

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

**Σχολή Χρηματοοικονομικής και
Στατιστικής**



**Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΠΑΡΑΜΕΤΡΙΚΑ ΜΟΝΤΕΛΑ
ΕΠΙΒΙΩΣΗΣ ΑΣΘΕΝΩΝ ΓΙΑ
ΔΙΑΦΟΡΑ ΕΙΔΗ ΚΑΡΚΙΝΟΥ**

Άννα Εργαζάκη

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

**Πειραιάς
Σεπτέμβριος 2024**

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

**Σχολή Χρηματοοικονομικής και
Στατιστικής**



**Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΠΑΡΑΜΕΤΡΙΚΑ ΜΟΝΤΕΛΑ
ΕΠΙΒΙΩΣΗΣ ΑΣΘΕΝΩΝ ΓΙΑ
ΔΙΑΦΟΡΑ ΕΙΔΗ ΚΑΡΚΙΝΟΥ**

Άννα Εργαζάκη

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

*Πειραιάς
Σεπτέμβριος 2024*

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

Τα μέλη της Επιτροπής ήταν:

- Τζαβελάς Γεώργιος, Αναπληρωτής Καθηγητής (Επιβλέπων)
- Λαβδανίτη Μαρία, Καθηγήτρια
- Ρακιτζής Αθανάσιος, Επίκουρος Καθηγητής

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

UNIVERSITY OF PIRAEUS
School of Finance and Statistics



Department of Statistics and Insurance Science

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

**PARAMETRIC SURVIVAL MODELS
FOR VARIOUS TYPES OF CANCER**

By

Anna Ergazaki

MSc Dissertation

submitted to the Department of Statistics and
Insurance Science of the University of Piraeus in
partial fulfilment of the requirements for the
degree of Master of Science in Applied Statistics

Piraeus, Greece
September 2024

Στους γονείς μου

Γιώργος και Αμαλία

Περίληψη

Η παρούσα διπλωματική εργασία εστιάζει στη μελέτη των παραγόντων που επηρεάζουν την επιβίωση ασθενών με καρκίνο του εγκεφάλου, του μαστού και του παγκρέατος. μέσα από τη χρήση περιγραφικής στατιστικής και παραμετρικών μοντέλων επιβίωσης. Για το σκοπό αυτό χρησιμοποιήθηκε η βάση δεδομένων από το SEER database. Το ενδιαφέρον εστιάστηκε στην εύρεση των παραγόντων εκείνων, οι οποίοι επηρεάζουν τη θνησιμότητα.

Στο πρώτο κεφάλαιο, πραγματοποιείται μια εισαγωγή στους συγκεκριμένους τύπους καρκίνου και στη σημασία της στατιστικής ανάλυσης στην κατανόηση της επιβίωσης.

Στο δεύτερο κεφάλαιο, παρουσιάζεται αναλυτικά η μεθοδολογία που ακολουθήθηκε, περιλαμβάνοντας την θεωρητική εφαρμογή μοντέλων αναλογικού κινδύνου καθώς και παραμετρικών μοντέλων επιβίωσης.

Στο τρίτο κεφάλαιο, διεξάγεται η περιγραφική ανάλυση στα σύνολα δεδομένων παρουσιάζοντας διάφορα γραφήματα και ποσοστά.

Το τέταρτο κεφάλαιο επικεντρώνεται στον υπολογισμό συσχετίσεων και στους ελέγχους τους. Σε πρώτη φάση δόθηκε έμφαση στις ποσοτικές και ποιοτικές μεταβλητές βάση της μεταβλητής «Μήνες επιβίωσης», όπου και εκτός των ελέγχων παρουσιάστηκαν και κατάλληλα γραφήματα για την καλύτερη κατανόηση του συνόλου δεδομένων. Στην δεύτερη φάση αναλύθηκαν οι ίδιες μεταβλητές βάση της μεταβλητής «Κατάσταση επιβίωσης ασθενή». Προφανώς αυτές οι μέθοδοι παρουσιάστηκαν για τα τρία σύνολα δεδομένων.

Στο πέμπτο κεφάλαιο υπάρχει η προσαρμογή των μοντέλων αναλογικού κινδύνου και επιταχυνόμενου χρόνου αποτυχίας με μεταβλητές απόκρισης τις μεταβλητές των μηνών επιβίωσης και την κατάσταση επιβίωσης του ασθενή. Στόχος του κεφαλαίου είναι η εύρεση των καταλληλότερων μοντέλων βάση των στατιστικών ελέγχων που διεξάγονται.

Τέλος, το έκτο κεφάλαιο συνοψίζει τα βασικά ευρήματα της έρευνας και προτείνει κατευθύνσεις για μελλοντική έρευνα και βελτίωση των στρατηγικών αντιμετώπισης των συγκεκριμένων τύπων καρκίνου.

Abstract

This thesis investigates the factors influencing the survival rates of patients with brain, breast, and pancreatic cancer through the application of descriptive statistics and parametric survival models, utilizing the SEER database. The primary objective is to identify determinants of mortality in these cancer types.

The first chapter provides a comprehensive introduction to brain, breast, and pancreatic cancers, highlighting their clinical significance and the role of statistical analysis in elucidating survival patterns. It sets the stage for understanding how statistical methodologies contribute to insights into patient outcomes.

The second chapter delineates the methodological framework employed in the study, focusing on the theoretical underpinnings and practical applications of proportional hazard models and parametric survival models. This section elucidates the analytical techniques utilized to interpret survival data effectively.

In the third chapter, a detailed descriptive analysis of the datasets is presented. This includes various graphical representations and statistical summaries that illustrate the distribution and characteristics of the data related to the survival of patients across the three cancer types.

The fourth chapter is dedicated to the calculation and testing of correlations within the datasets. It explores the relationships between quantitative and qualitative variables with respect to the "Survival Months" variable, accompanied by appropriate graphical depictions. Subsequently, the analysis extends to investigate correlations with the "Survival Status" variable, incorporating similar methodological approaches.

The fifth chapter focuses on the adaptation and application of proportional hazard models and accelerated failure time models. The response variables examined include survival months and survival status. The chapter aims to identify the most appropriate statistical models based on rigorous statistical testing and evaluation.

The final chapter summarizes the principal findings of the research, synthesizing the results from the statistical analyses and model fitting procedures. It also proposes avenues for future research and recommendations for enhancing strategies in the management and treatment of brain, breast, and pancreatic cancers.

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1°	5
1. Εισαγωγή	5
1.1. Καρκίνος του Εγκεφάλου	6
1.2. Καρκίνος του Μαστού	7
1.3. Καρκίνος του Παγκρέατος.....	7
1.4. Στατιστική και Καρκίνος	8
ΚΕΦΑΛΑΙΟ 2°	10
2. Μεθοδολογία.....	10
2.1. Μοντέλο αναλογικού κινδύνου.....	10
2.2. Παραμετρικά Μοντέλα επιβίωσης.....	11
ΚΕΦΑΛΑΙΟ 3°	16
3. Περιγραφική Στατιστική.....	16
3.1. Περιγραφή δεδομένων καρκίνου του εγκεφάλου	16
3.1.1. Περιγραφική ανάλυση ποσοτικών μεταβλητών και ποιοτικών μεταβλητών καρκίνου του εγκεφάλου.....	17
3.2. Περιγραφή δεδομένων καρκίνου του μαστού.....	23
3.2.1. Περιγραφική ανάλυση ποσοτικών μεταβλητών και ποιοτικών μεταβλητών καρκίνου του μαστού	25
3.3. Περιγραφή δεδομένων καρκίνου του παγκρέατος.....	31
3.3.1. Περιγραφική ανάλυση ποσοτικών μεταβλητών και ποιοτικών μεταβλητών καρκίνου του παγκρέατος	32
ΚΕΦΑΛΑΙΟ 4°	38
4. Έλεγχοι συσχετίσεων και ανεξαρτησίας των μεταβλητών	38
4.1. Δεδομένα Καρκίνου Εγκεφάλου	39
4.1.1. Ανάλυση βάσει της μεταβλητής «Survival Months»	39
4.1.2. Ανάλυση βάσει της μεταβλητής «Vital Status»	44
4.2. Δεδομένα Καρκίνου Μαστού.....	50
4.2.1. Ανάλυση βάσει της μεταβλητής «Survival Months»	50
4.2.2 Ανάλυση βάσει της μεταβλητής «Vital Status»	58
4.3 Δεδομένα Καρκίνου Παγκρέατος	71
4.3.1 Ανάλυση βάσει της μεταβλητής «Survival Months»	71
4.3.2 Ανάλυση βάσει της μεταβλητής «Vital Status»	75
ΚΕΦΑΛΑΙΟ 5°	83
5. Παραμετρικά Μοντέλα επιβίωσης.....	83

5.1. Εισαγωγικά	83
5.2. Μοντέλο επιβίωσης καρκίνος του Εγκεφάλου	85
5.2.1. Το μοντέλου επιταχυνόμενου κινδύνου του καρκίνου του Εγκεφάλου.....	85
5.2.2. Το μοντέλου αναλογικού κινδύνου του καρκίνου του Εγκεφάλου.....	89
5.3. Μοντέλο επιβίωσης καρκίνος του Μαστού	92
5.3.1. Το μοντέλου επιταχυνόμενου κινδύνου του καρκίνου του Μαστού.....	92
5.3.2. Το μοντέλου αναλογικού κινδύνου του καρκίνου του Μαστού.	95
5.4. Μοντέλο επιβίωσης καρκίνος του Παγκρέατος.....	99
5.4.1. Το μοντέλου επιταχυνόμενου κινδύνου του καρκίνου του Παγκρέατος	99
5.4.2. Το μοντέλου αναλογικού κινδύνου του καρκίνου του Παγκρέατος	102
ΚΕΦΑΛΑΙΟ 6°	106
6. Συμπεράσματα	106
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	108

ΚΕΦΑΛΑΙΟ 1^ο

1. Εισαγωγή

Με τον όρο *καρκίνος* περιγράφεται μια ομάδα ασθενειών που χαρακτηρίζεται από συσσώρευση γενετικών μεταλλαγών που οδηγούν σε απώλεια ελέγχου και ανώμαλη κυτταρική ανάπτυξη με αποτέλεσμα το σχηματισμό κακοήθων όγκων. Σε αντίθεση με τους καλοήθεις όγκους, οι οποίοι είναι συνήθεις και, ως επί το πλείστο, μη επικίνδυνος για την ανθρώπινη ζωή, οι κακοήθεις όγκοι έχουν τη δυνατότητα εισβολής ή εξάπλωσης σε διάφορες περιοχές του σώματος. Τα συμπτώματα ποικίλουν ανάλογα τον ιστό/όργανο που εμφανίζεται ο καρκίνος: ένα εξόγκωμα, μια ασυνήθιστη αιμορραγία, ένας επίμονος βήχας, μια ανεξήγητη απώλεια βάρους και μια αλλαγή στις συνήθειες του εντέρου είναι όλα πιθανά προειδοποιητικά σημάδια και συμπτώματα. Οι κύριες μέθοδοι θεραπείας περιλαμβάνουν τη χειρουργική επέμβαση, την ακτινοθεραπεία, τη χημειοθεραπεία, την ανοσοθεραπεία και τη στοχευμένη θεραπεία. Η επιλογή της θεραπείας εξαρτάται από τον τύπο και το στάδιο του καρκίνου, καθώς και από τη γενική υγεία του ασθενούς.

Ενδείξεις για την ύπαρξη της ασθένειας του καρκίνου, τόσο στον άνθρωπο όσο και στα υπόλοιπα ζώα υπήρχαν ήδη από την αρχαιότητα. Συνεπώς, δεν αποτελεί έκπληξη το γεγονός ότι βιβλιογραφία για τον καρκίνο εμφανίζεται από τότε. Αρχαία χειρόγραφα, αιγυπτιακές ανθρώπινες μούμιες και απολιθωμένοι όγκοι οστών περιέχουν μερικές από τις πρώτες αποδείξεις για τον καρκίνο. Αν και η λέξη "καρκίνος" δεν χρησιμοποιούνταν, η πρώτη περιγραφή της νόσου εντοπίστηκε στην Αίγυπτο και χρονολογείται περίπου στο 3000 π.Χ. Ο "Πάπυρος Edwin Smith" αποτελεί αρχαίο αιγυπτιακό εγχειρίδιο χειρουργικής, όπου περιγράφονται λεπτομερώς περιπτώσεις όγκων του μαστού που αφαιρέθηκαν με τη χρήση καυτηριασμού, αν και η θεραπεία θεωρούνταν ανέφικτη. Ο όρος "καρκίνος" αποδίδεται στον Ιπποκράτη (460-370 π.Χ.), που θεωρείται ο "πατέρας της Ιατρικής", ο οποίος χρησιμοποίησε τον όρο για να περιγράψει όγκους, πιθανόν λόγω της ομοιότητας με το σχήμα του καβουριού (οι δακτυλοειδείς προεξοχές που απλώνονταν από έναν καρκίνο θύμιζαν στους ανθρώπους το κέλυφος των καβουριών). Αργότερα, ο ελληνικός όρος μεταφράστηκε στη λατινική λέξη για τον κάβουρα, από τον Ρωμαίο γιατρό Celsus (25 π.Χ. - 50 μ.Χ.). Ένας άλλος Έλληνας ιατρός, ο Γαληνός (130-200 μ.Χ.) εισήγαγε τον όρο «όγκος». Κατά την Αναγέννηση, η επιστημονική μέθοδος, που εφαρμόστηκε από επιστήμονες όπως ο Γαλιλαίος και ο Νεύτωνας, επέτρεψε την πρόοδο στη μελέτη του ανθρώπινου σώματος και των ασθενειών. Ο Harvey διαλεύκανε τη ροή του αίματος μέσω αυτοψιών, ενώ ο Giovanni Morgagni το 1761 σύνδεσε παθολογικά ευρήματα με ασθένειες, θεμελιώνοντας την επιστημονική ογκολογία. Ο John Hunter πρότεινε τη χειρουργική αντιμετώπιση ορισμένων καρκίνων, ενώ η εφεύρεση της αναισθησίας και η χρήση μικροσκοπίου τον 19ο αιώνα οδήγησαν σε σημαντικές προόδους. Ο Rudolf Virchow, πατέρας της κυτταρικής παθολογίας, συνέδεσε τη μικροσκοπική παθολογία με την ασθένεια, καθορίζοντας τα θεμέλια της σύγχρονης ογκολογίας.

Στη σύγχρονη εποχή οι ερευνητές έχουν εντοπίσει διάφορα στάδια καρκίνου, γεγονός που υποδηλώνει ότι στην παθογένεια του καρκίνου εμπλέκονται διάφορες γονιδιακές μεταλλάξεις και διεργασίες, αποτέλεσμα τόσο περιβαλλοντικών όσο και κληρονομικών παραγόντων. Ο καρκίνος κατατάσσεται στη δεύτερη θέση ως κύρια αιτία θνησιμότητας παγκοσμίως,

ακολουθώντας τα καρδιαγγειακά νοσήματα. Σύμφωνα με στοιχεία του Παγκόσμιου Οργανισμού Υγείας, ο καρκίνος είναι υπεύθυνος για περίπου 10 εκατομμύρια θανάτους ετησίως, που αντιπροσωπεύει περίπου το 1/6 του συνόλου των θανάτων παγκοσμίως. Οι πλέον συχνές μορφές καρκίνου που οδηγούν σε θάνατο περιλαμβάνουν τον καρκίνο του πνεύμονα, του παχέος εντέρου, του στομάχου, του ήπατος και του μαστού. Η πρόγνωση για τους ασθενείς με καρκίνο έχει βελτιωθεί σημαντικά τα τελευταία χρόνια λόγω της προόδου στη διάγνωση και τη θεραπεία, με σημαντική αύξηση στα ποσοστά επιβίωσης για ορισμένους τύπους καρκίνου. (Οικονόμου Β., 2023).

1.1. Καρκίνος του Εγκεφάλου

Ο εγκέφαλος είναι το πιο πολύπλοκο τμήμα του ανθρώπινου σώματος. Αν και ο εγκέφαλος αποτελείται από πολλούς διαφορετικούς τύπους κυττάρων, ο κυριότερος είναι οι νευρώνες, οι οποίοι είναι νευρικά κύτταρα, και τα νευρογλία, τα οποία είναι μη νευρικά κύτταρα με ποικίλα σχήματα. Η νευρογλία υποστηρίζει και θωρακίζει τους νευρώνες του εγκεφάλου βοηθώντας στη διατήρηση της ομοιόστασης. Τα αστροκύτταρα, τα ολιγοδενδροκύτταρα, τα επενδυτικά κύτταρα και η μικρογλία είναι μεταξύ των νευρογλοιακών κυττάρων που βρίσκονται στον ανθρώπινο εγκέφαλο, ο οποίος έχει ένα νευρογλοιακό κύτταρο για κάθε νευρώνα και τρία για κάθε δύο στην εγκεφαλική φαιά ουσία (η πλειονότητα των όγκων του εγκεφάλου είναι νευρογλοιακής προέλευσης).

Όταν σχηματίζονται ανώμαλα εγκεφαλικά κύτταρα, δημιουργείται όγκος στον εγκέφαλο και ανάλογα την σημαντικότητα αυτοί μπορούν επίσης να διαχωριστούν σε πρωτοπαθείς όγκους, οι οποίοι προέρχονται από το εσωτερικό του εγκεφάλου, και σε δευτεροπαθείς όγκους, οι οποίοι συνήθως εξαπλώνονται από όγκους εκτός του εγκεφάλου και αναφέρονται ως όγκοι *εγκεφαλικής μετάστασης*. Ένας όγκος εγκεφάλου, όπως συμβαίνει με κάθε όγκο στο ανθρώπινο σώμα, λέγεται καλοήθης όταν αναπτύσσεται αργά και χωρίς να εισχωρεί στον εγκέφαλο. Παράδειγμα το μηνιγγίωμα. Ένας όγκος εγκεφάλου λέγεται κακοήθης όταν αναπτύσσεται με γρήγορο ρυθμό και εισχωρεί στον εγκέφαλο που βρίσκεται γύρω του. Παράδειγμα το πολύμορφο γλοιοβλάστωμα. Συνέπεια αυτής ανάπτυξης, είναι η διασπορά καρκινικών κυττάρων σε περιοχές απομακρυσμένες του εγκεφάλου. Ανάλογα με το μέγεθος του όγκου και την περιοχή του εγκεφάλου που έχει προσβληθεί, τα συμπτώματα από τους τύπους όγκων του εγκεφάλου μπορεί να διαφέρουν. Μερικά από τα συμπτώματα μπορεί να είναι εμετός, πονοκέφαλος, επιληπτικές κρίσεις, προβλήματα όρασης και νοητικές αλλαγές. Άλλα συμπτώματα μπορεί να περιλαμβάνουν αδυναμία ομιλίας ή βάδισης, αισθήσεις ή απώλεια αισθήσεων. Για παράδειγμα, η ικανότητα σκέψης μπορεί να αλλάξει ως αποτέλεσμα μεγαλύτερων όγκων στον μετωπιαίο λοβό. Αλλά ένας μικρότερος όγκος στην περιοχή του Wernicke, μια μικροσκοπική περιοχή που εμπλέκεται στην κατανόηση της γλώσσας, μπορεί να προκαλέσει μεγαλύτερη απώλεια λειτουργίας.

Η θεραπεία εξαρτάται από πολλούς παράγοντες, όπως το είδος, το μέγεθος, η τοποθεσία του όγκου, καθώς και η γενική υγεία του ασθενούς. Συνήθως περιλαμβάνει συνδυασμό χειρουργικής επέμβασης, ακτινοθεραπείας και χημειοθεραπείας. Οι προοπτικές για τον καρκίνο του εγκεφάλου ποικίλλουν ευρέως, με κάποιους όγκους να είναι πιο επιθετικοί και να έχουν δυσμενή πρόγνωση, ενώ άλλοι μπορεί να αντιμετωπίζονται πιο αποτελεσματικά. Η

έρευνα συνεχίζεται για την ανάπτυξη νέων και πιο αποτελεσματικών θεραπειών, με στόχο τη βελτίωση της επιβίωσης και της ποιότητας ζωής των ασθενών.

1.2. Καρκίνος του Μαστού

Η εξωτερική μορφολογία του μαστού περιλαμβάνει τη θηλή και τη θηλαία άλω. Εσωτερικά ο μαστός αποτελείται από τον μαστικό ή μαζικό αδένα και το περιμαστικό λίπος. Ο πρώτος ανατομικά περιέχει 15 έως 20 λοβούς αδενικού ιστού, διατεταγμένους όπως τα πέταλα μιας μαργαρίτας. Οι λοβοί χωρίζονται περαιτέρω σε μικρότερους λοβούς που παράγουν γάλα για το θηλασμό και το περιμαστικό λίπος είναι η επέκταση του υποδόριου λίπους, δηλαδή του μέρους της πιο εσωτερικής στοιβάδας του δέρματος.

Ο καρκίνος του μαστού αφορά την ανεξέλεγκτη ανάπτυξη κυττάρων στον ιστό του μαστού και μπορεί να εμφανιστεί σε διάφορες μορφές, με τις πιο κοινές να είναι το διηθητικό καρκίνωμα και το μη διηθητικό καρκίνωμα (καρκίνωμα *in situ*). Μπορεί να ξεκινήσει από τους λοβούς (λοβιακό καρκίνωμα) ή τους γαλακτοφόρους πόρους (πορογενές καρκίνωμα). Τα γονίδια αυτών των κυττάρων υφίστανται μεταλλάξεις, οι οποίες επιταχύνουν τον ρυθμό πολλαπλασιασμού τους με αποτέλεσμα την εμφάνιση του καρκίνου με την πάροδο του χρόνου και να τροφοδοτούν τον όγκο αναπτύσσοντας νέα αιμοφόρα αγγεία. Ο όγκος επεκτείνεται και τα καρκινικά κύτταρα που αποτελούν τον όγκο μπορεί να αναπτύξουν την ικανότητα να εγκαταλείψουν την αρχική περιοχή και να εξαπλωθούν στους λεμφαδένες ή σε άλλα όργανα του σώματος (μεταστάσεις). Οι παράγοντες κινδύνου περιλαμβάνουν τη γενετική προδιάθεση (όπως μεταλλάξεις στα γονίδια BRCA1 και BRCA2), την ηλικία, το οικογενειακό ιστορικό, την πρόωμη έναρξη έμμηνου ρύσεως και την καθυστερημένη εμμηνόπαυση. Η διατροφή, η σωματική δραστηριότητα και η κατανάλωση αλκοόλ επίσης σχετίζονται με τον κίνδυνο ανάπτυξης καρκίνου του μαστού.

Ο καρκίνος του μαστού είναι ένας από τους πιο συχνά διαγνωσμένους καρκίνους παγκοσμίως και η κύρια αιτία θανάτου από καρκίνο στις γυναίκες. Κατατάσσεται στην πρώτη θέση παγκοσμίως από τις πιθανές μορφές καρκίνου που εμφανίζουν οι γυναίκες καθώς προσεγγιστικά 1 στις 8 γυναίκες θα παρουσιάσει καρκίνο του μαστού σε κάποια φάση της ζωής της. Ωστόσο, χάρη στην έγκαιρη ανίχνευση μέσω της μαστογραφίας και την πρόοδο στη θεραπεία, τα ποσοστά θανάτου από καρκίνο του μαστού μειώνονται. Στο αρχικό στάδιο του καρκίνου του μαστού τα συμπτώματα δεν είναι απολύτως εμφανή. Αργότερα, τα κυριότερα σημάδια που θα παρουσιαστούν είναι το εξόγκωμα ή σκλήρυνση στην περιοχή του μαστού ή και όπως μασχάλης, διόγκωση λεμφαδένων, έκκριση υγρού ή αίματος από την θηλή και διάφορα ακόμη κατά τα οποία εάν η γυναίκα δεν δώσει την απαραίτητα προσοχή θα εμφανιστούν σημάδια προχωρημένης νόσου, όπως ερυθρό μαστό και πόνος στα οστά. Η θεραπεία του καρκίνου του μαστού περιλαμβάνει τη χειρουργική αφαίρεση, την ακτινοθεραπεία και τη χημειοθεραπεία καθώς και το συνδυασμό των παραπάνω θεραπειών.

1.3. Καρκίνος του Παγκρέατος

Το πάγκρεας είναι από τα σημαντικότερα όργανα του ανθρώπινου σώματος καθώς παράγει ορμόνες όπως την ινσουλίνη και το γλυκαγόνο οι οποίες ρυθμίζουν το σάκχαρο και επιπροσθέτως διέρχεται από αυτό ο παγκρεατικός πόρος μέσω του οποίου το παγκρεατικό υγρό

εκκρίνεται στο έντερο, διεργασία απαραίτητη για την απορρόφηση και πέψη των θρεπτικών συστατικών από το λεπτό έντερο. Η κεφαλή, ο αυχένας, το σώμα και η ουρά είναι τα τέσσερα τμήματα από τα οποία απαρτίζεται το πάγκρεας και τοπογραφικά βρίσκεται πίσω από το στομάχι μεταξύ του δωδεκαδάκτυλου και της σπλήνας.

Ο καρκίνος του παγκρέατος είναι μια από τις πιο επιθετικές μορφές καρκίνου και χαρακτηρίζεται από υψηλή θνησιμότητα. Προσβάλλει συνήθως άτομα μετά την ηλικία των 65 ετών δίχως όμως να γνωρίζουμε επακριβώς τον κύριο παράγοντα της εμφάνισής του. Είναι η πέμπτη αιτία θανάτου από καρκίνο. Πάνω από το 80% των καρκίνων του παγκρέατος εμφανίζεται στην εξωκρινή μοίρα, δηλαδή στους αδένες που παράγουν τα διάφορα υγρά του παγκρέατος (το αδενοκαρκίνωμα είναι το πιο συχνό). Το 75% αυτών ανευρίσκεται στην κεφαλή και στον αυχένα του οργάνου. Κάποιοι κοινοί παράγοντες που συνδέονται με την ανάπτυξη καρκίνου έχουν θεωρηθεί τα γονίδια, το κάπνισμα, η παχυσαρκία και η χρόνια παγκρεατίτιδα καθώς και ο διαβήτης είναι πιθανό να αποτελεί πρόιμη εμφάνιση του παγκρεατικού καρκίνου. Η απώλεια βάρους, ο ίκτερος (στην κεφαλή του παγκρέατος), η στεατόρροια, ο κοιλιακός ή οσφυϊκός πόνος (στο σώμα ή στην ουρά του παγκρέατος), η δυσπεψία, η ναυτία και ενδεχομένως η κατάθλιψη είναι μερικά μόνο από τα συμπτώματα που μπορεί να προκαλέσει ο καρκίνος ανάλογα με το πού βρίσκεται στο σώμα, στην κεφαλή ή στα ούρα. Το πρώτο κρίσιμο βήμα για την ακριβή διάγνωση και θεραπεία του καρκίνου του παγκρέατος και τη μείωση της θνησιμότητάς του είναι η πρόληψη στην ομάδα αυξημένου κινδύνου (αιφνίδια έναρξη σακχαρώδη διαβήτη και ανεξήγητη οξεία παγκρεατίτιδα). Η θεραπεία του καρκίνου του παγκρέατος εξαρτάται από το στάδιο της νόσου και περιλαμβάνει συνήθως χειρουργική επέμβαση (όταν είναι εφικτό), χημειοθεραπεία και ακτινοθεραπεία. Παρά τη θεραπεία, η πρόγνωση παραμένει φτωχή, με το πενταετές ποσοστό επιβίωσης να είναι χαμηλότερο από 10%. Λόγω της καθυστερημένης διάγνωσης, της ανάγκης για εξειδικευμένη χειρουργική θεραπεία, του υψηλού ποσοστού υποτροπών και της μικρής συνολικής επιβίωσης, η διαχείριση του καρκίνου του παγκρέατος αποτελεί πρόκληση. (Τζαβελάς Γ., 2012)

1.4. Στατιστική και Καρκίνος

Η ανάπτυξη της στατιστικής έχει διαδραματίσει καθοριστικό ρόλο στην πρόληψη και θεραπεία του καρκίνου, συμβάλλοντας σε πολλαπλά επίπεδα:

1. **Επιδημιολογικές Μελέτες:** Μέσω στατιστικών αναλύσεων, είναι εφικτός ο εντοπισμός παραγόντων κινδύνου που συνδέονται με την εμφάνιση καρκίνου. Οι επιδημιολογικές μελέτες, βασισμένες σε στατιστικά δεδομένα, επιτρέπουν την αναγνώριση συμπεριφορών και περιβαλλοντικών παραγόντων που αυξάνουν την πιθανότητα ανάπτυξης της νόσου.
2. **Κλινικές Δοκιμές:** Η στατιστική είναι απαραίτητη για τον σχεδιασμό, την ανάλυση και την ερμηνεία των αποτελεσμάτων κλινικών δοκιμών, οι οποίες αξιολογούν την αποτελεσματικότητα νέων θεραπειών για τον καρκίνο. Αυτό διασφαλίζει ότι οι νέες θεραπείες υποβάλλονται σε αυστηρό έλεγχο πριν εφαρμοστούν κλινικά.
3. **Προγνωστικά Μοντέλα:** Η στατιστική επιτρέπει την ανάπτυξη προγνωστικών μοντέλων που προβλέπουν την πορεία της ασθένειας και την ανταπόκριση σε

θεραπείες. Τα μοντέλα αυτά χρησιμοποιούνται για την εξατομίκευση της θεραπείας, με στόχο την παροχή βέλτιστων θεραπειών σε κάθε ασθενή.

4. **Γενετικές Μελέτες:** Οι στατιστικές μέθοδοι είναι κρίσιμες για την ανάλυση γενετικών δεδομένων, βοηθώντας στον εντοπισμό γονιδίων που σχετίζονται με την ανάπτυξη καρκίνου. Αυτό ανοίγει τον δρόμο για την εξατομίκευση της θεραπείας με βάση τις γενετικές ιδιαιτερότητες του ασθενούς.
5. **Αξιολόγηση Προγραμμάτων Πρόληψης:** Η στατιστική χρησιμοποιείται για την παρακολούθηση και αξιολόγηση της αποτελεσματικότητας προγραμμάτων πρόληψης και έγκαιρης διάγνωσης, όπως ο έλεγχος για καρκίνο του μαστού ή του προστάτη, συμβάλλοντας έτσι στη βελτίωση των αποτελεσμάτων δημόσιας υγείας.

Η ενσωμάτωση της στατιστικής στην ογκολογική έρευνα και κλινική πράξη έχει οδηγήσει σε σημαντικές προόδους στη διάγνωση, πρόληψη, εξατομίκευση της θεραπείας και συνολική διαχείριση του καρκίνου. (Σαχλάς Α., 2016)

ΚΕΦΑΛΑΙΟ 2^ο

2. Μεθοδολογία

2.1. Μοντέλο αναλογικού κινδύνου

Σε τέτοιου είδους μελέτες υπάρχει η ανάγκη να βρούμε την σχέση μεταξύ της μεταβλητής που δηλώνει τον χρόνο ζωής και μιας ή περισσότερων συμμεταβλητών το οποίο το καταφέρνουμε συνήθως μέσω ενός μοντέλου παλινδρόμησης. Οι λόγοι δημιουργίας ενός τέτοιου μοντέλου είναι ο προσδιορισμός των μεταβλητών που επηρεάζουν περισσότερο την συνάρτηση κινδύνου $h(t)$, η εκτίμηση της ίδιας και κατά συνέπεια της συνάρτησης επιβίωσης. Στην ανάλυση επιβίωσης, η αναλογική παλινδρόμηση (PH) ή μοντέλο αναλογικού κινδύνου (Proportional hazard model) είναι το βασικότερο μοντέλο για τη μοντελοποίηση της συσχέτισης μεταξύ του χρόνου μέχρι το συμβάν με τις συνδιακυμάνσεις ή τις ανεξάρτητες μεταβλητές το οποίο αποτελεί μια μορφή γενικευμένων γραμμικών μοντέλων. Χρησιμοποιείται εκτενώς σήμερα στην ανάλυση λογοκριμένων δεδομένων επιβίωσης που αφορούν βιολογικές εφαρμογές, για την εξακρίβωση των διαφορών στην επιβίωση που οφείλονται στο είδος της θεραπείας και σε προγνωστικούς παράγοντες σε κλινικές δοκιμές.

Συνήθως, στη πράξη εκτός από τη μεταβλητή T που συμβολίζει τον χρόνο επιβίωσης και τη μεταβλητή Δ που συμβολίζει εάν ο χρόνος ζωής είναι πλήρης ή λογοκριμένος, έχουμε και άλλες μεταβλητές $Z_j, j=1,2,\dots,p$, οι οποίες μπορεί να επηρεάζουν τον χρόνο επιβίωσης των ατόμων. Ας υποθέσουμε ότι κάθε άτομο ανήκει σε μια από δυο κατηγορίες.

Για παράδειγμα, μπορεί να έχουμε άνδρες και γυναίκες, καπνιστές και μη καπνιστές, κλπ. Στο μοντέλο αναλογικού κινδύνου δεχόμαστε ότι ισχύει η σχέση:

$$h_1(t) = c h_2(t) \quad \text{ή} \quad \frac{h_1(t)}{h_2(t)} = c, \quad \text{για κάθε } t \geq 0$$

όπου $h_i(t), i=1,2$ είναι η συνάρτηση κινδύνου της i ομάδας και c είναι ένας θετικός αριθμός που δεν εξαρτάται από το χρόνο t . Το μοντέλο αναλογικού κινδύνου αναφέρει ότι η συνάρτηση κινδύνου της ομάδας 1 είναι c -πλάσια της συνάρτησης κινδύνου της ομάδας 2. Εάν συμβολίζουμε με β τον λογάριθμο του c $\log(c)$ τότε η παραπάνω σχέση παίρνει τη μορφή:

$$h_1(t) = e^\beta h_2(t), \quad \text{για κάθε } t \geq 0$$

Το μοντέλο αναλογικού κινδύνου στην περίπτωση που υπάρχει μια μόνο ανεξάρτητη μεταβλητή δίνεται από τη σχέση:

$$h(t|Z) = h_0(t)e^{\beta Z}, \quad \text{για κάθε } t \geq 0,$$

όπου η $h_0(t)$ ονομάζεται συνάρτηση αναφοράς κινδύνου. Η αναφορική συνάρτηση κινδύνου είναι η συνάρτηση ενός ατόμου όταν η ανεξάρτητη μεταβλητή ισούται με 0. Παρατηρούμε ότι η αναφορική συνάρτηση κινδύνου εξαρτάται μόνο από τον χρόνο και όχι από την ανεξάρτητη μεταβλητή. Αντίθετα, ο όρος $e^{\beta Z}$ εξαρτάται μόνο από την ανεξάρτητη μεταβλητή και όχι από τον χρόνο. Όταν η μεταβλητή Z παίρνει την τιμή 1, τότε ισχύει:

$$h(t|Z = 1) = h_0(t)e^{\beta \cdot 1} = h_0(t)e^\beta, \quad \text{για κάθε } t \geq 0,$$

ενώ όταν υπάρχει η τιμή 0 ισχύει:

$$h_0(t|Z = 0) = h_0(t)e^{\beta \cdot 0} = h_0(t), \quad \text{για κάθε } t \geq 0.$$

Στην περίπτωση που υπάρχουν περισσότερες από μια ανεξάρτητες μεταβλητές, το μοντέλο αναλογικού κινδύνου γράφεται ως:

$$h(t|Z) = h_0(t)e^{\beta Z} = h_0(t)e^{\sum_{i=1}^p \beta_i z_i}, \quad \text{για κάθε } t \geq 0,$$

όπου $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ είναι το διάνυσμα των p παραμέτρων του μοντέλου και $Z = (Z_1, Z_2, \dots, Z_p)'$ είναι το διάνυσμα των p ανεξάρτητων μεταβλητών. Στην περίπτωση αυτή, η αναφορική συνάρτηση κινδύνου $h_0(t)$ είναι η συνάρτηση κινδύνου ενός ατόμου όταν όλες οι ανεξάρτητες μεταβλητές ισούται με το 0.

Ας θεωρήσουμε τώρα δυο άτομα με τιμές των συμμεταβλητών Z_1 και Z_2 , αντίστοιχα. Ο σχετικός λόγος κινδύνου ισούται με

$$HR = \frac{h(t|Z_1)}{h(t|Z_2)} = \frac{h_0(t)e^{\beta Z_1}}{h_0(t)e^{\beta Z_2}} = \exp\{\beta(Z_1 - Z_2)\},$$

ο οποίος είναι σταθερός και δεν εξαρτάται από τον χρόνο t . Στην πράξη, ο σχετικός λόγος κινδύνου είναι ο κίνδυνος ένα άτομο με χαρακτηριστικά Z_1 να εμφανίσει το υπό μελέτη γεγονός σε σχέση με ένα άτομο με χαρακτηριστικά Z_2 . Από την παραπάνω σχέση, διαπιστώνουμε ότι ο κίνδυνος του ατόμου με χαρακτηριστικά Z_1 ισούται με $\exp\{\beta(Z_1 - Z_2)\}$ φορές τον κίνδυνο ενός ατόμου με χαρακτηριστικά Z_2 . Από την παραπάνω έκφραση του σχετικού λόγου κινδύνου, προκύπτει ότι ο λογάριθμος του ισούται με

$$\log(HR) = \log\left(\frac{h(t|Z_1)}{h(t|Z_2)}\right) = \sum_{i=1}^p \beta_i (Z_{1i} - Z_{2i}).$$

Παρατηρούμε δηλαδή ότι ο λογάριθμος του σχετικού λόγου κινδύνου συνδέεται με τις διαφορές $Z_{1i} - Z_{2i}$, $i=1,2$, με μια γραμμική σχέση.

Το παραπάνω μοντέλο ονομάζεται μοντέλο αναλογικού κινδύνου. Πρόκειται για ένα ημι-παραμετρικό μοντέλο καθώς γίνονται κάποιες υποθέσεις για τον όρο $e^{\beta Z}$ αλλά δεν προκαθορίζεται κάποια μορφή για την αναφορική συνάρτηση κινδύνου $h_0(t)$.

Στην περίπτωση του μοντέλου αναλογικού κινδύνου, η εκτίμηση των παραμέτρων γίνεται με τη μέθοδο μέγιστης πιθανοφάνειας. Διακρίνονται δυο περιπτώσεις. Η πρώτη περίπτωση είναι η απουσία δεσμών (ties) και η δεύτερη είναι η παρουσία δεσμών. Απουσία δεσμών έχουμε όταν σε κάθε χρονική στιγμή συμβαίνει μόνο ένα γεγονός. Αντίθετα, παρουσία δεσμών έχουμε εάν σε κάποια χρονική στιγμή παρατηρούνται περισσότερα από ένα γεγονότα. (Αντζουλάκος Δ., 2023).

2.2. Παραμετρικά Μοντέλα επιβίωσης

Έστω ότι $T_1, \dots, T_n; C_1, \dots, C_n$ είναι ανεξάρτητες τυχαίες μεταβλητές. C_i είναι ο χρόνος λογοκρισίας ο οποίος σχετίζεται με τον χρόνο επιβίωσης T_i . Παρατηρούμε $(Y_1, \delta_1), \dots, (Y_n, \delta_n)$ όπου:

$$Y_i = T_i \wedge C_i \cdot \delta_i = I(T_i \leq C_i).$$

Είναι επίσης διαθέσιμες συμμεταβλητές $\underline{x}_1, \dots, \underline{x}_n$, όπου:

$$\underline{x}_i = (x_{i1}, \dots, x_{ip}), \quad i = 1, \dots, n$$

οι οποίες σχετίζονται με την εξαρτημένη μεταβλητή T_i .

Η συνάρτηση κινδύνου ορίζεται ως:

$$\lambda(t; \underline{x}) = \frac{f(t; \underline{x})}{1 - F(t; \underline{x})}.$$

Το αναλογικό μοντέλο συνάρτησης κινδύνου είναι:

$$\lambda(t; \underline{x}) = e^{\underline{\beta}'\underline{x}} \lambda_0(t)$$

όπου:

$$\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$$

είναι οι συντελεστές παλινδρόμησης.

Τα παραμετρικά μοντέλα επιβίωσης (Π.Μ.Ε.) έχουν τα ίδια χαρακτηριστικά με το αναλογικό μοντέλο επιβίωσης με την διαφορά ότι η συνάρτηση κινδύνου βάσης εκτιμάται μαζί με τυχόν συντελεστές. Στην παρούσα διπλωματική θα ασχοληθούμε με δυο τύπους παραμετρικών μοντέλων.

- Το παραμετρικό, αναλογικό μοντέλο επιβίωσης
- Το μοντέλο της επιταχυνόμενης (accelerated) συνάρτησης κινδύνου.

Το παραμετρικό μοντέλο επιβίωσης

Το μοντέλο αναλογικού κινδύνου (PH) περιγράφεται από τη σχέση

$$h(t; x_1, \dots, x_p) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p) \quad (2.2.1)$$

για το οποίο δεν γίνεται καμιά αναφορά για την κατανομή της τ.μ. T .

Ο λόγος

$$HR = \frac{h(t; x_{i1}, \dots, x_{ip})}{h(t; x_{j1}, \dots, x_{jp})}$$

για δυο μονάδες i και j λέγεται λόγος κινδύνου και είναι ανεξάρτητο το χρόνου.

Δηλαδή αν το μοντέλο προσαρμόζεται καλά, θα πρέπει οι συναρτήσεις κινδύνου για δυο μονάδες να είναι περίπου “παράλληλες”.

Ως εδώ δεν έχει γίνει κάποια υπόθεση για την κατανομή του μοντέλου. Αν θα θέλαμε να υιοθετήσουμε κάποια. θα έπρεπε να λάβουμε υπόψη την ιδιότητα (2.2.1) η οποία λέει ότι οι δυο συναρτήσεις κινδύνου h και h_0 είναι ανάλογες. Έχουμε λοιπόν τον παρακάτω ορισμό.

Ορισμός 1: Αν δυο συναρτήσεις κινδύνου h_0 και h_1 έχουν την ιδιότητα ότι

$$h_1(t) = \psi h_0(t) \text{ για όλα τα } t > 0 \text{ και κάποιο } \psi > 0, \quad (2.2.2)$$

τότε λέμε ότι είναι ανάλογες.

Η παραπάνω σχέση μπορεί να γραφεί σε σχέση με τις συναρτήσεις επιβίωσης σαν

$$S_1(t) = \{S_0(t)\}^\psi \text{ για όλα τα } t > 0 \text{ και κάποιο } \psi > 0 \quad (2.2.3)$$

Οι κατανομές αυτές λέγονται Lehmann alternatives. Προφανώς στην εφαρμογή του μοντέλου θέτουμε

$$\psi = \exp(\beta_1 x + \dots + \beta_p x_p).$$

Επομένως κατάλληλες κατανομές είναι αυτές με την ιδιότητα (2.2.2) ή (2.2.3). Το PH μοντέλο εστιάζει στον ρυθμό κινδύνου και είναι κατάλληλο όταν υποθέτουμε ότι οι παράγοντες επηρεάζουν τον κίνδυνο συμβάντος τη συγκεκριμένη χρονική στιγμή.

Κατανομές οι οποίες έχουν τις παραπάνω ιδιότητες είναι η Weibull, η κατά τμήματα σταθερή συνάρτηση κινδύνου (pch) και η κατανομή Gompertz.

The accelerated failure time model (AFT)

Το AFT μοντέλο καλύτερα περιγράφεται με βάση τη σχέση δυο συναρτήσεων επιβίωσης όταν συγκρίνουμε δυο ομάδες.

- Ομάδα 0 : $P(T \geq t) = S_0(t)$ (ομάδα ελέγχου)
- Ομάδα 1: $P(T \geq t) = S(t) = S_0(\gamma t)$ (πειραματική ομάδα) (2.2.4)

Το μοντέλο λέει ότι η ασθένεια επιταχύνει τον χρόνο θανάτου με έναν παράγοντα γ . Αν $\gamma < 1$, τότε η θεραπεία είναι αποτελεσματική. στην αντίθετη περίπτωση είναι αναποτελεσματική. Μια άλλη ερμηνεία είναι ότι η διάμεση διάρκεια ζωής πολλαπλασιάζεται με έναν παράγοντα $1/\gamma$. Η (2.2.4) είναι αυτή που ορίζει το (AFT) μοντέλο. (Τζαβελάς Γ., 2012)

Ορισμός 2: Ένα μοντέλο επιβίωσης λέγεται επιταχυνόμενο μοντέλο κινδύνου (AFT) όταν υπάρχει παράμετρος γ έτσι ώστε η συνάρτηση κινδύνου να γράφεται:

$$h(t; \gamma) = \gamma h(\gamma t; 1) = \gamma h_0(\gamma t) \quad (2.2.5)$$

όπου γ είναι ο παράγοντας επιτάχυνσης. Τα επόμενα είναι ισοδύναμα:

$$h(t; \gamma) = \gamma h(\gamma t; 1) = \gamma h_0(\gamma t)$$

$$S(t; \gamma) = S(\gamma t; 1) = S_0(\gamma t)$$

$$f(t; \gamma) = \gamma f(\gamma t; 1) = \gamma f_0(\gamma t)$$

Παρατηρούμε ότι αν η f έχει παράμετρο κλίμακας λ , δηλαδή

$$f(t; \lambda) = \frac{1}{\lambda} f\left(\frac{1}{\lambda} t; 1\right) \text{ και ισοδύναμα } S(t; \lambda) = S\left(\frac{1}{\lambda} t; 1\right).$$

τότε είναι ένα AFT μοντέλο με συντελεστή επιτάχυνσης $\gamma = \frac{1}{\lambda}$.

Το AFT μοντέλο εστιάζει στην επίδραση των εξηγητικών μεταβλητών στον χρόνο επιβίωσης και είναι κατάλληλο όταν υποθέτουμε ότι οι παράγοντες επιταχύνουν ή επιβραδύνουν τον χρόνο επιβίωσης.

Κατανομές που μπορεί να χρησιμεύσουν σαν AFT μοντέλο είναι η Weibull ή λογαριθμοκανονική (log-normal) , η λογαριθμολογιστική (log-logistic).

Η Weibull είναι η μόνη κατανομή η οποία ικανοποιεί τα κριτήρια και για PH και για AFT μοντέλο.

Για να δημιουργήσουμε το AFT μοντέλο, ξεκινάμε από το γεγονός ότι υπάρχει παράμετρος κλίμακας. Έτσι αν η τυχαία μεταβλητή έχει T συνάρτηση επιβίωσης $S(t)$ τότε η

$$T_\gamma = \frac{T}{\gamma} \quad (2.2.6)$$

έχει συνάρτηση επιβίωσης $S(\gamma t)$. Αν $Y = \log T$ και $Y_\gamma = \log(T_\gamma)$ τότε ισχύει η σχέση (2.2.6) η οποία γράφεται σαν:

$$Y_\gamma = Y - \log \gamma.$$

Αν $Y = \varepsilon$.

$$Y_\gamma = Y$$

και

$$\log(\gamma) = -\underline{\beta x} = -(\beta_1 x_1 + \dots + \beta_p x_p) \quad (2.2.7)$$

Η παραπάνω σχέση μπορεί να γραφεί στην γνωστή μορφή

$$Y = \underline{\beta x} + \varepsilon.$$

Δηλαδή έχουμε την περίπτωση ενός γραμμικού μοντέλου για τους λογαριθμικούς χρόνους επιβίωσης. Από την (4.6) έχουμε ότι ο παράγοντας επιτάχυνσης είναι

$$\gamma = \exp\{-(\beta_1 x_1 + \dots + \beta_p x_p)\}.$$

Ο λόγος

$$\frac{Y_i}{Y_j}$$

ερμηνεύεται σαν ο παράγοντας επιτάχυνσης της μονάδας i σε σχέση με την μονάδα j . Για δυο μονάδες οι οποίες διαφέρουν μόνο σε μια μεταβλητή $x_{ik} \neq x_{jk}$ τότε

$$\frac{Y_i}{Y_j} = e^{-\beta_\kappa(x_{ik} - x_{jk})}$$

από το οποίο συμπεραίνουμε ότι η τιμή $e^{-\beta_\kappa}$ ερμηνεύεται σαν ο συντελεστής επιτάχυνσης μιας μονάδας όταν αυξάνεται η τιμή ίδιος x_κ κατά μια μονάδα. Όταν ίδιος οι ίδιος μεταβλητές παραμένουν ίδιος.

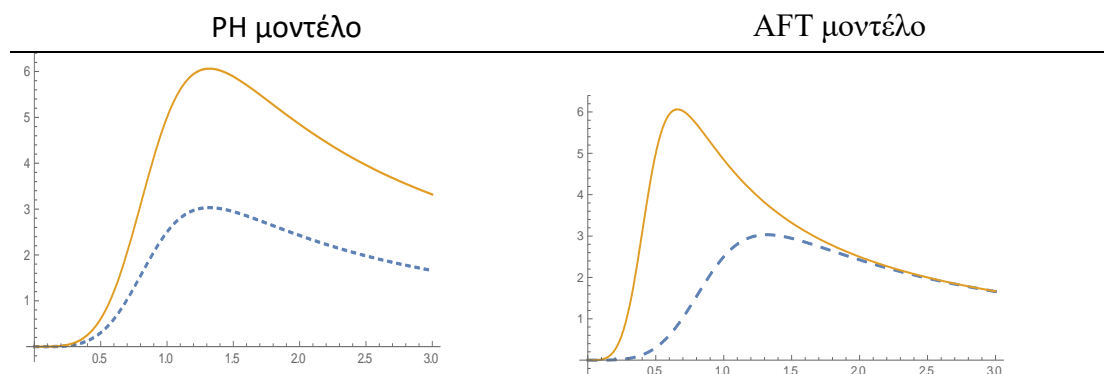
Η επιλογή μεταξύ των δυο αυτών μοντέλων γίνεται με την σύγκριση του AIC κριτηρίου. Επειδή ο αριθμός των παραμέτρων είναι ο ίδιος μπορούμε να συγκρίνουμε τις loglikelihood μέγιστες τιμές.

Για την καλύτερη κατανόηση των δυο εννοιών. Κατασκευάσαμε με κατανομή βάσης την loglogistic κατανομή με π.π.

$$f(x; \alpha, \beta) = \frac{\frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^\beta\right)^2}$$

Και συνάρτηση κινδύνου

$$h(t; \alpha, \beta) = \frac{\frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}}{1 + \left(\frac{t}{\alpha}\right)^\beta}$$



Σχήμα 2.2.1: Το αριστερό σχήμα το PH μοντέλο. Το δεξί σχήμα το AFT μοντέλο. Η διακεκομμένη γραμμή είναι η h_0 ενώ πλήρης γραμμή η $h(t)$.

Παρατηρούμε ότι στο PH μοντέλο, η συνάρτηση κινδύνου πολλαπλασιάζεται με 2 ενώ στο AFT μοντέλο γίνεται και μια μετατόπιση προς τα αριστερά. (Οικονόμου Π., 2010).

ΚΕΦΑΛΑΙΟ 3^ο

3. Περιγραφική Στατιστική

Στο παρόν κεφάλαιο της εργασίας παρουσιάζονται αναλυτικότερα κάποια περιγραφικά μέτρα. Πιο συγκεκριμένα για τις ποιοτικές μεταβλητές χρησιμοποιήθηκαν τα pie charts ή αλλιώς γνωστά και ως γραφήματα πίτας, τα οποία είναι κυκλικά διαγράμματα τα οποία διαμελίζονται σε κομμάτια για να διακρίνουμε τα ποσοστά που κατέχει το κάθε κομμάτι στις κατηγορίες που επιθυμούμε. Αντίστοιχα, για τις ποσοτικές μεταβλητές αναφέρουμε τα περιγραφικά μέτρα, μέση τιμή και τυπική απόκλιση.

3.1. Περιγραφή δεδομένων καρκίνου του εγκεφάλου

Για την συγκεκριμένη ανάλυση χρησιμοποιήθηκε δείγμα από το σύνολο δεδομένων SEERSTAT, το οποίο περιέχει 59.812 ασθενείς (33.623 άντρες και 26.189 γυναίκες). Οι μεταβλητές που χρησιμοποιούνται παρουσιάζονται στον παρακάτω πίνακα.

Μεταβλητή	Επεξήγηση
Patient ID	Κωδικός ασθενή
Age record	Ηλικιακό εύρος
Year of diagnosis	Χρονολογία διάγνωσης
Race record	Φυλή ασθενή
Site record	Είδος καρκίνου
Grade	Βαθμός καρκίνου
Diagnostic Confirmation	Μέθοδος διαγνωστικής επιβεβαίωσης
ICD-O-3	Κωδικοί ταξινόμησης όγκου
CS tumor size	Μέγεθος όγκου καρκίνου
Survival months	Μήνες επιβίωσης
Sequence number	Σειριακός αριθμός καρκίνου
Grade clinical	Κλινικός βαθμός καρκίνου
Grade pathological	Παθολογικός βαθμός καρκίνου
Total number of in situ/malignant tumors for patient	Συνολικός αριθμός διαγνωσμένων όγκων
Median household income	Μέσο οικιακό εισόδημα
Laterality	Πλευρικότητα όγκου
Vital status recode	Εν ζωή ή όχι (0=dead, 1=alive)

Πίνακα 3.1: Μεταβλητές καρκίνου εγκεφάλου

Για την καλύτερη κατανόηση της ανάλυσης του επιμέρους κεφαλαίου θα προστεθεί η παρακάτω κωδικοποίηση αναφορικά με το πεδίο «Clinical/ Pathological Grade». Πιο συγκεκριμένα:

- **Grade I** : Τέτοιοι όγκοι είναι συνήθως καλοήθεις, μοιάζουν σχεδόν με φυσιολογικά κύτταρα και επεκτείνονται αργά.

- **Grade II** : Είναι κύτταρα ενδιάμεσου βαθμού επιθετικότητας και φαίνονται λιγότερα φυσιολογικά από εκείνα με Grade I. Αναπτύσσονται και διαιρούνται λίγο πιο γρήγορα από τα κανονικά κύτταρα.

- **Grade III** : Τα κύτταρα αυτά είναι πιο επιθετικά, πολύ διαφορετικά σε σύγκριση με τα φυσιολογικά καθώς και η αφαίρεσή τους είναι αρκετά δύσκολη λόγω πιθανής μετάδοσης.

- **Grade IV** : Είναι κακοήθεις με πολλαπλές διαιρέσεις σε νέα καρκινικά κύτταρα και σπάνια η πλήρης αφαίρεση των όγκων αυτών

3.1.1. Περιγραφική ανάλυση ποσοτικών μεταβλητών και ποιοτικών μεταβλητών καρκίνου του εγκεφάλου

Στην παρούσα ενότητα θα μελετήσουμε τα περιγραφικά μέτρα τόσο των ποιοτικών όσο και των ποσοτικών μεταβλητών για τα δεδομένα που έχουμε για τον καρκίνο του εγκεφάλου. Οι ποσοτικές μεταβλητές που θα μας απασχολήσουν για την ανάλυσή μας είναι Survival Months και Tumor Size.

Ποσοτικές Μεταβλητές	Μέση τιμή	Τυπική Απόκλιση
Survival Months	43.20	69.72
Tumor Size	16.70	24.05

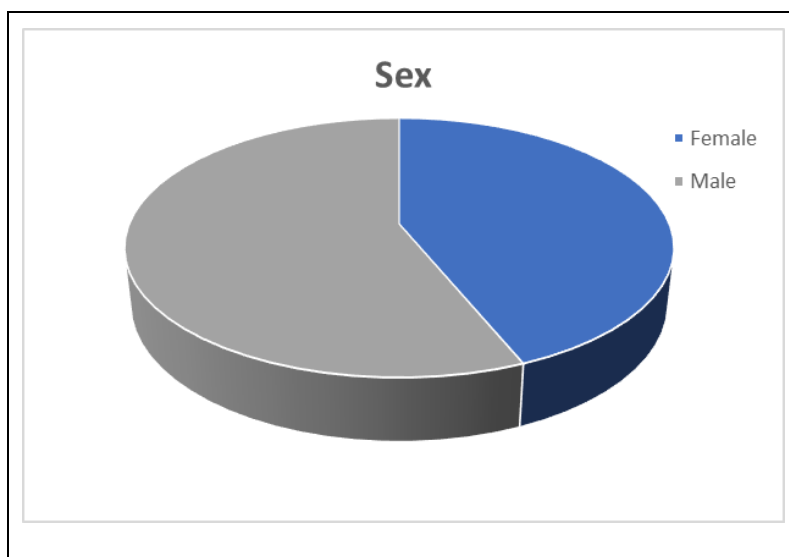
Πίνακας 3.2: Συγκεντρωτικός Πίνακας Συνεχών Μεταβλητών

Οι κατηγορικές μεταβλητές που αξίζει να εξετάσουμε είναι Sex, Age, Race, Diagnostic Confirmation, Laterality, Grade, Grade Clinical, Grade Pathological και Vital Status.

Στον παρακάτω πίνακα παρατηρούμε πως το δείγμα μας για τον καρκίνο του εγκεφάλου αποτελείται από 26.189 γυναίκες, δηλαδή 43.8% και από συνολικά 33.623 άντρες, με ποσοστό 56.2%.

Sex	Συχνότητα	Ποσοστό	Κατάσταση
Γυναίκα	26.189	43.8%	Επιβίωσαν : 6.889 (26.3%) Απεβίωσαν: 19.300 (73.7%)
Άντρας	33.623	56.2%	Επιβίωσαν : 8.137 (13.6%) Απεβίωσαν : 25.486 (86.4%)
Σύνολο	59.812	100%	

Πίνακας 3.3: Πίνακας Συχνοτήτων για την μεταβλητή Sex

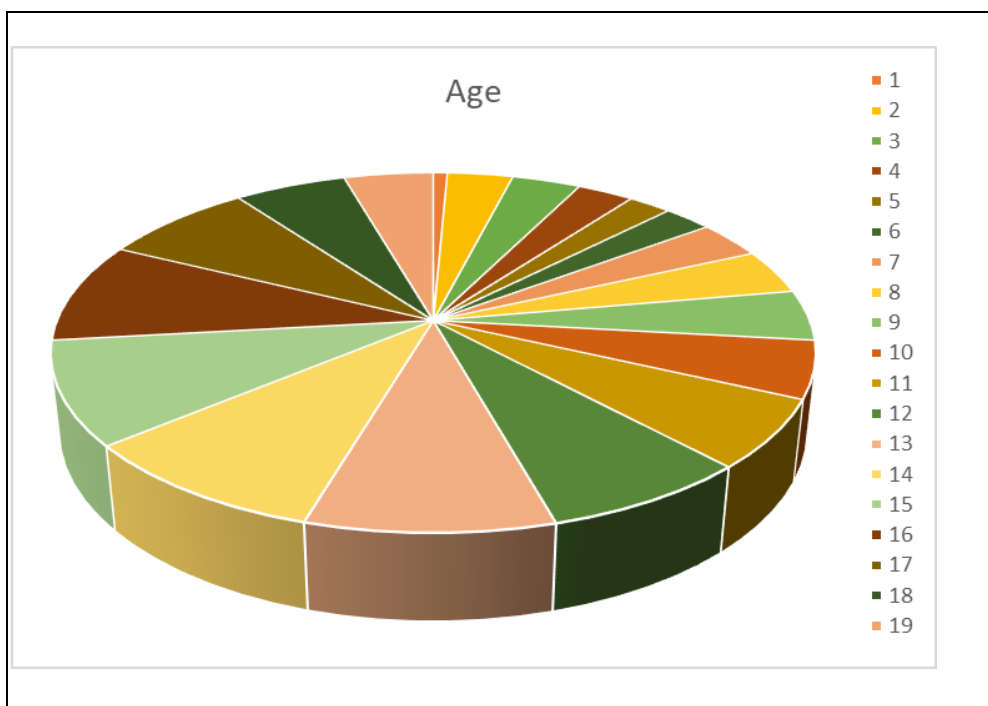


Διάγραμμα 3.1: Διάγραμμα της μεταβλητής Sex

Αναφορικά με την ηλικία παρατηρείται μεγαλύτερη συχνότητα στο εύρος των 60-64 ετών και 65-69 ετών καθώς και τα ποσοστά θνησιμότητας είναι 9.4% και 9.3% αντίστοιχα.

Age	Συχνότητα	Ποσοστό
(1) 0	420	0.7%
(2) 1-4	1.966	3.3%
(3) 5-9	2.056	3.4%
(4) 10-14	1.690	2.8%
(5) 15-19	1.296	2.2%
(6) 20-24	1.459	2.4%
(7) 25-29	2.008	3.4%
(8) 30-34	2.400	4%
(9) 35-39	2.715	4.5%
(10) 40-44	3.065	5.1%
(11) 45-49	3.690	6.2%
(12) 50-54	4.536	7.6%
(13) 55-59	5.351	8.9%
(14) 60-64	5.600	9.4%
(15) 65-69	5.549	9.3%
(16) 70-74	5.409	9%
(17) 75-79	4.566	7.6%
(18) 80-84	3.358	5.6%
(19) 85+	2.678	4.5%
Σύνολο	59.812	100%

Πίνακας 3.4: Πίνακας Συχνοτήτων για την μεταβλητή Age

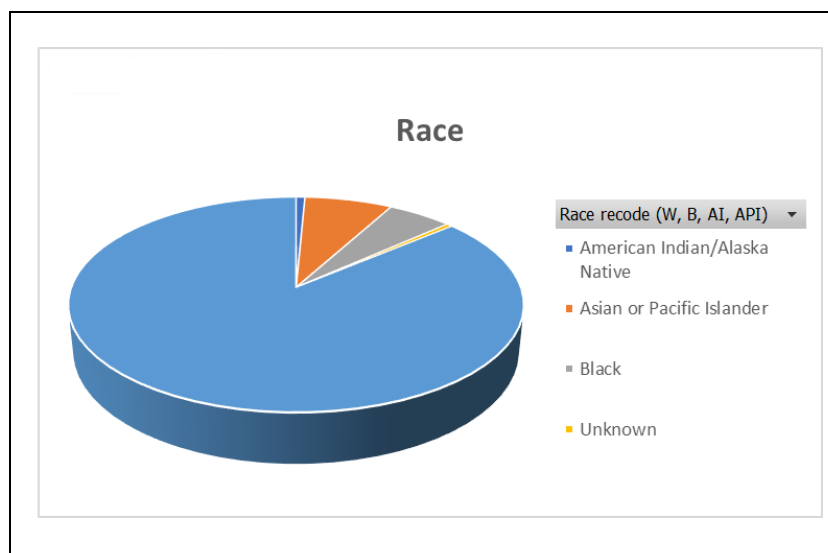


Διάγραμμα 3.2: Διάγραμμα της μεταβλητής Age

Στην συνέχεια, μέσω της μεταβλητής Race παρατηρείται ότι την μεγαλύτερη συχνότητα του καρκίνου του εγκεφάλου την έχουν οι λευκοί 51.529 (86.2%) καθώς οι έγχρωμοι είναι 3.205 (5.4%), οι Αμερικάνοι 439 (0.7%), οι Ασιάτες 4.346 (7.3%) και 293 (0.5%) αγνώστων εθνικοτήτων.

Race	Συχνότητα	Ποσοστό
Αμερικάνοι	439	0.7%
Ασιάτες	4.346	7.3%
Έγχρωμοι	3.205	5.4%
Λευκοί	51.529	86.2%
Δεν γνωρίζουμε	293	0.5%
Σύνολο	59.812	100%

Πίνακας 3.5: Πίνακας Συχνοτήτων για την μεταβλητή Race

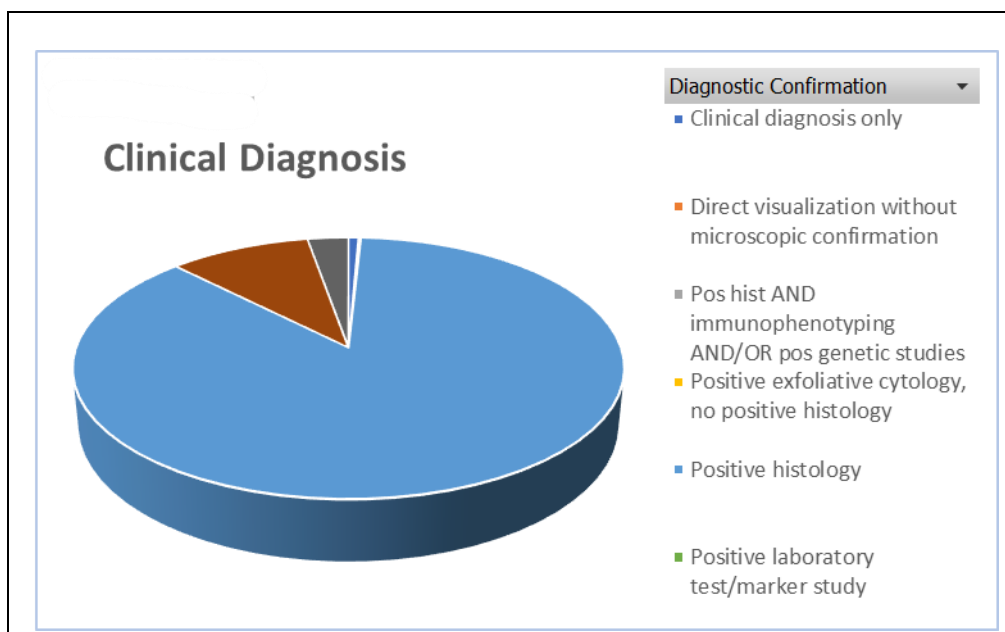


Διάγραμμα 3.3: Διάγραμμα της μεταβλητής Race

Στον παρακάτω πίνακα βλέπουμε την συχνότητα της οικογενειακής κατάστασης των γυναικών και πιο συγκεκριμένα υπάρχουν 3.439 (60.8%) παντρεμένες, 700 (12.4%) διαζευγμένες, 80 (1.4%) σε διάσταση, 736 (13%) ελεύθερες, 59 (1%) σε συμβίωση και 645 (11.4%) χήρες.

Diagnostic Confirmation	Συχνότητα	Ποσοστό
Κλινική Διάγνωση	403	0.7%
Απευθείας Οπτικοποίηση	38	0.1%
Θετική Αποφολιδωτική Κυτταρολογία	68	0.1%
Θετική Ιστολογία	51843	86.7%
Θετική Εργαστηριακή Δοκιμή	9	-
Θετική Μικροσκοπική Επιβεβαίωση	15	-
Ακτινογραφία	5768	9.6%
Δεν Γνωρίζουμε	1666	2.8%
Σύνολο	59.812	100%

Πίνακας 3.6: Πίνακας Συχνοτήτων για την μεταβλητή Diagnostic Confirmation

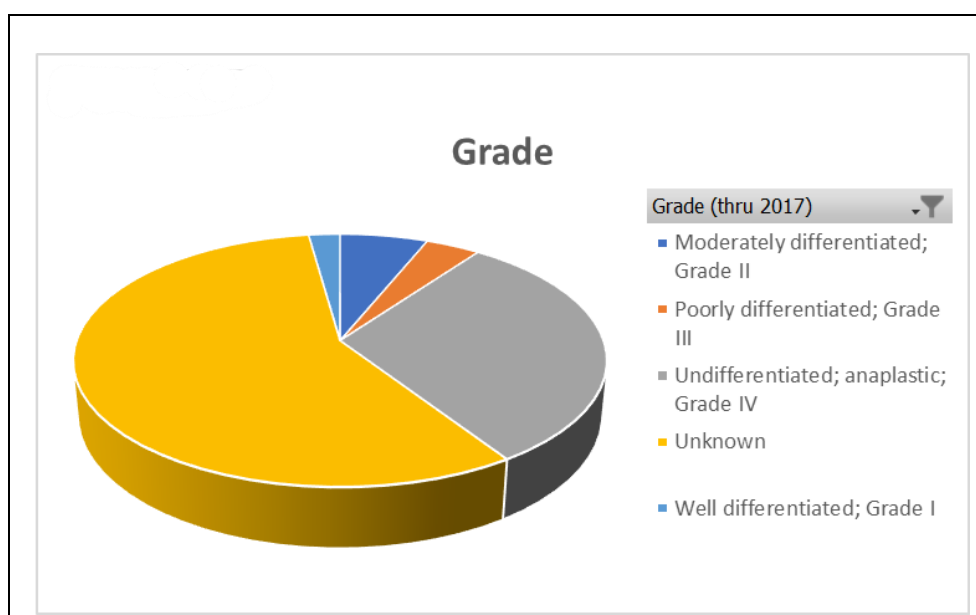


Διάγραμμα 3.4: Διάγραμμα για την μεταβλητή Diagnostic Confirmation

Για την μεταβλητή Grade που υποδηλώνει το στάδιο του καρκίνου που βρίσκεται ο ασθενής έχουμε τις συχνότητες : 1.210 (2%), 3.393 (5.7%), 2.114 (3.5%) και 16.853 (28.2%) για τους βαθμούς I, II, III και IV αντίστοιχα.

Grade	Συχνότητα	Ποσοστό
Βαθμός I	1.210	2%
Βαθμός II	3.393	5.7%
Βαθμός III	2.114	3.5%
Βαθμός IV	16.853	28.2%
Άγνωστο	36.242	60.6%
Σύνολο	59.812	100%

Πίνακας 3.7: Πίνακας Συχνοτήτων για την μεταβλητή Grade

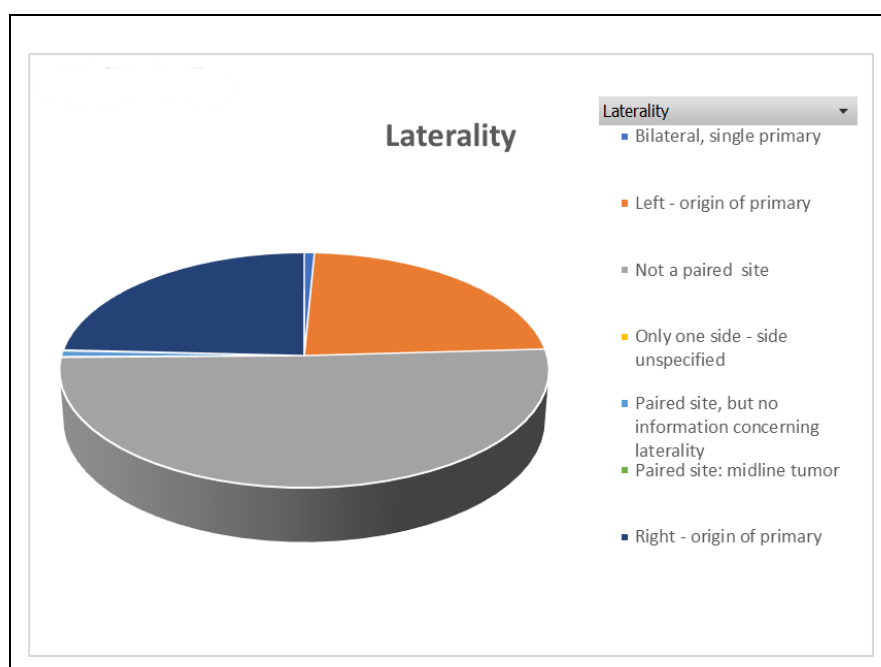


Διάγραμμα 3.5: Διάγραμμα για την μεταβλητή Grade

Αναφορικά με την πλευρικότητα του καρκίνου του εγκεφάλου και τον παρακάτω πίνακα παρατηρούμε το μεγαλύτερο ποσοστό να είναι δεξιάς προέλευσης (24,2%) και αριστερής (23.3%).

Laterality	Συχνότητα	Ποσοστό
Διμερής Πλευρικότητα	450	0.8%
Αριστερή Προέλευση	13.959	23.3%
Μη Ζευγοποιημένος	30.311	50.7%
Απροσδιόριστη	55	0.1%
Ζεύγος (χωρίς ξεκάθαρη πλευρικότητα)	487	0.8%
Μέσης Γραμμής Όγκος	64	0.1%
Δεξιά Προέλευση	14.486	34.2%
Σύνολο	59.812	100%

Πίνακας 3.8: Πίνακας Συχνοτήτων για την μεταβλητή Laterality

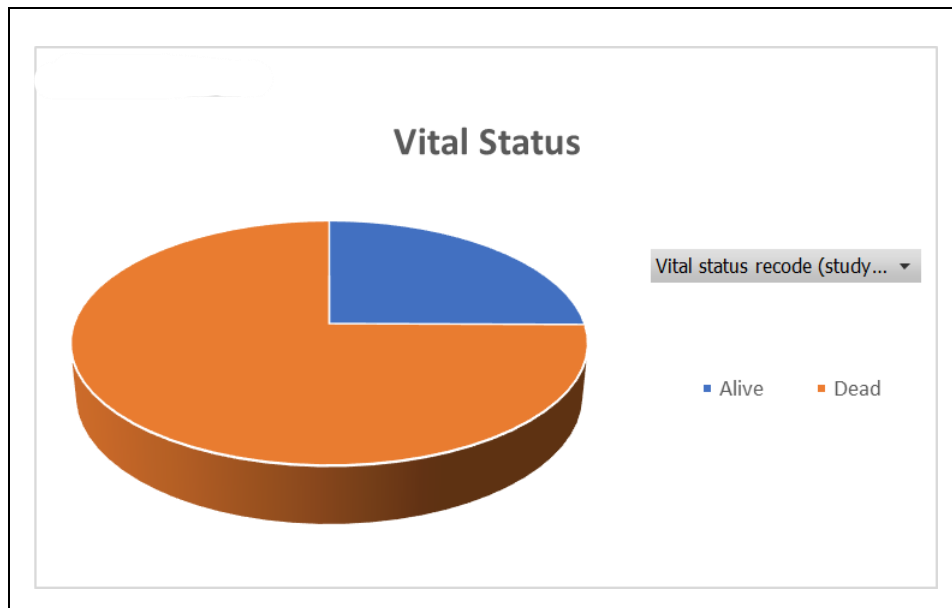


Διάγραμμα 3.6: Διάγραμμα για την μεταβλητή Laterality

Τέλος εξετάζοντας την κατάσταση των ασθενώς ερχόμαστε στο συμπέρασμα ότι 15.026 (25.1%) επιβίωσαν ενώ οι υπόλοιποι (74.9%) απεβίωσαν.

Vital Status	Συχνότητα	Ποσοστό
Επιβίωσαν	15.026	25.1%
Απεβίωσαν	44.786	74.9%
Σύνολο	59.812	100%

Πίνακας 3.9: Πίνακας Συχνοτήτων για την μεταβλητή Vital Status



Διάγραμμα 3.7: Διάγραμμα για την μεταβλητή Vital Status

3.2. Περιγραφή δεδομένων καρκίνου του μαστού

Χρησιμοποιήθηκε δείγμα από το SEERSTAT και πιο συγκεκριμένα έχουμε 5.659 ασθενείς με καρκίνο του μαστού (γυναίκες). Οι μεταβλητές που χρησιμοποιούνται παρουσιάζονται στον παρακάτω πίνακα :

Μεταβλητή	Επεξήγηση
Age	Ηλικία
Year of diagnosis	Χρονολογία διάγνωσης
Race record	Φυλή ασθενή
Marital Status	Οικογενειακή κατάσταση
ER Status	Υποδοχέας οιστρογόνου
PR Status	Υποδοχέας προγεστερόνης
Grade	Βαθμός καρκίνου
Regional Nodes Examined	Περιφερειακοί όγκοι που εξετάστηκαν
Regional Nodes Positive	Περιφερειακοί θετικοί όγκοι
Breast Subtype	Δευτερεύον τύπος καρκίνου μαστού
Diagnostic Confirmation	Μέθοδος διαγνωστικής επιβεβαίωσης
CS tumor size	Μέγεθος όγκου καρκίνου
Survival months	Μήνες επιβίωσης
Stage A	Εξάπλωση του καρκίνου
Stage T	Κλίμακα για το μέγεθος του όγκου και την θέση συμπεριλαμβανομένου του κατά πόσο ο όγκος έχει αναπτυχθεί σε κοντινούς ιστούς (0-4)

Stage N	Κλίμακα για τον αριθμό των λεμφαδένων με καρκίνο (0-3)
Laterality	Πλευρικότητα όγκου
Vital status recode	Εν ζωή ή όχι (0=dead, 1=alive)

Πίνακας 3.10: Μεταβλητές καρκίνου μαστού

Για την καλύτερη κατανόηση της ανάλυσης του επιμέρους κεφαλαίου θα προστεθεί η παρακάτω κωδικοποίηση αναφορικά με το πεδίο «Clinical/ Pathological Grade». Πιο συγκεκριμένα :

- **Grade I** : Τέτοιοι όγκοι είναι συνήθως καλοήθεις, μοιάζουν σχεδόν με φυσιολογικά κύτταρα και επεκτείνονται αργά.
- **Grade II** : Είναι κύτταρα ενδιάμεσου βαθμού επιθετικότητας και φαίνονται λιγότερα φυσιολογικά από εκείνα με Grade I. Αναπτύσσονται και διαιρούνται λίγο πιο γρήγορα από τα κανονικά κύτταρα.
- **Grade III** : Τα κύτταρα αυτά είναι πιο επιθετικά, πολύ διαφορετικά σε σύγκριση με τα φυσιολογικά καθώς και η αφαίρεσή τους είναι αρκετά δύσκολη λόγω πιθανής μετάδοσης.
- **Grade IV** : Είναι κακοήθεις με πολλαπλές διαιρέσεις σε νέα καρκινικά κύτταρα και σπάνια η πλήρης αφαίρεση των όγκων αυτών.

Επίσης ορθό θα ήταν να γίνει η επεξήγηση και για την μεταβλητή «Breast Subtype» καθώς ο καρκίνος του μαστού δεν είναι μία ομοιόμορφη νόσος, αλλά αντιθέτως, υπάρχουν διάφοροι τύποι καρκίνου του μαστού, οι οποίοι μπορούν να διαφέρουν σε ιστολογικά, μορφολογικά, και γενετικά χαρακτηριστικά, καθώς και στην ανταπόκρισή τους σε θεραπείες ως εξής :

- **Luminal A** (*Her2-/ HR+*) : τύπος καρκίνου μαστού θετικός στους υποδοχείς ορμονών και αρνητικός στο Her2 γεγονός που μας δείχνει ότι υπάρχουν χαμηλά επίπεδα πρωτεΐνης Ki-67, η οποία βοηθάει στον έλεγχο της ταχύτητας ανάπτυξης καρκινικών κυττάρων. Καρκίνοι χαμηλού βαθμού, τείνουν να αναπτύσσονται αργά και να έχουν καλύτερη πρόγνωση. Αυτός ο τύπος καρκίνου του μαστού εξαρτάται από ορμόνες, όπως το οιστρογόνο ή η προγεστερόνη, για την ανάπτυξή του.
- **Luminal B** (*Her2-/ HR+ ή Her2+/ HR+*) : τύπος θετικός στους υποδοχείς ορμονών (οιστρογόνου ή/και προγεστερόνης) και είτε θετικός είτε αρνητικός στο Her2. Τα καρκινικά κύτταρα αναπτύσσονται ελαφρώς ταχύτερα από εκείνα του Luminal A. Η θεραπεία με αντί-Her2 φάρμακα μπορεί να είναι αποτελεσματική σε αυτές τις περιπτώσεις.
- **Τριπλός Αρνητικός** (*Triple Negative : Her2-/ HR-*) : τύπος αρνητικός στους υποδοχείς ορμονών. Είναι συχνά πιο επιθετικός και πιο δύσκολος στη θεραπεία.
- **Her2-enriched** (*Her2+/ HR-*) : τύπος αρνητικός στους υποδοχείς ορμονών και θετικός στο Her2. Τείνουν να αναπτύσσονται γρηγορότερα από τους καρκίνους Luminal και μπορεί να έχουν χειρότερη πρόγνωση, αλλά συχνά αντιμετωπίζονται επιτυχώς με στοχευμένες θεραπείες που στοχεύουν στην πρωτεΐνη Her2.

3.2.1. Περιγραφική ανάλυση ποσοτικών μεταβλητών και ποιοτικών μεταβλητών καρκίνου του μαστού

Στους παρακάτω πίνακες παρουσιάζονται τα περιγραφικά μέτρα των συνεχών και μη μεταβλητών καθώς και τα διαγράμματα αυτών. Οι ποσοτικές μεταβλητές που θα μας απασχολήσουν για την ανάλυσή μας είναι Survival Months και Tumor Size.

Ποσοτικές Μεταβλητές	Μέση τιμή	Τυπική Απόκλιση
Survival Months	32.71	20.32
Tumor Size	33.35	23.31

Πίνακας 3.11: Συγκεντρωτικός Πίνακας Συνεχών Μεταβλητών

Οι κατηγορικές μεταβλητές που θα εξετάσουμε είναι Sex, Age, Race, Material Status, Grade, Laterality, Stage A, ER Status, PR Status, Stage T, Stage N και Status. Στον παρακάτω πίνακα παρατηρούμε πως το δείγμα μας για τον καρκίνο του μαστού αποτελείται μόνο από γυναίκες και το 89.8% από αυτές επιβίωσαν.

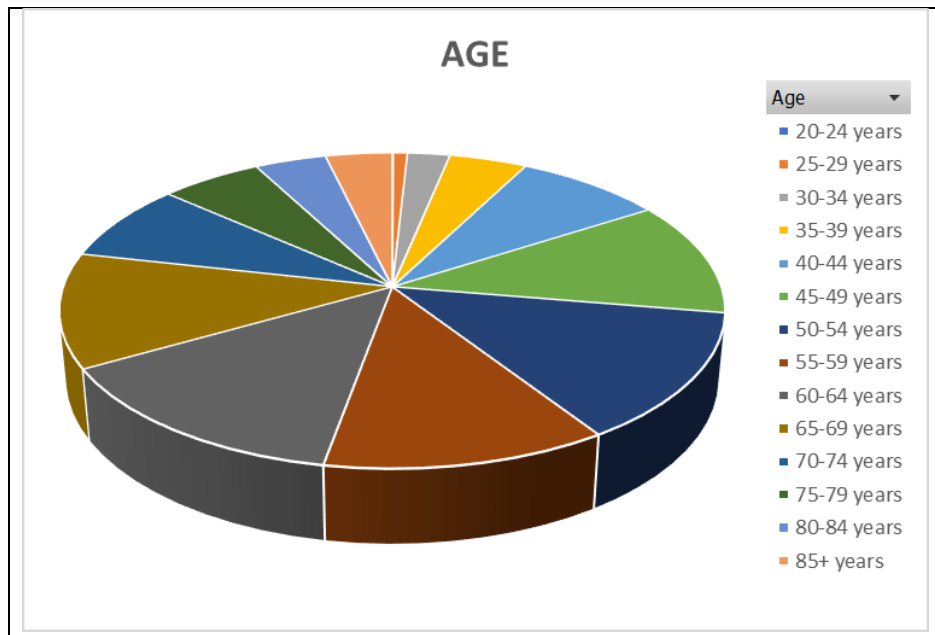
Sex	Συχνότητα	Ποσοστό	Κατάσταση
Γυναίκα	5.659	100	Επιβίωσαν : 5.085 (89.8%) Απεβίωσαν : 574 (10.1%)
Σύνολο	5.659	100	

Πίνακας 3.12: Πίνακας Συχνοτήτων για την μεταβλητή Sex

Αναφορικά με την ηλικία παρατηρείται μεγαλύτερη συχνότητα στο εύρος των 45-64 ετών.

Age	Συχνότητα	Ποσοστό
(1) 20-24	3	0.1%
(2) 25-29	45	0.8%
(3) 30-34	135	2.4%
(4) 35-39	247	4.4%
(5) 40-44	471	8.3%
(6) 45-49	663	11.7%
(7) 50-54	744	13.1%
(8) 55-59	683	12.1%
(9) 60-64	787	13.9%
(10) 65-69	665	11.8%
(11) 70-74	453	8%
(12) 75-79	324	5.7%
(13) 80-84	227	4%
(14) 85+	212	3.7%
Σύνολο	5.659	100

Πίνακας 3.13: Πίνακας Συχνοτήτων για την μεταβλητή Age

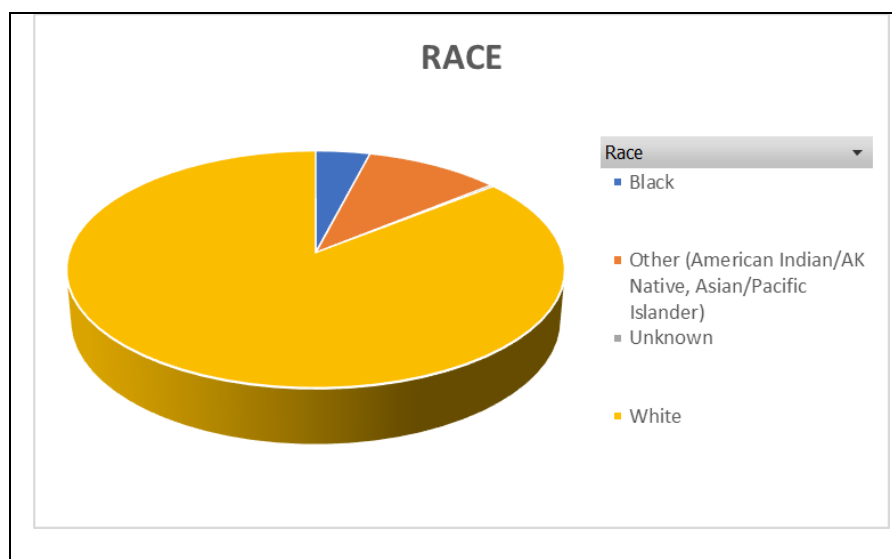


Διάγραμμα 3.8: Διάγραμμα της μεταβλητής Age

Στην συνέχεια παρατηρείται ότι την μεγαλύτερη συχνότητα του καρκίνου του μαστού την έχουν οι λευκές γυναίκες (85.7%), οι έγχρωμες γυναίκες είναι 228 (4%), 570 (10.1%) άλλως εθνικοτήτων και 14 (0.2%) αγνώστων εθνικοτήτων.

Race	Συχνότητα	Ποσοστό
Λευκοί	4.847	85.7%
Έγχρωμοι	228	4%
Άλλο	570	10.1%
Δεν γνωρίζουμε	14	0.2%
Σύνολο	5.659	100%

Πίνακας 3.14: Πίνακας Συχνοτήτων για την μεταβλητή Race

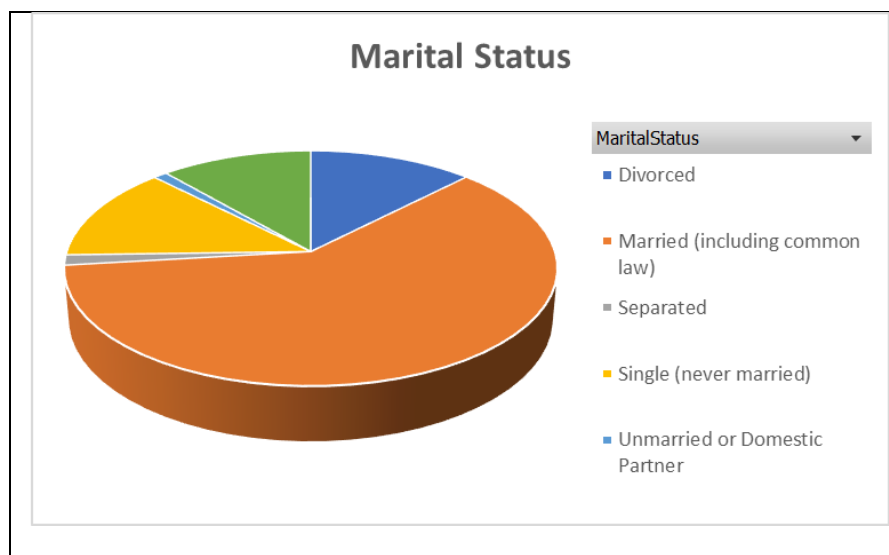


Διάγραμμα 3.9: Διάγραμμα της μεταβλητής Race

Στον παρακάτω πίνακα βλέπουμε την συχνότητα της οικογενειακής κατάστασης των γυναικών και πιο συγκεκριμένα υπάρχουν 3.439 (60.8%) παντρεμένες, 700 (12.4%) διαζευγμένες, 80 (1.4%) σε διάσταση, 736 (13%) ελεύθερες, 59 (1%) σε συμβίωση και 645 (11.4%) χήρες.

Marital Status	Συχνότητα	Ποσοστό
Διαζευγμένες	700	12.4%
Παντρεμένες	3.439	60.8%
Σε διάσταση	80	1.4%
Ελεύθερες	736	13%
Σε συμβίωση	59	1%
Χήρες	645	11.4%
Σύνολο	5.659	100%

Πίνακας 3.15: Πίνακας Συχνοτήτων για την μεταβλητή Marital Status

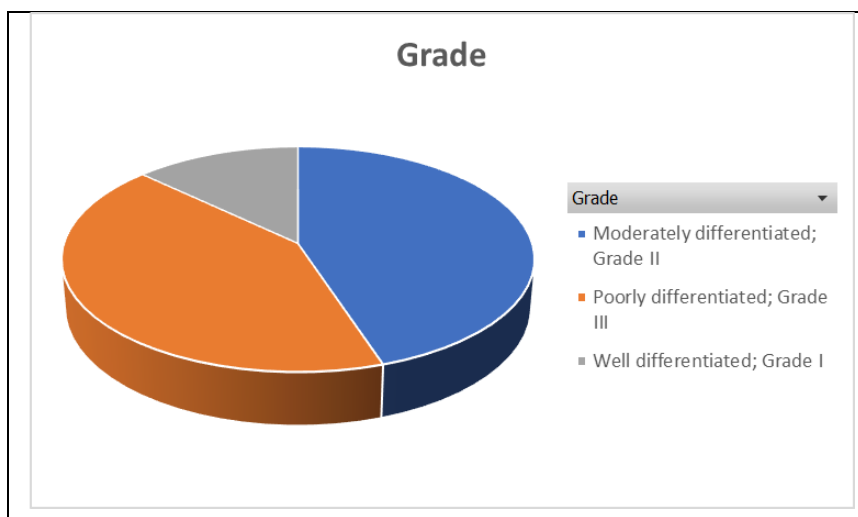


Διάγραμμα 3.10: Διάγραμμα για την μεταβλητή Marital Status

Για την μεταβλητή Grade που υποδηλώνει το στάδιο του καρκίνου που βρίσκεται η ασθενής έχουμε τις συχνότητες : 735 (13%), 2.542 (44.9%) και 2.383 (42.1%) για τους βαθμούς I, II και III αντίστοιχα.

Grade	Συχνότητα	Ποσοστό
Βαθμός I	735	13%
Βαθμός II	2.541	44.9%
Βαθμός III	2.383	42.1%
Σύνολο	5.659	100%

Πίνακας 3.16: Πίνακας Συχνοτήτων για την μεταβλητή Grade

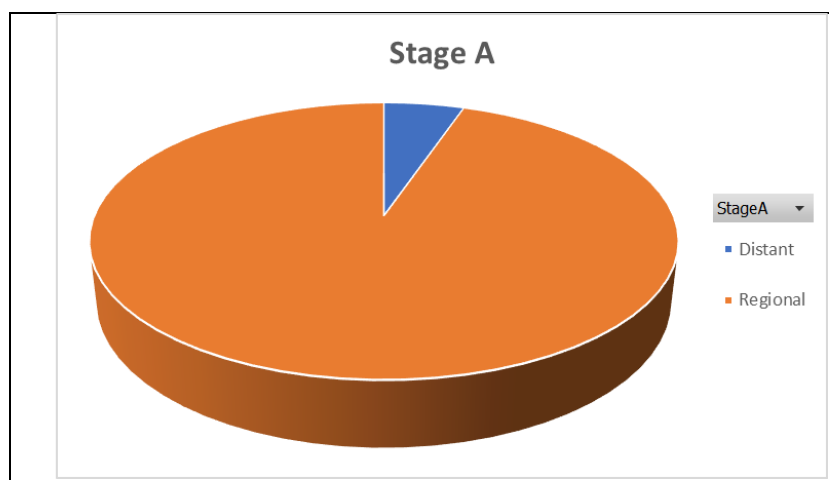


Διάγραμμα 3.11: Διάγραμμα για την μεταβλητή Grade

Αναφορικά με την μεταβλητή Stage A παρατηρήσαμε ότι σε 297 (5.2%) γυναίκες ο καρκίνος είναι μακρινός και σε 5.362 (94.8%) είναι σε κοντινούς λεμφαδένες.

Stage A	Συχνότητα	Ποσοστό
Μακρινός	297	5.2%
Περιφερικός	5.362	94.8%
Σύνολο	5.659	100%

Πίνακας 3.17: Πίνακας Συχνοτήτων για την μεταβλητή Stage A

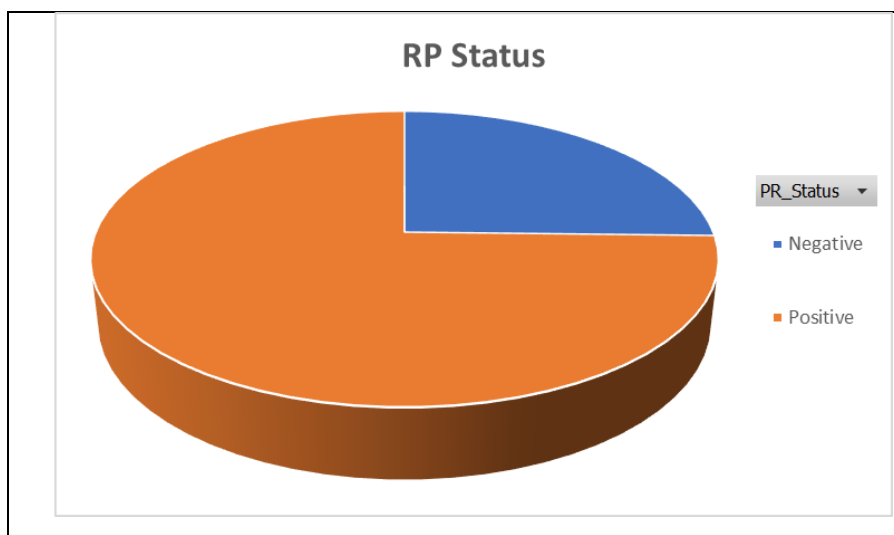


Διάγραμμα 3.12: Διάγραμμα για την μεταβλητή Stage A

Στην συνέχεια παρατηρείται ότι σε 1.436 (25.4%) γυναίκες ο υποδοχέας προγεστερόνης είναι θετικός ενώ σε 4.223 (74.6%) ο υποδοχέας προγεστερόνης είναι αρνητικός.

PR Status	Συχνότητα	Ποσοστό
Αρνητικό	1.436	25.4%
Θετικό	4.223	74.6%
Σύνολο	5.659	100%

Πίνακας 3.18: Πίνακας Συχνοτήτων για την μεταβλητή PR Status

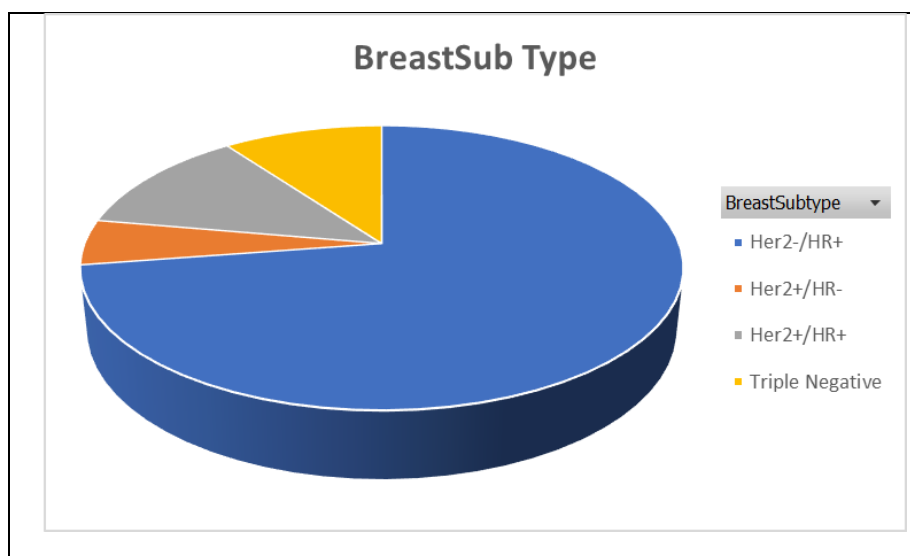


Διάγραμμα 3.13: Διάγραμμα για την μεταβλητή PR Status

Έπειτα παρατηρήσαμε ότι 4.108 (72.6%) γυναίκες έχουν Her2+/ HR+, 292 (5.2%) έχουν Her2+/ HR-, 688 (12.2%) έχουν Her2+/ HR+ και 571 (10.1%) έχουν αρνητικούς υποδοχείς ορμονών και αρνητικό Her2.

BreastSub Type	Συχνότητα	Ποσοστό
Her2-/HR+	4.108	72.6%
Her2+/HR-	292	5.2%
Her2+/HR+	688	12.2%
Τριπλό Αρνητικό	571	10.1%
Σύνολο	5.659	100%

Πίνακας 3.19: Πίνακας Συχνοτήτων για την μεταβλητή BreastSub Type

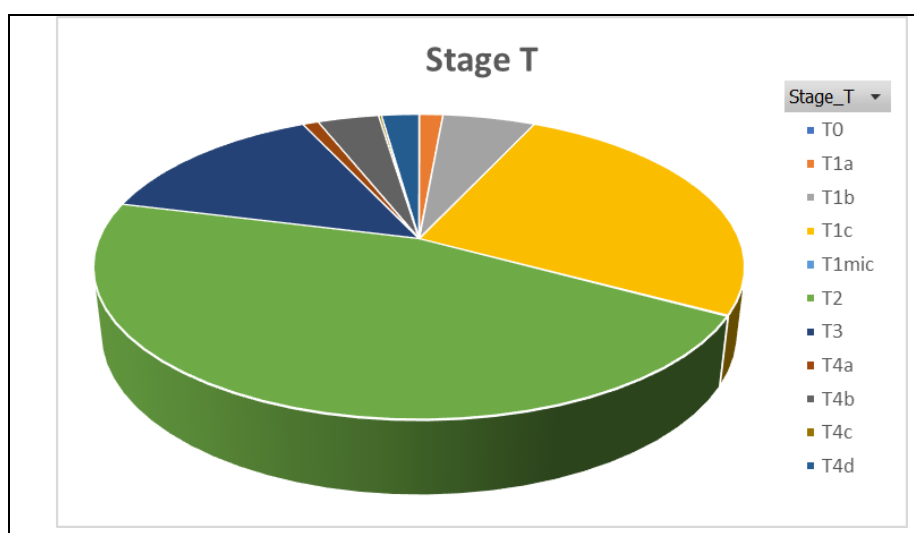


Διάγραμμα 3.14: Διάγραμμα για την μεταβλητή BreastSub Type

Από τον παρακάτω πίνακα φάνηκε ότι στις περισσότερες γυναίκες το μέγεθος και η θέση του όγκου περιγράφεται από την T2.

Stage T	Συχνότητα	Ποσοστό
T0	1	0
T1a	79	1.4%
T1b	311	5.5%
T1c	1.470	26%
T1mic	5	0.1%
T2	2.600	45.9%
T3	799	14.1%
T4a	53	0.9%
T4b	205	3.6%
T4c	10	0.2%
T4d	126	2.2%
Σύνολο	5.659	100%

Πίνακας 3.20: Πίνακας Συχνοτήτων για την μεταβλητή Stage T

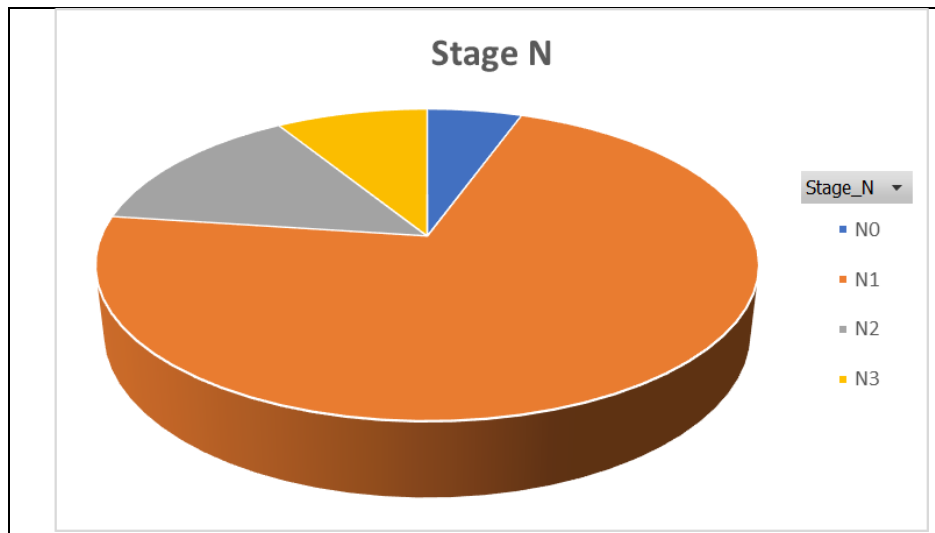


Διάγραμμα 3.15: Διάγραμμα για την μεταβλητή Stage T

Διαπιστώθηκε ότι οι περισσότερες γυναίκες παρουσιάζουν τον τύπο N1 όπως φαίνεται και στο πίνακα.

Stage N	Συχνότητα	Ποσοστό
N0	312	5.5%
N1	4.052	71.6%
N2	795	14%
N3	500	8.8%
Σύνολο	5.659	100%

Πίνακας 3.21: Πίνακας Συχνοτήτων για την μεταβλητή Stage N

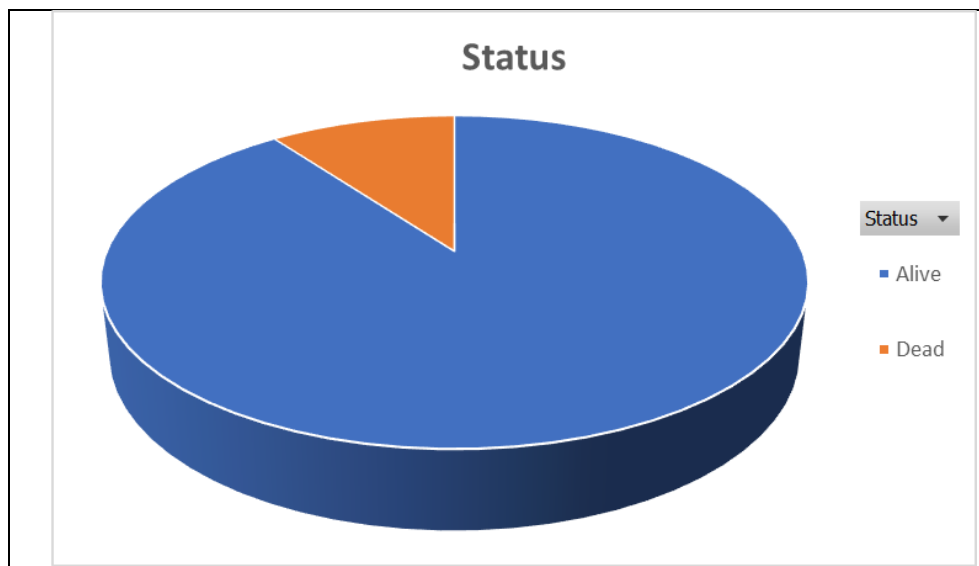


Διάγραμμα 3.16: Διάγραμμα για την μεταβλητή Stage N

Τέλος βρέθηκε ότι 5.085 (89.9%) γυναίκες επιβίωσαν από τον καρκίνο του μαστού ενώ 574 (10.1%) απεβίωσαν.

Status	Συχνότητα	Ποσοστό
Επιβίωσαν	5.085	89.9%
Απεβίωσαν	574	10.1%
Σύνολο	5.659	100%

Πίνακας 3.22: Πίνακας Συχνοτήτων για την μεταβλητή Status



Διάγραμμα 3.17: Διάγραμμα για την μεταβλητή Status

3.3. Περιγραφή δεδομένων καρκίνου του παγκρέατος

Για την τελευταία ανάλυση του καρκίνου του παγκρέατος χρησιμοποιήθηκε εξίσου δείγμα από SEERSTAT και οι ασθενείς που εντοπίστηκαν είναι 58.516 γυναίκες και 58.066 άντρες. Οι ανεξάρτητες μεταβλητές που θα μελετηθούν αναφέρονται στον κάτωθι πίνακα :

Μεταβλητή	Επεξήγηση
Patient ID	Κωδικός ασθενή
Age record	Ηλικιακό εύρος
Year of diagnosis	Χρονολογία διάγνωσης
Race record	Φυλή ασθενή
Site record	Είδος καρκίνου
Grade	Βαθμός καρκίνου
Diagnostic Confirmation	Μέθοδος διαγνωστικής επιβεβαίωσης
CS tumor size	Μέγεθος όγκου καρκίνου
Survival months	Μήνες επιβίωσης
Grade clinical	Κλινικός βαθμός καρκίνου
Grade pathological	Παθολογικός βαθμός καρκίνου
Historic Stage A	Έκταση νόσου
Vital status recode	Εν ζωή ή όχι (0=dead, 1=alive)

Πίνακας 3.23: Μεταβλητές καρκίνου παγκρέατος

3.3.1. Περιγραφική ανάλυση ποσοτικών μεταβλητών και ποιοτικών μεταβλητών καρκίνου του παγκρέατος

Στην παρούσα ενότητα θα ασχοληθούμε με περιγραφική ανάλυση των ποσοτικών και ποιοτικών μεταβλητών για το dataset του καρκίνου του παγκρέατος. Αρχικά οι ποσοτικές μεταβλητές είναι Survival Months και Tumor Size και στον παρακάτω πίνακα έχουμε την μέση τιμή και την τυπική απόκλιση τους.

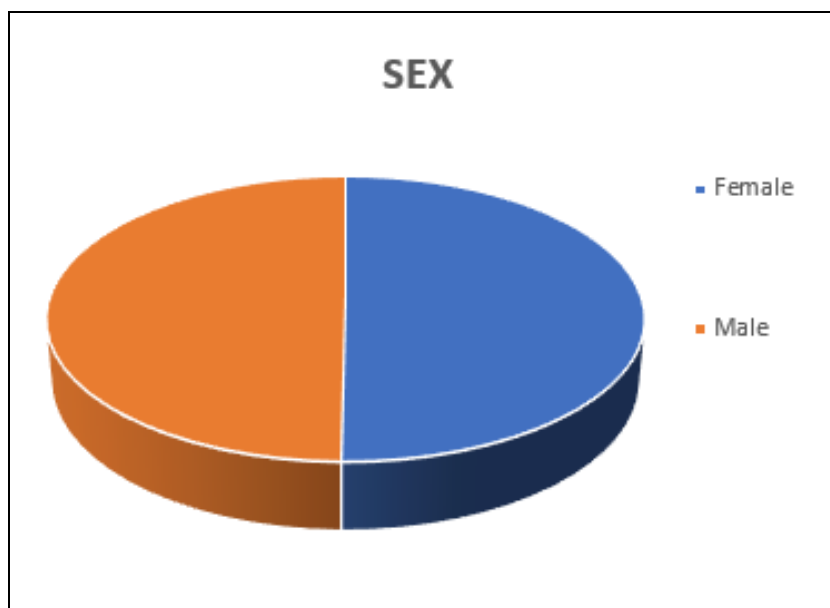
Ποσοτικές Μεταβλητές	Μέση τιμή	Τυπική Απόκλιση
Survival Months	16.95	26.73
Tumor Size	40.85	430.16

Πίνακας 3.24: Συγκεντρωτικός Πίνακας Συνεχών Μεταβλητών

Οι κατηγορικές μεταβλητές που αξίζει να εξετάσουμε είναι Sex, Age, Race, Diagnostic Confirmation, Grade, Grade Clinical, Grade Pathological και Vital Status. Στον παρακάτω πίνακα παρατηρούμε πως το δείγμα μας για τον καρκίνο του παγκρέατος αποτελείται από 50.2% γυναίκες και 49.8% άντρες.

Sex	Συχνότητα	Ποσοστό	Κατάσταση
Γυναίκα	58.516	50.2%	Επιβίωσαν : 5.523 (9.4%) Απεβίωσαν : 52.993 (90.6%)
Άντρας	58.066	49.8%	Επιβίωσαν : 5.510 (9.5%) Απεβίωσαν : 52.556 (90.5%)
Σύνολο	116.582	100%	

Πίνακας 3.25: Πίνακας Συχνοτήτων για την μεταβλητή Sex

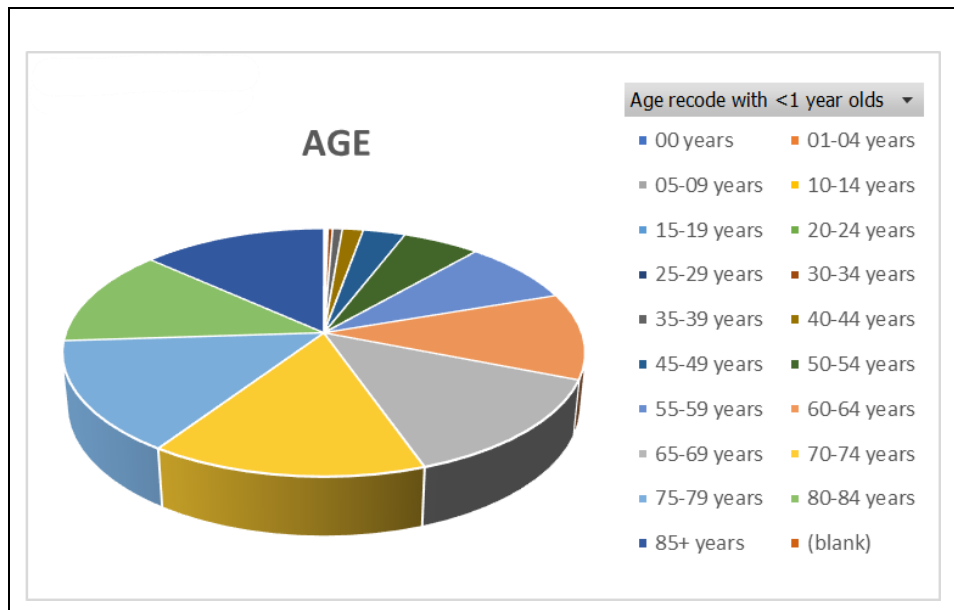


Διάγραμμα 3.18: Διάγραμμα της μεταβλητής Sex

Παρατηρείται ότι ο καρκίνος του παγκρέατος εντοπίζεται πιο συχνά στις ηλικίες των 70-74 ετών και 75-79 ετών σε ποσοστά 14.8% και 14.6% αντίστοιχα.

Age	Συχνότητα	Ποσοστό
(1) 0	2	0
(2) 1-4	6	0
(3) 5-9	14	0
(4) 10-14	28	0
(5) 15-19	51	0
(6) 20-24	90	0.1%
(7) 25-29	151	0.1%
(8) 30-34	393	0.3%
(9) 35-39	813	0.7%
(10) 40-44	1.753	1.5%
(11) 45-49	3.540	3%
(12) 50-54	6.484	5.6%
(13) 55-59	9.688	8.3%
(14) 60-64	13.134	11.3%
(15) 65-69	15.869	13.6%
(16) 70-74	17.230	14.8%
(17) 75-79	16.990	14.6%
(18) 80-84	14.797	12.7%
(19) 85+	15.549	13.3%
Σύνολο	116.582	100%

Πίνακας 3.26: Πίνακας Συχνοτήτων για την μεταβλητή Age

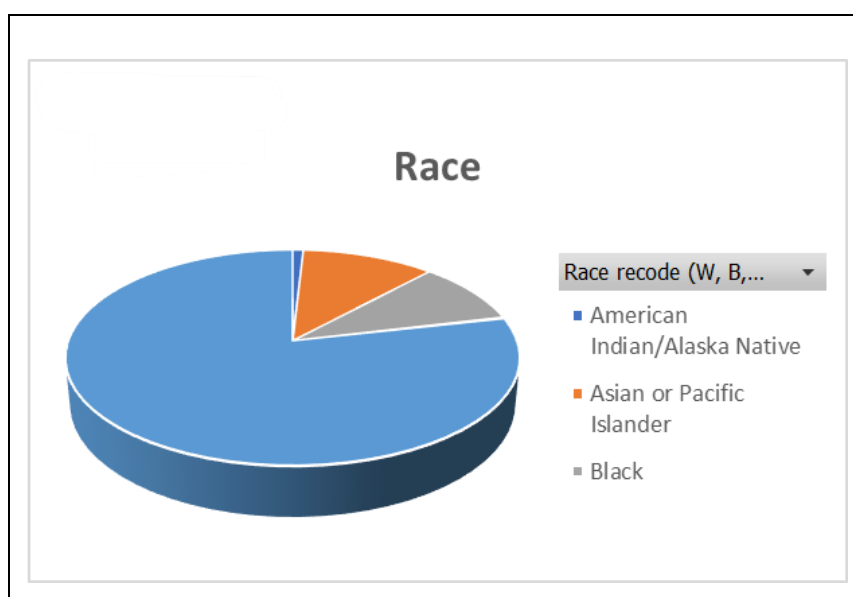


Διάγραμμα 3.19: Διάγραμμα της μεταβλητής Age

Εν συνεχεία παρατηρείται ότι την μεγαλύτερη συχνότητα του καρκίνου του παγκρέατος την έχουν οι λευκοί 91.508 (78.5%) καθώς οι έγχρωμοι είναι 10.934 (9.4%), οι Αμερικάνοι 1.207 (0.9%), οι Ασιάτες 12.831 (11%) και 282 (0.2%) αγνώστων εθνικοτήτων.

Race	Συχνότητα	Ποσοστό
Αμερικάνοι	1.207	0.9%
Ασιάτες	12.831	11%
Έγχρωμοι	10.934	9.4%
Λευκοί	91.508	78.5%
Δεν γνωρίζουμε	282	0.2%
Σύνολο	116.582	100%

Πίνακας 3.27: Πίνακας Συχνοτήτων για την μεταβλητή Race

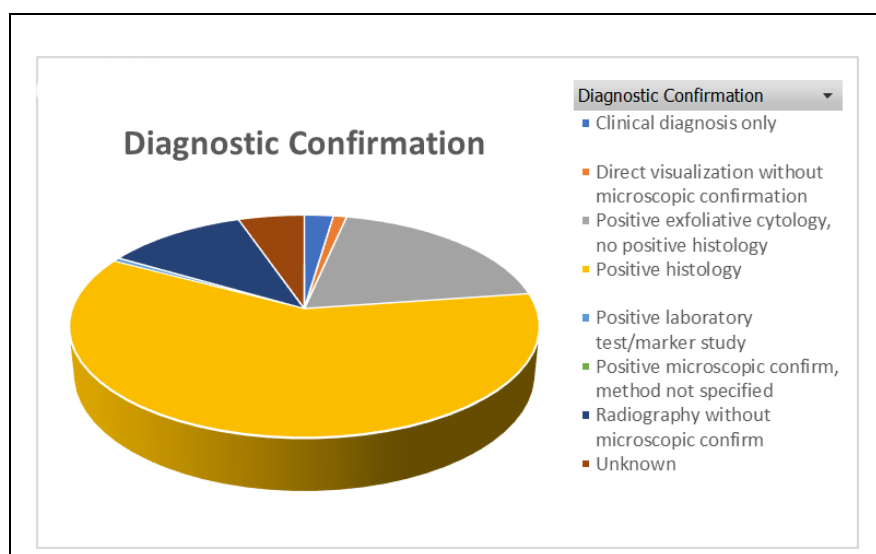


Διάγραμμα 3.20: Διάγραμμα της μεταβλητής Race

Επίσης στον παρακάτω πίνακα βλέπουμε την συχνότητα της μεθόδου διάγνωσης του καρκίνου καταλήγοντας ότι οι περισσότερες διαγνώσεις προήλθαν από θετική ιστολογία, θετική αποφολιδωτική κυτταρολογία και ακτινογραφία.

Diagnostic Confirmation	Συχνότητα	Ποσοστό
Κλινική Διάγνωση	2.704	2.3%
Απευθείας Οπτικοποίηση	1.237	1.1%
Θετική Αποφολιδωτική Κυτταρολογία	22.584	19.4%
Θετική Ιστολογία	70.043	60.1%
Θετική Εργαστηριακή Δοκιμή	604	0.5%
Θετική Μικροσκοπική Επιβεβαίωση	71	0.1%
Ακτινογραφία	12.175	11.3%
Δεν Γνωρίζουμε	6.164	5.3%
Σύνολο	116.582	100%

Πίνακας 3.28: Πίνακας Συχνοτήτων για την μεταβλητή Diagnostic Confirmation

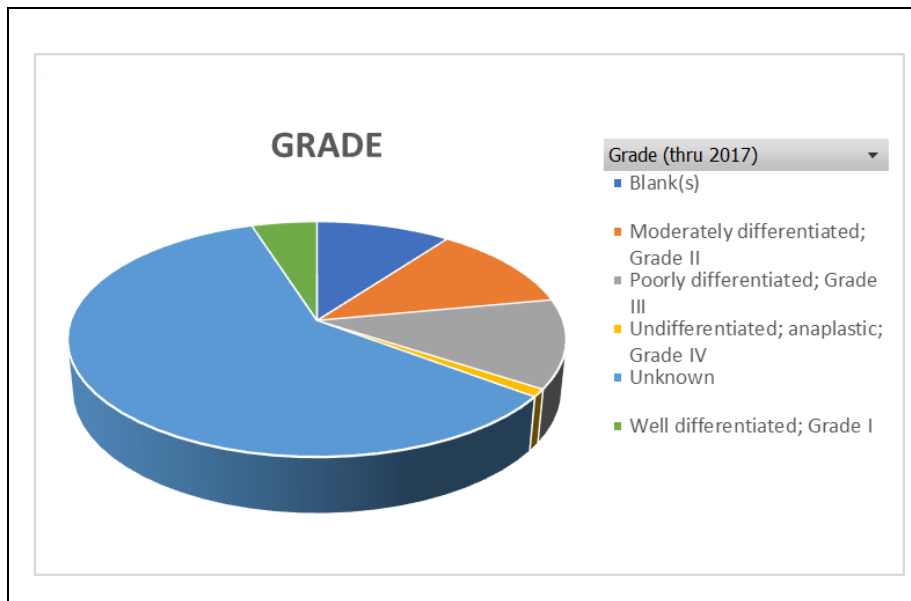


Διάγραμμα 3.21: Διάγραμμα για την μεταβλητή Diagnostic Confirmation

Για το στάδιο του καρκίνου του ασθενή έχουμε τις συχνότητες : 5.750 (4.9%), 13.839 (11.9%), 14.220 (12.2%) και 1.230 (1.1%) για τους βαθμούς I, II, III και IV αντίστοιχα.

Grade	Συχνότητα	Ποσοστό
Βαθμός I	5.750	4.9%
Βαθμός II	13.839	11.9%
Βαθμός III	14.220	12.2%
Βαθμός IV	1.230	1.1%
Άγνωστο	81.543	70%
Σύνολο	116.582	100%

Πίνακας 3.29: Πίνακας Συχνοτήτων για την μεταβλητή Grade

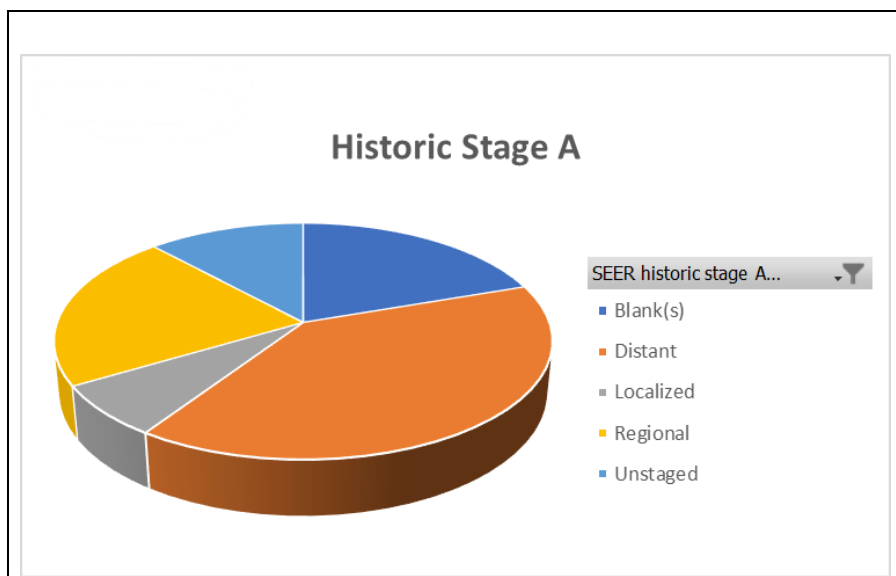


Διάγραμμα 3.22: Διάγραμμα για την μεταβλητή Grade

Παρατηρείται ότι στο δείγμα σε 46220 (39,6%) ασθενείς ο καρκίνος έχει εξαπλωθεί σε μακρινά μέρη τους σώματος, σε 8292 (7,1%) περιορίζεται στην αρχική του εστία και σε 25246 (21,7%) ο καρκίνος έχει εξαπλωθεί σε κοντινούς λεμφαδένες ή ιστούς.

Historical Stage A	Συχνότητα	Ποσοστό
Μακρινός	46.220	39.6%
Εντοπισμένος	8.292	7.1%
Περιφερικός	25.246	21.7%
Μη Κατηγοριοποιημένος	13.819	11.9%
Δεν γνωρίζουμε	23.005	19.7%
Σύνολο	116.582	100%

Πίνακας 3.30: Πίνακας Συχνότητας για την μεταβλητή Historical Stage A

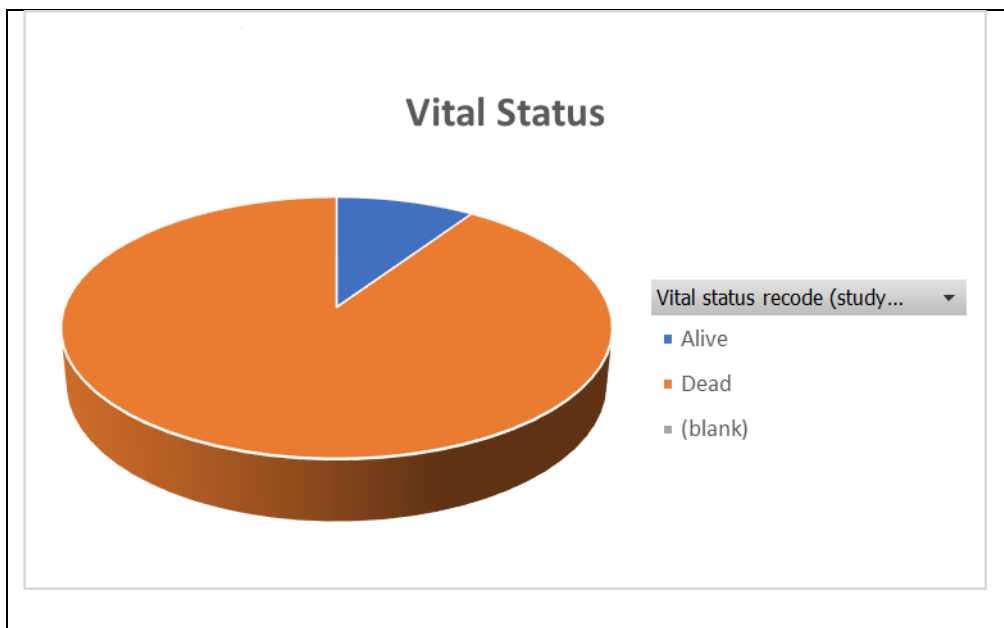


Διάγραμμα 3.23: Διάγραμμα για την μεταβλητή Historical Stage A

Στην τελευταία κατηγορική μεταβλητή που έχουμε καταλήξουμε ότι 11.033 (9.5%) του δείγματος επιβίωσαν ενώ οι 105.549 απεβίωσαν (90.5%).

Vital Status	Συχνότητα	Ποσοστό
Επιβίωσαν	11.033	9.5%
Απεβίωσαν	105.549	90.5%
Σύνολο	116.582	100%

Πίνακας 3.31: Πίνακας Συχνοτήτων για την μεταβλητή Vital Status



Διάγραμμα 3.24: Διάγραμμα για την μεταβλητή Vital Status

ΚΕΦΑΛΑΙΟ 4^ο

4. Έλεγχοι συσχετίσεων και ανεξαρτησίας των μεταβλητών

Στο παρόν κεφάλαιο θα μελετηθεί η συσχέτιση και η ανεξαρτησία των μεταβλητών του κάθε dataset ξεχωριστά. Προτού προχωρήσουμε στην εύρεση συσχετίσεων και ανεξαρτησίας ελέγξαμε την κανονικότητα των ποσοτικών μεταβλητών του συνόλου των δεδομένων. Με τον όρο κανονικότητα των δεδομένων εννοούμε πως το δείγμα που έχουμε προς ανάλυση προέρχεται από ένα σύνολο δεδομένων το οποίο ακολουθεί την κανονική κατανομή. Ο έλεγχος αυτός θεωρείται από τους σπουδαιότερους στη Στατιστική και η μορφή των ελέγχων αυτού του είδους είναι:

Μηδενική υπόθεση H_0 : η κατανομή από την οποία προέρχεται το δείγμα είναι η κανονική
vs

Εναλλακτική υπόθεση H_1 : η κατανομή από την οποία προέρχεται το δείγμα δεν είναι η κανονική. (Κούτρας Μ., 2018).

Στους παρακάτω πίνακες παρουσιάζονται τα αποτελέσματα του Lilliefors test για κάθε dataset ξεχωριστά.

Μεταβλητή	Τιμή p-value
Tumor Size	<2.2e-16
Survival Months	<2.2e-16
Total number of in situ / malignant tumors for patient	<2.2e-16

Πίνακας 4.1: Έλεγχοι κανονικότητας για το dataset του καρκίνου του Εγκεφάλου

Μεταβλητή	Τιμή p-value
Tumor Size	<2.2e-16
Survival Months	<2.2e-16
Regional Nodes Examined	<2.2e-16
Regional Nodes Positive	<2.2e-16

Πίνακας 4.2: Έλεγχοι κανονικότητας για το dataset του καρκίνου του Μαστού

Μεταβλητή	Τιμή p-value
Tumor Size	<2.2e-16
Survival Months	<2.2e-16
Total number of in situ / malignant tumors for patient	<2.2e-16

Πίνακας 4.3: Έλεγχοι κανονικότητας για το dataset του καρκίνου του Παγκρέατος

Σύμφωνα με τις παραπάνω τιμές των p-value και το επίπεδο σημαντικότητας με το οποίο εργαστήκαμε που ισούται με 5% καταλήξαμε στην απόρριψη της μηδενικής υπόθεσης, δηλαδή στην απόρριψη της κανονικότητας των μεταβλητών.

Προχωρώντας στην ανάλυσή μας και για τον έλεγχο των συσχετίσεων και της ανεξαρτησίας των μεταβλητών θα χρησιμοποιηθούν ο συντελεστής συσχέτισης Spearman για τις ποσοτικές μεταβλητές, καθώς είναι μη παραμετρικός έλεγχος και η τιμή Pearson (Chi-Squared) για την ανεξαρτησία των ποιοτικών μεταβλητών. Θα εφαρμοστούν και έλεγχοι της ποσοτικής μεταβλητής Survival Months με τις ποιοτικές μεταβλητές αφού θα χρησιμοποιηθεί στο επόμενο κεφάλαιο σαν την μεταβλητή απόκρισης. Τέτοιοι έλεγχοι θα είναι ο Man - Whitney και Kruskal - Wallis εφόσον δεν υπάρχει κανονικότητα. Πιο συγκεκριμένα θα υπολογιστεί η τιμή p-value των παραπάνω ελέγχων και οι υποθέσεις H_0 , H_1 θα είναι της μορφής:

Μηδενική υπόθεση H_0 : δεν υπάρχει συσχέτιση στις μεταβλητές που μελετάμε

vs

Εναλλακτική υπόθεση H_1 : υπάρχει συσχέτιση στις μεταβλητές που μελετάμε.

4.1. Δεδομένα Καρκίνου Εγκεφάλου

4.1.1. Ανάλυση βάσει της μεταβλητής «Survival Months»

Ξεκινώντας με τα δεδομένα του καρκίνου του εγκεφάλου και τις ποσοτικές μεταβλητές σε σχέση με την εξαρτημένη, Survival Months. Θα υπολογιστεί ο συντελεστής συσχέτισης Spearman καθώς και τα p-value. Η μηδενική υπόθεση θα μας εκφράσει την μη συσχέτιση των μεταβλητών και η εναλλακτική το αντίστροφο. Όπως διαπιστώνουμε και από τον παρακάτω συγκεντρωτικό πίνακα η μεταβλητή που υποδηλώνει τους μήνες επιβίωσης φαίνεται να έχει συσχέτιση με τις άλλες δύο ποσοτικές μεταβλητές. Ουσιαστικά οι τιμές του συντελεστή Spearman κυμαίνονται από το -1 έως το 1. Όσο πιο κοντά είναι η τιμή στο -1 τόσο πιο αρνητικά είναι συσχετισμένες οι μεταβλητές μας. Αντίστροφα για τις τιμές κοντά στο 1.

Μεταβλητή	Τιμή Συντελεστή Spearman	P-Value Ελέγχου
Tumor Size	-0.106108	<2.2e-16
Total number of in situ / malignant tumors for patient	-0.1154633	<2.2e-16

Πίνακας 4.4: Συντελεστές Spearman για τον καρκίνο του Εγκεφάλου

Για τις ποιοτικές μεταβλητές και την ποσοτική Survival Months αφότου έχουν γίνει έλεγχοι κανονικότητας και έχει απορριφθεί η μηδενική υπόθεση, δηλαδή η κανονικότητα των δεδομένων θα συνεχίσουμε με μη παραμετρικό έλεγχο συσχετίσεων. Δύο από τους πιο βασικούς ελέγχους είναι ο Mann – Whitney, όταν η ποιοτική μεταβλητή έχει δυο κατηγορίες, και ο έλεγχος Kruskal Wallis για περισσότερες από δυο κατηγορίες. Εφόσον υλοποιηθούν τα παραπάνω θα παρουσιαστούν και διαγράμματα boxplots με στόχο την γραφική απεικόνιση των συσχετίσεων.

Ξεκινώντας από την μεταβλητή «Φύλο» και τον έλεγχο Mann – Whitney εφόσον έχουμε δυο κατηγορίες παρατηρούμε μεγάλη τιμή p-value με αποτέλεσμα να δεχόμαστε την μηδενική υπόθεση, δηλαδή την μη συσχέτιση των μεταβλητών «Φύλο» και «Μήνες Επιβίωσης». Γεγονός που το βλέπουμε και από το boxplot καθώς οι μέσες τιμές δεν απέχουν.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Sex	49562757	0.8874

Πίνακας 4.5: Έλεγχος συσχέτισης μεταβλητών Survival Months - Sex



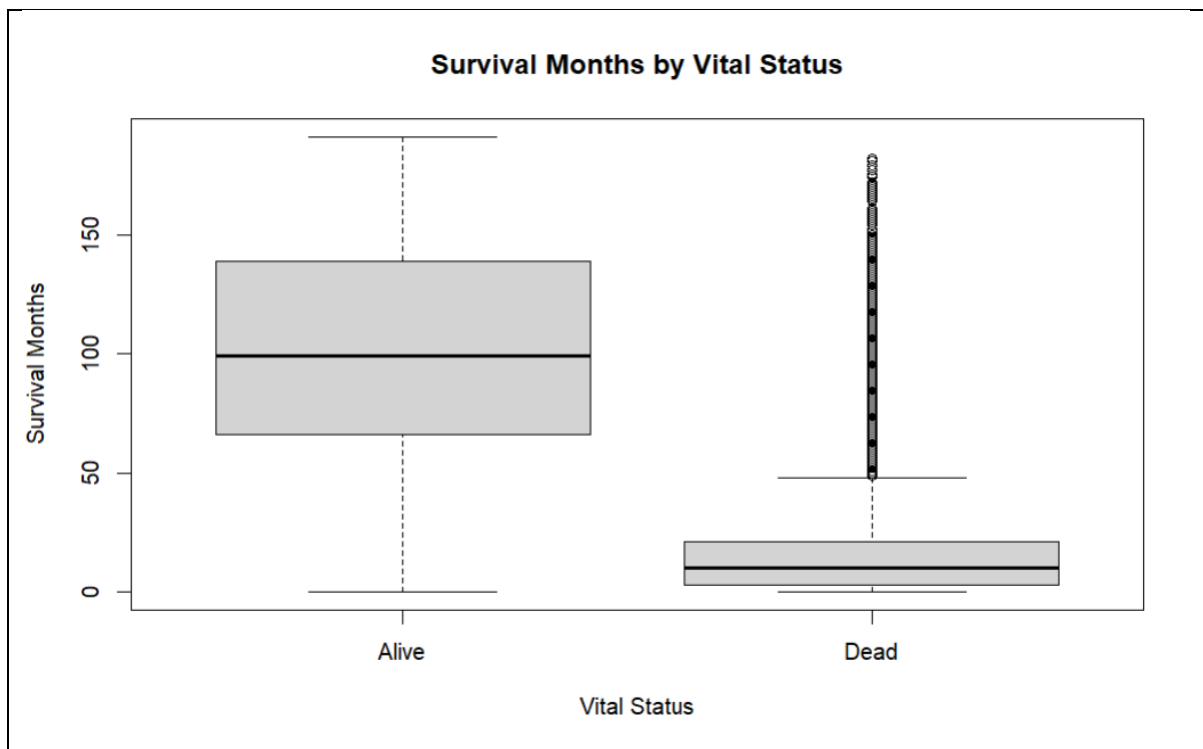
Διάγραμμα 4.1 : Boxplot μεταβλητών Survival Months – Sex

Συνεχίζουμε με τις μεταβλητή Vital Status που δηλώνει την κατάσταση του ασθενή στο τέλος της έρευνας. Να σημειώσουμε εδώ ότι οι εγγραφές με κατηγορία «Alive» θα χρησιμοποιηθούν και σαν λογοκριμένες παρατηρήσεις καθώς δεν γνωρίζουμε αν απεβίωσαν ή απλά έφυγαν από την έρευνα στην στιγμή που αναγράφεται στο πεδίο Survival Months.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Vital Status	4466399	< 2.2e-16

Πίνακας 4.6: Έλεγχος συσχέτισης μεταβλητών Survival Months – Vital Status

Όπως φαίνεται και από την τιμή του p-value αλλά και από το boxplot που ακολουθεί καταλήγουμε ότι οι παραπάνω μεταβλητές είναι συσχετισμένες.

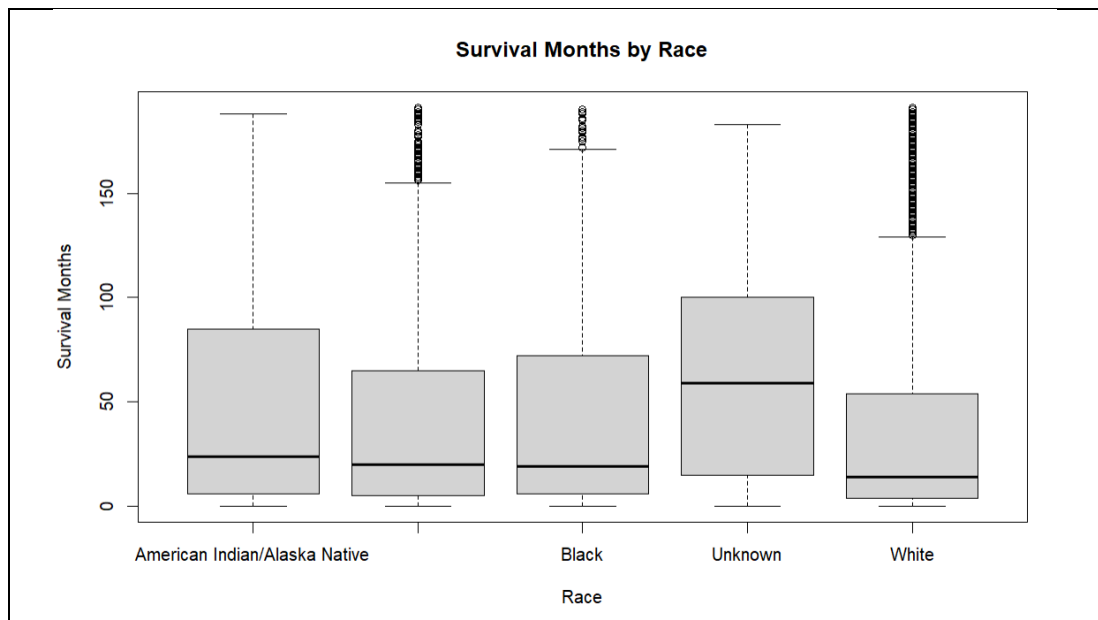


Διάγραμμα 4.2 : Boxplot μεταβλητών Survival Months – Vital Status

Για τις μεταβλητές με παραπάνω από δυο κατηγορίες θα εκτελεστεί ο έλεγχος Kruskal – Wallis, έτσι για το πεδίο «Race» παρατηρούμε πολύ μικρό p-value άρα έχουμε συσχέτιση στην φυλή και τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Race	85.641	< 2.2e-16

Πίνακας 4.7: Έλεγχος συσχέτισης μεταβλητών Survival Months – Race



Διάγραμμα 4.3 : Boxplot μεταβλητών Survival Months – Race

Η ηλικία, όπως βλέπουμε από το p-value του ελέγχου είναι συσχετισμένη με τους μήνες επιβίωσης.

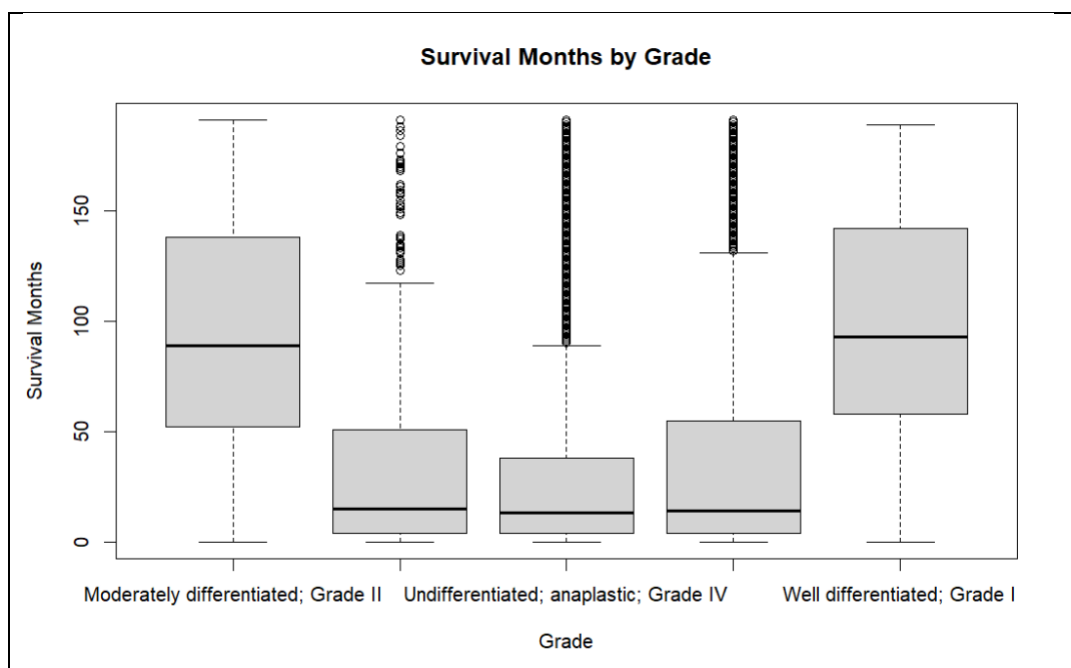
Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Age	7796.6	< 2.2e-16

Πίνακας 4.8: Έλεγχος συσχέτισης μεταβλητών Survival Months – Age

Ο έλεγχος που αφορά τον βαθμό του καρκίνου και τους μήνες επιβίωσης μαρτυρά την συσχέτισης τόσο από την τιμή του p-value όσο και από το boxplot.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Grade	1097.3	< 2.2e-16

Πίνακας 4.9: Έλεγχος συσχέτισης μεταβλητών Survival Months – Grade

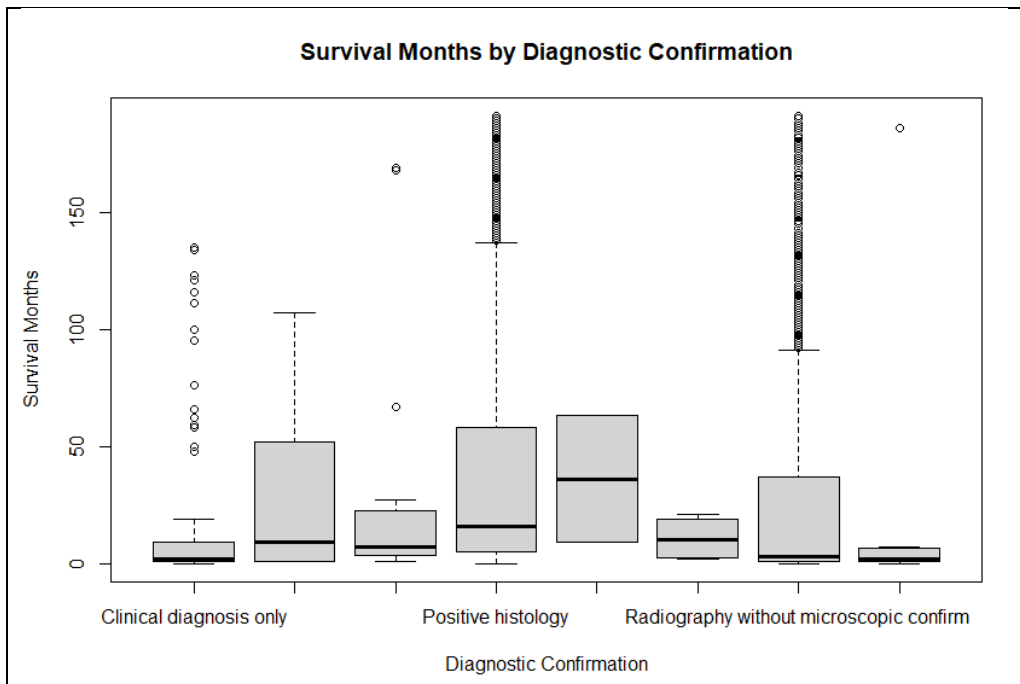


Διάγραμμα 4.4 : Boxplot μεταβλητών Survival Months – Grade

Τα ίδια αποτελέσματα παίρνουμε και για τις μεταβλητές Μήνες επιβίωσης και τον τρόπο διάγνωσης του καρκίνου, δηλαδή ότι είναι συσχετισμένες.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Diagnostic Confirmation	1097.3	< 2.2e-16

Πίνακας 4.10: Έλεγχος συσχέτισης μεταβλητών Survival Months – Diagnostic Confirmation

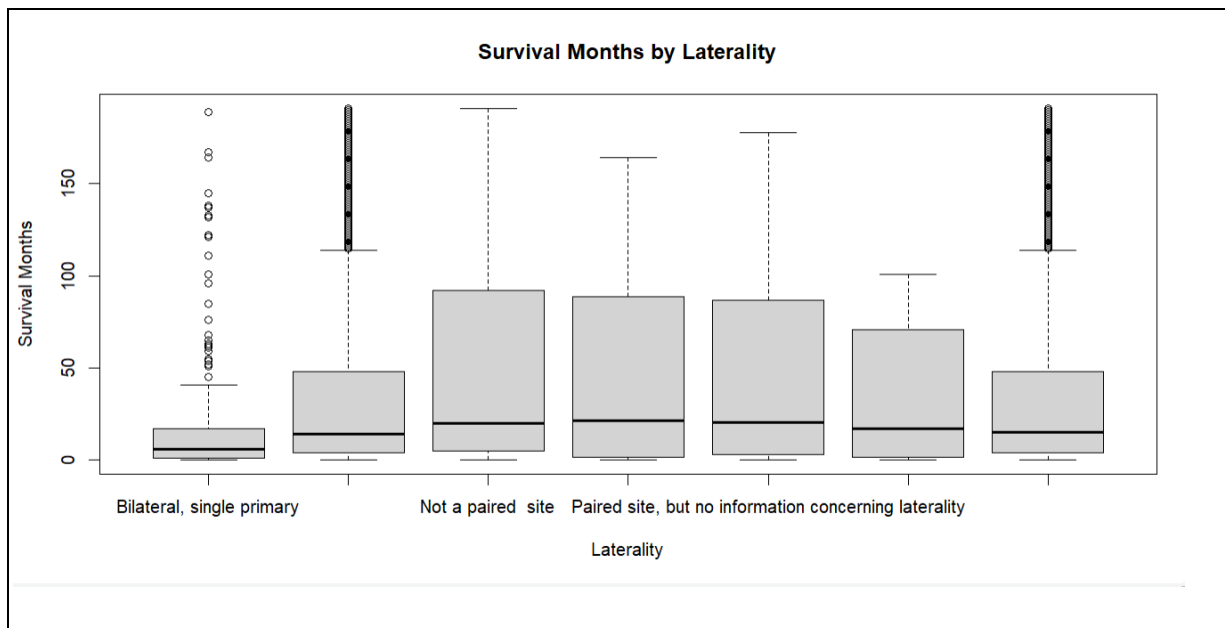


Διάγραμμα 4.5: Boxplot μεταβλητών Survival Months – Diagnostic Confirmation

Ομοίως και για την τελευταία ποιοτική μεταβλητή του καρκίνου του εγκεφάλου διαπιστώνεται συσχέτιση μεταξύ της πλευρικότητας και τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Laterality	226.36	< 2.2e-16

Πίνακας 4.11: Έλεγχος συσχέτισης μεταβλητών Survival Months – Laterality



Διάγραμμα 4.6 : Boxplot μεταβλητών Survival Months – Laterality

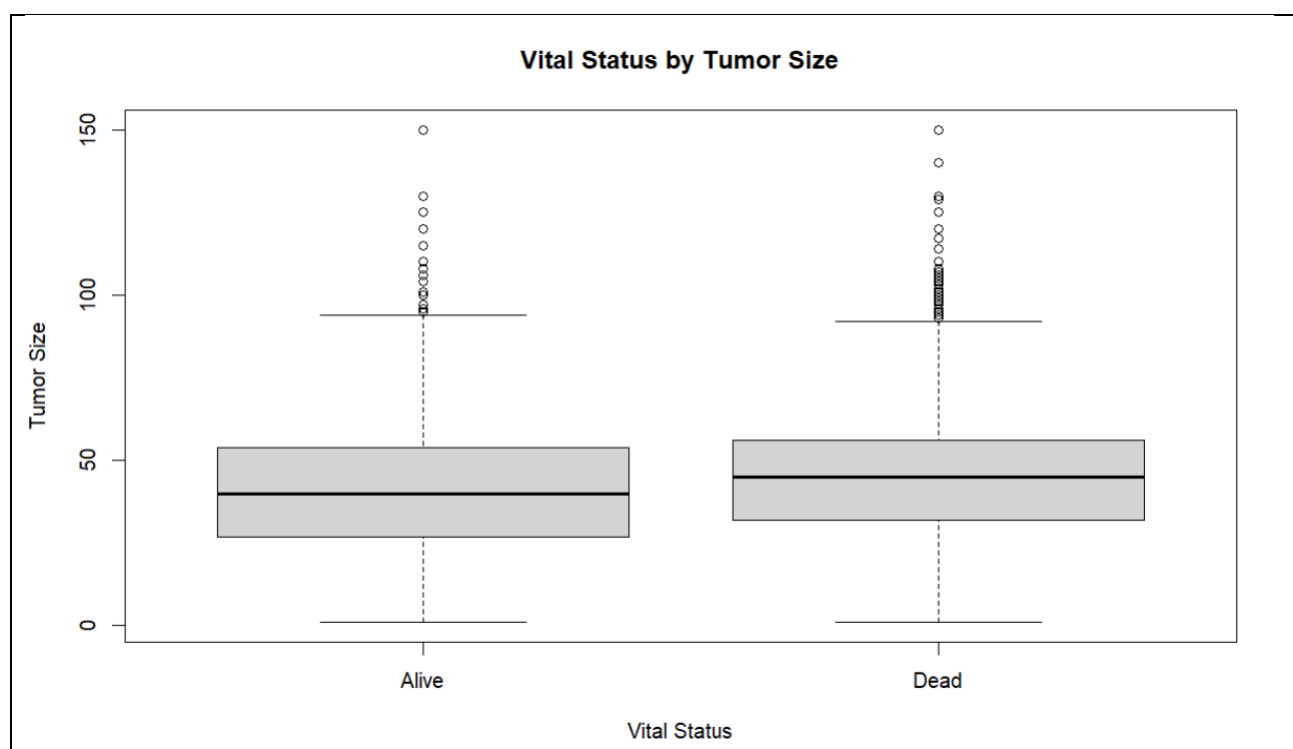
4.1.2. Ανάλυση βάσει της μεταβλητής «Vital Status»

Εξίσου σημαντική μεταβλητή του συνόλου δεδομένων του καρκίνου του εγκεφάλου είναι η δίτιμη Vital Status, η οποία χωρίζει τους ασθενείς σε αυτούς που απεβίωσαν και επιβίωσαν. Επομένως θεωρήθηκε ορθό να γίνει μια ανάλυση όπως προηγουμένως με την διαφορά ότι θα χρησιμοποιηθεί ο έλεγχος χ^2 του Pearson για τις ποιοτικές μεταβλητές και την «Vital Status».

Αρχικά θα δείξουμε την συσχέτιση της μεταβλητής που προαναφέραμε με τις ποσοτικές του δείγματος, δηλαδή «Tumor Size», «Survival Months» και «Total number of in situ / malignant tumors for patient». Σύμφωνα με τον Mann – Whitney έλεγχο για τις μεταβλητές έχουμε συσχέτιση στην θνησιμότητα και το μέγεθος του όγκου, γεγονός που φαίνεται και από το επόμενο boxplot.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Tumor Size	41785930	< 2.2e-16

Πίνακας 4.12: Έλεγχος συσχέτισης μεταβλητών Vital Status – Tumor Size



Διάγραμμα 4.7 : Boxplot μεταβλητών Vital Status – Tumor Size

Για τους μήνες επιβίωσης και την κατάσταση του ασθενή έχει γίνει ο σχετικός έλεγχος στην προηγούμενη ενότητα του κεφαλαίου. Στην συνέχεια για το συνολικό αριθμό των όγκων και το Vital Status διακρίνουμε εξίσου μια συσχέτιση.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Total number of in situ / malignant tumors for patient	41062227	< 2.2e-16

Πίνακας 4.13: Έλεγχος συσχέτισης μεταβλητών Vital Status – Number of tumors

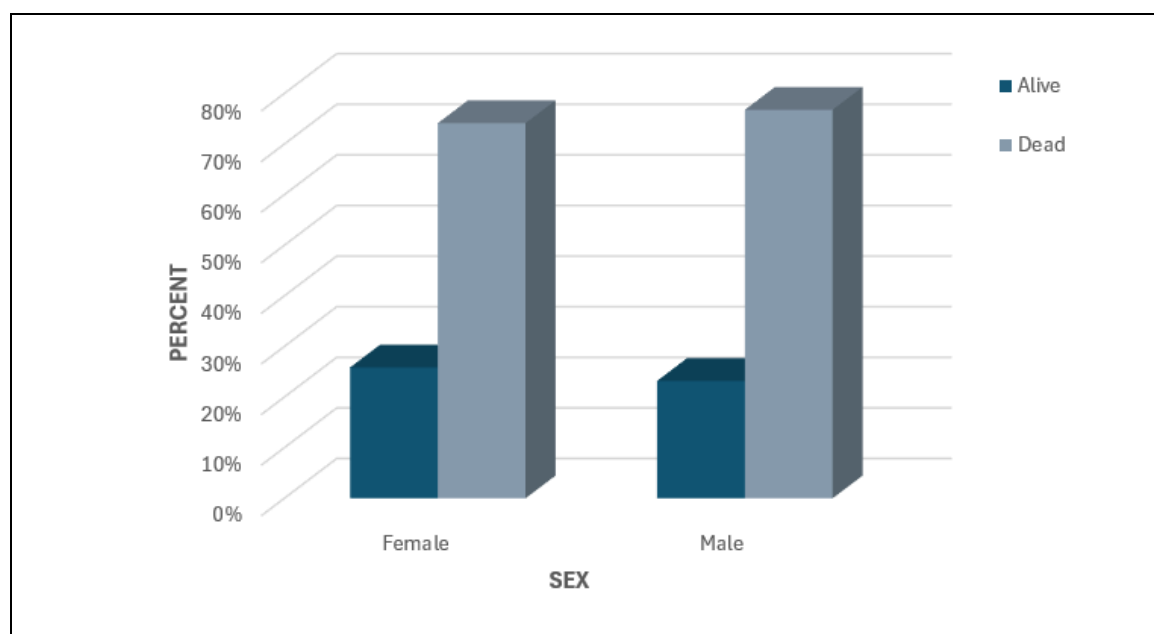
Περνώντας στις ποιοτικές μεταβλητές συναρτήσει της «Vital Status» θα ήταν ορθό να γίνει ο έλεγχος συσχετίσεων σε συνδυασμό με την γραφική απεικόνιση, η οποία θα αποτελείται από ραβδόγραμμα και το ποσοστό επιβίωσης ή θανάτου για κάθε μεταβλητή. Για το φύλο και την μεταβλητή Vital Status παρατηρείται σχετικά ίση θνησιμότητα για τους άντρες και τις γυναίκες, σχεδόν στο 75%.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Sex	18.643	1.577e-05

Πίνακας 4.14: Έλεγχος συσχέτισης μεταβλητών Vital Status – Sex

Sex	Επιβίωσαν	Απεβίωσαν
Γυναίκες	26%	74%
Άντρες	23%	77%
Σύνολο	24%	76%

Πίνακας 4.15: Ποσοστά θνησιμότητας της μεταβλητής Sex



Διάγραμμα 4.8: Ραβδόγραμμα μεταβλητών Vital Status – Sex

Αναπαριστώντας την ηλικίας συναρτήσει της κατάστασης του ασθενή καταλήγουμε στο συμπέρασμα ότι σε ποσοστό 84% και πάνω μετά τα 50 απεβίωσαν.

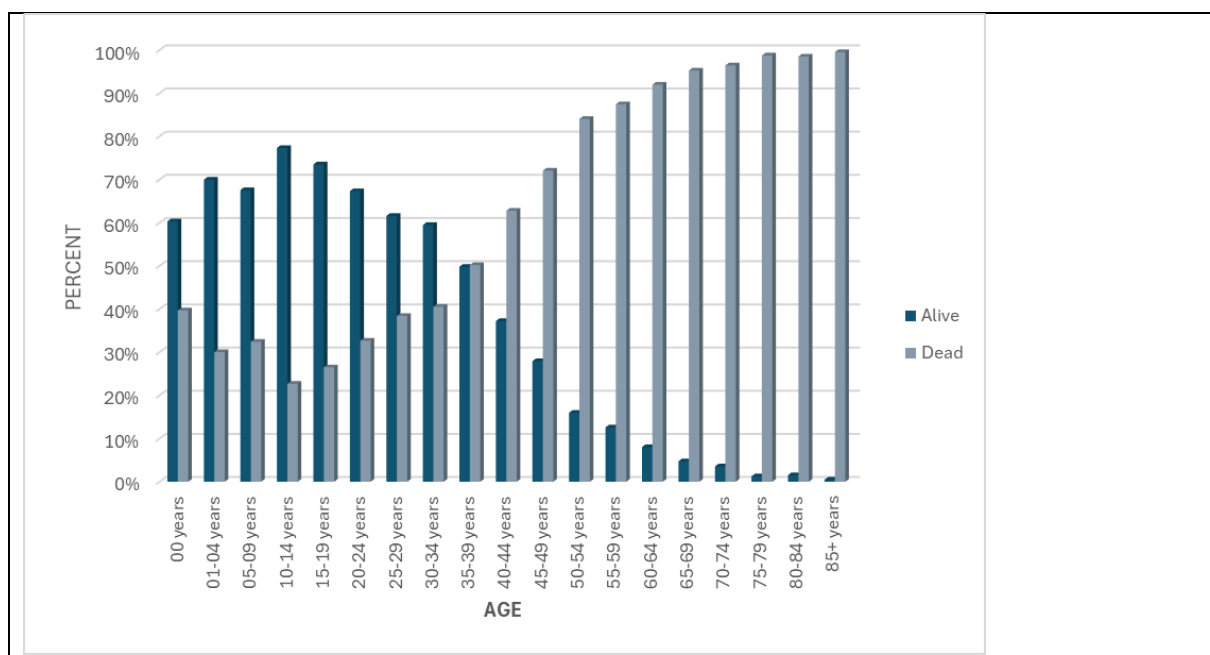
Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Age	7065.3	<2.2e-16

Πίνακας 4.16: Έλεγχος συσχέτισης μεταβλητών Vital Status – Age

Age	Επιβίωσαν	Απεβίωσαν
00 years	60%	40%
01-04 years	70%	30%
05-09 years	68%	32%
10-14 years	77%	23%

15-19 years	74%	26%
20-24 years	67%	33%
25-29 years	62%	38%
30-34 years	59%	41%
35-39 years	50%	50%
40-44 years	37%	63%
45-49 years	28%	72%
50-54 years	16%	84%
55-59 years	13%	87%
60-64 years	8%	92%
65-69 years	5%	95%
70-74 years	4%	96%
75-79 years	1%	99%
80-84 years	2%	98%
Σύνολο	24%	76%

Πίνακας 4.17: Ποσοστά θνησιμότητας της μεταβλητής Age



Διάγραμμα 4.9: Ραβδόγραμμα μεταβλητών Vital Status – Age

Συνεχίζοντας με τα πεδία Race και Vital Status, όπως βλέπουμε και στο παρακάτω σχήμα εκείνοι δεν τα κατάφεραν στο τέλος στην πλειοψηφία ήταν Λευκοί και Αμερικάνοι.

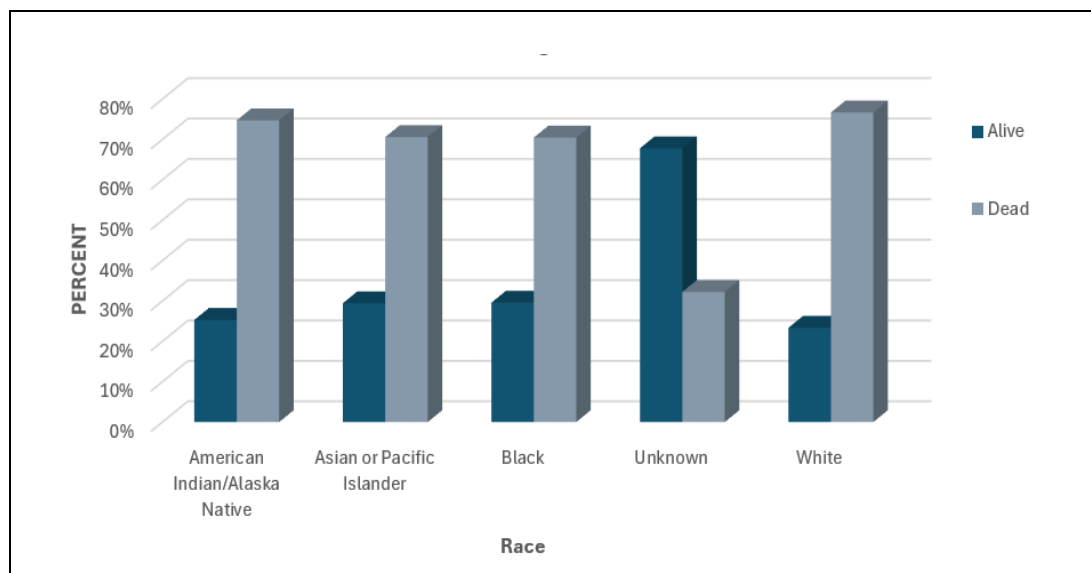
Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Race	138.9	<2.2e-16

Πίνακας 4.18: Έλεγχος συσχέτισης μεταβλητών Vital Status – Race

Race	Επιβίωσαν	Απεβίωσαν
Αμερικανοί	25%	75%
Ασιάτες	29%	71%

Έγχρωμοι	30%	70%
Δεν γνωρίζουμε	68%	32%
Λευκοί	23%	77%
Σύνολο	24%	76%

Πίνακας 4.19: Ποσοστά θνησιμότητας της μεταβλητής Race



Διάγραμμα 4.10: Ραβδόγραμμα μεταβλητών Vital Status – Race

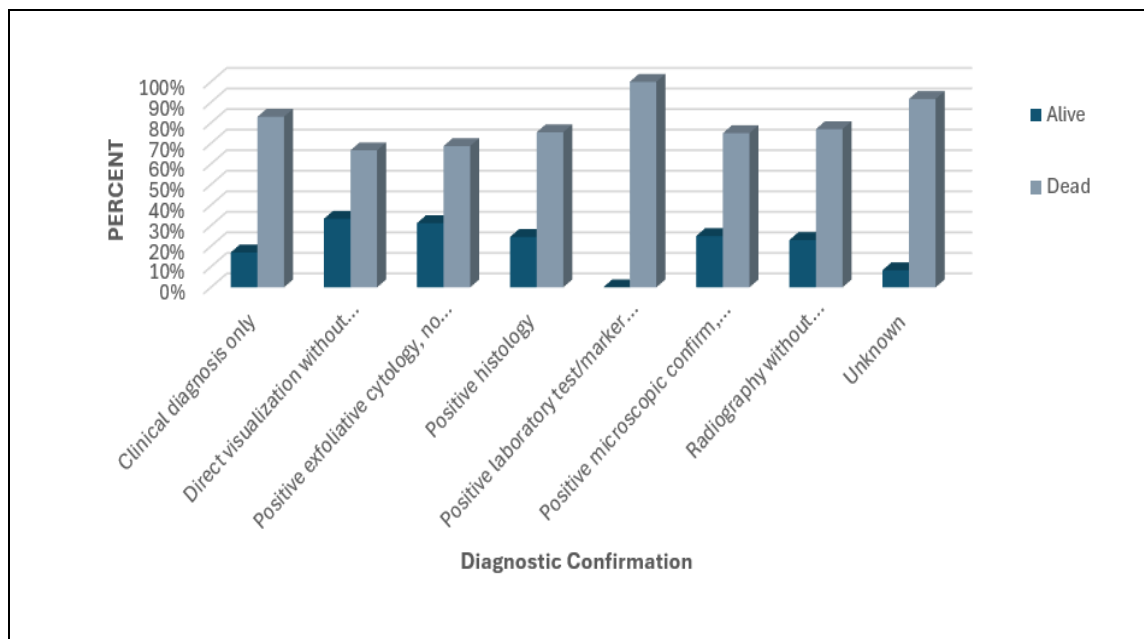
Για τον διαγνωστικό έλεγχο με το οποίο διαγνώστηκε ο ασθενής για το καρκίνο καταλαβαίνουμε ότι το μεγαλύτερο ποσοστό θνησιμότητας το βρίσκουμε σε εκείνους με την θετικής ιστολογική εξέταση.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Diagnostic Confirmation	7.8973	0.3417

Πίνακας 4.20: Έλεγχος συσχέτισης μεταβλητών Vital Status – Race

Diagnostic Confirmation	Επιβίωσαν	Απεβίωσαν
Κλινική Διάγνωση	17%	83%
Απευθείας Οπτικοποίηση	33%	67%
Θετική Αποφολιδωτική Κυτταρολογία	31%	69%
Θετική Ιστολογία	25%	75%
Θετική Εργαστηριακή Δοκιμή	0%	100%
Θετική Μικροσκοπική Επιβεβαίωση	25%	75%
Ακτινογραφία	23%	77%
Δεν Γνωρίζουμε	8%	92%
Σύνολο	24%	76%

Πίνακας 4.21: Ποσοστά θνησιμότητας της μεταβλητής Diagnostic Confirmation



Διάγραμμα 4.11 : Ραβδόγραμμα μεταβλητών Vital Status – Diagnostic Confirmation

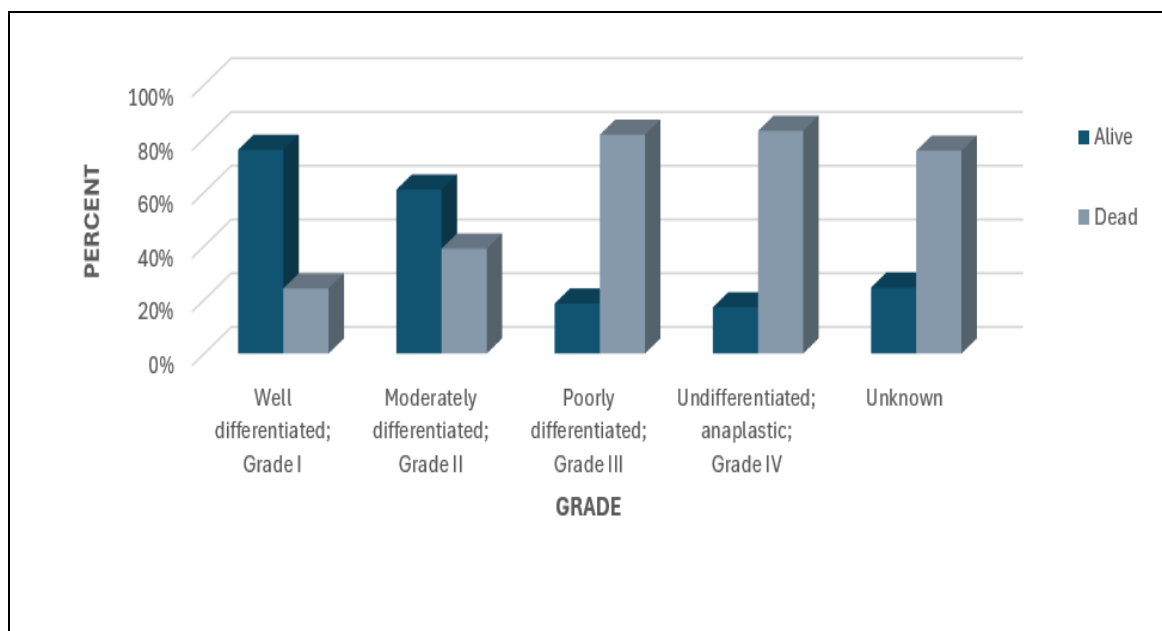
Από τον βαθμό του καρκίνου που είχαν οι ασθενείς του δείγματος όπως ήταν αναμενόμενο μικρότερο ποσοστό θνησιμότητας έχουν εκείνοι με τον μικρότερο βαθμό καρκίνου και αυξάνεται όσο αυξάνεται ο βαθμός καρκίνου.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Grade	1113.1	<2.2e-16

Πίνακας 4.22: Έλεγχος συσχέτισης μεταβλητών Vital Status – Grade

Grade	Επιβίωσαν	Απεβίωσαν
Βαθμός I	76%	24%
Βαθμός II	61%	39%
Βαθμός III	19%	81%
Βαθμός IV	17%	83%
Άγνωστο	25%	75%
Σύνολο	24%	76%

Πίνακας 4.23: Ποσοστά θνησιμότητας της μεταβλητής Grade



Διάγραμμα 4.12: Ραβδόγραμμα μεταβλητών Vital Status – Grade

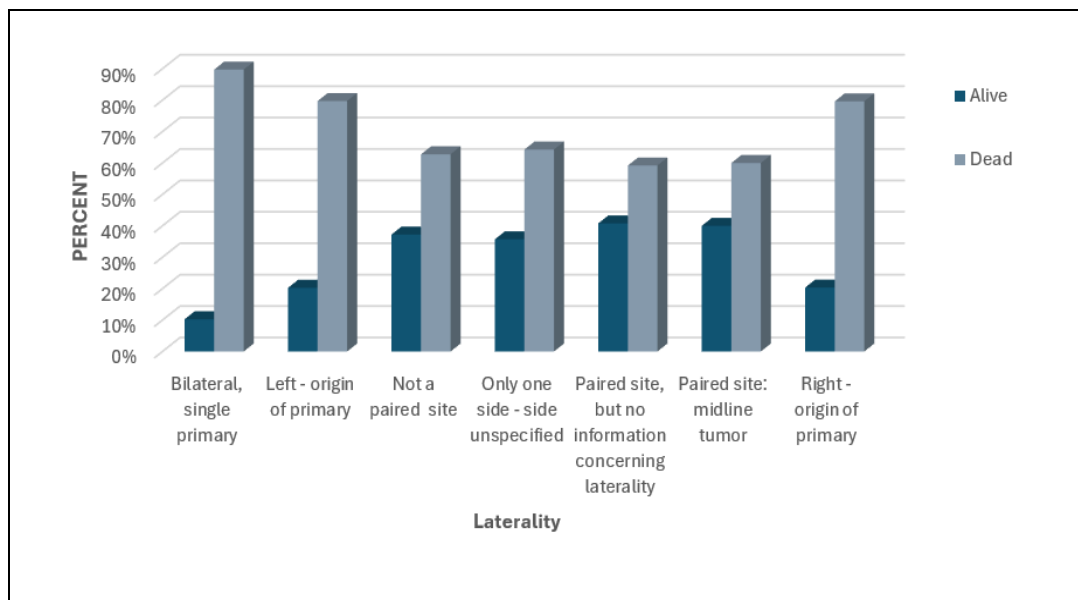
Στην συνέχεια παρουσιάζεται και ο πίνακας για την θνησιμότητα κάθε κατηγορίας της μεταβλητής Laterality όπου και βλέπουμε πως το μεγαλύτερο ποσοστό το έχει η κατηγορία «Διμερής Πλευρικότητα».

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Laterality	606.05	<2.2e-16

Πίνακας 4.24: Έλεγχος συσχέτισης μεταβλητών Vital Status – Laterality

Laterality	Επιβίωσαν	Απεβίωσαν
Διμερής Πλευρικότητα	10%	90%
Αριστερή Προέλευση	20%	80%
Μη Ζευγοποιημένος	37%	63%
Απροσδιόριστη	36%	64%
Ζεύγος (χωρίς ξεκάθαρη πλευρικότητα)	41%	59%
Μέσης Γραμμής Όγκος	40%	60%
Δεξιά Προέλευση	20%	80%
Σύνολο	24%	76%

Πίνακας 4.25: Ποσοστά θνησιμότητας της μεταβλητής Laterality



Διάγραμμα 4.13: Ραβδόγραμμα μεταβλητών Vital Status – Laterality

4.2. Δεδομένα Καρκίνου Μαστού

4.2.1. Ανάλυση βάσει της μεταβλητής «Survival Months»

Σε αυτή την υποενότητα θα εργαστούμε με τον ίδιο τρόπο για την ποσοτική μεταβλητή «Survival Months» και τις υπόλοιπες. Αρχίζουμε με τις ποσοτικές μεταβλητές και τον έλεγχο του Spearman.

Μεταβλητή	Τιμή Συντελεστή Spearman	P-Value Ελέγχου
Tumor Size	-0.08196594	9.575e-10
Regional Nodes Examined	0.1152165	<2.2e-16
Regional Nodes Positive	-0.02788634	0.03773

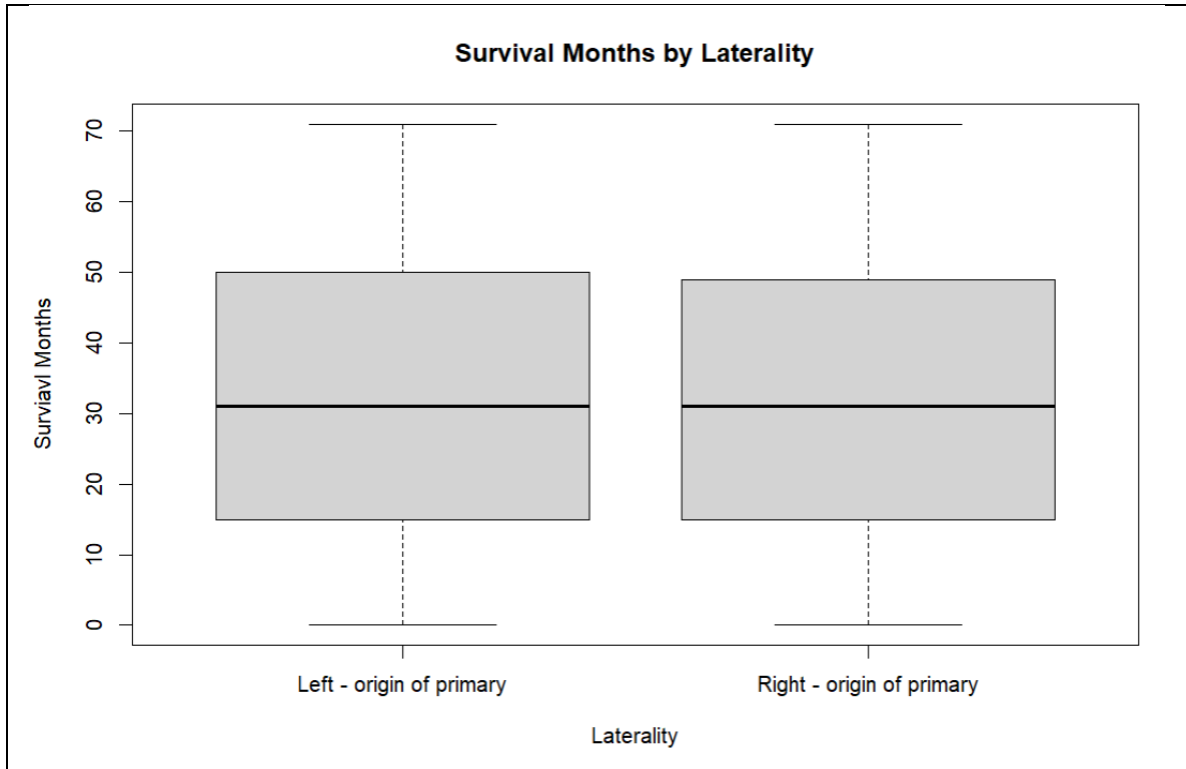
Πίνακας 4.26: Συντελεστές Spearman για τον καρκίνο του Μαστού

Αναφορικά με τις παραπάνω μεταβλητές και τους μήνες επιβίωσης παρατηρείται συσχέτιση καθώς οι τιμές των p-value είναι μικρότερες από το επίπεδο σημαντικότητας που έχουμε ορίσει για τους ελέγχους, το 5%.

Συνεχίζοντας με τις ποιοτικές μεταβλητές του dataset οι οποίες είναι Age, Race, Grade, Marital Status, Laterality, Stage A, ER Status, PR Status, Breast SubType, Stage N και Stage T θα εκτελεστούν οι έλεγχοι Mann – Whitney για τις μεταβλητές με δυο κατηγορίες και Kruskal – Wallis για παραπάνω. Επιπροσθέτως θα αναπαρασταθούν και boxplot. Για τις δίτιμες μεταβλητές τα αποτελέσματα παρουσιάζονται αναλυτικά για κάθε μεταβλητή στους παρακάτω πίνακες. Ξεκινώντας με την μεταβλητή της πλευρικότητας του καρκίνου του μαστού φαίνεται να μην έχει σχέση με τους μήνες επιβίωσης καθώς το p-value ισούται με 0.65 > 5%.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Laterality	3877727	0.6533

Πίνακας 4.27: Έλεγχος συσχέτισης μεταβλητών Survival Months – Laterality

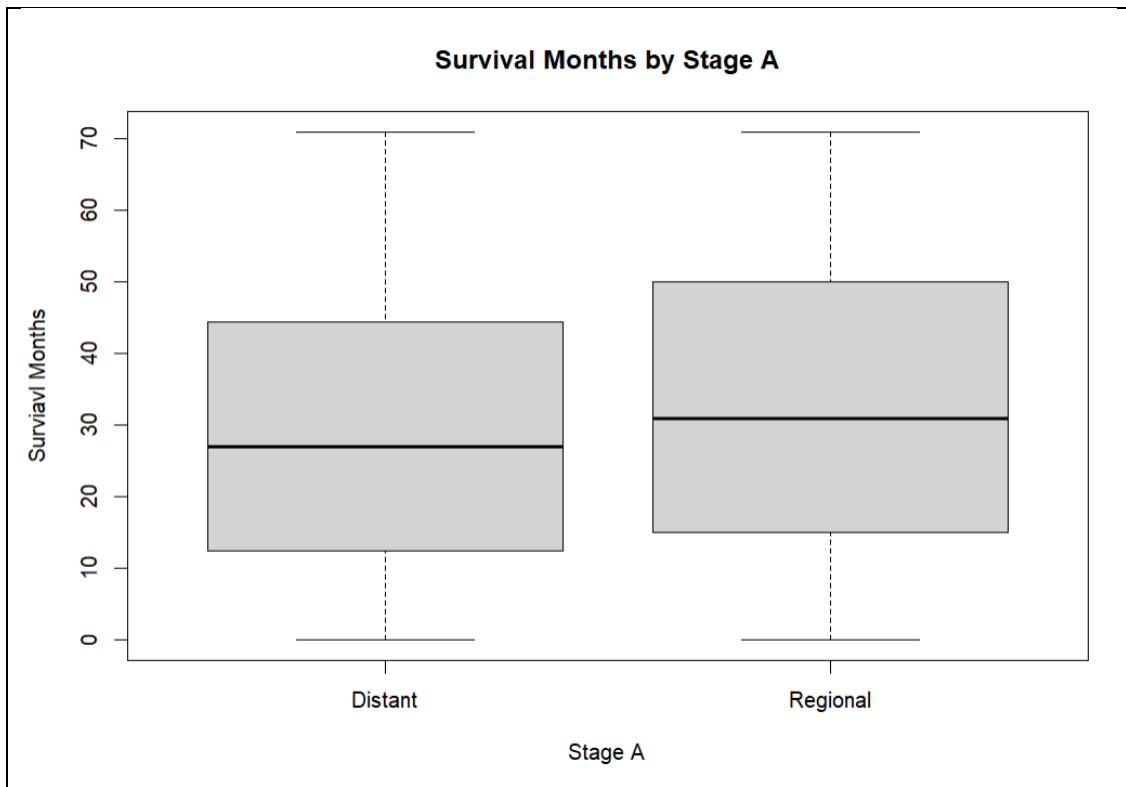


Διάγραμμα 4.14: Boxplot μεταβλητών Survival Months – Laterality

Η έκταση του καρκίνου, δηλαδή αν είναι τοπικός ή μακρινός, που εκφράζεται από το Stage A φαίνεται να έχει συσχέτιση με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Stage A	784808	0.002713

Πίνακας 4.28: Έλεγχος συσχέτισης μεταβλητών Survival Months – Stage A

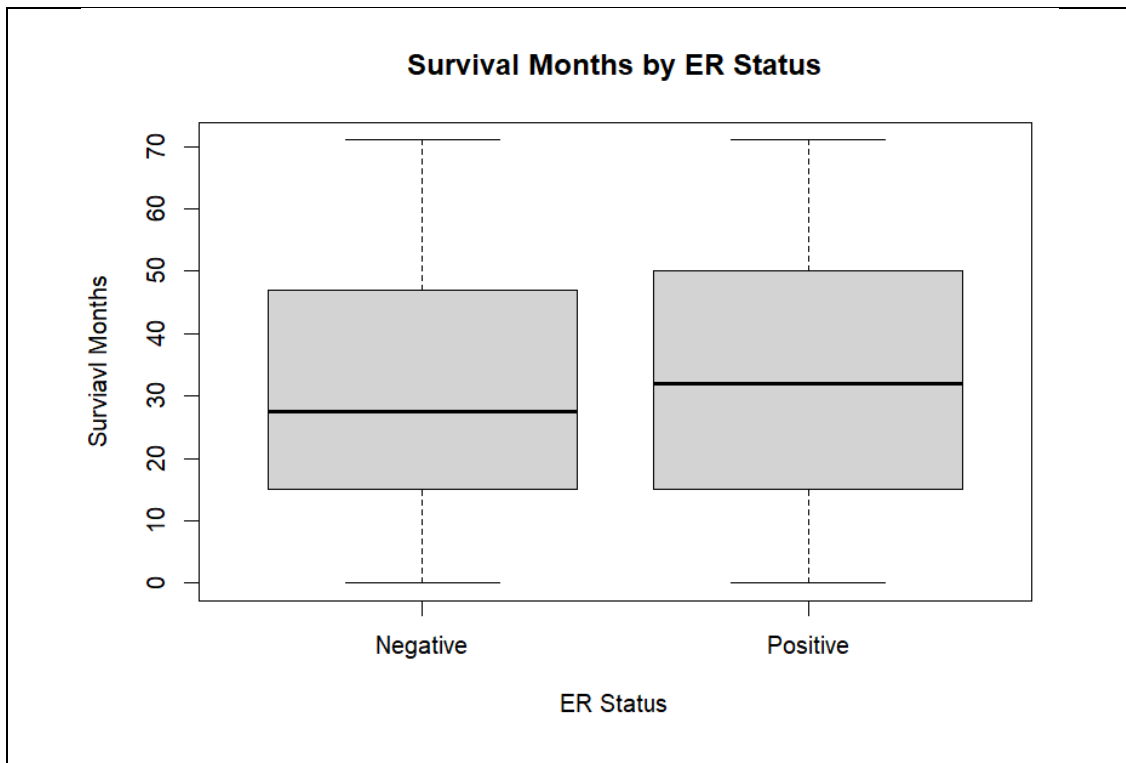


Διάγραμμα 4.15 : Boxplot μεταβλητών Survival Months – Stage A

Ο υποδοχέας οιστρογόνου μετά τον έλεγχο καταλαβαίνουμε ότι σχετίζεται με τους μήνες επιβίωσης της γυναίκας.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
ER Status	2221417	0.006267

Πίνακας 4.29: Έλεγχος συσχέτισης μεταβλητών Survival Months – ER Status

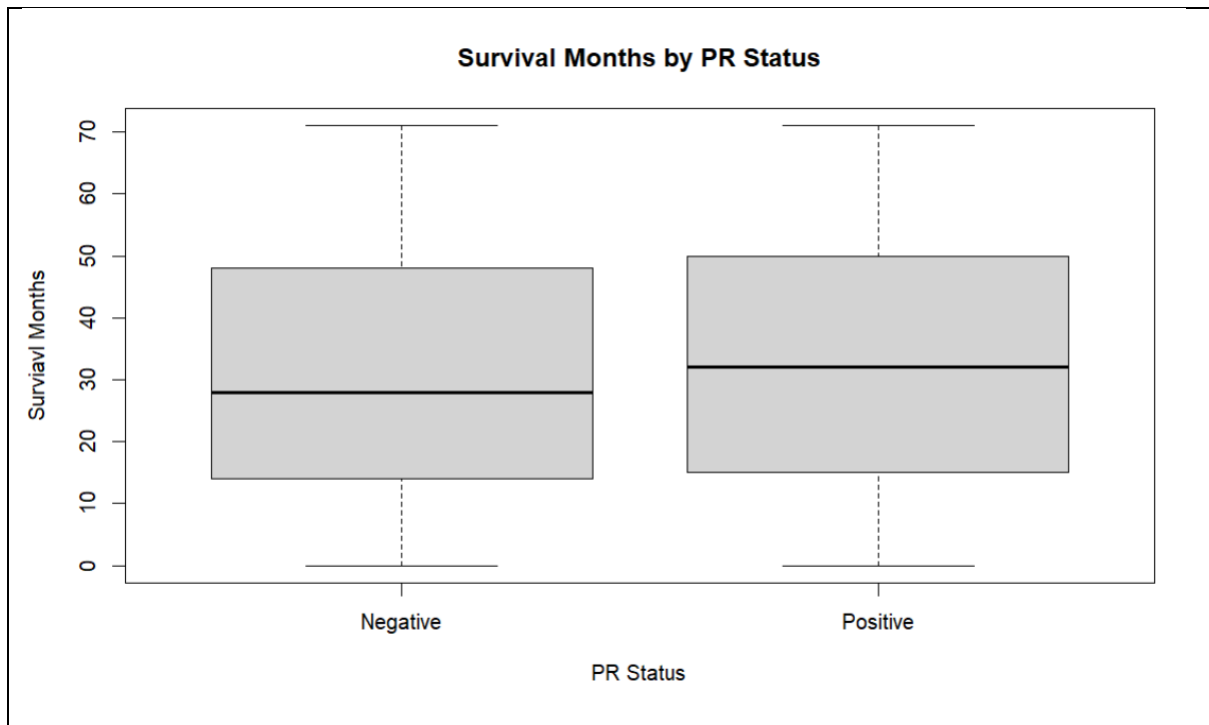


Διάγραμμα 4.16: Boxplot μεταβλητών Survival Months – ER Status

Για την μεταβλητή που υποδηλώνει για την υποδοχή ή όχι της προγεστερόνης ομοίως παρατηρούμε συσχέτιση με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
PR Status	3046563	0.002667

Πίνακας 4.30: Έλεγχος συσχέτισης μεταβλητών Survival Months – PR Status



Διάγραμμα 4.17: Boxplot μεταβλητών Survival Months – PR Status

Για τις ποιοτικές μεταβλητές με πάνω από 2 κατηγορίες θα υλοποιηθεί ο έλεγχος συσχέτισης Kruskal - Wallis. Όσον αφορά στην ηλικία των ασθενών και τους μήνες επιβίωσης, όπως το περιμέναμε υπάρχει συσχέτιση.

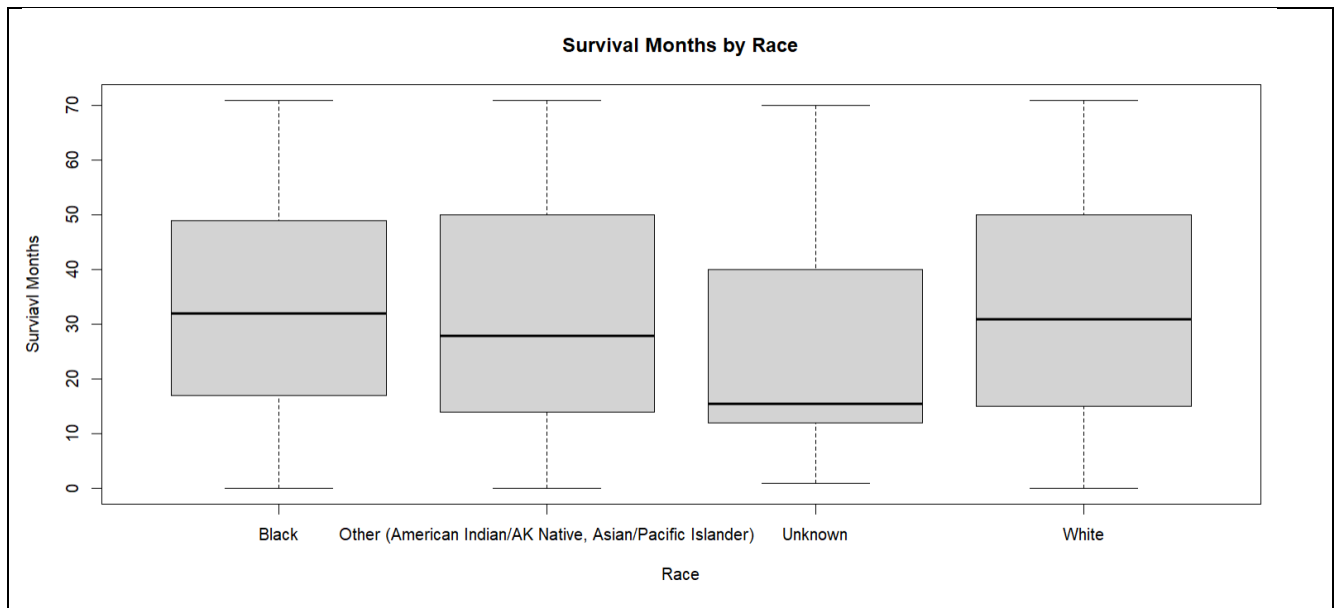
Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Age	60.135	4.97e-08

Πίνακας 4.31: Έλεγχος συσχέτισης μεταβλητών Survival Months – Age

Αντιθέτως, με την μεταβλητή Race ο έλεγχος μαρτυρά την μη συσχέτιση της με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Race	3.8188	0.2817

Πίνακας 4.32: Έλεγχος συσχέτισης μεταβλητών Survival Months – Race

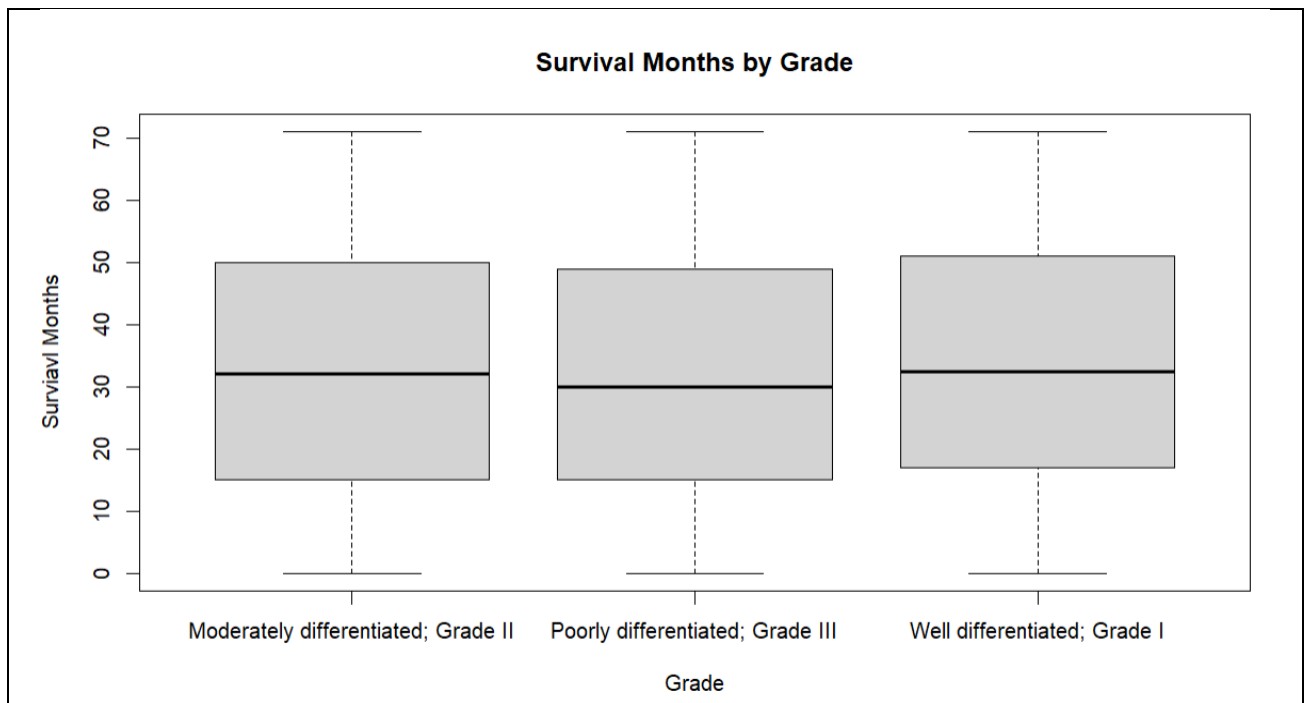


Διάγραμμα 4.18: Boxplot μεταβλητών Survival Months – Race

Για τον βαθμό του καρκίνου η συσχέτιση φαίνεται οριακή καθώς το p-value είναι λίγο μικρότερο από το 5%.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Grade	6.3576	0.04164

Πίνακας 4.33: Έλεγχος συσχέτισης μεταβλητών Survival Months – Grade

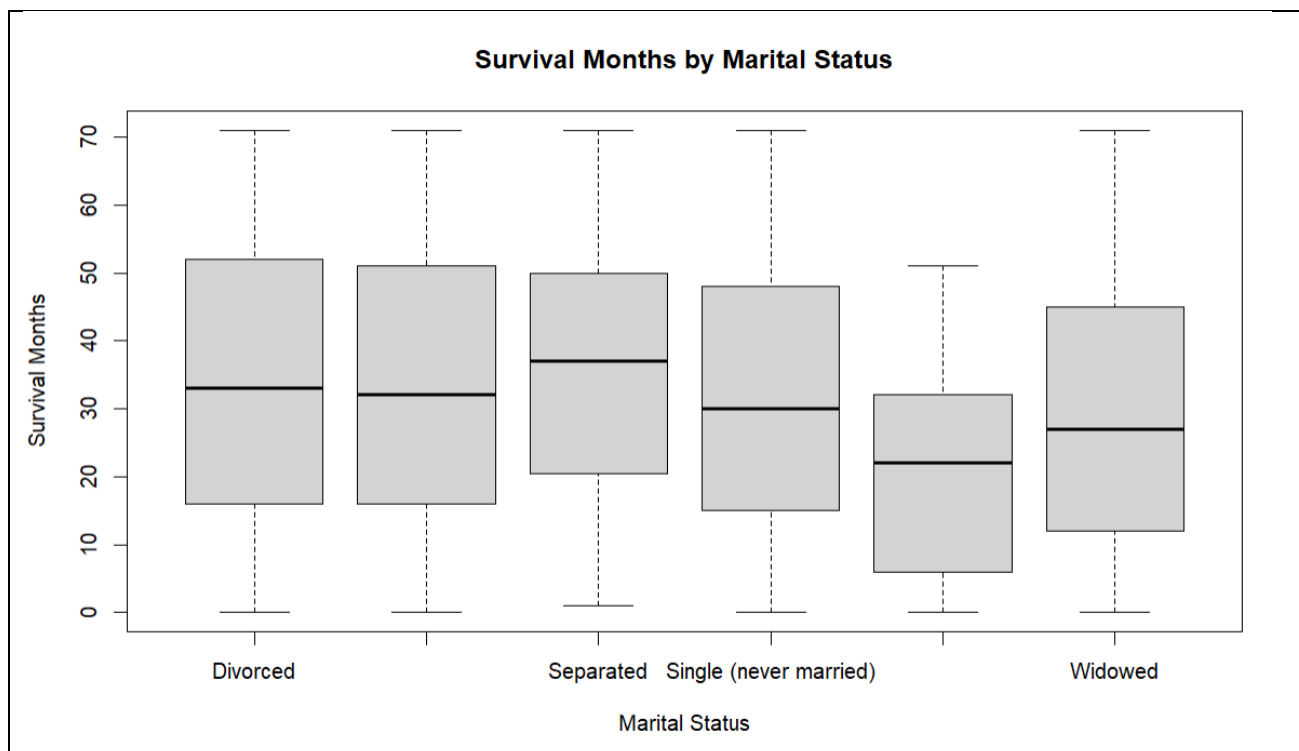


Διάγραμμα 4.19 : Boxplot μεταβλητών Survival Months – Grade

Η οικογενειακή κατάσταση των γυναικών δείχνει συσχέτιση με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Marital Status	40.475	1.198e-07

Πίνακας 4.34: Έλεγχος συσχέτισης μεταβλητών Survival Months – Marital Status

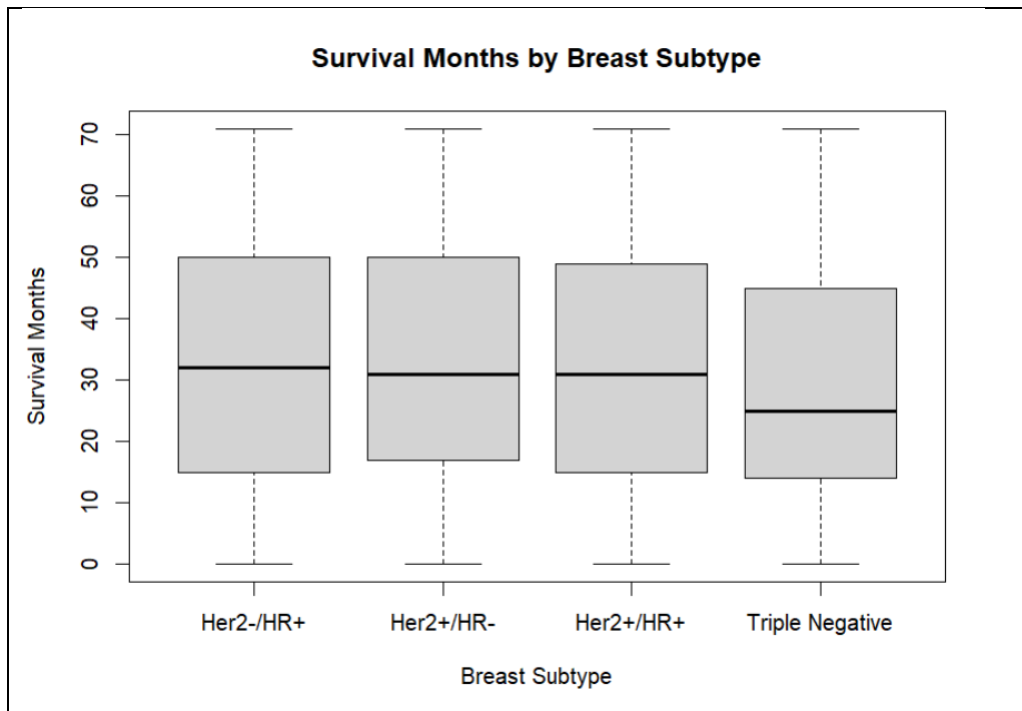


Διάγραμμα 4.20 : Boxplot μεταβλητών Survival Months – Marital Status

Ο δευτερεύον τύπος του μαστού φαίνεται να σχετίζεται με την μεταβλητή που αναφέρει τους μήνες επιβίωσης των γυναικών ασθενών.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Breast SubType	15.611	0.001363

Πίνακας 4.35: Έλεγχος συσχέτισης μεταβλητών Survival Months – Breast SubType

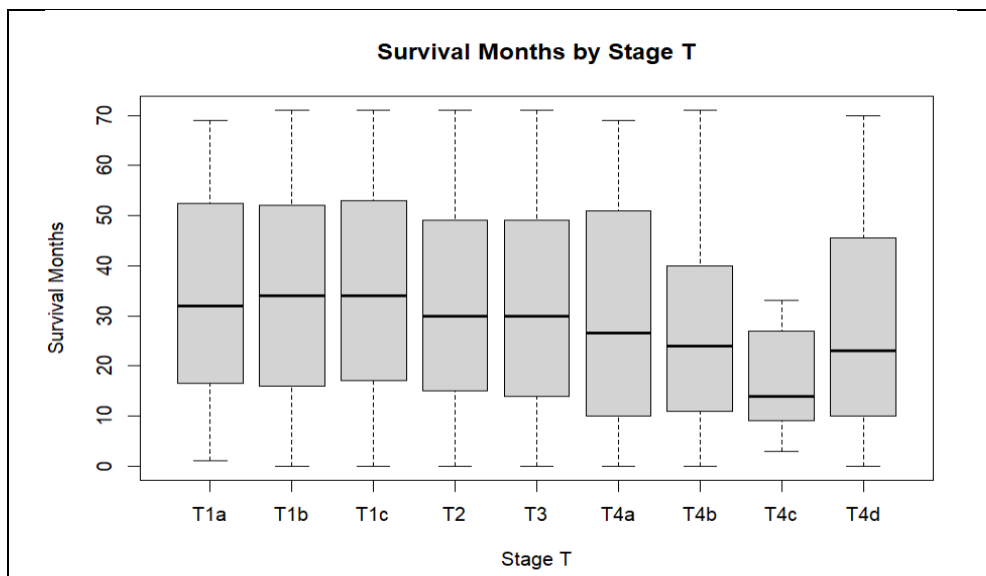


Διάγραμμα 4.21: Boxplot μεταβλητών Survival Months – Breast SubType

Η μεταβλητή που περιγράφει το μέγεθος και την θέση του όγκου, Stage T, από την στιγμή που ο έλεγχος δίνει πολύ χαμηλό p-value δεν μπορούμε παρά να δεχτούμε την εναλλακτική υπόθεση και την συσχέτισή της με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Stage T	54.409	5.751e-09

Πίνακας 4.36: Έλεγχος συσχέτισης μεταβλητών Survival Months – Stage T

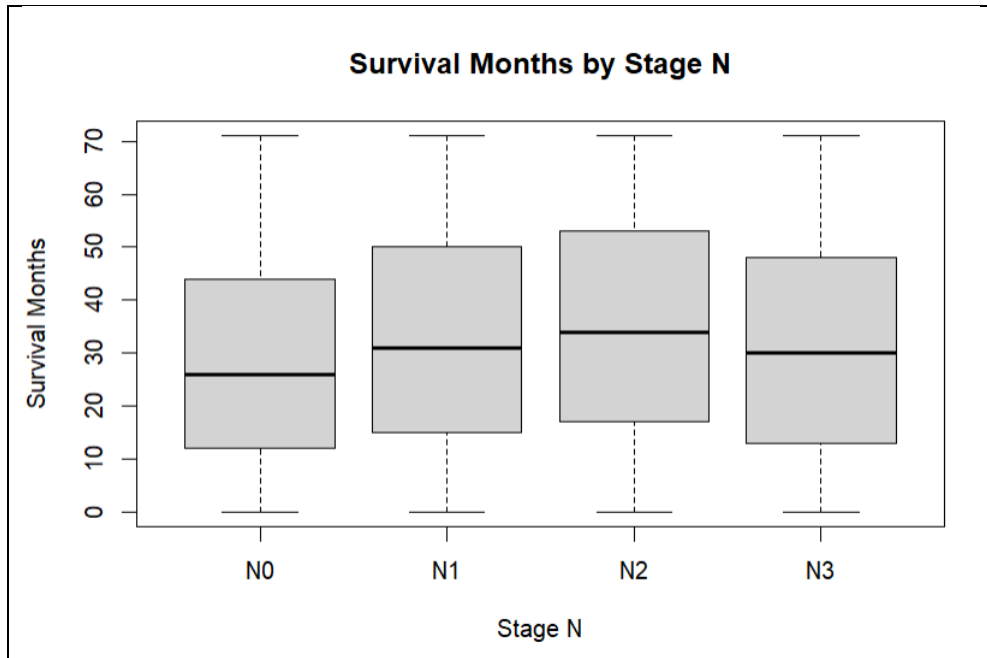


Διάγραμμα 4.22: Boxplot μεταβλητών Survival Months – Stage T

Τέλος, η μεταβλητή που καταγράφει τους «μολυσμένους» λεμφαδένες δείχνει εξίσου συσχέτιση με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Stage N	17.571	0.0005391

Πίνακας 4.37: Έλεγχος συσχέτισης μεταβλητών Survival Months – Stage N



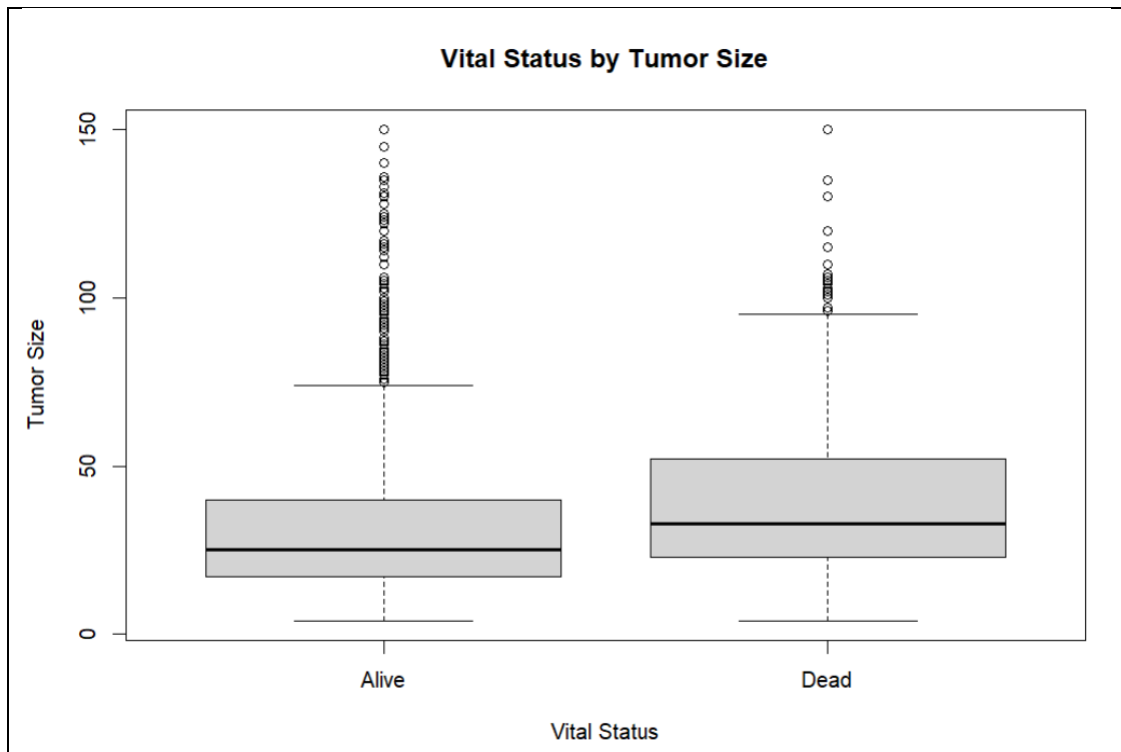
Διάγραμμα 4.23: Boxplot μεταβλητών Survival Months – Stage N

4.2.2 Ανάλυση βάσει της μεταβλητής «Vital Status»

Συνεχίζουμε την ανάλυση για του ελέγχους συσχετίσεων και ανεξαρτησίας με κύρια μεταβλητή την ποιοτική «Vital Status». Για τις ποσοτικές μεταβλητές θα εκτελεστούν έλεγχοι συσχετίσεων με την τιμή του συντελεστή Mann - Whitney και όπως φαίνεται και από τα αποτελέσματα των τιμών p-value όλες οι συσχετίσεις είναι σημαντικές καθώς είναι μικρότερες από το επίπεδο σημαντικότητας που έχει οριστεί (0.05). Θα δείξουμε την συσχέτιση της μεταβλητής που προαναφέραμε με τις ποσοτικές του δείγματος, δηλαδή «Tumor Size», «Regional Nodes Examined» και «Regional Nodes Positive». Σύμφωνα με τον Mann – Whitney έλεγχο για τις μεταβλητές έχουμε συσχέτιση στην θνησιμότητα και το μέγεθος του όγκου, γεγονός που φαίνεται και από το επόμενο boxplot.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Tumor Size	1714357	< 2.2e-16

Πίνακας 4.38: Έλεγχος συσχέτισης μεταβλητών Vital Status – Tumor Size

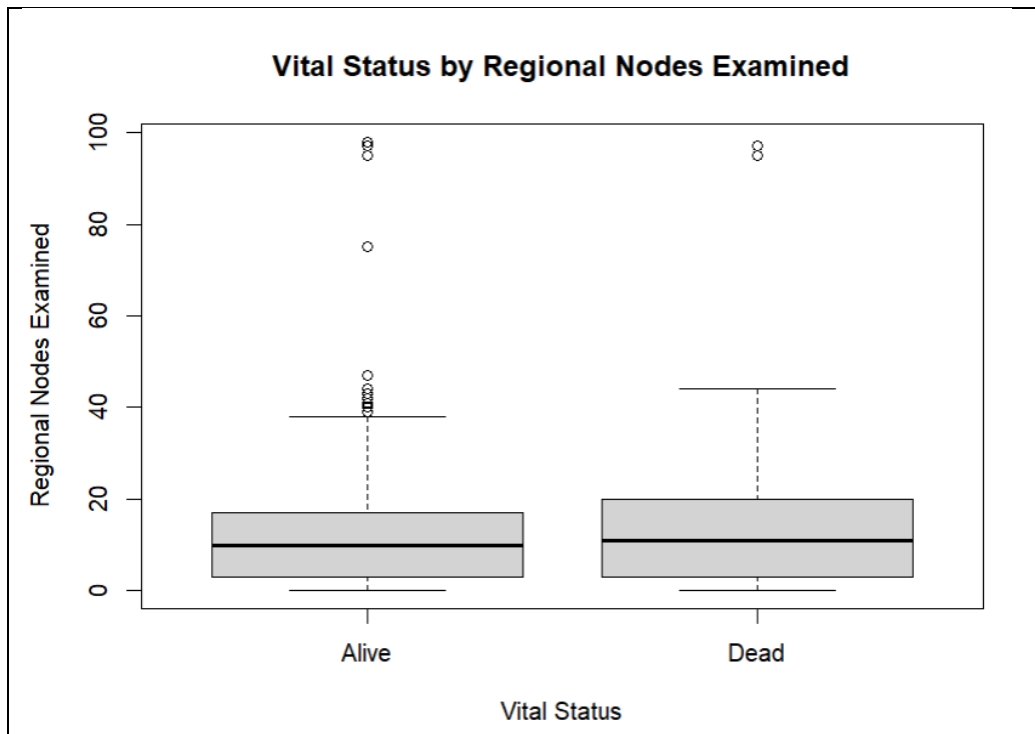


Διάγραμμα 4.24: Boxplot μεταβλητών Vital Status – Tumor Size

Παρατηρείται συσχέτιση και στους περιφερειακούς όγκους που εξετάστηκαν με την κατάσταση της ασθενούς.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Regional Nodes Examined	1470796	0.02616

Πίνακας 4.39: Έλεγχος συσχέτισης μεταβλητών Vital Status – Regional Nodes Examined

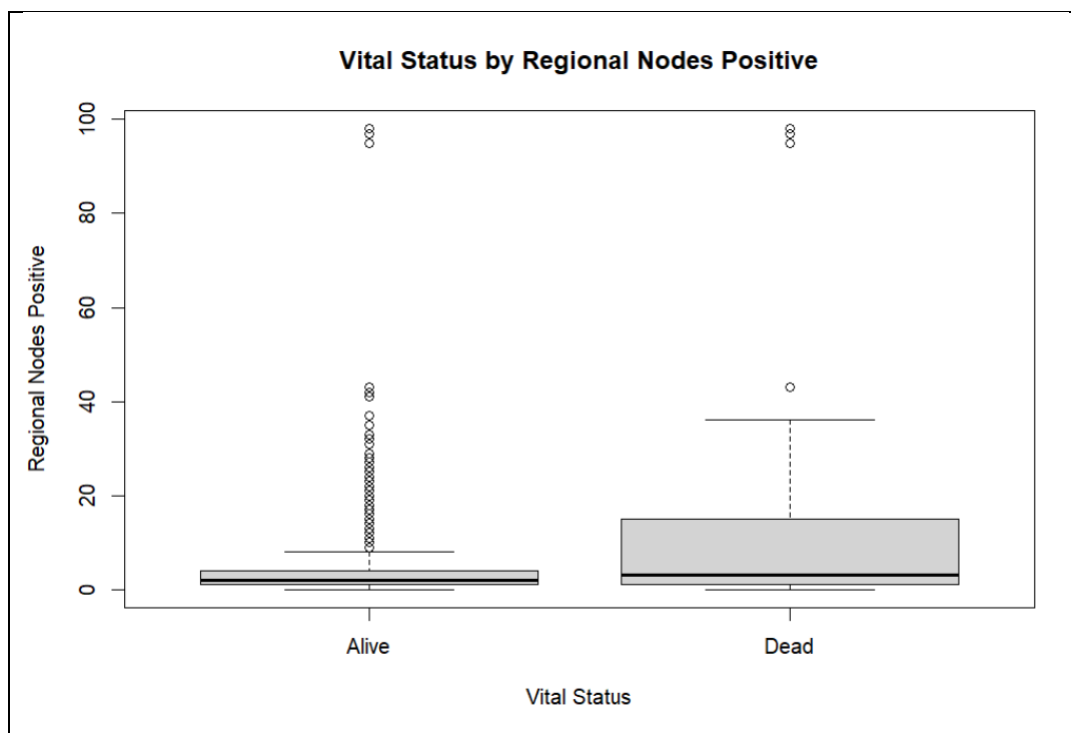


Διάγραμμα 4.25: Boxplot μεταβλητών Vital Status – Regional Nodes Examined

Επίσης για τους όγκους που αποδείχτηκαν θετικοί στον καρκίνο υπάρχει συσχέτιση με την μεταβλητή Vital Status.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Regional Nodes Positive	1713455	< 2.2e-16

Πίνακας 4.40: Έλεγχος συσχέτισης μεταβλητών Vital Status – Regional Nodes Positive



Διάγραμμα 4.26 : Boxplot μεταβλητών Vital Status – Regional Nodes Positive

Για τις ποιοτικές μεταβλητές του dataset θα ακολουθήσουμε τα ίδια βήματα ελέγχοντας την τιμή χ^2 του Pearson και το p-value. Γενικά ο καρκίνος του μαστού αποτελεί την πιο ήπια μορφή καρκίνου σε σχέση με τις άλλες δύο που αναλύονται στην διπλωματική καθώς μόνο το 10% του δείγματος απεβίωσε σε αντίθεση με το καρκίνο του εγκεφάλου (θνησιμότητα 76%) και τον καρκίνο του παγκρέατος (θνησιμότητα 92%). Διαβάζοντας τα παρακάτω αποτελέσματα καταλήγουμε στο γεγονός ότι όλες οι μεταβλητές έχουν συσχέτιση, είναι εξαρτημένες με την μεταβλητή απόκρισης εκτός από τις «Φυλή» και «Πλευρικότητα» εφόσον τα p-value των ελέγχων είναι μεγαλύτερα από το 5%.

Αρχικά παρατηρούμε την σχέση της ηλικίας με την κατάσταση της ασθενούς συμπεραίνοντας μεγαλύτερη θνησιμότητα στις ηλικίες 70+.

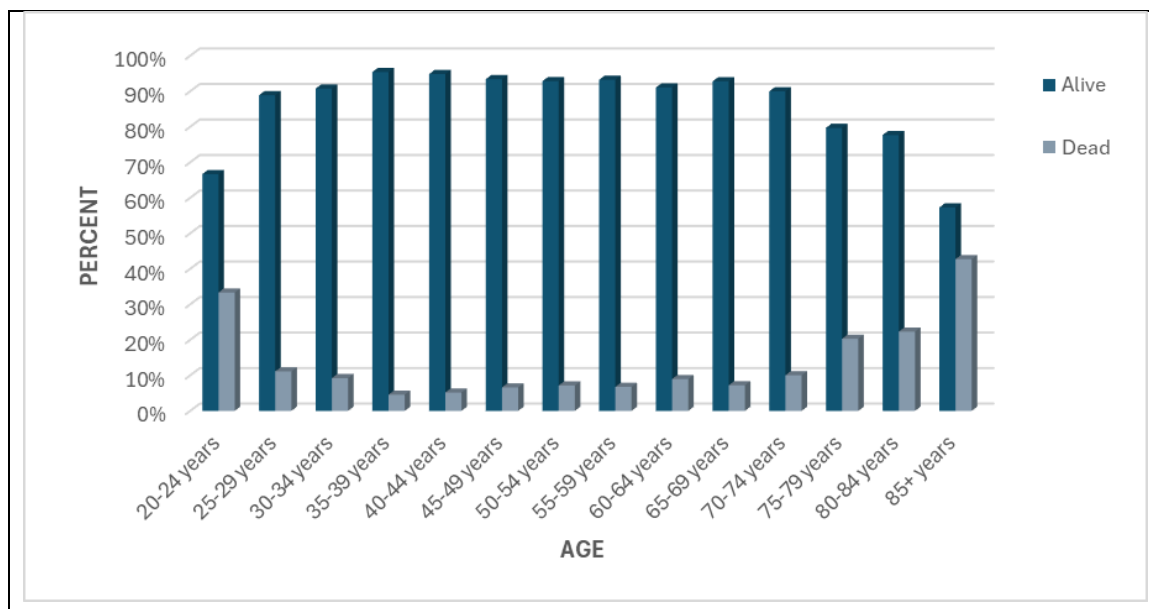
Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Age	370.9	<2.2e-16

Πίνακας 4.41: Έλεγχος συσχέτισης μεταβλητών Vital Status – Age

Age	Επιβίωσαν	Απεβίωσαν
20-24 years	67%	33%
25-29 years	89%	11%
30-34 years	91%	9%
35-39 years	95%	5%
40-44 years	95%	5%
45-49 years	93%	7%
50-54 years	93%	7%
55-59 years	93%	7%
60-64 years	91%	9%

65-69 years	93%	7%
70-74 years	90%	10%
75-79 years	80%	20%
80-84 years	78%	22%
85+ years	57%	43%
Σύνολο	90%	10%

Πίνακας 4.42: Ποσοστά θνησιμότητας της μεταβλητής Age



Διάγραμμα 4.27: Ραβδόγραμμα μεταβλητών Vital Status - Age

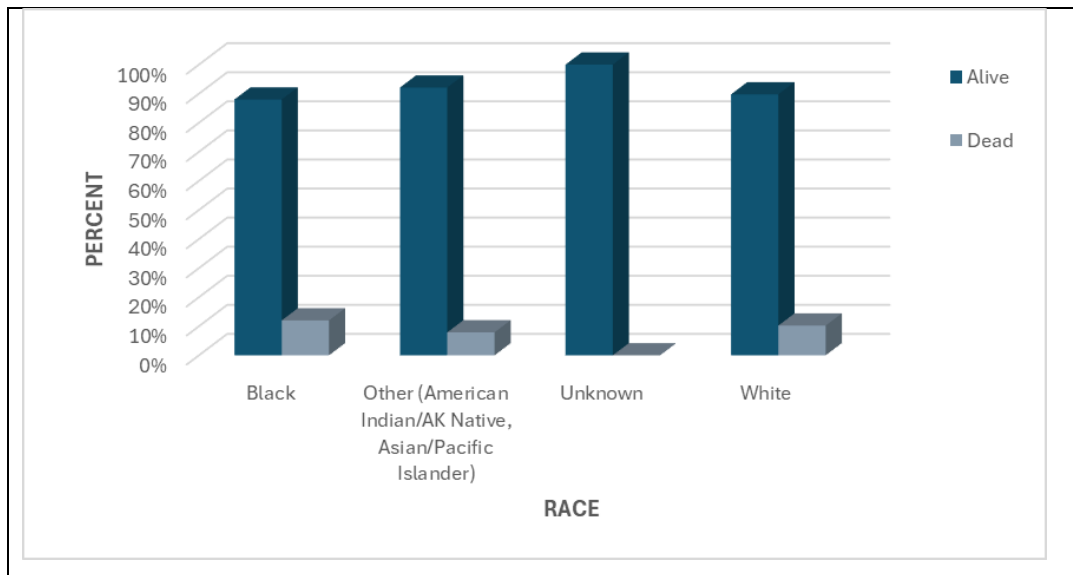
Προχωρώντας στην αναπαράσταση της φυλής και της μεταβλητής απόκρισης «Vital Status», παρότι δεν υπάρχει σημαντική συσχέτιση όπως είδαμε και παραπάνω παρατηρείται θνησιμότητα 12% στην κατηγορία των έγχρωμων γυναικών.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Race	5.5633	0.1349

Πίνακας 4.43: Έλεγχος συσχέτισης μεταβλητών Vital Status – Race

Race	Επιβίωσαν	Απεβίωσαν
Έγχρωμες	88%	12%
Άλλο	92%	8%
Δεν γνωρίζουμε	100%	0%
Λευκές	90%	10%
Σύνολο	90%	10%

Πίνακας 4.44: Ποσοστά θνησιμότητας της μεταβλητής Race



Διάγραμμα 4.28: Ραβδόγραμμα μεταβλητών Vital Status – Race

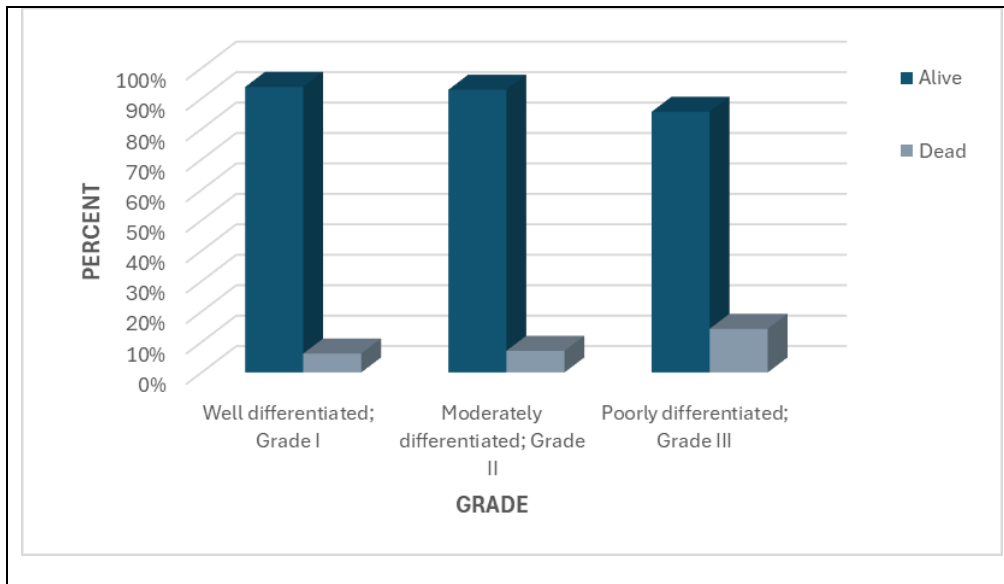
Για τον βαθμό του καρκίνου του μαστού φαίνεται το μεγαλύτερο ποσοστό θνησιμότητας να συγκεντρώνεται στην κατηγορία «Grade III».

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Grade	83.567	<2.2e-16

Πίνακας 4.45: Έλεγχος συσχέτισης μεταβλητών Vital Status – Grade

Grade	Επιβίωσαν	Απεβίωσαν
Grade I	94%	6%
Grade II	93%	7%
Grade III	86%	14%
Σύνολο	90%	10%

Πίνακας 4.46: Ποσοστά θνησιμότητας της μεταβλητής Grade



Διάγραμμα 4.29: Ραβδόγραμμα μεταβλητών Vital Status – Grade

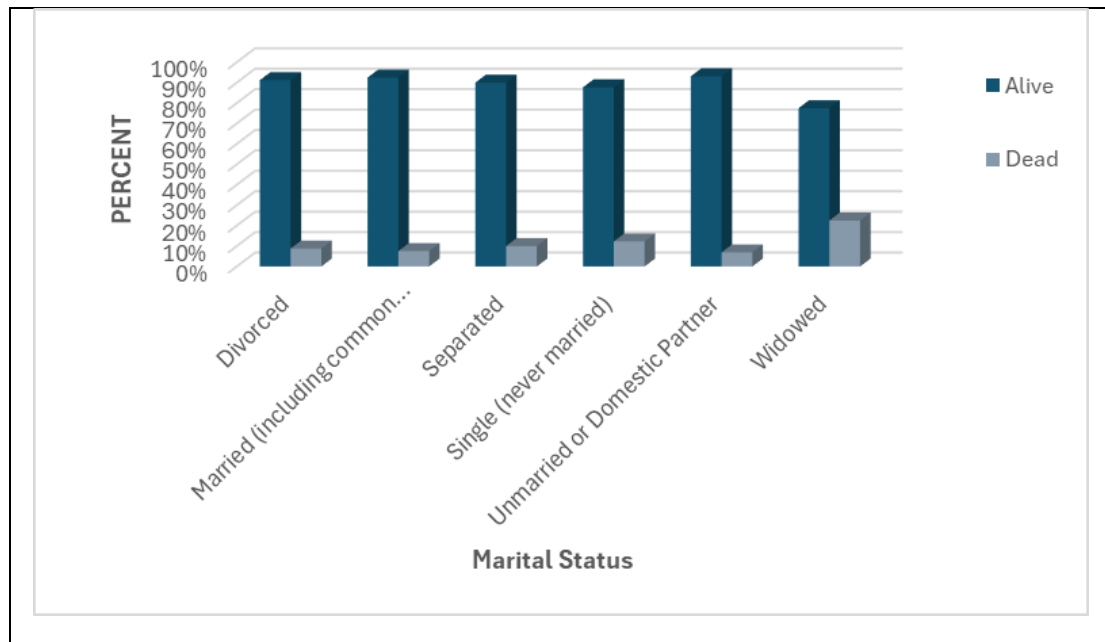
Από το ραβδόγραμμα που ακολουθεί βλέπουμε ότι το ποσοστό θνησιμότητας είναι υψηλότερο στις χήρες γυναίκες και έπειτα στις ελεύθερες.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Marital Status	137.84	<2.2e-16

Πίνακας 4.47: Έλεγχος συσχέτισης μεταβλητών Vital Status – Marital Status

Marital Status	Επιβίωσαν	Απεβίωσαν
Διαζευγμένες	91%	9%
Παντρεμένες	92%	8%
Σε διάσταση	90%	10%
Ελεύθερες	88%	12%
Σε συμβίωση	93%	7%
Χήρες	77%	23%
Σύνολο	90%	10%

Πίνακας 4.48: Ποσοστά θνησιμότητας της μεταβλητής Marital Status



Διάγραμμα 4.30: Ραβδόγραμμα μεταβλητών Vital Status – Marital Status

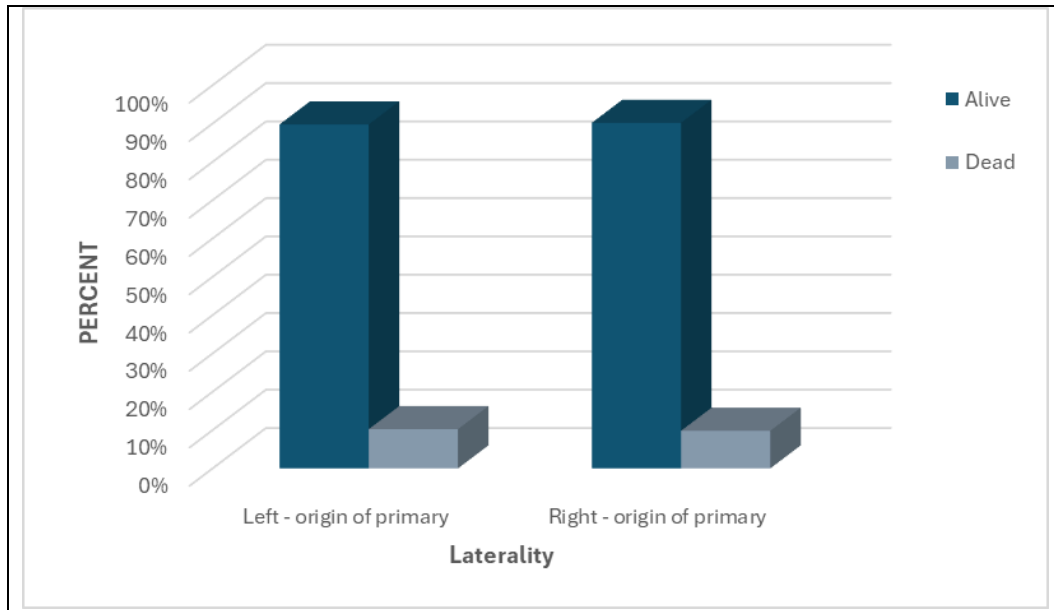
Για την πλευρικότητα και την κατάσταση επιβίωσης παρατηρείται ίσο ποσοστό είτε εντοπιστεί ο καρκίνος από την δεξιά πλευρά είτε την αριστερή (θνησιμότητα 10%).

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Laterality	0.23105	0.6307

Πίνακας 4.49: Έλεγχος συσχέτισης μεταβλητών Vital Status – Laterality

Laterality	Επιβίωσαν	Απεβίωσαν
Αριστερή Πλευρικότητα	90%	10%
Δεξιά Πλευρικότητα	90%	10%
Σύνολο	90%	10%

Πίνακας 4.50: Ποσοστά θνησιμότητας της μεταβλητής Laterality



Διάγραμμα 4.31: Ραβδόγραμμα μεταβλητών Vital Status – Lateralality

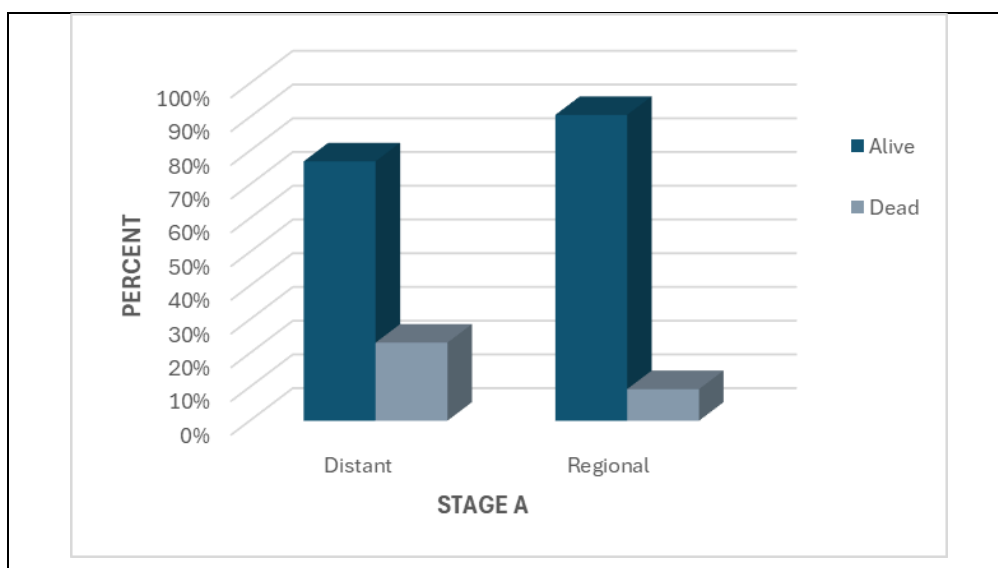
Για το Stage A βλέπουμε ότι ο καρκίνος που έχει αναπτυχθεί σε μακρινούς λεμφαδένες έχει υψηλότερο ποσοστό θνησιμότητας.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Stage A	52.04	5.437e-13

Πίνακας 4.51: Έλεγχος συσχέτισης μεταβλητών Vital Status – Stage A

Stage A	Επιβίωσαν	Απεβίωσαν
Μακρινός	77%	23%
Περιφερικός	91%	9%
Σύνολο	90%	10%

Πίνακας 4.52: Ποσοστά θνησιμότητας της μεταβλητής Stage A



Διάγραμμα 4.32: Ραβδόγραμμα μεταβλητών Vital Status – Stage A

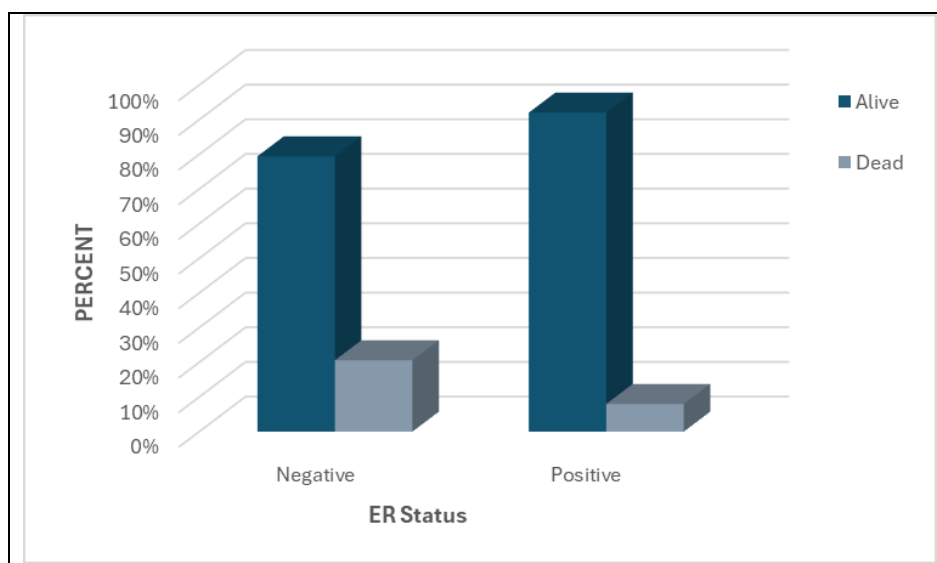
Όπως παρατηρείται και από τον παρακάτω πίνακα το υψηλότερο ποσοστό θνησιμότητας βρίσκεται στις γυναίκες που έχουν αρνητικό υποδοχέα οιστρογόνου.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
ER Status	131.58	<2.2e-16

Πίνακας 4.53: Έλεγχος συσχέτισης μεταβλητών Vital Status – ER Status

ER Status	Επιβίωσαν	Απεβίωσαν
Αρνητικό	79%	21%
Θετικό	92%	8%
Σύνολο	90%	10%

Πίνακας 4.54: Ποσοστά θνησιμότητας της μεταβλητής ER Status



Διάγραμμα 4.33: Ραβδόγραμμα μεταβλητών Vital Status – ER Status

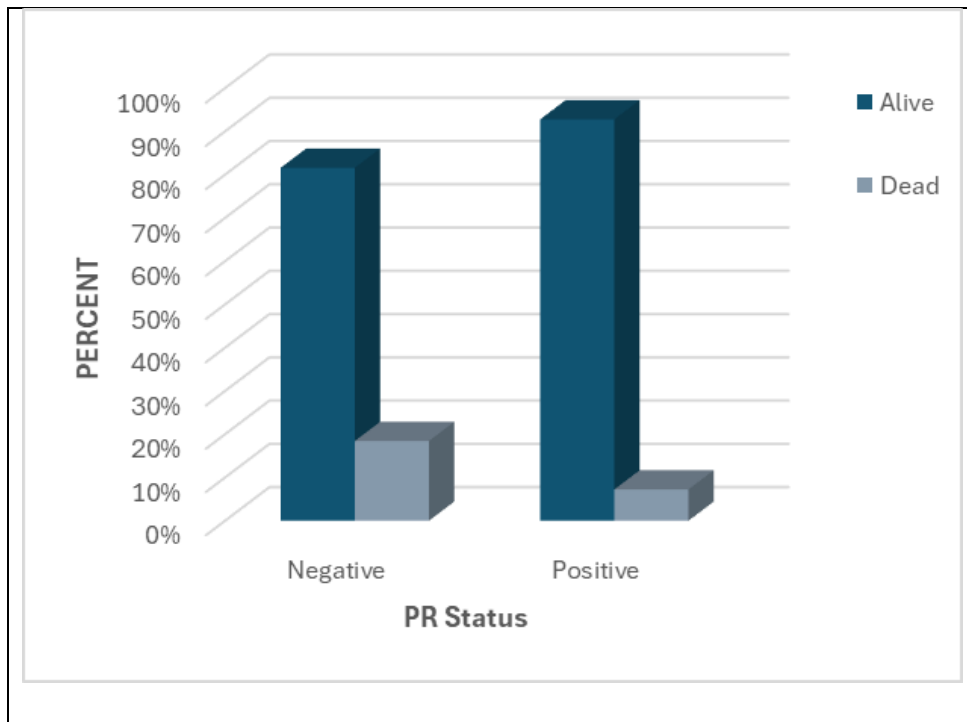
Αναφορικά με τον υποδοχέα προγεστερόνης την υψηλότερη θνησιμότητα την βρίσκουμε στην κατηγορία «Αρνητικό» με ποσοστό 18%.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
PR Status	143.51	<2.2e-16

Πίνακας 4.55: Έλεγχος συσχέτισης μεταβλητών Vital Status – PR Status

RR Status	Επιβίωσαν	Απεβίωσαν
Αρνητικό	82%	18%
Θετικό	93%	7%
Σύνολο	90%	10%

Πίνακας 4.56: Ποσοστά θνησιμότητας της μεταβλητής PR Status



Διάγραμμα 4.34: Ραβδόγραμμα μεταβλητών Vital Status – PR Status

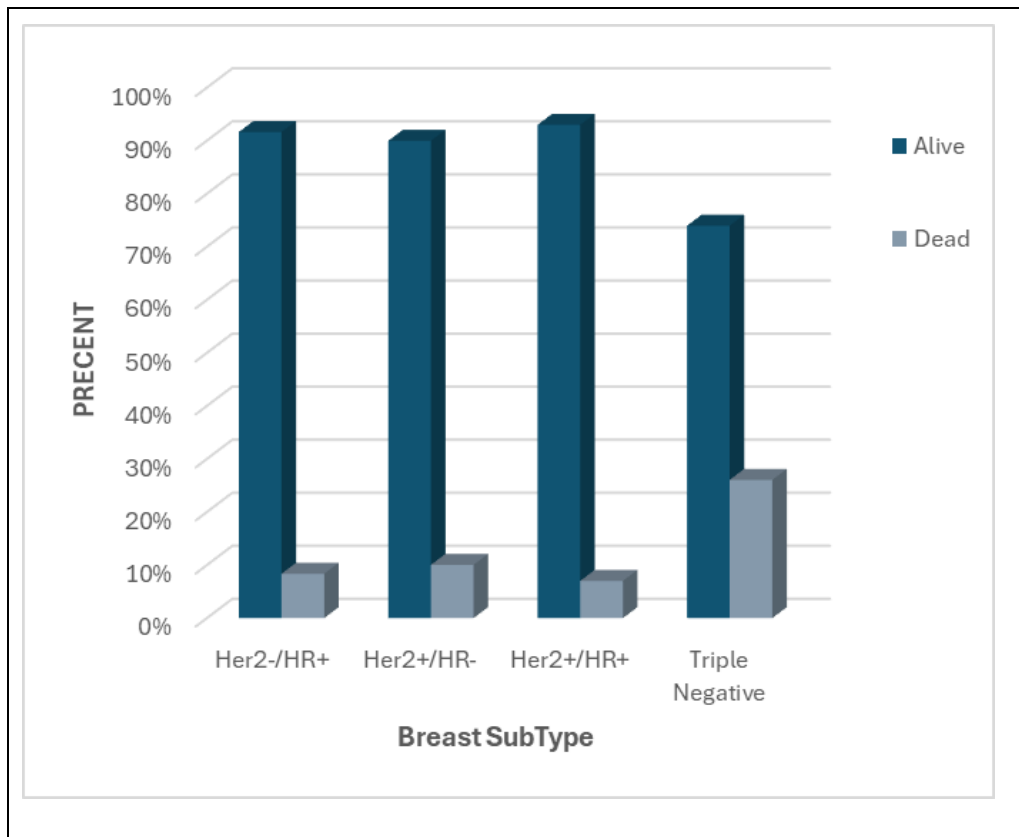
Επιπρόσθετα, για την υποκατηγορία του καρκίνου του μαστού διαπιστώνεται θανάσιμη κατάσταση για της γυναίκες με τριπλό αρνητικό.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Breast SubType	175.29	<2.2e-16

Πίνακας 4.57: Έλεγχος συσχέτισης μεταβλητών Vital Status – Breast SubType

Breast SubType	Επιβίωσαν	Απεβίωσαν
Her2- / HR+	92%	8%
Her2+ / HR-	90%	10%
Her2+ / HR+	93%	7%
Τριπλό Αρνητικό	74%	26%
Σύνολο	90%	10%

Πίνακας 4.58: Ποσοστά θνησιμότητας της μεταβλητής Breast SubType



Διάγραμμα 4.35: Ραβδόγραμμα μεταβλητών Vital Status – Breast SubType

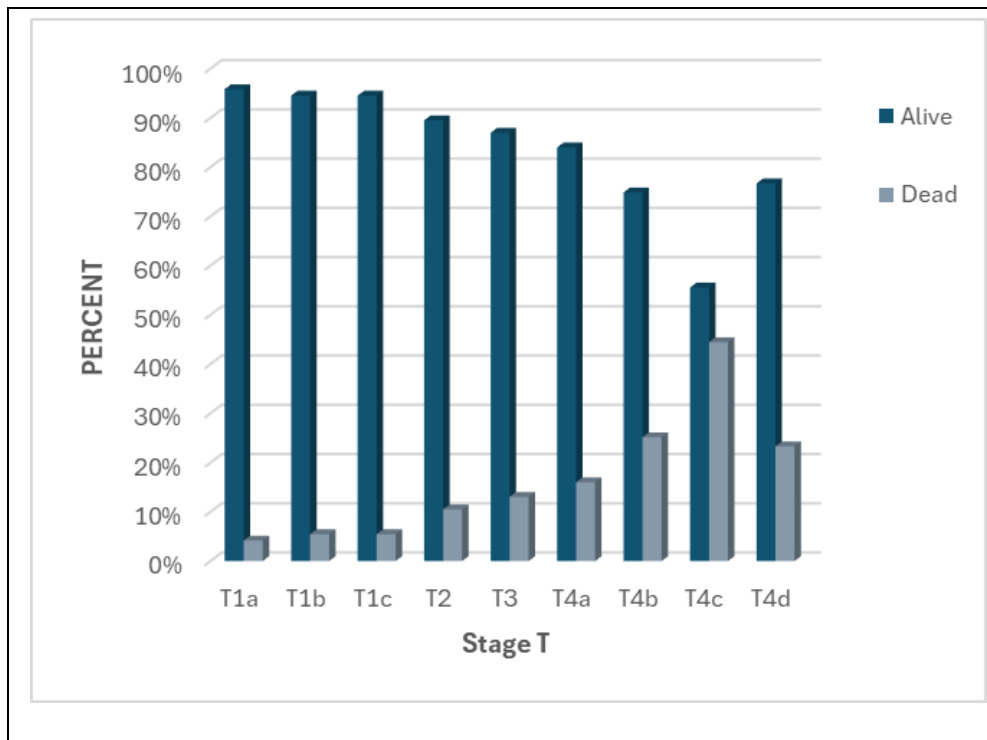
Για την κατάσταση επιβίωσης σε σχέση με το μέγεθος και την θέση του όγκου έχουμε υψηλότερο ποσοστό θνησιμότητας στον όγκο τύπου T4c.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Stage T	132.43	<2.2e-16

Πίνακας 4.59: Έλεγχος συσχέτισης μεταβλητών Vital Status – Stage T

Stage T	Επιβίωσαν	Απεβίωσαν
T1a	96%	4%
T1b	95%	5%
T1c	95%	5%
T2	90%	10%
T3	87%	13%
T4a	84%	16%
T4b	75%	25%
T4c	56%	44%
T4d	77%	23%
Σύνολο	90%	10%

Πίνακας 4.60: Ποσοστά θνησιμότητας της μεταβλητής Stage T



Διάγραμμα 4.36: Ραβδόγραμμα μεταβλητών Vital Status – Stage T

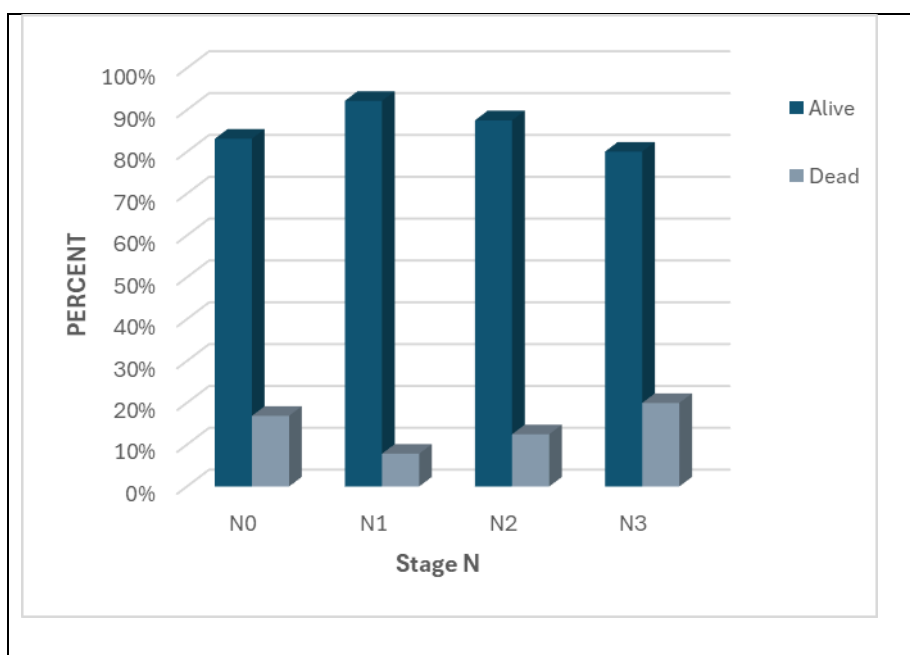
Για τον N τύπο του καρκίνου του μαστού το υψηλότερο ποσοστό θνησιμότητας υπάρχει στην κατηγορία N3 με 20%, ενώ με 17% είναι ο τύπος N0.

Μεταβλητή	Τιμή Chi-Square Pearson	P-Value Ελέγχου
Stage N	95.468	<2.2e-16

Πίνακας 4.61: Έλεγχος συσχέτισης μεταβλητών Vital Status – Stage N

Stage N	Επιβίωσαν	Απεβίωσαν
N0	83%	17%
N1	92%	8%
N2	88%	12%
N3	80%	20%
Σύνολο	90%	10%

Πίνακας 4.62: Ποσοστά θνησιμότητας της μεταβλητής Stage N



Διάγραμμα 4.37: Ραβδόγραμμα μεταβλητών Vital Status – Stage N

4.3 Δεδομένα Καρκίνου Παγκρέατος

4.3.1 Ανάλυση βάσει της μεταβλητής «Survival Months»

Για το τελευταίο dataset, εκείνο του καρκίνου του παγκρέατος θα ακολουθηθεί η ίδια διαδικασία για τις ποσοτικές μεταβλητές ελέγχοντας την συσχέτισή του με την «Survival Months». Από τον παρακάτω έλεγχο του Spearman διακρίνουμε συσχέτιση με την το μέγεθος του όγκου αλλά και τον συνολικό αριθμό όγκων σε σχέση με τους μήνες επιβίωσης.

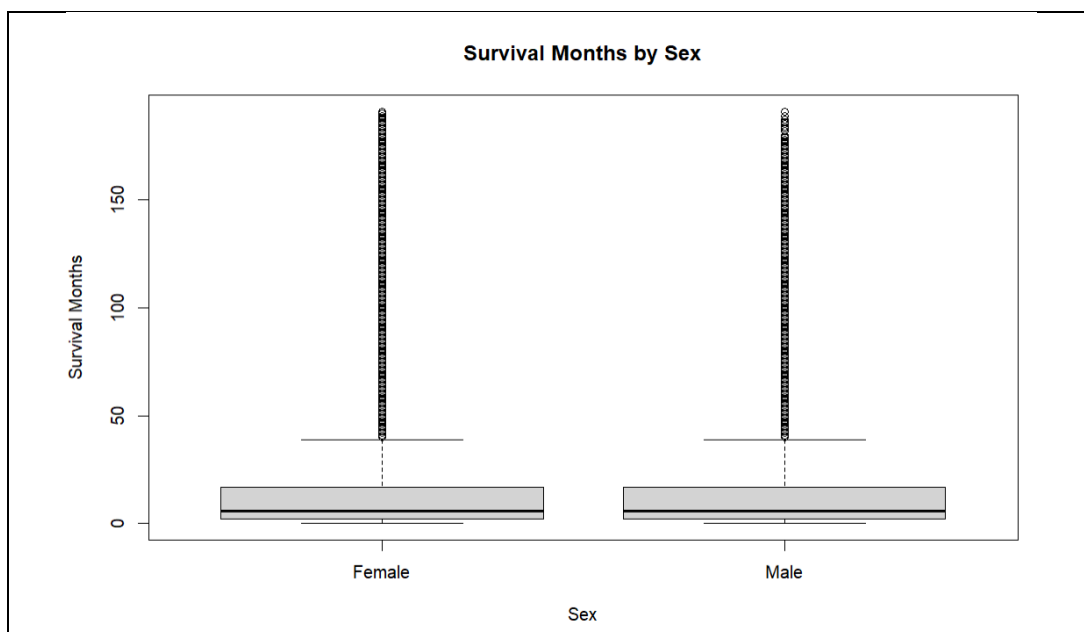
Μεταβλητή	Τιμή Συντελεστή Spearman	P-Value Ελέγχου
Tumor Size	-0.2047793	<2.2e-16
Total number of in situ / malignant tumors for patient	0.04607135	<2.2e-16

Πίνακας 4.63: Συντελεστές Spearman για τον καρκίνο του Παγκρέατος

Συνεχίζουμε με την ποιοτική μεταβλητή «Sex» και τον έλεγχο Mann – Whitney. Σύμφωνα με τον έλεγχο, εφόσον το p-value είναι οριακά μεγαλύτερο από το 5% καταλαβαίνουμε ότι δεν σχετίζονται οι μήνες επιβίωσης με το αν ο ασθενής είναι άντρας ή γυναίκα.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Sex	189704125	0.5103

Πίνακας 4.64: Έλεγχος συσχέτισης μεταβλητών Survival Months – Sex



Διάγραμμα 4.38: Boxplot μεταβλητών Survival Months – Sex

Για τις μεταβλητές «Age», «Race», «Grade», «Diagnostic Confirmation» και «Stage A» θα εκτελεστεί ο έλεγχος Kruskal – Wallis καθώς περιέχουν πάνω από δυο κατηγορίες. Για την ηλικία και τους μήνες επιβίωσης υπάρχει συσχέτιση, καθώς ο έλεγχος μας δίνει πολύ μικρό p-value.

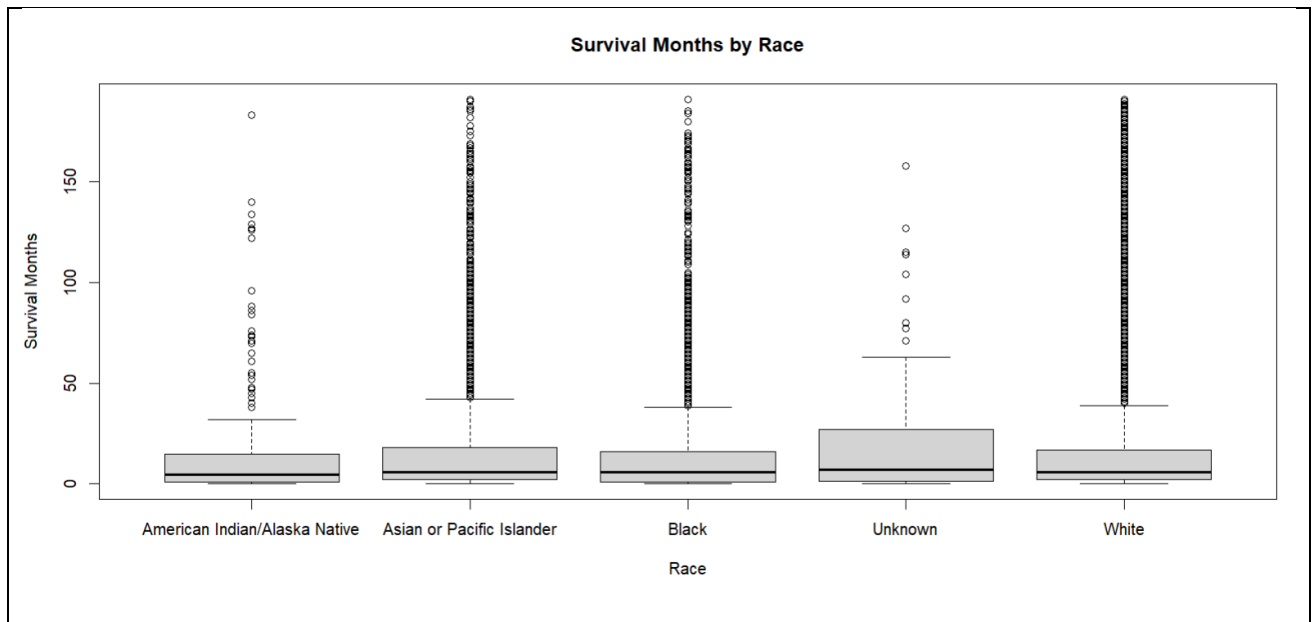
Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Age	4054.6	<2.2e-16

Πίνακας 4.65: Έλεγχος συσχέτισης μεταβλητών Survival Months – Age

Ομοίως για την φυλή του ασθενή μπορούμε να πούμε ότι έχει σχέση με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Race	20.943	0.000325

Πίνακας 4.66: Έλεγχος συσχέτισης μεταβλητών Survival Months – Race

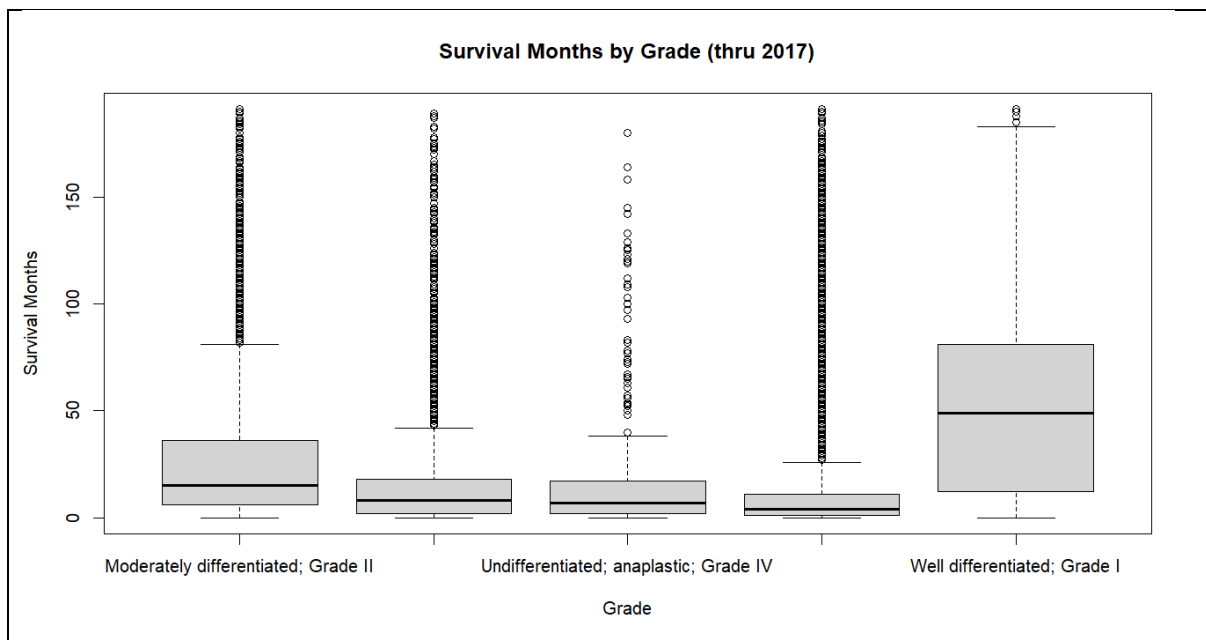


Διάγραμμα 4.39: Boxplot μεταβλητών Survival Months – Race

Έπειτα, παρατηρούμε ότι και ο βαθμός του καρκίνου του παγκρέατος έχει συσχέτιση με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Grade	5842.3	< 2.2e-16

Πίνακας 4.67: Έλεγχος συσχέτισης μεταβλητών Survival Months – Grade

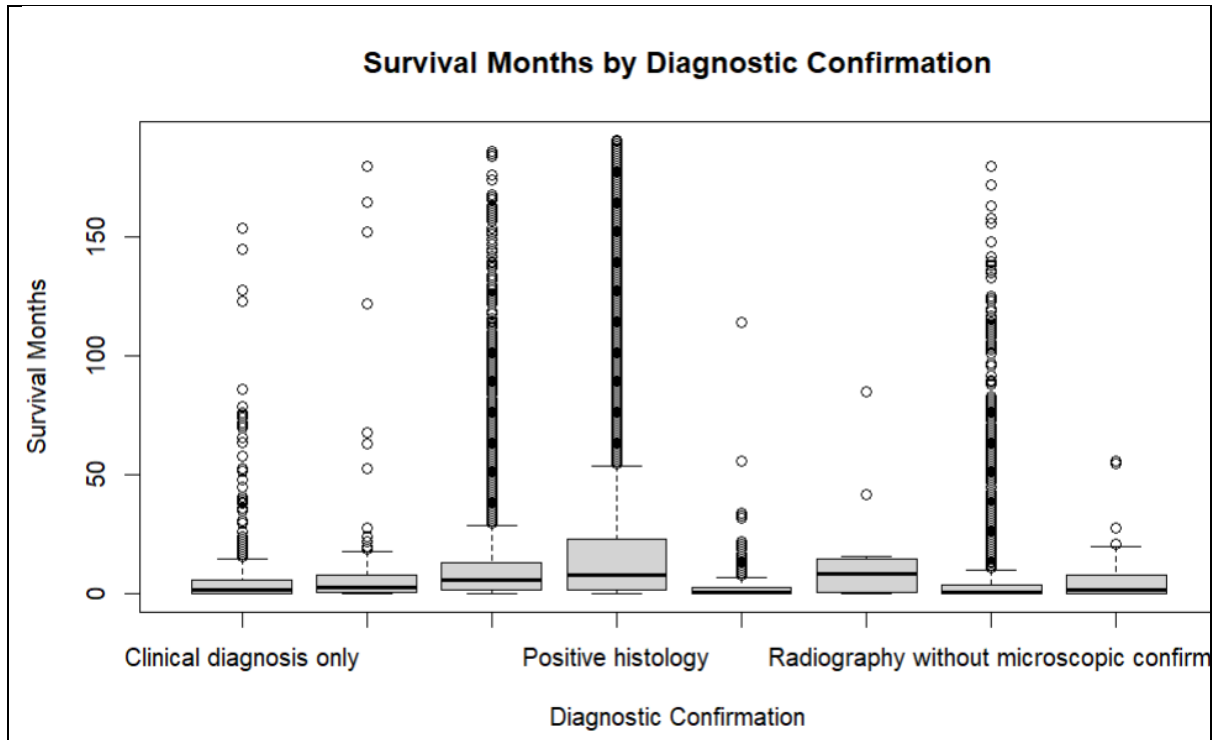


Διάγραμμα 4.40: Boxplot μεταβλητών Survival Months – Grade

Για τον τρόπο διάγνωσης του καρκίνου βλέπουμε συσχέτιση με την μεταβλητή «Survival Months».

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Diagnostic Confirmation	3684.8	< 2.2e-16

Πίνακας 4.68: Έλεγχος συσχέτισης μεταβλητών Survival Months – Diagnostic Confirmation

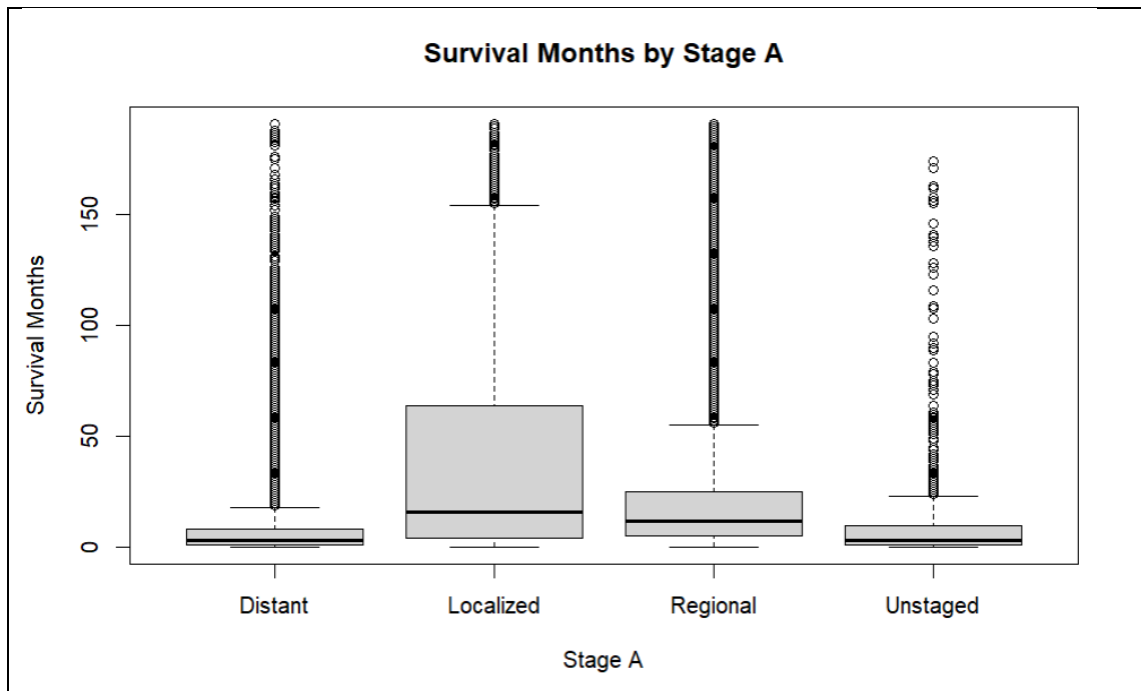


Διάγραμμα 4.41: Boxplot μεταβλητών Survival Months – Diagnostic Confirmation

Τελευταία ποιοτική μεταβλητή «Stage A» παρατηρούμε επίσης συσχέτιση με τους μήνες επιβίωσης.

Μεταβλητή	Τιμή Ελέγχου Kruskal – Wallis	P-Value Ελέγχου
Stage A	6811.1	< 2.2e-16

Πίνακας 4.69: Έλεγχος συσχέτισης μεταβλητών Survival Months – Stage A



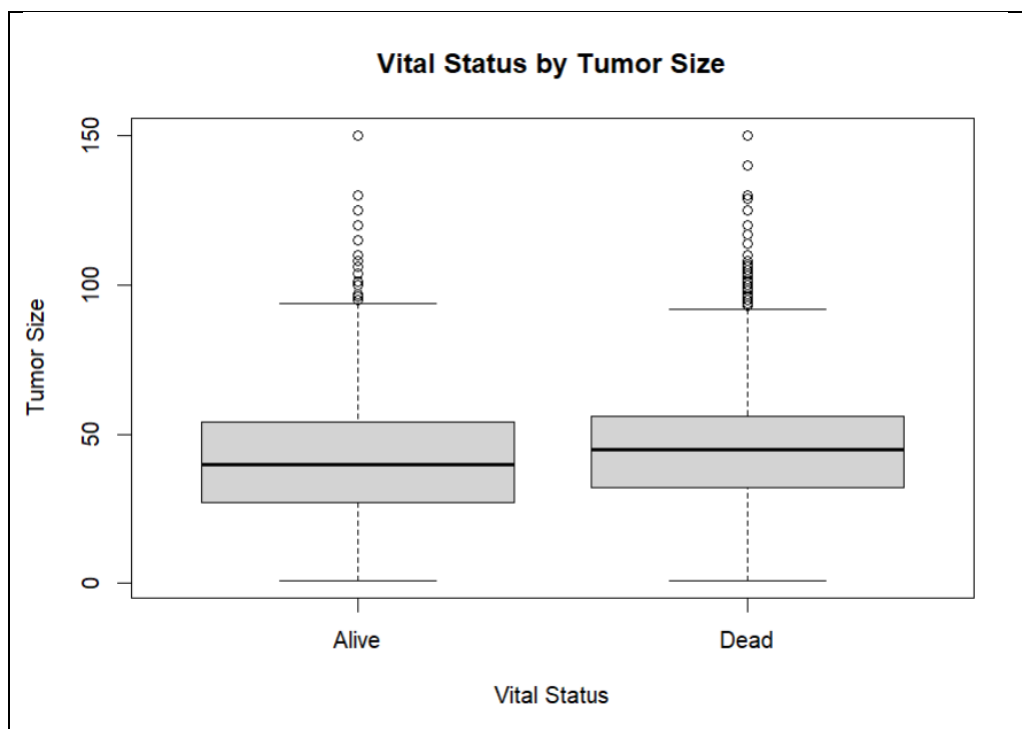
Διάγραμμα 4.42: Boxplot μεταβλητών Survival Months – Stage A

4.3.2 Ανάλυση βάσει της μεταβλητής «Vital Status»

Σε αυτή την τελευταία υποενότητα θα ανακαλύψουμε τις συσχετίσεις των μεταβλητών με εκείνη που ορίζει την τελική κατάσταση του ασθενή, «Vital Status». Για τις ποσοτικές μεταβλητές «Tumor Size» και «Total number of tumors» θα λειτουργήσουμε σύμφωνα με τον έλεγχο Mann – Whitney καθώς από την άλλη μεριά έχουμε την κατηγορική «Vital Status». Από τον παρακάτω πίνακα καταλαβαίνουμε ότι οι μεταβλητές Tumor Size και Vital Status είναι συσχετισμένες.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Tumor Size	1714357	< 2.2e-16

Πίνακας 4.70: Έλεγχος συσχέτισης μεταβλητών Vital Status – Tumor Size



Διάγραμμα 4.43: Boxplot μεταβλητών Vital Status – Tumor Size

Παρομοίως, έχουμε και συσχέτιση μεταξύ των μεταβλητών Vital Status και Total number of tumors.

Μεταβλητή	Τιμή Ελέγχου Mann – Whitney	P-Value Ελέγχου
Total number of in situ / malignant tumors for patient	41062227	< 2.2e-16

Πίνακας 4.71: Έλεγχος συσχέτισης μεταβλητών Vital Status – Number of tumors

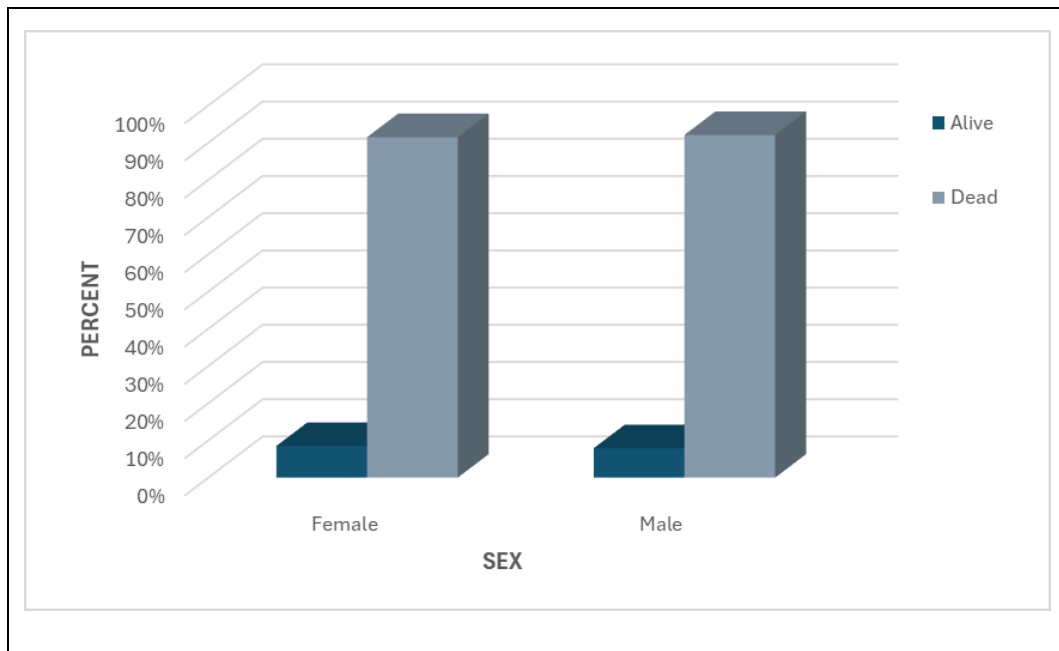
Ακολουθούν οι έλεγχοι Chi Squared του Pearson για τις συσχετίσεις των ποιοτικών μεταβλητών και τα ραβδογράμματα αυτών. Να αναφέρουμε ότι το συνολικό ποσοστό των ατόμων που απεβίωσαν φτάνει το 92% και εκείνων που έζησαν είναι 8%. Είναι φανερό ότι είναι ο πιο επιθετικός και επικίνδυνος τύπος καρκίνου για την συγκεκριμένη διπλωματική και δεδομένα. Πιο αναλυτικά για την μεταβλητή «Φύλο» παρατηρούνται ίσα ποσοστά για άντρες και γυναίκες συναρτήσει με την κατάσταση επιβίωσης.

Μεταβλητή	Τιμή Chi-Square Pearson	Τιμή P-Value
Sex	4.5657	0.03262

Πίνακας 4.72: Έλεγχος συσχέτισης μεταβλητών Vital Status – Sex

Sex	Επιβίωσαν	Απεβίωσαν
Γυναίκες	9%	91%
Άντρες	8%	92%
Σύνολο	8%	92%

Πίνακας 4.73: Ποσοστά θνησιμότητας της μεταβλητής Sex



Διάγραμμα 4.44: Ραβδόγραμμα μεταβλητών Vital Status – Sex

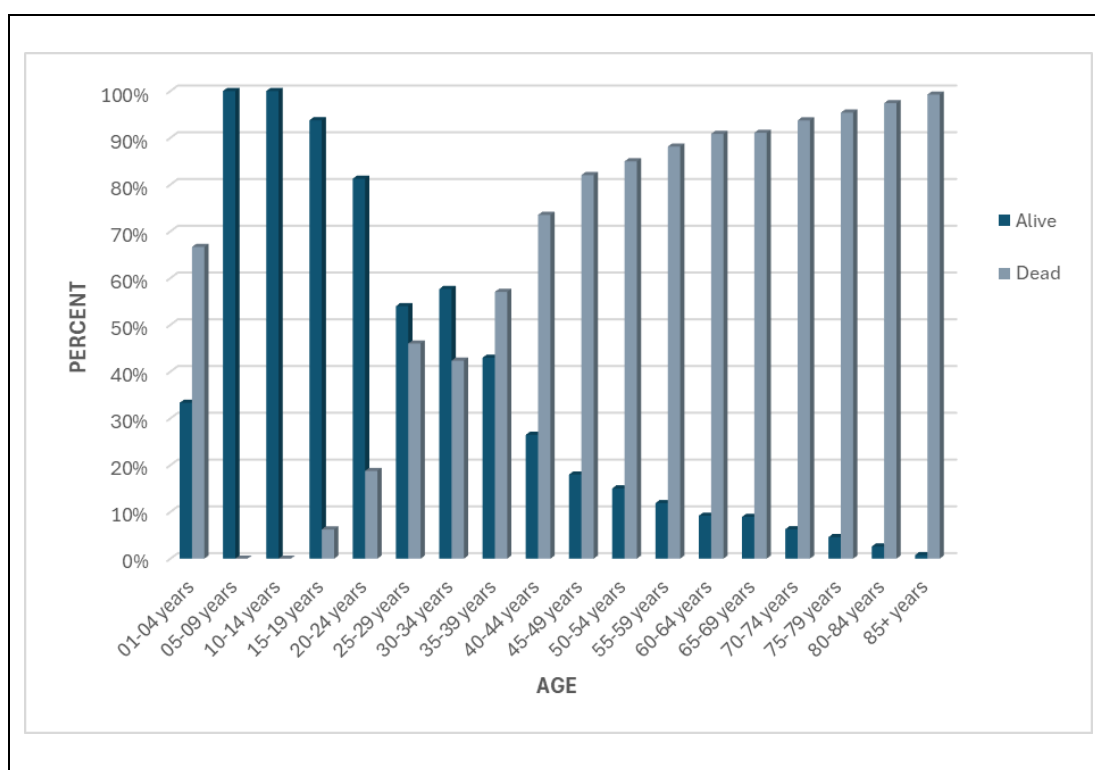
Μεγαλύτερη θνησιμότητα παρατηρείται στις ηλικιακές ομάδες μετά τα 50 έτη.

Μεταβλητή	Τιμή Chi-Square Pearson	Τιμή P-Value
Age	2846.6	< 2.2e-16

Πίνακας 4.74: Έλεγχος συσχέτισης μεταβλητών Vital Status – Age

Age	Επιβίωσαν	Απεβίωσαν
01-04 years	33%	67%
05-09 years	100%	0%
10-14 years	100%	0%
15-19 years	94%	6%
20-24 years	81%	19%
25-29 years	54%	46%
30-34 years	58%	42%
35-39 years	43%	57%
40-44 years	26%	74%
45-49 years	18%	82%
50-54 years	15%	85%
55-59 years	12%	88%
60-64 years	9%	91%
65-69 years	9%	91%
70-74 years	6%	94%
75-79 years	5%	95%
80-84 years	3%	97%
85+ years	1%	99%
Σύνολο	8%	92%

Πίνακας 4.75: Ποσοστά θνησιμότητας της μεταβλητής Sex



Διάγραμμα 4.45: Ραβδόγραμμα μεταβλητών Vital Status – Age

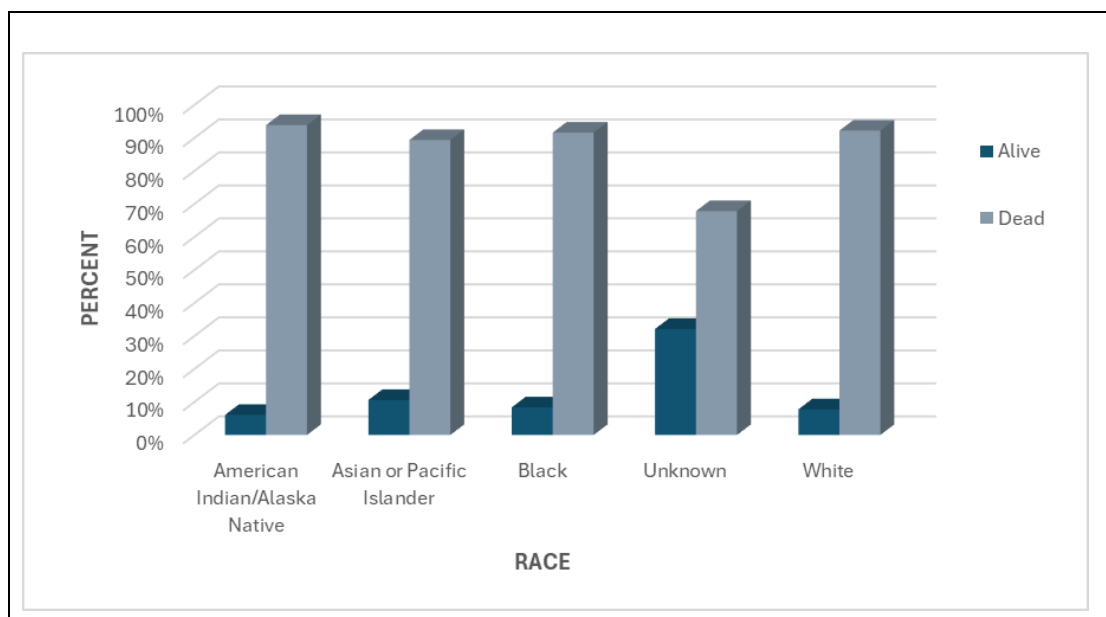
Ως προς την φυλή φαίνεται από τον παρακάτω πίνακα ότι Αμερικάνοι έχουν τα μεγαλύτερα ποσοστά θνησιμότητας ενώ ακολουθούν οι λευκοί και οι έγχρωμοι.

Μεταβλητή	Τιμή Chi-Square Pearson	Τιμή P-Value
Race	108.68	< 2.2e-16

Πίνακας 4.76: Έλεγχος συσχέτισης μεταβλητών Vital Status – Race

Race	Επιβίωσαν	Απεβίωσαν
Αμερικανοί	6%	94%
Ασιάτες	11%	89%
Έγχρωμοι	8%	92%
Δεν γνωρίζουμε	32%	68%
Λευκοί	8%	92%
Σύνολο	8%	92%

Πίνακας 4.77: Ποσοστά θνησιμότητας της μεταβλητής Race



Διάγραμμα 4.46: Ραβδόγραμμα μεταβλητών Vital Status – Race

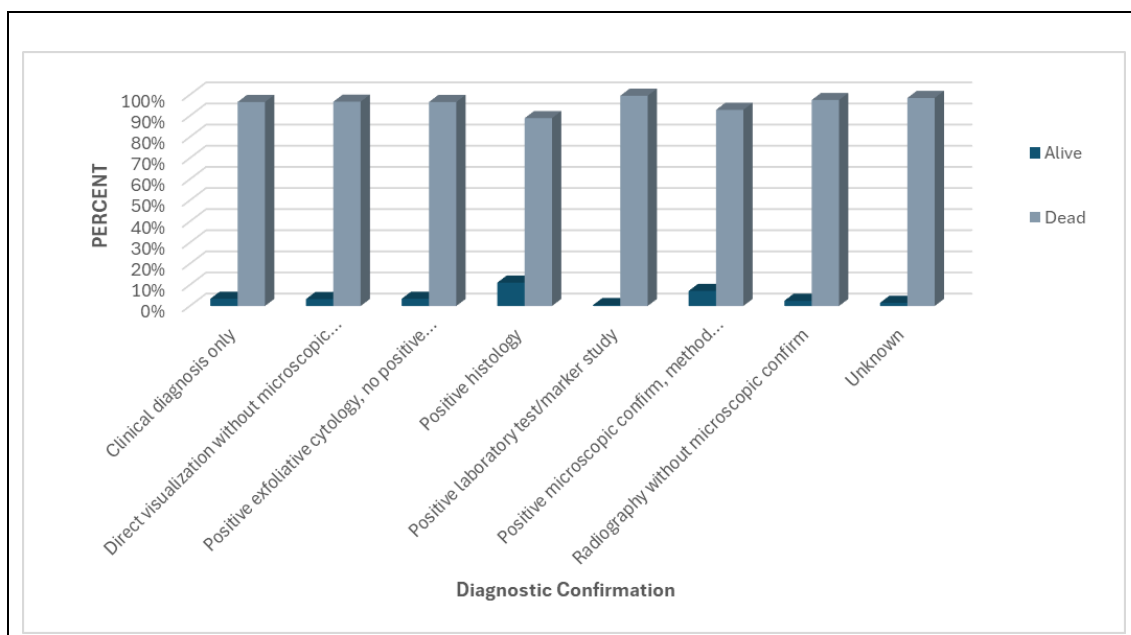
Εκείνοι με θετική εργαστηριακή δοκιμή παρατηρήθηκε με την απόλυτη θνησιμότητα και με την λιγότερη εκείνοι που διαγνώστηκαν με θετική ιστολογία.

Μεταβλητή	Τιμή Chi-Square Pearson	Τιμή P-Value
Diagnostic Confirmation	760.59	< 2.2e-16

Πίνακας 4.78: Έλεγχος συσχέτισης μεταβλητών Vital Status – Diagnostic Confirmation

Diagnostic Confirmation	Επιβίωσαν	Απεβίωσαν
Κλινική Διάγνωση	3%	97%
Απευθείας Οπτικοποίηση	3%	97%
Θετική Αποφολιδωτική Κυτταρολογία	3%	97%
Θετική Ιστολογία	11%	89%
Θετική Εργαστηριακή Δοκιμή	0%	100%
Θετική Μικροσκοπική Επιβεβαίωση	7%	93%
Ακτινογραφία	2%	98%
Δεν Γνωρίζουμε	2%	98%
Σύνολο	8%	92%

Πίνακας 4.79: Ποσοστά θνησιμότητας της μεταβλητής Diagnostic Confirmation



Διάγραμμα 4.47: Ραβδόγραμμα μεταβλητών Vital Status – Diagnostic Confirmation

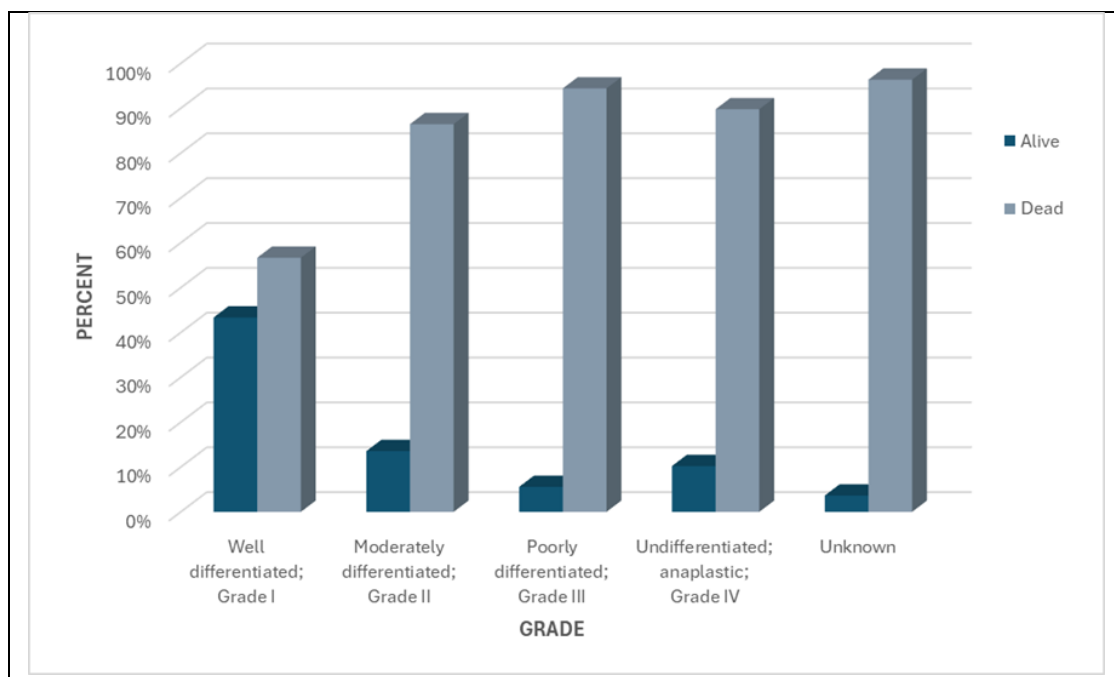
Σχετικά με τον βαθμό του καρκίνου υψηλότερο ποσοστό όπως είναι αναμενόμενο καταγράφεται στους ασθενείς με βαθμό III και IV.

Μεταβλητή	Τιμή Chi-Square Pearson	Τιμή P-Value
Grade	5336.3	< 2.2e-16

Πίνακας 4.80: Έλεγχος συσχέτισης μεταβλητών Vital Status – Grade

Grade	Επιβίωσαν	Απεβίωσαν
Βαθμός I	43%	57%
Βαθμός II	14%	86%
Βαθμός III	6%	94%
Βαθμός IV	10%	90%
Δεν γνωρίζουμε	4%	96%
Σύνολο	8%	92%

Πίνακας 4.81: Ποσοστά θνησιμότητας της μεταβλητής Grade



Διάγραμμα 4.48: Ραβδόγραμμα μεταβλητών Vital Status – Grade

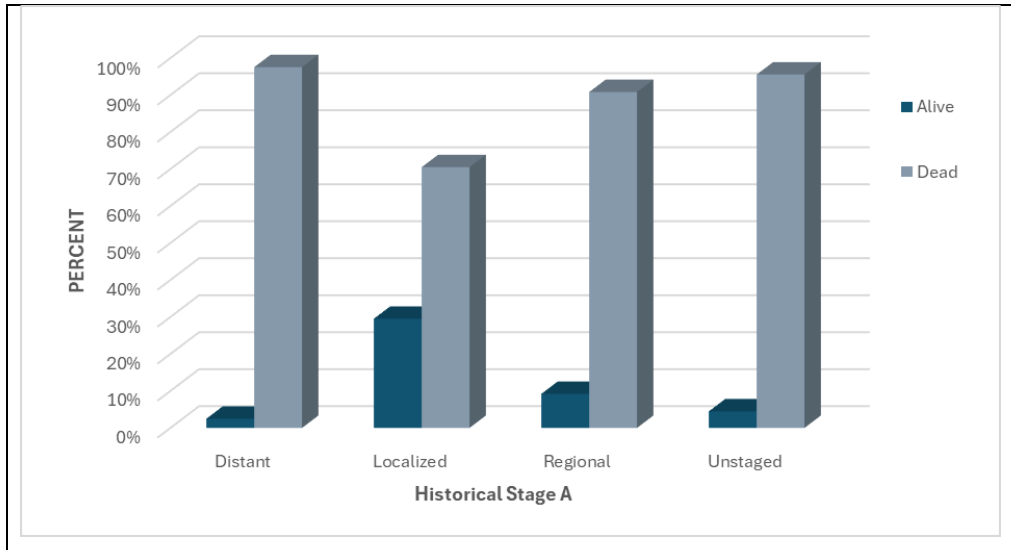
Τέλος, για την μεταβλητή Historical Stage A βλέπουμε υψηλότερο ποσοστό θνησιμότητας σε μεταστατικούς καρκίνους.

Μεταβλητή	Τιμή Chi-Square Pearson	Τιμή P-Value
Historical Stage A	3683.8	< 2.2e-16

Πίνακας 4.82: Έλεγχος συσχέτισης μεταβλητών Vital Status – Historical Stage A

Historical Stage A	Επιβίωσαν	Απεβίωσαν
Μακρινός	2%	98%
Εντοπισμένος	30%	70%
Περιφερικός	9%	91%
Δεν γνωρίζουμε	4%	96%
Σύνολο	8%	92%

Πίνακας 4.83: Ποσοστά θνησιμότητας της μεταβλητής Historical Stage A



Διάγραμμα 4.49: Ραβδόγραμμα μεταβλητών Vital Status – Historical Stage A

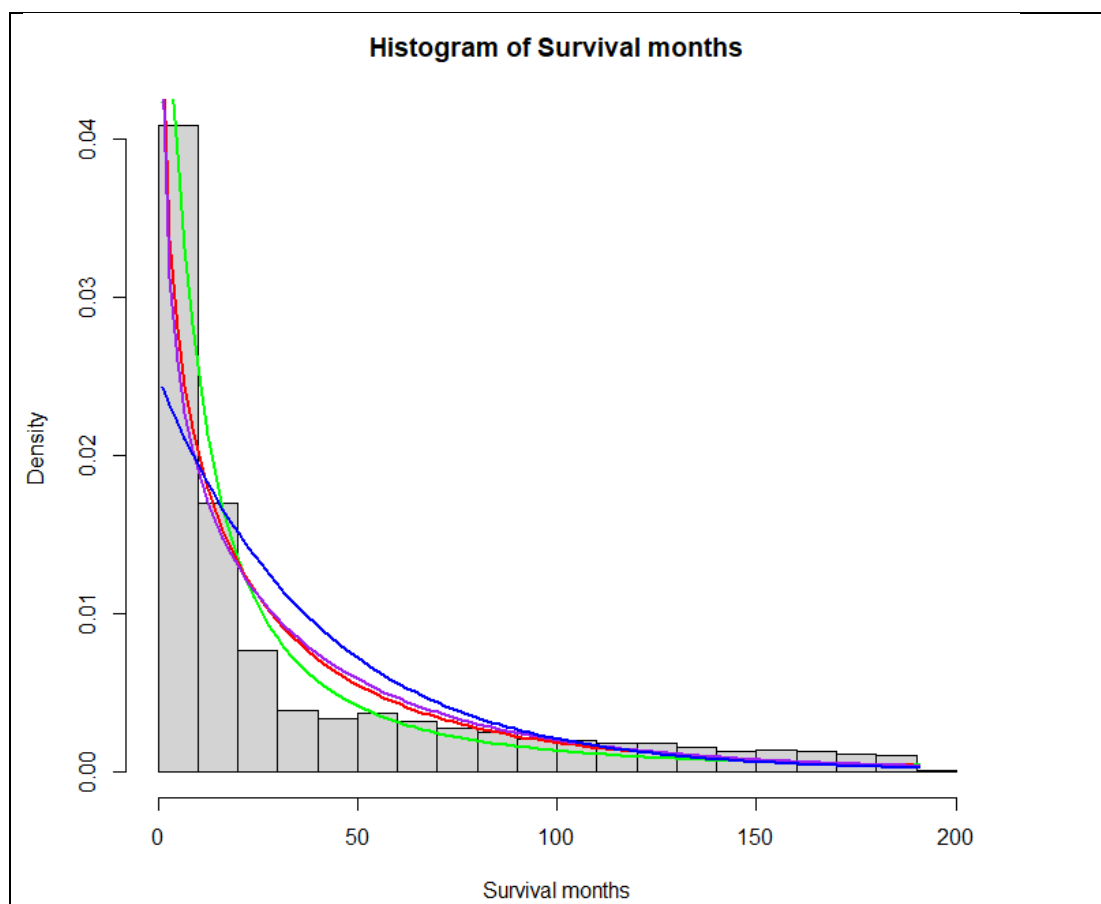
ΚΕΦΑΛΑΙΟ 5°

5. Παραμετρικά Μοντέλα επιβίωσης

5.1. Εισαγωγικά

Στο κεφάλαιο θα μας απασχολήσει η εύρεση του κατάλληλου παραμετρικού μοντέλου επιβίωσης για κάθε σύνολο δεδομένων με μεταβλητή απόκρισης «Survival Months». Για τον σκοπό αυτόν θα εργαστούμε με την γλώσσα προγραμματισμού R και προσαρμόζοντας διαφορετικές κατανομές θα καταλήξουμε ποια ταιριάζει καλύτερα στα δεδομένα μας. Οι κατανομές που θα ελεγχθούν είναι οι : *log-normal* (λογαριθμοκανονική), *weibull*, *gamma*, *exp* (εκθετική). Για την ορθότερη απόφασή μας θα συμβουλευτούμε τόσο το ιστόγραμμα της μεταβλητής που μας ενδιαφέρει με τις παραπάνω κατανομές όσο και τις τιμές AIC που θα πάρουμε για κάθε προσαρμογή μοντέλου. Για αρχή θα παρουσιαστούν συγκεντρωτικά τα ιστογράμματα για τους μήνες επιβίωσης και των τριών dataset και στην συνέχεια θα προσαρμόσουμε τα μοντέλα επιβίωσης που θεωρούμε ότι είναι πιο ικανά να περιγράψουν τα δεδομένα μας.

Για το καρκίνο του εγκεφάλου όπως βλέπουμε και από το παρακάτω σχήμα πιο “κοντά” είναι η κατανομή λογαριθμοκανονική με την πράσινη γραμμή. Παρατηρούμε επίσης κοντά την κόκκινη η οποία υποδηλώνει την Weibull κατανομή και την μωβ που δείχνει την gamma, ενώ πιο μακριά φαίνεται η μπλε, δηλαδή η εκθετική.



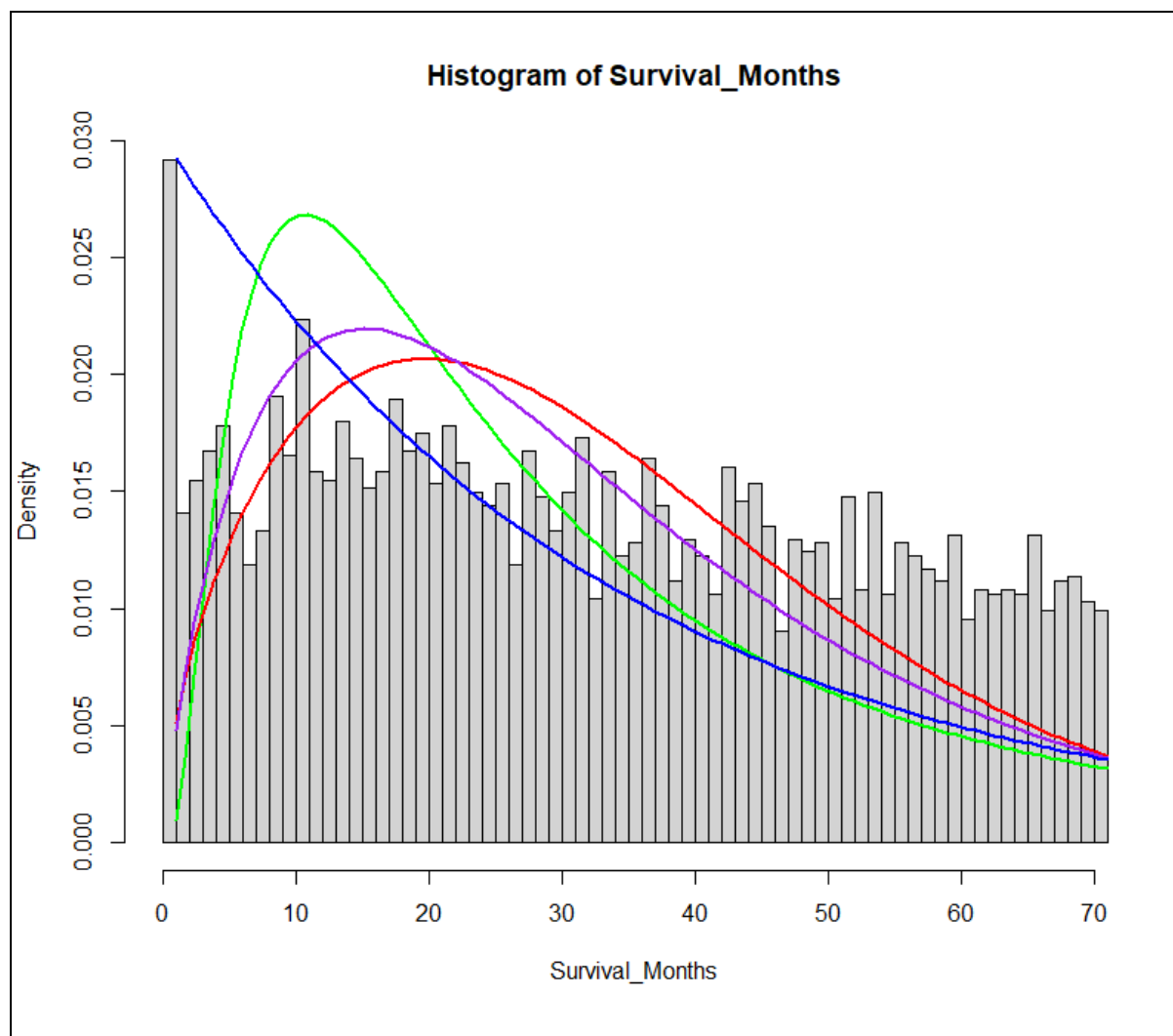
Διάγραμμα 5.1: Ιστόγραμμα χρόνου επιβίωσης (σε μήνες) ασθενών Καρκίνου του Εγκεφάλου

Τα ίδια αποτελέσματα θα παίρναμε καταγράφοντας τις τιμές AIC και από το αντίστοιχο κριτήριο θα καταλήγαμε στην λογαριθμοκανονική κατανομή καθώς έχει την μικρότερη τιμή.

	Log-normal	Weibull	Gamma	Exp
AIC	175084.8	175341.5	175660.7	177515.7

Πίνακας 5.1: Τιμές AIC για τα μοντέλα του καρκίνου του Εγκέφαλου

Ακολουθώντας την ίδια διαδικασία για τα δεδομένα του καρκίνου του μαστού, για την μεταβλητή απόκρισης «Survival Months» επιβεβαιώνουμε ότι η Weibull εφαρμόζει καλύτερα στα δεδομένα.

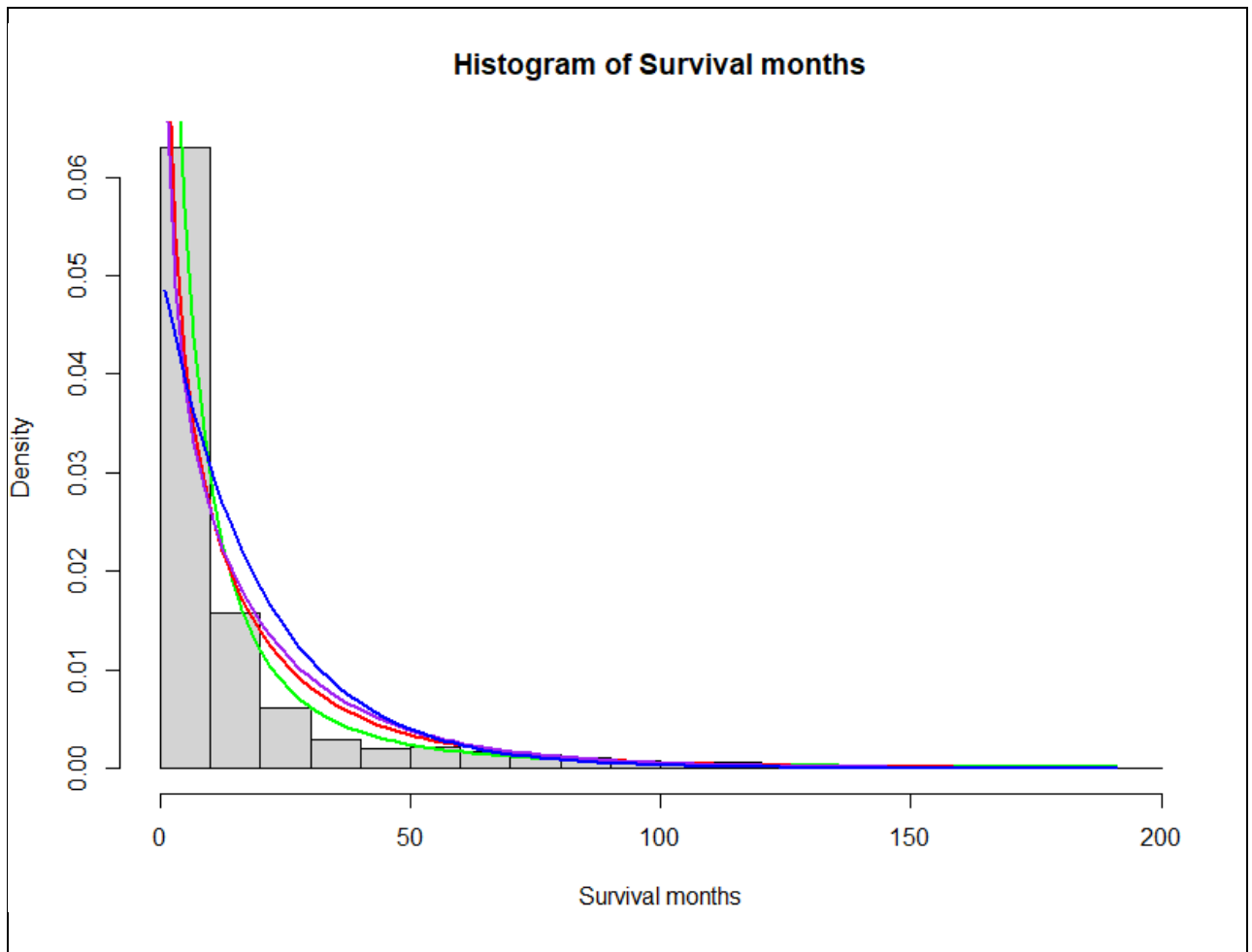


Διάγραμμα 5.2: Ιστόγραμμα χρόνου επιβίωσης (σε μήνες) ασθενών Καρκίνου του Μαστού

	Log-normal	Weibull	Gamma	Exp
AIC	49704.91	47906.68	48279.67	49306.38

Πίνακας 5.2: Τιμές AIC για τα μοντέλα του καρκίνου του Μαστού

Για το τελευταίο dataset, τον καρκίνο του παγκρέατος, έχουμε παρεμφερή αποτελέσματα με εκείνα του καρκίνου του εγκεφάλου. Από το ιστόγραμμα και τις προσαρμοσμένες κατανομές φαίνεται καταλληλότερη να είναι η log-normal κατανομή.



Διάγραμμα 5.3: Ιστόγραμμα χρόνου επιβίωσης (σε μήνες) ασθενών Καρκίνου του Παγκρέατος

	Log-normal	Weibull	Gamma	Exp
AIC	258005.9	263274.9	265238.8	268552.7

Πίνακας 5.3: Τιμές AIC για τα μοντέλα του καρκίνου του Παγκρέατος

5.2. Μοντέλο επιβίωσης καρκίνος του Εγκεφάλου

5.2.1. Το μοντέλου επιταχυνόμενου κινδύνου του καρκίνου του Εγκεφάλου

Όπως είδαμε στην παραπάνω προσαρμογή των παραμετρικών μοντέλων παραπάνω η κατανομή που ταιριάζει καλύτερα στα δεδομένα του καρκίνου του εγκεφάλου είναι η λογαριθμοκανονική. Για τον λόγο αυτόν θα ανατρέξουμε να βρούμε το μοντέλο επιταχυνόμενου χρόνου αποτυχίας (AFT) και τις μεταβλητές που είναι σημαντικές και επηρεάζουν τους μήνες επιβίωσης. Μετά από δοκιμές διαφορετικών μοντέλων AFT log-normal καταλήγουμε ότι εκείνο που ταιριάζει περισσότερο στα δεδομένα περιέχει τις μεταβλητές:

- Ηλικία,
- Φύλο,
- Φυλή,
- Βαθμός καρκίνου,
- τρόπος διάγνωσης καρκίνου,

- μέγεθος όγκου
- και πλευρικότητα.

Ενώ δεν συμπεριλαμβάνουμε στο μοντέλο την μεταβλητή που αναγράφει τον αριθμό των καρκινικών όγκων.

Θα παρουσιαστούν οι συντελεστές των μεταβλητών, το τυπικό σφάλμα καθώς και το p-value αυτών.

Variable	Value	Sig.	Std. Error
Intercept	4.90194	< 2e-16	0.278092
Age		< 2e-16	
Age (1-4 years)	0.249630	0.12886	0.164379
Age (5-9 years)	0.226591	0.16620	0.163661
Age (10-14 years)	0.644216	0.00015	0.169907
Age (15-19 years)	0.475233	0.00571	0.171946
Age (20-24 years)	0.359191	0.03335	0.168807
Age (25-29 years)	0.158235	0.33667	0.164697
Age (30-34 years)	0.050743	0.75520	0.162744
Age (35-39 years)	-0.261605	0.10603	0.161857
Age (40-44 years)	-0.754529	2.3e-06	0.159604
Age (45-49 years)	-1.112579	1.5e-12	0.157234
Age (50-54 years)	-1.585701	< 2e-16	0.156114
Age (55-59 years)	-1.832367	< 2e-16	0.155358
Age (60-64 years)	-2.088217	< 2e-16	0.155315
Age (65-69 years)	-2.319082	< 2e-16	0.155745
Age (70-74 years)	-2.620729	< 2e-16	0.156194
Age (75-79 years)	-2.882300	< 2e-16	0.156891
Age (80-84 years)	-3.193211	< 2e-16	0.158834
Age (85+ years)	-3.398871	< 2e-16	0.162328
Sex		0.00030	
Male	-0.061609	0.00335	0.021000
Race		0.00012	
Asian or Pacific	0.309751	0.00842	0.117566
Black	0.166661	0.16568	0.120228
Unknown	1.028310	5.1e-06	0.225384
White	0.218670	0.05078	0.111946
Grade		< 2e-16	
Grade I	0.303976	0.01884	0.129420
Grade III	-1.388926	< 2e-16	0.092567
Grade IV	-1.385773	< 2e-16	0.064968
Unknown	-1.243897	< 2e-16	0.063571
Diagnostic Confirmation		< 2e-16	
Άγνωστο	-0.279624	0.52817	0.443283

Απευθείας Οπτικοποίηση	-0.313476	0.60616	0.608031
Θετική Αποφολιδωτική Κυτταρολογία	0.187996	0.63504	0.396081
Θετική Ιστολογία	0.562777	0.00071	0.166146
Θετική Εργαστηριακή Δοκιμή	-0.317184	0.74379	0.970473
Θετική Μικροσκοπική Επιβεβαίωση	-1.385773	< 2e-16	0.064968
Ακτινογραφία	0.186318	0.26985	0.168859
Tumor Size		< 2e-16	
Tumor Size	-0.008438	< 2e-16	0.000585
Πλευρικότητα		5.585e-11	
Αριστερή Προέλευση	0.620578	7.3e-10	0.100748
Μη Ζευγοποιημένος	0.509515	6.1e-07	0.102129
Απροσδιόριστη	0.762719	0.02229	0.333732
Ζεύγος (χωρίς ξεκάθαρη πλευρικότητα)	0.469231	0.00367	0.161520
Μέσης Γραμμής Όγκος	0.493137	0.17054	0.359827
Δεξιά Προέλευση	0.643944	1.6e-10	0.100642

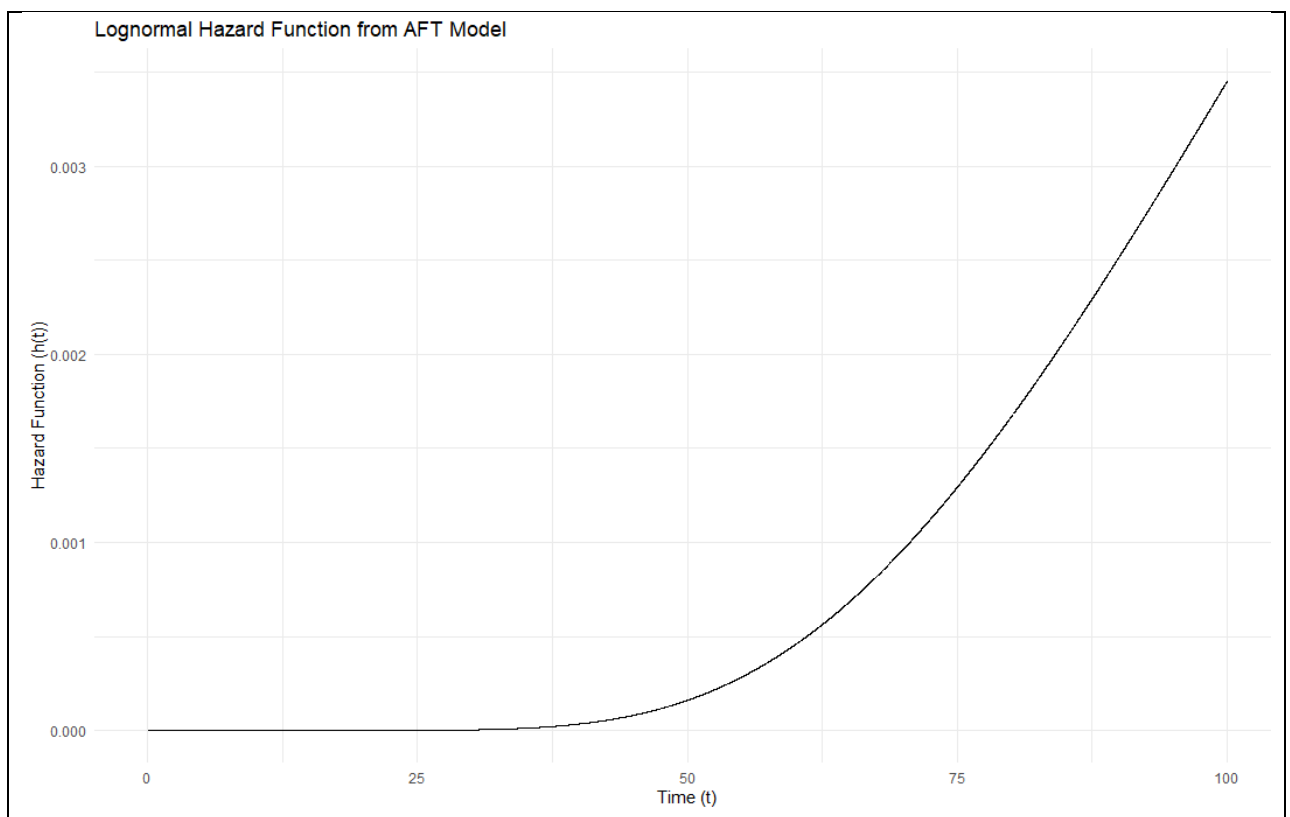
Πίνακας 5.4: Εκτιμήσεις του AFT μοντέλου για τον καρκίνο του Εγκεφάλου

Ερμηνεύοντας τους παραπάνω συντελεστές του λογαριθμοκανονικού μοντέλου μπορούμε να ισχυριστούμε σε ένα γενικό πλαίσιο ότι οι θετικοί συντελεστές δείχνουν μεγαλύτερο αναμενόμενο log-χρόνο επιβίωσης σε σχέση με την ομάδα αναφοράς ενώ οι αρνητικοί συντελεστές δείχνουν μικρότερη διάρκεια επιβίωσης. Πιο συγκεκριμένα:

- Η σταθερά (Intercept) είναι 4.90194 με p-value < 2e-16. Αυτός ο συντελεστής αντιπροσωπεύει την αναμενόμενη λογαριθμική επιβίωση όταν όλες οι άλλες μεταβλητές είναι μηδενικές.
- Για την ηλικιακή ομάδα 50-54 χρονών Ο συντελεστής -1.585701 (p-value < 2e-16) δείχνει ότι η επιβίωση αυτής της ηλικιακής ομάδας είναι σημαντικά μειωμένη κατά 1.585701 μονάδες σε σχέση με την ομάδα αναφοράς.
- Ο συντελεστής για το φύλο είναι -0.061609 (p-value 0.00335) για τους άνδρες. Αυτό δείχνει ότι οι άνδρες έχουν ελαφρώς χαμηλότερη επιβίωση από τις γυναίκες κατά 0.061609 μονάδες.
- Για την «Άγνωστη» φυλή, έχουμε ότι η επιβίωση αυτής της ομάδας είναι σημαντικά αυξημένη κατά 1.028310 μονάδες σε σχέση με την ομάδα αναφοράς, δηλαδή τους «Λευκούς».
- Ο συντελεστής του «Βαθμού Ι» είναι 0.303976 που δηλώνει ότι είναι σημαντικά αυξημένη κατά 0.303976 μονάδες σε σχέση με τους ασθενείς «Βαθμού ΙΙ».
- Ο συντελεστής 0.562777 (p-value 0.00071) δείχνει ότι η επιβίωση αυτής της ομάδας είναι σημαντικά αυξημένη κατά 0.562777 μονάδες σε σχέση με την ομάδα που διαγνώστηκε κλινικά.

- Για το μέγεθος του όγκου καταλαβαίνουμε ότι για κάθε μονάδα αύξησης στο μέγεθος του όγκου, η επιβίωση μειώνεται κατά 0.008438 μονάδες, υποδηλώνοντας αρνητική και σημαντική επίδραση.
- Αναφορικά με την πλευρικότητα του όγκου, ο συντελεστής «Αριστερή Προέλευση» μας δείχνει ότι η επιβίωση αυτής της ομάδας είναι σημαντικά αυξημένη κατά 0.620578 μονάδες σε σχέση με την διμερή πλευρικότητα.

Παρουσιάζεται η συνάρτηση κινδύνου για το AFT μοντέλο η οποία μας υποδεικνύει ότι η επιλογή της κατανομής είναι κατάλληλη για την μοντελοποίηση αυτών των δεδομένων. Στην αρχή (κοντά στο χρόνο 0) ο ρυθμός κινδύνου είναι πολύ χαμηλός, σχεδόν μηδενικός. Αυτό υποδηλώνει ότι η πιθανότητα να συμβεί ο θάνατος αμέσως μετά την έναρξη είναι ελάχιστη. Καθώς ο χρόνος προχωρά η συνάρτηση κινδύνου αυξάνεται σταδιακά, υποδεικνύοντας ότι ο κίνδυνος του θανάτου αυξάνεται με την πάροδο του χρόνου. Αυτό είναι τυπικό για τη λογαριθμική κανονική κατανομή, όπου ο ρυθμός κινδύνου αυξάνεται καθώς περνά περισσότερος χρόνος. Σε μακροπρόθεσμο ορίζοντα, η συνάρτηση κινδύνου συνεχίζει να αυξάνεται, δηλαδή ο κίνδυνος του γεγονότος γίνεται μεγαλύτερος όσο περισσότερο επιβιώνει το άτομο.



Διάγραμμα 5.4: Διάγραμμα Συνάρτησης Κινδύνου AFT μοντέλου καρκίνου εγκεφάλου

Τέλος, θα εκτελεστεί ο έλεγχος λόγου πιθανοφάνειας (Likelihood Ratio Test) για να αξιολογηθεί η καταλληλότητα του προσαρμοσμένου μοντέλου επιταχυνόμενης αποτυχίας χρόνου (Accelerated Failure Time - AFT) με την κατανομή lognormal. Συγκεκριμένα, το μοντέλο, το οποίο περιλάμβανε τις μεταβλητές Age, Sex, ace, Grade, Diagnostic Confirmation, Tumor Size, και Laterality, συγκρίθηκε με το μηδενικό μοντέλο, το οποίο

περιλάμβανε μόνο τη σταθερά (intercept) και τα αποτελέσματα φαίνονται στον παρακάτω πίνακα:

Μοντέλο	Resid. Df	-2*LL	Test Df	Deviance	Pr(>Chi)
Null Model	18911	130478.7	NA	NA	NA
AFT log-normal Model	18870	118836.4	41	11642.26	0

Πίνακας 5.5: Έλεγχος καταλληλότητας AFT μοντέλου για τον καρκίνο του Εγκεφάλου

Αναλύοντας τα στοιχεία από τον πίνακα καταλήγουμε ότι οι βαθμοί ελευθερίας για το μηδενικό μοντέλο είναι 18911 ενώ για το πλήρες είναι 18870, με διαφορά μεταξύ των 2 μοντέλων, δηλαδή οι βαθμοί ελευθερίας του Τεστ είναι 41. Το p-value για τον έλεγχο χ^2 είναι 0 (πολύ μικρό) το οποίο υποδηλώνει ότι το πλήρες μοντέλο είναι σημαντικά καλύτερο από το μηδενικό, υποστηρίζοντας την προσθήκη των μεταβλητών που είδαμε παραπάνω. Αυτό σημαίνει αυτές οι μεταβλητές συνεισφέρουν σημαντικά στην εξήγηση της μεταβλητής εξόδου (Survival Months και Status). Συνεπώς, το πλήρες μοντέλο είναι κατάλληλο και παρέχει σημαντικές πληροφορίες για τη μελέτη της επιβίωσης των ασθενών του καρκίνου του Εγκεφάλου.

5.2.2. Το μοντέλου αναλογικού κινδύνου του καρκίνου του Εγκεφάλου

Σε δεύτερο χρόνο για τον καρκίνο του εγκεφάλου θα ψάξουμε το αναλογικό μοντέλο του Gompertz που ταιριάζει καλύτερα στα δεδομένα μας. Παρατηρούμε πως οι σημαντικές μεταβλητές είναι οι ίδιες με το μοντέλου του επιταχυνόμενου κινδύνου, δηλαδή:

- ηλικία,
- φύλο,
- φυλή,
- βαθμός καρκίνου,
- τρόπος διάγνωσης καρκίνου,
- μέγεθος όγκου
- και πλευρικότητα.

Variable	B	Exp(B)	Sig.	Std. Error
Age				
Age (1-4 years)	0.093349655	0.910875	0.6172906	0.1868136570
Age (5-9 years)	0.028257139	1.02866	0.8785282	0.1848862491
Age (10-14 years)	0.364864441	0.694291	0.06212202	0.1955933240
Age (15-19 years)	0.194416082	0.823315	0.3215416	0.1961233612
Age (20-24 years)	0.031346168	1.031843	0.8684747	0.1892927331
Age (25-29 years)	0.217665710	1.243171	0.2361492	0.1837357909
Age (30-34 years)	0.265432445	1.303995	0.1443741	0.1818411334
Age (35-39 years)	0.537714262	1.712089	0.02819208	0.1800306612
Age (40-44 years)	0.901805084	2.464047	3.71727e-07	0.1774196109
Age (45-49 years)	1.175774558	3.240652	2.06945e-11	0.1754615727
Age (50-54 years)	1.574640885	4.829007	1.81733e-19	0.1744993128

Age (55-59 years)	1.738748329	5.690217	1.68948e-23	0.1740526310
Age (60-64 years)	1.951897979	7.042041	3.31147e-29	0.1739905096
Age (65-69 years)	2.128185629	8.399613	2.47509e-34	0.1741768453
Age (70-74 years)	2.362289551	1.615228	8.95783e-42	0.1744547658
Age (75-79 years)	2.609403924	1.590947	2.16936e-50	0.1748002633
Age (80-84 years)	2.879263314	1.801154	2.85536e-60	0.1758261116
Age (85+ years)	3.178734284	2.016336	1.07567e-71	0.1775321581
Sex				
Male	0.073583163	1.076358	1.84892e-05	0.0171827115
Race				
Asian or Pacific	0.279453966	0.756197	0.004019971	0.0971476403
Black	0.159602905	0.852482	0.108465	0.0994329543
Unknown	0.820663029	0.440140	0.000411523	0.2323125518
White	0.144530151	0.865429	0.1177898	0.0924035889
Grade				
Grade I	0.290283261	0.748052	0.03553789	0.1380870613D
Grade III	1.059533997	2.885026	2.98335e-38	0.0819333763
Grade IV	1.047109979	2.849404	8.80726e-63	0.0626123279
Unknown	0.958056407	2.606625	1.85499e-54	0.0616503285
Diagnostic Confirmation				
Άγνωστο	0.609494403	1.839501	0.07493298	0.3422444690
Απευθείας Οπτικοποίηση	0.605912058	1.832923	0.2419899	0.5178613389
Θετική Αποφολιδωτική Κυτταρολογία	0.282284727	0.754059	0.3916759	0.3295477595
Θετική Ιστολογία	0.353300366	0.702366	0.007666275	0.1324996529
Θετική Εργαστηριακή Δοκιμή	0.798171702	2.221476	0.269294	0.7225301318
Θετική Μικροσκοπική Επιβεβαίωση	0.259228270	1.29593	0.661718	0.5924643022
Ακτινογραφία	0.230510186	0.794128	0.08777024	0.1350161324
Tumor Size				
Tumor Size	0.006555515	1.006577	1.65929e-57	0.0004101369
Πλευρικότητα				
Αριστερή Προέλευση	0.439027801	0.644663	1.98499e-08	0.0782120085
Μη Ζευγοποιημένος	0.387569333	0.678705	1.14835e-06	0.0796755101
Απροσδιόριστη	0.646592295	0.523828	0.03083984	0.2994684755

Ζεύγος (χωρίς ξεκάθαρη πλευρικότητα)	0.294970001	0.744554	0.03266058	0.1380796254
Μέσης Γραμμής Όγκος	0.836993196	0.433011	0.01014087	0.3255530623
Δεξιά Προέλευση	0.449212515	0.638130	8.79356e-09	0.0780900239

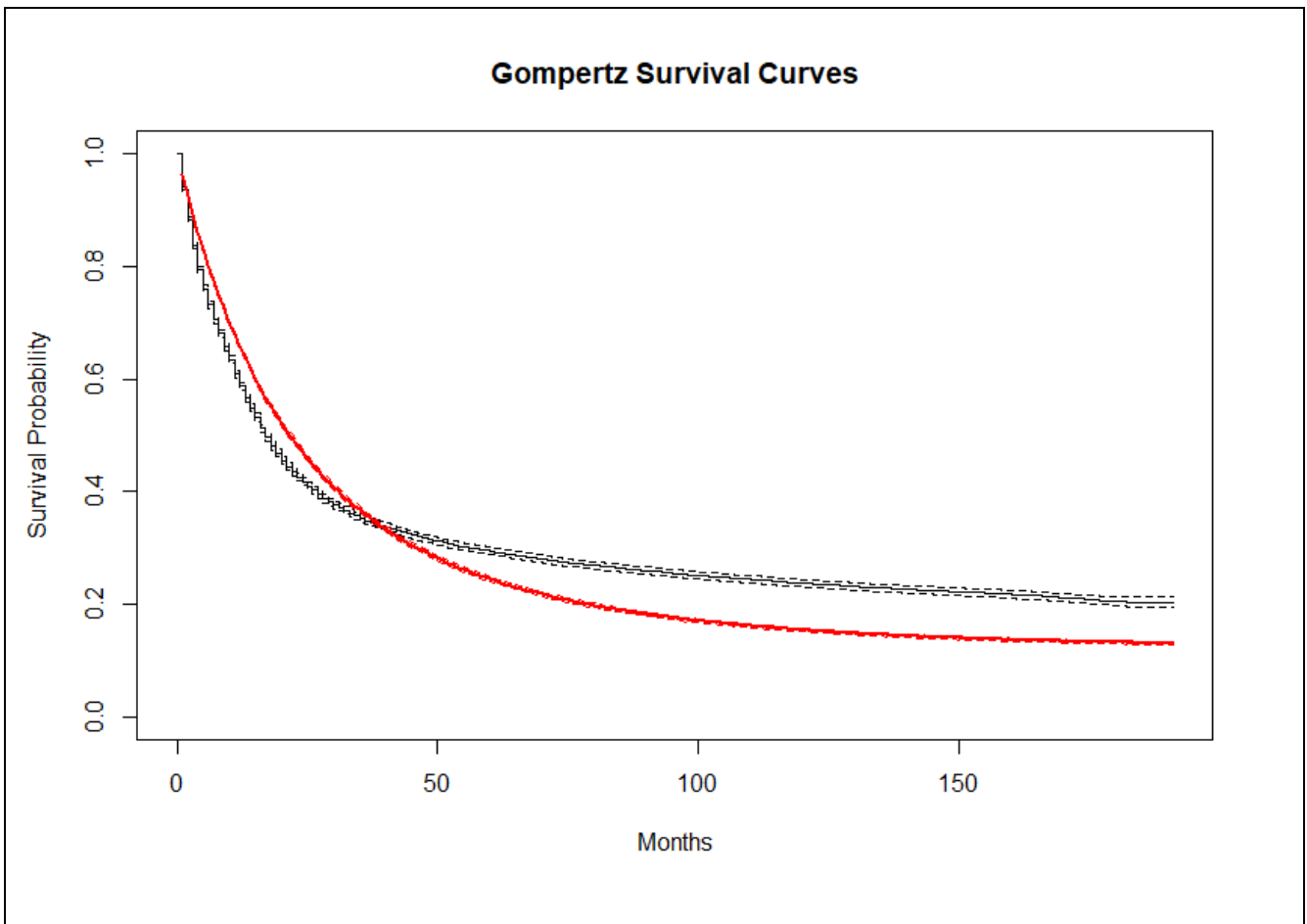
Πίνακας 5.6: Εκτιμήσεις του αναλογικού μοντέλου για τον καρκίνο του Εγκεφάλου

Για την ερμηνεία των μεταβλητών του αναλογικού μοντέλου του Gompertz θα συμβουλευτούμε το \exp των συντελεστών. Πιο συγκεκριμένα:

- Για την ηλικία και την κατηγορία 35-39 ετών προβλέπεται αύξηση κινδύνου κατά $\exp(0.5377) = 1.7120$, δηλαδή 71.21% περισσότερο.
- Για το φύλο έχουμε ότι οι άνδρες έχουν αυξημένο κίνδυνο σε σχέση με τις γυναίκες και μάλιστα $\exp(0.0735) = 1.0763$ φορές περισσότερες.
- Για την φυλή παρατηρούμε ότι εκείνοι με άγνωστη εγγραφή του πεδίου έχουν $\exp(-0.82) = 0.44$ φορές μικρότερο κίνδυνο από τους Αμερικάνους.
- Για τον βαθμό του καρκίνου διαπιστώθηκε ότι εκείνοι με Βαθμό I έχουν $\exp(-0.958) = 2.6$ λιγότερο κίνδυνο σε σχέση με εκείνους που έχουν διαγνωστεί με καρκίνο Βαθμού II, δηλαδή -25.19% μείωση στον κίνδυνο.
- Ομοίως, εκείνοι που έχουν διαγνωστική επιβεβαίωση με ακτινογραφία έχουν 0.79, 20.59% μικρότερο κίνδυνο σε σύγκριση με εκείνους που επιβεβαίωσαν με κλινική διάγνωση.
- Επίσης, όπως περιμέναμε όσο μεγαλύτερος είναι το μέγεθος του όγκου, τόσο μεγαλύτερος είναι ο κίνδυνος.
- Τέλος, εκείνοι με δεξιά προέλευση έχουν μειωμένο κίνδυνο κατά $\exp(-0.4492) = 0.63$ φορές σε σχέση με τους ασθενείς με διμερής πλευρικότητα.

Όπως και για το λογαριθμοκανονικό μοντέλο, πιο πάνω, θα εκτελεστεί ο έλεγχος για την καταλληλότητα του μοντέλου. Αναλυτικότερα, ο έλεγχος χ^2 για το Gompertz μοντέλο συγκρίνει την πιθανότητα του προσαρμοσμένου μοντέλου με την πιθανότητα του κενού μοντέλου. Η στατιστική χ^2 είναι ίση με το διπλάσιο της διαφοράς των λογαριθμικών πιθανοτήτων των δύο μοντέλων. Αν η στατιστική τιμή του X^2 είναι μεγάλη και το p-value μικρότερο από το 0.05, μπορούμε να πούμε ότι το προσαρμοσμένο μοντέλο είναι σημαντικά καλύτερο από το μοντέλο που περιέχει μόνο τον σταθερό όρο. Η τιμή X^2 Test Statistic είναι 11194.7, η οποίος αποτελεί ένα μέτρο της διαφοράς μεταξύ των παρατηρηθέντων και των αναμενόμενων συχνοτήτων στα δεδομένα μας. Σε αυτό το πλαίσιο, η τιμή αυτή υποδεικνύει ότι υπάρχει σημαντική απόκλιση από την μηδενική υπόθεση, η οποία συχνά προϋποθέτει ότι οι μεταβλητές που εξετάζονται είναι ανεξάρτητες. Το p-value που υπολογίστηκε είναι 0, γεγονός που σημαίνει ότι η πιθανότητα να παρατηρήσουμε ένα τόσο ακραίο ή πιο ακραίο αποτέλεσμα, δεδομένης της μηδενικής υπόθεσης, είναι ουσιαστικά μηδενική. Αυτό μας οδηγεί στο να απορρίψουμε τη μηδενική υπόθεση με υψηλό επίπεδο εμπιστοσύνης. Κατά συνέπεια, μπορούμε να συμπεράνουμε ότι υπάρχει ισχυρή στατιστική ένδειξη για την ύπαρξη σχέσης μεταξύ των εξεταζόμενων μεταβλητών.

Τέλος, για την ανάλυση των δεδομένων του καρκίνου του εγκεφάλου απεικονίζοντας τις καμπύλες επιβίωσης και πώς μεταβάλλεται η πιθανότητα επιβίωσης με την πάροδο του χρόνου. Οι μαύρες γραμμές αντιπροσωπεύουν την προβλεπόμενη καμπύλη επιβίωσης από το Gompertz μοντέλο. Επίσης, οι διακεκομμένες μαύρες γραμμές γύρω από την κεντρική μαύρη γραμμή δείχνουν το 95% διάστημα εμπιστοσύνης για την καμπύλη επιβίωσης, παρέχοντας μια ιδέα για την αβεβαιότητα των προβλέψεων του μοντέλου. Οι κόκκινες γραμμές αντιπροσωπεύουν τα παρατηρούμενα δεδομένα επιβίωσης από το dataset. Αυτές είναι οι πραγματικές τιμές επιβίωσης που έχουν καταγραφεί. Παρατηρείται ότι οι κόκκινες γραμμές βρίσκονται κοντά στις προβλέψεις του μοντέλου, γεγονός που υποδηλώνει καλή προσαρμογή του μοντέλου στα δεδομένα.



Διάγραμμα 5.5: Διάγραμμα Συνάρτησης Επιβίωσης PH Gompertz μοντέλου καρκίνου παγκρέατος

5.3. Μοντέλο επιβίωσης καρκίνος του Μαστού

5.3.1. Το μοντέλου επιταχυνόμενου κινδύνου του καρκίνου του Μαστού.

Όπως διαπιστώθηκε από τα παραπάνω για την κατανομή Weibull μπορεί να χρησιμοποιηθεί τόσο το AFT είτε το PH μοντέλο. Τρέχοντας διαφορετικά μοντέλα επιταχυνόμενου κινδύνου οι μεταβλητές που φάνηκαν σημαντικά συναρτησει με τους μήνες επιβίωσης είναι:

- η ηλικία,
- ο βαθμός,
- η οικογενειακή κατάσταση,

- ο δευτερεύων τύπος μαστού,
- ο υποδοχέας οιστρογόνου,
- ο υποδοχέας προγεστερόνης
- και ο αριθμός λεμφαδένων με καρκίνο.

Σύμφωνα με τις τιμές των p value παρατηρείται μη εξάρτηση από τις υπόλοιπες μεταβλητές όπως:

- η φυλή,
- το μέγεθος του όγκου
- και η πλευρικότητα.

Στο μοντέλο AFT με κατανομή Weibull, οι συντελεστές αντιπροσωπεύουν την αλλαγή στον χρόνο επιβίωσης (ή στην επιβίωση) ως αποτέλεσμα της αλλαγής ενός μοναδικού μονομεταβλητού. Οι συντελεστές αυτοί έχουν άμεση ανταπόκριση της μεταβλητής στον χρόνο επιβίωσης. Πιο συγκεκριμένα:

- Οι ασθενείς που έχουν ηλικία 75-79 έχουν 8.8% ($\exp(b)=0.9119$) μικρότερο κίνδυνο να αποβιώσουν από καρκίνο μαστού σε σχέση με τους ασθενείς που έχουν ηλικία 20-24, την κατηγορία αναφοράς.
- Οι ασθενείς που έχουν βαθμό καρκίνου III έχουν 38.7% ($\exp(b)=0.6127$) μικρότερο κίνδυνο να αποβιώσουν από καρκίνο μαστού σε σχέση με ασθενείς που έχουν βαθμό καρκίνου II.
- Οι ασθενείς που βρίσκονται σε χηρεία έχουν 13.9% ($\exp(b)= 0.8602$) μικρότερο κίνδυνο να αποβιώσουν σε σχέση με τους ασθενείς που είναι χωρισμένοι.
- Οι ασθενείς με υποτύπο μαστού Triple N έχουν 50.7% ($\exp(b)= 1.5067$) μεγαλύτερο κίνδυνο να πεθάνουν από καρκίνο μαστού σε σχέση με τους ασθενείς με υποτύπο μαστού Her2-/HR+.
- Οι ασθενείς με θετικό υποδοχέα οιστρογόνου έχουν 129.3% ($\exp(b)=2.2920$) μεγαλύτερο χρόνο επιβίωσης σε σχέση με ασθενείς που έχουν αρνητικό υποδοχέα οιστρογόνου.
- Οι ασθενείς με θετικό υποδοχέα προγεστερόνης έχουν 50.8% ($\exp(b)=1.5078$) μεγαλύτερο αναμενόμενο χρόνο επιβίωσης από καρκίνο μαστού σε σχέση με ασθενείς που έχουν αρνητικό υποδοχέα προγεστερόνης.
- Τέλος, οι ασθενείς του επιπέδου N3 έχουν 30.6% ($\exp(b)=0.1368$) λιγότερες πιθανότητες να έχουν περισσότερους μήνες επιβίωσης σε σχέση με εκείνους του σταδίου N0.

Variable	B	Exp(B)	Sig.	Std. Error
ER Status			<2.2e-16	
ER Status (Positive)	0.8300	2.2920	8.8e-05	0.2117
PR Status			1.094e-08	
PR Status (Positive)	0.4107	1.5078	3.9e-05	0.0998
Age			<2.2e-16	
Age (25-29 years)	0.8390	2.3142	0.3238	0.8504

Age (30-34 years)	0.8131	2.2541	0.3127	0.8054
Age (35-39 years)	1.4183	4.1301	0.0799	0.8099
Age (40-44 years)	1.2224	3.3953	0.1224	0.7914
Age (45-49 years)	0.9670	2.6307	0.2175	0.7841
Age (50-54 years)	0.9242	2.5206	0.2373	0.7821
Age (55-59 years)	0.9318	2.5397	0.2344	0.7836
Age (60-64 years)	0.6966	2.0069	0.3719	0.7802
Age (65-69 years)	0.8326	2.2992	0.2877	0.7831
Age (70-74 years)	0.5119	1.6688	0.5133	0.7831
Age (75-79 years)	-0.0921	0.9119	0.9061	0.7805
Age (80-84 years)	-0.0712	0.9313	0.9276	0.7842
Age (85+ years)	-0.5765	0.5616	0.4609	0.7818
Grade			2.722e-10	
Grade I	0.0278	1.0282	0.8318	0.1308
Grade III	-0.4898	0.6127	5.1e-09	0.0838
Marital Status			0.0008597	
Marital Status (Married)	0.0730	1.0757	0.5120	0.1114
Marital Status (Separated)	-0.0957	0.9088	0.7464	0.2961
Marital Status (Single)	-0.2853	0.7516	0.0298	0.1313
Marital Status (Unmarried)	-0.4267	0.6527	0.2848	0.3990
Marital Status (Widowed)	-0.1505	0.8602	0.2520	0.1314
Breast SubType			1.472e-09	
Breast SubType (Her2+/HR-)	1.2773	3.5869	3.4e-06	0.2750
Breast SubType (Her2+/HR+)	0.2969	1.3456	0.0191	0.1266
Breast SubType (Triple Negative)	0.4100	1.5067	0.0857	0.2386
Stage N			7.652e-10	
Stage N (N1)	0.2611	1.2981	0.0310	0.1211
Stage N (N2)	-0.0397	0.9611	0.7725	0.1373
Stage N (N3)	-0.3646	0.6947	0.0077	0.1368

Πίνακας 5.7: Εκτιμήσεις του AFT μοντέλου για τον καρκίνο του Μαστού

Για την αξιολόγηση του παραπάνω AFT Weibull μοντέλου εκτελέστηκε ο κατάλληλος έλεγχος με το κενό μοντέλο. Από τον πίνακα ANOVA συμπεραίνουμε ότι το κενό μοντέλο είχε $-2\text{Log-Likelihood} (-2LL)$ ίσο με 7496.693, ενώ το πλήρες μοντέλο, που περιλαμβάνει όλες τις μεταβλητές, είχε $-2*LL$ ίσο με 6823.698. Η διαφορά στη διακύμανση (Deviance) ήταν

672.9951 με 44 βαθμούς ελευθερίας και η τιμή του p-value ήταν 1.77e-113. Αυτά τα αποτελέσματα δείχνουν ότι το πλήρες μοντέλο είναι στατιστικά σημαντικά καλύτερο από το κενό μοντέλο (p-value < 0.001), υποδηλώνοντας ότι οι συμπεριλαμβανόμενοι παράγοντες βελτιώνουν σημαντικά την ακρίβεια πρόβλεψης της επιβίωσης. Συνολικά, το πλήρες μοντέλο προσφέρει μια πιο ακριβή εξήγηση της επιβίωσης σε σχέση με το κενό μοντέλο.

Μοντέλο	Resid. Df	-2*LL	Test Df	Deviance	Pr(>Chi)
Null Model	5474	7496.693	NA	NA	NA
AFT Weibull Model	5430	6823.698	44	672.9951	1.7665e-113

Πίνακας 5.8: Έλεγχος καταλληλότητας AFT μοντέλου για τον καρκίνο του Μαστού

5.3.2. Το μοντέλου αναλογικού κινδύνου του καρκίνου του Μαστού.

Από την άλλη μεριά, ακολουθώντας την ίδια διαδικασία για διαφορετικά μοντέλα αναλογικού κινδύνου καταλήγουμε ότι οι πιο σημαντικές μεταβλητές είναι:

- η ηλικία,
- ο βαθμός,
- η οικογενειακή κατάσταση,
- ο δευτερεύων τύπος μαστού,
- ο υποδοχέας οιστρογόνου,
- ο υποδοχέας προγεστερόνης,
- ο αριθμός λεμφαδένων με καρκίνο
- και ο αριθμός που εκφράζει κατά πόσο ο όγκος έχει αναπτυχθεί σε κοντινούς ιστούς.

Όπως και λίγο παραπάνω θα δοθεί η ερμηνεία των συντελεστών των πιο σημαντικών μεταβλητών. Στο PH μοντέλο, οι συντελεστές (ή αναλογίες κινδύνου - hazard ratio) αντιπροσωπεύουν την εκτίμηση του λογαρίθμου της αναλογίας κινδύνου μεταξύ δύο ομάδων που διαφέρουν κατά μία μονάδα στην τιμή της μεταβλητής, κρατώντας σταθερές τις τιμές των υπολοίπων μεταβλητών. Με άλλα λόγια, οι αναλογίες κινδύνου δείχνουν πώς μια αλλαγή στην τιμή μιας μεταβλητής σχετίζεται με την αλλαγή στον κίνδυνο (λογαριθμικά). Υπολογίζοντας ξανά το $\exp(b)$ των μεταβλητών έχουμε τις ακόλουθες ερμηνείες. Αρχικά.

- Οι ασθενείς με θετικό ER status έχουν κατά μέσο όρο 0.634 ($\exp(0.45565) = 0.63403$) φορές τον κίνδυνο θανάτου σε σχέση με εκείνους με αρνητικό ER status, διατηρώντας σταθερές τις άλλες μεταβλητές. Αυτό σημαίνει ότι ο κίνδυνος θανάτου είναι κατά 36% χαμηλότερος για τους ασθενείς με θετικό ER status.
- Για τους ασθενείς με θετικό PR status, ο κίνδυνος θανάτου είναι περίπου 0.714 ($\exp(-0.3372) = 0.7137$) φορές (ή 29% χαμηλότερος) σε σχέση με τους ασθενείς με αρνητικό PR status.
- Αναφορικά με την ηλικία, για παράδειγμα, η ηλικιακή ομάδα "Age35-39 years" έχει $\exp(-2.17893) = 0.11316$, που σημαίνει ότι αυτοί οι ασθενείς έχουν 0.113 φορές (ή περίπου 89% χαμηλότερο) τον κίνδυνο θανάτου σε σχέση με την ομάδα αναφοράς, 20–24 ετών.

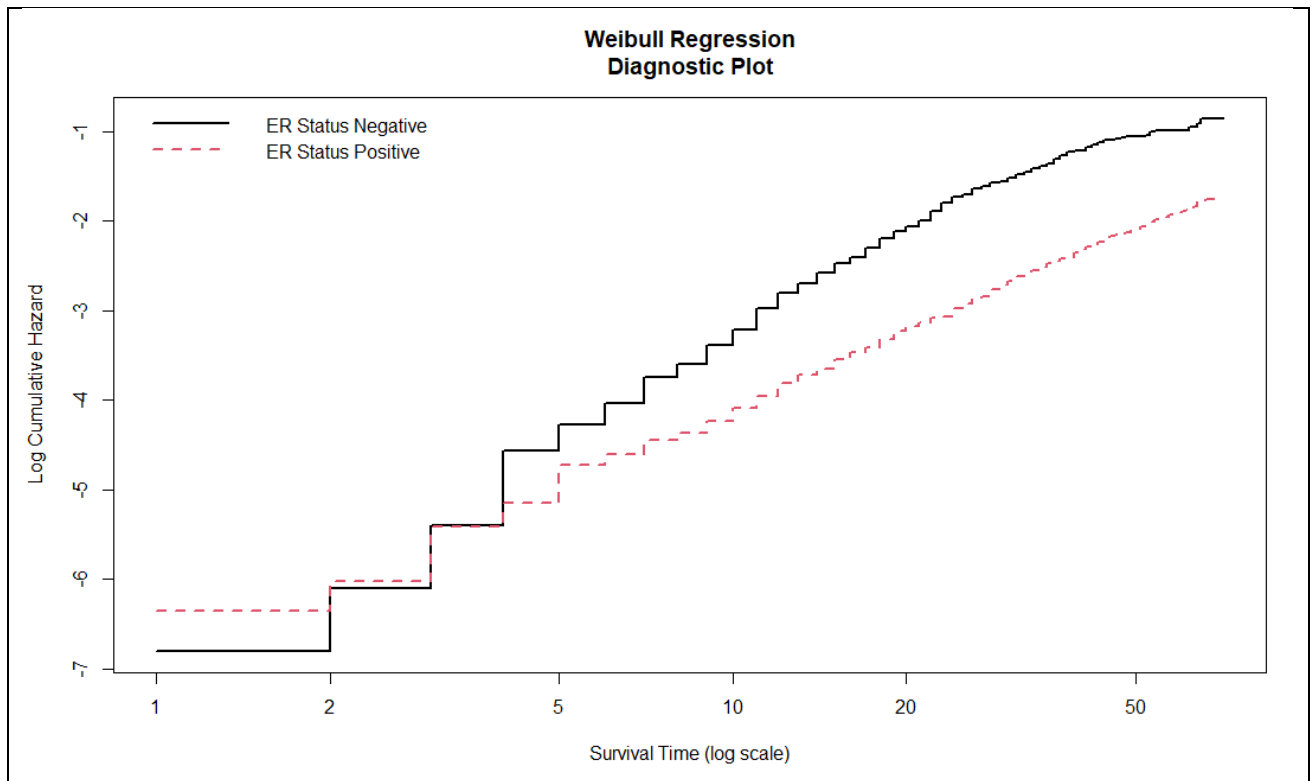
- Η κατηγορία "Grade III" έχει $\exp(0.5573)=1.7459$. δείχνει ότι οι ασθενείς με αυτό το βαθμό III καρκίνου έχουν περίπου 1.746 φορές μεγαλύτερο κίνδυνο θανάτου σε σύγκριση με τους ασθενείς στο επίπεδο II καρκίνου.
- Η κατηγορία του οικογενειακού καθεστώτος επηρεάζει επίσης τον κίνδυνο θανάτου. Για παράδειγμα, οι ασθενείς που είναι "Ελεύθερες" έχουν $\exp(0.44227)=1.55624$. που σημαίνει ότι ο κίνδυνος θανάτου είναι περίπου 1.556 φορές μεγαλύτερος σε σχέση με τους χωρισμένες ασθενείς.
- Η κατηγορία "T4c" έχει $\exp(2.14203)=8.51670$. που υποδεικνύει ότι οι ασθενείς σε αυτό το στάδιο έχουν περίπου 8.517 φορές μεγαλύτερο κίνδυνο θανάτου σε σύγκριση με αυτούς στο αναφερόμενο στάδιο. το T1.
- Τέλος, η κατηγορία Stage NN3 έχει συντελεστή $\exp(0.4641)=1.5905$, που σημαίνει ότι οι ασθενείς σε αυτή την κατηγορία έχουν 1.5905 φορές (ή περίπου 59% υψηλότερο) τον κίνδυνο θανάτου σε σύγκριση με την κατηγορία αναφοράς N0.

Variable	B	Exp(B)	Sig.	Std. Error
ER Status				
ER Status (Positive)	-0.45565	0.63403	0.000464	0.13017
PR Status				
PR Status (Positive)	-0.33722	0.71376	0.005779	0.12218
Age				
Age (25-29 years)	-133.229	0.26387	0.226105	110.065
Age (30-34 years)	-144.380	0.23603	0.166992	104.476
Age (35-39 years)	-217.893	0.11316	0.037640	104.819
Age (40-44 years)	-192.521	0.14584	0.060378	102.513
Age (45-49 years)	-161.445	0.19900	0.112106	101.614
Age (50-54 years)	-155.515	0.21116	0.124991	101.368
Age (55-59 years)	-155.080	0.21208	0.126677	101.536
Age (60-64 years)	-121.936	0.29542	0.228223	101.196
Age (65-69 years)	-143.088	0.23910	0.158883	101.565
Age (70-74 years)	-0.94459	0.38884	0.352315	101.556
Age (75-79 years)	-0.18843	0.82826	0.852458	101.317
Age (80-84 years)	-0.25130	0.77779	0.805035	101.809
Age (85+ years)	0.39551	148.515	0.696778	101.499
Grade				
Grade I	0.09683	110.167	0.573030	0.17180
Grade III	0.55732	174.598	1.23e-07	0.10537
Marital Status				
Marital Status (Married)	-0.02808	0.97231	0.846103	0.14465
Marital Status (Separated)	0.29857	134.793	0.434728	0.38223
Marital Status (Single)	0.44227	155.624	0.008904	0.16908

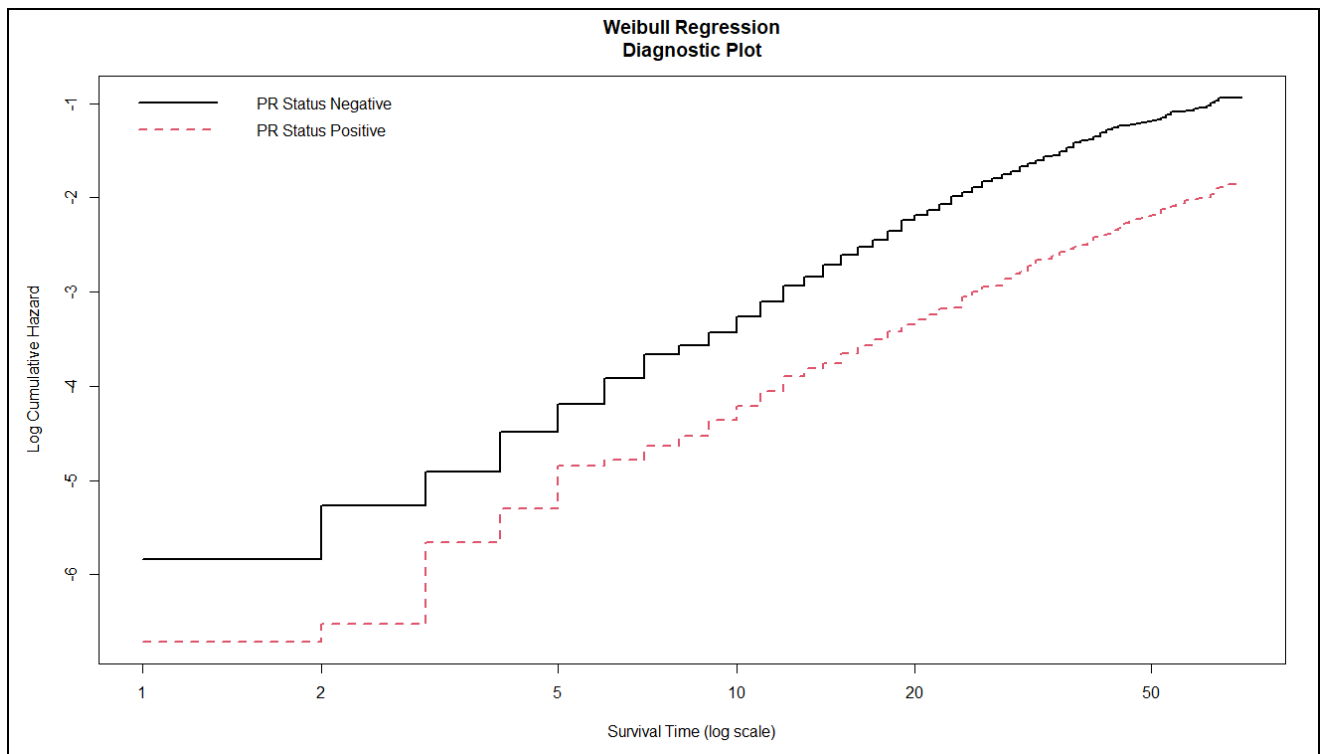
Marital Status (Unmarried)	0.67446	196.297	0.194658	0.52005
Marital Status (Widowed)	0.19818	121.918	0.247826	0.17149
Breast SubType				
Breast SubType (Her2+/HR-)	-161.135	0.19962	5.63e-06	0.35495
Breast SubType (Her2+/HR+)	-0.43086	0.64995	0.008856	0.16460
Breast SubType (Triple Negative)	-0.45362	0.63532	0.146943	0.31275
Stage N				
Stage N (N1)	-0.16867	0.84479	0.307183	0.16517
Stage N (N2)	0.09502	109.968	0.606340	0.18440
Stage N (N3)	0.46411	159.059	0.011397	0.18342
Stage T				
Stage TT1b	0.36700	144.339	0.624205	0.74913
Stage TT1c	0.19090	121.034	0.790388	0.71820
Stage TT2	0.69762	200.896	0.328359	0.71373
Stage TT3	0.88077	241.277	0.220164	0.71836
Stage TT4a	0.99848	271.416	0.211411	0.79898
Stage TT4b	133.067	378.359	0.067600	0.72807
Stage TT4c	214.203	851.670	0.014880	0.87958
Stage TT4d	139.268	402.564	0.060029	0.74056

Πίνακας 5.9: Εκτιμήσεις του αναλογικού μοντέλου για τον καρκίνο του Μαστού

Τα παρακάτω διαγνωστικά διαγράμματα με τον άξονα x να απεικονίζει το χρόνο επιβίωσης (σε λογαριθμική κλίμακα) και τον άξονα y να απεικονίζει το λογαριθμικό συσσωρευτικό κίνδυνο (log cumulative hazard) χρησιμοποιείται συχνά για να αξιολογήσει την καταλληλότητα ενός μοντέλου Weibull για τα δεδομένα επιβίωσης. Όπως θα παρατηρήσουμε καθώς οι κατηγορίες των μεταβλητών έρχονται παράλληλα συμπεραίνουμε ότι το μοντέλο που καταλήξαμε προσαρμόζεται καλά στα δεδομένα του καρκίνου του μαστού.



Διάγραμμα 5.6: Διαγνωστικό διάγραμμα ER Status



Διάγραμμα 5.7: Διαγνωστικό διάγραμμα PR Status

Η Ανάλυση Απόκλισης για το PH μοντέλο εξετάζει τη σημαντικότητα της προσθήκης μεταβλητών στο μοντέλο. Συγκεκριμένα, συγκρίνονται δύο μοντέλα, το κενό μοντέλο και εκείνο με τις μεταβλητές που αναφέρθηκαν και παραπάνω. Η ανάλυση επιστρέφει τα ακόλουθα αποτελέσματα:

Το log-Likelihood του κενού μοντέλου είναι -4453.8, ενώ το log-likelihood του πλήρους μοντέλου είναι -4132.9. Η διαφορά στην απόκλιση (Chisq) μεταξύ των δύο μοντέλων είναι 641.78 με 36 βαθμούς ελευθερίας. Η τιμή του p (Pr(>|Chi|)) είναι μικρότερη από 2.2e-16, υποδηλώνοντας ότι η προσθήκη των ανεξάρτητων μεταβλητών στο μοντέλο είναι εξαιρετικά σημαντική.

Συνεπώς, το πλήρες μοντέλο εξηγεί πολύ καλύτερα τα δεδομένα από το κενό μοντέλο, καθώς η προσθήκη των παραμέτρων βελτιώνει σημαντικά την προσαρμογή του μοντέλου στην παρατηρούμενη επιβίωση.

Μοντέλο	LogLik	Chisq	Df	Pr(> Chi)
Null Model	-4453.8	NA	NA	NA
PH Weibull Model	-4132.9	641.78	36	<2.2e-16

Πίνακας 5.10: Έλεγχος καταλληλότητας PH μοντέλου για τον καρκίνο του Μαστού

5.4. Μοντέλο επιβίωσης καρκίνος του Παγκρέατος

5.4.1. Το μοντέλου επιταχυνόμενου κινδύνου του καρκίνου του Παγκρέατος

Για το τελευταίο dataset της εργασίας, εκείνο του καρκίνου του παγκρέατος, θα εργαστούμε όμοια με τον καρκίνο του εγκεφάλου καθώς όπως είδαμε και στην αρχή του κεφαλαίου ταιριάζει η ίδια κατανομή, η λογαριθμοκανονική. Για το μοντέλο επιταχυνόμενου χρόνου αποτυχίας (AFT) και τις μεταβλητές που είναι σημαντικές και επηρεάζουν τους μήνες επιβίωσης καταλήγουμε μετά από δοκιμές διαφορετικών μοντέλων ότι εκείνο που ταιριάζει περισσότερο στα δεδομένα περιέχει τις μεταβλητές:

- ηλικία,
- φύλο,
- φυλή,
- βαθμός καρκίνου,
- τρόπος διάγνωσης καρκίνου,
- μέγεθος όγκου
- και η θέση του όγκου.

Θα παρουσιαστούν οι συντελεστές των μεταβλητών, το τυπικό σφάλμα καθώς και το p-value αυτών.

Variable	Value	Sig.	Std. Error
Intercept	3.71600	3.1e-07	0.72610
Age		< 2e-16	
Age (5-9 years)	6.49988	0.94470	93.70686
Age (10-14 years)	5.13007	0.93977	67.90050
Age (15-19 years)	2.02576	0.03646	0.96847
Age (20-24 years)	0.35281	0.65221	0.78283
Age (25-29 years)	-0.34146	0.64686	0.74533
Age (30-34 years)	-0.40934	0.57419	0.72849
Age (35-39 years)	-0.98258	0.17408	0.72291

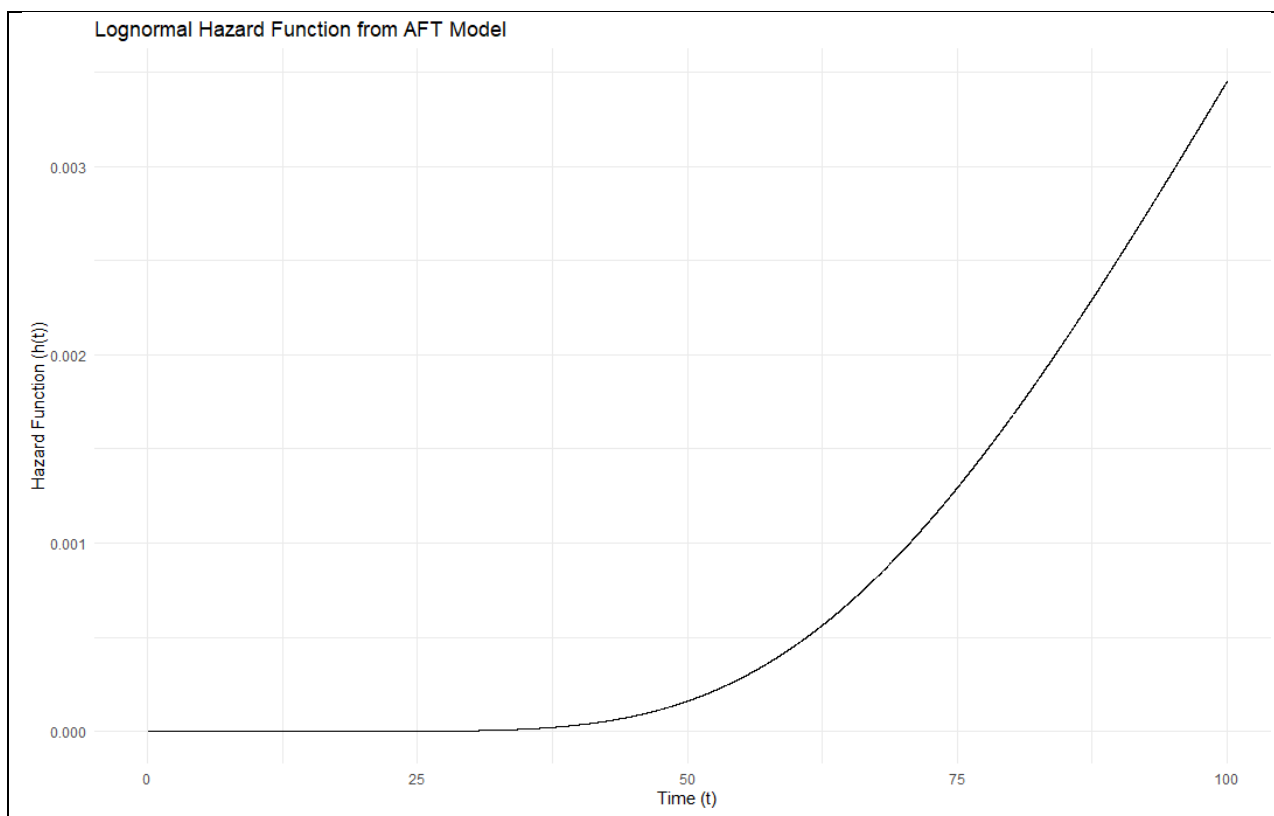
Age (40-44 years)	-1.25819	0.08071	0.72037
Age (45-49 years)	-1.51743	0.03492	0.71940
Age (50-54 years)	-1.59577	0.02645	0.71899
Age (55-59 years)	-1.68331	0.01920	0.71884
Age (60-64 years)	-1.83716	0.01059	0.71878
Age (65-69 years)	-1.87695	0.00902	0.71876
Age (70-74 years)	-2.05765	0.00420	0.71876
Age (75-79 years)	-2.22339	0.00198	0.71878
Age (80-84 years)	-2.44729	0.00066	0.71883
Age (85+ years)	-2.73827	0.00014	0.71894
Sex		9.699e-08	
Male	-0.05390	3.7e-05	0.01307
Race		6.387e-15	
Asian or Pacific	0.25584	0.00059	0.07448
Black	0.09560	0.20270	0.07505
Unknown	0.18757	0.25428	0.16453
White	0.25793	0.00037	0.07241
Grade		< 2e-16	
Grade I	0.80509	< 2e-16	0.03021
Grade III	-0.46094	< 2e-16	0.02328
Grade IV	-0.35516	4.0e-08	0.06470
Unknown	-0.51184	< 2e-16	0.01969
Diagnostic Confirmation		< 2e-16	
Άγνωστο	0.02624	0.88461	0.18077
Απευθείας Οπτικοποίηση	-0.07376	0.56599	0.12850
Θετική Αποφολιδωτική Κυτταρολογία	0.14889	0.02983	0.06854
Θετική Ιστολογία	0.39308	9.0e-09	0.06838
Θετική Εργαστηριακή Δοκιμή	-0.31759	0.00909	0.12173
Θετική Μικροσκοπική Επιβεβαίωση	0.74659	0.04040	0.36425
Ακτινογραφία	-0.05146	0.46995	0.07121
Tumor Size		< 2e-16	
Tumor Size	-0.00612	< 2e-16	0.00035
Stage A		< 2e-16	
Εντοπισμένος	1.35953	< 2e-16	0.02255
Περιφερικός	0.84591	< 2e-16	0.01484
Δεν γνωρίζουμε	0.66720	< 2e-16	0.04468

Πίνακας 5.11: Εκτιμήσεις του AFT μοντέλου για τον καρκίνο του Παγκρέατος

Συμπερασματικά, για τις ερμηνείες του AFT μοντέλου για τα δεδομένα του καρκίνου του παγκρέατος έχουμε :

- Ο συντελεστής του σταθερού όρου είναι 3.71600 με p-value 3.1e-07, ο οποίος αντιπροσωπεύει την αναμενόμενη λογαριθμική επιβίωση όταν όλες οι άλλες μεταβλητές είναι μηδενικές.
- Ο συντελεστής της ηλικιακής ομάδας 15-19 ετών δείχνει ότι η επιβίωση αυτής της ηλικιακής ομάδας είναι σημαντικά αυξημένη κατά 2.02576 μονάδες σε σχέση με την ομάδα αναφοράς, δηλαδή 0-4 ετών.
- Ο συντελεστής για το φύλο είναι -0.05390 (p-value 3.7e-05) για τους άνδρες. Αυτό δείχνει ότι οι άνδρες έχουν ελαφρώς χαμηλότερη επιβίωση από τις γυναίκες κατά 0.05390 μονάδες.
- Αναφορικά με την Φυλή εκείνοι που δηλώνουν «Άγνωστης» φυλής με συντελεστή 0.18757 φαίνεται να έχουν τόσες παραπάνω μονάδες επιβίωσης σε σχέση με τους «Λευκούς».
- Για τους ασθενείς με Βαθμό καρκίνου III κατεβαίνουμε ότι η επιβίωση τους είναι σημαντικά μειωμένη κατά 0.460694 σε σχέση με τους ασθενείς με καρκίνο Βαθμού II.
- Για εκείνους που διαγνώστηκαν με θετική ιστολογία μπορούμε να πούμε ότι η επιβίωσή τους είναι αυξημένη κατά 0.39308 σε σχέση με την ομάδα αναφοράς.
- Ο συντελεστής για το μέγεθος του όγκου είναι -0.00612 (p-value < 2e-16). Αυτό δείχνει ότι για κάθε μονάδα αύξησης στο μέγεθος του όγκου, η επιβίωση μειώνεται κατά 0.00612 μονάδες, υποδηλώνοντας αρνητική και σημαντική επίδραση.
- Για την τελευταία μεταβλητή, Stage A των ασθενών και πιο συγκεκριμένα για εκείνους που ο όγκος είναι «Εντοπισμένος» ο συντελεστής 1.359553 δηλώνει ότι η επιβίωση είναι αυξημένη κατά 1.359553 σε σχέση με την ομάδα αναφοράς.

Όπως βλέπουμε και στο επόμενο σχήμα, για τις πρώτες χρονικές στιγμές, ο κίνδυνος θανάτου είναι σχεδόν μηδενικός, ενώ αργότερα αυξάνεται εκθετικά. Αυτή η συμπεριφορά υποδηλώνει ότι ο κίνδυνος θανάτου αυξάνεται σημαντικά καθώς προχωρά ο χρόνος, κάτι που είναι χαρακτηριστικό σε πολλές περιπτώσεις καρκίνου. Η εκθετική αύξηση του κινδύνου θανάτου με την πάροδο του χρόνου είναι συμβατή με την κλινική εικόνα του καρκίνου του παγκρέατος, όπου τα πρώιμα στάδια μπορεί να είναι ασυμπτωματικά ή να παρουσιάζουν χαμηλό κίνδυνο, ενώ ο κίνδυνος αυξάνεται ραγδαία καθώς η ασθένεια εξελίσσεται.



Διάγραμμα 5.8: Διάγραμμα Συνάρτησης Κινδύνου AFT μοντέλου καρκίνου παγκρέατος

Για την αξιολόγηση της καταλληλότητας του λογαριθμικού κανονικού μοντέλου, εφαρμόστηκε ανάλυση διακύμανσης (ANOVA) για τη σύγκριση δύο ιεραρχικά διατεταγμένων μοντέλων. Το πρώτο μοντέλο αντιπροσωπεύει το απλοποιημένο μοντέλο αναφοράς, ενώ το δεύτερο μοντέλο περιλαμβάνει επιπλέον παράγοντες που μέσω ελέγχων κρίθηκαν στατιστικά σημαντικές. Τα αποτελέσματα του ANOVA παρουσιάζονται στον παρακάτω πίνακα:

Μοντέλο	Resid. Df	-2*LL	Test Df	Deviance	Pr(>Chi)
Null Model	33765	233424.4	NA	NA	NA
AFT log-normal Model	33728	218799.0	437	14625.41	0

Πίνακας 5.12: Έλεγχος καταλληλότητας AFT μοντέλου για τον καρκίνο του Παγκρέατος

Συμπερασματικά, η βελτίωση στη λογαριθμική πιθανοφάνεια στο προσαρμοσμένο μοντέλο υποδεικνύει καλύτερη προσαρμογή στα δεδομένα. Ωστόσο αυτό φαίνεται και από την τιμή p-value του ελέγχου που είναι εξαιρετικά μικρή σε σχέση με το επίπεδο σημαντικότητας που το συγκρίνουμε (0.05).

5.4.2. Το μοντέλου αναλογικού κινδύνου του καρκίνου του Παγκρέατος

Συνεχίζοντας την αναζήτηση του κατάλληλου μοντέλου για τα δεδομένα του καρκίνου του παγκρέατος τρέχουμε το μοντέλου αναλογικού κινδύνου με την κατανομή Gompertz με τις ίδιες μεταβλητές:

- ηλικία,
- φύλο,

- φυλή,
- βαθμός καρκίνου,
- τρόπος διάγνωσης καρκίνου,
- μέγεθος όγκου
- και η θέση του όγκου

και ακολουθούν τα αποτελέσματα, δηλαδή συντελεστές και ερμηνεία.

Variable	B	Exp(B)	Sig.	Std. Error
Age			< 2e-16	
Age (5-9 years)	-8.39902731	2.25e-04	8.27e+05	38.43332212
Age (10-14 years)	-6.15041321	2.13e-03	6.38e+05	13.0652532
Age (15-19 years)	-2.29201857	1.01e-01	63.108	1.233310975
Age (20-24 years)	-0.824774779	4.38e-01	325.164	0.838268880
Age (25-29 years)	0.115172476	1.12e+00	876.160	0.739049392
Age (30-34 years)	-0.012405078	9.88e-01	986.287	0.721763698
Age (35-39 years)	0.495750260	1.64e+00	4.87e+05	0.713666147
Age (40-44 years)	0.774770411	2.17e+00	2.76e+05	0.710689294
Age (45-49 years)	0.989100497	2.69e+00	1.63e+05	0.709143948
Age (50-54 years)	1.04542475	2.84e+00	1.40e+05	0.709060154
Age (55-59 years)	1.114283438	3.05e+00	1.16Ee+05	0.709032849
Age (60-64 years)	1.252054419	3.50e+00	7.74e+04	0.709021626
Age (65-69 years)	1.276608358	3.58e+00	71.778	0.709019385
Age (70-74 years)	1.412958406	4.11e+00	46.281	0.709023657
Age (75-79 years)	1.549568397	4.71e+00	28.854	0.709030166
Age (80-84 years)	1.739282399	5.69e+00	14.167	0.709040505
Age (85+ years)	2.008105271	7.45e+00	4.626	0.709074949
Sex			0.0002472	
Male	0.033767873	1.03e+00	0	0.003843336
Race			2.251e-1	
Asian or Pacific	-0.210679664	8.10e-01	1.009	0.064074787
Black	-0.097649351	9.07e-01	128.836	0.064297565
Unknown	-0.171194357	8.43e-01	2.66e+05	0.153816858
White	-0.213749036	8.08e-01	6.94e+02	0.063017950
Grade			< 2e-16	
Grade I	-0.669489263	5.12e-01	0	0.006296198
Grade III	0.357842305	1.43e+00	0	0.005274471
Grade IV	0.238600657	1.27e+00	3.57e+01	0.057719320
Unknown	0.397900671	1.49e+00	0	0.006006619
Diagnostic Confirmation			< 2e-16	
Άγνωστο	0.173603485	1.19e+00	2.65e+05	0.155848237

Απευθείας Οπτικοποίηση	0.230186565	1.26e+00	3.94e+04	0.111762105
Θετική Αποφολιδωτική Κυτταρολογία	0.033467890	1.03e+00	5.67e+05	0.058420226
Θετική Ιστολογία	-0.182543851	8.33e-01	1.85e+03	0.058639174
Θετική Εργαστηριακή Δοκιμή	0.364653657	1.44e+00	5.04e+02	0.104830431
Θετική Μικροσκοπική Επιβεβαίωση	-0.372173073	6.89e-01	2.47e+05	0.321609987
Ακτινογραφία	0.130246211	1.14e+00	2.81e+04	0.059315391
Tumor Size			< 2e-16	
Tumor Size	0.003119443	1.00e+00	4.17e+05	0.003843336
Stage A			< 2e-16	
Εντοπισμένος	-1.13423362	3.22e-01	0	0.004717426
Περιφερικός	-0.640416780	5.27e-01	0	0.003843336
Δεν γνωρίζουμε	-0.634472491	5.30e-01	1.93e-53	0.039022665

Πίνακας 5.13: Εκτιμήσεις του αναλογικού μοντέλου για τον καρκίνο του Παγκρέατος

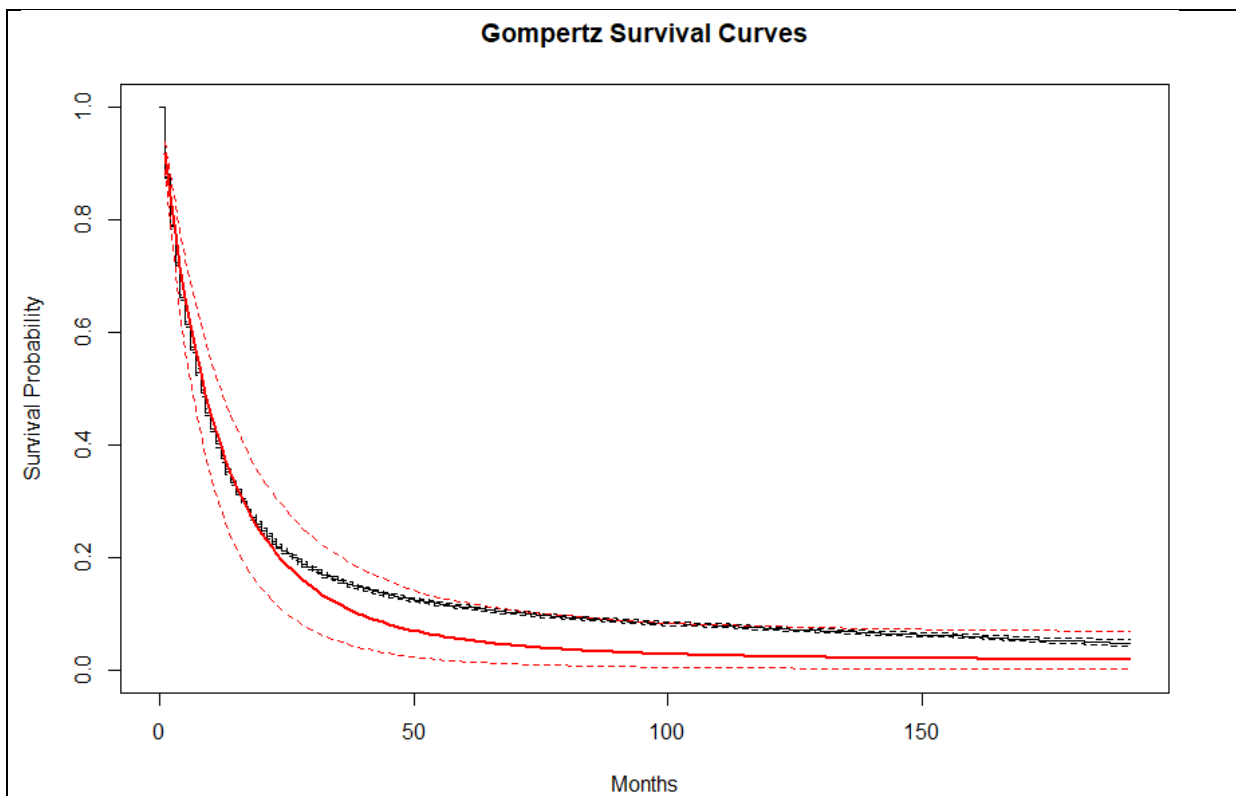
Όσον αφορά τις ερμηνείες των μεταβλητών για το αναλογικό μοντέλο του Gompertz έχουμε:

- Για την ηλικία 5-9 ετών, το αρνητικό «β» σημαίνει ότι τα παιδιά ηλικίας 5-9 ετών έχουν μικρότερο κίνδυνο σε σχέση με τα παιδιά κάτω των 5 ετών. δηλαδή την ομάδα αναφοράς και πιο συγκεκριμένα έχουν $\exp(-8.399) = 0.0000225$ φορές μεγαλύτερο κίνδυνο.
- Για το φύλο έχουμε ότι οι άνδρες έχουν αυξημένο κίνδυνο σε σχέση με τις γυναίκες και μάλιστα $\exp(0.036) = 1.03$ φορές περισσότερες.
- Για την φυλή παρατηρούμε ότι εκείνοι με την καταγραφή «Λευκοί» έχουν $\exp(-214) = 0.808$ ή 19.2% φορές μικρότερο κίνδυνο από τους Αμερικάνους.
- Για τον βαθμό του καρκίνου διαπιστώθηκε ότι εκείνοι με Βαθμό I έχουν $\exp(-669) = 0.512$ ή 49% λιγότερο κίνδυνο σε σχέση με εκείνους που έχουν διαγνωστεί με καρκίνο Βαθμού II.
- Επίσης, εκείνοι που έχουν διαγνωστεί με ακτινογραφία έχουν 14% φορές μεγαλύτερο κίνδυνο σε σύγκριση με εκείνους που επιβεβαίωσαν με κλινική διάγνωση.
- Όπως περιμέναμε και για το μέγεθος του όγκου όσο μεγαλύτερος είναι. τόσο μεγαλύτερος είναι ο κίνδυνος.
- Τέλος. για το στάδιο A των ασθενών εκείνοι που έχουν εντοπίσει το σημείο του όγκου έχουν μειωμένο κίνδυνο κατά $\exp(-1.134) = 0.322$ φορές σε σχέση με τους ασθενείς με που είναι απομακρυσμένος.

Ο έλεγχος πιθανολογικού λόγου (Likelihood Ratio Test - LRT) που εκτελέστηκε αποδεικνύει ότι το μοντέλο αναλογικού κινδύνου Gompertz, το οποίο περιλαμβάνει τους

παράγοντες ηλικία, φύλο, φυλή, βαθμό όγκου, επιβεβαίωση διάγνωσης, μέγεθος όγκου και ιστορικό στάδιο SEER A, προσφέρει σημαντικά καλύτερη προσαρμογή στα δεδομένα σε σχέση με το κενό μοντέλο. Η τιμή p , η οποία είναι σχεδόν μηδενική, υποδηλώνει τη βελτίωση στην καταλληλότητα του μοντέλου με την προσθήκη αυτών των παραγόντων είναι στατιστικά σημαντική.

Στο επόμενο σχήμα βλέπουμε το διάγραμμα επιβίωσης που προκύπτει από το μοντέλο Gompertz που έχει εφαρμοστεί στα δεδομένα. Η μαύρη καμπύλη αντιπροσωπεύει την προβλεπόμενη καμπύλη επιβίωσης που προκύπτει από το μοντέλο Gompertz καθώς οι γραμμές μαύρες διακεκομμένες αντιπροσωπεύουν τα 95% διαστήματα εμπιστοσύνης της προβλεπόμενης καμπύλης επιβίωσης. Η κόκκινη γραμμή αναπαριστά τα παρατηρούμενα δεδομένα επιβίωσης από το σύνολο των δεδομένων, ενώ οι κόκκινες διακεκομμένες γραμμές αντιπροσωπεύουν τα 95% διαστήματα εμπιστοσύνης των παρατηρούμενων δεδομένων επιβίωσης. Στην αρχή, υπάρχει μια απότομη πτώση στην πιθανότητα επιβίωσης, κάτι που είναι κοινό σε πολλές μελέτες επιβίωσης, ειδικά όταν η αρχική χρονική περίοδος περιλαμβάνει πιο σοβαρές περιπτώσεις ή υψηλότερη θνησιμότητα. Καθώς περνάει ο χρόνος, η καμπύλη αρχίζει να ισορροπεί, υποδηλώνοντας ότι η πιθανότητα επιβίωσης γίνεται πιο σταθερή. Αυτό μπορεί να σημαίνει ότι όσοι επιβιώνουν την αρχική περίοδο έχουν καλύτερες πιθανότητες μακροπρόθεσμης επιβίωσης. Καταλήγουμε ότι το γράφημα μας παρέχει μια σαφή σύγκριση μεταξύ των προβλεπόμενων τιμών επιβίωσης και των παρατηρούμενων δεδομένων, επιτρέποντας την εκτίμηση της ακρίβειας και της αξιοπιστίας των προβλέψεων του μοντέλου.



Διάγραμμα 5.9: Διάγραμμα Συνάρτησης Επιβίωσης PH Gompertz μοντέλου καρκίνου παγκρέατος

ΚΕΦΑΛΑΙΟ 6°

6. Συμπεράσματα

Στην παρούσα εργασία γίνεται αναφορά σε τρία είδη καρκίνων και στους παράγοντες που επηρεάζουν τη θνησιμότητά τους. Πιο συγκεκριμένα, τα είδη του καρκίνου που αναλύονται αφορούν δεδομένα του εγκεφάλου, του μαστού και του παγκρέατος. Αναφορικά με τον καρκίνο του εγκεφάλου το δείγμα αποτελείται από 59.812 ασθενείς, εκ των οποίων 43,8% είναι γυναίκες και το 56,2% άνδρες, με συνολικό ποσοστό θνησιμότητας 74,87% και μέσο όρο επιβίωσης τους 43 μήνες. Αντίστοιχα, στον καρκίνο μαστού το δείγμα μας περιλαμβάνει 5.659 γυναίκες ασθενείς με ποσοστό θνησιμότητας 10,1% και μέσο όρο επιβίωσης τους 33 μήνες. Τέλος, στον καρκίνο παγκρέατος τα δεδομένα περιέχουν 116.582 ασθενείς, εκ των οποίων 50,2% γυναίκες και 49,8% άνδρες, με συνολικό ποσοστό θνησιμότητας 90,5% και μέσο όρο επιβίωσης τους 17 μήνες.

Σε ότι αφορά τις συσχετίσεις της παρούσας εργασίας, κεφάλαιο 4°, παρατηρείται πως στο σύνολο δεδομένων του καρκίνου του εγκεφάλου οι μεταβλητές Tumor Size, Number of tumors, Vital Status, Race, Age, Grade, Diagnostic Confirmation και Laterality έχουν σημαντική συσχέτιση με την μεταβλητή απόκρισης Survival Months, ενώ η μεταβλητή Sex φάνηκε να είναι ανεξάρτητη από αυτή την μεταβλητή. Ακόμα, στην συνέχεια της ανάλυσης υπολογίστηκαν οι συσχετίσεις όλων των παραπάνω μεταβλητών σε σχέση με το Vital Status όπου βρέθηκαν παρόμοια αποτελέσματα. Στην συνέχεια του κεφαλαίου παρουσιάστηκαν αντίστοιχες αναλύσεις για τα άλλα δυο σύνολα δεδομένων.

Από το τελευταίο κεφάλαιο συμπεραίνουμε ότι το λογαριθμοκανονικό μοντέλο επιταχυνόμενου χρόνου αποτυχίας για τον καρκίνο του εγκεφάλου δείχνει ότι οι μεταβλητές ηλικία, φύλο, φυλή, βαθμός καρκίνου, μέγεθος όγκου, πλευρικότητα και μέθοδος διαγνωστικής επιβεβαίωσης επηρεάζουν σημαντικά το χρόνο επιβίωσης των ασθενών. Παράλληλα, παρατηρήθηκαν τα ίδια αποτελέσματα και στο μοντέλο αναλογικού κινδύνου Gompertz για το dataset με τα δεδομένα του καρκίνου του εγκεφάλου. Για τον καρκίνο του μαστού και το μοντέλο επιταχυνόμενου κινδύνου Weibull, καταλήγουμε ότι η ηλικία, ο βαθμός του καρκίνου, η οικογενειακή κατάσταση, ο υποδοχέας οιστρογόνου, ο υποδοχέας προγεστερόνης, ο υποτύπος μαστού και το Sage N είναι οι μεταβλητές που επηρεάζουν σημαντικά τους μήνες επιβίωσης της γυναίκας που νοσεί. Στην συνέχεια, για το μοντέλο αναλογικού κινδύνου Weibull πήραμε τα ίδια αποτελέσματα με την διαφορά ότι βρέθηκε σημαντική και η μεταβλητή που ορίζει τον αριθμό των όγκων έχουν αναπτυχθεί σε κοντινούς ιστούς (Stage T). Για το σύνολο δεδομένων του καρκίνου του παγκρέατος, τόσο το μοντέλο Log-normal επιταχυνόμενου κινδύνου όσο εκείνο του αναλογικού κινδύνου Gompertz έδειξαν τις μεταβλητές ηλικία, φύλο, φυλή, βαθμός καρκίνου, τρόπος διάγνωσης, μέγεθος όγκου και η θέση του όγκου να επιδρούν σημαντικά στο χρόνο επιβίωσης των ασθενών. Τέλος, παρατηρήσαμε ότι και στα τρία μοντέλα μη αναλογικού κινδύνου, δηλαδή μοντέλα Log-normal και Weibull, τα p-value των ελέγχων προσαρμογής είχαν αρκετά μικρές τιμές ώστε να καταλήξουμε στην καλή προσαρμογή τους στα δεδομένα μας. Για τα μοντέλα αναλογικού κινδύνου για τα διαφορετικά σύνολα δεδομένων με βάση τον έλεγχο πιθανοφάνειας καταλήξαμε στα ίδια συμπεράσματα.

Σύμφωνα με την διεθνή βιβλιογραφία, ο καρκίνος του εγκεφάλου εμφανίζεται πιο συχνά στους άνδρες από ότι στις γυναίκες, όπως παρατηρήθηκε και στην ανάλυση. Ωστόσο, ενώ πιο συχνά αποβιώνουν οι άντρες, στην ανάλυση φάνηκε ότι οι γυναίκες έχουν μεγαλύτερο κίνδυνο θνησιμότητας. Επίσης αναφέρεται ότι ο μέσος όρος επιβίωσης είναι περίπου 30 μήνες. Σύμφωνα και με την βιβλιογραφία, ο καρκίνος του μαστού εμφανίζεται ως επι το πλείστον σε γυναίκες. Ο μέσος όρος επιβίωσης είναι τα 5 έτη, ενώ τα ποσοστά θνησιμότητάς τους είναι πολύ χαμηλά. Για τον καρκίνο του παγκρέατος αναφέρεται ότι πιο συχνά εμφανίζεται στους άντρες, οι οποίοι παρουσιάζουν υψηλότερο ποσοστό θνησιμότητας, κάτι που έρχεται σε αντίθεση με τα αποτελέσματα της εργασίας. Ο μέσος όρος επιβίωσης αντιστοιχεί σε 8 μήνες, αλλά η πλειοψηφία δύναται να επιβιώσει και παραπάνω.

Επιπροσθέτως, η βιβλιογραφία αναφέρει ότι σε όλα τα είδη καρκίνων ο χρόνος επιβίωσης και η θνησιμότητα συσχετίζονται με την ηλικία, το φύλο, το βαθμό καρκίνου, το μέγεθος καρκίνου, τη μέθοδο διαγνωστικής επιβεβαίωσης, την πλευρικότητα και εξάπλωση του καρκίνου. Στα ίδια αποτελέσματα καταλήγουν τόσο η ανάλυση του αναλογικού μοντέλου όσο και εκείνη του επιταχυνόμενου χρόνου αποτυχίας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Αντζουλάκος, Δ., (2021), Σημειώσεις μαθήματος «Ανάλυση δεδομένων με τη χρήση στατιστικών πακέτων: Εισαγωγή στην R», Τμήμα Στατιστικής & Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς, ΠΜΣ «Εφαρμοσμένη Στατιστική».

Αντζουλάκος Δ., (2023), Σημειώσεις για το μάθημα «Ανάλυση επιβίωσης», Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πρόγραμμα Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Κούτρας Μ., Ευαγγελάρας Χ., (2018), Ανάλυση Παλινδρόμησης Θεωρία και Εφαρμογές, Εκδόσεις Τσότρας

Οικονόμου Β., (2023), Διπλωματική εργασία: Μοντέλα θνησιμότητας διάφορων ειδών καρκίνου.

Οικονόμου Π., Καρώνη Χ., (2010), Στατιστικά Μοντέλα Παλινδρόμησης, Εκδόσεις Συμεών Αθήνα.

Σαχλάς Α., Μπερσίμης Σ., (2016), Εφαρμοσμένη Στατιστική με Χρήση του IBM SPSS Statistics 23: Με Έμφαση στις Επιστήμες Υγείας, Εκδόσεις Τζιόλα.

Σαχλάς Α., Μπερσίμης Σ., (2016), Εφαρμοσμένη Στατιστική με Έμφαση στις Επιστήμες Υγείας, Εκδόσεις Τζιόλα.

Τζαβελάς Γ., (2012), Σημειώσεις μαθήματος «Κλινικές Δοκιμές», Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς ΠΜΣ στην Εφαρμοσμένη Στατιστική.

Ξένα

C.SIMON HERRINGTON, (2019), Παθολογική Ανατομική Muir's, Εκδόσεις Broken Hill

McCullagh P., Nelder J.A., (1989), Generalized Linear Models, Εκδόσεις Chapman and Hall/CRC

Rodrigues G., (2010), Parametric Survival Models

Διαδίκτυο

<https://www.neurocenter.gr/ogkoi-egkefalou-genika.html>

<https://tinyurl.com/543dyt88>

<https://www.oagiosdimitrios.gr/iatriki-ypiresia/cheiroyrgiko-tmima/pagkreas/karkinos-pagkreatos/>

<https://www.cancer.org/cancer/types/pancreatic-cancer.html>

<https://www.cancer.org/cancer/types/breast-cancer.html>

<https://www.breastcancer.org/types/molecular-subtypes>

<https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-grade>

<https://www.cancerresearchuk.org/about-cancer/breast-cancer/survival>

<https://www.cancerresearchuk.org/about-cancer/pancreatic-cancer/survival>

<https://www.cancerresearchuk.org/about-cancer/brain-tumours/survival>

<https://www.iarc.who.int/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570158/>

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3997531/?fbclid=IwAR3_Kg2JVTTesStOyDu6EKaAqjcuQc4rfUmSLG_UBMVQnkgmNCVRJtqlxMk

<https://www.cancer.gov/news-events/cancer-currents-blog/2019/glioblastoma-treatment-response-differs-by-sex>

<https://www.healthline.com/health/history-of-breast-cancer>

<https://www.tandfonline.com/doi/full/10.2217/fon-2016-0010>

<https://pubmed.ncbi.nlm.nih.gov/25340432/>

<https://www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-know-your-brain>

<https://www.sciencedirect.com/science/article/pii/S2311300617300125>

<https://www.cancer.org/cancer/understanding-cancer/history-of-cancer.html>

