CHAPTER 1
INTRODUCTION TO STATISTICS

Learning Objective:

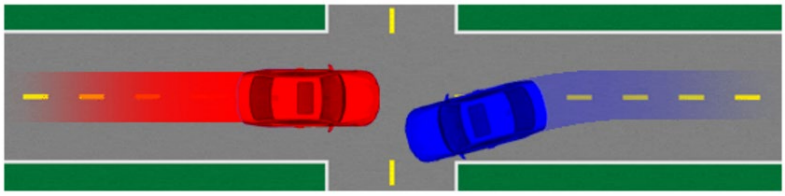
1. *To appreciate statistics as a useful tool to learn more about events around us.*

---

1. Introduction

In general terms, what a working engineer does is to design, build, operate, and/or improve physical systems and products. Civil engineers build highways, waterworks and large buildings; chemical engineers design and operate systems related to fertilizers to fuels, aeronautical and aerospace engineers design and improve aircraft systems, industrial engineers design and operate manufacturing facilities, etc.

As technology advances and new systems and products are encountered, engineers are often faced with questions and situations where their existing knowledge and experience may offer little or no help. Hence, what do engineers do?



Do autopilot self-driving cars generally have a significantly better (lower) reaction time compared to human beings in emergency situations?

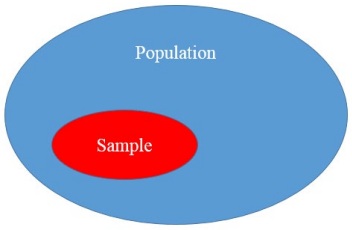
It is necessary then for engineers to ask the correct question, collect unbiased data, analyze data rigorously, and then interpret the results to help them understand how the new system or the new product works. If data are collected haphazardly or analyzed poorly, then valuable time and resources are wasted. Worse, sometimes erroneous conclusions made may lead to bad decisions which could cost lives.


Retrieved from: https://www.nasa.gov/multimedia/imagegallery/image_gallery_2437.html

The NASA family lost seven of its own on the morning of Jan. 28, 1986, when a booster engine failed, causing the shuttle Challenger to break apart just 73 seconds after launch.

2. Statistics

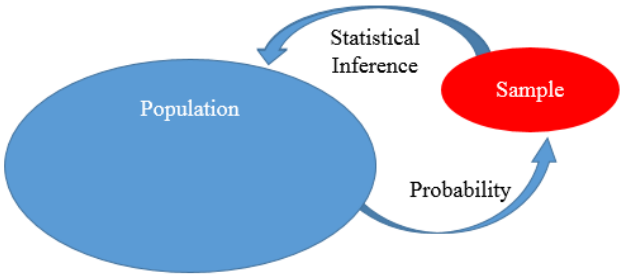
So what is statistics? It is a scientific way of thinking that helps us understand more about the phenomena that we encounter in our workplace or in our daily lives – performance of autopilot cars, likelihood of space shuttle exploding, the quality of potato chips, and many others.



To better understand what statistics is, let us first look at the concept of sample and population.
The **population**, whether it is people, grapes or marbles, is the entire group we want information about.
A **sample** is part of the whole. In other words, sample is a subset of population.

If we can access every member of the population to collect data, this will be a **census**.
It seems like a census would be a straightforward way to get the most accurate, thorough information, but taking an accurate census is not nearly as easy as one might think. It could be due to ignorance about size of population, or sheer size of population, or inaccessibility of its members.

Thus, the sample is the part we actually gather data from. Since the sample is only a part of the population, we can study it more extensively than we can all the members of the population. Then we can use the sample data to draw conclusions about the entire population (statistical inference). Conversely, knowledge about the population would tell us how likely we are to get the sample we obtained (probability).



For these conclusions to be valid though, the sample must be representative of the population. To make sure it is, statisticians rely on what is called simple **random sampling**. This means that the sample is chosen in such a way that each member has an equal chance of being selected. This helps eliminate bias in the study design and the conclusion drawn based on the sample could be generalized to the entire population with high level of confidence.

Retrieved from: "Census and Sampling: Against All Odds—Inside Statistics," director. Films Media Group, 2013, http://fod.infobase.com/portalplaylists.aspx?wid=151497&xtid=111535.

Frito Lay potato chip fans count on consistent appearance and taste from their favorite brand, and sampling is one way the company meets those expectations.

3. Outline of Content

The aim of statistics in this module is to provide you with a strong foundation in the scientific thinking of statistics that would be useful in solving engineering problems involving statistical data.

Hence in Chapter 2, you will be introduced to the statistical problem-solving process:

You will also learn important skills in collecting, managing, summarizing and presenting data in an informative manner through tables and charts using a statistical software called **Minitab Express**.

In Chapter 3, you will be taught the basic concept of probability as a measure of likelihood of an event happening. Then, the concept of probability will be extended to probability distributions where real-life scientific case study such as the explosion of space shuttle Challenger will be discussed. The idea of "rare-event" will also be introduced.

Finally, in Chapter 4, the very important concept of sampling distribution and the Central Limit Theorem will be presented. Using samples, you will learn how to estimate the value of an unknown population parameter, specifically the population mean. Instead of using only one value to estimate the unknown population mean, you will be taught the concept of confidence interval to construct an interval of estimates for the unknown population mean. You will also look at a case study involving claims about Subway selling less than 12-inch foot-long subway sandwiches.

LAB 1 : Minitab Express

Task: Install Minitab Express. Instructions can be found in eSP module site.