DOCUMENT RESUME

ED 033 022 SE 006 883

By-Goodman, A. F.; Blischke, W. R.

Probability and Statistics: A Prelude.

McDonnell Douglas Astronautics Co., Huntington Beach, Calif. Western Div.

Pub Date May 68

Note-39p.

EDRS Price MF-$0.25 HC-$2.05

Descriptors-Inservice Teacher Education, *Instructional Materials, *Probability Theory, *Secondary School Mathematics, *Statistics, Teacher Education

Probability and statistics have become indispensable to scientific, technical, and management progress. They serve as essential dialects of mathematics, the classical language of science, and as instruments necessary for intelligent generation and analysis of information. A prelude to probability and statistics is presented by examination of the important concepts that form their foundation. The brief written discussion of these concepts in outline form is augmented by examples and a bibliography. The outline forms the basis for both a series of lectures to eleventh grade students in a Mathematics Summer Honors Program, and a series of lectures to secondary mathematics teachers in a workshop on probability and statistics. (RP)

DOUGLAS PAPER 4613
MAY 1968

PROBABILITY AND STATISTICS: A PRELUDE

A.F. GOODMAN
W.R. BLISCHKE

Presented to
Workshop on Probability and Statistics
Division of Secondary Education
Los Angeles City School Districts
May 1968

DOUGLAS MISSILE & SPACE SYSTEMS DIVISION

MCDONNELL DOUGLAS
CORPORATION

ERRATA FOR "PROBABILITY AND STATISTICS: A PRELUDE"

1. Page 4, line 1: Underline "countably infinite"

2. Page 4, line 3: Underline "uncountably infinite"

3. Page 8, line 5: Underline "probability"

4. Page 8, line 18: Insert "of E" between "P(E)," and "should"

5. Page 11, line 2 from bottom: Delete ")" between "A_n" and "= S"

6. Page 14, line 12: Insert "fi" between "simpli" and "cations" to produce "simplifications"

7. Page 18, line 9: Delete ", and a and b being constants"

8. Page 18, line 2 from bottom: Change $[E\ (1/n \sum_{i=1}^{n} X_i - \mu)^2]$ to:

$$E\ [(1/n \sum_{i=1}^{n} X_i - \mu)^2]$$

9. Page "27" should be page "28"

10. Page "28" should be page "27"

PROBABILITY AND STATISTICS: A PRELUDE*

A. F. GOODMAN
Senior Technical Staff to Vice President
Information Systems Subdivision
Missile & Space Systems Division
Douglas Aircraft Company
Huntington Beach, California

and

W. R. BLISCHKE
Principal Statistician
Statistical Sciences Department
CEIR, The Professional Services Subsidiary
Control Data Corporation
Beverly Hills, California

ABSTRACT

Probability and statistics have become indispensable to scientific, technical, and management progress; they serve as essential dialects of mathematics, the classical language of science, and as instruments necessary for intelligent generation and analysis of information. Probability evolved from the investigation of gambling problems, and of problems in the analysis of information which contained observational errors. On the other hand, statistics evolved from the satisfaction of a governmental requirement for information, from the parallel and independent development of a framework in which to analyze information, and from the combination of the need for analysis created by the f rmer with the ability to perform analysis provided by the latter.

A prelude to probability and statistics is presented by examination of the important concepts that form their foundation. The brief written discussion of these concepts in outline form is augmented by examples and a bibliography.

*This outline was prepared for use in a cooperative program of the Southern California Chapter of the American Statistical Association, and the Division of Secondary Education in the Los Angeles City School Districts. It forms the basis for both a series of lectures to 11th grade students in the Mathematics Summer Honors Program, and a series of lectures to secondary mathematics teachers in the Workshop on Probability and Statistics.

I. INTRODUCTION

A. Probability and statistics have become indispensable to scientific, technical, and management progress; they serve as essential dialects of mathematics, the classical language of science, and as instruments necessary for intelligent generation and analysis of information.

1. Probability evolved from the investigation of gambling problems, and of problems in the analysis of information which contained observational errors; while statistics evolved from the satisfaction of a governmental requirement for information, from the parallel and independent development of a framework in which to analyze information, and from the combination of the need for analysis created by the former with the ability for analysis provided by the latter.
2. A prelude to probability and statistics is presented by an examination of the important concepts that form their foundation.
3. The brief written discussion of these concepts in outline form is augmented by examples and a bibliography.

B. An event is something that happens; it is an occurrence or an outcome.

1. Two types of events occur: deterministic and random.
2. A deterministic event occurs with certainty, and its occurrence can be predicted or determined in advance.

   a. A coin falls to rest when tossed.

   b. A fruit borne by an apple tree is an apple.

   c. The Earth rotates on its axis every 24 hours.

   d. The length of this table is ________ .

3. A random event does not occur with certainty, and its occurrence cannot be predicted or determined in advance.

   a. Will the tossed coin fall with heads up or tails up?

b. How many apples will be borne by an apple tree this year; and what will be the size, weight, color, flavor, and texture of a selected apple?

c. Will it rain tomorrow, and if so, how much rain will fall?

d. What will be the length of this table, as measured by a selected student?

C. *Probability* and *statistics* deal with random events and the random mechanisms which produce them by characterizing and utilizing the regularity of randomness.

1. *Probability* assumes knowledge concerning a random mechanism, and deduces statements concerning the random events which it produces.

2. *Statistics* assumes knowledge concerning random events, and infers statements concerning the random mechanism which produces them.

3. Based upon the above definitions, probability and statistics are inverse operations of each other.

4. Although Mendel's Law regarding plant genetics is probabilistic, it was inferred statistically from his experiments with sweet peas.

5. Probability and statistics, aided by computer science, are major elements of the modern technology or *modus operandi* of the scientific method (hypothesis to experiment to analysis to inference to hypothesis, and so forth); while the scientific disciplines are the raw material of the scientific method.

6. Currently, the secret of success in applying the scientific method is a dialogue between a scientist and a statistician.

D. Certain mathematical concepts are needed as prerequisites to the discussion of probability and statistics; they are:

1. A set, A, is a collection of elements, x, that are said to be contained in the set A.

   a. $A = \{x_1, x_2, x_3\}$, and $x_1 \varepsilon A$, $x_2 \varepsilon A$, and $x_3 \varepsilon A$.

   b. $A = \{x:$ mathematical statement concerning $x\}$ and $x \varepsilon A$.

2. If S and T are sets, then a function, f, from the domain space, S, into the range space, T, is a relationship that associates one and only one element, t, contained in set, T, with each element, s, contained in set, S.

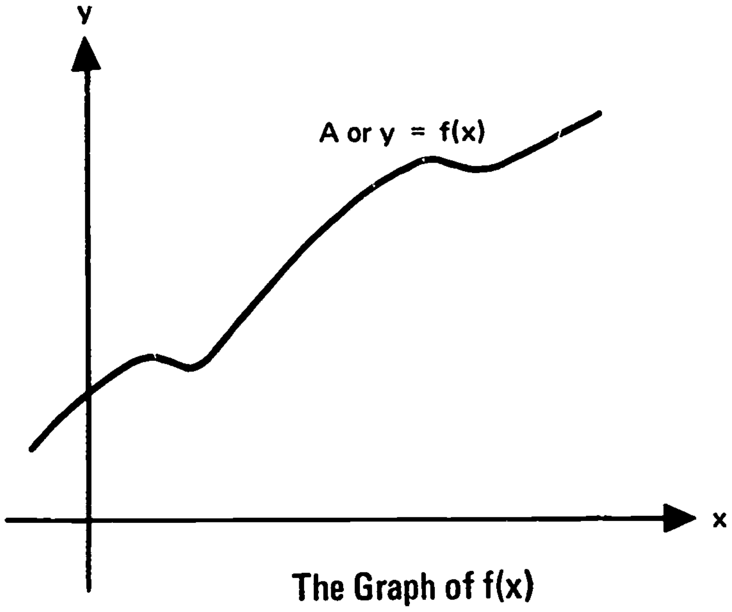
   a. $t = f(s)$.

   b. A function, f, is a function of a real variable if the elements of set, S, are real numbers.

   c. A function, f, is real-valued if the elements of set, T, are real numbers.

   d. A function, f, is a set function if the elements of set, S, are themselves sets.

3. A graph for the real-valued function, f, of the real variable, x, is a pictorial representation of the set, $A = \{(x, y): y = f(x)\}$, on a two-dimensional coordinate system.



The Graph of f(x)

4. The number of elements in a set is countably infinite if the elements are in one-to-one correspondence with the set of positive integers, and it is uncountably infinite if the elements are in one-to-one correspondence with the set of real numbers.

5. If the value of the real-valued function, f, becomes arbitrarily close to b as the value of the real variable, x, becomes sufficiently close to a, then the limit of f as x approaches a is said to be b.

a. $\lim_{x \to a} f(x) = b.$

b. $\lim_{n \to \infty} f_n = b$ (x is n and a is ∞).

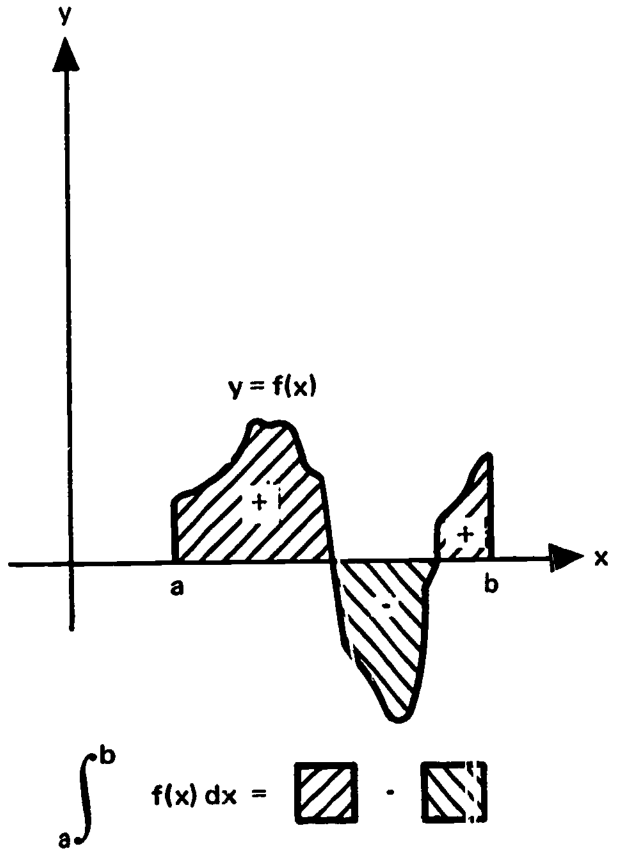
6. If $f_1, f_2, \ldots, f_i, \ldots$ constitute the range of a real-valued function, f, whose domain is the set of ordered positive integers (a sequence of real numbers), then the summation of the numbers, f_i, from the value of i being n to the value of i being N is the sum of the numbers, $f_n, f_{n+1}, \ldots, f_N$.

a. $\sum_{i=n}^{N} f_i = f_n + f_{n+1} + \cdots + f_N.$

b. $\sum_{i=1}^{\infty} f_i = f_1 + f_2 + \cdots + f_i + \cdots$ (n is 1 and N is ∞).

7. If f is a real-valued function of the real variable, x, then the integral of the function, f, from the value of x being a to the value of x being b is the area between the graph of the positive

values of the function, f, and the x-axis; minus the area between the graph of the negative values of the function, f, and the x-axis.



The Integral of f(x)

8. Calculus is the branch of mathematics concerned with these and related concepts for real numbers; topology is the branch of mathematics concerned with set, function, graph, limit, and related concepts in general; and measure theory is the branch of mathematics concerned with summation, integration, and related concepts in general.

DOCUMENT RESUME

ED 033 022 SE 006 883

By-Goodman, A. F.; Blischke, W. R.
Probability and Statistics: A Prelude.
McDonnell Douglas Astronautics Co., Huntington Beach, Calif. Western Div.
Pub Date May 68
Note-39p.
EDRS Price MF-$0.25 HC-$2.05
Descriptors-Inservice Teacher Education, *Instructional Materials, *Probability Theory, *Secondary School Mathematics, *Statistics, Teacher Education

Probability and statistics have become indispensable to scientific, technical, and management progress. They serve as essential dialects of mathematics, the classical language of science, and as instruments necessary for intelligent generation and analysis of information. A prelude to probability and statistics is presented by examination of the important concepts that form their foundation. The brief written discussion of these concepts in outline form is augmented by examples and a bibliography. The outline forms the basis for both a series of lectures to eleventh grade students in a Mathematics Summer Honors Program, and a series of lectures to secondary mathematics teachers in a workshop on probability and statistics. (RP)

DOUGLAS PAPER 4613
MAY 1968

PROBABILITY AND STATISTICS: A PRELUDE

A.F. GOODMAN
W.R. BLISCHKE

Presented to
Workshop on Probability and Statistics
Division of Secondary Education
Los Angeles City School Districts
May 1968

DOUGLAS MISSILE & SPACE SYSTEMS DIVISION

MCDONNELL DOUGLAS
CORPORATION

ERRATA FOR "PROBABILITY AND STATISTICS: A PRELUDE"

1.	Page 4, line 1:	Underline "countably infinite"
2.	Page 4, line 3:	Underline "uncountably infinite"
3.	Page 8, line 5:	Underline "probability"
4.	Page 8, line 18:	Insert "of E" between "P(E)," and "should"
5.	Page 11, line 2 from bottom:	Delete ")" between "A_n" and "= S"
6.	Page 14, line 12:	Insert "fi" between "simpli" and "cations" to produce "simplifications"
7.	Page 18, line 9:	Delete ", and a and b being constants"
8.	Page 18, line 2 from bottom:	Change $[E\ (1/n \sum_{i=1}^{n} X_i - \mu)^2]$ to: $E\ [(1/n \sum_{i=1}^{n} X_i - \mu)^2]$
9.	Page "27" should be page "28"	
10.	Page "28" should be page "27"	

PROBABILITY AND STATISTICS: A PRELUDE*

A. F. GOODMAN
Senior Technical Staff to Vice President
Information Systems Subdivision
Missile & Space Systems Division
Douglas Aircraft Company
Huntington Beach, California

and

W. R. BLISCHKE
Principal Statistician
Statistical Sciences Department
CEIR, The Professional Services Subsidiary
Control Data Corporation
Beverly Hills, California

ABSTRACT

Probability and statistics have become indispensable to scientific, technical, and management progress; they serve as essential dialects of mathematics, the classical language of science, and as instruments necessary for intelligent generation and analysis of information. Probability evolved from the investigation of gambling problems, and of problems in the analysis of information which contained observational errors. On the other hand, statistics evolved from the satisfaction of a governmental requirement for information, from the parallel and independent development of a framework in which to analyze information, and from the combination of the need for analysis created by the former with the ability to perform analysis provided by the latter.

A prelude to probability and statistics is presented by examination of the important concepts that form their foundation. The brief written discussion of these concepts in outline form is augmented by examples and a bibliography.

*This outline was prepared for use in a cooperative program of the Southern California Chapter of the American Statistical Association, and the Division of Secondary Education in the Los Angeles City School Districts. It forms the basis for both a series of lectures to 11th grade students in the Mathematics Summer Honors Program, and a series of lectures to secondary mathematics teachers in the Workshop on Probability and Statistics.

I. INTRODUCTION

A. Probability and statistics have become indispensable to scientific, technical, and management progress; they serve as essential dialects of mathematics, the classical language of science, and as instruments necessary for intelligent generation and analysis of information.

   1. Probability evolved from the investigation of gambling problems, and of problems in the analysis of information which contained observational errors; while statistics evolved from the satisfaction of a governmental requirement for information, from the parallel and independent development of a framework in which to analyze information, and from the combination of the need for analysis created by the former with the ability for analysis provided by the latter.
   2. A prelude to probability and statistics is presented by an examination of the important concepts that form their foundation.
   3. The brief written discussion of these concepts in outline form is augmented by examples and a bibliography.

B. An event is something that happens; it is an occurrence or an outcome.

   1. Two types of events occur: deterministic and random.
   2. A deterministic event occurs with certainty, and its occurrence can be predicted or determined in advance.

      a. A coin falls to rest when tossed.

      b. A fruit borne by an apple tree is an apple.

      c. The Earth rotates on its axis every 24 hours.

      d. The length of this table is __________ .

   3. A random event does not occur with certainty, and its occurrence cannot be predicted or determined in advance.

      a. Will the tossed coin fall with heads up or tails up?

b. How many apples will be borne by an apple tree this year; and what will be the size, weight, color, flavor, and texture of a selected apple?

c. Will it rain tomorrow, and if so, how much rain will fall?

d. What will be the length of this table, as measured by a selected student?

C. Probability and statistics deal with random events and the random mechanisms which produce them by characterizing and utilizing the regularity of randomness.

1. Probability assumes knowledge concerning a random mechanism, and deduces statements concerning the random events which it produces.

2. Statistics assumes knowledge concerning random events, and infers statements concerning the random mechanism which produces them.

3. Based upon the above definitions, probability and statistics are inverse operations of each other.

4. Although Mendel's Law regarding plant genetics is probabilistic, it was inferred statistically from his experiments with sweet peas.

5. Probability and statistics, aided by computer science, are major elements of the modern technology or modus operandi of the scientific method (hypothesis to experiment to analysis to inference to hypothesis, and so forth); while the scientific disciplines are the raw material of the scientific method.

6. Currently, the secret of success in applying the scientific method is a dialogue between a scientist and a statistician.

D. Certain mathematical concepts are needed as prerequisites to the discussion of probability and statistics; they are:

1. A set, A, is a collection of elements, x, that are said to be contained in the set A.

   a. $A = \{x_1, x_2, x_3\}$, and $x_1 \varepsilon A$, $x_2 \varepsilon A$, and $x_3 \varepsilon A$.

   b. $A = \{x\text{: mathematical statement concerning } x\}$ and $x \varepsilon A$.

2. If S and T are sets, then a function, f, from the domain space, S, into the range space, T, is a relationship that associates one and cnly one element, t, contained in set, T, with each element, s, contained in set, S.

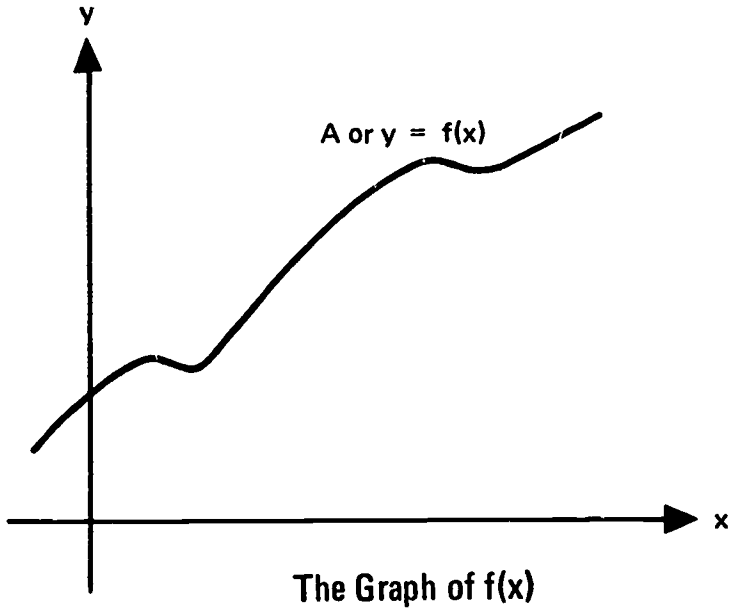
   a. $t = f(s)$.

   b. A function, f, is a function of a real variable if the elements of set, S, are real numbers.

   c. A function, f, is real-valued if the elements of set, T, are real numbers.

   d. A function, f, is a set function if the elements of set, S, are themselves sets.

3. A graph for the real-valued function, f, of the real variable, x, is a pictorial representation of the set, $A = \{(x, y)\text{: } y = f(x)\}$, on a two-dimensional coordinate system.



The Graph of f(x)

4. The number of elements in a set is countably infinite if the elements are in one-to-one correspondence with the set of positive integers, and it is uncountably infinite if the elements are in one-to-one correspondence with the set of real numbers.

5. If the value of the real-valued function, f, becomes arbitrarily close to b as the value of the real variable, x, becomes sufficiently close to a, then the limit of f as x approaches a is said to be b.

   a. $\lim_{x \to a} f(x) = b.$

   b. $\lim_{n \to \infty} f_n = b$ (x is n and a is ∞).

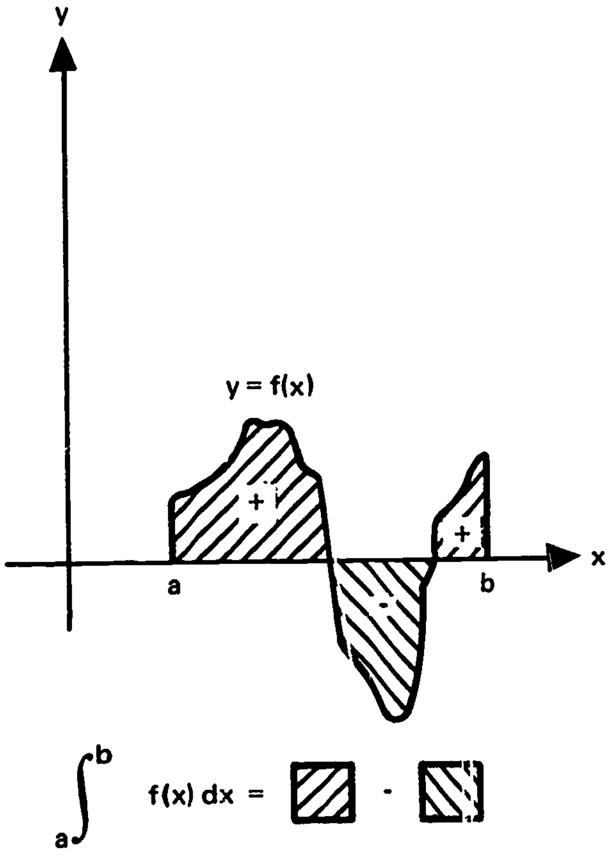
6. If $f_1, f_2, \ldots, f_i, \ldots$ constitute the range of a real-valued function, f, whose domain is the set of ordered positive integers (a sequence of real numbers), then the summation of the numbers, f_i, from the value of i being n to the value of i being N is the sum of the numbers, $f_n, f_{n+1}, \ldots, f_N$.

   a. $\sum_{i=n}^{N} f_i = f_n + f_{n+1} + \cdots + f_N.$

   b. $\sum_{i=1}^{\infty} f_i = f_1 + f_2 + \cdots + f_i + \cdots$ (n is 1 and N is ∞).

7. If f is a real-valued function of the real variable, x, then the integral of the function, f, from the value of x being a to the value of x being b is the area between the graph of the positive

values of the function, f, and the x-axis; minus the area between the graph of the negative values of the function, f, and the x-axis.



The Integral of f(x)

8. Calculus is the branch of mathematics concerned with these and related concepts for real numbers; topology is the branch of mathematics concerned with set, function, graph, limit, and related concepts in general; and measure theory is the branch of mathematics concerned with summation, integration, and related concepts in general.

II. PROBABILITY

A. Probability characterizes the uncertainty associated with random events by expressing the regularity of randomness in mathematical terms or numerical form.

1. Although the occurrence of a random event cannot be predicted and repeated trials of a random mechanism do not yield identical results regarding a random event, a large collection of such results does possess characteristics which are predictable in the long run.

2. This long-run predictability for characteristics associated with a random event or a random mechanism is what is meant by the regularity of randomness.

3. The uncertainty associated with a random event is characterized by expressing the likelihood or probability of occurrence of the event as a number between zero and one inclusive ($0 \leq$ probability ≤ 1).

4. Parzen* characterizes probability theory as the study of random phenomena and mathematical models of random phenomena.

5. As examples, consider the random events listed in I. B. 3 above.

B. Because probability is a mathematically primitive notion and is difficult (if not impossible) to define precisely and rigorously, the concept of probability will be characterized rather than defined.

1. Events are mutually exclusive if the occurrence of any one of them precludes or prevents the occurrence of all the others.

*E. Parzen. Modern Probability Theory and its Applications. John Wiley and Sons, Inc., 1960, Pages 1 and 5.

2. Events are equally likely, if each is as apt to occur as any other.

3. Classical characterization: If an event can occur in exactly N mutually exclusive and equally likely ways and M of these ways have an attribute, A, then the probability, P(A), of A should be M/N.

4. A preferable characterization of probability employs the notion of a random mechanism, called a random experiment.

   a. If the outcome of an experiment, ξ, with possible outcomes, E_α, is not predictable, then ξ is called a random experiment; E_α is called a simple event; any "meaningful" collection, E, of simple events is called an event; and the collection, S, of all simple events is called the sample space.

   b. Empirical characterization: If ξ is performed n times and the event, E, occurs m of these times, then the empirical probability, $P_n(E)$, of E should be m/n; and the probability, P(E), should be characterized as the limit, in some sense that exists, of $P_n(E)$ as n becomes infinite.

   c. Axiomatic characterization: Let there be associated with each event, E, a number, P(E), called the probability of E and having the following characteristics (which are the characteristics of $P_n(E)$ for a fixed n):

      (i) $P(E) \geqq 0$.

      (ii) $P(S) = 1$.

      (iii) $P(E) = P(E_1) + P(E_2)$, where E_1 and E_2 are two events which contain no common simple events (are mutually exclusive) and constitute E when combined.

d. A set of operating rules for probability, called the calculus of probability, may be derived by the application of mathematical logic to the axiomatic characterization of probability.

e. In this calculus of probability, the law of large numbers states that $P_n(E)$ essentially approaches $P(E)$ as n becomes infinite (specifically:

$$\lim_{n \to \infty} P\{|P_n(E) - P(E)| \geqq \varepsilon\} = 0 \text{ for } \varepsilon > 0, \text{ the proof}$$

of which is given under II. G. 13).

f. Hence, the empirical and axiomatic characterizations of probability are connected (the former being concerned with the calculation of probability; and the latter being concerned with the characterization of, and consequent operating rules for, probability) by making the requirement that P(E) be a number tl 7t satisfies the limit equation.

g. The ultimate measure of the "goodness" or "reasonableness" of the random-mechanism characterization of probability (the empirical and axiomatic characterizations of probability, and their connection) is the extent to which the resulting calculus of probability is applicable to real problems.

5. Probability is a function whose domain consists of events, and whose range is the interval, $0 \leqq P \leqq 1$ (a real-valued set function).

C. Given a characterization of the concept of probability, some definitions and notation that underlie the calculus of probability are introduced.

1. A diagram that depicts events and relationships among events in the sample space is called a Venn diagram.

2. S denotes the sample space, and $A_1, A_2, \ldots, A_n$ denote events.

3. $A_1, A_2, \ldots, A_n$ are exhaustive if at least one of them is certain to occur.

4. The empty or null event, Φ, is the event which is composed of no simple events.

5. The event which is composed of all simple events that are not contained in A_1 is called the complement of A_1, and denoted by A_1^c.

6. The event which is composed of all simple events that are simultaneously contained in $A_1, A_2, \ldots A_n$ is denoted by $A_1 \cap A_2 \cap \cdots \cap A_n$, and is called the intersection of $A_1, A_2, \ldots, A_n$.

7. $A_1 \cup A_2 \cup \cdots \cup A_n$ denotes the union of $A_1, A_2, \ldots, A_n$, which is the event composed of all simple events that are contained in at least one of $A_1, A_2, \ldots, A_n$.

8. If all simple events which are contained in A_1 also are contained in A_2, then A_1 is said to be contained in A_2; and this is denoted by $A_1 \subseteq A_2$.

9. $A_1 | A_2$ (read A_1, given A_2 has occurred) means to consider that part of A_1 which is contained in A_2 as an event in the restricted sample space, A_2.

10. $A_1, A_2, \ldots, A_n$ are independent if the occurrence of any one of them has no effect on the probability of occurrence of any other one.

D. Consequences of these definitions and notation regarding events (which are, of course, sets) are easily derived.

1. The group of events, $A_1, A_2, \ldots, A_n$, is mutually exclusive if and only if no two of them intersect--that is

$$A_i \cap A_j = \Phi \text{ for } i = 1, 2, \ldots, n, \text{ and } j = i + 1, i + 2, \ldots, n.$$

2. Some relationships among events are:

a. $(A_1^c)^c = A_1$.

b. $A_1 \cap A_1^c = \Phi$ and $A_1 \cup A_1^c = S$.

c. $A_1 \cap \Phi = \Phi$ and $A_1 \cup \Phi = A_1$.

d. $A_1 \cap A_2 = A_2 \cap A_1$ and $A_1 \cup A_2 = A_2 \cup A_1$.

e. $A_1 \cap (A_2 \cap A_3) = (A_1 \cap A_2) \cap A_3$ and
$A_1 \cup (A_2 \cup A_3) = (A_1 \cup A_2) \cup A_3$.

f. $A_1 \cap (A_2 \cup A_3) = (A_1 \cap A_2) \cup (A_1 \cap A_3)$ and
$A_1 \cup (A_2 \cap A_3) = (A_1 \cup A_2) \cap (A_1 \cup A_3)$.

g. $(A_1 \cap A_2)^c = A_1^c \cup A_2^c$ and $(A_1 \cup A_2)^c = A_1^c \cap A_2^c$.

h. $A_1 \cap A_2$ is that part of A_1 which is also contained in A_2, and is contained in both A_1 and A_2; and $A_1 \cap A_2^c$ is that part of A_1 which is not contained in A_2, and is contained in A_1.

i. $A_1 \cap A_2^c$, $A_1^c \cap A_2$, and $A_1 \cap A_2$ are mutually exclusive and $A_1 \cup A_2 = (A_1 \cap A_2^c) \cup (A_1^c \cap A_2) \cup (A_1 \cap A_2)$.

j. $A_1 \cap A_2^c \cap A_3^c$, $A_1^c \cap A_2 \cap A_3^c$, $A_1^c \cap A_2^c \cap A_3$, $A_1 \cap A_2 \cap A_3^c$, $A_1 \cap A_2^c \cap A_3$, $A_1^c \cap A_2 \cap A_3$, and $A_1 \cap A_2 \cap A_3$ are mutually exclusive; and $A_1 \cup A_2 \cup A_3 = (A_1 \cap A_2^c \cap A_3^c)$
$\cup (A_1^c \cap A_2 \cap A_3^c) \cup (A_1^c \cap A_2^c \cap A_3) \cup (A_1 \cap A_2 \cap A_3^c)$
$\cup (A_1 \cap A_2^c \cap A_3) \cup (A_1^c \cap A_2 \cap A_3) \cup (A_1 \cap A_2 \cap A_3)$.

k. As an example, consider $S = \{1, 2, 3, 4, 5, 6\}$, $A_1 = \{1, 2\}$, $A_2 = \{2, 4, 6\}$, and $A_3 = \{5\}$.

3. $A_1, A_2, \ldots, A_n$ are exhaustive if $A_1 \cup A_2 \cup \cdots \cup A_n) = S$, and if and only if $P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1$.

E. The following elementary results serve as examples of the calculus of probability:

1. $P(\Phi) = 0$ and $P(S) = 1$.

2. If $A_1 \subseteqq A_2$, then $P(A_1) \leq P(A_2)$.

3. $P(A_1^c) = 1 - P(A_1)$ and $P(A_1 \cap A_2^c) = P(A_1) - P(A_1 \cap A_2)$.

4. $P(A_1 \cup A_2 \cup \cdots \cup A_n) \leqq \sum_{i=1}^{n} P(A_i)$, with equality holding if and only if $A_1, A_2, \ldots, A_n$ are mutually exclusive.

5. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

6. $$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3).$$

7. If $P(A_2) > 0$, then
$$P(A_1 | A_2) = P(A_1 \cap A_2)/P(A_2) \text{ or}$$
$$P(A_1 \cap A_2) = P(A_1 | A_2)\, P(A_2).$$

8. Conditional probabilities satisfy the same relationships as (unconditional) probabilities, so long as all conditional probabilities are defined.

9. A_1 and A_2 are independent if and only if $P(A_1 \cap A_2) = P(A_1)\, P(A_2)$; whereas, $P(A_1 | A_2) = P(A_1)$ (or $P(A_2 | A_1) = P(A_2)$) if A_1 and A_2 are independent and $P(A_2) > 0$ (or $P(A_1) > 0$).

10. Mutual exclusiveness or lack of it is a property of events which is solely determined by the properties of events as sets; independence or lack of it is a property of events which is solely determined by the properties of the probability function defined over them.

11. Bayes Theorem: If $A_1, A_2, \ldots, A_{n-1}$ are mutually exclusive and A_n can occur only if one of $A_1, A_2, \ldots, A_{n-1}$ occurs ($A_n \subseteq A_1 \cup A_2 \cup \cdots \cup A_{n-1}$), then

$$P(A_n) = \sum_{i=1}^{n-1} P(A_n \cap A_i) = \sum_{i=1}^{n-1} P(A_n | A_i)\, P(A_i) \text{ and}$$

$$P(A_j | A_n) = P(A_n | A_j)\, P(A_j) / \sum_{i=1}^{n-1} P(A_n | A_i)\, P(A_i) \text{ for}$$

$$j = 1, 2, \ldots, n-1.$$

12. It is informative to reconsider the above example, with $P(x) = 1/6$ for $x \varepsilon S$.

F. The following comments concerning the calculation of probabilities are a complement to the above results concerning the manipulation of probabilities:

1. If $S = \{E_1, E_2, \ldots, E_n\}$, all E_i's are equally likely ($P(E_i) = 1/n$ for $i = 1, 2, \ldots, n$), and $A = \{E_1, E_2, \ldots, E_m\}$, then the size of A (number of E_i's which constitute A) is m and $P(A) = m/n$.

2. When previously selected objects are (are not) replaced before the next selection, the sampling procedure is called sampling with (without) replacement.

3. If an event, A_1, can occur in N_1 ways, for each of these ways a different event, A_2, can occur in N_2 ways, ..., and for each of these ways a different event, A_m, can occur in N_m ways, then the m events, $A_1, A_2, \ldots, A_m$, can occur in a combination in $N_1 N_2 \cdots N_m$ ways.

4. The number of permutations, $(n)_k$, of n objects selected k at a time is the number of arranged (with regard to order of selection) ways that k objects can be selected from a group of n objects.

5. The number of combinations, $\binom{n}{k}$, of n objects selected k at a time is the number of unarranged (with regard to order of selection) ways that k objects can be selected from a group of n objects.

6. If n is a positive integer, then the product of n, (n-1), ..., 2, 1, is called n factorial and written!. For convenience, the symbol, 0!, is defined to be 1.

7. Then $(n)_k = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$ and $\binom{n}{k} = (n)_k/k!$

$$= \frac{n!}{k!(n-k)!} = \binom{n}{n-k}.$$

8. Select the most accurate and complete sample space for the problem before attempting to compute probabilities, and be careful; simplications and shortcuts are justified only if they produce the same results.

9. Two illustrative examples are provided by the following problems:

   a. Calculate the probability of obtaining two heads and one tail in three tosses of a fair coin, with and without employing combinations.

   b. Calculate the probability of rolling a five when two fair dice are rolled.

G. Finally, additional probabilistic concepts are useful prerequisites for the discussion of statistics.

1. A random variable, X, is a real-valued function defined over the sample space, S, of a random experiment, ξ.

2. The probability associated with a value of the random variable, X, is the probability associated with the corresponding outcome(s) of the random experiment.

3. A discrete random variable, X, assumes only a countable number of isolated values, x_1, x_2, ...; a continuous random variable assumes every value in some interval.

4. The probability distribution of a random variable, X, is a description of the distribution of probability over the values associated with the random variable, X.

5. The cumulative distribution function, F, of a random variable, X, represents the probability that the random variable, X, does not exceed a given value, x:

$$F(x) = P(X \leqq x).$$

6. The probability density function, f, of a random variable, X, represents the probability that the random variable, X, assumes a given value, x:

   a. For a discrete random variable, X, $f(x) = P(X = x)$.

   b. For a continuous random variable, X, $f(x)\,dx \approx P(x < X \leqq x + dx)$.

7. If h is a real-valued function of a real variable and X is a random variable, then the expected value, $E[h(X)]$, of the random variable, $h(X)$, is the weighted average of the function, h, with respect to the probability distribution of the random variable, X.

   a. For a discrete random variable, X,

$$E[h(X)] = \sum_{i=1}^{\infty} h(x_i)\, f(x_i).$$

   b. For a continuous random variable, X,

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx.$$

8. The expected value, $E(X)$, of a random variable, X, is called its mean, μ; and μ may be thought of as the "center of probability mass."

9. The expected value, $E[(X - \mu)^2]$, of the squared deviation of a random variable, X, from its mean, μ, is called its variance, σ^2; and σ^2 may be thought of as the "moment of probability inertia."

10. The binomial distribution with parameters, $n = 1, 2, \ldots$ and $0 \leqq p \leqq 1$, has

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for $x = 0, 1, \ldots, n$; and it is an example of a probability distribution for a discrete random variable.

11. The normal distribution with parameters, $-\infty < \mu < \infty$ and $\sigma > 0$, has

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x - \mu)^2/2\sigma^2} \quad \text{for } -\infty < x < \infty;$$

and it is an example of a probability distribution for a continuous random variable.

12. Chebyshev's inequality: If X is a discrete random variable with mean, μ, and variance, σ^2, then

$$P(|X - \mu| \geqq t) \leqq \sigma^2/t^2 \text{ for } t > 0;$$

and the proof of this follows:

a. By definition,

$$\sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i).$$

b. Grouping x_1, x_2, ... into two sets, $A_1 = \{x_i : |x_i - \mu| < t\}$ and $A_2 = \{x_i : |x_i - \mu| \geq t\}$, produces

$$\sigma^2 = \sum_{x_i \varepsilon A_1} (x_i - \mu)^2 f(x_i) + \sum_{x_i \varepsilon A_2} (x_i - \mu)^2 f(x_i).$$

c. Because $(x_i - \mu)^2 \geqq 0$ and $f(x_i) \geqq 0$,

$$\sum_{x_i \varepsilon A_1} (x_i - \mu)^2 f(x_i) \geqq 0 \text{ and}$$

$$\sigma^2 \geqq \sum_{x_i \varepsilon A_2} (x_i - \mu)^2 f(x_i).$$

d. By definition,

$$\sum_{x_i \varepsilon A_2} (x_i - \mu)^2 f(x_i) \geqq t^2 \sum_{x_i \varepsilon A_2} f(x_i) \text{ and}$$

$$\sum_{x_i \varepsilon A_2} f(x_i) = P(|X - \mu| \geqq t).$$

e. Then

$$\sigma^2 \geqq t^2 P(|X - \mu| \geqq t).$$

f. Dividing both sides of this inequality by t^2 yields

$$P(|X - \mu| \geqq t) \leqq \sigma^2/t^2 \text{ for } t > 0.$$

13. Law of large numbers: If X_1, X_2, ..., X_n are independent, identically distributed, and discrete random variables with mean, μ, and variance, σ^2, and

$$\overline{X}_n = 1/n \sum_{i=1}^{n} X_i, \text{ then}$$

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| \geqq \varepsilon) = 0 \text{ for } \varepsilon > 0;$$

and the proof of this is given by:

a. If the discrete random variable, X, has probability density function, f, and a is a constant, then $E(a) = a$ and $E(aX) = aE(X)$; and its proof follows.

(i) By definition, $E(a) = a(1) = a$.

(ii) Also by definition,

$$E(aX) = \sum_{i=1}^{\infty} ax_i f(x_i) = a \sum_{i=1}^{\infty} x_i f(x_i) = aE(X).$$

b. It can be shown that $E(X + Y) = E(X) + E(Y)$ for X and Y being discrete random variables, and a and b being constants.

c. The combination of the above two properties of expectation implies that

$$E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i)$$

for $X_1, X_2, \ldots, X_n$ being discrete random variables and $a_1, a_2, \ldots, a_n$ being constants.

d. Also it can be demonstrated that $E[(X - \mu)(Y - \nu)] = 0$ for X and Y being independent and discrete random variables with means, μ and ν.

e. Using these properties of expectation produces

(i) $E(\overline{X}_n) = E(1/n \sum_{i=1}^{n} X_i) = 1/n \sum_{i=1}^{n} E(X_i) = 1/n\,(n\mu) = \mu.$

(ii) $E[(\overline{X}_n - \mu)]^2 = [E\ (1/n \sum_{i=1}^{n} X_i - \mu)^2]$

$$= E\{[1/n(\sum_{i=1}^{n} X_i - n\mu)]^2\} = E[1/n^2 (\sum_{i=1}^{n} X_i - n\mu)^2]$$

$$= 1/n^2 E\,[(\sum_{i=1}^{n} X_i - n\mu)^2] = 1/n^2\, E\,\{[\sum_{i=1}^{n} (X_i - \mu)]^2\}$$

$$= 1/n^2\, E\,[\sum_{i=1}^{n} (X_i - \mu)^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} (X_i - \mu)(X_j - \mu)]$$

$$= 1/n^2\, E\,[\sum_{i=1}^{n} (X_i - \mu)^2] + E\,[2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} (X_i - \mu)(X_j - \mu)]$$

$$= 1/n^2 \sum_{i=1}^{n} E[(X_i - \mu)^2] + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} E[(X_i - \mu)(X_j - \mu)]$$

$$= 1/n^2\,(n\sigma^2) + 2\,(0) = \sigma^2/n.$$

f. Since $X_1, X_2, \ldots, X_n$ are discrete random variables, $\overline{X}_n$ is also a discrete random variable.

g. Hence, $\overline{X}_n$ is a discrete random variable with mean, μ, and variance, σ^2/n.

h. By Chebyshev's inequality with t replaced by ε.

$$P(|\overline{X}_n - \mu| \geqq \varepsilon) \leqq \sigma^2/n\varepsilon^2 \text{ for } \varepsilon > 0.$$

i. Because $\sigma^2/n\varepsilon^2$ approaches zero as n approaches infinity, $\lim\limits_{n \to \infty} P(|X_n - \mu| \geqq \varepsilon) = 0$ for $\varepsilon > 0$.

14. The form of the law of large numbers employed in II. B. 4. e is a special case of the preceding result, in which:

a. Each X_i is 0 when the event, E, does not occur, and is 1 when the event, E, does occur.

b. Then

$$P(X_i = 0) = 1 - P(E) \text{ and}$$

$$P(X_i = 1) = P(E) \text{ for } i = 1, 2, \ldots, n.$$

c. In addition, $\overline{X}_n = P_n(E)$ and $\mu = P(E)$.

III. STATISTICS

A. Statistics is concerned mainly with the proper generation or collection of, description of, and inference based upon random data which represent outcomes of random experiments.

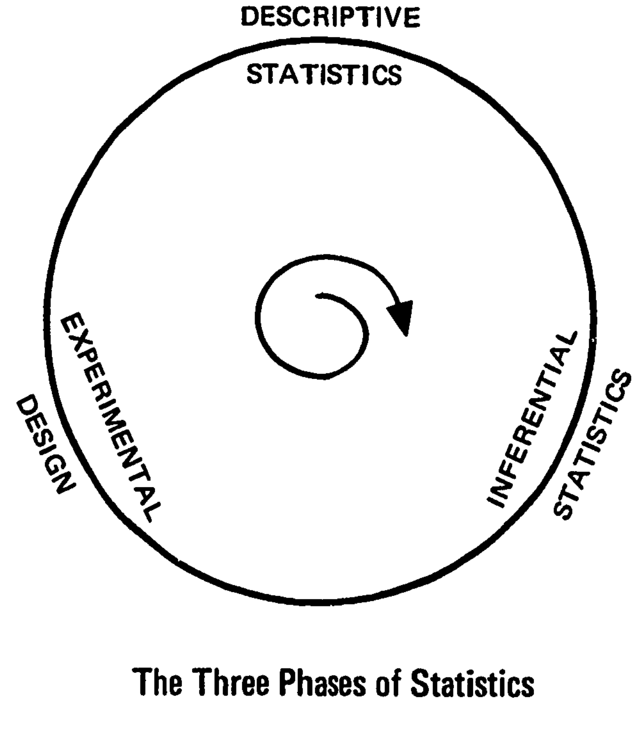
1. Descriptive statistics involves the description and summarization of random data.

2. Inferential statistics involves the drawing of inferences in the presence of uncertainty resulting from randomness or incomplete information; these inferences are based upon data, and serve as a basis for rational and objective decisions.

3. Experimental design involves the efficient performance of meaningful random experiments for either description or inference.

4. These three phases of statistics typically follow one another in an iterative cycle.

a. With the help of a scientist, the statistician first describes the data which he is investigating, then draws an inference based upon the data, and finally



The Three Phases of Statistics

designs additional experiments to investigate certain aspects of the random mechanism more thoroughly.

b. After these experiments are conducted, the second cycle begins with an analysis of the data from these experiments, and so forth.

B. Areas in which Statistics is Applied.

1. Four general areas in which statistics is applied are scientific research, government, management, and daily life.

2. Four stages occur in the evolution of a discipline from an art into a science: description, modeling, prediction, and control and optimization--Adolphe Quetelet, who may be viewed as the first modern statistician, said that "we can judge of the perfection to which a science has come by the facility more or less great, with which it may be approached by calculation."*

a. The first stage in the development of a discipline is the collection, structuring, and description of information by scientists.

(i) Because a discipline cannot even begin without such information or observations, technological advances are often required first (as for astronomy, which had progressed very little for thousands of years, until the invention of the telescope).

(ii) Examples of disciplines which currently are mainly in the descriptive stage are most social sciences

*H. M. Walker. Studies in the History of Statistical Method. The Williams and Wilkins Company, 1929, Page 39.

(except economics and psychology), oceanography, and information science.

(iii) Scientists in these fields use mainly descriptive statistics, but also draw some inferences and use some principles of experimental design.

b. Given a sufficient amount of structuring and description, scientists then construct and estimate models of relationships among component parts of the structured information.

(i) Models may be physical, electromechanical analog, mathematical, or digital computer.

(ii) Models are required generally for any type of analysis, and are required particularly for prediction, or control and optimization.

(iii) Examples of disciplines which currently are mainly in the modeling stage are most biological sciences (except agricultural science and genetics), medical science, management science, and psychology.

(iv) Scientists in these fields use mainly modeling techniques and inferential statistics.

c. Through the use of models (predominanatly mathematical, but increasingly electromechanical analog and digital computer), scientists are able to predict and simulate experiments.

(i) Because the accuracy and precision of such predictions and simulations depend upon the accuracy and precision of the models, a discipline may approach this stage of development when crude models exist, but only enters it when refined and representative models are constructed.

(ii) Examples of disciplines which currently are mainly in the prediction stage are economics, astronomy, engineering, meteorology, and genetics.

(iii) Scientists in these fields use mainly inferential statistics.

d. The final stage in the development of a discipline is reached when it is able to provide both the methods and the technology for control, improvement, or optimization.

(i) Examples of disciplines which currently are at least partially in the control and optimization stage are the physical and agricultural sciences (except in backward countries).

(ii) Scientists in these fields use mainly mathematical, electromechanical analog, and digital computer models, as well as inferential statistics.

3. Long a user of economic, labor, medical, and population statistics, government is becoming an increasing user of expanding types of statistical data, philosophy, and techniques.

4. Statistical philosophy and techniques are increasingly used as an aid to important management decisions.

5. Each of us is required to make many decisions every day (for example, should I take an umbrella this morning?); although most do not require formal analysis (in fact, many are subjective rather than objective), an understanding of statistical philosophy and techniques can be quite useful in many ways: interpretation of advertising, playing games of chance, comprehending scientific reports in the press,

understanding weather and other predictions, following the stock market and its analysis, and so forth.

C. Descriptive Statistics

1. Descriptive statistics concerns the collection and description of both numerical data (such as heights of boys 14 years old, or the number of voters preferring a specified candidate) and non-numerical data (such as whether a given brand of toothpaste tastes good, fair, or poor; or tastes better than, or not as good as, a competing brand).

   a. Although this collection of data is generally the layman's concept of statistics, statisticians are rarely involved in the collection of data; but they are involved in its analysis and interpretation, or in experimental design.

   b. To the statistician and the experimenter, the major use of descriptive statistics is to provide a basis for inferential statistics.

2. Concepts and Techniques

   a. A population is the collection of possible values for a random variable, X, which may be very large or even infinite (in this sense, the term refers to a conceptual, rather than an existing collection).

   b. A parameter, θ, is a measurable characteristic of the random variable, X, which frequently appears explicitly in the mathematical formula for its probability distribution.

   c. A sample is a collection of observed values, $X_1, X_2, \ldots, X_n$, for the random variable, X.

   d. A statistic is a measurable characteristic of the sample, $X_1, X_2, \ldots, X_n$.

   e. Thus, sample is to population as statistic is to parameter.

f. Probability assumes knowledge concerning the probability distribution of a random variable, X, and deduces statements concerning a sample, $X_1, X_2, \ldots, X_n$; whereas, statistics assumes knowledge concerning a sample, $X_1, X_2, \ldots, X_n$, and infers statements concerning the probability distribution of the random variable, X.

g. A histogram is a graph describing the distribution of probability over the sample, $X_1, X_2, \ldots, X_n$.

h. Measures of location are used to describe the location of the center of a sample probability distribution: if $X_1, X_2, \ldots, X_n$ is a sample, then the sample mean,

$$\bar{X} = (\sum_{i=1}^{n} X_i)/n,$$

is their average; the sample median is a value, $\tilde{X}$, such that one-half of the observations are greater than or equal to $\tilde{X}$, and one-half of them are less than or equal to $\tilde{X}$; and the sample mode is the most frequently occurring value in the sample (that is, the highest point of the histogram).

i. Measures of dispersion are used to describe the extent to which the sample probability distribution is dispersed over its range of values: if $X_1, X_2, \ldots, X_n$ is a sample, then the sample variance,

$$s^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n - 1),$$

especially for very large populations which change with time (when collection and analysis of such vast amounts of data may be so time-consuming that the population will have changed by the time the results are available).

c. Although inferential statistics has traditionally been viewed as being composed of theoretical statistics and applied statistics, a type of statistics, called developmental statistics (as in the research, development and production cycle), has been emerging between them.

(i) Theoretical statistics deals with research, mainly mathematical, on statistical methodology; and it concerns both investigation of the properties of existing methodologies, and derivation of new methodologies for the construction of efficient experimental designs, as well as for the efficient extraction and use of information contained in the data.

(ii) Developmental statistics deals with development of statistical techniques from existing methodologies for application to the design of experiments for, and the analysis and interpretation of data from, scientific or other investigations; this generally requires a familiarity with both theoretical and applied statistics.

(iii) Applied statistics deals with application of the best available statistical techniques to design experiments for, and analyze and interpret data from, scientific or other investigations; this is generally accomplished by the applied statistician working directly with the experimenter.

is essentially the average of $(X_1 - \overline{X})^2$, $(X_2 - \overline{X})^2$, ..., $(X_n - \overline{X})^2$; while the sample range is the difference between the largest and the smallest values in the sample.

D. Inferential Statistics

1. Inferential statistics involves the drawing of inferences in the presence of uncertainty resulting from randomness or incomplete information; these inferences are based upon data, and serve as a basis for rational and objective decisions.

   a. In a sense, probability and statistics, both of which are concerned with the concept of randomness, look in opposite directions from this central point: probability seeks to construct mathematical models of randomness, while statistics seeks to accomplish objective decision-making in the presence of randomness--much of which depends upon probability.

   b. Among the reasons for not obtaining complete information in a given problem are:

      (i) Complete information may be too expensive, expecially if the population is very large or resources are limited (as they usually are).

      (ii) Complete information may be impossible to obtain, especially for an infinite population (all possible corn plants cannot be grown).

      (iii) Complete information may be useless, especially if the population is used up (as in the testing of bullets).

      (iv) Complete information may be less accurate (although more precise) than partial information,

2. It is frequently both convenient and informative to view inferential statistics as a game against nature, in which the objective is to infer the true state of nature (probability distribution defined over the population) from the experimental data, so that rational and objective decisions may be made in the presence of uncertainty.

   a. The moves in this game are:

      (i) Nature selects a state of nature (population probability distribution) from the family of possible states of nature.

      (ii) The experimenter (with the statistician) precisely defines the population under study, including his population model for the family of possible states of nature, and formulates the objectives of the investigation (including theories, criteria and constraints).

      (iii) The statistician (with the experimenter) designs the experiment to achieve the objectives of the investigation.

      (iv) The experimenter performs the experiment and collects the data.

      (v) The statistician describes the data.

      (vi) The statistician infers the true state of nature, based upon the description of the data and the population model.

      (vii) The experimenter (with the statistician) makes a decision, based upon the inferred state of nature, and he sustains a loss (where a win is considered a negative loss), which depends upon the decision and the true state of nature.

(viii) In actuality, the second move frequently is postponed until after the fourth one, and the third move frequently is omitted; however, this unfortunate fact has often forced the development of new methodologies and techniques.

b. Estimation is a statistical inference that produces statements about the population model, usually in terms of its parameter(s).

(i) An estimator, $\hat{\theta}(X_1, X_2, \ldots, X_n)$, of the parameter, θ, is a recipe (or function) for combining the observed values, $X_1, X_2, \ldots, X_n$, in the sample to estimate the parameter.

(ii) An estimate, $\hat{\theta}(x_1, x_2, \ldots, x_n)$, is a particular outcome of the recipe (or value of the function), $\hat{\theta}(X_1, X_2, \ldots, X_n)$, for a given sample, $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$.

(iii) An interval estimator, $[\hat{\theta}_L(X_1, X_2, \ldots, X_n), \hat{\theta}_U(X_1, X_2, \ldots, X_n)]$, of the parameter, θ, is a pair of recipes, $\hat{\theta}_L(X_1, X_2, \ldots, X_n)$ and $\hat{\theta}_U(X_1, X_2, \ldots, X_n)$, for combining the observed values, $X_1, X_2, \ldots, X_n$, in the sample to estimate the lower and upper ends of an interval, $\hat{\theta}_L \leqq \theta \leqq \hat{\theta}_U$, that contains θ.

(iv) A confidence interval is the combination of an interval estimator and the amount of confidence placed in it, as measured by the probability, $1 - \alpha$, of its being correct:

(a) $P[\hat{\theta}_L(X_1, X_2, \ldots, X_n) \leqq \theta \leqq \hat{\theta}_U(X_1, X_2, \ldots, X_n)]$
$= 1 - \alpha$;

(b) $P[\theta \geqq \hat{\theta}_L(X_1, X_2, \ldots, X_n)] = 1 - \alpha$,
where $\hat{\theta}_U(X_1, X_2 \ldots, X_n)$is ∞; and

(c) $P[\theta \leqq \hat{\theta}_U(X_1, X_2, \ldots, X_n)] = 1 - \alpha$,
where $\hat{\theta}_L(X_1, X_2, \ldots, X_n)$ is $-\infty$.

c. Testing is a statistical inference that tests one hypothesis concerning the population model against a second hypothesis concerning it, usually in terms of its parameter(s).

(i) A statistical hypothesis, H, is an hypothesis concerning the parameter(s) of the population model.

(ii) The null hypothesis, H_0, is the statistical hypothesis which is to be tested against the alternative hypothesis, H_1:

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \grave{\theta}_0,$$

$$H_0 : \theta \leqq \theta_0 \text{ versus } H_1 : \theta > \theta_0, \text{ and}$$

$$H_0 : \theta \geqq \theta_0 \text{ versus } H_1 : \theta < \theta_0.$$

(iii) A test of the null hypothesis, H_0, against the alternative hypothesis, H_1, is the specification of all samples that are sufficiently critical of H_0, or have a sufficiently low probability of occurring if H_0 were true, for H_0 to be rejected in favor of H_1.

(iv) The null hypothesis, H_0, will then be rejected when the sample, $X_1, X_2, \ldots, X_n$, is sufficiently critical of it; and will be accepted when the sample, $X_1, X_2, \ldots, X_n$, is not sufficiently critical of it.

(v) The two types of errors that can be made are an error of rejecting the null hypothesis, H_0, when it is true, and an error of accepting H_0 when it is false.

d. Regression analysis treats statistical inference regarding the representation of a composite variable as a combination of component variables plus an error; and it is basic to many statistical applications, especially those concerned with the modeling, prediction, and control and optimization stages in the evolution of a discipline from an art into a science.

(i) The necessary notation is now introduced.

(a) t represents time or some other auxiliary variable.

(b) X_1 (t), X_2 (t), . . . , X_p (t) denote p component (factor, input or independent) variables at time, t.

(c) β_1, β_2, . . . , β_p are unspecified constants (coefficients).

(d) Y (t) denotes the composite (response, output or dependent) variable at time, t.

(e) ε (t) represents the error contained in the representation of Y (t).

(f) $(X_{1i}, X_{2i}, \ldots, X_{pi}, Y_i) = [X_1 (t_i), X_2 (t_i), \ldots, X_p (t_i), Y (t_i)]$ symbolizes corresponding observations of the p component variables and the composite variable at time, t_i.

(g) $f (X_1 (t), X_2 (t), \ldots, X_p (t); \beta_1, \beta_2, \ldots, \beta_p)$ denotes the combination of X_1 (t), X_2 (t), . . . , X_p (t).

(h) $\varepsilon_i = Y_i - f(X_{1i}, X_{2i}, \ldots, X_{pi}; \beta_1, \beta_2, \ldots, \beta_p)$ is an unknown, but implicit, observation of $\varepsilon(t)$ at time, t_i.

(ii) There are three equivalent statements of the regression problem.

(a) Physical statement: Based upon $(X_{1i}, X_{2i}, \ldots, X_{pi}, Y_i)$ for $i = 1, 2, \ldots, n$, obtain the optimum representation of Y (t) in the form,

$$Y(t) = f(X_1(t), X_2(t), \ldots, X_p(t); \beta_1, \beta_2, \ldots, \beta_p) + \varepsilon(t).$$

(b) Statistical statement: Based upon $(X_{1i}, X_{2i}, \ldots, X_{pi}, Y_i)$ for $i = 1, 2, \ldots, n$, obtain the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ in the representation of Y (t) in the form,

$$Y(t) = f(X_1(t), X_2(t), \ldots, X_p(t); \beta_1, \beta_2, \ldots, \beta_p) + \varepsilon(t).$$

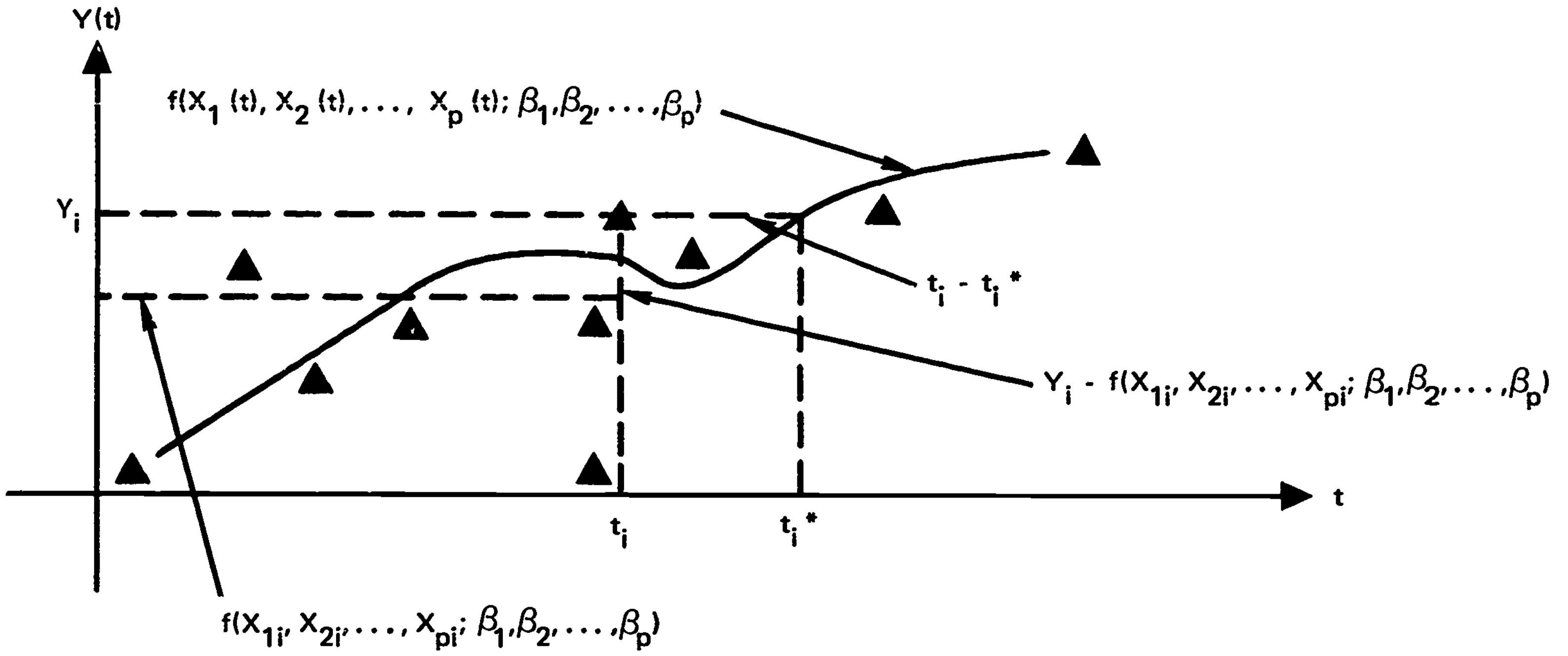
(c) Graphical statement: Based upon (t_i, Y_i) for $i = 1, 2, \ldots, n$, obtain the optimum representation of Y (t) in the form,

$$Y(t) = f(X_1(t), X_2(t), \ldots, X_p(t); \beta_1, \beta_2, \ldots, \beta_p) + \varepsilon(t).$$

(iii) Although many criteria may be utilized in selecting an optimum solution to the regression problem (one might select that solution which minimizes such quantities as

$$\sum_{i=1}^{n} | t_i - t_i^* |$$ for horizontal deviations,

$$\sum_{i=1}^{n} [Y_i - f(X_{1i}, X_{2i}, \ldots, X_{pi}; \beta_1, \beta_2, \ldots, \beta_p)]^2$$ for vertical deviations,


Y(t)
$f(X_1(t), X_2(t), \ldots, X_p(t); \beta_1, \beta_2, \ldots, \beta_p)$
Y_i
$t_i - t_i^*$
$Y_i - f(X_{1i}, X_{2i}, \ldots, X_{pi}; \beta_1, \beta_2, \ldots, \beta_p)$
t
t_i
t_i^*
$f(X_{1i}, X_{2i}, \ldots, X_{pi}; \beta_1, \beta_2, \ldots, \beta_p)$

$t_i - t_i^*$ = HORIZONTAL DEVIATION OF (t_i, Y_i) FROM THE CORRESPONDING POINT ON A REPRESENTATION OF Y(t) IN THE FORM, Y(t) = $f(X_1(t), X_2(t), \ldots, X_p(t); \beta_1, \beta_2, \ldots, \beta_p) + \epsilon(t)$

$Y_i - f(X_{1i}, X_{2i}, \ldots, X_{pi}; \beta_1, \beta_2, \ldots, \beta_p)$ = VERTICAL DEVIATION OF (t_i, Y_i) FROM THE CORRESPONDING POINT ON A REPRESENTATION OF Y (t) in THE FORM, Y (t) = $f(X_1(t), X_2(t), \ldots, X_p(t); \beta_1, \beta_2, \ldots, \beta_p) + \epsilon(t)$

The Regression Problem

or $\lambda \max \left| t_i - t_i^* \right| + (1 - \lambda)$ $\max \left| Y_i - f(X_{1i}, X_{2i}, \ldots, X_{pi}; \beta_1, \beta_2, \ldots, \beta_p) \right|$ for a linear combination of horizontal and vertical deviations), the least squares criterion selects that solution which minimizes

$$Q = \sum_{i=1}^{n} [Y_i - f(X_{1i}, X_{2i}, \ldots, X_{pi}; \beta_1, \beta_2, \ldots, \beta_p)]^2.$$

(a) Because Q is a standard mathematical measure of squared distance, the least squares criterion is meaningful mathematically.

(b) The least squares criterion also is relatively simple to implement mathematically.

(c) The appropriateness of the least squares criterion is determined actually by the probability distribution of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$.

(d) Adrien Marie Legendre first published the least squares criterion in 1806, but Carl Fredrich Gauss first formulated it in an unpublished manuscript of 1802.

(iv) When the combination of component variables is linear ($f(X_1 (t), X_2 (t), \ldots, X_p (t); \beta_1, \beta_2, \ldots, \beta_p)$ $= \sum_{j=1}^{p} \beta_j X_j (t)$), the solution of the regression problem is considerably easier to obtain mathematically.

(v) As an example, consider $Y (t) = \beta_1 + \beta_2 X (t) + \varepsilon (t)$ (that is, for all t: $X_1 (t) = 1$, $X_2 (t) = X (t)$, and $X_j (t) = 0$ for $j = 3, 4, \ldots, p$).

BIBLIOGRAPHY[*]

H. L. Alder and E. B. Roessler. Introduction to Probability and Statistics, 3rd Edition. W. H. Freeman and Co., 1964.

W. J. Dixon and F. J. Massey. Introduction to Statistical Analysis, 2nd Edition. McGraw-Hill Book Co., 1957.

H. F. Fehr, et al. Introduction to Sets, Probability and Hypothesis Testing. D. C. Heath and Co., 1964.

W. Feller. Introduction to Probability Theory and Its Applications, Volumes 1 (2nd Edition) and 2. John Wiley and Sons, Inc., 1957 and 1966.

L. G. Gotkin and L. S. Goldstein. Descriptive Statistics: A Programmed Textbook, Volumes 1 and 2. John Wiley and Sons, Inc., 1964 and 1965.

J. L. Hodges, Jr. and E. L. Lehmann. Basic Concepts of Probability and Statistics. Holden-Day, Inc., 1964.

P. G. Hoel. Elementary Statistics, 2nd Edition. John Wiley and Sons, Inc., 1966.

D. Huff and I. Geis. How to Lie with Statistics. W. W. Norton and Co., Inc., 1954.

D. Huff and I. Geis. How to Take a Chance. W. W. Norton and Co., Inc., 1959.

D. A. Johnson and W. H. Glenn. The World of Statistics, Exploring Mathematics on Your Own Series. Webster Publishing Co., 1961.

H. C. Levinson. Chance, Luck and Statistics, 2nd Edition. Dover Publications, Inc., 1963.

C. McCollough and L. Van Atta. Statistical Concepts: A Program for Self-Instruction. McGraw-Hill Book Co., 1963.

[*] The authors are indebted to R. L. McCornack of System Development Corporation for the compilation of this bibliography.

A. M. Mood and F. A. Graybill. Introduction to the Theory of Statistics, 2nd Edition. McGraw-Hill Book Co., 1963.

F. Mosteller. Fifty Challenging Problems in Probability. Addison-Wesley Publishing Co., Inc., 1965.

F. Mosteller, et al. Probability with Statistics. Addison-Wesley Publishing Co., Inc., 1961.

E. Parzen. Modern Probability Theory and Its Applications. John Wiley and Sons, Inc., 1960.

E. O. Thorp. Beat the Dealer: A Winning Strategy for the Game of 21, Revised Edition. Random House, Inc., 1966.

E. O. Thorp. Elementary Probability. John Wiley and Sons, Inc., 1966.

W. A. Wallis and H. V. Roberts. Statistics: A New Approach. The Free Press, 1956.

R. E. Walpole. Introduction to Statistics. MacMillan Co., 1968.

DOUGLAS MISSILE & SPACE SYSTEMS DIVISION
5301 Bolsa Avenue Huntington Beach, California 92646

MCDONNELL DOUGLAS
CORPORATION