ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ

ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

- ΈΡΕΥΝΑ ΓΙΑ ΤΟΥΣ ΠΟΛΙΤΕΣ ΤΟΥ ILLINOIS -

(ΕΡΓΑΣΙΑ 5)

Διδάσκων : Ι.Ντζουφράς- Ξ. Πεντελή

Φοιτητής:

Όνοματεπώνυμο: Τσέτσι Ιωάννα

Αριθμό Μητρώου: 6160095

Έτος σπουδών: 4ο

29 Μαίου,2020

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή - Περιγραφή μελέτης προβλήματος .

Στη πολιτεία Ilinois των ΗΠΑ πραγματοποιήθηκε μια ερεύνα με στόχο την εκτίμηση του πλούτου του εργατικού δυναμικού .Οι ερευνητές κατάφεραν να συγκεντρώσουν ένα δείγμα από το εργατικό πληθυσμό ίσο με 1000 ατόμων από το οποίο και συγκέντρωσαν πληροφορίες σχετικά με την οικογενειακή κατάσταση ,το μέγιστο ολοκληρωμένο εκτός εκπαίδευσης ,το φύλο το ζώδιο ,την κατάσταση υγείας την ηλικία και τον δείκτη πλούτου.

Στόχος της παρούσας μελέτης είναι να εξετάσει κατά πόσο το μέγιστο ολοκληρωμένο έτος εκπαίδευσης διαφέρει ανάμεσα σε άτομα διαφορετικού φύλου και άτομά διαφορετικού ζωδίου. Ακόμα ένας στόχος αποτελεί να εκτιμήσουμε τον δείκτη πλούτο σε σχέση με τις υπόλοιπες μεταβλητές. Από το σετ δεδομένων που δόθηκε οι μεταβλητές τροποποιήθηκαν ,για λογους που εξηγουνταί στο επομένο κεφάλιαο και έχουν την παρακάτω μορφή :

Πίνακάς 1 Πίνακας δεδομένων

μεταβλητή	**σημασία**	**τύπο μεταβλήτης**	**τιμές**
id		Αριθμητική	
marital	Οικογενειακή κατάσταση	Κατηγορική	MARRIED , WIDOWED DIVORCED, SEPARATED NEVER MARRIED
age	Ηλικία	Αριθμητική	
educ	Μέγιστο ολοκληρωμένο έτος εκπαίδευσης	Αριθμητική	
sex	Φύλο	Κατηγορική	MALE , FEMALE
health	Κατάσταση υγείας	Κατηγορική	EXCELLENT,GOOD ,FAIR ,POOR
wealth	Δείκτης που δείχνει το επίπεδο πλούτου κάθε ερωτώμενου	Αριθμητική	
zodiac	Ζώδιο	Κατηγορική	ARIES, TAURUS ,GEMINI , CANCER, LEO, VIRGO, LIBRA, SCORPIO SAGITTARIUS, CAPRICORN AQUARIUS ,PISCES

Κεφάλαιο 2

Περιγραφική ανάλυση.

Η ανάλυση θα γίνει με τη βοήθεια του στατιστικού πακέτου R. Αρχικά εισάγουμε τα δεδομένα στην R και τις βιβλιοθήκες οι οποίες είναι απαραίτητες για την ανάλυση. Παρατηρούμε ότι στα δεδομένα υπάρχουν κάποιές τιμές οι οποίες δεν έχουν καταχωρηθεί στα δεδομένα, ελλείπουσες τιμές (missing values, NA) .Επειδή δεν γνωρίζουμε τον ακριβή λόγο για τον οποίο δεν διαθέτουμε αυτήν την πληροφορία και ακόμα δεν έχουμε την δυνατότητα να την αναζητήσουμε την πληροφορία αυτήν δεν θα τις λάβουμε υπόψιν για την περιγραφική ανάλυση. Τέλος μετατρέπουμε κατάλληλα τις μεταβλητές wealth ,age και educ σε αριθμητικές.

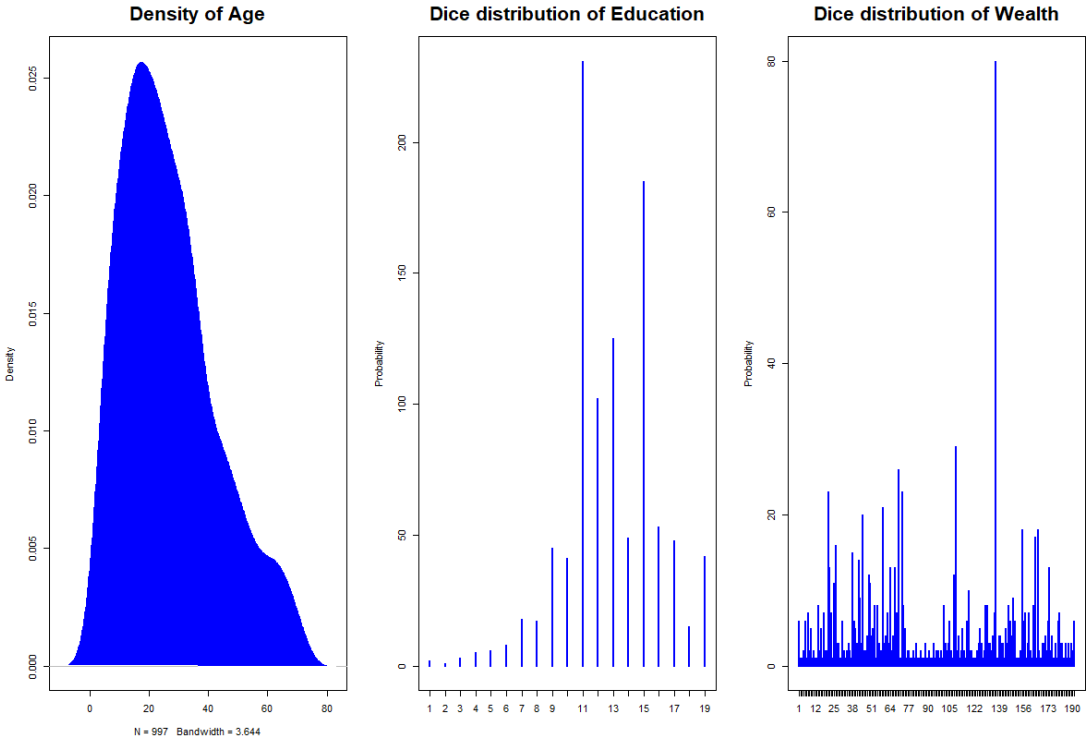
Στην συνέχεια θα περιγράψουμε κάθε μεταβλητή ξεχωριστά βασισμένοι στον εάν είναι ποιοτική ή ποσοτική .Ειδικότερα για τις ποσοτικές μεταβλητές ,wealth,age και educ τα κατάλληλα περιγραφικά είναι η μέση τιμή, η διάμεσος, τυπική απόκλιση, ασσυμετρία και η κύρτωση. Λαμβάνοντας τα αποτελέσματα παρατηρούμε ότι η μεταβλητή age με ασυμμετρία 0.72 και κύρτωση -0.09 φαίνεται να έχει δεξιά ασυμμετρία και να είναι πλατύκυρτή και επομένως αποκλίνει από την κανονικότητα.[1] Ακόμα βλέπουμε ότι και οι μεταβλητές educ και wealth αποκλίνουν από την κανονικότητα για αυτό και θα προχωρήσουμε σε περαιτέρω έλεγχο κανονικότητας με Shapiro – Wilk καθώς και τα αντίστοιχα QQ plots.[2] Τέλος βλέπουμε ότι καμία μεταβλητή δεν ακολουθεί την κανονική κατανομή (age: S-W p-value < 2.2e-16, educ:S-W p-value = 6.358e-15 , wealth: S-W p-value < 2.2e-16).

Στην συνεχεία θα ήταν χρήσιμο να δούμε και διαγραμματικά τα παραπάνω αποτελέσματα (Διάγραμμα 1). Από το γράφημα βλέπουμε ότι η μεταβλητή age έχει πολλές μικρές τιμές και λίγες υψηλές τιμές (θετική ασυμμετρία όπως είδαμε και παραπάνω).Οι μεταβλητές educ και wealth είναι διακριτές (παίρνουν τιμές στο αριθμήσιμο σύνολό) .Όπως βλέπουμε η μεταβλητή educ έχει πολλές υψηλές τιμές και λίγες χαμηλές(αρνητική ασυμμετρία) ενώ οι τιμές της μεταβλήτης wealth τείνουν να συσταδοποιούνται στις άκρες του άξονα x΄x . Τέλος για τις κατηγορικές μεταβλητές θα εκλέξουμε την συχνότητα με την οποία εμφανίζεται κάθε κατηγορία της ποιοτικής μεταβλήτης, ωστόσο αυτό δεν μας δίνει αρκετή πληροφορία και ποιο πολύ μας ενδιαφέρει η σχέση μεταξύ τους όπως και θα δούμε στο επόμενο κεφάλαιο .

[1] Παράρτημα Πίνακας 3

[2] Παράρτημα Διάγραμμα 7

Διάγραμμα 2 Distribution Plots



Κεφάλαιο 3

Σχέσεις μεταβλητών ανά δύο.

Έχοντας εξετάσει κάθε μεταβλητή ξεχωριστά θα πρέπει να μελετήσουμε και την μεταξύ τους σχέση. Αρχικά θα εξετάσουμε την συσχέτιση ανάμεσα στις ποσοτικές χρησιμοποιώντας τον συντελεστή συσχέτιση Kendal. Από οπού και συμπεραίνουμε ότι δεν έχουμε ισχυρές συσχετίσεις. Ακόμα για τις μεταβλητές age educ και wealth φτιάχνουμε μονό boxplots ανά κατηγορία των ποιοτικών μεταβλητών, διότι καμία δεν ακολουθεί την κανονική κατανομή. Οι σχέσεις των ποιοτικών μεταβλητών που έχουν νόημα να εξεταστούν είναι :

- Οικογενειακή κατάσταση ~φύλο (Marital~Sex)
- Οικογενειακή κατάσταση ~ Κατάσταση υγείας (Marital~ Health)
- Οικογενειακή κατάσταση ~ζώδιο(Marital~zodiac)
- Φύλο ~κατάσταση υγείας (Sex~health)
- Φύλο ~ ζώδιο (Sex ~zodiac)
- Κατάσταση υγείας ~ζώδιο (health~zodiac)

Για τους παραπάνω ελέγχους θα χρησιμοποιήσουμε ραβδόγραμμα, ραβδογράμματα ανά κατηγορίες των μεταβλητών και ελέγχους X^2 ανεξαρτησίας .Έχοντας ελέγξει τα παραπάνω γραφήματα και έχοντας κάνει τους παραπάνω ελέγχους ανεξαρτησίας καταλήγουμε ότι σημαντικές σχέσεις που παρατηρούνται είναι :

- ✓ Οικογενειακή κατάσταση ~φύλο (Marital~Sex)
- ✓ Οικογενειακή κατάσταση ~ κατάσταση υγείας (Marital~ Health)

Κεφάλαιο 4

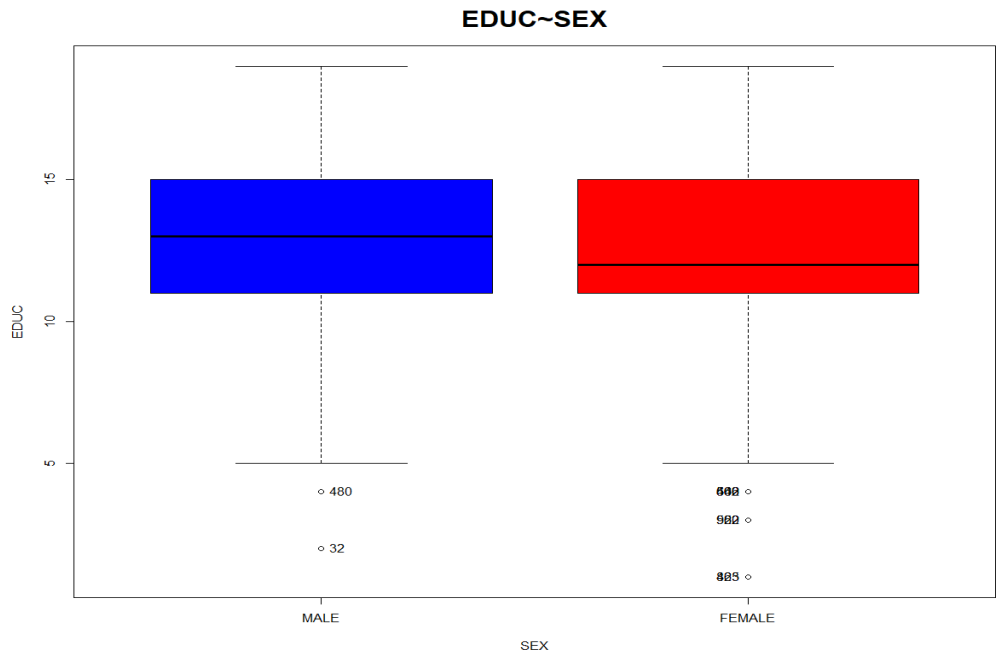
Προβλεπτικά ή ερμηνευτικά μοντέλα.

Έχοντας εξετάσει τις σχέσεις των μεταβλητών ανά δύο και έχοντας ξεχωρίσει αυτές που είναι ισχυρές θα μπορούσαμε να κατασκευάσουμε ένα μοντέλο για να εξετάσουμε πως σχετίζεται ο πλούτος των κατοίκων της πολιτεία Ilinois των ΗΠΑ σε σχέσεις με τις υπόλοιπες μεταβλητές που έχουμε.

Αρχικά θα πραγματοποιήσουμε κάποιους ελέγχους υποθέσεων για να έχουμε κάποια πλησιέστερη εικόνα για τα δεδομένα μας. Ένας πρώτος έλεγχος που μας ενδιαφέρει είναι εάν το μέγιστο ολοκληρωμένο έτος εκπαίδευσης διαφέρει ανάμεσα σε άτομα διαφορετικού φύλου. Για να δούμε ένα υπάρχει κάποια διαφοροποίηση ανάμεσα στις δύο ομάδες θα πρέπει να ελέγξουμε την κανονικότητα του μέγιστου έτους εκπαίδευσης σε κάθε κατηγορία της μεταβλήτης sex. Εκτελώντας τον έλεγχο Shapiro – Wilk για κάθε κατηγορία της μεταβλήτης sex έχουμε ισχυρές στατιστικές ενδείξεις για να απορρίψουμε την κανονικότητα για τους άντρες (S-W p-value = 1.484e-09) και για τις γυναίκες (S-W p-value = 6.009e-10). Έπειτα λόγω ότι έχουμε μεγάλο δείγμα θα εξετάσουμε αν ο μέσος είναι κατάλληλο περιγραφικό μετρό για τις κάθε κατηγορίες . Από τα αποτελέσματα που μας που παίρνουμε από την R () προκύπτει ότι δεν είναι κατάλληλο περιγραφικό μετρό για αυτό και προχωράμε στον μη παραμετρικό έλεγχο Kruskal-Wallis Test για να ελέγξουμε την ισότητα των διαμέσων.

Τελικά από την ανάλυση της ισότητας των διαμέσων συμπεραίνουμε ότι υπάρχει σημαντική στατιστική διαφορά (K-W p-value = 0.0005111) ,δηλαδή το μέγιστο ολοκληρωμένο έτος εκπαίδευσης διαφέρει ανάμεσα σε άνδρες και γυναίκες ,αυτό επιβεβαιώνεται και από το παρακάτω διάγραμμα πλαισίου απολήξεων (boxplot).

Διάγραμμα 3 Boxplot for Educ ~Sex.



Ένας έλεγχός ο οποίος είναι ακόμα ενδιαφέρον να εξετάσουμε είναι εάν διαφοροποιείται το μέγιστο ολοκληρωμένο έτος εκπαίδευσης ανάμεσα σε άτομα διαφορετικού ζωδίου. Έχοντας προσαρμόσει το κατάλληλο μοντέλο με την βοήθεια της R πρώτα θα εξετάσουμε εάν τα κατάλοιπα του μοντέλου ακολουθούν την κανονική κατανομή για να προβούμε και στον αντίστοιχό έλεγχο. Έχοντας υλοποιήσει με τον έλεγχο Shapiro -Wilk την κανονικότητα των καταλοίπων συμπεραίνουμε ότι υπάρχουν ισχυρές στατιστικές ενδείξεις για να απορρίψουμε την κανονικότητα (S-W p -value = 2.534e-10). Έπειτα όπως και προηγμένος λογού μεγάλου δείγματος θα πρέπει να εξετάσουμε εάν ο μέσος αποτελεί κατάλληλο περιγραφικό μετρό κεντρικής τάσης, οπού και παρατηρούμε ότι σε κανένα επίπεδο της μεταβλήτης zodiac η συμμετρία είναι είτε θετική είτε αρνητική αλλά όχι μηδενική(δηλαδή συμμετρική) και αντίστοιχα για την κύρτωση βλέπουμε ότι καμία κατανομή δεν είναι μεσοκυρτη (δηλαδή δεν πλησιάζει την κανονική κατανομή).Επομένως θα κάνουμε μη παραμετρικό έλεγχο για την ισότητα των διαμέσων του μέγιστου ολοκληρωμένου έτους εκπαίδευσης για κάθε ζώδιο. Από το τελικό αποτέλεσμα του ελέγχου Kruskal-Wallis Test συμπεραίνουμε ότι δεν είναι στατιστικά σημαντική η διαφορά των διαμέσων (K-W p-value = 0.4112). Επομένως αυτό που συμπεραίνουμε είναι ότι το ζώδιο δεν επηρεάζει στατιστικά σημαντικά το μέγιστο ολοκληρωμένο έτος εκπαίδευσης .

Επιπλέον μας ενδιαφέρει να προχωρήσουμε στην προσαρμογή ενός μοντέλου ώστε να δούμε την επιρροή των μεταβλητών στην μεταβλητή wealth. Στην ανάλυση δεν θα χρησιμοποιηθεί η μεταβλητή με τους κωδικούς της κάθε ερωτώμενου που συμμετείχε στην ερευνά εφόσον δεν προσθέτει κάποια πληροφορία στο μοντέλο . Αρχικά παίρνουμε το πλήρες μοντέλο για τον δείκτη πλούτου που είναι:

Wealth=bo+b1*(Οικογενειακή κατάσταση)+b2*(Ηλικία)+ b3*(μέγιστο ολοκληρωμένο έτος εκπαίδευσης)+b4*(Φύλο)+b5*(κατάσταση υγείας) +b6*(Ζώδιο).

Από το μοντέλο που έχουμε ορίσει στην R έχουν αφαιρεθεί οι ελλείπεις τιμές (missing values). Ελέγχοντας της προϋποθέσεις απορρίπτουμε την κανονικότητα σφαλμάτων (S-W p-value = 0.002881)και την ομοσκεδαστηκότητα (Levene Test p value =8.739e-05) . Επομένως δεν μπορούμε να προβούμε στην προσαρμογή του μοντέλου που περιλαμβάνει όλες τις μεταβλητές και για αυτό τον λόγω θα προχωρήσουμε στην επιλογή μεταβλητών με την μέθοδο Stepwise selection με βάση το κριτήριο επιλογής BIC. Από τον έλεγχο προκύπτει ότι το μοντέλο που επεξηγεί καλύτερα την μεταβλητή wealth είναι το

Wealth=bo+b1*(Ηλικία) +b2*(μέγιστό ολοκληρωμένο εκτός εκπαίδευσης).

Όπως και παραπάνω εξετάζουμε εάν ισχύουν οι έλεγχοι υποθέσεων. Από τα αποτελέσματα συμπεραίνουμε ότι εξασφαλίσουμε μόνο την ανεξαρτησία κατάλοιπων (dwt pvalue=0.6)την γραμμικότητα και η ομοσκεδαστικότητα (leveneTest pvalue= 0.03791 >0.01) .Παρατηρούμε ότι την κανονικότητα των καταλοίπων δεν μπορέσουμε να την εξασφαλίσουμε(S-W pavalue= 0.000572) για αυτό και δεν μπορούμε να είμαστε απολυτά σίγουροι για το ποσό αξιόπιστα θε είναι τα συμπεράσματα με βάση το μοντέλο που εξετάσαμε.

Αρχικά θα πρέπει να κεντροποιήσουμε τις ποσοτικές μας μεταβλητές , δηλαδή θα αφαιρέσουμε από αυτές τους μέσους τους για να μας βοηθήσει στην ερμηνεία της σταθερά .Το μόνο που αλλάζει κεντροποιώντας τις είναι οι συντελεστές στο μοντέλο μας καθώς και η ερμηνεία των μεταβλητών Το τελικό μοντέλο είναι :

Wealth=bo+b1*(educ_cen)+b2*(age_cen) οπού

age_cen=age- mean(age) και educ_cen=educ-mean(educ)

από το οποίο και προκύπτει ότι

Πίνακας 1 Μοντέλο πολλαπλής παλίνδρομης μετά από Stepwise Selection.

	ESTIMATE	***STD. ERROR***	***PR(>|T|)***
(INTERCEPT)	94.1137	60.801	<2e-16
EDUC_CEN	11.2743	21.038	<2e-16
AGE_CEN	0.4010	0.0970	3.96e-05
	Observed :774	**AIC:**	5842.36
	Adjusted R-squared:	0.3644	

Βλέπω ότι η σταθερά bo ισούται με 94.1137.Αυτο σημαίνει ότι ο αναμενόμενος δείκτης πλούτου είναι ίσος με το δειγματικό μέσο του μέγιστου ολοκληρωμένου έτος εκπαίδευσης και τη δειγματική μέση ηλικία όλων των συμμετεχόντων. Τότε ο δείκτης πλούτου θα είναι ίσος με 94.1137.

Για την παράμετρο b1 εάν αυξηθεί το αναμενόμενο μέγιστο ολοκληρωμένο έτος εκπαίδευσης ενώ η ηλικία είναι ίση με την δειγματική μέση ηλικία των συμμετεχόντων τότε ο αναμενόμενος διετής πλούτου θα αυξηθεί κατά 11.2743. Αντίστοιχα για την παράμετρο b2 συμπεραίνουμε ότι εάν αυξηθεί η αναμενόμενη ηλικία των συμμετεχόντων και το μέγιστο ολοκληρωμένο έτος εκπαίδευσης είναι ίσό με το αναμενόμενο μέγιστο ολοκληρωμένο έτος εκπαίδευσης τότε ο δείκτης πλούτου θα αυξηθεί κατά 0.4010. Τέλος βλέπουμε ότι το ποσοστό μεταβλητότητάς που

εξηγείται από το μοντέλο είναι χαμηλό (Radj=0.36) είναι κάτι το οποίο περιμέναμε δοθέντος ότι λόγω τον ελλιπών τιμών που έχουν αφαιρεθεί δεδομένα και από το γεγονός ότι δεν ικανοποιείται η προϋπόθεση της κανονικότητας.

Κεφάλαιο 5

Συμπεράσματά και συζήτηση .

Η παραπάνω μελέτη είχε ως στόχο τη μελέτη του τρόπου επίδρασής των κατηγοριών του φύλου και του ζωδίου στο μέγιστο ολοκληρωμένο έτος εκπαίδευσης αλλά και τη δυνατότητα ερμηνείας του δείκτη πλούτου.

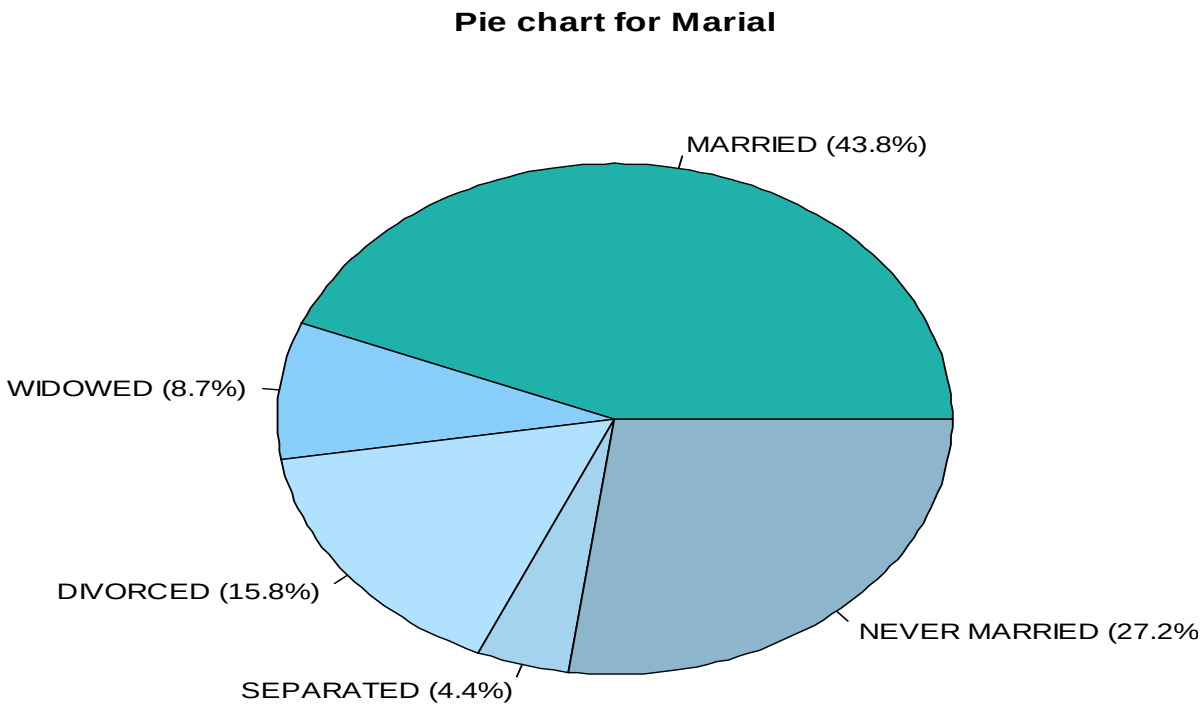
Με βάση την παραπάνω ανάλυση συμπεραίνουμε ότι το μέγιστό ολοκληρωμένο έτος εκπαίδευσης δεν διαφοροποιείται ανάλογα με το ζώδιο του συμμετέχοντος αλλά υπάρχει διαφοροποίηση ανάμεσα σε άνδρες και γυναίκες. Ωστόσο δεν καταφέραμε να εφαρμόσουμε ένα μοντέλο που να μπορεί να ερμηνεύει τον δείκτη του πλούτου και να ισχύουν όλες οι προϋποθέσεις του μοντέλου. Το τελικό μοντέλο που παρουσιάσαμε παρόλο την έλλειψή κανονικότητας καταλοίπων είναι αρκετά εύκολό στην ερμηνεία και έχει μια καλή προσαρμογή (Radj=0.36).

ΠΑΡΑΡΤΗΜΑ

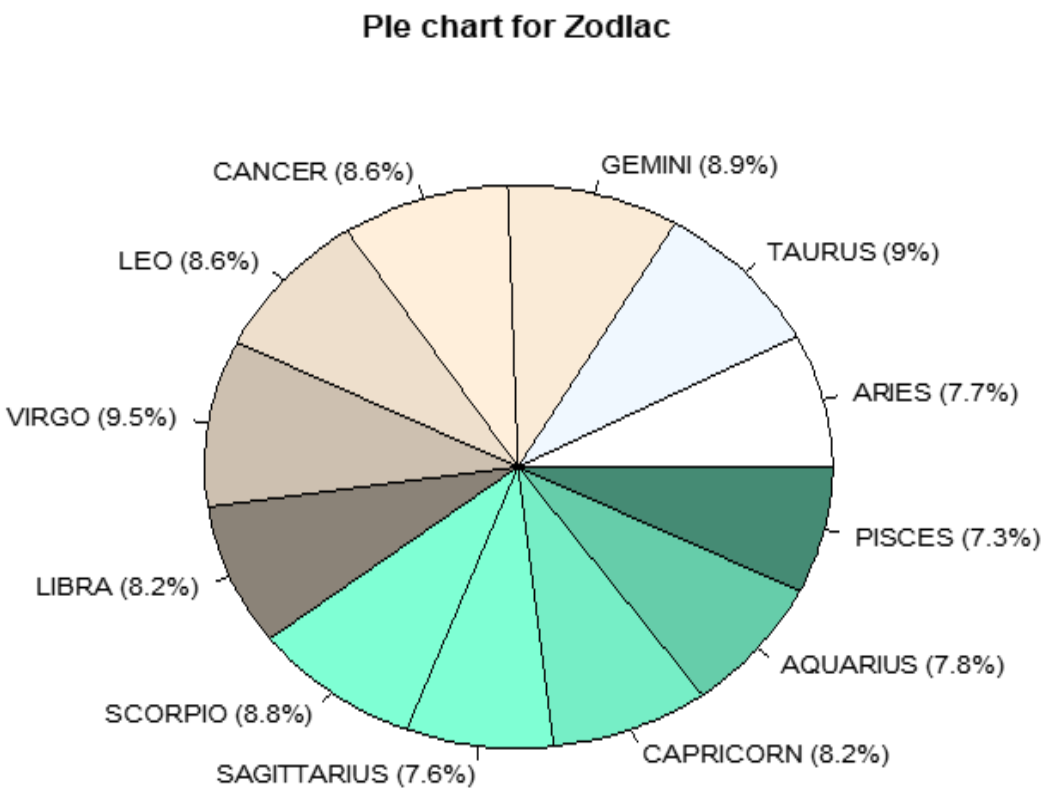
Πίνακας 3πίνακας περιγραφικών μέτρων για τις ποσοτικές μεταβλητές.

ΜΕΤΑΒΛΗΤΗ	ΕΛΛΙΠΗΣ ΤΙΜΕΣ	ΜΕΣΟΣ	ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ	ΔΙΑΜΕΣΟΣ	ΜΙΚΡΟΤΕΡΗ ΤΙΜΗ	ΜΕΓΑΛΥΤΕΡΗ ΤΙΜΗ	ΑΣΥΜΜΕΤΡΙΑ	ΚΥΡΤΩΣΗ
AGE	3	26.69	16.11	24	1	72	0.72	-0.09
EDUC	4	12.82	3	13	1	19	-0.28	0.61
WEALTH	43	94.61	53.63	96	1	191	0.03	-1.32

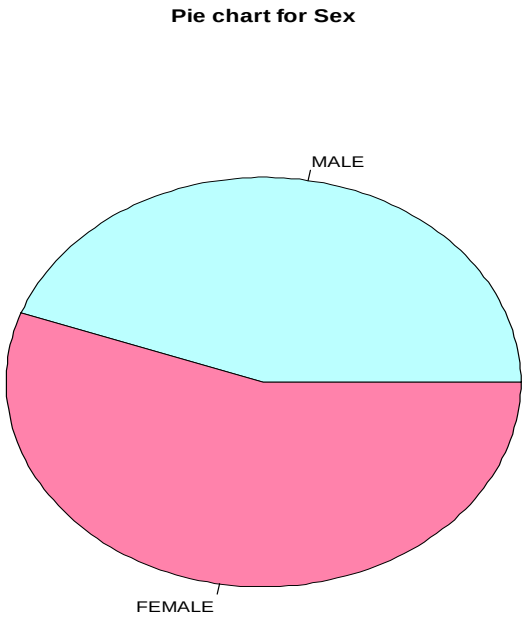
Διάγραμμα3Διάγραμμα πίτας για την μεταβλητή Marital(οικογενειακή κατάσταση)



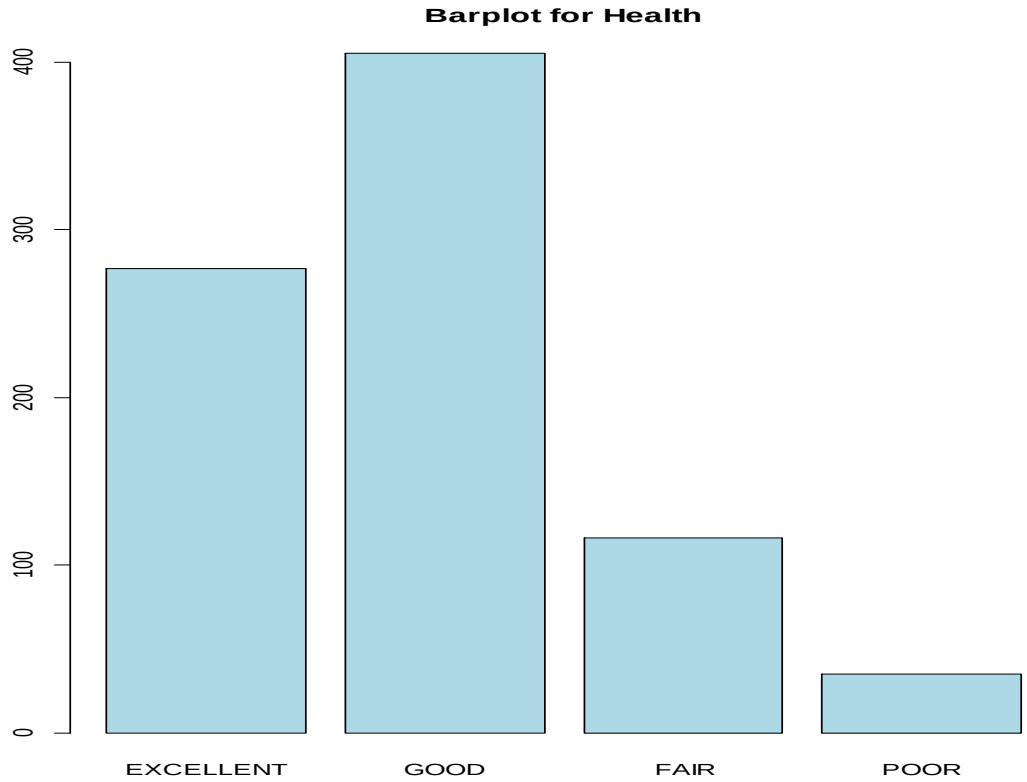
Διάγραμμα 4 Διάγραμμα πίτας για το ζώδιο των εργαζόμενων



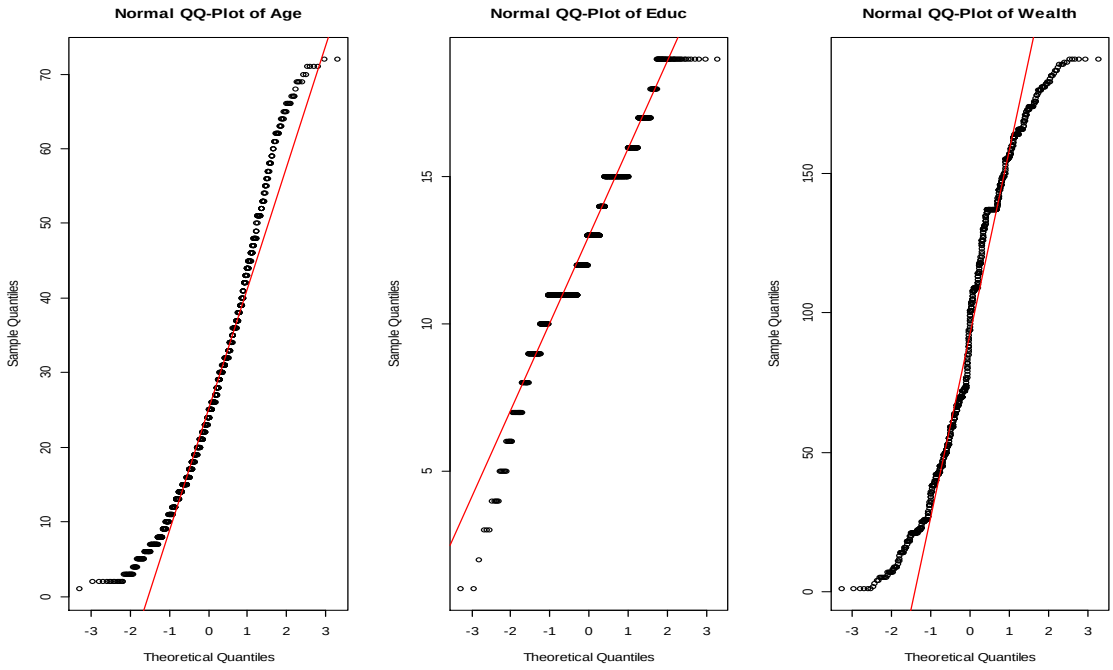
Διάγραμμα5 Διάγραμμα πίτας για την μεταβλητή Sex



Διάγραμμα 6 Ραβδόγραμμα για την μεταβλητή health

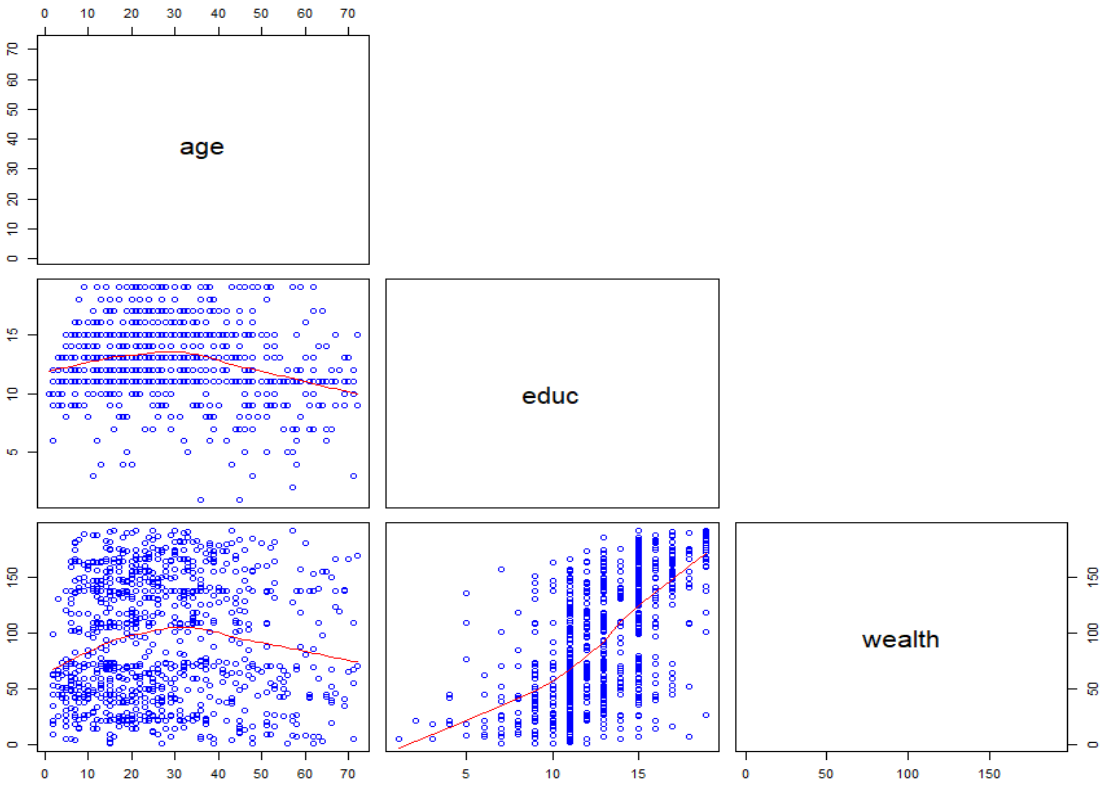


Normal QQ-Plots.

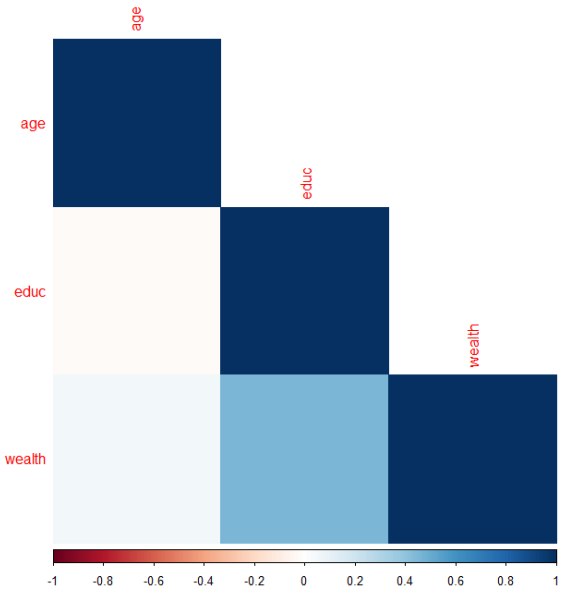


Διάγραμμά 7 QQ plots διαγράμματα για τις ποσοτικές μεταβλητές οπού και βλέπουμε απόκλιση από την κανονικότητα.

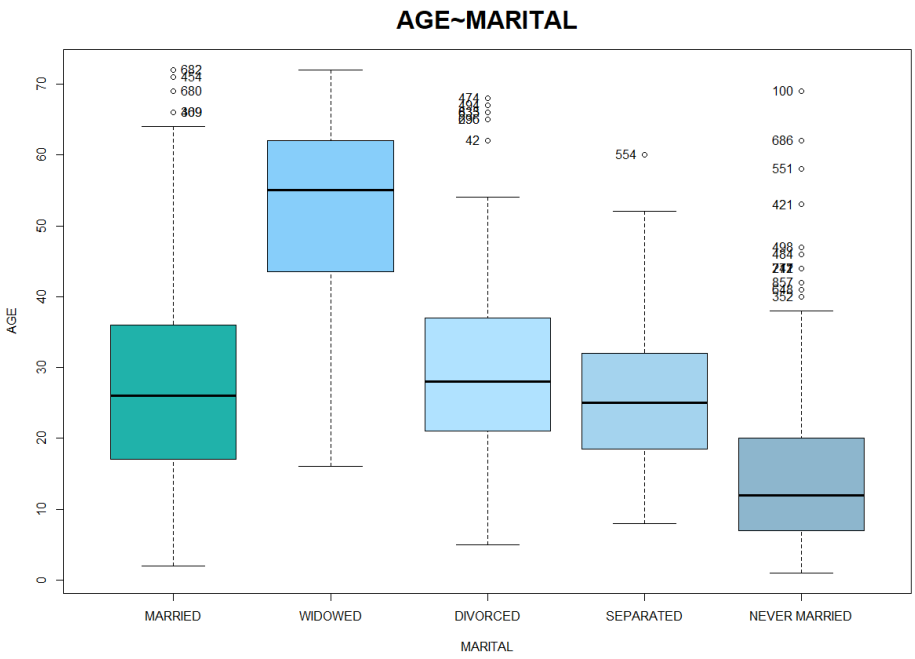
Διάγραμμα 8 Scatterplot ανά δύο για τις ποσοτικές μεταβλητές



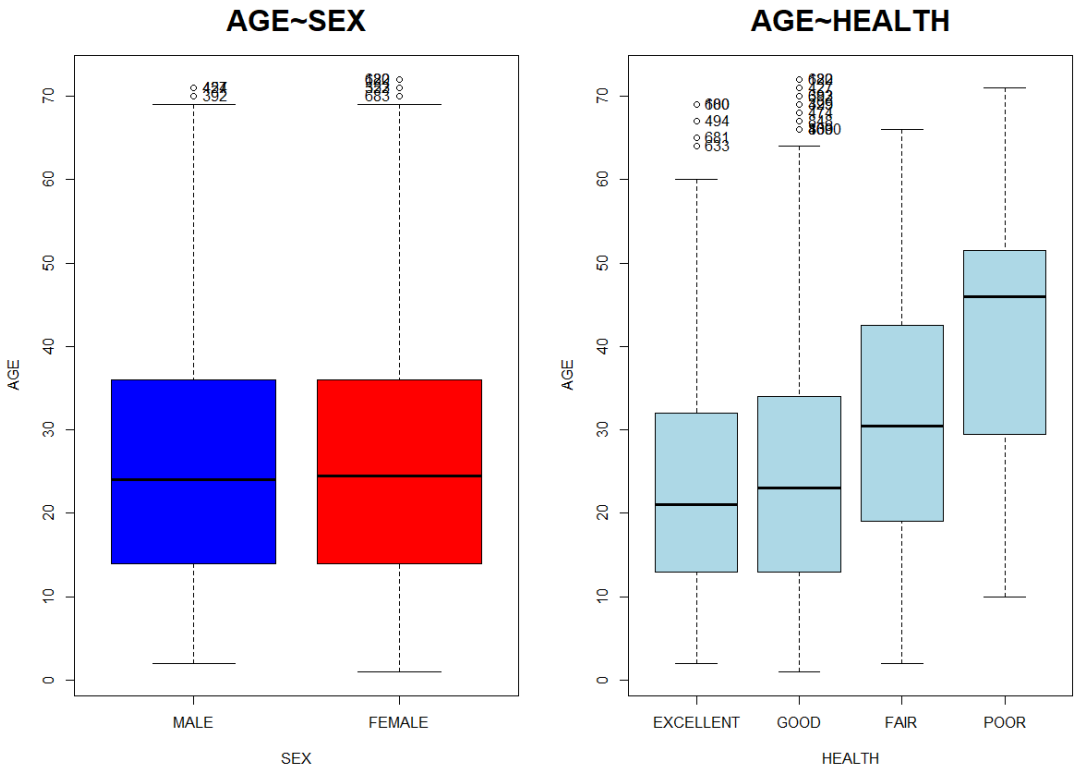
Διάγραμμα 8 Διάγραμμα συσχέτισης (Method=Kendal)



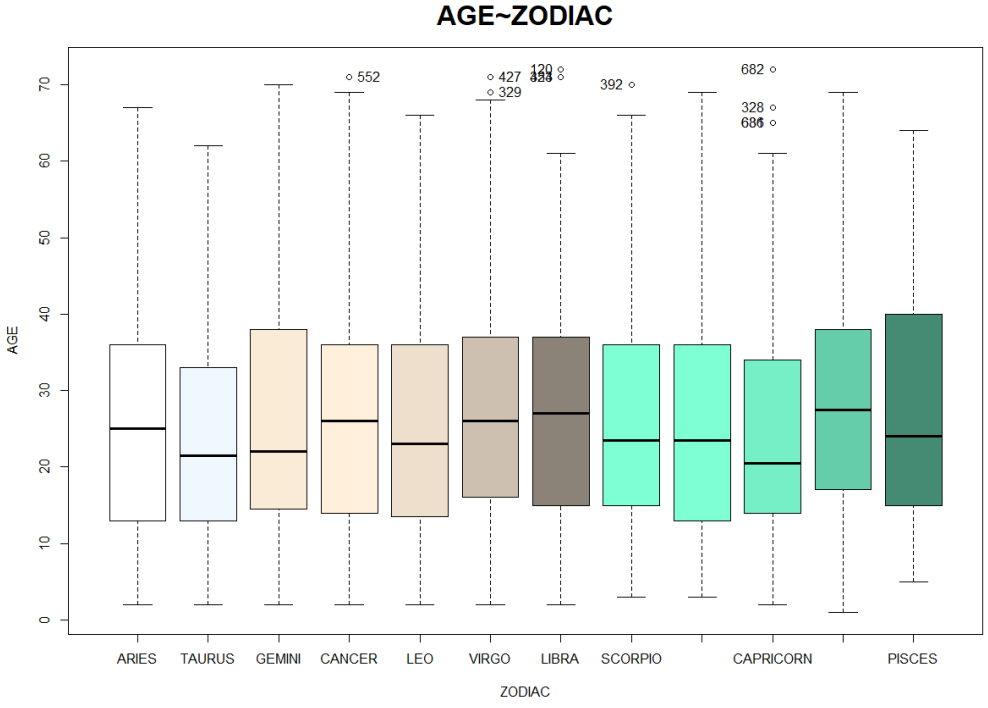
Διάγραμμα 9 AGE ~Marital



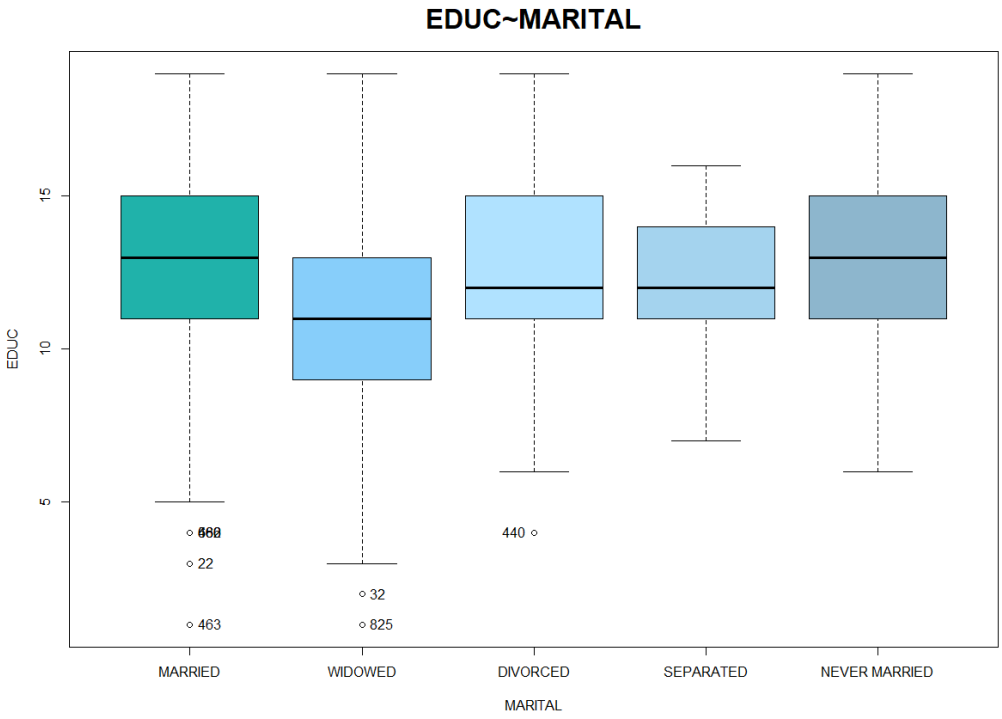
Διάγραμμα 10 .Boxplot για την age και κάθε κατηγορία της sex και health



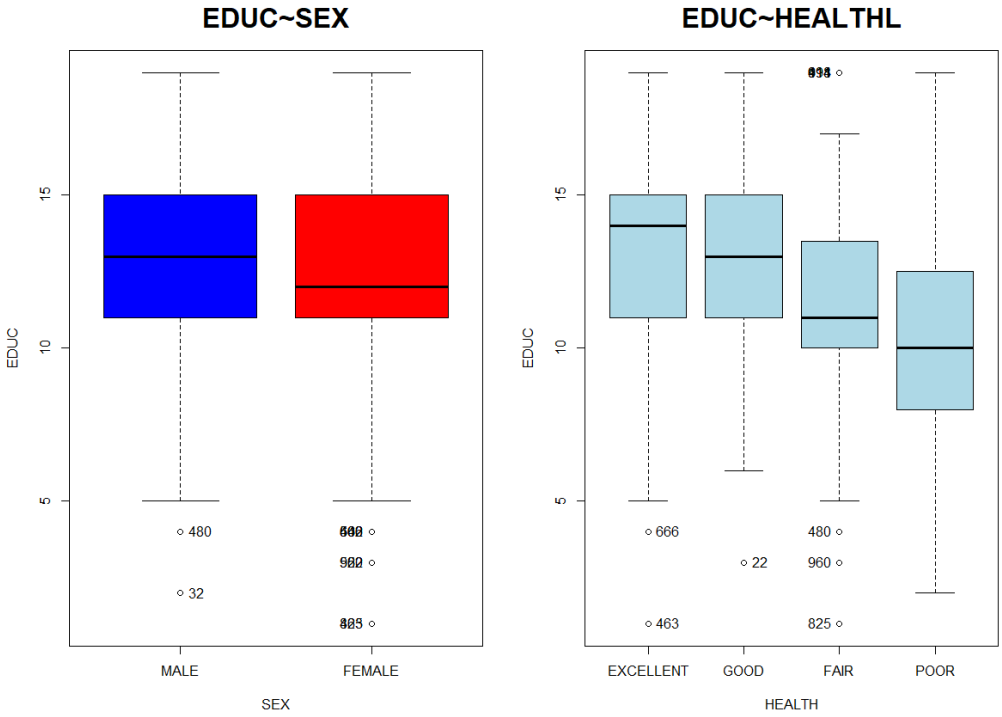
Διάγραμμα 11 Boxplot της age για κάθε κατηγορία της zodiac



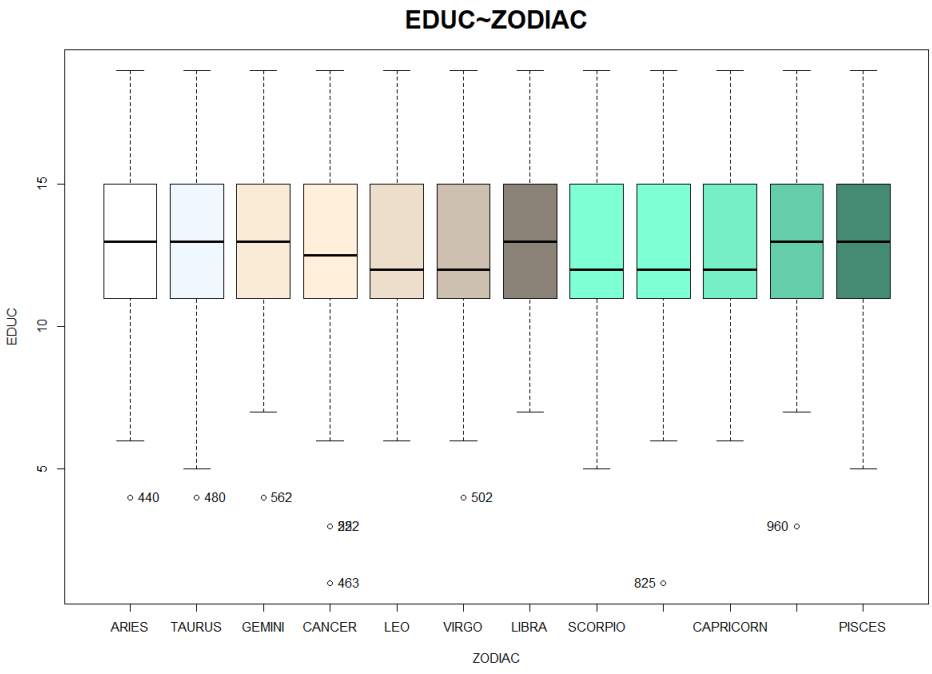
Διάγραμμα 12 boxplot για την educ και κάθε κατηγορία της marital



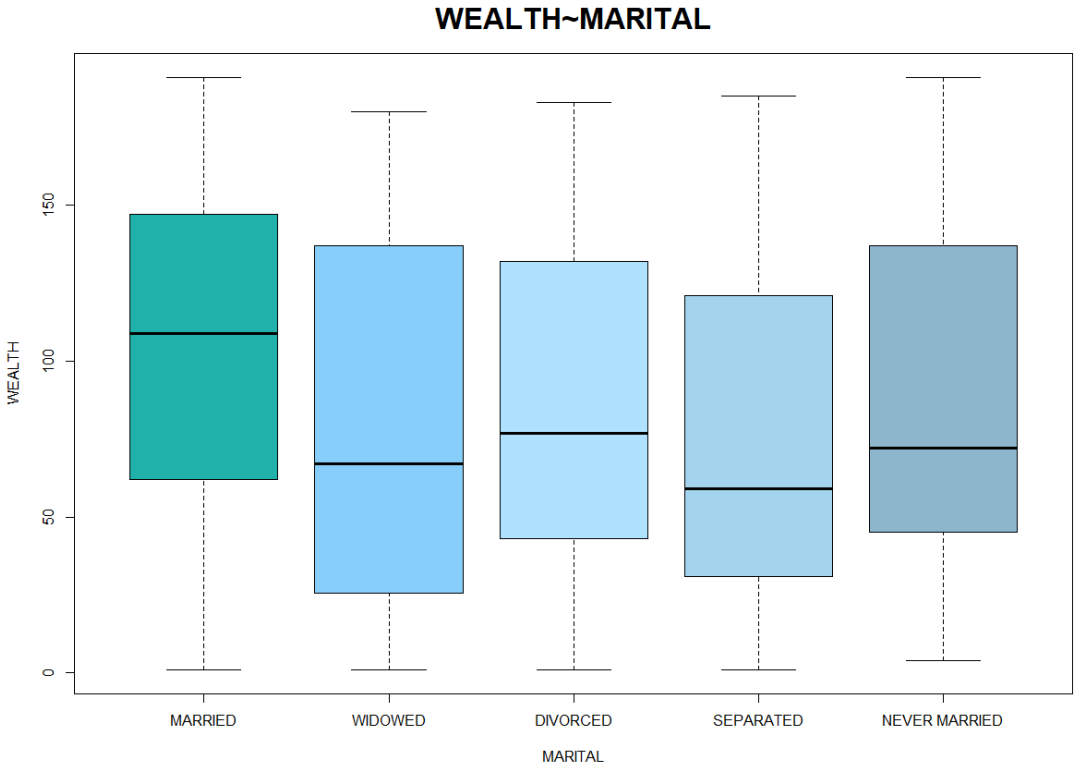
Διάγραμμα 13 Boxplot για την educ και κάθε κατηγορία της sex και health

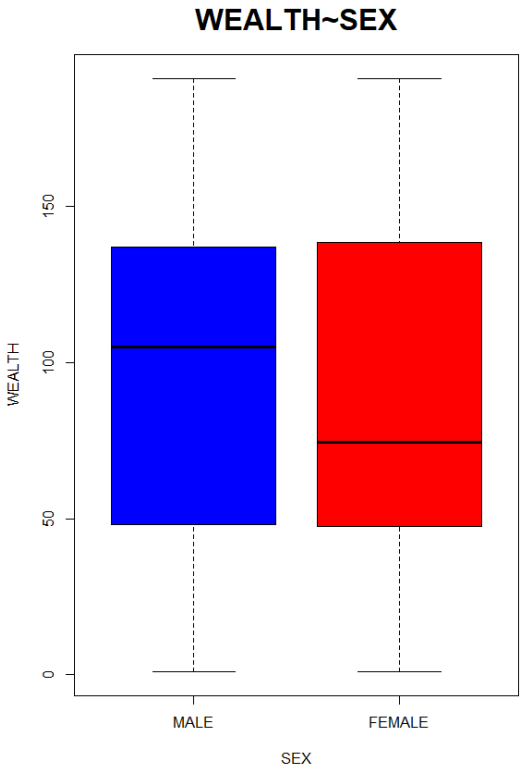


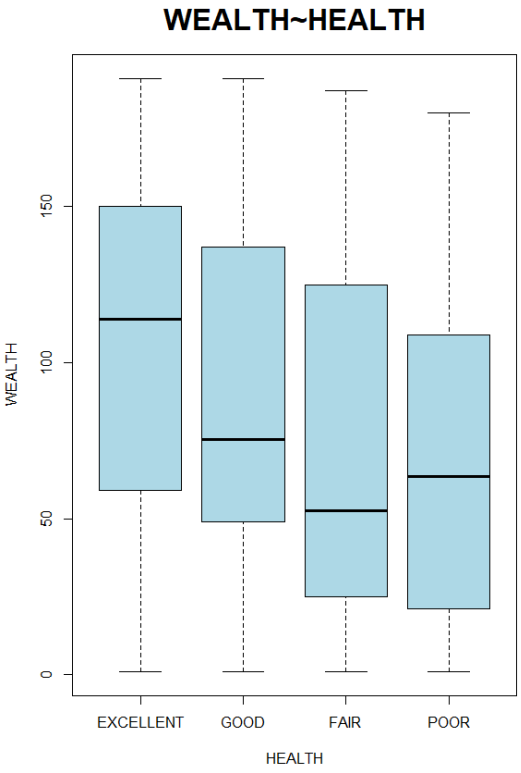
Διάγραμμα 14 Boxplot για την educ και κάθε κατηγορία της zodiac



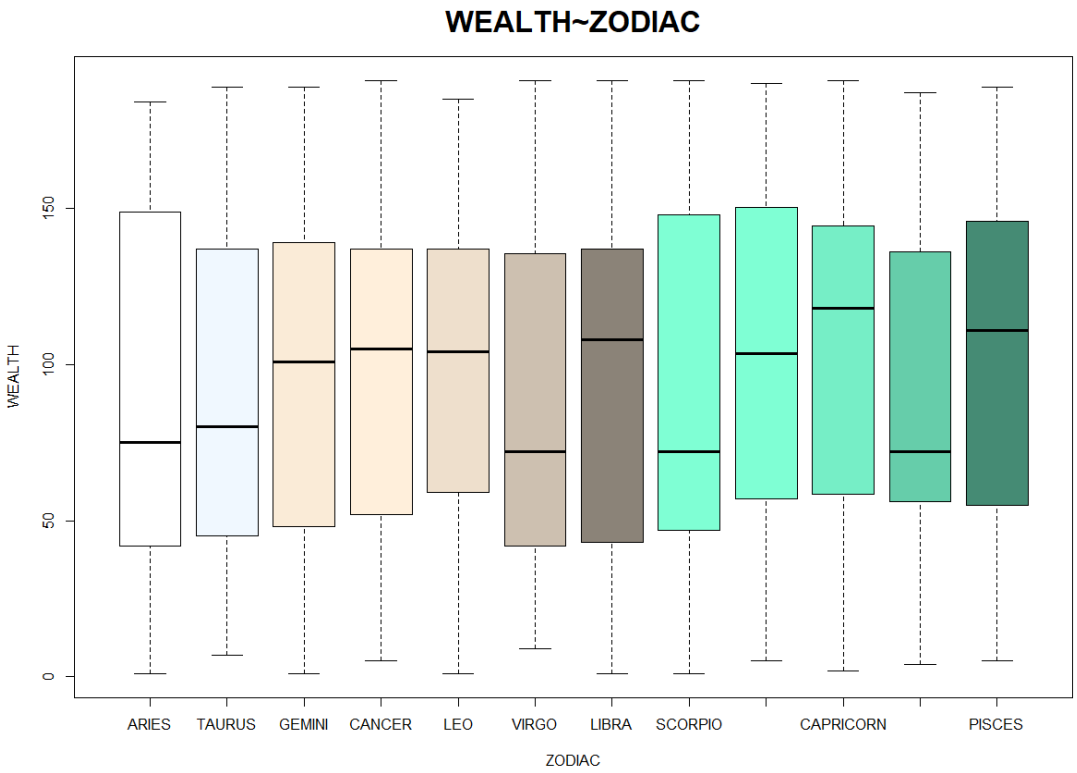
Διάγραμμα 15 Boxplot για την wealth και κάθε κατηγορία της marital,sex kai health



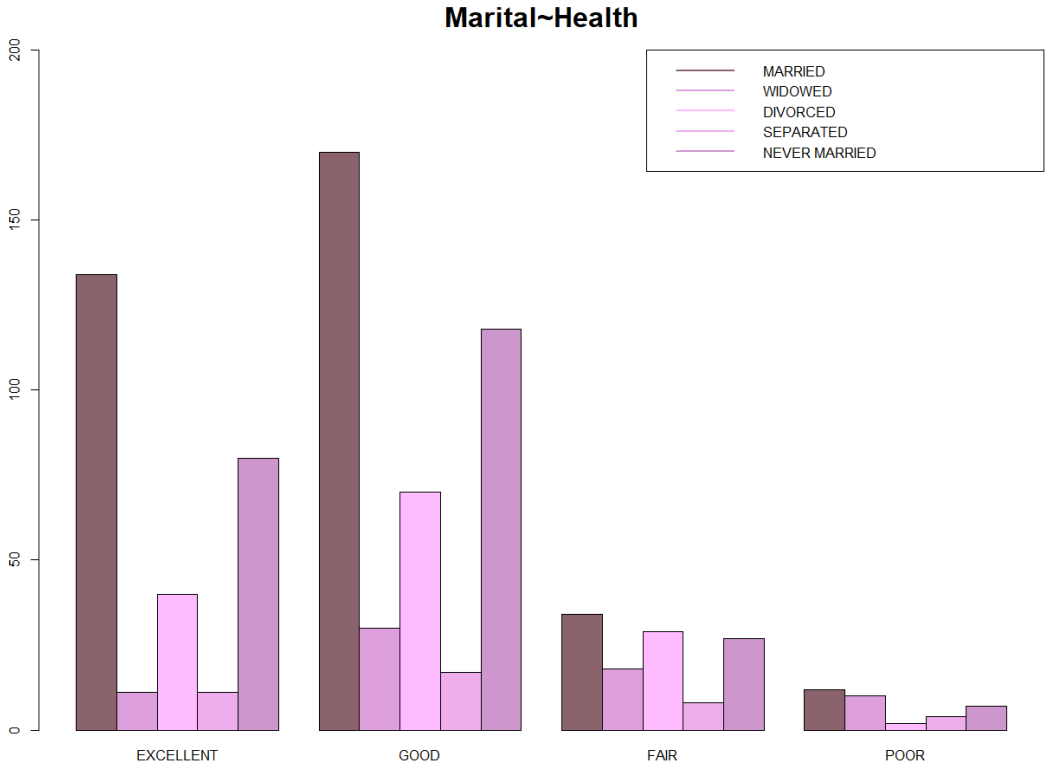




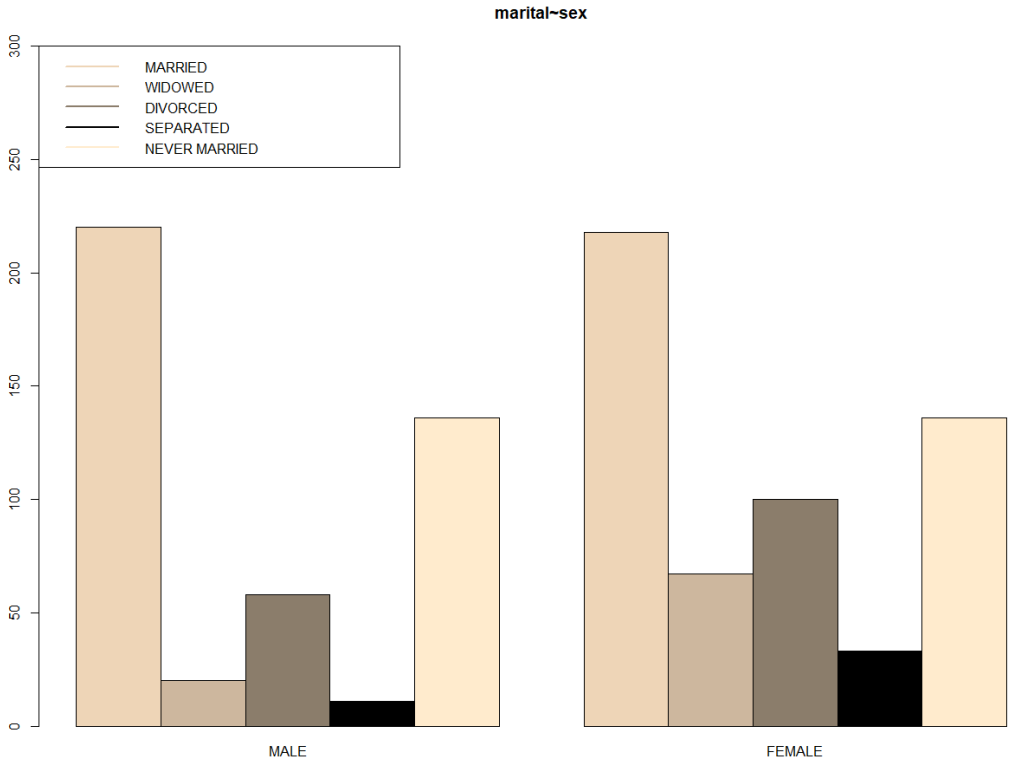
Διάγραμμα 16 Boxplot για την wealth και κάθε κατηγορία της zodiac



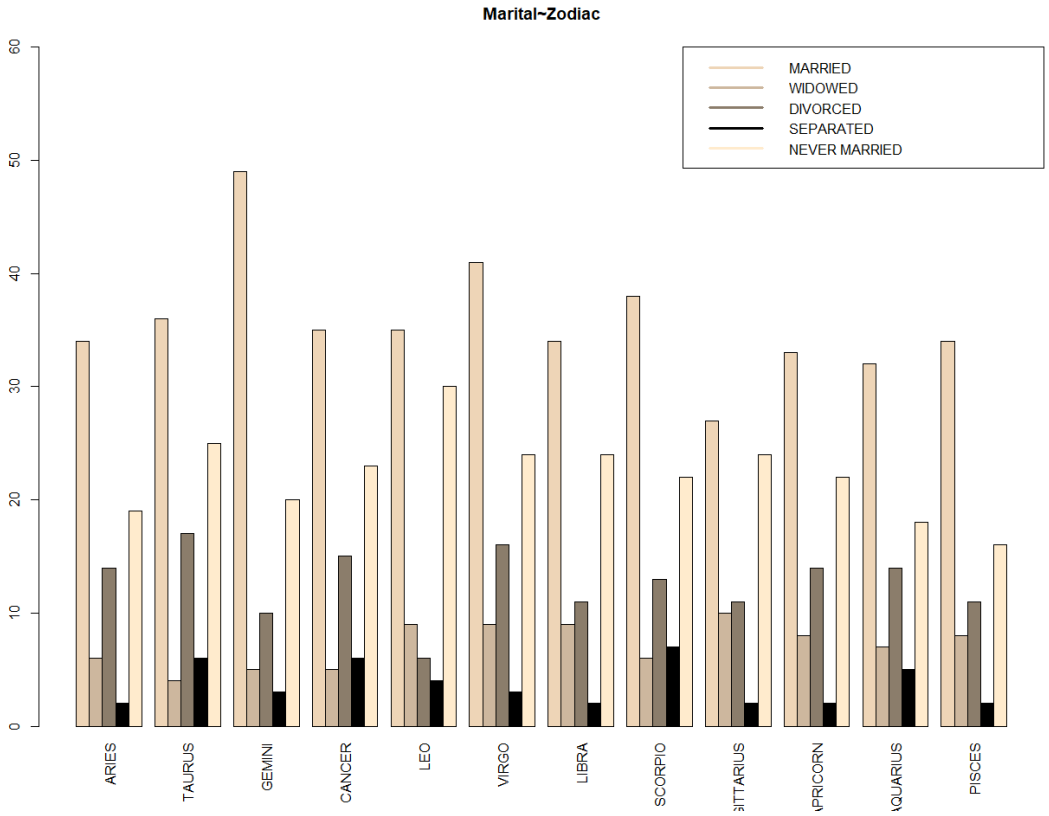
Διάγραμμα 17 Barplot για Marital και για health



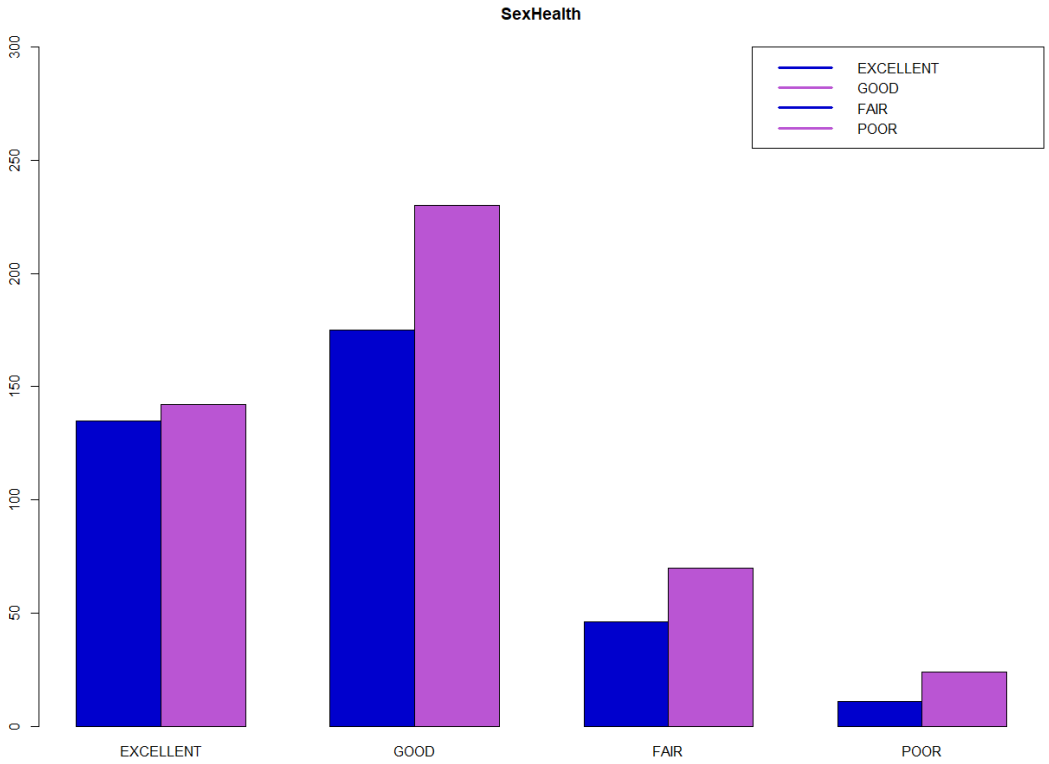
Διάγραμμα 18 Barplot για Marital και για sex



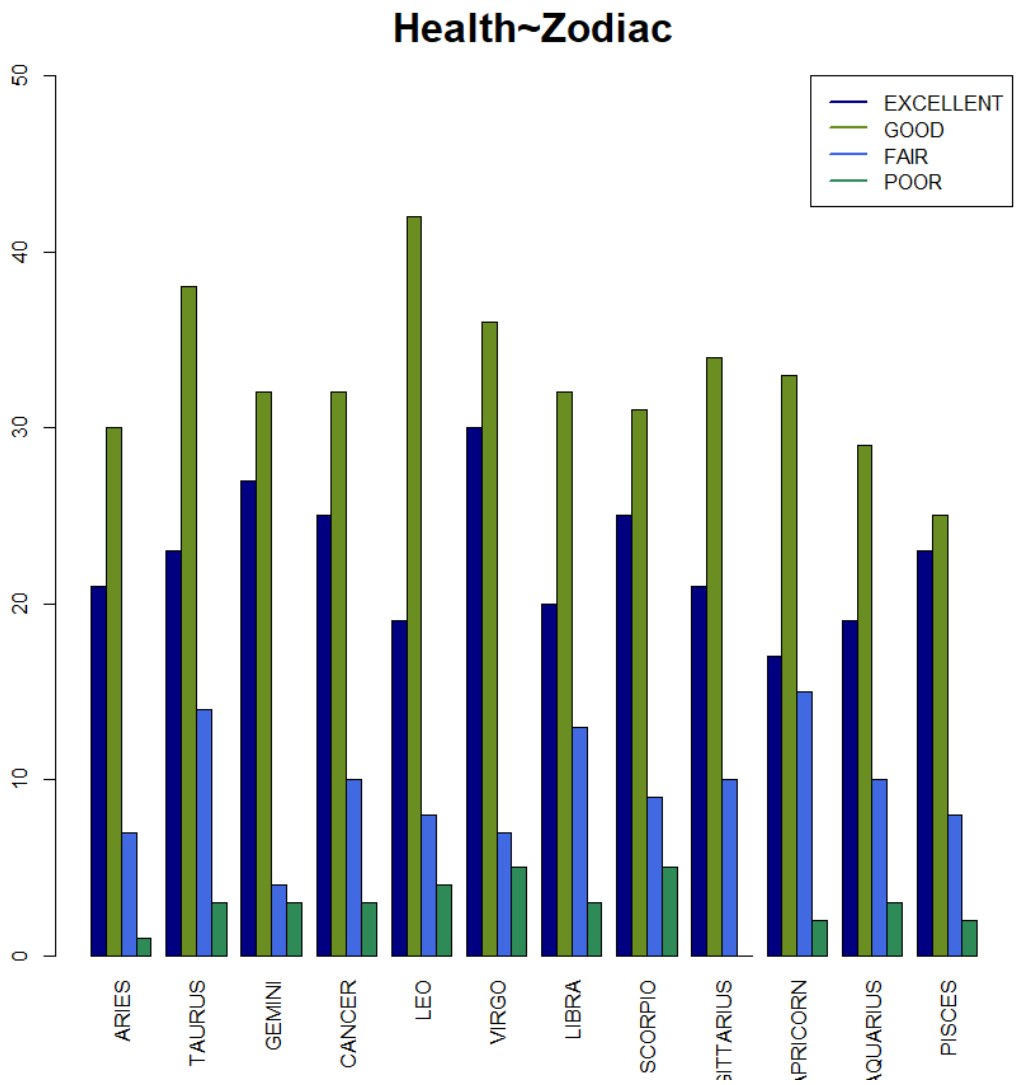
Διάγραμμα 19 Barplot για Marital και για zodiac



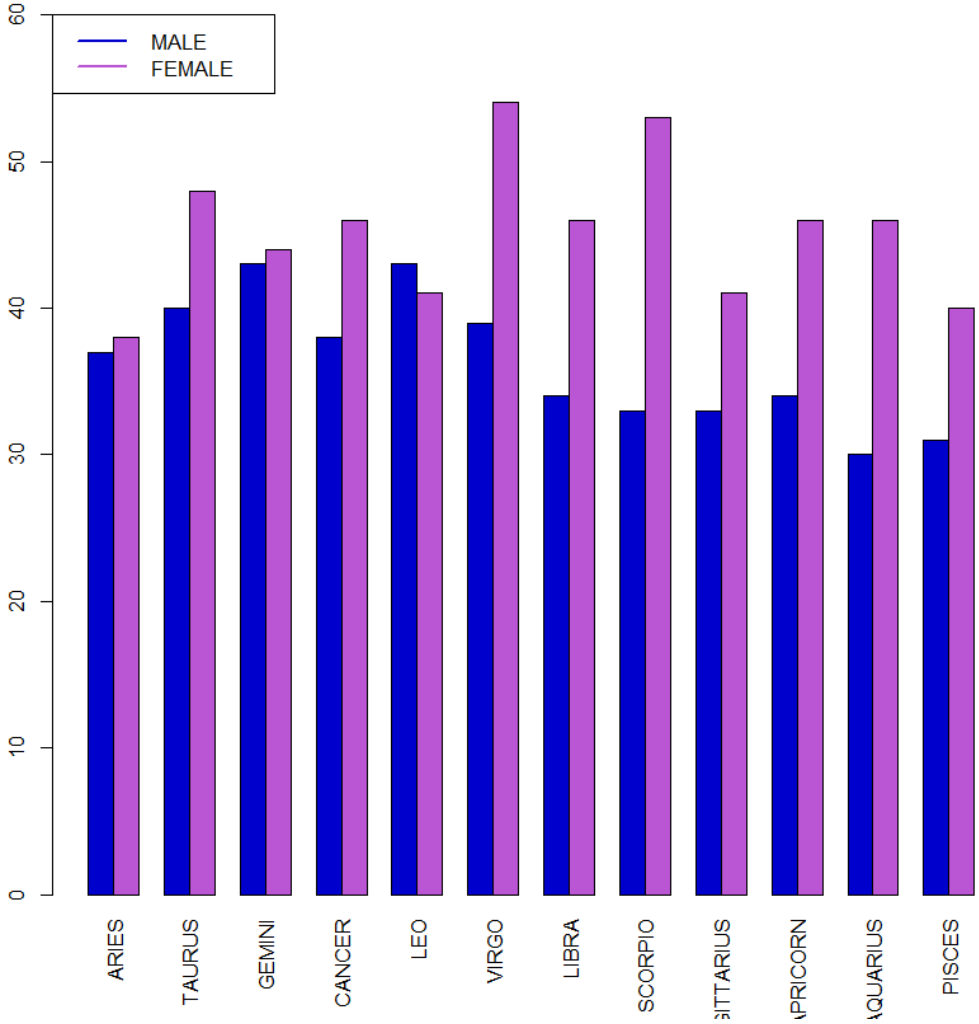
Διάγραμμα 20 Barplot για sex και για health



Διάγραμμα 21 Barplot για Health και κάθε κατηγορία της zodiac



Διάγραμμα 22 Barplot για sex και για zodiac.



Πίνακάς 4 Έλεγχος Shapiro-Wilk για κανονικότητα για της μεταβλήτης educ για τους κατηγορία MALE της μεταβλήτης sex .

Shapiro-Wilk normality test
data: educ_male$educ
W = 0.96043 , p-value = 1.484e-09

Πίνακας 5 Έλεγχος Shapiro-Wilk για κανονικότητα για της μεταβλήτης educ για τους κατηγορία FEMALE της μεταβλήτης sex.

Shapiro-Wilk normality test
data: educ_female$educ
W = 0.96627 , p-value = 6.009e-10

Πίνακάς 6 Ελέγχουμε εάν ο μέσος είναι κατάλληλο μετρό περιγραφής της κεντρικής θέσης για τις 2 κατηγορίες της μεταβλήτης sex. Εξετάζουμε αν ισχύει skew=0 και kurtosis=0 .Εδώ ισχύει skew<0(αρνητική ασσυμετρία) και kurtosis>0 (λεπτοκαρυές).

data$sex: MALE													
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	444	13.21	2.86	13	13.2	2.97	2	19	17	-0.13	0.34	0.14
data$sex: FEMALE													
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	552	12.51	3.08	12	12.57	2.97	1	19	18	-0.34	0.66	0.13

Πίνακας 7 Μη παραμετρικός έλεγχος για την ισότητα των διάμεσων για το μέγιστο ολοκληρωμένο έτος εκπαίδευσης ανάμεσα στα δύο φύλα . Το φύλο επηρεάζει στατιστικά σημαντικά για το μέγιστο ολοκληρωμένο έτος εκπαίδευσης .

Kruskal-Wallis rank sum test		
data: data$educ by data$sex		
Kruskal-Wallis chi-squared = 12.075 ,	df = 1,	p-value = 0.0005111

Πίνακας 8 Προσαρμογή του μοντέλου για τον έλεγχο της υπόθεσης κατά ποσό διαφέρει το ολοκληρωμένο έτος εκπαίδευσης ανάμεσα σε άτομά διαφορετικού ζωδίου (educ ~zodiac)

Call : aov(formula = lm(educ ~ zodiac, data = data))		
Terms:		
	Zodiac	Residuals
Sum of Squares	99.129	8550.509
Deg. of Freedom	11	964
Residual standard error:	2.978225	
Estimated effects may be unbalanced		
24 observations deleted due to missingness		

Πίνακάς 9 Έλεγχος κανονικότητας καταλοίπων του προσαρμοσμένου μοντέλου educ ~zodiac.

Shapiro-Wilk normality test	
data: m2$residuals	
W = 0.97999,	p-value = 2.534e-10

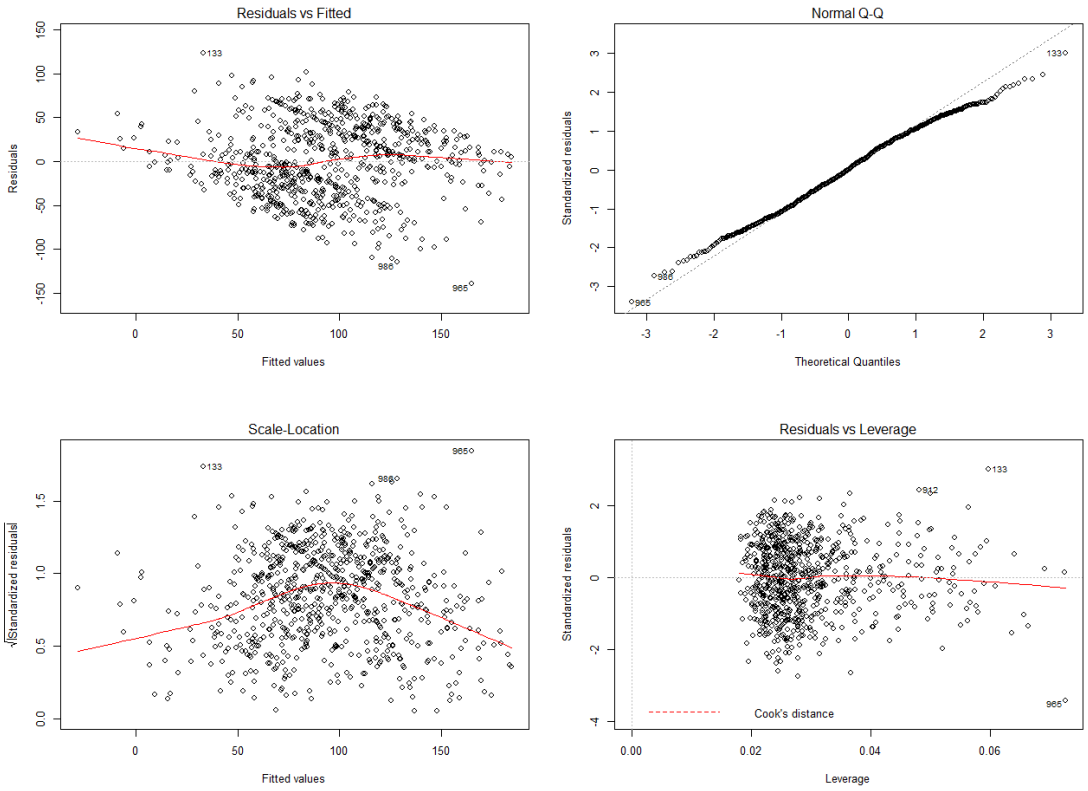
Πίνακάς 10 Μη παραμετρικός έλεγχος για την ισότητα των διαμέσων του μέγιστου ολοκληρωμένου έτος ανάμεσα στις κατηγορίες της μεταβλήτης zodiac.*Με βάση το p-value συμπεραίνουμε ότι οι διάμεσος του μέγιστου ολοκληρωμένου έτους εκπαίδευσης δεν διαφοροποιείται ανάμεσα στα ζώδια .*

Kruskal-Wallis rank sum test
data: data$educ by data$zodiac
Kruskal-Wallis chi-squared = 11.39, df = 11 , p-value = 0.4112

Πίνακάς 11 Έλεγχός κανονικότητας κατάλοιπων για το πλήρες μοντέλο

Shapiro-Wilk normality test
data: rstandard(fullmodel)
W = 0.99382, p-value = 0.002881

Διάγραμμά 23 διαγραμματικός έλεγχος προϋποθέσεων μοντέλου



Πίνακάς 12 Ελεγχος ομοσκεδαστηκότητας για το πλήρες μοντέλο

Levene's Test for Homogeneity of Variance (center = median)			
	Df	F value	Pr(>F)
group	3	7.225	8.739e-05 ***
	770		

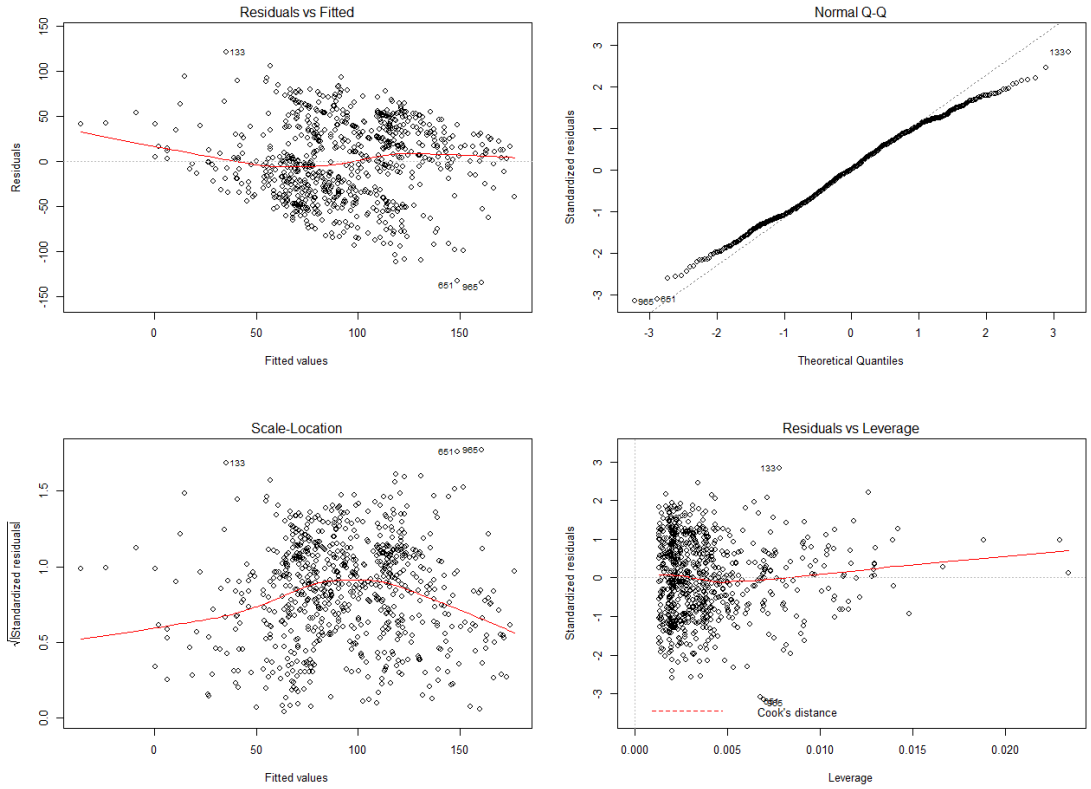
Πίνακας 13 Έλεγχος ανεξαρτησίας κατάλοιπων για το πλήρες μοντέλο.

lag	Autocorrelation	D-W Statistic	p-value
1	-0.01704596	2.031566	0.67
Alternative hypothesis: rho != 0			

Πίνακας 14 Έλεγχός κανονικότητας καταλοίπων για το μοντέλων wealth~educ_cen+age_cen

Shapiro-Wilk normality test	
data: rstandard(model2)	
W = 0.99245,	p-value = 0.000572

Διάγραμμα 4 Διαγραμματικός έλεγχός υποθέσεων για το μοντέλο wealth~educ_cen+age_cen



Πίνακας 15 έλεγχός ομοσκέδαστικότητας για το μοντέλο wealth~educ_cen+age_cen

Levene's Test for Homogeneity of Variance (center = median)			
	Df	F value	Pr(>F)
group	3	2.8236	0.03791 *
	770		

Πίνακας 16 Έλεγχός κανονικότητας καταλοίπων για το μοντέλων wealth~educ_cen+age_cen

Lag	Autocorrelation	D-W Statistic	p-value
1	-0.02051654	2.038378	0.6