Γιώργος Π. Στάμου

ΟΙΚΟΛΟΓΙΑ

Μέθοδοι Ανάλυσης και Σύνθεσης Δεδομένων

ΕΚΔΟΣΕΙΣ ΖΗΤΗ

Κάθε γνήσιο αντίτυπο φέρει την υπογραφή του συγγραφέα

ISBN 978-960-456-188-9

Φωτοστοιχειοθεσία
Εκτύπωση
Βιβλιοδεσία

Π. ΖΗΤΗ & Σια ΟΕ
18ο χλμ Θεσ/νίκης-Περαίας
Τ.Θ. 4171 • Περαία Θεσσαλονίκης • Τ.Κ. 570 19
Τηλ.: 2392.072.222 - Fax: 2392.072.229 • e-mail: info@ziti.gr

ΕΚΔΟΣΕΙΣ ΖΗΤΗ
www.ziti.gr

ΒΙΒΛΙΟΠΩΛΕΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ - ΚΕΝΤΡΙΚΗ ΔΙΑΘΕΣΗ:
Αρμενοπούλου 27, 546 35 Θεσσαλονίκη
Τηλ.: 2310.203.720, Fax: 2310.211.305 • e-mail: sales@ziti.gr

ΒΙΒΛΙΟΠΩΛΕΙΟ ΑΘΗΝΩΝ - ΕΝΩΣΗ ΕΚΔΟΤΩΝ ΒΙΒΛΙΟΥ ΘΕΣΣΑΛΟΝΙΚΗΣ:
Στοά του Βιβλίου (Πεσμαζόγλου 5), 105 64 Αθήνα • Τηλ.-Fax: 210.3211.097

ΑΠΟΘΗΚΗ ΑΘΗΝΩΝ - ΠΩΛΗΣΗ ΧΟΝΔΡΙΚΗ:
Ασκληπιού 60, 114 71 Αθήνα
Τηλ.-Fax: 210.3816.650 • e-mail: athina@ziti.gr

ΗΛΕΚΤΡΟΝΙΚΟ ΒΙΒΛΙΟΠΩΛΕΙΟ: www.ziti.gr

Πρόλογος

Η βασική θέση πάνω στην οποία βασίστηκε η συγγραφή τούτου του βιβλίου είναι ότι το επιστημονικό δεδομένο με βάση το οποίο θα εκτυλιχθεί η επιστημονική συζήτηση δεν είναι το δεδομένο της απλής εμπειρίας. Αντίθετα, το επιστημονικό δεδομένο θα παραχθεί στο πλαίσιο μιας θεωρίας και με βάση συγκεκριμένες μεθόδους. Με τις μεθόδους που χρησιμοποιούμε στο χώρο της βιολογίας και ειδικότερα στο πεδίο της οικολογίας ασχολείται τούτο το βιβλίο. Αποφάσισα να το γράψω γιατί τα τελευταία χρόνια διαπιστώνω να συμβαίνουν ραγδαίες αλλαγές στον τρόπο που προσδιορίζουμε το αντικείμενο της δουλειάς μας, που συλλέγουμε και αναλύουμε τα δεδομένα μας, που διαχειριζόμαστε τα συμπεράσματά μας. Ισχυρίζομαι μάλιστα ότι βρισκόμαστε στη φάση όπου ένα νέο μεθοδολογικό πρόγραμμα ξετυλίγεται αργόσυρτα μπροστά μας, με τα στοιχεία της εμπειρίας να υποχωρούν διαρκώς και τα στοιχεία των νοητικών κατασκευών να κερδίζουν έδαφος. Σε αυτό το πλαίσιο το βάρος της έρευνας μετατοπίζεται συνεχώς από τις καταγραφές πεδίου στα οργανωμένα πειράματα στο πεδίο και το εργαστήριο. Την ίδια στιγμή, πρωτογενή δεδομένα αποθηκευμένα σε απέραντες βάσεις δεδομένων μένει ακόμα να αναλυθούν. Νέες τεχνικές έρχονται στο προσκήνιο και παλιές εξελίσσονται και αναλαμβάνουν νέους ρόλους στο πλαίσιο της νέας μεθοδολογίας που πάει να εγκαθιδρυθεί. Ο στόχος μου λοιπόν ήταν να γράψω ένα βιβλίο όπου θα μιλάω για το παλιό και το καινούργιο κάτω από το πρίσμα αυτής της νέας μεθοδολογίας. Προπάντων θέλησα να γράψω ένα βιβλίο χρήσιμο κατά την καθημερινή πράξη.

Τούτο το βιβλίο ασχολείται με τις μεθόδους που εγώ και οι άνθρωποι που συνεργάστηκα μαζί τους χρησιμοποιούμε εδώ και τριάντα τόσα χρόνια. Πρόκειται με άλλα λόγια για ένα βιβλίο εμπειρίας και έτσι είναι γραμμένο. Τα περιεχόμενα του οργανώθηκαν σε στενή σύνδεση με τα ερωτήματα που έθετε κάθε φορά η πράξη. Για το λόγο αυτό, περισσότερο από κάθε άλλο βιβλίο τούτο δω τα χρωστάει όλα σχεδόν στους κατά καιρούς συνεργάτες. Σε αυτούς λοιπόν τους συνεργάτες θα ήθελα να το αφιερώσω. Όμως σε έναν ανάμεσά τους χρωστάω το πιο πολύ. Εννοώ το δάσκαλό μου, τον J.P. Cancela Da Fonseca που με εισήγαγε στη

στατιστική ανάλυση και μάλιστα σε μια εποχή που οι πιο πολλοί άνθρωποι του χώρου δεν υποπτεύονταν καν την ύπαρξή της. Σε αυτόν λοιπόν διαλέγω τελικά να το αφιερώσω. Πρόκειται για το δεύτερο βιβλίο που του αφιερώνω όμως είμαι της γνώμης ότι το χρέος στο δάσκαλο δεν ξεπληρώνεται ποτέ.

Χρωστάω στους κατά καιρούς ανθρώπους του Τομέα Οικολογίας, το Ν. Μάργαρη, την Ε. Φουσέκη, τη Ρ. Αριανούτσου, τον Γ. Διαμαντόπουλο, τη Δ. Βώκου, τη Μ. Πυροβέτση, το Γ. Παντή και τον Σ. Σγαρδέλη. Οι κουβέντες που είχα με όλους αυτούς συνέβαλαν στο ξεκαθάρισμα των ιδεών που περιλαμβάνονται στο βιβλίο.

Ύστερα έρχονται οι μαθητές/συνεργάτες χωρίς τη δουλειά των οποίων δεν μπορώ να φανταστώ τον εαυτό μου ως αυτό που είμαι σήμερα. Με τους πρώτους από αυτούς το Μ. Ασικίδης και το Γ. Ιατρού προβληματιστήκαμε πολύ πάνω στη μεθοδολογία και ανοίξαμε το δρόμο για αυτούς που ακολούθησαν. Με τους επόμενους, τη Μ. Αργυροπούλου και την Ε. Παπαθεοδώρου ο προβληματισμός πήγε μακριά και εξακολουθεί να πηγαίνει. Σε αυτές ανήκει το μέγιστο από τα περιεχόμενα του βιβλίου αλλά και από τα δεδομένα που αναλύονται εδώ.

Εκτός από τρία παραδείγματα που άντλησα από τη βιβλιογραφία και δύο που δανείστηκα από τους Θ. Καλλιμάνη και Κ. Τριάντη –και τους ευχαριστώ ιδιαίτερα γι’ αυτό– όλα τα άλλα δεδομένα που αναλύονται εδώ είναι προϊόν της πρόσφατης δραστηριότητας του Τομέα Οικολογία. Σε τούτο το βιβλίο συμπεριλήφθηκε μικρό μέρος από τα δεδομένα που έχει αποδώσει η δραστηριότητα του Τομέα. Όμως η επεξεργασία όλων - και αυτών που συμπεριέλαβα αλλά και εκείνων που έμειναν απέξω - συνεισέφερε στη διαμόρφωση των ιδεών που παρουσιάζω εδώ. Εκτός από τους παραπάνω μνημονευθέντες συνεργάτες, στη συλλογή και την επεξεργασία των δεδομένων αυτών είχαν συμμετοχή πολλοί άλλοι όπως ο Κ. Κορφιάτης, ο Σ. Παρασκευόπουλος, ο Θ. Μαρδίρης, ο Δ. Πατουχέας, ο Α. Ματθαίου, ο Σ. Πυρίντσος, η Α. Δαλάκα, ο Γ. Μπλιώνης, ο J. Bailey, ο Θ. Καλλιμάνης, ο Τ. Χοβαρδάς, ο Γ. Β. Στάμου, ο Ν. Μονοκρούσος, η Μ. Τσιαφούλη, η Α. Γιαννοτάκη, η Σ. Μιχαήλ, ο Γ. Μπούτσης, η Π. Καπαγιάννη, ο Κ. Τουλούμης, ο Δ. Σχίζας, η Α. Λευκαδίτη, ο Ν. Νικίσιανης, η Ε. Κάτρανα και άλλοι πολλοί που αυτή τη στιγμή διαφεύγει το όνομά τους. Όλους αυτούς θάθελα να ευχαριστήσω εκ βαθέων και να τους εξομολογηθώ ότι απόλαυσα τη συνεργασία με τον καθένα τους.

Γ. Στάμου

Αγ. Θεόδωροι, Σεπτέμβριος 2009

Περιεχόμενα

Μέρος Α Στατιστική ανάλυση

Μέρος Β Παραγωγή και διαχείριση δεδομένων

10. Πρωτόκολλα δειγματοληψίας

11. Μελέτη χωροδιανομής

12. Ανάλυση βιοκοινότητας

13. Αριθμητική ταξινόμηση

Μέρος Γ Εξόρυξη πληροφοριών

14. Νευρωνικά δίκτυα

15. Δένδρα απόφασης

16. Ανάλυση δικτύων

ΜΕΡΟΣ Α

Στατιστική Ανάλυση

1 Θεωρία και μέθοδος

1.1 Τα δομικά στοιχεία της επιστημονικής πρακτικής

Όπως συμβαίνει με το κάθε πεδίο της επιστήμης έτσι και η οικολογία διαθέτει το θεωρητικό της πυρήνα, έχει τις κατάλληλες μεθόδους συλλογής και ανάλυσης δεδομένων, καθώς και μεθόδους παραγωγής συμπερασμάτων. Επιπλέον, όπως ισχύει και για κάθε άλλη επιστημονική πρακτική, στόχος της οικολογικής έρευνας είναι η κατά το δυνατόν πληρέστερη γνώση του αντικειμένου της.

Το αντικείμενο της οικολογίας θα μπορούσε χονδρικά να οριστεί ως η μελέτη των σχέσεων που αναπτύσσονται εντός ενός πλαισίου που συγκροτούν διάφορα στοιχεία του βιολογικού και του φυσικού κόσμου και οι οποίες (σχέσεις) διασφαλίζουν τη διαιώνιση της ζωής.

Προδήλως, μιλώντας για διαιώνιση της ζωής είναι ωσάν να έχουμε ήδη εντάξει την οικολογία στο πλαίσιο που ορίζουν οι διάφορες εκδοχές της δαρβινισμού. Να υπενθυμίσω εδώ ότι όπως και πολλές άλλες θεωρητικές κατασκευές έτσι και ο δαρβινισμός, υπό οιανδήποτε εκδοχή και αν εκφέρεται, προϋποθέτει ότι:

1. Ανεξάρτητα από τη γνώση που έχουμε γι' αυτόν, ένας πραγματικός κόσμος υπάρχει εκεί έξω αντικειμενικά
2. Ο αντικειμενικός αυτός κόσμος είναι πειραματικά διαχειρίσιμος
3. Η γνώση του αντικειμενικού κόσμου είναι αποτέλεσμα γνωστικών διαδικασιών και είναι πάντα ατελής
4. Γνώση του αντικειμενικού κόσμου σημαίνει τη θεωρητική κατασκευή μιας εικόνας για το πώς περίπου είναι φτιαγμένος και το πώς λειτουργεί αυτός ο κόσμος

Ανάμεσα στα βασικά εργαλεία που χρησιμοποιούνται ώστε να γίνει γνωστός ο αντικειμενικός κόσμος εκεί έξω συγκαταλέγονται και οι έννοιες. Οι πιο αφηρη-

μένες από αυτές (π.χ. ενδοειδικός ανταγωνισμός, βιοχωρητικότητα) είναι πλούσιες σε θεωρητικό περιεχόμενο και συνδέονται με τη δομή της θεωρητικής κατασκευής που λέγαμε πιο πάνω. Αντίθετα, οι πιο συγκεκριμένες έννοιες (π.χ. γεννήσεις και θάνατοι) είναι πλουσιότερες σε εμπειρικό περιεχόμενο και γεφυρώνουν τον κόσμο της θεωρίας με εκείνον της εμπειρίας.

Τα στοιχεία που συγκροτούν το σώμα της επιστήμης, δηλαδή οι θεωρητικές κατασκευές, οι μέθοδοι και τα εργαλεία, αλλά και τα αντικείμενα της επιστήμης, συγκροτούν διαλεκτική ενότητα. Πάνω από όλα, αυτό θέλει να πει ότι όλα τα στοιχεία της επιστήμης (θεωρίες, εργαλεία, αντικείμενο) βρίσκονται σε μια ατελεύτητη διαδικασία συνεξέλιξης, πράγμα άλλωστε που διαπιστώνει και η καθημερινή εμπειρία. Έπειτα, σημαίνει ότι τα διάφορα στοιχεία για τα οποία έγινε λόγος πιο πάνω είναι μεταξύ τους αλληλεξαρτώμενα. Έτσι, επειδή ο δαρβινικός θεωρητικός πυρήνας αφορά –χονδρικά τουλάχιστον– στη διατήρηση της ζωής, η οικολογία θέτει ως αντικείμενο της τη γνώση των μηχανισμών και των προϋποθέσεων που διασφαλίζουν τη διατήρηση της ζωής στην αλληλεξάρτηση της με τον υπαρκτό φυσικό κόσμο. Προφανώς, το θεωρητικό ερώτημα που αφορά στη διατήρηση της ζωής έχει επηρεάσει αποφασιστικά το αντικείμενο της έρευνας που είναι η διερεύνηση των προϋποθέσεων ώστε αυτό να συμβεί.

Ακόμη, το ότι η επιστημονική διαδικασία ανελίσσεται διαλεκτικά σημαίνει ότι η γνώση που αποκτάται την κάθε φορά λειτουργεί αναδραστικά, γυρίζει δηλαδή πίσω, επιδρά και αναδιαμορφώνει το θεωρητικό πυρήνα. Το σπουδαίο είναι ότι τούτη η διαλεκτική διαδικασία συνεχίζεται αέναα, πράγμα που πάει να πει ότι η επιστήμη αποκαλύπτει πάντα μερικές όψεις του αντικειμενικού κόσμου εκεί έξω χωρίς ποτέ να μπορεί να φτάσει στην πλήρη γνώση, γι' αυτό άλλωστε και η προσπάθειά της δεν έχει τέλος.

Η αντικειμενικότητα της διαδικασίας διασφαλίζεται από το γεγονός ότι οι τελικές συγκεκριμένες έννοιες στις οποίες φτάνει η επιστήμη δεν είναι απλό παράγωγο της θεωρίας. Αντίθετα, παράγονται στο πλαίσιο μιας συνεχούς διαδικασίας σκέψης και αντιπαραβολής των στοιχείων της θεωρίας με εκείνα της εμπειρίας με στόχο τη συνεχή αναθεώρηση της θεωρίας, την εξέλιξη της μεθοδολογίας και τον επαναπροσδιορισμό του αντικειμένου.

Μια άλλη παραδοχή που γίνεται σε τούτο το βιβλίο είναι ότι οποιαδήποτε επιστημονική πρακτική, ακόμα και εκείνη που βρίσκεται περισσότερο κοντά στον κόσμο εκεί έξω (για παράδειγμα η παρατήρηση πελεκάνων στην Πρέσπα) πραγματοποιείται πάντοτε εντός της θεωρίας (είναι αυτό που λέμε theory laden). Αυτό θα πει ότι, σε αντίθεση με τον οίκο-τουρίστα της Πρέσπας που παρατηρεί ότι πέφτει στην αντίληψη του, ο επιστήμονας θα παρατηρήσει και θα καταγράψει συγκεκριμένα πράγματα, τέτοια που να έχουν νόημα στο πλαίσιο της δαρβινικής θεωρίας.

Ας υποθέσουμε προς στιγμήν ότι μας ενδιαφέρει να μελετήσουμε την ηθολογία του πελεκάνου. Μας ενδιαφέρει συγκεκριμένα να μάθουμε πότε και πώς πετάνε τα πουλιά στον αέρα, πότε και πώς ψαρεύουν, πότε και πώς ταΐζουν τα μικρά τους κ.λ.π. Ως βιολόγους όλα αυτά μας ενδιαφέρουν επειδή αποτελούν στοιχεία της απάντησης που θα δοθεί στη βασική ερώτηση που θέτει η οικολογική θεωρία και αφορά στους τρόπους με τους οποίους ο πελεκάνος διασφαλίζει την επιβίωσή του στην Πρέσπα. Είναι με άλλα λόγια ερωτήσεις που πηγάζουν από τις ανάγκες της θεωρίας.

Να παρατηρηθεί εδώ ότι ακόμα και η μέθοδος, αλλά και τα εργαλεία που θα χρησιμοποιηθούν σχετίζονται με το θεωρητικό ερώτημα. Ειδικά σε ότι αφορά τα εργαλεία να σημειωθεί ότι η χρήση τους απαιτεί πρόσθετη θεωρητική γνώση από άλλα πεδία της επιστήμης. Για παράδειγμα, η χρησιμοποίηση του τηλεσκοπίου φωτογράφησης των πουλιών προϋποθέτει σχετική γνώση από το πεδίο της οπτικής ώστε να γίνουν οι απαραίτητες ρυθμίσεις του οργάνου (φωτισμοί, γωνίες, φίλτρα κ.λ.π.). Αυτό γιατί μόνο τότε μπορεί κανείς να σταθμίσει το τί ακριβώς θα φωτογραφίσει, αλλά και το πώς θα ερμηνεύσει τη φωτογραφία που έλαβε ώστε να απαντήσει στο ερώτημα που θέτει η οικολογική θεωρία.

Μια άλλη σημαντική παραδοχή εδώ είναι ότι οι θεωρητικές έννοιες είναι ιδεολογικά φορτισμένες. Θεωρείται βέβαιο, για παράδειγμα, ότι η έννοια της προσαρμογής που χρησιμοποιεί ο Δαρβίνος είναι θεολογικό δάνειο. Γενικότερα, η βιολογία χρησιμοποιεί πληθώρα εννοιών, όπως η εξέλιξη μέσω του ανταγωνισμού, η ισορροπία, η παραγωγικότητα κ.ά. με πρόδηλο ή λανθάνον ιδεολογικό φορτίο.

Στόχος της επιστημονικής εργασίας είναι η αποκάθαρση των εννοιών από το ιδεολογικό τους φορτίο. Μάλιστα, θεωρείται ότι η νέα δαρβινική σύνθεση υπήρξε το αποτέλεσμα της προσπάθειας να αποκαθαρθεί το σώμα της βιολογίας από ιδεολογικά στοιχεία όπως η αυτόματη γένεση, ο βιταλισμός κ.λ.π. που απομείωναν την αξιοπιστία των πορισμάτων της επιστήμης. Ωστόσο, αυτή η διαδικασία κάθαρσης είναι αλυσιτελής καθόσον το λανθάνον ιδεολογικό φορτίο των εννοιών ανανεώνεται συνεχώς.

1.2 Θεωρίες και μοντέλα

1.2.1. Τύποι θεωρίας

Διακρίνονται δύο τύποι θεωρητικής εργασίας που απολήγουν το ένα στη διατύπωση γενικών και το άλλο στη διατύπωση ειδικών θεωριών. Συγκεκριμένα οι γενικές θεωρίες είναι κατασκευές που περιγράφουν σε γενικές γραμμές το πώς εί-

ναι φτιαγμένο και το πώς λειτουργεί το σύμπαν που αποτελεί το αντικείμενο μελέτης μιας συγκεκριμένης επιστήμης. Για παράδειγμα, η γενική θεωρία της βιολογίας ταυτοποιείται με τη Θεωρία της Δαρβινικής Εξέλιξης. Πράγματι, στο πλαίσιο αυτής της θεωρίας θα βρει την ερμηνεία της το κάθε στοιχείο που αποκαλύπτει η έρευνα. Επίσης στο δαρβινικό πλαίσιο θα βρει το τελικό της νόημα η κάθε πρόταση που γίνεται στο όνομα της επιστήμης της βιολογίας.

Στις γενικές θεωρίες υπόκεινται οι ειδικές θεωρίες. Πρόκειται για θεωρητικές κατασκευές που διατυπώνονται εντός των ορίων της γενικής θεωρίας την οποία και εξειδικεύουν σε έναν ορισμένο χώρο του πεδίου. Για παράδειγμα, η θεωρία αριστοποίησης της τροφικής συμπεριφοράς εξειδικεύει τη δαρβινική κατασκευή στο επίπεδο των τροφικών σχέσεων, ενώ η ειδική θεωρία που οδηγεί στη διατύπωση στρατηγικών ζωής εξειδικεύει το δαρβινισμό στο πεδίο της δημογραφίας. Επιπρόσθετα, οι συνοδοί θεωρίες αποτελούν δάνεια από άλλες επιστήμες που βρίσκουν χρησιμότητα κυρίως στην πειραματική εργασία. Θέση συνοδού υπέχουν για παράδειγμα τα στοιχεία από τη θεωρία της φυσικής που χρειάζεται ο φυσιολόγος ώστε να ερμηνεύσει το γράφημα που θα του δώσει ο παλμογράφος.

1.2.2. Τύποι μοντέλων

Τα μοντέλα είναι εργαλεία που στέκουν ανάμεσα στη θεωρία και τον πραγματικό κόσμο εκεί έξω και αναλαμβάνουν το καθήκον να γεφυρώσουν τον κόσμο της θεωρίας με τον κόσμο της εμπειρίας. Στην πράξη η αντιστοίχηση αυτή πραγματοποιείται βαθμιαία με τη διαδοχική διαμεσολάβηση διαφόρων μοντέλων. Συγκεκριμένα διακρίνουμε τις παρακάτω κατηγορίες μοντέλων:

► **Μοντέλα-τύποι ή θεωρητικά σχήματα**: Βρίσκονται κοντύτερα στη θεωρία και περιγράφουν τις συνθήκες κάτω από τις οποίες μπορεί να υπάρξει ένας από όλους τους δυνατούς κόσμους που είναι ικανή να παράγει η θεωρία. Για παράδειγμα το μοντέλο Lotka-Voltera

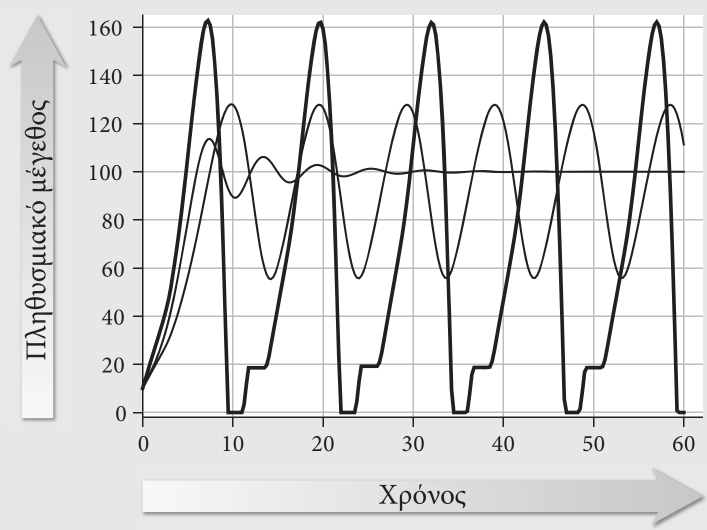
$$1 = \int l_x m_x e^{-rx}$$

όπου x = ηλικιακή κλάση, l_x = βιωσιμότητα, m_x = γεννητικότητα και r = ενδογενής ρυθμός αύξησης, εξειδικεύει τη θεωρία των αντισταθμίσεων στο χώρο της δημογραφίας. Μας λέει συγκεκριμένα ότι, επειδή η τιμή του ολοκληρώματος είναι πάντα σταθερή, ίση με τη μονάδα, τότε η οποιαδήποτε αύξηση στην τιμή της μιας δημογραφικής παραμέτρου (ας πούμε της γεννητικότητας) θα πρέπει να αντισταθμιστεί από κατάλληλου εύρους μείωση στην τιμή της άλλης παραμέτρου (εδώ της βιωσιμότητας).

► **Μοντέλα διερεύνησης**: Συνήθως πρόκειται για παραμετρικά εργαλεία που χρησιμοποιούνται για τη διερεύνηση της σημασίας που έχουν διάφοροι παράμετροι για τη συμπεριφορά του συστήματος που περιγράφουν. Χρησιμοποιούνται ακόμα και για τη διερεύνηση εναλλακτικών υποθέσεων. Για παράδειγμα το μοντέλο

$$\frac{dN_t}{dt} = r\left(1 - \frac{N_{t-\varphi}}{K}\right)$$

περιγράφει τη δυναμική πληθυσμών όταν η πληθυσμιακή ρύθμιση γίνεται με χρονική καθυστέρηση φ. Η προσομοίωση στον υπολογιστή αυτού του μοντέλου για διάφορες τιμές των παραμέτρων r, φ και K επιτρέπει τη διερεύνηση υποθέσεων σχετικά με την κυκλικότητα των φαινόμενων και την ισορροπία, τη χαοτική τους συμπεριφορά κ.λ.π. Έτσι, ανάλογα με την τιμή της παραμέτρου φ στο Διάγραμμα 1.1 παρατηρούμε τη δυναμική του πληθυσμού να εμφανίζει άλλοτε συγκλίνουσες, άλλοτε σταθερές και άλλοτε πολύπλοκες ταλαντώσεις.

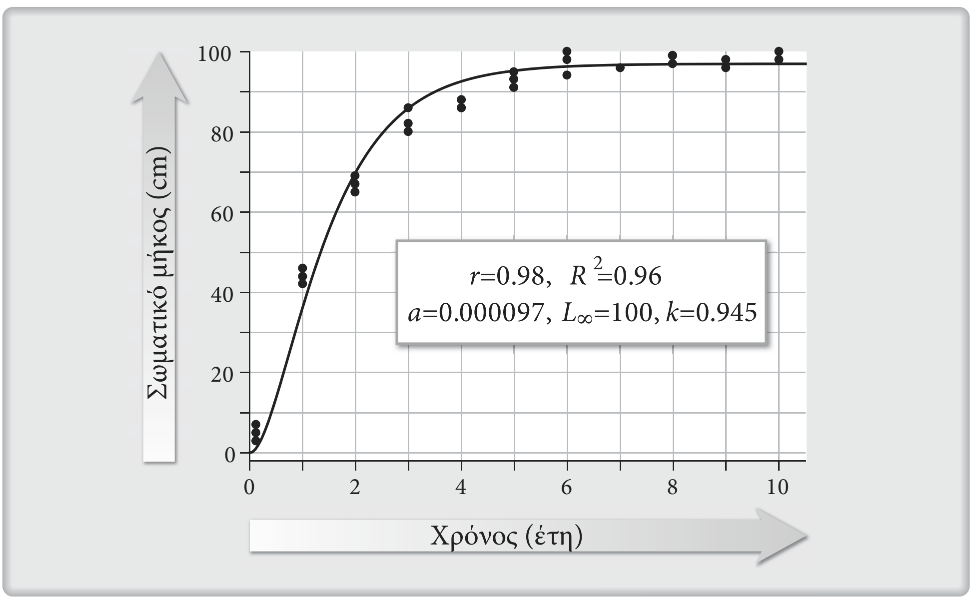


Διάγραμμα 1.1. Προσομοίωση μοντέλου χρονικής υστέρησης για διάφορες τιμές των παραμέτρων του.

► **Μοντέλα προσομοίωσης δεδομένων ή εμπειρικά μοντέλα**: Είναι αυτά που βρίσκονται κοντύτερα στα πειραματικά δεδομένα και χρησιμοποιούνται προκειμένου να τα περιγράψουν. Ένα τέτοιο μοντέλο είναι εκείνο του von Bertalanfy:

$$L_t = aL_{\infty}^{3}\left(1 - e^{-k(t-t_0)}\right)^2$$

το οποίο προσαρμόστηκε σε δεδομένα αύξησης του μήκους ενός ψαριού με το πέρασμα του χρόνου. Οι συμβολισμοί έχουν ως εξής: L_t είναι το μήκος, t είναι ο χρόνος, L_∞ είναι το μέγιστο μήκος που ασυμπτωτικά θα φτάσει το ψάρι και α και k είναι παράμετροι. Το μοντέλο προσαρμόστηκε σε αλιευτικά δεδομένα ψαριού που ζει 11 έτη και το τελικό του μήκος αγγίζει τα 100 cm. Στο Διάγραμμα 1.2. απεικονίζονται με συμπαγείς κύκλους τα πρωτογενή δεδομένα. Στο ίδιο διάγραμμα η συνεχής γραμμή αντιστοιχεί στις τιμές του μοντέλου. Στο γράφημα παρατίθενται επίσης οι τιμές των παραμέτρων a και k. Παρατίθενται ακόμα και οι τιμές των συντελεστών r και R^2. Ο πρώτος συντελεστής παίρνει τιμές από -1 ως 1 και μετρά τη συσχέτιση που υφίσταται ανάμεσα στην ανεξάρτητη μεταβλητή χρόνος και την εξαρτημένη μήκος. Όπως φαίνεται από την τιμή $r = 0.98$ η συσχέτιση ανάμεσα στις δύο μεταβλητές είναι θετική και πολύ ψηλή. Τέλος η παράμετρος R^2 μετρά το ποσοστό της διακύμανσης των δεδομένων (εδώ 96%) που αποδίδει το μοντέλο που προσαρμόστηκε στα δεδομένα.



Διάγραμμα 1.2. Προσαρμογή του εμπειρικού μοντέλου von Bertalanfy σε δεδομένα μεταβολής του βάρους ενός ψαριού με το πέρασμα του χρόνου.

Με το παραπάνω παράδειγμα επιχειρήθηκε να δειχθεί ότι το πρωτογενές πειραματικό δεδομένο (εδώ το μήκος του ψαριού) σπάνια αποτελεί και το επιστημονικό δεδομένο. Συνήθως ο επιστήμονας θα επεξεργαστεί κατάλληλα το αρχικό

δεδομένο και θα παράγει νέα δεδομένα (εδώ τις τιμές των παραμέτρων a και k, καθώς και τις τιμές των εκτιμητών r και R^2) με βάση τις οποίες θα γίνει και η επιστημονική συζήτηση.

► **Πειραματικά σχέδια**: Πρόκειται για σχέδια με βάση τα οποία θα πραγματοποιηθεί η πειραματική εργασία. Στον Πίνακα 1.1. παρατίθεται ένα τέτοιο σχέδιο που αφορά στην επίδραση επί της συλλεγόμενης βιομάζας ενός καλλιεργούμενου φυτού τριών διαφορετικών μορφών λίπανσης, επιπέδων άρδευσης και τύπων φυτοπροστασίας.

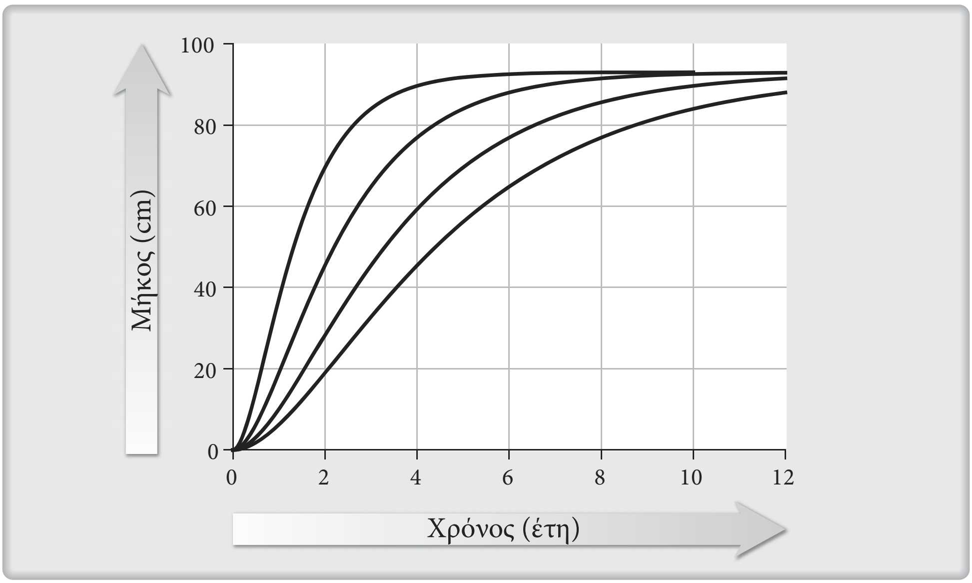
Πίνακας 1.1. *Πειραματικό σχέδιο μελέτης της επίδρασης τριών διαφορετικών καλλιεργητικών πρακτικών επί της παραγωγικότητας μιας καλλιέργειας.*

πειραματικές επιφάνειες	λίπανση	άρδευση	φυτοπροστασία
1	χημική	χαμηλή	χημική
2	χημική	κανονική	βιολογική
3	χημική	αυξημένη	καμία
4	κομπόστ	χαμηλή	καμία
5	κομπόστ	κανονική	χημική
6	κομπόστ	αυξημένη	βιολογική
7	χλωρή	χαμηλή	βιολογική
8	χλωρή	κανονική	καμία
9	χλωρή	αυξημένη	χημική

Το πειραματικό σχέδιο που υιοθετήθηκε είναι ένα λατινικό τετράγωνο και προβλέπει δειγματοληψίες σε 9 πειραματικές επιφάνειες. Στην πρώτη από αυτές εφαρμόζεται χημική λίπανση, χαμηλή άρδευση και χημική φυτοπροστασία. Αντίστοιχα, η έκτη επιφάνεια λιπαίνεται με κομπόστ, δέχεται αυξημένα ποσά νερού και εφαρμόζεται σε αυτήν βιολογική φυτοπροστασία. Να σημειωθεί επίσης ότι, όπως θα δούμε και αργότερα, εκτός από την οργάνωση των πειραματικών επιφανειών το σχέδιο προδιαγράφει το είδος των επιδράσεων που είναι δυνατόν να μελετηθούν, τα φύλλα καταχώρησης των δεδομένων καθώς και τη μέθοδο ανάλυσης των αποτελεσμάτων.

Ανάλογα με το σκοπό που καλούνται να εξυπηρετήσουν, τα διάφορα μοντέλα μπορούν να εκτελέσουν και διαφορετικές λειτουργίες. Στην προηγούμενη παράγραφο συζητήθηκε η χρήση του μοντέλου von Bertalanfy ως εργαλείου περι-

γραφής αρχικών δεδομένων. Το ίδιο μοντέλο ωστόσο μπορεί να χρησιμοποιηθεί και ως εργαλείο διερεύνησης. Συγκεκριμένα οι τιμές σωματικού μήκους με το πέρασμα του χρόνου προσομοιώθηκαν για τέσσερις διαφορετικές τιμές της παραμέτρου k (1, 0.6, 0.4 και 0.3), ενώ οι τιμές των άλλων παραμέτρων παρέμεναν σταθερές. Τα αντίστοιχα αποτελέσματα απεικονίζονται με τη σειρά από αριστερά προς τα δεξιά με τη βοήθεια των καμπύλων του Διαγράμματος 1.3. Από την εξέταση του διαγράμματος προκύπτει ότι η παράμετρος k μπορεί να ταυτοποιηθεί με το ρυθμό αύξησης του σωματικού βάρους (ταχύτερη αύξηση για $k=1$ και βραδύτερη για $k=0.3$).



Διάγραμμα 1.3. Χρησιμοποίηση του μοντέλου von Bertalanfy ως εργαλείου διερεύνησης. Εδώ έγινε προσομοίωση του μοντέλου για 4 διαφορετικές τιμές της παραμέτρου *k*.

Εκτός από εργαλείο προσομοίωσης και διερεύνησης το ίδιο μοντέλο μπορεί να χρησιμοποιηθεί και ως μοντέλο-τύπος ή θεωρητικό σχήμα που υλοποιεί τη θεωρία της αλλομετρίας στο επίπεδο του σωματικού μήκους. Ορίζει λοιπόν το μοντέλο-τύπος ότι η αύξηση του μήκους σχετίζεται με την ηλικία μόνο που η συσχέτιση αυτή δεν είναι γραμμική. Πράγματι όταν τα ψάρια βρίσκονται σε νεαρή ηλικία τότε από τη μια χρονιά στην άλλη το μήκος αυξάνει πολύ. Αντίθετα, σε μεγαλύτερες ηλικίες αντιστοιχούν όλο και μικρότερες αυξήσεις στο μήκος, ώσπου από μια ηλικία και μετά το μήκος παραμένει ασυμπτωτικά σταθερό. Προκειμένου να περιγραφούν αυτού του είδους τα φαινόμενα χρησιμοποιείται

ο όρος αλλομετρική αύξηση ώστε να εννοηθεί ότι η συμμεταβολή δύο μεταβλητών (π.χ. μήκος-βάρος, μήκος-πλάτος, μήκος-χρόνος κ.ά.) δεν υπακούει σε σταθερή σχέση.

Τα στατιστικά μοντέλα που υπόκεινται των διαδικασιών στατιστικής εκτίμησης της τιμής διαφόρων μεγεθών, ελέγχου υποθέσεων και εν γένει κάθε διαδικασίας που εμπλέκει στατιστικά μεγέθη εμπίπτουν στην κατηγορία των μοντέλων προσομοίωσης.

1.3 Θεμελιώδη ερωτήματα

Τα ερωτήματα στα οποία καλείται συνήθως να απαντήσει η οικολογική έρευνα μπορούν να συνοψιστούν ως εξής:

1. τί καθορίζει την αφθονία ενός είδους στο πεδίο και πώς θα εξελιχθεί αυτό το μέγεθος στο πέρασμα του χρόνου
2. πόσο σημαντικές για την αφθονία τους είναι οι σχέσεις ανάμεσα στα είδη (ανταγωνισμός, θήρευση, σχέσεις συνεργασίας). Ποια είναι σχετική σημασία κάθε μιας από αυτές για την ανέλιξη των φαινόμενων
3. έχουν οι διαειδικές σχέσεις σημασία για ευρύτερα επίπεδα οργάνωσης πέρα από τις τροφικές ομάδες (guilds)
4. με ποιους τρόπους διαμεσολαβούν οι σχέσεις αυτές τις ροές ύλης και ενέργειας εντός των οικολογικών σχηματισμών, αλλά και μεταξύ διαφορετικών οικολογικών σχηματισμών

Όλες οι παραπάνω ερωτήσεις (κι άλλες ακόμα) μπορούν ωστόσο να θεωρηθούν ως μερικές όψεις της κεντρικής ερώτησης που αφορά στο πώς όλοι οι παραπάνω παράγοντες καθορίζουν την κατεύθυνση που θα ακολουθήσει η φυσική επιλογή στο πλαίσιο ενός συγκεκριμένου οικολογικού σχηματισμού.

Η μεθοδολογία απάντησης των παραπάνω ερωτήσεων εμπλέκει παρατηρήσεις και καταγραφές πεδίου, σχεδιασμένα πειράματα πεδίου και εργαστηρίου καθώς και τη διατύπωση μαθηματικών αρχικά και ποιοτικών στη συνέχεια μοντέλων. Σε αυτό το πλαίσιο ο πρώτος στόχος αυτού του βιβλίου είναι να παρουσιάσει μεθόδους παραγωγής πρωτογενών δεδομένων, όπως είναι, μεταξύ άλλων, οι μέθοδοι δειγματοληψίας και τα πειραματικά σχέδια. Ο δεύτερος στόχος είναι να παρουσιάσει μεθόδους ανάλυσης και κωδικοποίησης των αρχικών δεδομένων, όπως είναι οι μέθοδοι προσαρμογής εμπειρικών μοντέλων

στα πρωτογενή δεδομένα, οι πολυμεταβλητές μέθοδοι ταξινόμησης κ.ά. Ο τρίτος τέλος στόχος είναι να παρουσιάσει μεθόδους διαχείρισης εκτεταμένων συνόλων δεδομένων, όπως είναι η προσαρμογή νευρωνικών δικτύων, η δικτυακή ανάλυση κ.ά.

2 Από τη θεωρία πιθανοτήτων

2.1 Πειραματικός σχεδιασμός

Ένα από τα βασικότερα χαρακτηριστικά του κάθε σχεδίου παραγωγής δεδομένων είναι η καθαρότητα των στόχων του. Αυτό σημαίνει τρία πράγματα: α) καλοδιατυπωμένες υποθέσεις εργασίας επειδή αυτές καθορίζουν εν πολλοίς όλα τα επόμενα βήματα της πειραματικής εργασίας, β) καθορισμό των αρχικών συνθηκών του πειράματος, καθώς και σαφή περιγραφή του είδους των δεδομένων που πρόκειται να προκύψουν αφού από αυτό εξαρτάται σε μεγάλο βαθμό η ερμηνευτική εμβέλεια του πειράματος, γ) καθορισμό του τρόπου με τον οποίο θα αναλυθούν αυτά τα δεδομένα γιατί αυτό ενισχύει την αξιοπιστία των συμπερασμάτων.

Ένα δεύτερο σημαντικό στοιχείο του πειραματικού σχεδίου είναι ο πραγματισμός του. Αυτό σημαίνει ότι πριν την εκτέλεση του σχεδίου θα πρέπει να γίνουν ακριβείς εκτιμήσεις σχετικά με την επάρκεια της διαθέσιμης υποδομής. Θα πρέπει επίσης να καθοριστεί ο χρόνος που αναμένεται ότι θα διαρκέσει το πείραμα και να γίνει μια κατά το δυνατόν ακριβέστερη εκτίμηση της ερευνητικής προσπάθειας που θα απαιτηθεί ώστε να επιτελεστεί το έργο. Τέλος, θεωρείται αναγκαίο να εκτιμηθεί και το κατά πόσον η διαθέσιμη τεχνογνωσία εγγυάται την αποτελεσματική εκτέλεση αυτού του έργου.

Το τρίτο σπουδαίο στοιχείο του κάθε πειραματικού σχεδιασμού είναι το αυστηρό σχέδιο καταχώρησης των ανεπεξέργαστων δεδομένων. Αυτό επιβάλει α) προεπιλογή ενός κατάλληλου λογισμικού (software), β) σχεδιασμό φύλλων καταχώρησης των δεδομένων γ) σχεδιασμό φύλλων αρχικής επεξεργασίας από την οποία θα προκύψουν βασικά στατιστικά μεγέθη, όπως οι μέσοι όροι, οι διακυμάνσεις των δεδομένων, κ.λ.π. Πολλοί χρησιμοποιούν για τη δουλειά αυτή το

πρόγραμμα Excel, ωστόσο και πολλά άλλα προγράμματα κρίνονται εξίσου κατάλληλα, δ) τη λεπτομερή περιγραφή, υπό τη μορφή ημερολογίου, των φάσεων εκτέλεσης του σχεδίου.

2.2 Αξιοπιστία μετρήσεων

Ας υποθέσουμε ότι σε κάποια στιγμή ένας ερευνητής ενδιαφέρεται να μάθει το πόσα πεύκα υπάρχουν στο δάσος του Σείχ-Σου. Ένας τρόπος για την εκπλήρωση αυτού του στόχου είναι να μετρηθούν όλα τα δένδρα του δάσους ένα προς ένα. Η μέθοδος αυτή δίνει μεν ακριβή αποτελέσματα, πλην όμως η εφαρμογή της παρουσιάζει ανυπέρβλητες πρακτικές δυσκολίες. Ένας άλλος, ευκολότερος τρόπος, είναι να προχωρήσουμε σε έμμεσες εκτιμήσεις. Προς τούτο επιλέγονται μερικές δασικές περιοχές ορισμένου μεγέθους εντός των οποίων και μόνον θα πραγματοποιηθούν εξαντλητικές καταγραφές. Στη συνέχεια θα υπολογιστεί ο μέσος αριθμός δένδρων ανά δασική περιοχή και το αποτέλεσμα θα προβληθεί (αναχθεί) στη συνολική έκταση του δάσους. Προφανώς, το έλλειμμα ακρίβειας των μεθόδων έμμεσης εκτίμησης αντισταθμίζεται από τις ευκολίες τους.

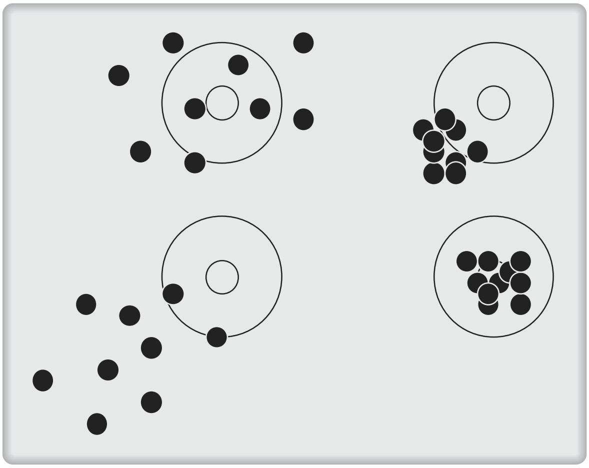
Στο πεδίο της οικολογίας εργαζόμαστε εν πολλοίς με αριθμητικές εκτιμήσεις της τιμής των διαφόρων μεγεθών. Άλλο όμως είναι η εκτίμηση της τιμής ενός μεγέθους και άλλο η αληθής του τιμή. Ενδέχεται, πράγματι, το αποτέλεσμα των αριθμητικών εκτιμήσεων να αποκλίνει από τις πραγματικές τιμές των διαφόρων μεταβλητών στις ακόλουθες περιπτώσεις:

- Αν πρόκειται για αριθμητική σειρά (π.χ. χρονοσειρά) και για τις εκτιμήσεις έχει ληφθεί υπόψη ένας περιορισμένος αριθμός αρχικών στοιχείων της σειράς.
- Αν για οποιονδήποτε λόγο παραλείπονται κατά σύστημα από τους υπολογισμούς αρκετά δεκαδικά ψηφία.
- Αν τα αναπόφευκτα ανθρώπινα λάθη ή τα λάθη των εργαλείων ή ακόμα και τα λάθη κατά τις καταχωρήσεις των δεδομένων συμβαίνουν με σχετικά υψηλή συχνότητα.

Για τους παραπάνω λόγους, προκειμένου να εκτιμηθεί η τιμή μιας μεταβλητής, σπάνια μία και μόνο μέτρηση επαρκεί. Το σχέδιο προβλέπει συνήθως την λήψη επαναληπτικών μετρήσεων, οπότε, η τελική τιμή της μεταβλητής υπολογίζεται με βάση μια στατιστική ποσότητα, ας πούμε το μέσο όρο των διαδοχικών μετρήσεων. Γενικά, η εκτίμηση της τιμής μιας μεταβλητής θεωρείται αξιόπιστη αν οι διαδοχικές μετρήσεις χαρακτηρίζονται από ακρίβεια και έλλειψη απόκλισης. Η ακρίβεια (precision) αναφέρεται στην επαναληψιμότητα των διαδοχικών μετρή-

σεων, ή πράγμα που είναι το ίδιο στη χαμηλή διακύμανση των αποτελεσμάτων των επαναληπτικών μετρήσεων. Αντίθετα με τον όρο απόκλιση (bias) εννοούμε τη συστηματική υπέρ- ή υπό-εκτίμηση της τιμής της μεταβλητής.

Προκειμένου να διευκρινιστούν οι παραπάνω κατηγορίες στο Διάγραμμα 2.1 σκιαγραφείται το αποτέλεσμα βολών κατά στόχων. Οι βολές του σκοπευτή που έβαλε στον πάνω αριστερά στόχο δεν είναι μεν ακριβείς αφού αποκλίνουν η μία από την άλλη, πλην όμως δεν αποκλίνουν συστηματικά προς τα πάνω ή προς κάτω, αλλά ούτε προς τα δεξιά ή τα αριστερά του στόχου. Επομένως, η απόκλιση δεν θα χαρακτηριστεί συστηματική.



Διάγραμμα 2.1. Σκαρίφημα όπου απεικονίζεται το αποτέλεσμα βολών επί στόχου.

Αντίθετα, οι βολές του σκοπευτή που έβαλε στον πάνω δεξιά στόχο χαρακτηρίζονται από ακρίβεια αφού διαθέτουν επαναληψιμότητα, πλην όμως αποκλίνουν συστηματικά προς τα κάτω και αριστερά του στόχου.
Με ανάλογο τρόπο οι βολές του σκοπευτή που σημάδευε στον κάτω αριστερά στόχο χαρακτηρίζονται από έλλειψη ακρίβειας αλλά και μεγάλη απόκλιση, ενώ εκείνες που έπεφταν στον κάτω δεξιά στόχο ήταν ακριβείς και εμφάνιζαν ελάχιστη απόκλιση.

2.3 Κατανομές πιθανοτήτων

Είναι γεγονός ότι ο βαθμός αβεβαιότητας που συνδέεται με τις εκτιμήσεις των βιολογικών δεδομένων είναι ιδιαίτερα ψηλός. Πέρα από τα προβλήματα αξιοπιστίας των εκτιμήσεων για τα οποία έγινε λόγος προηγούμενα, άλλη πηγή αβεβαιότητας αποτελεί και το γεγονός ότι οι εκτιμήσεις βασίζονται στη θεώρηση δείγματος (sample) και για το λόγο αυτό πρόκειται για ατελείς προσεγγίσεις της πραγματικής τιμής της μεταβλητής. Ακόμη όμως και δύο διαδοχικές τέλειες προσεγγίσεις της αληθούς τιμής μιας μεταβλητής σπάνια θα συμπέσουν λόγω της τεράστιας ποικιλομορφίας που χαρακτηρίζει τα βιολογικά φαινόμενα.

Για τη μελέτη φαινόμενων που ενέχουν αβεβαιότητα, γνώσεις από τη θεωρία των πιθανοτήτων είναι ιδιαίτερα χρήσιμες. Πράγματι, το σύνολο σχεδόν των τεχνικών που χρησιμοποιούνται για την παραγωγή και τη διαχείριση οικολογικών δεδομένων εμπλέκουν με τον ένα ή τον άλλον τρόπο στοιχεία από τη θεωρία πιθανοτήτων.

Η πιθανότητα να συμβεί ένα γεγονός X, να γεννήσει ας πούμε ένα πουλί 4 αβγά συμβολίζεται ως $P(4)$. Οι τιμές της ποσότητας $P(4)$ κυμαίνονται από 0 ως 1. Αν $P(4) = 0$ τότε το γεγονός (εδώ η περίπτωση να γεννήσει το πουλί μας 4 αβγά) δεν μπορεί να συμβεί. Αντίθετα, αν $P(4) = 1$ τότε το γεγονός είναι βέβαιο, πράγμα που θα πει ότι το πουλί μας θα γεννήσει 4 αβγά το δίχως άλλο.

Ο κλασσικός τρόπος μέτρησης της πιθανότητας με την οποία συμβαίνει ένα γεγονός είναι να καταγράψουμε έναν μεγάλο αριθμό γεγονότων (εδώ τον αριθμό αυγών που γέννησαν ας πούμε 141 πουλιά) και να λογαριάσουμε το πόσες φορές εμφανίστηκε το κάθε επιμέρους γεγονός. Έστω λοιπόν ότι από τα 141 πουλιά τα 41 δεν γέννησαν, τα 30 γέννησαν 1 αβγό, τα 43 γέννησαν 2 αβγά, τα 16 γέννησαν 3 και τα υπόλοιπα 11 γέννησαν 4 αβγά. Αυτό που κάναμε ως τώρα ήταν να λογαριάσουμε τη συχνότητα (frequency) εμφάνισης (καταγραφής) του κάθε πιθανού γεγονότος (41 φορές δεν καταγράφηκε κανένα αβγό, 30 φορές καταγράφηκε 1 αβγό, 43 φορές καταγράφηκαν 2 αβγά κ.ο.κ). Αν τώρα διαιρέσουμε την κάθε μια από αυτές τις συχνότητες με το σύνολο των καταμετρήσεων, δηλαδή το 141, τότε θα υπολογίσουμε τις πιθανότητες (probabilities) εμφάνισης του κάθε γεγονότος. Έτσι, η πιθανότητα να μη βρούμε κανένα αβγό σε μια φωλιά είναι $41/141 = 0.29$ ή 29%. Αντίστοιχα η πιθανότητα να βρούμε 1 αβγό είναι $30/141 = 0.21$ ή 21%, ενώ η πιθανότητα να καταγραφούν 3 αβγά είναι 0.11 ή 11%.

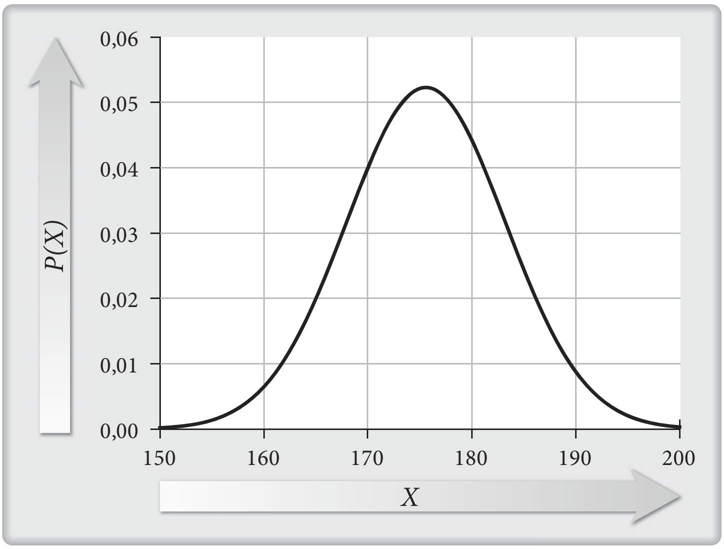
Ιδιαίτερη χρησιμότητα βρίσκουν διαγράμματα όπου στον οριζόντιο άξονα εμφανίζονται όλα τα πιθανά γεγονότα X και στον κάθετο η πιθανότητα να συμβεί καθένα από αυτά. Ας υποθέσουμε για παράδειγμα ότι σε κάποια χώρα το ύψος των ανθρώπων κυμαίνεται από 150 ως 200 cm. Τότε μπορούμε να κατασκευάσουμε γράφημα στον οριζόντιο άξονα του οποίου θα αναγράψουμε κατά αύξουσα σειρά τους αριθμούς 150-200 (Διάγραμμα 2.2). Στον κάθετο άξονα του ίδιου διαγράμματος θα απεικονιστούν οι πιθανότητες που αντιστοιχούν στα ενδεχόμενα γεγονότα X $(=150-200)$. Έτσι, για παράδειγμα η πιθανότητα συναντήσουμε σε αυτή τη χώρα έναν άνθρωπο με ύψος 170 cm είναι περίπου 0.04 ή 4%. Τα γραφήματα για τα οποία γίνεται λόγος εδώ ονομάζονται διαγράμματα πυκνότητας πιθανοτήτων. Η πυκνότητα πιθανοτήτων συμβολίζεται ως:

$$f_X(x) = P(X = x)$$

και διαβάζεται ως η πιθανότητα ώστε η μεταβλητή X να λάβει τη συγκεκριμένη τιμή x.

2.3.1. Κανονική κατανομή

Στο Διάγραμμα 2.2 απεικονίζεται η πυκνότητα πιθανοτήτων της κανονικής κατανομής. Η κατανομή αυτή είναι γνωστή και ως κατανομή Gauss και εμφανίζει τη γνωστή συμμετρική κωδωνοειδή μορφή.



Διάγραμμα 2.2. Διάγραμμα πυκνότητας πιθανοτήτων της κατανομής του ύψους ανθρώπινου πληθυσμού.

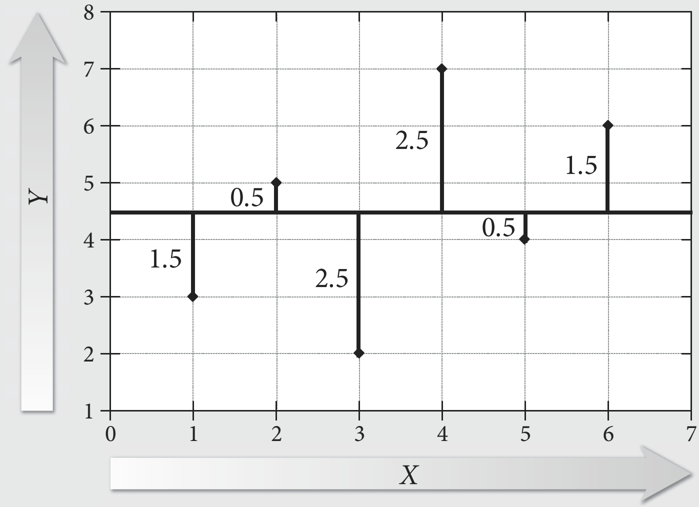
Η κανονική κατανομή (normal distribution) προσδιορίζεται πλήρως με τη βοήθεια δύο παραμέτρων, του μέσου όρου (μ) και της διακύμανσης (σ^2). Ο μέσος όρος αντιστοιχεί στο ενδεχόμενο X που έχει τη μεγαλύτερη πιθανότητα εμφάνισης (εδώ το ύψος των 175.6 cm) και αποτελεί μέτρο της κεντρικής τάσης της κατανομής. Η διακύμανση με τη σειρά της (Διάγραμμα 2.3) αποτελεί μέτρο της διασποράς των τιμών πάνω και κάτω από το μέσο όρο (στην περίπτωσή μας $\sigma^2 = 58.21$). Η κανονική κατανομή συμβολίζεται ως $N(\mu, \sigma^2)$, οπότε εδώ $N(176.5, 58.21)$.

Σημειώνω παρενθετικά ότι η διακύμανση είναι ίση με το μέσο όρο των τετραγωνισμένων αποκλίσεων των επιμέρους ατομικών παρατηρήσεων από το μέσο όρο αυτών των παρατηρήσεων. Για παράδειγμα για το διάγραμμα 2.3 υπολογίζουμε:

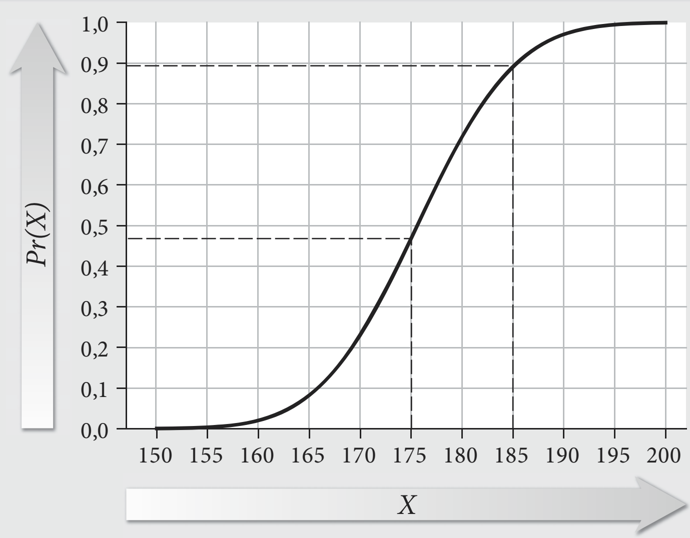
$$\boldsymbol{\sigma^2 = \{(-1.5)^2 + 0.5^2 + (-2.5)^2 + 2.5^2 + (-0.5)^2 + 1.5^2\}/6}$$

Η ερώτηση που αφορά στο ύψος των ανδρών της χώρος για την οποία γίνεται λόγος μπορεί να τεθεί και με διαφορετικό τρόπο: *ποια είναι η πιθανότητα ένας άνθρωπος από αυτή τη χώρα να έχει ύψος μεταξύ 175 και 185 εκατοστών;* Προ-

κειμένου να απαιτηθεί η ερώτηση θα χρησιμοποιηθεί ένα διαφορετικό γράφημα γνωστό ως διάγραμμα κατανομής πιθανοτήτων (Διάγραμμα 2.4).



Διάγραμμα 2.3. Επί του άξονα Y φέρονται οι ατομικές τιμές 6 παρατηρήσεων (1-6) που απεικονίζονται επί του άξονα X. Ο αριθμητικός μέσος όρος των 6 ατομικών μετρήσεων είναι ίσος με 4.5 και συμβολίζεται με την ευθεία γραμμή στη θέση Y = 4.5. Τέλος απεικονίζονται και οι αποκλίσεις των επιμέρους ατομικών μετρήσεων από το μέσο όρο.



Διάγραμμα 2.4. Διάγραμμα κανονικής κατανομής πιθανοτήτων.

Στον οριζόντιο άξονα φέρονται και πάλι τα ενδεχόμενα γεγονότα *X*, (εδώ το ύψος από 150 ως 200 cm), αλλά στον κάθετο απεικονίζονται οι αθροιστικές πιθανότητες $F(x)$. Στο γράφημα αυτό διαβάζουμε ότι η πιθανότητα ένα άτομο αυτής της χώρας να έχει ύψος έως το πολύ 175 εκ. είναι περίπου 47%, ενώ η αντίστοιχη πιθανότητα το άτομο αυτό να έχει ύψος ως το πολύ 185 εκ. είναι περίπου 89%. Έτσι, η πιθανότητα το άτομο αυτό να έχει ύψος μεταξύ 175 και 185 εκ. είναι $89-47=42\%$. Η κατανομή πιθανοτήτων συμβολίζεται ως:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)$$

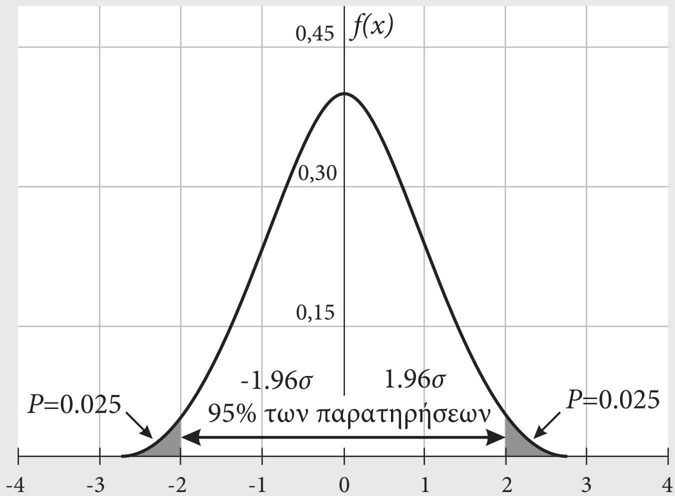
και διαβάζεται ως η πιθανότητα η μεταβλητή X να λάβει μια οποιαδήποτε τιμή μικρότερη ή το πολύ ίση με χ.

Στις παραπάνω παραγράφους περιγράφηκε μια μέθοδος που απαντά σε ερωτήματα που αφορούν στις πιθανότητες να συμβούν ορισμένα γεγονότα, όταν αυτές οι πιθανότητες ακολουθούν την κανονική κατανομή. Οπως όμως έχει αναφερθεί ήδη η κανονική κατανομή πιθανοτήτων ορίζεται απόλυτα από τις δύο παραμέτρους της, το μέσο όρο και τη διακύμανση. Αυτό θα πει ότι για να περιγράψουμε το φαινόμενο στη Γερμανία και την Ιαπωνία θα χρειαστούμε δύο ξεχωριστές καμπύλες αφού στις δύο χώρες καταγράφονται διαφορετικές τιμές μέσων όρων και διακυμάνσεων.

Η πολυπλοκότητα επιτείνεται όταν, όπως συνήθως συμβαίνει, χρησιμοποιούνται δείγματα και παράγονται εκτιμήσεις. Στην περίπτωση αυτή σε κάθε δειγματοληψία (sampling) αντιστοιχεί και διαφορετική καμπύλη. Για την υπέρβαση δυσκολιών αυτού του είδους χρησιμοποιείται μέθοδος στάθμισης (ή άλλως μετασχηματισμού *Z* ή κανονικοποίησης ή και τυποποίησης, standardization) των δεδομένων. Συγκεκριμένα, από κάθε επιμέρους τιμή της μεταβλητής, X_i αφαιρείται ο μέσος όρος. Στη συνέχεια το αποτέλεσμα που θα προκύψει διαιρείται με την ποσότητα σ (τυπική απόκλιση). Έτσι, προκύπτει η σταθμισμένη (κανονικοποιημένη ή τυποποιημένη) κανονική μεταβλητή:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

οι τιμές της οποίας ακολουθούν την κανονική κατανομή πιθανοτήτων με παραμέτρους $\mu=0$ και $\sigma=\sigma^2=1$. Έτσι λοιπόν, ανεξάρτητα από το γεγονός ότι οι διάφορες μεταβλητές εμφανίζουν διαφορετικούς μέσους όρους και διακυμάνσεις, μετά το μετασχηματισμό που περιγράφηκε εδώ θα παραχθεί μία και η αυτή σταθμισμένη κανονική κατανομή με συμβολισμό $N(0, 1)$ (Διάγραμμα 2.5).



Διάγραμμα 2.5. Διάγραμμα πυκνότητας πιθανοτήτων της σταθμισμένης κανονικής κατανομής. Οι σκιασμένες περιοχές στα δύο άκρα της καμπύλης καταλαμβάνουν και οι δύο μαζί εμβαδόν ίσο με το 5 του συνολικού εμβαδού που περιέχεται ανάμεσα στην καμπύλη της σταθμισμένης κανονικής κατανομής και τον οριζόντιο άξονα. Αυτό σημαίνει ότι η πιθανότητα να συμβεί ένα γεγονός με τιμή ίση ή μικρότερη του –1.96 είναι ίση ή μικρότερη του 2.5%. Αντίστοιχη είναι και η πιθανότητα να συμβεί γεγονός με τιμή ίση ή μεγαλύτερη του 1.96

Ιδιαίτερο ενδιαφέρον για τις διάφορες στατιστικές δοκιμασίες που θα παρουσιαστούν αργότερα έχει το γεγονός ότι η πιθανότητα ένα γεγονός να έχει τιμή μεταξύ +1.96*σ* και –1.96*σ* είναι ίση με το 95%. Με άλλα λόγια το 95% των τιμών βρίσκονται μεταξύ 1.96*σ* και –1.96*σ*. Αντίστοιχα, η πάνω και η κάτω ουρά της καμπύλης αντιστοιχούν στο επίπεδο πιθανότητας 2.5% η κάθε μια και 5% οι δύο μαζί.

Ως σήμερα η σταθμισμένη κανονική κατανομή αποτελεί τη βάση σχεδόν όλων των μεθόδων στατιστικού ελέγχου. Μάλιστα, όπως θα αναφερθεί και αργότερα, η εφαρμογή των διαφόρων δοκιμασιών προ-απαιτεί το μετασχηματισμό των αρχικών δεδομένων όταν αυτά δεν ακολουθούν την κανονική κατανομή.

Σήμερα ωστόσο αναπτύσσονται μέθοδοι ελεύθερες από τους περιορισμούς της κανονικότητας. Πρόκειται είτε για επαναληπτικές αριθμητικές μεθόδους (δένδρα απόφασης, νευρωνικά δίκτυα κ.λ.π.), ή για μεθόδους που διαχειρίζονται μη κανονικά δεδομένα. Στην τελευταία κατηγορία ανήκουν οι διαρκώς δι-

ευρυνόμενες τεχνικές των γενικευμένων γραμμικών μοντέλων (Generalized Linear Models).

2.3.2. Άλλες συνεχείς κατανομές πιθανοτήτων

Ανάμεσα στις πολλές κατανομές πιθανοτήτων ιδιαίτερο ενδιαφέρον έχει η συμμετρική κατανομή Student (Διάγραμμα 2.I, δες παράρτημα στο τέλος του κεφαλαίου) η οποία, όπως θα δούμε αργότερα χρησιμοποιείται για τον ορισμό διαστημάτων εμπιστοσύνης διαφόρων εκτιμητών, το στατιστικό έλεγχο κ.λ.π. Μια άλλη ενδιαφέρουσα κατανομή είναι η ασύμμετρη κατανομή X^2 (Διάγραμμα 2.II) η οποία χρησιμοποιείται ευρέως στο στατιστικό έλεγχο. Στο Διαγράμματα 2.III (ομοίως στο παράρτημα), επιχειρείται να απεικονιστεί το 95% διάστημα εμπιστοσύνης στην περίπτωση της κατανομής X^2 για τα οποία θα γίνει λόγος αργότερα. Τέλος, στο παράρτημα περιλαμβάνονται και δύο ασύμμετρες κατανομές, η πλέον χρησιμοποιούμενη στο στατιστικό έλεγχο κατανομή F (Διάγραμμα 2.IV) και η λογαριθμοκανονική κατανομή (Διάγραμμα 2.V) που η χρήση της μεγαλώνει τα τελευταία χρόνια.

2.3.3. Ασυνεχείς κατανομές

Υπάρχουν ποσότητες, όπως το ανθρώπινο ύψος για το οποίο έγινε λόγος προηγούμενα, οι οποίες μέσα στα όρια του πεδίου ορισμού τους (στο παράδειγμά μας 150-200 cm) μπορούν να πάρουν οποιαδήποτε τιμή (συνεχείς μεταβλητές). Πράγματι, αν διαθέταμε το κατάλληλο εργαλείο θα μπορούσαμε να μετρήσουμε ανθρώπους το ύψος των οποίων θα διέφερε κατά όσα δεκαδικά επέτρεπε το εργαλείο μας. Όμως, υπάρχουν και μεγέθη που μπορούν να πάρουν απολύτως διακριτές μεταξύ τους τιμές. Αυτό συμβαίνει με τον αριθμό αβγών που θα γεννήσει ένα πουλί. Έτσι, στο παράδειγμα που παρουσιάστηκε στην αρχή του κεφαλαίου τα αβγά αυτά μπορεί να είναι 0, 1, 2, 3 …….. 7. Η γεννητικότητα λοιπόν του πουλιού αποδίδεται με τη βοήθεια ασυνεχούς μεταβλητής.

2.3.3.1 Διωνυμική κατανομή πιθανοτήτων

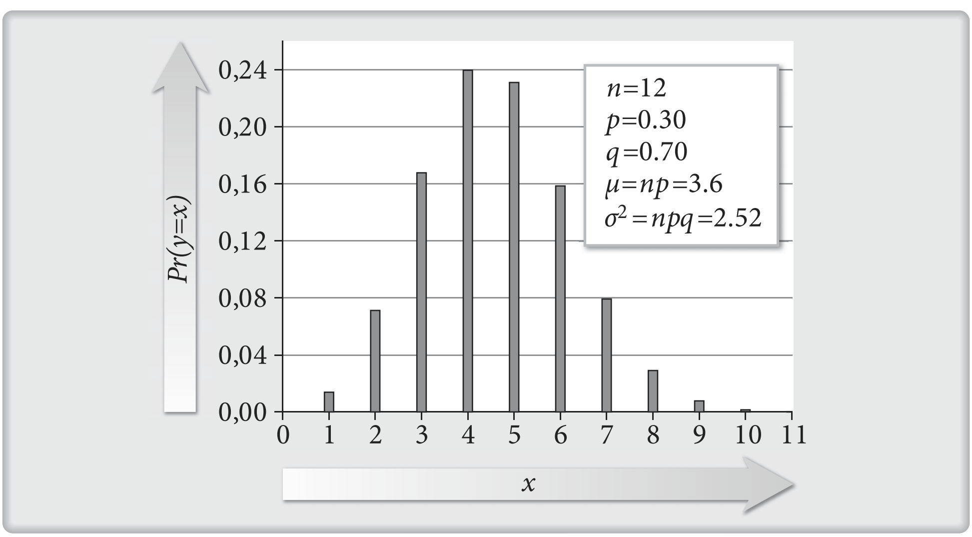
Ασυνεχής είναι και η μεταβλητή που μετρά το αποτέλεσμα που θα προκύψει από το στρίψιμο στον αέρα ενός νομίσματος. Επιπλέον, η μεταβλητή αυτή θα χαρακτηριστεί ως δίτιμη επειδή μπορεί να πάρει δύο μόνο τιμές, κορώνα ή γράμματα. Άλλες δίτιμες μεταβλητές είναι εκείνες που μετρούν ποιότητες π.χ. ζεστό-κρύο, γλυκό-πικρό κ.λ.π. Για να περιγράψουμε τέτοιου είδους φαινόμενα χρησιμοποιούμε συχνά τη διωνυμική κατανομή πιθανοτήτων.

Στην περίπτωση του νομίσματος το πρόβλημα τίθεται ως εξής: *πόσες πιθανότητες έχω να φέρω 1, 2, 3, ……. 10 φορές κορώνα αν στρίψω συνολικά 10 φορές*

το νόμισμα; Για να απαντηθεί το ερώτημα εκτός από το σύνολο των προσπαθειών (εδώ $n=10$) χρειάζομαι και την πιθανότητα με την οποία το στρίψιμο φέρνει κορώνα. Προφανώς, αν το νόμισμα δεν έχει ατέλειες και το στρίψιμο γίνει κανονικά τότε η πιθανότητα να είναι το αποτέλεσμα κορώνα είναι $p=0.5$, οπότε και η πιθανότητα το νόμισμα να πέσει γράμματα είναι $q=1-p=1-0.5=0.5$, δηλαδή όσο και η πιθανότητα να έρθει γράμματα. Η διωνυμική κατανομή ορίζεται πλήρως από τις παραμέτρους n (= σύνολο προσπαθειών, παρατηρήσεων, καταμετρήσεων κ.λ.π.) και p (=πιθανότητα να συμβεί το ένα από τα δύο ενδεχόμενα).

Ας υποθέσουμε παραδειγματικά ότι 12 αγελάδες τρέφονται σε περιφραγμένο λιβάδι 1000 m^2 με ομοιόμορφα διανεμημένο γρασίδι. Οι αγελάδες δεν δείχνουν καμιά τάση συγκέντρωσης. Ζητείται να βρούμε την πιθανότητα να συναντήσουμε εντός μιας επιφάνειας 300 m^2 0, 1, 2, 3, … , 12 αγελάδες.

Για να απαντήσουμε στο ερώτημα θα πρέπει να εξετάσουμε τη θέση που βρίσκεται η κάθε αγελάδα. Συνεπώς ο συνολικός αριθμός παρατηρήσεων θα είναι $n=12$. Για την κάθε αγελάδα ισχύουν δύο ενδεχόμενα, να βρίσκεται ή όχι εντός μιας επιφάνειας 300 m^2. Η πιθανότητα να βρεθεί μια αγελάδα εντός μιας επιφάνειας 300 m^2 είναι $p=300/1000=0.3$, επομένως η πιθανότητα να μη βρεθεί καμία αγελάδα σε εκείνη την επιφάνεια είναι $q=1-0.3=0.7$. Η κατανομή πιθανοτήτων στην περίπτωση που συζητάμε εδώ απεικονίζεται στο Διάγραμμα 2.6. Στον οριζόντιο άξονα φέρονται τα ενδεχόμενα γεγονότα x κατά σειρά αύξουσα, ενώ στον κάθετο άξονα απεικονίζεται η πιθανότητα με την οποία θα συμβεί το γεγονός y να είναι το x.

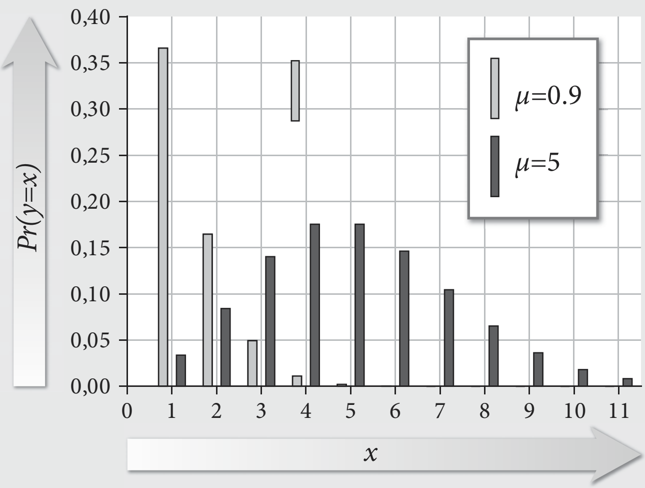


Διάγραμμα 2.6. Διωνυμική κατανομή. Απεικονίζεται επίσης ο τρόπος υπολογισμού του μέσου και της διακύμανσης.

Σημειώστε ότι αν οι πιθανότητες p και q είναι περίπου ίσες (≈0.5) και ο αριθμός των παρατηρήσεων αρκετά μεγάλος, ας πούμε n>25, τότε η κατανομή πιθανοτήτων τείνει ασυμπτωτικά προς την κανονική.

2.3.3.2 Κατανομή Poisson

Μια άλλη, ιδιαιτέρως χρήσιμη, ασυνεχής κατανομή είναι γνωστή ως κατανομή Poisson (Διάγραμμα 2.7). Χαρακτηρίζεται από μία μονάχα παράμετρο την μ. Χαρακτηριστικό της κατανομής αυτής είναι ότι ο μέσος όρος μ και η διακύμανση σ^2 έχουν την ίδια τιμή. Θεωρείται ως η κατανομή που ταιριάζει καλύτερα σε περιπτώσεις όπου τα γεγονότα συμβαίνουν σπάνια (δηλαδή οταν $p \to 0$ και επομένως $q \to 1$). Χρησιμοποιείται για παράδειγμα προκειμένου να περιγράψει μεταβλητές που αποδίδουν αριθμούς ατόμων στη μονάδα επιφανείας ή τη μονάδα του χρόνου.



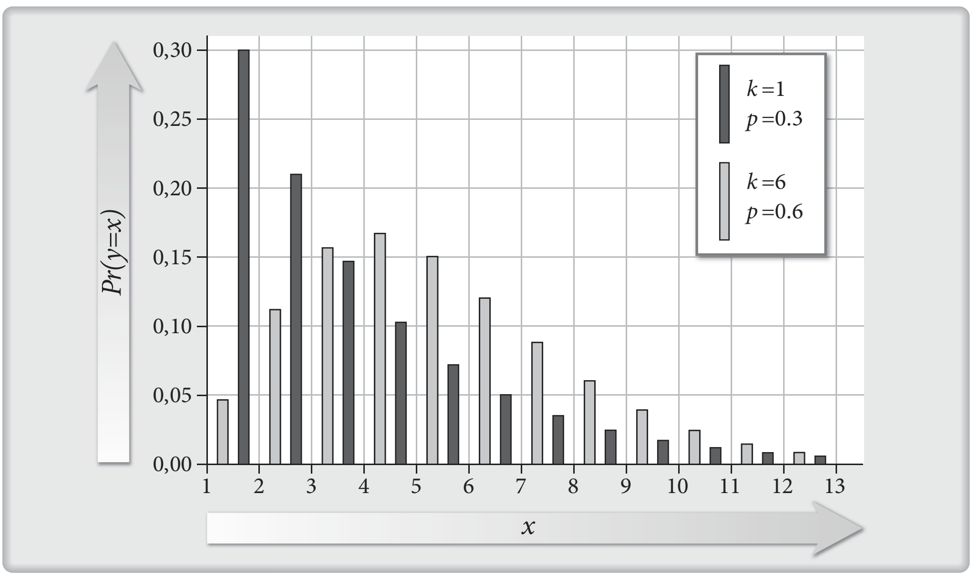
Διάγραμμα 2.7. Κατανομή Poisson. Για μεγαλύτερες τιμές μ η κατανομή τείνει ασυμπτωτικά προς τη διωνυμική.

Για σχετικά μεγάλες τιμές της παραμέτρου μ, πάντως, η κατανομή Poisson τείνει ασυμπτωτικά προς τη διωνυμική. Αυτό συχνά επιτρέπει τη χρήση είτε της μιας είτε της άλλης κατανομής. Ανάμεσα στις δύο, η κατανομή Poisson χρησιμοποιείται σε περιπτώσεις που θέλουμε να καταγράψουμε τις πιθανότητες εμφάνισης ενός γεγονότος *1, 2, 3, … n* φορές όταν είναι γνωστός ο μέσος αριθμός με τον οποίο συμβαίνουν αυτά τα γεγονότα. Αν για παράδειγμα γνωρίζουμε ότι ο αριθμός ατόμων ενός δένδρου ανά μονάδα επιφανείας είναι π.χ. $\mu = 0.9$ και θέλουμε

να μάθουμε με ποια πιθανότητα θα συναντήσουμε *0, 1, 2, 3 … n* δένδρα ανά μονάδα επιφανείας, τότε μπορούμε να χρησιμοποιήσουμε την κατανομή Poisson. Το παραπάνω πρόβλημα μπορεί να τεθεί και με διαφορετικό τρόπο: '*πόσες επιφάνειες με 1, 2, 3, … n δένδρα θα συναντήσουμε αν μετρηθούν 100 από αυτές τις επιφάνειες*'. Κατά προτίμηση την απάντηση στην τελευταία αυτή ερώτηση θα δώσει η διωνυμική κατανομή με $p = 0.9$, $q = 0.1$ και $n = 100$.

2.3.3.3 Αρνητική διωνυμική κατανομή

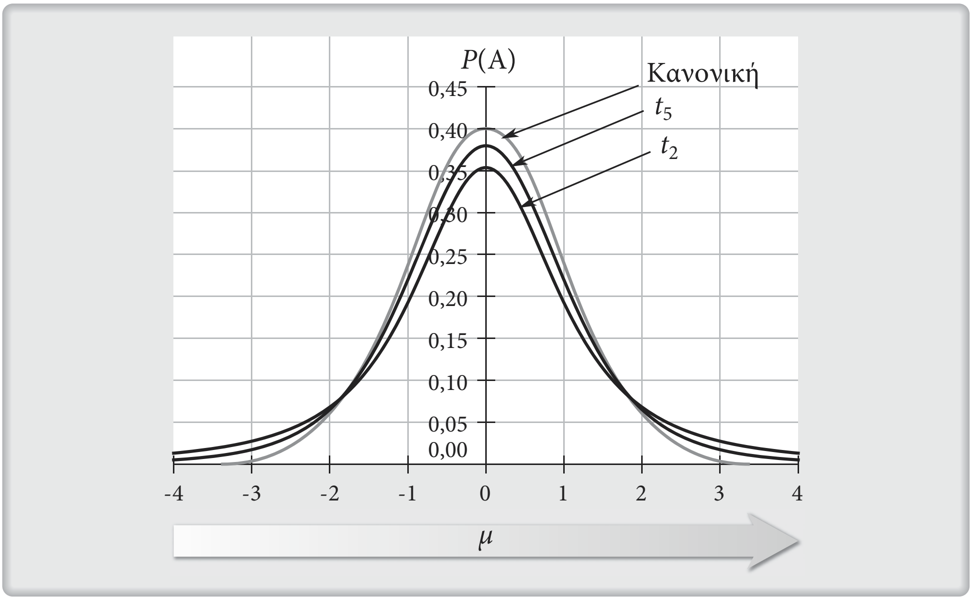
Μια άλλη ασυνεχής κατανομή που βρίσκει μεγάλη χρησιμότητα σε προβλήματα διανομής πληθυσμών στο χώρο είναι η αρνητική διωνυμική κατανομή (Διάγραμμα 2.8). Αυτή χρησιμοποιείται, για παράδειγμα, σε περιπτώσεις που λαμβάνονται εδαφικά καρώτα και γίνεται καταμέτρηση του αριθμού των εδαφόζωων σε καθένα από αυτά. Μάλιστα, είναι η συνιστώμενη κατανομή όταν η χωροδιανομή των ατόμων είναι συσωματική.



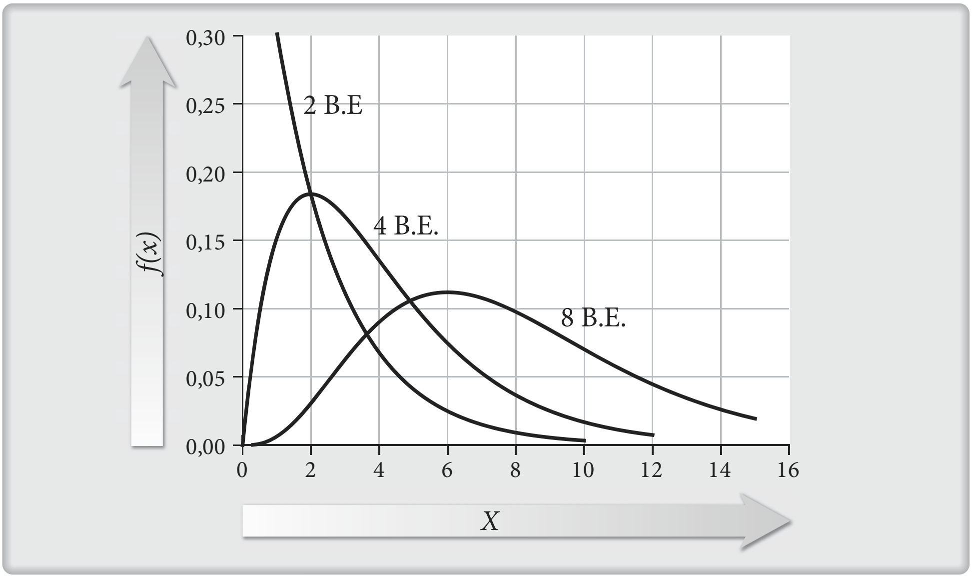
Διάγραμμα 2.8. Αρνητική διωνυμική κατανομή.

Η αρνητική διωνυμική κατανομή ορίζεται πλήρως από δύο παραμέτρους. Η παράμετρος k παίρνει θετικές ακέραιες τιμές. Αυτή αποτελεί μέτρο του μέσου μεγέθους των συσωματώσεων που σχηματίζουν τα άτομα. Η άλλη παράμετρος είναι η p που παίρνει τιμές μεταξύ 0 και 1. Ισχύει ότι μέσος $= kq/p$ και διακύμανση $= kq/p^2$, όπου $q = 1\text{-}p$.

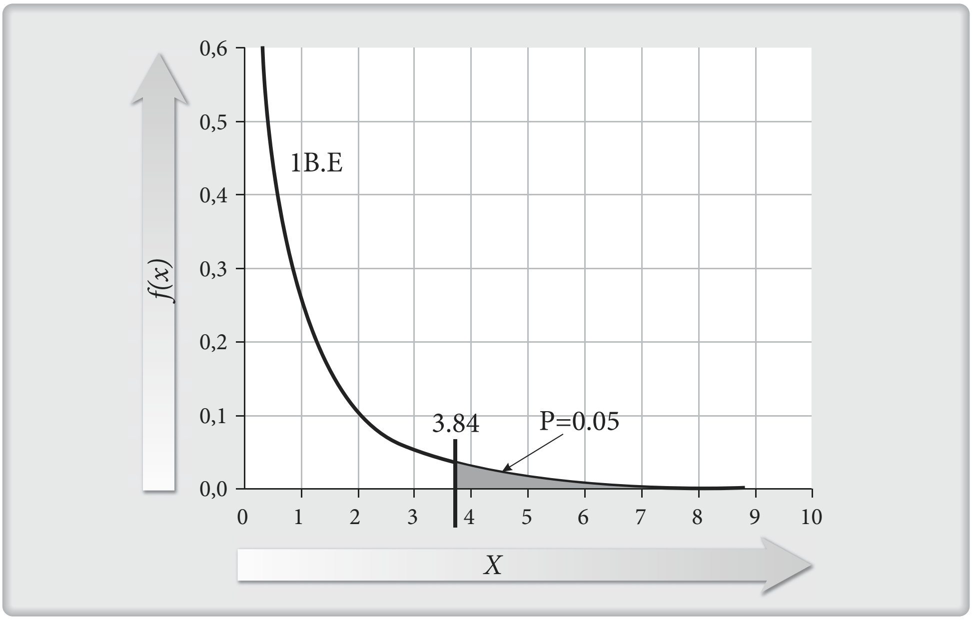
2.4 Παράρτημα



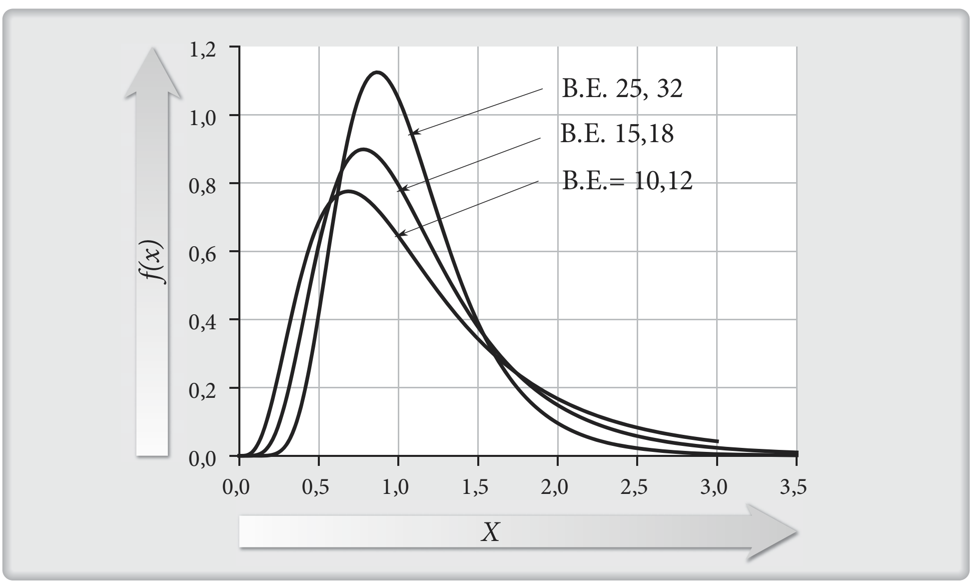
Διάγραμμα 2.Ι. Κατανομή Student με 5 και 2 βαθμούς ελευθερίας. Για χάρη συγκρίσεων οι δύο κατανομές Student περιβάλλονται από την *N*(0,1) που απεικονίζεται με αχνή γραμμή.



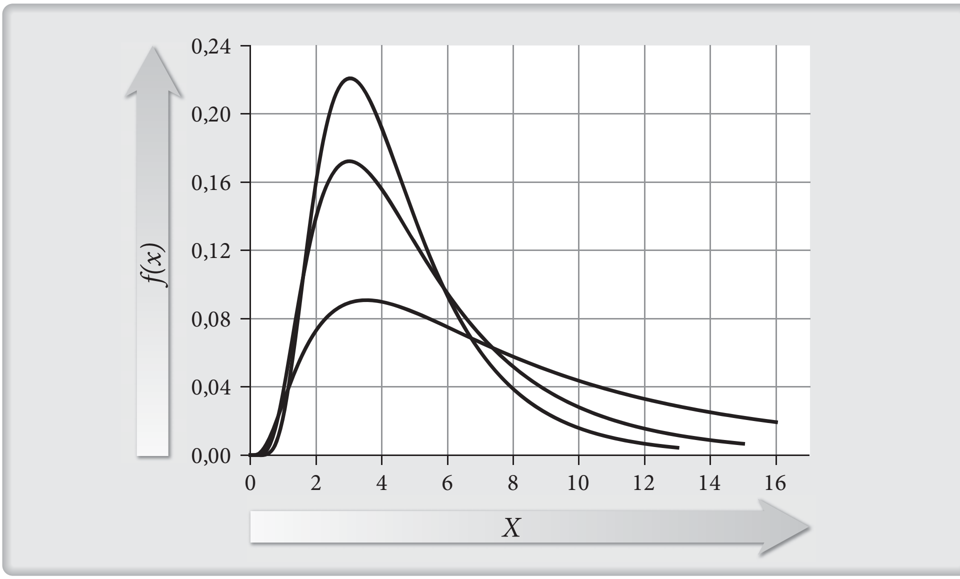
Διάγραμμα 2.ΙΙ. X^2 κατανομή με 2, 4 και 8 βαθμούς ελευθερίας.



Διάγραμμα 2.ΙΙΙ. Κατανομή X^2 με Β.Ε. = 1. Γεγονότα με τιμές X^2 = 3.84 και πάνω συμβαίνουν με πιθανότητα < 5%.



Διάγραμμα 2.ΙV. Κατανομή F (Fisher). Η μορφή της κατανομής ορίζεται από δύο παραμέτρους, τους βαθμούς ελευθερίας v_1 και v_2. Όσο μεγαλύτερες είναι οι τιμές των παραμέτρων τόσο μικρότερη είναι και η ασυμμετρία της καμπύλης.



Διάγραμμα 2.V. Δεξιά ασύμμετρη λογαριθμοκανονική κατανομή με παραμέτρους (με τη σειρά από τα πάνω προς τα κάτω) $\mu = 0.38$ και $\sigma = 0.52$, $\mu = 1.5$ και $\sigma = 0.63$, $\mu = 2$ και $\sigma = 0.86$. Για χαμηλές τιμές σ η κατανομή τείνει ασυμπτωτικά προς την κανονική.