CHAPTER I
INTRODUCTION
SIGNIFICANCE OF STATISTICS IN ENGINEERING

EXERCISES I

PROBLEM 1

Annual flow volumes at Keban station on Fırat river were measured from 1937 to 1967 thirty-one recorded values in 10^9 m^3 are given below.

YEAR	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947
FLOW	20,2	24,7	19,3	27,2	27,9	22,7	22,4	24,5	16,7	20,7	15,8
YEAR	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958
FLOW	25,1	15,5	16,8	15,8	22,9	21,6	24,3	13,1	19,7	18,8	15
YEAR	1959	1960	1961	1962	1963	1964	1965	1966	1967		
FLOW	12,5	19,9	10,1	15,1	30,8	20,8	18,5	26,6	27,6		

a) Plot the ***histogram*** of stream-flows, dividing the data into seven equal class intervals.
b) Plot the ***frequency histogram*** of stream-flows dividing the data into seven equal class intervals.
c) Estimate ***arithmetic mean*** $(\underline{y} = ?)$
d) Plot the ***cumulative frequency distributions***
e) Estimate ***variance*** $(Var(X) = ?)$
f) Estimate ***standard deviation*** $(s_X = ?)$
g) Estimate ***coefficient of variation*** $(C_{vX} = ?)$
h) Estimate ***coefficient of skewness*** $(C_{sX} = ?)$

PROBLEM 2

The data from a set of experiments where concrete beams were tested are given below. In this tests 30 crack loads (kg) were measured.

635	810	1045	890	520	800	710	760	860	990
660	730	790	570	810	740	940	860	840	595
930	840	790	740	810	685	780	610	850	1080

a) Plot the ***histogram*** of crack loads, dividing the data into six equal class intervals.
b) Plot the ***frequency histogram*** of crack loads, dividing the data into six equal class intervals.
c) Estimate ***arithmetic mean*** $(\underline{y} = ?)$
d) Plot the ***cumulative frequency distributions***
e) Estimate ***variance*** $(Var(X) = ?)$
f) Estimate ***standard deviation*** $(s_X = ?)$
g) Estimate ***coefficient of variation*** $(C_{vX} = ?)$
h) Estimate ***coefficient of skewness*** $(C_{sX} = ?)$

PROBLEM 3

The annual precipitation data (mm) recorded at Kumköy, Istanbul in period 1952-1990 are given below

a) Plot the ***histogram*** and ***frequency histogram*** dividing the data into six equal class intervals.
b) Investigate the effect of the number of class intervals using wider and narrower intervals.
c) Plot the ***cumulative frequency distributions***.
d) Estimate the parameters for the central value of the annual precipitation.
e) Estimate the ***parameters representing the dispersion*** of the data.
f) Compute the skewness coefficient. Can the data be considered to be symmetrical?

661	676	555	892	679	556	760	716	585	733
867	893	520	913	648	639	908	707	696	725
675	748	893	1139	815	685	907	932	919	1230
720	658	737	837	731	1046	926	471	763	

CHAPTER I
INTRODUCTION
SIGNIFICANCE OF STATISTICS IN ENGINEERING

EXERCISES I **SOLUTIONS** **05.02.2002**

These examples are expected to serve the significance of statistics in engineering problems to be better understood.

SOLUTION 1

a) As a first step we can classify the data into class intervals of equal length ($3x10^9$ m^3). Plotting the number of observations in each class interval as a horizontal line, we obtain a step diagram called the ***histogram***.

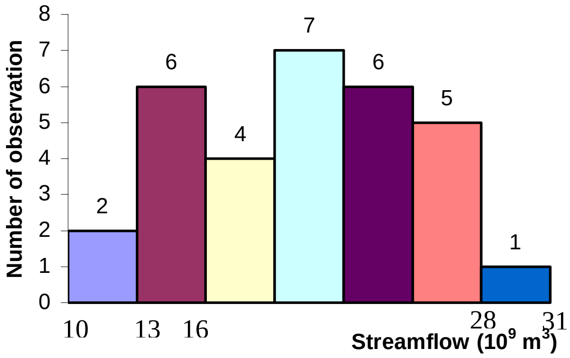

Figure 1: Histogram of streamflows

b)

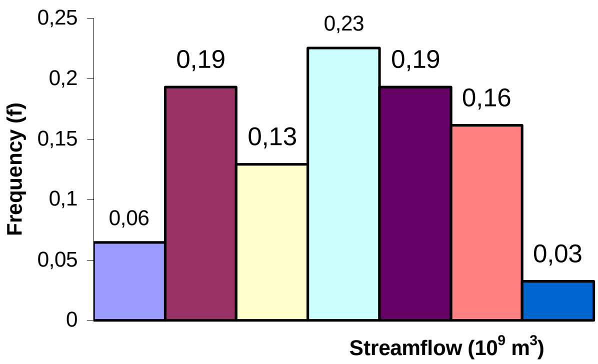

Figure 2: Frequency histogram of streamflows

c) Arithmetic mean computed as:

$$x_m=[\sum_{i=1}^{N} x_i]/N$$

$$=[(20,2+24,7+19,3+27,2+27,9+22,7+22,4+24,5+16,7+20,7+18,8+25,1+15,5+16,8+15,8+22,9+21,6+24,3+13,1+19,7+18,8+15,0+12,5+19,9+10,1+15,1+30,8+20,8+18,5+26,6+27,6]/31$$

$x_m = 20,4x10^9\ m^3$

Median is another parameter that characterizes the central value. It is one half of observations remain. Arranging the data in the increasing order, the observation in the center (for this problem 16th in rank) is estimate of the Median. In this problem two parameters related to the central value have almost equal values.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
10,1	12,5	13,1	15,0	15,1	15,5	15,8	15,8	16,7	16,8	18,5	18,8	19,3	19,7	19,9	20,2	20,7	20,8	21,6	22,4	22,7	22,9	24,3	24,5	24,7	25,1	26,6	27,2	27,6	27,9	30,8

$M_x = 20,2x10^9\ m^3$

d)

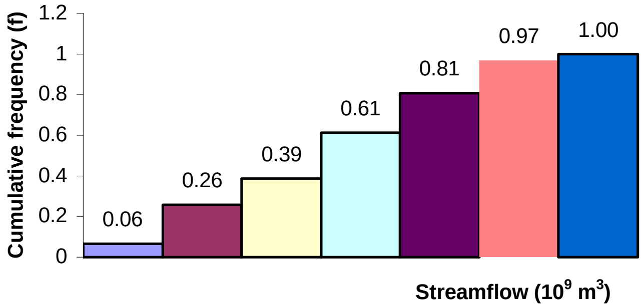

Figure 3: Cumulative frequency distributions of streamflows

e) The mean of the squares of the differences, called the ***Variance*** is a measure of the scatter of the data around its mean. Computed as:

$$Var(X) = [\sum_{i=1}^{N}(x_i - x_m)^2]/N$$

Var(X) = **[**$(20,2-20,4)^2+(24,7-20,4)^2+(19,3-20,4)^2+(27,2-20,4)^2+(27,9-20,4)^2+(22,7-20,4)^2+(22,4-20,4)^2+(24,5-20,4)^2+(16,7-20,4)^2+(20,7-20,4)^2+(18,8-20,4)^2+(25,1-20,4)^2+(15,5-20,4)^2+(16,8-20,4)^2+(15,8-20,4)^2+(22,9-20,4)^2+(21,6-20,4)^2+(24,3-20,4)^2+(13,1-20,4)^2+(19,7-20,4)^2+(18,8-20,4)^2+(15,0-20,4)^2+(12,5-20,4)^2+(19,9-20,4)^2+(10,1-20,4)^2+(15,1-20,4)^2+(30,8-20,4)^2+(20,8-20,4)^2+(18,5-20,4)^2+(26,6-20,4)^2+(27,6-20,4)^2$**]/31**

$Var(X) = 24,86x10^{18}\ m^6$

f) To obtain a parameter that has the same dimension as our variable we can take the square root of variance, called the ***Standard Deviation***. Computed as:

$$s_X = [Var(X)]^{1/2} = \{[\sum_{i=1}^{N}(x_i - x_m)^2]/N\}^{1/2}$$

$S_X = 4{,}98x10^9\ m^3$

g) To compare the two variables we must use a nondimensional parameter. The ***Coefficient Variation*** is defined as the ratio of the standard deviation of a variable to its mean:

$$C_{vX} = s_X / x_m$$

$C_{vX} = 4{,}98/20{,}4 = 0{,}24$

h) Another significant parameter that characterizes the skew of distribution is ***Coefficient of Skewness.*** We can measure the skew by the mean of the cubes of the differences x_i-x_m . If the data were perfectly symmetrical, the cube of a positive difference would be canceled by the cube of an equal negative difference, and the mean of the cubes would vanish. To have a nondimensional parameter we divide the mean by the cube of the standard deviation.

$$C_{sX} = \{[\sum_{i=1}^{N}(x_i - x_m)^3]/N\}/ s_x^3$$

$C_{sX} = \{[(20{,}2-20{,}4)^3+(24{,}7-20{,}4)^3+(19{,}3-20{,}4)^3+(27{,}2-20{,}4)^3+(27{,}9-20{,}4)^3+(22{,}7-20{,}4)^3+(22{,}4-20{,}4)^3+(24{,}5-20{,}4)^3+(16{,}7-20{,}4)^3+(20{,}7-20{,}4)^3+(18{,}8-20{,}4)^3+(25{,}1-20{,}4)^3+(15{,}5-20{,}4)^3+(16{,}8-20{,}4)^3+(15{,}8-20{,}4)^3+(22{,}9-20{,}4)^3+(21{,}6-20{,}4)^3+(24{,}3-20{,}4)^3+(13{,}1-20{,}4)^3+(19{,}7-20{,}4)^3+(18{,}8-20{,}4)^3+(15{,}0-20{,}4)^3+(12{,}5-20{,}4)^3+(19{,}9-20{,}4)^3+(10{,}1-20{,}4)^3+(15{,}1-20{,}4)^3+(30{,}8-20{,}4)^3+(20{,}8-20{,}4)^3+(18{,}5-20{,}4)^3+(26{,}6-20{,}4)^3+(27{,}6-20{,}4)^3]/31\}/ s_X^3$

$C_{sX} = 0{,}027$

SOLUTION 2

a) As a first step we can classify the data into class intervals of equal length (100 kg). Plotting the number of observations in each class interval as a horizontal line, we obtain a step diagram called the ***histogram***.

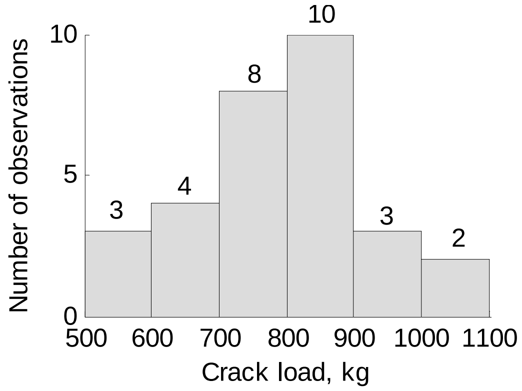

Figure 4. Histogram of crack load

b)

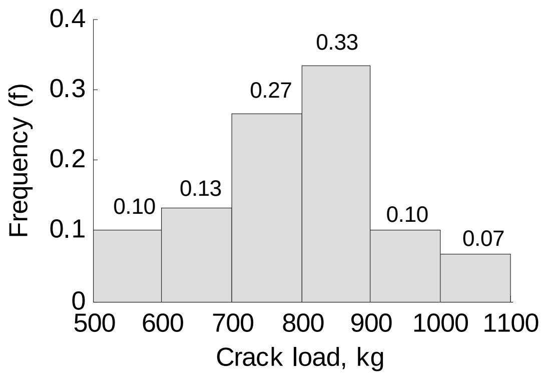

Figure 5. Frequency histogram of crack load

c) Arithmetic mean computed as:

$$x_m = \left[\sum_{i=1}^{N} x_i\right]/N$$

$=[(635+810+1045+890+520+800+710+760+860+990+660+730+790+570+810+740+940+860+840+595+930+840+790+740+810+685+780+610+850+1080]/30$

x_m = 789 kg

Median is another parameter that characterizes the central value. It is one half of observations remain. Arranging the data in the increasing order, the observation in the center [for this problem (15^{th} +16^{th})/2] is estimate of the Median.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	**15**
635	520	570	595	610	660	685	710	730	740	740	760	780	790	**790**
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
800	810	810	810	840	840	850	860	860	890	930	940	990	1045	1080

M_x = (790+800)/2 = 795 kg

d)

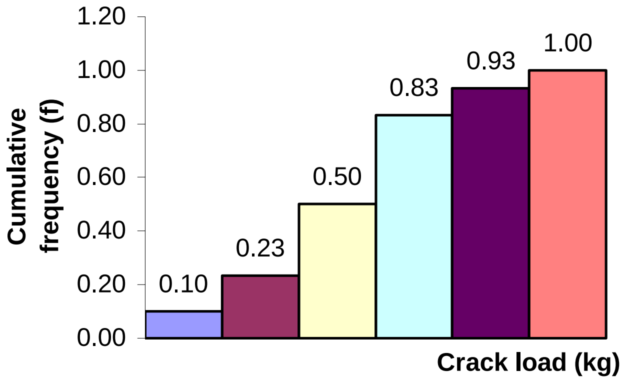

Figure 6: Cumulative frequency histogram of crack loads

e) The mean of the squares of the differences, called the ***Variance*** is a measure of the scatter of the data around its mean. Computed as:

$$\mathbf{Var(X) = [\sum_{i=1}^{N}(x_i - x_m)^2]/N}$$

Var(X) = [$(635-789)^2+(810-789)^2+(1045-789)^2+(890-789)^2+(520-789)^2+(800-789)^2+(710-789)^2+(760-789)^2+(860-789)^2+(990-789)^2+(660-789)^2+(730-789)^2+(790-789)^2+(570-789)^2+(810-789)^2+(740-789)^2+(940-789)^2+(860-789)^2+(840-789)^2+(595-789)^2+(930-789)^2+(840-789)^2+(790-789)^2+(740-789)^2+(810-789)^2+(685-789)^2+(780-789)^2+(610-789)^2+(850-789)^2+(1080-789)^2$**]/30**

Var(X) = 17352,3 kg²

f) To obtain a parameter that has the same dimension as our variable we can take the square root of variance, called the ***Standard Deviation***. Computed as:

$$\mathbf{s_X = [Var(X)]^{1/2} = \{[\sum_{i=1}^{N}(x_i - x_m)^2]/N\}^{1/2}}$$

S_X = 131,7 kg

g) To compare the two variables we must use a nondimensional parameter. The ***Coefficient Variation*** is defined as the ratio of the standard deviation of a variable to its mean:

$$\mathbf{C_{vX} = s_X / x_m}$$

C_{vX} = 131,7/789 = 0,167

h) Another significant parameter that characterizes the skew of distribution is ***Coefficient of Skewness.*** We can measure the skew by the mean of the cubes of the differences x_i-x_m . If the data were perfectly symmetrical, the cube of a positive difference would be canceled by the cube of an equal negative difference, and the mean of the cubes would vanish. To have a nondimensional parameter we divide the mean by the cube of the standard deviation.

$$C_{sX} = \{[\sum_{i=1}^{N}(x_i - x_m)^3]/N\}/\ s_x^3$$

$C_{sX}=\{[(635-789)^3+(810-789)^3+(1045-789)^3+(890-789)^3+(520-789)^3+(800-789)^3+(710-789)^3+(760-789)^3+(860-789)^3+(990-789)^3+(660-789)^3+(730-789)^3+(790-789)^3+(570-789)^3+(810-789)^3+(740-789)^3+(940-789)^3+(860-789)^3+(840-789)^3+(595-789)^3+(930-789)^3+(840-789)^3+(790-789)^3+(740-789)^3+(810-789)^3+(685-789)^3+(780-789)^3+(610-789)^3+(850-789)^3+(1080-789)^3]/30\}/\ s_X^3$

C_{sX} = 0,104

SOLUTION 3

a) As a first step we can classify the data into six class intervals of equal length (109 mm). Plotting the number of observations in each class interval as a horizontal line, we obtain a step diagram called the ***histogram***.

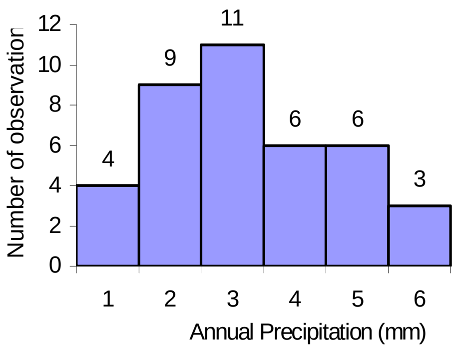

Figure 7:Histogram of Annual Precipitation

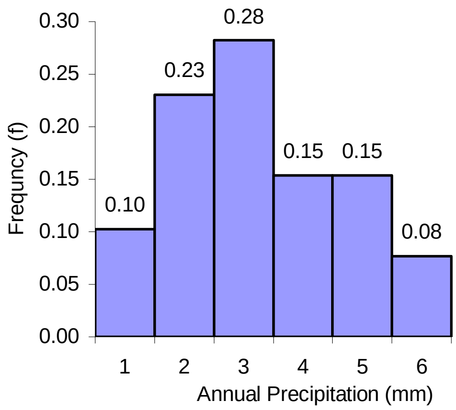

Figure 8: Frequency Histogram of Annual Precipitation

b)

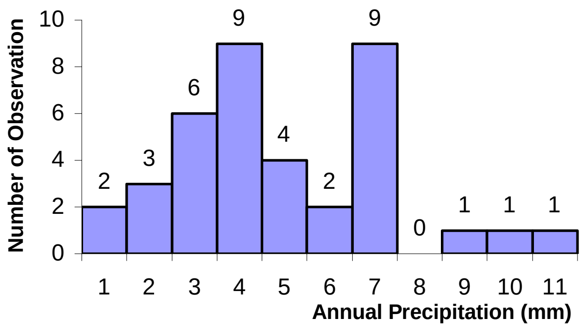

Figure 9: Histogram of Annual Precipitation with 11 class interval

Figures 7 and 9 are show the effect of the number of class intervals using wider and narrower intervals. The number of the class intervals in Figure 9 is more than that in Figure 7.

c)

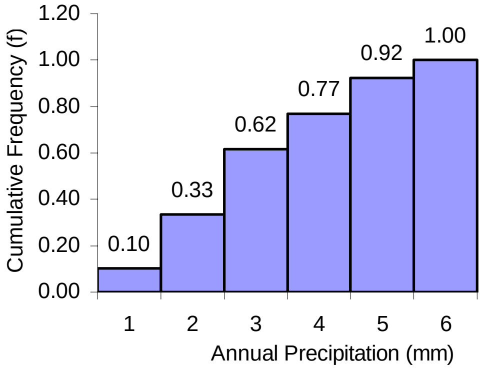

Figure 10: Cumulative Frequency Histogram of Annual Precipitation

d) Arithmetic mean computed as:

$$x_m = [\sum_{i=1}^{N} x_i]/N$$

$$x_m = [\sum_{i=1}^{39} x_i]/39$$

x_m = 773,36 kg

Median is another parameter that characterizes the central value. It is one half of observations remain. Arranging the data in the increasing order, the observation in the center (for this problem 19th = 731) is estimate of the Median.

471	520	555	556	585	639	648	658	661	675
1	2	3	4	5	6	7	8	9	10
676	679	685	696	707	716	720	725	**731**	733
11	12	13	14	15	16	17	18	**19**	20
737	748	760	763	815	837	867	892	893	893
21	22	23	24	25	26	27	28	29	30
907	908	913	919	926	932	1046	1139	1230	
31	32	33	34	35	36	37	38	39	

M_x = 731 mm

e) The mean of the squares of the differences, called the ***Variance*** is a measure of the scatter of the data around its mean. Computed as:

$$Var(X) = [\sum^{N}(x_i - x_m)^2]/N$$

$$\mathbf{Var(X) = [\sum_{i=1}^{39}(x_i - x_m)^2]/39}$$

Var(X) ***= 25771,46 mm²***

To obtain a parameter that has the same dimension as our variable we can take the square root of variance, called the ***Standard Deviation***. Computed as:

$$\mathbf{s_X = [Var(X)]^{1/2} = \{[\sum_{i=1}^{N}(x_i - x_m)^2]/N\}^{1/2}}$$

$S_X = 160{,}53\ mm$

To compare the two variables we must use a nondimensional parameter. The ***Coefficient Variation*** is defined as the ratio of the standard deviation of a variable to its mean:

$$\mathbf{C_{vX} = s_X / x_m}$$

$C_{vX} = 160{,}53/773{,}36 = 0{,}21$

f)

Another significant parameter that characterizes the skew of distribution is ***Coefficient of Skewness.*** We can measure the skew by the mean of the cubes of the differences x_i-x_m . If the data were perfectly symmetrical, the cube of a positive difference would be canceled by the cube of an equal negative difference, and the mean of the cubes would vanish. To have a nondimensional parameter we divide the mean by the cube of the standard deviation.

$$\mathbf{C_{sX} = \{[\sum_{i=1}^{N}(x_i - x_m)^3]/N\}/\ s_x^3}$$

$$\mathbf{C_{sX} = \{[\sum_{i=1}^{39}(x_i - 773{,}36)^3]/39\}/\ 162{,}63^3}$$

$C_{sX} = 0{,}653$

We can not say symmetrical data. Because the coefficient of skewness is significantly different from zero.