NORMAL DISRIBUTION

INTRODUCTION

The normal probability distribution is one of the most commonly and frequently used continuous distribution in applied research. It plays a prominent role in both parametric and non-parametric inferential procedures for many estimators and test statistics. An important application of normal distribution is in approximating the binomial probabilities when the number of trials is large. Further, as the sample size increases, the sampling distribution of the mean approaches normality, regardless of the shape of the population distribution. Due to this fact mean of all the human traits like height, weight, blood pressure, cardiovascular endurance, strength, speed etc. follows a normal distribution provided a large sample is obtained in a random manner.

Normal Distribution:

Normal distribution or Gaussian distribution is the most widely used continuous distribution. A continuous random variable X is said to follow the normal distribution if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

NORMAL CURVE

Normal curve can be defined as follows. Normal curve is properly known as bell shaped curve because of its shape similar to that of a bell. This curve was discovered by Gauss in 1909 and hence it is also known as Gaussian curve. Various other names given to this curve are theoretical curve, error curve and ideal curve.

CHARACTERISTICS OF NORMAL CURVE

The Normal curve has the following properties

1. It is a bell shaped curve
2. It is symmetrical about its peaks and its tails extend indefinitely in both directions that is although the curve is closer and close to the horizontal axis, it never touches it.
3. Mean, Median and Mode are all equal. They are located at the abscissa of the highest ordinate. It is Uni-model, it has only one mode.
4. The quartiles are equidistant from the median. In other words the first and third Quartiles are at equal distance from the Median.
5. Near the mean value, the Normal curve is concave to the horizontal axis. Near ± 3, the normal curve is convex to the horizontal axis.
6. A linear combination of independent Normal varieties also a Normal variate.
7. The point of inflexion of the curve are given by $\mu \pm \sigma$.
8. If the total area of the normal curve is considered to be 100% then the following limits cover the corresponding are in the curve.

Limits	Area
$\mu - \sigma$ to $\mu + \sigma$	68.26%
$\mu - 2\sigma$ to $\mu + 2\sigma$	95.44%
$\mu - 3\sigma$ to $\mu + 3\sigma$	99.73%

Properties of Normal Curve

Properties of the normal curve could be used in solving the given problem only when it is established that the distribution of the scores in question is normal or nearly normal. There are two such properties on which this fact could be verified. These are skewness and kurtosis.

Skewness

Skewness means lack of symmetricity. If the curve is not sysmmetrical it is skewed. We study skewness to measure symmetricity of the curve for the given data. In a skewed curve, mean median and mode fall at different points. Skewness is measured by the following formula:

$$S_k = \frac{3(M - M_d)}{\sigma}$$

Where symbol have their usual meanings. The limits of skewness is from -3 to 3. For a symmetrical distribution S_k =0. Skewness is of two types: positive and negative. A curve is said to be positively skewed if its tail is extending towards positive direction and in that case value of S_k is greater than 0. Similarly, in a negatively skewed curve its tail is extending towards negative direction. In that case value of S_k is less than 0. The above three situations can be shown graphically in Figure. Further skweness is positive if mean is greater than mode or mean is greater than median and negative otherwise.

Kurtosis

Kurrtosis means convexity of a curve. Kurtosis enables us have an idea about peakedness or flatness of the curve. It is measured by an index given by

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

Where $\mu_4 = \frac{1}{N}\sum(x-\bar{x})^4$ and $\mu_2^2 = \frac{1}{N}\sum(x-\bar{x})^2$

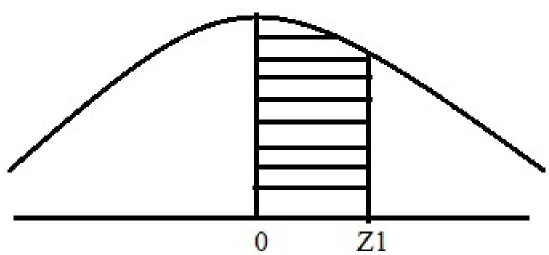
Curve of type A is neither peak nor flat and is called normal or measokurtic and for such a curve $\beta_2 = 3$. The type B curve is more peaked than the normal curve and is called leptokurtic curve, for such a curve β_2 is greater than the normal curve and type C which is flatter than the normal curve is known as platykurtic and for such a curve β_2 less than 3.

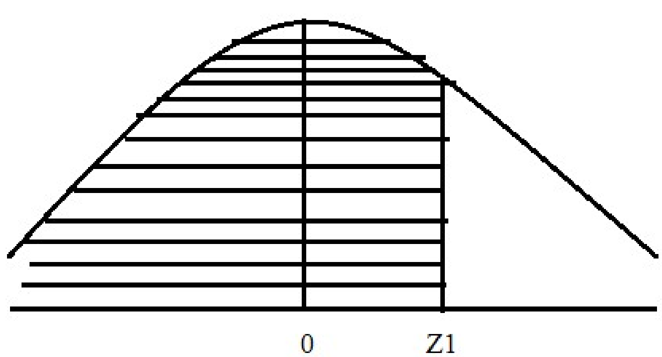
Standard normal Distribution:

The random variable Z = $\frac{X-\mu}{\sigma}$ follows the standard normal distribution. Z is called the standard normal variate. Its mean is 0 and the variance is 1.

Using Normal tables

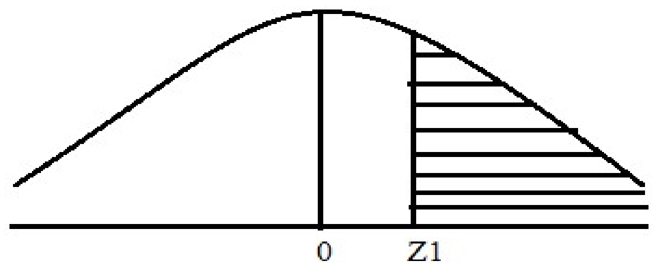
The normal tables give the area of the standard normal curve between Z = 0 and Z = Z_0 that is $P(0 < Z < Z_0)$





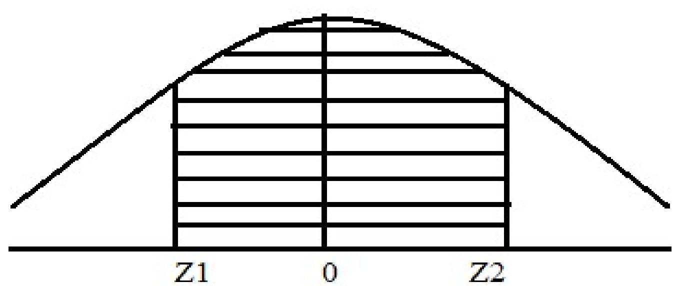
Shaded area = P $(Z < Z_1)$

$= 0.5 + P\ (0 < Z < Z_1)$



Shaded area = P $(Z > Z_1)$

$= 0.5 – P\ (0 < Z < Z_1)$



Shaded area = $P(Z_1 < Z < Z_2)$

$= P(0 < Z < Z_1) + P(0 < Z < Z_2)$

Problem Based on Normal Distribution

X is normally distributed with mean 12 and S.D 4. Find the probability that

(i) $X \geq 20$
(ii) $X \leq 20$,
(iii) $0 \leq X \leq 12$

Solution:

Given $\mu = 12$, $\sigma = 4$

$$Z = \frac{X-\mu}{\sigma} = \frac{X-12}{4}$$

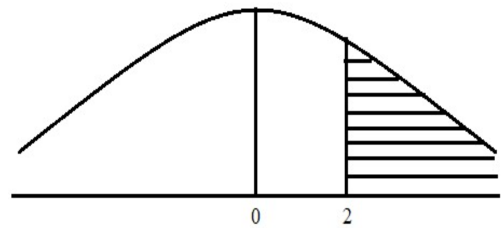
(i) $P(X \geq 20) = P\left(\frac{X-\mu}{\sigma} \geq \frac{20-12}{4}\right)$

$= P(Z \geq 2)$

$= 0.5 - P(0 \leq Z \leq 2)$

$= 0.5 - 0.4772$

$= 0.0228$



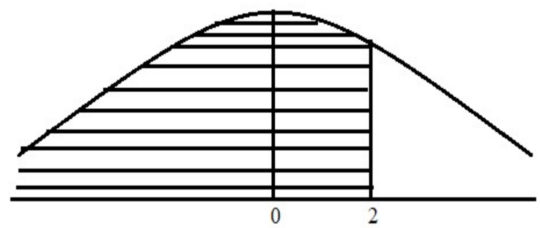
(ii) $P(X \leq 20) = P\left(\frac{X-\mu}{\sigma} \leq \frac{20-12}{4}\right)$

$= P(Z \geq 2)$

$= 0.5 - P(0 \leq Z \leq 2)$

$= 0.5 + 0.4772$

$= 0.9772$

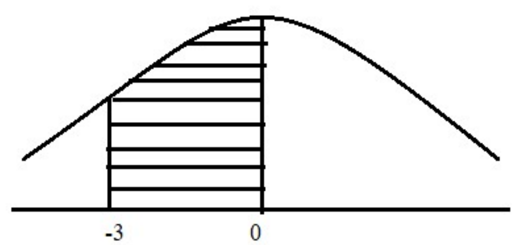


(iii) $P(0 \leq X \leq 12) = P\left(\frac{0-12}{4} \leq Z \leq \frac{12-12}{4}\right)$

$= P(-3 \leq Z \leq 0)$

$= P(3 \leq Z \leq 0)$

$= 0.4987$



Solved problem 1.4.9:

X is a normal variate with mean 30 and S.D = 5. Find (1) P(26 ≤ X ≤ 40), (2) P(X ≥ 45), (3) P($|X - 30| > 5$)

Solution:

Given μ = 30, $\sigma = 5$

$$Z = \frac{X-\mu}{\sigma} = \frac{X-30}{5}$$

(1) P (26 ≤ X ≤ 40)

$= P(\frac{26-30}{5} \leq X \leq \frac{40-30}{5})$

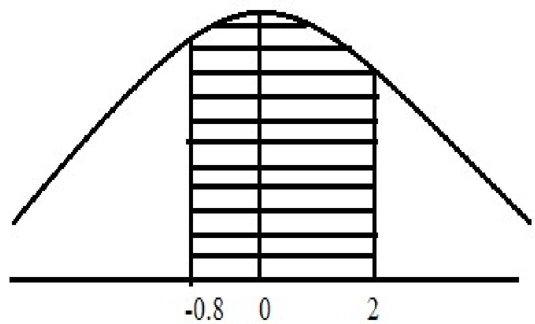
$= P(-0.8 \leq X \leq 2)$

$= P(-0.8 \leq X \leq 0) + P(0 \leq X \leq 2)$

$= P(0 \leq X \leq 0.8) + P(0 \leq X \leq 2)$

$= 0.2881 + 0.4772$

$= 0.7653$



(2) P(X ≥ 45)

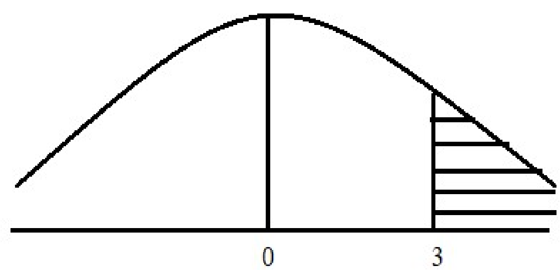
$= P (Z \geq \frac{45-}{5})$

$= P (Z \geq 3)$

$= 0.5 - P(0 \leq X \leq 3)$

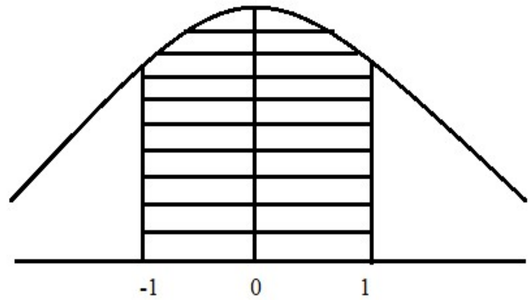
$= 0.5 - 0.4987$

$= 0.0013$



(3) P($\left| X - 30 \right| > 5$)

$= 1 - P(\left| X - 30 \right| < 5)$
$= 1 - P(-5 < X - 30 < 5)$
$= 1 - P(25 < X < 35)$
$= 1 - P(\frac{25-30}{5} \leq Z \leq \frac{35-3}{5})$
$= 1 - P(-1 < Z < 1)$
$= 1 - 2 \times P0 \; < Z < 1)$
$= 1 - 2 \times 0.3413$
$= 1 - 0.6826$
$= 0.3174$



Solved Problem

Assume that the mean height of soldiers is 68.22 inches with a variance of 10.8 inches. How many Soldiers in a regiment of thousand would you expect to be over 6 feet ball?

Solution:

$\mu = 68.22$, $\sigma^2 = 10.8$, $\sigma = 3.286$

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 68.22}{3.286}$$

P(X > 6 feet) = P (X > 72 inches)

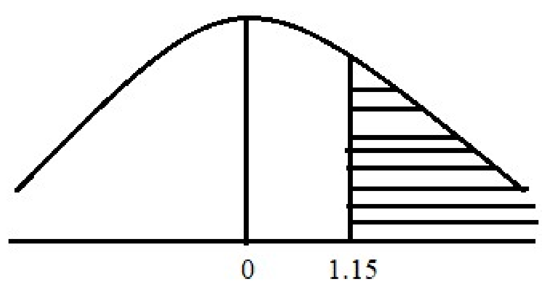
$= P(Z > \frac{72 - 68.22}{3.286})$

$= P(Z > 1.15)$

$= 0.5 - P(0 < Z < 1.15)$

$= 0.5 - 0.3749$

$= 0.1251$



Number of soldiers expected to be over 6 feet tall = 1000 × 0.1251 ≈ 125.

Solved Problem

In a test of 2000 electric bulbs it was found that the life of a particular make was normally distributed with an average life of 2040 hours and S.D of 60 hours. Estimate the number of bulbs likely to burn for

(1) More than 2150 hours

(2) Less than 1950 hours

(3) More than 1920 hours but less than 2160 hours

Solution:

$$\mu = 2040, \ \sigma = 60$$

$$Z = \frac{X-\mu}{\sigma} = \frac{X-204}{60}$$

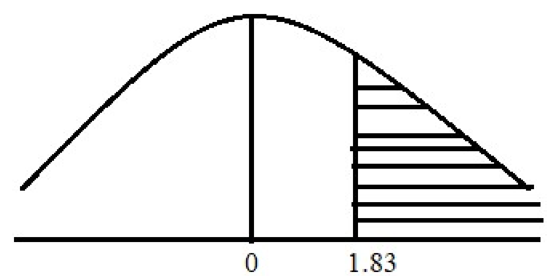
(1) $P(X > 2150) = P(\frac{X-2040}{60} > \frac{2150-2040}{60})$

$= P(Z > 1.83)$

$= 0.5 - P(0< Z < 1.83)$

$= 0.5 - 0.4664$

$= 0.0336$



Number of bulbs likely to burn for more than 2150 hours = 2000 × 0.0336 = 67.2≈ 67

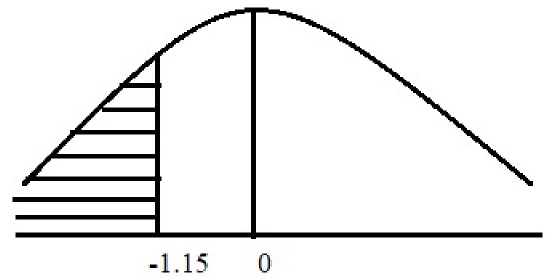
(2) $P(X < 1950) = P\left(Z < \frac{2150-2040}{60}\right)$

$= P(Z < -1.15)$

$= 0.5 - P(0 < Z < 1.15)$

$= 0.5 - 0.4332$

$= 0.0668$



Number of bulbs likely to burn for less than 1950 hours = $2000 \times 0.0668 = 133.6 \approx 134$

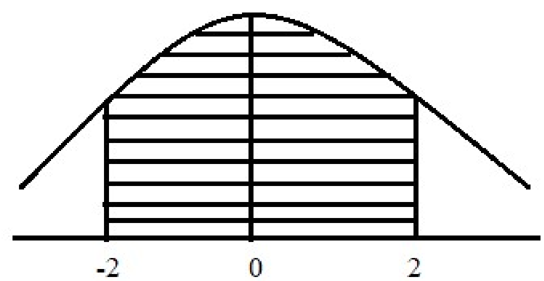
(3) $P(1920 < Z < 2160)$

$= P\left(\frac{1920}{60} < Z < \frac{2160-2040}{60}\right)$

$= P(-2 < Z < 2)$

$= 2 \times P(0 < Z < 2)$

$= 2 \times 0.4772 = 0.9544$



Number of bulbs likely to burn for more than 1920 hours but less than 2160 hours

$= 2000 \times 0.9544 = 1908.8 \approx 1909$

USES OF NORMAL DISTRIBUTION

The normal distribution is the most important probability distribution in statistics because it fits many natural phenomena. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution. It is also known as the Gaussian distribution and the bell curve.

****