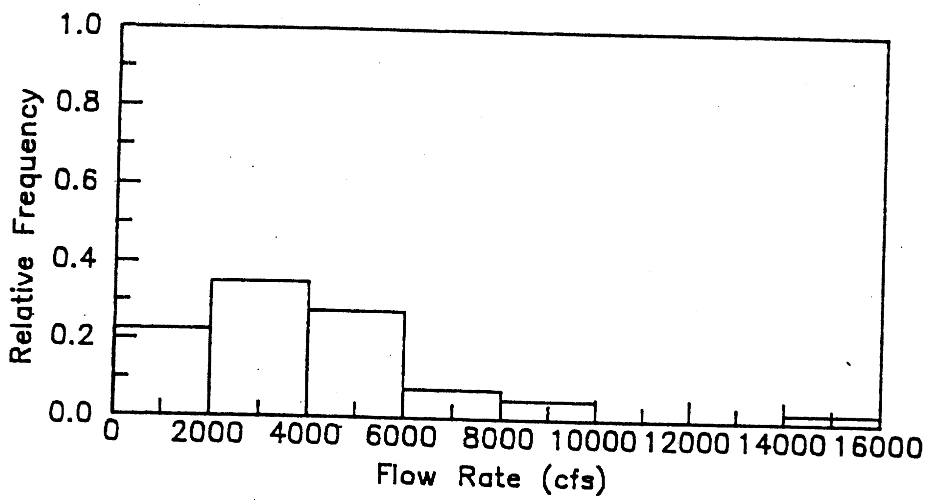
Solutions Chapter 3

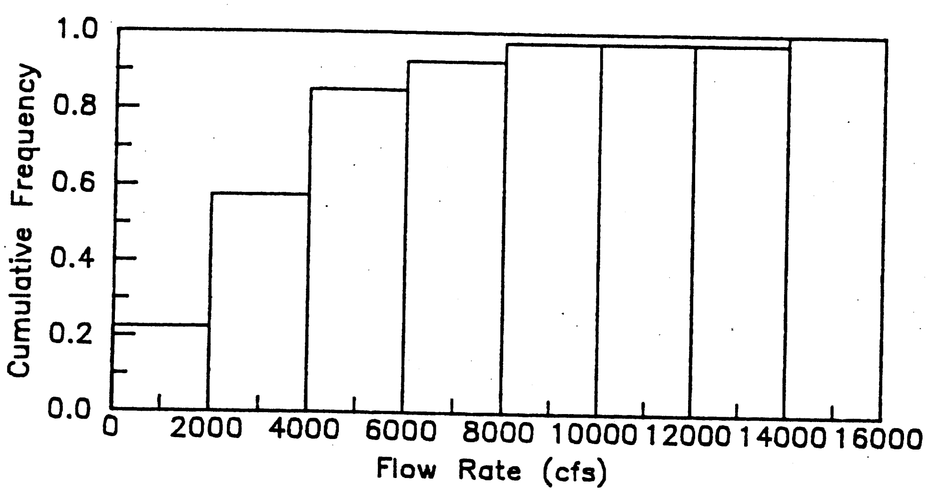
3.1. Data for Cypress Creek for the period 1945–1984 are listed in the following table. Using these data, develop a relative frequency histogram and a cumulative frequency histogram for Cypress Creek, 1945–1984. Use a class interval of 2000 cfs.

UNRANKED DATA		RANKED DATA	
Year	Flow (cfs)	Rank	Flow (cfs)
1945	9840	1	15,600
1946	5170	2	10,300
1947	1620	3	9840
1948	235	4	7760
1949	15,600	5	6560
1950	4740	6	6260
1951	427	7	5730
1952	3310	8	5440
1953	4400	9	5230
1954	7760	10	5170
1955	2520	11	5060
1956	340	12	4740
1957	5440	13	4710
1958	3000	14	4590
1959	3690	15	4400
1960	10,300	16	4300
1961	6260	17	4210
1962	1360	18	3980
1963	1000	19	3860
1964	2770	20	3830
1965	1400	21	3690
1966	3210	22	3460

1967	1110	23	3310
1968	5230	24	3210
1969	4300	25	3150
1970	2820	26	3080
1971	1900	27	3000
1972	3980	28	2820
1973	6560	29	2770
1974	4710	30	2730
1975	3460	31	2520
1976	3080	32	1900
1977	2730	33	1620
1978	3860	34	1400
1979	4210	35	1360
1980	3150	36	1110
1981	5730	37	1000
1982	3830	38	427
1983	5060	39	340
1984	4590	40	235

Both histograms are plotted below:





3.2. a) Use the data found in problem 3.1 to calculate the mean, standard deviation, and skew coefficient (Eqs. 3.37, 3.38, and 3.40) of the Cypress Creek data (1945–1984).

b) Repeat part (a) using the log (base 10) of the Cypress Creek data.

Using either a calculator or spreadsheet software, we get:

a) $\bar{x} = 4118$ cfs with n = 40

$S_x^2 = 8.6327 \times 10^6$

$S_x = 2938.1$

$C_{s2} = 2.098$

b) Assume $y_i = \log x_i$

$\bar{y} = 3.4922$

$S_y^2 = 0.14338$

$S_y = 0.3787$

$C_{s2} = -1.39176$

3.3. A temporary cofferdam is being designed to protect a 5-yr construction project from the 25-yr flood. What is the probability that the cofferdam will be overtopped

a) at least once during the 5-yr project,

b) not at all during the project,

c) in the first year only,

d) in the fourth year and fifth year exactly?

a) Risk $= 1 - (1\text{-}1/T)^n$

$n = 5$

Risk $= 1 - (1\text{-}1/25)^5$

$= 1 - 0.96^5$

Risk = 0.185

b) Reliability = 1 – Risk

$= 1 - (1 - (1\text{-} 1/T)^n)$

$= (1 - 1/T)^n$

$= (1 - 1/25)^5$

Reliability = 0.815

c) Prob $= (p)(1\text{-}p)^4$

$= (0.04)(0.96)^4$

Prob = 0.034

d) Prob $= (1\text{-}p)(1\text{-}p)(1\text{-}p)(p)(p)$

$= (1\text{-}p)^3\, p^2$

$= (0.96)^3\, (0.04)^3$

Prob = 0.00142

3.4. A recreational park is built near Buffalo Creek. The stream channel can carry 200 m^3/s, which is the peak flow of the 5-yr storm of the watershed. Find the following.

a) The probability that the park will flood next year

b) The probability that the park will flood at least once in the next 10 yr

c) The probability that the park will flood 3 times in the next 10 yr

d) The probability that the park will flood 10 times in the next 10 yr

a) Prob = 1/T

= 1/5

Prob = 0.20

b) n = 10

$$\text{Risk} = 1 - (1 - 1/T)^n$$

$$= 1 - (1 - 1/5)^{10}$$

$$= 1 - 0.8^{10}$$

Risk = 0.893

c) Use the binomial equation with n = 10, x = 3.

$$P(3) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \frac{10!}{3!7!}(0.2)^3 (0.8)^7$$

P(3) = 0.2013

d) x = 10, n = 10

e) $P(10) = \frac{10!}{10!0!}(0.2)^{10} (0.8)^0$

$P(10) = 1.024 \times 10^{-7}$

Problems 3.5 through 3.8 refer to the Cypress Creek data found in problem 3.1.

3.5. Assume that the Cypress Creek data for the period 1945–1984 are normally distributed. Find the following.

a) Peak flow of the 100-yr flood

b) Peak flow of the 50-yr flood

c) Probability that a flood will be less than or equal to 2000 cfs

d) Return period of the 2000-cfs flood

From Problem 3.2, we have the normal distribution parameters:

$\bar{Q} = \mu = 4118$ cfs

$S_a = 2938$

a) T = 50

F(z) = 1 – 1/T

= 1 – 1/50

F(z) = 0.980

From the table in Appendix D.1

z = 2.054

Then

$Q = \bar{Q} + zS$

= 4118 + (2.054)(2938)

$Q_{50} =$ 10,153 cfs

b) T = 25

F(z) = 1 – 1/T = 1 – 1/25 = 0.960

z = 1.751

$Q = \bar{Q} + zS$

= 4118 + (1.751)(2938)

3.5. (cont)

Q_{25} = 9262 cfs

c) $2000 = 4118 + z(2938) \Rightarrow z = -0.7209$

$F(-z) = 1 - F(z) = 1 - 0.7645 = 0.2355$

Prob $(Q \leq 2000) = 0.2355$

d) T = 1/Probability of exceedance = 1/0.7645

T = 1.31 year

3.6. Assume that the Cypress Creek data for 1945–1984 are lognormally distributed. Find the following.

a) Peak flow of the 100-yr flood

b) Peak flow of the 50-yr flood

c) Probability that a flood will be less than or equal to 2000 cfs

d) Return period of the 2000-cfs flood

The lognormal distribution is the same as the Log Pearson III with $C_s = 0$. Therefore, we may use Table 3.4 to get values for k. From Problem 3.2, we have:

$$\overline{y} = 3.4922$$

$$S_y = 3.3787$$

a) T = 100

K = 2.326 (From Table 3.4)

$$y = \overline{y} + KS_y$$

3.4922 + (2.326)(0.3787)

= 4.3731

$$Q_{100} = 23,608 cfs$$

b) T = 50

K = 2.054

y = 3.4922 + (2.054)(0.3787) = 4.2700

$$Q_{50} = 18,623 cfs$$

c) Q = 2000

y = 3.3010

3.3010 = 3.4922 + z (0.3787)

z = -0.5049

F(-z) = 1 – F(z) = 1 – 0.6932 = 0.3068

$$P(Q \leq 2000) = 0.3068$$

d) $T = 1 / P(Q \geq 2000)$

3.6. (cont)

$= 1/(1 - 0.3068)$

$T = 1.44$ year

3.7. Assume that the Cypress Creek data for 1945–1984 fit a log Pearson 3 distribution. Find the following.

a) Peak flow of the 100-yr flood

b) Peak flow of the 50-year flood

Use 1-F in GAMMAINV because of negative skew

average	StdDev	Skew =Cs	alpha	beta	xo	abs(beta)
3.4922	0.3787	-1.39176	2.06505342	-0.26353	4.036403	0.26353

T	F	1-F	x = GAMMAINV	logQ= xo - x	Q (cfs)
100	0.99	0.01	0.04304184	3.993361	9848
50	0.98	0.02	0.0615853	3.974818	9437

3.8. Assume that the Cypress Creek data for 1965–1984 fit a log Pearson 3 distribution (statistics of base 10 logs are $C_s = -1.15$, mean = 3.5375, Var = 0.03849, and standard deviation = 0.1962). Find the peak flow of the 100-yr flood and compare it with the value found in part (a) of problem 3.7. Explain the difference, knowing that 95% of residential development along Cypress Creek occurred after 1965.

Using the values given, we find

$$K = 1.484$$

$$y = \overline{y} + KS_y$$

$$= 3.5375 + (1.484)(0.1962)$$

$$y = 3.8286$$

$$Q_{100} = 6739 cfs$$

The 100 year peak flow is less for the time period 1965-1984 than for the period 1945-1984. Since most of the development occurred after 1965, we should have expected an increase in the peak flow. This decrease can be explained by the fact that there was a general decrease in the annual rainfall during the years 1965-1984. Another reason for the decrease is that the period 1945-1984 included two extremely high floods in the data. The removal of those two values affects the mean and standard deviation of the remaining data.

3.9. Assume that the Cypress Creek data for 1945–1984 fit a 3-parameter gamma distribution. Find the following.

a) Peak flow of the 100-yr flood

b) Peak flow of the 50-yr flood

Problem 3.9: Cypress Creek, Gamma-3, use statistics of flows
Use F in GAMMAINV

average	StdDev	Skew =Cs	alpha	beta	xo	abs(beta)
4118	2938	2.098	0.90875962	3081.962	1317	3081.962

T	F	x = GAMMAINV	Q = x + xo
100	0.99	13539.1951	14856 cfs
50	0.98	11442.0801	12759 cfs

3.10. Using results from problems 3.7 and 3.9 and additional computations as appropriate, estimate the return period and non-exceedance probability, $F(Q)$, of a flood of magnitude 2000 cfs for Cypress Creek, 1945–84. Perform this estimate graphically for the 3-parameter gamma distribution and for the log Pearson 3 distribution.

Distribution	average	SD	Skew	alpha	beta	xo	abs(beta)
LP3	3.4922	0.3787	-1.39176	2.065053	-0.26353	4.036403	0.26353

Distribution	average	SD	Skew	alpha	beta	xo	abs(beta)
Gamma3	4118	2938	2.098	0.90876	3081.962	1317	3081.962

Q (cfs)	logQ	x = xo-logQ	1-F = GAMMADIST	F	T (yr)
2000	3.30103	0.73537303	0.75290922	0.247	1.3

Q (cfs)	x = Q - xo	F = GAMMADIST	F	T (yr)
2000	683	0.23759337	0.238	1.3

3.11. Match the letters on the right with the numbers on the left to complete the mathematical statements about PDF properties. Assume that x is a normally distributed annual occurrence.

1. $\int_{\mu}^{\mu} f(x)dx = \square$

2. $\int_{\square}^{\infty} f(x)dx = 0.02$

3. $\int_{\square}^{\mu+\square} f(x)dx = 0.34$

4. $\int_{-\infty}^{\square} f(x)dx = 0.5$

5. $\int_{m_1}^{m_2} f(x)dx = \square$

A. Standard deviation

B. Median

C. 0

D. $P(m_1 \leq x \leq m_2)$

E. 50-yr magnitude

F. Variance

G. $F(x)$

1. C

2. E

3. A

4. B

5. D

3.12.

a) Compute frequency factors for $T = 2$ yrs and 100 yrs and for $C_s = +0.5$ and -0.5. Compare the four values with the values given in Table 3.4.

b) Repeat problems 3.7, 3.9, and 3.10 using Excel functions GAMMADIST and GAMMAINV.

Part a:

Frequency factor K (F, Cs) = Ginv*Cs/2-2/Cs

Ginv = GAMMAINV (probability, alpha, beta)

To evaluate K, set beta = 1 and alpha = 4/Cs^2

Probability = F for positive Cs and 1-F for negative Cs

Retain sign of Cs in equation for K

T	F	Cs	alpha	beta	Prob	Ginv	K Equation	K Table 3.4
2	0.5	0.5	16	1	0.5	15.6679	-0.083	-0.083
100	0.99	0.5	16	1	0.99	26.7429	2.686	2.686
2	0.5	-0.5	16	1	0.5	15.6679	0.083	0.083
100	0.99	-0.5	16	1	0.01	8.1811	1.955	1.955

Part B:

Here, must estimate parameters alpha and beta for each problem.

Then use function GAMMAINV to get magnitudes (peak flows) and function GAMMADIST to get probabilities.

See above for arguments of GAMMAINV

F = CDF = GAMMADIST (x, alpha, beta, true)

Parameter "true" means computation of the CDF

If Cs (and beta) are negative, use abs(beta) in function, and function returns 1-F

Then, with negative skew, compute X = xo-x, where x is value retained from GAMMAINV.

x = X – xo (positive skew)

x = xo – X (negative skew)

X = desired magnitude

xo = average- alpha*beta

3.12. (cont)

alpha = 4/Cs^2

beta = SD*Cs/2

Use statistics of logs for LP3

Use statistics of flows for Gamma3

3.13. The total annual runoff for a small watershed was determined to be approximately normal with a mean of 360 mm and a variance of 2900 mm^2. Determine the probability that the total runoff from the basin will exceed 250 mm in all four of the next consecutive 4 yr.

We have

$$x = \overline{x} + zS$$
$$S = \sqrt{\text{var}}$$
$$= \sqrt{29 cm^2}$$
$$S = 5.385 cm$$

Then

$$25 = 36 + z(5.385)$$
$$z = 2.043$$
$$F(-z) = 1 - F(+z)$$
$$= 1 - 0.9795$$
$$F(z) = 0.0205$$

Prob (4 exceedances):

$$= \binom{4}{4}(0.9795)^4(0.0205)^0$$
$$= \frac{4!}{4!0!}(0.9795)^4(0.0205)^0$$
$$= 0.9205$$

Problems 3.14 through 3.17 refer to the following data on Spring Creek.

YEAR	FLOW (cfs)	RANK	FLOW (cfs)
1940	3420	1	42,700
1941	42,700	2	31,100
1942	14,200	3	20,700
1943	8000	4	19,300
1944	5260	5	19,300
1945	31,100	6	14,200
1946	12,200	7	12,200
1947	10,000	8	12,100
1948	1430	9	10,700
1949	3850	10	10,300
1950	19,300	11	10,000
1951	0*	12	8760
1952	4130	13	8000
1953	8760	14	7560
1954	1400	15	7340
1955	3570	16	7340
1956	0*	17	6720
1957	4600	18	5260
1958	5260	19	5260
1959	6720	20	4660
1960	20,700	21	4600
1061	10,700	22	4130
1962	0*	23	3850
1963	1590	24	3570
1964	1770	25	3420

1965	2430	26	2430
1966	4660	27	1770
1967	1010	28	1590
1968	12,100	29	1430
1969	10,300	30	1400
1970	1400	31	1400
1971	1300	32	1300
1972	7560	33	1010
1973	19,300		
1974	7340		
1975	7340		

*The years where flow = 0 cfs should not be used in computations since they are outliers.

3.14.

a) Find the mean, standard deviation, and skew coefficient (Eqs. 3.37, 3.38, and 3.40) for the Spring Creek data for 1940–1975.

b) Repeat part (a) for the log (base 10) of the Spring Creek data for 1940–1975.

c) Use the Weibull formula (Eqs. 3.75 and 3.76) to determine the plotting positions of the Spring Creek data. Plot the data on normal probability paper. Graphically fit a normal distribution.

d) Repeat part (c) for the log (base 10) of the data or use lognormal paper. Graphically fit a lognormal distribution.

a) $\overline{Q} = 8952$

$S_Q = 9122$

$Q_{S_1} = 2.55$

b) $y_i = \log Q_i$

$\overline{y} = 3.7612$

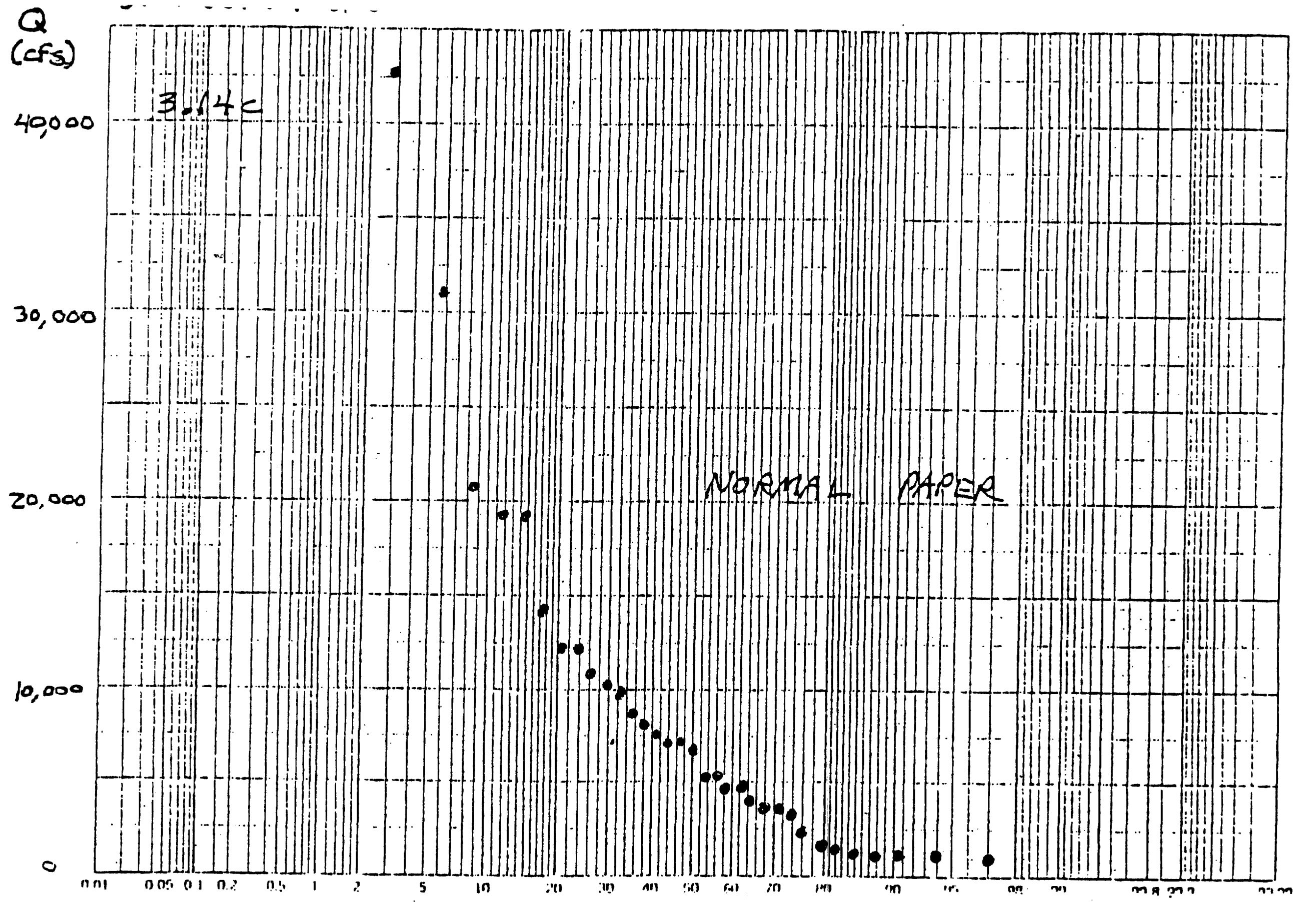
$S_y = 0.4257$

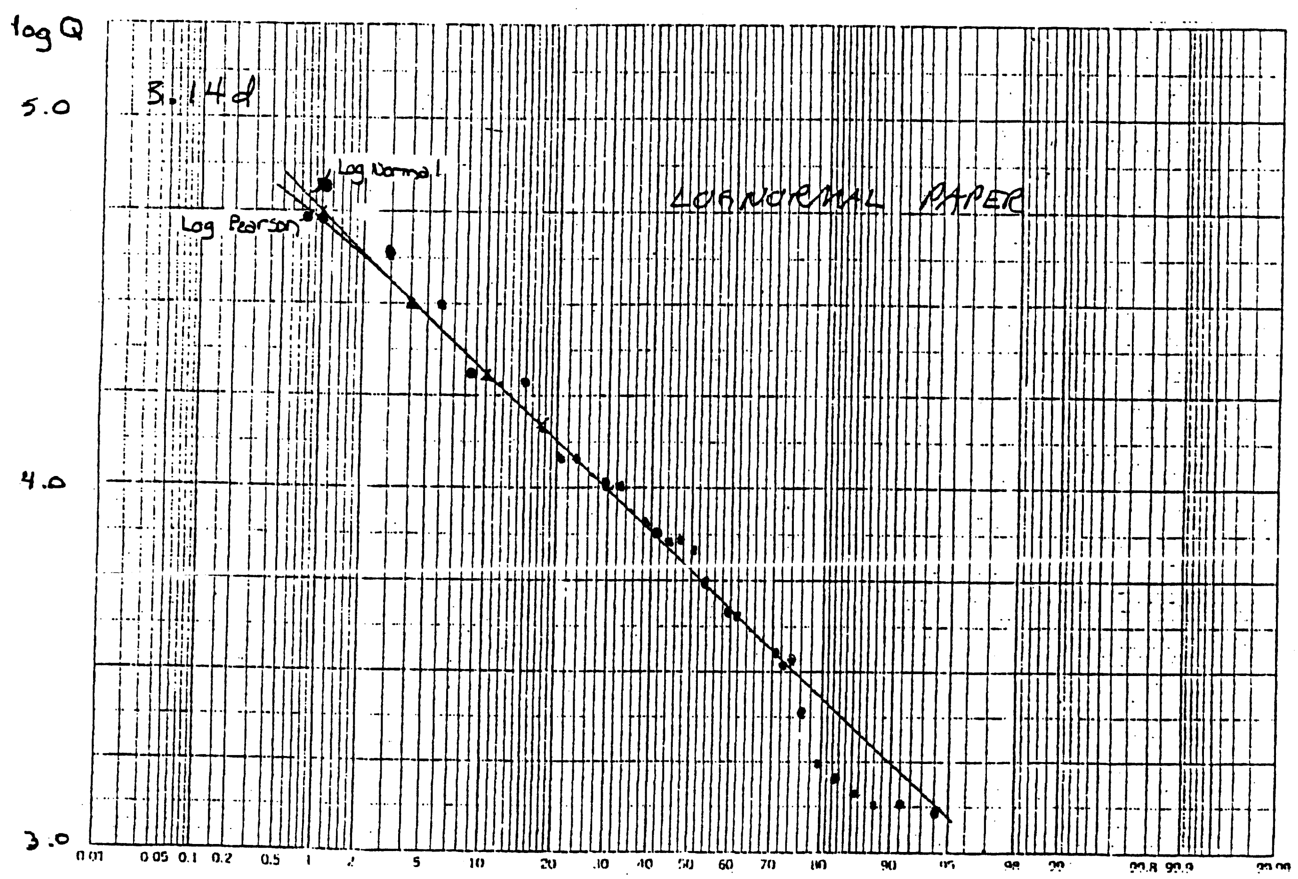
$C_{S_1} = -0.0459$

c) see plot

d) see plot next page

3.14. (cont)


Q
(cfs)
3.14c
40,000
30,000
20,000
10,000
0
NORMAL PAPER


log Q
3.14d
5.0
4.0
3.0
Log Normal
Log Pearson
LOGNORMAL PAPER

3.15. Assume that the Spring Creek data for 1940–1975 are lognormally distributed. What is the

a) peak flow of the 100-yr flood,

b) peak flow of the 25-yr flood,

c) peak flow of the 10-yr flood,

d) probability that a flood will be less than or equal to 6000 cfs,

e) return period of the 6000-cfs flood,

f) return period of the 15,000-cfs flood?

a) From the Log Pearson III table with C_S=0,

$K = 2.326$ (See 3.14b)

$y_{100} = 3.7612 + 2.326(0.4257) = 4.7514$

$Q_{100} = 56,413 cfs$

b) $K = 1.751$

$y = 3.7612 + 1.751\ (0.4257) = 4.5066$

$Q_{25} = 37,107 cfs$

c) $K = 1.282$

$y = 3.7612 + 1.282(0.4257) = 4.3073$

$Q_{10} = 20,274 cfs$

d) $Q = 6000 cfs$

$y = 3.7782$

$K = (y - \overline{y}) / s = (3.7782 - 3.7612) / 0.4257 = 0.0399$

Prob $(Q \geq 6000) = 0.49$ from plot of lognormal

3.15. (cont)

e) $T = 1/0.49$

$$\frac{1}{1-0.49}$$

$T = 2.04$ yr

f) From plot, $P(Q \geq 15000) = 0.17$

$T = 1/0.17$

$T = 5.9$ yr

$$\frac{1}{1-0.8377}$$

3.16. a) Assume that the Spring Creek data for 1940–1975 fit the log Pearson 3 distribution. Repeat problem 3.15.

b) Repeat part (a) for the 2-parameter gamma distribution.

c) Repeat part (a) for the 3-parameter gamma distribution.

a) $C_{S_1} = -0.039$ (from Problem 3.14)

(1) $K = 2.297$ by interpolation (Table 3.4)

$y = 3.7162 + 2.297(0.4257) = 4.7391$

$Q_{100} = 54,839 cfs$

(2) $K = 1.737$

$y = 4.5008$

$Q_{25} = 31,680 cfs$

(3) $K = 1.2773$

$y = 4.3050$

$Q_{10} = 20,182$ cfs

(4) $Q = 6000 \rightarrow y = 3.7782 \rightarrow K = 0.0399$

a linear interpolation gives

Prob $(Q \geq 6000) \approx 0.486$

Plotted data should be used for more accuracy.

(5) $T \approx 2$ yr

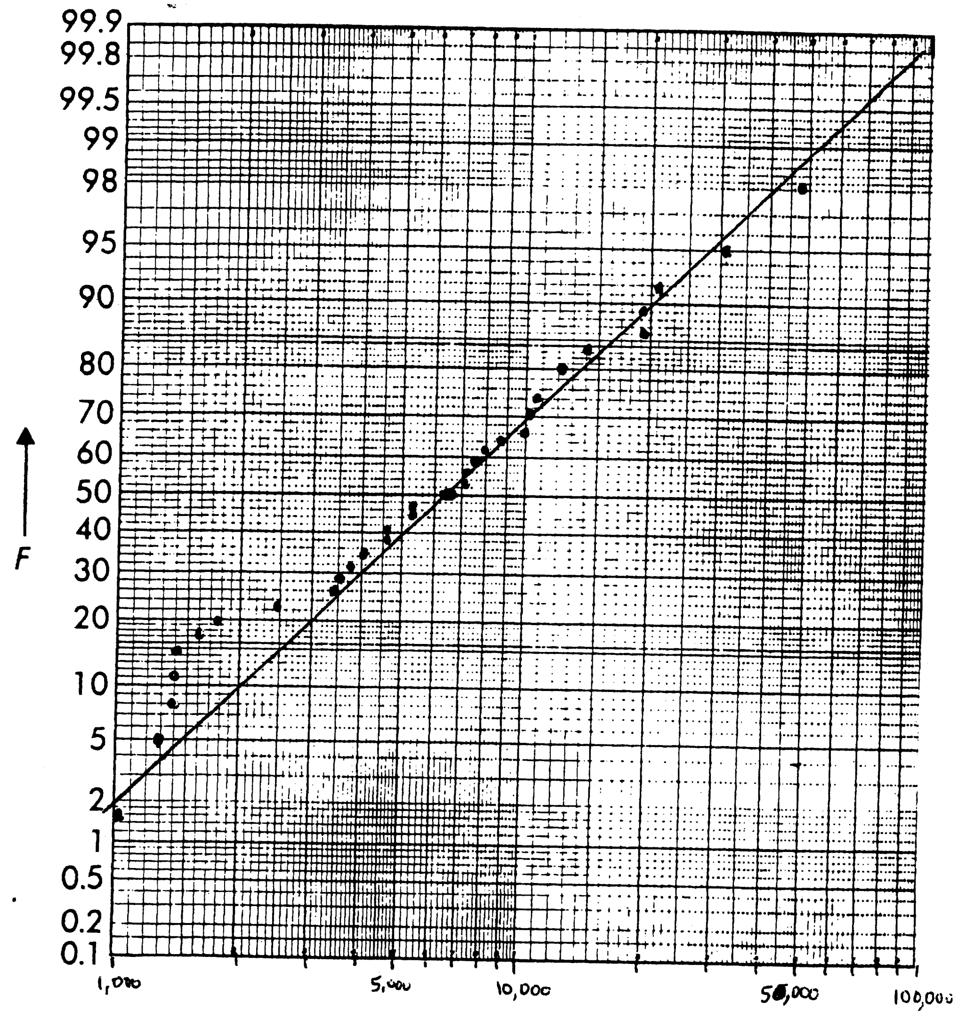
(6) Prob $(Q \geq 15000) \approx 0.1696$ - Plotted data would give more accurate probability than a linear interpolation.

$T \approx 5$ yr

b) Similar results will be attained using Eq. 3.68 for $C_s = 2.077$ and Table 3.4. For example, $Q_{100} = 42,155$ cfs.

c) Table 3.4 is used with $C_s = 2.55$ and similar results are obtained. For example, $Q_{100} = 44,162$ cfs.

3.17. Repeat problem 3.14(d) using the Cunnane plotting position (Eqs. 3.77 and 3.78) with parameter a = 0.4.



Note: In this solution, F vs. Flow is graphed. In Problem 3.14, I-F vs. Flow is graphed.

3.18. The following parameters were computed for a stream near Dallas, Texas, for 1940–1959, inclusive. The data were transformed to $\log_{10} Q = y$.

$$\bar{y} = 3.52 \text{ (mean)}$$
$$S_y = 0.50 \text{ (standard deviation)}$$
$$C_s = 0.50 \text{ (skewness coefficient)}$$

Find the magnitude of the 25-yr flood assuming that the annual peak flow follows (a) log Pearson 3 distribution and (b) lognormal distribution.

(a) $K = 1.91$

$y = 3.52 + 1.91\ (0.50) = 4.475$

$Q_{25} = 29{,}854$ cfs

(b) $K = 1.751$

$y = 3.52 + 1.751\ (0.50) = 4.3955$

$Q_{25} = 24{,}860$ cfs

3.19. A probability plot of 66 yr of peak discharges for the Kentucky River near Salvisa, Kentucky, is shown in Fig. P3.19.

a) What probability distribution is being used?

b) What are the mean and standard deviation of the peak discharges?

c) If the distribution has other parameters, what are their values?

d) What is the 25-yr flow?

e) What is the 100-yr flow?

f) What is the probability that the annual peak flow will be greater than or equal to 50,000 cfs for all of the next consecutive 3 yr?

g) What is the probability that at least one 100-yr event will occur in the next 33 yr? In the next 100 yr?

h) Which plotting position has been used to plot the data points?

i) Do the Kentucky River data appear to be skewed?

(a) Normal

(b) $\overline{Q} = 67,000 cfs$

$S_a = 23,000 cfs$

(c) None

(d) $Q_{25} = 102,000$ cfs

(e) $Q_{100} = 116,000$ cfs

(f) $P(Q \geq 50,000 \text{ in all of next 3 years}) = (0.8)^3 = 0.512$

(g) $R(33) = 1 - (0.99)^{33} = 0.282$

$R(100) = 1 - (0.99)^{100} = 0.634$

(h) Weibull

(i) No

3.20. A probability plot of 19 yr of peak discharges for the West Branch of the Mahoning River near Newton Falls, Ohio, is shown in Fig. P3.20. This is an example of Gumbel, or extreme value I, probability paper. Simply use the straight, fitted line to obtain the answers.

a) What is the 25-yr flow?

b) What is the 100-yr flow?

c) What is the return period of a flow of 4000 cfs?

d) What is the probability that the annual peak discharge will fall between 5000 and 7000 cfs?

a) $Q_{25} = 6400$ cfs

b) $Q_{100} = 8000$ cfs

c) 4000 => 0.8

0.8 => 20% flood = 5 year flood

d) $P(5000 < x \leq 7000) = 0.98 - 0.90 = 0.08$

3.21. The following annual total rainfall data for Houston Intercontinental Airport were collected over a 21-yr period.

YEAR	RAINFALL (in.)	YEAR	RAINFALL (in.)	YEAR	RAINFALL (in.)
1970	48.19	1977	34.94	1984	48.19
1971	37.83	1978	44.93	1985	49.14
1972	50.80	1979	58.97	1986	44.93
1973	70.16	1980	38.99	1987	40.60
1974	49.29	1981	55.98	1988	22.93
1975	50.97	1982	42.87	1989	52.73
1976	54.62	1983	53.21	1990	40.37

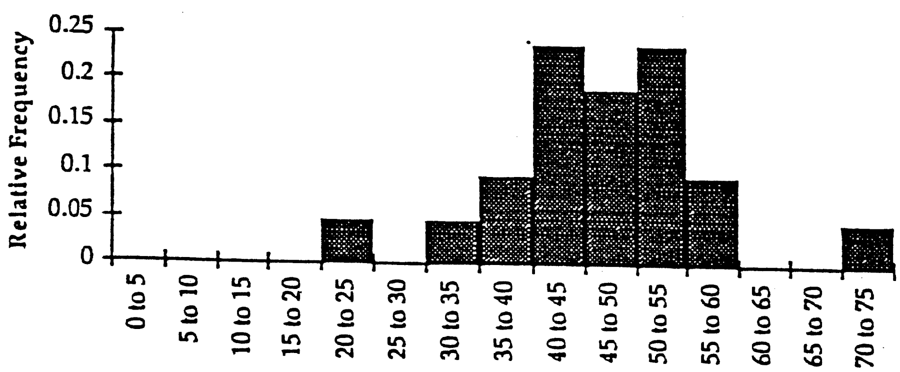
a) Compute the mean, variance, and the skewness coefficient (C_s).

$$\mu = \frac{\Sigma x_i}{n} = \frac{990.64}{21} = 47.17 \text{ inches}$$

$$Var(x) = \sigma^2 = \frac{\Sigma(x_i - \mu)^2}{n-1} = \frac{1915.6}{20} = 95.78$$

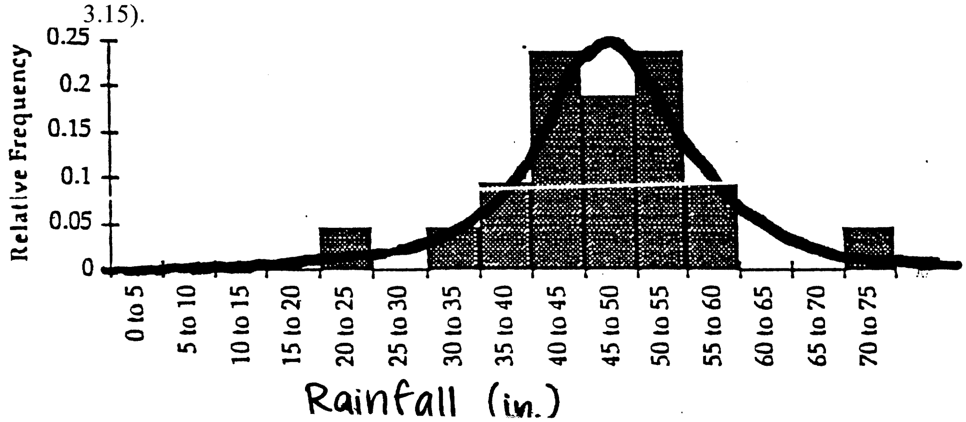
$$C_{S_i} = \frac{n}{(n-1)(n-2)} \times \frac{\Sigma(x_i - \mu)^3}{S_x^3} = \frac{21}{20 \cdot 19} \times \frac{-2747.76}{Var(x)^{3/2}} = -0.162$$

b) Plot a histogram using 5-in. intervals.



3.21. (cont)

c) Fit the data with the normal distribution. Sketch the normal PDF on the histogram of part (b), scaling such that the area under the histogram and under the PDF are the same (e.g., see Figure 3.15).



d) Find the value of the 10-yr annual rainfall total.

$F(x) = e^{-e^{-\alpha(x-u)}}$ For Gumbel

$$x = \frac{\ln(-\ln F(x))}{-\alpha} + u$$

$$F(x) = 1 - \frac{1}{T} = 1 - 0.1 = 0.9$$

$$\alpha = \frac{\pi}{\sigma\sqrt{6}} = 0.1311$$

$$u = \mu - \frac{0.5772}{\alpha} = 42.77$$

$$x = \frac{\ln(-\ln(0.9))}{-0.1311} + 42.77 = 59.94 \text{ inches}$$

e) Which years most closely represent the mean annual and 10-yr rainfalls for Houston?

1970 and 1984 most closely represent the mean annual rainfall.

1979 most closely represents the 10-year rainfall.

3.22. Explain how IDF curves (see Fig. 1.8) are statistically developed for any urban rainfall gage. Assume that data are available for 5-, 15-, 30-, and 60-min intervals up to 24 hr.

Solution: Perform a Gumbel Analysis for each of the data sets (ie. 5 minute intervals). For each interval, determine the 2, 5, 10, 25, 50, and 100 year rainfall intensities. Plot these points on log-log paper. Once this has been done for each interval, connect the points for each return period with a smooth curve.

3.23. Annual rainfall data for the Alvin gage are given below. The data should be fit using a log Pearson type 3 distribution. Decide if 1979 is an outlier by performing the analysis with and without the data point included.

YEAR	RAINFALL (in.)	YEAR	RAINFALL (in.)	YEAR	RAINFALL (in.)
1970	48.82	1977	34.53	1984	45.99
1971	38.27	1978	41.43	1985	59.12
1972	53.34	1979	102.58	1986	51.75
1973	71.93	1980	41.15	1987	67.70
1974	51.85	1981	52.79	1988	34.19
1975	43.73	1982	42.89	1989	48.02
1976	54.52	1983	60.48	1990	41.45

To determine if a value is an outlier, perform the following analysis, as presented by the Interagency Committee on Water Data (1982): Determine the high and low outlier thresholds of the distribution. If an outlier occurs, then discard it from the data set and repeat the analysis. These can be calculated from the following equations:

$$y_H = \mu + K_n \sigma$$
$$y_L = \mu - K_n \sigma$$

where K_n is the one-sided 10% significance level for the normal distribution, a function of n. (For $n = 21$, $K_n = 2.408$. For $n = 20$, $K_n = 2.385$.) The value y_H is the high outlier threshold (in log units for the lognormal or LP3 distributions), y_L is the low outlier threshold (in log units for the lognormal or LP3 distributions), μ is the mean (of the log-transformed data for the lognormal or LP3 distributions), and σ is the standard deviation.

Using log-converted data, perform a Log-Pearson analysis both with and without the 1979 data point.

WITH
$\mu_{21} = 1.69$
$\sigma_{21} = 0.11$
$Y_H = 1.92$
$Y_L = 1.43$

WITHOUT
$\mu_{20} = 1.59$
$\sigma_{20} = 0.11$
$Y_H = 1.85$
$Y_L = 1.33$

In both cases, the 1979 data ($Y = 2.01$) lies outside the upper outlier limit. This indicates that the 1979 data is an outlier, and should not be used in the analysis of the data.

3.24. The random variable x represents the depth of rainfall in June, July, and August in Houston. The whole PDF is *symmetric* and is shaped as an isosceles triangle, with base 0–60 in. Between values of $x = 0$ and $x = 30$, the probability density function has the equation

$$f(x) = \frac{x}{900}, \quad 0 \le x \le 30.$$

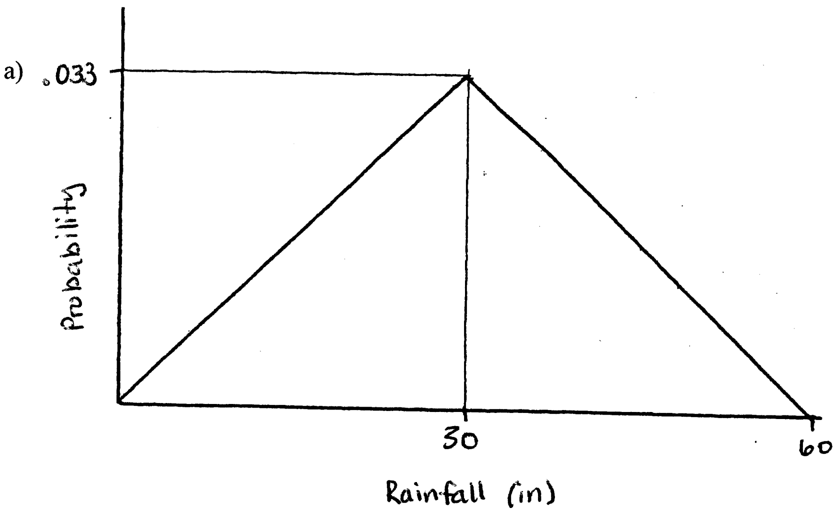
a) Sketch the complete PDF. Demonstrate that $\int f(x)\, dx = 1.0$.

b) Find the probability that next summer's rainfall will not exceed 20 in.

c) Find the probability that summer rainfall will equal or exceed 30 in. for the next three consecutive summers.

d) For the above PDF, what is the mean value of summer rainfall?

a)



$$f(x) = \frac{x}{900}, \quad 0 \le x \le 30$$

$$f(x) = .067 - \frac{x}{900}, \quad 30 \le x \le 60$$

$$\int_0^{60} f(x)dx = \text{Area inside triangle } = \frac{1}{2}(60)(.033) = 1$$

Alternately:

$$\int_0^{60} f(x)dx = \int_0^{30} \frac{x}{900}dx + \int_{30}^{60}\left(.067 - \frac{x}{900}\right)dx$$

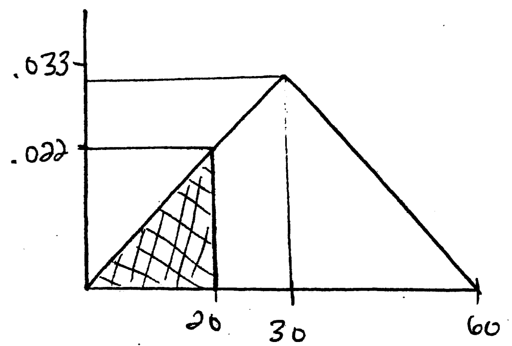
3.24. (cont)

$$= \left[\frac{x^2}{1800}\right]_0^{30} + \left[.067x - \frac{x^2}{1800}\right]_{30}^{60}$$

$$= .5 + (2.00 - 1.5) = 1$$

b) $P(x \le 20) = \int_0^{20} f(x)dx = \left[\frac{x^2}{1800}\right]_0^{20} = .222 = 22.2\%$

OR:



Area in triangle from 0 to 20 inches = ½bh

= ½ (.022)(20) = .222 = 22.2%

c) $P(x \le 30)$ for 3 summers

$$= \left[\int_{30}^{60} f(x)dx\right]^3 = \left(\left[.067x - \frac{x^2}{1800}\right]_{30}^{60}\right)^3 = .5^3 = .125 = 12.5\%$$

OR:

Area in triangle from:

$$\left[\left(\frac{1}{2}\right)(30)(.033)\right]^3 = .5^3 = .125 = 12.5\%$$

d) Mean = 30 inches

3.25. Repeat problem 3.21 developing a spreadsheet to solve the statistics.

INSERT TABLE

Year	Rain (inches)	Std Dev (X - μ)	Variance (X - μ)^2	Skewness (X - μ)^3	Freq F(X)	T (years)
1970	48.19	1.02	1.03	1.05	0.6118	2.6
1971	37.83	-9.34	87.30	-815.65	0.1479	1.2
1972	50.80	3.63	13.15	47.70	0.7054	3.4
1973	70.16	22.99	528.39	12145.85	0.9728	36.8
1974	49.29	2.12	4.48	9.48	0.6535	2.9
1975	50.97	3.80	14.41	54.73	0.7109	3.5
1976	54.62	7.45	55.45	412.94	0.8094	5.2
1977	34.94	-12.23	149.65	-1830.77	0.0613	1.1
1978	44.93	-2.24	5.03	-11.29	0.4708	1.9
1979	58.97	11.80	139.16	1641.64	0.8873	8.9
1980	38.99	-8.18	66.97	-548.01	0.1937	1.2
1981	55.98	8.81	77.56	683.02	0.8378	6.2
1982	42.87	-4.30	18.52	-79.69	0.3727	1.6
1983	53.21	6.04	36.44	219.98	0.7753	4.5
1984	48.19	1.02	1.03	1.05	0.6118	2.6
1985	49.14	1.97	3.87	7.61	0.6480	2.8
1986	44.93	-2.24	5.03	-11.29	0.4708	1.9
1987	40.60	-6.57	43.21	-284.03	0.2647	1.4
1988	22.93	-24.24	587.74	-14248.76	0.0000	1.0
1989	52.73	5.56	30.88	171.57	0.7626	4.2
1990	40.37	-6.80	46.29	-314.89	0.2542	1.3
Total	990.64	0.00	1915.60	-2747.76		

$\mu = 47.17 in.$

$\sigma = 9.79$

$\alpha = 0.1311$

$u = 42.77$

For equations and solutions for these values, refer to Problem 3.21.

3.26. This problem asks you to perform a descriptive analysis using real data of interest to you. The problem should be done using spreadsheet or similar software.

a) Download a series of annual maximum flows for a river of interest to you. USGS data may be obtained starting at the Web site *http://water.usgs.gov/nwis/.* Import the data into your spreadsheet or similar software. Convert the lines of text data into columnar data.

b) Note the characteristics of the basin from its description in the USGS files. What are the basin area and latitude and longitude of the gage? Are there diversions, controls, or storage (e.g., reservoirs) upstream?

c) Plot the time series of peak flows and $\log_{10}$(flows) vs. water year. The series of $\log_{10}$(flows) should have a lower coefficient of variation. Does the shape of the time-series plot of flows suggest that the time-series of river peak flows is nonstationary? If so, discuss possible reasons.

d) Compute and plot relative-frequency histograms for the flows and log(flows). Discuss any difference in skewness evident from the two plots.

e) Compute the following statistics for the series of flows and for the series of $\log_{10}$(flows): number, average, unbiased variance, unbiased standard deviation, coefficient of variation, unbiased skewness, maximum, and minimum.

3.27. For the data of problem 3.26, compute a weighted skew coefficient according to the Bulletin 17B method.

3.28. For the data of problem 3.26, fit a log Pearson 3 distribution to the peak flows, using the method of moments method described in this text. Use the weighted skew coefficient computed in problem 3.27.

a) Compute estimated flows for return periods listed in Table 3.4.

b) Plot the fitted CDF on lognormal probability paper.

c) Using the Cunnane plotting position formula (with parameter a = 0.4), plot enough of the measured flows to provide a comparison similar to Figure 3.20. Discuss the fit.

3.29. Repeat problem 3.28 using the lognormal and 3-parameter gamma distributions.

3.30. For a station of interest to you, download 10 years of daily average streamflow data from the USGS Web site *http://water.usgs.gov/nwis/.* Paste the data into a spreadsheet and convert the text data to columns. Construct and plot a flow-duration curve for these data. From the table and chart, what are the flows equaled or exceeded 20%, 50%, and 90% of the time?

The best example is the Siletz River example in the text.

3.31. Interevent times for winter storms arriving at Corvallis, Oregon for the months November through April for the winters of 1996–97, 1997–98, and 1998–99 were determined, and a frequency histogram prepared as shown in the table below. The average interevent time was 2.59 days.

a) Fit an exponential distribution to these data by finding the parameter λ.

b) Plot the relative frequency histogram and the fitted exponential PDF on the same chart. Care may need to be taken to be sure that the histogram and PDF are properly aligned. Each day (0–1, 1–2, etc.) is a class interval.

c) From the relative frequency histogram, compute the cumulative frequency histogram and plot on arithmetic graph paper.

d) On 2-cycle semilog paper (or using spreadsheet options for log-scales), plot the empirical CDF from part (c) and the fitted CDF. On this "probability paper," values should be plotted at the class mark, centering on half-days. The empirical values from part (c) should be plotted as individual points, and 1 – fitted CDF (exceedance probability) should be plotted as a straight line.

e) What is the probability that the time between winter storms is $\leq$ 3 days? Compute using both the empirical CDF and the fitted CDF.

Subtract 1/2 day from histogram values

						Part c:	Part d:		
Bin	Class Mark	Frequency	Relative Frequency	Subtr 1/2d Expon pdf	Expon pdf	Empirical Cum Freq	1- Emprical CDF	1 - fitted CDF (class mark)	Fitted CDF (bin)
0	-0.5		0.000	0.468	0.468	0	1.000	1.213	0.000
1	0.5	35	0.449	0.318	0.318	0.449	0.551	0.824	0.320
2	1.5	12	0.154	0.216	0.216	0.603	0.397	0.560	0.538
3	2.5	6	0.077	0.147	0.147	0.679	0.321	0.381	0.686
4	3.5	6	0.077	0.100	0.100	0.756	0.244	0.259	0.787
5	4.5	7	0.090	0.068	0.068	0.846	0.154	0.176	0.855
6	5.5	3	0.038	0.046	0.046	0.885	0.115	0.120	0.901
7	6.5	1	0.013	0.031	0.031	0.897	0.103	0.081	0.933
8	7.5	1	0.013	0.021	0.021	0.910	0.090	0.055	0.954
9	8.5	3	0.038	0.015	0.015	0.949	0.051	0.038	0.969
10	9.5	2	0.026	0.010	0.010	0.974	0.026	0.026	0.979
11	10.5	1	0.013	0.007	0.007	0.987	0.013	0.017	0.986
12	11.5	0	0.000	0.005	0.005	0.987	0.013	0.012	0.990
13	12.5	0	0.000	0.003	0.003	0.987	0.013	0.008	0.993
14	13.5	1	0.013	0.002	0.002	1.000	0.000	0.005	0.996
15	14.5	0	0.000	0.001	0.001	1.000	0	0.004	0.997
More	15.5	0							
		78	1.000						

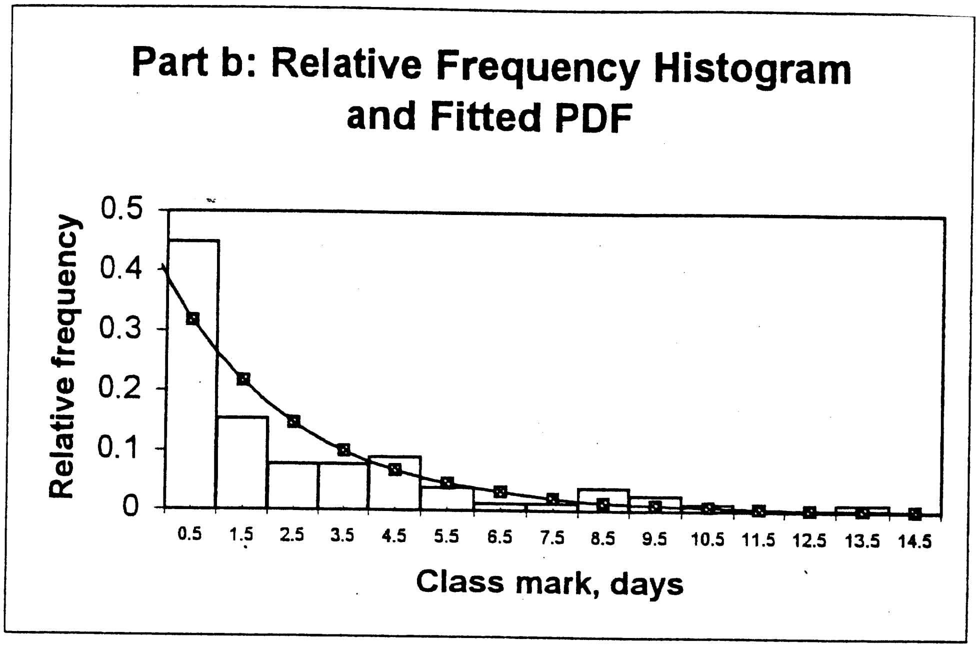
Part a:

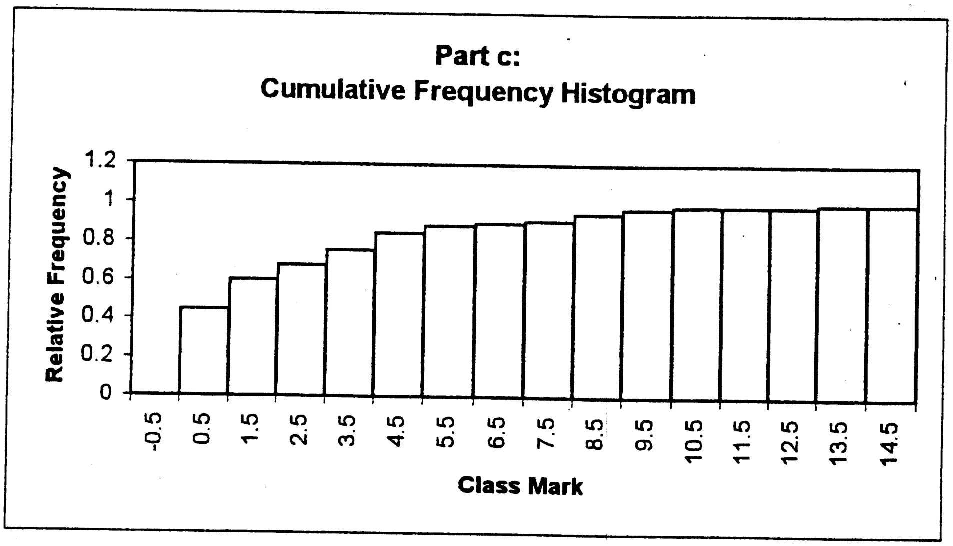
avg	2.59	days
lambda	0.3861	1/day (=1/avg)

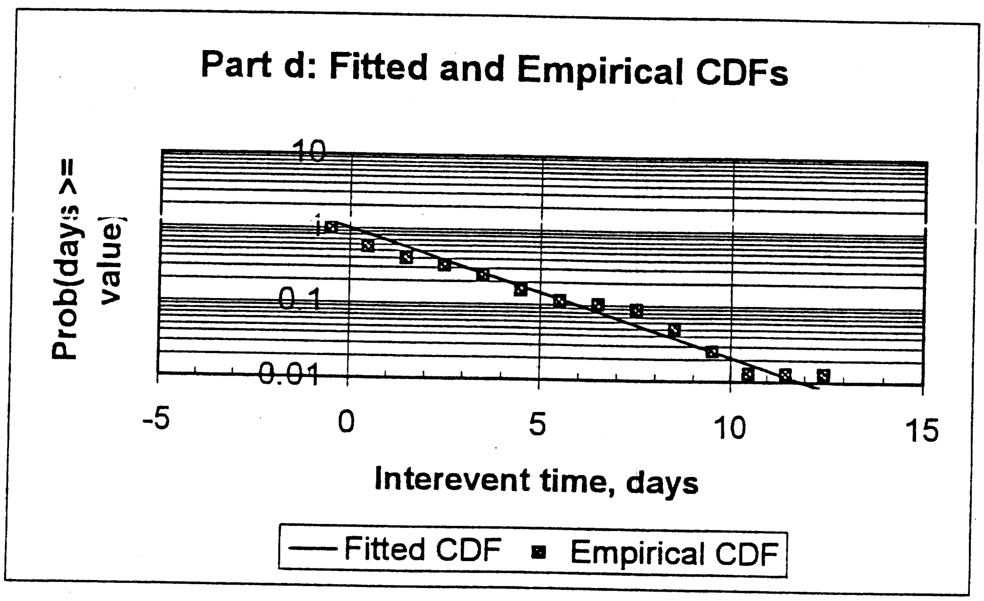
Part e:

Prob(t<=5)	0.846	empirical CDF
Prob(t<=5)	0.855	fitted CDF = 1 - exp(-5*lambda)

3.31. (cont)


Part b: Relative Frequency Histogram
and Fitted PDF
0.5
0.4
0.3
0.2
0.1
0
Relative frequency
0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5 10.5 11.5 12.5 13.5 14.5
Class mark, days


Part c:
Cumulative Frequency Histogram
1.2
1
0.8
0.6
0.4
0.2
0
Relative Frequency
-0.5 0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5 10.5 11.5 12.5 13.5 14.5
Class Mark


Part d: Fitted and Empirical CDFs
10
1
0.1
0.01
Prob(days >= value)
-5 0 5 10 15
Interevent time, days
Fitted CDF
Empirical CDF

3.32. The data presented in the table for problem 3.31 are known as **grouped data**, of the type that are developed in order to plot a frequency histogram. The mean of such data can be determined as a weighted average of the class marks, as follows,

$$\bar{t} = \frac{\sum_{i=1}^{k} f_i t_i}{\sum_{i=1}^{k} f_i}$$

where f_i and t_i are the frequency and class mark, respectively, for k class intervals.

a) Demonstrate that the mean of the interevent times of problem 3.31 is as stated.

b) How many interevent time values were used in the analysis?

Bin	ti = Class *Mark*	*Frequency*	*freq*ti*
0	-0.5		
1	0.5	35	17.5
2	1.5	12	18
3	*2.5*	6	15
4	3.5	6	21
5	*4.5*	7	31.5
6	5.5	3	16.5
7	*6.5*	1	6.5
8	7.5	1	7.5
9	*8.5*	3	25.5
10	9.5	2	19
11	*10.5*	1	10.5
12	11.5	0	0
13	*12.5*	0	0
14	13.5	1	13.5
15	*14.5*	0	0
More	15.5	0	0
	Sums:	78	202

Part a: average = 202/78 = 2.59

Part b: Sum freq = sum of sample values = 78

3.33. Using statistics for Siletz River peak flows, 1979–1999 (Table 3.1), generate a series of normally distributed synthetic streamflows by following these guidelines.

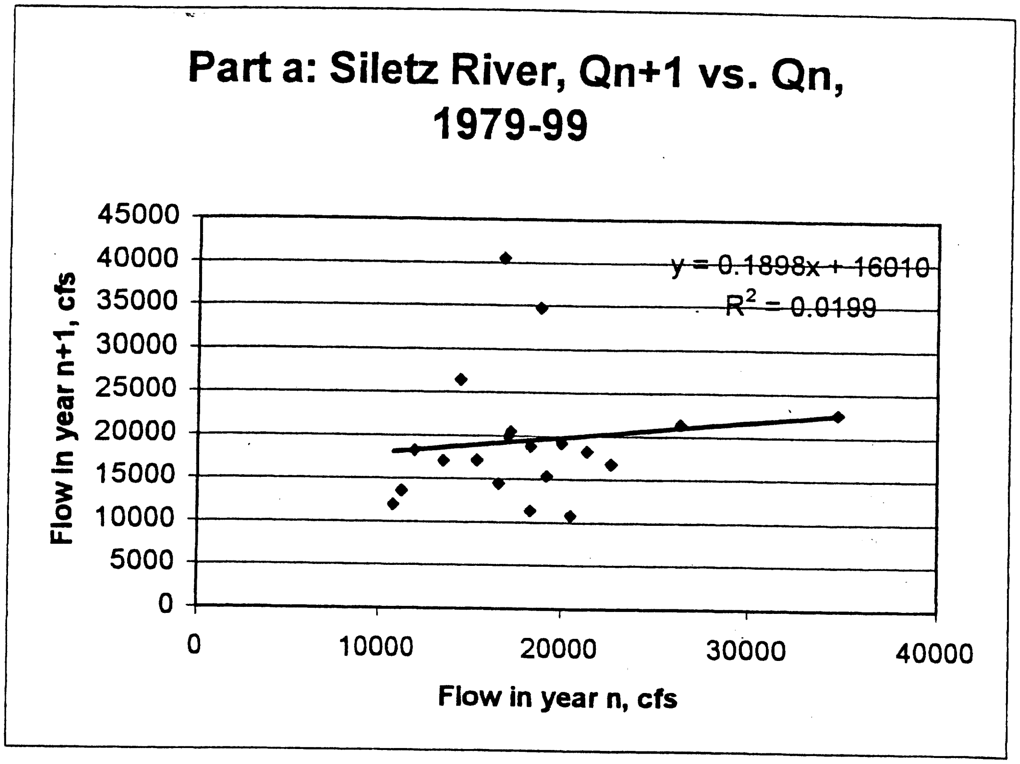
a) By performing a regression of flows for the period 1980–1999 vs. flows during 1979–1998, verify that the serial correlation coefficient for this time period is 0.1411.

b) Verify that the mean and unbiased standard deviation for the full 21-year period are 19,343 cfs and 7,376 cfs, respectively. Use these values for part (c).

c) The list of 21 $N(0,1)$ random numbers below was generated in Excel using the Tools/Data Analysis/Random number generation option with a seed of 12345. (The option for a seed allows one to generate identical sequences of random numbers.) Assuming that the initial flow = mean (at "step 0"), generate a sequence of 21 random flows using Eq. (3.82). Compute the mean, standard deviation, and serial correlation coefficient of the synthetic flow sequence to see how well these statistics are preserved. Optional: Create a new series of $N(0,1)$ random numbers and repeat the generation. Notice that as the mean and standard deviation of the random numbers differs from 0 and 1, respectively, so do the mean and standard deviation of the synthetic sequence differ from their historic values.

Year	Flow	Flow(n+1)
1979	16600	14500
1980	14500	26500
1981	26500	21400
1982	21400	18300
1983	18300	11300
1984	11300	13600
1985	13600	17100
1986	17100	20000
1987	20000	19200
1988	19200	15400
1989	15400	17200
1990	17200	20500
1991	20500	10800
1992	10800	12000
1993	12000	18300
1994	18300	18800
1995	18800	34700
1996	34700	22700
1997	22700	16800
1998	16800	40500
1999	40500	

3.33. (cont)



Part b: Use Excel functions AVERAGE and STDEV on 1979-99 data.

Qmean	19343	cfs
S	7376	cfs

	Column 1	*Column 2*
Column 1	1	
Column 2	0.141113	1

= output of Correlation from tools/Data Analysis