Probability Distributions

Chapter 6

GOALS

- Define the terms probability distribution and random variable.
- Distinguish between discrete and continuous probability distributions.
- Calculate the mean, variance, and standard deviation of a discrete probability distribution.
- Describe the characteristics of and compute probabilities using the binomial probability distribution.
- Describe the characteristics of and compute probabilities using the hypergeometric probability distribution.
- Describe the characteristics of and compute probabilities using the Poisson

What is a Probability Distribution?

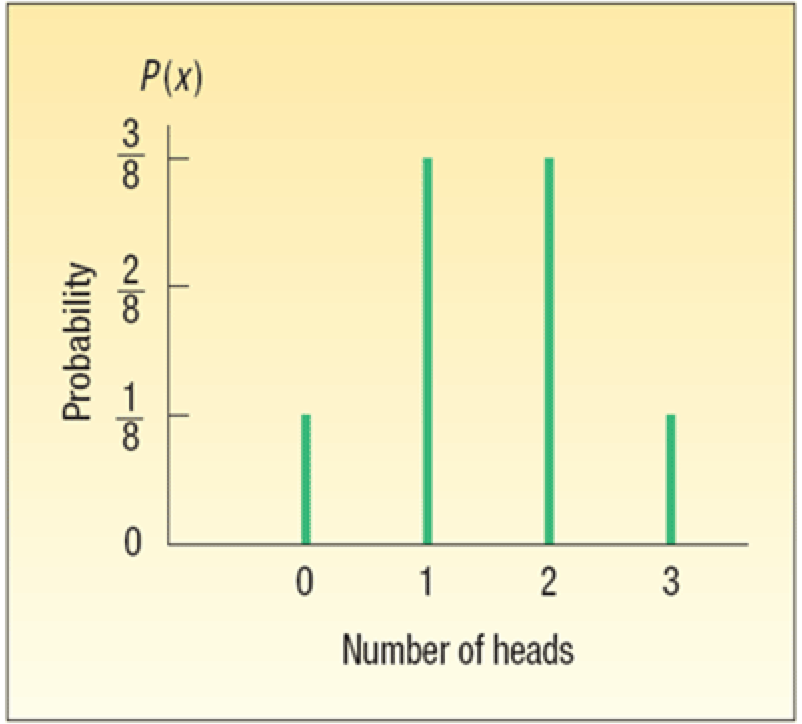
PROBABILITY DISTRIBUTION A listing of all the outcomes of an experiment and the probability associated with each outcome.

Experiment: Toss a coin three times. Observe the number of heads. The possible results are: zero heads, one head, two heads, and three heads.
What is the probability distribution for the number of heads?

Possible Result	Coin Toss First	Coin Toss Second	Coin Toss Third	Number of Heads
1	T	T	T	0
2	T	T	H	1
3	T	H	T	1
4	T	H	H	2
5	H	T	T	1
6	H	T	H	2
7	H	H	T	2
8	H	H	H	3

Probability Distribution of Number of Heads Observed in 3 Tosses of a Coin

Number of Heads, x	**Probability of Outcome, $P(x)$**
0	$\frac{1}{8} = .125$
1	$\frac{3}{8} = .375$
2	$\frac{3}{8} = .375$
3	$\frac{1}{8} = .125$
Total	$\frac{8}{8} = 1.000$



Characteristics of a Probability Distribution

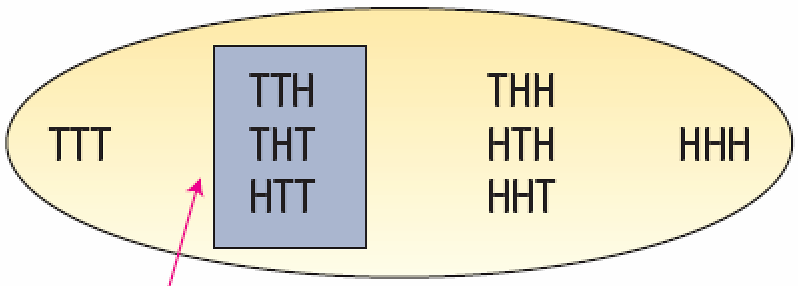
CHARACTERISTICS OF A PROBABILITY DISTRIBUTION

1. The probability of a particular outcome is between 0 and 1 inclusive.
2. The outcomes are mutually exclusive events.
3. The list is exhaustive. So the sum of the probabilities of the various events is equal to 1.

Random Variables

Random variable - a quantity resulting from an experiment that, by chance, can assume different values.

Possible *outcomes* for three coin tosses



The *event* {one head} occurs and the *random variable* $x = 1$.

Types of Random Variables

- Discrete Random Variable can assume only certain clearly separated values. It is usually the result of <u>counting</u> something

- Continuous Random Variable can assume an infinite number of values within a given range. It is usually the result of some type of <u>measurement</u>

Discrete Random Variables - Examples

- The number of students in a class.
- The number of children in a family.
- The number of cars entering a carwash in a hour.
- Number of home mortgages approved by Coastal Federal Bank last week.

Continuous Random Variables - Examples

- The distance students travel to class.
- The time it takes an executive to drive to work.
- The length of an afternoon nap.
- The length of time of a particular phone call.

Features of a Discrete Distribution

The main features of a discrete probability distribution are:

- The sum of the probabilities of the various outcomes is 1.00.
- The probability of a particular outcome is between 0 and 1.00.
- The outcomes are mutually exclusive.

The Mean of a Probability Distribution

MEAN

- The mean is a typical value used to represent the central location of a probability distribution.
- The mean of a probability distribution is also referred to as its **expected value.**

MEAN OF A PROBABILITY DISTRIBUTION $\mu = \Sigma[xP(x)]$ **[6–1]**

The Variance, and Standard Deviation of a Probability Distribution

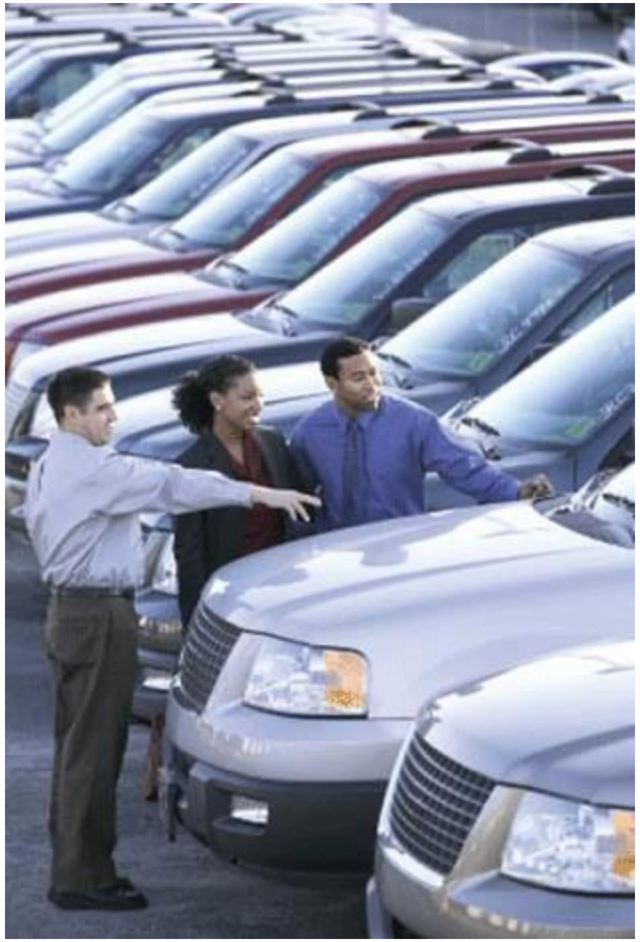
Variance and Standard Deviation

- Measures the amount of spread in a distribution
- The computational steps are:
  1. Subtract the mean from each value, and square this difference.
  2. Multiply each squared difference by its probability.
  3. Sum the resulting products to arrive at the variance.

  The standard deviation is found by taking the positive square root of the variance.

VARIANCE OF A PROBABILITY DISTRIBUTION $\sigma^2 = \Sigma[(x - \mu)^2 P(x)]$ **[6–2]**

Mean, Variance, and Standard Deviation of a Probability Distribution - Example



John Ragsdale sells new cars for Pelican Ford. John usually sells the largest number of cars on Saturday. He has developed the following probability distribution for the number of cars he expects to sell on a particular Saturday.

Number of Cars Sold, x	Probability, $P(x)$
0	.10
1	.20
2	.30
3	.30
4	.10
Total	1.00

Mean of a Probability Distribution - Example

$$\mu = \Sigma[xP(x)]$$
$$= 0(.10) + 1(.20) + 2(.30) + 3(.30) + 4(.10)$$
$$= 2.1$$

Number of Cars Sold, x	Probability, $P(x)$	$x \cdot P(x)$
0	.10	0.00
1	.20	0.20
2	.30	0.60
3	.30	0.90
4	.10	0.40
Total	1.00	μ = 2.10

Variance and Standard Deviation of a Probability Distribution - Example

Number of Cars Sold, x	Probability, $P(x)$	$(x - \mu)$	$(x - \mu)^2$	$(x - \mu)^2 P(x)$
0	.10	0 − 2.1	4.41	0.441
1	.20	1 − 2.1	1.21	0.242
2	.30	2 − 2.1	0.01	0.003
3	.30	3 − 2.1	0.81	0.243
4	.10	4 − 2.1	3.61	0.361
				$\sigma^2 = 1.290$

Binomial Probability Distribution

Characteristics of a Binomial Probability Distribution

- There are only two possible outcomes on a particular trial of an experiment.
- The outcomes are mutually exclusive,
- The random variable is the result of counts.
- Each trial is *independent* of any other trial

Binomial Probability Formula

BINOMIAL PROBABILITY FORMULA $$P(x) = {}_nC_x\, \pi^x (1 - \pi)^{n - x} \quad [6\text{–}3]$$

where:

C denotes a combination.
n is the number of trials.
x is the random variable defined as the number of successes.
π is the probability of a success on each trial.

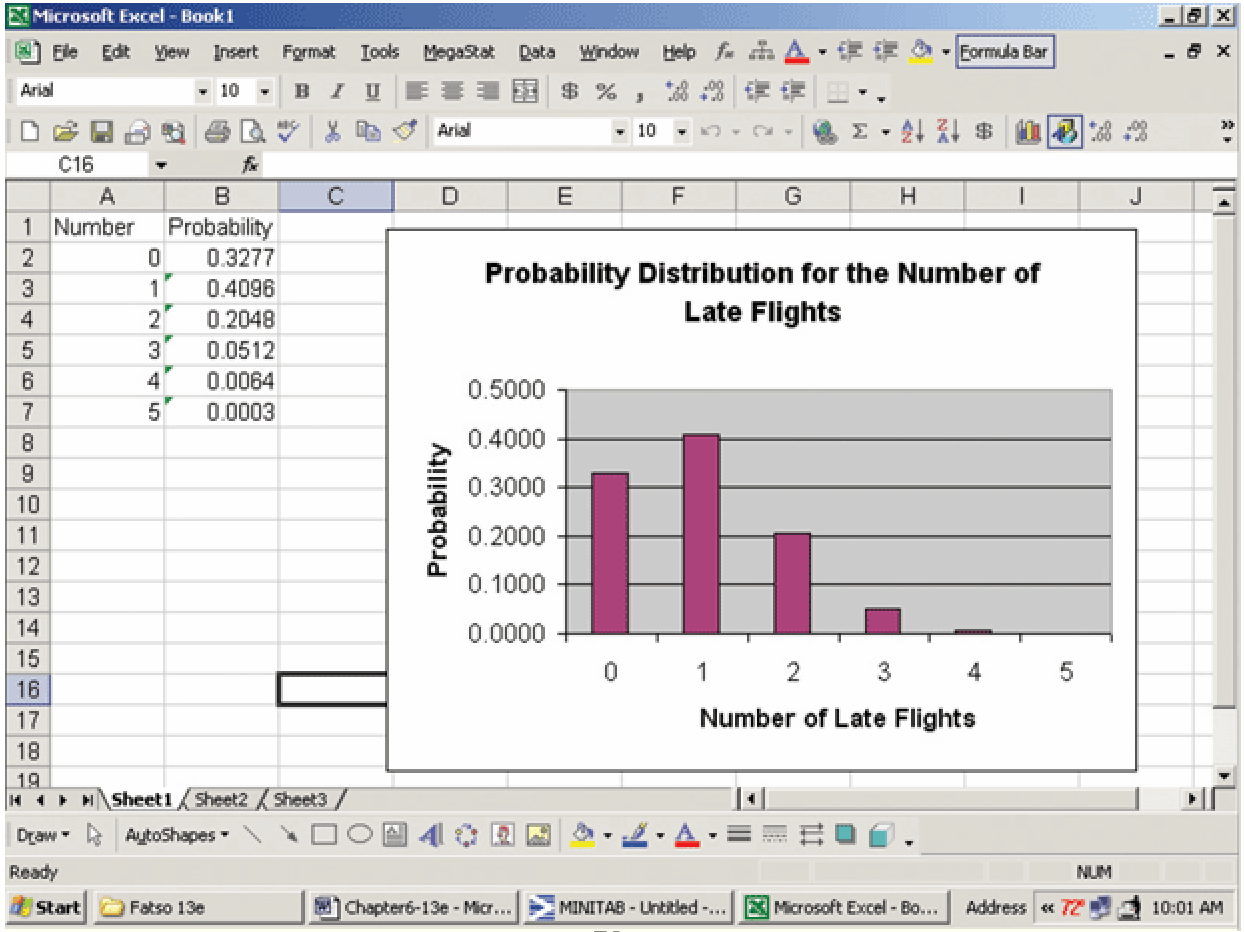
Binomial Probability - Example

There are five flights daily from Pittsburgh via US Airways into the Bradford, Pennsylvania, Regional Airport. Suppose the probability that any flight arrives late is .20.

What is the probability that none of the flights are late today?

$$
\begin{aligned}
P(x=0) &= {}_nC_x \pi^x (1-\pi)^{n-x} \\
&= {}_5C_0 (.20)^0 (1\text{-}.20)^{5\text{-}0} \\
&= (1)(1)(.3277) \\
&= 0.3277
\end{aligned}
$$

Binomial Probability - Excel



Binomial Dist. – Mean and Variance

MEAN OF A BINOMIAL DISTRIBUTION	$\mu = n\pi$	**[6–4]**

VARIANCE OF A BINOMIAL DISTRIBUTION	$\sigma^2 = n\pi(1 - \pi)$	**[6–5]**

Binomial Dist. – Mean and Variance: Example

For the example regarding the number of late flights, recall that π =.20 and n = 5.

What is the average number of late flights?

What is the variance of the number of late flights?

$$\mu = n\pi$$

$$= (5)(0.20) = 1.0$$

$$\sigma^2 = n\pi(1-\pi)$$

$$= (5)(0.20)(1-0.20)$$

$$= (5)(0.20)(0.80)$$

$$= 0.80$$

Binomial Dist. – Mean and Variance: Another Solution

Number of Late Flights, X	$P(X)$	$XP(X)$	$X - \mu$	$(X - \mu)^2$	$(X - \mu)^2 P(X)$
0	0.3277	0.0000	−1	1	0.3277
1	0.4096	0.4096	0	0	0
2	0.2048	0.4096	1	1	0.2048
3	0.0512	0.1536	2	4	0.2048
4	0.0064	0.0256	3	9	0.0576
5	0.0003	0.0015	4	16	0.0048
		$\mu = 1.0000$			$\sigma^2 = 0.7997$

Binomial Distribution - Table

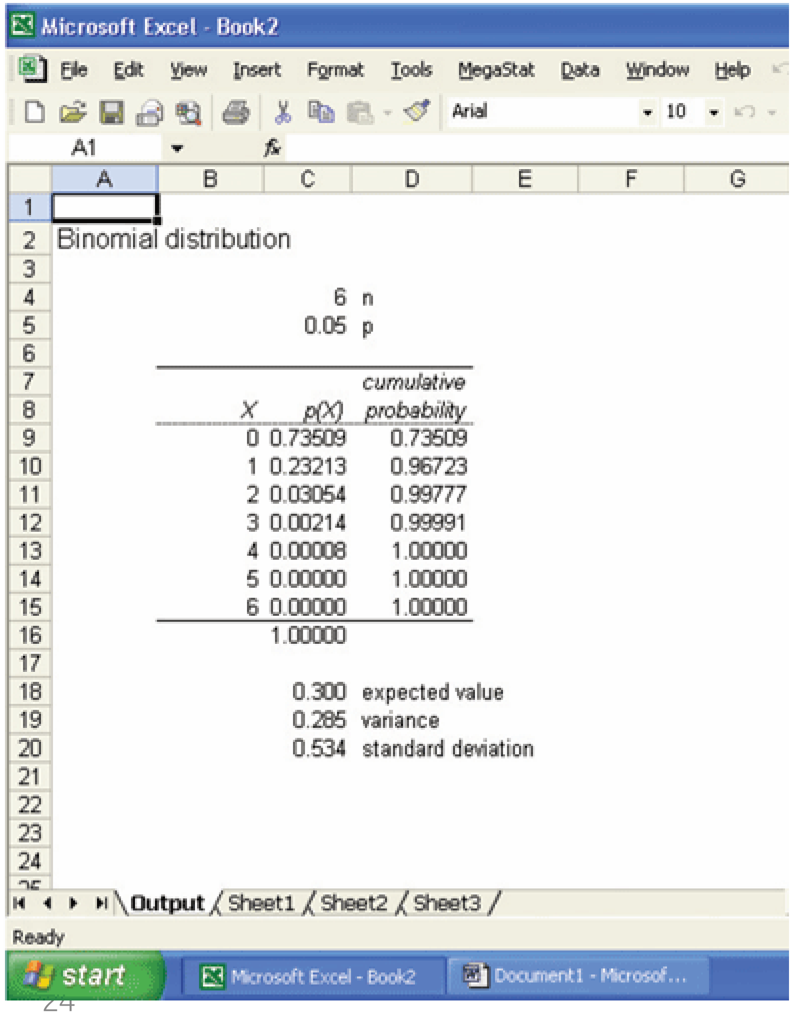
Five percent of the worm gears produced by an automatic, high-speed Carter-Bell milling machine are defective. What is the probability that out of six gears selected at random none will be defective? Exactly one? Exactly two? Exactly three? Exactly four? Exactly five? Exactly six out of six?

TABLE 6–2 Binomial Probabilities for $n = 6$ and Selecte Values of π

$n = 6$ Probability											
$x \backslash \pi$	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95
0	$\boxed{.735}$	.531	.262	.118	.047	.016	.004	.001	.000	.000	.000
1	.232	.354	.393	.303	.187	.094	.037	.010	.002	.000	.000
2	.031	.098	.246	.324	.311	.234	.138	.060	.015	.001	.000
3	.002	.015	.082	.185	.276	.313	.276	.185	.082	.015	.002
4	.000	.001	.015	.060	.138	.234	.311	.324	.246	.098	.031
5	.000	.000	.002	.010	.037	.094	.187	.303	.393	.354	.232
6	.000	.000	.000	.001	.004	.016	.047	.118	.262	.531	.735

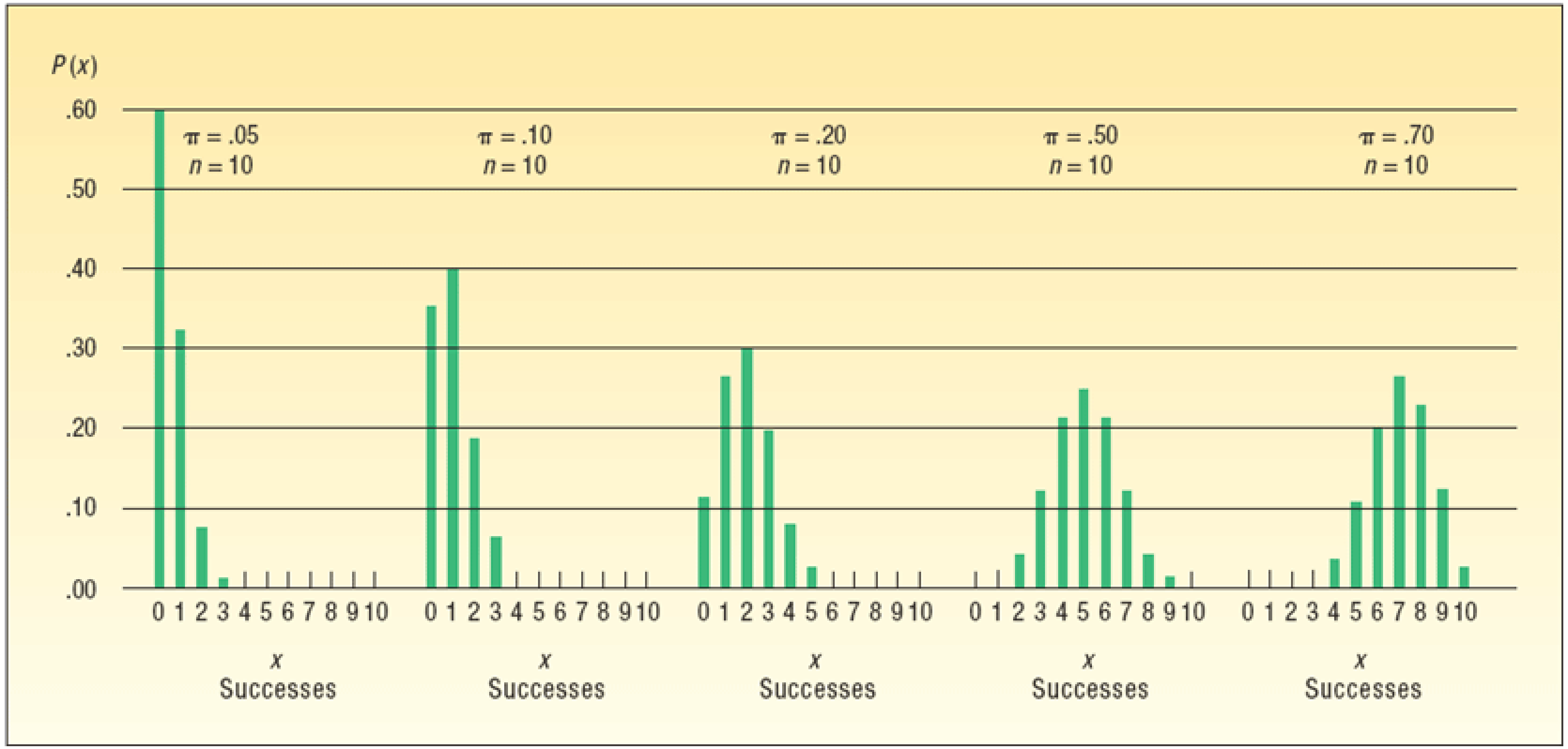
Binomial Distribution - MegaStat

Five percent of the worm gears produced by an automatic, high-speed Carter-Bell milling machine are defective. What is the probability that out of six gears selected at random none will be defective? Exactly one? Exactly two? Exactly three? Exactly four? Exactly five? Exactly six out of six?

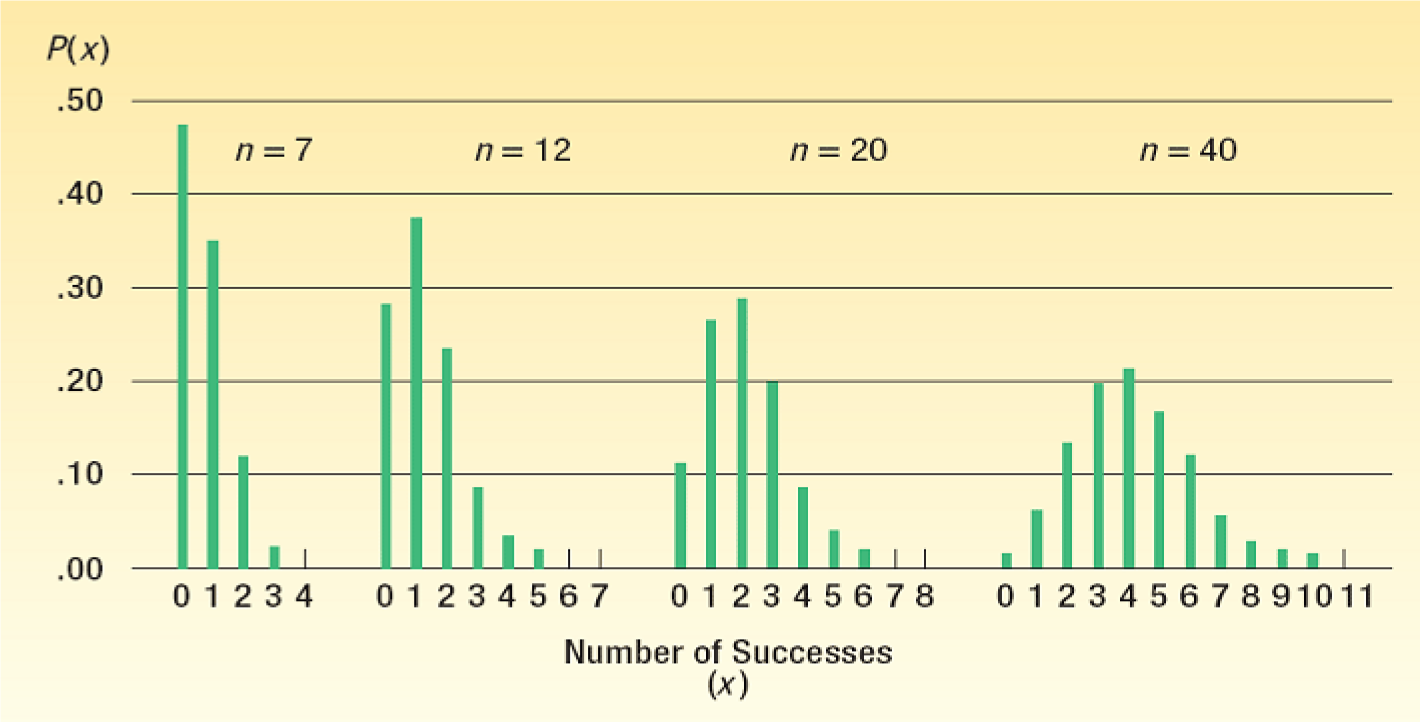


Binomial – Shapes for Varying π (*n* constant)

CHART 6–2 Graphing the Binomial Probability Distribution for a π of .05, .10, .20, .50, and .70 and an *n* of 10



Binomial – Shapes for Varying *n* (π constant)



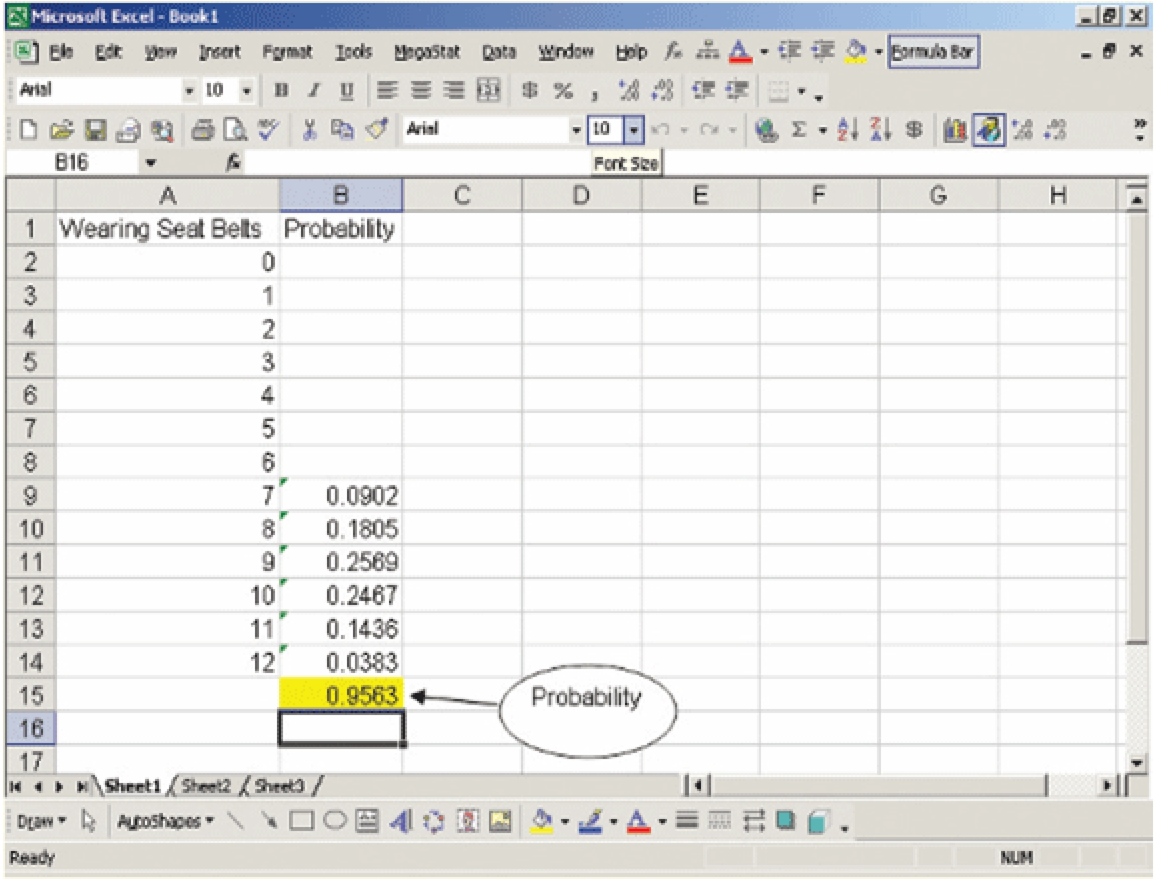
Cumulative Binomial Probability Distributions

A study in June 2003 by the Illinois Department of Transportation concluded that 76.2 percent of front seat occupants used seat belts. A sample of 12 vehicles is selected. What is the probability the front seat occupants in at least 7 of the 12 vehicles are wearing seat belts?

$$
\begin{aligned}
P(X \geq 7 \mid n = 12 \text{ and } \pi = .762)& \\
&= P(X=7) + P(X=8) + P(X=9) + P(X=10) + P(X=11) + P(X=12) \\
&= .0902 + .1805 + .2569 + .2467 + .1436 + .0383 \\
&= .9563
\end{aligned}
$$

$$
\begin{aligned}
P(x = 7 \mid n = 12 \text{ and } \pi = .762)& \\
&= {}_{12}C_7(.762)^7(1 - .762)^{12-7} = 792(.149171)(.000764) = .0902
\end{aligned}
$$

Cumulative Binomial Probability Distributions - Excel



Finite Population

A finite population is a population consisting of a fixed number of known individuals, objects, or measurements. Examples include:

- The number of students in this class.
- The number of cars in the parking lot.
- The number of homes built in Blackmoor

Hypergeometric Distribution

The hypergeometric distribution has the following characteristics:

- There are only 2 possible outcomes.
- The probability of a success is not the same on each trial.
- It results from a count of the number of successes in a fixed number of trials.

Hypergeometric Distribution

Use the hypergeometric distribution to find the probability of a specified number of successes or failures if:

- the sample is selected from a finite population without replacement
- the size of the sample n is greater than 5% of the size of the population N *(i.e. $n/N \geq .05$)*

Hypergeometric Distribution

HYPERGEOMETRIC DISTRIBUTION

$$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n} \qquad \textbf{[6–6]}$$

where:

N is the size of the population.
S is the number of successes in the population.
x is the number of successes in the sample. It may be 0, 1, 2, 3, ...
n is the size of the sample or the number of trials.
C is the symbol for a combination.

Hypergeometric Distribution - Example

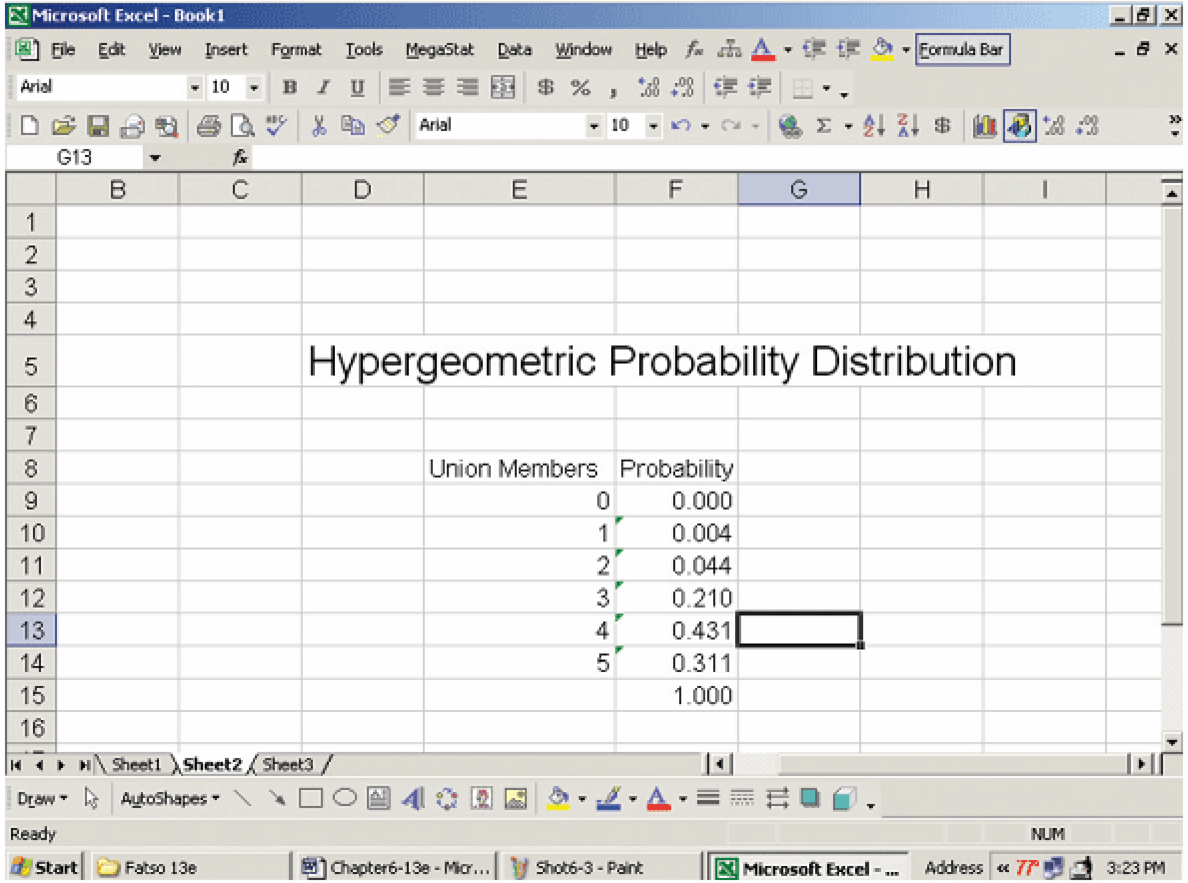
PlayTime Toys, Inc., employs 50 people in the Assembly Department. Forty of the employees belong to a union and ten do not. Five employees are selected at random to form a committee to meet with management regarding shift starting times. What is the probability that four of the five selected for the committee belong to a union?

Hypergeometric Distribution - Example

N is 50, the number of employees.
S is 40, the number of union employees.
x is 4, the number of union employees selected.
n is 5, the number of employees selected.

$$P(4) = \frac{({}_{40}C_4)({}_{50-40}C_{5-4})}{{}_{50}C_5} = \frac{\left(\frac{40!}{4!36!}\right)\left(\frac{10!}{1!9!}\right)}{\frac{50!}{5!45!}} = \frac{(91{,}390)(10)}{2{,}118{,}760} = .431$$

Hypergeometric Distribution - Excel



Poisson Probability Distribution

The **Poisson probability distribution** describes the number of times some event occurs during a specified interval. The interval may be time, distance, area, or volume.

- Assumptions of the Poisson Distribution
  (1) The probability is proportional to the length of the interval.
  (2) The intervals are independent.

Poisson Probability Distribution

The Poisson distribution can be described mathematically using the formula:

POISSON DISTRIBUTION $$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad [6\text{–}7]$$

where:

μ (mu) is the mean number of occurrences (successes) in a particular interval.
e is the constant 2.71828 (base of the Napierian logarithmic system).
x is the number of occurrences (successes).
$P(x)$ is the probability for a specified value of x.

Poisson Probability Distribution

- The mean number of successes can be determined in binomial situations by $n\pi$, where n is the number of trials and π the probability of a success.
- The variance of the Poisson distribution is also equal to $n\pi$.

Poisson Probability Distribution - Example

Assume baggage is rarely lost by Northwest Airlines. Suppose a random sample of 1,000 flights shows a total of 300 bags were lost. Thus, the arithmetic mean number of lost bags per flight is 0.3 (300/1,000). If the number of lost bags per flight follows a Poisson distribution with u = 0.3, find the probability of not losing any bags.

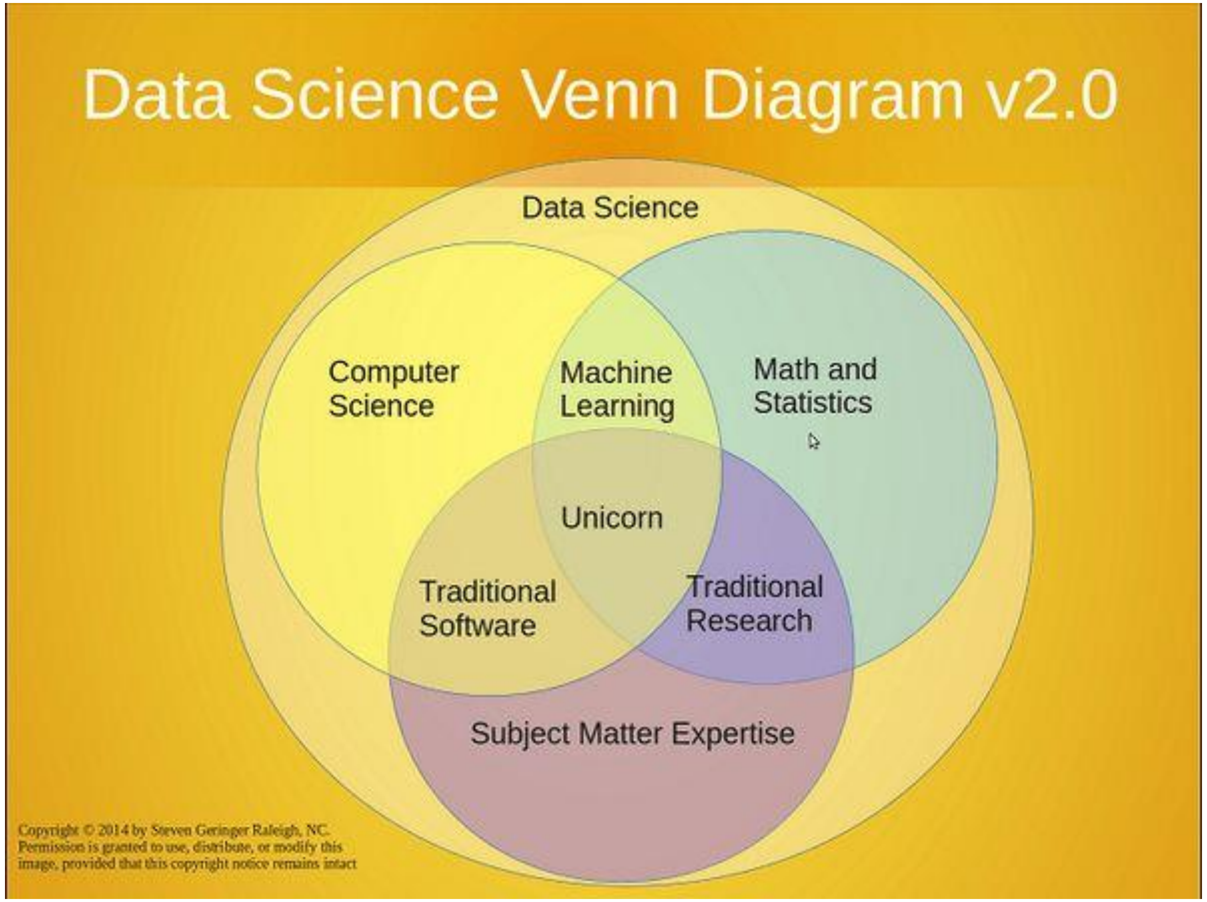
$$P(0) = \frac{\mu^{x} e^{-u}}{x!} = \frac{0.3^{0} e^{-.3}}{0!} = .7408$$

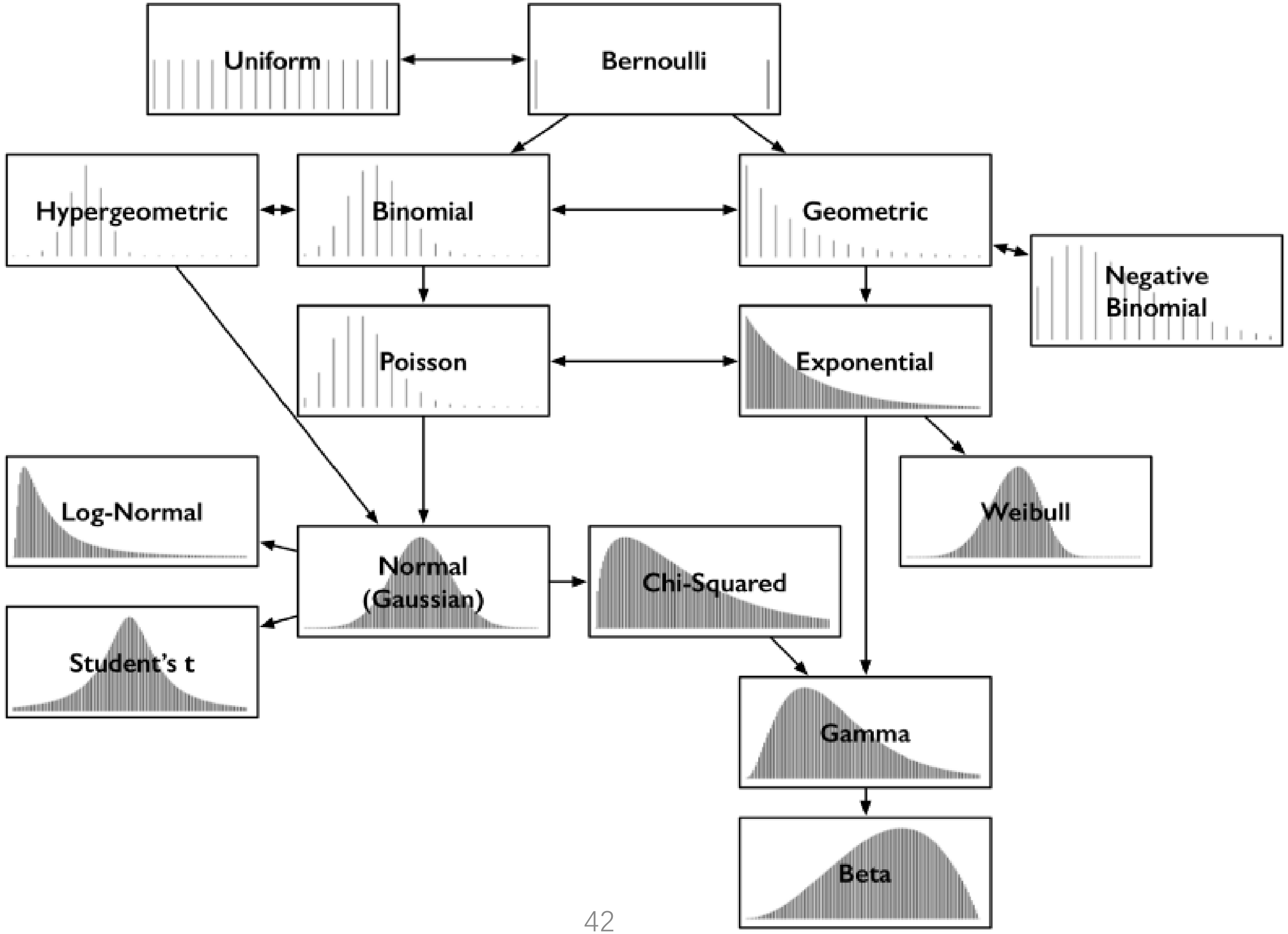
Poisson Probability Distribution - Table

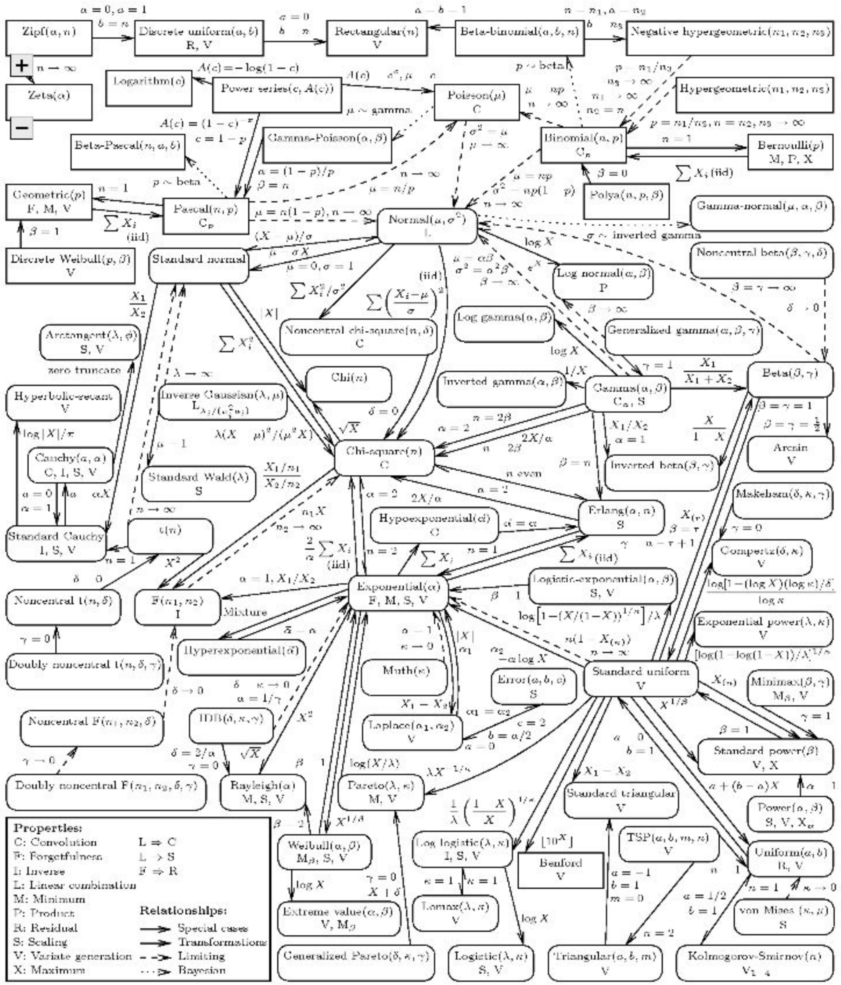
Assume baggage is rarely lost by Northwest Airlines. Suppose a random sample of 1,000 flights shows a total of 300 bags were lost. Thus, the arithmetic mean number of lost bags per flight is 0.3 (300/1,000). If the number of lost bags per flight follows a Poisson distribution with mean = 0.3, find the probability of not losing any bags

TABLE 6–6 Poisson Table for Various Values of μ (from Appendix B.5)

	μ								
x	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000


Data Science Venn Diagram v2.0
Data Science
Computer Science
Machine Learning
Math and Statistics
Unicorn
Traditional Software
Traditional Research
Subject Matter Expertise
Copyright © 2014 by Steven Geringer Raleigh, NC. Permission is granted to use, distribute, or modify this image, provided that this copyright notice remains intact


Uniform
Bernoulli
Hypergeometric
Binomial
Geometric
Negative Binomial
Poisson
Exponential
Weibull
Log-Normal
Normal (Gaussian)
Chi-Squared
Student's t
Gamma
Beta



End of Chapter 6

The Normal Probability Distribution

Chapter 7

GOALS

- Understand the difference between discrete and continuous distributions.
- Compute the mean and the standard deviation for a uniform distribution.
- Compute probabilities by using the uniform distribution.
- List the characteristics of the normal probability distribution.
- Define and calculate z values.
- Determine the probability an observation is between two points on a normal probability distribution.
- Determine the probability an observation is above (or below) a point on a normal probability distribution.
- Use the normal probability distribution to approximate the binomial distribution.

The Uniform Distribution

The uniform probability distribution is perhaps the simplest distribution for a continuous random variable.

This distribution is rectangular in shape and is defined by minimum and maximum values.

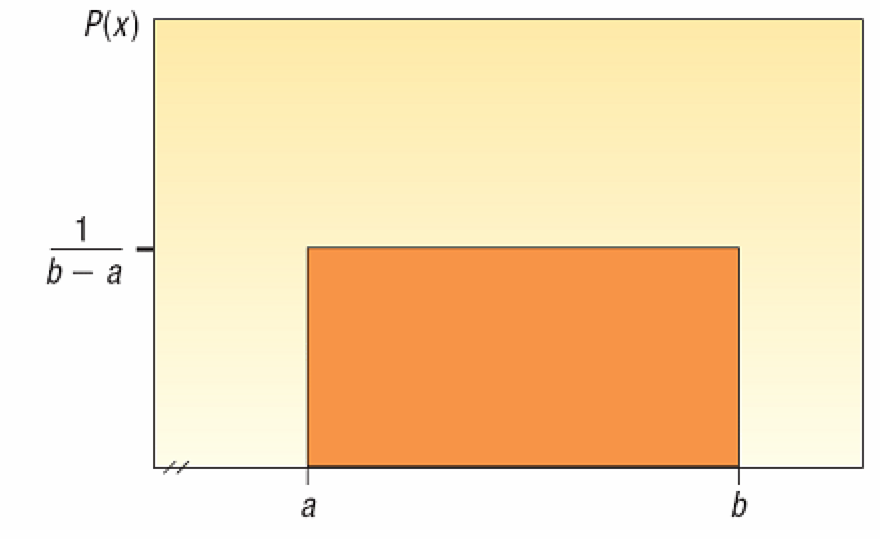


CHART 7–1 A Continuous Uniform Distribution

The Uniform Distribution – Mean and Standard Deviation

MEAN OF THE UNIFORM DISTRIBUTION $$\mu = \frac{a + b}{2} \qquad [7\text{–}1]$$

STANDARD DEVIATION OF THE UNIFORM DISTRIBUTION $$\sigma = \sqrt{\frac{(b - a)^2}{12}} \qquad [7\text{–}2]$$

UNIFORM DISTRIBUTION $$P(x) = \frac{1}{b - a} \quad \text{if } a \le x \le b \text{ and } 0 \text{ elsewhere} \qquad [7\text{–}3]$$

The Uniform Distribution - Example

Southwest Arizona State University provides bus service to students while they are on campus. A bus arrives at the North Main Street and College Drive stop every 30 minutes between 6 A.M. and 11 P.M. during weekdays. Students arrive at the bus stop at random times. The time that a student waits is uniformly distributed from 0 to 30 minutes.

1. Draw a graph of this distribution.

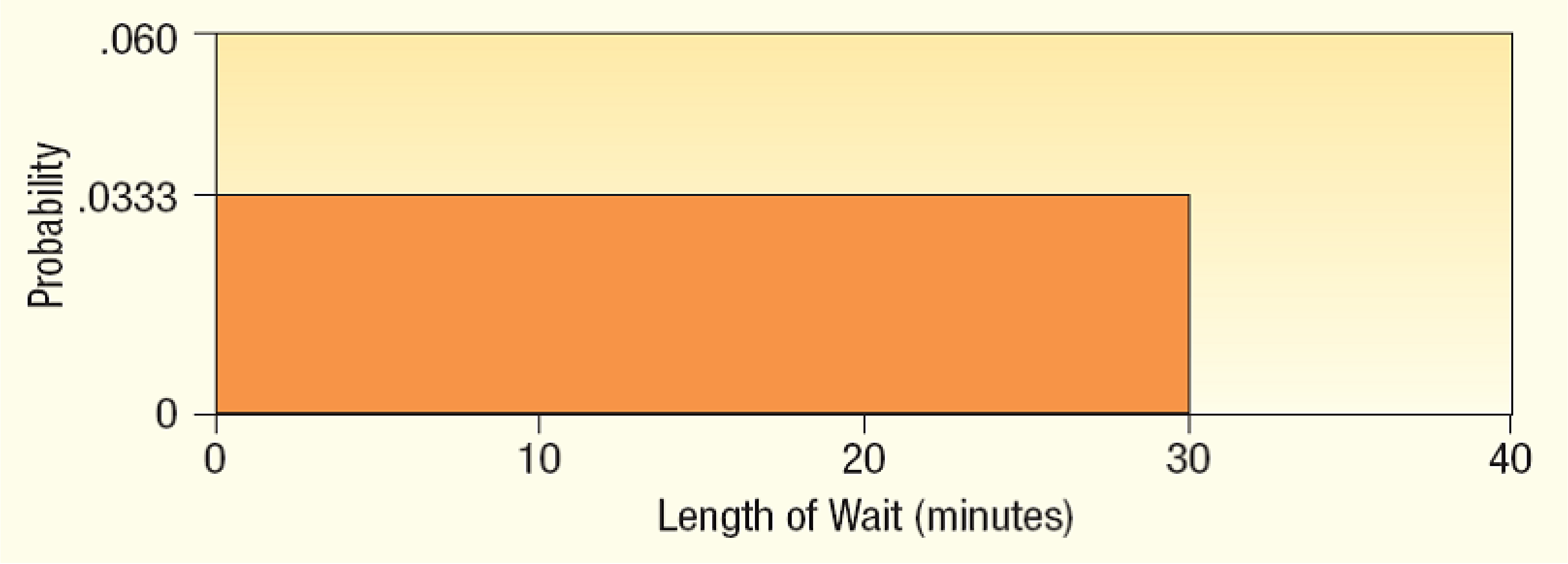
2. How long will a student “typically” have to wait for a bus? In other words what is the mean waiting time? What is the standard deviation of the waiting times?

3. What is the probability a student will wait more than 25 minutes?

4. What is the probability a student will wait between 10 and 20 minutes?

The Uniform Distribution - Example

Draw a graph of this distribution.



The Uniform Distribution - Example

How long will a student "typically" have to wait for a bus? In other words what is the mean waiting time? What is the standard deviation of the waiting times?

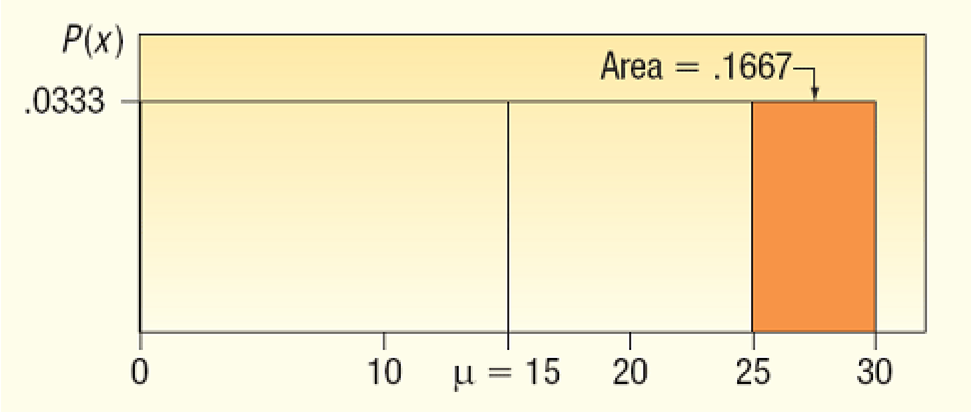
$$\mu = \frac{a+b}{2} = \frac{0+30}{2} = 15$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(30-0)^2}{12}} = 8.66$$

The Uniform Distribution - Example

What is the probability a student will wait more than 25 minutes?

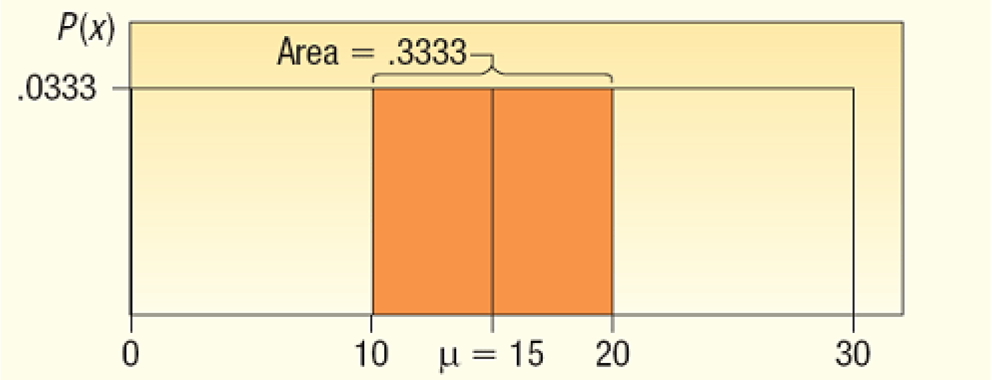
$$P(25 < \text{wait time} < 30) = (\text{height})(\text{base})$$
$$= \frac{1}{(30-0)}(5) = 0.1667$$



The Uniform Distribution - Example

What is the probability a student will wait between 10 and 20 minutes?

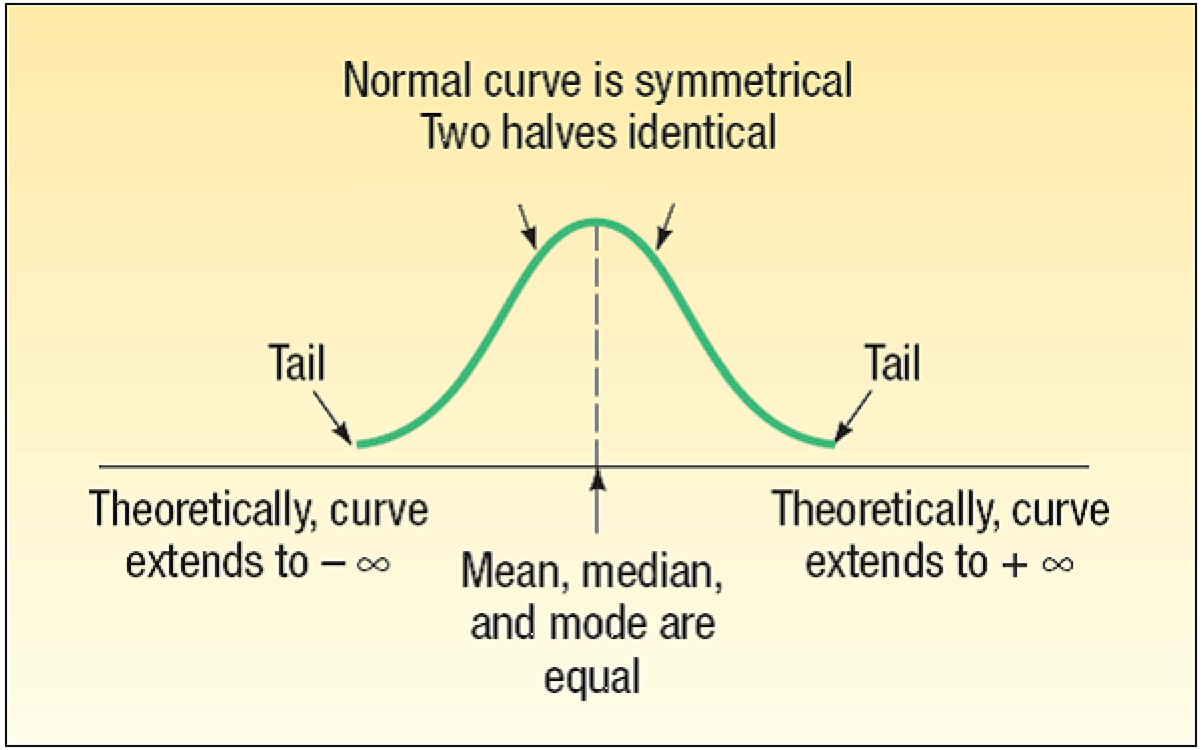
$$P(10 < \text{wait time} < 20) = (\text{height})(\text{base})$$
$$= \frac{1}{(30-0)}(10) = 0.3333$$



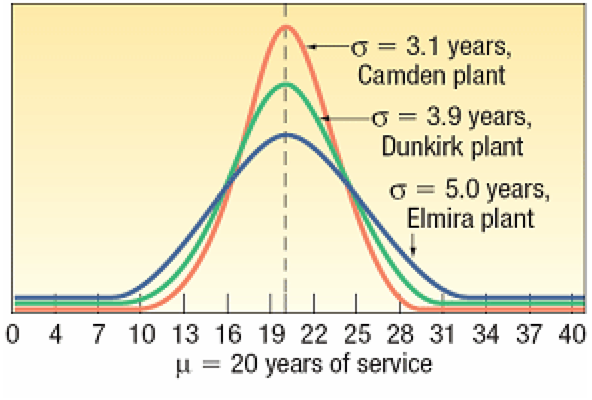
Characteristics of a Normal Probability Distribution

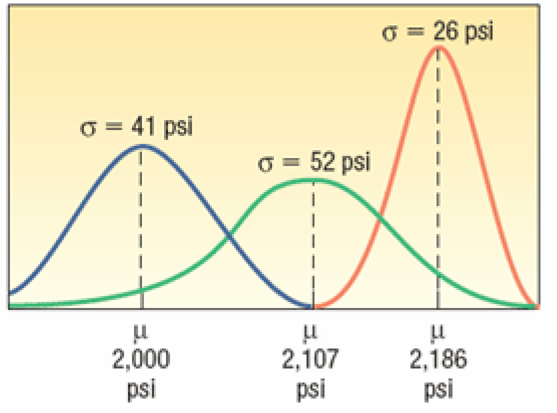
- It is **bell-shaped** and has a single peak at the center of the distribution.
- The arithmetic mean, median, and mode are equal
- The total area under the curve is 1.00; half the area under the normal curve is to the right of this center point and the other half to the left of it.
- It is **symmetrical** about the mean.
- It is **asymptotic:** The curve gets closer and closer to the *X*-axis but never actually touches it. To put it another way, the tails of the curve extend indefinitely in both directions.
- The location of a normal distribution is determined by the mean,μ, the dispersion or spread of the distribution is determined by the standard deviation,σ .

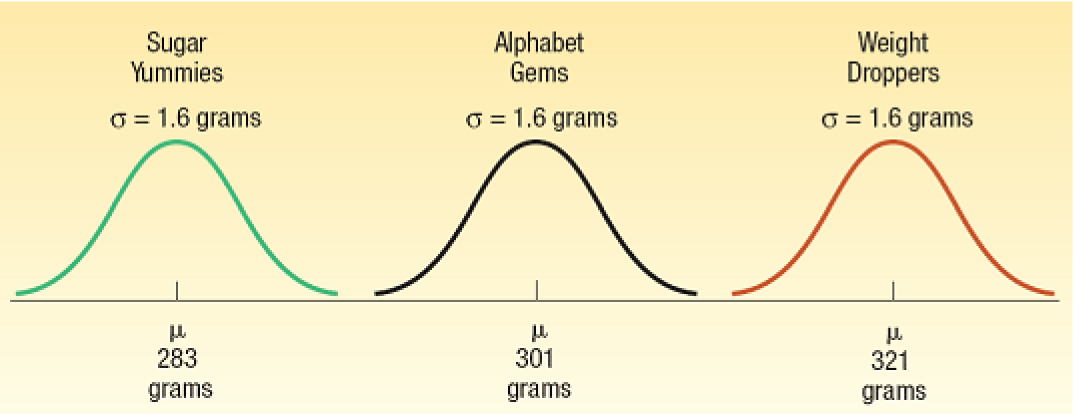
The Normal Distribution - Graphically



The Normal Distribution - Families







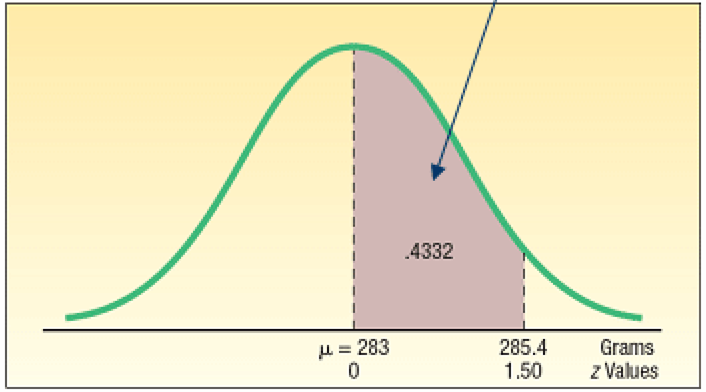
The Standard Normal Probability Distribution

- The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1.
- It is also called the *z* distribution.
- A *z*-value is the distance between a selected value, designated *X*, and the population mean μ, divided by the population standard deviation, σ.
- The formula is:

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

Areas Under the Normal Curve

z	0.00	0.01	0.02	0.03	0.04	0.05	. . .
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	
.							
.							
.							



The Normal Distribution – Example

The weekly incomes of shift foremen in the glass industry follow the normal probability distribution with a mean of \$1,000 and a standard deviation of \$100. What is the *z* value for the income, let's call it *X*, of a foreman who earns \$1,100 per week? For a foreman who earns \$900 per week?

For X = \$1,100:

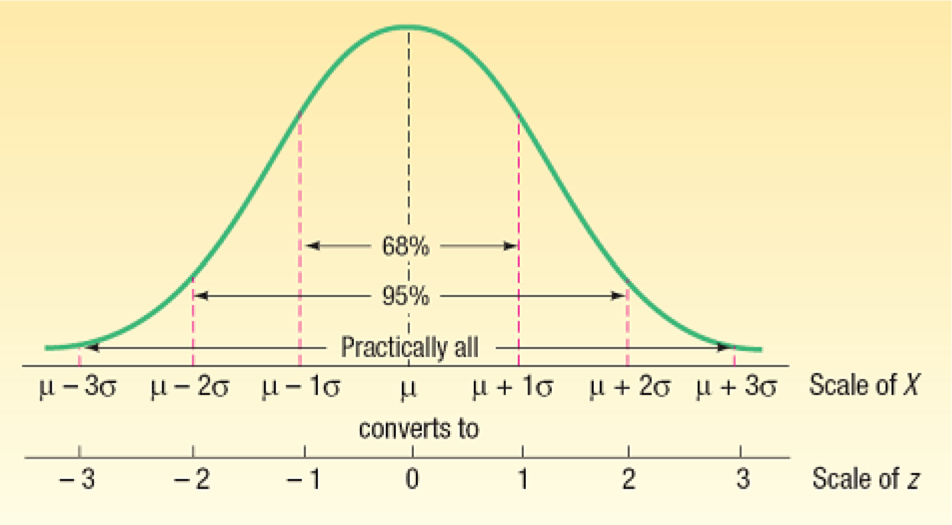
$$z = \frac{X - \mu}{\sigma} = \frac{\$1{,}100 - \$1{,}000}{\$100} = 1.00$$

For X = \$900:

$$z = \frac{X - \mu}{\sigma} = \frac{\$900 - \$1{,}000}{\$100} = -1.00$$

The Empirical Rule

- About 68 percent of the area under the normal curve is within one standard deviation of the mean.
- About 95 percent is within two standard deviations of the mean.
- Practically all is within three standard deviations of the mean.



The Empirical Rule - Example

As part of its quality assurance program, the Autolite Battery Company conducts tests on battery life. For a particular D-cell alkaline battery, the mean life is 19 hours. The useful life of the battery follows a normal distribution with a standard deviation of 1.2 hours.

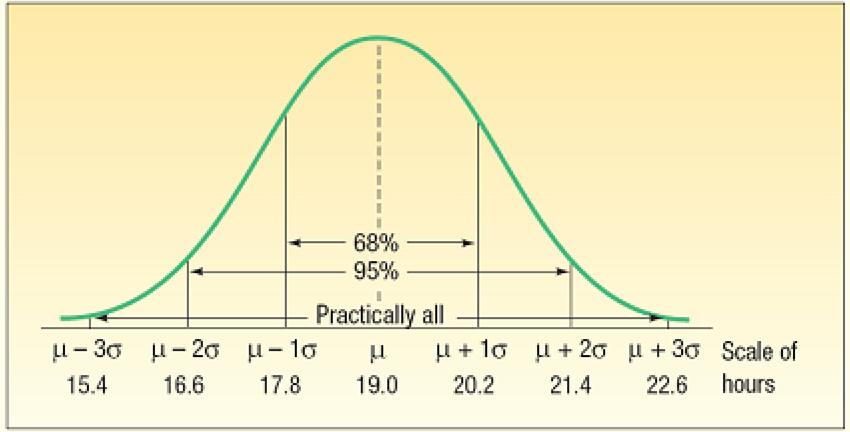
Answer the following questions.

1. About 68 percent of the batteries failed between what two values?
2. About 95 percent of the batteries failed between what two values?
3. Virtually all of the batteries failed between what two values?

We can use the results of the Empirical Rule to answer these questions.

1. About 68 percent of the batteries will fail between 17.8 and 20.2 hours by $19.0 \pm 1(1.2)$ hours.
2. About 95 percent of the batteries will fail between 16.6 and 21.4 hours by $19.0 \pm 2(1.2)$ hours.
3. Virtually all failed between 15.4 and 22.6 hours, found by $19.0 \pm 3(1.2)$

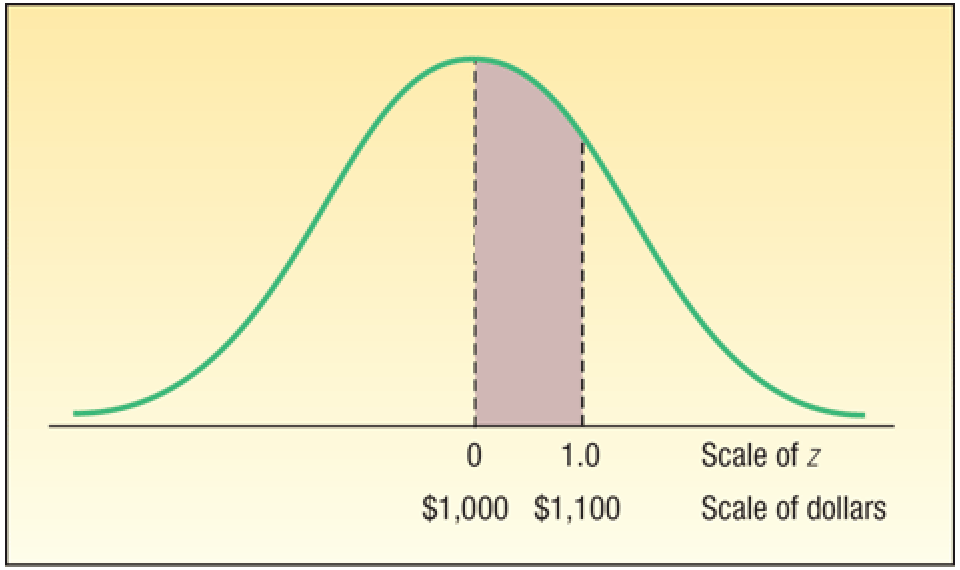
This information is summarized on the following chart.



Normal Distribution – Finding Probabilities

In an earlier example we reported that the mean weekly income of a shift foreman in the glass industry is normally distributed with a mean of $1,000 and a standard deviation of $100.

What is the likelihood of selecting a foreman whose weekly income is between $1,000 and $1,100?



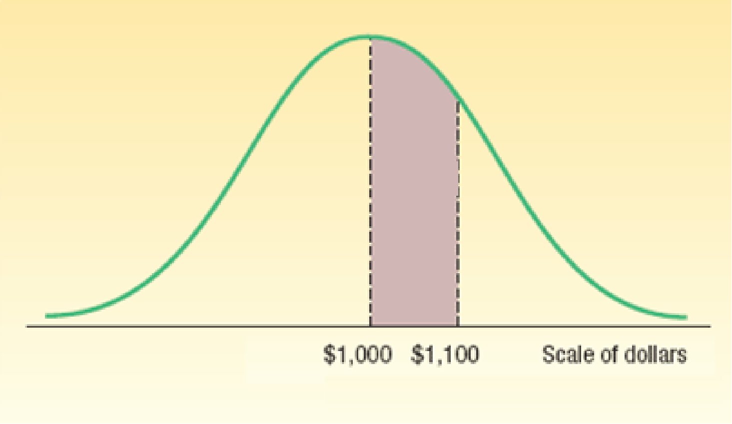
Normal Distribution – Finding Probabilities

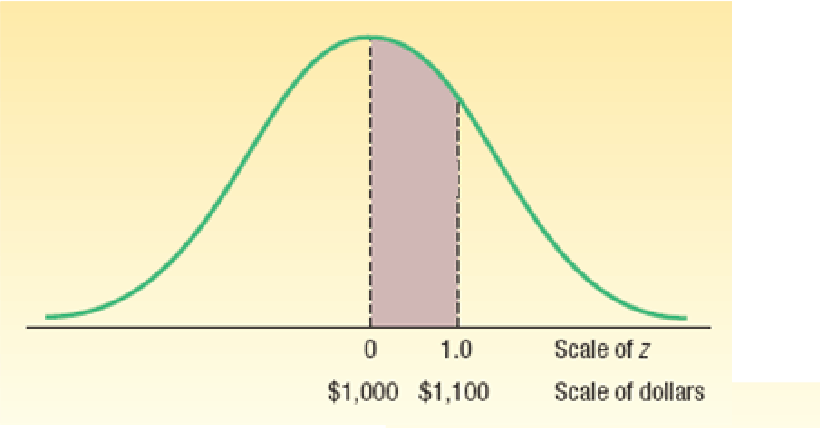
For $X = \$1{,}000$:

$$z = \frac{X - \mu}{\sigma} = \frac{\$1{,}000 - \$1{,}000}{\$100} = 0.00$$

For $X = \$1{,}100$:

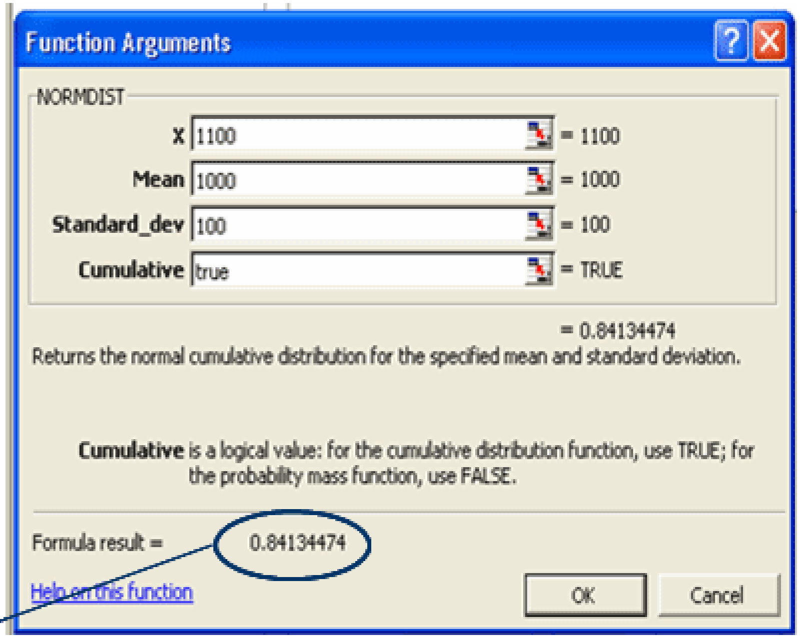
$$z = \frac{X - \mu}{\sigma} = \frac{\$1{,}100 - \$1{,}000}{\$100} = 1.00$$

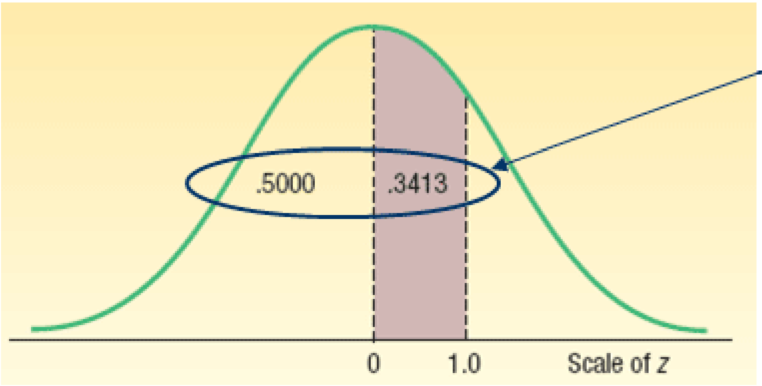




Finding Areas for Z Using Excel

The Excel function
=NORMDIST(x,Mean,Standard_dev,Cumu)
=NORMDIST(1100,1000,100,true)
generates area (probability) from
Z=1 and below





Normal Distribution – Finding Probabilities (Example 2)

Refer to the information regarding the weekly income of shift foremen in the glass industry. The distribution of weekly incomes follows the normal probability distribution with a mean of $1,000 and a standard deviation of $100.

What is the probability of selecting a shift foreman in the glass industry whose income is:

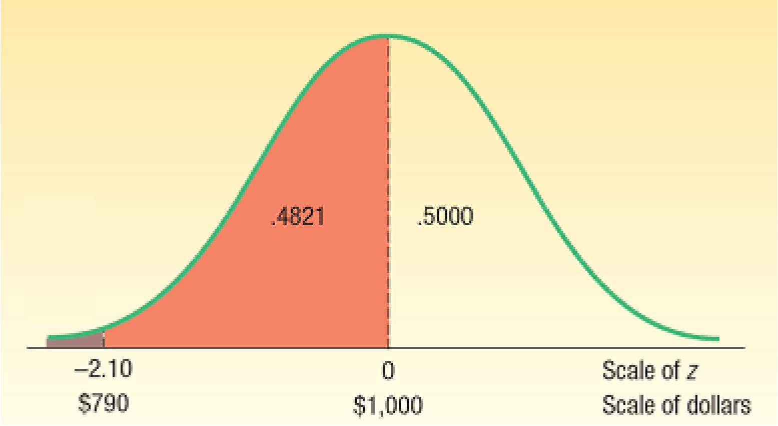
Between $790 and $1,000?

For $X = \$790$:

$$z = \frac{X - \mu}{\sigma} = \frac{\$790 - \$1{,}000}{\$100} = -2.10$$

For $X = \$1{,}000$:

$$z = \frac{X - \mu}{\sigma} = \frac{\$1{,}000 - \$1{,}000}{\$100} = 0.00$$



Normal Distribution – Finding Probabilities (Example 3)

Refer to the information regarding the weekly income of shift foremen in the glass industry. The distribution of weekly incomes follows the normal probability distribution with a mean of $1,000 and a standard deviation of $100.

What is the probability of selecting a shift foreman in the glass industry whose income is:

Less than $790?

Find Z for X = $790 :

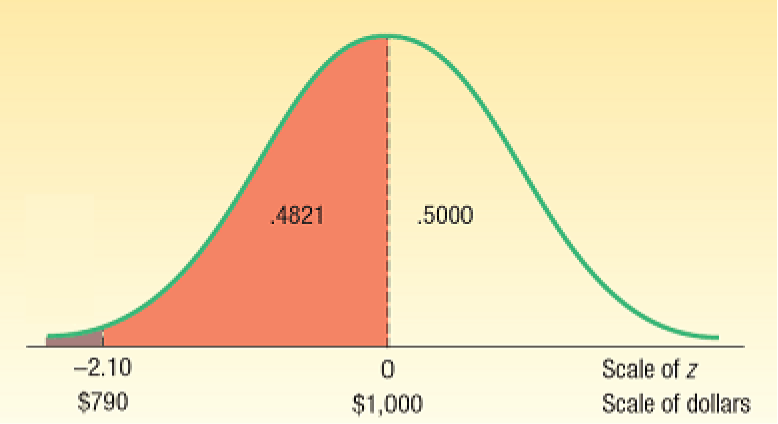
$$z = \frac{X - \mu}{\sigma} = \frac{\$790 - \$1,000}{\$100} = -2.10$$

To find the are below - 2.10,

subtract from 0.50 the area from - 2.10 to 0

= 0.50-0.4821

= 0.0179



Normal Distribution – Finding Probabilities (Example 4)

Refer to the information regarding the weekly income of shift foremen in the glass industry. The distribution of weekly incomes follows the normal probability distribution with a mean of \$1,000 and a standard deviation of \$100.

What is the probability of selecting a shift foreman in the glass industry whose income is:

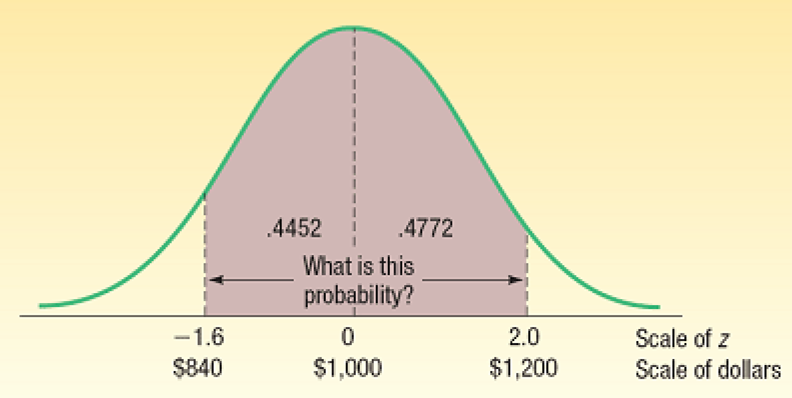
Between \$840 and \$1,200?

For $X = \$840$:

$$z = \frac{X-\mu}{\sigma} = \frac{\$840 - \$1{,}000}{\$100} = -1.60$$

For $X = \$1{,}200$:

$$z = \frac{X-\mu}{\sigma} = \frac{\$1{,}200 - \$1{,}000}{\$100} = 2.00$$



Normal Distribution – Finding Probabilities (Example 5)

Refer to the information regarding the weekly income of shift foremen in the glass industry. The distribution of weekly incomes follows the normal probability distribution with a mean of $1,000 and a standard deviation of $100.

What is the probability of selecting a shift foreman in the glass industry whose income is:

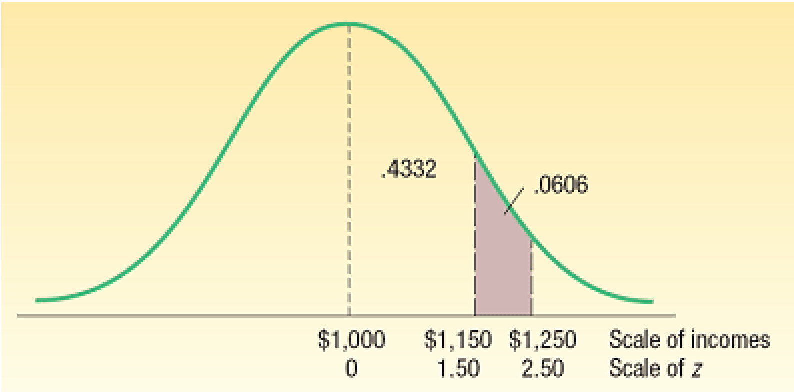
Between $1,150 and $1,250

For X = $1,150:

$$z = \frac{X - \mu}{\sigma} = \frac{\$1{,}150 - \$1{,}000}{\$100} = 1.50$$

For X = $1,250:

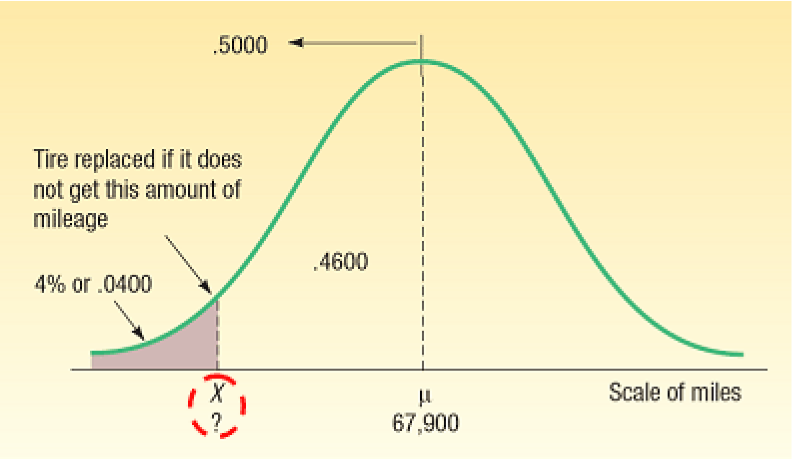
$$z = \frac{X - \mu}{\sigma} = \frac{\$1{,}250 - \$1{,}000}{\$100} = 2.50$$



Using Z in Finding X Given Area - Example

Layton Tire and Rubber Company wishes to set a minimum mileage guarantee on its new MX100 tire. Tests reveal the mean mileage is 67,900 with a standard deviation of 2,050 miles and that the distribution of miles follows the normal probability distribution. It wants to set the minimum guaranteed mileage so that no more than 4 percent of the tires will have to be replaced. What minimum guaranteed mileage should Layton announce?

Using Z in Finding X Given Area - Example



Solve X using the formula :

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

$$z = \frac{X - 67{,}900}{2{,}050}$$

The value of z is found using the 4% information

The area between 67,900 and X is .4600, found by .5000 - .0400.

Using Appendix D, the area closest to .4600 is .4599, which gives a z value of 1.75.

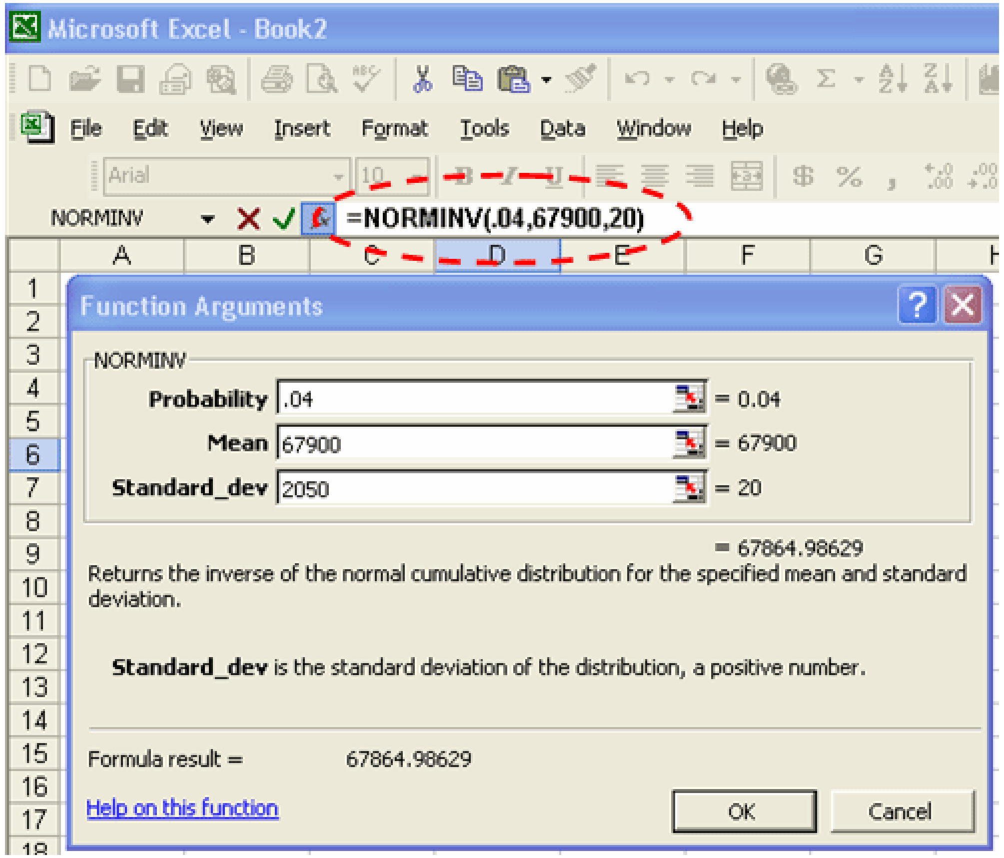
$$1.75 = \frac{X - 67{,}900}{2{,}050} \text{ then solving for X}$$

$$1.75(2{,}050) = X - 67{,}900$$

$$X = 67{,}900 - 1.75(2{,}050)$$

$$X = 64{,}312$$

Using Z in Finding X Given Area - Excel

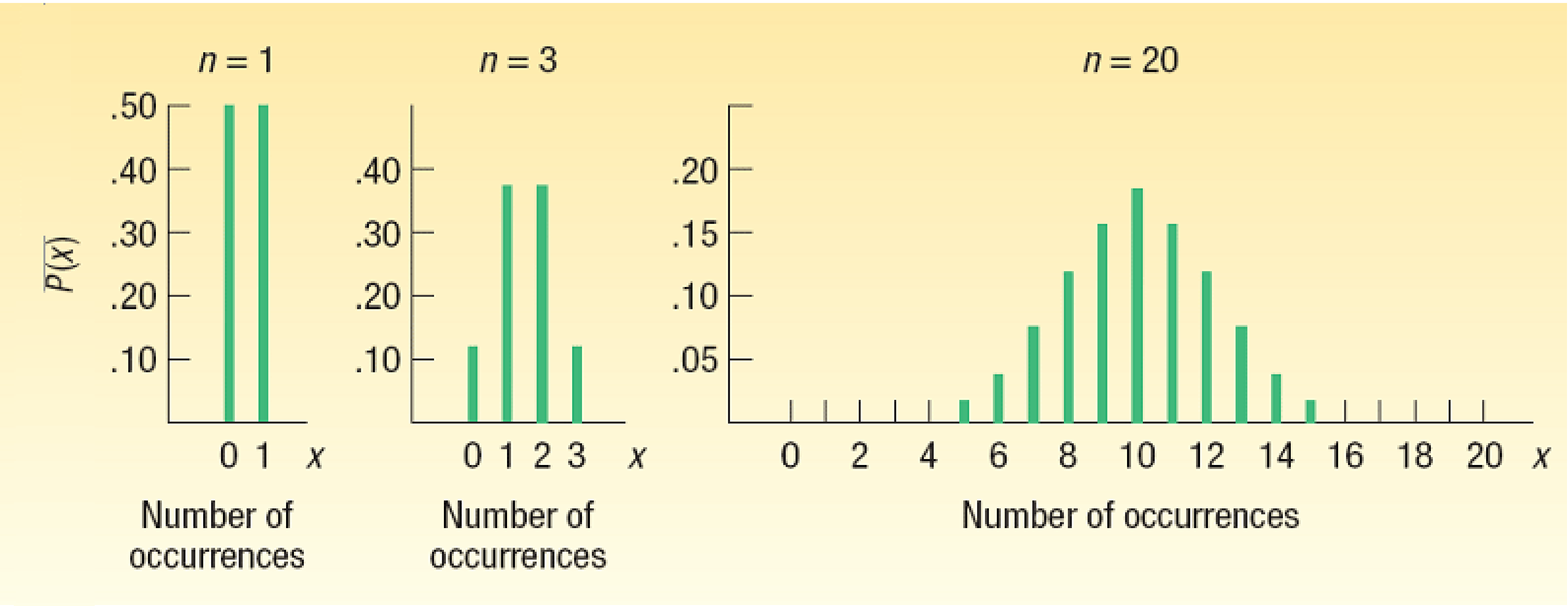


Normal Approximation to the Binomial

- The normal distribution (a continuous distribution) yields a good approximation of the binomial distribution (a discrete distribution) for large values of *n.*
- The normal probability distribution is generally a good approximation to the binomial probability distribution when $n\pi$ and $n(1-\pi)$ are both greater than 5.

Normal Approximation to the Binomial

Using the normal distribution (a continuous distribution) as a substitute for a binomial distribution (a discrete distribution) for large values of *n* seems reasonable because, as *n* increases, a binomial distribution gets closer and closer to a normal distribution.



Continuity Correction Factor

The value .5 subtracted or added, depending on the problem, to a selected value when a binomial probability distribution (a discrete probability distribution) is being approximated by a continuous probability distribution (the normal distribution).

How to Apply the Correction Factor

Only four cases may arise. These cases are:

1. For the probability *at least X* occurs, use the area *above* (X -.5).
2. For the probability that *more than X* occurs, use the area *above* (X+.5).
3. For the probability that *X or fewer* occurs, use the area *below* (X +.5).
4. For the probability that *fewer than X* occurs, use the area *below* (X-.5).

Normal Approximation to the Binomial - Example

Suppose the management of the Santoni Pizza Restaurant found that 70 percent of its new customers return for another meal. For a week in which 80 new (first-time) customers dined at Santoni's, what is the probability that 60 or more will return for another meal?

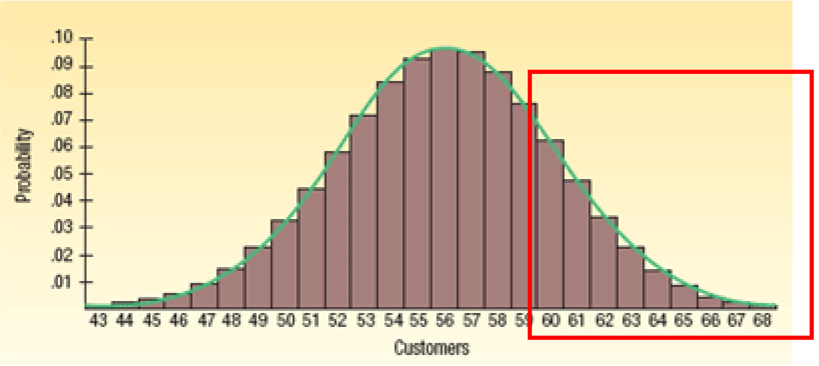
Normal Approximation to the Binomial - Example

$$P(x) = {}_nC_x\,(\pi)^x\,(1-\pi)^{n-x}$$

$$P(x=60) = {}_{80}C_{60}\,(.70)^{60}\,(1-.70)^{20} = .063$$

$$P(x=61) = {}_{80}C_{61}\,(.70)^{61}\,(1-.70)^{19} = .048$$

Number Returning	Probability	Number Returning	Probability
43	.001	56	.097
44	.002	57	.095
45	.003	58	.088
46	.006	59	.077
47	.009	60	.063
48	.015	61	.048
49	.023	62	.034
50	.033	63	.023
51	.045	64	.014
52	.059	65	.008
53	.072	66	.004
54	.084	67	.002
55	.093	68	.001



P(X ≥ 60) = 0.063+0.048+ … + 0.001) = 0.197

Normal Approximation to the Binomial - Example

Step 1. Find the mean and the variance of a binomial distribution and find the *z* corresponding to an *X* of 59.5 (x-.5, the correction factor)

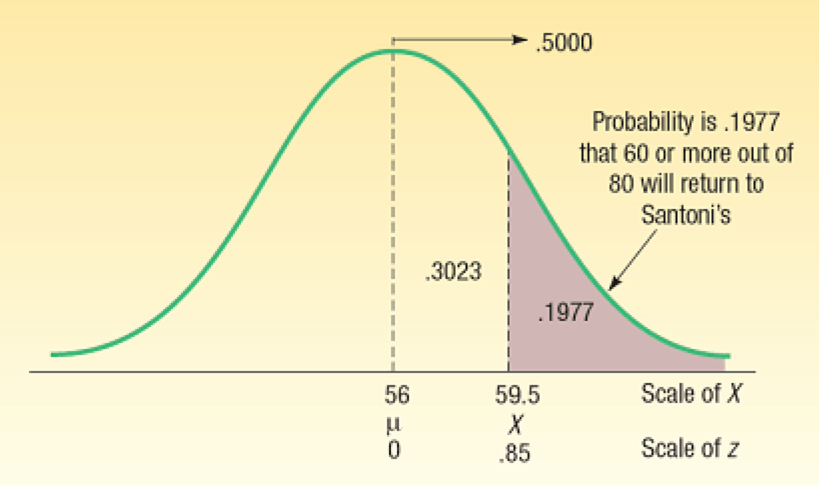
Step 2: Determine the area from 59.5 and beyond

$$\mu = n\pi = 80(.70) = 56$$

$$\sigma^2 = n\pi(1-\pi) = 80(.70)(1-.70) = 16.8$$

$$\sigma = \sqrt{16.8} = 4.10$$

$$z = \frac{X-\mu}{\sigma} = \frac{59.5-56}{4.10} = 0.85$$



End of Chapter 7

Sampling Methods and the Central Limit Theorem

Chapter 8

GOALS

- Explain why a sample is the only feasible way to learn about a population.
- Describe methods to select a sample.
- Define and construct a sampling distribution of the sample mean.
- Explain the central limit theorem.
- Use the Central Limit Theorem to find probabilities of selecting possible sample means from a specified population.

Why Sample the Population?

- The physical impossibility of checking all items in the population.
- The cost of studying all the items in a population.
- The sample results are usually adequate.
- Contacting the whole population would often be time-consuming.
- The destructive nature of certain tests.

Probability Sampling

- A **probability sample** is a sample selected such that each item or person in the population being studied has a known likelihood of being included in the sample.

Methods of Probability Sampling

- Simple Random Sample: A sample formulated so that each item or person in the population has the same chance of being included.

- Systematic Random Sampling: The items or individuals of the population are arranged in some order. A random starting point is selected and then every *k*th member of the population is selected for the sample.

Methods of Probability Sampling

- Stratified Random Sampling: A population is first divided into subgroups, called strata, and a sample is selected from each stratum.

- Cluster Sampling: A population is first divided into primary units then samples are selected from the primary units.

Methods of Probability Sampling

- In nonprobability sample inclusion in the sample is based on the judgment of the person selecting the sample.

- The sampling error is the difference between a sample statistic and its corresponding population parameter.

Sampling Distribution of the Sample Means

- The sampling distribution of the sample mean is a probability distribution consisting of all possible sample means of a given sample size selected from a population.

Sampling Distribution of the Sample Means - Example

Tartus Industries has seven production employees (considered the population). The hourly earnings of each employee are given in the table below.

TABLE 8–2 Hourly Earnings of the Production Employees of Tartus Industries

Employee	Hourly Earnings	Employee	Hourly Earnings
Joe	$7	Jan	$7
Sam	7	Art	8
Sue	8	Ted	9
Bob	8		

1. What is the population mean?
2. What is the sampling distribution of the sample mean for samples of size 2?
3. What is the mean of the sampling distribution?
4. What observations can be made about the population and the sampling distribution?

Sampling Distribution of the Sample Means - Example

1. The population mean is \$7.71, found by:

$$\mu = \frac{\Sigma X}{N} = \frac{\$7 + \$7 + \$8 + \$8 + \$7 + \$8 + \$9}{7} = \$7.71$$

2. To arrive at the sampling distribution of the sample mean, we need to select all possible samples of 2 without replacement from the population, then compute the mean of each sample. There are 21 possible samples, found by using formula (5–10) on page 173.

$$_{N}C_{n} = \frac{N!}{n!(N-n)!} = \frac{7!}{2!(7-2)!} = 21$$

Sampling Distribution of the Sample Means - Example

TABLE 8–3 Sample Means for All Possible Samples of 2 Employees

Sample	Employees	Hourly Earnings	Sum	Mean	Sample	Employees	Hourly Earnings	Sum	Mean
1	Joe, Sam	\$7, \$7	\$14	\$7.00	12	Sue, Bob	\$8, \$8	\$16	\$8.00
2	Joe, Sue	7, 8	15	7.50	13	Sue, Jan	8, 7	15	7.50
3	Joe, Bob	7, 8	15	7.50	14	Sue, Art	8, 8	16	8.00
4	Joe, Jan	7, 7	14	7.00	15	Sue, Ted	8, 9	17	8.50
5	Joe, Art	7, 8	15	7.50	16	Bob, Jan	8, 7	15	7.50
6	Joe, Ted	7, 9	16	8.00	17	Bob, Art	8, 8	16	8.00
7	Sam, Sue	7, 8	15	7.50	18	Bob, Ted	8, 9	17	8.50
8	Sam, Bob	7, 8	15	7.50	19	Jan, Art	7, 8	15	7.50
9	Sam, Jan	7, 7	14	7.00	20	Jan, Ted	7, 9	16	8.00
10	Sam, Art	7, 8	15	7.50	21	Art, Ted	8, 9	17	8.50
11	Sam, Ted	7, 9	16	8.00					

Sampling Distribution of the Sample Means - Example

$$\mu_{\bar{X}} = \frac{\text{Sum of all sample means}}{\text{Total number of samples}} = \frac{\$7.00 + \$7.50 + \cdots + \$8.50}{21}$$

$$= \frac{\$162}{21} = \$7.71$$

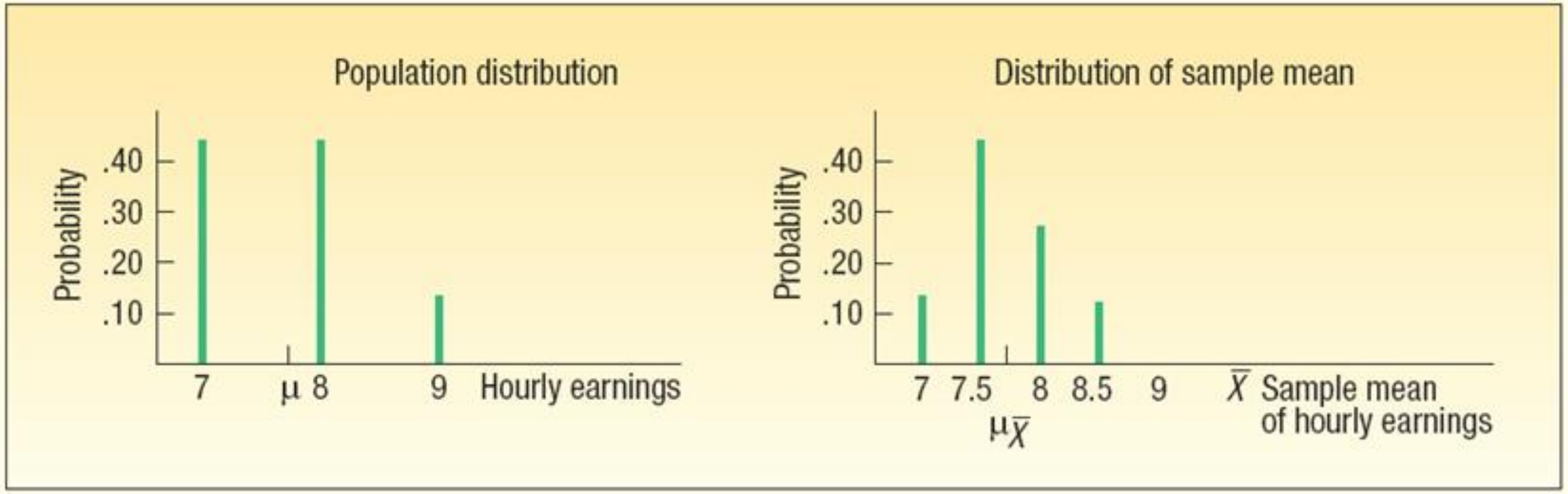


CHART 8–1 Distributions of Population Values and Sample Mean

Central Limit Theorem

- For a population with a mean μ and a variance σ^2 the sampling distribution of the means of all possible samples of size *n* generated from the population will be approximately normally distributed.
- The mean of the sampling distribution equal to μ and the variance equal to σ^2/n.

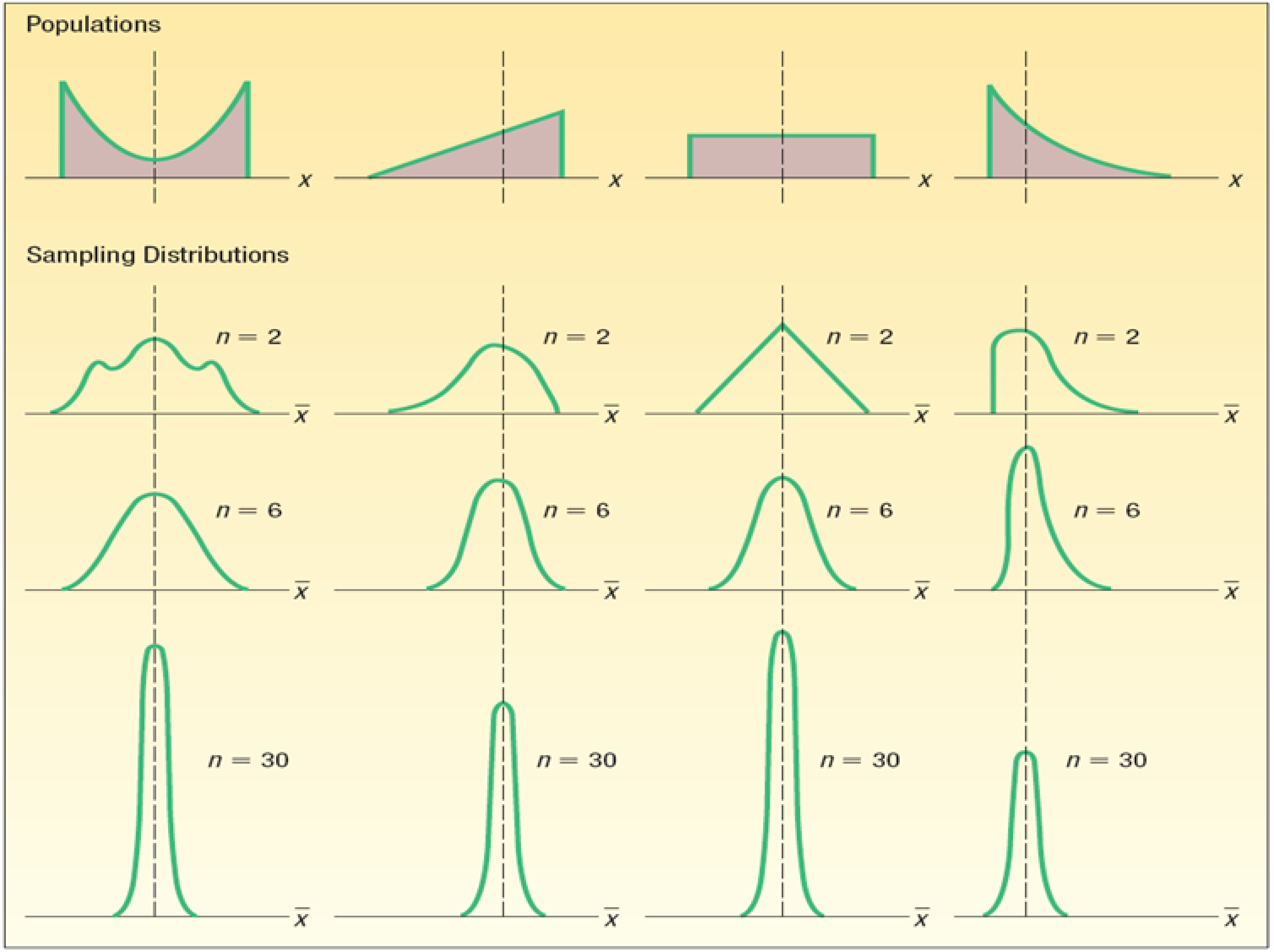


CHART 8–2 Results of the Central Limit Theorem for Several Populations

Using the Sampling Distribution of the Sample Mean (Sigma Known)

- If a population follows the normal distribution, the sampling distribution of the sample mean will also follow the normal distribution.
- To determine the probability a sample mean falls within a particular region, use:

$$z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Using the Sampling Distribution of the Sample Mean (Sigma Unknown)

- If the population does not follow the normal distribution, but the sample is of at least 30 observations, the sample means will follow the normal distribution.
- To determine the probability a sample mean falls within a particular region, use:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Using the Sampling Distribution of the Sample Mean (Sigma Known) - Example

The Quality Assurance Department for Cola, Inc., maintains records regarding the amount of cola in its Jumbo bottle. The actual amount of cola in each bottle is critical, but varies a small amount from one bottle to the next. Cola, Inc., does not wish to underfill the bottles. On the other hand, it cannot overfill each bottle. Its records indicate that the amount of cola follows the normal probability distribution. The mean amount per bottle is 31.2 ounces and the population standard deviation is 0.4 ounces. At 8 A.M. today the quality technician randomly selected 16 bottles from the filling line. The mean amount of cola contained in the bottles is 31.38 ounces.

Is this an unlikely result? Is it likely the process is putting too much soda in the bottles? To put it another way, is the sampling error of 0.18 ounces unusual?

Using the Sampling Distribution of the Sample Mean (Sigma Known) - Example

Step 1: Find the *z*-values corresponding to the sample mean of 31.38

$$z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{31.38-32.20}{\$0.2/\sqrt{16}}=1.80$$

Using the Sampling Distribution of the Sample Mean (Sigma Known) - Example

Step 2: Find the probability of observing a Z equal to or greater than 1.80

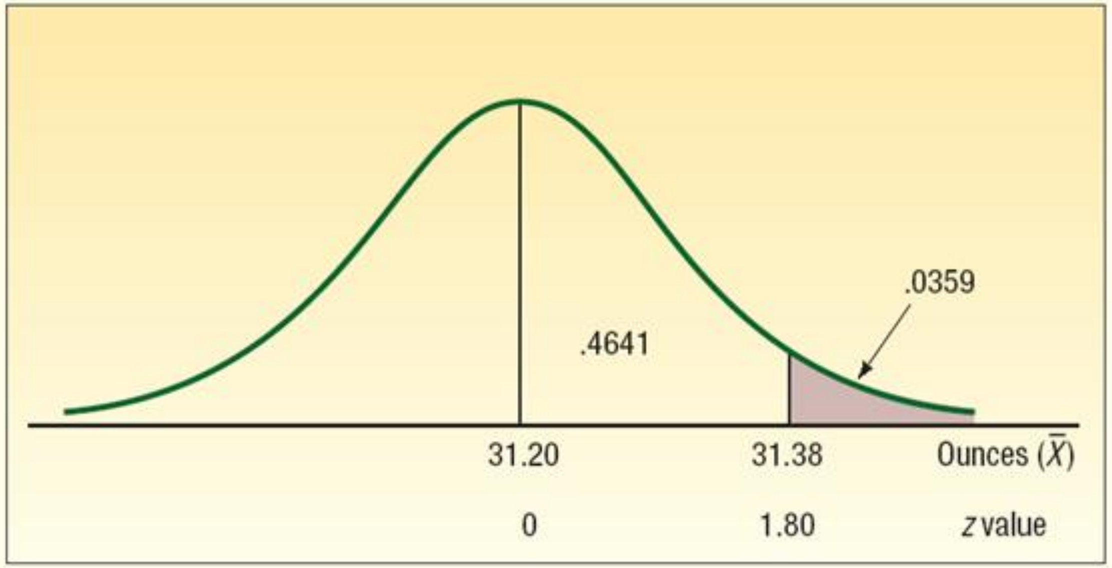


CHART 8–6 Sampling Distribution of the Mean Amount of Cola in a Jumbo Bottle

Using the Sampling Distribution of the Sample Mean (Sigma Known) - Example

What do we conclude?

It is unlikely, less than a 4 percent chance, we could select a sample of 16 observations from a normal population with a mean of 31.2 ounces and a population standard deviation of 0.4 ounces and find the sample mean equal to or greater than 31.38 ounces.

We conclude the process is putting too much cola in the bottles.

End of Chapter 8