A Review of Basic Statistical Concepts

DS 632: System Simulation, Summer I 2014

Professor C. Chandra
College of Business
University of Michigan-Dearborn

Basic Vocabulary of Statistics

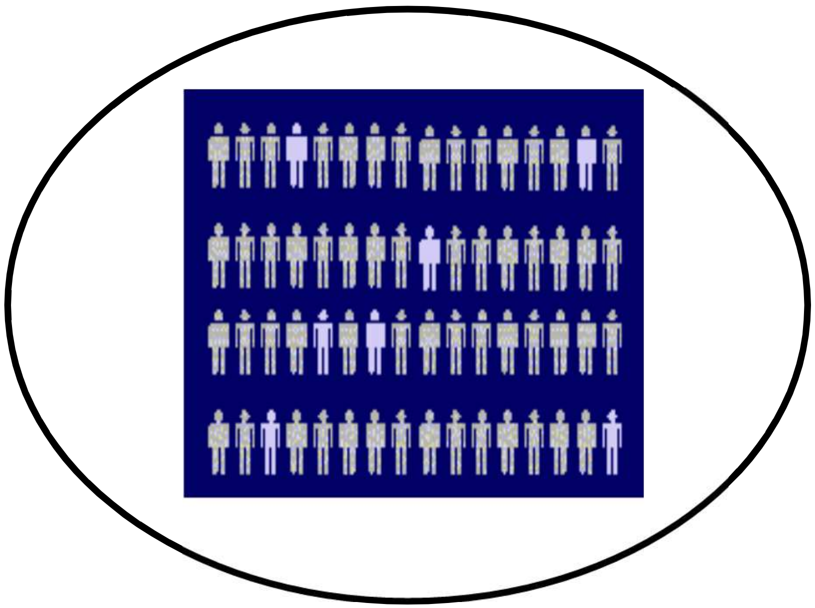
- **VARIABLE**
  - A **variable** is a characteristic of an item or individual.
- **DATA**
  - **Data** are the different values associated with a variable.
- **OPERATIONAL DEFINITIONS**
  - Variable values are meaningless unless their variables have **operational definitions,** universally accepted meanings that are clear to all associated with an analysis.

Basic Vocabulary of Statistics

- **POPULATION**
  - A **population** consists of all the items or individuals about which you want to draw a conclusion.
- **SAMPLE**
  - A **sample** is the portion of a population selected for analysis.
- **PARAMETER**
  - A **parameter** is a numerical measure that describes a characteristic of a population.
- **STATISTIC**
  - A **statistic** is a numerical measure that describes a characteristic of a sample.

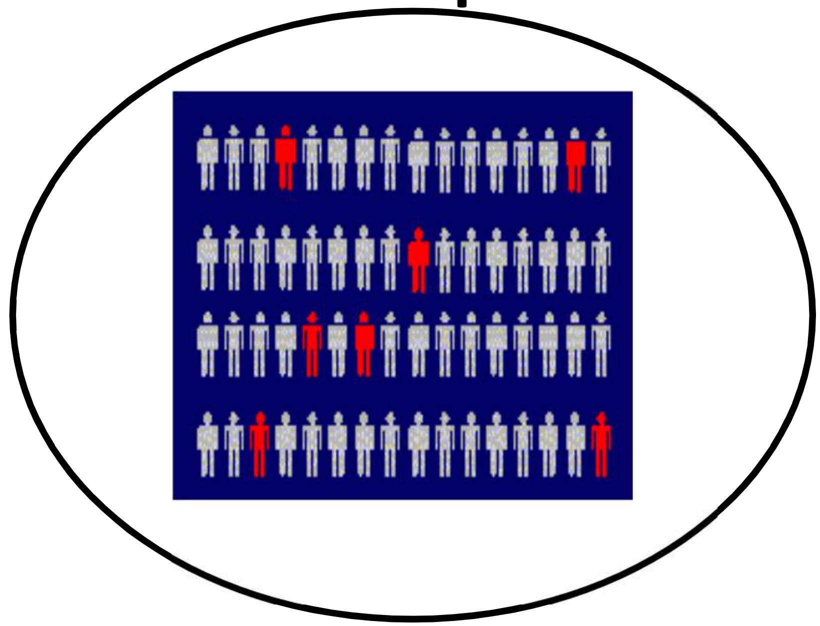
Population vs. Sample

Population



Measures used to describe the population are called **parameters**

Sample



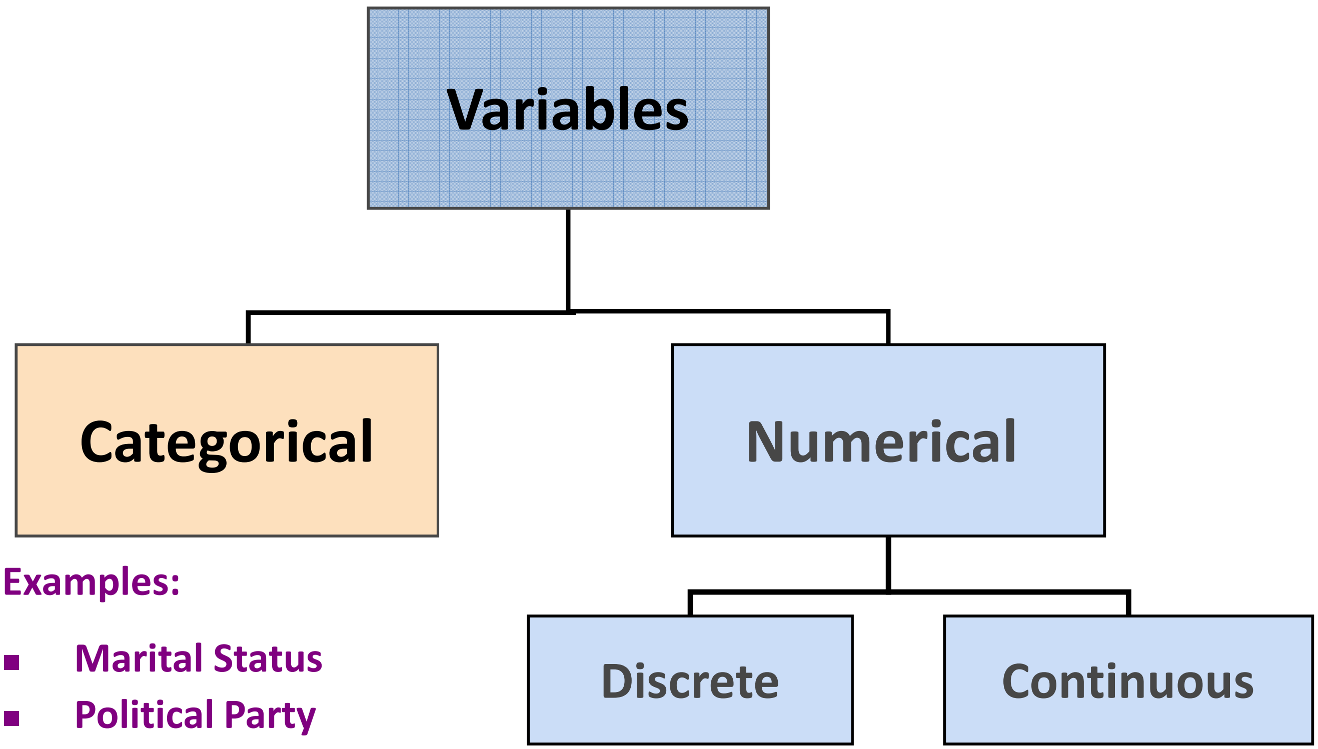
Measures computed from sample data are called **statistics**

Organizing and Visualizing Data

Types of Variables

- **Categorical** (***qualitative***) variables have values that can only be placed into categories, such as “yes” and “no”
- **Numerical** (***quantitative***) variables have values that represent quantities.
  - **Discrete** variables arise from a ***counting process***
  - **Continuous** variables arise from a ***measuring process***

Types of Variables



Examples:

- **Marital Status**
- **Political Party**
- **Eye Color**

(Defined categories)

Examples:

- **Number of Children**
- **Defects per hour**

(Counted items)

Examples:

- **Weight**
- **Voltage**

(Measured characteristics)

Basic Probability

Basic Probability Concepts

- **Event:** each possible outcome of a variable is an event.
- **Probability**: the chance, likelihood, or possibility that a particular event will occur (always between 0 and 1).
- **Impossible Event:** an event that has no chance of occurring (probability = 0).
- **Certain Event:** an event that is sure to occur (probability = 1).

Assessing Probability

Three approaches to assessing the probability of an uncertain event:

1. ***a priori* probability**: based on prior knowledge of the process involved.

2. **empirical probability**: based on observed data.

3. **subjective probability**: based on a combination of an individual's past experience, personal opinion, and analysis of a particular situation

Calculating Probability

1. ***a priori* probability**

$$\text{Probability of Occurrence} = \frac{X}{T} = \frac{\text{number of ways the event can occur}}{\text{total number of possible outcomes}}$$

2. **empirical probability**

$$\text{Probability of Occurrence} = \frac{\text{number of favorable outcomes observed}}{\text{total number of outcomes observed}}$$

These equations assume all outcomes are equally likely.

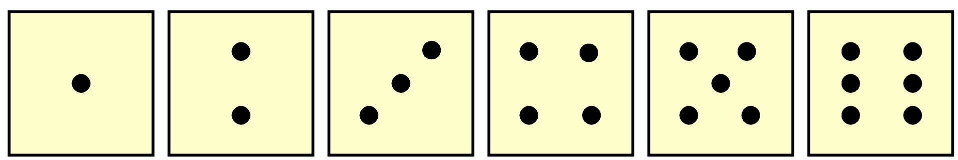
Events

- Simple event
  - An event described by a single characteristic
  - e.g., A red card from a deck of cards
- Joint event
  - An event described by two or more characteristics
  - e.g., An ace that is also red from a deck of cards
- Complement of an event A (denoted A')
  - All events that are not part of event A
  - e.g., All cards that are not diamonds

Sample Space

The Sample Space is the collection of all possible events

ex. All 6 faces of a die:



ex. All 52 cards in a deck of cards

Independence

- Two events are **independent** if and only if:

$$P(A \mid B) = P(A)$$

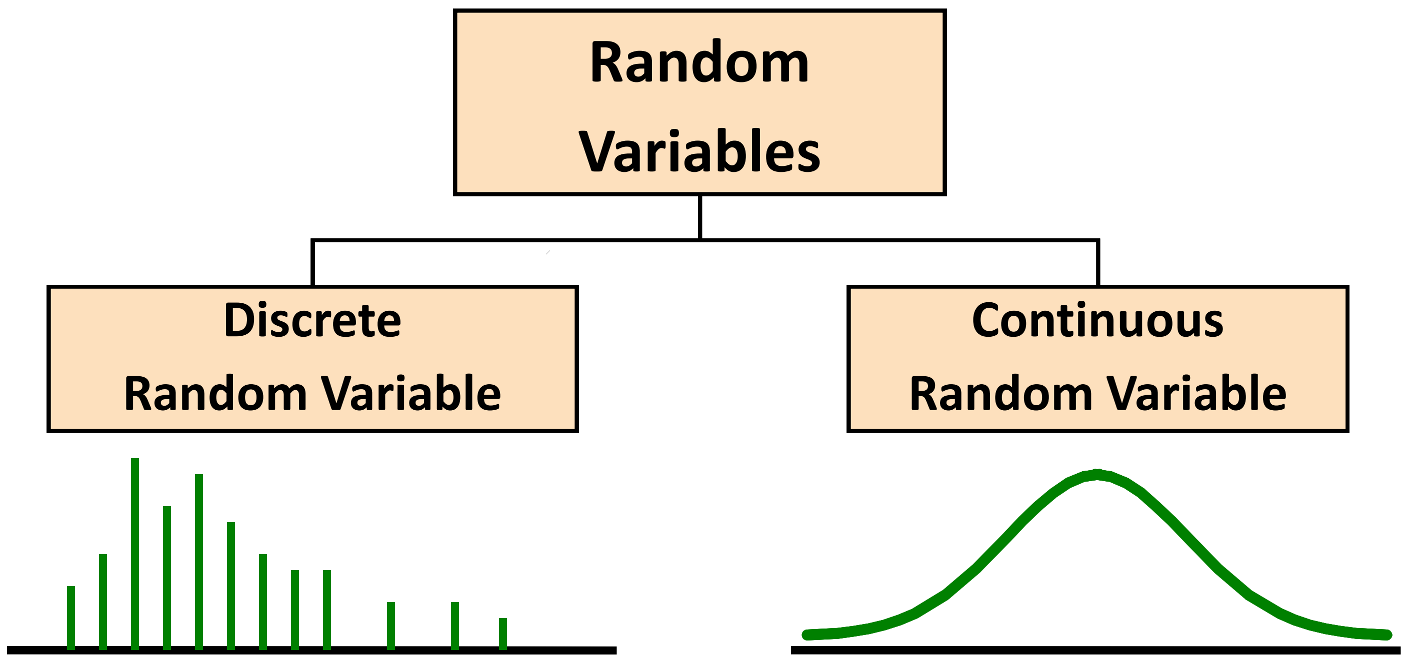
- Events A and B are independent when the probability of one event is not affected by the fact that the other event has occurred

Discrete Probability Distributions

Definitions: Random Variables

- A **random variable** represents a possible numerical value from an uncertain event.
- **Discrete** random variables produce outcomes that come from a counting process (i.e. number of classes you are taking).
- **Continuous** random variables produce outcomes that come from a measurement (i.e. your weight).

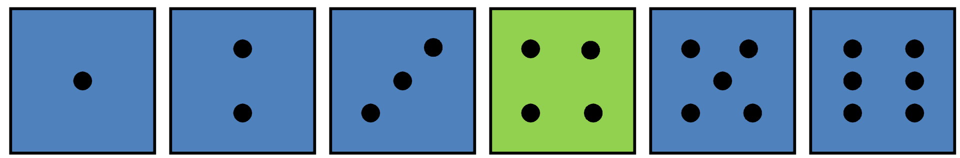
Definitions: Random Variables



Discrete Random Variables

Can only assume a countable number of values

- Examples:
  - Roll a die twice



    - Let X be the number of times 4 comes up (then X could be 0, 1, or 2 times)

  - Toss a coin 5 times.
    - Let X be the number of heads (then X = 0, 1, 2, 3, 4, or 5)

Probability Distribution For A Discrete Random Variable

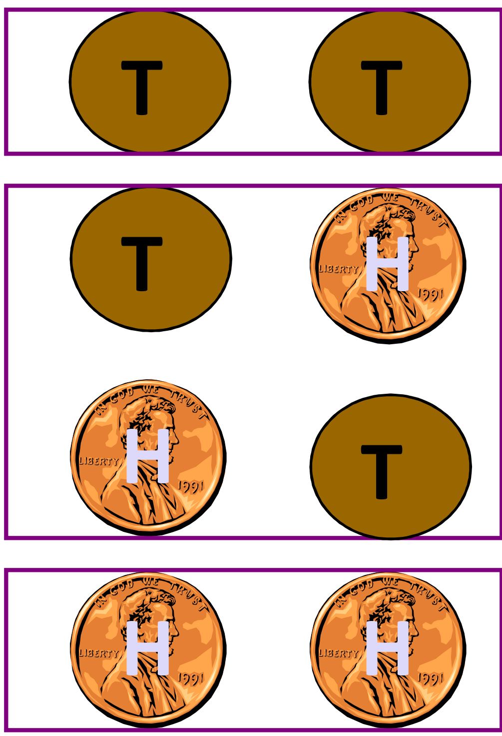
- A **probability distribution for a discrete random variable** is a mutually exclusive listing of all possible numerical outcomes for that variable and a particular probability of occurrence associated with each outcome.

Number of Classes Taken	Probability
2	0.2
3	0.4
4	0.24
5	0.16

Example of a Discrete Random Variable Probability Distribution

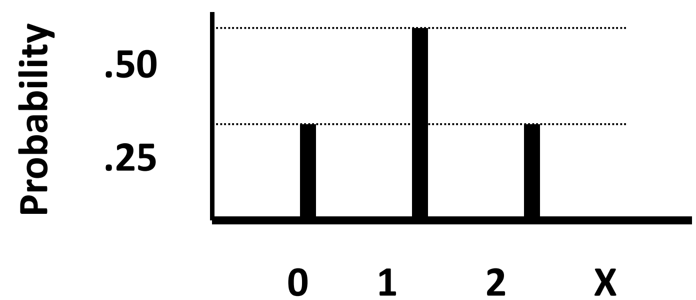
Experiment: Toss 2 Coins. Let X = # heads.

4 possible outcomes



Probability Distribution

X Value	**Probability**
0	**1/4 = .25**
1	**2/4 = .50**
2	**1/4 = .25**



Discrete Random Variables: Expected Value

- Expected Value (or mean) of a discrete distribution (Weighted Average)

$$\mu = \mathrm{E(X)} = \sum_{i=1}^{N} X_i P(X_i)$$

- Example: Toss 2 coins, X = # of heads,

  Compute expected value of X:

  E(X) = (0)(.25) + (1)(.50) + (2)(.25)

  = 1.0

X Value	Probability
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25

Discrete Random Variables: Measuring Dispersion

- Variance of a discrete random variable

$$\sigma^2 = \sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i)$$

- Standard Deviation of a discrete random variable

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i)}$$

where:

E(X) = Expected value of the discrete random variable X

X_i = the ith outcome of X

$P(X_i)$ = Probability of the ith occurrence of X

Discrete Random Variables: Measuring Dispersion

- Example: Toss 2 coins, X = # heads, compute standard deviation (recall that E(X) = 1)

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N}[X_i - E(X)]^2 P(X_i)}$$

$$\sigma = \sqrt{(0-1)^2(.25) + (1-1)^2(.50) + (2-1)^2(.25)} = \sqrt{.50} = .707$$

Possible number of heads = 0, 1, or 2

Covariance

- The covariance measures the strength of the linear relationship between two numerical random variables X and Y.
- A positive covariance indicates a positive relationship.
- A negative covariance indicates a negative relationship.

The Covariance Formula

- Covariance formula:

$$\sigma_{XY} = \sum_{i=1}^{N} [X_i - E(X)][(Y_i - E(Y)]\, P(X_i Y_i)$$

where: X = discrete variable X
X_i = the i^{th} outcome of X
Y = discrete variable Y
Y_i = the i^{th} outcome of Y
$P(X_i Y_i)$ = probability of occurrence of the condition affecting the i^{th} outcome of X and the i^{th} outcome of Y

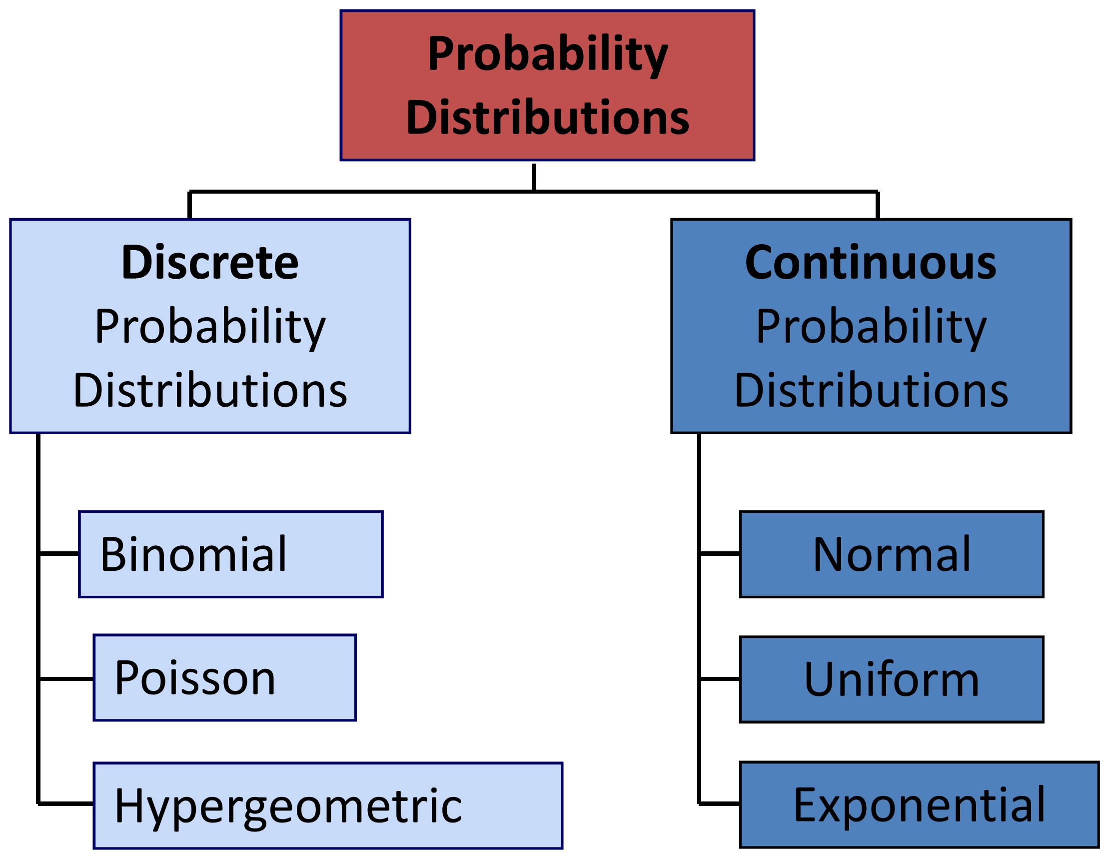
The Sum of Two Random Variables

- Expected Value: $E(X+Y) = E(X) + E(Y)$

- Variance: $\mathrm{Var}(X+Y) = \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$

- Standard deviation: $\sigma_{X+Y} = \sqrt{\sigma_{X+Y}^2}$

Probability Distribution Overview



Binomial Distribution

- A fixed number of observations, n
  - e.g., 15 tosses of a coin
- Each observation is categorized as to whether or not the "event of interest" occurred
  - e.g., head or tail in each toss of a coin
  - these two categories are mutually exclusive and collectively exhaustive
    - when the probability of the event of interest is represented as π, then the probability of the event of interest not occurring is $1 - \pi$

Binomial Distribution *(continued)*

- Constant probability for the event of interest occurring (π) for each observation
  - probability of getting a tail is the same each time we toss the coin
- Observations are independent
  - The outcome of one observation does not affect the outcome of the other
  - Two sampling methods deliver independence
    - Infinite population without replacement
    - Finite population with replacement

Possible Applications of the Binomial Distribution

- A manufacturing plant labels items as either defective or acceptable
- A firm bidding for contracts will either get a contract or not
- A marketing research firm receives survey responses of “yes I will buy” or “no I will not”
- New job applicants either accept the offer or reject it

The Binomial Distribution Counting Techniques

- Suppose success is defined as flipping heads at least two times out of three with a fair coin. How many ways is success possible?
- Possible Successes: HHT, HTH, THH, HHH, So, there are four possible ways.
- This situation is extremely simple. We need a way of counting successes for more complicated and less trivial situations.

Counting Techniques
Rule of Combinations

- The number of combinations of selecting X objects out of n objects is:

Excel Formula:
=COMBIN(n,X)

$$ {}_{n}C_{X} = \binom{n}{X} = \frac{n!}{X!(n-X)!} $$

Excel formula:
=FACT(n)

where:

$n! = n(n - 1)(n - 2) \ldots (2)(1)$

$X! = X(X - 1)(X - 2) \ldots (2)(1)$

$0! = 1$ (by definition)

Counting Techniques
Rule of Combinations

- How many possible 3 scoop combinations could you create at an ice cream parlor if you have 31 flavors to select from?
- The total choices is n = 31, and we select X = 3.

$$ {}_{31}C_3 = \binom{31}{3} = \frac{31!}{3!(31-3)!} = \frac{31!}{3!28!} = \frac{31 \bullet 30 \bullet 29 \bullet 28!}{3 \bullet 2 \bullet 1 \bullet 28!} = 31 \bullet 5 \bullet 29 = 4495 $$

Binomial Distribution Formula

$$P(X) = \frac{n!}{X!(n-X)!}\pi^{X}(1-\pi)^{n-X}$$

P(X) = probability of **X** successes in **n** trials, with probability of success **π** on each trial

X = number of 'successes' in sample, (X = 0, 1, 2, ..., n)

n = sample size (number of trials or observations)

π = probability of "success"

Example: Flip a coin four times, let x = # heads:

n = 4

π = 0.5

1 - **π** = (1 - .5) = .5

X = 0, 1, 2, 3, 4

The Binomial Distribution Example 1

What is the probability of one success in five observations if the probability of success is .1?

$X = 1$, $n = 5$, and $\pi = .1$

$$P(X=1) = \frac{n!}{X!(n-X)!}\pi^X(1-\pi)^{n-X}$$

$$= \frac{5!}{1!(5-1)!}(.1)^1(1-.1)^{5-1}$$

$$= (5)(.1)(.9)^4$$

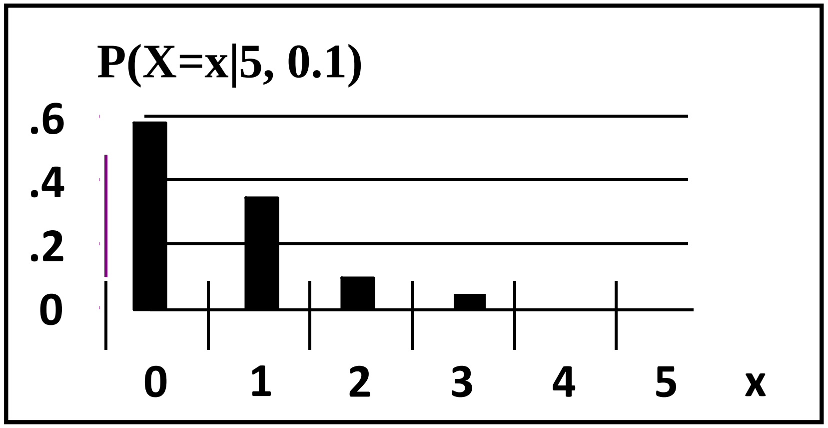
$$= .32805$$

=BINOMDIST(1,5,0.1,FALSE)

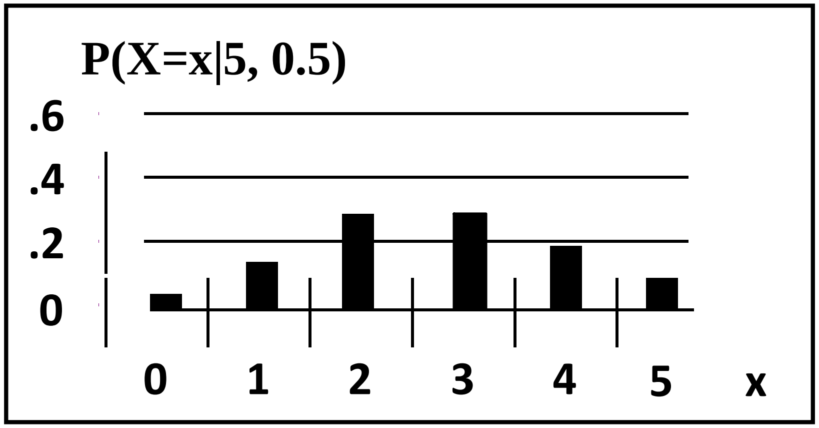
The Binomial Distribution Shape

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

- Here, n = 5 and π = .1



- Here, n = 5 and π = .5



The Binomial Distribution Using Binomial Tables

n = 10									
x	…	p=.20	p=.25	p=.30	**p=.35**	p=.40	p=.45	p=.50	
0	…	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010	10
1	…	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098	9
2	…	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439	8
3	…	0.2013	0.2503	0.2668	**0.2522**	0.2150	0.1665	0.1172	7
4	…	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051	6
5	…	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461	5
6	…	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051	4
7	…	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172	3
8	…	0.0001	**0.0004**	0.0014	0.0043	0.0106	0.0229	0.0439	**2**
9	…	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098	1
10	…	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0
	…	p=.80	**p=.75**	p=.70	p=.65	p=.60	p=.55	p=.50	**x**

Examples:

n = 10, π = .35, x = 3: P(x = 3|n =10, π = .35) = .2522

n = 10, π = .75, x = 2: P(x = 2|n =10, π = .75) = .0004

The Binomial Distribution Characteristics

- Mean

$$\mu = E(x) = n\pi$$

- Variance and Standard Deviation

$$\sigma^2 = n\pi(1-\pi) \qquad \sigma = \sqrt{n\pi(1-\pi)}$$

Where n = sample size

π = probability of success

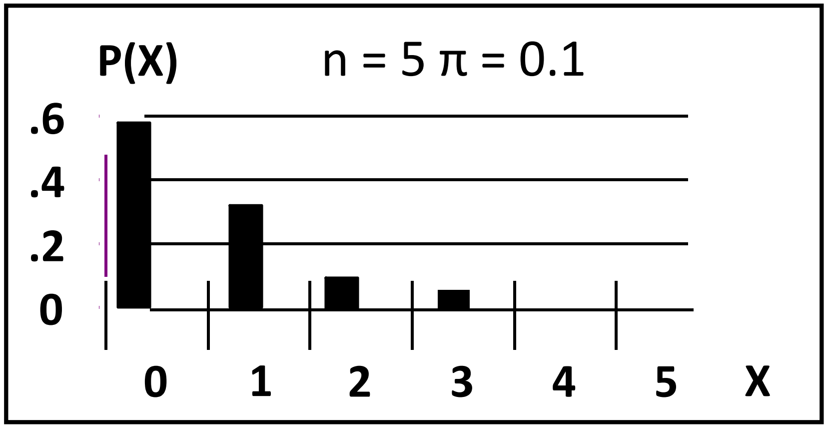
$(1 - \pi)$ = probability of failure

The Binomial Distribution Characteristics

Examples

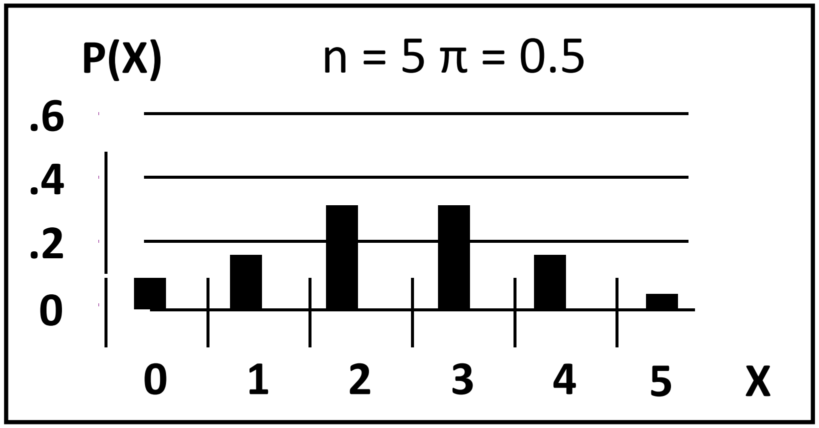
$$\mu = n\pi = (5)(.1) = 0.5$$

$$\sigma = n\pi(1-\pi) = \sqrt{(5)(.1)(1-.1)}$$
$$= 0.6708$$



$$\mu = n\pi = (5)(.5) = 2.5$$

$$\sigma = n\pi(1-\pi) = \sqrt{(5)(.5)(1-.5)}$$
$$= 1.118$$



The Poisson Distribution: Definitions

- You use the Poisson distribution when you are interested in the number of times an event occurs in a given **area of opportunity (AO) or domain**.
- **An area of opportunity** is a continuous unit or interval of time, volume, or such area in which more than one occurrence of an event can occur.
  - The number of scratches in a car's paint
  - The number of mosquito bites on a person
  - The number of computer crashes in a day

The Poisson Distribution: When to Apply

Apply the Poisson Distribution when:

- You wish to count the number of times an event occurs in a given area of opportunity
- The probability that an event occurs in one AO is the same for all AOs
- The number of events that occur in one AO is independent of the number of events that occur in other AOs
- The probability that two or more events occur in an AO approaches zero as the AO becomes smaller
- The average number of events per unit is λ (lambda)

The Poisson Distribution: Formula

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

where:

X = number of events in an area of opportunity

λ = expected number of events

e = base of the natural logarithm system (2.71828...)

The Poisson Distribution Example

- Suppose that, on average, 5 cars enter a parking lot per minute. What is the probability that in a given minute, 7 cars will enter?
- So, X = 7 and $\lambda = 5$

$$P(7) = \frac{e^{-\lambda}\lambda^{x}}{X!} = \frac{e^{-5}5^{7}}{7!} = 0.104$$

- So, there is a 10.4% chance 7 cars will enter the parking in a given minute. =POISSON(7,5,FALSE)

Poisson Distribution Characteristics

- Mean

$$\boxed{\mu = \lambda}$$

- Variance and Standard Deviation

$$\boxed{\sigma^2 = \lambda}$$

$$\boxed{\sigma = \sqrt{\lambda}}$$

where λ = expected number of events

The Poisson Distribution
Using Poisson Tables

X	λ								
	0.10	0.20	0.30	0.40	**0.50**	0.60	0.70	0.80	0.90
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	**0.0758**	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Example: Find P(X = 2) if $\lambda = .50$

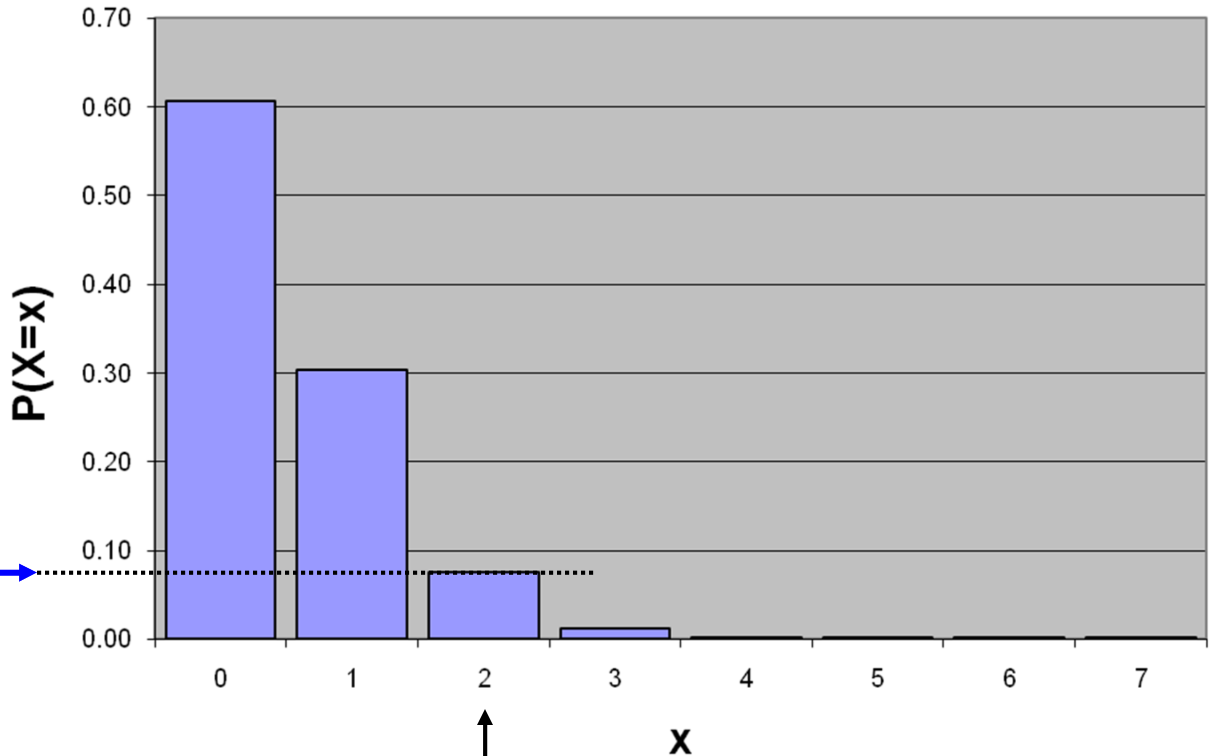
$$P(X=2) = \frac{e^{-\lambda}\lambda^{X}}{X!} = \frac{e^{-0.50}(0.50)^2}{2!} = .0758$$

Graph of Poisson Probabilities

Graphically:

λ = 0.50

X	λ = 0.50
0	0.6065
1	0.3033
2	0.0758
3	0.0126
4	0.0016
5	0.0002
6	0.0000
7	0.0000

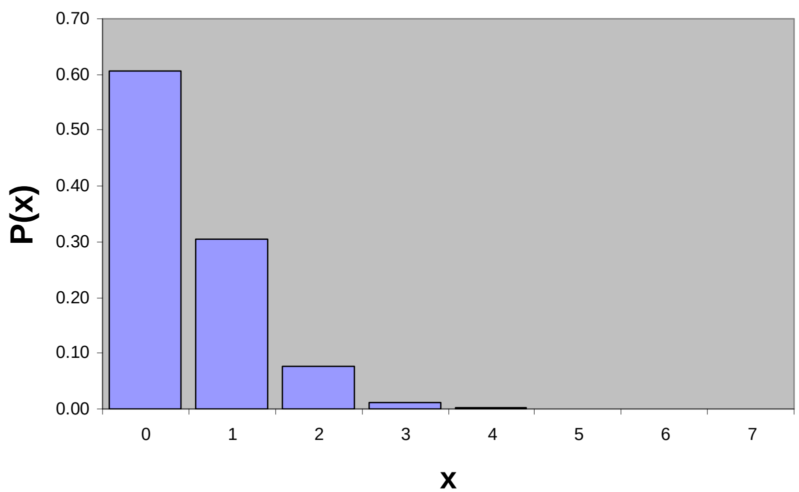


P(X = 2 | λ=0.50) = 0.0758

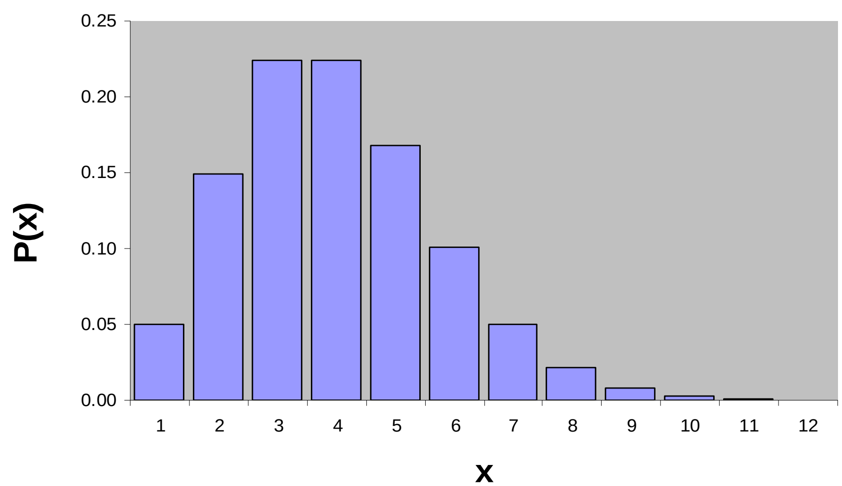
The Poisson Distribution Shape

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

$\lambda = 0.50$



$\lambda = 3.00$



The Hypergeometric Distribution

- The **binomial distribution** is applicable when selecting from a finite population with replacement or from an infinite population without replacement.
- The **hypergeometric distribution** is applicable when selecting from a finite population without replacement.

The Hypergeometric Distribution

- "n" trials in a sample taken from a **finite population** of size N
- Sample taken **without replacement**
- Outcomes of trials are **dependent**
- Concerned with finding the probability of "X" successes in the sample where there are "A" successes in the population

The Hypergeometric Distribution

$$P(X = x \mid n, N, A) = \frac{[{}_A C_X][{}_{N-A} C_{n-X}]}{{}_N C_n} = \frac{\binom{A}{X}\binom{N-A}{n-X}}{\binom{N}{n}}$$

Where

N = population size

A = number of successes in the population

N – A = number of failures in the population

n = sample size

X = number of successes in the sample

n – X = number of failures in the sample

The Hypergeometric Distribution Characteristics

- The mean of the hypergeometric distribution is:

$$\mu = E(x) = \frac{nA}{N}$$

- The standard deviation is:

$$\sigma = \sqrt{\frac{nA(N-A)}{N^2}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Where $\sqrt{\frac{N-n}{N-1}}$ is called the "Finite Population Correction Factor"

from sampling without replacement from a finite population

The Hypergeometric Distribution Example

- Different computers are checked from 10 in the department. 4 of the 10 computers have illegal software loaded. What is the probability that 2 of the 3 selected computers have illegal software loaded?
- So, N = 10, n = 3, A = 4, X = 2

$$P(X=2)=\frac{\binom{A}{X}\binom{N-A}{n-X}}{\binom{N}{n}}=\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}=\frac{(6)(6)}{120}=0.3$$

=HYPGEOMDIST(2,3,4,10)

- The probability that 2 of the 3 selected computers have illegal software loaded is .30, or 30%.

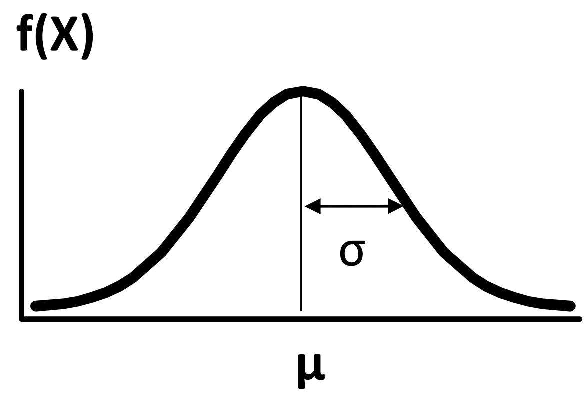
The Normal Distribution and Other Continuous Distributions

Continuous Probability Distributions

- A **continuous random variable** is a variable that can assume any value on a continuum (can assume an uncountable number of values)
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
- A **probability density function** is a mathematical expression that defines the distribution of the values of a continuous random variable.

The Normal Distribution Properties

- Bell Shaped
- Symmetrical
- Mean, Median and Mode are equal
- Location is characterized by the mean, μ
- Spread is characterized by the standard deviation, σ
- The random variable has an infinite theoretical range: $-\infty$ to $+\infty$



Mean
= Median
= Mode

The Normal Distribution Density Function

The formula for the normal probability density function (PDF) is

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

Where e = the mathematical constant approximated by 2.71828

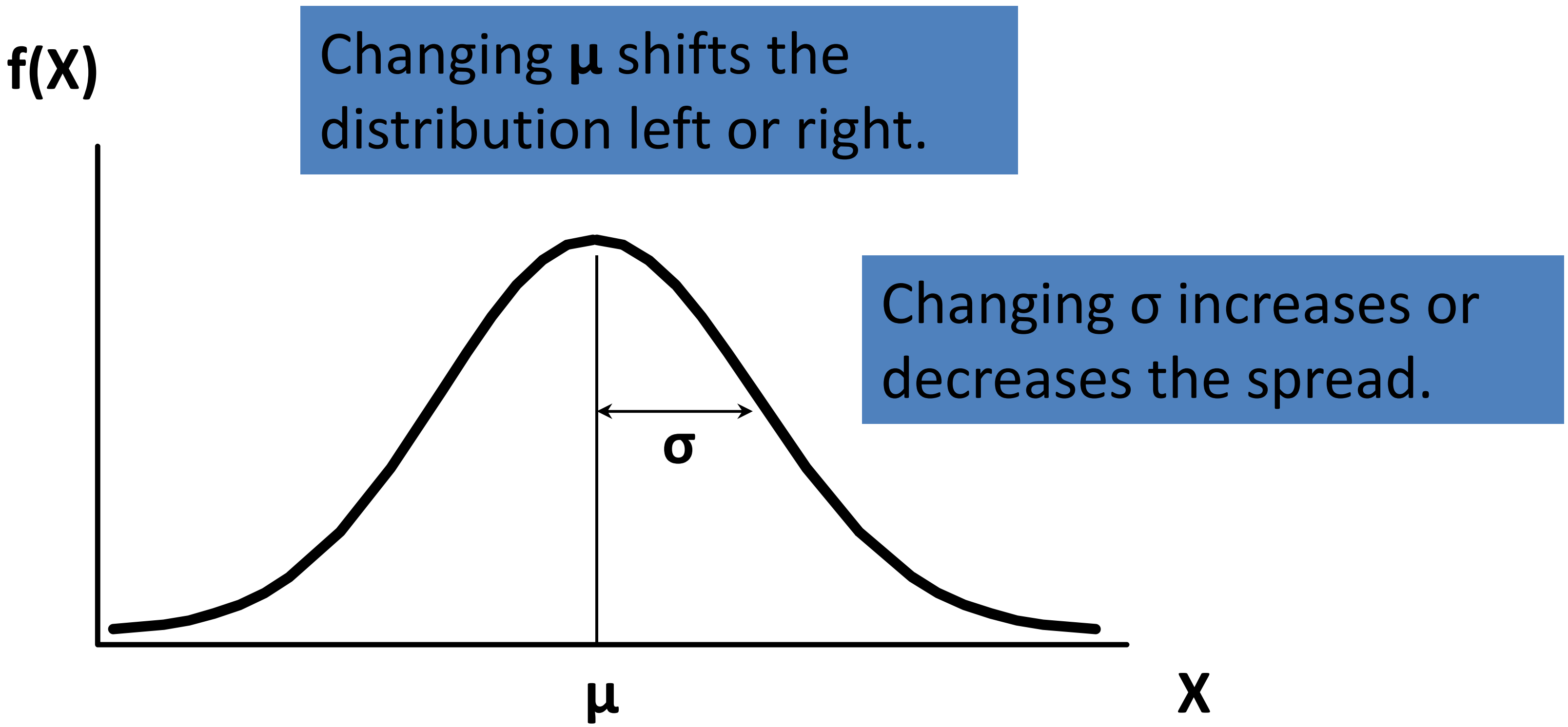
π = the mathematical constant approximated by 3.14159

μ = the population mean

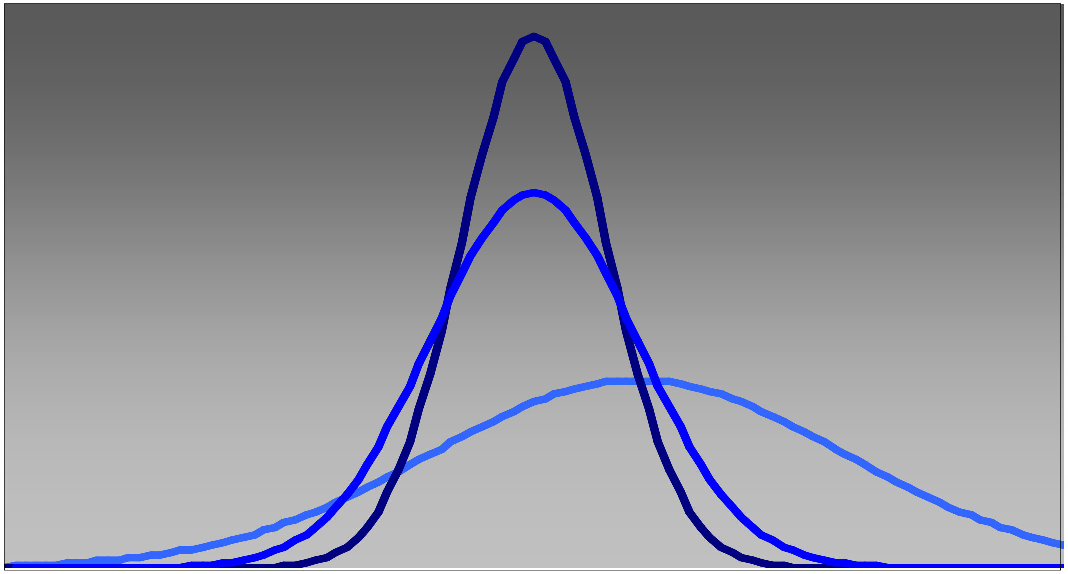
σ = the population standard deviation

X = any value of the continuous variable

The Normal Distribution Shape



The Normal Distribution
Different Shapes



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

The Standardized Normal Distribution

- 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.
- The standardized normal distribution has a mean of 0 and a standard deviation of 1.

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 Standardized Normal Distribution: Density Function

The formula for the standardized normal probability density function is

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

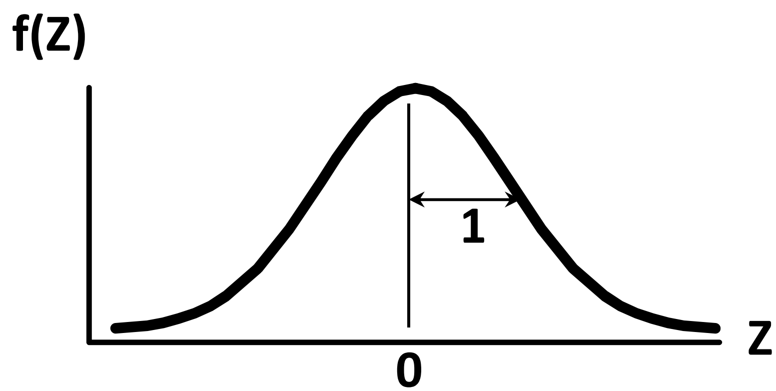
Where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

Z = any value of the standardized normal distribution

The Standardized Normal Distribution: Shape

- 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

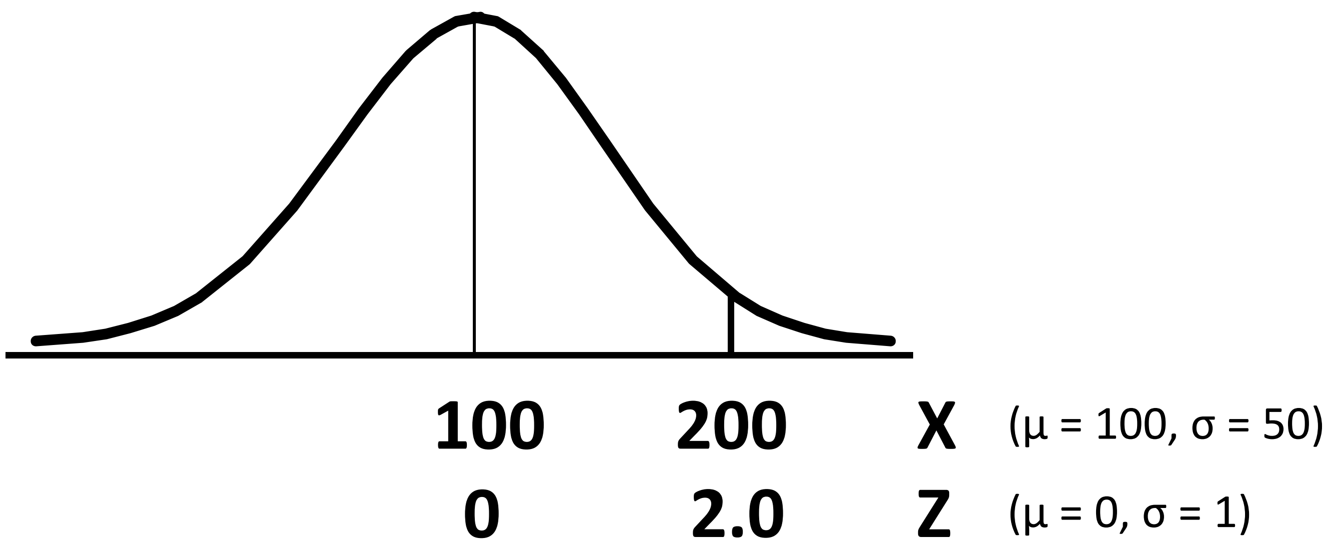
The Standardized Normal Distribution: Example

- 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.

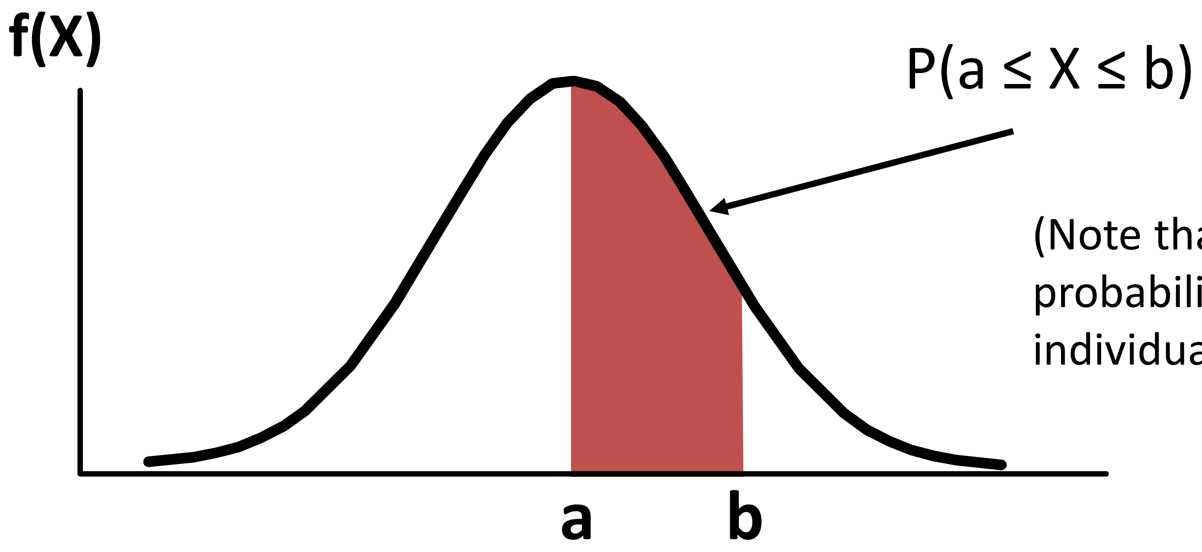
The Standardized Normal Distribution: Example



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)

Normal Probabilities

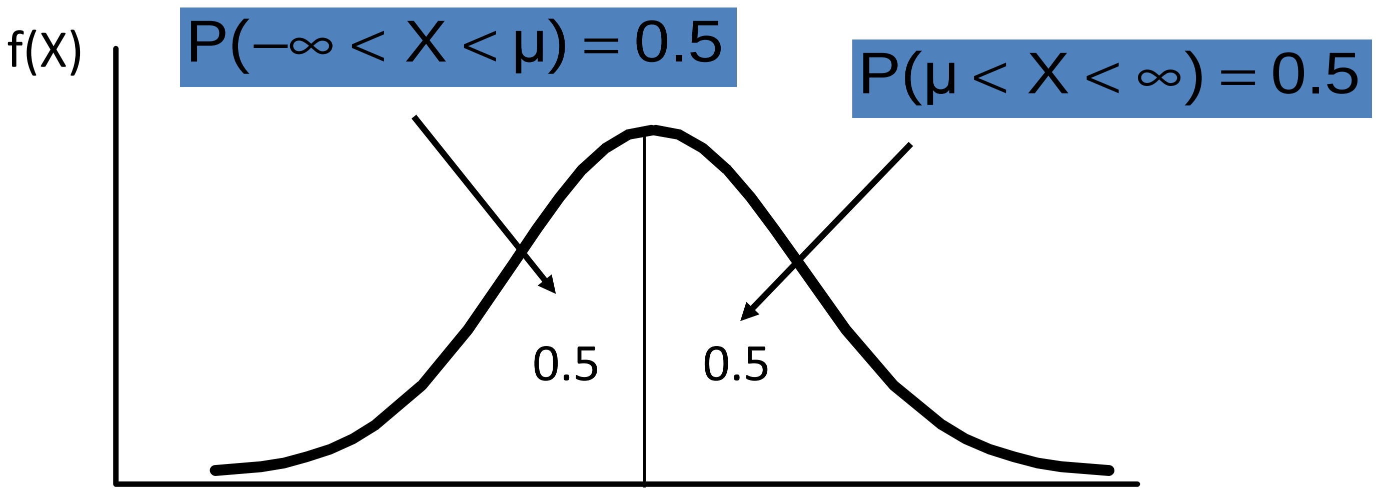
Probability is measured by the area under the curve



(Note that the probability of any individual value is zero)

Normal Probabilities

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



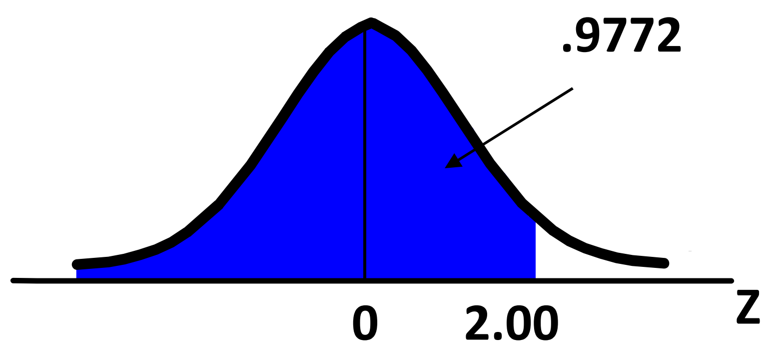
$$P(-\infty < X < \infty) = 1.0$$

Normal Probability Tables

The Standardized Normal table gives the probability less than a desired value for Z (i.e., from negative infinity to Z)

Example:

$P(Z < 2.00) = .9772$



Normal Probability Tables

=NORMSDIST

The column gives the value of Z to the second decimal point

Z	0.00	0.01	0.02 …
0.0			
0.1			
.			
.			
.			
2.0	**.9772**		

The row shows the value of Z to the first decimal point

The value within the table gives the probability from $Z = -\infty$ up to the desired Z value.

P(Z < 2.00) = .9772

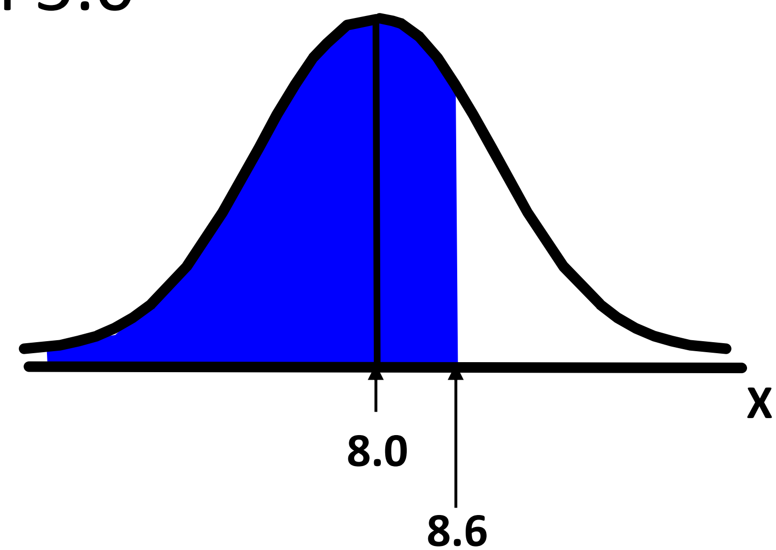
Finding Normal Probability Procedure

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

- Draw the normal curve for the problem in terms of X.
- Translate X-values to Z-values.
- Use the Standardized Normal Table.

Finding Normal Probability Example

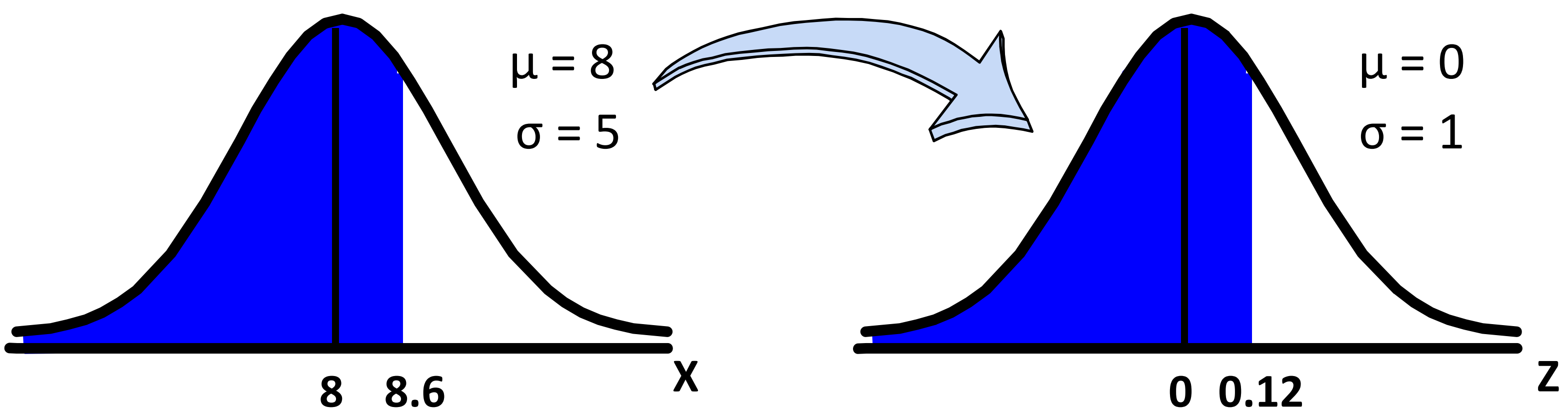
- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find $P(X < 8.6)$



Finding Normal Probability Example

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

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



P(X < 8.6)

P(Z < 0.12)

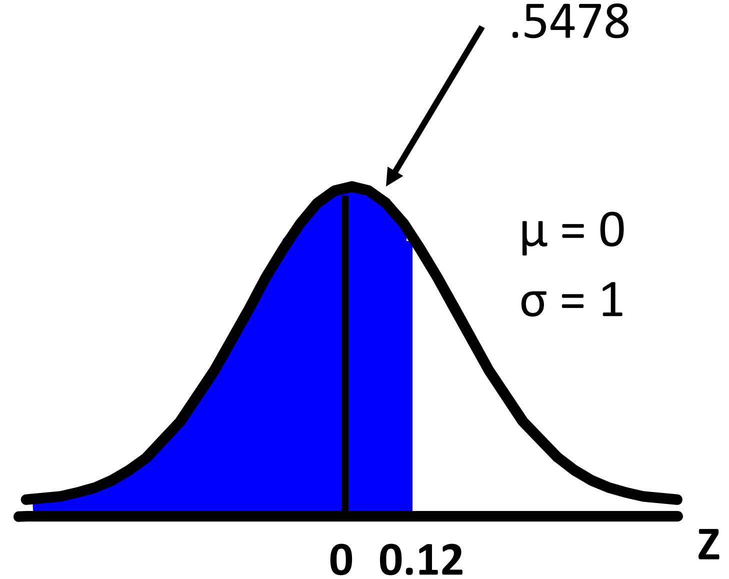
Finding Normal Probability Example

Standardized Normal Probability Table (Portion)

Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$P(X < 8.6)$

$= P(Z < 0.12)$



=NORMDIST(8.6,8,5,TRUE)

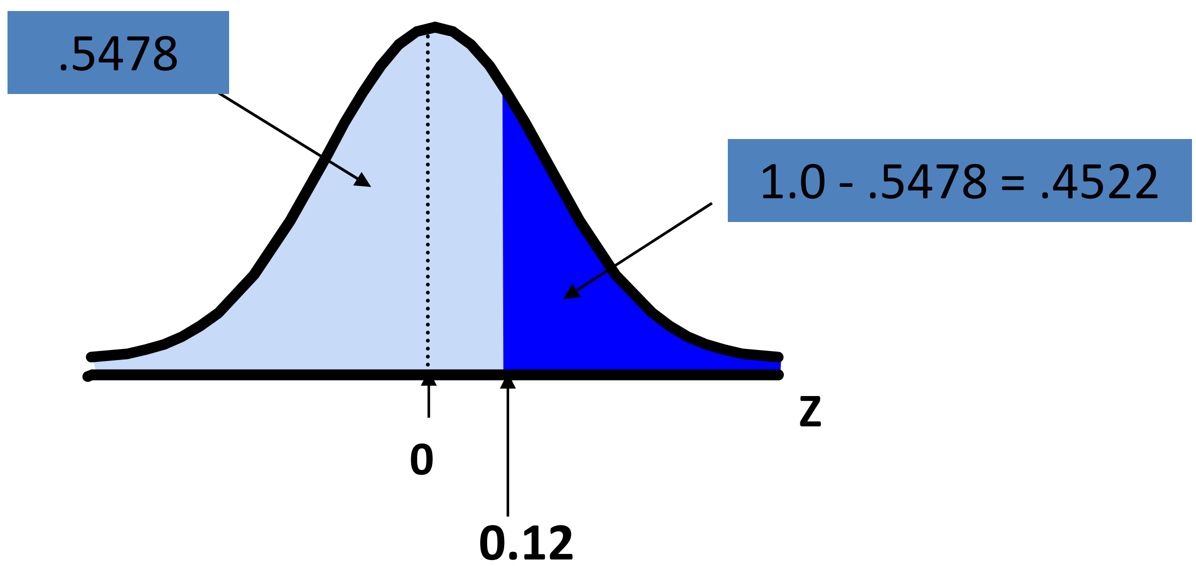
=NORMSDIST(0.12)

Finding Normal Probability Example

Find $P(X > 8.6)$…

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

$$= 1.0 - .5478 = .4522$$



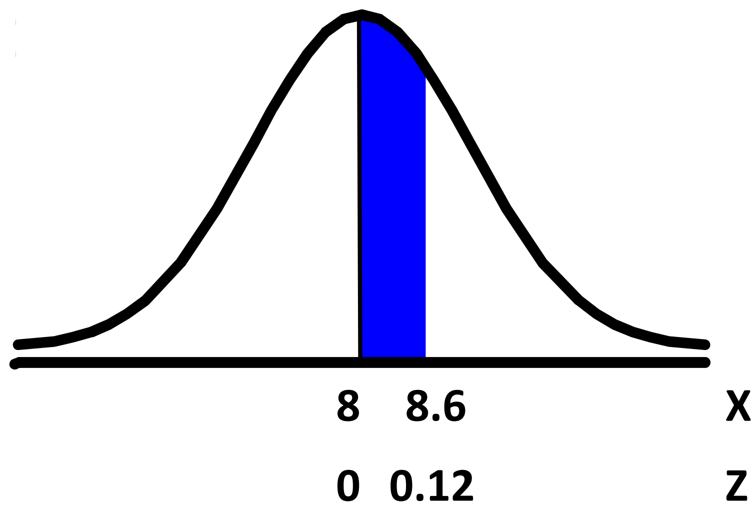
Finding Normal Probability Between Two Values

Suppose X is normal with mean 8.0 and standard deviation 5.0. Find P(8 < X < 8.6)

Calculate Z-values:

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 8}{5} = 0$$

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8}{5} = 0.12$$



P(8 < X < 8.6)

= P(0 < Z < 0.12)

Finding Normal Probability Between Two Values

Standardized Normal Probability Table (Portion)

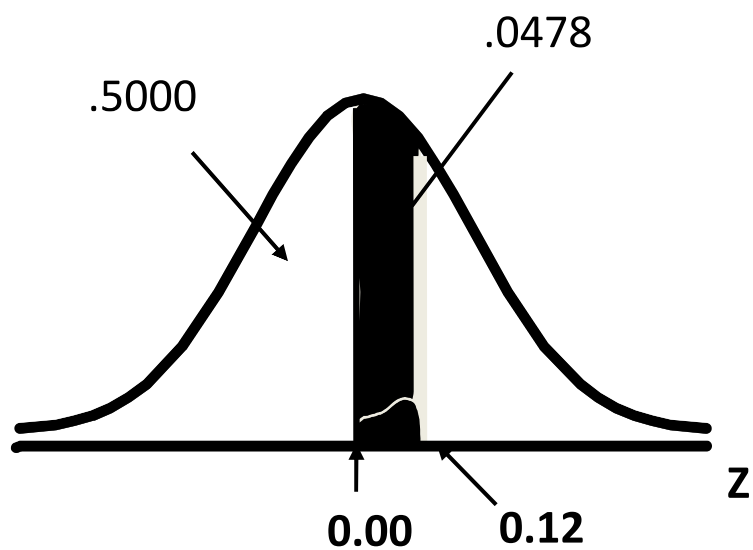
Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$P(8 < X < 8.6)$

$= P(0 < Z < 0.12)$

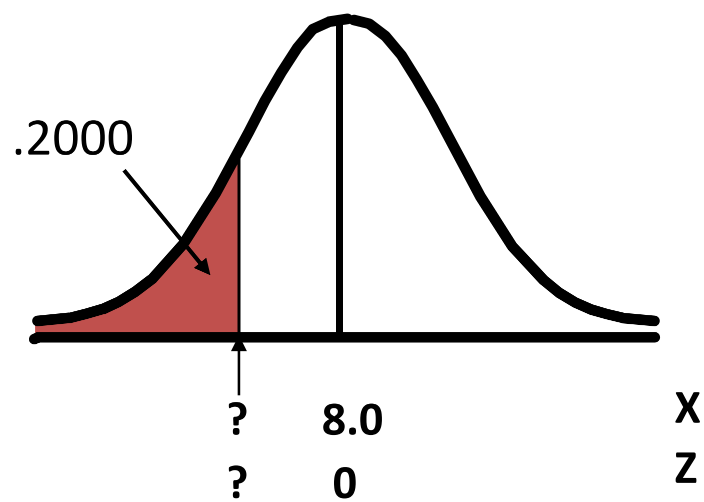
$= P(Z < 0.12) – P(Z \le 0)$

$= .5478 - .5000 = .0478$



Given Normal Probability, Find the X Value

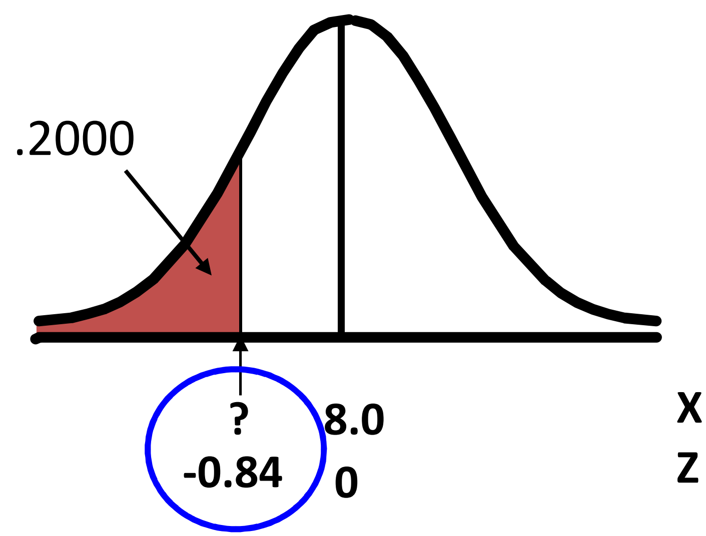
- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find X such that 20% of download times are less than X.



Given Normal Probability, Find the X Value

- First, find the Z value corresponds to the known probability using the table.

Z	….	.03	.04	.05
-0.9	….	.1762	.1736	.1711
-0.8	….	.2033	.2005	.1977
-0.7	….	.2327	.2296	.2266



Given Normal Probability, Find the X Value

- Second, convert the Z value to X units using the following formula.

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

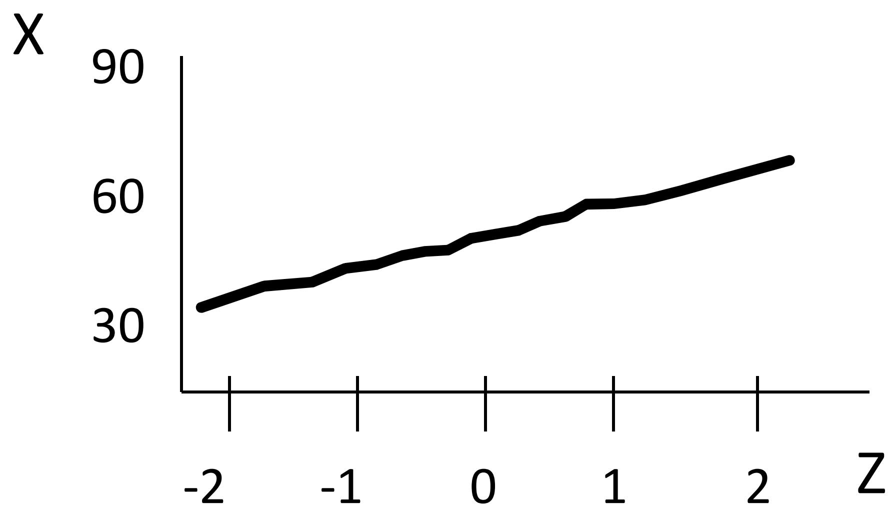
So 20% of the download times from the distribution with mean 8.0 and standard deviation 5.0 are less than 3.80 seconds.

=NORMSINV(0.2)=-0.84

=NORMINV(0.2,8,5)=3.8

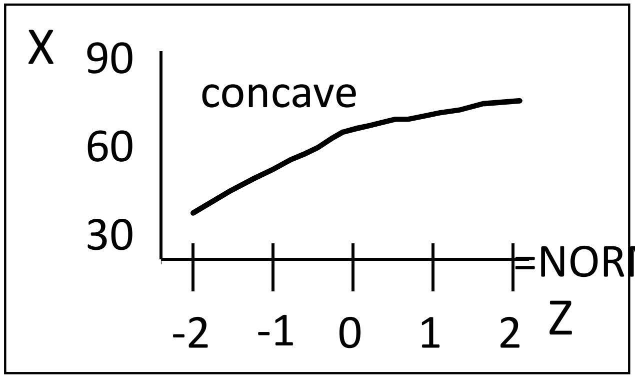
The Normal Probability Plot

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



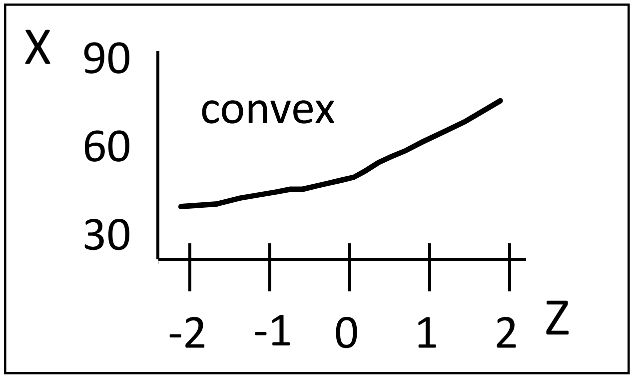
The Normal Probability Plot

Left-Skewed



=NORMSINV(0.2)

Right-Skewed



Nonlinear plots indicate a deviation from normality

The Uniform Distribution

- The **uniform distribution** is a probability distribution that has **equal probabilities** anywhere between the smallest value, a, and the largest value, b.

- Because of its shape, the uniform distribution is also called the **rectangular distribution.**

The Uniform Distribution

Uniform probability density function:

$$f(X) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le X \le b \\ 0 & \text{otherwise} \end{cases}$$

Where

a = the minimum value of X

b = the maximum value of X

The Uniform Distribution

- The mean of a uniform distribution is:

$$\mu = \frac{a+b}{2}$$

- The standard deviation is:

$$\sigma = \sqrt{\frac{(b-a)^2}{12}}$$

The Uniform Distribution Example

Example: Uniform probability distribution where a=2, b=6.

$$f(X) = \frac{1}{6 - 2} = .25 \quad \text{for } 2 \le X \le 6$$

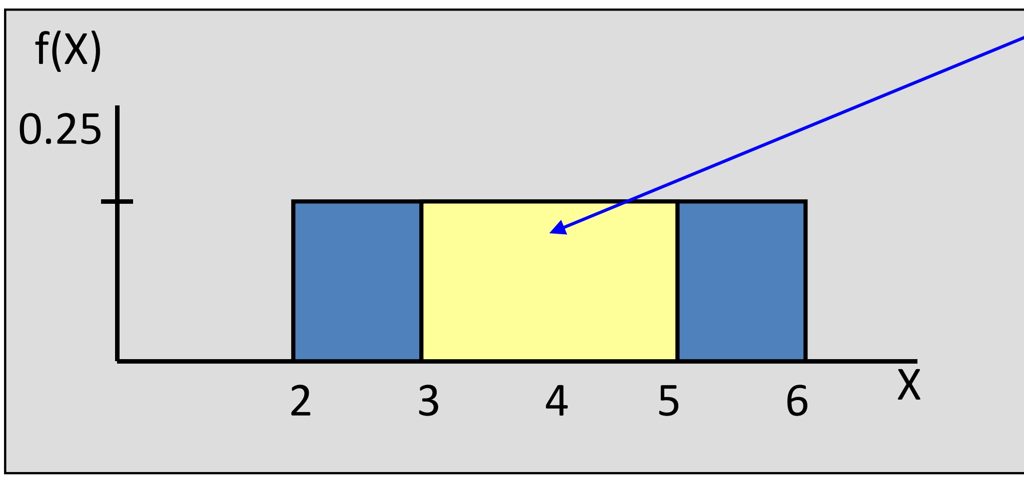
$$\mu = \frac{a+b}{2} = \frac{2+6}{2} = 4$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(6-2)^2}{12}} = 1.1547$$

The Uniform Distribution Example

Example: Using the uniform probability distribution to find $P(3 \le X \le 5)$:

$$P(3 \le X \le 5) = (\text{Base})(\text{Height}) = (2)(0.25) = 0.5$$



The Exponential Distribution

- Is a continuous distribution that is right-skewed and ranges from zero to positive infinity.
- Widely used in waiting-line (i.e. queuing) theory to model the length of time between arrivals in processes such as
  - customers at a bank’s ATM
  - patient entering a hospital emergency room
  - hits on a web site

The Exponential Distribution

- Defined by a single parameter, λ, the mean number of arrivals per unit of time.
- The cumulative probability that the length of time before the next arrival is less than or equal to X:

$$P(\text{arrival time} \leq X) = 1 - e^{-\lambda X}$$

where e = mathematical constant approximated by 2.71828

λ = the mean number of arrivals per unit

X = any value of the continuous variable where $0 < X < \infty$

The Exponential Distribution

Example: Customers arrive at the service counter at the rate of 15 per hour. What is the probability that the arrival time between consecutive customers is less than three minutes?

- The mean number of arrivals per hour is 15, so $\lambda = 15$
- Three minutes is .05 hours
- P(arrival time < .05) = $1 - e^{-\lambda X} = 1 - e^{-(15)(.05)} = .5276$
- So there is a 52.76% probability that the arrival time between successive customers is less than three minutes

=EXPONDIST(0.05,15,TRUE)