Chapter 4

Probability: The Study of Randomness

Introduction to the Practice of

STATISTICS EIGHTH EDITION

Moore / McCabe / Craig

Lecture Presentation Slides

Chapter 4
Probability: The Study of Randomness

4.1 Randomness

4.2 Probability Models

4.3 Random Variables

4.4 Means and Variances of Random Variables

4.5 General Probability Rules*

4.1 Randomness

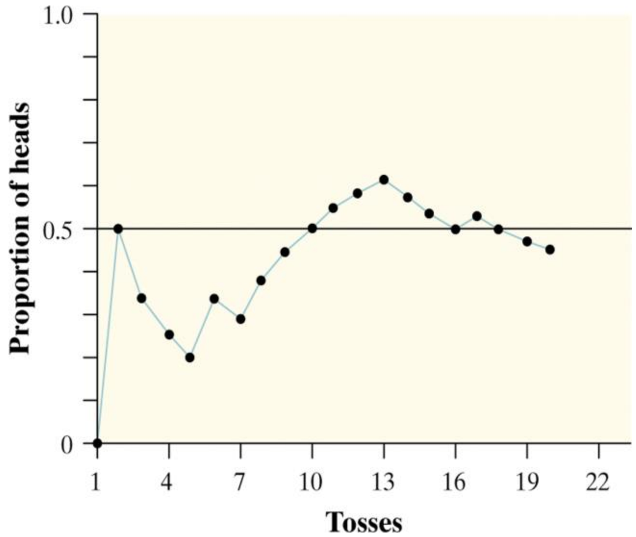
- The language of probability
- Thinking about randomness
- The uses of probability

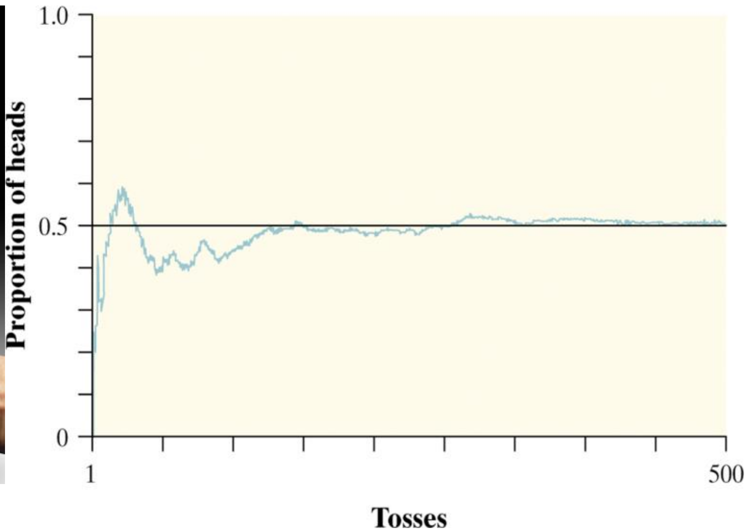
The Language of Probability

Chance behavior is unpredictable in the short run but has a regular and predictable pattern in the long run.

We call a phenomenon **random** if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.

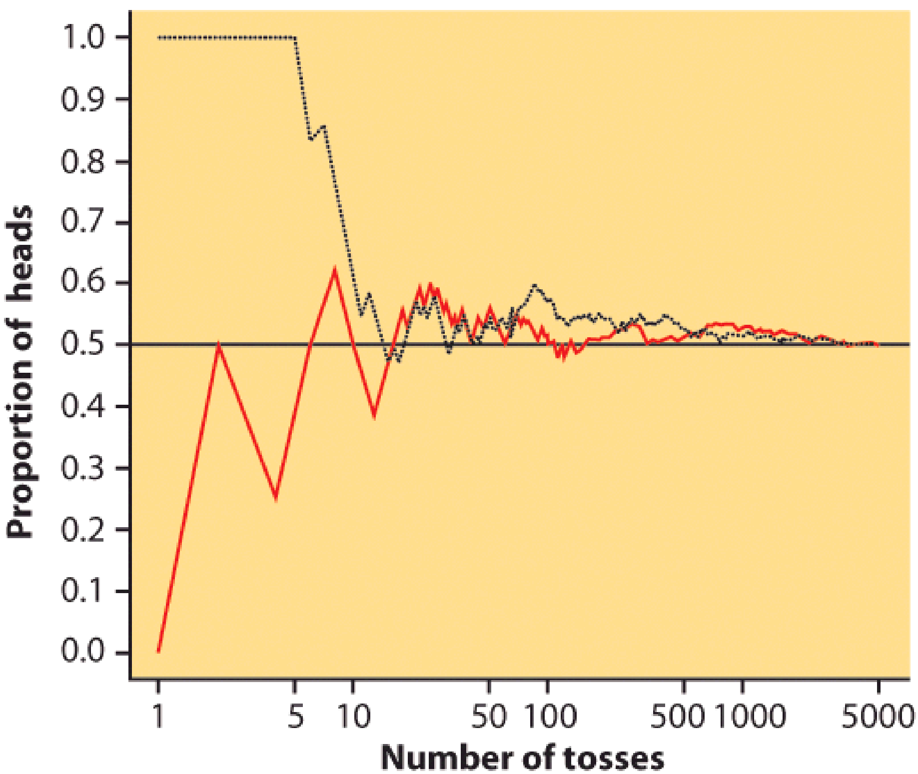
The **probability** of any outcome of a chance process is the proportion of times the outcome would occur in a very long series of repetitions.





Thinking About Randomness

The result of any single coin toss is random. But the result over many tosses is predictable, as long as the trials are **independent** (i.e., the outcome of a new coin flip is not influenced by the result of the previous flip).



The probability of heads is 0.5 = the proportion of times you get heads in many repeated trials.

First series of tosses

Second series

4.3 Random Variables

- Random variable
- Discrete random variables
- Continuous random variables
- Normal distributions as probability distributions

Random Variables

A numerical variable that describes the outcomes of a chance process is called a **random variable.**

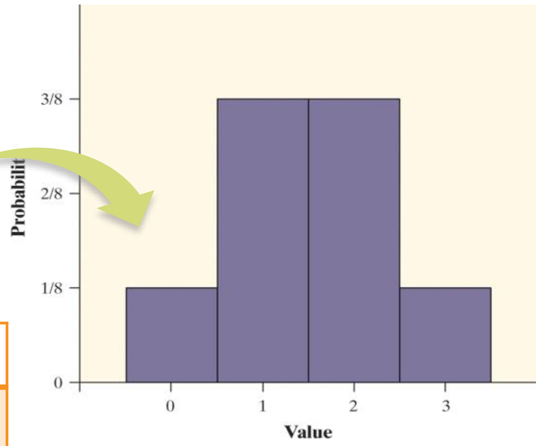
A **random variable** takes numerical values that describe the outcomes of some chance process.

The **probability distribution** of a random variable gives its possible values and their probabilities.

Example: Consider tossing a fair coin 3 times.
Define X = the number of heads obtained

X = 0: TTT
X = 1: HTT THT TTH
X = 2: HHT HTH THH
X = 3: HHH

Value	0	1	2	3
Probability	1/8	3/8	3/8	1/8



Random Variables

Discrete random variables commonly arise from situations that involve counting something, as in the previous example.

Situations that involve measuring something often result in a **continuous random variable.**

A **continuous random variable** ***Y*** takes on all values in an interval of numbers. The probability distribution of *Y* is described by a **density curve.** The probability of any event is the area under the density curve and above the values of *Y* that make up the event.

Normal Probability Models

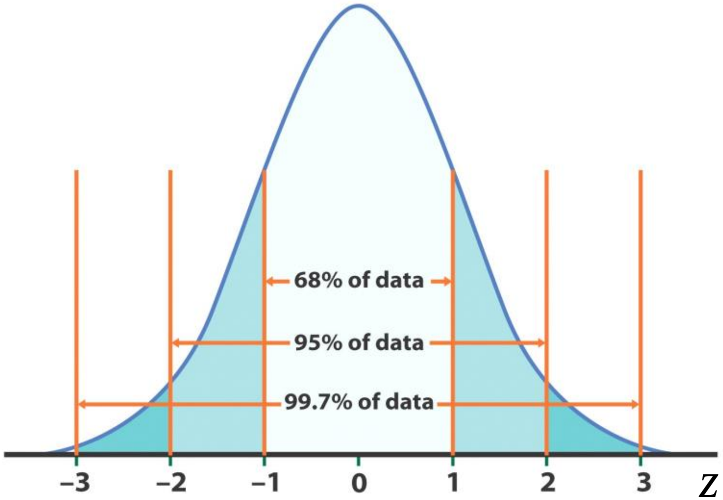
Often, the density curve used to assign probabilities to intervals of outcomes is the Normal curve.

<u>Normal distributions are probability models</u>:

Probabilities can be assigned to intervals of outcomes using the Standard Normal probabilities in Table A.

We **standardize** normal data by calculating *z*-scores so that any Normal curve $N(\mu,\sigma)$ can be transformed into the standard Normal curve $N(0,1)$.

$$\boxed{z = \frac{(x-\mu)}{\sigma}}$$



4.4 Means and Variances of Random Variables

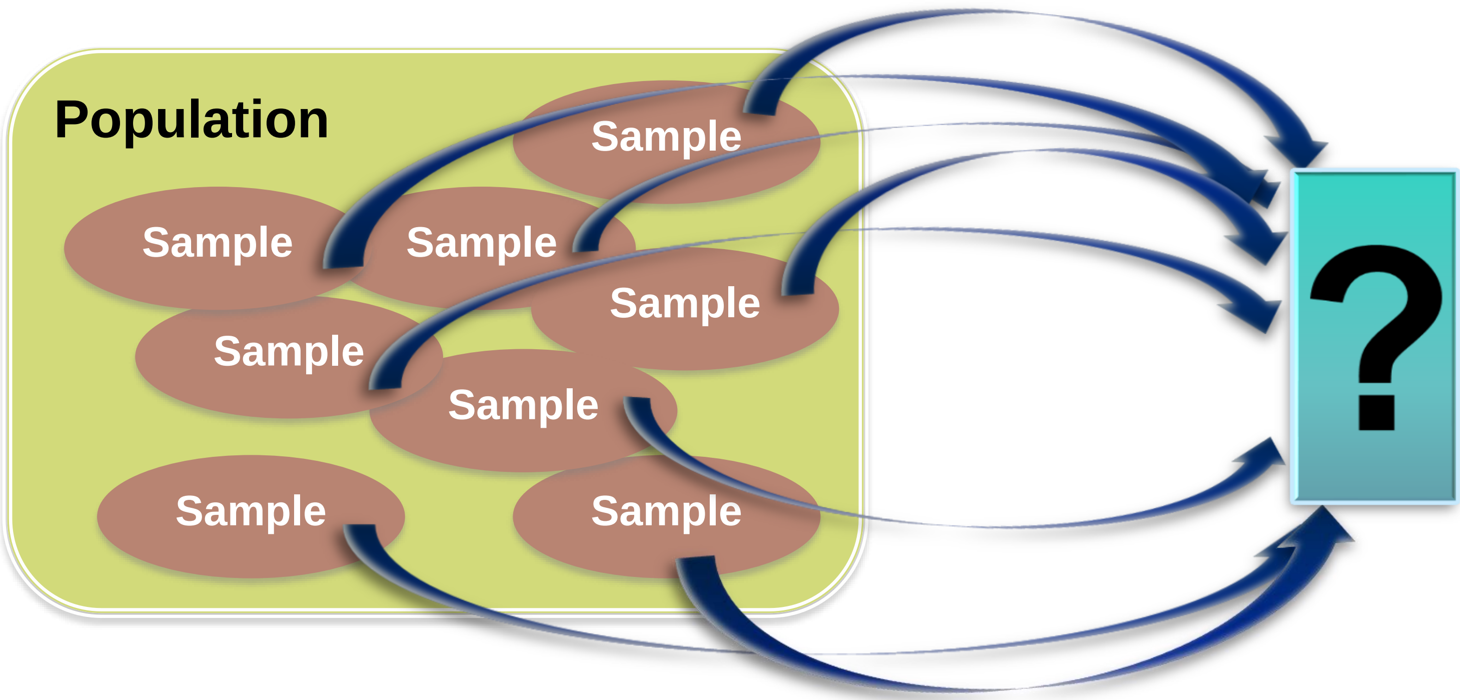
- The mean of a random variable
- The law of large numbers
- Rules for means
- The variance of a random variable
- Rules for variances and standard deviations

The Law of Large Numbers

Suppose we would like to estimate an unknown μ. We could select an SRS and base our estimate on the sample mean. However, a different SRS would probably yield a different sample mean.

This basic fact is called **sampling variability:** The value of a statistic varies in repeated random sampling.

To make sense of sampling variability, we ask, "What would happen if we took many samples?"



The Law of Large Numbers

How can $\bar{x}$ be an accurate estimate of μ? After all, different random samples would produce different values of $\bar{x}$.

If we keep on taking larger and larger samples, the statistic $\bar{x}$ is guaranteed to get closer and closer to the parameter μ.

Draw independent observations at random from any population with finite mean μ. The **law of large numbers** says that, as the number of observations drawn increases, the sample mean of the observed values gets closer and closer to the mean μ of the population.

Chapter 4
Probability: The Study of Randomness

4.1 Randomness

4.2 Probability Models

4.3 Random Variables

4.4 Means and Variances of Random Variables

4.5 General Probability Rules*

Chapter 5

Sampling Distributions

Introduction to the Practice of
STATISTICS EIGHTH EDITION

Moore / McCabe / Craig

Lecture Presentation Slides

Chapter 5
Sampling Distributions

5.1 The Sampling Distribution of a Sample Mean

5.2 Sampling Distributions for Counts and Proportions

5.1 The Sampling Distribution of a Sample Mean

- Population distribution vs. sampling distribution
- The mean and standard deviation of the sample mean
- Sampling distribution of a sample mean
- Central limit theorem

Sampling Distributions

The law of large numbers assures us that if we measure enough subjects, the statistic $\bar{X}$ will eventually get very close to the unknown parameter μ.

If we took every one of the possible samples of a certain size, calculated the sample mean for each, and graphed all of those values, we'd have a **sampling distribution.**

The **population distribution** of a variable is the distribution of values of the variable among all individuals in the population.

The **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population.

Mean and Standard Deviation of a Sample Mean

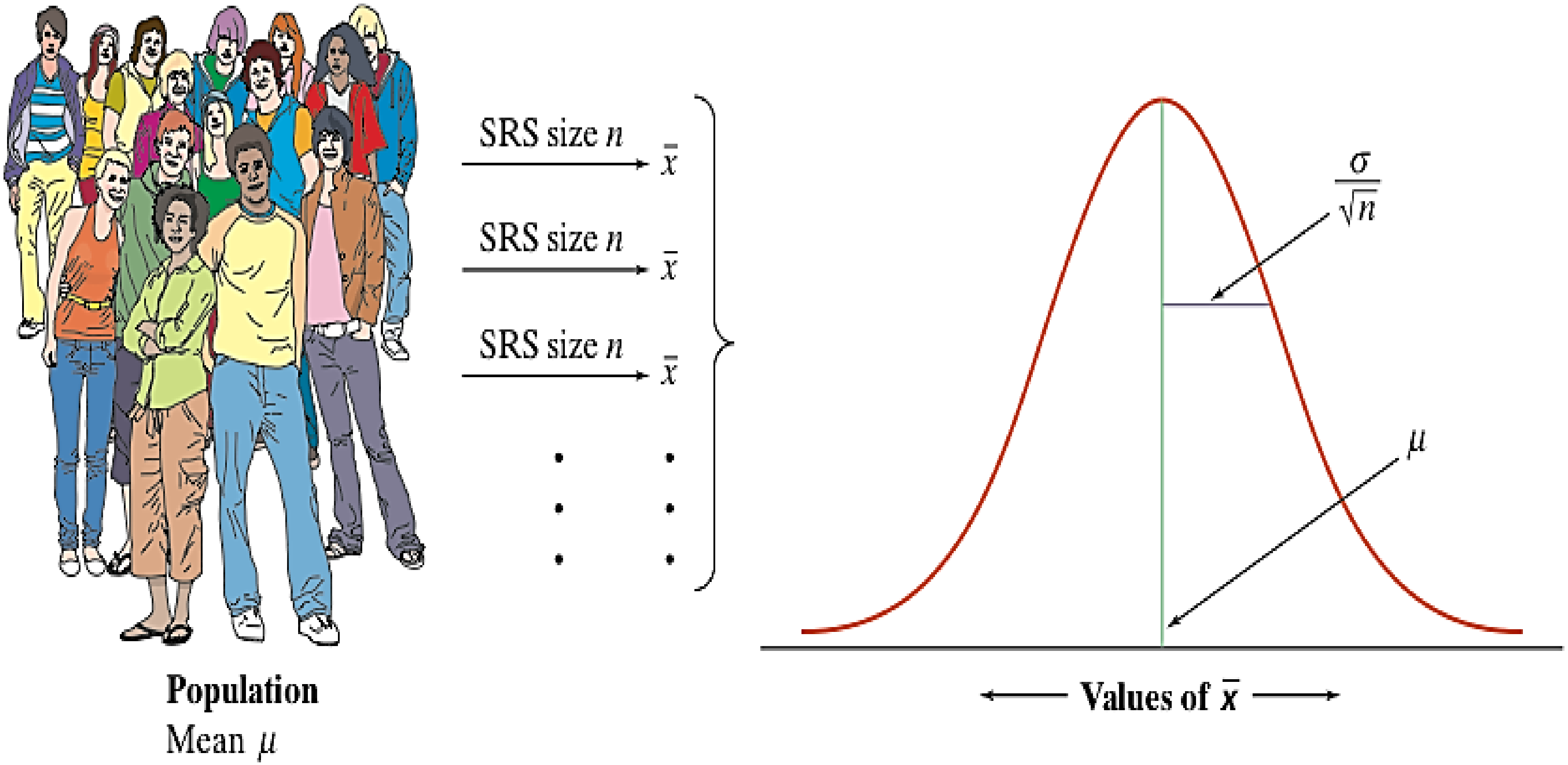
Mean of a sampling distribution of a sample mean
There is no tendency for a sample mean to fall systematically above or below μ, even if the distribution of the raw data is skewed. Thus, the sample mean is an **unbiased estimate** of the population mean μ.

Standard deviation of a sampling distribution of a sample mean
The standard deviation of the sampling distribution measures how much the sample statistic varies from sample to sample. It is smaller than the standard deviation of the population by a factor of $\sqrt{n}$.

➔ **Averages are less variable than individual observations.**

The Sampling Distribution of a Sample Mean

When we choose many SRSs from a population, the sampling distribution of the sample mean is centered at the population mean μ and is less spread out than the population distribution. Here are the facts.



The Central Limit Theorem

Most population distributions are not Normal. What is the shape of the sampling distribution of sample means when the population distribution isn't Normal?

It is a remarkable fact that, *as the sample size increases, the distribution of sample means begins to look more and more like a Normal distribution!*

When the sample is large enough, the distribution of sample means is very close to Normal, *no matter what shape the population distribution has*.

Draw an SRS of size n from any population with mean μ and standard deviation σ. The *central limit theorem* (CLT) says that when n is sufficiently large, the sampling distribution of the sample mean is approximately Normal, specifically,

$$\bar{X} \dot{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

5.2 Sampling Distributions for Counts and Proportions

- Binomial distributions for sample counts
- Binomial distributions in statistical sampling
- Finding binomial probabilities
- Binomial mean and standard deviation
- Sample proportions
- Normal approximation for counts and proportions
- Binomial formula

Sampling Distribution of a Sample Proportion

In many situations in education and the social sciences, the random variable of interest takes on only <u>two</u> possible values (e.g., pass/fail, yes/no, or *generically*, success/failure)

For simplicity, we can make the following assignment:

success → $X = 1$ and failure → $X = 0$

The population has made up of 0s and 1s, and if the true proportion of successes in the population is p, it's standard deviation is $\sqrt{p(1- \ p)}$

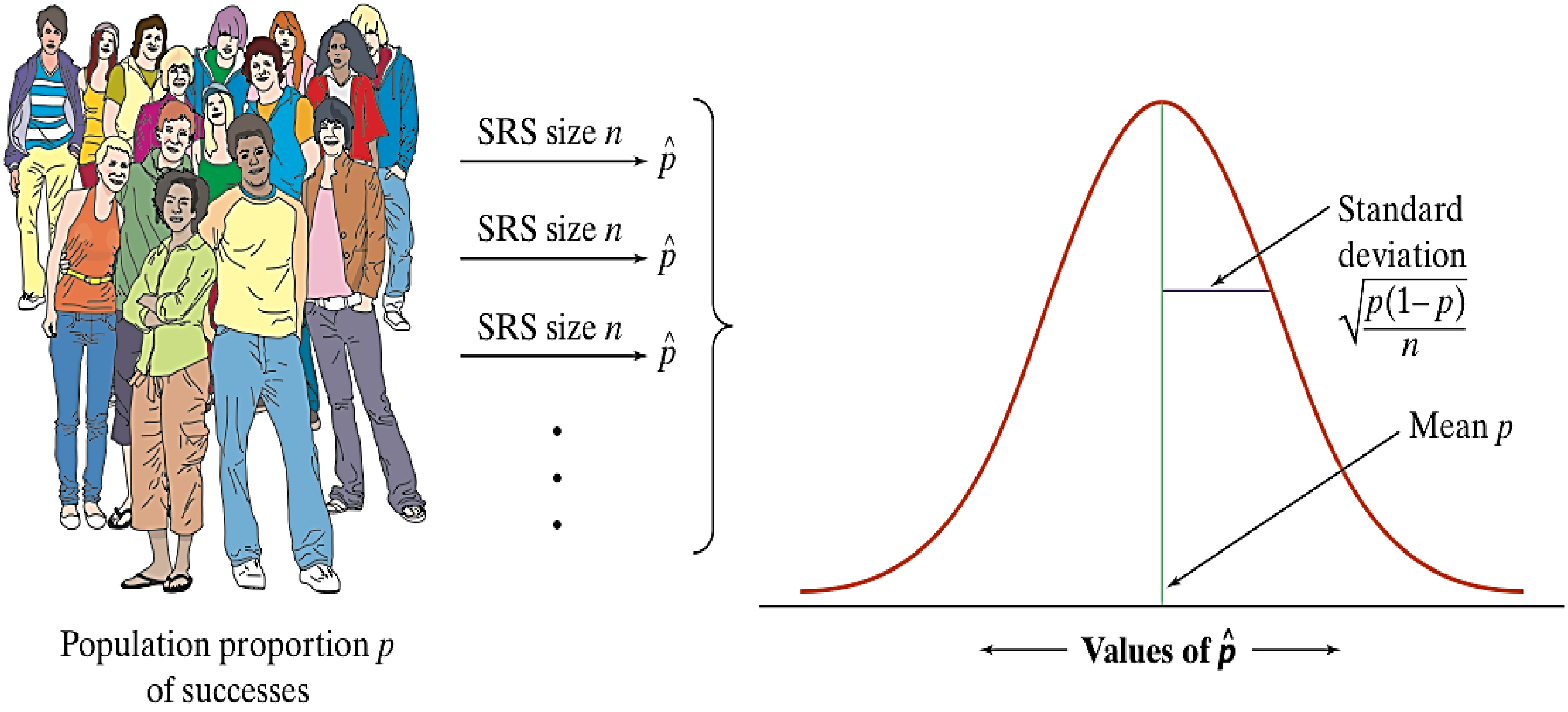
If we take a random sample of size n from this population, and average the observed values, we get the sample proportion, as in,

$$\hat{p} = \frac{\sum x}{n =} \bar{X}$$

Because the sample proportion is just a sample mean, its sampling distribution should have the same properties as the sample mean

Specifically, the CLT also applies of the sample proportion!

Sampling Distribution of a Sample Proportion



Sampling Distribution of a Sample Proportion

Sampling Distribution of the Sample Proportion

Choose an SRS of size *n* from a population with *p* as the true proportion of success → it follows that the population standard deviation is $\sqrt{p(1-p)}$

The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$

The standard deviation of the sampling distribution of $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$$

As *n* increases, the sampling distribution becomes **approximately Normal**.

For sufficiently large *n*:

$$\hat{p} \dot{\sim} N\left(p, \sqrt{p(1-p)/n}\right)$$

Chapter 5
Sampling Distributions

5.1 The Sampling Distribution of a Sample Mean

5.2 Sampling Distributions for Counts and Proportions