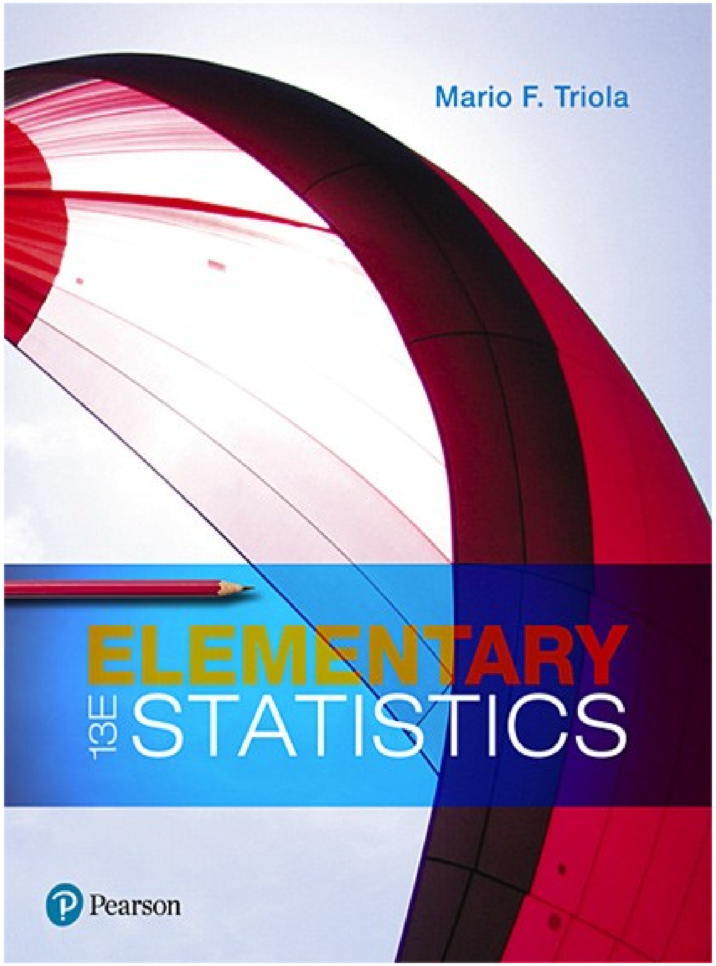
Elementary Statistics

Thirteenth Edition



Chapter 6

Normal Probability Distributions

Normal Probability Distributions

6-1 The Standard Normal Distribution

6-2 Real Applications of Normal Distributions

6-3 Sampling Distributions and Estimators

6-4 The Central Limit Theorem

6-5 Assessing Normality

6-6 Normal as Approximation to Binomial

Key Concept

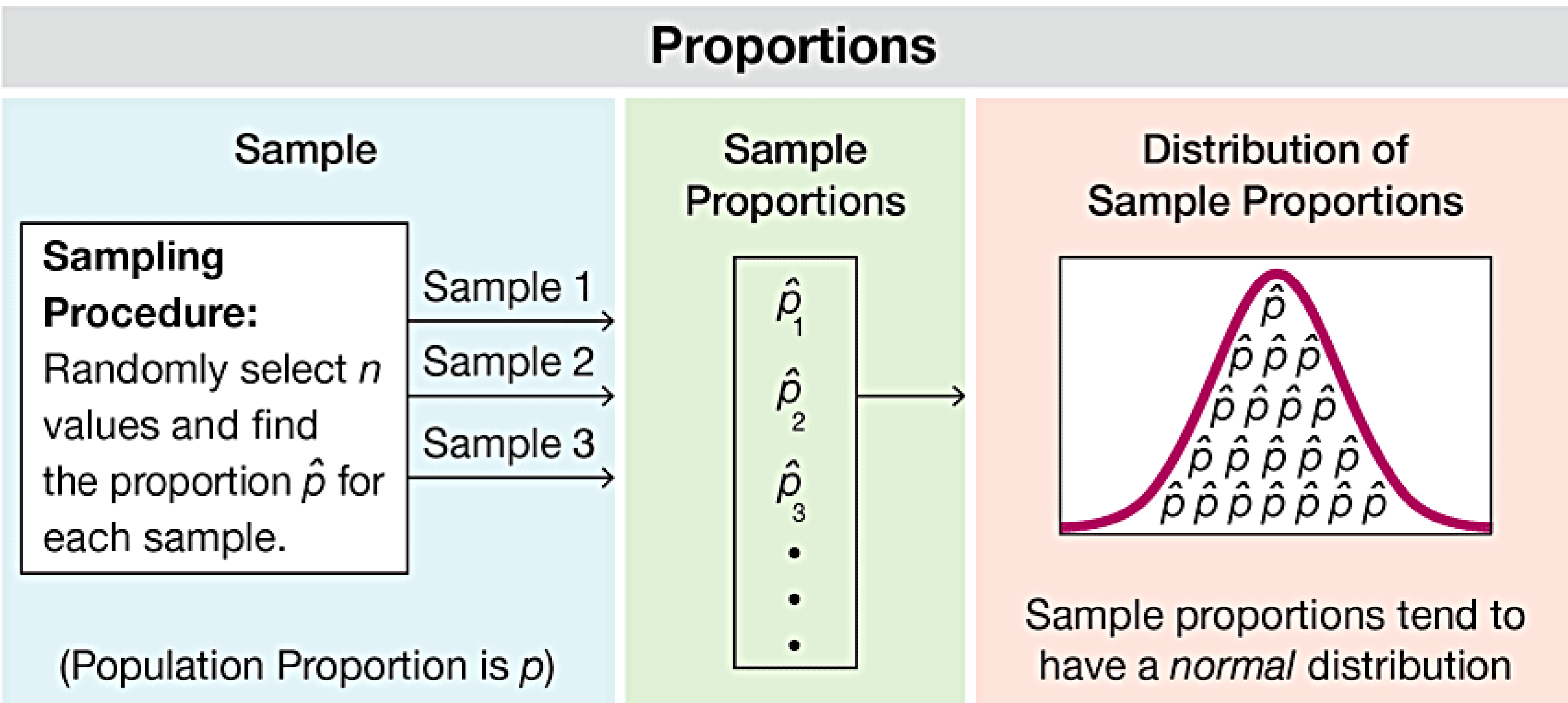
We now consider the concept of a **sampling distribution of a statistic.** Instead of working with values from the original population, we want to focus on the values of **statistics** (such as sample proportions or sample means) obtained from the population.

General Behavior of Sampling Distributions (1 of 4)

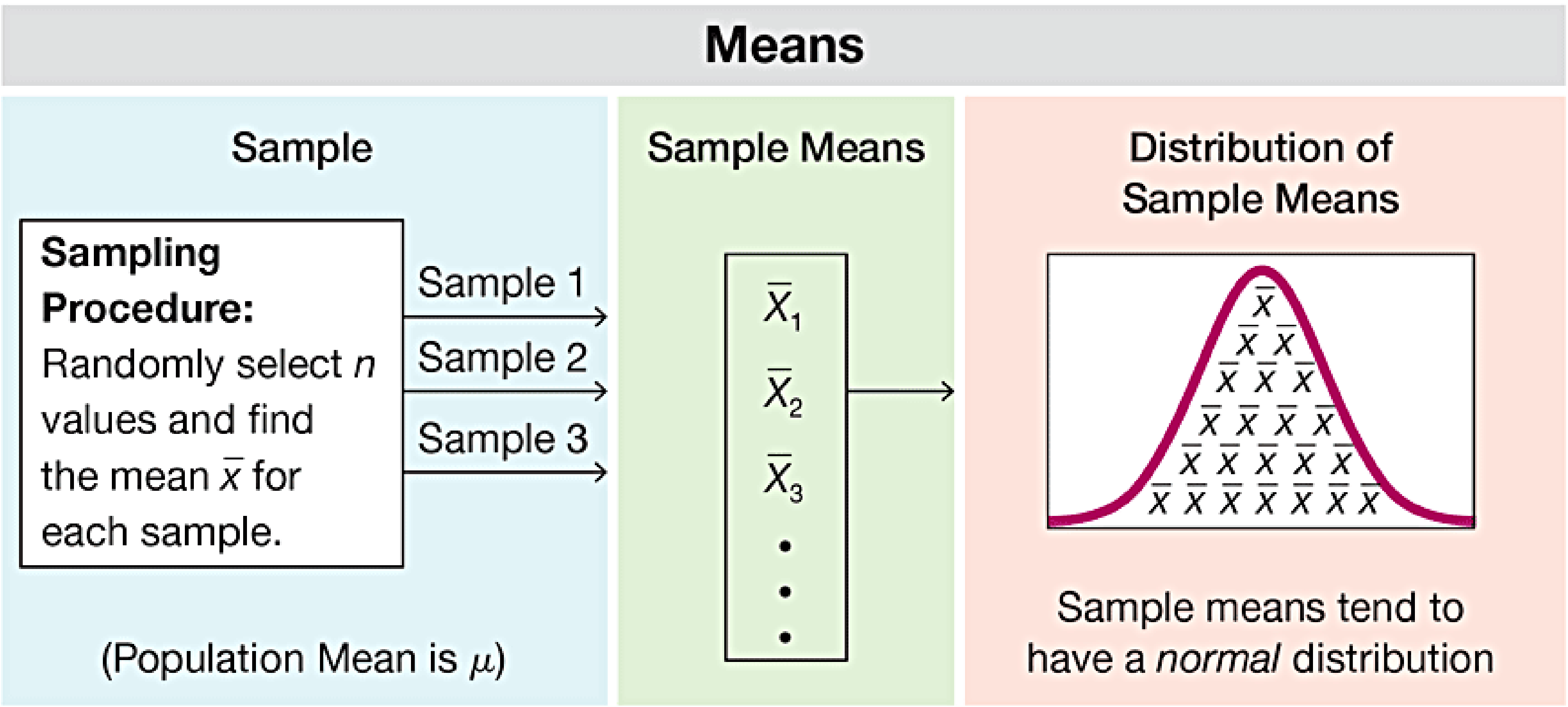
When samples of the same size are taken from the same population, the following two properties apply:

1. Sample proportions tend to be normally distributed.
2. The mean of sample proportions is the same as the population mean.

General Behavior of Sampling Distributions (2 of 4)

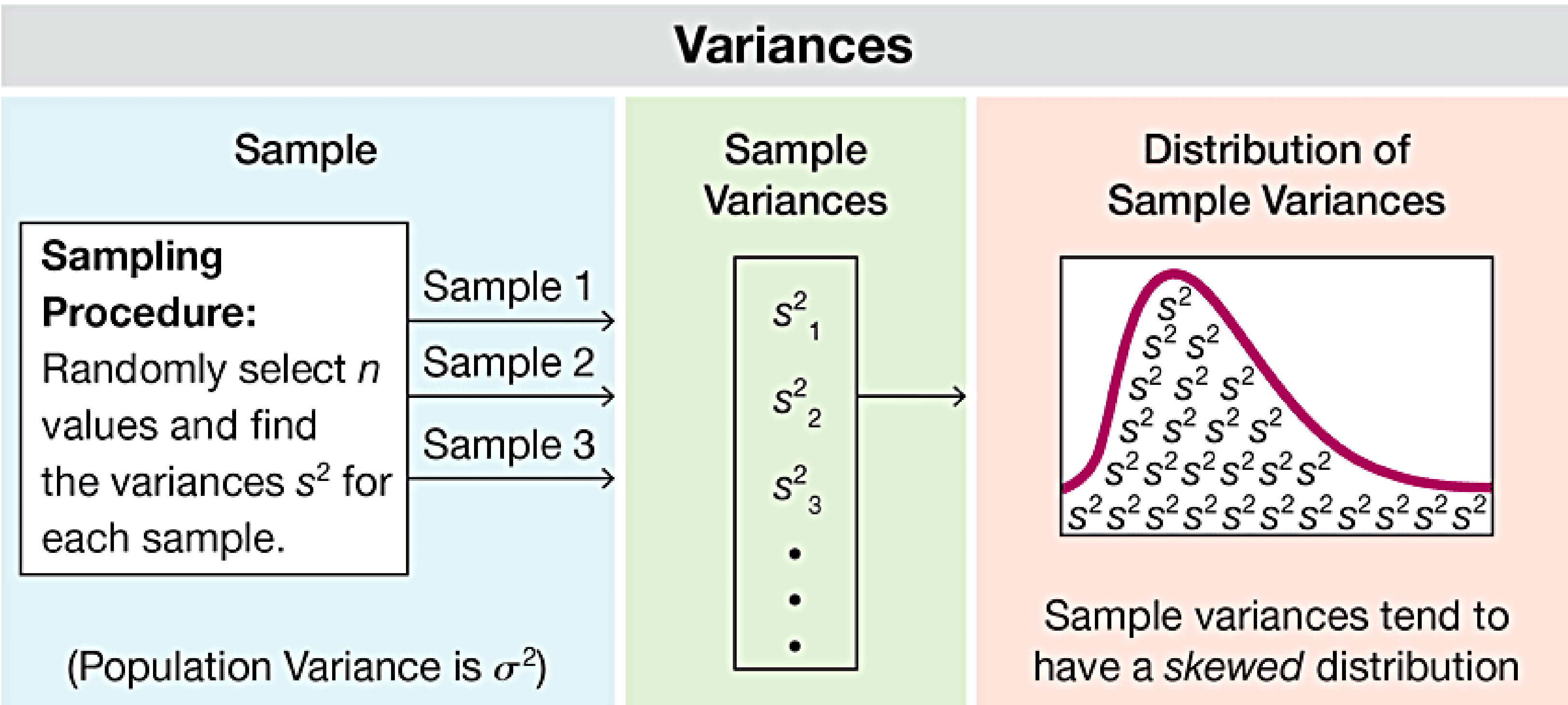


General Behavior of Sampling Distributions (3 of 4)



Pearson

General Behavior of Sampling Distributions (4 of 4)



Pearson

Sampling Distribution of a Statistic

- Sampling Distribution of a Statistic
  - The **sampling distribution of a statistic** (such as a sample proportion or sample mean) is the distribution of all values of the statistic when all possible samples of the same size *n* are taken from the same population. (The sampling distribution of a statistic is typically represented as a probability distribution in the format of a probability histogram, formula, or table.)

Sampling Distribution of the Sample Proportion

- Sampling Distribution of the Sample Proportion
  - The **sampling distribution of the sample proportion** is the distribution of sample proportions (or the distribution of the variable $\hat{p}$), with all samples having the same sample size n taken from the same population. (The sampling distribution of the sample proportion is typically represented as a probability distribution in the format of a probability histogram, formula, or table.)

Notations for Proportions

We need to distinguish between a population proportion p and some sample proportion:

p = **population** proportion

$\hat{p}$ = **sample** proportion

HINT $\hat{p}$ is pronounced "p-hat." When symbols are used above a letter, as in $\bar{x}$ and $\hat{p}$, they represent **statistics**, not parameters.

Behavior of Sample Proportions

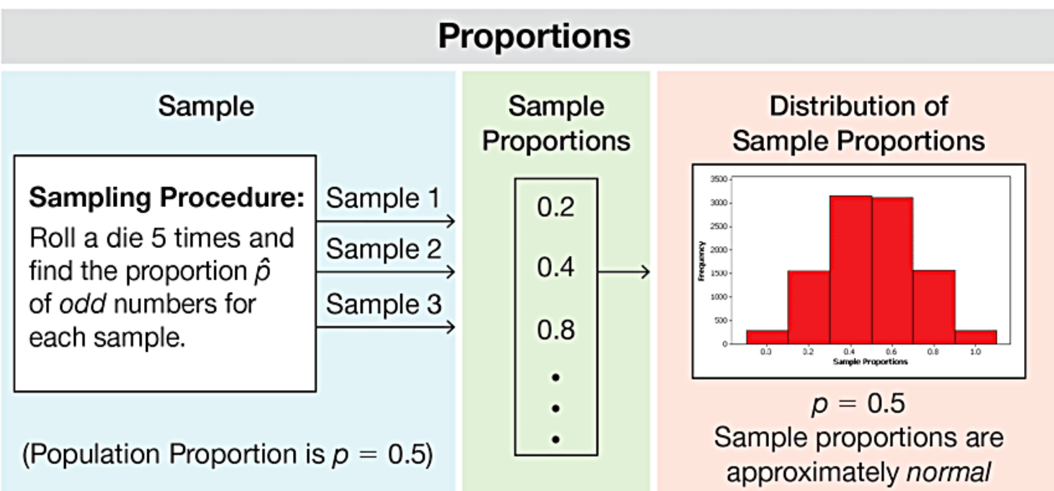
1. The distribution of sample proportions tends to approximate a normal distribution.
2. Sample proportions **target** the value of the population proportion in the sense that the mean of all of the sample proportions $\hat{p}$ is equal to the population proportion p; the expected value of the sample proportion is equal to the population proportion.

Example: Sampling Distributions of the Sample Proportion (1 of 3)

Consider repeating this process: Roll a die 5 times and find the proportion of **odd** numbers (1 or 3 or 5). What do we know about the behavior of all sample proportions that are generated as this process continues indefinitely?

Example: Sampling Distributions of the Sample Proportion (2 of 3)

The figure illustrates a process of rolling a die 5 times and finding the proportion of odd numbers. (The figure shows results from repeating this process 10,000 times, but the true sampling distribution of the sample proportion involves repeating the process indefinitely.)

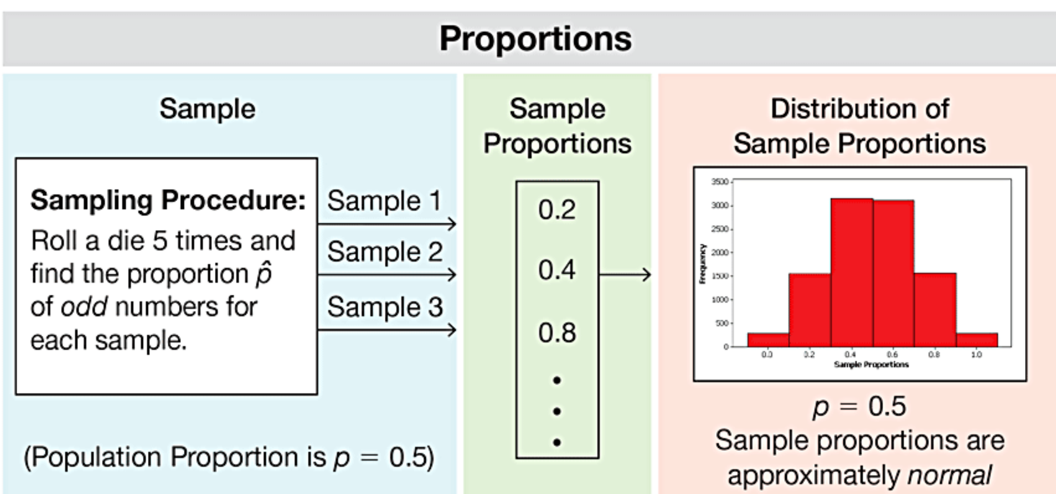


Sample Proportions from 10,000 Trials

Pearson

Example: Sampling Distributions of the Sample Proportion (3 of 3)

The figure shows that the sample proportions are approximately normally distributed. (Because the values of 1, 2, 3, 4, 5, 6 are all equally likely, the proportion of odd numbers in the population is 0.5, and the figure shows that the sample proportions have a mean of 0.50.)



Sample Proportions from 10,000 Trials

Pearson

Sampling Distribution of the Sample Mean

- Sampling Distribution of the Sample Mean
  - The **sampling distribution of the sample mean** is the distribution of all possible sample means (or the distribution of the variable $\bar{x}$), with all samples having the same sample size n taken from the same population. (The sampling distribution of the sample mean is typically represented as a probability distribution in the format of a probability histogram, formula, or table.)

Behavior of Sample Means

1. The distribution of sample means tends to be a normal distribution. (This will be discussed further in the following section, but the distribution tends to become closer to a normal distribution as the sample size increases.)
2. The sample means **target** the value of the population mean. (That is, the mean of the sample means is the population mean. The expected value of the sample mean is equal to the population mean.)

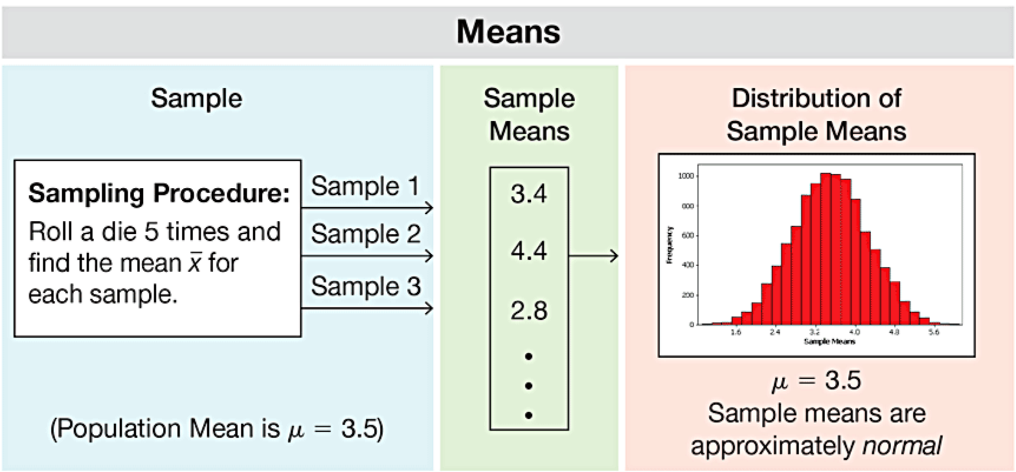
Example: Sampling Distribution of the Sample Mean (1 of 3)

Consider repeating this process: Roll a die 5 times to randomly select 5 values from the population {1, 2, 3, 4, 5, 6}, then find the mean $\bar{x}$ of the results.

What do we know about the behavior of all sample means that are generated as this process continues indefinitely?

Example: Sampling Distribution of the Sample Mean (2 of 3)

The figure illustrates a process of rolling a die 5 times and finding the mean of the results. The figure shows results from repeating this process 10,000 times, but the true sampling distribution of the mean involves repeating the process indefinitely.



Sample Means from 10,000 trials

Example: Sampling Distribution of the Sample Mean (3 of 3)

Because the values of 1, 2, 3, 4, 5, 6 are all equally likely, the population has a mean of $\mu = 3.5$. The 10,000 sample means included in the figure have a mean of 3.5. If the process is continued indefinitely, the mean of the sample means will be 3.5. Also, the figure shows that the distribution of the sample means is approximately a normal distribution.

Sampling Distribution of the Sample Variance

- Sampling Distribution of the Sample Variance
  - The **sampling distribution of the sample variance** is the distribution of sample variances (the variable s^2), with all samples having the same sample size *n* taken from the same population. (The sampling distribution of the sample variance is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

Population Standard Deviation and Population Variance

Population standard deviation:

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

Population variance:

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$$

Behavior of Sample Variances

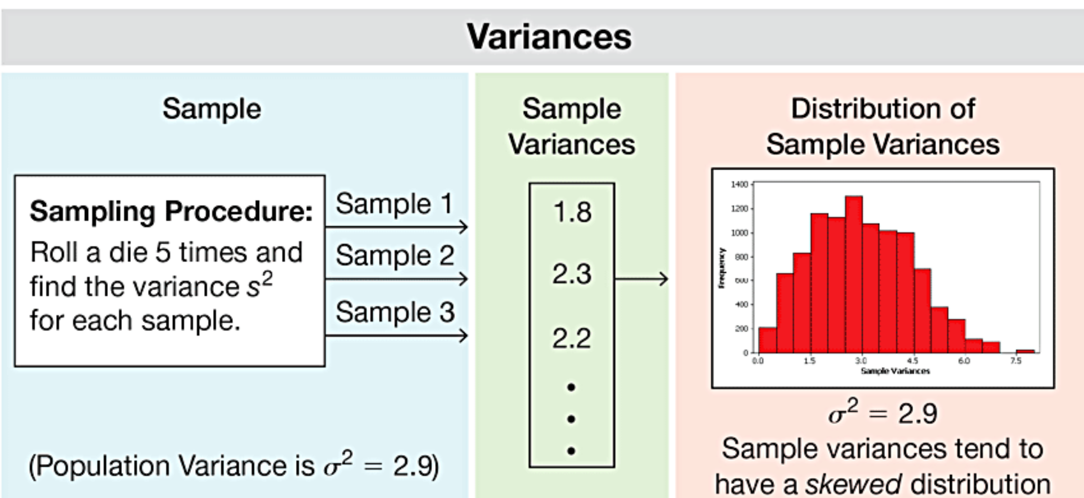
1. The distribution of sample variances tends to be a distribution skewed to the right.
2. The sample variances **target** the value of the population variance. (That is, the mean of the sample variances is the population variance. The expected value of the sample variance is equal to the population variance.)

Example: Sampling Distributions of the Sample Variances (1 of 4)

Consider repeating this process: Roll a die 5 times and find the variance s^2 of the results. What do we know about the behavior of all sample variances that are generated as this process continues indefinitely?

Example: Sampling Distributions of the Sample Variances (2 of 4)

The figure illustrates a process of rolling a die 5 times and finding the variance of the results. The figure shows results from repeating this process 10,000 times, but the true sampling distribution of the sample variance involves repeating the process indefinitely.

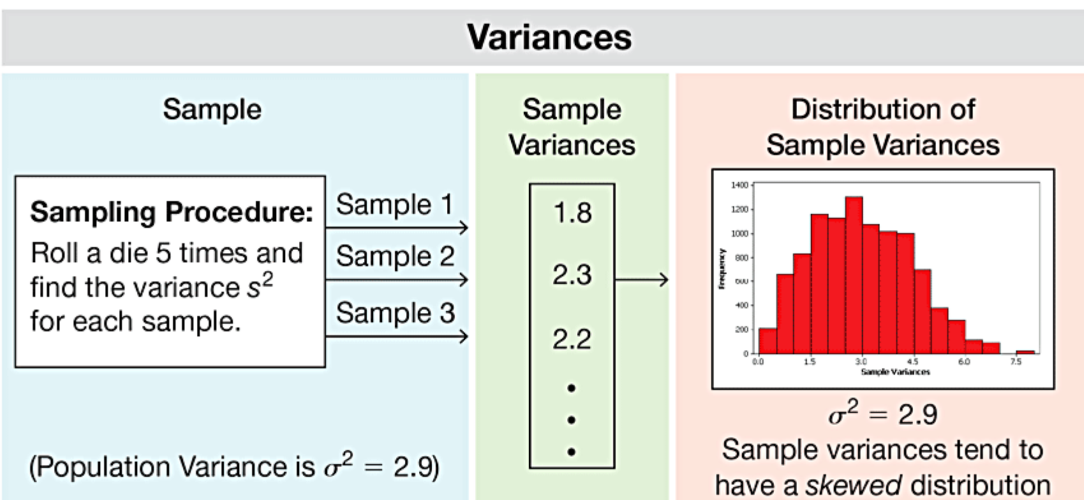


Sample Variances from 10,000 trials

Pearson

Example: Sampling Distributions of the Sample Variances (3 of 4)

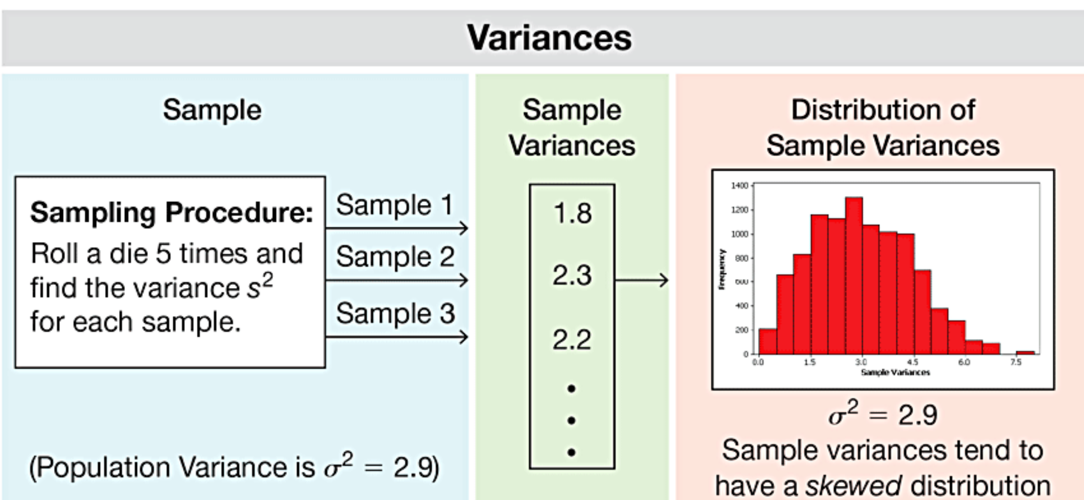
Because the values of 1, 2, 3, 4, 5, 6 are all equally likely, the population has a variance of s^2 = 2.9, and the 10,000 sample variances included in the figure have a mean of 2.9.



Sample Variances from 10,000 trials

Example: Sampling Distributions of the Sample Variances (4 of 4)

If the process is continued indefinitely, the mean of the sample variances will be 2.9. Also, the figure shows that the distribution of the sample variances is a skewed distribution, not a normal distribution with its characteristic bell shape.



Sample Variances from 10,000 trials

Pearson

Estimator

- Estimator
  - An **estimator** is a statistic used to infer (or estimate) the value of a population parameter.

Unbiased Estimator

- Unbiased Estimator
  - An **unbiased estimator** is a statistic that targets the value of the corresponding population parameter in the sense that the sampling distribution of the statistic has a mean that is equal to the corresponding population parameter.

Estimators: Unbiased and Biased (1 of 2)

Unbiased Estimator

These statistics are unbiased estimators. That is, they each target the value of the corresponding population parameter (with a sampling distribution having a mean equal to the population parameter):

- Proportion $\hat{p}$
- Mean $\bar{x}$
- Variance s^2

Estimators: Unbiased and Biased (2 of 2)

Biased Estimator

These statistics are biased estimators. That is, they do **not** target the value of the corresponding population parameter:

- Median
- Range
- Standard deviation s

Why Sample with Replacement?

Sampling is conducted with replacement because of these two very important reasons:

1. When selecting a relatively small sample from a large population, it makes no significant difference whether we sample with replacement or without replacement.
2. Sampling with replacement results in **independent** events that are unaffected by previous outcomes, and independent events are easier to analyze and result in simpler calculations and formulas.