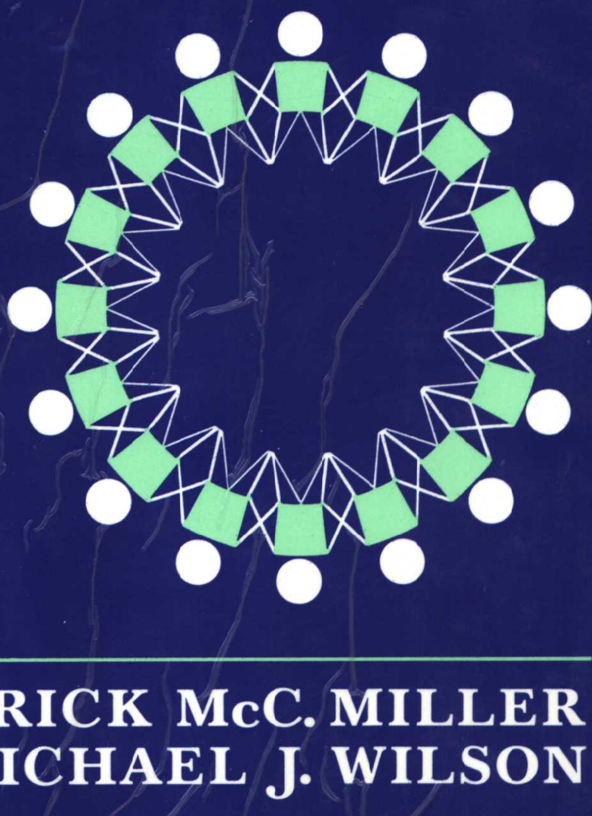
A DICTIONARY OF SOCIAL SCIENCE METHODS



PATRICK McC. MILLER
MICHAEL J. WILSON

A Dictionary of Social Science Methods

P. McC. Miller
Medical Research Council Epidemiology Unit,
Royal Edinburgh Hospital

and

M. J. Wilson
Faculty of Social Sciences
The Open University

JOHN WILEY & SONS
Chichester . New York . Brisbane . Toronto . Singapore

Reprinted March 1984

Library of Congress Cataloging in Publication Data:

Miller, P. McC.
A dictionary of social science methods.

Includes index.
1. Social sciences—Methodology—Dictionaries.
I. Wilson, M. J. (Michael John), 1939- . II. Title.
H41.M54 1983 300'.1'8 82-13681

ISBN 0 471 90035 4 (cloth)
ISBN 0 471 90036 2 (paper)

British Library Cataloguing in Publication Data:

Miller, P.McC.
A dictionary of social science methods.
1. Social sciences—Methodology—Dictionaries
2. Social sicences—Fieldwork—Dictionaries
I. Title II. Wilson, M. J.
300'.72 H61

ISBN 0 471 90035 4 (cloth)
ISBN 0 471 90036 2 (paper)

Typeset by Activity, New Canal, Salisbury.
Printed by Page Bros. (Norwich) Ltd.

Acknowledgements

The idea for this dictionary first arose from our work in preparing materials for a course on research methods in the social sciences for The Open University. We are grateful to the University for allowing us to re-use a small part of that material and to those colleagues with whom we worked in preparing the course. We should emphasize that neither the University nor our colleagues are in any way responsible for this book and our thanks to them are due for what we learnt in our collaboration. Especial thanks to: Martin Bulmer, John Bynner, Judith Calder, Peter Coxhead, Jim Davis, Jeff Evans, Martyn Hammersley, Cathie Marsh, Desmond Nuttall, Bram Oppenheim, Albert Pilliner, and Betty Swift.

We are grateful to Tiiu Miller and Glen McDougall who both gave sustained and thoughtful advice throughout the preparation of the dictionary and to Aillinn Wilson for editorial help in the planning of a long and complex work.

Yvonne Honeywell at The Open University, Caroline Urquhart and Joyce Greig at the MRC Unit have all helped with the lengthy and difficult typing which the manuscript required and we are in their debt.

Finally, our families deserve our particular thanks for their forbearance and help during the dictionary's preparation.

P.McC.M.
M.J.W.
April 1982

Preface

The major purpose of this dictionary is to collect in one convenient source accounts of the current methods of inquiry which the empirical social sciences share in common, together with accounts of methods peculiar to particular disciplines and which are (or ought to be) of wider interest. Mere definitions have a limited usefulness since they may give the essence of a term or method but lose sight of the controversies, plural meanings, and above all of the context which surrounds its use in the social sciences. It is, then, a second important purpose of this book to extend beyond definition only and to try to explain, illustrate, and to set in context the majority of the terms which this book contains. Setting in context has been achieved in two ways: within the entries themselves, although the extent to which this has been done depends on the length of the entry, and by a system of internal cross-referencing which indicates to the reader where to find terms relevant to a particular term under discussion. The system of cross-referencing is fully explained in 'How to use this Dictionary' (page ix).

The selection of terms for a dictionary of methods is doubly difficult because of the indistinct line to be drawn between methods as such and substantive theory, and again between methodological issues and philosophical terms and debates. We have resolved these choices as well as we can and kept the book to a reasonable length by excluding for the most part substantive theories and techniques for intervention. There is no coverage, for example, of concepts drawn from learning theory or from behaviour therapy.

We have also left out terms which are commonly confined to discourse within the philosophy of science but we have thought it best to include as entries those terms and concepts which, though philosophical, frequently enter into debates about research practice or the interpretation of research. Thus, we include a full entry on CAUSATION as a question actively debated in the findings of substantive research but we exclude any reference to primitive terms on the grounds that the problems which they raise are still of concern mainly to philosophers of theory construction.

The large body of statistical models and terms which readers will expect to find in a work such as this has been kept to manageable proportions by omitting the more obscure statistics and concentrating on the more commonly used ones. Although, doubtless, our choices on what to include and what to exclude will offend some, we hope to have reflected the general practice of research in the social sciences. When in doubt as to whether to include a term or not we have used the criterion of its relevance to the current practice of research rather than its potential theoretical importance.

We have dealt with statistical theory as far as possible from the practitioners' point of view and we have made no attempt to be mathematically rigorous or to derive expressions strictly. We have, however, given formulae and equations where useful (as is the usual case) and all the symbols used are fully explained. Note that we have adopted the convention of using greek symbols for population statistics and roman ones for sample statistics.

Lastly, we are certainly aware that a work such as this will contain errors; some in the choice of what to include and what to leave out in the way of entries, and some in what we write about particular concepts and methods. We leave it to the reader to judge.

P.Mc.C. Miller
Medical Research Council Epidemiology Unit
Royal Edinburgh Hospital

M.J. Wilson
Faculty of Social Sciences,
The Open University

How to use this Dictionary

The headings of entries in the text are of three different kinds, each of which is shown by a distinct typeface. Bold capitals (e.g. **FACTOR ANALYSIS**) signify a long entry which defines a method or term of central importance in the methodology of the social sciences. Such entries place a term in context, list and explain the vocabulary associated with it, discuss its uses in the literature, and assess its relative strengths and weaknesses. Under **FACTOR ANALYSIS** will be found a general account of the logic of the method followed by a briefer one of the different types of factor analysis and their specific terms, and ending with a discussion of the main uses and limitations of the technique.

Ordinary bold type (e.g. **Anchor Definition**) signifies an entry which has more or less a standard dictionary definition, although usually more expansive in form and frequently containing an illustrative example.

Bold italics (e.g. ***Biassed Estimator***) represent an entry which is simply a heading cross-referenced to another entry where it will be found to be explained as part of a discussion of related terms.

The *Dictionary* has been written on a system of internal cross-referencing in order to avoid excessive repetition. In an entry, terms which are defined elsewhere in the book are shown by the corresponding typeface so that the reader will know at a glance how long an entry to expect if he or she chooses to look it up. The reader is invited to follow up such cross-references until satisfied that no more useful knowledge is to be gained.

For example, under **ESTIMATION** will be found in part:

'... estimation of **regression weights** or coefficients in simple or multiple **REGRESSION ANALYSIS** follows the principles above but ...'

The reader who is unsure of either or both of the cross-referenced terms will know that a long entry for regression analysis and a short one for regression weights will be found in the appropriate place and may be quickly consulted if desired.

A

Abscissa The horizontal axis in a two-dimensional graph. The vertical axis is the *ordinate*. More modern usage is to call the horizontal axis the X-axis and the vertical one the Y-axis. See also **Graphic Presentation**

Absolute Difference The difference between two quantities taken irrespective of the sign, so that both negative and positive differences are treated as if positive. The absolute difference between two quantities x and y is symbolized as $|x - y|$.

Access Legitimate entry to a research field with reasonable freedom of action within it.

Achievement Test A test which measures how much has been learnt or is known by a subject, as distinct from an *aptitude test* which measures *ability to learn*. Achievement tests are published commercially. Many of them have **norms** relative to other people rather than to an absolute standard.

Acquiescence Response Set A tendency to agree with any statement put, regardless of its content. The effects of possible acquiescence in questionnaires may be controlled — usually by keying items in opposing directions so that agreement sometimes scores and sometimes does not score on the dimension being measured.

Action Research Research where instead of minimizing the impact of the investigations on the subjects or individuals under study, changes are purposefully introduced in order to study what, if any, effects occur.
Examples of action research include the use of intensive reading development programmes in classrooms, treatments for alcoholism, new types of penal institutions, and community development programmes in inner city districts. Ideally, changes (the 'action') should not be introduced before the existing field of study has been fully investigated so that some accurate 'baseline' is available against which to measure the results of the planned intervention. This intervention itself should be carefully designed so that unknown factors (or variables) are not brought into play accidentally and the changes may be attributed unequivocally to the researchers' deliberate manipulations. If these points are adequately met then action research becomes a form of **field experiment**, but with the reservation that historical or **maturational** changes are not allowed for. That is to say, the field of study may have changed because of the elapse of time whether or not 'action' or intervention by the researchers had taken place. In reading development programmes, for example, the fact that the group under study may be significantly older after the programmes has been applied might well, in itself, lead to improvements in reading attainment.
Action research reports frequently show a gulf between the researchers who design the study and the practitioners who put it into effect. Such a lack of co-ordination between the two central tasks of action research may weaken its validity.

Additivity In MULTIPLE REGRESSION analysis, if the value of the **dependent variable** can be accurately estimated by summing the particular values of the **independent variables** (each weighted by a **regression coefficient**) and the constant term then the equation is an additive one.
The converse of additivity is **interaction** where the value of the dependent variable will depend on how the value of one independent variable may interact with the value of another.
More generally, if an effect is produced by several variables acting simultaneously, then the statistical model is additive if and only if the portion of the effect contributed by an independent variable does not depend on the value of any other independent variable. In Mendelian genetics, for example, the contribution of father's height to offspring's height will not depend on mother's height and vice versa, and offspring's height can be estimated simply by $H_0 = \alpha + \beta\,(H_f + H_m)$.
Additivity may be an asumption (which can be tested) of models or a requirement of a test or analytical technique.

Adjusted Mean A mean which has been adjusted, using the technique of **covariance analysis**, in order to allow for the effects of an extraneous uncontrolled variable.

Age-Cohort-Survival Method Forecast of the future size and composition (age and sex) of a

human population, either national or regional, based on assumptions about future rates of fertility, mortality, and migration.
The projection starts from a base year (t) whose *age-specific fertility* and *death-rates* are known and are applied to age cohorts of the current population in order to calculate survivors (from death-rates) who will pass into the next age cohort, and live births (fertility) who will progress through the age cohorts and be subjected to successive death-rates and fertility-rates. The effect of in- and out-migration on both the size of the population and on its fertility and death-rates is also added to the forecast. Forecasts may be made for successive time-intervals ahead of the base year ($t + 1$, $t + 2, \ldots, t + n$) allowing for age cohorts changing through migration and ageing steadily and thus requiring different age-specific rates. Projection errors are cumulative and systematic because rates are fixed in the last year of observation (t) and may change, particularly in the case of fertility rates. Population forecasts are continually revised to use the most recent rates of death, birth, and migration, as they are incresingly prone to error the further ahead a projection is made.

Age-specific Death-rate See under **Death-Rate**

Age-specific Fertility-Rate See under **Birth-Rate**

Age-specific Sex Ratio See **Sex Ratio**

Agreement Coefficient An agreement coefficient measures not only the extent to which two variables are correlated or associated but also the extent to which they have the same values in correspondence. Agreement in this sense implies correlation, but correlation does not imply agreement. Agreement is, therefore, much the more demanding concept of the two. Fisher used an agreement coefficient for his analysis of data from sets of twins, where the theoretical demands of the study required twins to show identical measurements in mental and physical traits.
Various agreement coefficients are available depending on the level of measurement — in the **nominal** case **Scott's** π, in the **ordinal** case **Kendall's Tau**, and in the **interval** case, Fisher's **Intra-class correlation** coefficient.

AGGREGATION
The placing of discrete observations on individuals (companies, subjects, etc.) into groups so that individual observations can no longer be recovered except by going back to the original sources of data. This is frequently not practicable or possible in the case of official statistics or in any case where the original data before aggregation have been lost or have been withheld. *Disaggregation* is either re-partitioning the aggregated groups into new and smaller groups, or at the extreme, recovering data at the individual level.
All official statistics aggregate data. This leads to clarity of presentation (by removing excessive detail) especially in tabulated forms, protection of individuals from possible identification, and to easier analysis. Unfortunately, in many cases, the aggregated groups do not match the investigator's theoretical objectives. The **Standard Industrial Classification** (SIC), for example, is widely used in the social sciences where it forms the variable of the industrial sector in a number of statistical series, but it reports data for very large aggregates of companies or sections of the labour-force. SIC data has high **variances** on useful variables such as size of work-force, capital intensity, and technological level both within and between aggregate sectors. Thus, comparison of sectors on a variable such as man-days lost in industrial disputes (from the **Employment Gazette**) will be difficult because extraneous variables such as company size will be confounded with industrial sector.
Aggregation may be a source of special difficulties in economic analysis over and above the general problems noted above. Many economic theories are based on models of microeconomic behaviour and yet can be tested only on data which is highly aggregated and which has to 'represent' microeconomic activity, perhaps not accurately if the aggregate group is not a homogeneous one with respect to the variable or variables under study.

AH5 Test A **group-administered** intelligence test devised for use with highly intelligent people. It aims to discriminate amongst these more precisely than other intelligence tests would. It can be used from age 13 years and there are **norms** for several groups including university students, grammar school children, and engineering apprentice applicants.

Alpha Coefficient (Cronbach's Alpha) The mean value of all possible **split half reliability** coefficients for a test. It is given by

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\text{sum of variances of question scores}}{\text{variance of total scores}}\right)$$

where n is the number of questions in the test.
See also RELIABILITY

Alpha (α) Probability of Type I Error See **Type I Error**

Alternate Forms Reliability RELIABILITY assessed using two equivalent forms of a test on different occasions.

Alternative Hypothesis An hypothesis which is 'alternative' to the **null hypothesis** and states that differences will be found between two or more groups on some particular variable. A second meaning is an hypothesis which explains results equally as well as the hypothesis being tested.

Ambiguity of Testing When an hypothesis is tested it is generally possible that the results obtained may be explained at least in part by some hypothesis other than the one tested. This is ambiguity of testing.

ANALYSIS OF VARIANCE (ANOVA)
A flexible technique which has many uses. Two of the most important are to test whether there are significant differences between the means of several different groups of observations, and to test the significance of simple and MULTIPLE CORRELATION coefficients. The test is particularly useful where there are several means or correlations to test at the same time. Testing each separately by means of a **Z-test** or a ***t*-test** would increase the probability of falsely rejecting the **null hypothesis** (probabilities of error sum across tests, and with twenty tests and a 5 per cent significance level it is virtually certain that a spuriously significant result will be reported); analysis of variance helps to overcome this problem.

An analysis of variance is termed simple or one-way if only one independent variable is involved, e.g. if one wished to see whether reading ability (**dependent variable**) was different for pupils of high, middle, and low socio-economic status (**independent variable**). It is said to be multiple or complex if there is more than one independent variable, e.g. if it was desired to test whether both sex and socio-economic status affected reading ability. In multiple analysis of variance, **interactions** may arise. That is to say, the various levels of one independent variable may not affect the dependent variable in the same way within all the levels of another independent variable. For example, in the two-way analysis cited above socio-economic status might affect the reading ability of girls and boys in different ways. Interactions are tested after the **main effects** (i.e. overall effects) of the independent variables have been assessed.

Theoretically, analysis of variance is a special case of the general **linear model** which also underlies MULTIPLE REGRESSION ANALYSIS. The first step in an analysis of variance is to obtain *sums of squares*. The *total sum of squares* (TSS) is the sum of the squares of the deviations of all the measurements from the mean of all the measurements (the grand mean). The TSS may be split into components, the most basic of which are an *explained sum of squares* (ESS) (also known as the *between-groups sum of squares*) and a *residual* or *within-groups sum of squares* (RSS). ESS is the proportion of TSS which may be explained by a small number of factors/variables such as sex, socio-economic status, etc. in which a researcher may be interested. RSS is the proportion of TSS which may not be so explained. TSS = ESS - + RSS (and ESS/TSS = R^2, the square of the multiple correlation coefficient for predicting the dependent variable from the explanatory variables). ESS can often, if desired, be split up into further parts corresponding to each single explanatory variable and/or the interaction(s) between the explanatory variables. To each sum of squares there is a corresponding *mean square* obtained by dividing the sum of squares by the appropriate **degrees of freedom**. There is thus an *explained mean square* (EMS) and a *residual mean square* (RMS), and, if desired, mean squares corresponding to single variables or interactions. The point of computing mean squares is that they are **variance** estimates and, as variance estimates, they may be compared using the ***F*-ratio**. The RMS is taken as the standard of comparison and if any of the other mean squares are shown to be significantly greater than the RMS then the variables corresponding to these mean squares explain a significant proportion of the total variance.

When this is so, such conclusions may be drawn as that there *are* differences between group means, that a multiple correlation is significant, etc.

Once it is established that differences between means are significant, the researcher may wish to know which means differ from the rest (e.g. whether it is the high socio-economic status pupils which differ from the middle and low status pupils or whether it is the middle group that is different, etc.). There are several techniques available for this purpose, and which one to use depends on whether it was hypothesized in advance that a particular difference would exist, i.e. there was a *planned comparison* or whether the researcher was simply exploring the data, i.e. making **unplanned comparisons**. For the planned comparison situation a modified form of the *t*-test is appropriate. Examples of procedures for use in unplanned comparisons are the **least significant difference test**, the **Tukey test**, the **Scheffé method**, the **Newman–Keuls test**, and the **Duncan multiple range test**.

Analysis of variance makes powerful assumptions about **interval scale** measurement, NORMALITY OF DISTRIBUTION of scores, and homogeneity of certain variances (**homoscedasticity**). Furthermore, complex analyses of variance become tricky to perform and require special procedures if the numbers of cases in all the cells are not equal or if there are missing values. The normality of distribution problem may sometimes be overcome by an appropriate TRANSFORMATION of the scores, and, in any case, it has been shown that analysis of variance is robust to violations of its basic assumptions, and particularly so when cell numbers are equal. It is uniquely appropriate to analysing EXPERIMENTS, where the experimenter can control the conditions, randomly assign cases, arrange equal numbers in the cells, and make useful and accurate measurements. Survey data, where such direct manipulation is impossible, are usually more easily and appropriately analysed using multiple regression directly.

Analytic Categories In PARTICIPANT OBSERVATION, analytic categories are categories referring to the structure of the field developed by the observer during the course of field-work.

Analytic Concept A concept developed by a researcher to assimilate some phenomenon to a structure. It is to be distinguished from 'a members' concept', i.e. one employed by the people studied. In social science there is always some danger that analytic concepts bear no relationship to ordinary peoples' consciousness. Therefore one of the major requirements of ethnography is that analytic concepts be explicitly related to members' concepts.

Analytic Induction See under PARTICIPANT OBSERVATION

Anchor Definition In order to improve reliability of a rating scale some points on it may be given brief descriptions or **anchor definitions**. For instance, on a 10-point rating scale of intelligence, scale point 10 might be described as 'very intelligent indeed — a genius', point 7 might be called 'highly intelligent', point 3 'below average', and point 2 'extremely unintelligent'.

Anonymity If a research subject is promised anonymity, it means that the researcher has promised that the research is to be conducted in such a way that the identity of the subject is unknown, even to the researcher.

Aptitude Test See under **Achievement Test**

Arcsine Transformation A TRANSFORMATION which may be applied to proportions (fractions, percentages, cases per thousand population, etc.).

The DISTRIBUTION of a proportion (P) tends to be very bunched at the extreme values and its **variance**, i.e. $P(1 - P)$ depends on the value of P. The **arcsine transformation** spreads out the extreme values giving an approximately normal sampling distribution and usually renders the variance of the scores approximately constant, no matter what the original value of P.

The arcsine transformation is given by:

$A = 2 \text{ arcsine } \sqrt{P}$,

i.e. twice the angle (measured in radians) whose trigonometric sine is the square root of the proportion being transformed.

It has little effect and may not need to be applied if the proportions being measured all lie between 0.25 and 0.75.

Area Sampling See under SAMPLING THEORY AND METHODS

Arithmetic Mean The sum of values of a variable for all the observations in a data set divided by the total number of observations, i.e.

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

As a **measure of central tendency** the arithmetic mean takes into account all the observations in the set and can be influenced by atypical extreme values. It is appropriate only when measurement is at least on an **interval** or **ratio scale**. The sum of deviations of all observations from the mean is zero; from the knowledge of the mean and the number of observations, the total of all the observations may be found.

Array All the values in a set of data set out together. An *array mean*, in the context of a **joint distribution**, is the mean of a set of data for one variable for a constant value of the other.

Artefact The result of human activity rather than of natural processes. In the social sciences the term most commonly denotes a spurious result, which arises purely or at least partly, out of some feature of the design of a study.

Artificiality Artificiality enters into experiments and surveys which rely on the use of artificial contexts, laboratory situations, and *structured interviews* when collecting data. Since these contexts are established and controlled by the researcher they may be unrepresentative of everyday life.

Association See under **RELATIONSHIP**

Asymmetric Measure of Association Some measures of association give different numerical results depending on which variable is taken as the **independent variable**. These are known as **asymmetric measures of association** — an example being the ***d*-statistic.** A **symmetric measure of association** (e.g. the **Pearson *r***) gives the same numerical result no matter which variable is chosen as the independent variable.

Asymmetric Relation A relation which holds strongly in one direction only. Causation is an example. To say 'smoking causes lung cancer' does not imply that 'lung cancer causes smoking'.

Asymptotically Unbiased Estimator See under ESTIMATION

Atomism (Social) The assumption that discrete elements in the field of investigation may be joined together into larger wholes, these larger wholes being describable in terms of the discrete elements and nothing else.
See also **Gestalt**

Attenuation The correlation between two variables is always somewhat lessened or **attenuated** by the errors of measurement in both. When it may be assumed that the **error scores** are uncorrelated with the **true scores** and with each other a correction for attenuation is sometimes made, given by:

$$r_{tt} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}}$$

where r_{tt} is the corrected correlation, r_{xy} is the observed correlation, and r_{xx}, r_{yy} are the reliabilities of the tests. Corrected correlations cannot be used in REGRESSION ANALYSIS and the technique requires caution as corrected correlations can sometimes exceed unity.

Attitude Dimension A unitary and continuous **dimension** along which it is postulated that an attitude may vary. An **attitude scale** is a set of items designed to assign scores to individuals on this dimension.

Attitude Scale A scale which assigns a numerical score to a person on an underlying **attitude dimension** and thus makes comparisons between people possible. There are several established procedures for setting up attitude scales including the **Thurstone, Likert**, and **Guttman** methods and the McKennell procedure (alpha scaling).
See also ATTITUDE SCALING

ATTITUDE SCALING
The process of empirically constructing a method for measuring a particular concept, attitude, or trait within an individual. There are several properties which are desirable in an ideal scale. It should be **unidimensional**, measuring only one concept and not influenced by others. In particular, there should be no **response bias**. It should be graduated into a number of equal intervals and thus be linear. It should be RELIABLE, giving the same answer when applied to the same individual on different occasions, unless the concept being measured has itself altered. It should be VALID, measuring exactly what it is intended to measure, and it should be *reproducible*, that is, there should be only one way of obtaining a given score.
In the social sciences such ideal scales are virtually unobtainable for a number of reasons. There are three common methods of scale construction.
The *Thurstone Scaling Method* produces a set of items with which the subject is invited to agree or disagree. There may be, perhaps, 20 or 30 of these items in a typical scale and each one has been assigned a score showing its position on the scale. The subject scores on the item only if he agrees with it. The **median score** of the items to which the subject agrees is taken as his score on the whole scale. The steps in establishing a Thurstone scale are as follows: first, a large pool of items, perhaps 100–150, are written reflecting the whole range of the concept to be scaled; secondly, a number of judges, preferably people similar to the subjects on whom the scale will be used, are asked to order the statements or items into a number of categories, usually eleven. The judgements for each item are then examined and the median rating is taken as the item score. A small number of items on which there is good rating agreement between judges and which together cover all the eleven categories is then chosen to constitute the final scale.
In the *Likert Scaling Method* the subjects have to register their responses to each item on a five (or seven) point scale, ranging from strongly agree to strongly disagree. Each point of the item-scale carries a numerical value. The score on the whole scale is the sum of the scores on each of the items. The scale is constructed by trying out a large number of items on a sample of individuals similar to the ones on whom the final scale is to be used. **Item**

analysis, usually in the form of correlating the grouped scores on each item with the total score for the individual on all the trial items, is then used to select the best items, i.e. those with the highest correlation coefficients.

The *Guttman Scaling method* results in a series of items arranged in a hierarchy or cumulative scale such that agreement with an item implies agreement with all the items below it in the hierarchy. An individual's score is the number of items with which he agrees. Suitable items for a concept are selected by scalogram analysis on a large pool of items which have been tried out on a sample of respondents. The items do not have to be dichotomous but, if they are not, a cut-off point is usually established to make them so. It is uncommon to find a set of items which form a perfect Guttman scale and the *coefficient of reproducibility (R)* is an indication of how good the final scale is. It is given by:

$$R = 1 - \frac{\text{number of errors}}{\text{number of items} \times \text{number of respondents}}$$

An error is a deviation from the ideal Guttman pattern of a perfect hierarchy of items. (See under **Guttman Scale**.) A perfect Guttman scale is necessarily unidimensional in scaling the concept. It is desirable to establish a scale on at least one sample of 100 or more respondents and the minimum criterion for satisfactory reproducibility is usually taken as $R = 0.9$

Thurstone scales on the whole tend to be only moderately good at meeting the desirable criteria for an ideal scale and they need considerable labour to construct, unlike Likert scales which are much easier to construct and which often have higher reliability. However, Likert scales may not always be unidimensional and they have no reproducibility — the same score on a Likert scale may be obtained in a large number of different ways. Guttman scales have excellent reproducibility and unidimensionality but this is often to be attained only by severely limiting the content area which the items cover. There is no guarantee in advance that a Guttman scale to measure a certain attitude can be constructed. If it can, its construction tends to be long and laborious.

There are many modifications to these basic methods of scale construction. In particular, FACTOR ANALYSIS may be used to construct several scales simultaneously from a large pool of items, and there are many types of precaution which may be taken against response bias. Items may be keyed so that for about half of them a 'yes' answer (and for the other half a 'no' answer) indicates presence of the attitude being measured. *Buffer items*, which have nothing to do with the attitude being measured but are there to camouflage the purpose of the questionnaire, may be included. There may also be *lie-detector* or *social desirability items* included which, if scored too often, mean that the respondent is answering so as to leave a good impression on the tester. This type of bias may also be reduced by using *a forced choice format* in which items have two or more alternative answers each of which is equal on social desirability but for which the respondent has to say which is most and which least characteristic of himself. One relatively new technique, the *M'Kennell procedure* takes as its starting point the opinions/attitudes of a sample of people interviewed freely in a preliminary study. A large pool of Likert-type items is then written and this is administered to a new sample. **Elementary linkage analysis** is applied to determine how many CLUSTERS there are in the responses. This is followed by **principal components analysis** to extract as many components as the number of clusters which have been found and rotating by the **varimax method**. This sets up a number of **orthogonal** attitude dimensions. The final step is for the researcher to decide on the minimum level of reliability which is acceptable and is measured by **Cronbach's alpha**. The researcher then chooses a small number of items which together will attain the desired level of alpha. This method meets many, but not all, of the criteria for an ideal scale.

Attitude Survey A study, on a properly drawn sample of a specified population, designed to find out what people in that population feel about some particular issue. Attitude surveys usually use carefully constructed, well STANDARDIZED questionnaires.
See also **Sample Survey**

Attitudinal Item See **Item**

Attribute A qualitative measurement assigned to objects or to individuals. Useage of the term varies greatly. **Nominal categories** are certainly attributes, e.g. religious affiliation classified as Protestant, Catholic, and Jew is a set of attributes without any imputation of scaling. Some regard even **ordinal measures** as attributes, e.g. examination grades. Any set of categories which implies an ordering is an **ordered attribute. Interval** and **ratio scales** are unquestionably true variables rather than attributes.
See also SCALE OF MEASUREMENT, **Categorical Variable**

Auto-correlation See under ESTIMATION

Auxiliary Hypothesis An hypothesis which may afford an alternative explanation of some phenomenon and which needs to be ruled out before conclusions are reached about the main hypothesis under test.
See also **Ambiguity of Testing**

Average Often another term for **arithmetic mean**, but is more generally an undefined **measure of central tendency** of the values in a data set or **frequency distribution.**

Average Deviation See **Mean Deviation**

Axis See under **Graphic Presentation**

B

Slope of Least Squares Regression Line See **Regression Coefficient**

Bales Interaction Analysis A system which enables an observer to classify the behaviour of people in groups. Each social interaction between group members is clasified into one or more of twelve categories grouped into four broad areas — social–emotional area positive, social–emotional area negative, task area attempted answers, and task area questions. Examples of the twelve categories are: 'Shows tension release, jokes, laughs, shows satisfaction' (social–emotional area positive) and 'asks for orientation, information, repetition, confirmation' (task area questions).

Bar Chart A diagram for displaying **nominal scale** data. For each category in the data a bar is drawn from the axis, usually the horizontal axis, to a height proportional to the number of observations in the category. The bars are all made the same width, are separated from each other along the horizontal axis, and may be placed in any order along the axis.

Base Data A set of data, e.g. retail prices for a particular month in a particular year, which serves as a baseline against which changes can be measured.

Baseline A standard of performance which exists prior to an experimental treatment and against which that treatment may be measured.

BAYESIAN STATISTICS

Classical statistics are basically concerned with describing and comparing samples of data and making valid inferences about the underlying population parameters. Bayesian statistics afford a means for combining new information with other information already known in order to refine the conclusion to be drawn.

The backbone of Bayesian statistics is Bayes theorem. This takes various forms according to the nature of the problem. The simplest is:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/\bar{A})P(\bar{A})},$$

where

$P(A/B)$	is the probability that A is true given that B has occurred,
$P(B/A)$	is the probability that B will occur given that A is true,
$P(A)$	is the probability that A is true before knowing whether or not B occurred,
$P(B/\bar{A})$	is the probability that B will occur if A is *not* true,
$P(\bar{A})$	is the probability that A is not true before knowing about B.

For example, suppose political party A has a chance of 0.5 of winning the next election [$P(A)$]. If they win, the probability of tax increases [$P(B/A)$] is 0.8. If they lose, the probability of tax increases [$P(B/\bar{A})$] is 0.5. A few months later taxes *have*, in fact, gone up. The probability that party A won is now

$$\frac{(0.8)(0.5)}{(0.8)(0.5) + (0.5)(0.5)}$$

$$= 0.62$$

That is, the probability that party A won should be revised from 0.5 to 0.62, given the information that taxes have gone up.

The original probability estimate [$P(A)$], before the addition of new information, is known as the *prior probability* of event A. The conditional probabilities of event B depending on A, [$P(B/A)$ and $P(B/\bar{A})$] are known as *likelihoods*. The final probability estimate that A has occurred [$P(A/B)$] is termed the *posterior probability* of event A.

Both the prior and the posterior probabilities sum to one over all possible event outcomes. The likelihoods, however, do not.

Bayes theorem may be extended to cover more complex cases, e.g. where there are several different discrete outcomes (several parties with a chance of winning an election) and many separate pieces of information to be used in refining the prior probability. It may also be generalized to the case where all the variables are measured on **continuous** scales.

Bayesian statistics are useful in situations where the prior probabilities and the likelihoods are easily inferred or even approximated. In these cases, use of Bayesian methods to update existing information may be much preferred to the trouble and expense of data collection on a new **random sample**. Furthermore, it has been shown that these methods are much sounder than informal subjective refining of the prior probabilities. In most situations informal refinement tends to be much too conservative.

Beck Depression Inventory (BDI) An inventory developed to measure the severity of a depression.
There are 21 items each scored from 0 to 3. Each item consists of a series of statements related to depression, one of which is to be chosen. These alternatives are usually read out to the patient who then indicates his response. The BDI has satisfactory RELIABILITY and VALIDITY and has been widely used. There is a short form available.

Before–after Study A study which aims to assess the effects of some given **treatment**, e.g. a propaganda exercise or training programme, by taking measurements before and after its occurrence. Often the study will include a **control group** measured at the same time as the **experimental group** but not exposed to the treatment.
See also EXPERIMENT

Best Linear Unbiased Estimator (BLUE) See under ESTIMATION

Beta Coefficient See **Standardized Regresion Coefficient**

Beta* (β) *Probability of Type II Error See **Type II Error**

Beta Weight See **Standardized Regression Coefficient**

Between-Groups Variance The variance of a group of mean values about the mean of these means. In the ANALYSIS OF VARIANCE technique the between-groups variance is compared to the *within-groups variance*, i.e. the average variance of the observations about their group means, in order to determine whether there are significant differences between any of the means.

Between-Groups Sum of Squares See under ANALYSIS OF VARIANCE

BIAS
Deviation from a correct conclusion because of a flaw in the procedures which led to it.
The main identified sources of bias in social science research lie in the researcher, in the data-gathering methods, and in the sampling procedures.
The researcher may be biased in the ordinary sense of the term. He or she may, often quite unconsciously, hold tenaciously and unreasonably to erroneous attitudes and beliefs. These may influence the hypotheses selected for testing, the manner in which the research is conducted, and the interpretation of the findings. In particular, bias may be transmitted in various ways during the testing procedures. In the *experimenter bias effect*, in knowing the hypothesis to be tested the experimenter (perhaps quite unconsciously) uses subtle influences, such as extra encouragement, tone of voice, or selective probing, to extract answers from the subject which confirm the hypothesis. Where global judgements are to be made there may be a *halo effect* in which the rater's general attitude dictates the ratings for particular items — for example, the prettiest girl is rated the most intelligent. When a survey is being done *interviewer bias* will occur when the interviewer's social class (perhaps effected through his or her manner of speaking) influences what the respondent says. Where the researcher is a participant-observer there may be *over-rapport*, in which case he or she identifies so much with the subject that he or she adopts a particular perspective as obvious and beyond question.
Research procedures in themselves may often introduce bias. In general, people who know that they are being observed may not behave in their normal manner. In the *Hawthorne effect* the extra attention which is given by the researcher has an effect quite distinct from that predicted by the experimental manipulation. More specifically, the interview as a technique affords many opportunities of introducing bias, e.g. by the use of leading questions or selective coverage of the field. There are a number of systematic ways in which the answers to questionnaires may be biased (*response biases*). *Social desirability responding* occurs when answers are altered to show the subject in a more favourable light *vis-à-vis* the interviewer. *Acquiescence* is where 'yes' is the answer no matter what the question. In *extreme responding* the subject exaggerates his or her answers by always choosing the extreme amongst alternatives. There may also be an *order effect* in which the order of the items making up a test or questionnaire influences the responses to the items taken as individual elements of the test.
Finally, bias may occur because of faulty sampling. A *biased sample* is one which fails to reflect adequately the population it is meant to reflect. This may occur because proper sampling procedures have not been followed or

because certain types of people refuse to be interviewed, i.e. where the non-respondents are not a crosssection of the population.

Available methods to combat bias include:

(i) having people who dó not know the hypothesis under test conduct the data-gathering,

(ii) having **ratings** made by several different judges,

(iii) careful STANDARDIZATION of interviews,

(iv) the inclusion of suitable **control groups**,

(v) **counterbalancing** the item order,

(vi) keying responses to items so that the answers which indicate a particular attitude or trait are sometimes 'yes' and sometimes 'no',

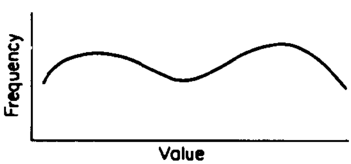
(vii) using **forced choice items** in which the items have equal social desirability,

(viii) vigorously searching for respondents who have been selected in the sample but are not easy to contact.

Biased Estimator See under ESTIMATION

Biased Sample A sample which fails to reflect adequately the characteristics of the population it is meant to reflect.
See also BIAS

Bimodal Distribution A distribution which has two different values around which observations tend to cluster. When plotted it exhibits a characteristic double hump shape thus:



See also DISTRIBUTION

Binomial Distribution A distribution which describes the frequencies of outcome for an event which is limited to occurring with a known probability (p). For instance, if an unbiased six-sided die is rolled 30 times the event 'a six' will occur about five times. If the 30 rolls are now repeated an infinite number of times the binominal distribution will describe the frequency of the event 'a six'— occasionally there will be no sixes at all and occasionally many more than five sixes, but usually about five. The binomial distribution has a mean Np and a standard deviation $\sqrt{Npq}$, where N = sample size (number of die rolls), p = probability that the event will occur, and $q = 1 - p$ = probability the event will not occur.

It has a similar shape to the NORMAL DISTRIBUTION but is skewed—particularly when p differs markedly from 0.5 and N is small. As N becomes large the binomial distribution becomes indistinguishable from the normal distribution.
See also DISTRIBUTION

Binomial Test When a sample can be divided into two discrete classes (e.g. men and women, rich and poor, achievers and non-achievers) the binomial test is appropriate for seeing whether the proportion in one of the classes corresponds to some expected proportion. This is done using the binomial theorem. For example, supposing it is known that one-third of a population are high achievers and in a sample of 20 low status people only two high achievers are observed, then, according to the binomial theorem, the chances of observing no high achievers is:

$$\frac{20!}{0!20!} \cdot \left(\frac{2}{3}\right)^{20};$$

the chances of observing one high achiever is:

$$\frac{20!}{1!19!} \left(\frac{1}{3}\right)^{1} \cdot \left(\frac{2}{3}\right)^{19};$$

the chances of observing two high achievers is:

$$\frac{20!}{2!18!} \left(\frac{1}{3}\right)^{2} \cdot \left(\frac{2}{3}\right)^{18}.$$

Therefore the probability of observing two *or less* high achievers by chance is the sum of all these, i.e. 0.0176 or 1.76 per cent. Thus, it may be concluded that those of low social status tend to be low achievers. As sample size increases, working out the probabilities by the binomial theorem becomes cumbersome and the normal approximation

$$Z = \frac{x - NP}{\sqrt{NPQ}}$$

may be used where N = sample size, x = observed number with a given outcome, P = probability of this outcome, and $Q = 1 - p$ is the probability of the other outcome. $\sqrt{NPQ}$ should be at least 9 before the approximation is used.

Bipolar Factor See under FACTOR ANALYSIS

Birth-Rate The crude birth-rate is the total of