Probability and Statistics

eng. Fawzia faraag

Course Syllabus

- Chapter 1: Probability.
- Chapter 2: Random Variables.
- Chapter 3: Probability Distributions.
- Chapter 4: Descriptive Statistics.
- Chapter 5: Estimation (Confidence Interval).
- Chapter 6: Hypothesis Tests.

Chapter 1: Probability

- Sample Space.
- Events.
- Counting Techniques.
- Probability of an Event.
- Additive Rules.
- Conditional Probability.
- Independence, and the Product Rule.
- Bayes’ Rule.

Sample Space (1/9)

Random (Statistical) Experiment:

- An experiment <with known outcomes> whose outcome cannot be predicted with certainty, before the experiment is run.

Sample Space (1/9)

Random (Statistical) Experiment:

- An experiment <with known outcomes> whose outcome cannot be predicted with certainty, before the experiment is run.

The roll of a dice

The toss of (flipping) a coin

Sample Space (2/9)

Sample Space (S):

- Set of ALL possible outcomes of a random experiment.
- A sample space is discrete if it consists of a finite or countable infinite set of outcomes.
- A sample space is continuous if it contains an interval (either finite or infinite) of real numbers.

Sample Space (3/9)

Sample Space (S):

- Set of ALL possible outcomes of a random experiment.

The roll of a dice

Sample Space (3/9)

Sample Space (S):

- Set of ALL possible outcomes of a random experiment.

The roll of a dice

$$S = \{1,2,3,4,5,6\}$$

Sample Space (3/9)

Sample Space (S):

Discrete

- Set of ALL possible outcomes of a random experiment.

The roll of a dice

$$S = \{1,2,3,4,5,6\}$$

Each outcome in a sample space is called an **element** or a **member** of the sample space, or simply a **sample point**.

Sample Space (4/9)

Sample Space (S):

- Set of ALL possible outcomes of a random experiment.

Flipping a coin

Sample Space (4/9)

Sample Space (S):

- Set of **ALL** possible outcomes of a random experiment.

Flipping a coin

$$S = \{Head, Tail\}$$
$$S = \{H, T\}$$

Sample Space (5/9)

Example1:

Find the sample space for the random experiments (flipping) a coin of two times?

Sample Space (5/9)

Example1:

Find the sample space for the random experiments (flipping) a coin of two times?

Answer:

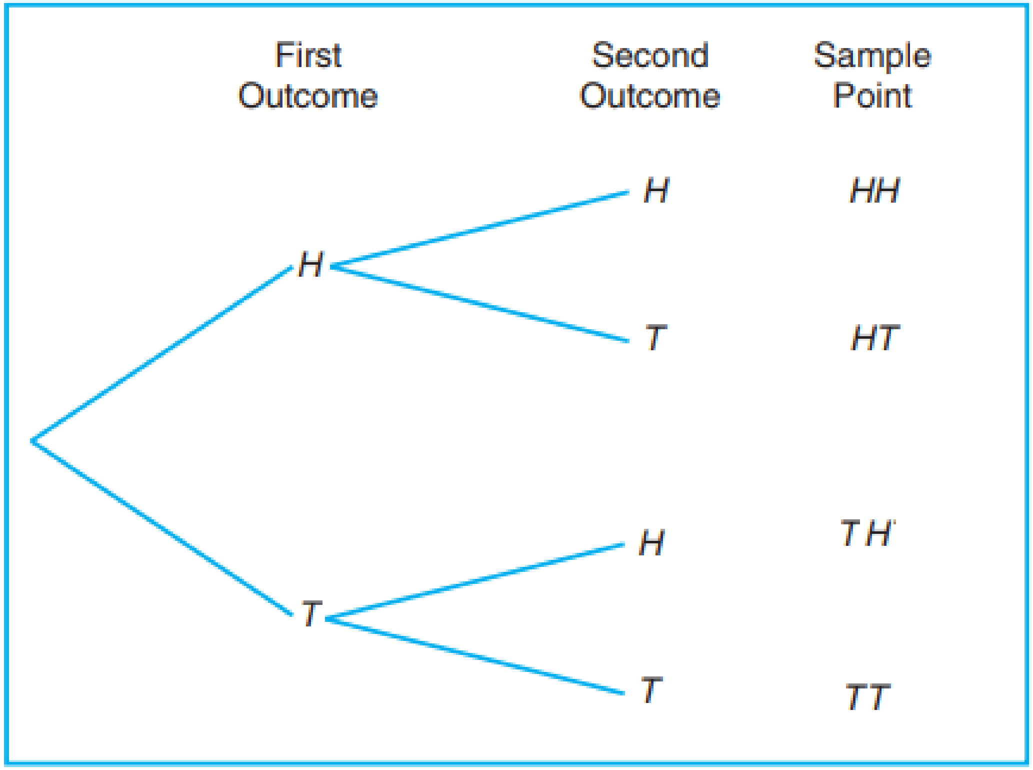
$S = \{HH, HT, TH, TT\}$

Sample Space (6/9)

Tree Diagrams:

Sample spaces can also be described graphically with *tree diagrams*.

$$S = \{HH, HT, TH, TT\}$$



Sample Space (7/9)

Example2:

An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once.

Sample Space (7/9)

Example2:

An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once.

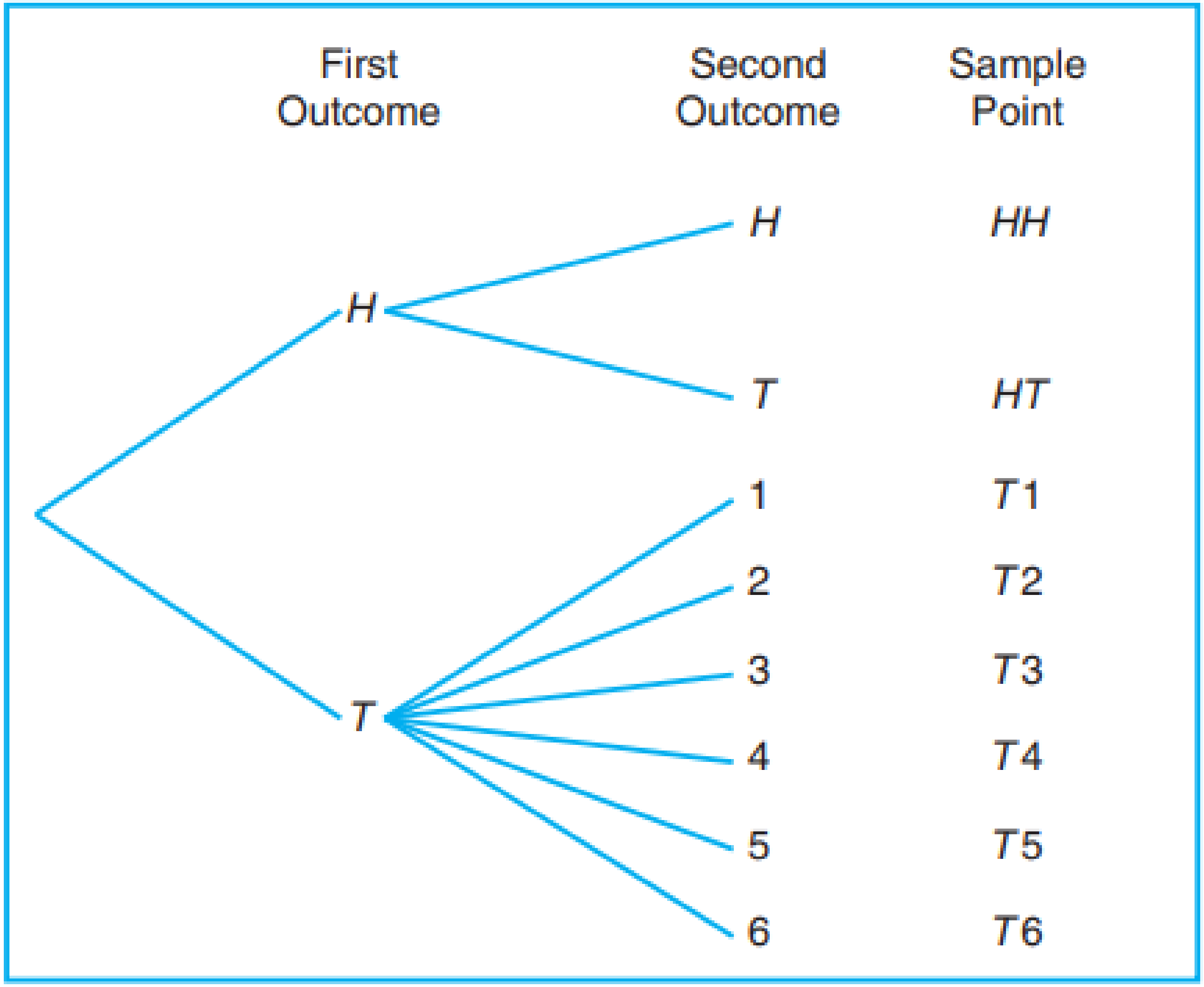
Answer:

$S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$

Sample Space (8/9)

Example2:

$S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$



Sample Space (9/9)

Example3:

Continuous

Consider an experiment that selects a cell phone camera and records the recycle time of a flash (the time taken to ready the camera for another flash).

$$\boldsymbol{S} = R^{+} = \{x \mid x > 0\}$$

If it is known that all recycle times are between 1.5 and 5 seconds, the sample space can be

$$\boldsymbol{S} = \{x \mid 1.5 < x < 5\}$$

Events (1/19)

Event (*E*):

- A result of *none* , *one* , or *more* outcomes in the sample space. An event is a subset of the sample space of a random experiment.

Events (2/19)

Event (E):

- A result of *none* , *one* , or *more* outcomes in the sample space. An event is a subset of the sample space of a random experiment.

The roll of a dice

$$S = \{1,2,3,4,5,6\}$$
$$E = \{2,4,6\}$$

Even Numbers

Events (3/19)

Example1:

A dice is rolled twice. What is the Event that the sum of the faces is greater than 7, given that the first outcome was a 4?

Events (4/19)

Example1:

A dice is rolled twice. What is the Event that the sum of the faces is greater than 7, given that the first outcome was a 4?

Answer:

$S = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26,$

$31, 32, 33, 34, 35, 36, \mathbf{41}, \mathbf{42}, \mathbf{43}, \mathbf{44}, \mathbf{45}, \mathbf{46},$

$51, 52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\}$

$E = \{44, 45, 46\}$

Events (5/19)

We can also be interested in describing new events from combinations of existing events. Because events are subsets, we can use basic set operations such as *unions*, *intersections*, and *complements* to form other events of interest. Some of the basic set operations are summarized here in terms of events:

1. The union of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as $E_1 \cup E_2$.

Events (6/19)

We can also be interested in describing new events from combinations of existing events. Because events are subsets, we can use basic set operations such as *unions*, *intersections*, and *complements* to form other events of interest. Some of the basic set operations are summarized here in terms of events:

2. The intersection of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$.

Events (7/19)

We can also be interested in describing new events from combinations of existing events. Because events are subsets, we can use basic set operations such as *unions*, *intersections*, and *complements* to form other events of interest. Some of the basic set operations are summarized here in terms of events:

3. The complement of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the complement of the event E as E'. The notation E^C is also used in other literature to denote the complement.

Events (8/19)

Example2:

In the tossing of a die, we might let A be the event that an even number occurs and B the event that a number greater than 3 shows.

Events (9/19)

Example2:

In the tossing of a die, we might let A be the event that an even number occurs and B the event that a number greater than 3 shows.

Then the subsets $A = \{2, 4, 6\}$ and $B = \{4, 5, 6\}$ are subsets of the same sample space $S = \{1, 2, 3, 4, 5, 6\}$.

Example2:

Then the subsets $A = \{2, 4, 6\}$ and $B = \{4, 5, 6\}$ are subsets of the same sample space $S = \{1, 2, 3, 4, 5, 6\}$.

$A \cap B = \{4, 6\}$

$A \cup B = \{2, 4, 5, 6\}$

$A' = \{1, 3, 5\}$

$B' = \{1, 2, 3\}$

Events (11/19)

Mutually Exclusive, or Disjoint:

Two events A and B are *mutually exclusive*, or *disjoint*, if $A \cap B = \emptyset$, that is, if A and B have **no** elements in common.

$A = \{2, 4, 6\}$ and $B = \{1, 3, 5\}$

$A \cap B = \{\ \} = \emptyset$

Events (12/19)

Venn Diagrams:

Diagrams are often used to portray relationships between sets, and these diagrams are also used to describe relationships between events. We can use Venn diagrams to represent a sample space and events in a sample space.

Events (13/19)

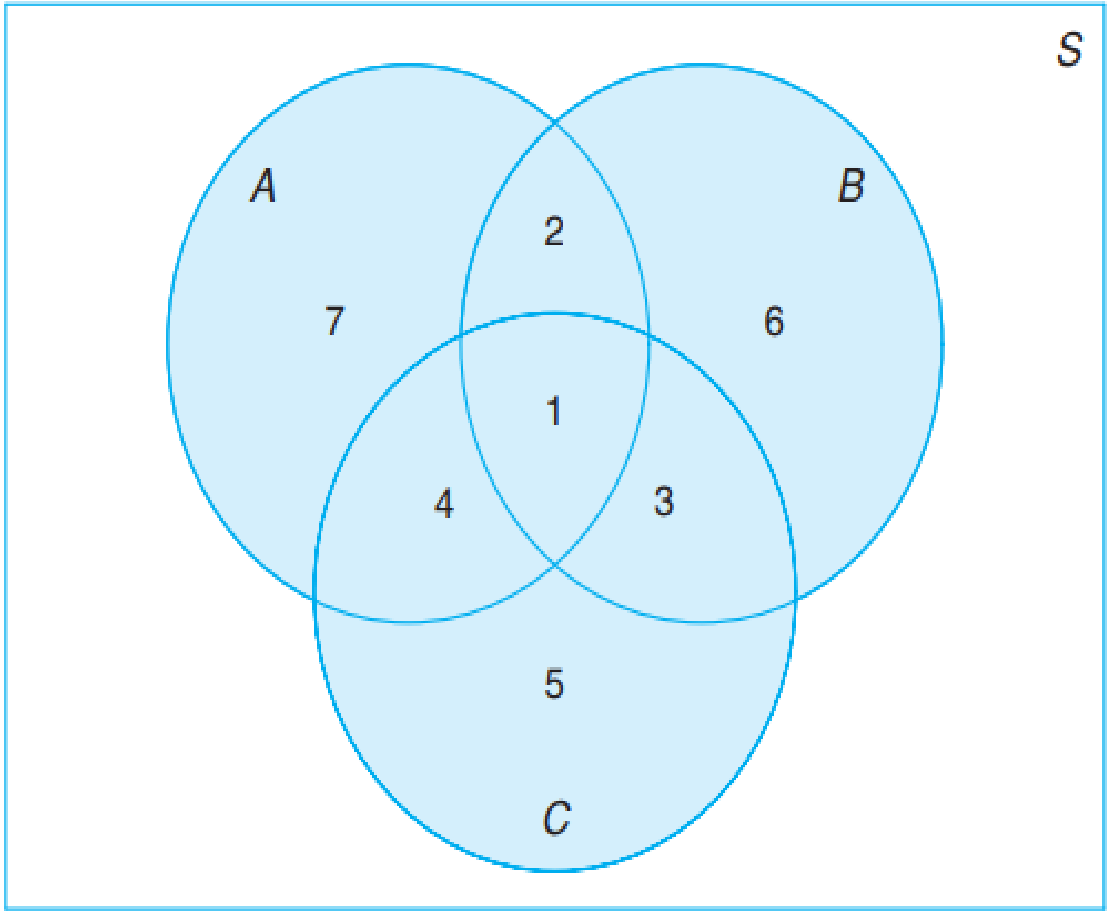
Example1:

$S = \{1, 2, 3, 4, 5, 6, 7\}$

$A = \{1, 2, 4, 7\}$

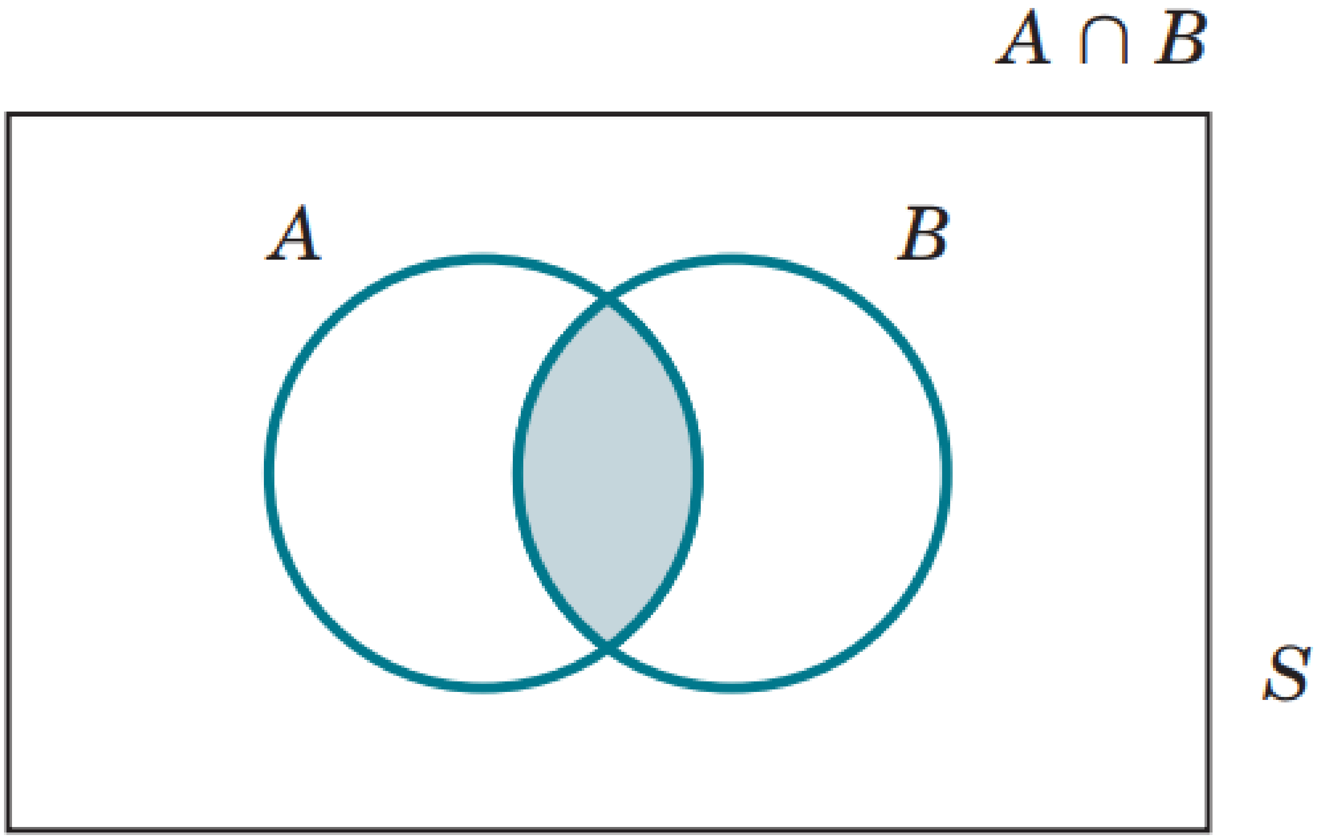
$B = \{1, 2, 3, 6\}$

$C = \{1, 3, 4, 5\}$



Events (14/19)

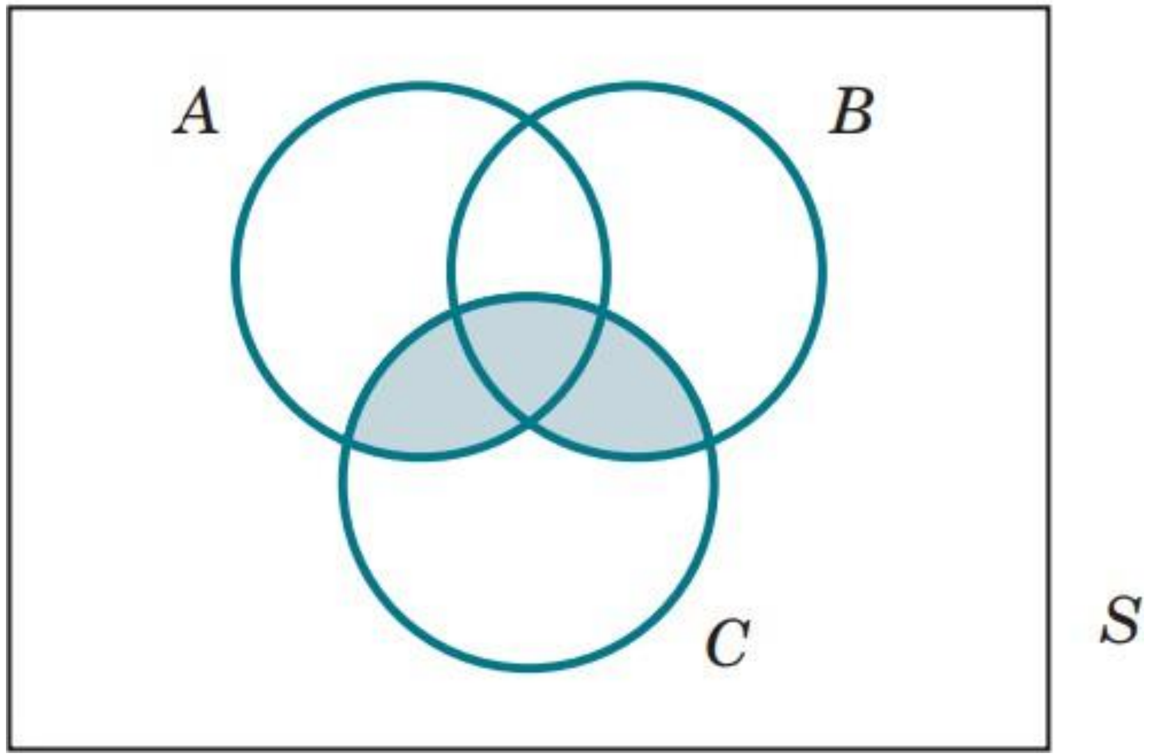
Example2:



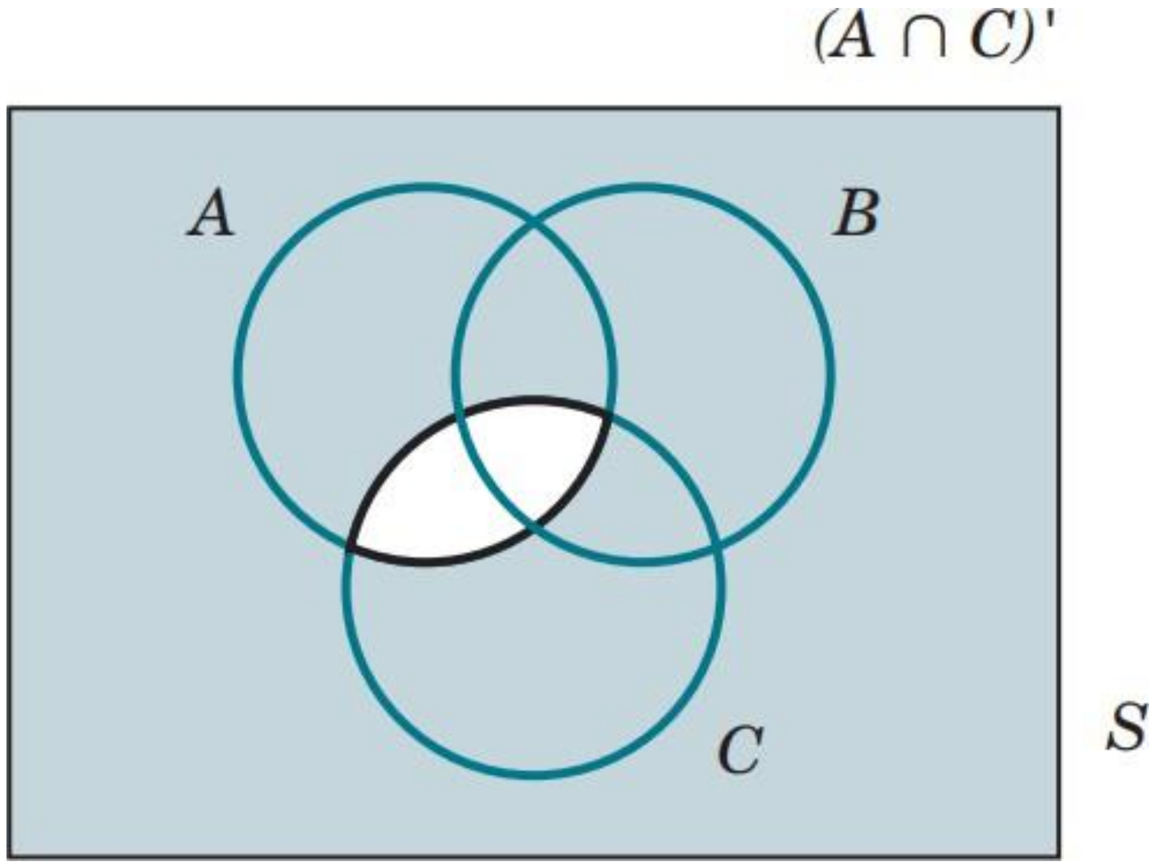
Events (15/19)

Example3:

$(A \cup B) \cap C$

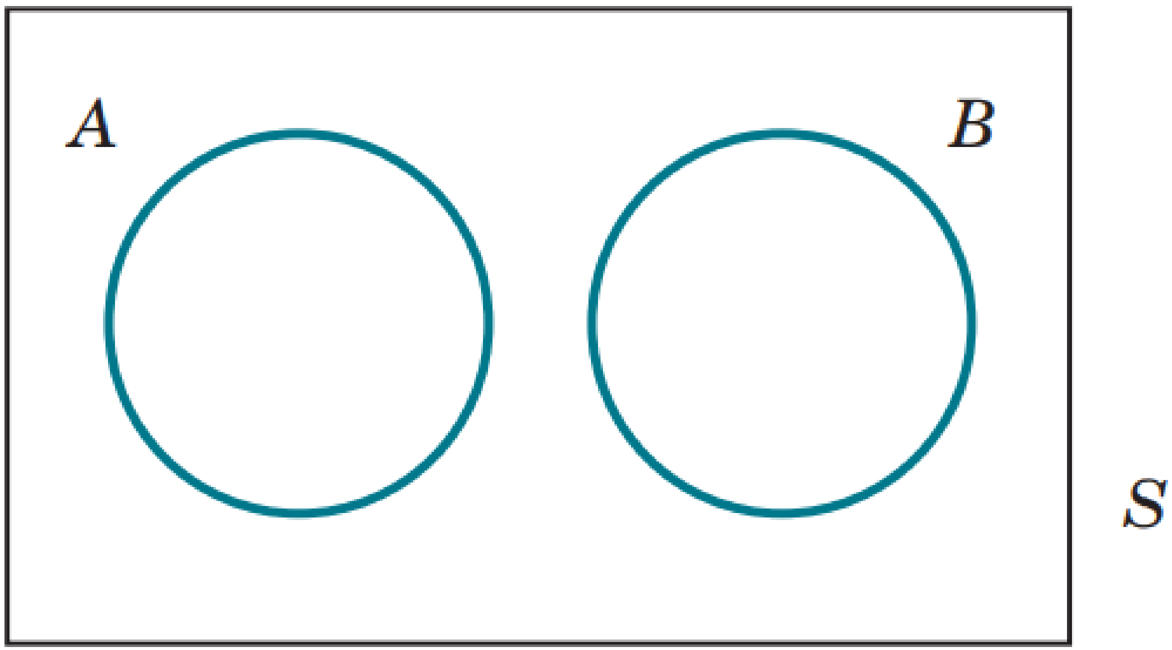


Example4:



Events (17/19)

Example5:



Mutually exclusive events.

Events (18/19)

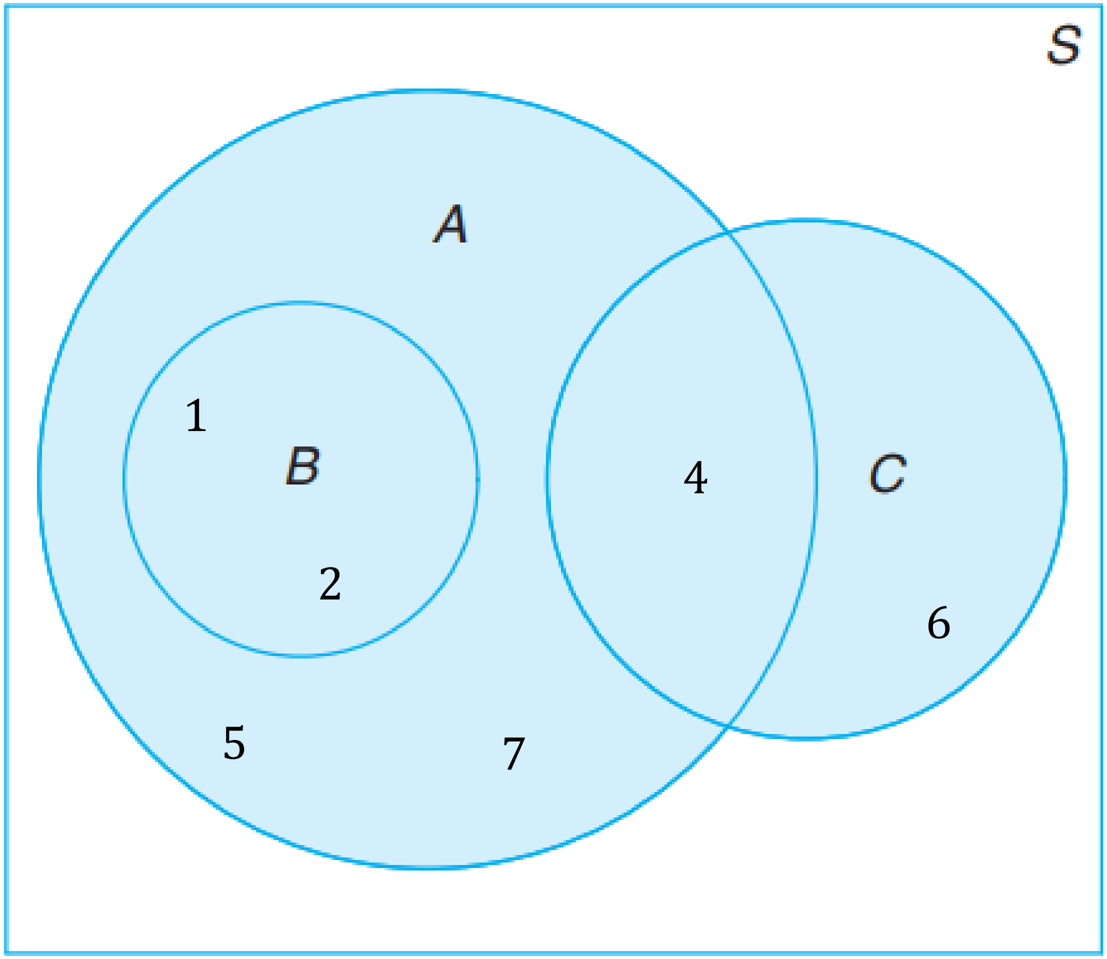
Example6:

$S = \{1, 2, 3, 4, 5, 6, 7\}$

$A = \{1, 2, 4, 5, 7\}$

$B = \{1, 2\}$

$C = \{4, 6\}$



Events (19/19)

Several Results:

1. $A \cap \phi = \phi.$
2. $A \cup \phi = A.$
3. $A \cap A' = \phi.$
4. $A \cup A' = S.$
5. $S' = \phi.$
6. $\phi' = S.$
7. $(A')' = A.$
8. $(A \cap B)' = A' \cup B'.$
9. $(A \cup B)' = A' \cap B'.$

Chapter 1: Probability

- Sample Space.
- Events.
- Counting Techniques.
- Probability of an Event.
- Additive Rules.
- Conditional Probability.
- Independence, and the Product Rule.
- Bayes' Rule.

Counting Techniques (1/31)

Introduction (1/2):

In more complicated examples, determining the outcomes in the sample space (or an event) becomes more difficult. In these cases, counts of the numbers of outcomes in the sample space and various events are used to analyze the random experiments.

Introduction (2/2):

Suppose that a **password** on a computer system consists of **eight** characters. Each of these characters must be a **digit** or a **letter** of the alphabet. Each password must contain **at least one digit**.

How many such passwords are there?!

Counting Techniques (2/31)

Multiplication (Product) Rule:

Suppose that a procedure can be **broken down** into a **sequence** of **two tasks**. If there are $\boldsymbol{n_1}$ ways to do the first task and for each of these ways of doing the first task, there are $\boldsymbol{n_2}$ ways to do the second task, then there are $\boldsymbol{n_1 n_2}$ ways to do the procedure and so forth.

The total number of ways to complete the operation is

$$n_1 \times n_2 \times \cdots \times n_k$$

Example1:

How many sample points are there in the sample space when a pair of dice is thrown once?

Example1:

How many sample points are there in the sample space when a pair of dice is thrown once?

Solution: The first die can land face-up in any one of $n_1 = 6$ ways. For each of these 6 ways, the second die can also land face-up in $n_2 = 6$ ways. Therefore, the pair of dice can land in $n_1 n_2 = (6)(6) = 36$ possible ways.

Counting Techniques (5/31)

Example2:

The design for a Website is to consist of *four colors*, *three fonts*, and *three positions for an image*.

How many different designs are possible?

Counting Techniques (6/31)

Example2:

The design for a Website is to consist of *four colors*, *three fonts*, and *three positions for an image*.

How many different designs are possible?

Solution: From the multiplication rule, $4 \times 3 \times 3 = 36$ different designs are possible.

Counting Techniques (7/31)

Example3:

In how many different ways can a true-false test consisting of 10 questions be answered?

Example3:

In how many different ways can a true-false test consisting of 10 questions be answered?

Solution: Each of the 10 questions can be chosen in two ways, because each question is either true or false. Therefore, the product rule shows there are:

$2 \times 2 \times \cdots \times 2 = 2^{10} = 1024$ ways to answer the test.

Example4:

How many different bit strings of *length three* are there?

Example4:

How many different bit strings of *length three* are there?

Solution: Each of the seven bits can be chosen in two ways, because each bit is either 0 or 1. Therefore, the product rule shows there are a total of $2^3 = 8$ different bit strings of length three.

$$2 \times 2 \times 2 = 8$$

Counting Techniques (11/31)

Example5:

How many bit strings of length 5, start and end with 1's?

Example5:

How many bit strings of length 5, start and end with 1's?

Solution:

	Bit #5	Bit #4	Bit #3	Bit #2	Bit #1
Prob.	1	0, 1	0, 1	0, 1	1
Count	1	2	2	2	1

There are $(1 \cdot 2 \cdot 2 \cdot 2 \cdot 1) = 8$ bit strings of length 5, start and end with 1's

Example6:

How many different license plates are available if each plate contains a sequence of *three letters* followed by *three digits*.

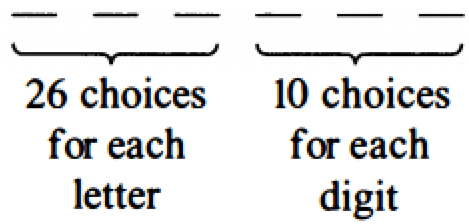
Counting Techniques (13/31)

Example6:

How many different license plates are available if each plate contains a sequence of *three letters* followed by *three digits*.

Example6:

How many different license plates are available if each plate contains a sequence of *three letters* followed by *three digits*.



Solution:

There are 26 choices for each of the three letters and ten choices for each of the three digits. Hence, by the product rule there are a total of $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17{,}576{,}000$ possible license plates.

Example7:

If a 22-member club needs to elect a chair and a treasurer, how many different ways can these two to be elected?

Example7:

If a 22-member club needs to elect a chair and a treasurer, how many different ways can these two to be elected?

Solution: For the chair position, there are 22 total possibilities. For each of those 22 possibilities, there are 21 possibilities to elect the treasurer. Using the multiplication rule, we obtain $n_1 \times n_2 = 22 \times 21 = 462$ different ways.

Permutations (1/3):

Another useful calculation finds the number of ordered sequences of the elements of a set. Consider a set of elements, such as $S = \{a, b, c\}$.

A permutation of the elements is an ordered sequence of the elements. For example, abc, acb, bac, bca, cab, and cba are all of the permutations of the elements of S.

$$3 \times 2 \times 1 = 6$$

Counting Techniques (17/31)

Permutations (2/3):

The number of **permutations** of n different elements is $n!$ where

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

Ex. The number of permutations of the four letters a, b, c, and d will be $4! = 24$.

Permutations (3/3):

In some situations, we are interested in the number of arrangements of only some of the elements of a set.

The number of permutations of subsets of r elements selected from a set of n different elements is

$$P_r^n = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

$$_nP_r = \frac{n!}{(n-r)!}$$

Example1:

Consider a set of elements, such as $S = \{a, b, c, d, e\}$.

What is the number of permutation of subsets of **3** elements selected from S is?

Example1:

Consider a set of elements, such as $S = \{a, b, c, d, e\}$.

What is the number of permutation of subsets of **3** elements selected from S is?

<u>**Solution:**</u> $\boldsymbol{r = 3}, \quad \boldsymbol{n = 5}$

$$P_r^n = {}_n P_r = \frac{n!}{(n-r)!} = \frac{5!}{(2)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = 60$$

Example2:

In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?

Example2:

In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?

Solution: $\boldsymbol{r = 3, \ \ n = 25}$

$$P_r^n = {}_nP_r = \frac{n!}{(n-r)!} = \frac{25!}{(22)!}$$

$$= \frac{25 \times 24 \times \cdots \times 3 \times 2 \times 1}{22 \times 21 \times \cdots \times 2 \times 1} = 13,800$$

Counting Techniques (22/31)

Permutations of Similar Objects:

The number of permutations of $n = n_1 + n_2 + \cdots + n_r$ objects of which n_1 are of one type, n_2 are of a second type, … , and n_r are of an rth type is

$$\frac{n!}{n_1!n_2!n_3!\ldots n_r!}$$

Example3:

In a Statistics class, the teacher needs to have 20 students standing in a row. Among these 20 students, there are 12 boy, and 8 girl. How many different ways can they be arranged in a row if only their class level will be distinguished?

Example3:

In a Statistics class, the teacher needs to have 20 students standing in a row. Among these 20 students, there are 12 boy, and 8 girl. How many different ways can they be arranged in a row if only their class level will be distinguished?

Solution: $\boldsymbol{n = 20}$, $\boldsymbol{n_1 = 12}$, $\boldsymbol{n_2 = 8}$

$$= \frac{n!}{n_1!\, n_2!} = \frac{20!}{12!\, 8!} = 125{,}970$$

Counting Techniques (25/31)

Combinations (1/2):

Another counting problem of interest is the number of subsets of r elements that can be selected from a set of n elements. Here, order is not important. These are called *combinations*.

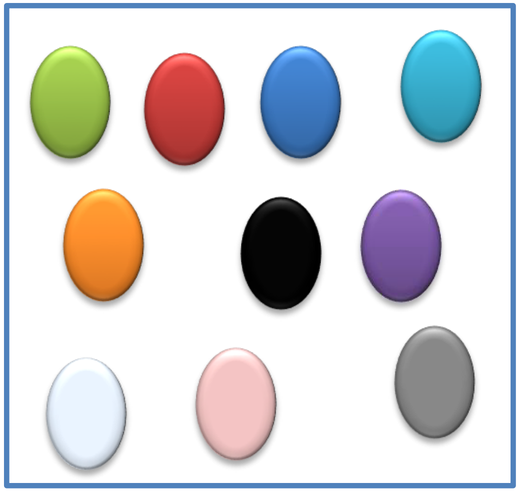
Combinations (2/2):

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example1:

How many possible selections of 3 balls from a box contains 10 colored balls?



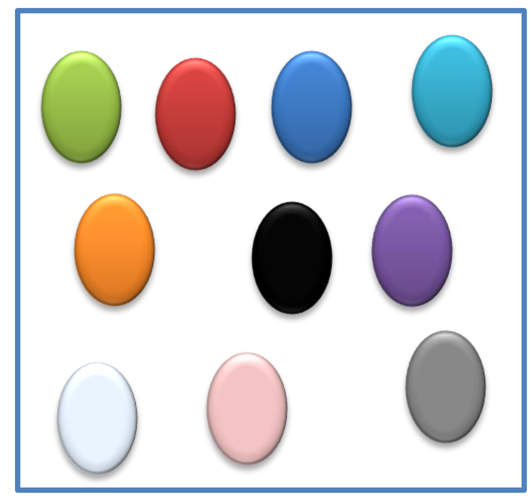
Counting Techniques (27/31)

Example1:

How many possible selections of 3 balls from a box contains 10 colored balls?

Solution: $\boldsymbol{n = 10}$, $\boldsymbol{r = 3}$

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{10!}{3!\,7!} = 120$$



Example2:

How many ways are there to select 3 candidates from 7 equally qualified recent graduates for work?

Example2:

How many ways are there to select 3 candidates from 7 equally qualified recent graduates for work?

<u>Solution:</u> $\boldsymbol{n = 7}\ ,\quad \boldsymbol{r = 3}$

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{7!}{3!\,4!} = 35$$

Example3:

A bin of 50 manufactured parts. A sample of 6 parts is selected without replacement. That is, each part can be selected only once. How many different samples are there of size 6?

Counting Techniques (31/31)

Example3:

A bin of 50 manufactured parts. A sample of 6 parts is selected without replacement. That is, each part can be selected only once. How many different samples are there of size 6?

Solution: $\boldsymbol{n = 50}$, $\boldsymbol{r = 6}$

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{50!}{6!\,44!} = 15{,}890{,}700$$

Probability of an Event (1/20)

Probability is used to quantify the likelihood, or chance, that an outcome of a random experiment will occur.

"The chance of rain today is 30%" is a statement that quantifies our feeling about the possibility of rain.

Probability of an Event (2/20)

The likelihood of an outcome is quantified by assigning a number from the interval [0, 1] to the outcome (or a percentage from 0 to 100%).

Higher numbers indicate that the outcome is more likely than lower numbers. A 0 indicates an outcome will not occur. A probability of 1 indicates that an outcome will occur with certainty.

Probability of an Event (3/20)

Equally Likely Outcomes:

Whenever a sample space consists of N possible outcomes that are equally likely, the probability of each outcome is $1/N$.

Probability of an Event (3/20)

Equally Likely Outcomes:

Whenever a sample space consists of N possible outcomes that are equally likely, the probability of each outcome is $1/N$.

Ex. $S = \{1, 2, 3, 4, 5, 6\}$

$$P(1) = \frac{1}{6}, \qquad P(2) = \frac{1}{6}, \qquad \ldots \qquad , \qquad P(6) = \frac{1}{6}$$

Probability of an Event (4/20)

Equally Likely Outcomes:

- S Is a sample space, E is an event
- $E \subseteq S$

$$P(E) = \frac{\#\text{of outcomes in event}(E)}{\text{Total \# of outcomes in sample space}(S)} = \frac{n(E)}{n(S)}$$

$$1 \geq P(E) \geq 0$$

Probability of an Event (5/20)

Example1:

A dice is rolled once. What is the probability of the Event that contains a prime number from the sample space?

Probability of an Event (6/20)

Example1:

A dice is rolled once. What is the probability of the Event that contains a prime number from the sample space?

Solution:

$S = \{1,2,3,4,5,6\}$

$E = \{2,3,5\}$

$$P(E) = \frac{3}{6} = 0.5$$

Probability of an Event (7/20)

Example2:

A coin is tossed twice. What is the probability that at least 1 head occurs?

Probability of an Event (8/20)

Example2:

A coin is tossed twice. What is the probability that at least 1 head occurs?

Solution:

$S = \{HH, HT, TH, TT\}$

$E = \{HH, HT, TH\}$

$$P(E) = \frac{3}{4}$$

Probability of an Event (9/20)

NOT Equally Likely Outcomes:

A random experiment can result in one of the outcomes $\{a, b, c, d\}$ with probabilities 0.1, 0.3, 0.5, and 0.1, respectively. Let A denote the event $\{a, b\}$, B the event $\{b, c, d\}$, and C the event $\{d\}$. Then,

$$P(A) = 0.1 + 0.3 = 0.4$$

$$P(B) = 0.3 + 0.5 + 0.1 = 0.9$$

$$P(C) = 0.1$$

Probability of an Event (10/20)

Probability of an Event:

For a discrete sample space, the probability of an event E, denoted as $P(E)$, equals the sum of the probabilities of the outcomes in E.

Probability of an Event (11/20)

Example3:

A dice is loaded in such a way that an even number is twice as likely to occur as an odd number. If E is the event that a number less than 4 occurs on a single toss of the dice, find $P(E)$.

Probability of an Event (12/20)

Example3:

A dice is loaded in such a way that an even number is twice as likely to occur as an odd number. If E is the event that a number less than 4 occurs on a single toss of the dice, find $P(E)$.

Solution:

We assign a probability of w to each odd number and a probability of $2w$ to each even number. Since the sum of the probabilities must be 1, we have $9w = 1$ or $w = 1/9$.

$S = \{1, 2, 3, 4, 5, 6\}$,

$E = \{1, 2, 3\}$

$$P(E) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$$

Probability of an Event (13/20)

Example4:

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that

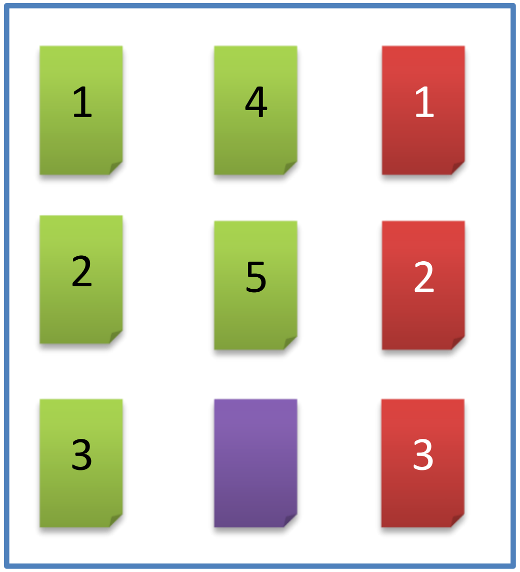
(a) the dictionary is selected?

(b) 2 novels and 1 book of mathematics are selected?

Probability of an Event (14/20)

Example4:

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that



Solution:

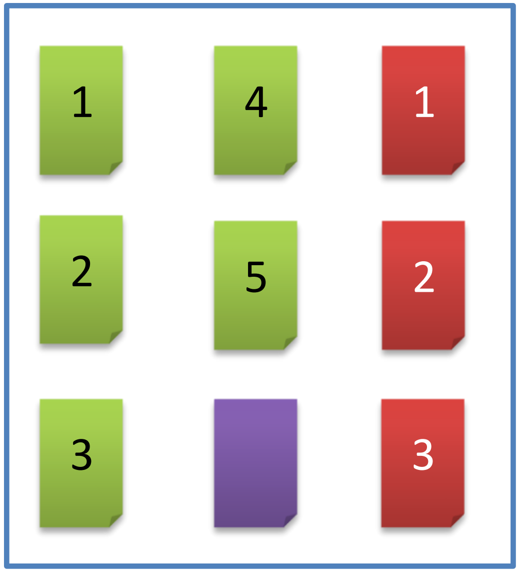
$$\#\text{of members of } S = \binom{9}{3} = \frac{9!}{3!\,6!} = 84,$$

Probability of an Event (15/20)

Example4:

(a) the dictionary is selected?

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that



Solution:

$$\#\text{of members of } S = \binom{9}{3} = \frac{9!}{3!\,6!} = 84,$$

$$\#\text{of members of } E = \binom{1}{1}\binom{8}{2}$$

Probability of an Event (16/20)

Example4:

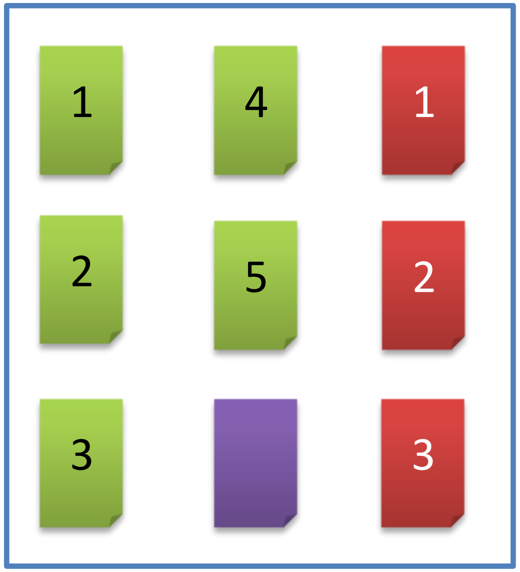
(a) the dictionary is selected?

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that

Solution:

$$\#\text{of members of } E = \binom{1}{1}\binom{8}{2}$$

$$= 1 * \frac{8!}{2!\ 6!} = 28$$

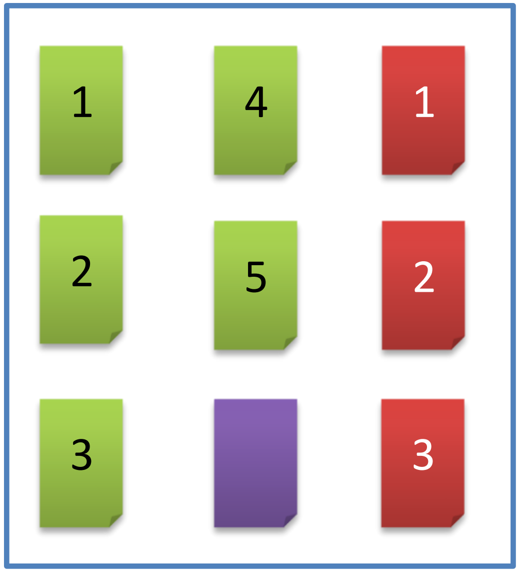


Probability of an Event (17/20)

Example4:

(a) the dictionary is selected?

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that



Solution:

The probability that the dictionary is selected

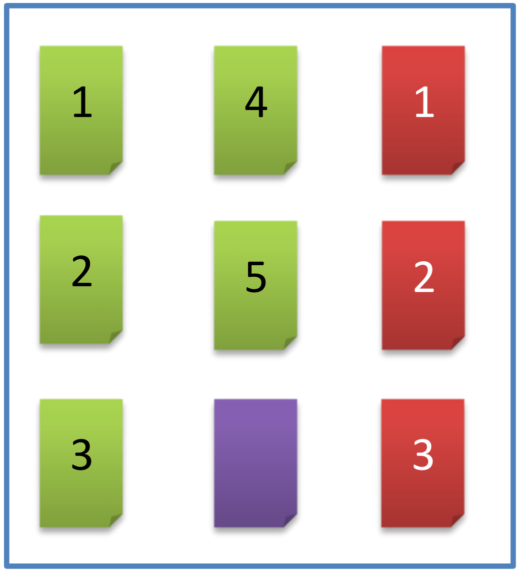
$$= \frac{28}{84} = \frac{1}{3} = 0.333$$

Probability of an Event (18/20)

Example4: (b) 2 novels and 1 book of mathematics are selected?

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that

Solution:



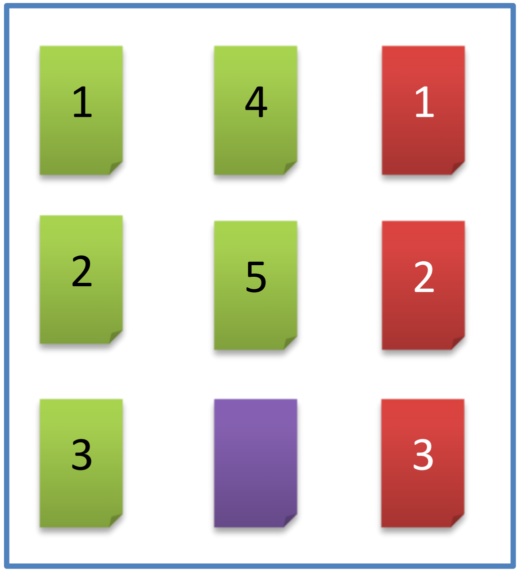
$$\#\text{of members of } E = \binom{5}{2}\binom{3}{1}$$

$$= \frac{5!}{2!\,3!} * \frac{3!}{1!\,2!} = 10 * 3 = 30$$

Probability of an Event (19/20)

Example4: (b) 2 novels and 1 book of mathematics are selected?

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that



Solution:

The probability that 2 novels and

1 book of mathematics are selected

$$= \frac{30}{84} = \frac{5}{14} = 0.357$$

Probability of an Event (20/20)

Axioms of Probability:

S is a sample space, A is an event

$A \subseteq S$

$$\boldsymbol{P(S) = 1}$$

$$\boldsymbol{P(\emptyset) = 0}$$

$$\boldsymbol{0 \leq P(A) \leq 1}$$

$$\boldsymbol{P(A') = 1 - P(A)}$$

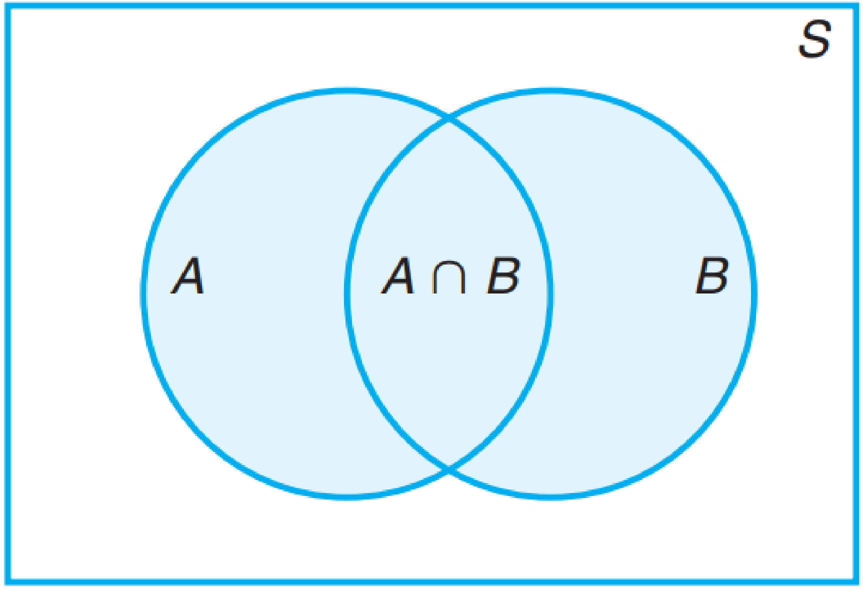
Chapter 1: Probability

- Sample Space.
- Events.
- Counting Techniques.
- Probability of an Event.
- Additive Rules.
- Conditional Probability.
- Independence, and the Product Rule.
- Bayes' Rule.

Additive Rules (1/13)

If A and B are two events, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$



Additive Rules (2/13)

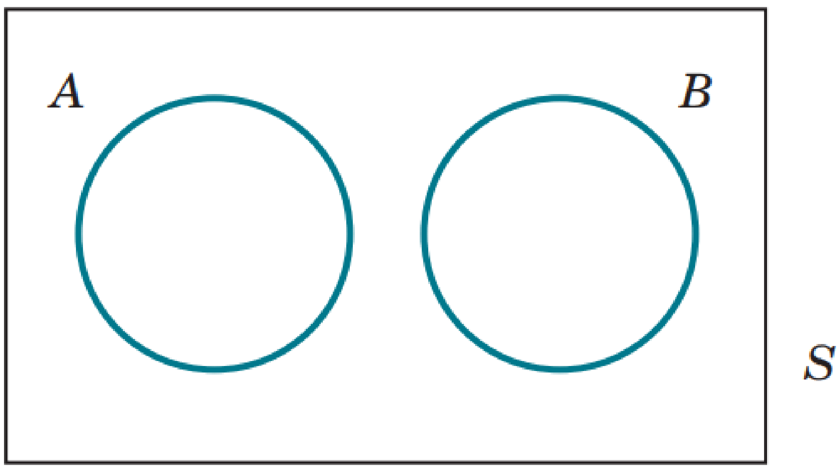
For three events A, B, and C,

$$
\begin{aligned}
P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\
& - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).
\end{aligned}
$$

Additive Rules (3/13)

If A and B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B).$$



Additive Rules (4/13)

If $A_1,\ A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

Additive Rules (5/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

1) $P(A')$
2) $P(A \cup B)$
3) $P(A' \cap B)$
4) $P[(A \cup B)']$

Additive Rules (6/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

1) $P(A') = 1 - P(A) = 1 - 0.3 = 0.7$

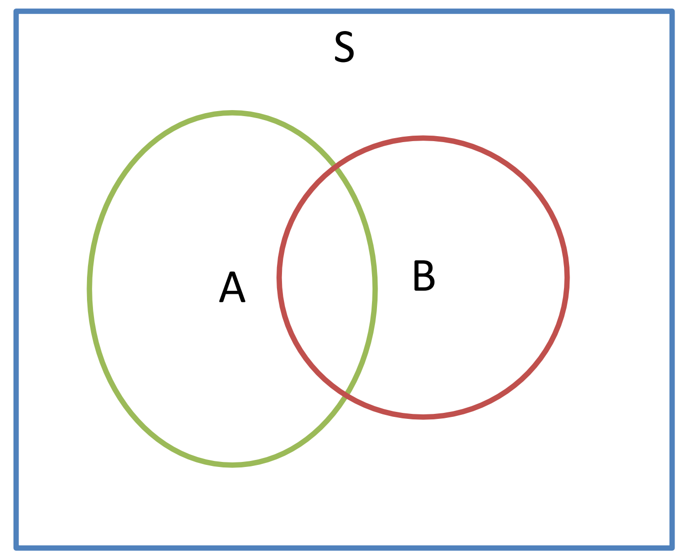
2) $P(A \cup B) = 0.3 + 0.2 - 0.1 = 0.4$

Additive Rules (7/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

3) $P(A' \cap B)$

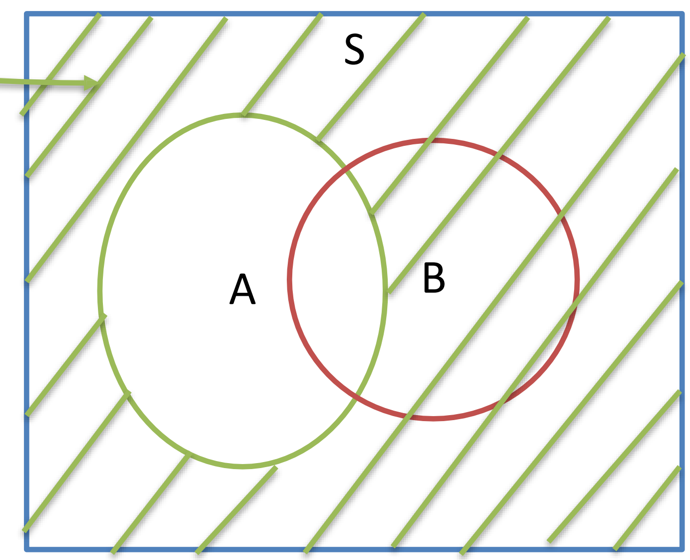


Additive Rules (7/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

3) $P(A' \cap B)$

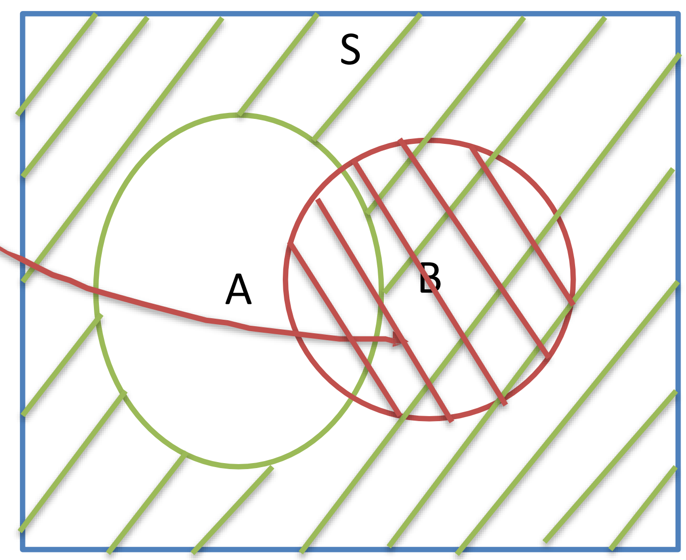


Additive Rules (7/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

3) $P(A' \cap B)$



Additive Rules (7/13)

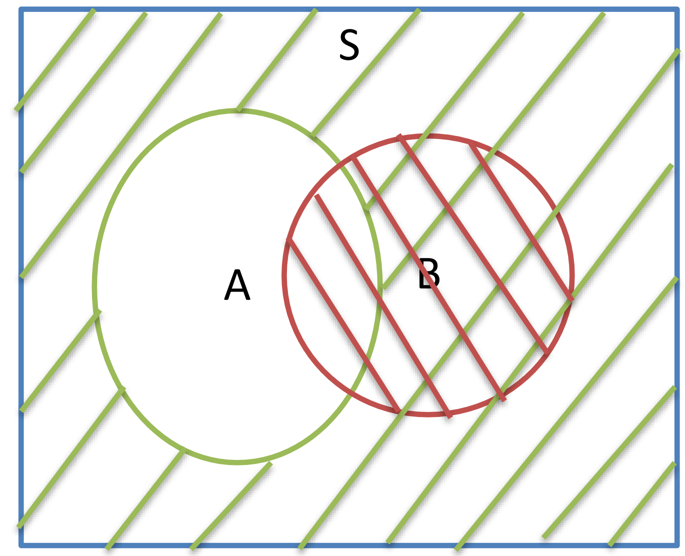
Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

3) $P(A' \cap B)$

$= P(B) - P(A \cap B) = 0.2 - 0.1$

$= 0.1$



Additive Rules (8/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

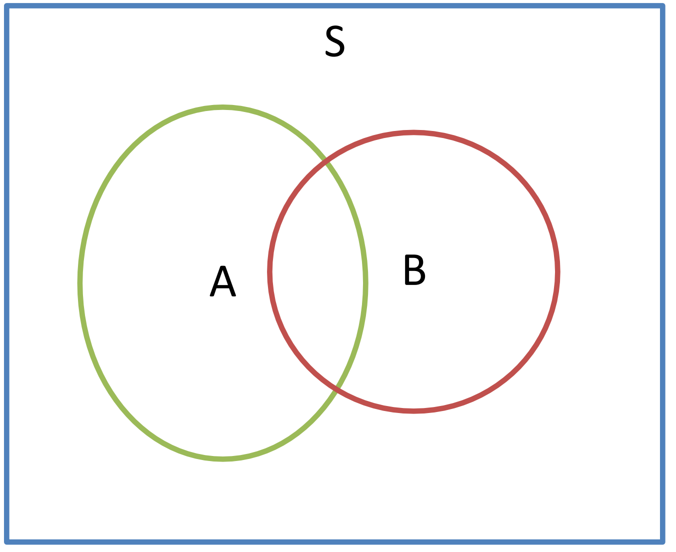
4) $P[(A \cup B)']$

Additive Rules (8/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

4) $P[(A \cup B)'] = P(A' \cap B')$



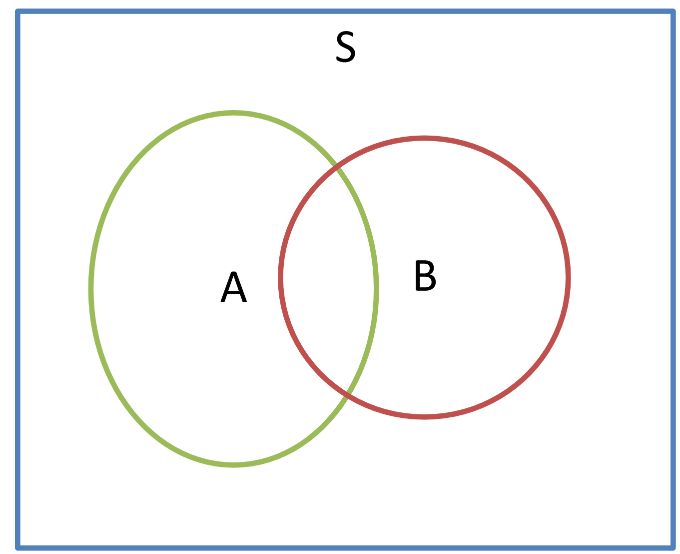
Additive Rules (8/13)

Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

4) $P[(A \cup B)'] = P(A' \cap B')$

$= 1 - P(A \cup B) = 1 - 0.4 = 0.6$



Additive Rules (9/13)

Example5:

What is the probability of getting a total of 7 or 11 when a pair of fair dice is tossed?

Additive Rules (10/13)

Example5:

What is the probability of getting a total of 7 or 11 when a pair of fair dice is tossed?

<u>**Solution:**</u>

Let A be the event that 7 occurs and B the event that 11 comes up.

$S = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26,$

$31, 32, 33, 34, 35, 36, 41, 42, 43, 44, 45, 46, 51,$

$52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\}$

Additive Rules (11/13)

Example5:

What is the probability of getting a total of 7 or 11 when a pair of fair dice is tossed?

Solution:

Let A be the event that 7 occurs and B the event that 11 comes up.

$$P(A) = \frac{6}{36}$$

$S = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26,$

$31, 32, 33, 34, 35, 36, 41, 42, 43, 44, 45, 46, 51,$

$52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\}$

Additive Rules (12/13)

Example5:

What is the probability of getting a total of 7 or 11 when a pair of fair dice is tossed?

Solution:

Let A be the event that 7 occurs and B the event that 11 comes up.

$$S = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 31, 32, 33, 34, 35, 36, 41, 42, 43, 44, 45, 46, 51, 52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\}$$

$$P(B) = \frac{2}{36}$$

Additive Rules (13/13)

Example5:

What is the probability of getting a total of 7 or 11 when a pair of fair dice is tossed?

Solution:

The probability of getting a total of 7 or 11 $= P(A \cup B)$

The events A and B are mutually exclusive, since a total of 7 and 11 cannot both occur on the same toss. Therefore,

$$P(A \cup B) = P(A) + P(B) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}$$

Conditional Probability (1/31)

Introduction (1/2):

Sometimes probabilities need to be reevaluated as additional information becomes available. The probability of an event B under the knowledge that the outcome will be in event A is denoted as

$$P(B \mid A)$$

and this is called the conditional probability of B given A.

Conditional Probability (1/31)

Introduction (2/2):

The conditional probability of B, given A, denoted by $P(B \mid A)$, is defined by

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad , \qquad \text{provided} \quad P(A) > 0$$

Conditional Probability (1/31)

Introduction (2/2):

The conditional probability of B, given A, denoted by $P(B \mid A)$, is defined by

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad , \qquad \text{provided} \quad P(A) > 0$$

Equally Likely Outcomes

$$\frac{P(A \cap B)}{P(A)} = \frac{\text{number of outcomes in } A \cap B}{\text{number of outcomes in } A}$$

Conditional Probability (2/31)

Example1:

As an additional illustration, suppose that our sample space S is the population of adults in a small town who have completed the requirements for a college degree.
We shall categorize them according to gender and employment status. The data are given in the following.

Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Conditional Probability (3/31)

Example1:

One of these individuals is to be selected at random. What is the probability that he is a male and given that he is employed?

Categorization of the Adults in a Small Town

	Employed	**Unemployed**	**Total**
Male	460	40	500
Female	140	260	400
Total	600	300	900

Conditional Probability (3/31)

Example1:

One of these individuals is to be selected at random. What is the probability that he is a male and given that he is employed?

$$P(M|E)$$

Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Conditional Probability (3/31)

Example1:

One of these individuals is to be selected at random. What is the probability that he is a male and given that he is employed?

$$P(M|E) = \frac{P(M \cap E)}{P(E)}$$

Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Example1:

One of these individuals is to be selected at random. What is the probability that he is a male and given that he is employed?

$$P(M \cap E) = \frac{460}{900}$$

Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Conditional Probability (5/31)

Example1:

One of these individuals is to be selected at random. What is the probability that he is a male and given that he is employed?

$$P(E) = \frac{600}{900}$$

Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Conditional Probability (5/31)

Example1:

One of these individuals is to be selected at random. What is the probability that he is a male and given that he is employed?

$$P(M|E) = \frac{460/900}{600/900} = \frac{460}{600}$$

Categorization of the Adults in a Small Town

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

Conditional Probability (6/31)

Example2:

Consider an industrial process in the textile industry in which strips of a particular type of cloth are being produced. These strips can be defective in two ways, length and nature of texture. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Conditional Probability (7/31)

Example2:

$P(L) = 0.1$

Consider an industrial process in the textile industry in which strips of a parti[illegible]lar type of cloth are being produced. These [illegible] be defective in two ways, length and nature of texture. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Conditional Probability (7/31)

Example2:

Consider an industrial process in the textile industry in which str… of a particular type of cloth are being pr… These strips …fective in two ways, length and nature of …ure. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

$P(L) = 0.1$

$P(T)$ .05

Conditional Probability (7/31)

Example2:

Consider an industrial process in the textile industry in which stri… of a particular type of cloth are being pr… These strips … fective in t… ngth and nature of … are. It is … from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

$P(L) = 0.1$

$P(T)$ .05

$P(L$ $T)$ .008

Example2:

Consider an industrial process in the textile industry in which stri... of a particular type of cloth are being pr... . These strips ... fective in t... ngth and nature of ...ure. It is ... from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests.

$P(L) = 0.1$

$P(T) = 0.05$

$P(L \cap T) = 0.008$

If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

$$P(T|L) = 0.008/0.1 = 0.08$$

Conditional Probability (8/31)

Example3:

A die is rolled twice. What is the probability that the sum equal 10, if you know that the 1st element equal 6?

Conditional Probability (9/31)

Example3:

A die is rolled twice. What is the probability that the sum equal 10, if you know that the 1st element equal 6?

Solution: $A = \{46, 55, 64\}$, $B = \{61, 62, 63, 64, 65, 66\}$

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

Conditional Probability (9/31)

Example3:

A die is rolled twice. What is the probability that the sum equal 10, if you know that the 1st element equal 6?

Solution: $A = \{46, 55, 64\}$, $B = \{61, 62, 63, 64, 65, 66\}$

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

$P(A) = 3/36$
$P(B) = 6/36$

$P(A \cap B) = 1/36$

$(A \cap B) = \{64\}$

Conditional Probability (9/31)

Example3:

A die is rolled twice. What is the probability that the sum equal 10, if you know that the 1^{st} element equal 6?

Solution: $A = \{46, 55, 64\}$, $B = \{61, 62, 63, 64, 65, 66\}$

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{6/36} = 1/6}$$

$P(A) = 3/36$
$P(B) = 6/36$

$P(A \cap B) = 1/36$

$(A \cap B) = \{64\}$

Conditional Probability (10/31)

Disjoint (or mutually exclusive):

S is a sample space, A, B are two events

$A, B \subseteq S$ and A, B **Disjoint** *or* **mutually exclusive**

$$\therefore P(A \cap B) = 0$$

$$\therefore P(A \cup B) = P(A) + P(B)$$

$$\therefore P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

Conditional Probability (11/31)

Independence:

S is a sample space, A, B are two events

$A, B \subseteq S$ and A, B are **independent**

$$\therefore P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B)}{P(B)} = P(A)$$

$$\therefore P(A \cap B) = P(A) * P(B)$$

$$\therefore P(A \cup B) = P(A) + P(B) - P(A) * P(B)$$

Example4:

- If $P(A) = 0.2$, $P(B) = 0.3$ determine the following probabilities:

If A, and B are **disjoint** (mutually exclusive)

1) $P(A \cap B)$
2) $P(A \cup B)$
3) $P(A|B)$

Conditional Probability (13/31)

Example4:

- If $P(A) = 0.2$, $P(B) = 0.3$ determine the following probabilities:

If A, and B are **disjoint** (mutually exclusive)

Solution:

$$P(A \cap B) = 0$$

$$P(A \cup B) = P(A) + P(B) = 0.2 + 0.3 = 0.5$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

Example5:

- If $P(A) = 0.2$, $P(B) = 0.3$ determine the following probabilities:

If A, and B are **independent**

1) $P(A \cap B)$
2) $P(A \cup B)$
3) $P(A|B)$

Conditional Probability (15/31)

Example5:

- If $P(A) = 0.2$, $P(B) = 0.3$ determine the following probabilities:

If A, and B are **independent**

Solution:

$$P(A \cap B) = P(A) * P(B) = 0.2 * 03 = 0.06$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 - 0.06 = 0.44$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B)}{P(B)} = P(A) = 0.2$$

Example6:

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?

Example6:

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?

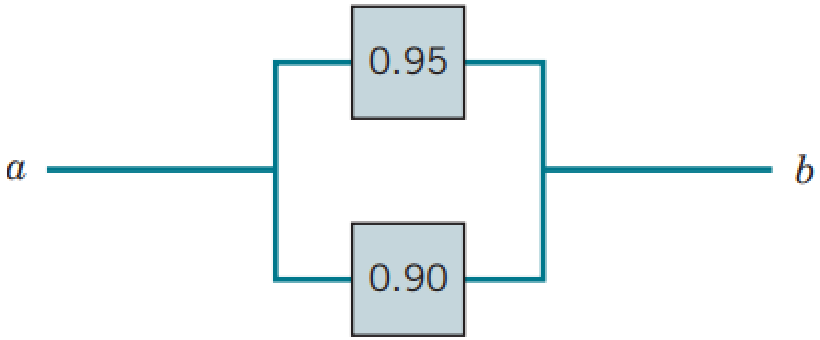
Let L and R denote the events that the left and right devices operate, respectively.

$$P(L \cap R) = P(L)P(R) = 0.80(0.90) = 0.72$$

Conditional Probability (18/31)

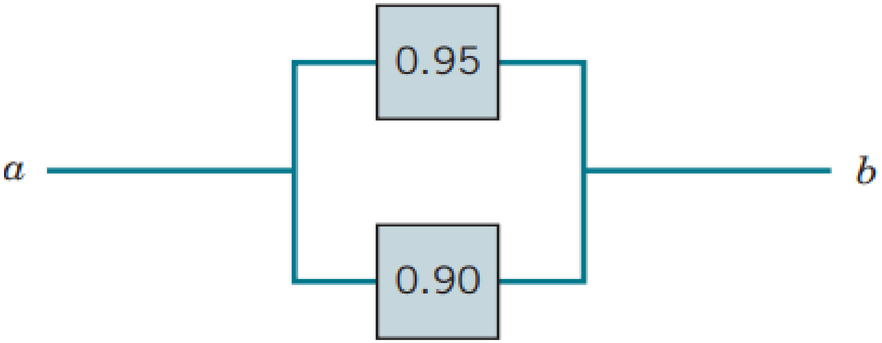
Example7:

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



Example7:

What is the probability that the circuit operates?



Let T and B denote the events that the top and bottom devices operate, respectively.

$$P(T \cup B) = P(T) + P(B) - P(T)P(B)$$

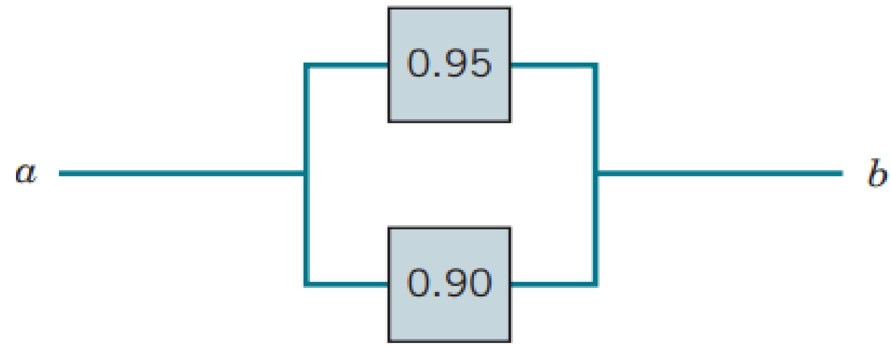
$$= 0.95 + 0.90 - (0.95)(0.90) = 0.995$$

Conditional Probability (20/31)

Example7:

Another Solution

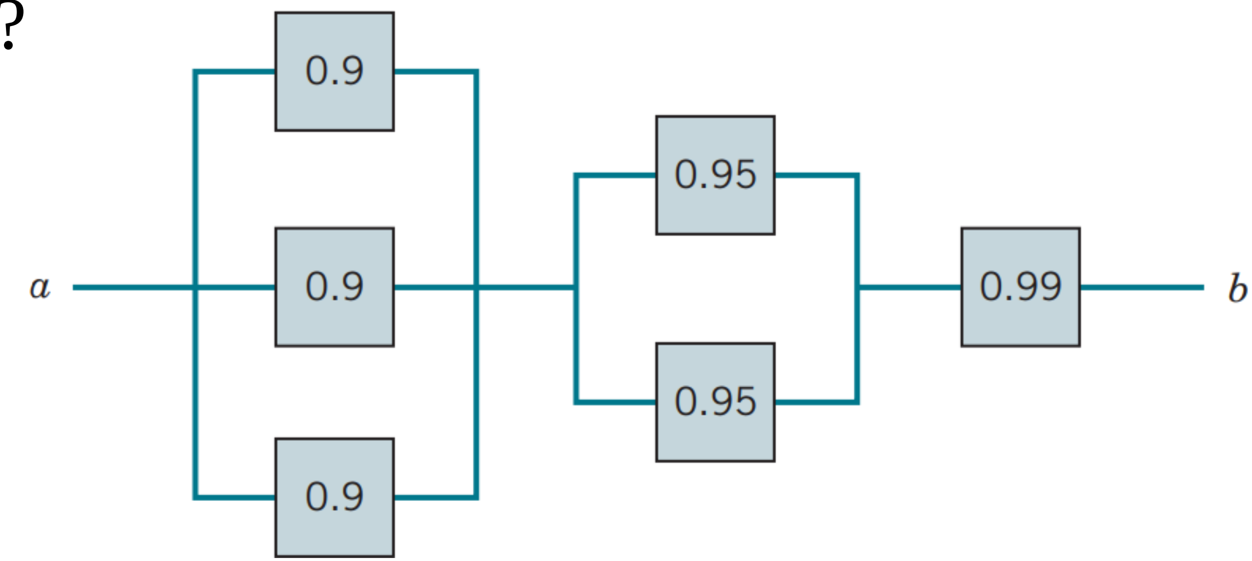
What is the probability that the circuit operates?



$$P(T \cup B) = 1 - P(T \cup B)'$$
$$= 1 - P(T' \cap B') = 1 - P(T')P(B')$$
$$= 1 - (0.05)(0.10) = 0.995$$

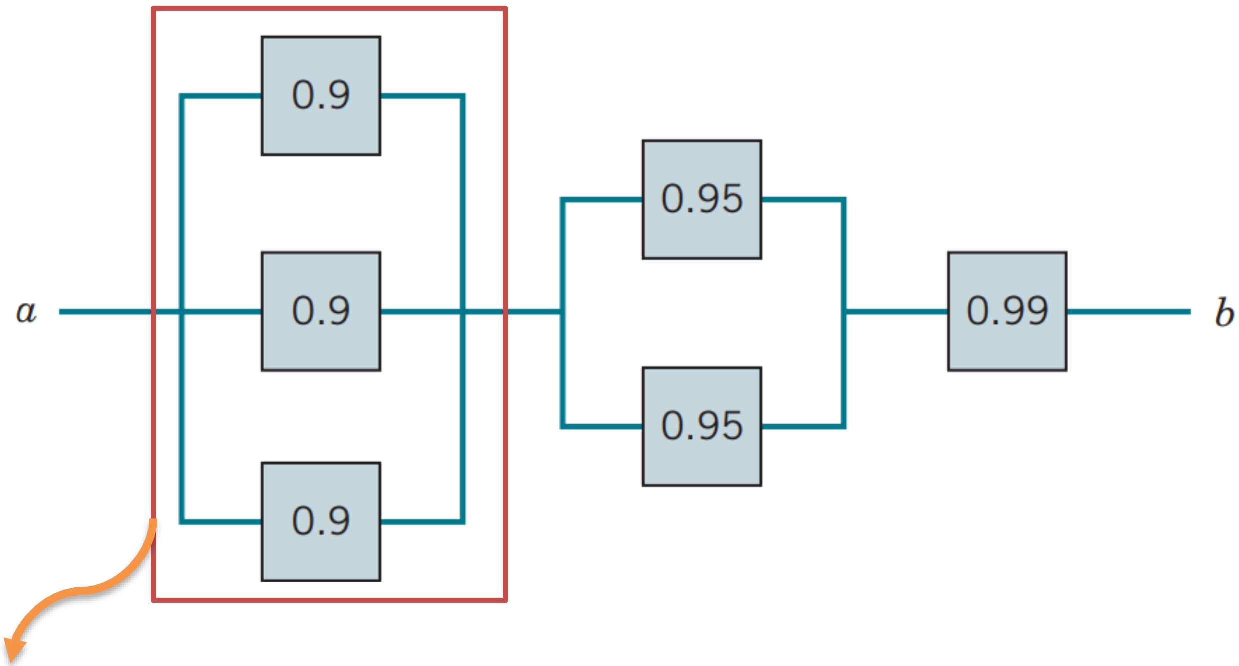
Example8:

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



Conditional Probability (22/31)

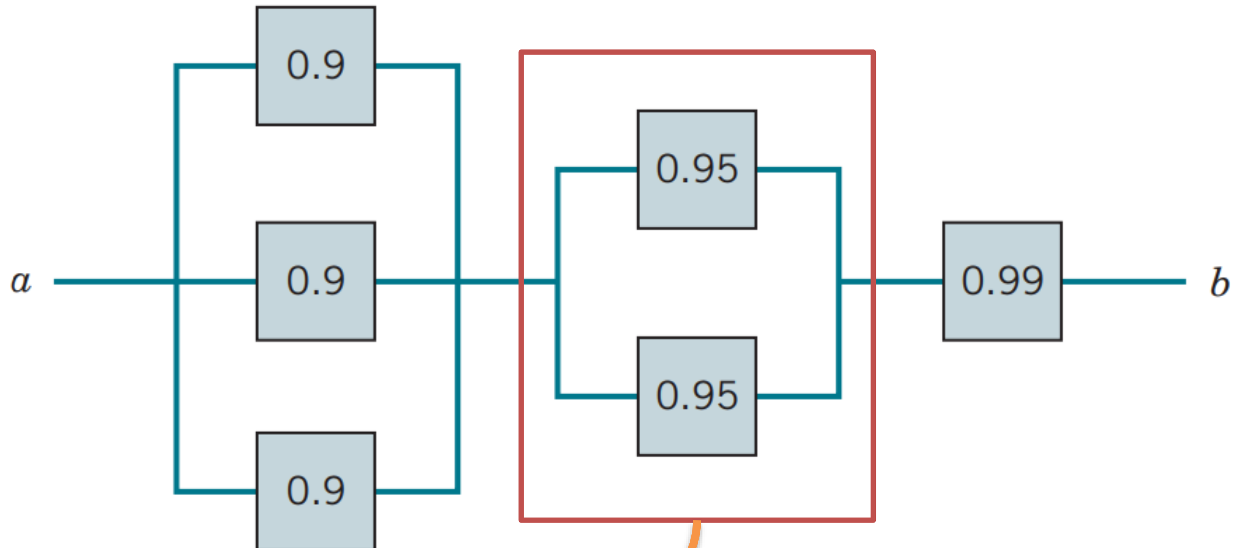
Example8:



$$= 1 - (0.10)(0.10)(0.10) = 0.999$$

Conditional Probability (23/31)

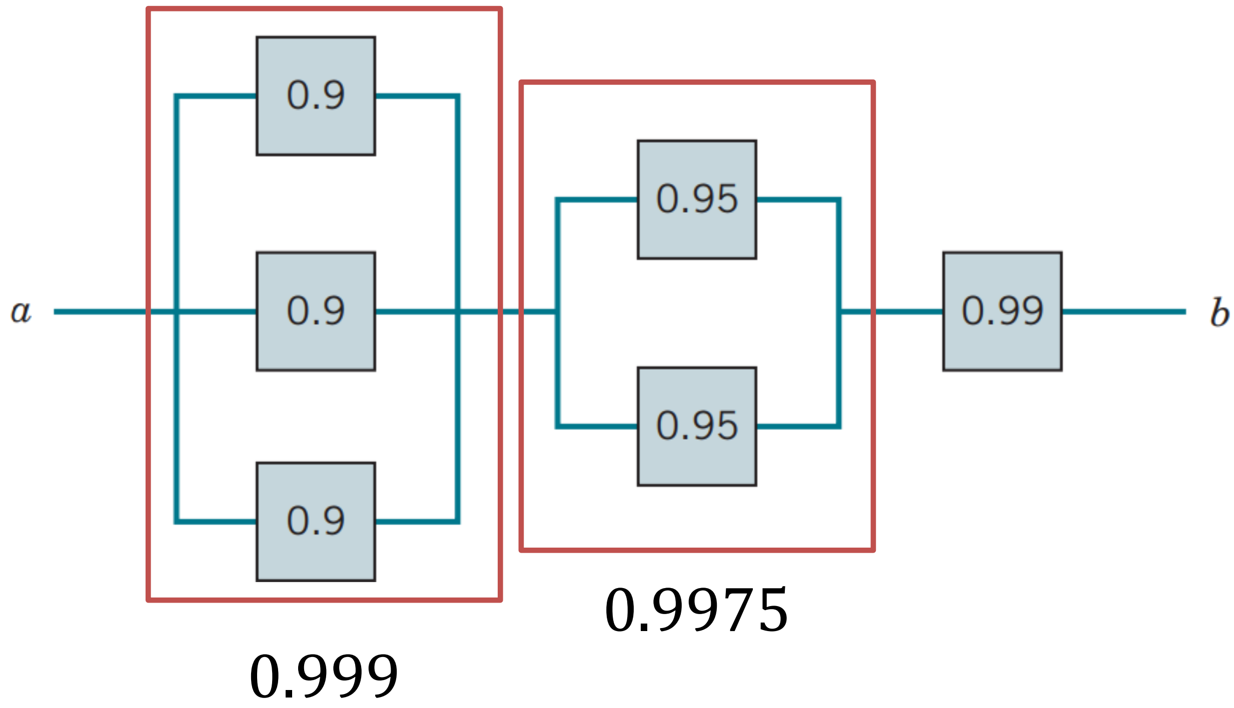
Example8:



$$= 1 - (0.05)(0.05) = 0.9975$$

Conditional Probability (24/31)

Example8:



$$= (0.999)(0.9975)(0.99) = 0.9865$$

Conditional Probability (25/31)

Multiplication Rule:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad , \qquad for \;\; P(B) > 0$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad , \qquad for \;\; P(A) > 0$$

$$\boxed{\boldsymbol{P(A \cap B) = \; P(A|B)P(B) = P(B|A)P(A)}}$$

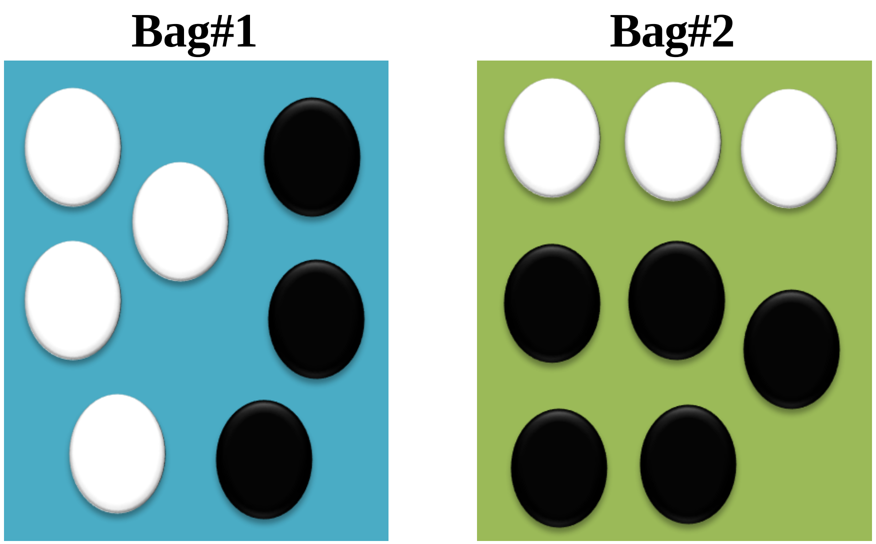
Example9:

One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?

Conditional Probability (27/31)

Example9:

One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?



Solution:

B_1: Black from bag#1

W_1: White from bag#1

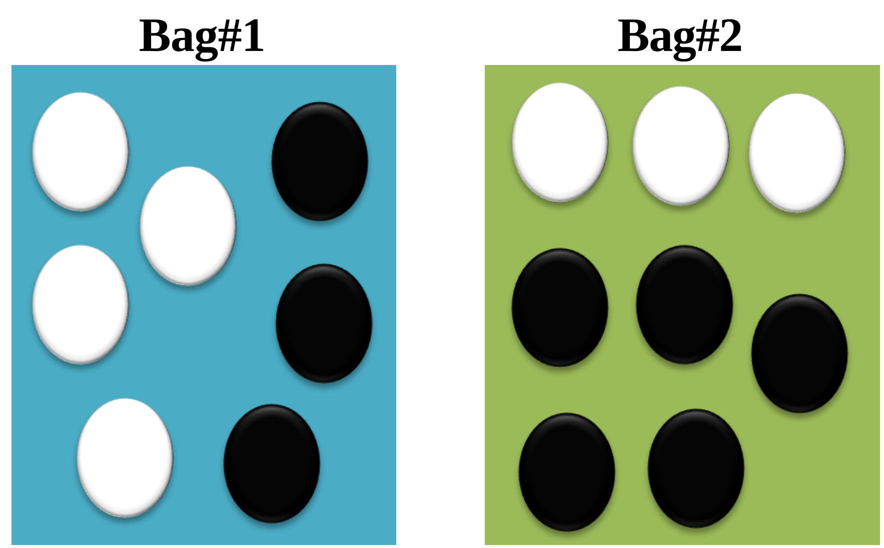
B_2: Black from bag#2

W_2: White from bag#2

Conditional Probability (27/31)

Example9: B_1 and B_2 or W_1 and B_2

One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?



Solution:

B_1: Black from bag#1

W_1: White from bag#1

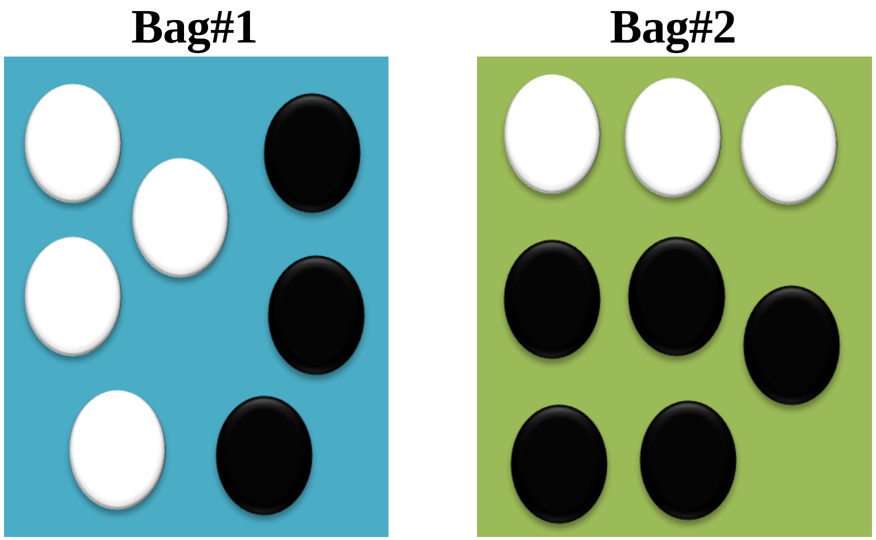
B_2: Black from bag#2

W_2: White from bag#2

Conditional Probability (28/31)

Example9: $\boxed{B_1 \text{ and } B_2}$ or W_1 and B_2

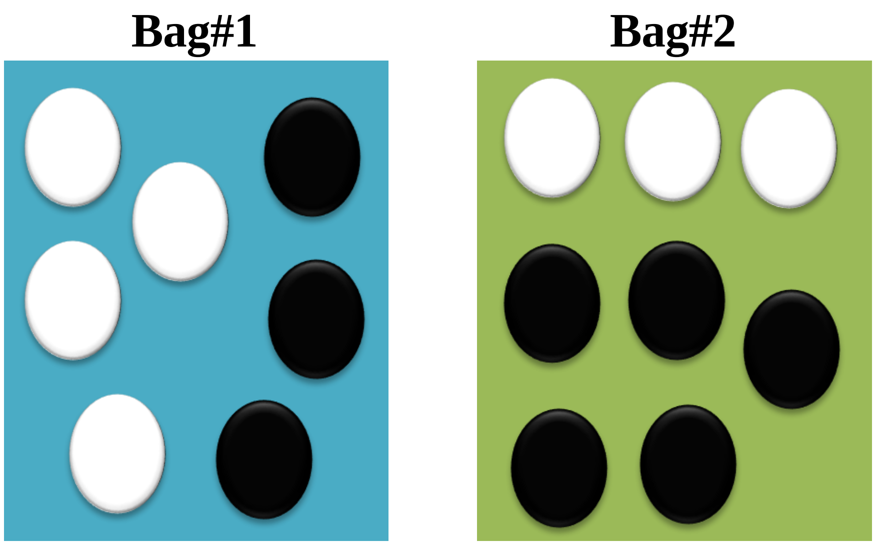
$$P(B_1 \cap B_2) = P(B_2|B_1)P(B_1) = \left(\frac{6}{9}\right)\left(\frac{3}{7}\right)$$



Conditional Probability (29/31)

Example9: $\boxed{B_1 \text{ and } B_2}$ or $\boxed{W_1 \text{ and } B_2}$

$$P(W_1 \cap B_2) = P(B_2|W_1)P(W_1) = \left(\frac{5}{9}\right)\left(\frac{4}{7}\right)$$



Conditional Probability (30/31)

Example9: B_1 and B_2 or W_1 and B_2 **Disjoint**

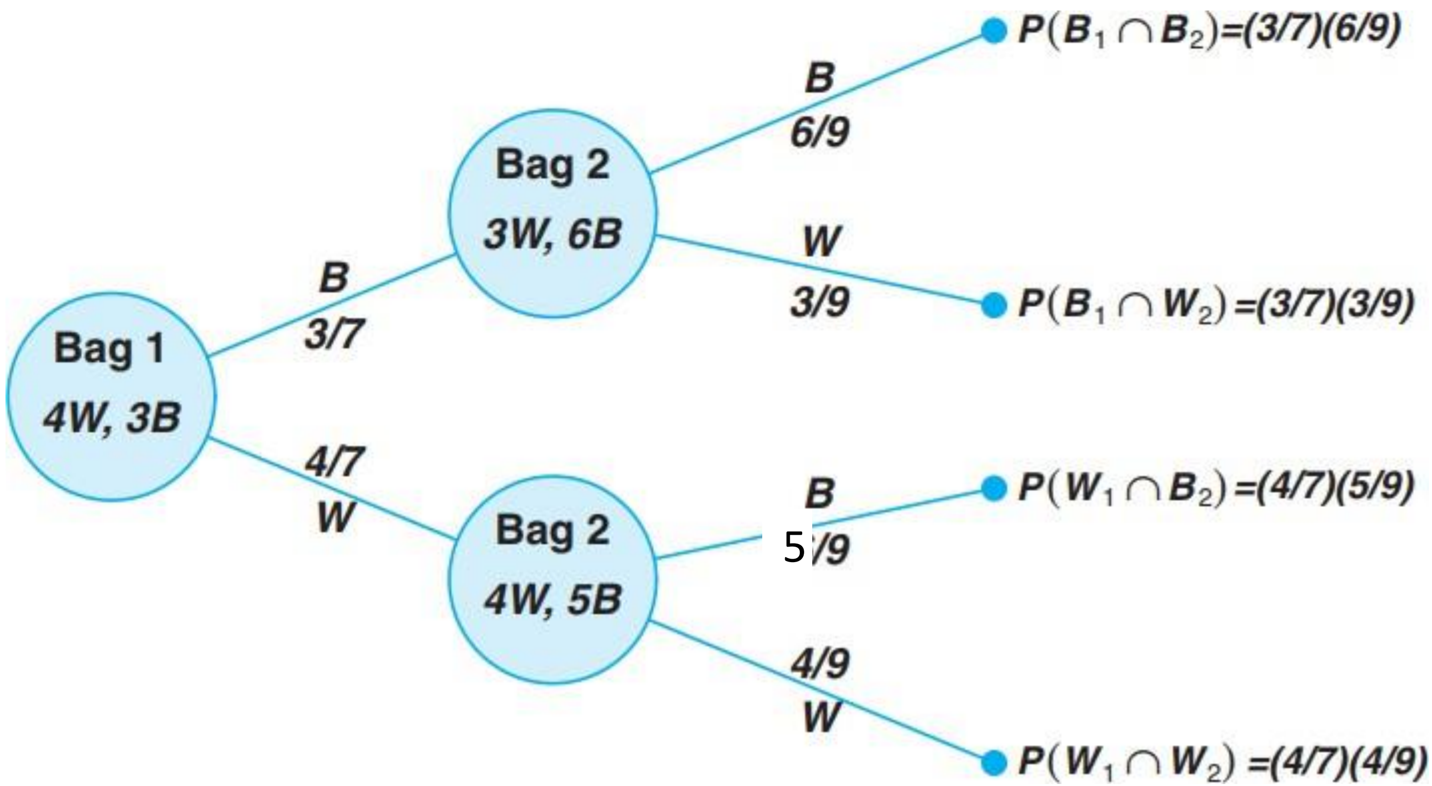
$$P(B_1 \cap B_2) = P(B_2|B_1)P(B_1) = \left(\frac{6}{9}\right)\left(\frac{3}{7}\right)$$

$$P(W_1 \cap B_2) = P(B_2|W_1)P(W_1) = \left(\frac{5}{9}\right)\left(\frac{4}{7}\right)$$

What is the probability that a ball now drawn from the second bag is black = $\left(\frac{6}{9}\right)\left(\frac{3}{7}\right) + \left(\frac{5}{9}\right)\left(\frac{4}{7}\right) = \frac{38}{63}$

Conditional Probability (31/31)

Example9: B_1 and B_2 or W_1 and B_2 **Disjoint**

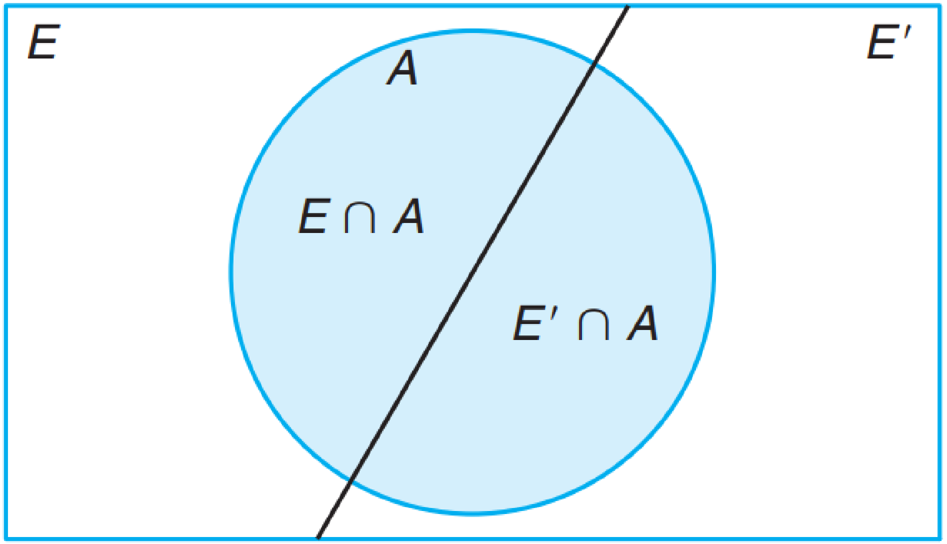


Chapter 1: Probability

- Sample Space.
- Events.
- Counting Techniques.
- Probability of an Event.
- Additive Rules.
- Conditional Probability.
- Independence, and the Product Rule.
- Total Probability Rule.
- Bayes’ Rule.

Total Probability Rule (1/10)

Total Probability Rule:



$$
\begin{aligned}
P(A) &= P(E \cap A) \cup P(E' \cap A) \\
&= P(E \cap A) + P(E' \cap A) \\
&= P(A|E)P(E) + P(A|E')P(E')
\end{aligned}
$$

Total Probability Rule (2/10)

Example10:

Consider the information about contamination in the following table.

$P(F|H) = 0.1$

$P(F|H') = 0.005$

$P(H) = 0.2$

$P(H') = 0.8$

$P(F)$?

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not high	0.8

Let F denote the event that the product fails, and let H denote the event that the chip is exposed to high levels of contamination.

Total Probability Rule (3/10)

Example10:

Consider the information about contamination in the following table.

$P(F|H) = 0.1$

$P(F|H') = 0.005$

$P(H) = 0.2$

$P(H') = 0.8$

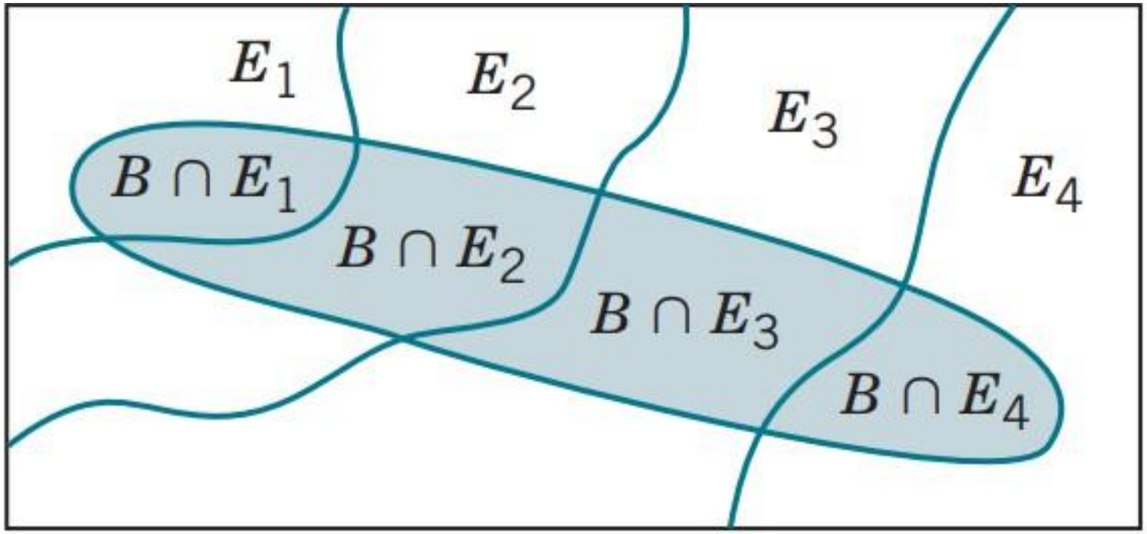
$P(F)$?

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not high	0.8

$$P(F) = P(F|H)P(H) + P(F|H')P(H')$$
$$= (0.1)(0.2) + (0.005)(0.8) = 0.024$$

Total Probability Rule (4/10)

Total Probability Rule (Multiple Events):



$$B = (B \cap E_1) \cup (B \cap E_2) \cup (B \cap E_3) \cup (B \cap E_4)$$

$$P(B) = P(B \cap E_1) + P(B \cap E_2) + P(B \cap E_3) + P(B \cap E_4)$$

Total Probability Rule (5/10)

Example11:

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Total Probability Rule (5/10)

Example11:

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

***D*: the product is defective. Find $P(D)$?**

Total Probability Rule (6/10)

Example11: $P(B_1) = 0.3, P(B_2) = 0.45, P(B_3) = 0.25$

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

$$P(D|B_1) = 0.02,$$
$$P(D|B_2) = 0.03,$$
$$P(D|B_3) = 0.02.$$

Total Probability Rule (7/10)

Example11: $P(B_1) = 0.3, P(B_2) = 0.45, P(B_3) = 0.25$

$P(D|B_1) = 0.02,$
$P(D|B_2) = 0.03,$
$P(D|B_3) = 0.02.$

Applying the total probability rule , we can write

$$P(D)$$

$$= P(D|B_1)P(B_1) + P(D|B_2)P(B_2) + P(D|B_3)P(B_3)$$

$$= 0.02(0.3) + 0.03(0.45) + 0.02(0.25) = 0.0245$$

Total Probability Rule (8/10)

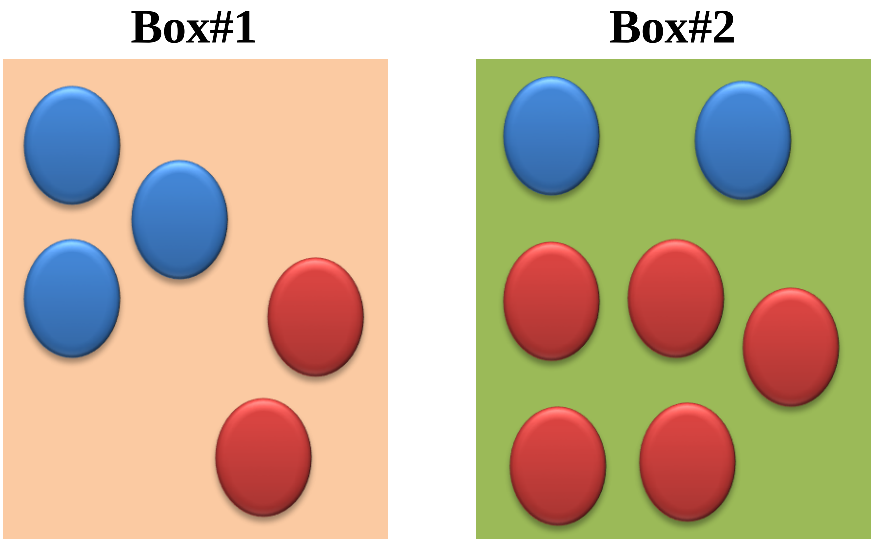
Example12:

Box#1 contains 2 red balls and 3 blue balls; Box#2 contains 5 red balls and 2 blue balls. If the selection of two boxes is equally likely, and you selected one ball, what is the probability that it is red?

Total Probability Rule (8/10)

Example12:

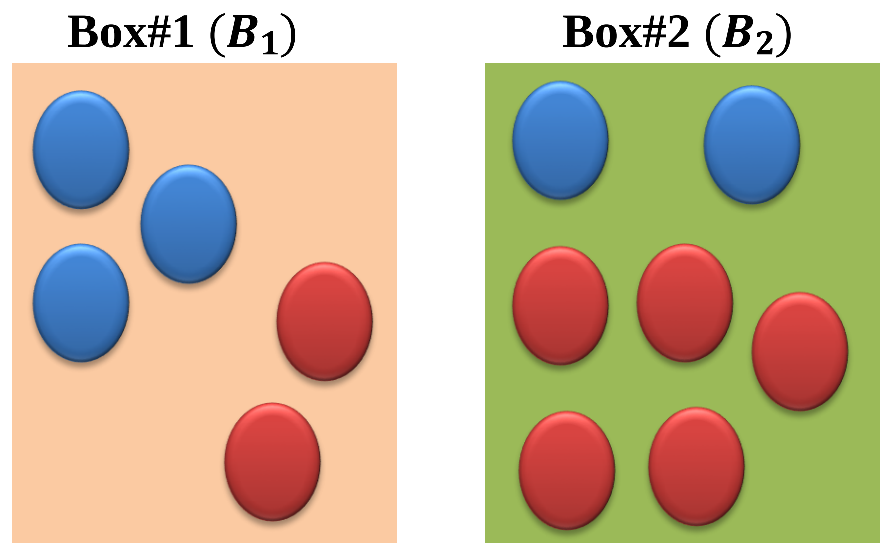
Box#1 contains 2 red balls and 3 blue balls; Box#2 contains 5 red balls and 2 blue balls. If the selection of two boxes is equally likely, and you selected one ball, what is the probability that it is red?



Total Probability Rule (9/10)

Example12:

Box#1 contains 2 red balls and 3 blue balls; Box#2 contains 5 red balls and 2 blue balls**. If the selection of two boxes is equally likely**, and you selected one ball, what is the probability that it is red?



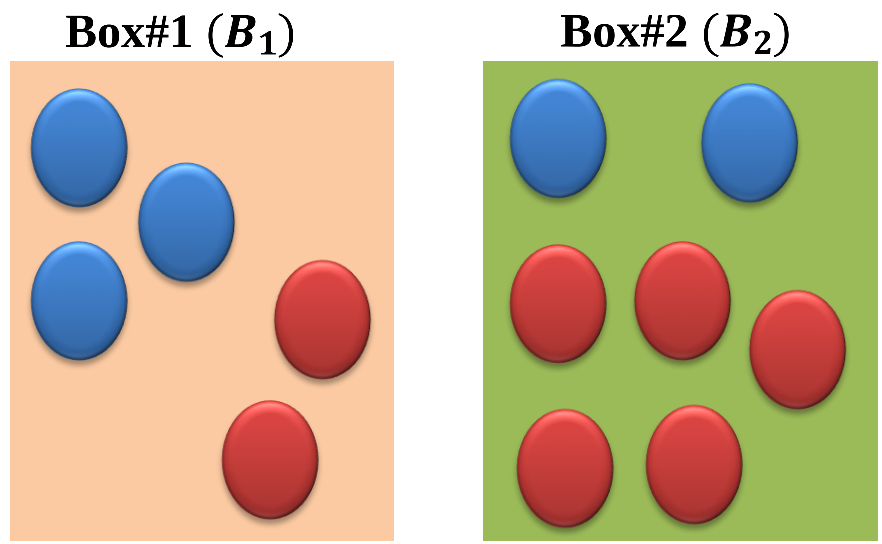
$P(B_1) = P(B_2) = 0.5$

$R: read,$ $B: blue$

Find $P(R)$?

Total Probability Rule (10/10)

Example12:



$P(B_1) = P(B_2) = 0.5$

$R: read, \qquad B: blue$

$P(R|B_1) = {}^{2}/_{5} = 0.4$

$P(R|B_2) = {}^{5}/_{7} = 0.7143$

$$P(R) = P(R|B_1)P(B_1) + P(R|B_2)P(B_2)$$

$$= (0.4)(0.5) + (0.7143)(0.5) = 0.55715$$

Bayes' Rule (1/11)

From the definition of conditional probability,

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Now, considering the second and last terms in the preceding expression, we can write

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad \text{for} \quad P(B) > 0$$

Bayes' Rule (2/11)

If $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P(E_1 \mid B) = \frac{P(B \mid E_1)P(E_1)}{P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k)}$$

for $P(B) > 0$

Bayes' Rule (3/11)

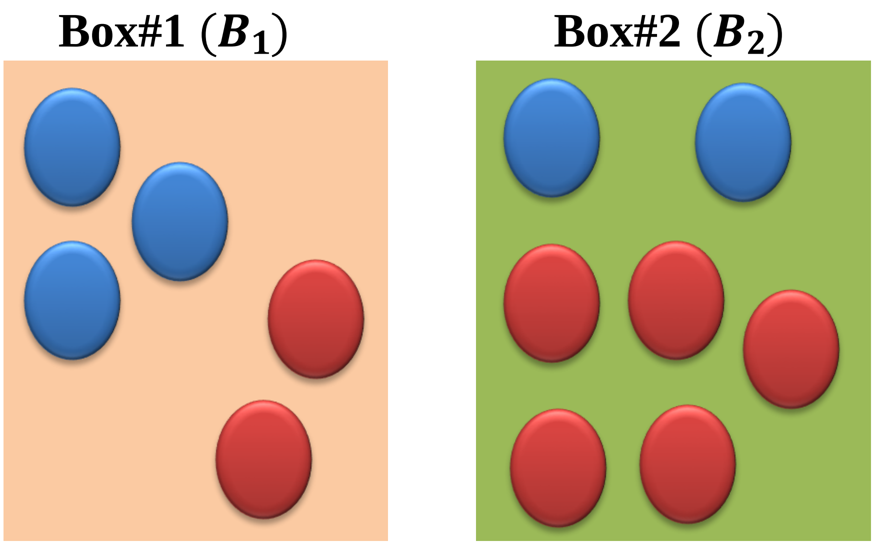
Example1:

Box#1 contains 2 red balls and 3 blue balls; Box#2 contains 5 red balls and 2 blue balls. If the selection of two boxes is equally likely, and the selected ball was red, what is the probability that it is from Box#1?

$P(B_1) = P(B_2) = 0.5$

$R: read, \qquad B: blue$

Find $\boldsymbol{P(B_1|R)}$ **?**



Bayes' Rule (4/11)

Example1:

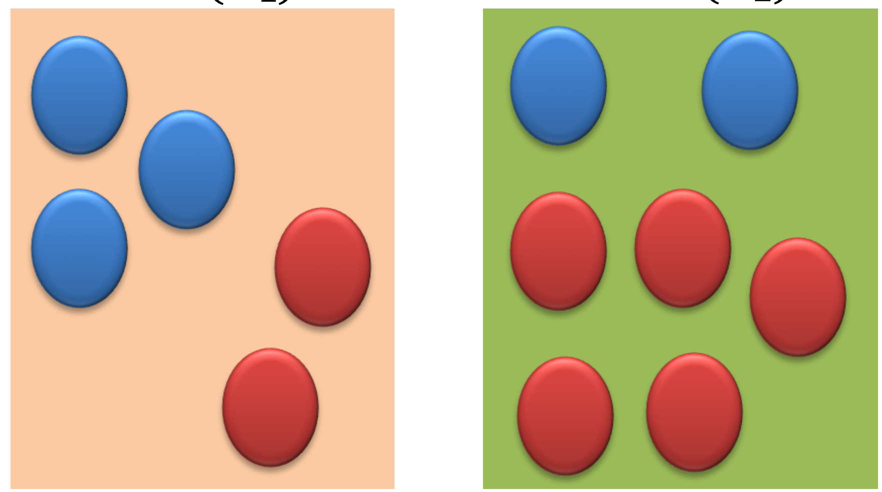
$P(B_1) = P(B_2) = 0.5$

$R: read, \qquad B: blue$

$P(R|B_1) = {}^{2}/_{5} = 0.4$

$P(R|B_2) = {}^{5}/_{7} =$
0.7143

Box#1 (B_1) **Box#2 (B_2)**



Bayes' Rule (5/11)

Example1:

$P(B_1) = P(B_2) = 0.5$

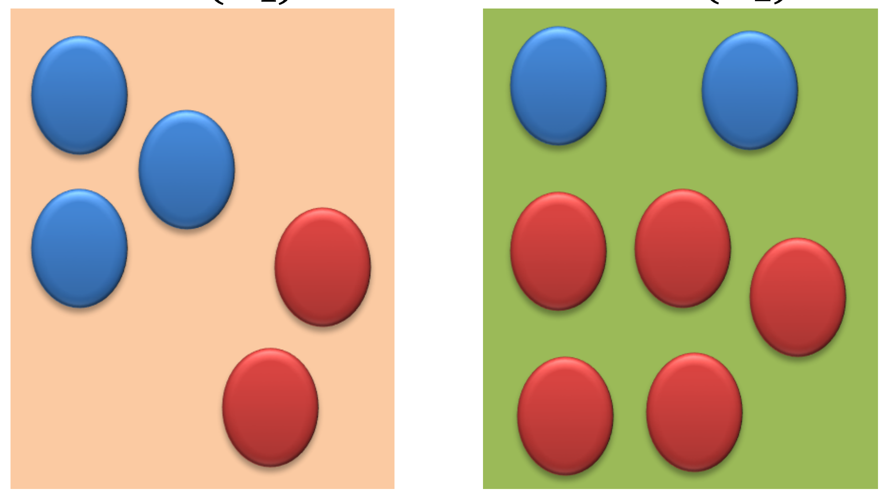
$R: read, \qquad B: blue$

$P(R|B_1) = {}^{2}/_{5} = 0.4$

$P(R|B_2) = {}^{5}/_{7} = 0.7143$

$$P(B_1|R) = \frac{P(R|B_1)P(B_1)}{P(R)} = \frac{(0.4)(0.5)}{P(R)}$$

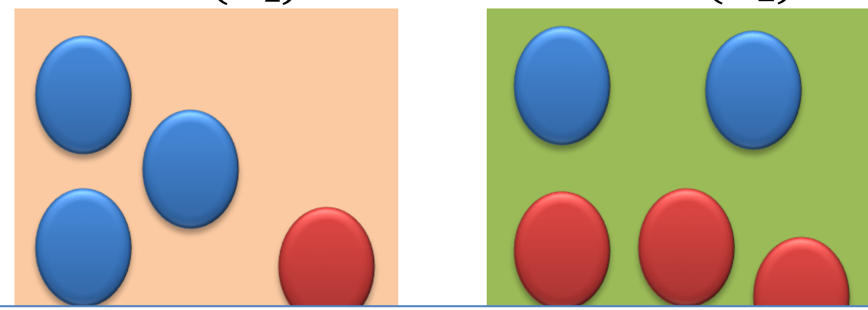
Box#1 (B_1) **Box#2 (B_2)**



Bayes' Rule (6/11)

Example1:

Box#1 (B_1) **Box#2 (B_2)**



$P(B_1) = P(B_2) = 0.5$

$R: read, \qquad B: blue$

$P(R|B_1) = 2$
0.4

$P(R|B_2) = 5$
0.7143

$$P(R) = P(R|B_1)P(B_1) + P(R|B_2)P(B_2)$$
$$= (0.4)(0.5) + (0.7143)(0.5) = 0.55715$$

$$P(B_1|R) = \frac{P(R|B_1)P(B_1)}{P(R)} = \frac{(0.4)(0.5)}{P(R)}$$

Bayes' Rule (7/11)

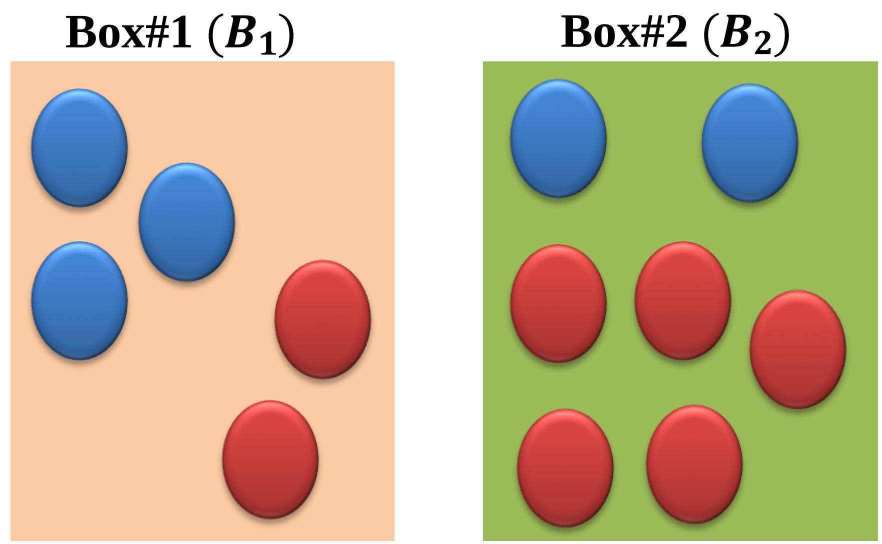
Example1:

$P(B_1) = P(B_2) = 0.5$

$R: read, \qquad B: blue$

$P(R|B_1) = {}^{2}/_{5} = 0.4$

$P(R|B_2) = {}^{5}/_{7} = 0.7143$



$$P(B_1|R) = \frac{P(R|B_1)P(B_1)}{P(R)} = \frac{(0.4)(0.5)}{P(R)} = \frac{0.2}{0.55715} = 0.35897$$

Bayes' Rule (8/11)

Example2:

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. If a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3?

***D*: the product is defective. Find $P(B_3|D)$?**

Bayes' Rule (8/11)

Example2: $P(B_1) = 0.3, P(B_2) = 0.45, P(B_3) = 0.25$

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. If a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3?

$$P(D|B_1) = 0.02,$$
$$P(D|B_2) = 0.03,$$
$$P(D|B_3) = 0.02.$$

Bayes' Rule (9/11)

Example2:

$$P(B_1) = 0.3, P(B_2) = 0.45, P(B_3) = 0.25$$

$$P(D|B_1) = 0.02,$$
$$P(D|B_2) = 0.03,$$
$$P(D|B_3) = 0.02.$$

Using Bayes' rule to write

$$P(B_3|D) = \frac{P(D|B_3)P(B_3)}{P(D)}$$

Bayes' Rule (10/11)

Example2:

$P(B_1) = 0.3, P(B_2) = 0.45, P(B_3) = 0.25$

$P(D|B_1) = 0.02,$
$P(D|B_2) = 0.03,$
$P(D|B_3) = 0.02,$

Applying the total probability rule, we can write:

$$P(D) = P(D|B_1)P(B_1) + P(D|B_2)P(B_2) + P(D|B_3)P(B_3)$$
$$= 0.02(0.3) + 0.03(0.45) + 0.02(0.25) = 0.0245$$

Using Bayes' rule to write

$$P(B_3|D) = \frac{P(D|B_3)P(B_3)}{P(D)}$$

Bayes' Rule (11/11)

Example2: $P(B_1) = 0.3, P(B_2) = 0.45, P(B_3) = 0.25$

$$P(D|B_1) = 0.02,$$
$$P(D|B_2) = 0.03,$$
$$P(D|B_3) = 0.02.$$

Using Bayes' rule to write

$$P(B_3|D) = \frac{P(D|B_3)P(B_3)}{0.0245} = \frac{(0.02)(0.25)}{0.0245} = 0.2041$$

Chapter 2: Random Variable

- Discrete Random Variables.
- Probability Mass Functions.
- Cumulative Distribution Functions.
- Discrete R.V. (Mean and Variance).
- Continuous Random Variables.
- Probability Density Functions.
- Continuous R.V. (Mean and Variance).
- Joint Probability Distributions.

Discrete Random Variables (1/3)

Random Variable

- Is a function that assigns a real number to each outcome in the sample space of random experiment. Denoted by an uppercase letter such as X

A Discrete Random Variable

- Is a random variable with a finite (or countable infinite) range.

- The possible values of X may be listed as $x_1, x_2, \ldots$

Discrete Random Variables (2/3)

Example1

- Flipping a coin of two times. Let X is the number of heads.

Discrete Random Variables (3/3)

Example1

- Flipping a coin of two times. Let X is the number of heads.

Answer:

$$S = \{HH, HT, TH, TT\}$$

2 1 1 0

$$x = 0, 1, 2$$

$$P(0) = \frac{1}{4}, \qquad P(1) = \frac{2}{4}, \qquad P(2) = \frac{1}{4}$$

Probability Mass Function

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$

x_i	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$	$\boldsymbol{x_5}$
$f(x_i) = P(x_i)$	$\boldsymbol{P(x_1)}$	$\boldsymbol{P(x_2)}$	$\boldsymbol{P(x_3)}$	$\boldsymbol{P(x_4)}$	$\boldsymbol{P(x_5)}$

Probability Mass Fun. (2/14)

Example1

Verify that the function is a probability mass function:

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Probability Mass Fun. (3/14)

Example1

Verify that the function is a probability mass function:

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Answer:

$\sum P(x_i) = 1$, $\quad P(x_i) \geq 0$

Probability Mass Fun. (4/14)

Example2

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Find:

a. $P(X \leq 2)$

b. $P(X > -2)$

c. $P(-1 \leq X \leq 1)$

d. $P(X \leq -1 \text{ or } X = 2)$

Probability Mass Fun. (5/14)

Example2

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Answer:

a. $P(X \le 2) = 1$

b. $P(X > -2) = \frac{7}{8}$

c. $P(-1 \le X \le 1) = \frac{6}{8}$

d. $P(X \le -1 \text{ or } X = 2) = \frac{3}{8} + \frac{1}{8} = \frac{4}{8}$

Probability Mass Fun. (6/14)

Example3

Two balls are drawn in succession without replacement from a box containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y , where y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Probability Mass Fun. (7/14)

Example3

Sample Space	y
RR	2
RB	1
BR	1
BB	0

No Red Balls

$$f(0) = P(Y = 0) = \frac{\binom{4}{0}\binom{3}{2}}{\binom{7}{2}} = \frac{3}{21} = \frac{1}{7}$$

Probability Mass Fun. (7/14)

Example3

Sample Space	y
RR	2
RB	1
BR	1
BB	0

One Red Ball

$$f(1) = P(Y = 1) = \frac{\binom{4}{1}\binom{3}{1}}{\binom{7}{2}} = \frac{12}{21} = \frac{4}{7}$$

Probability Mass Fun. (7/14)

Example3

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Two Red Balls

$$f(2) = P(Y = 2) = \frac{\binom{4}{2}\binom{3}{0}}{\binom{7}{2}} = \frac{6}{21} = \frac{2}{7}$$

Probability Mass Fun. (7/14)

Example3

Sample Space	y
RR	2
RB	1
BR	1
BB	0

y	0	1	2
$f(y) = P(Y = y)$	1/7	4/7	2/7

Example4

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Note: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school.

Probability Mass Fun. (8/14)

Example4

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Note: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. **Then x can only take the numbers 0, 1, and 2.**

Example4

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{136}{190}$$

Probability Mass Fun. (10/14)

Example4

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

$$f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$$

Example4

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$$

Example4

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

x	0	1	2
$f(x) = P(X = x)$	136/190	51/190	3/190

Probability Mass Fun. (13/14)

Example5

There is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$.

Suppose that the probabilities are

$$P(X = 0) = 0.6561 \quad P(X = 1) = 0.2916$$
$$P(X = 2) = 0.0486 \quad P(X = 3) = 0.0036$$
$$P(X = 4) = 0.0001$$

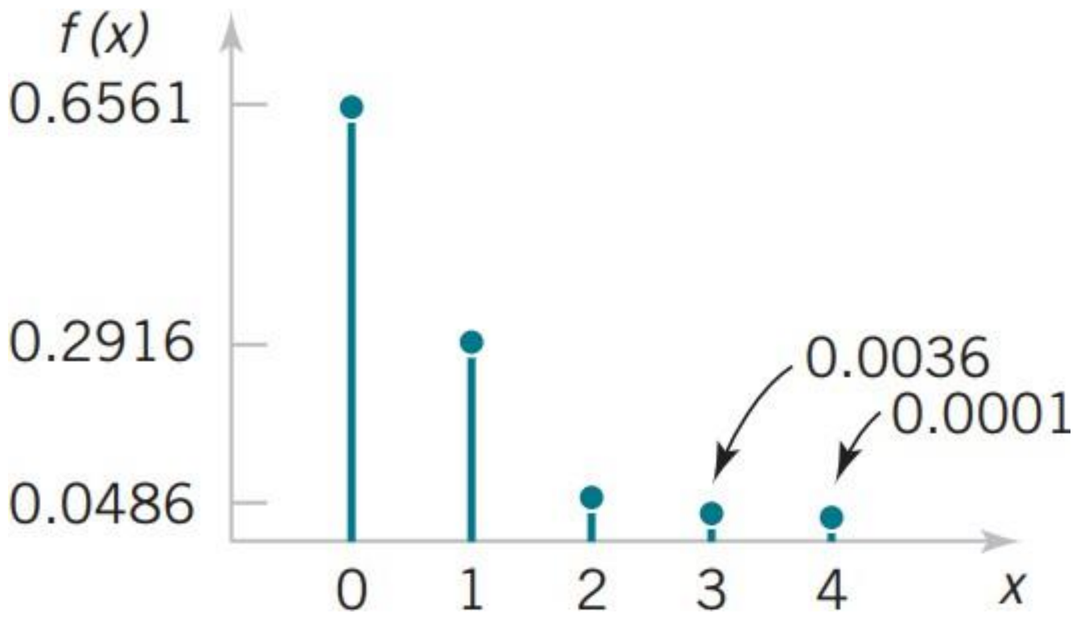
Probability Mass Fun. (14/14)

Example5

$P(X = 0) = 0.6561 \quad P(X = 1) = 0.2916$

$P(X = 2) = 0.0486 \quad P(X = 3) = 0.0036$

$P(X = 4) = 0.0001$



Cumulative Distribution (1/9)

The cumulative distribution function (cdf), denoted by $F(x)$, measures the probability that the random variable X assumes a value less than or equal to x, that is,

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

Cumulative Distribution (2/9)

If X is discrete, then

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

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Cumulative Distribution (3/9)

If X is discrete, then

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8
$\boldsymbol{F(x) = P(X \le x)}$	**1/8**	**3/8**	**5/8**	**7/8**	**8/8**

Cumulative Distribution (4/9)

Example1

$$P(X=0)=0.6561 \quad P(X=1)=0.2916$$
$$P(X=2)=0.0486 \quad P(X=3)=0.0036$$
$$P(X=4)=0.0001$$

x	0	1	2	3	4
$f(x)=P(X=x)$	0.6561	0.2916	0.0486	0.0036	0.0001

Cumulative Distribution (5/9)

Example1

$$P(X = 0) = 0.6561 \quad P(X = 1) = 0.2916$$
$$P(X = 2) = 0.0486 \quad P(X = 3) = 0.0036$$
$$P(X = 4) = 0.0001$$

x	0	1	2	3	4
$f(x) = P(X = x)$	0.6561	0.2916	0.0486	0.0036	0.0001
$\boldsymbol{F(x) = P(X \le x)}$	**0.6561**	**0.9477**	**0.9963**	**0.9999**	**1**

Cumulative Distribution (6/9)

Example1

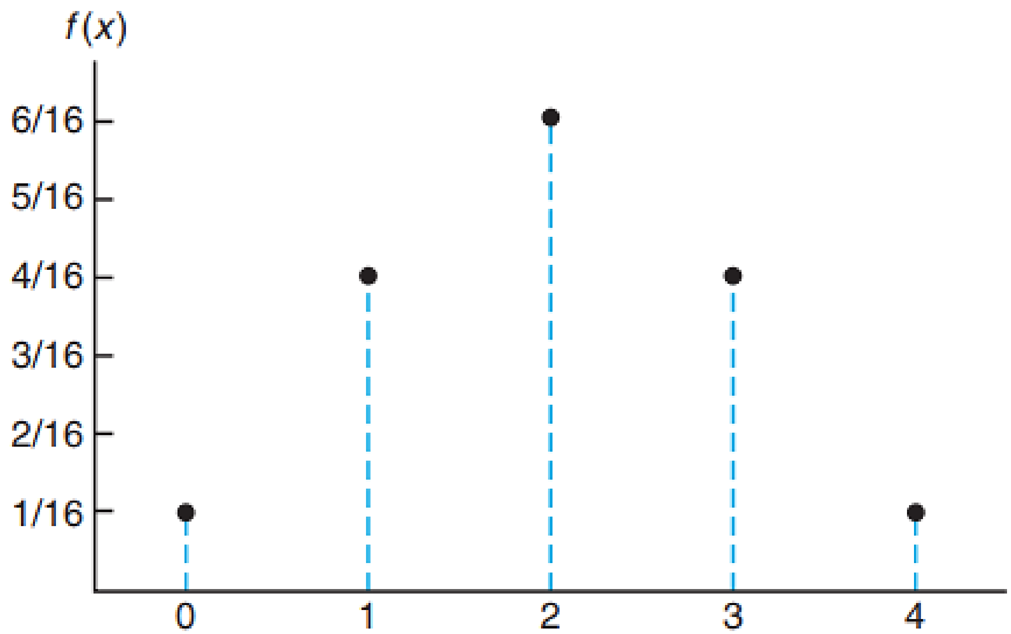
x	0	1	2	3	4
$f(x) = P(X = x)$	0.6561	0.2916	0.0486	0.0036	0.0001
$\boldsymbol{F(x) = P(X \leq x)}$	**0.6561**	**0.9477**	**0.9963**	**0.9999**	**1**

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.6561 & 0 \leq x < 1 \\ 0.9477 & 1 \leq x < 2 \\ 0.9963 & 2 \leq x < 3 \\ 0.9999 & 3 \leq x < 4 \\ 1 & 4 \leq x \end{cases}$$

Cumulative Distribution (7/9)

Example2

x	0	1	2	3	4
$f(x) = P(X = x)$	1/16	4/16	6/16	4/16	1/16



Probability mass function plot.

Cumulative Distribution (8/9)

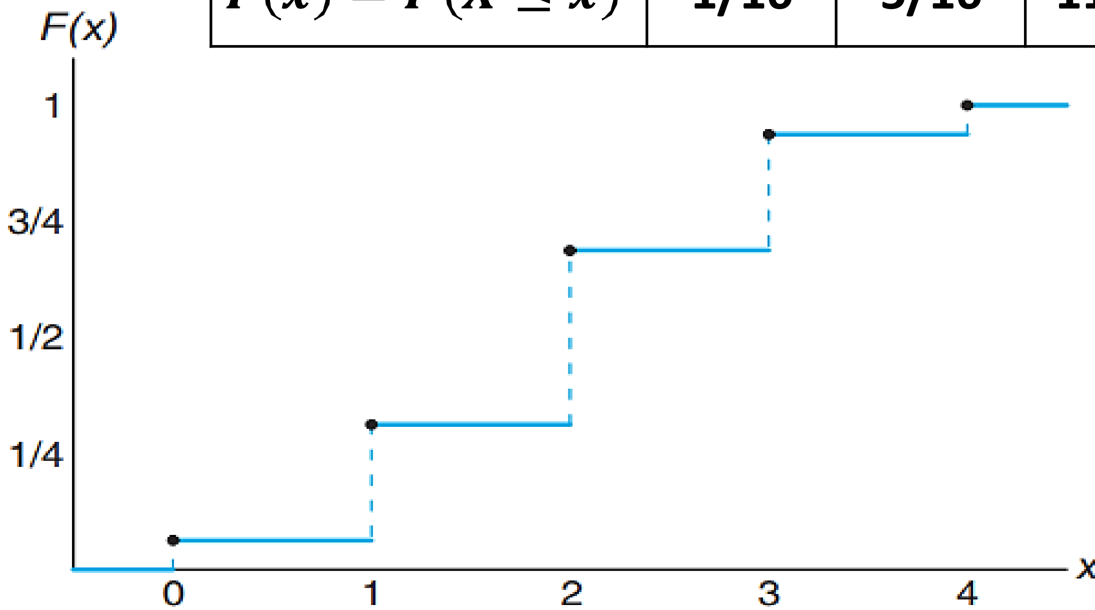
Example2

x	0	1	2	3	4
$f(x) = P(X = x)$	1/16	4/16	6/16	4/16	1/16
$\boldsymbol{F(x) = P(X \leq x)}$	**1/16**	**5/16**	**11/16**	**15/16**	**16/16**

Cumulative Distribution (9/9)

Example2

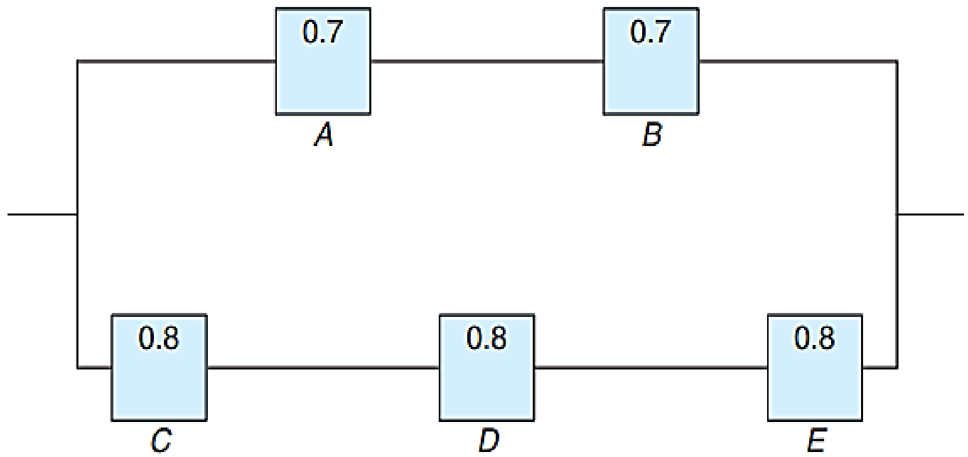
x	0	1	2	3	4
$f(x) = P(X = x)$	1/16	4/16	6/16	4/16	1/16
$\boldsymbol{F(x) = P(X \leq x)}$	**1/16**	**5/16**	**11/16**	**15/16**	**16/16**



Discrete cumulative distribution function.

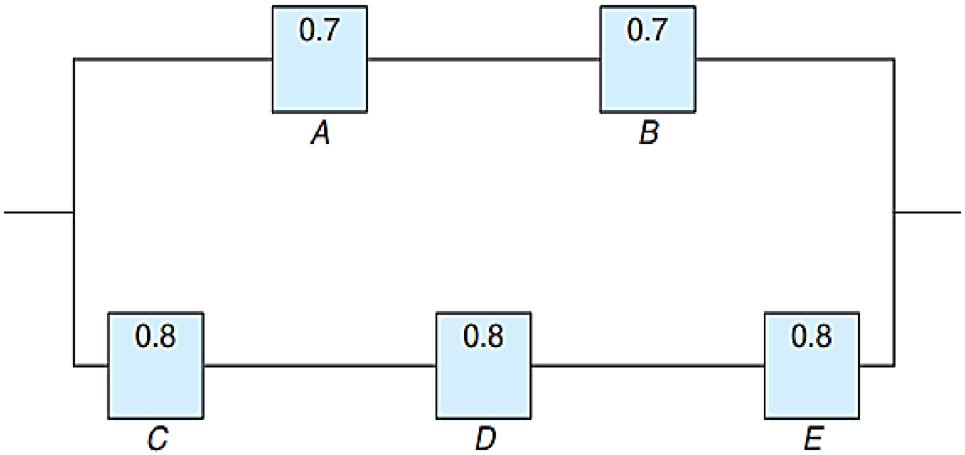
Quiz (1) – Q1

1. A circuit system is given in the following figure. Assume the components fail independently. What is the probability that the entire system works?



Quiz (1) – Q1

1. A circuit system is given in the following figure. Assume the components fail independently. What is the probability that the entire system works?



<u>UP:</u> (0.7) (0.7) = 0.49

<u>Down:</u> (0.8) (0.8) (0.8) = 0.512

$P(W)$ = 1 – [(0.51) (0.488)] = 0.75112

Quiz (1) – Q2

2. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many ways can a student check off one answer to each question and get all the answers wrong?

Quiz (1) – Q2

2. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many ways can a student check off one answer to each question and get all the answers wrong?

Ways to get all the answers wrong $= (3)^6 = 729$

Quiz (1) – Q3

3. If A, and B are two independent events. And $P(A) = 0.2$, $P(A \cap B) = 0.06$, determine the following probabilities:

   a) $P(A|B)$

   b) $P(B|A)$

Quiz (1) – Q3

3. If A, and B are two independent events. And $P(A) = 0.2$, $P(A \cap B) = 0.06$, determine the following probabilities:

a) $P(A|B)$

$$P(A|B) = P(A) = 0.2$$

b) $P(B|A)$

$$P(B|A) = P(\mathrm{B}) = \frac{P(A \cap B)}{P(A)} = \frac{0.06}{0.2} = 0.3$$

Quiz (1) – Q4

4. In how many ways can 3 boys and 4 girls sit in a row?

Quiz (1) – Q4

4. In how many ways can 3 boys and 4 girls sit in a row?

$$\text{number of ways} = \frac{n!}{n_1!\, n_2!} = \frac{7!}{3!\, 4!} = 35$$

Quiz (1) – Q5

5. The likely location of a mobile device in the home is as follows:

Adult bedroom: 0.10, Child bedroom: 0.20,

Office: 0.40, Other rooms: 0.30

(a) What is the probability that a mobile device is in a bedroom?

(b) What is the probability that it is not in a bedroom?

Quiz (1) – Q5

5. The likely location of a mobile device in the home is as follows:

Adult bedroom: 0.10, Child bedroom: 0.20,

Office: 0.40, Other rooms: 0.30

(a) What is the probability that a mobile device is in a bedroom?

$= P(Adult\ bedroom \cup Child\ bedroom) = 0.10 + 0.20 = 0.30$

(b) What is the probability that it is not in a bedroom?

$= 1 - 0.30 = 0.70$

Quiz (1) – Q6

6. How many distinct permutations can be made from the letters of the word COMPUTER?

Quiz (1) – Q6

6. How many distinct permutations can be made from the letters of the word COMPUTER?

$$n! = 8! = 40{,}320$$

Quiz (1) – Q7

7. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many different ways can a student check off one answer to each question?

Quiz (1) – Q7

7. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many different ways can a student check off one answer to each question?

$$= (4)^6 = 4{,}096$$

Chapter 2: Random Variable

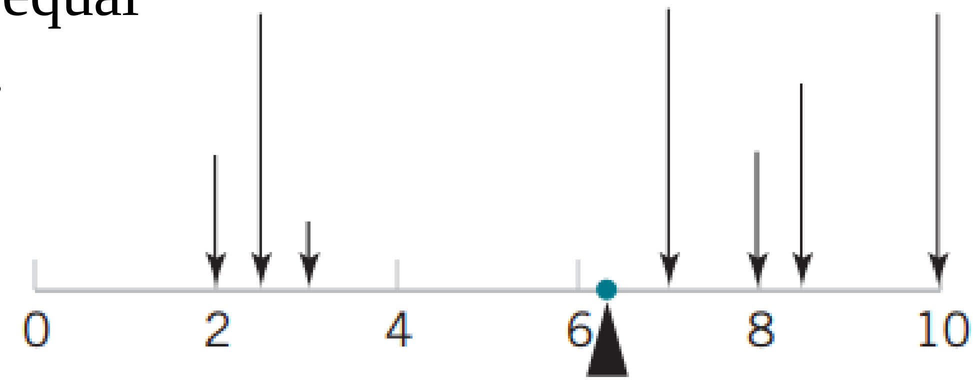
- Discrete Random Variables.
- Probability Mass Functions.
- Cumulative Distribution Functions.
- Discrete R.V. (Mean and Variance).
- Continuous Random Variables.
- Probability Density Functions.
- Continuous R.V. (Mean and Variance).
- Joint Probability Distributions.

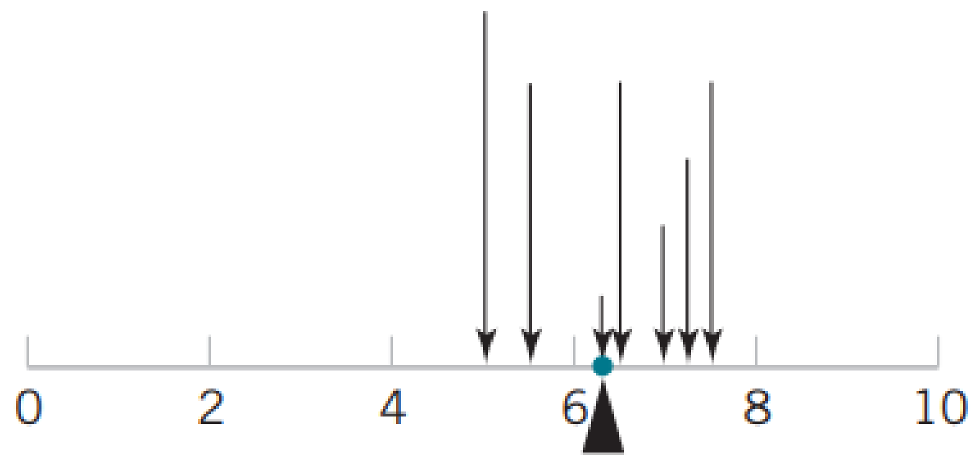
Mean and Variance (1/15)

Two numbers are often used to summarize a probability distribution for a random variable *X*. The mean is a measure of the center or middle of the probability distribution, and the variance is a measure of the dispersion, or variability in the distribution.

Mean and Variance (2/15)

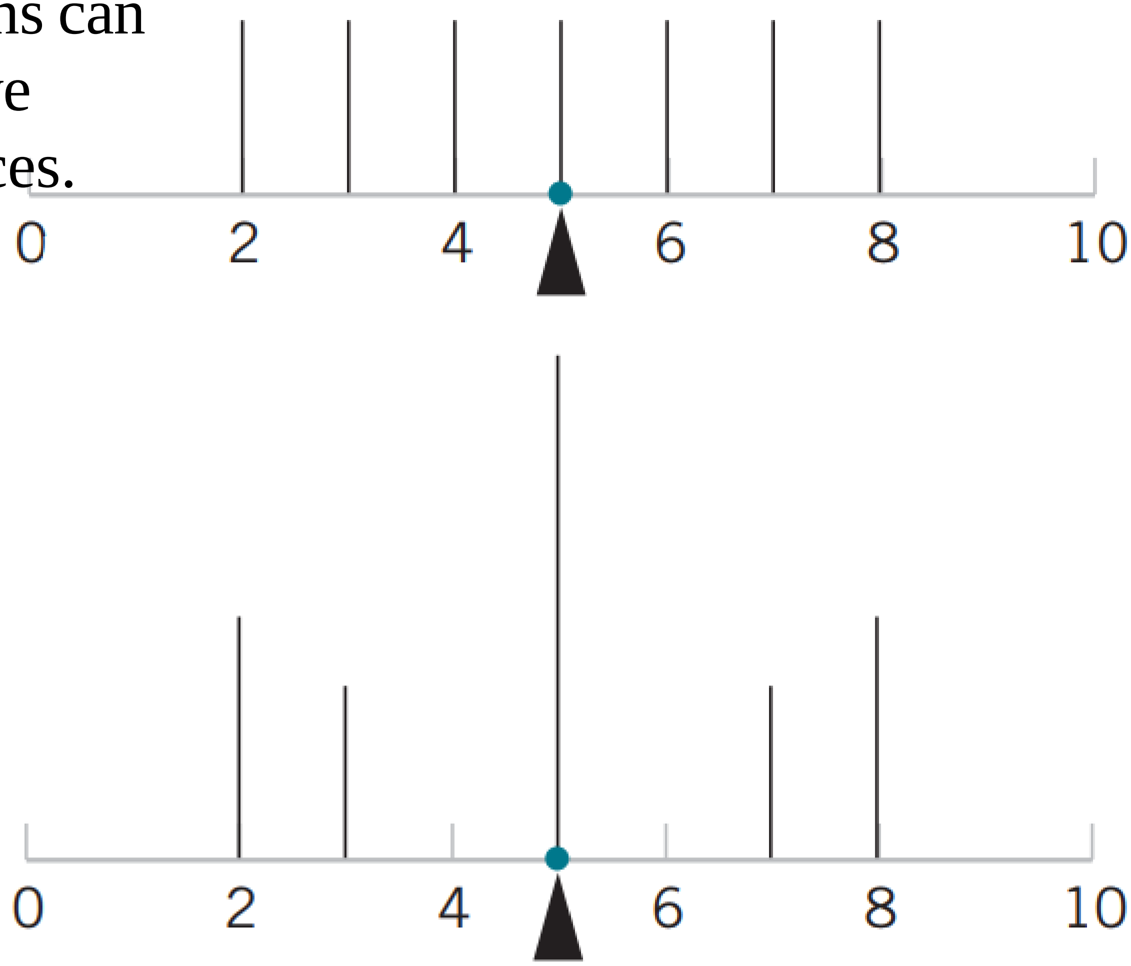
Probability distributions with equal means but different variances.





Mean and Variance (3/15)

Two probability distributions can differ even though they have identical means and variances.



Mean and Variance (4/15)

Mean, Variance, and Standard deviation

The **mean** or **expected value** of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x x f(x)$$

The **variance** of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X - \mu)^2 = \sum_x (x - \mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

$$\boxed{E(X^2) - \left(E(X)\right)^2}$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Mean and Variance (5/15)

Example1

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Find:

Determine the mean and variance of the random variable X

Mean and Variance (6/15)

Example1

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Find:

Determine the mean and variance of the random variable X

Answer: (1/2)

$E(X) =$

$$\sum x_i P(x_i) = (-2)\left(\frac{1}{8}\right) + (-1)\left(\frac{2}{8}\right) + (0)\left(\frac{2}{8}\right) + (1)\left(\frac{2}{8}\right) + (2)\left(\frac{1}{8}\right)$$

$$= 0$$

Mean and Variance (6/15)

Example1

$\boldsymbol{x}$	-2	-1	0	1	2
$\boldsymbol{f(x) = P(X = x)}$	1/8	2/8	2/8	2/8	1/8

Answer: (2/2)

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

$$E(X) = 0$$

$$E(X^2)$$

$$= \sum x_i^2 P(x_i) = (4)\left(\frac{1}{8}\right) + (1)\left(\frac{2}{8}\right) + (0)\left(\frac{2}{8}\right) + (1)\left(\frac{2}{8}\right) + (4)\left(\frac{1}{8}\right) = 1.5$$

$V(X) = 1.5 - (0)^2 = 1.5,$ Standard Deviation $(\sigma) = \sqrt{1.5}$

Mean and Variance (7/15)

Example2:

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Mean and Variance (8/15)

Example2 – Answer (1/9)

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Let X represent the number of good components in the sample. **Then $\boldsymbol{x}$ can only take the numbers 0, 1, 2 and 3.**

Mean and Variance (8/15)

Example2 – Answer (2/9)

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

The probability distribution of *X* is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

Mean and Variance (8/15)

Example2 – Answer (3/9)

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

$$f(0) = P(X = 0) = \frac{\binom{4}{0}\binom{3}{3}}{\binom{7}{3}} = \frac{1}{35}$$

Mean and Variance (8/15)

Example2 – Answer (4/9)

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

$$f(1) = P(X = 1) = \frac{\binom{4}{1}\binom{3}{2}}{\binom{7}{3}} = \frac{12}{35}$$

Mean and Variance (8/15)

Example2 – Answer (5/9)

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

$$f(2) = P(X = 2) = \frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}} = \frac{18}{35}$$

Example2 – Answer (6/9)

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

$$f(3) = P(X = 3) = \frac{\binom{4}{3}\binom{3}{0}}{\binom{7}{3}} = \frac{4}{35}$$

Mean and Variance (8/15)

Example2 – Answer (7/9)

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

x	0	1	2	3
$f(x) = P(X = x)$	1/35	12/35	18/35	4/35

Example2 – Answer (8/9)

Find the expected value of the number of good components in this sample.

x	0	1	2	3
$f(x) = P(X = x)$	1/35	12/35	18/35	4/35

$$E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7.$$

Mean and Variance (8/15)

Example2 – Answer (9/9)

$\boxed{E(X) = 1.7}$

Determine the variance of the random variable X

x	0	1	2	3
$f(x) = P(X = x)$	1/35	12/35	18/35	4/35

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

$$E(X^2) = \sum x_i^2 P(x_i) = 0\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (4)\left(\frac{18}{35}\right) + (9)\left(\frac{4}{35}\right) = \frac{120}{35} = 3.43$$

$V(X) = 3.43 - (1.7)^2 = 0.54,$ Standard Deviation $(\sigma) = \sqrt{0.54} = 0.74$

Mean and Variance (9/15)

For any constants *a* and *b*:

Mean

1. $E(a) = a\,, \qquad a \in \mathbb{R}$
2. $E(aX + b) = aE(X) + b\,, \qquad a, b \in \mathbb{R}$

Variance

1. $V(a) = 0\,, \ a \in \mathbb{R}$
2. $V(aX + b) = a^2 V(X)\,, \qquad a, b \in \mathbb{R}$

Mean and Variance (10/15)

Example3:

A discrete random variable with $V\,(X) = 2.5$
Evaluate $V(2X + 1)$

Mean and Variance (11/15)

Example3 – Answer

A discrete random variable with $V(X) = 2.5$
Evaluate $V(2X + 1)$

$$V(aX + b) = a^2 V(X), \qquad a, b \in \mathbb{R}$$

$$V(2X + 1) = 4V(X) = 4 \times 2.5 = 10$$

Mean and Variance (12/15)

Example4:

A discrete random variable with $E\,(X) = 2.5$

Evaluate $E\,(2X + 1)$

Mean and Variance (13/15)

Example4 – Answer

A discrete random variable with $E\,(X) = 2.5$
Evaluate $E\,(2X + 1)$

$$E\,(aX + b) = aE\,(X) + b\,, \qquad\qquad a, b\ \in \mathbb{R}$$

$$E\,(2X + 1) = 2E\,(X) + 1$$

$$E\,(2X + 1) = 2 \times 2.5 + 1 = 6$$

Mean and Variance (14/15)

Example5:

Let X is a random variable with mean 6 and variance 100. Consider another random variable Y such that $Y = 3X + 6$, evaluate the mean and variance of Y?

Mean and Variance (15/15)

Example5 – Answer

Let X is a random variable with mean 6 and variance 100. Consider another random variable Y such that $Y = 3X + 6$, evaluate the mean and variance of Y?

$$E(X) = 6 \quad , \quad V(X) = 100$$

$$E(Y) = E(3X + 6)$$

$$V(Y) = V(3X + 6)$$

Mean and Variance (15/15)

Example5 – Answer

Let X is a random variable with mean 6 and variance 100. Consider another random variable Y such that $Y = 3X + 6$, evaluate the mean and variance of Y?

$$E(X) = 6 \quad , \quad V(X) = 100$$

$$E(Y) = E(3X + 6) = 3E(X) + 6 = 3(6) + 6 = 24$$

$$V(Y) = V(3X + 6) = 9V(X) = 9(100) = 900$$

Continuous R. V. (1/3)

Continuous Random Variable:

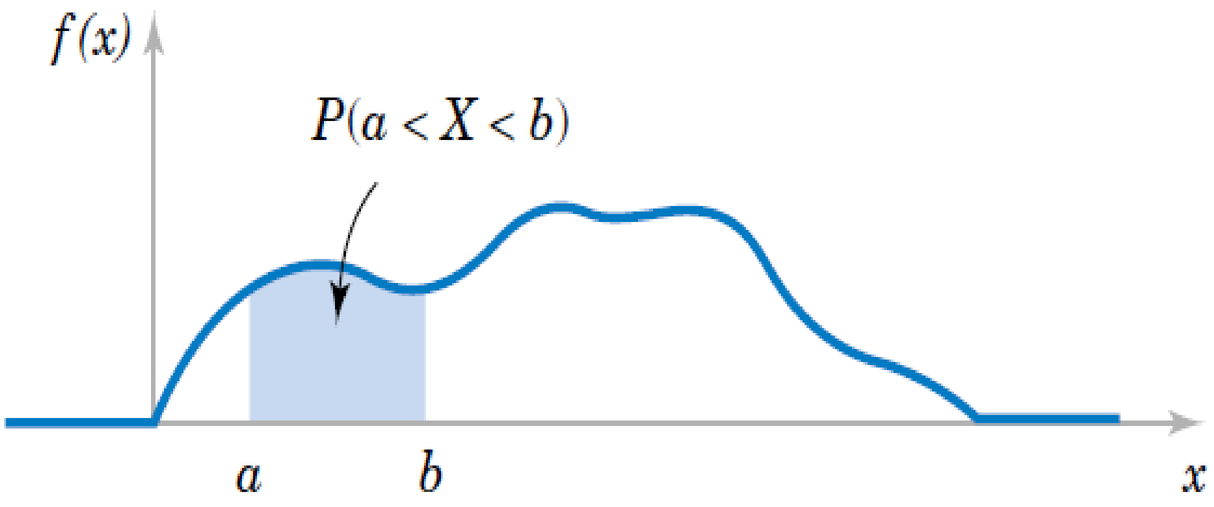
If the range space R_X of the random variable X is an interval or a collection of intervals, X is called a *continuous random variable*.

A continuous random variable has a probability of **0** of assuming *exactly* any of its values. Consequently, its probability distribution cannot be given in tabular form.

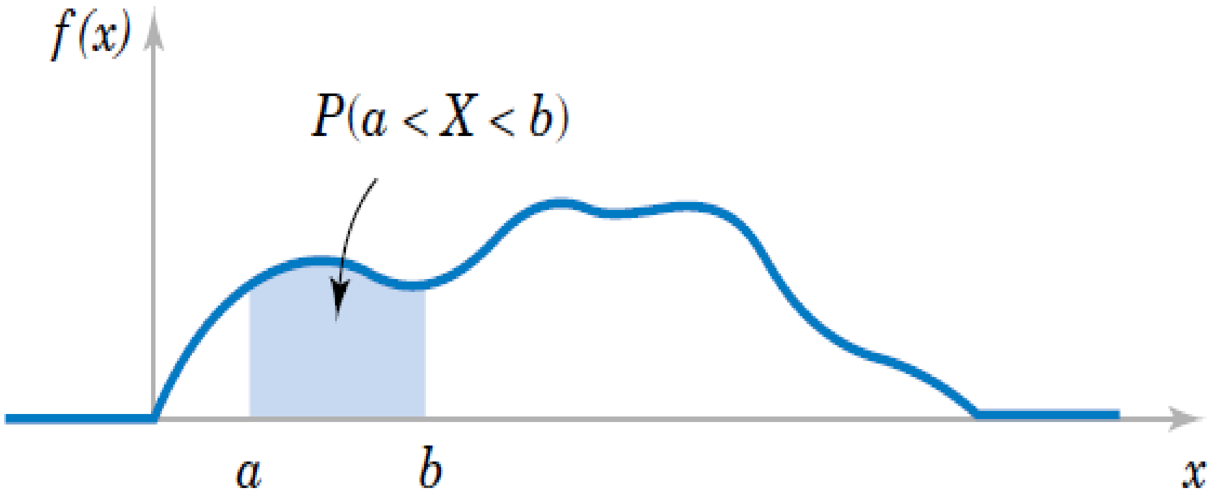
Continuous R. V. (2/3)

Example:

If we talk about the probability of selecting a person who is at least 163 centimeters but not more than 165 centimeters tall. Now we are dealing with an interval rather than a point value of our random variable.



Continuous R. V. (3/3)



If X is a **continuous random variable,** for any x_1 and x_2,

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2)$$

Prob. Density Functions (1/6)

Probability Density Function

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $\displaystyle P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx =$ area under $f(x)$ from a to b

for any a and b

Prob. Density Functions (2/6)

Definite Integral:

$$\boxed{\int_a^b f(x)\,dx = F(b) - F(a)}$$

$$\int_1^3 x^2\,dx$$

$$\int_1^3 x^2\,dx = \left[\frac{x^3}{3}\right]_1^3 = \left[\frac{3^3}{3}\right] - \left[\frac{1^3}{3}\right] = \left[9 - \left(\frac{1}{3}\right)\right] = \frac{26}{3}$$

Prob. Density Functions (3/6)

Example1:

Suppose that $\boldsymbol{f(x) = e^{-x}}$ for $\boldsymbol{x} > 0$
Check the probability density function, then determine the following probabilities:

1. $P(X < 1)$
2. $P(1 \leq X < 2.5)$
3. $P(X = 3)$
4. $P(X \geq 3)$

Example1 – Answer (1/5)

Check the probability density function:

$$\int_{0}^{\infty} e^{-x}\, dx$$

$$\int_{0}^{\infty} e^{-x}\, dx = -e^{-x}\Big|_{0}^{\infty} = \left(-e^{-\infty}\right) - \left(-e^{0}\right) = 0 + 1 = 1$$

Prob. Density Functions (4/6)

Example1 – Answer (2/5)

1) $P(X < 1)$

$$\boldsymbol{P(X < 1) = \int_0^1 e^{-x} dx = -e^{-x} \Big|_0^1 = \left(-e^{-1}\right) - \left(-e^{0}\right)}$$

$$\boldsymbol{= -0.367879 + 1 = 0.632121}$$

Prob. Density Functions (4/6)

Example1 – Answer (3/5)

2) $\boldsymbol{P(1 \le X < 2.5)}$

$$\boldsymbol{P(1 \le X < 2.5) = \int_{1}^{2.5} e^{-x}\, dx = -e^{-x}\Big|_{1}^{2.5}}$$

$$\boldsymbol{= \left(-e^{-2.5}\right) - \left(-e^{-1}\right) = -0.082085 + 0.367879}$$

$$\boldsymbol{= 0.285794}$$

Prob. Density Functions (4/6)

Example1 – Answer (4/5)

3) $\boldsymbol{P(X = 3)}$

$$\boldsymbol{P(X} = 3) = \boldsymbol{0}$$

Prob. Density Functions (4/6)

Example1 – Answer (5/5)

4) $\boldsymbol{P(X \geq 3)}$

$$\boldsymbol{P(X \geq} \ 3) = \int_{\boldsymbol{3}}^{\infty} \boldsymbol{e^{-x}dx} \ = \ \boldsymbol{-e^{-x}}\Big|_{3}^{\infty} \ = \ \left(\boldsymbol{-e^{-\infty}}\right) - \left(\boldsymbol{-e^{-3}}\right)$$

$$= \boldsymbol{0 + 0.049787 = 0.049787}$$

Example2:

Suppose that the error in the reaction temperature, in °C (Celsius), for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

Prob. Density Functions (6/6)

Example2 – Answer (1/2)

Check the probability density function:

$$\int_{-1}^{2} \frac{x^2}{3}\, \boldsymbol{dx}$$

$$\int_{-1}^{2} \frac{x^2}{3}\, \boldsymbol{dx} = \frac{x^3}{9}\Big|_{-1}^{2} = \left(\frac{(2)^3}{9}\right) - \left(\frac{(-1)^3}{9}\right) = \frac{8}{9} + \frac{1}{9} = \mathbf{1}$$

Example2 – Answer (2/2)

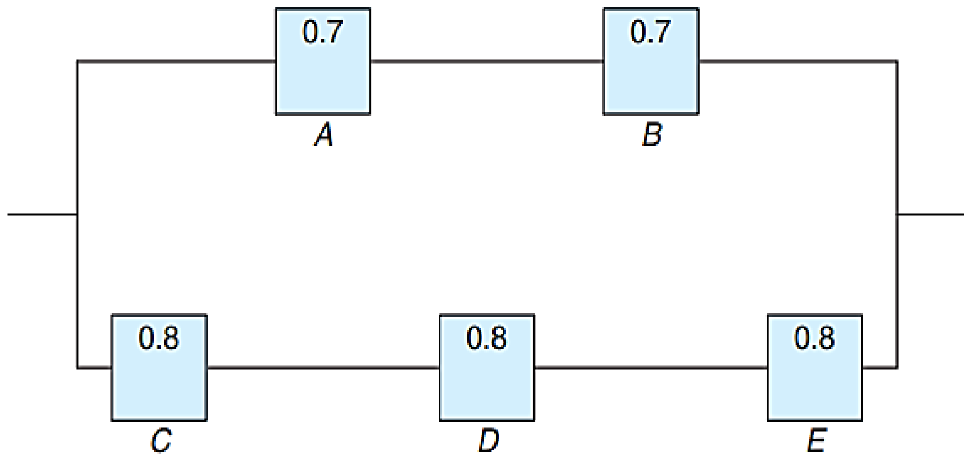
2) $\boldsymbol{P(0 < X \leq 1)}$

$$\boldsymbol{P(0 < X \leq 1)} = \int_0^1 \frac{x^2}{3}\, \boldsymbol{dx} = \frac{x^3}{9}\Big|_0^1$$

$$= \left(\frac{(1)^3}{9}\right) - \left(\frac{(0)^3}{9}\right) = \frac{1}{9} + \frac{0}{9} = \boldsymbol{\frac{1}{9}}$$

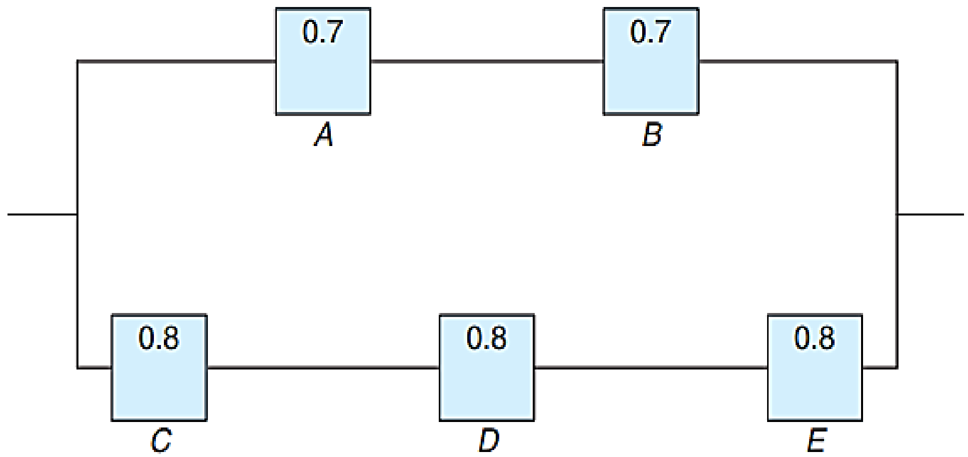
Quiz (1) – Q1

1. A circuit system is given in the following figure. Assume the components fail independently. What is the probability that the entire system works?



Quiz (1) – Q1

1. A circuit system is given in the following figure. Assume the components fail independently. What is the probability that the entire system works?



UP: (0.7) (0.7) = 0.49

Down: (0.8) (0.8) (0.8) = 0.512

$P(W)$ = 1 − [(0.51) (0.488)] = 0.75112

Quiz (1) – Q2

2. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many ways can a student check off one answer to each question and get all the answers wrong?

Quiz (1) – Q2

2. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many ways can a student check off one answer to each question and get all the answers wrong?

Ways to get all the answers wrong $= (3)^6 = 729$

Quiz (1) – Q3

3. If A, and B are two independent events. And $P(A) = 0.2$, $P(A \cap B) = 0.06$, determine the following probabilities:

   a) $P(A|B)$

   b) $P(B|A)$

Quiz (1) – Q3

3. If A, and B are two independent events. And $P(A) = 0.2$, $P(A \cap B) = 0.06$, determine the following probabilities:

   a) $P(A|B)$

$$P(A|B) = P(A) = 0.2$$

   b) $P(B|A)$

$$P(B|A) = P(\mathrm{B}) = \frac{P(A \cap B)}{P(A)} = \frac{0.06}{0.2} = 0.3$$

Quiz (1) – Q4

4. In how many ways can 3 boys and 4 girls sit in a row?

Quiz (1) – Q4

4. In how many ways can 3 boys and 4 girls sit in a row?

$$\text{number of ways} = \frac{n!}{n_1!\, n_2!} = \frac{7!}{3!\, 4!} = 35$$

Quiz (1) – Q5

5. The likely location of a mobile device in the home is as follows:

Adult bedroom: 0.10, Child bedroom: 0.20,

Office: 0.40, Other rooms: 0.30

(a) What is the probability that a mobile device is in a bedroom?

(b) What is the probability that it is not in a bedroom?

Quiz (1) – Q5

5. The likely location of a mobile device in the home is as follows:

Adult bedroom: 0.10, Child bedroom: 0.20,

Office: 0.40, Other rooms: 0.30

(a) What is the probability that a mobile device is in a bedroom?

$= P(Adult\ bedroom \cup Child\ bedroom) = 0.10 + 0.20 = 0.30$

(b) What is the probability that it is not in a bedroom?

$= 1 - 0.30 = 0.70$

Quiz (1) – Q6

6. How many distinct permutations can be made from the letters of the word COMPUTER?

Quiz (1) – Q6

6. How many distinct permutations can be made from the letters of the word COMPUTER?

$n! = 8! = 40{,}320$

Quiz (1) – Q7

7. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many different ways can a student check off one answer to each question?

Quiz (1) – Q7

7. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many different ways can a student check off one answer to each question?

$$= (4)^6 = 4{,}096$$

Chapter 2: Random Variable

- Discrete Random Variables.
- Probability Mass Functions.
- Cumulative Distribution Functions.
- Discrete R.V. (Mean and Variance).
- Continuous Random Variables.
- Probability Density Functions.
- Continuous R.V. (Mean and Variance).
- Joint Probability Distributions.

Prob. Density Functions (1/2)

Rev. Example:

Suppose that the error in the reaction temperature, in °C (Celsius), for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

Prob. Density Functions (2/2)

Rev. Example – Answer (1/2)

Check the probability density function:

$$\int_{-1}^{2} \frac{x^2}{3}\, \boldsymbol{dx}$$

$$\int_{-1}^{2} \frac{x^2}{3}\, \boldsymbol{dx} = \frac{x^3}{9}\Big|_{-1}^{2} = \left(\frac{(2)^3}{9}\right) - \left(\frac{(-1)^3}{9}\right) = \frac{8}{9} + \frac{1}{9} = \mathbf{1}$$

Prob. Density Functions (2/2)

Rev. Example – Answer (2/2)

2) $\boldsymbol{P(0 < X \leq 1)}$

$$\boldsymbol{P(0 < X \leq 1)} = \int_0^1 \frac{x^2}{3}\,\boldsymbol{dx} = \frac{x^3}{9}\Big|_0^1$$

$$= \left(\frac{(1)^3}{9}\right) - \left(\frac{(0)^3}{9}\right) = \frac{1}{9} + \frac{0}{9} = \boldsymbol{\frac{1}{9}}$$

Cumulative Distribution Fun. (1/4)

Cumulative Distribution Function

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du$$

for $-\infty < x < \infty$.

Probability Density Function from the Cumulative Distribution Function

Given $F(x)$,

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

Cumulative Distribution Fun. (2/4)

Cumulative Distribution Function

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du$$

for $-\infty < x < \infty$.

$$\boxed{P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx}, \text{ if the derivative exists.}}$$

Cumulative Distribution Fun. (3/4)

Example1:

Find the cumulative distribution function $F(x)$ and use it to evaluate $P(0 < X \leq 1)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Example1 – Answer (1/3)

Find the cumulative distribution function $F(x)$

$$F(x) = \int_{-\infty}^{x} \frac{u^2}{3}\, \boldsymbol{du}$$

$$\boxed{\frac{x^2}{3}, \quad -1 < x < 2,}$$

Example1 – Answer (1/3)

Find the cumulative distribution function $F(x)$

$$F(x) = \int_{-\infty}^{x} \frac{u^2}{3}\, du = \int_{-1}^{x} \frac{u^2}{3}\, du$$

$$\boxed{\frac{x^2}{3}, \quad -1 < x < 2,}$$

Example1 – Answer (2/3)

Find the cumulative distribution function $F(x)$

$$F(x) = \int_{-\infty}^{x} \frac{u^2}{3}\, \boldsymbol{du} = \int_{-1}^{x} \frac{u^2}{3}\, \boldsymbol{du} = \frac{u^3}{9}\Big|_{-1}^{x} = \left(\frac{(x)^3}{9}\right) - \left(\frac{(-1)^3}{9}\right)$$

$$= \frac{x^3}{9} + \frac{1}{9} = \frac{x^3 + 1}{9}$$

Cumulative Distribution Fun. (4/4)

Example1 – Answer (2/3)

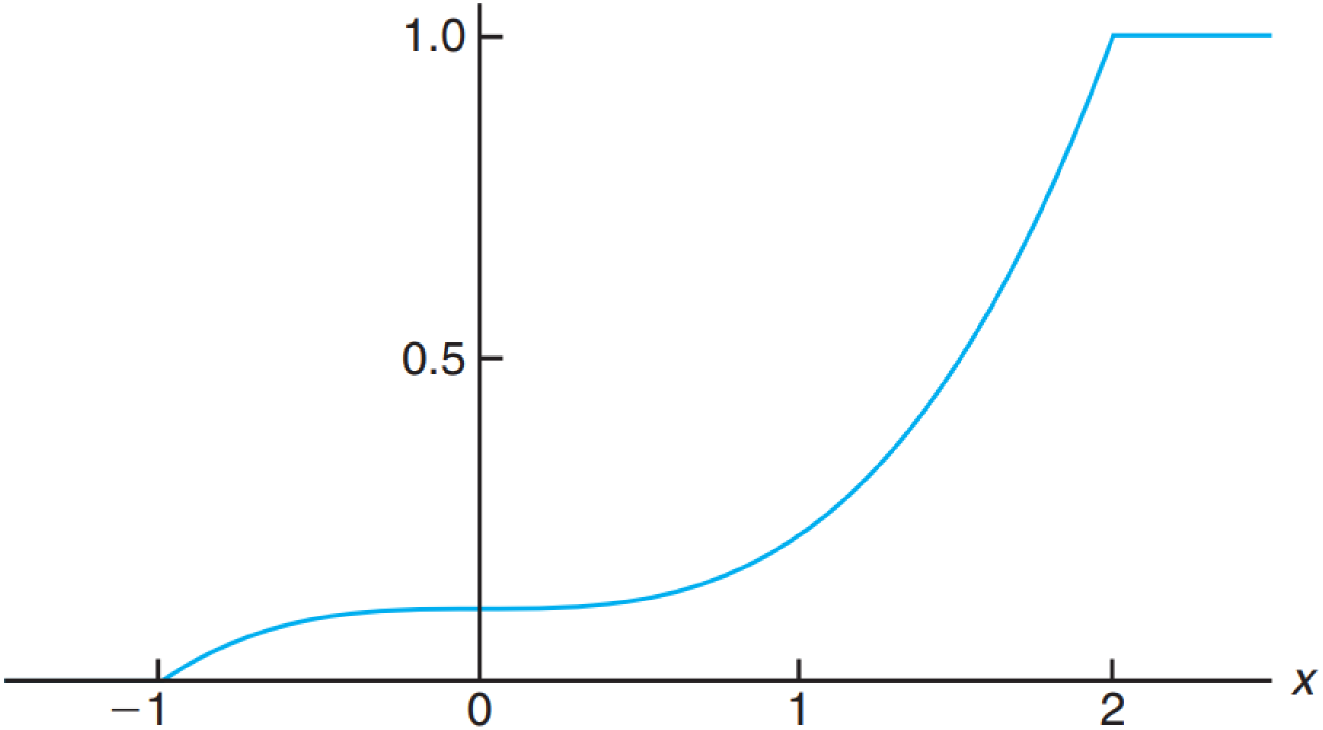
Find the cumulative distribution function $F(x)$

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$

Cumulative Distribution Fun. (4/4)

Example1 – Answer (2/3)

Find the cumulative distribution function $F(x)$



Cumulative Distribution Fun. (4/4)

Example1 – Answer (3/3)

Find the cumulative distribution function $F(x)$ and use it to evaluate $P(0 < X \leq 1)$.

$$\boxed{F(x) = \frac{x^3 + 1}{9}}$$

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

Mean and Variance (1/3)

Mean and Variance

Suppose that X is a continuous random variable with probability density function $f(x)$. **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Mean and Variance (2/3)

Example1:

Find the expected value of X, $E(X)$ and the variance $V(X)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Mean and Variance (3/3)

Example1 – Answer (1/3)

Find the expected value of X, $E(X)$ and the variance $V(X)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Mean and Variance (3/3)

Example1 – Answer (1/3)

Find the expected value of X, $E(X)$ and the variance $V(X)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-1}^{2} \frac{x^3}{3}\,dx = \frac{x^4}{12}\Big|_{-1}^{2}$$

Mean and Variance (3/3)

Example1 – Answer (1/3)

Find the expected value of X, $E(X)$ and the variance $V(X)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

$$E(X) = \int_{-1}^{2} \frac{x^3}{3} dx = \frac{x^4}{12} \Big|_{-1}^{2} = \frac{(2)^4}{12} - \frac{(-1)^4}{12} = \frac{15}{12}$$

Mean and Variance (3/3)

Example1 – Answer (2/3)

Find the expected value of X, $E(X)$ and the variance $V(X)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

$$E(X^2) = \int_{-1}^{2} \frac{x^4}{3}\, dx = \frac{x^5}{15}\Big|_{-1}^{2} = \frac{(2)^5}{15} - \frac{(-1)^5}{15} = \frac{33}{15}$$

Mean and Variance (3/3)

Example1 – Answer (3/3)

Find the expected value of X, $E(X)$ and the variance $V(X)$.

$$E(X) = \frac{15}{12}$$

$$E(X^2) = \frac{33}{15}$$

$$V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{33}{15} - \left(\frac{15}{12}\right)^2 = 0.6375$$

Mean and Variance (3/3)

Example1 – Answer (3/3)

Find the expected value of X, $E(X)$ and the variance $V(X)$.

$$E(X) = \frac{15}{12}$$

$$E(X^2) = \frac{33}{15}$$

Standard Deviation (σ)
$= \sqrt{0.6375} = 0.798$

$$V(X) = E(X^2) - \big(E(X)\big)^2 = \frac{33}{15} - \left(\frac{15}{12}\right)^2 = 0.6375$$

Joint Prob. Distributions (1/9)

Definition:

If X and Y are two random variables, the probability distribution that defines their simultaneous behavior is called a joint probability distribution.

Joint Prob. Distributions (2/9)

If *X* and *Y* are discrete random variables:

Joint Probability Mass Function

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

(1) $f_{XY}(x, y) \geq 0$

(2) $\sum_X \sum_Y f_{XY}(x, y) = 1$

(3) $f_{XY}(x, y) = P(X = x, Y = y)$

Joint Prob. Distributions (3/9)

Example1:

A coin is tossed two times. Calculate the joint probability mass function of the discrete random variables X and Y, where X denote the number of heads appear and Y denote the number of tails appear.

Joint Prob. Distributions (4/9)

Example1 – Answer (1/2)

$S = \{$	$HH,$	$HT,$	$TH,$	$TT\}$
X	2	1	1	0
Y	0	1	1	2

Joint Prob. Distributions (4/9)

Example1 – Answer (2/2)

$$S = \{HH, HT, TH, TT\}$$

X	2	1	1	0
Y	0	1	1	2

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Example2:

A coin is tossed two times. Calculate the joint probability mass function of the discrete random variables X and Y, where X denote the number of heads appear and Y denote the number of tails appear.

Find:

1) $f_{XY}(1,2) = P(X = 1,\ Y = 2)$

2) $f_{XY}(2,0) = P(X = 2,\ Y = 0)$

3) $P(X = 1,\ Y \le 2)$

4) $P(Y = 2)$

Joint Prob. Distributions (6/9)

Example2 – Answer (1/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

1) $f_{XY}(1,2) = P(X = 1,\ Y = 2)$

Joint Prob. Distributions (6/9)

Example2 – Answer (1/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

1) $f_{XY}(1,2) = P(X = 1,\ Y = 2) = 0$

Joint Prob. Distributions (6/9)

Example2 – Answer (2/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

2) $f_{XY}(2,0) = P(X = 2,\ Y = 0)$

Joint Prob. Distributions (6/9)

Example2 – Answer (2/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

2) $f_{XY}(2,0) = P(X = 2,\ Y = 0) = 1/4$

Joint Prob. Distributions (6/9)

Example2 – Answer (3/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

3) $P(X = 1,\ Y \leq 2)$

Joint Prob. Distributions (6/9)

Example2 – Answer (3/4)

$y \backslash x$	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

3) $P(X = 1,\ Y \leq 2) = 0 + \frac{2}{4} + 0 = \frac{2}{4}$

Joint Prob. Distributions (6/9)

Example2 – Answer (4/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

4) $P(Y = 2)$

Joint Prob. Distributions (6/9)

Example2 – Answer (4/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

4) $P(Y = 2) = \frac{1}{4} + 0 + 0 = \frac{1}{4}$

Marginal Probability Distributions

The marginal distributions of the random variable X alone is:

$$f_X(x) = \sum_{y} f_{XY}(x, y)$$

The marginal distributions of the random variable Y alone is:

$$f_Y(y) = \sum_{x} f_{XY}(x, y)$$

Example3:

A coin is tossed two times. Calculate the joint probability mass function of the discrete random variables X and Y, where X denote the number of heads appear and Y denote the number of tails appear.

Find:

1) $f_X(x)$

2) $f_Y(y)$

Joint Prob. Distributions (9/9)

Example3 – Answer (1/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

1) $f_X(x)$

Joint Prob. Distributions (9/9)

Example3 – Answer (1/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0
$f_X(x)$	**1/4**	**2/4**	**1/4**

Find:

1) $f_X(x)$

Example3 – Answer (2/4)

x	0	1	2
$f_X(x)$	1/4	2/4	1/4

Find:

1) $f_X(x)$

Joint Prob. Distributions (9/9)

Example3 – Answer (3/4)

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

2) $f_Y(y)$

Joint Prob. Distributions (9/9)

Example3 – Answer (3/4)

y \ x	**0**	**1**	**2**	$f_Y(y)$
0	0	0	1/4	**1/4**
1	0	2/4	0	**2/4**
2	1/4	0	0	**1/4**

Find:

2) $f_Y(y)$

Example3 – Answer (4/4)

y	0	1	2
$f_Y(y)$	1/4	2/4	1/4

Find:

2) $f_Y(y)$

Chapter 2: Random Variable

- Discrete Random Variables.
- Probability Mass Functions.
- Cumulative Distribution Functions.
- Discrete R.V. (Mean and Variance).
- Continuous Random Variables.
- Probability Density Functions.
- Continuous R.V. (Mean and Variance).
- Joint Probability Distributions.

Joint Prob. Mass Fun. (1/15)

Joint Probability Mass Function:

If X and Y are two discrete random variables, the joint probability mass function is denoted as $f_{XY}(x, y)$, satisfies

1) $f_{XY}(x, y) \geq 0$

2) $\sum_{X} \sum_{Y} f_{XY}(x, y) = 1$

3) $f_{XY}(x, y) = P(X = x, Y = y)$

Joint Prob. Mass Fun. (2/15)

Review Example

$$S = \{HH, HT, TH, TT\}$$

X	2	1	1	0
Y	0	1	1	2

number of heads

number of tails

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

joint probability mass function $f_{XY}(x, y)$

Joint Prob. Mass Fun. (3/15)

Marginal Probability Distributions (Discrete):

The marginal distributions of the random variable X alone is:

$$f_X(x) = \sum_{y} f_{XY}(x, y)$$

The marginal distributions of the random variable Y alone is:

$$f_Y(y) = \sum_{x} f_{XY}(x, y)$$

Joint Prob. Mass Fun. (4/15)

Review Example (1/4):

$f_X(x)$

y \ x	**0**	**1**	**2**
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0
$f_X(x)$	**1/4**	**2/4**	**1/4**

Joint Prob. Mass Fun. (4/15)

Review Example (2/4):

$f_X(x)$

x	0	1	2
$f_X(x)$	1/4	2/4	1/4

Joint Prob. Mass Fun. (4/15)

Review Example (3/4):

$f_Y(y)$

y \ x	**0**	**1**	**2**	$f_Y(y)$
0	0	0	1/4	**1/4**
1	0	2/4	0	**2/4**
2	1/4	0	0	**1/4**

Joint Prob. Mass Fun. (4/15)

Review Example (4/4):

$f_Y(y)$

y	0	1	2
$f_Y(y)$	1/4	2/4	1/4

Joint Prob. Mass Fun. (5/15)

Mean from a Joint Distribution (Discrete):

Mean for the random variable X alone is:

$$E(X) = \sum_{x} x f_X(x)$$

Mean for the random variable Y alone is:

$$E(Y) = \sum_{y} y f_Y(y)$$

Joint Prob. Mass Fun. (6/15)

Variance from a Joint Distribution (Discrete):

Variance for the random variable X alone is:

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

and s.d. of $X = \sigma_X = \sqrt{V(X)}$

Variance for the random variable Y alone is:

$$V(Y) = E(Y^2) - \left(E(Y)\right)^2$$

and s.d. of $Y = \sigma_Y = \sqrt{V(Y)}$

Joint Prob. Mass Fun. (7/15)

Example1 (1/4): (from the previous example)

$f_X(x)$

x	**0**	**1**	**2**
$f_X(x)$	**1/4**	**2/4**	**1/4**

Find:

$E(X)$

$V(X)$

σ_X

Joint Prob. Mass Fun. (7/15)

Example1 (2/4): (from the previous example)

$f_X(x)$

x	0	1	2
$f_X(x)$	1/4	2/4	1/4

1:

$$E(X) = \sum_{x} x f(x) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{2}{4}\right) + (2)\left(\frac{1}{4}\right) = 1$$

Joint Prob. Mass Fun. (7/15)

Example1 (3/4): (from the previous example)

$f_X(x)$

x	**0**	**1**	**2**
$f_X(x)$	**1/4**	**2/4**	**1/4**

2:

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

$$E(X^2) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{2}{4}\right) + (4)\left(\frac{1}{4}\right) = \frac{6}{4} = 1.5$$

Joint Prob. Mass Fun. (7/15)

Example1 (3/4): (from the previous example)

$f_X(x)$

x	0	1	2
$f_X(x)$	1/4	2/4	1/4

2:

$$E(X^2) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{2}{4}\right) + (4)\left(\frac{1}{4}\right) = \frac{6}{4} = 1.5$$

$$V(X) = 1.5 - (1)^2 = 0.5$$

Joint Prob. Mass Fun. (7/15)

Example1 (4/4): (from the previous example)

$f_X(x)$

x	**0**	**1**	**2**
$\boldsymbol{f_X(x)}$	**1/4**	**2/4**	**1/4**

3:

$$V(X) = 1.5 - (1)^2 = 0.5$$

$$\sigma_X = \sqrt{0.5} = 0.7071$$

Joint Prob. Mass Fun. (8/15)

Conditional Probability Distributions (Discrete):

The random variables X and Y are discrete, the conditional probability mass function of X given $Y = y$ is:

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

for $f_Y(y) > 0$

Joint Prob. Mass Fun. (9/15)

Conditional Probability Distributions (Discrete):

The random variables X and Y are discrete, the conditional probability mass function of X given $Y = y$ is:

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

1

$$\boldsymbol{f_{X|2}(x) = f_{X|Y=2}(x) = f_{XY}(x, 2)\mathrm{T}f_Y(2)}$$

The conditional probability distribution of X given that Y = 2.

Joint Prob. Mass Fun. (10/15)

Conditional Probability Distributions (Discrete):

The random variables X and Y are discrete, the conditional probability mass function of X given $Y = y$ is:

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

2

$$\boldsymbol{f_{Y|3}(y) = f_{Y|X=3}(y) = f_{XY}(3, y)} \mathrm{T} \boldsymbol{f_X(3)}$$

The conditional probability distribution of Y given that X = 3.

Joint Prob. Mass Fun. (11/15)

Example2:

$f_{XY}(x, y)$

Find:

$f_X(x)$

$E(Y)$

σ_X

$f_{X|3}(x)$

$f_{Y|1.5}(y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (1/17):

$f_{XY}(x, y)$

<u>**1:**</u>

$f_X(x) \rightarrow$ marginal X

y \ x	$\mathbf{1}$	$\mathbf{1.5}$	$\mathbf{2.5}$	$\mathbf{3}$
$\mathbf{1}$	$1/4$	0	0	0
$\mathbf{2}$	0	$1/8$	0	0
$\mathbf{3}$	0	$1/4$	0	0
$\mathbf{4}$	0	0	$1/4$	0
$\mathbf{5}$	0	0	0	$1/8$
$\boldsymbol{f_X(x)}$	$\mathbf{1/4}$	$\mathbf{3/8}$	$\mathbf{1/4}$	$\mathbf{1/8}$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (1/17):

x	1	1.5	2.5	3
$f_X(x)$	$1/4$	$3/8$	$1/4$	$1/8$

<u>1:</u>

$f_X(x) \rightarrow$ marginal X

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (2/17):

$f_{XY}(x, y)$

y \ x	$\mathbf{1}$	$\mathbf{1.5}$	$\mathbf{2.5}$	$\mathbf{3}$
$\mathbf{1}$	$1/4$	0	0	0
$\mathbf{2}$	0	$1/8$	0	0
$\mathbf{3}$	0	$1/4$	0	0
$\mathbf{4}$	0	0	$1/4$	0
$\mathbf{5}$	0	0	0	$1/8$

<u>**2:**</u>

$E(Y)$ $\rightarrow$ Mean Y

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (2/17):

$f_{XY}(x, y)$

$y \backslash x$	$\mathbf{1}$	$\mathbf{1.5}$	$\mathbf{2.5}$	$\mathbf{3}$
1	$1/4$	0	0	0
2	0	$1/8$	0	0
3	0	$1/4$	0	0
4	0	0	$1/4$	0
5	0	0	0	$1/8$

<u>2:</u>

$$E(Y) = \sum_{y} y f_Y(y)$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (3/17):

$f_{XY}(x, y)$

2:

$$E(Y) = \sum_{y} y f_Y(y)$$

y \ x	**1**	**1.5**	**2.5**	**3**	$f_Y(y)$
1	1/4	0	0	0	**1/4**
2	0	1/8	0	0	**1/8**
3	0	1/4	0	0	**1/4**
4	0	0	1/4	0	**1/4**
5	0	0	0	1/8	**1/8**

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (3/17):

2:

$$E(Y) = \sum_{y} y f_Y(y)$$

y	$f_Y(y)$
1	1/4
2	1/8
3	1/4
4	1/4
5	1/8

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (4/17):

y	$f_Y(y)$
1	1/4
2	1/8
3	1/4
4	1/4
5	1/8

2:

$$E(Y) = \sum_{y} y f_Y(y)$$

$$E(Y) = (1)\left(\frac{1}{4}\right) + (2)\left(\frac{1}{8}\right) + (3)\left(\frac{1}{4}\right) + (4)\left(\frac{1}{4}\right) + (5)\left(\frac{1}{8}\right) = \frac{23}{8}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (4/17):

y	$f_Y(y)$
1	1/4
2	1/8
3	1/4
4	1/4
5	1/8

2:

$$E(Y) = \sum_{y} y f_Y(y)$$

$$E(Y) = \frac{23}{8} = 2.875$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (5/17):

$f_{XY}(x, y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

<u>3:</u>

$\sigma_X \rightarrow$ s.d. of X

$= \sqrt{V(X)}$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (5/17):

$f_{XY}(x, y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

<u>**3:**</u>

$\sigma_X \rightarrow$ s.d. of X

$= \sqrt{V(X)}$

$= E(X^2) - \left(E(X)\right)^2$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (6/17):

Recall

1:

x	1	1.5	2.5	3
$f_X(x)$	1/4	3/8	1/4	1/8

$f_X(x) \rightarrow$ marginal X

3:

$\sigma_X \rightarrow$ s.d. of X

$$= \sqrt{V(X)}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (7/17):

Recall

x	1	1.5	2.5	3
$f_X(x)$	1/4	3/8	1/4	1/8

1:

$f_X(x) \rightarrow$ marginal X

3:

$\sigma_X \rightarrow$ s.d. of X

$= \sqrt{V(X)}$

$$E(X) = \sum_x x f_X(x)$$

$$= (1)\left(\frac{1}{4}\right) + (1.5)\left(\frac{3}{8}\right) + (2.5)\left(\frac{1}{4}\right) + (3)\left(\frac{1}{8}\right) = 1.8125$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (8/17):

Recall

<u>3:</u>

$\sigma_X \rightarrow$ s.d. of X

$= \sqrt{V(X)}$

<u>1:</u>

$f_X(x) \rightarrow$ marginal X

x	1	1.5	2.5	3
$f_X(x)$	1/4	3/8	1/4	1/8

$$E(X^2) = \sum_x x^2 f_X(x)$$

$$= (1)\left(\frac{1}{4}\right) + (2.25)\left(\frac{3}{8}\right) + (6.25)\left(\frac{1}{4}\right) + (9)\left(\frac{1}{8}\right)$$

$$= 3.78125$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (9/17):

$$E(X) = 1.8125$$

$$E(X^2) = 3.78125$$

3:

$\sigma_X \rightarrow$ s.d. of X

$$V(X) = 3.78125 - (1.8125)^2 = 0.49609375$$

$$\sigma_X = \sqrt{0.49609375} = 0.704339$$

Example2 – Answer (10/17):

4:

$f_{X|3}(x)$

$The\ conditional\ probability\ distribution\ of\ X\ given\ that\ Y = 3.$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (11/17):

$f_{XY}(x, y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	$1/4$	0	0	0
2	0	$1/8$	0	0
3	0	$1/4$	0	0
4	0	0	$1/4$	0
5	0	0	0	$1/8$

<u>**4:**</u>

$f_{X|3}(x)$

$$\boxed{f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (11/17):

$f_{XY}(x, y)$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**
1	$1/4$	0	0	0
2	0	$1/8$	0	0
3	0	$1/4$	0	0
4	0	0	$1/4$	0
5	0	0	0	$1/8$

<u>4:</u>

$f_{X|3}(x)$

$y = 3$

$$\boxed{f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (11/17):

$f_{XY}(x, y)$

4:

$f_{X|3}(x)$

$$f_{X|3}(x) = \frac{f_{XY}(x, 3)}{f_Y(3)}$$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (12/17):

$f_{XY}(x, y)$

4:

$f_{X|3}(x)$

$$\boxed{f_{X|3}(x) = \frac{f_{XY}(x,3)}{f_Y(3)}}$$

y \ x	**1**	**1.5**	**2.5**	**3**	$f_Y(y)$
1	1/4	0	0	0	**1/4**
2	0	1/8	0	0	**1/8**
3	0	1/4	0	0	**1/4**
4	0	0	1/4	0	**1/4**
5	0	0	0	1/8	**1/8**

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (12/17):

$f_{XY}(x, y)$

<u>4:</u>

$f_{X|3}(x)$

$y = 3$

$y \backslash x$	$\mathbf{1}$	$\mathbf{1.5}$	$\mathbf{2.5}$	$\mathbf{3}$	$f_Y(y)$
1	$1/4$	0	0	0	$\mathbf{1/4}$
2	0	$1/8$	0	0	$\mathbf{1/8}$
3	0	$1/4$	0	0	$\mathbf{1/4}$
4	0	0	$1/4$	0	$\mathbf{1/4}$
5	0	0	0	$1/8$	$\mathbf{1/8}$

$$\boxed{f_{X|3}(x) = \frac{f_{XY}(x, 3)}{f_Y(3)}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (13/17):

$f_{XY}(x,y)$

4:

$f_{X|3}(x)$

$y = 3$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**	$f_Y(y)$
1	1/4	0	0	0	**1/4**
2	0	1/8	0	0	**1/8**
3	0	1/4	0	0	**1/4**
4	0	0	1/4	0	**1/4**
5	0	0	0	1/8	**1/8**
$f_{X|3}(x)$	**0**				

$$\boxed{f_{X|3}(1) = \frac{f_{XY}(1,3)}{f_Y(3)} = \frac{0}{1/4}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (13/17):

$f_{XY}(x,y)$

4:

$f_{X|3}(x)$

$y = 3$

$y \backslash x$	$\mathbf{1}$	$\mathbf{1.5}$	$\mathbf{2.5}$	$\mathbf{3}$	$\boldsymbol{f_Y(y)}$
1	$1/4$	0	0	0	$\mathbf{1/4}$
2	0	$1/8$	0	0	$\mathbf{1/8}$
3	0	$1/4$	0	0	$\mathbf{1/4}$
4	0	0	$1/4$	0	$\mathbf{1/4}$
5	0	0	0	$1/8$	$\mathbf{1/8}$
$\boldsymbol{f_{X\|3}(x)}$	**0**	**1**			

$$f_{X|3}(1.5) = \frac{f_{XY}(1.5,3)}{f_Y(3)} = \frac{1/4}{1/4}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (13/17):

$f_{XY}(x, y)$

4:

$f_{X|3}(x)$

$y = 3$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**	$f_Y(y)$
1	1/4	0	0	0	**1/4**
2	0	1/8	0	0	**1/8**
3	0	1/4	0	0	**1/4**
4	0	0	1/4	0	**1/4**
5	0	0	0	1/8	**1/8**
$f_{X\|3}(x)$	**0**	**1**	**0**		

$$f_{X|3}(2.5) = \frac{f_{XY}(2.5,3)}{f_Y(3)} = \frac{0}{1/4}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (13/17):

$f_{XY}(x,y)$

<u>**4:**</u>

$f_{X|3}(x)$

$y = 3$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**	$f_Y(y)$
1	1/4	0	0	0	**1/4**
2	0	1/8	0	0	**1/8**
3	0	1/4	0	0	**1/4**
4	0	0	1/4	0	**1/4**
5	0	0	0	1/8	**1/8**
$f_{X\|3}(x)$	**0**	**1**	**0**	**0**	

$$\boxed{f_{X|3}(3) = \frac{f_{XY}(3,3)}{f_Y(3)} = \frac{0}{1/4}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (13/17):

$f_{XY}(x, y)$

4:

$f_{X|3}(x)$

x	1	1.5	2.5	3
$f_{X\|3}(x)$	0	1	0	0

$$\boxed{f_{X|3}(x) = \frac{f_{XY}(x, 3)}{f_Y(3)}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (14/17):

5:

$f_{Y|1.5}(y)$

$The\ conditional\ probability\ distribution\ of\ Y\ given\ that\ X = 1.5.$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (15/17):

$f_{XY}(x,y)$

<u>5:</u>

$f_{Y|1.5}(y)$

$$f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}$$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (15/17):

$f_{XY}(x, y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

<u>5:</u>

$f_{Y|1.5}(y)$

$x = 1.5$

$$\boxed{f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (15/17):

$f_{XY}(x, y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

5:

$f_{Y|1.5}(y)$

$$\boxed{f_{Y|1.5}(y) = \frac{f_{XY}(1.5, y)}{f_X(1.5)}}$$

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (16/17):

$x = 1.5$

$f_{XY}(x, y)$

5:

$f_{Y|1.5}(y)$

$$f_{Y|1.5}(y) = \frac{f_{XY}(1.5, y)}{f_X(1.5)}$$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8
$f_X(x)$	**1/4**	**3/8**	**1/4**	**1/8**

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (17/17):

$f_{XY}(x, y)$

<u>5:</u>

$f_{Y|1.5}(y)$

$$f_{Y|1.5}(y) = \frac{f_{XY}(1.5, y)}{f_X(1.5)}$$

$x = 1.5$

y \ x	**1**	**1.5**	**2.5**	**3**	$f_{Y\|1.5}(y)$
1	1/4	0	0	0	**0**
2	0	1/8	0	0	**1/3**
3	0	1/4	0	0	**2/3**
4	0	0	1/4	0	**0**
5	0	0	0	1/8	**0**
$f_X(x)$	**1/4**	**3/8**	**1/4**	**1/8**	

Joint Prob. Mass Fun. (12/15)

Example2 – Answer (17/17):

5:

$f_{Y|1.5}(y)$

$$f_{Y|1.5}(y) = \frac{f_{XY}(1.5, y)}{f_X(1.5)}$$

y	$f_{Y\|1.5}(y)$
1	0
2	1/3
3	2/3
4	0
5	0

Conditional Mean and Variance (Discrete):

The conditional mean of X given $Y = y$:

$$E(X \mid y) = \sum_{x} x f_{X|y}(x)$$

The conditional variance of X given $Y = y$:

$$V(X \mid y) = \sum_{x} x^2 f_{X|y}(x) - \left(\sum_{x} x f_{X|y}(x) \right)^2$$

Joint Prob. Mass Fun. (14/15)

Example3:

$f_X(x)$

$y \backslash x$	$\mathbf{1}$	$\mathbf{1.5}$	$\mathbf{2.5}$	$\mathbf{3}$
$\mathbf{1}$	$1/4$	0	0	0
$\mathbf{2}$	0	$1/8$	0	0
$\mathbf{3}$	0	$1/4$	0	0
$\mathbf{4}$	0	0	$1/4$	0
$\mathbf{5}$	0	0	0	$1/8$

Find:

$E(X \mid 3)$

$V(X \mid 3)$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (1/7):

$f_{XY}(x,y)$

<u>1:</u>

$E(X \mid 3)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

$$\boxed{E(X \mid y) = \sum_{x} x f_{X|y}(x)}$$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (1/7):

$f_{XY}(x,y)$

<u>1:</u>

$E(X \mid 3)$

$y = 3$

$$E(X \mid 3) = \sum_{x} x f_{X|3}(\ )$$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (2/7):

Recall

4:

$f_{X|3}(x)$

x	1	1.5	2.5	3
$f_{X\|3}(x)$	0	1	0	0

1:

$E(X \mid 3)$

$y = 3$

$$E(X \mid 3) = \sum_{x} x f_{X|3}(\ )$$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (3/7):

Recall

1:

$E(X \mid 3)$

4:

$f_{X|3}(x)$

x	1	1.5	2.5	3
$f_{X\mid 3}(x)$	0	1	0	0

$$E(X \mid 3) = \sum_{x} x f_{X|3}(x)$$

$$= (1)(0) + (1.5)(1) + (2.5)(0) + (3)(0) = 1.5$$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (4/7):

$f_{XY}(x, y)$

2:

$V(X \mid 3)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

$$V(X \mid y) = \sum_{x} x^2 f_{X|y}(x) - \left(\sum_{x} x f_{X|y}(x) \right)^2$$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (4/7):

$f_{XY}(x, y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

<u>**2:**</u>

$V(X \mid 3)$

$y = 3$

$$V(X \mid 3) = \sum_{x} x^2 f_{X|3}(x) - \left(\sum_{x} x f_{X|3}(\ ) \right)^2$$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (5/7):

Recall

$$E(X \mid 3) = \sum_{x} x f_{X|3}(x) = 1.5$$

2:

$V(X \mid 3)$

$y = 3$

$$\boxed{V(X \mid 3) = \sum_{x} x^2 f_{X|3}(x) - (1.5)^2}$$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (6/7):

Recall

<u>2:</u>

$V(X \mid 3)$

<u>4:</u>

$f_{X|3}(x)$

x	1	1.5	2.5	3
$f_{X\|3}(x)$	0	1	0	0

$$\sum_x x^2 f_{X|3}(x) = (1)(0) + (2.25)(1) + (6.25)(0) + (9)(0) = 2.25$$

$$V(X \mid 3) = \sum_x x^2 f_{X|3}(x) - (1.5)^2$$

Joint Prob. Mass Fun. (15/15)

Example3 – Answer (7/7):

2:

$V(X \mid 3)$

$$\sum_{x} x^2 f_{X|3}(x) = (1)(0) + (2.25)(1) + (6.25)(0) + (9)(0) = 2.25$$

$$\boxed{V(X \mid 3) = 2.25 - (1.5)^2 = 0}$$

Chapter 2: Random Variable

- Discrete Random Variables.
- Probability Mass Functions.
- Cumulative Distribution Functions.
- Discrete R.V. (Mean and Variance).
- Continuous Random Variables.
- Probability Density Functions.
- Continuous R.V. (Mean and Variance).
- Joint Probability Distributions.

Independence (1/3)

Independence:

If X and Y are two random variables, the X and Y are independent if any one of the following properties is true:

1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

2) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

3) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

Independence (2/3)

Example1:

$f_{XY}(x, y)$

$y \setminus x$	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Find:

$f_X(x)$

$f_Y(y)$

Are X and Y independent?

Independence (3/3)

Example1 – Answer (1/3):

$f_{XY}(x, y)$

Find:

$f_X(x)$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8
$f_X(x)$	**1/4**	**3/8**	**1/4**	**1/8**

Independence (3/3)

Example1 – Answer (1/3):

$f_{XY}(x,y)$

y \ x	1	1.5	2.5	3
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8
$f_X(x)$	**1/4**	**3/8**	**1/4**	**1/8**

Same as

Find:

$f_X(x)$

x	1	1.5	2.5	3
$f_X(x)$	**1/4**	**3/8**	**1/4**	**1/8**

Independence (3/3)

Example1 – Answer (2/3):

$f_{XY}(x, y)$

Find:

$f_Y(y)$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**	$f_Y(y)$
1	1/4	0	0	0	**1/4**
2	0	1/8	0	0	**1/8**
3	0	1/4	0	0	**1/4**
4	0	0	1/4	0	**1/4**
5	0	0	0	1/8	**1/8**

Independence (3/3)

Example1 – Answer (2/3):

$f_{XY}(x, y)$

y \ x	1	1.5	2.5	3	$f_Y(y)$
1	1/4	0	0	0	1/4
2	0	1/8	0	0	1/8
3	0	1/4	0	0	1/4
4	0	0	1/4	0	1/4
5	0	0	0	1/8	1/8

Same as

Find:

$f_Y(y)$

y	$f_Y(y)$
1	1/4
2	1/8
3	1/4
4	1/4
5	1/8

Independence (3/3)

Example1 – Answer (2/3):

$f_{XY}(x, y)$

Find:

$f_Y(y)$

y	$f_Y(y)$
1	1/4
2	1/8
3	1/4
4	1/4
5	1/8

Same as

y	1	2	3	4	5
$f_Y(y)$	1/4	1/8	1/4	1/4	1/8

Independence (3/3)

Example1 – Answer (3/3):

$f_{XY}(x, y)$

y \ x	$\mathbf{1}$	$\mathbf{1.5}$	$\mathbf{2.5}$	$\mathbf{3}$
$\mathbf{1}$	$1/4$	0	0	0
$\mathbf{2}$	0	$1/8$	0	0
$\mathbf{3}$	0	$1/4$	0	0
$\mathbf{4}$	0	0	$1/4$	0
$\mathbf{5}$	0	0	0	$1/8$

Are X and Y independent?

Independence (3/3)

Example1 – Answer (3/3):

$f_{XY}(x, y)$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Are X and Y independent?

If:

$$\boxed{f_{XY}(x, y) = f_X(x) f_Y(y) \text{ for all } x \text{ and } y}$$

Independence (3/3)

Example1 – Answer (3/3):

$$f_{XY}(x, y)$$

y \ x	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Same as

x	1	1.5	1.5	2.5	3
y	1	2	3	4	5
$f_{XY}(x, y)$	**1/4**	**1/8**	**1/4**	**1/4**	**1/8**

Independence (3/3)

Example1 – Answer (3/3):

x	1	1.5	1.5	2.5	3
y	1	2	3	4	5
$f_{XY}(x,y)$	**1/4**	**1/8**	**1/4**	**1/4**	**1/8**

x	1	1.5	2.5	3
$f_X(x)$	1/4	3/8	1/4	1/8

y	1	2	3	4	5
$f_Y(y)$	1/4	1/8	1/4	1/4	1/8

$$f_{XY}(1,1) = \frac{1}{4}$$

$$f_X(1) f_Y(1) = \frac{1}{4} * \frac{1}{4} = \frac{1}{16}$$

Independence (3/3)

Example1 – Answer (3/3):

x	1	1.5	1.5	2.5	3
y	1	2	3	4	5
$f_{XY}(x,y)$	**1/4**	**1/8**	**1/4**	**1/4**	**1/8**

x	1	1.5	2.5	3
$f_X(x)$	1/4	3/8	1/4	1/8

y	1	2	3	4	5
$f_Y(y)$	1/4	1/8	1/4	1/4	1/8

$$f_{XY}(1,1) = \frac{1}{4} \quad \neq \quad f_X(1) f_Y(1) = \frac{1}{4} * \frac{1}{4} = \frac{1}{16}$$

Then X and Y are not independent.

Linear Relationship (1/6)

Covariance (1/5):

The covariance between the random variables X and Y is the measure of *linear relationship* between them, denoted as $cov(X, Y)$ or σ_{XY}, where

$$cov(X, Y) = \sigma_{XY} = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big)$$

$$\boxed{= E(XY) - E(X)E(Y)}$$

Linear Relationship (1/6)

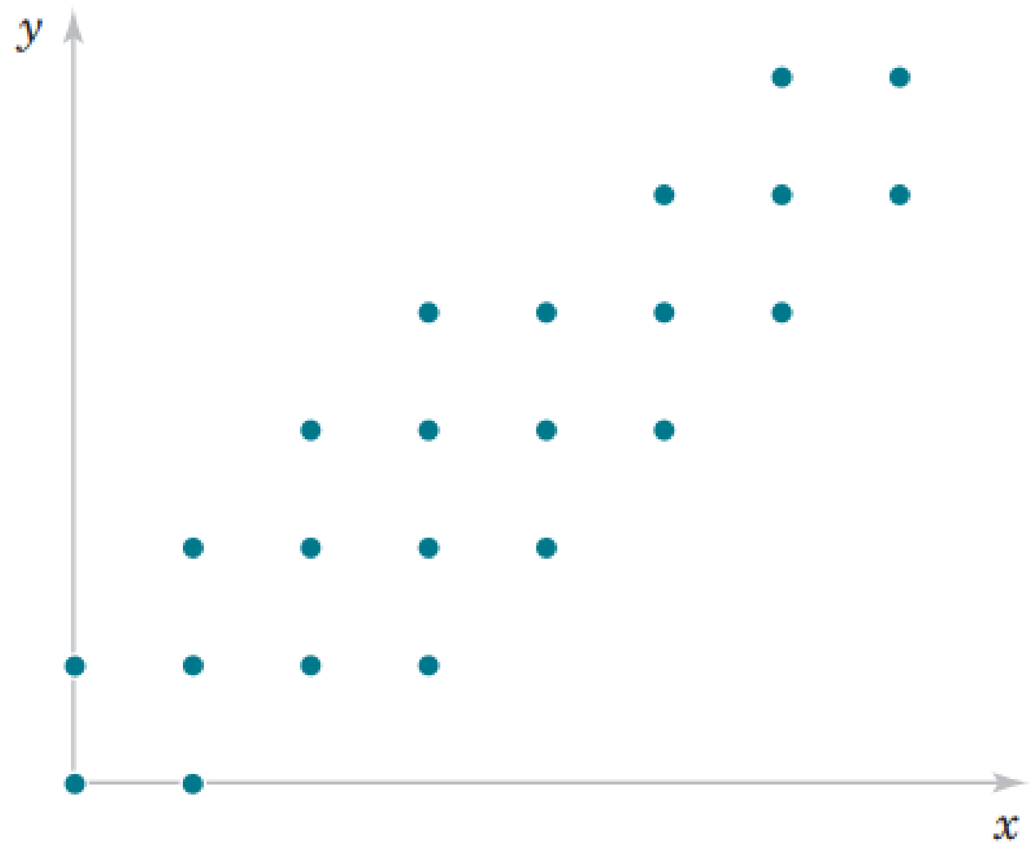
Covariance (2/5):

Covariance measures the total variation of two random variables from their expected values. Using covariance, we can only standard the direction of the relationship. However, it does not indicate the strength of the relationship.

Linear Relationship (1/6)

Covariance (3/5):

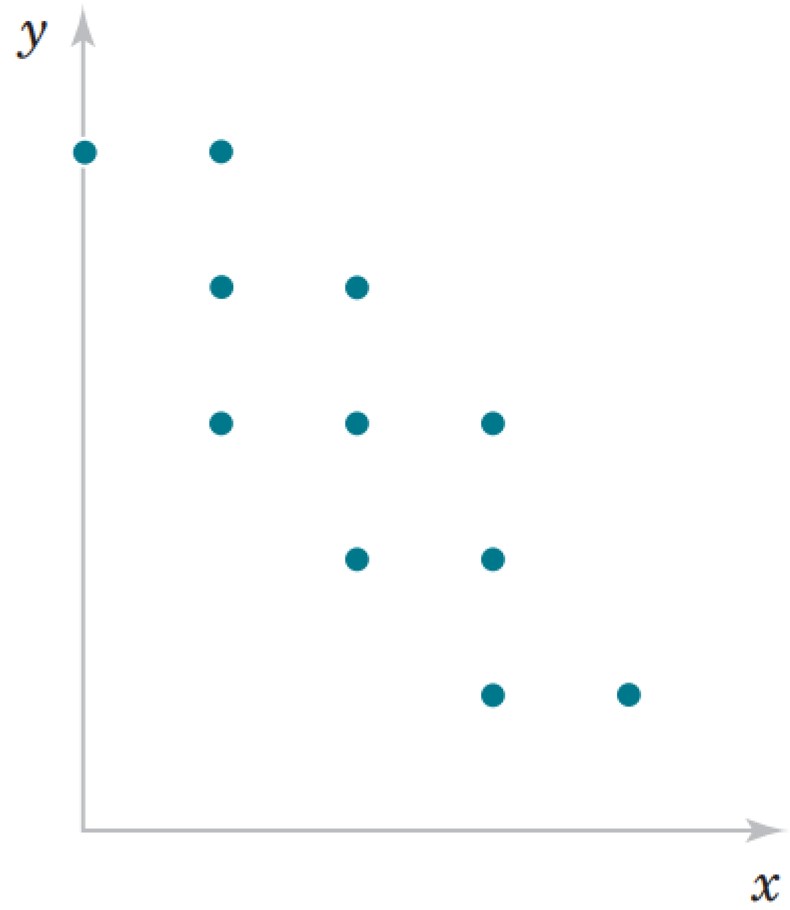
Positive covariance



Linear Relationship (1/6)

Covariance (4/5):

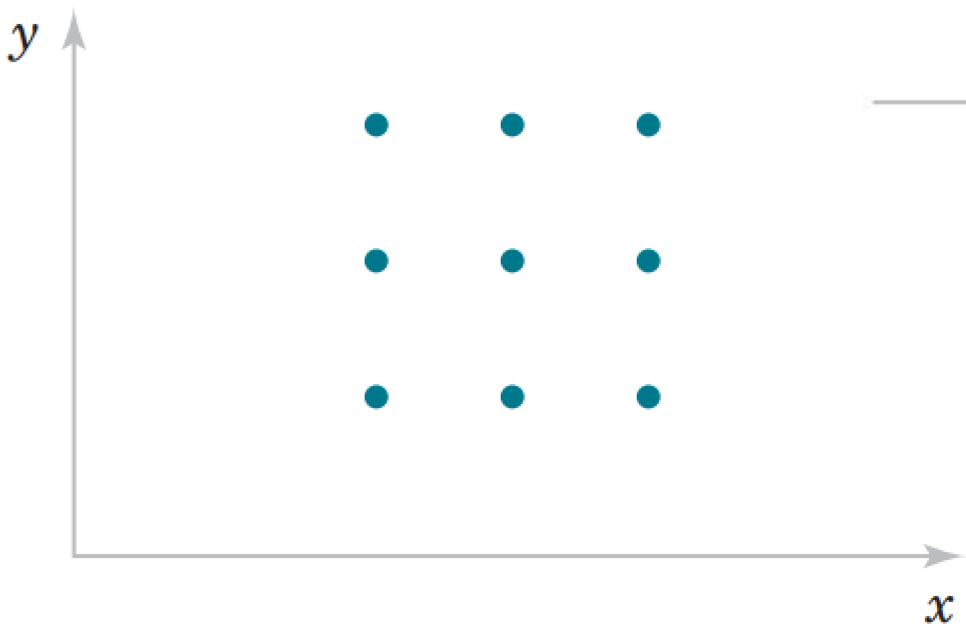
Negative covariance

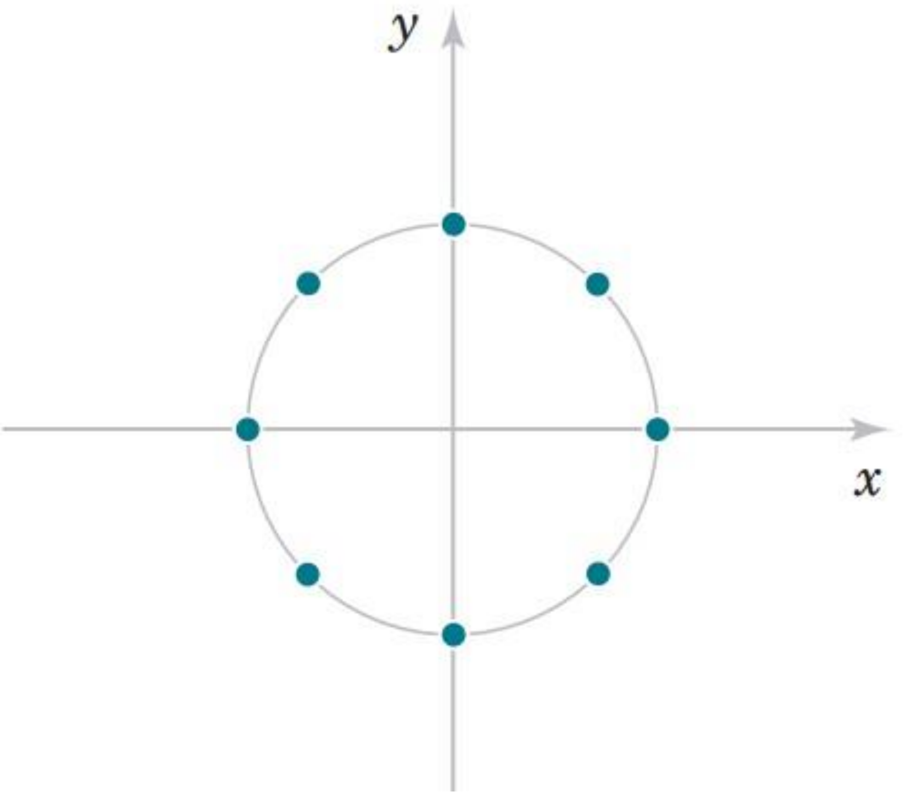


Linear Relationship (1/6)

Covariance (5/5):

<u>Zero covariance</u>





Linear Relationship (2/6)

Example1:

$f_{XY}(x, y)$

$y \backslash x$	**1**	**2**	**4**
3	1/8	0	0
4	1/4	0	0
5	0	1/2	0
6	0	0	1/8

Determine the covariance $\boldsymbol{\sigma_{XY}}$?

$$cov(X, Y) = \sigma_{XY} = E(XY) - E(X)E(Y)$$

Linear Relationship (2/6)

Example1:

$f_{XY}(x, y)$

y \ x	**1**	**2**	**4**
3	1/8	0	0
4	1/4	0	0
5	0	1/2	0
6	0	0	1/8

Same as

x	1	1	2	4
y	3	4	5	6
$f_{XY}(x, y)$	**1/8**	**1/4**	**1/2**	**1/8**

Linear Relationship (3/6)

Example1 – Answer (1/6):

$f_{XY}(x, y)$

$\boldsymbol{x}$	1	1	2	4
$\boldsymbol{y}$	3	4	5	6
$\boldsymbol{f_{XY}(x, y)}$	**1/8**	**1/4**	**1/2**	**1/8**

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{f_{XY}(x, y)}$	$\boldsymbol{x\, f_{XY}(x, y)}$	$\boldsymbol{y\, f_{XY}(x, y)}$	$\boldsymbol{xy\, f_{XY}(x, y)}$
1	3	1/8			
1	4	1/4			
2	5	1/2			
4	6	1/8			

Linear Relationship (3/6)

Example1 – Answer (2/6):

x	y	$f_{XY}(x,y)$	$x\,f_{XY}(x,y)$	$y\,f_{XY}(x,y)$	$xy\,f_{XY}(x,y)$
1	3	1/8	1/8		
1	4	1/4	1/4		
2	5	1/2	2/2		
4	6	1/8	4/8		

Linear Relationship (3/6)

Example1 – Answer (3/6):

x	y	$f_{XY}(x,y)$	$x\,f_{XY}(x,y)$	$y\,f_{XY}(x,y)$	$xy\,f_{XY}(x,y)$
1	3	$1/8$	$1/8$	$3/8$	
1	4	$1/4$	$1/4$	$4/4$	
2	5	$1/2$	$2/2$	$5/2$	
4	6	$1/8$	$4/8$	$6/8$	

Linear Relationship (3/6)

Example1 – Answer (4/6):

x	y	$f_{XY}(x,y)$	$x\, f_{XY}(x,y)$	$y\, f_{XY}(x,y)$	$xy\, f_{XY}(x,y)$
1	3	$1/8$	$1/8$	$3/8$	$3/8$
1	4	$1/4$	$1/4$	$4/4$	$4/4$
2	5	$1/2$	$2/2$	$5/2$	$10/2$
4	6	$1/8$	$4/8$	$6/8$	$24/8$

Linear Relationship (3/6)

Example1 – Answer (5/6):

x	y	$f_{XY}(x,y)$	$x\,f_{XY}(x,y)$	$y\,f_{XY}(x,y)$	$xy\,f_{XY}(x,y)$
1	3	1/8	1/8	3/8	3/8
1	4	1/4	1/4	4/4	4/4
2	5	1/2	2/2	5/2	10/2
4	6	1/8	4/8	6/8	24/8
Sum			**15/8**	**37/8**	**75/8**
			$E(X)$	$E(Y)$	$E(XY)$

Linear Relationship (3/6)

Example1 – Answer (6/6):

$\boldsymbol{E(X) = 15/8}$ $\boldsymbol{E(Y) = 37/8}$ $\boldsymbol{E(XY) = 75/8}$

$$cov(X,Y) = \sigma_{XY} = E(XY) - E(X)E(Y)$$

$$cov(X,Y) = \sigma_{XY} = \frac{75}{8} - \frac{(15)(37)}{64} = 0.703125$$

Positive covariance

Linear Relationship (4/6)

Correlation (1/2):

The correlation between the random variables X and Y is just scales the covariance by the product of the standard deviation of each variable. Correlation measures the strength of the relationship between variables and denoted as ρ_{XY}, where

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$-1 \leq \rho_{XY} \leq +1$$

Linear Relationship (4/6)

Correlation (2/2):

If X and Y are independent random variables,

$$\rho_{XY} = \sigma_{XY} = 0$$

However, if the correlation between two random variables is zero, we cannot conclude that the random variables are independent.

Linear Relationship (5/6)

Example2:

$f_{XY}(x, y)$

y \ x	**1**	**2**	**4**
3	$1/8$	0	0
4	$1/4$	0	0
5	0	$1/2$	0
6	0	0	$1/8$

Determine the correlation ρ_{XY}?

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Linear Relationship (6/6)

Example2 – Answer (1/4): <u>(From the previous example)</u>

x	y	$f_{XY}(x,y)$	$x\, f_{XY}(x,y)$	$y\, f_{XY}(x,y)$	$xy\, f_{XY}(x,y)$
1	3	1/8	1/8	3/8	3/8
1	4	1/4	1/4	4/4	4/4
2	5	1/2	2/2	5/2	10/2
4	6	1/8	4/8	6/8	24/8
Sum			**15/8**	**37/8**	**75/8**
			$E(X)$	$E(Y)$	$E(XY)$

Linear Relationship (6/6)

Example2 – Answer (2/4):

New

x	y	$f_{XY}(x,y)$	$x\,f_{XY}(x,y)$	$y\,f_{XY}(x,y)$	$xy\,f_{XY}(x,y)$	$x^2\,f_{XY}(x,y)$	$y^2\,f_{XY}(x,y)$
1	3	1/8	1/8	3/8	3/8	1/8	9/8
1	4	1/4	1/4	4/4	4/4	1/4	16/4
2	5	1/2	2/2	5/2	10/2	4/2	25/2
4	6	1/8	4/8	6/8	24/8	16/8	36/8
Sum			**15/8**	**37/8**	**75/8**	**35/8**	**177/8**
			$E(X)$	$E(Y)$	$E(XY)$	$E(X^2)$	$E(Y^2)$

Linear Relationship (6/6)

Example2 – Answer (3/4):

$\boldsymbol{E(X) = 15/8}$

$\boldsymbol{E(Y) = 37/8}$

$\boldsymbol{E(XY) = 75/8}$

$\boldsymbol{E(X^2) = 35/8}$

$\boldsymbol{E(Y^2) = 177/8}$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{E(X^2) - \left(E(X)\right)^2} = \sqrt{0.859375}$$

$$\sigma_Y = \sqrt{V(Y)} = \sqrt{E(Y^2) - \left(E(Y)\right)^2} = \sqrt{0.734375}$$

$$\sigma_{XY} = E(XY) - E(X)E(Y) = 0.703125$$

Linear Relationship (6/6)

Example2 – Answer (4/4):

$\boldsymbol{E(X) = 15/8}$ $\boldsymbol{E(Y) = 37/8}$ $\boldsymbol{E(XY) = 75/8}$

$\boldsymbol{E(X^2) = 35/8}$ $\boldsymbol{E(Y^2) = 177/8}$

$$\rho_{XY} = \frac{0.703125}{\sqrt{(0.859375)(0.734375)}} = 0.885079$$

Joint Probability Density Function:

If X and Y are two continuous random variables, the joint probability density fun. is denoted as $f_{XY}(x, y)$, satisfies

1) $f_{XY}(x, y) \geq 0$ for all x, y

2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\ dxdy = 1$

3) For any region R of two-dimensional space,

$$P\big((X, Y) \in R\big) = \iint_R f_{XY}(x, y)\, dxdy$$

Joint Prob. Density Fun. (2/19)

Double Integration:

$$\int_0^1 \int_1^2 \underbrace{xy^2}_{f(x,y)} \, dx\,dy$$

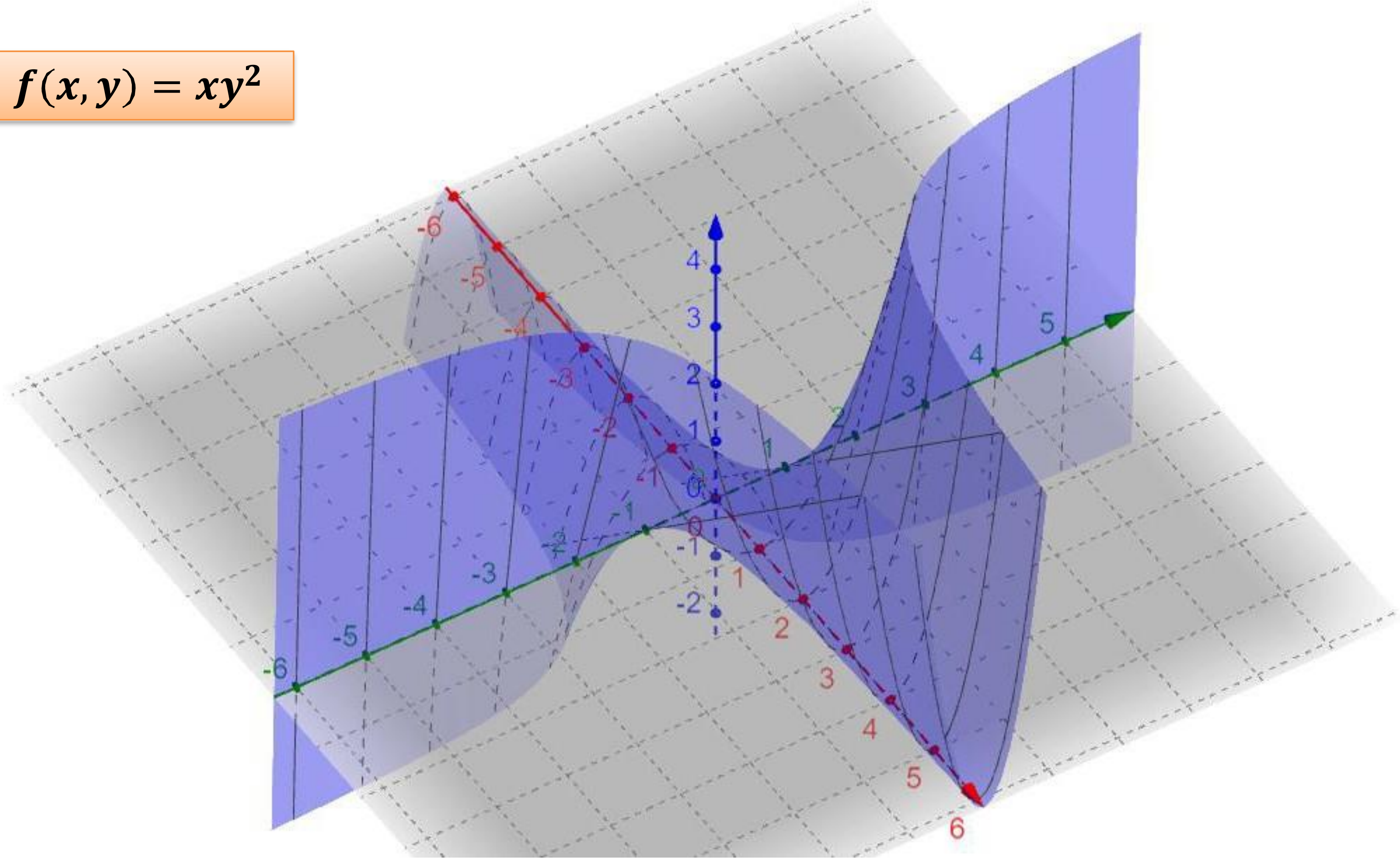
Same as:

$$\iint_R xy^2 \, dx\,dy$$

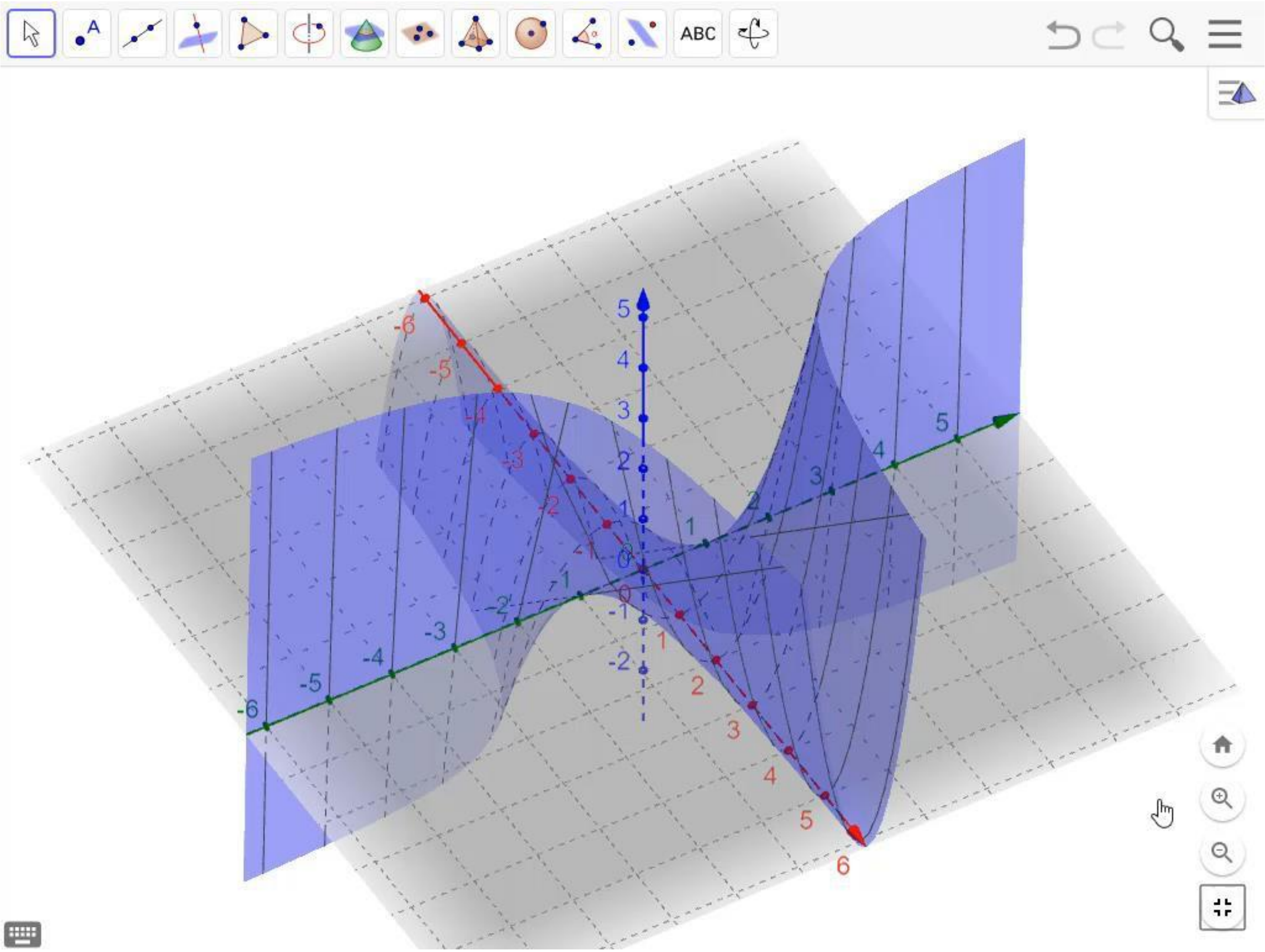
Where $R = \{(x, y) \mid 1 < x < 2, \;\; 0 < y < 1\}$

Joint Prob. Density Fun. (3/19)

$f(x, y) = xy^2$



Joint Prob. Density Fun. (4/19)



Joint Prob. Density Fun. (5/19)

Double Integration:

$$\int_0^1 \boxed{\int_1^2 xy^2 \, dx} dy$$

$$\int_1^2 xy^2 \, dx = \left. \frac{x^2}{2} y^2 \right|_{x=1}^{x=2}$$

Joint Prob. Density Fun. (6/19)

Double Integration:

$$\int_0^1 \boxed{\int_1^2 xy^2 \, dx} \, dy$$

$$\int_1^2 xy^2 \, dx = \frac{x^2}{2} y^2 \Big|_{x=1}^{x=2} = \left(\frac{4}{2} y^2\right) - \left(\frac{1}{2} y^2\right) = \frac{3}{2} y^2$$

Joint Prob. Density Fun. (7/19)

Double Integration:

$$\int_0^1 \boxed{\frac{3}{2}y^2}\,dy$$

$$\int_0^1 \frac{3}{2}y^2\,dy = \frac{y^3}{2}\bigg|_{y=0}^{y=1} = \left(\frac{1}{2}\right) - (0) = \frac{1}{2}$$

Joint Prob. Density Fun. (8/19)

Double Integration:

$$\therefore \int_0^1 \int_1^2 xy^2 \; dxdy = \frac{1}{2}$$

Example1:

Let X and Y are two continuous random variables with a joint probability distribution $f_{XY}(x, y)$, where

$$f_{XY}(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Verify that $f_{XY}(x, y)$ is a joint density function.

b) Find $P[(X, Y) \in A]$,

$$\text{where } A = \{(x,y) \mid 0 < x < \tfrac{1}{2},\ \tfrac{1}{4} < y < \tfrac{1}{2}\}$$

Example1 – Answer (1/5):

$$\int_0^1 \int_0^1 \frac{2}{5}(2x+3y)\ dxdy$$

$$\int_0^1 \left[\frac{2}{5}(x^2+3xy)\right]_{x=0}^{x=1} dy$$

$$\int_0^1 \frac{2}{5}[(1+3y)-(0)]\,dy = \int_0^1 \frac{2}{5}+\frac{6y}{5}\ dy = \left[\frac{2}{5}y+\frac{3y^2}{5}\right]_{y=0}^{y=1}$$

Joint Prob. Density Fun. (10/19)

Example1 – Answer (2/5):

$$\int_0^1 \int_0^1 \frac{2}{5}(2x + 3y)\ dxdy$$

$$\left[\frac{2}{5}y + \frac{3y^2}{5}\right]_{y=0}^{y=1} = \left(\frac{2}{5} + \frac{3}{5}\right) - (0) = \frac{5}{5} = 1$$

$$\therefore \int_0^1 \int_0^1 f_{XY}(x, y)\ dxdy = 1$$

is a joint density function

Example1 – Answer (3/5):

Find $P[(X, Y) \in A]$,

$$\text{where } A = \{(x, y) \mid 0 < x < \tfrac{1}{2},\ \tfrac{1}{4} < y < \tfrac{1}{2}\}$$

$$\int_{1/4}^{1/2} \int_{0}^{1/2} \frac{2}{5}(2x + 3y)\ dxdy$$

$$\int_{1/4}^{1/2} \left[\frac{2}{5}(x^2 + 3xy)\right]_{x=0}^{x=1/2} dy$$

Example1 – Answer (4/5):

$$\int_{1/4}^{1/2}\int_{0}^{1/2}\frac{2}{5}(2x+3y)\ dxdy$$

$$\int_{1/4}^{1/2}\left[\frac{2}{5}(x^2+3xy)\right]_{x=0}^{x=1/2} dy=\int_{1/4}^{1/2}\frac{2}{5}\left(\frac{1}{4}+\frac{3}{2}y\right)-0\ dy$$

$$\int_{1/4}^{1/2}\left(\frac{1}{10}+\frac{3}{5}y\right)dy=\left[\frac{1}{10}y+\frac{3}{10}y^2\right]_{y=1/4}^{y=1/2}$$

Joint Prob. Density Fun. (10/19)

Example1 – Answer (5/5):

$$\int_{1/4}^{1/2}\int_{0}^{1/2}\frac{2}{5}(2x+3y)\ dxdy$$

$$\int_{1/4}^{1/2}\left(\frac{1}{10}+\frac{3}{5}y\right)dy=\left[\frac{1}{10}y+\frac{3}{10}y^2\right]_{y=1/4}^{y=1/2}$$

$$=\left(\frac{1}{20}+\frac{3}{40}\right)-\left(\frac{1}{40}+\frac{3}{160}\right)=0.125-0.04375=\boxed{0.08125}$$

Example2:

Determine the value of $\boldsymbol{c}$ that makes the function $f_{XY}(x, y) = \boldsymbol{c}(x + y)$ a joint probability density function over the range $0 < x < 3$ and $x < y < x + 2$.

Example2 – Answer (1/4):

Determine the value of $\boldsymbol{c}$ that makes the function $f_{XY}(x, y) = \boldsymbol{c}(x + y)$ a joint probability density function over the range $0 < x < 3$ and $x < y < x + 2$.

$$\int_0^3 \int_x^{x+2} \boldsymbol{c}(x + y) \; dydx = 1$$

Joint Prob. Density Fun. (12/19)

Example2 – Answer (2/4):

$$\int_0^3 \int_x^{x+2} \boldsymbol{c}(x+y)\; dydx = 1$$

$$\int_0^3 \left[\boldsymbol{c}\left(xy+\frac{y^2}{2}\right)\right]_{y=x}^{y=x+2} dx$$

$$\int_0^3 \boldsymbol{c}\left[\left(x^2+2x+\frac{(x+2)^2}{2}\right)-\left(x^2+\frac{x^2}{2}\right)\right]dx$$

Joint Prob. Density Fun. (12/19)

Example2 – Answer (3/4):

$$\int_0^3 \int_x^{x+2} \boldsymbol{c}(x+y)\ dydx = 1$$

$$\int_0^3 \boldsymbol{c}\left[\left(x^2+2x+\frac{(x+2)^2}{2}\right)-\left(x^2+\frac{x^2}{2}\right)\right]dx$$

$$\int_0^3 (\boldsymbol{c})\ 4x+2\ dx = [\boldsymbol{c}(2x^2+2x)]_{x=0}^{x=3}$$

Joint Prob. Density Fun. (12/19)

Example2 – Answer (4/4):

$$\int_{0}^{3}\int_{x}^{x+2} \boldsymbol{c}(x+y)\ dydx = 1$$

$$\left[\boldsymbol{c}(2x^2+2x)\right]_{x=0}^{x=3} = \boldsymbol{c}[(18+6)-(0)] = 24\,\boldsymbol{c}$$

$$24\,\boldsymbol{c} = 1 \quad \boxed{\therefore \boldsymbol{c} = \frac{1}{24}}$$

Joint Prob. Density Fun. (13/19)

Marginal Probability Distributions (Continuous):

The marginal distributions of the random variable X alone is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\ dy$$

The marginal distributions of the random variable Y alone is:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\ dx$$

Example3:

Let X and Y are two continuous random variables with a joint probability distribution $f_{XY}(x, y)$, where

$$f_{XY}(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Find $f_X(x)$

b) Find $f_Y(y)$

Joint Prob. Density Fun. (15/19)

Example3 – Answer (1/2):

Find $f_X(x)$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\ dy = \int_0^1 \frac{2}{5}(2x + 3y)\ dy$$

$$\int_0^1 \frac{2}{5}(2x + 3y)\ dy = \left[\frac{2}{5}\left(2xy + \frac{3y^2}{2}\right)\right]_{y=0}^{y=1} = \left(\frac{4x}{5} + \frac{3}{5}\right) - (0)$$

$$f_X(x) = \frac{4x}{5} + \frac{3}{5}$$

Example3 – Answer (2/2):

Find $f_Y(y)$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\ dx = \int_0^1 \frac{2}{5}(2x + 3y)\ dx$$

$$\int_0^1 \frac{2}{5}(2x + 3y)\ dx = \left[\frac{2}{5}(x^2 + 3xy)\right]_{x=0}^{x=1} = \left(\frac{2}{5} + \frac{6y}{5}\right) - (0)$$

$$f_Y(y) = \frac{2}{5} + \frac{6y}{5}$$

Mean from a Joint Distribution (Continuous):

Mean for the random variable X alone is:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y)\, dy dx$$

Mean for the random variable Y alone is:

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{XY}(x, y)\, dy dx$$

Joint Prob. Density Fun. (17/19)

Variance from a Joint Distribution (Continuous):

<u>Variance for the random variable X alone is:</u>

$$V(X) = \boxed{E(X^2)} - \left(E(X)\right)^2$$

and s.d. of $X = \sigma_X = \sqrt{V(X)}$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\; dx$$

<u>Variance for the random variable Y alone is:</u>

$$V(Y) = \boxed{E(Y^2)} - \left(E(Y)\right)^2$$

and s.d. of $Y = \sigma_Y = \sqrt{V(Y)}$

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y)\; dy$$

Joint Prob. Density Fun. (18/19)

Example4:

Let X and Y are two continuous random variables with a joint probability distribution $f_{XY}(x, y)$, where

$$f_{XY}(x,y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Find $E(X)$

b) Find $E(Y)$

c) Find σ_X

Joint Prob. Density Fun. (19/19)

Example4 – Answer (1/4):

$$\boldsymbol{f_X(x) = \frac{4x}{5} + \frac{3}{5}}$$

Find $E(X)$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^1 \frac{4x^2}{5} + \frac{3x}{5}\, dx$$

$$\int_0^1 \frac{4x^2}{5} + \frac{3x}{5}\, dx = \left[\frac{4x^3}{15} + \frac{3x^2}{10}\right]_{x=0}^{x=1} = \left(\frac{4}{15} + \frac{3}{10}\right) - (0)$$

$$E(X) = \frac{4}{15} + \frac{3}{10} = 0.5667$$

Joint Prob. Density Fun. (19/19)

Example4 – Answer (2/4):

$$f_Y(y) = \frac{2}{5} + \frac{6y}{5}$$

Find $E(Y)$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \int_0^1 \frac{2y}{5} + \frac{6y^2}{5}\ dy$$

$$\int_0^1 \frac{2y}{5} + \frac{6y^2}{5}\ dy = \left[\frac{y^2}{5} + \frac{2y^3}{5}\right]_{y=0}^{y=1} = \left(\frac{1}{5} + \frac{2}{5}\right) - (0)$$

$$E(Y) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5} = 0.6$$

Joint Prob. Density Fun. (19/19)

Example4 – Answer (3/4):

$\boldsymbol{E(X) = 0.5667}$

Find σ_X

$$\sigma_X = \sqrt{V(X)} = \sqrt{E(X^2) - \left(E(X)\right)^2}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\ dx = \int_0^1 \frac{4x^3}{5} + \frac{3x^2}{5}\ dx$$

$$E(X^2) = \left[\frac{x^4}{5} + \frac{x^3}{5}\right]_{x=0}^{x=1} = \left(\frac{1}{5} + \frac{1}{5}\right) - (0) = \frac{2}{5} = 0.4$$

Joint Prob. Density Fun. (19/19)

Example4 – Answer (4/4):

$\boldsymbol{E(X) = 0.5667}$

$\boldsymbol{E\left(X^2\right) = 0.4}$

Find σ_X

$$\sigma_X = \sqrt{V(X)} = \sqrt{E(X^2) - \left(E(X)\right)^2}$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{0.4 - (0.5667)^2} = \sqrt{0.07885111}$$

$$\sigma_X = 0.2808044$$

Chapter 3: Prob. Distributions

- Discrete Uniform Distribution.
- Binomial Distribution.
- Continuous Uniform Distribution.
- Normal Distribution.

Discrete Uniform Dist. (1/6)

Definition:

A random variable X has a discrete uniform distribution if each of the n values in its range, x_1, x_2, ... , x_n, has equal probability. Then

$$f(x_i) = \frac{1}{n}$$

Discrete Uniform Dist. (2/6)

Example1 (1/2):

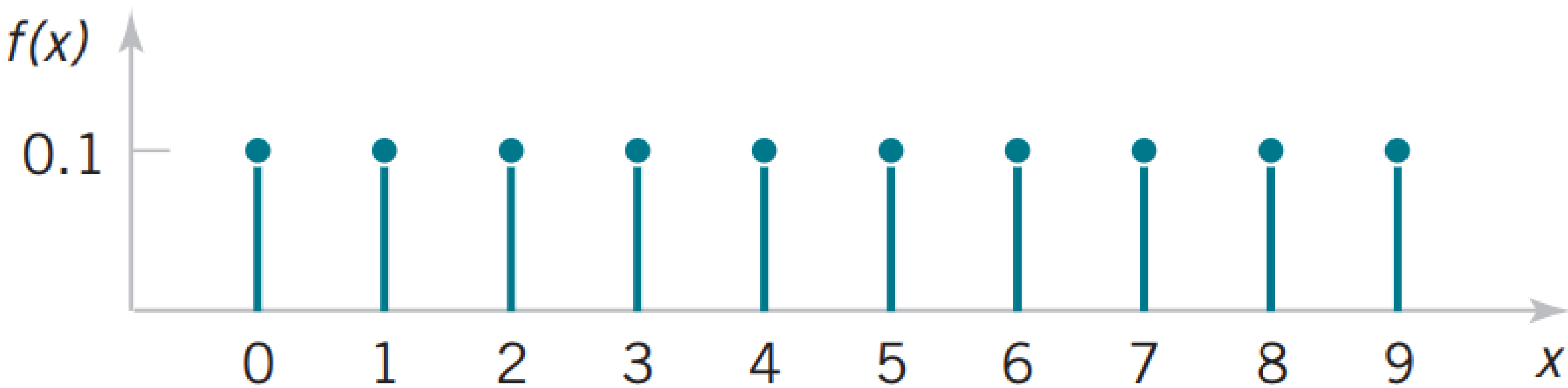
The first digit of a part's serial number is equally likely to be any one of the digits 0 through 9. If one part is selected randomly from a large batch and X is the first digit of the serial number, X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \ldots, 9\}$. That is,

$$f(x) = \frac{1}{10} = 0.1$$

Discrete Uniform Dist. (2/6)

Example1 (2/2):

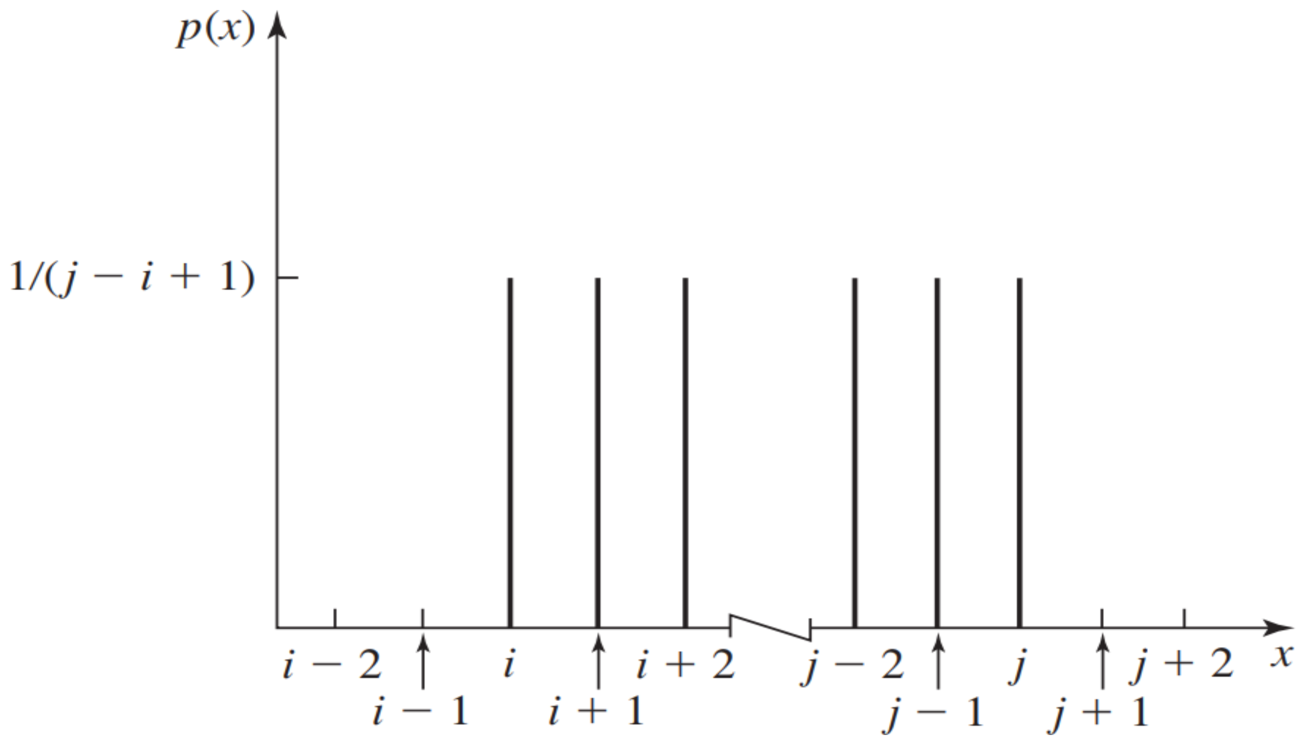
Probability mass function for a discrete uniform random variable



Discrete Uniform Dist. (3/6)

Discrete Uniform $DU(i, j)$:

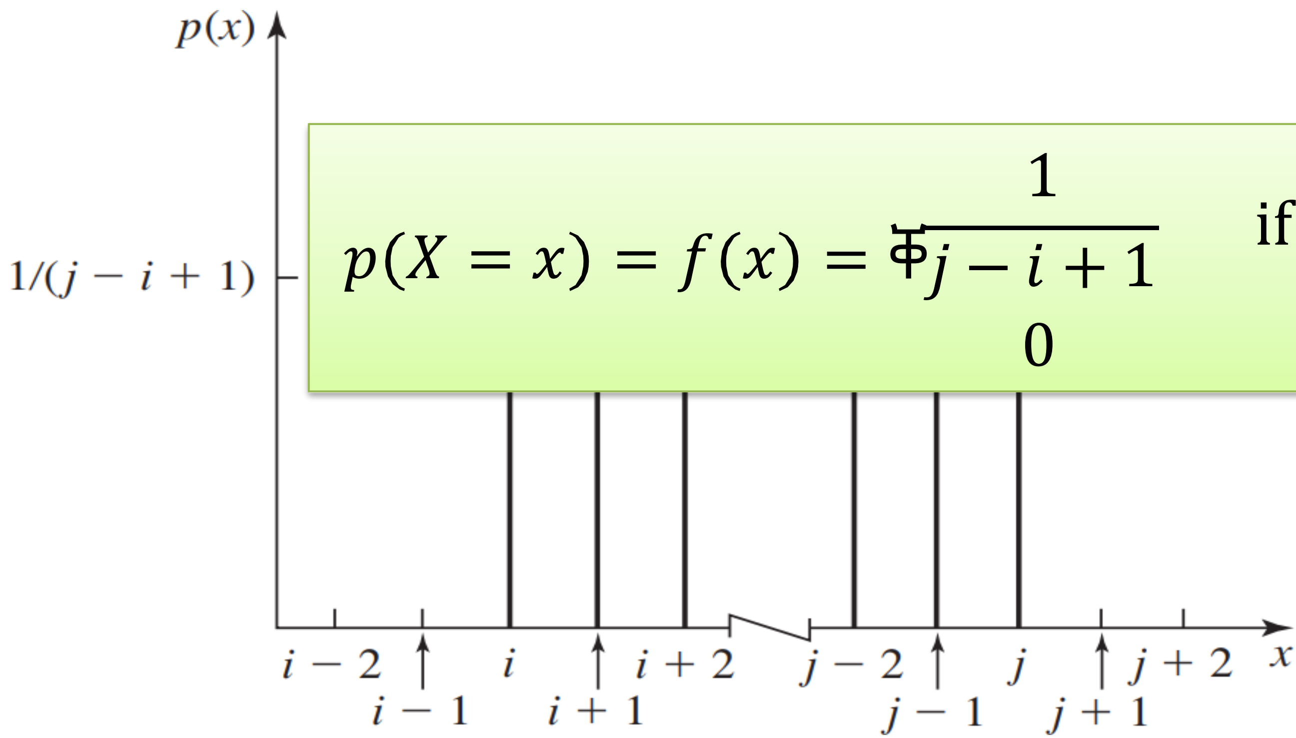
Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \ldots, j$, for $i \leq j$



Discrete Uniform Dist. (3/6)

Discrete Uniform $DU(i, j)$:

Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \ldots, j$, for $i \le j$

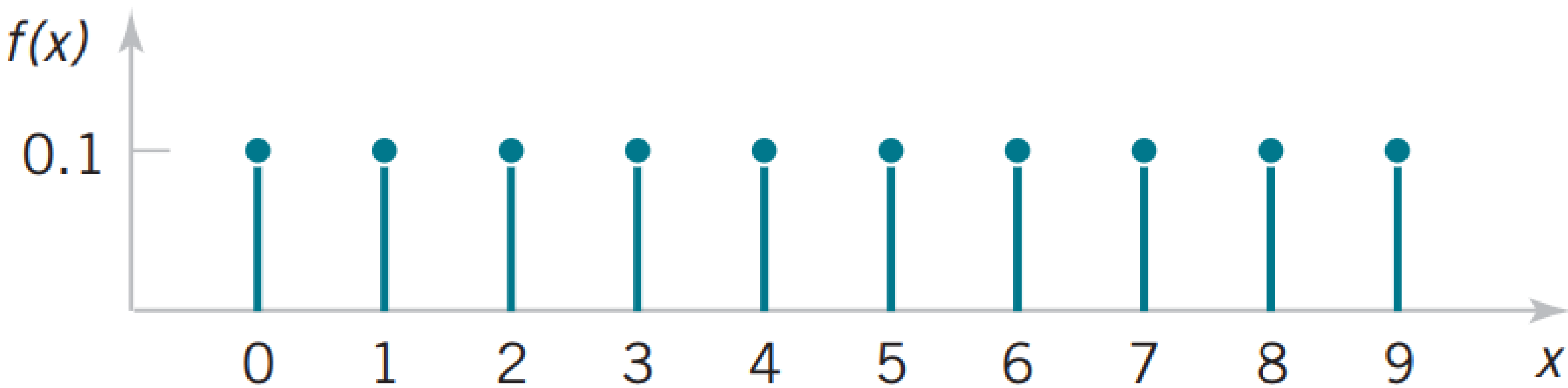


$$p(X = x) = f(x) = \begin{cases} \dfrac{1}{j - i + 1} & \text{if } x \in \{i, i+1, \ldots, j\} \\ 0 & \text{otherwise} \end{cases}$$

Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

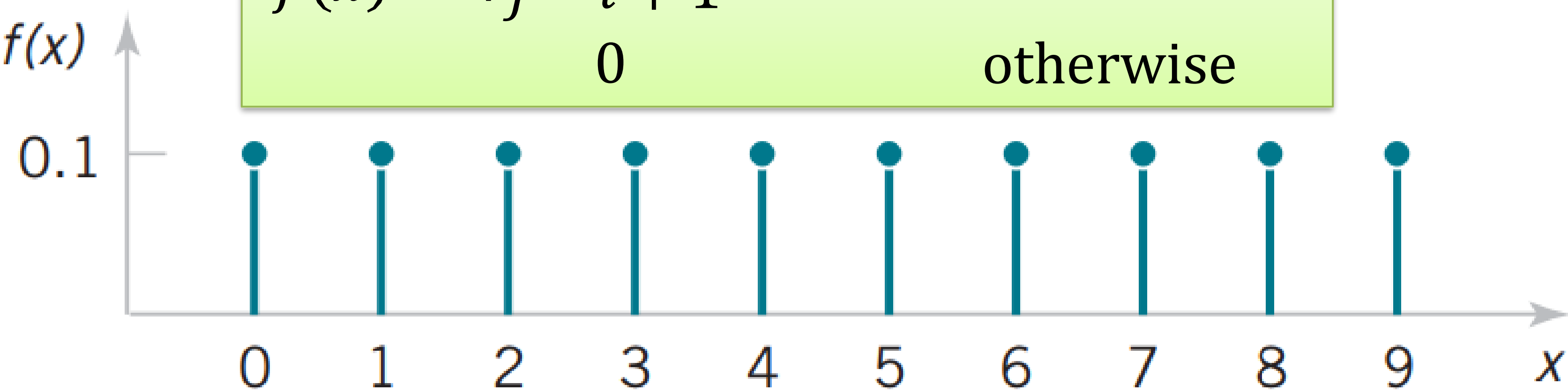


Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} \dfrac{1}{j-i+1} & \text{if } x \in \{i, i+1, \ldots, j\} \\ 0 & \text{otherwise} \end{cases}$$

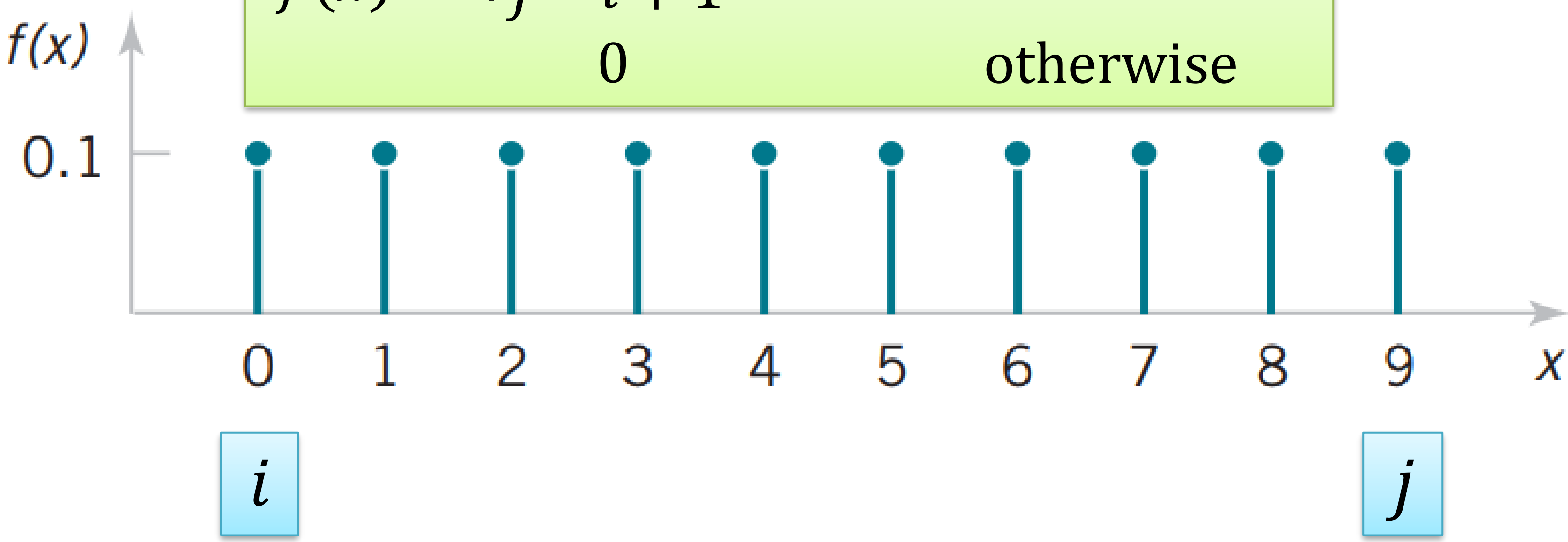


Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} \dfrac{1}{j - i + 1} & \text{if } x \in \{i, i+1, \ldots, j\} \\ 0 & \text{otherwise} \end{cases}$$

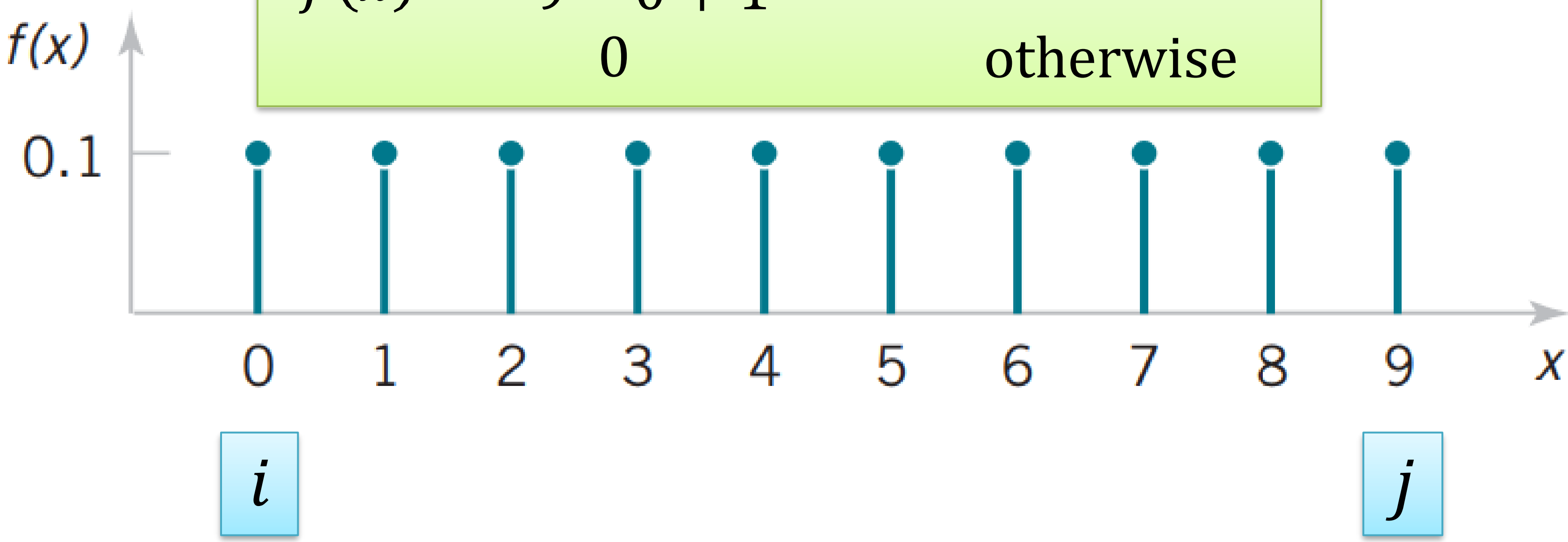


Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} \dfrac{1}{9-0+1} & \text{if } x \in \{0, 1, \dots, 9\} \\ 0 & \text{otherwise} \end{cases}$$

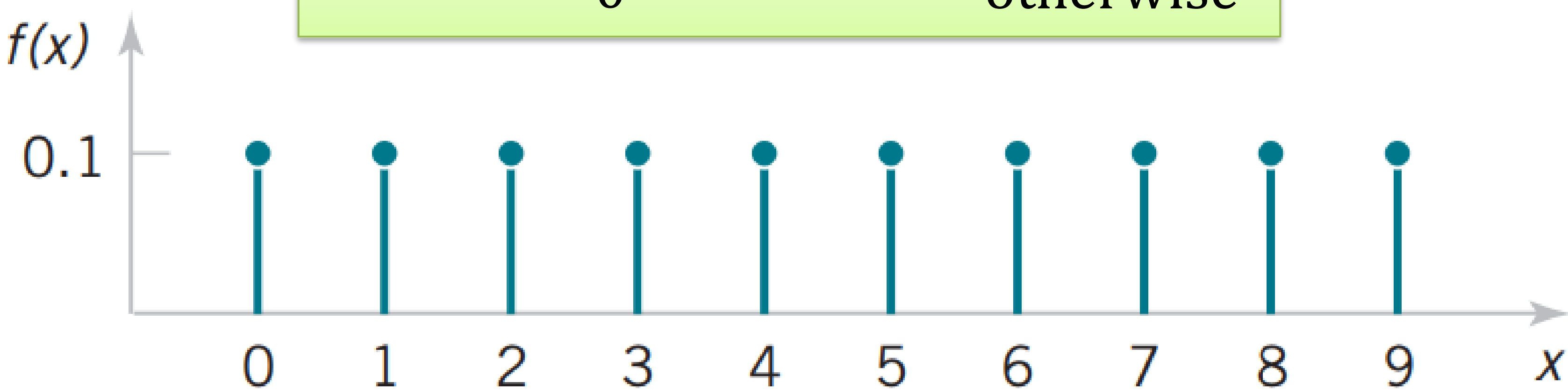


Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} 1/10 & \text{if } x \in \{0, 1, \ldots, 9\} \\ 0 & \text{otherwise} \end{cases}$$



i j

Discrete Uniform Dist. (5/6)

Mean and Variance:

Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \ldots, j$, for $i \leq j$

$$\mu = E(X) = \frac{j+i}{2}$$

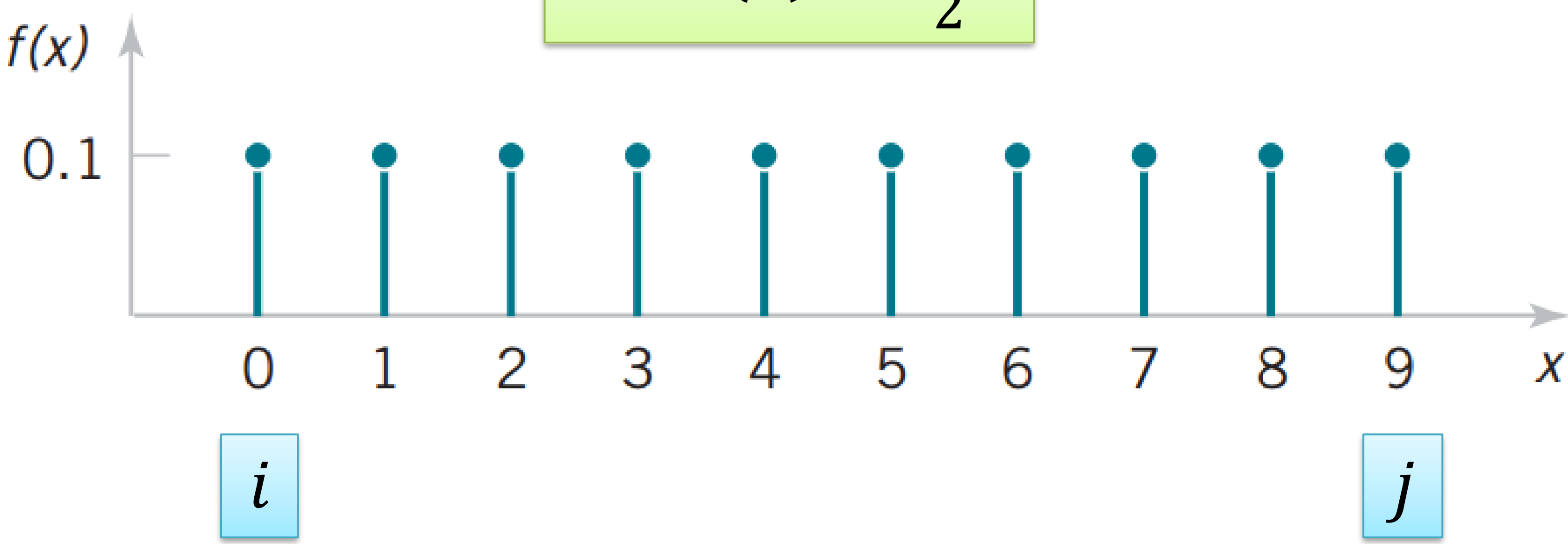
$$\sigma^2 = V(X) = \frac{(j-i+1)^2 - 1}{12}$$

Discrete Uniform Dist. (6/6)

Example2 (1/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\mu = E(X) = \frac{j+i}{2}$$

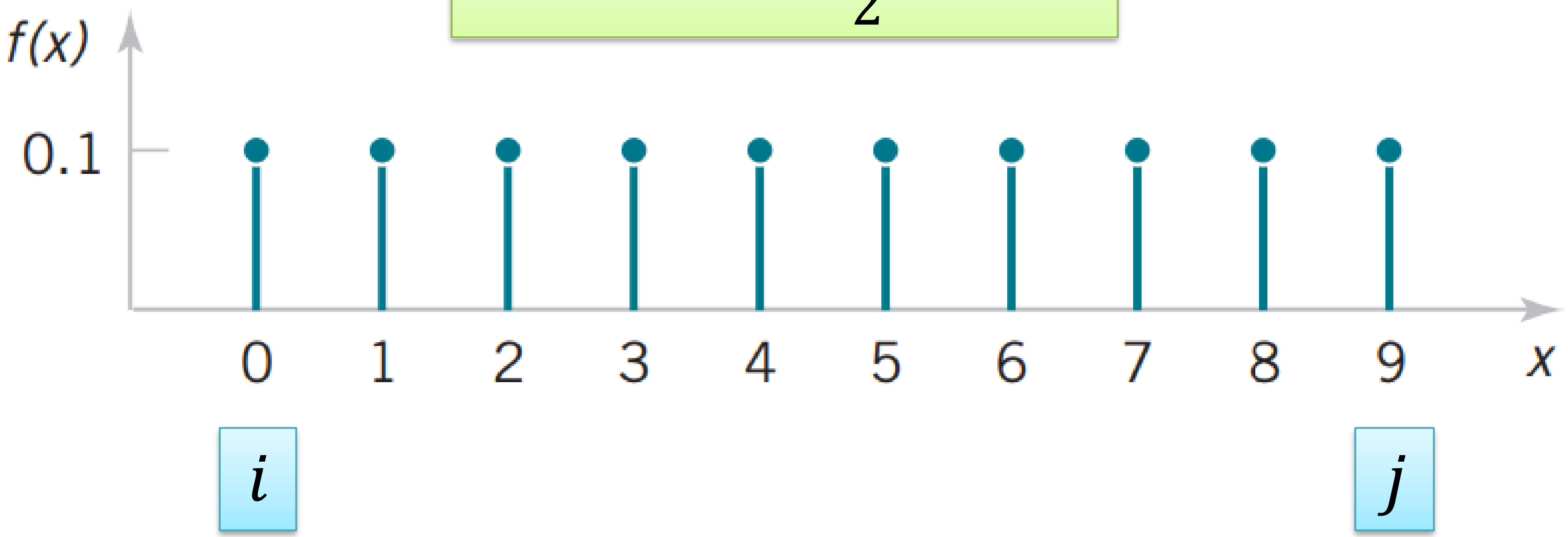


Discrete Uniform Dist. (6/6)

Example2 (2/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\mu = E(X) = \frac{9+0}{2} = 4.5$$

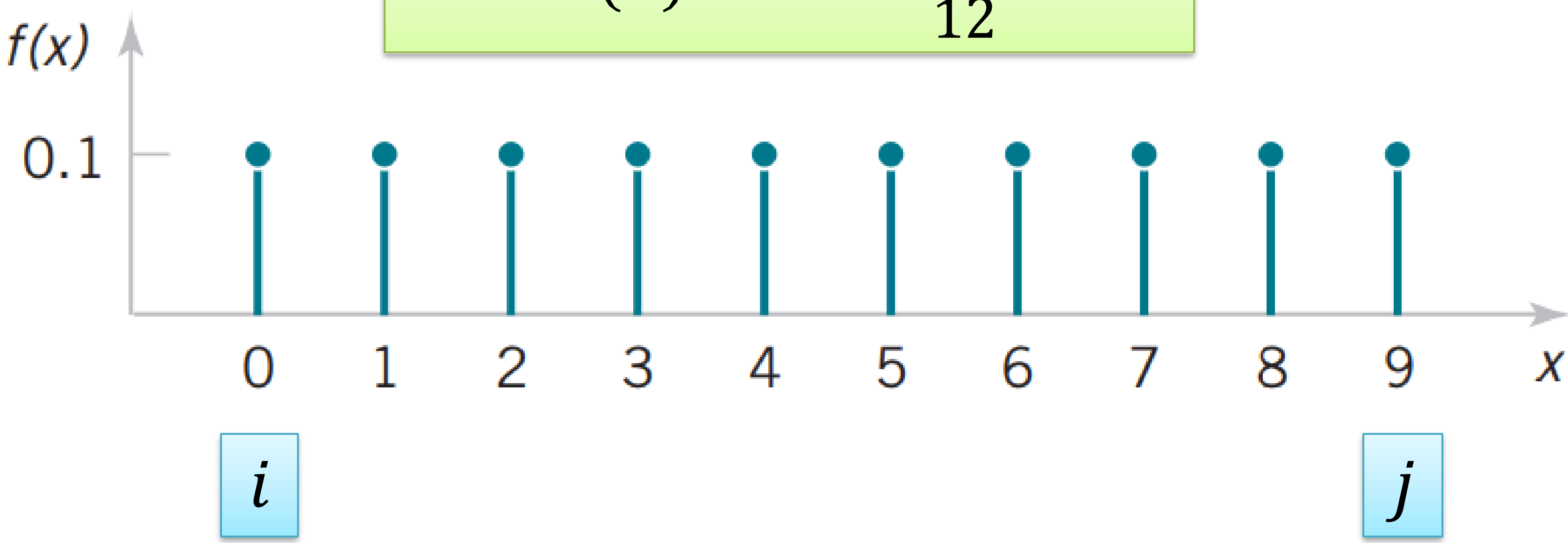


Discrete Uniform Dist. (6/6)

Example2 (3/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\sigma^2 = V(X) = \frac{(j - i + 1)^2 - 1}{12}$$

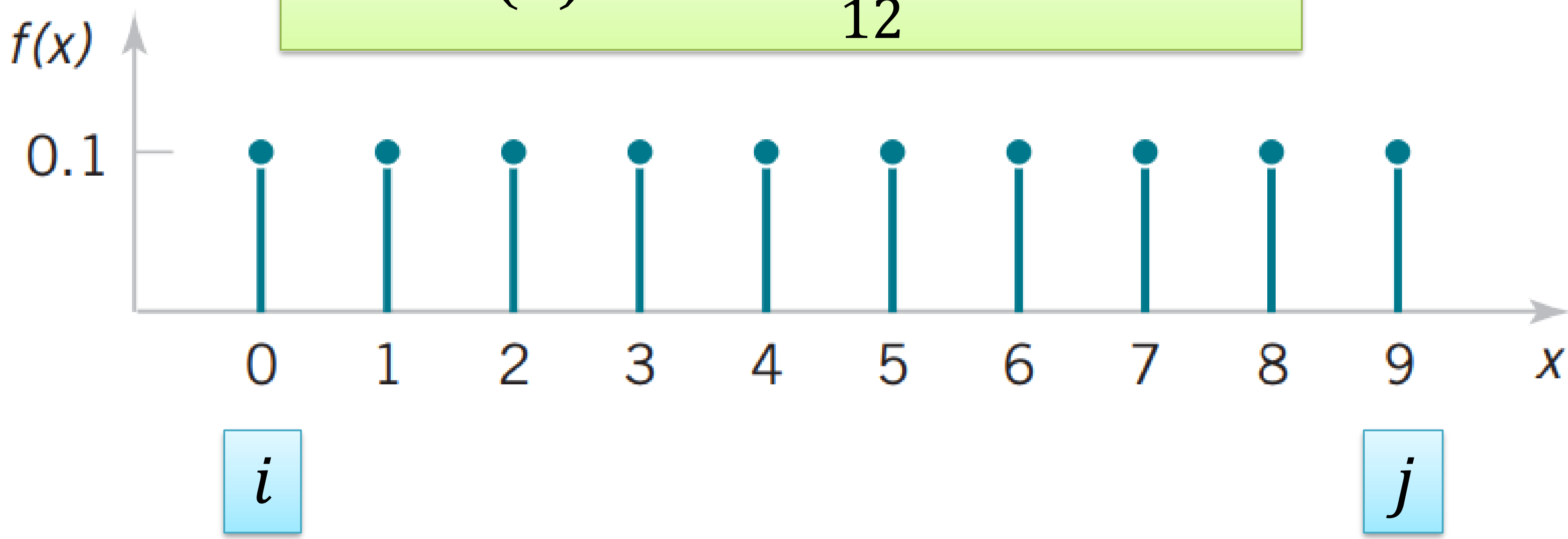


Discrete Uniform Dist. (6/6)

Example2 (4/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\sigma^2 = V(X) = \frac{(9-0+1)^2 - 1}{12} = 8.25$$



Binomial Distribution (1/10)

The Bernoulli Process:

1. The experiment consists of ***repeated*** trials and each trial is called a ***Bernoulli trial***.
2. Each trial results in an outcome that may be classified as a ***success*** or a ***failure***.
3. The probability of success, denoted by $\boldsymbol{p}$, remains constant from trial to trial.
4. The repeated trials are ***independent***.

Binomial Distribution (2/10)

The Bernoulli Process: (Examples)

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A multiple-choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
3. In the next 20 births at a hospital, let X = the number of female births.

Binomial Distribution (3/10)

Binomial Distribution $b(n, p)$:

A random experiment consists of n Bernoulli trials. The random variable X that equals the number of trials that result in a success is a *binomial random variable* with parameters $0 < p < 1$ and finite $n = 1, 2, \ldots$.

The probability mass function of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

Binomial Distribution (4/10)

Example 1 (1/5):

Flip a coin 10 times, what is the probability that 3 heads occurs?

Binomial Distribution (4/10)

Example 1 (2/5):

Flip a coin 10 times, what is the probability that 3 heads occurs?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (4/10)

Example 1 (3/5):

Flip a coin 10 times, what is the probability that 3 heads occurs?

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

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (4/10)

Example 1 (4/5):

Flip a coin 10 times, what is the probability that 3 heads occurs?

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (4/10)

Example 1 (5/5):

Flip a coin 10 times, what is the probability that $\boxed{\text{3 heads}}$ occurs?

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

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

$$f(3) = \binom{10}{3} (0.5)^3 (0.5)^{10-3}$$

$$= 120\,(0.5)^3 (0.5)^7$$

$$= 0.1171875$$

Binomial Distribution (5/10)

Example 2 (1/5):

Flip a coin 10 times, what is the probability that:

a) No head occurs.

b) At least 2 heads occurs.

Binomial Distribution (5/10)

Example 2 (2/5):

Flip a coin 10 times, what is the probability that:

a) No head occurs.

b) At least 2 heads occurs.

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

Binomial Distribution (5/10)

Example 2 (3/5):

Flip a coin 10 times, what is the probability that:

a) No head occurs.

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$f(0) = \binom{10}{0} (0.5)^0 (0.5)^{10-0}$$

Binomial Distribution (5/10)

Example 2 (4/5):

Flip a coin 10 times, what is the probability that:

b) At least 2 heads occurs.

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

Binomial Distribution (5/10)

Example 2 (4/5):

Flip a coin 10 times, what is the probability that:

b) At least 2 heads occurs.

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$P(X \geq 2)$

Binomial Distribution (5/10)

Example 2 (4/5):

Flip a coin 10 times, what is the probability that:

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

b) At least 2 heads occurs.

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$P(X \geq 2) = f(2) + f(3) + \cdots + f(10)$$

Binomial Distribution (5/10)

Example 2 (4/5):

Flip a coin 10 times, what is the probability that:

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

b) At least 2 heads occurs.

$$P(X \geq 2) = \cancel{f(2) + f(3) + \cdots} \\ \cancel{+ f(10)}$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [f(1) + f(0)]$$

Binomial Distribution (5/10)

Example 2 (5/5):

Flip a coin 10 times, what is the probability that:

b) At least 2 heads occurs.

$$f(0) = \binom{10}{0}(0.5)^0(0.5)^{10-0}$$

$$f(1) = \binom{10}{1}(0.5)^1(0.5)^{10-1}$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [f(1) + f(0)]$$

Binomial Distribution (6/10)

Mean and Variance $b(n, p)$:

If X is a binomial random variable with parameters p and n

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

$$\sigma^2 = V(X) = np(1-p)$$

Example 3 (1/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

Binomial Distribution (7/10)

Example 3 (2/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (7/10)

Example 3 (3/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \dots, 10$

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

$$\sigma^2 = V(X) = np(1-p)$$

Binomial Distribution (7/10)

Example 3 (4/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

$$\mu = E(X) = np$$
$$= 10 \times 0.5 = 5$$

$$\sigma^2 = V(X) = np(1 - p)$$

Binomial Distribution (7/10)

Example 3 (5/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

$$\mu = E(X) = np = 10 \times 0.5 = 5$$

$$\sigma^2 = V(X) = np(1-p) = 10 \times 0.5 \times 0.5 = 2.5$$

Example 4 (1/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

Example 4 (2/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
    - Values $x = 0, 1, 2, \dots, 15$

Binomial Distribution (8/10)

Example 4 (3/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \dots, 15$

Binomial Distribution (8/10)

Example 4 (3/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

$$f(x) = \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \ldots, 15$

Binomial Distribution (8/10)

Example 4 (4/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

$$f(x) = \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

(a) exactly 5 survive

$$f(5) = \binom{15}{5} (0.4)^5 (0.6)^{15-5}$$

$$= 0.1859$$

Binomial Distribution (8/10)

Example 4 (5/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

$$f(x) = \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

(b) at least 3 survive

$$P(X \geq 3) = 1 - P(X < 3)$$
$$= 1 - [f(2) + f(1) + f(0)]$$

Example 5 (1/4):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

Binomial Distribution (9/10)

Example 5 (2/4):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

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

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \dots, 15$

$$\sigma^2 = V(X) = np(1-p)$$

Example 5 (3/4):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

$$\boxed{\begin{aligned}\mu &= E(X) = np \\ &= 15 \times 0.4 = 6\end{aligned}}$$

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \ldots, 15$

$$\boxed{\sigma^2 = V(X) = np(1-p)}$$

Binomial Distribution (9/10)

Example 5 (4/4):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

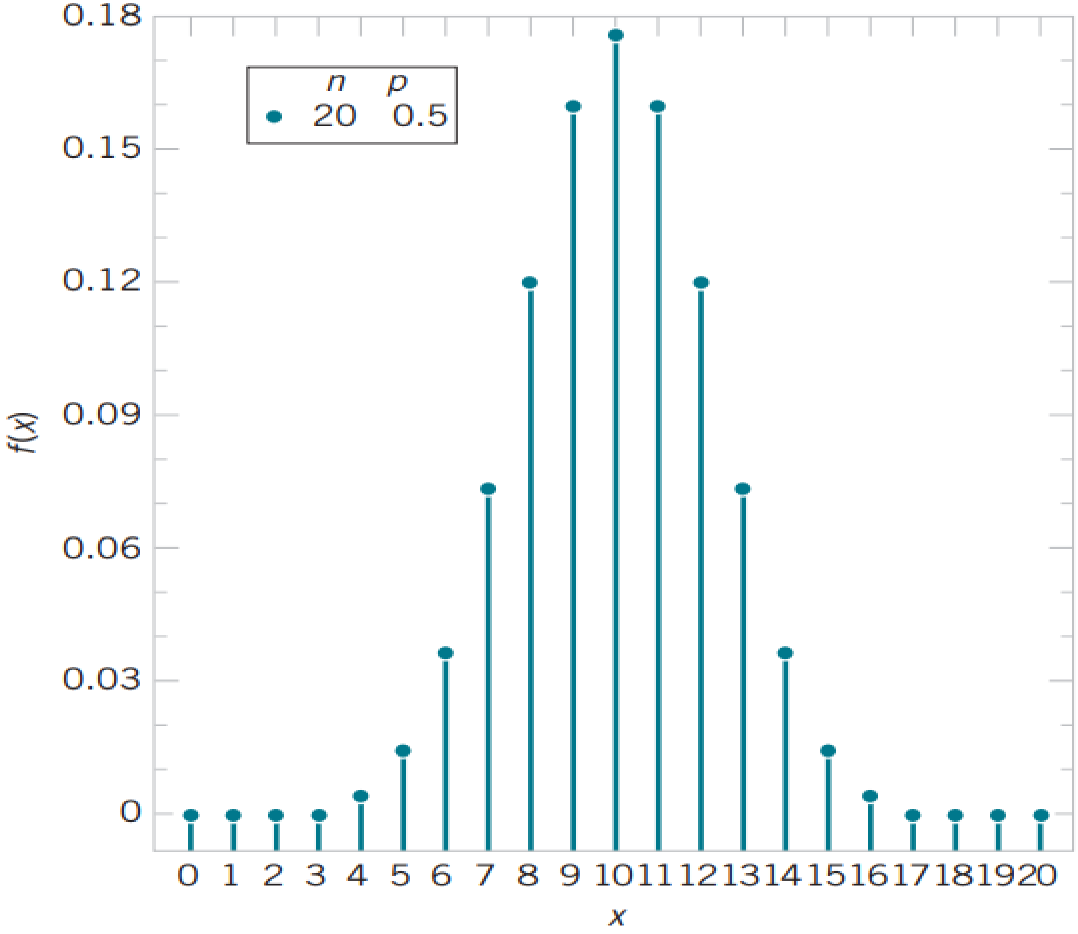
- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \ldots, 15$

$$\mu = E(X) = np$$
$$= 15 \times 0.4 = 6$$

$$\sigma^2 = V(X) = np(1-p)$$
$$= 15 \times 0.4 \times 0.6$$
$$= 3.6$$

Binomial Distribution (10/10)

Binomial distribution for selected values of $n = 20$ and $p = 0.5$



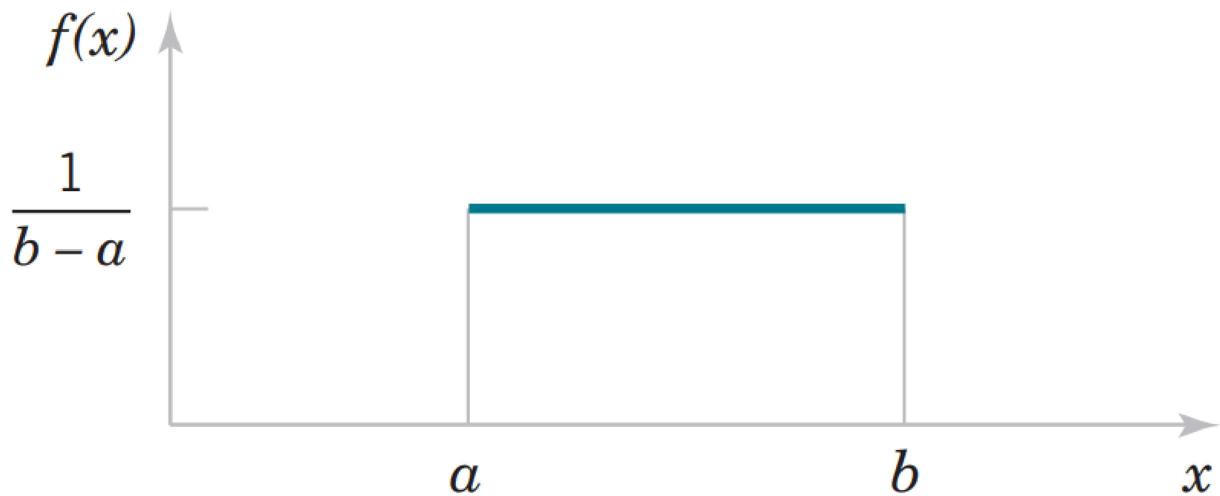
Continuous Uniform Dist. (1/2)

Continuous Uniform Distribution

A continuous random variable X with probability density function

$$f(x) = 1/(b - a), \qquad a \le x \le b$$

is a **continuous uniform random variable**.



Continuous Uniform Dist. (2/2)

Mean and Variance

If X is a continuous uniform random variable over $a \le x \le b$,

$$\mu = E(X) = \frac{a+b}{2}$$

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

Normal Distribution (1/2)

Normal Distribution

A random variable X with probability density function

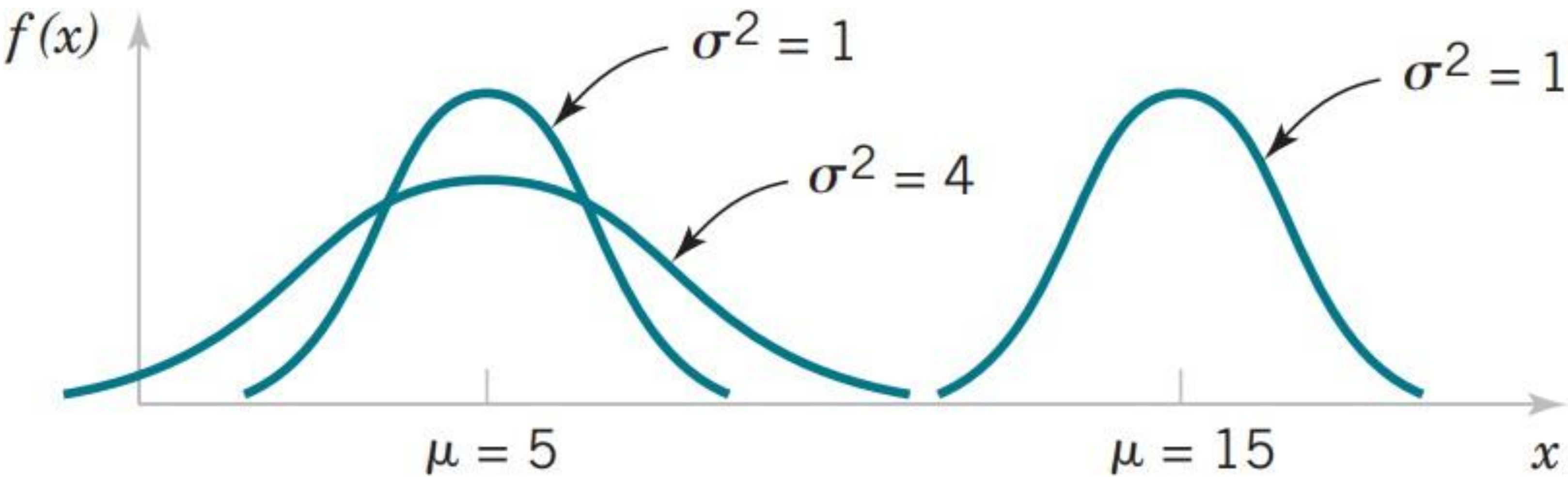
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

is a **normal random variable** with parameters μ where $-\infty < \mu < \infty$ and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution.

Normal Distribution (2/2)



Ch 3.1: Discrete Prob. Distr.

- Discrete Uniform Distribution.
- Binomial and Multinomial Distributions.
- Negative Binomial Distributions.
- Geometric Distributions.
- Poisson Distribution.

Discrete Uniform Dist. (1/6)

Definition:

A random variable X has a discrete uniform distribution if each of the n values in its range, x_1, x_2, ... , x_n, has equal probability. Then

$$f(x_i) = \frac{1}{n}$$

Discrete Uniform Dist. (2/6)

Example1 (1/2):

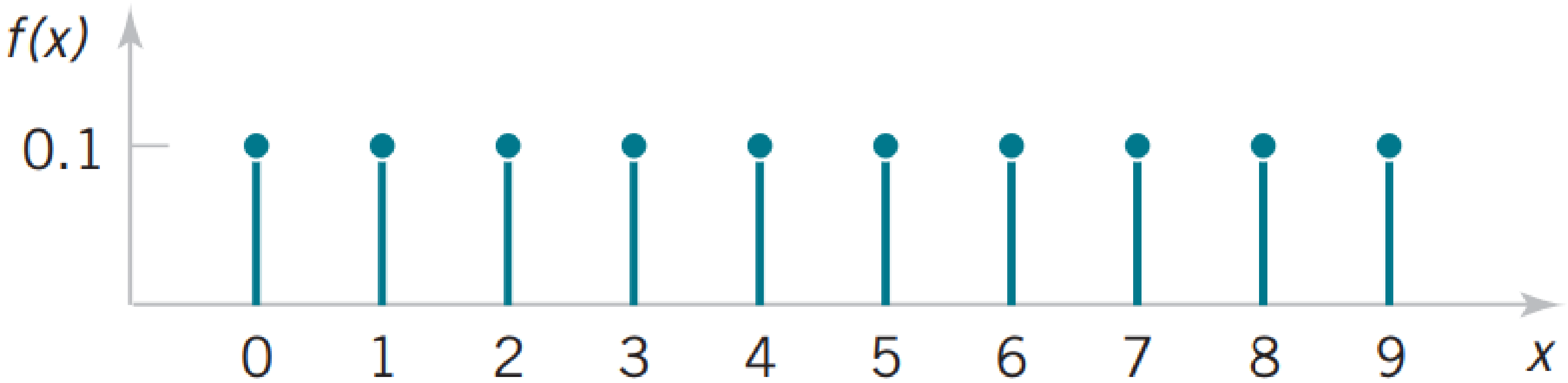
The first digit of a part's serial number is equally likely to be any one of the digits 0 through 9. If one part is selected randomly from a large batch and X is the first digit of the serial number, X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \dots, 9\}$. That is,

$$f(x) = \frac{1}{10} = 0.1$$

Discrete Uniform Dist. (2/6)

Example1 (2/2):

Probability mass function for a discrete uniform random variable

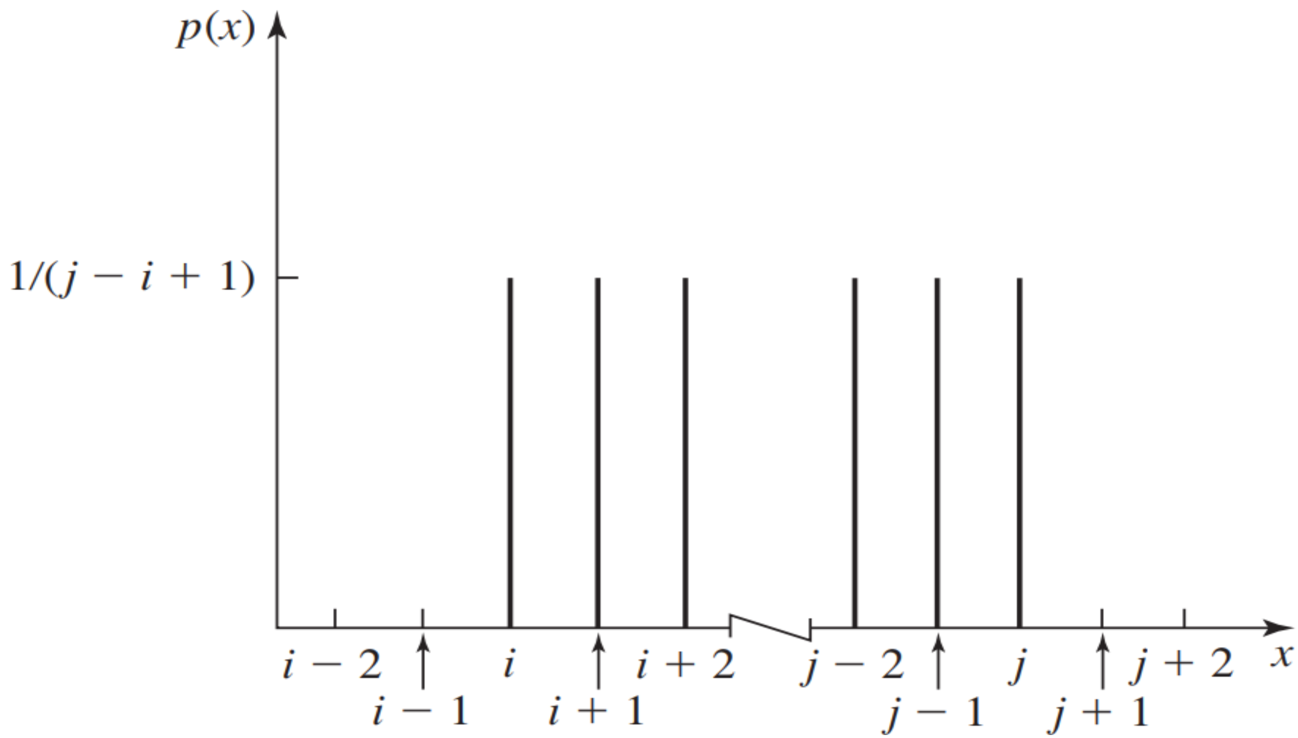


Discrete Uniform Dist. (3/6)

Discrete Uniform DU(i, j):

$X \sim \mathrm{DU}(i, j)$

Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \dots, j$, for $i \leq j$



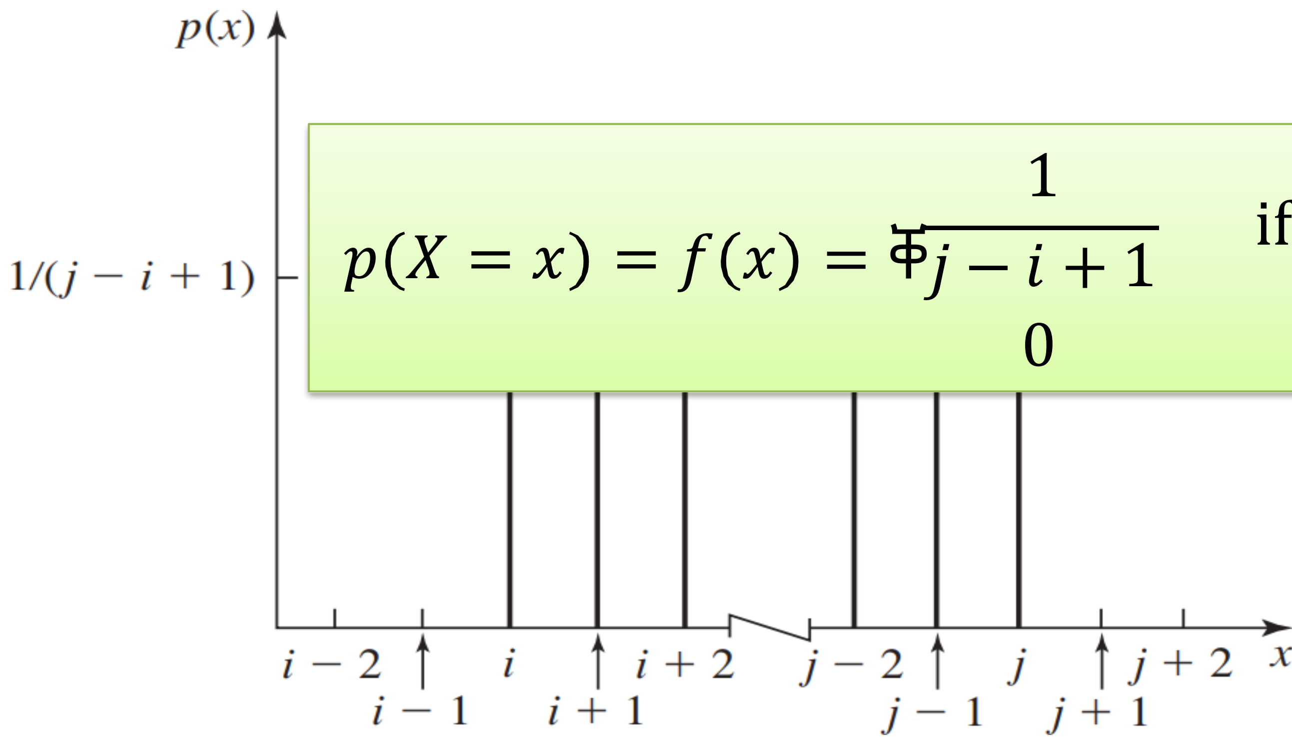
Discrete Uniform Dist. (3/6)

Discrete Uniform DU(i, j):

$X \sim \mathrm{DU}(i, j)$

Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \ldots, j$, for $i \le j$

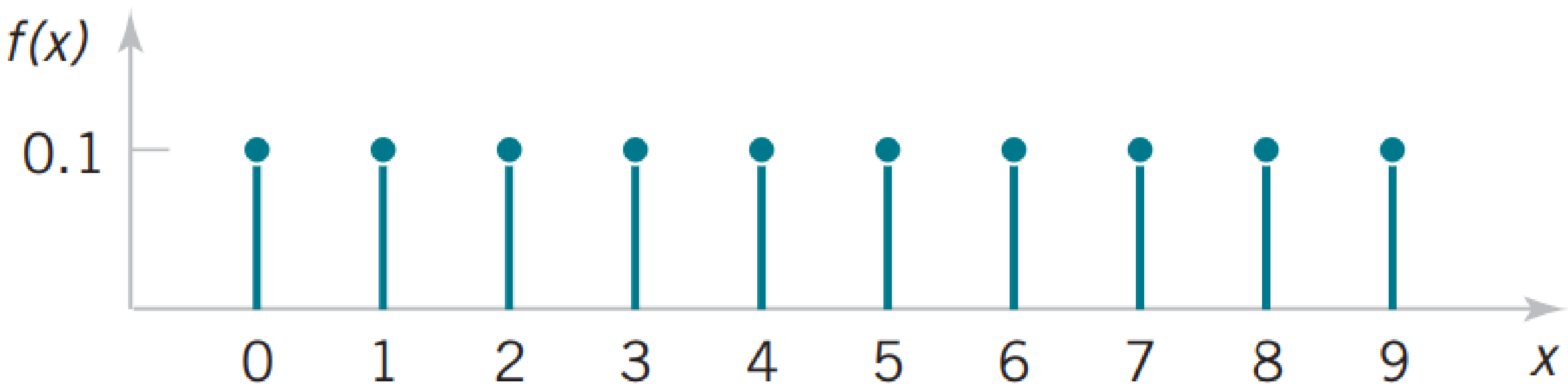
$$p(X = x) = f(x) = \begin{cases} \dfrac{1}{j-i+1} & \text{if } x \in \{i, i+1, \ldots, j\} \\ 0 & \text{otherwise} \end{cases}$$



Discrete Uniform Dist. (4/6)

Recall Example1:

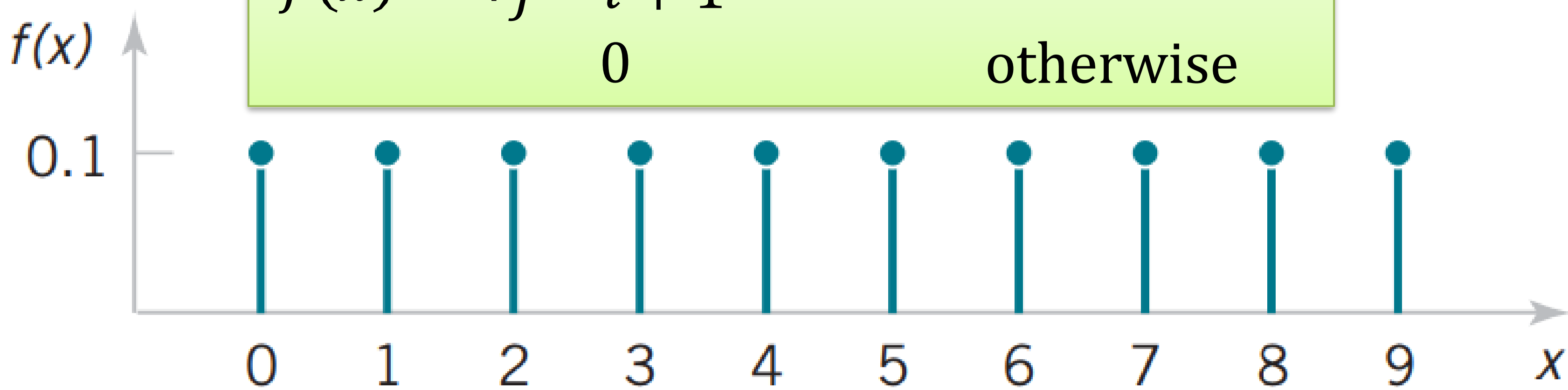
Probability mass function for a discrete uniform random variable



Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} \dfrac{1}{j-i+1} & \text{if } x \in \{i, i+1, \dots, j\} \\ 0 & \text{otherwise} \end{cases}$$

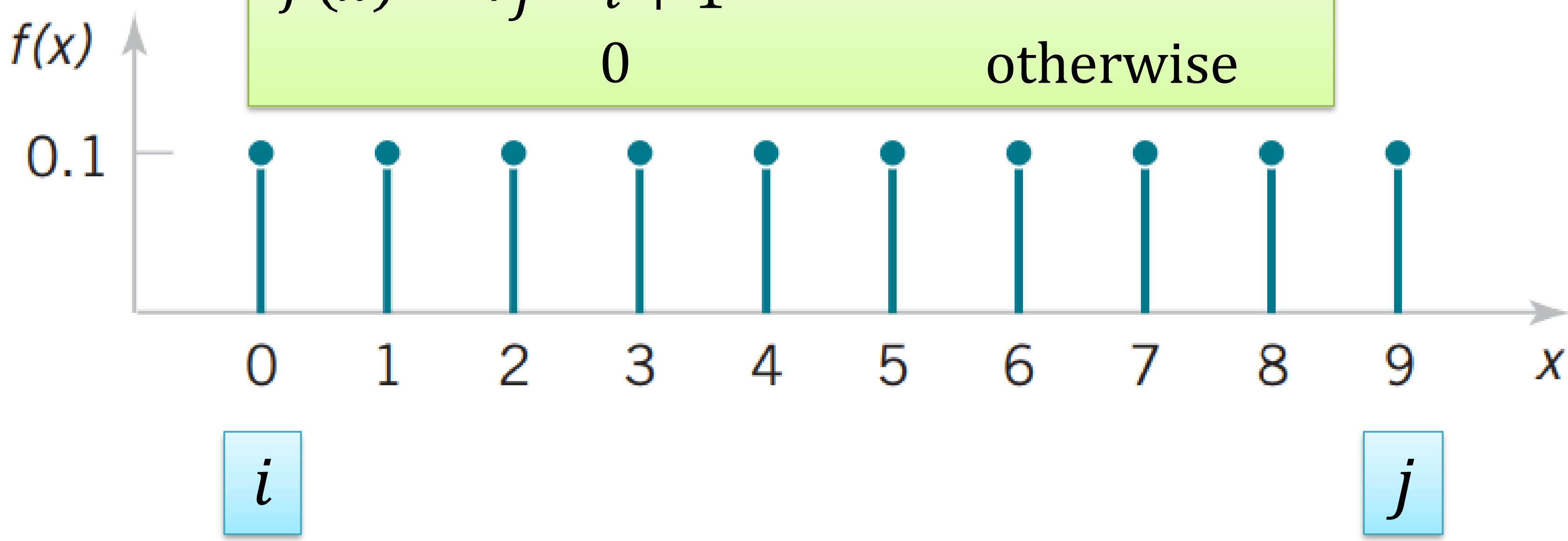


Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} \dfrac{1}{j - i + 1} & \text{if } x \in \{i, i+1, \ldots, j\} \\ 0 & \text{otherwise} \end{cases}$$

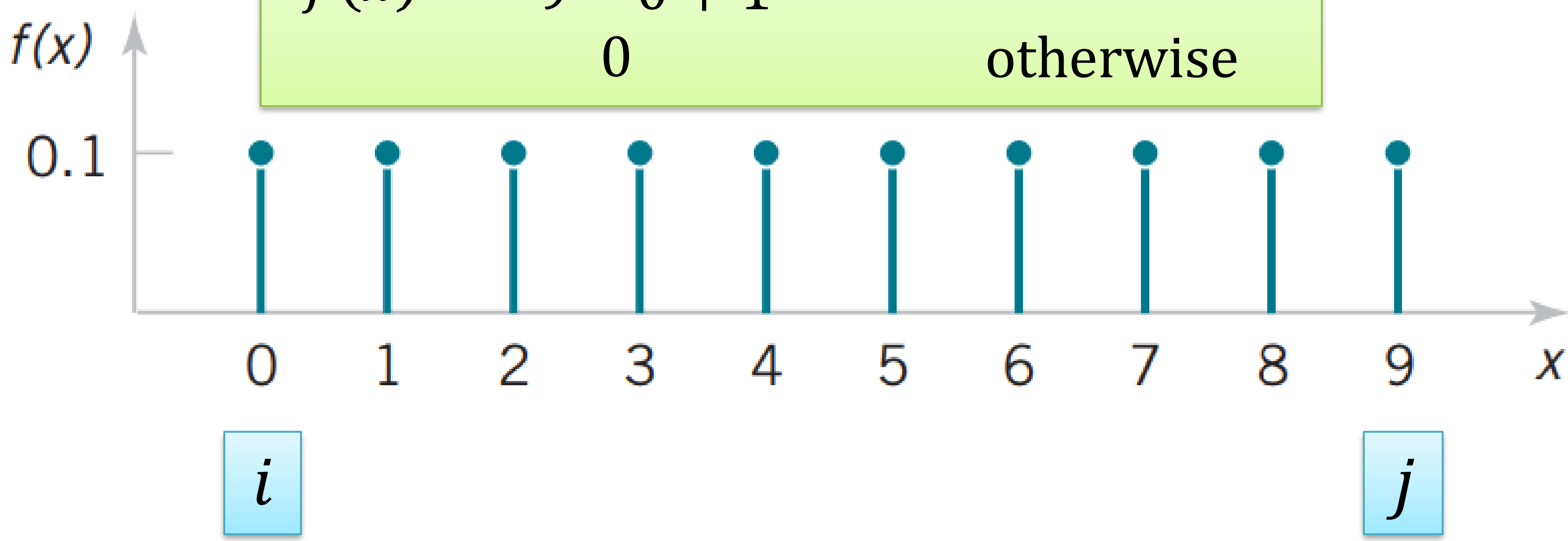


Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} \dfrac{1}{9-0+1} & \text{if } x \in \{0, 1, \dots, 9\} \\ 0 & \text{otherwise} \end{cases}$$

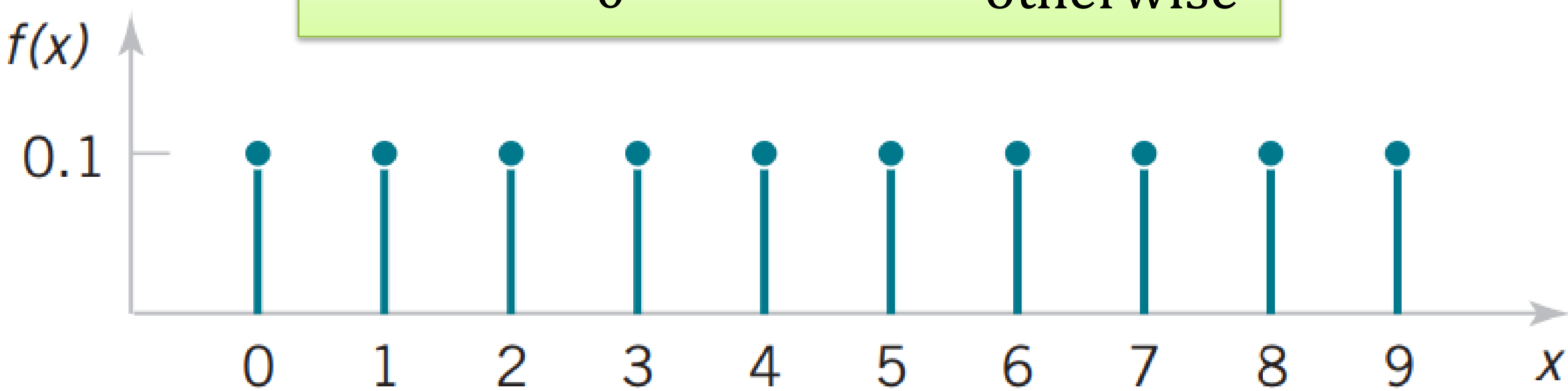


Discrete Uniform Dist. (4/6)

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} 1/10 & \text{if } x \in \{0, 1, \ldots, 9\} \\ 0 & \text{otherwise} \end{cases}$$



i j

Discrete Uniform Dist. (5/6)

Mean and Variance:

Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \ldots, j$, for $i \leq j$

$$\mu = E(X) = \frac{j+i}{2}$$

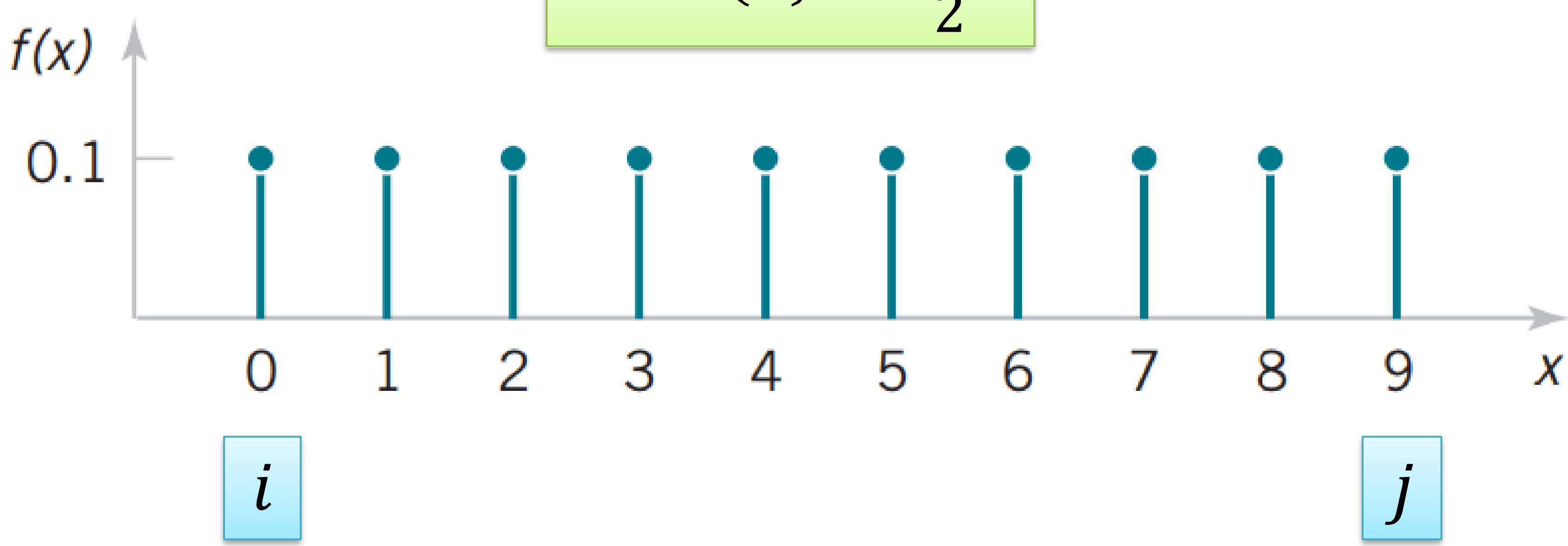
$$\sigma^2 = V(X) = \frac{(j-i+1)^2 - 1}{12}$$

Discrete Uniform Dist. (6/6)

Example2 (1/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\mu = E(X) = \frac{j+i}{2}$$

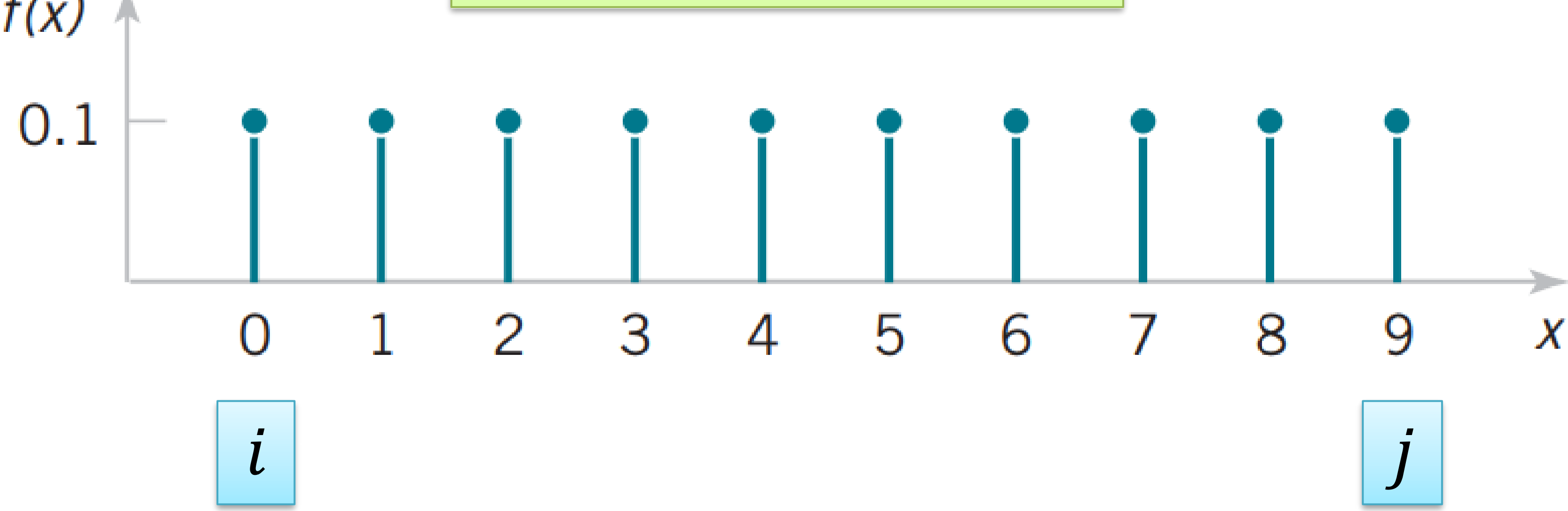


Discrete Uniform Dist. (6/6)

Example2 (2/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\mu = E(X) = \frac{9+0}{2} = 4.5$$

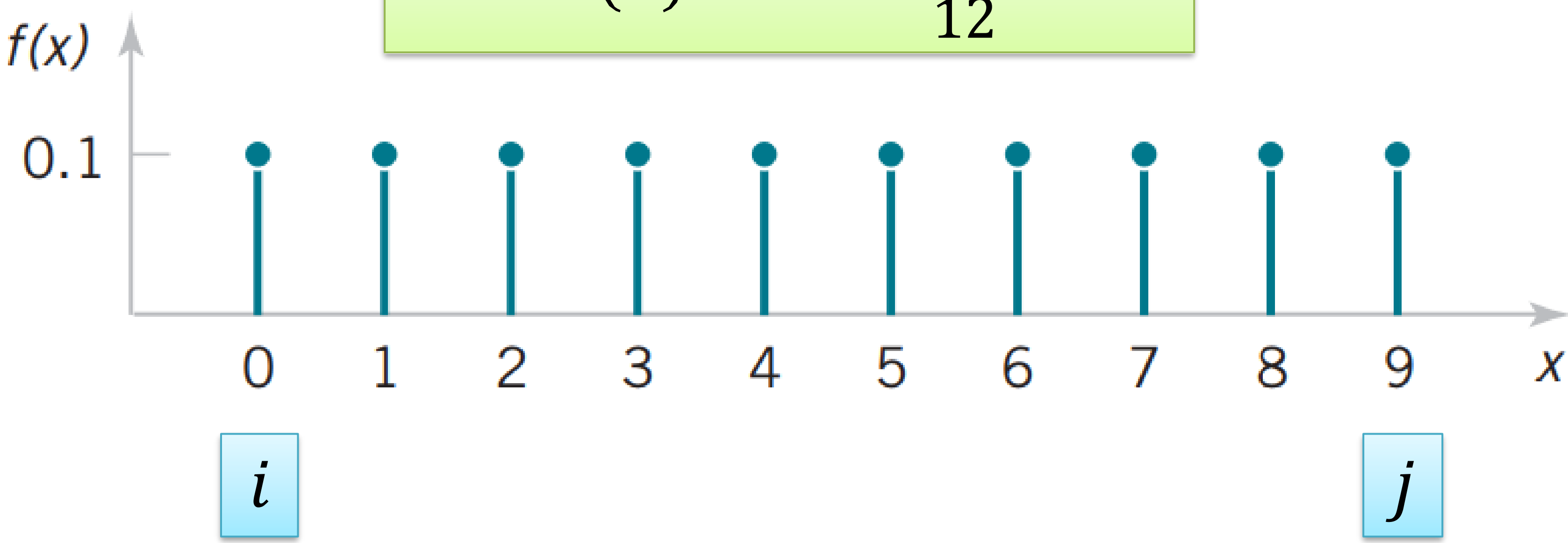


Discrete Uniform Dist. (6/6)

Example2 (3/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\sigma^2 = V(X) = \frac{(j-i+1)^2 - 1}{12}$$

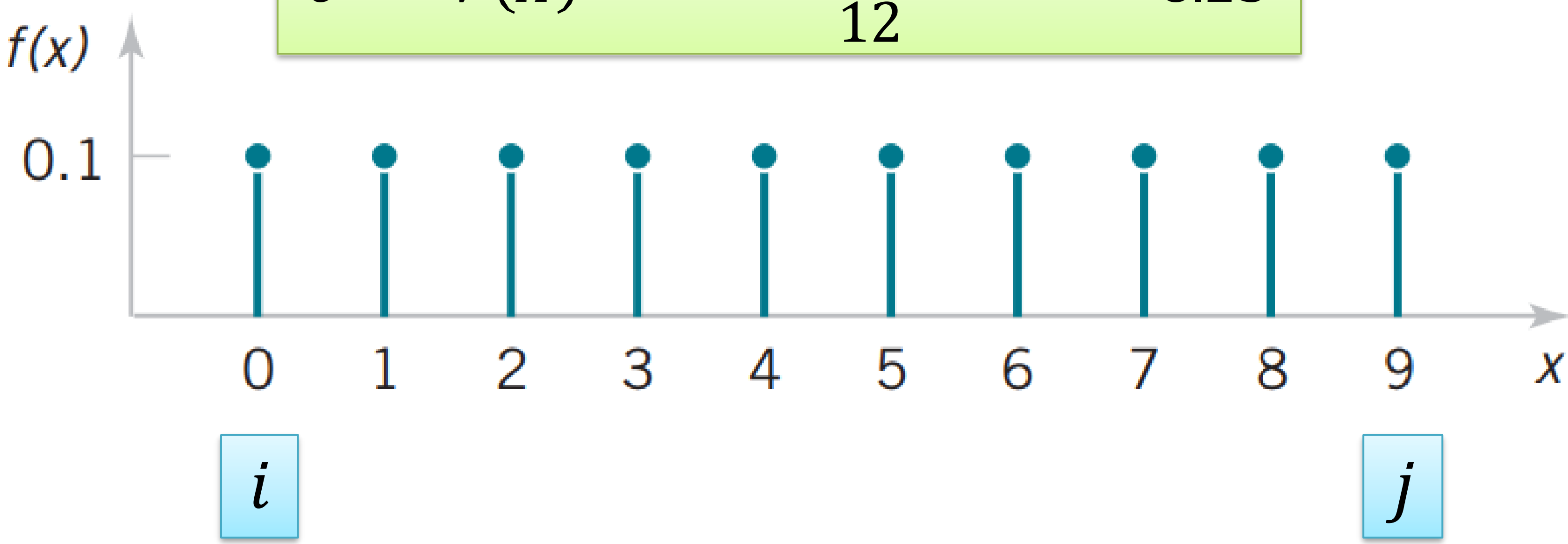


Discrete Uniform Dist. (6/6)

Example2 (4/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\sigma^2 = V(X) = \frac{(9-0+1)^2-1}{12} = 8.25$$



Binomial Distribution (1/11)

The Bernoulli Process:

1. The experiment consists of ***repeated*** trials and each trial is called a ***Bernoulli trial***.
2. Each trial results in an outcome that may be classified as a ***success*** or a ***failure***.
3. The probability of success, denoted by $\boldsymbol{p}$, remains constant from trial to trial.
4. The repeated trials are ***independent***.

Binomial Distribution (2/11)

The Bernoulli Process: (Examples)

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A multiple-choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
3. In the next 20 births at a hospital, let X = the number of female births.

Binomial Distribution (3/11)

Binomial Distribution $b(n, p)$: $X \sim b(n, p)$ $X \sim \text{Bin}(n, p)$

A random experiment consists of n Bernoulli trials. The random variable X that equals the number of trials that result in a success is a *binomial random variable* with parameters $0 < p < 1$ and finite $n = 1, 2, \ldots$.

The probability mass function of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

Binomial Distribution (4/11)

Example 1 (1/5):

Flip a coin 10 times, what is the probability that 3 heads occurs?

Binomial Distribution (4/11)

Example 1 (2/5):

Flip a coin 10 times, what is the probability that 3 heads occurs?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
    - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (4/11)

Example 1 (3/5):

$X \sim \mathrm{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that 3 heads occurs?

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

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (4/11)

Example 1 (4/5):

$X \sim \text{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that 3 heads occurs?

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (4/11)

Example 1 (5/5):

$X \sim \mathrm{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that 3 heads occurs?

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

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
    - Values $x = 0, 1, 2, \ldots, 10$

$$f(3) = \binom{10}{3} (0.5)^3 (0.5)^{10-3}$$

$$= 120\,(0.5)^3 (0.5)^7$$

$$= 0.1171875$$

Binomial Distribution (5/11)

Example 2 (1/5):

Flip a coin 10 times, what is the probability that:

a) No head occurs.

b) At least 2 heads occurs.

Binomial Distribution (5/11)

Example 2 (2/5):

$X \sim \mathrm{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that:

a) No head occurs.

b) At least 2 heads occurs.

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

Binomial Distribution (5/11)

Example 2 (3/5):

$X \sim \mathrm{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that:

a) No head occurs.

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$f(0) = \binom{10}{0} (0.5)^0 (0.5)^{10-0}$$

Binomial Distribution (5/11)

Example 2 (4/5):

$X \sim \text{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that:

b) At least 2 heads occurs.

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

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

Binomial Distribution (5/11)

Example 2 (4/5):

$X \sim \text{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that:

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

b) At least 2 heads occurs.

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$P(X \geq 2)$

Binomial Distribution (5/11)

Example 2 (4/5):

$$X \sim \mathrm{Bin}(10, 0.5)$$

Flip a coin 10 times, what is the probability that:

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

b) At least 2 heads occurs.

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$P(X \geq 2) = f(2) + f(3) + \cdots + f(10)$$

Binomial Distribution (5/11)

Example 2 (4/5):

$X \sim \text{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that:

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

b) At least 2 heads occurs.

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$P(X \geq 2) = \cancel{f(2) + f(3) + \cdots + f(10)}$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [f(1) + f(0)]$$

Binomial Distribution (5/11)

Example 2 (5/5):

$X \sim \text{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that:

b) At least 2 heads occurs.

$$f(0) = \binom{10}{0}(0.5)^0(0.5)^{10-0}$$

$$f(1) = \binom{10}{1}(0.5)^1(0.5)^{10-1}$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [f(1) + f(0)]$$

Binomial Distribution (6/11)

Mean and Variance $b(n, p)$:

If X is a binomial random variable with parameters p and n

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

$$\sigma^2 = V(X) = np(1-p)$$

Example 3 (1/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

Example 3 (2/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (7/11)

Example 3 (3/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \dots, 10$

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

$$\sigma^2 = V(X) = np(1-p)$$

Binomial Distribution (7/11)

Example 3 (4/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \dots, 10$

$$\mu = E(X) = np = 10 \times 0.5 = 5$$

$$\sigma^2 = V(X) = np(1-p)$$

Binomial Distribution (7/11)

Example 3 (5/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

$$\mu = E(X) = np = 10 \times 0.5 = 5$$

$$\sigma^2 = V(X) = np(1-p) = 10 \times 0.5 \times 0.5 = 2.5$$

Example 4 (1/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

Binomial Distribution (8/11)

Example 4 (2/5):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
    - Values $x = 0, 1, 2, \ldots, 15$

Binomial Distribution (8/11)

Example 4 (3/5):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
    - Values $x = 0, 1, 2, \dots, 15$

Binomial Distribution (8/11)

Example 4 (3/5):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

$$f(x) = \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \ldots, 15$

Binomial Distribution (8/11)

Example 4 (4/5):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

$$f(x) = \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

(a) exactly 5 survive

$$f(5) = \binom{15}{5} (0.4)^5 (0.6)^{15-5}$$

$$= 0.1859$$

Binomial Distribution (8/11)

Example 4 (5/5):

$X \sim \mathrm{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

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

(b) at least 3 survive

$$f(x) = \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

$$P(X \geq 3) = 1 - P(X < 3)$$
$$= 1 - [f(2) + f(1) + f(0)]$$

Binomial Distribution (9/11)

Example 5 (1/4):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

Binomial Distribution (9/11)

Example 5 (2/4):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

$\mu = E(X) = np$

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \dots, 15$

$\sigma^2 = V(X) = np(1 - p)$

Binomial Distribution (9/11)

Example 5 (3/4):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \ldots, 15$

$$\mu = E(X) = np$$
$$= 15 \times 0.4 = 6$$

$$\sigma^2 = V(X) = np(1 - p)$$

Binomial Distribution (9/11)

Example 5 (4/4):

$X \sim \text{Bin}(15, 0.4)$

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
    - Values $x = 0, 1, 2, \dots, 15$

$$\mu = E(X) = np$$
$$= 15 \times 0.4 = 6$$

$$\sigma^2 = V(X) = np(1 - p)$$
$$= 15 \times 0.4 \times 0.6$$
$$= 3.6$$

Binomial Distribution (10/11)

Example 6 (1/4):

The probability that a patient recovers from a COVID-19 virus is 0.8. What is the probability that at least 3 of the next 10 patients having this virus will survive?

Binomial Distribution (10/11)

Example 6 (2/4):

$X \sim \text{Bin}(10, 0.8)$

The probability that a patient recovers from a COVID-19 virus is 0.8. What is the probability that at least 3 of the next 10 patients having this virus will survive?

- Bernoulli trial → Patient
- Number of trials ($n = 10$)
- Success → Patient recover ($p = 0.8$)
- Let X = the number of people who survive
    - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (10/11)

Example 6 (2/4):

$X \sim \mathrm{Bin}(10, 0.8)$

The probability that a patient recovers from a COVID-19 virus is 0.8. What is the probability that at least 3 of the next 10 patients having this virus will survive?

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

$$f(x) = \binom{10}{x} (0.8)^x (0.2)^{10-x}$$

- Bernoulli trial → Patient
- Number of trials ($n = 10$)
- Success → Patient recover ($p = 0.8$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \ldots, 10$

Binomial Distribution (10/11)

Example 6 (3/4):

$X \sim \text{Bin}(10, 0.8)$

The probability that a patient recovers from a COVID-19 virus is 0.8. What is the probability that **at least 3** of the next 10 patients having this virus will survive?

$$P(X \geq 3) = 1 - P(X \leq 2)$$
$$= 1 - [f(2) + f(1) + f(0)]$$

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

$$f(x) = \binom{10}{x} (0.8)^x (0.2)^{10-x}$$

Binomial Distribution (10/11)

Example 6 (4/4):

$X \sim \mathrm{Bin}(10, 0.8)$

The probability that a patient recovers from a COVID-19 virus is 0.8. What is the probability that **at least 3** of the next 10 patients having this virus will survive?

$$P(X \geq 3) = 1 - P(X \leq 2)$$

You can find $P(X \leq 2)$ from the **Table**. $P(X \leq 2) = 0.0001$

$$\therefore P(X \geq 3) = 1 - 0.0001$$
$$= 0.9999$$

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

$$f(x) = \binom{10}{x} (0.8)^x (0.2)^{10-x}$$

Example 7 (1/4):

A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that fewer than 4 of the next 9 vehicles are from out of state?

Example 7 (2/4):

A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that fewer than 4 of the next 9 vehicles are from out of state?

- Bernoulli trial → Vehicle
- Number of trials ($n = 9$)
- Success → vehicle is from out of state ($p = 0.25$)
- Let X = the number of vehicles are from out of state
  - Values $x = 0, 1, 2, \dots, 9$

Binomial Distribution (11/11)

Example 7 (3/4):

$X \sim \text{Bin}(9, 0.25)$

A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that fewer than 4 of the next 9 vehicles are from out of state?

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

$$f(x) = \binom{9}{x} (0.25)^x (0.75)^{9-x}$$

- Bernoulli trial → Vehicle
- Number of trials ($n = 9$)
- Success → vehicle is from out of state ($p = 0.25$)
- Let X = the number of vehicles are from out of state
  - Values $x = 0, 1, 2, \ldots, 9$

Binomial Distribution (11/11)

Example 7 (4/4):

$X \sim \mathrm{Bin}(9, 0.25)$

A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that **fewer than 4** of the next 9 vehicles are from out of state?

$P(X < 4) = P(X \leq 3)$

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

$$f(x) = \binom{9}{x} (0.25)^x (0.75)^{9-x}$$

$P(X \leq 3)$, $n = 9$, $p = 0.25$

You can find $P(X \leq 3)$ from the **Table**.

$$P(X \leq 3) = 0.8343$$

Multinomial Distribution (1/5)

Multinomial Experiments:

The binomial experiment becomes a multinomial experiment if we let each trial have **more than two** possible outcomes.

Multinomial Distribution (2/5)

Multinomial Distribution:

If a given trial can result in the k outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in n independent trials, is

$$f(x_1, x_2, \ldots x_k) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

with

$$\sum_{i=1}^{k} x_i = n \quad \text{and} \quad \sum_{i=1}^{k} P_i = 1$$

Multinomial Distribution:

If a given trial can result in the k outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in n independent trials, is

$$f(x_1, x_2, \ldots x_k) = \boxed{\binom{n}{x_1, x_2, \ldots, x_k}} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

$$\boxed{= \frac{n!}{x_1!\, x_2! \cdots x_k!}}$$

Multinomial Distribution – Example 1 (1/5):

The complexity of arrivals and departures of planes at an airport is such that computer simulation is often used to model the "ideal" conditions. For a certain airport with three runways, it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet:

Runway1: $p_1 = 2/9$,
Runway2: $p_2 = 1/6$,
Runway3: $p_3 = 11/18$.

Multinomial Distribution (4/5)

Multinomial Distribution – Example 1 (2/5):

The individual runways are accessed by a randomly arriving commercial jet:

Runway1: $p_1 = 2/9$,
Runway2: $p_2 = 1/6$,
Runway3: $p_3 = 11/18$.

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

Runway1: 2 airplanes,
Runway2: 1 airplane,
Runway3: 3 airplanes.

Multinomial Distribution (4/5)

Multinomial Distribution – Example 1 (3/5):

The individual runways are accessed by a randomly arriving commercial jet:

Runway1: $p_1 = 2/9$,
Runway2: $p_2 = 1/6$,
Runway3: $p_3 = 11/18$.

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

Runway1: 2 airplanes, $x_1 = 2$
Runway2: 1 airplane, $x_2 = 1$
Runway3: 3 airplanes. $x_3 = 3$

Multinomial Distribution (4/5)

Multinomial Distribution – Example 1 (4/5):

The individual runways are accessed by a randomly arriving commercial jet:

$$f(x_1, x_2, x_3) = \binom{n}{x_1, x_2, x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

Runway1: $p_1 = 2/9$,

Runway2: $p_2 = 1/6$,

Runway3: $p_3 = 11/18$.

$$f(2,1,3) = \binom{6}{2,1,3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3$$

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

Runway1: 2 airplanes,

Runway2: 1 airplane,

Runway3: 3 airplanes.

$x_1 = 2$

$x_2 = 1$

$x_3 = 3$

Multinomial Distribution (4/5)

Multinomial Distribution – Example 1 (5/5):

$$f(2,1,3) = \binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3$$

$$= \frac{6!}{2!\,1!\,3!}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3 = 0.1127$$

Multinomial Distribution (5/5)

Multinomial Distribution – Example 2 (1/4):

The probabilities are 0.4, 0.2, 0.3, and 0.1, respectively, that a delegate to a certain convention arrived by air, bus, automobile, or train. What is the probability that among 9 delegates randomly selected at this convention, 3 arrived by air, 3 arrived by bus, 1 arrived by automobile, and 2 arrived by train.

Multinomial Distribution (5/5)

Multinomial Distribution – Example 2 (2/4):

The probabilities are 0.4, 0.2, 0.3, and 0.1, respectively, that a delegate to a certain convention arrived by air, bus, automobile, or train. What is the probability that among 9 delegates randomly selected at this convention, 3 arrived by air, 3 arrived by bus, 1 arrived by automobile, and 2 arrived by train.

$$p_1 = 0.4$$
$$p_2 = 0.2$$
$$p_3 = 0.3$$
$$p_4 = 0.1$$

$$x_1 = 3$$
$$x_2 = 3$$
$$x_3 = 1$$
$$x_4 = 2$$

Multinomial Distribution (5/5)

Multinomial Distribution – Example 2 (3/4):

$$f(x_1, x_2, x_3, x_4) = \binom{n}{x_1, x_2, x_3, x_4} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

$$f(3,3,1,2) = \binom{9}{3,3,1,2} (0.4)^3 (0.2)^3 (0.3)^1 (0.1)^2$$

$$p_1 = 0.4$$
$$p_2 = 0.2$$
$$p_3 = 0.3$$
$$p_4 = 0.1$$

$$x_1 = 3$$
$$x_2 = 3$$
$$x_3 = 1$$
$$x_4 = 2$$

Multinomial Distribution (5/5)

Multinomial Distribution – Example 2 (4/4):

$$f(3,3,1,2) = \binom{9}{3,3,1,2}(0.4)^3(0.2)^3(0.3)^1(0.1)^2$$

$$= \frac{9!}{3!\,3!\,1!\,2!}(0.4)^3(0.2)^3(0.3)^1(0.1)^2 = 0.0077$$

Negative Binomial (1/3)

Negative Binomial Distribution:

$X \sim nb(k, p)$

If repeated independent trials can result in a success with probability p, then the probability distribution of the random variable X, the number of the trial on which the kth success occurs, is

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$$

with

$$k = 1,2,\ldots \quad \text{and} \quad x = k, k+1, k+2, \ldots$$

Negative Binomial (2/3)

Negative Binomial Distribution – Example (1/8):

In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B.

What is the probability that team A will win the series in 6 games?

Negative Binomial (2/3)

Negative Binomial Distribution – Example (2/8):

In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that **team A has probability 0.55 of winning a game over team B**.

What is the probability that team A will win the series in 6 games?

$p = 0.55$

Negative Binomial Distribution – Example (3/8):

In an NBA (National Basketball Association) championship series, the team that wins **four games** out of seven is the winner. Suppose that teams A and B face each other in the championship games and that **team A has probability 0.55 of winning a game over team B**.

What is the probability that team A will win the series in **6 games**?

$p = 0.55, \quad x = 6, \quad k = 4$

Negative Binomial (2/3)

Negative Binomial Distribution – Example (4/8):

What is the probability that team A will win the series in **6 games**?

$$p = 0.55, \quad x = 6, \quad k = 4$$

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$$

$$f(6) = \binom{5}{3} (0.55)^4 (1 - 0.55)^{6-4} = 0.1853$$

Negative Binomial (2/3)

Negative Binomial Distribution – Example (5/8):

What is the probability that team A will win the series?

Negative Binomial (2/3)

Negative Binomial Distribution – Example (6/8):

What is the probability that team A will win the series?

$p = 0.55,$

1) $x = 4, \quad k = 4$
2) $x = 5, \quad k = 4$
3) $x = 6, \quad k = 4$
4) $x = 7, \quad k = 4$

Negative Binomial (2/3)

Negative Binomial Distribution – Example (7/8):

What is the probability that team A will win the series?

$p = 0.55,$

1) $x = 4, \quad k = 4$
2) $x = 5, \quad k = 4$
3) $x = 6, \quad k = 4$
4) $x = 7, \quad k = 4$

$f(4) + f(5) + f(6) + f(7)$

Negative Binomial (2/3)

Negative Binomial Distribution – Example (8/8):

What is the probability that team A will win the series?

$p = 0.55,$

1) $x = 4, \quad k = 4$
2) $x = 5, \quad k = 4$
3) $x = 6, \quad k = 4$
4) $x = 7, \quad k = 4$

$f(4) + f(5) + f(6) + f(7)$
$= 0.6083$

$$f(4) = \binom{3}{3}(0.55)^4(1-0.55)^{4-4} = 0.0915$$

$$f(5) = \binom{4}{3}(0.55)^4(1-0.55)^{5-4} = 0.1647$$

$$f(6) = \binom{5}{3}(0.55)^4(1-0.55)^{6-4} = 0.1853$$

$$f(7) = \binom{6}{3}(0.55)^4(1-0.55)^{7-4} = 0.1668$$

Negative Binomial (3/3)

Negative Binomial Distribution: (Mean and Variance)

If X is a negative binomial random variable with parameters p and k,

$$\mu = E(X) = k/p \text{ and}$$

$$\sigma^2 = V(X) = k(1-p)/p^2$$

Geometric Distribution (1/3)

Geometric Distribution:

$X \sim g(p)$

If repeated independent trials can result in a success with probability p, then the probability distribution of the random variable X, the number of the trial on which the **first** success occurs, is

$$f(x) = p(1-p)^{x-1}$$

with

$$x = 1, 2, 3, \ldots$$

Geometric Distribution (2/3)

Geometric Distribution – Example (1/5):

For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Geometric Distribution (2/3)

Geometric Distribution – Example (2/5):

For a certain manufacturing process, it is known that, on the average, **1 in every 100** items is defective. What is the probability that the fifth item inspected is the first defective item found?

$$p = 0.01$$

Geometric Distribution (2/3)

Geometric Distribution – Example (3/5):

For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the **fifth item inspected** is the **first** defective item found?

$p = 0.01, \quad x = 5, \quad k = 1$

Geometric Distribution – Example (4/5):

For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

$p = 0.01, \quad x = 5, \quad k = 1$

$f(x) = p(1-p)^{x-1}$

Geometric Distribution (2/3)

Geometric Distribution – Example (5/5):

For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

$$p = 0.01, \quad x = 5, \quad k = 1$$

$$f(x) = p(1-p)^{x-1}$$

$$f(5) = (0.01)(1-0.01)^{5-1} = 0.0096$$

Geometric Distribution (3/3)

Geometric Distribution: (Mean and Variance)

If X is a geometric random variable with parameter p,

$$\mu = E(X) = 1/p \text{ and}$$

$$\sigma^2 = V(X) = (1 - p)/p^2$$

Poisson Distribution (1/7)

Poisson Distribution:

$X \sim p(\lambda t)$ | $X \sim \text{Poisson}(\lambda t)$

The mean number of outcomes is computed from $\mu = \lambda t$, where t is the specific "time," "distance," "area," or "volume" of interest. Since the probabilities depend on λ, the rate of occurrence of outcomes.

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

with

$$\lambda > 0 \text{ and } \quad x = 0, 1, 2, \ldots$$

Poisson Distribution (2/7)

Poisson Distribution – Example 1 (1/4):

During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

Poisson Distribution (2/7)

Poisson Distribution – Example 1 (2/4):

During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

$$\lambda t = 4$$
$$x = 6$$

Poisson Distribution (2/7)

Poisson Distribution – Example 1 (3/4):

During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

$$\lambda t = 4$$
$$x = 6$$

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

Poisson Distribution (2/7)

Poisson Distribution – Example 1 (4/4):

During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

$$\lambda t = 4$$
$$x = 6$$

$$f(6) = \frac{e^{-4}(4)^6}{6!} = 0.1042$$

Poisson Distribution (3/7)

Poisson Distribution:

$X \sim p(\lambda)$

$X \sim \text{Poisson}(\lambda)$

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

with

$$\lambda > 0 \text{ and } \quad x = 0, 1, 2, \ldots$$

Poisson Distribution (4/7)

Poisson Distribution – Example 2 (1/6):

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

Poisson Distribution (4/7)

Poisson Distribution – Example 2 (2/6):

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$$\lambda = 10,\ x > 15$$

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

Poisson Distribution (4/7)

Poisson Distribution – Example 2 (3/6):

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$$\lambda = 10,\ x > 15$$

$$f(X > 15) = 1 - f(X \leq 15)$$

Poisson Distribution (4/7)

Poisson Distribution – Example 2 (4/6):

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$$\lambda = 10,\ x > 15$$

$$f(X > 15) = 1 - \boxed{f(X \leq 15)}$$

$$\boxed{\begin{array}{c} f(X = 0) \\ f(X = 1) \\ f(X = 2) \\ \ldots \\ f(X = 15) \end{array}}$$

Poisson Distribution (4/7)

Poisson Distribution – Example 2 (5/6):

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$$\lambda = 10,\ x > 15$$

$$f(X > 15) = 1 - \boxed{f(X \leq 15)}$$

From the Table $= 0.9513$

Poisson Distribution (4/7)

Poisson Distribution – Example 2 (6/6):

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$$\lambda = 10, \; x > 15$$

$$f(X > 15) = 1 - 0.9513 = 0.0487$$

Poisson Distribution (5/7)

Poisson Distribution: (Mean and Variance)

Both the mean and the variance of the Poisson distribution are

$$E(X) = V(X) = \lambda t$$

Poisson Distribution (6/7)

Poisson Distribution: (From Binomial)

Let X be a binomial random variable with probability distribution $b(n, p)$.

When $n \to \infty$, $p \to 0$,

and $np \xrightarrow{n\to\infty} \mu$ remains constant,

$b(n, p) \xrightarrow{n\to\infty} p(\mu)$.

Poisson Distribution (7/7)

Poisson Distribution: (From Binomial): Example (1/6)

In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

Poisson Distribution (7/7)

Poisson Distribution: (From Binomial): Example (2/6)

In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, **1 in every 1000** of these items produced has one or more bubbles. What is the probability that a random **sample of 8000** will yield **fewer than 7** items possessing bubbles?

$p = 0.001,$ and $n = 8000$

$f(x < 7)$ from Binomial

Poisson Distribution (7/7)

Poisson Distribution: (From Binomial): Example (3/6)

In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, **1 in every 1000** of these items produced has one or more bubbles. What is the probability that a random **sample of 8000** will yield **fewer than 7** items possessing bubbles?

$p = 0.001$, and $n = 8000$

$f(x < 7) = f(x \leq 6)$ from Binomial

Poisson Distribution (7/7)

Poisson Distribution: (From Binomial): Example (4/6)

In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

$p = 0.001,$ and $n = 8000$

$$f(x < 7) = f(x \leq 6) = \sum_{x=0}^{6} Bin(x; n, p)$$

Poisson Distribution (7/7)

Poisson Distribution: (From Binomial): Example (5/6)

$p = 0.001$, and $n = 8000$

$$f(x < 7) = f(x \leq 6) = \sum_{x=0}^{6} Bin(x; 8000, 0.001)$$

Since p is very close to 0 and n is quite large, we shall approximate with the Poisson distribution using

$$\mu = np = 8$$

$$\sum_{x=0}^{6} p(x; 8) = 0.3134$$

From the Table

Poisson Distribution (7/7)

Poisson Distribution: (From Binomial): Example (6/6)

$p = 0.001$, and $n = 8000$

$$f(x < 7) = f(x \leq 6) = \sum_{x=0}^{6} Bin(x; 8000, 0.001) \approx 0.3134$$

Since p is very close to 0 and n is quite large, we shall approximate with the Poisson distribution using

$$\mu = np = 8$$

$$\sum_{x=0}^{6} p(x; 8) = 0.3134$$

Ch 3.2: Continuous Prob. Distr.

- Continuous Uniform Distribution.
- Normal Distribution.

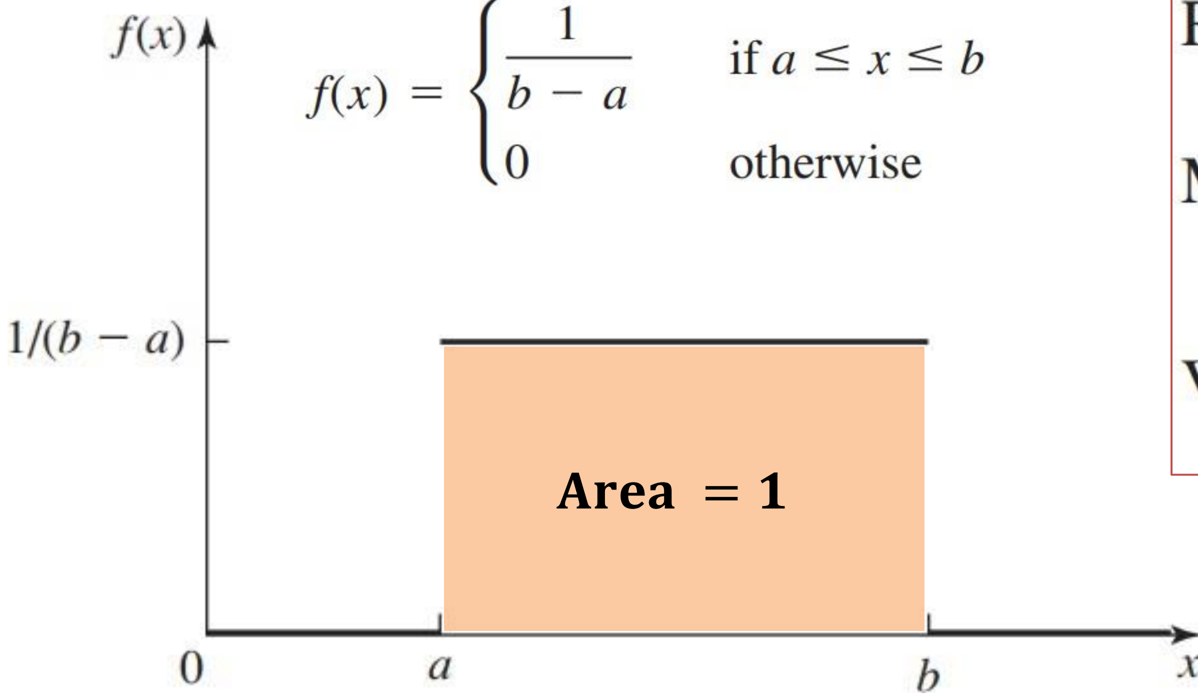
Continuous Uniform Dist. (1/4)

Probability Density Function:

$U(a, b)$

- a and b real numbers with $a < b$.

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



Range	$[a, b]$
Mean	$\dfrac{a+b}{2}$
Variance	$\dfrac{(b-a)^2}{12}$

Continuous Uniform Dist. (2/4)

Cumulative Distribution Function:

$\boldsymbol{U(a, b)}$

Density $\quad f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

$$F(x) = \int_a^x \frac{1}{b-a} du = \frac{x-a}{b-a}$$

cumulative distribution function
$P(X \leq x)$

Continuous Uniform Dist. (2/4)

Cumulative Distribution Function:

$\boldsymbol{U(a,b)}$

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

$$\text{Distribution} \qquad F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$$

Continuous Uniform Dist. (3/4)

Example1 (1/6):

Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform distribution on the interval $[0, 4]$.

(a) What is the probability density function?

(b) What is the probability that any given conference lasts at least 3 hours?

Continuous Uniform Dist. (3/4)

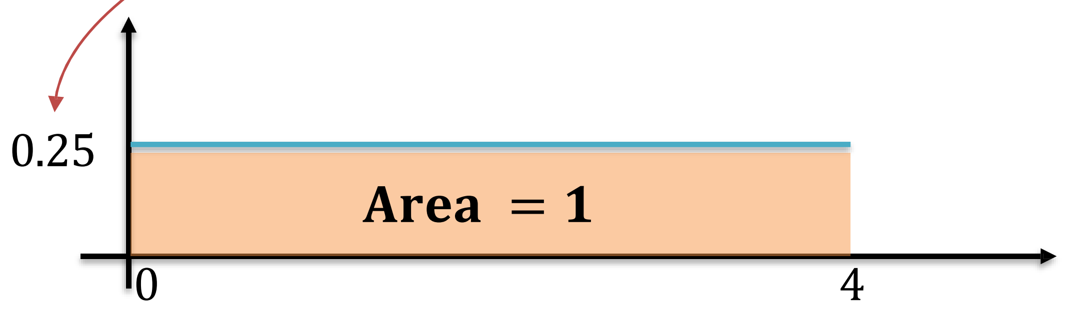
Example1 (2/6):

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

$\boldsymbol{X \sim U(0, 4)}$

a) What is the probability density function?

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \le x \le 4, \\ 0, & \text{elsewhere.} \end{cases}$$



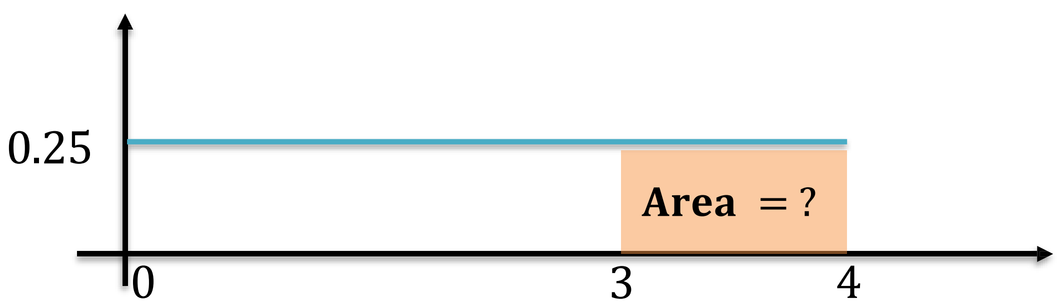
Continuous Uniform Dist. (3/4)

Example1 (3/6):

$\boldsymbol{X \sim U(0, 4)}$

b) What is the probability that any given conference lasts at least 3 hours?

$$P(X \geq 3) = \int_{3}^{4} \frac{1}{4}\, dx$$



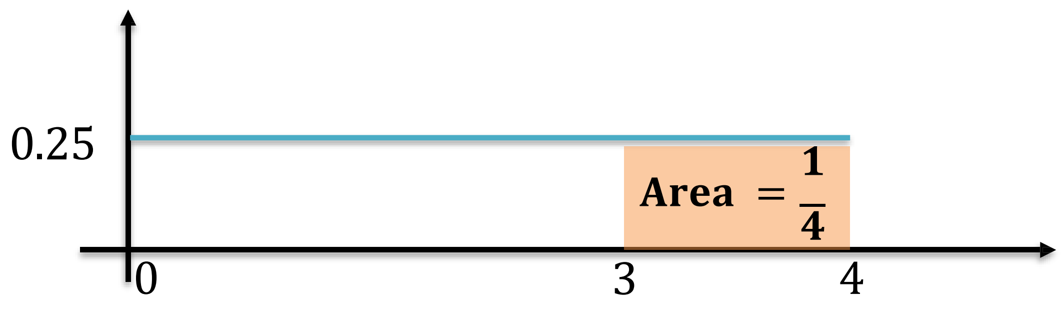
Continuous Uniform Dist. (3/4)

Example1 (4/6):

$\boldsymbol{X \sim U(0, 4)}$

b) What is the probability that any given conference lasts at least 3 hours?

$$P(X \geq 3) = \int_{3}^{4} \frac{1}{4}\, dx = \frac{1}{4} x \Big|_{x=3}^{x=4} = (1) - \left(\frac{3}{4}\right) = \frac{1}{4}$$



Continuous Uniform Dist. (3/4)

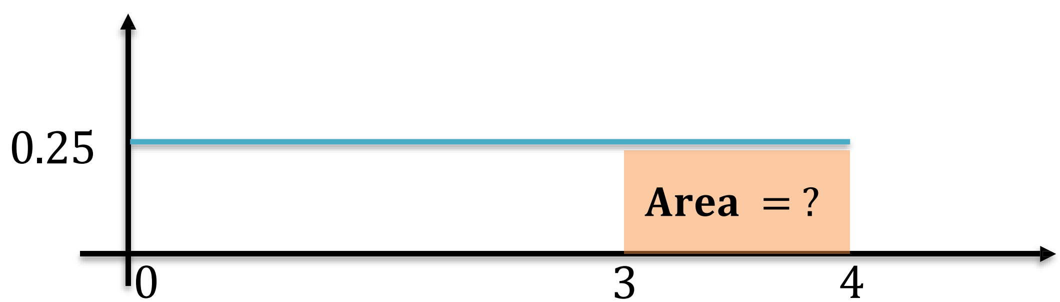
Example1 (5/6):

$\boldsymbol{X \sim U(0, 4)}$

Distribution $\quad F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$

b) What is the probability that any given conference lasts at least 3 hours?

$$P(X \geq 3) = 1 - P(X < 3) = 1 - F(3)$$



Continuous Uniform Dist. (3/4)

Example1 (6/6):

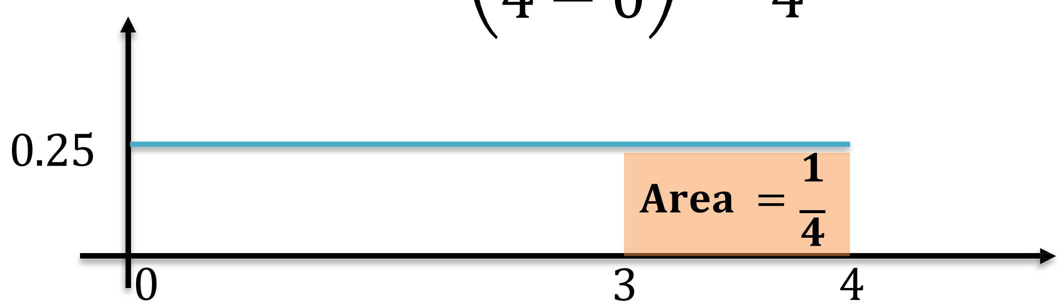
$\boldsymbol{X \sim U(0, 4)}$

Distribution $F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$

b) What is the probability that any given conference lasts at least 3 hours?

$$P(X \geq 3) = 1 - P(X < 3) = 1 - F(3)$$

$$= 1 - \left(\frac{3-0}{4-0}\right) = \frac{1}{4}$$



Example2 (1/10):

A bus arrives every 10 minutes at a bus stop. It is assumed that the waiting time for a particular individual is a random variable with a continuous uniform distribution.

(a) What is the probability that the individual waits at most 4 minutes?

(b) What is the probability that the individual waits more than 6 minutes (at least 6 minutes)?

(c) What is the probability that the individual waits between 1 and 4 minutes?

(d) What is the average waiting time?

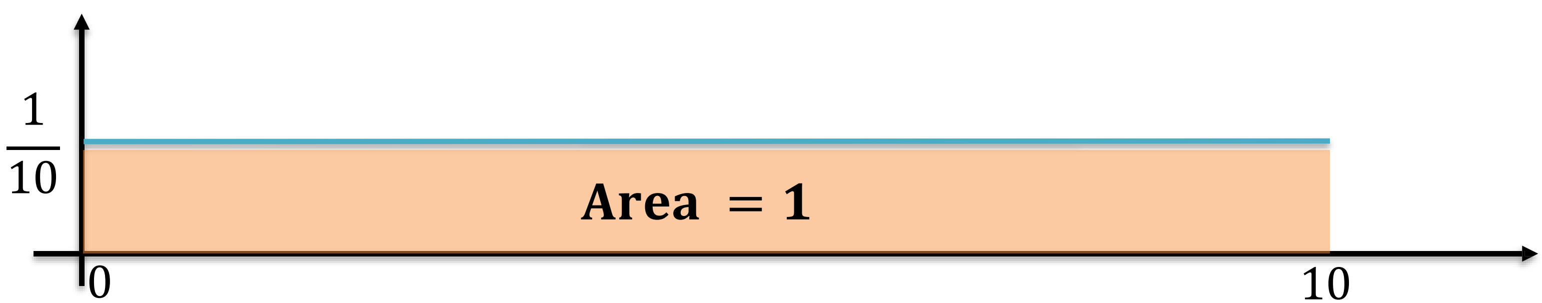
Continuous Uniform Dist. (4/4)

Example2 (2/10):

$$X \sim U(0, 10)$$

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

$$f(x) = \begin{cases} \dfrac{1}{10} & \text{if } 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$



Continuous Uniform Dist. (4/4)

Example2 (3/10):

$\boldsymbol{X \sim U(0, 10)}$

$$\text{Distribution} \qquad F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$$

a) What is the probability that the individual waits at most 4 minutes?

$$P(X \le 4) = F(4)$$

Continuous Uniform Dist. (4/4)

Example2 (4/10):

$\boldsymbol{X \sim U(0, 10)}$

Distribution $F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$

a) What is the probability that the individual waits at most 4 minutes?

$$P(X \le 4) = F(4) = \frac{4-0}{10-0} = \frac{4}{10} = 0.4$$

Continuous Uniform Dist. (4/4)

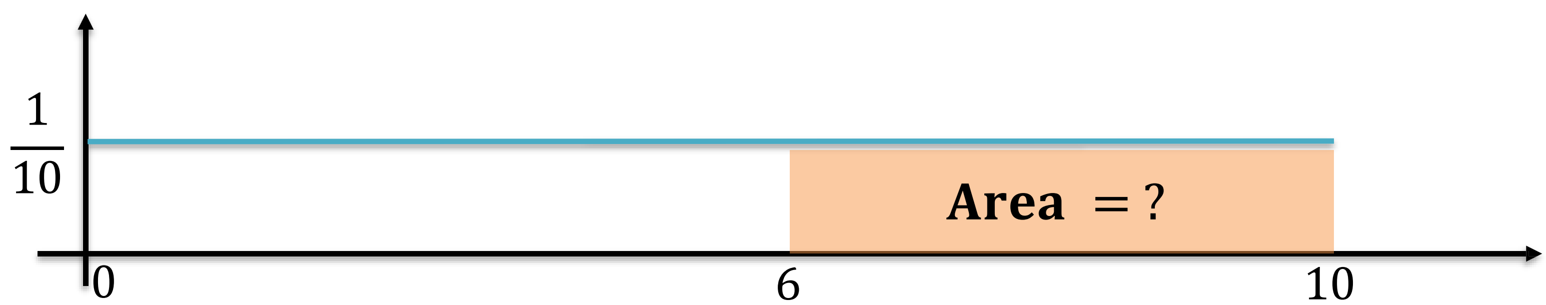
Example2 (5/10):

$\boldsymbol{X \sim U(0, 10)}$

Distribution $\quad F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$

b) What is the probability that the individual waits more than 6 minutes (at least 6 minutes)?

$$P(X \geq 6) = 1 - F(6)$$



Continuous Uniform Dist. (4/4)

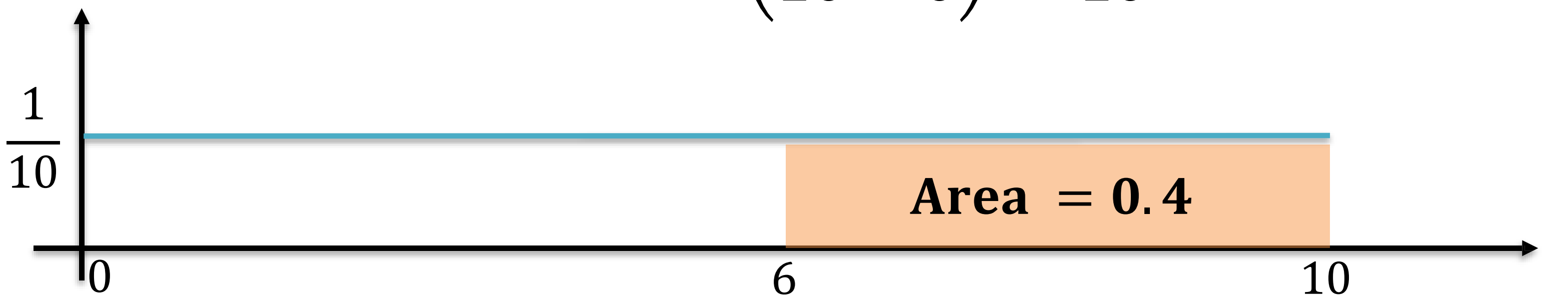
Example2 (6/10):

$\boldsymbol{X \sim U(0, 10)}$

Distribution $\quad F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$

b) What is the probability that the individual waits more than 6 minutes (at least 6 minutes)?

$$P(X \geq 6) = 1 - \left(\frac{6-0}{10-0}\right) = \frac{4}{10} = 0.4$$



Continuous Uniform Dist. (4/4)

Example2 (7/10):

$\boldsymbol{X \sim U(0, 10)}$

c) What is the probability that the individual waits between 1 and 4 minutes?

$$P(1 \leq X \leq 4)$$

Continuous Uniform Dist. (4/4)

Example2 (8/10):

$\boldsymbol{X \sim U(0, 10)}$

c) What is the probability that the individual waits between 1 and 4 minutes?

$$P(1 \leq X \leq 4) = \int_{1}^{4} \frac{1}{10}\, dx = \frac{1}{10} x \Big|_{x=1}^{x=4} = \left(\frac{4}{10}\right) - \left(\frac{1}{10}\right) = \frac{3}{10}$$

Continuous Uniform Dist. (4/4)

Example2 (9/10):

$\boldsymbol{X \sim U(0, 10)}$

Distribution	$F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$

c) What is the probability that the individual waits between 1 and 4 minutes?

$$P(1 \le X \le 4) = F(4) - F(1) = \left(\frac{4-0}{10-0}\right) - \left(\frac{1-0}{10-0}\right) = \frac{3}{10}$$

Continuous Uniform Dist. (4/4)

Example2 (10/10):

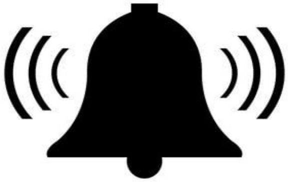
$$\boldsymbol{X \sim U(0, 10)}$$

$$\boxed{\mu = E(X) = \frac{a+b}{2}}$$

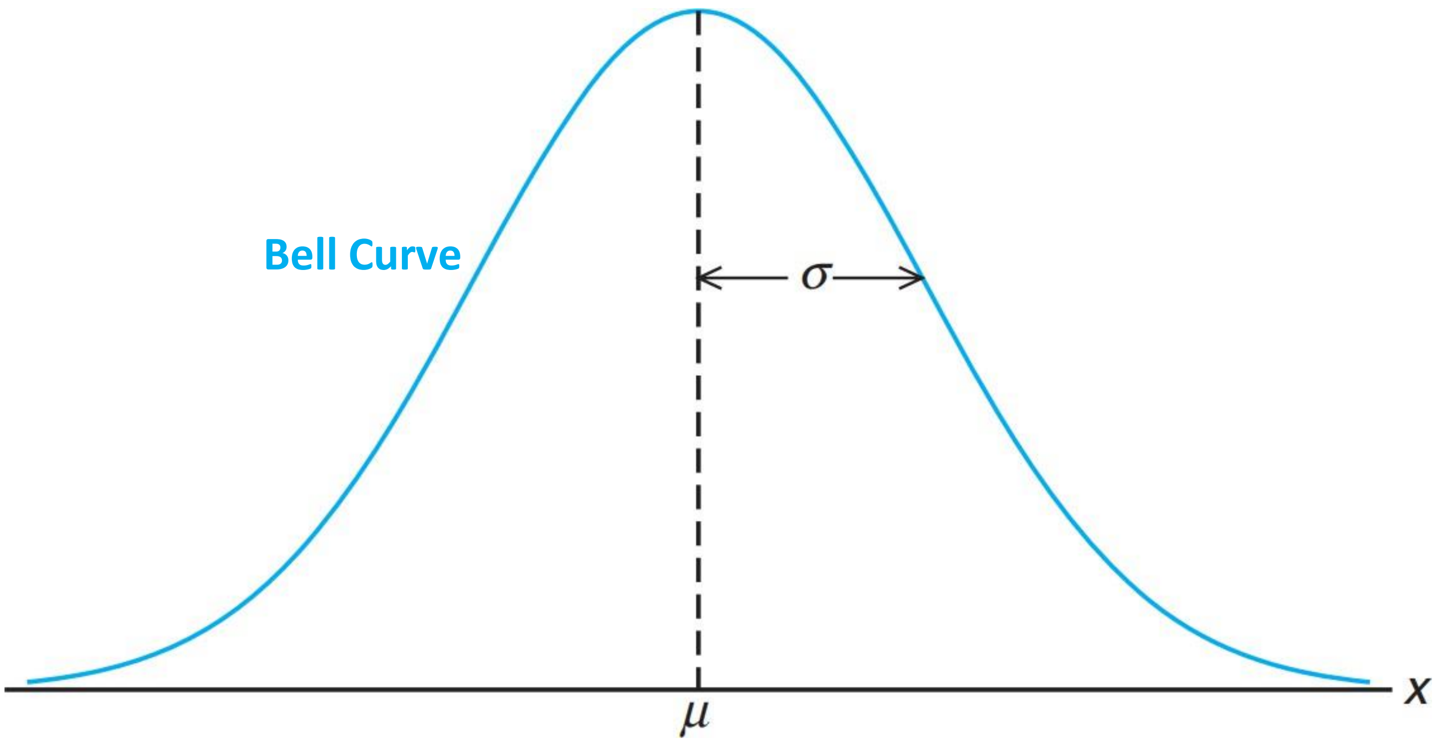
d) What is the average waiting time?

$$\mu = E(X) = \frac{0+10}{2} = 5 \text{ minutes}$$

Normal Distribution (1/41)



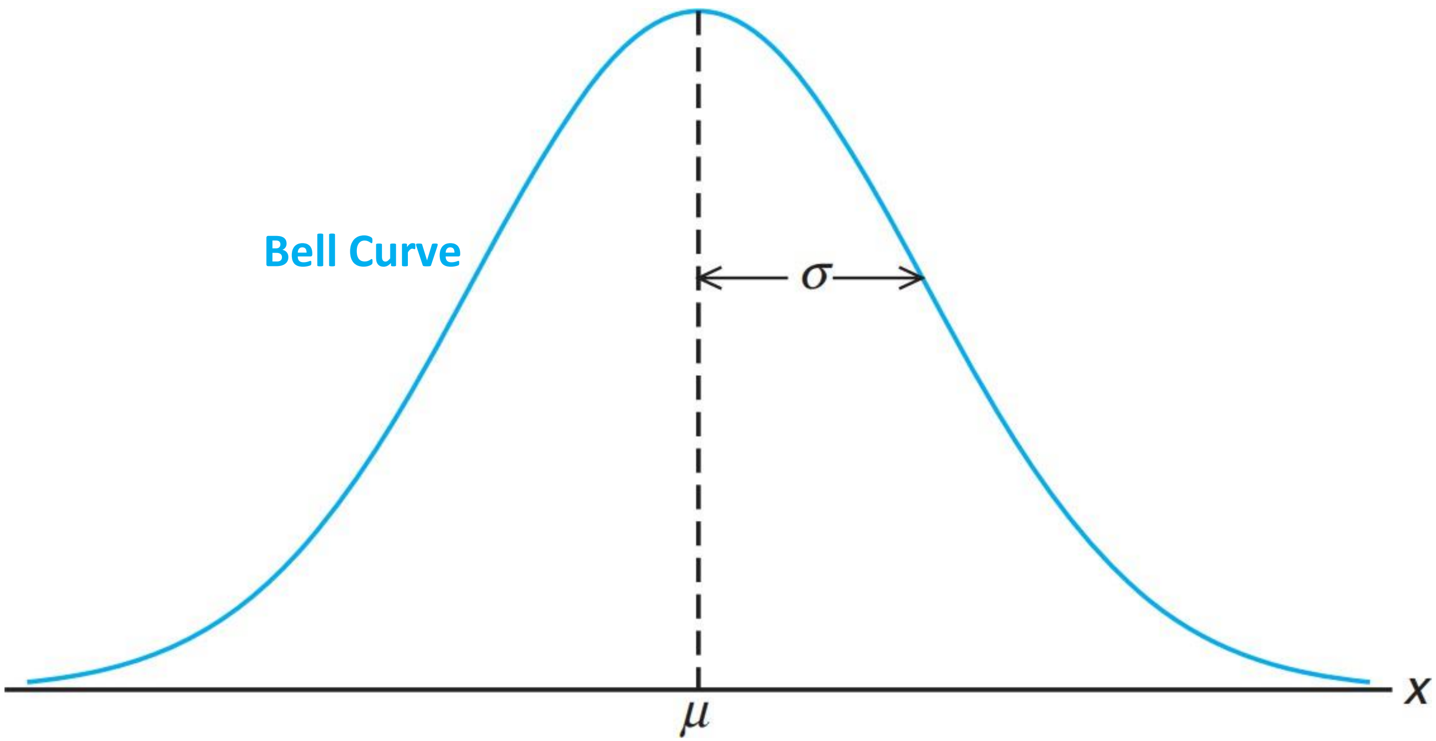
Bell



Normal Distribution (2/41)

A continuous random variable X having the bell-shaped distribution of the following figure is called a *normal random variable*.

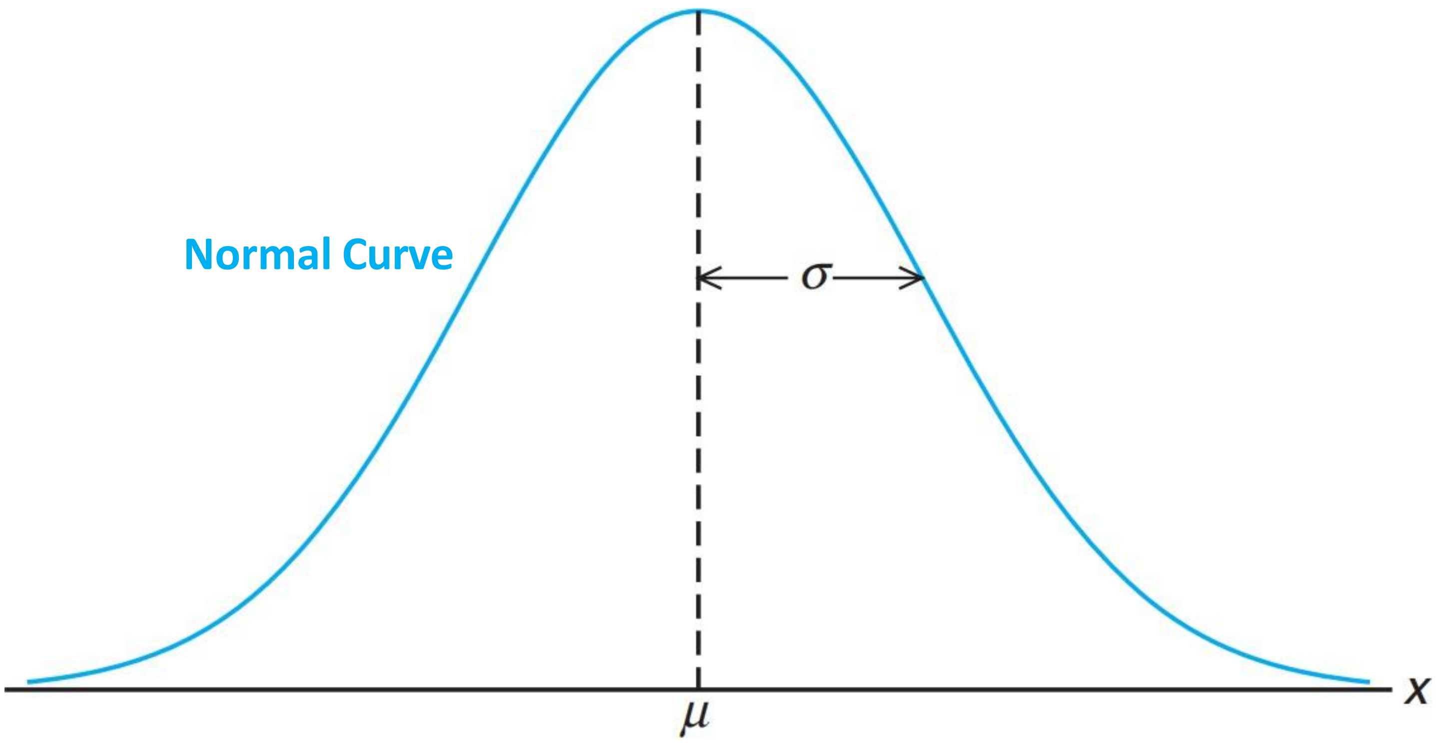
Its graph, called the *normal curve*.



Normal Distribution (2/41)

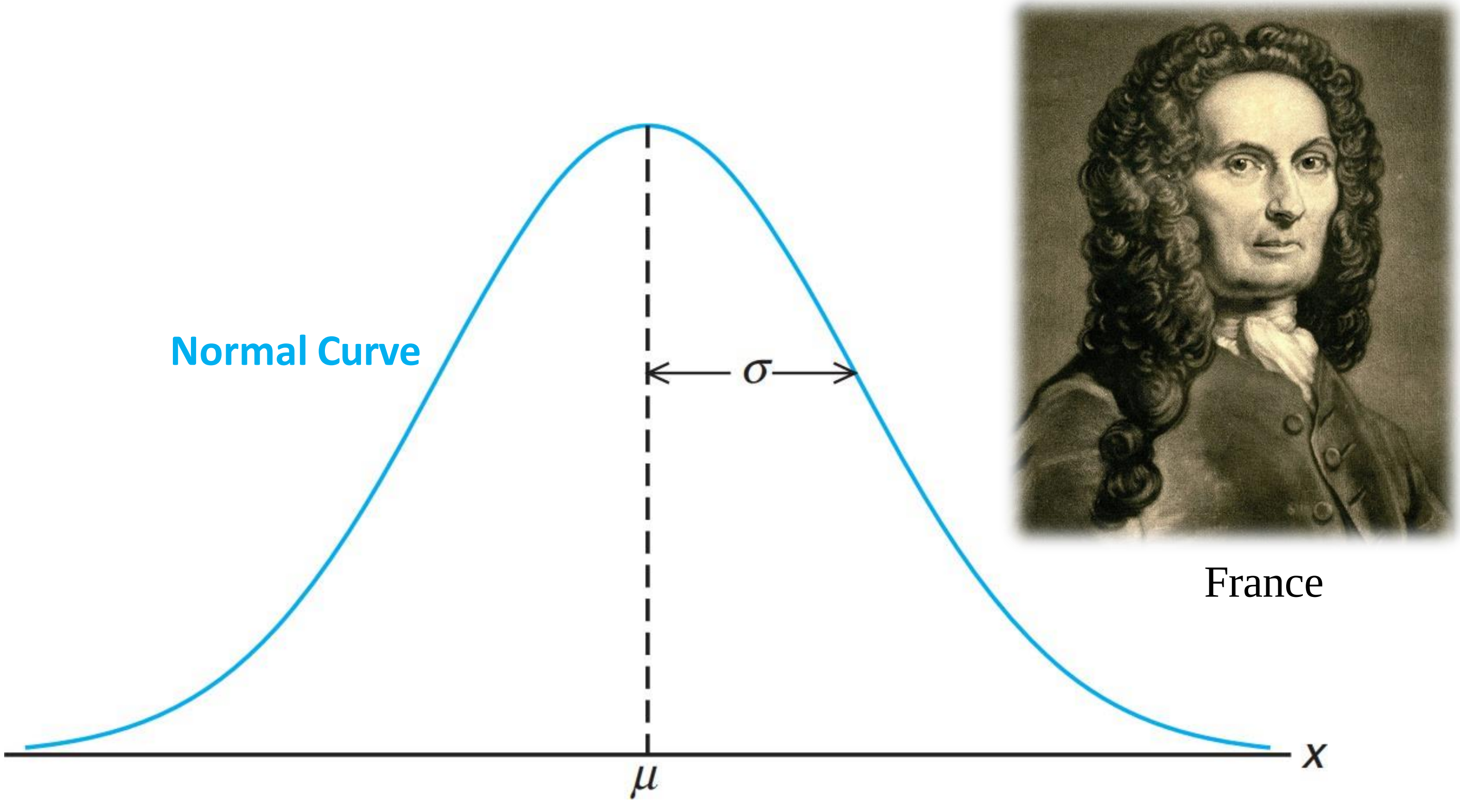
A continuous random variable X having the bell-shaped distribution of the following figure is called a *normal random variable*.

Its graph, called the *normal curve*.



Normal Distribution (3/41)

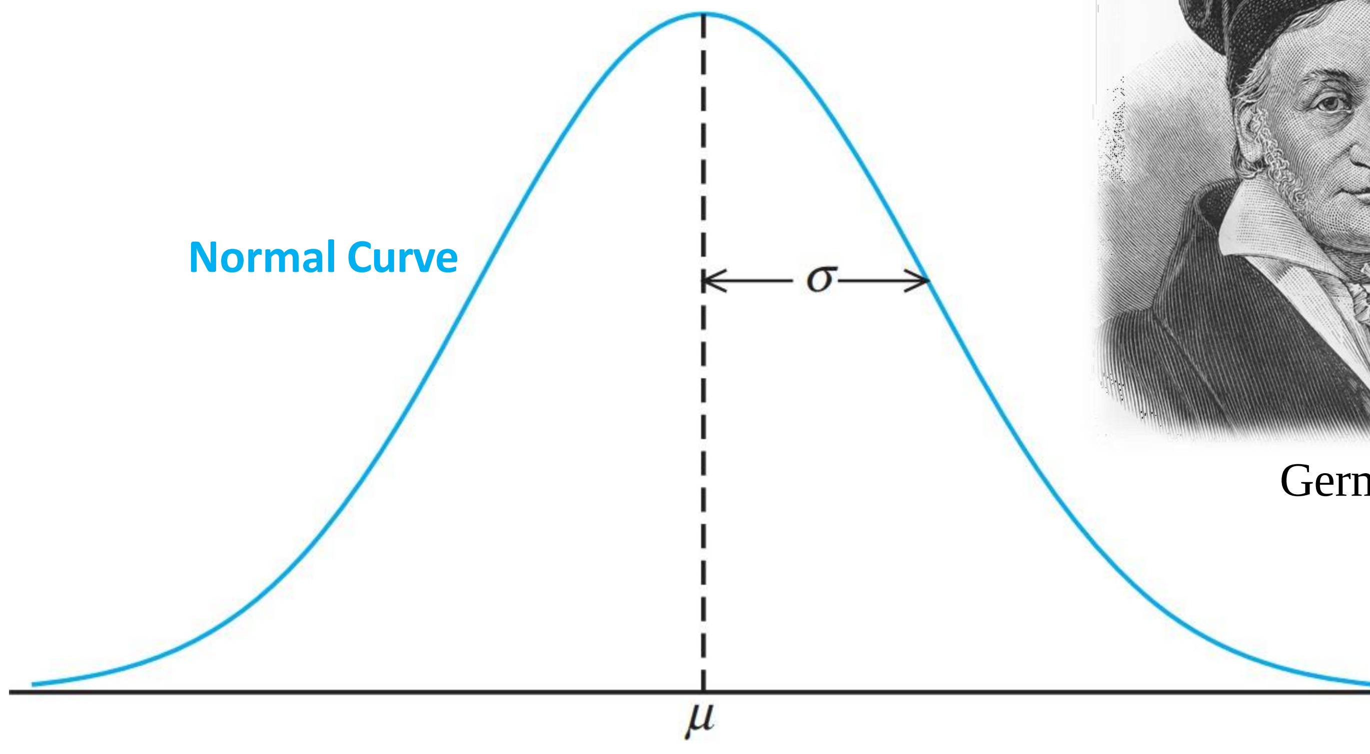
In 1733, **Abraham DeMoivre** developed the mathematical equation of the normal curve.



France

Normal Distribution (4/41)

The normal distribution is often referred to as the **Gaussian distribution**, in honor of Karl Friedrich Gauss, who discovered the normal distribution in 1823.



Germany

Normal Distribution (5/41)

X : continuous random variable

μ : mean

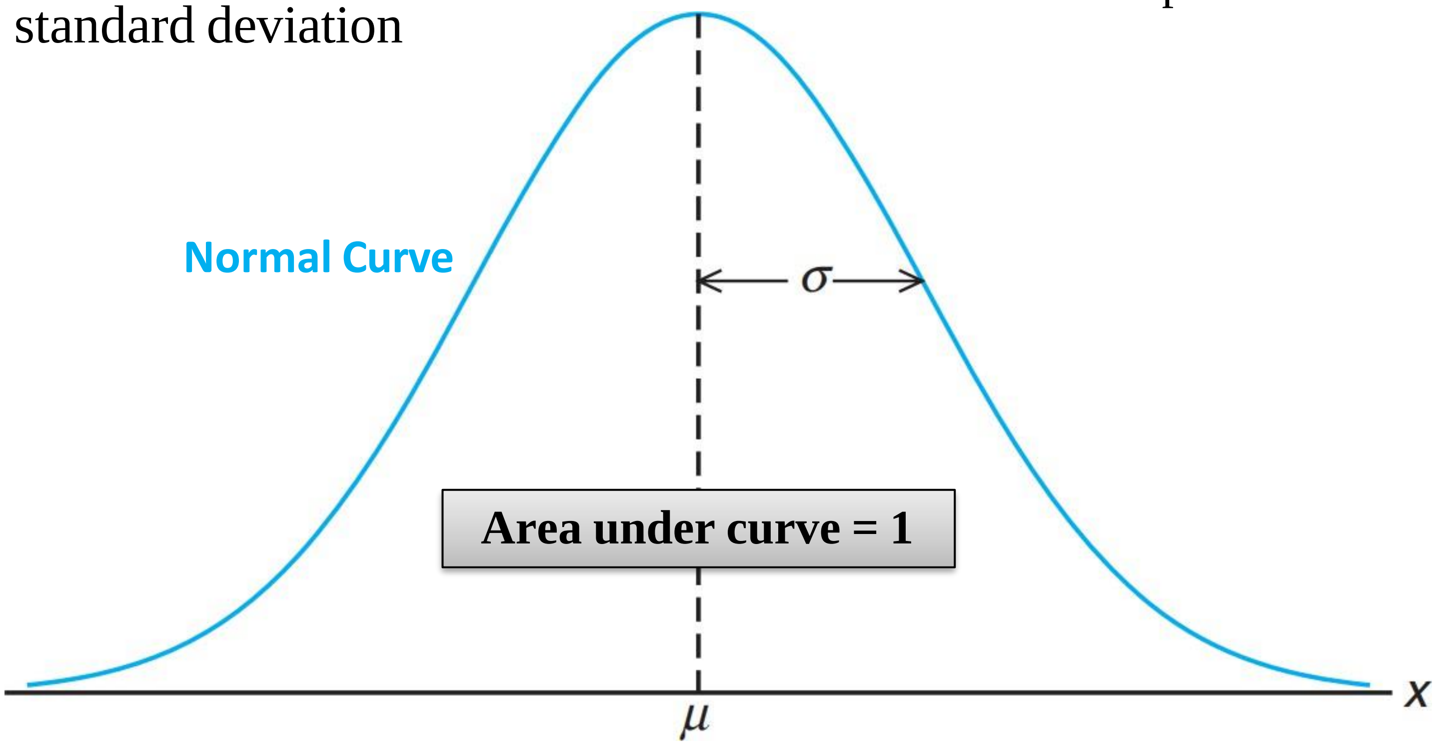
σ^2: variance

σ : standard deviation

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

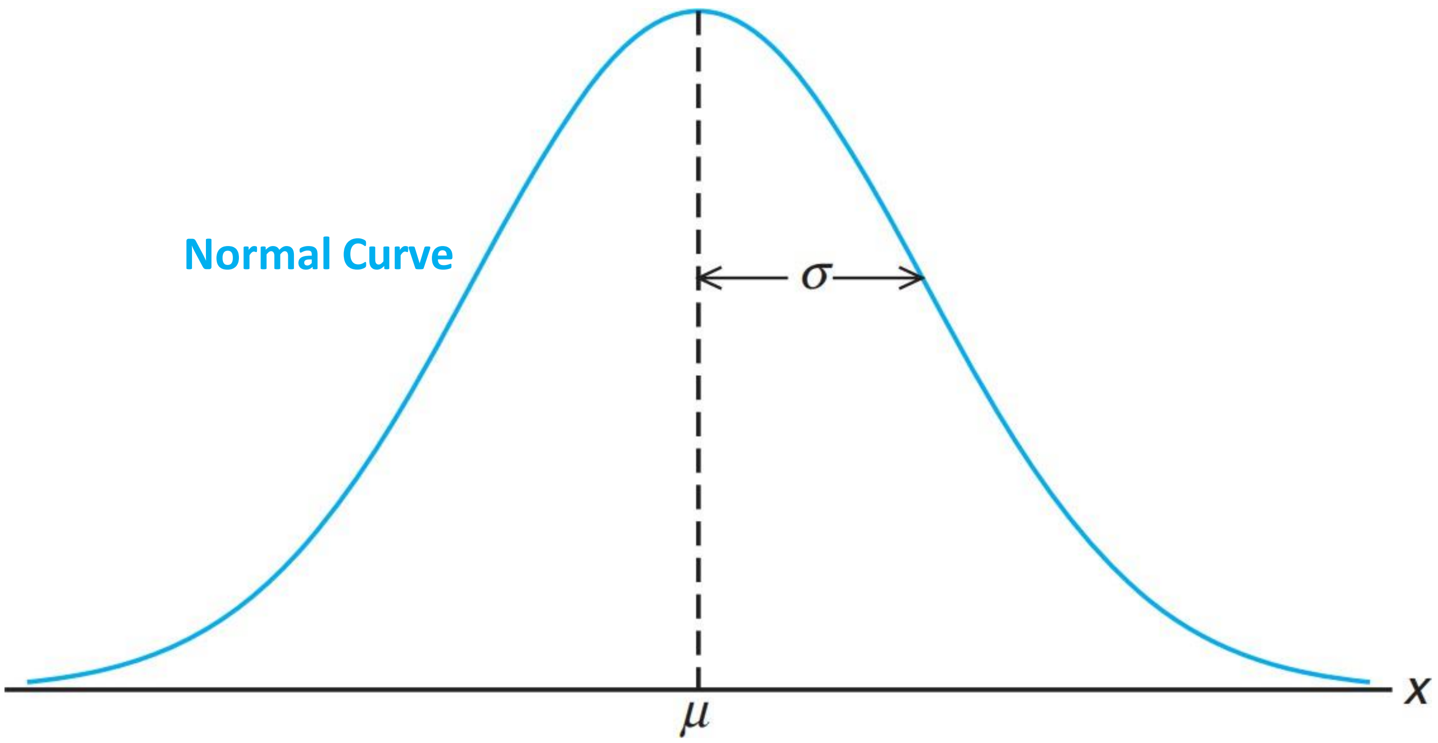
Location parameter $\mu \in \mathbb{R}$,

Scale parameter $\sigma > 0$.



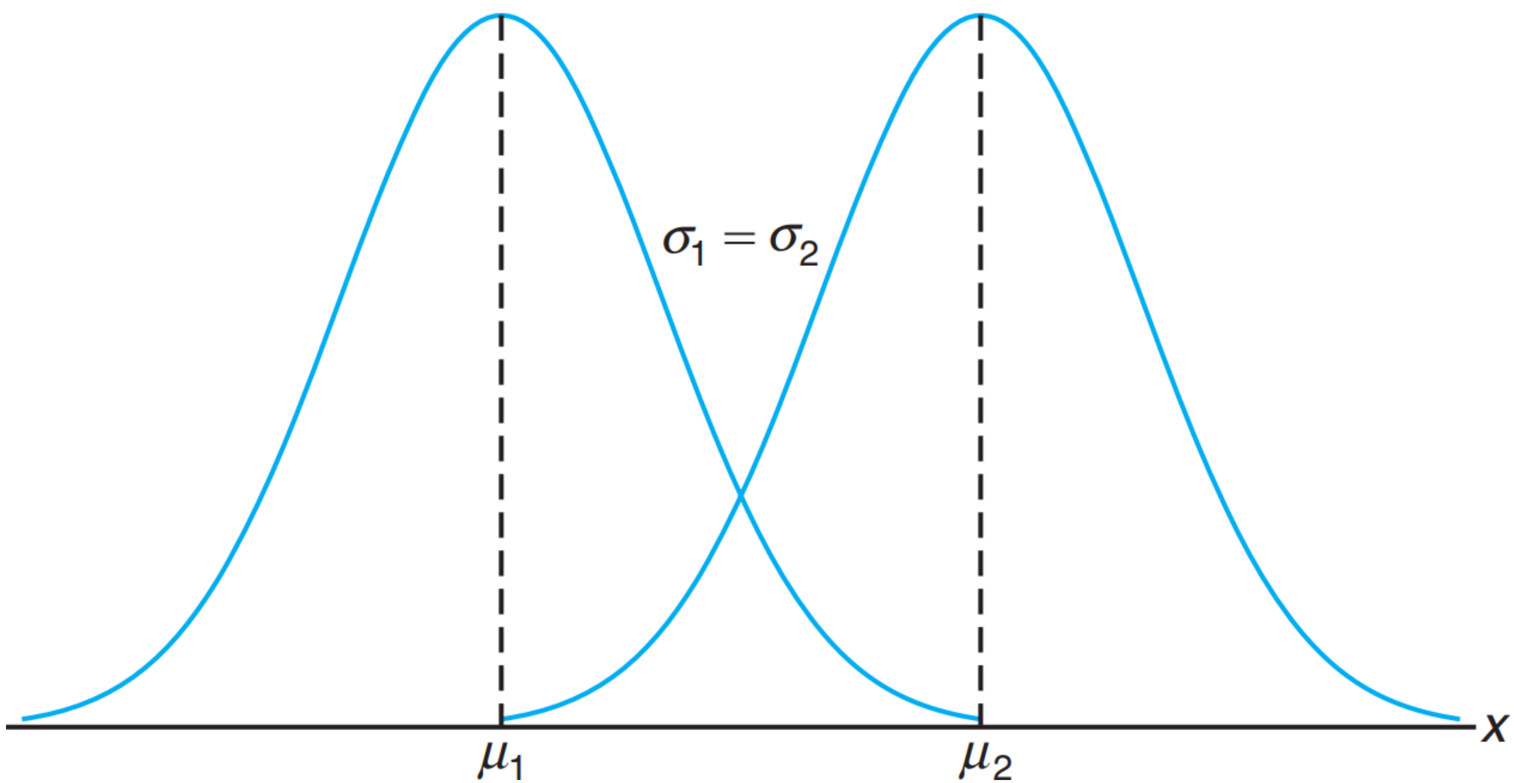
Normal Distribution (6/41)

Note that the graph of the normal distribution is ***symmetric*** about its mean μ (i.e., for any $c > 0$, $P(X > \mu + c) = P(X < \mu - c)$, and that σ^2 is a measure of the width of the bell shape.



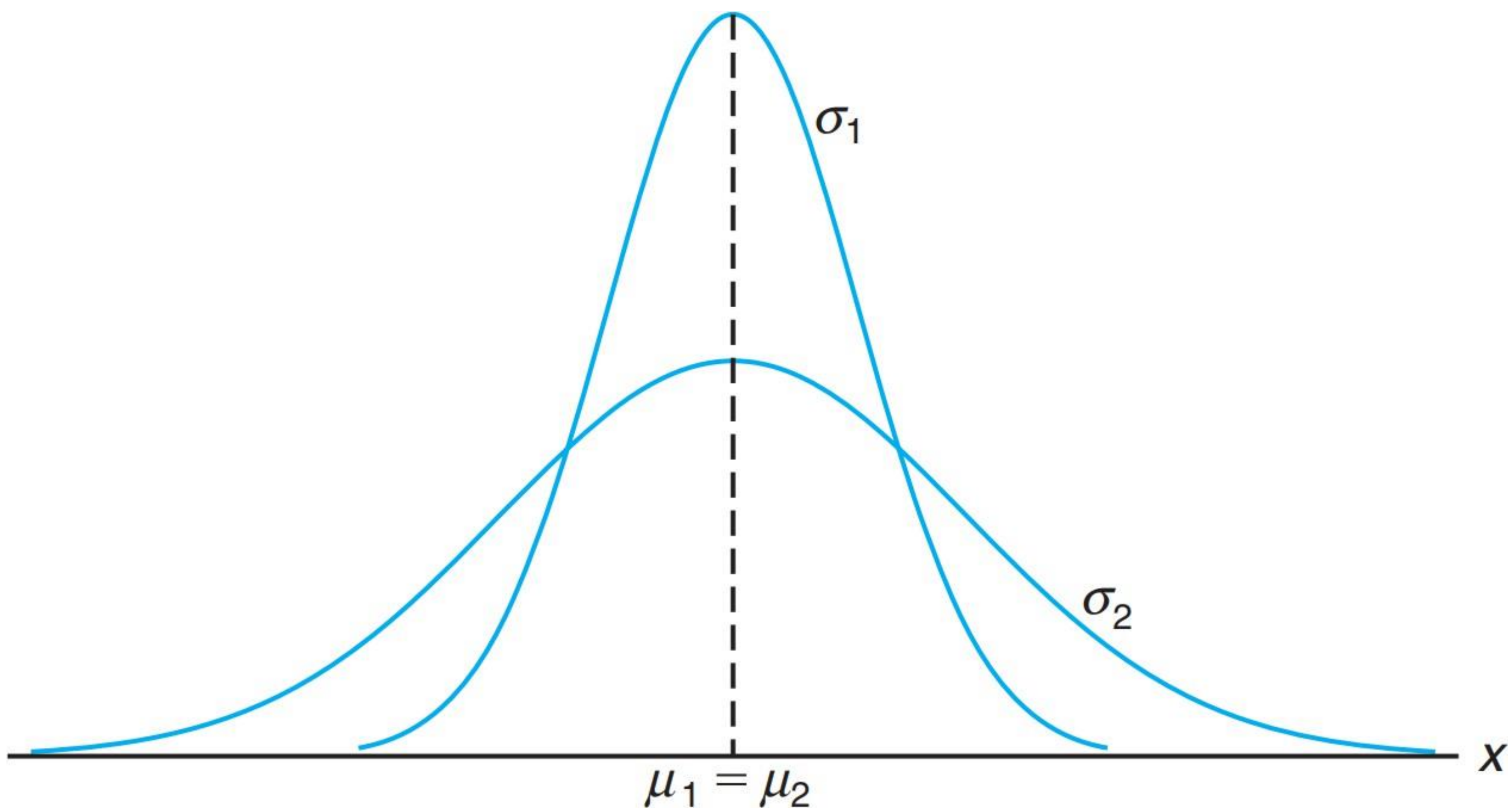
Normal Distribution (7/41)

Same standard deviation but different means $(\mu_1 < \mu_2)$.



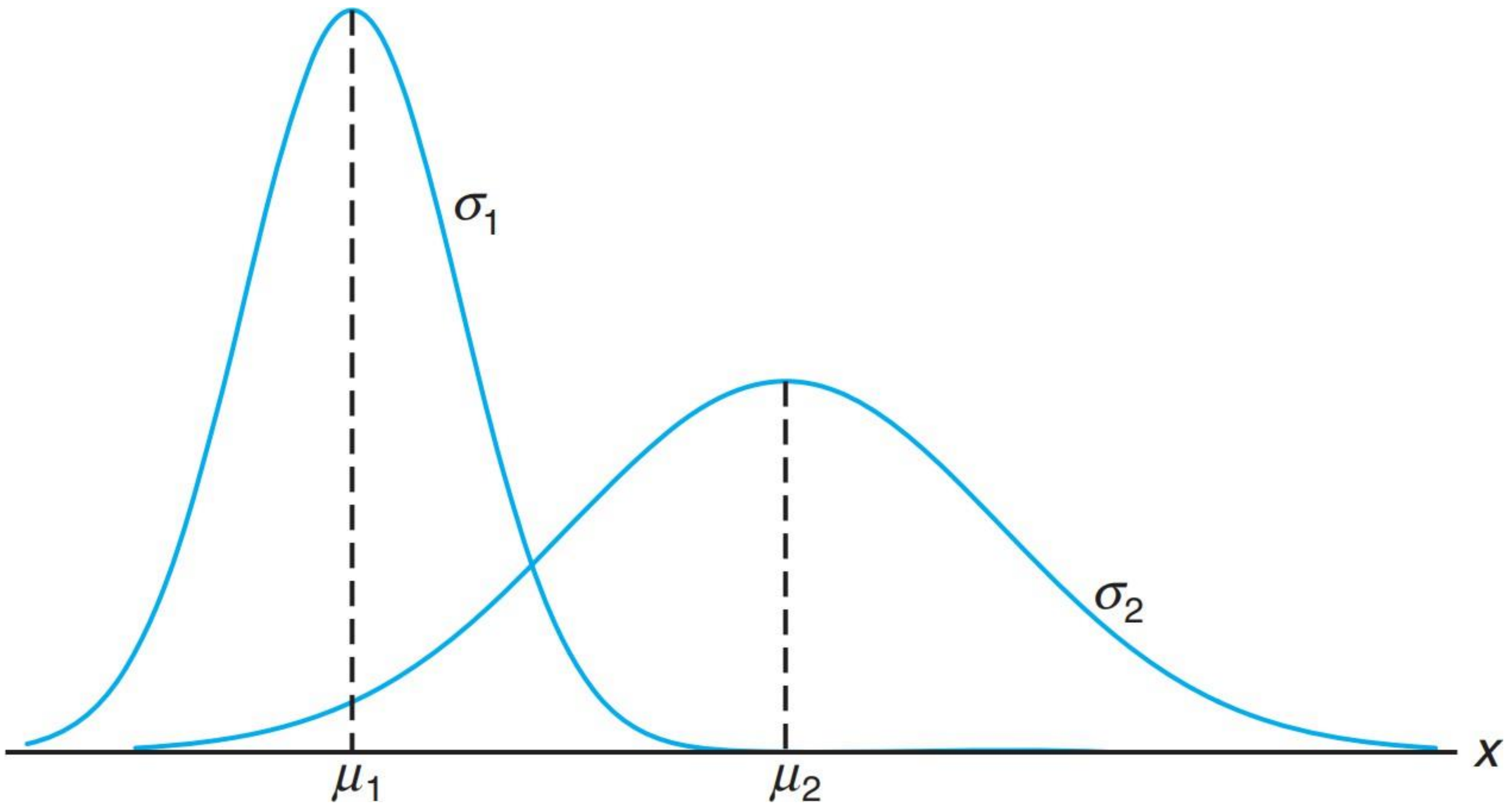
Normal Distribution (8/41)

Same mean but different standard deviations $(\sigma_1 < \sigma_2)$.



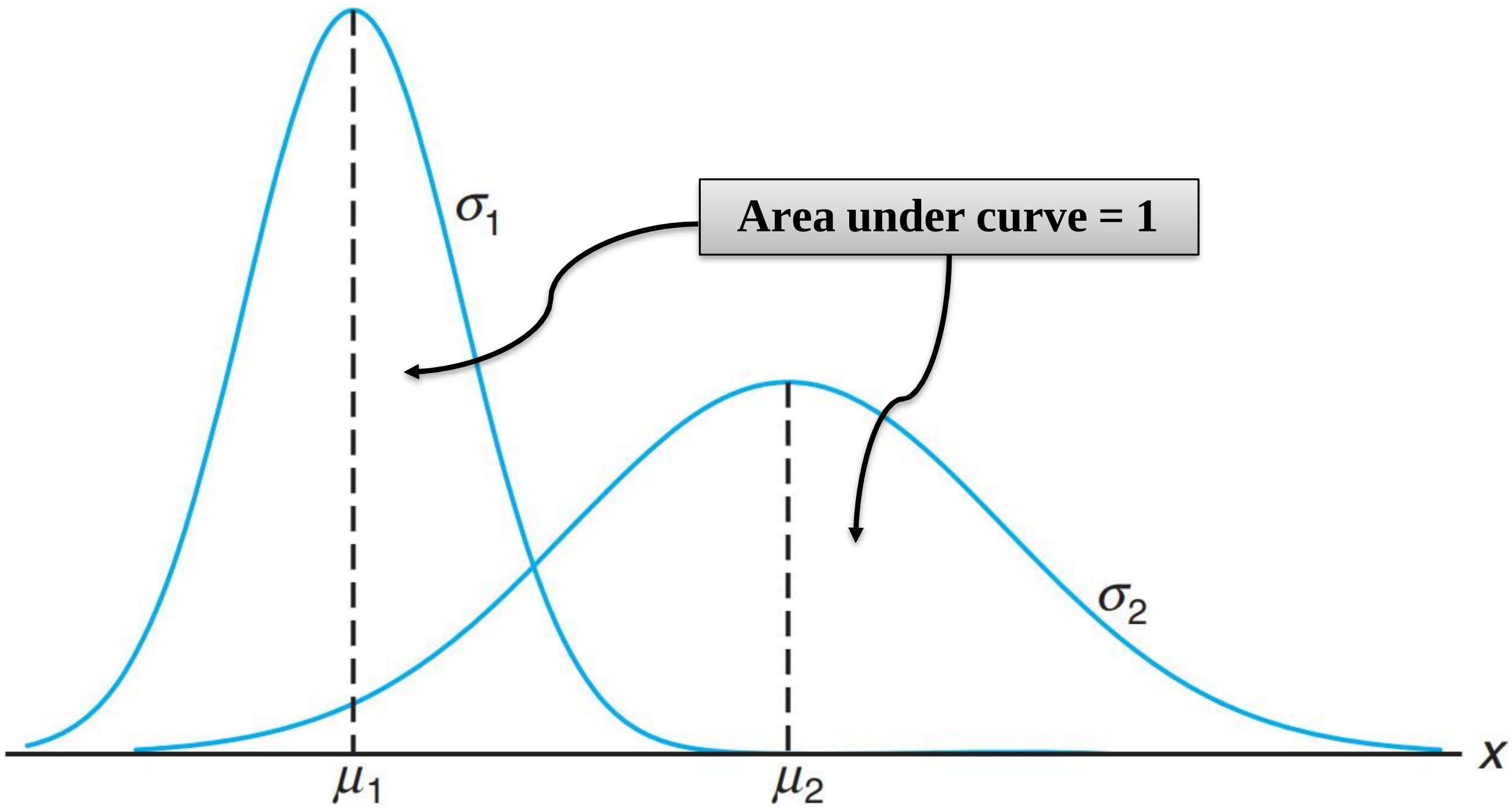
Normal Distribution (9/41)

Different means $(\mu_1 < \mu_2)$ and standard deviations $(\sigma_1 < \sigma_2)$.



Normal Distribution (10/41)

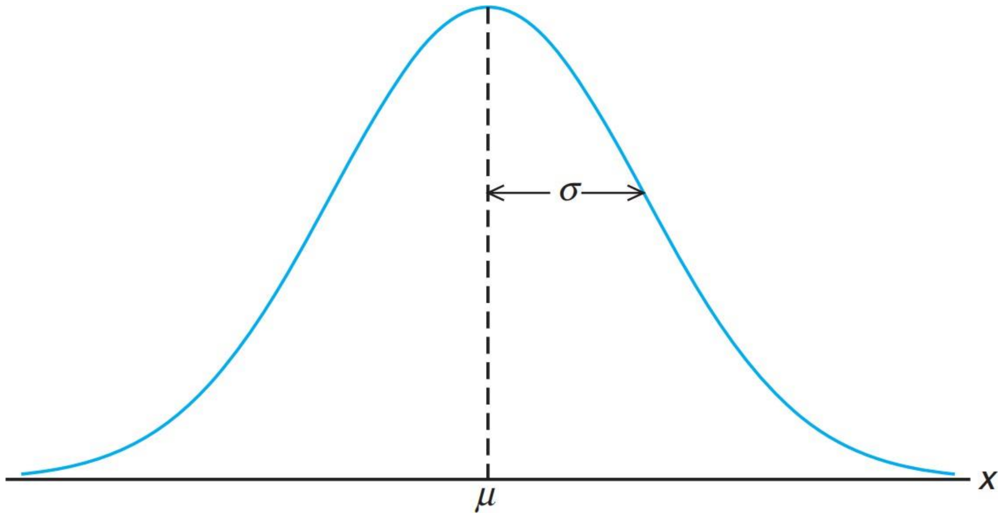
Different means $(\mu_1 < \mu_2)$ and standard deviations $(\sigma_1 < \sigma_2)$.



Normal Distribution (11/41)

Properties of the normal curve (1/2):

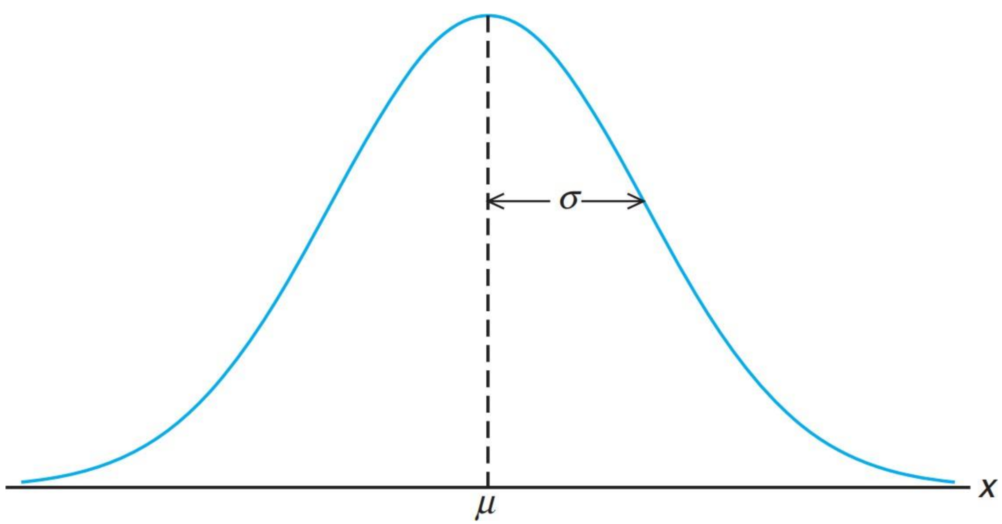
1) Mode = Mean (μ)
2) The curve is symmetric about a vertical axis through the mean (μ).
3) The curve has its points of inflection at $x = \mu \pm \sigma$; it is concave downward if $\mu - \sigma < X < \mu + \sigma$ and is concave upward otherwise.



Normal Distribution (11/41)

Properties of the normal curve (2/2):

4) The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
5) The total area under the curve and above the horizontal axis is equal to 1.



Normal Distribution (12/41)

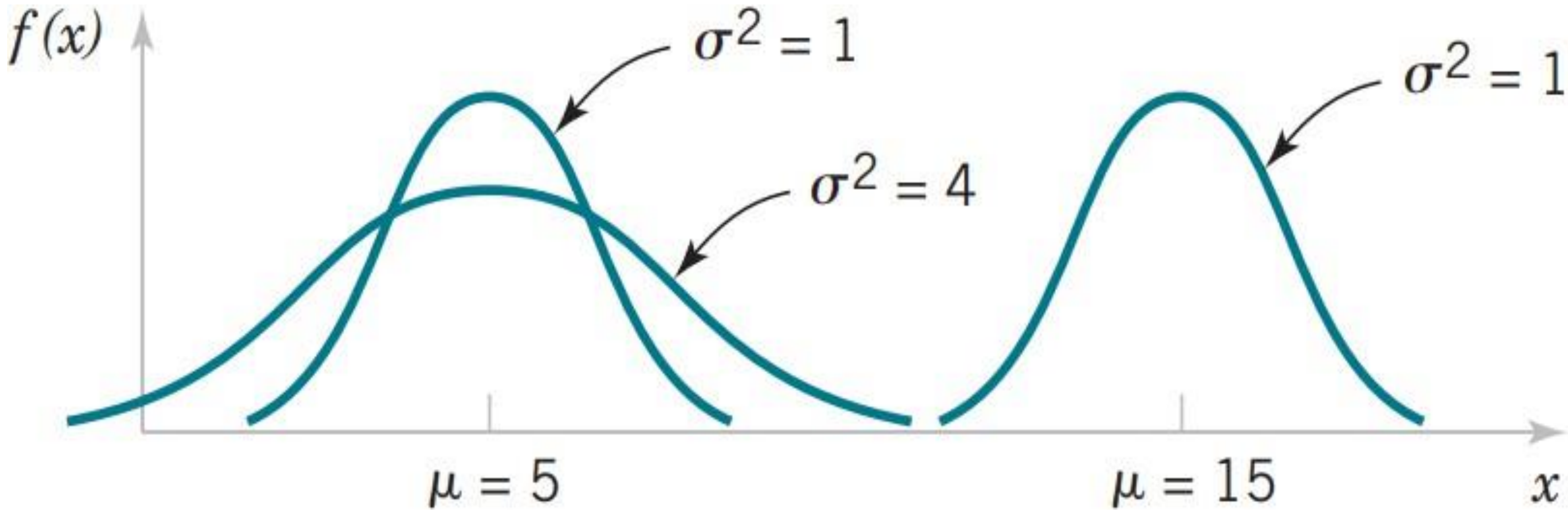
Probability density function of the normal random variable *X*:

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

The mean and variance of X are μ and σ^2, respectively,

where $-\infty < \mu < \infty$ and $\sigma > 0$

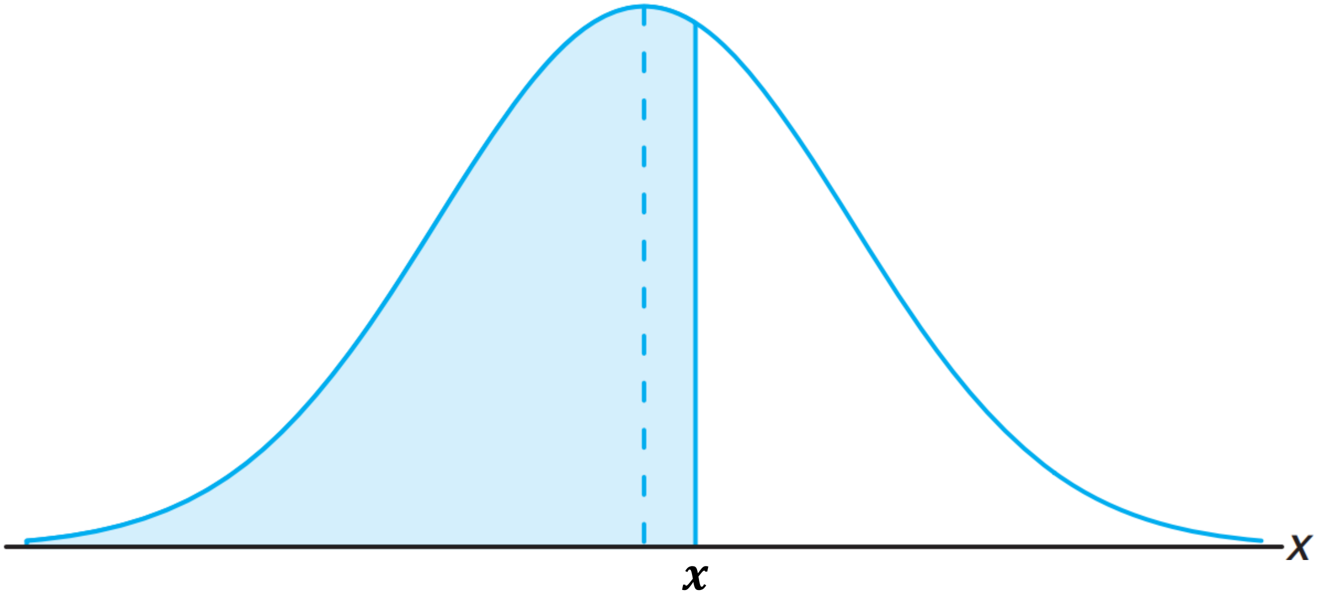
Normal Distribution (13/41)



Normal Distribution (14/41)

Cumulative distribution function:

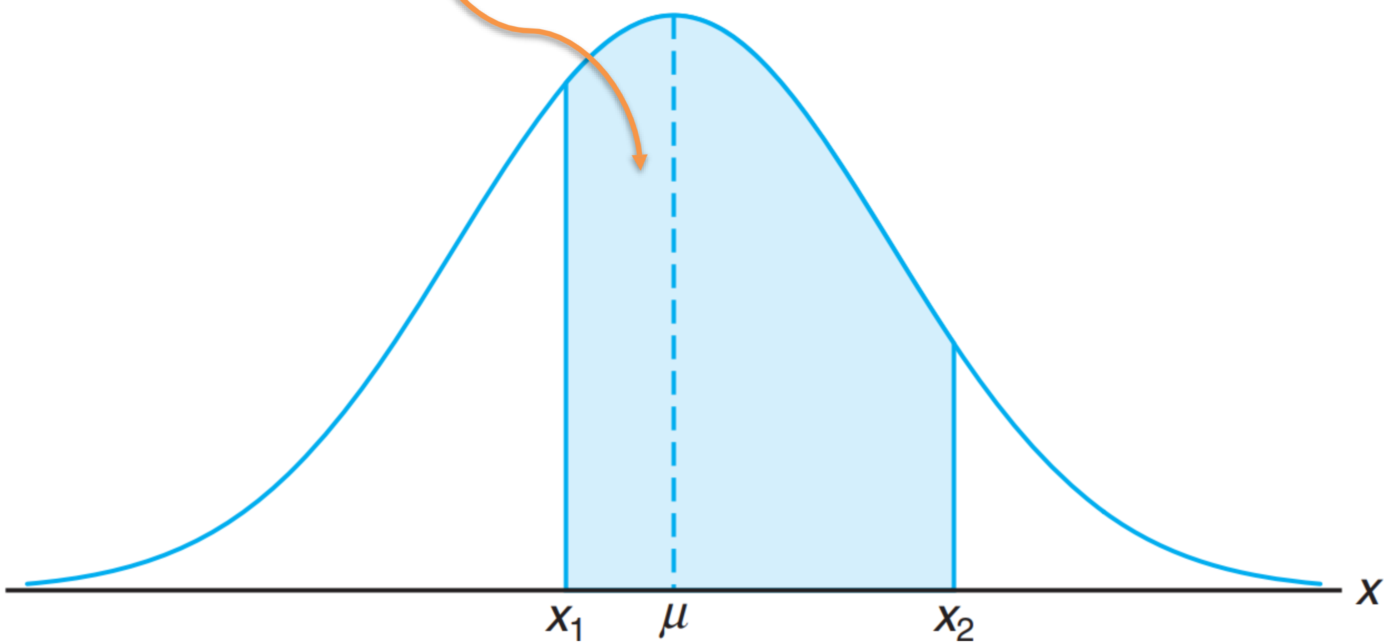
$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(v-\mu)^2}{2\sigma^2}}\, dv$$



Normal Distribution (15/41)

Areas under the Normal Curve:

$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx$$

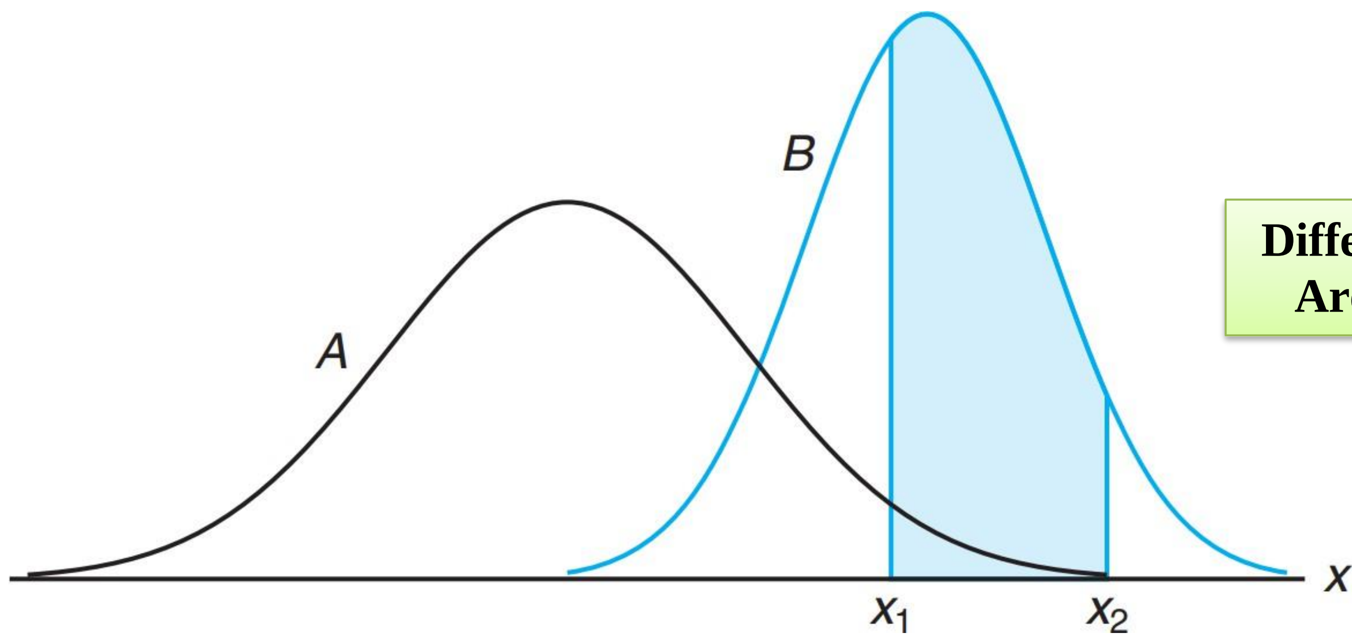


Normal Distribution (16/41)

Areas under the Normal Curve:

Different means and standard deviations

$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx$$



Different Areas

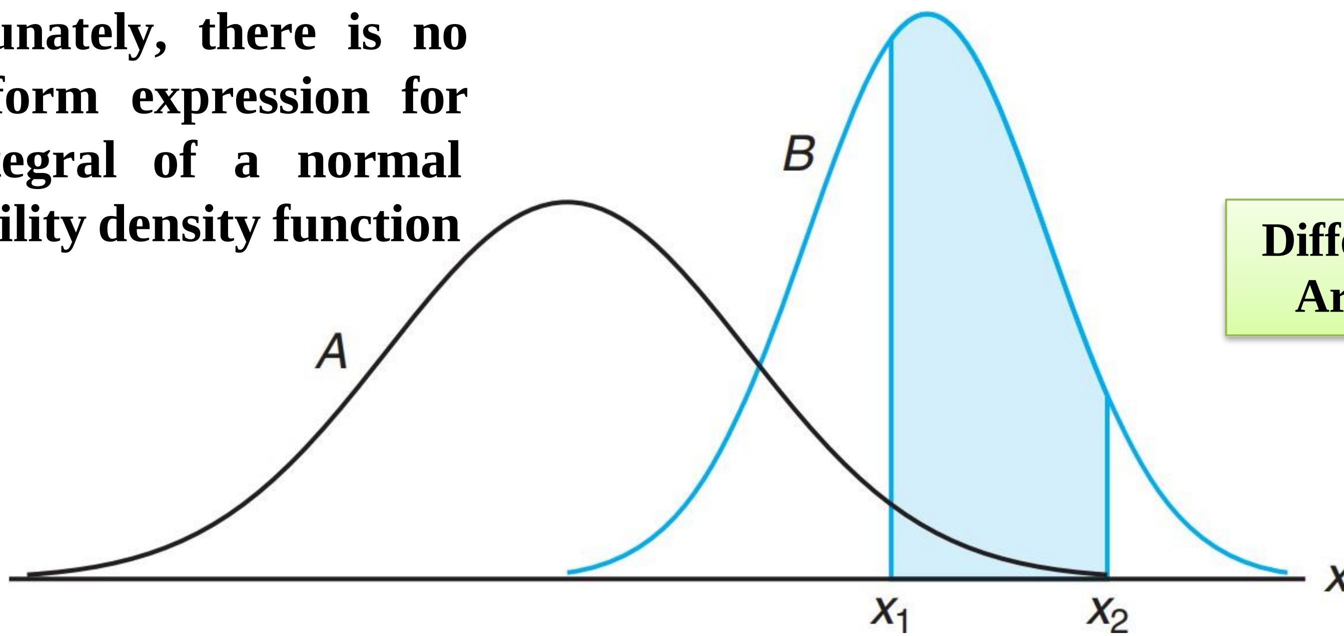
Normal Distribution (16/41)

Areas under the Normal Curve:

Different means and standard deviations

$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx$$

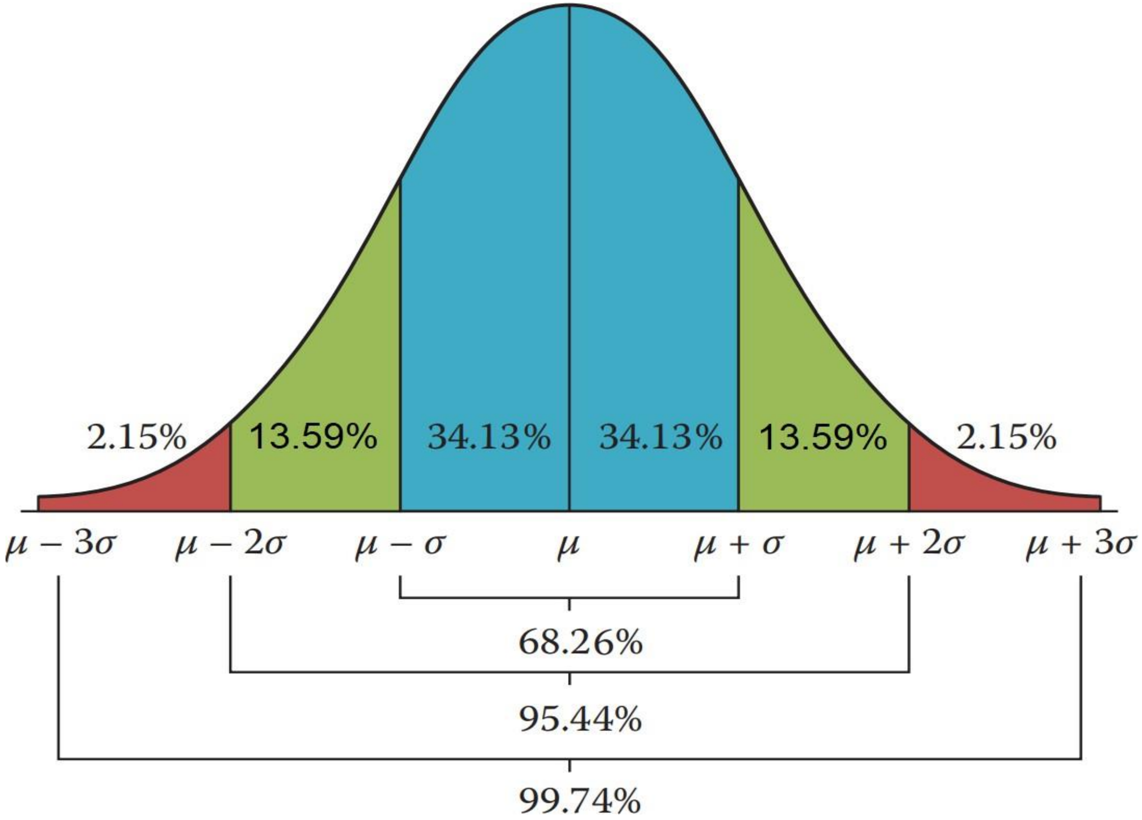
Unfortunately, there is no closed-form expression for the integral of a normal probability density function



Different Areas

Normal Distribution (17/41)

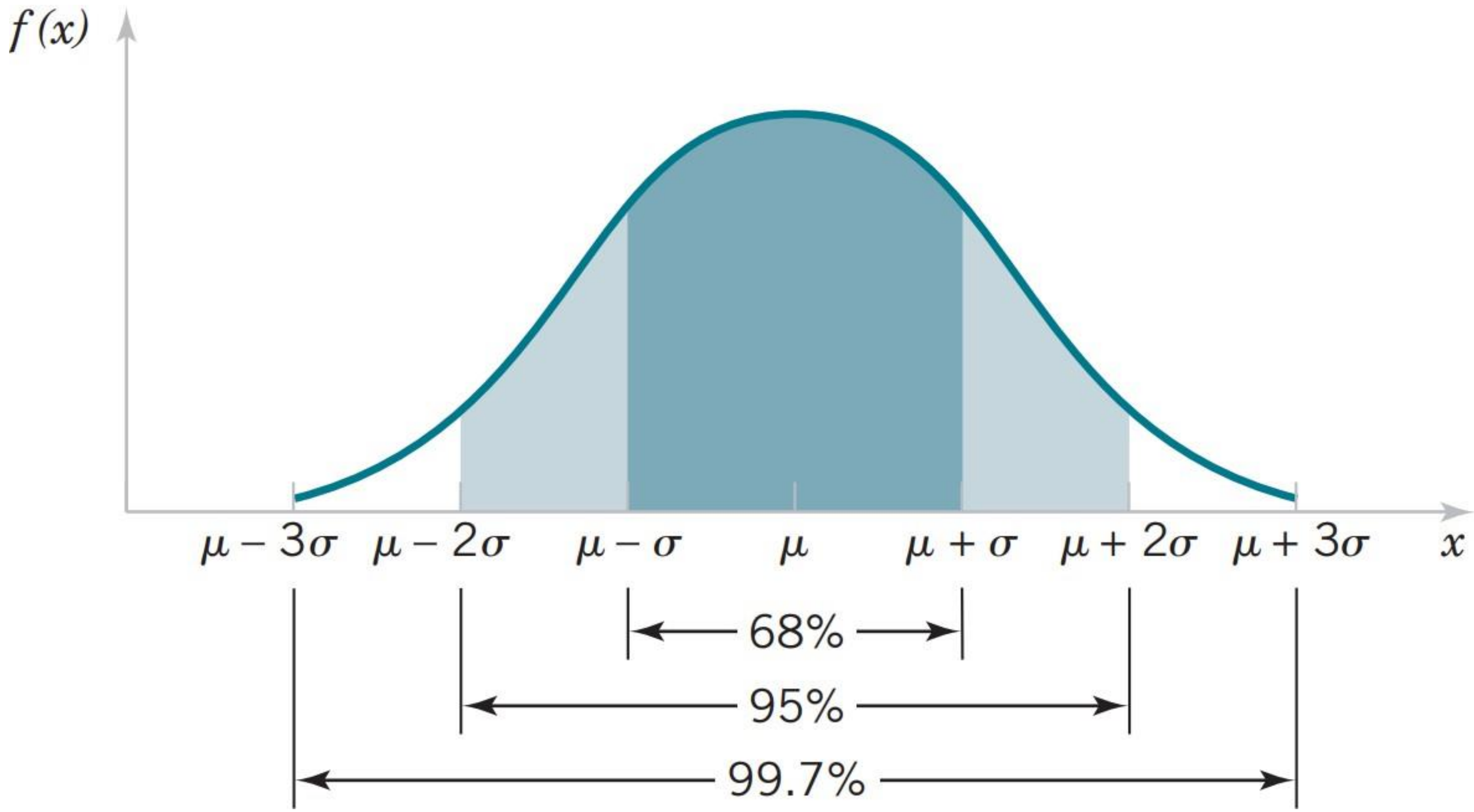
Empirical Rule (1/2):
68-95-99.7 rule *or* three-sigma rule



Normal Distribution (17/41)

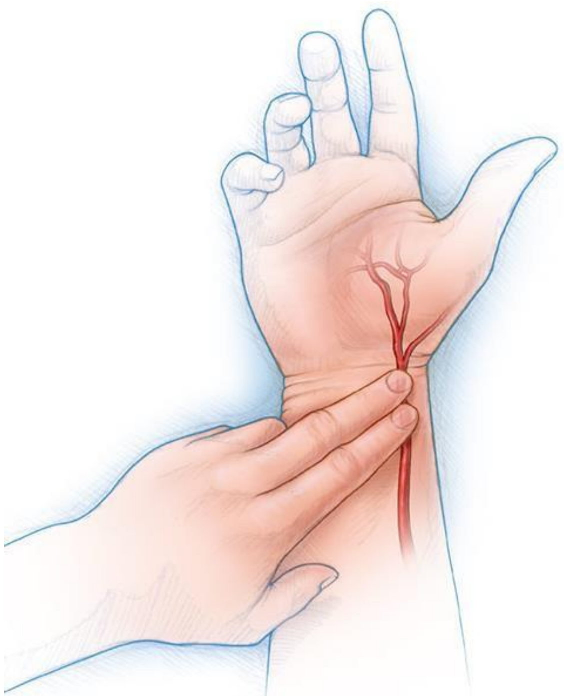
Empirical Rule (2/2):

68-95-99.7 rule *or* three-sigma rule



Example1 (1/5):

Suppose the pulse rates of 200 college men are bell-shaped with a mean of 72 beats per minute (bpm) and a standard deviation of 6 bpm.

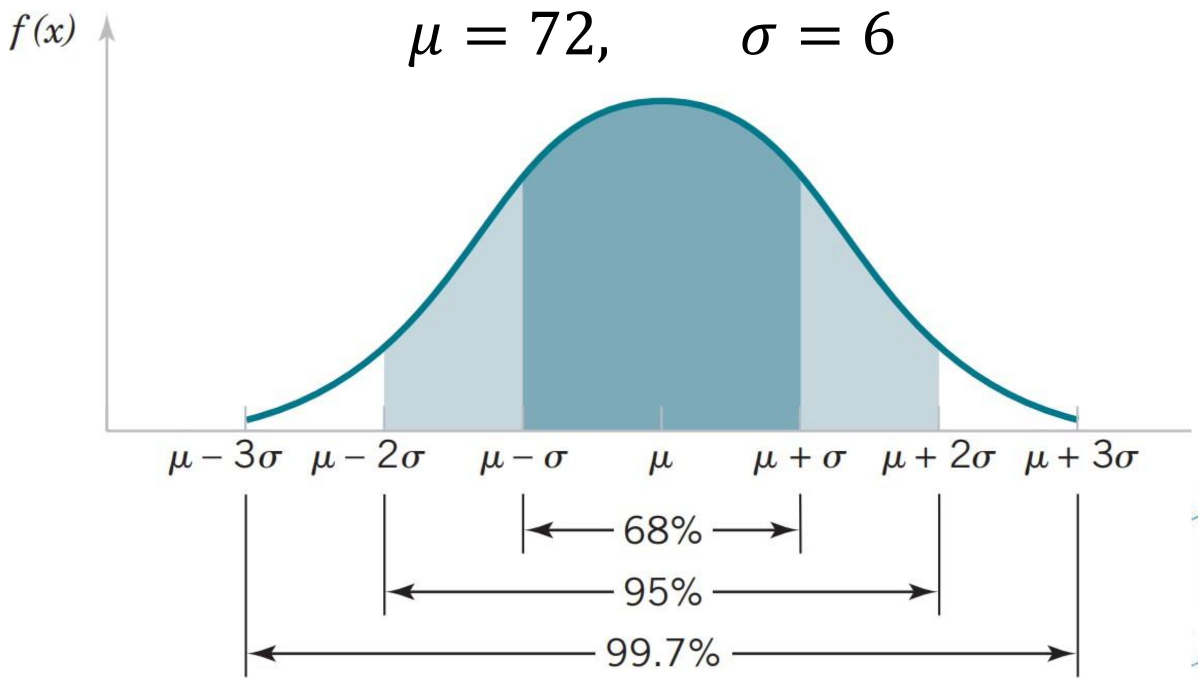


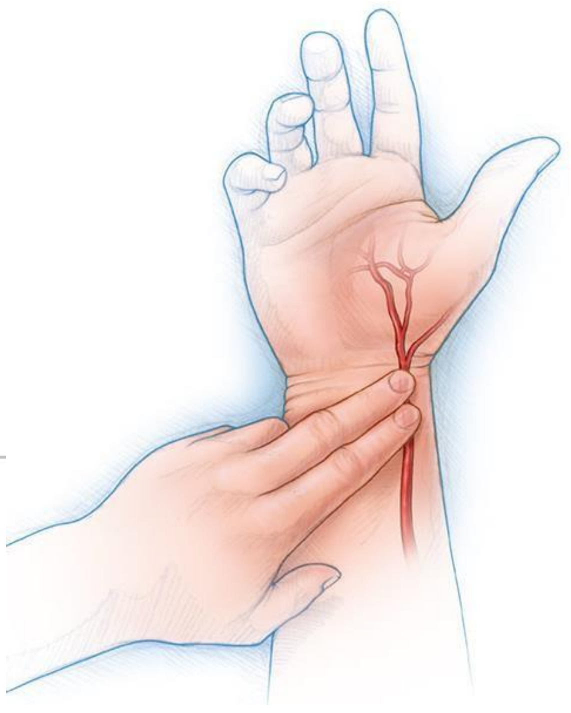
Normal Distribution (18/41)

Example1 (2/5):

Suppose the pulse rates of 200 college men are bell-shaped with a mean of 72 bpm and standard deviation of 6 bpm.

$$\mu = 72, \qquad \sigma = 6$$

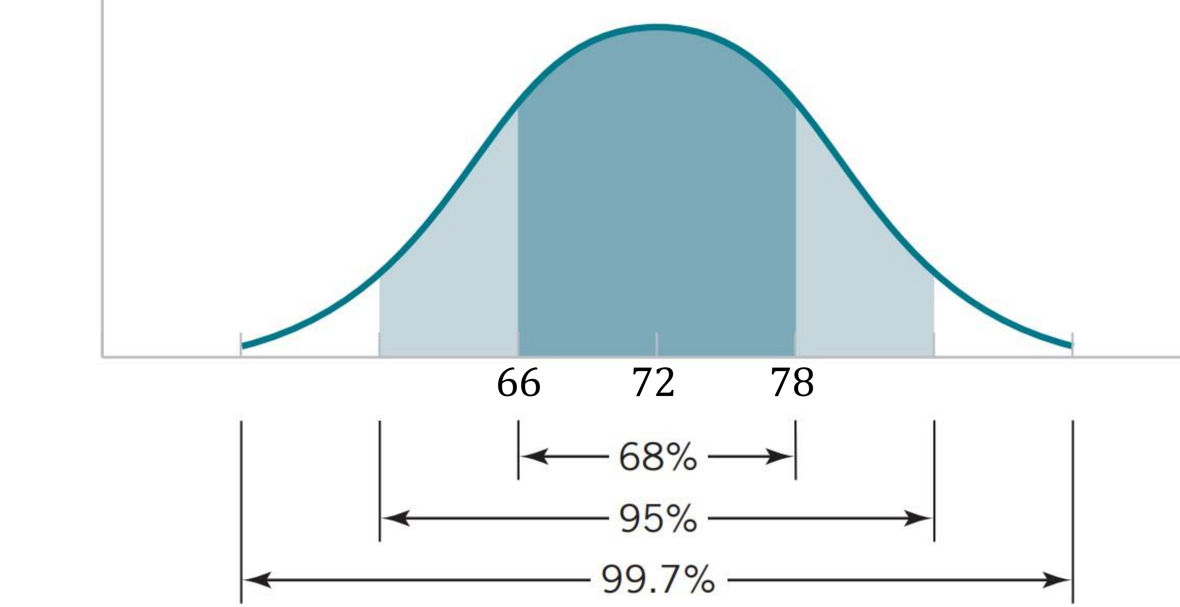


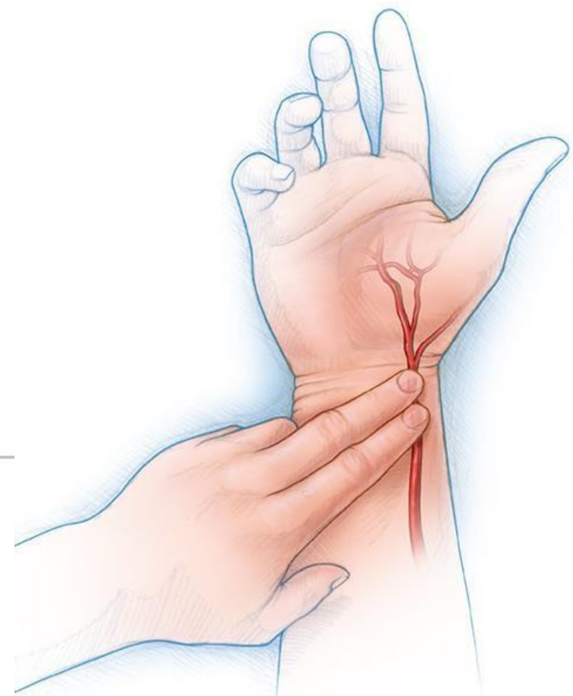


Example1 (3/5):

About 68% of the men have pulse rates in the interval $[66, 78]$ bpm.

$\mu = 72, \quad \sigma = 6$



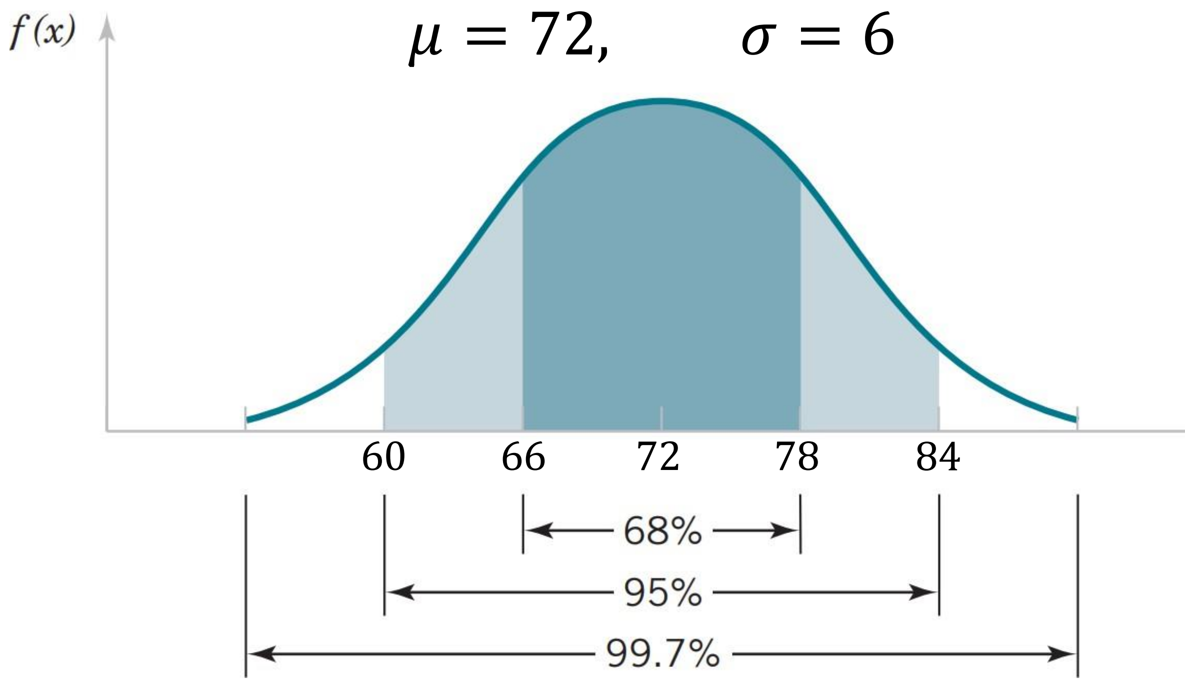


Normal Distribution (18/41)

Example1 (4/5):

About 95% of the men have pulse rates in the interval $[60, 84]$ bpm.

$$\mu = 72, \qquad \sigma = 6$$

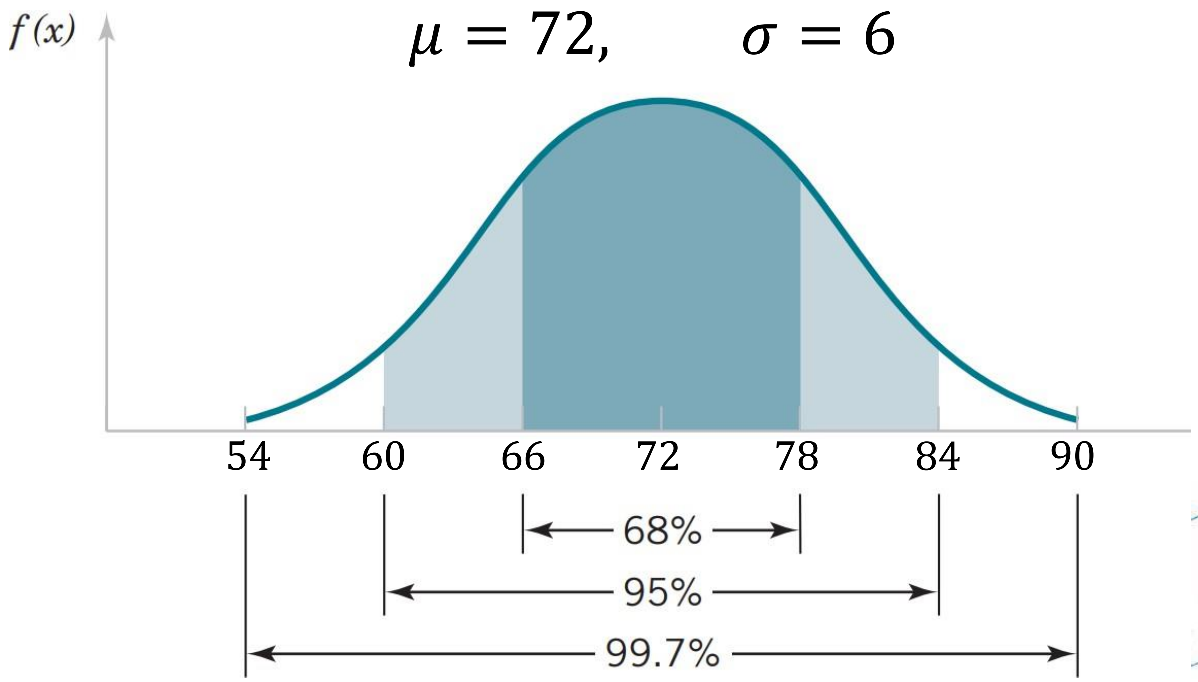


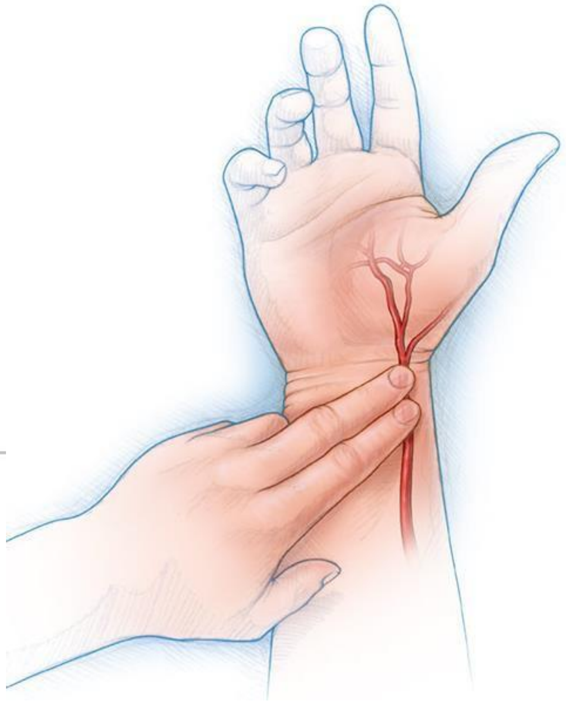
Normal Distribution (18/41)

Example1 (5/5):

About 99.7% of the men have pulse rates in the interval $[54, 90]$ bpm.

$\mu = 72, \qquad \sigma = 6$



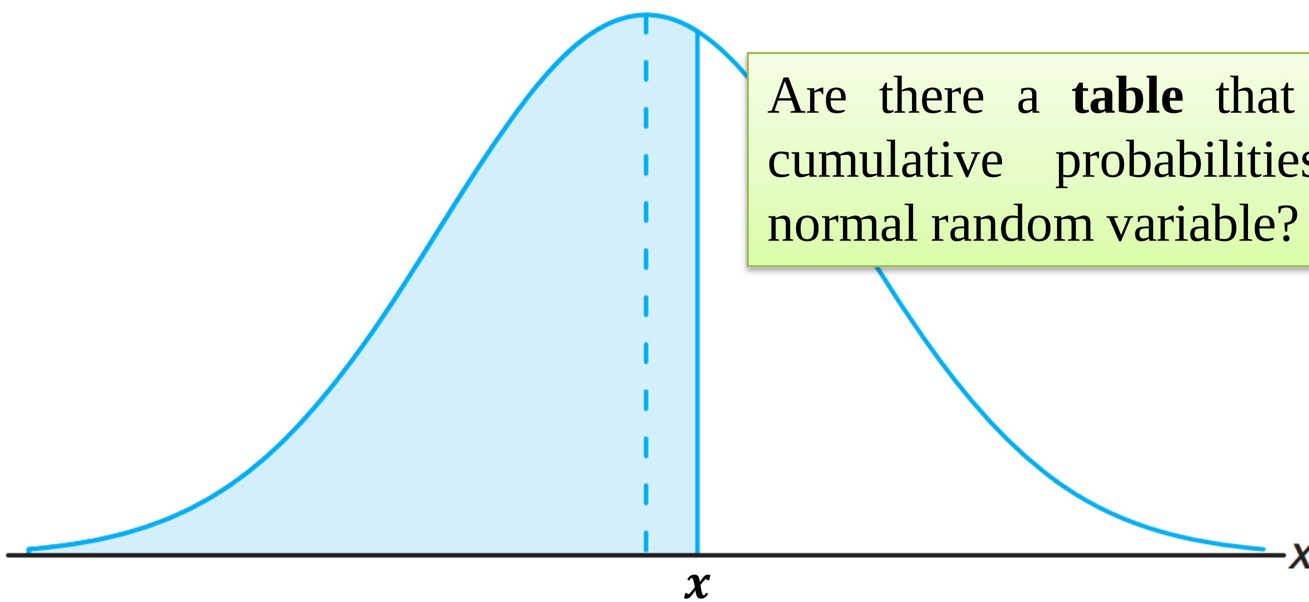


Normal Distribution (19/41)

Recall

Cumulative distribution function:

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(v-\mu)^2}{2\sigma^2}}\, dv$$



Are there a **table** that provides cumulative probabilities for a normal random variable?

Normal Distribution (19/41)

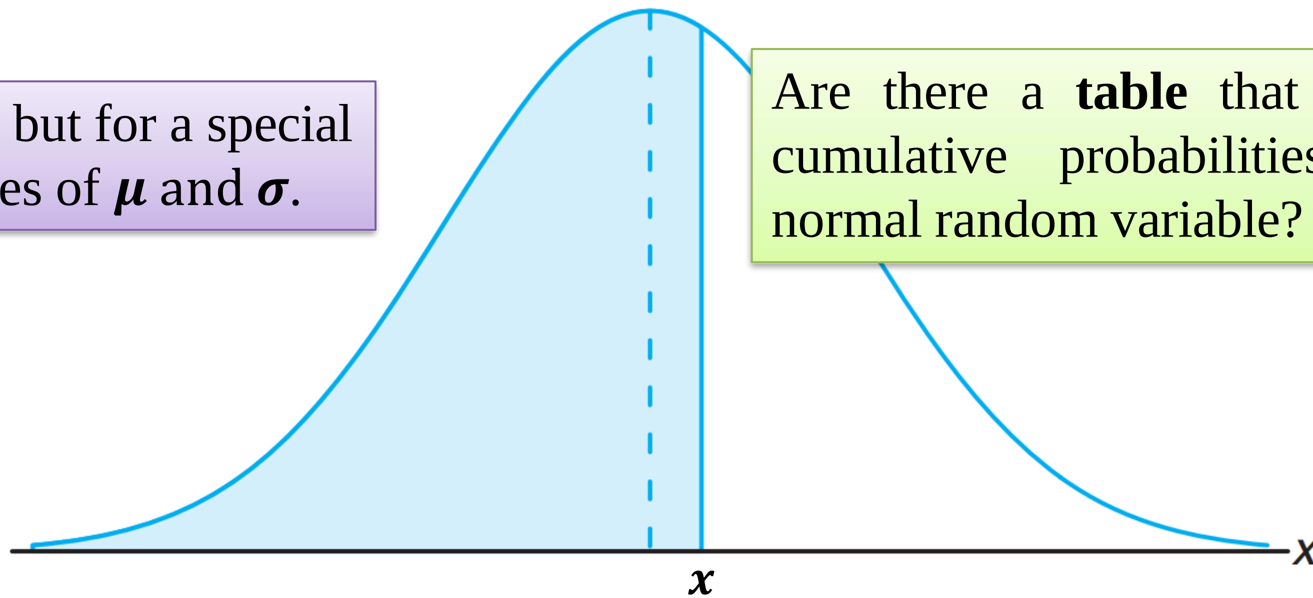
Recall

Cumulative distribution function:

$$F(x) = P(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(v-\mu)^2}{2\sigma^2}}\, dv$$

Yes, but for a special values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.

Are there a **table** that provides cumulative probabilities for a normal random variable?

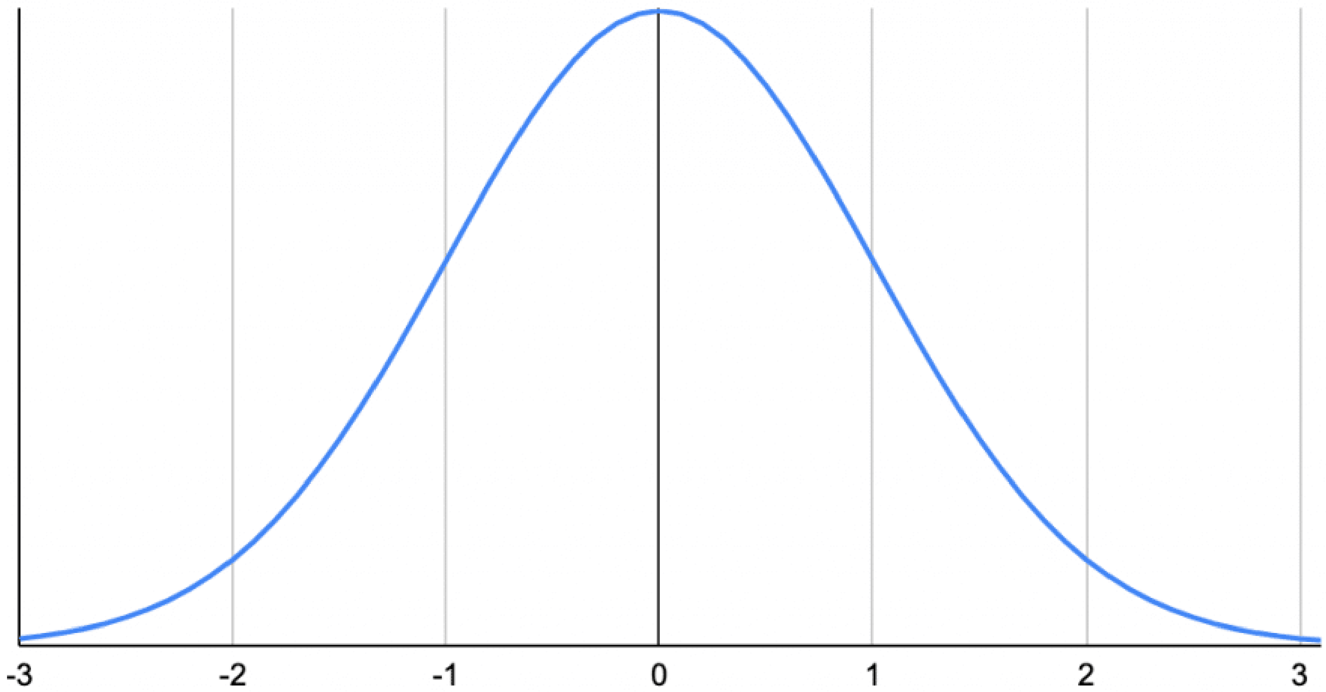


Normal Distribution (20/41)

Standard Normal Random Variable (Z):

$$\mu = 0 \quad \sigma^2 = 1$$

The distribution of a normal random variable with **mean 0** and **variance 1** is called a ***standard normal distribution***.



Normal Distribution (21/41)

Density function of standard normal random variable (Z):

$$\boldsymbol{Z \sim N(0,1)}$$

μ : mean = 0

σ^2: variance = 1

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

Normal Distribution (22/41)

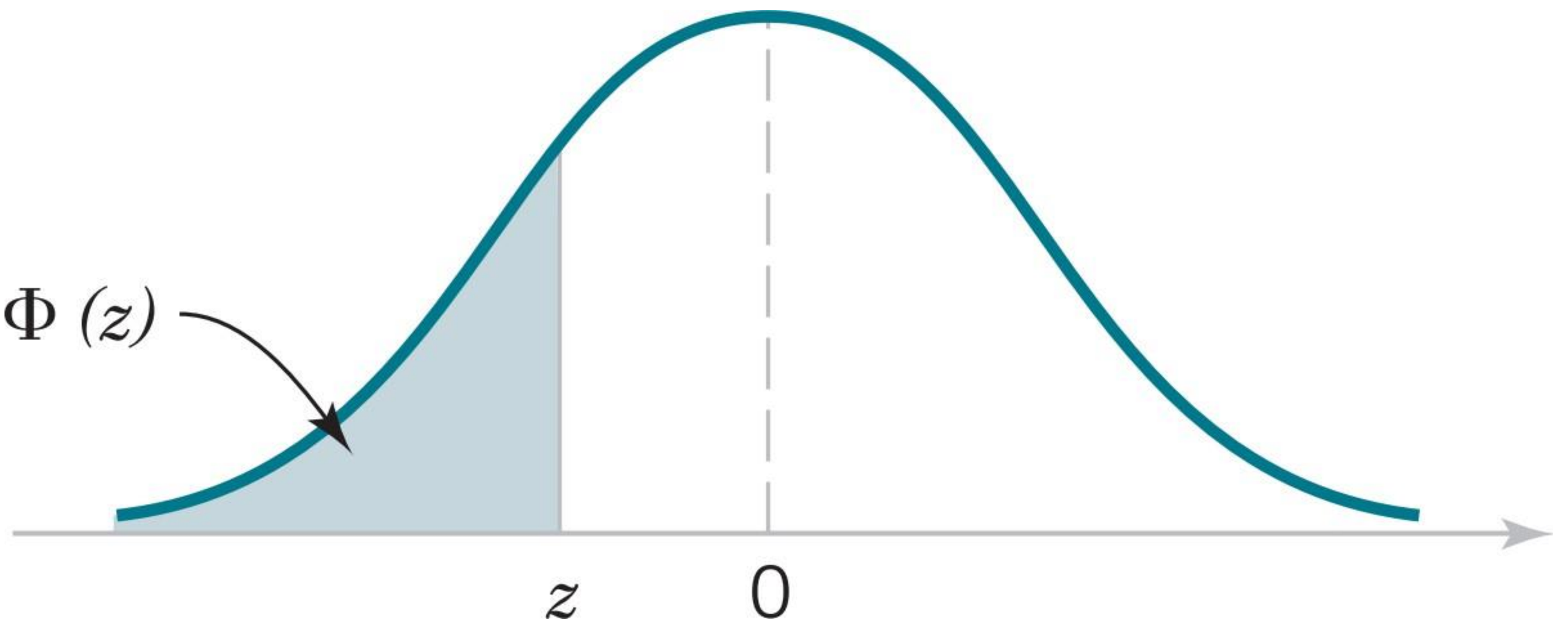
Cumulative distribution function of standard normal random variable $(\boldsymbol{Z})$:

$$\boldsymbol{Z} \sim \boldsymbol{N}(\mathbf{0}, \mathbf{1})$$

μ : mean = 0

σ^2: variance = 1

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} \, dv$$



Normal Distribution (22/41)

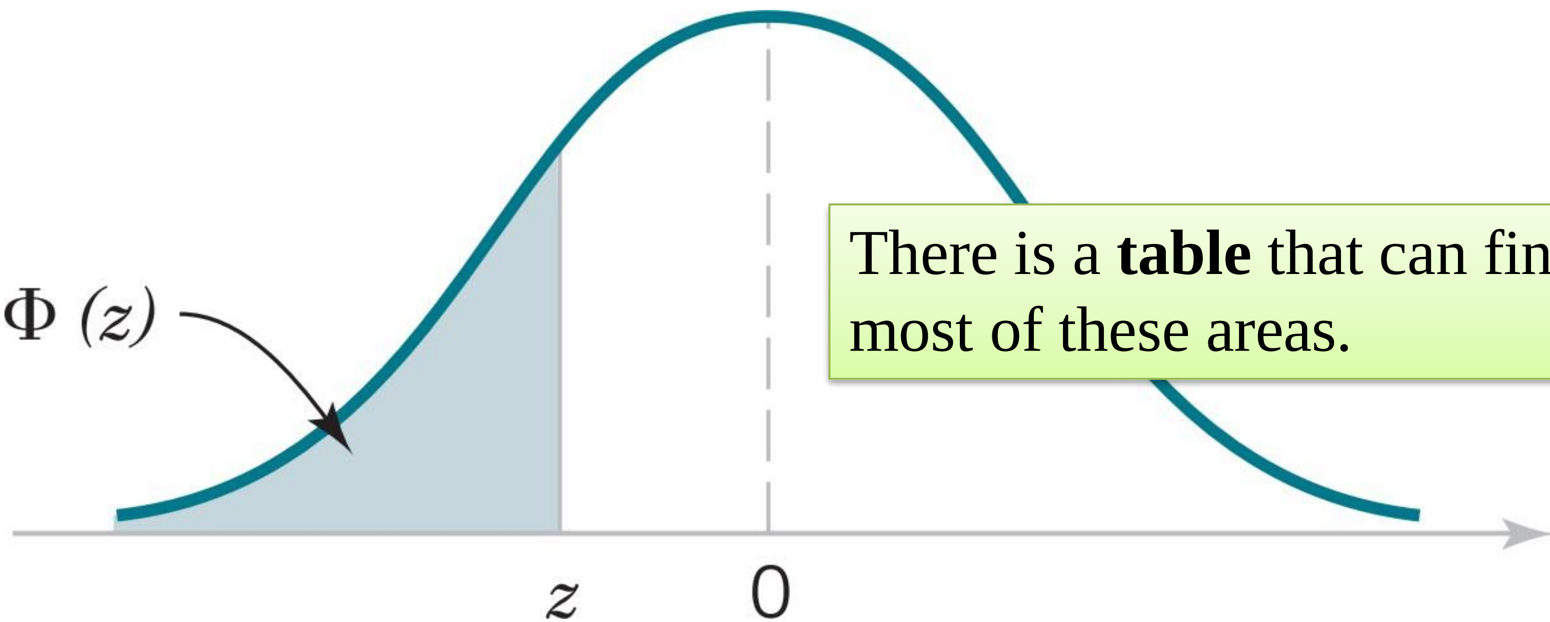
Cumulative distribution function of standard normal random variable (Z):

$$\boldsymbol{Z \sim N(0, 1)}$$

μ : mean = 0

σ^2: variance = 1

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} \, dv$$

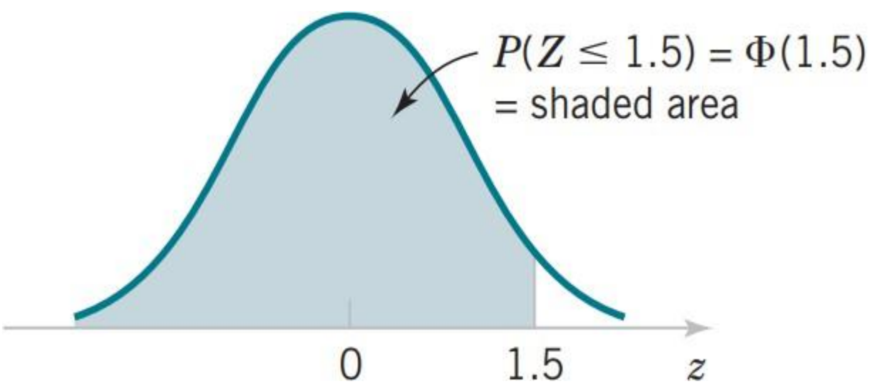


There is a **table** that can find most of these areas.

Example2 (1/2):

Assume that Z is a standard normal random variable. The Table provides probabilities of the form $\Phi(z) = P(Z \leq z)$.

The use of Table to find $P(Z \leq 1.5)$ is illustrated as follows:

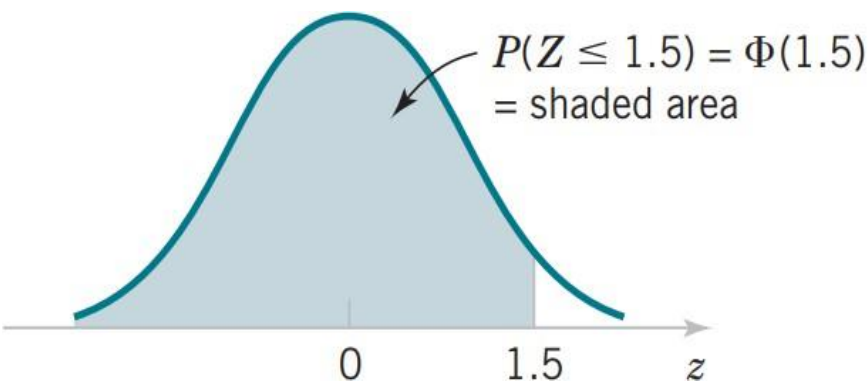


z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50800	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Example2 (2/2):

Assume that Z is a standard normal random variable. The Table provides probabilities of the form $\Phi(z) = P(Z \leq z)$.

The probability is 0.93319.



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50800	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Normal Distribution (23/41)

$\Phi(-2.35)$

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Normal Distribution (23/41)

$$\Phi(-2.35) = P(Z \leq -2.35)$$

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Normal Distribution (23/41)

$\Phi(-2.35) = P(Z \leq -2.35)$

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Normal Distribution (23/41)

$\Phi(-2.35) = P(Z \leq -2.35)$

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Normal Distribution (23/41)

$$\Phi(-2.35) = P(Z \leq -2.35) = 0.0094$$

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Normal Distribution (23/41)

Table A.3 (continued) Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Normal Distribution (23/41)

$$\Phi(1.27) = P(Z \leq 1.27)$$

Table A.3 (continued) Areas under the Norm

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Normal Distribution (23/41)

$$\Phi(1.27) = P(Z \le 1.27) = 0.8980$$

Table A.3 (continued) Areas under the Norm

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Normal Distribution (23/41)

TABLE III Cumulative Standard Normal Distribution

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
−3.1	0.000711	0.000736	0.000762	0.000789	0.000816	0.000845	0.000874	0.000904	0.000935	0.000968
−3.0	0.001001	0.001035	0.001070	0.001107	0.001144	0.001183	0.001223	0.001264	0.001306	0.001350
−2.9	0.001395	0.001441	0.001489	0.001538	0.001589	0.001641	0.001695	0.001750	0.001807	0.001866
−2.8	0.001926	0.001988	0.002052	0.002118	0.002186	0.002256	0.002327	0.002401	0.002477	0.002555
−2.7	0.002635	0.002718	0.002803	0.002890	0.002980	0.003072	0.003167	0.003264	0.003364	0.003467
−2.6	0.003573	0.003681	0.003793	0.003907	0.004025	0.004145	0.004269	0.004396	0.004527	0.004661
−2.5	0.004799	0.004940	0.005085	0.005234	0.005386	0.005543	0.005703	0.005868	0.006037	0.006210
−2.4	0.006387	0.006569	0.006756	0.006947	0.007143	0.007344	0.007549	0.007760	0.007976	0.008198
−2.3	0.008424	0.008656	0.008894	0.009137	0.009387	0.009642	0.009903	0.010170	0.010444	0.010724
−2.2	0.011011	0.011304	0.011604	0.011911	0.012224	0.012545	0.012874	0.013209	0.013553	0.013903
−2.1	0.014262	0.014629	0.015003	0.015386	0.015778	0.016177	0.016586	0.017003	0.017429	0.017864
−2.0	0.018309	0.018763	0.019226	0.019699	0.020182	0.020675	0.021178	0.021692	0.022216	0.022750
−1.9	0.023295	0.023852	0.024419	0.024998	0.025588	0.026190	0.026803	0.027429	0.028067	0.028717

Normal Distribution (24/41)

Some Properties:

$$\Phi(a) = P(Z \leq a) \quad \rightarrow \quad \text{From Table}$$

$$P(Z < -a) = 1 - P(Z \leq a)$$

$$P(a < Z < b) = P(Z < b) - P(Z < a)$$

$$P(Z > a) = 1 - P(Z \leq a)$$

$$P(Z > -a) = P(Z < a)$$

Example3 (1/16):

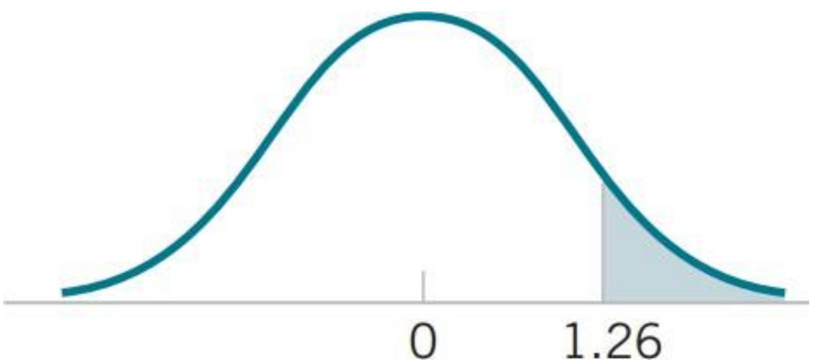
Assume that Z is a standard normal random variable. Find the following probabilities:

1. $P(Z > 1.26)$
2. $P(Z < -0.86)$
3. $P(Z > -1.37)$
4. $P(-1.25 < Z < 0.37)$
5. $P(Z \leq -4.6)$

Example3 (2/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

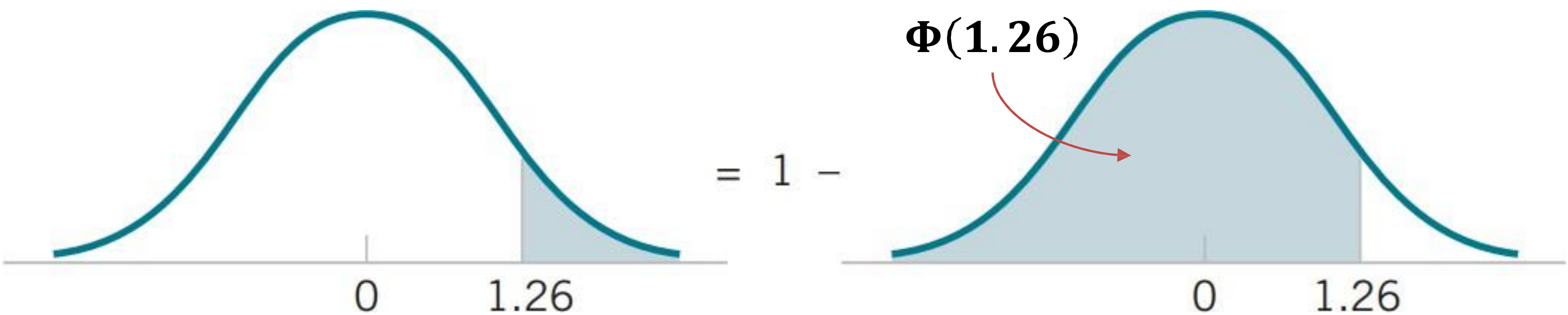
1. $P(Z > 1.26)$



Example3 (3/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

1. $P(Z > 1.26) = 1 - \Phi(1.26)$



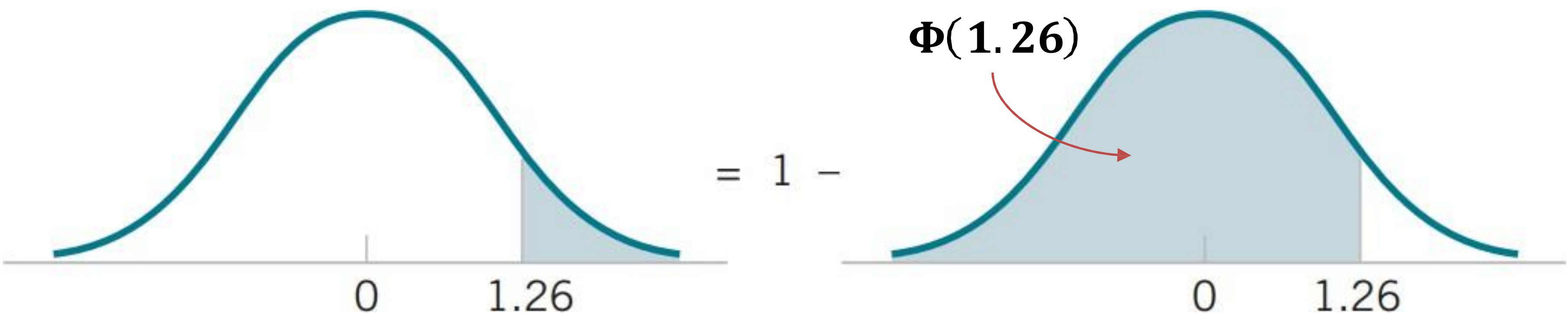
Normal Distribution (25/41)

Example3 (4/16):

Assume that Z is a standard normal random variable. Find the following probabilities

$\Phi(1.26) = P(Z \leq 1.26)$
From the Table

1. $P(Z > 1.26) = 1 - \Phi(1.26)$



Normal Distribution (25/41)

Example3 (5/16):

$$\Phi(1.26) = 0.896165$$

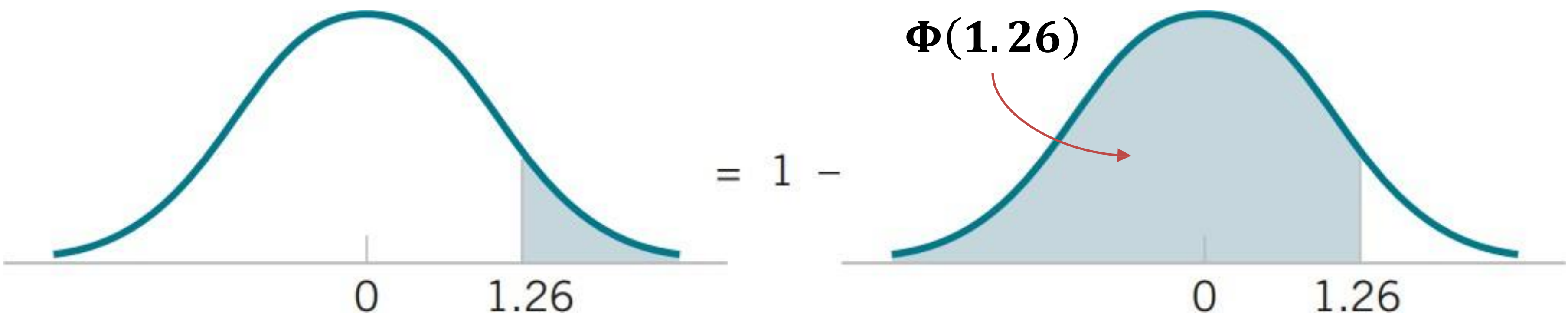
TABLE III **Cumulative Standard Normal Distribution** (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083

Example3 (6/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

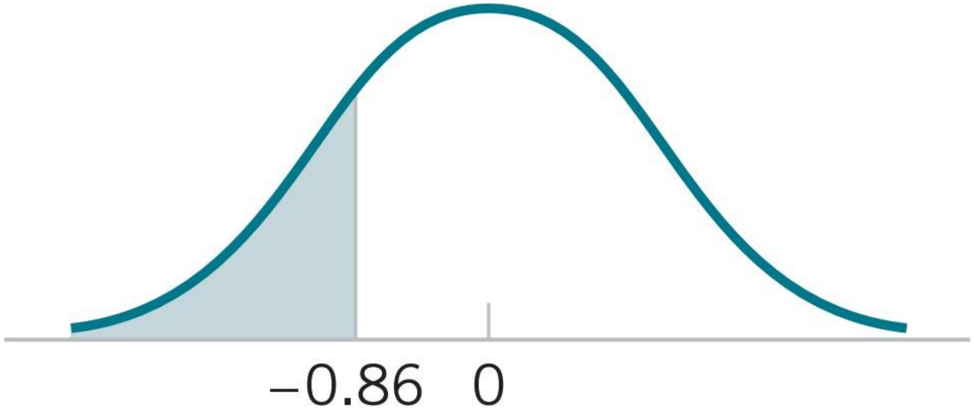
1. $P(Z > 1.26) = 1 - 0.896165 = 0.103835$



Example3 (7/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

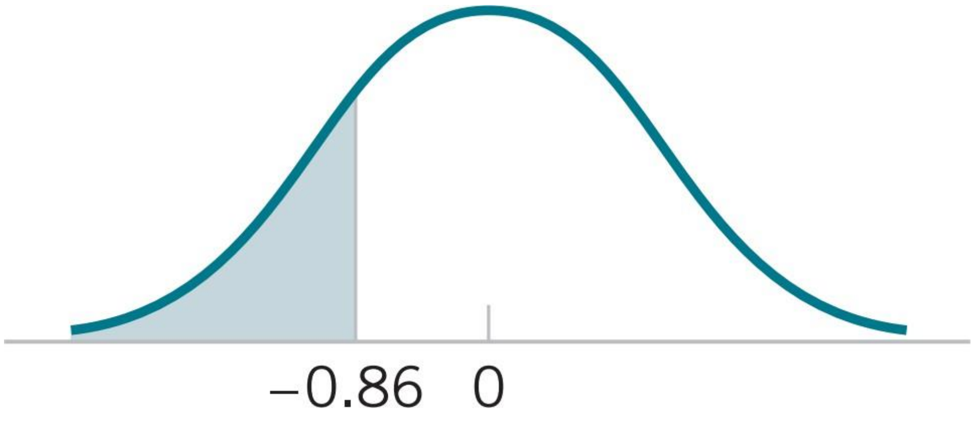
2. $P(Z < -0.86)$



Example3 (8/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

2. $P(Z < -0.86) = \Phi(-0.86) = 0.19490$



Normal Distribution (25/41)

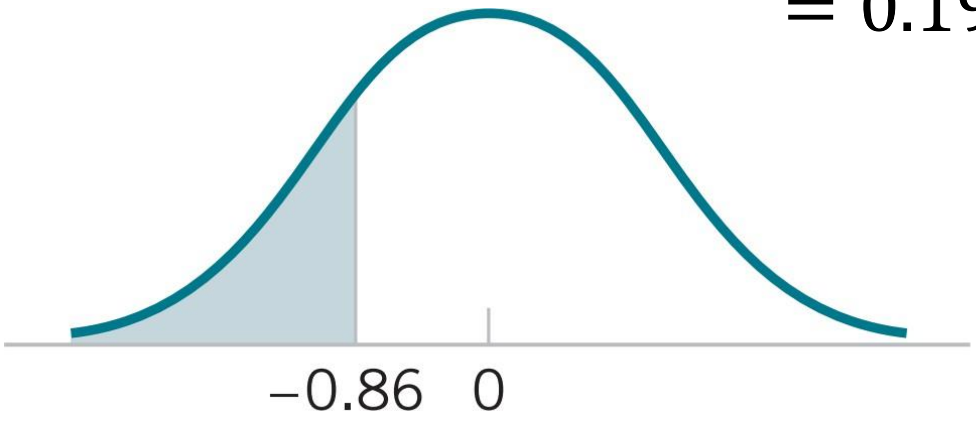
Example3 (9/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

2. $P(Z < -0.86) = \Phi(-0.86) = 0.19490$

or $P(Z < -0.86) = 1 - P(Z < 0.86) = 1 - \Phi(0.86)$

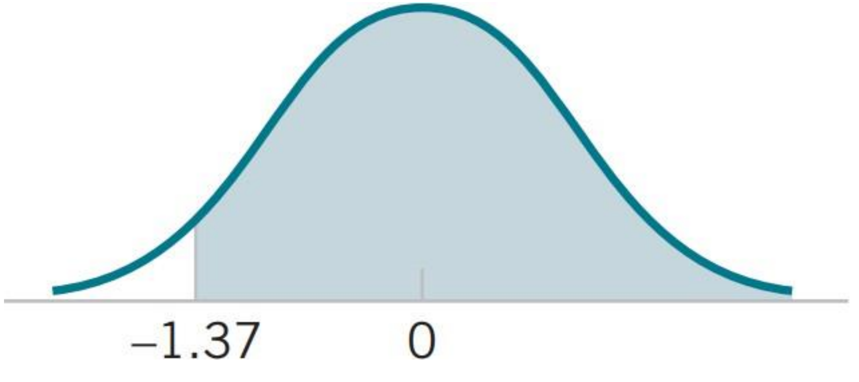
$= 0.19490$



Example3 (10/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

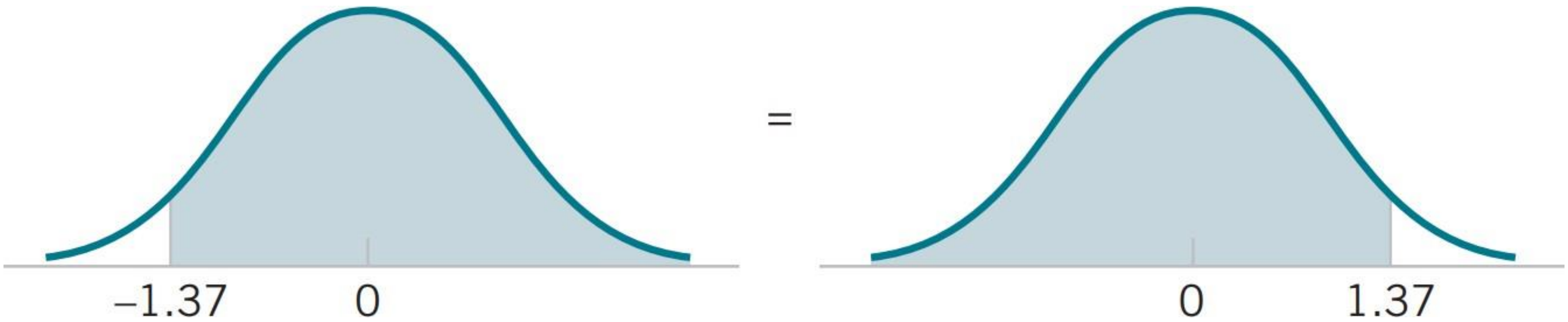
3. $P(Z > -1.37)$



Example3 (11/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

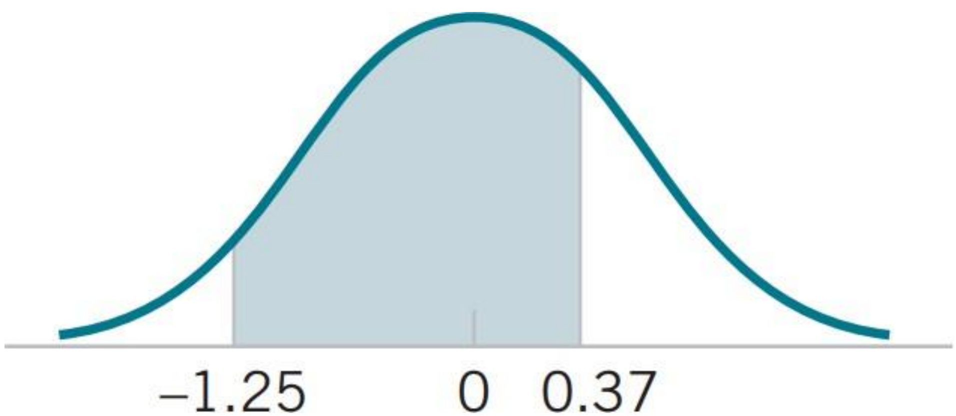
3. $P(Z > -1.37) = P(Z < 1.37) = 0.91465$



Example3 (12/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

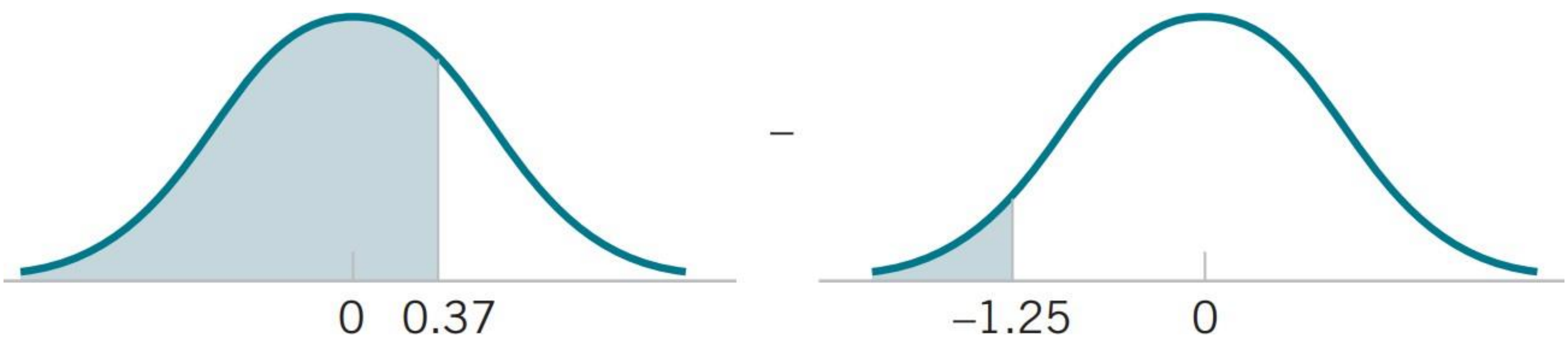
4. $P(-1.25 < Z < 0.37)$



Example3 (13/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

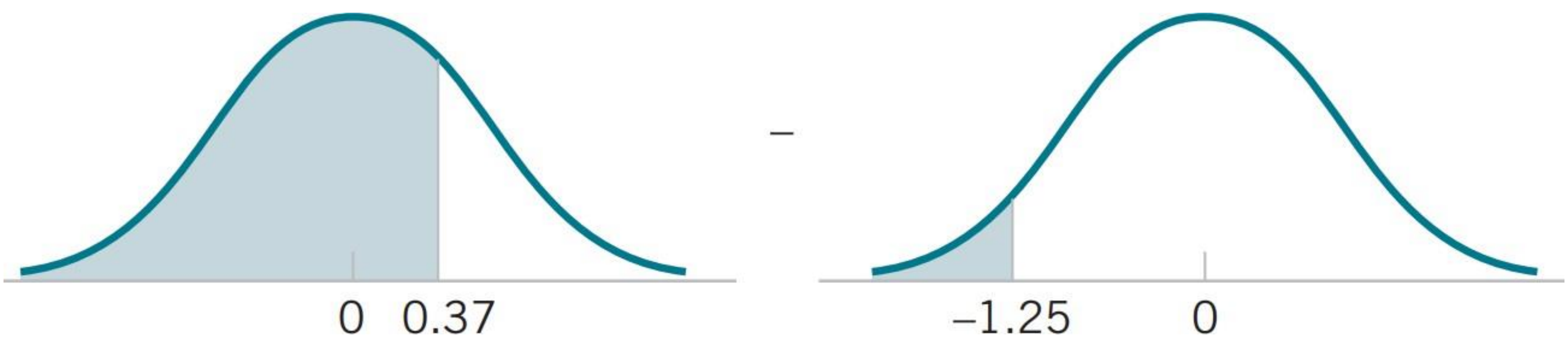
4. $P(-1.25 < Z < 0.37) = \Phi(0.37) - \Phi(-1.25)$



Example3 (14/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

4. $$P(-1.25 < Z < 0.37) = \Phi(0.37) - \Phi(-1.25)$$
$$= 0.64431 \; - \; 0.10565$$
$$= 0.53866$$

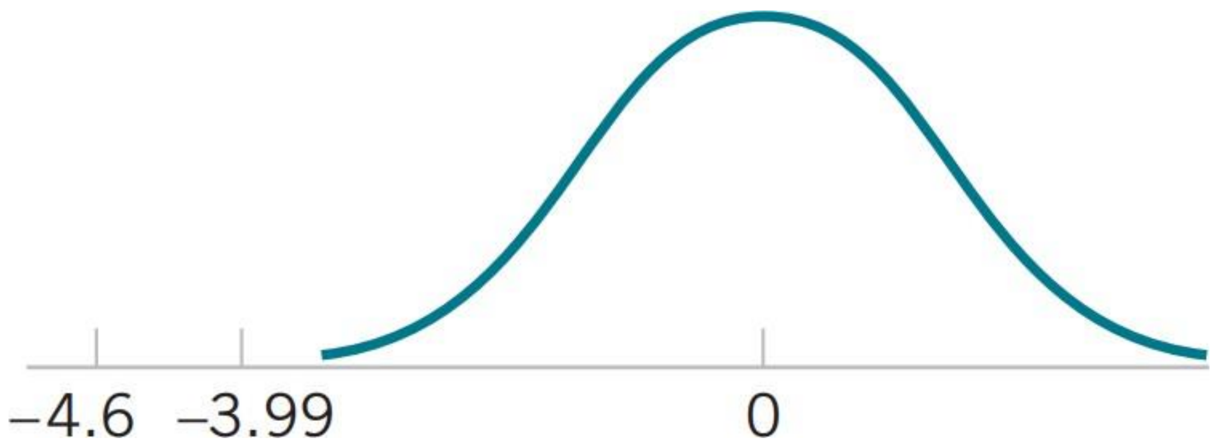


Example3 (15/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

5. $P(Z \leq -4.6)$ is nearly zero

The last entry in the table can be used to find that $P(Z \leq -3.99) = 0.00003$.



Normal Distribution (25/41)

Example3 (16/16):

Assume that Z is a standard normal random variable. Find the following probabilities:

5. $P(Z \leq -4.6)$ is nearly zero

The last entry in the table can be used to find that $P(Z \leq -3.99) = 0.00003$.

TABLE III Cumulative Standard Normal Distribution

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108

Example4 (1/8):

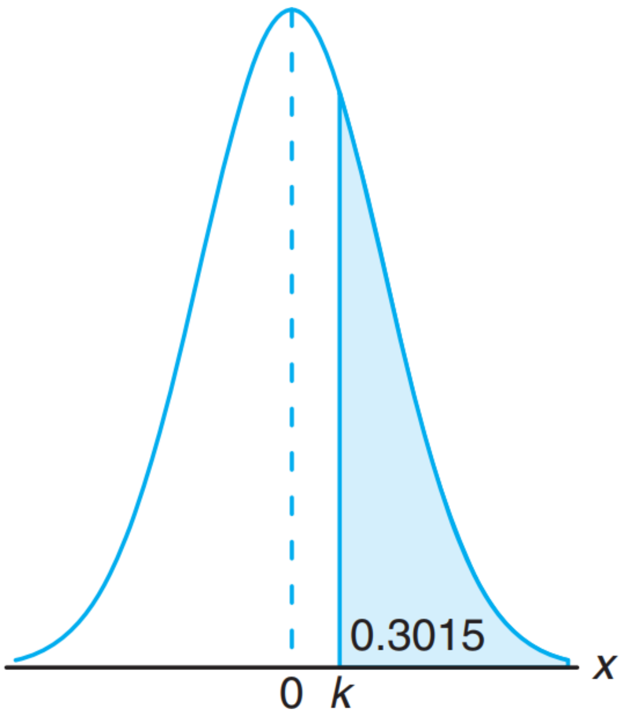
Given a standard normal distribution Z, find the value of k such that

(a) $P(Z > k) = 0.3015$ and

(b) $P(k < Z < -0.18) = 0.4197$

Example4 (2/8):

(a) $P(Z > k) = 0.3015$

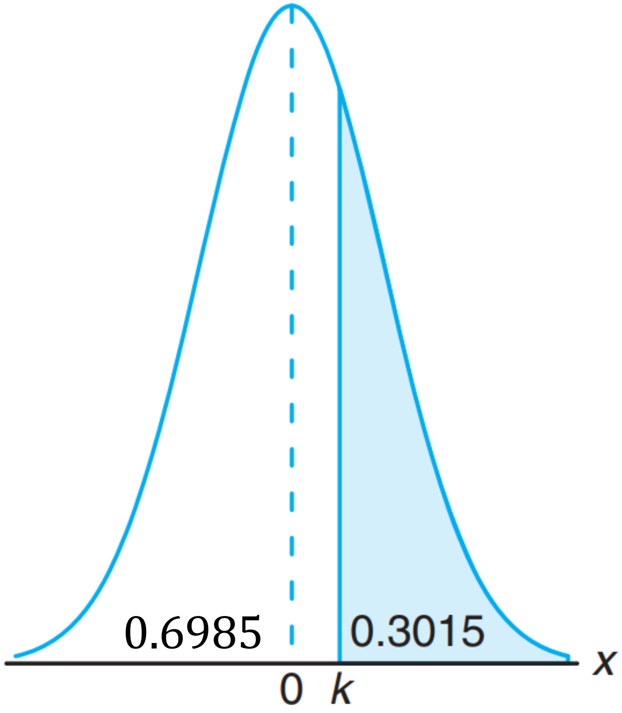


Example4 (3/8):

(a) $P(Z > k) = 0.3015$

$P(Z < k) = 1 - 0.3015 = 0.6985$

From the Table



Normal Distribution (26/41)

Table A.3 (continued) Areas under the Normal Curve

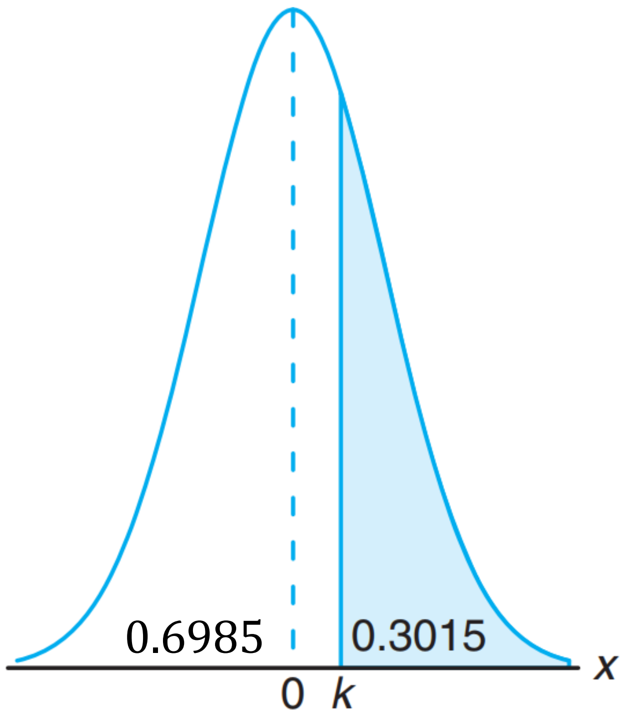
z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	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	0.8238	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

Example4 (4/8):

(a) $P(Z > k) = 0.3015$

$P(Z < k) = 1 - 0.3015 = 0.6985$

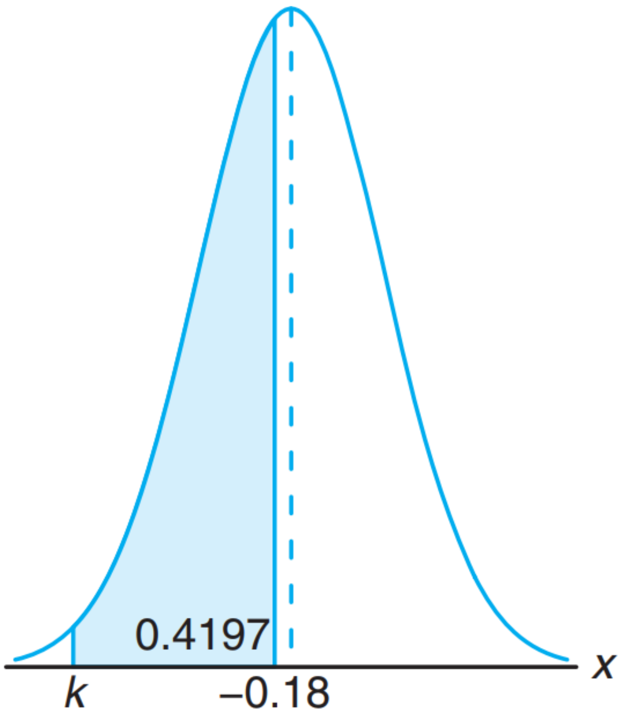
From the Table
it follows that $k = 0.52$.



Normal Distribution (26/41)

Example4 (5/8):

(b) $P(k < Z < -0.18) = 0.4197$

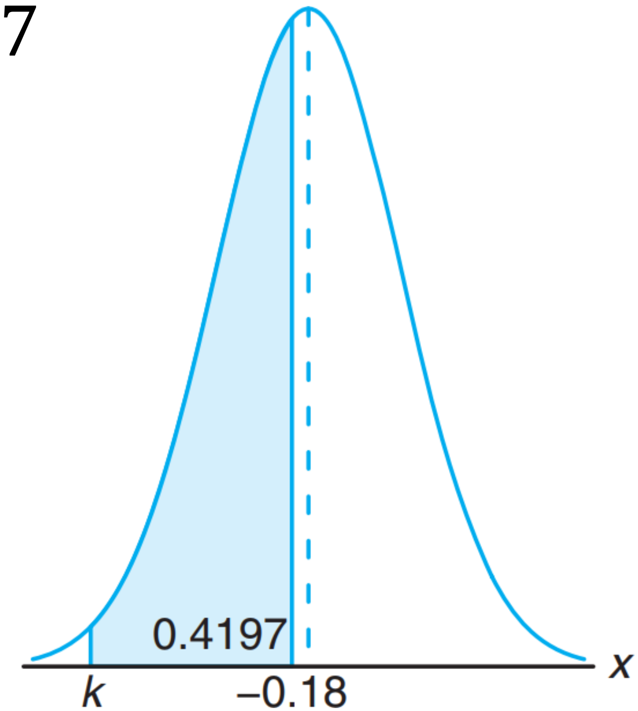


Example4 (6/8):

(b) $P(k < Z < -0.18) = 0.4197$

$P(Z < -0.18) - P(Z < k) = 0.4197$

From the Table
$P(Z < -0.18) = 0.4286.$



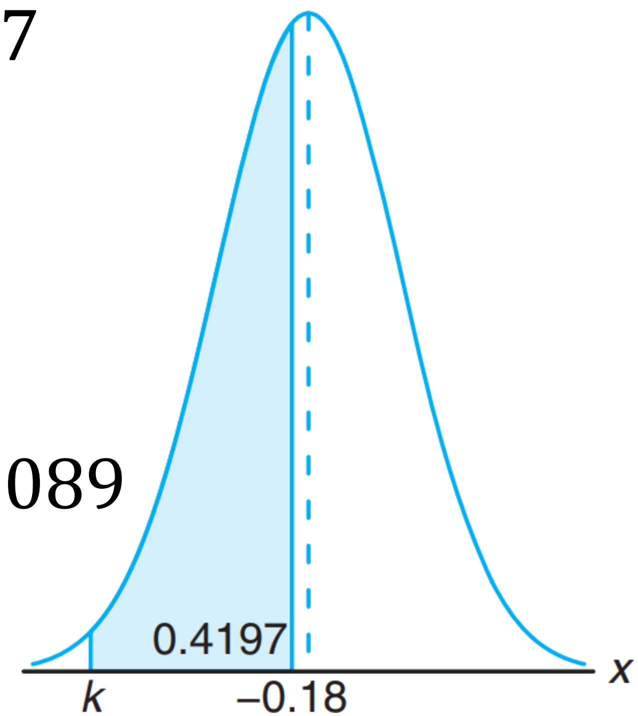
Example4 (7/8):

(b) $P(k < Z < -0.18) = 0.4197$

$P(Z < -0.18) - P(Z < k) = 0.4197$

From the Table
$P(Z < -0.18) = 0.4286.$

$\therefore\ 0.4286 - P(Z < k) = 0.4197$
$P(Z < k) = 0.4286 - 0.4197 = 0.0089$



Normal Distribution (26/41)

Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146

Example4 (8/8):

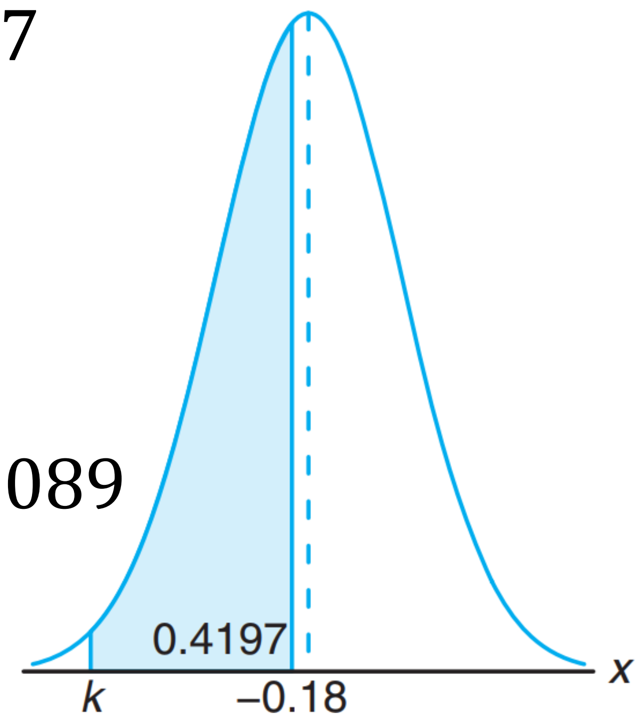
(b) $P(k < Z < -0.18) = 0.4197$

$P(Z < -0.18) - P(Z < k) = 0.4197$

From the Table
$P(Z < -0.18) = 0.4286.$

$\therefore 0.4286 - P(Z < k) = 0.4197$
$P(Z < k) = 0.4286 - 0.4197 = 0.0089$

From the Table
it follows that $k = -2.37$.



Example5 (1/6):

Given a standard normal distribution Z, find the value of k such that

(a) $P(Z > k) = 0.05$ and

(b) $P(-k < Z < k) = 0.99$.

Example5 (2/6):

(a) $P(Z > k) = 0.05$

$P(Z > k) = 1 - P(Z < k) = 0.05$
$P(Z < k) = 0.95$

Normal Distribution (27/41)

TABLE III Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273

Normal Distribution (27/41)

TABLE III Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273

Example5 (3/6):

(a) $P(Z > k) = 0.05$

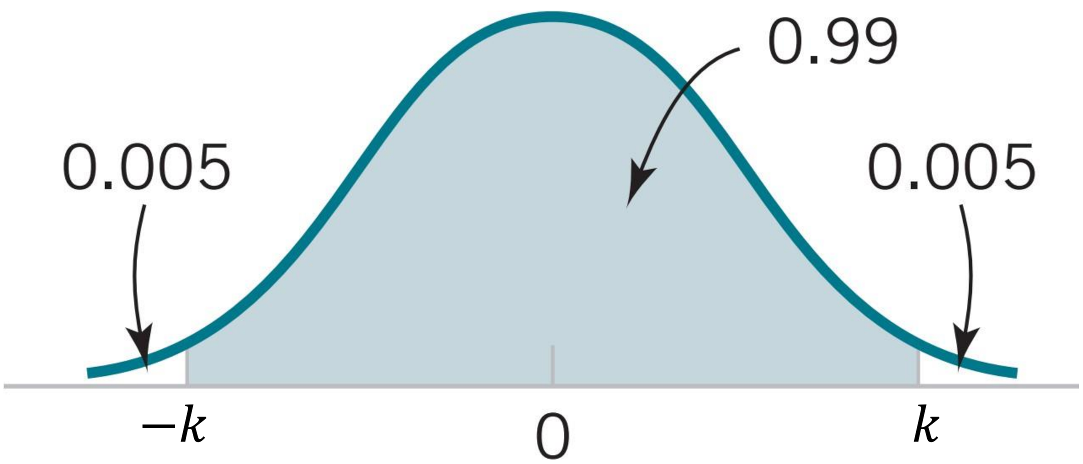
$P(Z > k) = 1 - P(Z < k) = 0.05$
$P(Z < k) = 0.95$

From the Table
We do not find 0.95 exactly; the nearest value is 0.95053, corresponding to $k \cong 1.65$.

Example5 (4/6):

(b) $P(-k < Z < k) = 0.99.$

$$P(Z < k) - P(Z < -k) = 0.99$$

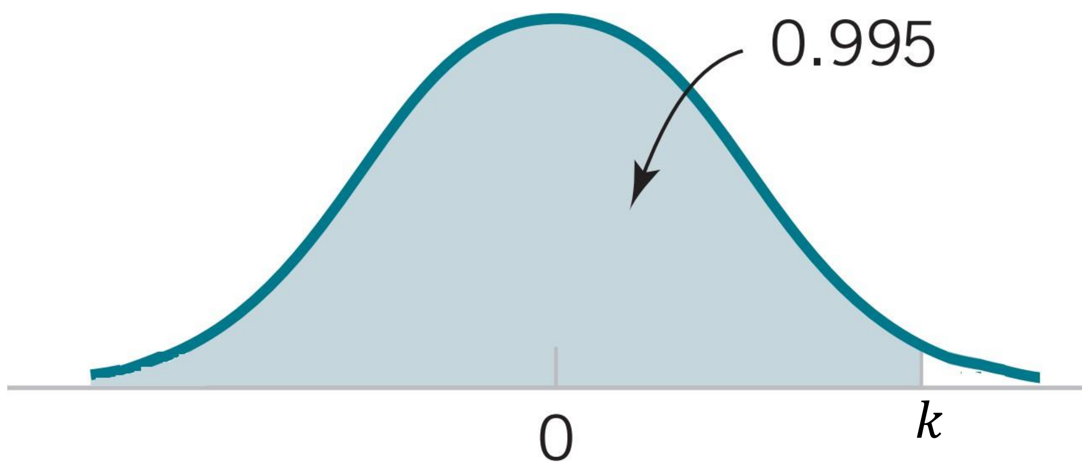


Example5 (5/6):

(b) $P(-k < Z < k) = 0.99.$

$$P(Z < k) - P(Z < -k) = 0.99$$

$$P(Z < k) = 0.995$$



Normal Distribution (27/41)

TABLE III Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365

Normal Distribution (27/41)

TABLE III **Cumulative Standard Normal Distribution (*continued*)**

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365

Example5 (6/6):

(b) $P(-k < Z < k) = 0.99.$

$$P(Z < k) - P(Z < -k) = 0.99$$

$$P(Z < k) = 0.995$$

From the Table
We do not find 0.995 exactly; the nearest value is 0.99506, corresponding to $k \cong 2.58$.

Normal Distribution (28/41)

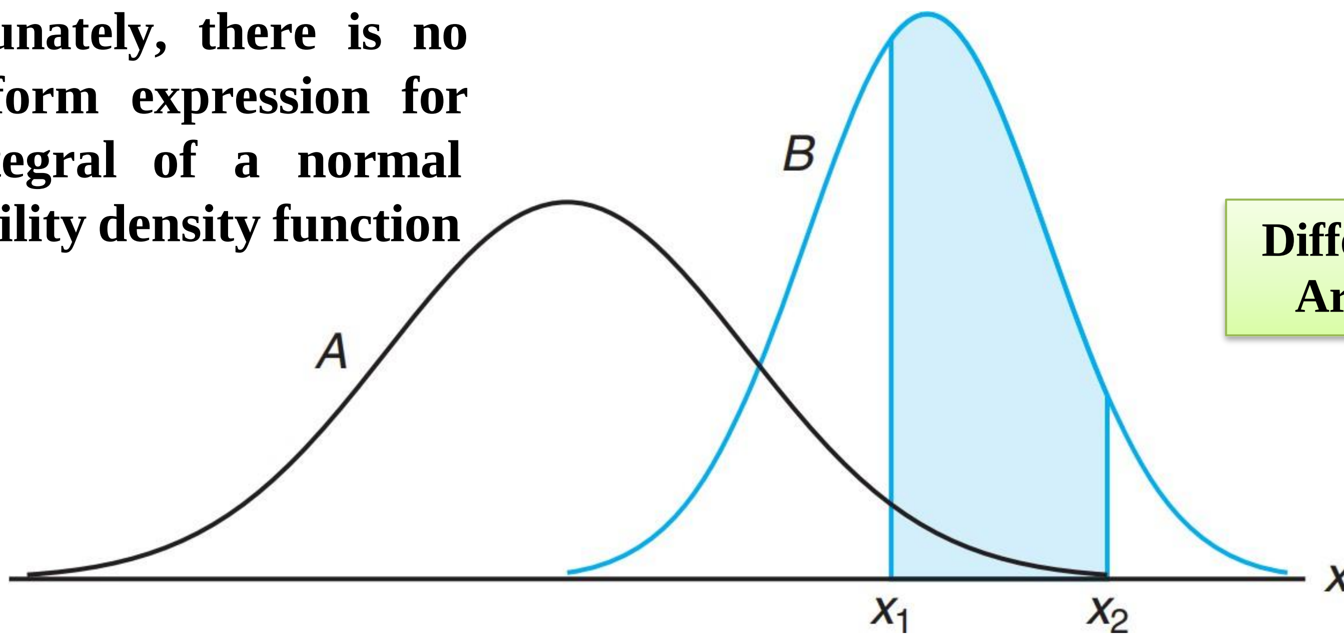
Recall

Areas under the Normal Curve:

Different means and standard deviations

$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx$$

Unfortunately, there is no closed-form expression for the integral of a normal probability density function



Different Areas

Normal Distribution (29/41)

Standardizing a Normal Random Variable:

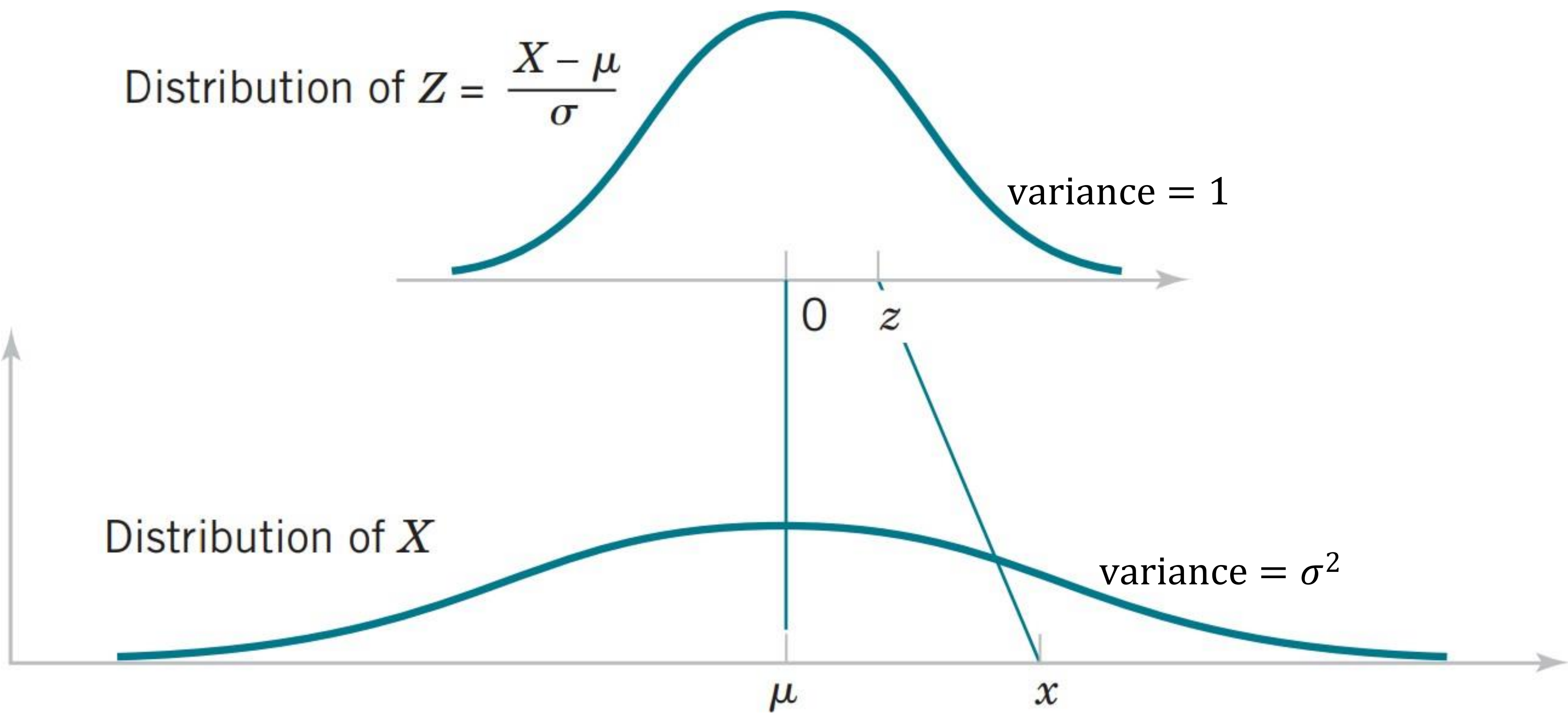
If X is a normal random variable with mean μ and variance σ^2, the random variable

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

is a normal random variable with mean = 0 and variance = 1. That is, Z is a standard normal random variable.

Normal Distribution (30/41)

$$\boldsymbol{X \sim N(\mu, \sigma^2) \implies Z = \frac{X-\mu}{\sigma} \sim N(0, 1)}$$



Normal Distribution (31/41)

The relationship between the values of X and the transformed values of Z:

$$\boldsymbol{X \sim N(\mu, \sigma^2) \implies Z = \frac{X-\mu}{\sigma} \sim N(0, 1)}$$

<u>The value z is</u>

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

<u>And the value x is</u>

$$x = \sigma z + \mu$$

Example1 (1/7):

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement exceeds 13 milliamperes?

Example1 (2/7):

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement exceeds 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$.

$\mu = 10$ and $\sigma^2 = 4 \quad \rightarrow \quad \sigma = 2$

Example1 (3/7):

$P(X > 13)$

$\mu = 10$ and $\sigma^2 = 4 \quad \rightarrow \quad \sigma = 2$

$$P(X > 13) = P\left(Z > \frac{13 - \mu}{\sigma}\right) = P\left(Z > \frac{13 - 10}{2}\right)$$

$$= P(Z > 1.5)$$

Example1 (4/7):

$$P(X > 13) = P(Z > 1.5)$$

$$= 1 - P(Z \leq 1.5)$$

Normal Distribution (32/41)

Example1 (5/7):

$$P(X > 13) = P(Z > 1.5)$$

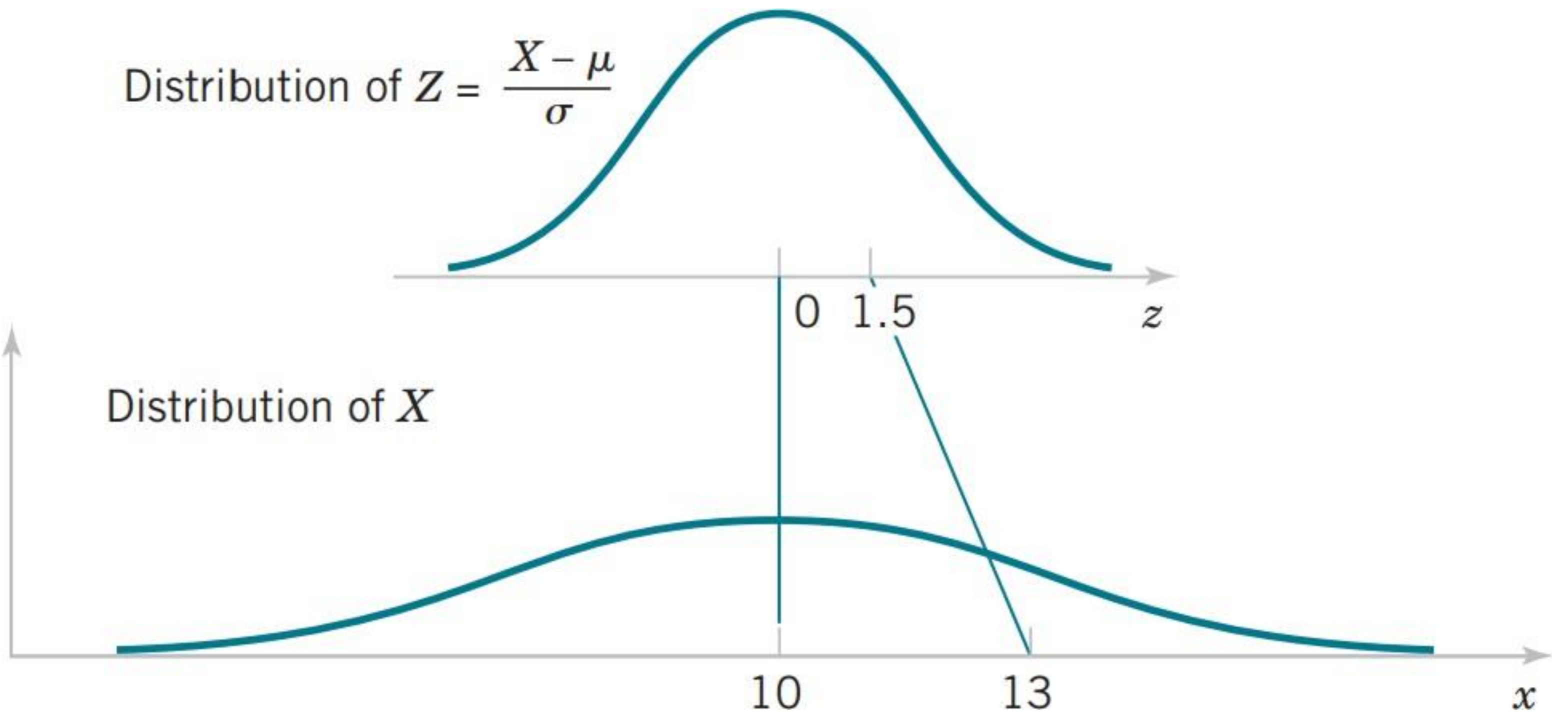
$$= 1 - \boxed{P(Z \leq 1.5)}$$

From the Table $P(Z \leq 1.5) = 0.93319$

$$\therefore P(X > 13) = 1 - 0.93319 = 0.06681$$

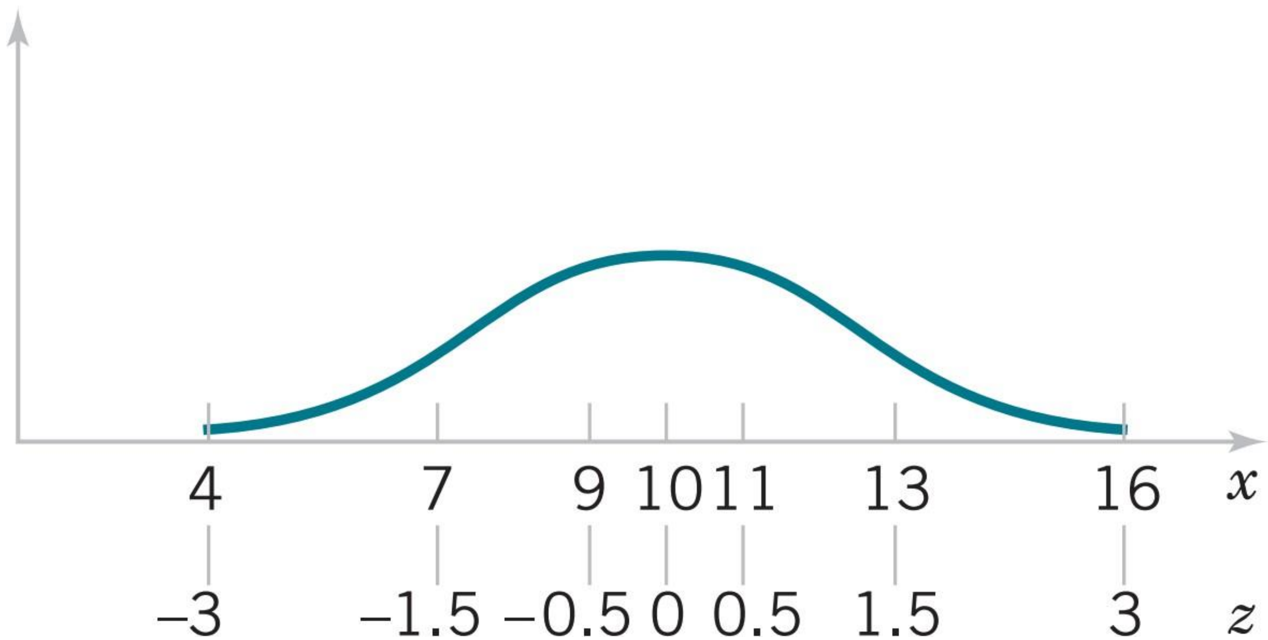
Example1 (6/7):

$$P(X > 13) = P(Z > 1.5) = 0.06681$$



Example1 (7/7):

$$P(X > 13) = P(Z > 1.5) = 0.06681$$



Example2 (1/6):

Continuing Example 1, what is the probability that a current measurement is between 9 and 11 milliamperes?

$$P\,(9 < X < 11)$$

Then, determine the value for which the probability that a current measurement is less than this value is 0.98.

$$P(X < x) = 0.98$$

Example2 (2/6):

Continuing Example 1, what is the probability that a current measurement is between 9 and 11 milliamperes?

$$P(9 < X < 11) = P\left(\frac{9-10}{2} < \frac{X-10}{2} < \frac{11-10}{2}\right)$$

Example2 (3/6):

Continuing Example 1, what is the probability that a current measurement is between 9 and 11 milliamperes?

$$
\begin{aligned}
P(9 < X < 11) &= P\left(\frac{9-10}{2} < \frac{X-10}{2} < \frac{11-10}{2}\right) \\
&= P(-0.5 < Z < 0.5) \\
&= P(Z < 0.5) - P(Z < -0.5) \\
&= 0.69146 - 0.30854 = 0.38292
\end{aligned}
$$

Example2 (4/6):

Determine the value for which the probability that a current measurement is less than this value is 0.98.

$$P(X < x) = 0.98$$

$$P(X < x) = P\left(\frac{X - 10}{2} < \frac{x - 10}{2}\right)$$
$$= P\left(Z < \frac{x - 10}{2}\right) = 0.98$$

Example2 (5/6):

$$P(X < x) = P\left(\frac{X-10}{2} < \frac{x-10}{2}\right)$$
$$= P\left(Z < \frac{x-10}{2}\right) = 0.98$$

Table is used to find the z-value such that $P(Z < z) = 0.98$. The nearest probability from the Table results in $P(Z < 2.06) = 0.980301$.

Example2 (6/6):

$$P(X < x) = P\left(\frac{X - 10}{2} < \frac{x - 10}{2}\right)$$
$$= P\left(Z < \frac{x - 10}{2}\right) = 0.98$$

Table is used to find the z-value such that $P\,(Z < z) = 0.98$. The nearest probability from the Table results in $P(Z < 2.06) = 0.980301$.

$\because\ x = \sigma z + \mu$

$\therefore\ x = (2)(2.06) + 10 = 14.1\ \mathrm{mA}$

Example3 (1/8):

The loaves of rye bread distributed to local stores by a certain bakery have an average length of 30 centimeters and a standard deviation of 2 centimeters. Assuming that the lengths are normally distributed,

what percentage of the loaves are

(a) longer than 31.7 centimeters?

(b) between 29.3 and 33.5 centimeters in length?

Normal Distribution (34/41)

Example3 (2/8):

$$X \sim N(\mu, \sigma^2) \implies Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

The loaves of rye bread distributed to local stores by a certain bakery have an **average length of 30** centimeters and a **standard deviation of 2** centimeters. Assuming that the lengths are *normally distributed*, what percentage of the loaves are

(a) longer than 31.7 centimeters?
(b) between 29.3 and 33.5 centimeters in length?

Normal Distribution (34/41)

Example3 (3/8):

$$\boldsymbol{X \sim N(30, 2^2) \implies Z = \frac{X-30}{2} \sim N(0, 1)}$$

The loaves of rye bread distributed to local stores by a certain bakery have an **average length of 30** centimeters and a **standard deviation of 2** centimeters. Assuming that the lengths are *normally distributed*, what percentage of the loaves are

(a) longer than 31.7 centimeters?
(b) between 29.3 and 33.5 centimeters in length?

Normal Distribution (34/41)

Example3 (4/8):

$$\boldsymbol{X \sim N\left(30, 2^2\right) \implies Z = \frac{X-30}{2} \sim N(0,1)}$$

(a) longer than 31.7 centimeters?

$$P(X > 31.7) = P\left(Z > \frac{31.7-\mu}{\sigma}\right) = P\left(Z > \frac{31.7-30}{2}\right)$$

$$P(Z > 0.85) = 1 - P(Z \leq 0.85)$$

Normal Distribution (34/41)

$$P(Z \leq 0.85) = 0.8023$$

Table A.3 (continued) Areas under the Normal Curve

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	0.8238	0.8264	0.8289	0.8315

Normal Distribution (34/41)

Example3 (5/8):

$$\boldsymbol{X \sim N\left(30, 2^2\right) \implies Z = \frac{X-30}{2} \sim N(0, 1)}$$

(a) longer than 31.7 centimeters?

$$P(X > 31.7) = P\left(Z > \frac{31.7-\mu}{\sigma}\right) = P\left(Z > \frac{31.7-30}{2}\right)$$

$$P(Z > 0.85) = 1 - P(Z \leq 0.85) = 1 - 0.8023 = 0.1977$$

Then, 19.77% of the loaves are longer than 31.7 centimeters

Normal Distribution (34/41)

Example3 (6/8):

$$\boldsymbol{X \sim N\left(30, 2^2\right) \Longrightarrow Z = \frac{X-30}{2} \sim N(0,1)}$$

(b) between 29.3 and 33.5 centimeters in length?

$P(29.3 < X < 33.5)$

Normal Distribution (34/41)

Example3 (7/8):

$$\boldsymbol{X \sim N(30, 2^2) \implies Z = \frac{X-30}{2} \sim N(0, 1)}$$

(b) between 29.3 and 33.5 centimeters in length?

$$P(29.3 < X < 33.5) = P\left(\frac{29.3 - 30}{2} < Z < \frac{33.5 - 30}{2}\right)$$

$$= P(-0.35 < Z < 1.75)$$

Normal Distribution (34/41)

Example3 (8/8):

$$\boldsymbol{X \sim N\left(30, 2^2\right) \Rightarrow Z = \frac{X-30}{2} \sim N(0,1)}$$

(b) between 29.3 and 33.5 centimeters in length?

$$P(29.3 < X < 33.5) = P\left(\frac{29.3-30}{2} < Z < \frac{33.5-30}{2}\right)$$

$$= P(-0.35 < Z < 1.75)$$

$$P(-0.35 < Z < 1.75) = P(Z < 1.75) - P(Z < -0.35)$$

$$P(-0.35 < Z < 1.75) = 0.9599 - 0.3632 = 0.5967$$

Then, 59.67% of the loaves are between 29.3 and 33.5 centimeters.

Example5 (1/8):

The average grade for an exam is 74, and the standard deviation is 7. If 12% of the class is given *A*s, and the grades are curved to follow a normal distribution, what is the lowest possible integer grade for *A* and the highest possible integer grade for *B*?

Normal Distribution (35/41)

Example5 (2/8):

$$\boldsymbol{X \sim N(74, 7^2) \Rightarrow Z = \frac{X-74}{7} \sim N(0, 1)}$$

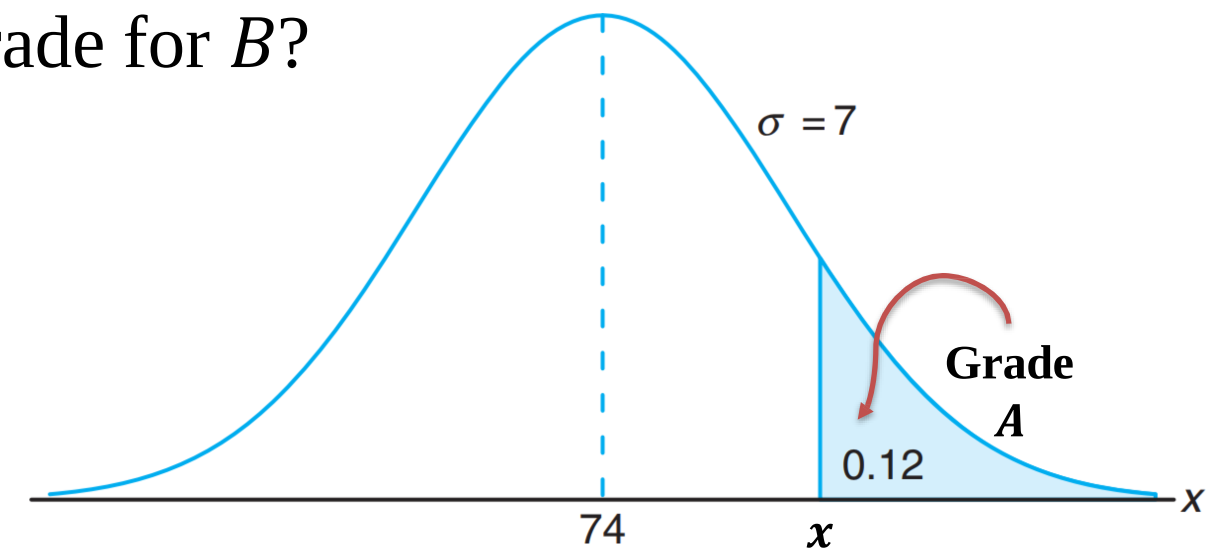
The **average grade for an exam is 74**, and the **standard deviation is 7**. If 12% of the class is given As, and the grades are curved to follow a normal distribution, what is the lowest possible integer grade for A and the highest possible integer grade for B?

Normal Distribution (35/41)

Example5 (3/8):

$$\boldsymbol{X \sim N\left(74, 7^2\right) \Rightarrow Z = \frac{X-74}{7} \sim N(0,1)}$$

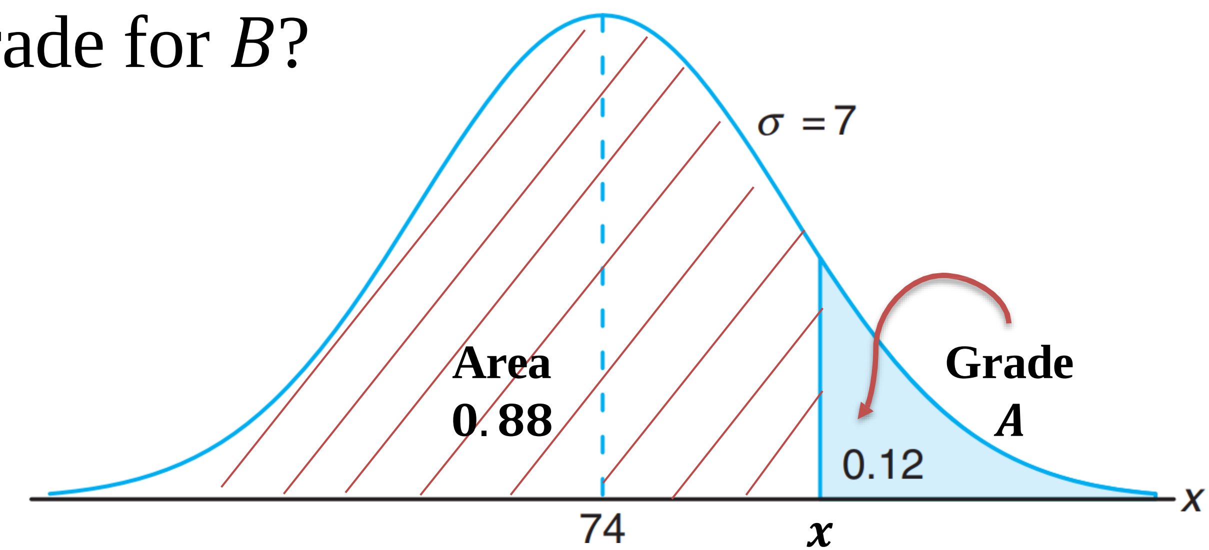
The average grade for an exam is 74, and the standard deviation is 7. If **12% of the class is given *A*s**, and the grades are curved to follow a normal distribution, what is the lowest possible integer grade for A and the highest possible integer grade for B?



Example5 (4/8):

$$X \sim N(74, 7^2) \Rightarrow Z = \frac{X-74}{7} \sim N(0,1)$$

The average grade for an exam is 74, and the standard deviation is 7. If **12% of the class is given *A*s**, and the grades are curved to follow a normal distribution, what is the lowest possible integer grade for A and the highest possible integer grade for B?

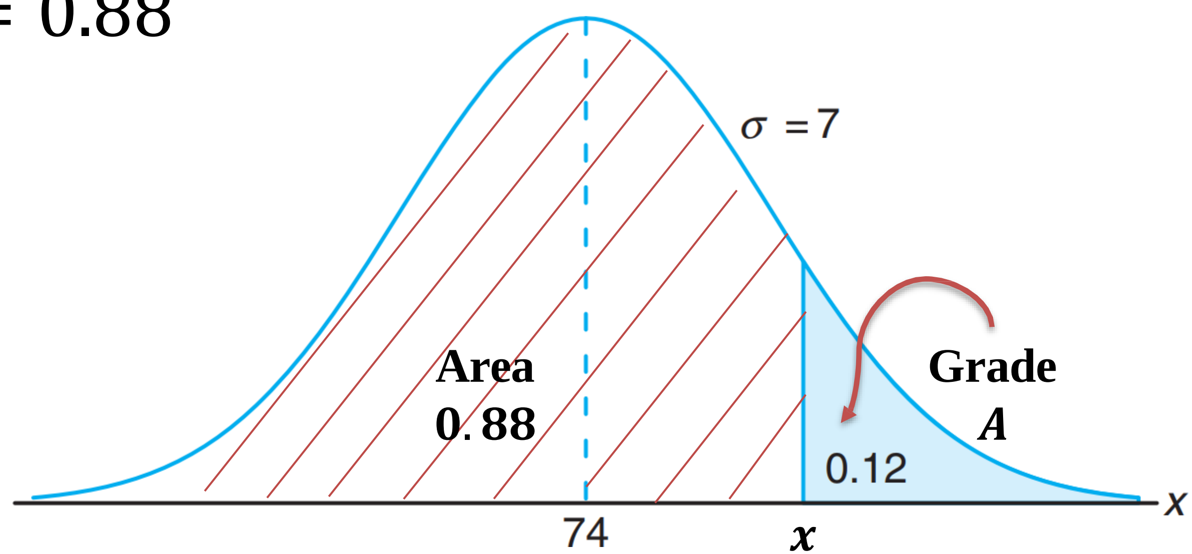


Example5 (5/8):

$$\boldsymbol{X \sim N(74, 7^2) \implies Z = \frac{X-74}{7} \sim N(0,1)}$$

$$P(X < x) = 0.88$$

$$P\left(Z < \frac{x-74}{7}\right) = 0.88$$



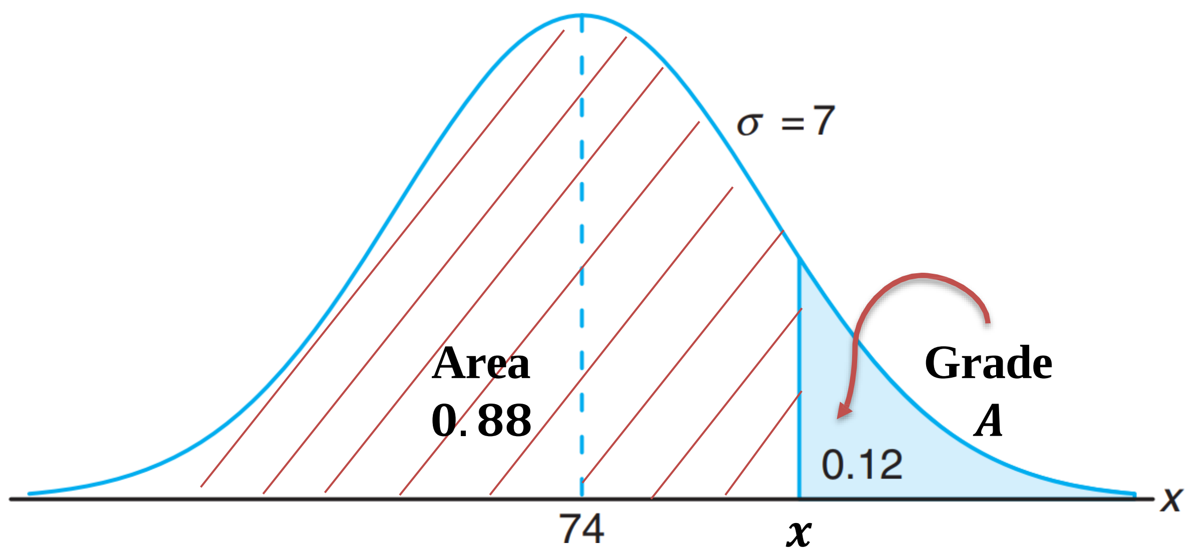
Normal Distribution (35/41)

Example5 (6/8):

$$\boldsymbol{X \sim N\left(74, 7^2\right) \implies Z = \frac{X-74}{7} \sim N(0,1)}$$

$$P(X < x) = 0.88$$

$$P(Z < z) = 0.88$$



Normal Distribution (35/41)

$$P(Z < z) = 0.88$$

z	.06	.07	.08	.09
0.0	0.5239	0.5279	0.5319	0.5359
0.1	0.5636	0.5675	0.5714	0.5753
0.2	0.6026	0.6064	0.6103	0.6141
0.3	0.6406	0.6443	0.6480	0.6517
0.4	0.6772	0.6808	0.6844	0.6879
0.5	0.7123	0.7157	0.7190	0.7224
0.6	0.7454	0.7486	0.7517	0.7549
0.7	0.7764	0.7794	0.7823	0.7852
0.8	0.8051	0.8078	0.8106	0.8133
0.9	0.8315	0.8340	0.8365	0.8389
1.0	0.8554	0.8577	0.8599	0.8621
1.1	0.8770	0.8790	0.8810	0.8830
1.2	0.8962	0.8980	0.8997	0.9015

Normal Distribution (35/41)

$$z \cong 1.18$$

z	.06	.07	.08	.09
0.0	0.5239	0.5279	0.5319	0.5359
0.1	0.5636	0.5675	0.5714	0.5753
0.2	0.6026	0.6064	0.6103	0.6141
0.3	0.6406	0.6443	0.6480	0.6517
0.4	0.6772	0.6808	0.6844	0.6879
0.5	0.7123	0.7157	0.7190	0.7224
0.6	0.7454	0.7486	0.7517	0.7549
0.7	0.7764	0.7794	0.7823	0.7852
0.8	0.8051	0.8078	0.8106	0.8133
0.9	0.8315	0.8340	0.8365	0.8389
1.0	0.8554	0.8577	0.8599	0.8621
1.1	0.8770	0.8790	0.8810	0.8830
1.2	0.8962	0.8980	0.8997	0.9015

Example5 (7/8):

$$\boldsymbol{X \sim N\left(74, 7^2\right) \implies Z = \frac{X-74}{7} \sim N(0, 1)}$$

$$P(X < x) = 0.88$$

$$P(Z < z) = 0.88$$

From the Table, $P(Z < 1.18)$ has the closest value to 0.88, so the desired z value is 1.18.

$$\because x = \sigma z + \mu$$

$$\therefore x = (7)(1.18) + 74 = 82.26$$

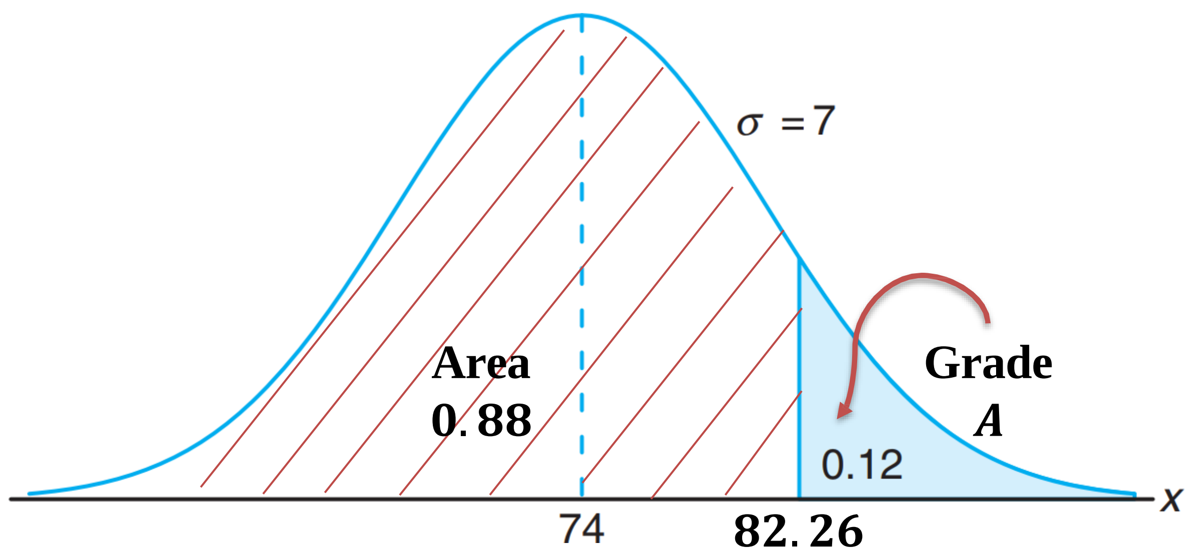
Normal Distribution (35/41)

Example5 (8/8):

$$X \sim N(74, 7^2) \Longrightarrow Z = \frac{X-74}{7} \sim N(0,1)$$

$\therefore x = (7)(1.18) + 74 = 82.26.$

Therefore, the lowest A is 83 and the highest B is 82.



Example6 (1/6):

The average life of a certain type of small motor is 10 years with a standard deviation of 2 years. The manufacturer replaces free all motors that fail while under guarantee. If he is willing to replace only 3% of the motors that fail, how long a guarantee should be offered? Assume that the lifetime of a motor follows a normal distribution.

Example6 (2/6):

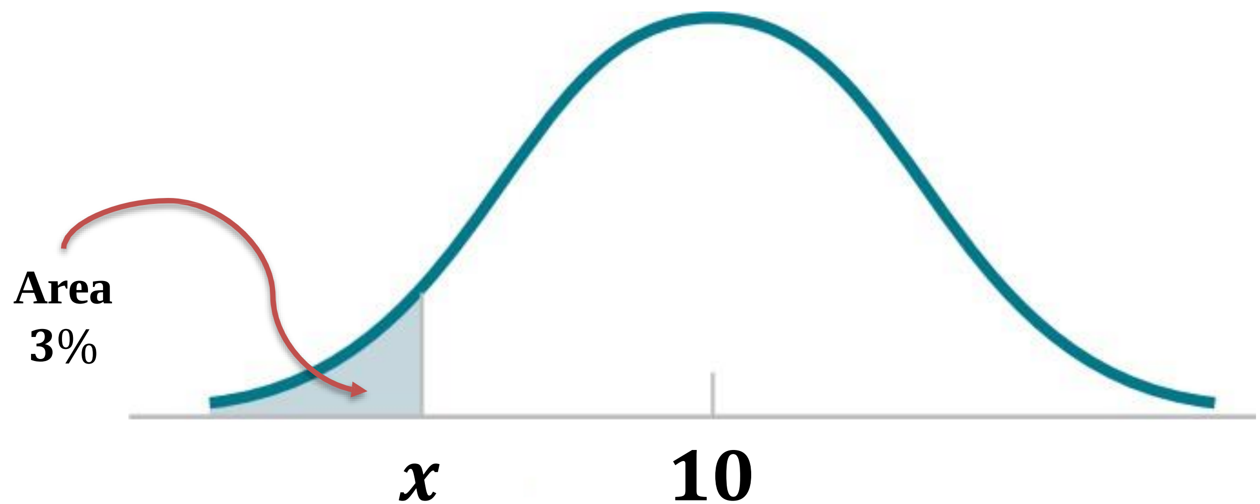
$$\boldsymbol{X \sim N(10, 2^2) \implies Z = \frac{X-10}{2} \sim N(0,1)}$$

The average life of a certain type of small motor is 10 years with a **standard deviation of 2 years**. The manufacturer replaces free all motors that fail while under guarantee. If he is willing to replace only 3% of the motors that fail, how long a guarantee should be offered? Assume that the lifetime of a motor follows a normal distribution.

Example6 (3/6):

$$X \sim N(10, 2^2) \Rightarrow Z = \frac{X-10}{2} \sim N(0,1)$$

The manufacturer replaces free all motors that fail while under guarantee. If he is willing to replace only 3% of the motors that fail, how long a guarantee should be offered? Assume that the lifetime of a motor follows a normal distribution.



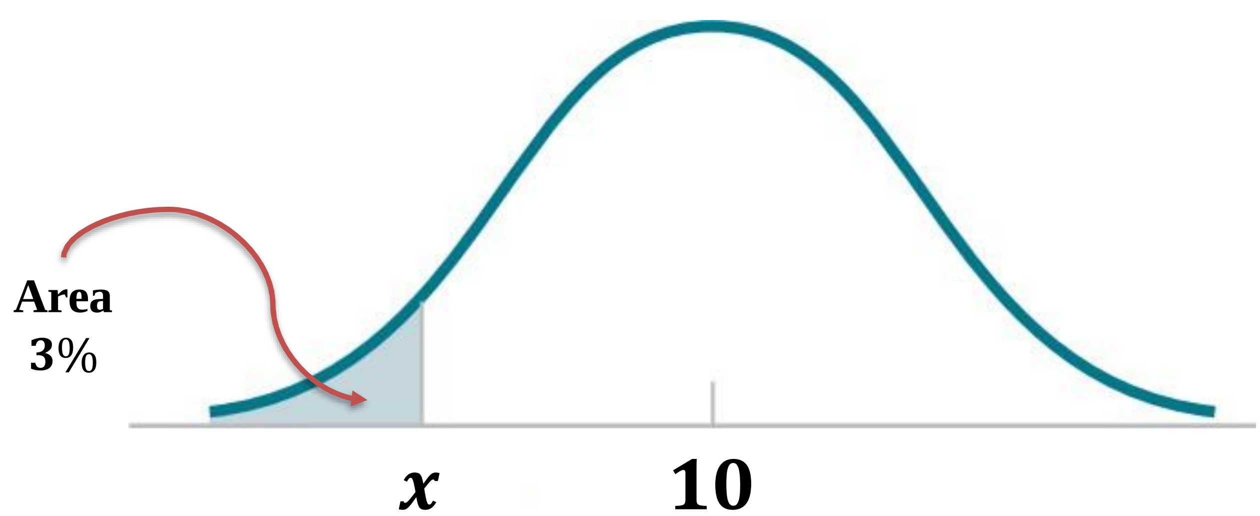
Normal Distribution (36/41)

Example6 (4/6):

$$\boldsymbol{X \sim N\left(10, 2^2\right) \implies Z = \frac{X-10}{2} \sim N(0,1)}$$

$$P(X < x) = 0.03$$

$$P(Z < z) = 0.03$$



Normal Distribution (36/41)

$$P(Z < z) = 0.03$$

z	.04	.05	.06	.07	.08	.09
−2.4	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367

Normal Distribution (36/41)

$$z \cong -1.88$$

z	.04	.05	.06	.07	.08	.09
−2.4	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367

Example6 (5/6):

$$\boldsymbol{X \sim N(10, 2^2) \implies Z = \frac{X-10}{2} \sim N(0,1)}$$

$P(X < x) = 0.03$

$P(Z < z) = 0.03$

From the Table, $P(Z < -1.88)$ has the closest value to 0.03, so the desired z value is -1.18.

$\because x = \sigma z + \mu$

$\therefore x = (2)(-1.18) + 10 = 6.24$ years

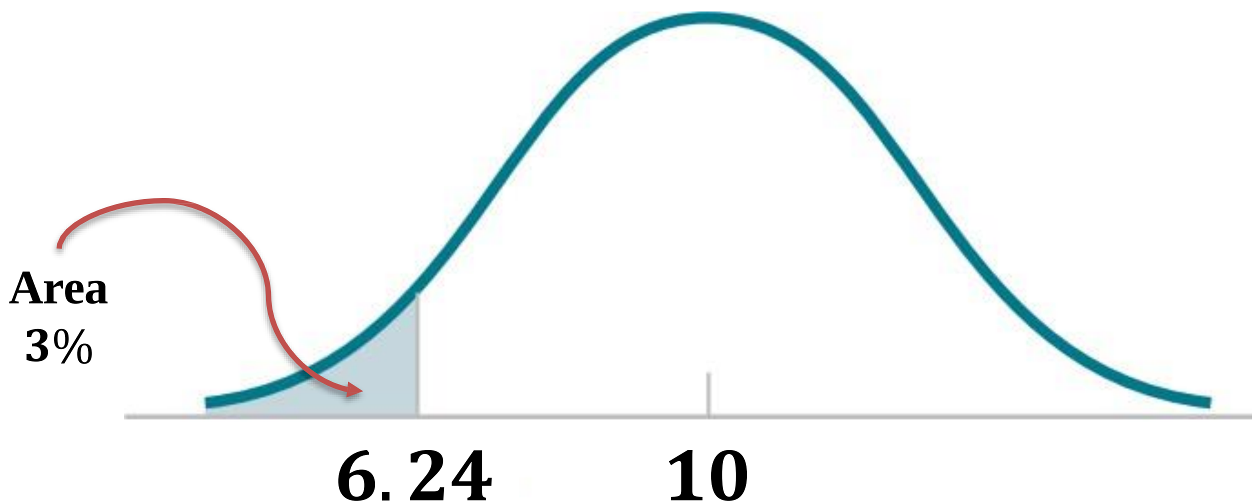
Normal Distribution (36/41)

Example6 (6/6):

$$X \sim N\left(10, 2^2\right) \Rightarrow Z = \frac{X-10}{2} \sim N(0,1)$$

$P(X < x) = 0.03$

$x = (2)(-1.18) + 10 = 6.24$ years



Normal Approximation to the Binomial Distribution (1/20):

Provides a very accurate approximation to the binomial distribution when n is large and p is not extremely close to 0 or 1

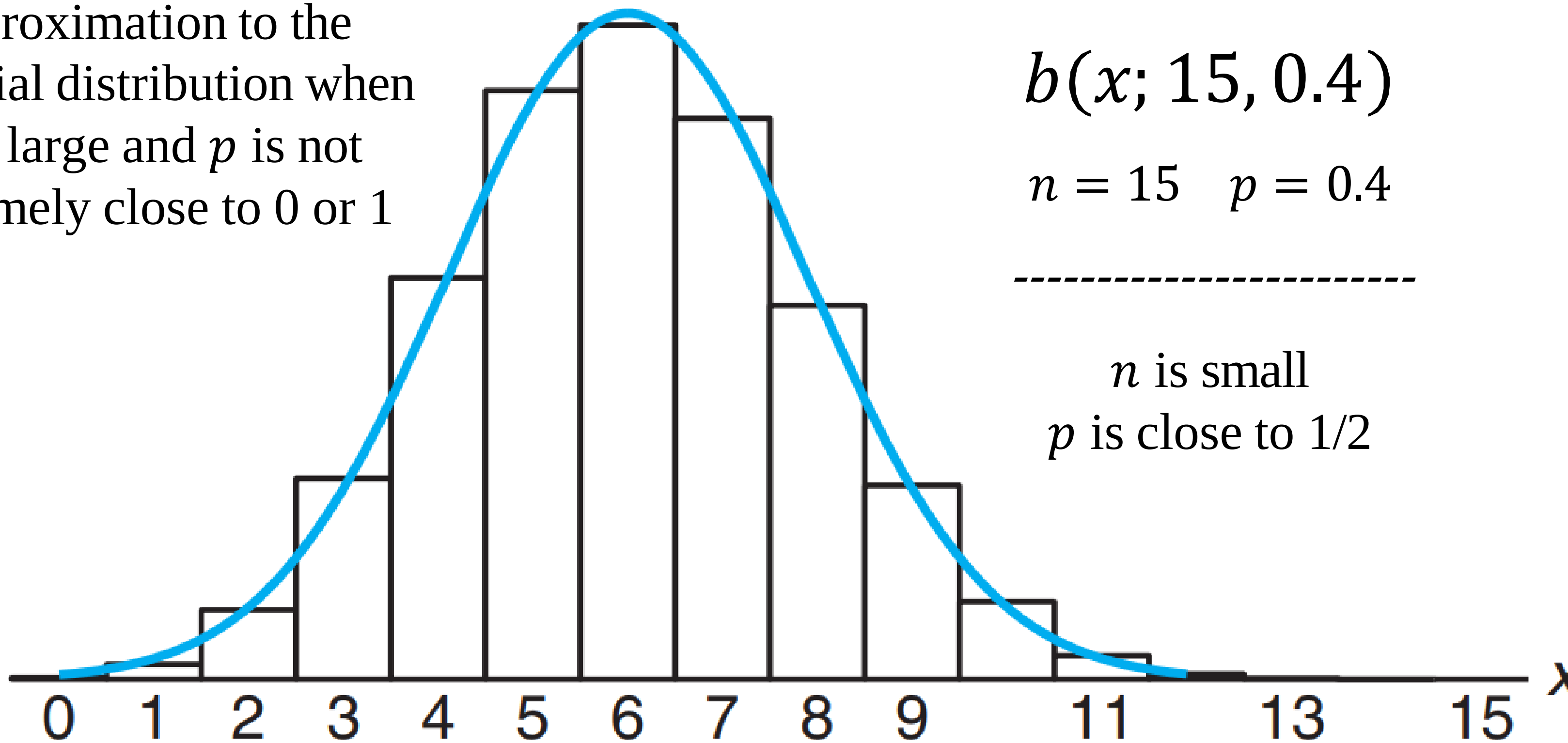
$b(x; 15, 0.4)$

$n = 15 \quad p = 0.4$

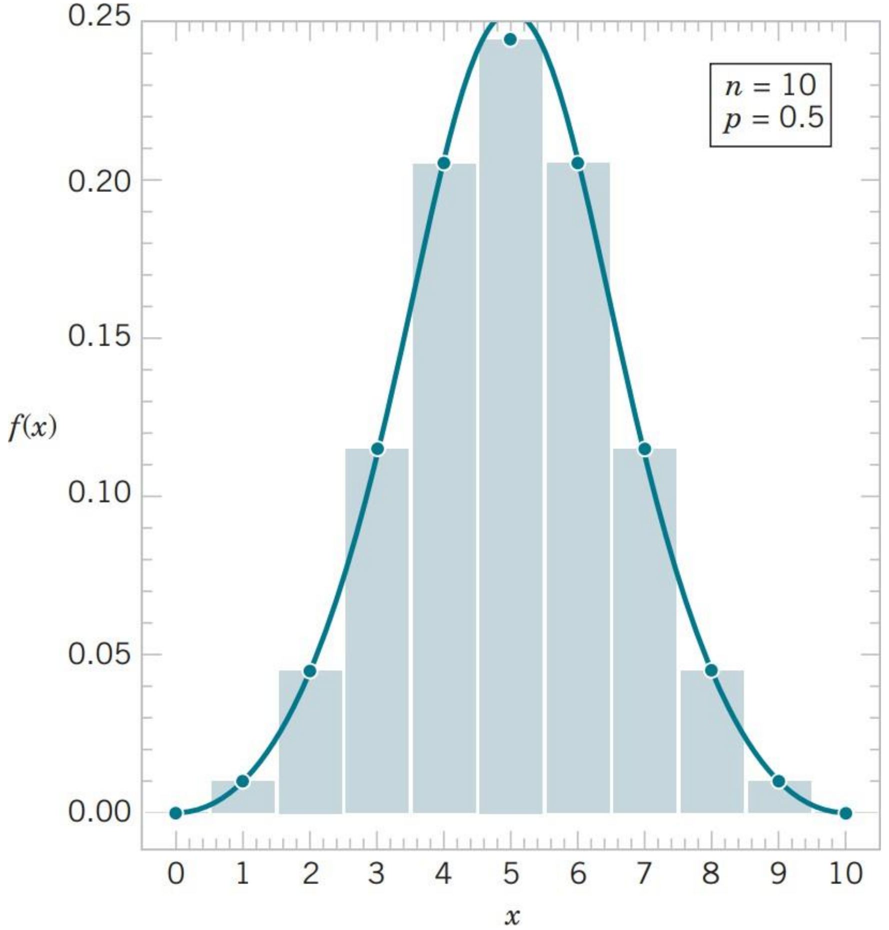
---

n is small

p is close to 1/2

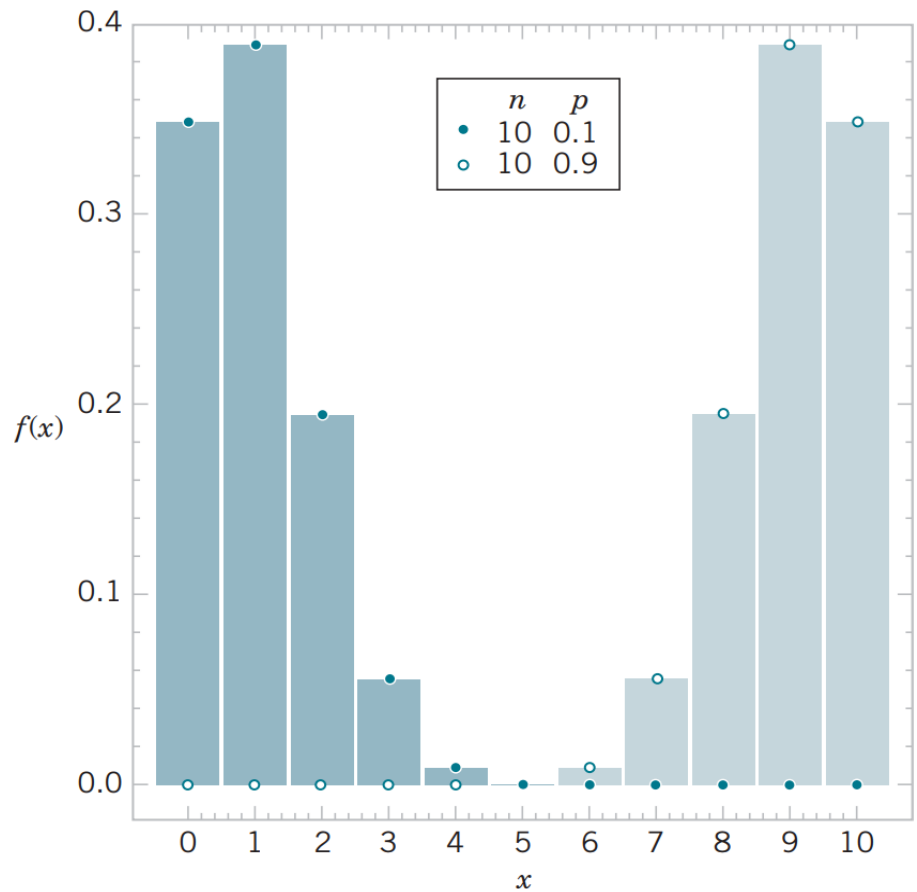


Normal Approximation to the Binomial Distribution (2/20):

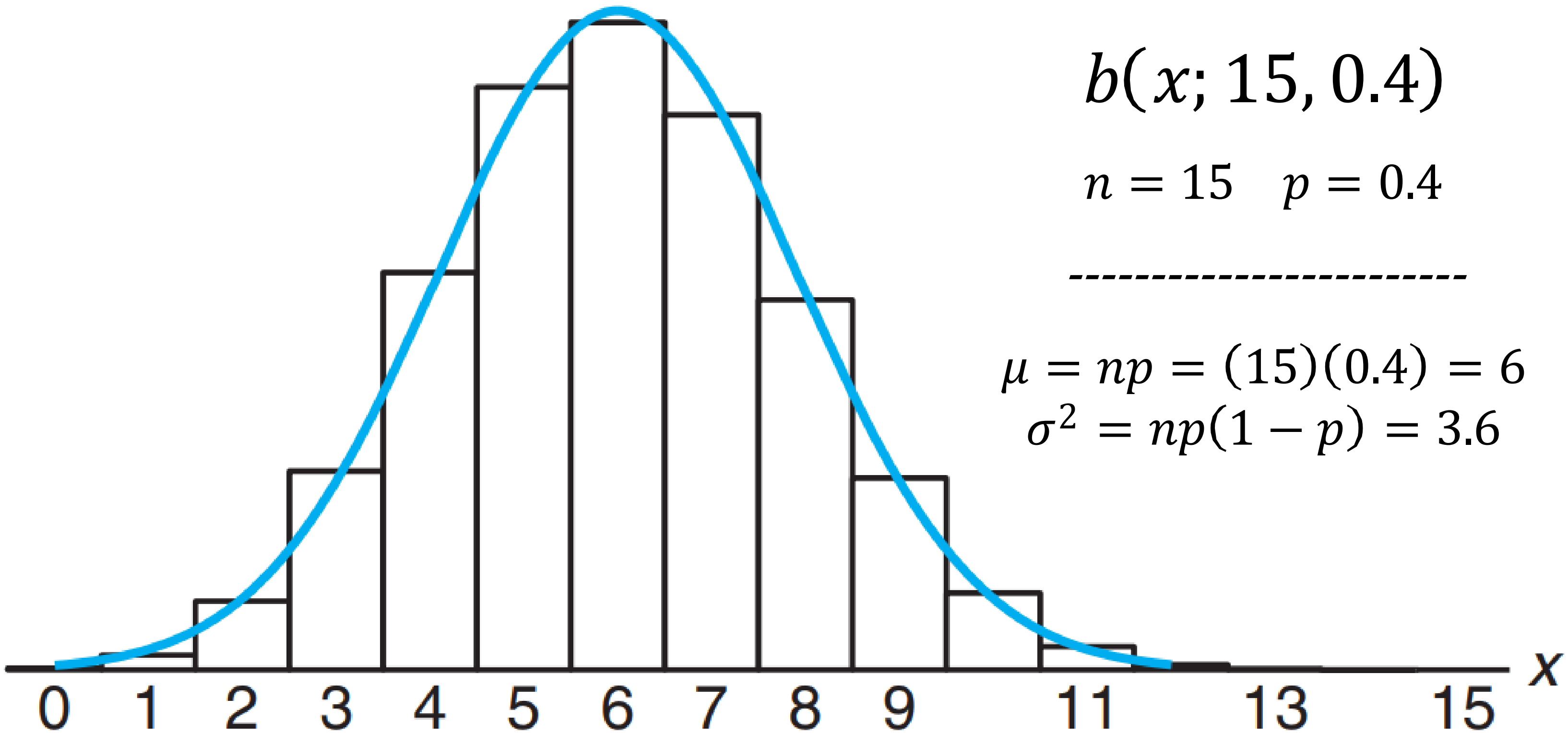


Normal Approximation to the Binomial Distribution (3/20):

Binomial distribution is not symmetrical if n is small and p is near 0 or 1



Normal Approximation to the Binomial Distribution (4/20):



$$b(x; 15, 0.4)$$

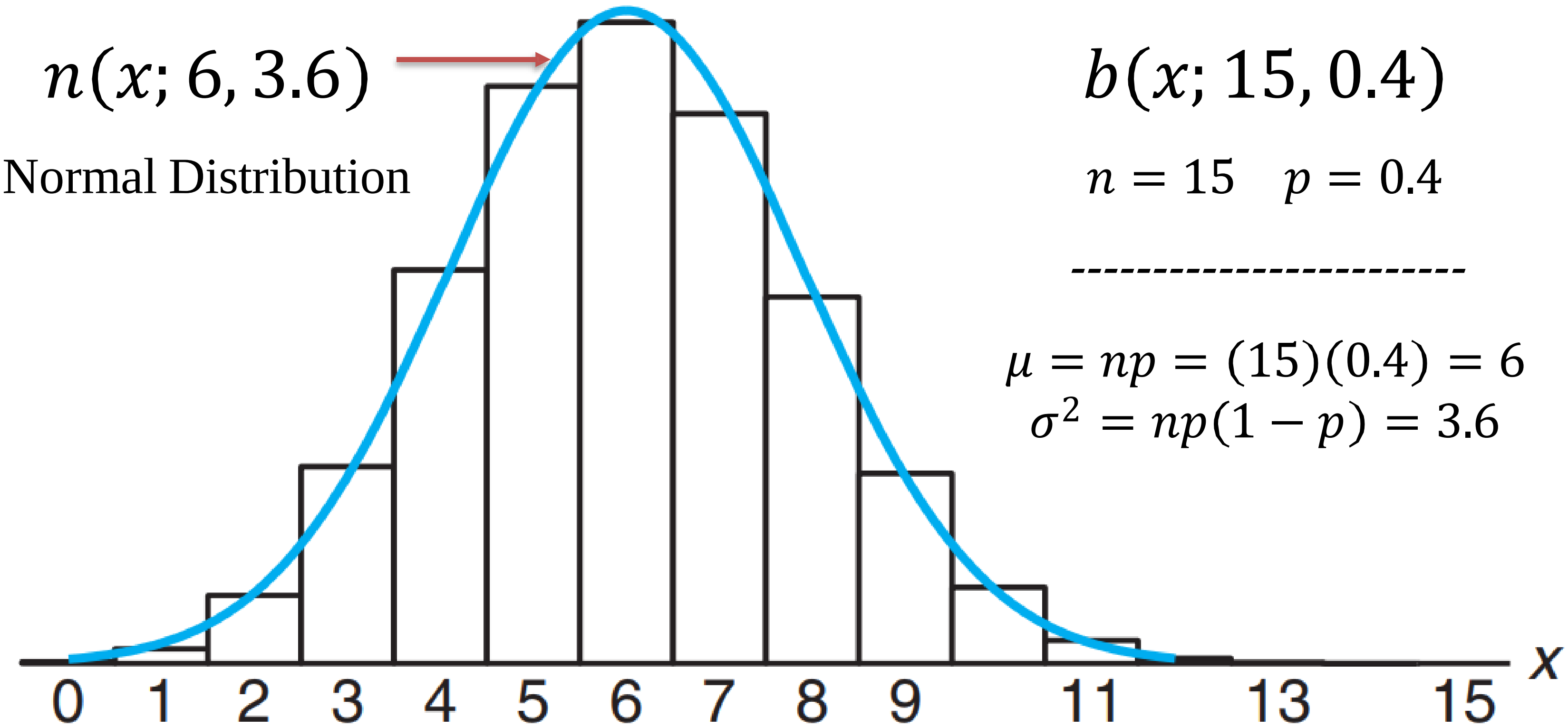
$$n = 15 \quad p = 0.4$$

$$\mu = np = (15)(0.4) = 6$$
$$\sigma^2 = np(1-p) = 3.6$$

Normal Approximation to the Binomial Distribution (5/20):

$n(x; 6, 3.6)$

Normal Distribution



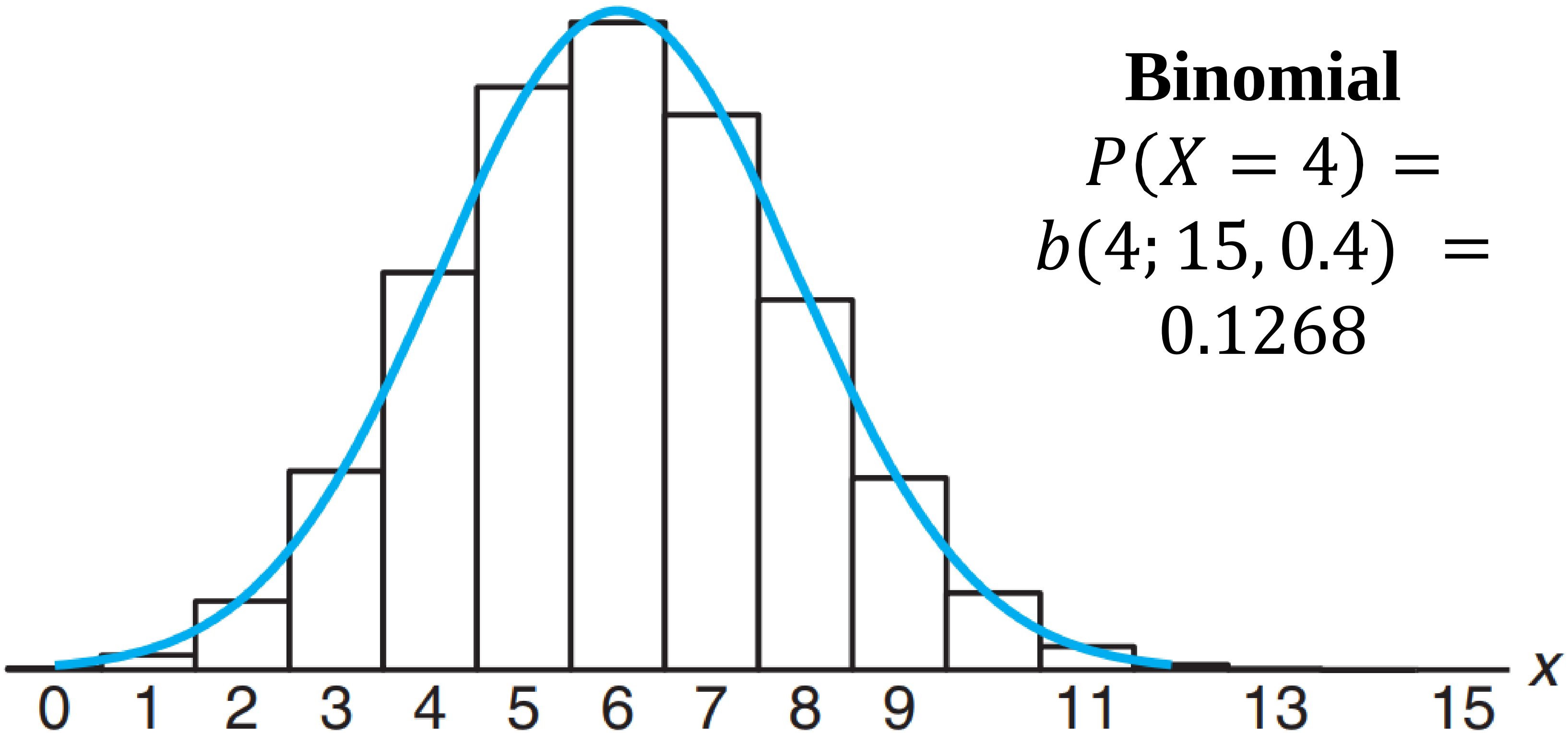
$b(x; 15, 0.4)$

$n = 15 \quad p = 0.4$

$$\mu = np = (15)(0.4) = 6$$
$$\sigma^2 = np(1-p) = 3.6$$

Normal Distribution (37/41)

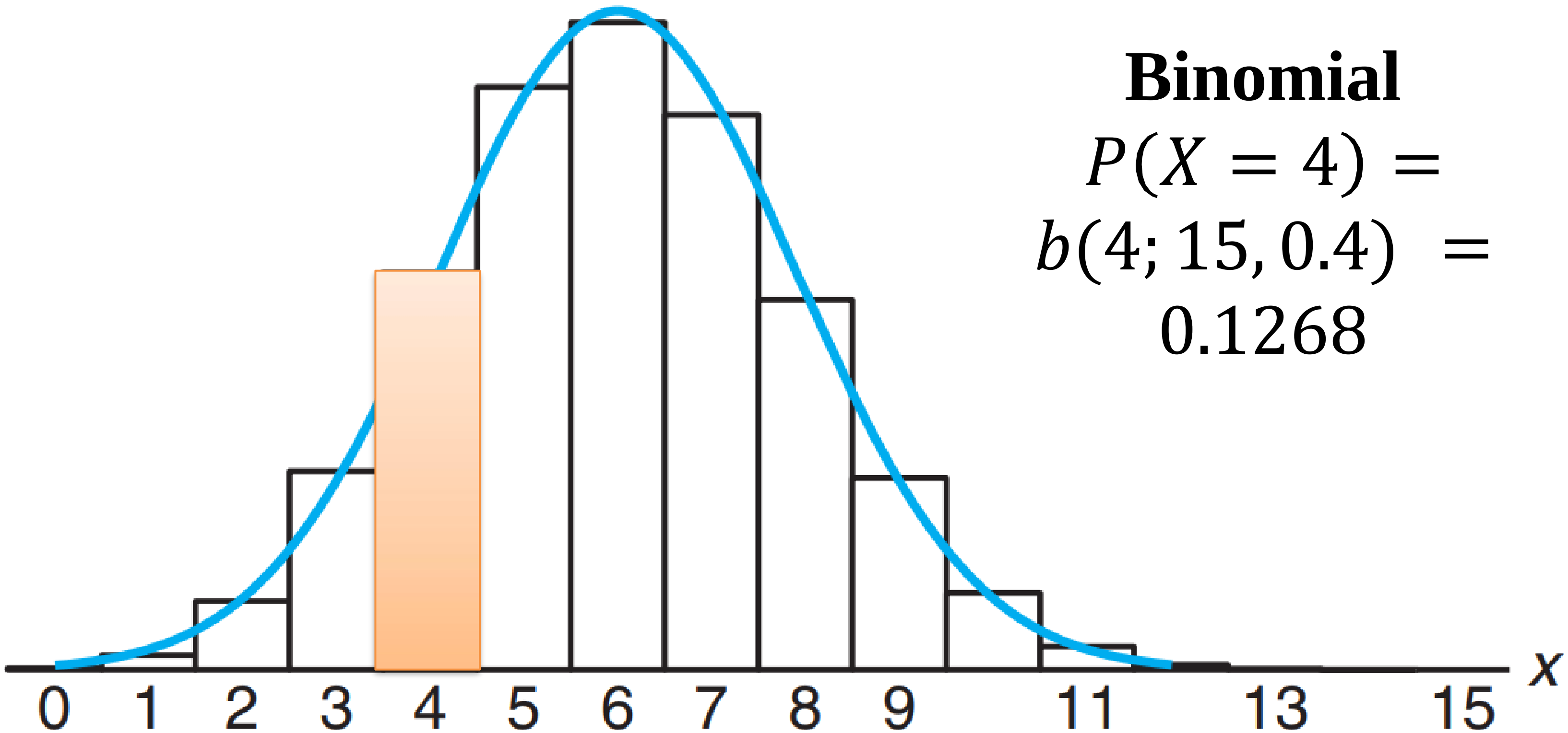
Normal Approximation to the Binomial Distribution (6/20):



Binomial

$P(X = 4) =$
$b(4; 15, 0.4) =$
0.1268

Normal Approximation to the Binomial Distribution (7/20):

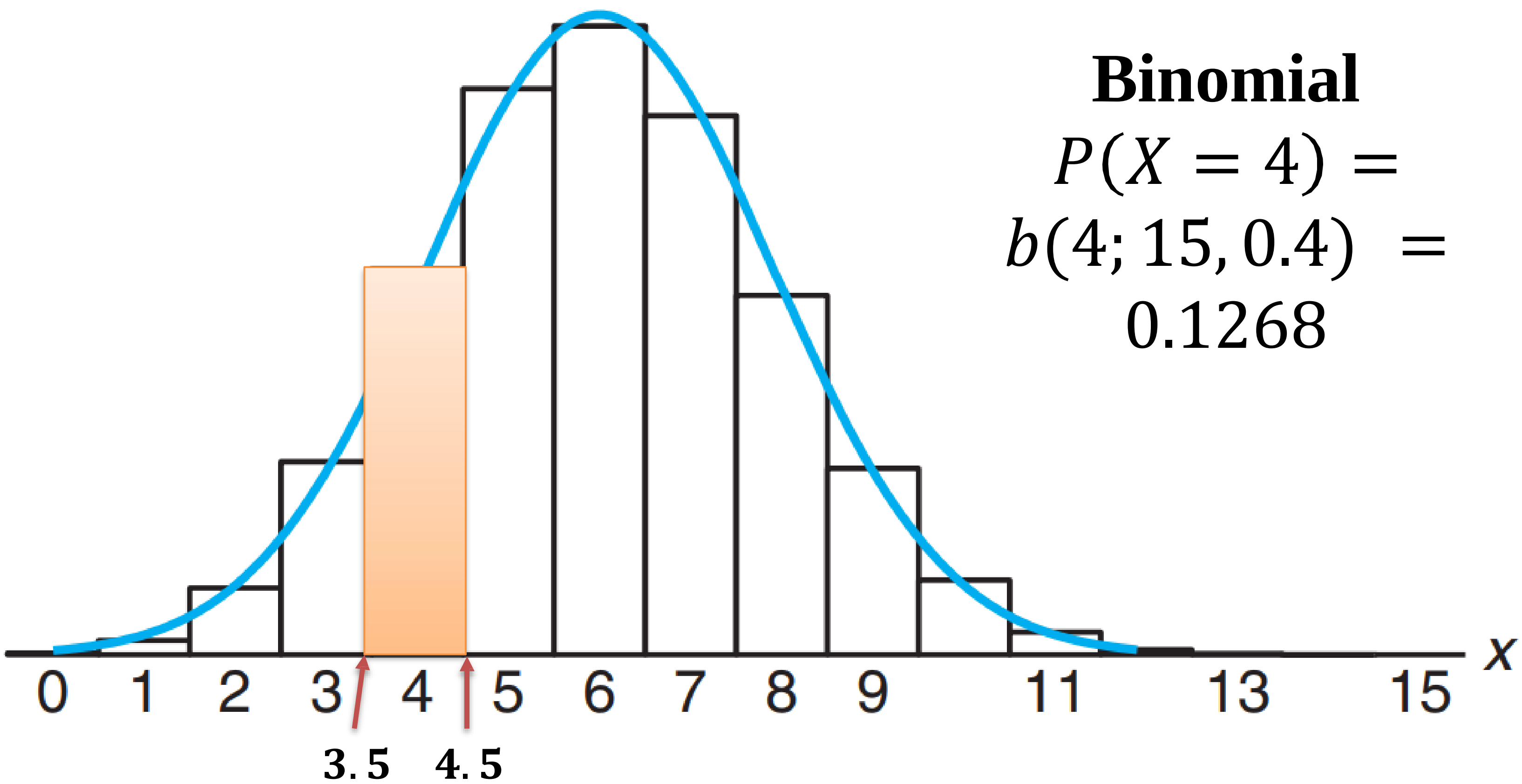


Binomial

$P(X = 4) = b(4; 15, 0.4) = 0.1268$

Normal Distribution (37/41)

Normal Approximation to the Binomial Distribution (8/20):



Binomial

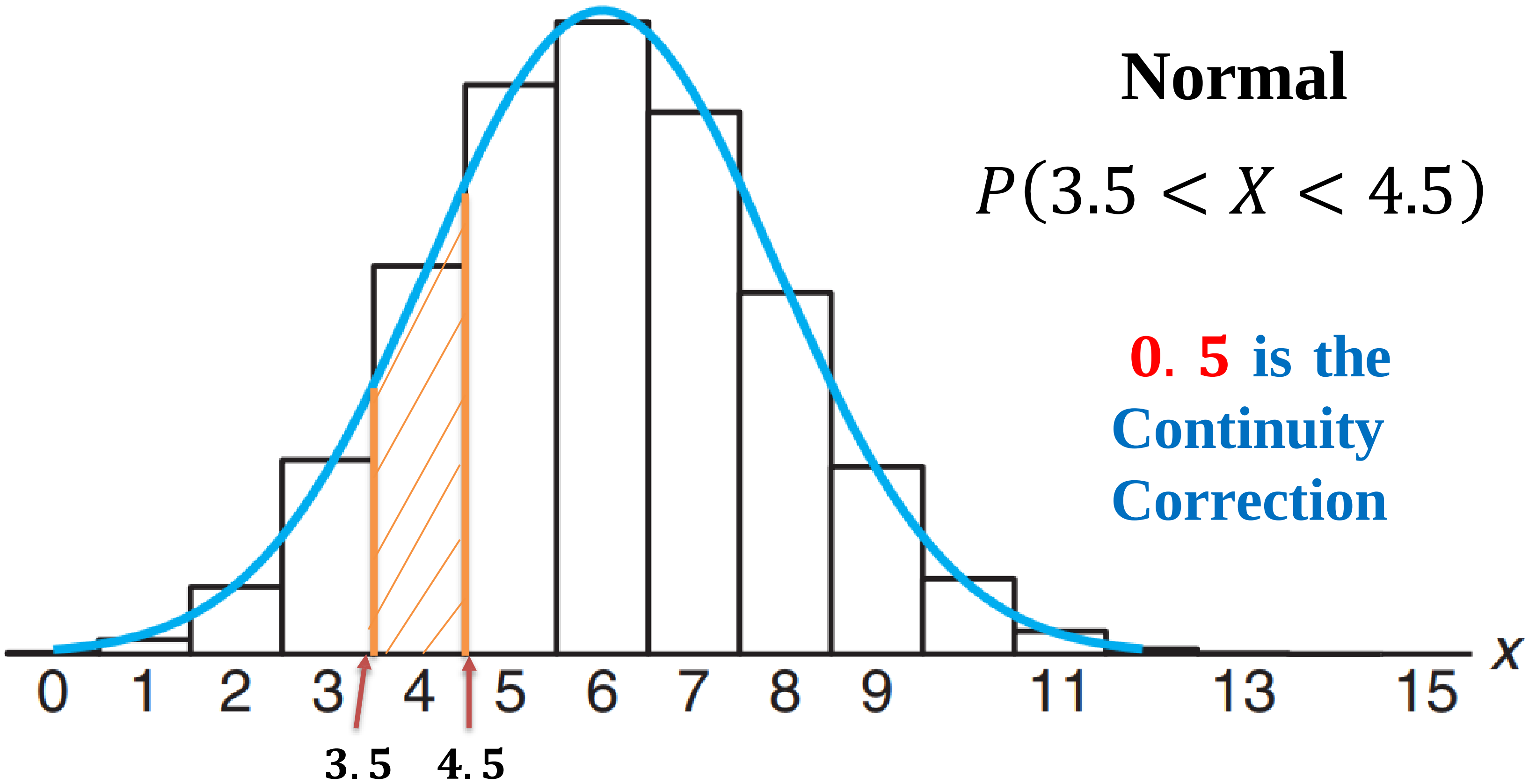
$P(X = 4) =$

$b(4; 15, 0.4) =$

0.1268

Normal Distribution (37/41)

Normal Approximation to the Binomial Distribution (9/20):

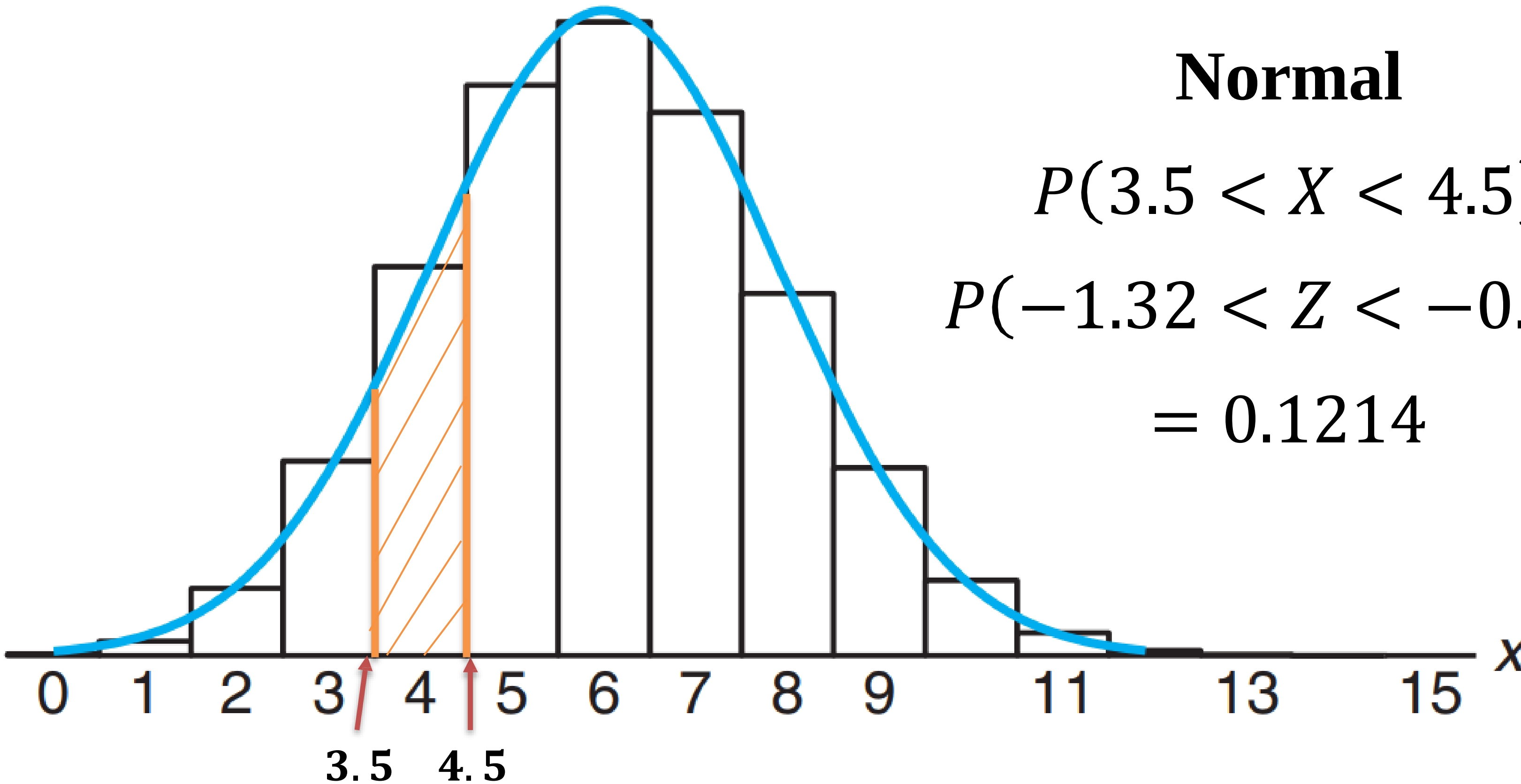


Normal

$$P(3.5 < X < 4.5)$$

0. 5 is the Continuity Correction

Normal Approximation to the Binomial Distribution (10/20):



Normal

$$P(3.5 < X < 4.5)$$

$$P(-1.32 < Z < -0.79)$$

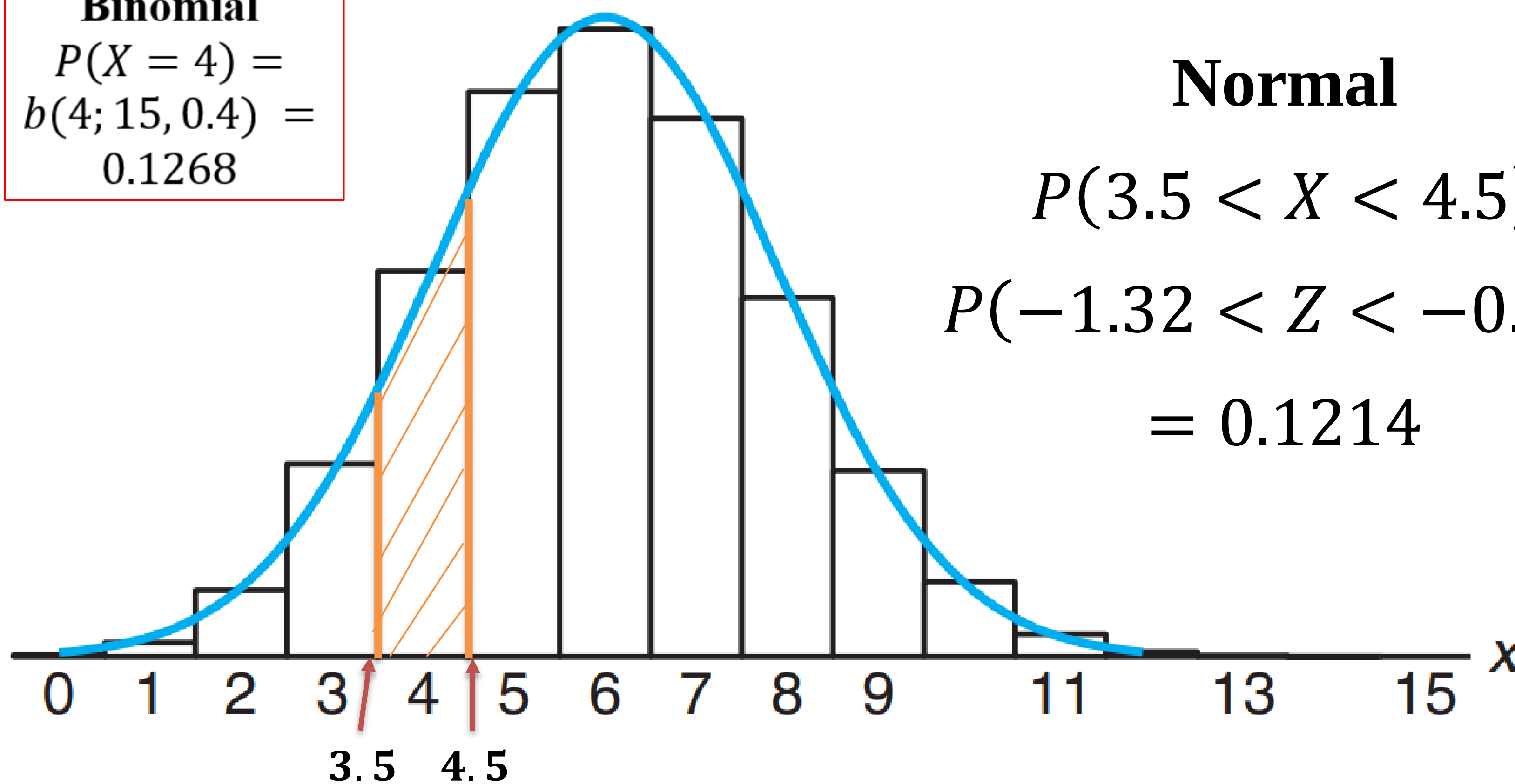
$$= 0.1214$$

Normal Distribution (37/41)

Normal Approximation to the Binomial Distribution (10/20):

Binomial

$P(X = 4) = b(4; 15, 0.4) = 0.1268$



Normal

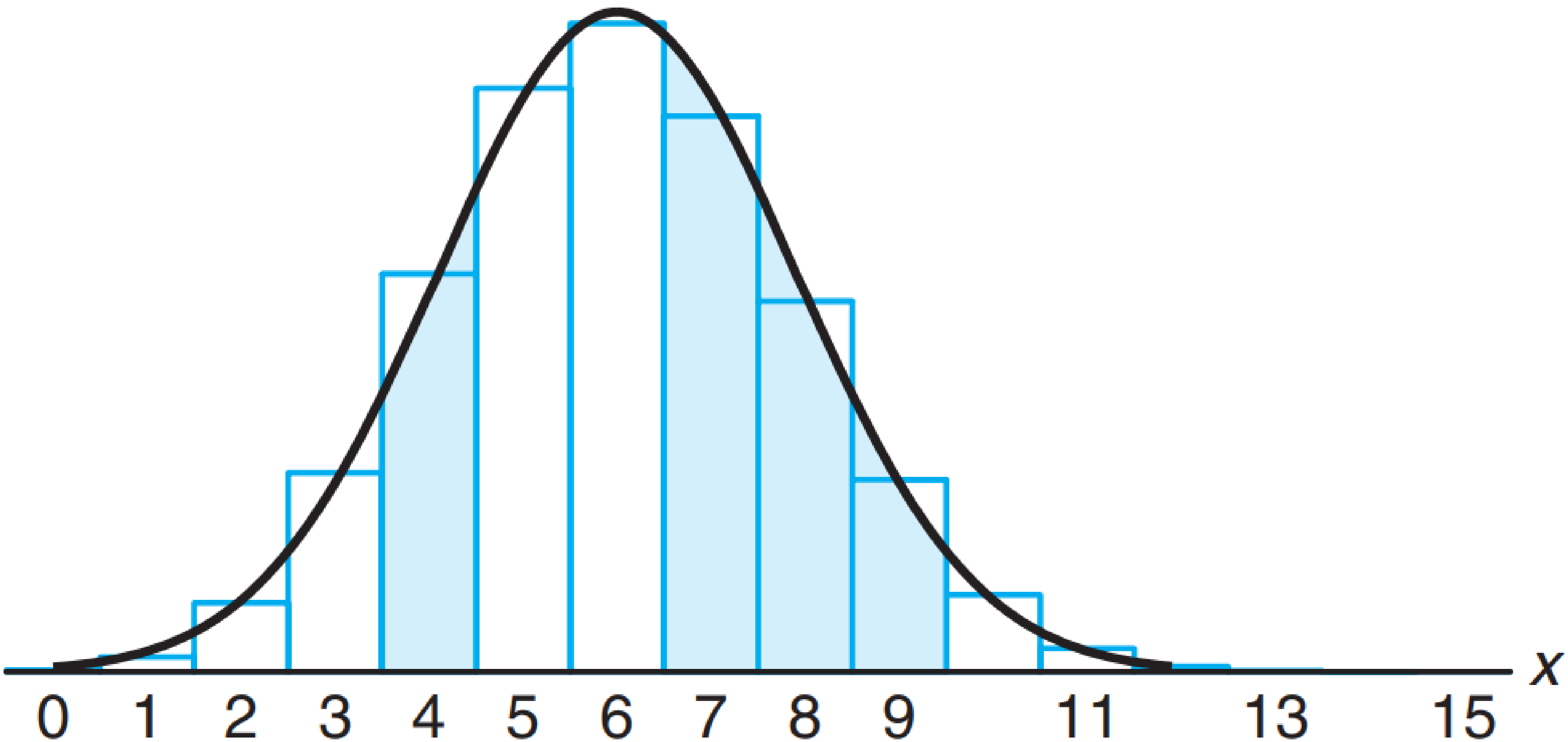
$$P(3.5 < X < 4.5)$$

$$P(-1.32 < Z < -0.79)$$

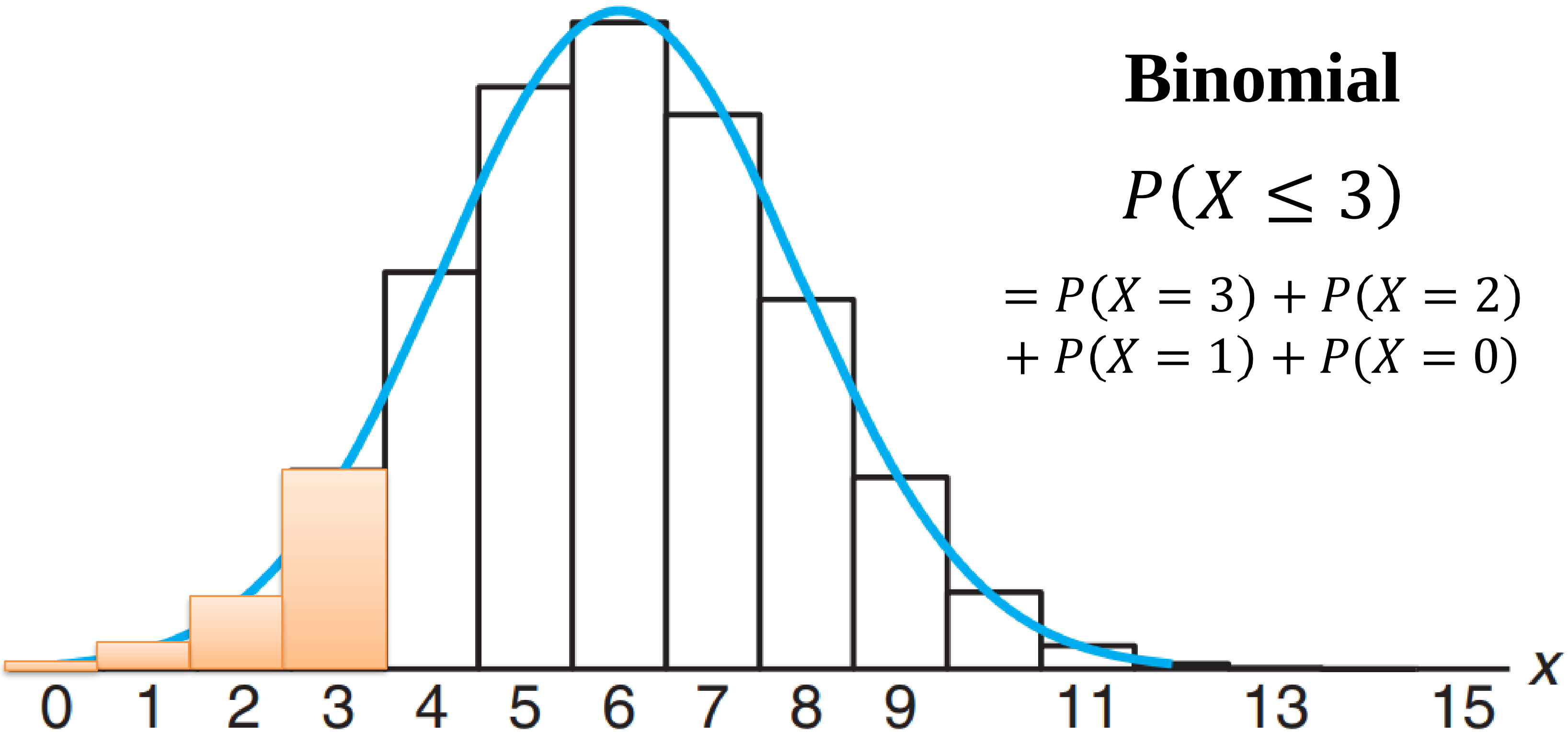
$$= 0.1214$$

Normal Distribution (37/41)

Normal Approximation to the Binomial Distribution (11/20):



Normal Approximation to the Binomial Distribution (12/20):

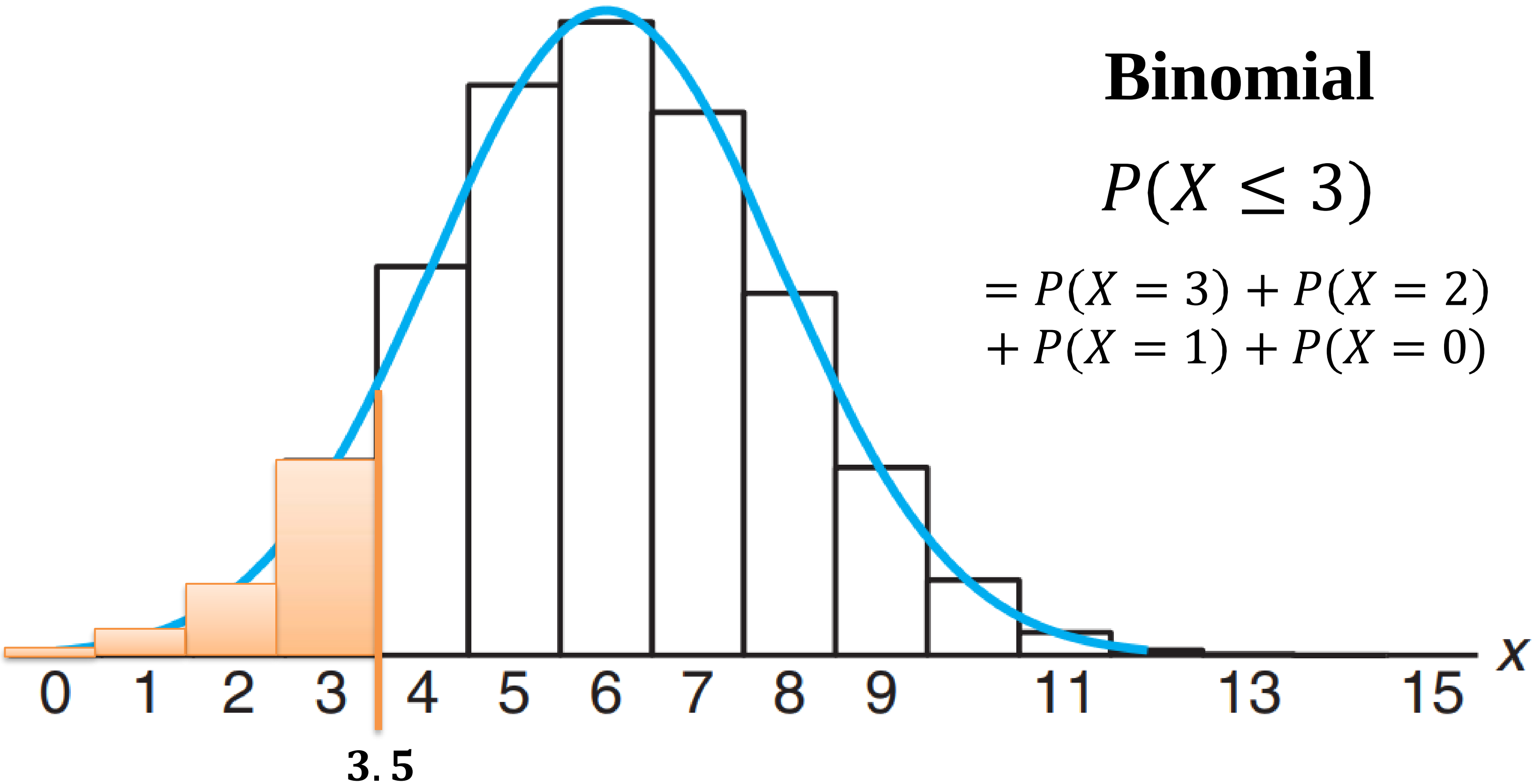


Binomial

$$P(X \leq 3)$$

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

Normal Approximation to the Binomial Distribution (13/20):

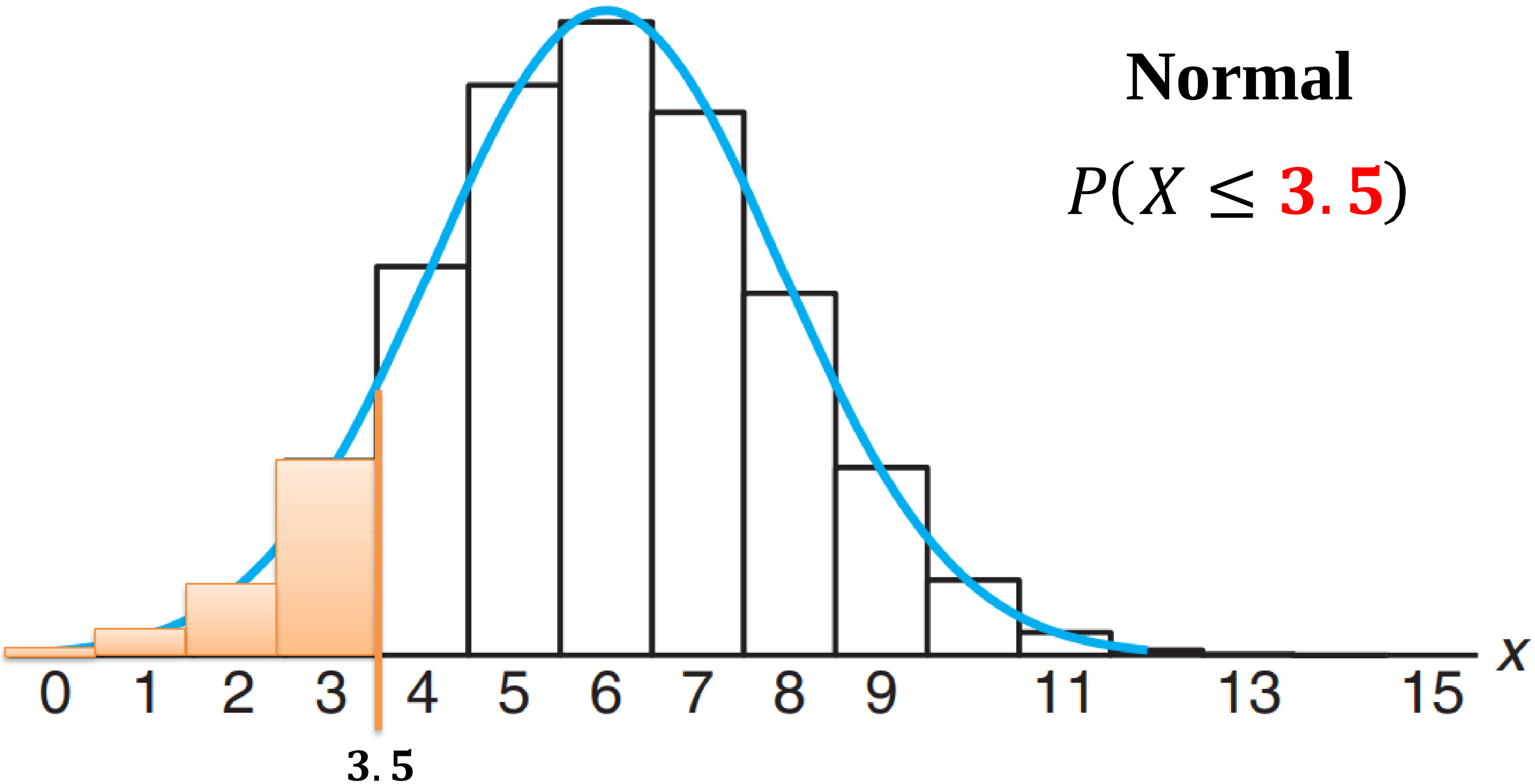


Binomial

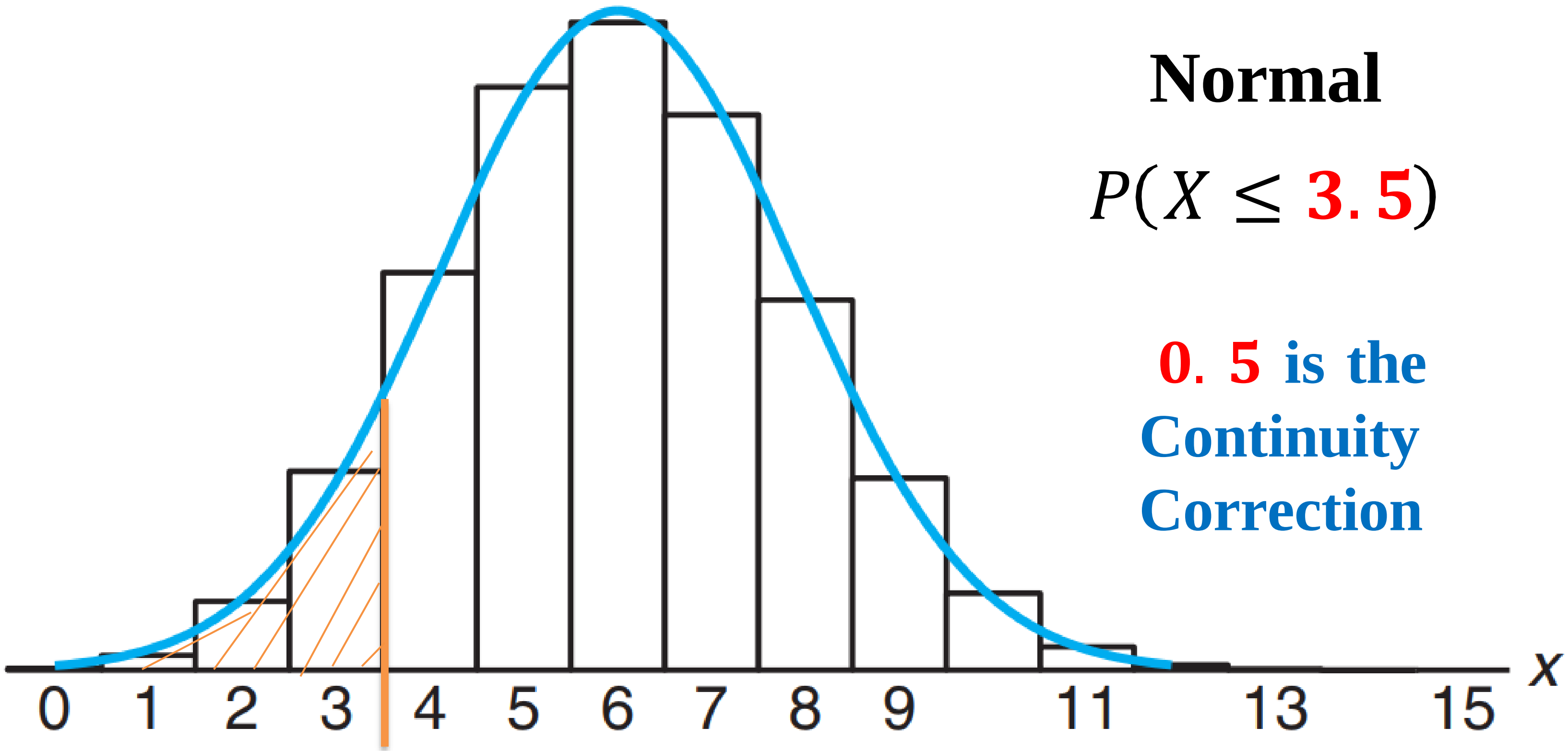
$$P(X \leq 3)$$

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

Normal Approximation to the Binomial Distribution (14/20):



Normal Approximation to the Binomial Distribution (15/20):



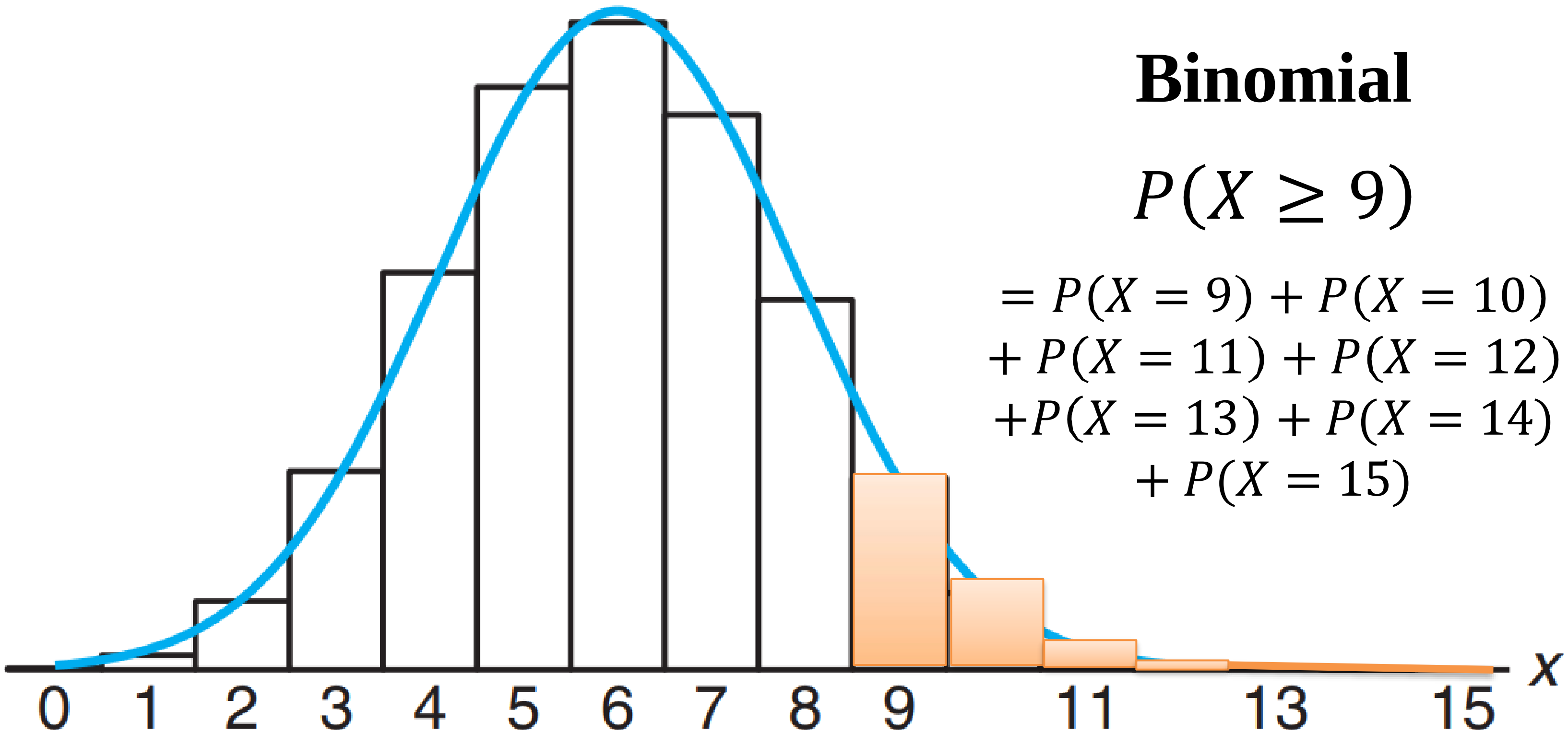
Normal

$P(X \leq 3.5)$

0.5 is the Continuity Correction

Normal Distribution (37/41)

Normal Approximation to the Binomial Distribution (16/20):

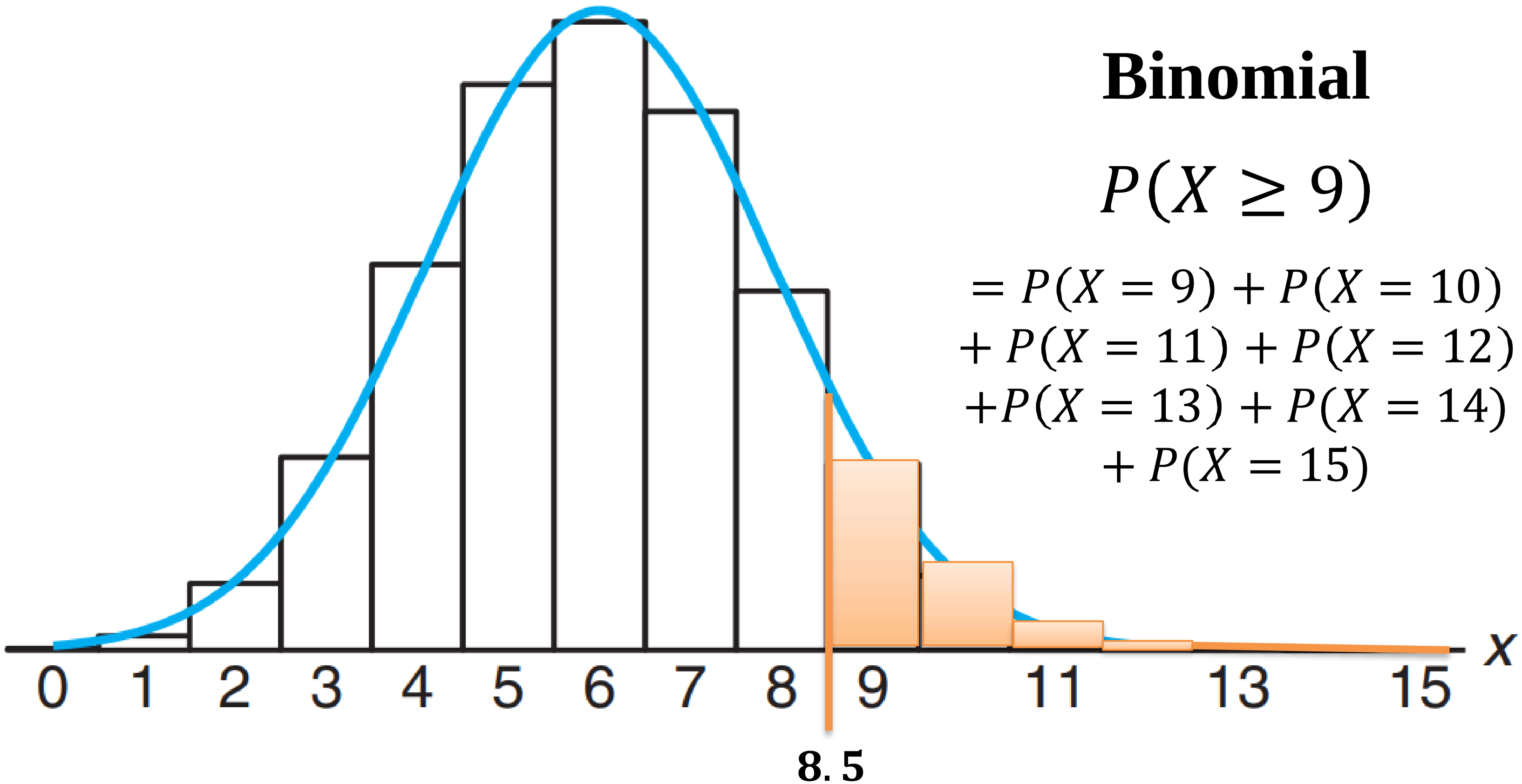


Binomial

$$P(X \geq 9)$$

$$= P(X = 9) + P(X = 10) \\ + P(X = 11) + P(X = 12) \\ + P(X = 13) + P(X = 14) \\ + P(X = 15)$$

Normal Approximation to the Binomial Distribution (17/20):



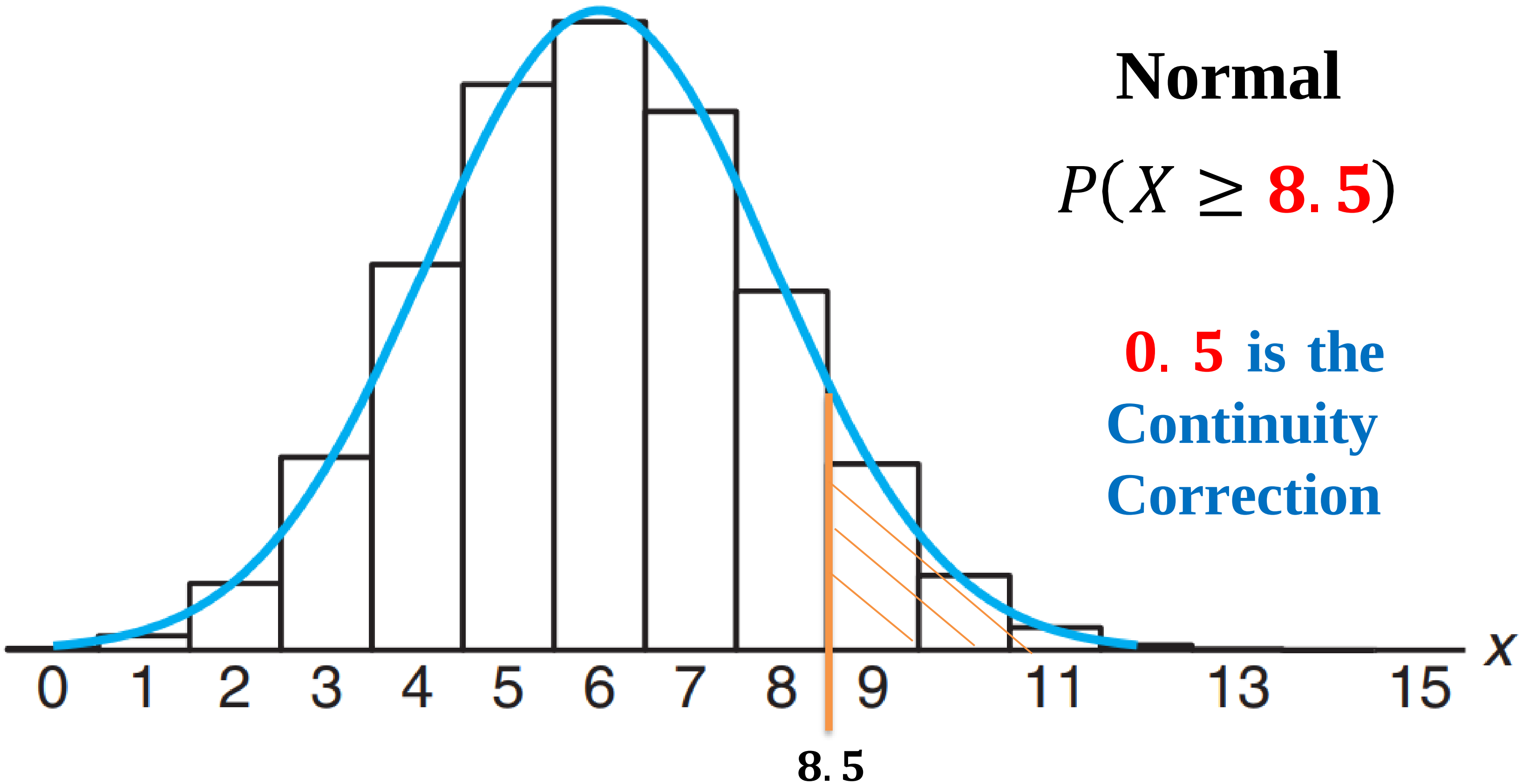
Binomial

$$P(X \geq 9)$$

$$= P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12) + P(X = 13) + P(X = 14) + P(X = 15)$$

Normal Distribution (37/41)

Normal Approximation to the Binomial Distribution (18/20):



Normal

$P(X \geq \mathbf{8.5})$

0.5 is the Continuity Correction

Normal Approximation to the Binomial Distribution (19/20):

If X is a binomial random variable with parameters n and p,

Because $\mu = np$ and $\sigma^2 = np(1-p)$ for binomial, then

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is a standard normal random variable

with mean = 0 and variance = 1.

Provides a very accurate approximation to the binomial distribution when n is large and p is not extremely close to 0 or 1

Normal Distribution (37/41)

Normal Approximation to the Binomial Distribution (20/20):

To approximate a binomial probability with a normal distribution, a *continuity correction* is applied as follows:

$$P(X \leq x) \approx P(X \leq x + 0.5) \approx P\left(Z \leq \frac{(x + 0.5) - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(X \geq x) \approx P(X \geq x - 0.5) \approx P\left(Z \geq \frac{(x - 0.5) - np}{\sqrt{np(1-p)}}\right)$$

Example1 (1/6):

A coin is tossed 400 times. Use the normal curve approximation to find the probability of obtaining

(a) between 185 and 210 heads inclusive;
(b) exactly 205 heads;
(c) fewer than 176 or more than 227 heads.

Normal Distribution (38/41)

Example1 (2/6):

$$\boldsymbol{b(n = 400, p = 0.5)}$$

A coin is tossed 400 times. Use the normal curve approximation to find the probability of obtaining

(a) between 185 and 210 heads inclusive;
(b) exactly 205 heads;
(c) fewer than 176 or more than 227 heads.

$$\mu = np = (400)(0.5) = 200$$

$$\sigma^2 = np(1-p) = (400)(0.5)(0.5) = 100$$

$$\sigma = \sqrt{100} = 10$$

Normal Distribution (38/41)

Example1 (3/6): $\boldsymbol{b(n = 400, p = 0.5),\ \mu = 200,\ \sigma = 10}$

(a) between 185 and 210 heads inclusive.

$$P(185 \leq X \leq 210) \approx P\left(\frac{184.5 - 200}{10} \leq Z \leq \frac{210.5 - 200}{10}\right)$$

$$\approx P(-1.55 \leq Z \leq 1.05) = P(Z \leq 1.05) - P(Z \leq -1.55)$$

$$= 0.8531 - 0.0606 = 0.7925$$

Normal Distribution (38/41)

Example1 (4/6): $\boldsymbol{b(n = 400, p = 0.5),\ \mu = 200,\ \sigma = 10}$

(b) exactly 205 heads.

$$P(X = 205) \approx P\left(\frac{204.5 - 200}{10} < Z < \frac{205.5 - 200}{10}\right)$$

$$\approx P(0.45 < Z < 0.55) = P(Z < 0.55) - P(Z < 0.45)$$

$$= 0.7088 - 0.6736 = 0.0352$$

Normal Distribution (38/41)

Example1 (5/6): $\boldsymbol{b(n = 400, p = 0.5),\ \mu = 200,\ \sigma = 10}$

(c) fewer than 176 or more than 227 heads.

$$P(X < 176) = P(X \leq 175)$$
$$+$$
$$P(X > 227) = P(X \geq 228)$$

Binomial is Discrete

Normal Distribution (38/41)

Example1 (5/6): $\boldsymbol{b(n=400, p=0.5),\ \mu=200,\ \sigma=10}$

(c) fewer than 176 or more than 227 heads.

$$P(X < 176) = P(X \le 175)$$
$$+$$
$$P(X > 227) = P(X \ge 228)$$

Binomial is Discrete

Normal Distribution (38/41)

Example1 (6/6): $\boldsymbol{b(n = 400, p = 0.5),\ \mu = 200,\ \sigma = 10}$

(c) fewer than 176 or more than 227 heads.

$$P(X \le 175) + P(X \ge 228)$$

$$\approx P\left(Z \le \frac{175.5 - 200}{10}\right) + P\left(Z \ge \frac{227.5 - 200}{10}\right)$$

$$\approx P(Z \le -2.45) + P(Z \ge 2.75)$$

$$= P(Z \le -2.45) + \big(1 - P(Z < 2.75)\big)$$

$$= 0.0071 + (1 - 0.9970) = 0.0101$$

Normal Distribution (39/41)

Example2 (1/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that fewer than 30 survive?

Example2 (2/5): $\boldsymbol{b(n = 100, p = 0.4),\ \mu = 40,\ \sigma = 4.899}$

The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that fewer than 30 survive?

Let the binomial variable X represent the number of patients who survive.

$$\mu = np = (100)(0.4) = 40$$

$$\sigma^2 = np(1 - p) = (100)(0.4)(0.6) = 24$$

$$\sigma = \sqrt{24} = 4.899$$

Example2 (3/5): $\boldsymbol{b(n = 100, p = 0.4),\ \mu = 40,\ \sigma = 4.899}$

The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that fewer than 30 survive?

$$P(X < 30) = P(X \leq 29)$$

Binomial is Discrete

Normal Distribution (39/41)

Example2 (4/5): $\boldsymbol{b(n=100, p=0.4),\ \mu=40,\ \sigma=4.899}$

The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that fewer than 30 survive?

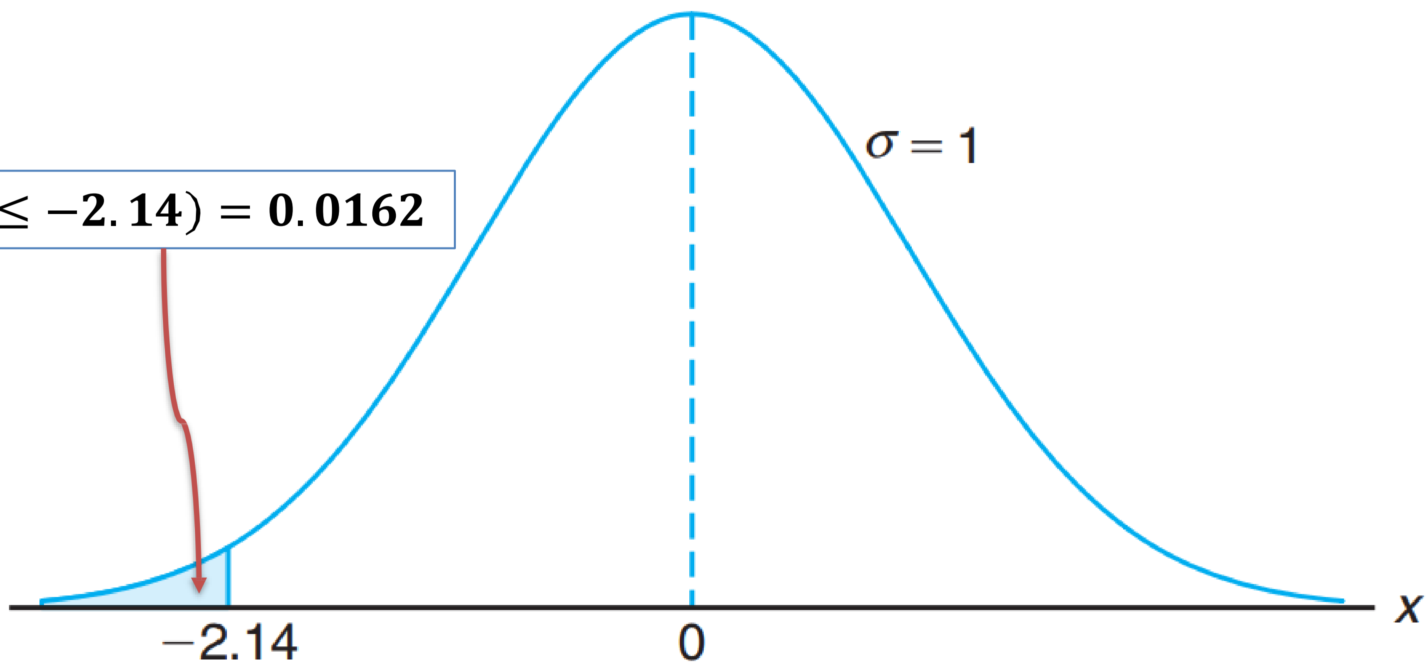
$$P(X<30)=P(X\le 29)\approx P\left(Z\le \frac{29.5-40}{4.899}\right)$$

$$\approx P(Z\le -2.14)=0.0162$$

Normal Distribution (39/41)

Example2 (5/5): $b(n = 100, p = 0.4),\ \mu = 40,\ \sigma = 4.899$

$P(Z \leq -2.14) = 0.0162$



Normal Distribution (40/41)

Normal Approximation to the Poisson Distribution (1/2):

If X is a Poisson random variable with parameter λ,

Because $\mu = \lambda$ and $\sigma^2 = \lambda$ for Poisson, then

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is a standard normal random variable

with mean = 0 and variance = 1.

Provides a very accurate approximation to the Poisson distribution when $\lambda > 5$

Normal Approximation to the Poisson Distribution (2/2):

To approximate a Poisson probability with a normal distribution, a *continuity correction* is applied as follows:

$$P(X \leq x) = P(X \leq x + 0.5) \approx P\left(Z \leq \frac{(x + 0.5) - \lambda}{\sqrt{\lambda}}\right)$$

and

$$P(X \geq x) = P(X \geq x - 0.5) \approx P\left(Z \geq \frac{(x - 0.5) - \lambda}{\sqrt{\lambda}}\right)$$

Example1 (1/4):

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

Example1 (2/4):

$\textbf{Poisson}(\lambda = 1000)$

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a **Poisson distribution with a mean of 1000**. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

Example1 (3/4):

$\mathbf{Poisson}(\lambda = 1000)$

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that **950 or fewer** particles are found?

$$P(X \leq 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^{x}}{x!}$$

Normal Distribution (41/41)

Example1 (4/4):

Poisson($\lambda = 1000$)

$$P(X \leq 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^x}{x!}$$

$$P(X \leq 950) \approx P\left(Z \leq \frac{(950 + 0.5) - 1000}{\sqrt{1000}}\right)$$

$$\approx P(Z \leq -1.57) = 0.058$$

Ch 4: Descriptive Statistics

Data Description (1/2)

- Population and Sample in a Statistical Study.
- Types of Data.
- Representation of Data.
- Numerical Summaries of Quantitative Data.
  - Measures of Centrality (Central Tendency/location)
  - Measures of Variability (Dispersion/Spread)
  - Measures of Shape. (Skewness, Kurtosis)
  - Measures of Relative Position. (Percentiles, Quartiles, IQR, Coefficient of Variation).

Data Description (2/2)

- Stem-and-Leaf Diagrams.
- Box-Whisker Plot (Box Plot).
- Time Sequence Plots.
- Scatter Diagrams.

Population and Sample (1/3)

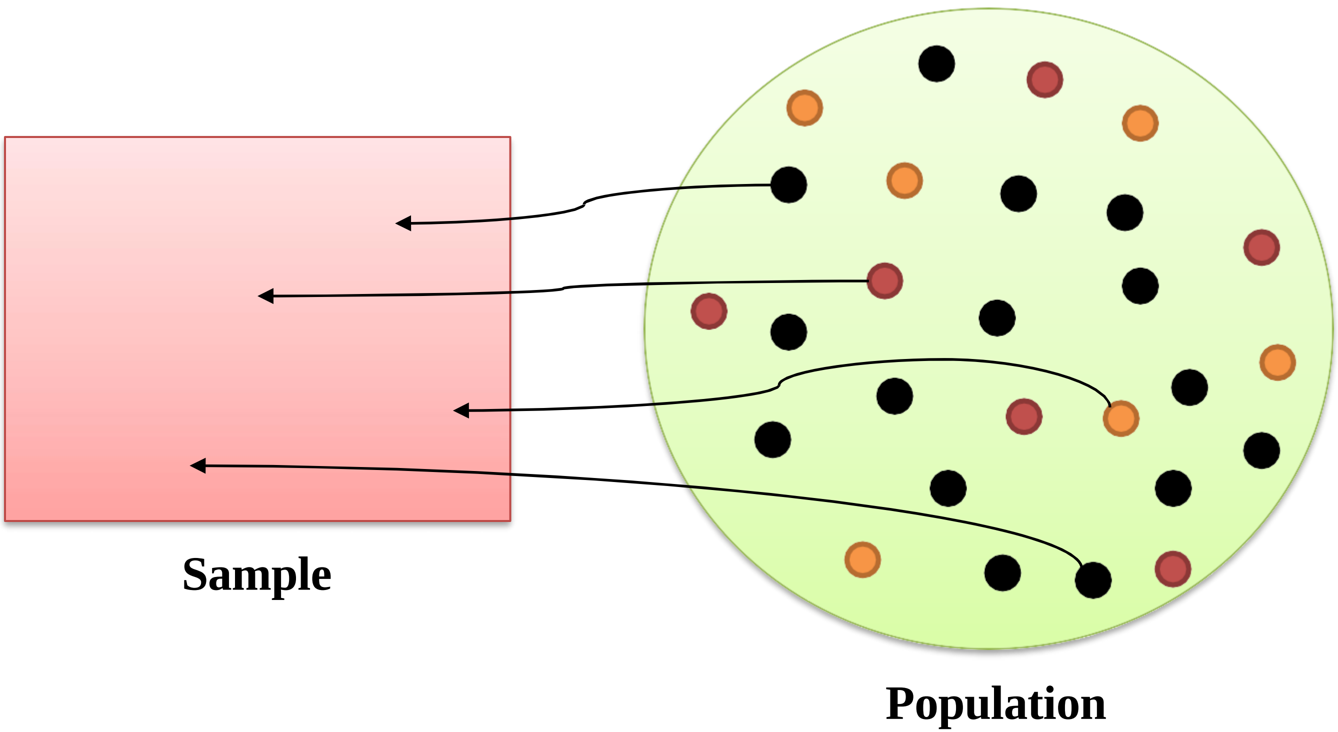
Definitions (1/9):

- A **population** is a collection of **all** elements that possess a characteristic of interest.
- Populations can be *finite* or *infinite*. A population where all the elements are easily countable may be considered as finite, and a population where all the elements are not easily countable as infinite.

Population and Sample (1/3)

Definitions (2/9):

- A **<u>portion</u>** of a population selected for study is called a **sample**.



Population and Sample (1/3)

Definitions (3/9):

First type is called **biased samples**. When one or more parts of the population are favored over others.

- Convenience Sample.
- Voluntary Responses Sample.

Population and Sample (1/3)

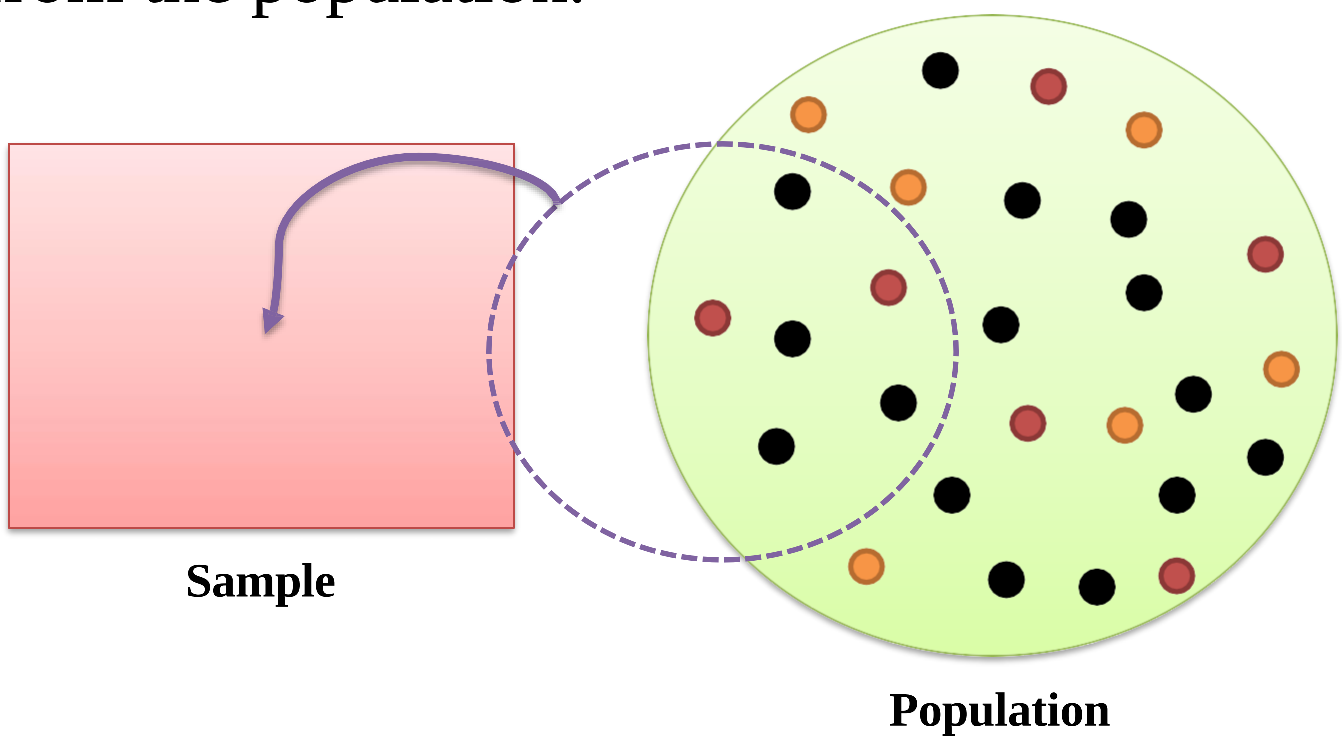
Definitions (4/9):

***Convenience Sample*: only include elements that are easy to reach from the population.**

Population and Sample (1/3)

Definitions (4/9):

***Convenience Sample*:** only include elements that are easy to reach from the population.



Population and Sample (1/3)

Definitions (5/9):

***Voluntary Responses Sample*:** consist of people that have chosen to include themselves (e.g., survey: people with a strong interest for the survey topic are the ones who are most likely to respond).

Population and Sample (1/3)

Definitions (6/9):

In order to avoid having bias, we want our sample to be random, called **unbiased samples**.

- Simple Random Sample (SRS).
- Stratified Random Sample.
- Clustering (Multistage) Random Sample.

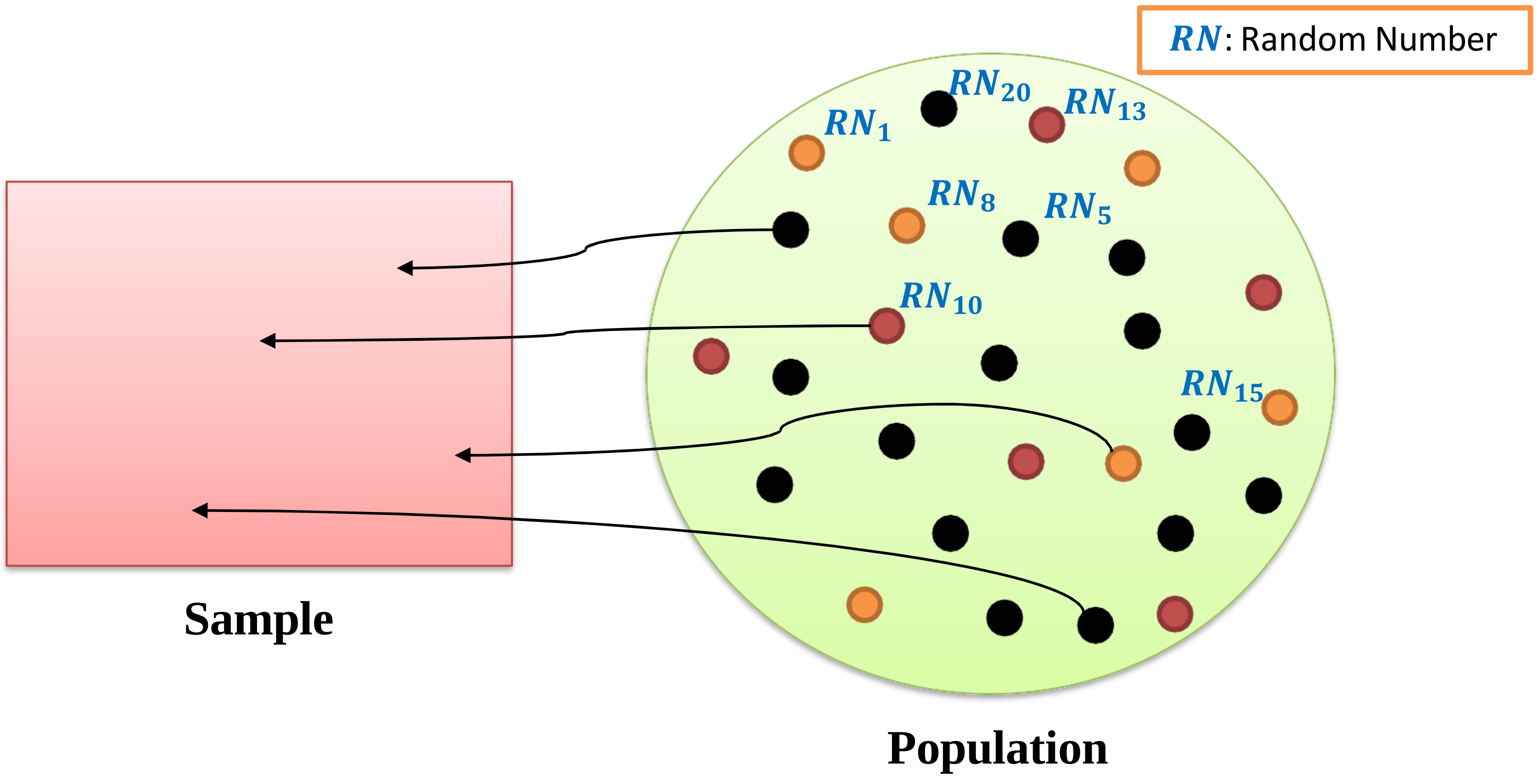
Population and Sample (1/3)

Definitions (7/9):

***Simple Random Sample (SRS)*:** if each element of the population has the same chance of being included in the sample.

Definitions (7/9):

***Simple Random Sample (SRS)*:** if each element of the population has the same chance of being included in the sample.



Population and Sample (1/3)

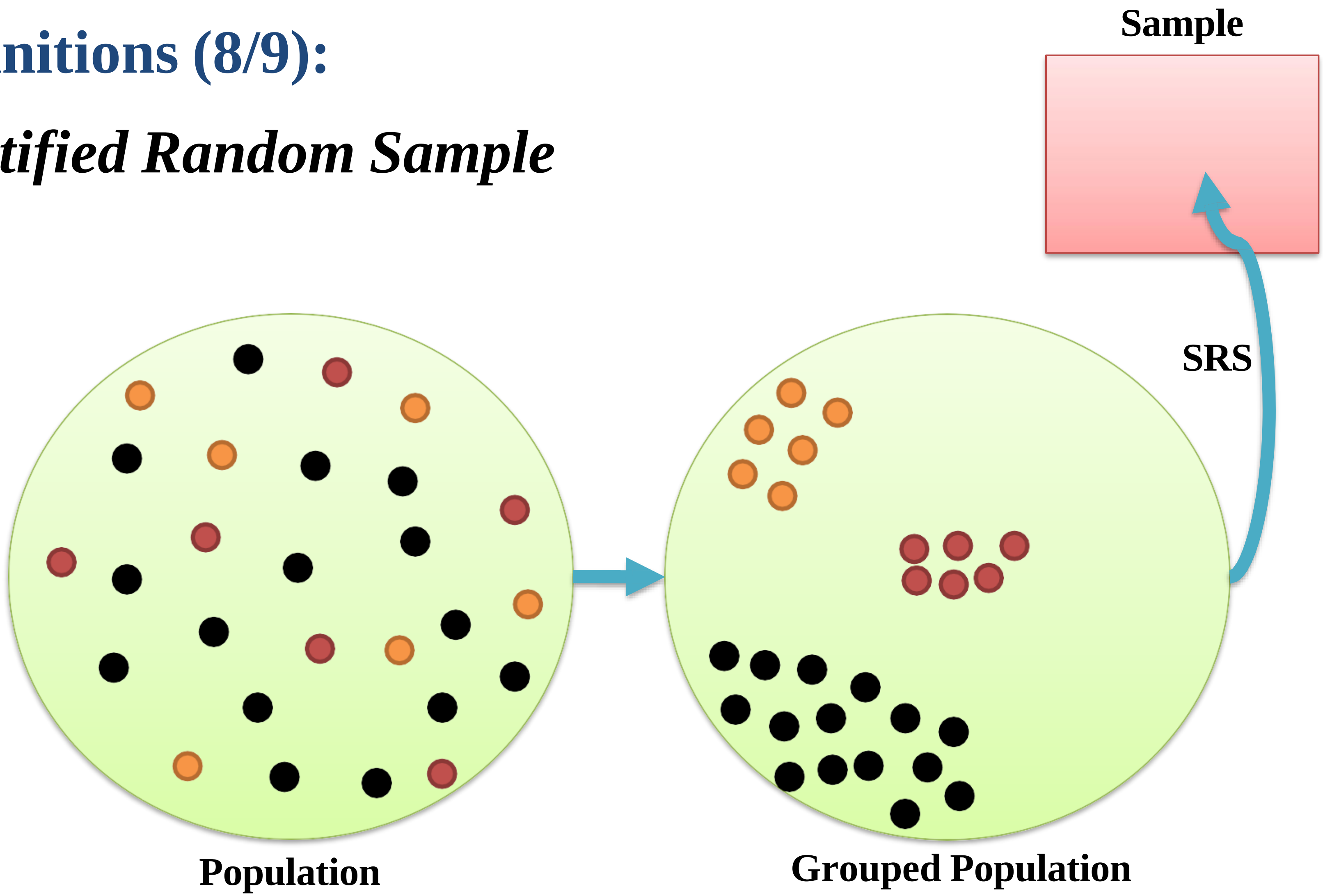
Definitions (8/9):

***Stratified Random Sample*:** elements of the population are divided into groups represent the similar types (called Stratum). Then, The SRS is taken within each stratum to select some items and combine the results to get the sample.

Population and Sample (1/3)

Definitions (8/9):

Stratified Random Sample



Population and Sample (1/3)

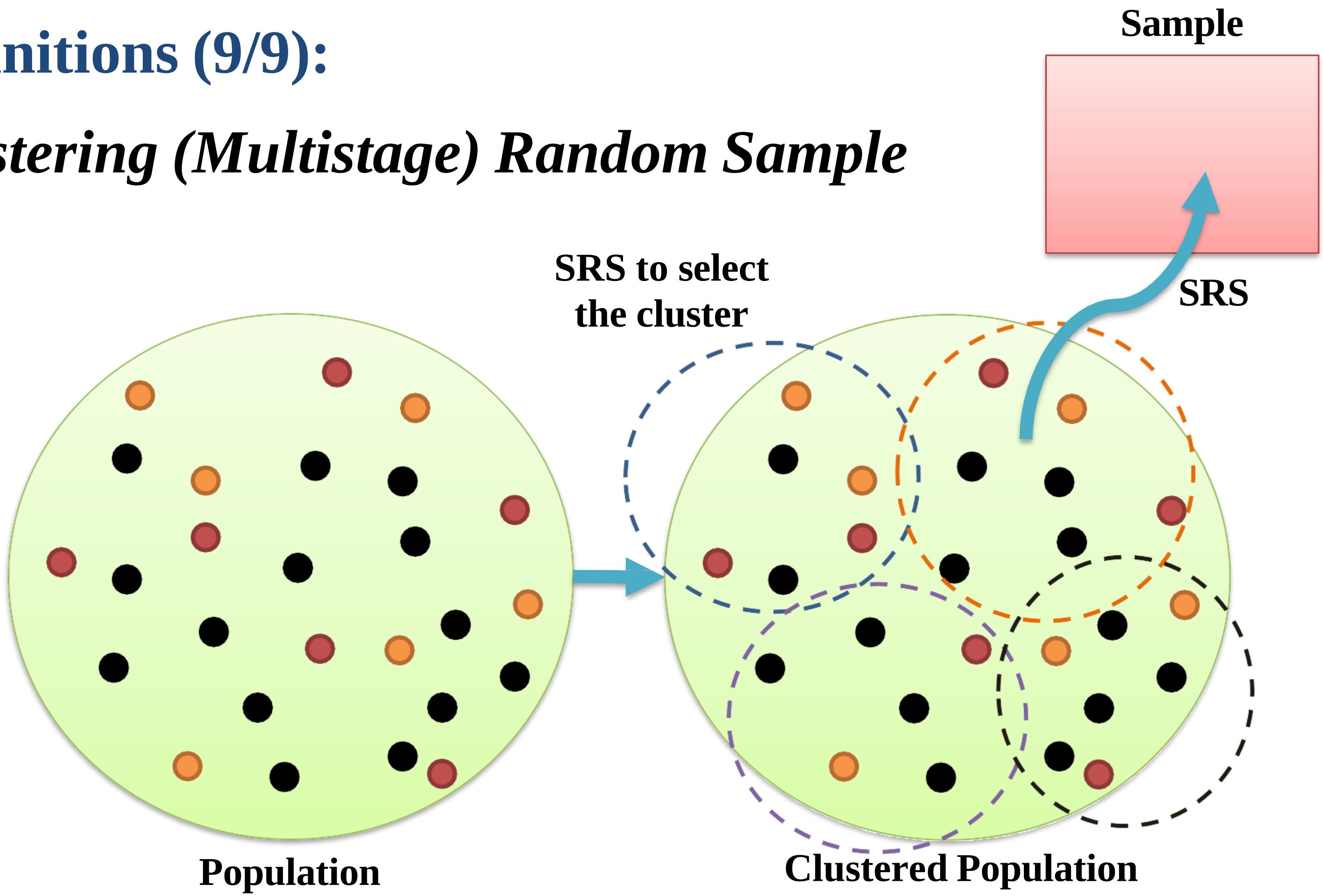
Definitions (9/9):

Clustering (Multistage) Random Sample**:** elements of the population are divided into different clusters. Then, the SRS is taken to select one cluster. After that, the SRS is taken again within the selected cluster to select some items.

Population and Sample (1/3)

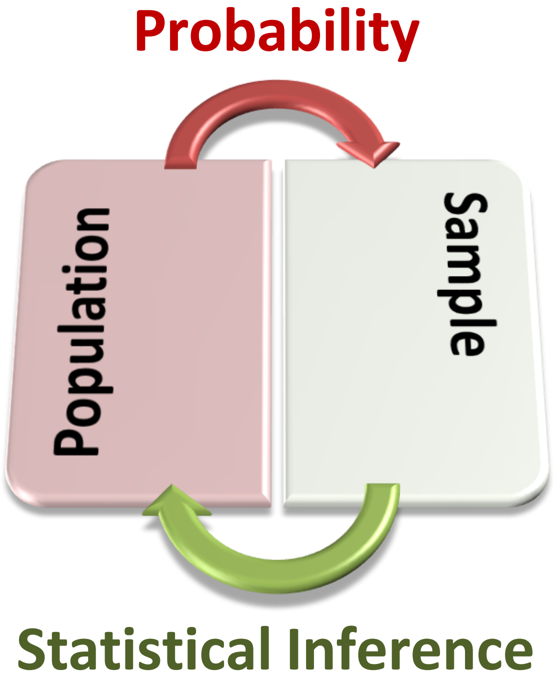
Definitions (9/9):

Clustering (Multistage) Random Sample



Population and Sample (2/3)

Probability And Inferential Statistics:



Population and Sample (3/3)

Statistics:

- Statistics concerns data; their collection, analysis, and interpretation.
- Descriptive statistics concerns the summarization of data. We have a dataset and we would like to describe the data set in multiple ways.
- Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

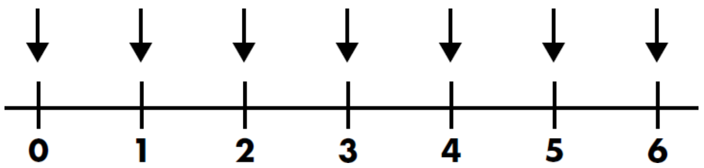
Types of Data (1/6)

Types of Data

- **Quantitative**
  - *Represent numerical quantities*
    - ***Discrete***: take values in a finite or countably infinite set of numbers.
      Examples: counts, number of arrivals, or number of cars.
    - ***Continuous***: take values in an interval of numbers.
      Examples: height, weight, length, time.
- **Qualitative**
  - *Do not represent numerical quantities (e.g., gender, color, ...)*
- **Logical**
  - *True/False*
- **Missing**
  - *Data that should be there but are not*
- **Other Types**

Types of Data (2/6)

Discrete Quantitative Data:



Example:

12 people were asked about their own cars and the results were recorded as follows:

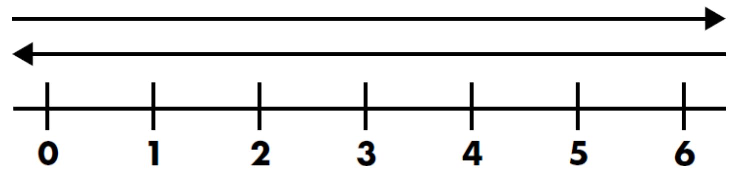
2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Car(s)

The measuring: *numerical*.

Types of Data (3/6)

Continuous Quantitative Data:



<u>Example:</u>

10 people were asked about their height and the results were recorded as follows:

$$\mathbf{175, 184, 160, 170, 173, 170, 175, 165, 171, 176}$$

cm

The measuring: *numerical*.

Types of Data (4/6)

Qualitative Data:

<u>Example:</u>

11 people were asked about their gender and the results were recorded as follows:

$$\boldsymbol{M, F, M, M, M, F, F, M, M, M, M}$$

The measuring: *nominal*.

Types of Data (5/6)

Logical Data:

The value is either TRUE or FALSE (note that equivalently you can use 1 = TRUE, 0 = FALSE).

<u>Example:</u>

Using R

```
> x <- 5:9
> y <- (x < 7.3)
> y
[1]  TRUE  TRUE  TRUE  FALSE  FALSE
```

Types of Data (6/6)

Missing Data:

Data that should be there but are not. **R** reserves the special symbol `NA` to representing missing data.

Example:

Using R

```
> x <- c(3, 7, NA, 4, 7)
> y <- c(5, NA, 1, 2, 2)
> x + y
[1]  8  NA  NA  6  9
> sum(x, na.rm = TRUE)
[1]  21
```

Revision Questions

Determine whether each statement is true or false:

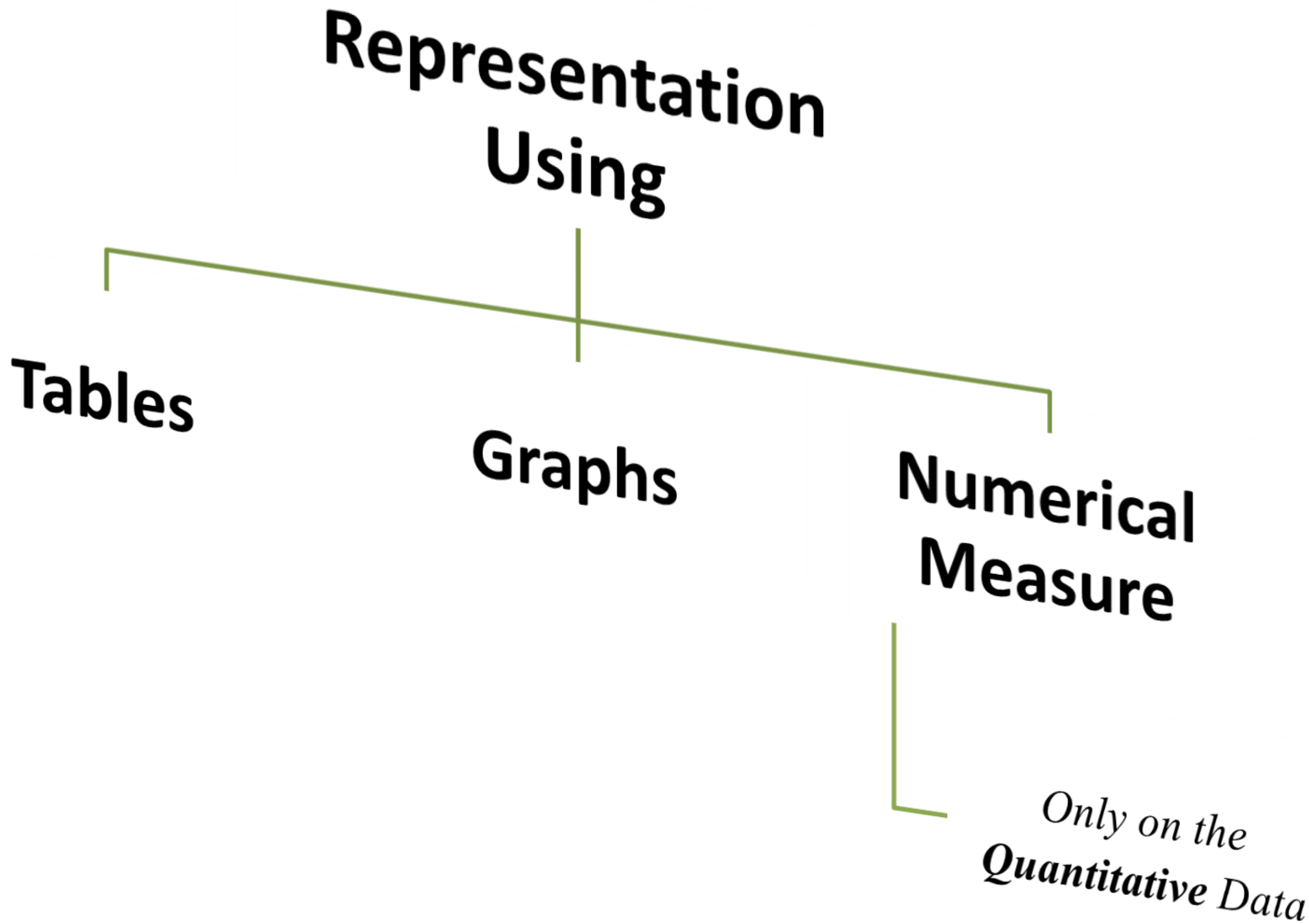
1. A sample consists of all subjects that are being studies. (   )
2. Descriptive statistics consists of the collection, organization, summarization, and presentation of data. (   )
3. The number of absences per year that a worker has is an example a continuous variable. (   )
4. The variable age is an example of discrete variable. (   )
5. Data that can be classified according to color are measured as nominal. (   )

Revision Questions

Determine whether each statement is true or false:

1. A sample consists of all subjects that are being studies. (F) (population)
2. Descriptive statistics consists of the collection, organization, summarization, and presentation of data. (T)
3. The number of absences per year that a worker has is an example a continuous variable. (F) (discrete)
4. The variable age is an example of discrete variable. (F) (continuous)
5. Data that can be classified according to color are measured as nominal. (T)

Representation of Data (1/8)



Representation of Data (2/8)

Example1: (1/6)

11 people were asked about their gender and the results were recorded as follows:

$$\boldsymbol{M, F, M, M, M, F, F, M, M, M, M}$$

The type of this data raw is: ***Qualitative Data***

Representation of Data (2/8)

Example1: (2/6)

$$M, F, M, M, M, F, F, M, M, M, M$$

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)
M			
F			
Sum			

Representation of Data (2/8)

Example1: (2/6)

$$M, F, M, M, M, F, F, M, M, M, M$$

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)
M	~~||||~~ |||	8	8/11
F	|||	3	3/11
Sum	~~||||~~ ~~||||~~ |	**11**	**1**

Representation of Data (2/8)

Example1: (2/6)

$M, F, M, M, M, F, F, M, M, M, M$

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)
M	~~||||~~ |||	8	8/11
F	|||	3	3/11
Sum	~~||||~~ ~~||||~~ |	**11**	**1**

sample size

Representation of Data (2/8)

Example1: (3/6)

$$\boldsymbol{M, F, M, M, M, F, F, M, M, M, M}$$

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
M	$\cancel{||||}\ |||$	8	8/11	
F	$|||$	3	3/11	
Sum	$\cancel{||||}\ \cancel{||||}\ |$	**11**	**1**	

Representation of Data (2/8)

Example1: (3/6)

$$\boldsymbol{M, F, M, M, M, F, F, M, M, M, M}$$

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
M	$\cancel{||||}\ |||$	8	8/11	72.73%
F	$|||$	3	3/11	27.27%
Sum	$\cancel{||||}\ \cancel{||||}\ |$	**11**	**1**	**100%**

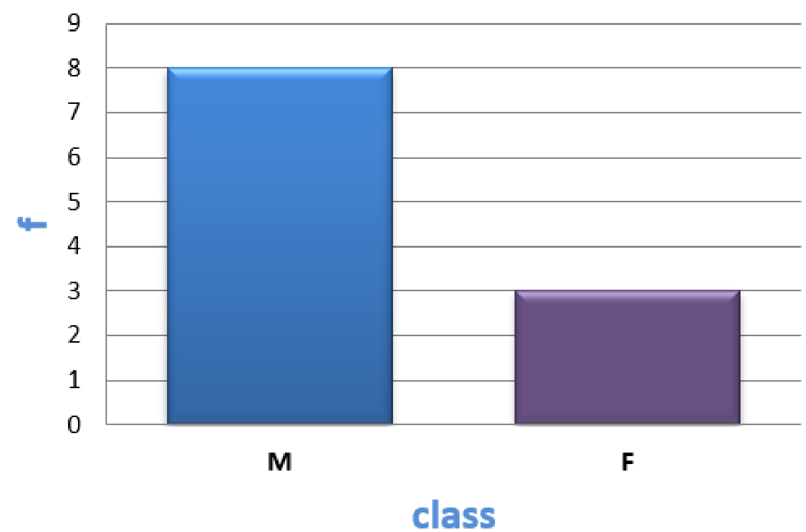
Representation of Data (2/8)

Example1: (4/6)

$$M, F, M, M, M, F, F, M, M, M, M$$

Using Bar Graphs

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
M	~~////~~ ///	8	8/11	72.73%
F	///	3	3/11	27.27%
Sum	~~////~~ ~~////~~ /	**11**	**1**	**100%**



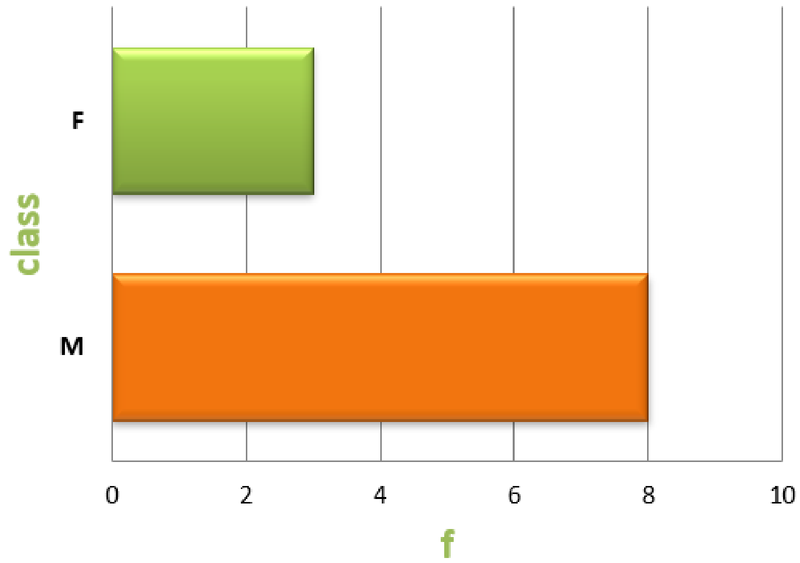
Representation of Data (2/8)

Example1: (5/6)

$$M, F, M, M, M, F, F, M, M, M, M$$

Using Bar Graphs

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
M	$\cancel{||||}\,|||$	8	8/11	72.73%
F	$|||$	3	3/11	27.27%
Sum	$\cancel{||||}\,\cancel{||||}\,|$	**11**	**1**	**100%**



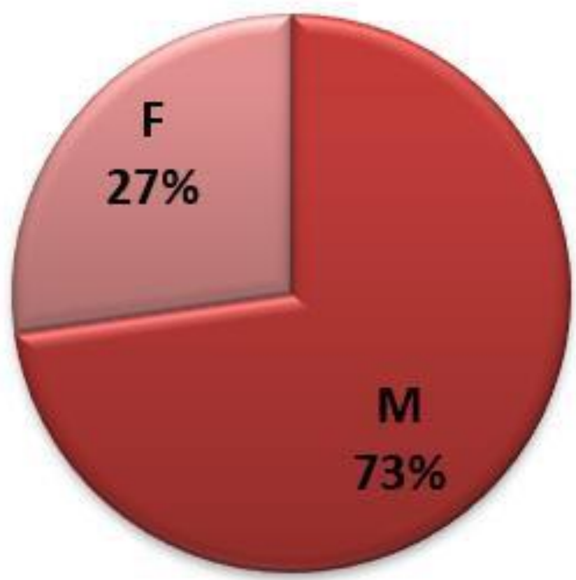
Representation of Data (2/8)

Example1: (6/6)

$M, F, M, M, M, F, F, M, M, M, M$

Using Pie Charts

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
M	𝍸 |||	8	8/11	72.73%
F	|||	3	3/11	27.27%
Sum	𝍸 𝍸		**11**	**1**	**100%**



Angle of a slice = (Relative frequency of the given class) × 360

Ex. Angle of a slice for class M = 8/11 * 360 = 261.81°

Representation of Data (3/8)

Example2: (1/4)

A study of 552 first-year college students asked about their preferences for online resources. One question asked them to pick their favorite. Here are the results

Resource	Count (n)
Google or Google Scholar	406
Library database or website	75
Wikipedia or online encyclopedia	52
Other	19
Total	552

The type of this data raw is: ***Qualitative Data***

Representation of Data (3/8)

Example2: (2/4)

Resource	Count (n)
Google or Google Scholar	406
Library database or website	75
Wikipedia or online encyclopedia	52
Other	19
Total	552

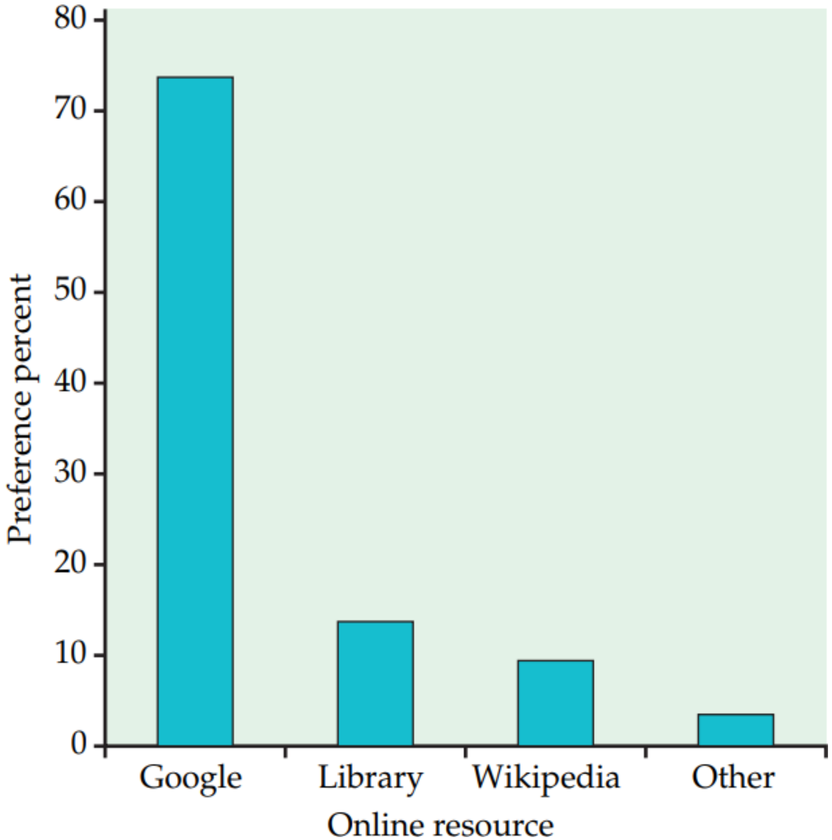
Resource	Percent (%)
Google or Google Scholar	73.6
Library database or website	13.6
Wikipedia or online encyclopedia	9.4
Other	3.4
Total	100.0

Representation of Data (3/8)

Example2: (3/4)

Using Bar Graphs

Resource	Percent (%)
Google or Google Scholar	73.6
Library database or website	13.6
Wikipedia or online encyclopedia	9.4
Other	3.4
Total	100.0

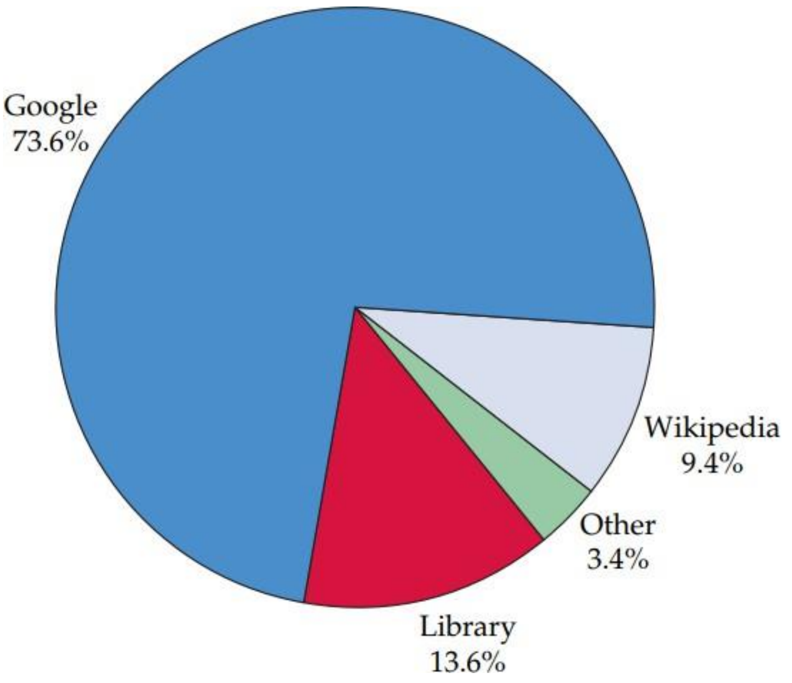


Representation of Data (3/8)

Example2: (4/4)

Using Pie Charts

Resource	Percent (%)
Google or Google Scholar	73.6
Library database or website	13.6
Wikipedia or online encyclopedia	9.4
Other	3.4
Total	100.0



Angle of a slice = (Percent%) × 3.6

Ex. Angle of a slice for class Wikipedia = 9.4* 3.6 = 33.84°

Representation of Data (4/8)

Example3: (1/7)

12 people were asked about their own cars and the results were recorded as follows:

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Car(s)

The type of this data raw is: ***Discrete Quantitative Data***

Representation of Data (4/8)

Example3: (2/7)

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)
0			
2			
3			
4			
Sum			

Representation of Data (4/8)

Example3: (2/7)

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)
0	/	1	1/12
2	~~////~~ ///	8	8/12
3	/	1	1/12
4	//	2	2/12
Sum	~~////~~ ~~////~~	12	1

sample size

Representation of Data (4/8)

Example3: (3/7)

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Using Frequency Table (Grouped Data)

Called: *Frequency distribution for the data.*

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
0	/	1	1/12	08.33 %
2	~~////~~///	8	8/12	66.67 %
3	/	1	1/12	08.33 %
4	//	2	2/12	16.67 %
Sum	~~////~~ ~~////~~	12	1	100 %

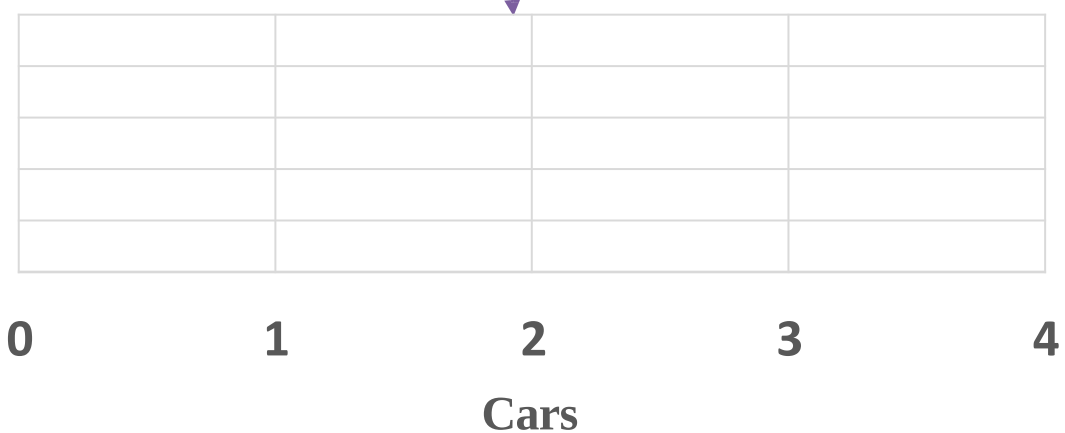
Representation of Data (4/8)

Example3: (4/7)

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Dot Plot

A dot plot is one of the simplest graphs.



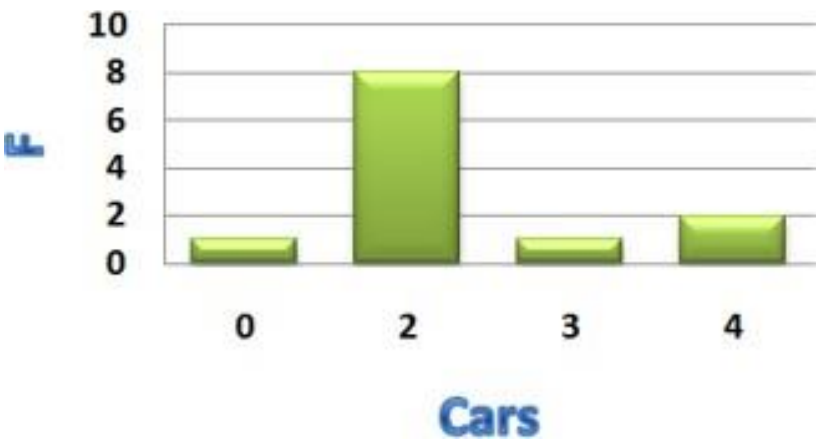
Representation of Data (4/8)

Example3: (5/7)

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Using Bar Graphs

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
0	/	1	1/12	08.33 %
2	~~////~~///	8	8/12	66.67 %
3	/	1	1/12	08.33 %
4	//	2	2/12	16.67 %
Sum	~~////~~ ~~////~~	12	1	100 %



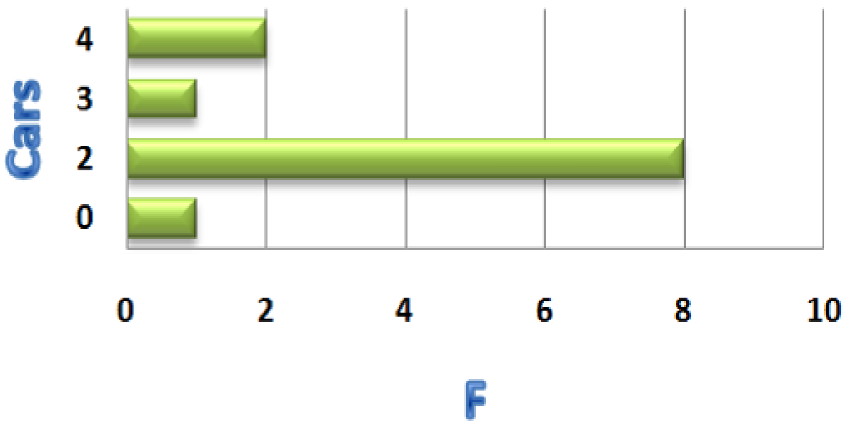
Representation of Data (4/8)

Example3: (6/7)

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Using Bar Graphs

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
0	/	1	1/12	08.33 %
2	~~////~~///	8	8/12	66.67 %
3	/	1	1/12	08.33 %
4	//	2	2/12	16.67 %
Sum	~~////~~ ~~////~~	12	1	100 %



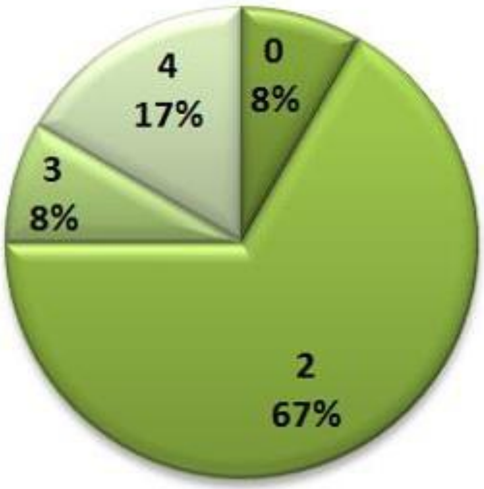
Representation of Data (4/8)

Example3: (7/7)

2, 0, 4, 2, 2, 3, 2, 2, 4, 2, 2, 2

Using Pie Charts

Class	Tally	Frequency (F)	Relative Frequency (R.F.)	Percentage
0	/	1	1/12	08.33 %
2	~~////~~ ///	8	8/12	66.67 %
3	/	1	1/12	08.33 %
4	//	2	2/12	16.67 %
Sum	~~////~~ ~~////~~	12	1	100 %



Angle of a slice = (Relative frequency of the given class) × 360

Ex. Angle of a slice for class 0 = 1/12 * 360 = 30°

Representation of Data (5/8)

Example4: (1/2)

The following data give the number of defective motors received in 20 different shipments:

8	12	10	16	10	25	21	15	17	5
26	21	29	8	6	21	10	17	15	13

Construct a dot plot for these data.

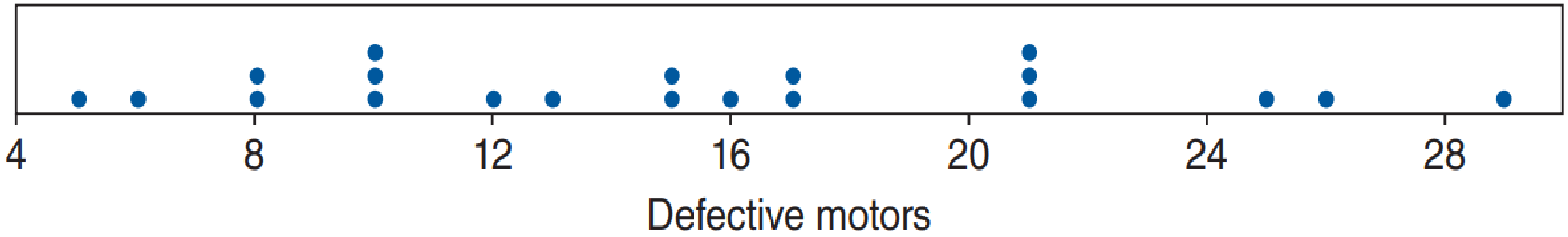
The type of this data raw is: ***Discrete Quantitative Data***

Representation of Data (5/8)

Example4: (2/2)

8	12	10	16	10	25	21	15	17	5
26	21	29	8	6	21	10	17	15	13

Dot Plot



Representation of Data (6/8)

Example5: (1/9)

10 people were asked about their height and the results were recorded as follows:

$$\mathbf{175}, \mathbf{184}, \mathbf{160}, \mathbf{170}, \mathbf{173}, \mathbf{170}, \mathbf{175}, \mathbf{165}, \mathbf{171}, \mathbf{176}$$

cm

The type of this data raw is: ***Continuous Quantitative Data***

Representation of Data (6/8)

Example5: (2/9)

$$\mathbf{175, 184, 160, 170, 173, 170, 175, 165, 171, 176}$$

Because the type of this data is continuous quantitative, we follow the following steps:

1. Find the range R of the data that is defined as:

$$\boxed{\textbf{Range} = R = \text{largest data point} - \text{smallest data point}}$$

Representation of Data (6/8)

Example5: (3/9)

$$\boxed{\mathbf{175, 184, 160, 170, 173, 170, 175, 165, 171, 176}}$$

2. Divide the data set into an appropriate number of ***classes***. The classes are also sometimes called *intervals*, *categories*, *cells*, or *bins*.

The number of classes is k. There is no "best" number of classes. However, Sturges's formula is often used, given by:

$$\boxed{\text{Number of classes} = k = \lfloor 1 + 3.322 \log n \rfloor}$$

Or we can use a simple formula as follows:

$$\boxed{k = \sqrt{n}}$$

where n is the total number of data points in a given data set.

Representation of Data (6/8)

Example5: (4/9)

$$\mathbf{175, 184, 160, 170, 173, 170, 175, 165, 171, 176}$$

3. Determine the width of classes as follows:

$$\boxed{\text{Class width} = \text{Range / Number of Classes} = R/k}$$

4. Finally, preparing the frequency distribution table is achieved by assigning each data point to an appropriate class.

Representation of Data (6/8)

Example5: (5/9)

$\boxed{\mathbf{175, 184, 160, 170, 173, 170, 175, 165, 171, 176}}$

1. Range $= R = 184 - 160 = 24$

2. The number of classes $k = \lfloor 1 + 3.322 \log 10 \rfloor = 4$

3. Class width $= R\text{T}k = 24\text{T}4 = 6$

4. The four classes used to prepare the frequency distribution table are as follows:

$$[160 - 166), [166 - 172), [172 - 178), [178 - 184]$$

Representation of Data (6/8)

Example5: (6/9)

175, 184, 160, 170, 173, 170, 175, 165, 171, 176

Using Frequency Table (Grouped Data)

Class	Tally	Frequency (F)	Relative Frequency (R.F.)
$[160 - 166)$			
$[166 - 172)$			
$[172 - 178)$			
$[178 - 184]$			
Sum			

Representation of Data (6/8)

Example5: (6/9)

175, 184, 160, 170, 173, 170, 175, 165, 171, 176

Using Frequency Table (Grouped Data)

Class	Tally	Frequency (F)	Relative Frequency (R.F.)
[160 – 166)	//	2	2/10
[166 – 172)	///	3	3/10
[172 – 178)	////	4	4/10
[178 – 184]	/	1	1/10
Sum	~~////~~ ~~////~~	**10**	**1**

Representation of Data (6/8)

Example5: (7/9)

$$175, 184, 160, 170, 173, 170, 175, 165, 171, 176$$

Histograms

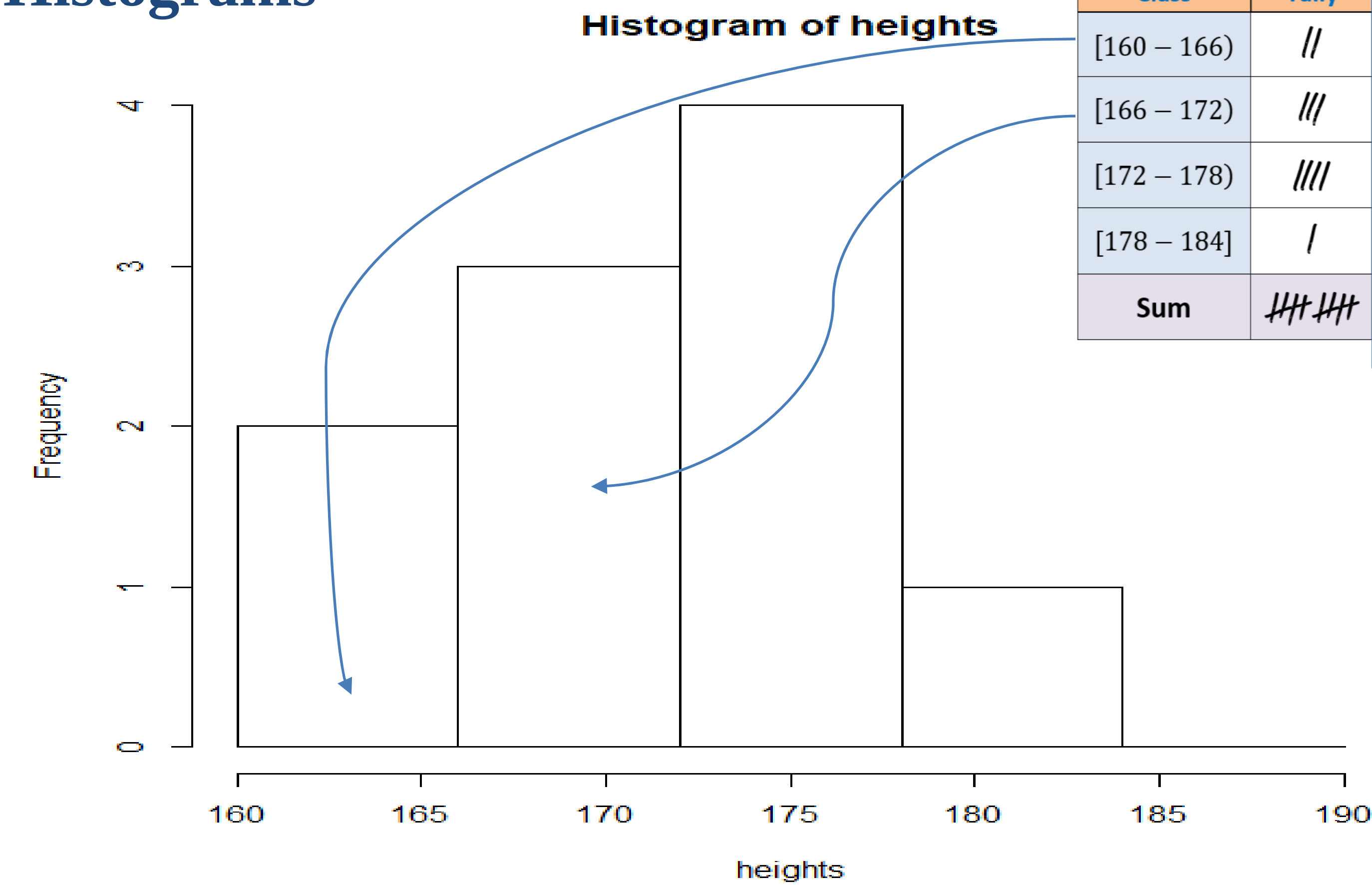
A histogram is a graphical tool consisting of bars placed side by side on a set of intervals (classes, bins, or cells) of equal width. The bars represent the frequency or relative frequency of classes. The height of each bar is proportional to the frequency or relative frequency of the corresponding class.

Representation of Data (6/8)

Example5: (7/9)

175, 184, 160, 170, 173, 170, 175, 165, 171, 176

Histograms



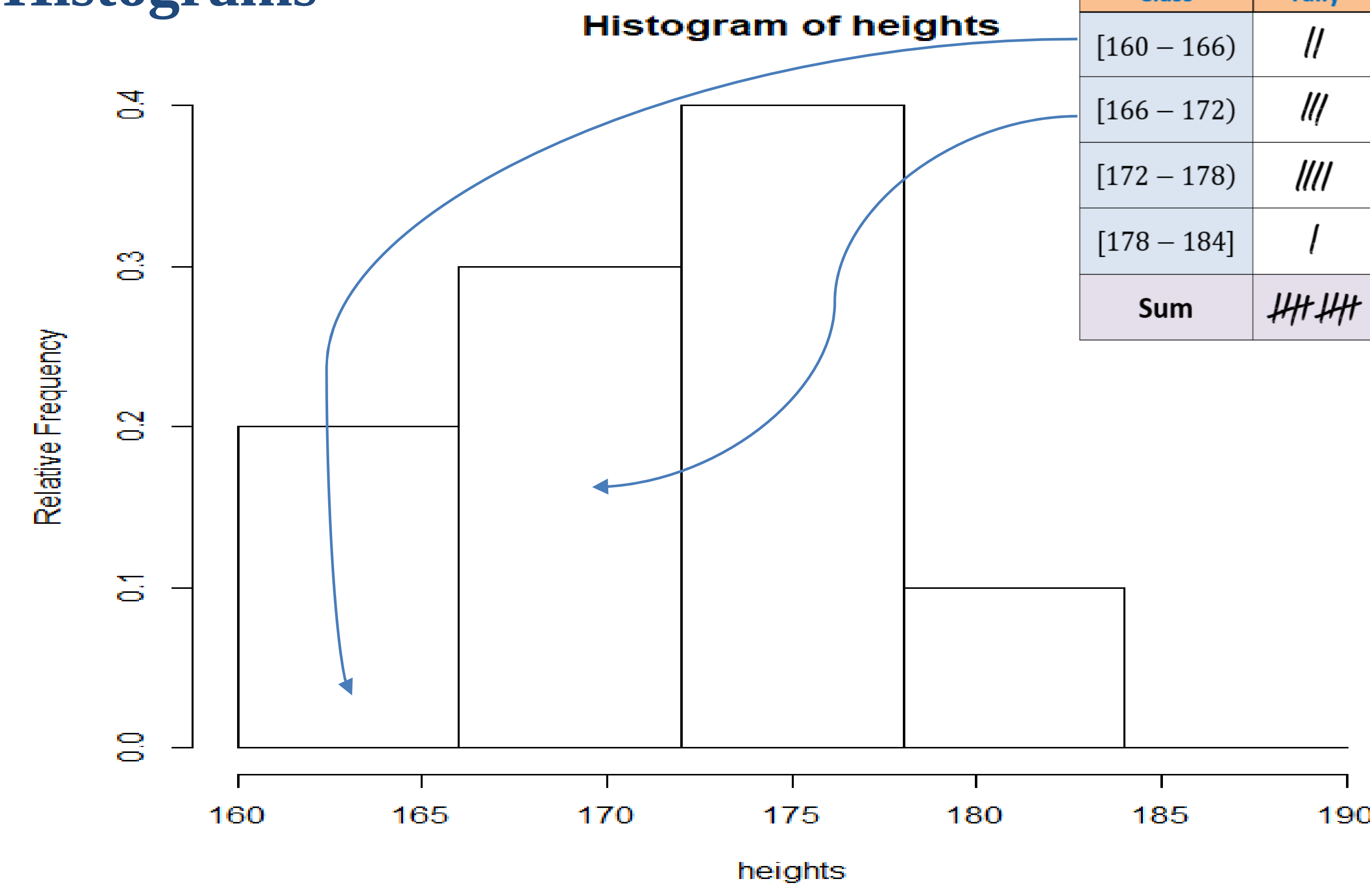
Class	Tally	Frequency (F)	Relative Frequency (R.F.)
[160 – 166)	//	2	2/10
[166 – 172)	///	3	3/10
[172 – 178)	////	4	4/10
[178 – 184]	/	1	1/10
Sum	~~////~~ ~~////~~	**10**	**1**

Representation of Data (6/8)

Example5: (8/9)

175, 184, 160, 170, 173, 170, 175, 165, 171, 176

Histograms



Class	Tally	Frequency (F)	Relative Frequency (R.F.)
[160 – 166)	//	2	2/10
[166 – 172)	///	3	3/10
[172 – 178)	////	4	4/10
[178 – 184]	/	1	1/10
Sum	~~////~~ ~~////~~	**10**	**1**

Representation of Data (6/8)

Example5: (9/9)

175, 184, 160, 170, 173, 170, 175, 165, 171, 176

<u>Example:</u>

Using R

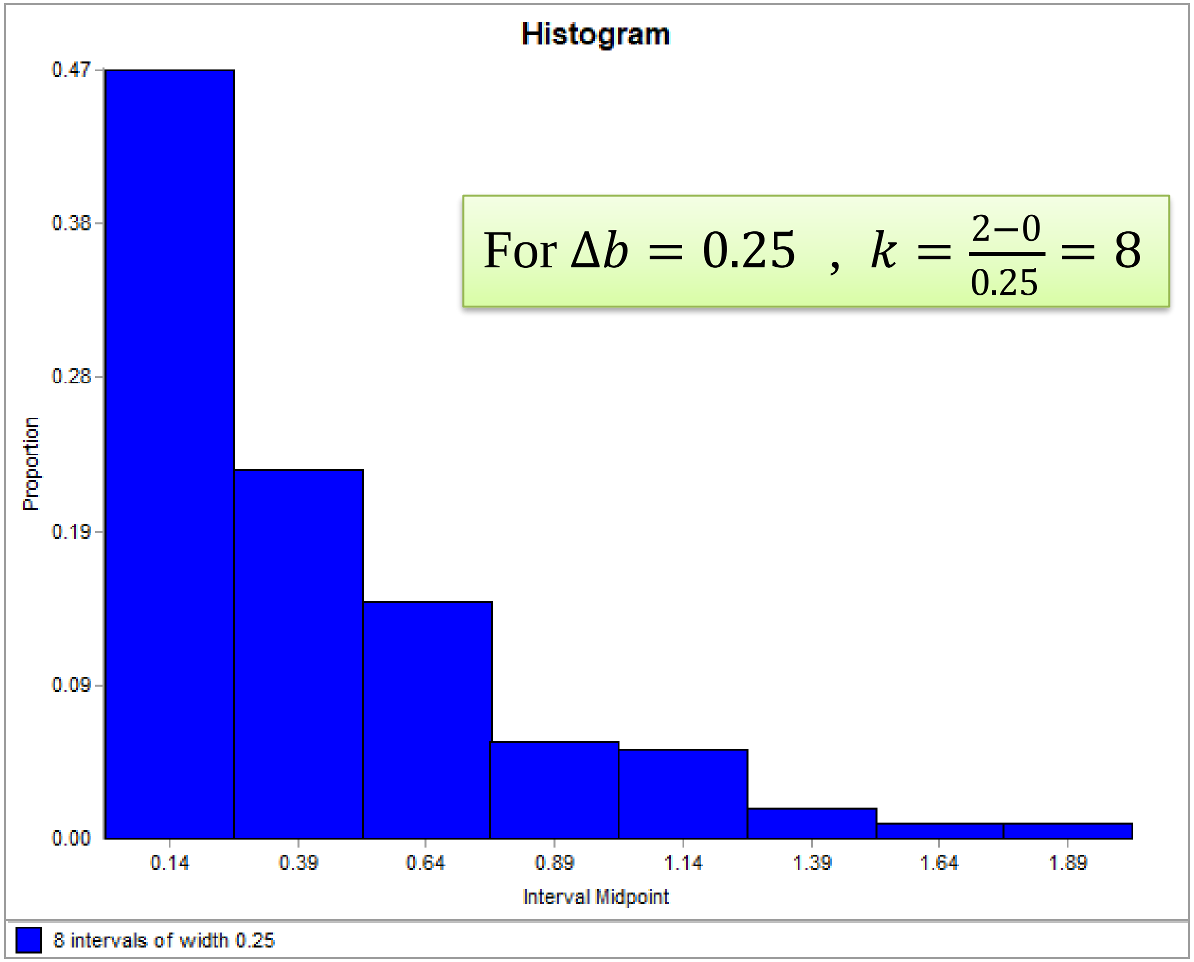
```
> heights <- c(175,184,160,170,173,170,165,175,171,176)

> hist(heights, breaks=seq(160, 190, by=6))

> install.packages(“HistogramTools”)
> library(HistogramTools)
> PlotRelativeFrequency(hist(heights, breaks=seq(160, 190, by=6)))
```

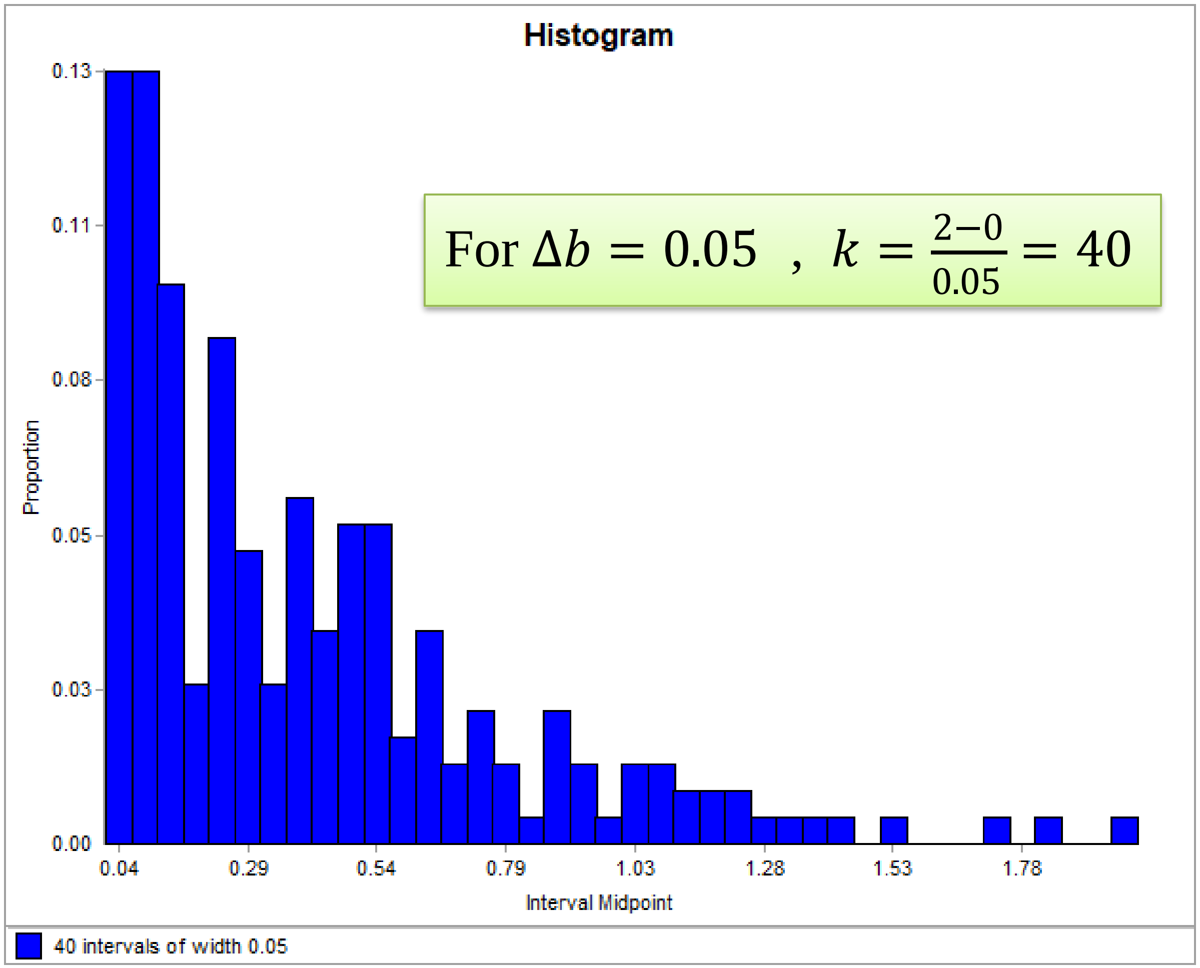
Representation of Data (7/8)

Histograms:



Representation of Data (7/8)

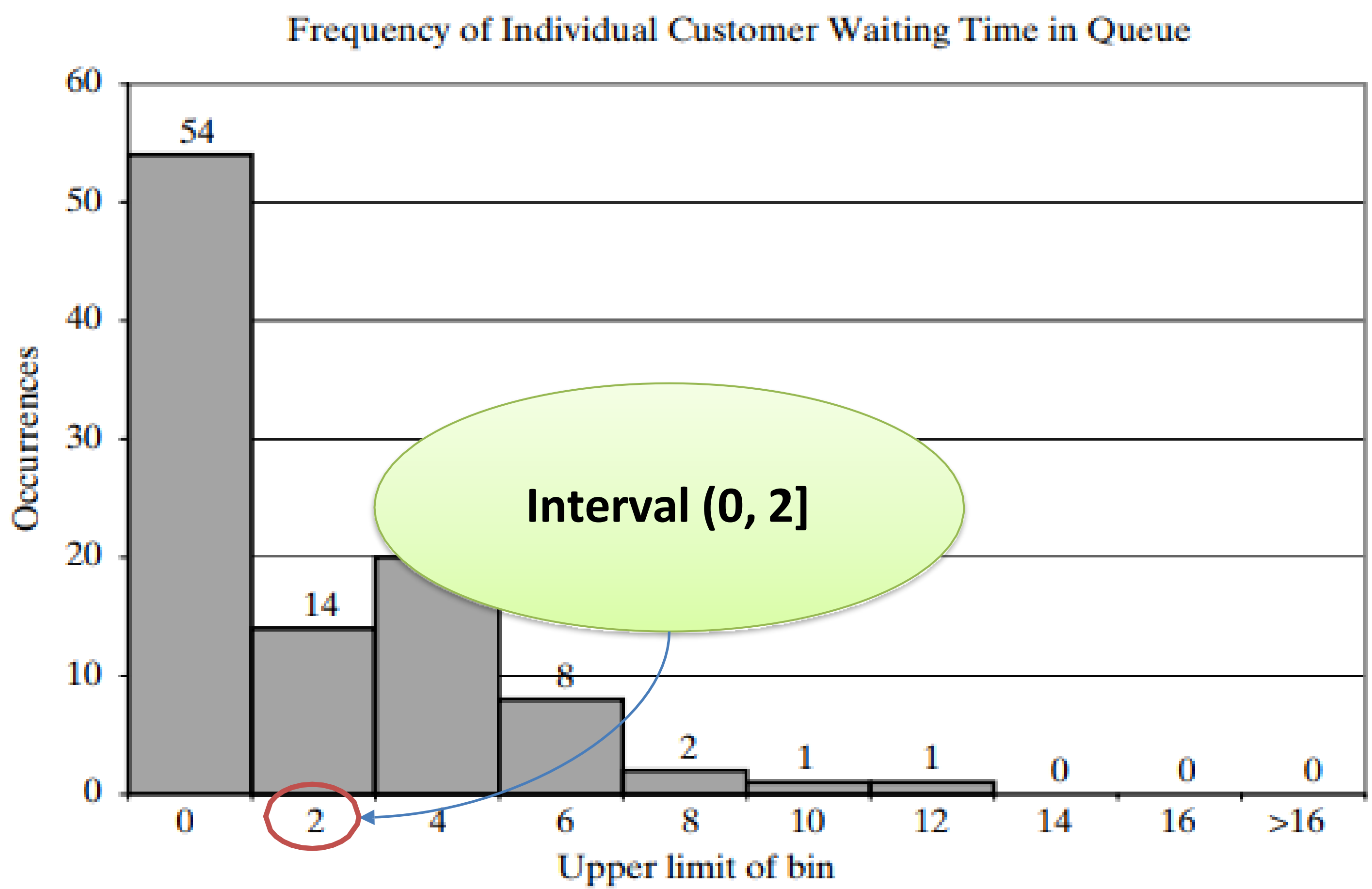
Histograms:



For $\Delta b = 0.05$, $k = \frac{2-0}{0.05} = 40$

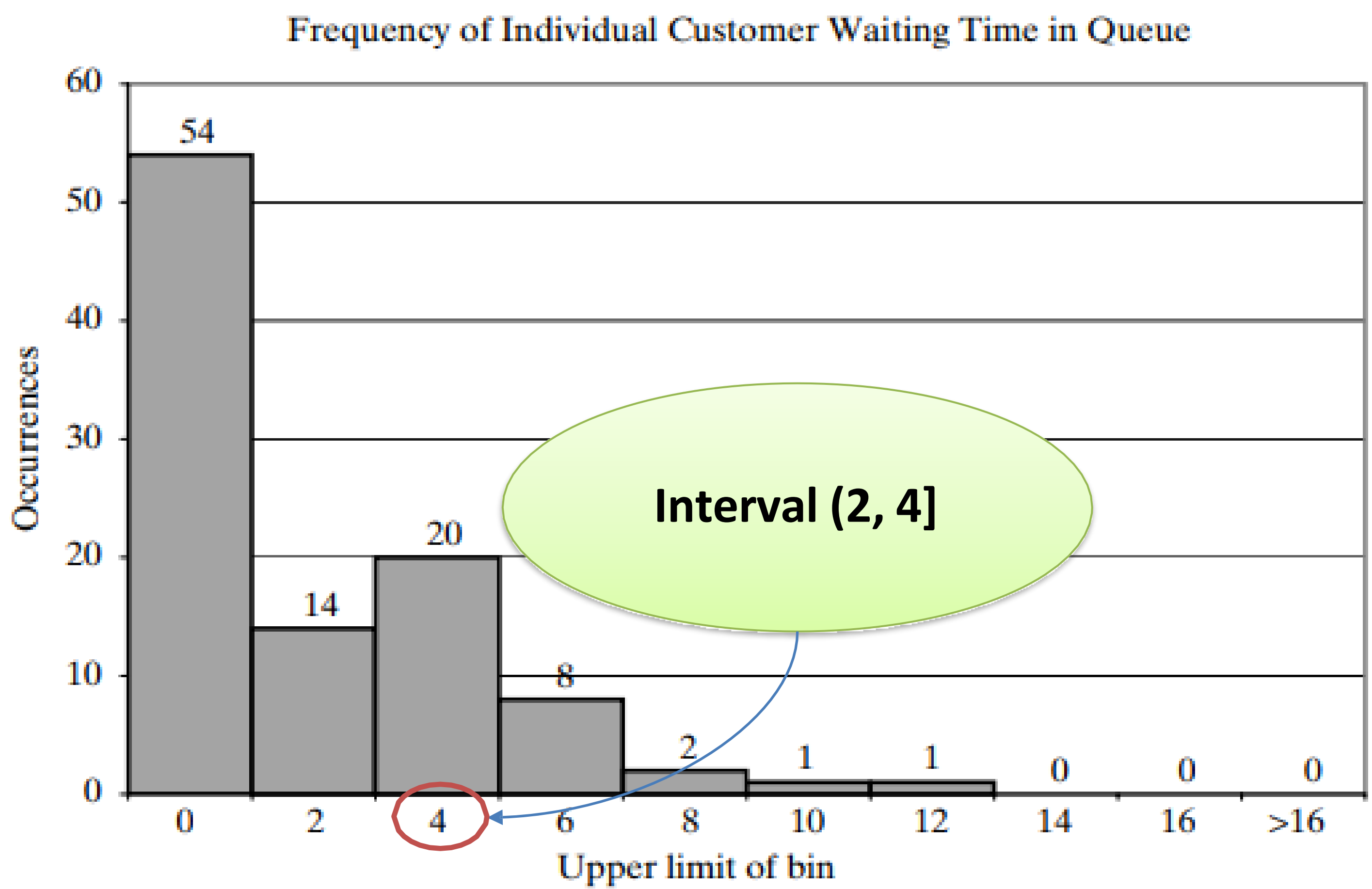
Representation of Data (8/8)

Histograms:



Representation of Data (8/8)

Histograms:



Revision Questions (1/2)

Given the following frequency table represent the colors represented in a sample:

Colors	Frequency
Red	5
Blue	8
Green	2

1. What is the type of data?
2. What is the sample size?
3. What graph can use to represent these data?
4. What is the most frequency colored?

Revision Questions (1/2)

Given the following frequency table represent the colors represented in a sample:

Colors	Frequency
Red	5
Blue	8
Green	2

1. What is the type of data? (Qualitative Data)
2. What is the sample size? = 15
3. What graph can use to represent these data? (Bar/Pie Charts)
4. What is the most frequency colored? “Blue”

Revision Questions (2/2)

Given the following table:

Classes	Frequency	Relative frequency
15 – 20	8	
20 – 25	10	
25 – 30	7	
30 – 35	5	

1. What is the type of data? ------
2. Complete the table
3. The sample size is ------
4. The table name is ------
5. The class interval = ------
6. The most frequency class = ------
7. The midpoint for the second class = ------
8. The name of graph represent this table is ------

Revision Questions (2/2)

Given the following table:

Classes	Frequency	Relative frequency
15 – 20	8	8/30
20 – 25	10	10/30
25 – 30	7	7/30
30 – 35	5	5/30

1. What is the type of data? (Continuous Quantitative Data)
2. Complete the table
3. The sample size = (30)
4. The table name is (Frequency Table)
5. The class interval = (5)
6. The most frequency class = [20 – 25)
7. The midpoint for the second class = (22.5)
8. The name of graph represent this table is (Histogram)

Ch 4: Descriptive Statistics

Data Description (1/2)

- Population and Sample in a Statistical Study.
- Types of Data.
- Representation of Data.
- Numerical Summaries of Quantitative Data.
  - Measures of Centrality (Central Tendency/location)
  - Measures of Variability (Dispersion/Spread)
  - Measures of Shape. (Skewness, Kurtosis)
  - Measures of Relative Position. (Percentiles, Quartiles, IQR, Coefficient of Variation).

Data Description (2/2)

- Stem-and-Leaf Diagrams.
- Box-Whisker Plot (Box Plot).
- Time Sequence Plots.
- Scatter Diagrams.

Recall

Representation Using

Tables

Graphs

Numerical Measure

Only on the ***Quantitative*** *Data*

Numerical Summaries (1/3)

Measures of Centrality (1/2):

(Central Tendency / location)

- Measures of location are designed to provide the analyst with some quantitative values of where the center, or some other location, of data is located.
- We need **one** value to represent the data.

Numerical Summaries (1/3)

Measures of Centrality (2/2):

The following measures are of primary importance:

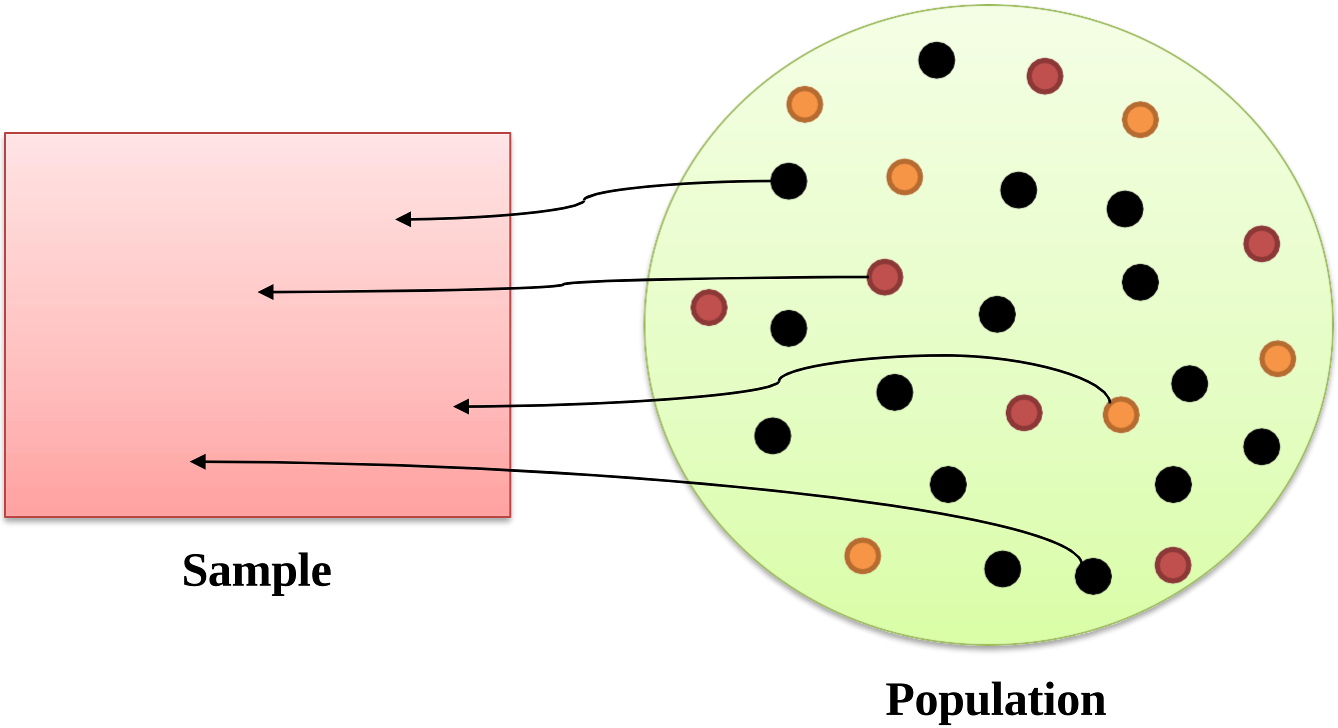
1. Mean
2. Median
3. Mode

The ***mean***, also sometimes referred to as the arithmetic mean, is the most useful and most commonly used measure of centrality. The ***median*** is the second most used, and the ***mode*** is the least used measure of centrality.

Numerical Summaries (2/3)

Measures of Centrality – Mean (1/9):

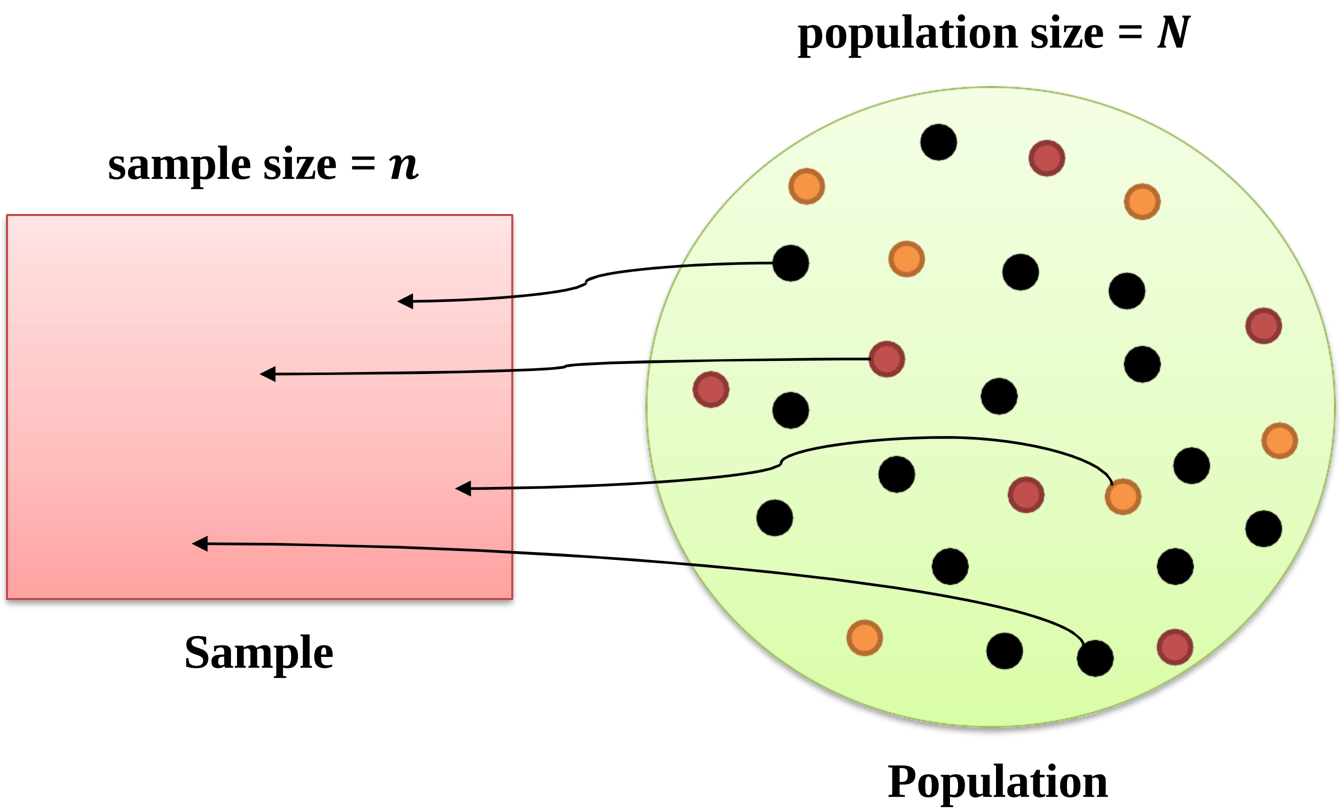
Sample or Population



Numerical Summaries (2/3)

Measures of Centrality – Mean (1/9):

Sample or Population



Numerical Summaries (2/3)

Measures of Centrality – Mean (2/9):

population mean

The population mean is denoted by the Greek letter μ (read as meu), for a finite population with N equally likely values:

$$\mu = \sum_{i=1}^{N} x_i f(x_i) = \frac{\sum_{i=1}^{N} x_i}{N}$$

Numerical Summaries (2/3)

Measures of Centrality – Mean (3/9):

sample mean "sample average"

The **sample mean** is the **average** value of all observations in the data set. Usually, these data are a sample of observations that have been selected from some larger population of observations.

It is denoted by $\bar{x}$ (read as x bar), for n observations:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Numerical Summaries (2/3)

Measures of Centrality – Mean (4/9):

sample mean "sample average"

The sample mean is an **estimate** of the population mean.

the ***degrees of freedom*** associated with the mean estimate

It is denoted by xn(read as x bar), for $\boxed{n \text{ observations:}}$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Numerical Summaries (2/3)

Measures of Centrality – Mean (5/9):

Example 1:

The eight observations are:

$x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$.

The sample mean is:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{8} x_i}{8}$$

$$= \frac{12.6 + 12.9 + \cdots + 13.1}{8} = \frac{104}{8} = 13.0$$

Numerical Summaries (2/3)

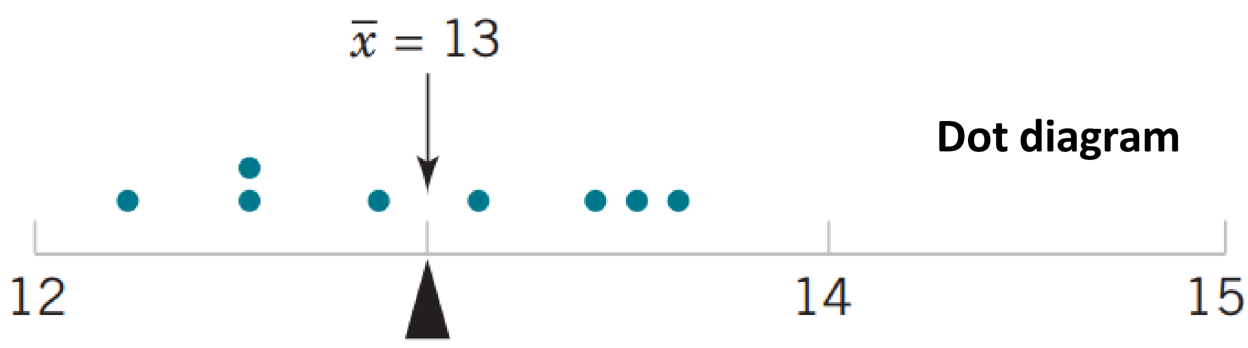
Measures of Centrality – Mean (5/9):

Example 1:

The eight observations are:

$x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$.

The sample mean is:



Measures of Centrality – Mean (6/9):

Example 2:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10 \ 8 \ 15 \ 12 \ 17 \ 7 \ 20 \ 19 \ 22 \ 25 \ 16 \ 15 \ 18 \ 250 \ 300 \ 12$$

Find the mean sale of these individuals.

Numerical Summaries (2/3)

Measures of Centrality – Mean (6/9):

Example 2:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10\ 8\ 15\ 12\ 17\ 7\ 20\ 19\ 22\ 25\ 16\ 15\ 18\ 250\ 300\ 12$$

The mean sale of these individuals is:

$$\bar{x} = \frac{(10+8+15+12+17+7+20+19+22+25+16+15+18+250+300+12)}{16}$$

$$= 47.875$$

Numerical Summaries (2/3)

Measures of Centrality – Mean (6/9):

Mean is sensitive to extreme values (**outliers**)

Example 2:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10\ 8\ 15\ 12\ 17\ 7\ 20\ 19\ 22\ 25\ 16\ 15\ 18\ 250\ 300\ 12$$

The mean sale of these individuals is:

$\bar{x} = 47.875$

In this case, the mean does not adequately represent the measure of centrality of the data set.

Measures of Centrality – Mean (7/9):

trimmed mean

A trimmed mean is computed by "trimming away" a certain percent of both the largest and the smallest set of values. For example, the 10% trimmed mean is found by eliminating the largest 10% and smallest 10% and computing the average of the remaining values. We will denote it by $\bar{x}_{tr0.1}$ or $\bar{x}_{tr(10)}$

Numerical Summaries (2/3)

Measures of Centrality – Mean (8/9):

<u>*Example 1:*</u>

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10 \ 8 \ 15 \ 12 \ 17 \ 7 \ 20 \ 19 \ 22 \ 25 \ 16 \ 15 \ 18 \ 250 \ 300 \ 12$$

Find the 13% trimmed mean sale of these individuals.

Numerical Summaries (2/3)

Measures of Centrality – Mean (8/9):

Example 1:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10\ 8\ 15\ 12\ 17\ 7\ 20\ 19\ 22\ 25\ 16\ 15\ 18\ 250\ 300\ 12$$

Find the 13% trimmed mean sale of these individuals.

order	7	8	10	12	12	15	15	16	17	18	19	20	22	25	250	300
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Remove 16*13/100 = 2.08 from each side.

Numerical Summaries (2/3)

Measures of Centrality – Mean (8/9):

<u>*Example 1:*</u>

order		10	12	12	15	15	16	17	18	19	20	22	25	

$$\bar{x}_{tr(13)} = \bar{x}_{t=0.13} = 16.75$$

$$\frac{(10 + 12 + 12 + 15 + 15 + 16 + 17 + 18 + 19 + 20 + 22 + 25)}{12}$$

Numerical Summaries (2/3)

Measures of Centrality – Mean (9/9):

weighted mean

Sometimes, we are interested in finding the sample average of a data set where each observation is given a relative importance expressed numerically by a set of values called weights.

$$\bar{X}_w = \frac{w_1 X_1 + w_2 X_2 + \cdots + w_n X_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

Numerical Summaries (2/3)

Measures of Centrality – Median (1/4):

Median (M)

The **median *M*** is the midpoint of a distribution. Half the observations are smaller than the median, and the other half are larger than the median.

Numerical Summaries (2/3)

Measures of Centrality – Median (2/4):

Median (M)

Here is a rule for finding the median:

1. Arrange all observations in order of size, from smallest to largest.
2. Find the location (rank) of the median of a data set of size n

$$\text{Rank} = \begin{cases} (n+1)/2 & \text{if } n \text{ odd} \\ n/2 \quad \text{and} \quad n/2+1 & \text{if } n \text{ even} \end{cases}$$

3. Find the value of the observation corresponding to the rank of the median founded in step 2. i.e., $x_{\left(\frac{n+1}{2}\right)}$ or $\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)/2$

Measures of Centrality – Median (3/4):

Example 1:

The nine observations are:

$x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, $x_8 = 13.1$ and $x_9 = 13.2$.

Arrange all observations in order of size $n = 9$:

Order	**12.3**	**12.6**	**12.6**	**12.9**	**13.1**	**13.2**	**13.4**	**13.5**	**13.6**
Ranks	1	2	3	4	5	6	7	8	9

Numerical Summaries (2/3)

Measures of Centrality – Median (3/4):

Example 1:

Arrange all observations in order of size $n = 9$:

Order	**12.3**	**12.6**	**12.6**	**12.9**	**13.1**	**13.2**	**13.4**	**13.5**	**13.6**
Ranks	1	2	3	4	5	6	7	8	9

$$\text{Rank} = \begin{cases} (n+1)/2 & \text{if } n \text{ odd} \\ n/2 \quad \text{and} \quad n/2+1 & \text{if } n \text{ even} \end{cases}$$

The size $n = 9$ is odd, so the median rank is $(n+1)/2 = 10/2 = 5$

Numerical Summaries (2/3)

Measures of Centrality – Median (3/4):

Example 1:

Arrange all observations in order of size $n = 9$:

Order	**12.3**	**12.6**	**12.6**	**12.9**	**13.1**	**13.2**	**13.4**	**13.5**	**13.6**
Ranks	1	2	3	4	5	6	7	8	9

$$\text{Rank} = \begin{cases} (n+1)/2 & \text{if } n \text{ odd} \\ n/2 \quad \text{and} \quad n/2+1 & \text{if } n \text{ even} \end{cases}$$

The size $n = 9$ is odd, so the median rank is $(n + 1)/2 = 10/2 = 5$

The median $= x_5 = 13.1$

Numerical Summaries (2/3)

Measures of Centrality – Median (4/4):

Example 2:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10 \; 8 \; 15 \; 12 \; 17 \; 7 \; 20 \; 19 \; 22 \; 25 \; 16 \; 15 \; 18 \; 250 \; 300 \; 12$$

Find the median sale of these individuals.

Numerical Summaries (2/3)

Measures of Centrality – Median (4/4):

Example 2:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10\ 8\ 15\ 12\ 17\ 7\ 20\ 19\ 22\ 25\ 16\ 15\ 18\ 250\ 300\ 12$$

Arrange all observations in order of size $n = 16$:

order	7	8	10	12	12	15	15	16	17	18	19	20	22	25	250	300
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Numerical Summaries (2/3)

Measures of Centrality – Median (4/4):

Example 2:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10\ 8\ 15\ 12\ 17\ 7\ 20\ 19\ 22\ 25\ 16\ 15\ 18\ 250\ 300\ 12$$

Arrange all observations in order of size $n = 16$:

order	7	8	10	12	12	15	15	16	17	18	19	20	22	25	250	300
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

$$\text{Rank} = \begin{cases} (n+1)/2 & \text{if } n \text{ odd} \\ n/2 \quad \text{and} \quad n/2+1 & \text{if } n \text{ even} \end{cases}$$

Numerical Summaries (2/3)

Measures of Centrality – Median (4/4):

Example 2:

The following data describe the sales (in thousands of dollars) for 16 randomly selected sales personnel distributed:

$$10\ 8\ 15\ 12\ 17\ 7\ 20\ 19\ 22\ 25\ 16\ 15\ 18\ 250\ 300\ 12$$

Arrange all observations in order of size $n = 16$:

order	7	8	10	12	12	15	15	16	17	18	19	20	22	25	250	300
Ranks	1	2	3	4	5	6	7	$\boxed{8}$	$\boxed{9}$	10	11	12	13	14	15	16

The size $n = 16$ is even, so the median ranks are $n/2$ and $n/2 + 1$ $= 16/2$ and $16/2 + 1$ $= 8$ and 9

Numerical Summaries (2/3)

Measures of Centrality – Median (4/4):

Example 2:

The size $n = 16$ is even, so the median ranks are $n/2$ and $n/2 + 1$ $= 16/2$ and $16/2 + 1$ $= 8$ and 9

order	7	8	10	12	12	15	15	16	17	18	19	20	22	25	250	300
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

The median = $(x_8 + x_9)/2 = (16 + 17)/2 = 16.5$

Numerical Summaries (2/3)

Measures of Centrality – Mode (1/4):

Mode

The mode of a data set is the value that occurs most frequently. There may be no mode, or conversely, there may be multiple modes.

- If all values without repeated, then NO mode.
- If founded more than one value with the same most frequency, then here more than one mode.

Numerical Summaries (2/3)

Measures of Centrality – Mode (2/4):

Example 1:

Find the mode for the following data set

$$3,\ 8,\ 5,\ 6,\ 10,\ 17,\ 19,\ 20,\ 3,\ 2,\ 11$$

Numerical Summaries (2/3)

Measures of Centrality – Mode (2/4):

Example 1:

Find the mode for the following data set

$$3, 8, 5, 6, 10, 17, 19, 20, 3, 2, 11$$

In the data set of this example, each value occurs once except the value 3, which occurs twice. Thus, *the mode* = 3.

Measures of Centrality – Mode (3/4):

Example 2:

Find the mode for the following data set

$$1,\ 7,\ 19,\ 23,\ 11,\ 12,\ 1,\ 12,\ 19,\ 7,\ 11,\ 23$$

Numerical Summaries (2/3)

Measures of Centrality – Mode (3/4):

Example 2:

Find the mode for the following data set

$$1, 7, 19, 23, 11, 12, 1, 12, 19, 7, 11, 23$$

Note that in this data set, each value occurs twice. Thus, this data set does *not have any mode*.

Numerical Summaries (2/3)

Measures of Centrality – Mode (4/4):

Example 3:

Find the mode for the following data set

$$5,\ 7,\ 12,\ 13,\ 14,\ 21,\ 7,\ 21,\ 23,\ 26,\ 5$$

Numerical Summaries (2/3)

Measures of Centrality – Mode (4/4):

Example 3:

Find the mode for the following data set

$$5, 7, 12, 13, 14, 21, 7, 21, 23, 26, 5$$

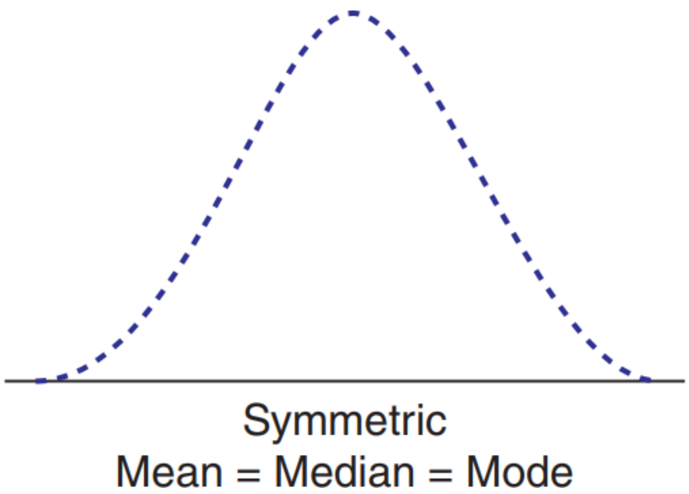
In this data set, values 5, 7, and 21 occur twice, and the rest of the values occur only once. Thus, in this example, there are *three modes* 5, 7, and 21.

Numerical Summaries (2/3)

Measures of Centrality – Shape (1/3):

Symmetric:

A data set is symmetric when the values in the data set that lie equidistant from the mean, on either side, occur with equal frequency.

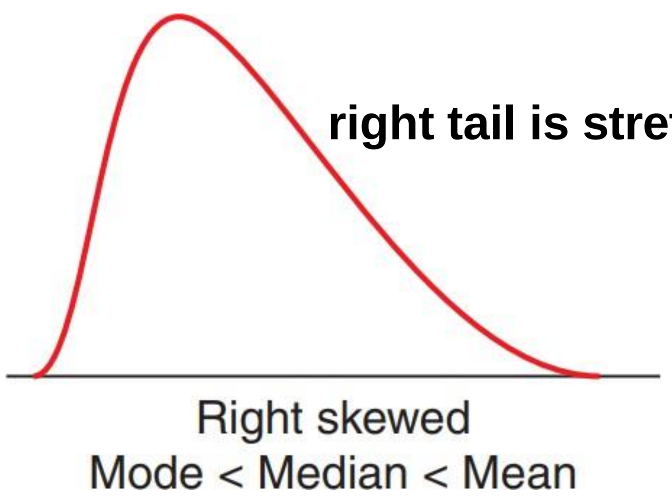


Numerical Summaries (2/3)

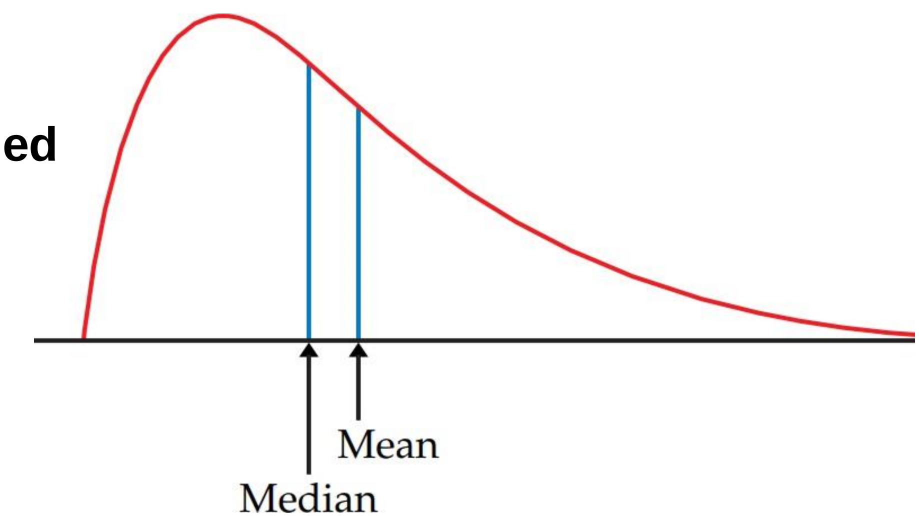
Measures of Centrality – Shape (2/3):

right-skewed:

A data set is right-skewed when values in the data set that are smaller than the median occur with relatively higher frequency than those values that are greater than the median. The values greater than the median are scattered to the right far from the median.



right tail is stretched

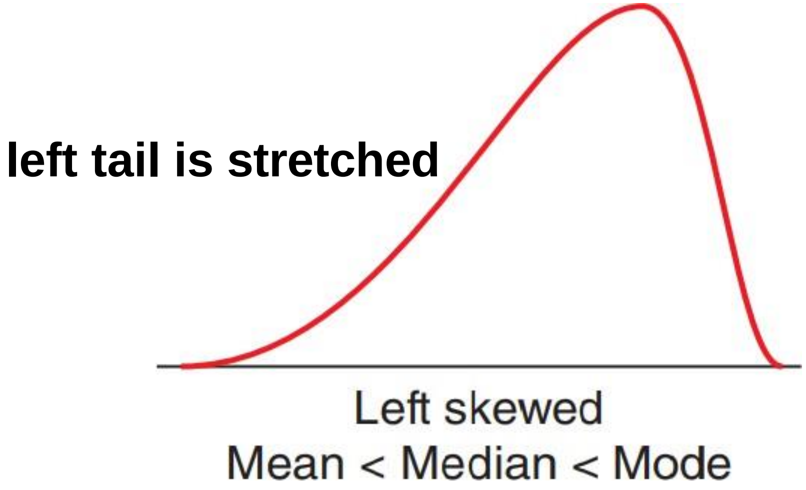


Numerical Summaries (2/3)

Measures of Centrality – Shape (3/3):

left-skewed:

A data set is left-skewed when values in the data set that are greater than the median occur with relatively higher frequency than those values that are smaller than the median. The values smaller than the median are scattered to the left far from the median.



Numerical Summaries (2/3)

Measures of Variability (Dispersion/Spread) (1/3):

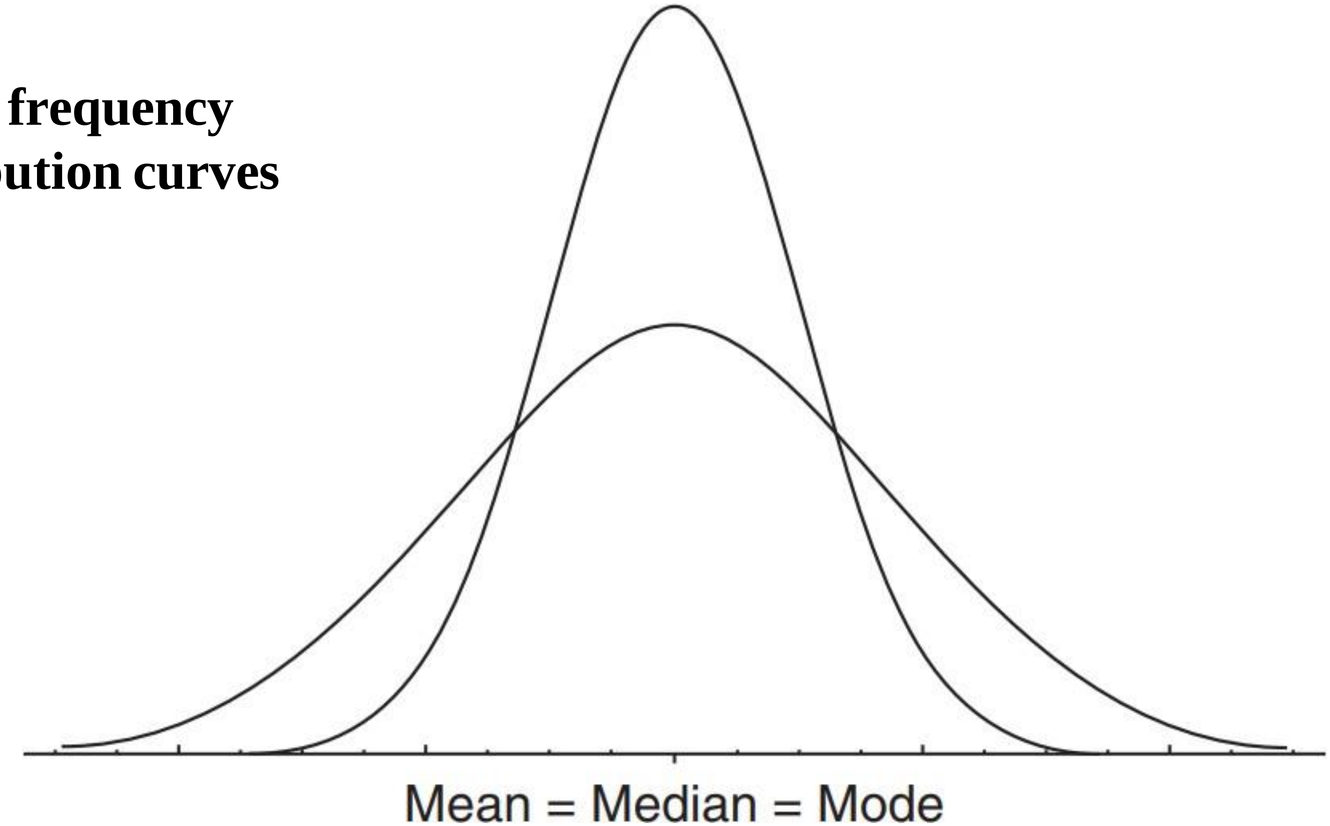
Although the sample mean is useful, it does not convey all of the information about a sample of data. Therefore, sample variability plays an important role in data analysis.

For example, a drug manufactured with the correct mean concentration of active ingredient is dangerous if some batches are much too high and others much too low. We are interested in the spread or variability of incomes and drug potencies as well as their centers.

Numerical Summaries (2/3)

Measures of Variability (Dispersion/Spread) (2/3):

Two frequency distribution curves



Numerical Summaries (2/3)

Measures of Variability (Dispersion/Spread) (3/3):

We study three measures of dispersion:

1. range,
2. variance, and
3. standard deviation.

Numerical Summaries (2/3)

Measures of Variability – Range (1/2):

Range is defined as:

$$\text{Range} = \text{Largest value} - \text{Smallest value}$$

Sample → Sample Range

Population → Population Range

Numerical Summaries (2/3)

Measures of Variability – Range (2/2):

Example 1:

Find the sample range for the following data set

$$1, 7, 19, 23, 11, 12, 1, 12, 19, 7, 11, 23$$

Largest value = 23

Smallest value = 1

$$\text{Sample Range} = 23 - 1 = 22$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (1/9):

Variance and Standard Deviation:

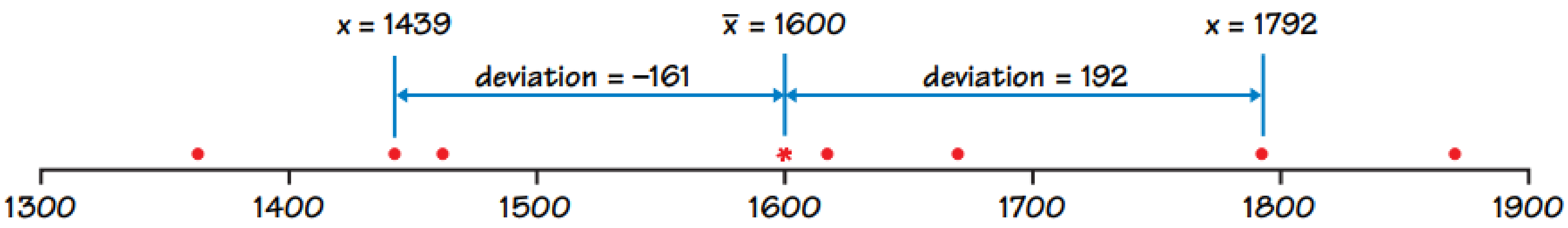
To better understand variability, we rely on more powerful indicators such as the **variance**, the **standard deviation**.

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (2/9):

Variance and Standard Deviation:

The standard deviation measures spread by looking at how far the observations are from their mean.



Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

The sample variance s^2 of a set of observations is the average of the squares of the deviations of the observations from their mean.

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the sample variance is

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the sample variance is

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}$$

The ***degrees of freedom*** associated with the variance estimate

Note: the degrees of freedom depict the number of independent pieces of information available for computing variability.

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If x_1, x_2, x_3, x_4, x_5 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$

x_1	x_2	x_3	x_4	x_5	$\bar{x}$
					90

$n = 5$, we have $(n - 1) = (4)$ degrees of freedom

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If x_1, x_2, x_3, x_4, x_5 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$

$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$	$\boldsymbol{x_5}$	$\bar{x}$
Free	Free	Free	Free	Not Free	Given
					90

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If x_1, x_2, x_3, x_4, x_5 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$

x_1	x_2	x_3	x_4	x_5	$\bar{x}$
Free	Free	Free	Free	Not Free	Given
80	85	90	95		90

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If x_1, x_2, x_3, x_4, x_5 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$

x_1	x_2	x_3	x_4	x_5	$\bar{x}$
Free	Free	Free	Free	Not Free	Given
80	85	90	95		90

$$90 = \frac{80 + 85 + 90 + 95 + x_5}{5} \quad \rightarrow \quad x_5 = 100$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If 80, 85, 90, 95, 100 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$. For example, suppose that we wish to compute the sample variance.

x_1	x_2	x_3	x_4	x_5	$\bar{x}$
80	85	90	95	100	90

In general,

$$\sum_{i=1}^{5} (x_i - \bar{x}) = 0$$

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If 80, 85, 90, 95, 100 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$. For example, suppose that we wish to compute the sample variance.

x_1	x_2	x_3	x_4	x_5	$\bar{x}$
80	85	90	95	100	90

In general,

$$\sum_{i=1}^{5} (x_i - \bar{x}) = (80 - 90) + (85 - 90) + (90 - 90) + (95 - 90) + (100 - 90) = 0$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If 80, 85, 90, 95, 100 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$. For example, suppose that we wish to compute the sample variance.

x_1	x_2	x_3	x_4	x_5	$\bar{x}$
80	85	90	95	100	90

In general,

$$\sum_{i=1}^{5} (x_i - \bar{x}) = (-10) + (-5) + (0) + (5) + (10) = 0$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If 80, 85, 90, 95, 100 is a sample of 5 observations, and the sample mean is $\bar{x} = 90$. For example, suppose that we wish to compute the sample variance.

x_1	x_2	x_3	x_4	x_5	$\bar{x}$
80	85	90	95	100	90

The computation of the variance involves,

$$\sum_{i=1}^{5} (x_i - \bar{x})^2 = (-10)^2 + (-5)^2 + (0)^2 + (5)^2 + (10)^2$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

The computation of the variance involves,

$$\sum_{i=1}^{5}(x_i - \bar{x})^2 = (-10)^2 + (-5)^2 + (0)^2 + (5)^2 + (10)^2$$

$\sum_{i=1}^{5}(x_i - \bar{x}) = 0$, then the computation of a sample variance does not involve n independent squared deviations from the mean $\bar{x}$. In fact, since the last value of $x - \bar{x}$ is determined by the initial $n-1$ of them, we say that these are $n-1$ "pieces of information" that produce s^2. Thus, there are $n-1$ degrees of freedom rather than n degrees of freedom for computing a sample variance.

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

Another way to think about $(n - 1)$: In practice, the value of μ is almost never known, and so the sum of the squared deviations about the sample average $\bar{x}$ must be used instead. However, the observations x_i tend to be closer to their average, $\bar{x}$, than to the population mean, μ.

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

Therefore, we use $(n-1)$ as the divisor rather than n. If we used n as the divisor in the sample variance, we would obtain a measure of variability that is on the average consistently smaller than the true population variance σ^2.

Of course, if n is large, the difference in the denominator is inconsequential.

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (3/9):

Sample Variance:

If $x_1, x_2, \ldots, x_n$ is a sample of n observations, the sample variance is

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}$$

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}.$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (4/9):

Sample Standard Deviation:

The sample standard deviation, denoted by s, is the positive square root of s^2, that is

$$s = \sqrt{s^2}$$

Note: It is important to remember the value of s^2, and therefore of s, is always greater than zero, except when all the data values are equal, in which case $s^2 = 0$.

Measures of Variability – Standard Deviation (5/9):

Example 1: *Find sample variance and sample standard deviation?*

The eight observations are:

$x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. (meters)

Measures of Variability – Standard Deviation (5/9):

Example 1:

The eight observations are:

$x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. (meters)

The sample mean is:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{8} x_i}{8}$$

$$= \frac{12.6 + 12.9 + \cdots + 13.1}{8} = \frac{104}{8} = 13.0$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (5/9):

Example 1:

The eight observations are:

$x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. (meters)

The sample mean is:

$$\bar{x} = 13 \text{ (meters)}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (5/9):

<u>Example 1:</u>

$\bar{x} = 13$ (meters)

The eight observations are:

$x_1 = 12.6$, $x_2 = 12.9$, $x_3 = 13.4$, $x_4 = 12.3$, $x_5 = 13.6$, $x_6 = 13.5$, $x_7 = 12.6$, and $x_8 = 13.1$. (meters)

The sample variance is:

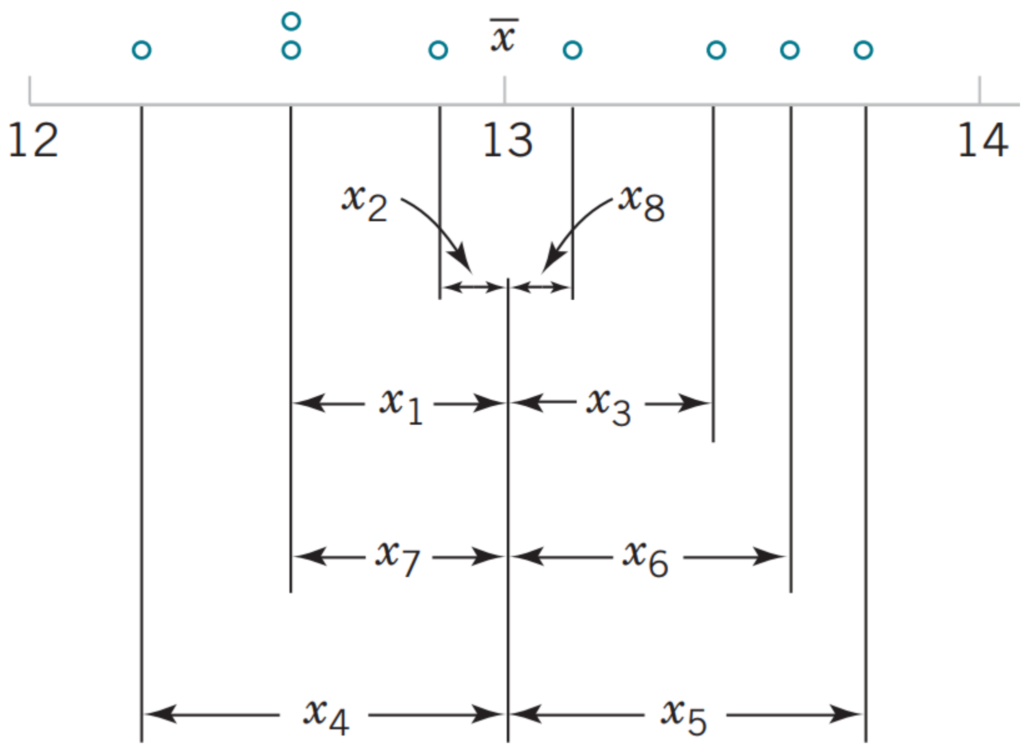
$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}.$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (5/9):

Example 1:

$\bar{x} = 13$ (meters)



$(x_i - \bar{x})$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (5/9):

Example 1:

$\bar{x} = 13$ (meters)

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	12.6	−0.4	0.16
2	12.9	−0.1	0.01
3	13.4	0.4	0.16
4	12.3	−0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	−0.4	0.16
8	13.1	0.1	0.01
Total	104.0	0.0	1.60

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (5/9):

Example 1:

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	12.6	−0.4	0.16
2	12.9	−0.1	0.01
3	13.4	0.4	0.16
4	12.3	−0.7	0.49
5	13.6	0.6	0.36
6	13.5	0.5	0.25
7	12.6	−0.4	0.16
8	13.1	0.1	0.01
Total	104.0	0.0	1.60

$$\bar{x} = 13 \text{ (meters)}$$

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 1.60$$

the sample variance is

$$s^2 = \frac{1.60}{8-1} = \frac{1.60}{7} = 0.2286 \ (\text{meters})^2$$

and the sample standard deviation is

$$s = \sqrt{0.2286} = 0.478 \ \text{(meters)}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (6/9):

Sample Variance:

If $x_1, x_2, \dots, x_n$ is a sample of n observations, the sample variance is

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}.$$

The computation of s^2 requires calculation of $\bar{x}$, n subtractions, and n squaring and adding operations. If the original observations or the deviations $x_i - \bar{x}$ are not integers, the deviations $x_i - \bar{x}$ may be tedious to work with, and several decimals may have to be carried to ensure numerical accuracy.

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (6/9):

Sample Variance:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}.$$

For computational purposes, we give below the simplified form for the sample variance.

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right)$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (7/9):

<u>Recall Example 1:</u>

i	x_i	x_i^2
1	12.6	158.76
2	12.9	166.41
3	13.4	179.56
4	12.3	151.29
5	13.6	184.96
6	13.5	182.25
7	12.6	158.76
8	13.1	171.61
Total	**104.0**	**1,353.6**

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right)$$

the sample variance is

$$s^2 = \frac{1}{8-1}\left(1353.6 - \frac{(104)^2}{8}\right)$$

$$= 0.2286 \text{ (meters)}^2$$

and the sample standard deviation is

$$s = \sqrt{0.2286} = 0.478 \text{ (meters)}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (8/9):

Population Variance:

The variability in the population is defined by the population variance. When the population is finite and consists of N equally likely values, we may define the population variance as

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (8/9):

Population Variance:

N is used because we assume that the population mean μ is given. Moreover, the number N is very large, the differences will be very small when use N or $N-1$.

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Numerical Summaries (2/3)

Measures of Variability – Standard Deviation (9/9):

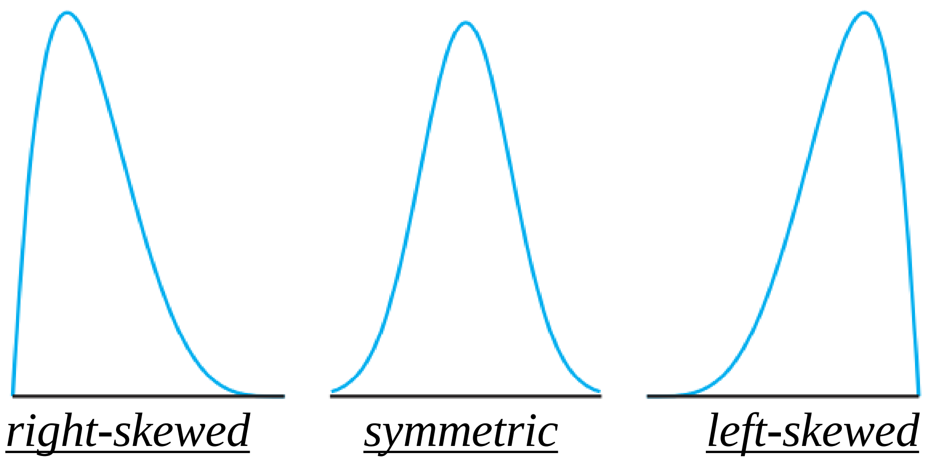
Population Standard Deviation:

The positive square root of σ^2 is the population standard deviation σ

Numerical Summaries (2/3)

Measures of Shape – Skewness (1/4):

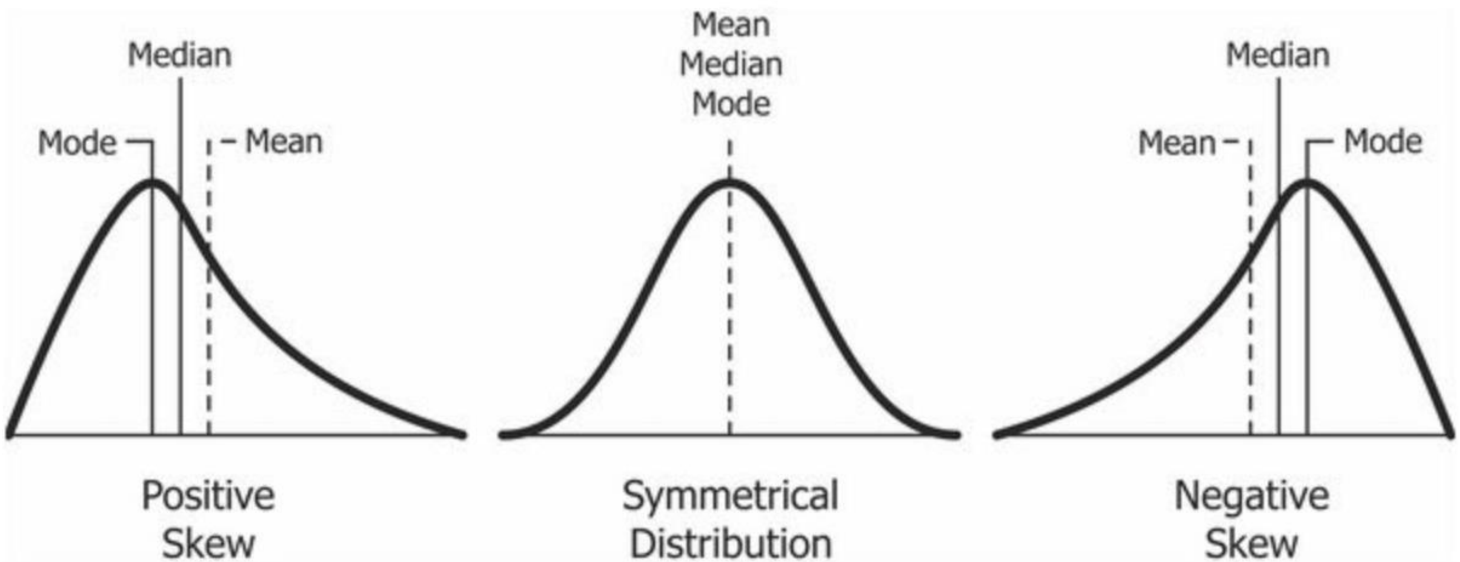
***Skewness**:* is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.



Numerical Summaries (2/3)

Measures of Shape – Skewness (2/4):

- Right skewness is positive skewness, which means skewness > 0.
- Left skewness is negative skewness, which means skewness < 0.
- Symmetrical distribution implies zero skewness.



Numerical Summaries (2/3)

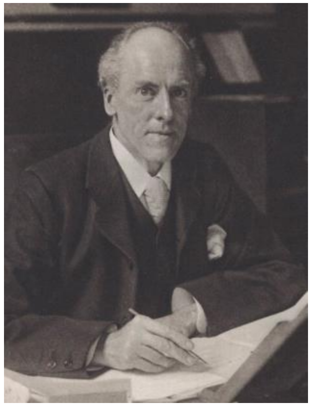
Measures of Shape – Skewness (3/4):

- If skewness is less than -1 or greater than 1, the distribution is highly skewed.
- If skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.
- If skewness is between -0.5 and 0.5, the distribution is approximately symmetric.

Numerical Summaries (2/3)

Measures of Shape – Skewness (4/4):

How to calculate?



Karl Pearson, British

There are several ways to measure skewness. Pearson's first and second coefficients of skewness are two common ones.

- *Mode skewness*: Pearson's first coefficient of skewness.

$$Sk_1 = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

- *Median skewness*: Pearson's second coefficient of skewness.

$$Sk_2 = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

Numerical Summaries (2/3)

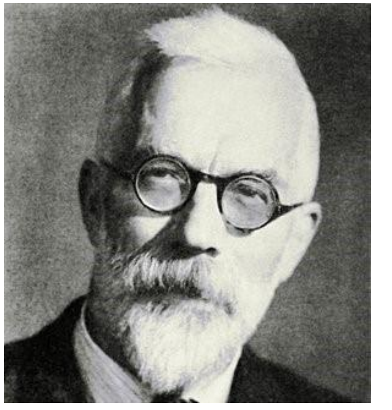
Measures of Shape – Skewness (4/4):

How to calculate?

Sample Skewness: for a sample of n values

$$g_1 = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_i - \overline{x})^3}{s^3}$$

The above formula for skewness is referred to as the Fisher-Pearson coefficient of skewness.



Sir Ronald Fisher, British

Measures of Shape – Skewness (4/4):

How to calculate?

Sample Skewness: Many software programs actually compute the **adjusted** Fisher-Pearson coefficient of skewness. For a sample of n values

$$G_1 = \frac{n^2}{(n-1)(n-2)} \left(\frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3}{s^3} \right)$$

Numerical Summaries (2/3)

Measures of Shape – Skewness (4/4):

How to calculate?

Sample Skewness: Many software programs actually compute the **adjusted** Fisher-Pearson coefficient of skewness. For a sample of n values

$$G_1 = \frac{n^2}{(n-1)(n-2)} \underbrace{\boxed{\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \right)}}_{g_1}$$

Numerical Summaries (2/3)

Measures of Shape – Skewness (4/4):

How to calculate?

Sample Skewness: Alternative definition of G_1

$$G_1 = \frac{\sqrt{n(n-1)}}{(n-2)}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{3/2}}\right)$$

Numerical Summaries (2/3)

Measures of Shape – Skewness (4/4):

How to calculate?

Sample Skewness:

The adjusted Fisher–Pearson standardized moment coefficient G_1 is the version found in Excel and several statistical packages including Minitab, SAS and SPSS.

Numerical Summaries (2/3)

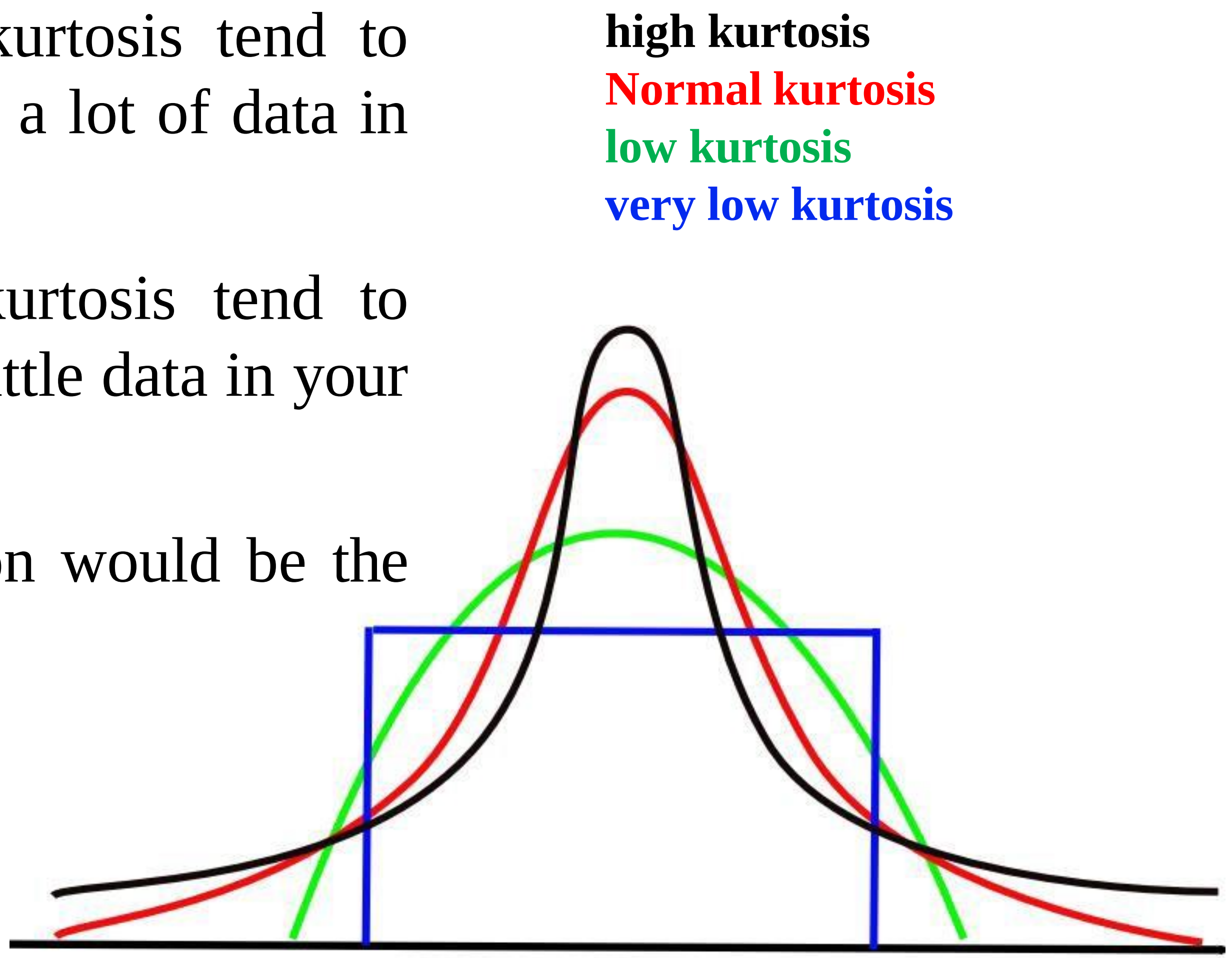
Measures of Shape – Kurtosis (1/6):

Kurtosis: is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.

Numerical Summaries (2/3)

Measures of Shape – Kurtosis (2/6):

- Datasets with high kurtosis tend to have heavy tails (i.e., a lot of data in your tails).
- Datasets with low kurtosis tend to have light tails (i.e., little data in your tails).
- A uniform distribution would be the extreme case.



Measures of Shape – Kurtosis (3/6):

How to calculate?

Kurtosis, κ, can be calculated as: for a sample of n values

$$\kappa = \frac{\sigma_{i=1}^{n}(x_i - \bar{x})^4/n}{s^4}$$

Greater values of κ indicates a "heavy-tailed" distribution whereas smaller κ indicates a "light-tailed" distribution.

Numerical Summaries (2/3)

Measures of Shape – Kurtosis (3/6):

How to calculate?

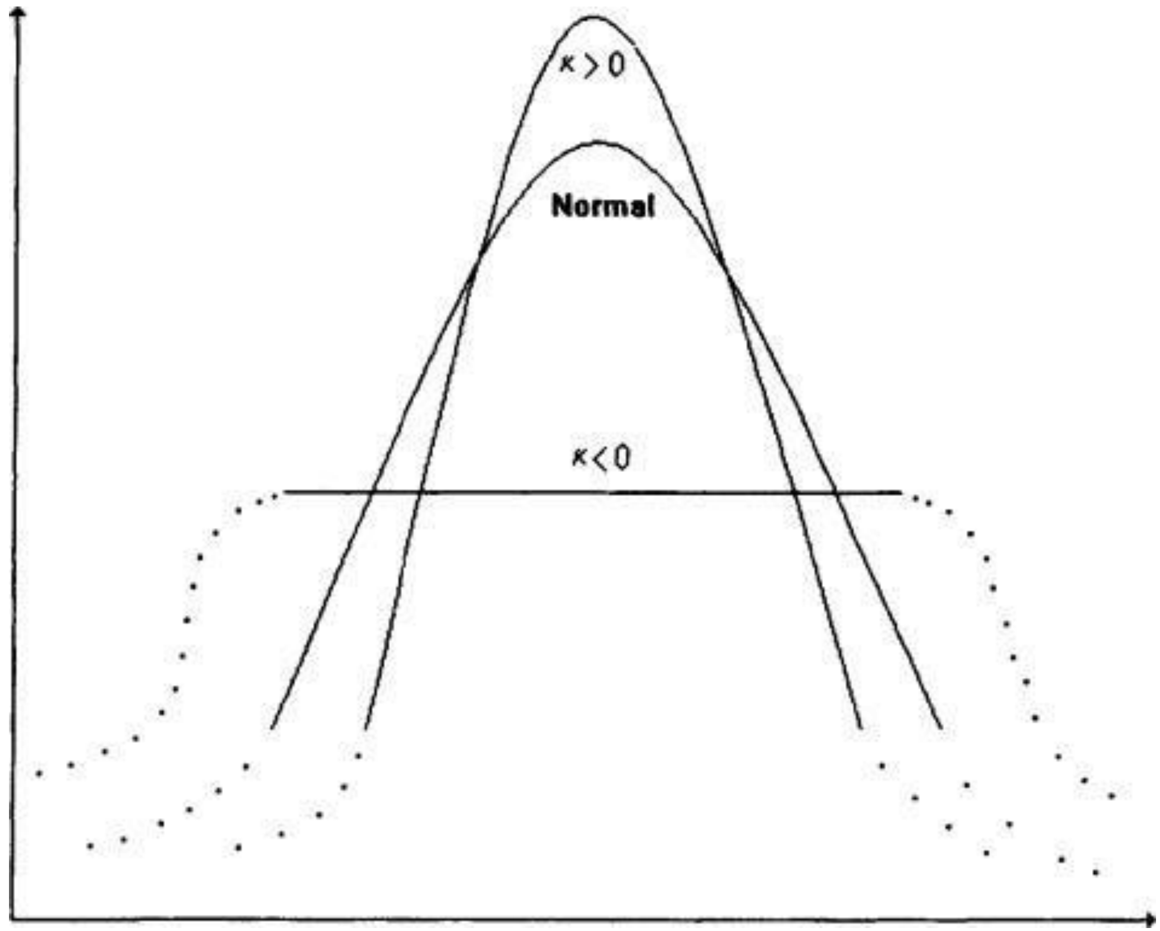
The kurtosis value for a *normal curve* often approximates **three**. For this reason, some sources use the following definition of kurtosis (often referred to as "***excess kurtosis***"): for a sample of n values

$$\kappa = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4/n}{s^4} - 3$$

This definition is used so that the normal distribution has a kurtosis of zero. In addition, with the second definition **positive** kurtosis indicates a "**heavy-tailed**" distribution and **negative** kurtosis indicates a "**light tailed**" distribution.

Numerical Summaries (2/3)

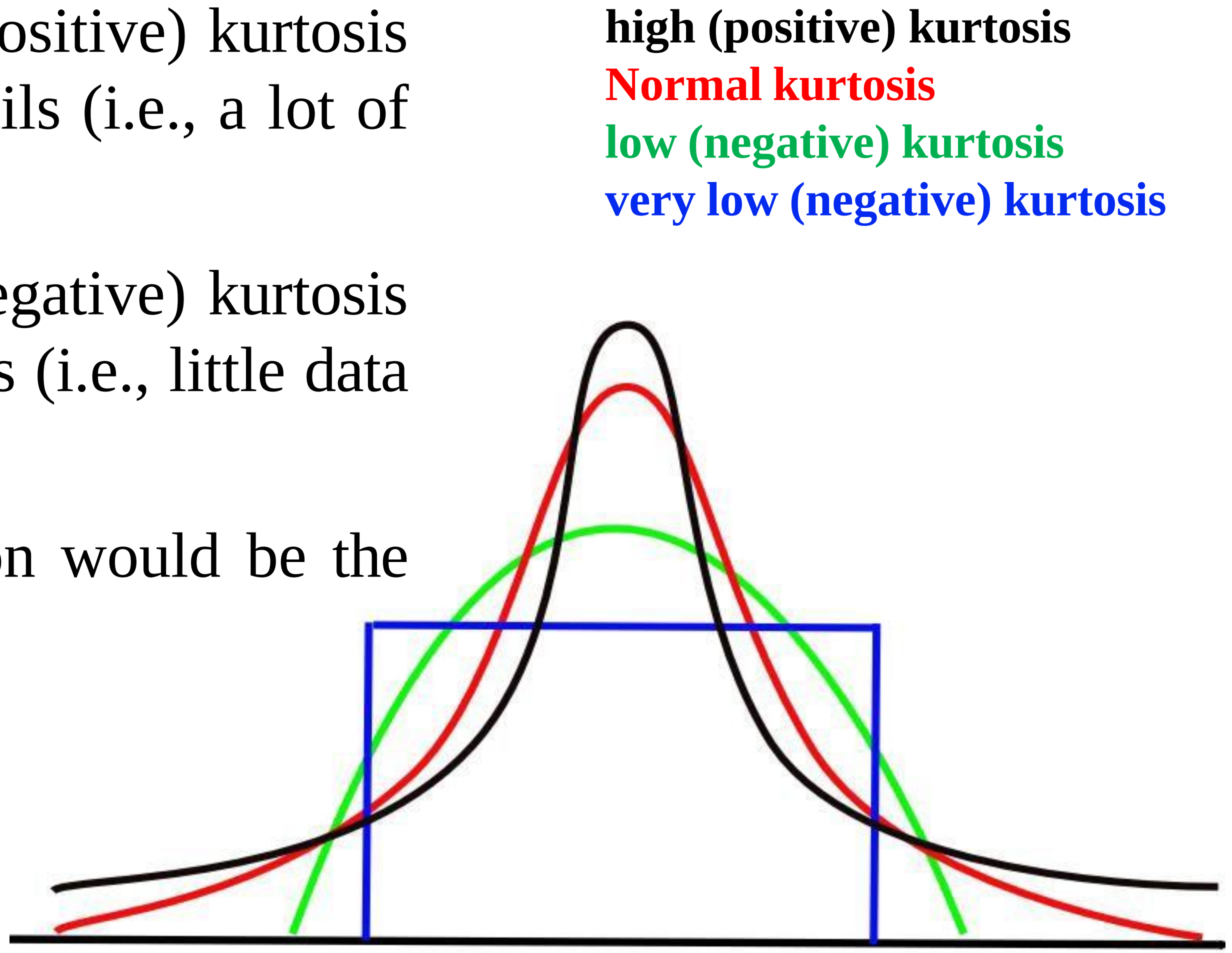
Measures of Shape – Kurtosis (4/6):



Numerical Summaries (2/3)

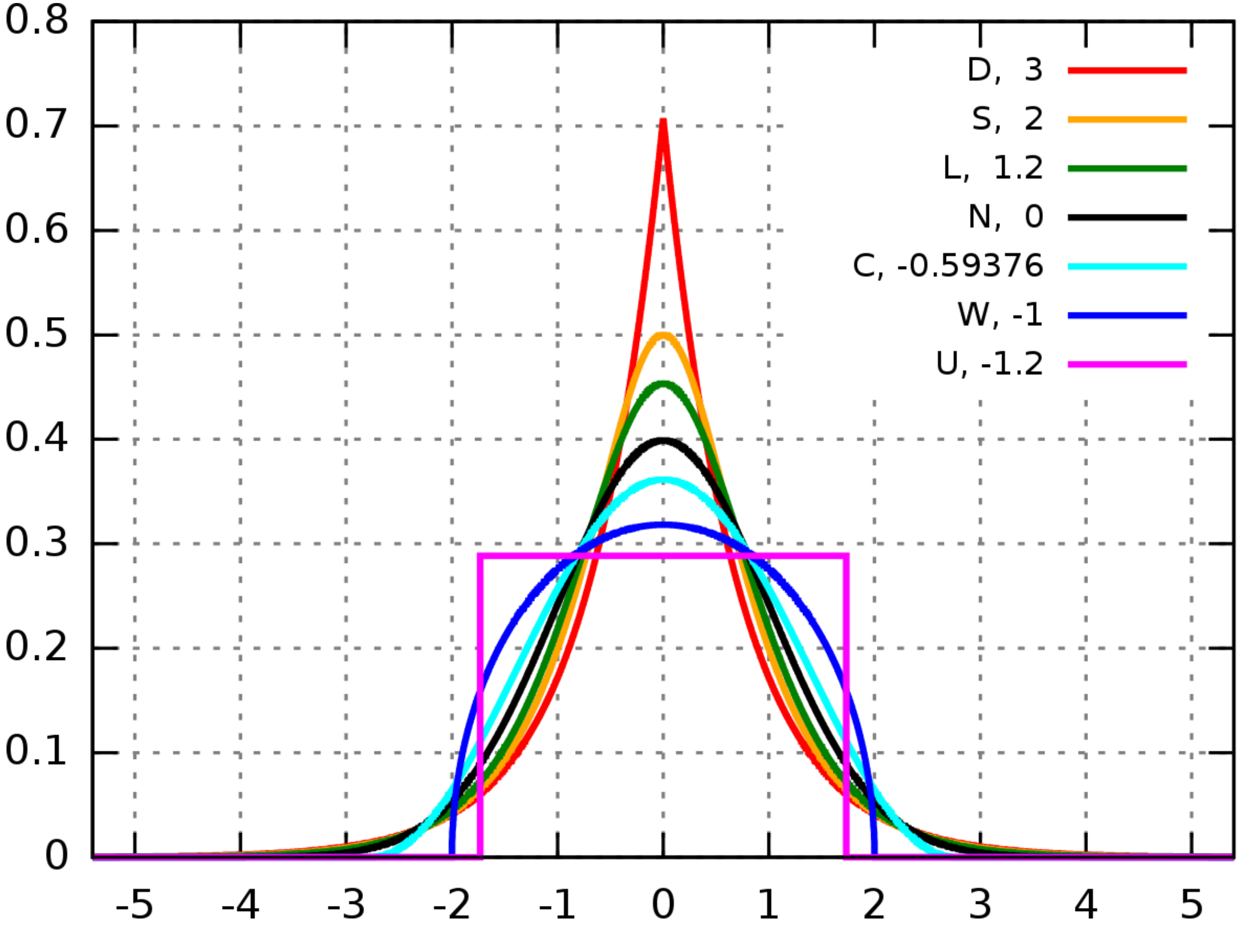
Measures of Shape – Kurtosis (5/6):

- Datasets with high (positive) kurtosis tend to have heavy tails (i.e., a lot of data in your tails).
- Datasets with low (negative) kurtosis tend to have light tails (i.e., little data in your tails).
- A uniform distribution would be the extreme case.



Numerical Summaries (2/3)

Measures of Shape – Kurtosis (6/6):



Numerical Summaries (2/3)

Measures of Relative Position:

This section introduces measures of relative position that divide the data into percentages to help locate any data value in the whole data set. Commonly used measures of relative position are ***percentiles*** and ***quartiles***.

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (1/4):

Percentiles: divide the data into **one hundred equal parts**; each part contains at the most 1% of the data and is numbered from 1 to 99.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	.	.	.	x_{95}	x_{96}	x_{97}	x_{98}	x_{99}	x_{100}

1 2 99

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (2/4):

Percentiles: divide the data into **one hundred equal parts**; each part contains at the most 1% of the data and is numbered from 1 to 99.

For example, the median of a data set is the 50th percentile, which divides the data into two equal parts so that at most 50% of the data fall below the median and at most 50% of the data fall above it. The procedure for determining the percentiles is similar to the procedure used for determining the median.

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (3/4):

We compute the percentiles as follows:

Step 1. Write the data values in an ascending order and rank them from 1 to n.

Step 2. Find the rank of the pth percentile (p = 1, 2, ... , 99), which is given by Rank of the pth percentile = $p \times [(n+1)/100]$

Step 3. Find the data value that corresponds to the rank of the pth percentile.

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

<u>*Example 1:*</u>

(Engineers' salaries) The following data give the salaries (in thousands of dollars) of 15 engineers in a company:

62 48 52 63 85 51 95 76 72 51 69 73 58 55 54

(a) Find the 50th percentile for these data (i.e., median).

(b) Find the 70th percentile for these data.

(c) Find the percentile corresponding to the salary of $60,000.

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

<u>*Example 1:*</u> (a) Find the 50th percentile for these data (i.e., median).

Step 1. Write the data values in an ascending order and rank them from 1 to 15.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Measures of Relative Position – Percentiles (4/4):

Example 1: (a) Find the 50th percentile for these data (i.e., median).

Step 2. Find the rank of the pth percentile (p = 1, 2, ... , 99), which is given by Rank of the pth percentile = $p \times [(n+1)/100]$.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 50th percentile = $50 \times [(15+1)/100]$ = 8

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

<u>*Example 1:*</u> (a) Find the 50th percentile for these data (i.e., median).

Step 3. Find the data value that corresponds to the rank of the pth percentile.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 50th percentile = $50 \times [(15 + 1)/100]$ = 8.

The 50th percentile is 62.

That is, at most 50% of the engineers are making less than \$62,000 and at most 50% of the engineers are making more than \$62,000.

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

Example 1: (b) Find the 70th percentile for these data.

Step 2. Find the rank of the pth percentile (p = 1, 2, ... , 99), which is given by Rank of the pth percentile = $p \times [(n+1)/100]$.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 70th percentile = $70 \times [(15+1)/100]$ = 11.2.

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

Example 1: (b) Find the 70th percentile for these data.

Step 3. Find the data value that corresponds to the rank of the *p*th percentile.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 70th percentile = $70 \times [(15 + 1)/100] = 11.2$.

The 70th percentile is $\big(72 + (73 - 72)(0.2)\big) = 72.2$.

That is, at most 70% of the engineers are making less than \$72,200 and at most 30% of the engineers are making more than \$72,200.

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

<u>*Example 1:*</u> (c) Find the percentile corresponding to the salary of $60,000.

Now we want to find the percentile p corresponding to a given value x. This can be done by using the following formula:

$$p = \frac{\text{Number of data values} \leq x}{n+1} \times 100$$

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Measures of Relative Position – Percentiles (4/4):

Example 1: (c) Find the percentile corresponding to the salary of $\boxed{\$60{,}000.}$

Now we want to find the percentile p corresponding to a given value x. This can be done by using the following formula:

$$p = \frac{\text{Number of data values} \le x}{n+1} \times 100$$

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

Example 1: (c) Find the percentile corresponding to the salary of $\boxed{\$60{,}000.}$

Now we want to find the percentile p corresponding to a given value x. This can be done by using the following formula:

$$p = \frac{\text{Number of data values} \leq x}{n+1} \times 100$$

7 values

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Numerical Summaries (2/3)

Measures of Relative Position – Percentiles (4/4):

Example 1: (c) Find the percentile corresponding to the salary of $60,000.

Now we want to find the percentile p corresponding to a given value 60. This can be done by using the following formula:

$$p = \left(\frac{7}{15+1}\right) \times 100 = 43.75 \approx 44$$

7 values

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Measures of Relative Position – Percentiles (4/4):

Example 1: (c) Find the percentile corresponding to the salary of $60,000.

Now we want to find the percentile p corresponding to a given value 60. This can be done by using the following formula:

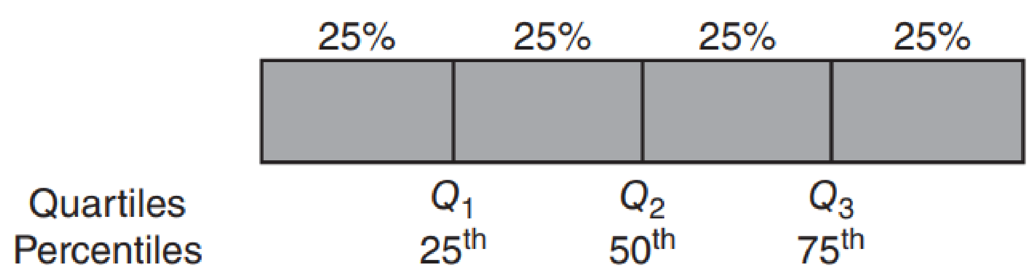
$$p = \left(\frac{7}{15+1}\right) \times 100 = 43.75 \approx 44$$

Hence, the engineer who makes a salary of $60,000 is at the 44th percentile. In other words, at most 44% of the engineers are making less than $60,000, or at most 56% are making more than $60,000.

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (1/2):

Quartiles: In the percentiles, we divide the data into 100 equal parts. Some of the percentiles have special importance, such as the 25th, 50th, and 75th percentiles. These percentiles are also known as the first, second, and third quartiles (denoted by Q_1, Q_2, and Q_3), respectively. Sometimes, they are also known as the lower, middle, and upper quartiles, respectively. The second quartile is the same as the median.



Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

Example 1:

(Engineers' salaries) The following data give the salaries (in thousands of dollars) of 15 engineers in a company:

```
62   48   52   63   85   51   95   76   72   51   69   73   58   55   54
```

(a) Find the 25th percentile (1st quartile) for these data.

(b) Find the 50th percentile (2nd quartile) for these data (i.e., median).

(c) Find the 75th percentile (3rd quartile) for these data.

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

<u>*Example 1:*</u> (a) Find the 25th percentile for these data (i.e., Q_1).

Step 1. Write the data values in an ascending order and rank them from 1 to 15.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

Example 1: (a) Find the 25th percentile for these data (i.e., Q_1).

Step 2. Find the rank of the pth percentile (p = 1, 2, ... , 99), which is given by Rank of the pth percentile = $p \times [(n+1)/100]$.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 25th percentile

$= 25 \times [(15+1)/100] = 1/4(15+1) = 4$

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

<u>*Example 1:*</u> (a) Find the 25th percentile for these data (i.e., Q_1).

Step 3. Find the data value that corresponds to the rank of the pth percentile.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 25th percentile = $25 \times [(15 + 1)/100] = 4$.

The 25th percentile is 52.

That is, the 1st quartile, $Q_1 = 52$

Measures of Relative Position – Quartiles (2/2):

Example 1: (b) Find the 50th percentile for these data (i.e., median).

Step 1. Write the data values in an ascending order and rank them from 1 to 15.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

<u>*Example 1:*</u> (b) Find the 50th percentile for these data (i.e., median).

Step 2. Find the rank of the pth percentile (p = 1, 2, ... , 99), which is given by Rank of the pth percentile = $p \times [(n + 1)/100]$.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 50th percentile

$= 50 \times [(15 + 1)/100] = 1/2(15 + 1) = 8$

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

Example 1: (b) Find the 50th percentile for these data (i.e., median).

Step 3. Find the data value that corresponds to the rank of the pth percentile.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 50th percentile = $50 \times [(15+1)/100] = 8$.

The 50th percentile is 62.

That is, the 2nd quartile, $Q_2 = 62$

Measures of Relative Position – Quartiles (2/2):

Example 1: (c) Find the 75th percentile for these data (i.e., Q_3).

Step 1. Write the data values in an ascending order and rank them from 1 to 15.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Measures of Relative Position – Quartiles (2/2):

Example 1: (c) Find the 75th percentile for these data (i.e., Q_3).

Step 2. Find the rank of the pth percentile (p = 1, 2, ... , 99), which is given by Rank of the pth percentile = $p \times [(n + 1)/100]$.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 75th percentile

$= 75 \times [(15 + 1)/100] = 3/4(15 + 1) = 12$

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

Example 1: (c) Find the 75th percentile for these data (i.e., Q_3).

Step 3. Find the data value that corresponds to the rank of the pth percentile.

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The rank of the 75th percentile = $75 \times [(15 + 1)/100] = 12$.

The 75th percentile is 73.

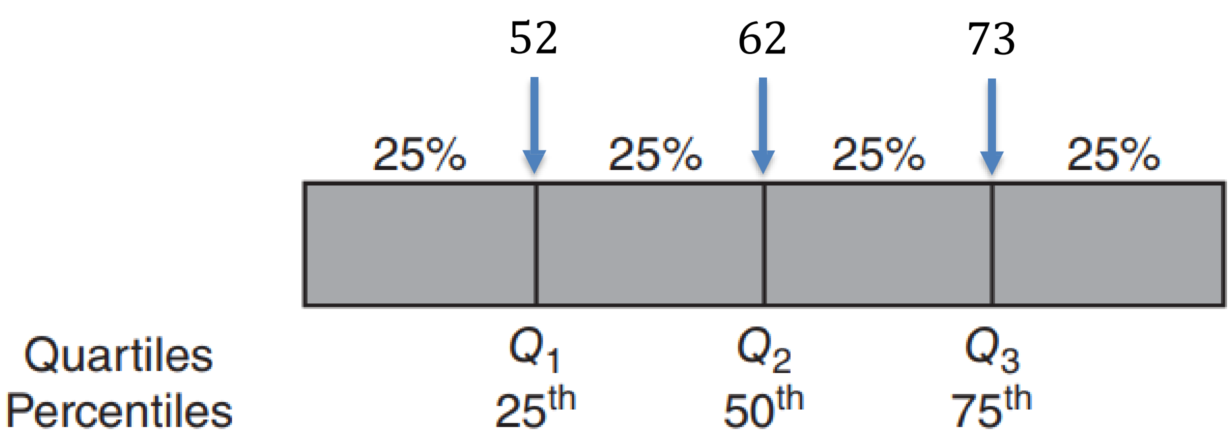
That is, the 3rd quartile, $Q_3 = 73$

Numerical Summaries (2/3)

Measures of Relative Position – Quartiles (2/2):

Example 1:

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15



Numerical Summaries (2/3)

Measures of Relative Position – IQR (1/3):

Interquartile Range (IQR): the spread between the first quartile and the third quartile, which is IQR and is defined as:

$$IQR = Q_3 - Q_1$$

Numerical Summaries (2/3)

Measures of Relative Position – IQR (2/3):

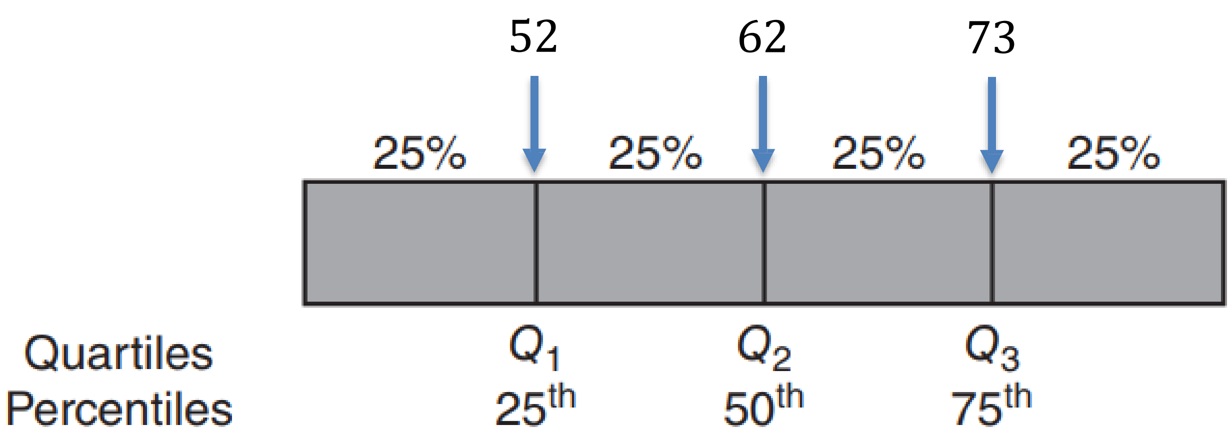
- The IQR gives an estimate of the range of the middle 50% of the population.
- The IQR is potentially a more meaningful measure of dispersion than the range as it is not affected by the extreme values that may be present in the data. By trimming 25% of the data from the bottom and 25% from the top, we eliminate the extreme values that may be present in the data set. Thus, the IQR is often used as a measure of comparison between two or more data sets on similar studies.

Numerical Summaries (2/3)

Measures of Relative Position – IQR (3/3):

Example 1:

Salaries	48	51	51	52	54	55	58	62	63	69	72	73	76	85	95
Ranks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15



$$IQR = Q_3 - Q_1 = 73 - 52 = 11$$

Numerical Summaries (2/3)

Measures of Relative Position – *cv* (1/4):

Coefficient of Variation (cv): The coefficient of variation is usually denoted by cv and is defined as the ratio of the standard deviation to the mean expressed as a percentage:

$$cv = \frac{standard\ deviation}{|mean|} = \frac{s}{|\bar{x}|} \times 100\%$$

the absolute value of the mean

Numerical Summaries (2/3)

Measures of Relative Position – cv (2/4):

The coefficient of variation is a relative comparison of a standard deviation to its mean and is unitless. The cv is commonly used to compare the variability in two populations.

Numerical Summaries (2/3)

Measures of Relative Position – cv (3/4):

For example, we might want to compare the disparity of earnings for technicians who have the same employer but work in two different countries. In this case, we would compare the coefficient of variation of the two populations rather than compare the variances, which would be an invalid comparison. The population with a greater coefficient of variation, generally speaking, has more variability than the other.

Numerical Summaries (2/3)

Measures of Relative Position – *cv* (4/4):

Example 1:

A company uses two measuring instruments, one to measure the diameters of ball bearings and the other to measure the length of rods it manufactures. The quality control department of the company wants to find which instrument measures with more precision. To achieve this goal, a quality control engineer takes several measurements of a ball bearing by using one instrument and finds the sample average $\bar{x}$ and the standard deviation s to be 3.84 and 0.02 mm, respectively. Then, he takes several measurements of a rod by using the other instrument and finds the sample average $\bar{x}$ and the standard deviation s to be 29.5 and 0.035 cm, respectively. *Estimate the coefficient of variation from the two sets of measurements.*

Measures of Relative Position – cv (4/4):

Example 1:

Instrument 1: $\bar{x} = 3.84$ and $s = 0.02$

$$cv_1 = \frac{s}{|\bar{x}|} \times 100\% = \frac{0.02}{3.84} \times 100\% = 0.521\%$$

Instrument 2: $\bar{x} = 29.5$ and $s = 0.035$

$$cv_2 = \frac{s}{|\bar{x}|} \times 100\% = \frac{0.035}{29.5} \times 100\% = 0.119\%$$

The measurements of the lengths of rod are relatively less variable than of the diameters of the ball bearings. Therefore, we can say the data show that instrument 2 is more precise than instrument 1.

Numerical Summaries (3/3)

Software (1/8):

*A numerical **summary** of a distribution*:

The summary consisting of the median, the quartiles, and the smallest and largest individual observations provides a quick overall description of a distribution.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6.00	9.75	15.50	15.10	20.25	23.00

Numerical Summaries (3/3)

Software (2/8):

R is a programming language and free software environment for statistical computing and graphics. Moreover, software such as Minitab, SAS, SPSS, Stata, *etc*. can be used.

Numerical Summaries (3/3)

Software (3/8):

Calculate numerical measures for the following sample data: 6, 8, 12, 9, 14, 18, 17, 23, 21, 23.

Using R

```
> data <- c(6, 8, 12, 9, 14, 18, 17, 23, 21, 23)
> summary(data)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6.00	9.75	15.50	15.10	20.25	23.00

Numerical Summaries (3/3)

Software (4/8):

Calculate numerical measures for the following sample data: 6, 8, 12, 9, 14, 18, 17, 23, 21, 23.

Using R

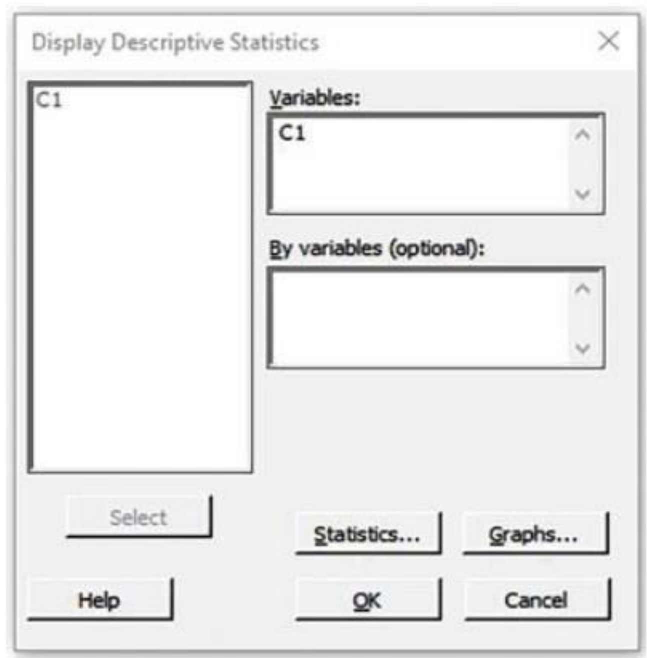
```
> data <- c(6, 8, 12, 9, 14, 18, 17, 23, 21, 23)
> mean(data)
> Median(data)
> sd(data)
> var(data)
```

Numerical Summaries (3/3)

Software (5/8):

<u>MINITAB</u>: is a powerful statistical software everyone can use.

1. Enter the data in column C1.
2. From the Menu bar, select **Stat** > **Basic Statistics** > **Display Descriptive Statistics**. This prompts the following dialog box to appear on the screen:

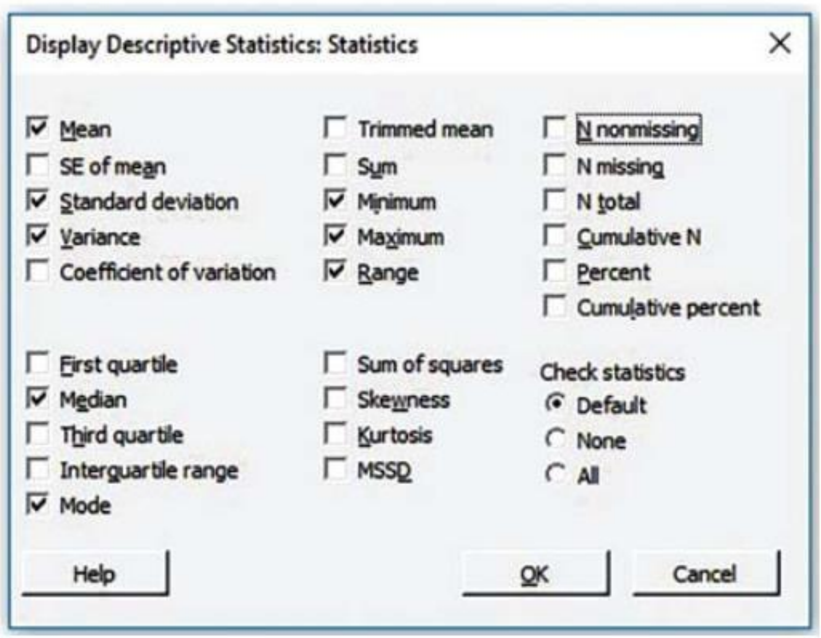


Numerical Summaries (3/3)

Software (6/8):

<u>*MINITAB*</u>: *is a powerful statistical software everyone can use.*

3. In this dialog box, enter C1 in the box under variables and click at the box **Statistics**.... Then, the following dialog box appears:



Numerical Summaries (3/3)

Software (7/8):

<u>MINITAB</u>: is a powerful statistical software everyone can use.

Statistics

Variable	Mean	StDev	Variance	Minimum	Median	Maximum	Range	Mode	N for Mode
C1	15.10	6.26	39.21	6.00	15.50	23.00	17.00	23	2

Statistics

Variable	Mean	StDev	Variance	CoefVar	Q1	Median	Q3	Range
C7	20.125	4.090	16.728	20.32	17.000	20.000	23.000	16.000

Numerical Summaries (3/3)

Software (8/8):

Using R

```
> data <- c(17,12,12,14,15,16,16,16,16,17,17,18,18,
18,19,19,20,20,20,20,20,20,20,20,21,21,21,22,22,23,
23,23,24,24,25,26,26,28,28,28)

> quantile(data)
```

0%	25%	50%	75%	100%
12	17	20	23	28