1 The Role of Statistics in Engineering

CHAPTER OUTLINE

Learning Objectives for Chapter 1

After careful study of this chapter, you should be able to do the following:

1. Identify the role that statistics can play in the engineering problem-solving process.
2. Discuss how variability affects the data collected and used for engineering decisions.
3. Explain the difference between enumerative and analytical studies.
4. Discuss the different methods that engineers use to collect data.
5. Identify the advantages that designed experiments have in comparison to the other methods of collecting engineering data.
6. Explain the differences between mechanistic models & empirical models.
7. Discuss how probability and probability models are used in engineering and science.

What Do Engineers Do?

An engineer is someone who solves problems of interest to society with the efficient application of scientific principles by:

- Refining existing products
- Designing new products or processes

The Creative Process

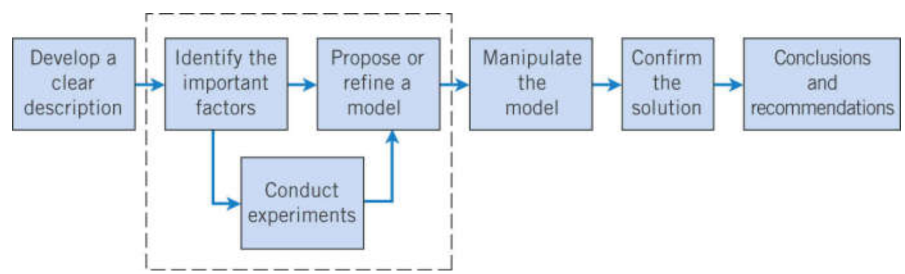

Figure 1.1 The engineering method

Statistics Supports The Creative Process

The field of statistics deals with the collection, presentation, analysis, and use of data to:

- Make decisions
- Solve problems
- Design products and processes

It is the science of learning information from data.

Experiments & Processes Are Not Deterministic

- Statistical techniques are useful for describing and understanding variability.
- By variability, we mean successive observations of a system or phenomenon do *not* produce exactly the same result.
- Statistics gives us a framework for describing this variability and for learning about potential sources of variability.

An Engineering Example of Variability-1

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 (in pounds):

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

A Engineering Example of Variability-2

- The **dot diagram** is a very useful plot for displaying a small body of data - say up to about 20 observations.
- This plot allows us to see easily two features of the data; the **location**, or the middle, and the **scatter** or **variability**.

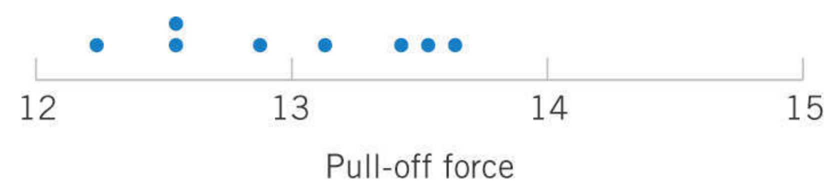

Figure 1-2 Dot diagram of the pull-off force data when wall thickness is 3/32 inch.

A Engineering Example of Variability-3

- The engineer considers an alternate design and eight prototypes are built and pull-off force measured.
- The dot diagram can be used to compare two sets of data.

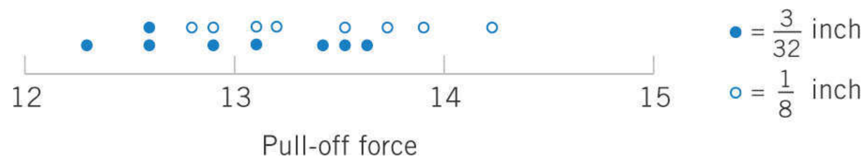

Figure 1-3 Dot diagram of pull-off force for two wall thicknesses.

A Engineering Example of Variability-4

- Since pull-off force varies or exhibits variability, it is a **random variable**.
- A random variable, X, can be modeled by:

$$X = \mu + \varepsilon \tag{1-1}$$

where μ is a constant and ε is a random disturbance.

Two Directions of Reasoning

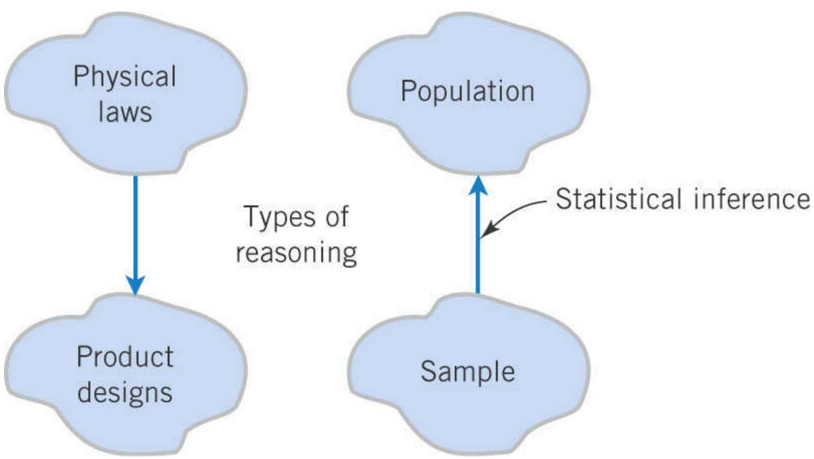

Figure 1-4 Statistical inference is one type of reasoning.

Basic Types of Studies

Three basic methods for collecting data:

- A **retrospective** study using historical data
  - Data collected in the past for other purposes.
- An **observational** study
  - Data, presently collected, by a passive observer.
- A **designed experiment**
  - Data collected in response to process input changes.

Hypothesis Tests

Hypothesis Test
- A statement about a process behavior value.
- Compared to a claim about another process value.
- Data is gathered to support or refute the claim.

One-sample hypothesis test:
- Example: Ford avg mpg = 30 vs. avg mpg < 30

Two-sample hypothesis test:
- Example: Ford avg mpg – Chevy avg mpg = 0 vs. > 0.

Factor Experiment Example-1

Consider a petroleum distillation column:
- Output is acetone concentration
- Inputs (factors) are:
  1. Reboil temperature
  2. Condensate temperature
  3. Reflux rate
- Output changes as the inputs are changed by experimenter.
- Each factor is set at 2 reasonable levels (-1 and +1)
- 8 (2^3) runs are made, at every combination of factors, to observe acetone output.
- Resultant data is used to create a mathematical model of the process representing cause and effect.

Factor Experiment Example-2

Table 1-1 The Designed Experiment (Factorial Design) for the Distillation Column

Reboil Temp.	Condensate Temp.	Reflux Rate
−1	−1	−1
+1	−1	−1
−1	+1	−1
+1	+1	−1
−1	−1	+1
+1	−1	+1
−1	+1	+1
+1	+1	+1

Factor Experiment Example-3

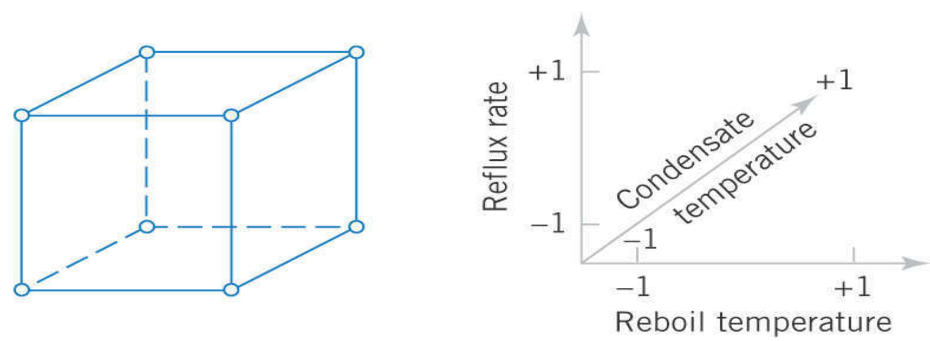

Figure 1-5 The factorial experiment for the distillation column.

Factor Experiment Example-4

Now consider a new design of the distillation column:

- Repeat the settings for the new design, obtaining 8 more data observations of acetone concentration.
- Resultant data is used to create a mathematical model of the process representing cause and effect of the new process.
- The response of the old and new designs can now be compared.
- The most desirable process and its settings are selected as optimal.

Factor Experiment Example-5

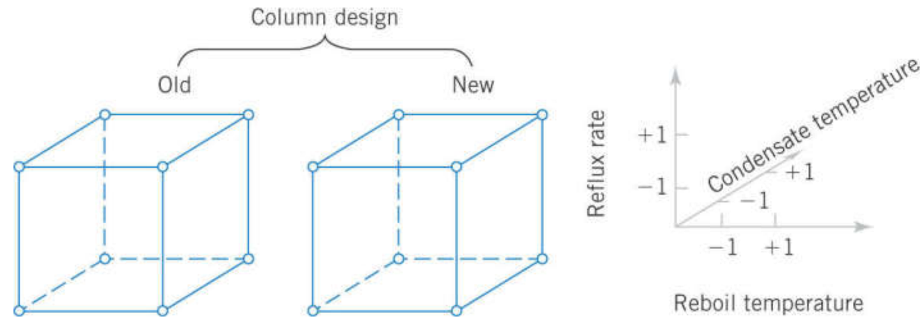

Figure 1-6 A four-factorial experiment for the distillation column $2^4 = 16$ settings.

Factor Experiment Considerations

- Factor experiments can get too large. For example, 8 factors will require $2^8 = 256$ experimental runs of the distillation column.
- Certain combinations of factor levels can be deleted from the experiments without degrading the resultant model.
- The result is called a **fractional factorial experiment**.

Factor Experiment Example-6

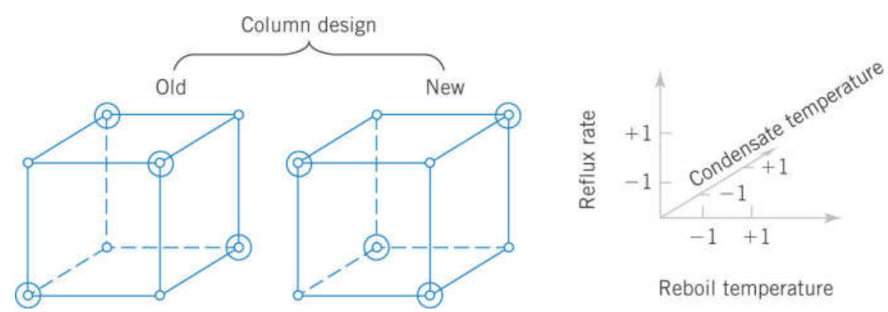

Figure 1-7 A fractional factorial experiment for the distillation column (one-half fraction) $2^4 / 2 = 8$ circled settings.

Distribution of 30 Distillation Column Runs

Whenever data are collected over time, it is important to plot the data over time. Phenomena that might affect the system or process often become more visible in a time-oriented plot and the concept of stability can be better judged.

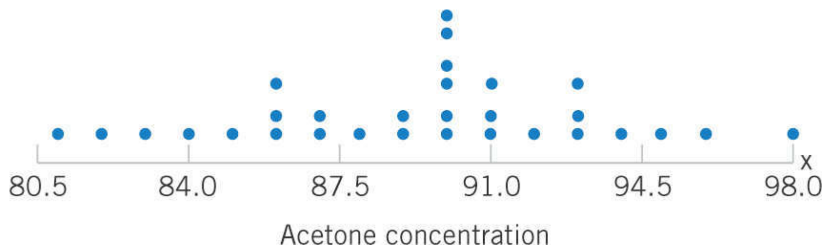

Figure 1-8 The dot diagram illustrates data **centrality** and **variation**, but does not identify any time-oriented problem.

30 Observations, Time Oriented

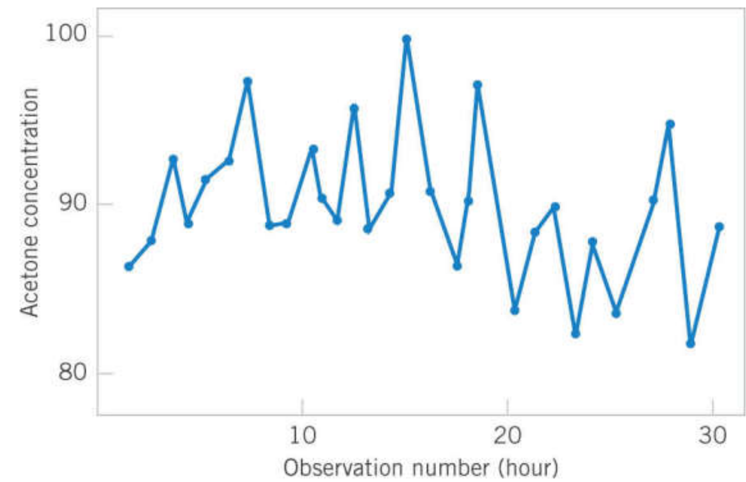

Figure 1-9 A time series plot of concentration provides more information than a dot diagram – shows a developing **trend**.

An Experiment in Variation

W. Edwards Deming, a famous industrial statistician & contributor to the Japanese quality revolution, conducted a illustrative experiment on process **over-control** or **tampering**.

Let's look at his apparatus and experimental procedure.

Deming's Experimental Set-up

Marbles were dropped through a funnel onto a target and the location where the marble struck the target was recorded.

Variation was caused by several factors:

Marble placement in funnel & release dynamics, vibration, air currents, measurement errors.

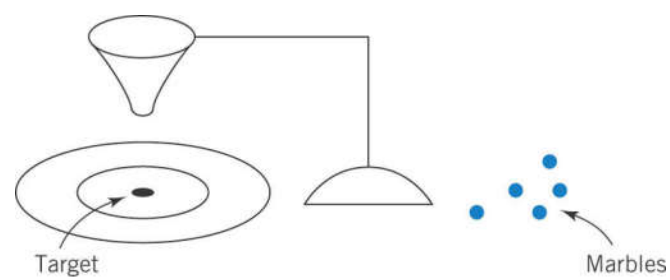

Figure 1-10 Deming's Funnel experiment

Deming's Experimental Procedure

- The funnel was aligned with the center of the target. Marbles were dropped. The distance from the strike point to the target center was measured and recorded
- Strategy 1: The funnel was not moved. Then the process was repeated.
- Strategy 2: The funnel was moved an equal distance in the opposite direction to compensate for the error. Then the process was repeated.

Adjustments Increased Variability

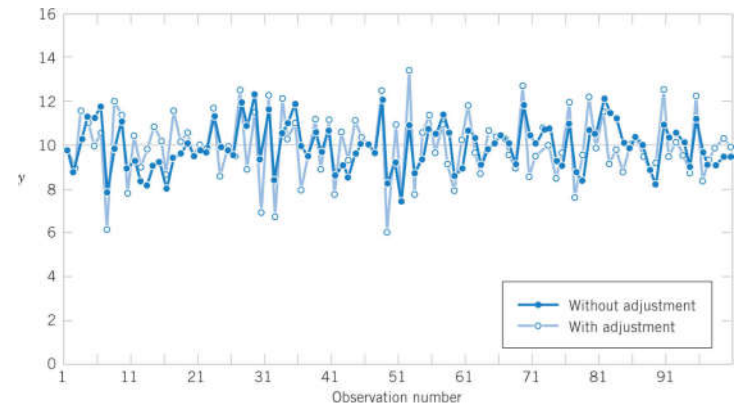

Figure 1-11 Adjustments applied to random disturbances **over-controlled** the process and increased the deviations from the target.

Conclusions from the Deming Experiment

The lesson of the Deming experiment is that a process should not be adjusted in response to random variation, but only when a clear shift in the process value becomes apparent.

Then a process adjustment should be made to return the process outputs to their normal values.

To identify when the shift occurs, a control chart is used. Output values, plotted over time along with the outer limits of normal variation, pinpoint when the process leaves normal values and should be adjusted.

Detecting & Correcting the Process

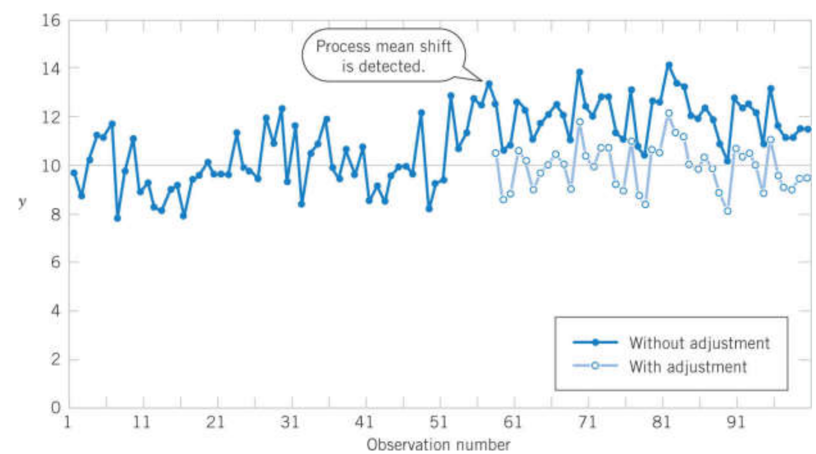

Figure 1-12 Process mean shift is detected at observation #57, and an adjustment (a decrease of two units) reduces the deviations from target.

How Is the Change Detected?

- A control chart is used. Its characteristics are:
  - Time-oriented horizontal axis, e.g., hours.
  - Variable-of-interest vertical axis, e.g., % acetone.
- Long-term average is plotted as the center-line.
- Long-term usual variability is plotted as an upper and lower control limit around the long-term average.
- A sample of size n is taken hourly and the averages are plotted over time. If the plot points are between the control limits, then the process is normal; if not, it needs to be adjusted.

How Is the Change Detected Graphically?

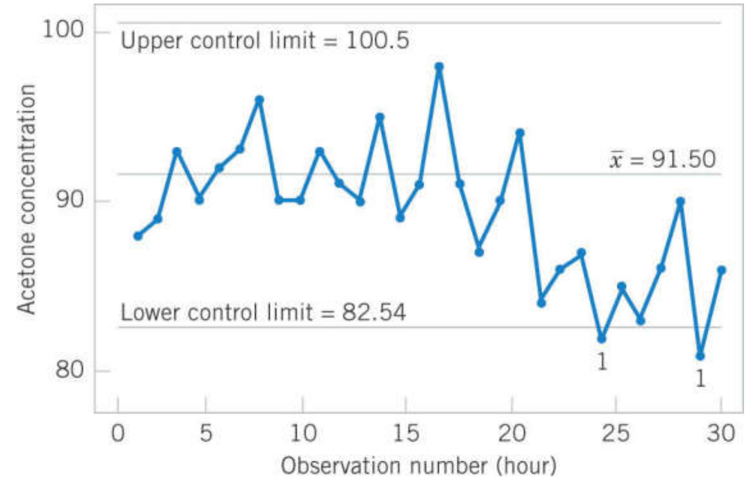

Figure 1-13 A control chart for the chemical process concentration data. Process steps out at hour 24 &29. Shut down & adjust process.

Use of Control Charts

Deming contrasted two purposes of control charts:

1. **Enumerative studies**: Control chart of past production lots. Used for lot-by-lot acceptance sampling.

2. **Analytic studies**: Real-time control of a production process.

Visualizing Two Control Chart Uses

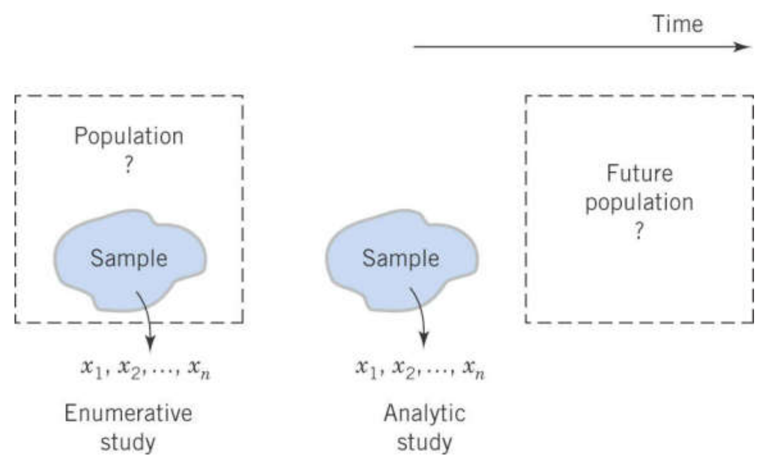

Figure 1-14 Enumerative versus analytic study.

Understanding Mechanistic & Empirical Models

- A **mechanistic model** is built from our underlying knowledge of the basic physical mechanism that relates several variables.

Example: Ohm's Law

Current = voltage/resistance

$$I = E/R$$

$$I = E/R + \varepsilon$$

- The form of the function is known.

Mechanistic and Empirical Models

An **empirical model** is built from our engineering and scientific knowledge of the phenomenon, but is not directly developed from our theoretical or first-principles understanding of the underlying mechanism.

The form of the function is not known *a priori*.

An Example of an Empirical Model

- We are interested in the numeric average molecular weight (M_n) of a polymer. Now we know that M_n is related to the viscosity of the material (V), and it also depends on the amount of catalyst (C) and the temperature (T) in the polymerization reactor when the material is manufactured. The relationship between M_n and these variables is

$$M_n = f(V,C,T)$$

say, where the form of the function f is unknown.

- We estimate the model from experimental data to be of the following form where the b's are unknown parameters.

$$M_n = \beta_0 + \beta_1 V + \beta_2 C + \beta_3 T + \epsilon$$

Another Example of an Empirical Model

- In a semiconductor manufacturing plant, the finished semiconductor is wire-bonded to a frame. In an observational study, the variables recorded were:
  - Pull strength to break the bond (y)
  - Wire length (x_1)
  - Die height (x_2)
- The data recorded are shown on the next slide.

Table 1-2 Wire Bond Pull Strength Data

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50
2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	69.00	20	600
18	10.30	1	585
19	34.93	10	540
20	46.59	15	250
21	44.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400

Empirical Model That Was Developed

$$\text{Pull strength} = \beta_0 + \beta_1(\text{wire length}) + \beta_2(\text{die height}) + \epsilon$$

In general, this type of empirical model is called a **regression model**.

The **estimated** regression relationship is given by:

$$\widehat{\text{Pull strength}} = 2.26 + 2.74(\text{wire length}) + 0.0125(\text{die height})$$

Visualizing the Data

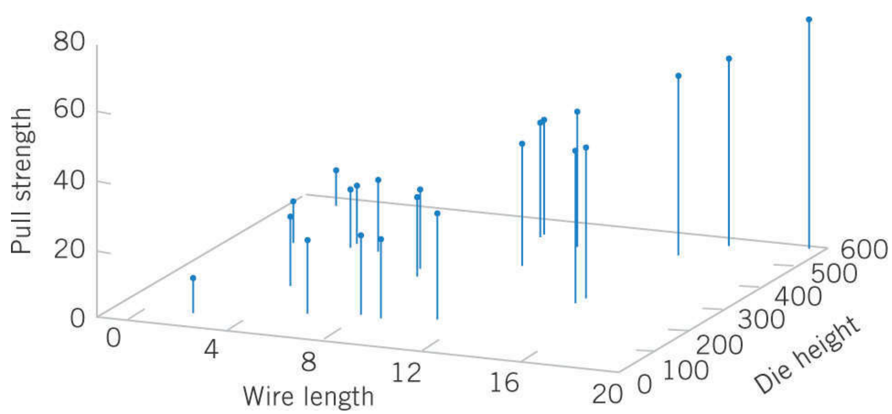

Figure 1-15 Three-dimensional plot of the pull strength (y), wire length (x_1) and die height (x_2) data.

Visualizing the Resultant Model Using Regression Analysis

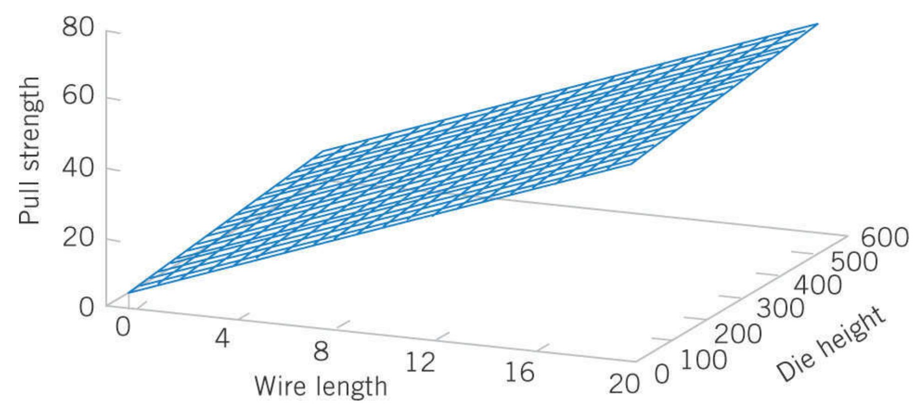

Figure 1-16 Plot of the predicted values (a plane) of pull strength from the empirical regression model.

Models Can Also Reflect Uncertainty

- **Probability models** help quantify the risks involved in statistical inference, that is, risks involved in decisions made every day.
- Probability provides the **framework** for the study and application of statistics.
- Probability concepts will be introduced in the next lecture.

Important Terms & Concepts of Chapter 1

Analytic study
Cause and effect
Designed experiment
Empirical model
Engineering method
Enumerative study
Factorial experiment
Fractional factorial experiment
Hypothesis testing
Interaction
Mechanistic model
Observational study
Overcontrol
Population
Probability model
Problem-solving method
Randomization
Retrospective study
Sample
Statistical inference
Statistical process control
Statistical thinking
Tampering
Time series
Variability