
***Κεφάλαιο 5* Τυπική στατιστική ανάλυση μιας υδρολογικής μεταβλητής**

Στο κεφάλαιο αυτό θα εφαρμόσουμε τις αρχές και μεθόδους της στατιστικής, τις οποίες παρουσιάσαμε ήδη στο κεφάλαιο 3, σε ένα από τα πιο τυπικά προβλήματα της στατιστικής υδρολογίας: στην ανάλυση μιας υδρολογικής μεταβλητής. Η υπό μελέτη υδρολογική μεταβλητή μπορεί να είναι οποιαδήποτε: βροχόπτωση ή απορροή, μέση ή μέγιστη, ετήσια ή μηνιαία, κτλ. Η ανάλυση περιλαμβάνει συνήθως τα εξής τρία βήματα:

1. *Συμπύκνωση του δείγματος.* Αυτή γίνεται είτε με αριθμητικό τρόπο, χρησιμοποιώντας ορισμένους αντιπροσωπευτικούς στατιστικούς δείκτες, είτε με γραφικό τρόπο, χρησιμοποιώντας ιστογράμματα και εμπειρικές συναρτήσεις κατανομής.
2. *Προσαρμογή θεωρητικού μοντέλου.* Αυτή περιλαμβάνει την επιλογή κατάλληλης συνάρτησης κατανομής για τη μεταβλητή, εκτίμηση των παραμέτρων της και στατιστικό έλεγχο της προσαρμογής.
3. *Στατιστική πρόγνωση.* Ανάλογα με το σκοπό της μελέτης, η πρόγνωση αποσκοπεί στην εκτίμηση της τιμής της μεταβλητής για δεδομένη περίοδο επαναφοράς. Η εκτίμηση αυτή μπορεί να είναι είτε σημειακή είτε εκτίμηση διαστήματος.

Παρόλο που η τυπική αυτή στατιστική ανάλυση εφαρμόζεται για κάθε τύπο θεωρητικού μοντέλου, στα παραδείγματα αυτού του κεφαλαίου θα κάνουμε χρήση μόνο της κανονικής κατανομής, την οποία έχουμε ήδη μελετήσει στο εδάφιο 2.8.2. Στο αμέσως επόμενο κεφάλαιο θα δώσουμε

το θεωρητικό υλικό καθώς και αντίστοιχα παραδείγματα για όλες τις άλλες συναρτήσεις κατανομής που χρησιμοποιούνται στην τεχνική υδρολογία.

5.1 Στατιστικά χαρακτηριστικά δείγματος

Με τον όρο *στατιστικά χαρακτηριστικά δείγματος* εννοούμε διάφορους στατιστικούς δείκτες που συνοψίζουν σε λίγους αριθμούς τις πιο χαρακτηριστικές ιδιότητες του δείγματος. Διακρίνουμε δύο κατηγορίες στατιστικών χαρακτηριστικών. Η πρώτη περιλαμβάνει τις *δειγματικές ροπές και τα παράγωγά τους χαρακτηριστικά*. Συγκεκριμένα περιλαμβάνει: (α) τη δειγματική μέση τιμή, η οποία, όπως είδαμε, αποτελεί παράμετρο θέσης· (β) τη δειγματική διασπορά και τους παράγωγους δείκτες διασποράς (τυπική απόκλιση και συντελεστή μεταβλητότητας)· (γ) την τρίτη κεντρική ροπή και το συντελεστή ασυμμετρίας. Η δεύτερη κατηγορία περιλαμβάνει απλούστερους στατιστικούς δείκτες που ο υπολογισμός τους προϋποθέτει τη διάταξη του δείγματος σε φθίνουσα ή αύξουσα σειρά και γι' αυτό εδώ αναφέρονται ως *συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος*. Πρόκειται για την ελάχιστη και μέγιστη τιμή του δείγματος, τη διάμεσο (παράμετρος θέσης), το άνω και κάτω τεταρτημόριο και το διατεταρτημοριακό πλάτος (παράμετρος διασποράς).

Οι δειγματικές ροπές και τα παράγωγά τους χαρακτηριστικά υπολογίζονται με εφαρμογή των αντίστοιχων εκτιμητριών που έχουν δοθεί στο κεφάλαιο 3. Οι αντίστοιχοι τύποι δίνονται επίσης συγκεντρωμένοι στον Πίν. 5.1, και μάλιστα σε μορφή απλούστερη για υπολογισμούς. Δίνονται ακόμη οι συντελεστές διόρθωσης μεροληψίας (στήλη 3), με τους οποίους πρέπει να πολλαπλασιαστούν οι απλές εκτιμήσεις (στήλη 2) προκειμένου να υπολογιστούν οι αντίστοιχες αμερόληπτες εκτιμήσεις. Στον Πίν. 5.1 δίνεται επίσης και ο τρόπος εύρεσης των συνοπτικών χαρακτηριστικών του διατεταγμένου δείγματος.

Πίν. 5.1 Τυπικά χαρακτηριστικά δείγματος και τρόπος υπολογισμού τους

Στατιστικός δείκτης	Απλή εκτίμηση	Συντελεστής διόρθωσης μεροληψίας
<i>1. Δειγματικές ροπές και παράγωγα χαρακτηριστικά</i>		
Μέση τιμή	$\bar{x} = \frac{1}{n} \sum x_i$	—
Διασπορά	$s_X^2 = \frac{1}{n} \sum x_i^2 - \frac{1}{n^2} (\sum x_i)^2$ $= \frac{1}{n} \sum x_i^2 - \bar{x}^2$	$\frac{n}{n-1}$
Τυπική απόκλιση	s_X	$\approx \sqrt{\frac{n}{n-1}}$
Συντελεστής μεταβλητότητας	$\hat{C}_{v_X} = \frac{s_X}{\bar{x}}$	$\approx \sqrt{\frac{n}{n-1}}$
Τρίτη κεντρική ροπή	$\hat{\mu}_X^{(3)} = \frac{1}{n} \sum x_i^3 - \frac{3}{n^2} (\sum x_i)(\sum x_i^2) + \frac{2}{n^3} (\sum x_i)^3$ $= \frac{1}{n} \sum x_i^3 - 3\bar{x}s_X^2 - \bar{x}^3$	$\frac{n^2}{(n-1)(n-2)}$
Συντελεστής ασυμμετρίας	$\hat{C}_{s_X} = \frac{\hat{\mu}_X^{(3)}}{s_X^3}$	βλ. εδάφιο 3.2.4
<i>2. Συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος</i>		
Ελάχιστη τιμή	$\hat{x}_{\min} = \min(x_1, x_2, \dots, x_n)$	—
Μέγιστη τιμή	$\hat{x}_{\max} = \max(x_1, x_2, \dots, x_n)$	—
Διάμεσος	$\hat{x}_{0.5}$: Ο μεσαίος όρος του διατεταγμένου δείγματος, ή για άρτιο μέγεθος δείγματος, ο μέσος όρος των δύο μεσαίων όρων του.	—
Κάτω τεταρτημόριο	$\hat{x}_{0.25}$: Η διάμεσος του τμήματος του δείγματος που περιλαμβάνει τις τιμές $x_i \leq \hat{x}_{0.5}$.	—
Άνω τεταρτημόριο	$\hat{x}_{0.75}$: Η διάμεσος του τμήματος του δείγματος που περιλαμβάνει τις τιμές $x_i \geq \hat{x}_{0.5}$.	—
Διατεταρτημοριακό εύρος	$\hat{\delta}_X = \hat{x}_{0.75} - \hat{x}_{0.25}$	—

Τα συνοπτικά χαρακτηριστικά του διατεταγμένου δείγματος μπορούν να απεικονιστούν και με τη μορφή απλού διαγράμματος, της *συνοπτικής παράστασης δείγματος* (στη διεθνή βιβλιογραφία αναφέρεται ως *box plot*), όπως αυτό του Σχ. 5.1 (σ. 111). Ένα τέτοιο διάγραμμα περιλαμβάνει ένα κεντρικό ορθογώνιο “κουτί” και δύο κατακόρυφες “κεραίες” πάνω και κάτω από το κουτί. Όλα αυτά τα στοιχεία απεικονίζονται σε κατάλληλη κατακόρυφη κλίμακα. Για τη σχεδίαση ακολουθούνται οι εξής οδηγίες (Hirsch et al., 1993, σ. 17.10):

1. Η μεσαία γραμμή του κουτιού αντιπροσωπεύει τη διάμεσο του δείγματος.
2. Η κάτω γραμμή (βάση) του κουτιού αντιπροσωπεύει το κάτω τεταρτημόριο του δείγματος.
3. Η πάνω γραμμή (οροφή) του κουτιού αντιπροσωπεύει το άνω τεταρτημόριο του δείγματος.
4. Ορίζεται το βοηθητικό μέγεθος *βήμα* το οποίο είναι 1.5 φορές το διατεταρτημοριακό πλάτος.
5. Η κάτω κεραία εκτείνεται από τη βάση του κουτιού μέχρι τη μικρότερη τιμή του δείγματος που βρίσκεται μέσα στα όρια ενός βήματος από τη βάση του κουτιού.
6. Η άνω κεραία εκτείνεται από την οροφή του κουτιού μέχρι τη μεγαλύτερη τιμή του δείγματος που βρίσκεται μέσα στα όρια ενός βήματος από την οροφή του κουτιού.
7. Τιμές του δείγματος που βρίσκονται έξω από τα όρια των κεραιών και απέχουν 1-2 βήματα από τη βάση ή την οροφή του κουτιού χαρακτηρίζονται ως *εξωτερικές* και σημειώνονται στο διάγραμμα με το σύμβολο \times .
8. Τιμές του δείγματος που βρίσκονται έξω από τα όρια των κεραιών και απέχουν περισσότερο από 2 βήματα από τη βάση ή την οροφή του κουτιού χαρακτηρίζονται ως *μακρινές εξωτερικές* και σημειώνονται στο διάγραμμα με το σύμβολο \circ .

Σύμφωνα με τα παραπάνω, η ελάχιστη και η μέγιστη τιμή του δείγματος σημειώνονται στη συνοπτική παράσταση δείγματος είτε ως όρια των κεραιών, εφόσον δεν απέχουν περισσότερο από ένα βήμα από τα όρια

του κουτιού, είτε διαφορετικά ως εξωτερικές ή μακρινές εξωτερικές τιμές.

Η συνοπτική παράσταση ενός δείγματος δίνει με πολύ σαφή τρόπο τη γενική στατιστική εικόνα του, απεικονίζοντας ταυτόχρονα χαρακτηριστικά θέσης (διάμεσος), διασποράς (διατεταρτημοριακό πλάτος) και ασυμμετρίας. Η συμμετρία ή ασυμμετρία του δείγματος αναγνωρίζεται από θέση της μεσαίας γραμμής σε σχέση με τις βάσεις του κουτιού, καθώς και από τη σύγκριση των μηκών των κεραιών. Ακόμη το διάγραμμα δίνει πληροφορίες για το πόσο κοντά στην κανονική κατανομή είναι ένα δείγμα. Αν η μεταβλητή του δείγματος είναι κανονική, τότε αναμένεται κατ' αρχήν συμμετρική εικόνα του διαγράμματος, και επίσης δεν αναμένονται εξωτερικές και μακρινές εξωτερικές τιμές, παρά μόνο με συχνότητες 1 στα 100 και 1 στα 300 000 σημεία, αντίστοιχα.

Εφαρμογή 5.1

Στον Πίν. 5.2 δίνεται το δείγμα των ετήσιων απορροών της λεκάνης του ποταμού Ευήνου ανάντη του υδρομετρικού σταθμού Πόρος Ρηγανίου. Ζητείται ο υπολογισμός των στατιστικών χαρακτηριστικών του και η σχεδίαση της συνοπτικής παράστασής του.*

Πίν. 5.2 Δείγμα ετήσιου όγκου απορροής (σε hm^3)[†] στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

Υδρολογ. έτος	Όγκος απορροής	Υδρολογ. έτος	Όγκος απορροής	Υδρολογ. έτος	Όγκος απορροής
1970-71	807	1977-78	715	1984-85	588
1971-72	695	1978-79	1064	1985-86	874
1972-73	788	1979-80	942	1986-87	552
1973-74	705	1980-81	1042	1987-88	529
1974-75	462	1981-82	1037	1988-89	469
1975-76	580	1982-83	674	1989-90	217
1976-77	807	1983-84	906	1990-91	772

* Η θέση Πόρος Ρηγανίου βρίσκεται σε αρκετή απόσταση κατάντη του φράγματος Αγίου Δημητρίου, το οποίο εντάσσεται στο σύστημα υδατικών πόρων που χρησιμοποιείται για την ύδρευση της Αθήνας. Το δείγμα των απορροών του Πόρου Ρηγανίου χρησιμοποιήθηκε (μετά από κατάλληλη αναγωγή) για τη μελέτη του ταμιευτήρα Αγίου Δημητρίου.

[†] Υπενθυμίζεται ότι η μονάδα hm^3 σημαίνει κυβικά εκατόμετρα ($1 \text{ hm}^3 = (100 \text{ m})^3 = 1\,000\,000 \text{ m}^3$).

α. Δειγματικές ροπές και παράγωγα χαρακτηριστικά

Αν και σήμερα ο υπολογισμός των ροπών γίνεται αυτόματα με υπολογιστές ή αριθμομηχανές, εδώ εκθέτουμε για διδακτικούς λόγους τον υπολογισμό των στατιστικών χαρακτηριστικών με το χέρι.

Πίν. 5.3 Βοηθητικοί υπολογισμοί για την Εφαρμογή 5.1.

i	x_i	x_i^2	x_i^3
1	807	651 249	525 557 943
2	695	483 025	335 702 375
3	788	620 944	489 303 872
4	705	497 025	350 402 625
5	462	213 444	98 611 128
6	580	336 400	195 112 000
7	807	651 249	525 557 943
8	715	511 225	365 525 875
9	1064	1 132 096	1 204 550 144
10	942	887 364	835 896 888
11	1042	1 085 764	1 131 366 088
12	1037	1 075 369	1 115 157 653
13	674	454 276	306 182 024
14	906	820 836	743 677 416
15	588	345 744	203 297 472
16	874	763 876	667 627 624
17	552	304 704	168 196 608
18	529	279 841	148 035 889
19	469	219 961	103 161 709
20	217	47 089	10 218 313
21	772	595 984	460 099 648
Άθροισμα	15 225	11 977 465	9 983 241 237

Ο υπολογισμός των αθροισμάτων $\sum x$, $\sum x^2$ και $\sum x^3$ γίνεται στον Πίν. 5.3. Έχουμε $\sum x = 15\,225$, $\sum x^2 = 11\,977\,465$ και $\sum x^3 = 9\,983\,241\,237$. Η μέση τιμή του δείγματος είναι

$$\bar{x} = \sum x / n = 15\,225 / 21 = 725.0 \text{ hm}^3$$

Η διασπορά είναι

$$s_x^2 = \sum x^2 / n - \bar{x}^2 = 11\,977\,465 / 21 - 725.0^2 = 44\,730.5 \text{ (hm}^3)^2$$

η τυπική απόκλιση

$$s_X = \sqrt{44\,730.5} = 211.5 \text{ hm}^3$$

και ο συντελεστής μεταβλητότητας

$$\hat{C}_{v_X} = s_X / \bar{x} = 211.5 / 725.0 = 0.29$$

Η τρίτη κεντρική ροπή είναι

$$\begin{aligned} \hat{\mu}_X^{(3)} &= \sum x^3 / n - 3 \bar{x} s_X^2 - \bar{x}^3 = \\ &= 9\,983\,241\,237 / 21 - 3 \times 725.0 \times 44\,730.5 - 725.0^3 \\ &= -2\,974\,523 \text{ (hm}^3)^3 \end{aligned}$$

και ο συντελεστής ασυμμετρίας

$$\hat{C}_{s_X} = \hat{\mu}_X^{(3)} / s_X^3 = -2\,974\,523 / 211.5^3 = -0.31$$

Οι συντελεστές διόρθωσης μεροληψίας είναι: (α) για τη διασπορά

$$n / (n - 1) = 21 / 20 = 1.05.$$

(β) για την τυπική απόκλιση και το συντελεστή μεταβλητότητας (κατά προσέγγιση)

$$\sqrt{n / (n - 1)} = \sqrt{1.05} = 1.025$$

και (γ) για την τρίτη κεντρική ροπή (και κατά προσέγγιση για το συντελεστή ασυμμετρίας, βλ. εξίσωση 3.28)

$$n^2 / [(n-1)(n-2)] = 21^2 / (20 \times 19) = 1.16$$

β. Συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος

Το δείγμα, διατεταγμένο σε φθίνουσα σειρά, φαίνεται στον Πίν. 5.5. Από τον πίνακα αυτό υπολογίζονται άμεσα τα συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος, τα οποία δίνονται στον Πίν. 5.4. Η διάμεσος είναι η 11η στη σειρά (μεσαία) τιμή του διατεταγμένου δείγματος, ενώ το κάτω και άνω τεταρτημόριο είναι η 16η και 6η στη σειρά τιμή, αντίστοιχα.

Πίν. 5.4 Στατιστικά χαρακτηριστικά (ροπές και παράγωγα χαρακτηριστικά) του δείγματος του ετήσιου όγκου απορροής (σε hm^3) στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

Στατιστικός δείκτης	Απλή εκτίμηση	Συντελεστής διόρθωσης μεροληψίας	Αμερόληπτη εκτίμηση
Μέση τιμή	$\bar{x} = \sum x / n = 725.0$	—	$\bar{x} = 725.0$
Διασπορά	$s_X^2 = \sum x^2 / n - \bar{x}^2 = 44\,730$	$\frac{n}{n-1} = 1.05$	$s_X^{*2} = 46\,967$
Τυπική απόκλιση	$s_X = 211.5$	$\approx \sqrt{\frac{n}{n-1}} = 1.025$	$s_X^* \approx 216.7$
Συντελεστής μεταβλητότητας	$\hat{C}_{v_X} = s_X / \bar{x} = 0.29$	$\approx \sqrt{\frac{n}{n-1}} = 1.025$	$\hat{C}_{v_X}^* \approx 0.29$
Τρίτη κεντρική ροπή	$\hat{\mu}_X^{(3)} = \sum x^3 / n - 3\bar{x}s_X^2 - \bar{x}^3 = -2\,974\,523$	$\frac{n^2}{(n-1)(n-2)} = 1.16$	$\hat{\mu}_X^{*(3)} = -3\,542\,012$
Συντελεστής ασυμμετρίας	$\hat{C}_{s_X} = \hat{\mu}_X^{(3)} / s_X^3 = -0.31$	$\approx \frac{n^2}{(n-1)(n-2)} = 1.16$	$\hat{C}_{s_X}^* \approx -0.36$

Πίν. 5.5 Διατεταγμένο (σε φθίνουσα σειρά) δείγμα ετήσιου όγκου απορροής (σε hm^3) στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

A/A	Όγκος απορροής	A/A	Όγκος απορροής	A/A	Όγκος απορροής
1	1064	8	807	15	588
2	1042	9	788	16	580
3	1037	10	772	17	552
4	942	11	715	18	529
5	906	12	705	19	469
6	874	13	695	20	462
7	807	14	674	21	217

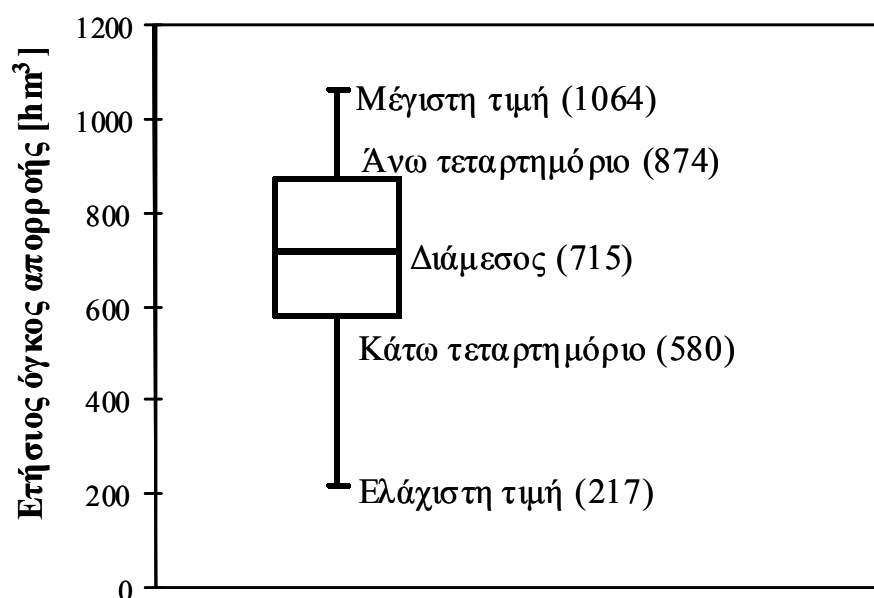
γ. Σχεδίαση συνοπτικής παράστασης δείγματος

Ακολουθώντας τη διαδικασία που περιγράφηκε στην ενότητα 5.1 και χρησιμοποιώντας τα συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος (Πίν. 5.4) σχεδιάζουμε εύκολα το διάγραμμα του Σχ. 5.1. Το βήμα της συνοπτικής παράστασης είναι $1.5 \times 294 = 441 \text{ hm}^3$ και επομένως η μέγιστη τεταγμένη της άνω κεραίας είναι $874 + 441 = 1315$

hm^3 . Δεδομένου, όμως, ότι η μέγιστη τιμή του δείγματος είναι 1064 hm^3 , η άνω κεραία πρέπει να εκτείνεται μέχρι αυτή την τιμή. Αντίστοιχα, η ελάχιστη τεταγμένη της κάτω κεραίας είναι $580 - 441 = 139 \text{ hm}^3$. Δεδομένου, όμως, ότι η ελάχιστη τιμή του δείγματος είναι 217 hm^3 , η κάτω κεραία πρέπει να εκτείνεται μέχρι αυτή την τιμή.

Πίν. 5.6 Συνοπτικά χαρακτηριστικά του διατεταγμένου δείγματος του ετήσιου όγκου απορροής (σε hm^3) στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

Στατιστικός δείκτης	Εκτίμηση
Ελάχιστη τιμή	$\hat{x}_{\max} = \min(x_1, \dots, x_n) = 217$
Μέγιστη τιμή	$\hat{x}_{\max} = \max(x_1, \dots, x_n) = 1064$
Διάμεσος	$\hat{x}_{0.5} = x_{(11)} = 715$
Κάτω τεταρτημόριο	$\hat{x}_{0.25} = x_{(16)} = 580$
Άνω τεταρτημόριο	$\hat{x}_{0.75} = x_{(6)} = 874$
Διατεταρτημοριακό εύρος	$\hat{d}_X = \hat{x}_{0.75} - \hat{x}_{0.25} = 294$



Σχ. 5.1 Συνοπτική παράσταση δείγματος (box plot) για το παράδειγμα της εφαρμογής 5.1

5.2 Ιστογράμματα

Ένα άλλο τρόπο γραφικής παράστασης ενός δείγματος παρέχουν τα ιστογράμματα, τα οποία προϋποθέτουν την κατάταξη των τιμών του δείγμα-

τος σε k διαστήματα μεγέθους Δ . Έστω ότι το i διάστημα έχει όρια c_i και $c_{i+1} = c_i + \Delta$ και ότι ο αριθμός των τιμών του δείγματος που βρίσκονται στο i διάστημα, δηλαδή αυτών που ικανοποιούν τη σχέση $c_i \leq x < c_{i+1}$, είναι n_i . Ορίζουμε ως ιστόγραμμα την κλιμακωτή συνάρτηση

$$\varphi(x) = \frac{n_i}{n\Delta} \quad c_i \leq x < c_{i+1} \quad i = 1, \dots, k \quad (5.1)$$

της οποίας η γραφική παράσταση έχει τη μορφή του Σχ. 5.2. Συχνά η συνάρτηση του ιστογράμματος ορίζεται με απλούστερους τρόπους ως $\varphi(x) = n_i/n$, ή και $\varphi(x) = n_i$. Για τις τελευταίες χρησιμοποιούμε τους όρους *ιστόγραμμα σχετικής συχνότητας* και *ιστόγραμμα απόλυτης συχνότητας*, αντίστοιχα. Προς αποφυγή σύγχυσης, για τη συνάρτηση της (5.1) χρησιμοποιούμε και τον όρο *ιστόγραμμα πυκνότητας συχνότητας*.

Προκειμένου να καταρτίσουμε το ιστόγραμμα, αρχικά επιλέγουμε τον αριθμό των διαστημάτων k . Κατά κανόνα λαμβάνουμε $k = \ln n / \ln 2$ και την τιμή που προκύπτει τη στρογγυλεύουμε προς τα πάνω. Το μέγεθος των διαστημάτων Δ το παίρνουμε ίσο για όλα τα διαστήματα (αν και για το ιστόγραμμα πυκνότητας συχνότητας είναι επιτρεπτά και άνισα διαστήματα).

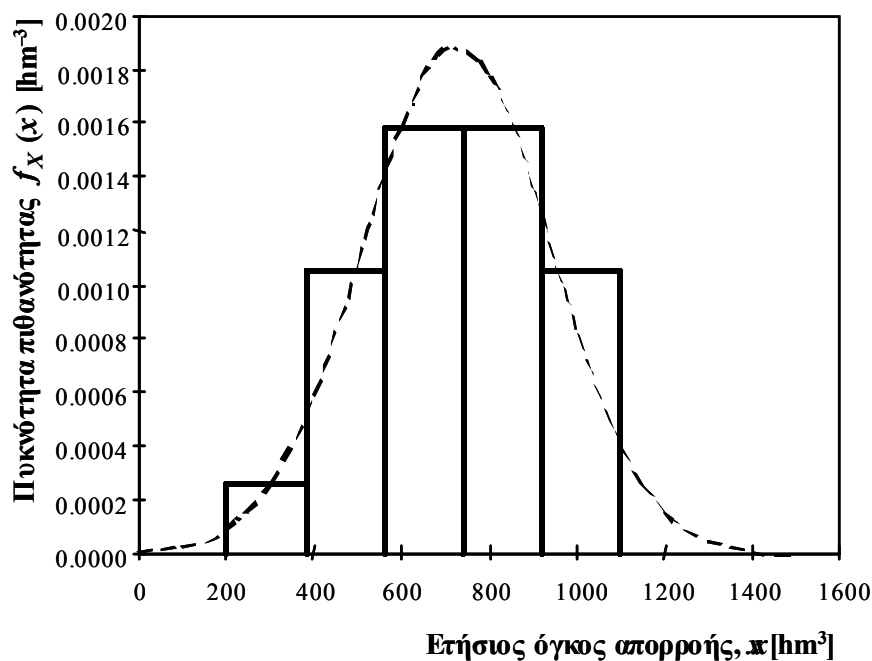
Εφαρμογή 5.2

Να σχεδιαστεί το ιστόγραμμα για το δείγμα της Εφαρμογής 5.1.

Ο αριθμός των κλάσεων θα πρέπει να ληφθεί $k = \ln 21 / \ln 2 = 4.4$. Στρογγυλεύοντας προς τα πάνω, επιλέγουμε 5 κλάσεις. Το διάστημα μεταβολής των τιμών της μεταβλητής στο δείγμα είναι το [217, 1064]. Στρογγυλεύοντας, παίρνουμε για διάστημα μεταβολής το [200, 1100] με $\Delta = (1100 - 200) / 5 = 180$. Οι υπόλοιποι υπολογισμοί δίνονται σε πινακοποιημένη μορφή στον Πίν. 5.7 και το ιστόγραμμα παρουσιάζεται στο Σχ. 5.2. Για λόγους σύγκρισης έχει χαραχτεί και η θεωρητική συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής (βλ. ενότητα 5.4).

Πίν. 5.7 Υπολογισμοί για το ιστόγραμμα της Εφαρμογής 5.2.

A/A	Όρια κλάσεων	Απόλυτη συχνότητα n_i	Σχετική συχνότητα n_i / n	Πυκνότητα συχνότητας $\phi = n_i / (n \Delta)$
	200			
1	380	1	0.048	0.00026
2	560	4	0.190	0.00106
3	740	6	0.286	0.00159
4	920	6	0.286	0.00159
5	1100	4	0.190	0.00106



Σχ. 5.2 Ιστόγραμμα για το παράδειγμα της εφαρμογής 5.2. Με διακεκομμένη γραμμή έχει χαραχτεί (για σύγκριση) η πυκνότητα πιθανότητας της κανονικής κατανομής $N(725, 211.5)$.

5.3 Εμπειρική συνάρτηση κατανομής

Όπως το ιστόγραμμα είναι το εμπειρικό αντίστοιχο της συνάρτησης πυκνότητας πιθανότητας, έτσι έχουμε και εμπειρικό αντίστοιχο της συνάρτησης κατανομής, το οποίο λέγεται *εμπειρική συνάρτηση κατανο-*

μής. Μια τέτοια εμπειρική συνάρτηση μπορεί κατ' αρχήν να κατασκευαστεί από ένα ιστόγραμμα, με ολοκλήρωσή του ως προς x , οπότε προκύπτει μια αύξουσα τεθλασμένη που αντιστοιχεί σε κάποιο είδος συνάρτησης κατανομής. Ωστόσο, η εισαγωγή της εμπειρικής συνάρτησης κατανομής μπορεί να γίνει με αμεσότερο και αντικειμενικότερο τρόπο, χωρίς τη μεσολάβηση του ιστογράμματος, το οποίο έχει κάποιο βαθμό υποκειμενικότητας λόγω της αυθαίρετης επιλογής των διαστημάτων και των ορίων τους.

5.3.1 Κλασική εμπειρική συνάρτηση κατανομής

Έστω η τυχαία μεταβλητή X με συνάρτηση κατανομής $F(x)$ και το δείγμα της X_1, X_2, \dots, X_n . Συμβολίζουμε με $X_{(1)}$ τη μεγαλύτερη από τις μεταβλητές του δείγματος, $X_{(2)}$ την αμέσως μικρότερη, κοκ., έτσι ώστε οι μεταβλητές $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ να παριστάνουν το δείγμα διατεταγμένο σε φθίνουσα σειρά. Οι $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ λέγονται και *διατεταγμένες στατιστικές συναρτήσεις*.

Σύμφωνα με τον κλασικό τρόπο, η εμπειρική συνάρτηση κατανομής είναι κλιμακωτή και ορίζεται από τη σχέση

$$\hat{F}(x) = \frac{n_x}{n} \quad (5.2)$$

όπου n_x είναι ο αριθμός των τιμών του δείγματος που δεν υπερβαίνουν την τιμή x . Η $\hat{F}(x)$ αποτελεί σημειακή εκτίμηση της άγνωστης συνάρτησης κατανομής του πληθυσμού $F(x)$.

5.3.2 Θέσεις σχεδίασης

Ονομάζουμε *θέση σχεδίασης* (plotting position) q_i της τιμής $x_{(i)}$ του διατεταγμένου δείγματος την εμπειρική πιθανότητα υπέρβασης της τιμής αυτής. Με βάση τον κλασικό τρόπο ορισμού της εμπειρικής κατανομής, για $x = x_{(1)}$ θα έχουμε $n_x = n$, και γενικότερα για $x = x_{(i)}$ θα έχουμε $n_x = n + 1 - i$. Κατά συνέπεια η εμπειρική συνάρτηση κατανομής είναι

$$\hat{F}(x_{(i)}) = \frac{n+1-i}{n} \quad i = 1, \dots, n \quad (5.3)$$

οπότε η θέση σχεδίασης, δηλαδή η εμπειρική πιθανότητα υπέρβασης είναι

$$q_i = \hat{F}_1(x_{(i)}) = 1 - \hat{F}(x_{(i)}) = \frac{i-1}{n} \quad i = 1, \dots, n \quad (5.4)$$

Παρατηρούμε ότι για $n = 1$ η πιο πάνω εξίσωση δίνει μηδενική πιθανότητα υπέρβασης. Έτσι, αν για παράδειγμα έχουμε ένα δείγμα ετήσιων βροχοπτώσεων με μέγιστη τιμή $x_{(1)} = 1800$ mm, οδηγούμαστε στο συμπέρασμα ότι η πιθανότητα ετήσιας βροχόπτωσης μεγαλύτερης από 1800 mm είναι μηδενική. Αυτό φυσικά είναι εσφαλμένο συμπέρασμα.

Για την άρση του παραπάνω προβλήματος χρησιμοποιούμε την τυχαία μεταβλητή

$$U_i = F_1(X_{(i)}) = 1 - F(X_{(i)}) \quad (5.5)$$

Μια σημειακή εκτίμηση[†] αυτής της μεταβλητής αποτελεί ταυτόχρονα και εκτίμηση της q_i . Κατ' αρχήν δεν μπορούμε να υπολογίσουμε τιμές της U_i από το δείγμα δεδομένου ότι η $F(x)$ είναι άγνωστη συνάρτηση. Ωστόσο, αποδεικνύεται[‡] ότι η κατανομή της U_i είναι ανεξάρτητη της $F(x)$ και έχει μέση τιμή*

$$E[U_i] = \frac{i}{n+1} \quad (5.6)$$

και διασπορά

$$\text{Var}[U_i] = \frac{i(n-i+1)}{(n+1)^2(n+2)} \quad (5.7)$$

[†] Ακριβέστερα και σύμφωνα με την ορολογία του κεφαλαίου 3, πρόκειται για πρόγνωση της μεταβλητής, εφόσον η U_i είναι τυχαία μεταβλητή και όχι παράμετρος.

[‡] Προκύπτει από τη συνάρτηση κατανομής των διατεταγμένων στατιστικών συναρτήσεων (βλ. Papoulis, 1990, σ. 207-208) μετά από κατάλληλη αλλαγή μεταβλητής.

* Για την ακρίβεια η U_i έχει συνάρτηση κατανομής βήτα (βλ. εδάφιο 6.4.1 με παραμέτρους i και $n - i + 1$).

Η απλούστερη εκτίμηση της U_i είναι η μέση τιμή της, δηλαδή

$$q_i = \frac{i}{n+1} \quad (5.8)$$

η οποία είναι γνωστή στη βιβλιογραφία ως *θέση σχεδίασης Weibull* ή *αμερόληπτη ως προς την πιθανότητα υπέρβασης θέση σχεδίασης*. Ο δεύτερος χαρακτηρισμός οφείλεται στην ιδιότητά της: $q_i = E[U_i] = E[F(X_{(i)})]$. Παρατηρούμε ότι με αυτή την εκτίμηση έχει αρθεί το πρόβλημα του μη-δενισμού της q_i για $i = 1$. Πράγματι για $i = 1$ έχουμε $q_i = 1 / (n + 1)$ και για $i = n$, $q_i = n / (n + 1)$.

Η εξίσωση (5.8) είναι η πιο διαδεδομένη για την εκτίμηση εμπειρικών πιθανοτήτων υπέρβασης στην τεχνική υδρολογία, αλλά δεν είναι η μοναδική. Άλλες παρόμοιες εξισώσεις έχουν αναπτυχθεί με στόχο να δίνουν αμερόληπτες εκτιμήσεις ποσοστημορίων, δηλαδή να ικανοποιούν (κατά προσέγγιση) τη συνθήκη

$$F^{-1}(q_i) = E[X_{(i)}] = E[F^{-1}(U_i)] \quad (5.9)$$

Στην περίπτωση αυτή η εκτίμηση, αντίθετα από αυτήν της (5.8), δεν είναι ανεξάρτητη της συνάρτησης κατανομής $F(x)$. Πάντως στην πράξη οι διάφορες εξισώσεις που έχουν αναπτυχθεί εκφράζονται με το γενικό τύπο

$$q_i = \frac{i - a}{n + 1 - 2a} \quad (5.10)$$

όπου a σταθερά (< 1). Ο τύπος αυτός είναι αντισυμμετρικός με την έννοια ότι $q_i = 1 - q_{n+1-i}$ και εμπεριέχει, ως ειδική περίπτωση ($a = 0$), και την (5.8). Στον Πίν. 5.8 δίνονται οι συνηθέστεροι στην τεχνική υδρολογία τύποι υπολογισμού της θέσης σχεδίασης μαζί με τις αντίστοιχες τιμές της σταθεράς a . Η εφαρμογή των διάφορων σχέσεων δίνει πρακτικά τα ίδια αποτελέσματα, εκτός από τις πολύ μικρές του i και ιδίως για $i = 1$, όπου διαφοροποιούνται αισθητά. Η τιμή για $i = 1$ έχει μεγάλη σημασία για την τεχνική υδρολογία, γιατί δίνει την εμπειρική περίοδο επαναφοράς της μεγαλύτερης τιμής του δείγματος, δηλαδή την $T_1 = 1 / q_1$. Η διαφοροποίηση του T_1 για τους διάφορους τύπους φαίνεται στον Πίν. 5.8.

Πίν. 5.8 Εναλλακτικοί τύποι υπολογισμού εμπειρικών πιθανοτήτων υπέρβασης *

Όνομα	Τύπος $q_i =$	Σταθερά $a =$	Περίοδος επα- ναφοράς μέγ. τιμής $T_1 =$	Εφαρμοσιμότητα
Weibull	$\frac{i}{n+1}$	0	$n+1$	Όλες οι κατανομές, αμερόληπτη εκτίμηση πιθανότητας υπέρβασης
Blom	$\frac{i-0.375}{n+0.25}$	0.375	$1.6n+0.4$	Κανονική κατανομή, αμερόληπτη εκτίμηση ποσοστημορίων
Cunnane	$\frac{i-0.4}{n+0.2}$	0.4	$1.667n+0.33$	Μεγάλο εύρος κατανομών, κατά προσέγγιση αμερόληπτη εκτίμηση ποσοστημορίων
Gringorten	$\frac{i-0.44}{n+0.12}$	0.44	$1.786n+0.21$	Κατανομή Gumbel †
Hazen	$\frac{i-0.5}{n}$	0.5	$2n$	Εμπειρική εκτίμηση ιστορικής αξίας· σήμερα τείνει να εγκαταλειφθεί

5.3.3 Χαρτιά πιθανότητας

Έχοντας εκτιμήσει τη θέση σχεδίασης κάθε τιμής του δείγματος με κάποιον από τους παραπάνω τρόπους, διαθέτουμε ένα σύνολο από n σημεία $(x_{(i)}, q_i)$ ή $(x_{(i)}, 1 - q_i)$, τα οποία μπορούν να παρασταθούν γραφικά ώστε να πάρουμε μια εικόνα της συνάρτησης κατανομής. Αυτό μπορεί κατ' αρχήν να γίνει σε κοινό ή μιλιμετρέ χαρτί, οπότε θα προκύψει ένα γράφημα παρόμοιο με αυτό του Σχ. 2.1 ή του Σχ. 2.3β, με τη διαφορά ότι, αντί της κλιμακωτής ή της συνεχούς γραμμής, θα έχουμε απλώς ένα σύνολο από σημεία. Όμως, στην τεχνική υδρολογία, επειδή η πληροφορία που παρέχει ένα τέτοιο διάγραμμα είναι πολύ ουσιαστική, είμαστε πιο συστηματικοί στη σχεδίασή του. Συγκεκριμένα, επιδιώκουμε, μέσω κατάλληλων μετασχηματισμών των αξόνων, να δώσουμε ευθύγραμμη

* Βλ. και Stedinger et al. (1993) όπου δίνονται και άλλοι τύποι.

† Βλ. κεφάλαιο 6.

διάταξη στα σημεία. Αυτό εξυπηρετεί διάφορους σκοπούς, όπως την ευκολότερη σχεδίαση, τη σαφέστερη σύγκριση θεωρητικής και εμπειρικής κατανομής, την ευκολότερη γραφική επέκταση έξω από τα όρια του δείγματος κ.ά. Χαρτιά στα οποία έχουν σχεδιαστεί οι άξονες με κατάλληλους μετασχηματισμούς, ώστε τα γραφήματα συγκεκριμένων συναρτήσεων κατανομής να παριστάνονται ως ευθείες, λέγονται *χαρτιά πιθανότητας*. Τέτοια χαρτιά συνήθως υπάρχουν έτοιμα (όπως το λογαριθμικό χαρτί) αλλά μπορούν και να κατασκευαστούν εύκολα.

Ας πάρουμε για παράδειγμα την κανονική κατανομή $N(\mu, \sigma)$. Αν παραστήσουμε γραφικά τη συνάρτηση $F(x)$ με οριζόντιο άξονα $h = F$ και κατακόρυφο $v = x$, τότε θα προκύψει ένα γράφημα μορφής \mathcal{J} . Γνωρίζουμε όμως ότι $x = \mu + \sigma z_F$, όπου z_F το F -ποσοστημόριο της τυποποιημένης κανονικής κατανομής $N(0, 1)$. Κατά συνέπεια, αν θέσουμε στον οριζόντιο άξονα $h = z_F$, τότε η εξίσωση του γραφήματος θα είναι $v = \mu + \sigma h$, η οποία παριστάνει ευθεία. Αυτό ισοδυναμεί με το μετασχηματισμό του οριζόντιου άξονα $h = z_F = F_0^{-1}(F)$, όπου $F_0^{-1}(\cdot)$ η αντίστροφη της τυποποιημένης κανονικής συνάρτησης κατανομής. Με άλλους κατάλληλους μετασχηματισμούς του οριζόντιου ή/και του κατακόρυφου άξονα μπορούμε να πετύχουμε ευθειοποίηση των γραφημάτων άλλων συναρτήσεων κατανομής, όπως αναλυτικότερα θα δούμε στο κεφάλαιο 6.

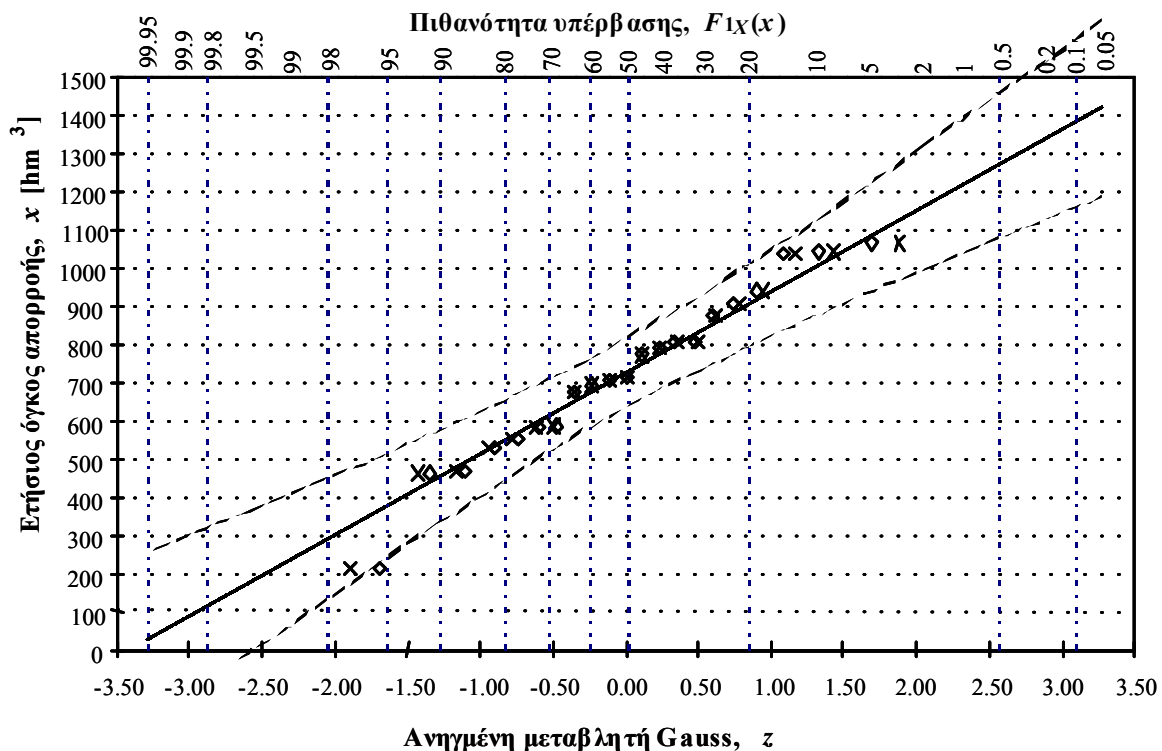
Επειδή υπάρχει ένα-προς-ένα αντιστοιχία μεταξύ των μεγεθών F και z_F , η βαθμονόμηση του οριζόντιου άξονα μπορεί να γίνει σε μονάδες F αντί z_F , πράγμα που διευκολύνει τη διαδικασία κατάρτισης της γραφικής παράστασης. Ακόμη, η βαθμονόμηση του οριζόντιου άξονα μπορεί να γίνει ως προς την πιθανότητα υπέρβασης $F_1 = 1 - F$ ή ως προς την περίοδο επαναφοράς $T = 1 / F_1$. Ένα παράδειγμα χαρτιού κανονικής κατανομής φαίνεται στο Σχ. 5.3, όπου μάλιστα φαίνονται ταυτόχρονα δύο βαθμονομήσεις του οριζόντιου άξονα, ως προς z_F και ως προς F_1 .

Η γραφική παράσταση του σημειοσυνόλου της εμπειρικής συχνότητας υπέρβασης $(x_{(i)}, q_i)$ σε χαρτί κανονικής κατανομής (δηλαδή $h = z_{1-q_i}$, $v = x_{(i)}$) θα δίνει μια περίπου ευθύγραμμη διάταξη των σημείων αν η κατανομή του πληθυσμού X είναι κανονική. Έτσι η γραφική παράσταση αυτή μας προσφέρει ένα γραφικό τρόπο ελέγχου της κανονικότητας ως προς την κατανομή ενός δείγματος. Τα παραπάνω διευκρινίζονται με το παράδειγμα που ακολουθεί.

Εφαρμογή 5.3

Να υπολογιστούν και να σχεδιαστούν σε χαρτί κανονικής κατανομής οι εμπειρικές πιθανότητες υπέρβασης για το δείγμα της Εφαρμογής 5.1.

Για τον υπολογισμό των εμπειρικών πιθανοτήτων υπέρβασης θα χρησιμοποιήσουμε τους τύπους Weibull (αμερόληπτη εκτίμηση πιθανότητας υπέρβασης) και Blom (αμερόληπτη εκτίμηση ποσοστημορίων κανονικής κατανομής, βλ. και Πίν. 5.8). Οι υπολογισμοί είναι απλούστατοι και προϋποθέτουν την κατάταξη του δείγματος σε φθίνουσα σειρά (βλ. Πίν. 5.5). Πινακοποίησή τους δίνεται στον Πίν. 5.9. Οι δύο τελευταίες στήλες κανονικά δεν είναι απαραίτητες αλλά έχουν συμπεριληφθεί για λόγους πληρότητας. Χρειάζονται μόνο στην περίπτωση που δεν διατίθεται κατάλληλα γραμμογραφημένο χαρτί κανονικής κατανομής (όπως αυτό του Σχ. 5.3), οπότε απεικονίζονται οι παρατηρημένες τιμές της μεταβλητής x_i συναρτήσει των τιμών της τυποποιημένης κανονικής μεταβλητής z_{1-q_i} . Οι τελευταίες είτε βρίσκονται από τον πίνακα της κανονικής κατανομής (Πίν. Π1), είτε υπολογίζονται με αριθμητικές μεθόδους (βλ. εδάφιο 6.1.1 και Παράρτημα 6.Α). Η απεικόνιση των εμπειρικών πιθανοτήτων υπέρβασης του δείγματος έγινε στο Σχ. 5.3. Για λόγους σύγκρισης έχει χαραχτεί και η θεωρητική συνάρτηση κανονικής κατανομής (βλ. ενότητα 5.5.4).



Σχ. 5.3 Εμπειρική συνάρτηση κατανομής κατά Weibull (ρόμβοι) και Blom (σύμβολα \times) σε χαρτί κανονικής κατανομής, για το παράδειγμα της Εφαρμογής 5.3. Με συνεχή γραμμή απεικονίζεται η κανονική συνάρτηση κατανομής $N(725, 211.5)$ και με διακεκομμένες τα αντίστοιχα όρια εμπιστοσύνης 95% (βλ. και Εφαρμογή 5.4 καθώς και Εφαρμογή 5.6).

Πίν. 5.9 Υπολογισμοί εμπειρικών πιθανοτήτων υπέρβασης για την Εφαρμογή 5.3.

A/A	Τιμή μεταβλητής	Εμπειρική πιθανότητα υπέρβασης		Τιμή τυποποιημένης κανονικής μεταβλητής	
		Weibull	Blom	Weibull	Blom
i	x_i	$q_i = \frac{i}{n+1}$	$q_i = \frac{i-0.375}{n+0.25}$	z_{1-q_i}	z_{1-q_i}
1	1064	0.045	0.029	1.691	1.890
2	1042	0.091	0.076	1.335	1.429
3	1037	0.136	0.124	1.097	1.158
4	942	0.182	0.171	0.908	0.952
5	906	0.227	0.218	0.748	0.780
6	874	0.273	0.265	0.605	0.629
7	807	0.318	0.312	0.473	0.491
8	807	0.364	0.359	0.349	0.362
9	788	0.409	0.406	0.230	0.238
10	772	0.455	0.453	0.114	0.118
11	715	0.500	0.500	0.000	0.000
12	705	0.545	0.547	-0.114	-0.118
13	695	0.591	0.594	-0.230	-0.238
14	674	0.636	0.641	-0.349	-0.362
15	588	0.682	0.688	-0.473	-0.491
16	580	0.727	0.735	-0.605	-0.629
17	552	0.773	0.782	-0.748	-0.780
18	529	0.818	0.829	-0.908	-0.952
19	469	0.864	0.876	-1.097	-1.158
20	462	0.909	0.924	-1.335	-1.429
$n=21$	217	0.955	0.971	-1.691	-1.890

5.4 Επιλογή και προσαρμογή θεωρητικής συνάρτησης κατανομής

Αφού εκτελέσουμε τα προηγούμενα, τα οποία αφορούν στη συμπύκνωση του δείγματος, είμαστε σε θέση να κάνουμε και κάποιες στατιστικές εκτιμήσεις ή προγνώσεις. Για παράδειγμα από ένα δείγμα ετήσιων παροχών μήκους 20 ετών μπορούμε να βρούμε τη μέση υπερετήσια παροχή (ίση με το μέσο όρο του δείγματος) ή την παροχή περιόδου επαναφοράς 10

ετών (περίπου ίση με τη δεύτερη σε μέγεθος τιμή του δείγματος), κτλ. Ωστόσο η εμβέλεια αυτών των εκτιμήσεων ή προγνώσεων είναι περιορισμένη. Έτσι δεν μπορούμε, στο παράδειγμά μας, να εκτιμήσουμε την παροχή περιόδου επαναφοράς 100 ετών, αφού δεν υπάρχει τρόπος επέκτασης έξω από τα όρια των 20 ετών που καλύπτει το δείγμα. Επίσης δεν μπορούμε να κάνουμε εκτιμήσεις διαστήματος για τα μεγέθη που μας ενδιαφέρουν.

Τα παραπάνω προβλήματα δεν θα υπήρχαν αν ξέραμε τη συνάρτηση κατανομής του πληθυσμού. Γενικά, η κατανομή του πληθυσμού μπορεί να είναι οποιαδήποτε συνάρτηση που υπακούει στους περιορισμούς (2.4). Η ακριβής γνώση της συνάρτησης κατανομής του πληθυσμού θα απαιτούσε να έχουμε μετρήσει τον ίδιο τον πληθυσμό στο σύνολό του, ή πρακτικώς να έχουμε ένα δείγμα πολύ μεγαλύτερο από τη χρονική εμβέλεια των εκτιμήσεων που πρόκειται να κάνουμε. Βεβαίως αυτό δεν μπορεί ποτέ να πραγματοποιηθεί, και έτσι η λύση που μένει είναι να υποθέσουμε ένα πιθανοτικό μοντέλο για τον πληθυσμό. Με τον όρο πιθανοτικό μοντέλο εννοούμε μια από τις τυπικές συναρτήσεις κατανομής της θεωρίας πιθανοτήτων που έχουν ένα συγκεκριμένο και σχετικώς απλό μαθηματικό τύπο. Το πιο τυπικό παράδειγμα είναι η κανονική κατανομή με συνάρτηση που δίνεται από τη (2.61). Άλλα παραδείγματα θα δοθούν στο κεφάλαιο 6. Βεβαίως, η χρήση ενός πιθανοτικού μοντέλου αποτελεί πάντα μια προσέγγιση της πραγματικότητας. Οι κατανομές των υδρολογικών μεταβλητών ποτέ δεν ταυτίζονται με τις απλές κατανομές της θεωρίας πιθανοτήτων.

Στην επιλογή του κατάλληλου μοντέλου μας οδηγούν τα ακόλουθα:

1. *Η πιθανοθεωρία.* Σε μερικές περιπτώσεις υπάρχουν θεωρητικοί λόγοι για τους οποίους μια συγκεκριμένη υδρολογική μεταβλητή αναμένεται να ακολουθεί συγκεκριμένο τύπο κατανομής. Για παράδειγμα, σύμφωνα με το κεντρικό οριακό θεώρημα, η ετήσια βροχόπτωση μιας περιοχής με πλούσιες βροχοπτώσεις αναμένεται να ακολουθεί κανονική κατανομή (βλ. και εδάφιο 2.8.1).
2. *Η γενική υδρολογική εμπειρία.* Σε πολλές περιπτώσεις η υδρολογική εμπειρία έχει δείξει ότι συγκεκριμένες υδρολογικές μεταβλητές ακολουθούν συγκεκριμένους τύπους κατανομής, αν και δεν υπάρχουν σαφείς πιθανοθεωρητικοί λόγοι που να οδηγούν στις τελευταίες. Για

παράδειγμα, οι μηνιαίες απορροές πολύ συχνά ακολουθούν κατανομές γάμα ή λογαριθμοκανονικές (βλ. κεφάλαιο 6).

3. *Η εξέταση του συγκεκριμένου δείγματος.* Τα στατιστικά χαρακτηριστικά του διαθέσιμου ιστορικού δείγματος μας βοηθούν στην επιλογή ή την απόρριψη συγκεκριμένου τύπου κατανομής. Για παράδειγμα, αν ο δειγματικός συντελεστής ασυμμετρίας έχει τιμή κοντά στο μηδέν, τότε μπορεί να επιλεγεί η κανονική (ή κάποια άλλη συμμετρική) κατανομή. Αντίθετα, αν ο συντελεστής ασυμμετρίας διαφέρει αρκετά από το μηδέν, θα πρέπει να αποκλειστεί η χρήση της κανονικής κατανομής.

Βεβαίως, η καταλληλότητα ενός συγκεκριμένου τύπου κατανομής δεν εξασφαλίζεται με τα παραπάνω κριτήρια, τα οποία αποτελούν απλώς ενδείξεις καταλληλότητας. Ο ουσιαστικός έλεγχος της καταλληλότητας της κατανομής γίνεται εκ των υστέρων. Αφού υπολογιστούν και οι παράμετροί της, ελέγχεται ο βαθμός προσαρμογής της στην εμπειρική συνάρτηση κατανομής του δείγματος. Αυτό μπορεί να γίνει κατ' αρχήν εμπειρικά, με βάση τη γραφική απεικόνιση της εμπειρικής και θεωρητικής συνάρτησης κατανομής στο κατάλληλο χαρτί πιθανότητας. Όμως, τα πλέον αξιόπιστα συμπεράσματα επιτυγχάνονται με τη βοήθεια των τυπικών στατιστικών ελέγχων, όπως διεξοδικά αναλύεται στην επόμενη ενότητα.

Ενδείξεις καταλληλότητας της κανονικής κατανομής για υδρολογικές μεταβλητές

Μέχρι τώρα έχουμε αναφερθεί αρκετές φορές σε ενδεικτικούς ελέγχους της καταλληλότητας της κανονικής κατανομής για την περιγραφή μιας υδρολογικής μεταβλητής. Παρακάτω κωδικοποιούμε όλες τις ενδείξεις καταλληλότητας της κανονικής κατανομής.

1. *Αξιοποίηση κεντρικού οριακού θεωρήματος.* Ελέγχεται κατά πόσον η υπό έλεγχο μεταβλητή μπορεί να θεωρηθεί ως άθροισμα πολλών επιμέρους συνιστωσών, οι οποίες θα πρέπει να εκπληρώνουν (έστω προσεγγιστικά) τις προϋποθέσεις του κεντρικού οριακού θεωρήματος. Το κριτήριο αυτό είναι θεωρητικό και δεν απαιτεί αριθμητικούς υπολογισμούς
2. *Αριθμητικός έλεγχος με βάση το συντελεστή ασυμμετρίας.* Αν ο δειγματικός συντελεστής ασυμμετρίας είναι περίπου μηδενικός,

αυτό αποτελεί μια σοβαρή ένδειξη για την καταλληλότητα της κανονικής κατανομής.

3. *Αριθμητικός έλεγχος με βάση το συντελεστή μεταβλητότητας.* Έστω X μια τυχούσα υδρολογική μεταβλητή. Κατά κανόνα η X παίρνει μόνο θετικές τιμές ή μηδέν, γιατί οι αρνητικές τιμές δεν έχουν φυσικό νόημα. Αντίθετα, η κανονική κατανομή προβλέπει ότι η X μπορεί να πάρει και αρνητικές τιμές. Θεωρητικά, λοιπόν, η κανονική κατανομή δεν μπορεί να αντιπροσωπεύει υδρολογικές μεταβλητές, παρά μόνο προσεγγιστικά. Για να είναι ικανοποιητική η προσέγγιση θα πρέπει η πιθανότητα $P(X < 0)$ να είναι πολύ μικρή, ώστε να μπορεί να αγνοηθεί, ήτοι $P(X < 0) \leq q$ όπου q μια αποδεκτά μικρή τιμή πιθανότητας, π.χ. $q < 0.02$. Αν $Z = (X - \mu_X) / \sigma_X$ είναι η αντίστοιχη τυποποιημένη κανονική μεταβλητή τότε θα πρέπει $P(Z < -\mu_X / \sigma_X) \leq q$. Αν z_q είναι το q -ποσοστημόριο της τυποποιημένης κανονικής κατανομής, τότε ισοδύναμα $C_{vX} = \sigma_X / \mu_X \leq -1/z_q$. Για $q = 0.02$ έχουμε $z_q \approx -2$ οπότε $C_{vX} \leq 0.5$. Αντίστοιχα, για $q = 0.00005$ έχουμε $z_q \approx -4$ οπότε $C_{vX} \leq 0.25$. Καταλήγουμε λοιπόν στο συμπέρασμα ότι αν $C_{vX} \leq 0.25$ έχουμε μια πολύ σοβαρή ένδειξη καταλληλότητας της κανονικής κατανομής. Αν $C_{vX} > 0.5$ τότε θα πρέπει να αποκλειστεί η χρήση της κανονικής κατανομής. Για ενδιάμεσες τιμές του συντελεστή μπορεί να είναι αποδεκτή η κανονική κατανομή αλλά με μικρότερο βαθμό προσέγγισης.
4. *Γραφικός έλεγχος με βάση τη συνοπτική παράσταση δείγματος.* Όπως αναφέρθηκε στο εδάφιο 5.1 τυχόν συμμετρικό διάγραμμα της συνοπτική παράστασης δείγματος, χωρίς αδικαιολόγητα μεγάλο αριθμό εξωτερικών ή μακρινών τιμών, αποτελεί ένδειξη καταλληλότητας της κανονικής κατανομής
5. *Γραφικός έλεγχος με βάση την εμπειρική συνάρτηση κατανομής.* Η ευθύγραμμη διάταξη της σημειοσειράς της εμπειρικής συνάρτησης κατανομής, σχεδιασμένης σε χαρτί κανονικής κατανομής, αποτελεί σοβαρή ένδειξη της καταλληλότητας της κανονικής κατανομής.

Τονίζουμε και πάλι ότι τα κριτήρια αυτά είναι απλές ενδείξεις και δεν μπορούν να θεωρηθούν ως στατιστική απόδειξη της καταλληλότητας της κανονικής κατανομής.

Εφαρμογή 5.4

Να προσαρμοστεί η κανονική κατανομή στο δείγμα της Εφαρμογής 5.1.

Η λύση αυτού του προβλήματος είναι πολύ απλή. Οι παράμετροι της κατανομής είναι (βλ. Εφαρμογή 5.1) $\mu = \bar{x} = 725.0$, $\sigma = s_X = 211.5$ (αποδεκτή είναι και η τιμή $\sigma = s_X^* = 216.7$). Η κανονική κατανομή με αυτές τις παραμέτρους έχει απεικονιστεί στο Σχ. 5.3 και η αντίστοιχη πυκνότητα πιθανότητας στο Σχ. 5.2.

Ο αναγνώστης μπορεί να διαπιστώσει ότι στο παράδειγμα που εξετάζουμε υπάρχουν όλες οι ενδείξεις καταλληλότητας της κανονικής κατανομής που αναφέρθηκαν στο παραπάνω ένθετο εδάφιο. Στην Εφαρμογή 5.5 θα δώσουμε τον πληρέστερο στατιστικό έλεγχο της καταλληλότητας της κανονικής κατανομής.

5.5 Έλεγχος προσαρμογής συνάρτησης κατανομής

Αφού υιοθετήσουμε μια συγκεκριμένη συνάρτηση κατανομής για την περιγραφή μιας φυσικής μεταβλητής και υπολογίσουμε τις παραμέτρους της, ελέγχουμε στη συνέχεια την προσαρμογή της κατανομής αυτής στο συγκεκριμένο δείγμα που διαθέτουμε. Ο έλεγχος βασίζεται στη στατιστική θεωρία ελέγχου υποθέσεων, την οποία περιγράψαμε πολύ συνοπτικά στην ενότητα 3.5. Έχουν αναπτυχθεί διάφοροι τύποι στατιστικών δοκιμών που μπορούν να εφαρμοστούν στον έλεγχο προσαρμογής μιας κατανομής. Θα αναφερθούμε στην πιο κλασική από αυτές, τη δοκιμή χ^2 . Άλλες στατιστικές δοκιμές με σημαντική εφαρμογή στην τεχνική υδρολογία είναι η δοκιμή *Kolmogorov-Smirnov* (βλ. π.χ. Benjamin and Cornell, 1970, σ. 466· Kottegoda, 1980, σ. 89) και η πιο πρόσφατη δοκιμή του συντελεστή συσχέτισης εμπειρικής πιθανότητας (βλ. π.χ. Stedinger et al., 1993, σ. 18.27).

5.5.1 Η δοκιμή χ^2 .

Η δοκιμή χ^2 στηρίζεται στη σύγκριση της θεωρητικής συνάρτησης κατανομής με την κλασική εμπειρική συνάρτηση κατανομής. Η σύγκριση γίνεται σε ένα πεπερασμένο σύνολο επιλεγμένων σημείων x_j , και όχι σε κάθε παρατηρημένη τιμή x_i του δείγματος. Η μηδενική υπόθεση H_0 και η εναλλακτική H_1 είναι

$$H_0: F(x_j) = F_0(x_j) \text{ για κάθε } j \quad H_1: F(x_j) \neq F_0(x_j) \text{ για κάποια } j \quad (5.11)$$

όπου $F(x)$ η άγνωστη πραγματική συνάρτηση κατανομής και $F_0(x)$ η κατανομή που έχει υποτεθεί. Η $F_0(x)$ μπορεί να είναι γνωστή πλήρως, και ως προς τη μαθηματική της μορφή και ως προς τις τιμές των παραμέτρων της, πριν από την εξέταση του συγκεκριμένου δείγματος. Στην περίπτωση αυτή η μηδενική υπόθεση λέγεται *τέλεια*. Συνηθέστερα όμως, οι τιμές των παραμέτρων υπολογίζονται από το δείγμα, οπότε μιλούμε για *ατελή μηδενική υπόθεση*.

Τα σημεία ελέγχου $x_j, j = 0, \dots, k$ χωρίζουν το πεδίο μεταβολής της τυχαίας μεταβλητής σε k κλάσεις, δηλαδή διαστήματα της μορφής $(x_0, x_1], (x_1, x_2], \dots, (x_{k-1}, x_k]$. Για τη δεδομένη συνάρτηση κατανομής $F_0(x)$ η πιθανότητα να βρεθεί ένα τυχαία επιλεγμένο σημείο στην κλάση $K_j = (x_{j-1}, x_j]$ προφανώς είναι

$$p_j = F_0(x_j) - F_0(x_{j-1}) \quad (5.12)$$

και κατά συνέπεια το αναμενόμενο πλήθος σημείων του δείγματος που θα βρίσκονται στην εν λόγω κλάση είναι $l_j = n p_j$, όπου n το μέγεθος του δείγματος. Το μέγεθος l_j λέγεται θεωρητικό δυναμικό της κλάσης, ενώ το πραγματικό πλήθος των σημείων του δείγματος, n_j , που βρίσκονται μέσα στα όρια της κλάσης, λέγεται πραγματικό δυναμικό της κλάσης. Προφανώς, η μικρή απόσταση μεταξύ των μεγεθών n_j και l_j , δηλαδή η μικρή διαφορά $|n_j - n p_j|$, αποτελεί συνηγορία υπέρ της καταλληλότητας της κατανομής $F_0(x_j)$ και της μη απόρριψης της μηδενικής υπόθεσης. Η *στατιστική συνάρτηση του Pearson* που ορίζεται από την εξίσωση

$$q := \sum_{j=1}^k \frac{(n_j - n p_j)^2}{n p_j} \quad (5.13)$$

συνεκτιμά τις διαφορές πραγματικού και θεωρητικού δυναμικού σε όλες τις κλάσεις. Αν η μηδενική υπόθεση είναι τέλεια, τότε η κατανομή της είναι η χ^2 με $k - 1$ βαθμούς ελευθερίας. Στη συνηθέστερη περίπτωση της ατελούς μηδενικής υπόθεσης, ο αριθμός των βαθμών ελευθερίας είναι

$k - r - 1$, όπου r είναι ο αριθμός των παραμέτρων που εκτιμώνται από το δείγμα.*

Η παραλλαγή της δοκιμής χ^2 που αναπτύσσουμε εδώ είναι αυτή των *ισοπίθανων κλάσεων*, σύμφωνα με την οποία το θεωρητικό δυναμικό είναι ίδιο σε όλες τις κλάσεις. Σε αυτή την περίπτωση η εξίσωση (5.13) απλοποιείται σε

$$q := \frac{k}{n} \sum_{j=1}^k n_j^2 - n \quad (5.14)$$

Η έκδοση αυτή έχει το πλεονέκτημα ότι δίνει μονοσήμαντα καθορισμένα όρια κλάσεων για δεδομένο αριθμό κλάσεων k . Για την επιλογή του αριθμού των κλάσεων k ισχύουν οι ακόλουθοι δύο αλληλοαντικρουόμενοι κανόνες:

- Υποχρεωτικά θα πρέπει $k \geq r + 2$, όπου r είναι ο αριθμός των παραμέτρων της κατανομής που εκτιμώνται από το δείγμα.
- Γενικά συνιστάται (βλ. π.χ. Benjamin and Cornell, 1970, σ. 465· Kottegoda, 1980, σ. 88) το θεωρητικό δυναμικό των κλάσεων είναι μεγαλύτερο από 5, πράγμα που συνεπάγεται ότι $k \leq n / 5$.

Για μικρά δείγματα είναι δυνατό οι κανόνες αυτοί να μη μπορούν να ικανοποιηθούν ταυτόχρονα, οπότε ικανοποιούμε μόνο τον πρώτο.†

Ο αλγόριθμος εφαρμογής της δοκιμής χ^2 περιγράφεται από τα ακόλουθα βήματα:

1. Επιλέγουμε τον αριθμό των κλάσεων k σύμφωνα με τους παραπάνω κανόνες.

* Κανονικά, στην περίπτωση αυτή οι εκτιμήσεις των παραμέτρων πρέπει να γίνονται με τη μέθοδο της μέγιστης πιθανοφάνειας· ωστόσο, ο όρος αυτός δεν τηρείται στις εφαρμογές.

† Η επιλογή του κατάλληλου αριθμού κλάσεων γίνεται με τον τύπο (Mann and Wald, 1942· Williams, 1955· βλ. και Kottegoda, 1980, σ. 88):

$$k = 2^{1.2} [(n - 1) / z_{1-\alpha}]^{0.4}$$

όπου $z_{1-\alpha}$ το $(1 - \alpha)$ -ποσοστημόριο της κανονικής κατανομής και α το επίπεδο σημαντικότητας του ελέγχου. Οι Kendall and Stuart (1973, σ. 455) δίνουν αναλυτικότερο τρόπο υπολογισμού του αριθμού κλάσεων, ο οποίος όμως ισχύει για μεγάλα δείγματα που σπάνια είναι διαθέσιμα στην τεχνική υδρολογία.

2. Χωρίζουμε το διάστημα πιθανότητας $[0, 1]$ σε k ίσα υποδιαστήματα με όρια $u_j = j / k$, ($j = 0, \dots, k$).
3. Υπολογίζουμε τα όρια x_j των κλάσεων (η τιμή x_j είναι το u_j -ποσοστημόριο της μεταβλητής).
4. Υπολογίζουμε το πραγματικό δυναμικό n_j της κάθε κλάσης (το βήμα αυτό απλοποιείται αν το δείγμα έχει προηγουμένως καταταγεί σε φθίνουσα ή αύξουσα σειρά).
5. Από την (5.14) (ή την (5.13)) υπολογίζουμε την τιμή της στατιστικής συνάρτησης q του Pearson.
6. Για ένα αποδεκτό επίπεδο σημαντικότητας α υπολογίζουμε την κρίσιμη τιμή της στατιστικής συνάρτησης ελέγχου $q_c = q_{1-\alpha}$. Για το σκοπό αυτό χρησιμοποιούμε την κατανομή χ^2 για $k - r - 1$ βαθμούς ελευθερίας όπου r είναι ο αριθμός των παραμέτρων της κατανομής που εκτιμώνται από το δείγμα (βλ. Πίν. Π2 στο Παράρτημα).
7. Απορρίπτουμε τη μηδενική υπόθεση αν $q > q_c$.

Ο αλγόριθμος διευκρινίζεται στο παράδειγμα που ακολουθεί.

Εφαρμογή 5.5

Να ελεγχθεί η καταλληλότητα της κανονικής κατανομής για το δείγμα της Εφαρμογής 5.1.

Στο συγκεκριμένο δείγμα προσαρμόστηκε ήδη (βλ. Εφαρμογή 5.4) η κανονική κατανομή με παραμέτρους $\mu = \bar{x} = 725.0$, $\sigma = s_X = 211.5$.

Ο αριθμός των παραμέτρων της κατανομής είναι $r = 2$ και το μέγεθος του δείγματος $n = 21$. Σύμφωνα με τα παραπάνω, ο αριθμός των κλάσεων k για τη δοκιμή χ^2 θα πρέπει να ικανοποιεί τις σχέσεις

$$k \geq 2 + 2 = 4, \quad k \leq 21 / 5 = 4.2$$

οι οποίες συναληθεύουν για $k = 4$. Κατά συνέπεια επιλέγουμε $k = 4$.

Οι υπολογισμοί των βημάτων 2-4 του παραπάνω αλγορίθμου κωδικοποιούνται στον Πίν. 5.10. Για τον υπολογισμό των ορίων της μεταβλητής χρησιμοποιήσαμε τον Πίν. Π1 (βλ. Παράρτημα) της κανονικής κατανομής. Για παράδειγμα, το άνω όριο της πρώτης κλάσης είναι

$$x_1 = 725.0 - 0.675 \times 211.5 = 528.3$$

Στο Σχ. 5.4 δίνουμε για διδακτικούς λόγους γραφική απεικόνιση των κλάσεων και του πραγματικού δυναμικού τους πάνω σε χαρτί κανονικής κατανομής.

Πίν. 5.10 Βασικοί υπολογισμοί για τη δοκιμή χ^2 της Εφαρμογής 5.5.

Κλάση	1	2	3	4	
Όρια πιθανότητας	0	0.25	0.5	0.75	1.0
Όρια μεταβλητής	$-\infty$	582.3	725.0	867.7	$+\infty$
Πραγματικό δυναμικό	6	5	4	6	

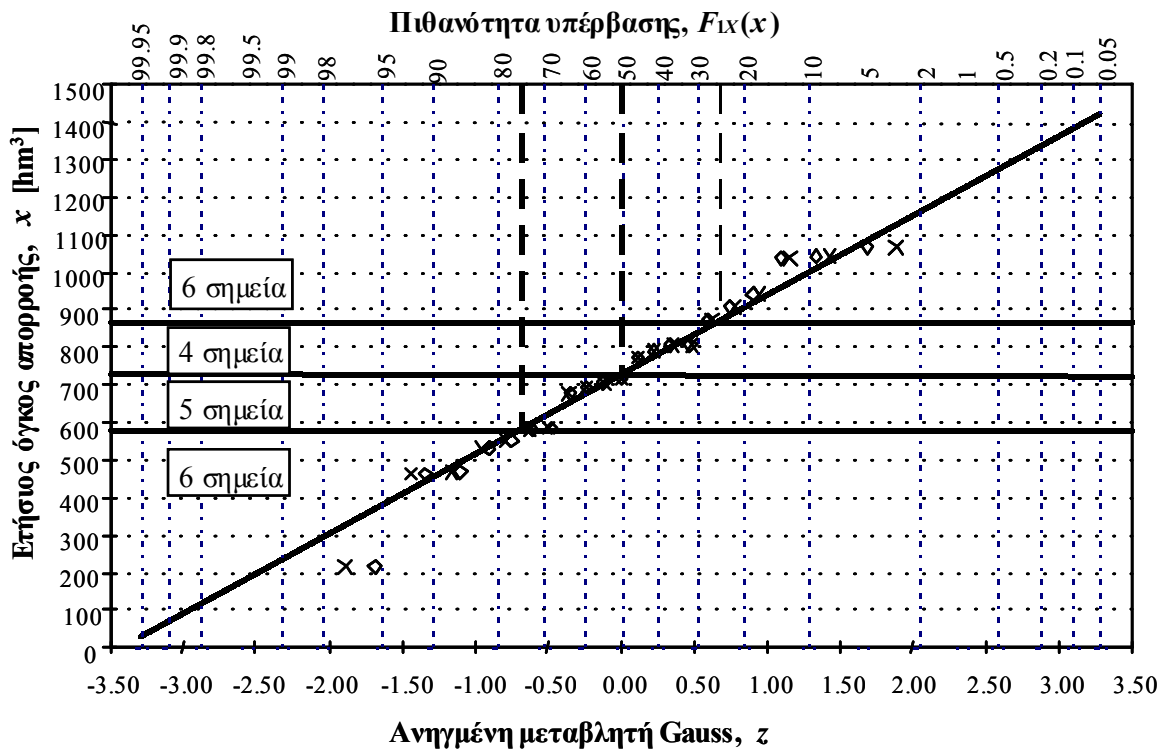
Από την (5.14) παίρνουμε

$$q = (4/21) \times (6^2 + 5^2 + 4^2 + 6^2) - 21 = 0.52$$

Εξ άλλου για επίπεδο σημαντικότητας $\alpha = 0.05$ η κρίσιμη τιμή της μεταβλητής είναι

$$q_c = \chi_{0.95}^2(1) = 3.84$$

(όπως προκύπτει από τον Πίν. Π2 για $u = 1 - \alpha = 0.95$ και βαθμούς ελευθερίας $= 4 - 2 - 1 = 1$). Κατά συνέπεια ισχύει $q < q_c$ και δεν απορρίπτεται η κανονική κατανομή.



Σχ. 5.4 Επεξηγηματικό σκαρίφημα για το παράδειγμα της Εφαρμογής 5.5.

5.6 Στατιστική πρόγνωση

Η στατιστική πρόγνωση στην τεχνική υδρολογία αφορά στην εκτίμηση της τιμής ενός υδρολογικού μεγέθους για δεδομένη πιθανότητα υπέρβασης (ή περίοδο επαναφοράς). Η πρόγνωση αυτή είναι αρκετά εύκολη από υπολογιστική άποψη, εφόσον έχει ήδη υιοθετηθεί και προσαρμοστεί ένα συγκεκριμένο πιθανοτικό μοντέλο για το μέγεθος που ενδιαφέρει. Για το σκοπό αυτό εφαρμόζονται όσα έχουν αναφερθεί στο κεφάλαιο 3. Ειδικότερα, η πρόγνωση μπορεί να είναι είτε σημειακή, είτε διαστήματος, όπως αναλυτικά φαίνεται στο παράδειγμα που ακολουθεί.

Εφαρμογή 5.6

Να υπολογιστεί ο μέγιστος και ο ελάχιστος ετήσιος όγκος απορροής εκατονταετίας στη λεκάνη του ποταμού Ευήνου ανάντη του υδρομετρικού σταθμού Πόρος Ρηγανίου, καθώς και τα όρια εμπιστοσύνης του 95% (βλ. και προηγούμενες Εφαρμογές αυτού του κεφαλαίου).

Εφαρμόζουμε την ίδια διαδικασία που είχαμε εφαρμόσει στην Εφαρμογή 3.3.4. Σημειώνουμε ότι το μέγεθος του δείγματος είναι πολύ μικρό σε σχέση με τον υπολογιστικό ορίζοντα της εκατονταετίας και έτσι δεν περιμένουμε πολύ καλό βαθμό ακρίβειας στους υπολογισμούς μας.

Θα υπολογίσουμε πρώτα τις σημειακές εκτιμήσεις. Για τον μέγιστο όγκο απορροής εκατονταετίας η πιθανότητα μη υπέρβασης είναι $u = 1 - 1/100 = 0.99$ και $z_u = 2.326$ (από τον Πίν. Π1 της κανονικής κατανομής του Παραρτήματος). Άρα η σημειακή εκτίμηση είναι

$$x_u = 725.0 + 2.326 \times 211.5 = 1216.9$$

Αντίστοιχα, για τον ελάχιστο όγκο απορροής εκατονταετίας η πιθανότητα μη υπέρβασης είναι $u = 1 / 100 = 0.01$ και $z_u = -2.326$, άρα

$$x_u = 725.0 - 2.326 \times 211.5 = 233.1$$

Προχωρούμε τώρα στον υπολογισμό των ορίων εμπιστοσύνης. Για $\gamma = 95\%$ και $z_{(1+\gamma)/2} = 1.96$, τα όρια για τον μέγιστο όγκο απορροής εκατονταετίας θα είναι:

$$\hat{x}_{u1} = 1216.9 - 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 1042.8$$

$$\hat{x}_{u2} = 1216.9 + 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 1391.0$$

Αντίστοιχα, τα όρια για τον ελάχιστο όγκο απορροής εκατονταετίας θα είναι:

$$\hat{x}_{u1} = 233.1 - 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 59.0$$

$$\hat{x}_{u2} = 233.1 + 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 407.2$$

Παρατηρούμε ότι ο μέγιστος και ο ελάχιστος όγκος απορροής που έχουν καταγραφεί στο δείγμα (Πίν. 5.5) βρίσκονται ανάμεσα στα αντίστοιχα όρια εμπιστοσύνης του μέγιστου και ελάχιστου όγκου εκατονταετίας. Ειδικότερα, ο ελάχιστος όγκος απορροής των 217 hm³ είναι μικρότερος από τη σημειακή πρόγνωση εκατονταετίας που δίνει η κανονική κατανομή (233.1 hm³), γεγονός που αντανακλά το σημαντικό μέγεθος της ξηρασίας του υδρολογικού έτους 1989-90.

Στο Σχ. 5.3 έχουμε απεικονίσει τα πιο πάνω αποτελέσματα, καθώς και μια σειρά σημειακών εκτιμήσεων και ορίων εμπιστοσύνης για άλλες τιμές της περιόδου επαναφοράς. Οι υπολογισμοί έγιναν με τον ίδιο τρόπο όπως παραπάνω. Συνδέοντας τις τιμές των ορίων εμπιστοσύνης στο διάγραμμα αποκτούμε τις καμπύλες εμπιστοσύνης 95% της κατανομής.