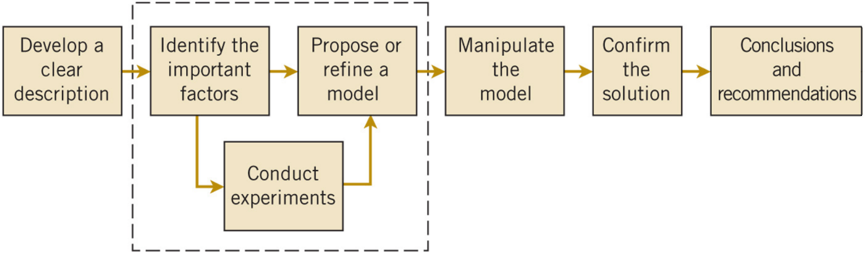
STAT 509: Statistics for Engineers

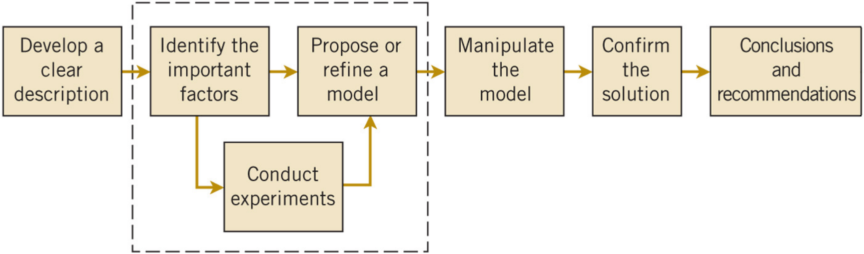
Chapter 1: The Role of Statistics in Engineering

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

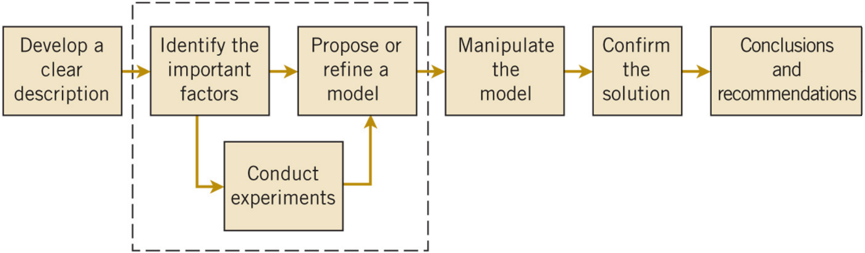


An engineer is someone who solves problems of interest to society by the efficient application of scientific principles. Engineers accomplish this by either refining an existing product or process or by designing a new product or process that meets customers' needs. The engineering, or scientific, method is the approach to formulating and solving these problems. The steps in the engineering method are as follows:

1. Develop a clear and concise description of the problem.
2. Identify, at least tentatively, the important factors that affect this problem or that may play a role in its solution.



3. Propose a model for the problem, using scientific or engineering knowledge of the phenomenon being studied. State any limitations or assumptions of the model.
4. Conduct appropriate experiments and collect data to test or validate the tentative model or conclusions made in steps 2 and 3.
5. Refine the model on the basis of the observed data.



6. Manipulate the model to assist in developing a solution to the problem.
7. Conduct an appropriate experiment to confirm that the proposed solution to the problem is both effective and efficient.
8. Draw conclusions or make recommendations based on the problem solution.

Chapter 1: The Role of Statistics in Engineering

The field of statistics deals with the collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes. In simple terms,

statistics is the science of data.

Because many aspects of engineering practice involve working with data, obviously knowledge of statistics is just as important to an engineer as are the other engineering sciences. Specifically, statistical techniques can be powerful aids in designing new products and systems, improving existing designs, and designing, developing, and improving production processes.

Variability

Statistical methods are used to help us describe and understand variability. By variability, we mean that successive observations of a system or phenomenon do not produce exactly the same result. Statistical thinking can give us a useful way to incorporate this variability into our decision-making processes. For example, consider the gasoline mileage performance of your car. Do you always get exactly the same mileage performance on every tank of fuel? Of course not. This observed variability depends on many factors:

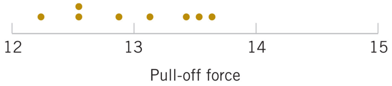
- the type of driving that has occurred most recently (city versus highway),
- the changes in the vehicle's condition over time (which could include factors such as tire inflation, engine compression, or valve wear),
- the brand and/or octane number of the gasoline used,
- or possibly even the weather conditions that have been recently experienced. These factors represent potential sources of variability in the system.

Variability

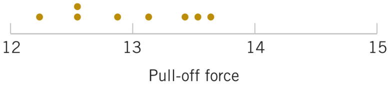
We also encounter variability in dealing with engineering problems. For example, suppose that an engineer is designing a nylon connector to be used in an automotive engine application. The engineer is considering establishing the design specification on wall thickness at 3/32 inch but is somewhat uncertain about the effect of this decision on the connector pull-off force. If the pull-off force is too low, the connector may fail when it is installed in an engine.

Eight prototype units are produced and their pull-off forces measured, resulting in the following data (in pounds):

$$12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1$$



Variability



As we anticipated, not all of the prototypes have the same pull-off force. We say that there is variability in the pull-off force measurements. Because the pull-off force measurements exhibit variability, we consider the pull-off force to be a random variable. A convenient way to think of a random variable, say X, that represents a measurement is by using the model

$$X = \mu + \epsilon$$

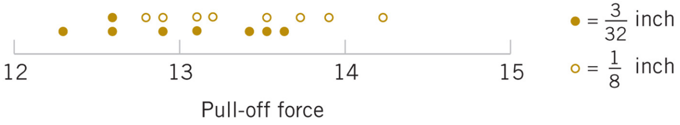
where μ is a constant and ϵ is a random disturbance. The constant remains the same with every measurement, but small changes in the environment, variance in test equipment, differences in the individual parts themselves, and so forth change the value of ϵ. Hence the actual measurements X exhibit variability. We often need to describe, quantify, and ultimately reduce variability.

Variability

The need for statistical thinking arises often in the solution of engineering problems. Consider the engineer designing the connector. From testing the prototypes, he knows that the average pull-off force is 13.0 pounds. However, he thinks that this may be too low for the intended application, so he decides to consider an alternative design with a thicker wall, $1/8$ inch in thickness. Eight prototypes of this design are built, and the observed pull-off force measurements are

$$12.9, 13.7, 12.8, 13.9, 14.2, 13.2, 13.5, 13.1$$

The average is 13.4. Results for both samples are plotted as dot diagrams below. This display gives the impression that increasing the wall thickness has led to an increase in pull-off force.



Variability

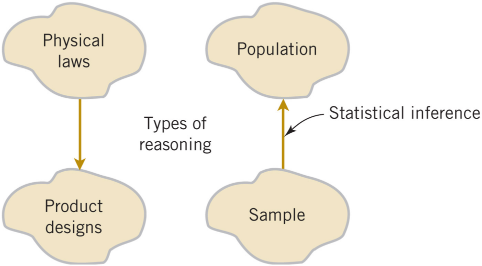
However, there are some obvious questions to ask. For instance,

- How do we know that **another sample** of prototypes will not give different results?
- Is a sample of **eight** prototypes adequate to give reliable results?
- If we use the test results obtained so far to conclude that increasing the wall thickness increases the strength, what **risks** are associated with this decision?
- For example, is it possible that the apparent increase in pull-off force observed in the thicker prototypes is due only to the **inherent variability** in the system and that increasing the thickness of the part (and its cost) **really has no effect** on the pull-off force?

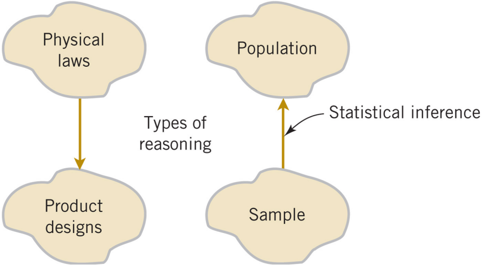
Population and Sample

Often, physical laws (such as Ohm's law and the ideal gas law) are applied to help design products and processes. We are familiar with this reasoning from general laws to specific cases. But it is also important to reason from a specific set of measurements to more general cases to answer the previous questions.

This reasoning comes from a **sample** (such as the eight connectors) to a **population** (such as the connectors that will be in the products that are sold to customers). The reasoning is referred to as statistical inference.



Population and Sample



Historically, measurements were obtained from a sample of people and generalized to a population, and the terminology has remained. Clearly, reasoning based on measurements from some objects to measurements on all objects can result in errors (called **sampling errors**). However, if the sample is selected properly, these risks can be quantified and an appropriate sample size can be determined.

Ways to collect data

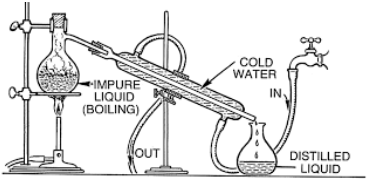
When the data are all of the observations in the population, this results in a **census**.

However, in the engineering environment, the data are almost always a **sample** that has been selected from the **population**.

Three basic methods of collecting data are

- A retrospective study using historical data
- An observational study
- A designed experiment

Montgomery, Peck, and Vining (2012) describe an acetone-butyl alcohol distillation column for which concentration of acetone in the distillate (the output product stream) is an important variable.



An illustrative picture taken from https://www.precisionwaterusa.com/blog/why-distillation

Retrospective Study

Factors that may affect the distillate are the reboil temperature, the condensate temperature, and the reflux rate. Production personnel obtain and archive the following records:

- The concentration of acetone in an hourly test sample of output product
- The reboil temperature log, which is a record of the reboil temperature over time
- The condenser temperature controller log
- The nominal reflux rate each hour

The reflux rate should be held constant for this process. Consequently, production personnel change this very infrequently.

Hazards of Using Historical Data

A retrospective study would use either all or a sample of the historical process data archived over some period of time. The study objective might be to discover the relationships among the two temperatures and the reflux rate on the acetone concentration in the output product stream. However, this type of study presents some problems:

1. We may not be able to see the relationship between the reflux rate and acetone concentration because **the reflux rate did not change much over the historical period**.
2. The archived data on the two temperatures (which are recorded almost **continuously**) do not correspond perfectly to the acetone concentration measurements (which are made **hourly**). It may not be obvious how to construct an approximate correspondence.

Hazards of Using Historical Data

3. Production maintains the two temperatures as closely as possible to desired targets or set points. Because the **temperatures change so little**, it may be difficult to assess their real impact on acetone concentration.
4. In the narrow ranges within which they do vary, the condensate temperature tends to increase with the reboil temperature. Consequently, **the effects of these two process variables on acetone concentration may be difficult to separate**.

Hazards of Using Historical Data

As you can see, a retrospective study may involve a significant amount of **data**, but those data may contain relatively little useful **information** about the problem. Furthermore, some of the relevant data may be missing, there may be transcription or recording errors resulting in **outliers** (or unusual values), or data on other important factors may not have been collected and archived. In the distillation column, for example, the specific concentrations of butyl alcohol and acetone in the input feed stream are very important factors, but they are not archived because the concentrations are too hard to obtain on a routine basis.

As a result of these types of issues, statistical analysis of historical data sometimes identifies interesting phenomena, but solid and reliable explanations of these phenomena are often difficult to obtain.

Observational Study

In an observational study, the engineer observes the process or population, disturbing it as little as possible, and records the quantities of interest. Because these studies are usually conducted for a relatively short time period, sometimes variables that are not routinely measured can be included. In the distillation column, the engineer would design a form to record the two temperatures and the reflux rate when acetone concentration measurements are made. It may even be possible to measure the input feed stream concentrations so that the impact of this factor could be studied.

Generally, an observational study tends to solve problems 1 and 2 and goes a long way toward obtaining accurate and reliable data. However, observational studies may not help resolve problems 3 and 4.

Designed Experiments

In a designed experiment, the engineer makes **deliberate or purposeful changes** in the controllable variables of the system or process, observes the resulting system output data, and then makes an inference or decision about which variables are responsible for the observed changes in output performance. The nylon connector example illustrates a **designed experiment**; that is, a deliberate change was made in the connector's wall thickness with the objective of discovering whether or not a stronger pull-off force could be obtained. Experiments designed with basic principles such as **randomization** are needed to establish **cause-and-effect** relationships.

Designed Experiments

Much of what we know in the engineering and physical-chemical sciences is developed through testing or experimentation. Often engineers work in problem areas **in which no scientific or engineering theory is directly or completely applicable**, so experimentation and observation of the resulting data constitute the only way that the problem can be solved. Even when there is a good underlying scientific theory that we may rely on to explain the phenomena of interest, **it is almost always necessary to conduct tests or experiments to confirm that the theory is indeed operative in the situation or environment in which it is being applied**.

Statistical thinking and statistical methods play an important role in planning, conducting, and analyzing the data from engineering experiments. Designed experiments play a very important role in engineering design and development and in the improvement of manufacturing processes.