



**Aalen–Johansen estimator:** An estimator of the *survival function* for a set of *survival times*, when there are competing causes of death. Related to the *Nelson–Aalen estimator*. [*Scandinavian Journal of Statistics*, 1978, **5**, 141–50.]

**Aalen’s linear regression model:** A model for the *hazard function* of a set of *survival times* given by

$$\alpha(t; \mathbf{z}(t)) = \alpha_0(t) + \alpha_1(t)z_1(t) + \cdots + \alpha_p(t)z_p(t)$$

where  $\alpha(t)$  is the hazard function at time  $t$  for an individual with covariates  $\mathbf{z}(t) = [z_1(t), \dots, z_p(t)]$ . The ‘parameters’ in the model are functions of time with  $\alpha_0(t)$  the baseline hazard corresponding to  $\mathbf{z}(t) = \mathbf{0}$  for all  $t$ , and  $\alpha_q(t)$ , the excess rate at time  $t$  per unit increase in  $z_q(t)$ . See also **Cox’s proportional hazards model**. [*Statistics in Medicine*, 1989, **8**, 907–25.]

**Abbot’s formula:** A formula for the proportion of animals (usually insects) dying in a toxicity trial that recognizes that some insects may die during the experiment even when they have not been exposed to the toxin, and among those who have been so exposed, some may die of natural causes. Explicitly the formula is

$$p_i^* = \pi + (1 - \pi)p_i$$

where  $p_i^*$  is the observable response proportion,  $p_i$  is the expected proportion dying at a given dose and  $\pi$  is the proportion of insects who respond naturally. [*Modelling Binary Data*, 2nd edition, 2003, D. Collett, Chapman and Hall/CRC Press, London.]

**ABC method:** Abbreviation for **approximate bootstrap confidence method**.

**Ability parameter:** See **Rasch model**.

**Absolute deviation:** Synonym for **average deviation**.

**Absolute risk:** Synonym for **incidence**.

**Absorbing barrier:** See **random walk**.

**Absorbing Markov chains:** A state of a *Markov chain* is absorbing if it is impossible to leave it, i.e. the probability of leaving the state is zero, and a Markov chain is labelled ‘absorbing’ if it has at least one absorbing state. [*International Journal of Mathematical Education in Science and Technology*, 1996, **27**, 197–205.]

**Absorption distributions:** Probability distributions that represent the number of ‘individuals’ (e.g. particles) that fail to cross a specified region containing hazards of various kinds. For example, the region may simply be a straight line containing a number of ‘absorption’ points. When a particle travelling along the line meets such a point, there is a probability  $p$  that it will be absorbed. If it is absorbed it fails to make any further progress, but also the point is incapable of absorbing any more particles. When there are  $M$  active absorption

points, the probability of a particle being absorbed is  $[1 - (1 - p)^M]$ . [*Naval Research Logistics Quarterly*, 1966, **13**, 35–48.]

**Abundance matrices:** Matrices that occur in ecological applications. They are essentially two-dimensional tables in which the classifications correspond to site and species. The value in the  $ij$ th cell gives the number of species  $j$  found at site  $i$ . [*Ecography*, 2006, **29**, 525–530.]

**Accelerated failure time model:** A general model for data consisting of *survival times*, in which explanatory variables measured on an individual are assumed to act multiplicatively on the time-scale, and so affect the rate at which an individual proceeds along the time axis. Consequently the model can be interpreted in terms of the speed of progression of a disease. In the simplest case of comparing two groups of patients, for example, those receiving treatment A and those receiving treatment B, this model assumes that the survival time of an individual on one treatment is a multiple of the survival time on the other treatment; as a result the probability that an individual on treatment A survives beyond time  $t$  is the probability that an individual on treatment B survives beyond time  $\phi t$ , where  $\phi$  is an unknown positive constant. When the end-point of interest is the death of a patient, values of  $\phi$  less than one correspond to an acceleration in the time of death of an individual assigned to treatment A, and values of  $\phi$  greater than one indicate the reverse. The parameter  $\phi$  is known as the *acceleration factor*. [*Modelling Survival Data in Medical Research*, 2nd edition, 2003, D. Collett, Chapman and Hall/CRC Press, London.]

**Accelerated life testing:** A set of methods intended to ensure product reliability during design and manufacture in which stress is applied to promote failure. The applied stresses might be temperature, vibration, shock etc. In order to make a valid inference about the normal lifetime of the system from the accelerated data (accelerated in the sense that a shortened time to failure is implied), it is necessary to know the relationship between time to failure and the applied stress. Often parametric statistical models of the time to failure and of the manner in which stress accelerates aging are used. [*Accelerated Testing*, 2004, W. Nelson, Wiley, New York.]

**Acceleration factor:** See **accelerated failure time model**.

**Acceptable quality level:** See **quality control procedures**.

**Acceptable risk:** The risk for which the benefits of a particular medical procedure are considered to outweigh the potential hazards. [*Acceptable Risk*, 1984, B. Fischhoff, Cambridge University Press, Cambridge.]

**Acceptance region:** A term associated with statistical significance tests, that gives the set of values of a *test statistic* for which the null hypothesis is not rejected. Suppose, for example, a *z-test* is being used to test the null hypothesis that the mean blood pressure of men and women is equal against the alternative hypothesis that the two means are not equal. If the chosen significance level of the test is 0.05 then the acceptance region consists of values of the test statistic  $z$  between  $-1.96$  and  $1.96$ . [*Encyclopedia of Statistical Sciences*, 2006, eds. S. Kotz, C. B. Read, N. Balakrishnan and B. Vidakovic, Wiley, New York.]

**Acceptance–rejection algorithm:** An algorithm for generating random numbers from some probability distribution,  $f(x)$ , by first generating a random number from some other distribution,  $g(x)$ , where  $f$  and  $g$  are related by

$$f(x) \leq kg(x) \text{ for all } x$$

with  $k$  a constant. The algorithm works as follows:

- let  $r$  be a random number from  $g(x)$ ;
- let  $s$  be a random number from a *uniform distribution* on the interval  $(0,1)$ ;
- calculate  $c = ksg(r)$ ;
- if  $c > f(r)$  reject  $r$  and return to the first step; if  $c \leq f(r)$  accept  $r$  as a random number from  $f$ . [*Statistics in Civil Engineering*, 1997, A. V. Metcalfe, Edward Arnold, London.]

**Acceptance sampling:** A type of *quality control procedure* in which a sample is taken from a collection or batch of items, and the decision to accept the batch as satisfactory, or reject them as unsatisfactory, is based on the proportion of defective items in the sample. [*Quality Control and Industrial Statistics*, 4th edition, 1974, A. J. Duncan, R. D. Irwin, Homewood, Illinois.]

**Accident proneness:** A personal psychological factor that affects an individual's probability of suffering an accident. The concept has been studied statistically under a number of different assumptions for accidents:

- pure chance, leading to the *Poisson distribution*;
- true contagion, i.e. the hypothesis that all individuals initially have the same probability of having an accident, but that this probability changes each time an accident happens;
- apparent contagion, i.e. the hypothesis that individuals have constant but unequal probabilities of having an accident.

The study of accident proneness has been valuable in the development of particular statistical methodologies, although in the past two decades the concept has, in general, been out of favour; attention now appears to have moved more towards risk evaluation and analysis. [*Accident Proneness*, 1971, L. Shaw and H. S. Sichel, Pergamon Press, Oxford.]

**Accidentally empty cells:** Synonym for **sampling zeros**.

**Accrual rate:** The rate at which eligible patients are entered into a *clinical trial*, measured as persons per unit of time. Often disappointingly low for reasons that may be both physician and patient related. [*Journal of Clinical Oncology*, 2001, **19**, 3554–61.]

**Accuracy:** The degree of conformity to some recognized standard value. See also **bias**.

**ACE:** Abbreviation for **alternating conditional expectation**.

**ACE model:** A biometrical genetic model that postulates additive genetic factors, common environmental factors, and specific environmental factors in a phenotype. The model is used to quantify the contributions of genetic and environmental influences to variation. [*Encyclopedia of Behavioral Statistics, Volume 1*, 2005, eds. B. S. Everitt and D. C. Howell, Wiley, Chichester.]

**ACES:** Abbreviation for **active control equivalence studies**.

**ACF:** Abbreviation for **autocorrelation function**.

**ACORN:** An acronym for 'A Classification of Residential Neighbourhoods'. It is a system for classifying households according to the demographic, employment and housing characteristics of their immediate neighbourhood. Derived by applying *cluster analysis* to 40 variables describing each neighbourhood including age, class, tenure, dwelling type and car ownership. [*Statistics in Society*, 1999, eds. D. Dorling and S. Simpson, Arnold, London.]

**Acquiescence bias:** The bias produced by respondents in a survey who have the tendency to give positive responses, such as ‘true’, ‘like’, ‘often’ or ‘yes’ to a question. At its most extreme, the person responds in this way irrespective of the content of the item. Thus a person may respond ‘true’ to two items like ‘I always take my medication on time’ and ‘I often forget to take my pills’. See also **end-aversion bias**. [*Journal of Intellectual Disability Research*, 1995, **39**, 331–40.]

**Action lines:** See **quality control procedures**.

**Active control equivalence studies (ACES):** *Clinical trials* in which the object is simply to show that the new treatment is at least as good as the existing treatment. Such studies are becoming more widespread due to current therapies that reflect previous successes in the development of new treatments. The studies rely on an implicit historical control assumption, since to conclude that a new drug is efficacious on the basis of this type of study requires a fundamental assumption that the active control drug would have performed better than a placebo, had a placebo been used in the trial. [*Statistical Issues in Drug Development*, 2nd edition, 2008, S. Senn, Wiley-Blackwell, Chichester.]

**Active control trials:** *Clinical trials* in which the trial drug is compared with some other active compound rather than a placebo. [*Annals of Internal Medicine*, 2000, **135**, 62–4.]

**Active life expectancy (ALE):** Defined for a given age as the expected remaining years free of disability. A useful index of public health and quality of life for populations. A question of great interest is whether recent trends towards longer *life expectancy* have been accompanied by a comparable increase in ALE. [*New England Journal of Medicine*, 1983, **309**, 1218–24.]

**Actuarial estimator:** An estimator of the *survival function*,  $S(t)$ , often used when the data are in grouped form. Given explicitly by

$$S(t) = \prod_{\substack{j \geq 0 \\ t_{(j+1)} \leq t}} \left[ 1 - \frac{d_j}{N_j - \frac{1}{2}w_j} \right]$$

where the ordered *survival times* are  $0 < t_{(1)} < \dots < t_{(n)}$ ,  $N_i$  is the number of people at risk at the start of the interval  $t_{(i)}$ ,  $t_{(i+1)}$ ,  $d_i$  is the observed number of deaths in the interval and  $w_i$  the number of censored observations in the interval. [*Survival Models and Data Analysis*, 1999, R. G. Elandt–Johnson and N. L. Johnson, Wiley, New York.]

**Actuarial statistics:** The statistics used by actuaries to evaluate risks, calculate liabilities and plan the financial course of insurance, pensions, etc. An example is *life expectancy* for people of various ages, occupations, etc. See also **life table**. [*Financial and Actuarial Statistics: An Introduction*, 2003, D. S. Borowiak and A. F. Shapiro, CRC Press, Boca Raton.]

**Adaptive cluster sampling:** A procedure in which an initial set of subjects is selected by some sampling procedure and, whenever the variable of interest of a selected subject satisfies a given criterion, additional subjects in the neighbourhood of that subject are added to the sample. [*Biometrika*, 1996, **84**, 209–19.]

**Adaptive designs:** *Clinical trials* that are modified in some way as the data are collected within the trial. For example, the allocation of treatment may be altered as a function of the response to protect patients from ineffective or toxic doses. [*Controlled Clinical Trials*, 1999, **20**, 172–86.]

**Adaptive estimator:** See **adaptive methods**.

**Adaptive lasso:** See **lasso**.

**Adaptive methods:** Procedures that use various aspects of the sample data to select the most appropriate type of statistical method for analysis. An *adaptive estimator*,  $T$ , for the centre of a distribution, for example, might be

$$\begin{aligned} T &= \text{mid-range when } k \leq 2 \\ &= \text{arithmetic mean when } 2 < k < 5 \\ &= \text{median when } k \geq 5 \end{aligned}$$

where  $k$  is the sample *kurtosis*. So if the sample looks as if it arises from a short-tailed distribution, the average of the largest and smallest observations is used; if it looks like a long-tailed situation the median is used, otherwise the mean of the sample is calculated. [*Journal of the American Statistical Association*, 1967, **62**, 1179–86.]

**Adaptive methods of treatment assignment:** Any method of treatment allocation in a *clinical trial* that uses accumulating outcome data to affect the treatment selection, for example, the *O'Brien-Fleming method*. [*Biometrika*, 1977, **64**, 191–199.]

**Adaptive sampling design:** A *sampling design* in which the procedure for selecting *sampling units* on which to make observations may depend on observed values of the variable of interest. In a survey for estimating the abundance of a natural resource, for example, additional sites (the sampling units in this case) in the vicinity of high observed abundance may be added to the sample during the survey. The main aim in such a design is to achieve gains in precision or efficiency compared to conventional designs of equivalent sample size by taking advantage of observed characteristics of the population. For this type of sampling design the probability of a given sample of units is conditioned on the set of values of the variable of interest in the population. [*Adaptive Sampling*, 1996, S.K. Thompson and G. A. F. Seber, Wiley, New York.]

**Added variable plot:** A graphical procedure used in all types of regression analysis for identifying whether or not a particular explanatory variable should be included in a model, in the presence of other explanatory variables. The variable that is the candidate for inclusion in the model may be new or it may simply be a higher power of one currently included. If the candidate variable is denoted  $x_i$ , then the *residuals* from the regression of the response variable on all the explanatory variables, save  $x_i$ , are plotted against the residuals from the regression of  $x_i$  on the remaining explanatory variables. A strong linear relationship in the plot indicates the need for  $x_i$  in the regression equation (Fig. 1). [*Regression Analysis*, Volume 2, 1993, edited by M. S. Lewis-Beck, Sage Publications, London.]

**Addition rule for probabilities:** For two events,  $A$  and  $B$  that are *mutually exclusive*, the probability of either event occurring is the sum of the individual probabilities, i.e.

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$

where  $\Pr(A)$  denotes the probability of event  $A$  etc. For  $k$  mutually exclusive events  $A_1, A_2, \dots, A_k$ , the more general rule is

$$\Pr(A_1 \text{ or } A_2 \dots \text{ or } A_k) = \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_k)$$

See also **multiplication rule for probabilities** and **Boole's inequality**. [KA1 Chapter 8.]

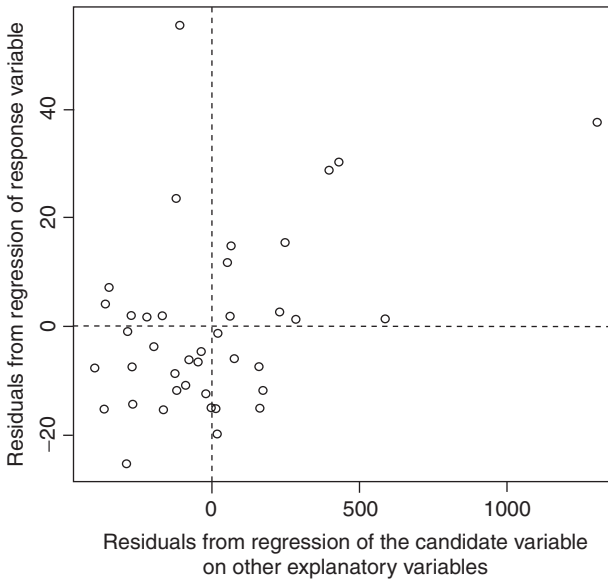


Fig. 1 Added variable plot indicating a variable that should be included in the model.

**Additive clustering model:** A model for *cluster analysis* which attempts to find the structure of a *similarity matrix* with elements  $s_{ij}$  by fitting a model of the form

$$s_{ij} = \sum_{k=1}^K w_k p_{ik} p_{jk} + \epsilon_{ij}$$

where  $K$  is the number of clusters and  $w_k$  is a weight representing the salience of the property corresponding to cluster  $k$ . If object  $i$  has the property of cluster  $k$ , then  $p_{ik} = 1$ , otherwise it is zero. [*Psychological Review*, 1979, **86**, 87–123.]

**Additive effect:** A term used when the effect of administering two treatments together is the sum of their separate effects. See also **additive model**. [*Journal of Bone Mineral Research*, 1995, **10**, 1303–11.]

**Additive genetic variance:** The variance of a trait due to the main effects of *genes*. Usually obtained by a factorial *analysis of variance* of trait values on the genes present at one or more loci. [*Statistics in Human Genetics*, 1998, P. Sham, Arnold, London.]

**Additive model:** A model in which the explanatory variables have an *additive effect* on the response variable. So, for example, if variable A has an effect of size  $a$  on some response measure and variable B one of size  $b$  on the same response, then in an assumed additive model for A and B their combined effect would be  $a+b$ .

**Additive outlier:** A term applied to an observation in a *time series* which is affected by a non-repetitive intervention such as a strike, a war, etc. Only the level of the particular observation is considered affected. In contrast an *innovational outlier* is one which corresponds to an extraordinary shock at some time point  $T$  which also influences subsequent observations in the series. [*Journal of the American Statistical Association*, 1996, **91**, 123–31.]

**Additive tree:** A connected, *undirected graph* where every pair of nodes is connected by a unique path and where the distances between the nodes are such that

## a. Dissimilarities

	A	B	C	D	E
Worker A	--				
Worker B	15	--			
Worker C	20	25	--		
Worker D	18	23	6	--	
Worker E	20	25	20	18	--

## b. Additive Tree

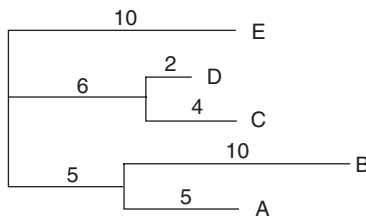


Fig. 2 An example of an additive tree.  
 (Reproduced by permission of Sage Publications  
 from *Tree Models of Similarity and Association*,  
 1996, J. E. Corter.)

$$d_{xy} + d_{uv} \leq \max[d_{xu} + d_{yv}, d_{xv} + d_{yu}] \text{ for all } x, y, u, \text{ and } v$$

An example of such a tree is shown in Fig. 2. See also **ultrametric tree**. [*Tree Models of Similarity and Association*, 1996, J. E. Corter, Sage University Papers 112, Sage Publications, Thousand Oaks.]

**Adelstein, Abe (1916–1993):** Born in South Africa, Adelstein studied medicine at the University of the Witwatersrand. In the 1960s he emigrated to Manchester where he worked in the Department of Social Medicine. Later he was appointed Chief Medical Statistician for England and Wales. Adelstein made significant contributions to the classification of mental illness and to the epidemiology of suicide and alcoholism.

**Adequate subset:** A term used in regression analysis for a subset of the explanatory variables that is thought to contain as much information about the response variable as the complete set. See also **selection methods in regression**.

**Adjacency matrix:** A matrix with elements,  $x_{ij}$ , used to indicate the connections in a *directed graph*. If node  $i$  relates to node  $j$ ,  $x_{ij} = 1$ , otherwise  $x_{ij} = 0$ . For a simple graph with no self-loops, the adjacency matrix must have zeros on the diagonal. For an undirected graph the adjacency matrix is symmetric. [*Introductory Graph Theory*, 1985, G. Chartrand, Dover, New York.]

**Adjusted correlation matrix:** A *correlation matrix* in which the diagonal elements are replaced by *communalities*. The basis of *principal factor analysis*.

**Adjusted treatment means:** Usually used for estimates of the treatment means in an *analysis of covariance*, after adjusting all treatments to the same mean level for the covariate(s), using the estimated relationship between the covariate(s) and the response variable. [*Biostatistics: A Methodology for the Health Sciences*, 2nd edn, 2004, G. Van Belle, L. D. Fisher, P. J. Heagerty and T. S. Lumley, Wiley, New York.]

**Adjusting for baseline:** The process of allowing for the effect of *baseline characteristics* on the response variable usually in the context of a *longitudinal study*. See also **Lord's paradox**

and **baseline balance**. [*Statistical Issues in Drug Development*, 2nd edition, 2008, S. Senn, Wiley-Blackwell, Chichester.]

**Administrative databases:** *Databases* storing information routinely collected for purposes of managing a health-care system. Used by hospitals and insurers to examine admissions, procedures and lengths of stay. [*Healthcare Management Forum*, 1995, **8**, 5–13.]

**Admissibility:** A very general concept that is applicable to any procedure of statistical inference. The underlying notion is that a procedure is admissible if and only if there does not exist within that class of procedures another one which performs uniformly at least as well as the procedure in question and performs better than it in at least one case. Here ‘uniformly’ means for all values of the parameters that determine the probability distribution of the random variables under investigation. [KA2 Chapter 31.]

**Admixture in human populations:** The inter-breeding between two or more populations that were previously isolated from each other for geographical or cultural reasons. Population admixture can be a source of spurious associations between diseases and *alleles* that are both more common in one ancestral population than the others. However, populations that have been admixed for several generations may be useful for mapping disease *genes*, because spurious associations tend to be dissipated more rapidly than true associations in successive generations of random mating. [*Statistics in Human Genetics*, 1998, P. Sham, Arnold, London.]

**Adoption studies:** Studies of the rearing of a nonbiological child in a family. Such studies have played an important role in the assessment of genetic variation in human and animal traits. [*Foundations of Behavior Genetics*, 1978, J.L. Fulker and W.R. Thompson, Mosby, St. Louis.]

**Adverse selection:** A term used in insurance when the insurer cannot distinguish between members of good- and poor-risk categories for a certain hazard and the poor-risks are the only purchasers of coverage with the consequence that the insurer expects to lose money on each policy sold. [*Quarterly Journal of Economics*, 1976, **90**, 629–650.]

**Aetiological fraction:** Synonym for **attributable risk**.

**Affine invariance:** A term applied to statistical procedures which give identical results after the data has been subjected to an *affine transformation*. An example is *Hotelling's  $T^2$  test*. [*Canadian Journal of Statistics*, 2003, **31**, 437–55.]

**Affine transformation:** The transformation,  $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$  where  $\mathbf{A}$  is a nonsingular matrix and  $\mathbf{b}$  is any vector of real numbers. Important in many areas of statistics particularly *multivariate analysis*.

**Age-dependent birth and death process:** A *birth and death process* where the birth rate and death rate are not constant over time, but change in a manner which is dependent on the age of the individual. [*Stochastic Modelling of Scientific Data*, 1995, P. Guttorp, Chapman and Hall/CRC Press, London.]

**Age heaping:** A term applied to the collection of data on ages when these are accurate only to the nearest year, half year or month. Occurs because many people (particularly older people) tend not to give their exact age in a survey. Instead they round their age up or down to the nearest number that ends in 0 or 5. See also **coarse data** and **Whipple index**. [*Population Studies*, 1991, **45**, 497–518.]



Cambridge University Press

978-0-521-76699-9 - The Cambridge Dictionary of Statistics, Fourth Edition

B. S. Everitt and A. Skrondal

Excerpt

[More information](#)

**Age–period–cohort model:** A model important in many *observational studies* when it is reasonable to suppose that age, number of years exposed to risk factor, and age when first exposed to risk factor, all contribute to disease risk. Unfortunately all three factors cannot be entered simultaneously into a model since this would result in *collinearity*, because ‘age first exposed to risk factor’+‘years exposed to risk factor’ is equal to ‘age’. Various methods have been suggested for disentangling the dependence of the factors, although most commonly one of the factors is simply not included in the modelling process. See also **Lexis diagram**. [*Statistics in Medicine*, 1984, **3**, 113–30.]

**Age-related reference ranges:** Ranges of values of a measurement that give the upper and lower limits of normality in a population according to a subject’s age. [*Archives of Disease in Childhood*, 2005, **90**, 1117–1121.]

**Age-specific death rates:** Death rates calculated within a number of relatively narrow age bands. For example, for 20–30 year olds,

$$DR_{20,30} = \frac{\text{number of deaths among 20 – 30 year olds in a year}}{\text{average population size in 20 – 30 year olds in the year}}$$

Calculating death rates in this way is usually necessary since such rates almost invariably differ widely with age, a variation not reflected in the *crude death rate*. See also **cause-specific death rates** and **standardized mortality ratio**. [*Biostatistics*, 2nd edition, 2004, G. Van Belle, L. D. Fisher, P. J. Heagerty and T. S. Lumley, Wiley, New York.]

**Age-specific failure rate:** A synonym for *hazard function* when the time scale is age. [*Statistical Methods for Survival Data Analysis*, 3rd edn, E. T. Lee and J. W. Wang, Wiley, New York.]

**Age-specific incidence rate:** *Incidence rates* calculated within a number of relatively narrow age bands. See also **age-specific death rates**. [*Cancer Epidemiology Biomarkers and Prevention*, 2004, **13**, 1128–1135.]

**Agglomerative hierarchical clustering methods:** Methods of *cluster analysis* that begin with each individual in a separate cluster and then, in a series of steps, combine individuals and later, clusters, into new, larger clusters until a final stage is reached where all individuals are members of a single group. At each stage the individuals or clusters that are ‘closest’, according to some particular definition of distance are joined. The whole process can be summarized by a *dendrogram*. Solutions corresponding to particular numbers of clusters are found by ‘cutting’ the dendrogram at the appropriate level. See also **average linkage**, **complete linkage**, **single linkage**, **Ward’s method**, **Mojena’s test**, **K-means cluster analysis** and **divisive methods**. [MV2 Chapter 10.]

**Agreement:** The extent to which different observers, raters or diagnostic tests agree on a binary classification. Measures of agreement such as the *kappa coefficient* quantify the relative frequency of the diagonal elements in a two-by-two contingency table, taking agreement due to chance into account. It is important to note that strong agreement requires strong *association* whereas strong association does not require strong agreement. [*Statistical Methods for Rates and Proportions*, 2nd edn, 2001, J. L. Fleiss, Wiley, New York.]

**Agresti’s  $\alpha$ :** A generalization of the *odds ratio* for  $2 \times 2$  contingency tables to larger contingency tables arising from data where there are different degrees of severity of a disease and differing amounts of exposure. [*Analysis of Ordinal Categorical Data*, 1984, A. Agresti, Wiley, New York.]

**Agronomy trials:** A general term for a variety of different types of agricultural field experiments including fertilizer studies, time, rate and density of planting, tillage studies, and pest and

weed control studies. Because the response to changes in the level of one factor is often conditioned by the levels of other factors it is almost essential that the treatments in such trials include combinations of multiple levels of two or more production factors. [*An Introduction to Statistical Science in Agriculture*, 4th edition, 1972, D.J. Finney, Blackwell, Oxford.]

**AI:** Abbreviation for **artificial intelligence**.

**AIC:** Abbreviation for **Akaike's information criterion**.

**Aickin's measure of agreement:** A chance-corrected measure of *agreement* which is similar to the *kappa coefficient* but based on a different definition of agreement by chance. [*Biometrics*, 1990, **46**, 293–302.]

**AID:** Abbreviation for **automatic interaction detector**.

**Aitchison distributions:** A broad class of distributions that includes the *Dirichlet distribution* and *logistic normal distributions* as special cases. [*Journal of the Royal Statistical Society, Series B*, 1985, **47**, 136–46.]

**Aitken, Alexander Craig (1895–1967):** Born in Dunedin, New Zealand, Aitken first studied classical languages at Otago University, but after service during the First World War he was given a scholarship to study mathematics in Edinburgh. After being awarded a D.Sc., Aitken became a member of the Mathematics Department in Edinburgh and in 1946 was given the Chair of Mathematics which he held until his retirement in 1965. The author of many papers on least squares and the fitting of polynomials, Aitken had a legendary ability at arithmetic and was reputed to be able to dictate rapidly the first 707 digits of  $\pi$ . He was a Fellow of the Royal Society and of the Royal Society of Literature. Aitken died on 3 November 1967 in Edinburgh.

**Ajne's test:** A *distribution free method* for testing the uniformity of a *circular distribution*. The test statistic  $A_n$  is defined as

$$A_n = \int_0^{2\pi} [N(\theta) - n/2]^2 d\theta$$

where  $(N\theta)$  is the number of sample observations that lie in the semicircle,  $\theta$  to  $\theta + \pi$ . Values close to zero lead to acceptance of the hypothesis of uniformity. [*Annals of Mathematical Statistics*, 1972, **43**, 468–479.]

**Akaike's information criterion (AIC):** An index used in a number of areas as an aid to choosing between competing models. It is defined as

$$-2L_m + 2m$$

where  $L_m$  is the maximized *log-likelihood* and  $m$  is the number of parameters in the model. The index takes into account both the statistical goodness of fit and the number of parameters that have to be estimated to achieve this particular degree of fit, by imposing a penalty for increasing the number of parameters. Lower values of the index indicate the preferred model, that is, the one with the fewest parameters that still provides an adequate fit to the data. See also **parsimony principle** and **Schwarz's criterion**. [MV2 Chapter 11.]

**ALE:** Abbreviation for **active life expectancy**.