# Chapter 6 Normal Probability Distributions
# Sampling Distributions and Estimators (6.3)

## Objectives

**In this lesson, we will learn how to**

- explain what a sampling distribution is;
- describe the sampling distribution of a sample proportion;
- describe the sampling distribution of a sample mean;
- describe the sampling distribution of a sample variance;
- distinguish between biased and unbiased estimators;
- explain why sampling with replacement is often used in theory.

## 1 Big Idea of Sampling Distributions

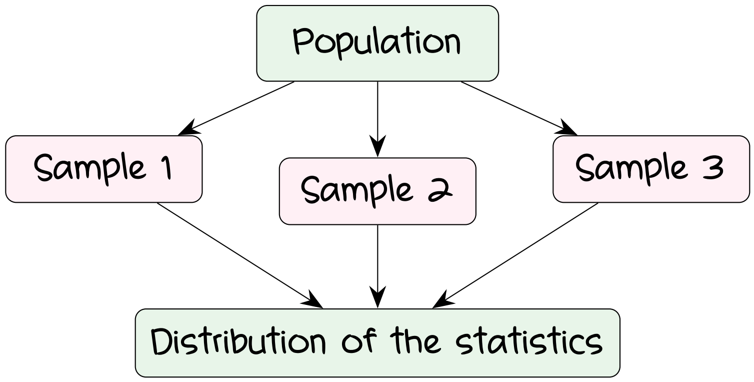
In earlier sections, we studied values from an original population. Now we study values of statistics computed from samples.

For example, instead of focusing on individual values $x$, we may focus on statistics such as

$$\hat{p}, \qquad \bar{x}, \qquad s^2.$$

**Key Concept**

A sampling distribution describes how a statistic behaves when we repeatedly take samples of the same size from the same population.




## 2 Sampling Distribution of a Statistic

**Sampling Distribution of a Statistic**

The sampling distribution of a statistic is the distribution of all values of the statistic when all possible samples of the same size $n$ are taken from the same population.

A sampling distribution is often shown as a probability distribution, histogram, table, or formula.

**Important Idea**

A sampling distribution is not the same as the distribution of the original data values.

- Original distribution: values such as $x$.
- Sampling distribution: statistics such as $\hat{p}$, $\bar{x}$, or $s^2$.

## 3 Notation for Sample Proportions

**Proportion Notation**

| | |
|---|---|
| $x$ | number of successes |
| $n$ | sample size |
| $N$ | population size |
| $\hat{p} = \dfrac{x}{n}$ | sample proportion, pronounced ``p-hat'' |
| $p$ | population proportion |

Hint: Symbols with a mark above the letter, such as $\hat{p}$ and $\bar{x}$, usually represent sample statistics, not population parameters.

## 4 Sampling Distribution of the Sample Proportion

**Sampling Distribution of $\hat{p}$**

The sampling distribution of the sample proportion is the distribution of all possible values of $\hat{p}$ when all samples have the same sample size $n$ and are taken from the same population.

**Behavior of Sample Proportions**

- The distribution of sample proportions tends to be approximately normal.
- The mean of all sample proportions is equal to the population proportion:
$$\mu_{\hat{p}} = p.$$
- In this sense, $\hat{p}$ targets the population proportion $p$.

Example 1: Sample Proportions from Rolling a Die: Consider repeating this process: roll a fair die 5 times and find the proportion of odd numbers. What do we know about the behavior of all sample proportions generated by repeating this process many times?

Solution

For a fair die, the odd numbers are

$$1, 3, 5.$$

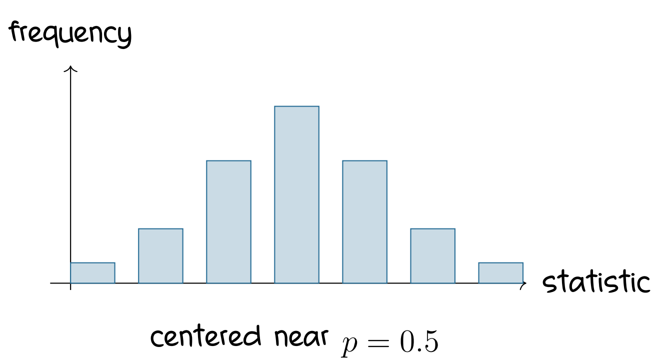
There are 3 odd numbers out of 6 possible outcomes, so the population proportion of odd numbers is

$$p = \frac{3}{6} = 0.5.$$

If we repeatedly roll the die 5 times and compute $\hat{p}$ each time, the values of $\hat{p}$ form a sampling distribution.
The sample proportions tend to be centered around

$$\mu_{\hat{p}} = p = 0.5.$$

They also tend to form an approximately normal distribution as the process is repeated many times.



# 5 Sampling Distribution of the Sample Mean

### Sampling Distribution of $\bar{x}$

The sampling distribution of the sample mean is the distribution of all possible sample means $\bar{x}$, with all samples having the same sample size $n$ and taken from the same population.

### Behavior of Sample Means

- The distribution of sample means tends to be normal.
- The distribution becomes closer to normal as the sample size increases.
- The mean of all sample means is equal to the population mean:

$$\mu_{\bar{x}} = \mu.$$

- Therefore, $\bar{x}$ targets the population mean $\mu$.

Example 2: Sample Means from Rolling a Die: Consider repeating this process: roll a fair die 5 times to randomly select 5 values from the population

$$\{1, 2, 3, 4, 5, 6\},$$

then find the sample mean $\bar{x}$ of the results. What do we know about the behavior of all sample means?

### Solution

For a fair die, the population mean is

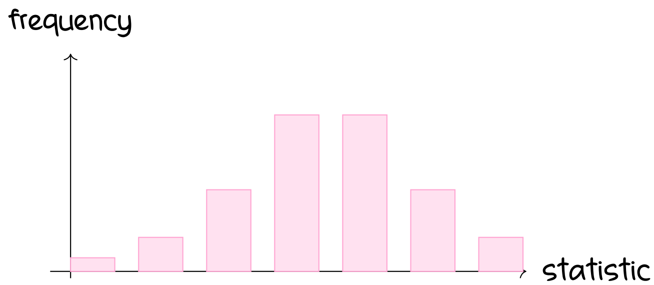
$$\mu = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5.$$

If we repeatedly roll the die 5 times and calculate $\bar{x}$ each time, the sample means form a sampling distribution.

The mean of all sample means is

$$\mu_{\bar{x}} = \mu = 3.5.$$

Thus, sample means target the population mean. The distribution of sample means is also approximately normal.




centered near $\mu = 3.5$

# 6 Sampling Distribution of the Sample Variance

**Sampling Distribution of $s^2$**

The sampling distribution of the sample variance is the distribution of all possible sample variances $s^2$, with all samples having the same sample size $n$ and taken from the same population.

**Population Variance and Standard Deviation**

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

### Behavior of Sample Variances

- The distribution of sample variances tends to be skewed to the right.
- The sample variances target the value of the population variance:

$$\mu_{s^2} = \sigma^2.$$

- Therefore, $s^2$ is useful for estimating $\sigma^2$.

Example 3: Sample Variances from Rolling a Die: Consider repeating this process: roll a fair die 5 times and find the sample variance $s^2$ of the results. What do we know about the behavior of all sample variances?

### Solution

For the population

$$\{1, 2, 3, 4, 5, 6\},$$

the population mean is

$$\mu = 3.5.$$

The population variance is

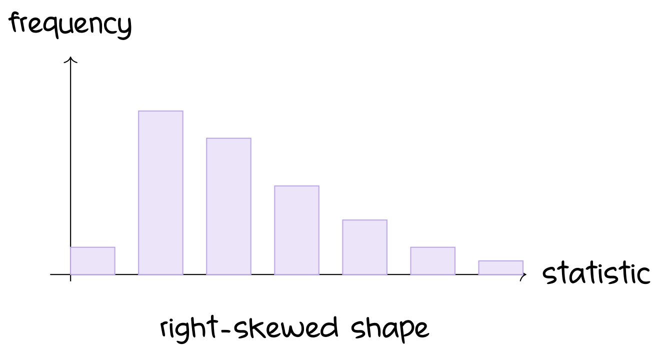
$$\sigma^2 = \frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6}.$$

$$\sigma^2 = \frac{6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25}{6} = \frac{17.5}{6} \approx 2.9.$$

If the process is continued indefinitely, the mean of all sample variances will be approximately

$$\sigma^2 \approx 2.9.$$

Unlike sample proportions and sample means, the distribution of sample variances is usually skewed to the right, not bell-shaped.




# 7 Estimators

**Estimator**

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

Examples:

$$\hat{p} \text{ estimates } p, \qquad \bar{x} \text{ estimates } \mu, \qquad s^2 \text{ estimates } \sigma^2.$$

**Unbiased Estimator**

An unbiased estimator is a statistic that targets the corresponding population parameter. This means the mean of the sampling distribution of the statistic is equal to the population parameter.

# 8 Unbiased and Biased Estimators

**Unbiased Estimators**

The following are unbiased estimators:

| Statistic | Parameter Estimated |
|---|---|
| $\hat{p}$ | $p$ |
| $\bar{x}$ | $\mu$ |
| $s^2$ | $\sigma^2$ |

### Biased Estimators

The following are biased estimators:

- median;
- range;
- sample standard deviation $s$.

### Important Note

Although $s$ is a biased estimator of $\sigma$, the bias is small for large samples. Therefore, $s$ is commonly used to estimate $\sigma$ in practice.

Example 4: Identifying Estimators: For each statistic, identify the parameter it is commonly used to estimate.

(a) $\hat{p}$

(b) $\bar{x}$

(c) $s^2$

(d) $s$

### Solution

(a) $\hat{p}$ estimates the population proportion $p$.

(b) $\bar{x}$ estimates the population mean $\mu$.

(c) $s^2$ estimates the population variance $\sigma^2$.

(d) $s$ estimates the population standard deviation $\sigma$.

The first three are unbiased estimators. The sample standard deviation $s$ is biased, but it is still widely used for estimating $\sigma$.

## 9 Why Sample with Replacement?

In theoretical sampling distributions, sampling is often described as being done with replacement.

Why Sampling with Replacement Is Helpful

- If the sample is small relative to a large population, sampling with or without replacement makes little practical difference.
- Sampling with replacement makes the selections independent.
- Independent events are easier to analyze and lead to simpler formulas.

Example 5: With Replacement or Without Replacement?: Suppose a population has 1,000,000 adults, and we randomly select 100 adults. Explain why sampling with replacement and sampling without replacement give very similar results.

Solution

The sample size is small compared with the population size:

$$\frac{100}{1,000,000} = 0.0001 = 0.01\%.$$

Because the sample is such a tiny part of the population, removing one person does not noticeably change the probabilities for the next selection.
Therefore, sampling without replacement behaves almost the same as sampling with replacement. In theory, using replacement makes the

events independent and the calculations much easier.

## Summary

**Key Ideas**

- A sampling distribution is the distribution of a statistic from all possible samples of the same size.
- Sample proportions $\hat{p}$ tend to be approximately normal and target $p$.
- Sample means $\bar{x}$ tend to be approximately normal and target $\mu$.
- Sample variances $s^2$ tend to be skewed right and target $\sigma^2$.
- $\hat{p}$, $\bar{x}$, and $s^2$ are unbiased estimators.
- The median, range, and sample standard deviation $s$ are biased estimators.
- Sampling with replacement makes selections independent and simplifies theory.