



SAPIENZA
UNIVERSITÀ DI ROMA



UNIVERSITÀ
DEGLI STUDI DELLA
TUSCIA

Statistics for business and decision making (SBDM)

Prof. Ilaria Benedetti

Random variables

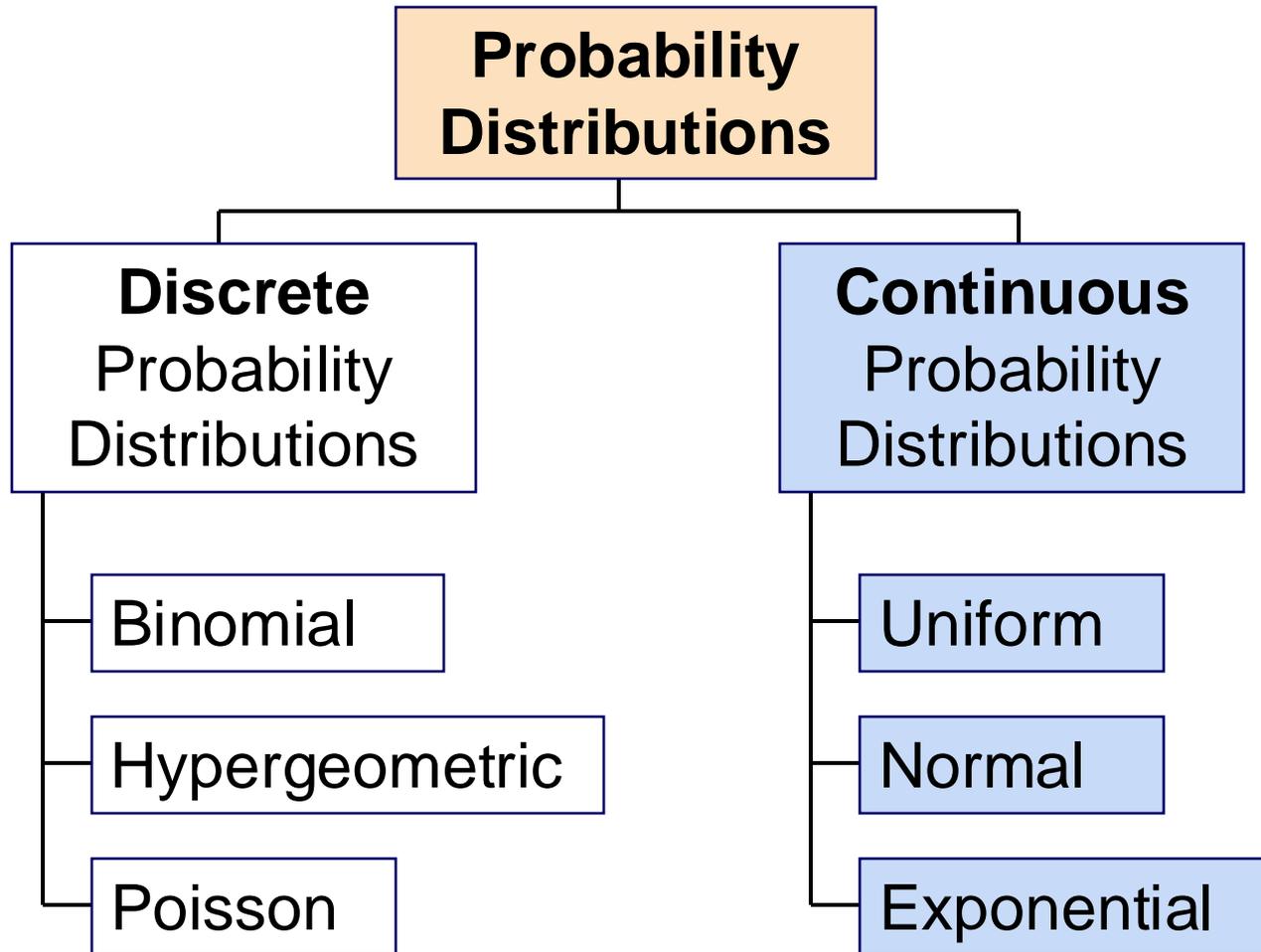
Random Variable

A **random variable** (r.v.) is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (and only one) numerical value is assigned to each sample point.

Random variables are **aleatory variables** or **stochastic variables**.

1. Random variables that can assume a countable number (finite) of values are called **discrete**. For convenience the values of a discrete r.v. X are sorted in ascending order.
2. Random variables that can assume values corresponding to any of the points contained in one or more intervals (i.e., values that are infinite and uncountable) are called **continuous**.

Probability Distributions



Discrete Random Variable

Examples

Experiment	Random Variable	Possible Values
Make 100 Sales Calls	# Sales	0, 1, 2, ..., 100
Inspect 70 Radios	# Defective	0, 1, 2, ..., 70
Answer 33 Questions	# Correct	0, 1, 2, ..., 33
Count Cars at Toll Between 11:00 & 1:00	# Cars Arriving	0, 1, 2, ..., ∞

Continuous Random Variable Examples

Experiment	Random Variable	Possible Values
Weigh 100 People	Weight	45.1, 78, ...
Measure Part Life	Hours	900, 875.9, ...
Amount spent on food	\$ amount	54.12, 42, ...
Measure Time Between Arrivals	Inter-Arrival Time	0, 1.3, 2.78, ...

Probability Distributions for Discrete Random Variables

Discrete Probability Distribution

The **probability distribution** of a **discrete random variable** is a graph, table, or formula that specifies the probability associated with each possible value the random variable can assume.

Discrete Probability Distribution Example

Experiment: Toss 2 coins. Count number of tails.



Probability Distribution

Values, x Probabilities, $p(x)$

0

$$1/4 = .25$$

1

$$2/4 = .50$$

2

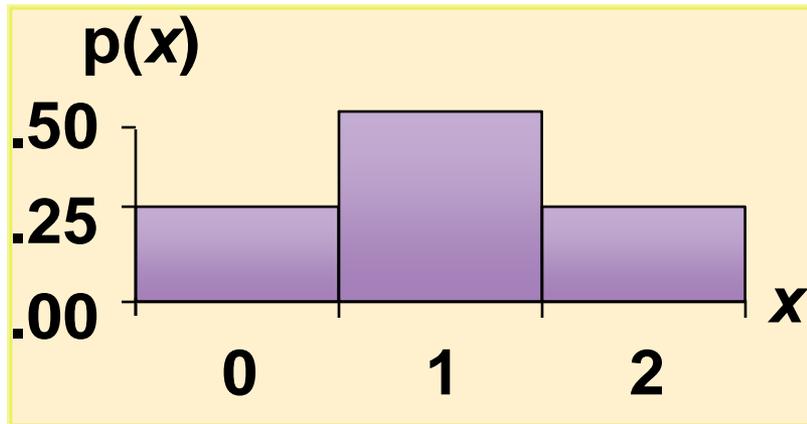
$$1/4 = .25$$

Visualizing Discrete Probability Distributions

Listing

$\{ (0, .25), (1, .50), (2, .25) \}$

Graph



Table

# Tails	$f(x)$ Count	$p(x)$
0	1	.25
1	2	.50
2	1	.25

The table can be described

by this function:

$$f(x) = P(X = x),$$
$$x = x_1, x_2, \dots, x_k$$

Requirements (rules) for the Probability Distribution of a Discrete Random Variable x

$f(x)$ is the probability function.

It satisfies the following properties:

1. $f(x) \geq 0$ for all values of x
2. $\sum f(x) = 1$

where the summation of $f(x)$ is over all possible values of x .

Requirements (rules) for the cumulative distribution function of a Discrete Random Variable x

Cumulative distribution function

is obtained by associating each value x the sum of the probabilities corresponding to x and to all values below it:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

The cumulative distribution function has the following properties:

- 1. $F(x)$ is defined for any x ;**
- 2. $0 \leq F(x) \leq 1$ for all values of x ;**
- 3. $F(x)$ is a non decreasing function of x $F(a) \leq F(b)$ for any pair of numbers a and b .**

Some parametric Probability distributions

■ Binomial distribution

- Number of 'successes' in a sample of n observations (trials):
 - Number of defective items in a batch of 5 items
 - Number correct on a 33 question exam
 - Number of customers who purchase out of 100 customers who enter store (each customer is equally likely to purchase)

■ Poisson distribution

- Number of events that occur in an interval: events per unit: Time, Length, Area, Space
 - Number of customers arriving in 20 minutes
 - Number of strikes per year in the U.S.
 - Number of defects per lot (group) of DVD's

Continuous Probability Distributions

- A **continuous random variable** is a variable that can assume any value in an interval
 - thickness of an item
 - time required to complete a task
 - temperature of a solution
 - height, in inches
- These can potentially take on any value, depending only on the ability to measure accurately.

Cumulative Distribution Function

- The **cumulative distribution function**, $F(x)$, for a continuous random variable X expresses the probability that X does not exceed the value of x

$$F(x) = P(X \leq x)$$

- Let a and b be two possible values of X , with $a < b$. The probability that X lies between a and b is

$$P(a < X < b) = F(b) - F(a)$$

Probability Density Function

The **probability density function**, $f(x)$, of random variable X has the following properties:

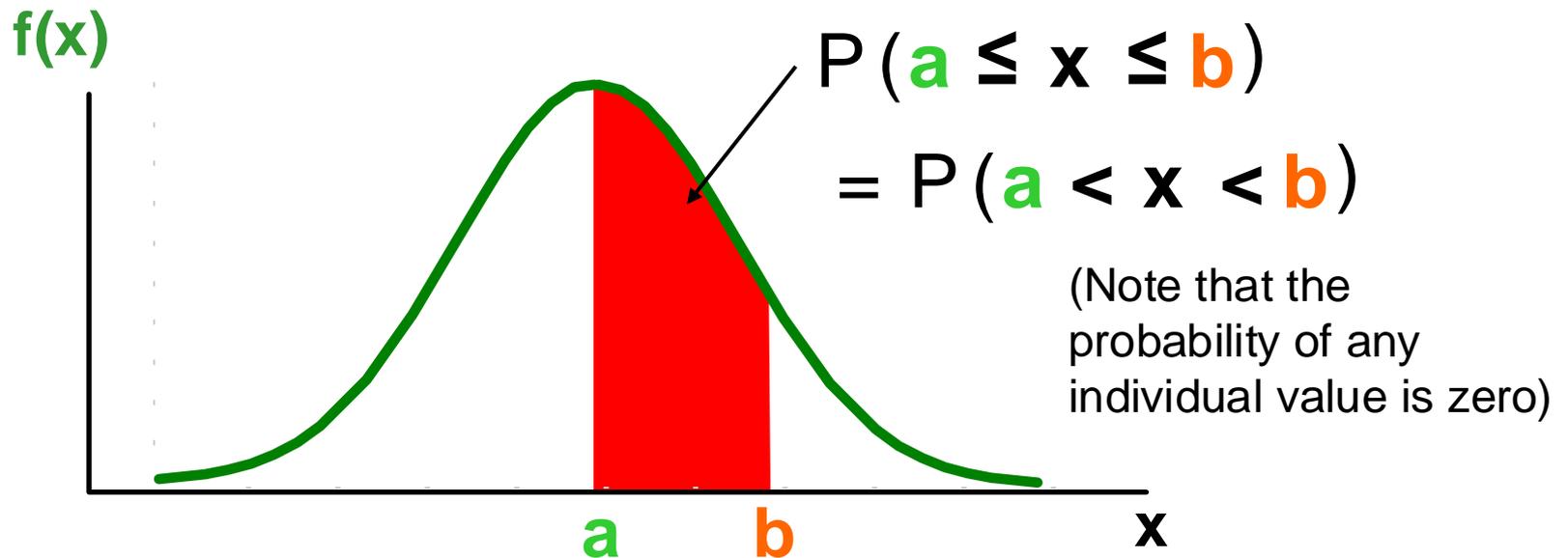
1. $f(x) > 0$ for all values of x
2. The area under the probability density function $f(x)$ over all values of the random variable X is equal to 1.0
3. The probability that X lies between two values is the area under the density function graph between the two values
4. The **cumulative density function** $F(x_0)$ is the area under the probability density function $f(x)$ from the minimum x value up to x_0

$$F(x_0) = \int_{x_m}^{x_0} f(x) dx$$

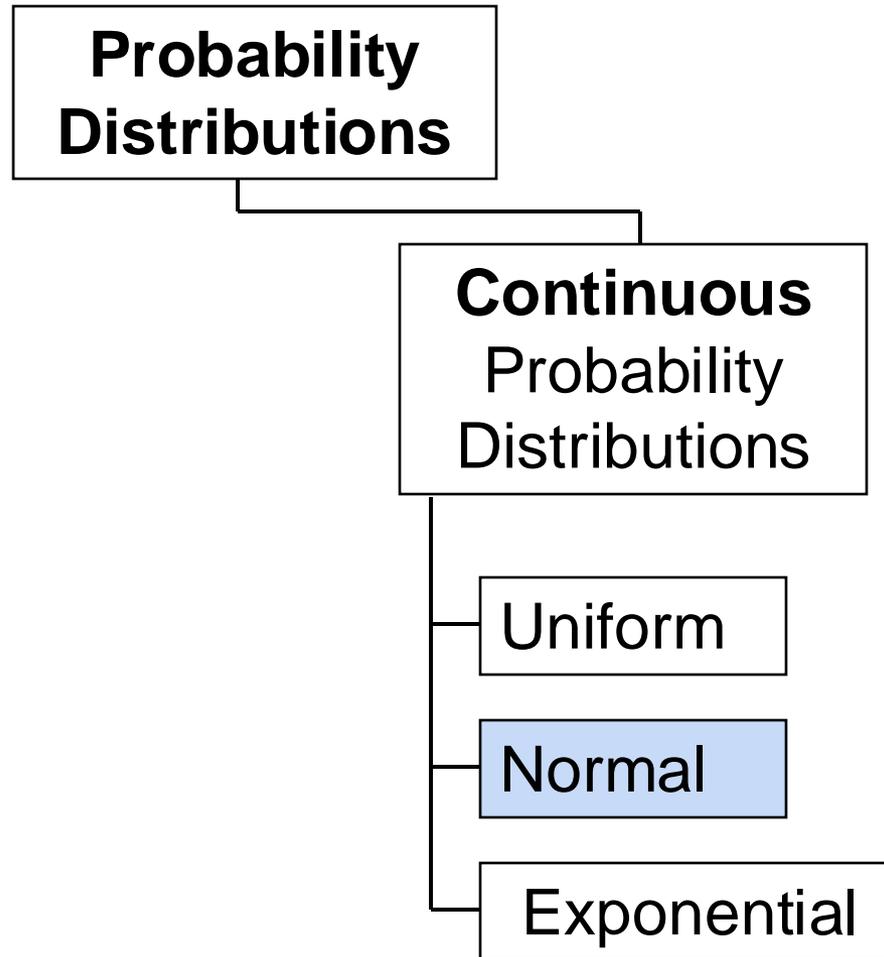
where x_m is the minimum value of the random variable x

Probability as an Area

Shaded area under the curve is the probability that X is between a and b



The Normal Distribution



The Normal Distribution

(continued)

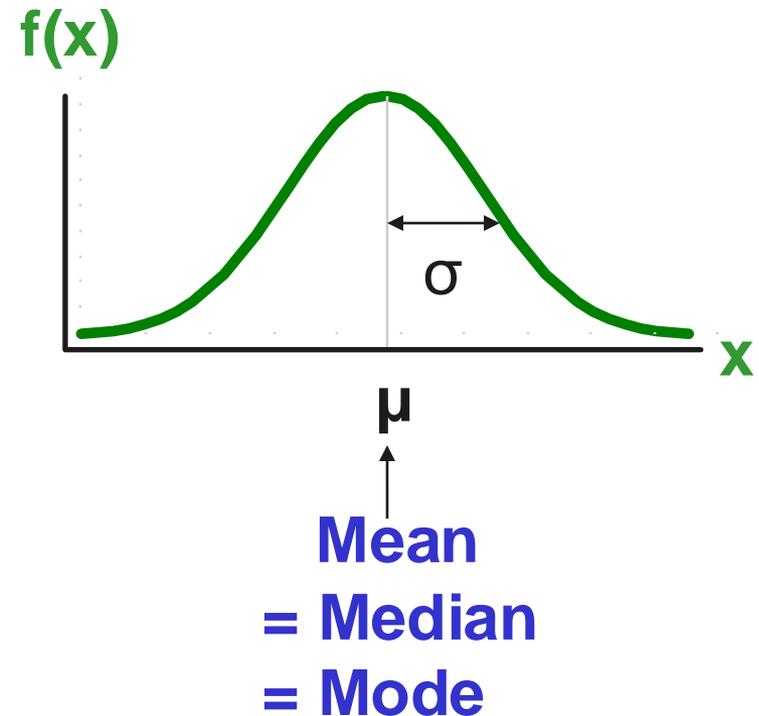
- Bell Shaped
- Symmetrical
- Mean, Median and Mode are Equal

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 $-\infty$

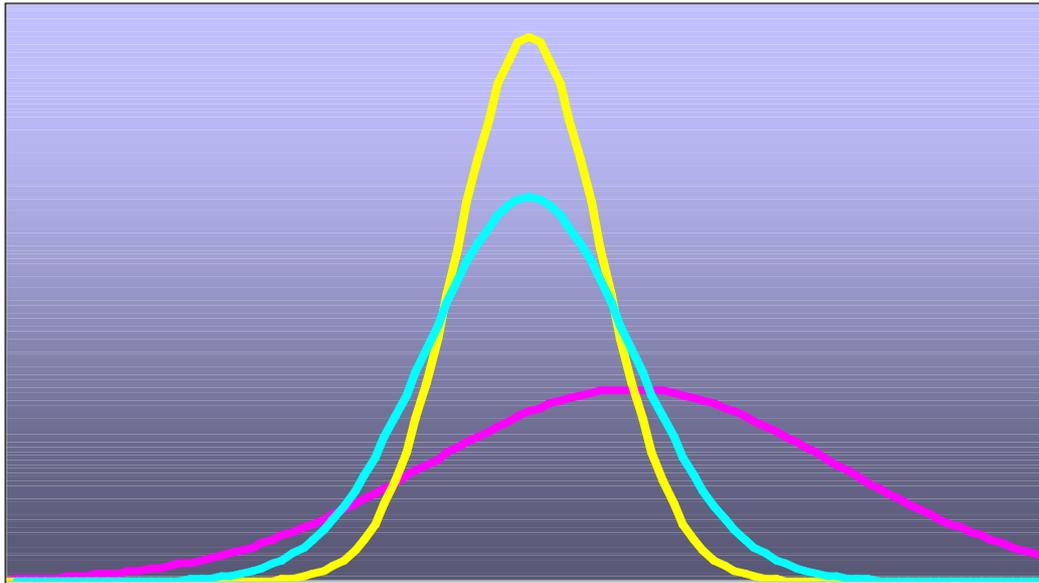


The Normal Distribution

)

- The normal distribution closely approximates the probability distributions of a wide range of random variables
- Distributions of sample means approach a normal distribution given a “large” sample size
- Computations of probabilities are direct and elegant
- The normal probability distribution has led to good business decisions for a number of applications

Many Normal Distributions

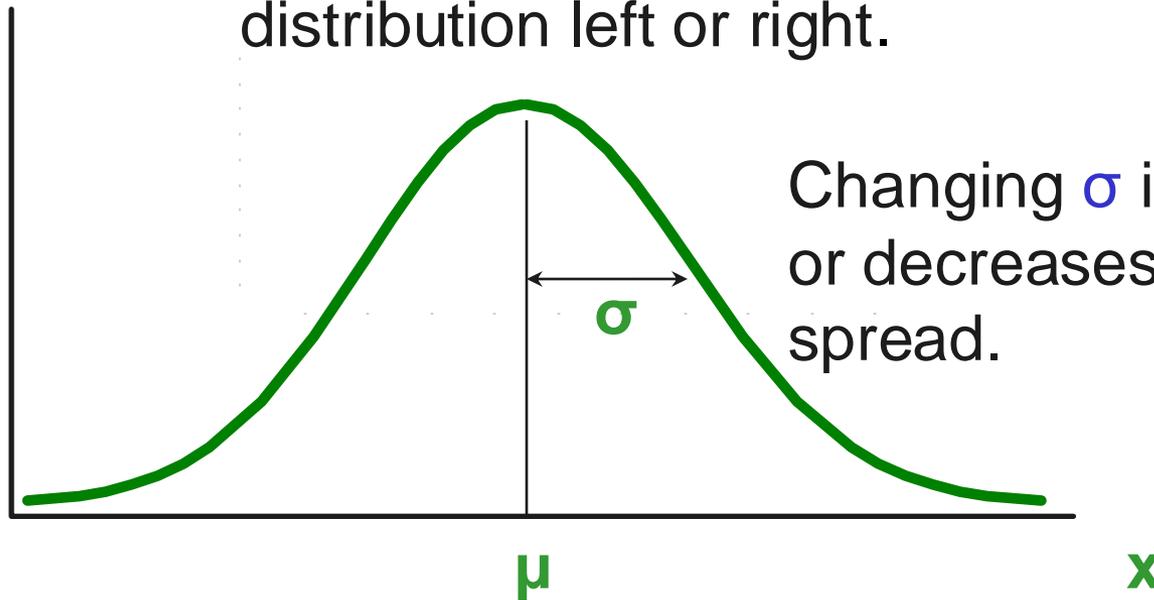


By varying the parameters μ and σ , we obtain different normal distributions

The Normal Distribution Shape

$f(x)$

Changing μ shifts the distribution left or right.



Given the mean μ and variance σ we define the normal distribution using the notation

$$X \sim N(\mu, \sigma^2)$$

The Normal Probability Density Function

- The formula for the normal probability density function is

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

Where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

μ = the population mean

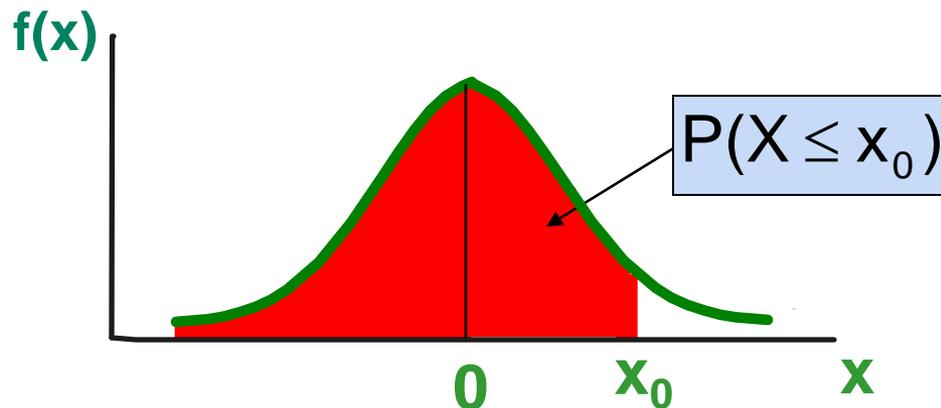
σ = the population standard deviation

x = any value of the continuous variable, $-\infty < x < \infty$

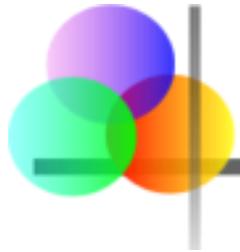
Cumulative Normal Distribution

- For a normal random variable X with mean μ and variance σ^2 , i.e., $X \sim N(\mu, \sigma^2)$, the **cumulative distribution function** is

$$F(x_0) = P(X \leq x_0)$$



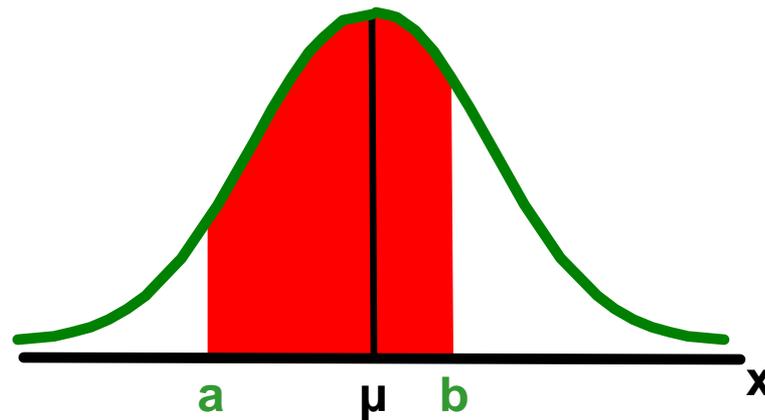
Is the area under the normal curve to the left of x



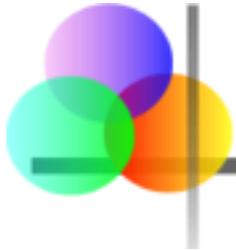
Finding Normal Probabilities

The probability for a range of values is measured by the area under the curve:

$$P(a < X < b) = F(b) - F(a)$$



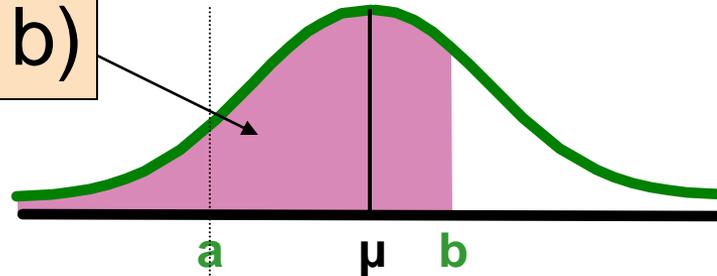
It shows the probability that X falls in any interval a and b .



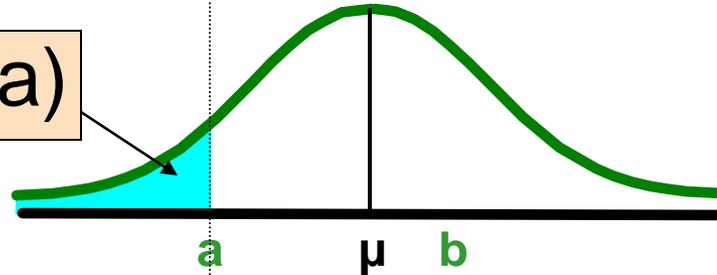
Finding Normal Probabilities

(continued)

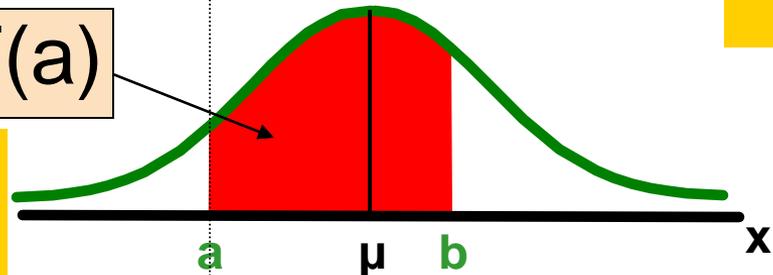
$$F(b) = P(X < b)$$



$$F(a) = P(X < a)$$

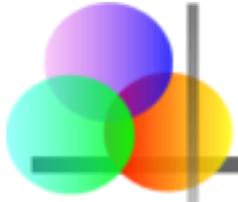


$$P(a < X < b) = F(b) - F(a)$$



The probability that $a < x < b$ is equal to the difference between the area under the normal curve to the left of b (pink area) and the area under curve to the left of a (light blue)

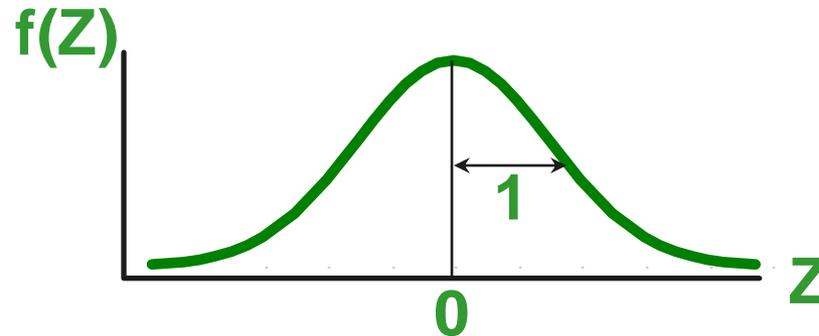
This probability can be computed by using Statistical softwares (eg. R, excel...) which provide the cumulative function for a given normal distribution $N(\mu; \sigma^2)$ for a given mean and variance. Alternatively, we one can be use the normal table



The Standardized Normal

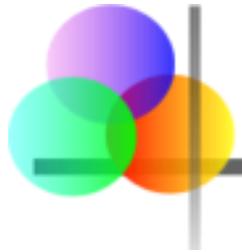
- Any normal distribution (with any mean and variance combination) can be transformed into the standardized normal distribution (Z), with mean 0 and variance 1

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



- Need to transform X units into Z units by subtracting the mean of X and dividing by its standard deviation

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

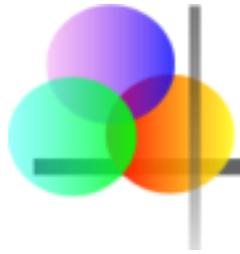


Example

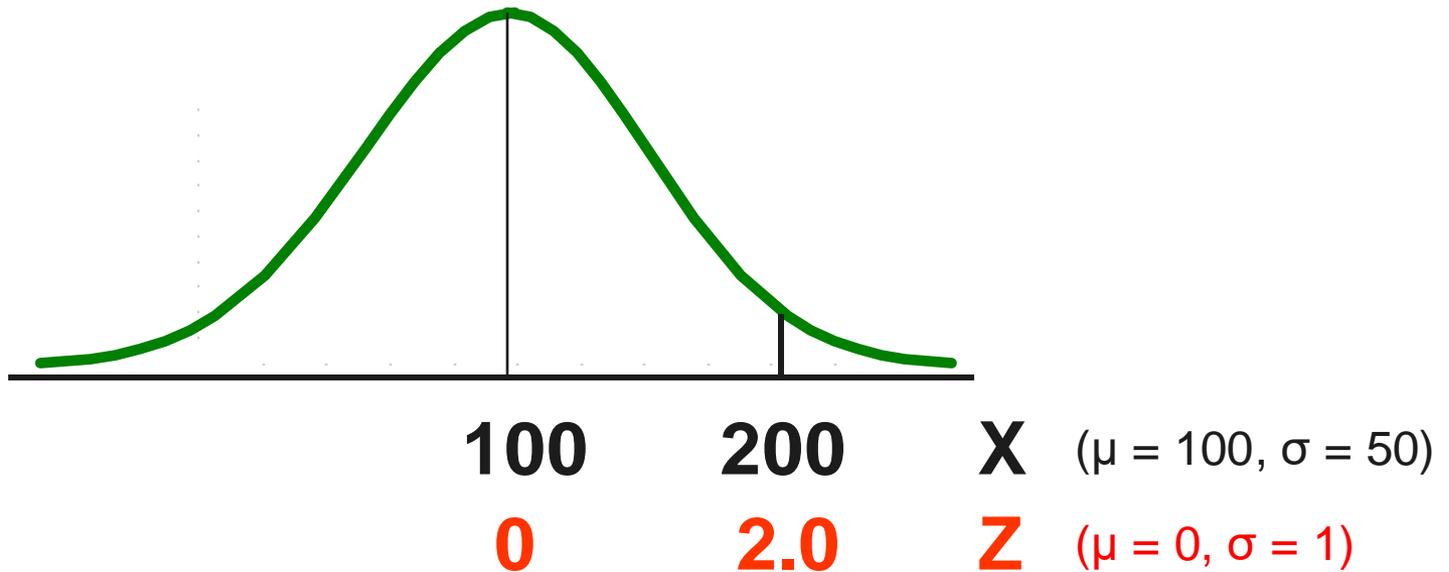
- If X is distributed normally with mean of 100 and standard deviation of 50, the Z value for $X = 200$ is

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

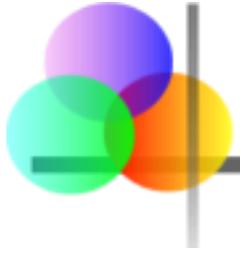
- This says that $X = 200$ is two standard deviations (2 increments of 50 units) above the mean of 100.



Comparing X and Z units

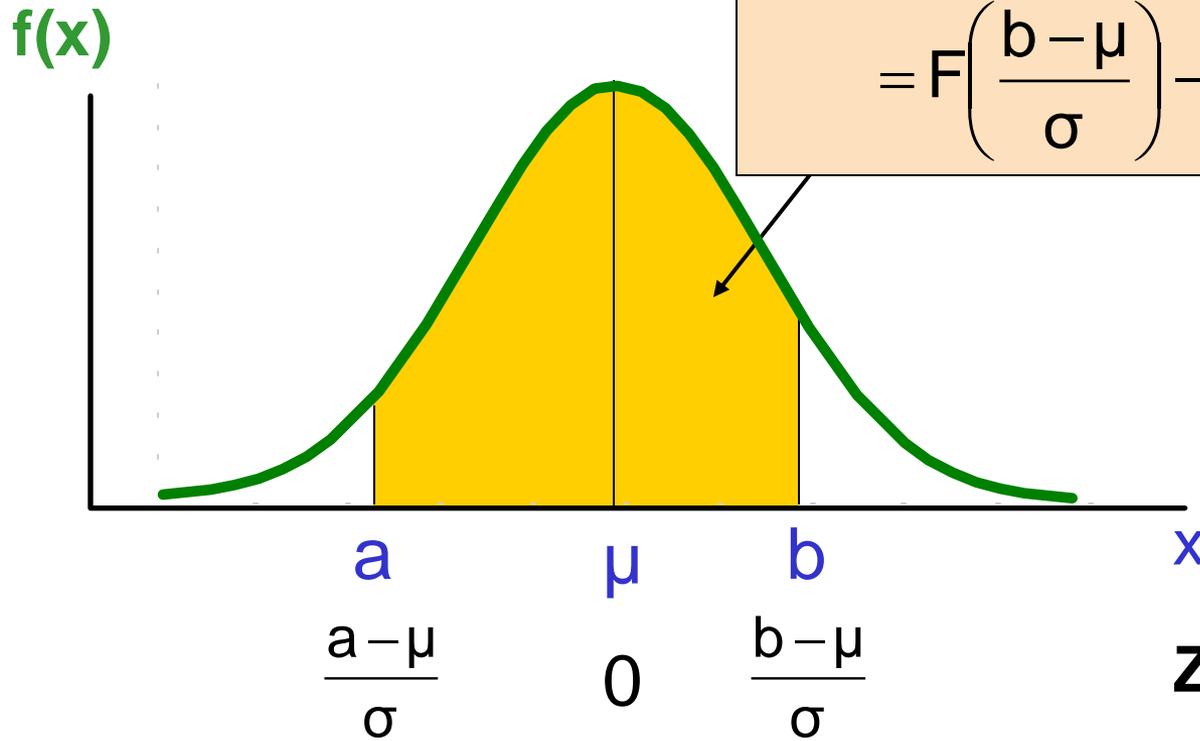


Note that the distribution is the same, only the scale has changed. We can express the problem in original units (X) or in standardized units (Z)



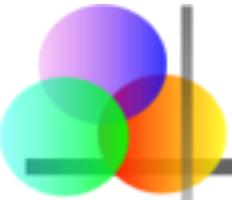
Finding Normal Probabilities

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) \\ &= F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$



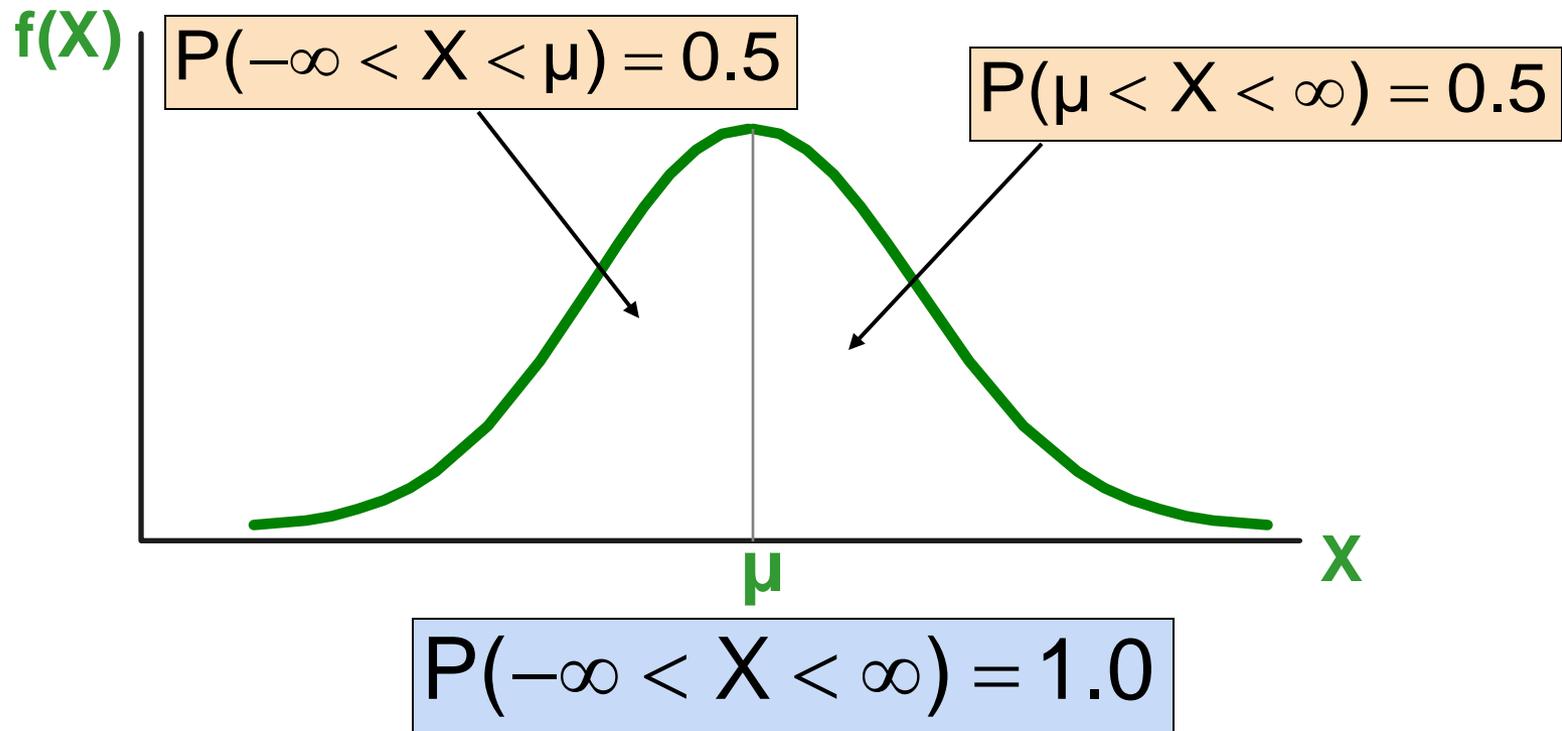
$F = \Phi$ it represents the cumulative distribution function of the standardized normal distribution

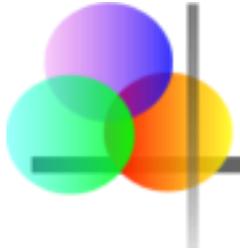
The cumulative distr. Function $\Phi(z)$ determines the probabilities related to the normal distribution $N(\mu; \sigma^2)$ with any mean and variance



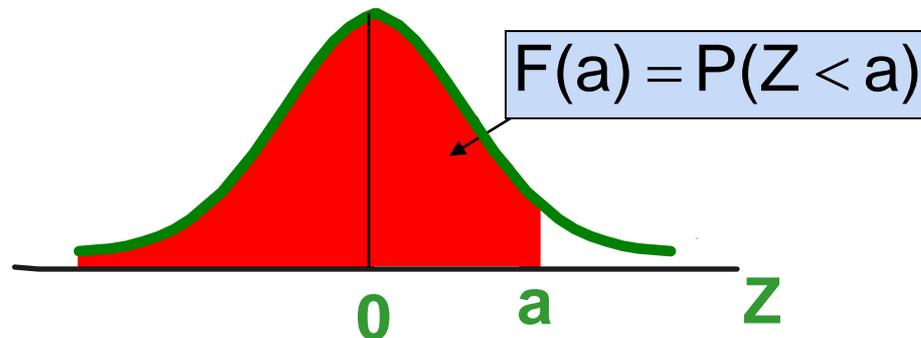
Probability as Area Under the Curve

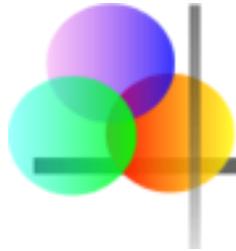
The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below





- The Standardized Normal table shows values of the cumulative normal distribution function
- For a given Z-value a , the table shows $F(a)$ (the area under the curve from negative infinity to a)

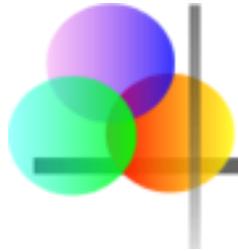




The Standardized Normal Table

- Standardized Normal Table gives the probability $F(a)$ for any value a
- The standardized Normal table contains the values of the cumulative distribution function of the normal standard distribution for non-negative values of z , ranging from 0 to 3,09.
- The values of $\Phi(z)$ for negatives values of z can be obtained using the following equation: $\Phi(-z) = 1 - \Phi(z)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177



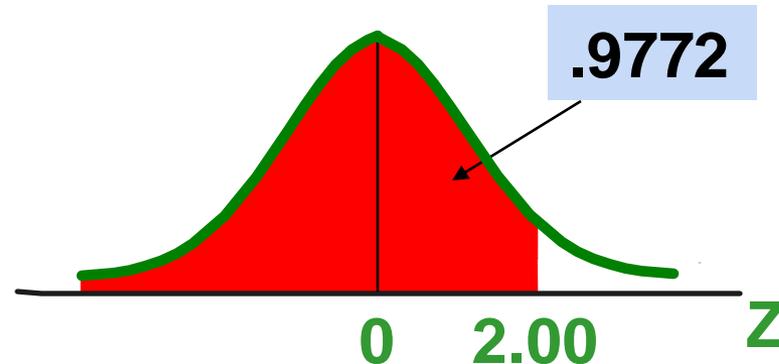
The Standardized Normal Table

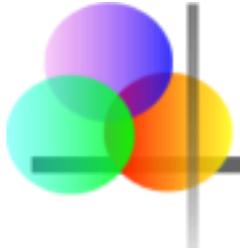
- **Standardized Normal Table** gives the probability $F(a)$ for any value a

$$\Phi(z) = \Phi(2) = 0.9772$$

Example:

$$P(Z < 2.00) = .9772$$





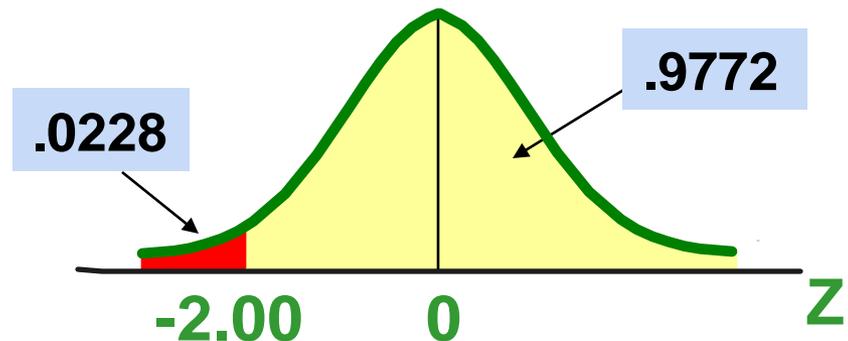
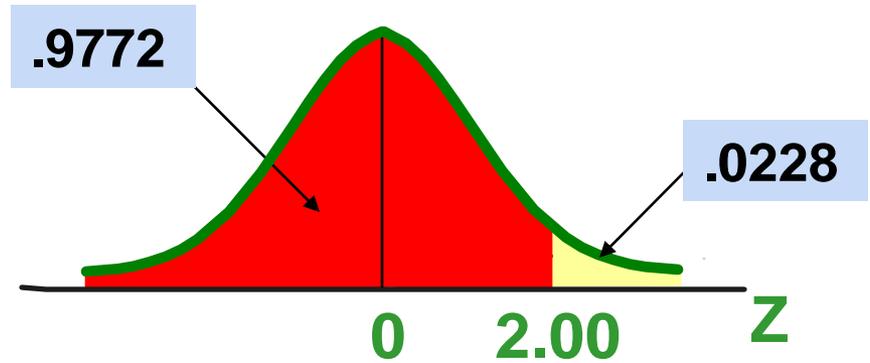
The Standardized Normal Table

(continued)

- For **negative Z-values**, use the fact that the distribution is symmetric to find the needed probability:

Example:

$$\begin{aligned} P(Z < -2.00) &= 1 - 0.9772 \\ &= 0.0228 \end{aligned}$$



General Procedure for Finding Probabilities

To find $P(a < X < b)$ when X is distributed normally:

1

Draw the normal curve for the problem in

2

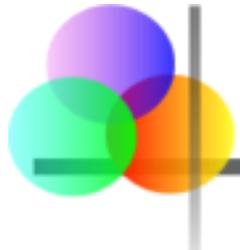
terms of X

3

Translate X -values to Z -values

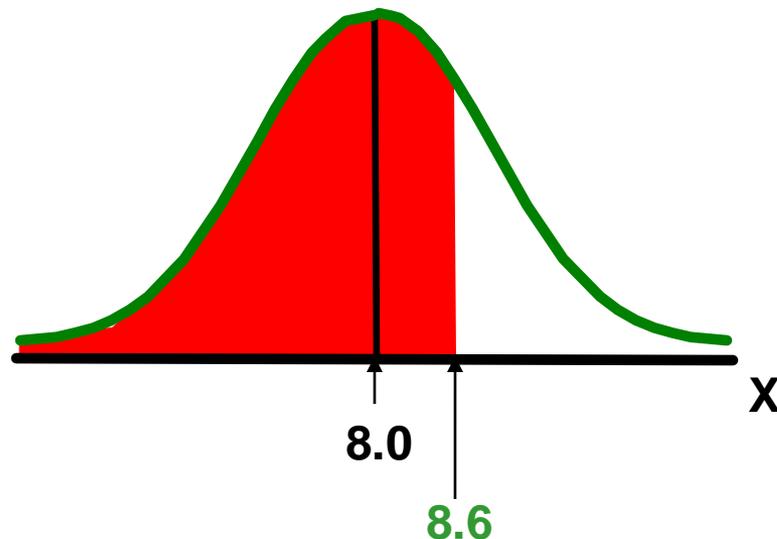


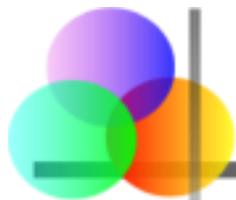
Use the Cumulative Normal Table



Finding Normal Probabilities

- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find $P(X < 8.6)$



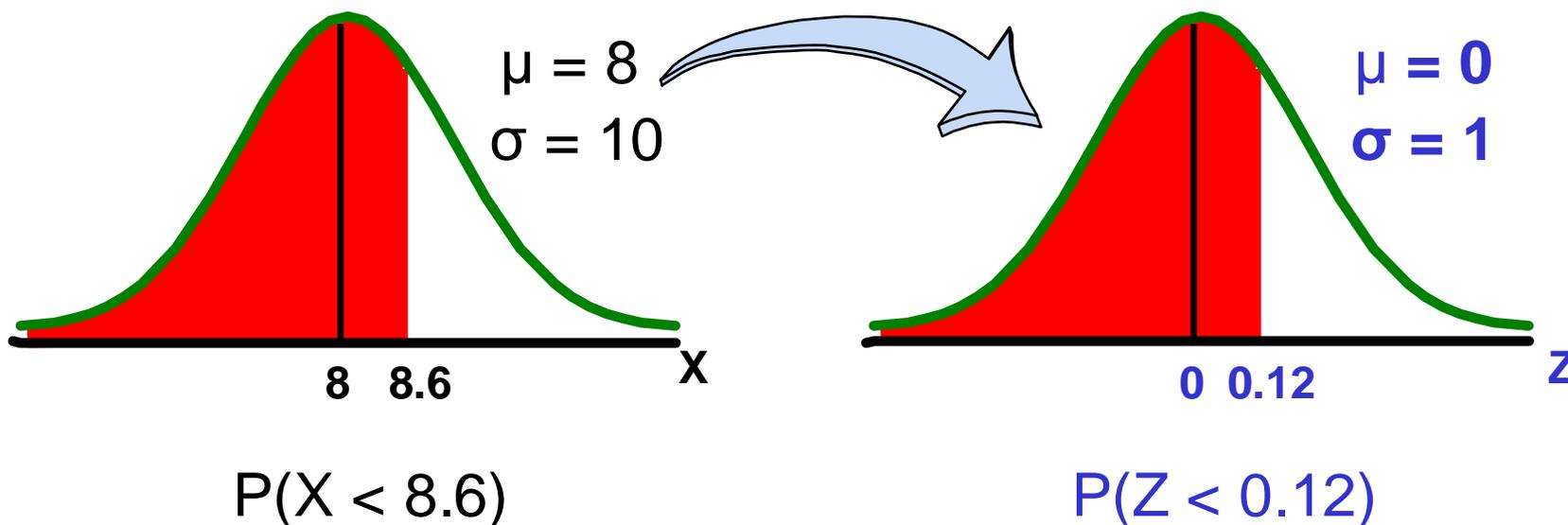


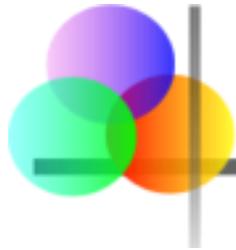
Finding Normal Probabilities

(continued)

- Suppose X is normal with mean 8.0 and standard deviation 5.0. Find $P(X < 8.6)$

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$



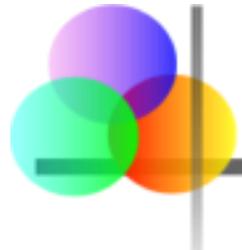


Finding Normal Probabilities

(continued)



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177

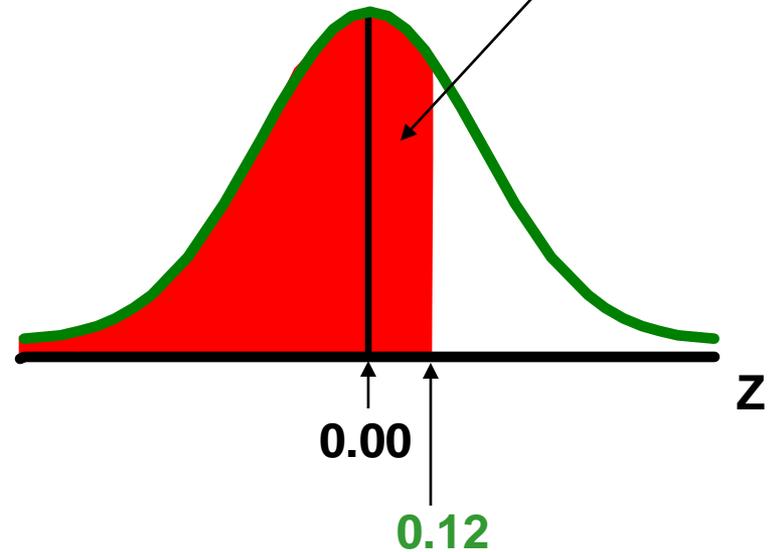


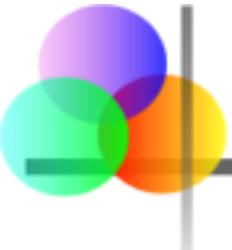
Solution: Finding $P(Z < 0.12)$

Standardized Normal Probability Table (Portion)

z	F(z)
.10	.5398
.11	.5438
.12	.5478
.13	.5517

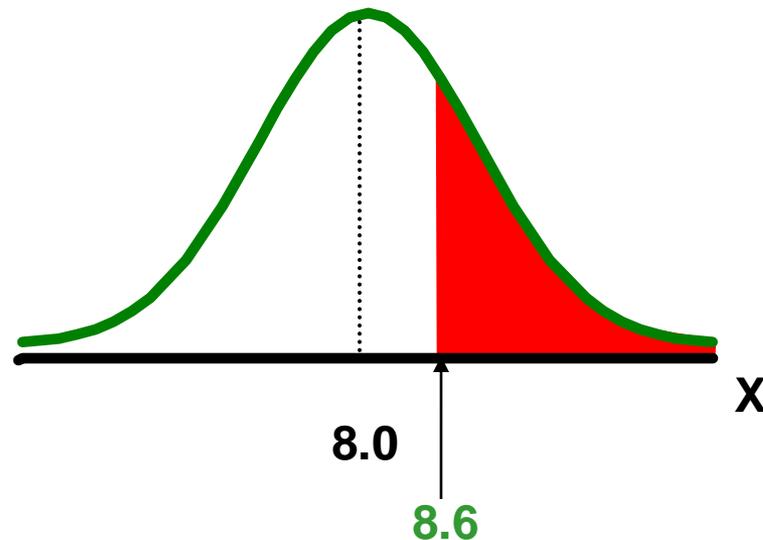
$$\begin{aligned} P(X < 8.6) \\ &= P(Z < 0.12) \\ &F(0.12) = 0.5478 \end{aligned}$$





Upper Tail Probabilities

- Suppose X is normal with mean 8.0 and standard deviation 5.0.
- Now Find $P(X > 8.6)$



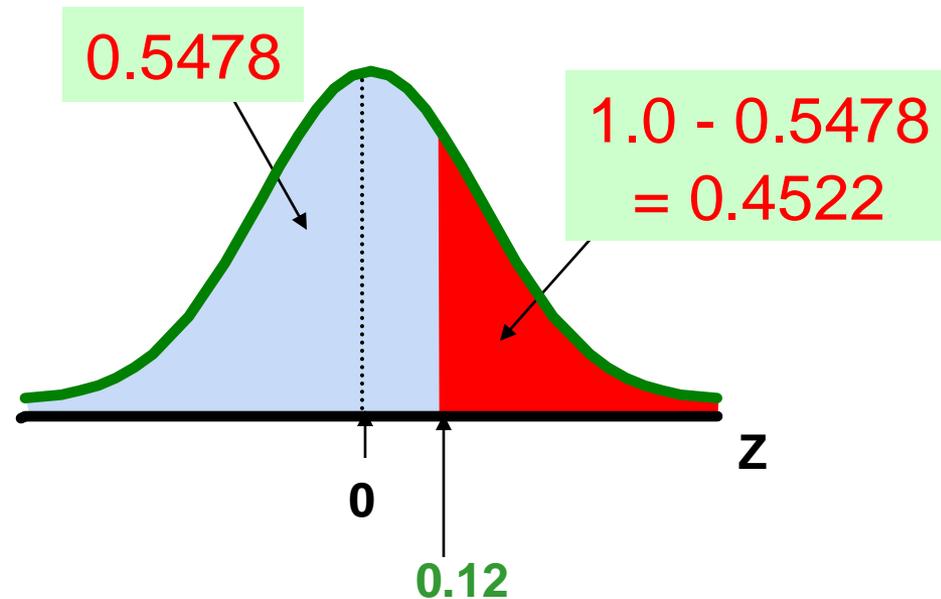
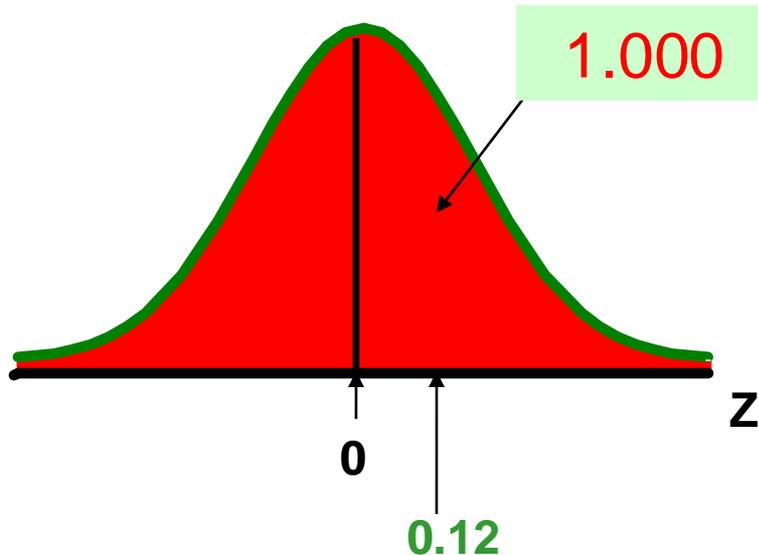
Upper Tail Probabilities

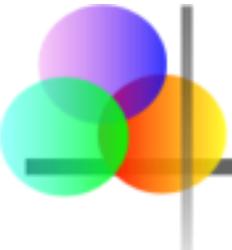
(continued)

- Now Find $P(X > 8.6)$...

$$P(X > 8.6) = P(Z > 0.12) = 1.0 - P(Z \leq 0.12)$$

$$= 1.0 - 0.5478 = 0.4522$$





Finding the X value for a Known Probability

- Steps to find the X value for a known probability:
 1. Find the Z value for the known probability
 2. Convert to X units using the formula:

$$X = \mu + Z\sigma$$

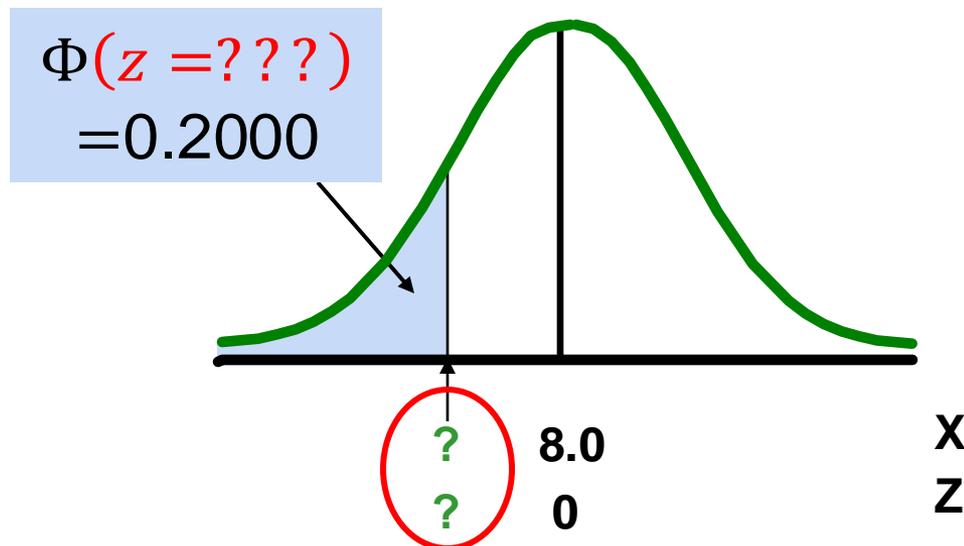
**Inverted problem...
from probability to z**

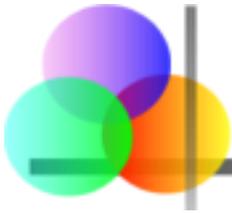
Finding the X value for a Known Probability

(continued)

Example:

- Suppose X is normal with mean 8.0 and standard deviation 5.0.
- Now find the X value so that only 20% of all values are below this X





Finding the X value for a Known Probability

(continued)

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177

From the table we can observe values of $\Phi(z) \geq 0.5$.

Simmety property of the normal r.v. $\Phi(-z) = 1 - \Phi(z)$

$$\Phi(-z = 0.20) = 1 - \Phi(z \approx 0.80)$$

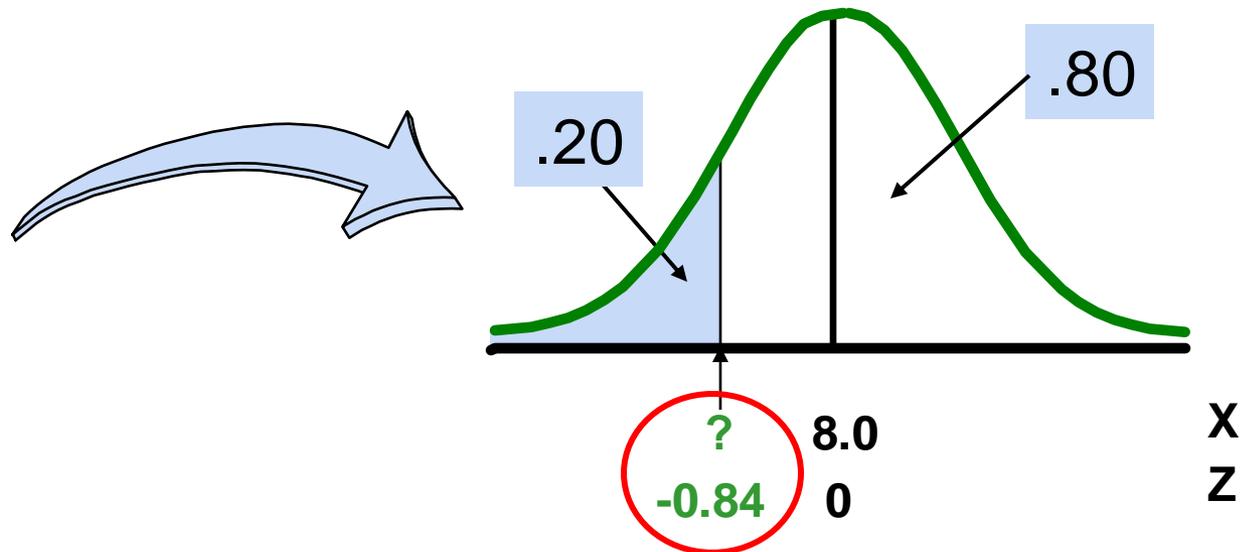
Find the Z value for 20% in the Lower Tail

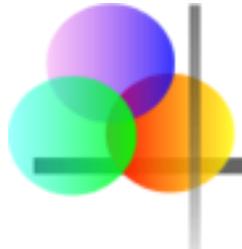
1. Find the Z value for the known probability

Standardized Normal Probability Table (Portion)

z	F(z)
.82	.7939
.83	.7967
.84	.7995
.85	.8023

- 20% area in the lower tail is consistent with a Z value of **-0.84**



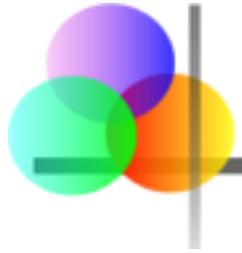


Finding the X value

2. Convert to X units using the formula:

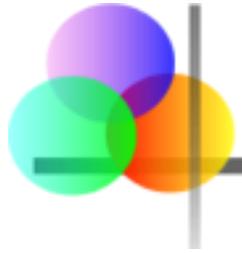
$$\begin{aligned} X &= \mu + Z\sigma \\ &= 8.0 + (-0.84)5.0 \\ &= 3.80 \end{aligned}$$

So 20% of the values from a distribution with mean 8.0 and standard deviation 5.0 are less than **3.80**



Assessing Normality

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



The Normal Probability Plot

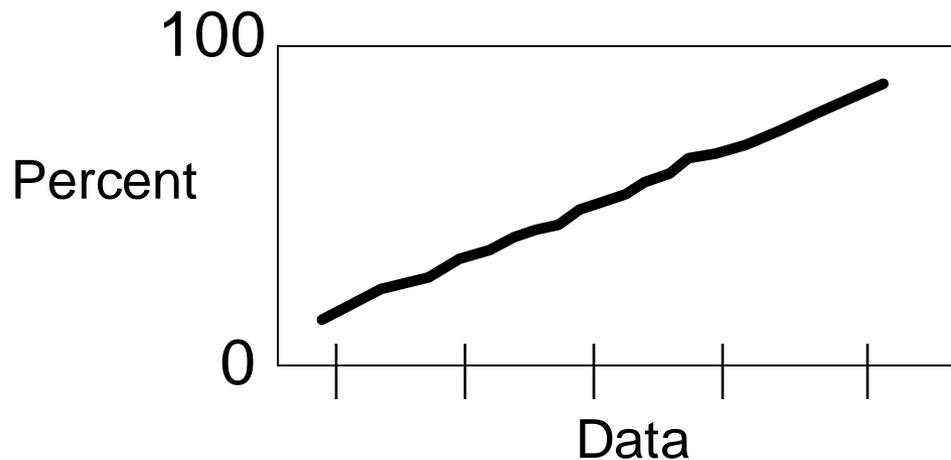
- Normal probability plot
 - Arrange data from low to high values
 - Find cumulative normal probabilities for all values
 - Examine a plot of the observed values vs. cumulative probabilities (with the cumulative normal probability on the vertical axis and the observed data values on the horizontal axis)
 - Evaluate the plot for evidence of linearity

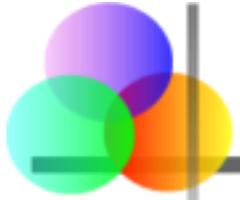


The Normal Probability Plot

(continued)

A normal probability plot for data from a normal distribution will be approximately linear:

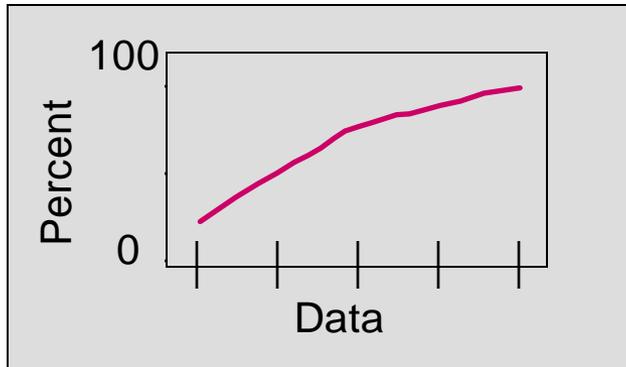




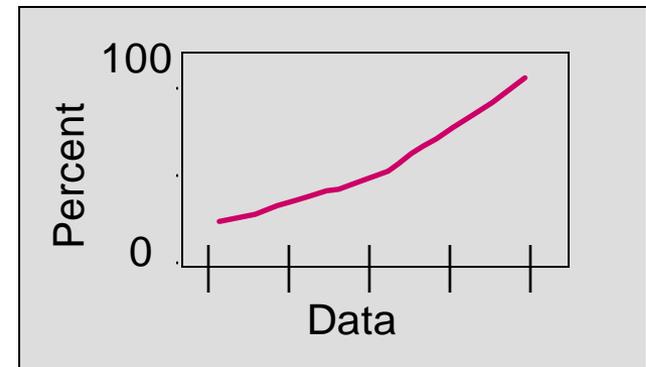
The Normal Probability Plot

(continued)

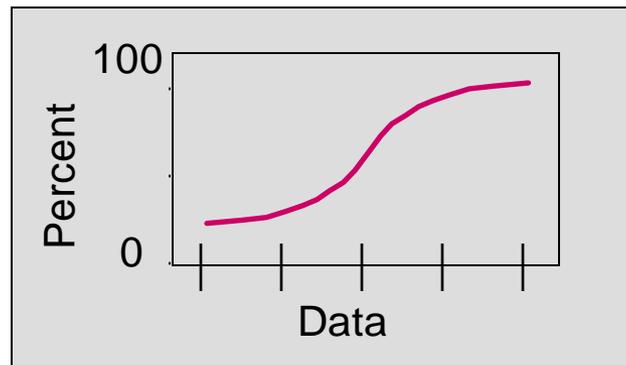
Left-Skewed



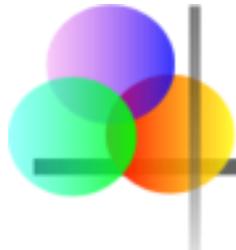
Right-Skewed



Uniform



Nonlinear plots indicate a deviation from normality

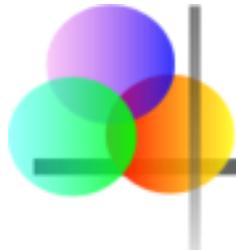


Normal Distribution Approximation for Binomial Distribution

- Recall the binomial distribution:
 - n independent trials
 - probability of success on any given trial = P
- Random variable X:
 - $X_i = 1$ if the i^{th} trial is “success”
 - $X_i = 0$ if the i^{th} trial is “failure”

$$E(X) = \mu = nP$$

$$\text{Var}(X) = \sigma^2 = nP(1-P)$$

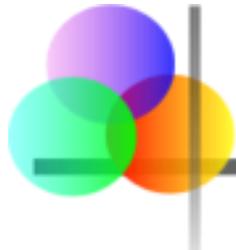


Normal Distribution Approximation for Binomial Distribution

(continued)

- The shape of the binomial distribution is **approximately normal** if n is large
- The normal is a good approximation to the binomial when $nP(1 - P) > 9$
- Standardize to Z from a binomial distribution:

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} = \frac{X - np}{\sqrt{nP(1-P)}}$$

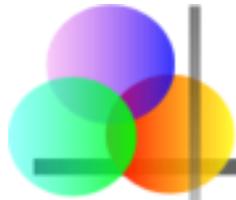


Normal Distribution Approximation for Binomial Distribution

(continued)

- Let X be the number of successes from n independent trials, each with probability of success P .
- If $nP(1 - P) > 9$,

$$P(a < X < b) = P\left(\frac{a - nP}{\sqrt{nP(1-P)}} \leq Z \leq \frac{b - nP}{\sqrt{nP(1-P)}}\right)$$



Binomial Approximation Example

- 40% of all voters support ballot proposition A. What is the probability that between 76 and 80 voters indicate support in a sample of $n = 200$?
 - $E(X) = \mu = nP = 200(0.40) = 80$
 - $\text{Var}(X) = \sigma^2 = nP(1 - P) = 200(0.40)(1 - 0.40) = 48$
(note: $nP(1 - P) = 48 > 9$)

$$\begin{aligned} P(76 < X < 80) &= P\left(\frac{76 - 80}{\sqrt{200(0.4)(1-0.4)}} \leq Z \leq \frac{80 - 80}{\sqrt{200(0.4)(1-0.4)}}\right) \\ &= P(-0.58 < Z < 0) \\ &= F(0) - F(-0.58) \\ &= 0.5000 - 0.2810 = 0.2190 \end{aligned}$$