# Sampling Distributions (Page 285-292, Chapter 11)

**TODAY YOU WILL BE ABLE TO…**

* Define and identify parameters and statistics
* Describe the law of large numbers
* Define and describe sampling distributions
* Describe the sampling distribution of sample means
* Describe and apply the central limit theorem

**RECALL**

* **Simple Random Sample (SRS) of Size n** – Consists of **n** individuals from the population chosen in such a way that every possible combination of **n** individuals has an equal chance of being the sample actually selected
* **Population Distribution** – Describes how individuals vary in the population; if we measure the mean, **μ**, or standard deviation, **σ**, these values encompass all individuals in the population
* **Distribution of Sample Data** – Describes how individuals vary in the sample; if we measure the mean, , or standard deviation, **sx**, these values encompass only individuals in the sample

**TERMINOLOGY AND NOTATION**

A **parameter** is a number that describes some characteristic of the population. In statistical practice, the value of a parameter is not known because we cannot examine the entire population. Remember **P**opulation **P**arameter.

μ represents the *mean* of a ***population***

σrepresents the *standard deviation* of a ***population***

A **statistic** is a number that describes some characteristic of a sample. The value of a statistic can be computed directly from the sample data. We often use a statistic to estimate an unknown parameter. Remember **S**ample **S**tatistic.

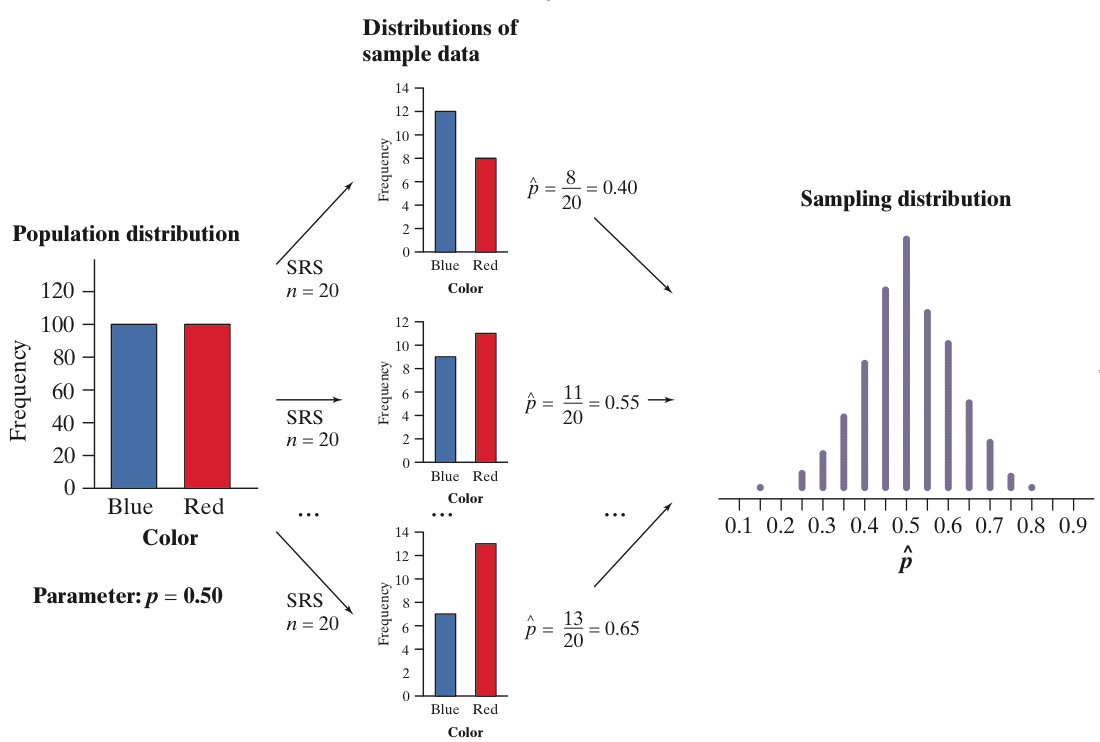
represents the *mean* of a set of ***sample*** values

sx represents the *standard deviation* of the set of ***sample*** values of x

**SAMPLING VARIABIITY**

The value of a statistic varies in repeated random sampling.

How can be an accurate estimate of **μ** if different samples produce different values of ?



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

**ACTIVITY**

Consider a small population of test grades given below for 10 students.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| **Score** | 74 | 83 | 71 | 77 | 78 | 71 | 75 | 80 | 74 | 69 |

PopulationDistribution

= 75.2

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| **Score** | 74 | 83 | 71 | 77 | 78 | 71 | 75 | 80 | 74 | 69 |

Range of the population: 69 to 83

Range of sample averages: 71.25 to 79.5

Select 10 random samples from this population with sample size of n=4.

Calculate the average test score of each sample and list below:

Sample average #1 =

Sample average #2 =

Sample average #3 =

Sample average #4 =

Sample average #5 =

Sample average #6 =

Sample average #7 =

Sample average #8 =

Sample average #9 =

Sample average #10 =

=

**LAW OF LARGE NUMBERS**

Draw 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.

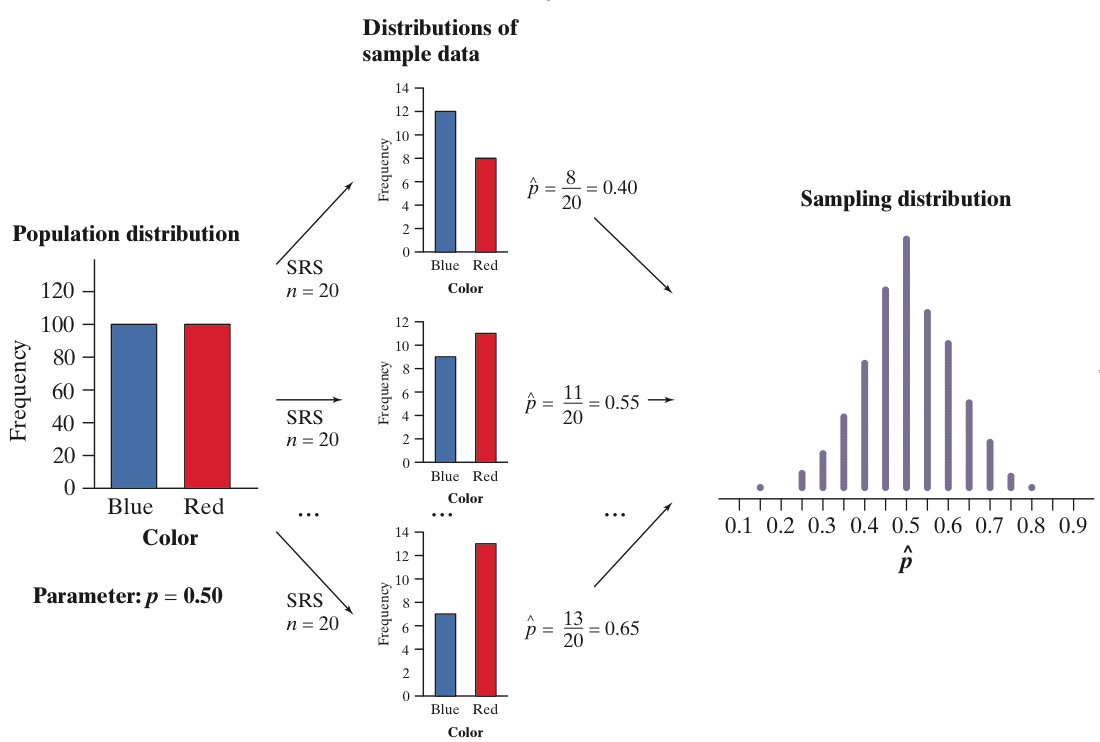
The law of large numbers assures us that if we measure enough subjects, the statistic x-bar will eventually get very close to the unknown parameter *µ*.

**SAMPLING DISTRIBUTIONS**

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.



In practice, it’s difficult to take all possible samples of size *n* to obtain the actual sampling distribution of a statistic. Instead, we can use simulation to imitate the process of taking many, many samples, which is what we did in the activity.

**SAMPLING DISTRIBUTION OF**

**The mean of the sampling distribution of is denoted as** .

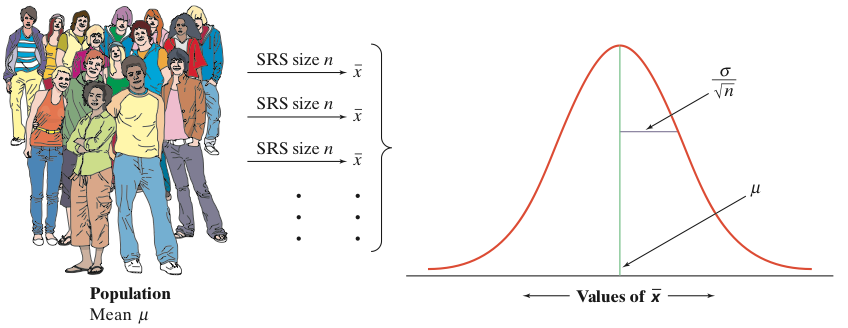
**The standard deviation of the sampling distribution of**  **is denoted as** .

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:

= *µ*

If n is large (n > 30) and the population has mean *µ* and standard deviation *σ,* then the sampling distribution of is Normally distributed with mean *µ* and standard deviation σ/√*n*, *no matter what shape the population distribution has*.

If the population is Normally distributed with mean *µ* and standard deviation *σ,* then the sampling distribution of is Normally distributed with mean *µ* and standard deviation σ/√*n*, regardless of the sample size *n*.



**THE CENTRAL LIMIT THEOREM**

Most population distributions are not Normal. It is a remarkable fact that as the sample size increases, the distribution of sample means changes its shape: it looks less like that of the population and more like a Normal distribution!

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