Chapter 3-Answer Key

3.1. This chapter employs Siletz River data for the continuous period 1925-1999. As of this writing, peak flow data through water year 2010 are available on the USGS Oregon surface water data webpage http://waterdata.usgs.gov/or/nwis/sw, including for Gage Number 14305500, which is the Siletz River. You will be asked in different problems in this chapter to update the Chapter 3 examples by using the 86-year record, 1925-2010.

a) Download the peak flow data for the Siletz River and enter the data into a spreadsheet.

b) Plot the 86 years of peak flows (1925-2010) and a 5-yr running mean vs. their water year, to update Fig. 3-2. Comment in general about the appearance of this time series, in the manner of the discussion of Fig. 3-2.

c) Using the updated data, develop new relative frequency and cumulative frequency histograms, that is, update Figs. 3-4 and 3-5.

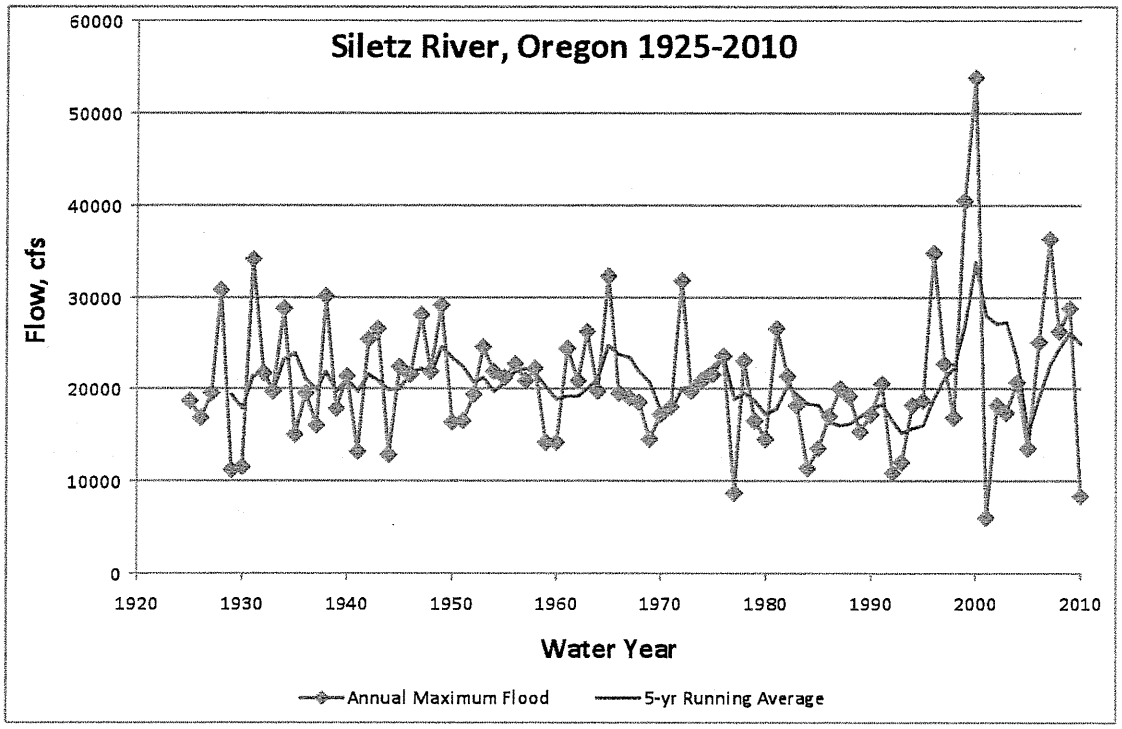
ANSWER:

a.

As reference, the first value should be for 1925 at 18800 cfs and the last value should be for year 2010 at 8410

3.1 cont'

b.

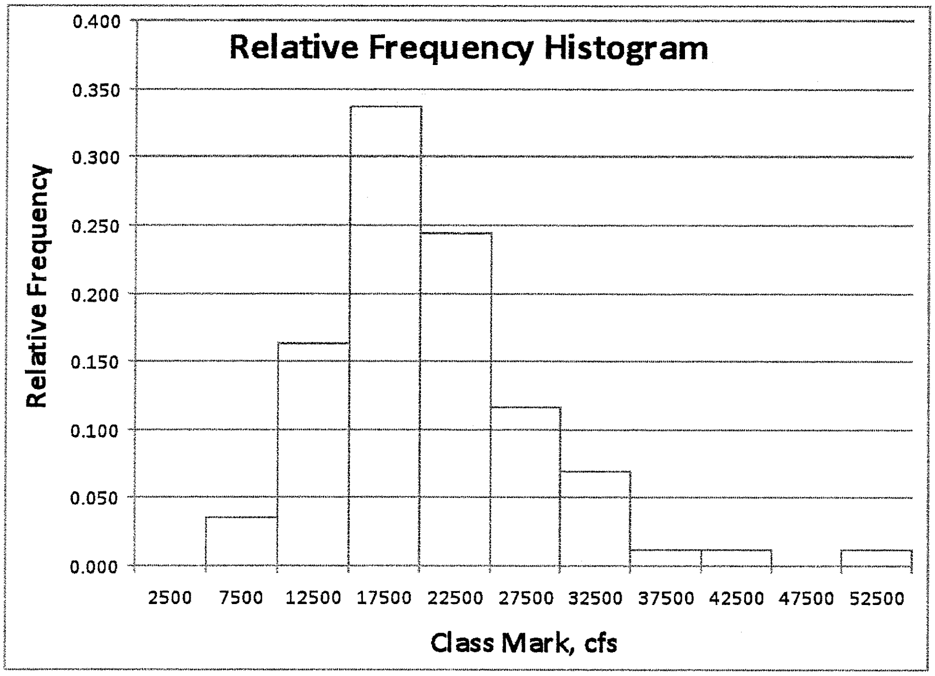


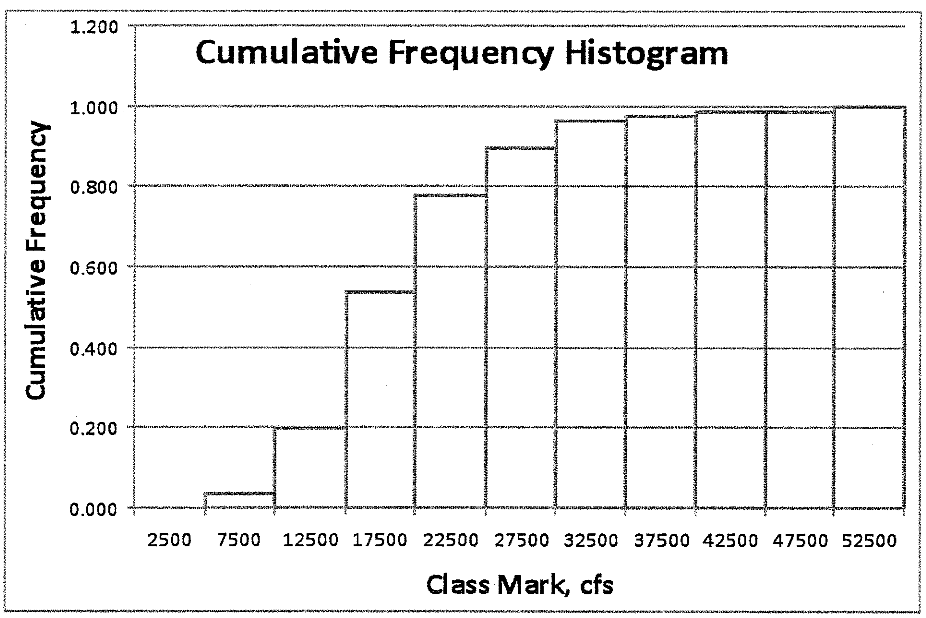
Discussion:

Variations in the decade 2000-2010 look different from those in earlier decades, but no reasonable explanation is apparent. Climate change possibilities, while newsworthy, cannot be said to be the cause without extensive study.

3.1 cont'

c.


Relative Frequency Histogram
Relative Frequency
0.400
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
2500 7500 12500 17500 22500 27500 32500 37500 42500 47500 52500
Class Mark, cfs


Cumulative Frequency Histogram
Cumulative Frequency
1.200
1.000
0.800
0.600
0.400
0.200
0.000
2500 7500 12500 17500 22500 27500 32500 37500 42500 47500 52500
Class Mark, cfs

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 updated Siletz River data (1925–2010).

b) Repeat part (a) using the logs (base 10) of the Siletz River data.

c) Develop the weighted skewness of the logs according to the Bulletin 17B protocol. That is, redo Example 3-3. At the conclusion of this problem you should have an updated version of Table 3-2.

ANSWER:

a)

Mean: **20796**

Standard Deviation: **7386**

Skew Coefficients: **1.34**

b)

Mean: **4.29**

Standard Deviation: **0.15**

Skew Coefficients: **-0.35**

c)

Weighted skewness according to the 17B protocol:

Use most recent data from 1925 to 2010

$C_w = WC_S + (1-W)C_m$

Where ,

C_S = Cstation = -0.3510 (part b)

C_m = Cmap = 0 for Siletz River on Oregon Central Coast

n = 86

3.2 cont'

$$W = \frac{V(C_m)}{V(C_S)+V(C_m)}$$

Where ,

$V(C_m) =$ Variance at Cmap = 0.302 from Bull 17B

$V(C_S) =$ Variance at Ċstation = $10^{A-B\log(n/10)}$

A and B are calculated in function of C_S, in this case it is <.90 so

A = -0.33+(0.08)(0.35) = -.302

B= 0.94-0.26(0.35) = 0.84

3.2 cont'

Plugging in A and B into equation

$V(C_S) = 0.080$

With known Variances W can be solved for,

W = 0.790 so,

$C_w = 0.790 \times -0.35 + (1-0.790)\times 0 =$ -.277

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 years exactly?

ANSWER:

T = 25

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

Reliability = 1 – risk

a)

For this scenario n = 5

$\text{Risk} = 1-(1-\frac{1}{25})^5 = 1-(.96)^5 = \mathbf{0.185}$

b.)

Use reliability

$=(1-\frac{1}{25})^5 = \mathbf{.815}$

c)

$\text{Probability} = p(1-p)^4$ where $p = \frac{1}{T}$

$P = 0.04\times(0.96)^4$

P = **0.034**

d)

$P = (1-p)\times(1-p)\times(1-p)\times p\times p$

$P = (1-p)^3\times p^2$

$P = (0.96)^3\times 0.04^2$

P = **0.000142**

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 three times in the next 10 yr

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

ANSWER:

a.

$P = \frac{1}{T} = \frac{1}{5} = \mathbf{0.20}$

b.

$\text{Risk} = 1-(1-\frac{1}{T})^n = 1-(1-\frac{1}{5})^{10}$

$= 1-(.80)^{10}$

$= \mathbf{0.893}$

c.

Use binomial equation where n = 10 x = 3

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

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

P(3) = **0.2013**

d.

n = 10, x = 10

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

P(10) = **1.204E-7**

Problems 3.5 through 3.8 refer to the updated 1925-2010 Siletz River data developed in

Problems 3.1 and 3.2. For consistency, assume the following moments are valid for the period water years 1925 – 2010:

	Original Data (cfs)	Log10 Data (log cfs)
Mean	20,796	4.29217
Standard Deviation	7,386	0.1527
Station Skewness	1.341	-0.3510
Weighted Skewness		-0.2773

3.5 Assume that the Siltez River data 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 30,000 cfs

(d) Return period of the 30,000-cfs flood

ANSWER:

a)

For a normal distribution and T = 100 and $F = 1-(1/T) = 0.99$

z = 2.326 (Table D2 or NORMSINV function in excel)

$Q = \overline{Q} + z \times Std$ = 20,796 + 2.326 * 7,386 = 37,975. 84 = **38,000 cfs**

b)

For T = 50, F = 0.98, z = 2.054

$Q = \overline{Q} + z \times Std$ = 20,796 + 2.054 * 7,386 = 35,966.84 = **36,000 cfs**

3.5 cont'

c)

Q = 30,000 cfs

$$z = \frac{(Q-\overline{Q})}{Std} = \frac{(30,000-20,796)}{7,386} = \mathbf{1.246}$$

Interpolating in Table D.1 or using NORMSDIST function in Excel F = **0.894**

d)

$$T = 1-(1/F) = \mathbf{9.4\ years}$$

3.6 Assume that the Siletz River data 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 30,000 cfs

(d) Return period of the 30,000-cfs flood

ANSWER:

a.

For Lognormal distribution and T = 100, $F = 1-(1/T)$, z = 2.326 (Table D.2 or NORMSINV in excel)

$\log(Q) = \log(\overline{Q}) + (z \times Std_{\log(Q)})$ = 4.2922 + 2.326 * 0.1527 = 4.6474

Q = $10^{\log(Q)}$ = 44,401.74 = **44,400 cfs**

b.

For T = 50, F = 0.98, z = 2.054

$\log(Q) = \log(\overline{Q}) + (z \times Std_{\log(Q)})$ = 4.2922 + 2.054*0.1527 = 4.6058

Q = $10^{\log(Q)}$ = 40,345.96 = **40,300 cfs**

c.

Q = 30,000 cfs $\log(Q)$ = 4.77121

$z = \log(Q) - \dfrac{\log(\overline{Q})}{std_{\log(Q)}}$ = 1.211

Interpolating from table D.1 or using NORMSDIST in Excel F = **0.887**

d.

$T = 1-(1/F)$ = **8.9 years**

3.7 Assume that the Siletz River data may be fit by a 3-parameter Gamma (Pearson 3) distribution. Find the following.

(a) Peak flow of the 100-yr flood

(b) Peak flow of the 50-year flood

ANSWER:

a.

Find K from Table 3-4 for given skewness of flows

Linear interpolation for T = 100 gives

K = 3.235

Using the GAMMAINV function in excel gives a K = 3.326 use the interpolation instead

Q = $\overline{Q} + (k \times Std)$ = 20,796 + 3.235*7,386 = 44,689.71 = **44,700 cfs**

b.

Find K from Table 3-4 for given skewness of flows

Linear interpolation for T = 50 gives

K = 2.682

Using the GAMMAINV function in excel gives a K = 2.683 use the interpolation instead

Q = $\overline{Q} + (k \times std)$ = 20,796 + 2.682*7,386 = 40,605.25 = **40,600 cfs**

3.8 Assume that the Siletz River data may be fit by a log-Pearson 3 (LP3) distribution. Find the following.

(a) Peak flow of the 100-yr flood

(b) Peak flow of the 50-yr flood

ANSWER:

a.

Find K from Table 3.4 for given weighted skewness of log(flows)

Linear interpolation for T = 100 gives

K = 2.121 (Same for Excel Funtion GAMMAINV but use interpolation)

$\log(Q) = \log(\overline{Q}) + (z \times Std_{\log(Q)})$ = 4.2922 + 2.121 * 0.1527 = 4.6160

Q = $10^{\log(Q)}$ = 41304.75 = **41,300 cfs**

b.

Find K from Table 3.4 for given weighted skewness of log(flows)

Linear interpolation for T = 100 gives

K = 1.903

Excel Funtion GAMMAINV = 1.902 but use interpolation

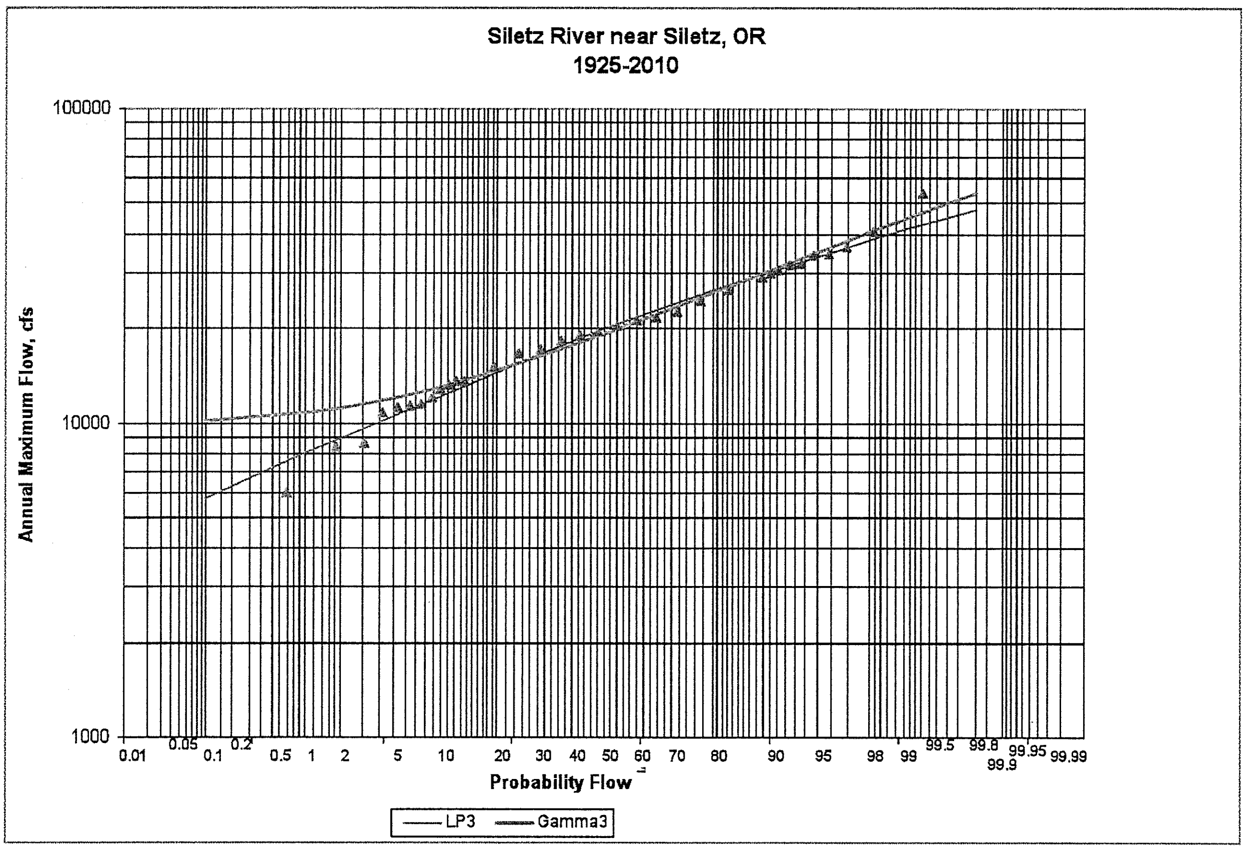
$\log(Q) = \log(\overline{Q}) + (z \times Std_{\log(Q)})$ = 4.2922 + 1.903 * 0.1527 = 4.5828

Q = $10^{\log(Q)}$ = 38,264.85 = **38,300 cfs**

3.9 Generate a new flood frequency plot for the updated Siletz River data, 1925-2010. That is, generate an updated version of Fig.3-20, but only plot the Gamma-3 and LP3 fits. You may omit some data points in the middle of the ordered series to ease crowding. If you are unable to obtain lognormal probability paper, a plot of magnitude vs. log(T) or magnitude vs. log(1-F) can serve as a substitute. That is, plot magnitude on the arithmetic scale vs. T or 1-F on the log scale on semi-log paper or in Excel.

ANSWER:

The plot is constructed exactly as shown in Example 3-16 in the text. Probability paper is obtained online (or "drawn" within excel).



3.10 Using graphs from Problem 3.9 and additional computations as appropriate, estimate the return period and nonexceedance probability, of a flood of magnitude 30,000 cfs for the Siletz River, 1925 - 2010.

ANSWER:

Both the LP3 and Gamma3 curves indicate a non-exceedance probability, F, of about **0.885.**

$$T = \frac{1}{(1-F)} = \mathbf{8.7\ yrs}$$

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. $\int_{\mu}^{\mu} f(x)dx =$ — a. Standard deviation

2. $\int_{[\,]}^{\mu} f(x)dx = 0.02$ — b. Median

3. $\int_{[\,]}^{\mu+[\,]} f(x)dx = 0.34$ — c. 0

4. $\int_{-\infty}^{[\,]} f(x)dx = 0.5$ — d. $P(m_1 \le x \le m_2)$

5. $\int_{m_1}^{m_2} f(x)dx = [\,]$ — e. 50-yr magnitude

f. Variance

g. F(x)

ANSWER:

1.C
2. E
3. A
4. B
5. D

3.12 The gamma distribution may be written in several different (but mathematically equivalent) forms. Excel uses the following form for the two-parameter gamma distribution in its functions GAMMADIST(x_1) $\equiv F(x_1)$ and GAMMAINV(F) $\equiv (x_1)$:

$$F(x_1,\alpha,\beta)=\int_0^{x_1}\frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}dx, \qquad x\geq 0.$$

The three-parameter gamma distribution with lower bound is easily evaluated by replacing the lower and upper limits by and respectively. Method-of-moments estimates for the three parameters are

$$\alpha=\frac{4}{C_S^2}$$

$$\beta=S_x\cdot\frac{C_S}{2}=\frac{S_x}{\sqrt{\alpha}}$$

$$x_0=\bar{x}-\alpha\beta=\bar{x}-\frac{2S_x}{C_S}$$

If $C_S<0$, β will be negative (not allowed in Excel). In this case, $|\beta|$ is used in the Excel functions, and GAMMADIST returns 1-F and the argument of GAMMAINV is 1-F Method-of-moments estimates for the two-parameter gamma distribution are:

$$\alpha=\bar{x}^2/S_x^2=1/CV^2$$

$$\beta=S_x^2/\bar{x}$$

The skewness of the two-parameter gamma distribution is

$$C_S = 2/\sqrt{\alpha} = 2CV$$

which provides a way of computing frequency factors, as described below.

Of course, moments are computed using logarithms of the data for the LP3 distribution. For given values of α and β, Excel function GAMMADIST returns $F(x_1)$ given x_1 and GAMMAINV returns x_1 given *F*. Letting $x_1 = \bar{x} + K \cdot S_x$ where $K(F, C_S)$ is the frequency factor, and using the parameter estimates given above, the upper limit of the integral can be manipulated to give

$$K(F, C_S) = G_{inv} \cdot C_S / 2 - 2 / C_S,$$

where $G_{inv} = GAMMAINV(F, \alpha, \beta)$ To evaluate *K*, set β = 1 and α $= 4 / C_S$ Hence, frequency factors can be computed only as a function of the CDF and skewness for any values of either parameter, and Table 3–4 can be avoided. For negative skewness, the symmetry of the distribution is exploited, and the same relationship holds for *K*, but with 1-F as the argument of GAMMAINV, instead of *F*. (The negative sign of C_S is retained in the equation.) For the special case of $C_S = 0$, K = z = standard normal variate.

(a) Compute frequency factors for T = 2yr and 100 yr and for C_S = +0.5 and C_S = -0.5 Compare the four values with the values given in Table 3–4.

(b) Repeat Problems 3.7, 3.8 and 3.10 using Excel functions GAMMADIST and GAMMAINV.

3.12 cont'

ANSWER:

a)

For C_s=-0.5

$\alpha = 4/0.5^2 = 16$

$\beta = 1$

$\Gamma = 0$

K (frequency factor) for 2 years = 0.083 (Table 3-4 = 0.083)

K (frequency factor) for 100 years = 2.686 (Table 3-4 = 2.686)

For C_s=+.05

α=4/0.5^2 = 16

$\beta = 1$

Γ=0

K (frequency factor) for 2 years = -0.083 (Table 3-4 = -0.83)

K (frequency factor) for 100 years = 1.955 (Table 3-4 = 1.955)

b)

3.7 repeat:

Gamma = lower boundary

Skewness = *2/*

Mean = *+ **

Standard Deviation ^{2} = ** 2*

For Siletz River Flows:

Mean = 20,796 cfs

SD = 7,386 cfs

Skew = 1.341

= 4 *2* = 2.226

= = 4,950.476 cfs

3.12 cont'

= − ∗ =9,776.624 cfs

For T = 100 **Q = 44,693 cfs** For T = 50 **Q= 40,608 cfs**

Q(normal) = 37,977 cfs Q(normal) = 35,964 cfs

3.8 repeat:

Same gamma, skewness, mean and SD equations

For Logs:

Mean = 4.292

SD = 0.153

Skew = -0.277

For Negative skewness

=4 2=52.035

= ∗ 2= −0.021

= − ∗ =5.394

Use in function

For negative skewness use 1-F

For T = 100 **Q = 41,307 cfs** For T= 50 **Q = 38,250 cfs**

Q(n) = 44,401 cfs Q(n) = 40,343 cfs

3.10 repeat:

GAMMA3

To Compute F(30,000)

Q = 30,000 cfs

Upper Bound = 30,000-gamma = 20,223.38

Lower Bound = 0

Use GAMMDIST

F = 0.889 **T = 9 years**

LP3

Q = 30,000 cfs

log =4.477

Upper Bound = 4.477-gamma = -0.917

3.12 cont'

Use Absolute Value of upper bound and return 1-F = 0.917

Lower Bound = 0

Use GAMMADIST with abs(beta)

1-F = 0.107552   F = 0.9

T = 9.3 years

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 years.

ANSWER:

$x = \bar{x} + zS$

$S = \sqrt{\text{var}}$

So $S = \sqrt{29cm^2}$ = 5.385

Then,

25 = 36 + z(5.385)

z = 2.043

Look in appendix D to get corresponding F(z) value about .9795

F(-z) = 1 - F(z)

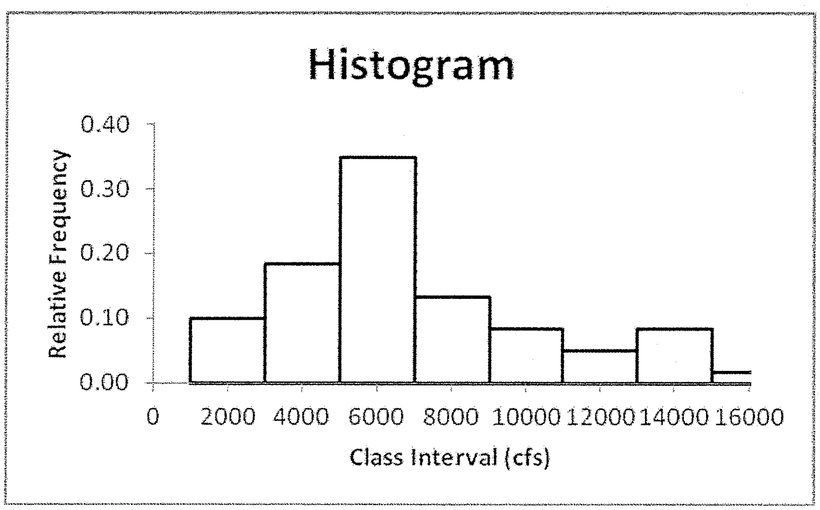
Probability 1-1/T = 1-.9795 = .0205

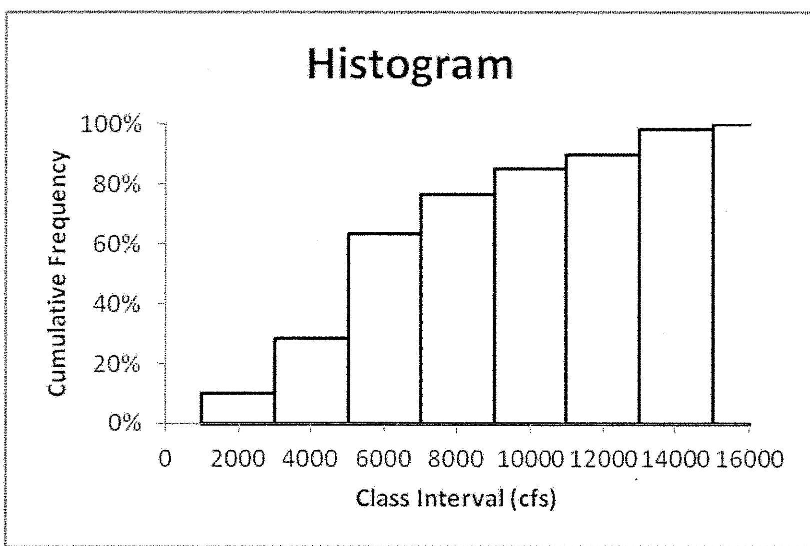
3.14 Data for Cypress Creek for the period 1951 – 2010 are listed on the USGS website. Following the link below, create a table of yearly peak streamflows, and then rank the data. Using these data, develop a relative frequency histogram and a cumulative frequency histogram for Cypress Creek, 1951 – 2010. Use a class interval of 2000 cfs.

*USGS National Water Information System for Gage # 0806900:*http://waterdata.usgs.gov/tx/nwis/dv

ANSWER:

Both Histograms are plotted below.





3.15 a) Use the data found in problem 3.14 to calculate the mean, standard deviation, and skew coefficient (Eqs. 3.37, 2.38, and 3.40) of the Cypress Creek data (1951 – 2010).

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

ANSWER:

Using either a calculator or spreadsheet software, we get:

a) = 5925
$S_x^2 = 1.1140 \times 10^7$
$S_x = 3377$
$C_s = 0.789$

b) = 3.69
$S_x^2 = 0.087$
$S_x = 0.295$
$C_s = -0.950$

Problems 3.16 and 3.17 refer to the Cypress Creek data found in problem 3.14

3.16 Assume that the Cypress Creek data for the period 1951 – 2010 are normally distributed. Find the following.

a) Peak flow of the 25-yr flood

b) Peak flow of the 5-yr flood

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

d) Return period of the 2000 cfs flood

ANSWER:

From Problem 3.15, we have the normal distribution parameters:

μ = 5925 cfs

Sa = 3377

T = 25

F(z) = 1-1/T

= 1-1/25

F(z) = 0.96

From the table in Appendix D.1

z = 1.701

Then,

Q = μ + z*S

= 5925 + 1.701*3377

Q100 = 11,669 cfs

T = 5

3.16 cont'

$F(z) = 0.8$

$z = 0.892$

Then,

Q50 = 5925 + 0.892*3377

=8,937 cfs

2000 = 5925 + z*3377

$z = -1.162$

$F(-z) = 1-F(z) = 1-0.8870 = 0.113$

Prob (Q <= 2000) = 0.113

T = 1/Probability of exceedance = 1/0.8870

T = 1.13 years

3.17 Assume that the Cypress Creek data for 1981 – 2010 fit a log Pearson 3 distribution (statistics of base 10 logs are C_s = -0.788, mean = 3.8262, Var = 0.05104, and st. dev. = 0.2259). Find the peak flow of the 100-yr flood and compare it with the peak flow using the entire data set (14,757 cfs peak 100-yr flood using entire data set). Explain the difference, knowing that 95% of residential development along Cypress Creek occurred after 1965.

ANSWER:

Using the values given, we find:

$K = 1.7418$

$y = \mu + K*S$

$= 3.8262 + 1.7418*0.2259$

$y = 4.2197$

$Q_{100} = 16{,}584$ cfs

The 100 yr peak flow is greater for the time period 1981 – 2010 than for the period 1951 – 2010. An increase in peak flow should have been expected as Cypress Creek watershed has been rapidly developed during the last part of the century. This increase can also be explained by a general decrease of rainfall during the 1950's through the 1980's, and several high intensity rain storms during the 2000's.

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

$$\bar{y}=3.52 \text{ (mean)}$$

$$S_y= 0.50 \text{ (standard deviation)}$$

$$C_s = \mathbf{0.50} \text{ (Weighted skewness)}$$

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

ANSWER:

(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 three years?

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

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

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

ANSWER:

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 in all of next 3 years) = $(0.8)^3 = 0.512$

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

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

3.19 cont'

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?

ANSWER:

a) $Q_{25} =$ 6400 cfs

b) Q_{100} =8000 cfs

c) 4000 =>0.8

0.8=>20% flood = 5 year flood

d) P ($5000 < x \leq 700$) = 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

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

(c) Fit the data with the normal distribution. Sketch the normal PDF on the histogram of part (b), scaling such that the areas 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.

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

ANSWER:

a)

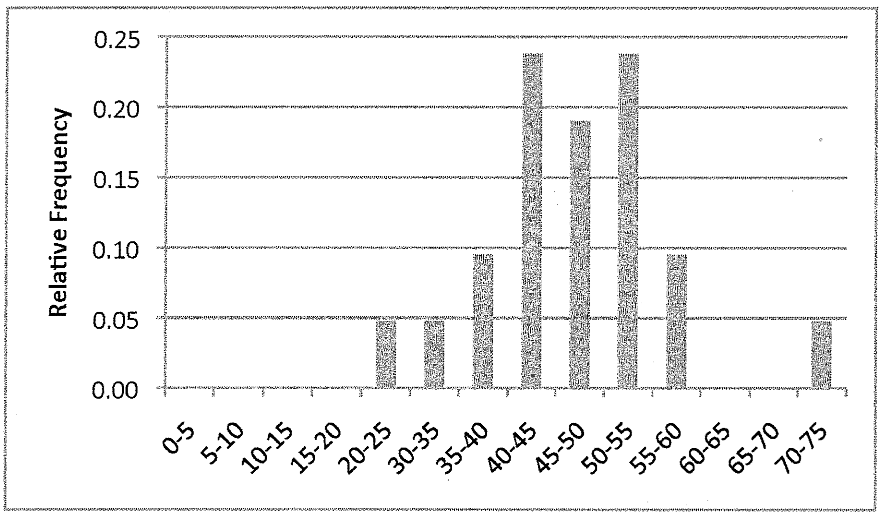
$$\mu = \frac{\sum x_i}{n} = \frac{990.64}{21} = 47.17 inches$$

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

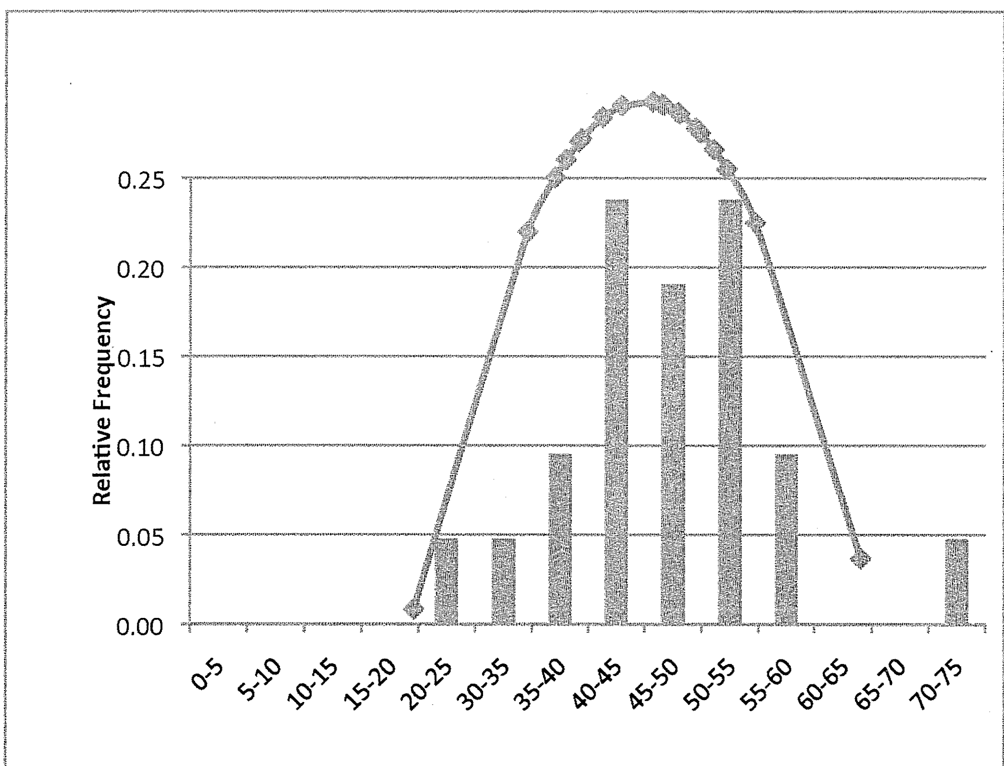
3.21 cont'

$$C_{s_1} = \frac{n}{(n-1)(n-2)} \times \frac{\sum (x_1 - \mu)^3}{S_x^3} = \frac{21}{20.19} \times \frac{-2747.76}{Var(x)^{3/2}} = -0.162$$

b)



c)



d)

$$F(x) = e^{-e^{(x-u)}}$$

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

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

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

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

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

3.21 cont'

e)

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.

ANSWER:

Perform a Gumbel Analysis for each of the data sets (ie. 5 minutes 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, Texas gage are given below. The data should be fitted 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 values 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 with and without the 1979 data point.

3.23 cont'

ANSWER:

WITH	WITHOUT
$\mu_{21} = 1.69$	$\mu_{20} = 1.59$
$\sigma_{21} = 0.11$	$\sigma_{20} = 0.11$
$Y_H = 1.92$	$Y_H = 1.85$
$Y_L = 1.43$	$Y_L = 1.33$

In both cases, the 1979 data (Y = 2.01) lies outside the upper and lower 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 and the probability density function has the equation

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

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

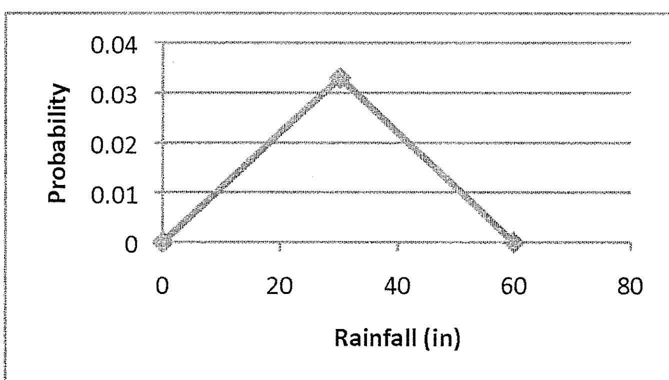
(b) Find the probability that next summer's rainfall will not exceed 20 in.

(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?

ANSWER:

a)



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

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

3.24 cont'

$$\int_0^{60} f(x)dx = \text{Area under the triangle} = (1/2)(60)(0.033) = 1$$

Alternatively,

$$\int_0^{60} f(x)dx = \int_0^{30} \frac{x}{900}dx + \int_{30}^{60}\left(0.067 - \frac{x}{900}\right)dx$$

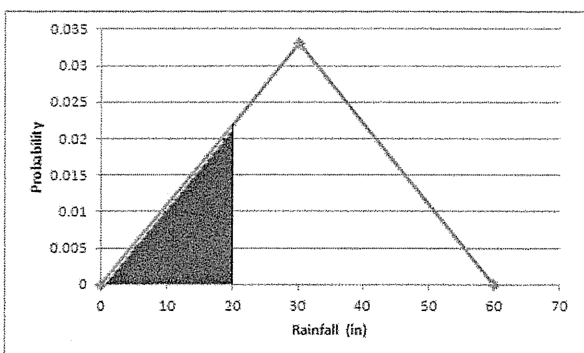
$$= \left[\frac{x^2}{1800}\right]_0^{30} + \left[0.067 - \frac{x^2}{1800}\right]_{30}^{60}$$

$$= 0.5 + (2 - 1.5) = 1$$

b)

$$P(x \le 20) = \int_0^{20} f(x)dx = \left[\frac{x^2}{1800}\right]_0^{20} = 0.222 = 22.2\%$$

OR



Area of triangle = (1/2)(0.022)(20)= 0.222 = 22.2%

c)

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

$$= \left[\int_{30}^{60} f(x)dx\right]^3 = \left(\left[0.067x - \frac{x^2}{1800}\right]_{30}^{60}\right)^3 = 0.5^3 = 0.125 = 12.5\%$$

3.24 cont'

OR

Area in the triangle = $\left[\frac{1}{2}(30)(0.033)\right] = 0.5^3 = .125 = 12.5\%$

d)

Mean = 30 inches

Numbers 3.25-3.29 do not have answers since they are open ended questions. To see solutions to similar problems please refer to the previous questions.

3.25 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 website 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 (flows) vs. water year. The series of (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 (flows): number, average, unbiased variance, unbiased standard deviation, coefficient of variation, unbiased skewness, maximum, and minimum.

3.26 For the data of Problem 3.25, compute a weighted skew coefficient according to the Bulletin 17B method.

3.27 For the data of Problem 3.25, fit a log Pearson 3 distribution to the peak flows, using the method of moments described in this text. Use the weighted skew coefficient computed in Problem 3.26.

(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), plot enough of the measured flows to provide a comparison similar to Figure 3–20. Discuss the fit.

3.28 Repeat Problem 3.27 using the lognormal and three-parameter gamma distributions.

3.29 For a station of interest to you, download 10 yr of daily average streamflow data from the USGS website 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?

3.30 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

(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, and so on) is a class interval.

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

(d) On two-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 one fitted CDF (exceedance probability) should be plotted as a straight line.

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

3.30 cont'

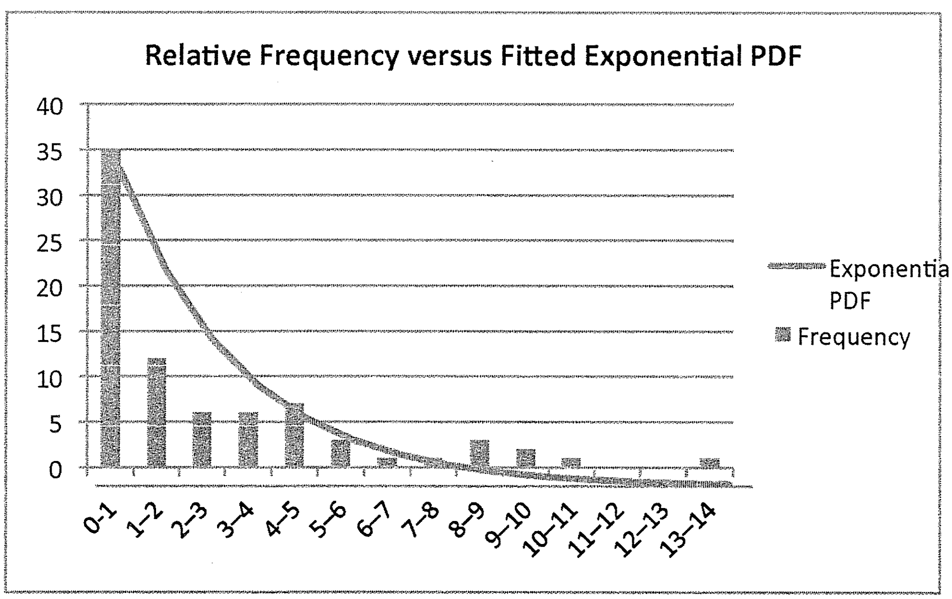
Interevent Time (days)	Frequency
0–1	35
1–2	12
2–3	6
3–4	6
4–5	7
5–6	3
6–7	1
7–8	1
8–9	3
9–10	2
10–11	1
11–12	0
12–13	0
13–14	1
	0

ANSWER

a)

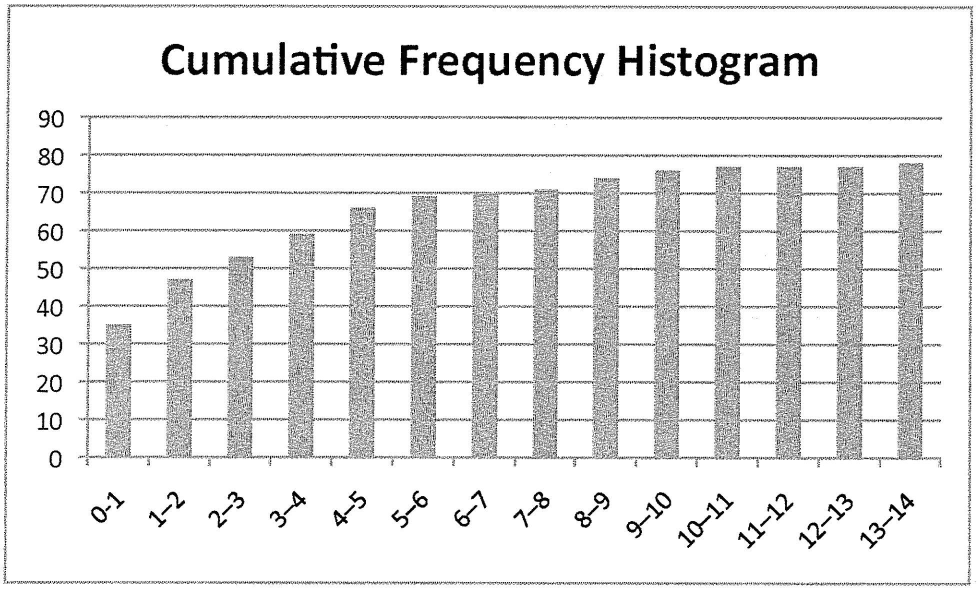
λ = 1/average = 1/2.59 = **0.3861**

b)

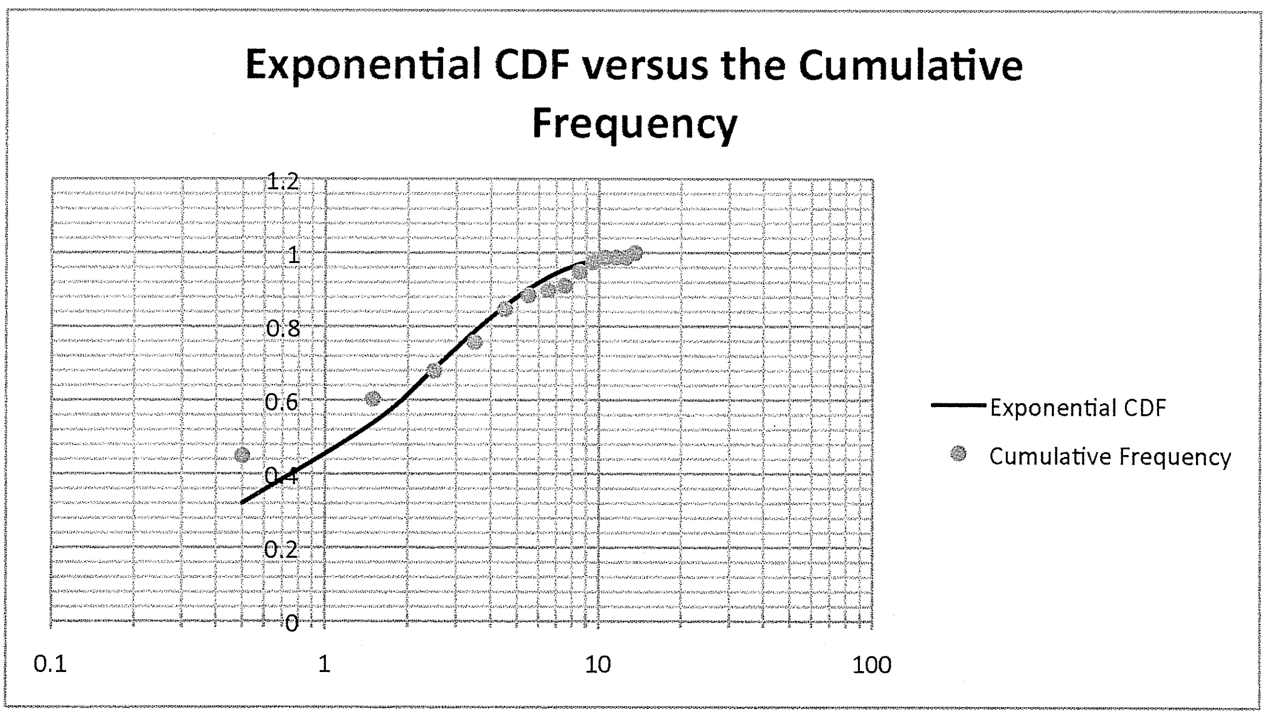


3.30 cont'

c)


Cumulative Frequency Histogram
90
80
70
60
50
40
30
20
10
0
0-1
1-2
2-3
3-4
4-5
5-6
6-7
7-8
8-9
9-10
10-11
11-12
12-13
13-14

d)


Exponential CDF versus the Cumulative Frequency
1.2
1
0.8
0.6
0.4
0.2
0
0.1
1
10
100
Exponential CDF
Cumulative Frequency

3.30 cont'

e) Using the fitted exponential distribution:

$\text{Prob}(t \leq 3) = F(3) = 1 - e^{(-0.3861*3)} = 1 - .3140 = .6860$

Using the empirical CDF:

$\text{Prob}(t \leq 3) = .7564$

3.31 The data presented in the table for Problem 3.30 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 and are the frequency and class mark, respectively, for k class intervals.

(a) Demonstrate that the mean of the interevent times of Problem 3.30 is as stated.

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

ANSWER:

	Bin	Mark	Frequency	Frequency* t_i
	0	-0.5	0	0
	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
	SUM		78	202

3.31 cont'

a) Average = 202/78 = 2.59

b) $\sum freq = \sum sampledvalues = 78$

3.32 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 7217 cfs, respectively. Use these values for part (c).

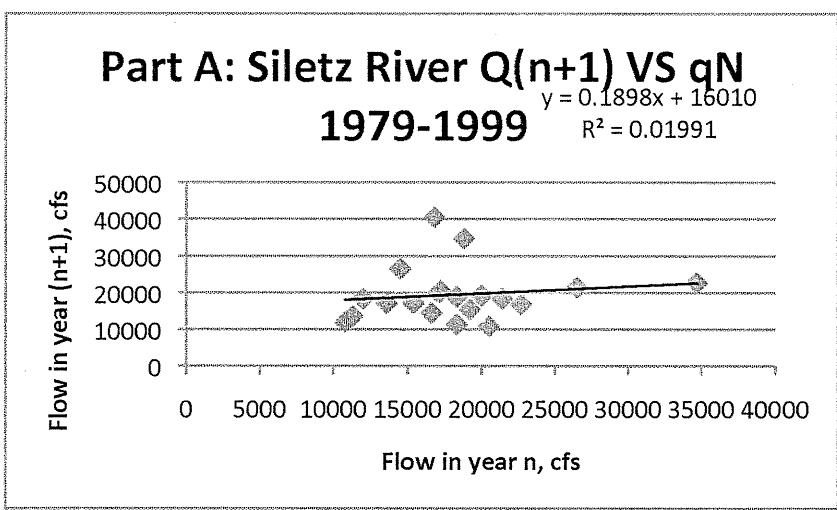
(c) The list of 21 $N(0,1)$ random numbers below was generated in Excel using the Tools/Data Analysis/Random Number Generation option with a seed of 12345. (The option for a seed allows one to generate identical sequences of random numbers.) Assuming that the initial flowmean (at "step 0"), generate a sequence of 21 random flows using Eq. (3–83). 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 differ from 0 and 1, respectively, so do the mean and standard deviation of the synthetic sequence differ from their historic values.

n	*z*	*n*	*z*	*n*	*z*
0	**n/a**				
1	-0.7341	8	-0.9733	15	-0.5506
2	0.2143	9	1.2119	16	-0.1774
3	0.7968	10	0.5659	17	0.4409
4	0.4544	11	-0.1092	18	-0.4908
5	-0.9235	12	-0.1214	19	0.8266
6	0.5659	13	0.3157	20	-1.1724
7	0.9885	14	0.4213	21	-0.406

3.32 cont'

ANSWER:

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	



3.32 cont'

b)

Using the Excel Functions AVERAGE, STDEV on 1979-1999 data

$\overline{Q} = 19{,}343$ cfs

S = 7,376 cfs

	Column 1	Column 2
Column 1	1	
Column 2	0.0141113	1

= output of Correlation from Tools/Data Analysis

c)

Synthetic Generation

$$Q_{n+1} = \overline{Q} + R \times (Q_n - \overline{Q}) + \sqrt{1 - R^2 \times S \times z}$$

Where R = 0.141113 = $\sqrt{R^2}$ on plot

$R^2 = 0.19913$

Ran no Seed = 12345

$Q_{start} = \overline{Q} = 19343$

n	z	Q		n	z	Q
0.0	n/a	19343.0		11.0	-0.109	19288.0
1.0	-0.734	13982.5		12.0	-0.121	18448.8
2.0	0.214	20151.4		13.0	0.316	21522.1
3.0	0.797	25275.5		14.0	0.421	22726.9
4.0	0.454	23498.3		15.0	-0.551	15799.9
5.0	-0.924	13185.8		16.0	-0.177	17547.6
6.0	0.566	22606.4		17.0	0.441	22309.2
7.0	0.989	27021.7		18.0	-0.491	16177.7
8.0	-0.973	13319.3		19.0	0.827	24932.3
9.0	1.212	27342.5		20.0	-1.172	11570.6
10.0	0.566	24604.1		21.0	-0.491	14663.8
			MEAN		0.050	19808.3
			STDEV		0.694	4888.4