



## Εργασία Ομάδας 4

SPS...S

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΔΔΠΜΣ στην Ιατρική Φυσική – Ακτινοφυσική

Ομάδα 4:

Βαρόκα Μαρία

Γιαννακούλη Παναγιώτα

Γκεζέρης Γιώργος

Τσελεπής Δημήτριος

15/12/2019

## Περίληψη

Η παρούσα εργασία αφορά τη στατιστική ανάλυση και επεξεργασία δεδομένων με τη χρήση του στατιστικού υπολογιστικού πακέτου SPSS.

Οι μεταβλητές που χρησιμοποιήθηκαν για τη στατιστική ανάλυση είναι:

- Πρώτο καρδιακό επεισόδιο (nominal)
- Ηλικία (scale)
- ΔΑΠ-Διαστολική Αρτηριακή Πίεση (scale)
- Χοληστερόλη (scale)
- Αριθμός τσιγάρων (scale)
- Επιβίωση μετά από 10 χρόνια (nominal)
- Οικογενειακό Ιστορικό ΚΕ (nominal)

Για τις κατηγορικές μεταβλητές (nominal) ως μέτρο θέσης δόθηκε η επικρατούσα τιμή (Mode) , ως σχετικά διαγράμματα τα Pie Chart και Histogram καθώς και οι πίνακες συχνοτήτων.

Για τις ποσοτικές μεταβλητές (scale) ως μέτρα θέσης δόθηκαν η επικρατούσα τιμή (Mode), η διάμεσος (Median) και η μέση τιμή (Mean) , ως μέτρα διασποράς δόθηκαν η τυπική απόκλιση (St.dev) και η διακύμανση (Variance) , ως μέτρα μορφής δόθηκαν οι συντελεστές κύρτωσης (Coefficient of Kurtosis) και ασυμμετρίας (Coefficient of Skewness), ως σχετικά διαγράμματα τα Histogram και θηκόγραμμα (Box plot).

# 1 Περιγραφικές διαδικασίες του SPSS (Α)

Κοιτώντας την τελευταία γραμμή του πίνακα Statistics μπορούμε να δούμε την επικρατούσα τιμή (Mode) για καθεμία από τις υπό μελέτη μεταβλητές (Πρώτο καρδιακό, Επιβίωση μετά από 10 χρόνια, Οικογενειακό ιστορικό ΚΕ). Παρατηρούμε πως για το πρώτο καρδιακό επεισόδιο η επικρατούσα τιμή ισούται με 1, ενώ για τις άλλες δύο κατηγορικές μας μεταβλητές η επικρατούσα τιμή είναι η 0.

Οι υπόλοιποι πίνακες με τίτλους Πρώτο καρδιακό επεισόδιο , Επιβίωση μετά από 10 χρόνια, Οικογενειακό ιστορικό ΚΕ είναι οι πίνακες συχνότητας των αντίστοιχων μεταβλητών, από τους οποίους στη στήλη Frequency βλέπουμε τις συχνότητες εμφάνισης της κάθε κατηγορίας κάθε μεταβλητής. Στη στήλη Valid Percent βλέπουμε τα έγκυρα ποσοστά τα οποία για να υπολογιστούν δεν λαμβάνονται υπόψη οι τιμές που δεν έχουν απαντηθεί (missing values). Σύμφωνα με τον πίνακα Statistics φαίνεται πως δεν υπάρχουν missing values. Η στήλη Cumulative Percent (αθροιστικό ποσοστό) εδώ δεν έχει νόημα, επειδή οι μεταβλητές είναι κατηγορικές.

**Statistics**

		πρώτο καρδιακό επεισόδιο	Επιβίωση μετά από 10 χρόνια	Οικογενειακό Ιστορικό ΚΕ
N	Valid	474	474	474
	Missing	0	0	0
Mode		1	0	0

**πρώτο καρδιακό επεισόδιο**

		<u>Frequency</u>	<u>Percent</u>	<u>Valid Percent</u>	<u>Cumulative Percent</u>
<u>Valid</u>	1 <u>Κανένα ΚΕ</u>	240	50,6	50,6	50,6
	2 <u>Ακίνδυνο</u>	144	30,4	30,4	81,0
	3 <u>Επικίνδυνο</u>	18	3,8	3,8	84,8
	4 <u>Ξαφνικός Θάνατος</u>	72	15,2	15,2	100,0
	<u>Total</u>	474	100,0	100,0	

**Επιβίωση μετά από 10 χρόνια**

		<u>Frequency</u>	<u>Percent</u>	<u>Valid Percent</u>	<u>Cumulative Percent</u>
<u>Valid</u>	0 Ζωντανός	358	75,5	75,5	75,5
	1 Πεθαμένος	116	24,5	24,5	100,0
	<u>Total</u>	474	100,0	100,0	

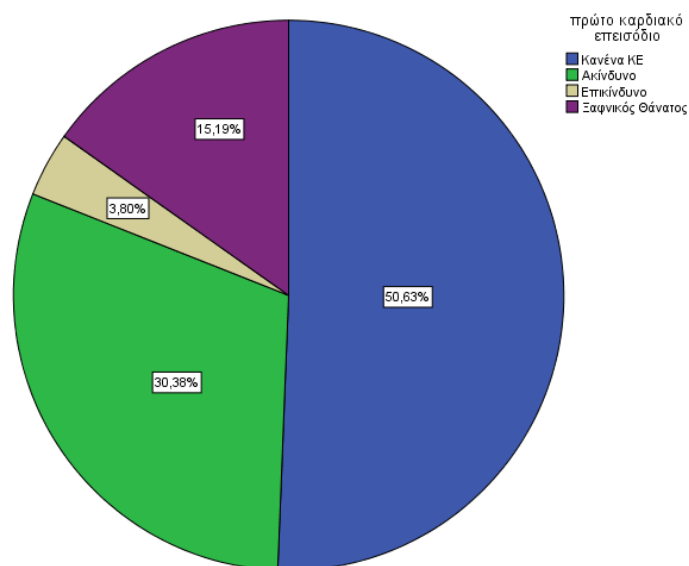
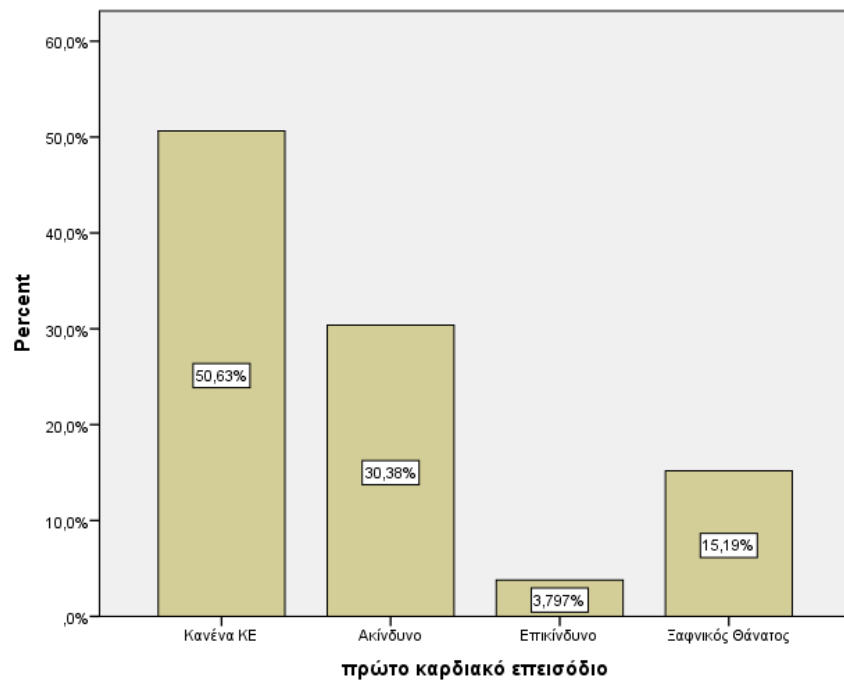
**Οικογενιακό Ιστορικό ΚΕ**

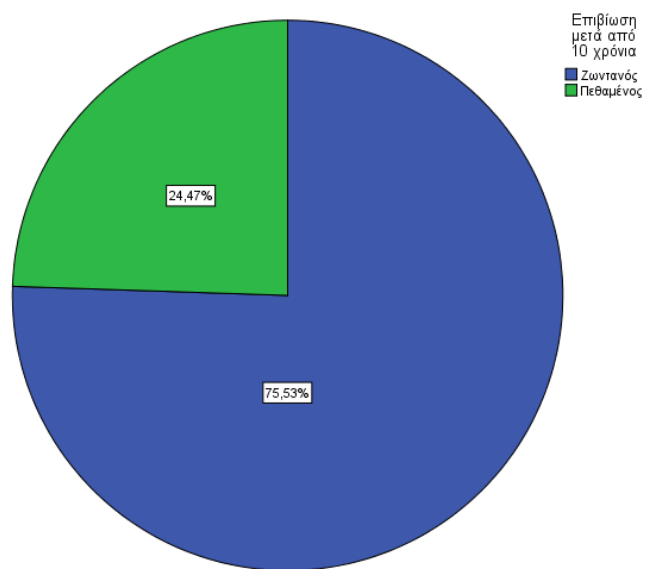
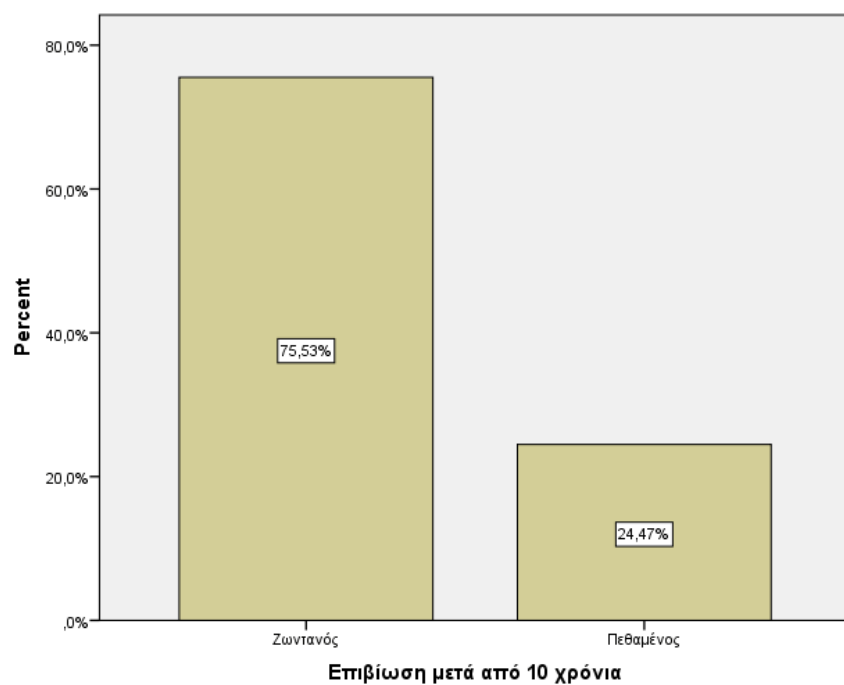
		<u>Frequency</u>	<u>Percent</u>	<u>Valid Percent</u>	<u>Cumulative Percent</u>
<u>Valid</u>	0 Όχι	354	74,7	74,7	74,7
	1 Ναι	120	25,3	25,3	100,0
	<u>Total</u>	474	100,0	100,0	

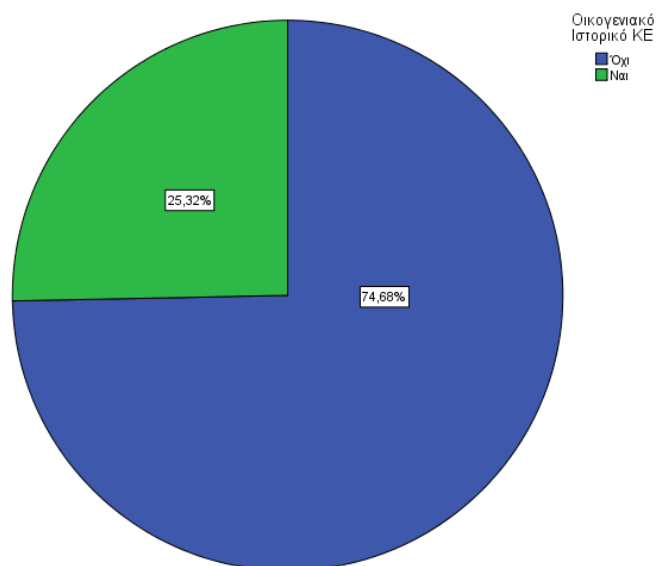
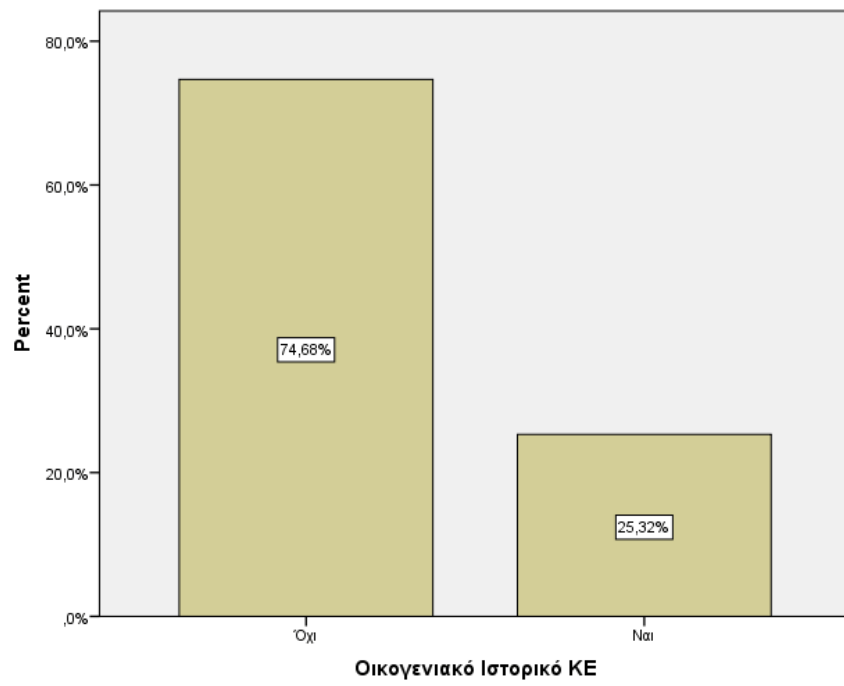
Από τους παραπάνω πίνακες παρατηρούμε τα εξής:

- Για το πρώτο καρδιακό επεισόδιο η μεταβλητή με τη μεγαλύτερη συχνότητα εμφάνισης είναι η κατηγορία με τον κωδικό 1 = Κανένα ΚΕ και ποσοστό 50,6 %.
- Για την Επιβίωση μετά από 10 χρόνια η μεταβλητή με τη μεγαλύτερη συχνότητα εμφάνισης είναι η κατηγορία με τον κωδικό 0 = Ζωντανός και ποσοστό 75,5 %.
- Για το Οικογενειακό ιστορικό ΚΕ η μεταβλητή με τη μεγαλύτερη συχνότητα εμφάνισης είναι η κατηγορία με τον κωδικό 0 = Όχι και ποσοστό 74,7 %.

Τα παραπάνω συμπεράσματα για τις κατηγορικές μεταβλητές επαληθεύονται και από τα ακόλουθα ραβδογράμματα και Pie charts.







Στον ακόλουθο πίνακα παρουσιάζονται όλα τα μέτρα θέσης , διασποράς και μορφής για τις scale μεταβλητές της στατιστικής μας ανάλυσης.

<b>Statistics</b>		ηλικία πρώτου επεισοδίου	Μέση διαστολική αρτηριακή πίεση	χοληστερόλη mg ανά dl	Αριθμός τσιγάρων ανά ημέρα	Σωματικό Βάρος
<b>N</b>	<u>Valid</u>	474	472	474	472	474
	<u>Missing</u>	0	2	0	2	0
<u>Mean</u>		47,80	88,53	264,31	11,69	173,15
<u>Median</u>		48,00	87,00	261,00	10,00	171,00
<u>Mode</u>		54	80	235 <sup>a</sup>	0	168
<u>Std. Deviation</u>		4,097	12,232	52,576	12,276	23,820
<u>Variance</u>		16,788	149,625	2764,231	150,704	567,399
<u>Skewness</u>		-,142	,764	,651	,850	,447
<u>Std. Error of Skewness</u>		,112	,112	,112	,112	,112
<u>Kurtosis</u>		-1,117	,675	1,804	,687	,064
<u>Std. Error of Kurtosis</u>		,224	,224	,224	,224	,224
<u>Range</u>		14	68	409	60	131
<u>Minimum</u>		40	65	106	0	123
<u>Maximum</u>		54	133	515	60	254

a. Multiple modes exist. The smallest value is shown

Για τη μεταβλητή ηλικία πρώτου επεισοδίου παρατηρούμε πως ο συντελεστής skewness είναι αρνητικός (-0,142) , για τη μέση διαστολική αρτηριακή πίεση είναι θετικός (+0,764) όπως και για τη χοληστερόλη (+0,651) , τον αριθμό τσιγάρων ανά ημέρα (+0,850) και το σωματικό βάρος (+0,447) . Όταν ο συγκεκριμένος συντελεστής έχει αρνητική τιμή η κορυφή της κατανομής μετατοπίζεται προς τα δεξιά ενώ όταν έχει θετική τιμή μετατοπίζεται προς τα αριστερά. Επιπλέον η συσώρευση των παρατηρήσεων προς τα αριστερά ή δεξιά της κατανομής εξετάζεται και μέσω της σύγκρισης των μέτρων της μέσης (mean) και της ενδιάμεσης τιμής (median). Εάν η μέση τιμή είναι μικρότερη της ενδιάμεσης τότε έχουμε αρνητική συμμετρία, ενώ εάν είναι μεγαλύτερη έχουμε θετική συμμετρία, πράγμα το οποίο αποδεικνύεται και από τον πίνακα.

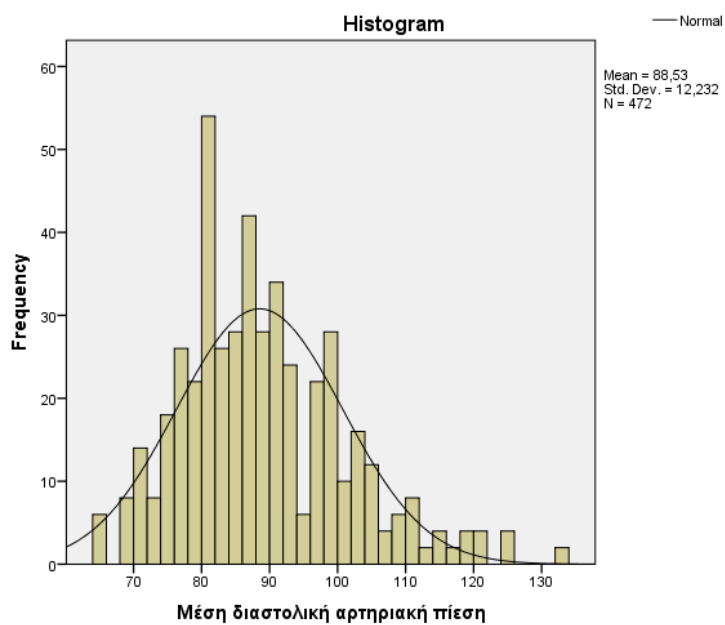
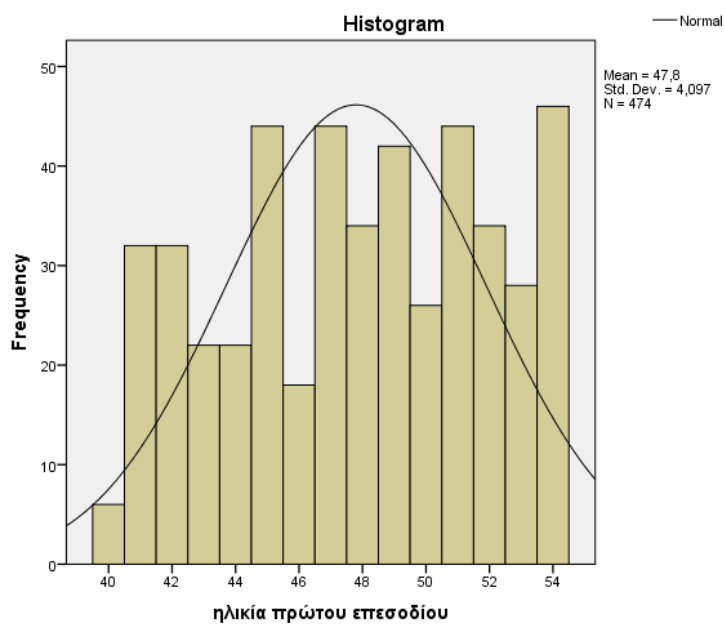
Για τη μεταβλητή ηλικία πρώτου επεισοδίου παρατηρούμε πως ο συντελεστής kurtosis είναι αρνητικός (-1,117) , για τη μέση διαστολική αρτηριακή πίεση είναι θετικός (+0,675) όπως και για τη χοληστερόλη (+1,804) , τον αριθμό τσιγάρων ανά ημέρα (+0,687) και το σωματικό βάρος (+0,064) . Όταν ο συγκεκριμένος συντελεστής έχει αρνητική τιμή η κατανομή μπορεί να θεωρηθεί πλατύκυρτη, ενώ όταν έχει θετική τιμή μπορεί να θεωρηθεί λεπτόκυρτη.

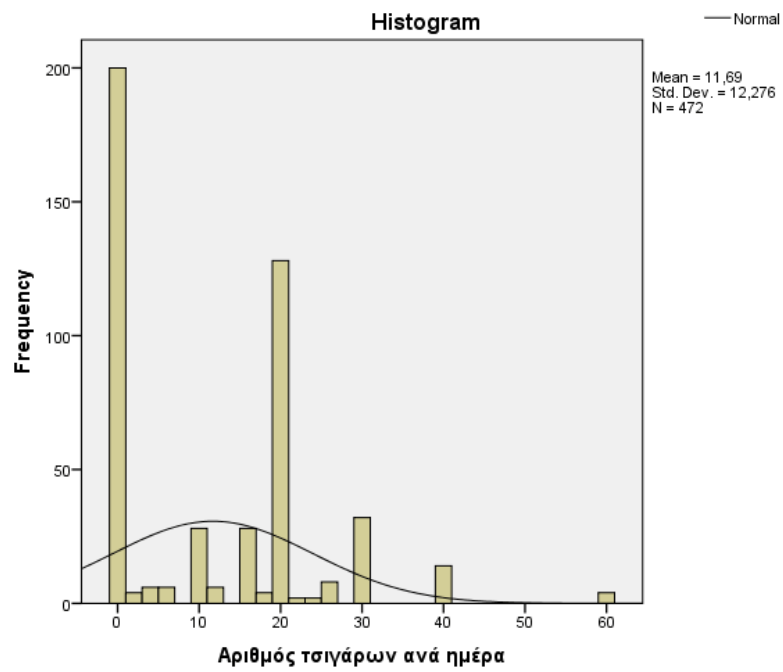
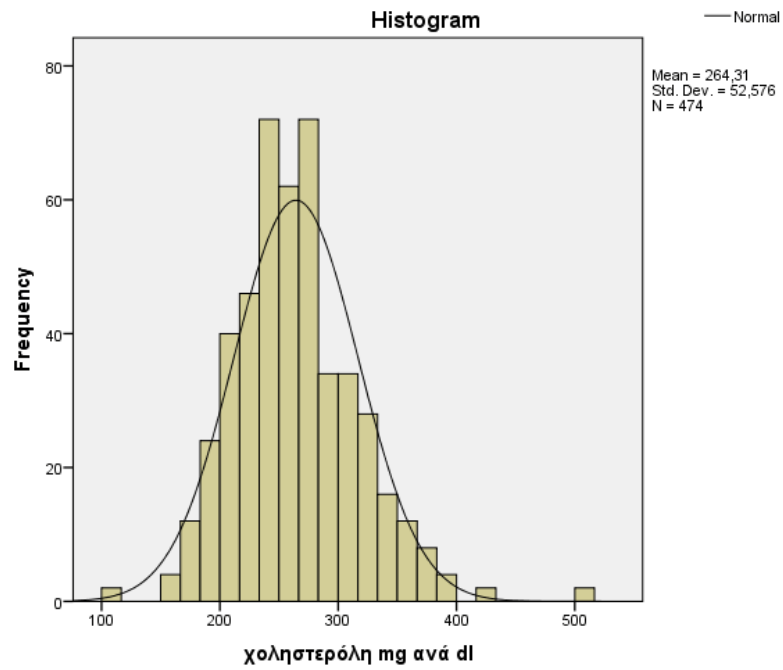
Σημειώνεται πως οι τιμές των συντελεστών ασυμμετρίας και κύρτωσης πρέπει να βρίσκονται μέσα στο διάστημα -2 έως 2, κάτι το οποίο τηρείται και από τα δεδομένα του πίνακα.

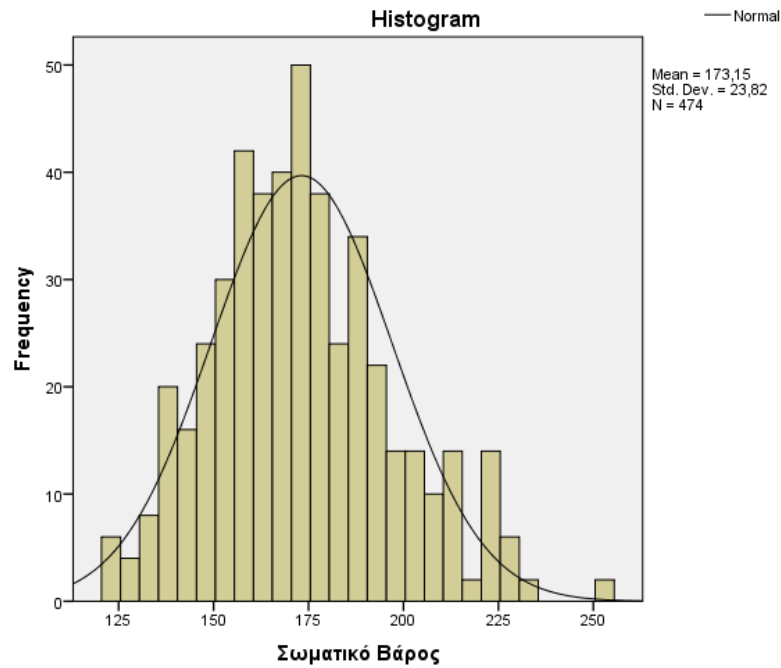
Τα παραπάνω συμπεράσματα για το συντελεστή skewness και kurtosis για τις ποσοτικές



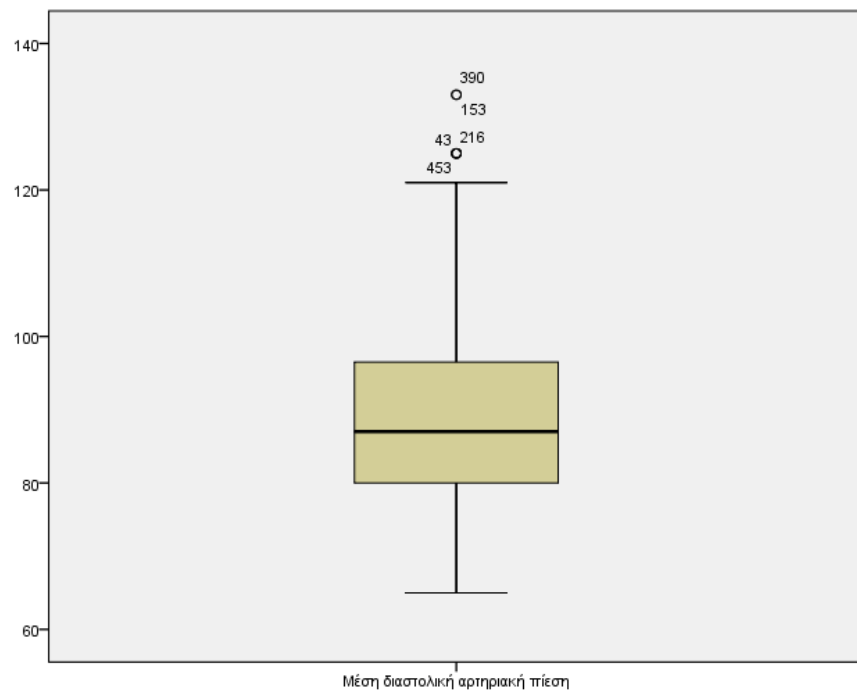
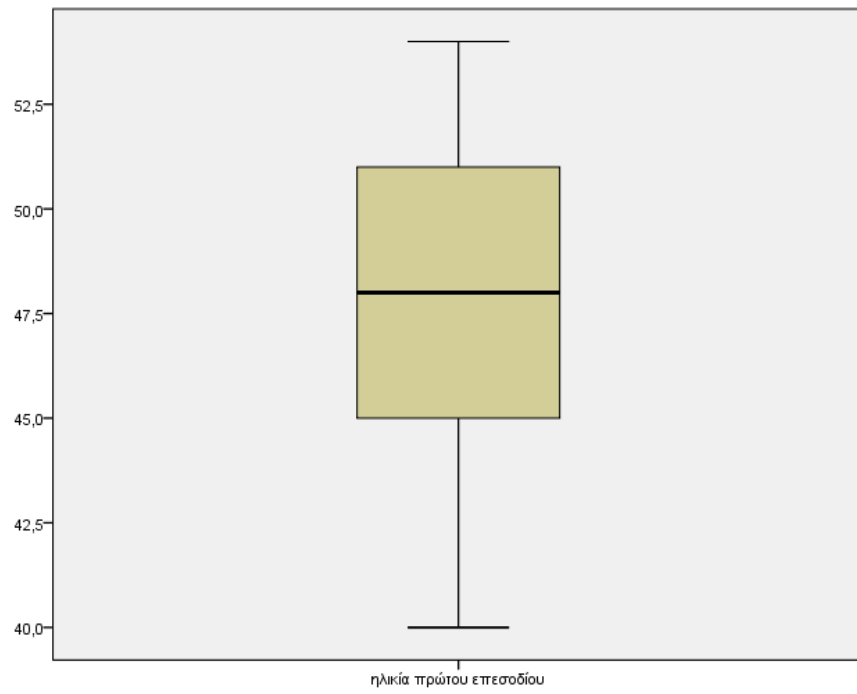
μεταβλητές επαληθεύονται και από τα ακόλουθα ιστογράμματα.

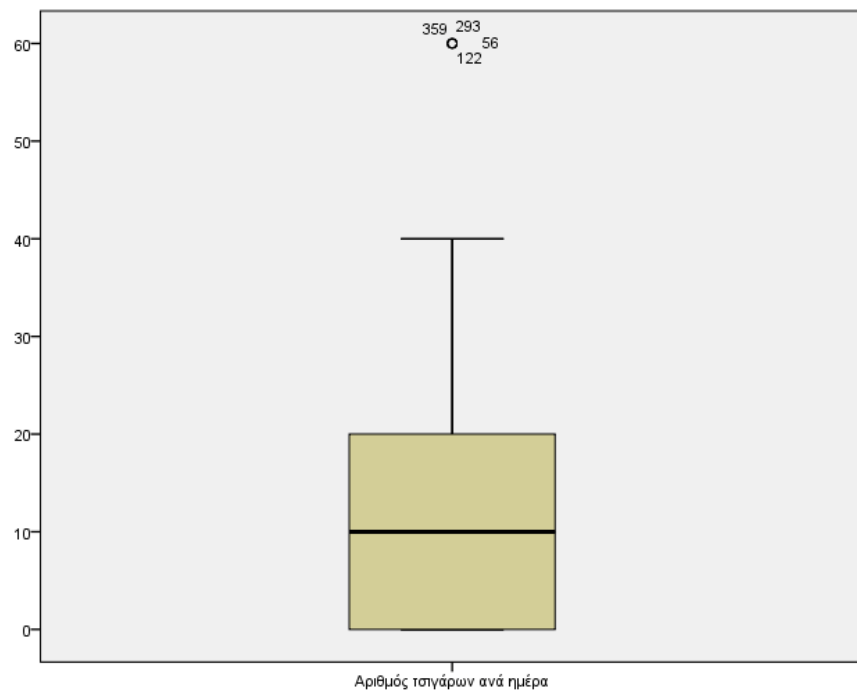
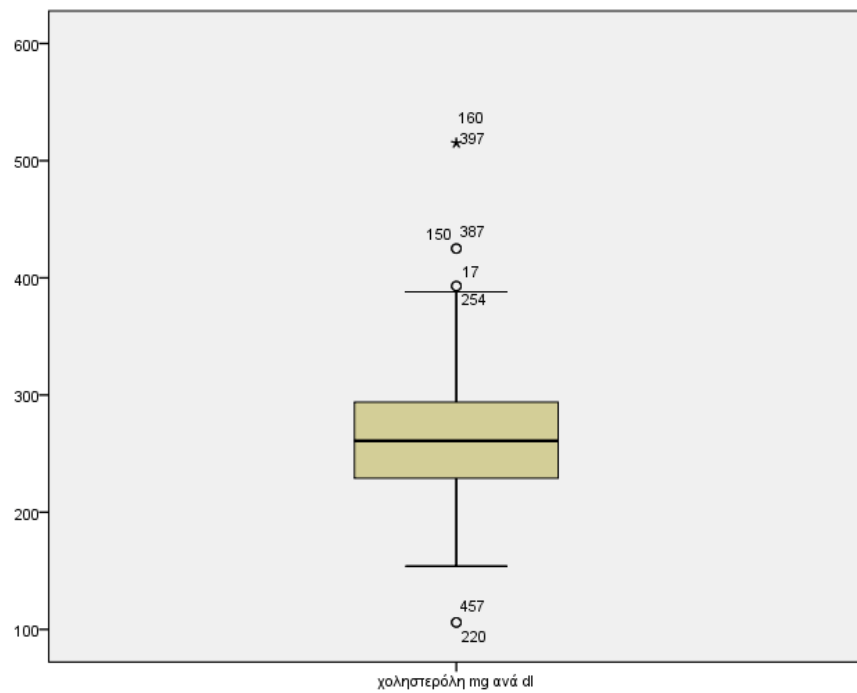


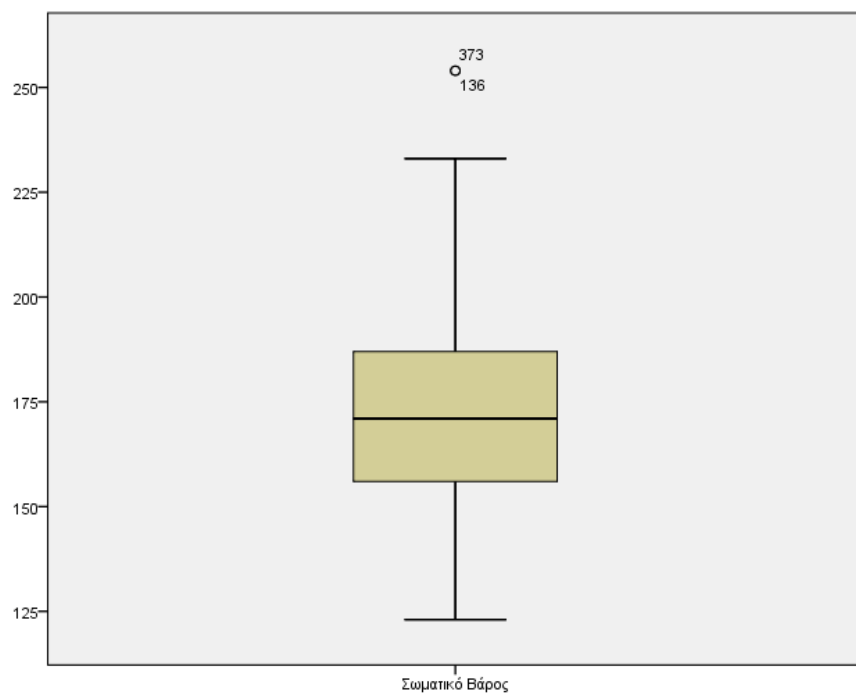




Στη συνέχεια παρατίθενται τα θηκογράμματα των ποσοτικών μεταβλητών. Όπως παρατηρείται δεν υπάρχουν εξαιρετικά ακραίες τιμές (outliers – δηλαδή τιμές που ξεπερνάνε κατά 1,5 ή κατά 3 φορές το ενδοτεταρτημοριακό εύρος) μόνο για τη μεταβλητή “Ηλικία πρώτου επεισοδίου”.







Στον ακόλουθο πίνακα παρουσιάζονται τα προεπιλεγμένα περιγραφικά στατιστικά στοιχεία για τη “Μέση διαστολική αρτηριακή πίεση” ανά κατηγορίες σοβαρότητας πρώτου καρδιακού επεισοδίου και οικογενειακού ιστορικού ΚΕ.

Report				
Μέση διαστολική αρτηριακή πίεση				
πρώτο καρδιακό επεισόδιο	Οικογενιακό Ιστορικό ΚΕ	Mean	N	Std. Deviation
Κανένα ΚΕ	Όχι	86,81	192	10,368
	Ναι	91,17	46	12,609
	Total	87,66	238	10,946
Ακίνδυνο	Όχι	87,54	92	12,777
	Ναι	89,73	52	11,298
	Total	88,33	144	12,269
Επικίνδυνο	Όχι	84,71	14	8,260
	Ναι	91,50	4	30,600
	Total	86,22	18	15,028
Ξαφνικός Θάνατος	Όχι	89,70	54	14,921
	Ναι	100,44	18	10,859
	Total	92,39	72	14,709
Total	Όχι	87,36	352	11,758
	Ναι	91,95	120	12,984
	Total	88,53	472	12,232

Από τα στοιχεία του πίνακα παρατηρούμε πως τα άτομα τα οποία δεν έχουν οικογενειακό ιστορικό παρουσιάζουν μικρότερες τιμές μέσης διαστολικής αρτηριακής πίεσης για όλες τις κατηγορίες της μεταβλητής “ πρώτο καρδιακό επεισόδιο “ .

Επιπλέον φαίνεται να μην υπάρχει κάποια συστηματική αύξηση της μέσης διαστολικής αρτηριακής πίεσης για τις κατηγορίες του πρώτου καρδιακού επεισοδίου στο υπό μελέτη δείγμα. Το μόνο που μπορούμε να παρατηρήσουμε είναι πως τα άτομα της κατηγορίας ξαφνικός θάνατος έχουν την μεγαλύτερη τιμή μέσης διαστολικής αρτηριακής πίεσης (Total = 92,39).

Τέλος, παρατηρώντας τη στήλη που παρουσιάζει την τυπική απόκλιση της μέσης διαστολικής αρτηριακής πίεσης αξίζει να αναφερθεί ότι η μεγαλύτερη διασπορά γύρω από τη μέση διαστολική αρτηριακή πίεση εμφανίζεται στα άτομα με οικογενειακό ιστορικό στην κατηγορία επικίνδυνο ΚΕ (st.dev = 30,60).

Στον ακόλουθο πίνακα παρουσιάζονται τα προεπιλεγμένα περιγραφικά στατιστικά στοιχεία για τη “Μέση διαστολική αρτηριακή πίεση” ανά κατηγορίες σοβαρότητας πρώτου καρδιακού επεισοδίου και επιβίωσης μετά από 10 χρόνια.

Report				
Μέση διαστολική αρτηριακή πίεση				
πρώτο καρδιακό επεισόδιο	Επιβίωση μετά από 10 χρόνια	Mean	N	Std. Deviation
Κανένα ΚΕ	Ζωντανός	87,66	238	10,946
	Total	87,66	238	10,946
Ακίνδυνο	Ζωντανός	87,37	118	12,395
	Πεθαμένος	92,69	26	10,855
Επικίνδυνο	Total	88,33	144	12,269
	Πεθαμένος	86,22	18	15,028
Ξαφνικός Θάνατος	Total	86,22	18	15,028
	Πεθαμένος	92,39	72	14,709
Total	Total	92,39	72	14,709
	Ζωντανός	87,56	356	11,430
Total	Πεθαμένος	91,50	116	14,062
	Total	88,53	472	12,232

Από τα στοιχεία του πίνακα παρατηρούμε πως στις κατηγορίες επικίνδυνο και ξαφνικός θάνατος της μεταβλητής πρώτο καρδιακό επεισόδιο δεν έχει επιβιώσει κανένας μετά από 10 χρόνια.

Επιπλέον φαίνεται να μην υπάρχει κάποια συστηματική αύξηση της μέσης διαστολικής αρτηριακής πίεσης για τις κατηγορίες του πρώτου καρδιακού επεισοδίου στο υπό μελέτη δείγμα. Το μόνο που μπορούμε να παρατηρήσουμε είναι πως τα άτομα της κατηγορίας ξαφνικός θάνατος έχουν την μεγαλύτερη τιμή μέσης διαστολικής αρτηριακής πίεσης (Total=92,39). Στην επόμενη παράγραφο θα εξεταστεί αν οι διαφορές είναι στατιστικά σημαντικές.

Τέλος, παρατηρώντας τη στήλη που παρουσιάζει την τυπική απόκλιση της μέσης διαστολικής αρτηριακής πίεσης αξίζει να αναφερθεί ότι η μεγαλύτερη διασπορά εμφανίζεται στα άτομα τα οποία δεν έχουν επιβιώσει μετά από 10 χρόνια στην κατηγορία επικίνδυνο ΚΕ (st.dev = 15,028).



Στον ακόλουθο πίνακα παρουσιάζονται τα προεπιλεγμένα περιγραφικά στατιστικά στοιχεία για τη “χοληστερόλη” ανά κατηγορίες σοβαρότητας πρώτου καρδιακού επεισοδίου και Οικογενειακό ιστορικό ΚΕ.

<b>Report</b>				
χοληστερόλη mg ανά dl				
πρώτο καρδιακό επεισόδιο	Επιβίωση μετά από 10 χρόνια	Mean	N	Std. Deviation
Κανένα ΚΕ	Ζωντανός	259,86	240	45,090
	Total	259,86	240	45,090
Ακίνδυνο	Ζωντανός	275,05	118	65,052
	Πεθαμένος	276,54	26	35,024
	Total	275,32	144	60,640
Επικίνδυνο	Πεθαμένος	249,33	18	55,920
	Total	249,33	18	55,920
Ξαφνικός Θάνατος	Πεθαμένος	260,86	72	55,089
	Total	260,86	72	55,089
Total	Ζωντανός	264,87	358	52,907
	Πεθαμένος	262,59	116	51,730
	Total	264,31	474	52,576

Από τα στοιχεία του πίνακα παρατηρούμε πως η μεγαλύτερη μέση τιμή για τη μεταβλητή χοληστερόλη παρατηρείται στα άτομα τα οποία δεν έχουν επιβιώσει μετά από 10 χρόνια για την κατηγορία ακίνδυνο της μεταβλητής πρώτο καρδιακό επεισόδιο (276,54) . Τέλος, παρατηρώντας τη στήλη που παρουσιάζει την τυπική απόκλιση της χοληστερόλης αξίζει να αναφερθεί ότι η μεγαλύτερη διασπορά εμφανίζεται στα άτομα τα οποία έχουν επιβιώσει μετά από 10 χρόνια στην κατηγορία ακίνδυνο ΚΕ (st.dev = 65,052).

Στον ακόλουθο πίνακα παρουσιάζονται τα προεπιλεγμένα περιγραφικά στατιστικά στοιχεία για τη “Σωματικό Βάρος” ανά κατηγορίες σοβαρότητας πρώτου καρδιακού επεισοδίου και επιβίωσης μετά από 10 χρόνια.

<b>Report</b>				
Σωματικό Βάρος				
πρώτο καρδιακό επεισόδιο	Επιβίωση μετά από 10 χρόνια	Mean	N	Std. Deviation
Κανένα ΚΕ	Ζωντανός	173,62	240	24,728
	Total	173,62	240	24,728
Ακίνδυνο	Ζωντανός	172,78	118	25,596
	Πεθαμένος	179,92	26	26,801
	Total	174,07	144	25,869
Επικίνδυνο	Πεθαμένος	169,89	18	11,235
	Total	169,89	18	11,235
Ξαφνικός Θάνατος	Πεθαμένος	170,56	72	18,209
	Total	170,56	72	18,209
Total	Ζωντανός	173,34	358	24,985
	Πεθαμένος	172,55	116	19,885
	Total	173,15	474	23,820

Από τον πίνακα παρατηρούμε πως η μεγαλύτερη μέση τιμή για το Σωματικό Βάρος αναφέρεται στα άτομα τα οποία δεν έχουν επιβιώσει μετά από 10 χρόνια στην κατηγορία ακίνδυνο της μεταβλητής πρώτο καρδιακό επεισόδιο.

Επιπλέον φαίνεται να μην υπάρχει κάποια συστηματική αύξηση του Σωματικού Βάρους για τις κατηγορίες του πρώτου καρδιακού επεισοδίου στο υπό μελέτη δείγμα. Το μόνο που μπορούμε να παρατηρήσουμε είναι πως τα άτομα της κατηγορίας ακίνδυνο έχουν την μεγαλύτερη τιμή Σωματικού Βάρους (Total = 174,07).

Τέλος, παρατηρώντας τη στήλη που παρουσιάζει την τυπική απόκλιση του Σωματικού Βάρους αξίζει να αναφερθεί ότι η μεγαλύτερη διασπορά εμφανίζεται στα άτομα τα οποία δεν έχουν επιβιώσει μετά από 10 χρόνια στην κατηγορία ακίνδυνο ΚΕ (st.dev = 26,801).

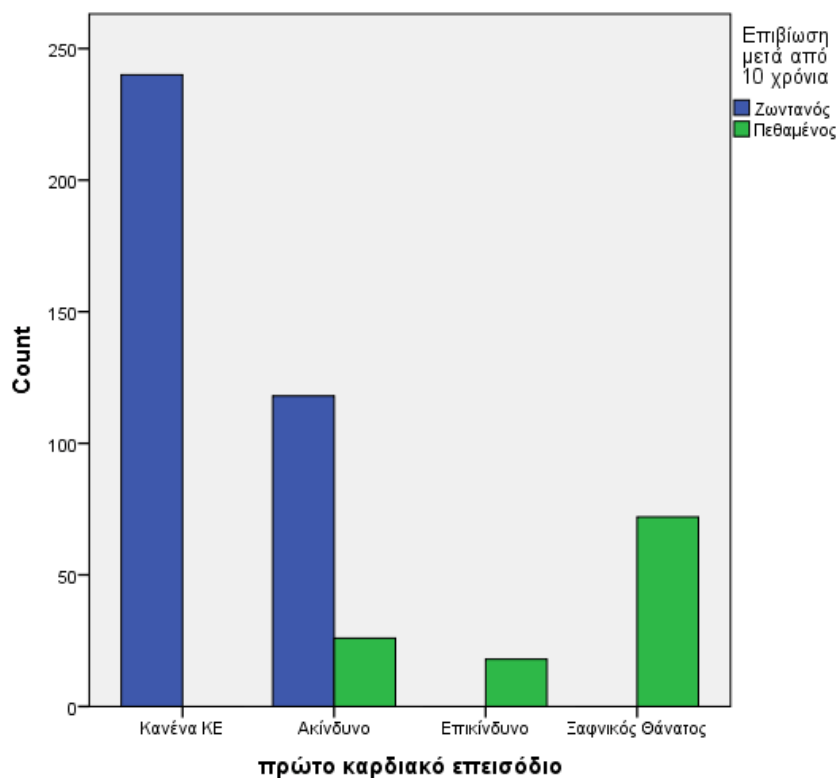
Στον ακόλουθο πίνακα παρουσιάζονται τα προεπιλεγμένα περιγραφικά στατιστικά στοιχεία για τον “Αριθμό τσιγάρων” ανά κατηγορίες σοβαρότητας πρώτου καρδιακού επεισοδίου και επιβίωσης μετά από 10 χρόνια.

<b>Report</b>				
Αριθμός τσιγάρων ανά ημέρα				
πρώτο καρδιακό επεισόδιο	Επιβίωση μετά από 10 χρόνια	Mean	N	Std. Deviation
Κανένα ΚΕ	Ζωντανός	9,62	238	12,158
	Total	9,62	238	12,158
Ακίνδυνο	Ζωντανός	13,36	118	12,998
	Πεθαμένος	17,00	26	9,269
Επικίνδυνο	Total	14,01	144	12,459
	Πεθαμένος	16,89	18	9,311
Ξαφνικός Θάνατος	Total	16,89	18	9,311
	Πεθαμένος	12,56	72	11,833
Total	Total	12,56	72	11,833
	Ζωντανός	10,86	356	12,548
Total	Πεθαμένος	14,22	116	11,070
	Total	11,69	472	12,276

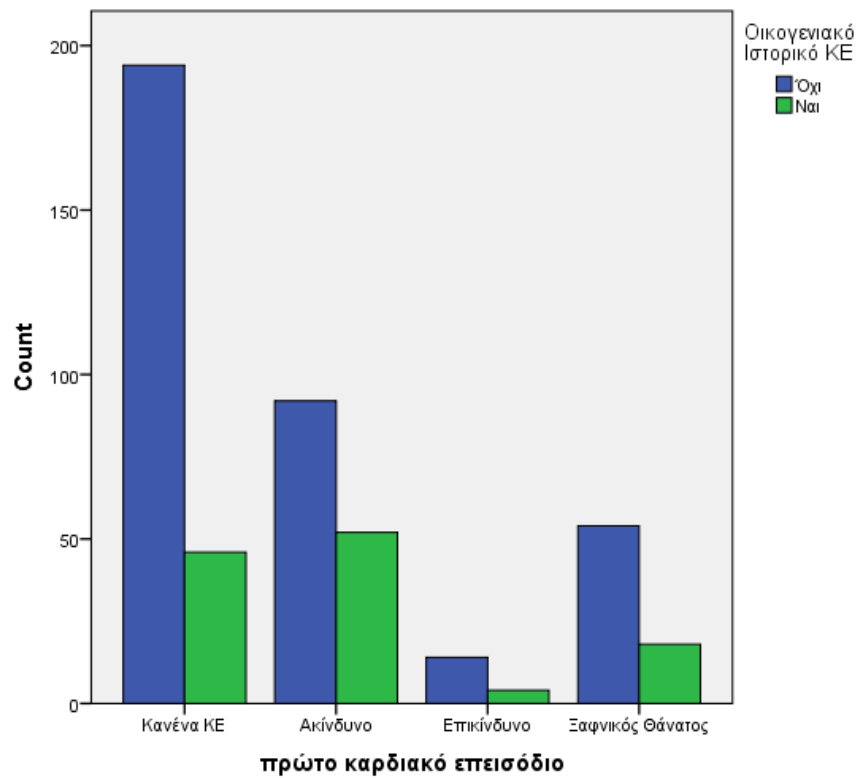
Από τον πίνακα φαίνεται να μην υπάρχει κάποια συστηματική αύξηση του Αριθμού τσιγάρων ανά ημέρα για τις κατηγορίες του πρώτου καρδιακού επεισοδίου στο υπό μελέτη δείγμα. Παρ’ όλα αυτά φαίνεται να υπάρχουν στατιστικά σημαντικές διαφορές στις μέσες τιμές, κάτι το οποίο θα εξεταστεί στη συνέχεια της εργασίας. Το μόνο που μπορούμε να παρατηρήσουμε είναι πως τα άτομα της κατηγορίας επικίνδυνο πρώτο καρδιακό επεισόδιο έχουν την μεγαλύτερη τιμή Αριθμού τσιγάρων ανά ημέρα (Total = 16,89).

Όσον αφορά τη σχέση της επιβίωσης μετά από 10 χρόνια με τον αριθμό των τσιγάρων ανά ημέρα, φαίνεται ότι όσοι δεν επιβίωσαν κάπνιζαν περισσότερο σε σχέση με αυτούς που επέζησαν. Η διαφορά αυτή στις μέσες τιμές θα εξεταστεί και παρακάτω.

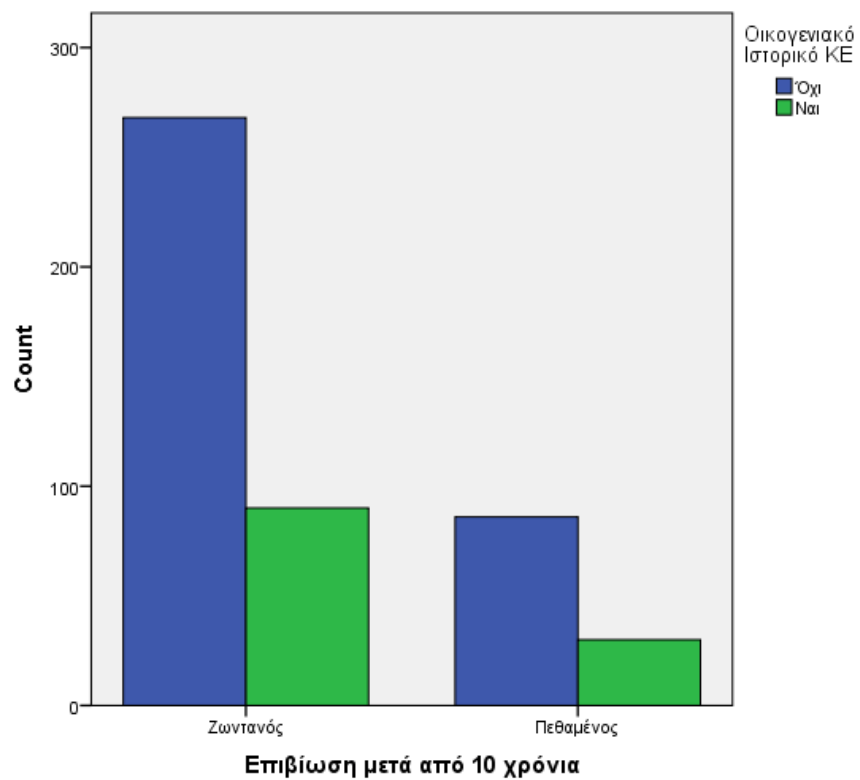
Ακολούθως, παρατίθενται τα συνδυαστικά διαγράμματα (clustered) μεταξύ των κατηγορικών μεταβλητών πρώτο καρδιακό επεισόδιο – επιβίωση μετά από 10 χρόνια – οικογενειακό ιστορικό ΚΕ.



Από το πρώτο συνδυαστικό διάγραμμα διαπιστώνουμε ότι το ποσοστό των ατόμων που έχουν επιβιώσει μετά από 10 χρόνια είναι πολύ μεγαλύτερο από αυτό των θανόντων για ακίνδυνο πρώτο καρδιακό επεισόδιο, ενώ στην υποκατηγορία επικίνδυνο δεν έχει επιβιώσει κανείς.



Στο δεύτερο συνδυαστικό διάγραμμα αξίζει να τονιστεί ότι τα άτομα τα οποία δεν έχουν οικογενειακό ιστορικό και έχουν υποστεί ακίνδυνο, επικίνδυνο και ξαφνικό θάνατο λόγω πρώτου καρδιακού επεισοδίου είναι περισσότερα σε σχέση με αυτά που έχουν οικογενειακό ιστορικό.



Στο τρίτο συνδυαστικό διάγραμμα διαπιστώνεται ότι στα άτομα που δεν έχουν επιβιώσει μετά από 10 χρόνια μετά από καρδιακό επεισόδιο τα περισσότερα δεν είχαν οικογενειακό ιστορικό ΚΕ.

## 2 Έλεγχοι με επαγωγικές διαδικασίες και η στατιστική ερμηνεία τους (B & Γ)

### 2.1 Έλεγχος ανεξαρτησίας $X^2$

Με βάση τα παραπάνω στοιχεία τα οποία είναι αποτέλεσμα του ερωτήματος Α απορρέουν μερικά ερωτήματα τα οποία μπορούν να ελεγχθούν με συγκεκριμένες επαγωγικές διαδικασίες.

Αρχικά, θα ήταν σκόπιμο να ελέγξουμε την ανεξαρτησία ορισμένων μεταβλητών οι οποίες θεωρούνται χαρακτηριστικές. Η σχέση ανάμεσα στην μεταβλητή "Πρώτο καρδιακό επεισόδιο" και την "Οικογενειακό ιστορικό" θα ήταν ενδιαφέρον να ερευνηθεί για τυχόν ύπαρξη σχέσης ανάμεσα τους.

Με την εφαρμογή του ελέγχου ανεξαρτησίας  $X^2$  παίρνουμε τις παρακάτω πληροφορίες σε μορφή πινάκων.

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Οικογενειακό Ιστορικό ΚΕ * πρώτο καρδιακό επεισόδιο	474	100,0%	0	0,0%	474	100,0%

Στον πίνακα **Case Processing Summary**, όπως και παραπάνω με τον πίνακα **Statistics** μπορούμε να πάρουμε πληροφορίες για τυχόν cases που δεν συμπεριλήφθηκαν από την στήλη Missing. Όπως είναι εμφανές, από ένα σύνολο 474 cases, όλες λήφθηκαν υπόψη.

Οικογενειακό Ιστορικό ΚΕ * πρώτο καρδιακό επεισόδιο Crosstabulation							
			πρώτο καρδιακό επεισόδιο				Total
			Κανένα ΚΕ	Ακίνδυνο	Επικίνδυνο	Ξαφνικός Θάνατος	
Οικογενειακό Ιστορικό ΚΕ	Όχι	Count	194	92	14	54	354
		% within Οικογενειακό Ιστορικό ΚΕ	54,8%	26,0%	4,0%	15,3%	100,0%
		% within πρώτο καρδιακό επεισόδιο	80,8%	63,9%	77,8%	75,0%	74,7%
	Ναι	Count	46	52	4	18	120
		% within Οικογενειακό Ιστορικό ΚΕ	38,3%	43,3%	3,3%	15,0%	100,0%
		% within πρώτο καρδιακό επεισόδιο	19,2%	36,1%	22,2%	25,0%	25,3%
Total		Count	240	144	18	72	474
		% within Οικογενειακό Ιστορικό ΚΕ	50,6%	30,4%	3,8%	15,2%	100,0%
		% within πρώτο καρδιακό επεισόδιο	100,0%	100,0%	100,0%	100,0%	100,0%

Από τον πίνακα συνάφειας παρατηρούμε τα εξής:

- Όσοι δεν είχαν κανένα καρδιακό επεισόδιο χωρίς οικογενειακό ιστορικό (80,8%) εμφανίζουν μεγαλύτερο ποσοστό σε σχέση με όσους είχαν οικογενειακό ιστορικό (19,2%).

- Όσοι υπέστησαν ακίνδυνο καρδιακό επεισόδιο χωρίς οικογενειακό ιστορικό (63,9%) είναι σχεδόν διπλάσιοι σε ποσοστό σε σχέση με αυτούς που είχαν οικογενειακό ιστορικό (36,1%).
- Επιπλέον παρατηρείται ότι το ποσοστό εμφάνισης καρδιακού επεισοδίου (από κανένα σε ακίνδυνο καρδιακό) αυξάνεται σχεδόν στο διπλάσιο για όσους έχουν οικογενειακό ιστορικό.

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13,770 <sup>a</sup>	3	,003
Likelihood Ratio	13,408	3	,004
Linear-by-Linear Association	1,871	1	,171
N of Valid Cases	474		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 4,56.

Στον πίνακα Chi-Square Tests παρατηρούμε ότι  $X^2 = 13,77$  και  $p=0,003$ . Δηλαδή, πρόκειται για ένα στατιστικά σημαντικό τεστ. Επιπλέον, δεδομένου ότι το  $p < 0,05$ , μπορούμε να συμπεράνουμε ότι οι μεταβλητές "Οικογενειακό Ιστορικό" και "Πρώτο Καρδιακό Επεισόδιο" συσχετίζονται η μία με την άλλη. Κατά συνέπεια, απορρίπτουμε την μηδενική υπόθεση καθώς υπάρχει εξάρτηση ανάμεσα στις δύο αυτές μεταβλητές.

Στην συνέχεια θα μελετήσουμε την σχέση εξάρτησης ανάμεσα στις μεταβλητές "Πρώτο Καρδιακό Επεισόδιο" και "Επιβίωση μετά από 10 χρόνια".

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
πρώτο καρδιακό επεισόδιο* Επιβίωση μετά από 10 χρόνια	474	100,0%	0	0,0%	474	100,0%

Στον πίνακα Case Processing Summary, όπως και προηγουμένως παρατηρούμε ότι όλες οι τιμές λήφθηκαν υπόψιν.



Επιβίωση μετά από 10 χρόνια * πρώτο καρδιακό επεισόδιο Crosstabulation							
			πρώτο καρδιακό επεισόδιο				Total
			Κανένα ΚΕ	Ακίνδυνο	Επικίνδυνο	Ξαφνικός Θάνατος	
Επιβίωση μετά από 10 χρόνια	Ζωντανός	Count	240	118	0	0	358
		% within Επιβίωση μετά από 10 χρόνια	67,0%	33,0%	0,0%	0,0%	100,0%
		% within πρώτο καρδιακό επεισόδιο	100,0%	81,9%	0,0%	0,0%	75,5%
	Πεθαμένος	Count	0	26	18	72	116
		% within Επιβίωση μετά από 10 χρόνια	0,0%	22,4%	15,5%	62,1%	100,0%
		% within πρώτο καρδιακό επεισόδιο	0,0%	18,1%	100,0%	100,0%	24,5%
Total		Count	240	144	18	72	474
		% within Επιβίωση μετά από 10 χρόνια	50,6%	30,4%	3,8%	15,2%	100,0%
		% within πρώτο καρδιακό επεισόδιο	100,0%	100,0%	100,0%	100,0%	100,0%

Από τον πίνακα συνάφειας παρατηρούμε τα εξής:

- Για όσους δεν υπέστησαν κανένα καρδιακό επεισόδιο η θνησιμότητα είναι μηδενική μετά από 10 χρόνια.
- Σε όσους υπέστησαν ακίνδυνο πρώτο καρδιακό επεισόδιο το μεγαλύτερο ποσοστό παρέμεινε ζωντανό μετά από 10 χρόνια (81,9%) σε σχέση με όσους απεβίωσαν (18,1%).
- Όσον αφορά τις υποκατηγορίες επικίνδυνο και ξαφνικός θάνατος τα αποτελέσματα που προέκυψαν είναι απολύτως λογικά, καθώς η θνησιμότητα φτάνει σε ποσοστό 100%.

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	358,732 <sup>a</sup>	3	,000
Likelihood Ratio	391,526	3	,000
Linear-by-Linear Association	332,062	1	,000
N of Valid Cases	474		

a. 1 cells (12,5%) have expected count less than 5. The minimum expected count is 4,41.

Στον πίνακα Chi-Square Tests παρατηρούμε ότι  $X^2 = 358,73$  και  $p=0,0001$ . Δηλαδή, πρόκειται για ένα στατιστικά σημαντικό τεστ. Επιπλέον, δεδομένου ότι το  $p < 0,05$ , μπορούμε να συμπεράνουμε ότι οι μεταβλητές "Πρώτο Καρδιακό Επεισόδιο" και "Επιβίωση μετά από 10 χρόνια" συσχετίζονται η μία με την άλλη. Κατά συνέπεια, απορρίπτουμε την μηδενική υπόθεση καθώς υπάρχει εξάρτηση ανάμεσα στις δύο αυτές μεταβλητές.

Στην συνέχεια θα μελετήσουμε την σχέση εξάρτησης ανάμεσα στις μεταβλητές "Οικογενειακό Ιστορικό ΚΕ" και "Επιβίωση μετά από 10 χρόνια".

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Οικογενειακό Ιστορικό ΚΕ * Επιβίωση μετά από 10 χρόνια	474	100,0%	0	0,0%	474	100,0%

Στον πίνακα **Case Processing Summary**, όπως και προηγουμένως παρατηρούμε ότι όλες οι τιμές λήφθηκαν υπόψη.

Οικογενειακό Ιστορικό ΚΕ * Επιβίωση μετά από 10 χρόνια Crosstabulation					
			Επιβίωση μετά από 10 χρόνια		Total
			Ζωντανός	Πεθαμένος	
Οικογενειακό Ιστορικό ΚΕ	Όχι	Count	268	86	354
		% within Οικογενειακό Ιστορικό ΚΕ	75,7%	24,3%	100,0%
		% within Επιβίωση μετά από 10 χρόνια	74,9%	74,1%	74,7%
	Ναι	Count	90	30	120
		% within Οικογενειακό Ιστορικό ΚΕ	75,0%	25,0%	100,0%
		% within Επιβίωση μετά από 10 χρόνια	25,1%	25,9%	25,3%
Total		Count	358	116	474
		% within Οικογενειακό Ιστορικό ΚΕ	75,5%	24,5%	100,0%
		% within Επιβίωση μετά από 10 χρόνια	100,0%	100,0%	100,0%

Από τον πίνακα συνάφειας παρατηρούμε τα εξής:

- Το ποσοστό των ατόμων που επιβίωσαν μετά από 10 χρόνια χωρίς να έχουν οικογενειακό ιστορικό ΚΕ είναι περισσότεροι σε σχέση με αυτούς που έχουν.
- Αυτό που αξίζει να τονιστεί είναι πως το ποσοστό των ατόμων που απεβίωσαν μετά από 10 χρόνια είναι μεγαλύτερο για όσους δεν έχουν οικογενειακό ιστορικό ΚΕ (74,1%) σε σχέση με όσους έχουν (25,9%).

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,024 <sup>a</sup>	1	,876		
Continuity Correction <sup>b</sup>	,001	1	,974		
Likelihood Ratio	,024	1	,877		
Fisher's Exact Test				,902	,483
Linear-by-Linear Association	,024	1	,877		
N of Valid Cases	474				
a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 29,37.					
b. Computed only for a 2x2 table					

Στον πίνακα Chi-Square Tests παρατηρούμε ότι  $X^2 = 0,024$  και  $p=0,876$ . Είναι προφανές ότι δεν υπάρχει ιδιαίτερη συσχέτιση ανάμεσα στις δύο μεταβλητές. Συγκεκριμένα, από την τιμή του p-value η οποία είναι μεγαλύτερη του 5% μπορούμε να πούμε ότι οι δύο αυτές μεταβλητές **είναι ανεξάρτητες μεταξύ τους**.

Αξίζει να σημειωθεί πως το ποσοστό που υπάρχει στην παρένθεση της υποσημείωσης σε όλους τους πίνακες  $X^2$  αναφέρεται στο ποσοστό των αναμενόμενων συχνοτήτων που είναι μικρότερες από 5 και δεν ξεπερνάει το 20%.

## 2.2 Έλεγχοι διαφορών ως προς μεταβλητές nominal

### 2.2.1 Έλεγχος κανονικότητας

Ξεκινώντας με τις μεταβλητές "Αριθμός Τσιγάρων ανά Ημέρα" και "Επιβίωση μετά από 10 χρόνια". Μέσω του ελέγχου ως προς την κανονική κατανομή με το Kolmogorov-Smirnov και Shapiro-Wilk παίρνουμε τους παρακάτω πίνακες και διαγράμματα.

Case Processing Summary							
	Επιβίωση μετά από 10 χρόνια	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Αριθμός τσιγάρων ανά ημέρα	Ζωντανός	356	99,4%	2	0,6%	358	100,0%
	Πεθαμένος	116	100,0%	0	0,0%	116	100,0%

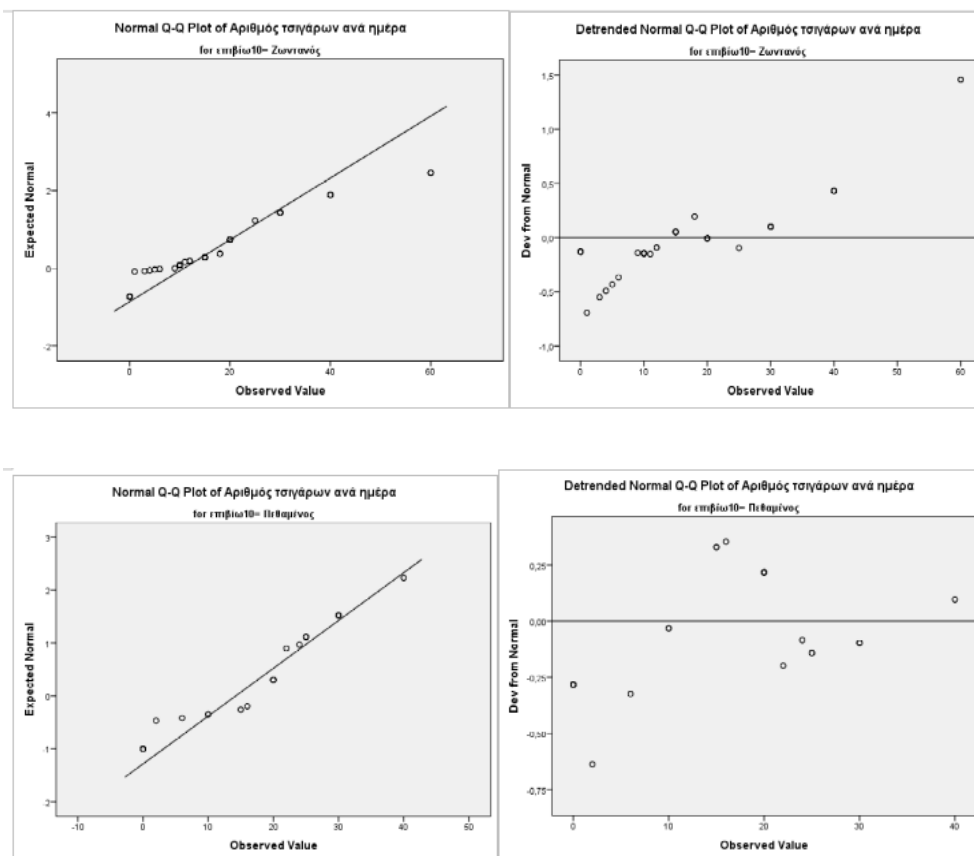
Ο πίνακας **Case Processing Summary**, όπως και προηγουμένως μας δίνει πληροφορίες για το πόσα από τα cases που υπήρχαν αρχικά στο αρχείο μας, λήφθηκαν υπόψιν.

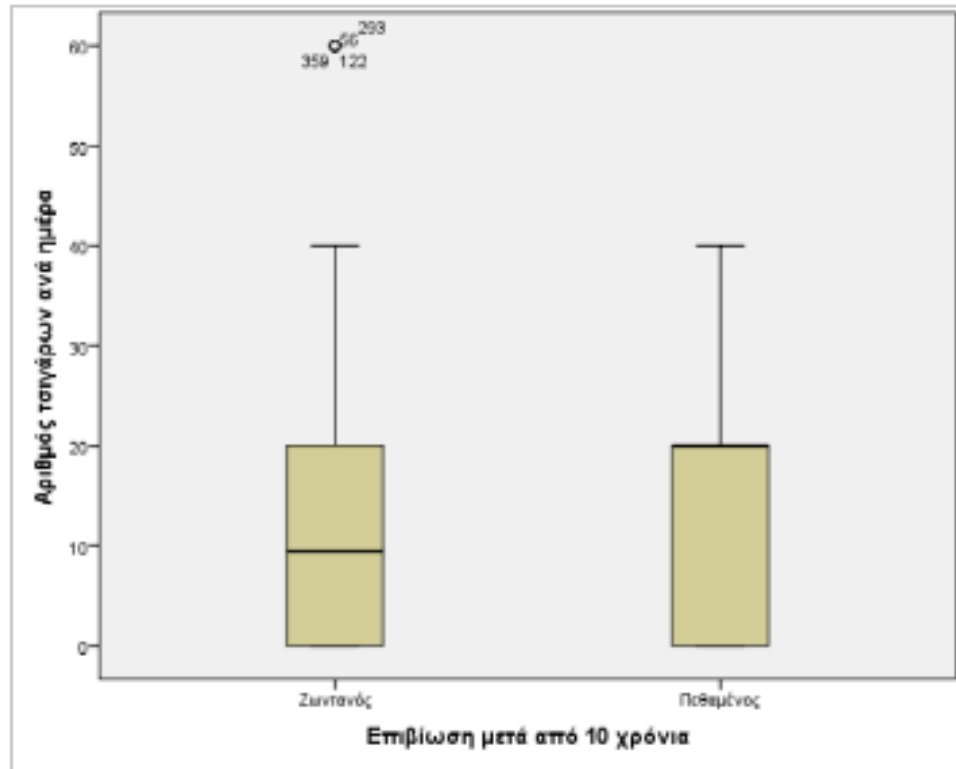
Tests of Normality							
		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Αριθμός τσιγάρων ανά ημέρα	Ζωντανός	,267	356	,000	,801	356	,000
	Πεθαμένος	,268	116	,000	,841	116	,000

a. Lilliefors Significance Correction

Από τον πίνακα Tests of Normality μπορούμε να πάρουμε αποτελέσματα από δύο τεστ κανονικότητας, τα Kolmogorov-Smirnov και Shapiro-Wilk. Το δεύτερο χρησιμοποιείται κυρίως για μικρά δείγματα ενώ το πρώτο δίνει καλά αποτελέσματα και για μεγάλα. Στην περίπτωση μας έχουμε δείγμα από 474 ασθενείς οπότε θα χρησιμοποιήσουμε το πρώτο. Παρατηρώντας τον πίνακα βλέπουμε ότι το p-value και στις δύο περιπτώσεις είναι κατά πολύ μικρότερο του 5%. Συνεπώς, τα δεδομένα μας αποκλίνουν από την κανονικότητα.

Αποτελέσματα για την κανονικότητα μπορούμε να πάρουμε και γραφικά μέσω των Q-Q plots. Όπως βλέπουμε και στα δύο διαγράμματα, τα δεδομένα μας έχουν την τάση να απομακρυνθούν από την διαγώνιο. Κατά συνέπεια, τα διαγράμματα όπως ήταν αναμενόμενο συμφωνούν με τον πίνακα.





Παρατηρούμε ότι από τον πίνακα **Test Statistics** υπάρχει σημαντική διαφορά ανάμεσα στις μέσες τιμές των μεταβλητών Επιβίωση μετά από 10 χρόνια και Αριθμού τσιγάρων σύμφωνα με το Mann-Whitney. Αυτό προκύπτει από το γεγονός ότι το p-value είναι  $<5\%$ . Συμπεραίνουμε ότι τα άτομα που δεν επιβίωσαν είχαν την τάση να καπνίζουν περισσότερα τσιγάρα.

Ranks				
	Επιβίωση μετά από 10 χρόνια	N	Mean Rank	Sum of Ranks
Αριθμός τσιγάρων ανά ημέρα	Ζωντανός	356	224.51	79924.00
	Πεθαμένος	116	273.31	31704.00
	Total	472		

Test Statistics	
	<u>Αριθμός τσιγάρων ανά ημέρα</u>
Mann-Whitney U	16378.000
<u>Wilcoxon W</u>	79924.000
Z	-3.521
<u>Asymp. Sig. (2-tailed)</u>	.000
a. Grouping Variable: Επιβίωση μετά από 10 χρόνια	

Ο πίνακας **Case Processing Summary** μας δίνει πληροφορίες για το πόσα από τα cases που υπήρχαν αρχικά στο αρχείο μας, λήφθηκαν υπόψιν.

Case Processing Summary							
	πρώτο καρδιακό επεισόδιο	Valid		Cases Missing		Total	
		N	Percent	N	Percent	N	Percent
Αριθμός τσιγάρων ανά ημέρα	Κανένα ΚΕ	238	99.2%	2	0.8%	240	100.0%
	Ακίνδυνο	144	100.0%	0	0.0%	144	100.0%
	Επικίνδυνο	18	100.0%	0	0.0%	18	100.0%
	Ξαφνικός Θάνατος	72	100.0%	0	0.0%	72	100.0%

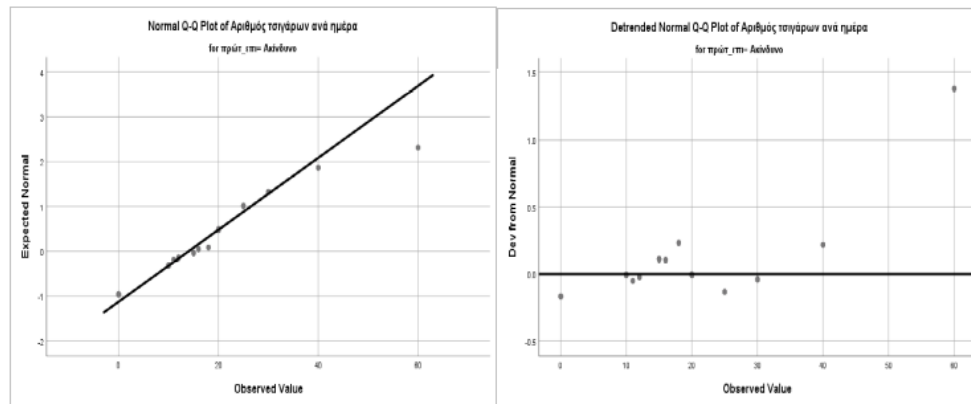
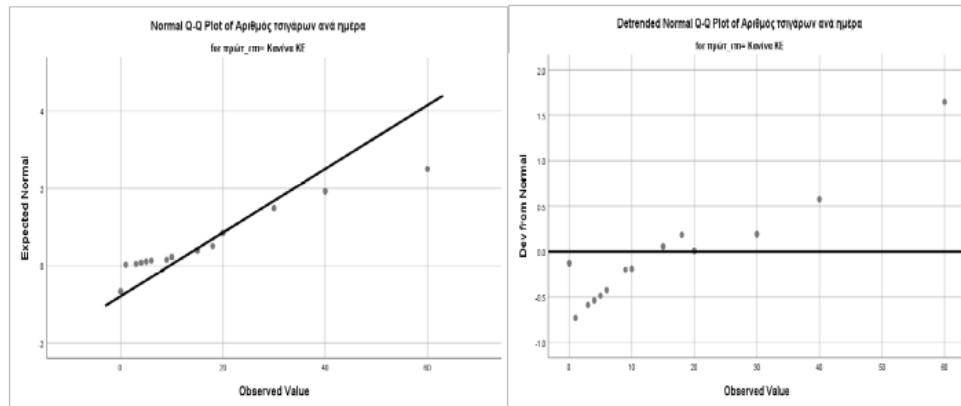
Από τον πίνακα Tests of Normality μπορούμε να πάρουμε αποτελέσματα από δύο τεστ κανονικότητας, τα Kolmogorov-Smirnov και Shapiro-Wilk. Παρατηρώντας τον πίνακα βλέπουμε ότι το p-value και στις δύο περιπτώσεις είναι κατά πολύ μικρότερο του 5%. Άρα οι μεταβλητές μας δεν ακολουθούν κανονική κατανομή.

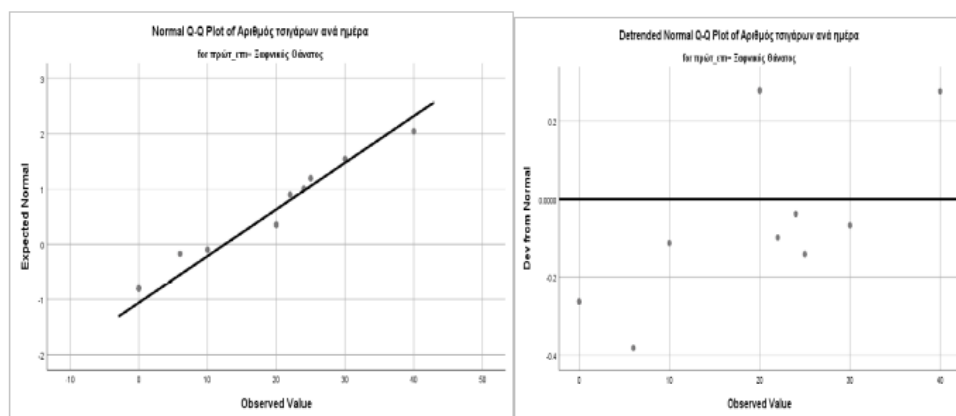
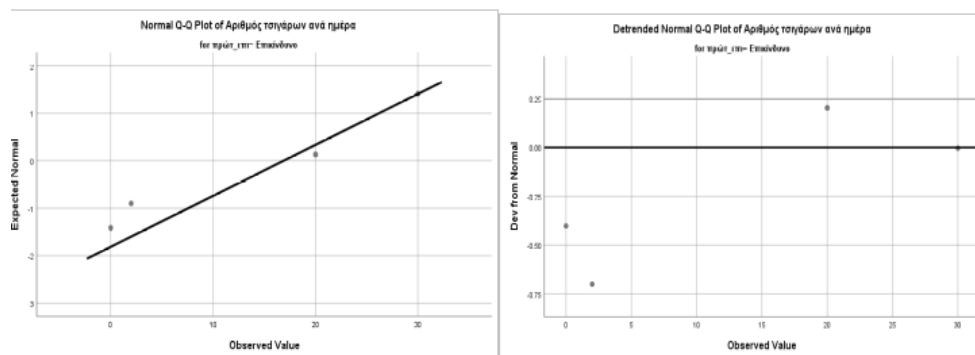
Όσο αναφορά την κατηγορία “Επικίνδυνο” έχουμε λίγες μετρήσεις και αρκετές ακραίες τιμές. Όμως η αφαίρεση τους αφήνει το δείγμα με 12 μόνο μετρήσεις με αποτέλεσμα τα tests κανονικότητας να μην δίνουν αποτελέσματα. Αξίζει να αναφέρουμε ότι τα δεδομένα για αυτήν την κατηγορία προφανώς δεν είναι αρκετά αλλά από ερευνητική σκοπιά κρίναμε αναγκαίο να μελετηθούν οι επιπτώσεις του τσιγάρου όσο αναφορά το “πρώτο καρδιακό επεισόδιο”. Επιπλέον οι μη παραμετρικοί έλεγχοι μπορούν να εφαρμοστούν σε περιπτώσεις πολύ μικρών δειγμάτων.

## Tests of Normality

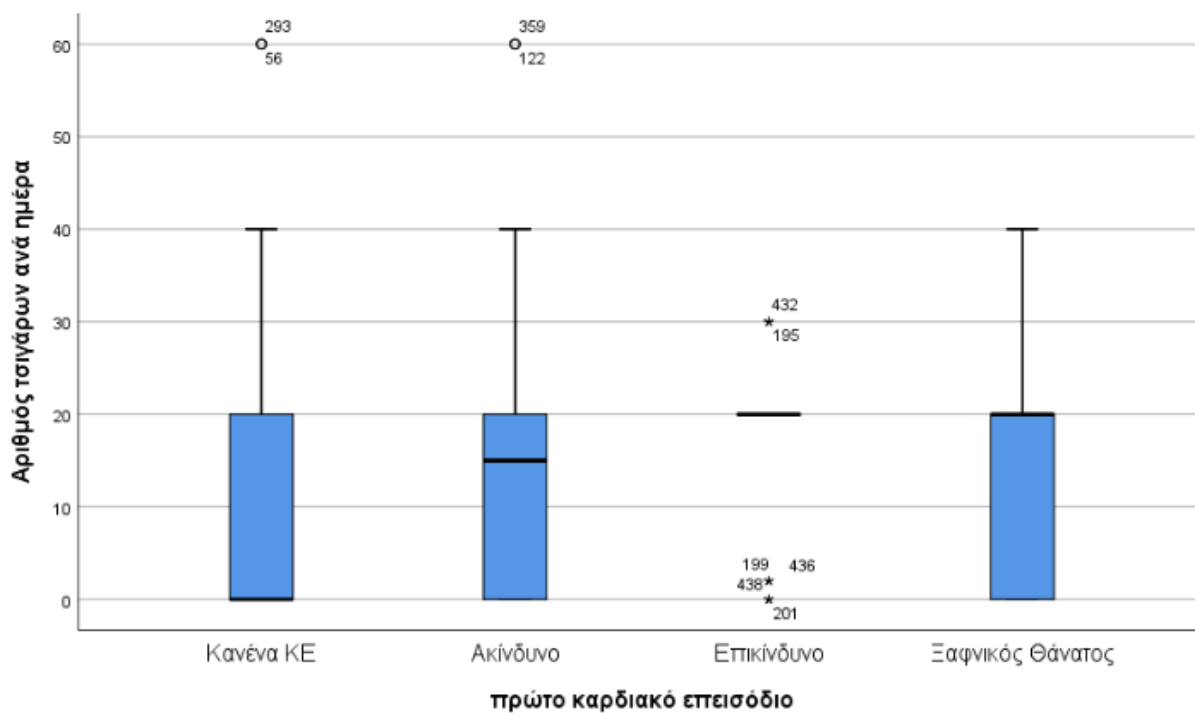
	πρώτο καρδιακό επεισόδιο	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Αριθμός τσιγάρων ανά ημέρα	Κανένα ΚΕ	.290	238	.000	.771	238	.000
	Ακίνδυνο	.203	144	.000	.867	144	.000
	Επικίνδυνο	.409	18	.000	.724	18	.000
	Ξαφνικός Θάνατος	.272	72	.000	.812	72	.000

a. Lilliefors Significance Correction









Παρατηρούμε ότι από τον πίνακα **Test Statistics** υπάρχει σημαντική διαφορά ανάμεσα στις μέσες τιμές των μεταβλητών "Πρώτο καρδιακό επεισόδιο" και τον "Αριθμό τσιγάρων". Αυτό προκύπτει από το γεγονός ότι το p-value είναι  $<5\%$ . Συμπεραίνουμε ότι ο αριθμός τσιγάρων φαίνεται να επηρεάζει την επικινδυνότητα του πρώτου καρδιακού επεισοδίου.

## Kruskal-Wallis Test

Ranks			
Αριθμός τσιγάρων ανά ημέρα	πρώτο καρδιακό επεισόδιο	N	Mean Rank
	Κανένα ΚΕ	238	211.33
	Ακίνδυνο	144	260.17
	Επικίνδυνο	18	313.50
	Ξαφνικός Θάνατος	72	253.11
	Total	472	

### Test Statistics<sup>a,b</sup>

	Αριθμός τσιγάρων ανά ημέρα
Kruskal-Wallis H	21.303
df	3
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: πρώτο  
καρδιακό επεισόδιο

Εφόσον το Kruskal-Wallis δείχνει ότι υπάρχουν σημαντικές διαφορές στις μέσες τιμές, θα πραγματοποιήσουμε τον έλεγχο Mann-Whitney, ώστε να εντοπιστούν οι κατηγορίες στις οποίες εμφανίζονται οι διαφορές.

### Mann-Whitney Test

Ranks				
	πρώτο καρδιακό επεισόδιο	N	Mean Rank	Sum of Ranks
Αριθμός τσιγάρων ανά ημέρα	Κανένα ΚΕ	238	176.40	41983.00
	Ακίνδυνο	144	216.46	31170.00
	Total	382		

Παρατηρούμε ότι από τον πίνακα **Test Statistics** υπάρχει σημαντική διαφορά ανάμεσα στις μέσες τιμές των μεταβλητών "Κανένα ΚΕ" και "Ακίνδυνο" σε σχέση με τον "Αριθμό τσιγάρων". Αυτό προκύπτει από το γεγονός ότι το p-value είναι <5%. Συμπεραίνουμε ότι τα άτομα που είναι στην κατηγορία "Κανένα ΚΕ" είχαν την τάση να καπνίζουν λιγότερα τσιγάρα σε σχέση με αυτά της κατηγορίας "Ακίνδυνο".

Test Statistics <sup>a</sup>	
	Αριθμός τσιγάρων ανά ημέρα
Mann-Whitney U	13542.000
Wilcoxon W	41983.000
Z	-3.622
Asymp. Sig. (2-tailed)	.000
a. Grouping Variable: πρώτο καρδιακό επεισόδιο	

Για τις διαφορές “Κανένα ΚΕ” και “Επικίνδυνο” φαίνεται και εδώ στατιστικά σημαντική διαφορά στις μέσες τιμές σε σχέση με τον “Αριθμό τσιγάρων”. Αυτό προκύπτει από το γεγονός ότι το p-value είναι <5%. Φαίνεται να καπνίζουν λιγότερο τα άτομα στην κατηγορία “Κανένα ΚΕ” σε σχέση με αυτούς στην κατηγορία “Επικίνδυνο”.

Ranks				
	πρώτο καρδιακό επεισόδιο	N	Mean Rank	Sum of Ranks
Αριθμός τσιγάρων ανά ημέρα	Κανένα ΚΕ	238	124.47	29625.00
	Επικίνδυνο	18	181.72	3271.00
	Total	256		

Test Statistics <sup>a</sup>	
	Αριθμός τσιγάρων ανά ημέρα
Mann-Whitney U	1184.000
Wilcoxon W	29625.000
Z	-3.377
Asymp. Sig. (2-tailed)	.001
a. Grouping Variable: πρώτο καρδιακό επεισόδιο	

Για τις διαφορές “Κανένα ΚΕ” και “Ξαφνικός θάνατος” φαίνεται και εδώ στατιστικά σημαντική διαφορά στις μέσες τιμές σε σχέση με τον “Αριθμό τσιγάρων”. Αυτό προκύπτει από το γεγονός ότι το p-value είναι  $<5\%$ . Φαίνεται να καπνίζουν λιγότερο τα άτομα στην κατηγορία “Κανένα ΚΕ” σε σχέση με αυτούς στην κατηγορία “Ξαφνικός θάνατος”.

### Mann-Whitney Test

Ranks				
	πρώτο καρδιακό επεισόδιο	N	Mean Rank	Sum of Ranks
Αριθμός τσιγάρων ανά ημέρα	Κανένα ΚΕ	238	149.46	35571.00
	Ξαφνικός Θάνατος	72	175.47	12634.00
	Total	310		

### Test Statistics<sup>a</sup>

	Αριθμός τσιγάρων ανά ημέρα
Mann-Whitney U	7130.000
Wilcoxon W	35571.000
Z	-2.310
Asymp. Sig. (2-tailed)	.021

a. Grouping Variable: πρώτο  
καρδιακό επεισόδιο

Για τις διαφορές “Ακίνδυνο” και “Επικίνδυνο” φαίνεται ότι δεν υπάρχει στατιστικά σημαντική διαφορά στις μέσες τιμές σε σχέση με τον “Αριθμό τσιγάρων”. Αυτό προκύπτει από το γεγονός ότι το p-value είναι  $>5\%$ . Δηλαδή ο αριθμός τσιγάρων που καπνίζεται από τις δυο κατηγορίες είναι ίδιος.

Ranks				
	πρώτο καρδιακό επεισόδιο	N	Mean Rank	Sum of Ranks
Αριθμός τσιγάρων ανά ημέρα	Ακίνδυνο	144	79.43	11438.00
	Επικίνδυνο	18	98.06	1765.00
	Total	162		

### Test Statistics<sup>a</sup>

	Αριθμός τσιγάρων ανά ημέρα
Mann-Whitney U	998.000
Wilcoxon W	11438.000
Z	-1.645
Asymp. Sig. (2-tailed)	.100

a. Grouping Variable: πρώτο  
καρδιακό επεισόδιο

Για τις διαφορές “Ακίνδυνο” και “Ξαφνικός θάνατος” φαίνεται ότι δεν υπάρχει στατιστικά σημαντική διαφορά στις μέσες τιμές σε σχέση με τον “Αριθμό τσιγάρων”. Αυτό προκύπτει από το γεγονός ότι το p-value είναι  $>5\%$ . Δηλαδή ο αριθμός τσιγάρων που καπνίζεται από τις δυο κατηγορίες είναι ίδιος.

### Ranks

	πρώτο καρδιακό επεισόδιο	N	Mean Rank	Sum of Ranks
Αριθμός τσιγάρων ανά ημέρα	Ακίνδυνο	144	109.28	15736.00
	Ξαφνικός Θάνατος	72	106.94	7700.00
	Total	216		

### Test Statistics<sup>a</sup>

	Αριθμός τσιγάρων ανά ημέρα
Mann-Whitney U	5072.000
Wilcoxon W	7700.000
Z	-.269
Asymp. Sig. (2-tailed)	.788

a. Grouping Variable: πρώτο  
καρδιακό επεισόδιο

Για τις διαφορές “Επικίνδυνο” και “Ξαφνικός θάνατος” φαίνεται ότι δεν υπάρχει στατιστικά σημαντική διαφορά στις μέσες τιμές σε σχέση με τον “Αριθμό τσιγάρων”. Αυτό προκύπτει από το γεγονός ότι το p-value είναι  $>5\%$ . Δηλαδή ο αριθμός τσιγάρων που καπνίζεται από τις δυο κατηγορίες είναι ίδιος.

### Mann-Whitney Test

Ranks				
	πρώτο καρδιακό επεισόδιο	N	Mean Rank	Sum of Ranks
Αριθμός τσιγάρων ανά ημέρα	Επικίνδυνο	18	52.72	949.00
	Ξαφνικός Θάνατος	72	43.69	3146.00
	Total	90		

### Test Statistics<sup>a</sup>

	Αριθμός τσιγάρων ανά ημέρα
Mann-Whitney U	518.000
Wilcoxon W	3146.000
Z	-1.389
Asymp. Sig. (2-tailed)	.165

a. Grouping Variable: πρώτο  
καρδιακό επεισόδιο

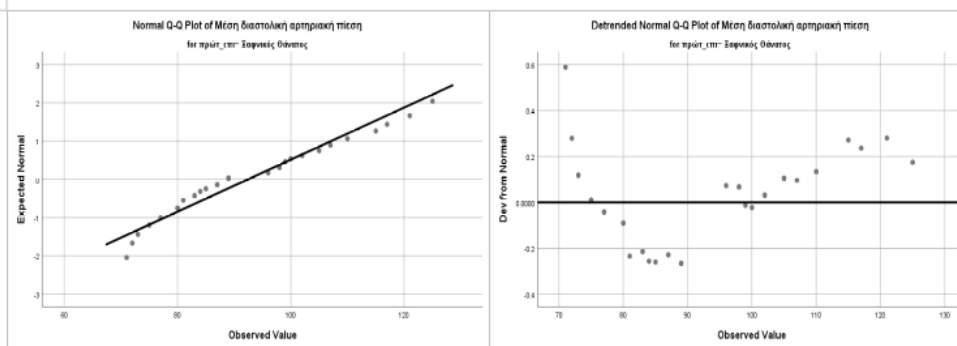
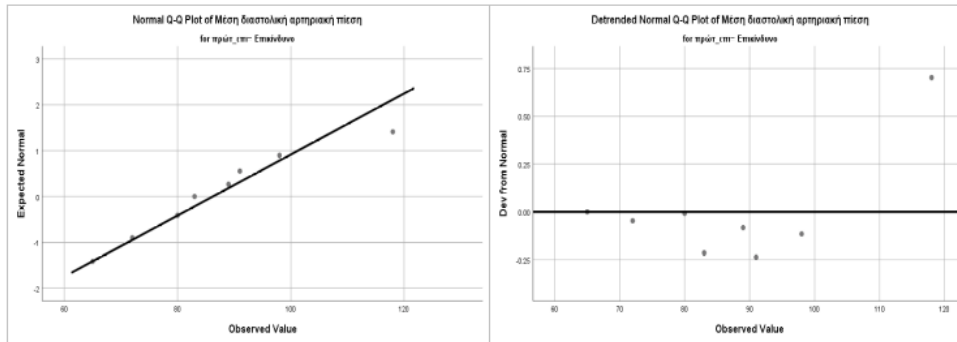
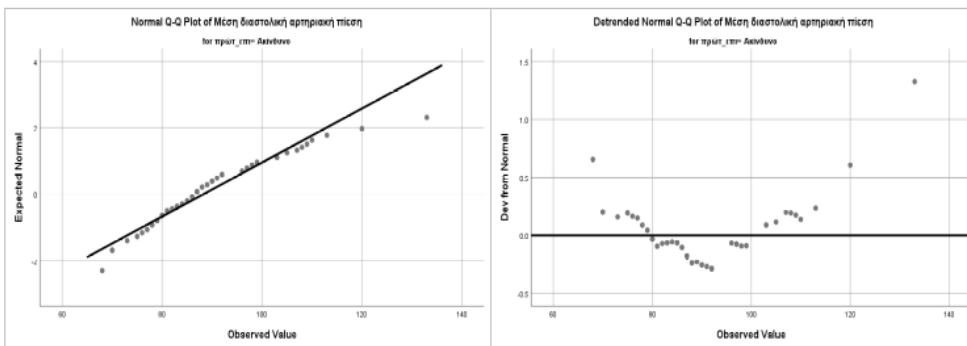
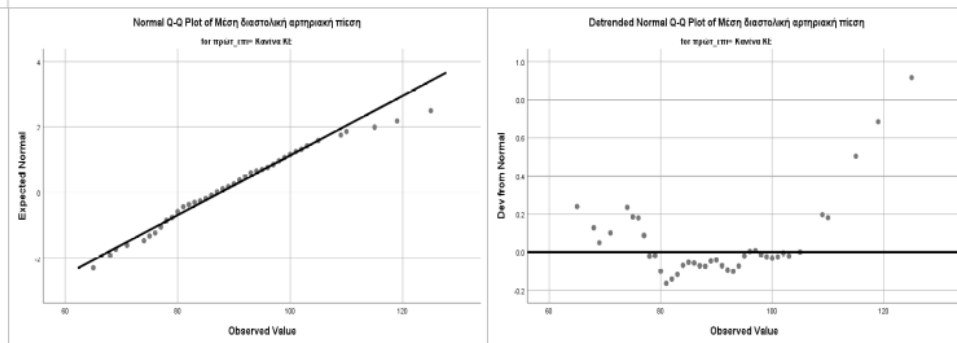
Ο πίνακας Case Processing Summary μας δίνει πληροφορίες για το πόσα από τα cases που υπήρχαν αρχικά στο αρχείο μας, λήφθηκαν υπόψιν, στην μελέτη των κατηγοριών του "Πρώτου καρδιακού επεισοδίου" σε σχέση με την "Μέση διαστολική αρτηριακή πίεση".

Case Processing Summary							
		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
<u>Μέση διαστολική αρτηριακή πίεση</u>	<u>Κανένα ΚΕ</u>	238	99.2%	2	0.8%	240	100.0%
	<u>Ακίνδυνο</u>	144	100.0%	0	0.0%	144	100.0%
	<u>Επικίνδυνο</u>	18	100.0%	0	0.0%	18	100.0%
	<u>Ξαφνικός Θάνατος</u>	72	100.0%	0	0.0%	72	100.0%

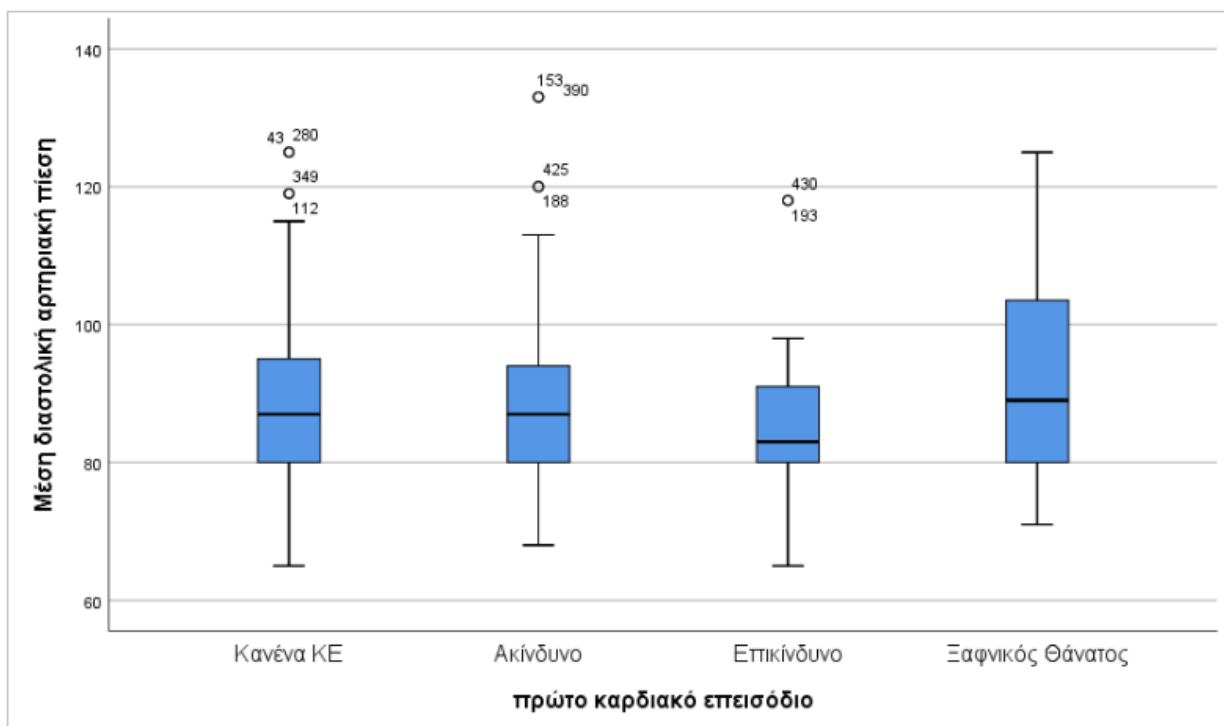
Από τον πίνακα **Tests of Normality** μπορούμε να πάρουμε αποτελέσματα από δύο τεστ κανονικότητας, τα Kolmogorov-Smirnov και Shapiro-Wilk. Παρατηρώντας τον πίνακα βλέπουμε ότι το p-value εκτός της υποομάδας που έχει τον χαρακτηρισμό Επικίνδυνο, είναι μικρότερο του 5%. Δηλαδή, δεν ακολουθούν κανονική κατανομή. Λόγω αυτού, θα χρησιμοποιήσουμε τον μη-παραμετρικό έλεγχο διαφορών ανάμεσα σε περισσότερες από δύο ανεξάρτητες ομάδες, Kruskal-Wallis H.

Tests of Normality							
	πρώτο καρδιακό επεισόδιο	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
<u>Μέση διαστολική αρτηριακή πίεση</u>	<u>Κανένα ΚΕ</u>	.077	238	.002	.976	238	.000
	<u>Ακίνδυνο</u>	.133	144	.000	.937	144	.000
	<u>Επικίνδυνο</u>	.153	18	.200*	.916	18	.111
	<u>Ξαφνικός Θάνατος</u>	.147	72	.001	.944	72	.003

\*. This is a lower bound of the true significance.  
a. Lilliefors Significance Correction







Από τον πίνακα **Test Statistics** προκύπτει ότι το p-value είναι  $>5\%$ . Συνεπώς δεν υπάρχουν σημαντικές διαφορές στη τιμή της Μέσης διαστολικής αρτηριακής πίεσης σε σχέση με το πρώτο καρδιακό επεισόδιο. Άρα, η μέση διαστολική αρτηριακή πίεση φαίνεται να μην επηρεάζεται παρόλο την κατηγορία επικινδυνότητας του πρώτου καρδιακού επεισοδίου.

Ranks			
	πρώτο καρδιακό επεισόδιο	N	Mean Rank
Μέση διαστολική αρτηριακή πίεση	Κανένα ΚΕ	238	231.48
	Ακίνδυνο	144	232.88
	Επικίνδυνο	18	207.83
	Ξαφνικός Θάνατος	72	267.50
	Total	472	

Test Statistics <sup>a,b</sup>	
	Μέση διαστολική αρτηριακή πίεση
Kruskal-Wallis H	4.946
df	3
Asymp. Sig.	.176
a. Kruskal-Wallis Test	
b. Grouping Variable: πρώτο καρδιακό επεισόδιο	

### 3 Λογαριθμική Παλινδρόμηση ( $\Delta$ )

Η λογαριθμική παλινδρόμηση χρησιμοποιείται για την εκτίμηση της πιθανότητας  $p$  της επιτυχίας μιας δίτιμης μεταβλητής για ένα σύνολο τιμών μίας ή περισσότερων ανεξάρτητων μεταβλητών. Για τη δημιουργία του λογαριθμικού μας μοντέλου χρησιμοποιήσαμε ως εξαρτημένη μεταβλητή την Επιβίωση μετά από 10 χρόνια και ως ανεξάρτητες το Πρώτο καρδιακό επεισόδιο, την Ηλικία και τη Μέση διαστολική αρτηριακή πίεση.

Σημειώνεται πως έγινε επανακωδικοποίηση της μεταβλητής Πρώτο καρδιακό επεισόδιο ως εξής:

Αρχικά	Τελικά
1 = "Κανένα ΚΕ"	0 = "Κανένα ΚΕ"
2 = "Ακίνδυνο"	1 = "Ακίνδυνο"
3 = "Επικίνδυνο"	1 = "Επικίνδυνο"
4 = "Ξαφνικός Θάνατος"	1 = "Ξαφνικός Θάνατος"

#### Case Processing Summary

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	472	99.6
	Missing Cases	2	.4
	Total	474	100.0
Unselected Cases		0	.0
Total		474	100.0

a. If weight is in effect, see classification table for the total number of cases.

Από τον πίνακα **Case Processing Summary** παρατηρούμε ότι για την εξαρτημένη μεταβλητή υπάρχουν 2 missing cases.

Στον πίνακα Dependent Variable Encoding φαίνεται η εσωτερική κωδικοποίηση της εξαρτημένης μεταβλητής και στον πίνακα Categorical Variables Codings φαίνεται η κωδικοποίηση της κατηγορικής ανεξάρτητης μεταβλητής.

### Dependent Variable Encoding

Original Value	Internal Value
Πεθαμένος	0
Ζωντανός	1

### Categorical Variables Codings

		Frequency	Parameter coding (1)
Πρώτο καρδιακό επεισόδιο	Κανένα ΚΕ	238	1.000
	Επεισόδιο	234	.000

Ο πίνακας Classification Table περιέχει μόνο το σταθερό όρο. Όλες οι τιμές της εξαρτημένης μεταβλητής κατατάσσονται στην κατηγορία με τη μεγαλύτερη συχνότητα, δηλαδή στην υποκατηγορία “Ζωντανός”.

### Classification Table<sup>a,b</sup>

Observed			Predicted		Percentage Correct
			Επιβίωση μετά από 10 χρόνια Πεθαμένος	Ζωντανός	
Step 0	Επιβίωση μετά από 10 χρόνια	Πεθαμένος	0	116	.0
		Ζωντανός	0	356	100.0
	Overall Percentage				75.4

a. Constant is included in the model.

b. The cut value is .500

Ο πίνακας **Variables in the Equation** μας δίνει πληροφορίες μόνο για το σταθερό όρο.

Ο πίνακας **Variables not in the Equation** μας δίνει πληροφορίες για την αξιολόγηση των ανεξάρτητων μεταβλητών που δεν έχουν εισέλθει ακόμη στο υπόδειγμα. Δηλαδή εξετάζει τη σημαντικότητα κάθε μιας από τις ανεξάρτητες μεταβλητές εάν έμπαινε μόνη της στο υπόδειγμα μαζί με το σταθερό όρο. Το κριτήριο είναι το **Score** το οποίο μας δείχνει ποια μεταβλητή έχει μεγαλύτερη βαρύτητα στην πρόγνωση της εξαρτημένης μεταβλητής. Στη δική μας περίπτωση μεγαλύτερη βαρύτητα φαίνεται να έχει η μεταβλητή Πρώτο καρδιακό επεισόδιο.

### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1.121	.107	110.012	1	.000	3.069

### Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	πρώτο καρδιακό επεισόδιο(1)	156.427	1	.000
		ηλικία πρώτου επεσοδίου	14.334	1	.000
		Μέση διαστολική αρτηριακή πίεση	9.088	1	.003
	Overall Statistics		164.536	3	.000

Ο πίνακας **Omnibus Tests of Model Coefficients** περιλαμβάνει τις τιμές για τα p-value οι οποίες μας δείχνουν το επίπεδο στατιστικής σημαντικότητας. Παρατηρούμε ότι  $p < 0,05$  επομένως οι ανεξάρτητες μεταβλητές συνδυαζόμενες μεταξύ τους συμβάλλουν σημαντικά στην πρόγνωση των τιμών της εξαρτημένης.

### Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	213.876	3	.000
	Block	213.876	3	.000
	Model	213.876	3	.000

Στον πίνακα Model Summary δίνεται η τιμή της συνάρτησης λογαριθμο-πιθανοφάνειας μαζί με το συντελεστή προσδιορισμού Cox & Snell και το συντελεστή Nagelkerke. Από αυτούς τους συντελεστές φαίνεται ότι η μεταβλητότητα της εξαρτημένης μεταβλητής ερμηνεύεται από τις τρεις ανεξάρτητες μεταβλητές του λογαριθμικού υποδείγματος σε ποσοστό 36% – 54%.

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	312.528 <sup>a</sup>	.364	.542

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Ο έλεγχος Hosmer and Lemeshow έχει ως υποθέσεις τις εξής:

- $H_0$  : Το μοντέλο έχει καλή προσαρμογή
- $H_1$  : Το μοντέλο δεν έχει καλή προσαρμογή

Από τον πίνακα παρατηρούμε πως το p – value είναι μεγαλύτερο από 0,05 επομένως το μοντέλο μας προσαρμόζεται ικανοποιητικά στα δεδομένα.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	1.289	8	.996

Ο πίνακας ταξινόμησης Classification Table δείχνει το ποσοστό στο οποίο συμφωνούν οι παρατηρούμενες και οι εκτιμώμενες από το υπόδειγμα τιμές της εξαρτημένης μεταβλητής. Στην περίπτωση μας το ποσοστό αυτό ανέρχεται σε 78,4%.

**Classification Table<sup>a</sup>**

Observed			Predicted		Percentage Correct
			Επιβίωση μετά από 10 χρόνια Πεθαμένος	Ζωντανός	
Step 1	Επιβίωση μετά από 10 χρόνια	Πεθαμένος	68	48	58.6
		Ζωντανός	54	302	84.8
	Overall Percentage				78.4

a. The cut value is .500

Ο πίνακας Variables in the Equation μας δείχνει τους συντελεστές του υποδείγματος μαζί με τους αντίστοιχους επαγωγικούς ελέγχους και τα διαστήματα εμπιστοσύνης τους .

- Η στήλη B αναγράφει τις τιμές των συντελεστών των ανεξάρτητων μεταβλητών που συνδέονται στατιστικά σημαντικά με την εξαρτημένη.
- Η στήλη S.E. δείχνει το τυπικό σφάλμα της εκτίμησης της τιμής του κάθε συντελεστή.
- Το κριτήριο Wald μας δείχνει κατά πόσο είναι σημαντική η επίδραση των ανεξάρτητων μεταβλητών στη διαμόρφωση των τιμών της εξαρτημένης.
- Η στήλη Sig. αποδεικνύει τη στατιστική σημαντικότητα των μεταβλητών που συμμετέχουν στο μοντέλο της παλινδρόμησης.
- Οι δύο τελευταίες στήλες του πίνακα μας δίνουν τα όρια ενός 95% διαστήματος εμπιστοσύνης για τον κάθε ένα συντελεστή των ανεξάρτητων μεταβλητών.

**Variables in the Equation**

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 <sup>a</sup>	πρώτο καρδιακό επεισόδιο(1)	21.119	2570.377	.000	1	.993	1485770991	.000	.
	ηλικία πρώτου επεισοδίου	-.085	.035	5.979	1	.014	.919	.859	.983
	Μέση διαστολική αρτηριακή πίεση	-.024	.010	5.421	1	.020	.976	.957	.996
	Constant	6.244	1.940	10.354	1	.001	514.751		

a. Variable(s) entered on step 1: πρώτο καρδιακό επεισόδιο, ηλικία πρώτου επεισοδίου, Μέση διαστολική αρτηριακή πίεση.

Ολοκληρώνοντας το λογαριθμικό μας μοντέλο καταλήξαμε στα εξής συμπεράσματα:

- Πριν εισαχθούν οι ανεξάρτητες μεταβλητές στο λογαριθμικό μοντέλο η μεταβλητή με τη μεγαλύτερη βαρύτητα στην πρόγνωση της εξαρτημένης μεταβλητής είναι το Πρώτο καρδιακό επεισόδιο ( $\text{Score} = 156,427$ ) . Ύστερα όμως από την εισαγωγή τους στο μοντέλο , παρατηρούμε ότι το πρώτο καρδιακό επεισόδιο δεν επιδρά σημαντικά στη διαμόρφωση τιμών της εξαρτημένης μεταβλητής ( $\text{Wald} = 0,000$ ) και αυτό φαίνεται επίσης από το γεγονός ότι το  $p - \text{value}$  είναι μεγαλύτερο από 0,05 .
- Παρ' όλα αυτά επιλέξαμε μία από τις ανεξάρτητες μεταβλητές μας να είναι το πρώτο καρδιακό επεισόδιο , διότι κάνοντας επιπλέον συνδυασμούς απουσίας της συγκεκριμένης ανεξάρτητης μεταβλητής , το λογαριθμικό μοντέλο δεν προσαρμόζοταν καθόλου ικανοποιητικά στα δεδομένα μας .
- Με βάση τον τελευταίο πίνακα το υπόδειγμα έχει τη μορφή :

(Πιθανότητα επιβίωσης μετά από 10 χρόνια) =  $6,244 + 21,119$  (πρώτο καρδιακό) –  $0,024$  (δαπ) –  $0,085$  (ηλικία πρώτου επεισοδίου)