

Μαθηματική μοντελοποίηση και ανάλυση  
αναπτυξιακών μοντέλων στην πρόληψη  
δημιουργίας καρκινικών κυττάρων

Η Μεταπτυχιακή Εργασία Εξειδίκευσης

υποβάλλεται στην ορισθείσα

από τη Γενική Συνέλευση Ειδικής Σύνθεσης  
του Τμήματος Μηχανικών Η/Υ και Πληροφορικής  
Εξεταστική Επιτροπή

από τον

Μπουχάρα Δημήτριο

ως μέρος των υποχρεώσεων για την απόκτηση του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ  
ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ  
ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ - ΕΦΑΡΜΟΓΕΣ

Πανεπιστήμιο Ιωαννίνων

2020

Εξεταστική Επιτροπή:

- **Γεώργιος Μανής**, Αναπλ. Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων (Επιβλέπων)
- **Χριστόφορος Νίκου**, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Λυσίμαχος-Πάυλος Κόντης**, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

# ΑΦΙΕΡΩΣΗ

---

*στην μητέρα μου Χαρούλα,  
που έφυγε νωρίς αλλά υπάρχει κάθε μέρα στην ζωή μου καθώς...  
«η μάνα εν κρύο νερόν»*

# ΕΥΧΑΡΙΣΤΙΕΣ

---

Η περάτωση της παρούσας διατριβής δεν θα ήταν δυνατή χωρίς την βοήθεια και την αμέριστη συμπαράσταση ανθρώπων που βρέθηκαν γύρω μου τα τελευταία ακαδημαϊκά έτη. Πιθανώς να ξεχάσω μερικούς, ας με συγχωρέσουν.

Πάνω από όλους, θα ήθελα να ευχαριστήσω τους γονείς μου, Γεώργιο και Χαρούλα, που μου πρόσφεραν τόσο τα πνευματικά όσο και τα οικονομικά εφόδια για να ανταπεξέλθω σε όλη την διάρκεια των σπουδών μου. Την αδερφή μου Ελένη, που με πίεσε να προσπαθώ όλο και περισσότερο. Για την σαφή υποστήριξή τους δεν αρκεί η απλή έκφραση ευχαριστιών μου.

Θα ήθελα να ευχαριστήσω όλους τους συναδέλφους που γνώρισα στο μεταπτυχιακό δίπλωμα. Λίγο παραπάνω, εκείνους που συνεργαστήκαμε σε εξαμηνιαίες εργασίες. Πιο συγκεκριμένα την Ελένη, τον Κωνσταντίνο, την Ειρήνη, την Ελένη και την Μαρίκα. Υπογραμμίζω την συνεργασία μου με την Ελένη για τις υπεράριθμες ώρες ανταλλαγής απόψεων, την κριτική σκέψη και ικανότητα πάνω σε εργασίες που αναλάβαμε από κοινού.

Θα ήθελα να συμπεριλάβω στις ευχαριστίες μου τους Σούρδους φίλους μου Ιωάννη, Χρήστο και Ιωάννη για τις ατελείωτες ώρες συζητήσεων, την αμέριστη συμπαράσταση και την ψυχολογική υποστήριξη που μου έδειξαν.

Θα ήθελα ιδιαίτερα να ευχαριστήσω τους Αναπληρωτή Καθηγητή κ. Γεώργιο Μανή και Καθηγητή κ. Σπυρίδων Καρκαμπούνα για την άψογη συνεργασία τα τελευταία δύο έτη. Η τεκμηριωμένη καθοδήγηση, το γνωσιολογικό τους εύρος και η φιλική τους διάθεση ήταν καθοριστικοί παράμετροι για την ολοκλήρωση της μεταπτυχιακής μου διατριβής. Η πίστη τους στο πρόσωπό μου ήταν απόδειξη ότι ποτέ δεν είναι αργά να αλλάξεις ρότα και να κυνηγάς το πραγματικό σου πάθος και αυτό είναι μάθημα ζωής...

Τέλος, θα ήθελα να υπογραμμίσω ότι ο Καθηγητής κ. Σπυρίδων Καρκαμπούνας της Ιατρικής Σχολής του Πανεπιστημίου Ιωαννίνων συνέβαλε στην επιστημονική καθοδήγηση της μεταπτυχιακής εργασίας και αποτελεί αναπόσπαστο κομμάτι αυτής.

Το θεωρητικό υπόβαθρο, η πειραματική προσέγγιση αλλά και τα πρώτα εξαγόμενα αποτελέσματα αποτελούν μέρος της μελέτης του Καθηγητή κ. Σπυρίδωνος Καρκαμπούνα, της επιστημονικής του ομάδας και του τμήματος Φυσιολογίας της Ιατρικής Σχολής του Πανεπιστημίου Ιωαννίνων.

Τα βιολογικά δεδομένα που παρουσιάζονται στην παρούσα εργασία είναι αδημοσίευτα και τελούν υπό την πνευματική ιδιοκτησία του Καθηγητή κ. Σπυρίδωνος Καρκαμπούνα και της επιστημονικής του ομάδας σύμφωνα με τους κείμενους νόμους.

Στενοχωρήθηκα ιδιαίτερα όταν πληροφορήθηκα ότι ο Καθηγητής κ. Σπυρίδων Καρκαμπούνας δεν θα μπορούσε να είναι μέρος της εξεταστικής επιτροπής λόγω της πρόσφατης συνταξιοδότησής του.

# ΠΕΡΙΕΧΟΜΕΝΑ

---

Κατάλογος Σχημάτων	iv
Κατάλογος Πινάκων	vii
Κατάλογος Αλγορίθμων	x
Γλωσσάρι	xi
Περίληψη	xii
Extended Abstract	xiv
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Στόχοι . . . . .	1
1.2 Δομή της Διατριβής . . . . .	2
<b>2 Βιολογικά Δεδομένα</b>	<b>4</b>
2.1 Η ιστορία του ποντικού NMRI . . . . .	4
2.2 Στοιχεία για τον ποντικό NMRI . . . . .	6
2.2.1 Ταξινόμηση και χρησιμότητα . . . . .	6
2.2.2 Εξωτερική Μορφολογία . . . . .	7
2.2.3 Αντιστοιχία με τον άνθρωπο . . . . .	7
2.2.4 Διμορφισμός ως προς το φύλο εν σχέσει με την ανάπτυξη του σώματος . . . . .	9
2.2.5 Μεταβολισμός . . . . .	10
2.2.6 Συμπεριφορά . . . . .	11
<b>3 Υλικά και Μέθοδος</b>	<b>14</b>
3.1 Πειραματόζωα . . . . .	14

3.1.1	Τύπος, Ποικιλία και Ομογενοποίηση . . . . .	14
3.1.2	Τοποθέτηση των ζώων και φροντίδα αυτών . . . . .	15
3.1.3	Χειρουργικοί χειρισμοί . . . . .	16
3.1.4	Παρακολούθηση των πειραματοζώων . . . . .	17
3.2	Πορεία του πειράματος . . . . .	17
3.2.1	Εξέταση των ζώων . . . . .	17
3.2.2	Ομάδες πειραματοζώων . . . . .	18
3.2.2.1	Βασικός θεωρητικός άξονας . . . . .	18
3.2.2.2	Γενικές Πληροφορίες . . . . .	19
3.2.2.3	Πολυαμίνες . . . . .	24
3.2.2.4	Θειόλες . . . . .	27
3.2.2.5	Βενζοπυρένιο . . . . .	28
<b>4</b>	<b>Μαθηματική Μοντελοποίηση</b>	<b>32</b>
4.1	Πληθυσμιακή και Εξατομικευμένη Προσέγγιση . . . . .	33
4.2	Μαθηματικά Μοντέλα Βιολογικής Ανάπτυξης Πληθυσμών . . . . .	35
4.2.1	Μαθηματικό Μοντέλο Ανάπτυξης Exponential . . . . .	38
4.2.2	Μαθηματικό Μοντέλο Ανάπτυξης Gompertz . . . . .	39
4.2.3	Μαθηματικό Μοντέλο Ανάπτυξης Bertalanffy . . . . .	41
4.2.4	Μαθηματικό Μοντέλο Ανάπτυξης Logistic . . . . .	41
4.2.5	Μαθηματικό Μοντέλο Ανάπτυξης Gauss . . . . .	42
4.2.6	Μαθηματικό Μοντέλο Ανάπτυξης Polynomial . . . . .	43
4.2.7	Μαθηματικό Μοντέλο Ανάπτυξης Brody . . . . .	45
4.2.8	Μαθηματικό Μοντέλο Ανάπτυξης Weibull . . . . .	45
4.2.9	Μαθηματικό Μοντέλο Ανάπτυξης PowerLaw . . . . .	46
4.3	Αλγόριθμος Προσαρμογής των Μαθηματικών Μοντέλων . . . . .	47
4.4	Αριθμητικές Μέθοδοι Υλοποίησης Ελαχίστων Τετραγώνων . . . . .	49
4.5	Αριθμητική Μέθοδος Marquardt-Levenberg . . . . .	51
4.6	Προσέγγιση αρχικών τιμών σε μη γραμμικά ελάχιστα τετράγωνα . . . . .	53
4.7	Μετρικές καλής προσαρμογής για μη γραμμικά μοντέλα . . . . .	53
4.8	Προβλεψιμότητα των μαθηματικών μοντέλων . . . . .	56
<b>5</b>	<b>Ταξινόμηση</b>	<b>59</b>
5.1	Το πρόβλημα της Ταξινόμησης . . . . .	60
5.2	Ταξινομητής Perceptron . . . . .	63

5.3	Ταξινομητής Support Vector Machine . . . . .	65
5.4	Ταξινομητής K - Nearest Neighbors . . . . .	68
5.5	Γενικές Μετρικές Εκτίμησης Ταξινομητών . . . . .	69
5.6	Εκτίμηση Χαρακτηριστικών . . . . .	72
<b>6</b>	<b>Στατιστική Προσέγγιση Καταλήξεων</b>	<b>75</b>
6.1	Κανονικότητα των δεδομένων . . . . .	76
6.2	Τεστ Κανονικότητας . . . . .	77
6.2.1	Το τεστ των Anderson-Darling . . . . .	77
6.2.2	Το τεστ των Jarque-Bera . . . . .	78
6.2.3	Το τεστ του Lilliefors . . . . .	80
6.2.4	Το τεστ των Shapiro-Wilk . . . . .	80
6.3	Q-Q γράφημα και κανονικότητα . . . . .	81
6.4	Μετατροπή των δεδομένων με κανονικοποίηση (z-score ή minmax-score)	82
6.5	Μη Παραμετρικό Τεστ των Kolmogorov-Smirnov Δύο Δειγμάτων . . .	83
6.6	Στατιστικές ροπές . . . . .	84
6.7	Το μη παραμετρικό τεστ των Mann - Whitney U Δύο Πληθυσμών . . .	87
<b>7</b>	<b>Κεφάλαιο: Αποτελέσματα Και Συμπεράσματα</b>	<b>88</b>
7.1	Μοντελοποίηση της ανάπτυξης της Χημικής Καρκινογένεσης . . . . .	89
7.2	Ταξινόμηση . . . . .	97
7.3	Στατιστική Προσέγγιση . . . . .	99
7.4	Μη αποδοτικά μοντέλα και διαδικασίες . . . . .	107
<b>8</b>	<b>Κεφάλαιο: Καινοτομίες και Μελλοντικοί Στόχοι</b>	<b>110</b>
8.1	Καινοτομίες . . . . .	110
8.2	Μελλοντικοί Στόχοι . . . . .	113
	<b>Βιβλιογραφία</b>	<b>114</b>
<b>A</b>	<b>Παράρτημα αποτελεσμάτων</b>	<b>121</b>
<b>B</b>	<b>Παράρτημα Σχημάτων Προσαρμογής Μοντέλων</b>	<b>130</b>



# ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

---

2.1	Η εξωτερική όψη του <i>Mus musculus domesticus varieta' albinus</i> , NMRI. .....	7
2.2	Κεφαλή του <i>Mus musculus domesticus varieta' albinus</i> , NMRI. Στοι- χεία της κεφαλής: Superciliary άνω βλέφαρο, Mystacial μύστακες, Sub- mental πώγων ,γένιο Interramal γενειάδα. ....	7
2.3	Κοιλιακή όψη. Διακρίνονται οι θηλές : Pectoral group : θωρακική ομάς θηλών Inguinal group : Κάτω κοιλιακή ομάς θηλών. ....	8
2.4	Συσχέτιση βάρους με φύλο ποντικού. ....	10
3.1	Μοριακή δομή του 3,4-βενζοπυρενίου. ....	29
3.2	Η καταβολική ενεργοποίηση του B[a]P μέσω του κυτοχρώμου P0450. ....	30
4.1	Διάγραμμα ροής των Μαθηματικών Μοντέλων Ανάπτυξης ανάλογα με τις τιμές των παραμέτρων τους ....	38
6.1	Παράδειγμα εμπειρικής κατανομής. ....	79
7.1	QQ γράφημα των δραστικών ουσιών και του βενζοπυρενίου ως προς την αντίστοιχη κανονική κατανομή. ....	101
7.2	QQ γράφημα των Πολυαμινών μονής και διπλής δόσης ως προς την αντίστοιχη κανονική κατανομή. ....	103
7.3	Σύγκριση Δόσεων με ιστόγραμμα σχετικών συχνοτήτων στην Αιθυλε- νοδιαμίνη. ....	105
7.4	Σύγκριση Δόσεων με ιστόγραμμα σχετικών συχνοτήτων στην Καντα- βερίνη. ....	106
7.5	Σύγκριση Δόσεων με ιστόγραμμα σχετικών συχνοτήτων στην Πουτρε- σίνη. ....	106
7.6	Περίπτωση κακής προσαρμογής με χρήση του μοντέλου Brody ....	107

7.7	Περίπτωση καλής προσαρμογής με χρήση του μοντέλου Brody . . . .	108
B.1	Παράδειγμα προσαρμογής προτύπου με Βενζοπυρένιο στο μοντέλο Exponential . . . . .	130
B.2	Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Exponential . . . . .	131
B.3	Παράδειγμα προσαρμογής προτύπου με Πουτρεσκίνη στο μοντέλο Ex- ponential . . . . .	131
B.4	Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Gompertz . . . . .	132
B.5	Παράδειγμα προσαρμογής προτύπου με Διθειοθρειτόλη στο μοντέλο Gompertz . . . . .	132
B.6	Παράδειγμα προσαρμογής προτύπου με αιθυλικό εστέρα της Κυστεΐ- νης στο μοντέλο Gompertz . . . . .	133
B.7	Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Bertalanffy . . . . .	133
B.8	Παράδειγμα προσαρμογής προτύπου με Διθειοθρειτόλη στο μοντέλο Bertalanffy . . . . .	134
B.9	Παράδειγμα προσαρμογής προτύπου με Κυστεΐνη στο μοντέλο Berta- lanffy . . . . .	134
B.10	Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Lo- gistic . . . . .	135
B.11	Παράδειγμα προσαρμογής προτύπου με Βισμούθειόλη στο μοντέλο Logistic . . . . .	135
B.12	Παράδειγμα προσαρμογής προτύπου με Μερκαπτοηλεκτρικό οξύ στο μοντέλο Logistic . . . . .	136
B.13	Παράδειγμα προσαρμογής προτύπου με Βενζοπυρένιο στο μοντέλο Gauss . . . . .	136
B.14	Παράδειγμα προσαρμογής προτύπου με Αιθυλενοδιαμίνη στο μοντέλο Gauss . . . . .	137
B.15	Παράδειγμα προσαρμογής προτύπου με Μερκαπτοηλεκτρικό οξύ στο μοντέλο Gauss . . . . .	137
B.16	Παράδειγμα προσαρμογής προτύπου με Διθειοθρειτόλη στο μοντέλο Polynomial . . . . .	138

B.17 Παράδειγμα προσαρμογής προτύπου με Πουτρεσκίνη στο μοντέλο Polynomial . . . . .	138
B.18 Παράδειγμα προσαρμογής προτύπου με Μερκαπτοηλεκτρικό οξύ στο μοντέλο Polynomial . . . . .	139
B.19 Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο PowerLaw . . . . .	139
B.20 Παράδειγμα προσαρμογής προτύπου με Αιθυλενοδιαμίνη στο μοντέλο PowerLaw . . . . .	140
B.21 Παράδειγμα προσαρμογής προτύπου με Κυστεΐνη στο μοντέλο Pow- erLaw . . . . .	140

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

---

2.1	Πίνακας παραμέτρων βάρους και αιτιολογικών παραμέτρων. . . . .	11
2.2	Πίνακας βιοχημικών παραμέτρων. . . . .	12
3.1	Αριθμοί πειραματοζώων ανά δραστική ουσία μονής και διπλής δόσης.	24
3.2	Αριθμοί πειραματοζώων που εμφάνισαν καρκινική κυτταρική ανάπτυξη στην οικογένεια των Πολυαμινών Μονής Δόσης. . . . .	26
3.3	Αριθμοί πειραματοζώων που εμφάνισαν καρκινική κυτταρική ανάπτυξη στην οικογένεια των Πολυαμινών Διπλής Δόσης. . . . .	26
3.4	Αριθμοί πειραματοζώων που εμφάνισαν καρκινική κυτταρική ανάπτυξη στην οικογένεια των Θειολών. . . . .	27
3.5	Αριθμός Παρατηρήσεων ανά κλάση . . . . .	29
5.1	Πίνακας Σύγκρισης . . . . .	69
5.2	Γενικές μετρικές απόδοσης των ταξινομητών. . . . .	69
5.3	Μετρικές Εκτίμησης Ταξινομητών με ειδίκευση στην ανισορροπία κλάσεων . . . . .	71
6.1	Πίνακας Μαθηματικών Τύπων των Στατιστικών Ροπών. . . . .	86
7.1	Μεταβλητότητα Παραμέτρων των Αναπτυξιακών Μοντέλων σε όλες τις δραστικές ουσίες. . . . .	93
7.2	Διαστήματα Παραμέτρων των Αναπτυξιακών Μοντέλων σε όλες τις δραστικές ουσίες. . . . .	94
7.3	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced SSE στο Βενζοπυρένιο. . . . .	94
7.4	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced RMSE στο Βενζοπυρένιο. . . . .	95

7.5	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced MAE στο Βενζοπυρένιο. . . . .	95
7.6	Προβλεψιμότητα των Αναπτυξιακών Μοντέλων. . . . .	96
7.7	Πίνακας Απόδοσης του ταξινομητή Γραμμικός Support Vector Machine. . . . .	97
7.8	Πίνακας Απόδοσης του ταξινομητή K-Nearest Neighbor. . . . .	98
7.9	Συνεισφορά Παραμέτρων Στην Ταξινόμηση. . . . .	99
7.10	Τεστ Ελέγχου Κανονικότητας Πολυαμινών. . . . .	100
7.11	Τεστ Ελέγχου Κανονικότητας Θειολών. . . . .	100
7.12	Τεστ Κοινής Κατανομής των Kolmogorov-Smirnov. . . . .	102
7.13	Στατιστικές Ροπές σε Πολυαμίνες και σύγκριση έναντι του Βενζοπυρένιου. . . . .	102
7.14	Στατιστικές Ροπές σε Πολυαμίνες και σύγκριση μονών και διπλών δόσεων. . . . .	104
7.15	Τεστ Ελέγχου Κανονικότητας Πολυαμινών Μονών Δόσεων με Μετασχηματισμούς. . . . .	109
A.1	Πίνακας Παραμέτρων των Αναπτυξιακών Μοντέλων στο Βενζοπυρένιο. . . . .	122
A.2	Πίνακας Παραμέτρων των Αναπτυξιακών Μοντέλων στις Πολυαμίνες. . . . .	123
A.3	Πίνακας Παραμέτρων των Αναπτυξιακών Μοντέλων στις Θειόλες. . . . .	124
A.4	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced SSE στις Πολυαμίνες. . . . .	125
A.5	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced RMSE στις Πολυαμίνες. . . . .	125
A.6	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced MAE στις Πολυαμίνες. . . . .	126
A.7	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced SSE στις Θειόλες. . . . .	126
A.8	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced RMSE στις Θειόλες. . . . .	127
A.9	Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced MAE στις Θειόλες. . . . .	127
A.10	Τεστ Ελέγχου Κανονικότητας Πολυαμινών Διπλών Δόσεων με Μετασχηματισμούς. . . . .	128

A.11 Τεστ Ελέγχου Κανονικότητας Θειολών Μονών Δόσεων με Μετασχηματισμούς. . . . .	129
---	-----

# ΚΑΤΑΛΟΓΟΣ ΑΛΓΟΡΙΘΜΩΝ

---

## ΓΛΩΣΣΑΡΙ

---

**DNA** - Νουκλειικό οξύ, χημικό μακρομόριο που αποτελεί το γενετικό υλικό όλων των έμβιων οργανισμών καθώς και αρκετών ιών. Εμπεριέχει πληροφορίες που καθορίζουν όλα τα χαρακτηριστικά ενός οργανισμού.

**RNA** - Ριβοζονουκλειικό οξύ, μία από τις δύο κατηγορίες των πολυμερών νουκλειικών οξέων στο κύτταρο. Παίζει σημαντικό ρόλο στην μετάφραση του γενετικού κώδικα.

**ODC** - Αποκαρβοξυλάση της ορνιθίνης, ένα ένζυμο που είναι καταλύτης της ορνιθίνης (προϊόν του κύκλου ουρίας) με σκοπό τον σχηματισμό πολυαμινών.

**QR** - Η QR αποτελεί διαδικασία παραγοντοποίησης ενός πραγματικού πίνακα. Στην ουσία αναλύουμε έναν πίνακα σε έναν ορθοκανονικό πίνακα  $Q$  και έναν άνω τριγωνικό πίνακα  $R$ , δηλαδή ισχύει  $A = Q * R$ . Σημειώνουμε ότι ο  $R$  είναι άνω τριγωνικός στο κομμάτι  $n \times n$  (άνω τμήμα) ενώ οι τελευταίες  $m-n$  γραμμές του είναι μηδενικές.

**SVD** - Η SVD αποτελεί διαδικασία παραγοντοποίησης ενός πίνακα  $A$  στην μορφή  $A = U * D * V^t$ . Οι πίνακες  $U$  και  $V$  είναι ορθοκανονικοί πίνακες ενώ ο  $D$  είναι ένας ορθογώνια διαγώνιος πίνακας.



# ΠΕΡΙΛΗΨΗ

---

Μπουχάρας Δημήτριος, Μ.Δ.Ε. στην Πληροφορική, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, 2020.

Μαθηματική μοντελοποίηση και ανάλυση αναπτυξιακών μοντέλων στην πρόληψη δημιουργίας καρκινικών κυττάρων.

Επιβλέπων: Γεώργιος Μανής, Αναπληρωτής Καθηγητής.

Ο καρκίνος είναι μία από τις σοβαρότερες ασθένειες που μαστίζουν σήμερα τις αναπτυγμένες χώρες στο σύνολό τους. Στατιστικές μελέτες που πραγματοποιούνται δείχνουν ότι αποτελεί τη δεύτερη πιο συχνή αιτία θανάτου μετά τις καρδιοπάθειες. Ως εκ τούτου την τελευταία δεκαετία παρατηρείται έντονα ότι η μάχη με τον καρκίνο να αποτελεί επίκεντρο αρκετών επιστημονικών ερευνητικών κέντρων. Μάλιστα οι τελευταίες έρευνες αφορούν στοχευμένες θεραπείες, εύρεση και μελέτη χαρακτηριστικών των κυττάρων που προκαλούν μετάσταση, την ανοσοθεραπεία και συσχετίσεις σωματικού βάρους με την εμφάνιση της νόσου. Αξίζει να σημειωθεί ότι παρόλο το έντονο επιστημονικό ενδιαφέρον εντούτοις οι επιστημονικές εργασίες που αφορούν την πρόληψη είναι περιορισμένες.

Η παρούσα μεταπτυχιακή εργασία επικεντρώνεται στην μελέτη μαθηματικών αναπτυξιακών μοντέλων. Πιο συγκεκριμένα, αναζητούμε το μοντέλο με την καλύτερη περιγραφική ικανότητα στο φαινόμενο της εξέλιξης της καρκινικής δραστηριότητας στο πεδίο του χρόνου. Η ανάπτυξη καρκινικών κυττάρων προκαλείται σε ένα συγκεκριμένο είδος πειραματόζων κατόπιν χημικής καρκινογένεσης με έναν πλήρως καρκινογόνο παράγοντα, γνωστό ως 3.4-βενζοπυρένιο. Σκοπός της εργασίας είναι η μελέτη του φαινομένου της χημικής καρκινογένεσης, της αναστολής της καρκινικής δραστηριότητας και της ανάπτυξης των κακοήθων όγκων. Απώτερος στόχος είναι η χημική πρόληψη.

Τα αποτελέσματα που αντλήθηκαν από την έρευνά μας είναι άκρως ενθαρρυντικά και αν συνεχιστεί θα μπορούσε ενδεχομένως να διευρύνει το πεδίο της έρευ-

νας όπως για παράδειγμα να οδηγήσει στην δημιουργία συμπληρωμάτων διατροφής ικανών να αναστέλλουν την χημική καρκινογένεση με βεβαιότητα.

# EXTENDED ABSTRACT

---

Boucharas Dimitrios, M.Sc. in Computer Science, Department of Computer Science and Engineering, University of Ioannina, Greece, 2020.

Mathematical modeling and growth model analysis for prevention of cancer cell development..

Advisor: Georgios Manis, Assistant Professor.

Cancer is one of the most serious diseases nowadays plaguing developed countries as a whole. It is highlighted in statistical studies as the second leading cause of death globally after heart disease. During the last decade it has been strongly observed that the fight against cancer has been the focus of several scientific research centers. In fact, recent researches refer to targeted therapies, metastatic cell characteristics, immunotherapy and weight correlations with disease onset. It is worth noting that despite the intense scientific interest, scientific work on prevention is limited.

The present dissertation focuses on the study of mathematical development models. More specifically, we look for the model with the best descriptive ability on the phenomenon of the evolution of cancer activity in the field of time. Cancer cell growth is induced in a specific species of experimental animal by chemical carcinogenesis with a fully carcinogenic agent known as 3,4-benzopyrene. The purpose of this study is to investigate the phenomenon of chemical carcinogenesis, the inhibition of cancer activity and the development of malignant tumors. The ultimate goal is chemical prevention.

The results of this research are extremely encouraging and in case it proceeds, it will potentially expand the scope of this research into areas such as developing nutritional supplements capable of inhibiting chemical carcinogenesis with certainty.

# ΚΕΦΑΛΑΙΟ 1

## ΕΙΣΑΓΩΓΗ

---

### 1.1 Στόχοι

### 1.2 Δομή της Διατριβής

---

### 1.1 Στόχοι

Στην παρούσα διατριβή έγινε μία προσπάθεια να παρατεθούν και να δοκιμαστούν όσα αναπτυξιακά μοντέλα ομοιάζουν της διαδικασίας της ομοιόστασης με σκοπό να καταδειχθεί η ανάπτυξη της χημικής καρκινογένεσης με ή χωρίς την ύπαρξη δραστικής ουσίας. Ο καρκίνος αποτελεί πληθυσμό και ως απόρροια αυτού επιλέχθηκαν αναπτυξιακά μοντέλα που έχουν χρησιμοποιηθεί κυρίως για την περιγραφική τους ικανότητα σε πληθυσμιακές αναπτύξεις.

Στην συνέχεια, θα πραγματοποιηθεί μία σύγκριση στα αναπτυξιακά μοντέλα ως προς το πόσο καλά προσεγγίζουν τα δεδομένα με στόχο να βρεθεί το καταλληλότερο ως προς την περιγραφική του ικανότητα. Επιπλέον, θα ελεγχθεί η διαχωριστική ικανότητα που προσφέρουν τα μοντέλα σε ένα πρόβλημα ταξινόμησης των διαφορετικών δραστικών ουσιών που δοκιμάστηκαν στα πειράματα.

Ακόμη, θα συγκριθεί το προσδόκιμο ζωής των πειραματοζώων μέσω στατιστικών μεθόδων για να υπογραμμίσουμε την σημαντικότητα των δραστικών ουσιών που δοκιμάστηκαν αλλά και αποδείξουμε την θωράκιση του ανοσοποιητικού συστήματος.

Τα δεδομένα μας είναι ενθαρρυντικά και αν η έρευνα συνεχιστεί θα μπορούσε

ενδεχομένως να οδηγήσει στην δημιουργία συμπληρωμάτων και φαρμάκων ικανών να αναστέλλουν την χημική καρκινογένεση πλήρως.

## 1.2 Δομή της Διατριβής

Η διατριβή αποτελείται από 8 κεφάλαια και 2 παραρτήματα. Το παρών κεφάλαιο είναι το πρώτο και δίνονται οι στόχοι και η δομή της παρούσας διατριβής. Το δεύτερο κεφάλαιο εξιστορεί την ποικιλία του ποντικού NMRI και αναφέρει βασικά βιολογικά του δεδομένα. Εν συνεχεία παραθέτει τους λόγους της συγκεκριμένης επιλογής ποντικιού και επικεντρώνεται στην αντιστοιχία του με τον άνθρωπο. Ενώ το κεφάλαιο κλείνει με την μελέτη συγκεκριμένων παραμέτρων όπως ο φυλετικός διμορφισμός, ο μεταβολισμός και η συμπεριφορά του ζώου. Το τρίτο κεφάλαιο αναφέρεται στον τρόπο ομογενοποίησης του πληθυσμού των ποντικιών και επεκτείνεται στον ειδικό χειρισμό και φροντίδα των πειραματοζώων ενώ περατώνει με μία αναφορά στις βασικές ομάδες των δραστικών ουσιών που δοκιμάστηκαν έναντι του βενζοπυρενίου καθώς δίνονται και αριθμητικές πληροφορίες για αυτές.

Το τέταρτο κεφάλαιο θέτει τα θεμέλια της μαθηματικής μοντελοποίησης της χημικής καρκινογένεσης. Το κεφάλαιο ξεκινάει με εκτενή βιβλιογραφική αναζήτηση ανάμεσα στην εξατομικευμένη και την πληθυσμιακή προσέγγιση. Στην συνέχεια, αναφέρονται τα μαθηματικά μοντέλα που είναι πιο πιθανά να προσεγγίσουν την ανάπτυξη της χημικής καρκινογένεσης μέσω της μορφολογίας τους ενώ τονίζεται η επιστημονική τους χρήση στον τομέα της πληθυσμιακής ανάπτυξης. Ακόμη, δίνεται το μαθηματικό υπόβαθρο, οι αριθμητικές μέθοδοι και οι μετρικές καλής προσαρμογής για την υλοποίηση και σύγκριση των αναπτυξιακών μοντέλων.

Το πέμπτο κεφάλαιο περιγράφει το πρόβλημα της ταξινόμησης, παραθέτει βασικούς ταξινομητές που αποτέλεσαν τους θεμέλιους λίθους για την επιστήμη των Υπολογιστών και διεύρυναν την επιστήμη των Μαθηματικών. Στην συνέχεια αναφέρονται οι μετρικές εκτίμησης των ταξινομητών και στοχεύετε το πρόβλημα της ανισορροπίας των κλάσεων. Το κεφάλαιο κλείνει με την εκτίμηση των χαρακτηριστικών ως προς την συμβολή τους στην ταξινόμηση.

Το έκτο κεφάλαιο έχει ως βασική του ενασχόληση το προσδόκιμο ζωής των πειραματοζώων μέσω στατιστικών προσεγγίσεων. Το κεφάλαιο ξεκινάει με την κανονικότητα των δεδομένων και τα βασικά μη παραμετρικά τεστ. Συνεχίζει με ανάλυση

των στατιστικών ροπών και αυτές επιφέρουν στην κατανομή ενώ κλείνει θέτοντας ερωτήσεις κοινής κατανομής ανάμεσα σε πληθυσμούς.

Το έβδομο κεφάλαιο συμπεριλαμβάνει τα αποτελέσματα της έρευνάς μας και τα αντίστοιχα εμπειριστατωμένα συμπεράσματα. Ξεκινάει με τα αποτελέσματα της μαθηματικής μοντελοποίησης, συνεχίζει τα αποτελέσματα της ταξινόμησης και κλείνει με την στατιστική συμπερασματολογία και κάποιες μη αποδοτικές διαδικασίες και μοντέλα. Το όγδοο κεφάλαιο υπογραμμίζει τις καινοτομίες που είχαν ως στόχο την βελτίωση της γενικής εικόνας της ανάπτυξης της χημικής καρκινογένεσης και θέτει μελλοντικούς στόχους που θα θέλαμε να εξετάσουμε διεξοδικά.

Στο πρώτο παράρτημα παραθέτουμε επιπλέον αποτελέσματα της έρευνάς μας για πληρότητα καθώς δεν ενείχαν σημαντικές αλλαγές ενώ στο δεύτερο παράρτημα παραθέτουμε εικόνες από την μοντελοποίηση με σκοπό να τονίσουμε την διαφορετικότητα των μοντέλων.

# ΚΕΦΑΛΑΙΟ 2

## ΒΙΟΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ

---

### 2.1 Η ιστορία του ποντικού NMRI

### 2.2 Στοιχεία για τον ποντικό NMRI

---

Σε αυτό το κεφάλαιο παρατίθεται η ιστορία του ποντικού NMRI και αναφέρονται κάποια βιολογικά του στοιχεία. Εν συνεχεία, επίκεντρο γίνεται η αντιστοιχία του με τον άνθρωπο και οι λόγοι που έγινε η συγκεκριμένη επιλογή του ποντικού. Τέλος, μελετώνται συγκεκριμένες παραμέτροι όπως ο φυλετικός διμορφισμός, ο μεταβολισμός και η συμπεριφορά του ζώου.

### 2.1 Η ιστορία του ποντικού NMRI

Ο αλφικός μυς (ποντικός) NMRI (Naval Medical Research Institute) είναι μια φυλή του είδους *Mus musculus* L που καταγράφηκε ταξινομικά από τον Λινναίο [1], ενώ ως ζώο είναι γνωστό στον άνθρωπο από αρχαιοτάτων χρόνων καθώς πολύ συχνά βρίσκεται στο άμεσο οικιακό του περιβάλλον. Ο ποντικός θεωρείται άγριο ζώο και όχι οικόσιτο, παρά το γεγονός ότι πολύ συχνά οι οικολογικοί θώκοι στους οποίους απαντάται συμπεριλαμβάνουν οι κατοικίες των ανθρώπων.

Πολλές παραλλαγές ποντικών αποτέλεσαν και αποτελούν ένα ζώο επιλογής για τον πειραματισμό που αφορά την ανάπτυξη φαρμάκων και την εν γένει μελέτη βιολογικών φαινομένων σχετικών με την τοξικολογία, τον καρκίνο, τις λοιμώξεις,

τα αυτοάνοσα νοσήματα, τις μεταμοσχεύσεις, το γήρας κ.α. Ο πειραματισμός με ποντικούς αρχίζει περί τα τέλη του 18ου αιώνα και συνεχίζεται μέχρι σήμερα. Για τις ανάγκες του βιοϊατρικού πειραματισμού έχουν δημιουργηθεί μερικές σημαντικές παραλλαγές του είδους *Mus musculus* μεταξύ των οποίων και ο αλφικός ποντικός NMRI. Ως γνωστόν τα αλφικά ζώα (albino) δεν διαθέτουν γονίδια παραγωγής χρωστικών για την χρώση της δοράς των τριχών και της ίριδος των οφθαλμών. Ως εκ τούτου έχουν απολέσει την προστασία που αποδίδει σε αυτά η παρουσία των εν λόγω χρωστικών. Ωστόσο είναι πιθανόν ο αλφισμός να προέκυψε και ως προσαρμογή πολλών ειδών ζώων στις συνθήκες των χιονοσκεπών περιοχών του Βορά και του αρκτικού κύκλου. Πολύ χαρακτηριστικά ο επίμυς (αρουραίος) Wistar, ο οποίος χρησιμοποιείται ως πειραματικό μοντέλο για πληθώρα μελετών, ανήκει στο είδος *Ratus norvegicus* (αρουραίος της Νορβηγίας) και συνεπώς η λευκή δορά πιθανόν να οφείλεται σε προσαρμογή στην λευκότητα του εδάφους εξ αιτίας του χιονιού. Ίσως αντίστοιχες προσαρμογές να έγιναν και σε άλλα ζώα των βόρειων περιοχών και του αρκτικού κύκλου, όπως π.χ στην περίπτωση της λευκής πολικής άρκτου. Ενδεχομένως έτσι να προέκυψε το είδος του ποντικού NMRI το οποίο χρησιμοποιήθηκε ως βιολογικό δοκιμαστικό μοντέλο για την χημική εξουδετέρωση του καρκινογόνου πολυκυκλικού υδρογονάνθρακα 3,4-βενζοπυρένιου.

Η εν λόγω ποικιλία ποντικού αναπτύχθηκε αρχικά στην Λοζάνη (Ελβετία) και από εκεί μια ομάδα εννέα ζώων μεταφέρθηκε στις ΗΠΑ το 1926 από την Clara Lynch. Το 1937 η εν λόγω ποικιλία ζώων ελήφθη από τον Poiley και αναπαρήχθη σε κανονικό και πλήρες εκτροφείο ώστε να τυποποιηθεί γενετικά και να μελετηθεί σε βάθος προκειμένου να χρησιμοποιείται ως τακτικό πειραματόζωο σε πολλούς τύπους πειραμάτων και έγινε γνωστή με την κωδική ονομασία NIH/Pl. Το 1955 εισάγεται στο Ινστιτούτο Ιατρικών Ερευνών του Ναυτικού των Ηνωμένων Πολιτειών όπου χρησιμοποιείται πλέον με τον κωδικό NMRI από την ακροστιχίδα των λέξεων Naval Medical Research Institute. Υπό αυτήν πλέον την κωδική ονομασία άτομα το 1955 θήλεις και άρρενες ποντικοί της συγκεκριμένης ποικιλίας μεταφέρθηκαν στο Bundes – Forschungsanstalt für Viruskrankheiten. Το 1958 η ποικιλία ποντικών NMRI μεταφέρθηκε στο Central Institute for Animal Breeding του Ανόβερου της τότε Δυτικής Γερμανίας. Το 1978 ένας αριθμός των εν λόγω ποντικών NMRI μεταφέρθηκαν στο Εργαστήριο Φυσιολογίας της Ιατρικής Σχολής του Πανεπιστημίου Ιωαννίνων και εγκαταστάθηκαν στο Πειραματικό Εκτροφείο Πειραματοζώων της Ιατρικής Σχολής. Στελέχη του εν λόγω μύς χρησιμοποιήθηκαν σταδιακά για την



μελέτη της εξουδετέρωσης της χημικής καρκινογένεσης της προκαλούμενης από 3,4-βενζοπυρένιο. Τα δεδομένα αυτών των μελετών ανήκουν στην επιστημονική ομάδα του Καθηγητή κ. Σπυρίδων Καρκαμπούνα και στο τμήμα Φυσιολογίας της Ιατρικής Σχολής του Πανεπιστημίου Ιωαννίνων. Ακόμη τα δεδομένα αναλύονται μαθηματικά με νέες μεθόδους στην ανά χειράς μεταπτυχιακή διατριβή με σκοπό την δημιουργία ενός μαθηματικού μοντέλου της χημικής καρκινογένεσης για περαιτέρω εξορθολογισμό και μαθηματικοποίηση της σχετικής έρευνας επί των αντικαρκινογόνων και εν γένει επί της ανάπτυξης μεθόδων αντιμετώπισης της νεοπλασματικής νόσου. Η επέκταση της μεταπτυχιακής αυτής μελέτης ευελπιστούμε ότι θα συνεχιστεί ως διδακτορική διατριβή με σκοπό την εφαρμογή στην πρόληψη του καρκίνου στον ανθρώπινο πληθυσμό.

## **2.2 Στοιχεία για τον ποντικό NMRI**

### **2.2.1 Ταξινόμηση και χρησιμότητα**

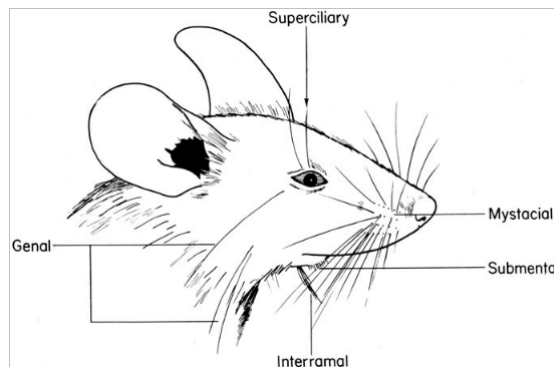
Ο ποντικός NMRI είναι ένα μικρό τρωκτικό θηλαστικό που ανήκει στο είδος *Mus musculus* και μπορεί να θεωρηθεί ότι ανήκει στο υποείδος *Mus musculus domesticus* και στην ποικιλία *Mus musculus domesticus albinus*, *varieta' NMRI*. Το ζώο αυτό είναι φυλογενετικώς συγγενές προς τα Πρωτεύοντα (Primates) στα οποία ανήκει ο Άνθρωπος ο έμφρων (*Homo sapiens*). Οπωσδήποτε ο μυς NMRI προέκυψε από σκόπιμες αναπαραγωγικές επιλογές ούτως ώστε να αποκτήσει ως πειραματόζωο εξειδικευμένα χαρακτηριστικά καταλληλότητας για την διεξαγωγή πειραμάτων κατάλληλων για την εκμαίευση συμπερασμάτων επί των δράσεων ουσιών και άλλων επεμβατικών δράσεων προκειμένου να υπάρξει η δυνατότητα χρήσης τους στο ανθρώπινο είδος προς όφελος των βελτιώσεων της θεραπείας και πρόληψης νόσων των ανθρώπων. Δηλαδή η χρησιμότητα του εν λόγω ποντικού είναι ο τοξικολογικός και φαρμακολογικός πειραματισμός, και εν γένει μελέτες που αφορούν την παθοφυσιολογία με συμπεράσματα που μπορούν κατόπιν να προβληθούν στο ανθρώπινο είδος.

## 2.2.2 Εξωτερική Μορφολογία

Η μορφολογία του σώματος του ποντικού NMRI αποσαφηνίζεται στις τρεις εικόνες, το Σχήμα 2.1, το Σχήμα 2.2 και το Σχήμα 2.3. Η μορφολογία του κατατάσσει το ζώο αυτό στην τάξη των Rodentia (τροκτικά), στο είδος *Mus musculus*, στο υποείδος *Mus musculus domesticus varieta' albinus*, NMRI.



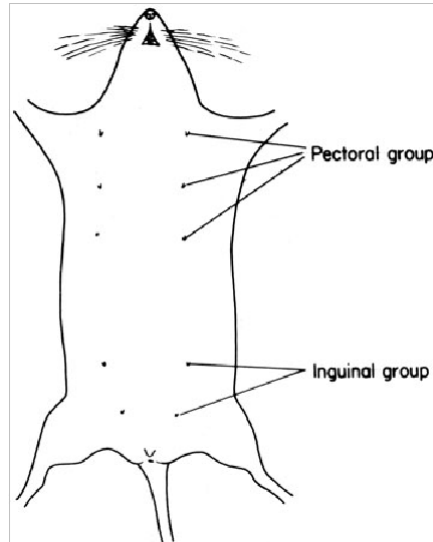
Σχήμα 2.1: Η εξωτερική όψη του *Mus musculus domesticus varieta' albinus*, NMRI.



Σχήμα 2.2: Κεφαλή του *Mus musculus domesticus varieta' albinus*, NMRI. Στοιχεία της κεφαλής: Superciliary άνω βλέφαρο, Mystacial μύστακες, Submental πώγων ,γένιο Interramal γενειάδα.

## 2.2.3 Αντιστοιχία με τον άνθρωπο

Υπάρχει πλήρης ανατομική αναλογία των οργάνων και των ιστών του ποντικού με τα όργανα του ανθρώπου, ανεξαρτήτως βεβαίως του μεγέθους και της μορφολογίας.



Σχήμα 2.3: Κοιλιακή όψη. Διακρίνονται οι θηλές : Pectoral group : θωρακική ομάς θηλών Inguinal group : Κάτω κοιλιακή ομάς θηλών.

Το ίδιο ισχύει επίσης με τις φυσιολογικές λειτουργίες τηρουμένων των αναλογιών μεταξύ του συγκεκριμένου ποντικού και του ανθρώπου. Μια γενική διαφορά μεταξύ των δύο αυτών ειδών είναι:

1. Οι μεγαλύτεροι λειτουργικοί ρυθμοί του επίμυος (π.χ ταχυκαρδικό και ταχυπνοϊκό ζώο) οι οποίοι αφορούν όλες τις λειτουργίες αυτού. Είναι πιθανόν οι αυξημένοι αυτοί ρυθμοί να σχετίζονται και με το μικρό προσδόκιμο επιβίωσης που παρουσιάζουν τα εν λόγω μικρόσωμα τρωκτικά.

2. Η διάρκεια ζωής (προσδόκιμο επιβίωσης) γενικά του είδους *Mus musculus* είναι για τα μεν άγρια ζώα (wild type) τα οποία ζουν ελεύθερα στο περιβάλλον γύρω στους 12 μήνες, ενώ τα ζώα που διαβιούν σε συνθήκες αιχμαλωσίας ή εκτροφής παρουσιάζουν μεγαλύτερη διάρκεια ζωής που κυμαίνεται μεταξύ ενός – δύο ετών. Το υποείδος NMRI είναι απόλυτα αναπτυγμένο με σκόπιμες διασταυρώσεις και επιλεγμένο από τον άνθρωπο, οπότε η μη γενετικά τροποποιημένες παραλλαγές του επιβιώνουν πάντα στο εσωτερικό εκτροφείων και η διάρκεια ζωής τους είναι επίσης από ενός έως δύο ετών.

## 2.2.4 Διμορφισμός ως προς το φύλο εν σχέσει με την ανάπτυξη του σώματος

Η ποικιλία ποντικών (μυών) NMRI εμφανίζει φυλετικό διμορφισμό, σε ό,τι αφορά το μέγεθος του σώματος και την διάρκεια ζωής. Ο άρρεν ποντικός αναπτύσσει σώμα μεγαλύτερων διαστάσεων και βάρους, ενώ ο θήλυς ποντικός έχει μικρότερες σωματικές διαστάσεις και μικρότερο βάρος. Το βάρος του ώριμου άρρενος (15 εβδομάδων ηλικίας από την γέννηση) κυμαίνεται μεταξύ 41 – 50 γραμμάρια αντιστοίχως στα μικρόσωμα και στα μεγαλόσωμα άτομα.

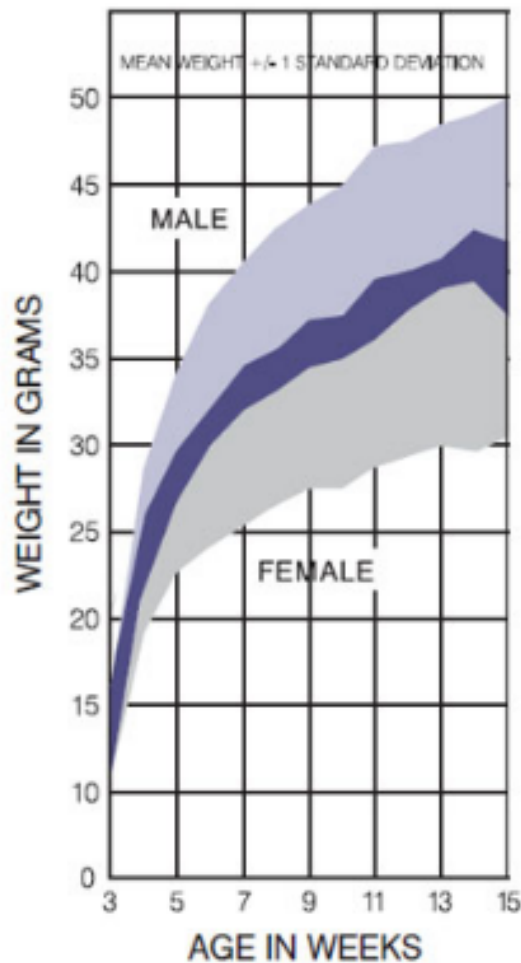
Ο ποντικός NMRI είναι ένα μικρό τρωκτικό θηλαστικό που ανήκει στο είδος *Mus musculus* και μπορεί να θεωρηθεί ότι ανήκει στο υποείδος *Mus musculus domesticus* και οπωσδήποτε προέκυψε από σκόπιμες αναπαραγωγικές επιλογές ούτως ώστε να αποκτήσει ως πειραματόζωο εξειδικευμένα χαρακτηριστικά καταλληλότητας για την διεξαγωγή πειραμάτων κατάλληλων για την εκμείευση συμπερασμάτων επί των δράσεων ουσιών και άλλων επεμβατικών δράσεων προκειμένου να υπάρξει η δυνατότητα χρήσης τους στο ανθρώπινο είδος προς όφελος των βελτιώσεων της θεραπείας και πρόληψης νόσων των ανθρώπων. Υπάρχει πλήρης ανατομική αντιστοιχία των οργάνων και των ιστών του ποντικού με τα όργανα του ανθρώπου, ανεξαρτήτως βεβαίως του μεγέθους και της μορφολογίας. Το ίδιο ισχύει επίσης με τις φυσιολογικές λειτουργίες τηρουμένων των αναλογιών μεταξύ του συγκεκριμένου ποντικού και του ανθρώπου. Μια γενική διαφορά μεταξύ των δύο αυτών ειδών είναι [2]:

1. Οι μεγαλύτεροι λειτουργικοί ρυθμοί του επίμυος (π.χ ταχυκαρδικό και ταχυπνοϊκό ζώο) οι οποίοι αφορούν όλες τις λειτουργίες αυτού. Είναι πιθανόν οι αυξημένοι αυτοί ρυθμοί να αντικατοπτρίζουν έναν ταχύτερο ενεργειακό μεταβολισμό, ο οποίος είναι απαραίτητος διότι τα εν λόγω ζώα χάνουν εύκολα θερμότητα εκ του γεγονότος ότι είναι μικρόσωμα. Σημειωτέον ότι στον *Mus musculus* δεν έχουμε την παρουσία λίπους στον υποδόριο ιστό.

2. Το μέγεθος του σώματος και των οργάνων. Ο ποντικός μπορεί να φτάσει τα 40 – 50 γραμμ. ενώ ο άνθρωπος έχει ένα μέσο βάρος σώματος 70 kgr, δηλαδή μεγαλύτερο κατά 1550 περίπου φορές. Η διαφορά αυτή του μεγέθους συνοδεύει όλα τα όργανα του ανθρώπου και του μυός.
3. Οι κινήσεις, η νευρική αγωγιμότητα είναι πολύ ταχύτερες στον ποντικό από ό,τι στον άνθρωπο.

4. Οι κυήσεις στον ποντικό είναι πολύδυμες (>4 – 8 εμβρύων ανά κύηση), ενώ

στον άνθρωπο οι πολυδυμία είναι σπάνια.



Σχήμα 2.4: Συσχέτιση βάρους με φύλο ποντικού.

### 2.2.5 Μεταβολισμός

Ο μεταβολισμός των ποντικών NMRI έχει πολλές αναλογίες με αυτόν του ανθρώπου, ωστόσο το αναβολικό και το καταβολικό του σκέλος εμφανίζει μεγαλύτερες ταχύτητες με αποτέλεσμα τα ζώα αυτά να διαβιούν σε μια κατάσταση μεγάλης μεταβολικής ταχύτητας, πράγμα το οποίο ερμηνεύει τους ταχείς χρόνους αναπλήρωσης που εμφανίζουν οι μεταβολίτες τους όταν απομακρύνονται μέσω αιμοληψιών και γενικότερα τους μικρούς χρόνους επούλωσης τραυμάτων ή αναπλήρωσης του αίματος σε περιπτώσεις μαζικής απώλειας αυτού για λόγους πειραματικούς. Οι καρδιαχικοί ρυθμοί ακολουθούν το πρότυπο της εποχιακής διαδοχής ημέρας και νύχτας, ωστόσο τα εν λόγω ζώα είναι συχνά ενεργά και κατά την περίοδο του σκό-

τους.

## 2.2.6 Συμπεριφορά

Η συμπεριφορά των ποντικών NMRI σε λαβύρινθο Skinner χαρακτηρίζεται από αποφυγή του κινδύνου [3]. Αλλαγές της συμπεριφοράς τους με την πάροδο της ηλικίας παρατηρούνται δείχνοντας ότι τα ζώα έχουν ικανό βαθμό μάθησης [4]. Θεωρείται ότι οι συμπεριφορά τους είναι ένα μείγμα διαφόρων στερεοτύπων τα οποία υιοθετούνται ανάλογα με τις λειτουργικές ανάγκες του ζώου όπως επίσης και τις δράσεις συγκεκριμένων ουσιών.

Παράμετρος	Άρρενα 6 εβδομάδων	Άρρενα 12 εβδομάδων	Θήλεα 6 εβδομάδων	Θήλεα 12 εβδομάδων
<i>Παράμετροι Βάρους</i>				
Σωματικό Βάρος	29,9 ± 1,8	41 ± 1,6	23 ± 1,9	32,8 ± 2,7
Καρδιά	0,167 ± 0,022	0,225 ± 0,038	0,127 ± 0,015	0,189 ± 0,030
Ήπαρ	1,182 ± 0,174	2,027 ± 0,122	1,145 ± 0,126	1,443 ± 0,151
Πνεύμονες	0,184 ± 0,016	0,220 ± 0,025	0,160 ± 0,009	0,203 ± 0,026
Νεφρός	0,419 ± 0,042	0,575 ± 0,043	0,257 ± 0,027	0,363 ± 0,033
<i>Αιτιολογικοί Παράμετροι</i>				
Ερυθρά	8,08 ± 0,25	8,36 ± 0,45	7,93 ± 0,33	8,8 ± 0,45
Αιμοπετάλια	1115 ± 166	1123 ± 284	864 ± 220	820 ± 127
Λευκοκύτταρα	6,3 ± 1,7	7,3 ± 3	6 ± 1,3	6,1 ± 2,5
Λεμφοκύτταρα	85 ± 2,9	73,5 ± 11,4	81,7 ± 7,3	85 ± 4,9
Ουδετερόφιλα	9,5 ± 4,1	16,4 ± 13	6,9 ± 2,2	7,2 ± 1,3
Μονοκύτταρα	3,3 ± 3,4	7,5 ± 4,4	8 ± 6,4	4,8 ± 2,7
Ηωσινόφιλα	2 ± 1,6	2 ± 1,1	2,8 ± 2,7	3 ± 2,9
Βλαστόφιλα	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Πίνακας 2.1: Πίνακας παραμέτρων βάρους και αιτιολογικών παραμέτρων.

Στους πίνακες 2.1 και 2.2 φαίνονται μερικές χρήσιμες παράμετροι που χαρακτηρίζουν τα σωματομετρικά δεδομένα, τα αιματολογικά στοιχεία, όπως επίσης τις φυσιολογικές τιμές μερικών βιοχημικών μεταβλητών του ποντικού NMRI. Οι τιμές

Παράμετρος	Άρρενα	Άρρενα	Θήλεα	Θήλεα
	6 εβδομάδων	12 εβδομάδων	6 εβδομάδων	12 εβδομάδων
<i>Βιοχημικοί Παράμετροι</i>				
Νάτριο	148 ± 1	148 ± 2	148 ± 1	149 ± 1
Κάλιο	3,8 ± 1,1	3,8 ± 0,3	3,2 ± 0,2	3,7 pm 0,6
Ασβέστιο	2,59 ± 0,23	2,46 ± 0,04	2,62 ± 0,07	2,56 ± 0,05
Φωσφορικά	2,19 ± 0,14	1,5 ± 0,2	2,16 ± 0,1	1,31 ± 0,3
Ουρία	7,7 ± 0,7	9,4 ± 1,5	6,3 ± 0,9	6,7 ± 0,9
Κρεατινίνη	27 ± 2	28 ± 2	28 ± 2	32 ± 2
Χολερυθρίνη	<10	<10	<10	<10
Χοληστερίνη	4,3 ± 0,5	3,9 ± 0,5	3,2 ± 0,5	3,8 ± 0,5
Τριγλυκερίδια	1,7 ± 0,16	1,67 ± 0,22	1,46 ± 0,43	2,58 ± 1,16
Γλυκόζη	12,8 ± 1,3	11 ± 1,2	11,3 ± 1,2	11,4 ± 1,1
AP	193 ± 20	71 ± 13	211 ± 45	125 ± 24
ASAT	73 ± 21	76 ± 13	80 ± 19	82 ± 18
ALAT	40 ± 5	43 ± 15	38 ± 20	38 ± 5
LD	532 ± 130	568 ± 139	485 ± 62	518 ± 170
Αιμοσφαιρίνη	8,5 ± 0,2	8,3 ± 0,4	8,3 ± 0,3	8,9 ± 0,5
Αιματοκρίτης	4,6 ± 1	4,6 ± 1	45 ± 2	46 ± 2

Πίνακας 2.2: Πίνακας βιοχημικών παραμέτρων.

στους πίνακες αυτούς αντιστοιχούν σε τιμές στο μετρικό σύστημα SI. Πιο συγκεκριμένα, οι βαρομετρικοί παράμετροι είναι σε γραμμάρια (grams), οι αιτιολογικοί παράμετροι είναι επί τοις εκατό εκτός από τα ερυθρά, τα αιμοπετάλια και λευκοκύτταρα που είναι σε  $10^{12}$  ανά λίτρο,  $10^9$  ανά λίτρο και  $10^9$  ανά λίτρο αντίστοιχα. Όσο αφορά τις βιοχημικούς παραμέτρους, η μονάδα μέτρησης είναι το γραμμαμόριο ανά λίτρο (mmol/l) εκτός από την γλυκόζη, τον αιματοκρίτη και τα ακρόνυμα AP, ASAT, ALAT και LD. Η γλυκόζη και τα προαναφερθέντα ακρόνυμα έχουν ως μονάδα μέτρησης την μονάδα ανά λίτρο (units/litre) ενώ για τον αιματοκρίτη έχουμε την επί τοις εκατό μέτρηση. Τα βάρη του σώματος, όπως επίσης και των αναφερόμενων σπλάχνων αντιστοιχούν στο νεαρό ζώο που έχει περατώσει φυσιολογικά την ανάπτυξή του και είναι ώριμο για αναπαραγωγή. Όλες οι τιμές των παραμέτρων

που φαίνονται στον πίνακα αφορούν τα εν λόγω ζώα, υπό φυσιολογικές συνθήκες και υπό απουσία κακουχιών.



# ΚΕΦΑΛΑΙΟ 3

## ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΣ

---

### 3.1 Πειραματόζωα

### 3.2 Πορεία του πειράματος

---

Σε αυτό το κεφάλαιο επεξηγείται πώς επιτυγχάθηκε η ομογενοποίηση του πληθυσμού των ποντικών. Επεκτείνεται έως τον ειδικό χειρισμό και την φροντίδα των ποντικών σε όλη την διάρκεια των πειραμάτων. Δίνονται λεπτομέρειες των πειραμάτων και το κεφάλαιο κλείνει με μία αναφορά στις τρεις βασικές ομάδες δραστικών ουσιών που δοκιμάστηκαν καθώς και σε πληροφορίες του πληθυσμού για κάθε δραστική ουσία ξεχωριστά και υπογραμμίζεται ο βασικός θεωρητικός άξονας των πειραμάτων.

### 3.1 Πειραματόζωα

#### 3.1.1 Τύπος, Ποικιλία και Ομογενοποίηση

Για την πραγματοποίηση των εν λόγω πειραμάτων χρησιμοποιήθηκαν 289 πειραματόζωα του μυός NMRI (Naval Medical Research Institute) για το πείραμα που αντιστοιχεί στην μονή δόση και επιπλέον 149 πειραματόζωα του μυός NMRI για το πείραμα που αντιστοιχεί στην διπλή δόση. Ειδικά για τα συγκεκριμένα πειράματα επιλέχθηκαν τα θήλεα ζώα, σύμφωνα με το ακολουθούμενο μοντέλο [5]. Τέσσερα

ζεύγη ποντικών NMRI ελήφθησαν από το Εκτροφείο Πειραματοζώων του Πανεπιστημίου Ιωαννίνων και εγκαταστάθηκαν σε κατάλληλο χώρο του Εργαστηρίου Πειραματικής Φυσιολογίας, ο οποίος είχε προδιαγραφεί έτσι ώστε να λειτουργεί ως εκτροφείο μικρών πειραματοζώων, υπό ελεγχόμενες εργαστηριακές συνθήκες. Από τα εν λόγω τέσσερα ζεύγη ζώων προέκυψε, μέσω κατάλληλων διασταυρώσεων, αιμομικτικού χαρακτήρα, όλος ο πληθυσμός των ποντικών NMRI, που χρησιμοποιήθηκε για την πραγματοποίηση των πειραμάτων μας. Τα προκύψαντα πειραματόζωα ήταν ομοιογενή από την πλευρά της γενετικής τους δομής, και συνεπώς ο εν λόγω πληθυσμός παρουσίαζε μεγάλη ομοιομορφία ως προς το προσδόκιμο επιβίωσης (διάρκεια ζωής), την αυθόρμητη παθολογία, τις αιτίες θανάτου, τα σωματικά μεγέθη (μήκος - βάρος), το μέγεθος των οργάνων, τις βιοχημικές παραμέτρους, τις αιματολογικές παραμέτρους και εν τα εν γένει βιολογικά χαρακτηριστικά των ποντικών. Συνεπώς χρησιμοποιήσαμε έναν γενετικά ομοιόμορφο πληθυσμό που εμφάνιζε όμοιες αντιδράσεις στον μεταβολισμό του χημικού καρκινογόνου [6].

### **3.1.2 Τοποθέτηση των ζώων και φροντίδα αυτών**

Τα ποντικοί NMRI, μετά την γέννηση και τον απογαλακτισμό τους, τοποθετούντο σε μεγάλους κλωβούς που περιείχαν συνολικά δέκα άτομα εξ αυτών, πάντα του ίδιου φύλου (άρρενα ή θήλεα). Τα νεαρά ζώα εκτρεφόταν σε αυτούς τους μεγάλους κλωβούς των δέκα ατόμων μέχρις ότου κλείσουν την ηλικία των 60 ημερών. Τα ζώα που χρησιμοποιούνται για τον εκάστοτε πειραματισμό ήταν της αυτής ηλικίας (ημερομηνίες γέννησης περιλαμβανόμενες σε χρονικό διάστημα λίγων ημερών). Οι ποντικοί NMRI ζυγίζονταν και ανά τρία άτομα του ίδιου φύλου τοποθετούνταν σε κλωβούς από πλεξιγκλάς μεγεθών 40 cm x 30 cm x 15 cm. Οι κλωβοί εντός των οποίων διέμεναν τα πειραματόζωα περιείχαν στο εσωτερικό τους μια στρωμνή από αδρά ροκανίδια με υψηλή απορροφητική ικανότητα ώστε τα απεκκρινόμενα ούρα να μην δημιουργούν στρώμα υγρασίας. Η αλλαγή της στρωμνής των πειραματοζώων γινόταν ανά τέσσερες ημέρες. Η πλύση κάθε κλωβού και η θερμική του απολύμανση γινόταν μία φορά ανά δύο εβδομάδες. Οι κλωβοί καλυπτόταν από μεταλλική γρίλια αποτελούμενη από κυλινδρικούς ανοξείδωτους χαλύβδινους ράβδους διαμέτρου 3 mm, με διάκενα μεταξύ τους ίσα προς 9 mm, ώστε να μην είναι δυνατή η έξοδος του ζώου από τον κλωβό. Επίσης η γρίλια του κλωβού περιείχε κατάλληλη εκβάθυνση στην οποία τοποθετούταν η τροφή τους, υπό την μορφή συμπιεσμένων ξηρών

βόλων. Τα ζώα μπορούσαν να παίρνουν την τροφή τους, δαγκώνοντας και τραβώντας με τα άνω άκρα και τις σιαγόνες τους τον κάθε βόλο. Η τροφή κάθε τριμελούς κλωβού ζυγιζόταν πριν την τοποθέτησή της και μετά από τρεις ημέρες ζυγιζόταν εκ νέου και καταγραφόταν η κατανάλωση της. Επίσης γινόταν συμπλήρωση της ώστε ανά τρεις ημέρες να υπάρχει σταθερό στοκ βάρους 200 γραμμαρίων βρώσιμων βόλων ανά κλωβό. Η ημερήσια κατανάλωση ανά πειραματόζωο ήταν περί τα  $10 \pm 0,5$  gr. Το καταναλισκόμενο ύδωρ λαμβανόταν από τον αστρικό αγωγό υδροδότησης και προστίθετο ελάχιστη ποσότητα αιθανολικού διαλύματος θυμόλης ώστε να μην δημιουργούνται αποικίες μικροβίων στην φιάλη του ύδατος. Σε κάθε κλωβό τοποθετούνταν δύο θήλαστρα με ύδωρ, ώστε να μην υπάρχει ανταγωνισμός μεταξύ των ζώων για την λήψη ύδατος. Η ημερήσια κατανάλωση νερού ανά πειραματόζωο ήταν ίση προς  $18 \pm 0,9$  ml/ημέρα. Γινόταν ζύγιση των ζώων ανά εβδομάδα και έλεγχος του σώματός τους οπτικά και διά ψηλάφησης. Επίσης στην έναρξη του πειράματος έγινε αιμοληψία για αιματολογικούς και βιοχημικούς ελέγχους. Αιμοληψία έγινε επίσης σε μεγάλη ηλικία (6 μηνών) για επανάληψη του αιματολογικού και βιοχημικού ελέγχου. Η αιμοληψία έγινε με ήπιο τρόπο και η ποσότητα αίματος ήταν πολύ μικρή, ώστε να μην διαταραχθεί η κυκλοφορική ισορροπία του κάθε ζώου, ούτε αυτό να υποστεί ισχυρό stress.

### 3.1.3 Χειρουργικοί χειρισμοί

Για την χορήγηση του καρκινογόνου ακολουθήσαμε το εξής πρωτόκολλο : Η χορήγηση του καρκινογόνου στα πειραματόζωα γινόταν χειρουργικά, υπό αναισθησία. Η αναισθητοποίηση του ζώου γινόταν μέσω ενδοπεριτοναϊκής, εφ' άπαξ, χορήγησης ενός ηρεμιστικού (μιδαζολάμη 3 mgr/kg b.w) και ενός αναισθητικού (κεταμίνη 3 mgr/kg b.w). Εφόσον το πειραματόζωο είχε αναισθητοποιηθεί γινόταν τοπική αντησηψία με βάμμα ιωδίου και υπεροξειδίου του υδρογόνου (οξυζενέ) και κατόπιν μικρή τομή ολικού πάχους δέρματος, ώστε να αποκαλυφθούν οι υποκείμενες μυϊκές μάζες στην περιοχή της δεξιάς ωμοπλάτης. Εν συνεχεία ενιόταν σταθερή ποσότητα 3,4-βενζοπυρενίου (2,52 mgr/ml), διαλυμένου σε τρικαπριλίνη, σε τρόπο ώστε να εισαχθεί κάτω από την μυϊκή περιτοναία εξ ολοκλήρου. Κατόπιν γινόταν πολύ βραδεία απόσυρση της βελόνας, ώστε να μην εκδιωχθεί προς τα έξω το ενεθέν διάλυμα του καρκινογόνου, αλλά να διαχυθεί στην περιοχή των ραχιαίων μυών της ωμοπλάτης. Η τομή συρράπτετο ταχέως και στεγανοποιείτο ισχυρώς ώστε να μην

υπάρχει διαρροή του καρκινογόνου διαλύματος εκτός του σώματος του ζώου. Μετά την χορήγηση του καρκινογόνου το πειραματόζωο ετίθετο μονήρες σε κλωβό, υπό παρακολούθηση, προκειμένου να ανανήψει πλήρως. Μετά την ανάνηψη του τοποθετείτο ξανά στον κλωβό από τον οποίο είχε ληφθεί.

### **3.1.4 Παρακολούθηση των πειραματοζώων**

Η παρακολούθηση των πειραματοζώων γινόταν τρεις φορές την ημέρα καθημερινά. Το πρωί στις 8 π.μ., το μεσημέρι στις 16:00 και το βράδυ στις 20:00. Τα πειραματόζωα επιβλεπόταν ένα προς ένα στους κλωβούς τους. Κάθε πειραματόζωο είχε ένα διακριτικό σημείο στην ουρά του, προκειμένου να γνωρίζουμε ποια είναι η ταυτότητά του ως κωδικός αριθμός. Το σύστημα κωδικοποίησης των πειραματοζώων περιλάμβανε τον αριθμό του κλωβού και τον αριθμό του ζώου προς παρακολούθηση. Ο αριθμός του ζώου σημειωνόταν με τον σχεδιασμό ενός κύκλου στην ουρά (αριθμός 1), δύο κύκλων (αριθμός 2), ενώ το τρίτο ζώο δεν έφερε κανένα σημείο, πράγμα που αντιστοιχούσε στον αριθμό 3. Π.χ για τον κλωβό K1 οι κωδικοί των ζώων του ήσαν KZ1, KZ2, KZ3, που σήμαινε αντίστοιχα «Κλωβός 1 – ζώο 1ο, Κλωβός 1 – ζώο 2, Κλωβός 1 – ζώο 3, ενώ για τον κλωβό K2, οι κωδικοί ήσαν K2Z1, K2Z2, K2Z3. Το ίδιο για τους υπόλοιπους κλωβούς.

## **3.2 Πορεία του πειράματος**

### **3.2.1 Εξέταση των ζώων**

Τα πειραματόζωα σε συγκεκριμένες χρονικές στιγμές, από την χορήγηση του καρκινογόνου, αναισθητοποιούνταν και πραγματοποιούσαμε υπερηχογραφικό τομογραφικό έλεγχο της περιοχής της ωμοπλάτης με δυνατότητα υπολογισμού του μεγέθους ενός όγκου και της αναγωγής του σε μονάδες βάρους (mgr, gr). Όταν το μέγεθος του όγκου αυξανόταν σημαντικά και η κατάσταση του πειραματόζωου επιβαρυνόταν, τότε γινόταν αλλαγή κλωβού λίγο πριν τον επικείμενο θάνατο, ώστε να μην κανιβαλιστεί το πτώμα, από τα άλλα ζώα που συγκατοικούσαν στον κλωβό. Με την επέλευση του θανάτου γινόταν ζύγιση και νεκροτομή. Ο όγκος παρασκευαζόταν ώστε να αφαιρεθεί πλήρως, ενώ όλα τα εσωτερικά όργανα αφαιρούνταν, ζυγίζόταν, παρατηρούσαν με μεγεθυντικό φακό ή στερεοσκόπιο, και φωτογραφιζόταν. Ο

όγκος και τα όργανα τυλιγόταν σε ένα τμήμα απορροφητικού χάρτου και κατόπιν τοποθετούνταν σε διάλυμα φορμαλδεΐδης 8 τοις εκατό, προκειμένου να μονιμοποιηθούν και κατόπιν να φτιαχτούν παρασκευάσματα διαφόρων τομών για μικροσκοπική ιστοπαθολογική διάγνωση. Η μικροσκόπηση και η ιστοπαθολογική διάγνωση, όπως επίσης η καταγραφή των μεταστάσεων, γινόταν με λεπτομερή μικροσκοπική παρατήρηση από εξειδικευμένο παθολογοανατόμο. Η επισκόπηση του πτώματος και οι τομές ήταν σταθερές ώστε να μπορούν να συγκριθούν μεταξύ τους, ως προς την ένταση και την έκταση της νεοπλασματικής νόσου. Η παθολογοανατομική μελέτη και ο χαρακτηρισμός των βιολογικών ευρημάτων (πρωτοπαθής όγκος, μέγεθος, βάρος αυτού, υφή, δομή, μεταστάσεις) ήταν κρίσιμης σημασίας για τον έλεγχο της αποτελεσματικότητας των αντικαρκινογόνων ουσιών. Στην ανά χείρας μελέτη οι κρίσιμες παράμετροι που ελήφθησαν υπ όψιν ήταν:

1. Η ανάπτυξη ή όχι κακοήθους όγκου.
2. Το μέγεθος του πρωτοπαθούς όγκου.
3. Ο ρυθμός ανάπτυξης του πρωτοπαθούς όγκου.

Η ποσότητα της αντικαρκινογόνου ουσίας που επωαζόταν με συγκεκριμένη ποσότητα βενζοπυρενίου και προκαλούσε συγκεκριμένα επίπεδα αναστολής, ήταν μια παράμετρος που ελήφθη υπ' όψιν, προκειμένου να βρεθεί αιτιακή συσχέτιση αναστολής με γνώμονα την σχέση βαρών καρκινογόνου και αντικαρκινογόνου (αναστέλλουσα ουσία).

## 3.2.2 Ομάδες πειραματοζώων

### 3.2.2.1 Βασικός θεωρητικός άξονας

Η πειραματική προσέγγιση που εμείς επιχειρήσαμε με τα πειράματά μας της χημικής καρκινογένεσης στους ποντικούς NMRI μέσω B[a]P ήταν σε μεγάλο βαθμό χημικού τύπου, καθόσον προσπαθήσαμε να καταστρέψουμε, ή να παραλλάξουμε την χημική δομή του B[a]P, επωάζοντας μια ποσότητά του με συγκεκριμένες ποσότητες διαφόρων ουσιών, οι οποίες θεωρητικά θα μπορούσαν να αντιδράσουν μαζί του και συγκεκριμένα να κάνουν αντιδράσεις προσθήκης στους διπλούς του δεσμούς. Η λογική αυτού του πειραματισμού έδειχνε ότι με αυτές τις αντιδράσεις θα ήταν δυνατή η αναγωγή των διπλών δεσμών και κυρίως αυτών της περιοχής L και της περιοχής K, οπότε το B[a]P θα αποκτούσε μια χημική δομή ανάλογη του χρυσενίου, το οποίο είναι πολυκυκλικός αρωματικός υδρογονάνθρακας, αλλά πολύ

πιο ασθενής ως καρκινογόνο μόριο σε σχέση με το B[a]P.

Συνεπώς στα εν λόγω πειράματα δοκιμάστηκε η αντοχή της θεωρίας της δομής των πολυκυκλικών αρωματικών υδρογονανθράκων, με κύρια σκέψη «αν μπορούσαμε να αλλοιώσουμε την δομή τους σε ορισμένα σημεία θα μπορούσαμε να άρουμε και την καρκινογόνο ισχύ αυτών».

### 3.2.2.2 Γενικές Πληροφορίες

Η χημική καρκινογένεση με το 3,4-βενζοπυρένιο πραγματοποιήθηκε με το εξής πρωτόκολλο : Ποσότητα 3,4-βενζοπυρενίου ίση προς 2,52 mgr διαλυόταν σε 0,5 ml τρικαπριλίνης. Η εν λόγω ποσότητα προκαλούσε καρκινογένεση στο 100 τοις εκατό των πειραματοζώων. Τα πειραματόζωα (μύες NMRI) ανέπτυσαν όγκο μέσα σε συγκεκριμένο χρονικό διάστημα μικρότερο των 60 ημερών. Σε αρκετές περιπτώσεις υπήρξαν κακοήγη μορφώματα σε χρόνο 20 ημερών από την έγχυση του βενζοπυρενίου. Οι όγκοι, μετά την εμφάνισή τους, αναπτυσσόταν ραγδαία, και μετά την προσέγγιση ενός οριακού μεγέθους προκαλούσαν τον θάνατο του ασθενούς ποντικού. Το χρονικό διάστημα όπου απέθνησκε το σύνολο των ζώων κυμαινόταν από 120 ημέρες έως 190 ημέρες – εξαιρουμένων ελάχιστων (δύο) περιπτώσεων. Το μοντέλο αυτό επιβεβαιώθηκε σε μεγάλους αριθμούς ζώων (συνολικά 6.000), ωστόσο εδώ χρησιμοποιούμε έναν αριθμό κοντά στα 50 ζώα ως ομάδα ελέγχου. Βάσει αυτών των δεδομένων τα πειράματα σχεδιάστηκαν ως εξής :

1. Ως διαλύτης του 3,4-βενζοπυρενίου χρησιμοποιήθηκε η τρικαπριλίνη. Ο όγκος διασποράς του διαλύτη ήταν μικρός (2,52 mgr διαλυόμενα σε 0,5 – 1,0 ml τρικαπριλίνης.

2. Το διάλυμα του 3,4-βενζοπυρενίου στην τρικαπριλίνη παρέμενε σε ένα ποτήρι ζέσεως στεγανό επί 12 ώρες σε θερμοκρασία δωματίου (20°C) σε εν λειτουργία απαγωγό εστία χημικής προστασίας και κατόπιν αναρροφείτο με γυάλινη 19 σύριγγα και ενιόταν στο πειραματόζωο υπό νάρκωση, όπως περιγράψαμε ανωτέρω.

3. Η προετοιμασία της εξουδετέρωσης του 3,4 – βενζοπυρενίου (συντομογραφικά B[a]P) γινόταν ως εξής : ποσότητα 2,52 mgr B[a]P (και ακέραια πολλαπλάσιά της, ίσα με τον εκάστοτε χρησιμοποιούμενο αριθμό πειραματοζώων, που επιλεγόταν ως απαραίτητα για την στατιστική αξιοπιστία κάθε πειραματισμού), ετίθετο με συγκεκριμένη ποσότητα της ουσίας, της οποίας μελετούσαμε την ενδεχόμενη αντικαρκινογόνα δράση της. Οι ποσότητες του καρκινογόνου και του αντικαρκινογόνου διαλυόταν επιμελώς σε γυάλινο ποτήριο ζέσεως και γινόταν μαγνητική ανάδευση επί

δωδεκάωρο (over night), υπό κανονικές συνθήκες (θερμοκρασία 20°C, δωδεκάωρη φωτοπερίοδος, φωτισμός ηλιακού φωτός, καθαρή ατμόσφαιρα απουσία οξειδωτικών και διαβρωτικών ατμιζόντων ουσιών και διαλυτών). Το ποτήριο ζέσεως ήταν αεροστεγώς πωματισμένο και μη ατμίζον, καθ όλη την διάρκεια του πειραματισμού, για την προστασία του προσωπικού και την μη ρύπανση των πειραματικών χώρων. Η επιμελής ανάμιξη και ανάδευση του καρκινογόνου με το υπό μελέτη αντικαρκινογόνο ήταν δυνατόν να δημιουργήσει αντιδράσεις μεταξύ αυτών ώστε η καρκινογόνος ισχύς του B[a]P να μειωθεί. Αυτό ήταν αναμενόμενο, αλλά πειράματα χημικού χαρακτήρα – που δεν περιλαμβάνονται στο ανά χείρας κείμενο – έδειξαν ότι δεν υπήρχαν τέτοιου είδους ποσοτικές αντιδράσεις ικανές να αποικοδομήσουν χημικά το 3,4-βενζοπυρένιο. Οπότε μπορούμε να θεωρήσουμε ότι το μείγμα των δύο ουσιών (βενζοπυρένιο + μελετούμενο αντικαρκινογόνο) λειτουργούσε σε βιολογικό επίπεδο στο εσωτερικό του πειραματόζωου, μετά την χειρουργική έγχυση του μείγματός τους. Διευκρινίζουμε ότι στην παρούσα μελέτη ερευνήσαμε αντικαρκινογόνες δράσεις και όχι δράσεις αντικαρκινικές. Ως αντικαρκινογόνες ονομάζονται οι δράσεις εκείνες οι οποίες αναστέλλουν την καρκινογένεση. Αντιθέτως ως αντικαρκινικές θεωρούνται οι δράσεις εκείνες οι οποίες καταστέλλουν η αναστέλλουν την κακοήγη νεοπλασματική νόσο. Ο όρος «καρκίνος» είναι εσφαλμένα γενικευμένος διότι ως καρκίνοι θεωρούνται τα κακοήγη νεοπλάσματα που προκύπτουν από επιθηλιακούς ιστούς, ενώ τα κακοήγη νεοπλάσματα τα προκύπτοντα από μεσεγχυματικούς ιστούς ονομάζονται «σαρκώματα». Οι ομάδες που χρησιμοποιήθηκαν για την εν λόγω μελέτη ήσαν οι εξής :

#### 1. Ομάδα Ελέγχου

Η ομάδα ελέγχου αποτελείτο από πενήντα (49) μύες NMRI. Η ομάδα αυτή ήταν κοινή για όλες τις λοιπές πειραματικές ομάδες. Για κάθε πειραματόζωο ζυγίστηκαν 2,52 mgr B[a]P (συνολικά 126 mgr για τους 49 μύες). Στην εν λόγω ομάδα ενέθηκε η δόση του B[a]P, με την χειρουργική μέθοδο που προαναφέραμε. Τα ζώα παρακολουθούντο όπως περιγράψαμε ανωτέρω μέχρι τον θάνατό τους και αμέσως κατόπιν ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, μακροσκοπική επισκόπηση, ζύγιση του νεοπλάσματος και των σπλάχνων κάθε ζώου, μονιμοποίηση και ιστολογική διάγνωση επιλεγμένων τομών του νεοπλάσματος και των οργάνων τους (έλεγχος μεταστάσεων).

2. Ομάδα αναστολής του B[a]P μέσω αιθυλενοδιαμίνης (ETHDMN-1) Η ομάδα αυτή αποτελείτο από δεκαεννέα (19) μύες NMRI για την μονή δόση και τριάντα

(30) μύες NMRI για την διπλή δόση. Για κάθε πειραματόζωο 10 mgr ETHDMN-1 ζυγίστηκαν με ζυγό ακριβείας (μέχρι τέταρτου δεκαδικού ψηφίου του χιλιοστογραμμαρίου - mgr), διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 320 mgr αιθυλενοδιαμίνης και 80,64 mgr B[a]P, τα οποία διαλύθηκαν σε 32 ml τρικαπριλίνης. Μετά την επώαση ενέθηκε 1 ml διαλύματος σε κάθε πειραματόζωο, με τον χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο όπως περιγράψαμε ανωτέρω μέχρι τον θάνατό τους και αμέσως κατόπιν ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, ζύγιση του νεοπλάσματος και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και ιστολογική διάγνωση επιλεγμένων τομών του νεοπλάσματος και των οργάνων τους (έλεγχος μεταστάσεων).

3. Ομάδα αναστολής του B[a]P μέσω κανταβερίνης (CDVRN-2) Η ομάδα αυτή περιλάμβανε τριάντα (30) μύες NMRI για την μονή δόση και πενήντα (50) μύες NMRI για την διπλή δόση. Για κάθε πειραματόζωο 10 mgr κανταβερίνης (πτωμαΐνης, CDVRN-2) ζυγίστηκαν με ζυγό ακριβείας (μέχρι τέταρτου δεκαδικού ψηφίου του χιλιοστογραμμαρίου - mgr), διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 500 mgr κανταβερίνης και 126 mgr B[a]P, τα οποία διαλύθηκαν σε 50 ml τρικαπριλίνης. Μετά την επώαση ενέθηκε 1 ml διαλύματος σε κάθε πειραματόζωο, με τον χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο, όπως περιγράψαμε ανωτέρω, μέχρι τον θάνατό τους και αμέσως κατόπιν ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, ζύγιση του νεοπλάσματος και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και ιστολογική διάγνωση επιλεγμένων τομών του νεοπλάσματος και των οργάνων τους (έλεγχος μεταστάσεων).

4. Ομάδα αναστολής του B[a]P μέσω πουτρεσκίνης (PTRSCN-3) Η ομάδα αυτή περιλάμβανε σαραντατρείς (43) μύες NMRI μονής δόσης και πενήντα (50) μύες διπλής δόσης. Για κάθε πειραματόζωο 10 mgr πουτρεσκίνης (PTRSCN-3) ζυγίστηκαν με ζυγό ακριβείας (μέχρι τέταρτου δεκαδικού ψηφίου του χιλιοστογραμμαρίου - mgr), διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 400 mgr πουτρεσκίνης και 100,8 mgr



B[a]P, τα οποία διαλύθηκαν σε 40 ml τρικαπριλίνης. Μετά την επώαση ενέθηκε, *ad arte medica*, 1 ml διαλύματος σε κάθε πειραματόζωο, με τον χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο, όπως περιγράψαμε ανωτέρω, μέχρι τον θάνατό τους και αμέσως κατόπιν ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση του νεοπλασματικού όγκου, εκτομή όλων των σπλάχνων ζύγιση του νεοπλασματος και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και μικροσκοπική ιστολογική διάγνωση επιλεγμένων τομών του νεοπλάσματος και των οργάνων τους (έλεγχος μεταστάσεων).

5. Ομάδα βισμούθειόλης-I (Bismuthiol-I) Η ομάδα αυτή περιλάμβανε τριάντα (30) μύες NMRI μονής δόσης. Για κάθε πειραματόζωο ζυγίστηκαν με ζυγό ακριβείας 10 mgr βισμούθειόλης-I, διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 300 mgr βισμούθειόλης-I και 75,6 mgr B[a]P, τα οποία διαλύθηκαν σε 30 ml τρικαπριλίνης. Μετά την επώαση ενέθηκε 1 ml διαλύματος σε κάθε πειραματόζωο, με τον χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο, όπως περιγράψαμε ανωτέρω, μέχρι τον θάνατό τους και όταν αυτός επήλθε, αμέσως ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, ζύγιση του νεοπλασματικού όγκου καθώς και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και ιστολογική διάγνωση τομών του νεοπλάσματος και των εσωτερικών οργάνων του κάθε πειραματόζωου (έλεγχος μεταστάσεων).

6. Ομάδα L-κυστεΐνης (L-Cystein) Η ομάδα αυτή περιλάμβανε τριάντα (30) μύες NMRI μονής δόσης. Για κάθε πειραματόζωο ζυγίστηκαν με ζυγό ακριβείας 10 mgr L-κυστεΐνης, διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 300 mgr L-κυστεΐνης και 75,6 mgr B[a]P, τα οποία διαλύθηκαν σε 30 ml τρικαπριλίνης. Μετά την επώαση ενέθηκε 1 ml διαλύματος σε κάθε πειραματόζωο, με τον *ad arte medica* χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο, όπως περιγράψαμε ανωτέρω, μέχρι τον θάνατό τους και όταν αυτός επήλθε, αμέσως ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, ζύγιση του νεοπλασματικού όγκου καθώς και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και ιστολογική διάγνωση τομών του νεοπλάσματος και των εσωτερικών οργάνων του κάθε πειραματοζώου (έλεγχος μεταστάσεων).

7. Ομάδα αιθυλικού εστέρος της L-κυστεΐνης (L-Cystein-Ethylester) Η ομάδα αυτή περιλάμβανε είκοσι οχτώ (28) μύες NMRI μονής δόσης. Για κάθε πειραματόζωο ζυγίστηκαν με ζυγό ακριβείας 10 mgr αιθυλικού εστέρα της L-κυστεΐνης, διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 280 mgr αιθυλικού εστέρα της L-κυστεΐνης και 70,56 mgr B[a]P, τα οποία διαλύθηκαν σε 28 ml τρικαπριλίνης. Μετά την επώαση ενέθηκε 1 ml διαλύματος σε κάθε πειραματόζωο, με τον ad arte medica χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο, όπως περιγράψαμε ανωτέρω, μέχρι τον θάνατό τους και όταν αυτός επήλθε, αμέσως ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, ζύγιση του νεοπλασματικού όγκου καθώς και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και, στο μικροσκόπιο, ιστολογική διάγνωση επιλεγμένων τομών του νεοπλάσματος και των εσωτερικών οργάνων του κάθε πειραματόζωου (έλεγχος μεταστάσεων).

8. Ομάδα μερκαπτοηλεκτρικού οξέος (Mercaptosuccinic Acid) Η ομάδα αυτή περιλάμβανε τριάντα (30) μύες NMRI μονής δόσης. Για κάθε πειραματόζωο ζυγίστηκαν με ζυγό ακριβείας 10 mgr μερκαπτοηλεκτρικού οξέος, διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 300 mgr μερκαπτοηλεκτρικού οξέος και 75,6 mgr B[a]P, τα οποία διαλύθηκαν σε 30 ml τρικαπριλίνης. Μετά την επώαση ενέθηκε 1 ml διαλύματος σε κάθε πειραματόζωο, με τον ad arte medica χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο, όπως ήδη περιγράψαμε, μέχρι τον θάνατό τους και όταν αυτός επήλθε, αμέσως ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, ζύγιση του νεοπλασματικού όγκου καθώς και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και, στο μικροσκόπιο, ιστολογική διάγνωση επιλεγμένων τομών του νεοπλάσματος και των εσωτερικών οργάνων του κάθε πειραματόζωου (έλεγχος μεταστάσεων).

9. Ομάδα διθειοθρεϊτόλης (Dithiothreitol) Η ομάδα αυτή περιλάμβανε τριάντα (30) μύες NMRI μονής δόσης. Για κάθε πειραματόζωο ζυγίστηκαν με ζυγό ακριβείας 10 mgr διθειοθρεϊτόλης, διαλύθηκαν σε τρικαπριλίνη μαζί με 2,52 mgr B[a]P και επώαστηκαν επί 12 ώρες, όπως περιγράψαμε ανωτέρω. Η διάλυση για κάθε πειραματόζωο έγινε σε 1 ml τρικαπριλίνης. Συνολικά ζυγίσαμε 300 mgr διθειοθρεϊτόλης και 75, 6 mgr B[a]P, τα οποία διαλύθηκαν σε 30 ml τρικαπριλίνης. Μετά την

επώαση ενέθηκε 1 ml διαλύματος σε κάθε πειραματόζωο, με τον ad arte medica χειρουργικό τρόπο που περιγράψαμε ανωτέρω. Τα ζώα παρακολουθούντο, όπως περιγράψαμε ανωτέρω, μέχρι τον θάνατό τους και όταν αυτός επήλθε, αμέσως ζυγίστηκαν, νεκροτομήθηκαν, έγινε απομόνωση, ζύγιση του νεοπλασματικού όγκου καθώς και των σπλάχνων κάθε ζώου, μακροσκοπική επισκόπηση όγκου και σπλάχνων, μονιμοποίηση και, στο μικροσκόπιο, ιστολογική διάγνωση επιλεγμένων τομών του νεοπλάσματος και των εσωτερικών οργάνων του κάθε πειραματόζωου (έλεγχος μεταστάσεων).

Τα παραπάνω δεδομένα απαρτίζουν τον πίνακα 3.1 ενώ αξίζει να αναφέρουμε ότι η ομάδα των δραστικών ουσιών αιθυλενοδιαμίνη, κανταβερίνη και πουτρεσκίνη ανήκουν στην οικογένεια των πολυαμινών ενώ η ομάδα των δραστικών ουσιών βισμούθειόλη, L-κυστεΐνη, μερκαπτοηλεκτρικό οξύ, διθειοθρεϊτόλη και αιθυλικού εστέρος της L-κυστεΐνης ανήκουν στην οικογένεια των θειολών.

<i>Substance Family</i>	<i>Substance Name</i>	<i>Single Dosage Obs.</i>	<i>Double Dosage Obs.</i>
Carcinogen	Benzopyrene	49	-
Polyamines	Ethylenediamine	19	49
	Cadaverine	30	50
	Putrescine	43	50
	Total	92	149
Thiols	Bismuthiol	30	-
	L-Cysteine Ethyl Ester	28	-
	Mercaptosuccinic Acid	30	-
	Dithiothreitol	30	-
	L-Cysteine	30	-
	Total	148	-

Πίνακας 3.1: Αριθμοί πειραματοζώων ανά δραστική ουσία μονής και διπλής δόσης.

### 3.2.2.3 Πολυαμίνες

Οι πολυαμίνες είναι ουσίες που παράγονται στον φυσιολογικό μεταβολισμό των αμινοξέων L-αργινίνη και L-ορνιθίνη και φαίνεται ότι προστατεύουν το DNA. Πιο συγκεκριμένα, είναι ουσίες που ανήκουν στον μικρόκοσμο και είναι απαραίτητες για

την φυσιολογική εξέλιξη των κυττάρων σε ευκαριώτες. Τόσο η κατανομή φορτίου όσο και η ευελιξία τις επιτρέπουν να προσδένονται σε ανιονικά μακρομόρια όπως είναι το RNA, DNA, πρωτεΐνες και φωσφολιπίδια. Η ζωτική σημασία των πολυαμινών για φυσιολογικές κυτταρικές διεργασίες τονίζεται περαιτέρω από τους ρόλους τους στην διατήρηση της δομής της χρωματίνης, στη ρύθμιση των καναλιών ιόντων, στη διατήρηση της σταθερότητας της μεμβράνης και στην απομάκρυνση ελεύθερων ριζών.

Υπό φυσιολογικές συνθήκες, οι πολυαμίνες ρυθμίζονται αυστηρά από ένα πολύπλοκο δίκτυο βιοσυνθετικών, καταβολικών και μη κατανοητών μηχανισμών μεταφοράς (δεν είναι γνωστοί όλοι οι παράμετροι μεταφοράς). Τέτοιοι έλεγχοι βοηθούν στη διατήρηση μίας συγκέντρωσης πολυαμινών στο κύτταρο, αν και στην πραγματικότητα η ποσότητα ελεύθερων πολυαμινών είναι πολύ χαμηλότερη. Ενώ έχει αποδειχθεί ότι η αύξηση της ενδοκυτταρικής συγκέντρωσης της πολυαμίνης συσχετίζεται με αυξημένο πολλαπλασιασμό των κυττάρων και κατ' επέκταση της δημιουργίας όγκων. Τα δεδομένα που έχουμε από την ιατρική κοινότητα μας δείχνουν ότι η οδός της πολυαμίνης είναι στόχος πολλών γονιδίων που προκαλούν όγκους όπως οι Myc και Ras. Καθ' αυτόν τον τρόπο επιβεβαιώνεται ότι η πολυαμίνη συντελεί χημειοπροληπτικό και χημειοθεραπευτικό παράγοντα [7].

Την δεκαετία του 1960, οι Russel και Snyder κατέδειξαν τα υψηλά επίπεδα ενζυμικής δραστηριότητας ODC στον καρκίνο του ανθρώπου [8] και άνοιξαν τον δρόμο στην επιστημονική κοινότητα για περαιτέρω διερεύνηση της καρκινογένεσης. Πιο αναλυτικά, το ένζυμο ODC έχει αποδειχθεί ότι είναι σε αυξημένα επίπεδα στον ανθρώπινο καρκίνο του δέρματος χωρίς μελάνωμα [9]. Η αλληλοεξάρτηση των υψηλών επιπέδων ODC και πολυαμινών έχει συσχετιστεί στον καρκίνο του μαστού [10] και του προστάτη [11].

Με αυτόν τον τρόπο, βλέπουμε ότι οι ενδείξεις είναι θεωρητικά τεκμηριωμένες και λογικά εμπεριστατωμένες για να δοκιμάσουμε αρχικά σε πειραματικό στάδιο την χρήση πολυαμινών σε πειραματόζωα και κατ' επέκταση να μοντελοποιήσουμε την χημική πρόληψη στον άνθρωπο. Δυστυχώς όμως, το κυριότερο πρόβλημα στην συγκεκριμένη έρευνα είναι οι δοκιμές και ο έλεγχος της θεωρίας στην πράξη, δηλαδή *in vivo*. Σύμφωνα με την συμφωνία του Ελσίνκι το 1979 και εκείνης της Λισσαβόνας το 2005, έχει απαγορευτεί ρητά η χρήση μεγάλων αριθμών πειραματόζωων. Ο μέγιστος αριθμός για μη στρατιωτική έρευνα είναι 50 πειραματόζωα. Καταστείτε σαφές ότι ο περιορισμός αυτός αποτελεί τόσο ηθικά όσο και πειραματικά μεγάλο

<i>Substance Family</i>	<i>Substance Name</i>	<i>Carcinogenesis Obs.</i>	<i>Total Obs.</i>
Carcinogen	Benzopyrene	49	49
Polyamines	Ethylenediamine	2	19
	Cadaverine	12	30
	Putrescine	2	43
	Total	16	92

Πίνακας 3.2: Αριθμοί πειραματοζώων που εμφάνισαν καρκινική κυτταρική ανάπτυξη στην οικογένεια των Πολυαμινών Μονής Δόσης.

φραγμό για την εξέλιξη της επιστήμης [12].

Στο πίνακα 3.2 βλέπουμε το πλήθος των πειραματοζώων που τους δόθηκε μονή δόση πολυαμίνης και εμφάνισαν καρκινική δραστηριότητα. Σημειώνεται ότι η κατάληξη ήταν λόγω της ανάπτυξης καρκίνου και παρατηρείται ότι μόλις το ένα δέκατο του πληθυσμού αναπτύσσει καρκινικά κύτταρα. Επομένως αντιλαμβανόμαστε γοργά την ισχυρή δράση των πολυαμινών. Τα συγκεκριμένα πρότυπα (16 από τα 149) θα χρησιμοποιηθούν για την μαθηματική μοντελοποίηση της χημικής καρκινογένεσης στο κεφάλαιο 4: Μαθηματική Μοντελοποίηση. Ο πίνακας 3.3 μας υπογραμμίζει ότι αυξάνοντας την δόση της πολυαμίνης που χορηγήθηκε μαζί με τον καρκιγόνο παράγοντα τα αποτελέσματα μπορούν να γίνουν ακόμη πιο ελπιδοφόρα. Τονίζουμε ότι οι καταλήξεις από καρκίνο και ποσοστό ένα στα δέκα μειώθηκαν κατά το ήμισυ (τέσσερα στα εκατό).

<i>Substance Family</i>	<i>Substance Name</i>	<i>Carcinogenesis Obs.</i>	<i>Total Obs.</i>
Carcinogen	Benzopyrene	49	49
Polyamines	Ethylenediamine	1	50
	Cadaverine	5	50
	Putrescine	1	50
	Total	7	150

Πίνακας 3.3: Αριθμοί πειραματοζώων που εμφάνισαν καρκινική κυτταρική ανάπτυξη στην οικογένεια των Πολυαμινών Διπλής Δόσης.

### 3.2.2.4 Θειόλες

Οι θειόλες ανήκουν στον μικρόκοσμο και βρίσκονται σε αμινοξέα. Η πιο συχνή θειόλη είναι η κυστεΐνη και υπολείμματά της περιέχονται είτε σε πρωτεΐνες είτε σε μικρότερα μόρια. Ένα τέτοιο μόριο είναι η γλουταθειόνη, ένα τριεπιπτιδίο που περιέχει κυστεΐνη και υπάρχει άφθονο μέσα στο κύτταρο. Για να αντιληφθούμε την ουσιαστική διαφορά με τις πολυαμίνες που είχαν ως σκοπό την θωράκιση του οργανισμού μέσω προστασίας του DNA, φωσφολιπιδίων κ.ά., θα χρησιμοποιήσουμε ως παράδειγμα το αμινοξύ της L-Κυστεΐνης.

Το αμινοξύ της L-Κυστεΐνης μπορεί να δώσει αντιδράσεις προσθήκης στους διπλούς δεσμούς του βενζοπυρενίου. Η δυναμοποίηση της φουλφυδριλικής ομάδας της L-Κυστεΐνης μέσω εστεροποίησης δείχνει ότι αυτό το φαινόμενο συμβαίνει και επιτείνεται τα μέγιστα μέσω της δημιουργίας του εστερικού δεσμού στο καρβοξύλιο του αμινοξέος της κυστεΐνης, οπότε η προσθήκη του -SH στους διπλούς δεσμούς του βενζοπυρενίου λαμβάνει χώρα με πολύ ποσοτικότερο τρόπο χρησιμοποιώντας τον αιθυλικό ή μεθυλικό εστέρα της L-κυστεΐνης, παρά την L-κυστεΐνη από μόνης της ως αντικαρκινογόνο μόριο. Βεβαίως η εικόνα θα ξεκαθαρίσει απολύτως αναλύοντας με φασματογράφο μάζας τις ουσίες οι οποίες προκύπτουν από την αντίδραση του βενζοπυρενίου με τα μόρια που εν γένει χρησιμοποιήθηκαν στα εν λόγω πειράματα.

<i>Substance Family</i>	<i>Substance Name</i>	<i>Carcinogenesis Obs.</i>	<i>Total Obs.</i>
Carcinogen	Benzopyrene	49	49
Thiols	Bismuthiol	21	30
	L-Cysteine Ethyl Ester	7	28
	Mercaptosuccinic Acid	12	30
	Dithiothreitol	8	30
	L-Cysteine	11	30
	Total	59	148

Πίνακας 3.4: Αριθμοί πειραματοζώων που εμφάνισαν καρκινική κυτταρική ανάπτυξη στην οικογένεια των Θειολών.

Ένας σαφής διαχωρισμός ανάμεσα σε Πολυαμίνες και Θειόλες θα ήταν ότι οι πρώτες δρουν ως ασπίδα του οργανισμού μέσω της θωράκισης του DNA, του RNA, των πρωτεϊνών και των φωσφολιπιδίων ενώ οι δεύτερες δρουν ως σπαθί του οργα-

νισμού προσπαθώντας να εξουδετερώσουν τον καρκινογόνο παράγοντα.

Στο πίνακα 3.4 βλέπουμε το πλήθος των πειραματοζώων που τους δόθηκε μονή δόση θειόλης και εμφάνισαν καρκινική δραστηριότητα. Σημειώνεται ότι η κατάληξη ήταν λόγω της ανάπτυξης καρκίνου και παρατηρείται ότι τέσσερα δέκατα του πληθυσμού αναπτύσσει καρκινικά κύτταρα. Επομένως αντιλαμβανόμαστε περιορισμένη δράση των θειολών σε σχέση με τις πολυαμίνες. Ωστόσο δεν παύει να είναι σημαντικό το γεγονός ότι οι θειόλες αποδίδουν στην πρόληψη της χημικής καρκινογένεσης. Τα συγκεκριμένα πρότυπα (59 από τα 149) θα χρησιμοποιηθούν για την μαθηματική μοντελοποίηση της χημικής καρκινογένεσης στο κεφάλαιο 4: Μαθηματική Μοντελοποίηση.

Συγκεντρωτικά από τις δύο τελευταίες υποενότητες, δημιουργούνται τέσσερις κλάσεις όπως φαίνονται στον πίνακα 3.5. Αυτές οι κλάσεις θα χρησιμοποιηθούν στο κεφάλαιο 5: "Ταξινόμηση". Στο πέμπτο κεφάλαιο θα προσπαθήσουμε να δημιουργήσουμε τρεις αποδοτικούς ταξινομητές για τις ακόλουθες τρεις περιπτώσεις. Η πρώτη περίπτωση αφορά τις κατηγορίες Βενζοπυρένιο και Βενζοπυρένιο μαζί με Πολυαμίνες και Θειόλες, η δεύτερη περίπτωση αφορά Βενζοπυρένιο και Βενζοπυρένιο μαζί με Πολυαμίνες ενώ η τρίτη περίπτωση αφορά Βενζοπυρένιο και Βενζοπυρένιο μαζί με Θειόλες. Σημειώνουμε ακόμη ότι ως χαρακτηριστικά θα δοθούν οι παράμετροι των μοντέλων από την μαθηματική μοντελοποίηση. Σκοπός της συγκεκριμένης διαδικασίας είναι να απαντήσουμε στο ερώτημα αν τα μοντέλα είναι ικανά να ξεχωρίσουν την δράση του βενζοπυρενίου από την δράση του βενζοπυρενίου μαζί με κάποια δραστική ουσία. Στην ουσία, προσπαθούμε να απαντήσουμε εάν στο ερώτημα αν οι καμπύλες των προτύπων που έλαβαν βενζοπυρένιο διαφέρουν αισθητά από τα πρότυπα που έλαβαν βενζοπυρένιο μαζί με την δραστική ουσία. Η διαχωριστική ικανότητα που μας δίνουν τα μαθηματικά μοντέλα θα τεθεί υπό έλεγχο σε κάθε μία από τις προαναφερθείσες περιπτώσεις. Τέλος, για την στατιστική προσέγγιση (κεφάλαιο 6) και κατά επέκταση Στατιστική Συμπερασματολογία χρησιμοποιήθηκε όλος ο πληθυσμός των προτύπων με σκοπό να μελετηθεί το προσδόκιμο ζωής των πειραματοζώων.

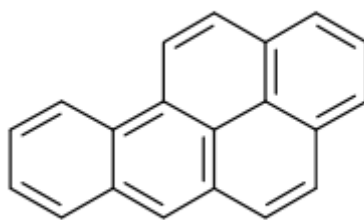
### 3.2.2.5 Βενζοπυρένιο

Το 3,4-βενζοπυρένιο (σχήμα 3.1) είναι μια ουσία περιέχουσα άνθρακα και υδρογόνο και που ανήκει στην χημική ομάδα των πολυκυκλικών αρωματικών υδρογονανθράκων (Π.Ο.Υ), η οποία περιλαμβάνει μια σειρά μορίων περιεχόντων διαφόρους

<i>Classes</i>	<i>Number of Observations</i>
Benzopyrene	49
Benzopyrene & Polyanimes	16
Benzopyrene & Thiols	59
Benzopyrene & Polyanimes OR Thiols	75

Πίνακας 3.5: Αριθμός Παρατηρήσεων ανά κλάση

αριθμούς βενζολικών δακτυλίων με εναλλάξ διπλούς δεσμούς μεταξύ των ατόμων άνθρακα που τους αποτελούν. Τα μόρια αυτά από την φύση τους και μέσω του φαινομένου της ταυτομέρειας μπορούν να υπάρχουν σε περισσότερες της μίας κατανομές των διπλών τους δεσμών, εκτός αν υπάρχει κάποιο δομικό κώλυμα στην στερεοδιάταξη των θεωρούμενων μορίων, που δεν επιτρέπει τον εύκολο απεντοπισμό των ηλεκτρονίων κάθε ουσίας σε πολλά άτομα άνθρακα. Εξαιτίας αυτών των καλουμένων «στερεοχημικών αποκλεισμών» σε μερικούς από τους Π.Ο.Υ παρουσιάζεται συγκέντρωση ηλεκτρονίων σε συγκεκριμένες περιοχές του μορίου, με αποτέλεσμα οι περιοχές αυτές να έχουν μια υψηλή δυνατότητα αντίδρασης με άλλα μόρια του χημικού περιβάλλοντος του θεωρούμενου πολυκυκλικού αρωματικού υδρογονάνθρακα.



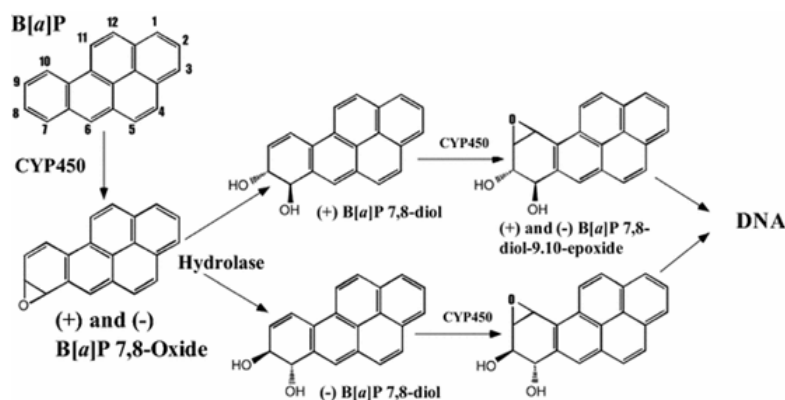
Σχήμα 3.1: Μοριακή δομή του 3,4-βενζοπυρενίου.

Εκείνα τα μόρια που εμφανίζουν δύο δομικά ιδιαίτερες ομάδες που ονομάζονται «περιοχή L» (bay region) και καρκινοφόρος περιοχή K (K region) έχουν την ιδιότητα να δημιουργούν καρκίνο στους οργανισμούς όταν χορηγούνται σε συγκεκριμένες ποσότητες υπερβαίνουσες μια ελάχιστη δόση. Η μετατροπή ενός κυττάρου σε κακόηθες νεοπλασματικό κύτταρο είναι γενετικό φαινόμενο, ονομαζόμενο κα-



κοήτης εξαλλαγή και προϋποθέτοντας ανεπανόρθωτη βλάβη του γενετικού υλικού, δηλαδή του DNA εκ μέρους του θεωρούμενου καρκινογόνου παράγοντα, ο οποίος στις περιπτώσεις των πειραμάτων μας είναι το 3,4-βενζοπυρένιο (B[a]P).

Στην περίπτωση του B[a]P το μόριο αυτό από μόνο του στην φυσική του μορφή δεν είναι σε θέση να δημιουργήσει κακοήθη εξαλλαγή σε οποιοδήποτε κύτταρο. Για να προκύψει αυτό πρέπει το B[a]P να υποστεί το λεγόμενο φαινόμενο της ενεργοποίησης, δηλαδή να μετατραπεί σε μια χημική ένωση η οποία να φέρει σε συγκεκριμένα άτομα άνθρακα ρίζες υδροξυλίου (OH•). Η ένωση αυτή ονομάζεται B[a]P-7,8-διόλη, είναι οπτικό ισομερές και συνεπώς έχει τις μορφές (+) B[a]P-7,8-διόλη και (-) B[a]P-7,8-διόλη και παράγεται από τον μεταβολισμό του B[a]P σε ένα σημαντικό ενζυμικό σύστημα των κυττάρων που ονομάζεται κυτόχρωμο P-450, αποτελούμενο από περισσότερα ένζυμα. Η (+) B[a]P-7,8-διόλη μεταβολίζεται κατόπιν σε ένα πιο ενεργό μόριο (υπό την μορφή ρακεμικού μείγματος, καθώς η προσθήκη του οξυγόνου στον δεσμό 9-10 είναι στερεοειδική, δηλαδή μπορεί να λάβει χώρα τόσο από την μία, όσο και από την άλλη πλευρά του επιπέδου του μορίου της (+)B[a]P-7,8-διόλης), σχηματίζοντας δύο εποξειδία που είναι μεταξύ τους οπτικά ισομερή. Τα εποξειδία αυτά της (+)B[a]P-7,8-διόλης είναι το (+) B[a]P-7,8-διόλη – 9,10 – εποξείδιο και το (-)B[a]P -διόλη-9,10-εποξείδιο, τα οποία δρουν κατ' ευθείαν πάνω στο DNA, δημιουργώντας σε αυτό μοριακές θραύσεις που μπορούν να μονιμοποιηθούν σε βλάβες αλληλουχίας και συνεπώς να δώσουν εξειδικευμένες μεταλλάξεις με κακοήθη χαρακτήρα στα κύτταρα στα οποία έχει εισέλθει και μεταβολιστεί το B[a]P. Η καταβολική ενεργοποίηση του B[a]P, μέσω του κυτοχρώμου P-450 φαίνεται στο σχήμα 3.2.



Σχήμα 3.2: Η καταβολική ενεργοποίηση του B[a]P μέσω του κυτοχρώμου P0450.

Το κυτόχρωμο P-450 περιέχει αιμικό σίδηρο, δηλαδή σίδηρο ενσωματωμένο σε

αιμικές ομάδες, σαν αυτές που υπάρχουν στην αιμοσφαιρίνη, η οποία αποτελεί και την πρωτεΐνη μεταφοράς οξυγόνου στους ιστούς των σωμάτων των εξελικτικά αναπτυγμένων ζώων, περιλαμβανομένων του ανθρώπου, όλων των πρωτευόντων και εν γένει όλων των σπονδυλωτών ζώων.

## ΚΕΦΑΛΑΙΟ 4

# ΜΑΘΗΜΑΤΙΚΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗ

---

- 4.1 Πληθυσμιακή και Εξατομικευμένη Προσέγγιση
  - 4.2 Μαθηματικά Μοντέλα Βιολογικής Ανάπτυξης Πληθυσμών
  - 4.3 Αλγόριθμος Προσαρμογής των Μαθηματικών Μοντέλων
  - 4.4 Αριθμητικές Μέθοδοι Υλοποίησης Ελαχίστων Τετραγώνων
  - 4.5 Αριθμητική Μέθοδος Marquardt-Levenberg
  - 4.6 Προσέγγιση αρχικών τιμών σε μη γραμμικά ελάχιστα τετράγωνα
  - 4.7 Μετρικές καλής προσαρμογής για μη γραμμικά μοντέλα
  - 4.8 Προβλεψιμότητα των μαθηματικών μοντέλων
- 

Στο κεφάλαιο αυτό τίθενται τα θεμέλια για την μοντελοποίηση της καρκινικής κυτταρικής ανάπτυξης. Στην αρχή του κεφαλαίου υλοποιείται με μία εκτενής βιβλιογραφική αναζήτηση με σκοπό να συγκριθεί η πληθυσμιακή και η εξατομικευμένη προσέγγιση. Στην συνέχεια αναφέρονται τα μαθηματικά μοντέλα που είναι πιο πιθανά να επιλύσουν το πρόβλημά μας μέσω της γραφικής τους μορφολογίας και υπογραμμίζεται η επιστημονική χρήση τους στον τομέα της πληθυσμιακής ανάπτυξης. Παρουσιάζεται πλήρως το μαθηματικό υπόβαθρο και οι αριθμητικές μέθοδοι για την προσαρμογή των μαθηματικών μοντέλων σε πειραματικά δεδομένα. Τέλος, επισημαίνονται μετρικές καλής προσαρμογής για την σύγκριση των μαθηματικών μοντέλων ενώ η προσοχή στρέφεται και στην προβλεψιμότητα που προσφέρει το εκάστοτε μοντέλο.

## 4.1 Πληθυσμιακή και Εξατομικευμένη Προσέγγιση

Στην αναπτυξιακή μοντελοποίηση, τα μαθηματικά παρέχουν μία ακριβή γλώσσα για να περιγραφεί η συσχέτιση μεταξύ μεταβλητών αλλά και οι αλλαγές στις βιολογικές καταστάσεις (λ.χ. έμβρυο, ανήλικος, ενήλικας). Στην ιατρική, μπορούμε να αναπαράστούμε μαθηματικά την κλινική πορεία της νόσου, την κατανομή της νόσου τόσο σε πληθυσμούς όσο και στην πάροδο του χρόνου αλλά και τους μηχανισμούς που δημιουργούν την ασθένεια [13]. Η ανάπτυξη ενός μοντέλου μίας λοίμωξης, μίας ασθένειας ή μίας πορείας κατάληξης απαιτεί από εμάς να ξεκινήσουμε με ακρίβεια τις παραδοχές μας σχετικά με τις παραμέτρους και τις διαδικασίες που επηρεάζουν την υγεία των προτύπων και μας επιτρέπει να υπολογίσουμε το αναμενόμενο αποτέλεσμα.

Τα μαθηματικά μοντέλα χωρίζονται σε μηχανιστικά και στατιστικά. Τα πρώτα χρησιμοποιούνται για να υλοποιηθούν προβλέψεις με βάση κάποια εκτίμηση παραμέτρων, ενώ τα δεύτερα χρησιμοποιούνται για άντληση των παραμέτρων από εμπειρικά δεδομένα. Η ανάπτυξη ενός μαθηματικού μοντέλου εμπεριέχει τον ορισμό κάποιων εξισώσεων ή κάποιων κανόνων που περιγράφουν την εξάπλωση μίας λοίμωξης ή την εξέλιξη μίας νόσου όπου παρευρίσκεται σε μία αναλογία μεταξύ της μαθηματικής αναπαράστασης του βιολογικού συστήματος του προτύπου. Οι προβλέψεις του μοντέλου αντλούνται είτε θεωρητικά μέσω μίας συσχέτισης των τιμών εισόδου και εξόδου στις τιμές των παραμέτρων, είτε αριθμητικά, με αποτελέσματα που δημιουργήθηκαν με λίγα βήματα από κάποιο υποσύνολο των παραμέτρων. Περισσότερες λεπτομέρειες θα δοθούν στην συνέχεια του κεφαλαίου όσο αφορά τον διαχωρισμό τους.

Σε αυτό το σημείο, θα μιλήσουμε για τις διαφορετικές προσεγγίσεις για την αναπαράσταση προτύπων. Οι κυριότερες προσεγγίσεις είναι η εξατομικευμένη προσέγγιση και η πληθυσμιακή προσέγγιση [14]. Στην εξατομικευμένη προσέγγιση κάθε άτομο στον πληθυσμό αντιπροσωπεύεται και παρακολουθείται ρητά μόνο από τον εαυτό του. Αντίθετα, στην πληθυσμιακή προσέγγιση, τα άτομα παρουσιάζονται μόνο ως μέλη μίας ομάδας και εκπροσωπούνται και παρακολουθούνται ρητά ως ομάδα. Σε μία εξατομικευμένη προσέγγιση μικρού πληθυσμού, τα χαρακτηριστικά του κάθε ατόμου μπορεί να είναι διαφορετικά και αυτή η προσέγγιση χρησιμοποιείται σε μοντελοποίηση νοικοκυριών ή εκδηλώσεις λοιμώξεων [15]. Συχνά, ωστόσο, σε πολλά άτομα αποδίδονται ίδια, αν όχι παρόμοια, χαρακτηριστικά που έχουν αντληθεί από

συγκεκριμένα άτομα. Αυτή η προσέγγιση επιτρέπει μεγάλους πληθυσμούς να μοντελοποιηθούν ως άτομα πιο αποδοτικά [16]. Η δυναμική αυτής της προσέγγισης έγκειται τόσο στο γεγονός ότι κάθε εμπειρία του κάθε διαφορετικού ατόμου καταγράφεται όσο και στο γεγονός ότι επιτρέπεται η είσοδος των πολύ μικρών λεπτομερειών στην ανάλυση. Ωστόσο, αποτελούν υπολογιστικά ακριβές προσεγγίσεις και είναι πολύ δύσκολες να αυτοματοποιηθούν και να ερμηνευτούν εξολοκλήρου. Η διερεύνηση της μετάδοσης και του ελέγχου σεξουαλικών μεταδιδόμενων νοσημάτων ακολουθεί ευρέως αυτήν την προσέγγιση.

Θα πρέπει να αναλογιστούμε έναν πολύ σημαντικό παράγοντα, την ετερογένεια. Σε μολυσματικές ασθένειες, η ετερογένεια στον κίνδυνο για απόκτηση ή μετάδοση μίας λοίμωξης μπορεί να επιτρέψει μία λοίμωξη να εισβάλει σε έναν πληθυσμό όταν ο ίδιος μέσος κίνδυνος σε έναν ομοιογενή πληθυσμό δεν θα το επέτρεπε αυτό. Ωστόσο, η ίδια ετερογένεια που καθιστά την εισβολή πιο πιθανή, θα περιορίσει την εξάπλωση της λοίμωξης. Ένα συχνό παράδειγμα που παρατίθεται είναι αυτό του HIV. Ένα μικρό σύνολο που πληθυσμού με πολλές επαφές όπως οι υπάλληλοι του έρωτα με τους πελάτες τους, μπορούν να μολυνθούν από την νόσο ενώ η μετάδοση στον υπόλοιπο πληθυσμό είναι περιορισμένη. Η ετερογένεια, λοιπόν, καταστείτε σαφές ότι αποτελεί μία κρυφή παράμετρο επικοινωνίας μεταξύ των ομάδων που μπορεί είτε να είναι τυχαία είτε όχι (άτομα κοινής κοινωνικής θέσης ή δημογραφικές ομάδες κ.ά.). Η ομοιογένεια υποθέτει ότι ο πληθυσμός έχει κοινό επίπεδο κινδύνου σε όλες τις ομάδες. Στην δική μας περίπτωση η ετερογένεια μπορεί να ληφθεί ως η περιοχή που αναπτύχθηκε ο όγκος, η προσωπική βιολογική αντίδραση του οργανισμού και κατ' επέκταση ίδια του η πολυπλοκότητα.

Στην πληθυσμιακή προσέγγιση, παρακολουθούνται αλλαγές που συμβαίνουν σε ομάδες εντός του πληθυσμού χωρίς να προσδιοριστεί ποιά άτομα εμπλέκονται ενώ η ετερογένεια μεταξύ των ατόμων παρουσιάζεται χωρίζοντας τον πληθυσμό σε υποομάδες. Η συγκεκριμένη προσέγγιση χρησιμοποιείται κυρίως για εξέταση πληθυσμιακών διαδικασιών και τάσεων και αντιπροσωπεύεται η εμπειρία των τυπικών ατόμων ενώ παραμένει υπολογιστικά αποδοτική. Το μεγαλύτερο της μειονέκτημα είναι η έλλειψη επιρροών από κάθε άτομο ξεχωριστά που μπορεί να είναι σημαντικό ορισμένες φορές.

Στην παρούσα εργασία θα ασχοληθούμε με την εξατομικευμένη προσέγγιση καθώς θέλουμε να αντλήσουμε περισσότερη πληροφορία από τις λεπτομέρειες ενώ παράλληλα αποδεχόμαστε την πολυπλοκότητα κάθε βιολογικού οργάνου και κατ'

επέκταση το οργανικό δίκτυο ως μοναδικό και ανεπανάληπτο. Τέλος, η εξατομικευμένη προσέγγιση μπορεί να ληφθεί ως υπερσύνολο των δύο προσεγγίσεων καθώς εύκολα μπορεί κάποιος να βγάλει γενικά συμπεράσματα έχοντας γνώση τι του συμβαίνει σε κάθε οργανισμό ξεχωριστά.

## 4.2 Μαθηματικά Μοντέλα Βιολογικής Ανάπτυξης Πληθυσμών

Η μέτρηση της βιολογικής ανάπτυξης είναι σημαντική σε πολλούς τομείς. Αρκετοί ερευνητές έχουν συμβάλει στην ανάπτυξη σχετικών μοντέλων: α)μοντέλο ανάπτυξης Gompertz [17], β)μοντέλο ανάπτυξης Logistic [18], γ)μοντέλο ανάπτυξης Bertalanffy [19], δ)μοντέλο ανάπτυξης Brody [20], ε)μοντέλο ανάπτυξης Weibull [21] και στ)μοντέλο ανάπτυξης Richards [22] είναι μόνο μερικά από τα πιο δημοφιλή. Η μαθηματική αναπαράσταση της σχετικής ανάπτυξης μπορεί να περιγραφεί ως μία πρώτης τάξης διαφορική εξίσωση όπως φαίνεται στην σχέση 4.1 :

$$\frac{df}{dt} = r_t f(t) \quad (4.1)$$

Στην σχέση 4.1 η  $f(t)$  αναπαριστά μία συνάρτηση ανάπτυξης και η  $r_t$  μία σχετική συνάρτηση ρυθμού με σεβασμό στο χρόνο. Αυτή η διαφορική εξίσωση έχει πολλές λύσεις και αρκετές από αυτές θα δούμε στην παρούσα μεταπτυχιακή εργασία. Τα μοντέλα ανάπτυξης έχουν χρησιμοποιηθεί ευρέως σε πολλά προβλήματα βιολογικής ανάπτυξης όπως σε επιστήμες ζώων [23] και φυτών [24]. Πιο συγκεκριμένα στην δασοκομία, οι περισσότερες ετήσιες γεωργικές καλλιέργειες είναι προκαθορισμένες και η ανάπτυξή τους σταματά όταν φτάσουν στην φυσιολογική ωριμότητα. Για κάποιες καλλιέργειες, η ανάπτυξη των μεμονωμένων οργάνων τους δεν είναι απεριόριστη, ακόμη και όταν οι περιβαλλοντικές συνθήκες παραμένουν ευνοϊκές. Η διάρκεια της περιόδου ανάπτυξης και το βάρος της τελικής ανάπτυξης, είτε για την καθορισμένη καλλιέργεια στο σύνολό της είτε για συγκεκριμένα όργανα, είναι δύο σημαντικά χαρακτηριστικά που εξαρτώνται από το περιβάλλον. Οι φυσιολόγοι και οι γενετιστές της καλλιέργειας επιθυμούν να ποσοτικοποιήσουν τα δύο χαρακτηριστικά για να χαρακτηρίσουν την γονοτυπική παραλλαγή σε απόκριση με τα περιβάλλοντα ανάπτυξης, βοηθώντας έτσι τους γεωργούς ή κτηνοτρόφους στον σχεδιασμό ποικιλιών καλλιέργειας για συγκεκριμένα περιβάλλοντα. Σε μοντέλα προσομοίωσης φυτών, οι “μοντελιστές” συχνά επιθυμούν να ποσοτικοποιήσουν την δυναμική της

ανάπτυξης, επιτρέποντας στους ημερήσιους ρυθμούς ανάπτυξης να είναι ίσοι με το αναμενόμενο τελικό βάρος στο τέλος του αναπτυσσόμενου κύκλου. Χρειάζεται, λοιπόν, μία εξίσωση για την μοντελοποίηση και τον χαρακτηρισμό της διάρκειας και του τελικού βάρους των καθορισμένων διαδικασιών ανάπτυξης [25].

Ωστόσο, μελέτες προσομοίωσης αποδεικνύουν ότι τα μοντέλα αυτά είναι τόσο ευέλικτα ώστε μπορούν να ταιριάζουν εσφαλμένα σε σύνολα δεδομένων και συνιστάται ιδιαίτερη προσοχή στην επιλογή αυτών [26]. Σε αυτό το σημείο αναφέρεται η ερώτηση: «Τί είναι όμως ένα μαθηματικό μοντέλο;».

Ένα μαθηματικό μοντέλο ανάπτυξης είναι μία μαθηματική έκφραση που περιγράφει ένα φαινόμενο στο πεδίο του χρόνου. Ένα ιδεατό μαθηματικό μοντέλο θα πρέπει να πληροί τα παρακάτω κριτήρια. Αρχικά θα πρέπει να έχει βάση του την φυσιολογία, να έχει όσον το δυνατόν λιγότερες αυθαίρετες σταθερές και οι μεταβλητές που χρησιμοποιούνται να είναι μετρήσιμες έτσι ώστε να δίνεται νόημα και έμφαση στο πείραμα. Επιπλέον, θα πρέπει να περιγράφει όσο το δυνατόν καλύτερα τα πειραματικά μας δεδομένα, να μπορεί να προβλέψει το φαινόμενο με εύλογη ακρίβεια, να βελτιώνει την γενική εικόνα που έχουμε για αυτό και να έχει πλάτος, υπό την έννοια ότι δύναται να εφαρμοστεί οπουδήποτε εμφανίζεται το φαινόμενο που μελετάτε εφόσον περιλαμβάνει κοινά χαρακτηριστικά [27].

Στην εργασία αυτή μελετάμε ως φαινόμενο την ανάπτυξη των καρκινικών κυττάρων σε πειραματόζωα και η μέτρηση αυτών γίνεται με βάση το βάρος τους σε γραμμάρια.

Ακόμη, πρέπει να γίνει αντιληπτό ότι η ανάπτυξη καρκινικών κυττάρων αποτελεί κομμάτι της μελέτης της ανάπτυξης κυττάρων ή αλλιώς της βιολογικής ανάπτυξης οργανισμών και για αυτόν τον λόγο τα βιολογικά αναπτυξιακά μοντέλα ταιριάζουν σε καρκινικές κυτταρικές αναπτύξεις. Ειδικότερα, μέσα στον κύκλο ζωής ενός οργανισμού, ενός φυτού ή μίας καλλιέργειας, η συνολική διάρκεια της ανάπτυξης μπορεί να χωριστεί σε τρεις επιμέρους φάσεις: μία πρώιμη φάση επιτάχυνσης, μία γραμμική φάση και μία φάση κορεσμού [28]. Επομένως, το μοτίβο ανάπτυξης όλων των οργανισμών ακολουθεί τυπικά μία σιγμοειδή καμπύλη και ο ρυθμός ανάπτυξης μία καμπύλη σε σχήμα καμπάνας. Ενώ το σιγμοειδές μοτίβο μπορεί να αναπαρασταθεί σε τμήματα χρησιμοποιώντας μία εκθετική, μία γραμμική και μία κυρτή εξίσωση διαδοχικά.

Μέσα στο κλίμα της μαθηματικής μοντελοποίησης γεννήθηκε και η ανάγκη για γενίκευση της λύσης της διαφορικής εξίσωσης που δόθηκε προηγουμένως. Η πιο

γνωστή γενίκευση αυτής δόθηκε από τους Koya και Goshu [29] και ικανοποιεί τα γνωστότερα μαθηματικά μοντέλα:

$$f(t) = A_L + (A - A_L)[1 - Be^{-k(\frac{t-\mu}{\delta})}]^m \quad (4.2)$$

Στην Σχέση 4.2 με  $A$  συμβολίζουμε την άνω ασύμπτωτη της  $f(t)$  δηλαδή ισχύει η εξής μαθηματική πρόταση:

$$\lim_{x \rightarrow +\infty} f(x) = A \quad (4.3)$$

Με  $\mu$  η σταθερά μετατόπισης του χρόνου και με  $\delta$  η σταθερά κλίμακας του χρόνου, και τέλος οι  $k$  και  $m$  μορφολογικοί παράμετροι του μαθηματικού μοντέλου ανάπτυξης. Για την μεταβλητή  $B$  έχουμε την Σχέση 4.3:

$$B = 1 - \left(\frac{A_\mu - A_L}{A - A_L}\right)^{\frac{1}{m}} \quad (4.4)$$

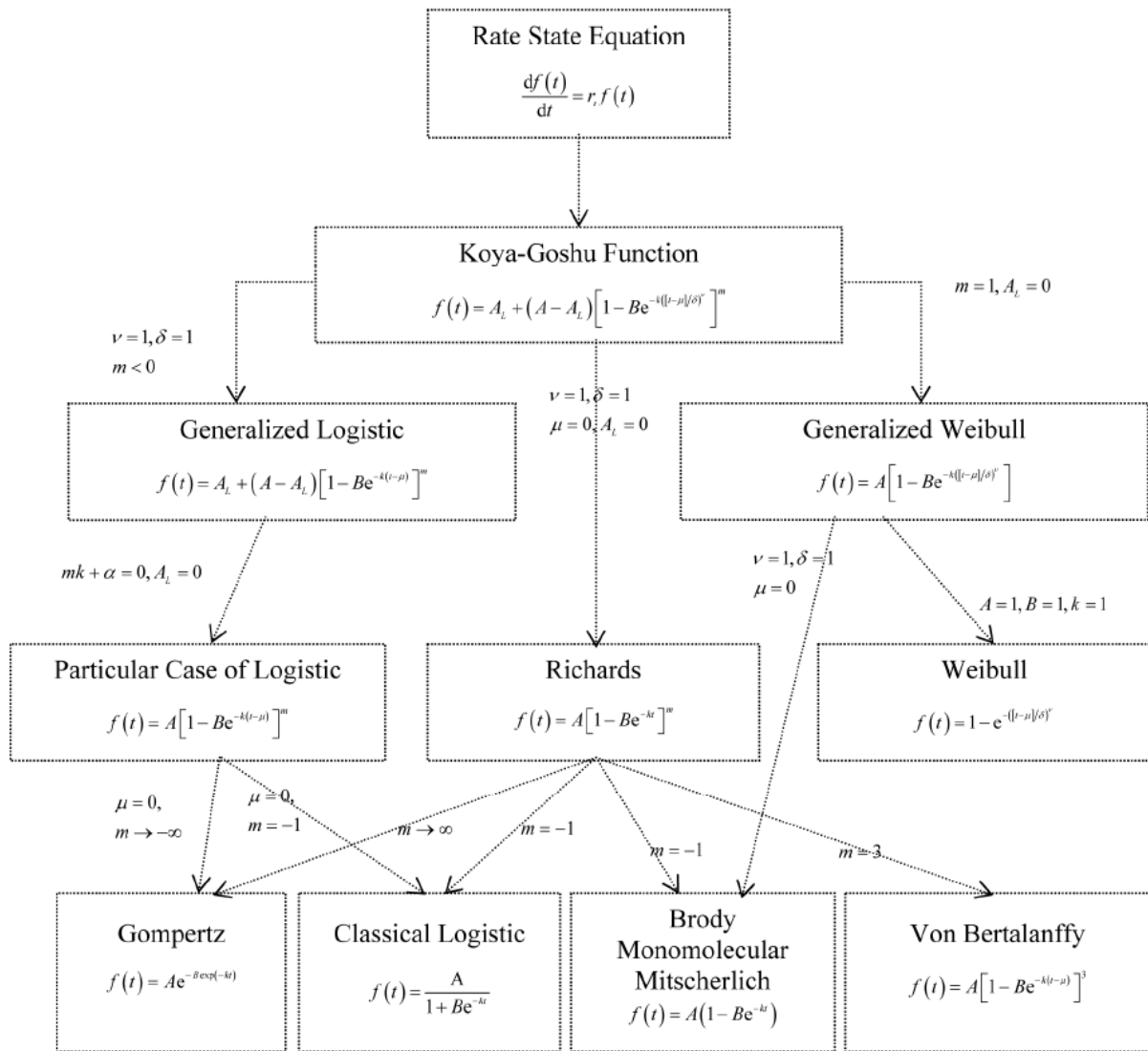
Όπου  $A_\mu$  η παράμετρος αναπτυξιακού ρυθμού. Ακόμη, διαλέγοντας συγκεκριμένες τιμές για τις παραμέτρους μπορούμε να αντλήσουμε όλα τα γνωστά μαθηματικά μοντέλα ανάπτυξης όπως βλέπουμε στο Σχήμα 4.1, στο οποίο υπάρχει το διάγραμμα ροής.

Τέλος, είναι σημαντικό να διακρίνουμε τα μοντέλα σε δύο βασικές κατηγορίες: τα μηχανιστικά και τα στατιστικά μοντέλα. Τα μηχανιστικά είναι εκείνα που προσπαθούν να περιγράψουν την εμφάνιση εμποδίων ή μη αναμενόμενων φαινομένων στις αναπτύξεις των οργανισμών. Ενώ τα στατιστικά προσπαθούν να συνοψίσουν την σχέση μεταξύ μεταβλητών μεταφέροντάς την σε μαθηματικές συναρτησιακές σχέσεις.

Τα μηχανιστικά μοντέλα προβλέπουν την επίπτωση και τον επιπολασμό μίας νόσου ή ενός μη αναμενόμενου φαινομένου (ως προς τη φυσιολογική κατάσταση του οργάνου ή του πληθυσμού) ενώ παράλληλα δύναται να κατανοήσει τις επιπτώσεις από αιτιακά μονοπάτια. Κάνει σαφείς υποθέσεις, επιτρέπει την πρόβλεψη πέρα από το εύρος των δεδομένων και παράγει δοκιμαστικές προβλέψεις. Παρόλα αυτά βασίζεται σε μία καλή κατανόηση των παραδοχών και μπορεί να μην εφαρμόζεται στα δεδομένα μας.

Τα στατιστικά μοντέλα χρησιμοποιούνται όλο και περισσότερο τα τελευταία χρόνια. Κύριος λόγος χρήσης τους είναι ο ποσοτικός προσδιορισμός της σχέσης μεταξύ μεταβλητών χωρίς καμία υπόθεση. Το τελευταίο προσφέρει τρομερή ευελιξία στο βωμό της κατανόησης των αιτιών καθώς τα μοντέλα αυτής της κατηγορίας είναι γρήγορα ως προς την υλοποίηση, χρησιμοποιούνται τακτικά και μπορούν να





Σχήμα 4.1: Διάγραμμα ροής των Μαθηματικών Μοντέλων Ανάπτυξης ανάλογα με τις τιμές των παραμέτρων τους

αιχμαλωτίσουν διαφορετικές αλληλεπιδράσεις σε δεδομένα. Οι ιδιότητές τους είναι γνωστές και κατανοητές ενώ η μεγάλη δύναμή τους έγκειται στην περιγραφή των δεδομένων και δεν βασίζεται σε διερεύνηση του οργανισμού με αιτιακά μονοπάτια.

#### 4.2.1 Μαθηματικό Μοντέλο Ανάπτυξης Exponential

Το εκθετικό μοντέλο υποθέτει την σταθερή αναλογία μεταξύ του ρυθμού αύξησης των καρκινικών κυττάρων και των συνολικών καρκινικών κυττάρων την δεδομένη χρονική στιγμή. Η σταθερή αναλογία συμβολίζεται με  $\lambda$  και πρόκειται στην ουσία

για την αυθαίρετη σταθερά. Έτσι έχουμε την Σχέση 4.5:

$$\frac{dV}{dt} = \lambda V \quad (4.5)$$

ο οποίος αποτελεί μία διαφορική εξίσωση με γενική λύση:

$$V = V_0 e^{\lambda t} \quad (4.6)$$

Εύκολα μπορεί να παρατηρήσει κάποιος ότι ο χρόνος διπλασιασμού της ποσότητας  $V$  είναι σταθερός καθώς το  $V_0$  αντιπροσωπεύει τον αρχικό αριθμό των καρκινικών κυττάρων και είναι κάτι σταθερό, όμοια για τα  $e$  και  $\lambda$ . Παραθέτουμε για διαφάνεια μία ενδεικτική απόδειξη.

$$\begin{aligned} \frac{V'}{V} = 2 &\Rightarrow \frac{V_0 e^{\lambda t'}}{V_0 e^{\lambda t}} = 2 \Rightarrow \frac{e^{\lambda t'}}{e^{\lambda t}} = 2 \Rightarrow \\ e^{\lambda t' - \lambda t} = 2 &\Rightarrow \lambda t' - \lambda t = \ln 2 \Rightarrow \\ \lambda(t' - t) = \ln 2 &\Rightarrow DT = \frac{\ln 2}{\lambda} \blacksquare \end{aligned}$$

Επομένως, το εκθετικό μοντέλο υποδηλώνει σταθερό ρυθμό διπλασιασμού των συνολικών κυττάρων. Το μοντέλο αυτό έδωσε έναυσμα στην επιστημονική κοινότητα να ασχοληθεί περαιτέρω με την έννοια του Χρόνου Διπλασιασμού ( $DT$ ) και ενώ δεν έχει ως βάση την φυσιολογία, κάνει μία πολύ λογική υπόθεση. Παρόλα αυτά τα μειονεκτήματά του θα φανούν στην πορεία της έρευνας. Γεωμετρικά, το εκθετικό μοντέλο προσεγγίζει την τυπική εκθετική συνάρτηση. Αναμένουμε ότι θα έχει καλή απόδοση σε δεδομένα με λίγες καταγραφές καθώς αιχμαλωτίζει την πρώτη φάση της ομοιόστασης. Η ραγδαία εξέλιξη στην αρχή και ο απότομος και συνεχώς αυξανόμενος ρυθμός αύξησης των τιμών είναι άλλωστε τα κυριότερα χαρακτηριστικά της εκθετικής συνάρτησης. Ωστόσο σε δεδομένα με αρκετές καταγραφές περιμένουμε περιορισμένη απόδοση καθώς λόγω της ομοιόστασης θεωρούμε ότι οι σιγμοειδείς καμπύλες τείνουν να προσαρμόζονται καλύτερα στο φαινόμενο αυτό.

## 4.2.2 Μαθηματικό Μοντέλο Ανάπτυξης Gompertz

Το Μοντέλο Gompertz αναπτύχθηκε το 1825 για ανθρώπινες καταλήξεις και το 1934 ο Albert Casey το χρησιμοποίησε για ταίριασμα δεδομένων με αναπτύξεις καρκινικών κυττάρων σε καμπύλη. Η διαφορική μορφή της καμπύλης διαφέρεται στην Σχέση 4.7:

$$\frac{dV}{dt} = V \left( \frac{A}{B} \right) (1 - (1 - B)e^{-bt}) \quad (4.7)$$

ενώ η γενική της λύση δίνεται από την Σχέση 4.8:

$$V = V_0 e^{\frac{A(1-e^{-bt})}{B}} \quad (4.8)$$

Όμοια με πριν  $V$  είναι ο όγκος των καρκινικών κυττάρων την δεδομένη στιγμή  $t$ ,  $V_0$  ο αρχικός όγκος των καρκινικών κυττάρων ενώ τα  $A$  και  $B$  αποτελούν τις αυθαίρετες σταθερές. Σχηματικά αυτό το μοντέλο αρχικά πλησιάζει το εκθετικό ωστόσο προς το τέλος της καμπύλης δεν υπάρχουν σημαντικές αλλαγές στο πεδίο του χρόνου.

Μορφολογικά θυμίζει μία σιγμοειδή καμπύλη και χρησιμοποιείται κυρίως για όγκους σε προχωρημένα στάδια καθώς είναι σαφώς ανεπαρκές για μικρότερους όγκους. Σε αρκετά μικρές ποσότητες όγκων, όταν ο ρυθμός πολλαπλασιασμού του όγκου πλησιάζει το άπειρο, το μοντέλο δεν δίνει ρεαλιστική εικόνα. Πιο συγκεκριμένα, οποιοσδήποτε νόμος ανάπτυξης ή μοντέλο που ακολουθεί έναν σχετικό ρυθμό ανάπτυξης της μορφής  $\frac{P}{Q}$  και έχοντας την ποσότητα  $P$  να τείνει στο άπειρο όσο η ποσότητα  $Q$  τείνει στο 0, δεν μπορεί να ερμηνεύσει ή να αντιπροσωπεύσει τις πολύπλοκες βιολογικές διαδικασίες στον κυτταρικό κύκλο και την απόπτωση. Το γεγονός αυτό συμβαίνει διότι οι τελευταίες διαδικασίες βασίζονται σε Χρόνους Διπλασιασμού που δεν μπορούν να λάβουν αυθαίρετα μικρές τιμές [30].

Αρκετές δουλειές άντλησαν αρκετά αποτελέσματα με την χρήση του συγκεκριμένου μοντέλου. Το 1964 και 1965 ο Laird ανέλυσε ιατρικά δεδομένα που έγιναν σε ποντικούς και κατέληξε ότι το μοντέλο είναι επαρκές και περιγράφει αρκετά καλά τα δεδομένα. Ο McCredie το 1965 συμφώνησε στην χρήση αυτού του μοντέλου έναντι του εκθετικού με βάση τα δικά του δεδομένα. Το 1970 οι Simpson και Herren συμφώνησαν και αυτοί με τα δικά τους δεδομένα σε αναπτύξεις καρκινικών κυττάρων σε διαφορετικά σημεία ανάπτυξης. Να σημειώσουμε σε αυτό το σημείο ότι καμία έρευνα δεν είχε πάνω από 20 πειραματόζωα, ενώ η συντριπτική πλειονότητα αυτών είχαν κάτω από 10. Τέλος, ενώ το μοντέλο Gompertz φαίνεται να έχει καλό ταίριασμα με τα δεδομένα, δεν απαλλάσσεται των μειονεκτημάτων. Το μεγαλύτερό του μειονέκτημα είναι ότι δεν αντλείται από καμία βάση Φυσιολογίας. Γεωμετρικά πρόκειται για μία σιγμοειδή συνάρτηση η οποία δεν είναι συμμετρική. Η ιδιαίτερή της ικανότητα είναι να αλλάζει την καμπυλότητά της πριν το σημείο συμμετρίας. Καθ' αυτόν τον τρόπο μπορεί να αιχμαλωτίζει αυξομειώσεις στις αρχικές τιμές με μεγαλύτερη ευκολία από τα άλλα μοντέλα.

### 4.2.3 Μαθηματικό Μοντέλο Ανάπτυξης Bertalanffy

Το μοντέλο Bertalanffy αρχικά χρησιμοποιήθηκε για την πληθυσμιακή ανάπτυξη των ψαριών και των ζώων γενικότερα ενώ έχει πολύ καλά αποτελέσματα και ως ένα μοντέλο περιγραφής της ανάπτυξης καρκινικών κυττάρων. Η ιδιαιτερότητα αυτού του μοντέλου είναι στο γεγονός ότι στηρίζεται στην θεωρία μεταβολισμού και μπορεί να επιβεβαιωθεί σε πειραματικό στάδιο. Οι υποθέσεις που στηρίχθηκαν ήταν οι ακόλουθες. Ο ρυθμός ανάπτυξης είναι ίσος της διαφοράς μεταξύ αναβολικού και καταβολικού ρυθμού και η αναλογία μεταξύ ρυθμού απώλειας κυττάρων και καρκινικών κυττάρων. Επιπλέον, η μορφολογία της νόσου δεν αλλάζει ενώ ο αναβολικός ρυθμός είναι ανάλογος των καρκινικών κυττάρων υψωμένος στο  $\frac{2}{3}$  και ο όγκος είναι ανάλογος της επιφάνειας. Με βάση τα παραπάνω καταλήγουν στην Σχέση 4.9:

$$\frac{dV}{dt} = nV^{\frac{2}{3}} - mV \quad (4.9)$$

ενώ η γενική της λύση δίνεται από την Σχέση 4.10:

$$V = \left[ \frac{n}{m} - \left( \frac{n}{m} - c \right) e^{-\frac{mt}{3}} \right]^3 \quad (4.10)$$

ή ισοδύναμα από την Σχέση 4.11:

$$V = b_1(1 - b_2 e^{-b_3 t})^3 \quad (4.11)$$

ενώ οι  $n$ ,  $m$  αποτελούν συντελεστές αναλογίας.

Γεωμετρικά το συγκεκριμένο μοντέλο πλησιάζει την αντίστροφη της εκθετικής συνάρτησης, την λογαριθμική. Αποτελεί μία κοίλη καμπύλη. Η κυρτότητα αυτή μας δίνει την προαίσθηση ότι το συγκεκριμένο μοντέλο θα προσαρμοστεί καλά στην δεύτερη και τρίτη φάση της κυτταρικής καρκινικής ανάπτυξης καθώς εμπεριέχει τόσο το γραμμικό κομμάτι στην αρχή όσο και την εξομάλυνση των τιμών στο τέλος.

### 4.2.4 Μαθηματικό Μοντέλο Ανάπτυξης Logistic

Το μοντέλο των Verhulst-Pearl ή γνωστότερο ως λογιστική εξίσωση είναι το μοντέλο που θα σχολιάσουμε σε αυτήν την παράγραφο. Το 1845 ο Verhulst ασχολήθηκε με το συγκεκριμένο μοντέλο ενώ το 1924 επανακαλύφθηκε από τον Pearl. Η διαφορική του εξίσωση δίνεται από την Σχέση 4.12:

$$\frac{dV}{dt} = \alpha V - \beta V^2 \quad (4.12)$$

ενώ η γενική της λύση δίνεται από την Σχέση 4.13:

$$V = \frac{\alpha}{e^{-at}(\frac{\alpha}{V_0} - b) + b} \quad (4.13)$$

όπου  $a$  και  $b$  αυθαίρετες σταθερές. Αξίζει να σημειωθεί ότι το συγκεκριμένο μοντέλο θυμίζει αρκετά αυτό του Bertalanffy, όμως είναι πιο ευέλικτο. Εάν οι παρατηρήσεις μας στο πρώιμο στάδιο εμφανίζουν μία παρεμφερή δραστηριότητα με την εκθετική αύξηση τότε διαλέγουμε μικρή τιμή στο  $b$ . Το λογιστικό μοντέλο, σε αυτήν την περίπτωση, έχει κοινή συμπεριφορά με το εκθετικό μοντέλο σε όλο το πεδίο του χρόνου. Σε άλλη περίπτωση, εάν οι παρατηρήσεις μας δεν έχουν εκθετική συμπεριφορά, διαλέγουμε μεγάλο  $b$  για να έχουμε την ζητούμενη καθυστέρηση στην ανάπτυξη των καρκινικών κυττάρων. Αναμένουμε καλή απόδοση από το συγκεκριμένο μοντέλο.

Η πιο τυπική χρήση της λογιστικής εξίσωσης είναι η μοντελοποίηση πληθυσμιακών αναπτύξεων. Η βάση του θεμελιωτή της ήταν ότι το ποσοστό αναπαραγωγής είναι ανάλογο τόσο του υπάρχοντος πληθυσμού όσο και του ποσού των διαθέσιμων πόρων ενώ η αρχική του χρήση ήταν να περιγράψει την αυτοπεριοριζόμενη ανάπτυξη ενός βιολογικού πληθυσμού. Γεωμετρικά, εφαρμόζοντας τις παρακάτω πράξεις μπορούμε να αντλήσουμε περισσότερη πληροφορία για την μορφολογία της σιγμοειδούς αυτής συνάρτησης [31].

$$\begin{aligned} \frac{dV}{dt} = \alpha V - \beta V^2 &\Rightarrow \frac{dV}{dt} = \alpha V \left(1 - \frac{\beta}{\alpha} V\right) \\ &\Rightarrow \frac{dV}{dt} = \alpha V \left(1 - \frac{V}{\frac{\alpha}{\beta}}\right) \blacksquare \end{aligned}$$

Η παράμετρος  $a$  ορίζει τον ρυθμό ανάπτυξης ενώ ο όρος  $\frac{a}{b}$  μας δίνει την μέγιστη πληθυσμιακή τιμή ενός είδους ή στην δική μας περίπτωση το μέγιστο βάρος καρκινικών κυττάρων. Καθ' αυτόν τον τρόπο η ευθεία  $y = \frac{a}{b}$  αποτελεί οριζόντια ασύμπτωτη της λογιστικής συνάρτησης. Σε αυτό το σημείο είναι σημαντικό να αναφερθεί ότι στην βιολογία, τα είδη προσδιορίζονται ως  $a$  - στρατηγικά ή  $\frac{a}{b}$  - στρατηγικά ανάλογα με την φυσική τους επιλογή στην ιστορική τους εξέλιξη. Το τελευταίο επηρεάζει την εξέλιξη και την κληρονομικότητα του είδους.

#### 4.2.5 Μαθηματικό Μοντέλο Ανάπτυξης Gauss

Στην επιστήμη των μαθηματικών, η γκαουσιανή συνάρτηση έχει την ακόλουθη μορφή που δίνεται στην Σχέση 4.14:

$$f(x) = \alpha e^{-\frac{(x-b)^2}{2c^2}} \quad (4.14)$$

Αντλήσε το όνομά της από τον μαθηματικό Carl Friedrich Gauss. Το γράφημα της συνάρτησης χαρακτηρίζεται από μία συμμετρική καμπύλη που έχει μορφή καμπάνας ενώ αποτελεί ένα στατιστικό μοντέλο. Η παράμετρος  $a$  δηλώνει το ύψος της καμπάνας. Η παράμετρος  $b$  την θέση του κέντρου της καμπάνας και η παράμετρος  $c$  αντιπροσωπεύει το πλάτος της. Η ομάδα αυτών των συναρτήσεων χρησιμοποιείται ευρέως στην Στατιστική, την επεξεργασία σήματος, την επεξεργασία εικόνας, την Φυσική και την Μηχανική. Στην Στατιστική και πιο συγκεκριμένα στην Θεωρία Πιθανοτήτων, χρησιμοποιείται για να περιγράψει κανονικές κατανομές. Στην επεξεργασία σήματος ορίζει Γκαουσιανά φίλτρα ενώ στην επεξεργασία εικόνας διδιάστατοι γκαουσιανοί χρησιμοποιούνται για την διαδικασία του θολώματος. Τέλος στην Φυσική και την Μηχανική, οι εξισώσεις θερμότητας αποτελούν μερικές διαφορικές εξισώσεις όπου περιγράφουν την κατανομή μίας ποσότητας (όπως της θερμότητας) εξελίσσεται στην πάροδο του χρόνου σε ένα στερεό μέσο, ενώ ρέει αυθόρμητα από μέρη που βρίσκονται ψηλότερα προς μέρη όπου είναι χαμηλότερα.

Το συγκεκριμένο μοντέλο έχει χρησιμοποιηθεί για πληθυσμιακές αναπτύξεις και ειδικότερα σε μικροβιακές αναπτύξεις [32]. Η λόγος που αποδίδει μία καμπανοειδής μορφολογία σε σιγμοειδή φαινόμενα είναι το γεγονός ότι κάποιος μπορεί να απομονώσει κομμάτια αυτής. Πιο αναλυτικά, διατηρώντας την αρχή της καμπύλης μέχρι το σημείο του μέγιστου ύψους η καμπύλη αποτελεί μία κλασική σιγμοειδής καμπύλη. Καθ' αυτόν τον τρόπο μπορούμε να αντιληφθούμε ότι τα στάδια της εκθετικότητας, γραμμικότητας και κυρτότητας πληρούνται. Ωστόσο το τρίτο στάδιο, αυτό της κυρτότητας, λόγω της συμμετρικότητας της καμπύλης είναι χρονικά περιορισμένο διότι ακολουθεί η απότομη πτώση της καμπύλης και για αυτόν τον λόγο φαίνεται συρρικνωμένο. Τέλος, το συγκεκριμένο μοντέλο δεν έχει χρησιμοποιηθεί για καρκινικές κυτταρικές αναπτύξεις ενώ χρησιμοποιείται ενδελεχώς για περιγραφή αναπτύξεων φυτών καθώς η γήρανση συντελεί μία πιο σταδιακή και συμμετρική διαδικασία.

#### 4.2.6 Μαθηματικό Μοντέλο Ανάπτυξης Polynomial

Το πολυωνυμικό μοντέλο χρησιμοποιείται ενδελεχώς σε γενικότερες αναπτύξεις και αποτελεί ένα εμπειρικό μοντέλο χωρίς κάποια αξιολογή βάση Φυσιολογίας. Με την πάροδο του χρόνου, η επιστημονική κοινότητα έδινε ολοένα και περισσότερη έμφαση στην ιατρική έρευνα πάνω σε αναπτυξιακή μοντελοποίηση σε όλη τη διάρκεια ζωής

και τον εντοπισμό παραγόντων που σχετίζονται με την ανάπτυξη ενός είδους λ.χ. του ανθρώπινου είδους. Τέτοιες έρευνες συνήθως χρησιμοποιούν πολυωνυμικά μοντέλα, μοντέλα που έρχονται από διαιρέσεις αυτών (ρητά μοντέλα) ή άλλες πράξεις μεταξύ τους, όπως πρόσθεση δύο πολυωνύμων με διαφορεικές παραμέτρους (πολυωνυμικό μοντέλο δύο όρων) [33]. Τα κυριότερα θετικά χαρακτηριστικά αυτών των μοντέλων αναφέρονται στο γεγονός της απλοϊκότητας, της ερμηνευσιμότητας, της ευελιξίας και της εύκολης υλοποίησης. Παρόλα αυτά υπάρχουν και αρκετά μειονεκτήματα για αυτήν την επιλογή μοντέλου. Τα πολυωνυμικά μοντέλα μεγάλου βαθμού δημιουργούν αναταράξεις και ταλαντώσεις μεταξύ τιμών που έχουν διέπονται από μεγάλη ακρίβεια. Τα πολυώνυμα είναι ικανά να προσαρμοστούν σε ένα εύρος των δεδομένων μας αλλά συχνά διαφεύγουν από το πλαίσιο των ορίων λόγω της προηγούμενης ταλαντούμενης συμπεριφοράς. Επιπλέον, εκ φύσεως, τα πολυώνυμα έχουν περιορισμένη αντίληψη της ασυμπτωτικής συμπεριφοράς και για αυτόν τον λόγο μπορεί να μην μοντελοποιούν τα ασυμπτωτικά φαινόμενα σε επαρκή βαθμό. Η περιορισμένη αυτή αντίληψη μπορεί να βασιστεί στο γεγονός ότι ένα πολυώνυμο δεν μπορεί να έχει συγκεκριμένη συμπεριφορά γύρω από συγκεκριμένες τιμές του άξονα  $x$ , αλλά μπορεί να έχει ασυμπτωτική συμπεριφορά (συγκεκριμένη συμπεριφορά στο όριο του απείρου). Τέλος, ενώ καμία διαδικασία δεν είναι απρόσβλητη ως προς το ισοζύγιο της μεταβλητότητας και του bias, εφόσον τα πολυώνυμα παρουσιάζουν μία ιδιαίτερα κακή ανταλλαγή μεταξύ μορφολογίας και βαθμού. Συνηθέστερος βαθμός των πολυωνύμων που χρησιμοποιείται είναι ο τρίτος και ο τέταρτος. Μεγαλύτερος βαθμός προσβάλλει ακόμη περισσότερο το *bias* και καταλήγουμε σε προϊόν υπερβολικής προσαρμογής δεδομένων.

Όταν η μοντελοποίηση μέσω πολυωνυμικών διαδικασιών κρίνεται ανεπαρκής λόγω οποιουδήποτε από τους παραπάνω περιορισμούς, η χρήση ρητών διαδικασιών μπορεί να δώσει καλύτερη επιλογή. Το ρητό μοντέλο πολυωνύμων εξομαλύνει κάποια προβλήματα των πολυωνυμικών όπως την ταλάντωση μεταξύ τιμών, την περιορισμένη μορφολογία της καμπύλης, την ασυμπτωτική συμπεριφορά τόσο στο άπειρο όσο σε συγκεκριμένες τιμές. Ωστόσο, η βιβλιογραφία για το ρητά πολυώνυμα είναι φτωχή καθώς η οικογένεια αυτού του μοντέλου είναι δύσκολα ερμηνεύσιμη και δεν ακολουθεί κάποιο μοτίβο. Η αύξηση των παραμέτρων δεν είναι ιδανική για προβλήματα πολλών παραγόντων. Ενώ σε περίπτωση που δεν τοποθετηθούν όρια στην μεταβλητή του χρόνου, μπορεί να εμφανιστούν κάθετες ασύμπτωτες λόγω ύπαρξης ριζών του πολυωνύμου του παρονομαστή.

### 4.2.7 Μαθηματικό Μοντέλο Ανάπτυξης Brody

Το μοντέλο ανάπτυξης Brody δημιουργήθηκε το 1945 για βιολογικές αναπτύξεις και ακολουθεί τον ακόλουθο τύπο της Σχέσης 4.15:

$$f(x) = A(1 - Be^{-kt}) \quad (4.15)$$

Το  $A$  αντιπροσωπεύει την άνω ασύμπτωτη της καμπύλης και δρα ως άνω φράγμα της  $f(t)$  ενώ το  $B = 1 - (\frac{f(0)}{A})$  [34]. Το μοντέλο του Brody χρησιμοποιείται ενδελεχώς σε βιολογικές πληθυσμιακές αναπτύξεις, με επίκεντρο τα ψάρια και άλλα ζώα [35]. Ενώ η περιγραφική του ικανότητα έχει χρησιμοποιηθεί για καρκινικές κυτταρικές αναπτύξεις με περιοχή ενδιαφέροντος τον μαστό [36]. Ωστόσο τα δεδομένα που χρησιμοποιήθηκαν ήταν συνθετικά και οι δοκιμές ήταν σε κατάσταση προσομοίωσης και όχι πειραματικού επιπέδου. Το συγκεκριμένο μοντέλο δεν έχει εφαρμοστεί σε προηγούμενη εργασία πάνω σε καρκινικές αναπτύξεις με πραγματικά δεδομένα οπότε δεν μπορούμε να είμαστε βέβαιοι για τα αποτελέσματα που θα αντληθούν εκ των προτέρων. Παρόλα αυτά, η καμπύλη του συγκεκριμένου μοντέλου είναι αρχικά λογαριθμική και με την πάροδο του χρόνου εξομαλύνει. Θεωρητικά, θα πρέπει να αποτυγχάνει να αιχμαλωτίσει αρχικές καταγραφές των πειραμάτων μας αλλά θα προσεγγίζει καλύτερα τις μεγαλύτερες σε χρόνο καταγραφές. Είναι σημαντικό σε αυτό το σημείο να αναφέρουμε, ότι έχει προταθεί ένας νέος σχεδιασμός για την καρκινική κυτταρική ανάπτυξη. Αυτό το νέο σχέδιο υποδηλώνει ότι ο καρκίνος αναπτύσσεται ως αποτέλεσμα μεταλλάξεων στα φυσιολογικά κύτταρα. Όλο το δίκτυο αυτών των μεταλλάξεων μπορεί να περιγράψει την συμπεριφορά των καρκινικών βλαστοκυττάρων καθώς όταν τα βλαστοκύτταρα διαιρούνται και παράγουν καρκινικά κύτταρα με βάση τις οδηγίες που λαμβάνουν από το δίκτυο. Γενικότερα, προτείνεται να θεωρήσουμε γραμμικά δίκτυα καρκίνου στα οποία τα καρκινικά βλαστοκύτταρα χωρίζονται σε κύτταρα ίδιου τύπου ή παράγουν ογκοκύτταρα. Όταν ένα τέτοιο δίκτυο εδραιωθεί μπορεί να επεκταθεί και να συμπεριλάβει δράσεις ουσιών όπως ορμονικές θεραπείες. Το μοντέλο Brody φαίνεται να έχει ιδιαίτερη χρησιμότητα σε ένα τέτοιο δίκτυο [37].

### 4.2.8 Μαθηματικό Μοντέλο Ανάπτυξης Weibull

Στην Θεωρία πιθανοτήτων και στατιστικής, η κατανομή Weibull είναι μία συνεχής κατανομή πιθανότητας. Πήρε το όνομά της από τον Δανό μαθηματικό Waloddi



Weibull, ο οποίος την περιέγραψε με πλήρη λεπτομέρεια το 1951. Ωστόσο ανακαλύφθηκε από τον Frechet το 1927 και χρησιμοποιήθηκε το 1933 από τους Rosin και Rammler για την κατανομή μεγέθους των μορίων. Σήμερα η συγκεκριμένη κατανομή και κατ' επέκταση το μαθηματικό της μοντέλο έχει εφαρμογές σε μία πληθώρα από επιστήμες. Η ανάλυση επιβίωσης, η κατανομή της υπέρτασης που συμβαίνει σε ένα ηλεκτρικό σύστημα, κατανομές χρόνων κατασκευής και παράδοσης σε εργοστάσια, κατανομές ταχύτητας σε ανεμογεννήτριες, κατανομές σε χρόνους παραμονής σε ιστοσελίδες και η κατανομή μεγέθους των μορίων είναι μόνο μερικές από τις εφαρμογές της.

Ως μαθηματικό μοντέλο και κατ' επέκταση αναπτυξιακό μοντέλο είναι αρκετά ευέλικτο καθώς μπορεί να αλλάξει σε σημαντικό βαθμό την μορφολογία του. Δύναται να ακολουθεί συμμετρική σιγμοειδή μορφή αλλά και λογαριθμική. Στην Δασολογία έχει χρησιμοποιηθεί ως αναπτυξιακό μοντέλο για να διερευνήσει το ποσοστό των παρασίτων σε διαφορετικά είδη ψαριών [38]. Τέλος, στην ίδια έρευνα συγκρίναν το μοντέλο Weibull δύο, τριών και τεσσάρων παραμέτρων καταλήγοντας ότι αυτό με τις τρεις παραμέτρους δίνει το καλύτερο αντίβαρο μεταξύ υπολογιστικού κόστους υλοποίησης, υπολογιστικού κόστους έρευνας αρχικών παραμέτρων του μοντέλου και τελικού αποτελέσματος μέσω μετρικών όπως  $RMSE$  και  $X^2$ . Το μοντέλο Weibull τριών παραμέτρων έχει τον ακόλουθο μαθηματικό τύπο που δίνεται από την Σχέση 4.16 και αποτελεί στατιστικό μοντέλο.

$$f(t) = A(1 - e^{-kt^m}) \quad (4.16)$$

Όμοια με προηγούμενα μοντέλα, Το  $A$  αντιπροσωπεύει την άνω ασύμπτωτη της καμπύλης και δρα ως άνω φράγμα της  $f(t)$  ενώ το  $k$  αποτελεί μορφολογικό παράγοντα όπως και το  $m$ .

#### 4.2.9 Μαθηματικό Μοντέλο Ανάπτυξης PowerLaw

Το δυναμονομικό μοντέλο υποθέτει την σχέση δύο ποσοτήτων όπου η σχετική μεταβολή της μίας ποσότητας έχει ως αποτέλεσμα ανάλογη μεταβολή στην άλλη ποσότητα, ανεξάρτητα από το αρχικό μέγεθος αυτών. Πιο συγκεκριμένα, η μία ποσότητα διαφοροποιείται ως δύναμη της άλλης. Ο μαθηματικός του τύπος δίνεται από τον τύπο:  $f(x) = ax^{-b}$ . Το δυναμονομικό μοντέλο θα μπορούσε να θεωρηθεί ως μία συνάρτηση άμεσης αναλογίας καθώς η  $f(cx) = f(x)c^{-b}$  έχει άμεση αναλογία με την

$f(x)$ . Η χρήση του μοντέλου αφορά κυρίως την επιστήμη της Αστρονομίας, της Φυσικής, της Βιολογίας και των Μαθηματικών. Η συχνότερη εμφάνισή της είναι στην συσχέτιση μεταβολισμού και μεγέθους ζώων, την αναλογία του μεγέθους των κρατήρων στην επιφάνεια της Σελήνης [39] αλλά και στην συχνότητα εμφάνισης λέξεων σε ένα κείμενο.

Θεωρούμε ότι το συγκεκριμένο μοντέλο θα έχει απόδοση σε δεδομένα με λίγες καταγραφές που θα αιχμαλωτίζουν κυρίως της φάση της εκθετικότητας ενώ δεν θα έχει εξίσου καλή απόδοση σε δεδομένα με πολλές καταγραφές. Θα θέλαμε να εστιάσουμε ωστόσο στην φυσιολογική του βάση, που δεν είναι άλλη από την ύπαρξη αναλογίας μεταξύ του χρόνου και καρκινικού βάρους. Πιθανώς να ήταν πιο ορθό να χρησιμοποιούσαμε τρεις διαφορετικές δυναμονομικές συναρτήσεις καθώς σε κάθε φάση της ομοιόστασης η αναλογία των ποσοτήτων διαφέρει. Όμως στην πραγματικότητα δεν γνωρίζουμε με ακρίβεια τις χρονικές στιγμές (ούτε κατά προσέγγιση λόγω μοναδικότητας των οργανισμών και των αντιδράσεών τους) όπου τέτοιες αλλαγές παίρνουν θέση.

### 4.3 Αλγόριθμος Προσαρμογής των Μαθηματικών Μοντέλων

Η μαθηματική μέθοδος που θα χρησιμοποιήσουμε για την προσαρμογή των μαθηματικών μοντέλων στα δεδομένα μας ήταν αυτή των ελαχίστων τετραγώνων. Η μέθοδος των ελαχίστων τετραγώνων χρησιμοποιείται ενδελεχώς στις περιπτώσεις που προσπαθούμε να περιγράψουμε κάποιο φαινόμενο στην πάροδο του χρόνου. Στόχος μας είναι η κατασκευή μίας γραφικής παράστασης που προσπαθεί να περιγράψει ένα φαινόμενο, όταν μοναδική μας αφετηρία είναι μία σειρά από πειραματικές τιμές των μεγεθών που περιγράφουν το φαινόμενο αυτό, ενώ οι τιμές αυτές έχουν ληφθεί σε διακριτές χρονικές στιγμές. Ακόμη, οι πειραματικές μας τιμές δεν είναι ακριβείς, δηλαδή δεν προκύπτουν από κάποιο μαθηματικό τύπο. Στην πραγματικότητα, προσπαθούμε να προσδιορίσουμε την άγνωστη μαθηματική καμπύλη που διαπερνά όλα μας τα σημεία, ενώ όταν δεν συμβαίνει αυτό, να διατρέχει σε κοντινή απόσταση από αυτά. Με άλλον τρόπο, θα μπορούσε κάποιος να πει ότι προσδιορίζουμε την μορφολογία μίας άγνωστης μαθηματικής σχέσης στην οποία ταιριάζουν όσο το δυνατόν καλύτερα τα πειραματικά μας δεδομένα.

Τα ελάχιστα τετράγωνα αποτελούν μία τυπική διαδικασία στην ανάλυση πα-

λινδρόμησης για να προσεγγίσουμε μία λύση από ένα σύστημα εξισώσεων όπου υπάρχουν περισσότερες εξισώσεις από ότι το πλήθος των αγνώστων. Σκοπός της μεθόδου είναι η ελαχιστοποίηση του αθροίσματος των τετραγωνικών διαφορών ανάμεσα στις παρατηρήσεις (πειραματικές τιμές) και στις εκτιμώμενες τιμές που προκύπτουν βάση της άγνωστης καμπύλης που προσπαθούμε να προσδιορίσουμε.

Η μέθοδος των ελαχίστων τετραγώνων είναι πολύ σημαντική για προσαρμογή δεδομένων σε καμπύλες ανάπτυξης. Οι κύριες κατηγορίες διαχωρισμού είναι τα γραμμικά ελάχιστα τετράγωνα και τα μη γραμμικά ελάχιστα τετράγωνα με βάση εάν οι διαφορές είναι γραμμικές προς όλους τους αγνώστους. Η γραμμική μέθοδος εμφανίζεται σε στατιστικές αναλύσεις παλινδρόμησης και έχει κλειστό τύπο ως προς την λύση της, αντίθετα η μη γραμμική εκτελείται με επαναλήψεις όπου σε κάθε επανάληψη το σύστημα προσεγγίζεται από ένα γραμμικό. Καταλαβαίνουμε ότι ο πυρήνας της μεθόδου είναι κοινός.

Η μέθοδος αποδίδεται στον Carl Friedrich Gauss το 1795 και δημοσιεύτηκε το 1805 από τον Adrien-Marie Legendre. Επιπλέον, η συγκεκριμένη μέθοδος αποδέχεται τις ακόλουθες υποθέσεις. Αρχικά, όλα τα πειραματικά σφάλματα των δεδομένων μπορούν να αποδοθούν στην τεταγμένη του καρτεσιανού, δηλαδή στον άξονα των  $y$ . Δεύτερον, όλα τα πειραματικά σφάλματα των δεδομένων μπορούν να περιγραφούν από μία Γκαουσιανή κατανομή ενώ τέλος, δεν ενέχεται συστηματικό σφάλμα στα δεδομένα. Και οι τρεις αυτές υποθέσεις πρέπει να ικανοποιούνται για την έγκυρη διαδικασία της μεθόδου των ελαχίστων τετραγώνων. Ακόμη, κάθε μία από αυτές τις παραδοχές έχει σημαντική επίπτωση στην ορθότητα της μεθόδου για την πειραματική ανάλυση των δεδομένων [40].

Όσο αφορά την πρώτη υπόθεση, η μέθοδος επιλέγει τιμές στις διάφορες παραμέτρους της συνάρτησης (σε εμάς τα μαθηματικά μοντέλα) με σκοπό να ελαχιστοποιήσει το άθροισμα των τετραγώνων των αποστάσεων από τις πειραματικές καταγραφές. Σε αυτό το σημείο, θα πρέπει να τονίσουμε ότι μόνο οι κάθετες αποστάσεις (οι αποστάσεις δηλαδή στον άξονα των  $y$ ) ελαχιστοποιούνται και όχι οι οριζόντιες. Για αυτόν τον λόγο και με σκοπό να χρησιμοποιηθεί σωστά η μέθοδος, το πρόβλημα θα πρέπει να διατυπωθεί με τέτοιο τρόπο ώστε η ακρίβεια του προσδιορισμού των τιμών κατά μήκος της τετμημένης (άξονας των  $x$ ) πρέπει να είναι σημαντικά μεγαλύτερες από την ακρίβεια της τεταγμένης (άξονας των  $y$ ). Επιπλέον, τα πειράματα πρέπει να σχεδιάζονται έτσι ώστε ο τυχαίος πειραματικός θόρυβος των εξαρτώμενων μεταβλητών να μην συσχετίζεται άμεσα με τον θόρυβο των ανε-

ξάρτητων μεταβλητών. Γενικότερα, δεν υπάρχει μέθοδος για την παράκαμψη αυτής της απαίτησης.

Η δεύτερη υπόθεση σημαίνει ότι εάν υπήρχε μία καταγραφή πολλές φορές στα πειραματικά δεδομένα τότε τα σφάλματα (δηλαδή οι αποστάσεις) θα μπορούσαν να περιγραφούν από μία Γκαουσιανή κατανομή. Το τελευταίο αποτελεί μία λογική υπόθεση για πολλές, αλλά όχι για όλες τις πειραματικές διαδικασίες. Η υπόθεση αυτή συνεπάγεται επίσης ότι ο αριθμός των δεδομένων μας είναι επαρκής για να διασφαλιστεί μία καλή στατιστική δειγματοληψία του τυχαίου λάθους. Ο ακριβής αριθμός των καταγραφών από τα πειράματα που απαιτείται για την εκπλήρωση αυτής της υπόθεσης είναι δύσκολο να προσδιοριστεί. Ωστόσο, με βεβαιότητα μπορεί να ειπωθεί ότι δεν υπάρχει πρακτική διαφορά προσδιορισμού έξι καταγραφών κατανεμημένων σε ένα εύρος τριπλάσιο ή έχοντας δεκαοχτώ καταγραφές σε όλο το εύρος. Καταλαβαίνουμε ότι μία ορθή επιστημονική δουλειά μπορεί να αντληθεί και με λίγες καταγραφές με ένα συγκεκριμένο εύρος.

Η τρίτη και τελευταία υπόθεση δηλώνει σαφώς ότι δεν μπορούν να υπάρχουν συστηματικά σφάλματα στα δεδομένα. Ο μόνος τρόπος για να μαρακάμψει κάποιος αυτήν την απαίτηση είναι να συμπεριλάβει μία σειρά όρων στην συνάρτηση για να περιγράψει στα συστηματικά σφάλματα. Για παράδειγμα, σε πειράματα ορμονικών δεσμών, μη συγκεκριμένοι δεσμοί είναι πηγές συστηματικών σφαλμάτων και οι αντίστοιχες επιστημονικές έρευνες έχουν δείξει ότι μπορούν να συμπεριληφθούν με τέτοιο τρόπο ώστε να μην αποτελούν συστηματικό σφάλμα στα δεδομένα.

#### 4.4 Αριθμητικές Μέθοδοι Υλοποίησης Ελαχίστων Τετραγώνων

Η μέθοδος των ελαχίστων τετραγώνων όπως είδαμε στην προηγούμενη παράγραφο αποτελεί μία διαδικασία ελαχιστοποίησης. Οι πιο συχνές αριθμητικές μέθοδοι που χρησιμοποιούνται για την υλοποίηση των ελαχίστων τετραγώνων είναι οι Nelder-Mead, Newton-Raphson, Gauss-Newton, Steepest Descent και Maequardt-Levenberg. Για την πλειονότητα των προβλημάτων, όλες οι μέθοδοι έχουν ισοδύναμα αποτελέσματα. Σε αυτήν την ενότητα θα δούμε τα δυνατά και τα τρωτά σημεία των αριθμητικών μεθόδων που αναφέραμε. Οι αριθμητικοί αλγόριθμοι θα παρουσιαστούν σε πρόβλημα δύο διαστάσεων  $Y$  εναντίον  $X$ . Πρέπει να σημειωθεί ότι οι ίδιοι αλγόριθμοι δουλεύουν και για εξαρτημένες και μη ματαβλητές. Για να επεκταθούν

οι μέθοδοι σε πολλές διαστάσεις, κάθε ζεύγος καταγραφής  $(X_i, Y_i)$  μπορεί να αντικατασταθεί με δύο διανύσματα: ένα για τις εξαρτημένες μεταβλητές και ένα για τις ανεξάρτητες ενώ και τα αθροίσματα με την σειρά τους επεκτείνονται για να περιλαμβάνουν επιπλέον όρους.

Ο αλγόριθμος Nelder-Mead είναι μία πολυδιάστατη διαδικασία εύρεσης ελαχίστου μίας νόρμας, συνήθως της  $X^2$  και υλοποιείται με προσεκτικά επιλεγμένες αναζητήσεις σε μία διάσταση κάθε φορά. Με σκοπό να προσδιορίσουμε δύο παραμέτρους χρειαζόμαστε τρία σημεία. Γενικότερα για  $\nu$  παραμέτρους απαιτούνται  $\nu + 1$  σημεία στον  $\nu$ -διάστατο χώρο των παραμέτρων. Το επόμενο βήμα είναι η αξιολόγηση. Σε κάθε ένα σημείο αξιολογούμε με βάση μία στατιστική νόρμα ή ένα μαθηματικό κανόνα, τα δεδομένα που θα πρέπει να ελαχιστοποιηθούν. Το σημείο που προσφέρει την καλύτερη απόδοση επιλέγεται και το κέντρο του υπόλοιπου συνόλου αξιολογεί εκ νέου τα εναπομείναντα σημεία. Στην ουσία, το κέντρο αυτό είναι ο μέσος όρος όλων των υπόλοιπων σημείων εκτός από το σημείο που επιλέχθηκε από τον κανόνα. Το επόμενο βήμα μας είναι να πραγματοποιήσουμε μία αναζήτηση κατά το μήκος της ευθείας που σχηματίζει το κέντρο με το σημείο που επιλέχθηκε. Η αναζήτηση αυτή τερματίζει όταν βρούμε ένα σημείο που ο κανόνας να δίνει μικρότερη τιμή από τον μέσο όρο. Σε αυτήν την περίπτωση η διαδικασία αλλάζει το σημείο την καλύτερης απόδοσης με αυτό που μόλις βρήκαμε. Η διαδικασία συνεχίζει κάνοντας νέες εκτιμήσεις σημείων ως προς την απόδοση του κανόνα, εκτιμώντας το νέο κέντρο από τα υπόλοιπα σημεία και ψάχνοντας στο μήκος της ευθείας αυτών μέχρι ένα νέο σημείο να δίνει καλύτερη απόδοση. Η συνεχόμενη αυτή διαδικασία μπορεί να συνεχιστεί μέχρι να όλα τα  $\nu + 1$  σημεία να συμπίπτουν. Μία οπτικοποίηση της μεθόδου θα μπορούσε να μας δείξει ότι πρόκειται στην ουσία για ένα τρίγωνο που αλλάζει το ένα ή τα δύο από τα τρία άκρα του προσπαθώντας να κλείσει μέσα του την λύση. Το μόνο μειονέκτημα της μεθόδου είναι ότι μερικές φορές η σύγκλιση έρχεται αρκετά αργά ενώ στα θετικά του συγκαταλέγεται η σύγκλιση του, η μη χρήση παραγώγων με σεβασμό στις παραμέτρους (καθώς αποτελεί τεράστιο υπολογιστικό κόστος) και η χρήση οποιουδήποτε κανόνα επιθυμούμε.

Ο αλγόριθμος Newton-Raphson αποτελεί μία μέθοδο επίλυσης συστημάτων εξισώσεων της μορφής  $\frac{dx^2}{da_i} = 0$  όπου έχουμε διαφορετική εξίσωση για κάθε εκτιμώμενη παράμετρο. Η βάση της μεθόδου είναι ότι τα ακρότατα της συνάρτησης εμφανίζονται σε σημεία που οι πρώτες παραγώγοι είναι μηδενικές. Τα θετικά στοιχεία αυτής της μεθόδου είναι ότι ελαχιστοποιεί οποιαδήποτε νόρμα θέσουμε, πέρα

από την X2 ενώ είναι πολύ γρήγορη μέθοδος. Ωστόσο απαιτείται εκτίμηση των πρώτων και δευτέρων παραγώγων της συνάρτησης. Παράλληλα η απόδοσή του είναι εξαρτημένη από τις αρχικές τιμές των παραμέτρων.

Ο αλγόριθμος Gauss-Newton είναι στην ουσία ένας αλγόριθμος που βασίζεται στην τυπική μέθοδο Gauss σε συνδυασμό με γραμμική αναζήτηση ενώ χρησιμοποιούμε μία συχνά αποδοτική προσέγγιση του Εσσιανού πίνακα [41] [42]. Έχοντας ως αντικειμενική συνάρτηση το άθροισμα των υπολοίπων, η κλασική μέθοδος του Newton επιλύει το σύστημα εξισώσεων, όπως παρουσιάζεται στην Σχέση 4.17.

$$\nabla^2 f(x_k)p = -\nabla f(x_k) \quad (4.17)$$

Ο αλγόριθμος των Gauss και Newton προσπαθεί να επιλύσει το σύστημα της Σχέσης 4.18, για να αντλήσει την κατεύθυνση  $p_k^{GN}$ .

$$J_k^T J_k p_k^{GN} = -J_k^T r_k \quad (4.18)$$

Αυτή η παραλλαγή στον αλγόριθμο δίνει μία πληθώρα από πλεονεκτήματα έναντι της αρχικής έκδοσης. Η αρχική απλοποίηση είναι στον Εσσιανό και συγκεκριμένα στην Σχέση 4.19, αποφεύγει το πρόβλημα του υπολογισμού των Εσσιανών πινάκων για κάθε ξεχωριστό υπόλοιπο.

$$\nabla^2 f_k \approx J_k^T J_k \quad (4.19)$$

## 4.5 Αριθμητική Μέθοδος Marquardt-Levenberg

Ο αλγόριθμος Levenberg-Marquardt χρησιμοποιεί την ίδια αποδοτική προσέγγιση όπως αντικαθιστά την γραμμική αναζήτηση με διαστήματα εμπιστοσύνης. Αυτή η προσέγγιση λύνει προβλήματα συμπεριφοράς που σχετίζονται με τον βαθμό του Ιακωβιανού πίνακα. Στην πράξη εάν υπολογίζαμε τον  $J_k$  για την εκτίμηση της παραγώγου μέσα από την σχέση 4.20.

$$\nabla^2 f_k = J_k^T J_k \quad (4.20)$$

η προσέγγισή μας δεν θα απαιτούσε επιπλέον εκτιμήσεις παραγώγων και η εξοικονόμηση του υπολογιστικού χρόνου είναι σημαντική για μία μεγάλη μερίδα εφαρμογών. Δεύτερον, η σύγκλιση της μεθόδου είναι παραπλήσια με αυτή του κλασικού

Newton, καθώς για τις τιμές κοντά στην λύση οι όροι  $J_k^T J_k$  και  $\nabla^2 f_k$  ομοιάζουν. Τρίτο πλεονέκτημα εμφανίζεται όταν ο  $J_k$  πίνακας έχει τον μέγιστο βαθμό και η  $\nabla f_k$  δεν είναι μηδενική, η κατεύθυνση της  $p_k^{GN}$  αποτελεί καθοδική κατεύθυνση για την  $f$  και επομένως αποτελεί ιδανική κατεύθυνση για γραμμική αναζήτηση. Τέλος, ένα ακόμη σημαντικό πλεονέκτημα είναι η ομοιότητα μεταξύ των σχέσεων 4.18 και 4.21.

$$J^T J_x^* = J^T y \quad (4.21)$$

Η Σχέση 4.21 αποτελεί το σύστημα εξισώσεων για το γραμμικό πρόβλημα ελαχίστων τετραγώνων. Η ομοιότητα αυτή μας υποδηλώνει στην ουσία ότι η κατεύθυνση  $p_k^{GN}$  είναι λύση του γραμμικού προβλήματος των ελαχίστων τετραγώνων. Απόρροια όλων αυτών είναι ότι μπορούμε να εντοπίσουμε την κατεύθυνση  $p_k^{GN}$  εφαρμόζοντας γραμμικούς αλγορίθμους ελαχίστων τετραγώνων στο υποπρόβλημα που αναφέρεται.

Σε αλγορίθμους που βασίζονται σε QR ή SVD παραγοντοποιήσεις, δεν χρειάζεται ο υπολογισμός του εσσιανού μέσω της προσέγγισης που δίνει η σχέση 1. Εάν ο αριθμός των υπολοίπων είναι πολύ μεγάλος και ο αριθμός των παραμέτρων πολύ μικρός, ίσως η αποθήκευση του πίνακα  $J$  δεν είναι ορθή. Αντ' αυτού θα ήταν πιο ορθό να υπολογιστούν ο πίνακας  $J_k^T J_k$  και το παράγωγο διάνυσμα  $J^T r$  εκτιμώντας τις ποσότητες  $r_j$  και  $\nabla r_j$  διαδοχικά και εν τέλει αθροίζοντάς τες όπως φαίνεται στις σχέσεις 4.22 και 4.23.

$$J^T J = \sum_{i=1}^m m(\nabla r_j)(\nabla r_j)^T \quad (4.22)$$

$$J^T r = \sum_{i=1}^m r_j(\nabla r_j) \quad (4.23)$$

Σε αυτήν την περίπτωση μπορούμε να λύσουμε πλέον το σύστημα από την σχέση 4.23.

Οι υλοποιήσεις της συγκεκριμένης μεθόδου συνήθως υλοποιούν μία γραμμική αναζήτηση στην κατεύθυνση της  $p_k^{GN}$  απαιτώντας ένα βήμα  $\alpha$  να ικανοποιήσει τις συνθήκες Armijo-Wolfe [43] [44].

## 4.6 Προσέγγιση αρχικών τιμών σε μη γραμμικά ελάχιστα τετράγωνα

Υπάρχει αρκετή δραστηριότητα στον τομέα της αναζήτησης μεθόδων για εύρεση αρχικών προσεγγίσεων σε μη γραμμικά προβλήματα ελαχίστων τετραγώνων. Ωστόσο οι βασικότερες κατευθύνσεις των μεθόδων αποτελούνται είτε από μεθόδους προσέγγισης είτε από μεθόδους ευθείας αναζήτησης σε συγκεκριμένες περιοχές. Ανεξάρτητα της μεθόδου, η αναζήτηση πρέπει να γίνει πολύ προσεκτικά λόγω της ύπαρξης τοπικών ελαχίστων και η αναζήτηση μπορεί να εγκλωβιστεί σε τέτοια σημεία δίνοντάς μας λάθος συμπεράσματα. Για τον λόγο αυτό, αρκετές επανηλήψεις της μεθόδου από διαφορετικά σημεία έναρξης είναι αναγκαίες. Το πρόβλημα έρευνας αρχικών προσεγγίσεων παραμένει ανοικτό στο σύνολό του.

Στην παρούσα μεταπτυχιακή εργασία χρησιμοποιήθηκε η αναζήτηση σε συγκεκριμένες περιοχές. Αρχικά, αντλήσαμε τις προτεινόμενες βέλτιστες τιμές παραμέτρων από προηγούμενες εργασίες και τις χρησιμοποιήσαμε ως αρχική προσέγγιση σε όλα τα πρότυπα. Μερικά από αυτά μοντελοποιήθηκαν επιτυχώς ενώ άλλα δεν κατόρθωσαν σύγκλιση. Σε αυτά που μοντελοποιήθηκαν κρατήσαμε τις τελικές τιμές των παραμέτρων και δημιουργήσαμε ένα εύρος τιμών για κάθε παράμετρο. Η τελική μοντελοποίηση έγινε σε αυτό το εύρος με μία διαφορά δέκα τις εκατό εκατέρωθεν των ορίων. Σημειώνουμε ότι δοκιμάστηκαν 250 τυχαία σημεία σε ένα πλεύγμα τιμών τριπλάσιο προς όλες τις κατευθύνσεις των παραμέτρων και δεν βρέθηκε αποτέλεσμα αναζήτησης με μικρότερη τιμή στην αντικειμενική συνάρτηση. Δεν μπορούμε να είμαστε βέβαιοι ότι οι τιμές που βρέθηκαν συντελούν το ολικό ελάχιστο, καθώς δεν μπορεί να ελεγχθεί όλο το πεδίο ορισμού σε λογικό υπολογιστικό χρόνο. Όμως είμαστε βέβαιοι ότι στην συγκεκριμένη περιοχή, δίνεται η βέλτιστη τιμή.

## 4.7 Μετρικές καλής προσαρμογής για μη γραμμικά μοντέλα

Καθώς πολλές μορφολογίες μπορούν να αιχμαλωτίσουν τα δεδομένα που έχουμε στην διάθεσή μας, αναφύεται η επιτακτική ανάγκη να βρεθεί η όσο το δυνατό καλύτερη. Έτσι βλέπουμε την ανάγκη ύπαρξης μίας μετρικής και επιπλέον μίας αντικειμενικής συνάρτησης. Η αντικειμενική συνάρτηση είναι στην ουσία ο κανόνας που



προσπαθούμε να ελαχιστοποιήσουμε κατά την μέθοδο των ελάχιστων τετραγώνων. Οι μετρικές είναι οι επιπλέον κανόνες που μπορούμε να ελέγξουμε στο τέλος της διαδικασίας και μας δίνουν επιπλέον πληροφορίες για την προσαρμογή που δημιουργήσαμε.

Η μετρική SSE ή Άθροισμα των τετραγωνικών εκτιμήσεων των σφαλμάτων είναι στην ουσία το άθροισμα των αποκλίσεων των πειραματικών τιμών από τις τιμές που δίνει ένα μαθηματικό μοντέλο. Αποτελεί ένα μέτρο της ασυμφωνίας μεταξύ των δεδομένων και ενός μοντέλου εκτίμησης. Μικρές τιμές στην μετρική SSE δηλώνουν στενή προσαρμογή του μοντέλου πάνω στα δεδομένα. Χρησιμοποιείται κυρίως ως κριτήριο βελτιστοποίησης για την επιλογή των παραμέτρων και του ίδιου του μαθηματικού μοντέλου κατ' επέκταση. Πρόκειται με άλλα λόγια για την πιο συνηθισμένη αντικειμενική συνάρτηση. Ενώ ο τύπος της δίνεται από την Σχέση 4.24.

$$SSE = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (4.24)$$

Όπου το  $x_i$  αντανακλά την  $i$ -οστή τετμημένη των πειραματικών δεδομένων ενώ  $y_i$  και  $f(x_i)$  είναι η  $i$ -οστή τεταγμένη των πειραματικών δεδομένων και η εκτιμώμενη τιμή μέσω του μοντέλου αντίστοιχα.

Η μετρική  $R^2$  (ή coefficient of determination), είναι το ποσοστό της μεταβλητότητας στην εξαρτημένη μεταβλητή που είναι προβλέψιμη από τις ανεξάρτητες μεταβλητές. Χρησιμοποιείται κυρίως για πρόβλεψη μελλοντικών αποτελεσμάτων ή για έλεγχο στατιστικών υποθέσεων. Η βασική πληροφορία που μας δίνει είναι η απάντηση στο πόσο καλά οι εκτιμώμενες τιμές αναπαράγονται από το μοντέλο, με βάση το ποσοστό της συνολικής μεταβλητότητας που εξηγείται από το μοντέλο [45]. Υπάρχουν αρκετοί ορισμοί για την μετρική αυτή και μόλις μερικές φορές είναι ισοδύναμοι. Μία κλάση περιλαμβάνει τις περιπτώσεις της γραμμικής παλινδρόμησης όπου χρησιμοποιούν την  $r^2$  αντί της  $R^2$ . Ενώ όταν σταθεροί όροι προστίθεται στην παλινδρόμηση ως θόρυβος τότε έχουμε στην ουσία τον συντελεστή συσχέτισης  $r$ . Εάν επιπλέον δεδομένα τοποθετηθούν στα πειραματικά μας δεδομένα, η μετρική τότε λέγεται συντελεστής πολλαπλής συσχέτισης. Σε κάθε περίπτωση όμως λαμβάνει τιμές από μηδέν έως ένα. Ωστόσο υπάρχουν φορές που λαμβάνει αρνητικές τιμές. Αυτό το φαινόμενο εμφανίζεται όταν η εκτίμηση δεν χρησιμοποιεί κάποιο μοντέλο για την προσαρμογή. Πιο σπάνια, σε περιπτώσεις που γίνεται προσαρμογή με κάποιο μοντέλο που δεν είναι γραμμικό ή δεν ταιριάζει στην μορφολογία των δεδομένων. Για αυτόν τον λόγο, ο ορισμός του Nash-Sutcliffe είναι ο πιο αποδεκτός

με τιμές από το μείον άπειρο έως το 1 στην πράξη ενώ ο πιο γενικός από θεωρία είναι ο  $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$  όπου  $SS_{res} = \sum (y_i - f(x_i))^2$  και  $SS_{tot} = \sum (\frac{y_i - 1}{n \sum y_i})^2$ . Τέλος, μία παραλλαγή αυτής της μετρικής είναι ο Προσαρμένος  $R^2$  όπου κοιτάμε πλέον το τυπικό σφάλμα της παλινδρόμησης αντί της τυπικής απόκλισης των σφαλμάτων και ο τύπος δίνεται από την Σχέση 4.25.

$$AdjR^2 = 1 - \frac{(n-1)}{(n-k-1)(1-R^2)} \quad (4.25)$$

όπου  $n$  το μέγεθος των πειραμάτων και  $k$  το πλήθος των ανεξάρτητων μεταβλητών. Η μετρική αυτή μπορεί να είναι αρνητική διότι το μοντέλο μπορεί να είναι πολύπλοκο για το μέγεθος των δεδομένων και πιθανώς οι ανεξάρτητες μεταβλητές να δίνουν μικρή αξία στην πρόβλεψη. Η επιστημονική κοινότητα δεν συμφωνεί όσο αφορά τις αποδεκτές τιμές του  $R^2$ . Οι Falk και Miller, το 1992 θεώρησαν ότι οι τιμές θα πρέπει να είναι άνω του 0.10 για να είναι αποδεκτές [46]. Ο Cohen το 1988 πρότεινε τον διαχωρισμό κοντά σε τιμές. Τιμές κοντά στο 0.02 θεωρούνται ασθενείς, κοντά στο 0.13 έχουν κάποιο βάρος ενώ άνω του 0.26 θεωρούνται αποδεκτές και ουσιώδεις [47]. Ο Chin το 1998 συμφώνησε με τον διαχωρισμό αλλά πρότεινε πιο μεγάλες τιμές. Συγκεκριμένα, τιμές κοντά στο 0.19 θεωρούνται ασθενείς. Τιμές κοντά στο 0.33 έχουν βάση ενώ κοντά στο 0.67 θεωρούνται αποδείξεις [48]. Τέλος, ο Hair τόσο το 2011 όσο και το 2013 με τα άρθρα που δημοσίευσε, θεωρεί ιδανικές τιμές τις 0.25, 0.50 και 0.75 για ασθενείς, κάποιο βάρος και σημαντικές αποδείξεις αντίστοιχα [49].

Η μετρική  $RMSE$  ή ρίζα του μέσου τετραγωνικού σφάλματος αποτελεί ένα συχνά χρησιμοποιούμενο μέτρο για τις διαφορές μεταξύ των τιμών που προέβλεψε ένα μοντέλο ή ένας εκτιμητής και τις πειραματικές τιμές. Ο μαθηματικός της τύπος είναι  $\sqrt{\frac{\sum_1^n (y_i - f(x_i))^2}{n}}$ . Η μετρική αυτή αντιπροσωπεύει την τετραγωνική ρίζα της δεύτερης στιγμής των διαφορών του δείγματος  $\frac{1}{n} \sum x_i^2$  ανάμεσα σε εκτιμώμενες και πειραματικές τιμές. Αυτές οι αποκλίσεις μεταξύ τους ονομάζονται υπόλοιπα όταν οι υπολογισμοί γίνονται στα δεδομένα που χρησιμοποιήθηκαν για εκτίμηση και σφάλματα όταν υπολογίζονται μακριά από το δείγμα. Αποτελεί μέτρο ακρίβειας, για την σύγκριση σφαλμάτων διαφορικών μοντέλων για ένα συγκεκριμένο σύνολο δεδομένων και όχι μεταξύ των συνόλων δεδομένων, καθώς αυτό εξαρτάται από άλλες κλίμακες. Συγκεντρώνει δηλαδή τα μεγέθη των σφαλμάτων στις προβλέψεις σε ένα μόνο μέτρο προγνωστικής ισχύος. Η μετρική αυτή λαμβάνει μόνο μη αρνητικές τιμές. Η μηδενική τιμή είναι απλησίαστη στην πράξη και υποδηλώνει τέλεια προσαρμογή της καμπύλης στα δεδομένα. Γενικότερα, μία χαμηλή τιμή σε αυτήν την

μετρική είναι πιο αποδοτική από ότι μία υψηλότερη. Παρόλα αυτά, θα πρέπει να αποφεύγουμε συγκρίσεις λόγω της διαφορετικότητας των δεδομένων. Από μαθηματικής σκοπιάς, η μετρική αυτή είναι η ρίζα των μέσων τετραγωνικών σφαλμάτων και η επιρροή κάθε σφάλματος είναι ανάλογη του μεγέθους του σφάλματος αυτού. Βλέπουμε λοιπόν ότι μεγάλα σφάλματα έχουν δυσανάλογα μεγάλη επιρροή στην τελική τιμή της μετρικής αυτής και προφανώς αντιλαμβανόμαστε την ευαισθησία της σε ακραίες τιμές.

Η μετρική  $MAE$  ή μέσο απόλυτο σφάλμα, αποτελεί ακόμη ένα συχνά χρησιμοποιούμενο μέτρο για τις διαφορές μεταξύ των τιμών που προέβλεψε ένα μοντέλο ή ένας εκτιμητής και τις πειραματικές τιμές. Πρόκειται για τον αριθμητικό μέσο των απόλυτων διαφορών των προαναφερθέντων τιμών. Είναι συνιφασμένη με την χρησιμοποιούμενη κλίμακα και για αυτό δεν αποτελεί ένα καθολικό μέτρο ανάμεσα σε διαφορετικά δείγματα. Είναι ο κύριος ανταγωνιστής της μετρικής  $RMSE$ . Η βασική τους διαφορά έγκειται στο γεγονός ότι η  $RMSE$  δίνει περισσότερο βάρος σε μεγάλες αποκλίσεις ανάμεσα σε πειραματικές και εκτιμηθήσες τιμές λόγω του τετραγώνου που υπάρχει στον μαθηματικό της τύπο. Αντίθετα, η μετρική  $MAE$  υποθέτει το ίδιο βάρος ανάμεσα σε αποκλίσεις. Ο μαθηματικός τύπος της  $MAE$  είναι  $\frac{1}{n} \sum_1^n |y_i - f(x_i)|$ . Ακόμη, στη μετρική αυτή κάθε σφάλμα επηρεάζει την τελική τιμή με άμεση αναλογία [50].

Σημειώνουμε ότι οι μετρικές  $R^2$  και  $Adj.R^2$  αφορούν μόνο γραμμικά μοντέλα ενώ οι  $SSE$ ,  $RMSE$  και  $MA$  χρησιμοποιούνται τόσο σε γραμμικά όσο και μη γραμμικά μοντέλα. Ο λόγος περιορισμού των μετρικών  $R^2$  και  $Adj.R^2$  κρύβεται στο γεγονός ότι ένα γραμμικό μοντέλο υποθέτει ότι η συνολική μεταβλητότητα μπορεί να διαχωριστεί σε μεταβλητότητα που εξηγείται από το μοντέλο και την μεταβλητότητα από τα σφάλματα. Από αυτό καταλήγουμε ότι το  $R^2$  λαμβάνει τιμές από 0 έως 1. Στα μη γραμμικά μοντέλα οι επιμέρους μεταβλητότητες δεν αθροίζουν στην συνολική μεταβλητότητα και αποτέλεσμα αυτού είναι οι τιμές της  $R^2$  να μην είναι από 0 έως 1.

#### 4.8 Προβλεψιμότητα των μαθηματικών μοντέλων

Όπως αναφέραμε στο συγκεκριμένο κεφάλαιο τα μαθηματικά μοντέλα ανάπτυξης χωρίζονται σε γραμμικά και μη γραμμικά. Η μετρική  $R^2$  αποτελεί, όπως είδαμε νω-

ρίτερα, ένα μέτρο προβλεψιμότητας για ένα μαθηματικό μοντέλο, ωστόσο η χρήση της περιορίζεται μόνο σε γραμμικά μαθηματικά μοντέλα. Δυστυχώς δεν έχει αναπτυχθεί ευρέως αποδεκτές μετρικές για την ικανότητα των μοντέλων να προβλέψουν μελλοντικές τιμές σε κοντινά ή μακρινά χρονικά διαστήματα. Οι περισσότερες προτάσεις πράττουν αρκετές παραδοχές ενώ δεν ασχολούνται με τη γενική συμπεριφορά του μοντέλου αλλά με την ασυμπτωτική του συμπεριφορά [51]. Η ασυμπτωτική συμπεριφορά βασίζεται σε ελέγχους υποθέσεων, μεγάλο πλήθος δεδομένων και στατιστικές ροπές. Η ασυμπτωτική συμπεριφορά των μαθηματικών μοντέλων που παρατέθηκαν είναι πέρα από την σκοπιά της έρευνάς μας καθώς στην παρούσα εργασία έχουμε χρονικούς περιορισμούς (λ.χ. 200 ημέρες ζωής σε περίπτωση ισχυρού καρκινογόνου) που δεν συνάδουν με την ασυμπτωτική συμπεριφορά.

Απόρροια αυτών των λόγων ήταν η ενασχόλησή μας με την ικανότητα πρόβλεψης που παρέχει ένα μαθηματικό μοντέλο ανάπτυξης μέσα από τα δεδομένα μας. Η διαδικασία που ακολουθήσαμε σκιαγραφείται στο αλγόριθμο που δίνουμε παρακάτω και η βάση του δόθηκε στην εργασία των Vinay G. Vaidya και Frank J. Alexandro Jr.. Ο αλγόριθμος ασχολείται με τα πρότυπα στα οποία υπάρχει επαρκής αριθμός από καταγραφές καρκινικού βάρους. Επαρκής αριθμός καταγραφών ορίστηκαν οι 15 καταγραφές. Επενθυμίζουμε ότι κάθε πρότυπο, με αυτόν τον τρόπο, έχει τουλάχιστον 15 ζεύγη από τιμές  $(x, y)$  όπου  $x$  είναι η ημέρα ελέγχου του βάρους και  $y$  το καρκινικό βάρος.

Στα πρότυπα, υλοποιούμε την μέθοδο των ελαχίστων τετραγώνων χωρίς τις τελευταίες τέσσερις (4) καταγραφές καρκινικού βάρους ενώ διατηρούμε σε ένα διάλυμα  $DV$  τις τιμές των τεσσάρων (4) τελευταίων ημερών. Επιπλέον, αποθηκεύουμε τις παραμέτρους της προσαρμογής αυτής. Άμεσα μπορούμε να υπολογίσουμε τις θεωρητικά τελευταίες 4 τιμές των καρκινικών βαρών με βάση το μαθηματικό μοντέλο, τις παραμέτρους από την προσαρμογή και το διάλυμα ημερών  $DV$ . Το τελευταίο συμβαίνει καθώς ένα διάλυμα παραμέτρων ενός μοντέλου ορίζει μοναδικά το μοντέλο και κατ' επέκταση αυτού, ορίζει μοναδική καμπύλη στο καρτεσιανό. Σκοπός μας είναι να βρούμε την ποσοστιαία διαφορά των τιμών του καρκινικού βάρους ανάμεσα στην τιμή που δίνει το αναπτυξιακό μοντέλο και την πραγματική τιμή του καρκινικού βάρους.

Επομένως, μπορούμε να ελέγξουμε εάν το εκάστοτε μοντέλο μπορεί να διατηρεί την μορφολογία του σε κάποιες συγκεκριμένες ημέρες με κάποια εύλογη αυξομείωση μεγέθους 5% έως 20% ή να συμπεράνουμε την αυθαιρετότητα του μοντέλου

σε αντίθετη περίπτωση. Στην ουσία προσπαθούμε να συγκρίνουμε την απόσταση των τιμών που προκύπτουν από την προσαρμογή με κάποιες από τις καταγραφές μας και τις πραγματικές καταγραφές. Εάν η απόσταση αυτή είναι μεγάλη, σημαίνει ότι το μοντέλο προσαρμόζεται διαφορετικά σε περισσότερες παρατηρήσεις και αλλάζει η μορφολογία του σε μεγάλο βαθμό, εάν όχι πλήρως. Με άλλα λόγια, το μοντέλο χαρακτηρίζεται από τα σημεία στην περιοχή συμμετρίας και κατάληξης. Εάν η απόσταση είναι μικρή, σημαίνει ότι το μοντέλο παραμένει σχετικά αμετάβλητο, χαρακτηρίζεται από τα σημεία έναρξης και μπορεί να προβλέψει σε εύλογο βαθμό την εξέλιξη του φαινομένου της ανάπτυξης χημικής καρκινογένεσης.

# ΚΕΦΑΛΑΙΟ 5

## ΤΑΞΙΝΟΜΗΣΗ

---

- 5.1 Το πρόβλημα της Ταξινόμησης
  - 5.2 Ταξινομητής Perceptron
  - 5.3 Ταξινομητής Support Vector Machine
  - 5.4 Ταξινομητής K - Nearest Neighbors
  - 5.5 Γενικές Μετρικές Εκτίμησης Ταξινομητών
  - 5.6 Εκτίμηση Χαρακτηριστικών
- 

Σε αυτό το κεφάλαιο θα περιγραφεί το πρόβλημα της ταξινόμησης, κυρίως από το στόχαστρο των βιοϊατρικών δεδομένων. Παρατίθενται κάποιοι βασικοί ταξινομητές που αποτέλεσαν τους θεμέλιους λίθους για την επιστήμη των Υπολογιστών αλλά διεύρυναν και την επιστήμη των Μαθηματικών. Στην συνέχεια θα υλοποιηθεί μία εκτενής αναζήτηση για εύρεση των σωστών μετρικών για την εκτίμηση των ταξινομητών και θα αναφερθεί το περίοπτο πρόβλημα της ανισορροπίας των κλάσεων. Τέλος, θα δοθεί ένας αλγόριθμος εκτίμησης χαρακτηριστικών με σκοπό να βρεθεί το πιο χρήσιμο χαρακτηριστικό που συμβάλει στην διαχωριστική ικανότητα του ταξινομητή.

## 5.1 Το πρόβλημα της Ταξινόμησης

Αρκετοί αλγόριθμοι σήμερα χαρακτηρίζονται ως Αλγόριθμοι «Εκπαίδευσης» ή Αλγόριθμοι Μηχανικής Μάθησης. Στόχος αυτών των αλγορίθμων είναι η εξόρυξη πληροφορίας από τα δεδομένα με απώτερο σκοπό να αντληθεί μία χρήσιμη σύνοψη των δεδομένων. Η σύνοψη χρησιμοποιείται για να ληφθεί αργότερα κάποια απόφαση. Όμως, οι αλγόριθμοι Μηχανικής Μάθησης δεν συνοψίζουν μόνο τα δεδομένα μας. Αντιθέτως, μπορούν να εκληφθούν ως ένα μοντέλο ή απλούστερα ως ένας ταξινομητής των δεδομένων και κατ' επέκταση να ανακαλύψουμε γεγονότα ή μοτίβα για τα δεδομένα μας που μπορούν να διαπιστωθούν στο μέλλον. Η Μηχανική Μάθηση μπορεί να χωριστεί σε δύο μεγάλες κατηγορίες: την Μηχανική Μάθηση με επίβλεψη και την Μηχανική Μάθηση χωρίς επίβλεψη. Η διαφορά τους αναφέρεται στο γεγονός της επίβλεψης, δηλαδή τα δεδομένα εισόδου δεν δίνουν κάποια πληροφορία στον αλγόριθμο μάθησης ως προς την κατηγορία που ανήκουν. Καθαυτόν τον τρόπο, όταν υπάρχει επίβλεψη, γνωρίζουμε τις κατηγορίες που ανήκουν τα δεδομένα μας και έχουμε Μηχανική Μάθηση με επίβλεψη, ενώ σε αντίθετη περίπτωση βρισκόμαστε σε Μηχανική Μάθηση χωρίς επίβλεψη [52].

Σε αυτό το σημείο, θα ήταν ουσιώδες να δώσουμε κάποιους βασικούς ορισμούς που θα μας βοηθήσουν στην συνέχεια του κειμένου. Τα δεδομένα στα οποία εφαρμόζουμε την αλγόριθμο μάθησης ονομάζονται σύνολο εκπαίδευσης. Ένα σύνολο εκπαίδευσης αποτελείται από ζεύγη  $(x, y)$  που ονομάζονται παρατηρήσεις του συνόλου εκπαίδευσης. Το διάνυσμα τιμών  $x$ , ονομάζεται διάνυσμα χαρακτηριστικών ή εισόδου. Κάθε τιμή του διανύσματος  $x$  μπορεί να είναι κατηγορική, διακριτή από ένα πεπερασμένο σύνολο ή αριθμητική (πραγματικός ή ακέραιος) ενώ η τιμή του  $y$  αντιστοιχεί στην κατηγορία που ανήκει η παρατήρηση ή αλλιώς τιμή ταξινόμησης και συντελεί την έξοδο. Το αντικείμενο της μαθησιακής διαδικασίας είναι να ανακαλυφθεί η συνάρτηση  $y = f(x)$  που μπορεί να προβλέψει την τιμή  $y$  σε σχέση με τις τιμές του  $x$  όσο το δυνατό καλύτερα. Η απειρία των τιμών του  $y$ , μπορεί να κατηγοριοποιηθεί στις παρακάτω ομάδες. Εάν ο τύπος του  $y$  είναι πραγματικός αριθμός το πρόβλημα καλείται παλινδρόμηση. Εάν ο τύπος του  $y$  λαμβάνει δύο τιμές, το πρόβλημα καλείται δυαδική ταξινόμηση. Εάν ο τύπος του  $y$  λαμβάνει πεπερασμένου πλήθους τιμές, το πρόβλημα καλείται ταξινόμηση πολλών κλάσεων και τέλος ο τύπος του  $y$  μπορεί λαμβάνει πιθανώς τιμές από ένα μη πεπερασμένο σύνολο αριθμών, όπως για παράδειγμα ένα αραιό δέντρο και σε αυτήν την περίπτωση δεν

έχει κάποια περίοπτη ονομασία.

Ένα γενικό πρόβλημα που αφορά την διαχείριση των δεδομένων, είναι η ανάγκη να κρατήσουμε κάποια δεδομένα από το αρχικό μας σύνολο εκτός της εκπαίδευσης. Το σύνολο αυτό το ονομάζουμε σύνολο δοκιμής. Σε κάποιες περιπτώσεις διατηρούμε δύο σύνολα από το αρχικό, το σύνολο επικύρωσης και το σύνολο δοκιμής. Η διαφορά είναι στο γεγονός ότι το σύνολο επικύρωσης βοηθάει στην σχεδίαση του μοντέλου, ενώ το σύνολο δοκιμής ελέγχει το πόσο καλό είναι το μοντέλο. Το πρόβλημα με το σύνολο επικύρωσης είναι ότι πολλοί αλγόριθμοι μάθησης έχουν την τάση να εκπαιδεύουν ακόμη και τον θόρυβο των δεδομένων (υπερμάθηση) ή αλλιώς δίνουν νόημα σε λεπτομέρειες που δεν είναι τυπικές στο μεγαλύτερο τμήμα του πληθυσμού. Χρησιμοποιώντας ένα σύνολο επικύρωσης και βλέποντας την απόδοση του ταξινομητή σε αυτόν τον χώρο, μπορούμε να παρατηρήσουμε πιθανές περιπτώσεις υπερμάθησης και να τις περιορίσουμε (π.χ. σε δέντρα απόφασης με αλλαγή του αριθμού των φύλλων).

Μία μέθοδος επικύρωσης που χρησιμοποιείται ενδελεχώς είναι η “k-fold cross-validation”. Σε αυτήν την μέθοδο το αρχικό σύνολο χωρίζεται σε  $k$  υποσύνολα ίσου πλήθους. Ένα από αυτά χρησιμοποιείται ως σύνολο επικύρωσης για το μοντέλου υπό έλεγχο και τα υπόλοιπα  $k - 1$  υποσύνολα χρησιμοποιούνται ως σύνολο εκπαίδευσης. Η διαδικασία αυτή επαναλαμβάνεται  $k$  φορές έτσι ώστε κάθε υποσύνολο να αποτελέσει σύνολο επικύρωσης ακριβώς μία φορά. Τα αποτελέσματα συναθροίζονται για να δοθεί τελικά η μέση εκτίμηση. Το πλεονέκτημα αυτής της διαδικασίας είναι ότι όλες οι παρατηρήσεις θα χρησιμοποιηθούν και ως σύνολο εκπαίδευσης αλλά και ως σύνολο επικύρωσης ενώ συνάμα κάθε παρατήρηση θα χρησιμοποιηθεί ακριβώς μία φορά για την επικύρωση. Η πιο συνηθισμένη επιλογή για την τιμή του  $k$  είναι το δέκα αλλά γενικότερα δεν αποτελεί μία προκαθορισμένη παράμετρο. Στην παρούσα εργασία δοκιμάστηκαν αρκετές τιμές και επιλέχθηκε η τιμή δέκα. Παραλλαγές αυτής της μεθόδου αποτελούν η «stratified k-fold cross-validation» και η «repeated cross-validation» [53].

Οι πιο μεγάλες κλάσεις αλγορίθμων είναι τα δέντρα απόφασης, perceptrons, νευρωνικά δίκτυα, μηχανές διανυσματικής στήριξης και αλγόριθμοι μάθησης με βάση κάποια υπόδειξη. Η κύρια διαφορά τους είναι η μορφή της συνάρτησης  $f$  που παρουσιάζεται σε κάθε αλγόριθμο. Στα δέντρα απόφασης η μορφή της  $f$  είναι ένα δέντρο και κάθε κόμβος του δέντρου έχει μία συνάρτηση του  $x$  που αποφασίζει σε ποιο παιδί ή παιδιά η έρευνα συνεχίζει. Τα δέντρα απόφασης ταιριάζουν για



δυναμικές ταξινομήσεις ή ταξινομήσεις πολλών κλάσεων, ειδικότερα εάν η διάσταση των χαρακτηριστικών δεν είναι πολύ μεγάλη (μεγάλος αριθμός χαρακτηριστικών μπορεί να οδηγήσει σε υπερμάθηση). Οι perceptrons είναι κατωφλικές συναρτήσεις που εφαρμόζονται στις συνιστώσες του διανύσματος  $x = [x_1, x_2 \dots x_n]$ . Ένα βάρος  $W_i$  συσχετίζεται με την  $i$ -οστή συνιστώσα, για όλες τις συνιστώσες του  $x$  και υπάρχει ένα κατώφλι  $\theta$ . Η έξοδος είναι  $+1$  εάν το  $\sum w_i x_i > \theta$  ενώ η έξοδος είναι  $-1$  εάν το άθροισμα είναι μικρότερο από το  $\theta$ . Οι perceptron είναι ιδανικοί για δυαδικές ταξινομήσεις, ακόμη και εάν ο αριθμός των χαρακτηριστικών είναι πολύ μεγάλος. Τα νευρωνικά δίκτυα είναι άκυκλα δίκτυα από perceptron, όπου οι έξοδοι κάποιων perceptron αποτελούν εισόδους για κάποιους άλλους. Είναι ιδανικοί για δυαδικές ταξινομήσεις ή ταξινομήσεις πολλών κατηγοριών καθώς πολλά perceptron μπορούν να χρησιμοποιηθούν ως έξοδοι, με έναν ή περισσότερους για κάθε κλάση. Οι αλγόριθμοι μάθησης με υπόδειξη χρησιμοποιούν ολόκληρο το σύνολο εκπαίδευσης για να αναπαραστήσουν την συνάρτηση  $f$ . Ο υπολογισμός της ταμπέλας  $y$  που συσχετίζεται με το διάνυσμα χαρακτηριστικών  $x$  μπορεί να εμπεριέχει μία πλήρη εξέταση του συνόλου εκπαίδευσης, ενώ συχνότερα κάποια προεπεξεργασία έχει υλοποιηθεί για απόδοση. Πιο ευρέως γνωστή μέθοδος αυτής της κατηγορίας είναι ο  $K - \text{Nearest Neighbor}$ . Τέλος, οι μηχανές διανυσματικής στήριξης αποτελούν μία εξέλιξη πάνω στους παραδοσιακούς αλγορίθμους που χρησιμοποιούνται για επιλογή τόσο των βαρών όσο και των κατωφλιών. Το αποτέλεσμα του ταξινομητή είναι ότι συνήθως έχει καλύτερα αποτελέσματα σε άγνωστα δεδομένα υπό το πρίσμα της ευστοχίας.

Έχοντας υλοποιήσει την προσαρμογή των μαθηματικών μοντέλων για κάθε παρατήρηση, όπως είδαμε στο προηγούμενο κεφάλαιο, αντλήσαμε τις παραμέτρους του κάθε μοντέλου και τις αποθηκεύσαμε με σκοπό να αποτελέσουν αντιπροσώπους των μοντέλων αυτών. Κάθε μαθηματικό μοντέλο από αυτά που χρησιμοποιήσαμε έχει δύο, τρεις ή τέσσερις παραμέτρους. Τις παραμέτρους αυτές μπορούμε να τις δούμε ως χαρακτηριστικά σε ένα πρόβλημα ταξινόμησης.

Στην παρούσα μεταπτυχιακή εργασία θα προσπαθήσουμε να δημιουργήσουμε τρεις αποδοτικούς ταξινομητές για τις ακόλουθες τρεις περιπτώσεις. Η πρώτη περίπτωση αφορά τις κατηγορίες Βενζοπυρένιο και Βενζοπυρένιο μαζί με Πολυαμίνες και Θειόλες, η δεύτερη περίπτωση αφορά Βενζοπυρένιο και Βενζοπυρένιο μαζί με Πολυαμίνες ενώ η τρίτη περίπτωση αφορά Βενζοπυρένιο και Βενζοπυρένιο μαζί με Θειόλες. Σκοπός της συγκεκριμένης διαδικασίας είναι να απαντήσουμε στο ερώτημα αν τα μοντέλα είναι ικανά να ξεχωρίσουν την δράση του βενζοπυραινίου από την

δράση του βενζοπυρραϊνίου μαζί με κάποια δραστηκή ουσία. Στην ουσία, προσπαθούμε να απαντήσουμε εάν στο ερώτημα αν οι καμπύλες των προτύπων που έλαβαν βενζοπυρένιο διαφέρουν αισθητά από τα πρότυπα που έλαβαν βενζοπυρένιο μαζί με την δραστηκή ουσία. Η διαχωριστική ικανότητα που μας δίνουν τα μαθηματικά μοντέλα θα τεθεί υπό έλεγχο σε κάθε μία από τις προαναφερθείσες περιπτώσεις.

## 5.2 Ταξινομητής Perceptron

Ο ταξινομητής Perceptron είναι ένας γραμμικός δυαδικός ταξινομητής [54]. Η είσοδος του αντιστοιχεί στο διάνυσμα  $x$  διάστασης  $d$  που οι συνιστώσες του είναι πραγματικοί αριθμοί και αποτελούν τα χαρακτηριστικά. Συσχετιζόμενο με τον perceptron είναι το διάνυσμα βαρών, που έχει και αυτό  $d$  διαστάσεις και οι συνιστώσες του είναι πραγματικοί αριθμοί. Κάθε perceptron έχει ένα κατώφλι  $\theta$ . Το αποτέλεσμα της διαδικασίας είναι  $+1$  εάν το  $w \cdot x > \theta$  και  $-1$  εάν το  $w \cdot x < \theta$ , όπου  $\cdot$  το εσωτερικό γινόμενο διανυσμάτων. Όταν το εσωτερικό γινόμενο είναι ίσο με το κατώφλι  $\theta$ , έχουμε μία ειδική περίπτωση και θα μελετηθεί στην συνέχεια.

Το διάνυσμα βαρών στην ουσία ορίζει ένα υπερεπίπεδο διάστασης  $d - 1$  : το σύνολο των  $x$  των οποίων το εσωτερικό γινόμενο με το αντίστοιχο διάνυσμα βαρών είναι  $\theta$ . Τα σημεία από την μία πλευρά του υπερεπιπέδου κατηγοριοποιούνται ως  $+1$  ενώ εκείνη της άλλης πλευράς με  $-1$ . Ο ταξινομητής αυτός δουλεύει όταν τα δεδομένα μας είναι γραμμικώς διαχωρίσιμα, με την έννοια ότι υπάρχει υπερεπίπεδο όπου χωρίζει όλα τα σημεία της κατηγορίας  $+1$  από εκείνα της κατηγορίας  $-1$ . Εάν υπάρχει έστω και ένα τέτοιο υπερεπίπεδο, ο αλγόριθμος συγκλίνει σε ένα από αυτά και διαχωρίζει τα σημεία. Σε αντίθετη περίπτωση που δεν υπάρχει τέτοιο υπερεπίπεδο, ο αλγόριθμος δεν συγκλίνει σε κανένα υπερεπίπεδο.

Για την εκπαίδευση αυτού του αλγορίθμου, ελέγχουμε το σύνολο εκπαίδευσης και προσπαθούμε να βρούμε το διάνυσμα βαρών και το κατώφλι  $\theta$  έτσι ώστε όλα τα χαρακτηριστικά που ανήκουν στην κατηγορία  $+1$  βρίσκονται στην μία πλευρά του υπερεπιπέδου και τα υπόλοιπα στην άλλη. Ίσως δεν είναι δυνατό να συμβεί αυτό, εφόσον δεν υπάρχει βεβαιότητα ως προς την ύπαρξη τέτοιου υπερεπιπέδου. Δεδομένου ότι υπάρχει ένα υπερεπίπεδο ο ακόλουθος αλγόριθμος σε φυσική γλώσσα θα μας οδηγήσει στην εύρεσή του. Ξεκινάμε υποθέτοντας ότι το κατώφλι είναι μηδενικό και το διάνυσμα βαρών είναι το μηδενικό διάνυσμα. Επιλέγουμε ένα ρυθμό

μάθησης  $\rho$ , που είναι ένας μικρός θετικός αριθμός και επηρεάζει την σύγκλιση του αλγορίθμου. Εάν ο  $\rho$  είναι πολύ μικρός, η σύγκλιση είναι αργή ενώ εάν είναι πολύ μεγάλος, τα όρια της απόφασης θα αλλάζουν συνέχεια και η σύγκλιση γίνεται πιο αργή, εάν κατορθωθεί. Στην συνέχεια, για όλες τις παρατηρήσεις του συνόλου εκπαίδευσης την μία μετά την άλλη, ορίζουμε  $y' = w \cdot x$ . Εάν το  $y'$  έχει κοινό πρόσημο με το  $y$ , τότε λέμε ότι η παρατήρηση έχει ταξινομηθεί σωστά ενώ σε αντίθετη περίπτωση αντικαθιστούμε το διάνυσμα βαρών  $w$  με  $w + \rho y \cdot x$ . Η τελευταία σχέση στην ουσία αλλάζει το διάνυσμα βαρών ελάχιστα προς την κατεύθυνση του  $x$  που ελέγχεται εκείνη την στιγμή. Με αυτόν τον τρόπο, είναι πιο πιθανό το  $x$  να τοποθετηθεί στην σωστή πλευρά του υπερεπίπεδου αλλά δεν είναι βέβαιο ότι θα καταλήξει στην σωστή λόγω των πιθανών πολλών αλλαγών στο διάνυσμα των βαρών.

Σε μη γραμμικώς διαχωρίσιμα δεδομένα, ο αλγόριθμος εισέρχεται σε δομή ανακύκλωσης που δεν μπορεί να τερματιστεί καθώς προσπαθεί συνέχεια να αλλάξει το διάνυσμα των βαρών. Επιπλέον πρόβλημα αποτελεί το γεγονός ότι ενώ το σύνολο εκπαίδευσης μπορεί να είναι γραμμικώς διαχωρίσιμο, το σύνολο των δεδομένων μπορεί να μην είναι. Μερικές λύσεις για αυτά τα προβλήματα είναι ο τερματισμός του αλγορίθμου μετά από κάποιες επαναλήψεις ελέγχου για όλες τις παρατηρήσεις, τερματισμός του αλγορίθμου όταν ένας συγκεκριμένος αριθμός από παρατηρήσεις σε λάθος κατηγορίες δεν αλλάζει μετά από κάποιες επαναλήψεις κ.ά.. Μία ακόμη τεχνική είναι η μείωση του  $\rho$  καθώς αυξάνονται οι επαναλήψεις.

Αρκετές φορές τα δεδομένα μας δεν είναι γραμμικώς διαχωρίσιμα εκ φύσεως. Σε τέτοιες περιπτώσεις γεννάται η ανάγκη μετατροπής των δεδομένων μας. Η συνηθέστερη τεχνική είναι ο μετασχηματισμός των δεδομένων μας σε έναν χώρο ίσης ή υψηλότερης διάστασης όπου εκεί είναι διαχωρίσιμα [55]. Ωστόσο, συχνά χάνεται η ερμηνευσιμότητα στην βωμό του μετασχηματισμού και σε βιολογικά δεδομένα ένα τέτοιο γεγονός δεν ενδείκνυται. Επιπλέον, καθώς μπορούμε πάντα να βρούμε ένα χώρο μεγαλύτερης διάστασης που θα είναι διαχωρίσιμα τα δεδομένα μας, είναι πιθανό να καταλήξουμε σε προϊόν υπερέπισης και να χάσουμε την γενικευτική μας ικανότητα. Ένα ακόμη πρόβλημα των perceptron είναι ότι το υπερεπίπεδο που θα καταλήξουμε δεν είναι μοναδικό. Κατά συνέπεια, τα υπερεπίπεδα που θα βρεθούν εάν τρέξουμε αρκετές φορές την διαδικασία αυτή δεν είναι ισάξια. Κάποια μπορεί να βρίσκονται κοντά σε παρατηρήσεις της μίας κατηγορίας ενώ κάποια άλλα να ισαπέχουν και από τις δύο ή ακόμη να βρίσκονται και κοντά στις παρατηρήσεις της άλλης κατηγορίας δημιουργώντας μία αίσθηση υπεροχής της μίας κλάσης έναντι της

άλλης. Έτσι όμως υπάρχουν περιοχές που με ένα υπερεπίπεδο μπορούν να ανήκουν σε μία κατηγορία ενώ με ένα άλλο υπερεπίπεδο ανήκουν στην άλλη κατηγορία. Θα ήταν ιδεατό εάν είχαμε ένα υπερεπίπεδο που χωρίζει δίκαια τις δύο μεριές έτσι ώστε σε πιθανά νέα δεδομένα που εμπίπτουν στην προηγούμενη περίπτωση να αντιμετωπίζονται ορθά. Στην παρούσα εργασία δεν χρησιμοποιήθηκε ο ταξινομητής perceptron αλλά δόθηκε η εκτενής αναφορά του καθώς αποτελεί βάση για τον ταξινομητή Support Vector Machine που θα δούμε ευθύς αμέσως.

### 5.3 Ταξινομητής Support Vector Machine

Ο ταξινομητής Support Vector Machine ή SVM γεννήθηκε από την ανάγκη βελτίωσης του ταξινομητή perceptron. Πρόκειται στην ουσία να μία αναβάθμιση που επιχειρεί να λύσει τα κύρια προβλήματα του perceptron. Ο ταξινομητής αυτός διαλέγει ένα συγκεκριμένο υπερεπίπεδο από αυτά του perceptron που όχι μόνο χωρίζει τα δεδομένα μας σε δύο κατηγορίες αλλά το πράττει αυτό μεγιστοποιώντας το περιθώριο [56]. Ως περιθώριο ορίζεται η απόσταση μεταξύ του υπερεπιπέδου και των κοντινότερων σημείων του συνόλου εκπαίδευσης. Η βασική ιδέα για τον ταξινομητή υπήρξε η ανάγκη για εύρεση του «καλύτερου» υπερεπιπέδου. Ωστόσο το να ορίσει κάποιος μία μετρική για το «καλύτερο» συντελεί μία τουλάχιστον προσεκτική προσπάθεια. Θα πρέπει να αναπτυχθεί μία συνάρτηση λάθους που να τιμωρεί τις παρατηρήσεις που βρίσκονται σε λάθος κατηγορία αλλά να μην τιμωρεί εξίσου εκείνα που βρίσκονται κοντά στο υπερεπίπεδο. Διαισθητικά, θα θέλαμε εκείνη τις παρατηρήσεις κοντά στο υπερεπίπεδο να τιμωρούνται λιγότερο από ότι εκείνα που πιο μακριά από το υπερεπίπεδο αλλά έχουν διαφορετική κατηγορία από αυτήν που πραγματικά φέρουν.

Στόχος, λοιπόν, του SVM είναι η επιλογή εκείνου του υπερεπιπέδου  $w \cdot x + b = 0$  που μεγιστοποιεί την απόσταση μεταξύ υπερεπιπέδου και οποιουδήποτε σημείου του συνόλου εκπαίδευσης. Θα θέλαμε οι παρατηρήσεις να βρίσκονται όσο το δυνατό πιο μακριά από το υπερεπίπεδο. Μία επιπλέον θετική ιδιότητα στην επιλογή του υπερεπιπέδου με όσο το δυνατό μεγαλύτερο περιθώριο είναι ότι ενώ μπορεί να υπάρχουν σημεία κοντά στο υπερεπίπεδο σε όλα τα δεδομένα μας, μπορεί να μην υπάρχουν στο σύνολο εκπαίδευσης. Εάν συμβεί αυτό, είναι πιο πιθανό να γίνει σωστά η ταξινόμηση αυτών των σημείων με αυτή τη μέθοδο από ότι εάν αφήναμε

σημεία να βρίσκονται κοντά στο υπερεπίπεδο. Στην περίπτωση όπου έχουμε δύο παράλληλα υπερεπίπεδα σε αποστάσεις  $A$  εκατέρωθεν του βέλτιστου υπερεπιπέδου, τα παράλληλα υπερεπίπεδα «ακουμπάνε» μία ή περισσότερες παρατηρήσεις. Αυτές οι παρατηρήσεις ονομάζονται διανύσματα στήριξης. Στις περισσότερες περιπτώσεις, για  $d$ -διάστατα σύνολα παρατηρήσεων χρειαζόμαστε  $d + 1$  διανύσματα στήριξης. Παρόλα αυτά, μπορούν να υπάρξουν περισσότερα διανύσματα στήριξης εάν αρκετά σημεία βρίσκονται πάνω στα παράλληλα υπερεπίπεδα.

Θα μπορούσαμε να πούμε ότι σε πρώτη φάση, ο ταξινομητής έχει ως είσοδο το σύνολο εκπαίδευσης και προσπαθεί να μεγιστοποιήσει την απόσταση, αλλάζοντας όποτε χρειάζεται το διάνυσμα βάρους  $w$  και το  $b$  ενώ ισχύει  $y_i(w \cdot x_i) \geq A$ . Το πρόσημο του  $y_i$  στην ουσία μας δίνει την πλευρά σε σχέση με το υπερεπίπεδο. Ωστόσο το πρόβλημα στον τύπο αυτό είναι ότι αυξάνοντας το  $w$  και το  $b$ , μπορούμε πάντα να βρούμε μεγαλύτερο  $A$ . Η λύση για αυτό το πρόβλημα είναι η κανονικοποίηση του υπερεπιπέδου μέσου του διανύσματος βαρών  $w$ . Καταλήγουμε έτσι στην ελαχιστοποίηση του  $\|w\| = \frac{1}{A}$  με κανόνα απόφασης  $y_i(w \cdot x_i + b) \geq 1$ .

Όλα τα παραπάνω ωστόσο δεν μπορούν να γενικεύσουν την εικόνα του προβλήματος της ταξινόμησης δύο κατηγοριών με την χρήση του υπερεπιπέδου. Για αυτόν τον λόγο, θα επιλέξουμε ένα υπερεπίπεδο που πιθανώς να έχει παρατηρήσεις σε λάθος μεριά ή επιπλέον να έχει και παρατηρήσεις κοντά στο υπερεπίπεδο με σκοπό να αποτελέσει λύση στο γενικευμένο πρόβλημα που επιχειρούμε να λύσουμε. Παρατηρήσεις που έχουν τα προαναφερθέντα χαρακτηριστικά θα τα ονομάσουμε παρατηρήσεις τιμωρίας. Κάθε μία παρατήρηση τιμωρίας θα δημιουργεί αλλαγές όταν προσπαθούμε να εκτιμήσουμε το υπερεπίπεδο. Όμως, το πόσο θα επηρεάζει αυτή η αλλαγή το υπερεπίπεδο εξαρτάται από την διαδικασία βελτιστοποίησης που θα χρησιμοποιήσουμε ενώ ταυτόχρονα θέλουμε οι ποινές αυτές να είναι μικρές. Η πιο συνηθισμένη συνάρτηση που χρησιμοποιείται αποτελείται από τον όρο  $\|w\|^2$  και έναν ακόμη όρο που σχετίζεται με ένα σταθερό πολλαπλάσιο του αθροίσματος των ποινών. Η ελαχιστοποίηση του πρώτου όρου συνάδει με την ελαχιστοποίηση του όρου  $\|w\|$  καθώς αποτελεί μονότονη συνάρτηση ως προς  $\|w\|$ . Ενώ, όσο αφορά τον δεύτερο όλο, η σταθερά αποτελεί παράμετρο κανονικοποίησης και αντανακλά την σημαντικότητα της μη σωστής ταξινόμησης. Η επιλογή μίας μεγάλης σταθεράς δίνει σημαίνει στην ουσία ότι δεν θέλουμε λάθος ταξινόμηση ενώ δεχόμαστε την πιθανότητα να έχουμε μικρό περιθώριο. Αντίθετα, με την επιλογή μικρής σταθεράς μας δίνεται ένα μεγάλο περιθώριο αλλά δεχόμαστε την ύπαρξη κάποιων παρα-

τηρήσεων που έχουν ταξινομηθεί λάθος. Μία τέτοια συνάρτηση είναι η ακόλουθη:  $f(x, b) = \frac{1}{2} \sum_{j=1}^d (w_j)^2 + C \sum_{i=1}^n (\max\{0, 1 - y_i(\sum_{j=1}^d w_j x_{ij} + b)\})$ . Ο πρώτος όρος  $\sum (w_i)^2$  ενθαρρύνει το γεγονός του μικρού  $\|w\|$ , ενώ ο δεύτερος δίνει νόημα στην σταθερά  $C$  και την συνάρτηση λάθους  $(\max\{0, 1 - y_i(\sum_{j=1}^d w_j x_{ij} + b)\})$ . Η συνάρτηση λάθους μειώνεται γραμμικά όταν ο δεύτερος όρος της συνάρτησης μεγίστου είναι μικρότερος από την μονάδα και παραμένει μηδέν από το σημείο ένα και μετά.

Ένας συνηθισμένος τρόπος επίλυσης της συνάρτησης  $f(x, b)$  είναι με χρήση της μεθόδου της απότομης καθόδου. Το πλεονέκτημα αυτής της μεθόδου είναι ότι σε πολύ μεγάλα δεδομένα σε όγκο, δεν χρειάζεται τα δεδομένα μας να είναι στην μνήμη του υπολογιστή οπότε εξοικονομούμε υπολογιστικές μονάδες. Η γενικά ιδέα είναι να υπολογίσουμε τις παραγώγους της  $f$  ως προς  $b$  και ως προς  $w$ . Εφόσον θέλουμε την ελαχιστοποίηση της  $f$ , αρκεί να μετακινήθουμε αντίθετα της διεύθυνσης της παραγώγου. Η μετακίνησή μας σε κάθε συνιστώσα είναι ανάλογη της παραγώγου με σεβασμό στην αντίστοιχη συνιστώσα. Συνοπτικά ο αλγόριθμος έχει ως εξής:

1. Επιλέγουμε τις παραμέτρους: πολλαπλασιαστική σταθερά της συνάρτησης λάθους  $C$  και απόσταση περιθωρίου

2. Αρχικοποίηση των τιμών του διανύσματος  $w$  επαυξημένο με το στοιχείο  $b$ .

3. Ακολουθιακά:

- a) Υπολογισμός μερικών παραγώγων με σεβασμό στο  $W_j$

- b) Αλλαγή των τιμών του διανύσματος  $w$  αφαιρώντας την ποσότητα  $A \frac{df}{dw_j}$  για κάθε  $w_j$ .

Τέλος, στην πράξη χρησιμοποιείται το δυικό του προβλήματος βελτιστοποίησης κατά Lagrange. Κατά συνέπεια, η συνάρτηση  $f(x, b)$  που αποτελούσε αντικείμενο ελαχιστοποίησης γίνεται πρόβλημα μεγιστοποίησης  $f(c_1, c_2, \dots, c_n) = \sum c_i - \frac{1}{2} \sum \sum y_i c_i x_i x_j y_j c_j$  υπό τις συνθήκες  $\sum c_i y_i = 0$  και  $0 < c_i < \frac{1}{2} n_\lambda$  ενώ  $w = \sum c_i y_i x_i$ . Όταν το  $i$ -οστή τιμή του  $c$  είναι μηδενική, βρισκόμαστε στην σωστή πλευρά του περιθωρίου. Όταν βρισκόμαστε ανάμεσα από το μηδέν και το  $\frac{1}{2} n_\lambda$ , είμαστε στα όρια του περιθωρίου. Ακολουθιακά, το διάνυσμα βαρών  $w$  μπορεί να γραφεί ως γραμμικός συνδυασμός των διανυσμάτων στήριξης. Υπάρχουν και πιο μοντέρνες τεχνικές που βασίζονται στην μέθοδο της απότομης καθόδου και δημιουργούν παραλλαγές ως προς την επιλογή κατεύθυνσης. Αντί να γίνει μετακίνηση αντίθετη της κατεύθυνσης της παραγώγου, γίνεται ένα βήμα προς ένα διάνυσμα που επιλέγεται από μερικές παραγώγους της συνάρτησης. Η επιλογή αυτή μας δίνει υπολογιστικές μονάδες καθώς δεν απαιτείται η επεξεργασία όλων των παρατηρήσεων.

Στην παρούσα εργασία χρησιμοποιήθηκε ο ταξινομητής Support Vector Machine ενώ δοκιμάστηκε ως αρχικός ρυθμός μάθησης η τιμή 0.1, 0.01 και 0.001.

## 5.4 Ταξινομητής K - Nearest Neighbors

Στην συγκεκριμένη κατηγορία ταξινόμησης, το σύνολο εκπαίδευσης χρησιμοποιείται για ειδική επεξεργασία που θα μας δώσει κάποια πληροφορία για την ταξινόμηση και για αυτόν τον λόγο απαιτείται η αποθήκευσή του δίσκο και η τοποθέτησή του στην μνήμη. Η πληροφορία αυτήν που αντλείται συνήθως χρησιμεύει για τον υπολογισμό της τελικής κατηγορίας στο πρόβλημα της ταξινόμησης. Στο διάνυσμα των χαρακτηριστικών, κάθε συνιστώσα γίνεται αντιληπτή ως ένα σημείο σε κάποιο χώρο. Όταν ένα νέο σημείο φτάσει και πρέπει να ταξινομηθεί, βρίσκουμε κάποιες παρατηρήσεις που είναι κοντά στο νέο σημείο. Με βάση τα κοντινά σημεία, λαμβάνεται η απόφαση σε ποιά κατηγορία θα πρέπει να τοποθετηθεί το νέο σημείο [57]. Ωστόσο λέγοντας «κοντινά» σημεία, ενέχουμε τον ορισμό της απόστασης, που δύναται να ποικίλει. Μερικές βασικές ερωτήσεις που αναφέρονται είναι οι επόμενες. Ποιά απόσταση χρησιμοποιούμε; Πόσους κοντινούς γείτονες χρειάζεται να εξετάσουμε για να ληφθεί μία απόφαση; Τί βάρος έχουν οι κοντινοί γείτονες; Είναι ισάξιοι;

Η πιο απλή περίπτωση μάθησης με κοντινότερους γείτονες είναι αυτήν όπου ελέγχουμε μόνο έναν γείτονα. Τυπικά, απαιτείται μία συνάρτηση απόστασης (συνάρτηση πυρήνα) μεταξύ της νέας παρατήρησης και των κοντινότερών της γεινόνων, ενώ χρησιμοποιούμε την συνάρτηση αυτή με σκοπό να δούμε το βάρος των γεινόνων. Σε αυτήν την περίπτωση, δεν χρειάζονται βάρη στους γείτονες οπότε αποφεύγουμε την χρήση συνάρτησης πυρήνα. Ενδιαφέρουσα πρόταση αποτελεί η χρήση της κανονικής κατανομής ως συνάρτηση πυρήνα. Επιπλέον, ο κανόνας μας γίνεται απλός, λαμβάνουμε την ετικέτα του γείτονά μας.

Ωστόσο, λέγοντας απόσταση, μέχρι αυτό το σημείο εννοούσαμε την Ευκλείδεια απόσταση. Παρόλα αυτά, οι περισσότερες τεχνικές δουλεύουν για οποιονδήποτε ορισμό της απόστασης. Κάποιος θα μπορούσε να χρησιμοποιήσει την απόσταση Jaccard ή την Manhattan.

Στην παρούσα εργασία χρησιμοποιήθηκε ο ταξινομητής K-Nearest-Neighbor με χρήση της Ευκλείδεια απόστασης.

## 5.5 Γενικές Μετρικές Εκτίμησης Ταξινομητών

Για την αποτίμηση των ταξινομητών απαιτούνται κάποιες μετρικές που θα προσπαθήσουν να αιχμαλωτίσουν την επιτυχία ή μη της διαδικασίας της ταξινόμησης. Η προφανής μετρική που θα μπορούσε κάποιος να σκεφτεί για την ταξινόμηση είναι η ευστοχία του ταξινομητή, δηλαδή το κατά πόσο μπορεί να προβλέψει σωστά το αποτέλεσμα. Γενικότερα όμως απαιτείται μία πιο σφαιρική εικόνα οπότε δημιουργήθηκε η ανάγκη για τον πίνακα σύγχυσης (πίνακας 5.1).

	<i>Actual Class A</i>	<i>Actual Class B</i>
<i>Predicted Class A</i>	TP	FP
<i>Predicted Class B</i>	FN	TN

Πίνακας 5.1: Πίνακας Σύγχυσης

Χρησιμοποιώντας τιμές του πίνακα θα δώσουμε και τους πρώτους ορισμούς:

1. TP = Μία παρατήρηση ήταν κατηγορίας A και προβλέφθηκε ορθά ως κατηγορίας A.
2. FN = Μία παρατήρηση ήταν κατηγορίας A και προβλέφθηκε λαθεμένα ως κατηγορίας B.
3. TN = Μία παρατήρηση ήταν κατηγορίας B και προβλέφθηκε ορθά ως κατηγορίας B.
4. FP = Μία παρατήρηση ήταν κατηγορίας B και προβλέφθηκε λαθεμένα ως κατηγορίας A.

<i>Metric</i>	<i>Formula</i>
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision (PPV)	$\frac{TP}{TP+FP}$
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$

Πίνακας 5.2: Γενικές μετρικές απόδοσης των ταξινομητών.

Εφόσον η ενασχόληση της παρούσας εργασίας εμπίπτει σε ιατρικά δεδομένα, θα χρησιμοποιούμε τους όρους «το πρότυπο έχει την πάθηση» ως κατηγορία A και «το πρότυπο δεν έχει την πάθηση» ως κατηγορία B για τους επόμενους ορισμούς. Οι



επόμενες τέσσερις μετρικές αποτελούν τις βασικότερες στον τομέα της αποτίμησης της απόδοσης ενός ταξινομητή [58] ενώ οι μαθηματικοί τύποι τους δίνονται στον πίνακα 5.2.

Η μετρική Accuracy αιχμαλωτίζει την εικόνα της επιτυχίας του ταξινομητή και αποτελεί ένα ουσιώδες μέτρο για να καταλάβουμε ποιά είναι η πιθανότητα της σωστής τοποθέτησης μίας παρατήρησης σε μία κατηγορία από τις δύο. Παρόλα αυτά, μπορεί να μας οδηγήσει σε λάθος συμπεράσματα όταν οι δύο κλάσεις μας δεν έχουν ισορροπία ως προς τον αριθμό των παρατηρήσεών τους. Ένα απλό παράδειγμα είναι η περίπτωση όπου έχουμε ένα σύνολο δεδομένων με 1000 παρατηρήσεις εκ των οποίων 990 είναι της μίας κατηγορίας και 10 της άλλης. Ένας απλός ταξινομητής που απλά θα επιλέξει ότι όλες οι παρατηρήσεις είναι της πρώτης κατηγορίας θα έχει 99% ευστοχία. Η ευστοχία αυτή όμως δεν συνάδει με την πραγματικότητα. Υπάρχουν αρκετές τεχνικές για την εξισορρόπηση των κατηγοριών όπως λ.χ. νέα δειγματοληψία με σκοπό να έχουμε λόγο ένα προς ένα ανάμεσα στις δύο κατηγορίες, νέα δειγματοληψία από τον μικρό πληθυσμό με σκοπό να έχουμε τυπικά διπλότυπα (απαιτείται αξιόλογο αρχικό δείγμα), χρήση πιο ανθεκτικών μετρικών που έχουν ιδιαίτερη ανένχια σε ανισοροπίες κλάσεων κ.ά..

Η μετρική Precision ή Positive Predictive Value μας δίνει την ευστοχία των προτύπων που έχουν την πάθηση και αναγνωρίστηκαν με την αυτήν από το σύνολο των προτύπων που αναγνωρίστηκαν με την πάθηση. Ύψηλές τιμές στην μετρική Precision μας δίνουν την ένδειξη ενός καλού ταξινομητή.

Η Recall ή Sensitivity ελέγχει την ευστοχία των προτύπων που αναγνωρίστηκαν με την πάθηση από το σύνολο των προτύπων που έχουν την πάθηση. Μπορούμε να καταλαβαίνουμε την σημαντικότητα αυτής της μετρικής καθώς υψηλές τιμές δείχνουν ηχηρή ισχύ στον ταξινομητή μας στον εντοπισμό των προτύπων με την πάθηση. Επιπλέον, γίνεται αντιληπτό ότι η κλάση με τις λιγότερες παρατηρήσεις θα είναι αυτή με την ασθένεια και κατ' επέκταση η μετρική Recall δεν επηρεάζεται το ίδιο από την ανισοροπία των κλάσεων όπως πράττει η Accuracy. Η μετρική Specificity, ελέγχει την ευστοχία των προτύπων χωρίς την πάθηση από το σύνολο των προτύπων που δεν έχουν την πάθηση. Σε πολλές αναλύσεις αναφέρεται το ισοζύγιο μεταξύ της μετρικής Sensitivity και της μετρικής Specificity ενώ επισημάνεται η απαίτηση προσεκτικής διαχείρισης στην σχέση αμφοτέρων.

Λόγω της παραπάνω ανάλυσης, θα πρέπει να ισορροπήσουμε τους ρυθμούς των  $FP$  και  $FN$ . Οι μετρικές  $G$ -Mean,  $F_{\beta}$  - Measure και Adjusted  $F$  - Measure θα προ-

<i>Metric</i>	<i>Formula</i>
Geometric Mean	$\sqrt{Sensitivity \times Specificity}$
F-Measure	$\frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision}$
$F_\beta$	$(1 + \beta^2) \frac{Precision \cdot Sensitivity}{(\beta^2 \cdot Sensitivity) + Sensitivity}$
Adjusted F-Measure	$\sqrt{F_2 \times InvF_{0.5}}$

Πίνακας 5.3: Μετρικές Εκτίμησης Ταξινομητών με ειδίκευση στην ανισορροπία κλάσεων

σπαθήσουν να πετύχουν ακριβώς αυτήν την ισορροπία [58] και οι μαθηματικοί τους τύποι δίνονται στον πίνακα 5.3. Η μετρική *G-Mean* θα προσπαθήσει να το πετύχει ισορροπώντας τους τιμές των μετρικών *Sensitivity* και *Precision*. Η τιμή της είναι μηδενική όταν η τιμή της *Sensitivity* ή της *Precision* είναι μηδενική. Μικρές τιμές στην μετρική *G-Mean* αποτελεί ένδειξη κακής απόδοσης του ταξινομητή στις θετικές περιπτώσεις πάθησης ακόμη και αν γίνεται σωστά η υπόλοιπη ταξινόμηση. Άρα το συγκεκριμένο μέτρο δρα ως ακρογωνιαίος λίθος για την αποφυγή υπερεκπαίδευσης της κλάσης με τις λιγότερες παρατηρήσεις και υποεκπαίδευσης της κλάσης με τις περισσότερες παρατηρήσεις. Με παρόμοιο τρόπο και κοινό σκοπό, την ισορροπία των *Sensitivity* και *Precision*, αναπτύχθηκε η *F-Measure*. Ωστόσο, δεν έχει περίοπτα αποτελέσματα με ύπαρξη ανισορροπίας κλάσεων. Για αυτό τον λόγο και διατηρώντας την ιδέα της ισορροπίας των ρυθμών FP και FN και κατ' επέκταση των *Sensitivity* και *Precision*, αναπτύχθηκε η μετρική *Adjusted F-Measure*. Η προσαρμοσμένη αυτή μετρική υπολογίζει αρχικά δύο ποσότητες, την  $F_2$  και την  $InvF_{0.5}$ . Η  $F_1$  συντελεί τον αρμονικό μέσο ανάμεσα στην *Precision* και *Sensitivity*, ενώ ο γενικότερος τύπος δίνεται παραπάνω. Η παράμετρος  $\beta$  αποτελεί έναν θετικό πραγματικό αριθμό που δηλώνει ότι η *Sensitivity* θεωρείται  $\beta$  φορές για να είναι το ίδιο σημαντική με την *Precision*. Καθαυτόν τον τρόπο καταλαβαίνουμε ότι η *Recall* είναι πιο σημαντική σε σχέση με την *Precision* σε τιμές του  $\beta$  που είναι μεγαλύτερες

του 1. Ενώ το αντίθετο συμβαίνει για τιμές μικρότερες του 1. Επιλέγοντας για  $\beta$  το 2, ο υπολογισμός της  $F_2$  είναι ευθύς ενώ για τον υπολογισμό της  $InvF_{0.5}$  αλλάζουμε τις ετικέτες των δύο κλάσεων με σκοπό η κλάση που αποτελεί μειονότητα στο σύνολο των παρατηρήσεων να γίνει πλειονότητα και επιπλέον η πλειονότητα να γίνει μειονότητα. Λαμβάνουμε άμεσα την αίσθηση ότι μία τακτική σαν αυτή αποτελεί μία καλή προσέγγιση για το πρόβλημα της ανισορροπίας. Τέλος, έχοντας υπολογίσει την  $F_2$  και την  $InvF_{0.5}$ , λαμβάνουμε τον γεωμετρικό τους μέσο που αποτελεί την Adjusted  $F - Measure$ . Η Adjusted  $F - Measure$  χρησιμοποιεί λοιπόν την πληροφορία τόσο του αρχικού πίνακα σύγχυσης όσο και εκείνης που αφορά την έρευνα μοτίβων στην κλάση που έχει τις λιγότερες παρατηρήσεις.

Στην παρούσα εργασία χρησιμοποιήθηκαν όλες οι προαναφερθείσες μετρικές με σκοπό την εκτίμηση των ταξινομητών Support Vector Machines με γραμμική συνάρτηση πυρήνα και K-Nearest-Neighbor.

## 5.6 Εκτίμηση Χαρακτηριστικών

Ο αλγόριθμος Relief αναπτύχθηκε από τους Kira Και Rendell το 1992 και υλοποιεί μία προσέγγιση φιλτραρίσματος με σκοπό την επιλογή των χαρακτηριστικών που είναι πιο ευαίσθητα σε αλληλεπιδράσεις χαρακτηριστικών [59]. Η πρώτη του εφαρμογή ήταν σε δυαδικά προβλήματα ταξινόμησης με αριθμητικά χαρακτηριστικά. Ο αλγόριθμος υπολογίζει ένα σκορ για κάθε χαρακτηριστικό με σκοπό να ταξινομήσει τα χαρακτηριστικά με βάση την συμβολή τους στην ταξινόμηση. Το τρόπο υπολογισμού του σκορ βασίζεται στην ανίχνευση διαφορών στις τιμές των χαρακτηριστικών των κοντινότερων γειτόνων. Εάν μία διαφορά εντοπιστεί ανάμεσα σε δύο χαρακτηριστικά της ίδιας κατηγορίας, το σκορ του χαρακτηριστικού μειώνεται. Αντίθετα, εάν μία διαφορά εντοπιστεί ανάμεσα σε δύο διαφορετικά διαφορετικής κατηγορίας, το σκορ του χαρακτηριστικού αυξάνεται.

Αρκετές παραλλαγές προτάθηκαν στην συνέχεια με βάση τον αλγόριθμο Relief οι οποίες ονομάστηκαν οικογένειες RBAs. Οι οικογένειες αυτές προσπαθούν να προσαρμοστούν πάνω σε προβλήματα θορύβου, να δώσουν γενίκευση σε προβλήματα πολλών κλάσεων, να προσφέρουν ανέχεια σε έλλειψη κάποιων τιμών των χαρακτηριστικών και να λύσουν προβλήματα που δεν έχουν αριθμητικά χαρακτηριστικά. Παρόλα αυτά η κεντρική ιδέα παραμένει ακέραια και η χρήση του τόσο για επι-

λογή των κορυφαίων χαρακτηριστικών σε πίνακες πολλών χαρακτηριστικών όσο και σε ταξινόμηση χαρακτηριστικών είναι αδιαμφισβήτητη.

Η μεγαλύτερη δύναμη της συγκεκριμένης οικογένειας αλγορίθμων είναι ότι δεν εξαρτώνται σε ερευνητικές μεθόδους ενώ η υλοποίησή τους γίνεται σε πολυωνυμικό χρόνο. Η ανέχειά τους σε θόρυβο και η ανθεκτικότητα τους παρά τις αλληλεπιδράσεις των χαρακτηριστικών αποτελούν ακόμη δύο εύρωστα πλεονεκτήματα. Ωστόσο, ένα βασικό μειονέκτημα της μεθόδου είναι ότι δεν δημιουργεί διακρίσεις μεταξύ περιττών χαρακτηριστικών και πιο χρήσιμων. Θεωρεί όλα τα χαρακτηριστικά ισάξια στην αρχικοποίησή του.

Τέλος, παραθέτουμε τον βασικό αλγόριθμο Relief. Έστω ένα σύνολο δεδομένων  $n$  παρατηρήσεων και  $z$  χαρακτηριστικών που ανήκουν στην κατηγορία  $A$  ή την κατηγορία  $B$ . Στο σύνολο των χαρακτηριστικών δημιουργούμε μία μεταφορά του κάθε χαρακτηριστικού στην κλίμακα από μηδέν έως ένα. Ο αλγόριθμος θα επαναληφθεί  $m$  φορές. Ξεκινάμε με ένα μηδενικό διάνυσμα βαρών  $w$ . Σε κάθε επανάληψη διαλέγεται τυχαία μία παρατήρηση  $x$ . Μαζί με αυτήν την παρατήρηση διαχωρίζουμε και δύο ακόμη παρατηρήσεις. Τις παρατηρήσεις που είναι πιο κοντά παρατήρηση που διαλέξαμε τυχαία και ανήκουν η μία στην ίδια κλάση και η άλλη στην διαφορετική κλάση. Η κοντινότερη παρατήρηση της ίδιας κλάσης ονομάζεται "near hit" ενώ η κοντινότερη παρατήρηση της διαφορετικής κλάσης ονομάζεται "near miss". Τέλος, ενημερώνουμε το διάνυσμα βαρών με τύπο:  $W_i = w_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2$ .

Κατά αυτόν τον τρόπο, το βάρος κάθε χαρακτηριστικού μειώνεται εάν διαφέρει περισσότερο από την κοντινότερη παρατήρηση της ίδιας κλάσης από ότι της διαφορετικής κλάσης. Ενώ το αντίθετο συμβαίνει στην άλλη περίπτωση. Μετά από  $m$  επαναλήψεις, διαιρούμε κάθε στοιχείο του διανύσματος βαρών με το  $m$ . Ενώ ορίζουμε ένα κατώφλι όπου όποια χαρακτηριστικά έχουν τιμή μεγαλύτερη από αυτήν, θεωρούνται σημαντικά. Μεγαλύτερη τιμή σημαίνει και μεγαλύτερη ισχύ του χαρακτηριστικού ως επιλογή.

Στην παρούσα εργασία χρησιμοποιήθηκε ο αλγόριθμος ReliefF με σκοπό να ταξινομήσουμε τα χαρακτηριστικά ως προς την συμβολή τους στην ταξινόμηση και όχι να βρούμε τα καλύτερα για να απομονώσουμε εκείνα για την ταξινόμηση [60]. Άλλωστε ο αριθμός των χαρακτηριστικών μας ήταν μικρός. Όσο αφορά τον αλγόριθμο ReliefF, η κύρια διαφορά του σε σχέση με τον βασικό αλγόριθμο είναι η χρήση της απόστασης Manhattan αλλά και η αντικατάσταση των τετραγώνων της διαφοράς των σημείων  $x_i$  και  $nearhit_i$  ή  $nearmiss_i$  με την απόλυτη διαφορά αυτών. Επιπλέον,

έναντι των  $m$  επαναλήψεων, έχουμε εξαντλητική επανάληψη (π.χ. 10000 φορές) για κάθε παρατήρηση ενώ αντί να ελέγχουμε το κοντινότερο miss και hit, ελέγχουμε τα  $k$  κοντινότερα *hits* και  $k$  κοντινότερα *misses* και βρίσκουμε τον μέσο όρο της συνεισφοράς τους στο διάλυσμα βαρών.

## ΚΕΦΑΛΑΙΟ 6

# ΣΤΑΤΙΣΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΚΑΤΑΛΗΞΕΩΝ

---

- 6.1 Κανονικότητα των δεδομένων
  - 6.2 Τεστ Κανονικότητας
  - 6.3 Q-Q γράφημα και κανονικότητα
  - 6.4 Μετατροπή των δεδομένων με κανονικοποίηση (z-score ή minmax-score)
  - 6.5 Μη Παραμετρικό Τεστ των Kolmogorov-Smirnov Δύο Δειγμάτων
  - 6.6 Στατιστικές ροπές
  - 6.7 Το μη παραμετρικό τεστ των Mann - Whitney U Δύο Πληθυσμών
- 

Αυτό το κεφάλαιο εστιάζει στο προσδόκιμο ζωής των πειραματοζώων που τους δόθηκε μία ουσία έναντι πειραματοζώων που τους δόθηκε μία άλλη ουσία. Στην αρχή του κεφαλαίου αναφέρεται η ιδιαίτερη ιδιότητα της κανονικότητας και παρουσιάζονται αρκετά μη παραμετρικά τεστ με σκοπό να δοθούν ενδείξεις κανονικότητας. Στόχος είναι οι ενδείξεις να γίνουν αποδείξεις προτίστως της αποτελεσματικότητας των δραστικών ουσιών ως προς το προσδόκιμο ζωής και δευτερευόντως να υπογραμμίσουμε ότι η οδός της αύξησης των δόσεων είναι η ορθή χωρίς να φέρει επιπλοκές. Τέλος, επισημαίνονται οι γνωστότεροι μετασχηματισμοί δεδομένων και κανονικοποιήσεις που πιθανώς επιφέρουν κάποια βελτίωση στο πρόβλημα της μη-κανονικότητας ενώ συνάμα μπορεί να μπερδεύουν περισσότερο την κατάσταση λόγω έλλειψης ερμηνευσιμότητας της διαδικασίας. Σκοπός του κεφαλαίου είναι να θέσει την βάση για να απαντηθούν τα παρακάτω βαρύγδουπα ερωτήματα:

α) «Υπάρχει ουσιαστική διαφορά ανάμεσα στους πληθυσμούς που έχουν λάβει μία ουσία έναντι μίας άλλης;»

β) «Υπάρχει ουσιαστική διαφορά ανάμεσα στους πληθυσμούς που έχουν λάβει μία ουσία σε μονή δόση έναντι της ίδιας ουσίας σε διπλή δόση;»

γ) «Υπάρχει ουσιαστική διαφορά ανάμεσα σε ουσίες με κοινά γνωρίσματα;»

Ερωτήματα που καταλήγουν στο βασικότερο στόχο όλων, την ποσοτικοποίηση των δόσεων δεδομένου ενός εύρους προσδόκιμου ζωής.

## 6.1 Κανονικότητα των δεδομένων

Η κανονική κατανομή αποτελεί την σπουδαιότερη κατανομή της Θεωρίας Πιθανοτήτων και Στατιστικής. Οι βασικότεροι λόγοι που την δίνουν αυτήν την ιδιαίτερη θέση είναι προτίστως το γεγονός ότι πολλές τυχαίες μεταβλητές περιγράφονται σε τεράστιο βαθμό από την κανονική κατανομή ή περιγράφονται από κατανομές που μπορούμε να τις προσεγγίσουμε με την χρήση της κανονικής κατανομής. Δευτερευόντως, οι ιδιότητες που περικλύουν την κανονική κατανομή δύναται να αξιοποιηθούν στην Στατιστική Συμπερασματολογία και να αποσαφηνίσουν το πρόβλημά μας σε ικανοποιητικό βαθμό.

Ο Abraham De Moivre το 1733 και αργότερα ο Laplace το 1812 έθεσαν τις βάσεις για το σημαντικότερο θεώρημα στην Θεωρία Πιθανοτήτων και Στατιστικής, το Κεντρικό Οριακό Θεώρημα. Το συγκεκριμένο Θεώρημα μας λέει ότι εάν από έναν πληθυσμό που έχει μέση τιμή  $\mu$  και διακύμανση  $\sigma^2$ , επιλέξουμε τυχαία  $\nu$  δείγματα και υπολίσουμε τους μέσους όρους και τις διακυμάνσεις αυτών, εάν το  $\nu$  είναι αρκετά μεγάλος αριθμός, η κατανομή των μέσων και των διακυμάνσεων θα ακολουθεί κανονική κατανομή μέσου όρου  $\mu$  και διακύμανση  $\frac{\sigma^2}{\nu}$ . Όσο μεγαλύτερο είναι το μέγεθος  $\nu$  των δειγμάτων, τόσο καλύτερη η προσέγγιση της κατανομής των δειγματικών μέσων από την κανονική κατανομή. Καθαυτόν τον τρόπο, καταστείται σαφές η σημαντικότητα της κανονικής κατανομής και κατά επέκταση η κανονικότητα των δεδομένων, δηλαδή η ιδιότητα των δεδομένων να προσεγγίζονται από την κανονική κατανομή.

## 6.2 Τεστ Κανονικότητας

Μεγάλος είναι ο αριθμός των τεστ κανονικότητας τα τελευταία χρόνια. Τα συγκεκριμένα τεστ αποτελούν στατιστικές ενδείξεις και απαντάνε στο ερώτημα ένας πληθυσμός μίας δειγματοληψίας ακολουθεί κανονική κατανομή δεδομένου ενός επιπέδου σημαντικότητας (συνήθως επιλέγεται το 5%). Το επίπεδο σημαντικότητας συντελεί τον κανόνα για την αποδοχή ή απόρριψη μίας στατιστικής υπόθεσης. Τα γνωστότερα τεστ αυτής την κατηγορίας είναι τα Anderson-Darling, Jarque-Bera, Lilliefors και Shapiro-Wilk. Σύμφωνα με έρευνες του 2011, το πιο δυνατό τεστ κανονικότητας αποτελεί το Shapiro-Wilk ενώ παρόμοια αποτελέσματα παρουσιάζει το Anderson-Darling [61]. Σε όλα τα τεστ που ακολουθούν ελέγχεται αν ένα δείγμα αντλήθηκε από την κανονική κατανομή (υπόθεση  $H_0$ ). Έτσι, εάν η τιμή  $p$ -value είναι μικρότερη από ένα  $\alpha$  που δηλώνει το επίπεδο σημαντικότητας (συνήθως 5%), τότε η υπόθεση  $H_0$  είναι αληθής. Επομένως, δεν έχουμε αρκετές αποδείξεις να απορρίψουμε την άντληση από την κανονική κατανομή. Αντίθετα εάν η τιμή  $p$ -value είναι μεγαλύτερη της  $\alpha$ , δεχόμαστε ότι έχουμε αρκετές αποδείξεις για να απορρίψουμε την  $H_0$ , επομένως το δείγμα μας δεν αντλήθηκε από την κανονική κατανομή.

### 6.2.1 Το τεστ των Anderson-Darling

Δημιουργήθηκε από τους Theodore Wilbur Anderson και Donald A. Darling το έτος 1952 [62]. Το συγκεκριμένο τεστ προσπαθεί να απαντήσει στο ερώτημα εάν ένα δείγμα έχει αντληθεί από μία συγκεκριμένη κατανομή πιθανότητας. Στην δική μας περίπτωση η κατανομή πιθανότητας μας είναι η κανονική κατανομή ενώ η γνώση μας για αυτήν είναι περιορισμένη. Το τεστ θα εκτιμήσει τις παραμέτρους της κανονικής κατανομής που είναι ο μέσος όρος και η διακύμανση με βάση το Κεντρικό Οριακό Θεώρημα. Η εκτίμηση των δύο παραμέτρων δίνεται από τις Σχέσεις 6.1 και 6.2 και γίνεται με βάση το δείγμα που δόθηκε. Με  $X_i$  συμβολίζεται η πειραματική παρατήρηση και βρίσκονται σε αύξουσα σειρά.

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i \quad (6.1)$$

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (6.2)$$

Οι τιμές  $X_i$  στην συνέχεια κανονικοποιούνται με βάση τον τύπο  $Y_i = \frac{(X_i - \mu)}{\sigma}$ . Δημιουργούμε την Αθροιστική Συνάρτηση Κανονικής Κατανομής  $\Phi$  μετά την εκτίμηση



των παραμέτρων που δημιουργήσαμε και υπολογίζουμε τις τιμές  $A^2$  και  $A^{*2}$ , όπως φαίνεται στις Σχέσεις 6.3 και 6.4.

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1(\ln\Phi(Y_i) + \ln(1 - \Phi(Y_{n+1-i})))) \quad (6.3)$$

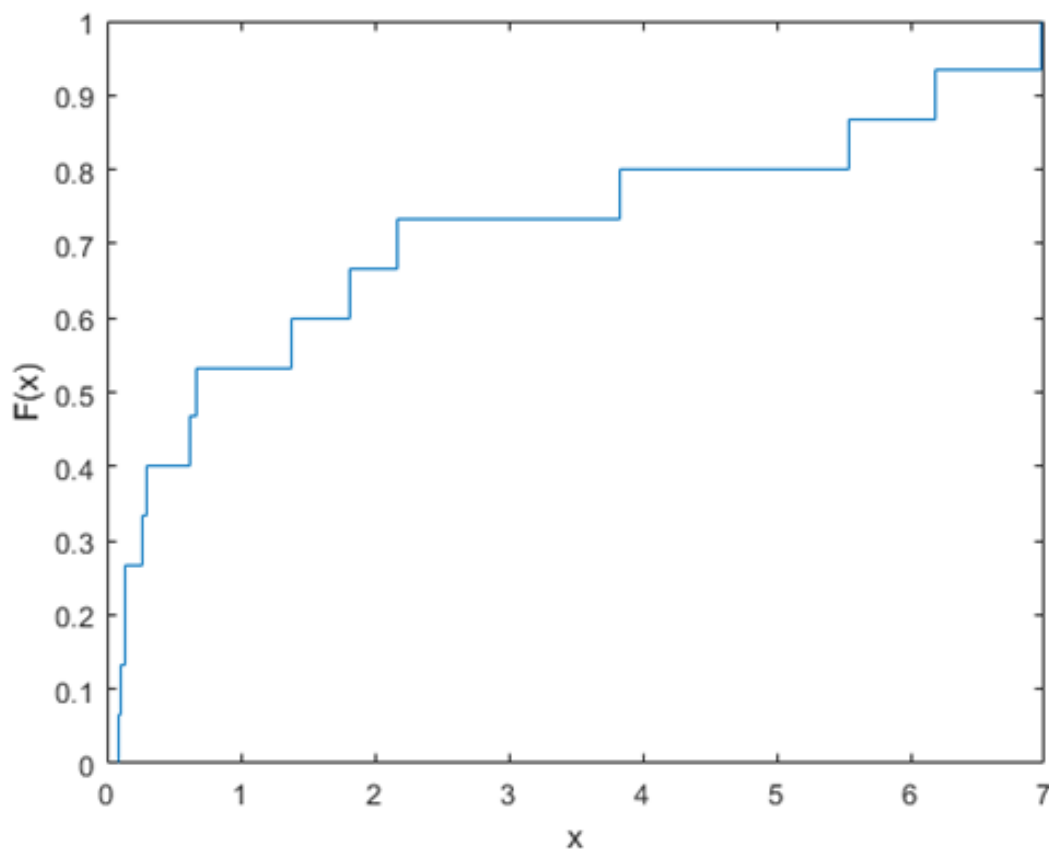
$$A^{*2} = A^2 \left(1 + \frac{4}{n} - \frac{25}{n^2}\right) \quad (6.4)$$

Τέλος, ο κανόνας απόρριψης που ακολουθείται είναι όταν τουλάχιστον μία ποσότητα από τις  $A^2$  και  $A^{*2}$  είναι μεγαλύτερη από την κρίσιμη τιμή, απορρίπτουμε την αρχική υπόθεση δηλαδή το δείγμα δεν αντλήθηκε από την κανονική κατανομή. Η κρίσιμη τιμή δίνεται με βάση από το επίπεδο σημαντικότητας και το μέγεθος του δείγματος από τους αντίστοιχους στατιστικούς πίνακες (πίνακες A). Επιπλέον, είναι σημαντικό να αναφέρουμε ότι το συγκεκριμένο τεστ προσπαθεί στην ουσία να μετρήσει την διαφορά απόστασης της υποτιθέμενης κανονικής κατανομής (αυτή που βρέθηκε κατά προσέγγιση από τα δεδομένα) και της εμπειρικής κατανομής των δεδομένων. Η εμπειρική κατανομή είναι στην ουσία μία αθροιστική συνάρτηση κατανομής που ακολουθεί μοτίβο συνάρτησης βημάτων ή συνάρτησης «σκάλα». Κάθε «σκαλοπάτι» έχει ύψος  $\frac{1}{n}$  όπου  $n$  ο αριθμός του δείγματος ενώ κάθε τιμή στον άξονα  $x$  αντιπροσωπεύει «πόσες παρατηρήσεις έχουν το πολύ ή ίση με την τιμή αυτή» ενώ στον άξονα  $y$  έχουμε την πιθανότητα να συμβεί αυτό. Ένα παράδειγμα δίνεται στην εικόνα 6.1.

## 6.2.2 Το τεστ των Jarque-Bera

Το συγκεκριμένο τεστ ελέγχει την κανονικότητα των δεδομένων μέσω της υπόθεσης ότι το δείγμα μας θα πρέπει να έχει λοξότητα και κυρτότητα κοινή με την κανονική κατανομή. Η ονομασία του δόθηκε από τους Carlos Jarque και Anil K. Bera [63] [64] [65]. Το στατιστικό αποτέλεσμα του τεστ είναι μη αρνητικό και ζητούνται τιμές κοντά στο μηδέν για την αποδοχή της υπόθεσης ότι το δείγμα μας αντλήθηκε από κανονική κατανομή. Θα χρειαστεί η προσέγγιση των τιμών  $S$  της λοξότητας και  $K$  της κυρτότητας. Οι τύποι δίνονται από τις Σχέσεις 6.5 και 6.6, όπως και το στατιστικό JB (Σχέση 6.7).

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{\frac{3}{2}}} \quad (6.5)$$



Σχήμα 6.1: Παράδειγμα εμπειρικής κατανομής.

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \quad (6.6)$$

$$JB = \frac{n}{6} \left( S^2 \frac{1}{4} (K - 3)^2 \right) \quad (6.7)$$

Οι τιμές  $\mu_3$  και  $\mu_4$  αποτελούν εκτιμητές των τρίτων και τέταρτων κεντρικών ροπών ενώ  $\bar{x}$  είναι ο μέσος όρος και  $\sigma$  η διακύμανση. Εάν το δείγμα μας ακολουθεί κανονική κατανομή, το στατιστικό JB θα ακολουθεί κατανομής  $x^2$  ασυμπτωτικά με δύο βαθμούς ελευθερίας, έτσι το στατιστικό JB μπορεί να χρησιμοποιηθεί για να ελέγξει εάν κάποιο δείγμα προέρχεται από κανονική κατανομή. Η μηδενική υπόθεσή μας στην ουσία είναι η σύμπραξη δύο γεγονότων, η μηδενική λοξότητα και η μηδενική υπερβατική κυρτότητα. Ως υπερβατική κυρτότητα ορίζουμε την κυρτότητα από την οποία έχουν αφαιρεθεί τρεις μονάδες, καθώς η κανονική κατανομή έχει κυρτότητα ίση με τρεις μονάδες.

### 6.2.3 Το τεστ του Lilliefors

Το συγκεκριμένο τεστ κανονικότητας βασίστηκε στο τεστ των Kolmogorov-Smirnov ενός δείγματος [66]. Τα βήματα του τεστ είναι τα ακόλουθα. Αρχικά γίνεται εκτίμηση του μέσου όρου και της διακύμανσης όπως είδαμε και στα προηγούμενα τεστ. Στην συνέχεια, δημιουργούμε το σχεδιάγραμμα της εμπειρικής συνάρτησης κατανομής και της κανονικής κατανομής με τις εκτιμηθήσες παραμέτρους και εντοπίζουμε την μεγαλύτερή τους διαφορά. Τέλος, αποφασίζουμε εάν η διαφορά αυτή είναι σημαντική ή όχι. Σε αυτό το σημείο διαφέρει το τεστ από αυτό των K-S καθώς η κανονική κατανομή που προέκυψε από τις εκτιμήσεις έχει μετακινηθεί πιο κοντά στα δεδομένα μας και η διαφορά ανάμεσά τους μειώθηκε. Ως διαφορά χρησιμοποιείται ο τύπος της συνάρτησης Supremum των απόλυτων τιμών, όπως φαίνεται στην Σχέση 6.8.

$$D_n = \sup_x |F_n(x) - F(x)| \quad (6.8)$$

### 6.2.4 Το τεστ των Shapiro-Wilk

Το συγκεκριμένο τεστ κανονικότητας δημιουργήθηκε το 1965 από τους Samuel Sanford Shapiro και Martin Wilk [67]. Το στατιστικό του συγκεκριμένου τεστ φαίνεται στην Σχέση 6.9.

$$W = \frac{(\sum_{i=1}^n a_i x(i))^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6.9)$$

όπου  $x_i$  οι παρατηρήσεις του δείγματος και  $\bar{x}$  ο μέσος όρος αυτών. Οι συντελεστές  $a$  δίνονται από την Σχέση 6.10.

$$(a_1, a_2, \dots, a_n) = \frac{m^T V^{-1}}{C} \quad (6.10)$$

ο  $C$  δίνεται από την Σχέση 6.11.

$$C = \|V^{-1}m\| = (m^T V^{-1} V^{-1} m)^{\frac{1}{2}} \quad (6.11)$$

με  $m = (m_1, m_2, \dots, m_n)^T$ . Ο πίνακας  $m$  απαρτίζεται στην ουσία από τους μέσους όρους των στατιστικών τάξεων των ανεξάρτητων και ομοίως κατανεμημένων τυχαίων μεταβλητών μίας τυπικής κανονικής κατανομής και  $V$  ο πίνακας συνδυακόμενης αυτών. Τέλος, δεν υπάρχουν στατιστικοί πίνακες για το στατιστικό  $W$  και οι κρίσιμες τιμές δίνονται από εξομοιώσεις Monte-Carlo.

### 6.3 Q-Q γράφημα και κανονικότητα

Πέρα από τα τεστ κανονικότητας, κάποιος θα μπορούσε να χρησιμοποιήσει το Q-Q γράφημα για να ελέγξει την κανονικότητα των δεδομένων [68]. Το συγκεκριμένο γράφημα αποτελεί ένδειξη όπως τα προηγούμενα τεστ αλλά συνάμα και έναν οπτικό έλεγχο κανονικότητας καθώς μπορούμε να παρατηρήσουμε εάν υπάρχει ταύτιση, προσέγγιση ή απόκλιση από την κανονική κατανομή. Γενικότερα, το συγκεκριμένο γράφημα μπορεί να χρησιμοποιηθεί έναντι οποιασδήποτε κατανομής, ωστόσο αυτό παραμένει αντικείμενο πέρα από τον σκοπό της παρούσας εργασίας καθώς τα δεδομένα μας δεν φαίνεται να ακολουθούν παρόμοιες κατανομές με απώτερο σκοπό την σύγκρισή τους. Αντίθετα ο έλεγχος έναντι της κανονικής κατανομής, εφόσον γίνει αποδοχή της κανονικότητας, μας επιτρέπει την χρήση μίας τεράστιας οικογένειας μεθόδων που είναι γνωστά ως παραμετρικά τεστ.

Στην ουσία το συγκεκριμένο γράφημα, αποτελείται από σημεία των οποίων η τετμημένη αντιστοιχεί σε εκατοστημόρια (percentiles) της κανονικής κατανομής ενώ η τεταγμένη αντιστοιχεί σε εκατοστημόρια του δείγματος που ελέγχουμε για κανονικότητα. Επιπλέον, οι τετμημένες αλλά και οι τεταγμένες έχουν τοποθετηθεί σε αύξουσα σειρά. Ιδεατά, θα θέλαμε όλα τα σημεία να βρίσκονται στην ευθεία  $y = x$  δηλαδή να έχουν κοινές τιμές μεταξύ τους. Ωστόσο, αυτό στην πράξη είναι αδύνατο οπότε και τιμές που δεν διαφέρουν αρκετά είναι αποδεκτές. Περιοχές με μεγάλες διαφορές αποτελούν κριτήριο απόρριψης της κανονικότητας του δείγματος που έχουμε για έλεγχο. Επιπροσθέτως, στην ειδική περίπτωση όπου τα σημεία βρίσκονται σε ευθεία παράλληλη της  $y = x$ , η κατανομή που ακολουθεί το δείγμα και η κανονική κατανομή έχουν γραμμική συσχέτιση. Καθαυτόν τον τρόπο παρατηρούμε ότι το Q-Q γράφημα μπορεί να αποτελέσει γραφικό μέσο εκτίμησης παραμέτρων σε οικογένειες κατανομών που ορίζονται από παραμέτρους θέσεων. Τέλος, για πληρότητα, ορίζουμε ως εκατοστημόρια κάποια σημεία του δείγματος τα οποία δρουν ως «σημεία αναφοράς» για να καταλάβουμε το ποσοστό των δεδομένων μας που έχουν μικρότερη τιμή από αυτή του εκατοστημόριου. Ως παράδειγμα ας πάρουμε την τυπική κανονική κατανομή (μέση τιμή 0 και διακύμανση 1), το 0.5-εκατοστημόριο αντιστοιχεί στην τιμή 0, καθώς το 50% του πληθυσμού έχει τιμές μικρότερες από την τιμή 0.

## 6.4 Μετατροπή των δεδομένων με κανονικοποίηση (z-score ή minmax-score)

Μία προσέγγιση που χρησιμοποιείται ως πιθανή λύση του προβλήματος της κανονικότητας των δεδομένων είναι η κανονικοποίηση. Στην παρούσα εργασία δοκιμάσαμε δύο κανονικοποιήσεις: την κανονικοποίηση του z-score και την minmax-score.

Στην κανονικοποίηση των τιμών μέσω της minmax-score φέρνει τις τιμές μας στο διάστημα  $[0, 1]$  [69]. Η ελάχιστη τιμή λαμβάνει την τιμή μηδέν ενώ η μέγιστη τιμή λαμβάνει την τιμή ένα. Οι υπόλοιπες τιμές λαμβάνουν δεκαδικές τιμές ανάμεσά τους. Ο μαθηματικός τύπος των νέων σημείων δίνεται από την Σχέση 6.12.

$$\frac{X_i - \min}{\max - \min} \quad (6.12)$$

Το μειονέκτημα αυτής της μεθόδου είναι ότι δεν μπορεί να συμπεριφερθεί σωστά σε ακραίες τιμές. Ως παράδειγμα, αναφέρουμε ένα αρχικό δείγμα 100 τιμών, με 98 τιμές στο διάστημα  $0 - 50$ , την τιμή 95 και την τιμή 100. Το τελικό αποτέλεσμα θα έχει πάλι 98 τιμές σε ένα διάστημα και δύο ακραίες ενώ η λοξότητα της κατανομής δεν έχει επιλυθεί στην βάση της.

Στην κανονικοποίηση των τιμών μέσω της z-score φέρνει πιο κοντά τις τιμές μας στην μέση τιμή ενώ δεν επηρεάζεται από ακραίες τιμές όσο η προηγούμενη κανονικοποίηση [70]. Ο τύπος της μετατροπής κάθε τιμής  $X_i$  του αρχικού μας συνόλου δίνεται από την Σχέση 6.13.

$$Z_i = \frac{Z_i - \bar{X}}{S} \quad (6.13)$$

όπου  $\bar{X}$  η μέση τιμή και  $S$  η τυπική απόκλιση της κανονικής κατανομής που υποθέτουμε ότι προσεγγίζει τα δεδομένα. Ενώ στην ουσία, το σκορ αυτό μας λέει πόσες φορές απομακρυνθήκαμε σε μονάδες τυπικής απόκλισης από την μέση τιμή, ενώ όταν πρόσημο του σκορ είναι θετικό δηλώνει ότι η τιμή μας είναι μεγαλύτερη από την μέση τιμή. Αντίθετα, εάν το πρόσημο του σκορ είναι αρνητικό, δηλώνει ότι η τιμή μας είναι μικρότερη από την μέση τιμή.

Τέλος, αξίζει να σημειωθεί μία άλλη προσέγγιση που χρησιμοποιείται για την λύση του προβλήματος της κανονικότητας. Αυτή η προσέγγιση αφορά την μετατροπή των δεδομένων μας μέσω ενός μαθηματικού μετασχηματισμού. Οι πιο συχνοί είναι ο λογάριθμος με βάση το δέκα ή το δύο, η ύψωση σε δύναμη και η εκθετοποίηση ενώ πιο περίπλοκοι τύποι που εμπλέκουν τόξα ημιτόνων έχουν προταθεί για κατανομές με αρνητική λοξότητα. Η προσέγγιση αυτή κρύβει έναν μεγάλο κίνδυνο - την έλλειψη

ερμηνευσιμότητας. Κατά επέπταση αυτού του γεγονότος, η ποσοτικοποίηση των μέτρων μπορεί να μην παραμένει το ίδιο ανάλογη ή δυσανάλογη στα νέα δεδομένα. Ωστόσο, δεν αποτελεί μονόδρομο και για αυτόν τον λόγο θα ασχοληθούμε με μη παραμετρικά τεστ.

## 6.5 Μη Παραμετρικό Τεστ των Kolmogorov-Smirnov Δύο Δειγμάτων

Σε αυτήν την παράγραφο θα ασχοληθούμε με το μη παραμετρικό τεστ των Kolmogorov και Smirnov που αφορά δύο δείγματα. Αυτό που προσπαθεί να απαντήσει το συγκεκριμένο τεστ είναι το γεγονός εάν τα δύο δείγματα έρχονται από κοινή κατανομή, ανεξάρτητα εάν η κοινή κατανομή τους είναι η κανονική, η εκθετική ή κάποια άλλη. Καταλαβαίνουμε ότι το τεστ αυτό αποτελεί ένα πολύ ισχυρό εργαλείο καθώς δεν έχει λάβει καμία υπόθεση.

Αρχικά, δημιουργούμε τις  $F_{1,n}$   $F_{2,m}$  που αντιστοιχούν στις εμπειρικές κατανομές των δειγμάτων 1 και 2 που έχουν  $n$  και  $m$  αριθμό παρατηρήσεων. Το στατιστικό του τεστ δίνεται από την Σχέση 6.14.

$$D_{n,m} = \sup |F_{1,n}(x) - F_{2,m}(x)| \quad (6.14)$$

όπου  $\sup$  η συνάρτηση Supremum. Ως Supremum ενός συνόλου  $S$  ορίζουμε το ελάχιστο άνω όριο ενός συνόλου, αντίστοιχα η συνάρτηση Supremum μας δίνει το ελάχιστο άνω φράγμα της συνάρτησης που ελέγχεται.

Για την απόρριψη της  $H_0$  με επίπεδο σημαντικότητας  $\alpha$  θα πρέπει να ισχύει η Σχέση 6.15.

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{nm}} \quad (6.15)$$

Αυτό που προσπαθεί να επιτύχει το συγκεκριμένο τεστ είναι να βγάλει ένα κανόνα απόφασης με βάση την μέγιστη απόσταση των δύο εμπειρικών κατανομών. Είναι προφανές ότι ακραίες τιμές είτε στο ένα δείγμα είτε στο άλλο θα επηρεάσουν το αποτέλεσμα μας καθώς θα δημιουργήσουν μία επιπλέον απόσταση ανάμεσα στις τιμές των εμπειρικών κατανομών.

## 6.6 Στατιστικές ροπές

Καθώς είναι λιγότερο πιθανό, τα δεδομένα μας να διέπονται από κοινές κατανομές, θα πρέπει να λάβουμε υπόψιν μας διαφορετικά μέτρα, τόσο για την σύγκρισή τους όσο και για την ποσοτικοποίησή τους. Η συχνότερη χρήση για την μορφολογική περιγραφή των στατιστικών κατανομών είναι οι στατιστικές ροπές. Στην πραγματικότητα, πρόκειται για μέτρα που ορίζουν μία κατανομή και της δίνουν σχήμα και μορφή. Για παράδειγμα μία κανονική κατανομή μπορεί να περιγραφεί μορφολογικά μέσω της πρώτης και της δεύτερης στατιστικής ροπής καθώς οι επόμενες δύο (τρίτη και τέταρτη) είναι σταθερές. Στο σημείο αυτό αξίζει να ασχοληθούμε περισσότερο με τις τέσσερις βασικές στατιστικές ροπές. Επιπλέον, να αναφέρουμε ότι στατιστικές ροπές μεγαλύτερου βαθμού θα πρέπει να χρησιμοποιούνται με μεγάλη προσοχή ή καθόλου λόγω της περιορισμένης δύναμής τους ως προς τα δεδομένα. Επιπλέον, είναι πολύ δύσκολο να ερμηνευτούν και απαιτείται η χρήση μικρότερων τάξεων ροπών για την ερμηνεία αυτή. Για παράδειγμα, η τέταρτη στατιστική ροπή μπορεί να ερμηνευθεί ως «η σχετική σημασία των ουρών έναντι των ώμων δεδομένης μίας διασποράς». Καταλήγουμε ότι μία υψηλή τέταρτη στατιστική ροπή έχει παχιές ουρές ενώ μία χαμηλή έχει ευρείς ώμους. Η πέμπτη όμως στατιστική ροπή μπορεί να ερμηνευθεί ως «η σχετική σημασία των ουρών έναντι της επικρατούσας τιμής με βάση την λοξότητα». Καταλήγουμε ότι μία υψηλή πέμπτη στατιστική ροπή έχει παχιές ουρές και μικρή μετακίνηση της επικρατούσας τιμής ενώ μία χαμηλή πέμπτη στατιστική ροπή μας δίνει την αίσθηση μεγάλης αλλαγής στους ώμους. Καταστείται, λοιπόν, σαφές ότι στατιστικές ροπές πέμπτου βαθμού και άνω γίνονται πολύ δύσκολα ερμηνεύσιμες ενώ για το εξαγώμενό τους απαιτείται συνδυασμός των παραμέτρων ή προηγούμενων στατιστικών ροπών. Για πληρότητα, επικρατούσα τιμή ενός δείγματος, είναι η τιμή που εμφανίζεται πιο συχνά στο δείγμα.

Στον πίνακα 6.1 μπορούμε να δούμε τους τύπους των τεσσάρων πρώτων στατιστικών ροπών. Η πρώτη στατιστική ροπή είναι η μέση τιμή και αποτελεί ένα μέτρο θέσης. Συχνά τα αριθμητικά δεδομένα παρουσιάζουν μία ιδιαίζουσα ιδιότητα να συγκεντρώνονται γύρω από κάποια τιμή, την λεγόμενη κεντρική τιμή. Την ιδιότητα αυτήν την ονομάζουμε κεντρική τάση. Η μέση τιμή μας δίνει την θέση της κεντρικής τιμής σε κάποιες περιπτώσεις. Σε αντίθετη περίπτωση, η μέση τιμή μας δίνει την αίσθηση της μέσης κατάστασης όταν δεν υπάρχουν μεγάλες αναταράξεις λόγω ακραίων τιμών.

Η δεύτερη στατιστική κεντρική ροπή είναι η διακύμανση και αποτελεί ένα μέτρο διασποράς. Στην ουσία πρόκειται για την αναμενόμενη τιμή των τετραγωνικών αποκλίσεων από την μέση τιμή. Η συγκεκριμένη ροπή μετρά πόσο μακριά από την μέση τιμή απλώνονται οι τιμές και έχει μία εξέχουσα θέση στην Στατιστική καθώς συντελεί επίκεντρο της περιγραφικής στατιστικής, του υποθετικού ελέγχου και της καλής προσαρμογής. Είναι πολύ σημαντικό εργαλείο διότι χρησιμοποιείται ενδελεχώς σε στατιστικές αναλύσεις.

Η τρίτη στατιστική κανονικοποιημένη ροπή είναι η λοξότητα. Η λοξότητα αποτελεί μέτρο ασυμμετρίας μίας κατανομής πιθανότητας μίας πραγματικής τυχαίας μεταβλητής ως προς την μέση τιμή. Η τιμή της λοξότητας μπορεί να είναι μηδενική, θετική, αρνητική ή αόριστη. Σε κατανομές με μία κορυφή, η αρνητική λοξότητα σημαίνει ότι η ουρά της κατανομής έγκειται στην αριστερή πλευρά της κατανομής ενώ το αντίθετο συμβαίνει στην θετική λοξότητα. Η μηδενική τιμή στο μέτρο της λοξότητας αντιστοιχεί στην πλήρη συμμετρία ως προς την μέση τιμή. Ωστόσο, υπάρχουν περιπτώσεις όπου ενώ έχουμε μηδενική τιμή στην λοξότητα, η κατανομή δεν ακολουθεί συμμετρική μορφολογία όπως για παράδειγμα η μία ουρά μίας κατανομής να είναι ψηλή και λεπτή ενώ η άλλη να είναι κοντή αλλά βαριά. Σε αυτό το σημείο είναι αρκετά σημαντικό να υπογραμμίσουμε κάποιες επιπλέον σχέσεις μεταξύ μέσης τιμής και ενδιάμεσου που αφορούν την παλαιότερο ορισμό της λοξότητας. Μέχρι το τέλος της παραγράφου, λέγοντας λοξότητα θα αναφερόμαστε στην παλαιά έκδοσή της, γνωστή ως παραδοσιακή λοξότητα και δίνεται από τον τύπο  $\frac{\mu-\nu}{\sigma}$  όπου  $\mu$  ο μέσος όρος,  $\nu$  ο ενδιάμεσος και  $\sigma$  η τυπική απόκλιση. Στις κατανομές με θετική λοξότητα, η επικρατούσα τιμή είναι μικρότερη από την ενδιάμεσο ενώ η ενδιάμεσος είναι μικρότερη από την μέση τιμή. Στις κατανομές με αρνητική λοξότητα, η μέση τιμή είναι μικρότερη από την ενδιάμεσο ενώ η ενδιάμεσος είναι μικρότερη από την επικρατούσα τιμή. Στις συμμετρικές κατανομές η ενδιάμεσος και η μέση τιμή ταυτίζονται. Στην ειδική περίπτωση όπου έχουμε μία κορυφή στην κατανομή, η επικρατούσα τιμή έχει την ίδια τιμή με την ενδιάμεσο και την μέση τιμή [71]. Η νέα εκδοχή της λοξότητας ως Τρίτη στατιστική ροπή και η παραδοσιακή της εκδοχή δεν έχουν πάντα κοινό πρόσημο, ενώ συμφωνούν για πολλές οικογένειες κατανομών.

Η τέταρτη στατιστική κανονικοποιημένη ροπή είναι η κύρτωση. Η κύρτωση είναι ένα μέτρο που ασχολείται με τις ουρές της κατανομής πιθανότητας μίας πραγματικής τυχαίας μεταβλητής. Όπως και η λοξότητα, η κύρτωση ορίζει την μορφολογία της κατανομής. Οι πρώτοι ορισμοί της κύρτωσης δόθηκαν από τον Karl Pearson, ο



οποίος ασχολήθηκε και με την λοξότητα [72]. Η κύρτωση της κανονικής κατανομής είναι ίση με τρία. Είναι συχνό φαινόμενο να συγκρίνουμε την κύρτωση με βάση την τιμή αυτή. Κατανομές με κύρτωση μικρότερη του τρία ονομάζονται πλατύκυρτες. Οι πλατύκυρτες κατανομές έχουν λίγα και σχετικά όχι πολύ μεγάλες ακραίες τιμές σε σχέση με την κανονική κατανομή. Η ομοιόμορφη κατανομή είναι ένα παράδειγμα μίας τέτοιας κατανομής, καθώς δεν έχει σχεδόν καθόλου ακραίες τιμές και όλες είναι «ομοιόμορφα» κατανομημένες. Κατανομές με κύρτωση μεγαλύτερη του τρία ονομάζονται λεπτόκυρτες. Ένα παράδειγμα μίας τέτοιας κατανομής είναι η κατανομή του Laplace. Η συγκεκριμένη κατανομή οι ουρές προσεγγίζουν ασυμπτωτικά το άπειρο πιο αργά σε σχέση με την κανονική κατανομή. Απόρροια αυτής της συμπεριφοράς είναι η εμφάνιση περισσότερων ακραίων τιμών σε σχέση με την κανονική κατανομή. Επιπλέον, αξίζει να σημειωθεί ότι σύμφωνα με τον Pearson, δόθηκε ο ορισμός της υπερβατικής κύρτωσης όπου αφορά την κύρτωση αφαιρώντας τρεις μονάδες με σκοπό να αποτελέσει βάση της σύγκρισης η κανονική κατανομή.

Όπως παρατηρούμε και στον πίνακα, η μέση τιμή, η διακύμανση, η λοξότητα και η κύρτωση έχουν κλειστούς τύπους τόσο για κατανομές όσο και για τυχαία δείγματα όπου απαιτείται η εκτίμηση. Ο όρος «κεντρική» σημαίνει ότι ενέχει τον ορισμό του μέσου όρου ενώ «κανονικοποιημένη» ενέχει τόσο τον μέσο όρο όσο και την κανονικοποίηση από την τυπική απόκλιση. Η ιδέα των στατιστικών ροπών αλλά και των εκτιμητών τους άρχισε από τον Pearson ενώ ορίστηκε πλήρως από τον Pafnuty Chebyshev το 1887 και χρησιμοποιήθηκαν για την απόδειξη του Κεντρικού Οριακού Θεωρήματος [73].

<i>Name Of Moment</i>	<i>Type</i>	<i>Formula</i>	<i>Estimator</i>
Mean	Raw	$\mu = \frac{1}{N} \sum_{i=1}^N A_i$	$\bar{x} = \frac{1}{N} \sum_{i=1}^N X_i$
Variance	Central	$V = \frac{1}{N-1} \sum_{i=1}^N  A_i - \mu ^2$	$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N  X_i - \mu ^2$
Skewness	Standarized	$s = \frac{E(x-\mu)^3}{\sigma^3}$	$s_0 = \frac{\sqrt{n(n-1)}}{n-2} s$
Kurtosis	Standarized	$k = \frac{E(x-\mu)^4}{\sigma^4}$	$k_0 = \frac{n-1}{(n-2)(n-3)} *$ $*((n+1)k - 3(n-1)) + 3$

Πίνακας 6.1: Πίνακας Μαθηματικών Τύπων των Στατιστικών Ροπών.

## 6.7 Το μη παραμετρικό τεστ των Mann - Whitney U Δύο Πληθυσμών

Το συγκεκριμένο τεστ είναι γνωστό ως Mann-Whitney-Wilcoxon ή Wilcoxon rank-sum test. Το τεστ των Mann και Whitney προσπαθεί να συγκρίνει δύο πληθυσμούς χωρίς κατά ανάγκη οι τιμές τους να έχουν αντληθεί από κανονικές κατανομές. Ωστόσο, οι κατανομές θα πρέπει να ομοιάζουν μορφολογικά. Η σύγκριση του τεστ βασίζεται στον έλεγχο της θέσης των δύο ενδιάμεσων. Η αναλογία του τεστ στα παραμετρικά τεστ είναι το t τεστ. Καταλήγουμε ότι η μηδενική μας υπόθεση είναι η ομοιότητα των δύο πληθυσμών ή η τοποθεσία των ενδιάμεσων των δύο πληθυσμών είναι κοινή. Επιπροσθέτως, απορρίπτοντας την μηδενική υπόθεση σημαίνει ότι ένας πληθυσμός τείνει να έχει μικρότερες ή μεγαλύτερες τιμές από τον άλλο στον δίπλευρο έλεγχο. Σε μονόπλευρο έλεγχο θα μπορούσαμε να εξετάσουμε ποιον πληθυσμό πιστεύουμε ότι έχει μικρότερες τιμές από τον άλλο (ή μεγαλύτερες).

Αρχικά θα τοποθετήσουμε τις παρατηρήσεις των δύο πληθυσμών σε ένα σύνολο ενώ στην συνέχεια θα τις βάλουμε έναν βαθμό για κάθε παρατήρηση του συνόλου που δημιουργήσαμε, που θα αντιπροσωπεύει την θέση του εάν όλες οι παρατηρήσεις είχαν διαταχθεί σε αύξουσα σειρά. Για τιμές που είναι κοινές, ο βαθμός θα είναι ο μέσος όρος των δύο βαθμών, π.χ. για τους βαθμούς (1,2,3,4,5), εάν ο βαθμός 3,4 είχε κοινή τιμή, θα γινόταν (1,2,3.5,3.5,5). Τέλος, προσθέτουμε όλους τους βαθμούς των παρατηρήσεων του πρώτου πληθυσμού και του δεύτερου πληθυσμού. Χρησιμοποιώντας το U στατιστικό και την κατανομή των πληθυσμών, βρίσκουμε το p-value που συγκρίνεται με το διάστημα εμπιστοσύνης  $\alpha$ . Για πληρότητα, η κατανομή των πληθυσμών έχει μέση τιμή και τυπική απόκλιση δίνονται από τις Σχέσεις 6.16 και 6.17.

$$\mu_w = \frac{n_1(n_1 + n_2 + 1)}{2} \quad (6.16)$$

$$\sigma_w = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \quad (6.17)$$

Το στατιστικό U δίνεται από τον τύπο  $U = \frac{W - \mu_w}{\sigma_w}$ , όπου W ο αθροιστικός βαθμός του πληθυσμού υπό έλεγχο ενώ η τιμή p-value βρίσκεται από την προσέγγιση της κανονικής κατανομής και το στατιστικό U. Τέλος, ελέγχουμε εάν το p-value είναι μικρότερο από ένα επίπεδο σημαντικότητας  $\alpha$ . Εάν όντως συμβαίνει αυτό, απορρίπτουμε την μηδενική υπόθεση ότι είναι ίδιες οι κατανομές στον δίπλευρο έλεγχο ή ότι είναι μία μεγαλύτερη της άλλης στο μονόπλευρο.

## ΚΕΦΑΛΑΙΟ 7

# ΚΕΦΑΛΑΙΟ: ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

- 
- 7.1 Μοντελοποίηση της ανάπτυξης της Χημικής Καρκινογένεσης
  - 7.2 Ταξινόμηση
  - 7.3 Στατιστική Προσέγγιση
  - 7.4 Μη αποδοτικά μοντέλα και διαδικασίες
- 

Σε αυτό το κεφάλαιο επισημαίνονται τα αποτελέσματα της έρευνας και εξάγονται κάποια εμπειριστατωμένα συμπεράσματα. Το κεφάλαιο ξεκινάει με την μαθηματική μοντελοποίηση της ανάπτυξης της χημικής καρκινογένεσης. Στην συνέχεια γίνεται ενασχόληση με την απόδοση των ταξινομητών που έχουν ως χαρακτηριστικά τις αυθαίρετες σταθερές των μαθηματικών αναπτυξιακών μοντέλων όπως προέκυψαν από την μέθοδο των ελαχίστων τετραγώνων και η προσοχή εστιάζεται στην ικανότητα τους να ξεχωρίζουν διαφορετικές οικογένειες των δραστικών ουσιών. Τέλος, παρουσιάζονται τα αποτελέσματα της στατιστικής προσέγγισης των καταλήξεων των πειραματοζώων και παρατίθενται τα συμπεράσματα που αντλήσαμε βάση της προσέγγισης αυτής. Ακόμη, αναφέρονται τα μη αποδοτικά μοντέλα και διαδικασίες με αποδείξεις.

## 7.1 Μοντελοποίηση της ανάπτυξης της Χημικής Καρκινογένεσης

Αρχικά θα ασχοληθούμε με τις παραμέτρους των αναπτυξιακών μοντέλων όπως αυτές προέκυψαν από την μέθοδο των ελαχίστων τετραγώνων. Πιο συγκεκριμένα, θα μελετήσουμε την μεταβλητότητά τους και για τον λόγο αυτό θα χρησιμοποιήσουμε τον μέσο όρο, την τυπική απόκλιση και τον συντελεστή μεταβλητότητας (Variation Coeff.). Ο συντελεστής μεταβλητότητας (CV) μας δείχνει το μέγεθος την μεταβλητότητας σε σχέση με τον μέσο όρο ή τυπικό πρότυπο του πληθυσμού και εκφράζεται συνήθως επί τοις εκατό. Ο μαθηματικός τύπος του συντελεστή μεταβλητότητας είναι  $CV = \frac{\sigma}{\mu}$  ενώ όταν οι τιμές που λαμβάνει, δεν ξεπερνάν το τριάντα τοις εκατό (30%) θεωρούνται ικανές να αποδώσουν ομοιογενές δείγματα ενώ πολύ μεγάλες τιμές υποδηλώνουν την μη ύπαρξη κάποιας συγκεκριμένης τάσης ή αλλιώς αυθαιρέτητα των τιμών. Στον πίνακα 7.1 παρατηρούμε ότι σε κάθε αναπτυξιακό μοντέλο εμπεριέχονται τιμές με πολύ υψηλούς δείκτες CV.

Μπορούμε πλέον να συμπεράνουμε δύο εύλογες προτάσεις. Πρώτον, η αυθαιρέτητα της παραμέτρου υποδηλώνει ότι το μοντέλο μπορεί να αιχμαλωτίσει οποιαδήποτε ιδιομορφία στις καταγραφές των καρκινικών βαρών και για αυτόν τον λόγο θα πρέπει να είμαστε πολύ προσεκτικοί στην επιλογή των αναπτυξιακών μοντέλων. Προτείνουμε η επιλογή να ξεκινάει από την ύπαρξη φυσιολογικής βάσης και στην συνέχεια να ελέγχεται η καλή προσαρμογή των δεδομένων. Με αυτόν τον τρόπο, μπορούμε να εξηγήσουμε ικανοποιητικά οποιοδήποτε αποτέλεσμα αντλήσουμε. Δεύτερον, βλέπουμε ότι υπάρχει ένα διάστημα που εντοπίζονται κάποιες από τις παραμέτρους που σημαίνει ότι θα μπορούσαμε να μετατρέψουμε τα μοντέλα τριών παραμέτρων σε μοντέλα δύο παραμέτρων, έχοντας δώσει ένα εύρος τιμών σε κάποια από τις παραμέτρους. Τα διαστήματα που προτείνουμε απαρτίζονται στον πίνακα 7.2 ενώ συνάμα παρατηρούμε ότι το μοντέλο Gaussian δεν έχει τέτοια διαστήματα και η μεταβλητότητα των παραμέτρων του δεν περιορίζεται επαρκώς με σκοπό την πρόβλεψή τους. Ο μη περιορισμός των παραμέτρων συμβαίνει πιθανότατα λόγω της μεγάλης αλλαγής στην μορφολογία της καμπύλης βάση του αρχικού όγκου  $V_0$ . Η συγκεκριμένη τιμή ορίζει το ύψος της Γκαουσιανής καμπύλης και επηρεάζει τις υπόλοιπες παραμέτρους.

Έτσι, η προσέγγισή μας μπορεί να αποτελέσει βάση για μελλοντικές έρευνες σε καρκινικές κυτταρικές αναπτύξεις και να εξοικονομήσει σημαντικό χρόνο ο επιστήμονας για την ενασχόλησή του σε πιο μεταβαλλόμενες παραμέτρους. Ακόμη, δεν

αναφύεται η περίπτωση εγκλωβισμού σε τοπικά ελάχιστα και μειώνεται το υπολογιστικό κόστος. Επιπροσθέτως, υπογραμμίζουμε ότι ο συντελεστής μεταβλητότητας απαιτεί ομόσημες παρατηρήσεις για να έχει ισχύ. Για τον λόγο αυτό, αφαιρέσαμε τα διανύσματα παραμέτρων που εμπειριείχαν τουλάχιστον μία αρνητική παράμετρο. Δεκαέξι (από τις 124) περιπτώσεις στο λογιστικό μοντέλο και πενήντα περιπτώσεις για το πολυωνυμικό. Ακόμη, δεν εντοπίζουμε σημαντικές διαφορές εάν η ίδια ανάλυση χρησιμοποιηθεί στις οικογένειες των δραστικών ουσιών ξεχωριστά έναντι όλου του πληθυσμού. Για πληρότητα τα αποτελέσματα αυτά δίνονται στους πίνακες A.1, A.2 και A.3 στο Παραρτήμα Αποτελεσμάτων και αφορούν την οικογένεια του βενζοπυρενίου, των πολυαμινών και των θειολών αντίστοιχα.

Στους πίνακες 7.3, 7.4 και 7.5 έχουμε παραθέσει την σύγκριση των αναπτυξιακών μοντέλων στην οικογένεια χωρίς καμία δραστική ουσία δηλαδή μόνο με λήψη βενζοπυρενίου. Διευκρινίζεται ότι η τιμή σε κάθε κελί στην πρώτη γραμμή του πίνακα αντιστοιχεί στον αριθμό των προτύπων που έχουν την βέλτιστη τιμή ως προς την εκάστοτε μετρική στο σύνολο των μοντέλων. Επομένως, κάθε πρότυπο εμφανίζεται μόνο μία φορά στην πρώτη γραμμή και συγκεκριμένα στο μοντέλο που η εκάστοτε μετρική είναι η ελάχιστη (καθώς όλες οι μετρικές αναφέρονται σε αποστάσεις). Ωστόσο, στην δεύτερη και μετά γραμμή του πίνακα, επιτρέπουμε μία ανέχεια εκφρασμένη επί τοις εκατό. Έτσι, εάν ένα πρότυπο διαφέρει στην τιμή της μετρικής 5% από την ελάχιστη δυνατή, θεωρείται επιτυχία και προσμετράται. Καθ'αυτόν τον τρόπο ένα πρότυπο μπορεί να ανήκει σε περισσότερα από ένα μοντέλα από την δεύτερη και μετά σειρά του πίνακα. Ως παράδειγμα, μπορούμε να σκεφτούμε ένα πρότυπο που έχει την τιμή 100 ως ελάχιστη τιμή στην μετρική RMSE στο μοντέλο Logistic ενώ λαμβάνει τιμή 103 και 102 στα μοντέλα Gaussian και Exponential αντίστοιχα και έχει τιμές μεγαλύτερες από 106 στα υπόλοιπα. Τότε ο αριθμός της Logistic, Gaussian και Exponential θα αυξηθούν κατά μία μονάδα στα αντίστοιχα κελιά που αφορούν την ανέχεια 5%. Ο λόγος που χρησιμοποιήσαμε αυτήν την παραλλαγή των μετρικών και συμπεριλάβαμε την ανέχεια είναι διότι δεν θέλουμε ένα μοντέλο να είναι πάρα πολύ αποδοτικό σε μία μεγάλη πλειονότητα των προτύπων και μη αποδοτικό στα υπόλοιπα αλλά ψάχνουμε το μοντέλο εκείνο που μπορεί να περιγράψει αποδοτικά το οποιοδήποτε πρότυπο ακόμη και εάν σε κάποιες περιπτώσεις δεν είναι το βέλτιστο. Τέλος, επενθυμίζουμε ότι η μετρική SSE είναι η αντικειμενική συνάρτηση της μεθόδου των ελαχίστων τετραγώνων ενώ οι μετρικές RMSE και MAE αποτελούν τους επιπρόσθετους ελέγχους. Παρατηρώντας το

σύνολο αυτών των πινάκων, αντιλαμβανόμαστε την τεράστια ισχύ που έχει το λογιστικό μοντέλο έναντι των υπολοίπων. Υπογραμμίζουμε ότι το λογιστικό μοντέλο σε όλες τις μετρικές κατορθώνει να αιχμαλωτίσει σχεδόν όλα τα πρότυπα με ανέχεια τάξεως 20%. Καλή απόδοση παρουσιάζουν τα μοντέλα Gompertz και Gaussian. Το μοντέλο Bertalanffy παρουσιάζει ικανοποιητική απόδοση, με έμφαση στις θειόλες ενώ τα υπόλοιπα μοντέλα έχουν κακή συμπεριφορά είτε στην ανέχεια τάξεως 0% είτε κοντά της 30%. Καταλήγουμε ότι το λογιστικό μοντέλο περιγράφει καλύτερα το σύνολο των δεδομένων μας, ενώ κρατάμε και την καλή απόδοση των Gompertz και Gaussian. Ο πίνακας 7.3 αντιστοιχεί στην μετρική SSE, ο πίνακας 7.4 στην μετρική RMSE και ο πίνακας 7.5 στην μετρική MAE. Οι αντίστοιχοι πίνακες της οικογένειας των Πολυαμινών και των Θειολών έχουν παρατεθεί στο παράρτημα καθώς δεν ενέχουν αξιοσημείωτες αλλαγές.

Όσο αφορά την προβλεψιμότητα των αναπτυξιακών μοντέλων, επιλέξαμε τα πρότυπα με περισσότερες από δεκαπέντε καταγραφές καρκινικού βάρους. Σε αυτά τα πρότυπα αποκρύψαμε τις τελευταίες τέσσερις καταγραφές. Υλοποιήσαμε την προσαρμογή στο εκάστοτε αναπτυξιακό μοντέλο χωρίς τις τελευταίες τέσσερις καταγραφές και υπολογίσαμε την διαφορά της τιμής που δίνει το μοντέλο με βάση τις παραμέτρους της προσαρμογής από την πραγματική τιμή. Η συγκεκριμένη τιμή μπορεί να εκληφθεί ως μέτρο αλλαγής του μοντέλου. Όταν το μοντέλο σε νέες καταγραφές προσαρμόζεται αλλάζοντας αρκετά την μορφολογία του δεν έχει ικανοποιητική ικανότητα πρόβλεψης. Ιδανικά θα θέλαμε να μην αλλάζει αρκετά μορφολογικά ενώ συνάμα να προσαρμόζεται στις νέες καταγραφές. Στην πραγματικότητα ένα τέτοιο γεγονός τείνει αδύνατο λόγω της μεγάλης ευελιξίας των μοντέλων, της τεράστιας δηλαδή προσαρμοστικότητάς τους σε μικρές αλλαγές όπως είναι ακόμη και μία νέα καταγραφή. Παραθέτουμε τις τέσσερις διαφορές ανάμεσα στις τιμές που προέβλεψε το αναπτυξιακό μοντέλο και τις πραγματικές τιμές του καρκινικού βάρους στον πίνακα 7.6 ενώ σημειώνουμε ότι χρησιμοποιήθηκαν τα τρία μοντέλα που προέκυψαν ως πιο αποδοτικά με βάση τις μετρικές SSE, RMSE και MAE. Τα υπόλοιπα δεν κρίνονται ικανά να περιγράψουν τα δεδομένα μας σε έναν επαρκή βαθμό οπότε δεν υπάρχει νόημα να μελετήσουμε την προβλεπτική τους ικανότητα. Στον πίνακα 7.6 παρατηρούμε ότι όσο μεγαλώνει ο χρόνος, η διαφορά στο καρκινικό βάρος μεγαλώνει με τρόπο που δεν μπορούμε να προβλέψουμε ικανοποιητικά. Ενώ στην αρχή βρισκόμαστε σε λογικά πλαίσια ( της τάξεως του 15%) όσο αφορά τις προβλέψεις, στην συνέχεια ξεφεύγουμε αρκετά από αυτά. Το γεγονός αυτό οφείλε-

ται προτίστως στην ευελιξία των μοντέλων, την ικανότητάς τους να αιχμαλωτίζουν οποιαδήποτε αλλαγή στις τιμές της ανάπτυξης των καρκινικών κυττάρων και δευτερευόντως στις λίγες καταγραφές στην εκθετική φάση της νόσου κατά την διεξαγωγή των πειραμάτων. Τονίζουμε ότι οι καταγραφές θα πρέπει να είναι αρκετά συχνότερες σε περιόδους έξαρσης του καρκινικού ρυθμού ανάπτυξης καθώς η εκθετικότητα διακρίνει αυτήν την περίοδο. Πολύ πιθανό να χρειάζονται περισσότερες από πέντε καταγραφές ανά ημέρα στην περίοδο της εκθετικότητας της νόσου ενώ και μία καταγραφή ανά ημέρα πριν από αυτήν την περίοδο θα είναι αντιπροσωπευτική. Ιδανικά θα θέλαμε κάθε καταγραφή να μην διαφέρει περισσότερο από 5% με 10% από την προηγούμενη.

<i>Growth Model</i>	<i>Parameter</i>	<i>Mean ± St.Deviation</i>	<i>Variation Coeff.</i>
Exponential	Vo	265,9979 ± 306,3857	115,1%
	$\lambda$	0,0495 ± 0,0837	169,2%
Gompertz	Vo	34,7636 ± 79,4489	228,5%
	A	0,3009 ± 0,2809	93,3%
	B	0,0204 ± 0,0098	71,8%
Logistic	A	0,0658 ± 0,0473	71,8%
	B	1,161e-05 ± 9,7972e-06	84,3%
	Vo	51,7463 ± 80,5408	155,6%
Bertalanffy	$\beta_1$	2398,9 ± 13971	582,3%
	$\beta_2$	1,3569 ± 0,3655	26,9%
	$\beta_3$	0,0083 ± 0,0077	93,4%
Gaussian	a	3,9035e-86 ± 4,3467e-87	1113,6%
	b	366,5738 ± 1,3640e-03	372,0%
	c	80,1319 ± 111,4567	139,0%
Powerlaw	a	1,4280 ± 7,3663	515,8%
	b	5,6291 ± 9,5748	170,0%
Polynomial of 3rd Deg	$p_1$	0,0153 ± 0,0184	120,1%
	$p_2$	1,9384 ± 1,6870	87,0%
	$p_3$	153,7386 ± 169,1828	110,0%
	$p_4$	1,7306e-03 ± 1,4499e-03	83,7%

Πίνακας 7.1: Μεταβλητότητα Παραμέτρων των Αναπτυξιακών Μοντέλων σε όλες τις δραστητικές ουσίες.



<i>Growth Model</i>	<i>Parameter</i>	<i>Interval</i>
Exponential	$\lambda$	[0, 0.5169]
Gompertz	C	[0, 0.0432]
Logistic	A	[0, 0.3412]
Bertalanffy	$\beta_2$	[0.9877, 2.2907]
Gaussian	-	-
Powerlaw	a	[0, 2.9033]
Polynomial of 3rd Deg	$p_1$	[0, 0.0841]

Πίνακας 7.2: Διαστήματα Παραμέτρων των Αναπτυξιακών Μοντέλων σε όλες τις δραστικές ουσίες.

<i>Growth Model</i>	<i>Toleranced SSE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	0	0	0	0	0	1	1	1	1
Gompertz	2	3	5	7	9	12	13	13	18
Logistic	30	31	31	31	33	34	38	39	39
Bertalanffy	4	5	6	9	11	12	13	15	15
Gaussian	0	0	1	2	2	2	2	3	3
Powerlaw	0	1	1	2	2	4	5	5	6
Polynomial 3rd Deg.	13	15	17	17	18	18	18	19	19

Πίνακας 7.3: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced SSE στο Βενζοπυρένιο.

<i>Growth Model</i>	<i>Toleranced RMSE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	0	0	1	1	2	2	3	4	4
Gompertz	3	7	14	19	21	23	24	24	24
Logistic	29	33	35	39	40	41	44	46	47
Bertalanffy	11	13	14	17	17	17	18	18	19
Gaussian	3	5	8	9	15	20	20	21	24
Powerlaw	3	5	6	7	10	12	15	16	18
Polynomial 3rd Deg.	0	4	10	16	17	19	19	19	21

Πίνακας 7.4: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced RMSE στο Βενζοπυρένιο.

<i>Growth Model</i>	<i>Toleranced MAE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	0	1	2	2	3	3	4	6	7
Gompertz	3	6	12	17	19	22	23	24	24
Logistic	32	35	36	38	39	40	43	45	46
Bertalanffy	9	11	12	16	17	17	18	18	19
Gaussian	0	1	2	3	6	8	10	13	15
Powerlaw	5	7	10	13	15	17	17	20	21
Polynomial 3rd Deg.	0	0	0	0	5	12	15	15	16

Πίνακας 7.5: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced MAE στο Βενζοπυρένιο.

<i>Substance Family</i>	<i>Growth Model</i>	<i>Mean Difference %</i>			
		<i>1st Value</i>	<i>2nd Value</i>	<i>3rd Value</i>	<i>4th Value</i>
Benzopyrene	Logistic	37.50	56.24	54.44	62.25
	Gompertz	19.03	66.35	69.39	75.07
	Gaussian	39.65	24.04	43.71	35.47
Polyamines	Logistic	0.61	11.38	29.87	53.81
	Gompertz	5.76	7.72	25.39	44.89
	Gaussian	2.82	10.81	21.83	32.94
Thiols	Logistic	3.59	13.95	18.82	23.96
	Gompertz	3.76	10.22	12.96	15.71
	Gaussian	14.46	17.38	17.13	9.54

Πίνακας 7.6: Προβλεψιμότητα των Αναπτυξιακών Μοντέλων.

## 7.2 Ταξινόμηση

Χρησιμοποιώντας τις αυθαίρετες σταθερές που προέκυψαν από την μέθοδο των ελαχίστων τετραγώνων δημιουργήσαμε ένα πρόβλημα ταξινόμησης ως προς τις δραστικές ουσίες με σκοπό να αιχμαλωτίσουμε την διαχωριστική ικανότητα που δίνουν τα αναπτυξιακά μοντέλα μέσα από τις σταθερές τους. Ανεξάρτητα από το αποτέλεσμα της αυθαιρετότητας των παραμέτρων, η ερώτημα παραμένει αν μέσα στην αυθαιρετότητα υπάρχουν ουσιαστικοί διαχωρισμοί.

<i>Metric</i>	<i>Cases</i>		
	<i>Case 1</i>	<i>Case 2</i>	<i>Case 3</i>
Accuracy	90%	89%	94%
Sensitivity	90%	94%	96%
Specificity	89%	70%	91%
Positive PV	87%	93%	91%
G-Mean	90%	81%	94%
F2-Measure	88%	93%	94%
Adj. F-Measure	89%	78%	93%

Πίνακας 7.7: Πίνακας Απόδοσης του ταξινομητή Γραμμικός Support Vector Machine.

Τα αποτελέσματα του ταξινομητή «γραμμικός SVM» απαρτίζουν τον πίνακα 7.7 για τις περιπτώσεις βενζοπυρένιο ενάντια όλων των δραστικών συνδυασμών (Περίπτωση 1), βενζοπυρένιο ενάντια πολυαμινών (Περίπτωση 2) και βενζοπυρένιο ενάντια θειολών (Περίπτωση 3). Όμοια για τον ταξινομητή «K-Nearest Neighbor» στον πίνακα 7.8. Και οι δύο ταξινομητές μας δίνουν απτά αποτελέσματα και μας αποδεικνύουν ότι οι δραστικές ουσίες μπορούν να διαχωριστούν με βάση τις παραμέτρους των αναπτυξιακών μοντέλων. Πιο συγκεκριμένα, όλες οι μετρικές έχουν ικανοποιητικές τιμές εκτός της μετρικής Specificity που είναι μόλις στο 54% και αποτελεί ένδειξη ανισορροπίας κλάσεων. Πράγματι, η Περίπτωση 2 έχει 59 παρατηρήσεις της μίας κλάσης έναντι 16 της άλλης και για αυτόν τον λόγο χρησιμοποιήσαμε επιπλέον μετρικές απόδοσης ταξινομητών όπως ο γεωμετρικός μέσος, η μετρική F2-Measure και η προσαρμοσμένη F-Measure. Ωστόσο οι επιπρόσθετες μετρικές επιβεβαιώνουν την καλή απόδοση του ταξινομητή και σε αυτήν την ιδιόρρυθμη περίπτωση.

Τέλος, για την εκτίμηση των χαρακτηριστικών ως προς την συμβολή τους στην

<i>Metric</i>	<i>Cases</i>		
	<i>Case 1</i>	<i>Case 2</i>	<i>Case 3</i>
Accuracy	91%	84%	90%
Sensitivity	91%	94%	94%
Specificity	89%	54%	86%
Positive PV	86%	87%	87%
G-Mean	90%	71%	90%
F2-Measure	89%	90%	90%
Adj. F-Measure	93%	94%	79%

Πίνακας 7.8: Πίνακας Απόδοσης του ταξινομητή K-Nearest Neighbor.

ταξινόμηση χρησιμοποιήσαμε τον αλγόριθμο Relief και τα κορυφαία πέντε χαρακτηριστικά απαρτίζουν τον πίνακα 7.9. Οι περιπτώσεις είναι όμοιες με πριν, όπου έχουμε βενζοπυρένιο ενάντια όλων των δραστικών συνδυασμών (Περίπτωση 1), βενζοπυρένιο ενάντια πολυαμινών (Περίπτωση 2) και βενζοπυρένιο ενάντια θειολών (Περίπτωση 3). Το απρόσμενο ήταν ότι δύο παράμετροι του λογιστικού μοντέλου προσφέρουν περισσότερο από οποιαδήποτε άλλη παράμετρο από το σύνολο των αναπτυξιακών μοντέλων ενώ η συμβολή τους είναι τετραπλάσια (44.47%) από την αμέσως επόμενη. Θεωρούμε ότι είναι ιδιαιτέρως σημαντικό ότι το λογιστικό μοντέλο τόσο στην απόδοσή του σε μετρικές προσαρμογής όσο και η διαχωριστική του ικανότητα κατέχει τόσο υψηλές θέσεις.

<i>Case</i>	<i>Growth Model</i>	<i>Model Parameter</i>	<i>Rank</i>	<i>Contribution</i>
	Bertalanffy	$\beta_2$	1	20,51%
Benzopyrene	Gaussian	c	2	18,53%
VS	Gaussian	b	3	18,09%
Polyamines	Gaussian	a	4	15,37%
	Bertalanffy	$\beta_3$	5	4,98%
Benzopyrene	Logistic	A	1	34,32%
VS	Logistic	Vo	2	10,17%
Polyamines	Exponential	Vo	3	8,91%
&	Gompertz	Vo	4	8,65%
Thiols	Gaussian	A	5	6,20%
	Logistic	A	1	27,64%
Benzopyrene	Bertalanffy	$\beta_1$	2	25,78%
VS	Powerlaw	a	3	7,10%
Thiols	Gompertz	Vo	4	6,01%
	Bertalanffy	$\beta_2$	5	5,60%

Πίνακας 7.9: Συνεισφορά Παραμέτρων Στην Ταξινόμηση.

### 7.3 Στατιστική Προσέγγιση

Η κανονικότητα των δεδομένων μας επιτρέπει ένα σύνολο τεστ που ονομάζονται παραμετρικά τεστ και διέπονται από κεντρικές τάσεις και την μέση τιμή. Ωστόσο, για να χρησιμοποιηθούν θα πρέπει να γίνει έλεγχος της κανονικότητας των δεδομένων, δηλαδή τα δεδομένα μας να αντλούνται από μία κανονική κατανομή. Τα τεστ ελέγχου κανονικότητας δίνονται στους πίνακες 7.10 για τις Πολυαμίνες και 7.11 για τις Θειόλες και μας κατευθύνουν σε μη παραμετρικά τεστ καθώς τα δεδομένα μας δεν ομοιάζουν της κανονικής κατανομής.

Ακόμη, τα τεστ των Lilliefor και Jarque-Bera ενώ μας δείχνουν οριακή κανονικότητα, δεν μπορούμε να βασιστούμε μόνο σε αυτά. Ο συνηθέστερος τρόπος είναι να υπάρχουν αρκετά τεστ που επιβεβαιώνουν καθώς όπως αναλύσαμε και θεωρητικά στο αντίστοιχο κεφάλαιο κάθε τεστ ελέγχει διαφορετικά πράγματα λ.χ. κύρτωση και λοξότητα ή μέγιστη απόκλιση στις εμπειρικές κατανομές. Ακόμη, το ισχυρότερο

<i>Normality Test</i>	<i>Single Dos. P-Value</i>	<i>Normality</i>	<i>Double Dos. P-Value</i>	<i>Normality</i>
Anderson-Darling	0.0294	0	0.0133	0
Jarque-Bera	0.2254	0	0.2152	1
Lilliefors	0.1602	0	0.0237	1
Shapiro-Wilk	0.0108	0	0.0072	0

Πίνακας 7.10: Τεστ Ελέγχου Κανονικότητας Πολυαμινών.

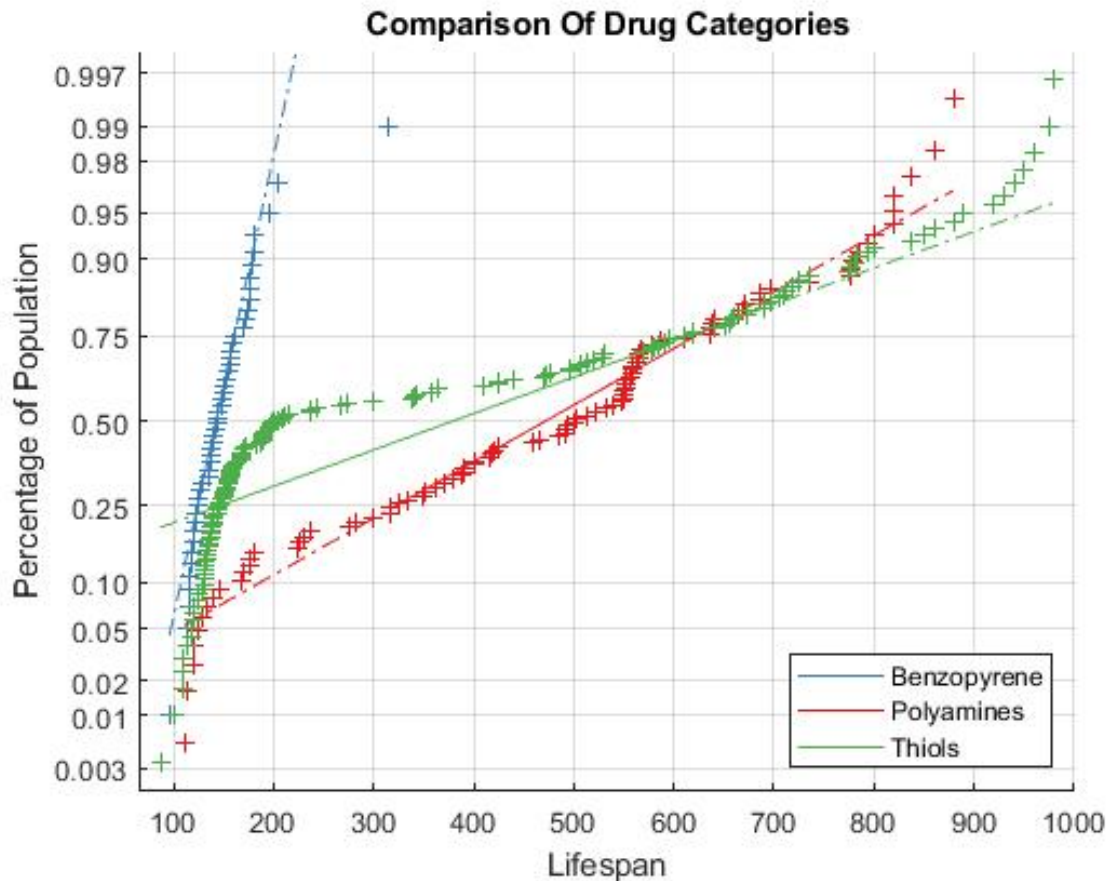
τεστ θεωρείται το τεστ των Shapiro-Wilk.

<i>Normality Test</i>	<i>Single Dosage P-Value</i>	<i>Normality</i>
Anderson-Darling	<0.001	0
Jarque-Bera	<0.001	0
Lilliefors	<0.001	0
Shapiro-Wilk	<0.001	0

Πίνακας 7.11: Τεστ Ελέγχου Κανονικότητας Θειολών.

Οπτικοποιώντας τον έλεγχο κανονικότητας, δώσαμε τον ορισμό του QQ γραφήματος έναντι της κανονικής κατανομής. Το συγκεκριμένο γράφημα μπορεί να παρατηρηθεί στο σχήμα 7.1. Αξίζει να σημειωθεί ότι παρατηρούμε την κανονικότητα του βενζοπυρενίου, ενώ οι πολυαμίνες και οι θειόλες έχουν αρκετές αποκλίσεις. Πιο συγκεκριμένα, η δράση των θειολών ομοιάζει του βενζοπυρενίου στις αρχικές τιμές των καταλήξεων ενώ στην συνέχεια δείχνει να βελτιώνεται ικανοποιητικά και να προσεγγίζει την κανονική κατανομή ενώ προς το τέλος παρατηρούμε ακραίες τιμές. Οι πολυαμίνες φαίνεται να έχουν πιο ομαλή κατανομή χωρίς επιρροές από το βενζοπυρένιο, ενώ είναι σημαντικό το γεγονός ότι το 50% του πληθυσμού καταλήγει μετά τις 480 ημέρες.

Στην συνέχεια παραθέτουμε τα αποτελέσματα του μη παραμετρικού τεστ των Kolmogorov-Smirnov κοινής κατανομής στον πίνακα 7.12. Παρατηρούμε ότι δεν υπάρχει κοινή κατανομή ανάμεσα σε βενζοπυρένιο και Πολυαμίνες αλλά ούτε σε βενζοπυρένιο και Θειόλες. Ακόμη, βλέπουμε ασυσχέτιση ανάμεσα σε μονές και διπλές δόσεις και για αυτόν τον λόγο ασχοληθήκαμε με τις στατιστικές ροπές. Οι στατιστικές ροπές αποτελούν μέτρα που αιχμαλωτίζουν τις κατανομές και μπο-



Σχήμα 7.1: QQ γράφημα των δραστικών ουσιών και του βενζοπυρενίου ως προς την αντίστοιχη κανονική κατανομή.

ρούν να τις συγκρίνουν και να τις περιγράψουν σε μεγάλο βαθμό.

Στον πίνακα 7.13 έχουμε τοποθετήσει τα αποτελέσματα και την σύγκριση των Πολυαμινών. Βλέπουμε ότι η μονή δόση καταφέρνει να βελτιώσει το μέσο προσδόκιμο ζωής κατά 320% ενώ η διπλή δόση κατά 399,5%. Οι τυπικές αποκλίσεις φαίνεται να μειώνονται όσο αυξάνουμε την δόση δίνοντάς μας την αίσθηση της εξομάλυνσης των τιμών, παρά το γεγονός ότι παραμένουν πολύ υψηλά επίπεδα. Η λοξότητα φαίνεται να μειώνεται και μερικές φορές να γίνεται αρνητική που είναι και το ζητούμενο διότι αυτό σημαίνει ότι έχουμε αρκετές καταλήξεις δεξιά της μέσης τιμής και λιγότερες αριστερά της μέσης τιμής. Η κύρτωση φαίνεται να πλησιάζει την τιμή τρία (3) που σημαίνει ότι ομοιάζουμε την κύρτωση της κανονικής κατανομής.

Σε αυτό το σημείο θα θέλαμε να παραθέσουμε ακόμη μία οπτικοποίηση που θεωρούμε αρκετά ξεχωριστή και σημαντική. Δημιουργήσαμε ακόμη ένα QQ γράφημα (σχέδιο 7.2) ανάμεσα στις μονές και διπλές δόσεις και παρατηρήσαμε την



<i>Case</i>	<i>Dosage</i>	<i>P-Value</i>	<i>Same Distribution</i>
Benzopyrene-Polyamines	Single	≪0.0001	0
Benzopyrene-Polyamines	Double	≪0.0001	0
Benzopyrene-Thiols	Single	≪0.0001	0
Polyamines-Thiols	Single-Single	≪0.0001	0
Polyamines-Polyamines	Single-Single	≪0.0001	0

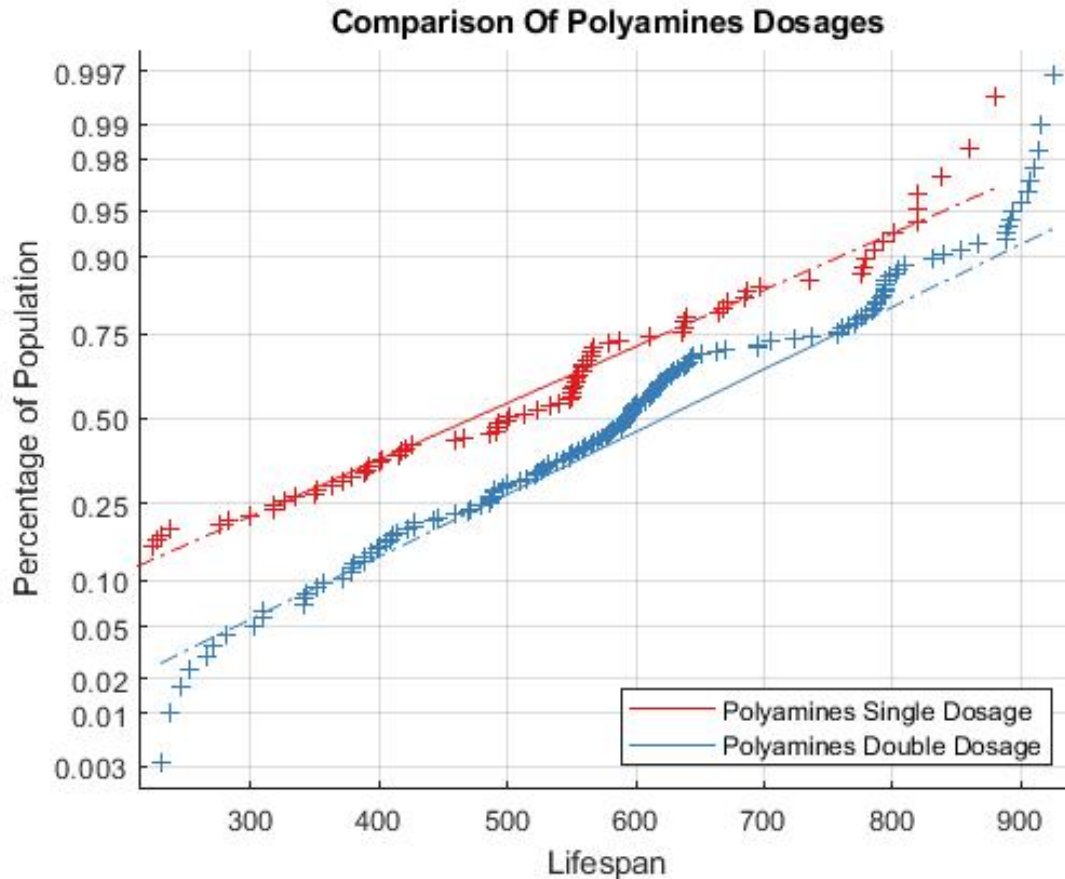
Πίνακας 7.12: Τεστ Κοινής Κατανομής των Kolmogorov-Smirnov.

<i>Substance</i> <i>Family</i>	<i>Moments</i>				Mortality Rate / Insufficient Reaction
	<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>4th</i>	
Benzopyrene	148.5800	1162	2.4475	13.5911	100.0%
Polyamines Single	475.8093	44456	-0.0697	2.1479	18,47%
Difference %	320.23%	3825%			18.47%
Polyamines Double	593.7181	30732	0.0230	2.2980	1,34%
Difference %	399.5%	2644%			1.34%

Πίνακας 7.13: Στατιστικές Ροπές σε Πολυαμίνες και σύγκριση έναντι του Βενζοπυρενίου.

παραλληλία των δύο δόσεων. Αυτό το γεγονός μας αποδεικνύει την ύπαρξη αναλογίας ανάμεσα σε δόσεις και αναστολή της χημικής καρκινογένεσης. Ωστόσο έχοντας μόνο δύο δείγματα, δεν μπορούμε να είμαστε βέβαιοι ότι αυτή η τάση θα συνεχιστεί λόγου χάρη στην τριπλή δόση. Ακόμη το τεστ των Mann-Whitney U επιβεβαιώνει τον ισχυρισμό βελτίωσης με την αύξηση της δόσης. Πιο συγκεκριμένα με P-Value  $3.8e-05$ , ισχυρίζεται ότι η τυχαία επιλογή μίας τιμής κατάληξης από το σύνολο Πολυαμινών διπλής δόσης να είναι μεγαλύτερη από μία τυχαία τιμή από το σύνολο Πολυαμινών μονής δόσης συμβαίνει πιο συχνά από το 50% των φορών (πιο συχνά από την τυχαία επιλογή).

Στον πίνακα 7.14, παραθέτουμε παρόμοια ανάλυση αλλά σε κάθε δραστική ουσία των πολυαμινών ξεχωστά ως προς τις μονές και διπλές δοσολογίες. Πρώτον, παρατηρούμε ότι η αιθυλενοδιαμίνη και η πουτρεσκίνη βελτιώνουν το μέσο προσδόκιμο ζωής κατά 10% και 13% ενώ η κανταβερίνη κατά 1%. Δεύτερον, τα δείγματα



Σχήμα 7.2: QQ γράφημα των Πολυαμινών μονής και διπλής δόσης ως προς την αντίστοιχη κανονική κατανομή.

της αιθυλενοδιαμίνης και της κανταβερίνης τείνουν να γίνουν πιο συμπαγή, με λιγότερες ακραίες τιμές σε σχέση με την πουτρεσκίνη. Όλες οι λοξότητες μικραίνουν και μερικές καταλήγουν ακόμη και σε αρνητικές που είναι ένα πολύ υποσχόμενο φαινόμενο όπως εξηγήσαμε προ ολίγου. Οι κυρτόσεις παραμένουν σε μικρά επίπεδα κοντά στην τιμή δύο (2) και τείνουν να ομοιάσουν στην τιμή της κανονικής κατανομής. Ενώ άκρως σημαντικό είναι το γεγονός ότι τα πρότυπα δεν καταλήγουν από την νόσο της καρκινικής κυτταρικής ανάπτυξης σε ποσοστό που πλησιάζει το 100% όταν λαμβάνουν διπλή δόση πολυαμίνης. Πιο συγκεκριμένα, υπάρχει μόνο μία περίπτωση που κατέλειξε από την χημική καρκινογένεση στα 50 πρότυπα της διπλής δόσης πουτρεσκίνης.

Τέλος, παραθέτουμε την οπτικοποίηση του τελευταίου πίνακα για να καταδείξουμε την σημαντικότητα των στατιστικών ροπών σε περιπτώσεις μη κανονικών δεδομένων μέσα από τα ιστογράμματα της Αιθυλενοδιαμίνης, Κανταβερίνης και Πουτρεσκίνης. Μελετώντας το ιστόγραμμα Α.3 της Αιθυλενοδιαμίνης, παρατηρούμε

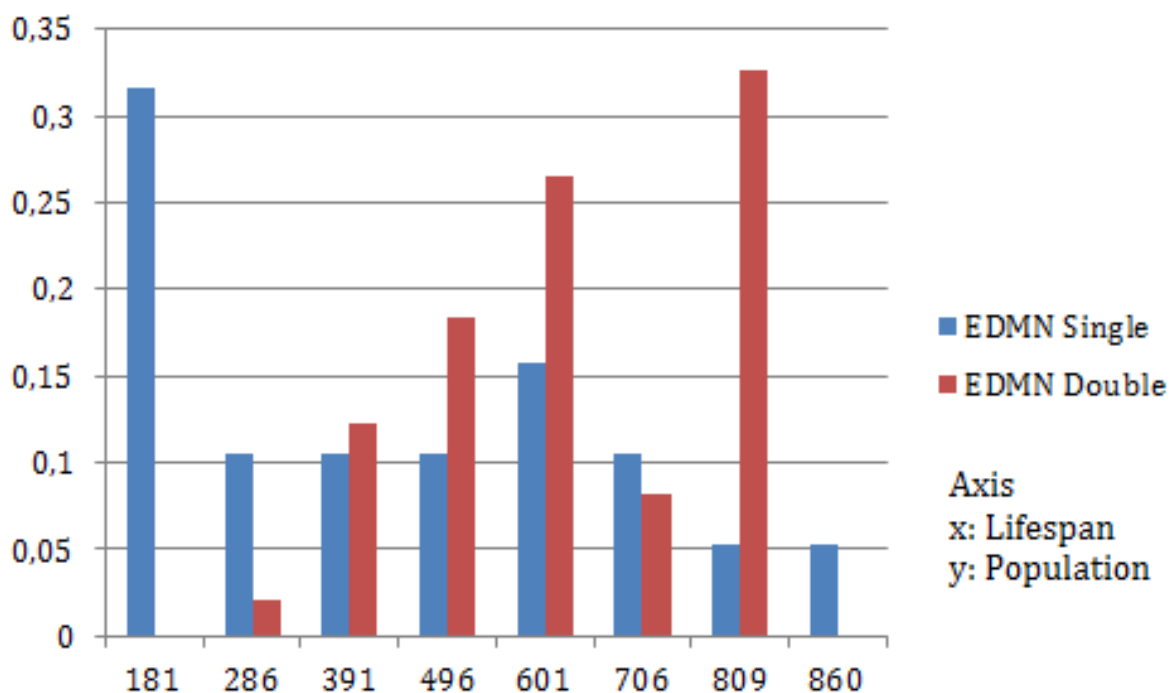
<i>Polyamine</i>	<i>Moments</i>				Mortality Rate / Insufficient Reaction
	<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>4th</i>	
Ethylened. Single	549.4615	31249	0.1733	1.8886	31.57%
Ethylened. Double	613.5306	22859	-0.2937	1.9499	0%
Difference %	11.66%	-26.84%			-100%
Cadaverine Single	555.5857	30981	-0.1685	2.0756	30%
Cadaverine Double	561.5200	16812	-0.4984	3.8224	0%
Difference %	1.08%	-45.73%			-100%
Putrescine Single	543.2902	23230	0.5159	2.7033	4.6512%
Putrescine Double	614.0816	50084	0.0681	1.5182	2%
Difference %	13.03%	115.60%			-57%
Benzopyrene	148.5800	1162	2.4475	13.5911	100.0%

Πίνακας 7.14: Στατιστικές Ροπές σε Πολυαμίνες και σύγκριση μονών και διπλών δόσεων.

την αρνητική λοξότητα της διπλής δόσης ενώ στην μονή δόση υπάρχει θετική λοξότητα που οφείλεται στις πρώιμες καταλήξεις (πριν τις 233 ημέρες). Η γενική εικόνα που λαμβάνουμε από τη συγκεκριμένη δραστική ουσία είναι ότι η μονή δόση έχει σημαντικό αριθμό καταλήξεων λόγω ανεπαρκούς αντίδρασης σε σχέση με το βενζοπυρένιο (30%). Οι καταλήξεις αυτές απλώνονται στην διπλή δόση δεδομένης μίας αρνητικής λοξότητας και ως απόρροια όλων αυτών έχουμε την βελτίωση του προσδόκιμου ζωής κατά 11.66%. Δεν υπάρχει κεντρική τάση τιμών και κατά επέκταση η τυπική απόκλιση δεν μας δίνει κάποια επιπλέον πληροφορία.

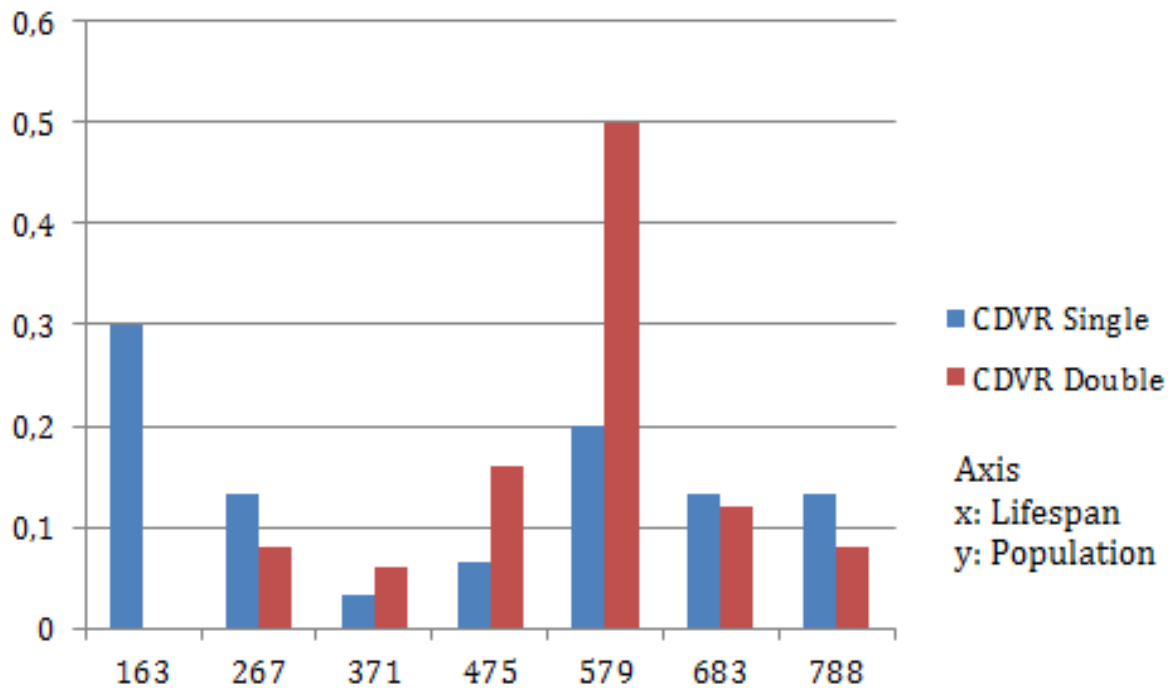
Μελετώντας το ιστόγραμμα A.4 της Κανταβερίνης, παρατηρούμε και σε αυτήν την περίπτωση ότι οι πρώιμες καταλήξεις εκμηδενίζονται στην διπλή δόση. Ωστόσο, ενώ στην μονή δόση φαίνεται να μην υπάρχει κάποιο μοτίβο στις καταλήξεις, στην διπλή δόση αυτό αλλάζει. Θα μπορούσαμε να πούμε ότι η διπλή δόση της Κανταβερίνης ομοιάζει της κανονικής κατανομής με μέση τιμή στο κάδο (527-631).

Μελετώντας το ιστόγραμμα A.5 της Πουτρεσκίνης, παρατηρούμε ότι δεν υπάρχουν πρώιμες καταλήξεις στην μονή αλλά ούτε και στην διπλή δόση. Αυτό σημαίνει ότι η Πουτρεσκίνη έχει ισχυρότερη δράση έναντι του βενζοπυρενίου στο πρώιμο στάδιο της χημικής καρκινογένεσης από ότι οι υπόλοιπες ουσίες. Όσο αφορά την

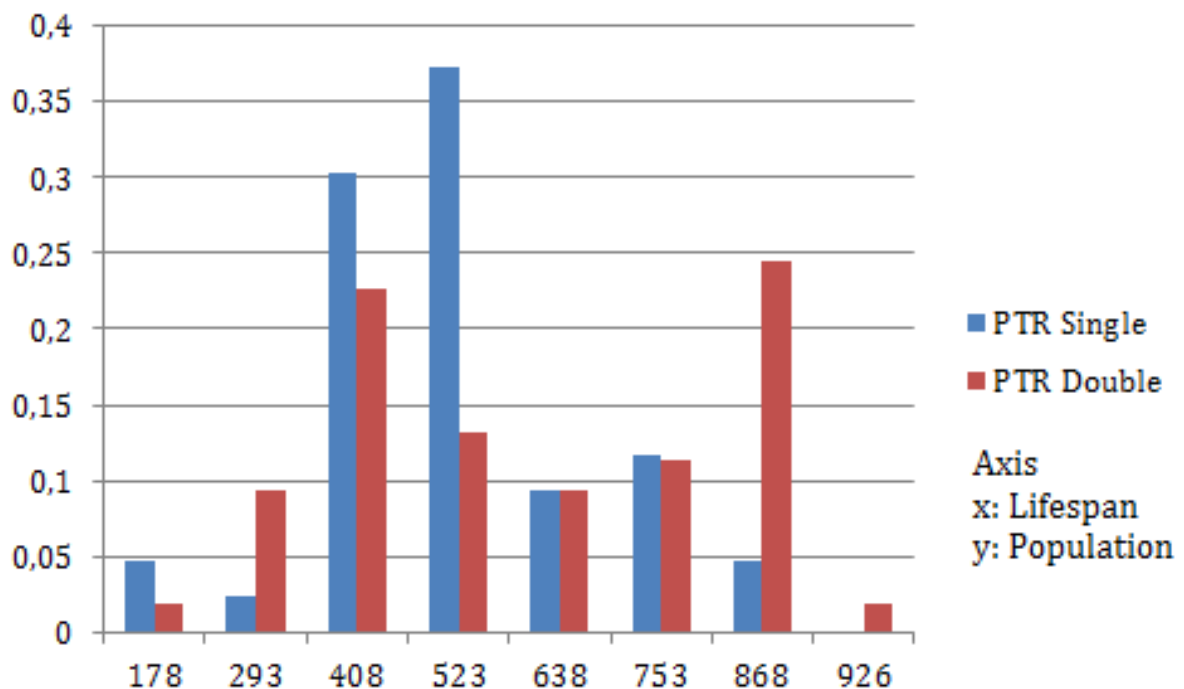


Σχήμα 7.3: Σύγκριση Δόσεων με ιστόγραμμα σχετικών συχνοτήτων στην Αιθυλενοδιαμίνη.

λοξότητα στην μονή δόση είναι θετική ενώ στην διπλή δόση λόγω εμφάνισης δύο πόλων (ισουψείς κάδοι) δεν καθορίζει σημαντικά την κατανομή. Ακόμη, καθώς ταυτίζεται η θέση του πρώτου πόλου στην διπλή δόση με αυτόν της μονής, θεωρούμε ότι πρόκειται για μία περιοχή μερικής δράσης της ουσίας. Για αυτόν τον λόγο έχουμε μείωση της σχετικής συχνότητας στην διπλή δόση. Πιθανολογείται ότι σε τριπλή δόση, θα μεωθεί επιπρόσθετα ενώ θα βελτιωθεί το ύψος του κάδου (810-926). Με βάση αυτά τα στοιχεία απορρέει η υπεροχή της Πουτρεσκίνης με βάση τις δύο δόσεις που έχουν δοκιμασθεί.



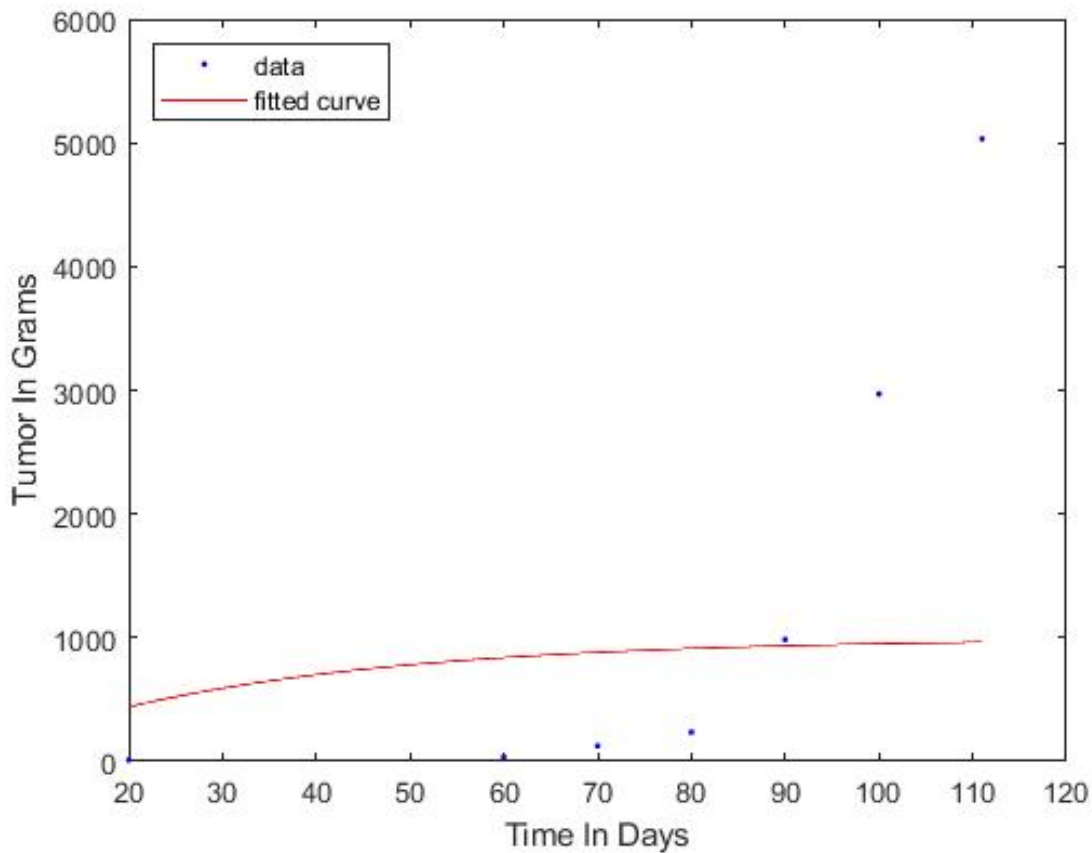
Σχήμα 7.4: Σύγκριση Δόσεων με ιστόγραμμα σχετικών συχνοτήτων στην Κανταβερίνη.



Σχήμα 7.5: Σύγκριση Δόσεων με ιστόγραμμα σχετικών συχνοτήτων στην Πουτρεσίνη.

## 7.4 Μη αποδοτικά μοντέλα και διαδικασίες

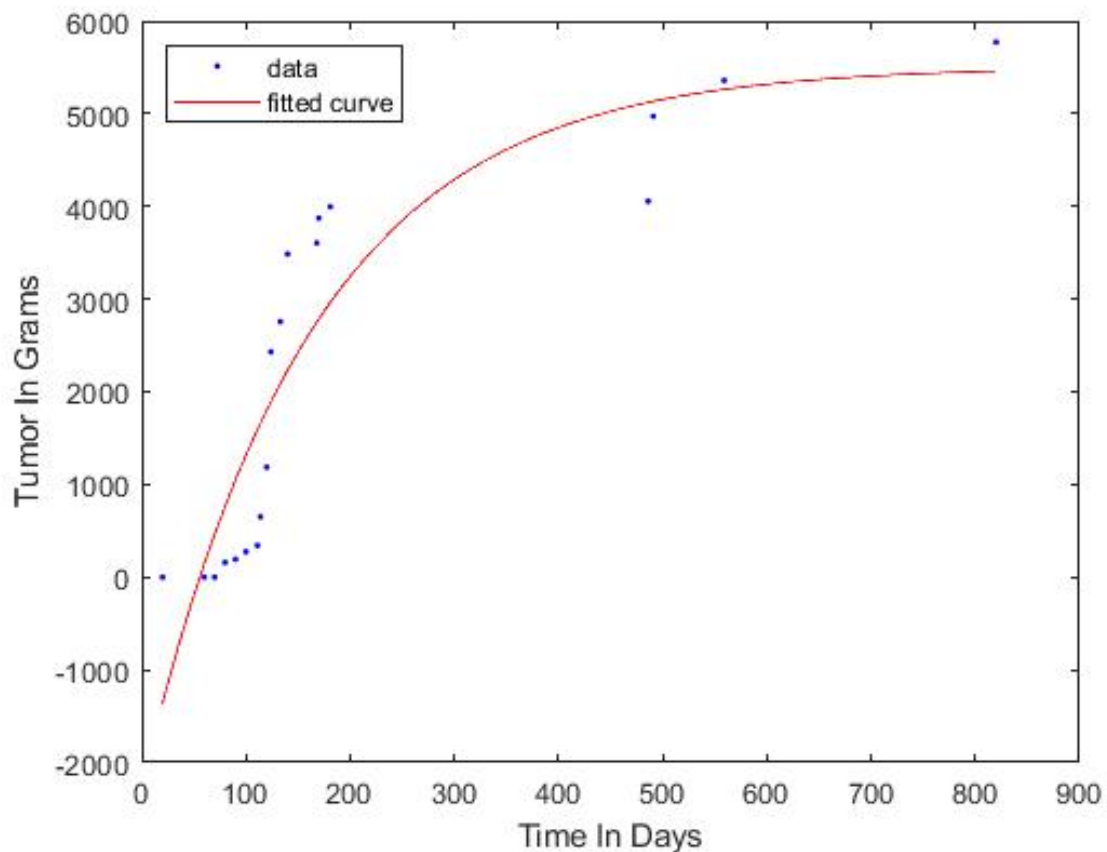
Τα δεδομένα μας δεν μπορούσαν να περιγραφούν σε ικανοποιητικό βαθμό από τα μοντέλα Brody και Weibull. Μορφολογικά τα δύο αυτά αναπτυξιακά μοντέλα ομοιάζουν και συντελούν μία κυρτή καμπύλη. Το πιο πιθανό ενδεχόμενο είναι ότι αυτή η συμπεριφορά δεν αρμόζει στα δεδομένα μας καθώς παρεκκλίνουν αρκετά από την αρχική ροή των καταγραφών. Το τελευταίο μπορεί να παρατηρηθεί στο σχήμα 7.6 όπου έχει χρησιμοποιηθεί το αναπτυξιακό μοντέλο του Brody.



Σχήμα 7.6: Περίπτωση κακής προσαρμογής με χρήση του μοντέλου Brody

Ενώ σε μόλις τέσσερις περιπτώσεις από το σύνολο των πειραματοζώων υπάρχει ικανοποιητική προσαρμογή όπως βλέπουμε στο σχήμα 7.7. Πρόκειται για πολύ μικρό ποσοστό προσαρμογής και έτσι καταλήγουμε ότι τα δεδομένα μας δεν προσεγγίζονται από το μοντέλο Brody. Όμοια, δεν υπήρξε επιτυχία προσαρμογής στο μοντέλο Weibull τριών παραμέτρων.

Όσο αφορά την μετασχηματισμό των δεδομένων αλλά και την κανονικοποίησή τους, δεν παρουσιάζονται ιδιαίτερες αλλαγές στα τεστ κανονικότητας. Πιο συγκε-



Σχήμα 7.7: Περίπτωση καλής προσαρμογής με χρήση του μοντέλου Brody

κριμένα, η κανονικοποίηση δεν έφερε καμία αλλαγή στα τεστ κανονικότητας ενώ οι μετασχηματισμοί είχαν μεταβαλλόμενη συμπεριφορά. Δεν έχουμε κανένα λόγο να αμφισβητήσουμε ότι τα δεδομένα μας δεν είναι κανονικά καθώς πολλά φαινόμενα δεν είναι κατά ανάγκη κανονικά. Έχουμε αρκετές ενδείξεις να πιστέψουμε ένα τέτοιο γεγονός και καμία ένδειξη για το αντίθετο. Ο πίνακας 7.15 απαρτίζεται από τα τεστ κανονικότητας με χρήση γνωστών μετασχηματισμών με έμφαση στην ύπαρξη αρνητικής λοξοτήτας. Ενώ συμπληρωματικοί πίνακες δίνονται στο παράρτημα Α και συγκεκριμένα στους πίνακες Α.10 και Α.11.

<i>Transform</i>	<i>Normality Test</i>	<i>Single Dosage P-Value</i>	<i>Normality</i>
C. Logarithm	Anderson-Darling	≪0.0001	0
	Jarque-Bera	0.0152	0
	Lilliefors	≪0.0001	0
	Shapiro-Wilk	≪0.0001	0
Square Root	Anderson-Darling	0.0010	0
	Jarque-Bera	0.0635	1
	Lilliefors	0.1602	0
	Shapiro-Wilk	0.0006	0
Cubic Root	Anderson-Darling	0.0001	0
	Jarque-Bera	0.0256	0
	Lilliefors	≪0.0001	0
	Shapiro-Wilk	0.0005	0

Πίνακας 7.15: Τεστ Ελέγχου Κανονικότητας Πολυαμινών Μονών Δόσεων με Μετασχηματισμούς.



## ΚΕΦΑΛΑΙΟ 8

# ΚΕΦΑΛΑΙΟ: ΚΑΙΝΟΤΟΜΙΕΣ ΚΑΙ ΜΕΛΛΟΝΤΙΚΟΙ ΣΤΟΧΟΙ

---

### 8.1 Καινοτομίες

### 8.2 Μελλοντικοί Στόχοι

---

Σε αυτό το κεφάλαιο υπογραμμίζονται οι καινοτομίες της παρούσας μεταπτυχιακής εργασίας που είχαν ως κυρίαρχο στόχο την βελτίωση της γενικής εικόνας της ανάπτυξης της χημικής καρκινογένεσης. Ακόμη, θα θέλαμε η εργασία αυτή να αποτελέσει έναυσμα για περισσότερη δραστηριότητα στον τομέα της χρήσης δραστικών ουσιών της οικογένειας των πολυαμινών τόσο ως προληπτικές ουσίες όσο και ως αναστολείς της καρκινικής ανάπτυξης. Παράλληλα έχουν τεθεί κάποιοι στόχοι για το μέλλον που θα θέλαμε να εξετάσουμε διεξοδικά.

### 8.1 Καινοτομίες

Η προσέγγιση που ακολουθήσαμε στην «Μοντελοποίηση της ανάπτυξης της χημικής καρκινογένεσης» ήταν η εξατομικευμένη ή ατομική. Στην ουσία μοντελοποιήσαμε κάθε πρότυπο ξεχωριστά. Αυτή η προσέγγιση δίνει τεράστια ισχύ στα εξαγόμενά μας παρά τον πολλαπλάσιο χρόνο ενασχόλησης που ενέχει. Το τελευταίο συμβαίνει διότι ο πληθυσμός μας δεν αντιπροσωπεύεται από το τυπικό πρότυπο. Αντίθετα,

κάθε πρότυπο είναι μοναδικό και ανεπανάληπτο ως προς την φυσιολογία του όπως κάθε έμβιος οργανισμός και αντιπροσωπεύεται μόνο από τον εαυτό του. Για αυτόν τον λόγο, θεωρούμε ότι όλα τα εξαγόμενα της παρούσας εργασίας δεν μπορούν να είναι τυχαία ούτε επηρεάζονται από κεντρικές τάσεις. Σε περίπτωση που υπήρχαν ιδιομορφίες ή γενικές τάσεις του πληθυσμού, καταγράφηκαν για επιπρόσθετη μελέτη. Οι περισσότερες έρευνες στον τομέα της αναπτυξιακής μοντελοποίησης βασίζονται σε τυπικά πρότυπα και δεν θεωρούν την ύπαρξη των κρυφών παραγόντων σημαντική όπως είναι η ετερογένεια των οργανισμών ως προς τις αντιδράσεις του οργανισμού αλλά και η ομοιόσταση σε περιπτώσεις εξέλιξης ασθενειών.

Στην συνέχεια πραγματοποιήθηκε η σύγκριση των αναπτυξιακών μοντέλων κάτω από διάφορες μετρικές όπως SSE, RMSE και MAE. Εκεί χρησιμοποιήσαμε την ανεχόμενη έκδοσή τους επί τοις εκατό (π.χ. Toleranced SSE 10%). Στην προκειμένη περίπτωση, αναζητάμε μία μετρική που μπορεί να αιχμαλωτίσει την απόδοση ενός μοντέλου με τις καλύτερες δυνατές τιμές σε μία μετρική λ.χ. SSE. Ωστόσο όταν αυτό δεν συμβαίνει, θα θέλαμε οι επιδόσεις στην μετρική SSE να μην απέχουν πάρα πολύ από τις βέλτιστες αλλά να είναι παραπλήσιες. Με άλλα λόγια αναζητούμε το μοντέλο εκείνο που έχει καθολική περιγραφική ικανότητα. Δεν αναζητούμε το μοντέλο εκείνο που έχει βέλτιστη απόδοση στην πλειονότητα των προτύπων και χειρίστη απόδοση στην μειονότητα. Για αυτόν τον λόγο χρησιμοποιήσαμε διαφορετικά ποσοστά στην ανέχεια στο εύρος 0% έως 40%. Σκοπό μας ήταν να επισημάνουμε την βελτίωση του κάθε μοντέλου όσο «χαλαρώνουμε» τα όρια της εκάστοτε μετρικής. Αν και η προσέγγιση φαίνεται αφελής και απλή, αντικατοπτρίζει και αιχμαλωτίζει πλήρως αυτό που αναζητούμε σε ένα αναπτυξιακό μοντέλο: την καθολικότητα όχι ως προς την βέλτιστη απόδοση της πλειονότητας των προτύπων, αλλά την καθολικότητα ως προς την βέλτιστη απόδοση με μία συγκεκριμένη απόκλιση, σε όλο τον πληθυσμό.

Στην ενότητα των αποτελεσμάτων παρατέθηκαν κάποια εύρη τιμών για τις παραμέτρους των αναπτυξιακών μοντέλων (πίνακας 7.2). Η εικόνα που είχαμε έως τώρα για τις παραμέτρους των αναπτυξιακών μοντέλων που χρησιμοποιήθηκαν ήταν επτά έως δέκα διανύσματα παραμέτρων. Η συγκεκριμένη εικόνα ήταν περιορισμένη και δεν θα μπορούσε σε καμία περίπτωση να αποδώσει κάποια τάση που να μην αμφισβητηθεί. Στην παρούσα εργασία δίνονται εύρη τιμών που βασίστηκαν στην μοντελοποίηση 150 πειραματοζώνων και καταστείτε σαφές ότι εάν υπάρχει κάποια τάση σε τιμές θα αποτυπωθεί. Παράλληλα, οι συγκεκριμένες τιμές μπορούν να μετατρέψουν ένα  $\nu$ -παραμετρικό μοντέλο σε  $\nu-1$ -παραμετρικό μοντέλο, δίνοντας

χρόνο στον ερευνητή να ασχοληθεί με τις πιο μεταβαλλόμενες παραμέτρους.

Η εικόνα που είχε διαμορφωθεί για την ανάπτυξη της χημικής καρκινογένεσης μέχρι και την τελευταία δεκαετία βασιζόταν σε εργασίες με πολύ μικρό αριθμό προτύπων λόγω της ισχύουσας νομοθεσίας. Οι πολυετείς έρευνες του Τμήματος Φυσιολογίας του Πανεπιστημίου Ιωαννίνων μας έδωσαν ώθηση να ξεπεραστεί αυτός ο φραγμός και να στηριχθεί ένα σύνολο δεδομένων άνω των 400 πειραματοζώων. Ο ικανοποιητικός αυτός αριθμός προτύπων επιτρέπει ένα σύνολο στατιστικών μεθόδων όπως ο έλεγχος κανονικότητας, τα παραμετρικά και μη-παραμετρικά τεστ, το τεστ κοινής κατανομής αλλά και τις στατιστικές ροπές. Δεν υπήρξε ικανοποιητικός αριθμός προτύπων σε παρεμφερείς έρευνες για να διεξαχθούν αντίστοιχες μελέτες. Τα αποτελέσματα αυτών των μεθόδων, μετατρέπουν τις ενδείξεις σε αποδείξεις γύρω από το φάσμα της αποδοτικότητας των δραστικών ουσιών τόσο των πολυαμινών όσο και των θειολών. Εκτός αυτού, μας υπογραμμίζουν ότι η πορεία που επιλέχθηκε ήταν ορθή. Ακόμη, αποδεικνύεται η διαφορετικότητα της συμπεριφοράς των πολυαμινών έναντι των θειολών και η παράλληλη μετατόπιση των κατανομών της διπλής δόσης δραστικής ουσίας έναντι της μονής δόσης δραστικής ουσίας. Το τελευταίο μας δίνει την αίσθηση της ύπαρξης αναλογίας ανάμεσα στο καρκινικό βάρος ενός προτύπου και την δοσολογία δραστικής ουσίας που του χορηγήθηκε.

Συνεχίζοντας στο ίδιο πλαίσιο, ο ικανοποιητικός αριθμός των προτύπων επιτρέπει και την δημιουργία ενός προβλήματος «Ταξινόμησης» με στόχο να διαχωριστούν οι χωροταξικές διαφορές των δραστικών ουσιών σε σχέση με το βενζοπυρένιο. Στην ταξινόμηση χρησιμοποιήσαμε ως χαρακτηριστικά τις παραμέτρους των αναπτυξιακών μοντέλων προσπαθώντας να ξεχωρίσουμε τις πολυαμίνες ή τις θειόλες από το βενζοπυρένιο. Καταλήξαμε ότι ο χώρος των παραμέτρων των αναπτυξιακών μοντέλων που περιγράφει την δράση των πολυαμινών διαφέρει ικανοποιητικά από τον αντίστοιχο χώρο του βενζοπυρενίου. Όμοια και για τις θειόλες. Ενώ παρατηρήσαμε την μεγάλη συνεισφορά της λογιστικής παραμέτρου A, στο πρόβλημα της ταξινόμησης, εδραιώνοντας για ακόμη μία φορά την περίοπτη θέση του συγκεκριμένου αναπτυξιακού μοντέλου.

## 8.2 Μελλοντικοί Στόχοι

Αρχικά θα θέλαμε η προσοχή να εστιαστεί περισσότερο στην προβλεψιμότητα των αναπτυξιακών μοντέλων καθώς αποτελεί σημαντικό παράγοντα επιλογής του μοντέλου. Πιο συγκεκριμένα, όπως έχει ήδη προταθεί θα θέλαμε συχνότερες καταγραφές του καρκινικού βάρους σε περιπτώσεις όπου η εξέλιξη της νόσου βρίσκεται σε στάδιο εκθετικότητας. Με περισσότερες καταγραφές θα μπορούσαμε να έχουμε μία πιο ισχυρή απάντηση στο θέμα της πρόβλεψης. Θεωρούμε ότι η γενική εικόνα που υπάρχει στο συγκεκριμένο θέμα είναι περιορισμένη, αν και θεωρητικά η ευελιξία των αναπτυξιακών μοντέλων έχει αποδειχθεί. Προτείνουμε δύο συνεχόμενες καταγραφές να μην διαφέρουν περισσότερο από δέκα τοις εκατό.

Σε δεύτερο χρόνο, θα θέλαμε να επαναλάβουμε τα πειράματα με χρήση πολυαμινών μονής και διπλής δόσης με σκοπό να βρεθούν παράλληλες κατανομές που ομοιάζουν με την κανονική κατανομή. Ακόμη, θα ήταν αξιοσημείωτο να υπάρχει μεγάλη ακρίβεια στις στατιστικές τους ροπές οπότε υπάρχει απαίτηση για μεγάλο δείγμα είτε σε έμβιους οργανισμούς είτε σε «οργανισμούς μικροσκοπίου». Υπό αυτές τις προτάσεις, θα μπορούσαν να αντληθούν δοσολογικοί τύποι που θα συσχετίζουν την δοσολογία των πολυαμινών με την αναστολή της χημικής καρκινογένεσης. Επιπλέον, είναι βέβαιο ότι θα υπάρχει ένα σημείο που η αύξηση της δοσολογίας δεν θα επιφέρει ανάλογα αποτελέσματα (Point of Deminishing Returns). Ωστόσο, οι ενδείξεις που έχουμε μέχρι τώρα δεν καταδεικνύουν τέτοιο φαινόμενο στις διπλές δόσεις.

Η οικογένεια των πολυαμινών είναι πολυάριθμη και χωρίζεται σε διάφορες κατηγορίες. Με βάση την προέλευση έχουμε τις φυσικές και τις συνθετικές πολυαμίνες. Θα ήταν αρκετά ενδιαφέρον να δοκιμάσουμε τόσο φυσικές πολυαμίνες όπως σπερμιδίνη ή σπερμίνη όσο και συνθετικές όπως διαιθυλενοτριάμινη ή μακροκυκλικές πολυαμίνες. Το περίοπτο με την συγκεκριμένη οικογένεια είναι ότι ο ρόλος τους παραμένει άγνωστος μέχρι και σήμερα ενώ η επιρροή τους σε βιολογικές διαδικασίες, όπως η ανάπτυξη των κυττάρων και η αλληλεπίδραση με βιοσυνθετικά ένζυμα, έχει αποδειχθεί.

Τέλος, θα θέλαμε να ελέγξουμε την δράση των πολυαμινών και των θειολών και σε περιπτώσεις που ήδη έχουν καρκινικές κυτταρικές αναπτύξεις καθώς θεωρητικά η πολυαμίνη μπορεί να δράσει τόσο προκαταβολικά όσο και σε αργότερο χρόνο. Οι πρώτες μας ενδείξεις είναι άκρως ενθαρρυντικές.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

---

- [1] Carolus Linnaeus et al. *Systema naturae*, vol. 1. *Systema naturae*, Vol. 1, 1758.
- [2] William A Sodeman Jr. The biology of schistosomes from genes to latrines. *The American Journal of Tropical Medicine and Hygiene*, 41(2):251–251, 1989.
- [3] H Schwegler and HP Lipp. Hereditary covariations of neuronal circuitry and behavior: correlations between the proportions of hippocampal synaptic fields in the regio inferior and two-way avoidance in mice and rats. *Behavioural brain research*, 7(1):1–38, 1983.
- [4] Yves Lamberty and Alma J Gower. Age-related changes in spontaneous behavior and learning in nmri mice from middle to old age. *Physiology & Behavior*, 51(1):81–88, 1992.
- [5] G Kallistratos and E Fasske. Inhibition of benzo (a) pyrene carcinogenesis in rats with vitamin c. *Journal of cancer research and clinical oncology*, 97(1):91–96, 1980.
- [6] John E Hall and Michael E Hall. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.
- [7] Shannon L Nowotarski, Patrick M Woster, and Robert A Casero Jr. Polyamines and cancer: implications for chemoprevention and chemotherapy. *Expert reviews in molecular medicine*, 15:e3, 2013.
- [8] Diane Russell and Solomon H Snyder. Amine synthesis in rapidly growing tissues: ornithine decarboxylase activity in regenerating rat liver, chick embryo, and various tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 60(4):1420, 1968.

- [9] Andrea Manni, David Mauger, Phyllis Gimotty, and Betty Badger. Prognostic influence on survival of increased ornithine decarboxylase activity in human breast cancer. *Clinical cancer research*, 2(11):1901–1906, 1996.
- [10] Andrea Manni, Rhea Grove, Susan Kunselman, and Marcelo Aldaz. Involvement of the polyamine pathway in breast cancer progression. *Cancer letters*, 92(1):49–57, 1995.
- [11] Sanjay Gupta, Nihal Ahmad, Susan R Marengo, Gregory T MacLennan, Norman M Greenberg, and Hasan Mukhtar. Chemoprevention of prostate carcinogenesis by  $\alpha$ -difluoromethylornithine in tramp mice. *Cancer research*, 60(18):5125–5133, 2000.
- [12] Arseniy S Karkach. Trajectories and models of individual growth. *Demographic research*, 15:347–400, 2006.
- [13] Roy M Anderson, B Anderson, and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.
- [14] Geoffrey P Garnett, Simon Cousens, Timothy B Hallett, Richard Steketee, and Neff Walker. Mathematical models in the evaluation of health programmes. *The Lancet*, 378(9790):515–525, 2011.
- [15] Kathryn Glass and Niels G Becker. Estimating antiviral effectiveness against pandemic influenza using household data. *Journal of the Royal Society Interface*, 6(37):695–703, 2009.
- [16] Jane J Kim, Jesse Ortendahl, and Sue J Goldie. Cost-effectiveness of human papillomavirus vaccination and cervical cancer screening in women older than 30 years in the united states. *Annals of internal medicine*, 151(8):538–545, 2009.
- [17] Charles P Winsor. The gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences of the United States of America*, 18(1):1, 1932.
- [18] JA Nelder. The fitting of a generalization of the logistic curve. *Biometrics*, 17(1):89–110, 1961.
- [19] Ludwig Von Bertalanffy. Quantitative laws in metabolism and growth. *The quarterly review of biology*, 32(3):217–231, 1957.

- [20] Samuel Brody and Henry A Lardy. Bioenergetics and growth. *The Journal of Physical Chemistry*, 50(2):168–169, 1946.
- [21] JO Rawlings and WW Cure. The weibull function as a dose-response model to describe ozone effects on crop yields 1. *Crop Science*, 25(5):807–814, 1985.
- [22] FJ Richards. A flexible growth function for empirical use. *Journal of experimental Botany*, 10(2):290–301, 1959.
- [23] JE Brown, HA Fitzhugh Jr, and TC Cartwright. A comparison of nonlinear models for describing weight-age relationships in cattle. *Journal of Animal Science*, 42(4):810–818, 1976.
- [24] Boris Zeide. Analysis of growth equations. *Forest science*, 39(3):594–616, 1993.
- [25] Jonathan M Read, Colin PD Birch, and John A Milne. Heathmod: a model of the impact of seasonal grazing by sheep on upland heaths dominated by *Calluna vulgaris* (heather). *Biological Conservation*, 105(3):279–292, 2002.
- [26] AT Goshu. Simulation study of the commonly used mathematical growth models. *Journal of the Ethiopian Statistical Association*, 17:44–53, 2008.
- [27] Vinay G Vaidya and Frank J Alexandro Jr. Evaluation of some mathematical models for tumor growth. *International journal of bio-medical computing*, 13(1):19–35, 1982.
- [28] J Goudriaan and HH Van Laar. Modelling potentiel crop growth process, 1994.
- [29] Purnachandra Rao Koya and Ayele Taye Goshu. Generalized mathematical model for biological growths. *Open Journal of Modelling and Simulation*, 2013, 2013.
- [30] Alberto d’Onofrio. A general framework for modeling tumor-immune system competition and immunotherapy: Mathematical analysis and biomedical inferences. *Physica D: Nonlinear Phenomena*, 208(3-4):220–235, 2005.
- [31] Anastasios Tsoularis and James Wallace. Analysis of logistic growth models. *Mathematical biosciences*, 179(1):21–55, 2002.

- [32] Peter D Tonner, Cynthia L Darnell, Barbara E Engelhardt, and Amy K Schmid. Detecting differential growth of microbial populations with gaussian process regression. *Genome research*, 27(2):320–333, 2017.
- [33] Kate Tilling, Corrie Macdonald-Wallis, Debbie A Lawlor, Rachael A Hughes, and Laura D Howe. Modelling childhood growth using fractional polynomials and linear splines. *Annals of Nutrition and Metabolism*, 65(2-3):129–138, 2014.
- [34] James D Murray. Mathematical biology: I. an introduction. interdisciplinary applied mathematics. *Mathematical Biology*, Springer, 2002.
- [35] ÜNAL Deniz, Hacer YELDAN, GÜL Eylem, N Deniz ERGÜÇ, and Merve ADIYAN. Gompertz, logistic and brody functions to model the growth of fish species siganus rivulatus. *Acta Biologica Turcica*, 30(4):140–145, 2017.
- [36] Abdulsamad Engida Sado and Purnachandra Rao Koya. Application of brody growth function to describe dynamics of breast cancer cells. *American Journal of Applied Mathematics (AJAM)*, 3(3):138–145, 2015.
- [37] Michelle McDuffie. A hormone therapy model for breast cancer using linear cancer networks. *Rose-Hulman Undergraduate Mathematics Journal*, 15(1):9, 2014.
- [38] Dimpal Jyoti Mahanta and Munindra Borah. Parameter estimation of weibull growth models in forestry. *parameters*, 1:5, 2014.
- [39] Mark EJ Newman. Power laws, pareto distributions and zipf’s law. *Contemporary physics*, 46(5):323–351, 2005.
- [40] Michael L Johnson and Susan G Frasier. [16] nonlinear least-squares analysis. In *Methods in enzymology*, volume 117, pages 301–342. Elsevier, 1985.
- [41] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.
- [42] Åke Björck. *Numerical methods for least squares problems*. SIAM, 1996.
- [43] Philip Wolfe. Convergence conditions for ascent methods. *SIAM review*, 11(2):226–235, 1969.
- [44] Larry Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of mathematics*, 16(1):1–3, 1966.



- [45] Robert George Douglas Steel, James Hiram Torrie, et al. Principles and procedures of statistics. *Principles and procedures of statistics.*, 1960.
- [46] R Frank Falk and Nancy B Miller. *A primer for soft modeling*. University of Akron Press, 1992.
- [47] Jacob Cohen. Set correlation and contingency tables. *Applied psychological measurement*, 12(4):425–434, 1988.
- [48] Wynne W Chin et al. The partial least squares approach to structural equation modeling. *Modern methods for business research*, 295(2):295–336, 1998.
- [49] Joe F Hair, Christian M Ringle, and Marko Sarstedt. Pls-sem: Indeed a silver bullet. *Journal of Marketing theory and Practice*, 19(2):139–152, 2011.
- [50] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250, 2014.
- [51] Kenneth D West. Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084, 1996.
- [52] Anand Rajaraman and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge University Press, 2011.
- [53] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.
- [54] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- [55] Martin Hofmann. Support vector machines-kernels and the kernel trick. *Notes*, 26(3), 2006.
- [56] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [57] NS Altman. The american statistician. *An introduction to kernel and nearest-neighbor nonparametric regression*, 46(3):175–185, 1992.

- [58] Mohammad Hossin and MN Sulaiman. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5:1, 2015.
- [59] Kenji Kira and Larry A Rendell. A practical approach to feature selection. In *Machine Learning Proceedings 1992*, pages 249–256. Elsevier, 1992.
- [60] Igor Kononenko, Edvard Šimec, and Marko Robnik-Šikonja. Overcoming the myopia of inductive learning algorithms with relieff. *Applied Intelligence*, 7(1):39–55, 1997.
- [61] Nornadiah Mohd Razali, Yap Bee Wah, et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1):21–33, 2011.
- [62] Theodore W Anderson and Donald A Darling. Asymptotic theory of certain” goodness of fit” criteria based on stochastic processes. *The annals of mathematical statistics*, pages 193–212, 1952.
- [63] Carlos M Jarque and Anil K Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259, 1980.
- [64] Anil K Bera and Carlos M Jarque. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte carlo evidence. *Economics letters*, 7(4):313–318, 1981.
- [65] Carlos M Jarque and Anil K Bera. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, pages 163–172, 1987.
- [66] Hubert W Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318):399–402, 1967.
- [67] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.
- [68] Martin B Wilk and Ram Gnanadesikan. Probability plotting methods for the analysis for the analysis of data. *Biometrika*, 55(1):1–17, 1968.

- [69] Jiawei Han, Micheline Kamber, and Jian Pei. Data transformation and data discretization. *Data Mining-Concepts and Techniques*, ed: Morgan Kaufmann, pages 111–112, 2011.
- [70] Erwin Kreyszig. *Advanced engineering mathematics*, 10th eddition, 2009.
- [71] Paul T Von Hippel. Mean, median, and skew: Correcting a textbook rule. *Journal of Statistics Education*, 13(2), 2005.
- [72] Karl Pearson. “das fehlergesetz und seine verallgemeiner-ungen durch fechner und pearson.” a rejoinder. *Biometrika*, 4(1-2):169–212, 1905.
- [73] KO Bowman and LR Shenton. Estimation: Method of moments. *Wiley StatsRef: Statistics Reference Online*, 2014.

# ΠΑΡΑΡΤΗΜΑ Α

## ΠΑΡΑΡΤΗΜΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

---

<i>Growth Model</i>	<i>Parameter</i>	<i>Mean ± St.Deviation</i>	<i>Variation Coeff.</i>
Exponential	V <sub>0</sub>	275,5166 ± 284,4711	103,2%
	λ	0,0825 ± 0,1249	151,3%
Gompertz	V <sub>0</sub>	26,8351 ± 64,5531	240,5%
	A	0,2140 ± 0,1695	79,2%
	B	0,0187 ± 0,0065	34,9%
Logistic	A	0,0818 ± 0,0589	72,0%
	B	1,1732e-05 ± 1,2365e-05	105,3%
	V <sub>0</sub>	43,3531 ± 81,6618	188,3%
Bertalanffy	β <sub>1</sub>	802,6117 ± 1553,1112	193,5%
	β <sub>2</sub>	1,3164 ± 0,3178	24,1%
	β <sub>3</sub>	0,0076 ± 0,0065	85,6%
Gaussian	a	8094,2857 ± 2491,8508	30,7%
	b	147,4576 ± 34,9361	23,6%
	c	39,251 ± 23,6274	60,1%
Powerlaw	a	0,1961 ± 0,5682	289,7%
	b	9,4914 ± 14,2985	150,6%
Polynomial of 3rd Deg	p <sub>1</sub>	0,02682 ± 0,0238	89,0%
	p <sub>2</sub>	2,2383 ± 1,7489	78,1%
	p <sub>3</sub>	305,3182 ± 203,4961	66,6%
	p <sub>4</sub>	2198,5581 ± 1587,9915	72,2%

Πίνακας Α.1: Πίνακας Παραμέτρων των Αναπτυξιακών Μοντέλων στο Βενζοπυρένιο.

<i>Growth Model</i>	<i>Parameter</i>	<i>Mean <math>\pm</math> St.Deviation</i>	<i>Variation Coeff.</i>
Exponential	V <sub>0</sub>	398,7308 $\pm$ 620,9677	155,7%
	$\lambda$	0,03531 $\pm$ 0,0342	96,8%
Gompertz	V <sub>0</sub>	13,7324 $\pm$ 40,0835	291,8%
	A	0,3488 $\pm$ 0,3121	89,4%
	B	0,0176 $\pm$ 0,0077	43,8%
Logistic	A	0,0616 $\pm$ 0,0430	69,7%
	B	9,2125e-06 $\pm$ 8,4190e-06	91,3%
	V <sub>0</sub>	19,8205 $\pm$ 30,8752	155,7%
Bertalanffy	$\beta_1$	11462,5 $\pm$ 37986,4	331,3%
	$\beta_2$	1,3017 $\pm$ 0,4066	31,2%
	$\beta_3$	0,0050 $\pm$ 0,0065	131,0%
Gaussian	a	10037,2 $\pm$ 7380,3	73,5%
	b	243,7215 $\pm$ 153,5575	63,0%
	c	76,0012 $\pm$ 68,5288	90,1%
Powerlaw	a	9,8932 $\pm$ 18,8620	190,6%
	b	4,5250 $\pm$ 3,4618	76,5%
Polynomial of 3rd Deg	$p_1$	0,0120 $\pm$ 0,0124	102,8%
	$p_2$	0,5265 $\pm$ 0,1539	29,2%
	$p_3$	119,7280 $\pm$ 94,3533	78,8%
	$p_4$	512,2167 $\pm$ 317,2647	61,9%

Πίνακας Α.2: Πίνακας Παραμέτρων των Αναπτυξιακών Μοντέλων στις Πολυαμίνες.

<i>Growth Model</i>	<i>Parameter</i>	<i>Mean <math>\pm</math> St.Deviation</i>	<i>Variation Coeff.</i>
Exponential	Vo	222,0970 $\pm$ 159,6270	71,8%
	$\lambda$	0,0258 $\pm$ 0,0089	34,7%
Gompertz	Vo	47,0515 $\pm$ 95,8809	203,7%
	A	0,3600 $\pm$ 0,3282	91,1%
	B	0,0224 $\pm$ 0,0119	53,3%
Logistic	A	0,0535 $\pm$ 0,0317	59,2%
	B	1,2221e-05 $\pm$ 7,2591e-06	59,3%
	Vo	67,3746 $\pm$ 86,0346	127,6%
Bertalanffy	$\beta_1$	1266,6193 $\pm$ 3551,9702	280,4%
	$\beta_2$	1,4045 $\pm$ 0,3911	27,8%
	$\beta_3$	0,0096 $\pm$ 0,0086	89,8%
Gaussian	a	8,2038e+86 $\pm$ 6,3015e+87	768,1%
	b	581,8676 $\pm$ 1961,2187	337,0%
	c	115,2032 $\pm$ 148,3245	128,7%
Powerlaw	a	0,1554 $\pm$ 0,3100	199,5%
	b	2,7208 $\pm$ 0,7607	27,9%
Polynomial of 3rd Deg	$p_1$	0,0063 $\pm$ 0,0053	83,7%
	$p_2$	1,8655 $\pm$ 1,6637	89,1%
	$p_3$	54,4448 $\pm$ 38,1660	70,1%
	$p_4$	1541,2671 $\pm$ 1298,0473	84,2%

Πίνακας Α.3: Πίνακας Παραμέτρων των Αναπτυξιακών Μοντέλων στις Θειόλες.

<i>Growth Model</i>	<i>Toleranced SSE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	0	0	0	1	1	1	2	2	3
Gompertz	3	5	5	7	8	9	10	10	10
Logistic	8	11	11	12	13	13	13	15	15
Bertalanffy	0	0	0	1	1	2	2	2	2
Gaussian	1	1	1	1	1	2	2	3	3
Powerlaw	1	1	1	3	3	3	3	3	3
Polynomial 3rd Deg.	3	3	3	3	3	4	5	6	6

Πίνακας A.4: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced SSE στις Πολυαμίνες.

<i>Growth Model</i>	<i>Toleranced RMSE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	0	0	2	3	4	4	4	5	5
Gompertz	4	6	8	10	10	11	11	11	12
Logistic	8	9	13	15	16	16	16	16	16
Bertalanffy	0	0	1	2	2	4	5	5	5
Gaussian	1	1	1	3	4	4	4	5	6
Powerlaw	2	3	3	4	4	5	5	6	6
Polynomial 3rd Deg.	1	2	3	4	5	5	6	8	8

Πίνακας A.5: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced RMSE στις Πολυαμίνες.



<i>Growth Model</i>	<i>Toleranced MAE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	1	1	3	3	4	4	4	4	4
Gompertz	4	7	9	11	11	12	12	12	13
Logistic	6	8	10	12	14	14	14	14	14
Bertalanffy	0	0	0	2	3	4	5	5	5
Gaussian	1	1	2	2	2	3	3	3	4
Powerlaw	3	3	4	4	4	4	5	5	5
Polynomial 3rd Deg.	1	1	1	2	3	4	4	4	5

Πίνακας A.6: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced MAE στις Πολυαμίνες.

<i>Growth Model</i>	<i>Toleranced SSE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	0	6	9	11	13	13	13	14	16
Gompertz	6	15	28	35	37	42	43	44	44
Logistic	23	32	34	37	40	43	46	47	51
Bertalanffy	7	14	14	18	20	21	22	22	22
Gaussian	7	19	24	30	33	36	39	39	41
Powerlaw	0	2	5	9	12	15	16	17	20
Polynomial 3rd Deg.	16	24	30	35	39	43	47	48	48

Πίνακας A.7: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced SSE στις Θειόλες.

<i>Growth Model</i>	<i>Toleranced RMSE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	11	13	14	18	18	22	23	24	26
Gompertz	6	22	39	44	44	44	46	47	47
Logistic	21	32	42	47	56	58	58	59	59
Bertalanffy	9	15	21	23	25	26	27	32	34
Gaussian	8	23	33	39	44	48	52	55	55
Powerlaw	3	13	19	20	25	29	33	35	36
Polynomial 3rd Deg.	1	10	18	30	43	48	48	49	49

Πίνακας A.8: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced RMSE στις Θειόλες.

<i>Growth Model</i>	<i>Toleranced MAE</i>								
	0%	5%	10%	15%	20%	25%	30%	35%	40%
Exponential	13	14	17	18	21	23	24	26	26
Gompertz	4	17	30	42	44	44	46	47	47
Logistic	18	25	37	46	53	56	58	59	59
Bertalanffy	7	15	21	21	23	26	27	29	33
Gaussian	8	18	25	36	43	48	52	55	55
Powerlaw	8	15	22	25	30	33	35	37	39
Polynomial 3rd Deg.	1	1	5	13	21	31	37	43	47

Πίνακας A.9: Πίνακας Σύγκρισης Αναπτυξιακών μοντέλων βάση της Toleranced MAE στις Θειόλες.

<i>Transform</i>	<i>Normality Test</i>	<i>Double Dosage P-Value</i>	<i>Normality</i>
C. Logarithm	Anderson-Darling	0.0004	0
	Jarque-Bera	0.0035	0
	Lilliefors	0.0073	0
	Shapiro-Wilk	≪0.0001	0
Square Root	Anderson-Darling	0.0116	0
	Jarque-Bera	0.1577	1
	Lilliefors	0.0302	0
	Shapiro-Wilk	0.0035	0
Cubic Root	Anderson-Darling	0.0053	0
	Jarque-Bera	0.0762	1
	Lilliefors	0.0371	0
	Shapiro-Wilk	0.0012	0

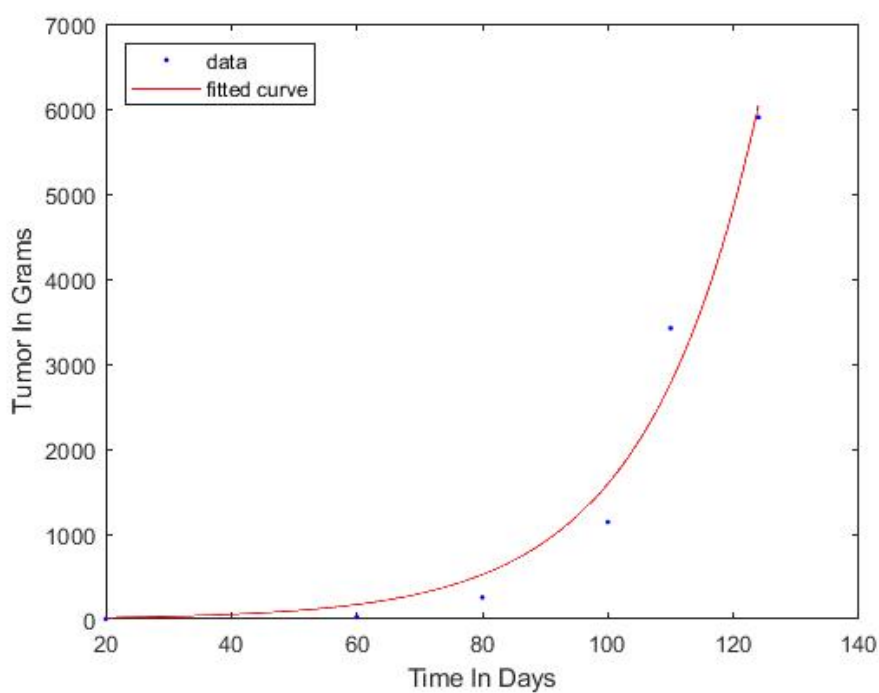
Πίνακας Α.10: Τεστ Ελέγχου Κανονικότητας Πολυαμινών Διπλών Δόσεων με Μετασχηματισμούς.

<i>Transform</i>	<i>Normality Test</i>	<i>Single Dosage P-Value</i>	<i>Normality</i>
C. Logarithm	Anderson-Darling	$\ll 0.0001$	0
	Jarque-Bera	0.0002	0
	Lilliefors	0.0010	0
	Shapiro-Wilk	$\ll 0.0001$	0
Square Root	Anderson-Darling	$\ll 0.0001$	0
	Jarque-Bera	0.0002	1
	Lilliefors	0.0010	0
	Shapiro-Wilk	$\ll 0.0001$	0
Cubic Root	Anderson-Darling	$\ll 0.0001$	0
	Jarque-Bera	0.0002	1
	Lilliefors	0.0010	0
	Shapiro-Wilk	$\ll 0.0001$	0

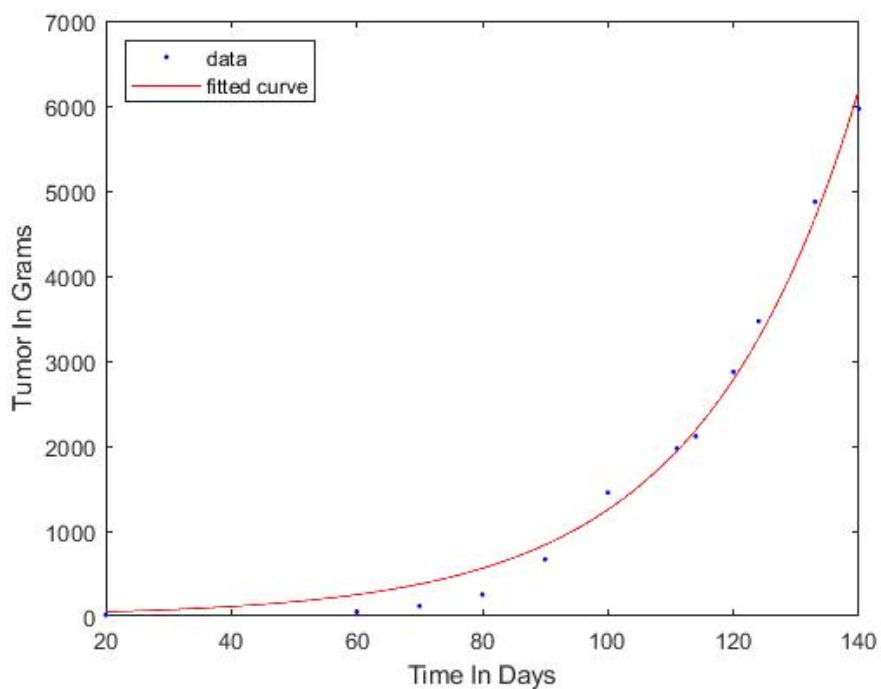
Πίνακας A.11: Τεστ Ελέγχου Κανονικότητας Θειολών Μονών Δόσεων με Μετασχηματισμούς.

# ΠΑΡΑΡΤΗΜΑ Β

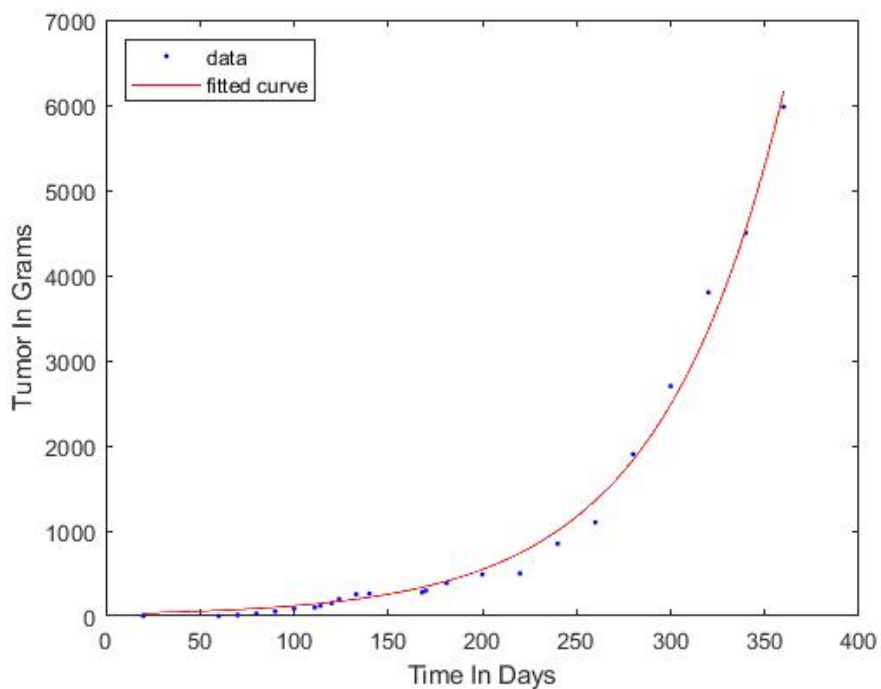
## ΠΑΡΑΡΤΗΜΑ ΣΧΗΜΑΤΩΝ ΠΡΟΣΑΡΜΟΓΗΣ ΜΟΝΤΕΛΩΝ



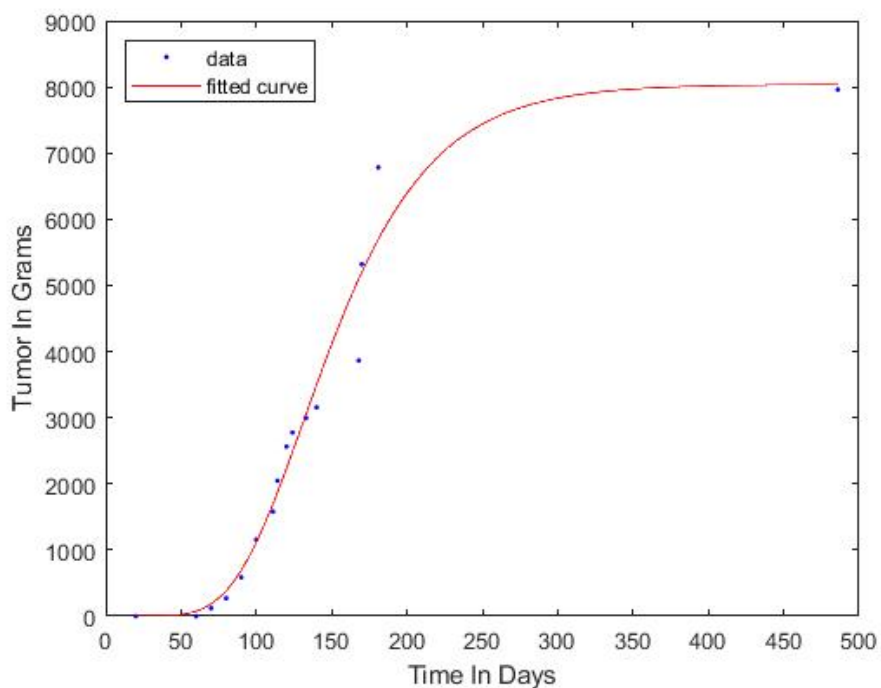
Σχήμα Β.1: Παράδειγμα προσαρμογής προτύπου με Βενζοπυρένιο στο μοντέλο Exponential



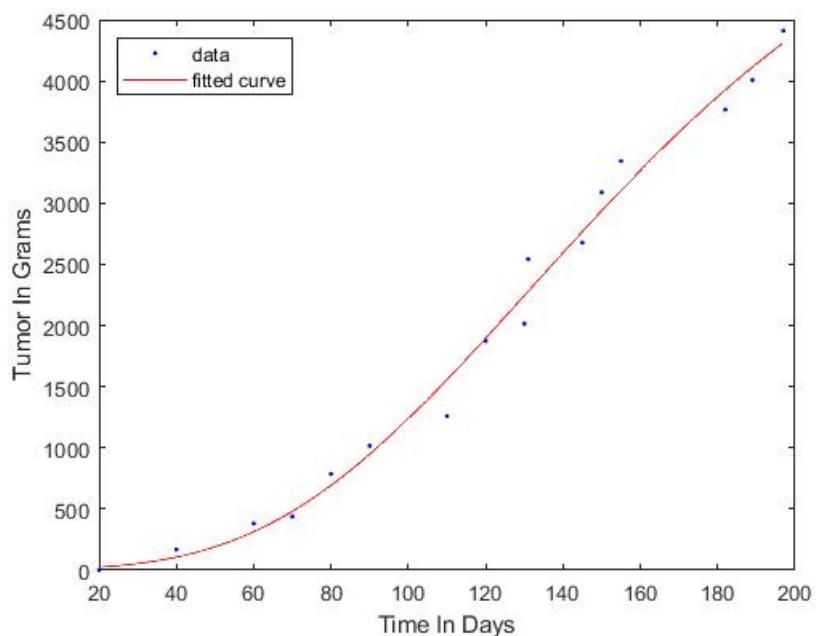
Σχήμα Β.2: Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Exponential



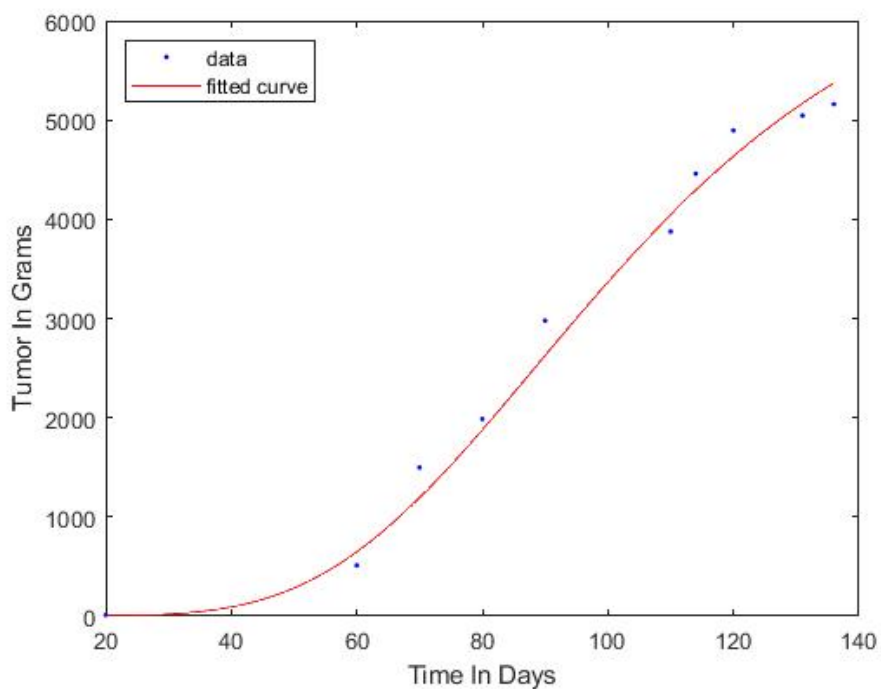
Σχήμα Β.3: Παράδειγμα προσαρμογής προτύπου με Πουτρεσκίνη στο μοντέλο Exponential



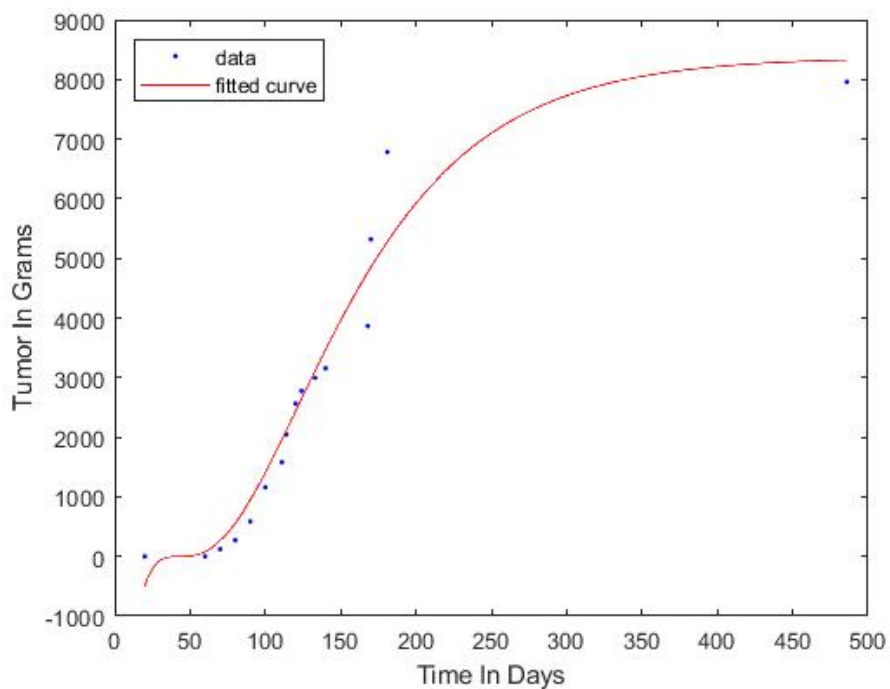
Σχήμα Β.4: Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Gompertz



Σχήμα Β.5: Παράδειγμα προσαρμογής προτύπου με Διθειοθρειτόλη στο μοντέλο Gompertz

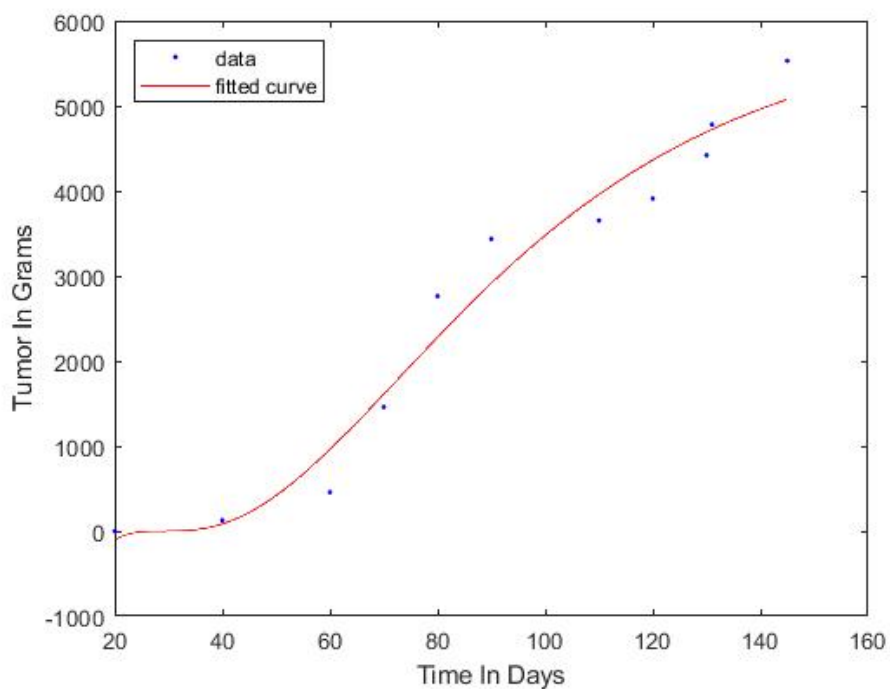


Σχήμα Β.6: Παράδειγμα προσαρμογής προτύπου με αιθυλικό εστέρα της Κυστεΐνης στο μοντέλο Gompertz

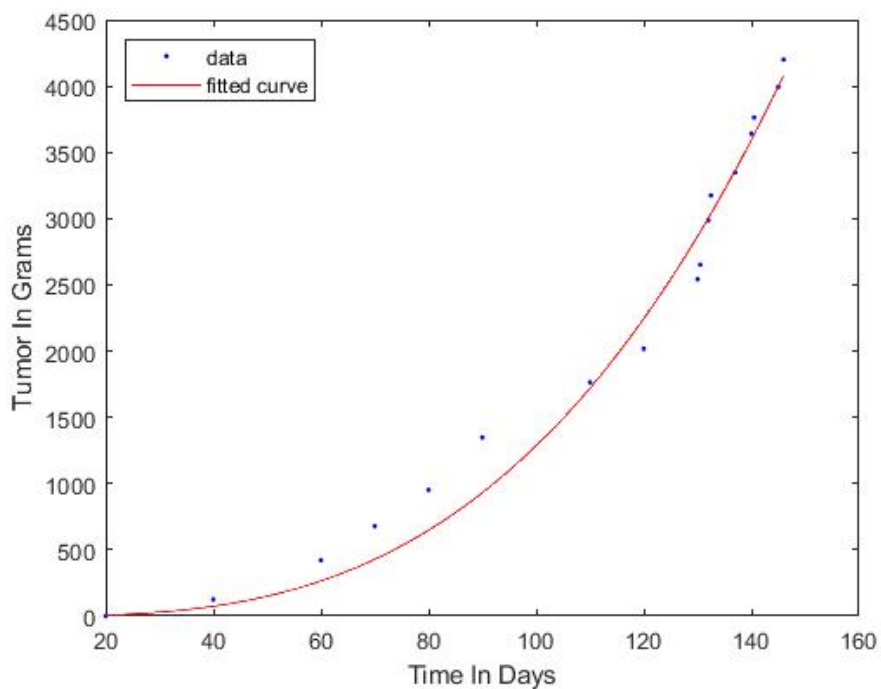


Σχήμα Β.7: Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Bertalanffy

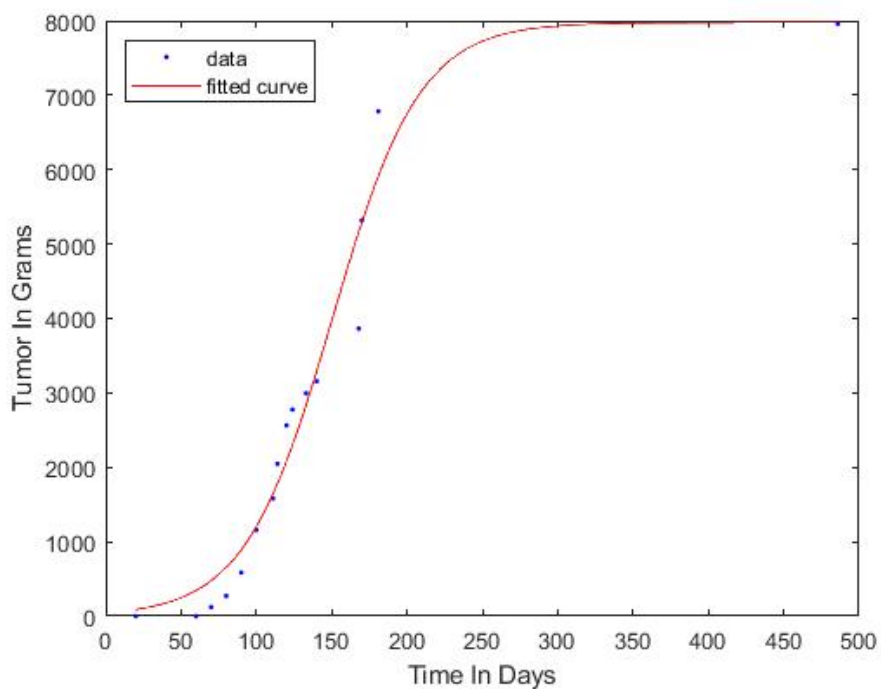




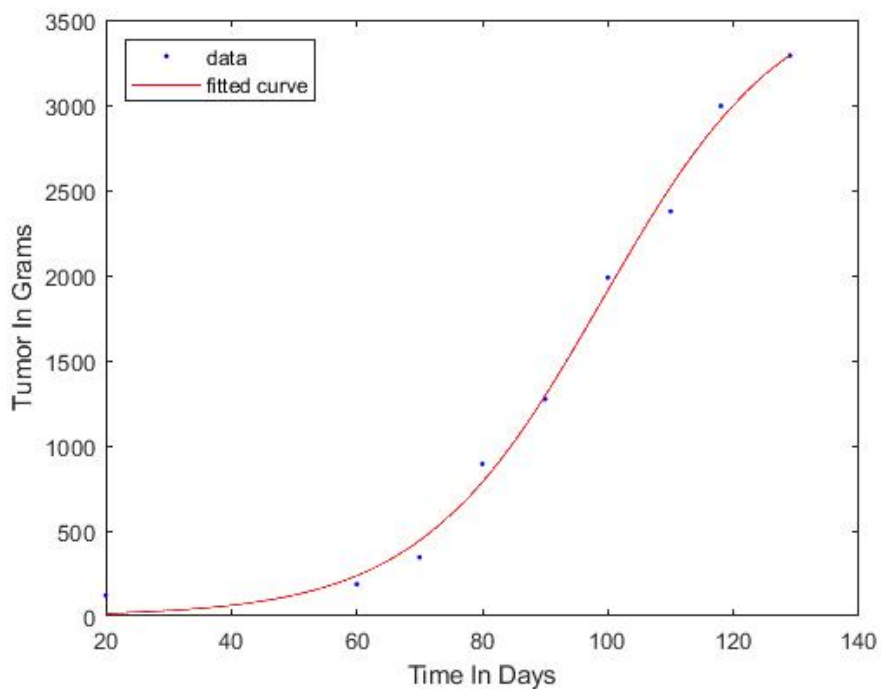
Σχήμα Β.8: Παράδειγμα προσαρμογής προτύπου με Διθειοθρεϊτόλη στο μοντέλο Bertalanffy



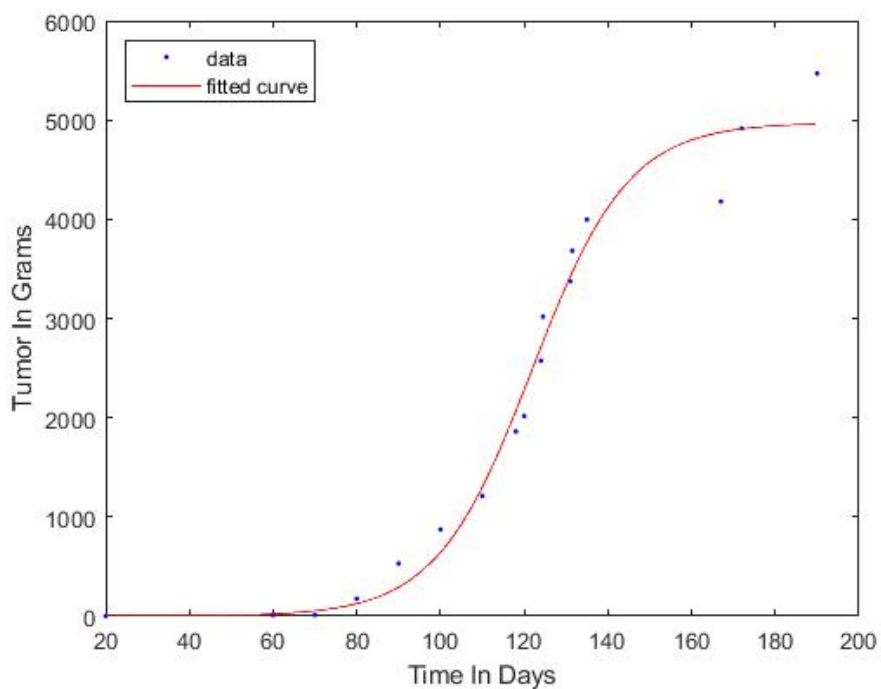
Σχήμα Β.9: Παράδειγμα προσαρμογής προτύπου με Κυστεΐνη στο μοντέλο Bertalanffy



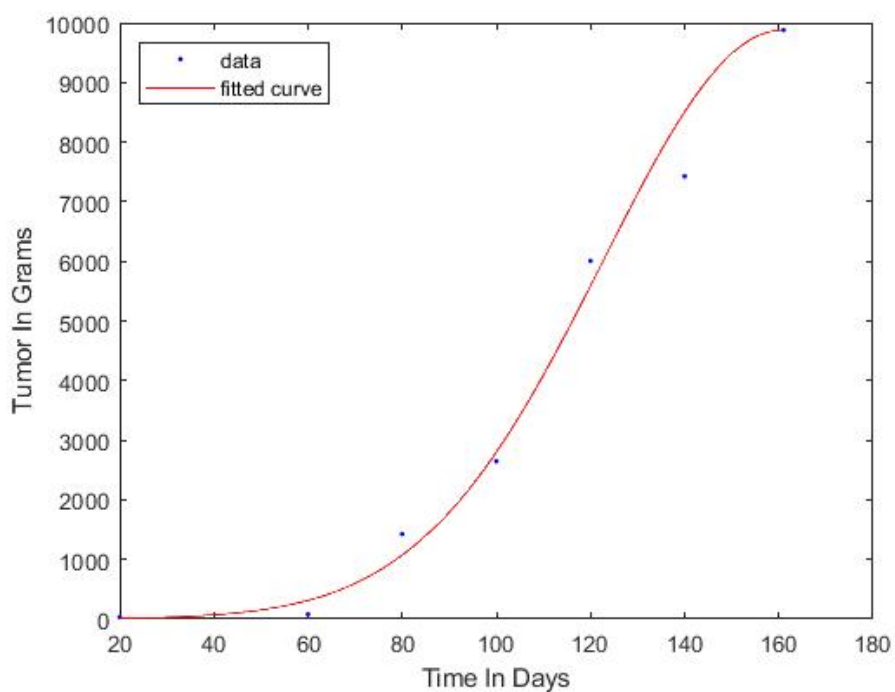
Σχήμα Β.10: Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο Logistic



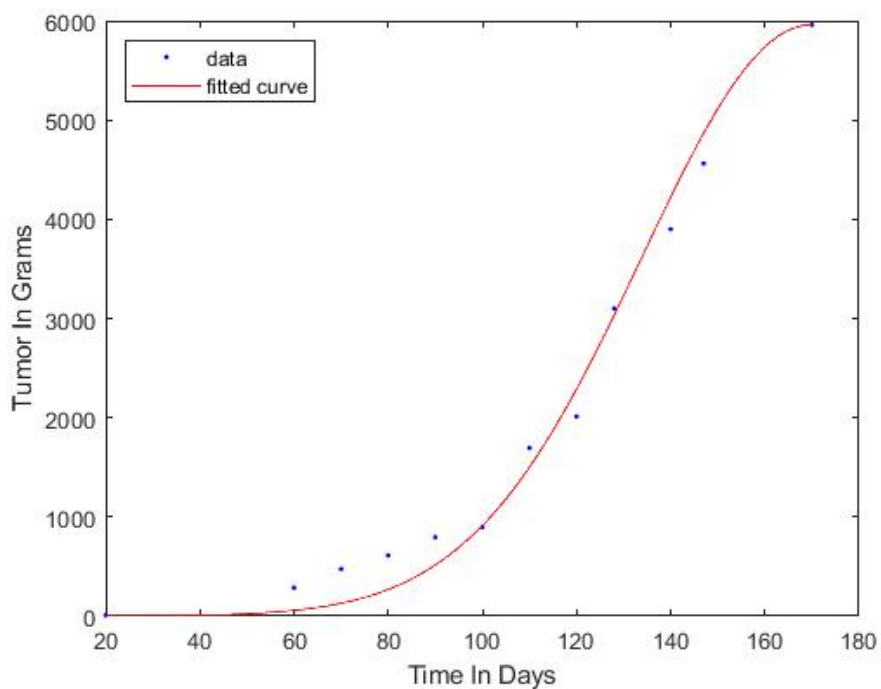
Σχήμα Β.11: Παράδειγμα προσαρμογής προτύπου με Βισμουθειόλη στο μοντέλο Logistic



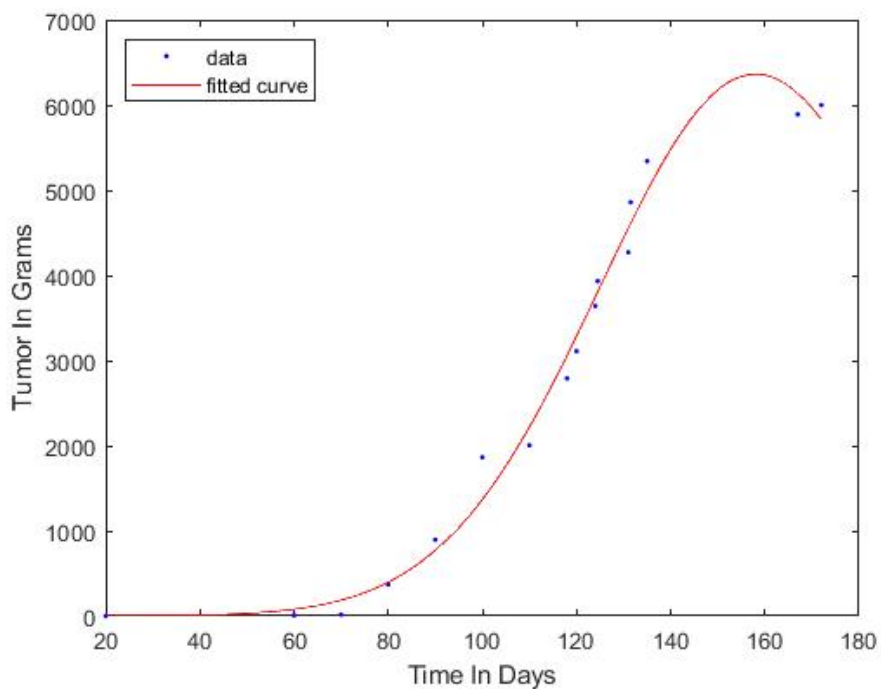
Σχήμα Β.12: Παράδειγμα προσαρμογής προτύπου με Μερκαπτοηλεκτρικό οξύ στο μοντέλο Logistic



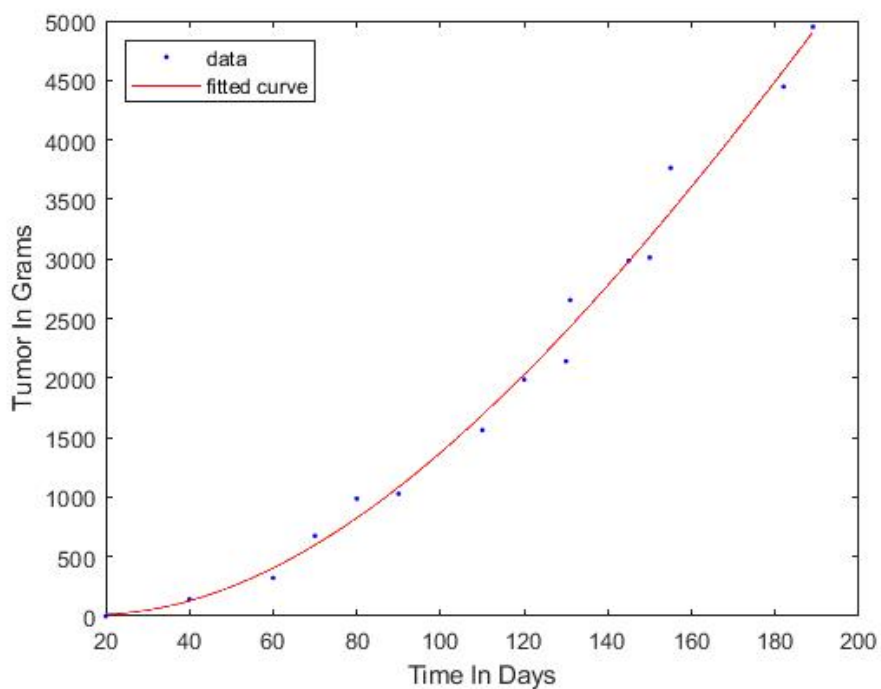
Σχήμα Β.13: Παράδειγμα προσαρμογής προτύπου με Βενζοπυρένιο στο μοντέλο Gauss



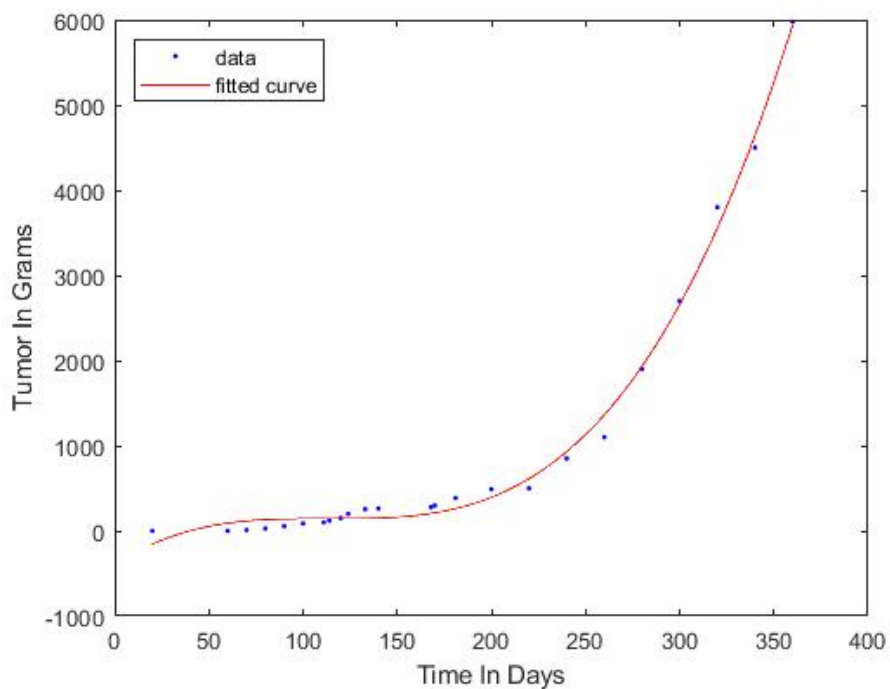
Σχήμα Β.14: Παράδειγμα προσαρμογής προτύπου με Αιθυλενοδιαμίνη στο μοντέλο Gauss



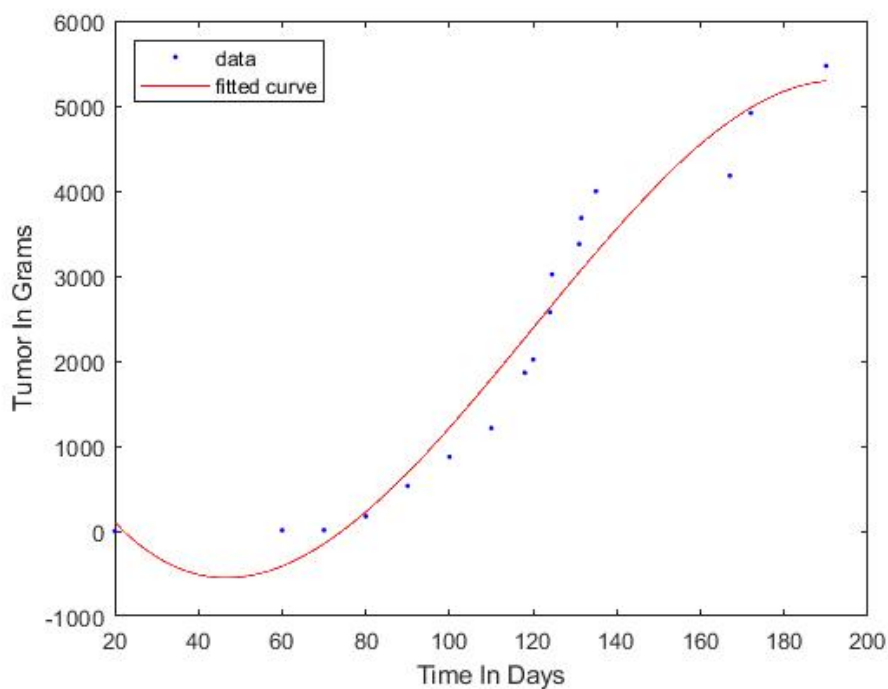
Σχήμα Β.15: Παράδειγμα προσαρμογής προτύπου με Μερκαπτοηλεκτρικό οξύ στο μοντέλο Gauss



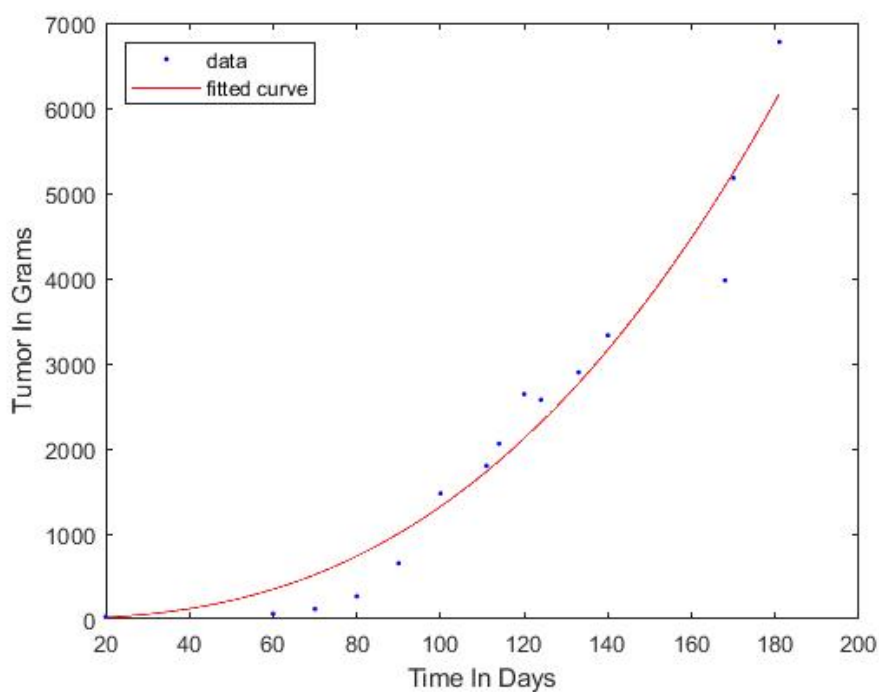
Σχήμα Β.16: Παράδειγμα προσαρμογής προτύπου με Διθειοθρεϊτόλη στο μοντέλο Polynomial



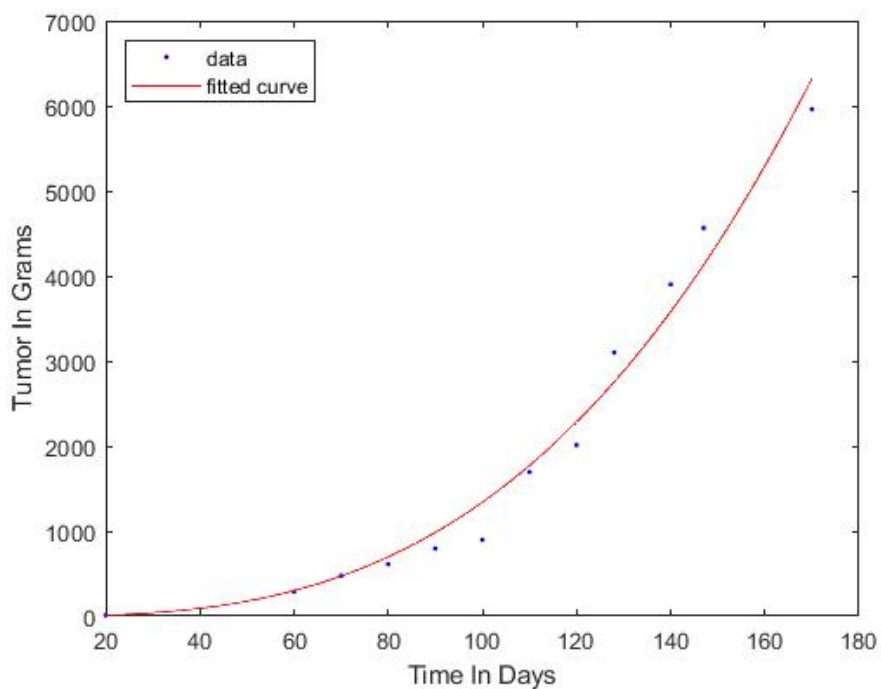
Σχήμα Β.17: Παράδειγμα προσαρμογής προτύπου με Πουτρεσκίνη στο μοντέλο Polynomial



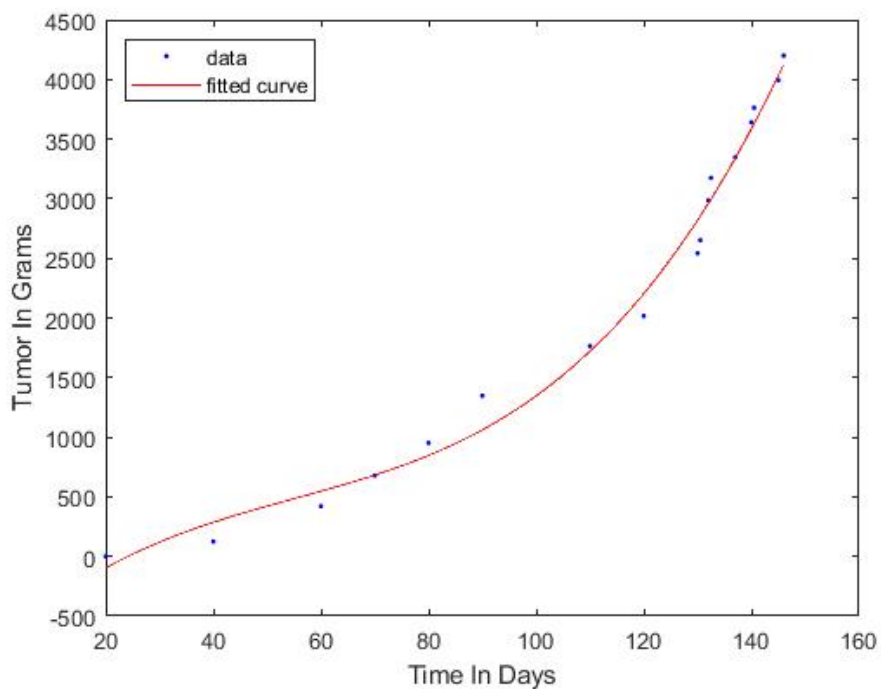
Σχήμα Β.18: Παράδειγμα προσαρμογής προτύπου με Μερκαπτοηλεκτρικό οξύ στο μοντέλο Polynomial



Σχήμα Β.19: Παράδειγμα προσαρμογής προτύπου με Κανταβερίνη στο μοντέλο PowerLaw



Σχήμα Β.20: Παράδειγμα προσαρμογής προτύπου με Αιθυλενοδιαμίνη στο μοντέλο PowerLaw



Σχήμα Β.21: Παράδειγμα προσαρμογής προτύπου με Κυστεΐνη στο μοντέλο PowerLaw

# ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

---

Ο Δημήτριος Μπουχάρας γεννήθηκε στις 22 Ιανουαρίου 1989 στην πόλη της Κοζάνης. Ολοκλήρωσε την βασική και μέση εκπαίδευσή του σε δημόσια σχολεία της πόλης της Κοζάνης. Κατόπιν, εισήχθη το 2007 μέσω Πανελληνίων Εξετάσεων στο τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων από το οποίο αποφοίτησε το 2015. Στην διάρκεια των σπουδών του ολοκλήρωσε την Κατεύθυνση Πληροφορικής με βαθμό 8,00. Στο επόμενο έτος ολοκλήρωσε τις στρατιωτικές του υποχρεώσεις σε παραμεθόριο περιοχή της χώρας. Από τον Οκτώβριο του 2017 εισήχθη κατόπιν συνεντεύξεως στο Πρόγραμμα Μεταπτυχιακών Σπουδών του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής και παρακολούθησε τις σχετικές παραδόσεις, όπως επίσης υποβλήθηκε με επιτυχία στις προβλεπόμενες εξετάσεις. Μέσα στην διαδικασία του Μεταπτυχιακού ανέλαβε ως θέμα "Μαθηματική μοντελοποίηση και ανάλυση αναπτυξιακών μοντέλων στην πρόληψη δημιουργίας καρκινικών κυττάρων". Τα ερευνητικά του ενδιαφέροντα εστιάζουν στο κομμάτι της έρευνας στην Ιατρική Πληροφορική και συνήθως αφορούν την ανάλυση δεδομένων και την μηχανική μάθηση.

- **Στοιχεία επικοινωνίας**

Διεύθυνση Κατοικίας: Ηρακλείου 10, Κοζάνη Κοζάνης, Ελλάδα. Τ.Κ.: 50131

Ηλεκτρονικό Ταχυδρομείο: DMpouxaras@cs.uoi.gr ή BoucharasJim@gmail.com

Κινητό Τηλέφωνο: +306981813683