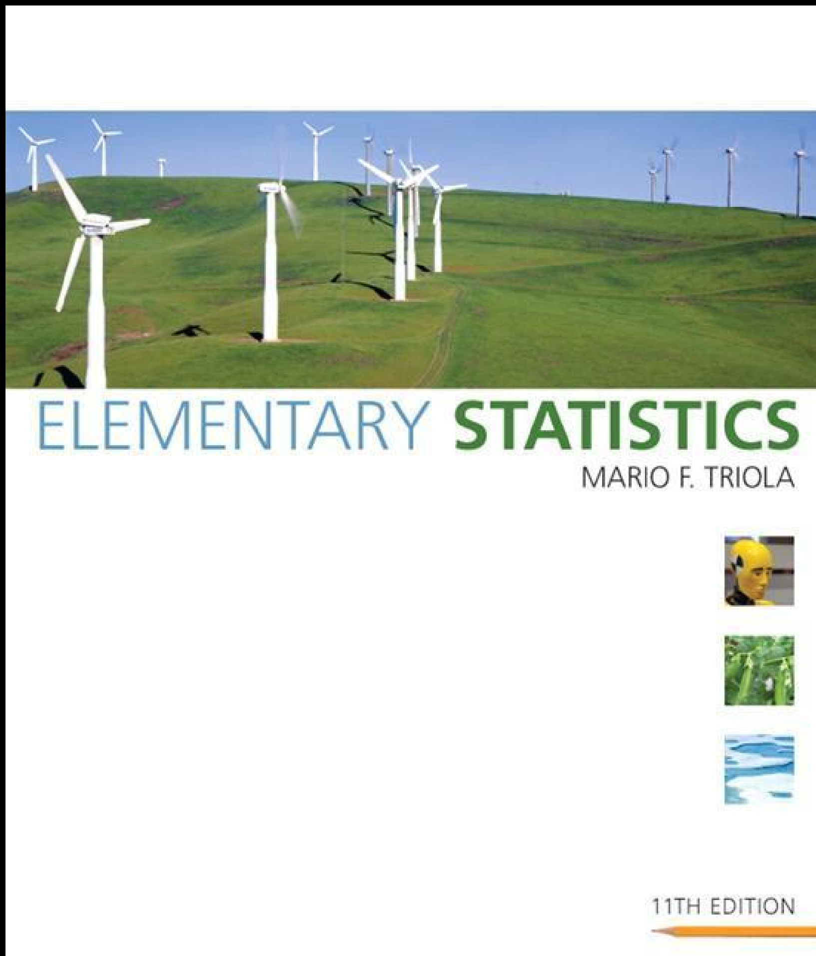
Lecture Slides



Elementary Statistics
Eleventh Edition

and the Triola Statistics Series

by Mario F. Triola

Chapter 5
Probability Distributions

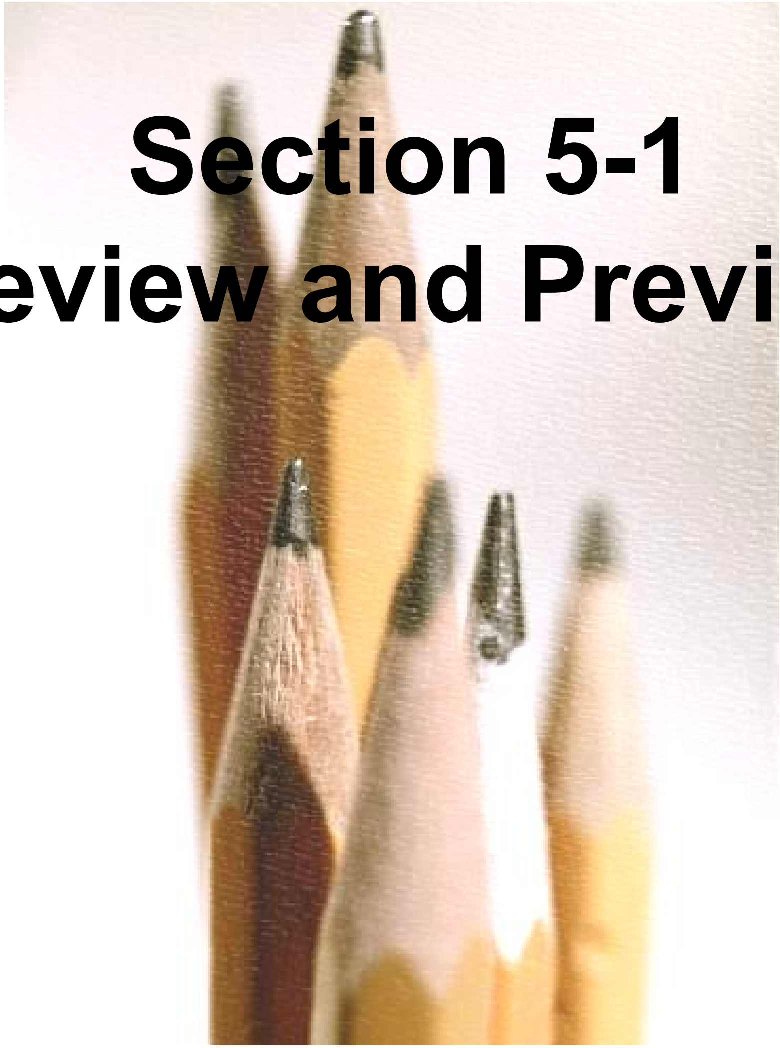
5-1 Review and Preview

5-2 Random Variables

5-3 Binomial Probability Distributions

5-4 Mean, Variance and Standard Deviation for the Binomial Distribution

5-5 Poisson Probability Distributions



Section 5-1
Review and Preview

Review and Preview

This chapter combines the methods of descriptive statistics presented in Chapter 2 and 3 and those of probability presented in Chapter 4 to describe and analyze

probability distributions.

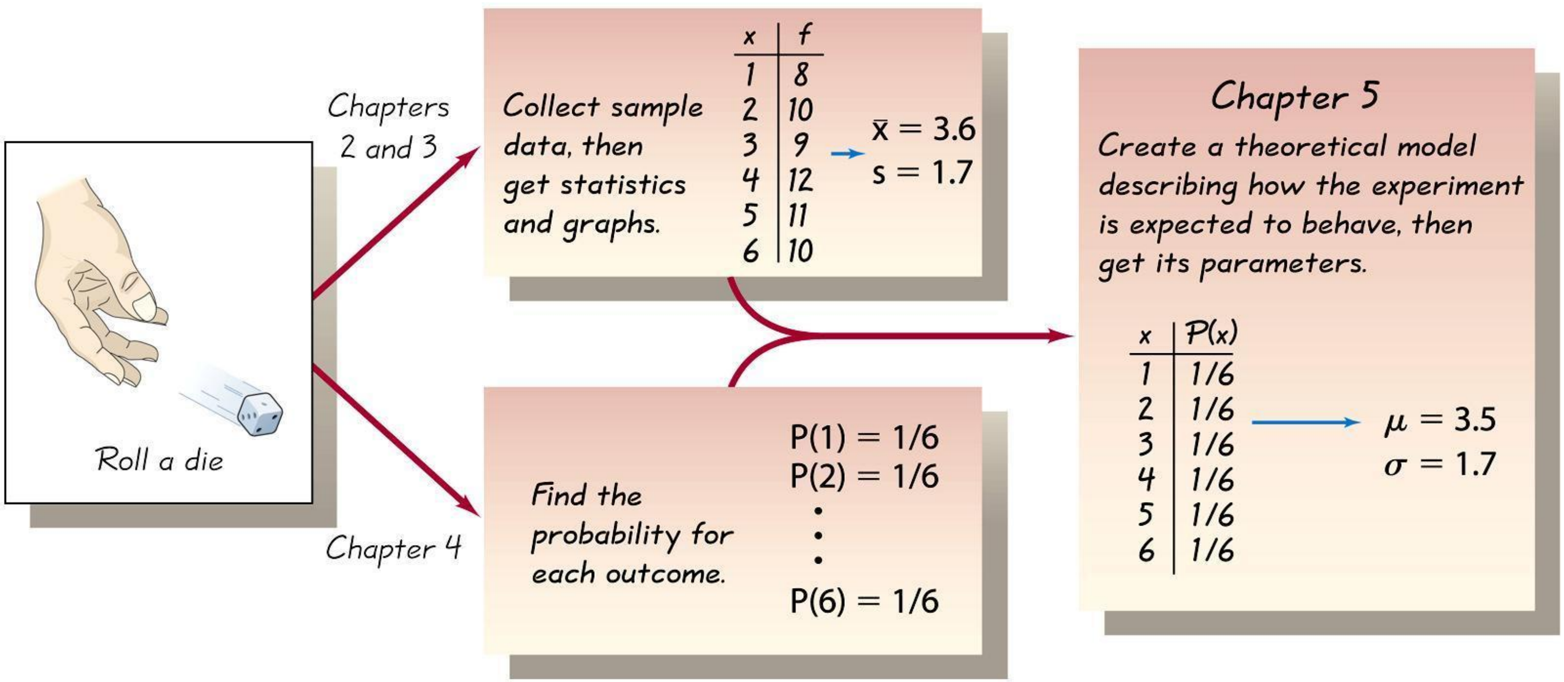
Probability Distributions describe what will probably happen instead of what actually did happen, and they are often given in the format of a graph, table, or formula.

Preview

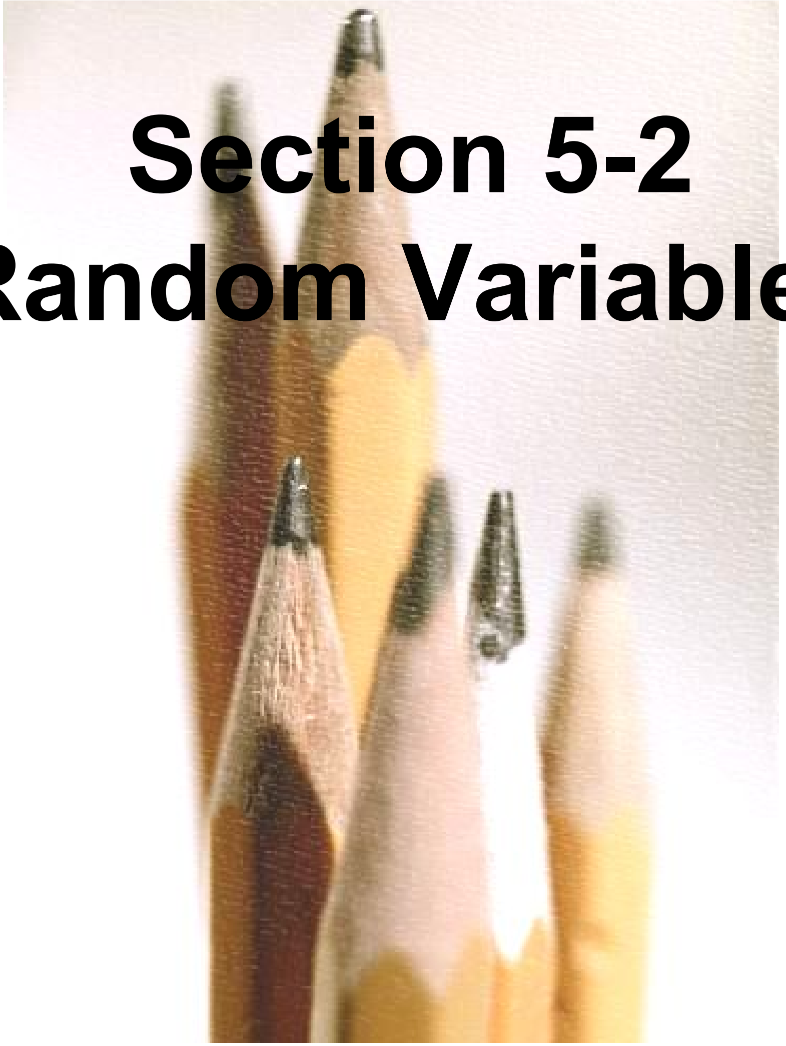
In order to fully understand probability distributions, we must first understand the concept of a random variable, and be able to distinguish between discrete and continuous random variables. In this chapter we focus on discrete probability distributions. In particular, we discuss binomial and Poisson probability distributions.

Combining Descriptive Methods and Probabilities

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



Section 5-2
Random Variables



Key Concept

This section introduces the important concept of a probability distribution, which gives the probability for each value of a variable that is determined by chance.

Give consideration to distinguishing between outcomes that are likely to occur by chance and outcomes that are "unusual" in the sense they are not likely to occur by chance.

Key Concept

- The concept of random variables and how they relate to probability distributions
- Distinguish between discrete random variables and continuous random variables
- Develop formulas for finding the mean, variance, and standard deviation for a probability distribution
- Determine whether outcomes are likely to occur by chance or they are unusual (in the sense that they are not likely to occur by chance)

Random Variable
Probability Distribution

- **Random variable**
  a variable (typically represented by *x*) that has a single numerical value, determined by chance, for each outcome of a procedure
- **Probability distribution**
  a description that gives the probability for each value of the random variable; often expressed in the format of a graph, table, or formula

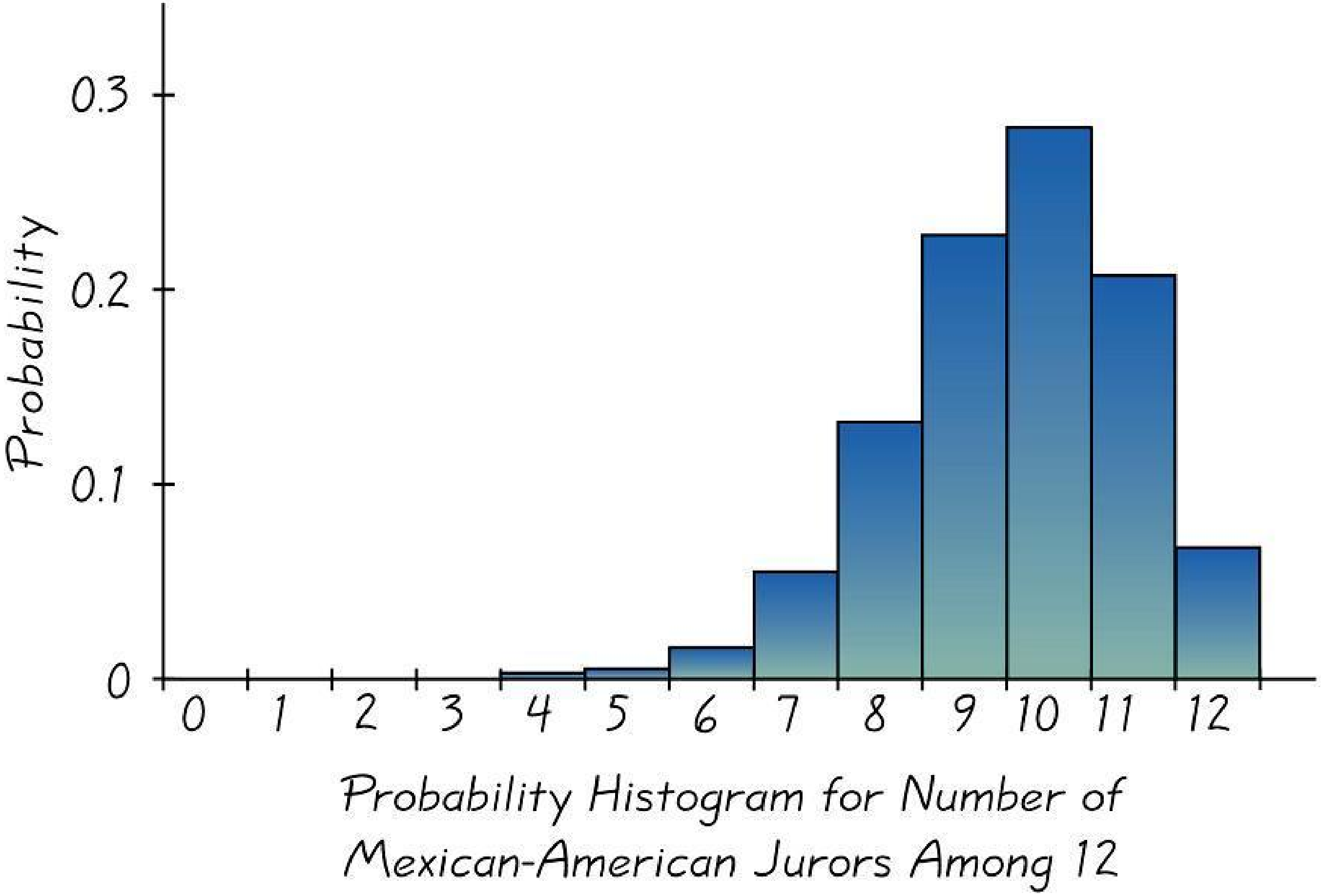
Discrete and Continuous Random Variables

- **Discrete random variable**
  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

- **Continuous random variable**
  infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions

Graphs

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



Requirements for Probability Distribution

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

where *x* assumes all possible values.

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

for every individual value of *x*.

Mean, Variance and Standard Deviation of a Probability Distribution

$$\mu = \Sigma\,[x \cdot P(x)]$$ **Mean**

$$\sigma^2 = \Sigma\,[(x - \mu)^2 \cdot P(x)]$$ **Variance**

$$\sigma^2 = \Sigma\,[x^2 \cdot P(x)] - \mu^2$$ **Variance (shortcut)**

$$\sigma = \sqrt{\Sigma\,[x^2 \cdot P(x)] - \mu^2}$$ **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 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 for Inferential Statistics

If, under a given assumption (such as the assumption that a coin is fair), the probability of a particular observed event (such as 992 heads in 1000 tosses of a coin) is extremely small, we conclude that the assumption is probably not correct.

Identifying Unusual Results Probabilities

Using Probabilities to Determine When Results Are Unusual

- **Unusually high**: x successes among n trials is an **unusually high** number of successes if $P(x \text{ or more}) \leq 0.05$.

- **Unusually low**: x successes among n trials is an **unusually low** number of successes if $P(x \text{ or fewer}) \leq 0.05$.

Expected Value

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

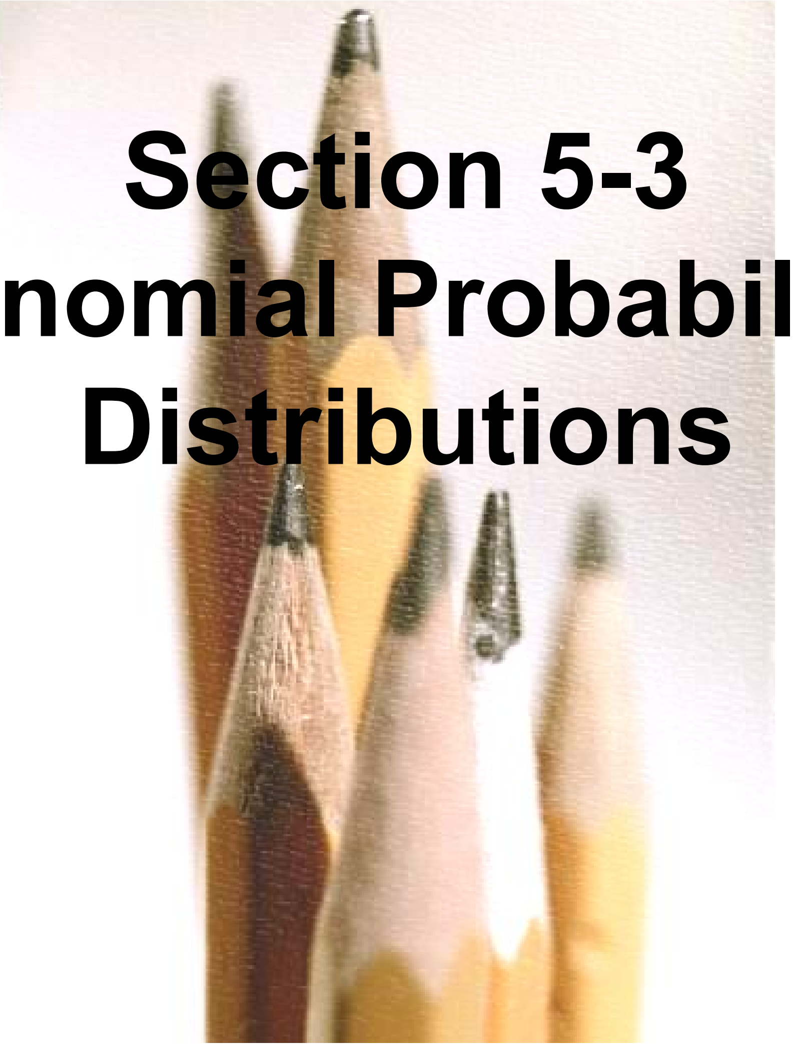
$$E = \Sigma\,[x \cdot P(x)]$$

Recap

In this section we have discussed:

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

Section 5-3 Binomial Probability Distributions



Key Concept

This section presents a basic definition of a binomial distribution along with notation, and methods for finding probability values.

Binomial probability distributions allow us to deal with circumstances in which the outcomes belong to two relevant categories such as acceptable/defective or survived/died.

Binomial Probability Distribution

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 (commonly referred to as success and failure).
4. The probability of a success remains the same in all trials.

Notation for Binomial Probability Distributions

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

$P(S) = p$ **(p = probability of success)**

$P(F) = 1 – p = q$ **(q = probability of failure)**

Notation (continued)

n denotes the fixed number of 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, consider events to be independent if $n < 0.05N$.

Methods for Finding Probabilities

We will now discuss 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 $x = 0, 1, 2, \ldots, n$

where

n = number of trials

x = number of successes among n trials

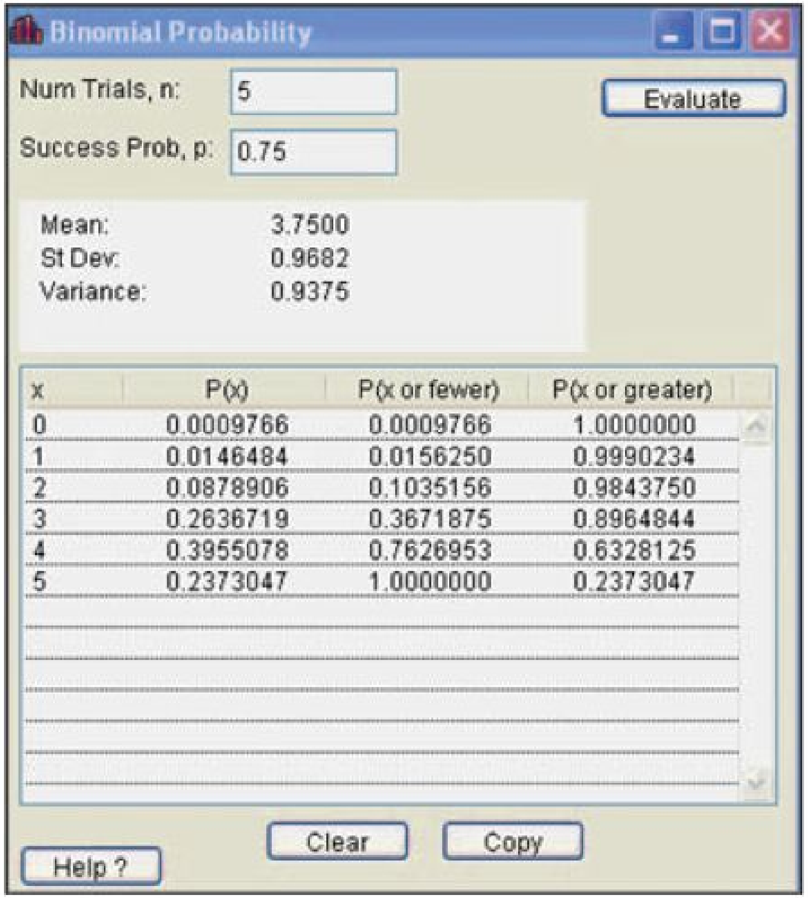
p = probability of success in any one trial

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

Method 2: Using Technology

STATDISK, Minitab, Excel, SPSS, SAS and the TI-83/84 Plus calculator can be used to find binomial probabilities.

STATDISK



x	P(x)	P(x or fewer)	P(x or greater)
0	0.0009766	0.0009766	1.0000000
1	0.0146484	0.0156250	0.9990234
2	0.0878906	0.1035156	0.9843750
3	0.2636719	0.3671875	0.8964844
4	0.3955078	0.7626953	0.6328125
5	0.2373047	1.0000000	0.2373047

MINITAB

x	P(x)
0	0.000977
1	0.014648
2	0.087891
3	0.263672
4	0.395508
5	0.237305

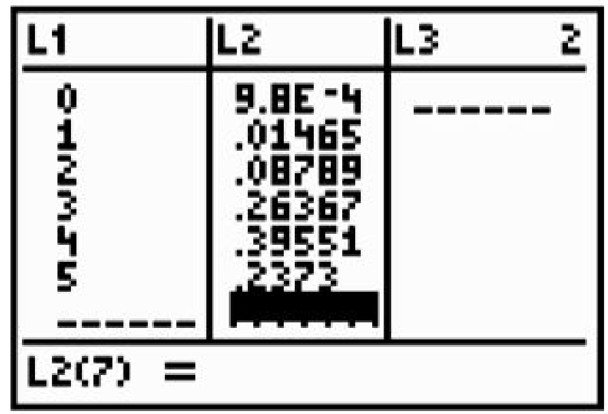
Method 2: Using Technology

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

EXCEL

	A	B
1	0	0.000977
2	1	0.014648
3	2	0.087891
4	3	0.263672
5	4	0.395508
6	5	0.237305

TI-83 PLUS Calculator



Method 3: Using Table A-1 in Appendix A

Part of Table A-1 is shown below. With *n* = 12 and *p* = 0.80 in the binomial distribution, the probabilities of 4, 5, 6, and 7 successes are 0.001, 0.003, 0.016, and 0.053 respectively.

n	x	p 0.80
	4	0.001
	5	0.003
	6	0.016
	7	0.053

→

x	p
4	0.001
5	0.003
6	0.016
7	0.053

Strategy for Finding Binomial Probabilities

- Use computer software or a TI-83 Plus calculator if available.
- If neither software nor 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

$$P(x) = \underbrace{\frac{n!}{(n-x)!x!}}_{\text{The number of outcomes with exactly } x \text{ successes among } n \text{ trials}} \cdot p^x \cdot q^{n-x}$$

Binomial Probability Formula

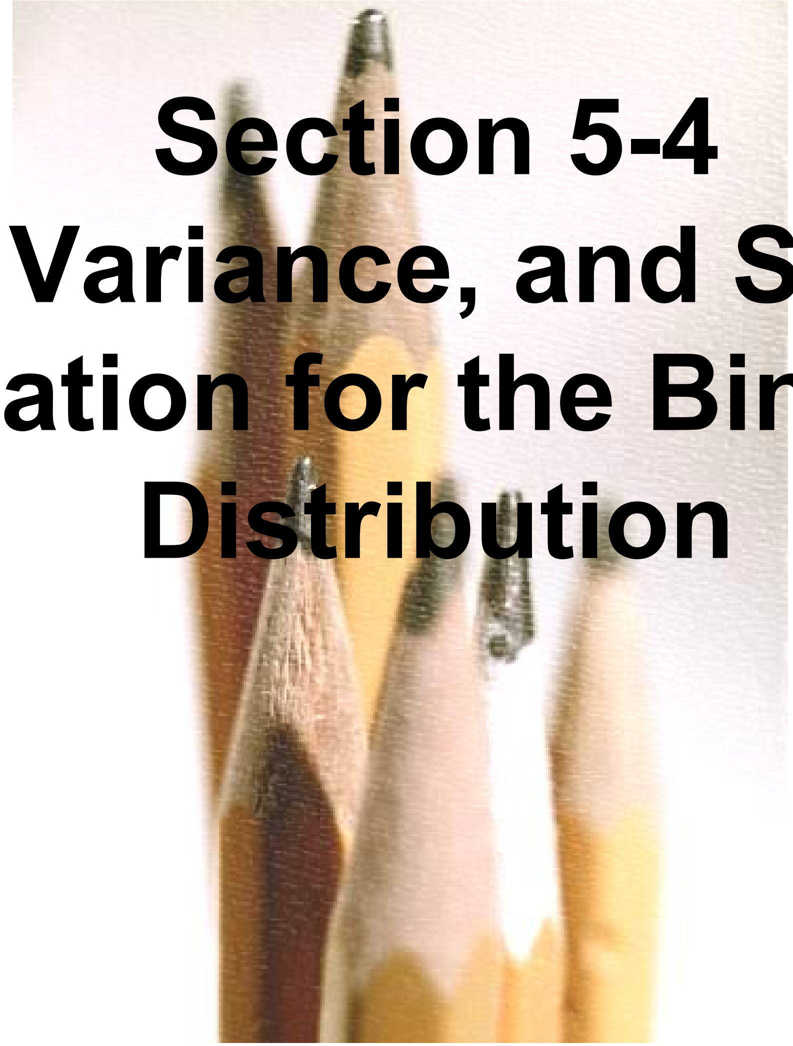
$$P(x) = \underbrace{\frac{n!}{(n-x)!x!}}_{\substack{\text{Number of outcomes with} \\ \text{exactly } x \text{ successes} \\ \text{among } n \text{ trials}}} \cdot \underbrace{p^{x} \cdot q^{n-x}}_{\substack{\text{The probability of} \\ x \text{ successes among } n \text{ trials} \\ \text{for any one particular order}}}$$

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 5-4
Mean, Variance, and Standard Deviation for the Binomial Distribution



Key Concept

In this section we consider important characteristics of a binomial distribution including center, variation and distribution. That is, given a particular binomial probability distribution we can find its mean, variance and standard deviation.

A strong emphasis is placed on interpreting and understanding those values.

For Any Discrete Probability Distribution: Formulas

Mean $\mu = \Sigma[x \cdot P(x)]$

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

Std. Dev $\sigma = \sqrt{[\Sigma x^2 \cdot P(x)] - \mu^2}$

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**

Interpretation of Results

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

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

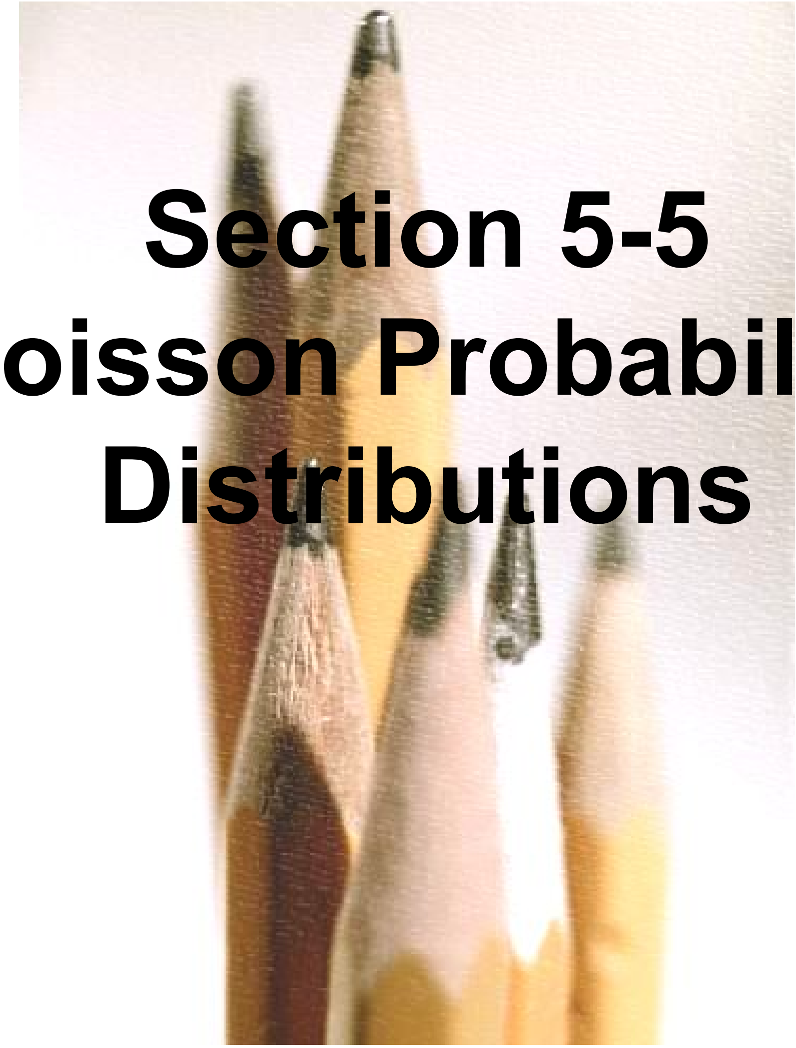
$$\text{Minimum usual values} = \mu - 2\,\sigma$$

Recap

In this section we have discussed:

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

Section 5-5
Poisson Probability Distributions



Key Concept

The Poisson distribution is another discrete probability distribution which is important because it is often used for describing the behavior of rare events (with small probabilities).

Poisson Distribution

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

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

Requirements of the Poisson Distribution

- 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 $\sigma = \sqrt{\mu}$.

Parameters of the Poisson Distribution

- 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 x are 0, 1, . . . n, but a Poisson distribution has possible x values of 0, 1, 2, . . . , with no upper limit.**

Poisson as an Approximation to the Binomial Distribution

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 an Approximation to the Binomial Distribution - μ

If both of the following requirements are met,

- $n \geq 100$
- $np \leq 10$

then use the following formula to calculate μ,

Value for μ

$$\mu = n \cdot p$$

Recap

In this section we have discussed:

- **Definition of the Poisson distribution.**
- **Requirements for the Poisson distribution.**
- **Difference between a Poisson distribution and a binomial distribution.**
- **Poisson approximation to the binomial.**