SCHAUM'S OUTLINE SERIES

THEORY AND PROBLEMS OF

STATISTICS and ECONOMETRICS

DOMINICK SALVATORE

INCLUDING 640 SOLVED PROBLEMS

SCHAUM'S OUTLINE SERIES IN ECONOMICS

McGRAW-HILL BOOK COMPANY

SCHAUM'S OUTLINE OF

THEORY AND PROBLEMS

of

STATISTICS
and
ECONOMETRICS

DOMINICK SALVATORE, Ph.D.

Professor of Economics
Fordham University

SCHAUM'S OUTLINE SERIES

McGRAW-HILL BOOK COMPANY

New York St. Louis San Francisco Auckland Bogotá Guatemala Hamburg Johannesburg Lisbon London Madrid Mexico Montreal New Delhi Panama Paris San Juan São Paulo Singapore Sydney Tokyo Toronto

DOMINICK SALVATORE received his Ph.D. in 1971 and is currently Professor of Economics at Fordham University. He is the author of the Schaum's Outlines of *Microeconomic Theory* and *International Economics,* and is the coauthor of *Development Economics* and *Principles of Economics.* His research has been published in numerous scholarly journals and presented at national and international conferences.

Schaum's Outline of Theory and Problems of
STATISTICS AND ECONOMETRICS

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 SH SH 8 6 5 4 3 2

ISBN 0-07-054505-7

Sponsoring Editor, John Aliano
Editing Supervisor, Susan Thomas
Production Manager, Nick Monti

Library of Congress Cataloging in Publication Data

Salvatore, Dominick.
Schaum's outline of theory and problems of statistics and econometrics.

(Schaum's outline series)
Includes index.
1. Social sciences—Statistical methods.
2. Statistics. 3. Econometrics. I. Title.
II. Series.
HA29.S239 519.5 81-3748
ISBN 0-07-054505-7 AACR2

Preface

This book presents a clear and concise introduction to statistics and econometrics. A course in statistics or econometrics is often one of the most useful but also one of the most difficult of the required courses in colleges and universities. The purpose of this book is to help overcome this difficulty by using a problem-solving approach.

Each chapter begins with a statement of theory, principles or background information, fully illustrated with examples. This is followed by numerous theoretical and practical problems with detailed, step-by-step solutions. While primarily intended as a supplement to all current standard textbooks of statistics and/or econometrics, the book can also be used as an independent text, as well as to supplement class lectures.

The book is aimed at college students in economics, business administration, and the social sciences taking a one-semester or a one-year course in statistics and/or econometrics. It also provides a very useful source of reference for M.A. and M.B.A. students and for all those who use (or would like to use) statistics and econometrics in their work. No prior statistical background is assumed.

The book is completely self-contained in that it covers the statistics (Chapters 1–5) required for econometrics (Chapters 6–10). It is applied in nature and all proofs appear in the problems section rather than in the text itself. Real-world socioeconomic and business data are used, whenever possible, to demonstrate the more advanced econometric techniques and models. An actual complete computer printout is included to show how to use and interpret the results obtained from the most common of the statistical programs *(The Statistical Package for the Social Sciences* or *SPSS).* Topics frequently encountered in econometrics, such as multicollinearity and autocorrelation, are clearly and concisely discussed as to the problems they create, the methods to test for their presence, and possible correction techniques. A sample statistics and econometrics examination is also included.

The methodology of this book and much of its content have been tested in undergraduate and graduate classes in statistics and econometrics at Fordham University. Students found the approach and content of the book extremely useful and made many valuable suggestions for improvement. I have also received very useful advice from Professors John Piderit and Edward Dowling of Fordham University and Joseph Kiernan and Richard Kjetsaa of Fairleigh Dickinson University. The following graduate and undergraduate students carefully read through the entire manuscript and made many useful comments: William Foote, Frank Altieri, Cecilia Winters, Thomas Lawder, Richard Michelfelder, Anita Pasmantier, and Connie and Maureen Reis. To all of them I am deeply grateful. I owe a great intellectual debt to my former professors of statistics and econometrics: Jack Johnston, Lawrence Klein, and Bernard Okun. Finally, I would like to express my gratitude to Joseph Middleton, Charles Barcellona, and Mary Grier of the Fordham Computing Center and to the entire Schaum's staff at McGraw-Hill, especially John Aliano and Nick Monti for their kind and skillful assistance.

I am indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to the Longman Group Ltd., London, for permission to adapt and reprint Tables III and IV from their book, *Statistical Tables for Biological, Agricultural and Medical Research*.

In addition to *Statistics and Econometrics*, the Schaum's Outline Series in Economics includes: *Microeconomic Theory, Macroeconomic Theory, International Economics, Development Economics, Mathematics for Economists*, and *Principles of Economics*.

DOMINICK SALVATORE

New York, 1982

Contents

Chapter 1

Introduction

1.1 THE NATURE OF STATISTICS

Statistics refers to the collection, presentation, analysis, and utilization of numerical data to make inferences and reach decisions in the face of uncertainty in economics, business, and other other social and physical sciences.

Statistics is subdivided into descriptive and inferential. *Descriptive statistics* is concerned with summarizing and describing a body of data. *Inferential statistics* is the process of reaching generalizations about the whole (called the *population*) by examining a portion (called the *sample*). In order for this to be valid, the sample must be *representative* of the population and the *probability* of error also must be specified.

Descriptive statistics is discussed in detail in Chap. 2. This is followed by (the more crucial) statistical inference, with Chap. 3 dealing with probability, Chap. 4 with estimation, and Chap. 5 with hypothesis testing.

EXAMPLE 1. Suppose that we have data on the incomes of 1,000 U.S. families. This body of data can be summarized by finding the average family income and the spread of these family incomes above and below the average. The data also can be described by constructing a table, chart, or graph of the number or proportion of families in each income class. This is descriptive statistics. If these 1,000 families are representative of all U.S. families, we can then estimate and test hypotheses about the average family income in the United States as a whole. Since these conclusions are subject to error, we also would have to indicate the probability of error. This is statistical inference.

1.2 STATISTICS AND ECONOMETRICS

Econometrics refers to the application of economic theory, mathematics, and statistical techniques for the purpose of testing hypotheses and estimating and forecasting economic phenomena. Econometrics has become strongly identified with *regression analysis*. This relates a dependent variable to one or more independent or explanatory variables. Since relationships among economic variables are generally inexact, a disturbance or error term (with well-defined probabilistic properties) must be included (see Prob. 1.8).

Chapters 6 and 7 deal with regression analysis; Chapter 8 extends the basic regression model; Chapter 9 deals with methods of testing and correcting for violations in the assumptions of the basic regression model; and Chapter 10 deals with simultaneous-equation methods. Thus Chaps. 1 to 5 deal with the statistics required for econometrics (Chaps. 6 to 10).

EXAMPLE 2. Consumption theory tells us that, in general, people increase their consumption expenditure C as their disposable (after-tax) income Y_d increases, but not by as much as the increase in their disposable income. This can be stated in explicit linear equation form as

$$C = b_0 + b_1 Y_d \tag{1.1}$$

where b_0 and b_1 are unknown constants called *parameters*. The parameter b_1 is the slope coefficient representing the marginal propensity to consume, MPC. Since even people with identical disposable income are likely to have somewhat different consumption expenditures, the theoretically exact and deterministic relationship represented by Eq. (1.1) must be modified to include a random disturbance or error term, u, making it stochastic:

$$C = b_0 + b_1 Y_d + u \tag{1.2}$$

1.3 THE METHODOLOGY OF ECONOMETRICS

Econometric research, in general, involves the following three stages:

Stage 1. Specification of the model or maintained hypothesis in explicit stochastic equation form, together with the a priori theoretical expectations about the sign and size of the parameters of the function.

Stage 2. Collection of data on the variables of the model and estimation of the coefficients of the function with appropriate econometric techniques (presented in Chaps. 6 to 8).

Stage 3. Evaluation of the estimated coefficients of the function on the basis of economic, statistical, and econometric criteria.

EXAMPLE 3. The *first stage* in econometric research on consumption theory is to state the theory in explicit stochastic equation form, as in Eq. (1.1), with the expectation that $b_0 > 0$ (i.e., at $Y_d = 0$, $C > 0$ as people dissave and/or borrow) and $0 < b_1 < 1$. The *second stage* involves the collection of data on consumption expenditure and disposable income and estimation of Eq. (1.1). The *third stage* in econometric research involves (1) checking to see if the estimated value of $b_0 > 0$ and if $0 < b_1 < 1$; (2) determining if a "satisfactory" proportion of the variation in C is "explained" by changes in Y_d and if b_0 and b_1 are "statistically significant at acceptable levels" [see Prob. 1.13(*c*) and Sec. 5.2]; and (3) testing to see if the assumptions of the basic regression model are satisfied or, if not, how to correct for violations. If the estimated relationship does not pass these tests, the hypothesized relationship must be modified and reestimated until a satisfactory estimated consumption relationship is achieved.

Solved Problems

THE NATURE OF STATISTICS

1.1 What is the purpose and function of: (*a*) The field of study of statistics? (*b*) Descriptive statistics? (*c*) Inferential statistics?

(*a*) Statistics is the body of procedures and techniques used to collect, present, and analyze data on which to base decisions in the face of uncertainty or incomplete information. Statistical analysis is used today in practically every profession. The economist uses it to test the efficiency of alternative production techniques; the businessperson may use it to test the product design or package that maximizes sales; the sociologist, to analyze the result of a drug rehabilitation program; the industrial psychologist, to examine workers' responses to plant environment; the political scientist, to forecast voting patterns; the medical doctor, to test the effectiveness of a new drug; the chemist, to produce cheaper fertilizers; and so on.

(*b*) Descriptive statistics summarizes a body of data with one or two pieces of information that characterize the whole data. It also refers to the presentation of a body of data in the form of tables, charts, graphs, and other forms of graphic display.

(*c*) Inferential statistics (both estimation and hypothesis testing) refers to the drawing of generalizations about the properties of the whole (called a *population*) from the specific or a sample drawn from the population. Inferential statistics thus involves inductive reasoning. (This is to be contrasted with deductive reasoning, which ascribes properties to the specific starting with the whole.)

1.2 (*a*) Is descriptive or inferential statistics more important today? (*b*) What is the importance of a representative sample in statistical inference? (*c*) Why is probability theory required?

(*a*) Statistics started as a purely descriptive science, but it grew into a powerful tool of decision making as its inferential branch was developed. Modern statistical analysis refers primarily to inferential or inductive statistics. However, deductive and inductive statistics are complementary. We must study how to generate samples from populations before we can learn to generalize from samples to populations.

(*b*) In order for statistical inference to be valid, it must be based on a sample that fully reflects the characteristics and properties of the population from which it is drawn. A representative sample is ensured by random sampling, whereby each element of the population has an equal chance of being included in the sample (see Sec. 4.1).

(*c*) Since the possibility of error exists in statistical inference, estimates or tests of a population property or characteristic are given together with the chance or probability of being wrong. Thus probability theory is an essential element in statistical inference.

1.3 How can the manager of a firm producing light bulbs summarize and describe to a board meeting the results of testing the life of a sample of 100 light bulbs produced by the firm?

Providing the (raw) data on the life of each in the sample of 100 light bulbs produced by the firm would be very inconvenient and time-consuming for the board members to evaluate. Instead, the manager might summarize the data by indicating that the average life of the bulbs tested is 360 h and that 95% of the bulbs tested lasted between 320 and 400 h. By doing this, the manager is providing two pieces of information (the average life and the spread in the average life) that characterize the life of the 100 bulbs tested. The manager also might want to describe the data with a table or chart indicating the number or proportion of bulbs tested that lasted within each 10-h classification. Such a tabular or graphic representation of the data is also very useful for gaining a quick overview of the data. In summarizing and describing the data in the ways indicated, the manager is engaging in descriptive statistics. It should be noted that descriptive statistics can be used to summarize and describe any body of data, be it a sample (as above) or a population (when all the elements of the population are known and its characteristics can be calculated).

1.4 (*a*) Why may the manager in Prob. 1.3 want to engage in statistical inference? (*b*) What would this involve and require?

(*a*) Quality control requires that the manager have a fairly good idea about the average life and the spread in the life of the light bulbs produced by the firm. However, testing all the light bulbs produced would destroy the entire output of the firm. Even when testing does not destroy the product, testing the entire output is usually prohibitively expensive and time-consuming. The usual procedure is to take a sample of the output and infer the properties and characteristics of the entire output (population) from the corresponding characteristics of a sample drawn from the population.

(*b*) Statistical inference requires first of all that the sample be representative of the population being sampled. If the firm produces light bulbs in different plants, with more than one work shift, and with raw materials from different suppliers, these must be represented in the sample in the proportion in which they contribute to the total output of the firm. From the average life and spread in the life of the bulbs in the sample, the firm manager might estimate, with 95% probability of being correct and 5% probability of being wrong, the average life of all the light bulbs produced by the firm to be between 320 and 400 h (see Sec. 4.3). Instead, the manager may use the sample information to test, with 95% probability of being correct and 5% probability of being wrong, that the average life of the population of all the bulbs produced by the firm is greater than 320 h (see Sec. 5.2). In estimating or testing the average for a population from sample information, the manager is engaging in statistical inference.

STATISTICS AND ECONOMETRICS

1.5 What is meant by (*a*) Econometrics? (*b*) Regression analysis? (*c*) Disturbance or error term? (*d*) Simultaneous-equations models?

(*a*) Econometrics is the integration of economic theory, mathematics, and statistical techniques for the purpose of testing hypotheses about economic phenomena, estimating coefficients of economic relationships, and forecasting or predicting future values of economic variables or phenomena. Econometrics is subdivided into theoretical and applied econometrics. *Theoretical econometrics* refers to the methods for measurement of economic relationships in general. *Applied econometrics* examines the problems encountered and the findings in particular fields of economics, such as demand theory, production, investment, consumption, and other fields of applied economic research. In any case, econometrics is partly an art and partly a science, because often the intuition and good judgment of the econometrician plays a crucial role.

(*b*) Regression analysis studies the causal relationship between one economic variable to be explained (the dependent variable) and one or more independent or explanatory variables. When there is only one independent or explanatory variable, we have *simple regression*. In the more usual case of more than one independent or explanatory variable, we have *multiple regression*.

(*c*) A (random) disturbance or error must be included in the exact relationships postulated by economic theory and mathematical economics in order to make them stochastic (i.e., in order to reflect the fact that in the real world, economic relationships among economic variables are inexact and somewhat erratic).

(*d*) Simultaneous-equations models refer to relationships among economic variables expressed with more than one equation and such that the economic variables in the various equations interact. Simultaneous-equations models are the most complex aspect of econometrics and are discussed in Chap. 10.

1.6 (*a*) What are the functions of econometrics? (*b*) What aspects of econometrics (and other social sciences) make it basically different from most physical sciences?

(*a*) Econometrics has basically three closely interrelated functions. The first is to test economic theories or hypotheses. For example, is consumption directly related to income? Is the quantity demanded of a commodity inversely related to its price? The second function of econometrics is to provide numerical estimates of the coefficients of economic relationships. These are essential in decision making. For example, a government policymaker needs to have an accurate estimate of the coefficient of the relationship between consumption and income in order to determine the stimulating (i.e., the multiplier) effect of a proposed tax reduction. A manager needs to know if a price reduction increases or reduces the total sales revenues of the firm and by how much. The third function of econometrics is the forecasting of economic events. This, too, is necessary in order for policymakers to take appropriate corrective action if the rate of unemployment or inflation is predicted to rise in the future.

(*b*) There are two basic differences between econometrics (and other social sciences), on the one hand, and most physical sciences (such as physics), on the other. One is that (as pointed out earlier) relationships among economic variables are inexact and somewhat erratic. The second is that most economic phenomena occur contemporaneously, so that laboratory experiments cannot be conducted. These differences require special methods of analysis (such as the inclusion of a disturbance or error term with the exact relationships postulated by economic theory) and multivariate analysis (such as multiple regression analysis). The latter isolates the effect of each independent or explanatory variable on the dependent variable in the face of contemporaneous change in all explanatory variables.

1.7 In what way and for what purpose are (*a*) economic theory, (*b*) mathematics, and (*c*) statistical analysis combined to form the field of study of econometrics?

(*a*) Econometrics presupposes the existence of a body of economic theories or hypotheses requiring testing. If the variables suggested by economic theory do not provide a satisfactory explanation, the researcher may experiment with alternative formulations and variables suggested by previous tests or opposing theories. In this way, econometric research can lead to the acceptance, rejection, and reformulation of economic theories.

(*b*) Mathematics is used to express the verbal statements of economic theories in mathematical form, expressing an exact or deterministic functional relationship between the dependent and one or more independent or explanatory variables.

(c) Statistical analysis applies appropriate techniques to estimate the inexact and nonexperimental relationships among economic variables by utilizing relevant economic data and evaluating the results.

1.8 What justifies the inclusion of a disturbance or error term in regression analysis?

The inclusion of a (random) disturbance or error term (with well-defined probabilistic properties) is required in regression analysis for three important reasons. First, since the purpose of theory is to generalize and simplify, economic relationships usually include only the most important forces at work. This means that numerous other variables with slight and irregular effects are not included. The error term can be viewed as representing the net effect of this large number of small and irregular forces at work. Second, the inclusion of the error term can be justified in order to take into consideration the net effect of possible errors in measuring the dependent variable or variable being explained. Finally, since human behavior usually differs in a random way under identical circumstances, the disturbance or error term can be used to capture this inherently random human behavior. This error term thus allows for individual random deviations from the exact and deterministic relationships postulated by economic theory and mathematical economics.

1.9 Consumer demand theory states that the quantity demanded of a commodity, D_X, is a function of or depends on its price, P_X, consumers' income, Y, and the price of other (related) commodities, say, commodity Z (that is, P_Z). Assuming that consumers' tastes remain constant during the period of analysis, state the preceding theory in (a) specific or explicit linear form or equation and (b) in stochastic form. (c) Which are the coefficients to be estimated? What are they called?

(a)
$$D_X = b_0 + b_1 P_X + b_2 Y + b_3 P_Z \tag{1.3}$$

(b)
$$D_X = b_0 + b_1 P_X + b_2 Y + b_3 P_Z + u \tag{1.4}$$

(c) The coefficients to be estimated are b_0, b_1, b_2, and b_3. They are called *parameters*.

THE METHODOLOGY OF ECONOMETRICS

1.10 With reference to the consumer demand theory in Prob. 1.9, indicate (a) what the first step is in econometric research and (b) what the a priori theoretical expectations are of the sign and possible size of the parameters of the demand function given by Eq. (1.4).

(a) The first step in econometric analysis is to express the theory of consumer demand in stochastic equation form, as in Eq. (1.4), and indicate the a priori theoretical expectations about the sign and possibly the size of the parameters of the function.

(b) Consumer demand theory postulates that in Eq. (1.4), $b_1 < 0$ (indicating that price and quantity are inversely related), $b_2 > 0$ if the commodity is a normal good (indicating that consumers purchase more of the commodity at higher incomes), $b_3 > 0$ if X and Z are substitutes, and $b_3 < 0$ if X and Z are complements.

1.11 Indicate the second stage in econometric research (a) in general and (b) with reference to the demand function specified by Eq. (1.4).

(a) The second stage in econometric research involves the collection of data on the dependent variable and on each of the independent or explanatory variables of the model and utilizing these data for the empirical estimation of the parameters of the model. This is usually done with multiple regression analysis (discussed in Chap. 7).

(b) In order to estimate the demand function given by Eq. (1.4), data must be collected on (1) the quantity demanded of commodity X by consumers, (2) the price of X, (3) consumers' incomes, and (4) the price of commodity Z per unit of time (i.e., per day, month, or year) and over a number of days, months, or years. Data on P_X, Y, and P_Z are then regressed against data on D_X and estimates of parameters b_0, b_1, b_2, and b_3 obtained.

1.12 How does the type of data required to estimate the demand function specified by Eq. (1.4) differ from the type of data that would be required to estimate the consumption function *for a group of families at one point in time*?

In order to estimate the demand function given by Eq. (1.4), numerical values of the variables are required over a period of time. For example, if we want to estimate the demand function for coffee, we need the numerical value of the quantity of coffee demanded, say, per year, over a number of years, say, from 1960 to 1980. Similarly, we need data on the average price of coffee, consumers' income, and the price of, say, tea (a substitute for coffee) per year from 1960 to 1980. Data that give numerical values for the variables of a function from period to period are called *time-series data*. However, to estimate the consumption function for a group of families at one point in time, we need *cross-sectional data* (i.e., numerical values for the consumption expenditures and disposable incomes of each family in the group at a particular point in time, say, in 1982).

1.13 What is meant by: (*a*) The third stage in econometric analysis? (*b*) A priori theoretical criteria? (*c*) Statistical criteria? (*d*) Econometric criteria? (*e*) The forecasting ability of the model?

(*a*) The third stage in econometric research involves the evaluation of the estimated model on the basis of the a priori economic criteria, statistical and econometric criteria, and the forecasting ability of the model.

(*b*) The a priori economic criteria refer to the sign and size of the parameters of the model postulated by economic theory. If the estimated coefficients do not conform to those postulated, the model must be revised or rejected.

(*c*) The statistical criteria refer to (1) the proportion of variation in the dependent variable "explained" by changes in the independent or explanatory variables and (2) verification that the dispersion or spread of each estimated coefficient around the true parameter is sufficiently narrow to give us "confidence" in the estimates.

(*d*) The econometric criteria refer to tests that the assumptions of the basic regression model, and particularly those about the disturbance or error term, are satisfied.

(*e*) The forecasting ability of the model refers to the ability of the model to accurately predict future values of the dependent variable based on known or expected future value(s) of the independent or explanatory variable(s).

1.14 How can the *estimated* demand function given by Eq. (1.4) be evaluated in terms of: (*a*) The a priori theoretical criteria? (*b*) The statistical criteria? (*c*) The econometric criteria? (*d*) The forecasting ability of the model?

(*a*) The estimated demand function given by Eq. (1.4) can be evaluated in terms of the a priori theoretical criteria by checking that the estimated coefficients conform to the theoretical expectations with regard to sign and possible size, as postulated in Prob. 1.10(*b*) The demand theory given by Eq. (1.4) is confirmed only if $b_1 < 0$, if $b_2 > 0$ (if X is a normal good), and if $b_3 > 0$ (if Z is a substitute for X), as postulated by demand theory.

(*b*) The statistical criteria are satisfied only if a "high" proportion of the variation in D_X over time is "explained" by changes in P_X, Y, and P_Z, and if the dispersion of estimated b_1, b_2, and b_3 around the true parameters are "sufficiently narrow." There is no generally accepted answer as to what is a "high" proportion of the variation in D_X "explained" by P_X, Y, and P_Z. However, because of common trends in time-series data, we would expect more than 50 to 70% of the variation in the dependent variable to be explained by the independent or explanatory variables for the model to be judged satisfactory. Similarly, in order for each estimated coefficient to be "statistically significant," we would expect the dispersion of each estimated coefficient about the true parameter (measured by its standard deviation; see Sec. 2.3) to be generally less than half the estimated value of the coefficient.

(c) The econometric criteria are used to determine if the assumptions of the econometric methods used are satisfied in the estimation of the demand function of Eq. (1.4). Only if these assumptions are satisfied will the estimated coefficients have the desirable properties of unbiasedness, consistency, efficiency, and so forth (see Sec. 6.4).

(d) One way to test the forecasting ability of the demand model given by Eq. (1.4) is to use the estimated function to predict the value of D_X for a period not included in the sample and checking that this predicted value is "sufficiently close" to the actual observed value of D_X for that period.

1.15 Present in schematic form the various stages of econometric research.

Stage 1: Economic theory
↓
Mathematical model
↓
Econometric (stochastic) model

Stage 2: Collection of appropriate data
↓
Estimation of the parameters of the model

Stage 3: Evaluation of the model on the basis of economic, statistical, and econometric criteria
↓

Accept theory if compatible with data	Reject theory if incompatible with data	Revise theory if incompatible with data
↓		↓
Prediction		Confrontation of revised theory with new data

Supplementary Problems

THE NATURE OF STATISTICS

1.16 (a) To which field of study is statistical analysis important? (b) What are the most important functions of descriptive statistics? (c) What is the most important function of inferential statistics?
Ans. (a) To economics, business, and other social and physical sciences (b) Summarizing and describing a body of data (c) Drawing inferences about the characteristics of a population from the corresponding characteristics of a sample drawn from the population

1.17 (a) Is statistical inference associated with deductive or inductive reasoning? (b) What are the conditions required in order for statistical inference to be valid?
Ans. (a) Inductive reasoning (b) A representative sample and probability theory

STATISTICS AND ECONOMETRICS

1.18 Express in the form of an explicit linear equation the statement that the level of investment spending I is inversely related to rate of interest R.
Ans. $I = b_0 + b_1 R$ with b_1 postulated to be negative (1.5)

1.19 What is the answer to Prob. 1.18 an example of?
Ans. An economic theory expressed in (exact or deterministic) mathematical form

1.20 Express Eq. (1.5) in stochastic form.
Ans.

$$I = b_0 + b_1 R + u \tag{1.6}$$

1.21 Why is a stochastic form required in econometric analysis?
Ans. Because the relationships among economic variables are inexact and somewhat erratic as opposed to the exact and deterministic relationships postulated by economic theory and mathematical economics

THE METHODOLOGY OF ECONOMETRICS

1.22 What are stages (*a*) one, (*b*) two, and (*c*) three in econometric research?
Ans. (*a*) Specification of the theory in stochastic equation form and indication of the expected signs and possible sizes of estimated parameters (*b*) Collection of data on the variables of the model and estimation of the coefficients of the function (*c*) Economic, statistical, and econometric evaluation of the estimated parameters

1.23 What is the first stage of econometric analysis for the investment theory in Prob. 1.18?
Ans. Stating the theory in the form of Eq. (1.6) and predicting $b_1 < 0$

1.24 What is the second stage in econometric analysis for the investment theory in Prob. 1.18?
Ans. Collection of time-series data on I and R and estimation of Eq. (1.6)

1.25 What is the third stage of econometric analysis for the investment theory in Prob. 1.18?
Ans. Determination that the estimated coefficient of $b_1 < 0$, that an "adequate" proportion of the variation in I over time is "explained" by changes in R, that b_1 is "statistically significant at customary levels," and that the econometric assumptions of the model are satisfied

Chapter 2

Descriptive Statistics

2.1 FREQUENCY DISTRIBUTIONS

It is often useful to organize or arrange a body of data into a *frequency distribution*. This breaks up the data into groups or classes and shows the number of observations in each class. The number of classes is usually between 5 and 15. A *relative frequency distribution* is obtained by dividing the number of observations in each class by the total number of observations in the data as a whole. The sum of the relative frequencies equals 1. A *histogram* is a bar graph of a frequency distribution, where classes are measured along the horizontal axis and frequencies along the vertical axis. A *frequency polygon* is a line graph of a frequency distribution resulting from joining the frequency of each class plotted at the class midpoint. A *cumulative frequency distribution* shows, for each class, the total number of observations in all classes up to and including that class. When plotted, this gives a *distribution curve*, or *ogive*.

EXAMPLE 1. A student received the following grades (measured from 0 to 10) on the 10 quizzes he took during a semester: 6, 7, 6, 8, 5, 7, 6, 9, 10, and 6. These grades can be arranged into frequency distributions as in Table 2.1 and shown graphically as in Fig. 2-1.

Table 2.1 Frequency Distributions of Grades

Grades	Absolute Frequency	Relative Frequency
5	1	0.1
6	4	0.4
7	2	0.2
8	1	0.1
9	1	0.1
10	1	0.1
	10	1.0

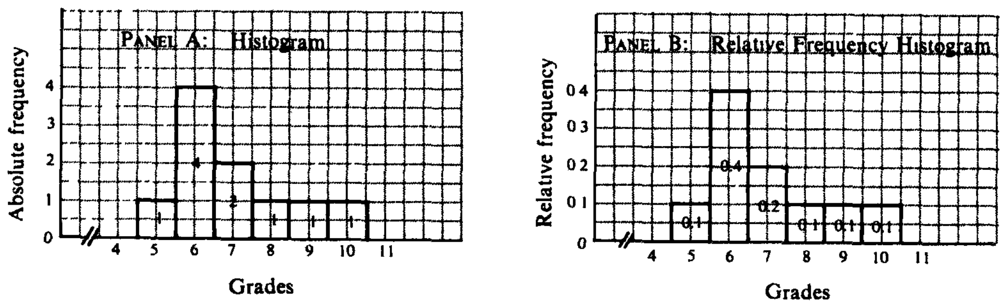

Fig. 2-1

EXAMPLE 2. The cans in a sample of 20 cans of fruit contain net weights of fruit ranging from 19.3 to 20.9 oz, as given in Table 2.2. If we want to group these data into 6 classes, we get *class intervals* of 0.3 oz [(21.0 − 19.2)/6 = 0.3 oz]. The weights given in Table 2.2 can be arranged into the frequency distributions given in Table 2.3 and shown graphically in Fig. 2-2.

Table 2.2 Net Weight in Ounces of Fruit

19.7	19.9	20.2	19.9	20.0	20.6	19.3	20.4	19.9	20.3
20.1	19.5	20.9	20.3	20.8	19.9	20.0	20.6	19.9	19.8

Table 2.3 Frequency Distribution of Weights

Weight, oz	Class Midpoint	Absolute Frequency	Relative Frequency	Cumulative Frequency
19.2–19.4	19.3	1	0.05	1
19.5–19.7	19.6	2	0.10	3
19.8–20.0	19.9	8	0.40	11
20.1–20.3	20.2	4	0.20	15
20.4–20.6	20.5	3	0.15	18
20.7–20.9	20.8	2	0.10	20
		20	1.00	

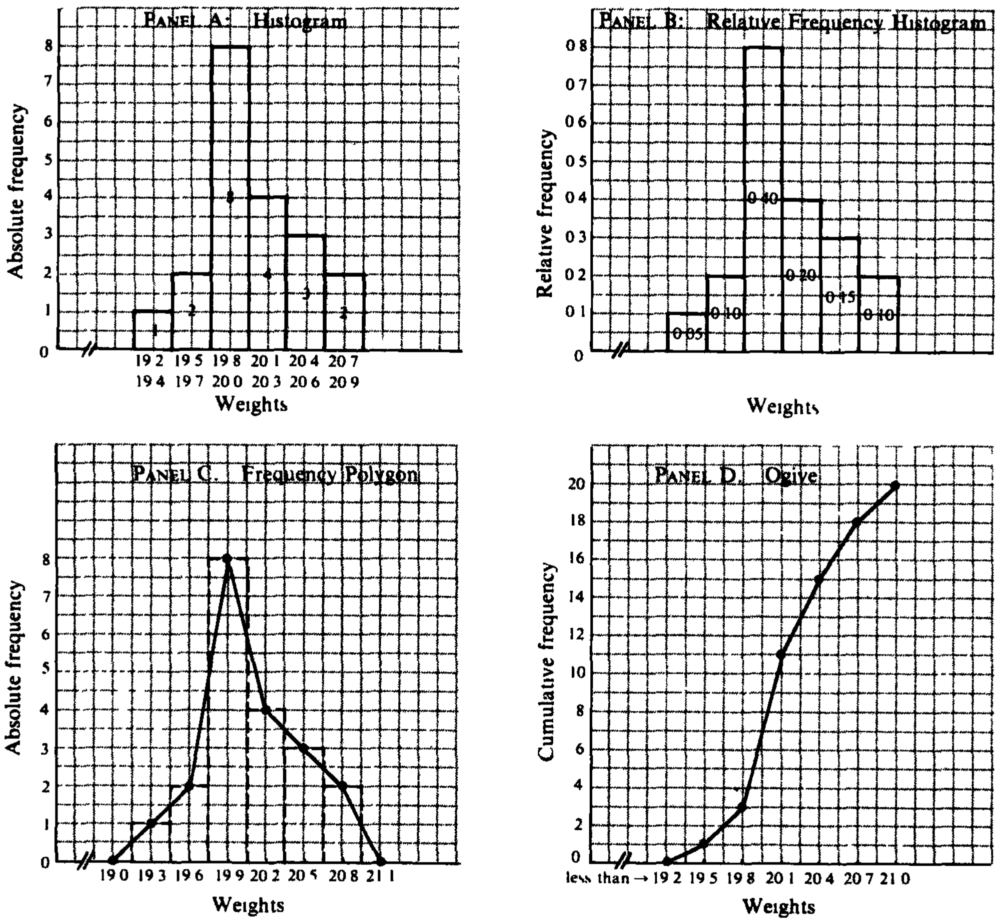

Fig. 2-2