

1 ΑΠΟΔΕΙΞΕΙΣ ΑΠΛΟΥ ΓΡΑΜΜΙΚΟΥ ΜΟΝΤΕΛΟΥ

Τα παρακάτω αφορούν αποκλειστικά το απλό γραμμικό μοντέλο, $E(y|x) = \beta_0 + \beta_1 x$, το οποίο δίνει εκτιμήσεις \hat{y} για τις οποίες ισχύει $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Οι συντελεστές $\hat{\beta}_0$ και $\hat{\beta}_1$ προσδιορίζονται βάσει της μεθόδου ελαχίστων τετραγώνων, η οποία έγκειται στην ελαχιστοποίηση της ποσότητας

$$S(\beta_0, \beta_1) = \sum_i (y_i - E(y_i|x_i))^2 = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2 \quad (1.1)$$

ως προς τα β_0 και β_1 . Συγκεκριμένα, οι

$$0 = \frac{\partial S}{\partial \beta_0} \Big|_{(\hat{\beta}_0, \hat{\beta}_1)} = \sum_i y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad (1.2)$$

και

$$0 = \frac{\partial S}{\partial \beta_1} \Big|_{(\hat{\beta}_0, \hat{\beta}_1)} = \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i \quad (1.3)$$

καλούνται οι κανονικές εξισώσεις του προβλήματος και η λύση τους [1] δίνει τους συντελεστές $\hat{\beta}_0$ και $\hat{\beta}_1$ ως

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \quad \text{και} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (1.4)$$

Προς απλοποίηση του συμβολισμού, ορίζονται οι ακόλουθες ποσότητες:

| | |
|----------------|---|
| SSR | $\sum_i (\hat{y}_i - \bar{y})^2$ |
| SST ή S_{yy} | $\sum_i (y_i - \bar{y})^2$ |
| S_{xx} | $\sum_i (x_i - \bar{x})^2$ |
| S_{xy} | $\sum_i (x_i - \bar{x})(y_i - \bar{y})$ |

Βάσει αυτών, για παράδειγμα, ο συντελεστής $\hat{\beta}_1$ μπορεί να γραφεί ως S_{xy}/S_{xx} .

1.1 Για το συντελεστή προσδιορισμού, R , ισχύει:

$$\begin{aligned} R^2 &= \frac{\text{SSR}}{\text{SST}} = \frac{1}{S_{yy}} \sum_i (\hat{y}_i - \bar{y})^2 = \frac{1}{S_{yy}} \sum_i (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2 = \\ &= \frac{1}{S_{yy}} \sum_i (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y})^2 = \frac{\hat{\beta}_1^2}{S_{yy}} \sum_i (x_i - \bar{x})^2 = \frac{S_{xx}}{S_{yy}} \cdot \hat{\beta}_1^2 = \\ &= \frac{S_{xx}}{S_{yy}} \cdot \left(\frac{S_{xy}}{S_{xx}} \right)^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} = \left(\frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \right)^2 \equiv r_{xy}^2, \end{aligned} \quad (1.5)$$

δηλαδή πράγματι προκύπτει πως ταυτίζεται με το συντελεστή συσχέτισης Pearson.

1.2 Η κανονική εξίσωση (1.2) μπορεί να γραφεί ως

$$0 = \sum_i y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = \sum_i y_i - \hat{y}_i = \sum_i y_i - \sum_i \hat{y}_i, \quad (1.6)$$

από όπου προκύπτει το ζητούμενο, δηλαδή

$$\sum_i y_i = \sum_i \hat{y}_i. \quad (1.7)$$

Αξίζει να αναφερθεί πως από τη Σχέση (1.7) μπορεί να αποδειχθεί πως η μέση τιμή των συνήθων υπολοίπων $e_i = y_i - \hat{y}_i$ είναι επίσης μηδέν, όπως εξάλλου αναμένουμε βάσει θεωρίας.

1.3 Η σχέση για το $\hat{\beta}_1$ μπορεί να γραφεί ισοδύναμα ως

$$\begin{aligned} \hat{\beta}_1 &= \frac{1}{S_{xx}} \sum_i (x_i - \bar{x}) (y_i - \bar{y}) = \frac{1}{S_{xx}} \sum_i (x_i - \bar{x}) y_i - \frac{\bar{y}}{S_{xx}} \sum_i (x_i - \bar{x}) = \\ &= \frac{1}{S_{xx}} \sum_i (x_i - \bar{x}) y_i, \end{aligned} \quad (1.8)$$

διότι

$$\sum_i (x_i - \bar{x}) = \sum_i x_i - n\bar{x} = \sum_i x_i - n \cdot \frac{1}{n} \sum_i x_i = 0. \quad (1.9)$$

Ισχύει, λοιπόν,

$$\begin{aligned} \text{Cov}(\bar{y}, \hat{\beta}_1) &= \text{Cov}\left(\frac{1}{n} \sum_i y_i, \frac{1}{S_{xx}} \sum_i (x_i - \bar{x}) y_i\right) = \frac{1}{n \cdot S_{xx}} \cdot \text{Cov}\left(\sum_i y_i, \sum_i (x_i - \bar{x}) y_i\right) = \\ &= \frac{1}{n \cdot S_{xx}} \sum_{i,j} \text{Cov}(y_j, (x_i - \bar{x}) y_i) = \frac{1}{n \cdot S_{xx}} \sum_{i,j} (x_i - \bar{x}) \text{Cov}(y_j, y_i) = \\ &= \frac{1}{n \cdot S_{xx}} \sum_{i,j} (x_i - \bar{x}) \text{Var}(y_i) \delta_{ij} = \frac{\sigma^2}{n \cdot S_{xx}} \sum_i (x_i - \bar{x}) = 0, \end{aligned} \quad (1.10)$$

λόγω της Σχέσης (1.9), καθώς και του γεγονότος πως $\text{Var}(y_i) = \sigma^2, \forall i$. Αξίζει να σημειωθεί πως στα πλαίσια της Σχέσης (1.10) τα x θεωρούνται σταθερές και δεν έχουν το ρόλο τυχαίας μεταβλητής. Αυτός είναι και ο λόγος για τον οποίο χρησιμοποιείται η γραμμικότητα της Cov ώστε να βγουν εκτός του ορίσμάτος της.

1.4 Ισχύει πως

$$\begin{aligned}
 \sum_i y_i \hat{y}_i - \sum_i \hat{y}_i^2 &= \sum_i \hat{y}_i (y_i - \hat{y}_i) = \sum_i (\hat{\beta}_0 + \hat{\beta}_1 x_i) (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \\
 &= \hat{\beta}_0 \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) + \hat{\beta}_1 \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = \\
 &= \hat{\beta}_0 \cdot 0 + \hat{\beta}_1 \cdot 0 = 0,
 \end{aligned} \tag{1.11}$$

όπου στην προτελευταία ισότητα ο συντελεστής του $\hat{\beta}_0$ μηδενίζεται λόγω της κανονικής εξίσωσης (1.2), ενώ ο συντελεστής του $\hat{\beta}_1$ μηδενίζεται λόγω της κανονικής εξίσωσης (1.3). Έτσι, αποδείχθηκε πως

$$\sum_i y_i \hat{y}_i = \sum_i \hat{y}_i^2. \tag{1.12}$$

1.5 Το άθροισμα μπορεί να γραφεί ως

$$\begin{aligned}
 \sum_i (y_i - \hat{y}_i) (\hat{y}_i - \bar{y}) &= \sum_i y_i \hat{y}_i - \sum_i \hat{y}_i^2 + \bar{y} \sum_i (y_i - \hat{y}_i) = \\
 &= 0 + \bar{y} \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0,
 \end{aligned} \tag{1.13}$$

αφενός λόγω της Σχέσης (1.11) και αφετέρου λόγω της κανονικής εξίσωσης (1.2).

1.6 Το τυπικό σφάλμα $se(\hat{\beta}_1)$ ορίζεται ως η τετραγωνική ρίζα της εκτιμημένης διασποράς $\hat{V}(\hat{\beta}_1)$, για την οποία ισχύει (βλ. [1])

$$\hat{V}(\hat{\beta}_1) = \frac{S^2}{S_{xx}}, \tag{1.14}$$

όπου το S^2 είναι η εκτιμήτρια της πραγματικής διασποράς σ^2 (η οποία είναι άγνωστη) και ορίζεται ως

$$S^2 = \frac{1}{n-2} \sum_i (y_i - \hat{y}_i)^2. \tag{1.15}$$

Υψώνοντας στο τετράγωνο την τετριμμένη ισότητα $y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$ και χρησιμοποιώντας τη Σχέση (1.13) προκύπτει

$$\begin{aligned}
 (y_i - \bar{y})^2 &= (y_i - \hat{y}_i)^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2 \Leftrightarrow \\
 \sum_i (y_i - \bar{y})^2 &= \sum_i (y_i - \hat{y}_i)^2 + 2 \underbrace{\sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})}_0 + \sum_i (\hat{y}_i - \bar{y})^2 \Leftrightarrow \\
 \sum_i (y_i - \hat{y}_i)^2 &= \sum_i (y_i - \bar{y})^2 - \sum_i (\hat{y}_i - \bar{y})^2.
 \end{aligned} \tag{1.16}$$

Βάσει αυτού, η εκτιμήτρια S^2 μπορεί να γραφεί ως

$$\begin{aligned} S^2 &= \frac{1}{n-2} \left(S_{yy} - \sum_i (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2 \right) = \frac{1}{n-2} \left(S_{yy} - \hat{\beta}_1^2 \sum_i (x_i - \bar{x})^2 \right) \\ &= \frac{1}{n-2} \left(S_{yy} - \left(\frac{S_{xy}}{S_{xx}} \right)^2 \cdot S_{xx} \right) = \frac{1}{n-2} \left(S_{yy} - \frac{S_{xy}^2}{S_{xx}} \right). \end{aligned} \quad (1.17)$$

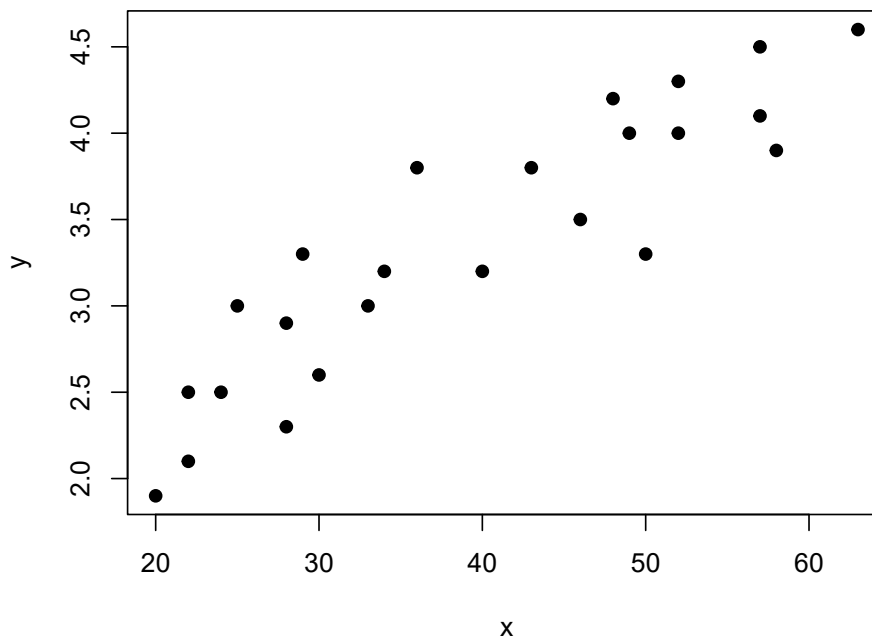
Από τον συνδυασμό των Σχέσεων (1.14) και (1.17) προκύπτει

$$\begin{aligned} \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} &= \frac{S_{xy}}{S_{xx}} \cdot \frac{\sqrt{S_{xx}}}{S} = \frac{\sqrt{n-2} S_{xy}}{\sqrt{S_{xx}}} \cdot \frac{\sqrt{S_{xx}}}{\sqrt{S_{yy} S_{xx} - S_{xy}^2}} = \\ &= \frac{\sqrt{n-2} S_{xy}}{\sqrt{S_{xx} S_{yy}}} \cdot \frac{1}{\sqrt{1 - \frac{S_{xy}^2}{S_{yy} S_{xx}}}} = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1 - r_{xy}^2}}, \end{aligned} \quad (1.18)$$

αποδεικνύοντας έτσι το ζητούμενο.

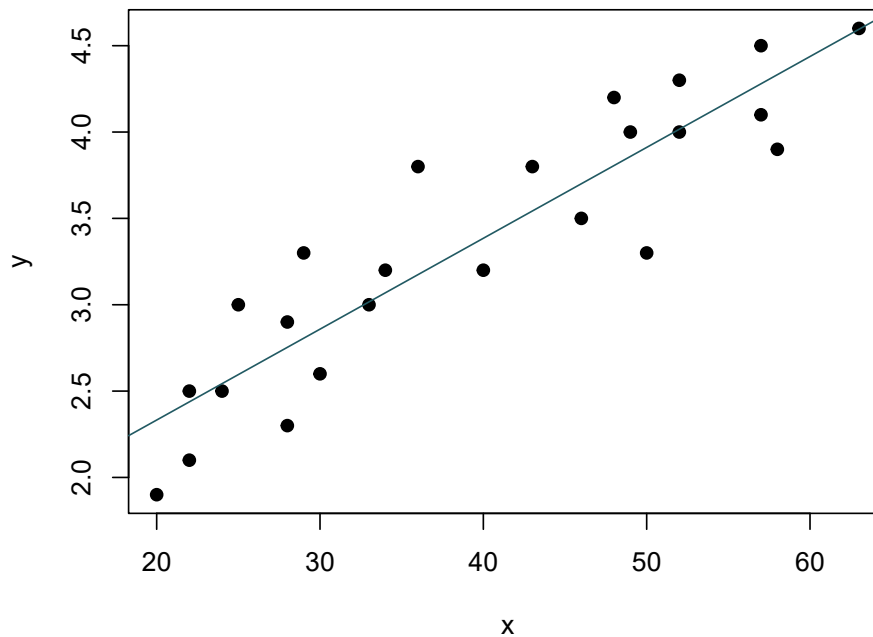
2 ΔΕΔΟΜΕΝΑ ΧΟΛΗΣΤΕΡΟΛΗΣ

2.1 Τα δεδομένα του Πίνακα Α του Παραρτήματος αφορούν τα επίπεδα ολικής χοληστερόλης (mg/ml) για 24 ασθενείς (τιμές y) και την ηλικία τους (τιμές x). Αφού τα δεδομένα φορτώθηκαν στο R Studio, σχεδιάστηκε το αντίστοιχο διάγραμμα διασποράς (βλ. Εικόνα 2.1).



Εικόνα 2.1: Διάγραμμα διασποράς για τα δεδομένα χοληστερόλης-ηλικίας.

Εκ πρώτης όψεως φαίνεται οι τιμές x και y να έχουν κάποια θετική συσχέτιση, η οποία ενδεχομένως να είναι γραμμική. Για το σκοπό αυτό, στα δεδομένα προσαρμόστηκε το απλό γραμμικό μοντέλο $E(y|x) = \beta_0 + \beta_1 x$, με την ευθεία που αντιστοιχεί στην εκτίμηση της εξίσωσης παλινδρόμησης να φαίνεται στην Εικόνα 2.2.



Εικόνα 2.2: Προσαρμογή του απλού γραμμικού μοντέλου στα δεδομένα.

Η εξίσωση της ευθείας δόθηκε ως

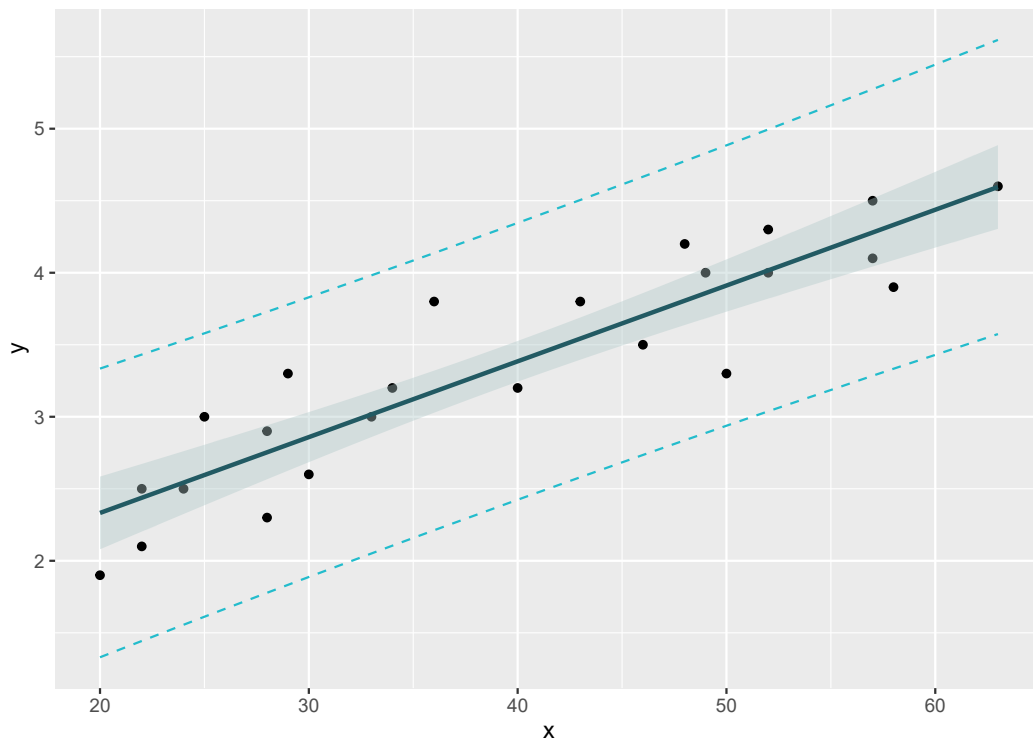
$$\hat{y} = 0.053(5)x + 1.28(22), \quad (2.1)$$

όπου σε παρένθεση φαίνεται το τυπικό σφάλμα¹.

2.2 Σχετικά με τον έλεγχο της υπόθεσης $H_0 : \beta_1 = 0$ έναντι της $H_1 : \beta_1 \neq 0$, η τιμή της ελεγχουσυνάρτησης t που αντιστοιχεί στην H_0 είναι 10.136, με αντίστοιχη p -value ίση με $9.43 \cdot 10^{-10}$ ($p < 0.001$). Το γεγονός αυτό υποδεικνύει πως η H_0 πρέπει να απορριφθεί, ή, ισοδύναμα, πως ο συντελεστής του x είναι στατιστικά διάφορος του μηδενός. Αξίζει να σημειωθεί εδώ πως, αν και συνήθως ο σταθερός όρος του μοντέλου δε μας ενδιαφέρει, στην προκειμένη περίπτωση επίσης φαίνεται να είναι στατιστικά σημαντικός, με p -value ίσο με $5.69 \cdot 10^{-6}$ ($p < 0.001$). Το 0.95-διάστημα εμπιστοσύνης για το συντελεστή του x υπολογίζεται ως (0.042, 0.063). Το γεγονός ότι το μηδέν δεν περιέχεται στο διάστημα εμπιστοσύνης έρχεται σε πλήρη συμφωνία με τους στατιστικούς ελέγχους που προηγήθηκαν για την υπόθεση $H_0 : \beta_1 = 0$. Τέλος, σε ό,τι αφορά την ερμηνεία του $\hat{\beta}_1$, από μαθηματικής πλευράς αυτό αποτελεί το συντελεστή βάρους της μεταβλητής x . Από πρακτικής πλευράς, το $\hat{\beta}_1$ ισοδυναμεί με τη μέση αύξηση στα επίπεδα ολικής χοληστερόλης που αναμένουμε να παρατηρηθεί σε έναν ασθενή εάν η ηλικία του αυξηθεί κατά 1 έτος.

¹ Ο συμβολισμός 0.053(5) ισοδυναμεί με $\hat{\beta}_1 = 0.053 \pm 0.005$, ενώ ο 1.28(22) με $\hat{\beta}_0 = 1.28 \pm 0.22$.

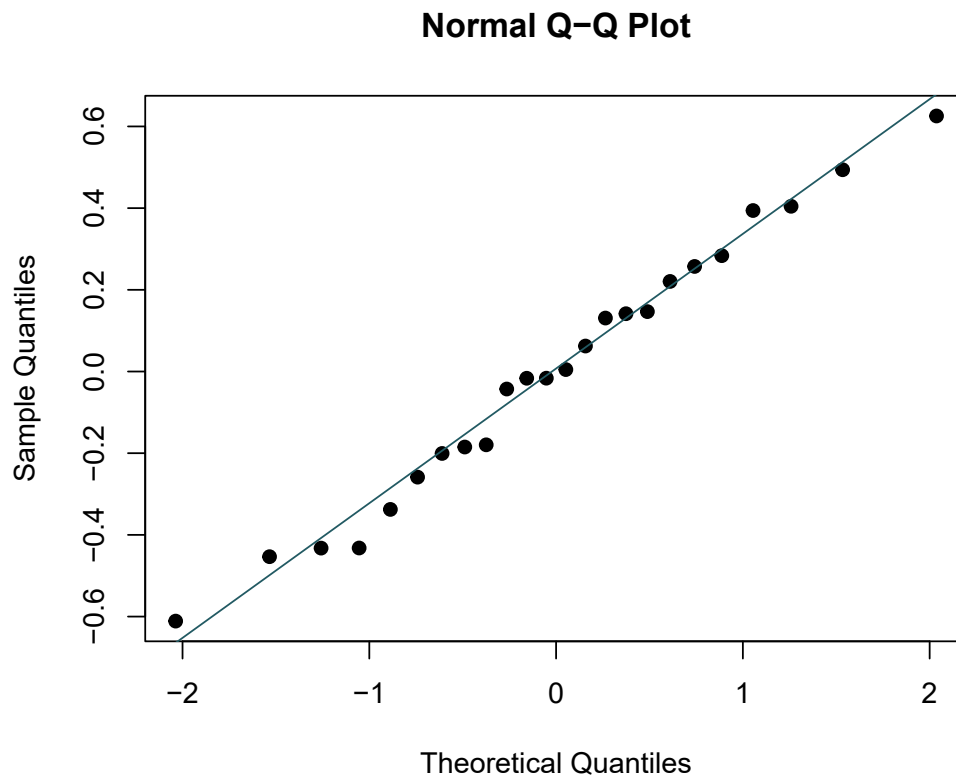
2.3 Σε ό,τι αφορά την πρόβλεψη για το επίπεδο χοληστερόλης ενός ασθενή ηλικίας 35 ετών, για $x_0 = 35$ έτη το μοντέλο προβλέπει μέσο επίπεδο ολικής χοληστερόλης $\hat{y}_0 = 3.122(23)$ mg/ml. Το 0.99-διάστημα εμπιστοσύνης για τη συγκεκριμένη πρόβλεψη ως μεμονωμένη νέα παρατήρηση είναι (2.159, 4.085), ενώ το 0.99-διάστημα εμπιστοσύνης για την αναμενόμενη τιμή της είναι (2.919, 3.325). Τα αποτελέσματα αυτά φαίνονται και στο γράφημα της Εικόνας 2.3, όπου το διάστημα (2.159, 4.085) οριοθετείται από τα σημεία τομής των διακεκομμένων γραμμών με την ευθεία $x = 35$, ενώ το διάστημα (2.919, 3.325) οριοθετείται από τα σημεία τομής της αχνής γραμμής με την ευθεία $x = 35$.



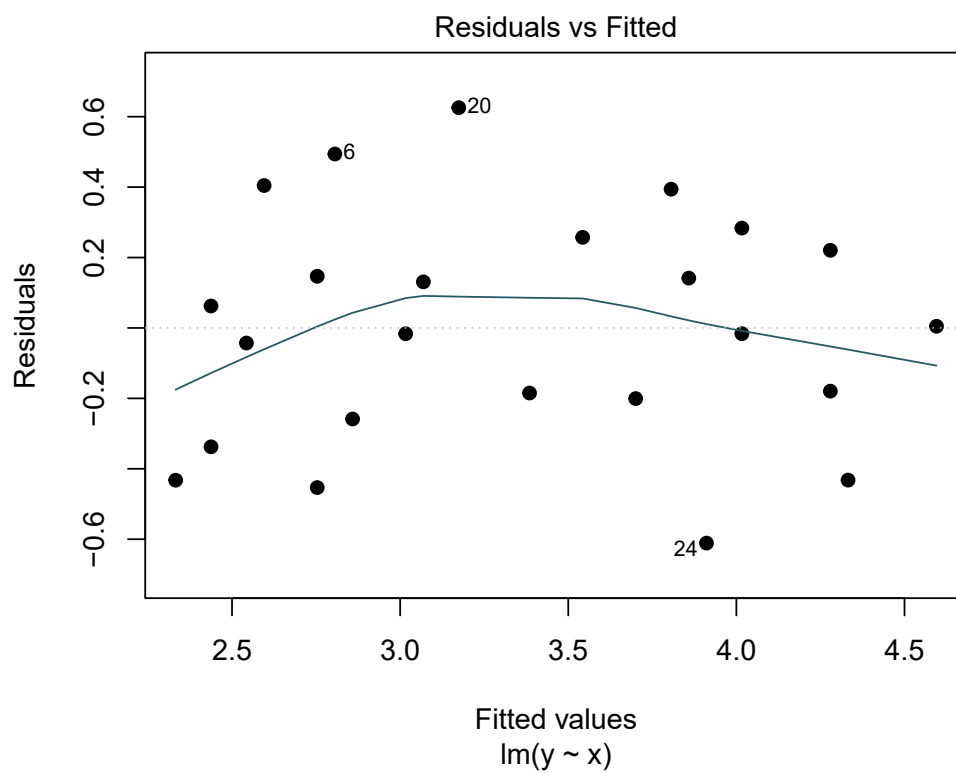
Εικόνα 2.3: Προσαρμοσμένη ευθεία παλινδρόμησης και 0.99-διάστημα εμπιστοσύνης.

2.4 Ο γραφικός έλεγχος της κανονικής κατανομής αντιστοιχεί στο σχεδιασμό του Q-Q Plot που φαίνεται στην Εικόνα 2.4. Βάσει της γραφικής αυτής παράστασης, το μάλλον ασφαλές συμπέρασμα είναι πως δεν υπάρχει λόγος αμφισβήτησης της καταλληλότητας του μοντέλου παλινδρόμησης, δεδομένου πως τα σημεία πράγματι κείτονται σε μια ευθεία, με πολύ μικρές αποκλίσεις.

Φυσικά, η καταλληλότητα του μοντέλου δεν τεκμαίρεται μονάχα από την παρατήρηση αυτή. Ένας επιπλέον έλεγχος είναι αυτός της ομοσκεδαστικότητας των τυχαίων σφαλμάτων, ο οποίος γίνεται μέσω του γραφήματος e_i (υπόλοιπα) vs \hat{y}_i (προσαρμοσμένες τιμές), το οποίο φαίνεται στην Εικόνα 2.5. Εκεί, φαίνεται πως τα τυχαία σφάλματα είναι πράγματι τυχαία κατανομημένα γύρω από το μηδέν, χωρίς να υπάρχει κάποια συσχέτιση με τις αντίστοιχες προσαρμοσμένες τιμές \hat{y}_i . Έτσι, η αποδοχή της υπόθεσης της ομοσκεδαστικότητας, επίσης συνηγεί υπέρ της καταλληλότητας του μοντέλου παλινδρόμησης.



Εικόνα 2.4: Γραφικός έλεγχος κανονικής κατανομής.



Εικόνα 2.5: Γραφικός έλεγχος ομοσκεδαστικότητας.

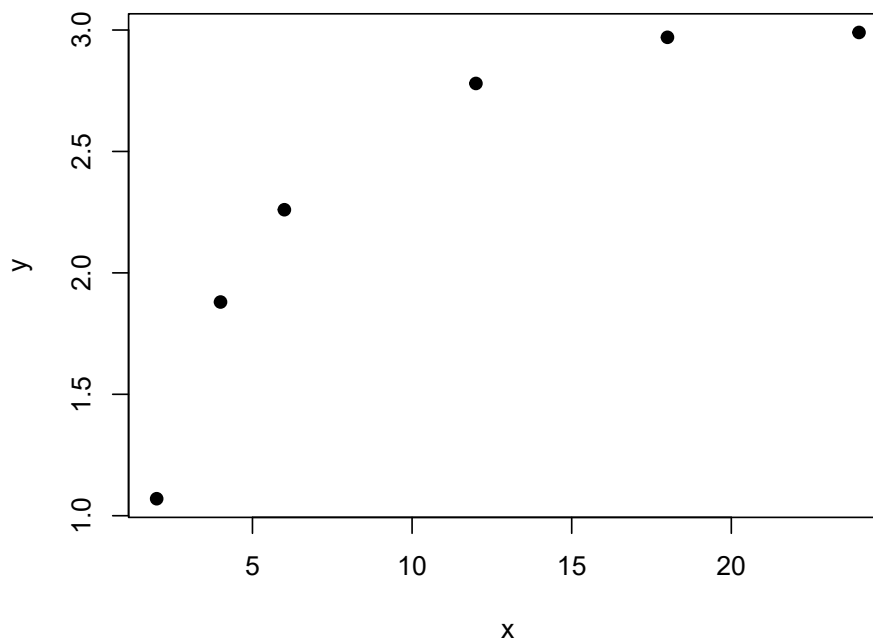
3 ΜΗ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ

3.1 Τα δεδομένα βάσει των οποίων θα προσαρμοστεί το μοντέλο της άσκησης αυτής παρουσιάζονται στον Πίνακα 3.1.

| | | | | | | |
|-----|------|------|------|------|------|------|
| x | 2 | 4 | 6 | 12 | 18 | 24 |
| y | 1.07 | 1.88 | 2.26 | 2.78 | 2.97 | 2.99 |

Πίνακας 3.1: Δεδομένα για προσαρμογή μοντέλου.

Το διάγραμμα διασποράς μεταξύ x και y φαίνεται στην Εικόνα 3.1.



Εικόνα 3.1: Διάγραμμα διασποράς για τα δεδομένα του Πίνακα 3.1.

Είναι εμφανές πως η σχέση μεταξύ των δεδομένων δεν είναι γραμμική. Το μοντέλο που καλούμαστε να προσαρμόσουμε είναι το

$$y = 3 - a \exp(\beta x). \quad (3.1)$$

Παρότι η Σχέση (3.1) δεν είναι γραμμική, μπορεί να μετασχηματιστεί σε μια γραμμική σχέση εφόσον η μεταβλητή y επανοριστεί. Ισχύει

$$y = 3 - a \exp(\beta x) \Leftrightarrow a \exp(\beta x) = 3 - y \Leftrightarrow \ln(3 - y) = \beta x + \ln a. \quad (3.2)$$

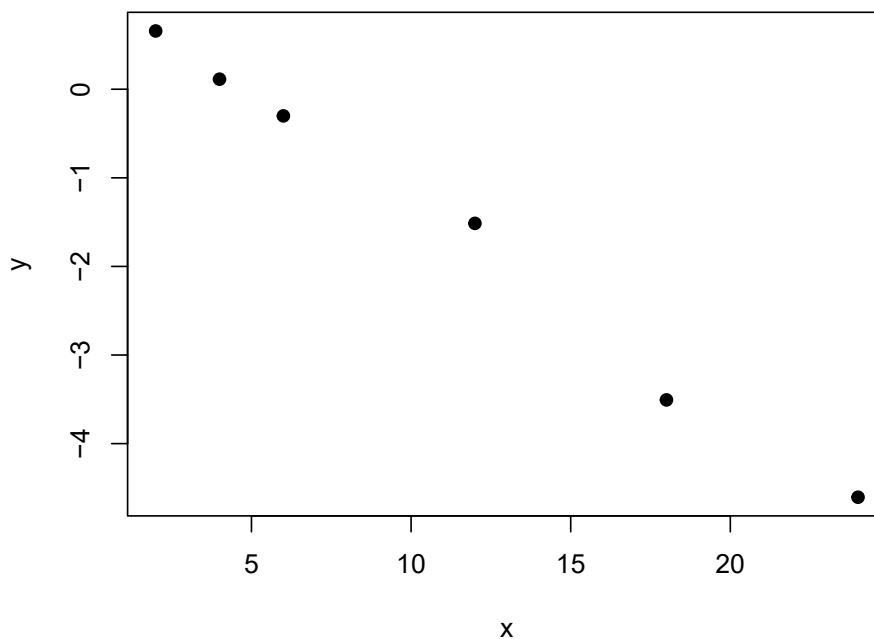
Η Σχέση (3.2) είναι καλά ορισμένη, δεδομένου πως οι τιμές του y στον Πίνακα 3.1 ικανοποιούν τη συνθήκη $y < 3$ και εφόσον ισχύει $a > 0$. Ορίζοντας τώρα τις νέες μεταβλητές

$$y^* = \ln(3 - y), \quad \beta_1 = \beta \quad \text{και} \quad \beta_0 = \ln a, \quad (3.3)$$

επιστρέφουμε στο απλό γραμμικό μοντέλο με εξίσωση $y^* = \beta_1 x + \beta_0$. Ο προσδιορισμός των παραμέτρων β_1 και β_0 θα μας επιτρέψει τον προσδιορισμό των παραμέτρων β και a , αφού η σχέση ορισμού τους είναι 1-1. Τα μετασχηματισμένα δεδομένα για την προσαρμογή του μοντέλου φαίνονται στον Πίνακα 3.2, ενώ το μετασχηματισμένο διάγραμμα διασποράς απεικονίζεται στην Εικόνα 3.2.

| | | | | | | |
|-------|--------|--------|---------|---------|---------|---------|
| x | 2 | 4 | 6 | 12 | 18 | 24 |
| y^* | 0.6575 | 0.1133 | -0.3011 | -1.5141 | -3.5066 | -4.6052 |

Πίνακας 3.2: Μετασχηματισμένα δεδομένα για προσαρμογή μοντέλου.



Εικόνα 3.2: Διάγραμμα διασποράς για τα δεδομένα του Πίνακα 3.2.

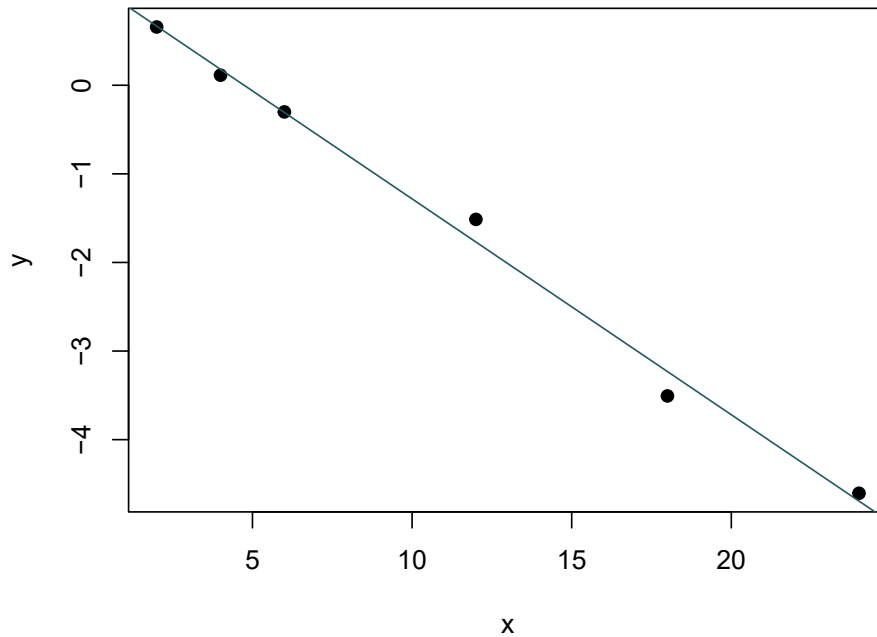
Το διάγραμμα διασποράς των μετασχηματισμένων δεδομένων υποδεικνύει πως η εφαρμογή του απλού γραμμικού μοντέλου είναι πλέον δικαιολογημένη. Για το σκοπό αυτό, χρησιμοποιώντας ξανά το R Studio, προσαρμόζουμε στα μετασχηματισμένα δεδομένα το απλό γραμμικό μοντέλο, όπως φαίνεται στην Εικόνα 3.3.

Η εξίσωση παλινδρόμησης δόθηκε ως

$$\hat{y}^* = -0.244(10)x + 1.15(14), \quad (3.4)$$

συνεπώς, χρησιμοποιώντας τη Σχέση (3.3), οι πραγματικοί συντελεστές του μοντέλου προκύπτουν ως $\beta = -0.244 \pm 0.010$ και $a = 3.172 \pm 0.444$ (πράγματι, το a προκύπτει μεγαλύτερο του μηδενός, συμπεριλαμβανομένου του τυπικού σφάλματός του). Έτσι, το προσαρμοσμένο μοντέλο είναι:

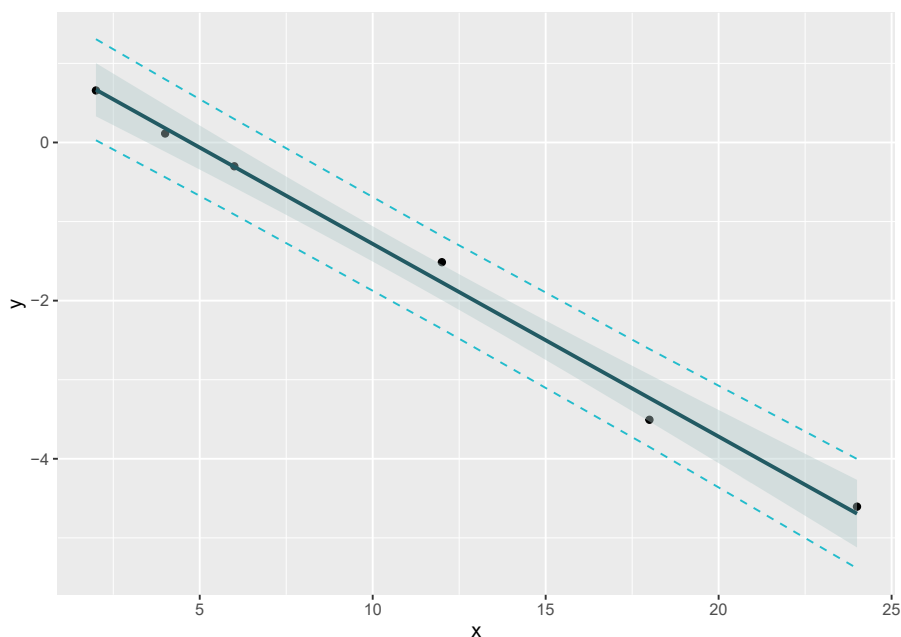
$$\hat{y} = 3 - 3.172(444) \exp(-0.244(10)x). \quad (3.5)$$



Εικόνα 3.3: Προσαρμογή του απλού γραμμικού μοντέλου στα μετασχηματισμένα δεδομένα.

Αξίζει να αναφερθεί πως για τη γραμμική προσαρμογή, η p-value που αντιστοιχεί στο συντελεστή β_1 προκύπτει ίση με $1.765 \cdot 10^{-5}$ (p-value < 0.001), δικαιολογώντας έτσι την επιλογή μας για το απλό γραμμικό μοντέλο.

3.2 Χρησιμοποιώντας τη Σχέση (3.5), η πρόβλεψη του μοντέλου για την τιμή $x_0 = 9$ είναι $\hat{y}_0 = 2.647 \pm 0.049$. Σε ό,τι αφορά το 0.95-διάστημα εμπιστοσύνης, αυτό θα υπολογιστεί βάσει του γραμμικού μοντέλου της Σχέσης (3.4) (βλ. Εικόνα 3.4).



Εικόνα 3.4: Προσαρμοσμένη ευθεία παλινδρόμησης και 0.95-διάστημα εμπιστοσύνης.

Στα πλαίσια του απλού γραμμικού μοντέλου, το 0.95-διάστημα εμπιστοσύνης για τη συγκεκριμένη πρόβλεψη ως μεμονωμένη νέα παρατήρηση y^* είναι $(-1.628, -0.449)$, ενώ το 0.95-διάστημα εμπιστοσύνης για την αναμενόμενη τιμή της είναι $(-1.268, -0.810)$. Δεδομένων των διαστημάτων αυτών για τη μεταβλητή y^* , μπορούμε χρησιμοποιώντας τη σχέση

$$y^* = \ln(3 - y) \Leftrightarrow y = 3 - \exp(y^*) \quad (3.6)$$

να προσδιορίσουμε προσεγγιστικά τα αντίστοιχα διαστήματα για τη μεταβλητή y . Επειδή η συνάρτηση $y(y^*)$ είναι γνησίως φθίνουσα, κρατούμε υπ' όψιν πως τα διαστήματα που θα βρούμε δε θα έχουν τη μορφή $(y(-1.628), y(-0.449))$ και $(y(-1.268), y(-0.810))$, αλλά $(y(-0.449), y(-1.628))$ και $(y(-0.810), y(-1.268))$, αντίστοιχα. Τα αποτελέσματα που προκύπτουν είναι πως το 0.95-διάστημα εμπιστοσύνης για τη συγκεκριμένη πρόβλεψη ως μεμονωμένη νέα παρατήρηση y είναι $(2.362, 2.804)$, ενώ το 0.95-διάστημα εμπιστοσύνης για την αναμενόμενη τιμή της είναι $(2.555, 2.719)$.

ΠΑΡΑΡΤΗΜΑ

Ακολουθεί ο πίνακας με τα δεδομένα που χρησιμοποιήθηκαν στη 2η άσκηση.

| x | y | | x | y |
|-----|-----|--|-----|-----|
| 46 | 3.5 | | 52 | 4.0 |
| 57 | 4.5 | | 28 | 2.9 |
| 22 | 2.1 | | 57 | 4.1 |
| 22 | 2.5 | | 40 | 3.2 |
| 28 | 2.3 | | 52 | 4.3 |
| 29 | 3.3 | | 24 | 2.5 |
| 20 | 1.9 | | 30 | 2.6 |
| 25 | 3.0 | | 36 | 3.8 |
| 43 | 3.8 | | 33 | 3.0 |
| 63 | 4.6 | | 48 | 4.2 |
| 49 | 4.0 | | 58 | 3.9 |
| 34 | 3.2 | | 50 | 3.3 |

Πίνακας Α: Δεδομένα χοληστερόλης ασθενών.

ΑΝΑΦΟΡΕΣ

- [1] Χ. Καρώνη και Π. Οικονόμου, *Στατιστικά Μοντέλα Παλινδρόμησης με χρήση MINITAB και R*. Εκδόσεις Συμεών, 2020.