# Chapter 7 Sampling Distributions of Estimates

We have studied probability, we have discussed random variables and their probability distributions. Now we turn to *Statistical Inference*: process of using data to draw conclusions about some wider population.

## Section 7.1 Parameters and Estimates

Some distinctions to keep in mind ...

- Population versus Sample
- Parameter versus Statistic and Estimate

Definitions:

- **Parameter**: A number that describes the **population**.
- **Statistic (or estimator)**: An abstraction that describes a **sample**.
- **Estimate**: A number **calculated from the data** to estimate an unknown parameter

Notation:

- Population proportion $p$ versus sample proportion $\widehat{P}$ or $\widehat{p}$
- Population mean $\mu$ versus sample mean $\overline{X}$ or $\overline{x}$

Since we hardly ever know the true population parameter value, we take a sample $X_1, X_2, ..., X_n$ and use the sample statistic to **estimate** the parameter. When we do this, the sample statistic may not be equal to the population parameter, in fact, it would change every time we take a new sample (that is why we call this process *ESTIMATION*). Will the observed sample statistic value be a reasonable estimate? If the sample is a **random sample**, then we will be able to say something about the accuracy of the estimation process. What do we mean by a **random sample**?

Recall that $X_1, X_2, ..., X_n$ is a random sample if ...

1.

2.

**Example:**

A poll was conducted by The Heldrich Center for Workforce Development (at Rutgers University).
A random sample of 1000 workers resulted in 460 (for 46%) stating they work more than 40 hours per week.

**Identify:**

Population =

Parameter =

Sample =

Estimate =

Can anyone say how close this observed sample proportion $\widehat{p}$ is to the true population proportion $p$? If we were to take another random sample of the same size $n = 1000$, would we get the same value for the sample proportion $\widehat{p}$ ?

The value of a statistic from a random sample, like $\widehat{p}$ , will vary from sample to sample. So a statistic (also called **estimator**) is a random variable and it will have a probability distribution. This probability distribution is called the **sampling distribution** of the statistic. We will study in this chapter the sampling distribution of the following statistics:

Sample Mean: $\overline{X}$ = the mean of a sample of size $n$

Sample Proportion: $\widehat{P} = \frac{X}{n}$ = the proportion of *successes* in a sample of size $n$

## Section 7.2 Sampling Distribution of the Sample Mean

**Scenario:**

A poll was conducted by The Heldrich Center for Workforce Development (at Rutgers University).
A **sample of 1000 workers** resulted in a **mean number of hours worked per week** of 43.

Population =

Parameter =

Sample =

Estimate =

Can anyone say how close this observed sample mean $\overline{x}$ is to the true population mean $\mu$?

If we were to take another random sample of the same size $n = 1000$, would we get the same value for the sample mean $\overline{x}$?

**Picture Model:**

The value of the sample mean from a random sample will vary from sample to sample. So the sample mean (a statistic) is a random variable and it will have a probability distribution. This probability distribution is called the **sampling distribution** of the statistic. In Section 7.2 we learn about the sampling distribution for a sample mean, $\overline{X}$ = the mean of a sample of size $n$. What can we say about the $\overline{X}$ basket (population)?

**Notation:**

- population mean $\mu$
- population standard deviation $\sigma$
- random sample of size $n$: $X_1, X_2, \ldots, X_n$
- sample mean $\overline{X} = \frac{X_1+X_2+\ldots+X_n}{n}$

**Main Results:**

1. The average of all of the possible sample mean values
(from all possible samples of the same size n) is equal to ________________

That is, the mean of the sample means is equal to the population mean.

Thus the sample mean is an ________________ estimator of the population mean.

2. The standard deviation of all of the possible sample mean values is equal to the original population standard deviation divided by $\sqrt{n}$.

This is **approximately the average distance of the possible sample mean values from the population mean**. It is a measure of the *accuracy of the process* of using a sample mean to estimate the population mean.

Would this quantity $\frac{\sigma}{\sqrt{n}}$ tell me how far away a particular $\overline{x}$ value is from $\mu$?

Note: If the sample size increases, the standard deviation decreases, which says the possible sample mean values will be closer to the true population mean (on average).

3. If the **population has a normal distribution**,
then the distribution of the sample mean is ________________________

Note: Sums and differences of independent normal r.v. also have a normal distribution (Section 6.4.3).

4. If the **population is not necessarily normally distributed**,
   but the **sample size $n$ is large**,
   then the distribution of the sample mean is __________________________

   This result is called the ______________________________________________

See Figures 7.2.3- 7.2.5 on pages 285-287.

**Rule of Thumb:** In general, the approximation is good if ________________

**Example**

ACT scores: **Normally distributed** with mean of 18.6 and standard deviation of 5.9.

a. What is the probability that a **single student** randomly selected will score 21 or higher?

b. A random sample of 50 students is obtained. What is the probability that the **mean score** for these 50 students will be 21 or higher?

c. What if the normal distribution assumption were not given? How would your answer to part b change? Explain.

**Example: Uniform Distribution**

Suppose the random variable X = actual flight time (in minutes) for Delta Airlines flights from Cincinnati to Tampa follows a *uniform distribution over the range of 110 minutes to 130 minutes.*

a. Sketch the distribution for X (include labels for the axes and some values on the axes).

b. Suppose we were to repeatedly take a random sample of size 100 from this distribution and compute the sample mean for each sample. What would the *histogram of the sample mean values* look like? Provide a sketch and any details about the distribution of the sample mean that you can.

**Example:**

Grinding Machine prepares axles: target diameter = 40.125 mm. The standard deviation of diameters is 0.002 mm. A (random) sample of 4 axles taken each hour and the sample mean diameter is recorded and plotted. What will be the mean and standard deviation of the numbers recorded?

**Note:**

When n = 1 we have:

When n = 4 we have:

So to reduce the standard deviation in half, we must ________________

**Question:** What sample size would be needed for the standard deviation of the sample mean to be 0.0005 mm?

**Standard Error of Sample Mean**

Recall that the s.d. of sample mean $\overline{X}$ is ________________

In practice, we don't know s.d. $\sigma$, so we have to estimate $\sigma$ and $sd(\overline{X})$.

We can estimate $\sigma$ by ________________________

and estimate $sd(\overline{X})$ by ________________________

The Standard Error of the Sample Mean is

The Standard Error is a measure of the ________________ of an estimate. You should always report an estimate and its standard error. One way to do this is to report a **confidence intervals**.

**A two-standard-error interval** for $\mu$ is