Απόδειξη 1

Θέλουμε να αποδείξουμε ότι $R^2 = r^2_{xy}$

Γενικά από θεωρία έχουμε ότι:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{SSR}{SST}$$

Επίσης, είναι γνωστό ότι:

i) $SSR = \hat{B}_1^2 \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$, όπου $\hat{B}_1^2 = \frac{S^2_{xy}}{S^2_{xx}}$

και $\sum_{i=1}^{n} (x_i - \bar{x}) = S_{xx}$

Συνεπώς, $SSR = \frac{S^2_{xy} \cdot S_{xx}}{S^2_{xx}} = \frac{S^2_{xy}}{S_{xx}}$

ii) $SST = \sum_{i=1}^{n} (y_i - \bar{y}) = S_{yy}$

Από (i) κ' (ii):

$$R^2 = \frac{SSR}{SST} = \frac{S^2_{xy}}{S_{xx} \cdot S_{yy}}$$

Από θεωρία έχουμε ότι: $r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$

Άρα: $r^2_{xy} = \frac{S^2_{xy}}{S_{xx} \cdot S_{yy}}$

Μόλις αποδείξαμε ότι: $R^2 = r^2_{xy}$

Απόδειξη 2

Θέλουμε να αποδείξουμε ότι:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

Γενικά έχουμε ότι: $\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i$ κ'

$$\sum_{i=1}^{n} \hat{y}_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

Για να αποδείξουμε το ζητούμενο της άσκησης, ξεκινάμε από την μέθοδο των ελαχίστων τετραγώνων που έγκειται στην ελαχιστοποίηση της παράστασης:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - E(y_i))^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2,$$

ως προς τα β_0 και β_1, δηλαδή στην ελαχιστοποίηση του αθροίσματος των τετραγώνων και αποκλίσεων μεταξύ των y_i και $\hat{y}_i$
Ακολουθώντας τη συνηθισμένη διαδικασία για τον εντοπισμό των ελαχίστων τιμών, δηλαδή παραγωγίζοντας το $S(\beta_0, \beta_1)$ ως προς β_0 και ως προς β_1 και θέτοντας στη συνέχεια τις μερικές αυτές παραγώγους ίσες με μηδέν σε ένα σημείο $(\hat{\beta}_0, \hat{\beta}_1)$, καταλήγουμε στις εξής εξισώσεις:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad \text{κ'} \quad \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Απόδειξη:

Μερική παράγωγος του C ως προς Bo

$C = \left(\sum_{i=1}^{n}(y - B_0 - B_1 x)^2\right)$, γιατί $E(y) = B_0 + B_1 x$

$$\frac{\partial C}{\partial B_0} = 0 \Rightarrow \left(\sum_{i=1}^{n}(y - B_0 - B_1 x)^2\right)' = 0 \Rightarrow$$

$$2\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} B_0 - \sum_{i=1}^{n} B_1 x_i\right)\left(\sum_{i=1}^{n}(y - B_0 - B_1 x)\right)' = 0$$

$2\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} B_0 - \sum_{i=1}^{n} B_1 x_i\right)(-1)$, διότι παραγωγίζουμε ως προς Bo

$$-2\sum_{i=1}^{n} y_i + 2\sum_{i=1}^{n} B_0 + 2\sum_{i=1}^{n} B_1 x_i = 0$$

$$\sum_{i=1}^{n} y_i = nB_0 + \cancel{2}B_1 \sum_{i=1}^{n} x_i$$

Τώρα, όπως ανέφερα και πριν αντικαθιστώ με το σημείο $(\hat{B}_0, \hat{B}_1)$

$$\sum_{i=1}^{n} y_i = n\hat{B}_0 + \hat{B}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \hat{y}_i$$

Συνεπώς, μόλις αποδείξαμε το ζητούμενο της άσκησης. Αποδείξαμε δηλαδή ότι:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

Ακολουθώ την ίδια διαδικασία για το B_1

$$\frac{\partial C}{\partial B_1} = \left(\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)^2\right)' = 0$$

$$2\left(\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)\right)\left(\sum_{i=1}^{n}(y_i - B_0 - B_1 x_i)\right)' = 0$$

$$\left(2\sum_{i=1}^{n} y_i - 2\sum_{i=1}^{n} B_0 - 2\sum_{i=1}^{n} B_1 x_i\right)\left(-\sum x_i\right) = 0$$

$$-2\sum_{i=1}^{n} x_i y_i + 2\sum_{i=1}^{n} B_0 x_i + 2\sum_{i=1}^{n} B_1 x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i = B_0 \sum_{i=1} x_i + \hat{B}_1 \sum_{i=1} x_i^2$$

Αντικαθιστώντας στο σημείο $(\hat{B}_0, \hat{B}_1)$ έχουμε:

$$\sum_{i=1}^{n} x_i y_i = \hat{B}_0 \sum_{i=1}^{n} x_i + \hat{B}_1 \sum_{i=1}^{n} x_i^2$$

Απόδειξη 3

Θέλουμε να δείξουμε ότι:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$$

Από θεωρία έχουμε ότι:

(i) $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ κ' (ii) $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 x_i$

Λύση: (1^{ος} τρόπος - Αυτόν βαθμολογήστε)

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) =$$

$$\sum_{i=1}^{n} (y_i \hat{y}_i - y_i \bar{y} - (\hat{y}_i)^2 + \hat{y}_i \bar{y}) =$$

$$\overset{\text{Από (i)}}{\Longrightarrow} \sum_{i=1}^{n} (y_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - y_i \bar{y} - (\hat{\beta}_0 + \hat{\beta}_1 x_i)^2 + (\hat{\beta}_0 + \hat{\beta}_1 x_i)\bar{y}) =$$

$$\overset{\text{Από (ii)}}{\Longrightarrow} \sum_{i=1}^{n} (y_i(\bar{y} - \hat{\beta}_1 x_i + \hat{\beta}_1 x_i) - y_i \bar{y} - (\bar{y} - \hat{\beta}_1 x_i + \hat{\beta}_1 x_i)^2$$

$$+ (\bar{y} - \hat{\beta}_1 x_i - \hat{\beta}_1 x_i)\bar{y}) =$$

$$\sum_{i=1}^{n} (y_i \bar{y} - y_i \bar{y} - (\bar{y})^2 + (\bar{y}) \cdot \bar{y}) =$$

$$\sum_{i=1}^{n} (y_i \bar{y} - y_i \bar{y} - \bar{y}^2 + \bar{y}^2) = \sum_{i=1}^{n} 0 = 0$$

Μόλις αποδείξαμε ότι: $\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$

Για την απόδειξη 3 σκέφτηκα και έναν ακόμη τρόπο, για τον οποίο όμως δεν είμαι σίγουρος και δεν ξέρω κατά πόσο αληθεύει

Αυτό που σκέφτηκα είναι το εξής:

Στην Απόδειξη 2 δείξατε ότι:

$$\sum_{i=1}^{n} y = \sum_{i=1}^{n} \hat{y}.$$

Συνεπώς ισχύει ότι: $\sum_{i=1}^{n} y - \sum_{i=1}^{n} y_i = 0 \Rightarrow \sum_{i=1}^{n} (y - y_i) = 0$

Συνεπώς το γινόμενο του $\sum_{i=1}^{n} (y - y_i)$ με οποιονδήποτε άλλο αριθμό θα δίνει αποτέλεσμα 0

Άρα: $\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} (y_i - \hat{y}_i) \sum_{i=1}^{n} (\hat{y}_i - \bar{y}) =$

$$= 0 \cdot \sum_{i=1}^{n} (\hat{y}_i - \bar{y}) = 0$$

Ωστόσο, αν πρέπει να δεχτείτε μόνο 1 απόδειξη, σας παρακαλώ πολύ λάβετε υπόψιν σας, τον 1ο τρόπο

Απόδειξη 4

Θέλουμε να αποδείξουμε ότι:

$$\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r^2_{xy}}}$$

Λύση:

Από Θεωρία έχουμε ότι $^{(i)}$ $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

Επίσης: $se(\hat{\beta}_1) = \sqrt{\hat{V}(\hat{\beta}_1)} = S\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{-1/2} = S \cdot S_{xx}^{-1/2}$

Συνεπώς: $se(\hat{\beta}_1) = \frac{S}{\sqrt{S_{xx}}}$, όπου S:

$$S = \sqrt{\frac{1}{n-2}S_{yy}(1-r^2_{xy})} = \frac{\sqrt{S_{yy}(1-r^2_{xy})}}{\sqrt{n-2}} = \frac{\sqrt{S_{yy}}\sqrt{(1-r^2_{xy})}}{\sqrt{n-2}}$$

Άρα $^{(ii)}$ $se(\hat{\beta}_1) = \frac{\frac{\sqrt{S_{yy}}\cdot\sqrt{(1-r^2_{xy})}}{\sqrt{n-2}}}{\sqrt{S_{xx}}} = \frac{\sqrt{S_{yy}}\cdot\sqrt{(1-r^2_{xy})}}{\sqrt{n-2}\cdot\sqrt{S_{xx}}}$

Από (i) κ' (ii):

$$\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{\frac{S_{xy}}{S_{xx}}}{\frac{\sqrt{S_{yy}}\cdot\sqrt{(1-r^2_{xy})}}{\sqrt{n-2}\cdot\sqrt{S_{xx}}}} = \frac{S_{xy}\cdot\sqrt{n-2}\cdot\sqrt{S_{xx}}}{S_{xx}\cdot\sqrt{S_{yy}}\cdot\sqrt{(1-r^2_{xy})}} =$$

$$= \frac{S_{xy}\cdot\sqrt{n-2}\cdot(S_{xx})^{\frac{1}{2}}}{S_{xx}\cdot\sqrt{S_{yy}}\cdot\sqrt{(1-r^2_{xy})}} = \frac{S_{xy}\cdot\sqrt{n-2}}{S_{xx}\cdot S_{xx}^{-\frac{1}{2}}\sqrt{S_{yy}}\cdot\sqrt{(1-r^2_{xy})}} =$$

$$= \frac{S_{xy}\cdot\sqrt{n-2}}{(S_{xx})^{\frac{3}{2}-\frac{1}{2}}\cdot\sqrt{S_{yy}}\cdot\sqrt{(1-r^2_{xy})}} = \frac{S_{xy}}{\sqrt{S_{xx}}\cdot\sqrt{S_{yy}}}\cdot\frac{\sqrt{n-2}}{\sqrt{(1-r^2_{xy})}}$$

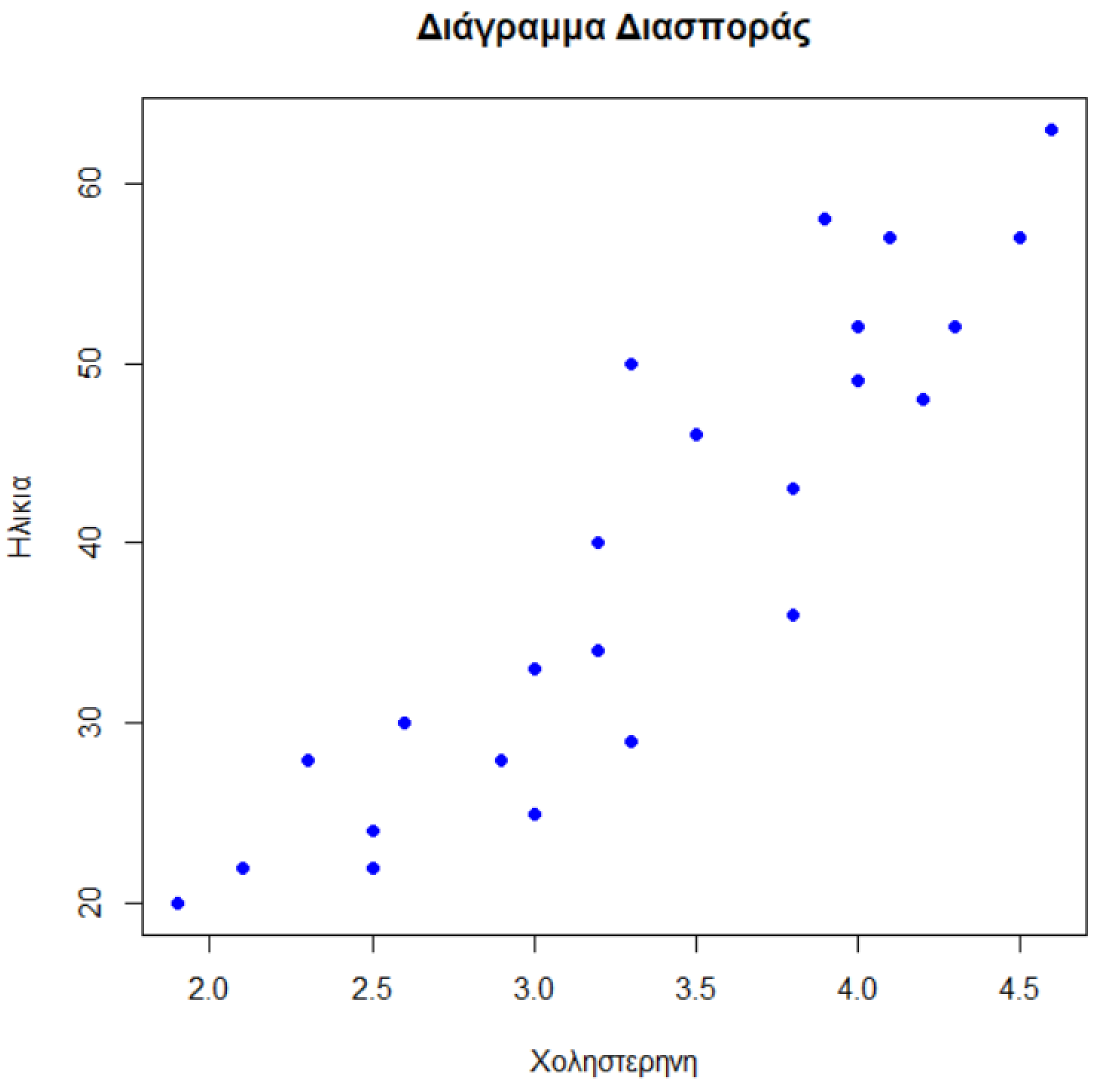
Από θεωρία έχουμε ότι: $r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$

Συνεπώς, μόλις αποδείξαμε ότι:

$$\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

ΆΣΚΗΣΗ Β

i) Διάγραμμα διασποράς μεταξύ των μεταβλητών y (χοληστερόλης) και x (ηλικίας)



Τα δεδομένα μας προσαρμόζονται στο μοντέλο y=b0+b1x ως εξής:

Y= 1.279868 + 0.052625x

```
> model<- lm(chol~age)
> summary(model)

Call:
lm(formula = chol ~ age)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.6111 -0.2151 -0.0058  0.2297  0.6256 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.279868   0.215699   5.934 5.69e-06 ***
age         0.052625   0.005192  10.136 9.43e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.334 on 22 degrees of freedom
Multiple R-squared:  0.8236,    Adjusted R-squared:  0.8156 
F-statistic: 102.7 on 1 and 22 DF,  p-value: 9.428e-10
```

ii) Εκτελούμε έλεγχο σημαντικότητας και βρίσκουμε p-value=5.707435e-47 , το οποίο είναι ένας αριθμός πολύ μικρότερος του 1%, άρα η μηδενική υπόθεση απορρίπτεται. Για τον υπολογισμό αυτόν κάναμε τις ακόλουθες πράξεις:
Z=mean(x)/(sd(y)/$\sqrt{y}$)
P value← 2*Φ(-|Z|) , επειδή θέλουμε Β1!=0

Συμπέρασμα:

Η πρόβλεψη για την τιμή $\widehat{\beta 1}$ είναι ότι είναι διάφορη του 0. ($\widehat{\beta 1}$!=0)

```
#Ελεγχος σημαντικότητας
z<- mean(age)/(sd(age)/sqrt(24))
> z
[1] 14.39322
> pvalue<- 2*pnorm(-abs(z))
> pvalue
[1] 5.707435e-47
```

Το διάστημα εμπιστοσύνης το υπολόγισα με 2 τρόπους.

Ο 1ος είναι χρησιμοποιώντας μια μεμονωμένη εντολή στην R και ο 2ος τρόπος είναι χρησιμοποιώντας τους κατάλληλους μαθηματικούς τύπους. Και στις 2 περιπτώσεις βρήκα ότι το διάστημα εμπιστοσύνης είναι: [33.75153 ,45.08180]. Παρακάτω είναι οι εντολές που χρησιμοποιήθηκαν:

```
#Υπολογισμός διαστήματος εμπιστοσύνης
> confidence_interval <- t.test(age)$conf.int
> confidence_interval
[1] 33.75153 45.08180
attr(,"conf.level")
[1] 0.95

> t_value <- qt(0.975, df = 23)
> t_value
[1] 2.068658
> x2<- mean(age)+t_value*(sd(age)/sqrt(24))
> x1<- mean(age)-t_value*(sd(age)/sqrt(24))
> x1
[1] 33.75153
> x2
[1] 45.0818
```

iii) Για να υπολογίσουμε ένα διάστημα εμπιστοσύνης επιπέδου 99% για την μεταβλητή y (χοληστερίνη) θα χρειαστεί πρώτα να υπολογίσουμε την τιμή του y για x=20. Έπειτα να βρω την τυπική απόκλιση της μεταβλητής y (χοληστερίνη) και τέλος να χρησιμοποιήσω τον τύπο:
$\hat{y}$ ∓t_critical*sd(chol)*sqrt(1/n+((x0-mean(age))^2)/Σ((age[i]-mean(age))^2))
Το διάστημα εμπιστοσύνης που παίρνουμε είναι:
[2.649497 , 3.593983]

Στην προκειμένη περίπτωση προβλεπόμενη και αναμενόμενη τιμή θα δώσουν το ίδιο αποτέλεσμα. Αυτό συμβαίνει διότι ο όρος "προβλεπόμενη τιμή" αναφέρεται στην τιμή που προβλέπεται από το μοντέλο παλινδρόμησης για μια συγκεκριμένη τιμή της ανεξάρτητης μεταβλητής. Από την άλλη πλευρά, η "αναμενόμενη τιμή" αναφέρεται στην τιμή που αναμένεται να προκύψει για την ανεξάρτητη μεταβλητή, με βάση το μοντέλο παλινδρόμησης και τα δεδομένα που έχουν χρησιμοποιηθεί για την εκπαίδευση του μοντέλου.

Οι δύο αυτές έννοιες ταυτίζονται εφόσον αναφέρονται στην ίδια τιμή, η οποία προκύπτει από το μοντέλο παλινδρόμησης. Όταν λοιπόν μιλάμε για την προβλεπόμενη τιμή, αναφερόμαστε στην τιμή που αναμένεται από το μοντέλο για μια συγκεκριμένη τιμή της ανεξάρτητης μεταβλητής, η οποία συμπίπτει με την αναμενόμενη τιμή.

Ο κώδικας που χρησιμοποιήθηκε ήταν ο εξής:

#Υπολογισμός προβλεπόμενης τιμής και αναμενόμενης τιμής
E(y)=b0 + b1*x , όπου b0=1.279868 , b1=0.052625 και x0=35.
Άρα, E(y)=1.279868 +0.052625*35

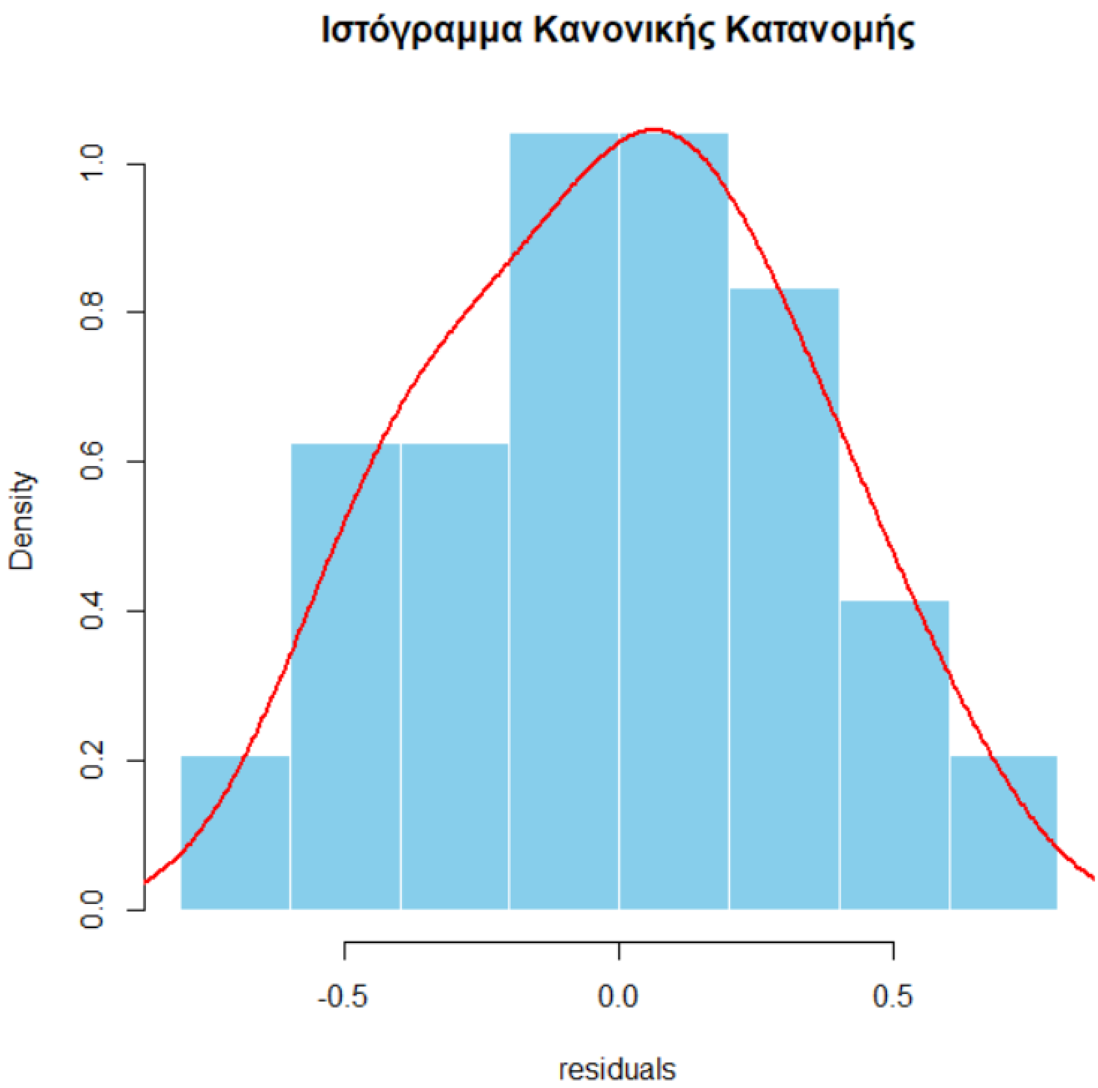
Χρησιμοποιώντας την R, έχουμε ότι:

```
> x0 <- 35
> E_y <- coef(model)[1] + coef(model)[2] * x0
> E_y
(Intercept)
   3.12174
```

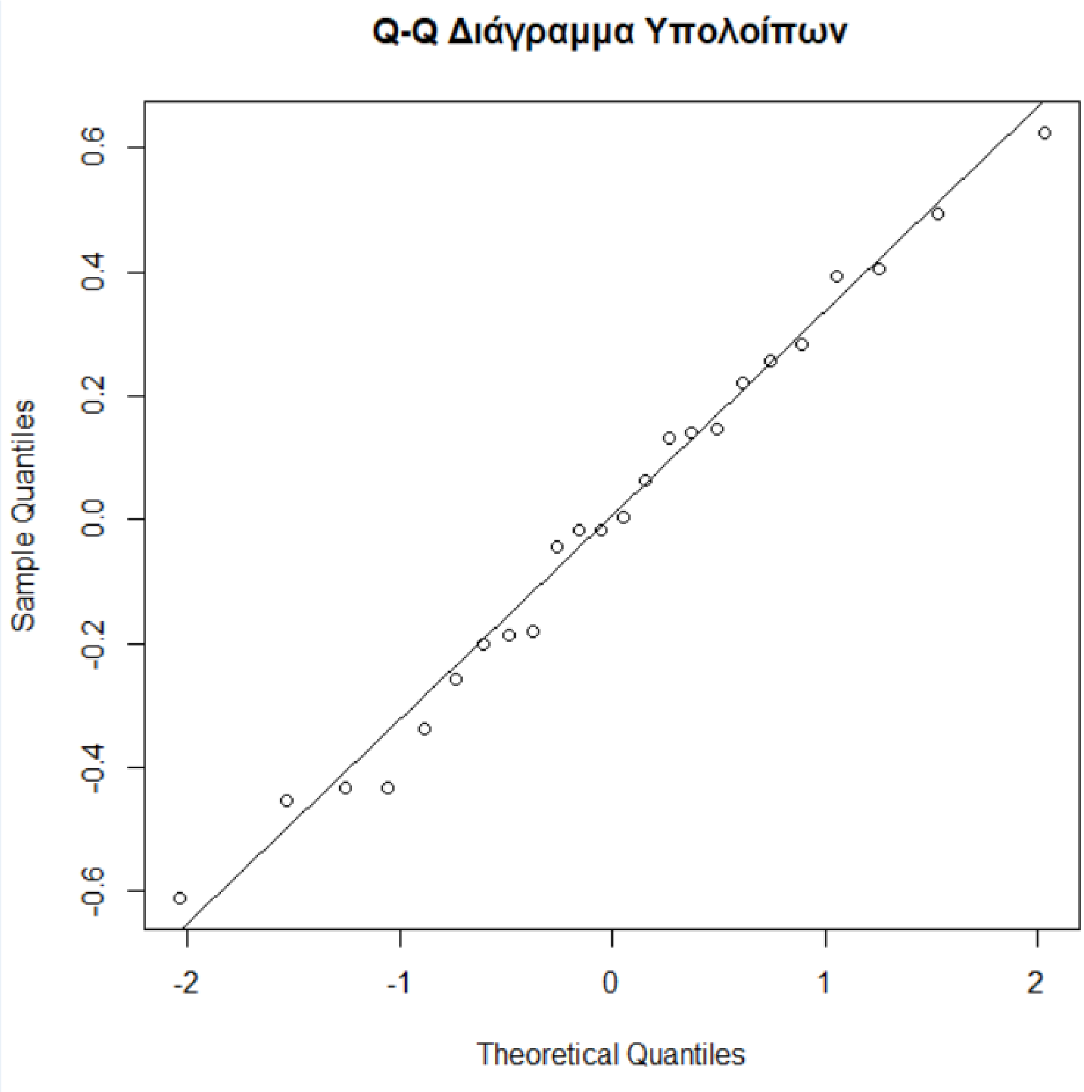
```
#Υπολογισμός T student
t_critical <- qt(0.995, 22)

> sum<-0
> for (i in 1:24){
+ sum<- sum+ (age[i]- mean(age))^2
+ }
> margin_error<- t_critical*sd(chol)*sqrt(1/n+((x0-
mean(age))^2)/sum)
```

```
> confidence_interval<- c(E_y - margin_error, E_y +
margin_error)
> confidence_interval
(Intercept) (Intercept)
  2.649497    3.593983
```

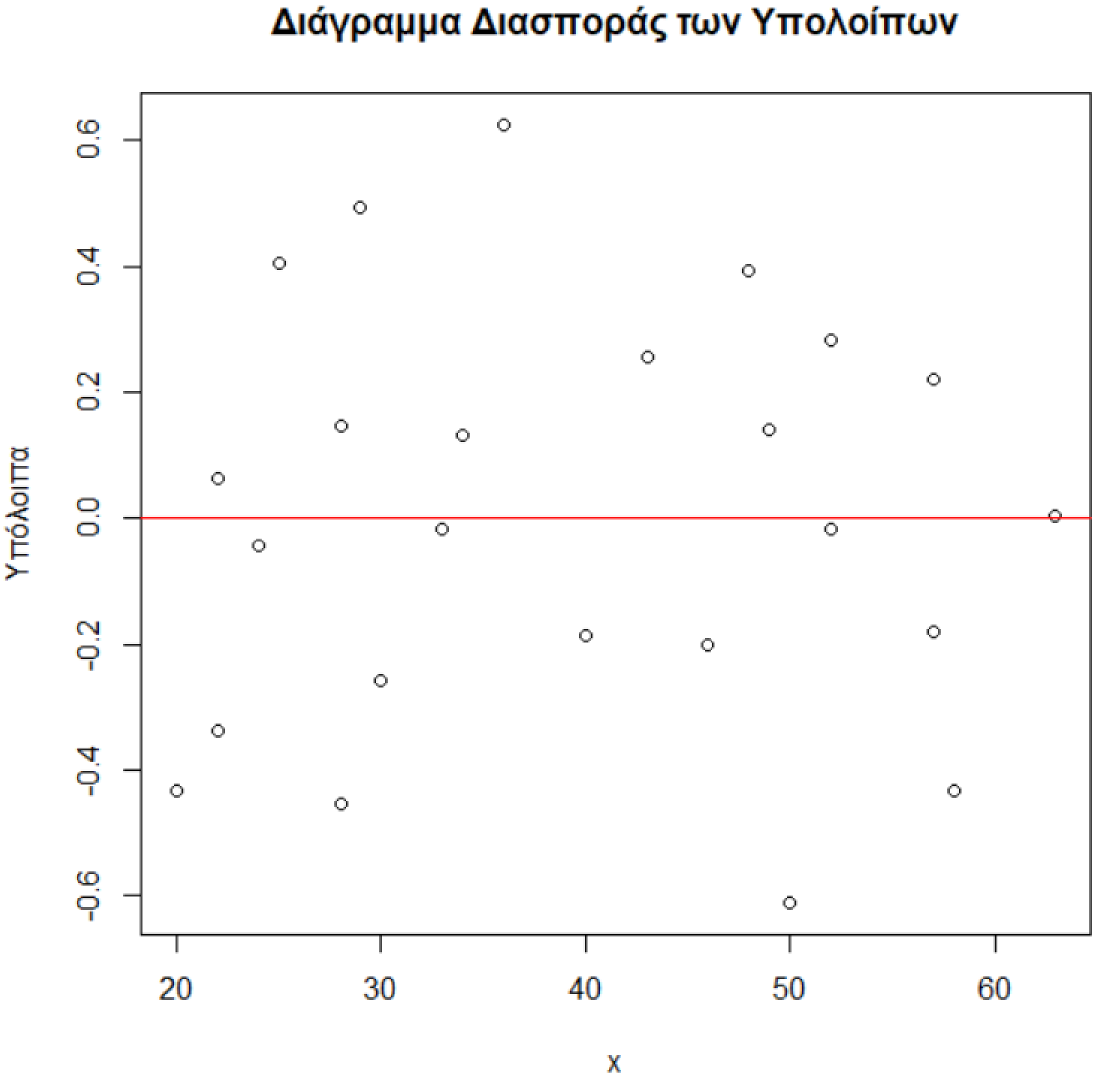
iv) Πραγματοποίησα τον γραφικό έλεγχο της κανονικής κατανομής δημιουργώντας ένα ιστόγραμμα:



Διάγραμμα υπολοίπων:



Η γραφική παράσταση ei με $\hat{yi}$ για τα υπόλοιπα ei:



Συμπέρασμα:

Μέσα από τα παραπάνω διαγράμματα συμπεραίνουμε ότι οι κατανομές των καταλοίπων είναι ομοιόμορφες παρά τις διαφορετικές τιμές των προβλεπόμενων τιμών. Αυτό σημαίνει ότι η διασπορά των καταλοίπων είναι ομοιόμορφη γύρω από το μηδέν σε όλο το εύρος των προβλεπόμενων τιμών. Το γεγονός αυτό είναι πολύ σημαντικό γιατί υποδηλώνει ότι το μοντέλο γραμμικής παλινδρόμησης προβλέπει τις τιμές με ακρίβεια και συνέπεια.

Ο κώδικας που χρησιμοποιήθηκε είναι:

```
#Διάγραμμα Κανονικής Κατανομής
> residuals <- resid(model)
> hist(residuals, probability = TRUE, col = "skyblue", border =
"white", main = "Ιστόγραμμα Κανονικής Κατανομής")
> lines(density(residuals), col = "red", lwd = 2)
```

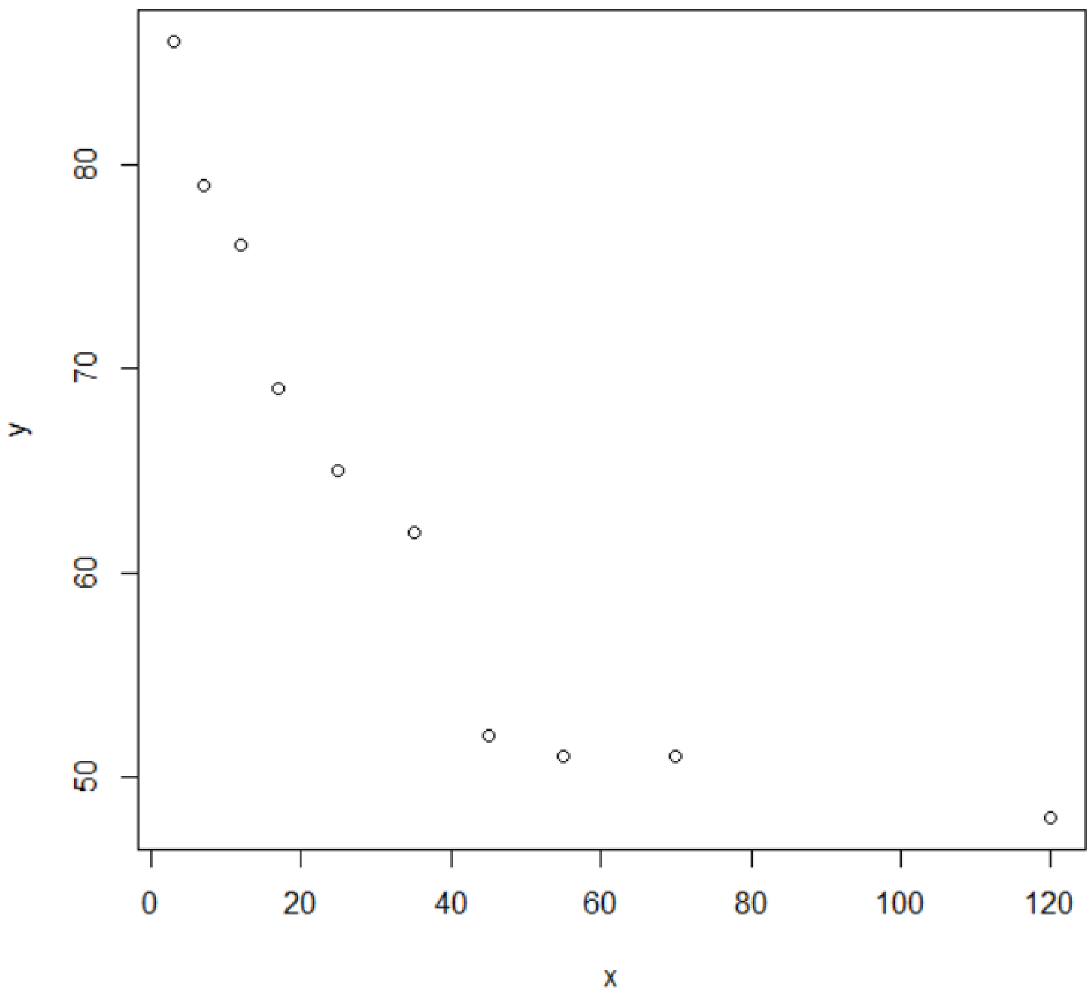
```
#Διάγραμμα υπολοίπων
> qqnorm(residuals, main = "Q-Q Διάγραμμα Υπολοίπων")
> qqline(residuals)
```

#Γραφική Παράσταση ει με $\widehat{yi}$ για τα υπόλοιπα ει

```
plot(data$x, residuals, xlab = "x", ylab = "Υπόλοιπα", main =
"Διάγραμμα Διασποράς των Υπολοίπων")
abline(h = 0, col = "red")
```

ΑΣΚΗΣΗ Γ

i) plot(y~x). Με τη συγκεκριμένη εντολή πραγματοποιώ διάγραμμα διασποράς μεταξύ της εξαρτημένης μεταβλητής y και της ανεξάρτητης x



Η άσκηση μας ζητάει να προσαρμόσουμε τα δεδομένα μας στο μοντέλο 100/(100-y)=a+b/x. Ωστόσο, το συγκεκριμένο μοντέλο είναι μη γραμμικό.
Για να το μετασχηματίσουμε σε γραμμικό εκτελούμε τις ακόλουθες ενέργειες.
Θέτουμε y*=100/(100-y) και x*=1/x και έτσι έχουμε:
y*=a+bx*

Αυτό σε κώδικα μεταφράζεται:

```
y_asteri<- 100/(100-y)
x_asteri<- 1/x
model_asteri<- lm(y_asteri~x_asteri)
summary(model_asteri)
```

```
> summary(model_asteri)

Call:
lm(formula = y_asteri ~ x_asteri)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.34675 -0.31185 -0.07989  0.18582  0.74362 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.0675     0.1596   12.96 1.19e-06 ***
x_asteri     16.2662     1.3232   12.29 1.78e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3952 on 8 degrees of freedom
Multiple R-squared:  0.9497,	Adjusted R-squared:  0.9434 
F-statistic: 151.1 on 1 and 8 DF,  p-value: 1.783e-06
```

ii)

1.Η εκτίμηση των σημείων γίνεται μέσω του μοντέλου γραμμικής παλλινδρόμησης y=a+bx. Στην προκειμένη περίπτωση πρόκειται για το μετασχηματισμένο μας μοντέλο.

Αυτό που έχουμε να κάνουμε είναι να προβλέψουμε για κάθε τιμή του Χ μία τιμή του y.
Με τις παρακάτω εντολές μπορούμε να εκτιμήσουμε σημειακά την άγνωστη μεταβλητή y

predicted_y <- predict(model_asteri)
print(predicted_y)

```
> predicted_y <- predict(model_asteri)
> print(predicted_y)
       1        2        3        4        5        6        7        8 
7.489603 4.391274 3.423046 3.024364 2.718177 2.532277 2.428999 2.363277 
       9       10 
2.299902 2.203079 
```

Τώρα την μετατρέπουμε στην κανονική της μορφή:

> y_estimates <- 100 - (100 / predicted_y)
> y_estimates

```
> y_estimates <- 100 - (100 / predicted_y)
> y_estimates
       1        2        3        4        5        6        7        8
86.64816 77.22757 70.78626 66.93520 63.21063 60.50985 58.83078 57.68588
       9       10
56.51989 54.60899
```

ii. 2)

Για να υπολογίσουμε ένα διάστημα εμπιστοσύνης της τάξης του 95% για τη μέση τιμή E(y) , όταν η αξία της κατοικίας είναι x=20,εκτελούμε τι σπαρακάτω ενέργεις:

1. Υπολογίζουμε την αναμενόμενη τιμή στο γραμμικό μοντέλο που δημιουργήσαμε αντικαθιστώντας όπου x=1/20
2. Για το y* που βρήκαμε , υπολογίζουμε το κανονικό y ως y=100-(100/y*)
3. Εφαρμόζουμε τον τύπο: y$\pm$ t_critical*sd(y)* $\sqrt{\frac{1}{n} + (\frac{\left(\frac{1}{x0} - mean(x)\right)^2}{\Sigma(\left(X - mean(x)\right)^2)}}$

Μέσα από την συγκεκριμένη διαδικασία υπολόγισα ότι το διάστημα εμπιστοσύνης για ένα επίπεδο του 95% είναι : [64.03446, 66.54132]

```
x0<- 1/20
n<- 10
> E_y <- coef(model_asteri)[1] + coef(model_asteri)[2] *
x0
> E_y_asteri<- 100-(100/E_y)
> E_y_asteri
(Intercept)
  65.28789

> sum<- 0
> for (i in 1:10){
+ sum<- sum+ (x_asteri[i]- mean(x_asteri))^2
```

+ }

> margin_error<-
t_critical*sd(y_asteri)*sqrt(1/n+(((1/x0)-
mean(x_asteri))^2)/sum)

> confidence_interval<- c(E_y_asteri - margin_error,
E_y_asteri + margin_error)
> confidence_interval
(Intercept) (Intercept)
  64.03446   66.54132

```
> confidence_interval
(Intercept) (Intercept)
   64.03446    66.54132
```

ii.3)

Σε αυτό το σημείο υπολογίζω ένα διάστημα εμπιστοσύνης 95% για τη πρόβλεψη τη ςπαρατήρησης y.

1ος Τρόπος:
> prediction_interval <- predict(model_asteri, interval =
"confidence", level = 0.95)
> prediction_interval<- 100-100/prediction_interval
> prediction_interval

```
> prediction_interval
        fit      lwr      upr
1  86.64816 84.96314 87.99359
2  77.22757 75.22559 78.93018
3  70.78626 68.08939 73.06281
4  66.93520 63.39677 69.84982
5  63.21063 58.52102 66.94754
6  60.50985 54.77517 64.95383
7  58.83078 52.35583 63.75637
8  57.68588 50.66564 62.95705
9  56.51989 48.91001 62.15667
10 54.60899 45.95635 60.87335
```

2ος Τρόπος:

> t.test(y_asteri)

```
> t.test(y_asteri)

        One Sample t-test

data:  y_asteri
t = 6.2562, df = 9, p-value = 0.0001485
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 2.098722 4.476078
sample estimates:
mean of x
   3.2874
```

> low_interv<- 2.098722

> max_interv<- 4.476078

> interval_of_confidence<- c(100-100/low_interv, 100-100/max_interv)

> interval_of_confidence

[1] 52.35196 77.65901

```
> interval_of_confidence
[1] 52.35196 77.65901
>
```