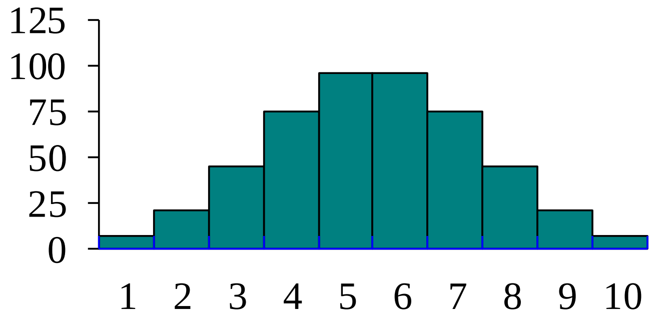
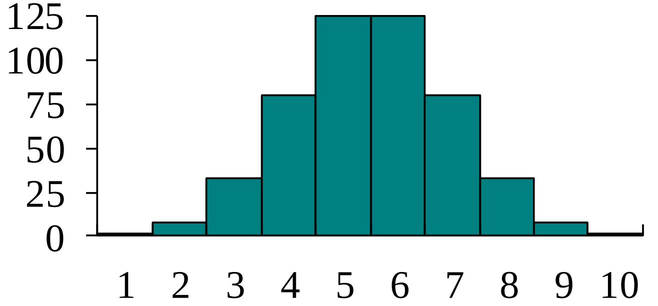
Measures of Dispersion

- *Measures of dispersion* are descriptive statistics that describe how dissimilar a set of scores are to each other i.e. measurement of degree of variability/spread.
  - The more similar the scores are to each other, the lower the measure of dispersion will be
  - The less similar the scores are to each other, the higher the measure of dispersion will be
  - In general, high dispersion means less reliability on control value leads to more risk.

Which of the distributions of scores has the larger dispersion?



- The upper distribution has more dispersion because the scores are more spread out
  - That is, they are less similar to each other



Measures of Dispersion

- Range (R)
- Quartile Deviation (QD)
- Mean Absolute Deviation (MAD)
- Standard Deviation (σ)

Range

- The *range* is defined as the difference between the largest score in the set of data and the smallest score in the set of data, X_{max} - X_{min}
- What is the range of the following data: 4 8 1 6 6 2 9 3 6 9
- The largest score (X_L) is 9; the smallest score (X_S) is 1; the range is $X_L - X_S = 9 - 1 = 8$

Quartile deviation

- The *quartile deviation* is defined as the difference of the first and third quartiles divided by two.
  - The first quartile is the 25^{th} percentile
  - The third quartile is the 75^{th} percentile
- *Quartile deviation* = $(Q_3 - Q_1) / 2$
- *Coefficient of quartile deviation* (using average size of quartile) = $(Q_3 - Q_1) / (Q_3 + Q_1)$

Quartile deviation Example

- What is the *quartile deviation* for the data to the right?
- 25 % of the scores are below 5
  - 5 is the first quartile
- 25 % of the scores are above 25
  - 25 is the third quartile
- Q.D. = $(Q_3 - Q_1) / 2 = (25 - 5) / 2 = 10$

2
4
6
8
10
12
14
20
30
60

← 5 = 25th %tile

← 25 = 75th %tile

Mean/average deviation

- M.D. or M.A.D. =

$$\frac{\sum \left| x_i - \overline{x} \right|}{n} \qquad \text{or} \qquad \frac{\sum f_i \left| x_i - \overline{x} \right|}{\sum f_i}$$

- Coefficient Of M.D. = $\dfrac{M.D.}{\overline{x}}$

Class	Midpoint x	Frequency f	f . x	|x - mean|	f|x - mean|
7 – 18	12.5	6	75	29.28	175.68
19 – 30	24.5	10	245	17.28	172.8
31 – 42	36.5	13	474.5	5.28	68.64
43 – 54	48.5	8	388	6.72	53.76
55 – 66	60.5	5	302.5	18.72	93.6
67 – 78	72.5	6	435	30.72	184.32
79 – 90	84.5	2	169	42.72	85.44
	SUM =	50	2089		834.24
		Mean =	41.78	MAD =	16.6848

Variance & Standard Deviation

- *Variance* is defined as the average of the square deviations:

$$\sigma^2 = \frac{\sum \left(x_i - \overline{x}\right)^2}{n} or = \frac{\sum f_i \left(x_i - \overline{x}\right)^2}{\sum f_i}$$

- *Standard deviation,* S.D. (σ) =

$$\sigma = \sqrt{\frac{\sum \left(x_i - \overline{x}\right)^2}{n}} or = \sqrt{\frac{\sum f_i \left(x_i - \overline{x}\right)^2}{\sum f_i}}$$

Standard Deviation & Coeff. Of Variation

- Standard deviation = √variance
- Variance = standard deviation^2

$$\sigma = \sqrt{\frac{\sum x_i^{\,2}}{n} - \bar{x}^2} \; or = \sqrt{\frac{\sum f_i x_i^{\,2}}{\sum f_i} - \bar{x}^2}$$

- *Coefficient of variation*

$$c.v. = \frac{\sigma}{\bar{x}}$$

Class	Midpoint, x	Frequency, f	f.x	f.x^2
7 – 18	12.5	6	75	937.5
19 – 30	24.5	10	245	6002.5
31 – 42	36.5	13	474.5	17319.25
43 – 54	48.5	8	388	18818
55 – 66	60.5	5	302.5	18301.25
67 – 78	72.5	6	435	31537.5
79 – 90	84.5	2	169	14280.5
	SUM =	50	2089	107196.5

Std. Dev. = 19.959

Skewness

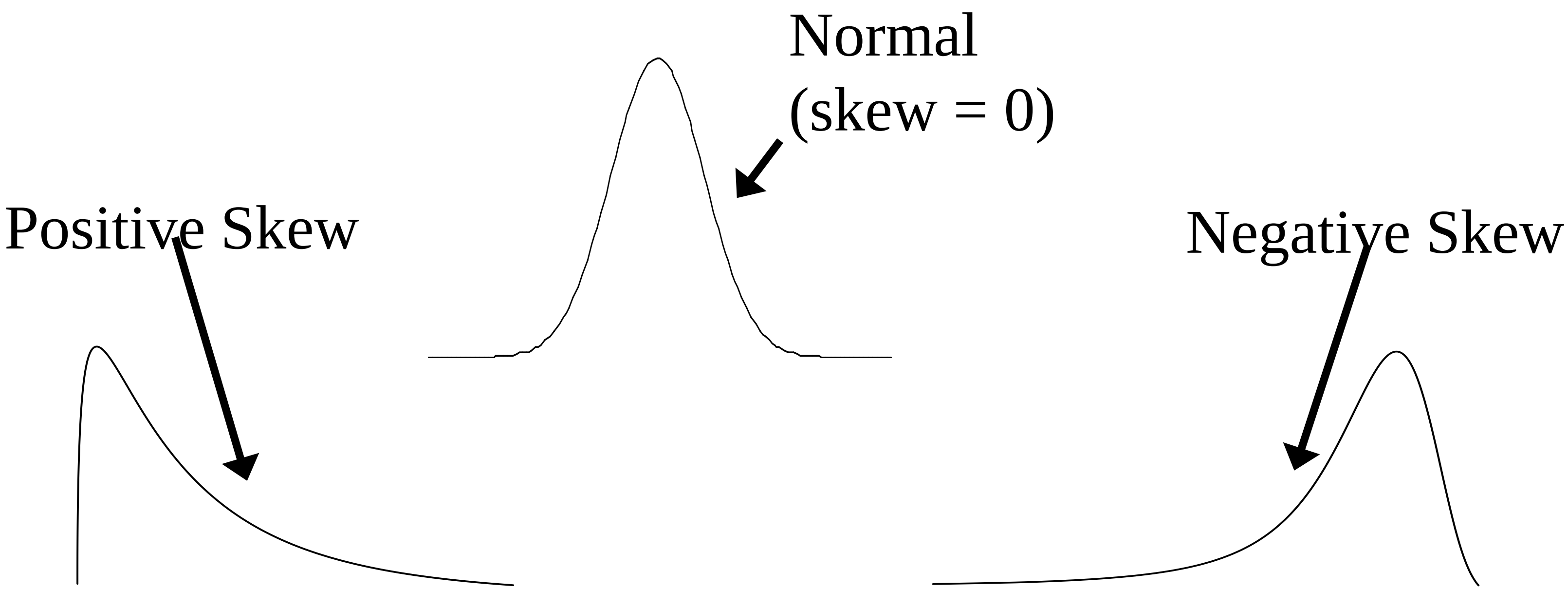
- Measurement of degree of asymmetry.
- +vely skewed : A.M. > Median > Mode
- -vely skewed : A.M. < Median < Mode
- Karl Pearson's coefficient of skewness

$$\frac{A.M. - Mode}{\sigma} \; or = \frac{3(A.M. - Median)}{\sigma}$$

- This value of skewness lies between -3 to +3
- For asymmetric curve, empirical relationship:
- A.M.- Mode = 3 (A.M. – Median)

Measure of Skew

- *Skew* is a measure of symmetry in the distribution of scores



Class	Midpoint, x	Frequency, f	Cumulative frequency
7 – 18	12.5	6	6
19 – 30	24.5	10	16
31 – 42	36.5	13	29
43 – 54	48.5	8	37
55 – 66	60.5	5	42
67 – 78	72.5	6	48
79 – 90	84.5	2	50
	SUM =	50	

Median Class: 30.5 – 42.5

Md = 30.5 + (25-16)*12/13 = 38.806

Skewness = 3(41.8 – 38.806)/19.959 = 0.45

Concept of Moment

- For representation of all statistical properties of a data set, r^{th} moment of about mean

$$\mu_r = \frac{\sum (x_i - \overline{x})^r}{n}$$

1st moment $\mu_1 = 0$

1st moment about origin = $\bar{x}$ → Mean

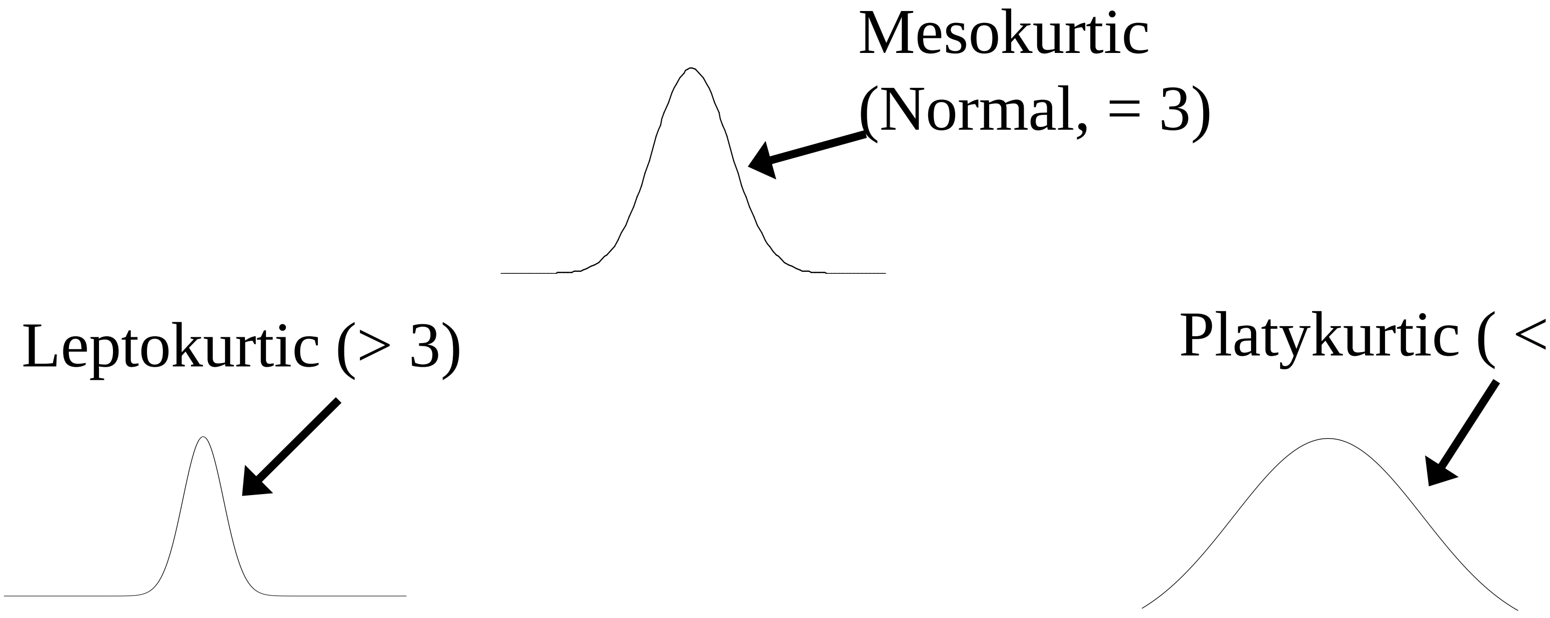
2nd moment = $$\mu_2 = \frac{\sum (x_i - \bar{x})^2}{n}$$ → Variance

3rd moment = $$\mu_3 = \frac{\sum (x_i - \bar{x})^3}{n}$$ → Degree of skewness $\mu_3 = 0$ for symmetric curve

4th moment = $$\mu_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$ → Kurtosis

Kurtosis

- Kurtosis measures whether the scores are spread out more or less than they would be in a normal (Gaussian) distribution
- Kurtosis = $\mu_4 / (\mu_2)^2$



Xi	μ1	μ2	μ3	μ4
50	8.04	64.6416	519.7185	4178.53645
40	-1.96	3.8416	-7.52954	14.7578906
41	-0.96	0.9216	-0.88474	0.84934656
17	-24.96	623.0016	-15550.1	388130.994
11	-30.96	958.5216	-29675.8	918763.658
7	-34.96	1222.202	-42728.2	1493776.75
22	-19.96	398.4016	-7952.1	158723.835
44	2.04	4.1616	8.489664	17.3189146
28	-13.96	194.8816	-2720.55	37978.838
21	-20.96	439.3216	-9208.18	193003.468
19	-22.96	527.1616	-12103.6	277899.353
23	-18.96	359.4816	-6815.77	129227.021
37	-4.96	24.6016	-122.024	605.238723
51	9.04	81.7216	738.7633	6678.41991
54	12.04	144.9616	1745.338	21013.8655
42	0.04	0.0016	6.4E-05	2.56E-06
86	44.04	1939.522	85416.53	3761744.04
41	-0.96	0.9216	-0.88474	0.84934656
78	36.04	1298.882	46811.69	1687093.41
56	14.04	197.1216	2767.587	38856.9252
72	30.04	902.4016	27108.14	814328.648
56	14.04	197.1216	2767.587	38856.9252
17	-24.96	623.0016	-15550.1	388130.994
7	-34.96	1222.202	-42728.2	1493776.75
69	27.04	731.1616	19770.61	534597.285
30	-11.96	143.0416	-1710.78	20460.8993
80	38.04	1447.042	55045.46	2093929.39
56	14.04	197.1216	2767.587	38856.9252
29	-12.96	167.9616	-2176.78	28211.0991
33	-8.96	80.2816	-719.323	6445.1353
46	4.04	16.3216	65.93926	266.394627
31	-10.96	120.1216	-1316.53	14429.1988
39	-2.96	8.7616	-25.9343	76.7656346
20	-21.96	482.2416	-10590	232556.961
18	-23.96	574.0816	-13755	329569.683
29	-12.96	167.9616	-2176.78	28211.0991
34	-7.96	63.3616	-504.358	4014.69235
59	17.04	290.3616	4947.762	84309.8588
73	31.04	963.4816	29906.47	928296.794
77	35.04	1227.802	43022.17	1507496.77
36	-5.96	35.5216	-211.709	1261.78407
39	-2.96	8.7616	-25.9343	76.7656346
30	-11.96	143.0416	-1710.78	20460.8993
62	20.04	401.6016	8048.096	161283.845
54	12.04	144.9616	1745.338	21013.8655
67	25.04	627.0016	15700.12	393131.006
39	-2.96	8.7616	-25.9343	76.7656346
31	-10.96	120.1216	-1316.53	14429.1988
53	11.04	121.8816	1345.573	14855.1244
44	2.04	4.1616	8.489664	17.3189146
	0.0	**394.5**	**2576.5**	**366622.7**

$\mu 1$	$\mu 2$	$\mu 3$	$\mu 4$
0	394.5	2576.5	366622.7

Kurtosis $= \mu_4 / (\mu_2)^2$

$= 366622.7 / (394.5 \times 394.5)$

$= 2.36$

The peakedness of the distribution is slightly below the normal distribution curve (< 3)