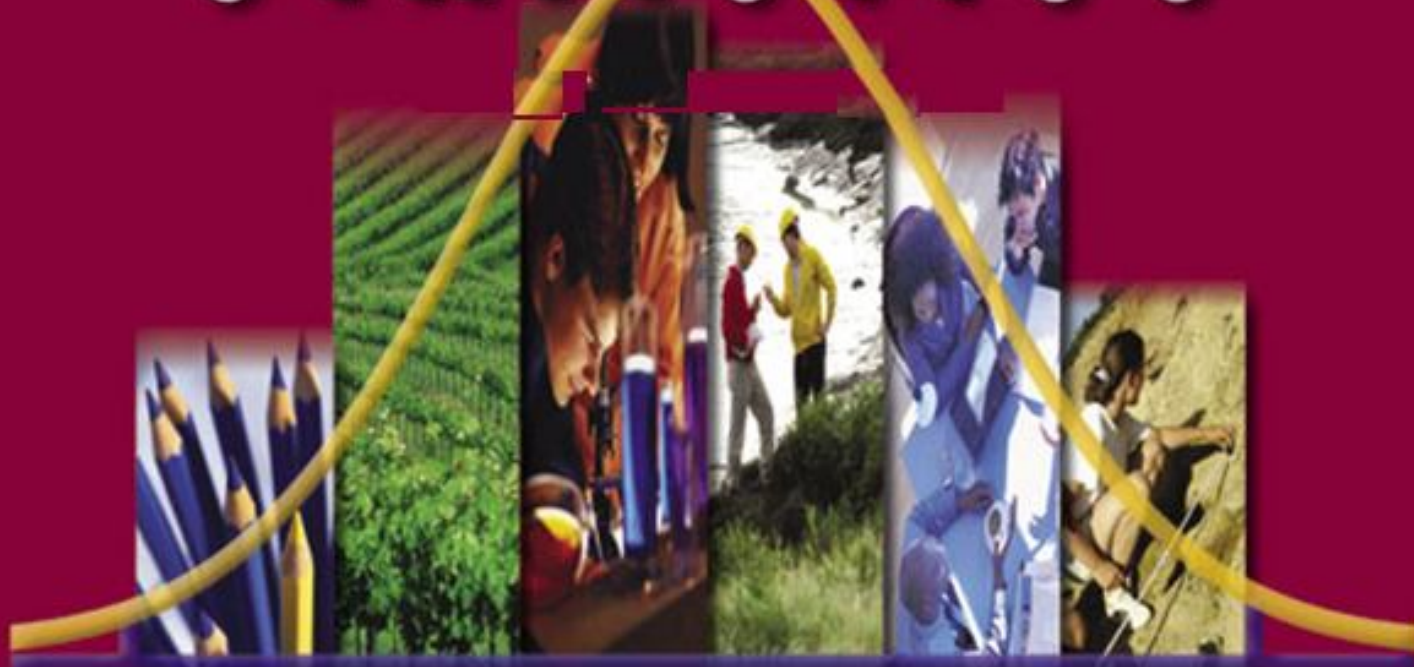


ELEMENTARY

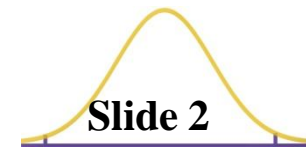
STATISTICS



Dadan Dasari

Chapter 4

Probability Distributions



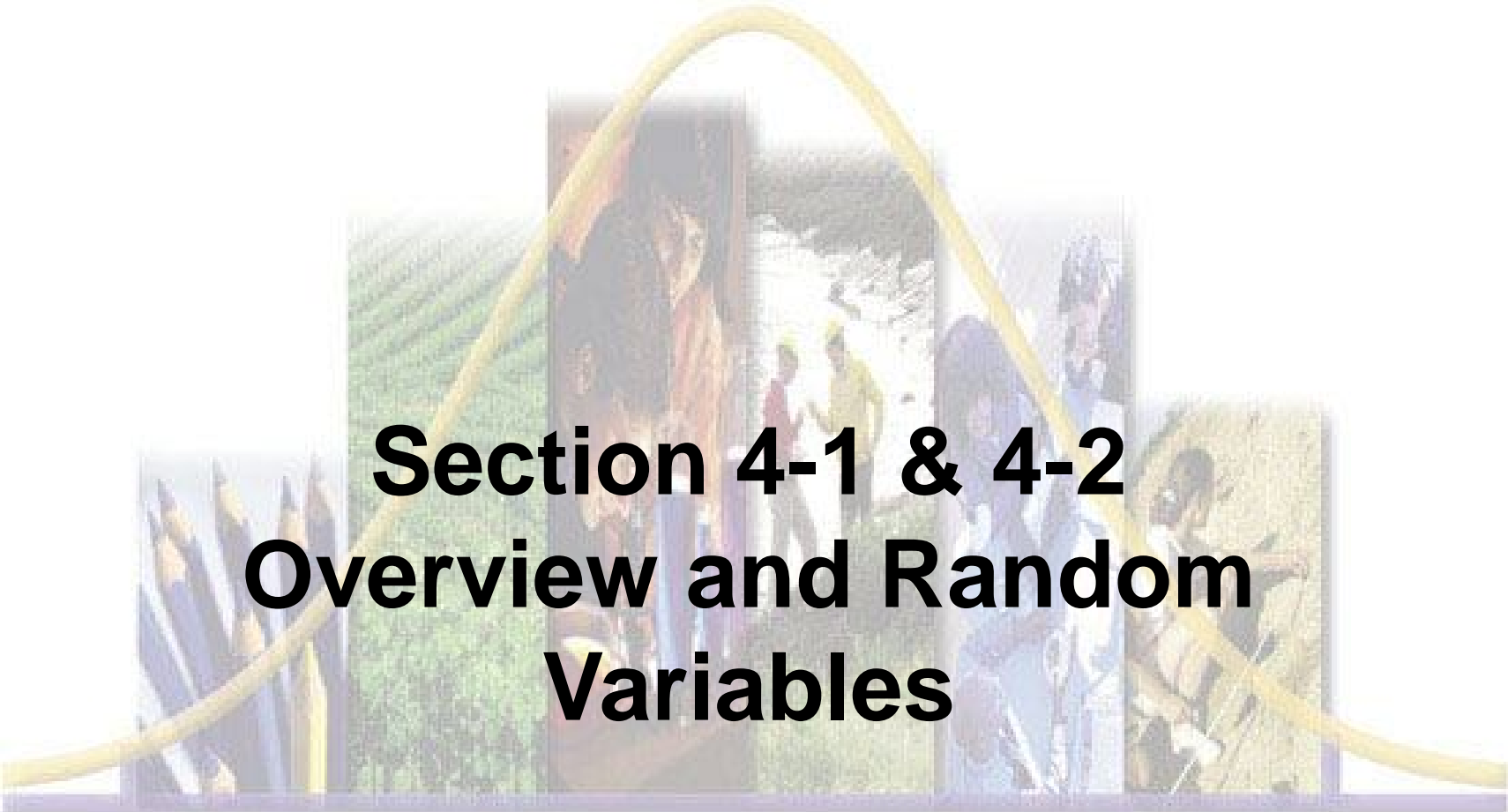
4-1 Overview

4-2 Random Variables

4-3 Binomial Probability Distributions

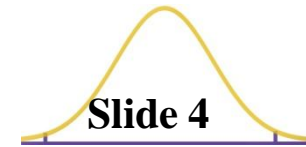
4-4 Mean, Variance, and Standard
Deviation for the Binomial Distribution

4-5 The Poisson Distribution



**Section 4-1 & 4-2
Overview and Random
Variables**

Overview



This chapter will deal with the construction of

probability distributions

by combining the methods of *descriptive statistics* presented in Chapter 2 and those of *probability* presented in Chapter 3.

Probability Distributions will describe what will *probably* happen instead of what actually *did* happen.

Combining Descriptive Methods and Probabilities

In this chapter we will construct probability distributions by presenting possible outcomes along with the relative frequencies we *expect*.

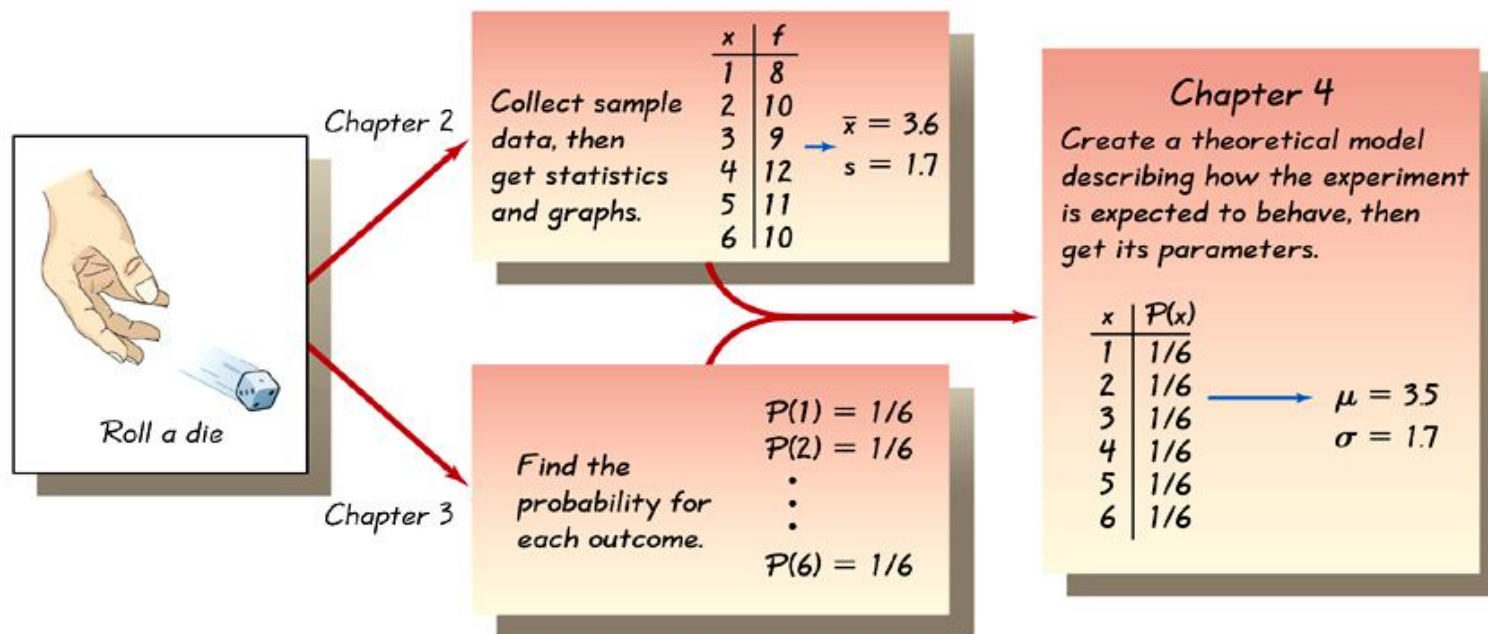
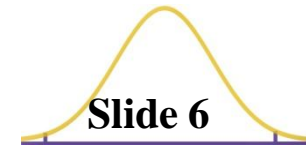


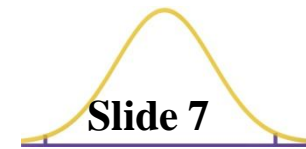
Figure 4-1

Definitions



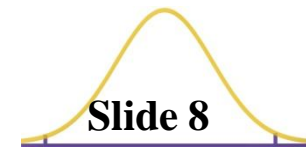
- ❖ A **random variable** is a variable (typically represented by **X**) that has a single numerical value, determined by chance, for each outcome of a procedure.
- ❖ A **probability distribution** is a graph, table, or formula that gives the probability for each value of the random variable.

Definitions



- ❖ A **discrete random variable** has either a finite number of values or countable number of values, where “countable” refers to the fact that there might be infinitely many values, but they result from a counting process.
- ❖ A **continuous random variable** has infinitely many values, and those values can be associated with measurements on a continuous scale in such a way that there are no gaps or interruptions.

Graphs



The **probability histogram** is very similar to a relative frequency histogram, but the vertical scale shows probabilities.

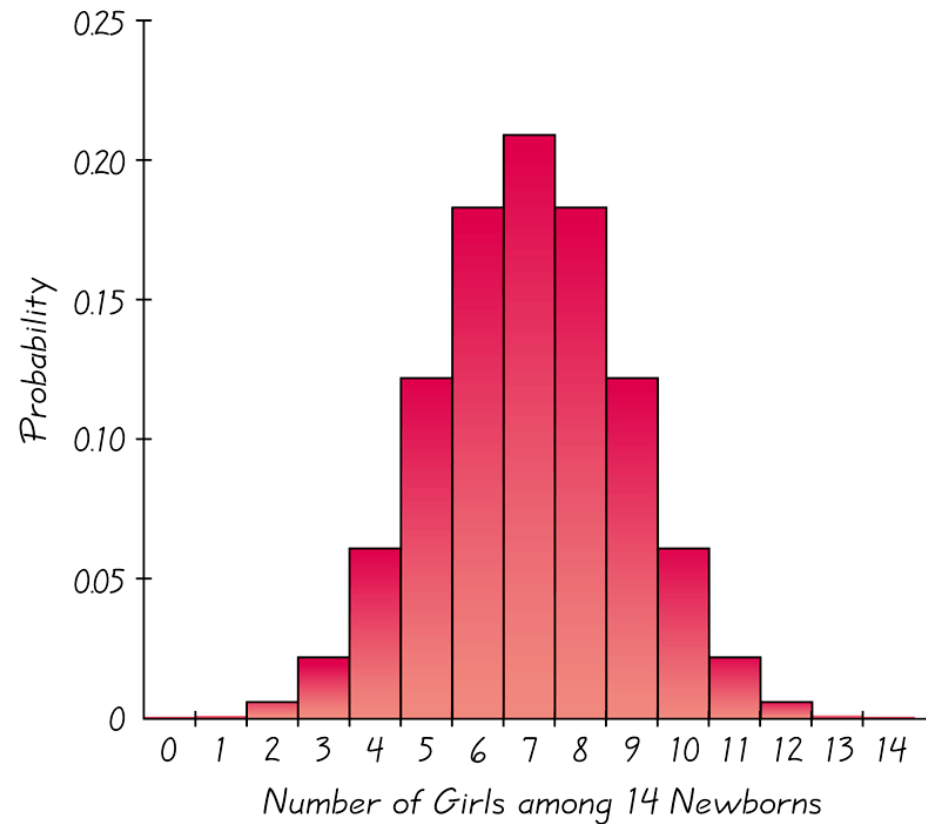
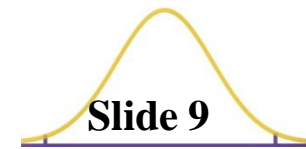


Figure 4-3

Requirements for Probability Distribution



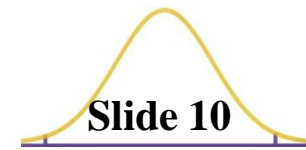
$$\sum P(x) = 1$$

where X assumes all possible values

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

for every individual value of X

Mean, Variance and Standard Deviation of a Probability Distribution



$$\mu = \sum [x \cdot P(x)]$$

Mean

$$\sigma^2 = \sum [(x - \mu)^2 \cdot P(x)]$$

Variance

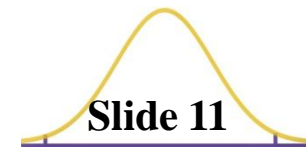
$$\sigma^2 = [\sum x^2 \cdot P(x)] - \mu^2$$

Variance (shortcut)

$$\sigma = \sqrt{\sum [x^2 \cdot P(x)] - \mu^2}$$

Standard Deviation

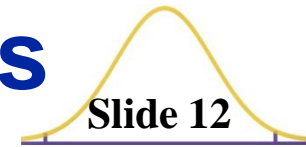
Roundoff Rule for μ , σ , and σ^2



Round results by carrying one more decimal place than the number of decimal places used for the random variable x . If the values of x are integers, round μ , σ , and σ^2 and to one decimal place.

Identifying Unusual Results

Range Rule of Thumb



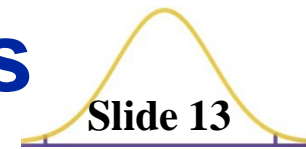
According to the **range rule of thumb**, most values should lie within 2 standard deviations of the mean.

We can therefore identify “unusual” values by determining if they lie outside these limits:

$$\text{Maximum usual value} = \mu + 2\sigma$$

$$\text{Minimum usual value} = \mu - 2\sigma$$

Identifying Unusual Results Probabilities

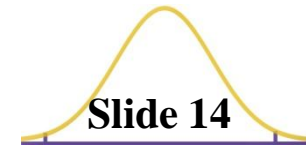


Rare Event Rule

If, under a given assumption (such as the assumption that boys and girls are equally likely), the probability of a particular observed event (such as 13 girls in 14 births) is extremely small, we conclude that the assumption is probably not correct.

- ❖ **Unusually high:** x successes among n trials is an *unusually high* number of successes if $P(x \text{ or more})$ is very small (such as 0.05 or less).
- ❖ **Unusually low:** x successes among n trials is an *unusually low* number of successes if $P(x \text{ or fewer})$ is very small (such as 0.05 or less).

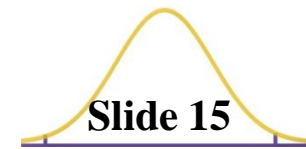
Definition



The **expected value** of a discrete random variable is denoted by E , and it represents the average value of the outcomes. It is obtained by finding the value of $\sum [x \cdot P(x)]$.

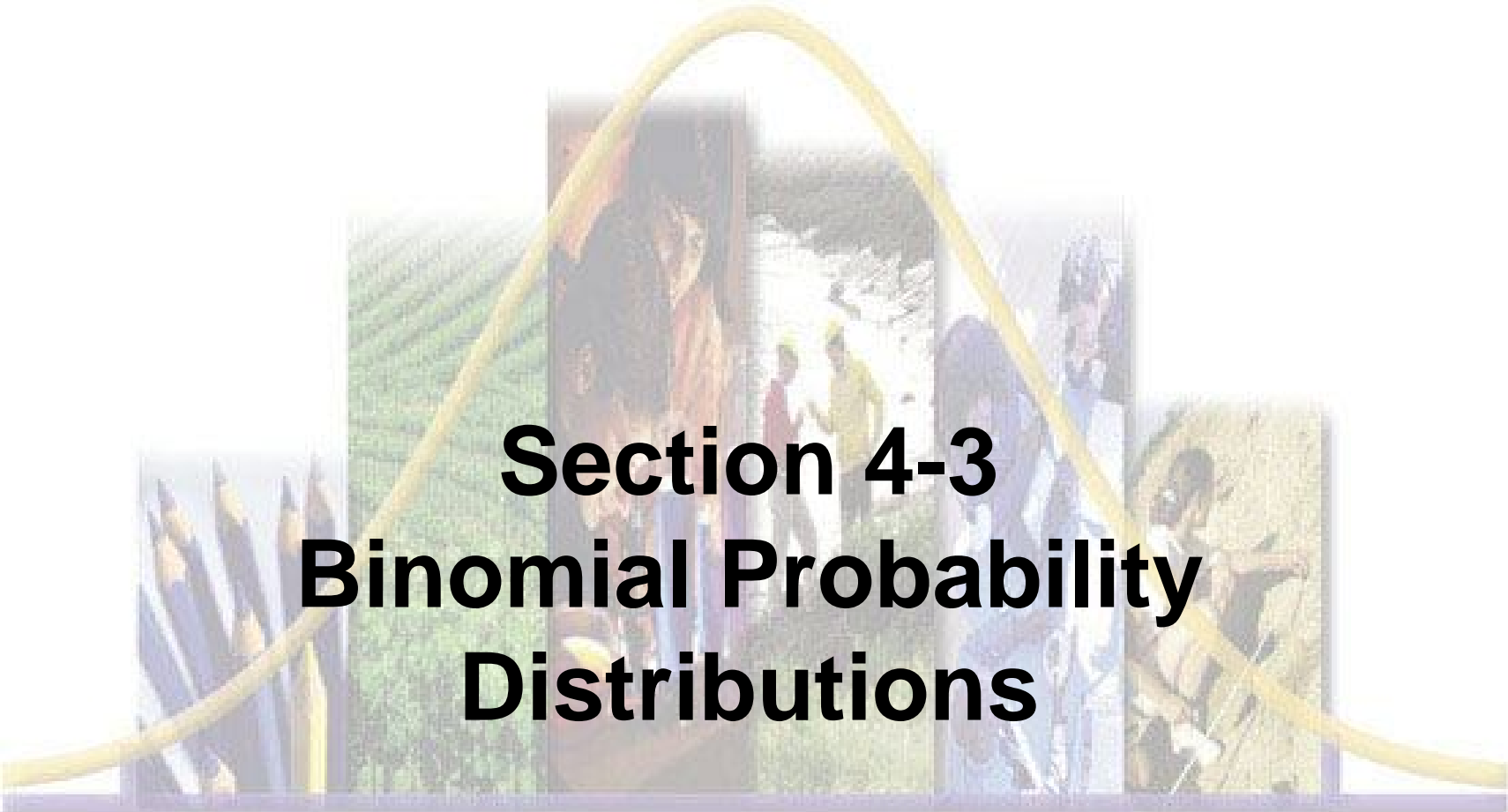
$$E = \sum [x \cdot P(x)]$$

Recap



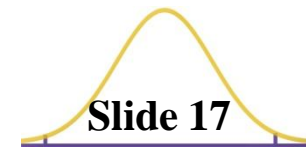
In this section we have discussed:

- ❖ Combining methods of descriptive statistics with probability.**
- ❖ Random variables and probability distributions.**
- ❖ Histograms**
- ❖ Requirements for a probability distribution.**
- ❖ Mean, variance and standard deviation of a probability distribution.**
- ❖ Identifying unusual results.**
- ❖ Expected value.**



Section 4-3
Binomial Probability
Distributions

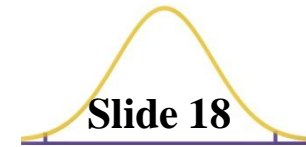
Definitions



A **binomial probability distribution** results from a procedure that meets all the following requirements:

1. The procedure has a ***fixed number of trials***.
2. The trials must be ***independent***. (The outcome of any individual trial doesn't affect the probabilities in the other trials.)
3. Each trial must have all outcomes classified into ***two categories***.
4. The probabilities must remain ***constant*** for each trial.

Notation for Binomial Probability Distributions

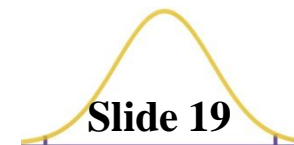


S and **F** (success and failure) denote two possible categories of all outcomes; ***p*** and ***q*** will denote the probabilities of **S** and **F**, respectively, so

$$P(S) = p \quad (p = \text{probability of success})$$

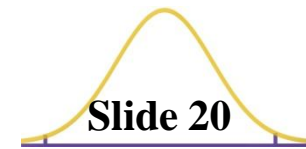
$$P(F) = 1 - p = q \quad (q = \text{probability of failure})$$

Notation (cont)



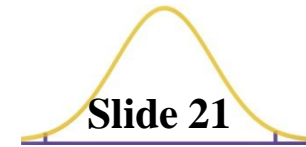
- n** denotes the number of fixed trials.
- x** denotes a specific number of successes in **n** trials, so **x** can be any whole number between **0** and **n** , inclusive.
- p** denotes the probability of **success** in *one* of the **n** trials.
- q** denotes the probability of **failure** in *one* of the **n** trials.
- $P(x)$** denotes the probability of getting exactly **x** successes among the **n** trials.

Important Hints



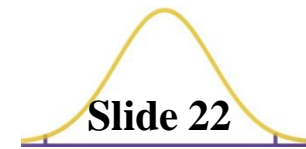
- ❖ Be sure that x and p both refer to the same category being called a success.
- ❖ When sampling without replacement, the events can be treated as if they were independent if the sample size is no more than 5% of the population size. (That is n is less than or equal to $0.05N$.)

Methods for Finding Probabilities



We will now present three methods for finding the probabilities corresponding to the random variable x in a binomial distribution.

Method 1: Using the Binomial Probability Formula



$$P(x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot q^{n-x}$$

for $x = 0, 1, 2, \dots, n$

where

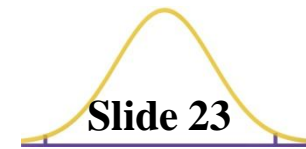
n = number of trials

x = number of successes among n trials

p = probability of success in any one trial

q = probability of failure in any one trial ($q = 1 - p$)

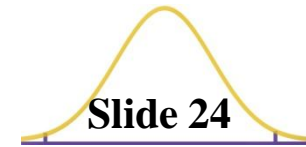
Method 2: Using Table A-1 in Appendix A



Part of Table A-1 is shown below. With $n = 4$ and $p = 0.2$ in the binomial distribution, the probabilities of 0, 1, 2, 3, and 4 successes are 0.410, 0.410, 0.154, 0.026, and 0.002 respectively.

From Table A-1:			Binomial probability distribution for $n = 4$ and $p = 0.2$	
n	x	p 0.20	x	$P(x)$
4	0	0.410	0	0.410
	1	0.410	1	0.410
	2	0.154	2	0.154
	3	0.026	3	0.026
	4	0.002	4	0.002

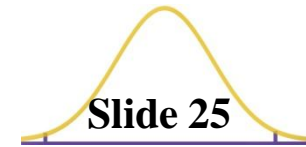
Method 3: Using Technology



STATDISK, Minitab, Excel and the TI-83 Plus calculator can all be used to find binomial probabilities.

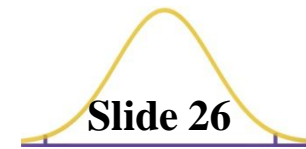


Strategy



- ❖ **Use computer software or a TI-83 calculator if available.**
- ❖ **If neither software or the TI-83 Plus calculator is available, use Table A-1, if possible.**
- ❖ **If neither software nor the TI-83 Plus calculator is available and the probabilities can't be found using Table A-1, use the binomial probability formula.**

Rationale for the Binomial Probability Formula

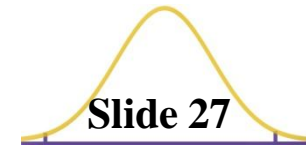


$$P(x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot q^{n-x}$$



Number of
outcomes with
exactly x
successes
among n trials

Binomial Probability Formula

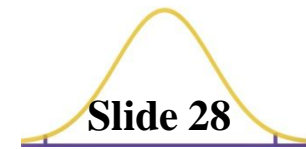


$$P(x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot q^{n-x}$$

Number of outcomes with exactly x successes among n trials

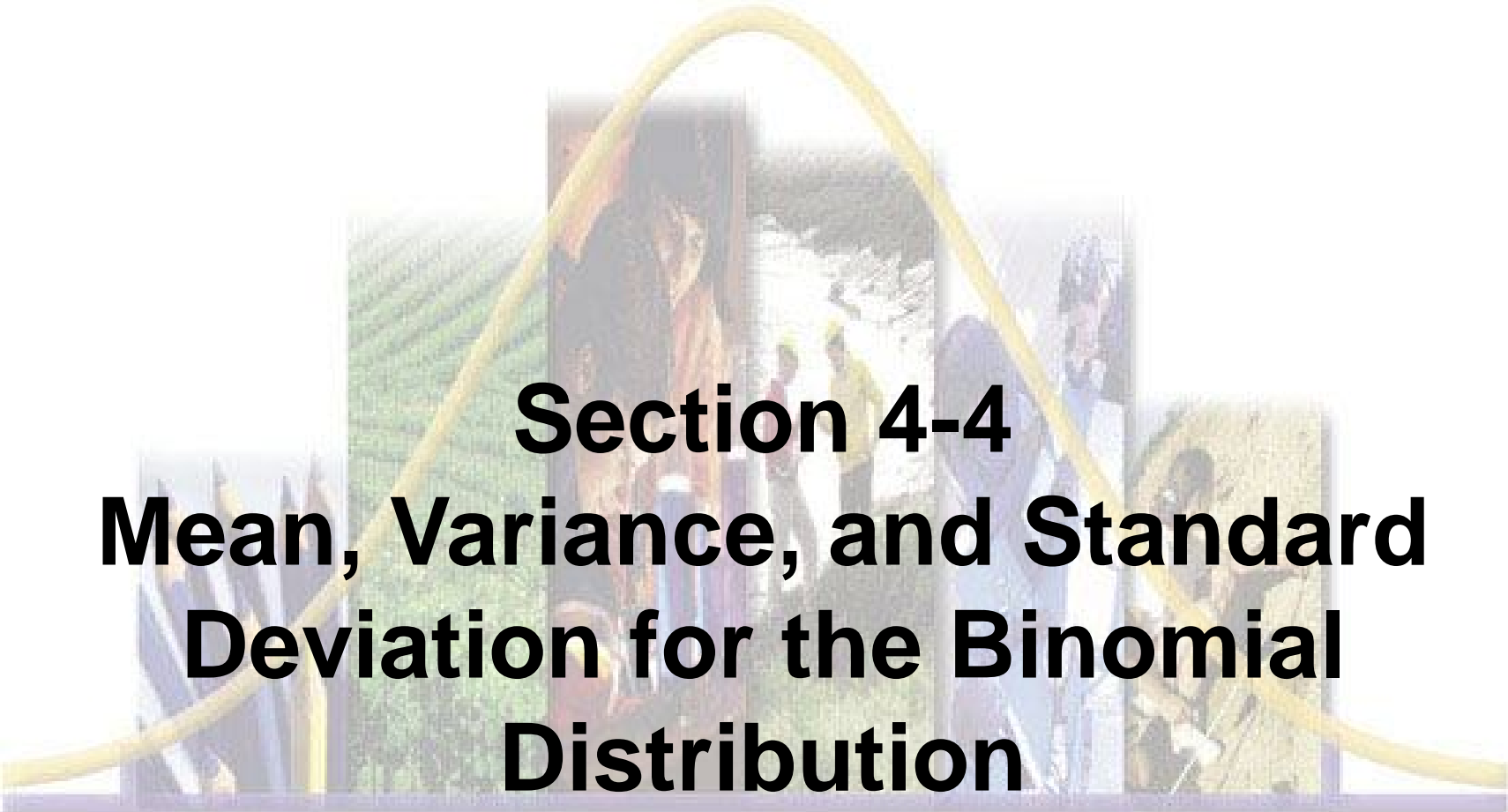
Probability of x successes among n trials for any one particular order

Recap



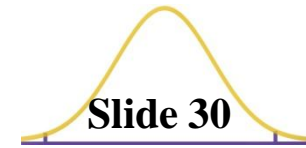
In this section we have discussed:

- ❖ **The definition of the binomial probability distribution.**
- ❖ **Notation.**
- ❖ **Important hints.**
- ❖ **Three computational methods**
- ❖ **Rationale for the formula.**



Section 4-4
Mean, Variance, and Standard
Deviation for the Binomial
Distribution

Any Discrete Probability Distribution: Formulas



Mean

$$\mu = \sum [x \cdot P(x)]$$

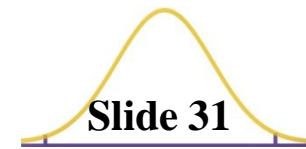
Variance

$$\sigma^2 = [\sum x^2 \cdot P(x)] - \mu^2$$

Std. Dev

$$\sigma = \sqrt{[\sum x^2 \cdot P(x)] - \mu^2}$$

Binomial Distribution: Formulas



Mean $\mu = n \cdot p$

Variance $\sigma^2 = n \cdot p \cdot q$

Std. Dev. $\sigma = \sqrt{n \cdot p \cdot q}$

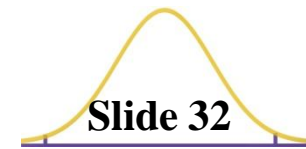
Where

n = number of fixed trials

p = probability of **success** in one of the **n** trials

q = probability of **failure** in one of the **n** trials

Example



Find the mean and standard deviation for the number of girls in groups of 14 births.

This scenario is a binomial distribution where:

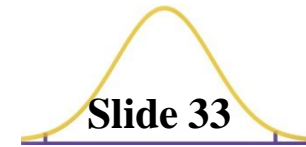
$$n = 14$$

$$p = 0.5$$

$$q = 0.5$$

Using the binomial distribution formulas:

Example (cont)



This scenario is a binomial distribution where

$$n = 14$$

$$p = 0.5$$

$$q = 0.5$$

Using the binomial distribution formulas:

$$\mu = (14)(0.5) = 7 \text{ girls}$$

$$\sigma = \sqrt{(14)(0.5)(0.5)} = 1.9 \text{ girls (rounded)}$$

Interpretation of Results

Slide 34

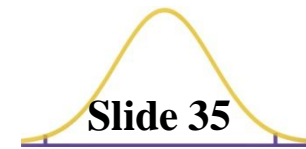


It is especially important to interpret results. The **range rule of thumb** suggests that values are unusual if they lie outside of these limits:

Maximum usual values = $\mu + 2\sigma$

Minimum usual values = $\mu - 2\sigma$

Example



Determine whether 68 girls among 100 babies could easily occur by chance.

For this binomial distribution,

$$\mu = 50 \text{ girls}$$

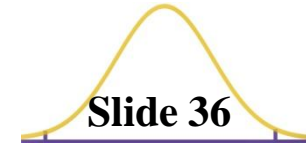
$$\sigma = 5 \text{ girls}$$

$$\mu + 2 \sigma = 50 + 2(5) = 60$$

$$\mu - 2 \sigma = 50 - 2(5) = 40$$

The usual number girls among 100 births would be from 40 to 60. So 68 girls in 100 births is an unusual result.

Recap



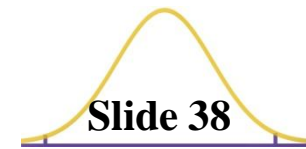
In this section we have discussed:

- ❖ **Mean, variance and standard deviation formulas for the any discrete probability distribution.**
- ❖ **Mean, variance and standard deviation formulas for the binomial probability distribution.**
- ❖ **Interpreting results**



Section 4-5
The Poisson Distribution

Definition

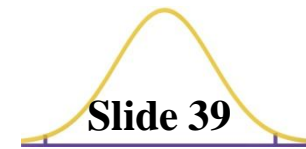


The **Poisson distribution** is a discrete probability distribution that applies to occurrences of some event *over a specified interval*. The random variable **x** is the number of occurrences of the event in an interval. The interval can be time, distance, area, volume, or some similar unit.

Formula

$$P(x) = \frac{\mu^x \cdot e^{-\mu}}{x!} \quad \text{where } e \approx 2.71828$$

Poisson Distribution Requirements

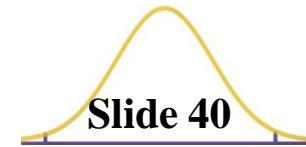


- ❖ The random variable X is the number of occurrences of an event *over some interval*.
- ❖ The occurrences must be *random*.
- ❖ The occurrences must be *independent* of each other.
- ❖ The occurrences must be *uniformly distributed* over the interval being used.

Parameters

- ❖ The mean is μ .
- ❖ The standard deviation is $\sigma = \sqrt{\mu}$.

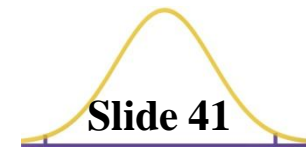
Difference from a Binomial Distribution



The Poisson distribution differs from the binomial distribution in these fundamental ways:

- ❖ The binomial distribution is affected by the sample size n and the probability p , whereas the Poisson distribution is affected only by the mean μ .
- ❖ In a binomial distribution the possible values of the random variable are x are $0, 1, \dots, n$, but a Poisson distribution has possible x values of $0, 1, \dots$, with no upper limit.

Example

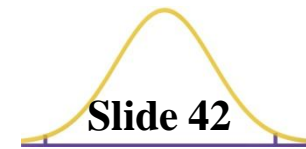


World War II Bombs In analyzing hits by V-1 buzz bombs in World War II, South London was subdivided into 576 regions, each with an area of 0.25 km^2 . A total of 535 bombs hit the combined area of 576 regions

If a region is randomly selected, find the probability that it was hit exactly twice.

The Poisson distribution applies because we are dealing with occurrences of an event (bomb hits) over some interval (a region with area of 0.25 km^2).

Example



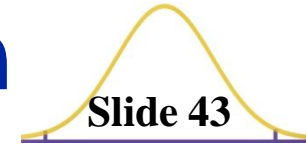
The mean number of hits per region is

$$\mu = \frac{\text{number of bomb hits}}{\text{number of regions}} = \frac{535}{576} = 0.929$$

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} = \frac{0.929^2 \cdot 2.71828^{-0.929}}{2!} = \frac{0.863 \cdot 0.395}{2} = 0.170$$

The probability of a particular region being hit exactly twice is $P(2) = 0.170$.

Poisson as Approximation to Binomial



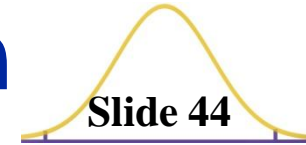
The Poisson distribution is sometimes used to approximate the binomial distribution when n is large and p is small.

Rule of Thumb

❖ $n \geq 100$

❖ $np \leq 10$

Poisson as Approximation to Binomial - μ



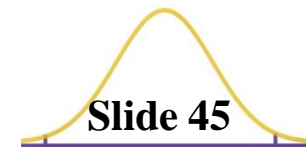
❖ $n \geq 100$

❖ $np \leq 10$

Value for μ

$$\mu = n \cdot p$$

Recap



In this section we have discussed:

- ❖ **Definition of the Poisson distribution.**
- ❖ **Requirements for the Poisson distribution.**
- ❖ **Poisson distribution requirements.**
- ❖ **Poisson approximation to the binomial.**