The Roles of statistics

- The Engineering Method and Statistical Thinking
- Collecting Engineering Data
- Mechanistic and Empirical Models
- Probability and Probability Models

Objectives

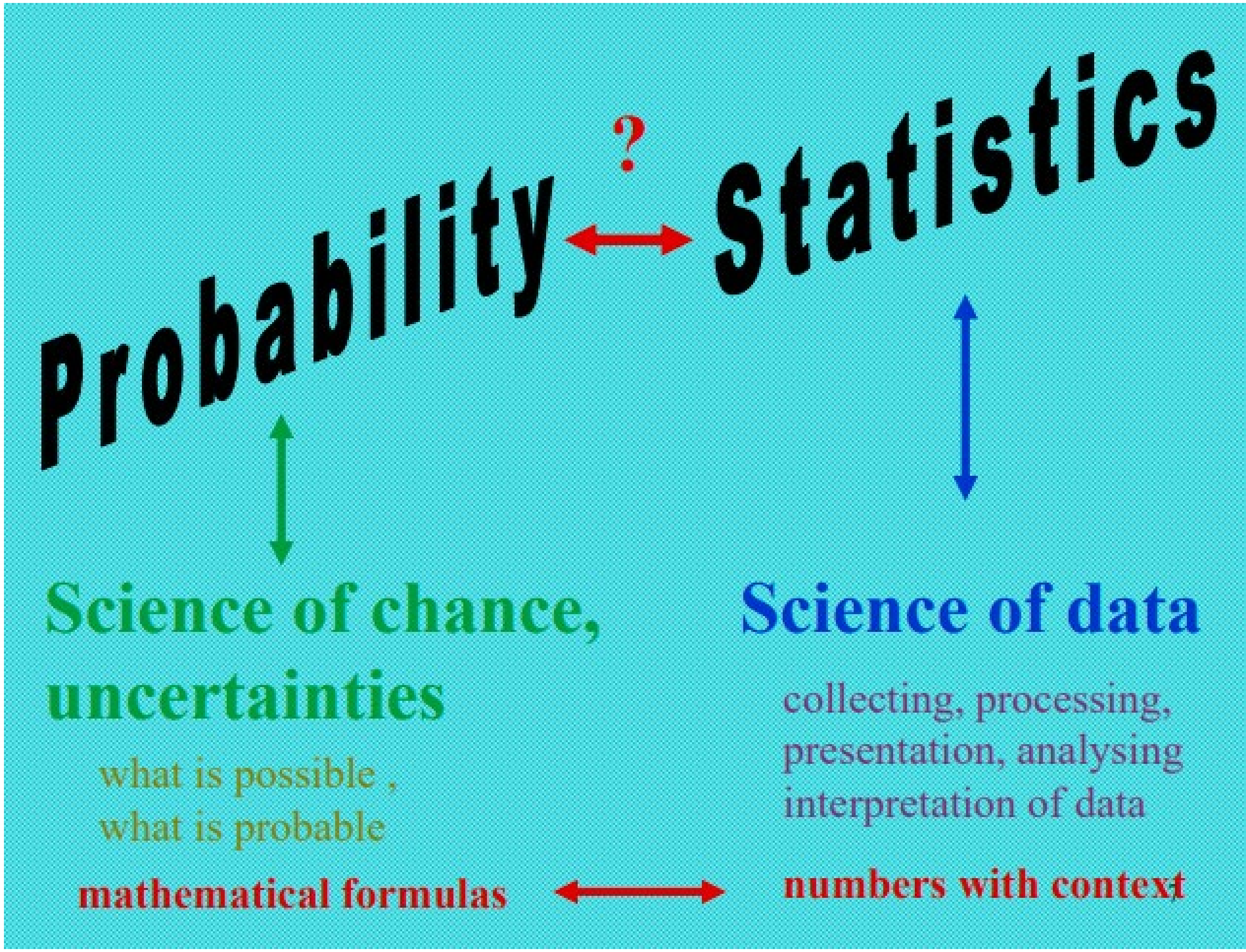
- The role that statistics
- Making engineering decisions
- Enumerative vs analytical studies
- Methods of data collection
- The advantages of designed experiments
- Mechanistic models vs empirical models
- Why probability models?

Descriptive Statistics

- Involves
  - Collecting data
  - Organizing or summarizing data
  - Presenting data
- Purpose
  - Describe the situation

Inferential Statistics

- Involves
  - Estimation
  - Hypothesis testing
- Purpose
  - Draw conclusions
  - or inferences
  - about population
  - characteristics


Probability
?
Statistics
Science of chance, uncertainties
what is possible ,
what is probable
Science of data
collecting, processing, presentation, analysing interpretation of data
mathematical formulas
numbers with context


Data
Statistical Tools
Information

Step 1
Experiment:
Compare two teaching methods

Population of first-grade children

Data
Test scores for the students in each sample

Sample A
Taught by Method A

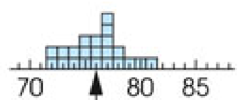
73	75	72	79
76	77	75	77
72	75	76	78
80	74	76	78
73	77	74	81
77	77		

Sample B
Taught by Method B

68	70	73	71
67	72	70	71
75	68	70	71
72	74	69	72
76	73	70	70
69			

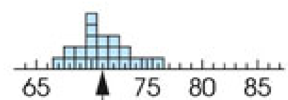
Step 2
Descriptive statistics:
Organize and simplify

Sample A



Average score = 76

Sample B



Average score = 71

Step 3
Inferential statistics:
Interpret results

The sample data show a 5-point difference between the two teaching methods. However, there are two ways to interpret the results:

1. There actually is no difference between the two teaching methods, and the sample difference is due to chance (sampling error).
2. There really is a difference between the two methods, and the sample data accurately reflect this difference.

The goal of inferential statistics is to help researchers decide between the two interpretations.

Statistics in Engineering

Problems of an engineer

- Refining existing products
- Designing new products or processes

Statistics in Engineering

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

- Make decisions
- Solve problems
- Design products and processes

1-1 Statistics in Engineering

- 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.

1-1 Statistics in Engineering

Engineering Example

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.

1-1 Statistics in Engineering

Engineering Example

•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.

1-1 Statistics in Engineering

Engineering Example

• 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

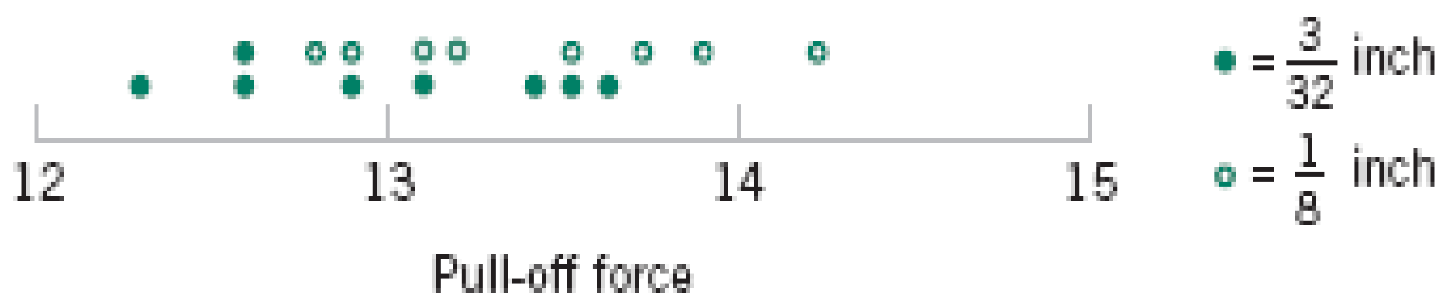


Figure 1-3 Dot diagram of pull-off force for two wall thicknesses.

1-1 Statistics in Engineering

Engineering Example

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

$$X = \mu + \varepsilon$$

where μ is a constant and ε a random disturbance.

1-1 Statistics in Engineering

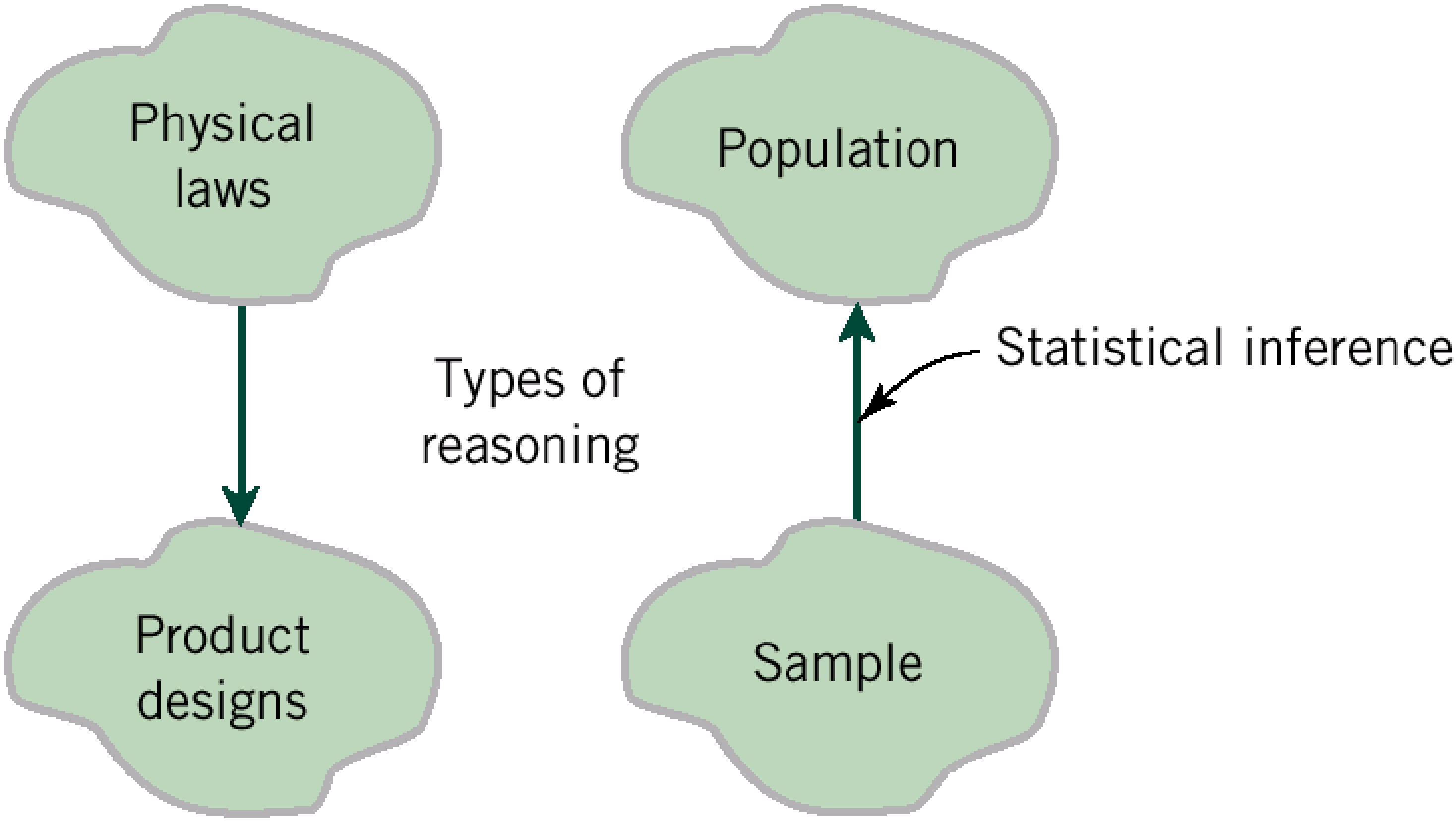


Figure 1-4 Statistical inference is one type of reasoning.

1-2 Collecting Engineering Data

Three basic methods for collecting data:

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

1-2 Collecting Engineering Data

– A **retrospective** study using historical data **would use either all or a sample of the historical process data archived over some period of time.**

1-2 Collecting Engineering Data

– An **observational** study: the engineer observes the process or population, disturbing it

1-2 Collecting Engineering Data

– A **designed experiment**: **the engineer makes deliberate in the controlable variables of the system, observes the resulting system output data**

1-2.5 Observing Processes Over Time

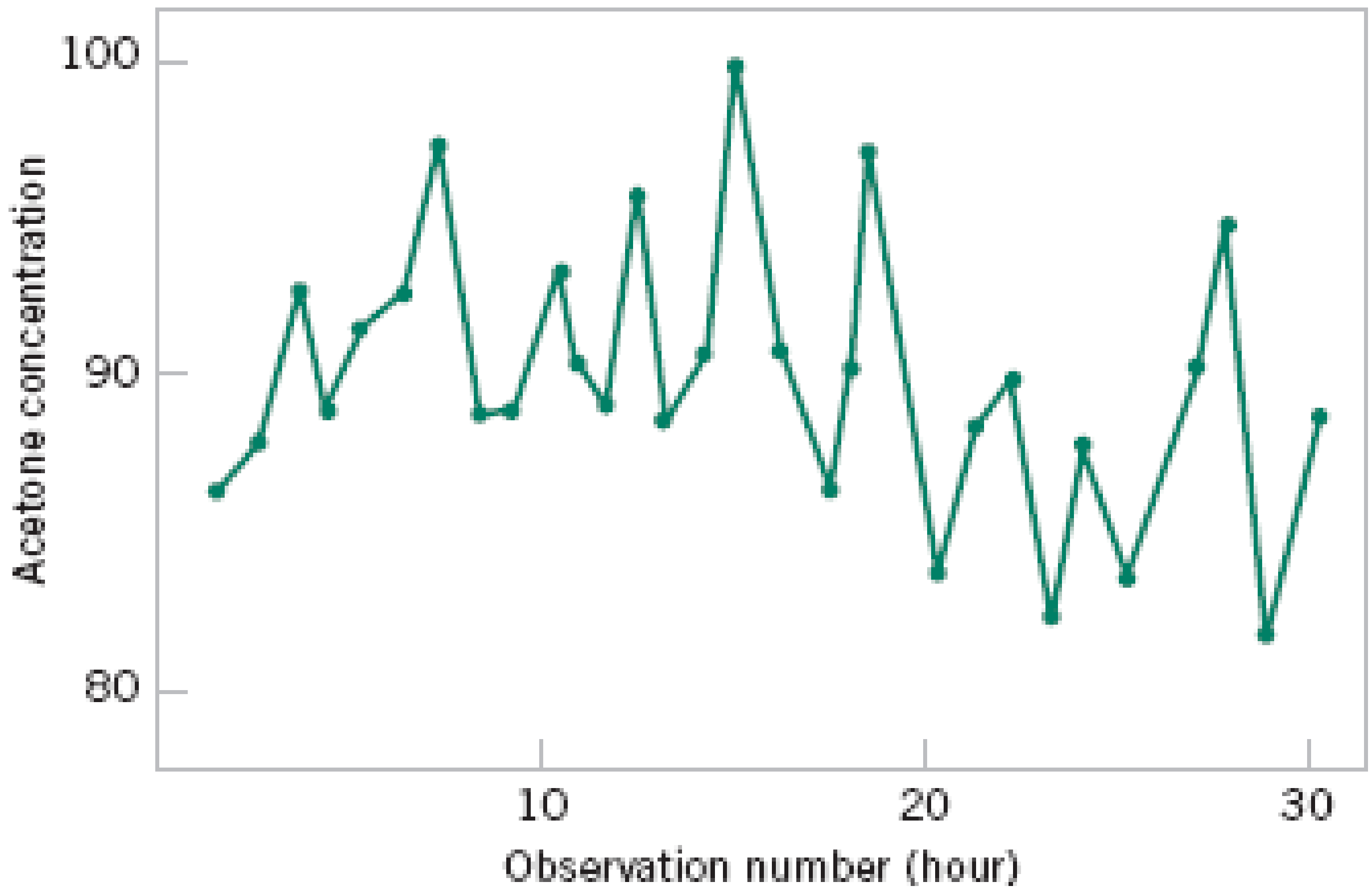


Figure 1-9 A time series plot of concentration provides more information than a dot diagram.

1-2 Observing Processes Over Time

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.

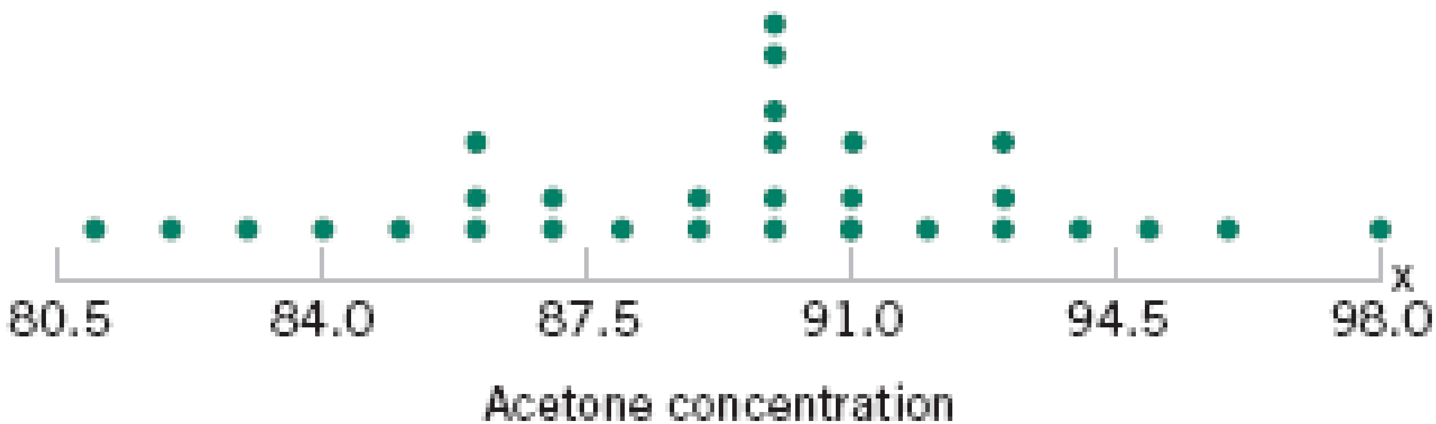


Figure 1-8 The dot diagram illustrates variation but does not identify the problem.

1-2 Observing Processes Over Time

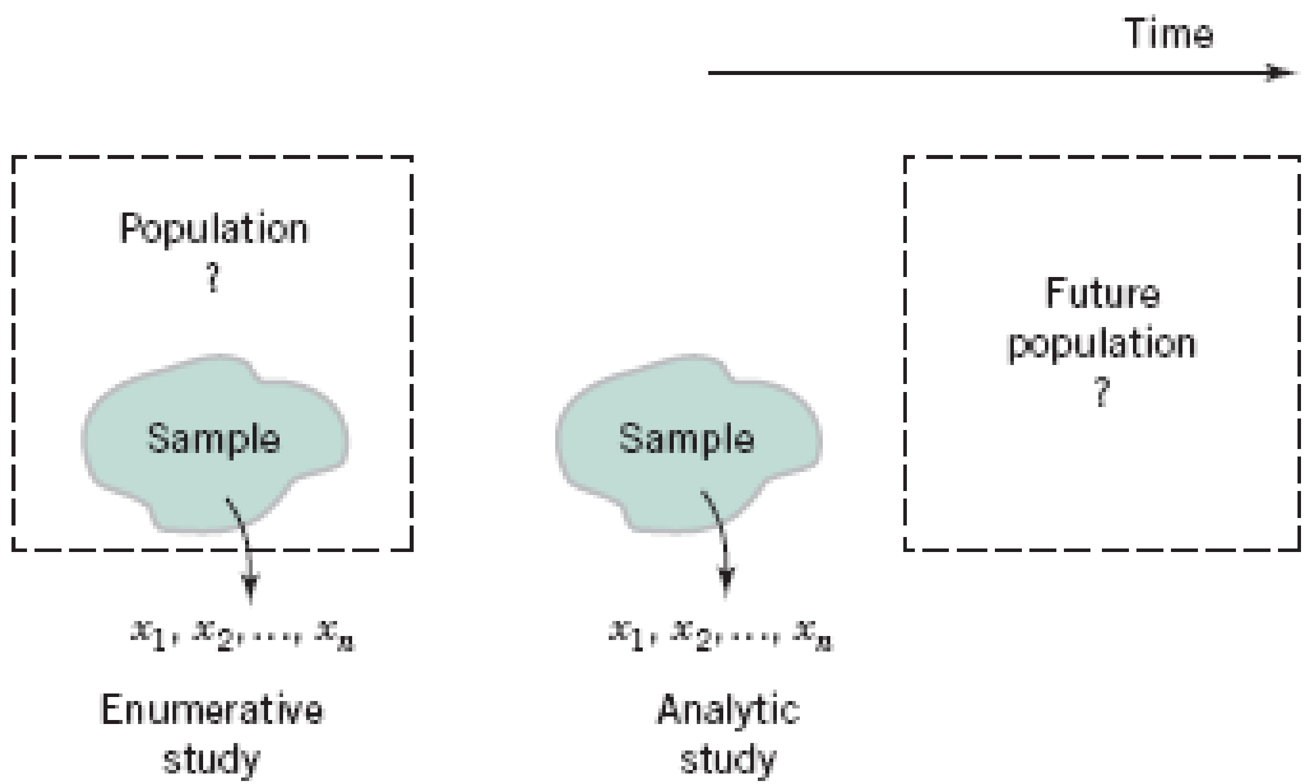


Figure 1-14 Enumerative versus analytic study.

1-3 Mechanistic and Empirical Models

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

Example: Ohm's Law

$$\text{Current} = \text{voltage/resistance}$$

$$I = E/R$$

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

1-3 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.

1-3 Mechanistic and Empirical Models

Example

Suppose we are interested in the number 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.

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

where the β's are unknown parameters.

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

1-3 Mechanistic and Empirical Models

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

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

The **estimated** regression line is given by

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

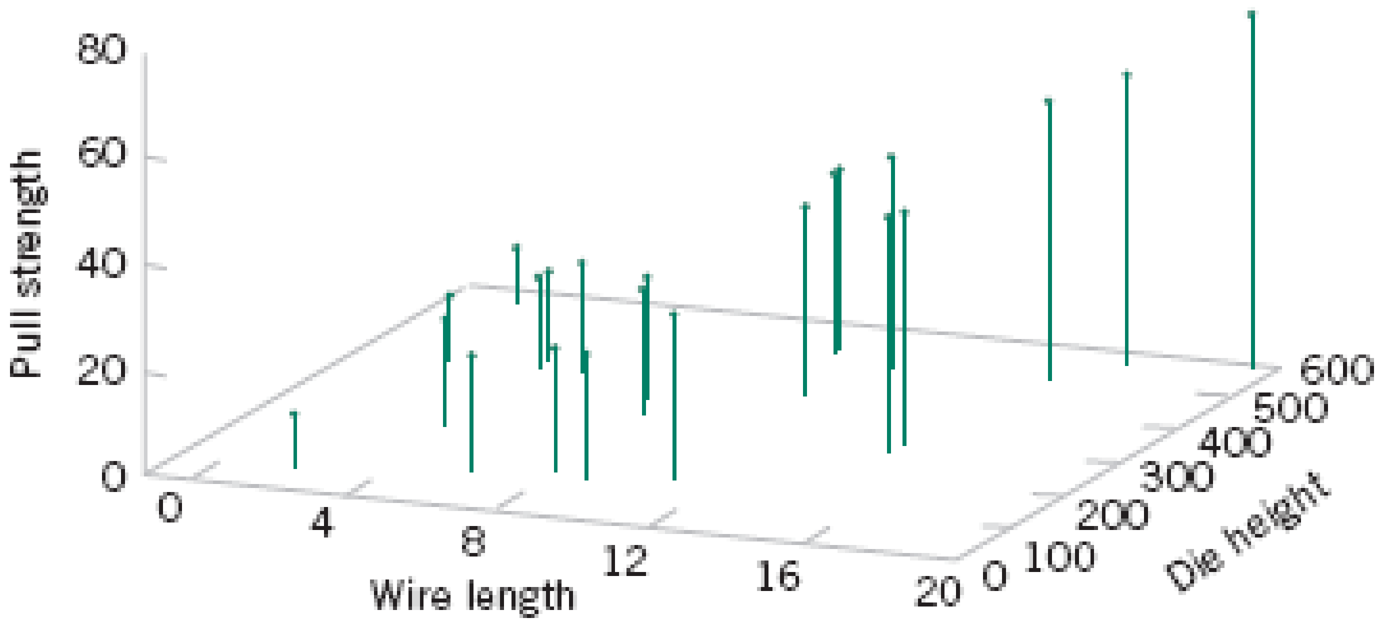


Figure 1-15 Three-dimensional plot of the wire and pull strength data.

1-4 Probability and Probability Models

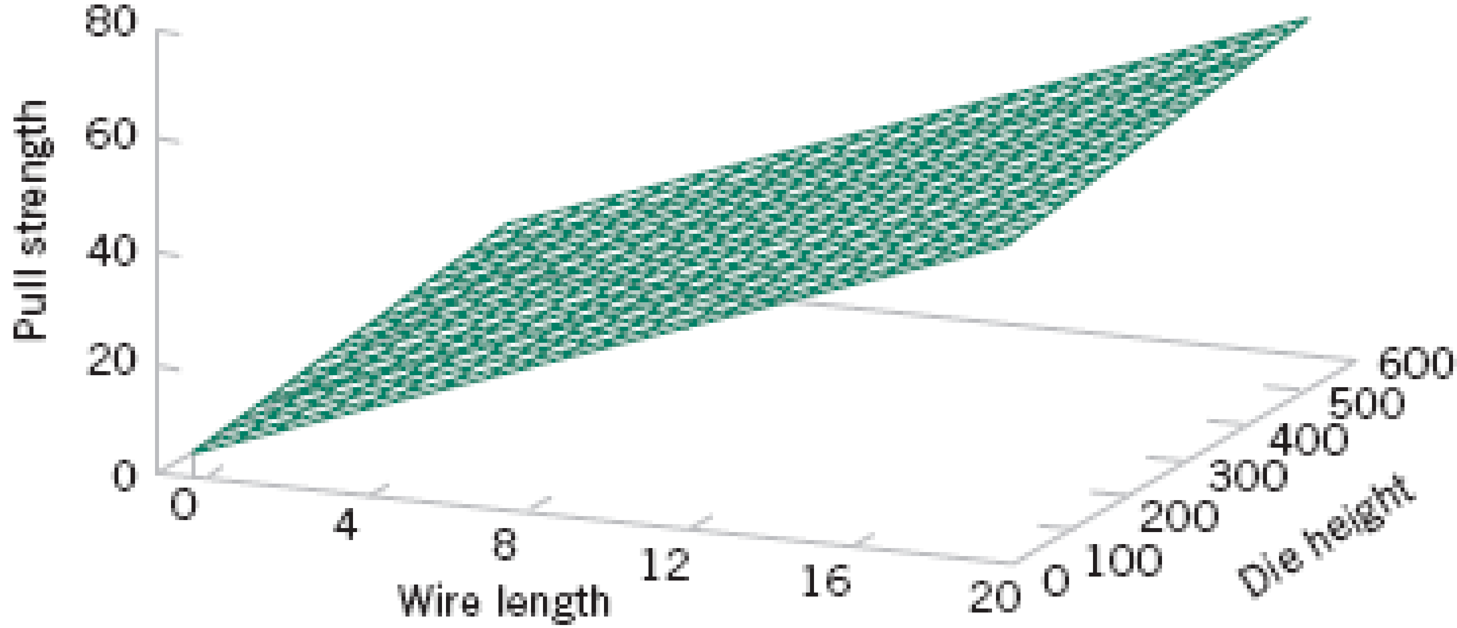


Figure 1-16 Plot of the predicted values of pull strength from the empirical model.

1-4 Probability and Probability Models

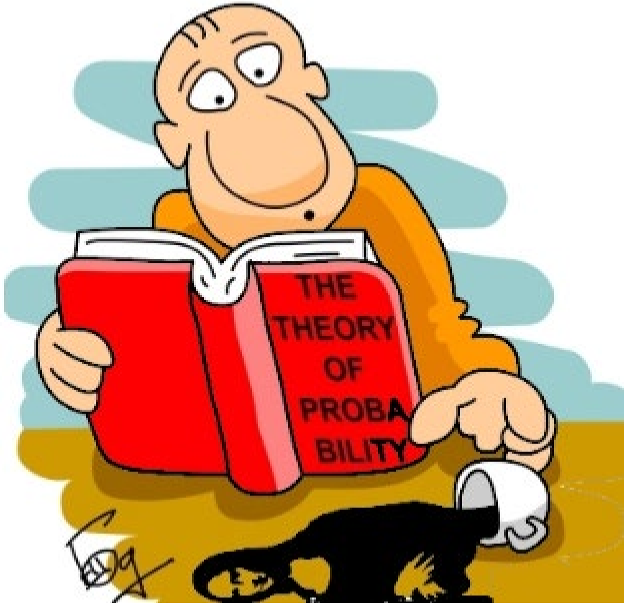
- **Probability models** help quantify the risks involved in statistical inference, that is, risks involved in decisions made every day.

- Probability provides the **framework** for the study and application of statistics.

IMPORTANT TERMS AND CONCEPTS

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

2 Probability



CHAPTER OUTLINE

2-1 SAMPLE SPACES AND EVENTS

- 2-1.1 Random Experiments
- 2-1.2 Sample Spaces
- 2-1.3 Events
- 2-1.4 Counting Techniques

2-2 INTERPRETATIONS OF PROBABILITY

- 2-2.1 Introduction
- 2-2.2 Axioms of Probability

2-3 ADDITION RULES

2-4 CONDITIONAL PROBABILITY

2-5 MULTIPLICATION AND TOTAL PROBABILITY RULES

- 2-5.1 Multiplication Rule
- 2-5.2 Total Probability Rule

2-6 INDEPENDENCE

2-7 BAYES' THEOREM

2-8 RANDOM VARIABLES

LEARNING OBJECTIVES

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

1. Understand and describe sample spaces and events for random experiments with graphs, tables, lists, or tree diagrams
2. Interpret probabilities and use probabilities of outcomes to calculate probabilities of events in discrete sample spaces
3. Use permutation and combinations to count the number of outcomes in both an event and the sample space.
4. Calculate the probabilities of joint events such as unions and intersections from the probabilities of individual events
5. Interpret and calculate conditional probabilities of events
6. Determine the independence of events and use independence to calculate probabilities
7. Use Bayes' theorem to calculate conditional probabilities
8. Understand random variables

2-1 Sample Spaces and Events

2-1.1 Random Experiments

Definition

An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a **random experiment.**

2-1 Sample Spaces and Events

2-1.2 Sample Spaces

Definition

The set of all possible outcomes of a random experiment is called the **sample space** of the experiment. The sample space is denoted as S.

2-1 Sample Spaces and Events

2-1.2 Sample Spaces

Example 2-1

Consider an experiment in which you select a molded plastic part, such as a connector, and measure its thickness. The possible values for thickness depend on the resolution of the measuring instrument, and they also depend on upper and lower bounds for thickness. However, it might be convenient to define the sample space as simply the positive real line

$$S = R^+ = \{x \mid x > 0\}$$

because a negative value for thickness cannot occur.

2-1 Sample Spaces and Events

Example 2-1 (continued)

If it is known that all connectors will be between 10 and 11 millimeters thick, the sample space could be

$$S = \{x \mid 10 < x < 11\}$$

If the objective of the analysis is to consider only whether a particular part is low, medium, or high for thickness, the sample space might be taken to be the set of three outcomes:

$$S = \{low, medium, high\}$$

If the objective of the analysis is to consider only whether or not a particular part conforms to the manufacturing specifications, the sample space might be simplified to the set of two outcomes

$$S = \{yes, no\}$$

that indicate whether or not the part conforms.

2-1 Sample Spaces and Events

Example 2-2

If two connectors are selected and measured, the extension of the positive real line R is to take the sample space to be the positive quadrant of the plane:

$$S = R^+ \times R^+$$

If the objective of the analysis is to consider only whether or not the parts conform to the manufacturing specifications, either part may or may not conform. We abbreviate *yes* and *no* as y and n. If the ordered pair yn indicates that the first connector conforms and the second does not, the sample space can be represented by the four outcomes:

$$S = \{yy, yn, ny, nn\}$$

2-1 Sample Spaces and Events

Example 2-2 (continued)

If we are only interested in the number of conforming parts in the sample, we might summarize the sample space as

$$S = \{0, 1, 2\}$$

As another example, consider an experiment in which the thickness is measured until a connector fails to meet the specifications. The sample space can be represented as

$$S = \{n, yn, yyn, yyyn, yyyyn, \text{and so forth}\}$$

2-1 Sample Spaces and Events

Tree Diagrams

- Sample spaces can also be described graphically with **tree diagrams.**
  - When a sample space can be constructed in several steps or stages, we can represent each of the n_1 ways of completing the first step as a branch of a tree.
  - Each of the ways of completing the second step can be represented as n_2 branches starting from the ends of the original branches, and so forth.

2-1 Sample Spaces and Events

Example 2-3

Each message in a digital communication system is classified as to whether it is received within the time specified by the system design. If three messages are classified, use a tree diagram to represent the sample space of possible outcomes.

Each message can either be received on time or late. The possible results for three messages can be displayed by eight branches in the tree diagram shown in Fig. 2-5.

2-1 Sample Spaces and Events

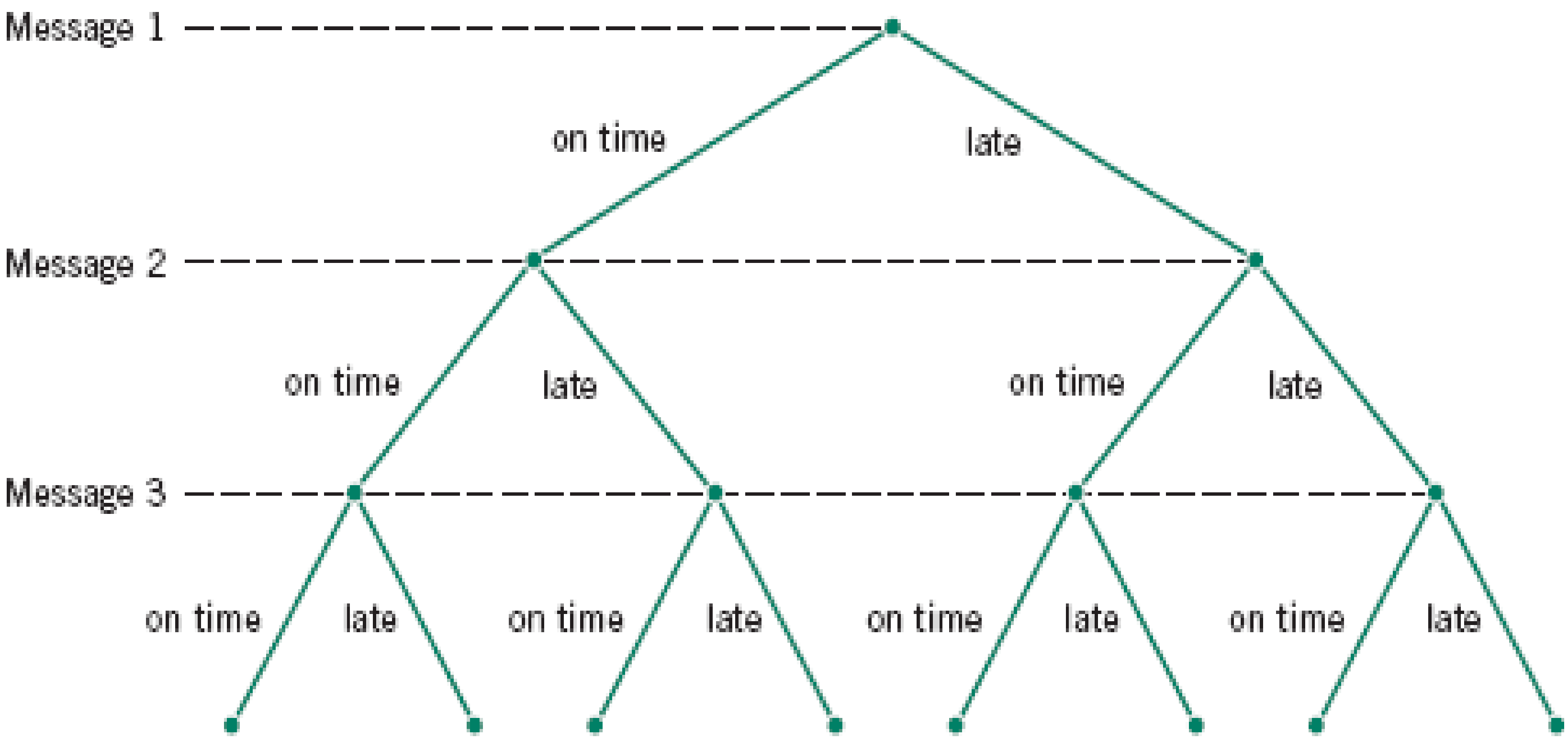


Figure 2-5 Tree diagram for three messages.

2-1 Sample Spaces and Events

2-1.3 Events

Definition

An **event** is a subset of the sample space of a random experiment.

2-1 Sample Spaces and Events

2-1.3 Events

Basic Set Operations

- The **union** of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as $E_1 \cup E_2$.
- The **intersection** of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$.
- The **complement** of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the component of the event E as E'.

2-1 Sample Spaces and Events

2-1.3 Events

Example 2-6

Consider the sample space $S = \{yy, yn, ny, nn\}$ in Example 2-2. Suppose that the set of all outcomes for which at least one part conforms is denoted as E_1. Then,

$$E_1 = \{yy, yn, ny\}$$

The event in which both parts do not conform, denoted as E_2, contains only the single outcome, $E_2 = \{nn\}$. Other examples of events are $E_3 = \varnothing$, the null set, and $E_4 = S$, the sample space. If $E_5 = \{yn, ny, nn\}$,

$$E_1 \cup E_5 = S \qquad E_1 \cap E_5 = \{yn, ny\} \qquad E_1' = \{nn\}$$

2-1 Sample Spaces and Events

Definition

Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \varnothing$$

are said to be **mutually exclusive.**

2-1 Sample Spaces and Events

Venn Diagrams

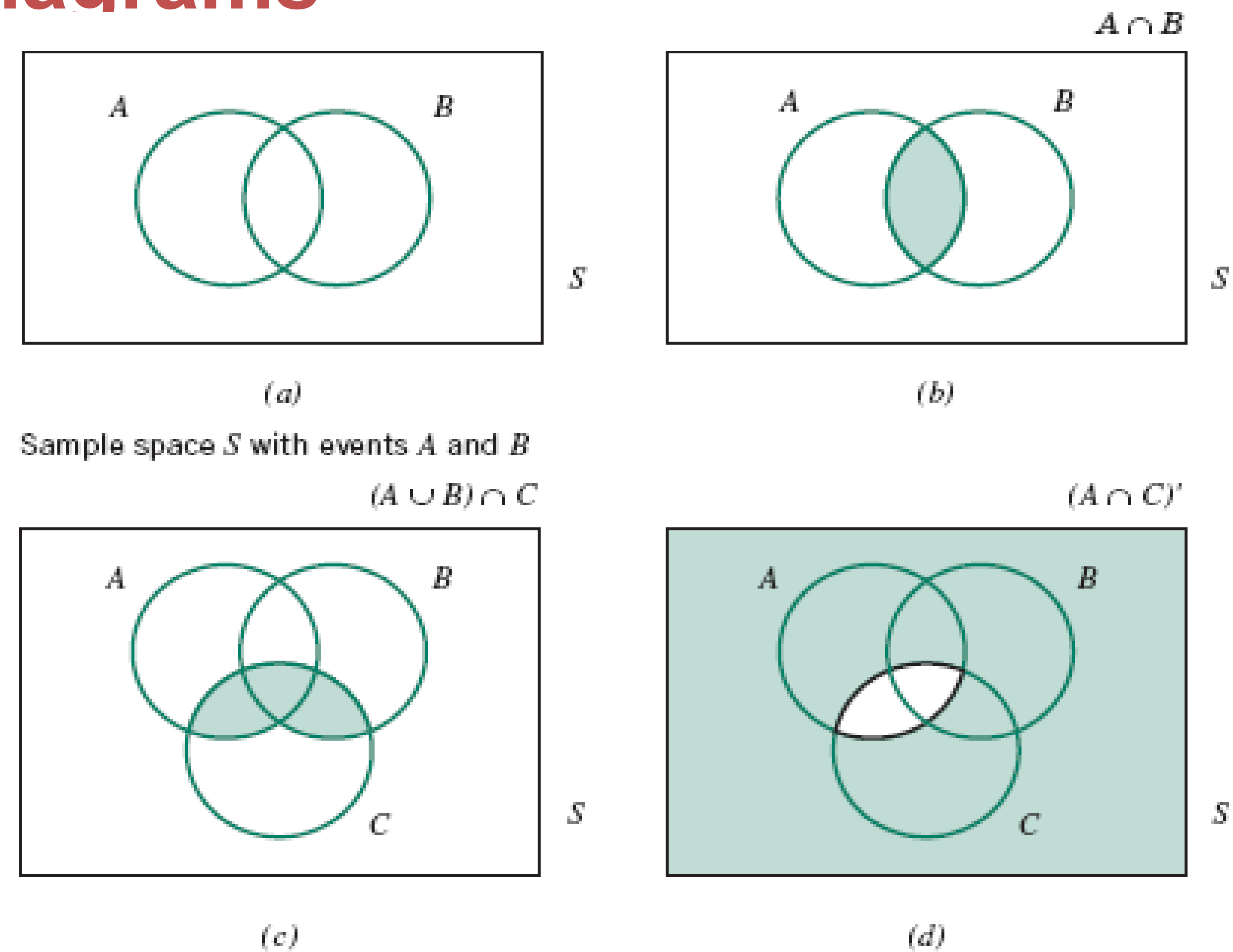


Figure 2-8 Venn diagrams.

2.1.4 Counting Techniques

Multiplication Rule

If an operation can be described as a sequence of k steps, and

If the number of ways of completing step 1 is n_1, and

If the number of ways of completing step 2 is n_2 for each way of completing step 1, and

If the number of ways of completing step 3 is n_3 for each way of completing step 2, and so forth

the total number of ways of completing the operation is

$$\boxed{n_1 \times n_2 \times \ldots \times n_k}$$

EXAMPLE 2_9

In the design of a casing for a gear housing, we can use four different types of fasteners, three different bolt lengths and three different bolt locations. From the multiplication rule, 4x3x3 = 36 different design are possiple.

Permutations

A permutation of the elements is an ordered sequence of the elements. For example, let S ={a,b,c}, then *abc, acb, bac, bca, cab* and *cba* are all of the permutations of the elements of S.

The number of permutations of n different elements is n! where

$$\boxed{n! = n\times(n-1)\times(n-2)\times\ldots\times 2\times 1} \qquad (2-1)$$

Permutations of Subsets

The number of permutations of subsets of r elements selected from a set of n different elements is

$$P_r^n = n \times (n-1) \times (n-2) \times \ldots \times (n-r+1) = \boxed{\frac{n!}{(n-r)!}} \qquad (2-2)$$

EXAMPLE 2 – 10 Printed Circuit Board

A printed circuit board has eight different locations in which a component can be placed. If four different components are to be placed on the board, how many different designs are possible?

Each design consists of selecting a location from the eight lacations for the first component, a location from the remaining seven for the second component, a location from the remaining six for the third component, and a location from the remaining five for the fourth component. Therefore,

$$P_4^8 = \frac{8!}{4!} = 1680$$ different designs are possible.

Permutations of Similar Objects

The number of permutations of $n = n_1 + n_2 + \ldots + n_r$ objects of which n_1 are of one types, n_2 are of a second type, ..., and n_r are of an r^{th} type is

$$\boxed{\frac{n!}{n_1!n_2!..n_r!}} \qquad (2-3)$$

EXAMPLE 2 – 12 Bar Codes

A part is labeled by printing with four thick lines, three medium lines, and two thin lines. If each ordering of the nine lines represents a different label, how many different labels can be generated by using this scheme?

From Equation 2 – 3, the number of possible part labels is

$$\frac{9!}{4!3!2!} = 1260$$

Combinations

The number of combinations, subsets of size r that can be selected from a set of n elements (order is not important), is denoted as $\binom{n}{r}$ or C_r^n and

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \qquad (2-4)$$

EXAMPLE 2 – 13

A printed circuit board has eight different locations in which a component can be placed. If five identical components are to be placed on the board, how many different designs are possible?

Each design is a subset of the eight locations that are to contain the components, so the number of possible designs is

$$\frac{8!}{5!3!} = 56$$

EXAMPLE 2 – 14 (Choosing Problem)

A bin of 50 manufactured parts contains three defective parts and 47 nondefective parts. A sample of six parts is selected from the 50 parts. Selected parts are not replaced. How many different samples are there of six that contain exactly two defective parts?

From the multiplication rule, the number of subsets of size six that contain exactly two defective items is

$$\binom{3}{2}\binom{47}{4} = 535095.$$