

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

*ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ
ΣΤΑΤΙΣΤΙΚΗ*

***ΤΟ ΜΟΝΤΕΛΟ ΔΟΣΗ - ΑΠΟΚΡΙΣΗ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥ
ΣΤΗΝ ΒΙΟΣΤΑΤΙΣΤΙΚΗ***

ΠΑΤΜΙΟΥ ΑΝΑΣΤΑΣΙΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΟΥ ΥΠΟΒΛΗΘΗΚΕ ΣΤΟ ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ ΤΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΠΕΙΡΑΙΩΣ ΩΣ ΜΕΡΟΣ ΤΩΝ
ΑΠΑΙΤΗΣΕΩΝ ΓΙΑ ΤΗΝ ΑΠΟΚΤΗΣΗ ΤΟΥ ΜΕΤΑΠΤΥΧΙΑΚΟΥ
ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

ΠΕΙΡΑΙΑΣ

ΣΕΠΤΕΜΒΡΙΟΣ 2023

UNIVERSITY OF PIRAEUS
SCHOOL OF FINANCE AND STATISTICS



DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN APPLIED STATISTICS

**THE DOSE RESPONSE MODEL AND ITS APPLICATION TO
BIOSTATISTICS**

by

PATMIOU ANASTASIA

MSC DISSERTATION

**SUBMITTED TO THE DEPARTMENT OF STATISTICS AND INSURANCE
SCIENCE OF UNIVERSITY OF PIRAEUS IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN
APPLIED STATISTICS.**

PIRAEUS

SEPTEMBER 2023

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ..... συνεδρίαση του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της επιτροπής ήταν:

1. Τζαβελάς Γεώργιος, Αναπληρωτής Καθηγητής (επιβλέπων)
2. Πολίτης Κωνσταντίνος, Αναπληρωτής Καθηγητής
3. Βερροπούλου Γεωργία, Καθηγήτρια

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

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω πολύ τον επιβλέποντα καθηγητή μου κ. Γεώργιο Τζαβελά για την βοήθεια του, καθώς επίσης και για την επιμονή και υπομονή που είχε μαζί μου όλο αυτό το διάστημα.

ΠΕΡΙΛΗΨΗ

Το μοντέλο δόσης-απόκρισης μελετά τον τρόπο με τον οποίο μια ουσία ή ένα φάρμακο επηρεάζει ένα βιολογικό σύστημα, σε σχέση με την ποσότητα της δόσης που χορηγείται. Βοηθάει στο να κατανοήσουμε κατά πόσο η απόκριση αυξάνεται ή μειώνεται ανάλογα με την χορήγηση της δόσης και εφαρμόζεται σε ποικίλους τομείς. Σκοπός του μοντέλου αυτού είναι η καλύτερη εκτίμηση της αποτελεσματικότητας και ασφάλειας των φαρμάκων και άλλων χημικών ουσιών, συμβάλλοντας στη βελτίωση της φαρμακευτικής αγωγής και της υγείας του ανθρώπου.

Στο πρώτο κεφάλαιο, εξετάζεται η έννοια και η μοντελοποίηση του μοντέλου δόσης-απόκρισης, η διερεύνηση των διαφορετικών μετρήσεων (ED_{50} κλπ) και η μελέτη των ποικίλων καμπυλών δόσης-απόκρισης. Όλα τα γραφήματα που περιλαμβάνονται σε αυτό το κεφάλαιο έχουν αναπαραχθεί με την χρήση των πακέτων Microsoft Excel και R.

Στο δεύτερο κεφάλαιο, γίνεται η μαθηματική ανάλυση των μοντέλων δόσης-απόκρισης. Σε αυτή την ενότητα εξετάζονται τα διαφορετικά μοντέλα δόσης-απόκρισης, οι συναρτήσεις σύνδεσης και η γενίκευσή τους, καθώς και η παλινδρόμηση τριών παραμέτρων.

Το τρίτο κεφάλαιο περιέχει εφαρμογές των διαφόρων μοντέλων με την χρήση του στατιστικού πακέτου R. Χρησιμοποιείται η βιβλιοθήκη `drm` από την οποία και αντλούμε το μεγαλύτερο ποσοστό των δεδομένων μας. Πιο συγκεκριμένα στην ενότητα αυτή γίνεται σύγκριση των μοντέλων `logit`, `probit` και `cloglog` (δεδομένα από τους : Jeske, D. R., Xu, H. K., Blessinger, T., Jensen, P. and Trumble, J. (2009)), μελετάται η συνεχής απόκριση όταν έχουμε μια καμπύλη δόσης-απόκρισης (δεδομένα από τους : Inderjit and J. C. Streibig, and M. Olofsdotter (2002)) και δύο καμπύλες δόσης-απόκρισης (δεδομένα από τους : Christensen, M. G. and Teicher, H. B., and Streibig, J. C. (2003)), και τέλος εξετάζεται η διωνυμική απόκριση με την χρήση μη γραμμικής παλινδρόμησης (δεδομένα από την : Nina Cedergreen (2013)).

Τέλος, γίνεται μια ανασκόπηση των αποτελεσμάτων που έχουμε βρεί από κάθε εφαρμογή, με σκοπό την σημείωση των βασικών συμπερασμάτων στα οποία οδηγηθήκαμε μετά την ανάλυση.

ABSTRACT

The dose-response model examines how a substance or drug influences a biological system concerning the amount of dose administered. It aids in determining whether the response increases or decreases in proportion to the given dose and finds applications across multiple domains. This model's objective is to enhance the appraisal of drug and chemical substance efficacy and safety, thereby contributing to refining pharmaceutical treatment and human well-being.

In the initial chapter, we delve into the concept and modeling of the dose-response model, explore different measurements like ED_{50} , and analyze various dose-response curves. All graphs featured in this chapter are generated utilizing Microsoft Excel and R software.

The subsequent chapter entails the mathematical scrutiny of dose-response models. This segment assesses distinct dose-response models, link functions, their extensions, and the regression involving three parameters.

The third chapter contains applications of various models using the statistical package R. We utilize the `drc` library, from which we extract the majority of our data. More specifically, in this section, we compare the logit, probit, and cloglog models (data from: Jeske, D. R., Xu, H. K., Blessinger, T., Jensen, P., and Trumble, J. (2009)). We examine continuous response when we have a dose-response curve (data from: Inderjit and J. C. Streibig, and M. Olofsdotter (2002)), and two dose-response curves (data from: Christensen, M. G. and Teicher, H. B., and Streibig, J. C. (2003)). Finally, we investigate binomial response using non-linear regression (data from: Nina Cedergreen (2013)).

Ultimately, we provide an overview of the outcomes derived from each application, aiming to encapsulate the fundamental findings reached post-analysis.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή στο Μοντέλο Δόσης-Απόκρισης	1
1.1 Η Έννοια του Μοντέλου Δόσης-Απόκρισης	1
1.2 Μοντελοποίηση Δόσης-Απόκρισης	3
1.3 Διερεύνηση Μετρήσεων LD_{50} , ED_{50} , EC_{50} , TD_{50} , IC_{50} και MIC.....	4
1.4 Καμπύλες Δόσης Απόκρισης	6
1.5 Γραφικές Παραστάσεις Διαφόρων Καμπυλών Δόσης-Απόκρισης.....	7
1.6 Εξίσωση Hill	10
1.7 Διάφοροι Τύποι Μονότονων Σχέσεων Δόσης-Απόκρισης	12
1.8 Μη Μονότονες Σχέσεις Δόσης-Απόκρισης	16

ΚΕΦΑΛΑΙΟ 2

Μαθηματική Ανάλυση Μοντέλων Δόσης-Απόκρισης	20
2.1 Μοντέλα Δόσης-Απόκρισης	20
2.2 Συνάρτηση Σύνδεσης	24
2.2.1 Οι Συνιστώσες Ενός Γενικευμένου Γραμμικού Μοντέλου	24
2.2.2 Η Γενίκευση	25
2.2.3 Συναρτήσεις Σύνδεσης.....	29
2.2.4 Επαρκής Στατιστική και Κανονικές Συνδέσεις	31
2.3 Συνάρτησεις Σύνδεσης για Μοντέλα με Δυαδικές Αποκρίσεις.....	32
2.4 Παλινδρόμηση Τριών Παραμέτρων	36

ΚΕΦΑΛΑΙΟ 3

Εφαρμογές Μοντέλων Δόσης-Απόκρισης με Πραγματικά Δεδομένα	41
3.1 Διωνυμική Απόκριση : Σύγκριση Μοντέλων logit, probit και cloglog	41
3.2 Συνεχής Απόκριση : Μία Καμπύλη Δόσης-Απόκρισης	47
3.3 Συνεχής Απόκριση : Δύο Καμπύλες Δόσης-Απόκρισης	54
3.4 Διωνυμική Απόκριση : Γενικευμένη Μη Γραμμική Παλινδρόμηση	61

Σύνοψη Αποτελεσμάτων	68
Βιβλιογραφία	70
Παράρτημα 1	72
Παράρτημα 2	73
Παράρτημα 3	74
Παράρτημα 4	76

Κατάλογος Σχημάτων

1.1 Παράδειγμα Μονότονης Καμπύλης	8
1.2 Παράδειγμα Πολλαπλών Μονότονων Καμπύλων	9
1.3 Παράδειγμα Μονότονης Καμπύλης με Καθοδική Κλίση	10
1.4 Καμπύλη Δόση-Απόκριση σε Σχέση με τον Συντελεστή Υπερευαισθησίας	11
1.5 Γραμμικό Μοντέλο	12
1.6 Γραμμικό Μοντέλο Λογαριθμικής Δόσης	13
1.7 Εκθετικό Μοντέλο	14
1.8 Υπερβολικό Μοντέλο E_{max}	15
1.9 Σιγμοειδές Μοντέλο E_{max}	16
1.10 Ενδεικτικό Σχήμα Μη Μονότονης Καμπύλης – Περίπτωση 1	17
1.11 Ενδεικτικό Σχήμα Μη Μονότονης Καμπύλης – Περίπτωση 2	18
1.12 Ενδεικτικό Σχήμα Μη Μονότονης Καμπύλης – Περίπτωση 3	18
2.1 Ομοιόμορφη κατανομή: $f(s)$ και π	21
2.2 Κανονική κατανομή: $f(s)$ και π	22
2.3 Γραφική Σύγκριση Συναρτήσεων g_1, g_2, g_3, g_4	35
3.1 Ιστόγραμμα Μυγών Που Απεβίωσαν	42
3.2 Καμπύλη Δόσης-Απόκρισης	43
3.3 Διάγραμμα Διασποράς Φερουλικού Οξέος-Μήκος Ρίζας Σίκαλης	50
3.4 Καμπύλη Δόσης – Απόκρισης Μήκους Ρίζας Σίκαλης Με Φερουλικό Οξύ.....	53
3.5 Καμπύλη Δόσης-Απόκρισης Χωρίς Διαχωρισμό Ζιζανιοκτόνων	56
3.6 Καμπύλες Δόσης-Απόκρισης ανά Ζιζανιοκτόνο	59

3.7 Καμπύλη Δόσης-Απόκρισης Γαιοσκωλήκων Που Παρέμειναν Στο Μολυσμένο Δοχείο64

Κατάλογος Πινάκων

2.1 Χαρακτηριστικά Ορισμένων Μονομετάβλητων Κατανομών Στην Εκθετική Οικογένεια (McCullagh, P. and Nelder, J.A.) 29

2.2 Ποσοστό Θνησιμότητας38

3.1 Αποτελέσματα μοντέλων logit, probit και cloglog46

ΚΕΦΑΛΑΙΟ 1^ο : Εισαγωγή στο Μοντέλο Δόσης-Απόκρισης

1.1 Η Έννοια του Μοντέλου Δόσης-Απόκρισης

Το μοντέλο δόσης απόκρισης περιγράφει το μέγεθος της απόκρισης ενός οργανισμού, ως μια συνάρτηση της έκθεσης (ή των δόσεων) σε ένα ερέθισμα ή στρεσογόνο παράγοντα (συνήθως μια χημική ουσία) μετά από ένα συγκεκριμένο χρονικό διάστημα. Μια συνάρτηση απόκρισης ερεθίσματος ή καμπύλη απόκρισης ερεθίσματος ορίζεται ευρύτερα ως η απόκριση από οποιοδήποτε τύπο ερεθίσματος, χωρίς να περιορίζεται σε χημικές ουσίες.

Η μελέτη της δόσης-απόκρισης και η ανάπτυξη μοντέλων δόσης-απόκρισης είναι ζωτικής σημασίας, δεδομένου ότι μας βοηθάει να κατανοήσουμε πως λειτουργεί ένα φάρμακο και να ερευνήσουμε κατα πόσο έχει τις επιθυμητές ενέργειες μιας πιθανής νέας θεραπείας. Πιο συγκεκριμένα μας ενδιαφέρει ο προσδιορισμός των ασφαλών, επικίνδυνων και (όπου χρειάζεται) ωφέλιμων επιπέδων και δόσεων για φάρμακα, ρύπους, τρόφιμα και άλλες ουσίες στις οποίες εκτίθενται άνθρωποι ή άλλοι οργανισμοί. Οι σχέσεις δόσης-απόκρισης μπορούν να χρησιμοποιηθούν σε άτομα ή σε πληθυσμούς. Όσον αφορά τα άτομα, μια μικρή ποσότητα τοξίνης μπορεί να μην έχει σημαντικό αποτέλεσμα, ενώ μια μεγάλη ποσότητα μπορεί να είναι θανατηφόρα. Αντίθετα αυτό που μελετάμε σε πληθυσμούς είναι ο τρόπος με τον οποίον επηρεάζονται ομάδες ανθρώπων ή οργανισμών σε διαφορετικά επίπεδα έκθεσης. Σε πολλές περιπτώσεις υπάρχει η ανάγκη διερεύνησης για περισσότερες από μια απόκριση. Αυτό συμβαίνει επειδή μια θεραπεία θα οδηγήσει γενικά σε πολλαπλές βιολογικές αντιδράσεις και χρειάζεται να προσπαθήσει κανείς να ξεχωρίσει τις άμεσες από τις έμμεσες αποκρίσεις και τις επιθυμητές από τις ανεπιθύμητες ενέργειες.

Ο προσδιορισμός των διαφορετικών επιπέδων της δόσης, αποτελεί συχνά τη βάση για τη δημόσια πολιτική. Η Υπηρεσία Προστασίας του Περιβάλλοντος των ΗΠΑ έχει αναπτύξει εκτενείς οδηγίες και εκθέσεις σχετικά με τη μοντελοποίηση και την αξιολόγηση δόσης-απόκρισης, καθώς και το λογισμικό. Ο Οργανισμός Τροφίμων και Φαρμάκων των ΗΠΑ έχει επίσης οδηγίες για την αποσαφήνιση των σχέσεων δόσης-απόκρισης κατά την ανάπτυξη φαρμάκων.

Τα μοντέλα δόσης-απόκρισης δεν περιορίζονται σε ένα ενιαίο μαθηματικό πλαίσιο. Έχουν αναπτυχθεί διάφορα μοντέλα, το καθένα καλύπτοντας συγκεκριμένα χαρακτηριστικά δεδομένων και ερευνητικούς στόχους. Τα γενικευμένα γραμμικά μοντέλα προσφέρουν μια απλή αναπαράσταση της σχέσης μεταξύ δόσης-απόκρισης και είναι αυτά που χρησιμοποιούνται στις περισσότερες περιπτώσεις λόγω του ότι είναι πιο εύκολα στην ανάλυση. Πιο πολύπλοκα μοντέλα, όπως το μοντέλο Emax, αντιπροσωπεύουν μια μέγιστη επιτεύξιμη απόκριση και μια δόση στην οποία η απόκριση είναι η μισή-μέγιστη (ED_{50}). Παρ'όλο που αυτά τα μοντέλα χρησιμεύουν ως θεμελιώδη δομικά στοιχεία, υπάρχουν και περιπτώσεις στις οποίες δεν μπορούμε να ισχυριστούμε ότι υπάρχει γραμμικότητα και άρα δεν μπορούμε να

χρησιμοποιήσουμε ένα γενικευμένο γραμμικό μοντέλο για την ανάλυση ενός προβλήματος. Για τον λόγο αυτό υπάρχουν και τα μη γενικευμένα γραμμικά, όπου μπορούν να χρησιμοποιηθούν για να συλλάβουν μοναδικά χαρακτηριστικά ή πολυπλοκότητες των δεδομένων, που τα γενικευμένα γραμμικά μοντέλα δεν μπορούν να εκτιμήσουν. Αξίζει να σημειωθεί ότι, η απόκριση μπορεί να μεταβληθεί με πολλούς διαφορετικούς τρόπους, ανάλογα με τη δόση που θα χορηγηθεί, και πολλές από τις σχέσεις δόσης-απόκρισης είναι πολύπλοκες ειδικά σε περιπτώσεις όπου δεν υπάρχει γραμμικότητα.

Η επιλογή του κατάλληλου μοντέλου δόσης-απόκρισης εξαρτάται από διάφορους παράγοντες, συμπεριλαμβανομένης της φύσης της μεταβλητής απόκρισης, των υποκείμενων βιολογικών μηχανισμών και του συγκεκριμένου ερευνητικού ερωτήματος. Χρησιμοποιούνται στατιστικές τεχνικές για την εκτίμηση των παραμέτρων του μοντέλου, την επικύρωση της καλής προσαρμογής του μοντέλου και τη διεξαγωγή συμπερασμάτων σχετικά με τη σχέση δόσης-απόκρισης. Αυτές οι τεχνικές επιτρέπουν στους ερευνητές να εξάγουν ουσιαστικά συμπεράσματα, να ποσοτικοποιούν την αβεβαιότητα που σχετίζεται με τις εκτιμήσεις παραμέτρων και να αξιολογούν τη στατιστική σημασία της σχέσης μεταξύ δόσης και απόκρισης.

Τέλος, ένα αναπόσπαστο κομμάτι της ανάλυσης τέτοιου είδους μοντέλων είναι οι καμπύλες δόσης-απόκρισης. Οι καμπύλες δόσης-απόκρισης είναι γραφικές αναπαραστάσεις που απεικονίζουν τη σχέση μεταξύ της δόσης ή του επιπέδου έκθεσης μιας ουσίας και της αντίστοιχης βιολογικής απόκρισης. Αυτές οι καμπύλες παρέχουν πολύτιμες πληροφορίες για το πώς η ένταση ή το μέγεθος της απόκρισης αλλάζει καθώς η δόση αυξάνεται ή μειώνεται. Οι καμπύλες δόσης-απόκρισης χρησιμοποιούνται ευρέως σε τομείς όπως η φαρμακολογία, η τοξικολογία, η περιβαλλοντική υγεία και η επιδημιολογία για την αξιολόγηση της ισχύος, της αποτελεσματικότητας, της ασφάλειας και των πιθανών δυσμενών επιπτώσεων των ουσιών.

Το σχήμα μιας καμπύλης δόσης-απόκρισης μπορεί να ποικίλλει ανάλογα με τη συγκεκριμένη ουσία, το υπό μελέτη βιολογικό σύστημα και τον τρόπο δράσης. Για παράδειγμα, ορισμένα κοινά μοτίβα που παρατηρούνται στις καμπύλες δόσης-απόκρισης είναι τα εξής : Γραμμική Απόκριση, Σιγμοειδής Απόκριση, Εκθετική Απόκριση, Μη Μονότονη Απόκριση κ.λ.π.

Όλα τα προαναφερθέντα, θα αναλυθούν εκτενέστερα σε επόμενες ενότητες τόσο θεωρητικά όσο και πρακτικά, ώστε να γίνουν πλήρως κατανοητά.

1.2 Μοντελοποίηση Δόσης-Απόκρισης

Τα μοντέλα δόσης-απόκρισης έχουν ως στόχο την περιγραφή της εξάρτησης μιας συγκεκριμένης απόκρισης από την δόση. Η δόση ποσοτικοποιεί την ποσότητα του φαρμάκου στο οποίο εκτίθεται το άτομο. Η δόση που καταναλώνεται από το στόμα είναι συνήθως σε χιλιοστόγραμμα, μικρογραμμάρια ή γραμμάρια ανά κιλό σωματικού βάρους. Ενώ, η δόση που καταναλώνεται μέσω της εισπνοής, είναι σε χιλιοστόγραμμα ανά κυβικό μέτρο αέρα περιβάλλοντος. Άλλες μονάδες δόσης περιλαμβάνουν γραμμομόρια ανά σωματικό βάρος, γραμμομόρια ανά ζώο και για δερματική έκθεση γραμμομόρια ανά τετραγωνικό εκατοστό.

Τα παραδείγματα κυμαίνονται από συνεχή δεδομένα, π.χ. δεδομένα βάρους σώματος, έως διωνυμικά δεδομένα όπου παρατηρείται η παρουσία ενός συμβάντος (ναι ή όχι) ή πολυωνυμικά δεδομένα όπου πολλαπλά επίπεδα είναι πιθανά, όπως διαφορετικές βαθμολογίες πόνου ή τύποι ανεπιθύμητων ενεργειών.

Αν και ο σκοπός της μοντελοποίησης της δόσης-απόκρισης για κάθε πιθανό τύπο απόκρισης είναι ο ίδιος, δηλαδή να βρεθεί η βέλτιστη δόση, η μεθοδολογία πίσω από αυτό διαφέρει πολύ ανα περίπτωση.

Οι Ruberg (1995a,b) και Chuang-Stein and Agresti (1997) διατύπωσαν τέσσερις κύριες ερωτήσεις που τίθενται συνήθως σε μελέτες δόσης-απόκρισης :

1. Υπάρχουν στοιχεία για την επίδραση του φαρμάκου;
2. Για ποιες δόσεις διαφέρει η απόκριση;
3. Ποιά είναι η φύση της σχέσης δόσης-απόκρισης;
4. Ποιά είναι η βέλτιστη δόση;

Για να δώσουμε απάντηση σε αυτές τις ερωτήσεις, θα χρησιμοποιήσουμε είτε την μονότονη τάση της απόκρισης σε σχέση με τη δόση ή την μοντελοποίηση της σχέσης δόσης-απόκρισης. Και στις δύο περιπτώσεις, η βασική μας υπόθεση είναι ότι η πραγματική σχέση μεταξύ της δόσης και της απόκρισης είναι μονότονη. Σε ορισμένες εφαρμογές, η υπόθεση της μονοτονίας δεν είναι κατάλληλη και άλλα μη μονότονα προφίλ όπως η μονοτροπική μερική σειρά και τα κυκλικά μοτίβα θα πρέπει να ληφθούν υπόψη. Στην περίπτωση της μονοτροπικής σειράς, η απόκριση αυξάνεται με την αύξηση της δόσης μόνο για ένα συγκεκριμένο εύρος δοσολογίας, και μετά αρχίζει να μειώνεται, δημιουργώντας μια "μονοτροπική" σχέση μόνο για αυτό το τμήμα της καμπύλης. Ενώ στα κυκλικά μοτίβα, η σχέση δόσης-απόκρισης μπορεί να παρουσιάζει εναλλαγές ανάμεσα σε αυξητικές και μειωτικές τάσεις με την αύξηση της δόσης.

Η μοντελοποίηση της δόσης-απόκρισης αναφέρεται στην εφαρμογή μιας μαθηματικής αναπαράστασης κάποιας αληθινής και άγνωστης σχέσης. Τα μοντέλα δόσης-απόκρισης μπορούν να ταξινομηθούν ως εμπειρικής ή μηχανιστικής φύσης. Ένα εμπειρικό μοντέλο, όπως το λογιστικό μοντέλο τεσσάρων παραμέτρων, εξυπηρετεί επαρκώς το παρατηρούμενο μοτίβο μεταξύ μιας δόσης και μιας απόκρισης χωρίς να κατανοεί την υποκείμενη βιολογική

διαδικασία. Δηλαδή, οι παράμετροι που υπάρχουν στο μοντέλο δεν αντιπροσωπεύουν βιολογικές διεργασίες. Ενώ, το μηχανιστικό μοντέλο χρησιμοποιεί μηχανιστικά μονοπάτια για να εξηγήσει το παρατηρούμενο μοτίβο.

1.3 Διερεύνηση Μετρήσεων LD_{50} , ED_{50} , EC_{50} , TD_{50} , IC_{50} και MIC

Αυτή η ενότητα στοχεύει στην διερεύνηση και ανάλυση των μέτρων δόσης-απόκρισης, τα οποία είναι αρκετά σημαντικά για την αξιολόγηση των ουσιών. Συνεπώς, θα εστιάσουμε σε έξι σημαντικά μέτρα: LD_{50} , ED_{50} , EC_{50} , TD_{50} , IC_{50} και MIC. Στις παρακάτω ενότητες γίνονται αναφορές στα συγκεκριμένα μέτρα, αν όχι σε όλα στα περισσότερα, καθώς χρησιμοποιούνται σε μαθηματικούς τύπους, καμπύλες αλλά και υπολογίζονται σε κάποια παραδείγματα μέσω της R.

LD_{50} (θανατηφόρα δόση 50%) : Το LD_{50} είναι ένα μέτρο οξείας τοξικότητας και αντιπροσωπεύει τη δόση μιας ουσίας που προκαλεί θάνατο στο 50% του πληθυσμού ή στα πειραματικά άτομα εντός συγκεκριμένου χρονικού πλαισίου. Εκφράζεται συνήθως σε χιλιοστόγραμμα ουσίας ανά κιλό σωματικού βάρους (mg/kg) για εκθέσεις από το στόμα ή στο δέρμα ή σε μέρη ανά εκατομμύριο (ppm) για εκθέσεις με εισπνοή. Το LD_{50} τυπικά προσδιορίζεται μέσω μελετών σε ζώα, όπου χορηγούνται διαφορετικές δόσεις μιας ουσίας και παρατηρούνται τα προκύπτοντα ποσοστά θνησιμότητας. Οι τιμές LD_{50} χρησιμοποιούνται για την αξιολόγηση της τοξικότητας και των πιθανών κινδύνων που σχετίζονται με την έκθεση σε χημικές ουσίες ή ουσίες.

ED_{50} (αποτελεσματική δόση 50%) : Το ED_{50} είναι ένα μέτρο της ισχύος ενός φαρμάκου ή μιας θεραπείας και αντιπροσωπεύει τη δόση που παράγει ένα συγκεκριμένο θεραπευτικό αποτέλεσμα στο 50% του πληθυσμού ή των πειραματιζόμενων ατόμων. Συνήθως εκφράζεται στις ίδιες μονάδες με τη χορηγούμενη δόση, όπως χιλιοστόγραμμα (mg) ή μικρογραμμάρια (μ g). Το ED_{50} προσδιορίζεται μέσω μελετών δόσης-απόκρισης, όπου χορηγούνται διαφορετικές δόσεις ή συγκεντρώσεις ενός φαρμάκου ή θεραπείας και παρατηρούνται τα προκύπτοντα αποτελέσματα. Οι τιμές ED_{50} χρησιμοποιούνται για την αξιολόγηση της αποτελεσματικότητας και της ισχύος των φαρμάκων και για τη δημιουργία κατάλληλων δοσολογικών σχημάτων.

EC_{50} (αποτελεσματική συγκέντρωση 50%) : Το EC_{50} είναι παρόμοιο με το ED_{50} , αλλά αναφέρεται συγκεκριμένα στη συγκέντρωση ενός φαρμάκου ή θεραπείας που παράγει ένα συγκεκριμένο θεραπευτικό αποτέλεσμα στο 50% του πληθυσμού ή των πειραματικών ατόμων. Εκφράζεται συνήθως ως γραμμομοριακή συγκέντρωση (π.χ. moles ανά λίτρο, M) ή άλλες μονάδες συγκέντρωσης που σχετίζονται με τη συγκεκριμένη μελέτη. Το EC_{50} προσδιορίζεται συχνά μέσω πειραμάτων *in vitro* ή μελετών κυτταροκαλλιέργειας, όπου διαφορετικές συγκεντρώσεις μιας ουσίας εφαρμόζονται σε κύτταρα ή δείγματα ιστών και μετρώνται τα προκύπτοντα αποτελέσματα. Οι τιμές EC_{50} χρησιμοποιούνται για την αξιολόγηση της ισχύος

των ουσιών in vitro και για την καθοδήγηση της επιλογής των κατάλληλων συγκεντρώσεων για περαιτέρω μελέτες ή κλινικές εφαρμογές.

TD_{50} (τοξική δόση 50%) : Το TD_{50} είναι ένα μέτρο τοξικότητας και αντιπροσωπεύει τη δόση μιας ουσίας που προκαλεί τοξική επίδραση στο 50% του πληθυσμού ή των πειραματιζόμενων ατόμων. Είναι παρόμοιο με το LD_{50} αλλά εστιάζει σε μη θανατηφόρες τοξικές επιδράσεις. Οι τιμές TD_{50} προσδιορίζονται μέσω τοξικολογικών μελετών, όπου χορηγούνται διαφορετικές δόσεις ή συγκεντρώσεις μιας ουσίας και παρατηρούνται οι προκύπτουσες τοξικές επιδράσεις. Οι τιμές TD_{50} χρησιμοποιούνται για την αξιολόγηση των πιθανών κινδύνων και κινδύνων που σχετίζονται με την έκθεση σε χημικές ουσίες ή ουσίες και για τον καθορισμό ασφαλών ορίων έκθεσης.

IC_{50} (ανασταλτική συγκέντρωση 50%) : Το IC_{50} είναι ένα μέτρο της ανασταλτικής ισχύος και αντιπροσωπεύει τη συγκέντρωση ενός φαρμάκου ή θεραπείας που αναστέλλει μια συγκεκριμένη βιολογική διεργασία ή δραστηριότητα ενζύμου κατά 50%. Εκφράζεται συνήθως ως γραμμομοριακή συγκέντρωση (π.χ. moles ανά λίτρο, M) ή άλλες σχετικές μονάδες συγκέντρωσης. Οι τιμές IC_{50} προσδιορίζονται μέσω πειραμάτων in vitro, όπου διαφορετικές συγκεντρώσεις μιας ουσίας εφαρμόζονται σε ένα σύστημα ανάλυσης και μετράται η προκύπτουσα αναστολή της διαδικασίας στόχου ή της ενζυμικής δραστηριότητας. Οι τιμές IC_{50} χρησιμοποιούνται στη φαρμακολογία και την ανάπτυξη φαρμάκων για την αξιολόγηση της ισχύος και της εκλεκτικότητας των αναστολέων και για την καθοδήγηση της ανάπτυξης θεραπευτικών παρεμβάσεων.

MIC (ελάχιστη ανασταλτική συγκέντρωση) : Το MIC είναι ένα μέτρο της αντιμικροβιακής ισχύος και αντιπροσωπεύει τη χαμηλότερη συγκέντρωση ενός αντιμικροβιακού παράγοντα που εμποδίζει την ορατή ανάπτυξη ενός μικροοργανισμού. Εκφράζεται συνήθως ως γραμμομοριακή συγκέντρωση (π.χ. moles ανά λίτρο, M) ή άλλες μονάδες συγκέντρωσης που σχετίζονται με τη συγκεκριμένη μελέτη. Οι τιμές MIC προσδιορίζονται μέσω μικροβιολογικών δοκιμασιών, όπου εφαρμόζονται διαφορετικές συγκεντρώσεις ενός αντιμικροβιακού παράγοντα σε καλλιέργειες μικροοργανισμών και παρατηρείται η προκύπτουσα αναστολή ανάπτυξης.

Όσον αφορά τα πειράματα in vitro που αναφέρθηκαν παραπάνω, πρόκειται για πειράματα που διεξάγονται εκτός ζωντανού οργανισμού, σε ελεγχόμενες εργαστηριακές συνθήκες. Ο όρος "in vitro" προέρχεται από τα λατινικά, που σημαίνει "σε γυαλί" και αναφέρεται σε πειράματα που πραγματοποιούνται σε δοκιμαστικούς σωλήνες, τρυβλία Petri, πλάκες καλλιέργειας ή άλλες τεχνητές διατάξεις. Στο πλαίσιο της βιολογικής έρευνας, τα πειράματα in vitro συνήθως περιλαμβάνουν τη χρήση απομονωμένων κυττάρων, ιστών ή βιολογικών μορίων για τη μελέτη συγκεκριμένων βιολογικών διεργασιών ή φαινομένων. Αυτά τα πειράματα προσφέρουν πολλά πλεονεκτήματα, συμπεριλαμβανομένων των ελεγχόμενων πειραματικών συνθηκών, των απλοποιημένων συστημάτων και της ικανότητας εστίασης σε συγκεκριμένες πτυχές ενός βιολογικού συστήματος.

1.4 Καμπύλες Δόσης-Απόκρισης

Όπως αναφέραμε και παραπάνω το μοντέλο δόσης-απόκρισης μοντελοποιείται από καμπύλες δόσης-απόκρισης, όπου χρησιμοποιούνται αρκετά στη φαρμακολογία και την ανάπτυξη φαρμάκων. Ειδικότερα, το σχήμα της καμπύλης δόσης-απόκρισης ενός φαρμάκου (ποσοτικοποιημένο με παραμέτρους EC_{50} , n_H και y_{max}) αντανακλά τη βιολογική δραστηριότητα και την ισχύ του φαρμάκου.

Η καμπύλη δόσης-απόκρισης είναι μια γραφική παράσταση συντεταγμένων που συσχετίζει το μέγεθος μιας δόσης (ερέθισμα) με την απόκριση ενός βιολογικού συστήματος. Μπορεί να μελετηθεί ένας αριθμός επιδράσεων (ή τελικών σημείων). Στον άξονα X απεικονίζεται η εφαρμοσμένη δόση και στον άξονα Y η επίδραση. Σε μερικές περιπτώσεις, στον άξονα X απεικονίζεται ο λογάριθμος της δόσης. Η καμπύλη είναι τυπικά σιγμοειδής, με το πιο απότομο τμήμα στη μέση, αν και παρακάτω θα μελετηθούν και άλλα είδη γραφικών απεικονίσεων.

Η στατιστική ανάλυση των καμπυλών δόσης-απόκρισης μπορεί να γίνει με την χρήση μεθόδων παλινδρόμησης, όπως το μοντέλο probit ή το μοντέλο logit, ή με άλλες μεθόδους όπως η μέθοδος Spearman-Kärber. Για τα μοντέλα probit και logit θα αναφερθούμε σε άλλο κεφάλαιο εκτενέστερα, καθώς αποτελούν κύριο κομμάτι αυτής της εργασίας τόσο θεωρητικά όσο και πρακτικά. Στην μέθοδο Spearman-Kärber θα αναφερθούμε περιληπτικά σε αυτή την ενότητα, καθώς δεν έχει γίνει χρήση της στα παραδείγματα που ακολουθούν. Να σημειωθεί ότι μερικές φορές είναι προτιμότερο να χρησιμοποιούνται εμπειρικά μοντέλα που βασίζονται σε μη γραμμική παλινδρόμηση, αντί της χρήσης κάποιου μετασχηματισμού των δεδομένων που γραμμικοποιεί τη σχέση δόσης-απόκρισης.

Η μέθοδος Spearman-Kärber, γνωστή και ως εξίσωση Spearman-Kärber ή μέθοδος μέσης θανατηφόρας δόσης (LD_{50}), είναι μια στατιστική μέθοδος που χρησιμοποιείται για την εκτίμηση της συγκέντρωσης ή της δόσης μιας ουσίας που απαιτείται για να προκαλέσει μια συγκεκριμένη απόκριση σε ένα βιολογικό σύστημα. Χρησιμοποιείται συνήθως στην τοξικολογία και τη φαρμακολογία για τον προσδιορισμό της δραστηριότητας ή της τοξικότητας μιας ουσίας. Η μέθοδος βασίζεται στην υπόθεση ότι η σχέση δόσης-απόκρισης ακολουθεί μια σιγμοειδή καμπύλη, όπου η απόκριση αυξάνεται σταδιακά για χαμηλές τιμές, και για μεγάλες δόσεις συγκλίνει σε μια οριζόντια ασύμπτωτη. Ο στόχος είναι να εκτιμηθεί η συγκέντρωση ή η δόση που παράγει ένα συγκεκριμένο επίπεδο απόκρισης, τυπικά τη διάμεση θανατηφόρα δόση (LD_{50}), η οποία είναι η δόση που προκαλεί θάνατο στο 50% των υποκειμένων της δοκιμής.

Η εξίσωση Spearman-Kärber προέρχεται από την έννοια της ανάλυσης probit και υποθέτει ότι η απόκριση ακολουθεί μια λογαριθμική-κανονική κατανομή.

Η εξίσωση έχει ως εξής:

$$LD_{50} = \frac{1}{2} \sum_i^k (p_{i+1} - p_i)(x_{i+1} - x_i)$$

Όπου x_1, x_2, \dots, x_k είναι οι δόσεις με $p_1 = 0$, $p_k = 1$.

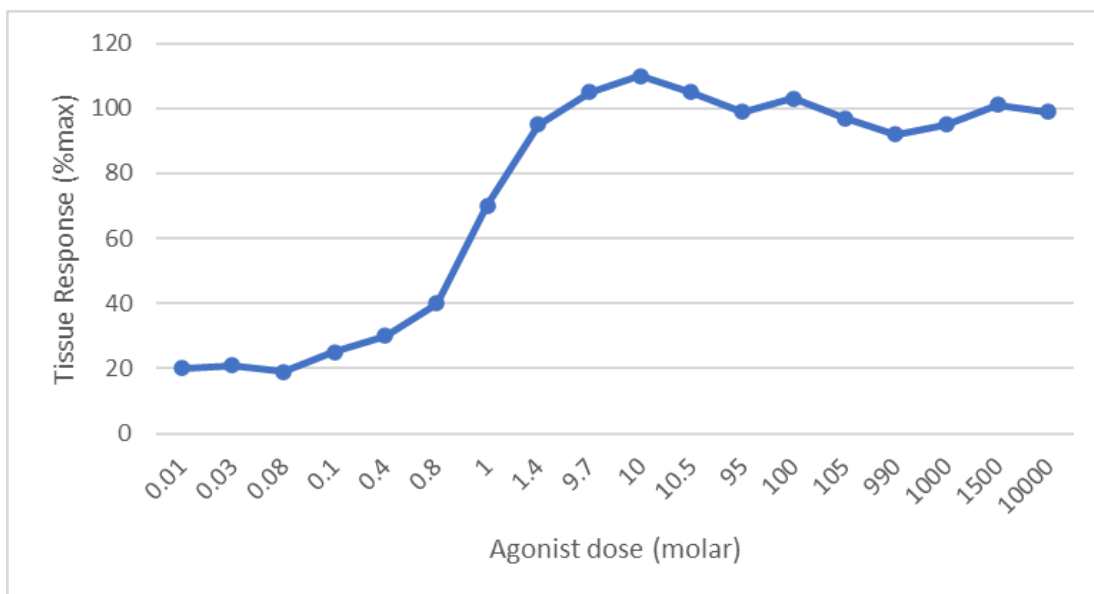
Διασθητικά η αύξηση $p_{i+1} - p_i$ αντιστοιχεί στη δόση $\frac{1}{2}(x_{i+1} - x_i)$.

Για την εκτίμηση του LD_{50} , η μέθοδος περιλαμβάνει τον έλεγχο μιας σειράς δόσεων ή συγκεντρώσεων και τον προσδιορισμό του ρυθμού απόκρισης σε κάθε δόση. Σχεδιάζοντας τη λογαριθμική δόση ή τη λογαριθμική συγκέντρωση έναντι των τιμών probit (το αντίστροφο της τυπικής κανονικής αθροιστικής συνάρτησης κατανομής), μπορεί να προσαρμοστεί μια ευθεία γραμμή μέσω των σημείων δεδομένων. Το LD_{50} μπορεί στη συνέχεια να παρεμβληθεί από αυτή τη γραμμή.

Είναι σημαντικό να σημειωθεί ότι η μέθοδος Spearman-Kärber έχει τους περιορισμούς και τις υποθέσεις της και είναι πιο ακριβής όταν οι δόσεις ή οι συγκεντρώσεις που δοκιμάστηκαν καλύπτουν ένα ευρύ φάσμα και οι αποκρίσεις είναι δυαδικές (π.χ. ζωντανές ή νεκρές). Στην πράξη, υπάρχουν παραλλαγές και βελτιώσεις της μεθόδου για την προσαρμογή διαφορετικών πειραματικών σχεδίων και μεταβλητών απόκρισης.

1.5 Γραφικές Παραστάσεις Διαφόρων Καμπυλών Δόσης-Απόκρισης

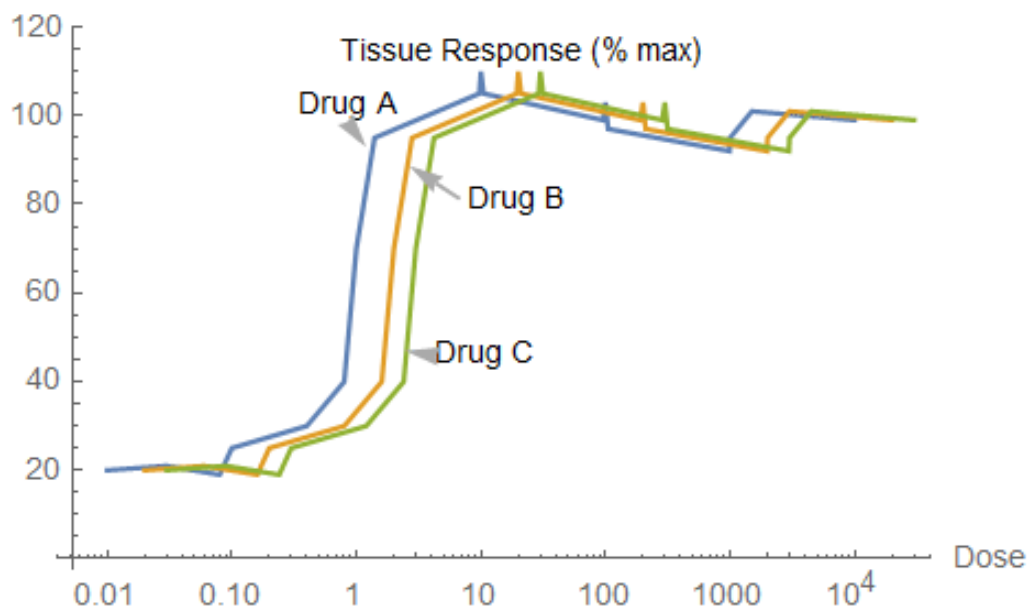
Ενδεικτικά, έχουμε δημιουργήσει το παρακάτω σχήμα μέσω του excel, προκειμένου να γίνει πιο κατανοητή η μορφή της καμπύλης δόσης-απόκρισης. Τα δεδομένα τα οποία χρησιμοποιήθηκαν για το παρακάτω σχήμα, προέρχονται από τον James Goodman (2019) ο οποίος τα έχει δημοσιεύσει στην βικιπαίδεια, στο κομμάτι της σχέσης δόσης-απόκρισης. Απο ότι βλέπουμε οι χαμηλές δόσεις είναι ανεπαρκείς για τη δημιουργία υψηλής απόκρισης, ενώ οι υψηλές δόσεις δημιουργούν τη μέγιστη απόκριση. Η μέγιστη απόκριση η οποία είναι ίση με 110 επιτυγχάνεται για δόση ίση με 10 molar. Αντίστοιχα το EC_{50} βρίσκεται εκεί όπου η δόση είναι ίση με 0,8 molar.



Σχήμα 1.1 Παράδειγμα Μονότονης Καμπύλης

Αφού είδαμε την απλοϊκή μορφή της καμπύλης δόσης απόκρισης, είναι εξίσου ενδιαφέρον να μελετήσουμε την απόκριση στην περίπτωση που υπάρχουν τρία διαφορετικά φάρμακα και όχι ένα. Πάλι με την χρήση του excel και χρησιμοποιώντας δεδομένα που έχουν δημοσιευθεί στην βικιπαίδεια στο κομμάτι της σχέσης δόσης-απόκρισης, δημιουργούμε το παρακάτω γράφημα, στο οποίο υποθέτουμε ότι έχουμε τα φάρμακα A, B και C.

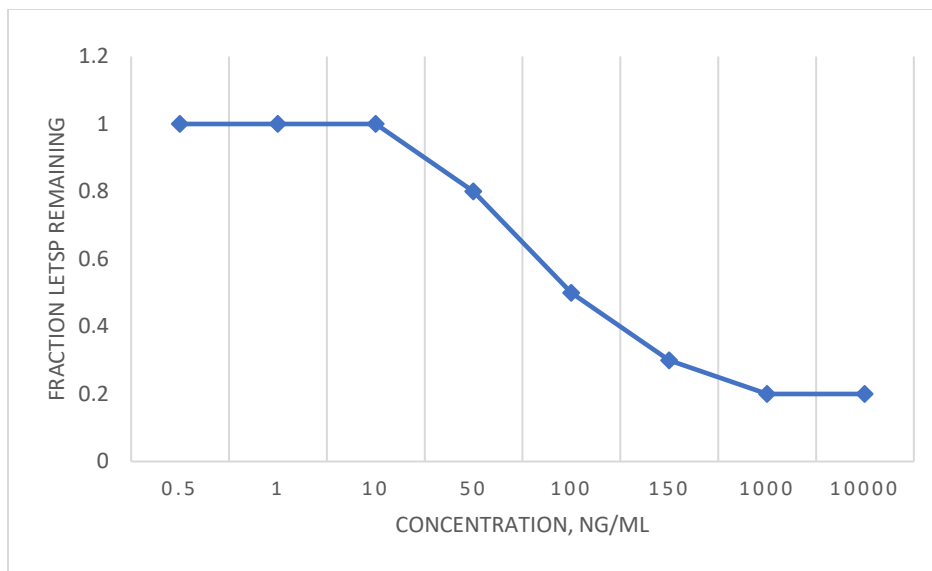
Στο παρακάτω σχήμα βλέπουμε ποιά είναι η απαιτούμενη δοσολογία για το κάθε φάρμακο προκειμένου να επιτευχθεί η συγκεκριμένη απόκριση. Παρατηρούμε ότι για να επιτευχθεί η συγκεκριμένη απόκριση με το φάρμακο A, χρειάζεται μικρότερη δοσολογία συγκριτικά με τα φάρμακα B και C. Αντίθετα, με την χρήση του φαρμάκου C χρειάζεται μεγαλύτερη δοσολογία συγκριτικά με φάρμακο B και ακόμη μεγαλύτερη συγκριτικά με το φάρμακο A. Η δοσολογία του φαρμάκου B βρίσκεται κάπου στην μέση συγκριτικά με τα φάρμακα A και C. Συνεπώς, αντιλαμβανόμαστε ότι υπάρχουν πολλά φάρμακα που μπορούν να επιτύχουν ένα συγκεκριμένο αποτέλεσμα, αλλά κάθε ένα απο αυτά έχει την δική του ιδανική δοσολογία.



Σχήμα 1.2 Παράδειγμα Πολλαπλών Μονότονων Καμπύλων

Επίσης, είναι σημαντικό να σημειωθεί ότι εκτός από την παραπάνω σχέση δόσης-απόκρισης, κατά την οποία όσο αυξάνεται η δόση, τόσο αυξάνεται και η απόκριση μέχρι την μέγιστη επίδραση, υπάρχει και η σχέση δόσης-απόκρισης όπου όσο αυξάνεται η δόση, τόσο μειώνεται η απόκριση.

Το παρακάτω γράφημα το οποίο απεικονίζει τη μείωση των επιπέδων φιμπρονεκτίνης μετά τη χορήγηση μεζερεΐνης σε κύτταρα CEF, μας δίνει μια ιδέα για το πως είναι η καμπύλη δόσης-απόκρισης με καθοδική κλίση. Τα δεδομένα τα οποία χρησιμοποιήθηκαν για αυτό το γράφημα, αντλήθηκαν από την βικιπαίδεια στο κομμάτι των καμπυλών δόσης-απόκρισης. Προφανώς, η συγκεκριμένη καμπύλη χρησιμοποιείται πολύ πιο σπάνια από την προηγούμενη με την ανοδική τάση, αλλά είναι καλό να γνωρίζουμε την ύπαρξη της.



Σχήμα 1.3 Παράδειγμα Μονότονης Καμπύλης με Καθοδική Κλίση

1.6 Εξίσωση Hill

Οι λογαριθμικές καμπύλες δόσης-απόκρισης είναι γενικά σιγμοειδείς και μονότονες και μπορούν να προσαρμοστούν σε μια κλασική εξίσωση Hill. Η εξίσωση Hill είναι μια λογιστική συνάρτηση σε σχέση με τον λογάριθμο της δόσης και είναι παρόμοια με ένα μοντέλο logit.

Η εξίσωση Hill δίνεται από τον παρακάτω τύπο, όπου το E είναι το μέγεθος της απόκρισης, το A είναι η συγκέντρωση του φαρμάκου, το EC_{50} είναι η συγκέντρωση φαρμάκου που παράγει η 50% μέγιστη απόκριση και το n είναι ο συντελεστής Hill.

$$\frac{E}{E_{max}} = \frac{1}{1 + \left(\frac{EC_{50}}{|A|}\right)^n}$$

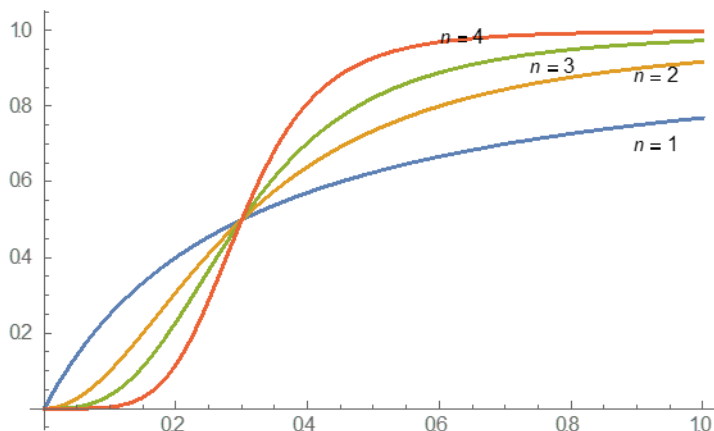
Η εξίσωση αυτή προκύπτει ως εξής :

Αν $p = \frac{E}{E_{max}}$ υποθέτουμε ότι ισχύει η παρακάτω σχέση,

$$\frac{p}{1-p} = \frac{|A|^n}{|A|^n + (EC_{50})^n}$$

Λύνοντας ως προς p βρίσκουμε την εξίσωση Hill.

Ο συντελεστής n λέγεται συντελεστής υπερεισθησίας γιατί καθορίζει την κλίση της σιγμοειδούς καμπύλης.



Σχήμα 1.4 Καμπύλη Δόση-Απόκριση σε Σχέση με τον Συντελεστή Υπερευαισθησίας

Μια καμπύλη που χρησιμοποιείται αρκετά συχνά είναι η καμπύλη EC_{50} , η μισή μέγιστη αποτελεσματική συγκέντρωση, όπου το σημείο EC_{50} είναι το σημείο καμψής της καμπύλης. Οι καμπύλες δόσης-απόκρισης προσαρμόζονται συνήθως στην εξίσωση Hill.

Το πρώτο σημείο κατά μήκος του γραφήματος όπου επιτυγχάνεται μια απόκριση πάνω από το μηδέν, αναφέρεται συνήθως ως κατώτατο όριο δόσης. Για τα περισσότερα ωφέλιμα φάρμακα, τα επιθυμητά αποτελέσματα βρίσκονται σε δόσεις που είναι ελαφρώς μεγαλύτερες από την οριακή δόση. Σε υψηλότερες δόσεις, εμφανίζονται ανεπιθύμητες παρενέργειες και ενισχύονται όσο αυξάνεται η δόση. Όσο πιο ισχυρή είναι μια ουσία, τόσο πιο απότομη θα είναι η καμπύλη. Σε ποσοτικές καταστάσεις, ο άξονας Y συχνά ορίζεται με ποσοστά, τα οποία αναφέρονται στο ποσοστό των εκτεθειμένων ατόμων που καταγράφουν μια τυπική απόκριση (η οποία μπορεί να είναι θάνατος, όπως στην LD_{50}). Πρόκειται λοιπόν για μια καμπύλη ποσοτικής δόσης-απόκρισης, η οποία διαφοροποιείται από μια διαβαθμισμένη καμπύλη δόσης-απόκρισης, όπου έχει συνεχής απόκριση.

1.7 Διάφοροι Τύποι Μονότονων Σχέσεων Δόσης-Απόκρισης

Εκτός απο το λογιστικό μοντέλο δόσης-απόκρισης που είδαμε παραπάνω το οποίο είναι και το πιο συνηθέστερο, υπάρχουν πολλά ακόμα μοντέλα που χρησιμοποιούνται συχνά σε μελέτες δόσης-απόκρισης. Παρακάτω παρουσιάζονται οι συναρτήσεις των διαφορετικών μονότονων μοντέλων δόσης-απόκρισης και αναπαριστώνται γραφικά. Τα δεδομένα που έχουν χρησιμοποιηθεί για τα γραφήματα που βρίσκονται σε αυτό το κεφάλαιο προέρχονται από τον Setia Pramana (2011) και αντλήθηκαν από πειράματα γονιδιακής έκφρασης. Συνεπώς, με την χρήση του excel, έγιναν οι παρακάτω γραφικές απεικονίσεις.

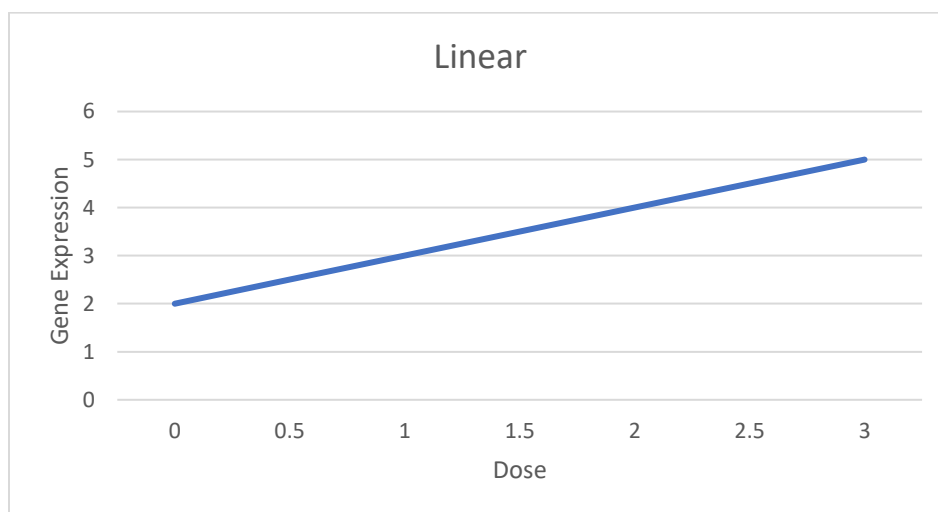
Γραμμικό Μοντέλο

Ένα απο τα πιο απλά μοντέλα που εξετάζονται σε μελέτες δόσης-απόκρισης είναι το γραμμικό μοντέλο. Στο συγκεκριμένο μοντέλο η απόκριση αυξάνεται ή μειώνεται γραμμικά με τη δόση ή το επίπεδο έκθεσης. Αυτή η γραμμική σχέση συνεπάγεται σταθερό ρυθμό μεταβολής στην απόκριση για κάθε μονάδα αύξησης της δόσης. Οι γραμμικές καμπύλες δόσης-απόκρισης παρατηρούνται τυπικά όταν η απόκριση του βιολογικού συστήματος είναι ευθέως ανάλογη με τη δόση. Ο τύπος του γραμμικού μοντέλου δίνεται παρακάτω :

$$f(x) = E_0 + \delta x$$

Στο συγκεκριμένο μοντέλο, η παράμετρος δ αντιπροσωπεύει ένα γραμμικό αποτέλεσμα της δόσης και η παράμετρος E_0 είναι το επίπεδο απόκρισης όταν η δόση του φαρμάκου είναι μηδενική.

Παρακάτω δίνεται το γράφημα του γραμμικού μοντέλου δόσης-απόκρισης :



Σχήμα 1.5 Γραμμικό Μοντέλο

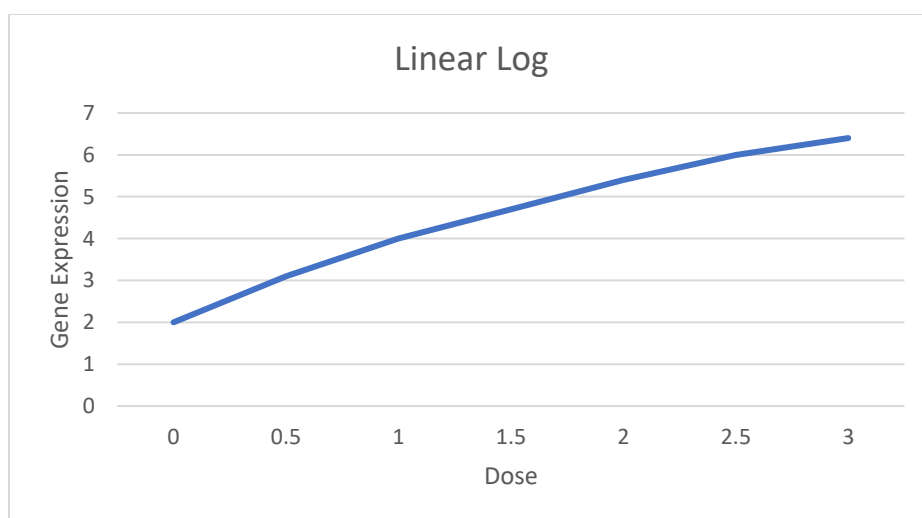
Γραμμικό Μοντέλο Λογαριθμικής Δόσης

Ακόμη ένα απλό μοντέλο δόσης-απόκρισης, το οποίο αποτελεί επέκταση του γραμμικού μοντέλου που είδαμε προηγουμένως είναι το γραμμικό μοντέλο λογαριθμικής δόσης. Το μοντέλο αυτό αντιπροσωπεύει μια γραμμική σχέση μεταξύ της απόκρισης και του λογάριθμου της δόσης. Ο τύπος του οποίου δίνεται παρακάτω :

$$f(x) = E_0 + \beta \log(x + c)$$

Όπου, η παράμετρος β είναι γραμμική επίδραση του $\log(\text{δόση}+c)$. Επίσης, η τιμή $c > 0$ προστίθεται, προκειμένου να αποφευχθεί το πρόβλημα σε περίπτωση που έχουμε δόση=0.

Παρακάτω δίνεται το γράφημα του γραμμικού μοντέλου λογαριθμικής δόσης :



Σχήμα 1.6 Γραμμικό Μοντέλο Λογαριθμικής Δόσης

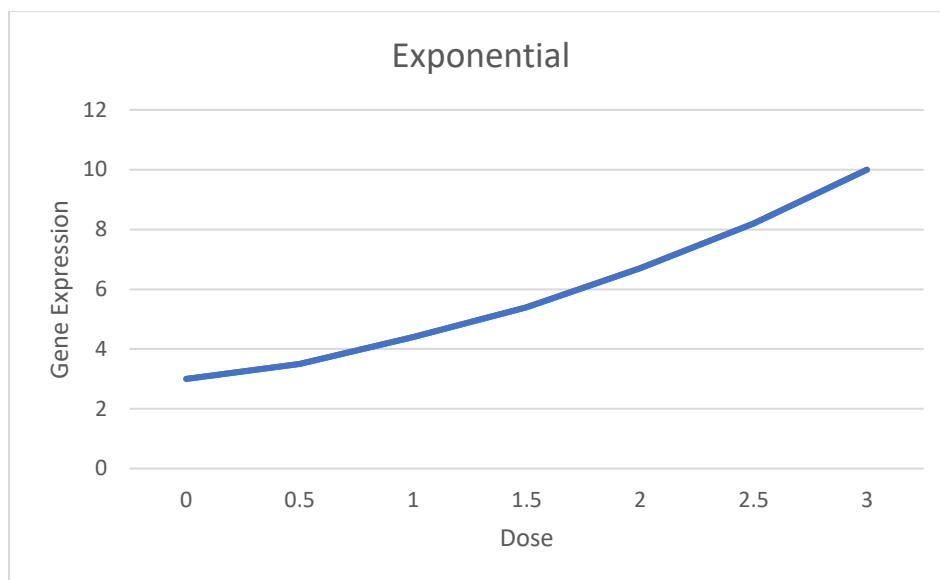
Εκθετικό Μοντέλο

Ένα εξίσου αξιοσημείωτο μοντέλο δόσης-απόκρισης είναι το εκθετικό. Σύμφωνα με τους Bretz et al. (2008) και Bornkamp et al. (2009), το συγκεκριμένο μοντέλο μπορεί να πάρει την μορφή μιας υπογραμμικής ή μιας κυρτής σχέσης δόσης-απόκρισης. Ουσιαστικά στο συγκεκριμένο μοντέλο οι αλλαγές στην δόση μιας ουσίας προκαλούν την μεταβολή της απόκρισης με εκθετικό ρυθμό. Ο τύπος του εκθετικού μοντέλου δίνεται παρακάτω :

$$f(x) = E_0 + E_1(e^{\frac{x}{\theta}} - 1)$$

Όπου, η παράμετρος E_1 συμβολίζει τον αριθμό αύξησης ή μείωσης της επίδρασης της δόσης και η παράμετρος θ καθορίζει την κυρτότητα του μοντέλου.

Παρακάτω δίνεται το γράφημα του εκθετικού μοντέλου :



Σχήμα 1.7 Εκθετικό Μοντέλο

E_{max} Μοντέλο

Το E_{max} μοντέλο, αποτελεί ένα ευρέως χρησιμοποιούμενο εργαλείο σε μελέτες δόσης-απόκρισης και ανήκει στην κατηγορία των μη γραμμικών μοντέλων. Η καμπύλη δόσης-απόκρισης E_{max} , υποθέτει ότι η σχέση δόσης-απόκρισης είναι μονότονη. Όσο αυξάνεται η δόση, η καμπύλη μπορεί να πάρει αύξουσα ή φθίνουσα μορφή. Εάν η απόκριση μειώνεται, η τιμή της παραμέτρου E_{max} θα είναι αρνητική και ο ορισμός του σημείου καμπής θα ορίζεται πλέον ως Ανασταλτική Δόση/Συγκέντρωση (ID_{50} ή IC_{50}) αντί για ED_{50} .

Σύμφωνα με τον Setia Pramana (2011), το μοντέλο E_{max} αποτελείται από τέσσερις παραμέτρους:

(α) τη μέγιστη επίδραση μιας δόσης στην απόκριση, για παράδειγμα την έκφραση γονιδίου (E_{max})

(β) τη βασική επίδραση που αντιστοιχεί στο επίπεδο έκφρασης, όταν η δόση του φαρμάκου είναι μηδενική (E_0)

(γ) το σημείο καμπής της συνάρτησης ή την δόση που παράγει το ήμισυ του E_{max} (ED_{50})

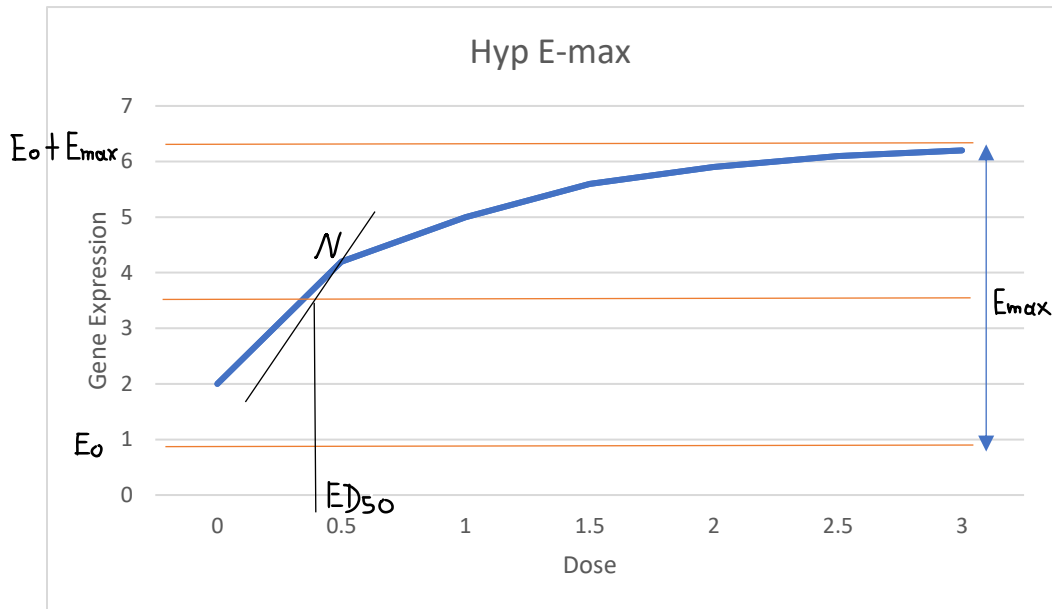
(δ) τον παράγοντα κλίσης ή παράγοντα Hill (N), που μετρά την ευαισθησία της απόκρισης στο εύρος δόσης του φαρμάκου, προσδιορίζοντας την κλίση της καμπύλης δόσης-απόκρισης

Ο MacDougall (2006) μελέτησε το συγκεκριμένο μοντέλο εκτενώς και αποφάσισε να το διασπάσει σε δύο μοντέλα, το ένα είναι το υπερβολικό μοντέλο E_{max} και το άλλο είναι το σιγμοειδές μοντέλο E_{max} , τα οποία παρουσιάζονται παρακάτω.

Υπερβολικό Μοντέλο E_{max}

Το υπερβολικό μοντέλο E_{max} , είναι το μοντέλο κατα το οποίο ο συντελεστής κλίσης (N) δεν περιλαμβάνεται στο μοντέλο E_{max} . Ο τύπος και το γράφημα του συγκεκριμένου μοντέλου δίνονται παρακάτω :

$$f(x) = E_0 + \frac{x \times (E_{max})}{x + ED_{50}}$$

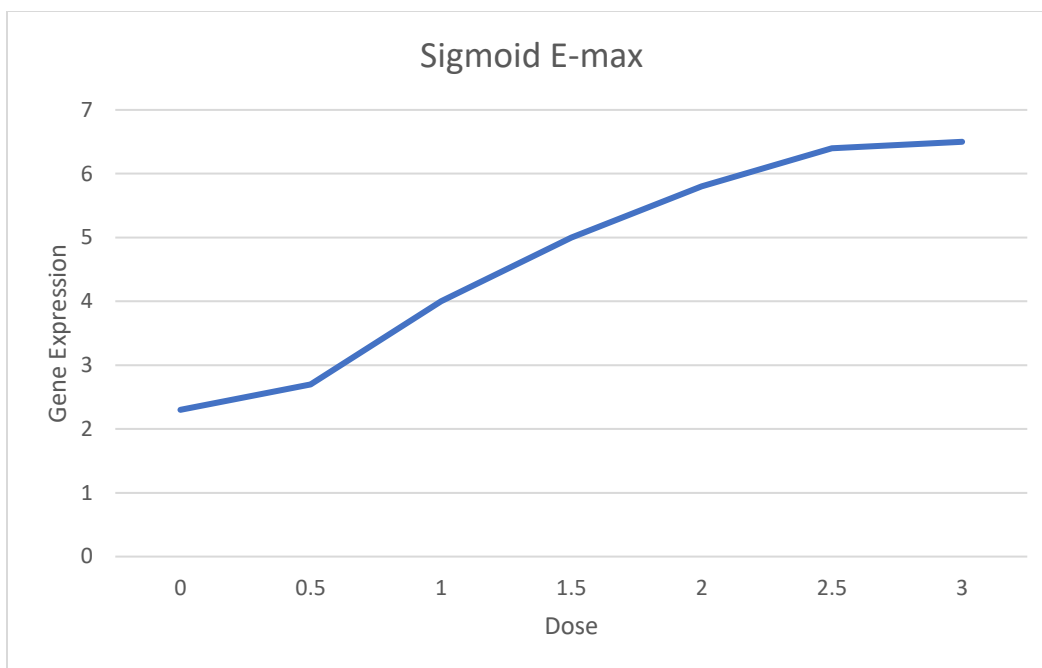


Σχήμα 1.8 Υπερβολικό Μοντέλο E_{max}

Σιγμοειδές Μοντέλο E_{max}

Το σιγμοειδές μοντέλο E_{max} , είναι το μοντέλο κατα το οποίο ο συντελεστής κλίσης (N) περιλαμβάνεται στο μοντέλο E_{max} . Ο τύπος και το γράφημα του συγκεκριμένου μοντέλου δίνονται παρακάτω :

$$f(x) = E_0 + \frac{x^N \times (E_{max})}{x^N + ED_{50}^N}$$



Σχήμα 1.8 Σιγμοειδές Μοντέλο E_{max}

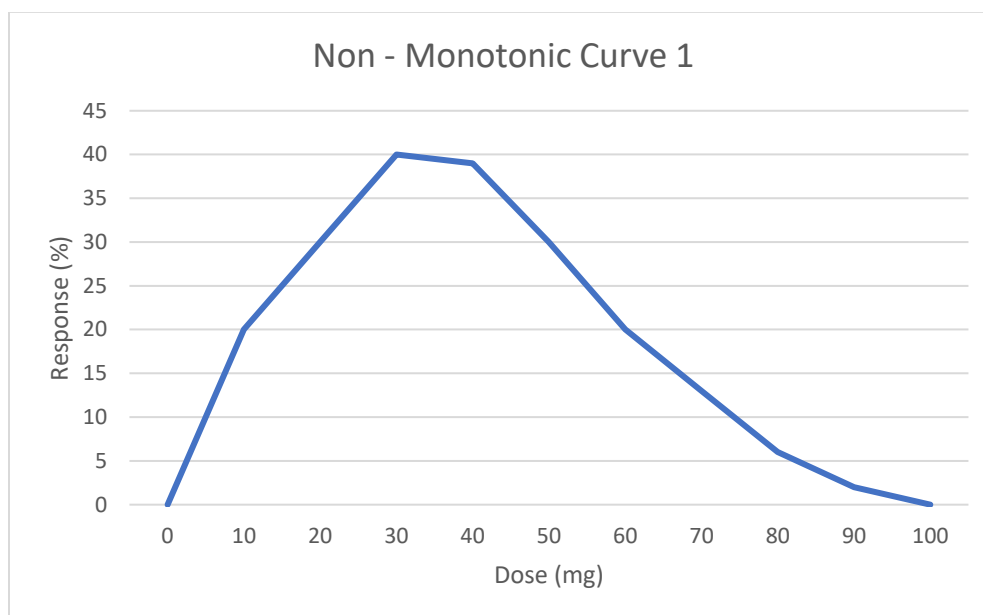
1.9 Μη Μονότονες Σχέσεις Δόσης-Απόκρισης

Παραπάνω αναφέραμε ότι στις περισσότερες περιπτώσεις οι καμπύλες δόσης-απόκρισης είναι μονότονες, παρ'όλα αυτά υπάρχουν και οι μη μονότονες οι οποίες δεν ακολουθούν τα τυπικά πρότυπα. Όταν λέμε τυπικά πρότυπα εννοούμε ότι η αύξηση της δόσης συνεπάγεται και αύξηση της απόκρισης. Στις μη μονότονες περιπτώσεις, η καμπύλη αλλάζει κατεύθυνση εκτός του εύρους των δόσεων που εξετάζονται, δημιουργώντας ασυνήθιστες καμπύλες όπου είναι δύσκολο να καθοριστούν από κάποιο συγκεκριμένο μαθηματικό μοντέλο, όπως κάνουν οι μονότονες (γραμμικό, σιγμοειδικό, εκθετικό κ.λ.π.). Το σχήμα μιας μη μονότονης καμπύλης δόσης-απόκρισης επηρεάζεται από διάφορους παράγοντες, συμπεριλαμβανομένης της φύσης του βιολογικού συστήματος, της συγκεκριμένης ουσίας ή παράγοντα που ελέγχεται και των υποκείμενων μηχανισμών που παίζουν.

Λόγω της πολυπλοκότητας και της ποικιλομορφίας των μη μονότονων σχέσεων δόσης-απόκρισης, συχνά αναλύονται και ερμηνεύονται ποιοτικά αντί να προσπαθούν να προσαρμοστούν σε ένα συγκεκριμένο μαθηματικό μοντέλο. Οι στατιστικές μέθοδοι μπορούν να εξακολουθούν να χρησιμοποιούνται για την περιγραφή και τη σύγκριση αποκρίσεων σε διαφορετικές δόσεις, αλλά η εστίαση είναι συνήθως στην κατανόηση των υποκείμενων μηχανισμών και στον προσδιορισμό του κρίσιμου εύρους δόσης που σχετίζεται με το παρατηρούμενο πρότυπο απόκρισης.

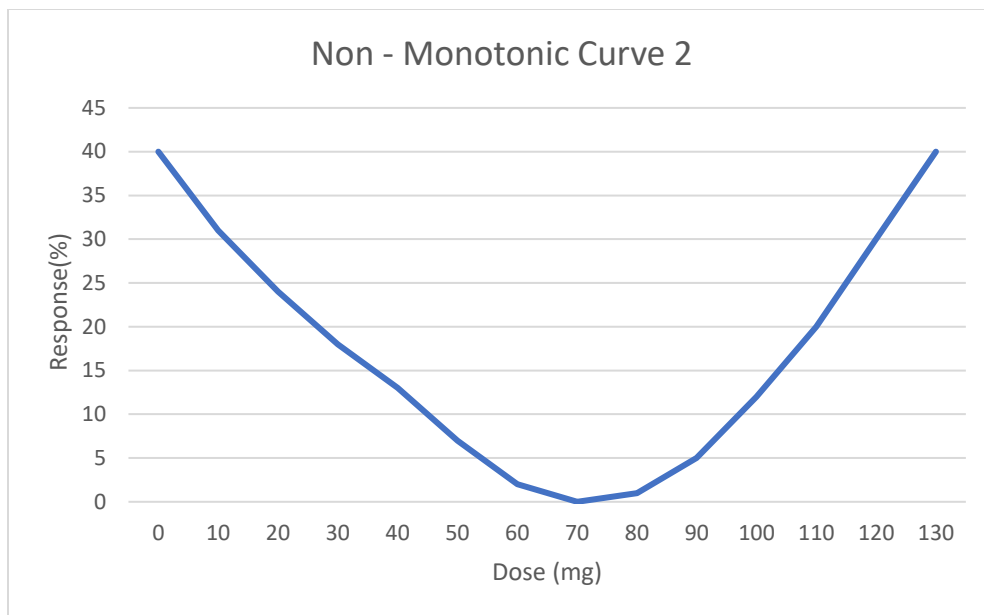
Είναι σημαντικό να σημειωθεί ότι οι μη μονότονες σχέσεις δόσης-απόκρισης έχουν κερδίσει την προσοχή τα τελευταία χρόνια και υπάρχει συνεχής έρευνα για την καλύτερη κατανόηση και τον χαρακτηρισμό της συμπεριφοράς τους. Διάφορες αναλυτικές προσεγγίσεις, όπως μοντέλα μη γραμμικής παλινδρόμησης ή στατιστικές Bayes, μπορούν να χρησιμοποιηθούν για την ανάλυση και την ερμηνεία μη μονοτονικών δεδομένων, αλλά η συγκεκριμένη προσέγγιση θα εξαρτηθεί από τη φύση των δεδομένων και το ερευνητικό ερώτημα.

Παρακάτω δίνονται μερικά ενδεικτικά γραφήματα που αναπαριστούν τέτοιου είδους περιπτώσεις.



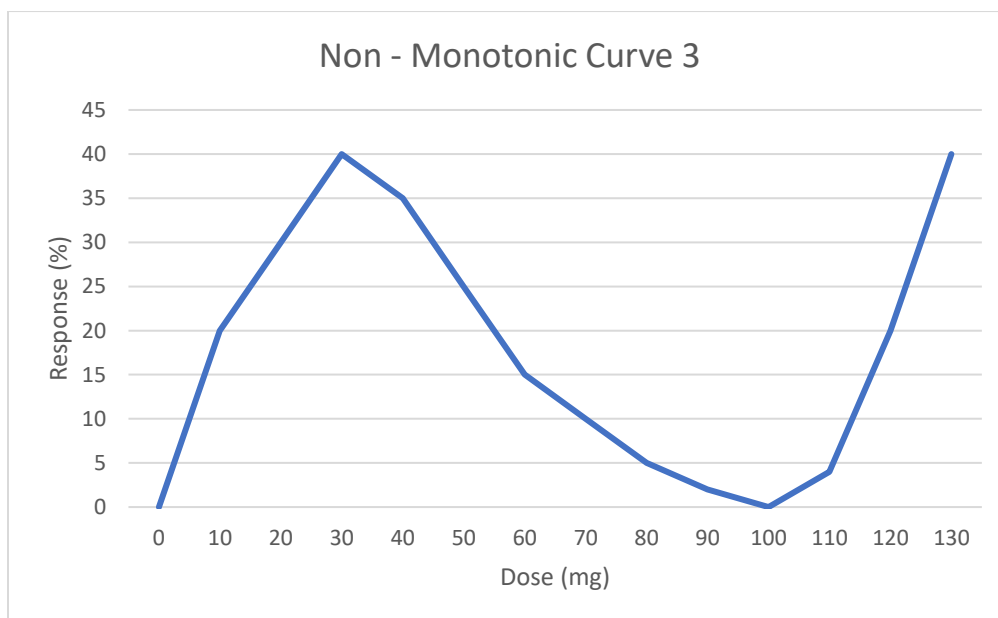
Σχήμα 1.10 Ενδεικτικό Σχήμα Μη Μονότονης Καμπύλης – Περίπτωση 1

Στο παραπάνω γράφημα, η απόκριση αρχικά αυξάνεται καθώς η δόση αυξάνεται από 0 σε 30 mg. Μετά την μέγιστη απόκριση όμως, όπου είναι ίση με 40 και επιτυγχάνεται με την δόση 30 mg, το γράφημα αρχίζει και παίρνει μια καθοδική μορφή. Πιο συγκεκριμένα, για δόση από 40 έως και 100, η απόκριση αρχίζει να μειώνεται μέχρι την στιγμή που γίνεται ίση με 0 στο επίπεδο δόσης 100 mg.



Σχήμα 1.11 Ενδεικτικό Σχήμα Μη Μονότονης Καμπύλης – Περίπτωση 2

Το παραπάνω γράφημα έχει ακριβώς την αντίθετη μορφή από το γράφημα που αναλύσαμε πιο πάνω. Πιο συγκεκριμένα, για δόση από 0 έως και 70 η απόκριση μειώνεται και βλέπουμε ότι για δόση ίση με 70 η απόκριση φτάνει στο κατώτατο της σημείο, δηλαδή το 0. Μετά το κατώτο σημείο, η απόκριση αρχίζει και αυξάνεται για δόση από 70 έως και 130.



Σχήμα 1.12 Ενδεικτικό Σχήμα Μη Μονότονης Καμπύλης – Περίπτωση 3

Το παραπάνω γράφημα αποτελεί έναν συνδυασμό των δύο προηγούμενων. Δηλαδή, για δόση από 0 έως και 30 η απόκριση αυξάνεται, για δόση από 30 έως και 100 η απόκριση μειώνεται και για δόση από 100 έως και 130 η απόκριση αυξάνεται πάλι. Άρα βλέπουμε ότι συνεχώς αλλάζει η μορφή του παραπάνω γραφήματος και στην συγκεκριμένη περίπτωση έχουμε δύο κατώτατα σημεία (δόση=0 και δόση=100), δηλαδή εκεί που η απόκριση γίνεται 0 και δύο ανώτατα σημεία (δόση=30 και δόση=130), δηλαδή εκεί που η απόκριση γίνεται ίση με 40.

Συμπερασματικά, για τέτοιου είδους περιπτώσεις μπορούμε να ισχυριστούμε ότι γενικά ο τύπος μη μονότονης καμπύλης υποδηλώνει ότι η ουσία ή ο παράγοντας που ελέγχεται έχει μια πολύπλοκη αλληλεπίδραση με το βιολογικό σύστημα, όπου η απόκριση επηρεάζεται από πολλούς παράγοντες. Το ακριβές σχήμα και τα χαρακτηριστικά της καμπύλης μπορεί να διαφέρουν ανάλογα με τη συγκεκριμένη ουσία και το βιολογικό σύστημα που μελετάται.

ΚΕΦΑΛΑΙΟ 2^ο : Μαθηματική Ανάλυση Μοντέλων Δόσης-Απόκρισης

2.1 Μοντέλα Δόσης-Απόκρισης

Σύμφωνα με τον Finney(1973), μια απο τις πρώτες φορές που χρησιμοποιήθηκαν μοντέλα παλινδρόμησης για διωνυμικά δεδομένα, ήταν για σκοπούς βιοδοκιμασίας, δηλαδή για την μέτρηση συγκέντρωσης μιας ουσίας από την επίδρασή της σε ζωντανά κύτταρα ή ιστούς. Οι αποκρίσεις εκφράστηκαν ως αναλογίες ή ποσοστά επιτυχιών, όπως το ποσοστό των πειραματόζωων που θανατώθηκαν από διάφορα επίπεδα δόσης τοξικών χημικών. Οι ‘ποσοτικές αποκρίσεις’ είναι ένα άλλο όνομα για αυτούς τους τύπους δεδομένων. Ο στόχος είναι να χαρακτηριστεί η πιθανότητα επιτυχίας, σε συνάρτηση με τη δόση, x . Για παράδειγμα, $g(\pi) = \beta_1 + \beta_2 x$.

Εάν η κατανομή tolerance $f(s)$ είναι η ομοιόμορφη κατανομή στο διάστημα $[c_1, c_2]$

$$f(s) = \begin{cases} \frac{1}{c_2 - c_1} & \text{if } c_1 \leq s \leq c_2 \\ 0 & \text{otherwise} \end{cases}$$

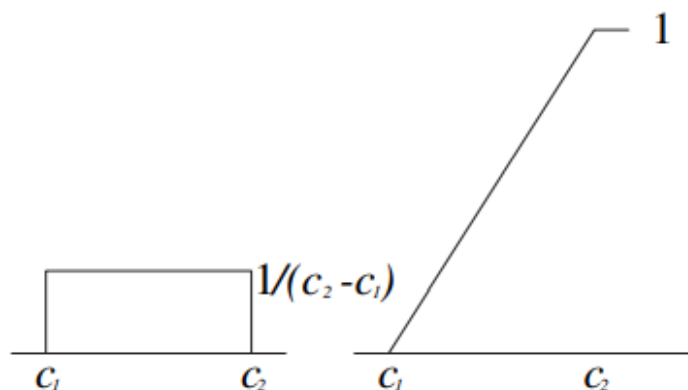
τότε για $c_1 \leq x \leq c_2$, το π είναι αθροιστικό

$$\pi = \int_{c_1}^x f(s) ds = \frac{x - c_1}{c_2 - c_1}$$

Η εξίσωση αυτή έχει τη μορφή $\pi = \beta_1 + \beta_2 x$, όπου

$$\beta_1 = \frac{-c_1}{c_2 - c_1} \quad \text{και} \quad \beta_2 = \frac{1}{c_2 - c_1}$$

Σχηματικά η ομοιόμορφη κατανομή $f(s)$ και π είναι ως εξής :



Σχήμα 2.1 Ομοιόμορφη κατανομή: $f(s)$ και π

Το παραπάνω γραμμικό μοντέλο είναι ισοδύναμο με την χρήση της ταυτοτικής συνάρτησης σαν συνάρτηση σύνδεσης g και έχοντας ως όρο ότι τα x , β_1 και β_2 είναι σύμφωνα με την προϋπόθεση $c_1 \leq x \leq c_2$. Οι τυπικές τεχνικές που χρησιμοποιούνται συνήθως για την εκτίμηση του β_1 και β_2 στα γενικευμένα γραμμικά μοντέλα, δεν μπορούν να χρησιμοποιηθούν άμεσα λόγω αυτών των πρόσθετων προϋποθέσεων. Στην πραγματικότητα, το συγκεκριμένο μοντέλο δεν χρησιμοποιείται ευρέως.

Ένα από τα πρωτότυπα μοντέλα που χρησιμοποιούνται όταν έχουμε δεδομένα βιοδραστικότητας και τοξικότητας, ονομάζεται μοντέλο probit. Το μοντέλο probit είναι ένας τύπος παλινδρόμησης όπου η εξαρτημένη μεταβλητή μπορεί να λάβει μόνο δύο τιμές, για παράδειγμα καπνίζοντες και μη καπνίζοντες. Επίσης, είναι σημαντικό να σημειωθεί ότι το συγκεκριμένο μοντέλο χρησιμοποιείται σε διάφορους τομείς των βιολογικών και κοινωνικών επιστημών στις οποίες υπάρχουν φυσικές ερμηνείες του μοντέλου. Σύμφωνα με τους A. J. Dobson και A. G. Barnett (2008), το $x = \mu$ ονομάζεται διάμεση θανατηφόρα δόση (LD_{50}) επειδή αντιστοιχεί στη δόση που αναμένεται να σκοτώσει τα μισά από τα ζώα.

Ως κατανομή tolerance, χρησιμοποιείται η κανονική κατανομή.

$$\begin{aligned} \pi &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2\right] ds \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

Όπου το Φ , υποδηλώνει την συνάρτηση αθροιστικής πιθανότητας για την τυπική κανονική κατανομή $N(0,1)$. Άρα έχουμε την παρακάτω σχέση, όπου

$$\beta_1 = -\frac{\mu}{\sigma}$$

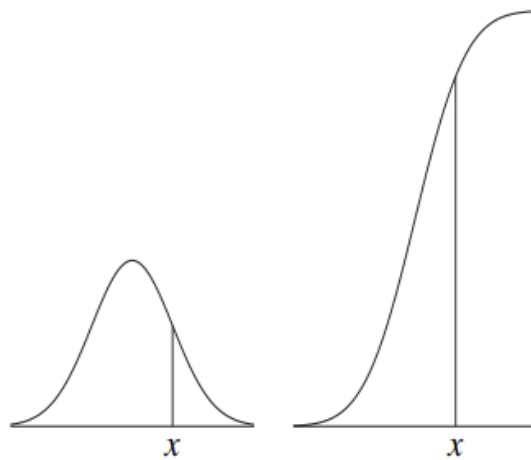
και

$$\beta_2 = \frac{1}{\sigma}$$

και η συνάρτηση σύνδεσης g είναι η Κανονική αντίστροφη αθροιστική συνάρτηση πιθανότητας Φ^{-1} .

$$\Phi^{-1}(\pi) = \beta_1 + \beta_2 x$$

Παρακάτω, απεικονίζεται η κανονική κατανομή $f(s)$ και π :



Σχήμα 2.2 Κανονική κατανομή: $f(s)$ και π

Σε πολλές περιπτώσεις, αντί για το μοντέλο probit χρησιμοποιείται το μοντέλο logit, το οποίο δίνει αρκετά παρόμοια αποτελέσματα με το probit και υπολογιστικά είναι πιο εύκολο.

Διαφορές μεταξύ logit και probit μοντέλου :

1. Το μοντέλο logit χρησιμοποιείται για τη μοντελοποίηση των πιθανοτήτων επιτυχίας ενός γεγονότος ως συνάρτηση ανεξάρτητων μεταβλητών, ενώ το μοντέλο probit χρησιμοποιείται για τον προσδιορισμό της πιθανότητας ότι ένα γεγονός θα εμπίπτει σε ένα εύρος κατηγοριών, εκτιμώντας την πιθανότητα ότι η παρατήρηση με συγκεκριμένα χαρακτηριστικά θα ανήκει σε μια συγκεκριμένη κατηγορία.
2. Στην περίπτωση του μοντέλου logit, χρησιμοποιούμε μια λογιστική ή σιγμοειδή συνάρτηση αντί της Φ που είναι μια αθροιστική τυπική συνάρτηση κανονικής κατανομής.
3. Το μοντέλο logit χρησιμοποιείται για τη μοντελοποίηση της κατανομής Logistic, ενώ το μοντέλο probit για τη μοντελοποίηση της αθροιστικής τυπικής κανονικής κατανομής.

Οπότε, για το μοντέλο logit η κατανομή tolerance δίνεται απο τον εξής τύπο :

$$f(s) = \frac{\beta_2 \exp(\beta_1 + \beta_2 s)}{[1 + \exp(\beta_1 + \beta_2 s)]^2}$$

για το π έχουμε :

$$\pi = \int_{-\infty}^x f(s) ds = \frac{\exp(\beta_1 + \beta_2 x)}{1 + \exp(\beta_1 + \beta_2 x)}$$

Και καταλήγουμε στην εξής συνάρτηση σύνδεσης :

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_1 + \beta_2 x$$

Ο όρος $\log\left(\frac{\pi}{1-\pi}\right)$, συνήθως ονομάζεται συνάρτηση logit και ερμηνεύεται ως ο λογάριθμος του λόγου των πιθανοτήτων π , $1-\pi$. Το λογιστικό μοντέλο χρησιμοποιείται συνήθως όταν έχουμε

διωνυμικά δεδομένα, δηλαδή όταν η εξαρτημένη μεταβλητή μπορεί να πάρει μόνο δύο τιμές όπως προαναφέραμε και έχει εφαρμογή σε πολλά στατιστικά προγράμματα, για παράδειγμα R, Python, SPSS, STATA και άλλα. Όσον αφορά τα σχήματα των συναρτήσεων $f(s)$ και $\pi(x)$, είναι παραπλήσια με αυτά που είδαμε στο μοντέλο probit παραπάνω. Σύμφωνα με τους Cox και Snell (1989) οι ουρές των κατανομών των probit και logit διαφέρουν.

Υπάρχουν πολλά ακόμα μοντέλα που χρησιμοποιούνται όταν έχουμε δεδομένα δόσης-απόκρισης. Για παράδειγμα, αν η κατανομή ακραίων τιμών (περιοριστική κατανομή για το ελάχιστο ή το μέγιστο μιας πολύ μεγάλης συλλογής τυχαίων παρατηρήσεων από την ίδια αυθαίρετη κατανομή) με τον παρακάτω τύπο

$$f(s) = \beta_2 \exp [(\beta_1 + \beta_2 s) - \exp(\beta_1 + \beta_2 s)]$$

χρησιμοποιείται ως κατανομή tolerance, τότε το π δίνεται από τον εξής τύπο :

$$\pi = 1 - \exp [-\exp (\beta_1 + \beta_2 x)]$$

Και άρα $\log[-\log(1 - \pi)] = \beta_1 + \beta_2 x$. Αυτός ο σύνδεσμος, $\log[-\log(1 - \pi)]$, ονομάζεται συμπληρωματική συνάρτηση log-log ή αλλιώς c-log-log. Το μοντέλο c-log-log είναι παρεμφερές με τα μοντέλα logit και probit, για τις τιμές όπου το π είναι κοντά στο 0,5. Αντιθέτως διαφέρει από αυτά όταν οι τιμές του π είναι κοντά στο 0 ή 1.

2.2 Συνάρτηση Σύνδεσης

2.2.1 Οι Συνιστώσες Ενός Γενικευμένου Γραμμικού Μοντέλου

Παραπάνω γίνεται αναφορά στην συνάρτηση σύνδεσης των probit, logit και cloglog. Συνεπώς, ο σκοπός μας σε αυτήν την ενότητα είναι να εμβαθύνουμε περισσότερο στην συνάρτηση σύνδεσης.

Όπως είδαμε και παραπάνω, για την ανάλυση σχέσεων δόσης-απόκρισης χρησιμοποιούμε κυρίως γενικευμένα γραμμικά μοντέλα, τα οποία είναι μια επέκταση των κλασικών γραμμικών μοντέλων. Ένα διάνυσμα παρατηρήσεων y που έχει n συνιστώσες, υποτίθεται ότι είναι η πραγματοποίηση μιας τυχαίας μεταβλητής Y , της οποίας οι συνιστώσες κατανέμονται ανεξάρτητα, με μέσο όρο μ . Το συστηματικό μέρος του μοντέλου είναι ένας προσδιορισμός για το διάνυσμα μ ως προς έναν μικρό αριθμό άγνωστων παραμέτρων β_1, \dots, β_p . Για τυπικά γραμμικά μοντέλα, αυτός ο προσδιορισμός έχει την ακόλουθη μορφή :

$$\mu = \sum_1^p x_j \beta_j,$$

Όπου, τα β είναι παράμετροι και εφόσον τις περισσότερες φορές είναι άγνωστοι πρέπει να εκτιμηθούν. Στην περίπτωση που χρησιμοποιηθεί και ο δείκτης i , τότε το συστηματικό μέρος του μοντέλου γράφεται ως εξής :

$$E(Y_i) = \mu_i = \sum_1^p x_{ij} \beta_j, \quad i = 1, \dots, n$$

Όπου, το x_{ij} είναι η τιμή της j_{th} συμμεταβλητής για την παρατήρηση i . Σε μορφή πίνακα, όπου ισχύει ότι, το μ είναι $n \times 1$, το X είναι $n \times p$ και το β είναι $p \times 1$, το μ δίνεται απο τον εξής τύπο :

$$\mu = X\beta$$

Όπου, το X είναι η μήτρα του μοντέλου και το β είναι το διάνυσμα που αποτελείται απο τις παραμέτρους.

Για το τυχαίο μέρος, υποθέτουμε ότι έχουμε ανεξαρτησία και ότι η διακύμανση των σφαλμάτων δεν μεταβάλλεται. Αυτές οι υποθέσεις είναι σημαντικές και πρέπει να ελέγχονται όταν είναι εφικτό. Αντίστοιχα, όσον αφορά την δομή του συστηματικού μέρους, γίνεται η υπόθεση ότι γνωρίζουμε τις συμμεταβλητές που επηρεάζουν την μέση τιμή και μπορούμε να τις μελετήσουμε αποτελεσματικά χωρίς να υπάρχουν σφάλματα. Αυτή η υπόθεση επίσης πρέπει να ελέγχεται. Μια επιπλέον υπόθεση που γίνεται είναι ότι τα σφάλματα ακολουθούν την Κανονική κατανομή με διακύμανση σ^2 .

Συνοψίζοντας λοιπόν, το κλασικό γραμμικό μοντέλο παίρνει την παρακάτω μορφή :

Οι συνιστώσες του Y , είναι ανεξάρτητες Κανονικές μεταβλητές με σταθερή διακύμανση σ^2 και ισχύει ότι :

$$E(Y) = \mu \quad \text{όπου} \quad \mu = X\beta$$

2.2.2 Η Γενίκευση

Για να απλοποιήσουμε τη μετάβαση στα γενικευμένα γραμμικά μοντέλα, θα αναδιατάξουμε ελάχιστα την παραπάνω σχέση, με σκοπό την αναπαραγωγή των ακόλουθων τριών προδιαγραφών :

1. Το τυχαίο στοιχείο : οι συνιστώσες του Y έχουν ανεξάρτητες Κανονικές κατανομές με $E(Y) = \mu$ και σταθερή διακύμανση σ^2 .

2. Το συστηματικό στοιχείο : οι συμμεταβλητές x_1, x_2, \dots, x_p παράγουν μια γραμμική πρόβλεψη, η οποία δίνεται από τον τύπο :

$$\eta = \sum_1^p x_j \beta_j,$$

3. Η σύνδεση μεταξύ του τυχαίου στοιχείου και του συστηματικού στοιχείου είναι η εξής :

$$\mu = \eta$$

Ουσιαστικά, η γενίκευση εισάγει ένα νέο σύμβολο η για τη γραμμική πρόβλεψη και στην συνέχεια το τρίτο συστατικό ορίζει ότι το μ και το η είναι στην πραγματικότητα ίδια.

Στην σχέση $n_i = g(\mu_i)$, το $g(\cdot)$ καλείται συνάρτηση σύνδεσης. Κατά αυτή την περίπτωση, τα κλασικά γραμμικά μοντέλα, ακολουθούν την Κανονική κατανομή στην πρώτη συνιστώσα και την ταυτοτική συνάρτηση για την σύνδεση στην τρίτη συνιστώσα.

Τα γενικευμένα γραμμικά μοντέλα, έχουν δύο επεκτάσεις. Η πρώτη επέκταση είναι ότι, η κατανομή στην πρώτη συνιστώσα προέρχεται από την εκθετική οικογένεια κατανομών και όχι από την Κανονική, ενώ η δεύτερη επέκταση αναφέρεται στο ότι η συνάρτηση σύνδεσης στην τρίτη συνιστώσα έχει την πιθανότητα να γίνει οποιαδήποτε μονότονη διαφοροποιούμενη συνάρτηση.

Παρακάτω θα δούμε την υπόθεση που γίνεται για την κατανομή.

Αρχικά, υποθέτουμε ότι κάθε στοιχείο του Y έχει μια κατανομή απο την εκθετική οικογένεια και για συγκεκριμένες συναρτήσεις $a(\cdot)$, $b(\cdot)$ και $c(\cdot)$ παίρνει την εξής μορφή :

$$f_Y(y; \theta, \varphi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\varphi) + c(y, \varphi)} \right\} \quad (2.1)$$

Αν το φ είναι γνωστό, τότε η παραπάνω σχέση είναι ένα μοντέλο εκθετικής οικογένειας, με κανονική παράμετρο θ . Αν το φ είναι άγνωστο η παραπάνω σχέση μπορεί να είναι, αλλά μπορεί και να μην είναι εκθετική οικογένεια δύο παραμέτρων.

Έτσι για την Κανονική κατανομή έχουμε :

$$f_Y(y; \theta, \varphi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y - \mu)^2}{2\sigma^2} \right\}$$

$$= \exp \left\{ \frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{1}{2} \left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right) \right\},$$

έτσι ώστε,

$$\theta = \mu, \quad \varphi = \sigma^2$$

και

$$\alpha(\varphi) = \varphi, \quad b(\theta) = \theta^2/2, \quad c(y, \varphi) = -\frac{1}{2} \left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right).$$

Γράφουμε $l(\theta, \varphi; y) = \log f_Y(y; \theta, \varphi)$ για την συνάρτηση log-πιθανότητας, που λαμβάνεται υπόψη ως μια συνάρτηση των θ και φ , δοθέντος του y . Η μέση τιμή και η διακύμανση του Y , μπορεί να υπολογιστεί εύκολα από τις παρακάτω σχέσεις.

$$E \left(\frac{\partial l}{\partial \theta} \right) = 0 \tag{2.2}$$

και

$$E \left(\frac{\partial^2 l}{\partial \theta^2} \right) + E \left(\frac{\partial l}{\partial \theta} \right)^2 = 0 \tag{2.3}$$

Από την σχέση (2.1) έχουμε,

$$l(\theta; y) = \frac{\{y\theta - b(\theta)\}}{\alpha(\varphi)} + c(y, \varphi)$$

Από όπου,

$$\frac{\partial l}{\partial \theta} = \frac{\{y - b'(\theta)\}}{\alpha(\varphi)} \tag{2.4}$$

και

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{b''(\theta)}{\alpha(\varphi)} \quad (2.5)$$

Όπου οι πρώτοι δηλώνουν διαφοροποίηση ως προς το θ .

Από τις σχέσεις (2.2) και (2.4) έχουμε :

$$0 = E\left(\frac{\partial l}{\partial \theta}\right) = \frac{\{\mu - b'(\theta)\}}{\alpha(\varphi)}$$

έτσι ώστε,

$$E(Y) = \mu = b'(\theta).$$

Αντίστοιχα απο τις σχέσεις (2.3), (2.4) και (2.5) έχουμε :

$$0 = -\frac{b''(\theta)}{\alpha(\varphi)} + \frac{var(Y)}{\alpha^2(\varphi)}$$

έτσι ώστε,

$$var(Y) = b''(\theta)\alpha(\varphi).$$

Συνεπώς, η διακύμανση του Y είναι προϊόν δύο συναρτήσεων. Η μια συνάρτηση είναι η $b''(\theta)$, η οποία εξαρτάται μόνο απο την κανονική παράμετρο και κατα συνέπεια απο τον μέσο όρο και καλείται συνάρτηση διακύμανσης. Αντίθετα, η δεύτερη εξαρτάται μόνο απο το φ και όχι απο το θ . Η συνάρτηση διακύμανσης θεωρείται ως μία συνάρτηση του μ και γράφεται ως $V(\mu)$.

Η συνάρτηση $\alpha(\varphi)$ συνήθως παίρνει της εξής μορφή :

$$\alpha(\varphi) = \frac{\varphi}{w}$$

όπου, το φ καλείται παράμετρος διασποράς, μπορεί να συμβολιστεί και ως σ^2 και είναι σταθερό από παρατήρηση σε παρατήρηση. Το w είναι ένα γνωστό βάρος, το οποίο ποικίλλει απο παρατήρηση σε παρατήρηση. Έτσι για ένα κανονικό μοντέλο, στο οποίο κάθε παρατήρηση είναι ο μέσος όρος των m ανεξάρτητων μετρήσεων, έχουμε τον εξής τύπο :

$$\alpha(\varphi) = \frac{\sigma^2}{m}$$

έτσι ώστε να ισχύει $w=m$.

Στον παρακάτω πίνακα, συνοψίζονται κάποιες από τις πιο σημαντικές κατανομές της μορφής 2.1.

Κατανομές	Κανονική	Πουασσόν	Διωνυμική	Γάμμα	Αντίστροφη Γκαουσιανή
Σύμβολο Κατανομής	$N(\mu, \sigma^2)$	$P(\mu)$	$\frac{B(m, \pi)}{m}$	$G(\mu, \nu)$	$IG(\mu, \sigma^2)$
Εύρος του y	$(-\infty, \infty)$	$0(1)\infty$	$\frac{0(1)m}{m}$	$(0, \infty)$	$(0, \infty)$
Παράμετρος Δισποράς φ	$\varphi = \sigma^2$	1	$\frac{1}{m}$	$\varphi = \nu^{-1}$	$\varphi = \sigma^2$
Αθροιστική Συνάρτηση $b(\theta)$	$\frac{\theta^2}{2}$	$\exp(\theta)$	$\log(1 + e^\theta)$	$-\log(-\theta)$	$-(-2\theta)^{\frac{1}{2}}$
$c(y;\varphi)$	$-\frac{1}{2}\left(\frac{y^2}{\varphi} + \log(2\pi\varphi)\right)$	$-\log y!$	$\log\left(\frac{m}{my}\right)$	$\nu \log(\nu y) - \log y$ $-\log \Gamma(\nu)$	$-\frac{1}{2}\left\{\log(2\pi\varphi y^3) + \frac{1}{\varphi y}\right\}$
$\mu(\theta)=E(Y;\theta)$	θ	$\exp(\theta)$	$\frac{e^\theta}{1 + e^\theta}$	$-\frac{1}{\theta}$	$(-2\theta)^{-\frac{1}{2}}$
Κανονική Σύνδεση $\theta(\mu)$	identity	log	logit	reciprocal	$\frac{1}{\mu^2}$
Συνάρτηση Διακύμανσης $V(\mu)$	1	μ	$\mu(1 - \mu)$	μ^2	μ^3

Η παράμετρος της μέσης τιμής συμβολίζεται με μ ή με π για την διωνυμική κατανομή.

Η παραμετροποίηση της κατανομής γάμμα είναι τέτοια, ώστε η διακύμανσή της να είναι ίση με $\frac{\mu^2}{\nu}$.

Η κανονική παράμετρος, που συμβολίζεται με θ , ορίζεται από την 2.1. Η σχέση μεταξύ μ και θ δίνεται στις γραμμές 6 και 7 του πίνακα.

Πίνακας 2.1 Χαρακτηριστικά Ορισμένων Μονομετάβλητων Κατανομών Στην Εκθετική Οικογένεια (McCullagh, P. and Nelder, J.A.)

2.2.3 Συναρτήσεις Σύνδεσης

Η συνάρτηση σύνδεσης συσχετίζει την αναμενόμενη τιμή π , ενός δεδομένου y με τον γραμμικό παράγοντα πρόβλεψης η . Στα κλασικά γραμμικά μοντέλα ο μέσος όρος και ο γραμμικός παράγοντας πρόβλεψης είναι πανομοιότυποι. Το η και το π , μπορούν να λάβουν οποιαδήποτε τιμή στην πραγματική γραμμή χάρις τον ταυτοτικό σύνδεσμο. Ωστόσο, όταν έχουμε να κάνουμε με μετρήσεις και η κατανομή είναι Poisson, πρέπει να ισχύει ότι $\pi > 0$, έτσι ώστε ο ταυτοτικός σύνδεσμος να είναι λιγότερο «ελκυστικός», εν μέρει επειδή το η μπορεί να είναι

αρνητικό, ενώ το π όχι. Τα μοντέλα μετρήσεων είναι πολλαπλασιαστικά μοντέλα και η συνάρτηση συνδέσμου είναι $\eta = \log(\pi)$ ή ισοδύναμα $\pi = e^\eta$. Η επιλογή της συγκεκριμένης συνάρτησης συνδέσμου στηρίζεται σε τρεις λόγους :

1^ο Είναι η κανονική συνάρτηση συνδέσμου η οποία μας παρέχει μια επαρκή σ.σ.

2^ο Είναι θετική

3^ο Μετατρέπει την προσθετική σχέση σε πολλαπλασιαστική.

Στην διωνυμική κατανομή ισχύει ότι $0 < \pi < 1$ και ο σύνδεσμος σε αυτήν την περίπτωση πρέπει να καθορίζει το διάστημα $(0,1)$ σε όλη την πραγματική γραμμή. Παρακάτω, παρατίθενται οι τύποι των συναρτήσεων σύνδεσης των logit, probit και complementary log-log, όπου αναφέρθηκαν και σε προηγούμενο κεφάλαιο.

1. logit

$$\eta = \log \left\{ \frac{\pi}{1 - \pi} \right\}$$

2. probit

$$\eta = \Phi^{-1}(\pi)$$

Όπου το $\Phi(\cdot)$ είναι η αθροιστική συνάρτηση Κανονικής κατανομής.

3. complementary log-log

$$\eta = \log \{ -\log(1 - \pi) \}$$

Η οικογένεια των συνδέσμων είναι σημαντική, ειδικότερα για παρατηρήσεις που έχουν θετικό μέσο όρο. Η οικογένεια συνδέσμου της συνάρτησης δύναμης (power) προσδιορίζεται είτε από τον παρακάτω τύπο :

$$\eta = \frac{\pi^\lambda - 1}{\lambda},$$

όπου έχει ως όριο

$$\eta = \log \pi \quad \text{καθώς} \quad \lambda \rightarrow 0$$

είτε απο τον παρακάτω τύπο

$$\eta = \begin{cases} \pi^\lambda, & \lambda \neq 0 \\ \log \pi, & \lambda = 0 \end{cases}$$

Το πλεονέκτημα του πρώτου τύπου, είναι ότι υπάρχει ομαλή μετάβαση όσο το λ διέρχεται απο το 0. Παρ'όλα αυτά, κατά την χρήση και των δύο τύπων, πρέπει να λαμβάνεται υπόψη το $\lambda=0$ σε όλους τους υπολογισμούς.

2.2.4 Επαρκής Στατιστική και Κανονικές Συνδέσεις

Κάθε μια από τις κατανομές που παρουσιάζονται στον παραπάνω πίνακα, έχει την δική της συνάρτηση σύνδεσης, για την οποία υπάρχει επαρκής στατιστική ίση σε διάσταση με το β στο γραμμικό παράγοντα πρόβλεψης $\eta = \sum x_j \beta_j$. Αυτές οι συνδέσεις ονομάζονται κανονικές συνδέσεις και εμφανίζονται όταν ισχύει $\theta = \eta$, όπου το θ είναι η κανονική παράμετρος όπως την έχουμε δει και στον παραπάνω πίνακα.

Οι κανονικές συνδέσεις των κατανομών που αναφέρονται στον πίνακα που είδαμε παραπάνω είναι οι εξής :

Κανονική	$\eta = \pi$
Πουασσόν	$\eta = \log \pi$
Διωνυμική	$\eta = \{\pi/(1 - \pi)\}$
Γάμμα	$\eta = \pi^{-1}$
Αντίστροφη Γκαουσιανή	$\eta = \pi^{-2}$

Για τις κανονικές συνδέσεις, η επαρκής στατιστική είναι $X^T Y$ σε μορφή διανύσματος, με συνιστώσες

$$\sum_i x_{ij} Y_i, \quad j = 1, \dots, p.$$

Η εύρεση μιας επαρκούς στατιστικής συνάρτησης είναι σημαντική για την κατασκευή αποτελεσματικών εκτιμητών, οι οποίοι εκφάζονται σε σχέση με μια επαρκή στατιστική συνάρτηση. Είναι σημαντικό να σημειωθεί ότι, παρόλο που οι κανονικές συνδέσεις οδηγούν σε επιθυμητές στατιστικές ιδιότητες του μοντέλου, κυρίως σε μικρά δείγματα, δεν υπάρχει λόγος για τον οποίο τα συστηματικά αποτελέσματα σε ένα μοντέλο θα πρέπει να είναι

αθροιστικά στην κλίμακα που δίνεται από αυτόν τον σύνδεσμο. Το να είναι, σίγουρα είναι θεμιτό, αλλά αυτό δεν πρέπει να αποτελεί κριτήριο επιλογής του μοντέλου.

2.3 Συναρτήσεις Σύνδεσης Για Μοντέλα Με Δυναμικές Αποκρίσεις

Μερικές φορές, προκειμένου να μελετηθεί η σχέση μεταξύ της πιθανότητας της απόκρισης π και του διανύσματος x_1, \dots, x_p , είναι πρακτικό να κατασκευάζεται ένα μοντέλο το οποίο να περιγράφει την επίδραση στο π , με την χρήση των αλλαγών στο x_1, \dots, x_p . Στην πραγματικότητα, το συγκεκριμένο μοντέλο περιλαμβάνει υποθέσεις που έχουν να κάνουν με την γραμμικότητα, την μηδενική συσχέτιση ή την ανεξαρτησία, της έλλειψη αλληλεπίδρασης κ.α. Αυτές οι υποθέσεις πρέπει να λαμβάνονται υπόψη και να ελέγχονται. Επίσης, η συμπεριφορά του μοντέλου πρέπει να συνάδει με τους γνωστούς νόμους των μαθηματικών της βιολογίας και της φυσικής, ιδιαίτερα όσον αφορά τα όρια του.

Τα Γραμμικά Μοντέλα, είναι αρκετά σημαντικά τόσο στην εφαρμογή όσο και στην θεωρία. Συνεπώς, υποθέτουμε ότι η εξάρτηση του π με το διάνυσμα x_1, \dots, x_p , μπορεί να μελετηθεί από τον ακόλουθο τύπο

$$\eta = \sum_{j=1}^p x_j \beta_j$$

για άγνωστους συντελεστές β_1, \dots, β_p .

Αν δεν επιβληθούν περιορισμοί στο β , έχουμε $-\infty < \eta < \infty$. Για να εκφράσουμε το π ως γραμμικό συνδυασμό, θα έρθουμε αντιμέτωποι με την ασυνέπεια ως προς τους νόμους των πιθανοτήτων. Προκειμένου να αποφύγουμε αυτή την ασυνέπεια, καλό είναι να χρησιμοποιήσουμε τον μετασχηματισμό $g(\pi)$, ο οποίος αντιστοιχίζει το μοναδιαίο διάστημα σε ολόκληρη την πραγματική γραμμή $(-\infty, \infty)$. Αυτό το βήμα θα μας οδηγήσει σε περιπτώσεις γενικευμένων γραμμικών μοντέλων, όπου το συστηματικό μέρος είναι το εξής :

$$g(\pi_i) = \eta_i = \sum_{j=1}^p x_{ij} \beta_j \quad i = 1, \dots, n$$

Μια ευρεία επιλογή συναρτήσεων σύνδεσης είναι διαθέσιμη. Παρακάτω βρίσκονται τρεις συναρτήσεις, που χρησιμοποιούνται αρκετά στην πράξη και έχει γίνει ήδη αναφορά σε αυτές σε προηγούμενα κεφάλαια.

1. logit ή λογιστική συνάρτηση

$$g_1(\pi) = \log\left\{\frac{\pi}{1-\pi}\right\}$$

2. probit ή αντίστροφη Κανονική συνάρτηση

$$g_2(\pi) = \Phi^{-1}(\pi)$$

3. complementary log-log συνάρτηση

$$g_3(\pi) = \log\{-\log(1-\pi)\}$$

Μια τέταρτη πιθανότητα, η log-log συνάρτηση

$$g_4(\pi) = -\log\{-\log(\pi)\}$$

Το οποίο είναι το φυσικό αντίγραφο της συμπληρωματικής συνάρτησης log-log και χρησιμοποιείται σπάνια, επειδή η συμπεριφορά του είναι ακατάλληλη για $\pi < \frac{1}{2}$, την περιοχή που συνήθως μας ενδιαφέρει να μελετήσουμε. Όλες οι συναρτήσεις παραπάνω, μπορούν να θεωρηθούν ως αντίστροφες πολύ γνωστών αθροιστικών συναρτήσεων κατανομής, που έχουν υποστήριξη σε ολόκληρο τον πραγματικό άξονα.

Κάποιες ιδιότητες των παραπάνω συναρτήσεων συνδέσμων είναι οι εξής :

Οι δύο πρώτες συναρτήσεις είναι συμμετρικές με την εξής έννοια :

$$g_1(\pi) = -g_1(1-\pi)$$

Η οποία σχέση προκύπτει από τα παρακάτω :

Για την g_1 έχουμε

$$g_1(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = -\log\left(\frac{1-\pi}{\pi}\right) = -g_1(1-\pi)$$

Για την g_2 έχουμε

$$\begin{aligned}g_2(\pi) &= \Phi^{-1}(\pi) \\ \Leftrightarrow \Phi(g_2(\pi)) &= \pi \\ \Leftrightarrow P(Z \leq g_2(\pi)) &= \pi\end{aligned}\tag{2.6}$$

Αλλά και

$$\begin{aligned}g_2(1 - \pi) &= \Phi^{-1}(1 - \pi) \\ \Leftrightarrow \Phi(g_2(1 - \pi)) &= 1 - \pi \\ \Leftrightarrow P(Z \leq g_2(1 - \pi)) &= 1 - \pi \\ \Leftrightarrow P(Z \leq -g_2(1 - \pi)) &= \pi\end{aligned}\tag{2.7}$$

Γιατί γνωρίζουμε ότι ισχύει $P(Z < -x) = 1 - P(Z < x)$.

Από τις σχέσεις (2.6) και (2.7) έχουμε ότι

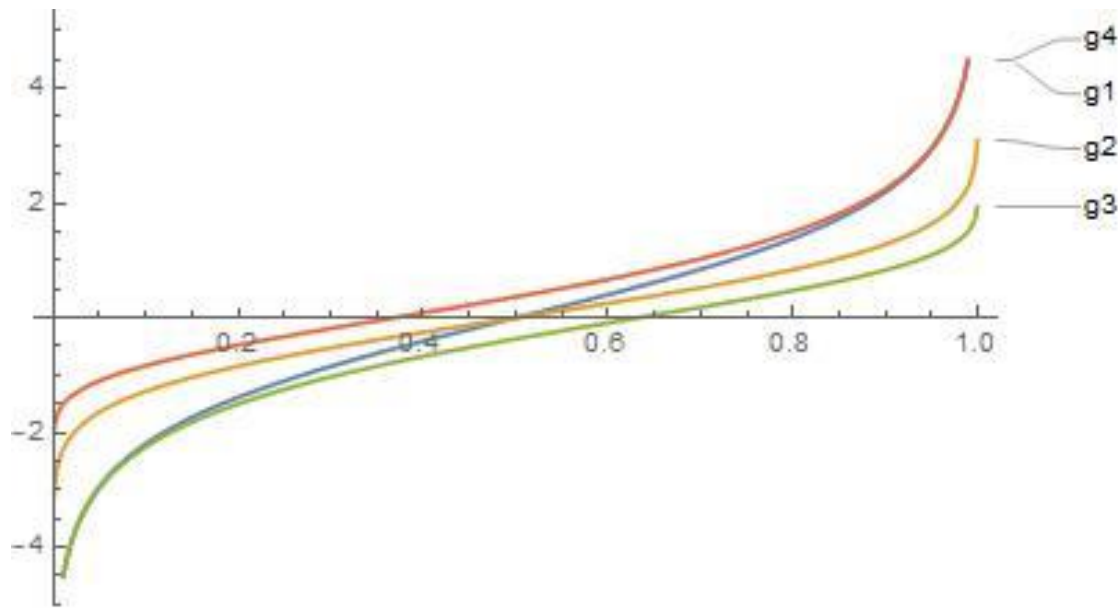
$$g_2(\pi) = -g_2(1 - \pi).$$

Ενώ, οι δύο τελευταίες συναρτήσεις δεν είναι συμμετρικές με αυτήν την έννοια, αλλά σχετίζονται μέσω της εξής σχέσης :

$$g_3(\pi) = -g_4(1 - \pi)$$

Και οι τέσσερις συναρτήσεις που έχουν παρουσιαστεί παραπάνω είναι συνεχείς και αυξανόμενες στο διάστημα (0,1).

Με την χρήση του excel και με δεδομένα από τους P. McCullagh(1983) και John A. Nelder(1983), απεικονίσαμε γραφικά τις τέσσερις συναρτήσεις που μελετήσαμε παραπάνω.



Σχήμα 2.3 Γραφική Σύγκριση Συναρτήσεων g_1, g_2, g_3, g_4

Το παραπάνω Σχήμα 2.3 συγκρίνει τις τέσσερις συναρτήσεις. Η λογιστική συνάρτηση λαμβάνεται ως πρότυπο και οι $g_2(\pi)$, $g_3(\pi)$, $g_4(\pi)$ σχεδιάζονται σε σχέση με την $g_1(\pi)$, για τιμές του π στο διάστημα $(0.01, 0.99)$.

Οι συναρτήσεις logit και probit είναι σχεδόν γραμμικά συσχετισμένες στο διάστημα $0.1 \leq \pi \leq 0.9$. Για αυτόν τον λόγο, συνήθως είναι δύσκολο να γίνει διάκριση μεταξύ αυτών των δύο συναρτήσεων με βάση την καλή προσαρμογή. Για μικρές τιμές του π , η complementary log-log συνάρτηση είναι κοντά στην logistic, και οι δύο είναι κοντά στο $\log(\pi)$. Όσο το π τείνει στο 1, η complementary log-log συνάρτηση τείνει στο άπειρο, αλλά με πολύ πιο αργό ρυθμό από ότι οι συναρτήσεις logit και probit.

Όλη η ασυμπτωτική και η προσεγγιστική θεωρία που παρουσιάζεται σε αυτό το κεφάλαιο ισχύει ανεξάρτητα από την επιλογή της συνάρτησης σύνδεσης. Ωστόσο, στην πραγματικότητα η λογιστική συνάρτηση είναι αυτή που χρησιμοποιείται περισσότερο. Αυτό συμβαίνει κυρίως λόγω της απλής ερμηνείας της ως ο λογάριθμος του λόγου πιθανοτήτων και όχι λόγω των απλούστερων θεωρητικών ιδιοτήτων της. Επιπρόσθετα, η λογιστική συνάρτηση έχει ένα σημαντικό πλεονέκτημα έναντι όλων των άλλων μετασχηματισμών, το οποίο έχει να κάνει με την καταλληλότητα της όσον αφορά την ανάλυση δεδομένων τα οποία συλλέγονται αναδρομικά.

2.4 Παλινδρόμηση Τριών Παραμέτρων

Όταν μελετώνται ποσοτικές αποκρίσεις σε ένα μεγάλο εύρος δοσολογίας, συχνά παρατηρείται μια σιγμοειδής σχέση μεταξύ της απόκρισης και της δοσολογίας. Με την κατάλληλη επιλογή κλίμακας δοσολογίας, συχνά με λογαριθμική κλίμακα, η σχέση μπορεί να προσεγγιστεί από την κανονική σιγμοειδή μορφή. Αν και ο μετασχηματισμός probit αρχικά αναπτύχθηκε για την ανάλυση δεδομένων διωνυμικής απόκρισης, η ιδιότητά του να μετατρέπει μια κανονική σιγμοειδή καμπύλη σε μια ευθεία γραμμή δεν εξαρτάται από την πηγή ή τη φύση των δεδομένων. Συνεπώς, ο ίδιος μετασχηματισμός εφαρμόζεται όταν η σιγμοειδής καμπύλη αντιπροσωπεύει τόσο τη σχέση δόσης-ποσοτικής απόκρισης όσο και την σχέση δόσης-διωνυμικής απόκρισης.

Ωστόσο, υπάρχει μια σημαντική διαφορά μεταξύ των στατιστικών διαδικασιών για τα δύο είδη δεδομένων. Για διωνυμικές αποκρίσεις, η αναλογία των ατόμων που ανταποκρίνονται σε μια συγκεκριμένη δοσολογία αξιολογείται με τον υπολογισμό της αναλογίας του αριθμού ατόμων που εμφανίζουν τη χαρακτηριστική απόκριση προς τον συνολικό αριθμό ατόμων που λαμβάνουν αυτήν τη δοσολογία. Από την άλλη πλευρά, για ποσοτικές αποκρίσεις, μπορεί να υπολογιστεί η μέση απόκριση για κάθε δοσολογία με βάση τα δεδομένα, αλλά γενικά δεν είναι γνωστή η μέγιστη δυνατή απόκριση και, συνεπώς, δεν μπορεί να υπολογιστεί απευθείας μια αναλογική απόκριση.

Η διακύμανση της απόκρισης όλων των δοκιμαζόμενων που λαμβάνουν την ίδια δοσολογία, μπορεί συνήθως να εκτιμηθεί εμπειρικά από τις ξεχωριστές παρατηρούμενες αποκρίσεις σε εκείνη τη δοσολογία, ανεξάρτητα από την αναμενόμενη απόκριση για εκείνη τη δοσολογία. Η διακύμανση της μέσης απόκρισης παίρνει τη μορφή v/n , όπου v είναι η διακύμανση μιας μεμονωμένης απόκρισης και n είναι ο αριθμός των αποκρίσεων που συντείνουν στη μέση τιμή. Σε χαμηλές δοσολογίες, η μέση απόκριση τείνει προς το μηδέν. Αν όμως, οι αποκρίσεις των ατόμων είναι αρνητικές, η διακύμανση (v) πρέπει να γίνεται πολύ μικρή. Επίσης, σε υψηλές δοσολογίες, όταν η μέση απόκριση πλησιάζει το μέγιστο της, η διακύμανση (v) μπορεί να γίνει ξανά μικρή. Σύμφωνα με τον Finney (1971), αν χρησιμοποιηθούν πολύ ακραίες δοσολογίες, η v μπορεί συχνά να θεωρηθεί ως σταθερή και να εκτιμάται συλλέγοντας τις εκτιμήσεις για διάφορα επίπεδα δοσολογίας.

Εάν η μέση απόκριση για τα n άτομα που δοκιμάστηκαν στη δοσολογία x (σε κανονικοποιημένη κλίμακα) είναι u , τότε η αναμενόμενη τιμή του u μπορεί να γραφεί ως :

$$U = HP.$$

Εδώ, η H αναπαριστά τη μέση περιοριστική απόκριση για υψηλές δοσολογίες, και το P αναπαριστά την αναλογική απόκριση σε αυτήν τη δοσολογία. Το P καθορίζεται από μια

κανονική κατανομή με μέση τιμή 5 και διακύμανση 1. Το probit του P είναι το Y . Ο τύπος του P δίνεται παρακάτω :

$$P = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{Y-5} e^{-\frac{1}{2}u^2} du$$

Όταν η τιμή του H είναι γνωστή, δεν απαιτείται να εκτιμηθεί από τα δεδομένα και μια εκτίμηση του P είναι η :

$$p = \frac{U}{H}$$

Τα probits των διάφορων τιμών του p , μπορούν να προσαρμοστούν από μια γραμμική παλινδρόμηση. Οι προτεινόμενες τυπικές αποκλίσεις των εκτιμηθέντων παραμέτρων από την ανάλυση είναι ανάλογες με \sqrt{u} και βασίζονται στον ίδιο αριθμό βαθμών ελευθερίας με το v αντί να αναφέρονται σε μια κανονική κατανομή. Αντί να χρησιμοποιείται ένα τεστ χ^2 για τη σημαντικότητα των αποκλίσεων των παρατηρήσεων από την προσαρμοσμένη παλινδρόμηση probit, πρέπει να χρησιμοποιηθεί ένα τεστ λόγου διακύμανσης.

Όταν το H είναι γνωστό, το γραμμικό μοντέλο που χρησιμοποιείται είναι το

$$y = a + bx$$

Στο οποίο χρησιμοποιείται ο συντελεστής βαρύτητας

$$w = Z^2.$$

Όταν το H δεν είναι γνωστό, τότε πρέπει να εκτιμηθεί από τα δεδομένα μας. Για τον λόγο αυτό εισάγεται μια βοηθητική μεταβλητή (τυποιημένο ποσοστό θνησιμότητας)

$$x' = \frac{P}{Z}$$

Οι τιμές της παραπάνω σχέσης είναι τυποποιημένες και μπορούν να βρεθούν από τον Πίνακα II (Finney 1971) μέρος του οποίου είναι ο Πίνακας 2.2 παρακάτω. Καθορίζονται από τον πίνακα 2.2, βλέποντας την στήλη Q/Z και $10-Y$ αντί για Y .

Y	Q/Z	Percentage natural mortality, C									
		21	22	23	24	25	26	27	28	29	30
6-6	0-4940	-18540	-18294	-18049	-17805	-17561	-17317	-17073	-16829	-16586	-16343
6-7	0-4739	-16252	-16039	-15826	-15613	-15401	-15188	-14976	-14764	-14552	-14341
6-8	0-4551	-14105	-13921	-13738	-13554	-13371	-13188	-13005	-12822	-12639	-12457
6-9	0-4376	-12119	-11962	-11805	-11649	-11492	-11336	-11179	-11023	-10866	-10710
7-0	0-4214	-10308	-10175	-10042	-9909	-9777	-9644	-9512	-9379	-9247	-9114
7-1	0-4062	-08677	-08566	-08455	-08343	-08232	-08121	-08010	-07899	-07787	-07676
7-2	0-3919	-07230	-07137	-07045	-06953	-06860	-06768	-06675	-06583	-06491	-06398
7-3	0-3786	-05962	-05886	-05809	-05733	-05657	-05581	-05505	-05429	-05353	-05277
7-4	0-3661	-04865	-04803	-04741	-04679	-04617	-04555	-04493	-04431	-04369	-04307
7-5	0-3543	-03928	-03878	-03828	-03778	-03728	-03678	-03628	-03578	-03528	-03479
7-6	0-3432	-03139	-03099	-03059	-03019	-02979	-02939	-02899	-02860	-02820	-02780
7-7	0-3327	-02481	-02450	-02418	-02387	-02355	-02324	-02292	-02261	-02229	-02198
7-8	0-3228	-01941	-01917	-01892	-01867	-01843	-01818	-01793	-01769	-01744	-01720
7-9	0-3134	-01502	-01483	-01464	-01445	-01426	-01407	-01388	-01369	-01350	-01331
8-0	0-3046	-01151	-01136	-01122	-01107	-01092	-01078	-01063	-01049	-01034	-01019
8-1	0-2962	-00872	-00861	-00850	-00839	-00828	-00817	-00806	-00795	-00784	-00773
8-2	0-2882	-00654	-00646	-00637	-00629	-00621	-00612	-00604	-00596	-00588	-00579
8-3	0-2806	-00485	-00479	-00473	-00467	-00461	-00454	-00448	-00442	-00436	-00430
8-4	0-2734	-00356	-00352	-00347	-00343	-00338	-00334	-00329	-00325	-00320	-00316
8-5	0-2666	-00259	-00255	-00252	-00249	-00246	-00242	-00239	-00236	-00232	-00229
8-6	0-2600	-00186	-00184	-00181	-00179	-00176	-00174	-00172	-00169	-00167	-00165
8-7	0-2538	-00132	-00130	-00129	-00127	-00125	-00124	-00122	-00120	-00119	-00117
8-8	0-2478	-00093	-00092	-00091	-00090	-00088	-00087	-00086	-00085	-00084	-00083
8-9	0-2421	-00065	-00064	-00063	-00062	-00062	-00061	-00060	-00059	-00058	-00057
9-0	0-2367	-00045	-00044	-00044	-00043	-00042	-00042	-00041	-00041	-00040	-00040

Πίνακας 2.2 Ποσοστό Θνησιμότητας

Στην περίπτωση αυτή το μοντέλο είναι

$$y = a + bx + \frac{\delta H}{H} x'$$

Αν το H αντιστοιχεί στην τιμή της μέγιστης απόκρισης και δH , μια προσαρμογή του H τότε οι συντελεστές της ευθείας παλινδρόμησης καθορίζονται από το παρακάτω σύστημα εξισώσεων.

$$\begin{cases} bS_{xx} + \frac{\delta H}{H} S_{xx'} = S_{xy} \\ bS_{xx'} + \frac{\delta H}{H} S_{x'x'} = S_{x'y} \end{cases}$$

και

$$\alpha = \bar{y} - b\bar{x} - \frac{\delta H}{H} \bar{x}',$$

όπου το σταθμισμένο άθροισμα $S_{xx'}$ δίνεται από τον παρακάτω τύπο :

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

το S_{xx} δίνεται από τον παρακάτω τύπο :

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2,$$

το S_{xy} δίνεται από τον παρακάτω τύπο :

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

το $S_{x'x'}$ δίνεται από τον παρακάτω τύπο :

$$S_{x'x'} = \sum_{i=1}^n (x'_i - \bar{x}')^2,$$

και το $S_{x'y}$ δίνεται από τον παρακάτω τύπο :

$$S_{x'y} = \sum_{i=1}^n (x'_i - \bar{x}')(y_i - \bar{y}).$$

ΚΕΦΑΛΑΙΟ 3^ο : Εφαρμογές Μοντέλων Δόσης-Απόκρισης με Πραγματικά Δεδομένα

3.1 Διωνυμική Απόκριση : Σύγκριση Μοντέλων logit, probit και cloglog

Προκειμένου να εμβαθύνουμε περισσότερο στα μοντέλα που αναφέραμε παραπάνω, θα χρησιμοποιούσουμε πραγματικά δεδομένα και μέσω της R θα τρέξουμε τρία γενικευμένα γραμμικά μοντέλα : 1) logit, 2) probit, 3) cloglog. Ο σκοπός αυτής της ανάλυσης είναι η σύγκριση των τριών βασικών μοντέλων που χρησιμοποιούνται όταν έχουμε δεδομένα δόσης-απόκρισης, η ερμηνεία των αποτελεσμάτων και η απόφαση για το ποιο μοντέλο προσαρμόζεται καλύτερα στα δεδομένα μας.

Τα δεδομένα που έχουν χρησιμοποιηθεί για το συγκεκριμένο παράδειγμα, προέρχονται από τους Jeske et.al (2009), οι οποίοι έκαναν κάποια πειράματα πάνω σε μύγες, με σκοπό την ανάλυση της τοξικότητας του ιχνοστοιχείου σεληνίου. Το σελήνιο μπορεί να χρησιμοποιηθεί για τον έλεγχο παρασίτων, καθώς απωθεί τα έντομα, μειώνει την ανάπτυξή τους ή προκαλεί τοξικές επιδράσεις. Αντιθέτως, η επίδραση του στην ανάπτυξη των φυτών είναι θετική. Οι 4 βασικές μορφές σεληνίου είναι οι εξής : 1) σεληνικό 2) σεληνίτης, 3) σεληνομεθειονίνη και 4) σεληνοκυστεΐνη. Στο συγκεκριμένο παράδειγμα θα επικεντρωθούμε στην ανάλυση της σεληνομεθειονίνης.

Πρόκειται για ένα πλαίσιο δεδομένων που αποτελείται από 8 παρατηρήσεις στις ακόλουθες τέσσερις μεταβλητές : Την κατηγορική μεταβλητή TYPE, όπου είναι ο τύπος του σεληνίου, δηλαδή η σεληνομεθειονίνη. Το αριθμητικό διάνυσμα CONC, όπου περιλαμβάνει την συγκέντρωση του σεληνίου. Το αριθμητικό διάνυσμα NS, όπου περιλαμβάνει τον αριθμό των μυγών πριν την έκθεση τους στο σελήνιο και το αριθμητικό διάνυσμα ND , όπου περιλαμβάνει τον αριθμό των μυγών που απεβίωσαν μετά την έκθεση τους στο σελήνιο.

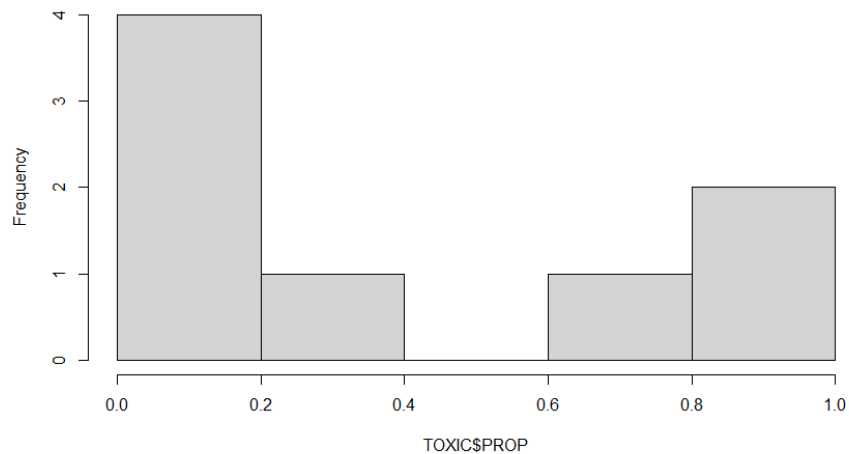
Τα δεδομένα τα οποία αναλύονται στην συγκεκριμένη άσκηση βρίσκονται στο παράρτημα 1.

✚ Αρχικά διαβάζουμε το αρχείο με τα δεδομένα και ορίζουμε την στήλη TYPE ως factor.

```
> setwd("C:\\anastasia\\Dose-Response")
> TOXIC <- read.table("SELENIUM.TXT",header=TRUE)
> TOXIC$TYPE <- as.factor(TOXIC$TYPE)
```

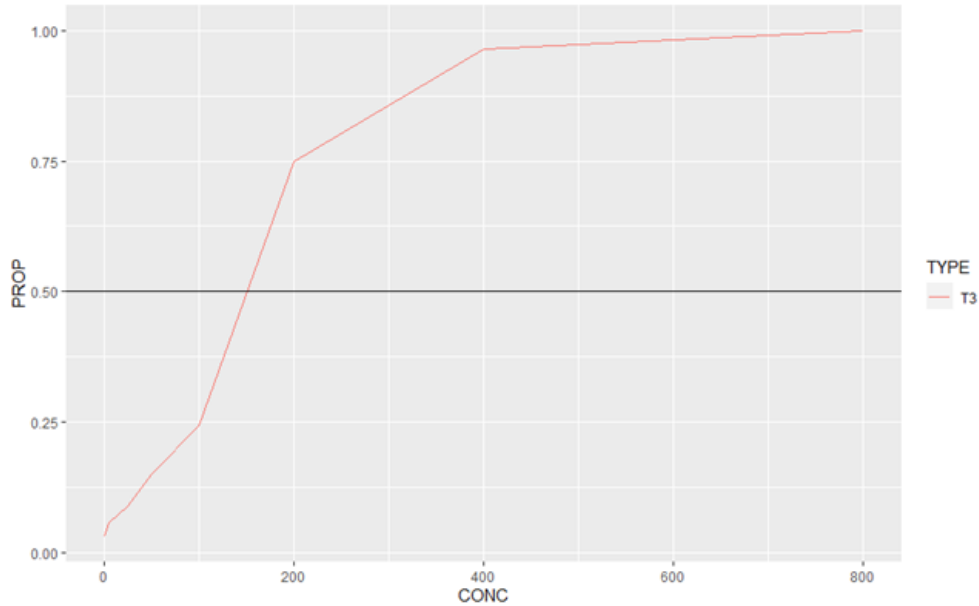
- ✚ Στην συνέχεια υπολογίζουμε τον αριθμό των θανάτων σε σχέση με τον συνολικό αριθμό των δειγμάτων και δημιουργούμε δύο γραφικές παραστάσεις για να έχουμε καλύτερη οπτική των δεδομένων.

```
> TOXIC$PROP <- TOXIC$ND/TOXIC$NS  
> hist(TOXIC$PROP)  
> library(ggplot2)  
> conc <- ggplot(data=TOXIC, aes(x=CONC, y=PROP)) + geom_line(aes(colour=TYPE))  
> conc + geom_hline(yintercept=0.5)  
> abline(h=0.5)
```



Σχήμα 3.1 Ιστόγραμμα Μυγών Που Απεβίωσαν

Απο το σχήμα 3.1 βλέπουμε πόσο πιθανό είναι να πεθάνει μια μύγα μετά την έκθεση της στο σελήνιο. Για την πλειοψηφία των μυγών είναι κατά 0%-20% πιο πιθανό να πεθάνουν μετά την έκθεση τους στο σελήνιο. Για λίγο μικρότερο ποσοστό μυγών είναι κατά 80-100% πιθανό να πεθάνουν μετά την έκθεση τους στο σελήνιο. Ενώ για την μειοψηφία των μυγών είναι κατά 20-40% ή 60-80% πιθανό να πεθάνουν μετά την έκθεση τους στο σελήνιο. Παρατηρούμε ότι δεν υπάρχει πιθανότητα θανάτου με ποσοστό από 40% έως και 60%.



Σχήμα 3.2 Καμπύλη Δόσης-Απόκρισης

Απο το σχήμα 3.2, βλέπουμε ότι τα δεδομένα μας ακολουθούν την τυπική καμπύλη δόσης-απόκρισης η οποία είναι σιγμοειδής και μονότονη. Όσο αυξάνεται η συγκέντρωση της ουσίας στον οργανισμό των μυγών τόσο αυξάνεται και το ποσοστό των θανάτων.

- ✚ Μετά την διαγραμματική απεικόνιση, διαμορφώνουμε τα δεδομένα με τέτοιο τρόπο έτσι ώστε να μπορούν να χρησιμοποιηθούν στα τρία διαφορετικά γραμμικά μοντέλα. Οπότε, από τον αρχικό αριθμό του δείγματος των μυγών αφαιρούμε τον αριθμό των μυγών που πέθαναν, για να βρούμε πόσες μύγες επέζησαν. Έπειτα, συνδυάζουμε σε δύο στήλες τον αριθμό των θανόντων και επιζώντων.

```
> attach(TOXIC)
> N=NS-ND
> N.mat=cbind(ND,N)
```

- ✚ Στην συνέχεια εφαρμόζουμε το μοντέλο logit, probit και cloglog.

```

> res.glm1=glm(N.mat~CONC, family=binomial(link="logit"))
> summary(res.glm1)

> res.glm2=glm(N.mat~CONC, family=binomial(link="probit"))
> summary(res.glm2)

> res.glm3=glm(N.mat~CONC, family=binomial(link="cloglog"))
> summary(res.glm3)

```

Μοντέλο Logit

```

> summary(res.glm1)

```

```

Call:
glm(formula = N.mat ~ CONC, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1014  -0.8945  -0.1847   0.2032   1.6226

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.752544   0.165357  -16.65  <2e-16 ***
CONC         0.017711   0.001191   14.87  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 902.649  on 7  degrees of freedom
Residual deviance: 10.031  on 6  degrees of freedom
AIC: 43.251

Number of Fisher Scoring iterations: 5

```

Μοντέλο Probit

```
> summary(res.glm2)

Call:
glm(formula = N.mat ~ CONC, family = binomial(link = "probit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.35391 -0.96681 -0.28895  0.08688  2.73499

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.5523880  0.0822369  -18.88  <2e-16 ***
CONC         0.0095847  0.0005828   16.45  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 902.649  on 7  degrees of freedom
Residual deviance:  15.978  on 6  degrees of freedom
AIC: 49.198

Number of Fisher Scoring iterations: 6
```

Μοντέλο Cloglog

```
> summary(res.glm3)

Call:
glm(formula = N.mat ~ CONC, family = binomial(link = "cloglog"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9215 -2.1973 -0.9865 -0.1446  5.5668

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.1715779  0.1141790  -19.02  <2e-16 ***
CONC         0.0093257  0.0005487   17.00  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 902.649  on 7  degrees of freedom
Residual deviance:  53.871  on 6  degrees of freedom
AIC: 87.091

Number of Fisher Scoring iterations: 10
```

- ✚ Τέλος, βρίσκουμε τις προσαρμοσμένες τιμές του κάθε μοντέλου και τις πολλαπλασιάζουμε με τον αρχικό αριθμό του δείγματος των μυγών.

```

> fit_p=c(fitted.values(res.glm1))
> fit_y=NS*fit_p
> fit_y
> fit_p2=c(fitted.values(res.glm2))
> fit_y2=NS*fit_p2
> fit_y2
> fit_p3=c(fitted.values(res.glm3))
> fit_y3=NS*fit_p3
> fit_y3

```

Στον παρακάτω πίνακα δίνονται τα αποτελέσματα του παραπάνω κώδικα, δηλαδή οι προσαρμοσμένες τιμές επί τον αρχικό αριθμό του δείγματος των μυγών. Επίσης, εκτός από αυτά δίνονται και κάποια χρήσιμα αποτελέσματα που έχουν προκύψει από την εντολή `summary`, προκειμένου να συγκρίνουμε τα μοντέλα μεταξύ τους.

Αριθμός Θανάτων	Μοντέλο Logit	Μοντέλο Probit	Μοντέλο Cloglog
3	5.994316	6.028471	10.77399
6	6.903983	7.020478	11.93368
11	11.018582	11.544889	16.35420
22	19.413676	20.532214	24.09388
31	34.619782	35.088090	31.93938
105	96.283412	89.919005	72.94091
166	169.768321	170.063377	170.51592
188	187.997928	188.000	188.00000
Summary	Μοντέλο Logit	Μοντέλο Probit	Μοντέλο Cloglog
Null Deviance	902.649	902.649	902.649
Residual Deviance	10.031	15.978	53.871
Estimate Intercept	-2.752544	-1.5523880	-2.1715779
Estimate CONC	0.017711	0.0095847	0.0093257
Std. Error Intercept	0.165357	0.0822369	0.1141790
Std. Error CONC	0.001191	0.0005828	0.0005487
Pr(> z) Intercept	<2e-16	<2e-16	<2e-16
Pr(> z) CONC	<2e-16	<2e-16	<2e-16
Akaike Criterion	43.251	49.198	87.091

Πίνακας 3.1 Αποτελέσματα μοντέλων logit, probit και cloglog

Από τα παραπάνω αποτελέσματα βλέπουμε ότι η μηδενική απόκλιση (902.649) και στα τρία μοντέλα είναι η ίδια, ενώ η απόκλιση του καταλοίπου διαφέρει από μοντέλο σε μοντέλο. Συγκριτικά με τα άλλα, το μοντέλο logit έχει την μικρότερη απόκλιση του καταλοίπου (10.031) και αυτό αποτελεί ένα στοιχείο υπέρ του logit. Ο σταθερός όρος και ο συντελεστής της μεταβλητής συγκέντρωσης είναι παρόμοιος μεταξύ των τριών μοντέλων. Το μοντέλο logit, έχει τα μεγαλύτερα τυπικά σφάλματα, το οποίο δεν είναι θεμιτό γενικά, χωρίς όμως να έχουν μεγάλη διαφορά από αυτά των probit και cloglog. Όσον αφορά το p-value της μεταβλητής συγκέντρωσης είναι στατιστικά σημαντικό σε όλα τα μοντέλα δεδομένου ότι είναι $<2e-16$, το οποίο είναι λογικό γιατί η συγκέντρωση του σεληνίου επηρεάζει την θνησιμότητα των μυγών. Τέλος, το κριτήριο του Akaike, το οποίο μας δείχνει το πόσο καλά προσαρμόζεται ένα μοντέλο στα δεδομένα μας, είναι μικρότερο στο μοντέλο logit (43.251). Όσο μικρότερη η τιμή του Akaike, τόσο καλύτερα προσαρμόζεται το μοντέλο στα δεδομένα μας και από ότι βλέπουμε τα μοντέλα probit και cloglog δεν προσαρμόζονται τόσο καλά όσο το logit. Συνεπώς, για όλους τους παραπάνω λόγους επιλέγουμε το μοντέλο logit ως το καταλληλότερο για το συγκεκριμένο παράδειγμα.

3.2 Συνεχής Απόκριση : Μία Καμπύλη Δόσης-Απόκρισης

Εξίσου σημαντικό είναι να μελετήσουμε την σχέση δόσης-απόκρισης, όταν η απόκριση είναι συνεχής. Αυτό θα επιτευχθεί μέσα από ένα παράδειγμα, για το οποίο έχουν χρησιμοποιηθεί δεδομένα όπου είναι ήδη περασμένα στο πακέτο `drc` της R.

Στην R, τα συγκεκριμένα δεδομένα έχουν την ονομασία `ryegrass` και προέρχονται από τους Inderjit και J. C. Streibig και M. Olofsdotter (2002), οι οποίοι έκαναν μελέτες σχετικά με την επίδραση ενός ζιζανιοκτόνου στην ανάπτυξη πολυετούς σίκαλης. Πρόκειται για ένα πλαίσιο δεδομένων με 24 παρατηρήσεις στις ακόλουθες δύο μεταβλητές : Το αριθμητικό διάνυσμα `Rootl` όπου είναι το μήκος της ρίζας της σίκαλης και το αριθμητικό διάνυσμα `Conc` όπου είναι η συγκέντρωση του φερουλικού οξέος σε mM.

Για την απόκριση του μήκους της ρίζας (U), σε δόσεις φαινολικών (z), το λογιστικό μοντέλο που χρησιμοποιήθηκε είναι το εξής :

$$U_0 = C + \frac{D - C}{1 + \exp[-b_i(\log(z_0) - \log(ED_{50}))]} \quad (2.8)$$

Η σχέση αυτή είναι μια παραλλαγή της εξίσωσης Hill για συνεχείς αποκρίσεις. Πράγματι εύκολα γράφεται στη μορφή

$$U_0 = C + \frac{D - C}{1 + \left(\frac{ED_{50}}{z_0}\right)^{b_i}}. \quad (2.9)$$

Για $C=0, D=1, b_1 = n$ έχουμε την κλασσική εξίσωση Hill.

Για κλίμακα από 0-100, δηλαδή $D=100, C=0$, η παραπάνω σχέση γράφεται ως εξής:

$$x = \frac{100}{1 + \left(\frac{ED_x}{ED_{50}}\right)^b}.$$

Λύνοντας ως προς ED_x έχουμε

$$x + x \left(\frac{ED_x}{ED_{50}}\right)^b = 100$$

ή ισοδυνάμως

$$\left(\frac{ED_x}{ED_{50}}\right)^b = \frac{100 - x}{x}$$

$$\Leftrightarrow \frac{ED_x}{ED_{50}} = \left(\frac{100 - x}{x}\right)^{1/b}.$$

Η τελευταία ισότητα συνεπάγεται την εξίσωση ED_{50} .

$$ED_{50} = \frac{ED_x}{\left(\frac{x}{100 - x}\right)^{\frac{1}{b}}}$$

Το U_{ij} υποδηλώνει το μήκος της ρίζας στην j^{th} ($j=1.7$) δόση της i^{th} ($i=1.3$) του φαινολικού παρασκευάσματος. Τα γράμματα D και C, εκφράζουν τα άνω και τα κάτω όρια του μήκους της ρίζας (στο μηδέν και σε άπειρες δόσεις φαινολικών οξέων). Το ED_{50i} , εκφράζει την δόση που απαιτείται από το i^{th} παρασκεύασμα φαινολικού οξέος για την μείωση του μήκους της ρίζας στα μισά των άνω και κάτω ορίων D και C. Ενώ το b_i είναι ανάλογο με την κλίση της καμπύλης γύρω από ED_{50i} για όλα τα φαινολικά σκευάσματα.

Σύμφωνα με τους Finney (1978) και Streibig et al. (1993), η διακύμανση των αποκρίσεων καθορίστηκε με την χρήση μετασχηματισμών και στις δύο πλευρές της μη γραμμικής παλινδρόμησης.

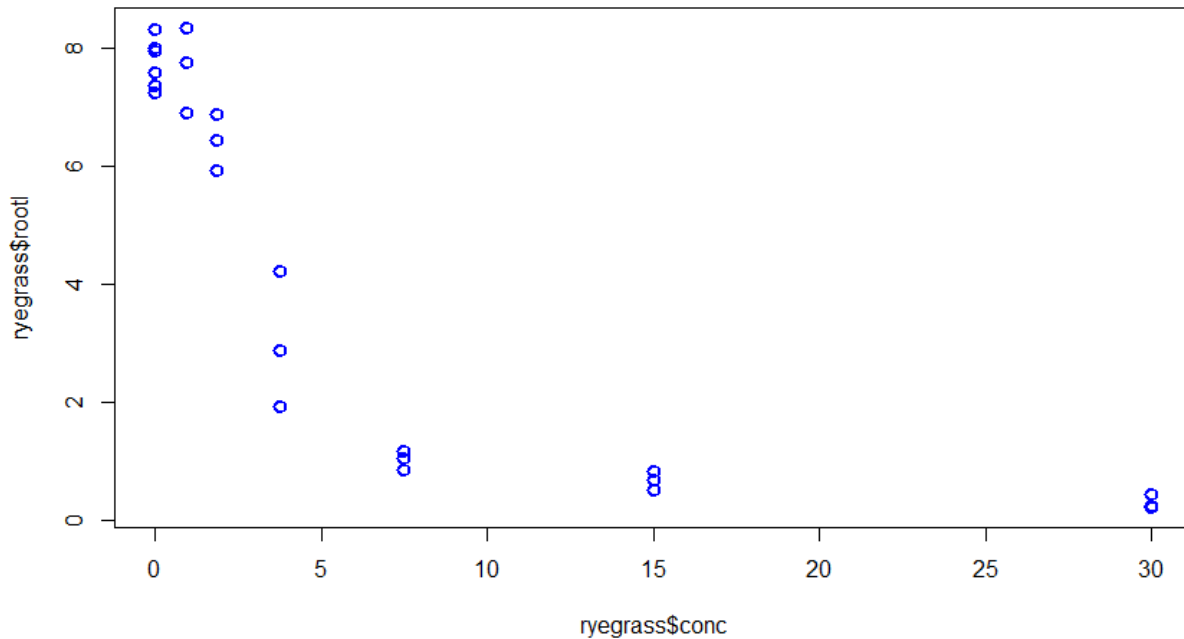
Το ED_{50} γενικά θεωρείται μια 'φυσική' παράμετρος της λογιστικής καμπύλης απόκρισης. Παρ' όλ' αυτά, οποιοδήποτε άλλο επίπεδο απόκρισης (π.χ. ED_{10} , ED_{60}) μπορεί εύκολα να εκφραστεί με μια συνάρτηση των παραμέτρων b και ED_{50} .

Συνεπώς, υπάρχει η δυνατότητα υπολογισμού και άλλων επιπέδων απόκρισης, το οποίο και θα κάνουμε στο συγκεκριμένο παράδειγμα παρακάτω.

Τα δεδομένα τα οποία αναλύονται στην συγκεκριμένη άσκηση βρίσκονται στο παράρτημα 2.

- ✚ Αρχικά φορτώνουμε την βιβλιοθήκη `drc` η οποία είναι κατάλληλη για δεδομένα δόσης-απόκρισης και δημιουργούμε ένα γράφημα έτσι ώστε να έχουμε καλύτερη οπτική των δεδομένων μας.

```
> library(drc)
> plot(ryegrass$conc,ryegrass$rootl,type='p',col='blue',lwd=2,cex=1.3)
```



Σχήμα 3.3 Διάγραμμα Διασποράς Φερουλικού Οξέος-Μήκος Ρίζας Σίκαλης

Στο παραπάνω γράφημα ο άξονας x απεικονίζει την συγκέντρωση του φερουλικού οξέος (δόση) και ο άξονας y απεικονίζει το μήκος της ρίζας της σίκαλης (απόκριση). Αυτό που αντιλαμβανόμαστε από αυτό το γράφημα, είναι ότι όσο αυξάνεται η συγκέντρωση του φερουλικού οξέος, τόσο μειώνεται το μήκος της ρίζας της σίκαλης. Επίσης, παρατηρείται μεγαλύτερη συγκέντρωση παρατηρήσεων στις μικρές συγκεντρώσεις συγκριτικά με τις μεγάλες.

- ✚ Στην συνέχεια, τρέχουμε ένα γενικευμένο μοντέλο με την συνάρτηση `drm` από την βιβλιοθήκη `drm`, η οποία χρησιμοποιείται για ανάλυση δεδομένων δόσης-απόκρισης. Η `drm` βασίζεται στην συνάρτηση `optim`, για την ελαχιστοποίηση της αρνητικής λογιστικής συνάρτησης πιθανοφάνειας. Για συνεχή απόκριση, αυτό είναι ίδιο με την μέθοδο των ελαχίστων τετραγώνων. Δηλαδή την :

$$\sum_{i=1}^N (w_i(y_i - f_i))^2$$

Συνεπώς, σε πρώτη φάση βρίσκουμε κάποια συνοπτικά στοιχεία του μοντέλου που θα αναλύσουμε με το `summary`. Τέλος, για να ελέγξουμε την υπόθεση της γραμμικότητας, χρησιμοποιούμε την εντολή `glht`, από την βιβλιοθήκη `multcomp`.

```
> library(drc)
> library(multcomp)
> ryegrass.LL.4 <- drm(rootl ~ conc, data = ryegrass, fct=LL.4(names = c("Slope", "Lower
Limit", "Upper Limit", "ED50")))
> summary(ryegrass.LL.4)
> summary(glht(ryegrass.LL.4))
```

```
> summary(ryegrass.LL.4)
```

```
Model fitted: Log-logistic (ED50 as parameter) (4 parms)
```

```
Parameter estimates:
```

	Estimate	Std. Error	t-value	p-value	
Slope:(Intercept)	2.98222	0.46506	6.4125	2.960e-06	***
Lower Limit:(Intercept)	0.48141	0.21219	2.2688	0.03451	*
Upper Limit:(Intercept)	7.79296	0.18857	41.3272	< 2.2e-16	***
ED50:(Intercept)	3.05795	0.18573	16.4644	4.268e-13	***

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

```
Residual standard error:
```

```
0.5196256 (20 degrees of freedom)
```

Από την παραπάνω σύνοψη του μοντέλου μας, βλέπουμε την εκτίμηση, το τυπικό σφάλμα, την τιμή t και το p -value για την κλίση, το κατώτατο όριο, το ανώτατο όριο και του ED_{50} . Τα p -values της κλίσης, του κατώτατου ορίου, του ανώτατου ορίου και του ED_{50} είναι μικρότερα του 0.05 που είναι το επίπεδο σημαντικότητας που χρησιμοποιούμε, άρα είναι στατιστικά σημαντικά για το μοντέλο. Επίσης, βλέπουμε ότι το residual standard error είναι αρκετά μικρό, πιο συγκεκριμένα είναι ίσο με 0.5196256, το οποίο σημαίνει ότι τα δεδομένα μας προσαρμόζονται αρκετά καλά στο συγκεκριμένο μοντέλο. Σημειώνεται ότι το residual standard error είναι διαθέσιμο μόνο όταν έχουμε συνεχή δεδομένα δόσης-απόκρισης, όπως στο συγκεκριμένο παράδειγμα.

```
> summary(glht(ryegrass.LL.4))
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: drm(formula = root1 ~ conc, data = ryegrass, fct = LL.4(names = c("Slope",  
"Lower Limit", "Upper Limit", "ED50")))
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)	
slope:(Intercept) == 0	2.9822	0.4651	6.413	<0.001	***
Lower Limit:(Intercept) == 0	0.4814	0.2122	2.269	0.083	.
Upper Limit:(Intercept) == 0	7.7930	0.1886	41.327	<0.001	***
ED50:(Intercept) == 0	3.0580	0.1857	16.464	<0.001	***

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)
```

Από τα παραπάνω αποτελέσματα, παρατηρούμε ότι τα p-values της κλίσης, του ανώτατου ορίου και του ED_{50} είναι αρκετά κοντά στο 0. Συνεπώς, για τα συγκεκριμένα η υπόθεση της γραμμικότητας απορρίπτεται σε επίπεδο 0.05. Όσον αφορά το p-value του κατώτατου ορίου είναι ίσο με 0.083, που σημαίνει ότι σε επίπεδο σημαντικότητας 0.05 δεν απορρίπτεται η υπόθεση της γραμμικότητας.

- ✚ Χρησιμοποιώντας το γενικευμένο γραμμικό μοντέλο που δημιουργήσαμε παραπάνω, υπολογίζουμε τα εξής επίπεδα απόκρισης : ED_5 , ED_{10} και ED_{50} . Αυτό επιτυγχάνεται με την εντολή ED και προκειμένου να λάβουμε υπόψιν τα διαστήματα εμπιστοσύνης, χρησιμοποιούμε το όρισμα "delta".

```
> ED(ryegrass.LL.4, c(5, 10, 50), interval = "delta")
```

```
> ED(ryegrass.LL.4, c(5, 10, 50), interval = "delta")
```

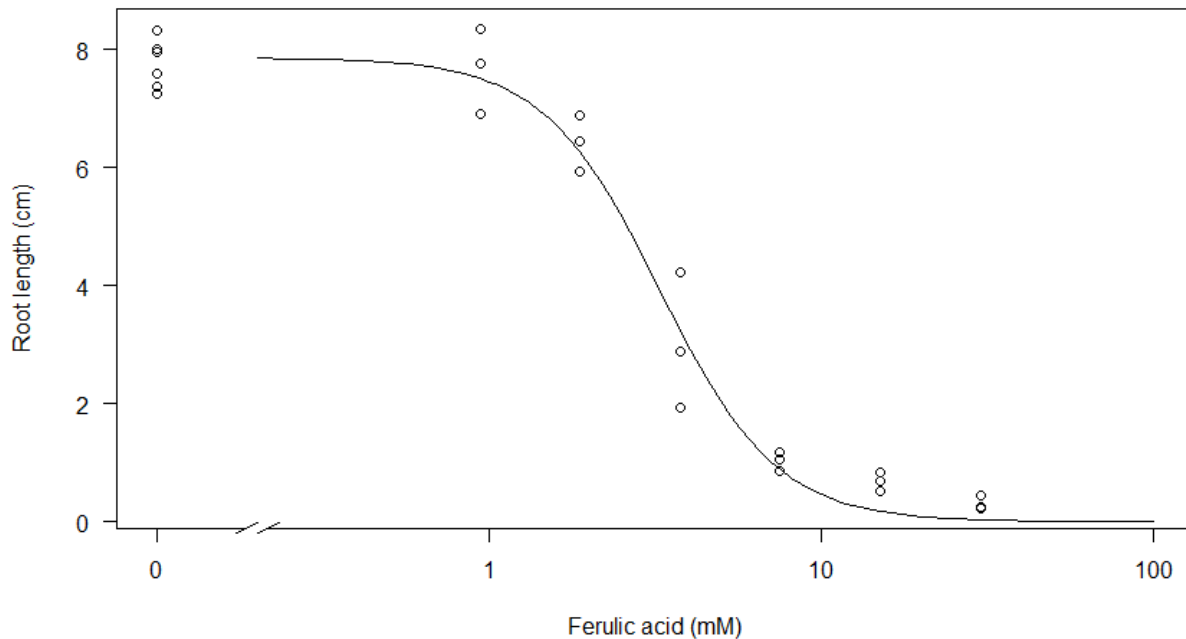
Estimated effective doses

	Estimate	Std. Error	Lower	Upper
e:1:5	0.99088	0.17808	0.62053	1.36122
e:1:10	1.34086	0.18904	0.94773	1.73398
e:1:50	3.26336	0.19641	2.85491	3.67181

Από το παραπάνω πινακάκι βλέπουμε ότι η εκτίμηση για το ED_5 είναι ίση με 0.99088 και το διάστημα εμπιστοσύνης του σε επίπεδο 95% είναι ίσο με [0.62053,1.36122]. Αντίστοιχα, η εκτίμηση για το ED_{10} είναι ίση με 1.34086 και το διάστημα εμπιστοσύνης του σε επίπεδο 95% είναι ίσο με [0.94773,1.73398]. Τέλος, η εκτίμηση του ED_{50} είναι ίση με 3.26336 και το διάστημα εμπιστοσύνης του σε επίπεδο 95% είναι ίσο με [2.85491,3.67181].

✚ Τέλος, απεικονίζουμε γραφικά την καμπύλη δόσης-απόκρισης.

```
> plot(ryegrass.LL.4, broken = TRUE, type = "all",  
+ xlab = "Ferulic acid (mM)", xlim = c(0, 100),  
+ ylab = "Root length (cm)")
```



Σχήμα 3.4 Καμπύλη Δόσης – Απόκρισης Μήκους Ρίζας Σίκαλης Με Φερουλικό Οξύ

Από το παραπάνω σχήμα βλέπουμε ότι η καμπύλη είναι μονότονη και σιγμοειδής. Επίσης, όσο αυξάνεται η δόση, τόσο μειώνεται η απόκριση, άρα παρατηρείται μια καθοδική τάση στην σχέση φερουλικού οξέος-μήκος ρίζας. Να σημειωθεί ότι η απόκριση έχει πιο έντονη καθοδική τάση για δόση από 1 έως 10.

3.3 Συνεχής Απόκριση : Δύο Καμπύλες Δόσης-Απόκρισης

Σε συνέχεια του προηγούμενου παραδείγματος, θα μελετήσουμε πάλι την σχέση δόσης-απόκρισης, όταν η απόκριση είναι συνεχής, αυτή την φορά όμως θα έχουμε δύο καμπύλες δόσης-απόκρισης. Αυτό θα επιτευχθεί μέσα από ένα παράδειγμα, για το οποίο έχουν χρησιμοποιηθεί δεδομένα όπου είναι ήδη περασμένα στο πακέτο `drm` της R.

Στην R, τα συγκεκριμένα δεδομένα έχουν την ονομασία `S.alba` και προέρχονται από τους Christensen, M. G. και Teicher, H. B. και Streibig, J. C. (2003), οι οποίοι έκαναν ένα πείραμα με σκοπό να συγκρίνουν την επίδραση που έχουν τα ζιζανιοκτόνα Glyphosate και Bentazone στην λευκή μουστάρδα. Πρόκειται για ένα σύνολο δεδομένων, όπου αποτελείται από 68 παρατηρήσεις στις ακόλουθες τρεις μεταβλητές : Το αριθμητικό διάλυσμα `Dose`, όπου περιέχει την δόση σε g/ha, την κατηγορική μεταβλητή `Herbicide` όπου αποτελείται από τα επίπεδα Glyphosate και Bentazone και τέλος από το αριθμητικό διάλυσμα `DryMatter` όπου περιέχει την απόκριση σε g/pot.

Για τον υπολογισμό της απόκρισης χρησιμοποιήθηκε η παραπάνω λογιστική καμπύλη (2.8).

Όπου U είναι η απόκριση και z η δόση. Το D είναι το ανώτατο όριο κατά το οποίο η δόση είναι ίση με 0 και C είναι το κατώτατο όριο σε άπειρες δόσεις. Το ED_{50} είναι η δόση που δίνει απόκριση μεταξύ D και C , και το b περιγράφει την κλίση γύρω από το ED_{50} .

Σύμφωνα με τους Christensen, M. G. και Teicher, H. B. και Streibig, J. C. (2003), το κατώτατο και το ανώτατο όριο για τα δύο ζιζανιοκτόνα είναι πανομοιότυπα, ενώ οι κλίσεις και οι τιμές των ED_{50} είναι διαφορετικές στο λογιστικό μοντέλο. Αυτό είναι μια υπόθεση που θα εξασφαλιστεί στην συνέχεια με την χρήση της R.

Το μοντέλο που θα χρησιμοποιηθεί είναι το Log-logistic μοντέλο.

Πράγματι η (2.9) γράφεται

$$\begin{aligned} U_0 &= C + (D - C) \frac{z_0^b}{z_0^b + (ED_{50})^b} \\ &= C + (D - C) \frac{\left(\frac{z_0}{ED_{50}}\right)^b}{\left(\frac{z_0}{ED_{50}}\right)^b + 1} \end{aligned}$$

Η σύγκριση με την συνάρτηση κατανομής της Log-logistic

$$F(x, \lambda, b) = \frac{(\lambda x)^b}{(\lambda x)^b + 1}$$

δικαιολογεί την επιλογή του μοντέλου.

Τα δεδομένα τα οποία αναλύονται στην συγκεκριμένη άσκηση βρίσκονται στο παράρτημα 3.

- ✚ Όπως και στο προηγούμενο παράδειγμα, φορτώνουμε την βιβλιοθήκη `drc` η οποία είναι κατάλληλη για δεδομένα δόσης-απόκρισης. Σε πρώτη φάση τρέχουμε ένα γενικευμένο μοντέλο με την συνάρτηση `drm`, χωρίς να λάβουμε υπόψιν την υπόθεση που αναφέραμε παραπάνω (το κατώτατο και το ανώτατο όριο για τα δύο ζιζανιοκτόνα είναι πανομοιότυπα, ενώ οι κλίσεις και οι τιμές των ED_{50} είναι διαφορετικές) και αυτό γίνεται με την παράλειψη του ορίσματος `pmodels`. Το μοντέλο που χρησιμοποιείται με την εντολή `drm`, έχει αναλυθεί στο προηγούμενο παράδειγμα. Παρακάτω, θα τρέξουμε το ίδιο μοντέλο λαμβάνοντας υπόψιν το όρισμα `pmodels`, για να δούμε τις διαφορές. Συνεπώς, βρίσκουμε κάποια συνοπτικά στοιχεία του μοντέλου με το `summary`. Τέλος, για να ελέγξουμε την υπόθεση της γραμμικότητας, χρησιμοποιούμε την εντολή `glht`, από την βιβλιοθήκη `multcomp`.

```
> library(drc)
> library(multcomp)
> S.alba.LL.4.1 <- drm(DryMatter~Dose, data=S.alba, fct = LL.4(names = c("Slope",
"Lower Limit", "Upper Limit", "ED50")))
> summary(S.alba.LL.4.1)
> summary(glht(S.alba.LL.4.1))
```

```
> summary(S.alba.LL.4.1)
```

```
Model fitted: Log-logistic (ED50 as parameter) (4 parms)
```

```
Parameter estimates:
```

	Estimate	Std. Error	t-value	p-value	
Slope:(Intercept)	2.61651	0.68431	3.8236	0.0003006	***
Lower Limit:(Intercept)	0.79037	0.14863	5.3177	1.428e-06	***
Upper Limit:(Intercept)	3.88832	0.12217	31.8282	< 2.2e-16	***
ED50:(Intercept)	39.45328	4.08723	9.6528	4.170e-14	***

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

```
Residual standard error:
```

```
0.5592301 (64 degrees of freedom)
```

Από την παραπάνω σύνοψη του μοντέλου μας, βλέπουμε την εκτίμηση, το τυπικό σφάλμα, την τιμή t και το p -value για την κλίση, το κατώτατο όριο, το ανώτατο όριο και του ED_{50} . Τα p -values της κλίσης, του κατώτατου ορίου, του ανώτατου ορίου και του ED_{50} είναι μικρότερα

του 0.05 που είναι το επίπεδο σημαντικότητας που χρησιμοποιούμε, άρα είναι στατιστικά σημαντικά για το μοντέλο. Επίσης, βλέπουμε ότι το residual standard error είναι αρκετά μικρό, πιο συγκεκριμένα είναι ίσο με 0.5592301, το οποίο σημαίνει ότι τα δεδομένα μας προσαρμόζονται καλά στο συγκεκριμένο μοντέλο, αλλά θα το συγκρίνουμε και με το residual standard error του επόμενου μοντέλου. Σημειώνεται ότι το residual standard error είναι διαθέσιμο μόνο όταν έχουμε συνεχή δεδομένα δόσης-απόκρισης, όπως στο συγκεκριμένο παράδειγμα.

```
> summary(gLht(S.alba.LL.4.1))
```

```

Simultaneous Tests for General Linear Hypotheses

Fit: drm(formula = DryMatter ~ Dose, data = S.alba, fct = LL.4(names = c("slope",
"Lower Limit", "Upper Limit", "ED50")))

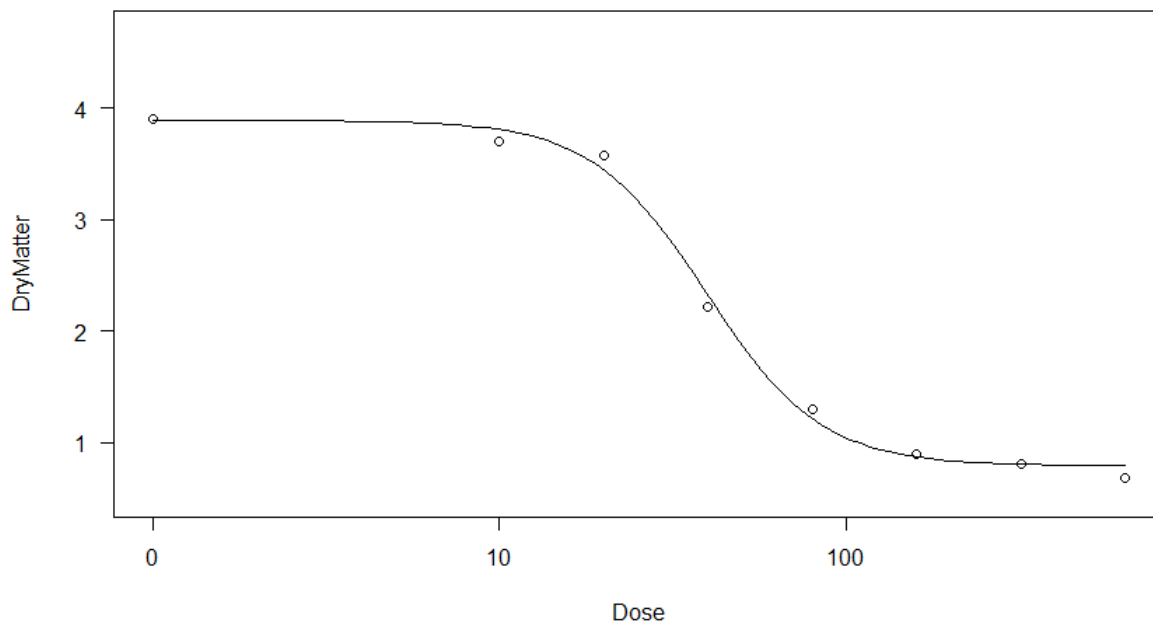
Linear Hypotheses:
              Estimate Std. Error z value Pr(>|z|)
slope:(Intercept) == 0      2.6165    0.6843   3.824 0.000514 ***
Lower Limit:(Intercept) == 0  0.7904    0.1486   5.318 < 1e-04 ***
Upper Limit:(Intercept) == 0  3.8883    0.1222  31.828 < 1e-04 ***
ED50:(Intercept) == 0      39.4533    4.0872   9.653 < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```

Από τα παραπάνω αποτελέσματα, παρατηρούμε ότι τα p-values της κλίσης, του κατώτατου ορίου, του ανώτατου ορίου και του ED_{50} είναι αρκετά κοντά στο 0. Συνεπώς, η υπόθεση της γραμμικότητας απορρίπτεται σε επίπεδο 0.05.

✚ Στην συνέχεια, φτιάχνουμε το γράφημα για το παραπάνω μοντέλο.

```
> plot(S.alba.LL.4.1)
```



Σχήμα 3.5 Καμπύλη Δόσης-Απόκρισης Χωρίς Διαχωρισμό Ζιζανιοκτόνων

Εφόσον, στο συγκεκριμένο μοντέλο δεν έχει γίνει διαχωρισμός των ζιζανιοκτόνων σε Glyphosate και Bentazone, το παραπάνω γράφημα απεικονίζει την σχέση-δόσης απόκρισης και για τα δύο ζιζανιοκτόνα μαζί. Πρόκειται για μια μονότονη και σιγμοειδής καμπύλη. Επίσης, βλέπουμε ότι όσο αυξάνεται η δόση, τόσο μειώνεται η απόκριση, άρα παρατηρείται μια καθοδική τάση στην σχέση ζιζανιοκτόνων-λευκής μουστάρδας. Να σημειωθεί ότι η απόκριση έχει πιο έντονη καθοδική τάση για δόση από 10 έως 100.

- ✚ Σε δεύτερη φάση ξανά τρέχουμε το ίδιο μοντέλο, αλλά κάνοντας χρήση του ορίσματος `pmodels`, έτσι ώστε να λάβουμε υπόψιν την υπόθεση που προαναφέραμε και να κάνουμε διαχωρισμό των ζιζανιοκτόνων. Με το `summary` βλέπουμε κάποια συνοπτικά στοιχεία του μοντέλου και με το `glht` την ύπαρξη γραμμικότητας.

```
> S.alba.LL.4.2 <- drm(DryMatter~Dose, Herbicide, data=S.alba, fct = LL.4(names =
c("Slope", "Lower Limit", "Upper Limit", "ED50")),
+           pmodels=list(~Herbicide-1, ~1, ~1, ~Herbicide-1))
> summary(S.alba.LL.4.2)
```

```

> summary(S.alba.LL.4.2)

Model fitted: Log-logistic (ED50 as parameter) (4 parms)

Parameter estimates:

              Estimate Std. Error t-value  p-value
Slope:HerbicideBentazone  5.046141  1.040135  4.8514 8.616e-06 ***
Slope:HerbicideGlyphosate 2.390218  0.495959  4.8194 9.684e-06 ***
Lower Limit:(Intercept)  0.716559  0.089245  8.0291 3.523e-11 ***
Upper Limit:(Intercept)  3.854861  0.076255 50.5519 < 2.2e-16 ***
ED50:HerbicideBentazone  28.632355  2.038098 14.0486 < 2.2e-16 ***
ED50:HerbicideGlyphosate 66.890545  5.968819 11.2067 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error:

0.3705151 (62 degrees of freedom)

```

Από την παραπάνω σύνοψη του μοντέλου μας, βλέπουμε την εκτίμηση, το τυπικό σφάλμα, την τιμή t και το p -value, αυτή την φορά ξεχωριστά για το Glyphosate και Bentazone όσον αφορά την κλίση, και το ED_{50} . Ενώ για το ανώτατο και κατώτατο όριο όπως προαναφέραμε δεν υπάρχει διαχωρισμός σε Glyphosate και Bentazone, για αυτό τα όρια είναι κοινά και για τα δύο ζιζανιοκτόνα. Όλα τα παραπάνω p -values είναι μικρότερα του 0.05 που είναι το επίπεδο σημαντικότητας που χρησιμοποιούμε, άρα είναι στατιστικά σημαντικά για το μοντέλο. Επίσης, βλέπουμε ότι το residual standard error είναι αρκετά μικρό, πιο συγκεκριμένα είναι ίσο με 0.3705151, το οποίο σημαίνει ότι τα δεδομένα μας προσαρμόζονται καλύτερα στο συγκεκριμένο μοντέλο συγκριτικά με το προηγούμενο όπου είχε residual standard error ίσο με 0.5592301. Άρα ο διαχωρισμός που έγινε στα ζιζανιοκτόνα ήταν χρήσιμος.

```

> summary(gLht(S.alba.LL.4.2))

Simultaneous Tests for General Linear Hypotheses

Fit: drm(formula = DryMatter ~ Dose, curveid = Herbicide, pmodels = list(~Herbicide -
1, ~1, ~1, ~Herbicide - 1), data = S.alba, fct = LL.4(names = c("Slope",
"Lower Limit", "Upper Limit", "ED50")))

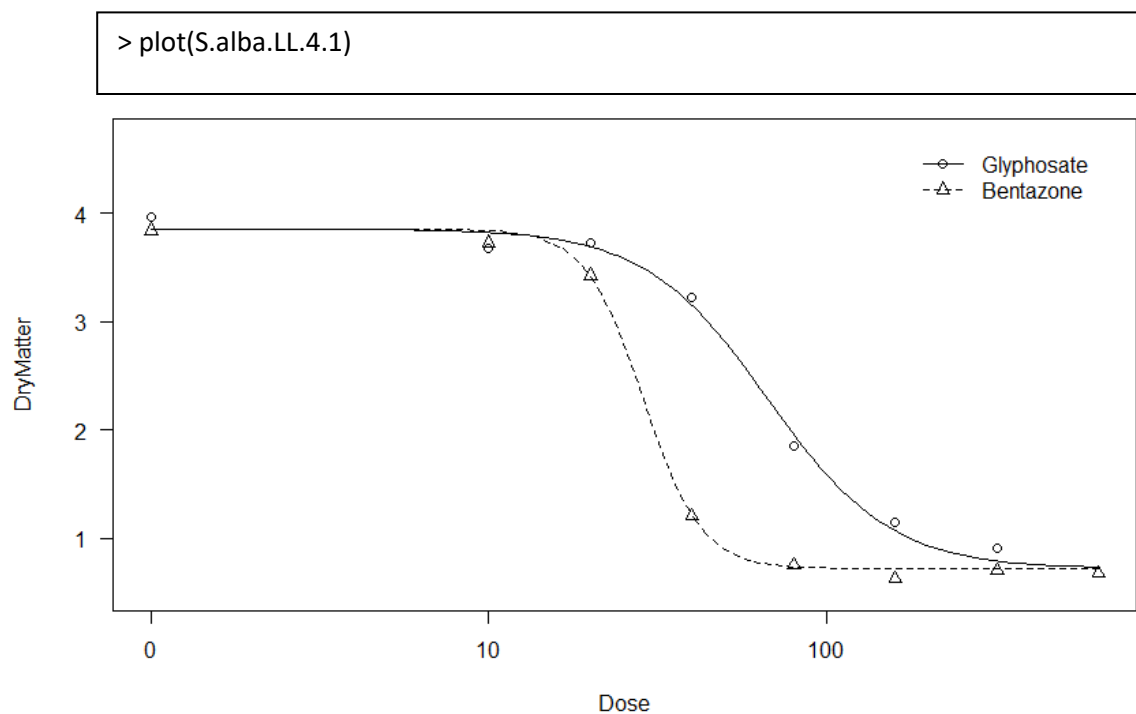
Linear Hypotheses:

              Estimate Std. Error z value Pr(>|z|)
Slope:HerbicideBentazone == 0  5.04614  1.04013  4.851 <1e-05 ***
Slope:HerbicideGlyphosate == 0  2.39022  0.49596  4.819 <1e-05 ***
Lower Limit:(Intercept) == 0  0.71656  0.08925  8.029 <1e-05 ***
Upper Limit:(Intercept) == 0  3.85486  0.07626 50.552 <1e-05 ***
ED50:HerbicideBentazone == 0  28.63235  2.03810 14.049 <1e-05 ***
ED50:HerbicideGlyphosate == 0  66.89054  5.96882 11.207 <1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```

Από τα παραπάνω αποτελέσματα, παρατηρούμε ότι τα p-values των κλίσεων, του κατώτατου ορίου, του ανώτατου ορίου και των ED_{50} είναι αρκετά κοντά στο 0. Συνεπώς, η υπόθεση της γραμμικότητας απορρίπτεται σε επίπεδο 0.05.

✚ Στην συνέχεια, φτιάχνουμε το γράφημα για το παραπάνω μοντέλο.



Σχήμα 3.6 Καμπύλες Δόσης-Απόκρισης ανά Ζιζανιοκτόνο

Τώρα που διαχωρίστηκαν τα ζιζανιοκτόνα, μπορούμε να μιλήσουμε για την σχέση που έχει το καθένα ξεχωριστά με την απόκριση. Και στα δύο υπάρχει καθοδική τάση στην σχέση δόση-απόκριση, δηλαδή όσο αυξάνεται η δόση μειώνεται η απόκριση, αλλά με διαφορετικό τρόπο στο καθένα. Οι δύο καμπύλες διαφοροποιούνται για δόση από 10 έως 150. Για παράδειγμα για δόση=50, βλέπουμε ότι το Bentazone πετυχαίνει απόκριση περίπου ίση με 2, ενώ το Glyphosate πετυχαίνει απόκριση περίπου ίση με 4. Άρα μπορούμε να ισχυριστούμε ότι για δόση από 10 έως 150, το Glyphosate έχει την διπλάσια απόκριση από το Bentazone, άρα είναι πιο αποτελεσματικό. Παρ'όλα αυτά, τον συγκεκριμένο ισχυρισμό θα τον δούμε και παρακάτω.

✚ Στην συνέχεια, συγκρίνουμε τα δύο παραπάνω μοντέλα με την εντολή anova, δηλαδή με το F-test.

```
> anova(S.alba.LL.4.1, S.alba.LL.4.2)
```

```
> anova(S.alba.LL.4.1, S.alba.LL.4.2)

1st model
 fct:    LL.4()
 pmodels: 1 (for all parameters)
2nd model
 fct:    LL.4()
 pmodels: ~Herbicide - 1, ~1, ~1, ~Herbicide - 1

ANOVA table

      ModelDf      RSS Df F value p value
1st model      64 20.0153
2nd model      62  8.5114  2  41.899  0.000
```

Από τα παραπάνω αποτελέσματα, βλέπουμε ότι το p-value είναι ίσο με 0.000, δηλαδή μικρότερο από το επίπεδο σημαντικότητας 0.05, συνεπώς απορρίπτουμε την μηδενική υπόθεση ότι οι δύο κλίσεις και οι τιμές των ED_{50} για τα ζιζανιοκτόνα Glyphosate και Bentazone είναι ίσα. Άρα οι δύο καμπύλες δεν είναι πανομοιότυπες, το οποίο το είδαμε και από το γράφημα παραπάνω.

- ✚ Για να βρούμε την εκτιμώμενη σχετική ισχύ με βάση τα δύο ED_{50} , χρησιμοποιούμε την συνάρτηση EDcomp.

```
> EDcomp(S.alba.LL.4.2, c(10, 50, 50), interval = "delta")
```

```
> EDcomp(S.alba.LL.4.2, c(10, 50, 50), interval = "delta")

Estimated ratios of effect doses

      Estimate Lower Upper
Bentazone/Glyphosate:10/50 0.27694 0.19936 0.35453
Bentazone/Glyphosate:10/50 0.27694 0.19936 0.35453
Bentazone/Glyphosate:50/50 0.42805 0.34026 0.51583
```

Από τα παραπάνω αποτελέσματα μπορούμε να συμπεράνουμε ότι το ζιζανιοκτόνο Bentazone είναι περίπου μισές φορές πιο ισχυρό συγκριτικά με το ζιζανιοκτόνο Glyphosate. Αυτό μπορούμε να το δούμε από την εκτίμηση Bentazone/Glyphosate η οποία είναι ίση με 0.42805. Άρα το συμπέρασμα που βγάλαμε για τα δύο ζιζανιοκτόνα από το γράφημα παραπάνω,

ενισχύεται με την εκτίμηση που κάναμε μέσω της συνάρτησης EDcomp. Συνεπώς, το Glyphosate είναι πιο αποτελεσματικό από το Bentazone.

- ✚ Η σύγκριση των κλίσεων των δύο ζιζανιοκτόνων μπορεί να γίνει και με άλλο τρόπο εκτός του F-test που είδαμε παραπάνω. Πιο συγκεκριμένα μπορεί να γίνει με Z-test, χρησιμοποιώντας την συνάρτηση compParm. Γενικά, το F-test (anova) χρησιμοποιείται για να ελεγχθεί εάν υπάρχουν σημαντικές διαφορές μεταξύ των μέσων δύο ή περισσότερων ομάδων, ενώ το Z-test χρησιμοποιείται για να ελεγχθεί εάν ένας μέσος όρος δείγματος διαφέρει σημαντικά από έναν μέσο όρο πληθυσμού.

```
> compParm(S.alba.LL.4.2, "b", "-")
```

```
> compParm(s.alba.LL.4.2, "b", "-")
```

```
Comparison of parameter 'b'
```

```
Estimate Std. Error t-value p-value
HerbicideBentazone-HerbicideGlyphosate 2.6559 1.0689 2.4847 0.01568 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Από τα αποτελέσματα παραπάνω βλέπουμε ότι, η τιμή του p-value είναι διαφορετική συγκριτικά με πριν, δηλαδή με το F-test είχαμε p-value=0.00 ενώ με το Z-test έχουμε 0.01568. Παρ'όλα αυτά το p-value εξακολουθεί να είναι σημαντικό σε επίπεδο σημαντικότητας 0.05. Άρα ακόμα και με διαφορετικό τρόπο προσέγγισης καταλήγουμε στα ίδια αποτελέσματα.

3.4 Διωνυμική Απόκριση : Γενικευμένη Μη Γραμμική Παλινδρόμηση

Στο συγκεκριμένο παράδειγμα πρόκειται να γίνει ανάλυση ενός μοντέλου δόσης-απόκρισης, αυτή την φορά έχοντας διακριτή απόκριση και όχι συνεχής όπως στα δύο προηγούμενα παραδείγματα, πιο συγκεκριμένα η απόκριση είναι δίτιμη. Τα δεδομένα που έχουν χρησιμοποιηθεί στο συγκεκριμένο παράδειγμα είναι ήδη περασμένα στο πακέτο drc της R.

Στην R, τα συγκεκριμένα δεδομένα έχουν την ονομασία earthworms/ γαιοσκώληκες και προέρχονται από την Nina Cedergreen, η οποία είναι καθηγήτρια στο Πανεπιστήμιο της Κοπεγχάγης στο τμήμα Επιστημών Ζωής. Τα πειράματα πάνω σε γαιοσκώληκες χρησιμοποιούνται συχνά για την αξιολόγηση της τοξικότητας στην οικοτοξικολογία. Οικοτοξικολογία είναι η μελέτη των επιδράσεων των τοξικών χημικών ουσιών στους βιολογικούς οργανισμούς, ειδικά σε επίπεδο πληθυσμού, κοινότητας, οικοσυστήματος και βιόσφαιρας. Οπότε, το συγκεκριμένο σύνολο δεδομένων περιέχει τον αριθμό των

γαιοσκωλήκων που έχουν παραμείνει σε ένα δοχείο, το οποίο ήταν μολυσμένο με μια τοξική ουσία (δεν αποκαλύπτεται) σε διάφορες δόσεις. Δηλαδή τον αριθμό των γαιοσκωλήκων που δεν μεταναστεύουν στο γειτονικό μη μολυσμένο δοχείο. Πρόκειται για ένα πλαίσιο δεδομένων που αποτελείται από 35 παρατηρήσεις στις ακόλουθες τρεις μεταβλητές : Το αριθμητικό διάνυσμα Dose, όπου περιέχει τις χορηγούμενες δόσεις, το αριθμητικό διάνυσμα Number, όπου περιέχει το πλήθος των γαιοσκωλήκων που έχουν απομείνει στο δοχείο και το αριθμητικό διάνυσμα Total, όπου περιέχει τον συνολικό αριθμό των γαιοσκωλήκων που έχουν τοποθετηθεί στα δοχεία.

Για δόση ίση με 0, αναμένουμε περίπου οι μισοί γαιοσκώληκες να κατανεμηθούν στο μολυσμένο δοχείο και οι άλλοι μισοί στο μη μολυσμένο δοχείο. Για τον λόγο αυτό, θα προσαρμόσουμε ένα log-logistic μοντέλο τριών παραμέτρων σε διωνυμικά δεδομένα, το οποίο σύμφωνα με τον Finney (1971) είναι ένα παράδειγμα γενικευμένου μη γραμμικού μοντέλου. Πιο συγκεκριμένα, πρόκειται για ένα μοντέλο λογιστικής παλινδρόμησης, όπου η συνάρτηση σύνδεσης του εξαρτάται από τις παραμέτρους.

Τα δεδομένα τα οποία αναλύονται στην συγκεκριμένη άσκηση βρίσκονται στο παράρτημα 4.

- ✚ Όπως και στο προηγούμενο παράδειγμα, φορτώνουμε την βιβλιοθήκη drc η οποία είναι κατάλληλη για δεδομένα δόσης-απόκρισης. Τρέχουμε το log-logistic μοντέλο τριών παραμέτρων και βρίσκουμε κάποια συνοπτικά στοιχεία με το summary. Όπως αναφέραμε παραπάνω, πρόκειται για ένα μη γραμμικό μοντέλο, παρ'όλα αυτά θα ελέγξουμε την υπόθεση της γραμμικότητας για να είμαστε σίγουροι, χρησιμοποιώντας την εντολή glht, από την βιβλιοθήκη multcomp.

```
> library(drc)
> library(multcomp)
> earthworms.m1 <- drm(number/total ~ dose, weights = total, data = earthworms,
+       fct = LL.3(names = c("Slope", "Upper Limit", "ED50")), type =
+       "binomial")
> summary(earthworms.m1)
> summary(glht(earthworms.m1))
```

```
> summary(earthworms.m1)
Model fitted: Log-logistic (ED50 as parameter) with lower limit at 0 (3 parms)
Parameter estimates:
      Estimate Std. Error t-value  p-value
Slope:(Intercept)      1.505679  0.338992  4.4416 8.928e-06 ***
Upper Limit:(Intercept) 0.604929  0.085800  7.0505 1.783e-12 ***
ED50:(Intercept)       0.292428  0.083895  3.4856 0.000491 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Από τα παραπάνω αποτελέσματα βλέπουμε ότι, οι τρεις παράμετροι που χρησιμοποιήθηκαν είναι η κλίση, το ανώτατο όριο και το ED_{50} . Οπότε για αυτές τις παραμέτρους βλέπουμε την εκτίμηση, το τυπικό σφάλμα, την τιμή t και το p -value. Τα p -values της κλίσης, του ανώτατου ορίου και του ED_{50} είναι μικρότερα του 0.05 που είναι το επίπεδο σημαντικότητας που χρησιμοποιούμε, άρα είναι στατιστικά σημαντικά για το μοντέλο. Σημειώνεται ότι το residual standard error είναι διαθέσιμο μόνο όταν έχουμε συνεχή δεδομένα δόσης-απόκρισης, το οποίο δεν ισχύει στο συγκεκριμένο παράδειγμα όπου έχουμε διακριτά δεδομένα. Άρα δεν μπορούμε να πούμε αν το συγκεκριμένο μοντέλο προσαρμόζεται καλά στα δεδομένα.

```
> summary(glht(earthworms.m1))
Simultaneous Tests for General Linear Hypotheses

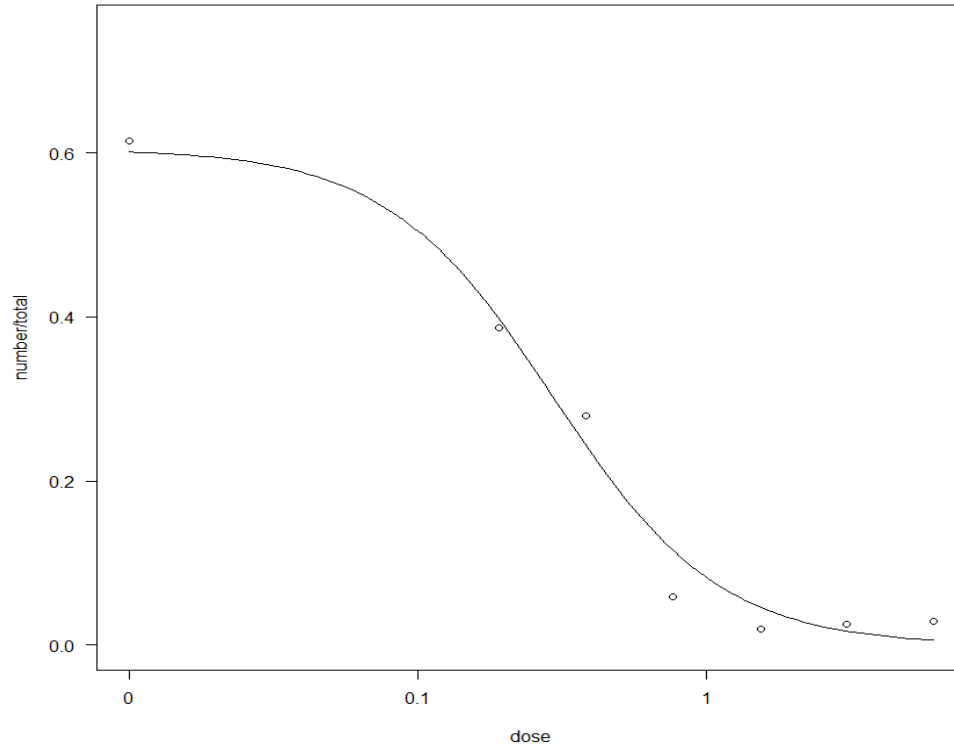
Fit: drm(formula = number/total ~ dose, weights = total, data = earthworms,
  fct = LL.3(names = c("Slope", "Upper Limit",
    "ED50")), type = "binomial")

Linear Hypotheses:
      Estimate Std. Error z value Pr(>|z|)
Slope:(Intercept) == 0      1.5057   0.3390  4.442 < 0.001 ***
Upper Limit:(Intercept) == 0  0.6049   0.0858  7.050 < 0.001 ***
ED50:(Intercept) == 0       0.2924   0.0839  3.486 0.00126 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Αντίστοιχα από τα παραπάνω αποτελέσματα μπορούμε να ελέγξουμε την υπόθεση της γραμμικότητας. Παρατηρούμε ότι τα p -values της κλίσης, του ανώτατου ορίου και του ED_{50} είναι μικρότερα του 0.05 που είναι το επίπεδο σημαντικότητας που χρησιμοποιούμε, άρα η υπόθεση της γραμμικότητας απορρίπτεται, όπως και αναμέναμε.

✚ Στην συνέχεια, φτιάχνουμε το γράφημα για το παραπάνω μοντέλο.

```
> plot(earthworms.m1)
```



Σχήμα 3.7 Καμπύλη Δόσης-Απόκρισης Γαιοσκωλήκων Που Παρέμειναν Στο Μολυσμένο Δοχείο

Από το παραπάνω σχήμα συμπεραίνουμε ότι πρόκειται για μια μονότονη και σιγμοειδής καμπύλη. Επίσης, βλέπουμε ότι όσο αυξάνεται η δόση, τόσο μειώνεται η απόκριση, το οποίο πρακτικά σημαίνει ότι όσο αυξάνεται η τοξική ουσία στο μολυσμένο δοχείο, τόσο μειώνεται ο αριθμός των γαιοσκωλήκων στο συγκεκριμένο δοχείο. Να σημειωθεί ότι η απόκριση έχει πιο έντονη καθοδική τάση για δόση από 0.1 έως 1.

- ✚ Ίσως, ακόμα καλύτερα θα ήταν να προσαρμόζαμε ένα log-logistic μοντέλο, στο οποίο το ανώτατο όριο δεν εκτιμάται αλλά αντ' αυτού είναι ίσο με 0.5. Αυτό γίνεται στις παρακάτω γραμμές.

```
> earthworms.m2 <- drm(number/total ~ dose, weights = total, data = earthworms,
+      fct = LL.3(fixed = c(NA, 0.5, NA)), type = "binomial")
> summary(earthworms.m2)
```

```
> summary(earthworms.m2)
```

```
Model fitted: Log-logistic (ED50 as parameter) with lower limit at 0 (2 parms)
```

```
Parameter estimates:
```

	Estimate	Std. Error	t-value	p-value	
b:(Intercept)	1.646689	0.376494	4.3737	1.221e-05	***
e:(Intercept)	0.377269	0.076785	4.9133	8.956e-07	***

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

Συγκρίνοντας αυτό το μοντέλο με το προηγούμενο, βλέπουμε ότι η εκτίμηση για το b(Slope) αυξήθηκε από 1.505679 σε 1.646689, όπως και για το e(ED_{50}) όπου αυξήθηκε από 0.2924 σε 0.377269. Όσον αφορά το τυπικό σφάλμα, για το b(Slope) αυξήθηκε από 0.338992 σε 0.376494, ενώ για το e(ED_{50}) μειώθηκε από 0.0839 σε 0.076785. Το t-value του b(Slope) μειώθηκε από 4.442 σε 4.3737, ενώ το t-value του e(ED_{50}) αυξήθηκε από 3.486 σε 4.9133. Τα p-values της κλίσης και του ED_{50} συνεχίζουν να είναι στατιστικά σημαντικά και σε αυτό το μοντέλο, δεδομένου ότι είναι μικρότερα του 0.05 που είναι το επίπεδο σημαντικότητας που χρησιμοποιούμε. Συνεπώς, με τον καθορισμό του ανώτατου ορίου σε 0.5, παρατηρούμε ότι υπάρχει ένα πολύ μικρό κέρδος στην ακρίβεια για την παράμετρο e, που αντιστοιχεί στο ED_{50} , δεδομένου ότι μειώθηκε το τυπικό σφάλμα. Αντίθετα, η ακρίβεια της παραμέτρου b, όπου αντιστοιχεί στην κλίση μειώνεται, δεδομένου ότι το τυπικό σφάλμα αυξήθηκε.

- ✚ Όπως προαναφέραμε στην αρχή του παραδείγματος, το residual standard error το οποίο μας δείχνει αν προσαρμόζονται καλά τα δεδομένα στο μοντέλο, είναι διαθέσιμο μόνο όταν έχουμε συνεχή δεδομένα δόσης-απόκρισης, το οποίο δεν ισχύει στο συγκεκριμένο παράδειγμα όπου έχουμε διακριτά δεδομένα. Συνεπώς, για να κάνουμε αυτό τον έλεγχο και για τα δύο μοντέλα που προσαρμόσαμε παραπάνω, με τις τρεις και τις δύο παραμέτρους θα χρησιμοποιήσουμε την συνάρτηση modelFit. Η συνάρτηση modelFit χρησιμοποιεί το F-test για τον έλεγχο.

```
> modelFit(earthworms.m1)
Goodness-of-fit test

              Df Chisq value p value
DRC model    32      43.13  0.0905
```

Από τα παραπάνω αποτελέσματα, που αναφέρονται στο πρώτο μοντέλο δηλαδή αυτό με τις τρεις παραμέτρους, βλέπουμε ότι σε επίπεδο σημαντικότητας 0.05 τα δεδομένα μας δεν προσαρμόζονται ιδιαίτερα καλά, δεδομένου ότι το p-value είναι ίσο με 0.0905.

```
> modelFit(earthworms.m2)
Goodness-of-fit test

              Df Chisq value p value
DRC model    33      49.859  0.0301
```

Από τα παραπάνω αποτελέσματα, που αναφέρονται στο δεύτερο μοντέλο δηλαδή αυτό στο οποίο το ανώτατο όριο δεν εκτιμάται αλλά αντ' αυτού είναι ίσο με 0.5, βλέπουμε ότι σε επίπεδο σημαντικότητας 0.05 τα δεδομένα μας προσαρμόζονται καλά, δεδομένου ότι το p-value είναι ίσο με 0.0301.

- ✚ Τέλος, θα συγκρίνουμε τα δύο παραπάνω μοντέλα με την χρήση της συνάρτησης anova, όπου γίνεται με βάση τον έλεγχο X^2 .

```
> anova(earthworms.m1, earthworms.m2)

1st model
fct: LL.3(names = c("slope", "Upper Limit", "ED50"))
2nd model
fct: LL.3(fixed = c(NA, 0.5, NA))

ANOVA-like table

      ModelDf  Loglik Df LR value p value
1st model     3 -36.155
2nd model     2 -36.907  1  1.5044  0.22
```

Από τα παραπάνω αποτελέσματα, βλέπουμε ότι το p-value είναι ίσο με 0.22, το οποίο είναι μεγαλύτερο από το επίπεδο σημαντικότητας 0.05, άρα συμπεραίνουμε ότι τα δύο μοντέλα δεν

διαφέρουν σημαντικά. Παρ'όλα αυτά, επειδή είδαμε από το `modelFit` ότι τα δεδομένα μας προσαρμόζονται καλύτερα στο δεύτερο μοντέλο από ότι στο πρώτο, καταλήγουμε στο ότι το δεύτερο μοντέλο είναι καλύτερη επιλογή για το συγκεκριμένο παράδειγμα.

ΣΥΝΟΨΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Όπως έχουμε αναφέρει, ένας από τους κυριότερους στόχους της παρούσας διπλωματικής εργασίας είναι η ανάλυση διαφορετικών μοντέλων δόσης-απόκρισης με την χρήση πραγματικών δεδομένων όπως και κάναμε, με σκοπό την εξαγωγή σημαντικών συμπερασμάτων. Συνεπώς, σε αυτή την ενότητα επικεντρωνόμαστε στα κύρια συμπεράσματα στα οποία καταλήξαμε για κάθε ένα από τα παραδείγματα παραπάνω.

Διωνυμική Απόκριση : Σύγκριση Μοντέλων logit, probit και cloglog

Στο συγκεκριμένο παράδειγμα (3.1) πραγματοποιήθηκε σύγκριση των μοντέλων logit, probit και cloglog, χρησιμοποιώντας πραγματικά δεδομένα από τη μελέτη των Jeske et.al (2009) σχετικά με την τοξικότητα του σεληνίου σε μύγες. Ο στόχος ήταν να αναλυθούν οι σχέσεις δόσης-απόκρισης και να προσδιοριστεί ποιο μοντέλο εφαρμόζεται καλύτερα στα δεδομένα.

Τα p-values για τη μεταβλητή συγκέντρωσης ήταν στατιστικά σημαντικά ($<2e-16$) για όλα τα μοντέλα, επιβεβαιώνοντας την επίδραση της συγκέντρωσης του σεληνίου στη θνησιμότητα των μυγών. Λαμβάνοντας όμως υπόψη τα συνολικά αποτελέσματα, συμπεριλαμβανομένης της χαμηλότερης απόκλισης του καταλοίπου (10.031) και του ανώτερου κριτηρίου Akaike (43.251), επιλέξαμε το μοντέλο logit ως το πλέον κατάλληλο για το συγκεκριμένο παράδειγμα. Παρέχει μια πιο ακριβή αναπαράσταση της σχέσης δόσης-απόκρισης μεταξύ της συγκέντρωσης σεληνίου και της θνησιμότητας των μυγών.

Συνεχής Απόκριση : Μία Καμπύλη Δόσης-Απόκρισης

Στο συγκεκριμένο παράδειγμα (3.2) επικεντρωθήκαμε στην ανάλυση μιας συνεχούς απόκρισης, πιο συγκεκριμένα της επίδρασης του φερουλικού οξέος στο μήκος της ρίζας της σίκαλης, χρησιμοποιώντας πραγματικά δεδομένα από την μελέτη των Inderjit, J. C. Streibig και M. Olofsdotter (2002). Στόχος αυτού του παραδείγματος ήταν ο προσδιορισμός της σχέσης δόσης-απόκρισης, ο έλεγχος της γραμμικότητας και προσαρμογής του μοντέλου και η εκτίμηση μερικών επιπέδων απόκρισης όπως ED_{50} .

Από το γράφημα και την ανάλυση του μοντέλου, συμπεράναμε ότι όσο αυξάνεται η συγκέντρωση του φερουλικού οξέος, τόσο μειώνεται το μήκος της ρίζας της σίκαλης. Η υπόθεση της γραμμικότητας απορρίπτεται σε επίπεδο σημαντικότητας 0.05, δείχνοντας ότι η καμπύλη δόσης-απόκρισης δεν είναι γραμμική. Αυτό υποδεικνύει ότι η σχέση μεταξύ της δόσης του φερουλικού οξέος και του μήκους της ρίζας δεν είναι γραμμική. Υπολογίζοντας τα επίπεδα ED_5 , ED_{10} και ED_{50} είδαμε την δόση του φερουλικού οξέος που απαιτείται για να επιτευχθούν συγκεκριμένες αλλαγές στο μήκος της ρίζας. Τέλος, Το γενικευμένο γραμμικό

μοντέλο που χρησιμοποιήθηκε προσαρμόζεται αρκετά καλά στα δεδομένα, όπως φαίνεται από το χαμηλό τυπικό σφάλμα (0.5196256).

Συνεχής Απόκριση : Δύο Καμπύλες Δόσης-Απόκρισης

Στο συγκεκριμένο παράδειγμα (3.3) μελετήσαμε την επίδραση δύο ζιζανιοκτόνων στην λευκή μουστάρδα, χρησιμοποιώντας πραγματικά δεδομένα από τους Christensen, M. G. και Teicher, H. B. και Streibig, J. C. (2003). Στόχος αυτού του παραδείγματος ήταν η αξιολόγηση του μοντέλου χωρίς τον διαχωρισμό των ζιζανιοκτόνων και με τον διαχωρισμό των ζιζανιοκτόνων, ο έλεγχος της γραμμικότητας του μοντέλου και η σύγκριση της αποτελεσματικότητας των δύο ζιζανιοκτόνων στην λευκή μουστάρδα.

Γενικά παρατηρήσαμε ότι όσο αυξάνεται η δόση των ζιζανιοκτόνων τόσο μειώνεται η απόκριση, δηλαδή υπάρχει καθοδική τάση στην σχέση ζιζανιοκτόνων-λευκής μουστάρδας. Παρ'όλα αυτά το Glyphosate είναι πιο αποτελεσματικό από το Bentazone, καθώς σε κάποιες δόσεις πετυχαίνει σχεδόν την διπλάσια απόκριση από το Bentazone. Η υπόθεση της γραμμικότητας απορρίπτεται για τα δύο ζιζανιοκτόνα, Glyphosate και Bentazone, σε επίπεδο σημαντικότητας 0.05. Αυτό σημαίνει ότι η αντίδραση της λευκής μουστάρδας στην αύξηση της δόσης των ζιζανιοκτόνων δεν είναι γραμμική. Επιπλέον, ο διαχωρισμός των ζιζανιοκτόνων βελτίωσε την ακρίβεια της ανάλυσης και των συμπερασμάτων που μπορούν να εξαχθούν.

Διωνυμική Απόκριση : Γενικευμένη Μη Γραμμική Παλινδρόμηση

Στο συγκεκριμένο παράδειγμα (3.4) εφαρμόσαμε ένα μοντέλο δόσης-απόκρισης σε διωνυμικά δεδομένα τα οποία σχετίζονται με την επίδραση τοξικής ουσίας σε γαιοσκώληκες, χρησιμοποιώντας πραγματικά δεδομένα που προέρχονται από έρευνα που πραγματοποιήθηκε από την καθηγήτρια Nina Cedergreen. Στόχος αυτού του παραδείγματος ήταν η αξιολόγηση της τοξικότητας και των επιπτώσεών της στους γαιοσκώληκες, ο έλεγχος γραμμικότητας του μοντέλου και η σύγκριση με το μοντέλο με προκαθορισμένο ανώτατο όριο.

Συνεπώς, επιβεβαιώθηκε η υπόθεση ότι το μοντέλο δόσης-απόκρισης δεν είναι γραμμικό, με τα p-values να υποδεικνύουν την απόρριψη της υπόθεσης της γραμμικότητας. Το γράφημα της καμπύλης δόσης-απόκρισης έδειξε μια σιγμοειδής καμπύλη, όπου η απόκριση μειώνεται όσο αυξάνεται η δόση της τοξικής ουσίας. Οι αναλύσεις μοντέλων και οι συγκρίσεις αποδείχθηκαν στατιστικά σημαντικές, αν και τα δύο μοντέλα δεν διαφέρουν σημαντικά μεταξύ τους, το δεύτερο μοντέλο (με προκαθορισμένο ανώτατο όριο) παρουσίασε καλύτερη προσαρμογή στα δεδομένα.

Βιβλιογραφία

- Altshuler, B. (1981) Modeling of Dose-Response Relationships. *Environ Health Perspect.* Dec;42:23-27.
- Bates, D.M., Watts, D.G. (1988) Nonlinear Regression Analysis and its Applications. John Wiley & Sons, Inc. N.Y.
- Christensen, M.G., Teicher, H.B and Streibig, J.C. (2003) Linking fluorescence induction curve and biomass in herbicide screenin. *Pest Management Science*, Vol. 59, No. 12, 2003, p. 1303-1310.
- Crump, K.S., Hoel, D.G., Langley, C.H., Peto, R.(1976) Fundamental Carcinogenic Processes and Their Implications for Low Dose Risk Assessment.*Cancer Res.* 1976 Sep;36(9 pt.1):2973-9.
- Demidenko, E. (2004) Mixed Models Theory and Applications. John Wiley & Sons, Inc.N.Y.
- Ding, S. and Sachs, F. (1999) Single Channel Properties of P2X2 Purinoceptors. *J Gen Physiol.* May;113(5):695-720.
- Dobson, A.J. and Barnett, A.G. (2018) An Introduction to Generalized Linear Models. 4th Ed. Chapman and Hall. N.Y.
- Finney, D.J. (1971) Probit Analysis. 3rd Ed. Cambridge University Press. N.Y.
- Hamilton, M.A., Russo, R.C. and Thurston, R.V. (1977) Trimmed Spearman-Kärber method for estimating median lethal concentrations in toxicity bioassays. *Environ. Sci. Technol.* 11(7) 714–719.
- Inderjit, Streibig, J. C. and Olofsdotter, M. (2002) Joint action of phenolic acid mixtures and its significance in allelopathy research. *Physiol. Plant Mar*;114(3):422-428.
- Jeske, D.R., Xu, H.K., Blessinger, T., Jensen, P. and Trumble, J. (2009) "Testing for the equality of EC_{50} values in the presence of unequal slopes with application to toxicity of selenium types". *Journal of Agricultural, Biological, and Environmental Statistics.* 14, 469–483.
- Lin, D., Shkedy, Z., Yekutieli, D., Amaratunga, D., Bijmens, L. (2012) Modeling Dose-Response Microarray Data in Early Drug Development Experiments Using R. Springer N.Y.
- Martin, M. (2009) Benchmark Dose Software (BMDS). User's Manual, 2.0. Arlington, V.A.
- McCullagh, P. and Nelder, J.A. (1983) Generalized Linear Models. Chapman and Hall. N.Y.

Pramana, S. (2011) "Statistical Methods for Microarray Experiments: Analysis Dose-response Studies and Software Development in R". Ph.D. Thesis Univ. of Hasselt Sweden.

Seber. G. and Wild, C.J. (2003). Nonlinear Regression. John Willey and Sons. N.Y.

Veroli, D.G.Y., Fornari, C., Goldlust, I., Mills, G., Koh, S.B., Bramhall, J.L., Richards, F.M., Jodrell, Duncan, I.J. (1 October 2015) An automated fitting procedure and software for dose-response curves with multiphasic features. *Sci Rep.* 5: 14701.

Παράρτημα 1

Στο συγκεκριμένο παράρτημα παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν στο παράδειγμα 3.1.

Πρόκειται για ένα πλαίσιο δεδομένων που αποτελείται από 8 παρατηρήσεις στις ακόλουθες τέσσερις μεταβλητές : Την κατηγορική μεταβλητή TYPE, όπου είναι ο τύπος του σεληνίου, δηλαδή η σεληνομεθειονίνη. Το αριθμητικό διάνυσμα CONC, όπου περιλαμβάνει την συγκέντρωση του σεληνίου. Το αριθμητικό διάνυσμα NS, όπου περιλαμβάνει τον αριθμό των μυγών πριν την έκθεση τους στο σελήνιο και το αριθμητικό διάνυσμα ND , όπου περιλαμβάνει τον αριθμό των μυγών που απεβίωσαν μετά την έκθεση τους στο σελήνιο.

TYPE	CONC	NS	ND
T3	0	100	3
T3	5	106	6
T3	25	122	11
T3	50	145	22
T3	100	127	31
T3	200	140	105
T3	400	172	166
T3	800	188	188

Παράρτημα 2

Στο συγκεκριμένο παράρτημα παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν στο παράδειγμα 3.2.

Πρόκειται για ένα πλαίσιο δεδομένων με 24 παρατηρήσεις στις ακόλουθες 2 μεταβλητές : Rootl είναι ένα αριθμητικό διάνυσμα που περιλαμβάνει διάφορα μήκοι ρίζας και Conc είναι ένα αριθμητικό διάνυσμα που περιλαμβάνει τις συγκεντρώσεις του φερουλικού οξέος.

Rootl	Conc
7.5800	0.00
8.0000	0.00
8.3285714	0.00
7.2500000	0.00
7.3750000	0.00
7.9625000	0.00
8.3555556	0.94
6.9142857	0.94
7.7500000	0.94
6.8714286	1.88
6.4500000	1.88
5.9222222	1.88
1.9250000	3.75
2.8857143	3.75
4.2333333	3.75
1.1875000	7.50
0.8571429	7.50
1.0571429	7.50
0.6875000	15.00
0.5250000	15.00
0.8250000	15.00
0.2500000	30.00
0.2200000	30.00
0.4400000	30.00

Παράρτημα 3

Στο συγκεκριμένο παράρτημα παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν στο παράδειγμα 3.3.

Πρόκειται για ένα πλαίσιο δεδομένων, όπου αποτελείται από 68 παρατηρήσεις στις ακόλουθες τρεις μεταβλητές : Το αριθμητικό διάλυμα Dose, όπου περιέχει την δόση σε g/ha, την κατηγορική μεταβλητή Herbicide όπου αποτελείται από τα επίπεδα Glyphosate και Bentazone και τέλος από το αριθμητικό διάλυμα DryMatter όπου περιέχει την απόκριση σε g/rot.

Dose	Herbicide	DryMatter
0	Glyphosate	4.7
0	Glyphosate	4.6
0	Glyphosate	4.1
0	Glyphosate	4.4
0	Glyphosate	3.2
0	Glyphosate	3.0
0	Glyphosate	3.8
0	Glyphosate	3.9
10	Glyphosate	3.8
10	Glyphosate	3.8
10	Glyphosate	3.8
10	Glyphosate	3.3
20	Glyphosate	4.1
20	Glyphosate	3.5
20	Glyphosate	4.1
20	Glyphosate	3.2
40	Glyphosate	2.8
40	Glyphosate	3.0
40	Glyphosate	3.7
40	Glyphosate	3.4
80	Glyphosate	1.8
80	Glyphosate	1.6
80	Glyphosate	2.3
80	Glyphosate	1.7
160	Glyphosate	1.1
160	Glyphosate	1.3
160	Glyphosate	1.0
160	Glyphosate	1.2
320	Glyphosate	0.9
320	Glyphosate	0.9

320	Glyphosate	0.9
320	Glyphosate	0.9
0	Bentazone	4.1
0	Bentazone	3.4
0	Bentazone	2.6
0	Bentazone	3.5
0	Bentazone	4.3
0	Bentazone	4.2
0	Bentazone	4.1
0	Bentazone	4.5
10	Bentazone	3.5
10	Bentazone	3.8
10	Bentazone	3.9
10	Bentazone	3.7
20	Bentazone	3.2
20	Bentazone	3.5
20	Bentazone	3.0
20	Bentazone	4.0
40	Bentazone	1.1
40	Bentazone	1.4
40	Bentazone	1.4
40	Bentazone	0.9
80	Bentazone	0.9
80	Bentazone	0.6
80	Bentazone	0.7
80	Bentazone	0.8
160	Bentazone	0.7
160	Bentazone	0.7
160	Bentazone	0.5
160	Bentazone	0.6
320	Bentazone	0.8
320	Bentazone	0.5
320	Bentazone	0.7
320	Bentazone	0.8
640	Bentazone	0.6
640	Bentazone	0.7
640	Bentazone	0.6
640	Bentazone	0.8

Παράρτημα 4

Στο συγκεκριμένο παράρτημα παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν στο παράδειγμα 3.4.

Πρόκειται για ένα πλαίσιο δεδομένων που αποτελείται από 35 παρατηρήσεις στις ακόλουθες τρεις μεταβλητές : Το αριθμητικό διάνυσμα Dose, όπου περιέχει τις χορηγούμενες δόσεις, το αριθμητικό διάνυσμα Number, όπου περιέχει το πλήθος των γαισκωλήκων που έχουν απομείνει στο μολυσμένο δοχείο και το αριθμητικό διάνυσμα Total, όπου περιέχει τον συνολικό αριθμό των γαισκωλήκων που έχουν τοποθετηθεί στα δοχεία.

Dose	Number	Total
0	3	5
0	3	5
0	3	4
0	5	8
0	4	8
19	4	11
19	4	9
19	2	8
19	3	8
19	2	4
38	2	9
38	0	4
38	6	11
38	4	11
38	4	15
76	0	11
76	1	6
76	1	8
76	0	9
76	0	8
1.53	0	6
1.53	1	10
1.53	0	10
1.53	0	10
1.53	0	8
3.05	0	5
3.05	0	9
3.05	0	7
3.05	0	8
3.05	1	8

6.11	0	3
6.11	0	7
6.11	0	11
6.11	0	10
6.11	1	7