NORMAL PROBABILITY CURVE

Dr. Kuldeep Kaur

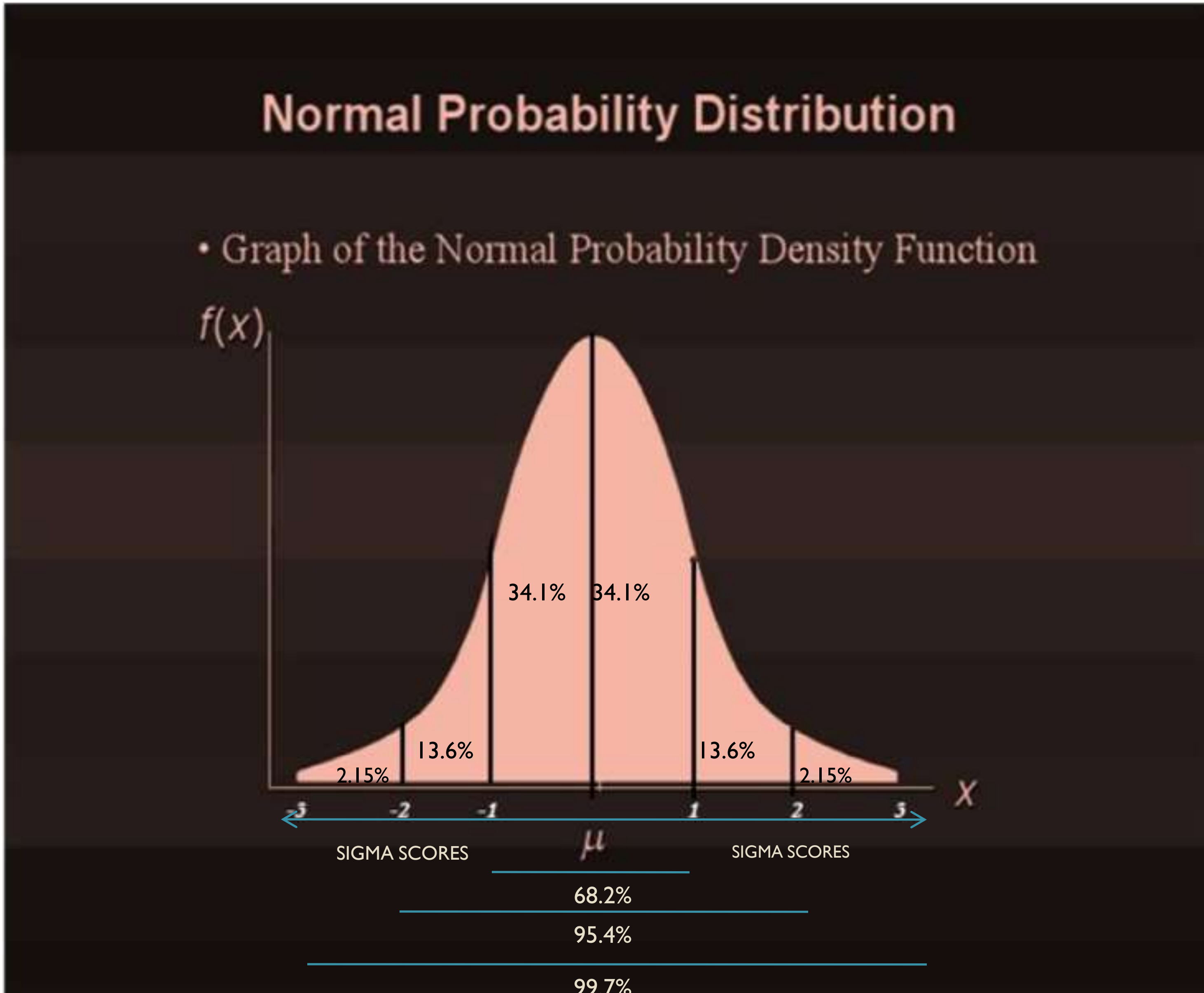
History

- ***Laplace and Gauss*** *(1777-1855) ,derived the* ***normal probability curve*** *independently,so the curve is also known as* ***gaussian curve*** *in the honor of Gauss.*

Introduction

- *NPC is the frequency polygon of any normal distribution.It is an ideal symmetrical frequency curve and is supposed to be based on the data of a population.*

- *Normal probability curve,is bell shaped curve and a graph representing a distribution of scores.*


Normal Probability Distribution
• Graph of the Normal Probability Density Function
f(x)
34.1%
34.1%
13.6%
13.6%
2.15%
2.15%
x
-3
-2
-1
1
2
3
μ
SIGMA SCORES
SIGMA SCORES
68.2%
95.4%
99.7%

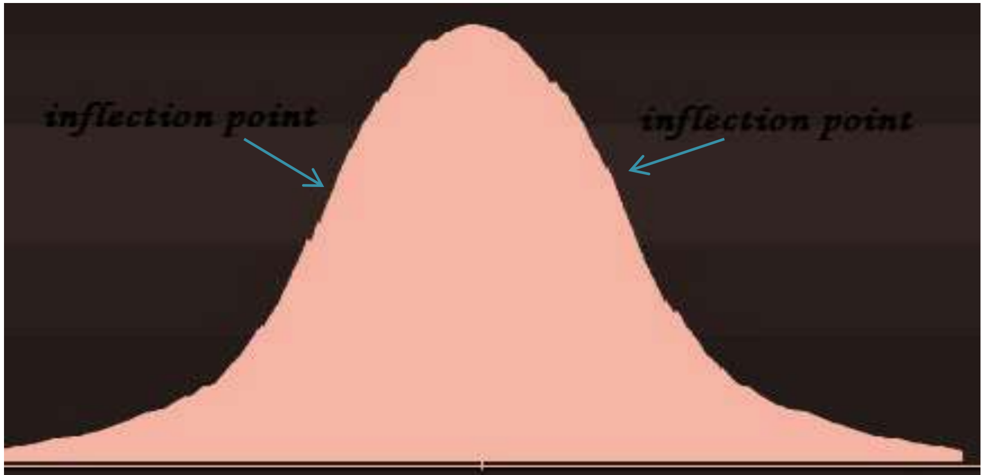
Characteristics/Properties of NPC

- NPC is a bell shaped curve.
- All the three central tendencies:mean,median and mode coincide in it and are equal.
- NPC is asymptotic.It approaches but never touches the base line.
- The NPC is bilateral symmetrical.It implies size,shape and slope of the curve on one side are identical to that of the other side.

The curve has its maximum height or ordinate at the starting point.ie.the mean of the distributiom.

The first and the third quartile(Q1&Q3) are at equal distance from Q2 or median.

The point of inflection(where the curvature changes its direction) is at point +- 1 σ,up and below the mean.

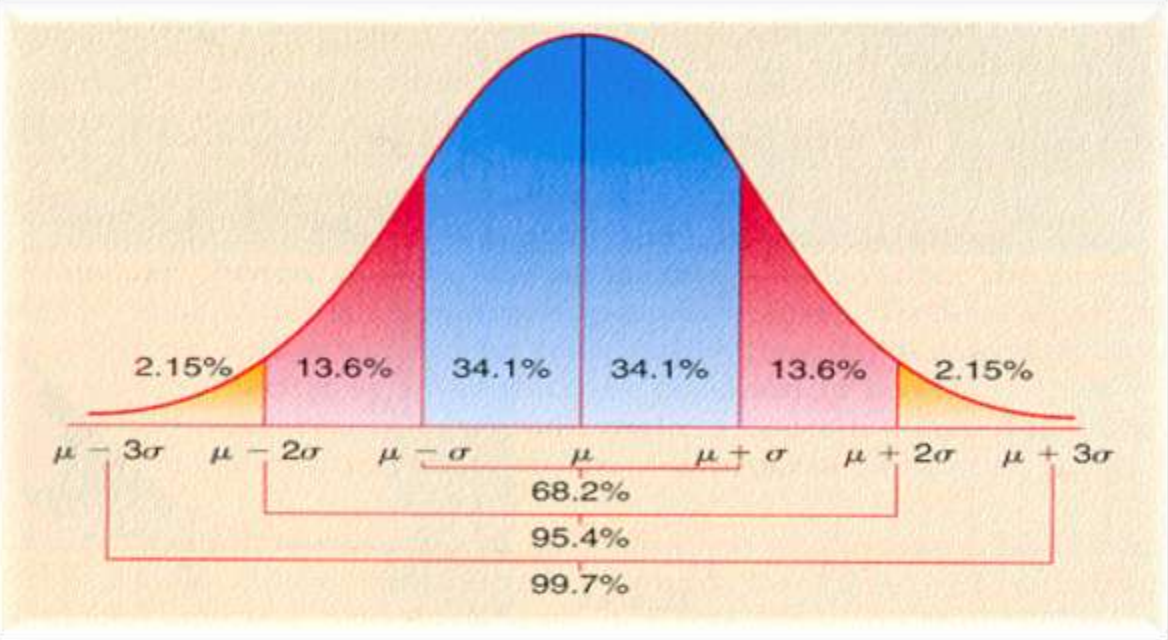


To find the deviations from the point of departure(mean),standard deviation of the distribution(σ) is used as a unit of measurement.

The total area under the curve is taken arbitarily to be 10,000,for a greater ease in the computation.

The curve extends on both sides,ie. -3σ distance on the left to +3σ on the right.

We may find that 3,413 cases out of 10,000 or 34.13% of the entire area of the curve lies between the mean and +1σ on the base line of the normal curve. Similarly another 34.13% of the cases lie between the mean and -1σ on the base line. Rest of the percentage divisions are shown in the diagram ahead:

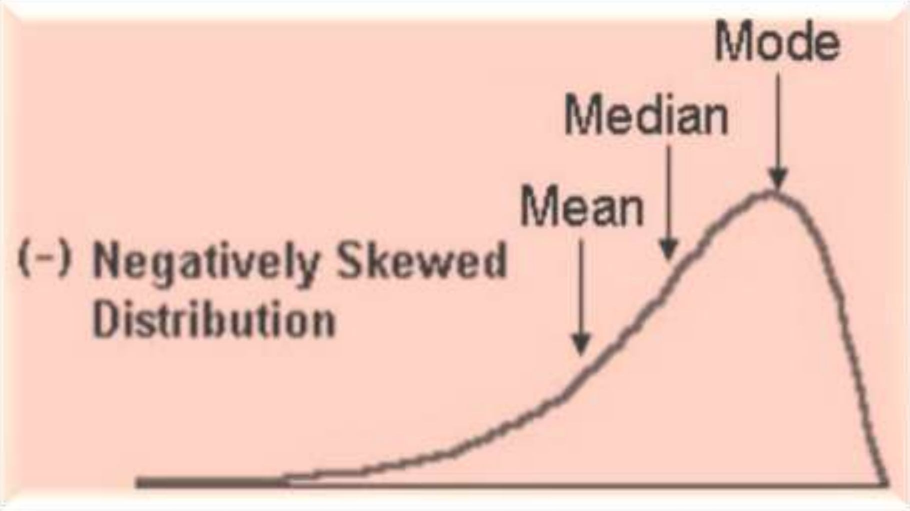


NPC in terms of skewness

Skewness refers to lack of symmetry.A normal curve is a perfect symmetrical curve.In many distributions which deviate from the normal,the value of mean,median and mode are different and there is no symmetry between the two halves of the curve.Such distributions are said to be skewed.

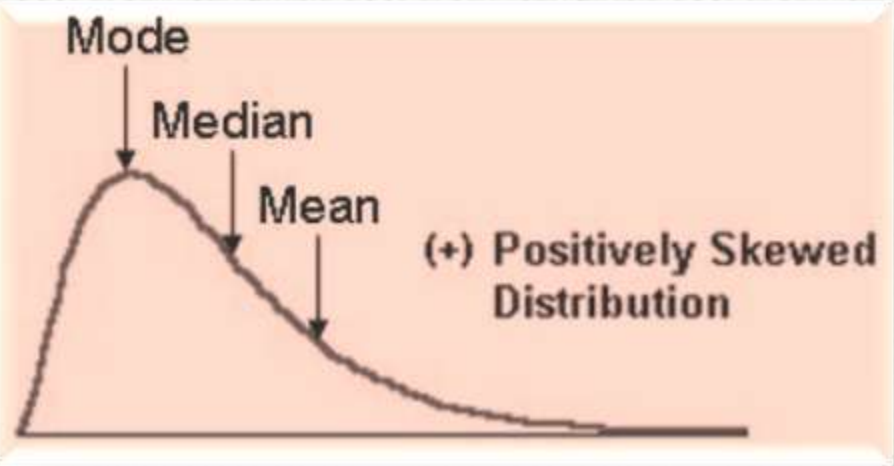
Negative skewness

In case of negative skewness,the curve is more inclined towards the left.



Positive skewness

In case of positive skewness,the curve is more inclined towards the right.



Formula for skewness

- $Sk = 3(M - md)/\sigma$
- (In terms of frequency distribution)

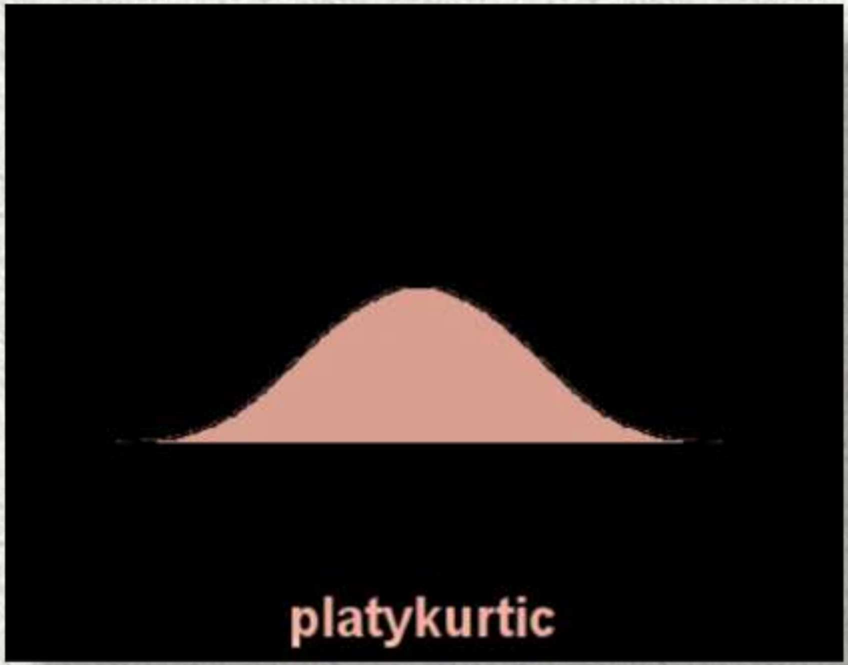
- $Sk = [P90+P10/2]-P50$
- (In terms of percentile)

NPC in terms of kurtosis

- When there are very few individuals whose scores are near to the average score for their group,the curve representing such a distribution becomes 'flattened' in the middle.On the other hand,when there are too many cases in the central area,the distribution curve becomes too 'peaked' in comparison with the normal curve.Both these characteristics of being flat or peaked ,are used to describe the term kurtosis.

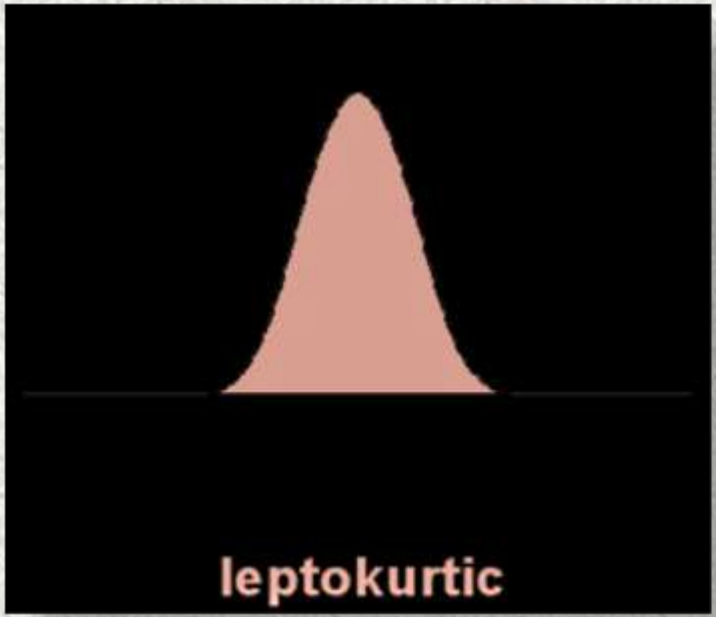
Platykurtic

- *A freuency distribution is said to be platykurtic when the curve is flatter than the normal curve.*



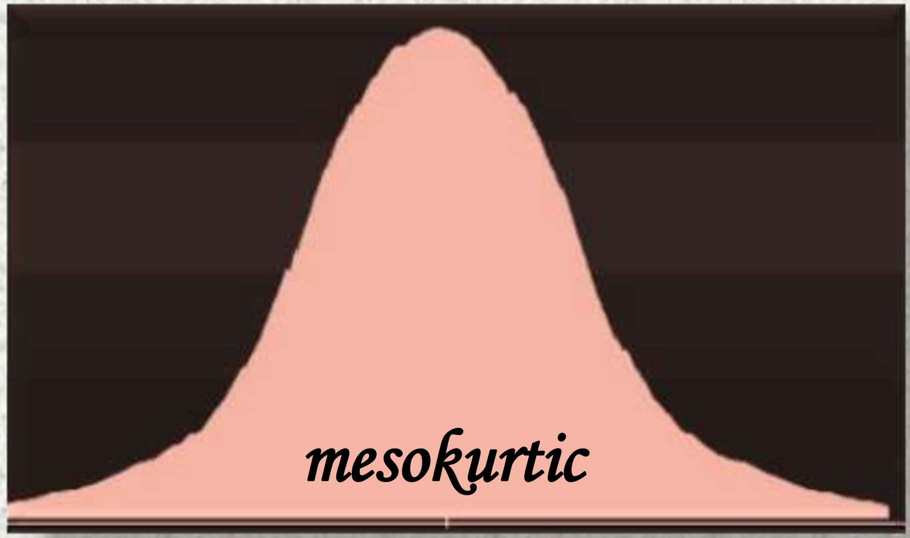
Leptokurtic

- *A frequency distribution is said to be leptokurtic when it is more peaked than the normal.*



Mesokurtic

- *A frequency distribution is said to be mesokurtic when it almost resembles the normal(neither too flat nor too peaked).*



Formula for kurtosis

- *Ku= Q/P90-P10*

Or;

- *Ku= [Q75-Q25/2]/P90-P10*

Applications of the normal curve

Use as a model -> Normal curve represents a model distribution.It can be used as a model to:

1)Compare various distributions with it,ie. To say,whether the distribution is normal or not and,if not,in what way it diverges from the normal.

2)Compare two or more distributions in terms of overlapping;and

3)Evaluate student's performance from their scores.

Computing percentiles and percentile ranks

- Normal probability curve may be conveniently used for computing percentiles and percentile ranks in a given normal distribution.

Ability grouping

- A group of individuals may be conveniently grouped into certain categories as A,B,C,D,E(very good,good,average,poor,very poor) in terms of some trait(assumed to be normally distributed),with the help of a normal curve.

Converting raw score into comparable standard normalized scores

With the help of a normal curve,we can convert the raw scores belonging to different tests into a standard normalized scores like sigma scores.For converting a given raw score into a z score,we subtract the mean of the scores of distribution from the respective raw scores and divide it by the standard deviation of the distribution.ie.

$$z=X-M/\sigma.$$

In this way a standard z score clearly indicates how many standard deviation units a raw score is above or below the mean and thus provides a standard scale for the purpose of valuable comparison.

Determining the relative difficulty of test items

- *Normal curve provides the simplest rational method of scaling test items for difficulty and therefore,may be conveniently employed for determining the relative difficulty of test questions,problems and other test items.*

Making use of the table of normal curve

- Table A of the normal curve provides the fractional parts of the total area (taken as 10,000) under the curve in relation to the respective sigma distances from the mean.This table may therefore be used to find the fractional part of the total area when z scores or sigma scores are given and also to find the sigma or z scores,when the fractional parts of the total area are given.

Statistical table-Table A

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4855	0.4956	0.4957	0.4959	0.496	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.497	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.498	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.499	0.499

We study some of the applications of the normal curve:

Application 1:

To determine the percentage of cases in a normal distribution within given limits.There can be three cases under this:

a) To find the percentage of cases below a given score point.

b) To find the percentage of cases above a given score point.

c)To find the percentage of cases lying between two given score points.

Illustration based on case (a)

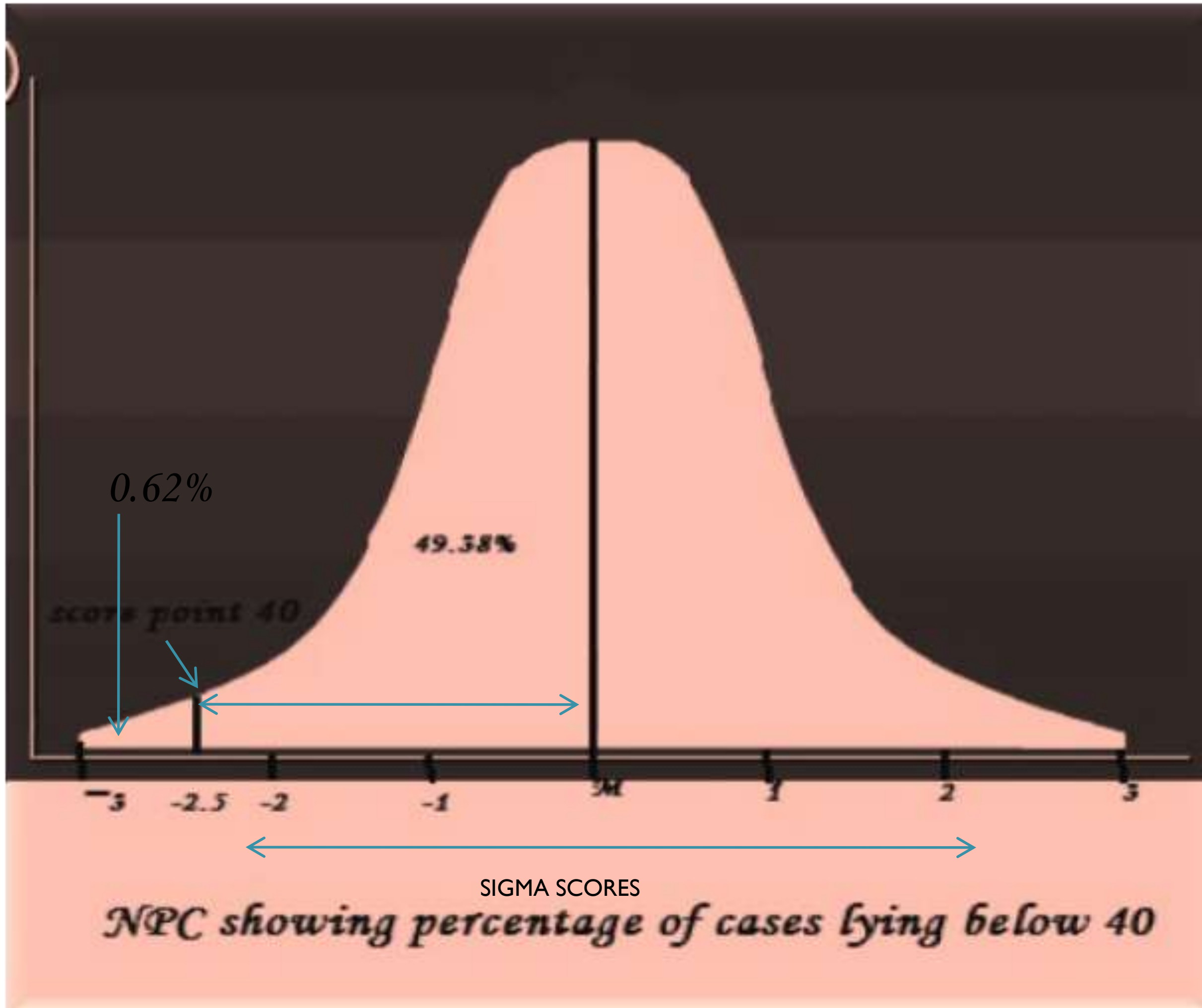
- Given a normal distribution,N=1,000;Mean=80 and SD=16.To find the percentage of individuals whose scores lie below the score point 40:
- The raw scores will be converted into z scores by using the formula-

  Z=X-M/σ ; z=40-80/16=> -2.5σ

  Referring to table A,the σ score -2.5 gives the value 4938.This value is further converted into percentage by dividing it with 10,000 and multiplying the same with 100.

  The value comes out to be 49.38%,which means 49.38% cases lie between mean and -2.5σ.

  In all,50-49.38=0.62% cases lie below the score point 40 and out of 1000,6 individuals achieve below the score point 40 which is also the **percentile rank.**



NPC showing percentage of cases lying below 40

Application 2:

To determine the limits of the scores between which a certain percentage of cases lie.

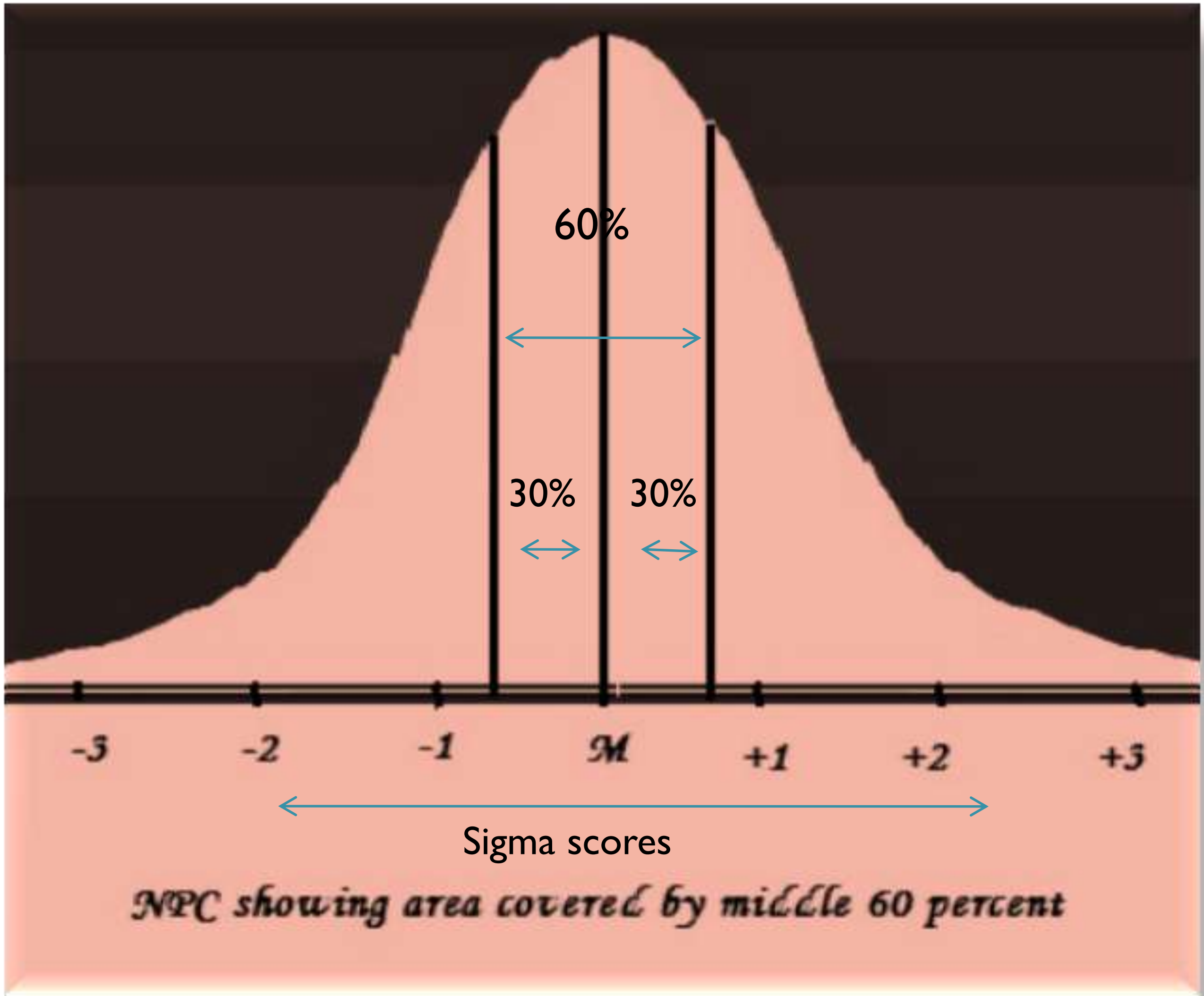
Illustration:If a distribution is normal with M=100,SD=20,find out the tow points between which the middle 60% of cases lie.

Solution:The middle 60% implies that 30% of the cases fall to the left and the rest 30% (3000 out of 10,000) to right of the mean.

Referring to table A,corresponding sigma distance for 3,000 fractional parts of total area would be calculated which comes out to be 0.84σ and -0.84σ for the cases falling to the left of mean.

The standard z scores would be converted to raw scores with help of same formula as discussed earlier[X-M/σ]

Value of raw score comes out to be 117 and 83(after rounding the figures) that include the middle 60% of cases.

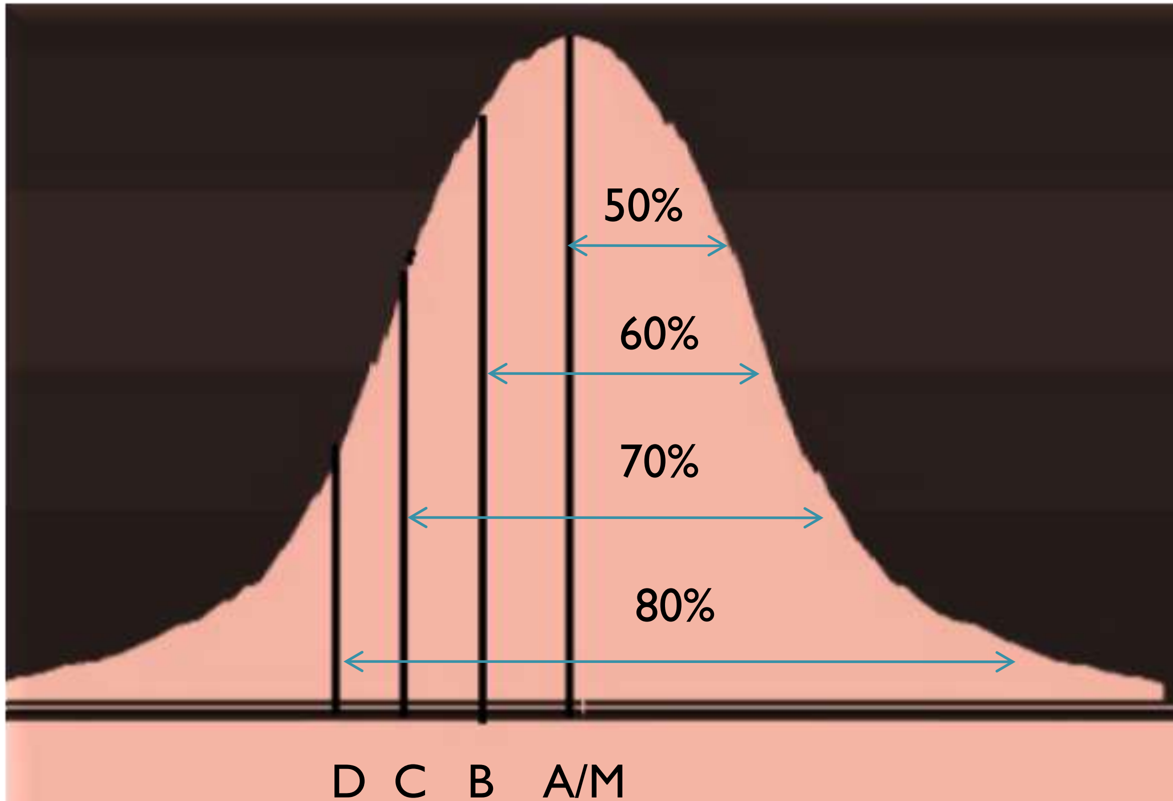

60%
30%
30%
-3
-2
-1
M
+1
+2
+3
Sigma scores
NPC showing area covered by middle 60 percent

Application 3

- To determine the relative difficulty value of the test items.

  Illustration:Four problems A,B,C and D have been solved by 50%,60%,70%,80% respectively of a large group.Compare the difficulty between A and B,with the difficulty between C and D.

Solution:The percentage of students who are able to solve the problem are counted from the extreme right.



NPC showing the problems solved by different percentages of cases

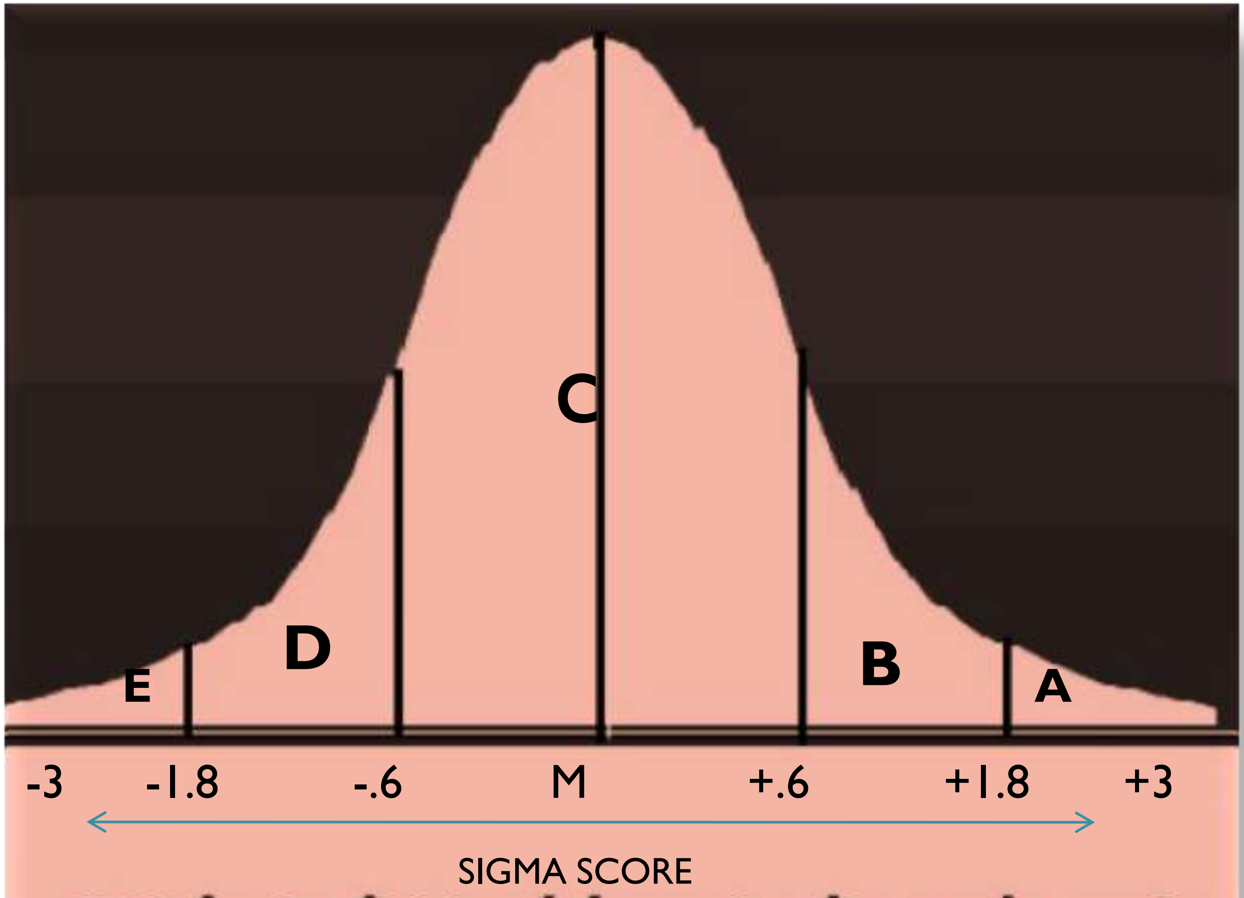
- *Problem A was solved by 50% of the group,also,it implies 50% of the group has not been able to solve it,means it was an average problem having 0 difficulty level.*
- *Problem B was solved by 60% of the group,also,it implies,in comparison to A,10% more individuals were able to solve it.*
- *Similarly,Problem C was passed by 70% of the group,also,it implies,20% more individuals were able to solve it than the average.*
- *Problem D was passed by 80% of the group,also,it implies,30% more individuals were able to solve it than the average.*

PROBLEM	SOLVED BY(%)	DIFFICULTY VALUE	RELATIVE DIFFICULTY
A	50	-0	
B	60	-0.253σ	-0.253σ
C	70	-0.252σ	
D	80	-0.840σ	-0.315σ

Problem B is simpler than problem A by having 0.253 sigma less difficulty value and problem D is simpler than problem C by having 0.315 sigma less difficulty value.

Application 4

- To divide a given group into categories according to an ability or trait assumed to be distributed normally.
- Illustration:There is a group of 200 students that has to be classified into 5 categories:A B C D and E according to ability,the range of ability being equal in each category.If trait under ability is normally distributed,tell how many students should be placed in each category?
- Solution:Base line of curve extend from -3 sigma to +3 sigma,ie,over a range of 6 sigma,may be divided into 5 equal parts.It gives 1.2 sigma to be alloted to each category.



NPC showing division of the area into five equal categories

- *Now we need to calculate the percentage of cases lying within each of these areas.*
- *Area under A extends from 1.8 sigma to 3 sigma , therefore,group A may said to be comprise 3.5% of the whole group.*
- *B will cover the cases lying between 0.6 sigma and 1.8 sigma,therefore group B may be said to comprise 23.84% of the entire group.*
- *Group C may be said to comprise 45.14% of the entire group.*
- *Similarly group D and E said to comprise 23.8 % and 3.5% respectively of the whole group.*

	A	B	C	D	E
Percentage of whole group in each category	3.5	23.8	45.0	23.8	3.5
No. of students in each category out of 200	7.0	47.6	90.0	47.6	7.0
No. of students in whole no.	7	48	90	48	7


THANKYOU..