# Chapter 5
# Probability Distribution

# OBJECTIVE

The objective of this chapter are:

1. To construct a probability distribution for a random variable.

2. To find the mean, variance, and expected value for a discrete random variable.

# RANDOM VARIABLE

i. **Discrete random variables**: A **random variable** is a variable whose values is determined by the outcome of a random experiment

ii. **Continuous random variables:** are obtained from data that can be measured rather than counted

# Probability Distribution

A list of all outcomes of an experiment and the probability associated with each outcome

## Requirements for A Probability Distribution

i. The sum of the probabilities of all the events in the sample space must equal 1;

$$\sum P(X) = 1$$

ii. The probability of each event in the sample space must be between or equal to 0 and 1;

$$0 \leq P(X) \leq 1$$

# Mean of A Discrete Probability Distribution

- The mean is a typical value used to represent the central location of a distribution.
- The mean of a discrete random variable x is the value that is expected to occur per repetition, on average, if an experiment is repeated a large number of times.
- The mean of a probability distribution is also referred to as its expected value;

$$\mu = X_1 \cdot P(X_1) + X_2 \cdot P(X_2) + X_3 \cdot P(X_3) + \ldots + X_n \cdot P(X_n)$$
$$= \sum X \cdot P(X)$$

# Variance of A Discrete Probability Distribution

- Measures the amount spread in a distribution
- Variance,

$$\sigma^2 = \sum \left[ X^2 \cdot P(X) \right] - \mu^2$$

- Standard deviation.

$$\sigma = \sqrt{\sigma^2} \quad \text{OR} \quad \sigma = \sqrt{\sum \left[ X^2 \cdot P(X) \right] - \mu^2}$$

# Constructing Discrete Probability Distribution

- E.g Probability experiment of tossing three coins.
- Sample space: TTT, TTH, THT, HTT, HHT, HTH, THH, HHH
- If X is the random variable for the number of heads, then X assumes the value 0, 1, 2 or 3.
- Probabilities values of X can be determined as follows:

| **No heads** | **1 head** | **2 heads** | **3 heads** |
|---|---|---|---|
| **TTT** | TTH,THT,HTT | HHT,HTH,THH | HHH |
| **1/8** | 3/8 | 3/8 | 1/8 |

| | | | | |
|---|---|---|---|---|
| X | | | | |
| P(X) | | | | |

# NORMAL DISTRIBUTION

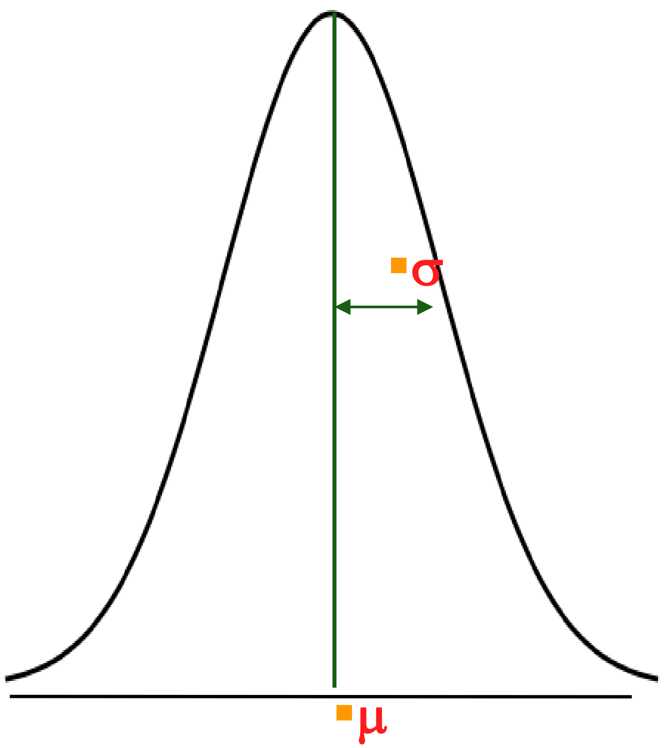
- Normal Probability Distribution
- Standard Normal Probability Distribution
- Continuous Probability Distribution
- Applications of the Normal Distribution

# The Normal Distribution

Mathematically defined as:

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

- Since $\pi$ and $e$ are constants, we only have to determine $\mu$ (the population mean) and $\sigma$ (the population standard deviation) to graph the mathematical function of any variable we are interested in.
- The Normal distribution has the shape of a "bell curve" with parameters $\mu$ and $\sigma^2$ that determine the center and spread.




Don't worry, understanding this is not necessary to understand the normal distribution, only a helpful aside for the mathematically inclined

# Properties of normal distribution

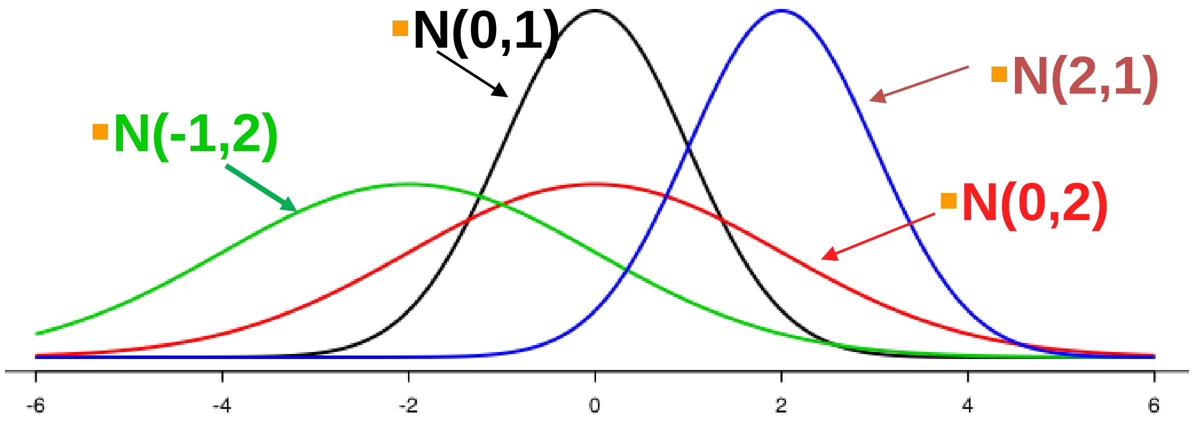
A normal probability distribution, when plotted, gives a bell-shaped curve such that

1. The total area under the curve is 1.0
2. The curve is symmetric about the mean
3. The two tails of the curve extend indefinitely

- **Normal Distribution** – specific type of distribution that assumes a characteristic **bell shape** and is perfectly **symmetrical**

# Different Normal Distributions

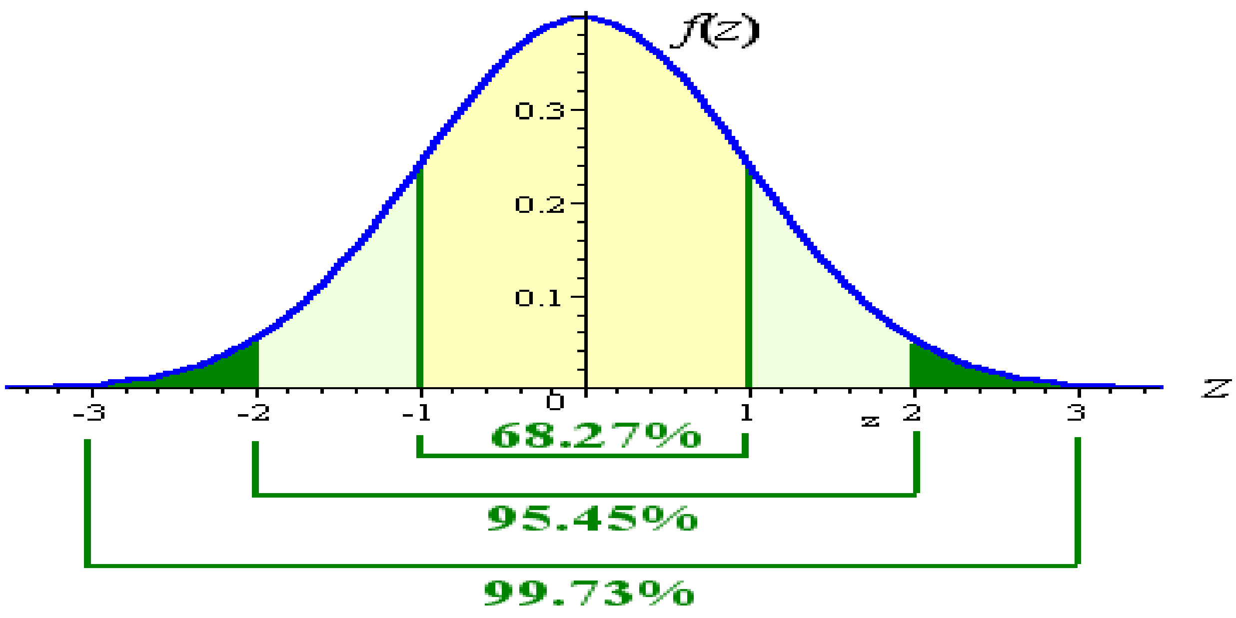
Each different value of $\mu$ and $\sigma^2$ gives a different Normal distribution, denoted $N(\mu,\sigma^2)$




- We can adjust values of $\mu$ and $\sigma^2$ to provide the best approximation to observed data
- If $\mu = 0$ and $\sigma^2 = 1$, we have the **Standard Normal** distribution

# The Standard Normal Distribution

- The standard normal distribution is a special case of the normal distribution.
- The standard normal distribution is the normal distribution with $\mu$=0 and $\sigma$=1.

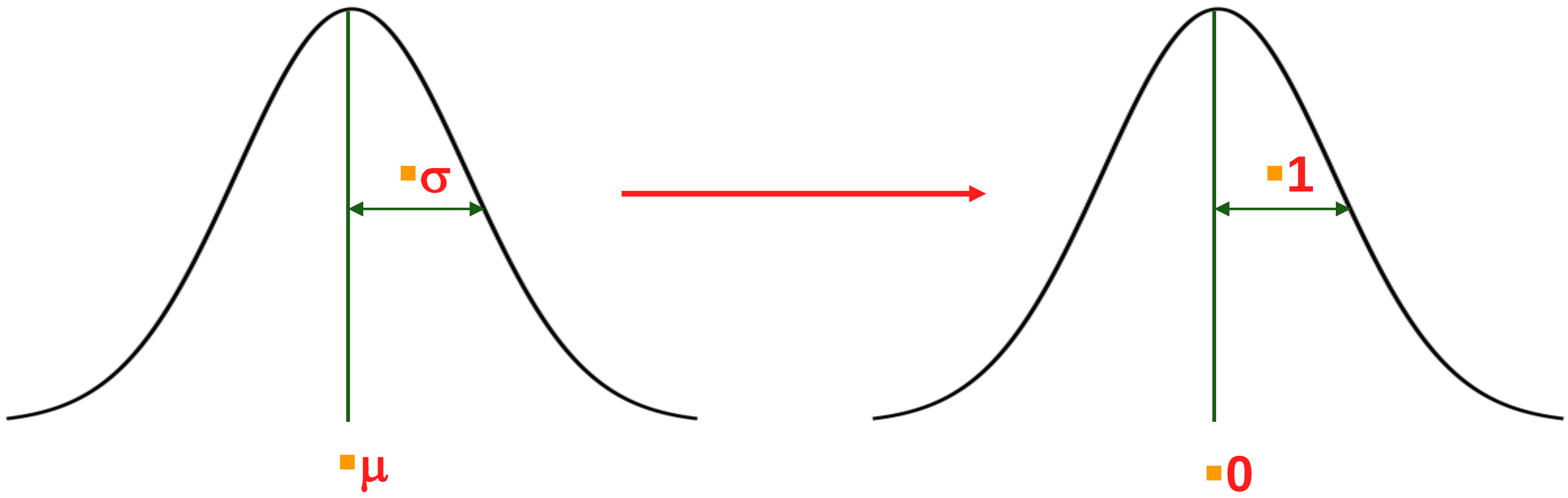


# The Standard Normal Distribution

- The units for the standard normal distribution curve are denoted by $z$ and are called the $z$ values or $z$ scores.
- They are also called standard units or standard scores.
- The $z$ value for a point on the horizontal gives the distance between the mean and that point in terms of the standard deviation.

# Standardization

If we only have a standard normal table, then we need to ***transform*** our non-standard normal distribution into a standard one. This process is called **standardization**




- The standard normal distribution is a normal distribution with a mean of 0 and standard deviation of 1.

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

# Standardizing A Normal Distribution

For a normal random variable $x$, a particular value of $x$ can be converted to its corresponding $z$ value by using the formula

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

where $\mu$ and $\sigma$ are the mean and standard deviation of the normal distribution of $x$, respectively.

# Continuous Probability Distribution

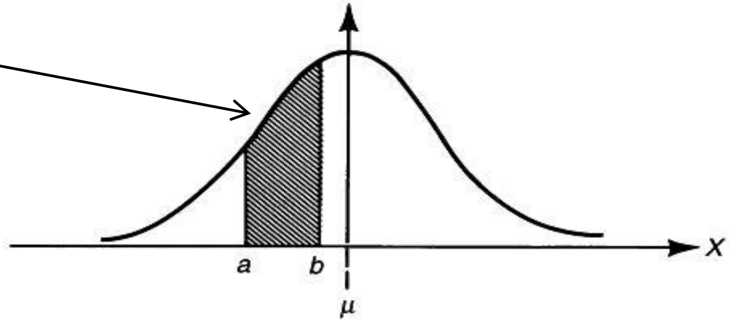
The probability distribution of a continuous random variable possesses the following two characteristics:

1. The probability that $x$ assumes a value in any interval lies in the range 0 to 1
2. The total probability of all the intervals within which $x$ can be assume a value is 1.0.

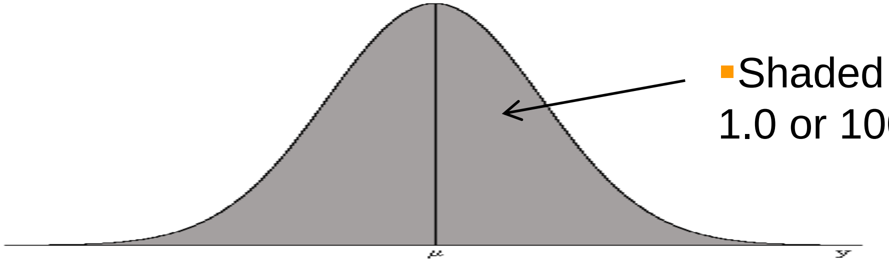
# Continuous Probability Distribution

1. The area under the probability distribution curve of a continuous random variable between any two points is between 0 and 1.

- Shaded area is between 0 and 1



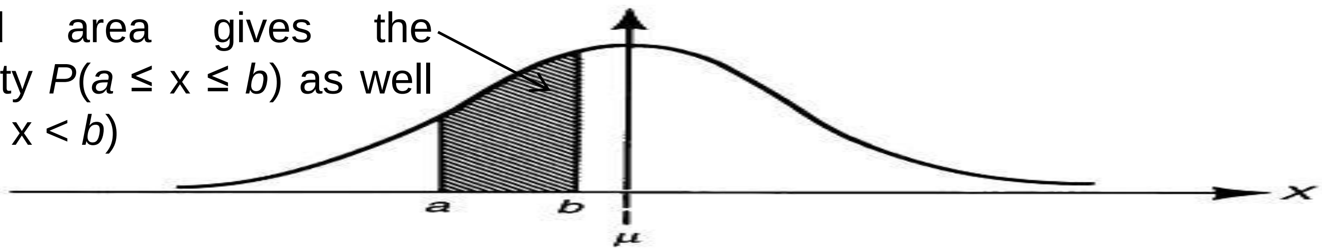

2. The total area under the probability distribution curve of a continuous random variable is always 1.0 or 100%.



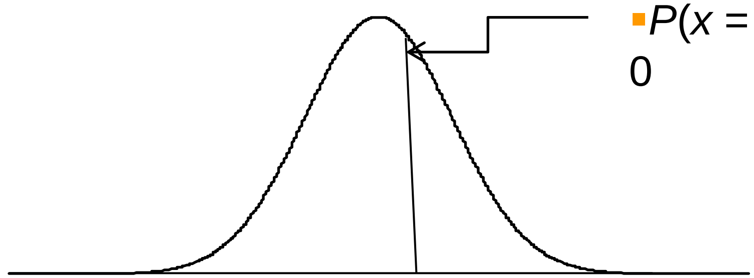
- Shaded area is 1.0 or 100%

# Continuous Probability Distribution

The probability that a continuous random variable *x* assumes a value within a certain interval is given by the area under the curve between the two limits of the interval.

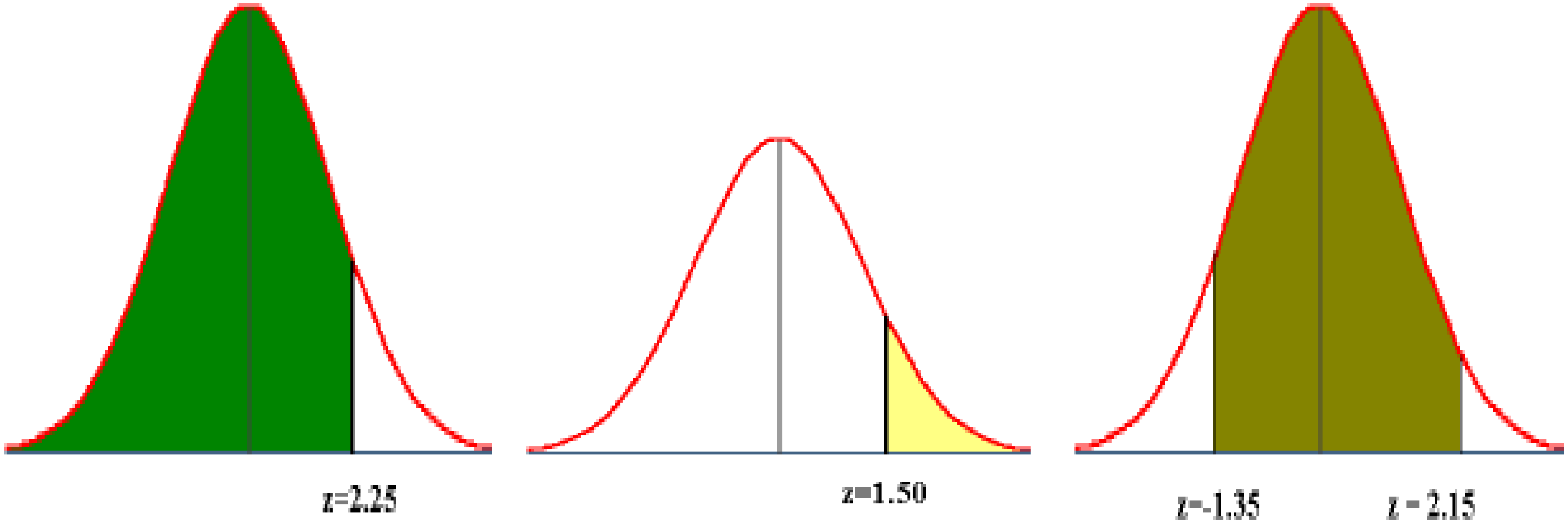



The probability that a continuous random variable *x* assumes a single value is always **zero**.



# Example

1. Find the area to the left of z = 2.25
2. Find the area to the right of z = 1.50
3. Find the area between z = -1.35 and z = 2.15



# Your Turn

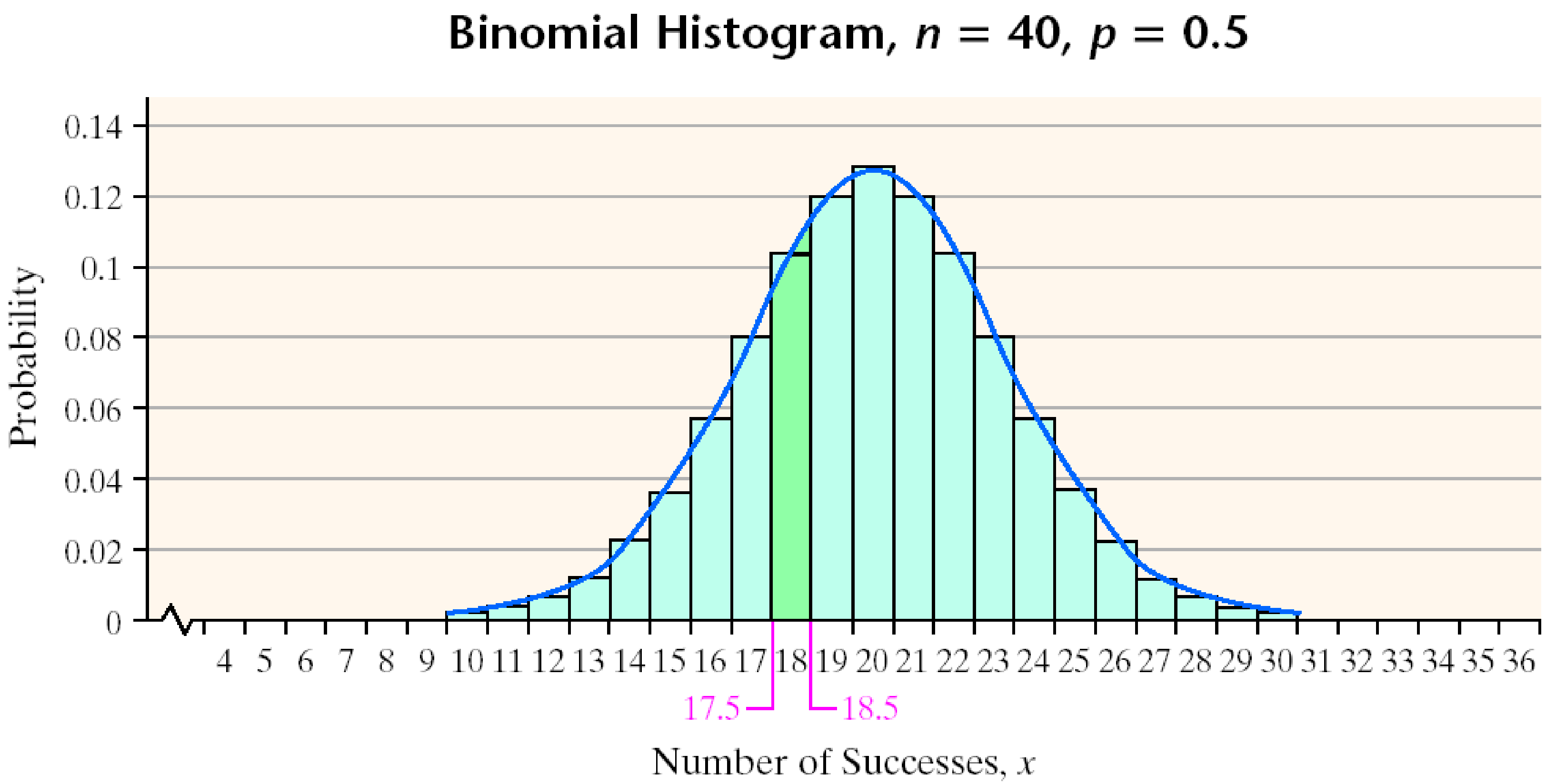
Find the area for each:

1. area to the left of $z = -1.04$ ___________________
2. area to the right of $z = 1.07$ ___________________
3. area between $z = 0$ and $z = 2.75$ _______________
4. area between $z = -1.00$ and $z = 1.00$ ___________

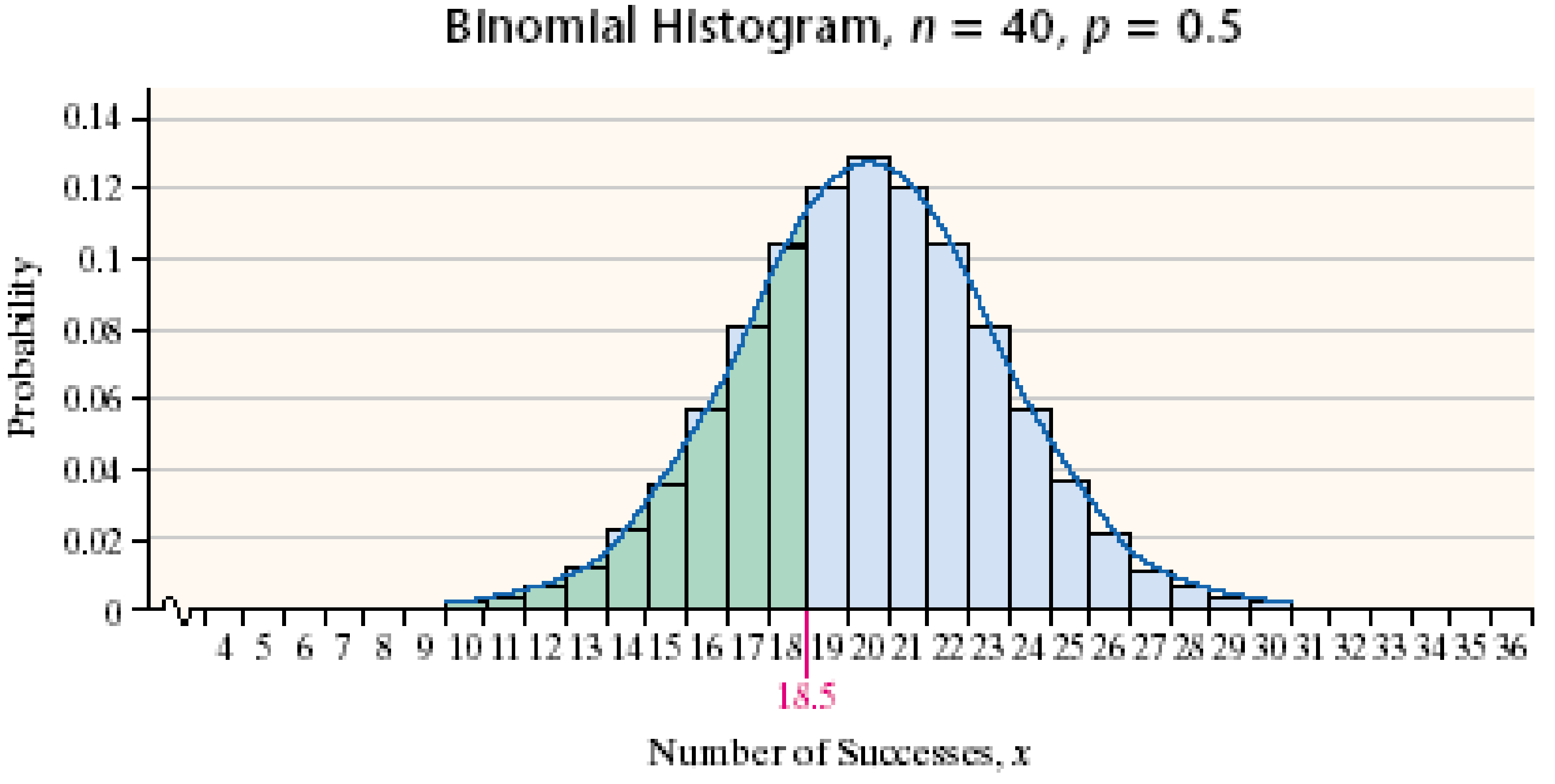
# The Normal Approximation to the Binomial Probability Distribution

- As the number of trials *n* in a binomial experiment increase, the probability distribution of the random variable *X* becomes symmetric and bell-shaped. As a general rule of thumb, if $np(1-p) \geq 10$, then the probability distribution will be approximately symmetric and bell-shaped.

- If $np(1-p) \geq 10$, then the binomial random variable X is approximately normally distributed with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$




- $P(X = 18) \approx P(17.5 \leq X \leq 18.5)$




- $P(X \leq 18) \approx P(X \leq 18.5)$

# SUMMARY

| Exact Binomial | Approximate Normal |
|---|---|
| $P(X = a)$ | $P(a - 0.5 < X < a + 0.5)$ |
| $P(X \leq a)$ | $P(X < a + 0.5)$ |
| $P(X \geq a)$ | $P(X > a - 0.5)$ |
| $P(a \leq X \leq b)$ | $P(a - 0.5 < X < b + 0.5)$ |

**NOTE:**

For $P(X < a) = P(X \leq a - 1)$, so rewrite $P(X < 5)$ as $P(X \leq 4)$

# Using the Binomial probability Distribution Function

## EXAMPLE ...

According to the United States Census Bureau, 18.3% of all households have 3 or more cars.

(a) In a random sample of 200 households, what is the probability that at least 30 have 3 or more cars?

(b) In a random sample of 200 households, what is the probability that less than 15 have 3 or more cars?

(c) Suppose in a random sample of 500 households, it is determined that 110 have 3 or more cars. Is this result unusual? What might you conclude?

# Binomial Probability Distribution

A binomial random variable X is defined to the number of "successes" in n independent trials where the P(success)=p is constant.

$$X \sim BIN(n, p)$$

In the definition above notice the following conditions need to be satisfied for a binomial experiment:

1. There is a fixed number of *n* trials carried out.
2. Each trial can have only two outcomes.
3. The trials must be independent.
4. The probabilities (*p* and *q*), must remain constant for each trial, $p + q = 1$

# Binomial Distribution

- If $X$ ~ BIN($n$, $p$), then

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad x = 0,1,\ldots,n.$$

where

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 1, \text{ also } 0! = 1 \text{ and } 1! = 1$$

$$\binom{\text{n}}{\text{x}} = \text{"n choose x"} = \text{the number of ways to obtain}$$

$$x \text{ "successes" in } n \text{ trials.}$$

$$P(\text{"}success\text{"}) = p$$

# Binomial Distribution

## Example…

a. A coin is tossed 3 times. Find the probability of getting exactly two head

b. A survey found that one out of five Americans says he or she has visited a doctor in any given month. If 10 people are selected at random, find the probability that exactly 3 will have visited a doctor last month.

c. A man sent 5 sets of application form for 5 jobs. If the probability of success for every application is 1/8, find the probability that:

   i. One application out of 5 sets of application are successful
   ii. At least one application out of 5 sets of application are successful
   iii. Less than two applications out of 5 set of application are successful.

# Mean, Variance and Standard Deviation for Binomial Distribution

Mean $$\mu = np$$

Variance $$\sigma^2 = npq$$

Standard Deviation $$\sigma = \sqrt{npq}$$

# Example

In Pittsburgh, 57% of the days in a year are cloudy. Find the mean, variance, and standard deviation for the number of cloudy days during the month of June. What can you conclude?

Solution: There are 30 days in June. Using n=30, p = 0.57, and q = 0.43, you can find the mean variance and standard deviation as shown.

- Mean: $\mu = np = 30(0.57) = 17.1$
- Variance: $\sigma 2 = npq = 30(0.57)(0.43) = 7.353$
- Standard Deviation: $\sigma = \sqrt{npq} = \sqrt{7.353} \approx 2.71$

# The Poisson Distribution

## Overview

- When there is a large number of trials, but a small probability of success, binomial calculation becomes impractical
  - Example: Number of deaths from horse kicks in the Army in different years

- Simeon D. Poisson (1781-1840)

- The binomial distribution approaches the Poisson distribution if n is large and p small

# Understanding Poisson Distribution

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

- The number of occurrences is a discrete random variable.
- The occurrences are random.
- The occurrences are independent.

# Poisson Probability Distribution

- The random variable X is said to follow the Poisson probability distribution if it has the probability function:

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad \text{for x} = 0, 1, 2, \ldots$$

- where
  - P(x) = the probability of x successes over a given period of time or space, given λ
  - λ = the expected number of successes per time or space unit; λ > 0
  - e = 2.71828 (the base for natural logarithms)

- The mean and variance of the Poisson probability distribution are:

$$\mu_x = E(X) = \lambda \quad and \quad \sigma_x^2 = E[(X-\mu)^2] = \lambda$$