



Probability Distribution – It is a total listing of the various values the random variable can take along with the corresponding probability of each value.

- **Discrete Probability distribution** – The probability distribution uses a discrete random variable.

Binomial distribution

Poisson distribution

- **Continuous Probability Distribution** – The probability distribution uses a continuous random variable.

Normal distribution.

Binomial Distribution

Conditions for Binomial Distribution (Bernoulli Process)

- Trials are independent and random
- There are fixed number of trials (n trials)
- There are only two outcomes of the trial designated as success or failure.
- The probability of success is uniform through out the n trials.



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

Binomial Distribution plays a major role in Quality control and quality assurance function.

It is also being used in service organisations like banks, insurance corporations to get an idea regarding customers satisfaction regarding the quality of service.

In production it can be used to decide whether to accept or reject the lot on statistically designed sampling plans.

Binomial probability function

Under the conditions of a Bernoulli process,

The probability of getting x successes out of n trials is indeed the definition of a Binomial Distribution. The Binomial Probability function is given as:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Where $P(x)$ is the probability of getting x successes from n trials.

n

is the no. of ways in which x successes can take place out

x

Of n trials. = $n! / x!(n-x)!$

P is the probability of success, which is same throughout the n trials.

X can take values $0, 1, 2, 3, \dots, n$.



- Mean and Standard Deviation of the Binomial Distribution

The mean μ of the distribution is

$$\mu = E(x) = np$$

The Standard Deviation σ is

$$\sigma = \sqrt{np(1-p)}$$

Example

- A bank issues credit cards to customers under the scheme of Master card. Based on the past data, the bank has found out that 60% of all accounts pay on time following the bill. If a sample of 7 accounts is selected at random from the current database, construct the binomial probability distribution of accounts paying ontime.



- Solution:
 - Here $p=0.6$, $n=7$, $x=0$ to 7.
1. Probability of not getting more than 3 accounts paying on time _____.
 2. Prob. Of at least 4 accounts paying on time _____.
 3. What is the probability of no account paying on time? _____.

x	P(x)	Cumm. Probability
0	0.0016384	0.0016384
1	0.0172032	0.0188416
2	0.0774144	0.0962560
3	0.1935360	0.2897920
4	0.2912736	0.5800960
5	0.2612736	0.8413639
6	0.1306368	0.9720064
7	0.0279936	1.0000000



- Characteristics of Binomial Distribution
 - Binomial distribution can be symmetrical or skewed.
 - When $p=0.5$, the distribution will be symmetrical; regardless of how large or small the value of n .
 - When p is not equal to 0.5 , it will be skewed. The closer the p is to 0.5 and larger the value of n , the less skewed the distribution will be.
 - For example if $p=0.1$, the distribution will be right skewed. For example if $p=0.9$, the distribution will be left skewed.



- Example:

From the shop of a manufacturing company, the quality control department selected a sample of 15 items. According to the requirement, if 3 or more of the items in the sample are found to be defective, the entire production lot will be rejected and then the lot will go for 100% inspection. From the past data it is known that the probability of an item being defective is 0.04.

1. What is the probability that the lot will be rejected?
2. Find the mean and standard deviation of the Binomial distribution (x = number of defectives).



Poisson Distribution

- It is a discrete probability distribution which is useful when we are interested in the number of times a certain event will occur in a given unit of area or time (area of opportunity).
- Example: a quality assurance manager is interested in the number of defects on the surface of the refrigerator, an accountant is interested in the number of numerical errors per 100 invoices, a check post officer is interested to know the number of cars arriving at a highway check post per hour, the number of customers visiting a bank per hour, etc.



Centurion
UNIVERSITY
Shaping Lives...
Empowering Communities...

Poisson distribution is used to model situations that have the following properties:

- The person is interested to count the number of times a particular event occurs in a given area of opportunity. The area of opportunity is defined by time, length, surface area etc.
- The probability that an event occurs in a given area of opportunity is the same for all areas of opportunity.
- The number of events that occur in area of opportunity is independent of the number of the events that occur in other areas of opportunity.
- The probability that two or more events will occur in an area of opportunity approaches zero as the area of opportunity becomes smaller.



- **Poisson Distribution Formula:**

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Where:

$P(x)$ = probability of x successes given an idea of λ

λ = average number of successes

$e = 2.71828$ (based on natural logarithm)

x = successes per unit which can take value $0,1,2,3,4,\dots,\infty$

λ is the parameter of the Poisson distribution

Mean of the Poisson Distribution = λ

Standard deviation of the Poisson distribution is = $\sqrt{\lambda}$



- **Example:**

If on an average 6 customers arrive every 2 minutes at a bank during the busy hours

1. What is the probability that exactly 4 customers arrive in a given minute?
2. What is the probability that more than 3 customers will arrive in a given minute?



Centurion
UNIVERSITY
Shaping Lives...
Empowering Communities...

Normal Distribution

- This is a continuous probability distribution which is also called as Gaussian distribution. It denotes a classical bell-shaped distribution curve. It forms the basis for all inferential statistics.



Properties of the Normal Distribution:

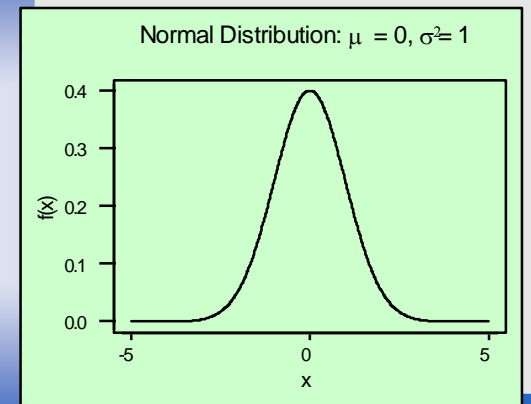
- It is bell shaped in its appearance
- It is symmetrical about its mean
- Its measures of central tendencies (mean, median, mode) are all same
- Its associated random variable has an infinite range ($-\infty < x < \infty$)
- If the tails of the normal distribution are extended, they will run parallel to the horizontal axis without actually touching it.

The Normal Probability Distribution

The normal probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad -\infty < x < \infty$$

where $e = 2.7182818 \dots$ and $\pi = 3.14159265 \dots$





Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

- Standard normal variable

$$z = \frac{x - \mu}{\sigma}$$

Where μ = mean

σ = standard deviation



Example:

- The mean weight of a morning breakfast cereal pack is 0.295kg with a standard deviation of 0.025kg. The random variable weight of the pack follows a normal distribution.
1. What is the probability that the pack weighs less than 0.280kg?
 2. What is the probability that the pack weighs more than 0.350kg?
 3. What is the probability that the pack weighs between 0.260kg to 0.340kg ?



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

- What is the weight which will be there in 90% of the packs.



Centurion
UNIVERSITY

*Shaping Lives...
Empowering Communities...*

- **Hypergeometric Distribution**

The Binomial Distribution and the Hypergeometric Distribution are having similar characteristics, it tells us about the number of successes in n observations if we know the probability of success.

What differentiates the two is the way in which data are obtained.

For Binomial Distribution- sample data are drawn with replacement from a finite population or without replacement from a infinite population.

For Hypergeometric distribution- sample data is drawn without replacement from a finite population.

It means in Binomial distribution outcome of any particular observation is independent of the other.

The same is not true for Hypergeometric distribution.



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

Example

An organisation is trying to create a team of 8 people from different department who have a knowledge of a particular process. A total of 30 people within the organisation have knowledge of the process and 10 of these are from design department. If the members of a team are selected at random, what is the probability that the team will contain 2 members from design department?

Normal approximation of Binomial Distribution

When the no. of trials n is large (for eg. >1000), then the Binomial distribution approximates to Normal distribution.

The mean is np

The standard deviation is $\sqrt{np(1-p)}$

Continued

- Since the Binomial distribution is discrete and Normal distribution is continuous a continuity correction is required.
- To do so, subtract 0.5 from the left limit and add 0.5 to the right limit.
- **For example**

A total student of 2,058 take a difficult test. Each student has an independent 0.6205 probability of passing the test.

1. What is the probability that between 1,250 and 1,300 students, both no.s inclusive will pass?
2. What is the probability that at least 1,300 students will pass?

The Normal Distribution

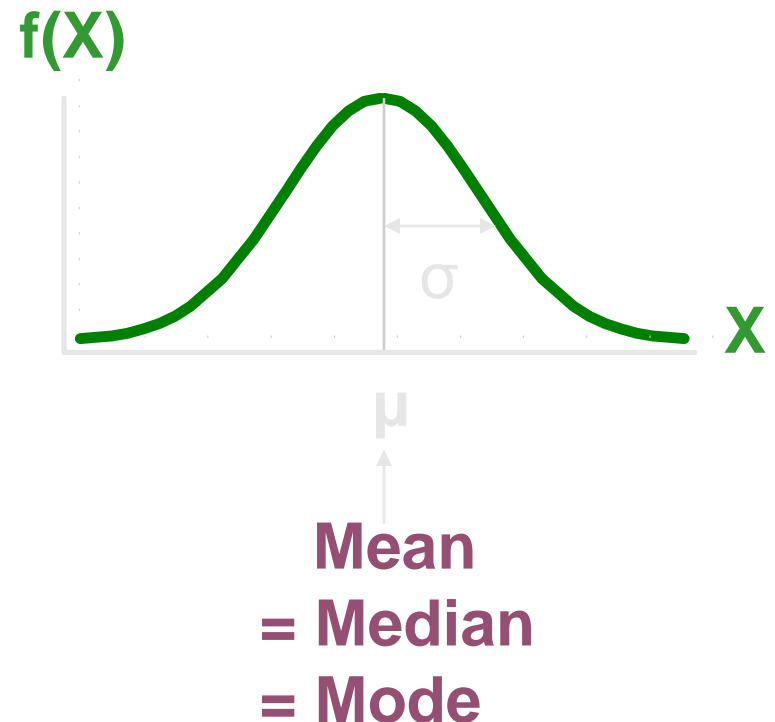
- **Bell Shaped**
- **Symmetrical**
- **Mean, Median and Mode are Equal**

Location is determined by the mean, μ

Spread is determined by the standard deviation, σ

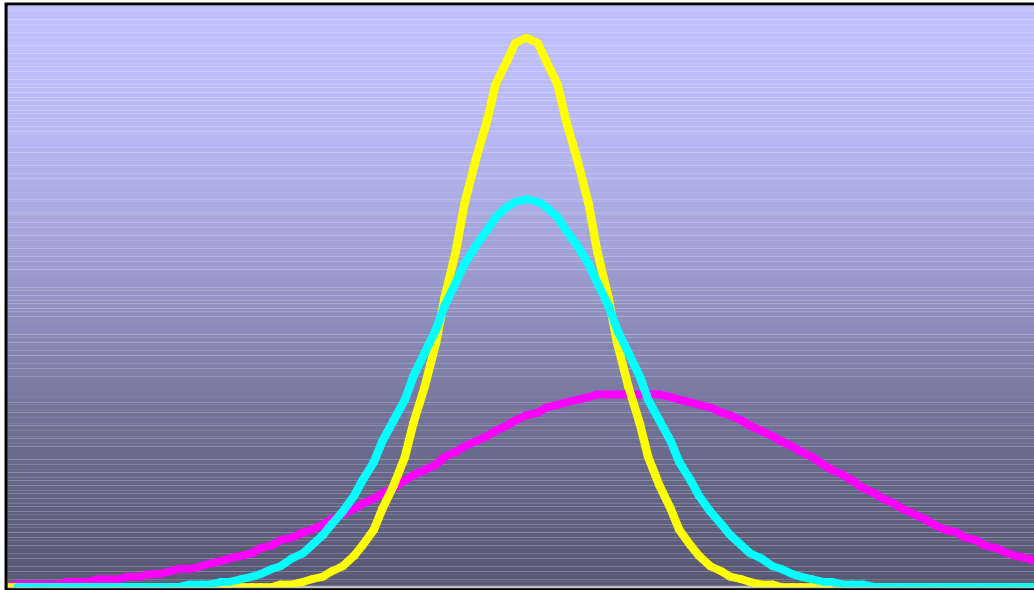
The random variable has an infinite theoretical range:

$+\infty$ to $-\infty$



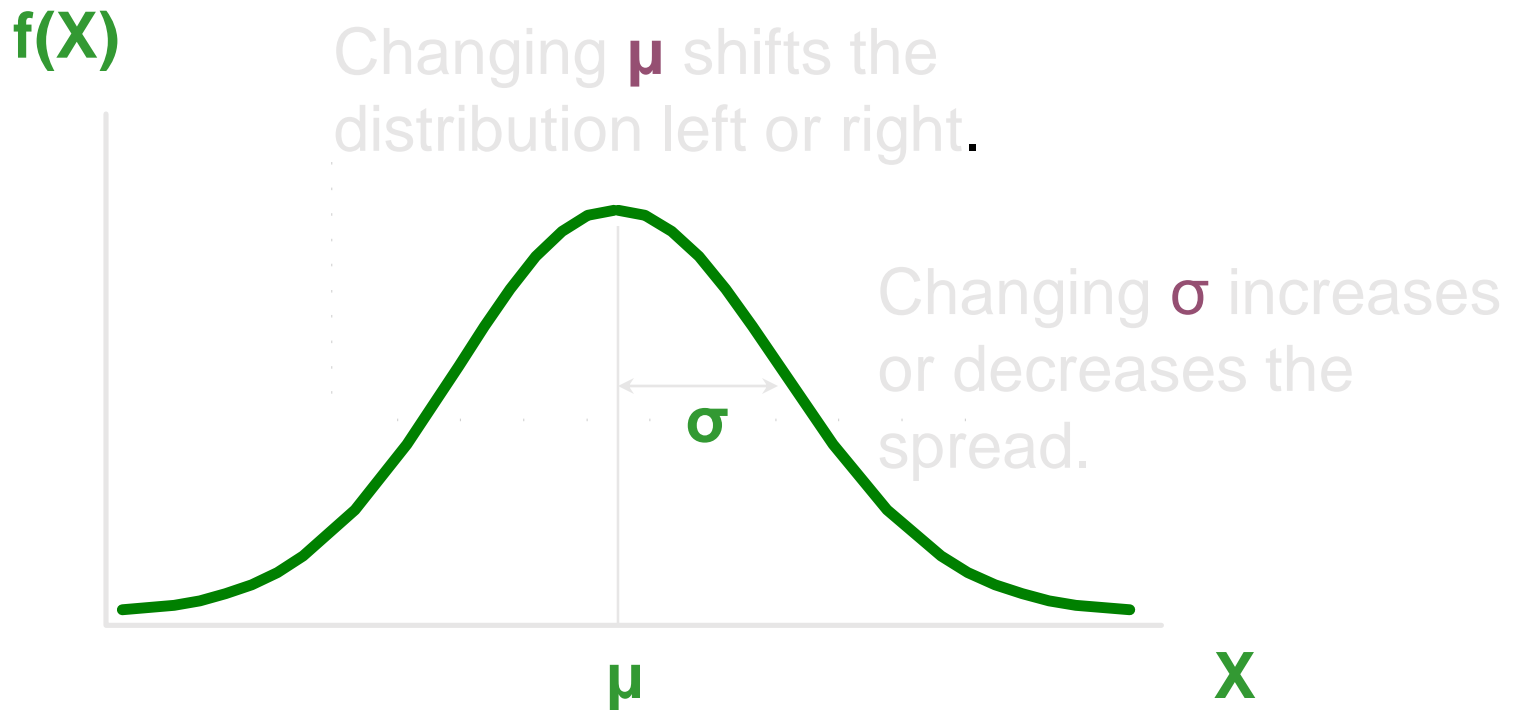


Many Normal Distributions



By varying the parameters μ and σ , we obtain different normal distributions

The Normal Distribution Shape



The Normal Probability Density Function

- The formula for the normal probability density function is

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(X-\mu)/\sigma]^2}$$

Where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

μ = the population mean

σ = the population standard deviation

X = any value of the continuous variable

The Standardized Normal

- Any normal distribution (with any mean and standard deviation combination) can be transformed into the standardized normal distribution (Z)
- Need to transform X units into Z units

Translation to the Standardized Normal Distribution

- Translate from X to the standardized normal (the “Z” distribution) by **subtracting the mean** of X and **dividing by its standard deviation**:

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

The Z distribution always has mean = 0 and standard deviation = 1

The Standardized Normal Probability Density Function

- The formula for the standardized normal probability density function is

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

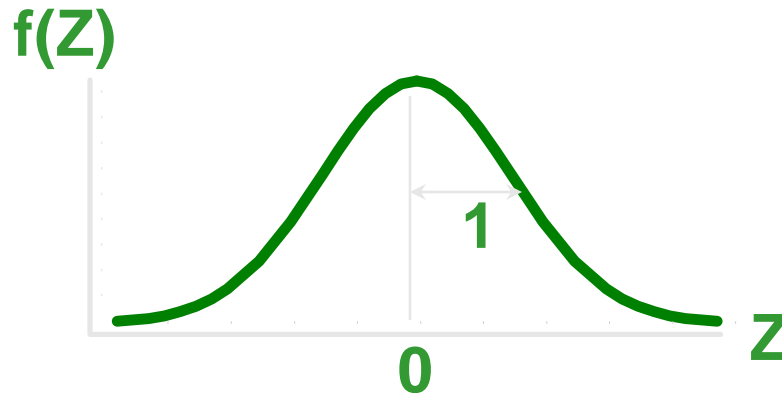
Where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

Z = any value of the standardized normal distribution

The Standardized Normal Distribution

- Also known as the “Z” distribution
- Mean is 0
- Standard Deviation is 1



Values above the mean have **positive** Z-values, values below the mean have **negative** Z-values



Example

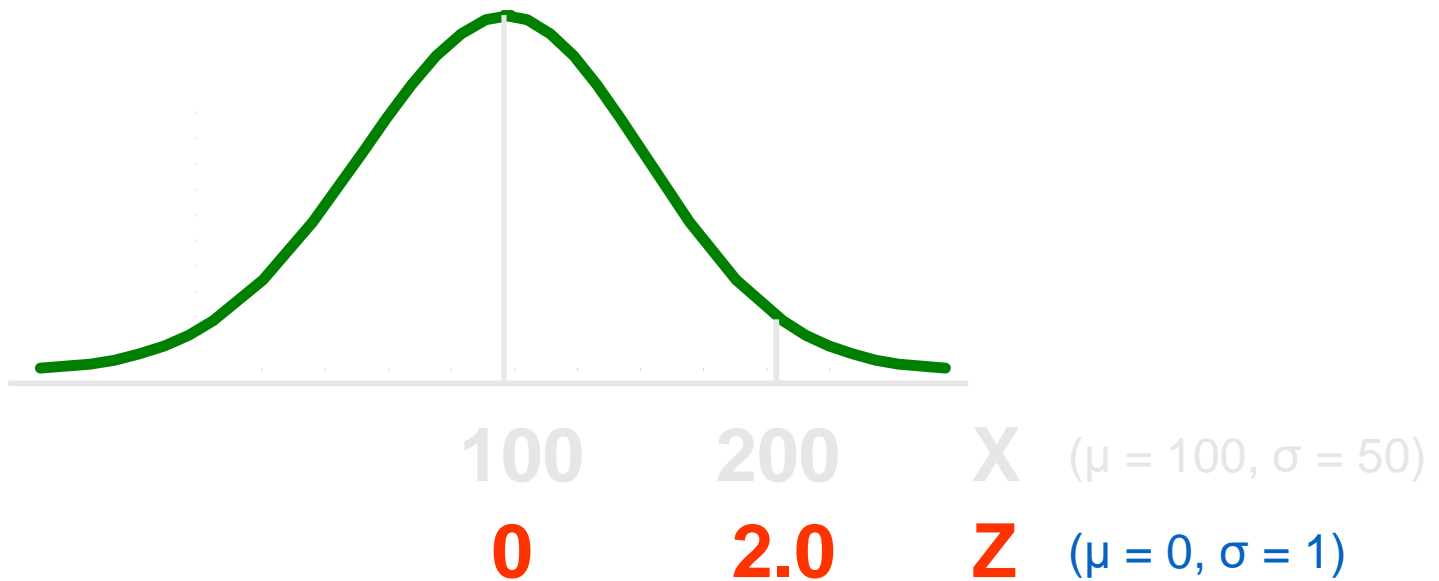
- If X is distributed normally with mean of 100 and standard deviation of 50, the Z value for $X = 200$ is

$$Z = \frac{X - \mu}{\sigma} = \frac{200 - 100}{50} = 2.0$$

- This says that $X = 200$ is two standard deviations (2 increments of 50 units) above the mean of 100.



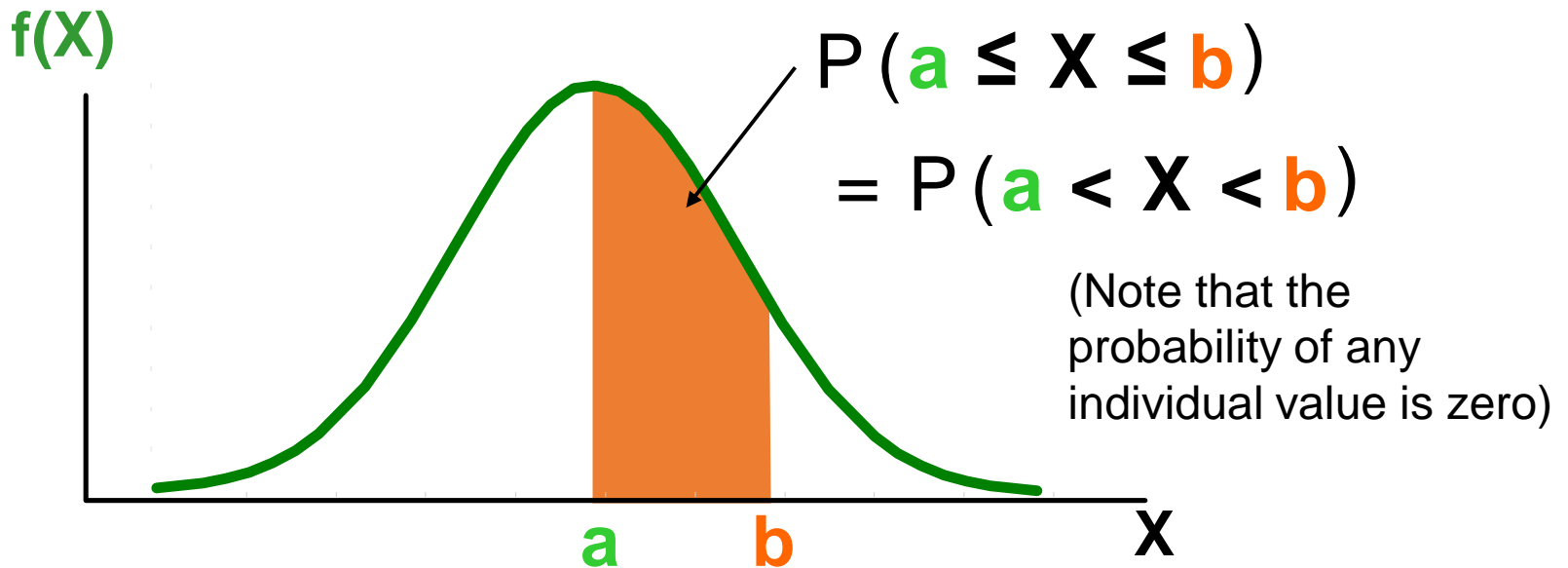
Comparing X and Z units



Note that the distribution is the same, only the scale has changed. We can express the problem in original units (X) or in standardized units (Z)

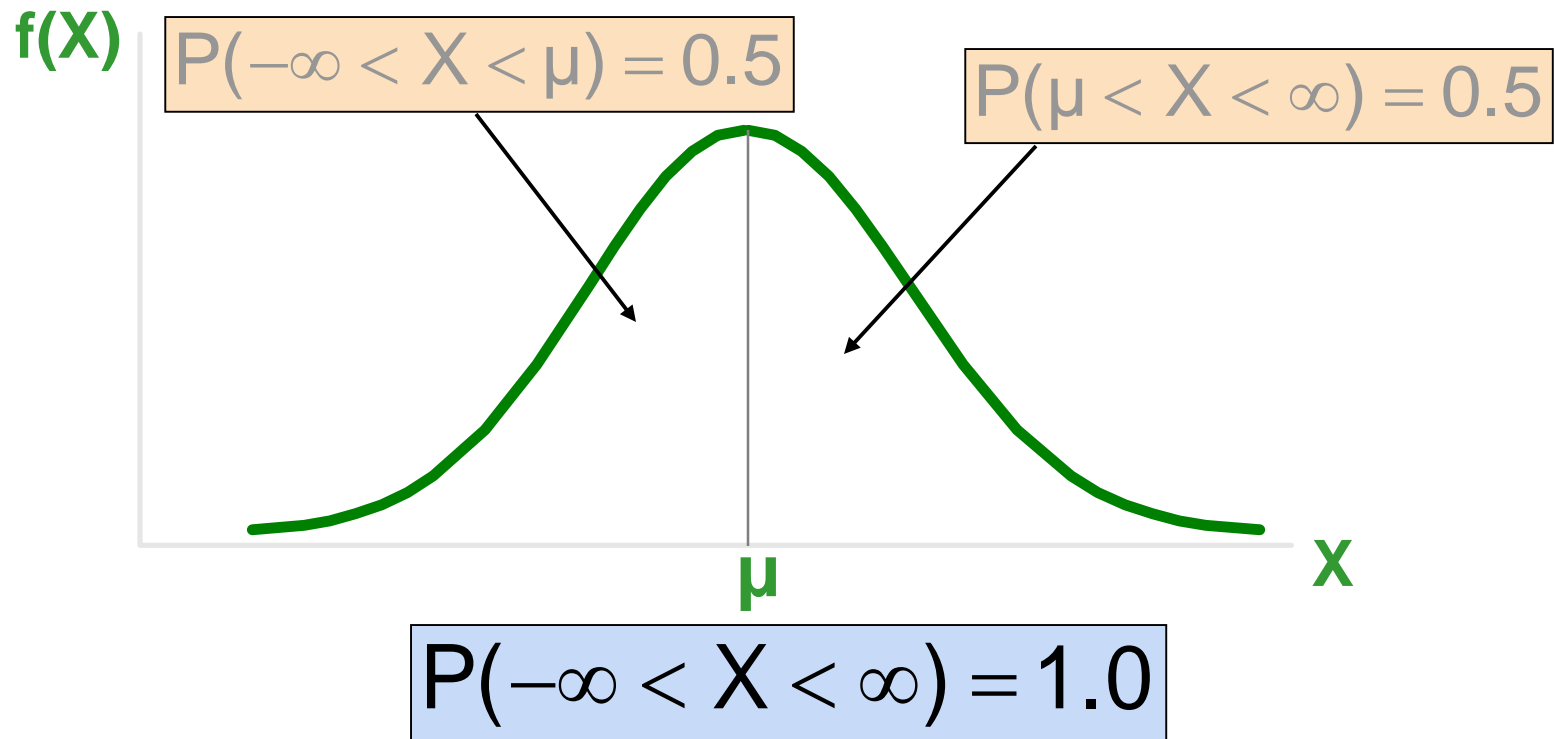
Finding Normal Probabilities

Probability is measured by the area under the curve



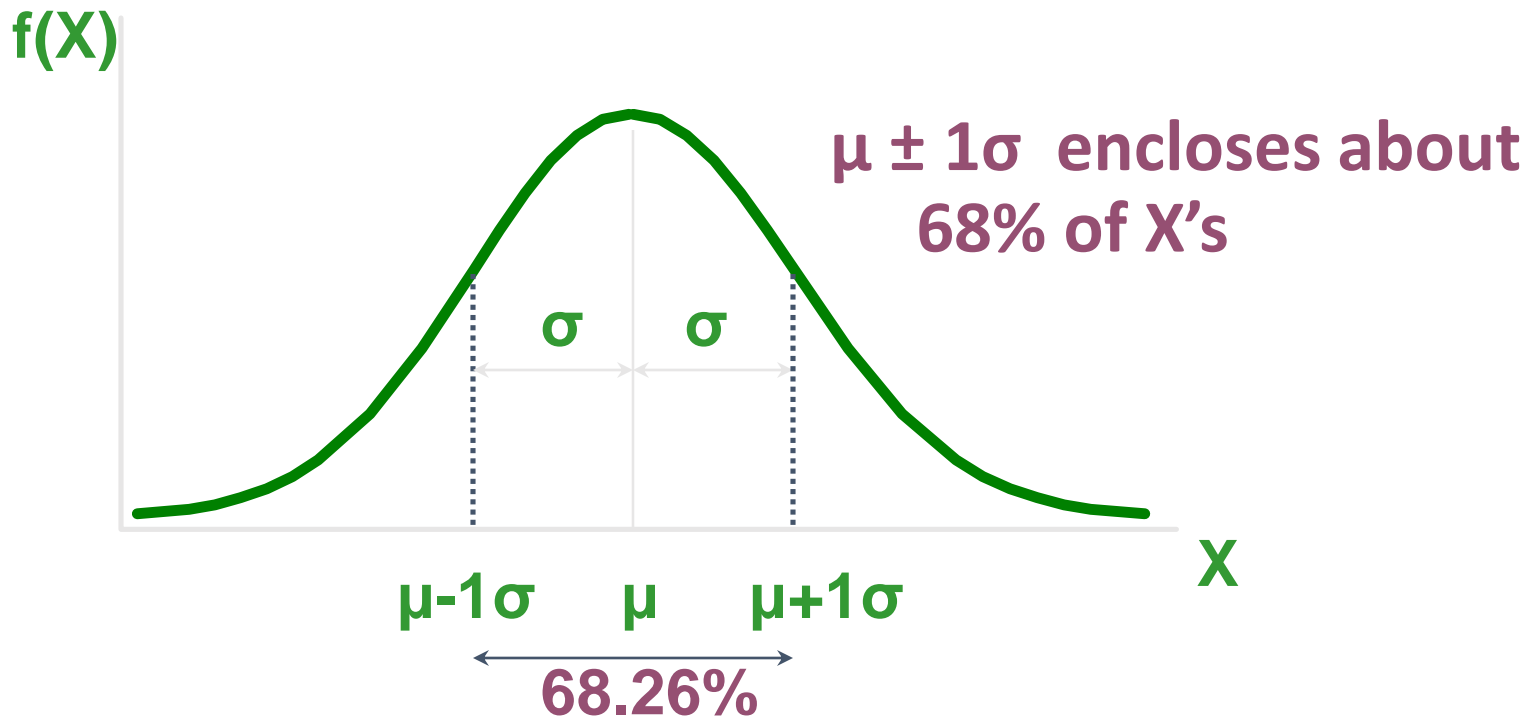
Probability as Area Under the Curve

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below



Empirical Rules

What can we say about the distribution of values around the mean? There are some general rules:





The Empirical Rule

(continued)

- $\mu \pm 2\sigma$ covers about **95%** of X 's
- $\mu \pm 3\sigma$ covers about **99.7%** of X 's

