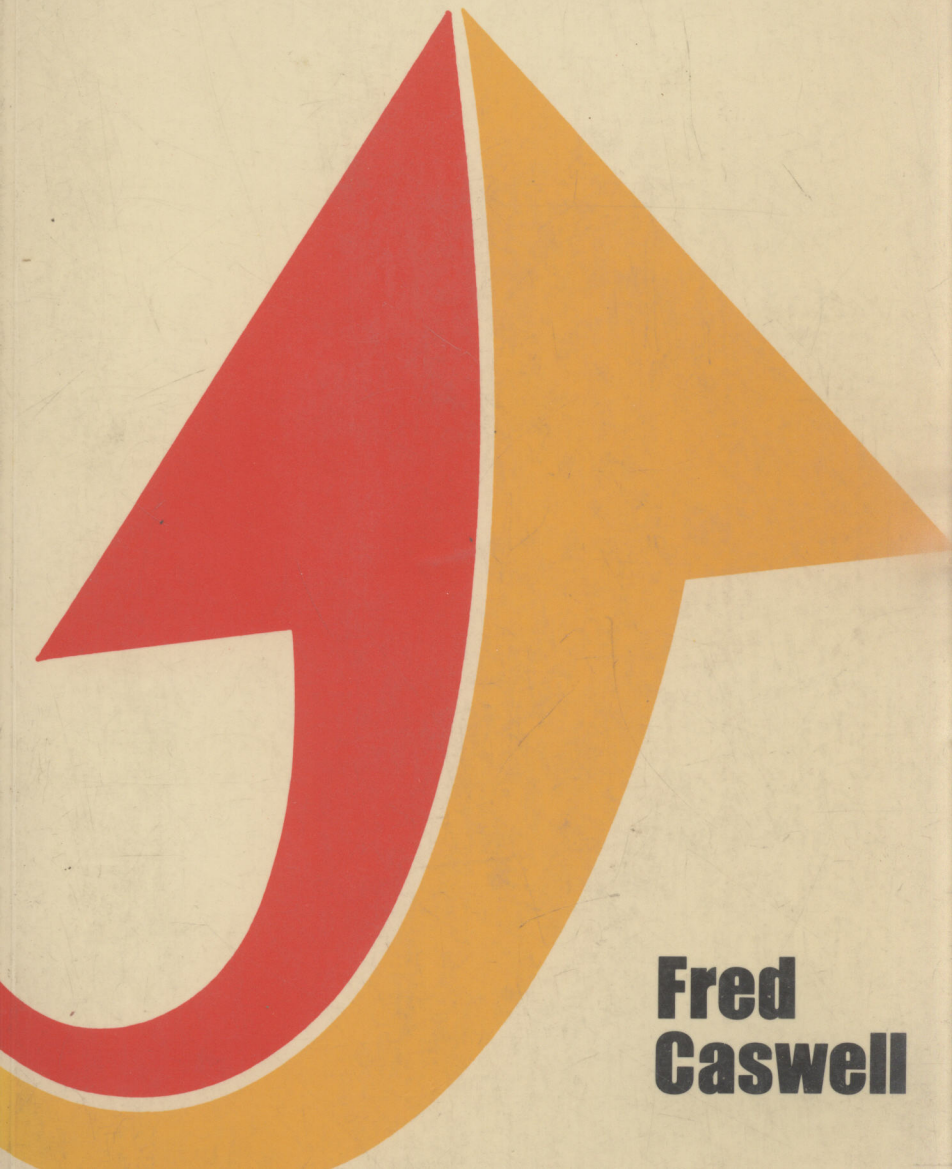
SUCCESS IN
Statistics



Fred Caswell

Success in STATISTICS

Fred Caswell, B.A.

Lanchester Polytechnic, Coventry

CONSULTANT EDITOR

Helen Wright, M.A., Ph.D.

John Murray

First published 1982
by John Murray (Publishers) Ltd
50 Albemarle Street, London W1X 4BD

Printed in Hong Kong by
Wing King Tong Ltd

British Library Cataloguing in Publication Data
Caswell, Fred
Success in statistics.—(Success studybooks)
1. Mathematical statistics
I. Title
519.5 QA276
ISBN 0-7195-3902-1

Foreword

In our modern society it has become increasingly necessary to be numerate. Numeracy at one time implied merely agility at formal arithmetic but the arrival of electronic calculators and computers has lifted this burden and enables us to concentrate on the interpretation of numerical data rather than on numerical procedures.

The rate of inflation, index linked savings, forecasts of trends in the economy, public opinion polls, suggestions of links between environmental factors and illness – all require some knowledge of statistical techniques in order to be fully understood and interpreted. It is hoped that this book will provide a useful introduction to statistics for anyone who wishes to appreciate published data in greater depth.

There are many people, however, for whom a study of statistics is an essential element of their working lives and *Success in Statistics* caters for them: accountants, market executives, managers, economists, scientists and others. The aim is not to produce specialist statisticians but to give an insight into basic statistical methods which can be applied in many diverse fields.

This Studybook is appropriate for all BEC National and Higher National modules with a statistical content. The Suggestions for Projects, which can be adapted for use as BEC assignments, and Unit 16, Interest and Investment Appraisal, are particularly intended for such courses. The examinations of many professional bodies now have a statistical element at an early stage and *Success in Statistics* provides the necessary material to master such syllabuses. It also covers the requirements for O-level statistics examinations and provides a useful introduction to statistics for those intending to study the subject at a higher level.

F.C. and H.W.

Acknowledgments

Thanks are due to the many people who have helped in the preparation of this book, especially those who have read, criticized and contributed to it constructively. In particular we are grateful to George O'Sullivan of Birmingham Polytechnic, Robin Gardner of Vauxhall College, Anne Knight and other statistics lecturers of Watford College, Anne Webster of John Murray, Jean Macqueen, the Watford College librarians, and Ann Caswell and Ronald Reed who typed the manuscript.

Questions from past examination papers are reproduced by permission of the following: Associated Examining Board; Association of Certified Accountants; Institute of Chartered Accountants; Institute of Chartered Secretaries and Administrators; Institute of Cost and Management Accountants; Institute of Marketing; Institute of Personnel Management; Institute of Statisticians; Institution of Industrial Managers. Answers, where given, are the responsibility of the author, not the examining body.

Data from the *Annual Abstract of Statistics*, *Census Monitors*, the *Family Expenditure Survey*, the *Monthly Digest of Statistics* and *Social Trends* are reproduced by permission of the Controller of Her Majesty's Stationery Office.

Mathematical tables are reproduced by permission of the following: Addison-Wesley (Table D, pages 326–7, from A. Haber and R. P. Runyon, *General Statistics*); McGraw-Hill (Table E, page 328, from M. R. Spiegel, *Statistics*); Pitman (Tables F and G, pages 329 and 330, from J. E. Freund and F. J. Williams, *Modern Business Statistics*); Department of Statistics, University College, London (Table H, page 331, from *Tracts for Computers, 24*); Institute of Cost and Management Accountants (Table I, pages 332–3, from *The Profitable Use of Capital in Industry*); Biometrika Trustees (Table J, page 334, from *Tables for Statisticians*).

F.C. and H.W.

Contents

Unit One
Introduction to Statistics

1.1 Why Statistics Matter

Are you aware of the extent to which statistics enter your life? You will probably have read in the newspaper or heard on news programmes, of such statistics as 'average earnings' and the 'Index of Retail Prices'. These kinds of statistics are used in negotiations that lead to decisions that can have major effects on living standards.

Have you ever seen people standing around, usually in busy shopping centres, clutching clip-boards and stopping the occasional passer-by to ask questions? Perhaps you may have had someone calling at your home asking for your help with a survey of some kind, or have been asked to complete a questionnaire. The results of such surveys may be used to influence the types and qualities of goods and services that you are offered in the future.

Every ten years in Great Britain all households must complete a census return which is used to compile statistics on population patterns and developments. These statistics have an effect on government planning for the provision of housing, services such as schools and hospitals, and the development of industry.

In an environment where statistics play such an important role it is in the interests of us all to know more about them: how they are collected, analysed and used.

1.2 The Meaning of Statistics

The word statistics has two meanings:

(*a*) In the plural sense it means collections of numerical facts and is widely used when reference is made to facts and figures on such things as population, crime and education. Statisticians call the figures which have been collected *data*.

(*b*) In the singular sense it means the science (or art) of dealing with statistical data. The collection, analysis and interpretation of data is called *statistical method*, and it is with this sense of the word that we are mainly concerned in this book.

There are two subdivisions of statistical method:

(i) **Descriptive statistics.** This deals with the compilation and presentation of data in various forms such as tables, graphs and diagrams. The purpose of descriptive statistics is to display and pass on information from which conclusions can be drawn and decisions made. Businesses, for example, use descriptive statistics when presenting their annual accounts and reports, and the Government is a particularly prolific provider of descriptive statistics.

(ii) **Mathematical or inductive statistics.** This deals with the tools of statistics, the

techniques that are used to analyse the data and to make estimates or draw conclusions from the data.

1.3 The Growth and Importance of Statistics

Statistics have been collected since the earliest times. Rulers needed to have information about the population and their possessions so that taxes could be levied to maintain the state and the court, and it was also essential for them to be aware of the military strength of the nation. In the sixteenth century the word 'statist' was used to describe a politician—a dealer in facts about the state, its government and its people.

With the growth of the population and the advent, in the eighteenth and nineteenth centuries, of the industrial revolution and the accompanying agricultural revolution there was a need for a greater volume of statistics on an increasing variety of subjects. In this country the government also began to intervene more and more in the affairs of the people and of business and to attempt to control the workings of the economy. It therefore required information on:

(i) production;
(ii) earnings;
(iii) expenditure;
(iv) imports and exports.

As time went on, the Government took over many of the activities that had been part of the private sector of the economy, such as education and health services. It was vital therefore for information to be available on:

(v) population growth or decline;
(vi) disease and its incidence;
(vii) housing conditions.

All this led to an enormous expansion in the volume of statistics that have been and are being collected by governments over the last few decades. For governments to make sensible decisions, however, these data need to be correctly collected, processed and analysed.

1.4 Statistics and the Business World

It is not only governments that have required more and more statistics: companies have grown to such an extent that some have an annual turnover as great as the annual budget of a national economy. Large firms therefore have to make decisions on the basis of data.

The statistics that are collected by the Government are vital to businesses, and in the United Kingdom the Government's Central Statistical Office provides a specialist statistical service for businesses.

Businesses also collect their own statistics or pay specialist companies to collect the data for them. They require information on:

(i) the reaction of other companies to their products;
(ii) the reaction of customers to their products;
(iii) the effect of other companies' activities on their own;
(iv) the need for new products.

A large company will also need to compile internal statistics on sales, production, purchases, costings, and personnel matters. The accounting records of a company can be considered as a set of statistics displayed in a particular manner.

The tools of statistics are essential to modern business in areas such as planning, forecasting, and quality control. The techniques of statistical method are analogous to those of any other business tool: they are designed to do particular jobs. The methods of tackling any job can be summarized in the following steps:

(i) assessing the problem;
(ii) selecting the correct tool;
(iii) collecting the materials to enable the tools to be used.

Problems that involve statistical method should be approached in the same way. The businessman has to use his knowledge of statistics to select the correct tool to deal with the job he is faced with.

1.5 Misuse of Statistics

No doubt you have heard an argument supported by the words 'statistics show that ...' Many people seem to believe that a case is proven if statistics can be produced to support it. Thus an endless stream of data is thrown at us in an attempt to impress, persuade or even coerce us into believing that a particular political party is wise and good, or that we should buy such and such a product or that we should hold certain opinions. What we are seldom told is how the statistics which are meant to impress us have been collected, where they were collected or from whom they were collected. The old quip about 'lies, damned lies and statistics' has some truth in it. The figures themselves cannot mislead, but the statisticians who present the figures certainly can.

It is hoped that this book will alert you to the pitfalls associated with the use of statistics and that you will be ready to question claims that are supported merely by statistics. Data can be misused in the following ways:

(i) They can be used for the wrong purpose, that is, one that is different from the purpose for which they were collected.
(ii) They can be collected incorrectly so that they are biased.
(iii) They can be analysed carelessly so that the results obtained from them are misleading.

Be ready to find out more about these matters before believing any claims. A good statistician will be eager to show you how he obtained and dealt with his data.

1.6 Using this Book

To understand the subject-matter fully, you must be prepared to *work* through this book rather than merely to read it from cover to cover. You will find it useful to have paper and pencil to hand. Squared paper or graph paper, plus a ruler, are required for some Units. You should check the working in the text and attempt as many of the exercises as possible. The answers to numerical questions are at the end of the book. Any necessary mathematical and statistical tables are also to be found there.

You will find an electronic calculator an invaluable aid. Before buying one, study carefully the regulations regarding calculators for your particular examination. For statistical work a calculator should have a square root facility and, if possible, a memory, an exponential key (e^x) and an exponent key (x^y). 'Statistical' calculators giving standard deviations, for example, are available and can be used in some examinations. It is not essential, however, to have such a model since examiners at this level want to see full details of your workings—they are not solely interested in the final answers.

The material in the first eleven Units is essential reading for most first examinations in statistics, whether for a BEC award, a professional qualification or O-level statistics. Later Units are not required by all examining bodies so look carefully at your syllabus and past examination papers to see which sections you need. Units 12–15, in particular, are essential if you are to continue your statistical studies at a higher level. The sections on official statistics refer mainly to UK sources. If you are not resident in the UK, knowledge of these publications may not be appropriate for your examinations. In this case you should try to obtain information from your nearest library about relevant publications on such topics as population, prices, employment, wages and production.

In 'real life', statistical work often involves extensive calculations but the prime objective here is to enable you to learn principles and methods rather than to do lots of arithmetic. Most of the examples and exercises contain far fewer numbers than would be encountered in practice. After Unit 17, Suggestions for Projects are given. You are recommended to attempt some of them so that you acquire an appreciation of the practical difficulties that are met with in actual statistical investigations.

Some knowledge of mathematical techniques is needed in statistical work. It may be some time since you did any formal mathematical study and you may find it useful to revise specific topics as and when they are needed in the book. References to the companion volume, *Success in Mathematics*, are given to assist you wherever possible. These are denoted by the letters *SM* followed by a page or section reference.

A Further Reading list is to be found at the end of the book to help you to pursue statistical topics of particular interest to you in greater depth.

Unit Two

The Collection of Data

Statistics is concerned with the analysis of numerical data, so the first stage in statistical method must be the collection of the data to be analysed. It is an important stage, since unless data are collected carefully and correctly they will be of little value.

2.1 Populations

The data that the statistician works with are a series of observations from a *population*. For example, a statistical quality control exercise may be conducted in a factory producing electric light bulbs. At regular intervals a light bulb is picked at random from the production line and tested to see how long it will burn. The population here is the lifetimes of all the electric light bulbs on that production line. The lifetimes of those bulbs selected for testing are a *sample* from that population.

Another example might be an auctioneer at a livestock market who keeps records of all the cattle that pass under his hammer, including details of weights and prices, so that he can produce statistics in the future. Here there are two populations: the weights of all the cattle and the prices of all the cattle.

Sometimes populations concern people. When the results of a general election are being predicted, the population under scrutiny will be the opinions of all the people eligible to vote in that election. Another example would be when statisticians are trying to find out how families at a certain level in the economy spend their weekly income so that the increase in the cost of living can be calculated; the population here is the expenditure of every family at that economy level. We shall often refer loosely to the people and the families as populations, although, strictly speaking, it is their opinions and their expenditures respectively which form the populations.

If an attempt is made to observe a complete population, human or otherwise, then we say a *census* is being carried out. The decennial Census of Population in Great Britain is an increasingly important source of data for commercial organizations as well as supplying vital information on the size, distribution and living conditions of the population for central government and local authorities. Other official UK censuses are the Census of Distribution which covers retail trade and the Census of Production which collects data about the productive activity of the country. A general election attempts to record the votes of all those on the electoral list. Whereas in the Census of Population each head of household is legally required to complete a census return, there is no compulsion to vote in a general election in the United Kingdom. Thus, 100% coverage of the

population is not achieved. The overall percentage of those on the electoral roll who actually vote in a general election is around 80%. Some of the discrepancy is accounted for by deaths and migration in addition to those people who do not wish to record their votes.

We can readily see why samples have to be taken from populations: it reduces the data to manageable proportions. When data are taken from every family in an economy, it costs so much in time (and money) to collect and analyse them that the statistics are out of date by the time they are published. It might *just* be possible for the livestock auctioneer to record details of all the cattle he deals with and to compile statistics from the whole population, but it would clearly be impracticable to test every light bulb on the production line to extinction: there would be none left to sell. So, no matter what the size of the population, samples are taken and the observations which are made from the samples provide the data to be used to calculate the statistics of interest. It should not be thought that a sample survey is necessarily inferior to a census. Complete coverage of a population is seldom achieved unless the population is very small. Sample surveys take less time and can be more closely controlled. Consequently, trained investigators can make a more thorough examination of a sample than would be possible with a census.

2.2 Problems of Bias

When we collect data for analysis, we try to take a *random* sample, that is, one taken in such a way that every member of the population stands an equal chance of being selected as part of the sample. There must be no *bias* towards any individual or any groups of individuals in the population. We wish the sample truly to represent the population from which it is taken so that any conclusions drawn from it can be extended to the population as a whole. The sample must also be large enough: the larger the sample the more accurately findings from it should reflect the findings that would be obtained from the whole population.

Of course, if we want to prove an argument by fair means or foul we could deliberately set out to obtain a biased sample, collecting the data in such a way that they will support our argument. Any conclusions, however, could soon be faulted by someone who looked carefully at the way the data were collected. When you are presented with statistics in support of an argument, remember to inquire how the data from which the statistics are drawn were collected.

Unfortunately, obtaining a random and unbiased sample is probably the biggest problem we face as statisticians. We have to rely on the skill of the people who collect the data, we have to hope that the sample fills in any questionnaires accurately and honestly, we have to hope that we get a good response from the sample, and, finally, we have to pick a suitable sampling method.

2.3 Sampling Frames

The first thing which has to be done before choosing the sample is to try to identify the population from which the sample is to be taken. To help in making

this identification we can use a *sampling frame*, if an appropriate one is available. This is a list of the entire population from which the sample is to be selected. Some sampling frames come easily to mind: one of the most commonly used is the telephone directory. If a random sample from the population of Norwich is required, names can be selected at random from the Norwich telephone directory. However, the telephone directory does not contain the names of everyone in Norwich: it omits people who do not have a telephone, a large and important section of the population. If we were surveying the life-style of the population of Norwich we should obviously need to include people without a telephone if we did not want our findings to be biased. The directory would probably only be an ideal sampling frame if we wanted to sample telephone owners in Norwich.

The electoral roll is another commonly used sampling frame. It contains the names of all members of the population who are eligible to vote—ideal for a survey of how people will vote in an election, or for assessing the impact of a new brand of beer perhaps, but not suitable for assessing the impact of a new kind of bubble-gum. The frame must suit the survey being conducted.

Another sampling frame for the population of a town is the local street map. This can be used to select streets at random, and then some system of selecting houses within those streets devised—every tenth house, for example—in order to collect a sample. Here, residences are being selected for the sample, rather than people, and the data are collected from the occupants of those residences.

There are many more possible sampling frames: trade directories, professional registers, club membership lists and the rating valuation list of dwellings at the local council offices. It is essential to try to find a suitable sampling frame before beginning to select a sample, although it will not always be possible to find such a frame. Suppose you are conducting a survey which involves interviewing housewives with two children: there are no directories listing housewives with two children. The only way of finding that population is to go out and look for it, either knocking on doors or stopping people in the street and asking them if they possess those characteristics.

The final, and obvious, point to be made about sampling frames is that they should be as up-to-date as possible.

2.4 Sampling Methods

There are various ways of actually selecting a sample. For probability, i.e. random sampling, a sampling frame is required. Quota sampling, which is an example of non-random sampling, does not need a sampling frame.

(i) **Simple random sampling**, sometimes known as the *lottery method*. This is the best method from a theoretical viewpoint of selecting a truly random sample. All the items in the population are given a number and pieces of paper each with one number on are placed in a drum or hat. The numbers are selected one at a time, with the drum being revolved or the hat shaken between each selection, until the required sample size is reached. If the population is large, then an alternative method using a computer-generated table of random numbers (see Table H) is