Έρευνα σχετικά με τη χρήση κινητών τηλεφώνων

Γκόγκος Χρίστος 72

Παπαδοπούλου Δανάη 60

Στατιστική Ανάλυση Δεδομένων

Π.Μ.Σ. Τεχνολογίες Διαδραστικών Συστημάτων

Επιβλέπων Καθηγητής: Δρ. Αγγελής Ελευθέριος

Θεσσαλονίκη,
Φεβρουάριος 2022

Περιεχόμενα

Εισαγωγή

Η εισαγωγή δεδομένων έγινε μόνο από το αρχείο Cell_Phones_labels με την χρήση import dataset -> from excel ->browse έπειτα όλες τις μεταβλητές από Q17a εως και την Q17g άλλαξα τον τύπο τους απο logical σε Character.
Στην παρούσα έκθεση γίνεται στατιστική ανάλυση των δεδομένων που συλλέχθηκαν έπειτα από έρευνα που πραγματοποιήθηκε σχετικά με τη χρήση των κινητών τηλεφώνων. Στον παρακάτω πίνακα αναφέρεται η περιγραφή των μεταβλητών που χρησιμοποιήθηκαν στην έρευνα.

Όνομα Μεταβλητής	Περιγραφή (label)
psraid	ID
usr_r	Περιοχή κατοικίας (1 = Urban, 2 = Suburban, 3 = Rural)
sex	Φύλο (1 = "Male"/ 2 = "Female")
q10c	Έχετε κινητό τηλέφωνο ή κάποια συσκευή που λειτουργεί και σαν κινητό; (1 = "Yes"/ 2 = "No")
q14a	Χρησιμοποιείτε το κινητό για να στέλνετε/λαμβάνετε email? (1 = "Yes"/ 2 = "No")
q14b	Χρησιμοποιείτε το κινητό για να στέλνετε/λαμβάνετε γραπτά μηνύματα; (1 = "Yes"/ 2 = "No")
q14c	Χρησιμοποιείτε το κινητό για να βγάλετε φωτογραφία; (1 = "Yes"/ 2 = "No")
q14d	Χρησιμοποιείτε το κινητό για να ακούσετε μουσική; (1 = "Yes"/ 2 = "No")
q14e	Χρησιμοποιείτε το κινητό για να στέλνετε/λαμβάνετε άμεσα μηνύματα; (1 = "Yes"/ 2 = "No")
q14g	Χρησιμοποιείτε το κινητό για να παίξετε παιχνίδια; (1 = "Yes"/ 2 = "No")
q14h	Χρησιμοποιείτε το κινητό για να μπείτε στο internet? (1 = "Yes"/ 2 = "No")
q17a	Χρησιμοποιείτε το κινητό για να στείλετε φωτογραφία σε κάποιον; (1 = "Yes, do this"/2 = "No, do not do this/Have not done this"/3 = "Cell phone can’t do this")
q17b	Χρησιμοποιείτε το κινητό για να ανεβάσετε φωτογραφίες/ βίντεο online; (1 = "Yes, do this"/2 = "No, do not do this/Have not done this"/3 = "Cell phone can’t do this")
q17c	Χρησιμοποιείτε το κινητό για να αγοράσετε κάτι; (1 = "Yes, do this"/2 = "No, do not do this/Have not done this"/3 = "Cell phone can’t do this")

q17d	Χρησιμοποιείτε το κινητό για να κάνετε μια φιλανθρωπική δωρεά μέσω γραπτού μηνύματος;
q17e	Χρησιμοποιείτε το κινητό για να έχετε πρόσβαση σε κάποιο κοινωνικό δίκτυο, όπως MySpace, Facebook ή LinkedIn.com? (1 = "Yes, do this"/2 = "No, do not do this/Have not done this"/3 = "Cell phone can't do this")
q17f	Χρησιμοποιείτε το κινητό για να μπείτε στο Twitter ή σε κάποια παρόμοια σελίδα; (1 = "Yes, do this"/2 = "No, do not do this/Have not done this"/3 = "Cell phone can't do this")
q17g	Χρησιμοποιείτε το κινητό για να παρακολουθήσετε ένα βίντεο; (1 = "Yes, do this"/2 = "No, do not do this/Have not done this"/3 = "Cell phone can't do this")
q18	Κατά μέσο όρο, πόσες κλήσεις περίπου κάνετε/λαμβάνετε μέσα σε μια μέρα;
q20	Κατά μέσο όρο, πόσα μηνύματα περίπου στέλνετε/λαμβάνετε μέσα σε μια μέρα;
q22a	Νιώθετε πιο ασφαλής επειδή μπορείτε να χρησιμοποιήσετε το κινητό σας για να καλέσετε βοήθεια; (1 = "Agree"/2 = "Disagree"/3 = "Neither agree nor disagree/Agree some-Disagree some")
q22b	Θεωρείτε ότι το κινητό διευκολύνει το να κάνετε σχέδια με άλλους; (1 = "Agree"/2 = "Disagree"/3 = "Neither agree nor disagree/Agree some-Disagree some")
q22c	Εκνευρίζεστε όταν σας διακόπτει μια κλήση ή ένα μήνυμα; (1 = "Agree"/2 = "Disagree"/3 = "Neither agree nor disagree/Agree some-Disagree some")
q22d	Χρησιμοποιείτε το κινητό για να περάσει η ώρα όταν βαριέστε; (1 = "Agree"/2 = "Disagree"/3 = "Neither agree nor disagree/Agree some-Disagree some")
q22e	Θεωρείτε αγένεια να διακόπτει κάποιος συνεχώς μια συζήτηση για να μιλήσει στο κινητό; (1 = "Agree"/2 = "Disagree"/3 = "Neither agree nor disagree/Agree some-Disagree some")
q24	Έχετε κατεβάσει κάποια εφαρμογή στο κινητό σας; (1 = "Yes, have done this"/2 = "No, have never done this"/3 = "Phone can NOT download apps")
q25	Είχε το κινητό σας ήδη εγκατεστημένες εφαρμογές; (1 = "Yes"/ 2 = "No")
q26	Πόσες εφαρμογές έχετε τώρα στο κινητό σας;
q29	Έχετε πληρώσει ποτέ για να αποκτήσετε κάποια

	εφαρμογή; (1 = "Yes, have paid for app"/2 = "Only download apps that are free")
age	Ηλικία
mar	Οικογενειακή κατάσταση 1 Married 2 Living with a partner 3 Divorced 4 Separated 5 Widowed 6 Never been married 7 Single
educ	Εκπαίδευση 1 None, or grades 1-8 2 High school incomplete (grades 9-11) 3 High school graduate (grade 12 or GED certificate) 4 Technical, trade or vocational school AFTER high school 5 Some college, no 4-year degree (includes associate degree) 6 College graduate (B.S., B.A., or other 4-year degree) 7 Post-graduate training/professional school after college (toward a Masters/Ph.D., Law or Medical school)
empl	Επάγγελμα 1 Employed full-time 2 Employed part-time 3 Retired 4 Not employed for pay 5 Have own business/self-employed 6 Disabled 7 Student 8 Other
inc	Εισόδημα 1 Less than $10,000 2 $10,000 to under $20,000 3 $20,000 to under $30,000 4 $30,000 to under $40,000 5 $40,000 to under $50,000 6 $50,000 to under $75,000 7 $75,000 to under $100,000 8 $100,000 to under $150,000 9 $150,000 or more
mobileprice	Τιμή κινητού

Η ανάλυση χωρίστηκε σε δύο στάδιο. Στο πρώτο στάδιο τέθηκαν ερευνητικά ερωτήματα για σχέσεις ανάμεσα σε μεταβλητές και με τη χρήση της R πραγματοποιήθηκε περιγραφική στατιστική στις μεταβλητές. Στο δεύτερο στάδιο αναπτύχθηκαν μοντέλα με εξαρτημένες και ανεξάρτητες μεταβλητές.

Στάδιο 1

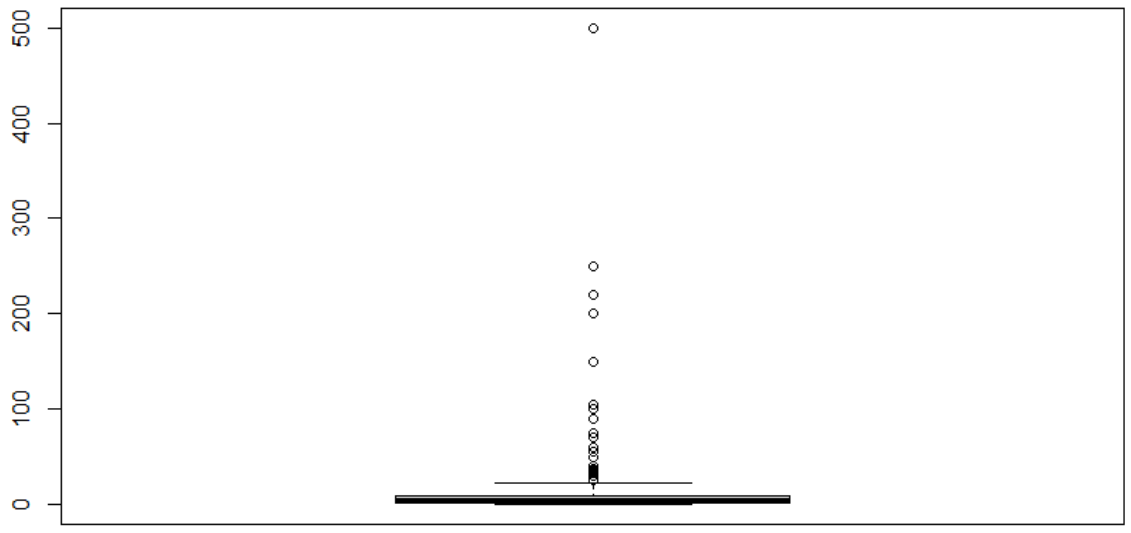
Αρχικά, ενδιαφέρον προέκυψε για τις μεταβλητές που έχουν αριθμητική φύση, δηλαδή οι q18, q20, age, και mobileprice, για τις οποίες μάλιστα έγινε αναλυτική περιγραφική στατιστική στην R, όπως επισυνάπτεται στο Παράρτημα.

Για κάθε μία από τις μεταβλητές, υπολογίστηκαν και αναλύθηκαν οι παρακάτω μετρικές και εξάγαμε γραφήματα για την καλύτερη κατανόηση των αποτελεσμάτων.
Αρχικά, χρησιμοποιήθηκαν οι συναρτήσεις summary( ) και describe( ) για να υπολογίσουμε το ελάχιστο και το μέγιστο, τη μέση και ενδιάμεση τιμή, το πρώτο και το τρίτο τεταρτημόριο, την τυπική απόκλιση, το τυπικό σφάλμα κτλ. Επιπλέον, εξάγαμε το ιστόγραμμα για τις τιμές κάθε μεταβλητής και υλοποιήσαμε την boxplot προκειμένου να ερμηνεύσουμε καλύτερα τα παραπάνω αποτελέσματα μέσω των γραφημάτων. Στη συνέχεια, υλοποιήσαμε μια qqnorm συνάρτηση προκειμένου να ελέγξουμε την κανονικότητα των τιμών της μεταβλητής και πιθανώς και το είδος της κατανομής (κανονική).

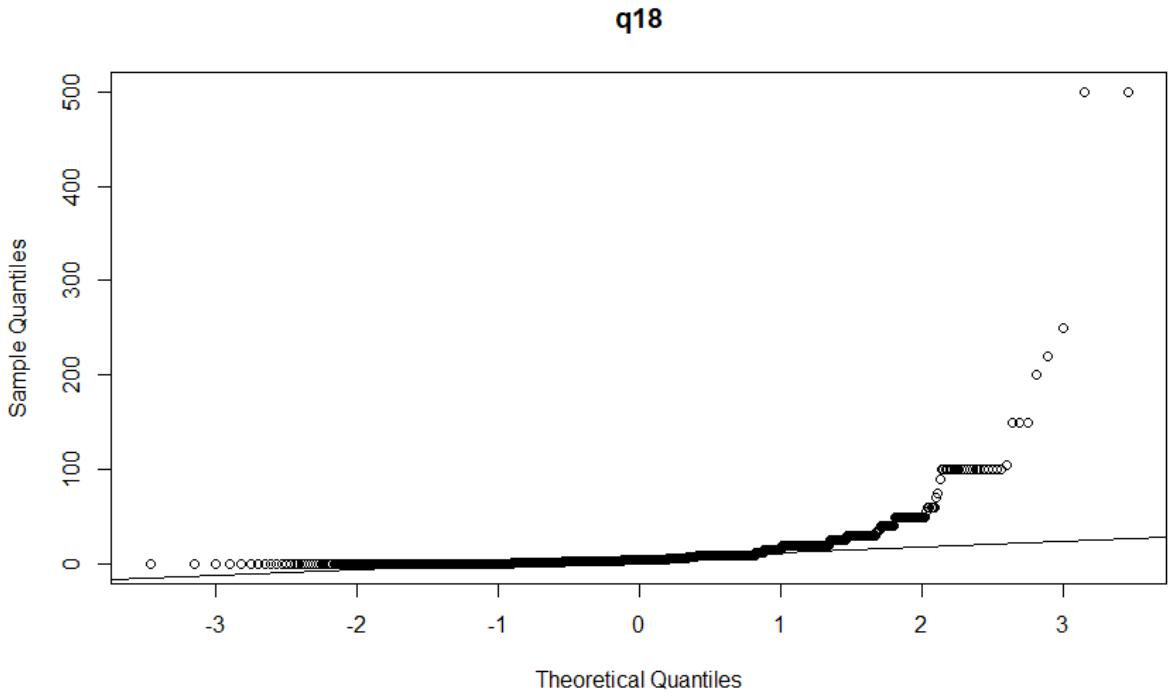
q18:

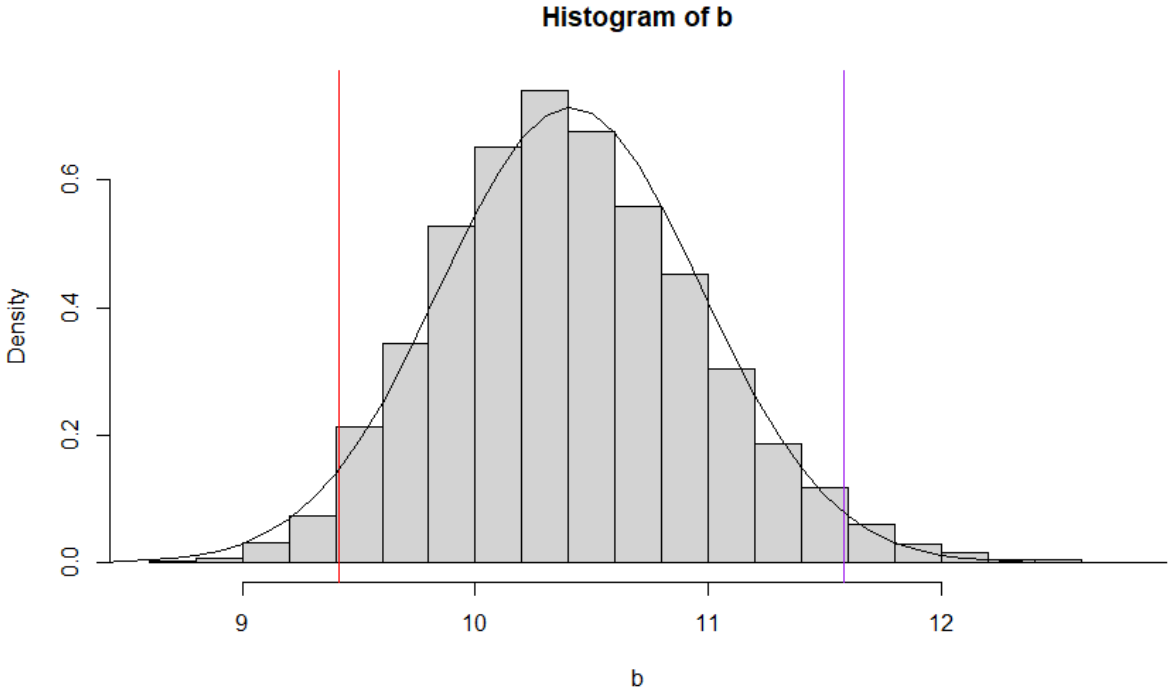
```
> psych::describe(q18)
   vars    n  mean    sd median trimmed  mad min max range skew kurtosis   se
X1    1 1846 10.41 23.81      5    6.47 5.93   0 500   500 11.7   205.43 0.55
> summary(q18)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00    2.00    5.00   10.41   10.00  500.00     406
```

Όπως παρουσιάζονται στα παραπάνω αποτελέσματα, η μέση τιμή της q18 είναι 10.41, η διάμεσος 5, η ελάχιστη τιμή 0 και η μέγιστη 500.



q18

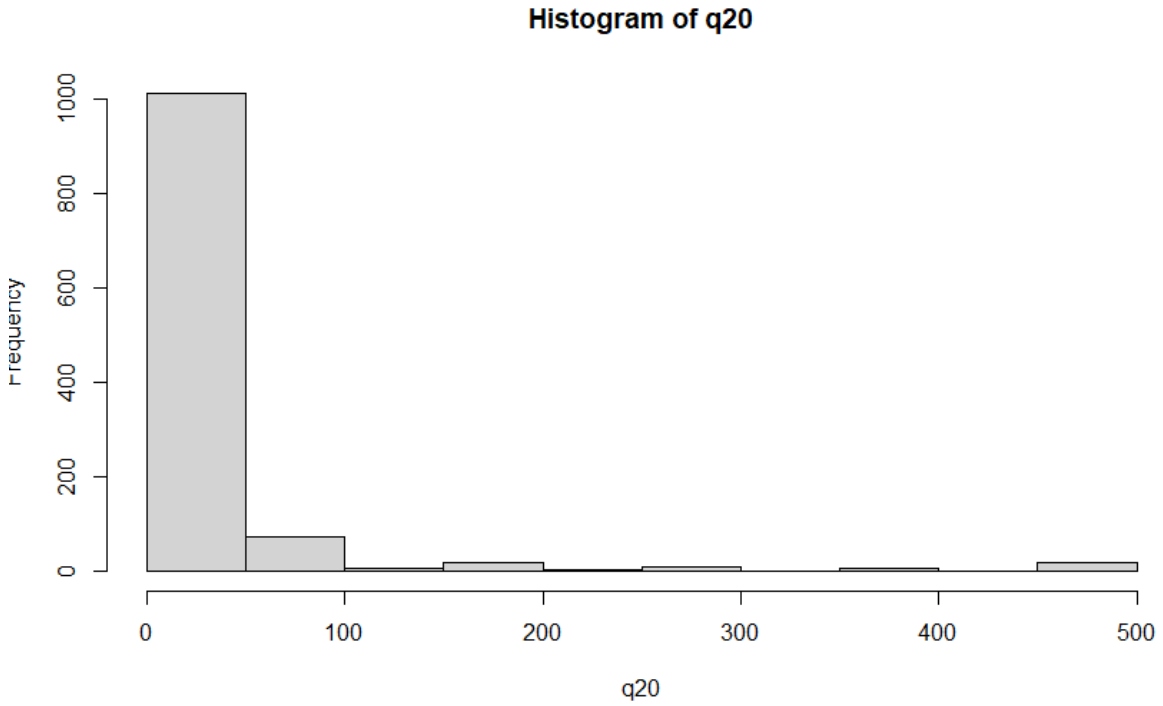




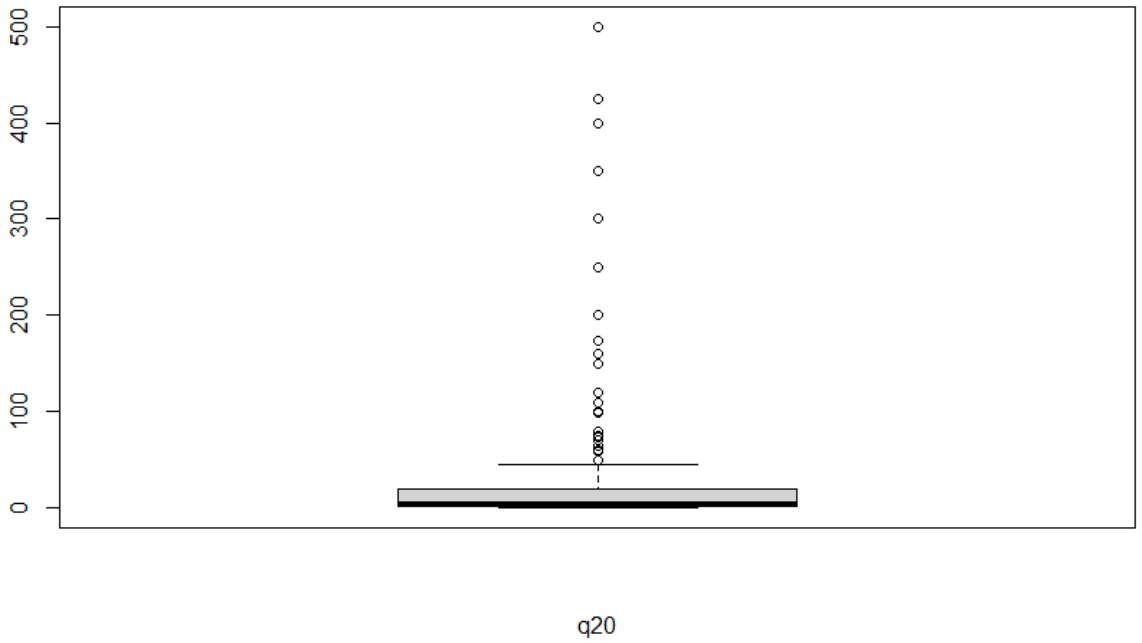
q20:

```
> #q20
> q20 = deframe(Cell_Phones_labels[20][,1])
> summary(q20)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00    2.00    5.00   32.14   20.00  500.00    1099
> psych::describe(q20)
   vars    n  mean    sd median trimmed  mad min max range skew kurtosis   se
X1    1 1153 32.14 80.94      5   12.14 7.41   0 500   500 4.27    19.32 2.38
```

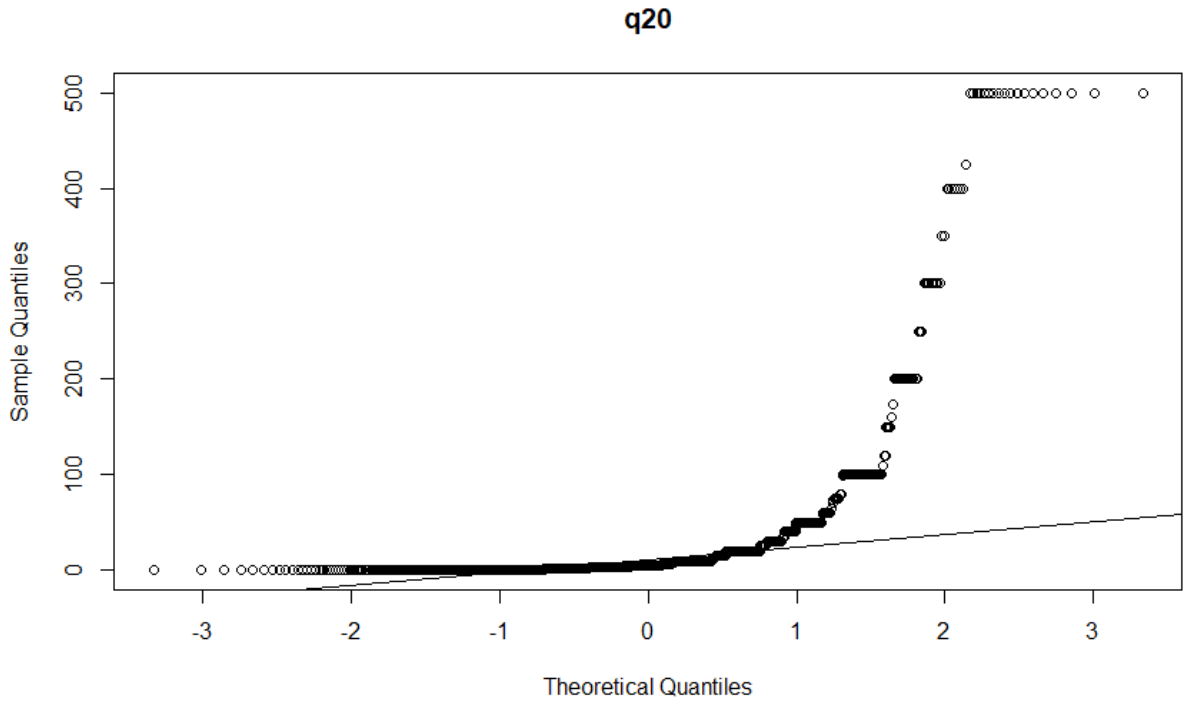
Όπως παρουσιάζονται στα παραπάνω αποτελέσματα, η μέση τιμή της q20 είναι 32.14, η διάμεσος 5, η ελάχιστη τιμή 0 και η μέγιστη 500.

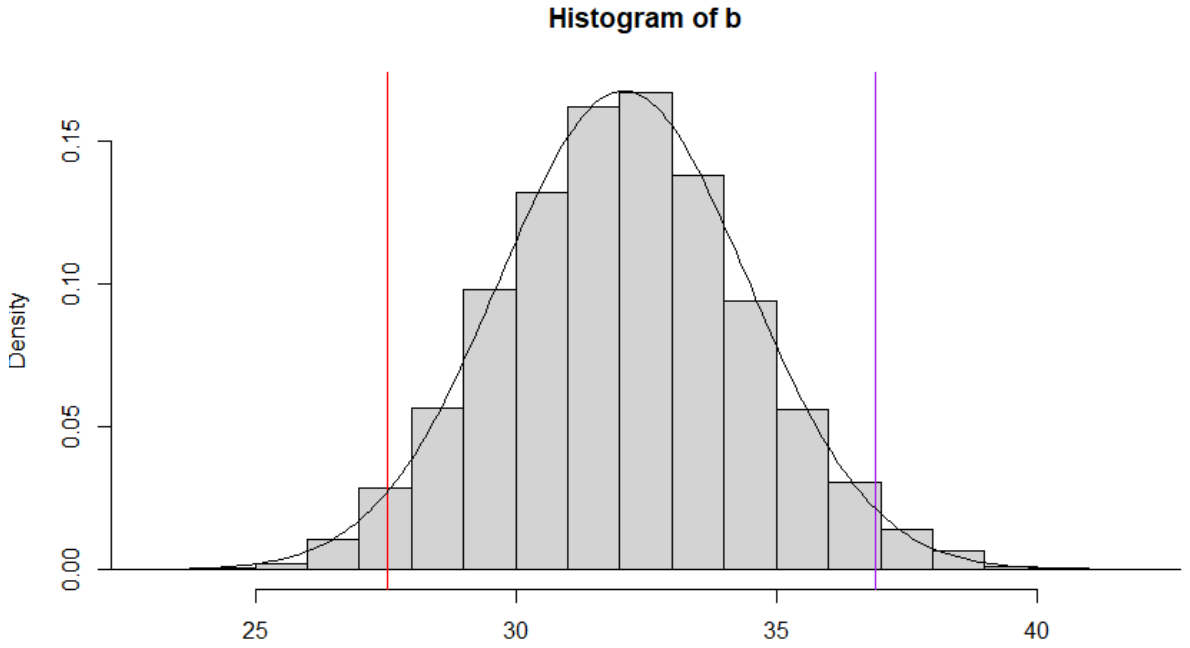


Παρατηρούμε ότι τα περισσότερα δείγματα ανήκουν μεταξύ 0-50.



Τα αποτελέσματα του ιστογράμματος επιβεβαιώνονται και από το boxplot.

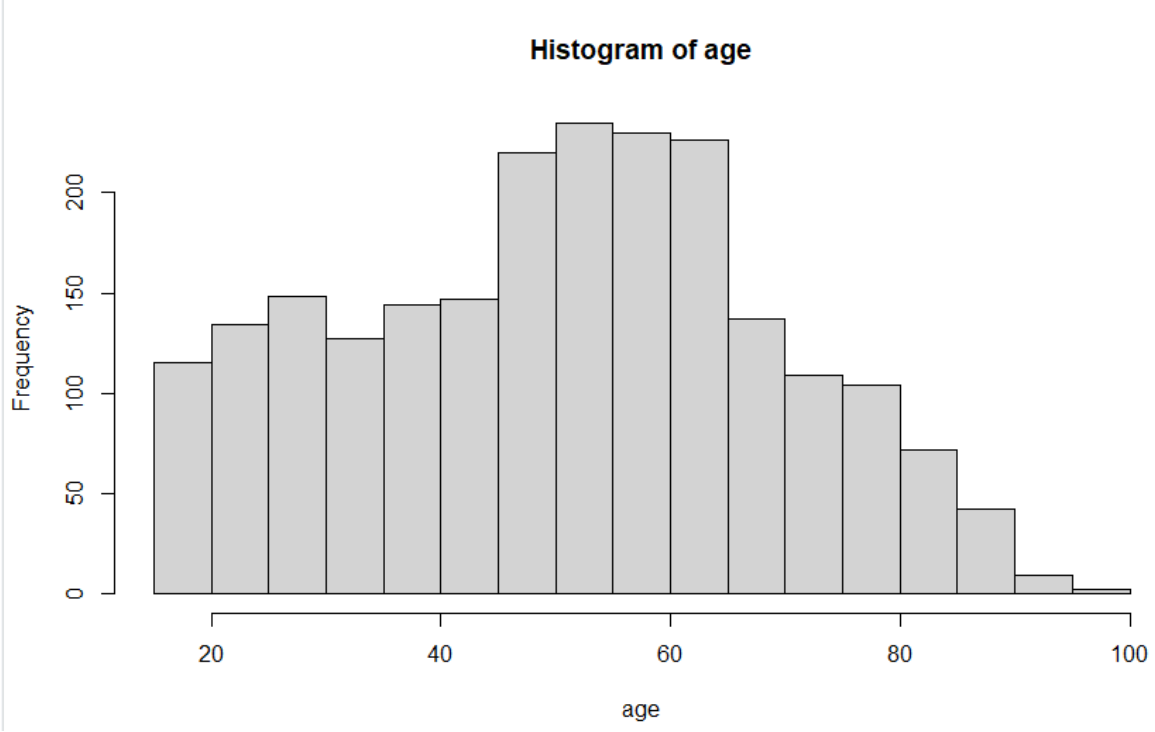




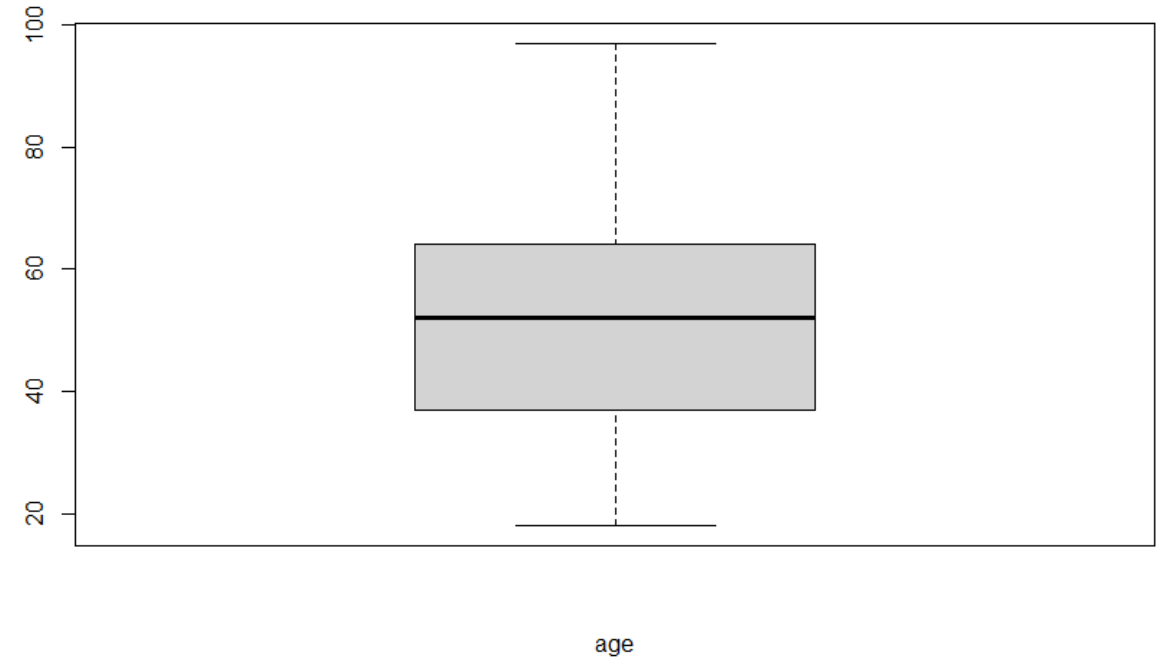
age:

```
> #age
> age = deframe(Cell_Phones_labels[30][,1])
> summary(age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  18.00   37.00   52.00   50.98   64.00   97.00      51
> psych::describe(age)
   vars    n  mean    sd median trimmed   mad min max range skew kurtosis   se
X1    1 2201 50.98 18.43     52   50.86 19.27  18  97    79    0     -0.8 0.39
```

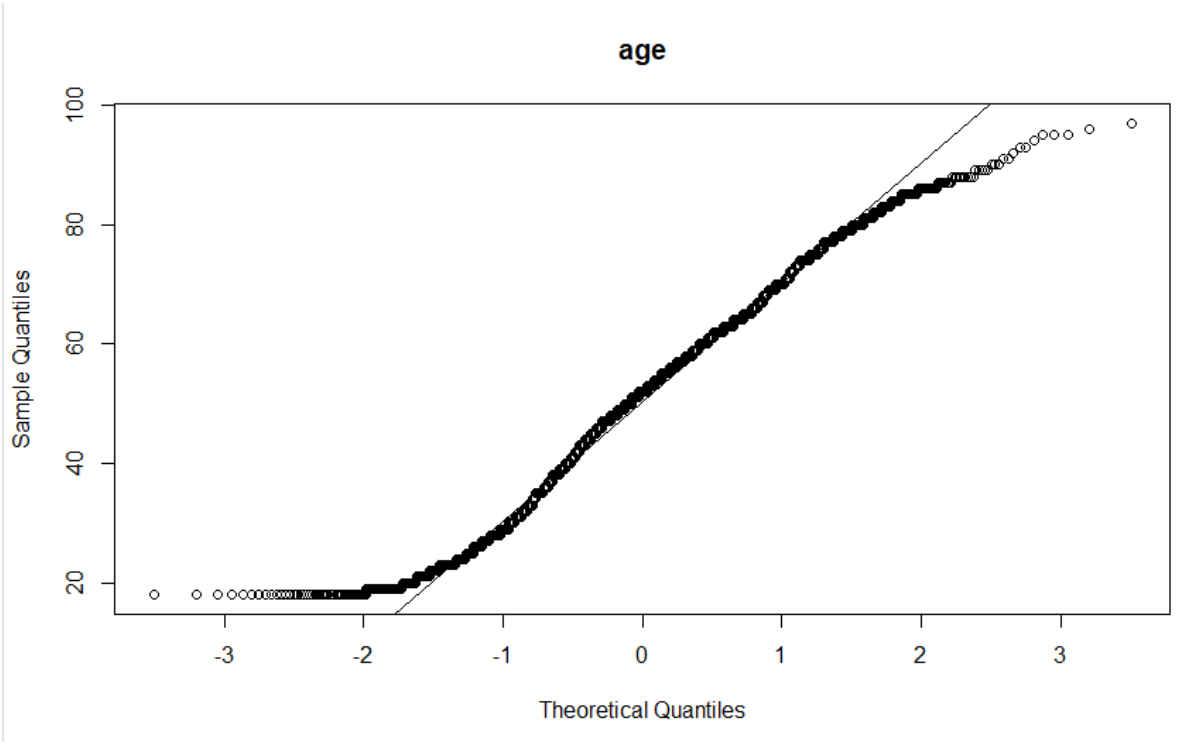
Όπως παρουσιάζονται στα παραπάνω αποτελέσματα, η μέση τιμή της age είναι 50.98, η διάμεσος 52, η ελάχιστη τιμή 18 και η μέγιστη 97.



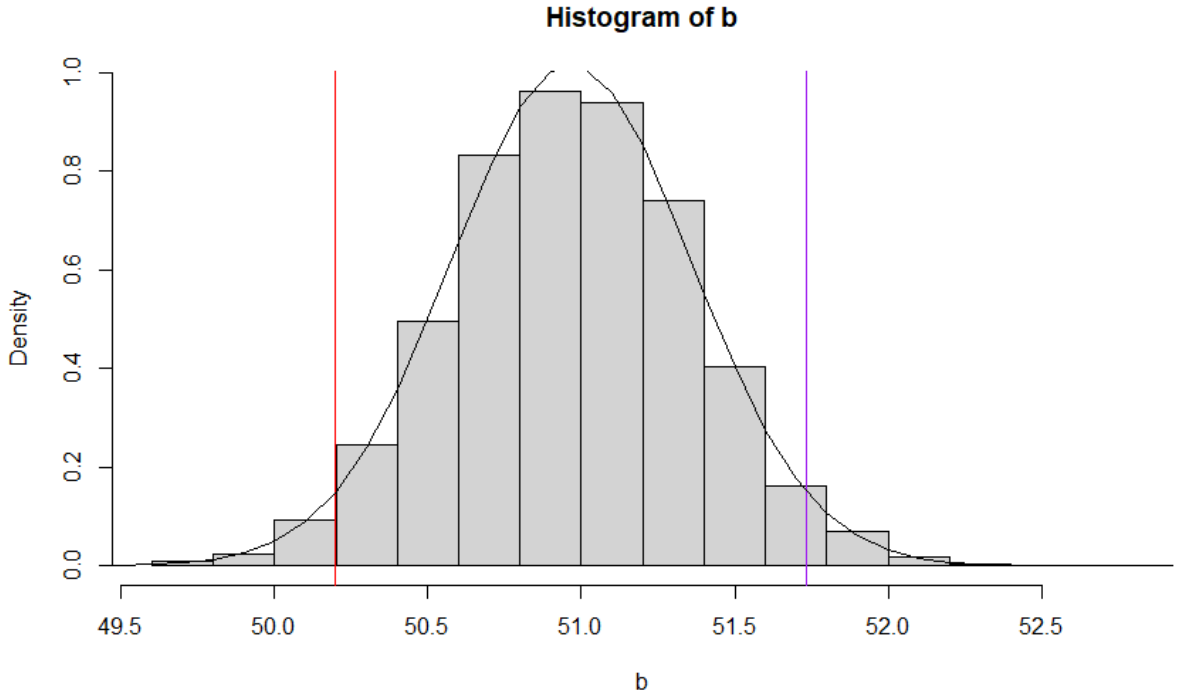
Τα δείγματα φαίνονται να ακολουθούν σε μεγάλο βαθμό την κανονική κατανομή.



Αδύνατον να έχουμε outliers εφόσον η ηλικία των ανθρώπων δεν ξεπερνάει εύκολα την τιμή 100, ιδιαίτερα σε αυτούς που χρησιμοποιούν τηλέφωνο.



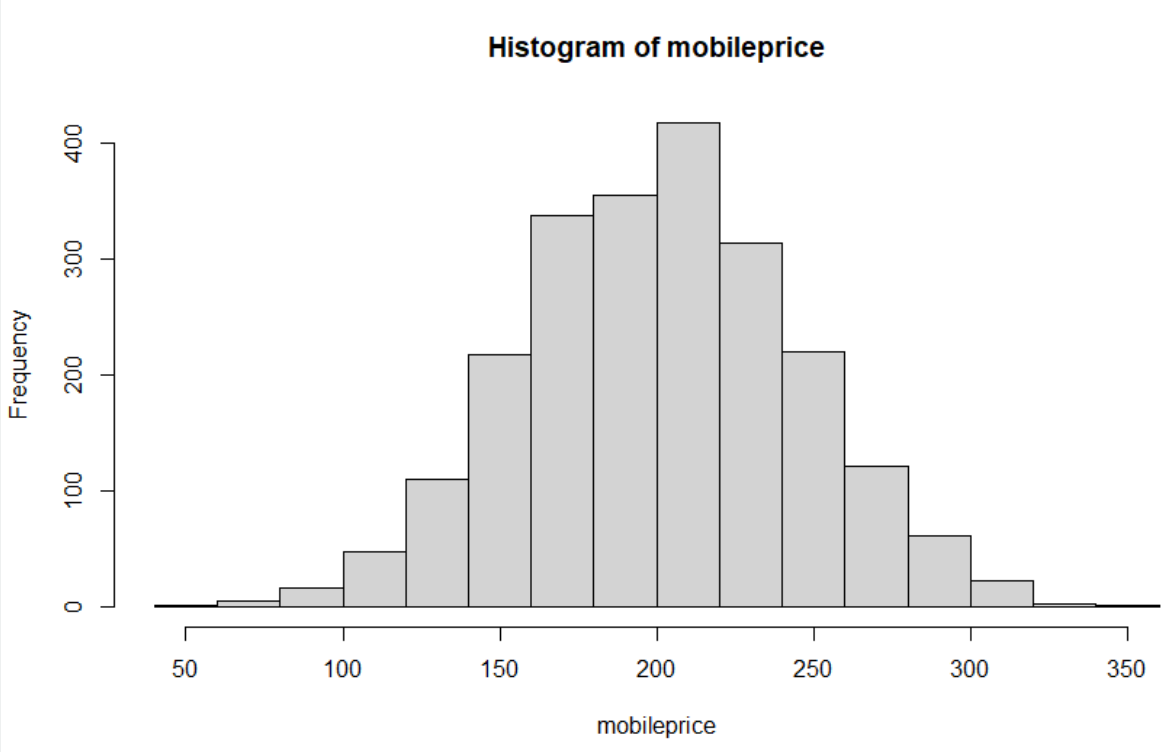
Τα δεδομένα, όπως φαίνονταν και στο ιστόγραμμα, είναι αρκετά κοντά στη γραμμή κανονικότητας.



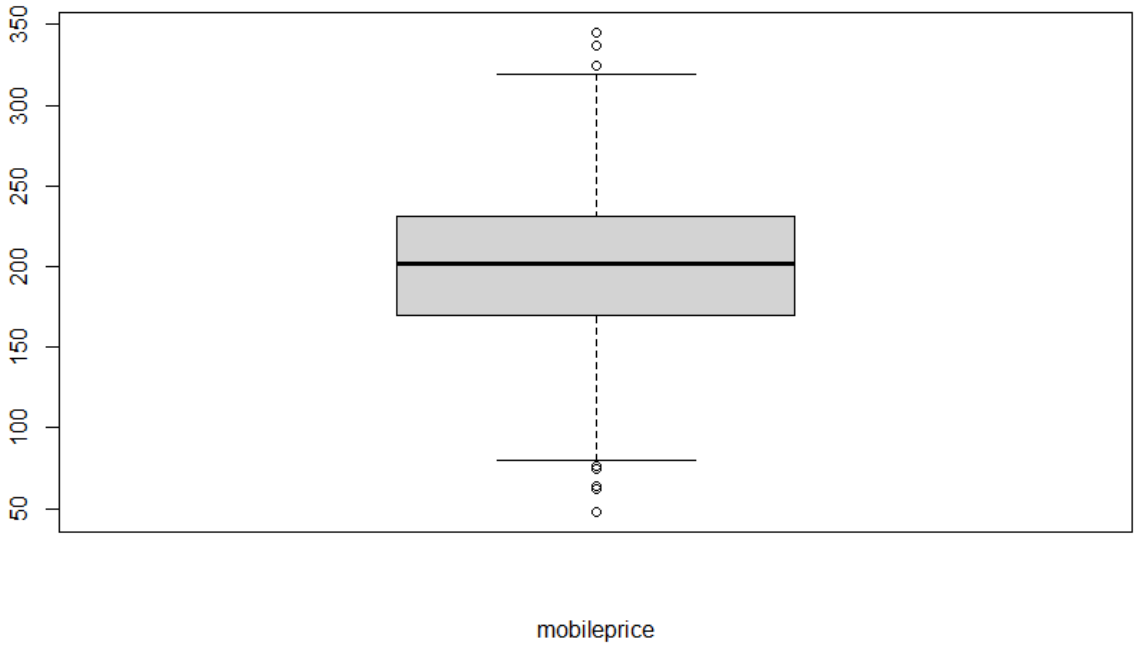
mobileprice:

```
> #mobileprice
> mobileprice = deframe(Cell_Phones_labels[35][,1])
> summary(mobileprice)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   47.8   170.2   201.4   201.3   231.2   345.3
> psych::describe(mobileprice)
   vars    n   mean    sd median trimmed   mad  min   max range skew kurtosis   se
X1    1 2252 201.31 44.28  201.4  201.19 44.92 47.8 345.3 297.5 0.01    -0.11 0.93
```

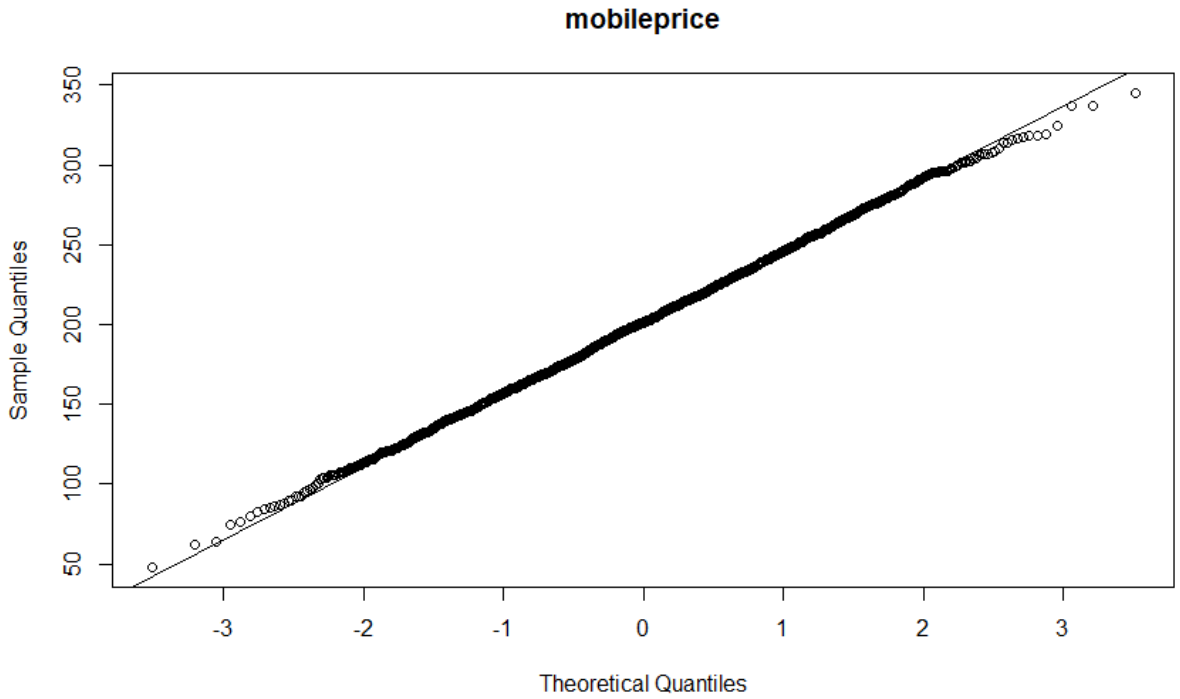
Όπως παρουσιάζονται στα παραπάνω αποτελέσματα, η μέση τιμή της mobileprice είναι 201.3, η διάμεσος 201.4, η ελάχιστη τιμή 47,8 και η μέγιστη 345.3.



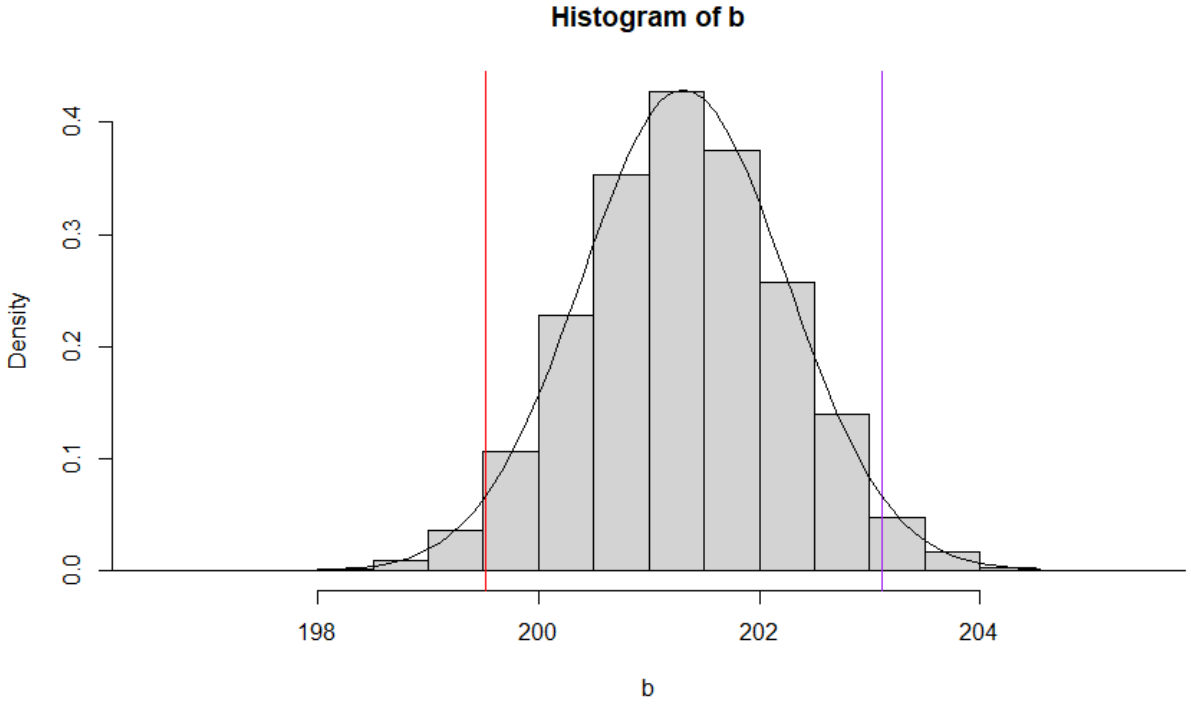
Οι τιμές της mobileprice δείχνουν ξεκάθαρα να ακολουθούν την κανονική κατανομή.



Λίγοι outliers, που δικαιολογούνται από την κανονική μεταβλητή

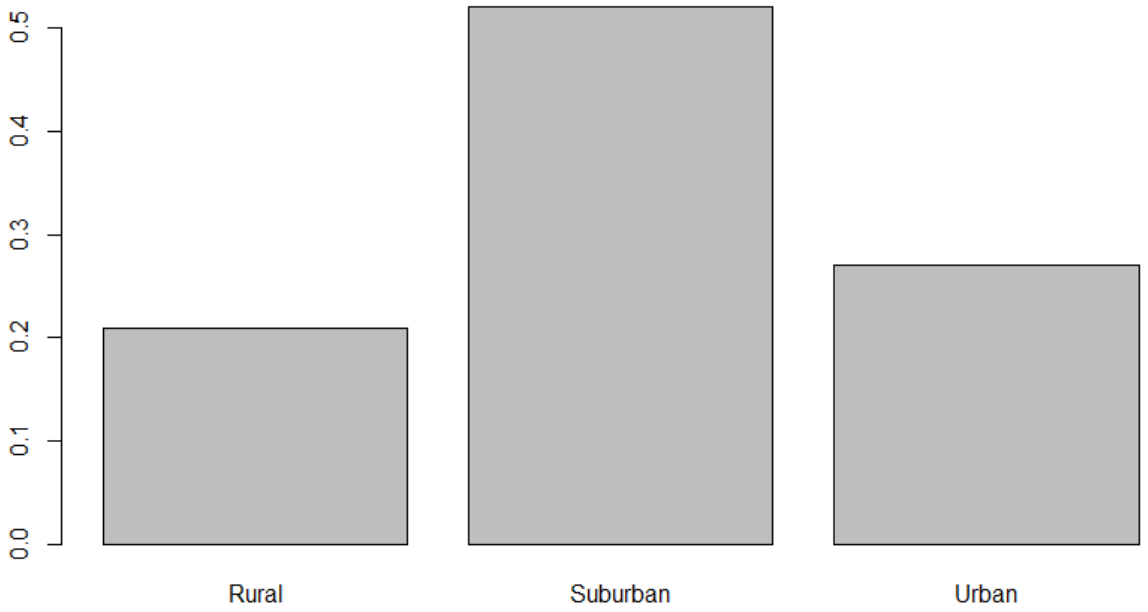


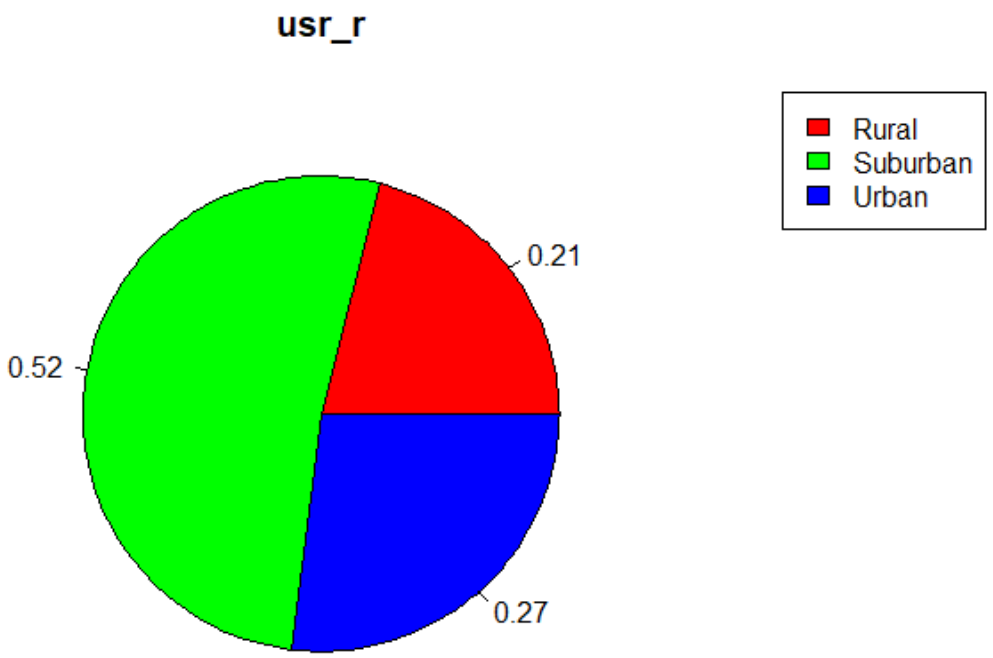
Τα δείγματα σχεδόν ταυτίζονται με τη γραμμή, επομένως η μεταβλητή mobileprice υπακούει στην κανονικότητα.



Για τις υπόλοιπες μεταβλητές, οι οποίες είναι κατηγορικές υλοποιήσαμε τις παρακάτω λειτουργίες. Αρχικά, υλοποιήσαμε τη συνάρτηση freq για να διεξάγουμε τον πίνακα συχνότητας για κάθε μεταβλητή και επίσης εξάγαμε barplots και pies για την απεικόνιση των τιμών μέσω γραφημάτων, Ενδεικτικά παρουσιάζονται παρακάτω οι usr_r, sex και q10c.

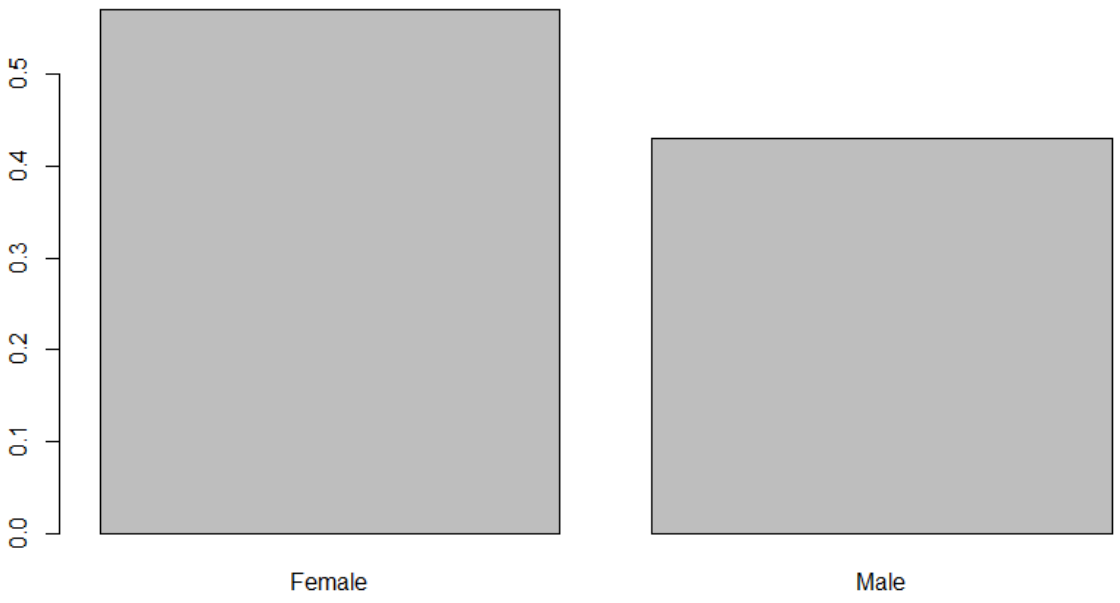
usr_r:

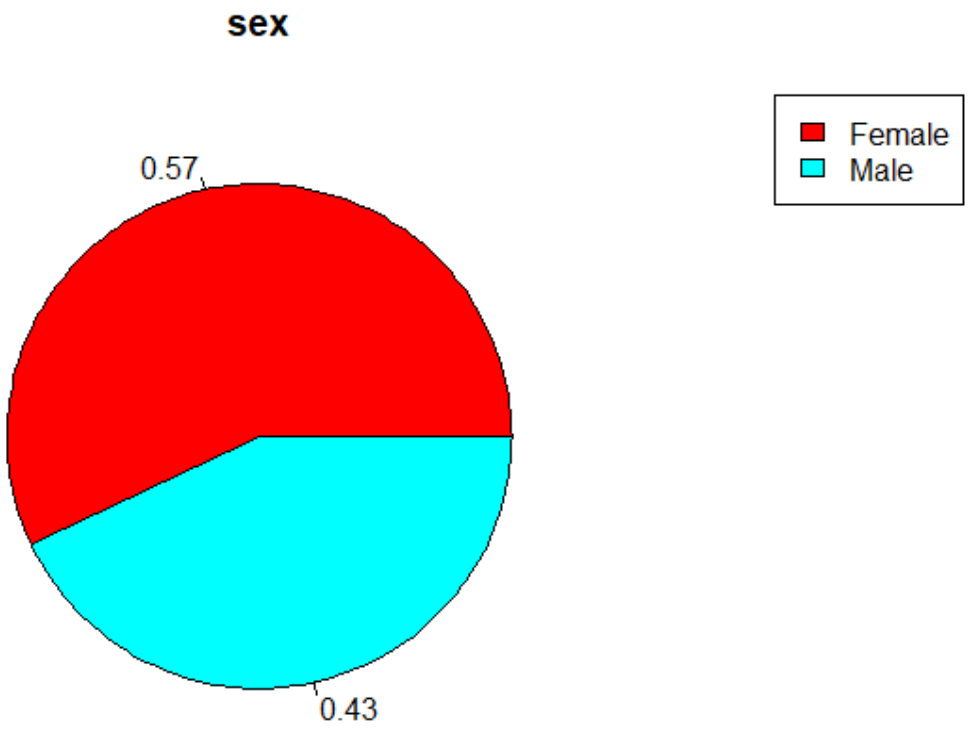




Το 52% των δειγμάτων μένουν σε προάστιες περιοχές, το 27% σε αστικές ενώ το 21% σε αγροτικές περιοχές.

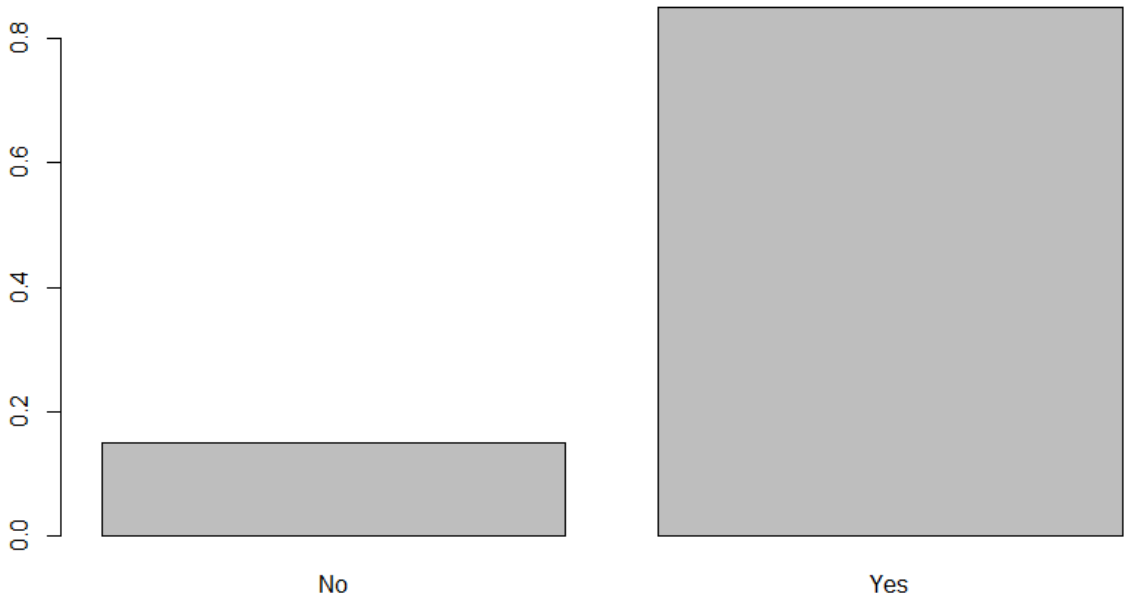
sex:

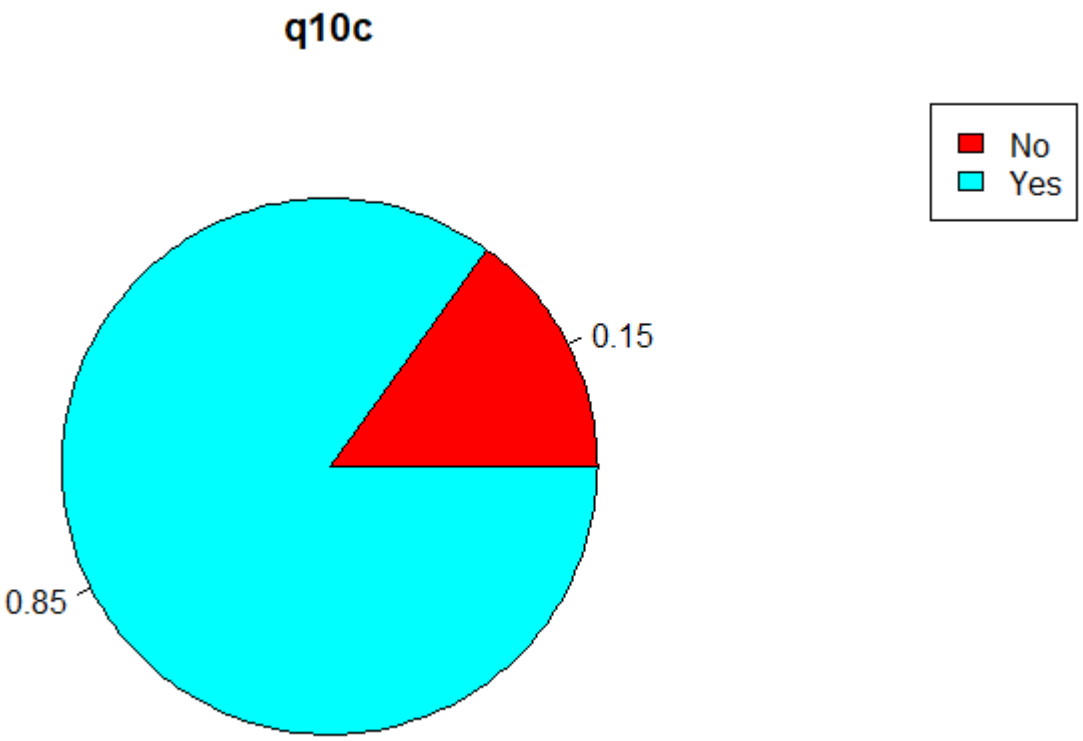




Το 57% των δειγμάτων είναι γυναίκες ενώ το 43% άντρες.

q10c:





Το 85% των δειγμάτων απάντησε ότι έχει κινητό τηλέφωνο ή κάποια συσκευή που λειτουργεί και σαν κινητό, ενώ το 15% ότι δεν έχει.

Επιπλέον, εξετάσαμε συσχετίσεις μεταξύ μεταβλητών (correlation) και συγκεκριμένα χρησιμοποιούμε τον συντελεστή συσχέτιση Spearman's μεταξύ των mobileprice και q20 και mobileprice και age. Τα αποτελέσματα φαίνονται παρακάτω.

mobileprice ~ q20:

```
> cor.test(mobileprice,q20,  method="spearman")

        Spearman's rank correlation rho

data:  mobileprice and q20
S = 192553493, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.2462713
```

Παρατηρούμε ότι το p-value < 0.05, επομένως υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μεταβλητών mobileprice και q20, υπάρχει δηλαδή συσχέτιση μεταξύ των δύο αυτών μεταβλητών.

mobileprice ~ age:

```
> cor.test(mobileprice,age,  method="spearman")

        Spearman's rank correlation rho

data:  mobileprice and age
S = 2487758055, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.3999075
```

Παρατηρούμε ότι το p-value < 0.05, επομένως υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μεταβλητών mobileprice και age, υπάρχει δηλαδή συσχέτιση μεταξύ των δύο αυτών μεταβλητών.

Στάδιο 2

Στο δεύτερο στάδιο, συνεχίσαμε με την κατασκευή μοντέλων για να ερμηνεύσουμε την επίδραση των ανεξάρτητων μεταβλητών στην εξαρτημένη mobileprice.

Το πρώτο μοντέλο που επιλέξαμε ήταν η παλινδρόμηση.

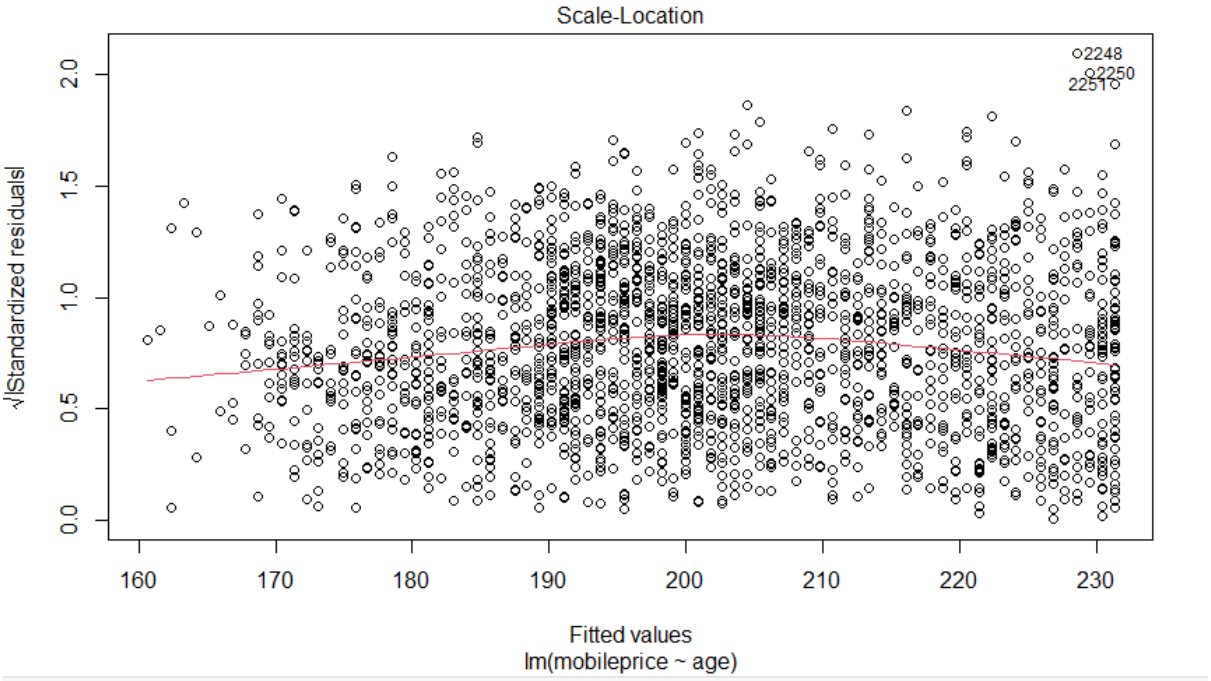
Γραμμική Παλινδρόμηση
Δοκιμάστηκαν όλοι οι συνδυασμοί (mobileprice με q18 , q20 , q26) αλλά τα καλύτερα αποτελέσματα τα έδωσε ο συνδυασμός mobileprice και age. Δηλαδή τηρούσε τις προυποθέσεις ότι το p-value<0.05 και Pr<0.05 (ότι το μοντέλο συνεισφέρει σημαντικά στην ερμηνεία της εξαρτημένης μεταβλητής) και είχε το μεγαλύτερο R-squared που σημαίνει ότι αυτό το μοντέλο ερμηνεύει καλύτερα την εξαρτημένη μεταβλητή (παρότι η τιμή του R-squared είναι μικρή).
Επομένως, επιλέχθηκε η γραμμική παλινδρόμηση για τις μεταβλητές mobileprice και age. Τα αποτελέσματα ήταν τα παρακάτω:

```
Call:
lm(formula = mobileprice ~ age)

Residuals:
     Min       1Q   Median       3Q      Max
-180.791  -25.193   -0.038   24.715  131.854

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 247.36536    2.57691   95.99   <2e-16 ***
age          -0.89404    0.04754  -18.81   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41.09 on 2199 degrees of freedom
  (51 observations deleted due to missingness)
Multiple R-squared:  0.1385,    Adjusted R-squared:  0.1382
F-statistic: 353.7 on 1 and 2199 DF,  p-value: < 2.2e-16
```



Πολυωνυμική παλινδρόμηση

Ακολούθησε η δημιουργία ενός πολυωνυμικού μοντέλου του mobileprice σε συνάρτηση με το age + age^2. Από τα αποτελέσματα φαίνεται ότι ο τετραγωνικός όρος λόγω το ότι το Pr = 0.45028<0.05 δεν συνεισφέρει σημαντικά στην ερμηνεία της εξαρτημένης μεταβλητής. Επομένως το γραμμικό μοντέλο είναι προτιμότερο από το πολυωνυμικό. Ενώ από το πολυωνυμικό μοντέλο του mobileprice ~ q18 + $q18^2$ συνεισφέρει σημαντικά στην ερμηνεία της εξαρτημένης μεταβλητής γιατί όλοι οι όροι του συνεισφέρουν σημαντικά (Pr < 0.05) όπως επίσης και το p-value< 2.2e-16<0.05. Εδώ να σημειωθεί ότι παρόλου που το γραμμικό μοντέλο mobileprice ~ q18 δεν ερμηνεύει καλά την εξαρτημένη μεταβλητή το πολυωνυμικό μοντέλο την ερμηνεύει. Βέβαια ακόμα και σε αυτή την περίπτωση το multiple R-squared παραμένει πολύ μικρό.

```
Call:
lm(formula = mobileprice ~ age + age2)

Residuals:
     Min       1Q   Median       3Q      Max
-179.793  -25.222    0.385   24.825  131.279

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 243.341834   5.919070  41.112  < 2e-16 ***
age          -0.712620   0.244917  -2.910  0.00365 **
age2         -0.001778   0.002355  -0.755  0.45028
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41.1 on 2198 degrees of freedom
  (51 observations deleted due to missingness)
Multiple R-squared:  0.1388,    Adjusted R-squared:  0.138
F-statistic: 177.1 on 2 and 2198 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = mobileprice ~ q18 + q18_2)

Residuals:
     Min       1Q   Median       3Q      Max
-168.714  -27.208    1.717   28.490  135.304

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.992e+02  1.190e+00 167.425   <2e-16 ***
q18          9.012e-01  7.464e-02  12.075   <2e-16 ***
q18_2       -1.799e-03  2.072e-04  -8.683   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42.39 on 1843 degrees of freedom
  (406 observations deleted due to missingness)
Multiple R-squared:  0.0761,    Adjusted R-squared:  0.0751
F-statistic:  75.9 on 2 and 1843 DF,  p-value: < 2.2e-16
```

Εκθετική παλινδρόμηση

Εφαρμόστηκε εκθετική παλινδρόμηση mobileprice-age,mobileprice-q18,mobileprice-q20, mobileprice-q26. Διαπιστώθηκε ότι τα τρία πρώτα μοντέλα συνεισφέρουν σημαντικά στην ερμηνεία της ανεξάρτητης μεταβλητής. Παρότι η το γραμμικό μοντέλο του mobileprice-q18 δεν ερμηνεύει καλά την εξαρτημένη μεταβλητή το εκθετικό της κάνει καλύτερη ερμηνεία. Βέβαια το multiple R-squared παραμένει πολύ μικρό και για τα τρία εκθετικά μοντέλα. Επιπλέον ένα θετικό με τα εκθετικά μοντέλα είναι ότι τα residuals είναι πολύ μικρά σε σχέση τα αντίστοιχα γραμμικά και πολυωνυμικά μοντέλα. Όσο αφορά το μοντέλο mobileprice-q26 δεν ερμηνεύει καλά την εξαρτημένη γιατί το p-value = 0.1143>0.05.

```
Call:
lm(formula = log(mobileprice) ~ age)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5481 -0.1145  0.0221  0.1404  0.5287

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.5089178  0.0137956  399.32   <2e-16 ***
age         -0.0044672  0.0002545  -17.55   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.22 on 2199 degrees of freedom
  (51 observations deleted due to missingness)
Multiple R-squared:  0.1229,    Adjusted R-squared:  0.1225
F-statistic: 308.1 on 1 and 2199 DF,  p-value: < 2.2e-16

> |
```

```
Call:
lm(formula = log(mobileprice[x]) ~ q18[x])

Residuals:
     Min       1Q   Median       3Q      Max
-1.45955 -0.12125  0.03576  0.15571  0.52473

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.2905418  0.0058244 908.335  < 2e-16 ***
q18[x]      0.0018017  0.0002242   8.037 1.62e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2293 on 1844 degrees of freedom
Multiple R-squared:  0.03385,   Adjusted R-squared:  0.03332
F-statistic:  64.6 on 1 and 1844 DF,  p-value: 1.622e-15

> |
```

Γενικευμένα προσθετικά μοντέλα (gam)

Τα γενικευμένα προσθετικά μοντέλα που εκτελέστηκαν είναι mobileprice~s(age) ,mobileprice ~ s(q18). Και τα δύο συνεισφέρουν σημαντικά στην ερμηνεία της εξαρτημένης μεταβλητής γιατί τα p-value <0.05 και Pr<0.05 (στα παρακάτω σχήματα φαίνεται). Βέβαια το multiple R-squared παραμένει πολύ μικρό ενώ παίζει πολύ σημαντικό ρόλο στο πόσο ερμηνεύει καλά το μοντέλο την εξαρτημένη μεταβλητή. Παρατηρείται ότι το multiple R-squared των γενικευμένων προσθετικών μοντέλων είναι μεγαλύτερο από όλα τα προηγούμενα μοντέλα που ερευνήθηκαν (γραμμικά,πολυωνυμικά,εκθετικά).

```
Family: gaussian
Link function: identity

Formula:
mobileprice ~ s(age)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 201.7899     0.8754   230.5   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
         edf Ref.df     F p-value
s(age) 4.029   4.99 71.71  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.139   Deviance explained = 14.1%
GCV = 1690.4  Scale est. = 1686.5    n = 2201
```

```
Family: gaussian
Link function: identity

Formula:
mobileprice[x] ~ s(q18[x])

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 207.3746     0.9433   219.8   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
            edf Ref.df     F p-value
s(q18[x]) 8.737  8.955 38.41  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.154   Deviance explained = 15.8%
GCV = 1651.5  Scale est. = 1642.8    n = 1846
> |
```

Ανάλυση Διακύμανσης

Μετέπειτα, υλοποιήσαμε ανάλυση διακύμανσης για να βρούμε πιθανή επίδραση κάποιας ανεξάρτητης κατηγορικής μεταβλητής στην εξαρτημένη. Έτσι, υλοποιήσαμε τη συνάρτηση ANOVA μεταξύ της εξαρτημένης μεταβλητής mobileprice με μία ανεξάρτητη, με πολλαπλές ανεξάρτητες και με παράγοντες. Στα αποτελέσματα φαίνεται ότι η anova συνεισφέρει σημαντικά στην ερμηνεία της εξαρτημένης μεταβλητής (αυτό ισχύει γιατί p-value<0.05 & τα Pr <0.05). Τέλος, με τη βοήθεια post hoc τεστ εξετάσαμε και την επιρροή των μεταβλητών όταν είχαμε μοντέλα με παράγοντες. Τα αποτελέσματα ήταν τα παρακάτω:

mobileprice ~ sex:

```
> #mobile price & sex
> sex = deframe(Cell_Phones_labels[3][,1])
> mobileprice = deframe(Cell_Phones_labels[35][,1])
> summary(aov(mobileprice~sex))
              Df  Sum Sq Mean Sq F value   Pr(>F)
sex            1   63261   63261   32.72 1.21e-08 ***
Residuals   2250 4350078    1933
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary.lm(model_an)

Call:
aov(formula = mobileprice ~ sex)

Residuals:
     Min       1Q   Median       3Q      Max
-159.555  -29.857   -0.509   28.537  140.537

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  196.663      1.232  159.58  < 2e-16 ***
sexMale       10.692      1.869    5.72 1.21e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.97 on 2250 degrees of freedom
Multiple R-squared:  0.01433,   Adjusted R-squared:  0.0139
F-statistic: 32.72 on 1 and 2250 DF,  p-value: 1.206e-08
```

Τα αποτελέσματα δείχνουν ότι $P<0.05$, επομένως υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μεταβλητών, δηλαδή υπάρχει εξάρτηση μεταξύ της mobileprice και sex.

mobileprice ~inc:

```
> inc = deframe(Cell_Phones_labels[34][,1])
> mobileprice = deframe(Cell_Phones_labels[35][,1])
> model_an<-(aov(mobileprice~inc))
> summary(model_an)
              Df  Sum Sq Mean Sq F value   Pr(>F)
inc            8   75118    9390   4.796 7.43e-06 ***
Residuals   1835 3592801    1958
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
408 observations deleted due to missingness
> summary.lm(model_an)

Call:
aov(formula = mobileprice ~ inc)

Residuals:
     Min       1Q   Median       3Q      Max
-156.839  -29.494    1.683   27.181  148.210

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   201.8951     3.2799  61.555  < 2e-16 ***
inc$100,000 to under $150,000   8.0647     4.6579   1.731 0.083547 .
inc$150,000 or more            17.1435     5.0260   3.411 0.000661 ***
inc$20,000 to under $30,000    -4.8051     4.3941  -1.094 0.274299
inc$30,000 to under $40,000    -0.5838     4.5117  -0.129 0.897061
inc$40,000 to under $50,000     1.6410     4.5494   0.361 0.718359
inc$50,000 to under $75,000     0.7353     4.0789   0.180 0.856966
inc$75,000 to under $100,000    6.7850     4.3651   1.554 0.120267
incLess than $10,000           -9.2897     4.8886  -1.900 0.057553 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 44.25 on 1835 degrees of freedom
  (408 observations deleted due to missingness)
Multiple R-squared:  0.02048,   Adjusted R-squared:  0.01621
F-statistic: 4.796 on 8 and 1835 DF,  p-value: 7.425e-06
```

Τα αποτελέσματα δείχνουν από το δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των δύο μεταβλητών, δηλαδή υπάρχει δεν υπάρχει εξάρτηση μεταξύ της mobileprice και inc, παρόλο που εύκολα θα υποθέταμε ότι η τιμή του τηλεφώνου που αγοράζει κάποιος επηρεάζεται από το εισόδημά του.

```
Coefficients: (5 not defined because of singularities)
                                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                212.322     13.872  15.306  < 2e-16 ***
emplEmployed full-time                                      -2.404     14.460  -0.166  0.86797
emplEmployed part-time                                     -19.506     16.266  -1.199  0.23060
emplHave own business/self-employed                        -16.544     19.618  -0.843  0.39913
emplNot employed for pay                                    -7.308     15.946  -0.458  0.64679
emplRetired                                                -28.926     15.142  -1.910  0.05622 .
emplStudent                                                 61.567     33.979   1.812  0.07014 .
marLiving with a partner                                   -30.337     20.972  -1.447  0.14818
marMarried                                                 -16.614     16.266  -1.021  0.30719
marNever been married                                      -48.762     23.212  -2.101  0.03578 *
marSeparated                                               -31.922     27.744  -1.151  0.25002
marSingle                                                  -60.289     27.744  -2.173  0.02988 *
marWidowed                                                 -33.332     19.121  -1.743  0.08144 .
emplEmployed full-time:marLiving with a partner             53.953     21.990   2.454  0.01422 *
emplEmployed part-time:marLiving with a partner             57.670     24.466   2.357  0.01850 *
emplHave own business/self-employed:marLiving with a partner 62.459    34.778   1.796  0.07265 .
emplNot employed for pay:marLiving with a partner           45.284     24.169   1.874  0.06112 .
emplRetired:marLiving with a partner                        49.041     25.867   1.896  0.05811 .
emplStudent:marLiving with a partner                           NA         NA      NA       NA
emplEmployed full-time:marMarried                           14.932     16.862   0.886  0.37597

emplEmployed part-time:marMarried                           12.529     18.737   0.669  0.50378
emplHave own business/self-employed:marMarried              25.166     22.538   1.117  0.26429
emplNot employed for pay:marMarried                         11.707     18.404   0.636  0.52477
emplRetired:marMarried                                      15.947     17.523   0.910  0.36290
emplStudent:marMarried                                     -32.150     40.739  -0.789  0.43010
emplEmployed full-time:marNever been married                60.435     23.824   2.537  0.01126 *
emplEmployed part-time:marNever been married                77.255     25.193   3.067  0.00219 **
emplHave own business/self-employed:marNever been married   82.868     31.935   2.595  0.00953 **
emplNot employed for pay:marNever been married              62.408     24.859   2.510  0.01213 *
emplRetired:marNever been married                           40.109     25.343   1.583  0.11365
emplStudent:marNever been married                          -11.327     42.303  -0.268  0.78892
emplEmployed full-time:marSeparated                         46.016     29.910   1.539  0.12407
emplEmployed part-time:marSeparated                         45.886     34.471   1.331  0.18328
emplHave own business/self-employed:marSeparated               NA         NA      NA       NA
emplNot employed for pay:marSeparated                       23.394     30.907   0.757  0.44919
emplRetired:marSeparated                                    57.316     31.301   1.831  0.06722 .
emplStudent:marSeparated                                   -32.167     58.853  -0.547  0.58474
emplEmployed full-time:marSingle                            66.866     29.301   2.282  0.02258 *
emplEmployed part-time:marSingle                            82.456     33.623   2.452  0.01427 *
emplHave own business/self-employed:marSingle                  NA         NA      NA       NA
emplNot employed for pay:marSingle                          57.160     30.907   1.849  0.06453 .
emplRetired:marSingle                                       58.893     35.207   1.673  0.09451 .
emplStudent:marSingle                                          NA         NA      NA       NA

emplEmployed full-time:marWidowed                           23.103     21.874   1.056  0.29101
emplEmployed part-time:marWidowed                           35.657     24.127   1.478  0.13957
emplHave own business/self-employed:marWidowed               8.654     37.735   0.229  0.81862
emplNot employed for pay:marWidowed                         18.322     22.352   0.820  0.41248
emplRetired:marWidowed                                      24.394     20.342   1.199  0.23057
emplStudent:marWidowed                                         NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41.62 on 2174 degrees of freedom
  (34 observations deleted due to missingness)
Multiple R-squared:  0.1303,    Adjusted R-squared:  0.1131
F-statistic: 7.577 on 43 and 2174 DF,  p-value: < 2.2e-16
```

Σύμφωνα με τα παραπάνω αποτελέσματα, στατιστικά σημαντική διαφορά παρουσιάζεται μόνο στους συνδυασμούς όπου το $p<0.05$ (γραμμές με αστερίσκο ή αστερίσκους).

mobileprice ~ usr_r*sex

```
> #mobilprice & usr_r*sex
> usr_r = deframe(Cell_Phones_labels[2][,1])
> sex = deframe(Cell_Phones_labels[3][,1])
> mobileprice = deframe(Cell_Phones_labels[35][,1])
> model <-(aov(mobileprice~usr_r*sex))
> summary.lm(model)

Call:
aov(formula = mobileprice ~ usr_r * sex)

Residuals:
     Min       1Q   Median       3Q      Max
-160.377  -29.430   -0.267   28.322  149.384

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           187.816      2.603  72.163  < 2e-16 ***
usr_rSuburban          10.252      3.127   3.279  0.00106 **
usr_rUrban             11.415      3.579   3.190  0.00144 **
sexMale                 7.575      4.198   1.805  0.07128 .
usr_rSuburban:sexMale   2.535      4.954   0.512  0.60893
usr_rUrban:sexMale      5.473      5.529   0.990  0.32235
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.94 on 2191 degrees of freedom
  (55 observations deleted due to missingness)
Multiple R-squared:  0.02812,   Adjusted R-squared:  0.0259
F-statistic: 12.68 on 5 and 2191 DF,  p-value: 3.638e-12
```

Σύμφωνα με τα παραπάνω αποτελέσματα, η mobileprice δεν παρουσιάζει εξάρτηση με τις usr_r*sex.

Για τα posthoc tests χρησιμοποιήσαμε το hsd που σημαίνει το Tukey HSD (Honestly Significant Difference) που είναι η πιο συνηθισμένη μέθοδος.

mobileprice ~ usr_r με παράγοντες

```
> PostHocTest(model,method="hsd")

  Posthoc multiple comparisons of means : Tukey HSD
    95% family-wise confidence level

$usr_r
                    diff    lwr.ci   upr.ci    pval
Suburban-Rural 11.721830  6.003026 17.44063 4.9e-06 ***
Urban-Rural    14.579818  8.162352 20.99728 3.3e-07 ***
Urban-Suburban  2.857988 -2.379444  8.09542  0.4067

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Στατιστικά σημαντική διαφορά υπάρχει μεταξύ των μεταβλητών suburban και rular και urban και rural.

mobileprice ~ q17c με παράγοντες

```
> #anova mobileprice & usr_r με παραγοντες
> empl = deframe(Cell_Phones_labels[33][,1])
> mobileprice = deframe(Cell_Phones_labels[35][,1])
> model <-(aov(mobileprice~empl))
> library(DescTools)
> PostHocTest(model,method="hsd")

  Posthoc multiple comparisons of means : Tukey HSD
    95% family-wise confidence level

$empl
                                                        diff      lwr.ci     upr.ci    pval
Employed full-time-Disabled                         24.10412911   7.5710966  40.637162 0.00035 ***
Employed part-time-Disabled                         13.56230136  -4.2290047  31.353607 0.26952
Have own business/self-employed-Disabled            17.73362146  -5.5284185  40.995661 0.26945
Not employed for pay-Disabled                       17.76306765   0.3981719  35.127963 0.04106 *
Retired-Disabled                                    -7.68480386 -24.5523511   9.182743 0.83080
Student-Disabled                                    28.39894614  -8.6748248  65.472717 0.26391
Employed part-time-Employed full-time              -10.54182775 -19.3018690  -1.781786 0.00716 **
Have own business/self-employed-Employed full-time  -6.37050765 -23.7293745  10.988359 0.93321
Not employed for pay-Employed full-time             -6.34106146 -14.1990101   1.516887 0.20688
Retired-Employed full-time                         -31.78893297 -38.4763766 -25.101489 5.8e-11 ***
Student-Employed full-time                           4.29481703 -29.3900794  37.979713 0.99978
Have own business/self-employed-Employed part-time   4.17132010 -14.3899209  22.732561 0.99449
Not employed for pay-Employed part-time              4.20076628  -6.0430985  14.444631 0.89045
Retired-Employed part-time                         -21.24710523 -30.6232084 -11.871002 6.6e-10 ***
Student-Employed part-time                          14.83664477 -19.4833443  49.156634 0.86315
Not employed for pay-Have own business/self-employed  0.02944619 -18.1234792  18.182372 1.00000
Retired-Have own business/self-employed            -25.41842532 -43.0961866  -7.740664 0.00046 ***
Student-Have own business/self-employed             10.66532468 -26.7840223  48.114672 0.98067
Retired-Not employed for pay                       -25.44787151 -33.9872125 -16.908531 5.8e-11 ***
Student-Not employed for pay                        10.63587849 -23.4650113  44.736768 0.96939
Student-Retired                                     36.08375000   2.2334142  69.934086 0.02792 *

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Σύμφωνα με τα παραπάνω αποτελέσματα, στατιστικά σημαντική διαφορά παρουσιάζεται μόνο στους συνδυασμούς όπου το $p<0.05$ (γραμμές με αστερίσκο ή αστερίσκους).

mobileprice ~ educ με παράγοντες

```
> educ = deframe(Cell_Phones_labels[31][,1])
> mobileprice = deframe(Cell_Phones_labels[35][,1])
> model <-(aov(mobileprice~educ))
> library(DescTools)
> PostHocTest(model,method="hsd")

  Posthoc multiple comparisons of means : Tukey HSD
    95% family-wise confidence level

$educ
                                              diff     lwr.ci     upr.ci    pval
Living with a partner-Divorced           23.053678   8.876447  37.230908 3.5e-05 ***
Married-Divorced                         -2.284849 -11.513736   6.944038 0.99072
Never been married-Divorced              15.649939   4.881989  26.417888 0.00037 ***
Separated-Divorced                        7.422484 -12.380645  27.225614 0.92643
Single-Divorced                           5.547077 -13.647587  24.741741 0.97916
Widowed-Divorced                        -21.370741 -33.369630  -9.371852 3.4e-06 ***
Married-Living with a partner           -25.338527 -37.276551 -13.400502 9.5e-09 ***
Never been married-Living with a partner  -7.403739 -20.567812   5.760334 0.64315
Separated-Living with a partner         -15.631194 -36.832798   5.570411 0.30917
Single-Living with a partner            -17.506601 -38.141019   3.127817 0.15830
Widowed-Living with a partner           -44.424419 -58.613055 -30.235782 5.6e-11 ***
Never been married-Married               17.934788  10.353805  25.515770 1.4e-10 ***
Separated-Married                         9.707333  -8.559755  27.974421 0.70279
Single-Married                            7.831926  -9.773689  25.437541 0.84618
Widowed-Married                         -19.085892 -28.332291  -9.839493 2.8e-08 ***
Separated-Never been married             -8.227455 -27.318365  10.863456 0.86490
Single-Never been married               -10.102862 -28.561845   8.356121 0.67256
Widowed-Never been married              -37.020680 -47.803642 -26.237718 5.6e-11 ***
Single-Separated                         -1.875407 -26.713871  22.963056 0.99999
Widowed-Separated                       -28.793225 -48.604522  -8.981928 0.00037 ***
Widowed-Single                          -26.917818 -46.120908  -7.714728 0.00072 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Στα παραπάνω αποτελέσματα, στατιστικά σημαντική διαφορά παρουσιάζεται μόνο στους συνδυασμούς όπου το p<0.05 (γραμμές με αστερίσκους).

Σύνθετα Μοντέλα

Το πρώτο σύνθετο μοντέλο που διενεργήθηκε είναι το γραμμικό μοντέλο μέχρι τέταρτης τάξης (όλες οι μεταβλητές, τα τετράγωνά τους, αλληλεπιδράσεις 2ης τάξης,3ης και 4ης τάξης). Στα αποτελέσματα φαίνεται ότι όλες οι αλληλεπιδράσεις δεν είναι στατιστικά σημαντικές για την ερμηνεία της εξαρτημένης μεταβλητής. Το ίδιο ισχύει και για τους όρους q18 και q26. Ενώ ο όρος q20 είναι μικρότερος από το 0.05 άρα θεωρείται στατιστικά σημαντικός για το μοντέλο. Επίσης το p-value>0.05 άρα δεν ερμηνεύει σημαντικά την εξαρτημένη. Επομένως αυτό το μοντέλο δεν είναι ικανοποιητικό.

```
     Min       1Q   Median       3Q      Max
-191.653  -15.850   -2.374   18.569  104.423

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.677e+02  1.702e+01  15.727   <2e-16 ***
age             -1.368e+00  7.379e-01  -1.854   0.0643 .
q18             -1.641e-01  4.242e-01  -0.387   0.6990
q20             -2.946e-01  1.334e-01  -2.209   0.0277 *
q26             -1.352e-01  4.456e-01  -0.303   0.7617
I(age^2)         1.006e-02  7.921e-03   1.270   0.2046
I(q18^2)        -3.978e-05  5.216e-04  -0.076   0.9392
I(q20^2)         2.062e-04  1.490e-04   1.384   0.1671
I(q26^2)        -5.832e-04  1.489e-03  -0.392   0.6954
age:q18          1.057e-02  1.204e-02   0.879   0.3801
age:q20          7.926e-03  4.717e-03   1.680   0.0936 .
q18:q20          1.814e-03  2.307e-03   0.786   0.4321
age:q26          1.101e-02  1.105e-02   0.996   0.3196
q18:q26          8.235e-03  1.387e-02   0.594   0.5530
q20:q26          6.578e-03  5.049e-03   1.303   0.1933
age:q18:q20     -7.524e-05  8.556e-05  -0.879   0.3797
age:q18:q26     -3.407e-04  4.319e-04  -0.789   0.4305
age:q20:q26     -2.598e-04  1.926e-04  -1.349   0.1781
q18:q20:q26     -9.281e-05  9.376e-05  -0.990   0.3228
age:q18:q20:q26  3.431e-06  3.257e-06   1.053   0.2927
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.86 on 462 degrees of freedom
  (1770 observations deleted due to missingness)
Multiple R-squared:  0.03073,   Adjusted R-squared:  -0.009133
F-statistic: 0.7709 on 19 and 462 DF,  p-value: 0.7425
```

Ένα άλλο σύνθετο μοντέλο που χρησιμοποιήθηκε με σκοπό να αυξηθεί το multiple R-squared είναι το mobileprice~*log(q18)*log(age). Βέβαια λόγω το ότι η μεταβλητή q18 έχει μηδενικά σε κάποιες θέσεις το log(0) δεν ορίζεται για αυτό αυτές οι τιμές τις θεωρούμε ότι είναι μηδέν. Αυτό έγινε λόγω το ότι η μεταβλητή q18 δεν παίρνει καθόλου αρνητικές τιμές και προφανώς δεν έχουν νόημα για αυτήν. Εφόσον η κατώτερη τιμή η μεταβλητή q18 που μπορεί να πάρει είναι μηδεν για αυτό θεωρούμε ότι στις θέσεις που έχει μηδενικά και λογαριθμίζεται (γίνεται -inf) η τιμή που παίρνει είναι η κατώτερη που είναι το μηδέν. Τα αποτελέσματα φαίνονται στο παρακάτω σχήμα. Εδώ παρατηρείται ότι όλοι οι όροι του μοντέλου επηρεάζουν σημαντικά την εξαρτημένη μεταβλητή mobileprice όπως επίσης η τιμή p-value <0.05 επομένως το μοντέλο συνεισφέρει σημαντικά στην ερμηνεία της mobileprice. Βέβαια παρότι το multiple R-squared είναι μεγαλύτερο σε σχέση με τα υπόλοιπα μοντέλα που αναλύθηκαν παραπάνω παραμένει μικρή σαν τιμή. Επίσης τα Residuals έχουν μεγάλες απόλυτες τιμές (δεν είναι κοντά στο μηδέν) αυτό ισχύει και στα προηγούμενα μοντέλα. Μπορεί να διορθωθεί ένα λογαριθμίσουμε το mobileprice.

```
Call:
lm(formula = mobileprice ~ log_q18 * log(age))

Residuals:
     Min       1Q   Median       3Q      Max
-188.873  -20.905   -0.699   23.463  135.426

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      328.068     15.967  20.547  < 2e-16 ***
log_q18          -17.287      7.555  -2.288   0.0222 *
log(age)         -36.475      4.044  -9.020  < 2e-16 ***
log_q18:log(age)   7.833      1.990   3.937 8.58e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.14 on 1804 degrees of freedom
  (444 observations deleted due to missingness)
Multiple R-squared:  0.2082,    Adjusted R-squared:  0.2069
F-statistic: 158.1 on 3 and 1804 DF,  p-value: < 2.2e-16
```

Τέλος δημιουργήθηκε ένα μοντέλο το οποίο έχει ως όρο την μεταβλητή inc στην οποία έγινε σύμπτυξη κάποιων επιπέδον. Αρχικά η μεταβλητή inc επηρεάζει σημαντικά την μεταβλητή αφου p-value<0.05(αυτό φαίνεται στο παρακάτω σχήμα). Επίσης η σύμπτυξη που έγινε είναι ότι υπάρχουν μόνο τρεις κατηγορίες αντι για εννιά οι οποίες είναι : “Less than $30,000”,”$30,000 to under $75,000”,”$75,000 or more”. Επίσης μετα την σύμπτυξη παραμένει η μεταβλητή inc να επηρεάζει σημαντικά την ερμηνεία της εξαρτημένης μεταβλητής (αφού Pr<0.05). Το μοντέλο που εξετάστηκε είναι το mobileprice~inc :age+log(age)+q18+age+inc. Τα αποτελέσματα φαίνονται παρακάτω.

```
> summary(model1)
             Df  Sum Sq Mean Sq F value   Pr(>F)
inc_factor    8   75118    9390   4.796 7.43e-06 ***
Residuals  1835 3592801    1958
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
408 observations deleted due to missingness
```

```
               Df  Sum Sq Mean Sq F value   Pr(>F)
inc_factor_2    2   57527   28764   14.67 4.79e-07 ***
Residuals    1841 3610392    1961
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
408 observations deleted due to missingness
> model1 <-aov(mobileprice~inc_factor)
```

```
Call:
lm(formula = mobileprice ~ inc_factor_2:age + log(age) + q18 +
    age + inc_factor_2)

Residuals:
     Min       1Q   Median       3Q      Max
-182.901  -22.928   -0.251   25.538  130.167

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        248.40086   36.03702   6.893 8.00e-12 ***
log(age)            -1.74012   13.29442  -0.131  0.89588
q18                  0.21277    0.04335   4.908 1.02e-06 ***
age                 -0.70312    0.31073  -2.263  0.02379 *
inc_factor_2.L      -3.06900    5.86905  -0.523  0.60111
inc_factor_2.Q     -11.66546    5.24536  -2.224  0.02630 *
inc_factor_2.L:age   0.13015    0.11528   1.129  0.25909
inc_factor_2.Q:age   0.30709    0.10293   2.984  0.00289 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41 on 1513 degrees of freedom
  (731 observations deleted due to missingness)
Multiple R-squared:  0.1324,    Adjusted R-squared:  0.1284
F-statistic: 32.99 on 7 and 1513 DF,  p-value: < 2.2e-16

> |
```

Οι όροι inc_factor_2.L , log(age) και inc_factor_2.L:age δεν συνεισφέρουν σημαντικά στην ερμηνεία του μοντέλου.