Dr Kofi A. Ababio

BSc MSc MPhil PhD

kaababs@gmail.com

+233 24 4980077

Probability Distributions

Week 1

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?

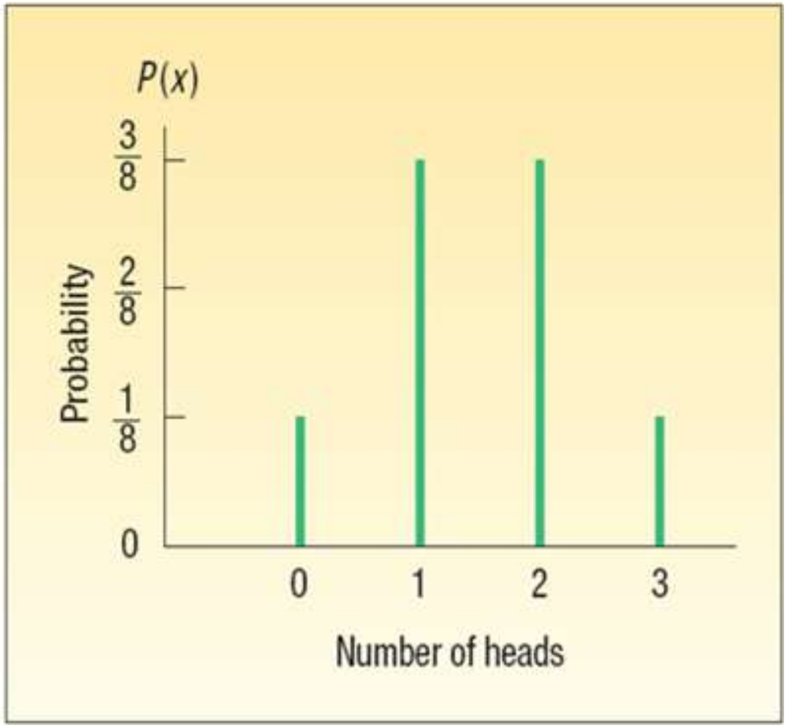
PROBABILITY DISTRIBUTION A listing of all the outcomes of an experiment and the probability associated with each outcome.

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			Number of Heads
	First	Second	Third	
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

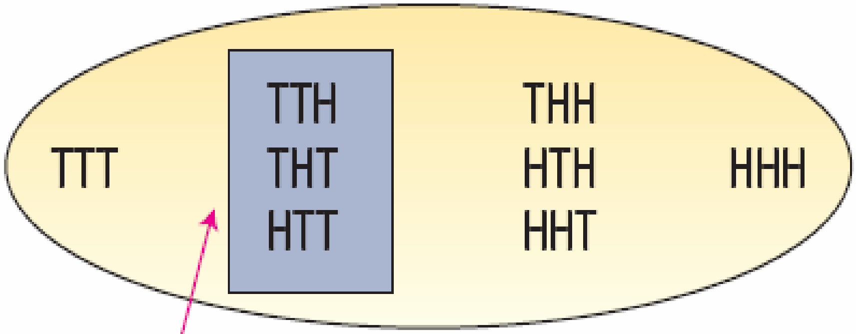
CHARACTERISTICS OF A PROBABILITY DISTRIBUTION

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 **counting** something

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

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 $$\mu = \Sigma[xP(x)] \qquad [6\text{–}1]$$

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

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

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	$\mu =$ 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**?

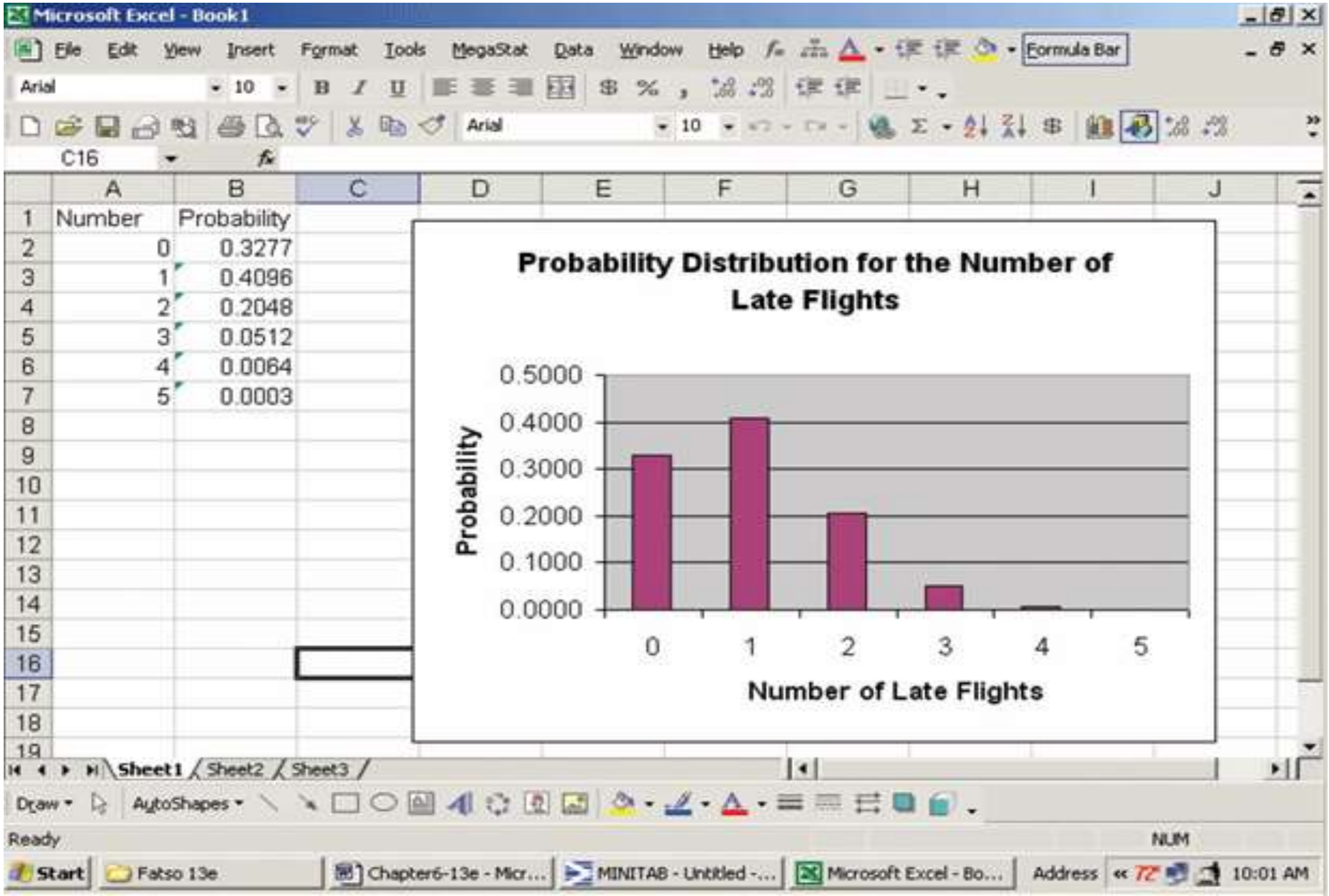
$$P(x=0) = {}_nC_x \pi^x (1-\pi)^{n-x}$$

$$= {}_5C_0 (.20)^0 (1-.20)^{5-0}$$

$$= (1)(1)(.3277)$$

$$= 0.3277$$

Binomial Probability - Excel



Number	Probability
0	0.3277
1	0.4096
2	0.2048
3	0.0512
4	0.0064
5	0.0003

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

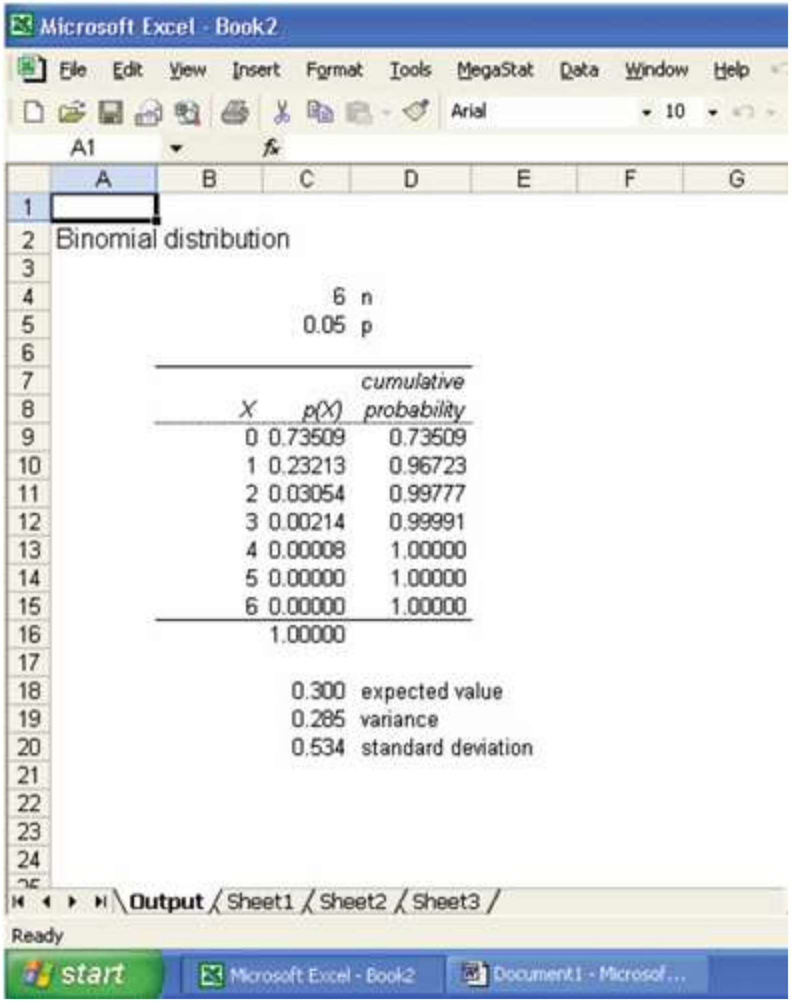
- **Five percent** of the worm gears produced by an automatic, high-speed Carter-Bell milling machine are **defective**. What is the **probability that out of six gears selected at random none will be defective? Exactly one? Exactly two? Exactly three? Exactly four? Exactly five? Exactly six out of six?**

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	.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

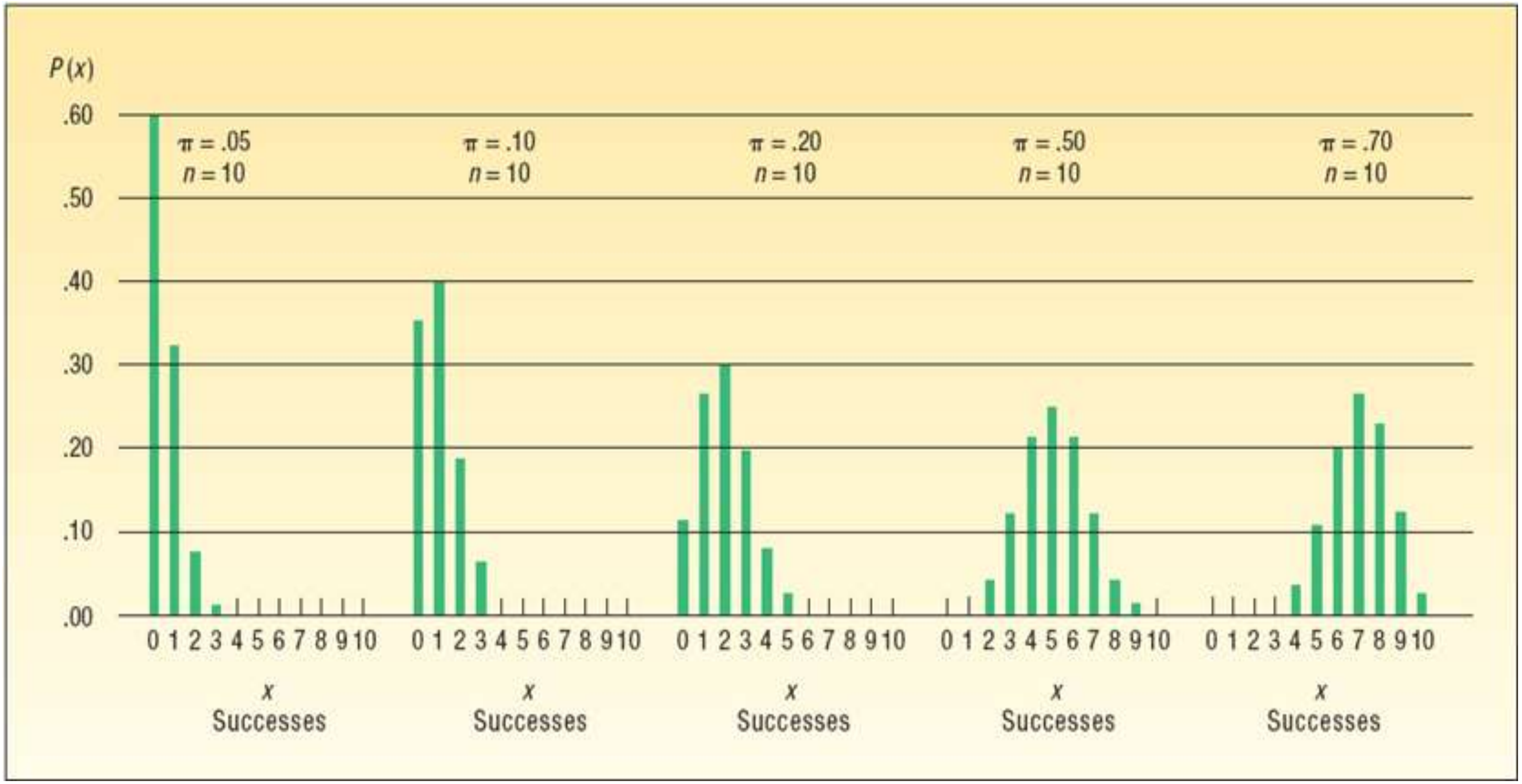
Binomial Distribution - MegaStat

- 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?

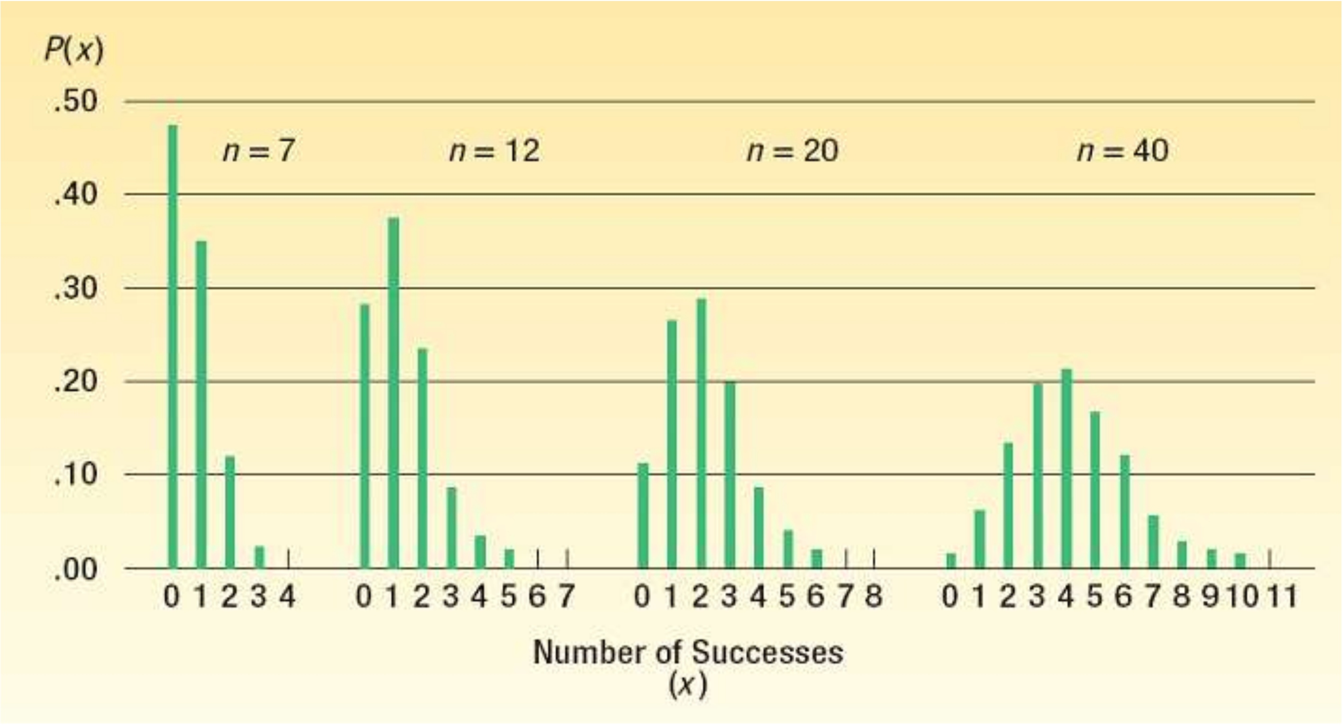


Binomial – Shapes for Varying π (n constant)

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)



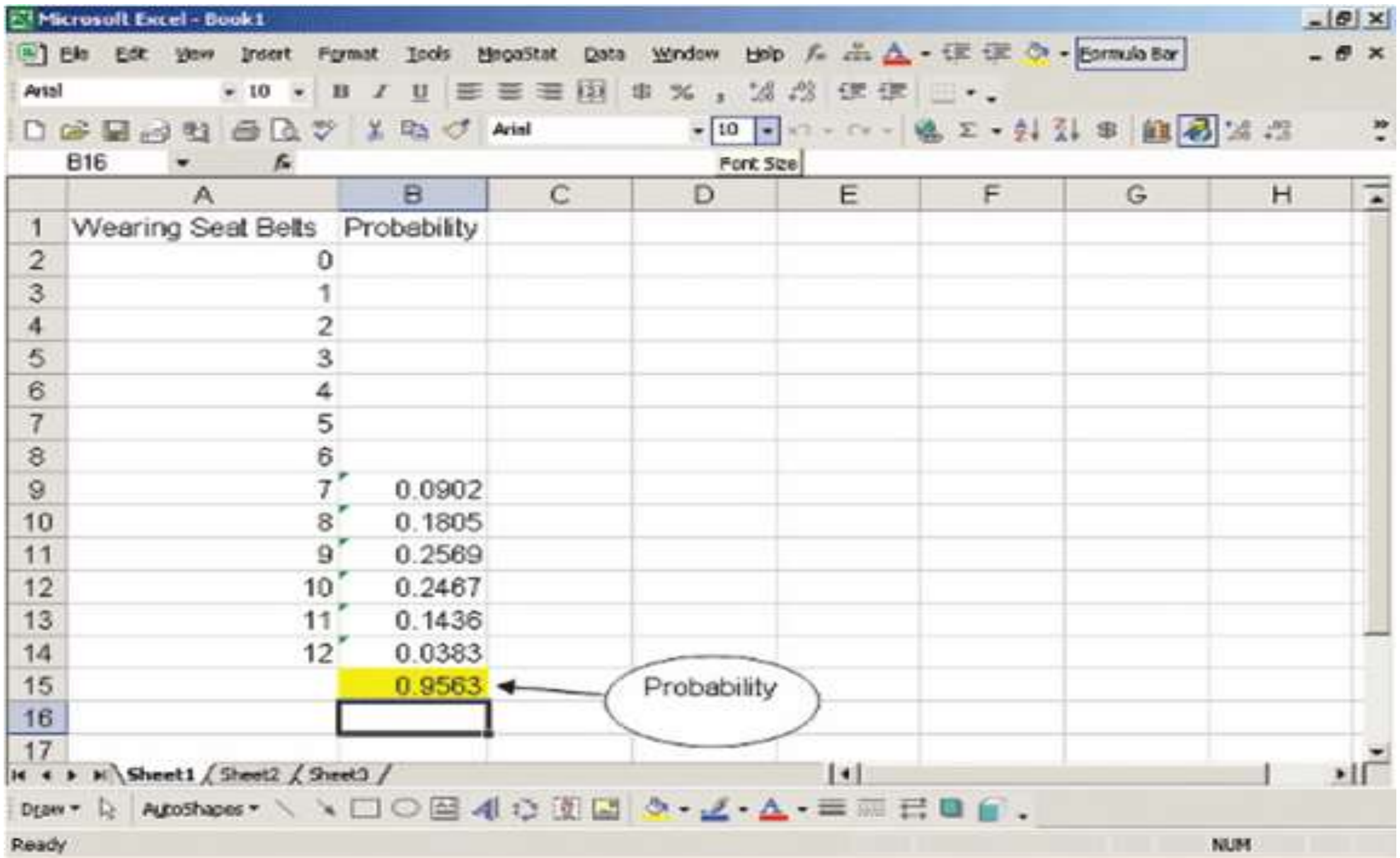
Cumulative Binomial Probability Distributions

- 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



	A	B
1	Wearing Seat Belts	Probability
2	0	
3	1	
4	2	
5	3	
6	4	
7	5	
8	6	
9	7	0.0902
10	8	0.1805
11	9	0.2569
12	10	0.2467
13	11	0.1436
14	12	0.0383
15		0.9563
16		
17		

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 **Ayigya**

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

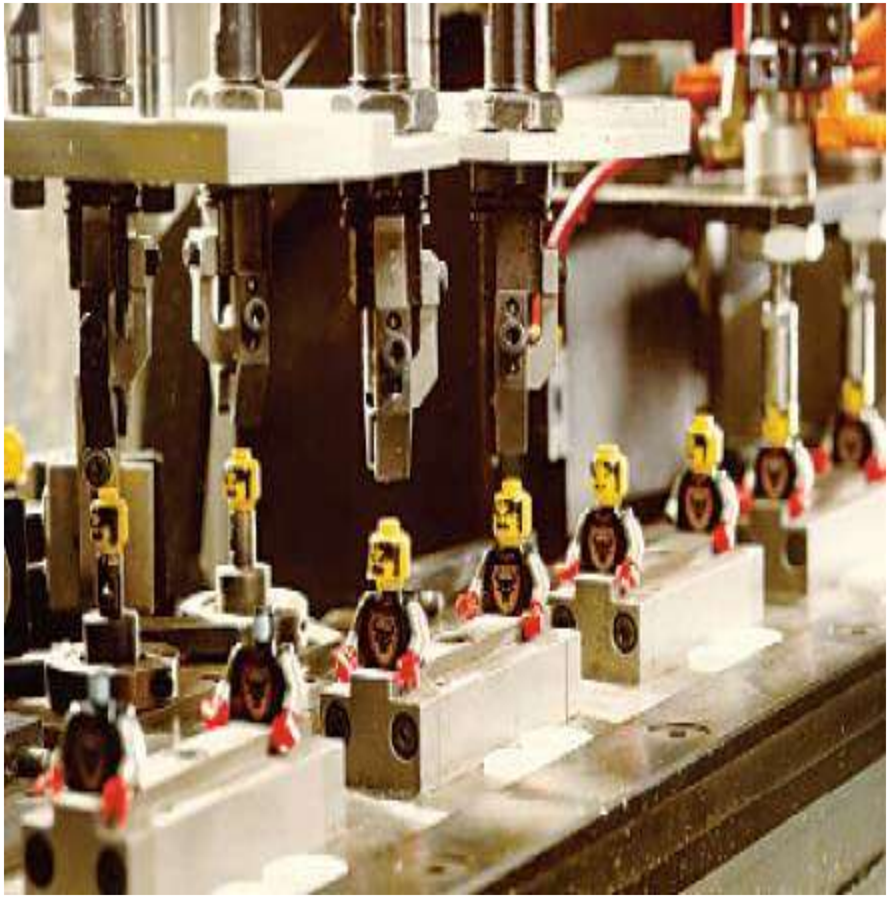
HYPERGEOMETRIC DISTRIBUTION $$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n} \qquad [6\text{–}6]$$

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**?

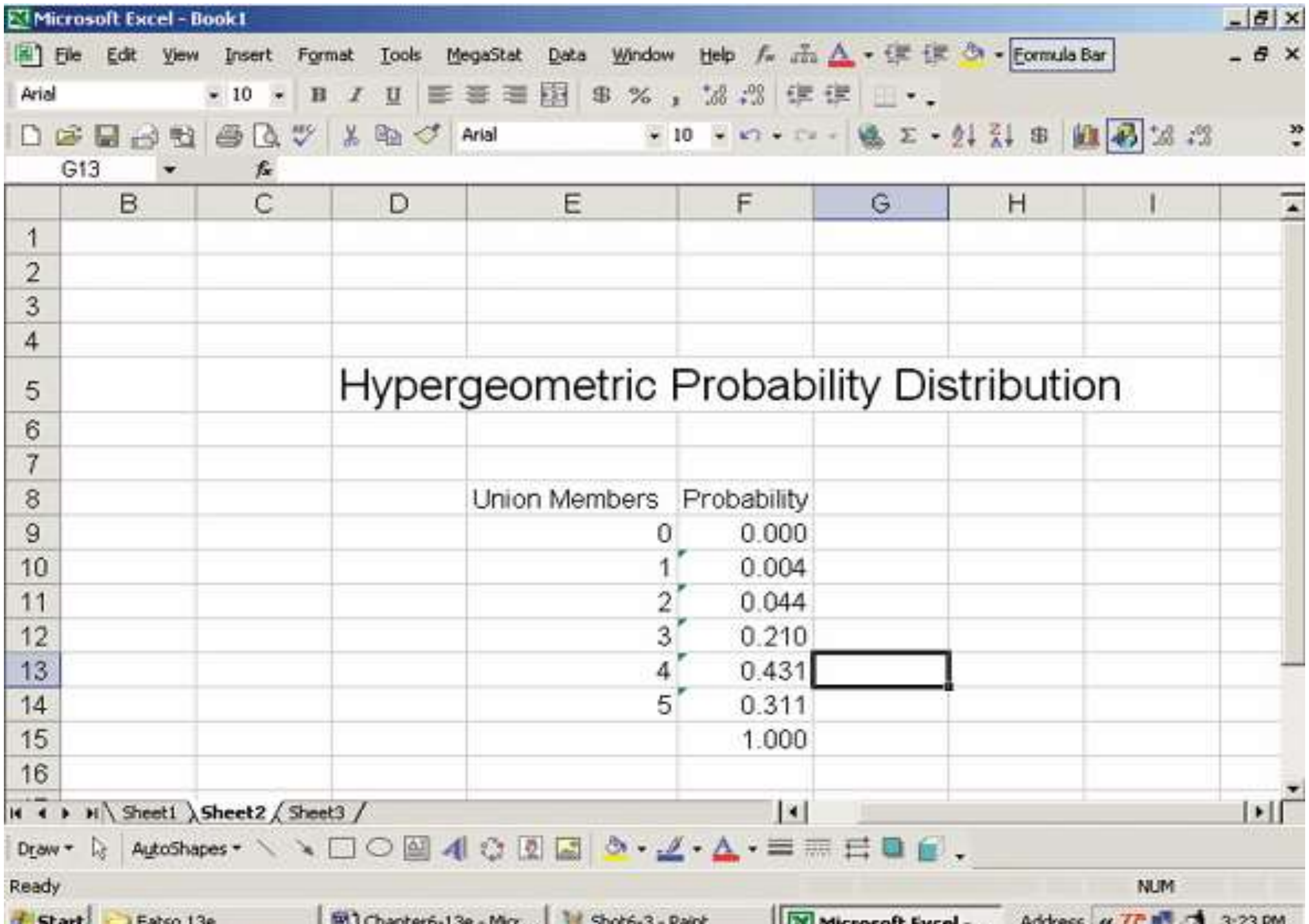


Hypergeometric Distribution - Example

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

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

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 $\boldsymbol{n\pi}$, where ***n*** is the **number of trials** and π the **probability of a success**.
- The **variance** of the **Poisson distribution** is also equal to $\boldsymbol{n}\,\pi$.

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

The Multinomial Distribution

The binomial experiment becomes a multinomial experiment, if we let each trial has more than two possible outcome.

Definition 4.4: **Multinomial distribution**

If a given trial can result in the ***k*** outcomes $E_1, E_2, \ldots\ldots, E_k$ with probabilities $p_1, p_2, \ldots\ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots\ldots, X_k$ representing the number of occurrences for $E_1, E_2, \ldots\ldots, E_k$ in *n* independent trials is

$$f(x_1, x_2, \ldots\ldots, x_k) = \binom{n}{x_1, x_2, \ldots\ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots\ldots p_k^{x_k}$$

where $\binom{n}{x_1, x_2, \ldots\ldots, x_k} = \frac{n!}{x_1! x_2! \ldots\ldots x_k!}$

$$\sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1$$

Multinomial example

If you are randomly choosing 8 people from an audience that contains 50% democrats, 30% republicans, and 20% green party, what's the probability of choosing exactly 4 democrats, 3 republicans, and 1 green party member?

$$P(D=4, R=3, G=1) = \frac{8!}{4!\,3!\,1!}(.5)^4(.3)^3(.2)^1$$

Geometric Distribution

A **geometric distribution** is a discrete probability distribution of a random variable x that satisfies the following conditions.

1. A trial is repeated until a success occurs.
2. The repeated trials are independent of each other.
3. The probability of a success p is constant for each trial.

The **probability that the first success will occur on trial *x*** is

$$P(x) = p(q)^{x-1}, \text{ where } q = 1 - p.$$

Geometric Distribution

Example:

A fast food chain puts a winning game piece on every fifth package of French fries. Find the probability that you will win a prize,

a.) with your third purchase of French fries,

b.) with your third or fourth purchase of French fries.

$p = 0.20$ $\quad q = 0.80$

a.) $x = 3$

$$P(3) = (0.2)(0.8)^{3-1}$$
$$= (0.2)(0.8)^2$$
$$= (0.2)(0.64)$$
$$= 0.128$$

b.) $x = 3, 4$

$$P(3 \text{ or } 4) = P(3) + P(4)$$
$$\approx 0.128 + 0.102$$
$$\approx 0.230$$

End of Week 1