Παράρτημα 3: Βασική Στατιστική στην Υδρολογία

Συχνά, οι υδρολόγοι διαθέτουν μόνο *σημειακές* μετρήσεις ποσοτήτων όπως η βροχόπτωση, οι οποίες διαφέρουν φανερά, τόσο χρονικά όσο και τοπικά. Λόγω της συνθετότητας των διεργασιών όπως η βροχόπτωση συχνά πρέπει να υιοθετήσουμε μια πιθανολογική προσέγγιση στην μελέτη τους που θα βασίζεται στις παρατηρήσεις μάλλον παρά στην φυσική θεωρία ή την ντετερμινιστική προσέγγιση.

Σε μια πιθανολογική προσέγγιση, οι διεργασίες θεωρούνται σαν τυχαίες, ή σαν να έχουν ένα στοιχείο αβεβαιότητας ή μη προβλεψιμότητας. Τα πιο κοινά παραδείγματα που χρησιμοποιούνται στην θεωρία των πιθανοτήτων είναι η ρίψη ενός ζαριού ή ενός νομίσματος. Σε αυτές τις περιπτώσεις, το αποτέλεσμα μιας τυχαίας διαδικασίας είναι ένας αριθμός (1 έως 6 στη κάθε ζαριά) ή "κορώνα" ή "γράμματα". Ο αριθμός των κουκίδων στην πλευρά ενός κύβου και οι "κορώνες" ή τα "γράμματα" είναι παραδείγματα *διακριτών τυχαίων μεταβλητών*, γιατί μόνο συγκεκριμένες τιμές των μεταβλητών μπορούν να σημειωθούν. Πολλές υδρολογικές μετρήσεις ή ποσότητες (πχ μέση ετήσια βροχόπτωση) θεωρούνται *συνεχείς τυχαίες μεταβλητές* γιατί μπορούν να πάρουν οποιαδήποτε (λογική) τιμή.

Δείτε τα δεδομένα στον πίνακα Α3.1. Φαίνεται να υπάρχει μια γενική τάση της ετήσιας βροχόπτωσης να βρίσκεται κοντά στο ένα μέτρο, με δύο "ακραίες" τιμές, 780 και 1192 mm. Πως να χρησιμοποιήσουμε αυτή την πληροφορία για να κάνουμε μια πρόβλεψη για τις μελλοντικές βροχοπτώσεις? Με συνεχείς τυχαίες μεταβλητές όπως είναι η ετήσια βροχόπτωση, η ερώτηση που θα μπορούσαμε να ρωτήσουμε είναι: Ποιά είναι η πιθανότητα η βροχόπτωση να είναι μεγαλύτερη από 1100 mm? Για να απαντήσουμε πρέπει να δούμε πως "κατανέμεται" η πιθανότητα, ή τι πιθανότητες συνδέονται με κάθε περιοχή (εύρος) αναμενόμενων τιμών.

Έτος	Ετήσια βροχόπτωση (mm)
1975	1020
1976	987
1977	894
1978	1040
1979	995
1980	780
1981	1004
1982	930
1983	1192
1984	950

Πίνακας Α3.1 Ετήσια βροχόπτωση για μια δεκαετή περίοδο

Οι συνεχείς τυχαίες μεταβλητές περιγράφονται με την συνάρτηση πυκνότητας πιθανότητας ή *συνάρτησης πιθανότητας* (pdf – probability density function), $f(x)$. Η συνάρτηση αθροιστικής κατανομής πιθανότητας ή *συνάρτηση κατανομής* (cdf – cumulative density function), $F(a)$, είναι το ολοκλήρωμα (ή η παράγουσα) της συνάρτησης πυκνότητας πιθανότητας:

$$F(a) = P(x \le a) = \int_{-\infty}^{a} f(x)dx$$

(A3.1)

Με άλλα λόγια, η *συνάρτηση κατανομής* για μια δεδομένη τιμή, *a* (1100 mm, στο παραπάνω παράδειγμα), είναι η πιθανότητα ότι το αποτέλεσμα μιας τυχαίας διεργασίας, *x* (η βροχόπτωση την επόμενη χρονιά) θα είναι *μικρότερο απο ή ίσο με* την δεδομένη τιμή. Παρατηρήστε ότι τα όρια της ολοκλήρωσης είναι το μείον άπειρο και η τιμή *a*; η ολοκλήρωση αθροίζει όλες τις πιθανότητες που συνδέονται με όλες τις τιμές τις μικρότερες από, και συμπεριλαμβανόμενης της, *a*.

Αντίστροφα, η *συμπληρωματική συνάρτηση κατανομής*, που συχνά συμβολίζεται σαν *G*(*a*), ορίζεται σαν:

$$G(a) = P(x \ge a) = 1 - F(a) = \int_{a}^{\infty} f(x)dx$$

(A3.2)

η οποία στο παράδειγμα, δίνει την πιθανότητα ότι η βροχόπτωση θα είναι μεγαλύτερη από **1100 mm**. Αναφερόμαστε σε αυτή την πιθανότητα σαν *πιθανότητα υπέρβασης*. Από τους παραπάνω ορισμούς προκύπτουν διάφορες σχέσεις:

$$P(a \le x \le b) = F(b) - F(a) = \int_{a}^{b} f(x)dx$$

(A3.3)

$$P(x = a) = \int_{a}^{a} f(x)dx \equiv 0$$

(A3.4)

$$P(-\infty \le x \le \infty) = \int_{-\infty}^{\infty} f(x)dx \equiv 1$$

(A3.5)

Η εξίσωση (A3.3) δίνει την πιθανότητα του αποτελέσματος μιας τυχαίας διεργασίας να βρίσκεται σε μια περιοχή τιμών. Η εξίσωση (A3.4) απλά δείχνει ότι με τις συνεχείς τυχαίες μεταβλητές, η πιθανότητα να εμφανιστεί μια συγκεκριμένη τιμή είναι μηδέν: γι' αυτό πάντα αναφερόμαστε στην πιθανότητα ενός αποτελέσματος να είναι μεγαλύτερο από, ή μικρότερο από, ή να βρίσκεται σε μια περιοχή τιμών. Η εξίσωση (A3.5) δείχνει ότι η συνολική πιθανότητα είναι ίση με 1 (η ετήσια βροχόπτωση πρέπει να έχει *κάποια* τιμή).

Η κατανομή ενός *δείγματος* τιμών μιας τυχαίας μεταβλητής μπορεί να πάρει μια ποικιλία από μορφές και συχνά μπορεί να εκφραστεί από σχετικά απλές συναρτήσεις πυκνότητας πιθανότητας. Τα παραδείγματα περιλαμβάνουν την κανονική (normal) κατανομή, την λογαριθμοκανονική (log-normal) κατανομή και την εκθετική (exponential) κατανομή.

Η *κανονική κατανομή* ορίζεται από δύο παραμέτρους, τον *μέσο* (μ_x) και την *τυπική απόκλιση* (σ_x). Αν έχουμε ένα δείγμα τιμών (πχ κάποιες δεκαετίες ετήσιας βροχόπτωσης), ο μέσος και η τυπική απόκλιση είναι περίπου $\bar{x}$ (για το μ_x) και s_x (για το σ_x):

(A3.6) $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

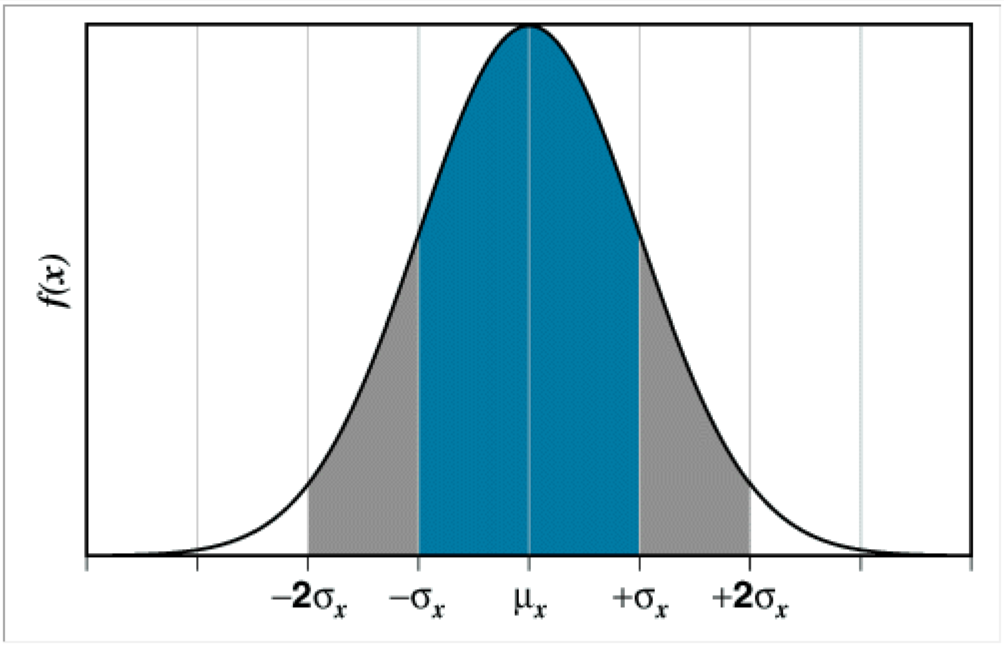
(A3.7) $$s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_x)^2}{n-1}}$$

όπου *n* είναι ο αριθμός των παρατηρήσεων. Για τα δεδομένα του πίνακα ο μέσος και η τυπική απόκλιση του δείγματος είναι αντίστοιχα 979 mm και 101 mm. Οι συνάρτηση πιθανότητας και η συνάρτηση κατανομής για την κανονική κατανομή είναι:

(A3.8) $$f(x;\mu_x,\sigma_x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$$

(A3.9) $$F(x;\mu_x,\sigma_x) = P(X \le x) = \int_{-\infty}^{x} f(x;\mu_x,\sigma_x)$$

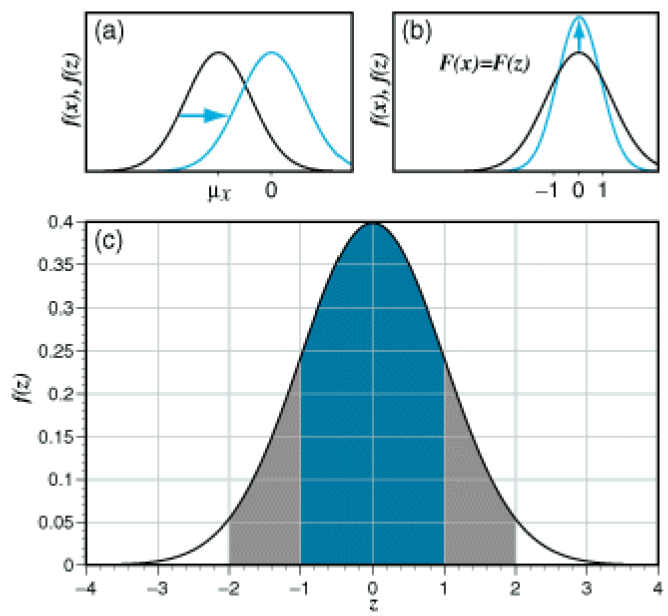
Η κανονική κατανομή είναι μια "καμπανόμορφη" καμπύλη. Ο μέσος δίνει ένα μέτρο της κεντρικής τάσης της μεταβλητής. Για δεδομένα που είναι κανονικά κατανεμημένα, ο μέσος ισούται με τον διάμεσο (την μεσαία τιμή, εάν όλες οι τιμές διαταχθούν από την μικρότερη προς την μεγαλύτερη ή ανάποδα). Η πιθανότητα ότι μια τιμή θα βρίσκεται εντός διαστήματος μιας τυπικής απόκλισης από τον μέσο είναι περίπου δύο-τρίτα (0.68) και η πιθανότητα ότι μια τιμή θα βρίσκεται εντός διαστήματος δύο τυπικών αποκλίσεων από το μέσο είναι περίπου 0.95 (δείτε τις γκρίζες και τις σκούρες μπλέ περιοχές του σχήματος A3.1).



Σχήμα Α3.1 Η κανονική κατανομή

Συχνά τα δεδομένα *τυποποιούνται* υπολογίζοντας τιμές z. Οι πίνακες της αθροιστικής κατανομής της *τυποποιημένης κανονικής κατανομής* (*standard normal distribution*) βασίζονται στις τιμές z. Η τυποποίηση των δεδομένων εμπεριέχει την μετάθεση ολόκληρης της κατανομής κατά την ποσότητα που απαιτείται για να φέρει τον μέσο στη τιμή μηδέν, και μετά την συρρίκνωση ή το άπλωμα της κατανομής έτσι ώστε η τυπική απόκλιση να είναι 1. Οι τιμές της συνάρτησης κατανομής ή της συνάρτησης της συμπληρωματικής κατανομής της τυπικής κανονικής κατανομής μπορούν να χρησιμοποιηθούν σε υπολογισμούς. Η τιμή z αντιπροσωπεύει το τυποποιημένο εξαγόμενο για το οποίο θέλουμε την πιθανότητα (1100 mm στο προηγούμενο παράδειγμα). Αυτή η τιμή υπολογίζεται ως εξής:

(A3.10) $$z = \frac{a - \bar{x}}{s_x}$$

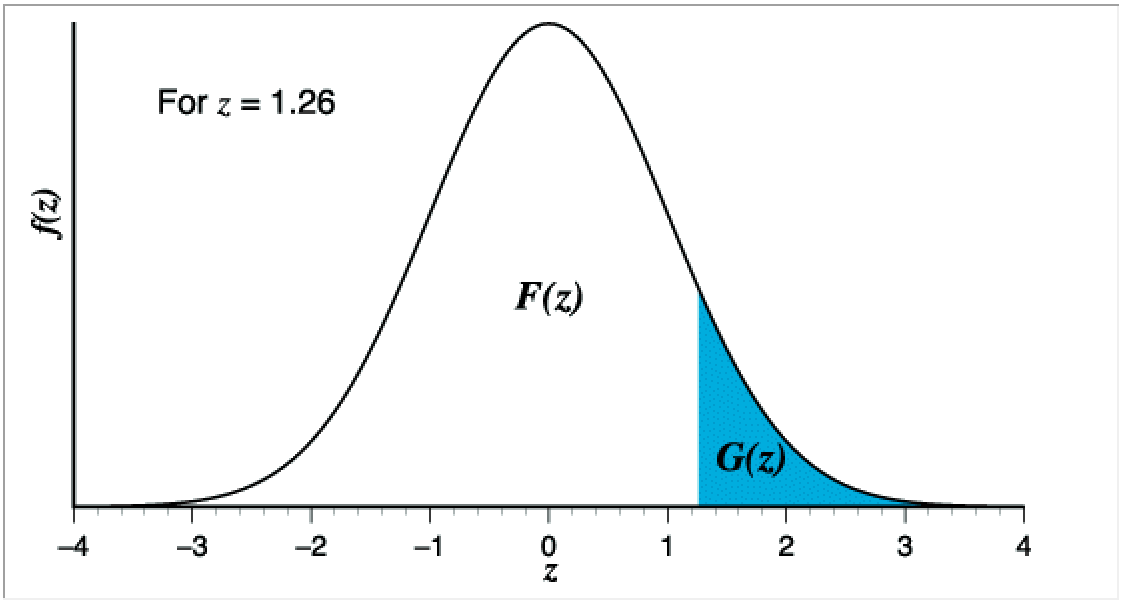


Σχήμα Α3.2 Ο μετασχηματισμός στην τυποποιημένη κανονική κατανομή (c) γίνεται με αλλαγή τόσο του μέσου (a) όσο και της τυπικής αποκλισης (b) της αρχικής κατανομής.

Για τα δεδομένα του παραδείγματος του πίνακα Α3.1 με a = 1100 mm, $\bar{x}$ = 979 mm, s_x = 101 mm, η υπολογισμένη τιμή z με χρήση της εξίσωσης (Α3.10) είναι 1.20 (σημειώστε ότι η τιμή z είναι αδιάστατη). Η συνάρτηση πιθανότητας για αυτή την τιμή z (Πίνακας Α3.2) είναι 0.8849 (δηλ. $F(z)$ = 0.8849), γεγονός που σημαίνει ότι η πιθανότητα σε κάθε χρονιά η βροχόπτωση να είναι μικρότερη ή ίση με 1100 mm είναι περίπου 0.9 (Σχήμα Α3.3). Για να βρούμε την πιθανότητα υπέρβασης αφαιρούμε την τιμή αυτή από το 1 [εξίσωση (Α3.2)], και βρίσκουμε την πιθανότητα η βροχόπτωση σε κάθε χρονιά να είναι ίση ή μεγαλύτερη από 1100 mm. Με άλλα λόγια υπάρχει 10% πιθανότητα η ετήσια βροχόπτωση να ξεπεράσει τα 1100 mm. Αυτό σημαίνει ότι, κατά μέσο όρο, 1100 χιλιοστά βροχόπτωσης ή περισσότερο συμβαίνει μια φορά στα 10 χρόνια. Η περίοδος επαναφοράς για ετήσια βροχόπτωση ίση ή μεγαλύτερη από 1100 mm είναι 10 χρόνια:

$$T_{return} = \frac{1}{\text{exceedance probability}} = \frac{1}{0.1\ \text{yr}^{-1}} = 10\ \text{yr} \qquad \textbf{(A3.11)}$$

Ολόκληρη η διαδικασία αυτή μπορεί να γίνει αντίστροφα για να προσδιορίσουμε την τιμή ή το εύρος των τιμών που συνδέονται με μια δεδομένη περίοδο επαναφοράς ή πιθανότητα υπέρβασης. Για παράδειγμα, χρησιμοποιώντας τα δεδομένα βροχόπτωσης του παραδείγματος, μπορούμε να υπολογίσουμε το ετήσιο ύψος βροχόπτωσης που περιμένουμε ότι θα ξεπερνιέται μια φορά στα τρία χρόνια. Τέτοιου είδους προβλήματα είναι συνηθισμένα στην διαχείριση υδατικών πόρων. Από τα δεδομένα ξέρουμε ότι η $G(z)$ είναι περίπου 0.33. Χρησιμοποιώντας τον πίνακα Α3.2 βρίσκουμε την τιμή για $F(z) = 1 - G(z) = 0.67$, που είναι 0.44. Αντικαθιστώντας την τιμή αυτή στην εξίσωση Α3.10 και αναδιατάσσοντάς την για να λύσουμε ως προς a (ξέρουμε τον μέσο και την τυπική απόκλιση του δείγματος) παίρνουμε την τιμή 1021 mm.



Σχήμα Α3.3 Η τυπική κανονική κατανομή στην οποία δείχνονται, για το πρόβλημα του παραδείγματος, τα τμήματα της κατανομής που δίνονται από την συνάρτηση κατανομής *F*(*z*), και την συμπληρωματική συνάρτηση κατανομής *G*(*z*).

Αυτή η σύντομη εισαγωγή ξύνει απλά την επιφάνεια των εφαρμογών της στατιστικής στα υδρολογικά προβλήματα, αλλά δίνει κάποια ιδέα του πως αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν για να προβλέψουν πιθανές συμπεριφορές. Κάποιες από τις έννοιες που εισάγαμε, όπως η πιθανότητα, η κατανομή πιθανότητας, η πιθανότητα υπέρβασης, και η περίοδος επαναφοράς χρησιμοποιούνται σε όλη την έκταση του βιβλίου.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Πίνακας Α3.2 Η συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής , $F(z)$, για $z > 0$. Για να χρησιμοποιήσετε τον Πίνακα, ψάξτε στην αριστερή στήλη για να βρείτε την τιμή του *z*-value στο πρώτο δεκαδικό, και μετά ψάξτε την πρώτη γραμμή για το δεύτερο δεκαδικό. Στην συμβολή της στήλης και της γραμμής παίρνετε την $F(z)$. Αν $z < 0$, χρησιμοποιήστε $F(z) = 1 - F(|z|)$.