

# 1 Introduction

First we start with a roadmap for the course. The subject of **engineering statistics** is defined, its importance is described and basic terminology is introduced.

## 1.1 What is statistics?

Statistics is the science of collecting, presenting, analyzing, and making decisions from data. Often, as an engineer, it is necessary to **collect and interpret data** that will help in understanding how a new system or product works.

Statistics has applications to engineering through quality control, process control, reliability, risk management, system identification, design of experiments, etc.

**Definition 1.1.** *Engineering statistics* is the study of how best to

1. collect engineering data,
2. summarize or describe engineering data, and
3. draw formal inference and practical conclusions on the basis of engineering data,

all while recognizing the reality of variation.

We can break down this study into three main tasks:

1. **Summary:**
2. **Inference:**
3. **Interpretation:**

**Example 1.1** (Heat treating gears, pg. 2). A process engineer is faced with the question, "How should gears be loaded into a continuous carburizing furnace in order to minimize distortion during heat treating?" The engineer conducts a well-thought-out study and obtains the runout values for 38 gears laid and 39 gears hung.

hung	laid
7, 8, 8, 10, 10,	5, 8, 8, 9, 9, 9,
10, 10, 11, 11,	9, 10, 10, 10,
11, 12, 13, 13,	11, 11, 11, 11,
13, 15, 17, 17,	11, 11, 11, 12,
17, 17, 18, 19,	12, 12, 12, 13,
19, 20, 21, 21,	13, 13, 13, 14,
21, 22, 22, 22,	14, 14, 15, 15,
23, 23, 23, 23,	15, 15, 16, 17,
24, 27, 27, 28,	17, 18, 19, 27
31, 36	

Table 1: Thrust face runouts (.0001 in.)

This data, as is, is hard to get insights from. Should the gears be hung or laid? We still cannot tell by looking at the table.

**Numerical summaries:**

**Variation:**

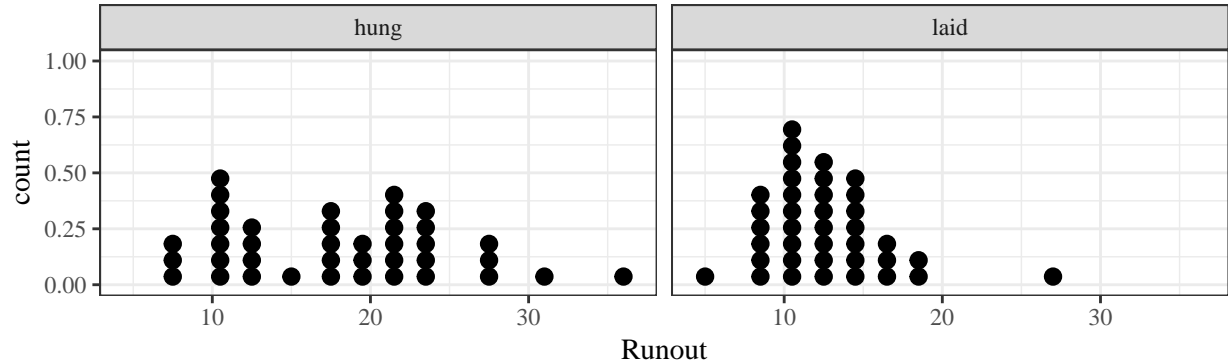


Figure 1: Dot diagrams of runouts.

**Inference:** Runout values are variable, is there any assurance that the difference seen in the present data would reappear in further testing?

Points to need for methods of formal statistical inference from data and translation into practical conclusions. We can use statistical inference to say

1. We are roughly 90% sure that the difference in long-run mean runout produced under conditions like those in the study is in the range 3.2 to 7.4.
2. We are roughly 95% sure that the difference in long-run mean runout produced under conditions like those in the study is in the range 3.0 to 22.2.

**Interpretation:** Explains those answers in layman's terms without all the probability theory, symbols and equations.

**Overview/schematic:**

## 1.2 Basic terminology

It's important we speak the same language. This section introduces common terminology related to statistical studies, types of data, and types of data structures.

### 1.2.1 Population vs. sample

**Definition 1.2.** A *population* is the entire group of objects about which one wishes to gather information in a statistical study.

**Definition 1.3.** A *sample* is the group of objects on which one actually gathers data.

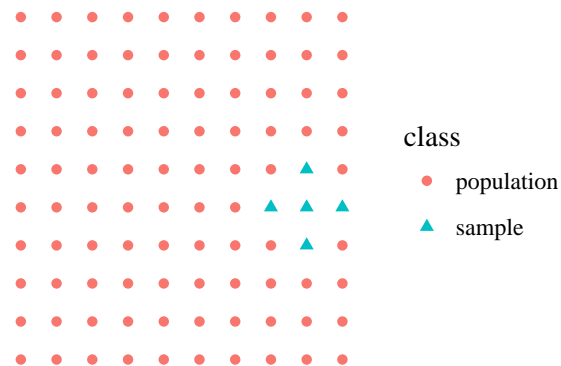


Figure 2: The relationship between a population and a sample. In this example, we have 100 parts and 5 are examined in order to verify acceptability. Notice we say “one sample”, not “five samples”.

**Example 1.2** (Heat treating gears, cont'd).

**Population:**

**Sample:**

**Example 1.3** (Fiscal cliff). On Dec. 1-2, 2012, Gallup conducted a study to find out what proportion of Americans prefer a compromise on the Fiscal Cliff issue. 1000 adults were randomly selected for telephone interviews. The adults were aged 18 and older and living in any of the 50 U.S. states or the District of Columbia.

**Population:**

**Sample:**

**Example 1.4.** Esbit manufactures fuel pellets out of compressed hexamine powder. Suppose a new shipment of 100 pelletizing machines arrives, and the goal of a new study is to determine the quality of this particular new shipment.

5 machines out of the 100 are randomly selected for comprehensive testing in which each produces 200 pellets, and each pellet's mass, volume, flash point, and rate of combustion are measured.

**Population:**

**Sample:**

### 1.2.2 Types of studies

When an engineer **collects data**, (s)he must decide how active to be. Should the engineer manipulate the process or let things happen and record the results?

**Definition 1.4.** An *experimental study* (or, more simply, an experiment) is one in which the investigator's role is active. Process variables are manipulated, and the study environment is regulated.

**Definition 1.5.** In a *randomized experiment*, investigators control the assignment of treatments to experimental units using a chance mechanism (like the flip of a coin or a computer's random number generator).

**Example 1.5** (Smoking). To answer the question of if smoking affects lung capacity in young men, an investigator does the following.

Finds 100 men age 20 who do not currently smoke. Randomly assigns 50 of the 100 men to the smoking treatment and the other 50 to the non-smoking treatment. Those in the smoking group smoke a pack a day for 10 years while those in the control group remain smoke free for 10 years. Measures lung capacity for each of the 100 men. Analyze, interpret, and draw conclusions from data.

**Definition 1.6.** An *observational study* is one in which the investigator's role is basically passive. A process or phenomenon is watched and data are recorded, but there is no intervention on the part of the person conducting the study.

**Example 1.6** (Smoking). To answer the question of if smoking affects lung capacity in young men, an investigator does the following.

Finds 100 men age 30 of which 50 have been smoking a pack a day for 10 years while the other 50 have been smoke free for 10 years. Measures lung capacity for each of the 100 men. Analyze, interpret, and draw conclusions from data.

### 1.2.3 Data

### 1.2.4 Data structures

## 1.3 Measurement difficulties