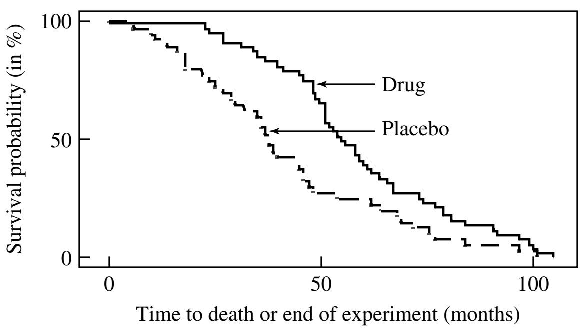
sah51693_K.qxd 8/31/01 10:26 AM Page 135

K

Kaplan–Meier estimator– A method for analyzing **survival data** containing **censored observations.** It uses exact survival times in the calculation of **probabilities** and provides an **estimate** of the **proportion** $S(t)$ of patients whose age at death would exceed t if no patients had been censored. The **estimator** consists of the product of a number of **conditional probabilities** resulting in an estimated **survival function** $\hat{S}(t)$ in the form of a step function. This is used to construct a **survival curve** in which the probability of survival remains **constant** between **events,** but drops at the time of occurrence of a new event. Censored observations are generally marked on the curve at the time of their occurrence. The Kaplan–Meier estimator is used to calculate an estimate of cumulative survival that can then be used to calculate the cumulative **hazard rate.** The Kaplan–Meier estimator differs from the method of **life table analysis** by grouping censored observations into intervals, in contrast to using exact end points in time when an event of interest has occurred. The procedure is also known as the product limit estimator. See also *survival analysis.*



Kaplan–Meier survival curves for a placebo and an active drug

kappa coefficient– Same as *kappa statistic.*

kappa statistic– A **statistic** used to measure agreement or **reliability** between two observers or raters for **nominal data.** It can also be used to assess the agreement between two alternative methods of diagnosis. It is defined as the agreement beyond **chance** divided by the amount of agreement possible beyond chance. Thus, the kappa statistic measures proportional agreement corrected for chance, that is, the **proportion** of agreements over and above what might be expected by chance alone. The formula for kappa statistic (κ) is

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where p_0 is the **probability** of observed occurrence and p_e is the probability of expected or chance agreement. It takes the value 1 when there is perfect agreement and 0 when observed agreement is equal to chance agreement. When the **data** involve **measurements** on **ordinal variables,** a modified procedure known as ordinal kappa statistic is employed.

Kendall's coefficient of concordance– A measure of agreement among two or more raters who rank a number of individuals according to certain criteria.

Kendall's rank correlation– Same as *Kendall's tau.*

Kendall's tau– A nonparametric **measure of association** between two **ordinal variables** proposed by M. G. Kendall in 1938. It is based on the number of inversions (interchanges of **ranks**) in one **ranking** compared with another. It is calculated as $P - Q$ where P is the number of concordant pairs, i.e., pairs with rankings in the same direction, and Q is the number of discordant pairs, i.e., pairs with rankings in the reverse direction. It is especially appropriate for small **sample sizes.** There are a number of modifications of τ introduced in the literature for measuring **associations** in a **contingency table** where both rows and columns represent natural ordered categories.

kernel density estimator– Same as *kernel estimator.*

kernel estimator– A **nonparametric method** for estimating the **density function** of a **probability distribution.** It is calculated from a **sample** of size n by replacing each **data value** by a "kernel" of area $1/n$ resulting in a curve similar to a smoothed **frequency polygon.**

Khinchin theorem– A theorem in mathematical statistics that states that the **sample mean** converges in **probability** to the **population mean** as the **sample size** tends to infinity.

Klotz test– A **nonparametric procedure** for testing the equality of **variances** of two **populations** having the same **median.** It is based on inverse **normal scores** and was developed by Jerome Klotz in 1962. If the populations are symmetrical, its **asymptotic relative efficiency** compared to the classical ***F* test** is one. In many cases its **efficiency** exceeds one. See also *Ansari–Bradley test, Barton–David test, Conover test, F test for two population variances, Mood test.*

Kolmogorov–Smirnov one-sample test– See *Kolmogorov-Smirnov tests.*

Kolmogorov–Smirnov tests– Nonparametric tests for testing significant differences between two **cumulative distribution functions.** The one sample test is used to test whether the **data** are consistent with a given **distribution function** and the two sample test is used to test the agreement between two observed **cumulative distributions.** The test is based on the maximum absolute difference between the two cumulative distribution functions. See also *goodness-of-fit test.*

Kolmogorov–Smirnov two-sample test– See *Kolmogorov–Smirnov tests.*

Kruskal–Wallis one-way analysis of variance by ranks– Same as *Kruskal-Wallis test.*

Kruskal–Wallis test– A **nonparametric procedure** used to compare three or more **independent samples** of **observations** that cannot be compared by means of an ***F* test for analysis of variance** either because the **data** are measured on **ordinal scale** or because the **normality** or **homogeneity of variance** assumptions cannot be satisfied. The method consists of **ranking** observations in all **samples** combined and the **test statistic** is based on the sum of the **ranks** assigned to the individual **treatment groups.** The test is a direct generalization of the **Wilcoxon rank-sum test** to three or more independent samples. When the **null hypothesis** is true, the test statistic can be approximated by a **chi-square distribution.** See also *Friedman's rank test.*

***k* statistics–** A set of symmetric functions calculated from the **sample data,** originally proposed by R. A. Fisher to determine the **moments** of **sample statistics.** The univariate *k* statistic of order *r* is defined as the **statistic** whose **mean** value is the *r*th **cumulant** of the **parent population.** The *k* statistics possess semi-invariant properties and their sampling cumulants can be determined directly from combinatorial methods.

kurtosis– The degree of "flatness" or "peakedness" of a univariate **frequency distribution.** A measure of kurtosis is obtained as the product moment ratio μ_4/μ_2^2, where μ_4 is the fourth **central moment** and μ_2 is the **variance.** For the **normal distribution,** it takes the value of 3. See also *coefficient of skewness, coefficient of kurtosis, leptokurtic, mesokurtic, platykurtic.*