Sampling Distributions and Estimators

Colby Community College

Example 1

Let us assume that among the population of all adults, exactly 70% do not feel comfortable in a self-driving car.

Example 1

Let us assume that among the population of all adults, exactly 70% do not feel comfortable in a self-driving car.

In a TE Connectivity survey of 1000 adults, 69% said they did not feel comfortable in a self-driving vehicle.

Example 1

Let us assume that among the population of all adults, exactly 70% do not feel comfortable in a self-driving car.

In a TE Connectivity survey of 1000 adults, 69% said they did not feel comfortable in a self-driving vehicle.

Empowered by visions of hordes of driverless cars, $50,000$ people became so enthusiastic that they each conducted their own survey of 1000 randomly selected adults on the same topic.

Example 1

Let us assume that among the population of all adults, exactly 70% do not feel comfortable in a self-driving car.

In a TE Connectivity survey of 1000 adults, 69% said they did not feel comfortable in a self-driving vehicle.

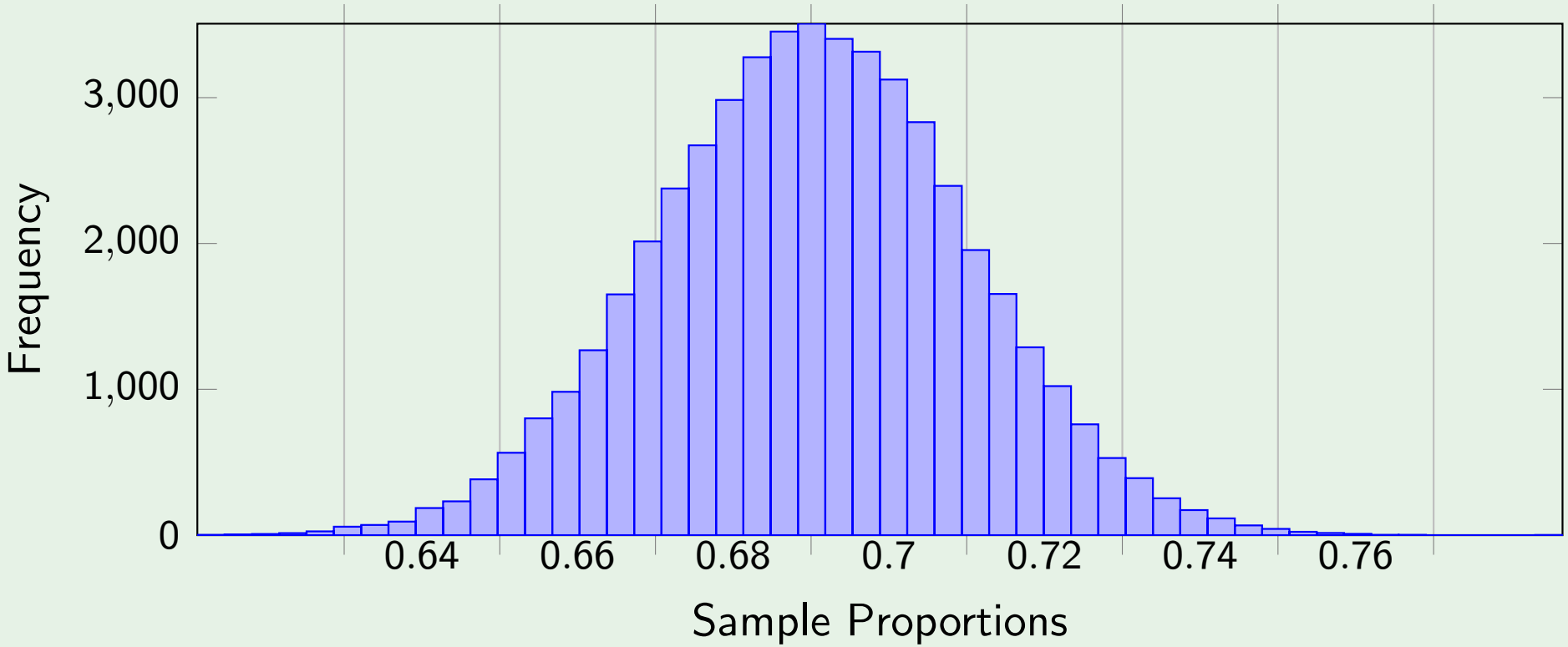
Empowered by visions of hordes of driverless cars, $50,000$ people became so enthusiastic that they each conducted their own survey of 1000 randomly selected adults on the same topic.

Each of these $50,000$ newbie surveyors reported the percentage that they found, with results such as 68%, 72%, 70%, etc. . .

Example 1

Let us assume that among the population of all adults, exactly 70% do not feel comfortable in a self-driving car.

Lets assume we gather all $50,000$ of the percentages and convert them to proportions. We can then construct the histogram:



Note

When samples of the same size are taken from the same population:

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

Note

When samples of the same size are taken from the same population:

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

Definition

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.

Note

When samples of the same size are taken from the same population:

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

Definition

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.

Note

The sampling distribution of a statistic is typically represented as a probability distribution in the format of a probability histogram, formula, or table.

Definition

The **sampling distribution of the sample proportion** is the distribution of sample proportions, with all samples having the same sample size n taken from the same population.

Definition

The **sampling distribution of the sample proportion** is the distribution of sample proportions, with all samples having the same sample size n taken from the same population.

Note

The sampling distribution of the sample proportion is typically represented as a probability distribution in the format of a probability histogram, formula or table.

Definition

The **sampling distribution of the sample proportion** is the distribution of sample proportions, with all samples having the same sample size n taken from the same population.

Note

The sampling distribution of the sample proportion is typically represented as a probability distribution in the format of a probability histogram, formula or table.

Notation

$$\begin{aligned} p &= \text{population proportion} \\ \hat{p} &= \text{sample proportion} \end{aligned}$$

Behavior of Sample Proportions

1. The distribution of sample proportions tends to approximate a normal distribution.

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 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.

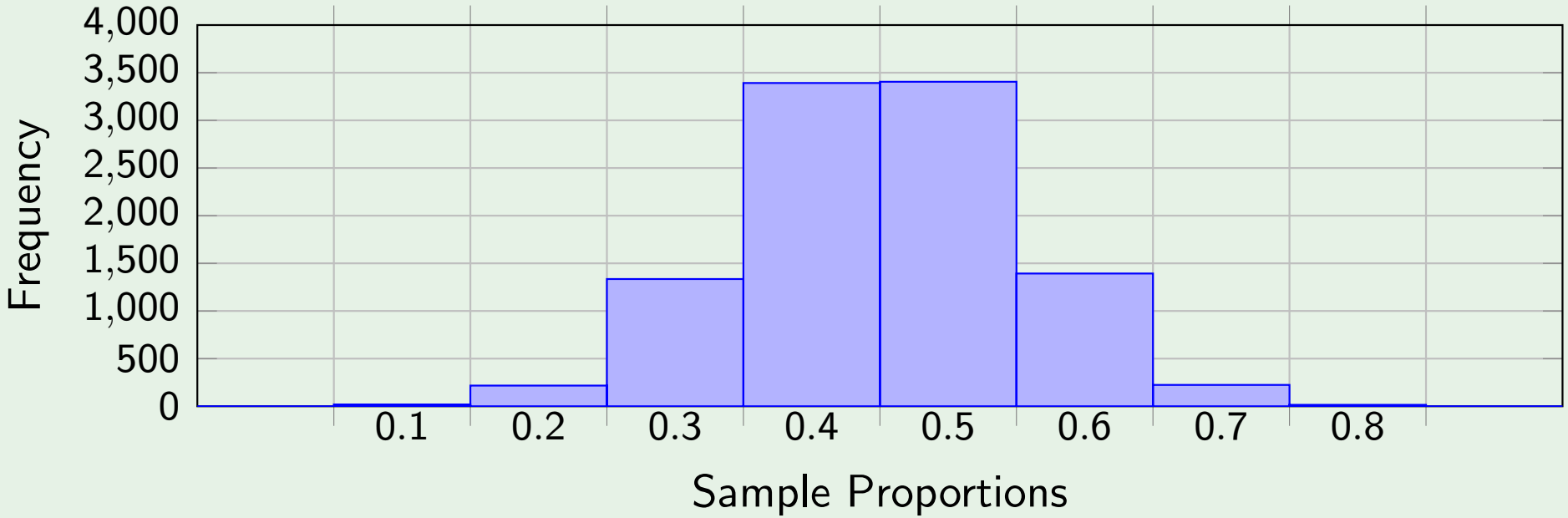
Example 2

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Find the proportion of odd numbers.

What can we say about the behavior of all sample proportions that are generated as this process continues indefinitely?

Repeating this process $10,000$ times gives the following distribution of sample proportions.



Definition

The **sampling distribution of the sample mean** is the distribution of all possible sample means, with all samples having the same sample size n taken from the same population.

Definition

The **sampling distribution of the sample mean** is the distribution of all possible sample means, with all samples having the same sample size n taken from the same population.

Note

The sampling distribution of the sample mean is typically represented as a probability distribution in the format of a probability histogram, formula or table.

Definition

The **sampling distribution of the sample mean** is the distribution of all possible sample means, with all samples having the same sample size n taken from the same population.

Note

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 approximate a normal distribution.

Definition

The **sampling distribution of the sample mean** is the distribution of all possible sample means, with all samples having the same sample size n taken from the same population.

Note

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 approximate a normal distribution.
2. Sample means target the value of the population mean in the sense that the mean of all the sample means is equal to the population mean. The expected value of the sample mean is equal to the population mean.

Example 3

A friend has three children with ages 4, 5, and 9. Let's consider the population consisting of $\{4, 5, 9\}$.

Example 3

A friend has three children with ages 4, 5, and 9. Let's consider the population consisting of $\{4, 5, 9\}$.

If two ages are randomly selected with replacement from the population, let's consider the sampling distribution of the sample mean by creating a table representing the probability distribution of the sample mean.

Example 3

A friend has three children with ages 4, 5, and 9. Let's consider the population consisting of $\{4, 5, 9\}$.

If two ages are randomly selected with replacement from the population, let's consider the sampling distribution of the sample mean by creating a table representing the probability distribution of the sample mean.

Sample	$\bar{x}$	**Probability**
4,4	4.0	1/9
4,5	4.5	1/9
4,9	6.5	1/9
5,4	4.5	1/9
5,5	5.0	1/9
5,9	7.0	1/9
9,4	6.5	1/9
9,5	7.0	1/9
9,9	9.0	1/9

Example 3

A friend has three children with ages 4, 5, and 9. Let's consider the population consisting of $\{4, 5, 9\}$.

If two ages are randomly selected with replacement from the population, let's consider the sampling distribution of the sample mean by creating a table representing the probability distribution of the sample mean.

Condensing the sample means gives:

$\bar{x}$	Probability	$\bar{x}$	Probability
4.0	1/9	6.5	2/9
4.5	2/9	7.0	2/9
5.0	1/9	9.0	1/9

Example 3

A friend has three children with ages 4, 5, and 9. Let's consider the population consisting of $\{4, 5, 9\}$.

If two ages are randomly selected with replacement from the population, let's consider the sampling distribution of the sample mean by creating a table representing the probability distribution of the sample mean.

Condensing the sample means gives:

$\bar{x}$	Probability	$\bar{x}$	Probability
4.0	1/9	6.5	2/9
4.5	2/9	7.0	2/9
5.0	1/9	9.0	1/9

Since the table has all possible values of the sample mean along with their corresponding probabilities, this is an example of a sampling distribution of a sample mean.

Example 3

A friend has three children with ages 4, 5, and 9. Let's consider the population consisting of $\{4, 5, 9\}$.

If two ages are randomly selected with replacement from the population, let's consider the sampling distribution of the sample mean by creating a table representing the probability distribution of the sample mean.

Condensing the sample means gives:

$\bar{x}$	Probability	$\bar{x}$	Probability
4.0	1/9	6.5	2/9
4.5	2/9	7.0	2/9
5.0	1/9	9.0	1/9

Since the table has all possible values of the sample mean along with their corresponding probabilities, this is an example of a sampling distribution of a sample mean.

The mean of the population $\{4, 5, 9\}$ is $\mu = 0.6$. Using either version of the table, we find that the mean of the sample values is 6.0.

Example 4

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Select 5 values from the population, then find the mean $\bar{x}$ of the results.

Example 4

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Select 5 values from the population, then find the mean $\bar{x}$ of the results.

Let us consider the behavior of all sample means that are generated as this process continues indefinitely.

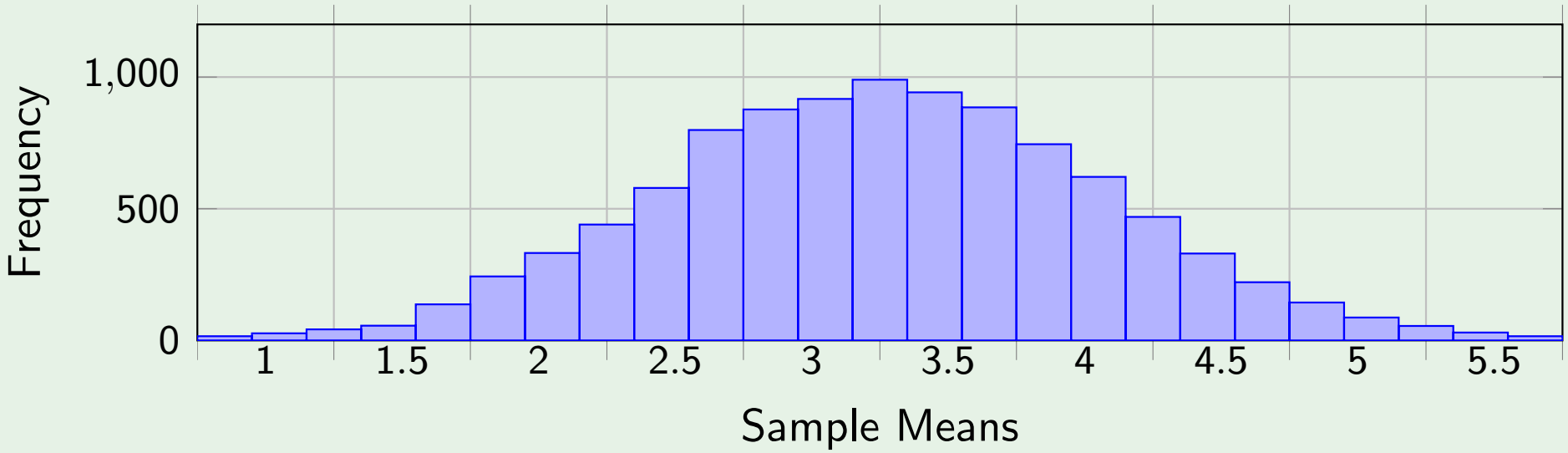
Example 4

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Select 5 values from the population, then find the mean $\bar{x}$ of the results.

Let us consider the behavior of all sample means that are generated as this process continues indefinitely.

Repeating this process $10,000$ times gives:



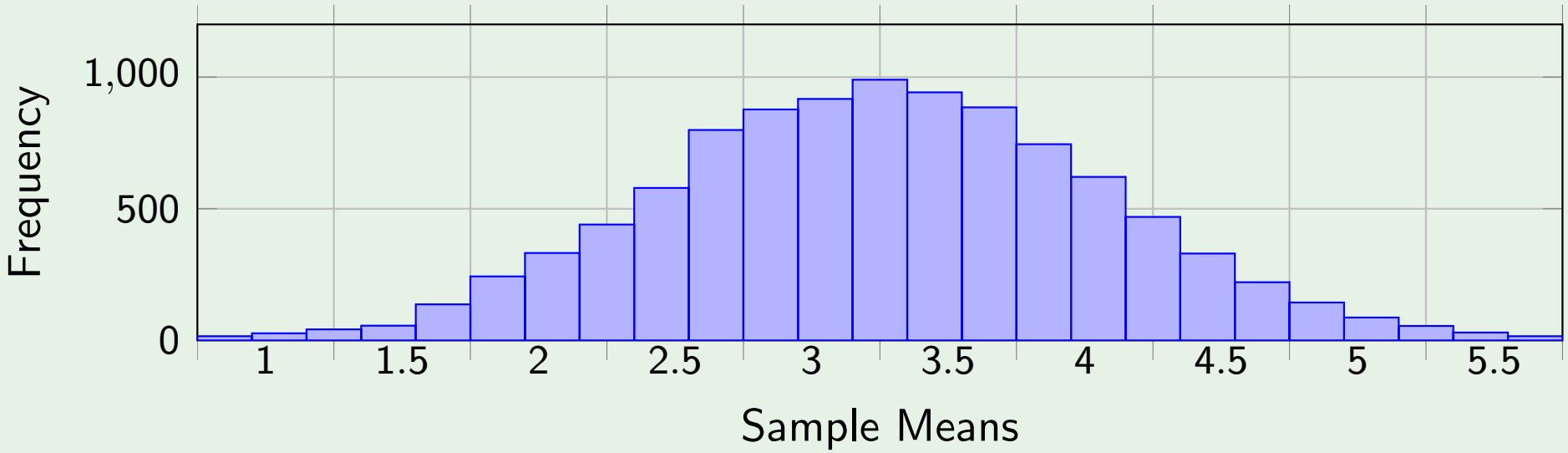
Example 4

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Select 5 values from the population, then find the mean $\bar{x}$ of the results.

Let us consider the behavior of all sample means that are generated as this process continues indefinitely.

Repeating this process $10,000$ times gives:



If the process continues, the mean of the sample means will be 3.5.

Definition

The **sampling distribution of the sample variance** is the distribution of sample variances, with all samples having the same sample size n taken from the same population.

Definition

The **sampling distribution of the sample variance** is the distribution of sample variances, with all samples having the same sample size n taken from the same population.

Note

The sampling distribution of the sample variance is typically represented as a probability distribution in the format of a table, probability histogram, or formula.

Definition

The **sampling distribution of the sample variance** is the distribution of sample variances, with all samples having the same sample size n taken from the same population.

Note

The sampling distribution of the sample variance is typically represented as a probability distribution in the format of a table, probability histogram, or formula.

Reminder

When working with population you divide by the population size N:

$$\text{Population standard deviation:} \quad \sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

$$\text{Population Variance:} \quad \sigma^2 = \frac{\sum (x-\mu)^2}{N}$$

Behavior of Sample Variances

1. The distribution of sample variances tend to be a distribution skewed to the right.

Behavior of Sample Variances

1. The distribution of sample variances tend to be a distribution skewed to the right.
2. The sample variances target the value of the population variance in the sense that mean of the sample variances is the population variance. The expected value of the sample variance is equal to the population variance.

Example 5

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Find the variance of the results.

Example 5

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Find the variance of the results.

Let us consider the behavior of all sample variances that are generated as this process continues indefinitely.

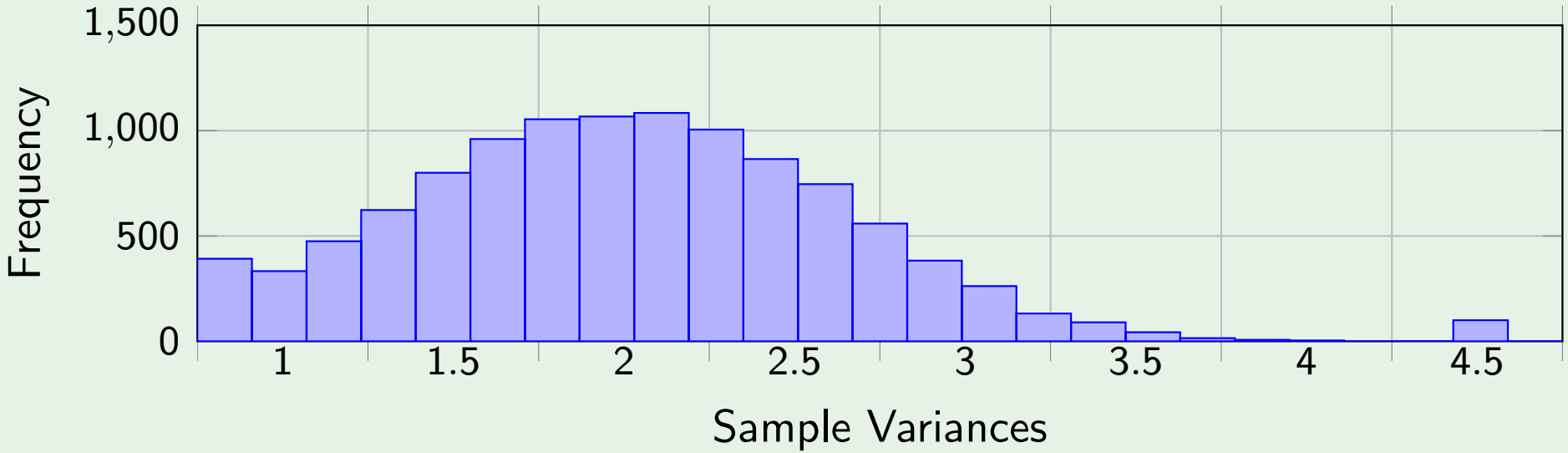
Example 5

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Find the variance of the results.

Let us consider the behavior of all sample variances that are generated as this process continues indefinitely.

Repeating this process $10,000$ times gives:



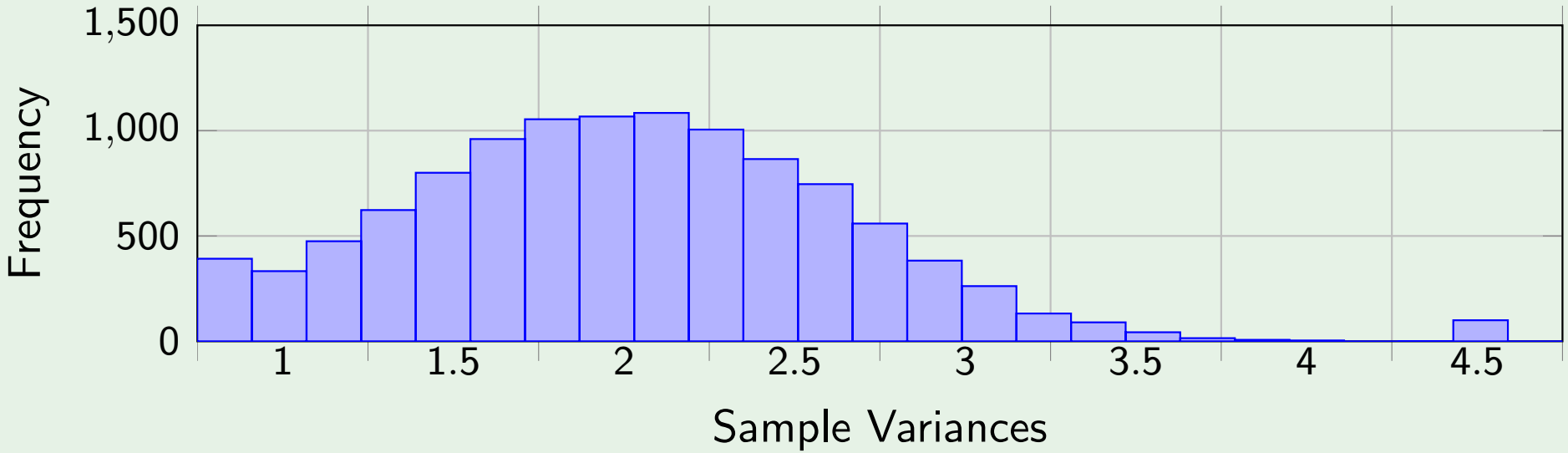
Example 5

Consider repeating the following process:

- Roll a six-sided die 5 times.
- Find the variance of the results.

Let us consider the behavior of all sample variances that are generated as this process continues indefinitely.

Repeating this process $10,000$ times gives:



The histogram shows that the distribution of the sample variances is a skewed distribution, not a normal distribution.

Definition

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

Definition

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

Definition

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.

Definition

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

Definition

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.

Definition

An **biased estimator** is a statistic that is not an unbiased estimator.

Unbiased Estimators

These statistics target the value of the corresponding population parameters.

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

Unbiased Estimators

These statistics target the value of the corresponding population parameters.

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

Biased Estimators

These statistics do not target the value of the corresponding population parameters.

- Median
- Range
- Standard deviation s

Why Sample with Replacement?

All of the examples we have looked at sample with replacement. Why did we do this?

Why Sample with Replacement?

All of the examples we have looked at sample with replacement. Why did we do this?

Sampling without replacement would have the very practical advantage of avoiding wasteful duplication whenever the same item is selected more than once.

Why Sample with Replacement?

All of the examples we have looked at sample with replacement. Why did we do this?

Sampling without replacement would have the very practical advantage of avoiding wasteful duplication whenever the same item is selected more than once.

Many statistical procedures are based on the assumption that sampling is conducted with replacement for the two important reasons:

Why Sample with Replacement?

All of the examples we have looked at sample with replacement. Why did we do this?

Sampling without replacement would have the very practical advantage of avoiding wasteful duplication whenever the same item is selected more than once.

Many statistical procedures are based on the assumption that sampling is conducted with replacement for the two 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.

Why Sample with Replacement?

All of the examples we have looked at sample with replacement. Why did we do this?

Sampling without replacement would have the very practical advantage of avoiding wasteful duplication whenever the same item is selected more than once.

Many statistical procedures are based on the assumption that sampling is conducted with replacement for the two 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 calculation and formulas.