Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

Σ χ ο λ ή Χ ρ η μ α τ ο ο ι κ ο ν ο μ ι κ ή ς κ α ι Σ τ α τ ι σ τ ι κ ή ς

Τ μ ή μ α Σ τ α τ ι σ τ ι κ ή ς κ α ι Α σ φ α λ ι σ τ ι κ ή ς Ε π ι σ τ ή μ η ς

Μ Ε Τ Α Π Τ Υ Χ Ι Α Κ Ο Π Ρ Ο Γ Ρ Α Μ Μ Α Σ Π Ο Υ Δ Ω Ν

Σ Τ Η Ν **Ε Φ Α Ρ Μ Ο Σ Μ Ε Ν Η Σ Τ Α Τ Ι Σ Τ Ι Κ Η**

ΜΗ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

ΑΝΑΣΤΑΣΙΟΣ ΧΟΥΝΤΑΛΑΣ

Δ ι π λ ω μ α τ ι κ ή Ε ρ γ α σ ί α

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς

ΣΕΠΤΕΜΒΡΙΟΣ 2020

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. 6^{η}/11.06.2018 συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Μ. Κούτρας (Επιβλέπων)

- Γ. Ηλιόπουλος, Καθηγητής

- Γ. Τζαβελάς, Επικ. Καθηγητής

Η έγκριση της Διπλωματική Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

UNIVERSITY OF PIRAEUS

School of Finance and Statistics

Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN

APPLIED STATISTICS

NON-LINEAR REGRESSION MODELS AND APPLICATIONS

By

Anastasios Chountalas

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Applied Statistics

Piraeus, Greece

September 2020

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής εργασίας θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Μάρκο Κούτρα, για την πολύτιμη βοήθεια του και για τη συνεχή υποστήριξη του κατά την πορεία της εργασίας μου.

Τέλος, θα ήθελα να αφιερώσω αυτή τη διπλωματική εργασία στους γονείς μου και στην αγαπημένη μου αδερφή, τη Πένυ, για τη διαρκή τους ψυχολογική υποστήριξη που μου παρείχαν, προκειμένου να στεφθεί με επιτυχία η ολοκλήρωση των σπουδών μου.

Περίληψη

Η μη γραμμική παλινδρόμηση είναι αρκετά παρόμοια με τη γραμμική παλινδρόμηση, η οποία έχει μελετηθεί διεξοδικά για πολλές δεκαετίες από τους στατιστικούς. Ωστόσο, το θέμα δεν είναι τόσο εύκολο να προσεγγιστεί, αφού η απώλεια της γραμμικότητας δημιουργεί αρκετά θεωρητικά προβλήματα, τα οποία πρέπει να αντιμετωπισθούν με διαφορετικούς τρόπους από τη γραμμική παλινδρόμηση. Η γνωστή ανάλυση και η γεωμετρία των ελάχιστων τετραγώνων δεν ισχύει πλέον, οπότε είναι αναπόφευκτο να καταφύγει κανείς σε ασυμπτωματικές μεθόδους και σε επαναληπτικούς αλγορίθμους. Ένα μη γραμμικό μοντέλο έχει τη γενική μορφή $Y_i = f(\boldsymbol{X}_i, \boldsymbol{\beta}) + \varepsilon_i$, όπου το $\boldsymbol{X}_i$ είναι το διάνυσμα των προβλεπουσών μεταβλητών και $\boldsymbol{\beta}$ το διάνυσμα των παραμέτρων. Μέχρι σήμερα, έχουν αναπτυχθεί αρκετά βασικά αποτελέσματα σχετικά με τη μη γραμμική παλινδρόμηση, αφού το συγκεκριμένο θέμα είναι ιδιαίτερα ενδιαφέρον και ουσιαστικής πρακτικής σημασίας που βρίσκει εφαρμογή σε προβλήματα πολλών επιστημονικών περιοχών.

Στην παρούσα εργασία περιγράφεται αρχικά το γενικό μοντέλο της μη γραμμικής παλινδρόμησης και στη συνέχεια γίνεται μια συστηματική παρουσίαση των τεχνικών εκτίμησης των παραμέτρων του μοντέλου. Στη πορεία, καλύπτονται θέματα στατιστικής συμπερασματολογίας και ύστερα γίνεται σύγκριση των διαφόρων μεθόδων που υπάρχουν. Τέλος, στην εργασία παρουσιάζονται συγκεκριμένες περιοχές που έχουν χρησιμοποιηθεί μοντέλα μη γραμμικής παλινδρόμησης.

Abstract

Nonlinear regression is quite similar to linear regression, which has been studied extensively by statisticians for many decades. However, the non-linear model is not so easy to deal with, since the loss of linearity creates several theoretical problems, which must be addressed in different ways than in linear regression. The well-known analysis and geometry of least squares are no longer valid, so it is inevitable to resort to asymptomatic methods and iterative algorithms. A non-linear model has the general form $Y_i = f(\boldsymbol{X}_i, \boldsymbol{\beta}) + \varepsilon_i$, where $\boldsymbol{X}_{\mathrm{i}}$ is the vector of the predicted variables and $\boldsymbol{\beta}$ is the vector of the parameters. To date, several key results on nonlinear regression have been developed, as this topic is of particular interest and practical importance with mumerous applications to problems in many scientific areas.

In the present Thesis, the general model of nonlinear regression is first described and then a systematic presentation of the model parameter estimation techniques is made. Statistical inference topics are covered and the various methods that exist are compared. Finally, the Thesis presents specific areas where nonlinear regression models have been used.

Περιεχόμενα

Κατάλογος σχημάτων

Κατάλογος πινάκων

Εισαγωγή

Σε πολλές στατιστικές εφαρμογές αναζητείται ο εντοπισμός ενός μοντέλου, το οποίο να εκφράζει τη σχέση μεταξύ δύο ή περισσοτέρων μεταβλητών. Τέτοια παραδείγματα είναι η σχέση ανάμεσα στο ύψος και το βάρος μιας ομάδας ανθρώπων, η σχέση που υπάρχει ανάμεσα στο εισόδημα και την κατανάλωση μιας κοινωνικής ομάδας, η σχέση που υπάρχει ανάμεσα στην εργασιακή εμπειρία και στο εισόδημα κ.ά. Το ζητούμενο σε τέτοιες καταστάσεις, είναι πρώτα από όλα, να εξεταστεί αν πράγματι υπάρχει κάποια σχέση μεταξύ δύο μεταβλητών και αν ναι να βρεθεί μια εξίσωση, ή αλλιώς μοντέλο, που να εκφράζει τη συγκεκριμένη σχέση.

Ο κύριος λόγος για τον οποίο αναζητείται ένα τέτοιο μοντέλο, είναι ότι η γνώση του παρέχει τη δυνατότητα πρόβλεψης ή καλύτερα εκτίμησης, των τιμών μιας μεταβλητής όταν είναι γνωστές η τιμή ή οι τιμές αυτών που την επηρεάζουν. Αυτό είναι πολύ συνηθισμένο σε οικονομικές εφαρμογές που αφορούν επιχειρήσεις, το χρηματιστήριο κ.ά., αλλά και ακόμα και τους υπεύθυνους χάραξης οικονομικής πολιτικής των κυβερνήσεων, αφού παρέχει τη δυνατότητα εκτιμήσεων που αφορούν μια σειρά από μεταβλητές, όπως είναι τα επιτόκια, ο πληθωρισμός, οι δείκτες ανεργίας κ.ά.

Ανεξάρτητα, όμως, από τους λόγους που οδηγούν στη μελέτη της σχέσεως μεταξύ δύο ή περισσοτέρων μεταβλητών, το πρώτο βήμα αποτελεί η κατασκευή μιας μαθηματικής εξίσωσης, ή ενός μοντέλου, το οποίο θα είναι σε θέση να περιγράψει, σε ικανοποιητικό βαθμό, τη φύση της σχέσης που υπάρχει ανάμεσα στις μεταβλητές που ενδιαφέρουν.

Η διαδικασία δημιουργίας αυτής της μαθηματικής εξίσωσης, η οποία θα είναι σε θέση να περιγράψει το συγκεκριμένο φαινόμενο, μπορεί να είναι αρκετά πολύπλοκη. Αυτό οφείλεται στο γεγονός ότι για την κατασκευή του μοντέλου είναι απαραίτητη, σε κάποιο βαθμό, της φύσης της σχέσης που συνδέει τις μεταβλητές του μοντέλου. Για παράδειγμα, αν μια επενδυτική εταιρεία ενδιαφέρεται να επενδύσει στην αγορά των ακινήτων, με σκοπό να αποκομίσει κέρδη σε δύο χρόνια, θα έπρεπε να λάβει υπόψη της μια σειρά από παράγοντες, όπως είναι τα επιτόκια, ο πληθωρισμός, η τιμή του πετρελαίου, η ζήτηση για ακίνητα, ο σχετικός χρηματιστηριακός δείκτης, η φορολογία κ.ά.

Αν, πιο συγκεκριμένα, η εταιρεία ενδιαφέρεται να μελετήσει τον τρόπο με τον οποίο τα επιτόκια επηρεάζουν τις τιμές των ακινήτων, θα μπορούσε να θεωρήσει ότι υπάρχει μια γραμμική σχέση, μεταξύ τους, όπως η παρακάτω:

$$Τιμές\ ακινήτων = \beta_0 + \beta_1 \cdot Επιτόκια$$

Στην περίπτωση που η τιμή του συντελεστή β_1 είναι αρνητική τότε κάθε αύξηση των επιτοκίων προκαλεί μείωση στις τιμές των ακινήτων, ενώ στην περίπτωση που είναι θετική, τότε αύξηση των επιτοκίων προκαλεί αύξηση στις τιμές των ακινήτων. Αν η εξίσωση ήταν τετραγωνικής μορφής, όπως η παρακάτω

$$Τιμές\ ακινήτων = \beta_0 + \beta_1 \cdot (Επιτόκια)^2.$$

Αν στη συνέχεια προστεθεί στο μοντέλο και κάποια άλλη μεταβλητή, για παράδειγμα η τιμή του πετρελαίου, τότε ο συνδυασμός των τιμών των επιτοκίων με τις τιμές του πετρελαίου, θα μπορούσε να επηρεάσει με διαφορετικό τρόπο τις τιμές των ακινήτων. Είναι προφανές ότι οι συνδυασμοί που μπορούν να γίνουν, μεταξύ και των υπολοίπων μεταβλητών, μπορεί να είναι άπειροι σε πλήθος.

Η παρούσα εργασία έχει ως σκοπό να παρουσιάσει τις βασικές έννοιες της μη γραμμικής παλινδρόμησης, καθώς και τις εφαρμογές της. Η μη γραμμική παλινδρόμηση, σε πολλά σημεία, είναι παρόμοια με τη γραμμική, η οποία έχει μελετηθεί διεξοδικά. Πλην όμως, η απώλεια της γραμμικότητας δημιουργεί αρκετά θεωρητικά προβλήματα, των οποίων η αντιμετώπιση γίνεται με διαφορετικούς τρόπους από ότι στη γραμμική παλινδρόμηση. Η γνωστή ανάλυση, καθώς και η γεωμετρία ελαχίστων τετραγώνων, δεν ισχύει πλέον, οπότε αναπόφευκτα χρειάζεται η χρήση ασυμπωτικών μεθόδων σε επαναληπτικούς αλγόριθμους.

Ένα μη γραμμικό μοντέλο έχει την παρακάτω γενική μορφή:

$$Y_i = f(\boldsymbol{X}_i, \boldsymbol{\beta}) + \varepsilon_i$$

όπου $\boldsymbol{X}_i$ είναι το διάνυσμα των προβλεπουσών μεταβλητών (predictors) και $\boldsymbol{\beta}$ το διάνυσμα των παραμέτρων.

Μέχρι σήμερα, έχουν αναπτυχθεί αρκετά βασικά αποτελέσματα, σχετικά με την εξαγωγή συμπερασμάτων, στην περίπτωση της μη γραμμικής παλινδρόμησης, αφού το συγκεκριμένο θέμα παρουσιάζει ιδιαίτερο ενδιαφέρον, αλλά και πρακτική σημασία, αφού βρίσκει εφαρμογή σε προβλήματα πολλών επιστημονικών περιοχών.

Συγκεκριμένα, ο κύριος στόχος της παρούσης εργασίας είναι:

- Η περιγραφή του γενικού μοντέλου της μη γραμμικής παλινδρόμησης, καθώς και οι ειδικές περιπτώσεις της, οι οποίες παρουσιάζουν ιδιαίτερο ενδιαφέρον.
- Η συστηματική παρουσίαση των τεχνικών που χρησιμοποιούνται για την εκτίμηση των παραμέτρων του συγκεκριμένου μοντέλου.
- Η κάλυψη των θεμάτων στατιστικής συμπερασματολογίας.
- Η παράθεση συγκεκριμένων περιοχών, όπου έχουν χρησιμοποιηθεί μοντέλα μη γραμμικής παλινδρόμησης.

Αρχικά, στο πρώτο κεφάλαιο, παρουσιάζονται τα απλά γραμμικά μοντέλα, ενώ στη συνέχεια παρουσιάζεται η έννοια της παλινδρόμησης. Στη συνέχεια, παρουσιάζεται η απλή και η πολλαπλή γραμμική παλινδρόμηση. Το δεύτερο κεφάλαιο, εστιάζει στα μη γραμμικά μοντέλα και τα είδη τους, ενώ γίνεται αναφορά στην εκτίμηση των παραμέτρων παλινδρόμησης, καθώς και στα πεδία εφαρμογής της μη γραμμικής παλινδρόμησης. Τέλος, στο τρίτο κεφάλαιο παρουσιάζεται μια εφαρμογή με χρήση της R.

Κεφάλαιο 1. Γραμμικά μοντέλα

1.1. Απλά γραμμικά μοντέλα

Ας υποθέσουμε ότι η μια τυχαία μεταβλητή Y, επηρεάζεται από μια άλλη μεταβλητή X με γραμμικό τρόπο, δηλαδή η σχέση που τις συνδέει είναι της μορφής (Montgomery & Peck, 1991):

$$Y = \beta_0 + \beta_1 \cdot X \tag{1}$$

Τότε η τυχαία μεταβλητή Y καλείται εξαρτημένη μεταβλητή (dependent variable) και η μεταβλητή X καλείται ανεξάρτητη μεταβλητή (independent variable) ή επεξηγηματική μεταβλητή (predictor variable). Η ερμηνεία των σταθερών πραγματικών αριθμών β_0 και β_1 θα παρουσιαστεί αναλυτικά στη πορεία της εργασίας. Κάθε εξίσωση της παραπάνω μορφής καλείται γραμμικό μοντέλο (linear model) (Montgomery & Peck, 1991; Kutner, Nachtsheim, Neter, & Li, 2005).

Ένα γραμμικό μοντέλο, όπως το παραπάνω, είναι πολύ χρήσιμο, αφού αν είναι γνωστή η τιμή της μεταβλητής X είναι δυνατόν να προβλεφθεί η τιμή της μεταβλητής Y, ενώ αποτυπώνεται και το είδος της εξάρτησης μεταξύ των τυχαίων μεταβλητών X και Y. Βέβαια, η συγκεκριμένη δυνατότητα πρόβλεψης, υφίσταται και στα μη γραμμικά μοντέλα, όπως θα αναφερθεί σε επόμενο κεφάλαιο. Εξισώσεις αυτού του τύπου συναντώνται σε πολλές επιστήμες και ιδιαίτερα στη Φυσική. Για παράδειγμα, η εξίσωση της μορφής

$$E = m \cdot c^2,$$

όπου E είναι η ενέργεια, m είναι η μάζα και c = ταχύτητα του φωτός είναι ένα απλό μη γραμμικό μοντέλο. Ένα άλλο παράδειγμα τέτοιου μοντέλου, το οποίο συναντάται στην Οικονομία, είναι το

$$Συνολικό\ κόστος = Στ.κόστος + Μετ.κόστος\ \cdot αριθμό\ μον.προιόντος$$

Και τα δύο παραπάνω παραδείγματα ανήκουν στην κατηγορία των ντετερμινιστικών μοντέλων (deterministic models). Τα συγκεκριμένα μοντέλα επιτρέπουν τον καθορισμό της τιμής της εξαρτημένης μεταβλητής, η οποία βρίσκεται στο αριστερό μέλος της εξίσωσης, αν είναι γνωστή η τιμή της ανεξάρτητης μεταβλητής, η οποία βρίσκεται στο δεξιό μέλος της εξίσωσης.

Παρά την ευρεία χρήση αυτών των μοντέλων, σε μια σειρά από επιστήμες, πρακτικά είναι δύσκολο να υπάρχει μια τέλεια γραμμική σχέση μεταξύ δύο μεταβλητών. Για παράδειγμα, είναι προφανές ότι η τιμή πώλησης ενός οχήματος εξαρτάται από τον κυβισμό του, αλλά όχι μόνο από αυτόν, αφού οχήματα του ίδιου κυβισμού πωλούνται σε διαφορετικές τιμές. Αυτό οφείλεται και σε άλλα χαρακτηριστικά (μεταβλητές), όπως είναι η μάρκα του αυτοκινήτου, οι οποίες είναι δύσκολο να εντοπιστούν ή ακόμα και αν εντοπιστούν είναι δύσκολο να μετρηθούν. Έτσι, στις περισσότερες των περιπτώσεων που αναφέρονται σε πρακτικά προβλήματα, είναι απαραίτητο να χρησιμοποιηθούν μοντέλα τα οποία θα λαμβάνουν υπόψη τους το στοιχείο της τυχαιότητας (randomness). Αυτού του είδους τα μοντέλα καλούνται μοντέλα πιθανότητας (probabilistic models) ή στοχαστικά μοντέλα (stochastic models).

Η κατασκευή ενός μοντέλου πιθανότητας ξεκινά από ένα ντετερμινιστικό μοντέλο, το οποίο προσεγγίζει ικανοποιητικά τη σχέση μεταξύ των υπό μελέτη μεταβλητών. Στη συνέχεια, προστίθεται ένας τυχαίος όρος, ο οποίος μετρά τις αποκλίσεις (σφάλματα) του ντετερμινιστικού όρου. Τέτοια μοντέλα είναι το μοντέλο της απλής γραμμικής παλινδρόμησης (simple linear regression) με εξίσωση (Kutner et al., 2005).

$$\Upsilon = \beta_0 + \beta_1 X + \varepsilon \tag{2}$$

το μοντέλο της πολλαπλής γραμμικής παλινδρόμησης (multiple linear regression) με εξίσωση (Kutner et al., 2005).

$$\Upsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{\mathrm{k}} X_{\mathrm{k}} + \varepsilon \tag{3}$$

και το μοντέλο πολυωνυμικής παλινδρόμησης (multinomial regression) με εξίσωση

$$\Upsilon = \beta_0 + \beta_1 \cdot X + \beta_2 \cdot X^2 + \cdots + \beta_k \cdot X^k + \varepsilon \tag{4}$$

Σε όλες τις παραπάνω περιπτώσεις, ο όρος ε που προστίθεται, αφορά τα σφάλματα του κάθε μοντέλου.

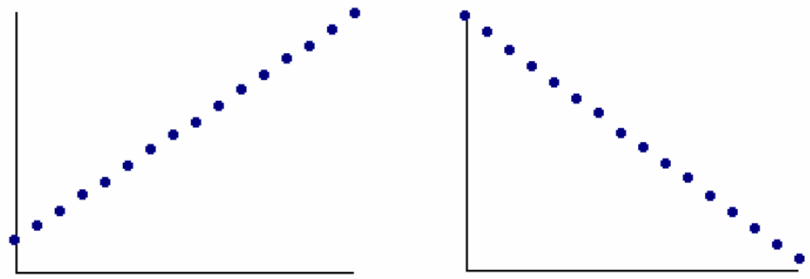
1.2. Παλινδρόμηση

Η παλινδρόμηση αποτελεί ένα σημαντικό εργαλείο για την ανάλυση διάφορων φαινομένων και ασχολείται με την περιγραφή και αξιολόγηση των σχέσεων μεταξύ μιας μεταβλητής, η οποία καλείται εξαρτημένη, με μία ή περισσότερες μεταβλητές, οι οποίες ονομάζονται ανεξάρτητες. Η παλινδρόμηση είναι εκείνη η μέθοδος η οποία εξετάζει τον τρόπο με τον οποίο οι μεταβολές των ανεξάρτητων μεταβλητών επηρεάζουν την εξαρτημένη.

Το πρώτο στάδιο σε μια ανάλυση παλινδρόμησης είναι η ανάλυση της συσχέτισης προκειμένου να διαπιστωθεί αν υπάρχει στατιστική σχέση μεταξύ δύο μεταβλητών. Εάν η συγκεκριμένη ανάλυση συσχέτισης καταλήξει στο συμπέρασμα ότι οι δύο μεταβλητές συσχετίζονται, τότε το επόμενο βήμα αφορά την περιγραφή της μεταξύ τους σχέσης. Το πιο απλό υπόδειγμα είναι αυτό που αντιστοιχεί σε μια ευθεία γραμμή, οπότε η μέθοδος καλείται ανάλυση γραμμικής παλινδρόμησης (linear regression).

Ο πιο απλός τρόπος για να διαπιστωθεί η ύπαρξη ή μη συσχέτισης ανάμεσα σε δύο μεταβλητών X και Y, είναι η κατασκευή του διαγράμματος διασποράς (scatter plot). Στη συνέχεια της εργασίας, η μεταβλητή Y θεωρείται ως εξαρτημένη και η μεταβλητή X ως ανεξάρτητη, εκτός αν αναφέρεται διαφορετικά (Kutner et al., 2005).

Παρακάτω παρουσιάζονται και ερμηνεύονται κάποια τέτοια διαγράμματα.



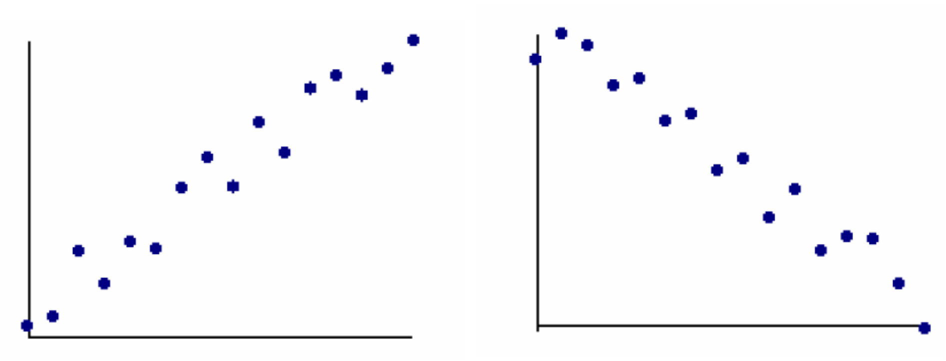
Σχήμα 1. Τέλεια θετική (αριστερά) και τέλεια αρνητική (δεξιά) γραμμική συσχέτιση

Στο Σχήμα 1 φαίνονται τα διαγράμματα διασποράς για δύο μεταβλητές X και Y όπου οι τιμές της X παριστάνονται στον οριζόντιο άξονα και της Y στον κατακόρυφο άξονα.

Στο αριστερό μέρος του Σχήματος 1 φαίνεται το διάγραμμα διασποράς μιας τέλειας θετικής γραμμικής συσχέτισης ανάμεσα στις δύο μεταβλητές X και Y, ενώ στο δεξιό μέρος του Σχήματος 1 φαίνεται το διάγραμμα διασποράς μιας τέλειας αρνητικής γραμμικής συσχέτισης ανάμεσα στις δύο μεταβλητές. Τα σημεία

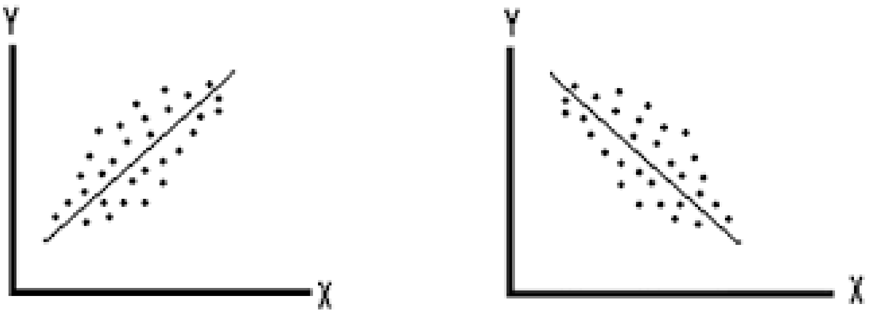
$$(x_i, y_i), \quad i = 1,2, \dots, k$$

τα οποία αναπαριστούν τα ζευγάρια των τιμών των δύο μεταβλητών, και στις δύο περιπτώσεις ανήκουν σε μια ευθεία. Στην πρώτη περίπτωση (αριστερά), η ευθεία έχει θετική κλίση, και ο συντελεστής συσχέτισης ισούται με r=+1. Στη δεύτερη περίπτωση (δεξιά), η κλίση της ευθείας είναι αρνητική και ο συντελεστής συσχέτισης ισούται με r=-1 (Montgomery & Peck, 1991).



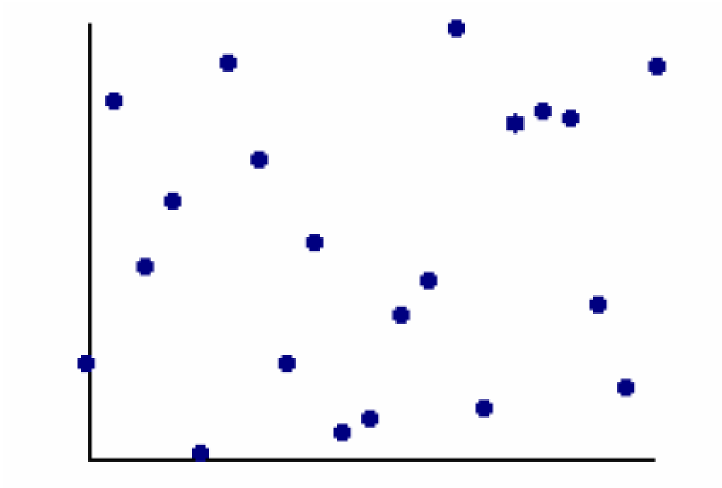
Σχήμα 2. Πολύ ισχυρή θετική (αριστερά) και αρνητική (δεξιά) γραμμική συσχέτιση

Στην πρώτη περίπτωση του Σχήματος 2 (αριστερά), η συσχέτιση παραμένει θετική, ενώ ο συντελεστής συσχέτισης λαμβάνει τιμές πολύ κοντά στο +1. Στη δεύτερη περίπτωση του ίδιου σχήματος (δεξιά), η συσχέτιση είναι αρνητική και ο συντελεστής συσχέτισης λαμβάνει τιμές πολύ κοντά στο -1.



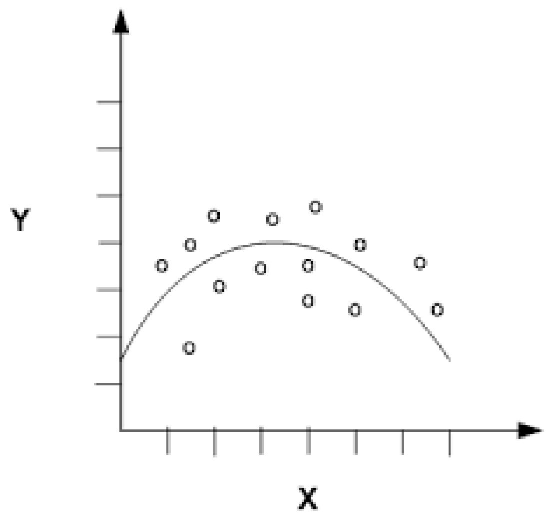
Σχήμα 3. Ισχυρή θετική (αριστερά) και αρνητική (δεξιά) γραμμική συσχέτιση

Στην πρώτη περίπτωση του Σχήματος 3 (αριστερά), η συσχέτιση παραμένει θετική, ενώ ο συντελεστής συσχέτισης λαμβάνει τιμές πολύ κοντά στο +0,7. Στη δεύτερη περίπτωση του ίδιου σχήματος (δεξιά), η συσχέτιση είναι αρνητική και ο συντελεστής συσχέτισης λαμβάνει τιμές πολύ κοντά στο -0,7 (Montgomery & Peck, 1991).



Σχήμα 4. Απουσία συσχέτισης

Τέλος, στη Σχήμα 4 δεν φαίνεται να υπάρχει κάποιου είδους συσχέτιση, με το συντελεστή συσχέτισης να λαμβάνει τιμές κοντά στο 0, ενώ στο Σχήμα 5, φαίνεται να υπάρχει συσχέτιση, αλλά είναι τετραγωνική (Montgomery & Peck, 1991).



Σχήμα 5. Τετραγωνική συσχέτιση

Η ποσοτική μέτρηση της γραμμικής σχέσης μεταξύ δύο μεταβλητών γίνεται με τη χρήση του συντελεστή συσχέτισης (correlation coefficient), ο οποίος λαμβάνει τιμές μεταξύ του -1 και του +1. Η τιμή του συγκεκριμένου δείκτη υπολογίζεται από τον τύπο (Montgomery & Peck, 1991):

$$r = \frac{\sigma_{XY}}{\sigma_x \sigma_Y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{(n(\sum X^2) - (\sum X)^2)(n(\sum Y^2) - (\sum Y)^2)}}$$

όπου είναι $\sum X = \sum_{i=1}^{k} X_i$ και $\sum Y = \sum_{i=1}^{k} Y_i$ είναι τα αθροίσματα των τιμών της μεταβλητής X και της μεταβλητής Y αντίστοιχα, $\sum XY = \sum_{i=1}^{k} X_i Y_i$, $\sum X^2 = \sum_{i=1}^{k} (X_i)^2$ και $\overline{X} = \frac{\sum_{i=1}^{k} X_i}{k}, \overline{Y} = \frac{\sum_{i=1}^{k} Y_i}{k}$ είναι οι μέσες τιμές τους.

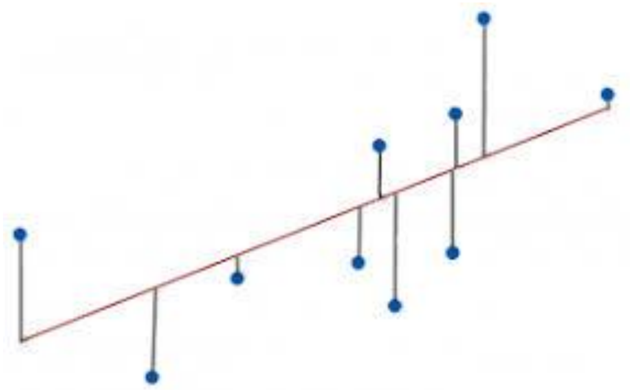
Όσο οι τιμές του *r* προσεγγίζουν το ένα, κατά απόλυτη τιμή, τόσο πιο ισχυρή είναι η γραμμική συσχέτιση, ενώ όσο οι τιμές του προσεγγίζουν το μηδέν, κατά απόλυτη τιμή, τόσο πιο ασθενής είναι η συσχέτιση.

1.3. Απλή Γραμμική Παλινδρόμηση

Το επόμενο βήμα είναι να εκτιμηθεί η γραμμική εξίσωση που περιγράφει τη σχέση μεταξύ της εξαρτημένης μεταβλητής Y και της ανεξάρτητης μεταβλητής X. Με δεδομένο ότι υπάρχουν μόνο δύο μεταβλητές, η στατιστική μέθοδος που χρησιμοποιείται ονομάζεται απλή ανάλυση παλινδρόμησης. Σκοπός της είναι η περιγραφή της σχέσεως μεταξύ των X και Y με ένα μοντέλο της μορφής (Montgomery & Peck, 1991; Kutner et al., 2005):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i. \tag{5}$$

Στο παραπάνω μοντέλο, με Y_i σημειώνεται η τιμή της εξαρτημένης μεταβλητής και με X_i η τιμή της ανεξάρτητης μεταβλητής. Η τιμή του β_0 μας δίνει το σημείο στο οποίο η ευθεία παλινδρόμησης τέμνει τον κατακόρυφο άξονα, δηλαδή το σημείο (0, β_0) το οποίο προκύπτει αν στην παραπάνω εξίσωση μηδενιστεί η τιμή της μεταβλητής X. Ο συντελεστής της μεταβλητής X, δηλαδή ο αριθμός β_1, ονομάζεται κλίση (slope) της ευθείας παλινδρόμησης και ε_i είναι το σφάλμα (error) ή το κατάλοιπο (residual). Το κάθε σφάλμα ε_i παριστάνει τη διαφορά μεταξύ της πραγματικής τιμής της Y και της τιμής της πρόβλεψης που προκύπτει από την παραπάνω εξίσωση ή υπόδειγμα (Kutner et al., 2005).



Σχήμα 6. Σφάλματα σε μια απλή γραμμική παλινδρόμηση

Αυτό γίνεται καλύτερα αντιληπτό από το παραπάνω σχήμα. Τα σημεία

$$(x_i, y_i), \quad i = 1,2, \dots, k$$

αναπαριστούν τα ζευγάρια των πραγματικών τιμών των δύο μεταβλητών, ενώ οι κατακόρυφες αποστάσεις από την ευθεία παλινδρόμησης είναι οι διαφορές των y_i από τις τιμές που εκτιμώνται με βάση το υπόδειγμα.

Για ένα υπόδειγμα απλής γραμμικής παλινδρόμησης, της μορφής (5) γίνονται οι επόμενες τέσσερις υποθέσεις (Kutner et al., 2005; Montgomery & Peck, 1991):

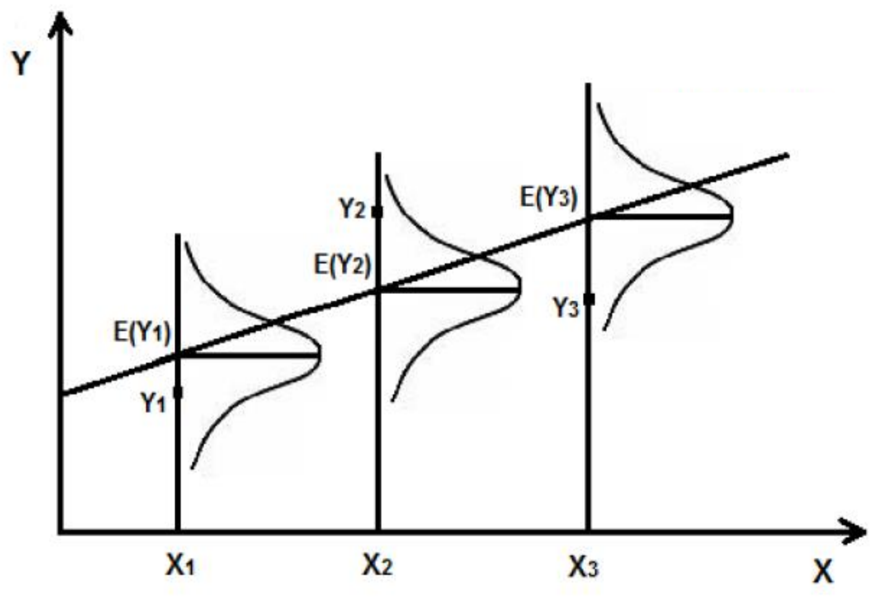
1. Οι ατομικές παρατηρήσεις της εξαρτημένης μεταβλητής είναι μεταξύ τους ανεξάρτητες.

2. Για κάθε συγκεκριμένη τιμή της μεταβλητής X αντιστοιχούν πολλές τιμές της μεταβλητής Y, οι οποίες κατανέμονται κανονικά.
3. Για μέγεθος δείγματος n αντιστοιχούν n κανονικές κατανομές της Y με την ίδια διακύμανση σ^2.
4. Ο μέσος της κάθε κανονικής κατανομής της Y_i ισούται με

$$E(Y_i) = \beta_0 + \beta_1 X_i. \tag{6}$$

Όλοι οι μέσοι βρίσκονται σε μια ευθεία γραμμή που αποτελεί την γραμμή παλινδρόμησης του πληθυσμού. Η ευθεία γραμμή της παλινδρόμησης συνδέει τους μέσους της μεταβλητής Y που αντιστοιχούν στις τιμές της X και έχει εξίσωση:

$$E(Y) = \beta_0 + \beta_1 X \tag{7}$$



Σχήμα 7. Οι μέσοι σε μια ευθεία παλινδρόμησης

Η ευθεία παλινδρόμησης προσδιορίζεται από 2 παραμέτρους, τους β_0 και β_1 οι οποίοι ονομάζονται συντελεστές παλινδρόμησης. Ο συντελεστής β_1 δίνει την κλίση της ευθείας παλινδρόμησης και είναι η μέση μεταβολή της εξαρτημένης μεταβλητής Y που αντιστοιχεί σε μεταβολή της X κατά μία μονάδα. Με απλά λόγια, αν η τιμή της μεταβλητής X αυξηθεί κατά μία μονάδα, τότε από το υπόδειγμα (5), αναμένεται ότι η τιμή της μεταβλητής Y να αυξηθεί ή να μειωθεί κατά β_1, ανάλογα με το πρόσημο του β_1.

Για παράδειγμα, στο υπόδειγμα

$$Y = -1,2 + 0,4X$$

έχουμε β_0=-1,2 και β_1=0,4. Αυτό σημαίνει ότι αν η τιμή της ανεξάρτητης μεταβλητής γίνει ίση με μηδέν (X=0) τότε αναμένεται να είναι Y=-1,2. Αν η τιμή της μεταβλητής X αυξηθεί κατά μία μονάδα τότε αναμένεται η μεταβλητή της Y να αυξηθεί (θετικό πρόσημο στο β_1) κατά 0,4 μονάδες.
Με δεδομένο ότι ισχύει η σχέση (6), το υπόδειγμα (5) λαμβάνει τη μορφή

$$Y_i = E(Y_i) + \varepsilon_i \tag{8}$$

όπου ε_i είναι η τυχαία συνιστώσα, δηλαδή το κατάλοιπο ή το σφάλμα που ενσωματώνει στο μοντέλο όλους τους άλλους παράγοντες, εκτός της X, που επηρεάζουν τη διαμόρφωση της τιμής της Y.

Επομένως, η τιμή της Y που παρατηρούμε προκύπτει ως το άθροισμα δύο συνιστωσών. Η μία είναι η συνιστώσα που προσδιορίζεται από την επίδραση της X πάνω στην Y, και επομένως είναι δυνατόν να προβλεφθεί από την εξίσωση της παλινδρόμησης, και από το κατάλοιπο ή υπόλοιπο που οφείλεται σε άλλους παράγοντες οι οποίοι είτε είναι άγνωστοι, είτε δεν είναι δυνατόν να μετρηθούν.

Τα τυχαία σφάλματα έχουν μηδενική μέση τιμή και θεωρείται ότι έχουν κοινή διακύμανση σ^2 και να είναι ανεξάρτητα μεταξύ τους ανά ζεύγη. Δεδομένου ότι το μόνο τυχαίο στοιχείο στο παραπάνω μοντέλο είναι τα ε_i, αυτές οι υποθέσεις υπονοούν ότι τα Y_i έχουν επίσης κοινή διακύμανση σ^2 και είναι ανεξάρτητα ανά ζεύγη. Για σκοπούς διεξαγωγής ελέγχου σημαντικότητας, θεωρείται ότι τα τυχαία σφάλματα κατανέμονται κανονικά, το οποίο σημαίνει ότι και τα Y_i κατανέμονται κανονικά. Οι συγκεκριμένες υποθέσεις για τα τυχαία σφάλματα αναφέρονται συχνά ως $\varepsilon_i \sim N.I.D.(0, \sigma^2)$ όπου το N.I.D. σημαίνει κανονικά και ανεξάρτητα κατανεμημένα (Normally and Independently Distributed), ενώ οι τιμές που βρίσκονται εντός της παρένθεσης δηλώνουν τη μέση τιμή και τη διακύμανση της κανονικής κατανομής (Kutner et al., 2005; Montgomery & Peck, 1991). Στη παρούσα εργασία θα χρησιμοποιείτε η απλούστερη γραφή $\varepsilon_i : \; N\left(0, \, \sigma^2\right)$.

Το απλό γραμμικό μοντέλο έχει δύο παραμέτρους β_0 και β_1, οι οποίες υπολογίζονται από τα δεδομένα. Εάν δεν υπήρχε το τυχαίο σφάλμα στις τιμές των Y_i, οποιαδήποτε δύο γνωστά σημεία θα μπορούσαν να χρησιμοποιηθούν για να υπολογιστούν με ακρίβεια οι τιμές των παραμέτρων. Η τυχαία μεταβολή της Y, ωστόσο, έχει ως αποτέλεσμα κάθε ζεύγος παρατηρημένων δεδομένων να δίνει διαφορετικά αποτελέσματα. Η μόνη περίπτωση στην οποία οι εκτιμήσεις θα ήταν ίδιες, είναι αυτή που τα παρατηρούμενα δεδομένα θα βρίσκονταν όλα πάνω στην ευθεία παλινδρόμησης Για τον παραπάνω λόγο, απαιτείται μια μέθοδος η οποία θα συνδυάζει όλες τις πληροφορίες προκειμένου να δώσει μια λύση, η οποία θα είναι η καλύτερη δυνατή με βάση κάποιο κριτήριο. Η συγκεκριμένη μέθοδος καλείται μέθοδος εκτίμησης ελαχίστων τετραγώνων (least squares estimation procedure) (Kutner et al., 2005; Montgomery & Peck, 1991).

Η μέθοδος εκτίμησης ελαχίστων τετραγώνων βασίζεται στο κριτήριο ότι η λύση πρέπει να δίνει το μικρότερο δυνατό άθροισμα τετραγωνικών αποκλίσεων μεταξύ των πραγματικών τιμών των Y_i και των εκτιμήσεων των πραγματικών τους μέσων που βρέθηκαν από τη λύση.

Έστω $\hat{\beta}_0$ και $\hat{\beta}_1$ οι αριθμητικές εκτιμήσεις των παραμέτρων β_0 και β_1, αντίστοιχα, και

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \tag{9}$$

είναι η εκτίμηση της μεταβλητής Y για κάθε X_i, i = 1,…, n. Οι τιμές των $\hat{Y}_i$ υπολογίζονται αντικαθιστώντας τους εκτιμητές των παραμέτρων στο μοντέλο που συνδέει το $E(Y_i)$ και X_i, δηλαδή στην εξίσωση

$$E(Y_i) = \beta_0 + \beta_1 X_i + \varepsilon_i. \tag{10}$$

Η μέθοδος ελαχίστων τετραγώνων επιλέγει εκείνα τα $\hat{\beta}_0$ και $\hat{\beta}_1$ τα οποία ελαχιστοποιούν τα υπόλοιπα της μορφής (Kutner et al., 2005; Montgomery & Peck, 1991):

$$SS(Res) = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \varepsilon_i^{\,2} \tag{11}$$

όπου $\varepsilon_i = Y_i - \hat{Y}$ είναι το υπόλοιπο που παρατηρείται για την i παρατήρηση, ενώ με το άθροισμα $\sum_{i=1}^{n} \varepsilon_i^2$ σημειώνεται το άθροισμα όλων αυτών των υπολοίπων (ή σφαλμάτων) για i=1,…,n.

Ο υπολογισμός των $\hat{\beta}_0$ και $\hat{\beta}_1$ γίνεται με χρήση μεθόδων της Ανάλυσης, με σκοπό την εύρεση της τιμής που ελαχιστοποιεί το $SS(Res)$. Οι μερικές παράγωγοι της συγκεκριμένης ποσότητας , ως προς $\hat{\beta}_0$ και $\hat{\beta}_1$, πρέπει να είναι ίσες με μηδέν. Η συγκεκριμένη συνθήκη παρέχει το παρακάτω σύστημα εξισώσεων με δύο αγνώστους, οι οποίες καλούνται κανονικές εξισώσεις.

$$n \cdot \hat{\beta}_0 + \sum X_i \cdot \hat{\beta}_1 = \sum Y_i \tag{12}$$

$$\sum X_i \cdot \hat{\beta}_0 + \sum X^2{}_i \cdot \hat{\beta}_1 = \sum X_i Y_i \tag{13}$$

Λύνοντας το παραπάνω σύστημα εξισώσεων, ως προς $\hat{\beta}_0$ και $\hat{\beta}_1$, βρίσκουμε (Kutner et al., 2005; Montgomery & Peck, 1991):

$$\hat{\beta}_1 = \frac{\sum \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum \left(X_i - \bar{X}\right)^2} \tag{14}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{15}$$

Οι ποσότητες $x_i = X_i - \bar{X}$ και $y_i = Y_i - \bar{Y}$ εκφράζουν τις αποκλίσεις των παρατηρήσεων των μεταβλητών *X* και *Y* από τη δειγματική μέση τιμή τους αντιστοίχως. Οι παρακάτω τύποι είναι πιο βολικοί για τους υπολογισμούς

$$\sum x_i^2 = \sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n} \tag{16}$$

$$\sum x_i y_i = \sum X_i Y_i - \frac{\left(\sum X_i\right)\left(\sum Y_i\right)}{n} \quad . \tag{17}$$

Με βάση τους παραπάνω τύπους, η εκτίμηση της κλίσης $\hat{\beta}_1$ είναι

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \frac{\left(\sum X_i\right)\left(\sum Y_i\right)}{n}}{\sum X_i^{\,2} - \frac{\left(\sum X_i\right)^2}{n}} \,. \tag{18}$$

Οι συγκεκριμένοι συντελεστές παρέχουν την παρακάτω εξίσωση παλινδρόμησης (Kutner et al., 2005; Montgomery & Peck, 1991)

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \,. \tag{19}$$

Με βάση την παραπάνω εξίσωση αντικαθιστώντας τιμές στην μεταβλητή X είναι δυνατόν να εκτιμηθεί η τιμής της μεταβλητής Y.

1.4. Πολλαπλή Γραμμική Παλινδρόμηση

Ένα γραμμικό μοντέλο πολλαπλής παλινδρόμησης, το οποίο συνδέει την εξαρτημένη μεταβλητή *Y* με p ανεξάρτητες μεταβλητές *X*, έχει την παρακάτω μορφή (Kutner et al., 2005; Cohen, Cohen, West & Aiken, 2003; Montgomery & Peck, 1991):

$$Y_{\mathrm{i}} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i. \tag{20}$$

Στο παραπάνω μοντέλο, ο δείκτης *i* υποδηλώνει την μονάδα, από την οποία προέρχονται οι παρατηρήσεις επί της μεταβλητής *Y*, καθώς και από την οποία ελήφθησαν οι *p* ανεξάρτητες μεταβλητές. Ο δεύτερος δείκτης υποδηλώνει την ανεξάρτητη μεταβλητή. Το μέγεθος του δείγματος υποδηλώνεται με n, και ο δείκτης p δηλώνει τον αριθμό των ανεξάρτητων μεταβλητών. Υπάρχουν p+1 παράμετροι β_j, j = 0, ..., p, οι οποίες πρέπει να υπολογιστούν,. Για ευκολία, θα χρησιμοποιείται από εδώ και εξής $p'=p+1$, όπου $n > p$.

Προκειμένου να περιγραφεί το παραπάνω μοντέλο με μορφή πινάκων, απαιτούνται τέσσερις πίνακες (Kutner et al., 2005):

- ο $n\times1$ πίνακας στήλη, $\boldsymbol{Y}$ των παρατηρήσεων επί της εξαρτημένης μεταβλητής Yi,
- ο $n\times p$ πίνακας $\boldsymbol{X}$ που αποτελείται από μια στήλη με μονάδες, η οποία φέρει την ένδειξη **1**, ακολουθούμενη από τις p στήλες των παρατηρήσεων επί των ανεξάρτητων μεταβλητών,
- ο $p\times1$ πίνακας $\boldsymbol{\beta}$ των παραμέτρων που πρέπει να εκτιμηθούν και
- ο $n\times1$ πίνακας $\boldsymbol{\varepsilon}$ των τυχαίων σφαλμάτων.

Με βάση τα παραπάνω, το γραμμικό μοντέλο μπορεί να γραφεί στη μορφή

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{21}$$

ή πιο αναλυτικά στη μορφή

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \qquad (22)$$

Κάθε στήλη του πίνακα $\boldsymbol{X}$ περιέχει τις τιμές για μια συγκεκριμένη ανεξάρτητη μεταβλητή. Τα στοιχεία μιας συγκεκριμένης γραμμής του πίνακα $\boldsymbol{X}$, έστω της γραμμής r, είναι οι συντελεστές των αντίστοιχων παραμέτρων στο β που δίνουν το $E(Y_r)$. Παρατηρήστε ότι το β_0 έχει τον ίδιο, σταθερό πολλαπλασιαστή, δηλαδή το 1, για όλες τις παρατηρήσεις. Επομένως, το διάνυσμα στήλη **1** είναι η πρώτη στήλη του πίνακα $\boldsymbol{X}$. Πολλαπλασιάζοντας την πρώτη γραμμή του πίνακα $\boldsymbol{X}$ με το διάνυσμα $\boldsymbol{\beta}$, και προσθέτοντας το πρώτο στοιχείο επιβεβαιώνει ότι το μοντέλο για την πρώτη παρατήρηση έχει τη μορφή

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_p X_{1p} + \varepsilon_i. \qquad (23)$$

Οι πίνακες στήλη, ή αλλιώς τα διανύσματα $\boldsymbol{Y}$ και $\boldsymbol{\varepsilon}$ είναι τυχαία διανύσματα, δηλαδή με τα στοιχεία τα οποία είναι τυχαίες μεταβλητές. Ο πίνακας $\boldsymbol{X}$ είναι ένας πίνακας με σταθερά στοιχεία. Ένα μοντέλο για το οποίο ο πίνακας $\boldsymbol{X}$ είναι πλήρους τάξης στήλης ονομάζεται μοντέλο πλήρους τάξης.

Το διάνυσμα $\boldsymbol{\beta}$ είναι ένα διάνυσμα άγνωστων παραμέτρων, το οποίο πρέπει να εκτιμηθεί με βάση τα δεδομένα. Κάθε στοιχείο του, β_j, είναι ένας συντελεστής μερικής παλινδρόμησης που αντανακλά την αλλαγή στην εξαρτημένη μεταβλητή ανά μεταβολή μονάδας στην ανεξάρτητη μεταβλητή που βρίσκεται στη στήλη j, με την προϋπόθεση ότι όλες οι υπόλοιπες ανεξάρτητες μεταβλητές διατηρούνται σταθερές. Ο ορισμός του κάθε συντελεστή μερικής παλινδρόμησης εξαρτάται από το σύνολο των ανεξάρτητων μεταβλητών στο μοντέλο. Όποτε απαιτείται σαφήνεια, η αναφορά στο δείκτη του στοιχείου β_{j} επεκτείνεται για να προσδιορίσει ρητά και την ανεξάρτητη μεταβλητή στην οποία εφαρμόζεται ο συντελεστής και τις άλλες ανεξάρτητες μεταβλητές στο μοντέλο. Για παράδειγμα, το στοιχείο $\beta_{2.13}$ θα υποδηλώνει το συντελεστή μερικής παλινδρόμησης για την ανεξάρτητη μεταβλητή $\boldsymbol{X_2}$ σε ένα μοντέλο που περιέχει τις ανεξάρτητες μεταβλητές $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ και $\boldsymbol{X_3}$.

Είναι συνηθισμένο να γίνεται η υπόθεση ότι, τα ε_i είμαι ανεξάρτητα και κανονικά κατανεμημένα (independent and identically distributed-i.i.d.) ως κανονικές τυχαίες μεταβλητές με μέση τιμή μηδέν και διακύμανση σ^2. Εφόσον τα ε_i θεωρούνται ανεξάρτητα μεταξύ τους, η συνδιακύμανση (covariance) μεταξύ των ε_i και ε_j ισούται με μηδέν, για κάθε $i \neq j$. Η από κοινού συνάρτηση πυκνότητας πιθανότητας (joint probability density function) των $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ είναι ίση με (Kutner et al., 2005; Montgomery & Peck, 1991):

$$\prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\varepsilon_i^2}{2\sigma^2}}\right]=\frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n}e^{-\sum_{i=1}^{n}\frac{\varepsilon_i^2}{2\sigma^2}} \tag{24}$$

Το τυχαίο διάνυσμα $\boldsymbol{\varepsilon}=(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ είναι ένα διάνυσμα το οποίο αποτελείται από τις τυχαίες μεταβλητές ε_i .

Δεδομένου ότι τα στοιχεία των πινάκων $\boldsymbol{X}$ και $\boldsymbol{\beta}$ θεωρούνται ότι είναι σταθερά, ο όρος $\boldsymbol{X\beta}$, στο μοντέλο $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, είναι ένα διάνυσμα που περιέχει σταθερές. Επομένως, το $\boldsymbol{Y}$ είναι ένα τυχαίο διάνυσμα, το οποίο προκύπτει ως το άθροισμα του σταθερού διανύσματος $\boldsymbol{X\beta}$ και του τυχαίου διανύσματος $\boldsymbol{\varepsilon}$. Με δεδομένο ότι τα ε_i είναι ανεξάρτητες τυχαίες μεταβλητές, οι οποίες ακολουθούν την κανονική κατανομή με μέση τιμή 0 και διακύμανση σ^2 (δηλαδή ε_i: $N(0, \sigma^2)$), θα ισχύει ότι (Kutner et al., 2005)

1. Κάθε μεταβλητή Y_i είναι μια κανονική τυχαία μεταβλητή με μέση τιμή $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}$ και διακύμανση σ^2.
2. Οι μεταβλητές Y_i είναι ανεξάρτητες μεταξύ τους.

Η συνδιακύμανση των Y_i και Y_j είναι μηδέν για $i \neq j$, ενώ η από κοινού συνάρτηση πυκνότητας πιθανότητας των μεταβλητών $Y_1, \ldots, Y_n$ είναι ίση με

$$\frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n}e^{-\frac{\sum_{i=1}^{n}\left[Y_i-\left(\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_p X_{ip}\right)\right]^2}{2\sigma^2}} . \tag{25}$$

Οι συμβατικοί έλεγχοι υποθέσεων και των εκτιμήσεων των διαστημάτων εμπιστοσύνης των παραμέτρων βασίζονται στην υπόθεση ότι οι εκτιμήσεις είναι κανονικά κατανεμημένες. Έτσι, η υπόθεση της κανονικότητας των ε_i είναι σημαντική για το σκοπό αυτό. Ωστόσο, η κανονικότητα δεν απαιτείται για τους εκτιμητές ελαχίστων τετραγώνων. Ακόμη και αν δεν υπάρχει η κανονικότητα, οι εκτιμητές των ελάχιστων τετραγώνων είναι οι καλύτεροι γραμμικοί και αμερόληπτοι εκτιμητές (Best Linear Unbiased Estimates- BLUE) (Kutner et al., 2005; Montgomery & Peck, 1991).

Οι συγκεκριμένοι εκτιμητές είναι οι καλύτερες με την έννοια ότι είναι αυτές που έχουν την ελάχιστη διακύμανση μεταξύ όλων των γραμμικών αμερόληπτων εκτιμητών. Εάν η κανονικότητα ισχύει, οι εκτιμητές μέγιστης πιθανότητας υπολογίζονται χρησιμοποιώντας το κριτήριο για την εύρεση αυτών των τιμών των παραμέτρων που μεγιστοποιηθούν τη πιθανότητα απόκτησης του συγκεκριμένου δείγματος, η οποία ονομάζεται συνάρτηση πιθανοφάνειας (likelihood function). Η μεγιστοποίηση της συνάρτησης πιθανοφάνειας στην εξίσωση (25) σε σχέση με το διάνυσμα $\boldsymbol{\beta} = (\beta_0, \beta_1,\ldots,\beta_p)^{\mathrm{T}}$ είναι ισοδύναμη με την ελαχιστοποίηση του αθροίσματος των τετραγώνων, το οποίο βρίσκεται στον εκθέτη, και συνεπώς οι εκτιμητές των ελαχίστων τετραγώνων συμπίπτουν με τους εκτιμητές μέγιστης πιθανοφάνειας.

Στη μορφή πινάκων, οι κανονικές εξισώσεις είναι οι παρακάτω (Rawlings, Pantula & Dickey 1998):

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} \tag{26}$$

όπου $\boldsymbol{X}^{T}$ είναι ο ανάστροφος του πίνακα $\boldsymbol{X}$.

Οι κανονικές εξισώσεις έχουν πάντα μία λύση της μορφής

$$\widehat{\boldsymbol{\beta}}= (\boldsymbol{X}^{T}\boldsymbol{X})^{-1}\, \boldsymbol{X}^{T}\boldsymbol{Y} \tag{27}$$

με την προϋπόθεση ότι ο πίνακας $\boldsymbol{X}^{T}\boldsymbol{X}$ αντιστρέφεται, δηλαδή υπάρχει ο πίνακας $(\boldsymbol{X}^{T}\boldsymbol{X})^{-1}$. Τότε οι κανονικές εξισώσεις έχουν μια μοναδική λύση που δίνεται από τον τύπο (27). Με το γινόμενο $\boldsymbol{X}^{T}\boldsymbol{X}$ δημιουργείται ένας πίνακας διαστάσεων $p'\times p'$, όπου $p'=p+1$, του οποίου η κύρια διαγώνιος αποτελείται από τα αθροίσματα των τετραγώνων κάθε ανεξάρτητης μεταβλητής και τα εκτός κυρίας διαγωνίου στοιχεία του είναι τα αθροίσματα των γινομένων μεταξύ ανεξάρτητων μεταβλητών. Η γενική μορφή του συγκεκριμένου πίνακα είναι η παρακάτω (Rawlings, Pantula & Dickey, 1998):

:

$$\begin{pmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{ip} \\ \sum X_{i1} & \sum X^2_{i1} & \sum X_{i1}X_{i2} & \cdots & \sum X_{i1}X_{ip} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X^2_{i2} & \cdots & \sum X_{i2}X_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{ip} & \sum X_{i1}X_{ip} & \sum X_{i2}X_{ip} & \cdots & \sum X^2_{ip} \end{pmatrix} \tag{28}$$

Στην περίπτωση εκείνη που το μοντέλο περιέχει μόνο μία ανεξάρτητη μεταβλητή, ο πίνακας $\boldsymbol{X^T X}$ αποτελείται μόνο από τον πάνω δεξιά 2×2 πίνακα.

Από την άλλη, τα στοιχεία του πίνακα $\boldsymbol{X^T Y}$ είναι τα αθροίσματα των γινομένων μεταξύ της κάθε ανεξάρτητης μεταβλητής και της εξαρτημένης, δηλαδή έχει τη μορφή:

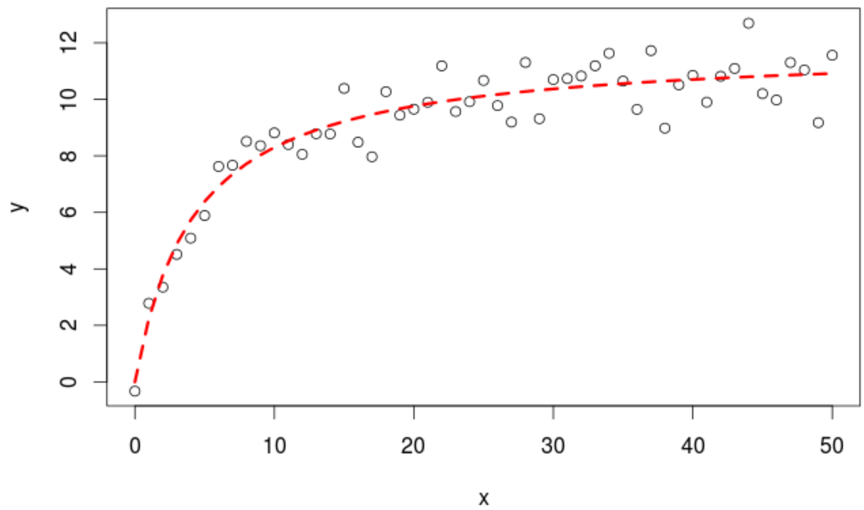
$$\boldsymbol{X^T Y} = \begin{pmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \sum X_{i2}Y_i \\ \vdots \\ \sum X_{ip}Y_i \end{pmatrix} \tag{29}$$

Το πρώτο στοιχείο του παραπάνω πίνακα, δηλαδή το $\sum Y_i$ είναι το άθροισμα των γινομένων μεταξύ διανύσματος που περιέχει τις μονάδες, δηλαδή την πρώτη στήλη του πίνακα $\boldsymbol{X}$ και του διανύσματος $\boldsymbol{Y}$. Στην περίπτωση που υπάρχει μόνο μία ανεξάρτητη μεταβλητή, ο πίνακας $\boldsymbol{X^T Y}$ αποτελείται μόνο από τα δύο πρώτα στοιχεία του παραπάνω πίνακα, οι οποίες είναι τα δεξιά μέλη των δύο κανονικών εξισώσεων στο μοντέλο απλής γραμμικής παλινδρόμησης (Rawlings, Pantula & Dickey, 1998).

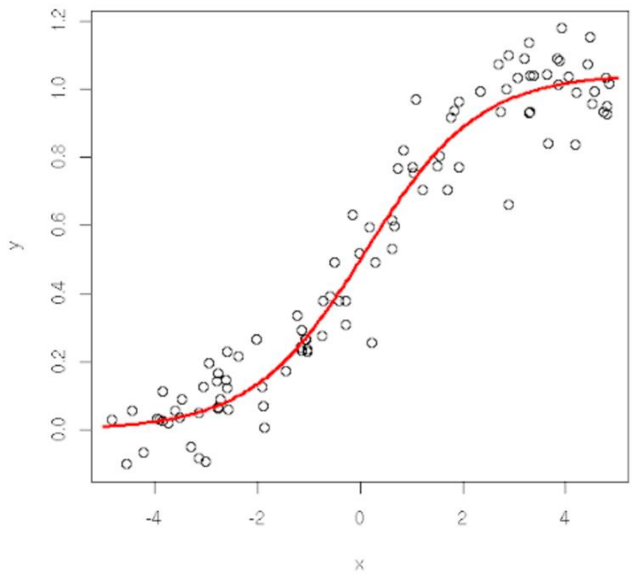
Κεφάλαιο 2. Μη γραμμικά μοντέλα

Ένα γραμμικό μοντέλο βασίζεται στην παραδοχή ότι υπάρχει μια γραμμική σχέση μεταξύ μιας εξαρτημένης και μια ανεξάρτητης μεταβλητής, το οποίο πρακτικά σημαίνει ότι υπάρχει ένας σταθερός ρυθμός μεταβολής. Κάτι τέτοιο όμως δεν είναι απαραίτητο να ισχύει. Για παράδειγμα η συγκέντρωση ενός φαρμάκου στο αίμα είναι πιθανό να μην είναι γραμμική σε σχέση με το χρόνο. Ένα ακόμα παράδειγμα αποτελούν και οι οικονομικές χρονολογικές σειρές, όπως είναι ο δείκτης πληθωρισμού και το Ακαθάριστο Εγχώριο Προϊόν, οι οποίες παρουσιάζουν τάσεις με την πάροδο του χρόνου, οι οποίες δεν είναι απαραίτητο να είναι γραμμικές. Σε περίπτωση λοιπόν, που ο ρυθμός μεταβολής της μέσης τιμής μιας εξαρτημένης μεταβλητής Y δεν είναι σταθερός σε σχέση με την ανεξάρτητη μεταβλητή X, τότε έχουμε ένα μη γραμμικό μοντέλο.

Ένα μη γραμμικό μοντέλο είναι εύκολο να γίνει αντιληπτό από το διάγραμμα διασποράς (scatter plot). Για παράδειγμα, στο Σχήμα 8 παριστάνεται το διάγραμμα διασποράς ενός λογαριθμικού μοντέλου (κορυφή), ενός εκθετικού (μέση) και ενός λογιστικού μοντέλου



Σχήμα 8. Διάγραμμα διασποράς εκθετικού μοντέλου



Σχήμα 9. Διάγραμμα διασποράς λογιστικής παλινδρόμησης

2.1. Είδη μη γραμμικών μοντέλων

2.1.1. Πολυωνυμικά μοντέλα

Η απλούστερη μορφή ενός μη γραμμικού μοντέλου, στο οποίο υπάρχει μια ανεξάρτητη μεταβλητή, είναι το πολυωνυμικό δευτέρου βαθμού (quadratic model), το οποίο έχει την παρακάτω μορφή (Bates & Watts, 1988; Rawlings, Pantula & Dickey, 1998):

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 \tag{1}$$

Η διαφορά με το γραμμικό μοντέλο, είναι ότι στο πολυωνυμικό μοντέλο, εκτός του όρου X, εμφανίζεται και ο όρος X^2. Αξίζει να σημειωθεί ότι το συγκεκριμένο μοντέλο μπορεί να θεωρηθεί και ως μια ειδική περίπτωση ενός μοντέλου πολλαπλής παλινδρόμησης, όπου είναι $X_1 = X$ και $X_2 = X^2$.

Γενικεύοντας, ένα πολυωνυμικό μοντέλο ανώτερης τάξης έχει τη μορφή:

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_p X^p \tag{2}$$

όπου $p \geq 3$.

Και το συγκεκριμένο μοντέλο μπορεί να θεωρηθεί ως μια ειδική περίπτωση ενός μοντέλου πολλαπλής παλινδρόμησης, όπου είναι $X_i=X^i$, για $i=1,...,p$. Ένα μοντέλο της τελευταίας μορφής καλείται πολυωνυμικό μοντέλο p βαθμού.

Μια σημαντική διαφορά ενός πολυωνυμικού μοντέλου, η οποία το διακρίνει από τα άλλα μοντέλα πολλαπλής παλινδρόμησης, είναι ότι ο μέσος όρος της εξαρτημένης μεταβλητής είναι συνάρτηση μιας μόνο ανεξάρτητης μεταβλητής. Ακόμα και αν οι ανεξάρτητες μεταβλητές σε ένα γενικό μοντέλο πολλαπλής παλινδρόμησης είναι πιθανό να σχετίζονται μεταξύ τους, συνήθως δεν θεωρούνται ότι η μία είναι συνάρτηση της άλλης. Αντιθέτως, σε ένα πολυωνυμικό μοντέλο p βαθμού, οι ανεξάρτητες μεταβλητές είναι συναρτήσεις μίας και μόνο ανεξάρτητης μεταβλητής, αφού είναι όλες δυνάμεις της μεταβλητής X (Bates & Watts, 1988).

Για παράδειγμα, στο μοντέλο της παρακάτω μορφής

$$E\left(Y\right)=\beta_0 +\beta_1 X_1+\beta_2 X_2 \qquad (3)$$

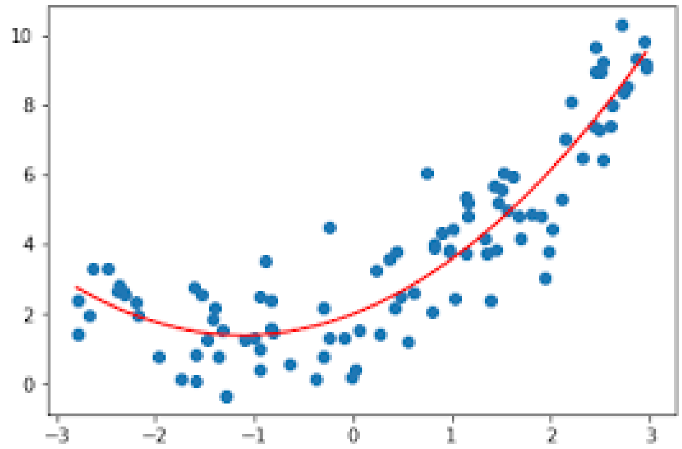
ο συντελεστής β_1 ερμηνεύει τη μεταβολή του μέσου όρου της εξαρτημένης μεταβλητής, όταν η τιμή της μεταβλητής X_1 αυξηθεί κατά μία μονάδα, θεωρώντας την τιμή της μεταβλητής X_2 σταθερή. Όμοια, ο συντελεστής β_2 τη μεταβολή του μέσου όρου της εξαρτημένης μεταβλητής, όταν η τιμή της μεταβλητής X_2 αυξηθεί κατά μία μονάδα, θεωρώντας την τιμή της μεταβλητής X_1 σταθερή. Ωστόσο, σε ένα πολυωνυμικό μοντέλο δευτέρου βαθμού, όπου είναι $X_2=X^2$, τότε μια μεταβολή κατά μία μονάδα στην τιμή της μεταβλητής X_1 οδηγεί και σε μεταβολή της τιμής και της μεταβλητής X_2.

Σε ένα μοντέλο δευτέρου βαθμού, της μορφής (3) ο ρυθμός μεταβολής της μέσης τιμής της εξαρτημένης μεταβλητής, ως συνάρτηση της μεταβλητής X, καλείται κλίση στο X ή παράγωγος στο X (Bates & Watts, 1988). Από την Ανάλυση, είναι γνωστό ότι η παράγωγος της παραπάνω σχέσης, ως προς X, ισούται με

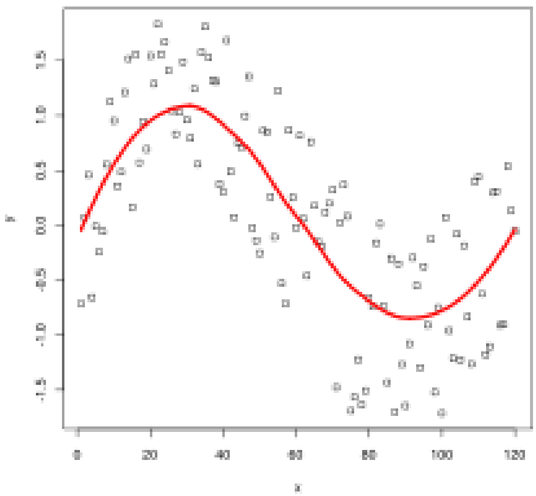
$$\frac{dE\left(Y\right)}{dX}=\beta_1+2\beta_2 X \qquad (4)$$

Η παραπάνω σχέση σημαίνει ότι η κλίση της *E(Y)* εξαρτάται από την τιμή της ανεξάρτητης μεταβλητής, δηλαδή της X. Η τιμή της παραμέτρου β_1 παρέχει την κλίση της μόνο για X = 0. Από την άλλη, η τιμή της παραμέτρου β_2 ισούται με το ήμισυ του ρυθμού μεταβολής της κλίσης του $E(Y)$.

Κάθε πολυωνυμικό μοντέλο μιας μεταβλητής μπορεί να παρασταθεί από ένα καμπυλόγραμμο γράφημα σε ένα σύστημα δύο αξόνων, παρά από μια επιφάνεια ενός χώρου μεγαλύτερης διάστασης, όπως σε ένα μοντέλο πολλαπλής παλινδρόμησης, αφού η εξαρτημένη μεταβλητή είναι συνάρτηση μιας μόνο ανεξάρτητης μεταβλητής (βλέπε Σχήμα 9 και 10).



Σχήμα 10. Διάγραμμα διασποράς ενός πολυωνυμικού μοντέλου δευτέρου βαθμού



Σχήμα 11. Διάγραμμα διασποράς ενός πολυωνυμικού μοντέλου ανωτέρου βαθμού

2.1.2. Τριγωνομετρικά μοντέλα

Είναι πιθανό, ιδιαίτερα σε χρονολογικές σειρές, οι τιμές της μεταβλητής απόκρισης (*Y*t), σε σχέση με τον χρόνο (*t*), όπως για παράδειγμα στην περίπτωση των θερμοκρασιών σε μια περιοχή, να παρουσιάζουν περιοδική συμπεριφορά, η οποία επαναλαμβάνεται κάθε *s* χρονικές περιόδους. Το ίδιο ισχύει και στις οικονομικές χρονολογικές σειρές, στις οποίες εμφανίζονται συχνά περιοδικές συμπεριφορές που αντικατοπτρίζουν τους λεγόμενους «επιχειρηματικούς κύκλους» (Rawlings, Pantula & Dickey, 1998).

Τριγωνομετρικές συναρτήσεις όπως είναι οι $sin(\omega t)$ και $cos(\omega t)$ είναι περιοδικές, σε σχέση με το χρόνο, περίοδο $T = \frac{2\pi}{\omega}$. Για παράδειγμα, αυτό σημαίνει ότι οι τιμές της συνάρτησης $sin(\omega t)$ είναι ίδιες με τις τιμές της συνάρτησης $sin\left[\omega\left(t + \frac{2\pi}{\omega}i\right)\right]$ για i=1, 2, ... Επομένως, χρονοσειρές με περιοδική συμπεριφορά είναι δυνατόν να μοντελοποιηθούν τριγωνομετρικά χρησιμοποιώντας ένα τριγωνομετρικό μοντέλο παλινδρόμησης.

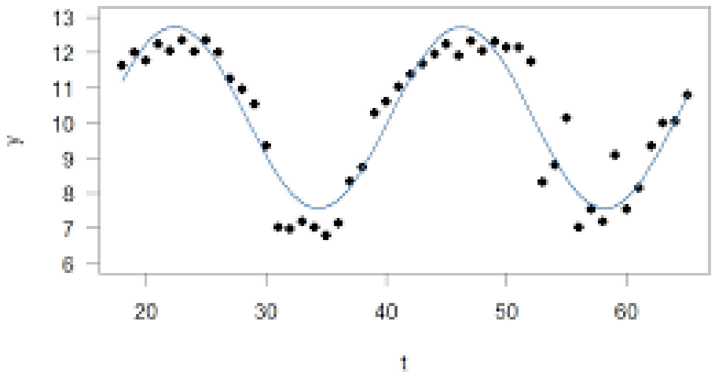
Για παράδειγμα, ένα τριγωνομετρικό μοντέλο παλινδρόμησης μπορεί να έχει τη μορφή

$$Yt = \beta_0 + \beta_1 \, cos\left(\frac{2\pi t}{4}\right) + \beta_2 \, sin\left(\frac{2\pi t}{4}\right) + \beta_3 \, cos(\pi t) \; + \varepsilon_t \tag{5}$$

Κανονικά, οι τριγωνομετρικοί όροι με τα sin και cos εμφανίζονται σε ζευγάρια. Στο παραπάνω μοντέλο, ο όρος $sin(\pi t)$ παραλείπεται, αφού είναι ίσος με 0. Ακόμα ο όρος β_0 (intercept) μπορεί να γραφεί και ως $\beta_0 = cos(0t)$. Και το παραπάνω μοντέλο μπορεί να θεωρηθεί ως μία ειδική περίπτωση του μοντέλου πολλαπλής παλινδρόμησης, με

$$X_{t1} = cos\left(\frac{2\pi t}{4}\right),\; X_{t2} = sin\left(\frac{2\pi t}{4}\right) \text{ και } X_{t3} = cos(\pi t)$$

Στο Σχήμα 12 φαίνεται η περιοδικότητα των τιμών, η οποία προσομοιάζει σε μεγάλο βαθμό, με ένα ημιτονοειδές μοντέλο.



Σχήμα 12. Διάγραμμα διασποράς ενός ημιτονοειδούς μοντέλου παλινδρόμησης

2.1.3. Εκθετικά και λογιστικά μοντέλα

Δύο παραδείγματα μοντέλων μη γραμμικής παλινδρόμησης που χρησιμοποιούνται ευρέως στην πράξη είναι τα εκθετικά μοντέλα παλινδρόμησης (exponential regression models) και τα λογιστικά μοντέλα παλινδρόμησης (logistic regression models) (Rawlings, Pantula & Dickey, 1998), τα οποία παρουσιάζονται αναλυτικά στη συνέχεια.

Ένα ευρέως χρησιμοποιούμενο μοντέλο μη γραμμικής παλινδρόμησης είναι το εκθετικό μοντέλο παλινδρόμησης. Όταν υπάρχει μόνο μία εξαρτημένη μεταβλητή, η γενική μορφή ενός τέτοιου μοντέλου παλινδρόμησης με τους όρους σφάλματος οι οποίοι είναι κανονικά κατανεμημένοι είναι η εξής:

$$Y_i = \beta_0 exp\left(\beta_1 X_i\right) + \varepsilon_i \quad (6)$$

όπου β_0 και β_1 είναι οι παράμετροι του μοντέλου, οι τιμές X_i είναι γνωστές και τα σφάλματα ε_i είναι ανεξάρτητα και ακολουθούν την κανονική κατανομή με μέση τιμή μηδέν και διακύμανση σ^2, δηλαδή είναι $\varepsilon_i :\ N\left(0,\ \sigma^2\right)$.

Η συνάρτηση απόκρισης (response function) ενός τέτοιου μοντέλου έχει τη μορφή: (Rawlings, Pantula & Dickey, 1998):

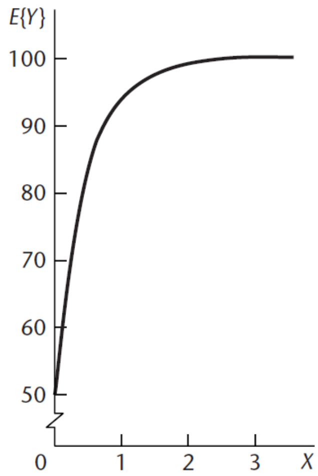
$$f\left(X, \boldsymbol{\beta}\right) = \beta_0 exp\left(\beta_1 X\right) \quad (7)$$

όπου $\boldsymbol{\beta} = \left(\beta_0, \beta_1\right)$.

Μια πιο γενική μορφή ενός τέτοιου μοντέλου έχει τη μορφή $Y_i = \beta_0 + \beta_1 exp(\beta_2 X_i) + \varepsilon_i$ (8), όπου είναι $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$, με συνάρτηση απόκρισης

$$f(X, \boldsymbol{\beta}) = \beta_0 + \beta_1 exp(\beta_2 X). \tag{9}$$

Ένα μοντέλο εκθετικής παλινδρόμησης χρησιμοποιείται συνήθως στις μελέτες ανάπτυξης όπου ο ρυθμός ανάπτυξης, σε δεδομένο χρόνο *X*, είναι ανάλογος με το μέγεθος της ανάπτυξης που απομένει. Μια άλλη χρήση αυτής της παλινδρόμησης είναι όταν αναζητούμε τη συσχέτιση της συγκέντρωσης μιας ουσίας (*Y)* με το χρόνο που έχει περάσει *(X)* (Rawlings, Pantula & Dickey, 1998). Στο Σχήμα 13 παριστάνεται το διάγραμμα του εκθετικού μοντέλου παλινδρόμησης $E(Y) = 100 - 50exp(-2X)$.



Σχήμα 13. Διάγραμμα του εκθετικού μοντέλου παλινδρόμησης

Ένα άλλο σημαντικό μοντέλο μη γραμμικής παλινδρόμησης είναι το μοντέλο λογιστικής ανάπτύξης. Αυτό το μοντέλο με μία ανεξάρτητη μεταβλητή και σφάλματα τα οποία είναι κανονικά κατανεμημένα, έχει την παρακάτω μορφή: (Rawlings, Pantula & Dickey, 1998).

$$Y_i = \frac{\beta_0}{1 + \beta_1 exp(\beta_2 X_i)} + \varepsilon_i \tag{10}$$

όπου και πάλι οι όροι σφάλματος ε_i είναι κανονικά κατανεμημένοι και με σταθερή διακύμανση σ^2. Σε αυτήν την περίπτωση η συνάρτηση απόκρισης έχει την παρακάτω μορφή:

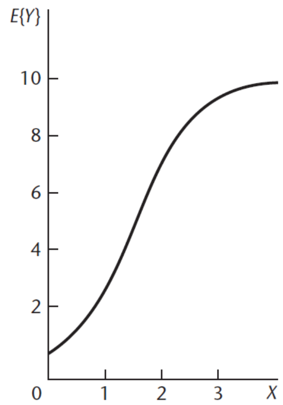
$$f\left(X,\boldsymbol{\beta}\right)=\frac{\beta_0}{1+\beta_1 exp\left(\beta_2 X\right)} \tag{11}$$

όπου $\boldsymbol{\beta}=\left(\beta_0,\beta_1,\beta_2\right)$.

Το μοντέλο λογιστικής ανάπτυξης έχει χρησιμοποιηθεί σε δημογραφικές μελέτες, όπως για παράδειγμα, για την μελέτη του πληθυσμού ενός είδους *(Y)* σε σχέση με το χρόνο *(X)* (Rawlings, Pantula & Dickey, 1998). Στο Σχήμα 14 παριστάνεται το διάγραμμα μιας συνάρτησης απόκρισης, με τιμές παραμέτρων β_0=10, β_1=20 και β_2=-2, δηλαδή πρόκειται για το μοντέλο

$$E(Y) = \frac{10}{1+20exp\left(-2\mathrm{X}\right)}$$

Σημειώστε ότι η παράμετρος β_0=10 αντιπροσωπεύει τη μέγιστη τιμή ανάπτυξης την οποία προσεγγίζει η μεταβλητή *Y*, και για αυτό το λόγο η καμπύλη πλησιάζει ασυμπωτικά την ευθεία *y*=10, καθώς η τιμή της μεταβλητής *X* τείνει στο άπειρο.



Σχήμα 14. Διάγραμμα του λογιστικού μοντέλου ανάπτυξης

Ένα μοντέλο λογιστικής ανάπτυξης χρησιμοποιείται επίσης ευρέως όταν η μεταβλητή απόκρισης είναι ποιοτική. Ένα παράδειγμα αυτής της χρήσης του μοντέλου λογιστικής ανάπτυξης μπορεί να προβλέψει αν ένα νοικοκυριό θα αγοράσει ένα νέο αυτοκίνητο το συγκεκριμένο έτος, δηλαδή αν θα το αγοράσει ή όχι βάσει ανεξάρτητων μεταβλητών (predictors) όπως είναι η ηλικία του αυτοκινήτου που ήδη έχει, το εισόδημα του νοικοκυριού, καθώς και ο αριθμός των μελών του.

Στο παραπάνω μοντέλο η μεταβλητή απόκρισης, ή αλλιώς η εξαρτημένη μεταβλητή, (θα ή δεν θα αγοράσει αυτοκίνητο) είναι ποιοτική και αντιπροσωπεύεται από μεταβλητή δείκτη με τιμές 0 και 1 αντίστοιχα. Συνεπώς, τα σφάλματα δεν κατανέμονται κανονικά με σταθερή διακύμανση.

2.2.Μη γραμμική παλινδρόμηση

Τα μοντέλα μη γραμμικής παλινδρόμησης έχουν την ίδια βασική μορφή με εκείνη που αναφέρθηκε στις ενότητες των μοντέλων γραμμικής παλινδρόμησης. Συγκεκριμένα, ένα μη γραμμικό μοντέλο έχει τη γενική μορφή (Seber & Wild, 1989):

$$Y_i = f(\boldsymbol{X}_i, \boldsymbol{\beta}) + \varepsilon_i \tag{11}$$

όπου $\boldsymbol{X}_i$ είναι το διάνυσμα των προβλεπουσών μεταβλητών (predictors) και $\boldsymbol{\beta}$ το διάνυσμα των παραμέτρων.

Μια παρατήρηση Y_i εξακολουθεί να είναι το άθροισμα μιας μέσης απόκρισης $f(\boldsymbol{X}_i, \boldsymbol{\beta})$ που δίνεται από το μη γραμμικό όρο $f(\boldsymbol{X}, \boldsymbol{\beta})$, ενώ πάλι περιλαμβάνει και τα σφάλματα ε_i. Τα σφάλματα θεωρούνται ότι έχουν μέση τιμή ίση με μηδέν, σταθερή διακύμανση και είναι ασυσχέτιστα, όπως και στα μοντέλα γραμμικής παλινδρόμησης. Συχνά χρησιμοποιείται ένα μοντέλο κανονικών σφαλμάτων (normal error model), το οποίο υποθέτει ότι οι όροι σφάλματος είναι ανεξάρτητες κανονικές τυχαίες μεταβλητές με σταθερή διακύμανση (Seber & Wild, 1989).

Δύο παραδείγματα μοντέλων μη γραμμικής παλινδρόμησης που χρησιμοποιούνται ευρέως στην πράξη είναι τα εκθετικά μοντέλα παλινδρόμησης (exponential regression models) και τα λογιστικά μοντέλα παλινδρόμησης (logistic regression models), τα οποία αναφέρθηκαν παραπάνω και θα παρουσιαστούν αναλυτικά στη συνέχεια.

Όπως φάνηκε από τα δύο τελευταία παραδείγματα μοντέλων μη γραμμικής παλινδρόμησης, αυτά είναι παρόμοια, σε γενική μορφή, με τα μοντέλα γραμμικής παλινδρόμησης. Κάθε παρατήρηση Y_i θεωρείται ότι είναι το άθροισμα μιας μέσης απόκρισης $f(\boldsymbol{X}_i, \boldsymbol{\beta})$ η οποία βασίζεται σε μια δεδομένη μη γραμμική συνάρτηση απόκρισης και ένα τυχαίο σφάλμα ε_i. Επιπλέον, τα σφάλματα ε_i συχνά θεωρούνται ότι είναι ανεξάρτητες και κανονικές τυχαίες μεταβλητές με σταθερή διασπορά.

Μια σημαντική διαφορά των μοντέλων μη γραμμικής παλινδρόμησης είναι ότι ο αριθμός των παραμέτρων ή συντελεστών παλινδρόμησης δεν σχετίζεται απαραίτητα άμεσα με τον αριθμό των ανεξάρτητων μεταβλητών στο μοντέλο. Σε μοντέλα γραμμικής παλινδρόμησης, αν υπάρχουν p-1 ανεξάρτητες μεταβλητές X, τότε υπάρχουν p συντελεστές παλινδρόμησης. Όμως, σε ένα μοντέλο εκθετικής παλινδρόμησης (8) έχουμε μία ανεξάρτητη μεταβλητή X, αλλά τρεις συντελεστές παλινδρόμησης, τους β_0, β_1 και β_2. Το ίδιο ισχύει και στο μοντέλο λογιστικής παλινδρόμησης της μορφής (10) (Seber & Wild, 1989).

Επομένως, υποδηλώνουμε τώρα τον αριθμό των μεταβλητών X σε ένα μοντέλο μη γραμμικής παλινδρόμηση με το δείκτη q, αλλά συνεχίζουμε να υποδηλώνουμε τον αριθμό των συντελεστών παλινδρόμησης στην συνάρτηση απόκρισης με το δείκτη p'. Έτσι λοιπόν, στο μοντέλο εκθετικής παλινδρόμησης (8) υπάρχουν p'=3 συντελεστές παλινδρόμησης και q=1 ανεξάρτητες μεταβλητές.

Με αυτές τις παραδοχές, η γενική μορφή ενός μη γραμμικού μοντέλου, είναι η παρακάτω:

$$Y_\mathrm{i} = f(\boldsymbol{X}_\mathrm{i}, \boldsymbol{\beta}) + \varepsilon_\mathrm{i} \tag{12}$$

όπου $\boldsymbol{X}_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{iq} \end{pmatrix}$ και $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p\text{-}1} \end{pmatrix}$. Πρακτικά, το πρώτο διάνυσμα μπορεί να θεωρηθεί ως ένας $q\times1$ πίνακας, ενώ το δεύτερο ως ένας $p\times1$ πίνακας (Rawlings, Pantula & Dickey, 1998).

2.3.Μετασχηματισμοί σε μη γραμμικά μοντέλα

Κάποιες μη γραμμικές συναρτήσεις απόκρισης μπορούν να γραμμικοποιηθούν με κατάλληλους μετασχηματισμούς, οι οποίες συχνά καλούνται ενδογενώς γραμμικές συναρτήσεις απόκρισης (intrinsically linear response functions).

Για παράδειγμα, η εκθετική συνάρτηση απόκρισης

$$f\left(X,\boldsymbol{\beta}\right) = \beta_0 exp\left(\beta_1 X\right) \tag{13}$$

είναι μια εγγενώς γραμμική συνάρτηση απόκρισης επειδή μπορεί να γραμμικοποιηθεί από το λογαριθμικό μετασχηματισμό

$$\ln f\left(X,\boldsymbol{\beta}\right)=\ln\beta_0+\beta_1 X \ . \tag{14}$$

Αυτή η μετασχηματισμένη συνάρτηση απόκρισης μπορεί να αναπαρασταθεί στη μορφή ενός γραμμικού μοντέλου όπως το παρακάτω:

$$g\left(X,\boldsymbol{\beta}\right)=\beta_0'+\beta_1'X \tag{15}$$

όπου $g\left(X,\boldsymbol{\beta}\right)=\ln f\left(X,\boldsymbol{\beta}\right)$, $\beta_0'=\ln\beta_0$ και $\beta_1'=\beta_1$.

Επειδή μια μη γραμμική συνάρτηση απόκρισης είναι εγγενώς γραμμική δεν σημαίνει απαραίτητα ότι η γραμμική παλινδρόμηση είναι κατάλληλη. Ο λόγος είναι ότι ο γραμμικός μετασχηματισμός της συνάρτησης απόκρισης επηρεάζει τα σφάλματα στο μοντέλο. Για παράδειγμα, ας θεωρήσουμε το ακόλουθο μοντέλο εκθετικής παλινδρόμησης, με τους συνήθεις όρους σφάλματος που έχουν σταθερή διακύμανση, δηλαδή ένα μοντέλο της μορφής (6). Ένας λογαριθμικός μετασχηματισμός της μεταβλητής *Y* για τη γραμμικοποίηση της συνάρτησης απόκρισης θα επηρεάσει τον κανονικά κατανεμημένο όρο σφάλματος, ε_i, έτσι ώστε ο όρος σφάλματος στο γραμμικοποιημένο μοντέλο να μην είναι πλέον κανονικά κατανεμημένος με σταθερή διακύμανση. Για αυτό είναι σημαντικό κάθε μη γραμμικό μοντέλο που έχει γραμμικοποιηθεί, να εξετάζεται ως προς τη συνθήκη της κανονικότητας των σφαλμάτων. Γενικά, μπορεί να αποδειχθεί ότι το μη γραμμικό μοντέλο παλινδρόμησης είναι προτιμότερο από την γραμμική μορφή του, η οποία προκύπτει μετά από ένα γραμμικό μετασχηματισμό.

Γενικά, στο μοντέλο (12), η συνάρτηση $f\left(X,\boldsymbol{\beta}\right)$ περιέχει τις παραμέτρους κατά ένα μη-γραμμικό τρόπο. Υπάρχουν περιπτώσεις στις οποίες η μη-γραμμική παλινδρόμηση, με τον κατάλληλο μετασχηματισμό ή συνδυασμό μετασχηματισμών, μπορεί να καταλήξει σε γραμμική. Τότε το μη-γραμμικό μοντέλο μετασχηματίζεται σε μοντέλο γραμμικής μορφής, όπως αυτό παρουσιάστηκε στο προηγούμενο κεφάλαιο. Οι Draper & Smith (1988) αναφέρονται σε ένα τέτοιο μοντέλο ως ενδογενώς γραμμικό μοντέλο (intrinsically linear).

Για παράδειγμα, αν έχουμε μια μη γραμμική συνάρτηση εκθετικής μορφής, όπως είναι η $Y=\beta_1 e^{\beta_2 X}$, (με $\beta_1>0$) λογαριθμίζοντας και τα δύο μέλη παίρνουμε

$$\ln Y = \ln \beta_1 e^{\beta_2 X} \Rightarrow \ln Y = \ln \beta_1 + \beta_2 X \ ,$$

οπότε ένας μετασχηματισμός που θα μας δώσει μια γραμμική συνάρτηση είναι να θέσουμε $Y' = \ln Y$, οπότε και καταλήγουμε στην $Y' = \ln \beta_1 + \beta_2 X$

Αν έχουμε μια μη γραμμική συνάρτηση, σε μορφή δύναμης, όπως είναι η $Y = \beta_1 X^{\beta_2}$, (με $\beta_1 > 0$) λογαριθμίζοντας και τα δύο μέλη παίρνουμε

$$\ln Y = \ln \beta_1 X^{\beta_2} \Rightarrow \ln Y = \ln \beta_1 + \beta_2 \ln X \ ,$$

οπότε ένας μετασχηματισμός που θα μας δώσει μια γραμμική συνάρτηση είναι να θέσουμε $Y' = \ln Y$ και $X' = \ln X$ οπότε και καταλήγουμε στην $Y' = \ln \beta_1 + \beta_2 X'$.

Αν έχουμε μια μη γραμμική συνάρτηση, σε μορφή κλάσματος, όπως είναι η $Y = \dfrac{X}{\beta_1 X + \beta_2}$, τότε αν αντιστρέψουμε έχουμε

$$\frac{1}{Y} = \frac{\beta_1 X + \beta_2}{X} \Rightarrow \frac{1}{Y} = \beta_1 + \beta_2 \frac{1}{X} \ ,$$

οπότε ένας μετασχηματισμός που θα μας δώσει μια γραμμική συνάρτηση είναι να θέσουμε $Y' = \dfrac{1}{Y}$ και $X' = \dfrac{1}{X}$ (προφανώς για μη μηδενικές τιμές των *Y* και *X*), οπότε και είναι καταλήγουμε στην $Y' = \beta_1 + \beta_2 X'$

Ωστόσο, δεν έχουν όλα τα μη-γραμμικά μοντέλα αυτή την «ικανότητα» και δεν είναι δυνατόν να εκφραστούν στη μορφή ενός γραμμικού, ως προς τις παραμέτρους, μοντέλου. Τα συγκεκριμένα μοντέλα ανήκουν στην κατηγορία των ενδογενώς μη-γραμμικών μοντέλων (intrinsically nonlinear) (Draper & Smith, 1988).

Ας δούμε όμως ποια είναι η σημαντική τους διαφορά. Ας θεωρήσουμε το μοντέλο με εξίσωση $Y = \beta_1 e^{\beta_2 X} \varepsilon$, με $\beta_1 > 0$ και *ε* να είναι τα υπόλοιπα ή τα σφάλματα, με τις συνήθεις υποθέσεις. Παρατηρείστε ότι στο παραπάνω μοντέλο, το *ε* είναι ένα πολλαπλασιαστικό τυχαίο σφάλμα. Αν λογαριθμίσουμε και τα δύο μέλη έχουμε:

$$Y = \beta_1 e^{\beta_2 X} \varepsilon \Rightarrow \ln Y = \ln\left(\beta_1 e^{\beta_2 X} \varepsilon\right) = \ln \beta_1 + \beta_2 X + \ln \varepsilon$$

Αν τώρα θέσουμε $Y' = \ln Y,\ \varepsilon' = \ln\varepsilon,\ \beta_1' = \ln\beta_1$, τότε η τελευταία εξίσωση γίνεται

$$Y' = \beta_1' + \beta_2 X + \varepsilon'$$

το οποίο είναι ένα πλήρως γραμμικό μοντέλο. Άρα είναι ένα ενδογενώς γραμμικό μοντέλο. Με δεδομένο ότι είναι γραμμικό, ακολουθούμε τις συνήθεις μεθόδους της γραμμικής παλινδρόμησης και λαμβάνουμε τα αντίστοιχα συμπεράσματα. Για να γίνει όμως αυτό, θα πρέπει να υποθέσουμε ότι το ε, στο αρχικό μοντέλο, ακολουθεί τη λογαριθμική κανονική κατανομή, έτσι ώστε το $\varepsilon' = \ln\varepsilon$ να ακολουθεί την κανονική κατανομή, όπως και πρέπει να συμβαίνει για τα σφάλματα σε ένα γραμμικό μοντέλο.

Από την άλλη, ας θεωρήσουμε το μοντέλο με εξίσωση $Y = \beta_1 e^{\beta_2 X} + \varepsilon$, με $\beta_1 > 0$ και ε να είναι τα υπόλοιπα ή τα σφάλματα, με τις συνήθεις υποθέσεις. Παρατηρείστε ότι στο παραπάνω μοντέλο, το ε είναι ένα προσθετικό τυχαίο σφάλμα. Αν λογαριθμίσουμε και τα δύο μέλη έχουμε

$$Y = \beta_1 e^{\beta_2 X} + \varepsilon \Rightarrow \ln Y = \ln\left(\beta_1 e^{\beta_2 X} + \varepsilon\right).$$

Από το τελευταίο μοντέλο μπορεί να προκύψει γραμμικό μοντέλο, όπως στην προηγούμενη περίπτωση, μόνο κατά προσέγγιση. Το συγκεκριμένο μοντέλο είναι ενδογενώς μη-γραμμικό μοντέλο.

Από το παραπάνω παράδειγμα, διαπιστώνεται ότι ο μετασχηματισμός ενός μη-γραμμικού μοντέλου σε γραμμικό, είναι άμεσα συνδεδεμένος με τη φύση του τυχαίου σφάλματος. Προκύπτει λοιπόν, ότι πρέπει να είμαστε συγκεκριμένοι για τον τρόπο με τον οποίο το σφάλμα εισέρχεται στο μοντέλο, πριν αποφασίσουμε αν μπορεί να γίνει ο μετασχηματισμός του σε γραμμικό ή όχι.

Προφανώς, η δυνατότητα χρήσης ενός μετασχηματισμού, μέσω του οποίου από ένα μη-γραμμικό μοντέλο προκύπτει ένα γραμμικό, είναι ιδιαίτερα χρήσιμη γιατί είναι δυνατόν στη συνέχεια να εφαρμοστεί η θεωρία της γραμμικής παλινδρόμησης. Αυτό είναι και το κύριο πλεονέκτημα των ενδογενώς γραμμικών μοντέλων. Έτσι η διαδικασία προσαρμογής ή ανάλυσης του συγκεκριμένου τύπου μοντέλου είναι πιο εύκολη, αφού οι άγνωστοι παράμετροι του είναι δυνατόν να εκτιμηθούν μέσω των απλών διαδικασιών της γραμμικής παλινδρόμησης, που περιεγράφηκαν στο προηγούμενο κεφάλαιο, παρά μέσω της μη-γραμμικής εκτίμησης, η οποία θα παρουσιαστεί σε επόμενη ενότητα.

Παρόλα αυτά, ο χειρισμός του μετασχηματισμένου μοντέλου απαιτεί προσοχή. Συγκεκριμένα, η εφαρμογή της μεθόδου των ελαχίστων τετραγώνων, στο μετασχηματισμένο μοντέλο, μας δίνει εκτιμηθέντες συντελεστές, οι οποίοι εκτιμητές ελαχίστων τετραγώνων, αλλά για το μετασχηματισμένο μοντέλο και όχι για το αρχικό, δηλαδή το μη-γραμμικό. Έτσι λοιπόν, μετά την προσαρμογή, είναι σημαντικό να ελεγχθούν τα υπόλοιπα έτσι ώστε να ελέγξουμε αν πράγματι εξακολουθούν να ισχύουν οι υποθέσεις της κλασσικής μεθόδου των ελαχίστων τετραγώνων στο μετασχηματισμένο μοντέλο (Rawlings, Pantula & Dickey, 1998).

Σε αυτό το σημείο θα πρέπει να αναφερθεί ότι είναι προτιμότερο να προσαρμόσουμε ένα μη-γραμμικό μοντέλο, εφόσον αυτό είναι δυνατόν, παρά να προσαρμόσουμε ένα μοντέλο το οποίο ναι μεν θα είναι γραμμικό, αλλά λιγότερο ρεαλιστικό. Έτσι λοιπόν, διατηρώντας την κανονική μορφή του μοντέλου, χωρίς να το μετασχηματίσουμε, επιτυγχάνεται μια καλύτερη ερμηνεία των φυσικών ιδιοτήτων των παραμέτρων του μοντέλου, το οποίο είναι και το ζητούμενο σε μια ανάλυση παλινδρόμησης (Rawlings, Pantula & Dickey, 1998).

2.4.Η εκτίμηση των παραμέτρων παλινδρόμησης

Η εκτίμηση των παραμέτρων ενός μοντέλου μη γραμμικής παλινδρόμησης πραγματοποιείται, συνήθως, με τη μέθοδο των ελαχίστων τετραγώνων (method of least squares) ή τη μέθοδο μέγιστης πιθανοφάνειας (method of maximum likelihood), όπως και στα μοντέλα γραμμικής παλινδρόμησης. Επίσης, όπως και στην γραμμική παλινδρόμηση, και οι δύο μέθοδοι εκτίμησης δίνουν την ίδια εκτίμηση των παραμέτρων όταν οι όροι σφάλματος στο μοντέλο μη γραμμικής παλινδρόμησης είναι ανεξάρτητοι, κανονικά κατανεμημένοι και με σταθερή διακύμανση.

Σε αντίθεση με τη γραμμική παλινδρόμηση, συνήθως δεν είναι δυνατό να βρεθούν οι αναλυτικές εκφράσεις των εκτιμητών ελαχίστων τετραγώνων και μέγιστης πιθανοφάνειας σε ένα μη γραμμικό μοντέλο παλινδρόμησης. Αντί αυτών, πρέπει να χρησιμοποιούνται αριθμητικές μέθοδοι αναζήτησης, και για τις δύο διαδικασίες εκτίμησης, απαιτώντας πολλούς υπολογισμούς. Επομένως, η ανάλυση των μοντέλων μη γραμμικής παλινδρόμησης, συνήθως εκτελείται με τη χρήση ειδικού λογισμικού .

Προκειμένου να γίνει αντιληπτή η προσαρμογή και η ανάλυση ενός μοντέλου μη γραμμικής παλινδρόμησης, θα χρησιμοποιηθεί ένα παράδειγμα όπου το μοντέλο έχει μόνο δύο παραμέτρους και το μέγεθος του δείγματος είναι σχετικά μικρό. Με τον τρόπο αυτό, θα γίνει μια προσπάθεια επεξήγησης των βασικών εννοιών και διαδικασιών, χωρίς «εξαντλητικές» λεπτομέρειες. Για αυτό χρησιμοποιούμε το παράδειγμα των Rawlings, Pantula & Dickey (1998).

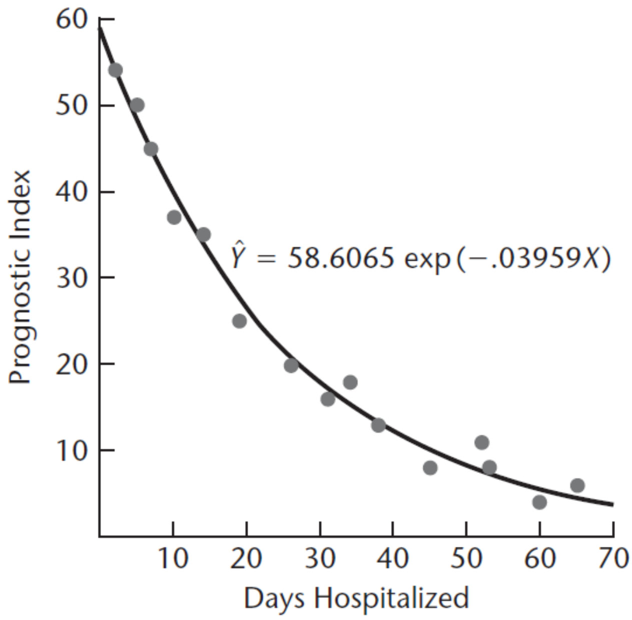
Η διοίκηση ενός νοσοκομείου επιθυμεί να αναπτύξει ένα μοντέλο παλινδρόμησης για την πρόβλεψη του επιπέδου της μακροχρόνιας ανάρρωσης μετά το εξιτήριο από το νοσοκομείο για σοβαρά τραυματισμένους ασθενείς. Η μεταβλητή πρόβλεψης, δηλαδή η ανεξάρτητη μεταβλητή, που πρόκειται να χρησιμοποιηθεί είναι ο αριθμός ημερών νοσηλείας (X) και η μεταβλητή απόκρισης, δηλαδή η εξαρτημένη μεταβλητή, είναι ένας δείκτης μακροχρόνιας ανάρρωσης (Y).

Τα δεδομένα για 15 ασθενείς παρουσιάζονται στον Πίνακα 1.

Ασθενής *i*	**Ημέρες νοσηλείας X_i**	**Δείκτης ανάρρωσης Y_i**
1	2	54
2	5	50
3	7	45
4	10	37
5	14	35
6	19	25
7	26	20
8	31	16
9	34	18
10	38	13
11	45	8
12	52	11
13	53	8
14	60	4
15	65	6

Πίνακας 1. Δεδομένα σοβαρά τραυματισμένων ασθενών

Μια γραφική απεικόνιση των δεδομένων φαίνεται στο Σχήμα 15.



Σχήμα 15. Διάγραμμα διασποράς και η καμπύλη προσαρμογής του εκθετικού μοντέλου παλινδρόμησης

Βασιζόμενοι σε προηγούμενες σχετικές μελέτες, βρέθηκε ότι η σχέση μεταξύ της μεταβλητής πρόβλεψης και της μεταβλητής απόκρισης είναι εκθετική. Ως εκ τούτου, αποφασίστηκε να διερευνηθεί η καταλληλότητα του μοντέλου μη γραμμικής εκθετικής παλινδρόμησης δύο παραμέτρων μορφής (6), όπου τα σφάλματα ε_i είναι ανεξάρτητα, κανονικά κατανεμημένα και με σταθερή διακύμανση. Εάν αυτό το μοντέλο είναι κατάλληλο, είναι επιθυμητή η εκτίμηση των παραμέτρων παλινδρόμησης β_0 και β_1.

Η μέθοδος ελαχίστων τετραγώνων για την απλή γραμμική παλινδρόμηση, της μορφής

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ απαιτεί την ελαχιστοποίηση της συνάρτησης Q:

$$Q = \sum_{i=1}^{n} \left[Y_i - \left(\beta_0 + \beta_1 X_i \right) \right]^2 \tag{16}$$

Οι τιμές των β_0 και β_1 που ελαχιστοποιούν το Q για τα ζεύγη των παρατηρήσεων του δείγματος, δηλαδή τα ζεύγη της μορφής (X_i, Y_i), είναι οι εκτιμήσεις των ελάχιστων τετραγώνων και σημειώνονται με τα $\hat{\beta}_0$ και $\hat{\beta}_1$.

Ένας τρόπος για την εύρεση των εκτιμητών ελαχίστων τετραγώνων είναι μέσω των κανονικών εξισώσεων ελαχίστων τετραγώνων. Οι κανονικές εξισώσεις των ελαχίστων τετραγώνων βρίσκονται με αναλυτικούς τρόπους, δηλαδή με τη μερική παραγώγιση της συνάρτησης Q, ως προς β_0 και β_1 αντίστοιχα, και βρίσκοντας τις τιμές που συγχρόνως τις μηδενίζουν. Η λύση του συστήματος των κανονικών εξισώσεων, δίνει τις εκτιμήτριες ελάχιστων τετραγώνων. Ένας ακόμη τρόπος για την εύρεση των εκτιμητών ελαχίστων τετραγώνων, είναι χρησιμοποιώντας μια αριθμητική μέθοδο αναζήτησης. Με αυτή την προσέγγιση, η συνάρτηση Q αξιολογείται για διαφορετικές τιμές των β_0 και β_1, μεταβάλλοντάς τα συστηματικά μέχρι να βρεθεί η ελάχιστη τιμή της. Οι τιμές των β_0 και β_1 που ελαχιστοποιούν τη συνάρτηση Q, είναι οι εκτιμητές ελάχιστων τετραγώνων $\hat{\beta}_0$ και $\hat{\beta}_1$ αντίστοιχα.

Οι μέθοδοι εκτίμησης ελαχίστων τετραγώνων της γραμμικής παλινδρόμησης επεκτείνονται και σε μοντέλα μη γραμμικής παλινδρόμησης. Σε αυτή την περίπτωση, η συνάρτηση αθροιστικών τετράγωνων έχει τη μορφή:

$$Q = \sum_{i=1}^{n} \left[Y_i - f\left(\boldsymbol{X}_i, \boldsymbol{\beta} \right) \right]^2 \tag{17}$$

όπου $f\left(\boldsymbol{X}_i, \boldsymbol{\beta}\right)$ είναι η μέση απόκριση για την περίπτωση i σύμφωνα με τη μη γραμμική συνάρτηση απόκρισης $f\left(\boldsymbol{X}, \boldsymbol{\beta}\right)$. Η συνάρτηση αθροιστικών τετραγώνων πρέπει να ελαχιστοποιείται σε σχέση με τις παραμέτρους μη γραμμικής παλινδρόμησης, δηλαδή τις β_0, β_1,. . . , $\beta_{p\text{-}1}$, ώστε να βρεθούν οι εκτιμητές ελαχίστων τετραγώνων.

Και στην περίπτωση της μη γραμμικής παλινδρόμησης μπορούν χρησιμοποιηθούν, για την εύρεση των εκτιμητών ελαχίστων τετραγώνων, η αριθμητική αναζήτηση όπως και οι κανονικές εξισώσεις. Η διαφορά από τη γραμμική παλινδρόμηση είναι ότι η επίλυση του συστήματος των κανονικών εξισώσεων απαιτεί, συνήθως, μια επαναληπτική διαδικασία αριθμητικής αναζήτησης, επειδή οι αναλυτικές λύσεις, γενικά, δεν μπορούν να βρεθούν.

Επανερχόμενοι στο παράδειγμα των σοβαρά τραυματισμένων ασθενών, η συνάρτηση απόκρισης φαίνεται να είναι της μορφής:

$$f\left(\boldsymbol{X}_i,\boldsymbol{\beta}\right)=\beta_0 exp\left(\beta_1 X_i\right). \tag{18}$$

Επομένως, η συνάρτηση αθροιστικών τετραγώνων Q θα είναι της μορφής:

$$Q=\sum_{i=1}^{n}\left[Y_i-exp\left(\beta_1 X_i\right)\right]^2 . \tag{19}$$

Η μέθοδος της μέγιστης πιθανοφάνειας οδηγεί στο ίδιο κριτήριο όταν οι όροι σφάλματος ε_i είναι ανεξάρτητοι κανονικά κατανεμημένοι και με σταθερή διακύμανση λαμβάνοντας υπόψη την παρακάτω συνάρτηση πιθανοφάνειας:

$$L\left(\boldsymbol{\beta},\sigma^2\right)=\frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n}\exp\left(-\frac{\sum_{i=1}^{n}\left[Y_i-exp\left(\beta_1 X_i\right)\right]^2}{2\sigma^2}\right). \tag{20}$$

Ακριβώς όπως και στη γραμμική παλινδρόμηση, η μεγιστοποίηση αυτής της συνάρτησης πιθανοφάνειας, σε σχέση με τις παραμέτρους παλινδρόμησης β_0 και β_1, ισοδυναμεί με την ελαχιστοποίηση του αθροίσματος στον εκθέτη, οπότε οι εκτιμητές μέγιστης πιθανότητας να είναι οι ίδιες με τις εκτιμήσεις των ελάχιστων τετραγώνων.

Στη συνέχεια παρουσιάζεται η εύρεση των εκτιμητριών των ελαχίστων τετραγώνων, πρώτα με τη χρήση του συστήματος των κανονικών εξισώσεων και στη συνέχεια με διαδικασίες άμεσης αριθμητικής αναζήτησης.

Για να βρεθούν οι κανονικές εξισώσεις για ένα μη γραμμικό μοντέλο παλινδρόμησης της μορφής (12) πρέπει να ελαχιστοποιηθεί η συνάρτηση αθροιστικών τετραγώνων (17) σε σχέση με τις παραμέτρους μη γραμμικής παλινδρόμησης, δηλαδή τις β_0, $\beta_1,\ldots,\beta_{p-1}$. Οι μερικές παράγωγοι της συνάρτησης Q ως προς τις παραμέτρους β_κ, με i=1,2,…,κ έχουν τη μορφή:

$$\frac{\partial Q}{\partial \beta_\kappa}=-2\sum_{i=1}^{n}\left[Y_i-f\left(\boldsymbol{X}_i,\boldsymbol{\beta}\right)\right]\left[\frac{\partial f\left(\boldsymbol{X}_i,\boldsymbol{\beta}\right)}{\partial \beta_\kappa}\right] \tag{21}$$

Αν θέσουμε τις παραπάνω μερικές παραγώγους ίσες με μηδέν και οι παράμετροι β_κ αντικατασταθούν με τους εκτιμητές ελαχίστων τετραγώνων $\hat{\beta}_\kappa$, τότε καταλήγουμε σε p κανονικές εξισώσεις της μορφής

$$\sum_{i=1}^{n} Y_i \left[\frac{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\beta}\right)}{\partial \beta_\kappa} \right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \sum_{i=1}^{n} f\left(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}}\right) \left[\frac{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\beta}\right)}{\partial \beta_\kappa} \right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0 \tag{22}$$

με κ=1,2,…,p-1 όπου

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{pmatrix} . \tag{23}$$

Οι κανονικές εξισώσεις για μη γραμμικά μοντέλα παλινδρόμησης είναι μη γραμμικές ως προς τις παραμέτρους των εκτιμητριών $\hat{\beta}_\kappa$ και συνήθως είναι δύσκολο να επιλυθούν, ακόμη και στην απλούστερη περίπτωση. Ως εκ τούτου, απαιτούνται συνήθως αριθμητικές διαδικασίες αναζήτησης για να ληφθεί μια λύση των κανονικών εξισώσεων, με επαναληπτικές μεθόδους. Μια επιπλέον δυσκολία είναι ότι είναι πιθανό να υπάρχουν περισσότερες από μία λύσεις.

Στο παράδειγμα με τους σοβαρά τραυματισμένους ασθενείς, οι μερικές παράγωγοι της συνάρτησης $f\left(\boldsymbol{X}_i, \boldsymbol{\beta}\right)$, ως προς β_0 και β_1, αντίστοιχα είναι ίσες με

$$\frac{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\beta}\right)}{\partial \beta_0} = exp\left(\beta_1 X_i\right) \tag{24}$$

$$\frac{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\beta}\right)}{\partial \beta_1} = \beta_0 X_i exp\left(\beta_1 X_i\right). \tag{25}$$

Αντικαθιστώντας τις παραμέτρους β_0, β_1 με τους εκτιμητές ελαχίστων τετραγώνων $\hat{\beta}_0$ και $\hat{\beta}_1$ στις παραπάνω εξισώσεις, έχουμε τις κανονικές εξισώσεις

$$\sum_{i=1}^{n} Y_i \exp\left(\hat{\beta}_i X_i\right) - \sum_{i=1}^{n} \hat{\beta}_0 \exp\left(2\hat{\beta}_i X_i\right) = 0 \tag{26}$$

$$\sum_{i=1}^{n} Y_i \hat{\beta}_0 X_i \exp\left(\hat{\beta}_i X_i\right) - \sum_{i=1}^{n} \hat{\beta}_0 \exp\left(\hat{\beta}_i X_i\right) \hat{\beta}_0 X_i \exp \hat{\beta}_i X_i \tag{27}$$

οι οποίες μετά από πράξεις μας δίνουν

$$\sum_{i=1}^{n} Y_i \exp\left(\hat{\beta}_i X_i\right) - \hat{\beta}_0 \sum_{i=1}^{n} \exp\left(2\hat{\beta}_i X_i\right) = 0 \tag{28}$$

$$\sum_{i=1}^{n} Y_i X_i \exp\left(\hat{\beta}_i X_i\right) - \hat{\beta}_0 \sum_{i=1}^{n} X_i \exp\left(2\hat{\beta}_i X_i\right) = 0 \ . \tag{29}$$

Οι παραπάνω εξισώσεις, ως προς $\hat{\beta}_0$ και $\hat{\beta}_1$, είμαι μη γραμμικές και δεν υπάρχει μια κλειστού τύπου λύση. Αυτό σημαίνει ότι απαιτούνται αριθμητικές μέθοδοι για την επαναληπτική εύρεση λύσης για τους εκτιμητές ελαχίστων τετραγώνων.

Σε πολλά προβλήματα μη γραμμικής παλινδρόμησης, είναι πιο πρακτικό να βρεθουν οι εκιμήτριες των ελάχιστων τετραγώνων με διαδικασίες άμεσης αριθμητικής αναζήτησης και όχι με τη χρήση των κανονικών εξισώσεων και στη συνέχεια η αριθμητική λύση του συστήματος τους. Τα στατιστικά πακέτα χρησιμοποιούν μία ή περισσότερες διαδικασίες άμεσης αριθμητικής αναζήτησης για την επίλυση προβλημάτων μη γραμμικής παλινδρόμησης. Μια από αυτές τις μεθόδους είναι η μέθοδος Gauss-Newton, η οποία ονομάζεται επίσης μέθοδος γραμμικοποίησης. Η συγκεκριμένη μέθοδος χρησιμοποιεί μια σειρά Taylor για να προσεγγίσει το μοντέλο μη γραμμικής παλινδρόμησης με γραμμικούς όρους και στη συνέχεια εφαρμόζει τη μέθοδο ελαχίστων τετραγώνων για την εκτίμηση των παραμέτρων. Η επανάληψη αυτών των βημάτων γενικά οδηγεί σε μια λύση σε ένα πρόβλημα μη γραμμικής παλινδρόμησης.

Η μέθοδος Gauss-Newton αρχίζει με αρχικές τιμές για τις παραμέτρους παλινδρόμησης β_0, β_1,…, $\beta_{p\text{-}1}$. Οι συγκεκριμένες αρχικές τιμές δηλώνονται ως εξής:

$$\hat{\beta}_0^{(0)}, \hat{\beta}_1^{(0)}, ..., \hat{\beta}_{p\text{-}1}^{(0)}$$

όπου ο δείκτης σε παρένθεση δηλώνει τον αριθμό επανάληψης. Οι αρχικές τιμές της μορφής $\hat{\beta}_\kappa^{(0)}$ μπορεί να ληφθούν από προηγούμενες ή σχετικές μελέτες, θεωρητικές αναμενόμενες τιμές, ή ακόμα από μια προκαταρκτική αναζήτηση τιμών των παραμέτρων οι οποίες οδηγούν σε μια σχετικά μικρή τιμή της συνάρτησης Q.

Αφού ληφθούν οι αρχικές τιμές για τις παραμέτρους, προσεγγίζονται οι μέσες αποκρίσεις για τις n περιπτώσεις από τους γραμμικούς όρους του αναπτύγματος της σειράς Taylor γύρω από τις αρχικές τιμές $\hat{\beta}_{\kappa}^{(0)}$. Για την i περίπτωση έχουμε:

$$f\left(\boldsymbol{X}_i,\boldsymbol{\beta}\right) \approx f\left(\boldsymbol{X}_i,\hat{\boldsymbol{\beta}}^{(0)}\right)+\sum_{\kappa=0}^{p-1}\left[\frac{\partial f\left(\boldsymbol{X}_i,\boldsymbol{\beta}\right)}{\partial \beta_{\kappa}}\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(0)}}\left(\beta_{\kappa}-\hat{\beta}_{\kappa}^{(0)}\right) \tag{30}$$

όπου

$$\hat{\boldsymbol{\beta}}^{(0)}=\begin{pmatrix}\hat{\beta}_0^{(0)}\\ \hat{\beta}_1^{(0)}\\ \vdots\\ \hat{\beta}_{p-1}^{(0)}\end{pmatrix}. \tag{31}$$

Το διάνυσμα $\hat{\boldsymbol{\beta}}^{(0)}$ είναι το διάνυσμα των αρχικών τιμών των παραμέτρων του μοντέλου μη γραμμικής παλινδρόμησης, με τους όρους του να είναι οι ίδιες οι μερικοί παράγωγοι της συνάρτησης παλινδρόμησης που παρουσιάζονται και στις κανονικές εξισώσεις, με τη διαφορά ότι εδώ υπολογίζονται για τις τιμές $\beta_{\kappa}=\hat{\beta}_{\kappa}^{(0)}$ για κ =0,1,. . ., p-1.

Αν στη συνέχεια θέσουμε

$$f_i^{(0)}=f\left(\boldsymbol{X}_i,\hat{\boldsymbol{\beta}}^{(0)}\right) \tag{32}$$

$$\beta_{\kappa}'^{(0)}=\beta_{\kappa}-\hat{\beta}_{\kappa}^{(0)} \tag{33}$$

$$D_{i\kappa}^{(0)}=\left[\frac{\partial f\left(\boldsymbol{X}_i,\boldsymbol{\beta}\right)}{\partial \beta_{\kappa}}\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(0)}} \tag{34}$$

η προσέγγιση Taylor για τη μέση απόκριση της i περίπτωσης έχει την παρακάτω μορφή

$$f\left(\boldsymbol{X}_i,\boldsymbol{\beta}\right)\approx f_i^{(0)}+\sum_{\kappa=0}^{p-1}D_{i\kappa}^{(0)}\beta_{\kappa}^{(0)} \tag{35}$$

και μια προσέγγιση του μη γραμμικού μοντέλου παλινδρόμησης είναι η

$$Y_i \approx f_i^{(0)} + \sum_{\kappa=0}^{p\text{-}1} D_{i\kappa}^{(0)} \beta_\kappa^{(0)} + \varepsilon_i \Rightarrow Y_i - f_i^{(0)} \approx + \sum_{\kappa=0}^{p\text{-}1} D_{i\kappa}^{(0)} \beta_\kappa^{(0)} + \varepsilon_i \ . \tag{36}$$

Αν η διαφορά $Y_i - f_i^{(0)}$ αντικατασταθεί με $Y_i^{(0)} = Y_i - f_i^{(0)}$ τότε παίρνουμε την ακόλουθη προσέγγιση ενός γραμμικού μοντέλου παλινδρόμησης

$$Y_i^{(0)} \approx \sum_{\kappa=0}^{p\text{-}1} D_{i\kappa}^{(0)} \beta_\kappa^{(0)} + \varepsilon_i \quad , \quad i = 1, 2, ..., n \ . \tag{37}$$

Η προσέγγιση του παραπάνω μοντέλου γραμμικής παλινδρόμησης έχει τη μορφή

$$Y_i \ = \beta_0 \mathrm{X}_{i0} + \ \beta_1 \mathrm{X}_{i1} + ... + \beta_{p\text{-}1} \mathrm{X}_{i,p\text{-}1} + \varepsilon_i \ . \tag{38}$$

Οι αποκρίσεις $Y_i^{(0)}$ είναι υπόλοιπα (residuals), δηλαδή, οι αποκλίσεις των παρατηρήσεων γύρω από τη συνάρτηση μη γραμμικής παλινδρόμησης, με τις παραμέτρους να αντικαθίστανται από τις αρχικές τιμές των εκτιμήσεων. Οι $D_{i\kappa}^{(0)}$ είναι οι μερικές παράγωγοι της μέσης απόκρισης για κάθε μία από τις περιπτώσεις, όπου οι παράμετροι αντικαθίστανται από τις αρχικές εκτιμήσεις τους. Κάθε συντελεστής παλινδρόμησης $\beta_\kappa^{(0)}$ αντιπροσωπεύει τη διαφορά μεταξύ της πραγματικής παραμέτρου παλινδρόμησης και της αρχικής εκτίμησης.

Έτσι, οι συντελεστές παλινδρόμησης αντιπροσωπεύουν τις προσαρμοσμένες ποσότητες (adjustment amounts) κατά τις οποίες πρέπει να διορθωθούν οι αρχικοί συντελεστές παλινδρόμησης. Ο σκοπός της προσαρμογής της προσέγγισης του μοντέλου γραμμικής παλινδρόμησης (37) είναι η εκτίμηση των συντελεστών παλινδρόμησης $\beta_\kappa^{(0)}$ και η χρήση τους για την προσαρμογή των αρχικών εκτιμήσεων των παραμέτρων της μη γραμμικής παλινδρόμησης. Για την προσαρμογή (fitting) αυτής της προσέγγισης γραμμικής παλινδρόμησης, δεν υπάρχει ο σταθερός όρος $\beta^{(0)}$ (intercept).

Σε μορφή πινάκων, η παραπάνω προσέγγιση του γραμμικού μοντέλου παλινδρόμησης γράφεται ως εξής

$$\boldsymbol{Y}^{(0)} \approx \boldsymbol{D}^{(0)} \boldsymbol{\beta}^{(0)} + \boldsymbol{\varepsilon} \tag{39}$$

όπου

$$\underset{n\times 1}{\boldsymbol{Y}} = \begin{pmatrix} Y_1 - f_1^{(0)} \\ Y_2 - f_2^{(0)} \\ \vdots \\ Y_n - f_n^{(0)} \end{pmatrix} \qquad \underset{n\times p}{\boldsymbol{D}} = \begin{pmatrix} D_{10}^{(0)} & D_{11}^{(0)} & \cdots & D_{1,\mathrm{p-1}}^{(0)} \\ D_{20}^{(0)} & D_{21}^{(0)} & \cdots & D_{2,\mathrm{p-1}}^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ D_{\mathrm{n0}}^{(0)} & D_{\mathrm{n1}}^{(0)} & \cdots & D_{\mathrm{n,p-1}}^{(0)} \end{pmatrix}$$

$$\underset{n\times 1}{\boldsymbol{\varepsilon}}^{(0)} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Σημειώστε ξανά ότι το μοντέλο προσέγγισης (39) έχει ακριβώς τη μορφή του γενικού μοντέλου παλινδρόμησης, με τον πίνακα $\boldsymbol{D}$ των μερικών παραγώγων να έχει το ρόλο του πίνακα $\boldsymbol{X}$, αλλά χωρίς τη στήλη με τις μονάδες για το σταθερό όρο (intercept). Ως εκ τούτου, είναι δυνατή η εκτίμηση των παραμέτρων $\boldsymbol{\beta}^{(0)}$ με τη συνήθη μέθοδο των ελαχίστων τετραγώνων και θα έχουμε

$$\boldsymbol{\beta}'^{(0)} = \left(\left(\boldsymbol{D}^{(0)}\right)^T \boldsymbol{D}^{(0)}\right)^{-1} \left(\boldsymbol{D}^{(0)}\right)^T \boldsymbol{Y}^{(0)} \tag{40}$$

όπου $\boldsymbol{\beta}'^{(0)}$ είναι το διάνυσμα των εκτιμημένων συντελεστών παλινδρόμησης των ελαχίστων τετραγώνων.

Στη συνέχεια, χρησιμοποιούνται αυτές οι εκτιμήσεις των ελαχίστων τετραγώνων για να λάβουμε τις αναθεωρημένες εκτιμήτριες των συντελεστών παλινδρόμησης με τη βοήθεια της σχέσης $\beta_\kappa'^{(0)} = \beta_\kappa - \hat{\beta}_\kappa^{(0)}$ οπότε παίρνουμε $\hat{\beta}_\kappa^{(1)} = \hat{\beta}_\kappa^{(0)} + \beta_\kappa'^{(0)}$ όπου η $\hat{\beta}_\kappa^{(1)}$ δηλώνει την αναθεωρημένη εκτίμηση της β_κ στο τέλος της πρώτης επανάληψης. Σε μορφή πινάκων, είναι

$$\boldsymbol{\beta}_\kappa^{(1)} = \boldsymbol{\beta}_\kappa^{(0)} + \boldsymbol{\beta}_\kappa'^{(0)} \tag{41}$$

Σε αυτό το σημείο, μπορούμε να εξετάσουμε αν οι αναθεωρημένοι συντελεστές παλινδρόμησης παρέχουν κατάλληλες προσαρμογές. Παρακάτω δίνεται η προσέγγιση της συνάρτησης Q, με τη μέθοδο των ελαχίστων τετραγώνων, η οποία αξιολογείται για τους αρχικούς συντελεστές παλινδρόμησης $\hat{\boldsymbol{\beta}}^{(0)}$ με χρήση του $SSE^{(0)}$ είναι

$$SSE^{(0)} = \sum_{i=1}^{n}\left[Y_i - f\left(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}}^{(0)}\right)\right]^2 = \sum_{i=1}^{n}\left(Y_i - f_i^{(0)}\right)^2 \quad .$$

Στο τέλος της πρώτης επανάληψης, οι αναθεωρημένοι συντελεστές παλινδρόμησης είναι οι $\hat{\boldsymbol{\beta}}^{(1)}$ και

$$SSE^{(1)} = \sum_{i=1}^{n}\left[Y_i - f\left(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}}^{(1)}\right)\right]^2 = \sum_{i=1}^{n}\left(Y_i - f_i^{(1)}\right)^2 .$$

Εάν η μέθοδος Gauss-Newton λειτουργεί αποτελεσματικά στην πρώτη επανάληψη, πρέπει να είναι $SSE^{(1)} < SSE^{(0)}$, δεδομένου ότι οι αναθεωρημένοι εκτιμητές των συντελεστών παλινδρόμησης, $\hat{\boldsymbol{\beta}}^{(1)}$, θα πρέπει να είναι καλύτεροι.

Στον υπολογισμό των $SSE^{(0)}$ και $SSE^{(1)}$ χρησιμοποιούνται οι συναρτήσεις μη γραμμικής παλινδρόμησης $f\left(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}}^{(0)}\right)$ και $f\left(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}}^{(1)}\right)$ και όχι οι γραμμικές προσεγγίσεις από το ανάπτυγμα της σειράς Taylor. Οι αναθεωρημένοι συντελεστές παλινδρόμησης $\hat{\boldsymbol{\beta}}^{(1)}$ δεν είναι οι εκτιμήσεις των ελάχιστων τετραγώνων για το μοντέλο μη γραμμικής παλινδρόμησης, επειδή το προσαρμοσμένο μοντέλο (39), είναι μόνο μια προσέγγιση του μη γραμμικού μοντέλου.

Επομένως, η μέθοδος Gauss-Newton επαναλαμβάνει τη διαδικασία, η οποία περιγράφηκε παραπάνω, με το διάνυσμα $\hat{\boldsymbol{\beta}}^{(1)}$ να είναι η νέα αρχική τιμή. Έτσι παράγεται ένα νέο σύνολο αναθεωρημένων εκτιμήσεων, οι οποίες σημειώνονται με $\hat{\boldsymbol{\beta}}^{(2)}$, όπως και ένα νέο άθροισμα τετραγώνων το $SSE^{(2)}$. Η επαναληπτική διαδικασία συνεχίζεται μέχρι οι διαφορές μεταξύ των διαδοχικών εκτιμήσεων των συντελεστών παλινδρόμησης, $\hat{\boldsymbol{\beta}}^{(s+1)} - \hat{\boldsymbol{\beta}}^{(s)}$, και/ή τη διαφορά μεταξύ των διαδοχικών αθροισμάτων ελαχίστων τετραγώνων, $SSE^{(s+1)} - SSE^{(s)}$, να καταστούν αμελητέες. Οι τελικές εκτιμήτριες των συντελεστών παλινδρόμησης συμβολίζονται απλά με $\hat{\boldsymbol{\beta}}$ και το τελικό άθροισμα ελαχίστων τετραγώνων, το οποίο είναι το άθροισμα των τετραγώνων των σφαλμάτων ως *SSE*.

Η μέθοδος Gauss-Newton λειτουργεί αποτελεσματικά σε πολλές εφαρμογές μη γραμμικής παλινδρόμησης, αλλά μπορεί να απαιτεί μεγάλο αριθμό επαναλήψεων, πριν τη σύγκλιση, ενώ υπάρχει και η περίπτωση να μην συγκλίνει καθόλου.

Στο παράδειγμα με τους σοβαρά τραυματισμένους ασθενείς, οι αρχικές εκτιμήσεις των παραμέτρων ελήφθησαν από το γραμμικό μετασχηματισμό της μορφής

$$\ln \beta_0 \left[\exp\left(\beta_1 X \right) \right] = \ln \beta_0 + \beta_1 X$$

το οποίο παίρνει τη γραμμική μορφή $Y_i' = \beta_0' + \beta_1' X + \varepsilon_i$ όπου $Y_i' = \ln Y_i$, $\beta_0' = \ln \beta_0$ και $\beta_1' = \beta_1$.

Το παραπάνω μοντέλο γραμμικής παλινδρόμησης, μετά την προσαρμογή του με τη μέθοδο των ελαχίστων τετραγώνων έδωσε τους παρακάτω συντελεστές παλινδρόμησης, $\beta_0' = 4,0371$ και $\beta_1' = -0,03797$. Έτσι, οι αρχικές τιμές για την επαναληπτική διαδικασία είναι $\hat{\beta}_0^{(0)} = \exp\left(\beta_0'\right) = \exp\left(4,0371\right) = 56,6646$ και $\hat{\beta}_1^{(0)} = \beta_1' = -0,03797$

Το κριτήριο της μεθόδου ελάχιστων τετραγώνων σε αυτό το στάδιο απαιτεί την αξιολόγηση του μοντέλου μη γραμμικής παλινδρόμησης $f\left(\boldsymbol{X}, \boldsymbol{\beta}\right)$ για κάθε διαθέσιμη τιμή της μεταβλητής *X*, χρησιμοποιώντας τις αρχικές τιμές των παραμέτρων, δηλαδή τις των τιμές παραμέτρων εκκίνησης $\hat{\beta}_0^{(0)}$ και $\hat{\beta}_1^{(0)}$.

Για παράδειγμα, για $X_1 = 2$, είναι

$$f\left(\boldsymbol{X}_1, \hat{\boldsymbol{\beta}}^{(0)}\right) = f_1^{(0)} = \hat{\beta}_0^{(0)} exp\left(\hat{\beta}_1^{(0)} X_1\right) = 56,6646 \cdot \exp\left(-0,03797\right) = 52,5208$$

και $Y_1 - f_1^{(0)} = 1,4792$

Συνεχίζοντας με τον ίδιο τρόπο, έχουμε:

$$\underset{15\times 1}{\boldsymbol{Y}} = \begin{pmatrix} Y_1 - f_1^{(0)} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ Y_{15} - f_{15}^{(0)} \end{pmatrix} = \begin{pmatrix} Y_1 - \hat{\beta}_0^{(0)} \exp\left(\hat{\beta}_1^{(0)} X_1\right) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ Y_{15} - \hat{\beta}_0^{(0)} \exp\left(\hat{\beta}_1^{(0)} X_{15}\right) \end{pmatrix} = \begin{pmatrix} 1,4792 \\ 3,1337 \\ 1,5609 \\ -1,7624 \\ 1,6996 \\ -2,5422 \\ -1,1139 \\ -1,4629 \\ 2,4172 \\ -0,3871 \\ -2,2625 \\ 3,1327 \\ 0,4259 \\ -1,8063 \\ 1,1977 \end{pmatrix}$$

$$\underset{15\times 2}{\boldsymbol{D}} = \begin{pmatrix} \exp\left(\hat{\beta}_1^{(0)} X_1\right) & \hat{\beta}_0^{(0)} X_1 \exp\left(\hat{\beta}_1^{(0)} X_1\right) \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \exp\left(\hat{\beta}_1^{(0)} X_{15}\right) & \hat{\beta}_0^{(0)} X_1 \exp\left(\hat{\beta}_1^{(0)} X_1\right) \end{pmatrix} = \begin{pmatrix} 0,92687 & 105,0416 \\ 0,82708 & 234,3317 \\ 0,76660 & 304,0736 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0,10247 & 348,3801 \\ 0,08475 & 312,1510 \end{pmatrix}$$

Το κριτήριο των ελαχίστων τετραγώνων σε αυτό το αρχικό στάδιο είναι απλά το άθροισμα των τετραγώνων των υπολοίπων στο στάδιο (0).

$$SSE^{(0)} = \sum_{i=1}^{15}\left(Y_i - f_i^{(0)}\right)^2 = \sum_{i=1}^{15}\left(Y_i^{(0)}\right)^2 = (1,4792)^2 + \ldots + (1,1977)^2 = 56,0869$$

Για τον υπολογισμό των αρχικών εκτιμήσεων των παραμέτρων, χρειάζεται και ο υπολογισμός του πίνακα $\boldsymbol{D}^{(0)}$, οπότε είναι απαραίτητο να υπολογιστούν οι μερικές παράγωγοι της συνάρτησης γραμμικής παλινδρόμησης για $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(0)}$. Θυμίζουμε ότι είναι

$$D_{i\kappa}^{(0)} = \left[\frac{\partial f\left(\boldsymbol{X}_i, \boldsymbol{\beta} \right)}{\partial \beta_\kappa} \right]_{\beta = \hat{\beta}^{(0)}}$$

οπότε έχουμε

$$D_{10}^{(0)} = \left[\frac{\partial f\left(\boldsymbol{X}_1, \boldsymbol{\beta} \right)}{\partial \beta_0} \right]_{\beta = \hat{\beta}^{(0)}} = \exp\left(\hat{\beta}_1^{(0)} \mathrm{X}_1 \right) = \exp\left(-0,3797 \right) = 0,92687$$

$$D_{11}^{(0)} = \left[\frac{\partial f\left(\boldsymbol{X}_1, \boldsymbol{\beta} \right)}{\partial \beta_1} \right]_{\beta = \hat{\beta}^{(0)}} = \hat{\beta}_0^{(0)} \mathrm{X}_1 \exp\left(\hat{\beta}_1^{(0)} \mathrm{X}_1 \right) = 56,6646 \exp\left(-0,03797 \right)$$
$$= 105,0416$$

Για να λάβουμε τις εκτιμήσεις των ελάχιστων τετραγώνων $\boldsymbol{\beta}'^{(0)}$, μέσω της παλινδρόμησης επί της μεταβλητής $Y^{(0)}$, χρησιμοποιούμε παλινδρόμηση χωρίς σταθερό όρο. Οι υπολογισμοί έδωσαν

$$\hat{\boldsymbol{\beta}}^{(0)} = \begin{pmatrix} 1,8932 \\ -0,001563 \end{pmatrix}$$

Στη συνέχεια είναι

$$\hat{\boldsymbol{\beta}}^{(1)} = \hat{\boldsymbol{\beta}}^{(0)} + \hat{\boldsymbol{\beta}}'^{(0)} = \begin{pmatrix} 56,6646 \\ -0,03797 \end{pmatrix} + \begin{pmatrix} 1,8932 \\ -0,001563 \end{pmatrix} = \begin{pmatrix} 58,5578 \\ -0,03953 \end{pmatrix}$$

Μετά από τρεις επαναλήψεις λάβαμε τον παρακάτω πίνακα:

Βήμα	$\hat{\boldsymbol{\beta}}_0$	$\hat{\boldsymbol{\beta}}_1$	SSE
0	58,6646	-0,03797	56,0869
1	58,5578	-0,03953	49,4638
2	58,6065	-0,03959	49,4593
3	58,6065	-0,03959	49,4593

Οι τελικοί εκτιμητές ελαχίστων τετραγώνων είναι

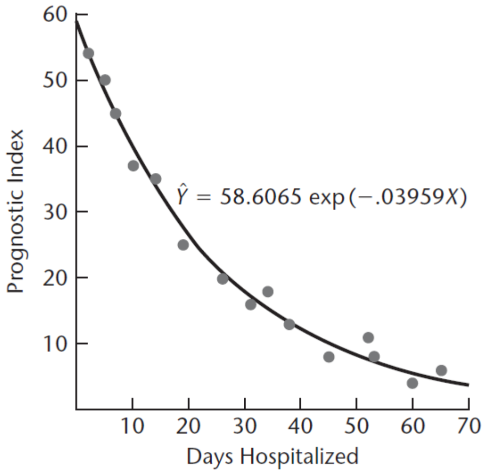
κ	$\hat{\beta}_\kappa$	$S(\hat{\beta}_\kappa)$
0	58,6065	1,472
1	-0,03959	0,00171

Ακόμα είναι $MSE = \dfrac{49,4593}{13} = 3,80456$.

Η διαδικασία αναζήτησης τερματίστηκε μετά από τρεις επαναλήψεις. Οι τελικοί εκτιμητές των συντελεστών παλινδρόμησης είναι $\hat{\beta}_0$=58,6065 και $\hat{\beta}_1$=-0.3959 και το προσαρμοσμένο μοντέλο παλινδρόμησης είναι

$$Y = (58.6065)\exp(-0.3959X)$$

Το άθροισμα των τετραγώνων των σφαλμάτων για αυτό το προσαρμοσμένο μοντέλο είναι *SSE*=49,4593. Στο παρακάτω διάγραμμα απεικονίζονται τα ζεύγη των παρατηρήσεων, καθώς και η καμπυλη παλινδρόμησης, η οποία φαίνεται να έχει αρκετά καλή προσαρμογή στο μοντέλο.



Σχήμα 16. Διάγραμμα διασποράς και η καμπύλη προσαρμογής του παραδείγματος

2.5.Περιοχές εφαρμογής μη γραμμικής παλινδρόμησης

Η επιλογή ενός μη-γραμμικού μοντέλου εξαρτάται από το θεωρητικό υπόβαθρο του κάθε επιστημονικού πεδίου, καθώς και από το αντικείμενο μελέτης. Αυτό σημαίνει ότι υπάρχουν θεωρητικές ενδείξεις ότι η εξάρτηση μιας μεταβλητής *Y* από μία ανεξάρτητη μεταβλητή *X*, είναι μιας συγκεκριμένης μη-γραμμικής μορφής. Οδηγούμαστε δηλαδή μέσω ενός «μηχανιστικού μοντέλου» ή «μοντέλου μηχανιστικής βάσης» στην εύρεση της κατάλληλης συνάρτησης παλινδρόμησης (Gallant, 1975).

Τα «μηχανιστικά μοντέλα» έχουν ως στόχο τη συσχέτιση των πειραματικών δεδομένων με θεμελιώδεις επιστημονικές αρχές ή έστω με κάποιες διεργασίες ή μηχανισμούς. Εξαιτίας αυτού του τρόπου σχεδιασμού των μοντέλων, προκύπτουν αντίστοιχα «κλάσεις μη-γραμμικών μοντέλων» για συγκεκριμένες περιπτώσεις ενός επιστημονικού πεδίου. Αυτό σημαίνει ότι η μη-γραμμική μορφή της συνάρτησης παλινδρόμησης ποικίλει ανάλογα με τη φύση του κάθε προβλήματος. Έτσι μπορεί να εμφανίζεται ως μία περιοδική συνάρτηση του χρόνου με ανεξάρτητη μεταβλητή το χρόνο *t*, όπως η $f\left(t,\boldsymbol{\beta}\right)=\beta_1+\beta_2\cos\beta_4 t+\beta_3\sin\beta_5 t$ (Gallant, 1975). Τέτοιου είδους μοντέλα είναι γνωστά ως «Μοντέλα Ανάπτυξης» (“Growth Models”), τα οποία βρίσκουν ευρεία μορφή σε πολλά πεδία.

Σε άλλες περιπτώσεις, η συνάρτηση παλινδρόμησης, μπορεί να προκύπτει ως λύση ενός συστήματος γραμμικών διαφορικών εξισώσεων. Τέτοια μοντέλα καλούνται «Διαμερισματικά Μοντέλα» (“Compartment Models”). Ένα «Διαμερισματικό Μοντέλο», αποτελείται από πεπερασμένο αριθμό διαμερισμάτων, τα οποία συνδέονται με καθορισμένο τρόπο. Ως «διαμέρισμα» (compartment) ορίζεται η ποσότητα μιας ουσίας, ενώ οι συνδέσεις των διαμερισμάτων αντιπροσωπεύουν τη ροή της ουσίας από το ένα διαμέρισμα στο άλλο. Αυτά τα μοντέλα βρίσκουν εφαρμογές στη Χημεία, αλλά και στη Φαρμακευτική, όπου συχνά χημικές αντιδράσεις περιγράφονται από γραμμικά συστήματα διαφορικών εξισώσεων πρώτης τάξης. Επιπλέον, χρησιμοποιούνται για τη μελέτη της κίνησης ουσιών σε διάφορα βιολογικά συστήματα, με τις ουσίες να είναι είτε εξωγενείς, όπως είναι για παράδειγμα τα φάρμακα, είτε ενδογενείς, όπως είναι οι ορμόνες (Gallant, 1975).

Ένα «Μη-Γραμμικό Μοντέλο Ανάπτυξης» μπορεί να χρησιμοποιηθεί προκειμένου να περιγράψει το τρόπο με τον οποίο αναπτύσσεται-μεταβαλλεται ένα μέγεθος, σε σχέση με τις αλλαγές των τιμών μιας ανεξάρτητης μεταβλητής *X*, που συνήθως είναι ο χρόνος ($X=t$). Τα συγκεκριμένα μοντέλα μπορούν να βρούν εφαρμογές σε πολλά πεδία, όπως είναι η Βιολογία, η Βοτανική, η Δασολογία, αφού σε όλα αυτά τα πεδία αποτελεί αντικείμενο μελέτης η ανάπτυξη ανθρώπων, ζώων και γενικά των οργανισμών, αλλά και των φυτών και των δέντρων. Σε όλες αυτές τις περιπτώσεις, η ανάπτυξη εκφράζεται ως μία συνάρτηση του χρόνου, μέχρι ο οργανισμός να φτάσει στο στάδιο ωρίμανσης.

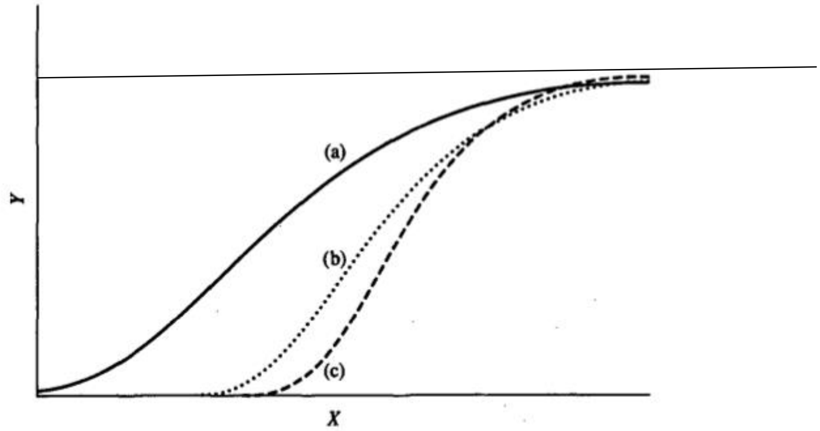
Από την άλλη στη Χημεία και στη Χημική μηχανική, η ανάπτυξη αφορά το αποτέλεσμα χημικών αντιδράσεων, και για αυτό είναι επίσης μια συνάρτηση του χρόνου t, κατά την πραγματοποίηση της αντίδρασης. Αλλά και σε άλλα πεδία, όπως είναι η Οικονομία και οι Πολιτικές Επιστήμες, είναι αντικείμενο μελέτης η ανάπτυξη των οργανισμών, αποθεμάτων τροφίμων, υλικών αλλά ακόμα και εθνών, πάλι ως μια συνάρτηση του χρόνου. Από τα παραπάνω γίνεται αντιληπτό ότι προκειμένου να είναι δυνατή η χρήση τους, είναι απαραίτητο να ερμηνευτούν οι παράμετροί τους.

Ο τύπος του μοντέλου που θα χρησιμοποιηθεί σε κάθε περίπτωση ή πρόβλημα εξαρτάται από τον τύπο της ανάπτυξης που εμφανίζεται. Έτσι, ένα μοντέλο ανάπτυξης είναι μάλλον «μηχανιστικό» παρά εμπειρικό και προκύπτει από τις υποθέσεις που γίνονται σχετικά με τον τύπο της ανάπτυξης, οι οποίες συνήθως εκφράζονται μέσω διαφορικών εξισώσεων ή εξισώσεων διαφορών, από τη λύση των οποίων προκύπτει το μοντέλο. Στη συνέχεια παρουσιάζονται μερικές ευρέως χρησιμοποιούμενες οικογένειες μοντέλων ανάπτυξης (Seber & Wild, 1989).

a. Σιγμοειδή μοντέλα ανάπτυξης (Sigmoidal Growth Models)

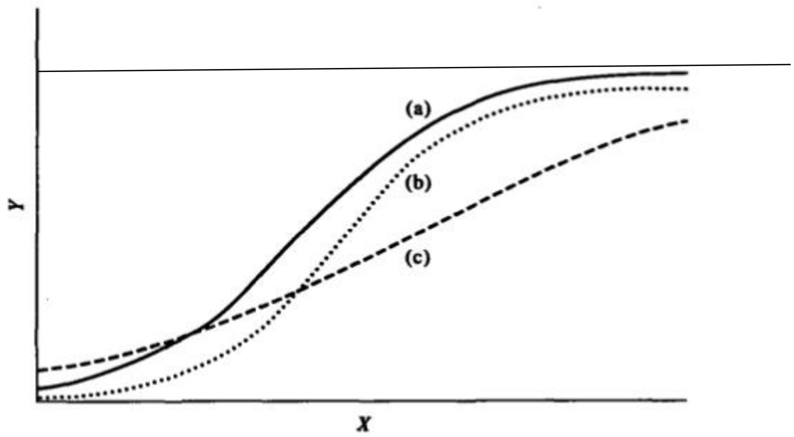
Στα μοντέλα αυτής της μορφής, τα δεδομένα έχουν την τάση να προσαρμόζονται καλύτερα σε «σιγμοειδείς καμπύλες» ή «καμπύλες S-μορφής». Χαρακηριστικό αυτών των καμπυλών είναι ότι αυξάνουν με μεταβαλλόμενο ρυθμό, ξεκινώντας από αρχική τιμή ανάπτυξης και πλησιάζουν ασυμπτωτικά μια τελική τιμή. Αρχικά, ο ρυθμός ανάπτυξης αυξάνεται συνεχώς, μέχρι να φτάσει μία μέγιστη τιμή, ενώ στη συνέχεια μειώνεται σταδιακά, το οποίο και προκαλεί ένα σημείο καμπής στο γράφημα. Τα πλέον γνωστά από αυτα τα μοντέλα είναι το ***μοντέλο Gompertz***, και το ***λογιστικό μοντέλο ανάπτυξης.*** Το πρώτο έχει τη μορφή $Y = \beta_1 e^{-e^{-(\beta_2+\beta_3 X)}}$. Η τιμή του β_1 μας δίνει την ασύμπτωτη, αφού $Y \to \beta_1$ καθώς $X \to \infty$. Η τιμή του β_2 καθορίζει πόσο θα μετατοπιστεί το γράφημα κατά μήκος του ορζόντιου άξονα και τέλος, η τιμή του β_3

καθορίζει τον κατακόρυφο ρυθμό ανάπτυξης. Στο Σχήμα 17, παρουσιάζεται το αντίστοιχο διάγραμμα για β_1=1, β_2=-1,5 και β_3=5 (a), για β_1=1, β_2=-3,5 και β_3=7 (b) και για β_1=1, β_2=-5 και β_3=9 (c). Παρατηρείται ότι μεγαλύτερες τιμές του β_2 οδηγούν σε γράφημα το οποίο είναι πιο αριστερα, ενώ μεγαλύτερες τιμές του β_3 οδηγούν σε μεγαλύτερο ρυθμό ανάπτυξης.



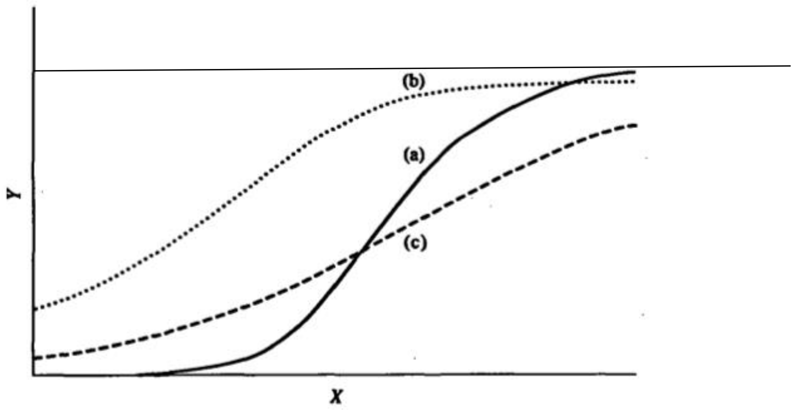
Σχήμα 17. Διαγράμματα μοντέλου Gompertz

Το δεύτερο έχει τη μορφή $Y = \dfrac{\beta_1}{1+e^{-(\beta_2+\beta_3 X)}}$. Η τιμή του β_1 μας δίνει την ασύμπτωτη, αφού $Y \to \beta_1$ καθώς $X \to \infty$. Στο Σχήμα 18, παρουσιάζεται το αντίστοιχο διάγραμμα για β_1=1, β_2=-3,22 και β_3=8 (a), για β_1=0,95, β_2=-4,61 και β_3=10 (b) και για β_1=1, β_2=-2,3 και β_3=4 (c). Παρατηρείται ότι μεγαλύτερες τιμές του β_2 οδηγούν σε γράφημα «πιο ψηλά», ενώ μεγαλύτερες τιμές του β_3 το οδηγούν «πιο δεξιά».



Σχήμα 18. Διαγράμματα λογιστικού μοντέλου

Τέλος, το ***μοντέλο Richard***, έχει τη μορφή $Y = \dfrac{\beta_1}{1 + e^{-(\beta_2 + \beta_3 X)^{\beta_4}}}$, το οποίο είναι γενίκευση του λογιστικού μοντέλου, που προκύπτει για β_4=1. Η τιμή του β_1 μας δίνει την ασύμπτωτη, αφού $Y \to \beta_1$ καθώς $X \to \infty$. Στο Σχήμα 19, παρουσιάζεται το αντίστοιχο διάγραμμα για β_1=1, β_2=-3,22, β_3=5 και β_4=3,33 (a), για β_1=0,95, β_2=-4,61, β_3=10 και β_4=0,33 (b) και β_1=1, β_2=-2,3, β_3=4 και β_4=1,25 (c) .



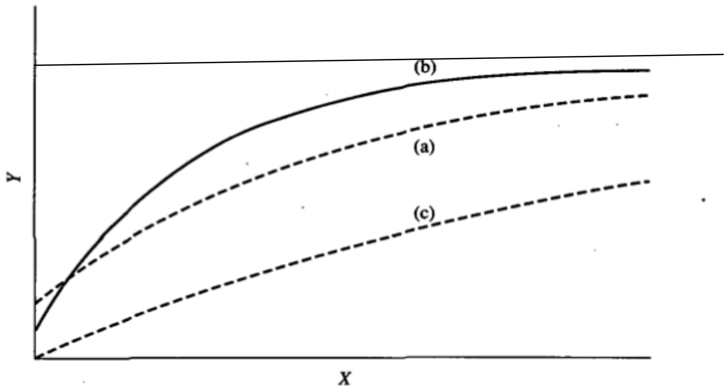
Σχήμα 19. Διαγράμματα μοντέλου Richard

β. Ασυμπτωτικά Μοντέλα Ανάπτυξης (Asymptotic Growth Models)

Σε αυτά τα μοντέλα, η εξαρτητημένη μεταβλητή *Y* αυξάνεται σταθερά, σε σχέση με την ανεξάρτητη μεταβλητή X, με ένα ρυθμό όμως ο οποίος φθίνει και σε αντίθεση με τα προηγούμενα μοντέλα, δεν παρουσιάζεται κάποιο σημείο καμπής στην καμπύλη τους. Τελικά, η μεταβλητή *Y* σταθεροποιείται σε μία τιμή, η οποία είναι και η οριακή τιμή ανάπτυξης. Ένα τέτοιο μοντέλο είναι το:

$$Y = \beta_1 + \beta_2 e^{-\beta_3 X}$$

Το β_1 είναι η οριακή τιμή ανάπτυξης, αφού $Y \to \beta_1$ καθώς $X \to \infty$. Ακόμα, η αρχική τιμή, για *x*=0, είναι ίση με $\beta_1 + \beta_2$, ενώ τέλος από τις τιμές των β_2 και β_3 εξαρτάται ο ρυθμός με τον οποίο φθίνει η μεταβλητή *Y*, αφού είναι $Y' = -\beta_2 \beta_3 e^{-\beta_3 X}$. Στο Σχήμα 20, παρουσιάζεται το αντίστοιχο διάγραμμα για β_1=1, β_2=-0,9 και β_3=4 (a), για β_1=1, β_2=-0,8 και β_3=2 (b) και β_1=1, β_2=-1 και β_3=0,9 (c) .



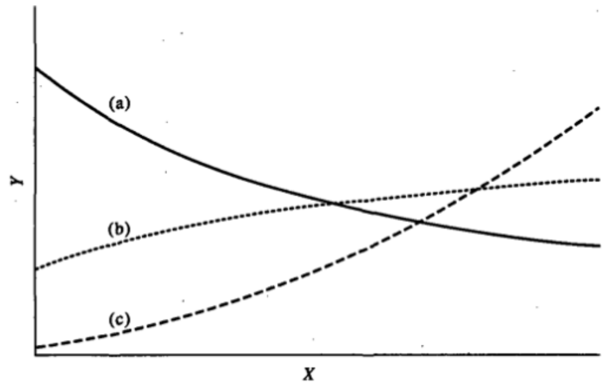
Σχήμα 20. Διαγράμματα ασυμπτωτικού μοντέλου

Μοντέλα Yield-Density (απόδοσης-πυκνότητας)

Τα μοντέλα Yield-Density έχουν εφαρμογή στη μελέτη της ανάπτυξης στον κλάδο της Γεωργίας, όπου και εμφανίστηκαν για πρώτη φορά. Ο σκοπός τους ήταν ο προσδιορισμός της σχέσης μεταξύ της απόδοσης μιας σοδειάς Y σε μία καλλιεργήσιμη έκταση, και της πυκνότητας φύτευσης X σε αυτή, για αυτό και αναφέρονται ως «μοντέλα απόδοσης-πυκνότητας».

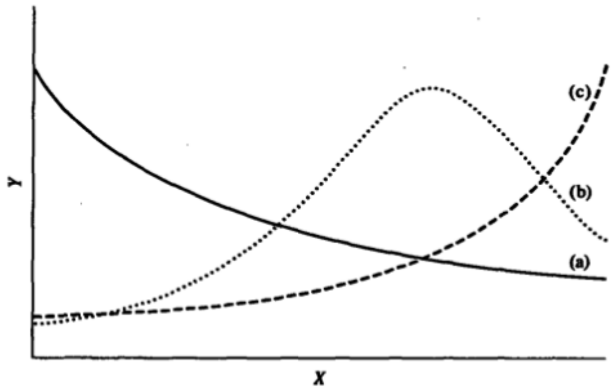
Αναφορικά με τις καμπύλες αυτών των μοντέλων, αυτές χωρίζονται σε δύο τύπους, στην «ασυμπτωτική καμπύλη» και στην «παραβολική καμπύλη». Στην πρώτη περίπτωση, η καμπύλη αυξάνεται σταθερά πλησιάζοντας ασυμπτωτικά μία τιμή και στη δεύτερη περίπτωση, η καμπύλη ενώ αυξάνεται μέχρι το βέλτιστο σημείο, στη συνέχεια ακολουθεί καθοδική πορεία. Παραδείγματα Yield-Density μοντέλων παρουσιάζονται παρακάτω, όπως και τα αντίστοιχα διαγράμματα τους.

- $Y = \dfrac{1}{\left(\beta_1 + \beta_2 X\right)^{\beta_3}}$. Στο Σχήμα 21 παρουσιάζονται τα αντίστοιχα διαγράμματα για β_1=0,6, β_2=1 και β_3=1(a), για β_1=0,1, β_2=1 και β_3=-0,3 (b) και β_1=0,2, β_2=1 και β_3=-2 (c). Και στις τρεις περιπτωσεις η αρχική τιμή, για X=0 είναι $Y = \dfrac{1}{\beta_1^{\beta_3}}$



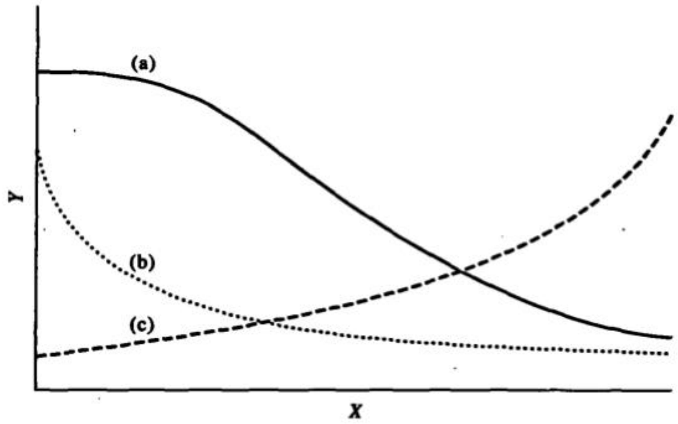
Σχήμα 21. Διαγράμματα Yield-Density μοντέλου I

- $Y = \dfrac{1}{\beta_1 + \beta_2 X + \beta_3 X^2}$. Στο Σχήμα 22 παρουσιάζονται τα αντίστοιχα διαγράμματα για β_1=0,1, β_2=3 και β_3=-0,2 (a), β_1=8,94, β_2=-22,4 και β_3=16 (b) και β_1=8, β_2=-8 και β_3=1 (c). Και στις τρεις περιπτωσεις η αρχική τιμή, για X=0 είναι $Y = \dfrac{1}{\beta_1}$.



Σχήμα 22. Διαγράμματα Yield-Density μοντέλου II

- $Y = \dfrac{1}{\beta_1 + \beta_2 X^{\beta_3}}$. Στο Σχήμα 23 παρουσιάζονται τα αντίστοιχα διαγράμματα για β_1=1, β_2=6 και β_3=3 (a), β_1=1,2, β_2=9 και β_3=0,9 (b) και β_1=10, β_2=-8,8 και β_3=0,5 (c). Και στις τρεις περιπτωσεις η αρχική τιμή, για X=0 είναι ίση με $Y = \dfrac{1}{\beta_1}$.



Σχήμα 23. Διαγράμματα Yield-Density μοντέλου ΙΙΙ

Ένα μη γραμμικό μοντέλο το οποίο θα χρησιμοποιηθεί στο επόμενο κεφάλαιο, είναι το μοντέλο Michaelis-Menten, το οποίο χρησιμοποιείται στη Βιοχημεία, προκειμένου να περιγράψει το «ρυθμό ενζυματικών αντιδράσεων» συσχετίζοντας τον «ρυθμό αντίδρασης» Y, που είναι η εξαρτημένη μεταβλητή, με τη «συγκέντρωση του υποστρώματος» X, η οποία είναι η ανεξάρτητη μεταβλητή. Το συγκεκριμένο μοντέλο έχει τη μορφή (Abedon, 2009):

$$Y = \frac{\beta_0 X}{\beta_1 + X} + \varepsilon .$$

Στην παραπάνω ισότητα, η παράμετρος β_0 εκφράζει το μέγιστο ρυθμό που επιτυγχάνει το σύστημα στο μέγιστο της συγκέντρωσης του υποστρώματος, αφού από την ισότητα $Y \approx \frac{\beta_0 X}{\beta_1 + X}$ έχουμε $Y \to \beta_0$ καθώς $X \to \infty$. Ακόμα, η παράμετρος β_1 εκφράζει τη συγκέντρωση του υποστρώματος στην οποία ο ρυθμός αντίδρασης είναι ίσος με $\frac{\beta_0}{2}$, αφού από την προηγούμενη ισότητα προκύπτει ότι $Y \approx \frac{\beta_0}{2}$ όταν είναι $X = \beta_1$. Τέλος, ε είναι τα κατάλοιπα.

Κεφάλαιο 3. Εφαρμογή στην R

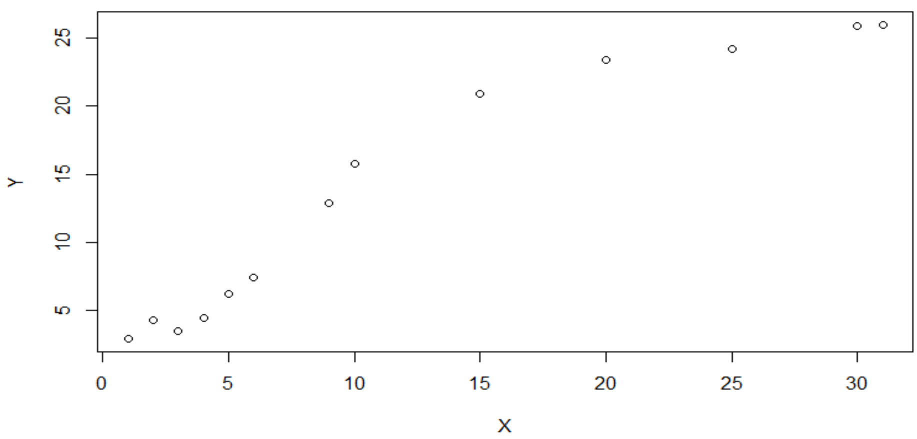
Για την εφαρμογή στην R θα χρησιμοποιηθεί το μη γραμμικό μοντέλο Michaelis-Menten, το οποίο περιγράφηκε παραπάνω. Για τις εντολές της R χρησιμοποιήθηκαν τα βιβλία των Fox & Weisberg (2011) και Ritz & Streibig (2008), ενώ θεωρήθηκαν τυχαία τα παρακάτω 13 ζεύγη τιμών, με τη λογική αυτών που βρέθηκαν στο Abedon (2009):

Y	2,9	4,3	3,5	4,4	6,2	7,4	12,9	15,8	20,9	23,4	25,2	25,9	26
X	1	2	3	4	5	6	9	10	15	20	25	30	31

Αρχικά εισάγουμε τις τιμές και κάνουμε και ένα διάγραμμα διασποράς

```
> Y=c(2.9,4.3,3.5,4.4,6.2,7.4,12.9,15.8,20.9,23.4,24.2,25.9,26)
> X=c(1,2,3,4,5,6,9,10,15,20,25,30,31)
> plot(X,Y)
```

Από το διάγραμμα διασποράς παρατηρείται απότομη αύξηση των τιμών της μεταβλητής Y η οποία προοδευτικά μειώνεται. Στην αρχή μικρές αυξήσεις της συγκέντρωσης του υποστρώματος οδηγούν σε μια σημαντική αύξηση του ρυθμού αντίδρασης, αλλά στη συνέχεια καθώς συνεχίζει να αυξάνεται η συγκέντρωση, η αύξηση του ρυθμού αντίδρασης είναι μικρότερη.



Σχήμα 24. Διάγραμμα διασποράς Yield-Density μοντέλου ΙΙ

Για την εκτίμηση των δύο παραμέτρων του μη γραμμικού μοντέλου η R χρησιμοποιεί τη συνάρτηση **nls**, με την μέθοδο Gauss-Newton, η οποία βασίζεται στη μέθοδο ελαχιστοποίησης της συνάρτησης

$$Q = \sum_{i=1}^{n} \left[Y_i - \frac{\beta_0 X_i}{\beta_1 + X_i} \right]^2$$

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, η συγκεκριμένη διαδικασία απαιτεί τον ορισμό κάποιων αρχικών τιμών τους. Αυτές μπορούν να βρεθουν από το Σχήμα 24. Η μεγαλύτερη τιμή της εξαρτημένης μεταβλητής Y είναι περίπου ίση με 25, οπότε μπορούμε να τη θεωρήσουμε ως αρχική τιμή για την παράμετρο β_0. Από την άλλη, για την παράμετρο β_1 πρέπει να θεωρήσουμε μια τιμή της συγκέντρωσης για την οποία η τιμή της μεταβλητής Y είναι ίση με $\frac{\beta_0}{2} = \frac{25}{2}$. Αυτή η τιμή είναι περίπου ίση με 9.

Πληκτρολογούμε την εντολή

```
> model=nls(Y~(b0*X/(b1+X)), start=list(b0=25,b1=9), trace=TRUE)
```

και παίρνουμε τα παρακάτω αποτελεσματα

```
227.9364 :  25   9
43.08237 :  42.40335 20.96836
32.51925 :  50.11358 26.40549
32.44118 :  50.16921 26.23440
32.44114 :  50.19643 26.26080
32.44114 :  50.19235 26.25697
32.44114 :  50.19295 26.25753
```

Όπως φαίνεται εκτελέστηκαν 6 επαναλήψεις μέχρι να συγκλίνει η μέθοδος. Οι εκτιμήσεις των παραμέτρων είναι β_0=50.19295 και β_1=26.25753 ενώ το άθροισμα τετραγώνων των καταλοίπων (residual sum-of-squares) βρέθηκε ίσο με 32.44114. Με στρογγυλοποίηση στα δύο δεκαδικά ψηφία, το εκτιμώμενο μοντέλο είναι το:

$$Y = \frac{50{,}19 \cdot X}{26{,}26 + X}$$

Για περισσότερες πληροφορίες πληκτρολογούμε:

```
> summary(model)
```

Και έχουμε:

```
Formula: Y ~ b0 * X/(b1 + X)

Parameters:
   Estimate Std. Error t value Pr(>|t|)
b0   50.193      7.079   7.090 2.02e-05 ***
b1   26.258      6.506   4.036  0.00196 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.717 on 11 degrees of freedom

Number of iterations to convergence: 6
Achieved convergence tolerance: 3.729e-06
```

Προκειμένου να εξεταστεί αν οι εκτιμήσεις των δύο παραμέτρων είναι στατιστικά σημαντικές, θεωρούμε αντίστοιχα τις υποθέσεις

$$(H_0: \beta_0=0 \text{ έναντι } H_1: \beta_0\neq 0) \quad \text{και} \quad (H_0: \beta_1=0 \text{ έναντι } H_1: \beta_1\neq 0)$$

Παρατηρώντας τις τιμές *Pr(>|t|)* για τις δύο παραμέτρους, οι μηδενικές υποθέσεις απορρίπτονται και στις δύο περιπτώσεις, αφού είναι p-value<0,05.

Για τον πίνακα διασποράς-συνδιασποράς των εκτιμητών, πληκτρολογούμε την εντολή

```
> vcov(model)
```

και έχουμε:

```
         b0       b1
b0 50.11742 45.05815
b1 45.05815 42.32735
```

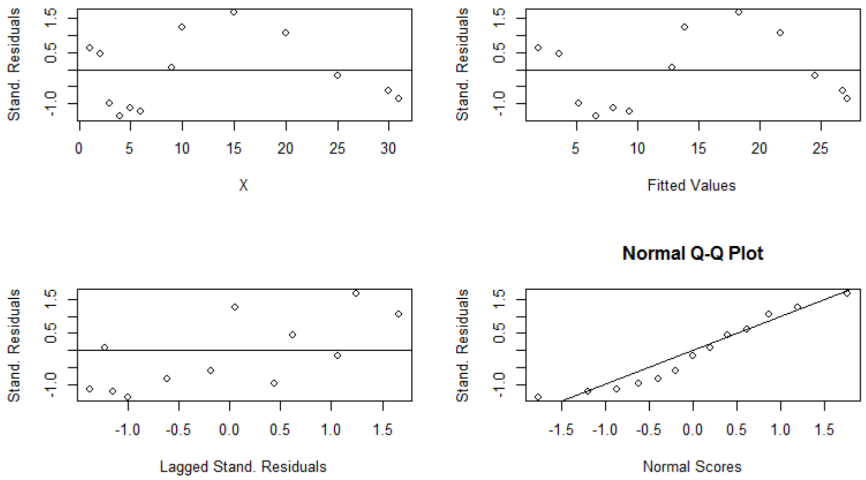
Ο υπολογισμός των τυποποιημένων καταλοίπων μπορεί να πραγματοποιηθεί ορίζοντας τον αριθμό των παραμέτρων p=2, με τις εντολές:

```
> v=diag(D%*%solve(crossprod(D))%*%t(D))
> p=2
> r=residuals(model)/(s*sqrt(1-v))
```

```
0.01264225             0.04007832             0.07163219
0.10135646             0.12626549             0.14519576
0.16814342             0.16718898             0.14033123
0.13755854             0.19843813             0.32871455
0.36245468
```

Για το γραφικό έλεγχο χρησιμοποιούμε τις παρακάτω εντολές

```
> par(mfrow=c(2,2))
> plot(r~X,ylab="Stand. Residuals",xlab="X")
> abline(a=0,b=0)
> plot(r~fitted(model),ylab="Stand. Residuals",xlab="Fitted Values")
> abline(a=0,b=0)
> plot(r,c(r[-1],NA),xlab="Lagged Stand. Residuals",ylab="Stand. Res
iduals")
> abline(a=0,b=0)
> qqnorm(r,ylab="Stand. Residuals",xlab="Normal Scores")
> abline(a=0,b=1)
```



Σχήμα 25. Γραφικοί έλεγχοι

Στα γραφήματα της πρώτης γραμμής έχουμε τις τιμές των τυποποιημένων καταλοίπων (residuals) σε σχέση με τις τιμές της ανεξάρτητης μεταβλητής X (αριστερά) και τις τιμές των τυποποιημένων καταλοίπων σε σχέση με τις προσαρμοσμένες τιμές της ανεξάρτητης μεταβλητής X (δεξιά). Στη δεύτερη γραμμή είναι τα τυποποιημένα κατάλοιπα ε_i έναντι των προηγούμενών τους ε_{i-1} (αριστερά), ενώ δεξιά είναι το Q-Q plot για το γραφικό έλεγχο της κανονικότητας.

Από τα διαγράμματα των καταλοίπων, δεν παρατηρείται κάποια τάση οπότε δεν παραβιάζονται οι υποθέσεις της ανεξαρτησίας και της ομοσκεδαστικότητας, ενώ υπάρχει και κανονικότητα. Για την κανονικότητα διενεργείται και ο έλεγχος Kolmogorov-Smirnov με την εντολή:

```
> ks.test(r,"pnorm",mean=mean(r),sd=sd(r),exact=FALSE)
```

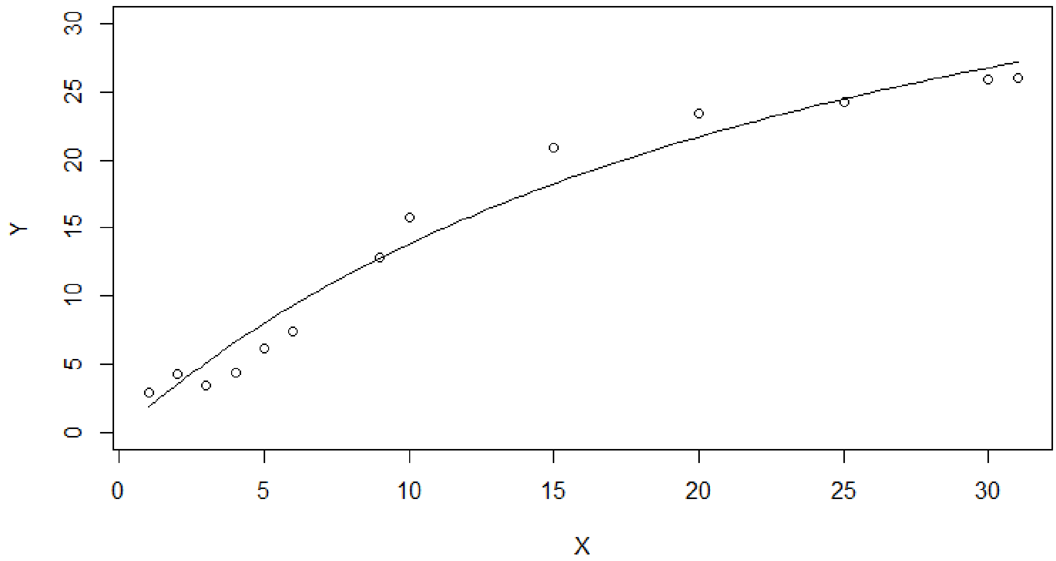
και έχουμε τα παρακάτω αποτελέσματα:

```
        One-sample Kolmogorov-Smirnov test

data:  r
D = 0.15381, p-value = 0.9182
alternative hypothesis: two-sided
```

Η μεγάλη τιμή για την p-value (0,9182) επιβεβαιώνει την υπόθεση της κανονικότητας.

Για να εξεταστεί κατά πόσο το μοντέλο Michaelis-Menten προσαρμόζεται στις παρατηρήσεις, προστίθεται στο αρχικό γράφημα των δεδομένων και η εκτιμώμενη καμπύλη του μοντέλου με τις εντολές:

```
> X.seq<-seq(range(X)[1],range(X)[2],length=100)
> Y.seq<-50.19*X.seq/(26.26+X.seq)
> plot(X,Y,ylim=c(0,30))
> lines(X.seq,Y.seq)
```



Σχήμα 26. Δεδομένα και εκτιμώμενη καμπύλη

Από το γράφημα φαίνεται πως η εκτιμώμενη καμπύλη προσαρμόζεται ικανοποιητικά στα δεδομένα. Στην περίπτωση που επιλέγαμε διαφορετικές αρχικές τιμές των παραμέτρων, θα είχαμε διαφορετικές εκτιμώμενες τιμές και τελικά ελαφρώς διαφορετική καμπύλη στο γράφημα. Για τη σύγκριση των εναλλακτικών μοντέλων χρησιμοποιείται το κριτήριο AIC. Καλύτερο μοντέλο είναι αυτό με την μικρότερη τιμή του κριτηρίου AIC. Για το συγκεκριμένο μοντέλο, και χρησιμοποιώντας την εντολή

```
> AIC(model)
```

βρίσκουμε τιμή ίση με

```
[1] 54.78062
```

Αν θέλουμε να υπολογίσουμε τα 95% διαστήματα εμπιστοσύνης για τις παραμέτρους παλινδρόμησης, χρησιμοποιούμε την εντολή:

```
> confint(model,level=0.95)
```

και βρίσκουμε

```
       2.5%    97.5%
b0 39.05654 72.02234
b1 16.44133 46.98837
```

Τέλος, αν θέλουμε από το μοντέλο να εκτιμήσουμε την τιμή του Y αν είναι γνωστό ότι είναι X=12, τότε χρησιμοποιούμε την εντολή:

```
> predict(model,newdata=data.frame(X=12))
```

και βρίσκουμε

```
[1] 15.74373
```

οπότε είναι Y=15.74

Προκειμένου να εκτιμηθεί η επίδραση των αρχικών τιμών στις τελικές εκτιμήσεις, θεωρήθηκαν τα παρακάτω ζεύγη αρχικών τιμών. Για κάθε ζεύγος τιμών υπολογίστηκαν οι συντελεστές β_0 και β_1 του ίδιου μοντέλου, ενώ υπολογίστηκαν επιπλέον και Residual standard errors και AIC. Τα αποτελέσματα παρουσιάζονται στον επόμενο πίνακα.

Αρχικές τιμές		Τελικές τιμές		Residual standard errors	AIC
β_0	β_1	β_0	β_1		
20	9	50.19283	26.25742	1.717322	54.78062
24	9	50.19302	26.25759	1.717322	54.78062
25	9	50.19295	26.25753	1.717322	54.78062
26	9	50.19301	26.25759	1.717322	54.78062
30	9	50.19291	26.25750	1.717322	54.78062
25	5	50.19292	26.25750	1.717322	54.78062
25	8	50.19274	26.25733	1.717322	54.78062
25	10	50.19305	26.25763	1.717322	54.78062
25	15	50.19291	26.25750	1.717322	54.78062

Πίνακας 2. Σύγκριση μοντέλων

Παρατηρώντας τις στήλες Residual standard errors και AIC στον Πίνακα 2 φαίνεται ότι δεν υπάρχει καμία διαφορά, οπότε τα μοντέλα θεωρούνται το ίδιο καλά. Μικρές διαφορές παρατηρούνται στις τιμές των συντελεστών β_0 και β_1, στο τρίτο δεκαδικό ψηφίο για τον πρώτο και στο τέταρτο δεκαδικό ψηφίο για τον δεύτερο.

Στη συνέχεια προχωρήσαμε σε προσαρμογή των ίδιων δεδομένων σε ένα μετασχηματισμένο γραμμικό μοντέλο. Για τον μετασχηματισμό έχουμε

$$Y \approx \frac{\beta_0 X}{\beta_1 + X} \Rightarrow \frac{1}{Y} \approx \frac{\beta_1 + X}{\beta_0 X} = \frac{\beta_1}{\beta_0} \cdot \frac{1}{X} + \frac{1}{\beta_0} \Rightarrow Y' = \beta_0' + \beta_1' X',$$

όπου είναι $Y' = \frac{1}{Y}$, $X' = \frac{1}{X}$, $\beta_0' = \frac{1}{\beta_0}$, $\beta_1' = \frac{\beta_1}{\beta_0}$.

Αρχικά μετασχηματίζουμε τα δεδομένα με τις παρακάτω εντολές:

```
> Y1=1/Y
> X1=1/X
```

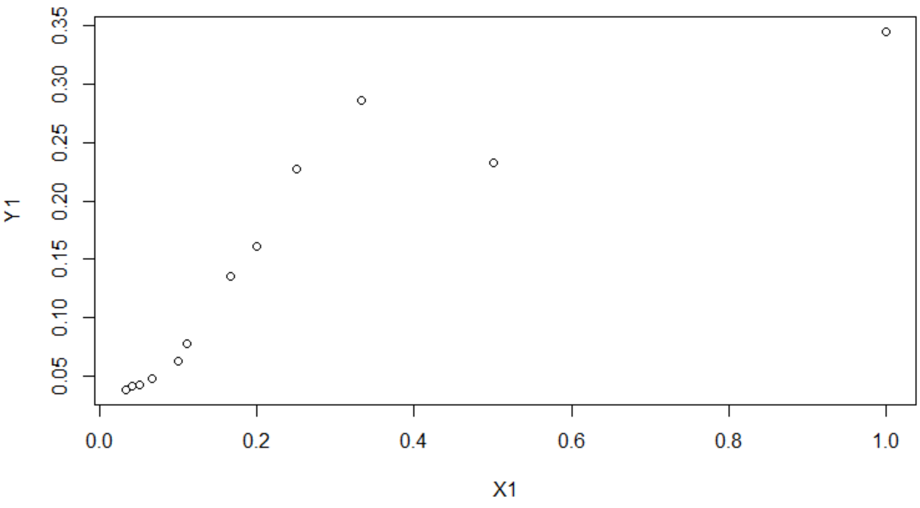
Τα δεδομένα έχουν την παρακάτω μορφή

$X' = \frac{1}{X}$	$Y' = \frac{1}{Y}$
1.00000000	0.34482759
0.50000000	0.23255814
0.33333333	0.28571429
0.25000000	0.22727273
0.20000000	0.16129032
0.16666667	0.13513514
0.11111111	0.07751938
0.10000000	0.06329114
0.06666667	0.04784689
0.05000000	0.04273504
0.04000000	0.04132231
0.03333333	0.03861004
0.03225806	0.03846154

Για το διάγραμμα διασποράς χρησιμοποιούμε την εντολή

```
>plot(X1,Y1)
```

και έχουμε το παρακάτω διάγραμμα



Σχήμα 27. Τα μετασχηματισμένα δεδομένα

Για τον υπολογισμό των συντελεστών χρησιμοποιούμε την παρακάτω εντολή

```
> lm(formula=Y1~X1)
```

και έχουμε

```
Coefficients:
(Intercept)           X1
    0.05658      0.34718
```

Τελικά έχουμε $\beta_0' = 0,05658$ $\beta_1' = 0,34718$ και $Y' = 0,05658 + 0.34718 \cdot X'$

Για περισσότερες πληροφορίες πληκτρολογούμε:

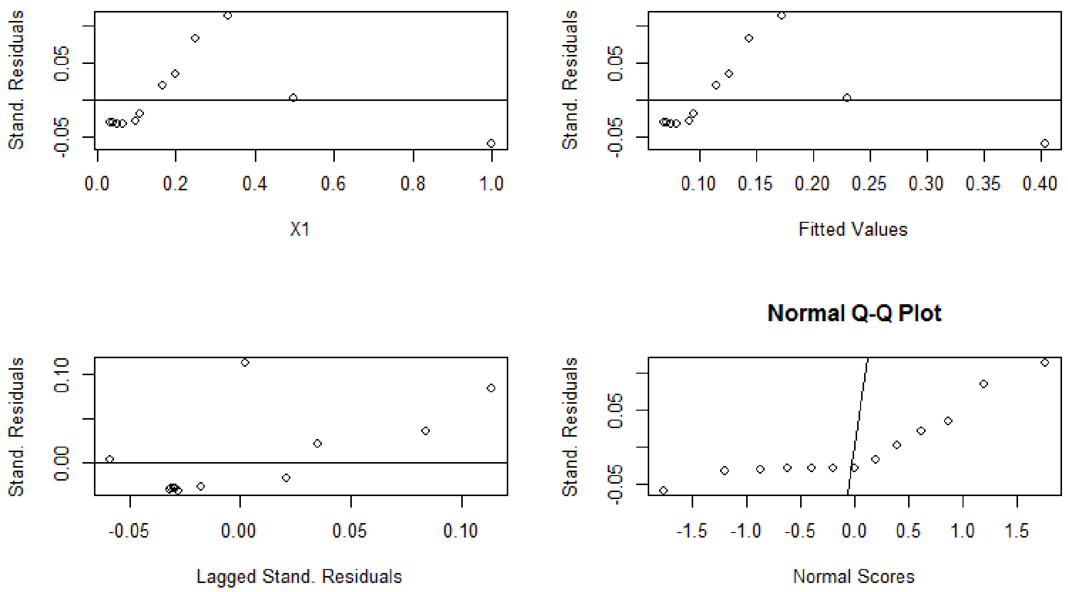
```
> summary(model_new)

Call:
lm(formula = Y1 ~ X1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.05893 -0.02954 -0.02801  0.02069  0.11341 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.05658    0.01924   2.941   0.0134 *  
X1           0.34718    0.05618   6.180 6.91e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05284 on 11 degrees of freedom
Multiple R-squared:  0.7764,	Adjusted R-squared:  0.7561 
F-statistic: 38.19 on 1 and 11 DF,  p-value: 6.906e-05
```

Και οι δύο συντελεστές είναι στατιστικά σημαντικοί σε επίπεδο 5%, αφού είναι `Pr(>|t|)= 0.0134` για το β_0' και `Pr(>|t|)= 6.91e-05` για το β_1'.

Ακόμα είναι R-squared=0,7764 το οποίο σημαίνει ότι το μοντέλο εξηγεί περίπου το 78% της διασποράς της μεταβλητής Y' και Residual standard error=0.05284 το οποίο είναι μικρότερο από το μη γραμμικό.

Για το γραφικό έλεγχο χρησιμοποιούμε τις παρακάτω εντολές

```
> par(mfrow=c(2,2))
> plot(r~X1,ylab="Stand. Residuals",xlab="X1")
> abline(a=0,b=0)
> plot(r~fitted(model_new),ylab="Stand. Residuals",xlab="Fitted Valu
es")
> abline(a=0,b=0)
> plot(r,c(r[-1],NA),xlab="Lagged Stand. Residuals",ylab="Stand. Res
iduals")
> abline(a=0,b=0)
> qqnorm(r,ylab="Stand. Residuals",xlab="Normal Scores")
> abline(a=0,b=1)
```



Σχήμα 28. Γραφικός έλεγχος για το μετασχηματισμένο μοντέλο

Στα γραφήματα της πρώτης γραμμής έχουμε τις τιμές των τυποποιημένων καταλοίπων (residuals) σε σχέση με τις τιμές της ανεξάρτητης μεταβλητής X' (αριστερά) και τις τιμές των τυποποιημένων καταλοίπων σε σχέση με τις προσαρμοσμένες τιμές της ανεξάρτητης μεταβλητής X' (δεξιά). Στη δεύτερη γραμμή είναι τα τυποποιημένα κατάλοιπα ε_i έναντι των προηγούμενών τους ε_{i-1} (αριστερά), ενώ δεξιά είναι το Q-Q plot για το γραφικό έλεγχο της κανονικότητας.

Από τα διαγράμματα των καταλοίπων, παρατηρείται ότι παραβιάζονται οι υποθέσεις της ανεξαρτησίας και της ομοσκεδαστικότητας, ενώ δεν υπάρχει και κανονικότητα.

Τέλος, αυτό που παρατηρούμε είναι πως δουλεύοντας, με μετασχηματισμό δεν παίρνουμε μοντέλο που ικανοποιεί τις απαραίτητες προϋποθέσεις για να προχωρήσουμε σε στατιστική συμπερασματολογία. Αντίθετα, η αντίστοιχη μη γραμμική προσέγγιση μας οδήγησε σε μοντέλο που μπορεί να χρησιμοποιηθεί για στατιστικά συμπεράσματα.

Βιβλιογραφία

Abedon, S.T. (2009). Kinetics of phage-mediated biocontrol of bacteria. *Foodborne Pathog Dis*. 6 (7): 807–15.

Bates, D.M. & Watts, D.G. (1988). *Nonlinear Regression Analysis & Its Applications*. New York: John Wiley & Sons.

Cohen, J., Cohen, P., West, S.G. & Aiken L.S. (2003). *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences*. 3rd ed. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Daniel, C. & Wood, F.S. (1980). *Fitting Equations to Data*. New York: John Wiley & Sons.

Draper, N.R. & Smith, H. (1998). *Applied Regression Analysis*. 3rd ed. New York: John Wiley & Sons.

Fox, J. & Weisberg, S. (2011). *An R Companion to Applied Regression* (2nd Ed.). New York: SAGE Publications, Inc.

Gallant, A.R. (1987). Nonlinear Statistical Models. New York: John Wiley & Sons.

Kutner, M.H., Nachtsheim, C.J., Neter, J. & Li, W. (2005). *Applied Linear Statistical Models* (Fifth Edition). New York: McGraw-Hill.

Montgomery, D.C. & Peck, D.G. (1991). *Introduction to Linear Regression Analysis*. (2nd Ed,) New York: John Wiley & Sons.

Rawlings, J.O., Pantula, S.G. & Dickey, D.A. (1998). *Applied Regression Analysis A Research Tool* (Second Edition). New York: Springer-Verlag, Inc.

Ritz, C. & Streibig, J.C. (2008). Nonlinear Regression with R. New York: Springer.

Seber, G. & Wild, C. (1989). *Nonlinear regression*. New York: John Wiley & Sons.