DESCRITPIVE STATISITICS

UNIT III

Measures of dispersion

It helps us to study the extent of the variation of the individual observations from the central value and among themselves

For example consider the three series

	Series A	Series B	Series C
	201	200	75
	198	200	405
	199	200	150
	200	200	270
	202	200	1**00**
total	1000	1000	1000
A.M	200	200	200

Measures of Dispersion

Absolute and Relative measures.

Absolute measures are given in the same units as the individual data whereas Relative measures are free from units of measurements and are called as co-efficients. To compare two or more series for their variability we use relative measures.

1) Range and co-efficient of range
2) Quartile deviation (Q.D) and co-efficient of Q.D.
3) Mean deviation (M.D) and co-efficient of M.D.
4) Standard deviation (S.D), co-efficient of S.D. and co-efficient of variation

1) **Range and co-efficient of range**

   Range is defined as the difference between the highest and lowest value of x

   Range R = largest value – smallest value =L – S

   Co-efficient of range = (L - S) /(L+S)

   Find the range and its co-efficient

   234, 22, 425, 325, 78, 236, 120, 422

Largest value L= 425
Smallest value S = 22
Range = 425 – 22 = 403
Co-efficient of range = (L - S) /(L+S)
= (425-22)/(425 +22)
= 403/ 447 = 0.90

Find range and its co-efficient

X	F
10 -20	4
20-30	12
30-40	25
40-50	16
50-60	6

Largest value of X , L = 60

Smallest value of X , S = 10

Range = L – S = 60 -10 =50

Co-efficient of range = (L - S) /(L+S) = (60 -10) /(60+10) = 50/70 =0.71

Merits:

1. It is simple to understand.
2. It is easy to calculate.
3. In certain types of problems like quality control, weather forecasts, share price analysis, etc., range is most widely used.

Demerits:

1. It is very much affected by the extreme items.
2. It is based on only two extreme observations.
3. It cannot be calculated from open-end class intervals.
4. It is not suitable for mathematical treatment.
5. It is a very rarely used measure.

2) **Quartile deviation (Q.D) and co-efficient of Q.D.**
Inter quartile range
Definition: Q.D. is defined as the difference between upper and lower quartile divided by 2
i.e., Q.D. = $(Q_3 - Q_1)/2$
Co-efficient of Q.D. = $(Q_3 - Q_1)/(Q_3 + Q_1)$

Find the Q.D. and its co-efficient
23, 32, 46, 67, 12, 89, 32, 45, 90,15
Sol.
Arrange the data in ascending order
12,15,23,32,32,45,46,67,89,90

Q_3 = value of 3(n+1)/4 th obsevation
= 8.25 th value
= 8^{th} value + 0.25 (9^{th} -8^{th} value)
=67 + 0.25 (89 - 67) = 67 + 0.25 x22 = 67 + 5.5 = 72.5
Q_1 = value of (n+1)/4 th obsevation
= 2.75 th value
= 2^{nd} value + 0.75 (3^{rd} - 2^{nd} value)
=15 + 0.75 (23 - 15) = 15 + 0.75 x 8 = 15 +6 = 21

Q.D. = **$(Q_3 - Q_1)/2$ = (72.5 – 21)/2 = 51.5/2 =25.75**

Co-efficient of Q.D. = $(Q_3 - Q_1)/(Q_3 + Q_1)$
= (72.5 -21) /(72.5 + 21)
= 51.5 / 93.5 = 0.55

Find Quartile deviation and its co-efficient

X	F
20	5
22	8
24	12
26	6
28	3
30	2

Solution

Q_3 = value of x corresponding to the c.f. just greater than or = 3N/4

Q_1 = value of x corresponding to the c.f. just greater than or = N/4

X	F	c.f.
20	5	5
22	8	13
24	12	25
26	6	31
28	3	34
30	2	36
	N=36	

N/4 = 36/4 = 9

3N/4 = 3 x 9 =27

Q_1 = value of x corresponding to the c.f. just greater than N/4 i.e. 13

$Q_1 = 22$

Q_3 = value of x corresponding to the c.f. just greater than 3N/4 i.e. 31

$Q_3 = 26$

Q.D. = $(Q_3 - Q_1)/2$ = (26 -22)/2 = 4/2 =2

Co-efficient of Q.D. = $(Q_3 - Q_1)/(Q_3 + Q_1)$

= (26-22) /(26 + 22)

= 4/48 = 1/12

Find the Q.D. and Its co-efficient

C.I	F
20-25	3
25-30	8
30-35	16
35-40	22
40-45	15
25-50	5

Sol.

C.I	F	c.f.
20-25	3	3
25-30	8	11
30-35	16	27
35-40	23	50
40-45	15	65
25-50	5	70
	N=70	

$N/4 = 70/4 = 17.5$

$3N/4 = 3 \times 17.5 = 52.5$

Q_1 class is 30-35, Q_3class = 40- 45

$Q_1 = L_1 + \{(N/4 – c.f_1) \times c_1/f_1\}$

Q_1 classis 30-35, $L_1 = 30$, $c_1 = 35-30 = 5$, $f_1 = 16$, $c.f_1 = 11$

$Q_1 = 30 + \{17.5 – 11) \times 5 / 16\}$

$= 30 + \{6.5 \times 5 / 16\} = 30 + 2.03 = 32.03$

$Q_3 = L_3 + \{(3N/4 – c.f_3) \times c_3/f_3\}$

Q_3 class is 40-45

$L_3 = 40$, $c_3 = 45 – 40 = 5$, $f_3 = 15$, $c.f_3 = 50$

$Q_3 = L_3 + \{(3N/4 – c.f_3) \times c_3/f_3\}$

$= 40 + \{(52.5 – 50) \times 5/ 15\}$

$= 40 + \{2.5 \times 5/15\} = 40 + 0.83 = 40.83$

Q.D. $= (Q_3 - Q_1) /2 = (40.83 – 32.03)/2 = 8.8/2 = 4.4$

Co-efficient of Q.D. $= (Q_3 - Q_1) / (Q_3 + Q_1)$

$= (40.83 – 32.03) / (40.83 + 32.03)$

$= 8.8 / 72.86 = 0.12$

Merits:

1. It is Simple to understand and easy to calculate
2. It is not affected by extreme values.
3. It can be calculated form data with open end classes also.

Demerits:

1. It is not based on all the items. It is based on two positional values Q1 and Q3 and ignores the extreme 50% of the items.
2. It is not amenable to further mathematical treatment.
3. It is affected by sampling fluctuations.

Mean Deviation M.D. or average deviation

Mean deviation is defined as the arithmetic mean of the absolute deviations of the observations from mean or median or mode. By absolute deviation we mean the positive deviation of the observations from average.

Calculation of M.D.

Raw data

M.D. about A = $\frac{\Sigma\ |x-A\ |}{n}$

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A}$

Where A is mean or median or mode

Calculate mean deviation about mean for the following data

10,20,30,40,50

Solution

A.M. = (10 + 20 + 30 + 40 + 50) /5 = 150/5 = 30

M.D. about A = $\frac{\Sigma\ |x-A\ |}{n} = \frac{\Sigma\ |x-30\ |}{5} =$ 60 /5 =12

Co-efficient of M.D. = $\frac{M.D.\ about\ mean}{mean}$ = 12/30 = 0.4

X	$\mid x - A \mid$ = ∣x – 30 I
10	20
20	10
30	0
40	10
50	20
	60

Daily earnings in Rs. (X) coolies are given. Calculate all three mean deviations and its coefficient

X: 32,51, 23,46, 20, 78, 57, 56,57,30

X	$\mid x - 45 \mid$	$\mid x - 48.5 \mid$	$\mid x - 57 \mid$
32	13	16.5	25
51	6	2.5	6
23	22	25.5	34
46	1	2.5	11
20	25	28.5	37
78	33	29.5	21
57	12	8.5	0
56	11	7.5	1
57	12	8.5	0
30	15	18.5	27
	150	148	162

A.M. = $\frac{\Sigma x}{n}$ = $\frac{450}{10}$ = 45

Median

Arranging in ascending order

20,23,30,32,46,51,56,57,57,78

Md =(n+1)/2 th item = [(10+1)/2 =5.5th item] = (46 + 51) /2 = 48.5

Mode, Z = value which repeats more often = 57

M.D. about mean= $\frac{\Sigma\ |x-A.mI}{n} = \frac{\Sigma\ |x-45\ |}{10} =$ 150/10 =15

Co-efficient of M.D. = $\frac{M.D.\ about\ mean}{mean}$ = 15/45 = 0.33

M.D. about median= $\frac{\Sigma\ |x-md\ |}{n} = \frac{\Sigma\ |x-48.5\ |}{10} =$148/10 =14.8

Co-efficient of M.D. = $\frac{M.D.\ about\ md}{md}$ = 14.8/48.5 = 0.31

M.D. about mode= $\frac{\Sigma\ |x-Z\ |}{n} = \frac{\Sigma\ |x-57\ |}{10} =$162/10 =16.2

Co-efficient of M.D. = $\frac{M.D.\ about\ mode}{mode}$ = 16.2 /57 = 0.28

Discrete Data

M.D. about A = $\frac{\Sigma[f\ |x-A\ |]}{\Sigma f}$

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A}$

Where A is mean or median or mode

Calculate the mean deviation

X	2	4	6	8	10
f	1	4	6	4	1

x	f	xf	ǀx – a.m. ǀ= ǀx - 6 ǀ	f ǀ x - 6 ǀ
2	1	2	4	4
4	4	16	2	8
6	6	36	0	0
8	4	32	2	8
10	1	10	4	4
	16	96		24

Mean = $\frac{\Sigma fx}{\Sigma f} = \frac{96}{16}$ = 6

M.D. about A = $\frac{\Sigma[f\ |x-A\ |]}{\Sigma f} = 24/16 = 1.5$

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A} = 1.5\ /6 = 0.25$

Continuous Data

M.D. about A = $\frac{\Sigma[f\ |\ m-A\ |]}{\Sigma f}$

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A}$

Where A is mean or median or mode

Calculate mean deviation about mean

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60 -70
F	6	5	8	15	7	6	3

Mean = $\frac{\Sigma(fm)}{\Sigma f}$ = 1670/50 =33.4

x	f	m	mf	$\lvert m - A \rvert = \lvert m - 33.4 \rvert$	$f \lvert m - A \rvert$
0-10	6	5	30	28.4	170.4
10-20	5	15	75	18.4	92.0
20-30	8	25	200	8.4	67.2
30-40	15	35	525	1.6	24.0
40-50	7	45	315	11.6	81.2
50-60	6	55	330	21.6	129.6
60-70	3	65	195	31.6	94.8
	50		1670		659.2

M.D. about mean = $\frac{\Sigma[f \lvert m-A \rvert]}{\Sigma f}$ = 659.2 /50 =13.18

Co-efficient of mean deviation = $\frac{M.D.\ about\ mean}{mean}$ = 13.18/ 33.4 = 0.39

Merits:

1. It is simple to understand and easy to compute.
2. It is rigidly defined.
3. It is based on all items of the series.
4. It is not much affected by the fluctuations of sampling.
5. It is less affected by the extreme items.
6. It is flexible, because it can be calculated from any average.
7. It is a better measure of comparison.

Demerits:

1. It is not a very accurate measure of dispersion.
2. It is not suitable for further mathematical calculation.
3. It is rarely used. It is not as popular as standard deviation.
4. Algebraic positive and negative signs are ignored. It is mathematically unsound and illogical.

Standard deviation

It is defined as the square root of the arithmetic mean of the squares of deviations of the observations from arithmetic mean.

Standard deviation is denoted by σ

Raw data

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{(3038/10)} = \sqrt{303.8} = 17.43$$

X: 32,51, 23,46, 20, 78, 57, 56,57,30

X	$(x-45)$	$(x-x)^2 = (x-45)^2$
32	-13	169
51	6	36
23	-22	484
46	1	1
20	-25	625
78	33	1089
57	12	144
56	11	121
57	12	144
30	-15	225
450		3038

$$\text{A.M.} = \frac{\Sigma x}{n} = \frac{450}{10} = 45$$

Discrete Data

$$\sigma = \sqrt{\frac{\Sigma[f(x-\bar{x})^2]}{\Sigma f}}$$

X	2	4	6	8	10
f	1	4	6	4	1

Mean = $\frac{\Sigma fx}{\Sigma f}$ = $\frac{96}{16}$ = 6

x	f	xf	(x – a.m.)=(x – 6)	$(x-6)^2$	$f(x-6)^2$
2	1	2	-4	16	16
4	4	16	-2	4	16
6	6	36	0	0	0
8	4	32	2	4	16
10	1	10	4	16	16
	16	96			64

σ =√ $\frac{\Sigma[f(x-\bar{x})^2]}{\Sigma f}$ = √ [64/16] =√4 =2

Continuous data

σ =√ $\frac{\Sigma[f(m-\bar{x})^2]}{\Sigma f}$

Calculate standard deviation

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60 -70
F	6	5	8	15	7	6	3

Mean = $\frac{\Sigma(fm)}{\Sigma f}$ = 1670/50 =33.4

X	f	M	mf	(m – A) = (m -33.4)	$(m-33.4)^2$	$f(m-33.4)^2$
0-10	6	5	30	-28.4	806.56	4839.36
10-20	5	15	75	-18.4	338.56	1692.8
20-30	8	25	200	-8.4	70.56	564.48
30-40	15	35	525	1.6	2.56	38.4
40-50	7	45	315	11.6	134.56	941.92
50-60	6	55	330	21.6	466.56	2799.36
60-70	3	65	195	31.6	998.56	2995.68
	50		1670			13872.00

$$\sigma = \sqrt{\frac{\Sigma[f(m-\bar{x})^2]}{\Sigma f}} = \sqrt{[13872.00/50]} = \sqrt{277.44} = 16.66$$

Step-deviation method.

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times C$$

Where d = (m-A)/C

Merits: It is rigidly defined and its value is always definite and based on all the observations and the actual signs of deviations are used.

1. As it is based on arithmetic mean, it has all the merits of an arithmetic mean.
2. It is the most important and widely used measure of dispersion.
3. It is possible for further algebraic treatment.
4. It is less affected by the fluctuations of sampling and hence stable.
5. It is the basis for measuring the coefficient of correlation and sampling.

Demerits:

1. It is not easy to understand and it is difficult to calculate.
2. It gives more weight to extreme values because the values are squared up.
3. As it is an absolute measure of variability, it cannot be used for the purpose of comparison.

COEFFICIENT OF VARIATION:

If we want to compare the variability of two or more series, we can use C.V. The series or groups of data for which the C.V. is greater indicate that the group is more variable, less stable, less uniform, less consistent or less homogeneous. If the C.V. is less, it indicates that the group is less variable, more stable, more uniform, more consistent or more homogeneous.

1.The means and standard deviation values for the number of runs of two players A and B are 55; 65 and 4.2; 7.8 respectively. Who is the more consistent player?

Given:	Player A	Player B
Mean($\overline{X}$)	55	65
S.D.(σ)	4.2	7.8

Coefficient of variation of Player A = $\frac{\sigma}{\overline{X}} X 100 = \frac{4.2}{55}$ X 100 = 7.64

Coefficient of variation of Player B = $\frac{\sigma}{\overline{X}} X 100 = \frac{7.8}{65}$ X 100 = 12

Coefficient of variation of player A is less. Therefore Player A is the more consistent player.

2.From the following data, find

(i) Which firm pays more amount as monthly wages

(ii) Which firm has greater variability in individual wages and

(iii) What are the mean monthly wage if the two firms merge.

	Firm I	Firm II
No of workers	100	200
Mean monthly wage (Rs.000)	7	8
S.D. of individual wages (Rs.000)	2	2.5

Solution:

Given: $N_1 = 100$ $N_2 = 200$

$\overline{X_1} = 7$ $\overline{X_2} = 8$

$\sigma_1 = 2$ $\sigma_2 = 2.5$

$$\overline{X} = \frac{\sum X}{N}$$

$$\sum X = N \text{ x } \overline{X}$$

(i) Total wage = Number x Mean

Total wage in firm I = N_1 x $\overline{X}_1$ = 100 x 7 = 700

Total wage in firm II = N_2 x $\overline{X}_2$ = 200 x 8 = 1600

Hence, firm II pays more amount as monthly wages.

(ii) Coefficient of variation(C.V.) = $\frac{\sigma}{\overline{X}} X100$

Coefficient of variation of firm I = $\frac{\sigma_1}{\overline{X_1}} X100 = \frac{2}{7} X100$ = 28.57

Coefficient of variation of firm II = $\frac{\sigma_2}{\overline{X_2}} X100 = \frac{2.5}{8} X100$= 31.25

Coefficient of variation of firm II is more. Hence, there is greater variabilityin individual wages in firm II

(iii) Combined Mean monthly wage:

$$\overline{X_{12}} = \frac{N_1\overline{X_1} + N_2\overline{X_2}}{N_1 + N_2} = \frac{100X7 + 200X8}{100 + 200} = \frac{700 + 1600}{300} = \frac{2300}{300} = \text{Rs.7.67}$$

(thousands)

4.From the following price of gold in a week, find the city in which the price was more stable.

Day	Mon	Tue	Wed	Thurs	Fri	Sat
City A	498	500	505	504	502	509
City B	500	505	502	498	496	505

Solution:

City A

X_1	X_1^2
498	248004
500	250000
505	255025
504	254016
502	252004
509	259081
$\sum X_1$ = 3018	$\sum X_1^2$ =1518130

$$\overline{X_1} = \frac{\sum X_1}{N_1} = 3018/6 = 503$$

$$\sigma_1 = \sqrt{\frac{\sum X_1^2}{N_1} - \left[\frac{\sum X_1}{N_1}\right]^2} = \sqrt{\frac{1518130}{6} - (503)^2} = \sqrt{253021.67 - 253009} = \sqrt{12.67} = 3.56$$

$$\text{C.V.}_A = \frac{\sigma_1}{\overline{X_1}} X100 = 3.56/503 \text{ X } 100 = 0.71$$

City B

X_2	X_2^2
500	250000
505	255025
502	252004
498	248004
496	246016
505	255025
$\sum X_2 = 3006$	$\sum X_2^2 = 1506074$

$$\overline{X_2} = \frac{\sum X_2}{N_2} = 3006/6 = 501$$

$$\sigma_2 = \sqrt{\frac{\sum X_2^2}{N_2} - \left[\frac{\sum X_2}{N_2}\right]^2} = \sqrt{\frac{1506074}{6} - (501)^2} = \sqrt{25101233 - 251001}$$

$$= \sqrt{11.33} = 3.37$$

$$\text{C.V.}_2 = \frac{\sigma_2}{\overline{X_2}} X100 = 3.37/501 \text{ X } 100 = 0.67$$

Coefficient of Variation of City B is less. Hence the price was more stable in City B.

4.Goals scored by two teams A and B in a series of football matches were observed as follows:

No of goals Scored in a match	No of Matches	
X	Team A (f_1)	Team B (f_2)
0	5	4
1	7	5
2	5	5
3	3	4
4	2	3
5	3	3

Which team A or B may be considered as a more consistent team?

Solution: Team A

Goals x	f_1	f_1x	x^2	f_1x^2
0	5	0	0	0
1	7	7	1	7
2	5	10	4	20
3	3	9	9	27
4	2	8	16	32
5	3	15	25	75
	$\sum f_1 = 25$	$\sum f_1x = 49$		$\sum f_1x^2 = 161$

$\overline{X_1} = \frac{\sum f_1 x}{\sum f_1}$ = 49/25 = 1.96

$$\sigma = \sqrt{\frac{\sum f_1 X^2}{\sum f_1} - \left[\frac{\sum f_1 X}{\sum f_1}\right]^2}$$

$$= \sqrt{\frac{161}{25} - (1.96)^2}$$

$= \sqrt{6.44 - 3.8416}$

$= \sqrt{2.5984} = 1.61$

C.V.$_A$ = $\frac{\sigma_1}{\overline{X_1}} X 100$ = 1.61/1.96 X 100 = 82.14

Team B

Goals x	f_2	f_2x	x^2	f_2x^2
0	4	0	0	0
1	5	5	1	5
2	5	10	4	20
3	4	12	9	36
4	3	12	16	48
5	3	15	25	75
	$\sum f_2 = 24$	$\sum f_2x = 54$		$\sum f_2x^2 = 184$

$\overline{X_2} = \frac{\sum f_2 x}{\sum f_2}$ = 54/24 = 2.25

$$\sigma = \sqrt{\frac{\sum f_2 X^2}{\sum f_2} - \left[\frac{\sum f_2 X}{\sum f_2}\right]^2}$$

$$= \sqrt{\frac{184}{24} - (2.25)^2}$$

$= \sqrt{7.6667 - 5.0625}$

$= \sqrt{2.6052} = 1.61$

C.V.$_B$ = $\frac{\sigma_2}{\overline{X_2}} X 100$ = 1.61/2.25 X 100 = 71.56

Coefficient of Variation of Team B is less. Hence Team B is the more consistent Team.

5.The marks in Business Mathematics of two sections of students of a college are given below. Find in which section the marks are more variable.

Marks	20-30	30-40	40-50	50-60	60-70
Section A	5	13	24	5	3
Section B	7	14	25	12	2

Section A

Marks	f_1	m	f_1m	m^2	f_1m^2
20-30	5	25	125	625	3125
30-40	13	35	455	1225	15925
40-50	24	45	1080	2025	48600
50-60	5	55	275	3025	15125
60-70	3	65	195	4225	12675
	$\sum f_1 = 50$		$\sum f_1 m = 2130$		$\sum f_1 m^2 = 95450$

$\overline{X_1} = \frac{\sum f_1 m}{\sum f_1}$ = 2130/50 = 42.6

$$\sigma_1 = \sqrt{\frac{\sum f_1 m^2}{\sum f_1} - \left[\frac{\sum f_1 m}{\sum f_1}\right]^2}$$

$$= \sqrt{\frac{95450}{50} - [42.6]^2} = \sqrt{1909 - 1814.76} = \sqrt{94.24} = 9.71$$

C.V.$_A$ = $\frac{\sigma_1}{\overline{X_1}} X100$=9.71/42.6 X 100 = 22.79

Section B

Marks	f_2	m	f_2m	m^2	f_2m^2
20-30	7	25	175	625	4375
30-40	14	35	490	1225	17150
40-50	25	45	1125	2025	50625
50-60	12	55	660	3025	36300
60-70	2	65	130	4225	8450
	$\sum f_2 = 60$		$\sum f_2 m = 2580$		$\sum f_2 m^2 = 116900$

$\overline{X_2} = \frac{\sum f_2 m}{\sum f_2}$ = 2580/60 = 43

$$\sigma = \sqrt{\frac{116900}{60} - [43]^2} = \sqrt{1948.33 - 1849} = \sqrt{99.33} = 9.97$$

C.V.$_B$ = $\frac{\sigma_2}{\overline{X_2}} X100$= 9.97/43 X 100 = 23.19

Coefficient of Variation of section B is more. Hence in section B, the marks are more variable.

Lorenz Curve:

Lorenz curve is a graphical method of studying dispersion. It was introduced by Max.O.Lorenz, a great Economist and a statistician, to study the distribution of wealth and income. It is also used to study the variability in the distribution of profits, wages, revenue, etc.

It is specially used to study the degree of inequality in the distribution of income and wealth between countries or between different periods. It is a percentage of cumulative values of one variable in combined with the percentage of cumulative values in other variable and then Lorenz curve is drawn.

The curve starts from the origin (0,0) and ends at (100,100). If the wealth, revenue, land etc are equally distributed among the people of the country, then the Lorenz curve will be the diagonal of the square. But this is highly impossible.

The deviation of the Lorenz curve from the diagonal, shows how the wealth, revenue, land etc are not equally distributed among people.

Example 16:

In the following table, profit earned is given from the number of companies belonging to two areas A and B. Draw in the same diagram their Lorenz curves and interpret them.

Profit earned (in thousands)	Number of Companies	
	Area A	Area B
5	7	13
26	12	25
65	14	43
89	28	57
110	33	45
155	25	28
180	18	13
200	8	6

Solution:

Profits			Area A			Area B		
In Rs.	Cumulati ve	Cumulati ve	No. of compani	Cumulati ve	Cumulati ve	No. of compani	Cumulati ve	Cumulati ve
5	5	1	7	7	5	13	13	6
26	31	4	12	19	13	25	38	17
65	96	12	14	33	23	43	81	35
89	185	22	28	61	42	57	138	60
110	295	36	33	94	65	45	183	80
155	450	54	25	119	82	28	211	92
180	630	76	18	137	94	13	224	97
200	830	100	8	145	100	6	230	100

LORENZ-CURVE