Discrete Random Variable and Discrete Probability Distribution

Prepared by

Sudip Pokhrel

Random Variables

Definition

- A ***random variable*** is a variable whose value is determined by the outcome of a random experiment.
- It is denoted by capital letters. For example X, Y, Z etc.

Types of random Variables

There are two types of random variables. They are

1. Continuous Random Variable
2. Discrete Random Variable

Continuous random variable

- A random variable that can assume any value contained in one or more intervals is called a ***continuous random variable***.

- Example of a continuous random variable
  - Height or weight of a randomly chosen person
  - Amount of rainfall
  - Temperature set in a machine
  - Duration of telephone calls
  - The time taken to commute from home to work

Discrete Random Variable

- A *random variable* that assumes range of countable values is called a ***discrete random variable***.
- Example of discrete random variable
  - Number of students present in class
  - Number of customers visited the store
  - Number of defective items in a box
  - Number of telephone calls per unit time

Probability distribution of a discrete random variable

The ***probability distribution of a discrete random variable*** lists all the possible values that the random variable can assume and their corresponding probabilities.

The probability distribution of a discrete random variable possesses the following *two characteristics*.

1. $0 \leq P(x) \leq 1$ for each value of x
2. $\Sigma P(x) = 1$.

Terminology

More than 4: $P(X > 4)$

Less than 4: $P(X < 4)$

At least 4: $P(X \geq 4)$

At most 4: $P(X \leq 4)$

Between 4 and 6: $P(4 < X < 6)$

Between 4 and 6 (inclusive): $P(4 \leq X \leq 6)$

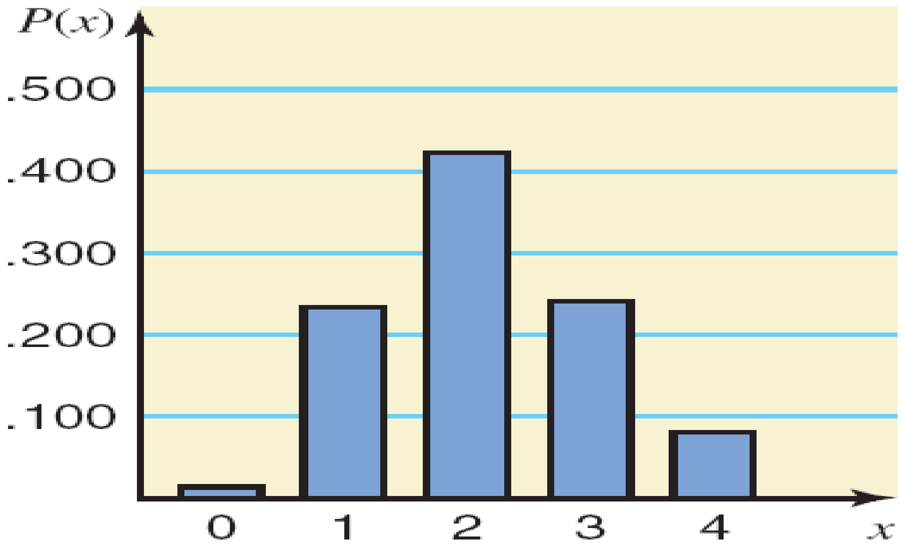
Frequency and Relative Frequency Distribution of the Number of Vehicles Owned by Families

Number of Vehicles Owned	Frequency	Relative Frequency
0	30	30/2000 = .015
1	470	470/2000 = .235
2	850	850/2000 = .425
3	490	490/2000 = .245
4	160	160/2000 = .080
	$N = 2000$	Sum = 1.000

Number of Vehicles Owned	Frequency	Relative Frequency
0	30	.015
1	470	.235
2	850	.425
3	490	.245
4	160	.080
	$N = 2000$	Sum = 1.000

Table 5.3 **Probability Distribution of the Number of Vehicles Owned by Families**

Number of Vehicles Owned x	Probability $P(x)$
0	.015
1	.235
2	.425
3	.245
4	.080
	$\Sigma P(x) = 1.000$



Example

The following table lists the probability distribution of the number of breakdowns per week for a machine based on past data

Breakdowns per week	0	1	2	3
Probability	.15	.20	.35	.30

Find the probability that the number of breakdowns for this machine during a given week is

a) exactly 2
b) 0 to 2
c) more than 1
d) at most 1

Example

Let X denote the number of breakdowns for this machine during a given week.

x	$P(x)$
0	.15
1	.20
2	.35
3	.30
	$\Sigma P(x) = 1.00$

Example

Using above table,

a. P(exactly 2 breakdowns) = $P(X = 2)$ = **.35**

b. P(0 to 2 breakdowns) = $P(0 \le X \le 2)$

$= P(X = 0) + P(X = 1) + P(X = 2)$

$= .15 + .20 + .35 =$ **.70**

c. P(more then 1 breakdown) = $P(X > 1)$

$= P(X = 2) + P(X = 3)$

$= .35 + .30 =$ **.65**

d. P(at most one breakdown) = $P(X \le 1)$

$= P(X = 0) + P(X = 1)$

$= .15 + .20 =$ **.35**

Mean of discrete random variable

The ***mean of a discrete variable*** x is the value that is expected to occur per repetition, on average, if an experiment is repeated a large number of times. It is denoted by μ, also known as the expected value, *E(X)*, and it is calculated as

$$\mu = E(X) = \Sigma x P(x)$$

Standard Deviation of discrete random variable

The ***standard deviation of a discrete random variable*** X measures the spread of its probability distribution and is computed as:

$$\sigma = \sqrt{\sum x^2 P(x) - \mu^2}$$

$$\sigma = \sqrt{E(X^2) - [E(X)]^2}$$

Example

x	0	1	2	3	4	5
$P(x)$	.02	.20	.30	.30	.10	.08

a) Find the mean number for the given probability distribution.

b) Find the standard deviation of X.

Example

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
0	.02	.00	0	.00
1	.20	.20	1	.20
2	.30	.60	4	1.20
3	.30	.90	9	2.70
4	.10	.40	16	1.60
5	.08	.40	25	2.00
		$\Sigma xP(x) = 2.50$		$\Sigma x^2P(x) = 7.70$

Mean (μ) = $\Sigma x\ P(x)$ = 2.50

$$\sigma = \sqrt{E(X^2) - [E(X)]^2}$$
$$= \sqrt{7.7 - 2.5^2}$$
$$= \sqrt{1.45} = \boxed{1.204}$$

Exercise 3

The following table records the probability distribution of a discrete random variable *X*.

x	**0**	**1**	**2**	**3**	**4**	**5**
$P(X = x)$	0.10	0.25	r	$2r$	0.15	0.05

(a) Determine the value of *r*.

(b) Determine: μ_X and σ_X.

(c) Determine: E(2x+1)

(d)Determine: V(2x+3)

Exercise 4

Let X represent the number of times a student visits a nearby pizza shop in one-month period. Assume that the following table is the probability distribution of x:

x	0	1	2	3
$P(X)$	0.1	0.3	0.4	0.2

a. Find the mean (μ) and the standard deviation (σ) of this distribution.
b. What is the probability that the student visits the pizza shop at least twice in a month?

Exercise 5

House sales per month	0	1	2	3
Probability	0.2	a	0.3	0.1

The number of houses sold in any given month by a local estate agent has the following probability distribution.

a) Find the value of a.
b) What is the probability of 3 houses sold in a given month?
c) What is the probability that one or two houses being sold in a given month?
d) Find mean and standard deviation of the distribution.
e) E(3x+9)
f) V(4x+9)

Exercise 7

After watching a number of children playing games at a video arcade, a statistics practitioner estimated the following probability distribution of x, the number of games played per visit.

x	1	2	3	4	5	6	7
$P(X)$	0.05	0.15	0.15	0.25	0.20	0.10	0.10

a. What is the probability that a child will play more than four games?
b. What is the probability that a child will play at least two games?

Discrete Probability Distribution

- Theoretical distributions are mathematical modals
- In a population, the values of the variables may be distributed according to some probability law which can be distributed mathematically and the corresponding probability distribution of random variable called theoretical probability distribution

Types

a) Discrete Theoretical Distribution- Binomial, Poisson, Negative Binomial Distribution, Hyper Geometric Distribution, Negative Binomial Distribution etc.

b) Continuous Theoretical Distribution: Normal, Student- t distribution, F-distribution, Gamma, Exponential, Chi square etc.

Binomial Probability Distribution

- Most widely used discrete distribution
- Having only two mutually exclusive outcomes i.e. success and failure
- Example: tossing a coin, Birth of a child, exam result etc.

Considerations:

- There are ***n*** identical trials
- Each trial has only two mutually exclusive outcomes, i.e. success and failure
- The trails are identical and independent
- The probability of success **(p)** remains constant for each trail

Example

Consider the experiment consisting of 10 tosses of a coin. Determine whether or not it is a binomial experiment.

1. There are a total of 10 trials (tosses), and they are all identical. Here, $n = 10$.

2. Each trial (toss) has only two possible outcomes: a head and a tail.

3. The probability of obtaining a head (a success) is ½ and that of a tail (a failure) is ½ for any toss. That is,

$$p = P(\mathrm{H}) = \tfrac{1}{2} \quad \text{and} \quad q = P(\mathrm{T}) = \tfrac{1}{2}$$

4. The trials (tosses) are independent.

Consequently, the experiment consisting of 10 tosses is a binomial experiment.

The Binomial Probability Distribution

For a Binomial experiment, the probability of exactly x successes in n trials is given by the following probability mass function (binomial formula)

$$P(X = x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

where

n = total number of trials

p = probability of success

$q = 1 - p$ = probability of failure

x = number of successes in n trials

$n - x$ = number of failures in n trials

Properties:

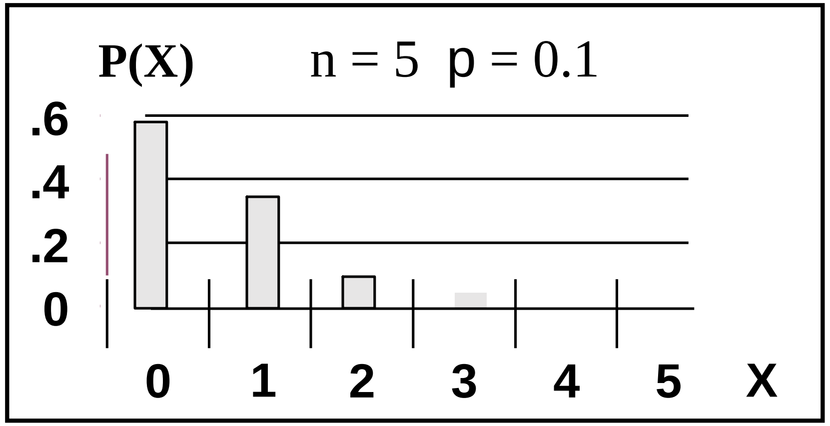
X~B(n, p) and it has following properties

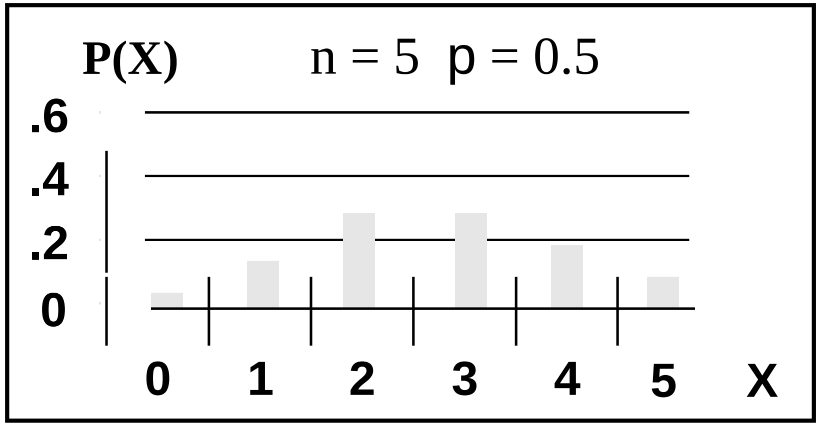
- Two parameters n and p
- Mean= np , variance= npq
- Mean> Variance (mean of the distribution is greater than variance)
- Distribution is symmetrical if p=q= 0.5 , if p>q left skewed, if p<q, Right skewed

The Binomial Distribution Shape

- The shape of the binomial distribution depends on the values of p and n

▪ Here, n = 5 and p = .1





▪ Here, n = 5 and p = .5

Example -1

At the Express House Delivery Service, providing high-quality service to customers is the top priority of the management. The company guarantees a refund of all charges if a package it is delivering does not arrive at its destination by the specified time. It is known from past data that despite all efforts, 2% of the packages mailed through this company do not arrive at their destinations within the specified time. Suppose a corporation mails 10 packages through Express House Delivery Service on a certain day.

(a) Find the probability that exactly one of these 10 packages will not arrive at its destination within the specified time.

(b) Find the probability that at most one of these 10 packages will not arrive at its destination within the specified time.

Example-1

n = total number of packages mailed = 10
p = P(success) = .02
q = P(failure) = 1 – .02 = .98

$$X \sim B(n = 10, p = 0.02)$$

$$\text{(a) } P(X = 1) = {}^{10}C_1(.02)^1(.98)^9 = 10 \times (.02)^1(.98)^9$$
$$= (10)(.02)(.83374776) = 0.1667$$

$$\text{(b) } P(X \le 1) = P(X = 0) + P(X = 1)$$
$$= {}^{10}C_0(.02)^0(.98)^{10} + {}^{10}C_1(.02)^1(.98)^9$$
$$= (1)(1)(.81707281) + (10)(.02)(.83374776)$$
$$= 0.8171 + 0.1667$$
$$= 0.9838$$

Example -2

In a Pew Research Centre nationwide telephone survey conducted in March through April 2011, 74% of college graduates said that college provided them intellectual growth (*Time*, May 30, 2011). Assume that this result holds true for the current population of college graduates. Let X denote the number in a random sample of three college graduates who hold this opinion. Write the probability distribution of X.

n = total college gradates in the sample = 3

p = *P*(a college graduate holds the said opinion) = .74

q = P(a college graduate does not hold the said opinion)

= 1 - .74 = .26

Example -2

$$P(X=0) = {}^{3}C_{0}(.74)^{0}(.26)^{3} = (1)(1)(.017576) = .0176$$

$$P(X=1) = {}^{3}C_{1}(.74)^{1}(.26)^{2} = (3)(.74)(.0676) = .1501$$

$$P(X=2) = {}^{3}C_{2}(.74)^{2}(.26)^{1} = (3)(.5476)(.26) = .4271$$

$$P(X=3) = {}^{3}C_{3}(.74)^{3}(.26)^{0} = (1)(.405224)(1) = .4052$$

x	$P(x)$
0	.0176
1	.1501
2	.4271
3	.4052

Poisson Probability Distribution

Conditions to Apply the Poisson Probability Distribution

The following three conditions must be satisfied to apply the Poisson probability distribution.

1. X is a discrete random variable.
2. The occurrences are random.
3. The occurrences are independent.

Examples:

1. The number of accidents that occur on a given highway during a 1-week period.
2. The number of customers entering a grocery store during a 1 –hour interval

Definition:

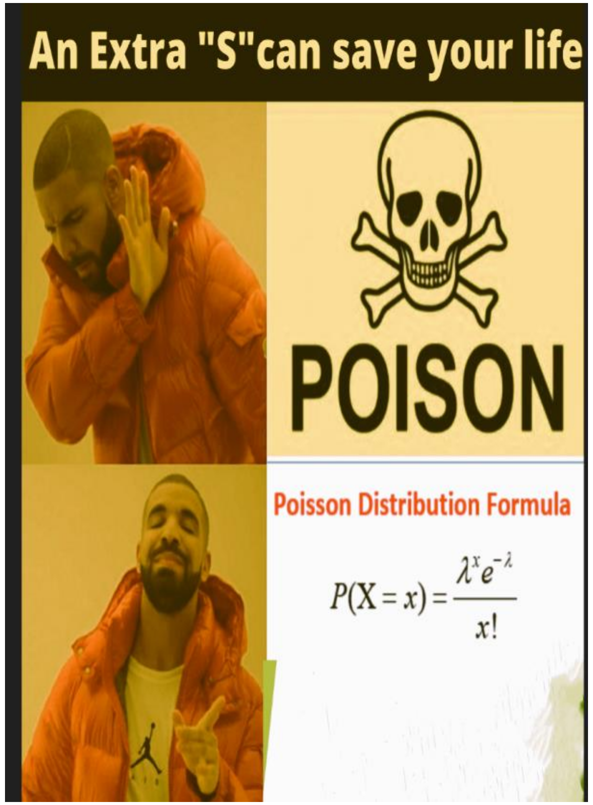
A discrete random variable X is said to have a Poisson distribution with parameters λ if

a) it assumes only non-negative values and

b) its probability mass function is given by

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \quad ; x=0,1,2,\ldots.,\infty$$

where λ (pronounced *lambda*) is the mean number of occurrences in that interval and the value of e is approximately 2.71828.



Properties:

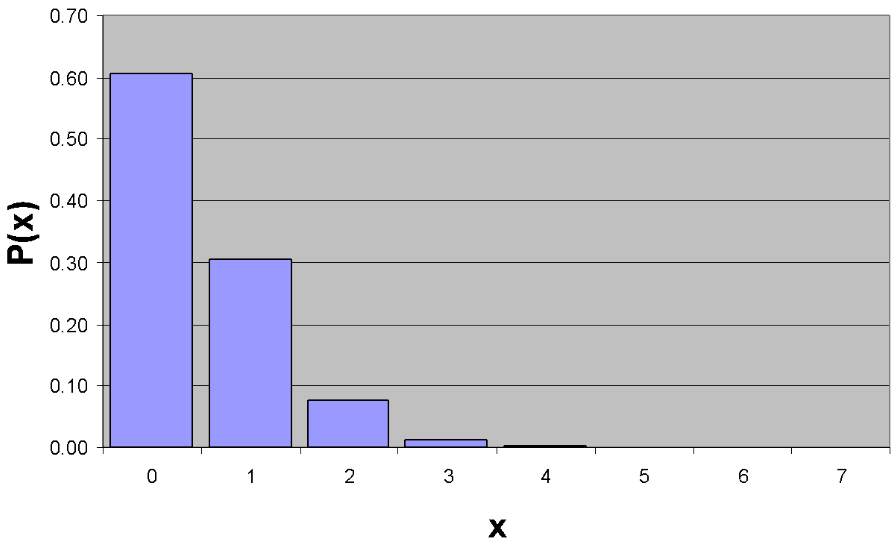
- It is uniparametric distribution.
- The average number of occurrences in a given interval of time or space should be known in advance.
- The mean and variance of the Poisson distribution is equal to λ.

  $E(x)=\lambda$ and $V(x)=\lambda$
- It is a positively skewed distribution.
- It tends to have normal distribution when λ is very large.

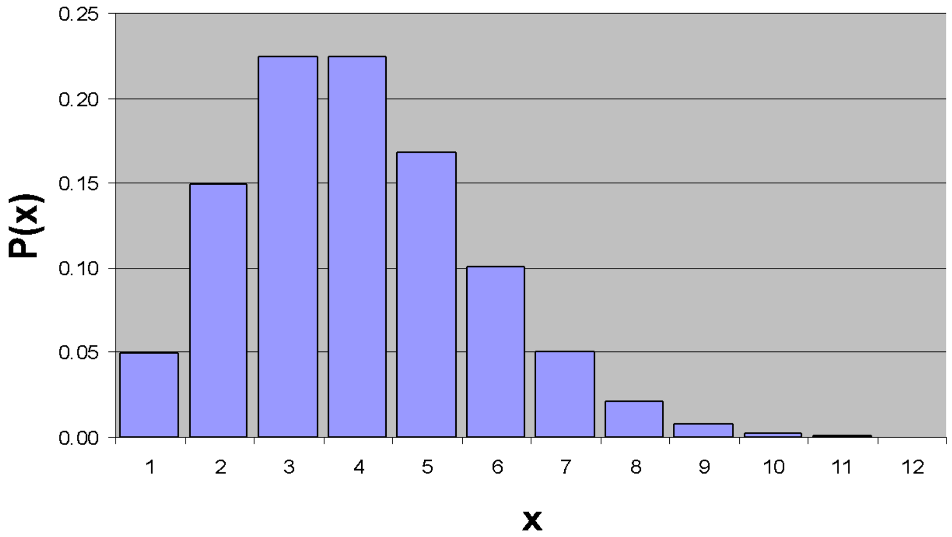
Poisson Distribution Shape

- The shape of the Poisson Distribution depends on the parameter λ :

$\lambda = 0.50$



$\lambda = 3.00$



Example: On average, two new accounts are opened per day at an Imperial Savings Bank branch. Assume that the events of account opening follow a Poisson distribution, find the probability that:

(a) exactly 6 accounts will be opened during a one-day period.

(b) at most 3 accounts will be opened during a one-day period.

(c) less than 3 accounts will be opened during a two-day period.

(d) between 4 and 6 (inclusive) accounts will be opened during a three-day period.

(e) Find the mean and standard deviation of the number of accounts opened during a five-day working week.

Let X be the number of new accounts.

(a) $X \sim \text{Po}(\lambda=2)$

$$P(X=6)=\frac{e^{-2}\cdot 2^{6}}{6!}=\boxed{0.01203}$$

(b) $P(X \leq 3)$

$$= P(X=0)+P(X=1)+P(X=2)+P(X=3)$$

$$= \frac{e^{-2}\cdot 2^{0}}{0!}+\frac{e^{-2}\cdot 2^{1}}{1!}+\frac{e^{-2}\cdot 2^{2}}{2!}+\frac{e^{-2}\cdot 2^{3}}{3!}$$

$$= e^{-2}\left(\frac{2^{0}}{0!}+\frac{2^{1}}{1!}+\frac{2^{2}}{2!}+\frac{2^{3}}{3!}\right)$$

$$= \boxed{0.8571}$$

Let X be the number of new accounts.

(c) $X \sim \mathrm{Po}(\lambda = 2 \times 2 = 4)$

$$P(X < 3)$$

$$= P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!}$$

$$= e^{-4} \left(\frac{4^0}{0!} + \frac{4^1}{1!} + \frac{4^2}{2!} \right)$$

$$= \boxed{0.2381}$$

Let X be the number of new accounts.

(d) $X \sim \text{Po}(\lambda = 3 \times 2 = 6)$

$$P(4 \le X \le 6)$$

$$= P(X = 4) + P(X = 5) + P(X = 6)$$

$$= \frac{e^{-6} \cdot 6^4}{4!} + \frac{e^{-6} \cdot 6^5}{5!} + \frac{e^{-6} \cdot 6^6}{6!}$$

$$= e^{-6}\left(\frac{6^4}{4!} + \frac{6^5}{5!} + \frac{6^6}{6!}\right)$$

$$= \boxed{0.4551}$$

(e) Let X be the number of new accounts.

$X \sim \text{Po}(\lambda = 5 \times 2 = 10)$

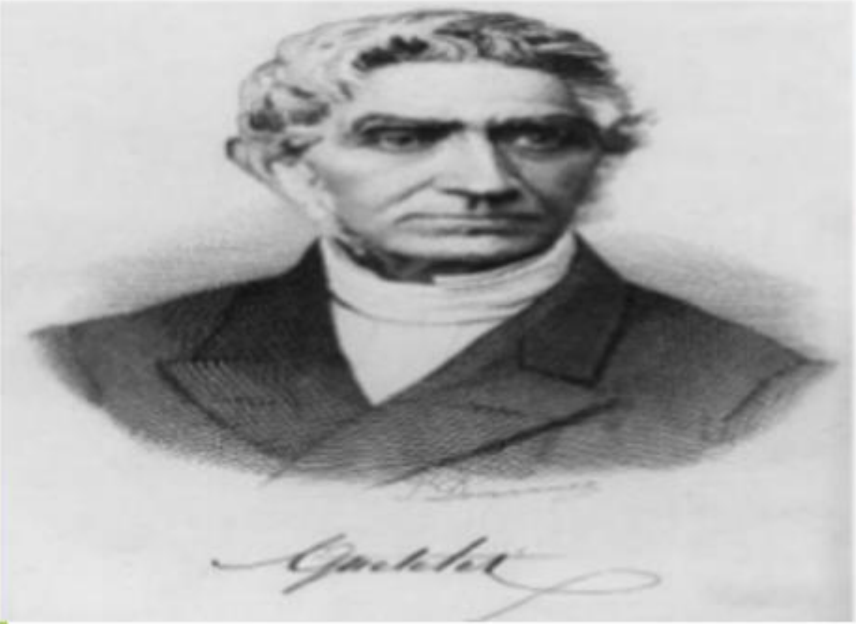
Mean, $\mu = \lambda = \boxed{10}$

Standard deviation, $\sigma = \sqrt{\lambda} = \sqrt{10} = \boxed{3.162}$

Normal Distribution

Continuous Probability Distribution

* **Abraham de Moivre** discovered the normal distribution in 1733
* French



* **Quetelet** noticed this in heights of army people.
* Belgian

* Gaussian distribution, after **Carl Friedrich Gauss.**
* German

* **Marquis de Laplace** proved the central limit theorem in 1810 , French
* For large sample size the sampling distribution of the mean follows normal distribution
* If sample studied is large enough normal distribution can be assumed for all practical purposes

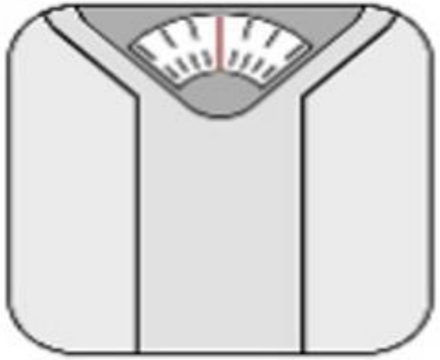
Normal Distribution

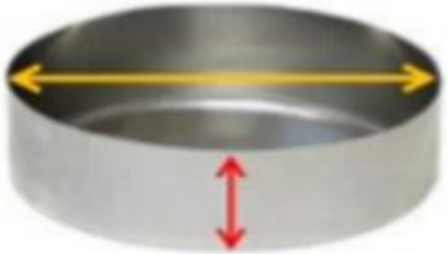
- Can be found practically everywhere:
  - In nature.
  - In engineering and industrial processes.
  - In social and human sciences.
- Many everyday data sets follow approximately the normal distribution.

- Normal Distribution

Examples:

- The body temperature for healthy humans.
- The heights and weights of adults.
- The thickness and dimensions of a product.
- IQ and standardized test scores.
- Quality control test results.
- Errors in measurements.





Importance of Normal Distribution

- Normal distributions are most important probability distributions in statistical field. Normal distributions can be used to describe many real-life situations and widely used in the field of engineering, science, business, psychology.
- The most incredible application of normal distribution lies in the Central Limit Theorem. This theorem states that no matter what type of distribution the population may have, as long as the sample size at least 30,the distribution of sample means will be normal. Hence the majority of problems and studies can be analyzed through normal distribution.

Importance of Normal Distribution

- It can be used to approximate other discrete distributions, such as Binomial distributions, Poisson distributions.
- For large value of n almost all the sampling distribution such as Student's t-distribution, Z-distribution, F-distribution and Chi-square distribution conform to normal distribution.
- It is used in statistical quality control for setting quality standards and to define quality limit.

Definition

- A Continuous random variable X is said to follow Normal distribution with Parameter mean μ and variance σ^2 if its probability density function p.d.f. is given by

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

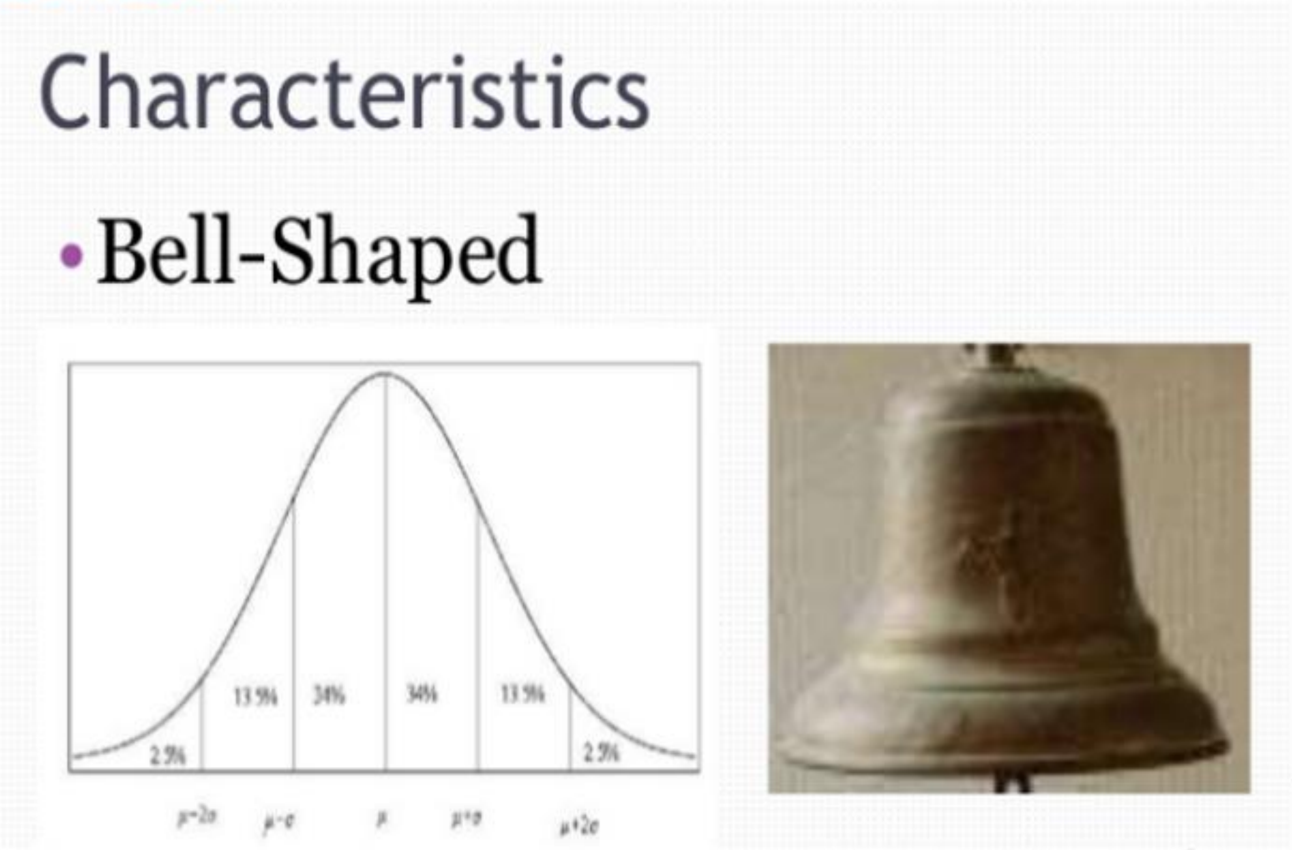
<u>Note constants:</u>
π=3.14159
e=2.71828

This is a bell shaped curve with different centers and spreads depending on μ and σ

- A graphical representation of normal distribution is called Normal curve.

Properties of Normal distribution

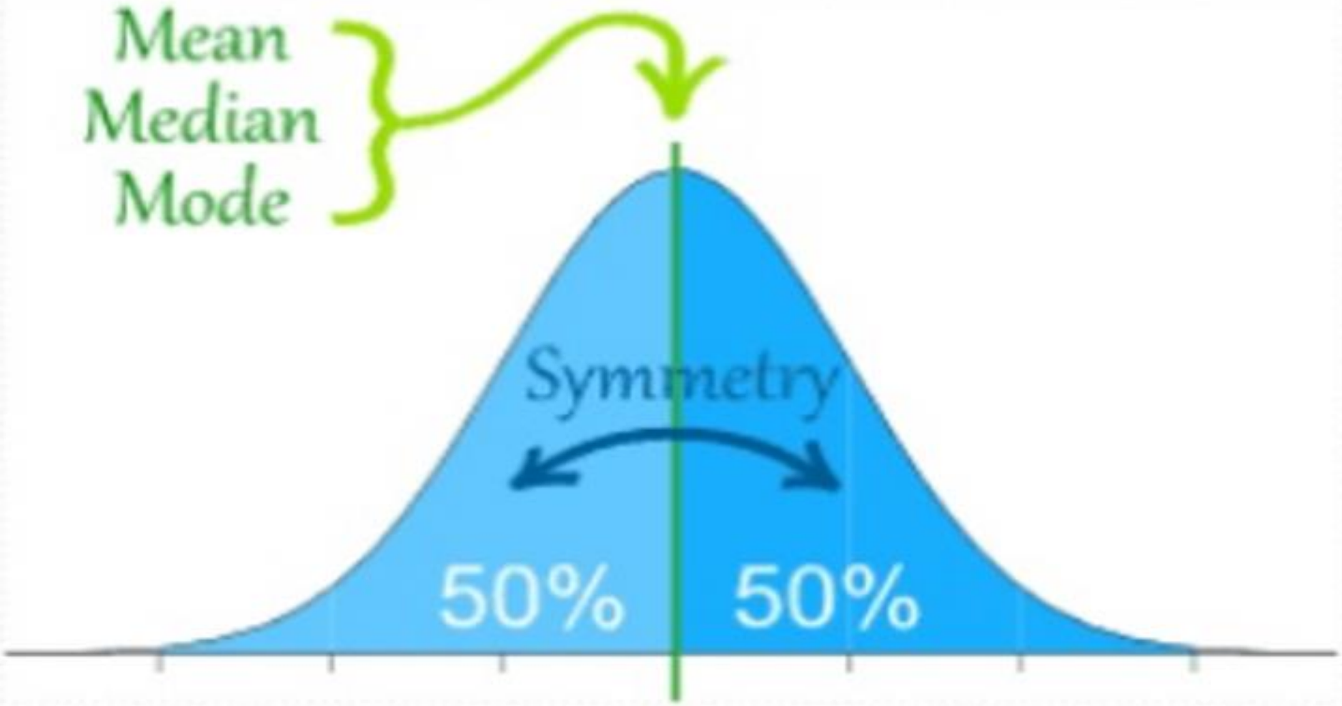
1. The curve of normal distribution is bell-shaped and symmetrical.



Properties of Normal distribution

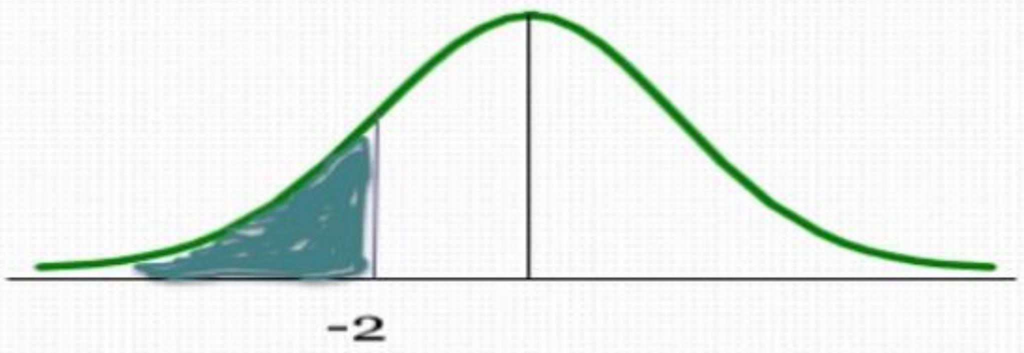
2. The mean, median and mode of Normal distribution lie at same point at center.

- Mean=Median=Mode



➢ Properties of Normal distribution

3.Since it is continuous distribution, area gives probabilities.



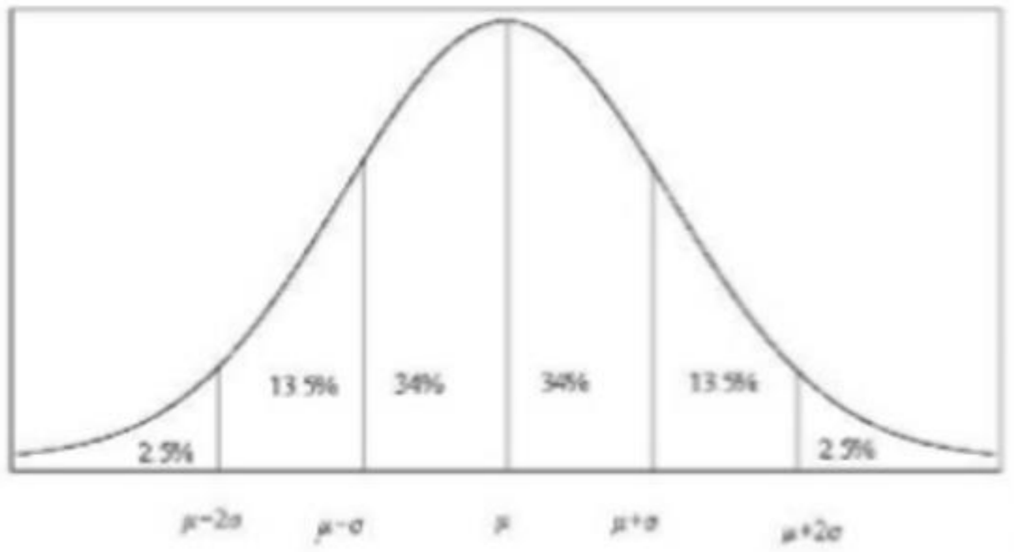
- $P\,[x<70] = P[z<-2] = 0.5 - P[0<z<2]$

  $= 0.5 - 0.4772$

  $= 0.0228$ or 2.28%

Properties of Normal distribution

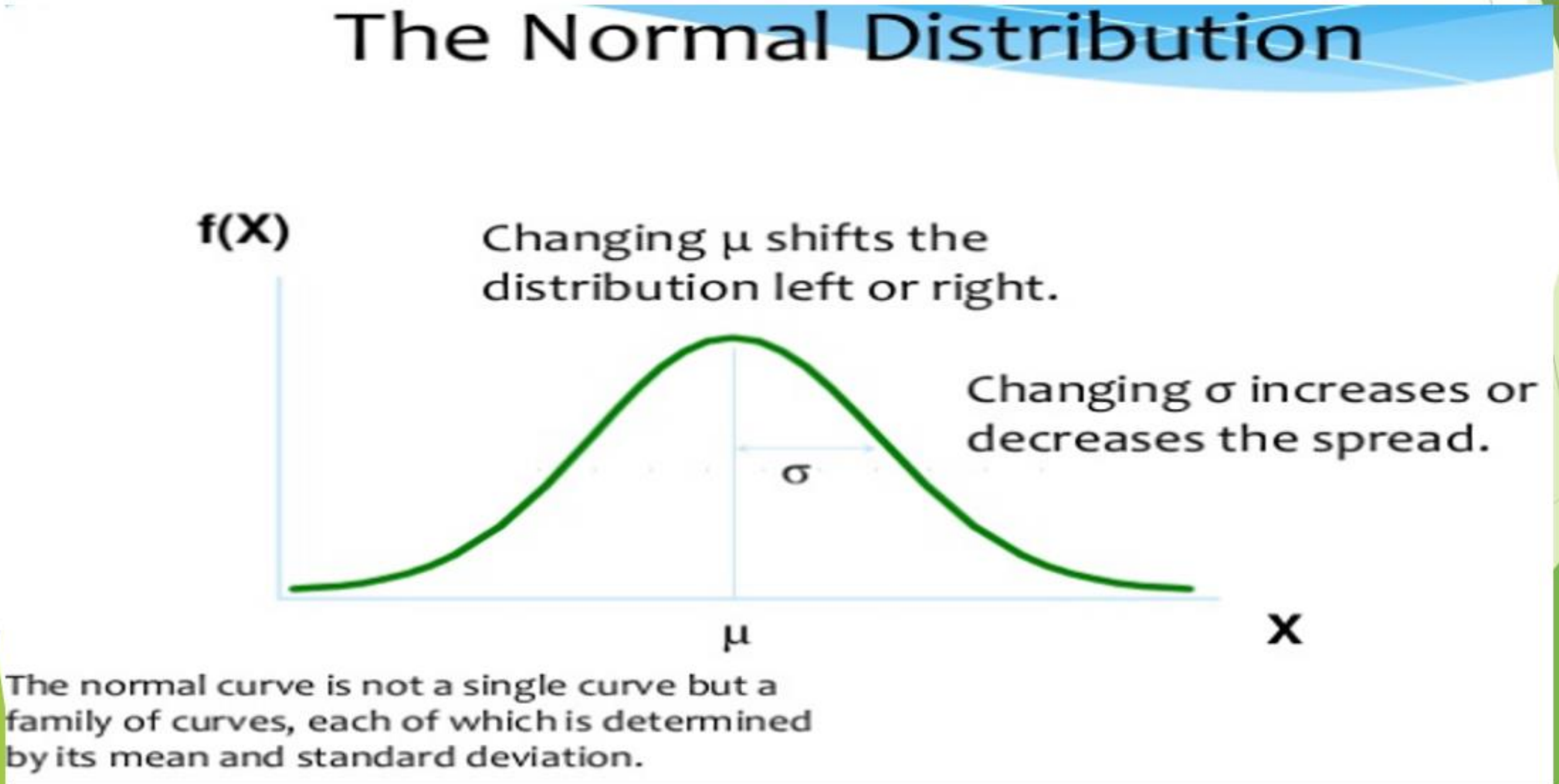
- Curve is asymptotic to the x-axis

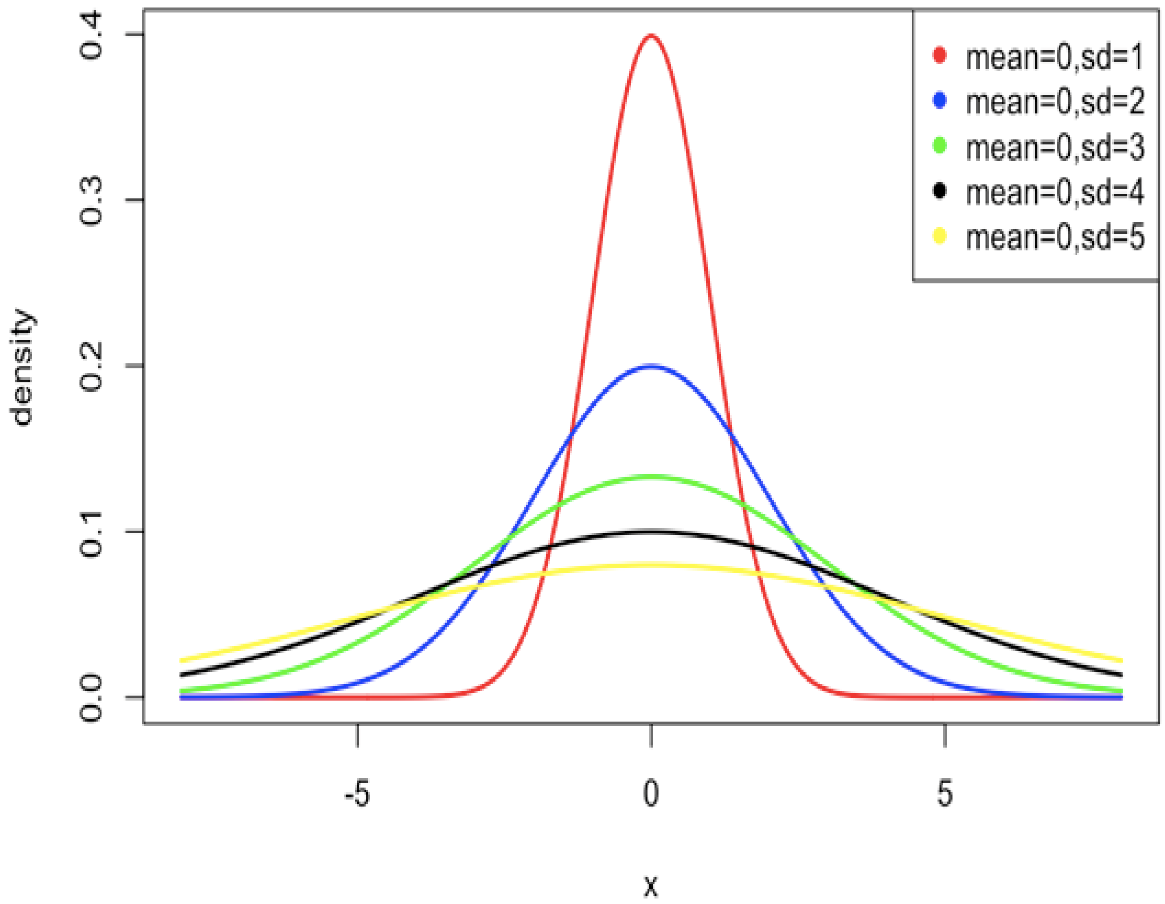


- Total area under the curve above the x-axis = 1 or 100%

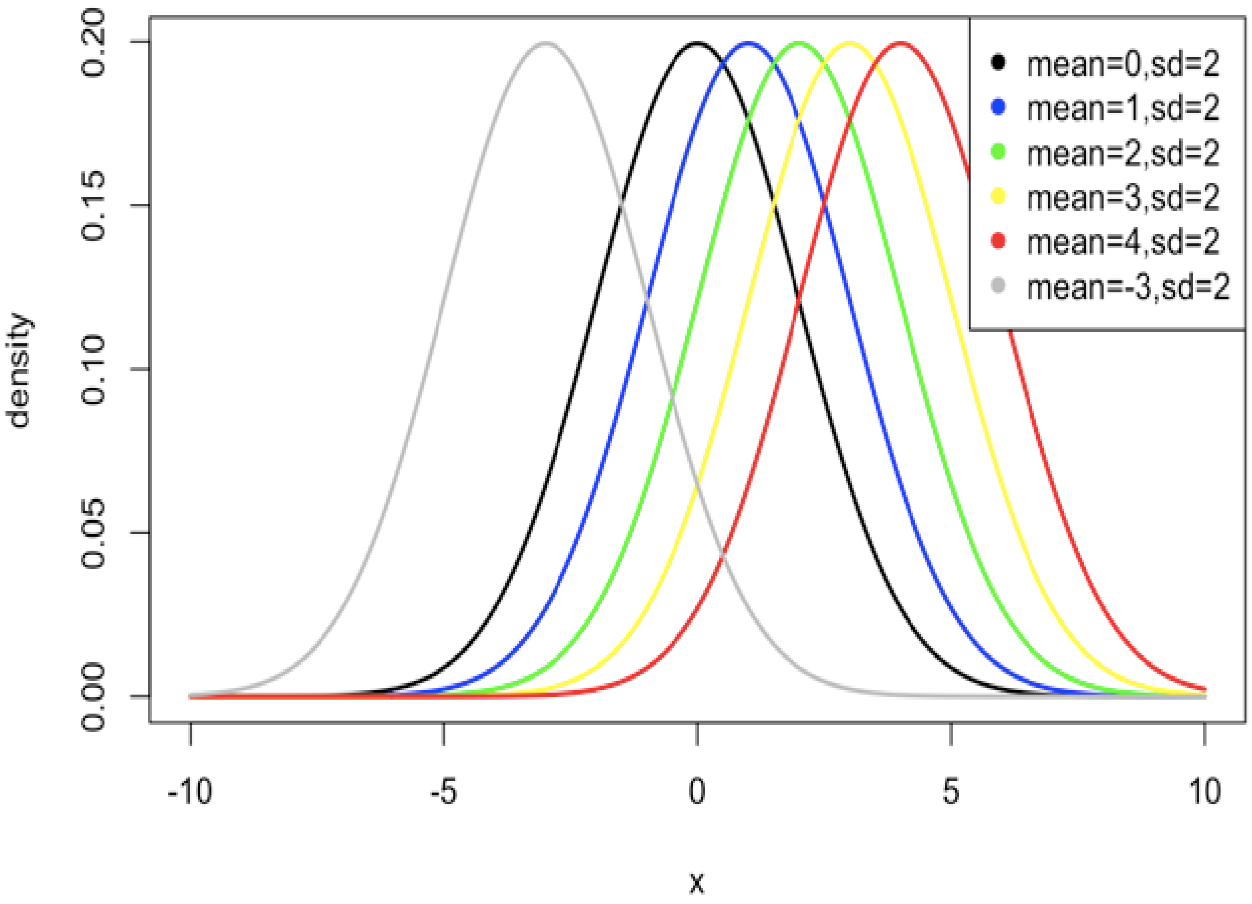
Properties of Normal distribution

6. The width of normal curve is controlled by standard deviation of data.



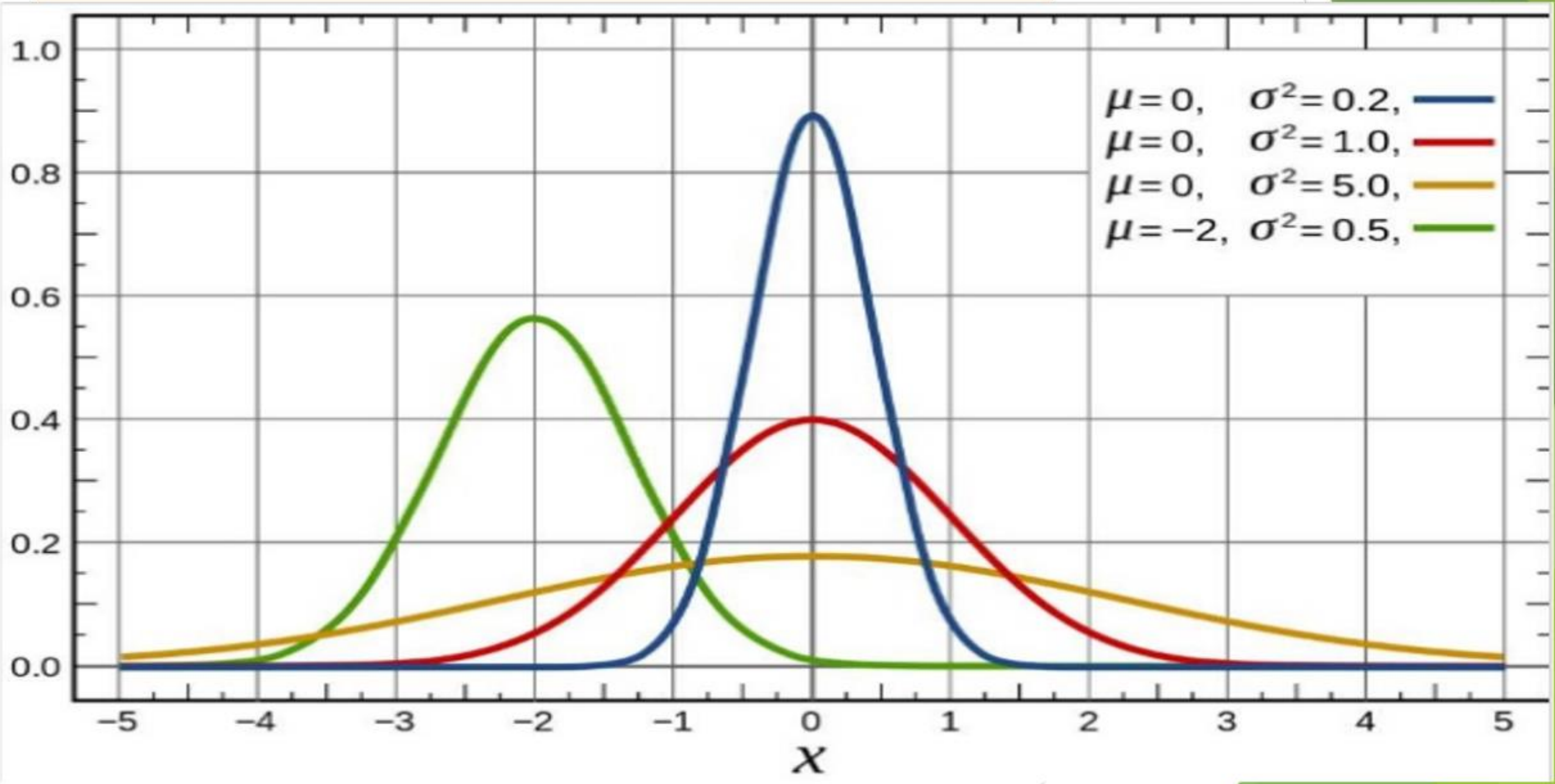


Mean equal
SD Unequal



Mean Unequal
SD Equal

Properties of Normal distribution



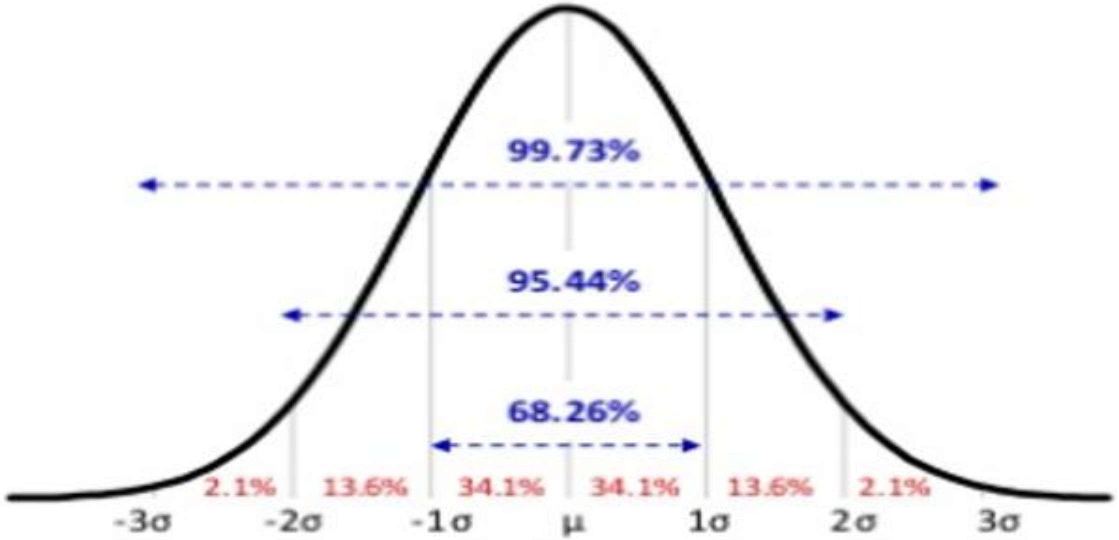
Properties of Normal distribution

7. Area Property:

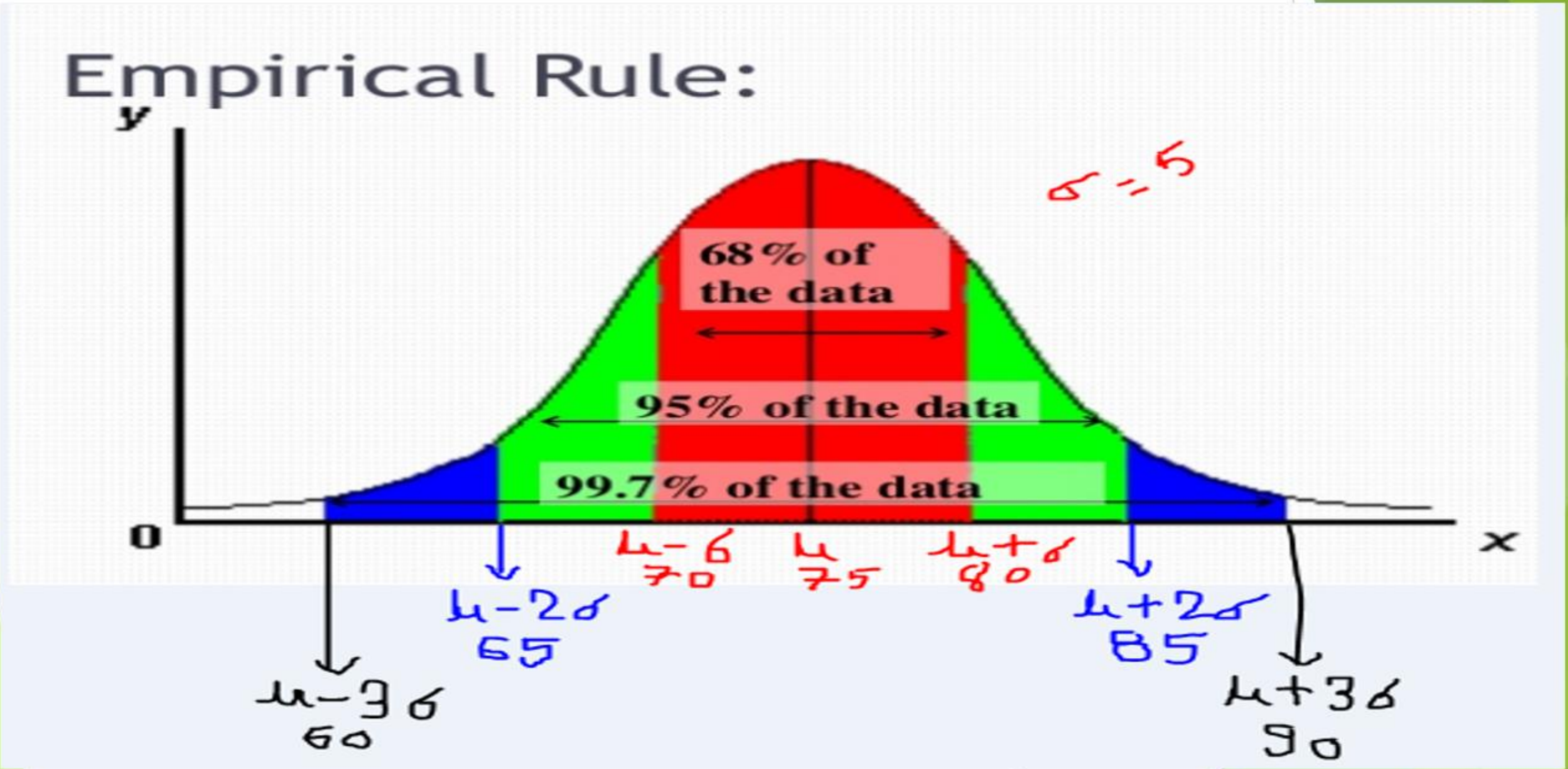
- Normal Distribution

Empirical Rule:

- For any normally distributed data:
  - **68%** of the data fall within **1** standard deviation of the mean.
  - **95%** of the data fall within **2** standard deviations of the mean.
  - **99.7%** of the data fall within **3** standard deviations of the mean.



Properties of Normal distribution



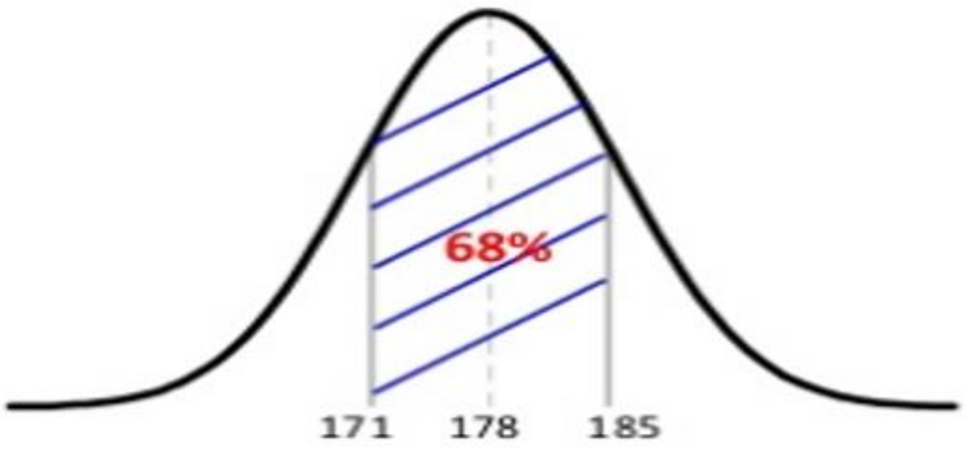
Area Property:

- Normal Distribution

Empirical Rule:

- Suppose that the heights of a sample men are normally distributed.
- The mean height is **178** cm and a standard deviation is **7** cm.

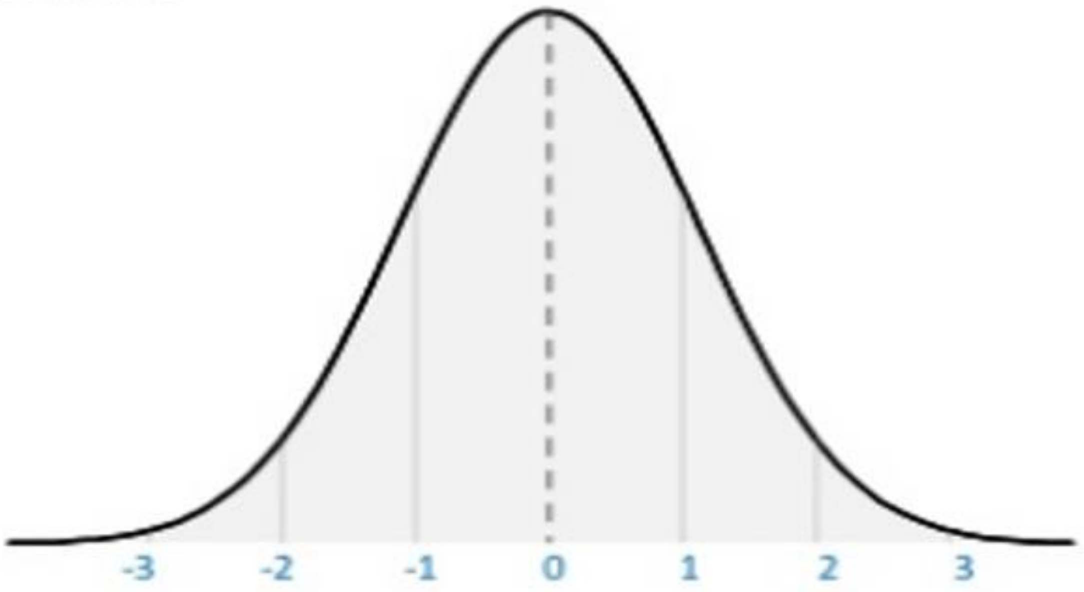
- **We can generalize that:**
  - 68% of population are between **171** cm and **185** cm.
  - This might be a generalization, but it's true if the data is normally distributed.



- Normal Distribution

Standard Normal Distribution:

- A common practice to convert any normal distribution to the standardized form and then use the standard normal table to find probabilities.
- The **Standard Normal Distribution** (Z distribution) is a way of standardizing the normal distribution.
- It always has a mean of **0** and a standard deviation of **1**.



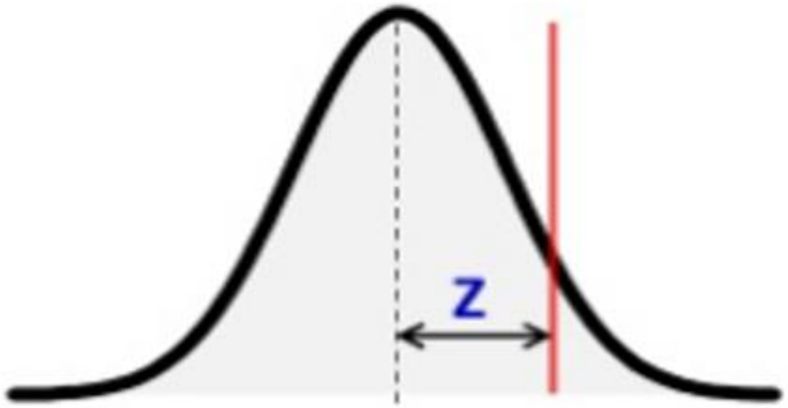
- Normal Distribution

Standard Normal Distribution:

- Any normally distributed data can be converted to the standardized form using the formula:

$$Z = (X - \mu) / \sigma$$

- where:
  - 'X' is the data point in question.
  - 'Z' (or **Z-score**) is a measure of the number of standard deviations of that data point from the mean.



So to convert a value to a Standard Score ("z-score"):

- first subtract the mean,
- then divide by the Standard Deviation

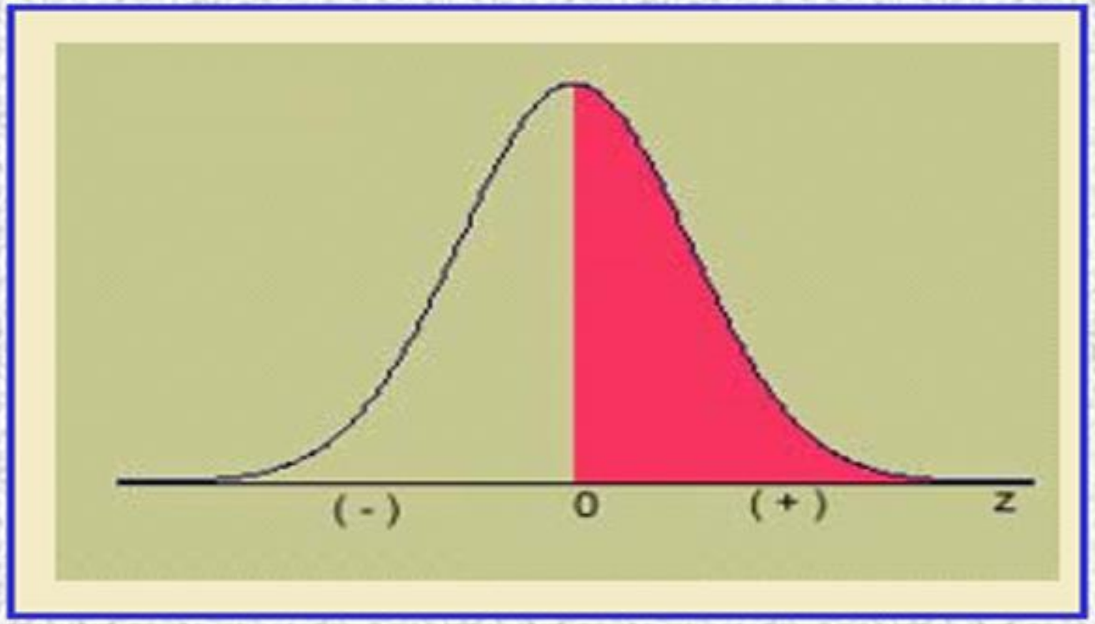
And doing that is called "Standardizing":

We can take any Normal Distribution and convert it to The Standard Normal Distribution.

The Standard Normal Distribution (Z)

- The mean (μ) = 0
- Standard deviation (σ) =1

$$X \sim N(\mu, \sigma) \quad \Rightarrow \quad Z = \frac{x-\mu}{\sigma} \sim N(0,1)$$



The Standard Normal (z) Distribution

- Mean = 0; Standard deviation = 1
- When $x = \mu$, $z = 0$
- Symmetric about $z = 0$
- Values of z to the left of center are negative
- Values of z to the right of center are positive
- Total area under the curve is 1.

Standard Normal Distribution

<u>Definition:</u>

If a continuous random variable X follows normal distribution with mean μ and variance σ^2 then the variable $Z=\frac{X-\mu}{\sigma}$ follows standard normal distribution with mean zero and variance one if its probability density function is given by

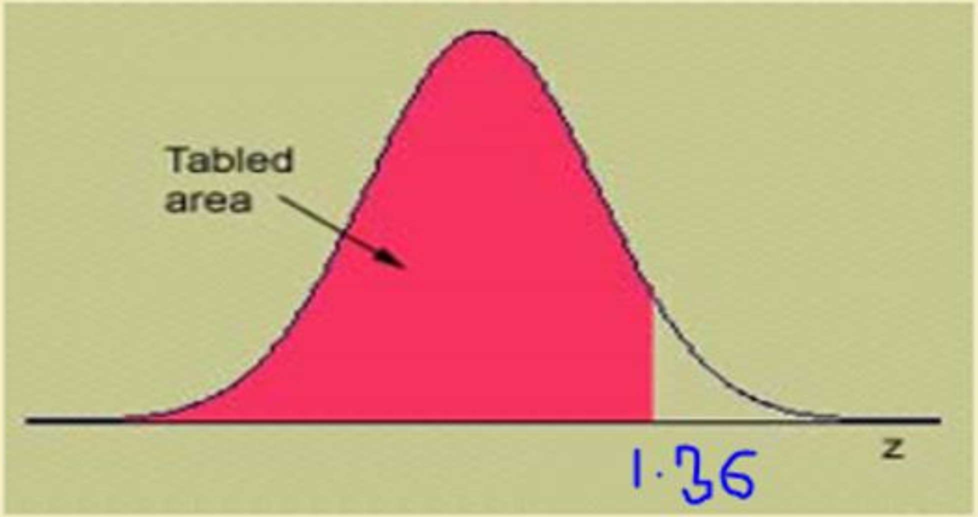
$$f(z)=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1z^2}{2}}$$

we say, $Z \sim N(0,1)$

Using Table 3

The four digit probability in a particular row and column of Table 3 gives the area under the z curve to the left that particular value of z.

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8328	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278



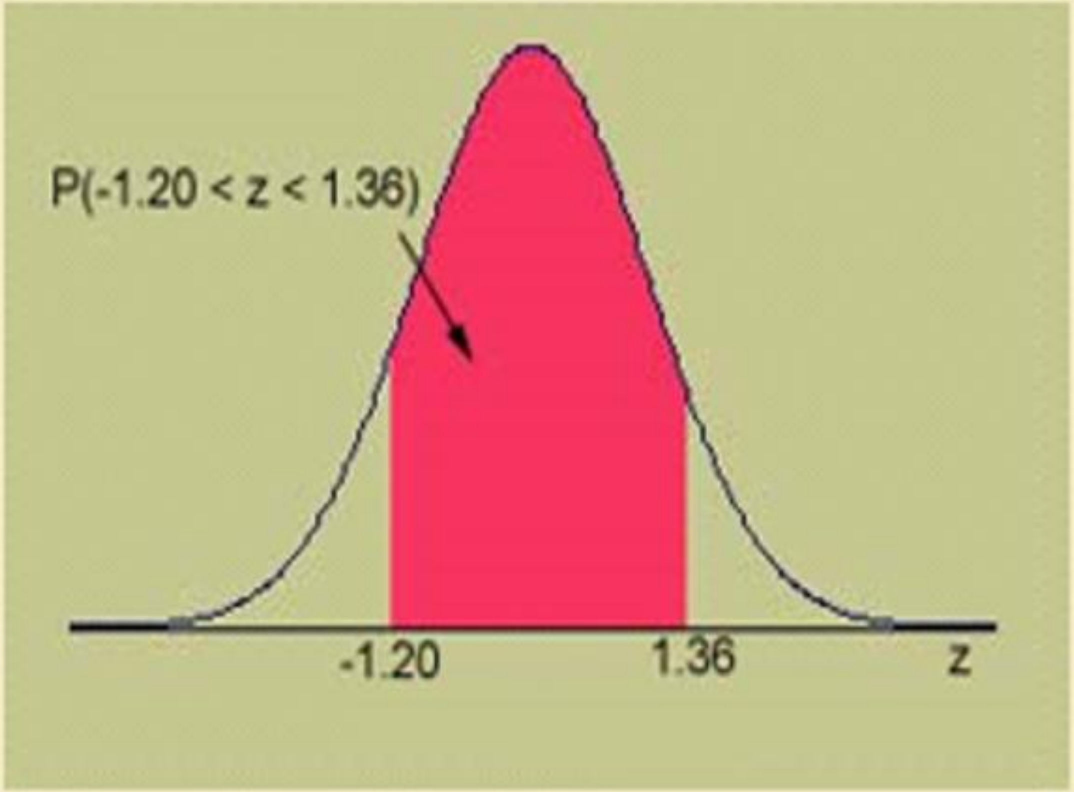
Area for $z = 1.36$

$P(z \leq 1.36) = .9131$

$P(z > 1.36)$

$= 1 - .9131 = .0869$

$P(-1.20 \leq z \leq 1.36) =$
$.9131 - .1151 = .7980$



NEGATIVE *z* Scores

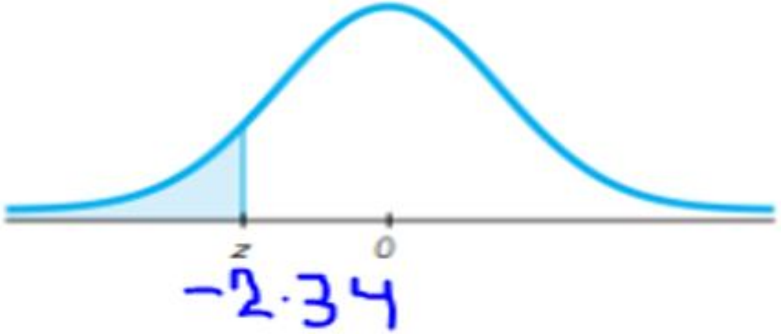
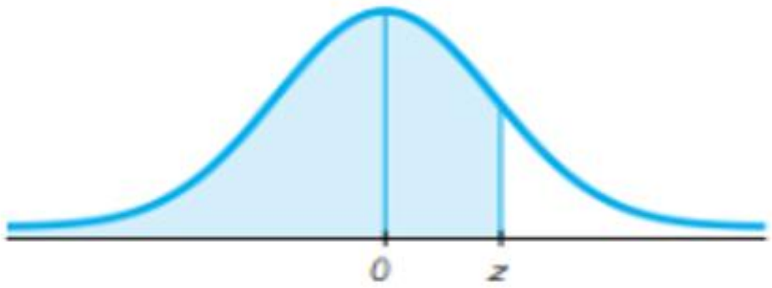


TABLE A-2 Standard Normal (*z*) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051 *	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233



1·54

POSITIVE *z* Scores

TABLE A-2 (*continued*) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495 *	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633

- Normal Distribution

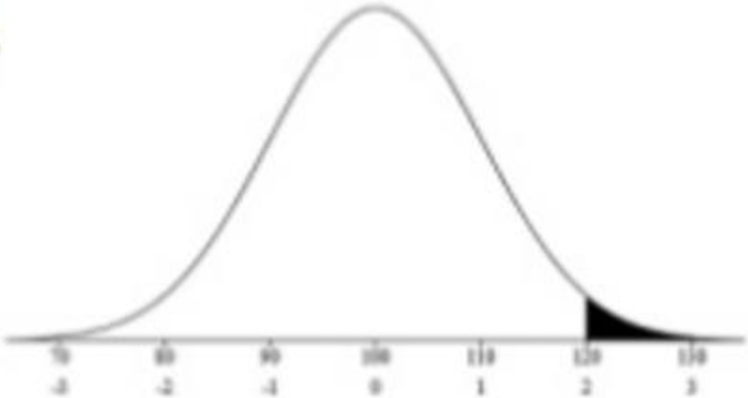
Exercise:

- **Question:** For a process with a mean of **100**, a standard deviation of **10** and an upper specification of **120**, what is the probability that a randomly selected item is defective (or beyond the upper specification limit)?

- Normal Distribution

Exercise:

- **Answer:**
  - The Z-score is equal to = **(120 – 100) / 10 = 2**.
  - This means that the upper specification limit is **2** standard deviations above the mean.
  - Now that we have the Z-score, we can use the Z-table to find the probability.



  - From the Z-table (the complementary cumulative table), the area under the curve for a Z-value of **2 = 0.02275** or **2.275%**.
  - This means that there is a chance of **2.275%** for any randomly selected item to be defective.

Evaluating Normality

- Not all continuous random variables are normally distributed
- It is important to evaluate how well the data set is approximated by a normal distribution

Evaluating Normality *(continued)*

- **Construct charts or graphs**
  - For small- or moderate-sized data sets, do stem-and-leaf display and box-and-whisker plot look symmetric?
  - For large data sets, does the histogram or polygon appear bell-shaped?
- **Compute descriptive summary measures**
  - Do the mean, median and mode have similar values?
  - Is the interquartile range approximately 1.33 σ?
  - Is the range approximately 6 σ?

(continued)

Assessing Normality

- **Observe the distribution of the data set**
  - Do approximately 2/3 of the observations lie within mean $\pm$ 1 standard deviation?
  - Do approximately 80% of the observations lie within mean $\pm$ 1.28 standard deviations?
  - Do approximately 95% of the observations lie within mean $\pm$ 2 standard deviations?
- **Kolmogorov-Smirnov Test:** In the case of a **large sample**, most researchers use K-S test to test the assumption of normality. This test should not be significant to meet the assumption of normality.

(If the **Sig.** value of the both tests is **>0.05**, the data is normal. If it is below 0.05, the data is significantly deviate form normal distribution.)

Thank You!