβλητής «βαθμού πτυχίου» στο χρόνο.

Στη συνέχεια και διατηρώντας το υπόλοιπο μοντέλο με τις ίδιες μεταβλητές όπως στο απλό μοντέλο Cox Regression (χωρίς χρονική εξάρτηση) εισάγουμε πέραν των υπολοίπων μεταβλητών και την μεταβλητή αλληλεπίδρασης «έτος * βαθμός πτυχίου». Η τιμή αύξησης του κινδύνου (HR) σε σχέση με τον βαθμό και τον χρόνο T είναι τώρα $HR = \exp(3,810 - 0,476 * T)$ με δεδομένο ότι και η μεταβλητή «βαθμός πτυχίου» και η μεταβλητή

«χρόνος** βαθμός πτυχίου» είναι στατιστικά σημαντικές (βλέπε πίνακα 59). Για παράδειγμα αν ο χρόνος είναι T=4 ο κίνδυνος είναι $HR = \exp(3,810 - 0,476 \cdot 4) = 6,72$. Όμοια υπολογίζουμε τις τιμές για τους χρόνους T={4, 5, 6, 7, 8, 9, 10} ότι είναι $HR = \{6.7, 4.2, 2.6, 1.6, 1.0, 0.6, 0.4\}$ ενώ στο απλό μοντέλο η αύξηση του κινδύνου για κάθε αύξηση ενός βαθμού ήταν σταθερά $HR=3.2$. Τέλος παρατηρούμε ότι οι εκτιμώμενες επιδράσεις των μεταβλητών έτος και φύλο είναι περίπου ίδιες με το προηγούμενο μοντέλο όταν δεν είχαμε λάβει υπόψη το παράγοντα χρόνο.

Variables in the Equation									
	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)		
							Lower	Upper	
Step 1	ptyxio	1,247	,086	211,635	1	,000	3,479	2,941	4,115
Step 2	T_COV_	-,469	,085	30,590	1	,000	,626	,530	,739
	ptyxio	3,783	,462	67,011	1	,000	43,943	17,764	108,704
Step 3	T_COV_	-,476	,085	31,383	1	,000	,621	,526	,734
	sex	-,188	,080	5,541	1	,019	,829	,708	,969
	ptyxio	3,810	,464	67,283	1	,000	45,145	18,166	112,193

Πίνακας 59: Μοντέλο Cox παλινδρόμησης με χρονικά εξαρτημένη την μεταβλητή «βαθμός πτυχίου»

Το τελικό μοντέλο πρόβλεψης προκύπτει από τη στατιστική διαδικασία "COX REGRESSION – TIME DEPENDENT COVARIATES" με την μέθοδο εισαγωγής μεταβλητών "Forward LR" τοποθετώντας τις ακόλουθες μεταβλητές στο SPSS:

Time -> ΧΡΟΝΟΣ ΦΟΙΤΗΣΗΣ

Status -> ΚΑΤΑΣΤΑΣΗ «χρόνος** βαθμός πτυχίου»

Strata-> ΕΤΟΣ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: 1999-2005)

Covariates -> ΧΡΟΝΟΣ * ΒΑΘΜΟΣ ΠΤΥΧΕΙΟΥ (χρονικά εξαρτημένη)

-> ΦΥΛΟ (κατηγορική μεταβλητή: Άνδρας – Γυναίκα)

-> ΥΠΗΚΟΟΤΗΤΑ (κατηγορική μεταβλητή: Ελληνική-Κυπριακή-Αλβανική)

-> ΕΓΓΡΑΦΗ ΜΕ (κατηγορική μεταβλητή: Εισιτήριες εξετάσεις-ΤΕΕ - Άλλο)

-> ΠΤΥΧΙΑΚΗ (ποσοτική μεταβλητή: Βαθμός πτυχιακής)

- > ΒΑΘΜΟΣ ΠΤΥΧΙΟΥ (ποσοτική μεταβλητή)
- > ΕΞΑΜΗΝΟ ΕΙΣΑΓΩΓΗΣ (κατηγορική μεταβλητή: Χειμερινό
εξάμηνο – Εαρινό εξάμηνο)

Συμπεράσματα

Στην την έρευνα που παρουσιάστηκε εξεταστήκαν παράγοντες που επηρεάζουν το χρόνο φοίτησης στη τριτοβάθμια εκπαίδευση. Τα στοιχεία αφορούν πτυχιούχους του τμήματος Επιχειρηματικού Σχεδιασμού και Πληροφορικών Συστημάτων του ΤΕΙ Πάτρας για τα έτη 2000 έως 2009.

Αναλυτικότερα στην παρούσα έρευνα χρησιμοποιήθηκε στατιστική μεθοδολογίας ανάλυσης επιβίωσης και συγκεκριμένα το μοντέλο της COX παλινδρόμησης προκειμένου να εξεταστεί η σχέση της διάρκειας φοίτησής (επιβίωση) με διάφορες επεξηγητικές μεταβλητές πρόβλεψης με βάση τα δημογραφικά χαρακτηριστικά των φοιτητών. Ως μεταβλητές πρόβλεψης του χρόνου φοίτησης χρησιμοποιήθηκαν οι εξής: φύλο, υπηκοότητα, τρόπος εισαγωγής, βαθμός πτυχιακής, βαθμός πτυχίου, έτος εισαγωγής και εξάμηνο εισαγωγής.

Τόσο με τον επιμέρους έλεγχο της επίδρασης του κάθε δημογραφικού χαρακτηριστικού ξεχωριστά όσο και συγκεντρωτικό μοντέλο το οποίο δημιουργήθηκε με τη μέθοδο stepwise Forward LR (Likelihood Ratio) δηλαδή με την σταδιακή είσοδο των μεταβλητών κατά σειρά διαπιστώθηκε ότι στατιστικά σημαντική επίδραση πάνω στο χρόνο φοίτησης είχαν οι μεταβλητές φύλο, έτος εισαγωγής και βαθμός πτυχίου.

Τέλος, έγινε έλεγχος της υπόθεσης για τη σταθερότητα των διαφόρων συμμεταβλητών στο χρόνο. Διαπιστώθηκε κατ' αρχήν ότι η αναλογία επικινδυνότητας μεταξύ των διαφορετικών ετών εισαγωγής δεν είναι σταθερή στο χρόνο και δεν ικανοποιείται η υπόθεση του μοντέλου επιβίωσης Cox Regression. Επίσης, για την μεταβλητή «βαθμός πτυχίου» προέκυψε ότι η μεταβλητή συστηματικά μεταβάλλεται σε σχέση με τον χρόνο (αρνητική συσχέτιση). Για να λυθεί το πρόβλημα δημιουργήθηκαν επεκτάσεις του αρχικού μοντέλου επιβίωσης με στρωματοποίηση (stata) και χρονικά εξαρτημένες μεταβλητές (time-varying covariate) που ικανοποιούν πλήρως τις προϋποθέσεις της Cox παλινδρόμησης παρέχοντας ασφαλή συμπεράσματα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

David Garson 2010. Cox Regression, from Statnotes: Topics in Multivariate Analysis (Available at:

<http://faculty.chass.ncsu.edu/garson/PA765/cox.htm>)

SPSS BI Greece A.E. 2008 : SPSS Medical Data Analysis

SPSS- PASW Statistics 18, IBM Company 2009. Help /Case studies/
Advance statistics option/Cox regression

Κ.Φωκιανός , Χ.Χαραλάμπους, 2001. Ανάλυση επιβίωσης. Παν/μιο Κύπρου,
Τμήμα Μαθηματικών & Στατιστικής (Διαθέσιμο στο:

<http://www2.ucy.ac.cy/~fokianos/GreekRbook/survival.pdf>)

Ελπίδα Θ.Παύλου, 2006 Το μοντέλο αναλογικού κινδύνου του Cox στην
Ανάλυση Επιβίωσης Τμήμα Μαθηματικών ΑΠΘ (Διαθέσιμο στο:

<http://invenio.lib.auth.gr/record/68568/files/gri-2007-222.pdf>)