Graphical Methods in Nonparametric Statistics: A Review and Annotated Bibliography
Author(s): Nicholas I. Fisher
Source: *International Statistical Review / Revue Internationale de Statistique*, Vol. 51, No. 1 (Apr. , 1983), pp. 25-58
Published by: International Statistical Institute (ISI)
Stable URL: http://www.jstor.org/stable/1402730
Accessed: 10-02-2016 20:39 UTC

International Statistical Review, **51** (1983), pp. 25–58. Longman Group Limited/Printed in Great Britain

Graphical Methods in Nonparametric Statistics: A Review and Annotated Bibliography

Nicholas I. Fisher

CSIRO Division of Mathematics & Statistics, Lindfield, N.S.W., Australia 2071

Summary

This paper reviews the range of graphical methods available for use with nonparametric procedures, and provides examples of the use of many of the methods under the broad groupings: two-sample procedures, one-sample procedures, association and regression procedures, and miscellaneous procedures. An annotated bibliography is also provided.

Key words: Ansari–Bradley test; Association; Butler test; Centre of symmetry; Hazard ratio; Hodges–Lehmann estimate; Kendall's tau; Kolmogorov–Smirnov test; Linked vector plot; Location difference; Mann–Whitney–Wilcoxon test; Olmstead–Tukey test; P plots; P–P plots; Pair chart; Parameter space plot; Plotting methods; Proportional hazards; Q plots; Q–Q plots; Rank correlation; Runs test; Scale difference; Shift function; Simple linear regression; Spearman's rho; Survival data; Symmetry function; Tukey quick test; Wilcoxon rank-sum test.

1 Introduction

Graphical methods for display of data and as computational devices have a long history (Fienberg 1979); with the ready availability of high-speed computers these applications have been magnified manyfold, the greatest impact being in the analysis of large data sets and of multivariate data. The field of nonparametric statistics† has been a particularly fertile place for graphical methods to flourish. Many of the early methods were designed to simplify computation of such statistics as an estimate of centre of symmetry (the median of all the sample midranges) or the sample value of Kendall's tau, both of which are relatively tedious to calculate otherwise, even from small samples. More recently, powerful statistical techniques based on the sample distribution function utilise graphical displays both for informal and formal inference about underlying models, for example Wilk & Gnanadesikan (1968), Doksum (1977), to elucidate the nature of association between variables (Taguri *et al.*, 1976), and to highlight the structure of multivariate distributions (Tukey & Tukey, 1981).

In this paper, we survey the range of graphical methods used in nonparametric statistics, and illustrate many of them. The descriptions of the methods in the next section are categorized as follows: § 2.1 Two-sample procedures, (i) Location difference, (ii) Scale difference, (iii) Location and scale difference, other two-sample comparisons, general difference; § 2.2 One-sample symmetry procedures, (i) Centre of symmetry, (ii) General assessment of symmetry; § 2.3 Association and regression procedures, (i) Association, (ii)

† This will be taken to mean the study of those statistical situations in which no underlying parametric family of distributions is assumed; that is, the *family* of distributions under consideration cannot be indexed by a finite number of parameters (and hence is a *nonparametric* family).

Regression; § 2.4 Miscellaneous procedures, (i) k-sample procedures, (ii) Contingency tables, (iii) Analysis of covariance, (iv) Angular data, (v) Other multivariate procedures, (vi) Time series.

Examples are given of almost all the two-sample procedures (§ 2.1). Because of the close relationship between the two-sample homogeneity problem and one-sample symmetry problem (see introduction to § 2.2) several of the procedures in § 2.2(i) are simple adaptations of those in § 2.1(i) and so have not been illustrated. Most of the methods referred to in § 2.3 are illustrated; however the various methods referenced in § 2.4 are not, partly due to space considerations, but also because some idea of how a particular method works can be gleaned from related material illustrated in other sections, § 2.4(i), (iii), (vi), and because one example would be hopelessly inadequate as an advertisement, § 2.4(v).

This last point brings us to the question of data sets. The decision of whether to create an artificial data set which attempts to demonstrate all possible uses of a given technique or whether to use an available data set (experimental or observational) which probably will not display the technique to best advantage is not easy to make. In this paper, the latter course has been adopted in all cases but one, because there is intrinsic interest in the data. Thus, many of the examples serve only to highlight a few aspects of the methods, and the reference sources in the Bibliography should be investigated to form a proper appreciation of them. (The only occasion on which artificial data have been used is during the discussion of some methods associated with simple linear regression, where a very small sample size (5) has been used to simplify the diagram.)

Following the survey of graphical methods is a section containing some general remarks, particularly about the broad applicability of the methods, and then a Bibliography, in which each reference has been annotated briefly. J.W. Tukey's book *Exploratory Data Analysis* is not included here, nor are any of its tabular/graphical methods discussed in the survey: it did not seem sensible to remove them from their natural habitat and display them in isolation, because of the unity of development of the book.

2 Survey of graphical procedures

2.1 *Two-sample procedures*

The class of procedures to be discussed can be divided conveniently into two groups, those designed for a particular estimate or test (e.g. the Mann–Whitney–Wilcoxon estimate of location shift), and those which are of more general applicability (P–P, Q–Q and H–H plots, the pair chart, and confidence procedure based on the shift function). All discussion of the latter group will be given in § 2.1(iii), although they are relevant to § 2.1(i), (ii).

The two random samples will be denoted by $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, drawn from underlying populations with distribution functions F and G respectively; the sample distribution functions are denoted by F_m and G_n respectively. The X-order statistics will be written as $X_{(1)}, \ldots, X_{(m)}$ (with $X_{(1)} < X_{(2)} < \ldots$) and similarly for the Y's.

(i) *Location difference.* Suppose it is known that $G(x) = F(x+\theta)$ for some unknown θ. The Hodges–Lehmann estimate of θ based on the Mann–Whitney–Wilcoxon test is simply $\hat{\theta}_w = \text{med}\{X_i - Y_j, 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$. A variety of graphical methods have been proposed to compute $\hat{\theta}_w$ and an associated confidence interval for θ. To illustrate them, we consider estimating the difference in bowling skills of two competitors in the 1935 Willimantic Duckpin Bowling Sweepstakes, based on the data in Table 1.

For the first method, construct the set of points $\{(X_i, Y_j), 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$ as the intersections of two families of parallel lines, one family parallel to the x axis and the

Table 1

Scores on various string numbers of two entrants in the Willimantic Duckpin Bowling Stakes; *Willimantic, Connecticut, February* 16, 1935

	String number							
	4	5	6	7	8	9	10	11
X: Entrant 8	107	135	116	120	136	117	102	131
Y: Entrant 12	131	121	118	110	130	108	114	101

Source: Hartford Courant, February 17, 1935. As reported by Waugh, A.E. (1944), *Laboratory Manual and Problems for Elements of Statistical Method*, New York: McGraw-Hill.

other to the y axis, as shown in Fig. 1a. (The data have been reduced modulo 100 for convenience.) Consider the line $x-y=c$, which makes a 45° angle with the x axis, and choose initially $c>36$. As c is decreased, the line passes sequentially through the points (X_i, Y_j) in an order corresponding to decreasing values of the elementary estimates $X_i - Y_j$. In this way, the median value(s) of $\{X_i - Y_j\}$ is (are) readily obtained; for mn even, as in this case, the 32nd and 33rd of the ordered differences are averaged to yield $\hat{\theta}_w = 4.5$. For a 95% confidence interval for θ, obtain the critical values 14 and 51 from tables of the Mann–Whitney–Wilcoxon test (see for example Noether, 1971) and count in from each end, as above. (If mn is large, a reasonable estimate of θ is the midpoint of the confidence interval; see for example Hollander & Wolfe 1973.)

This method was first proposed by Moses (1953) and subsequently reported by Moses (1965), and is discussed in many places: Conover (1971) (which includes some discussion of the continuity assumption in relation to estimation), Daniel (1978), Gibbons (1971), Hollander & Wolfe (1973), and Noether (1971) (which includes a discussion on ties). Jaeckel (1969) mentions generalizations of the estimator median $\{X_i - Y_j\}$, in which the differences $X_i - Y_j$ are differentially weighted: the graphical computation of such an estimator from Fig. 1a then involves cumulating the weights assigned to the intersections until half the total assignable weight has been obtained.

The second method for calculating $\hat{\theta}_w$ involves setting out the ordered samples as shown in Fig. 1b, and computing the leading diagonal of differences (entrant 1–entrant 2). Notice that, as one proceeds across a row to the right, or up a column, the differences cannot decrease (because of the ordered marginal numbers).

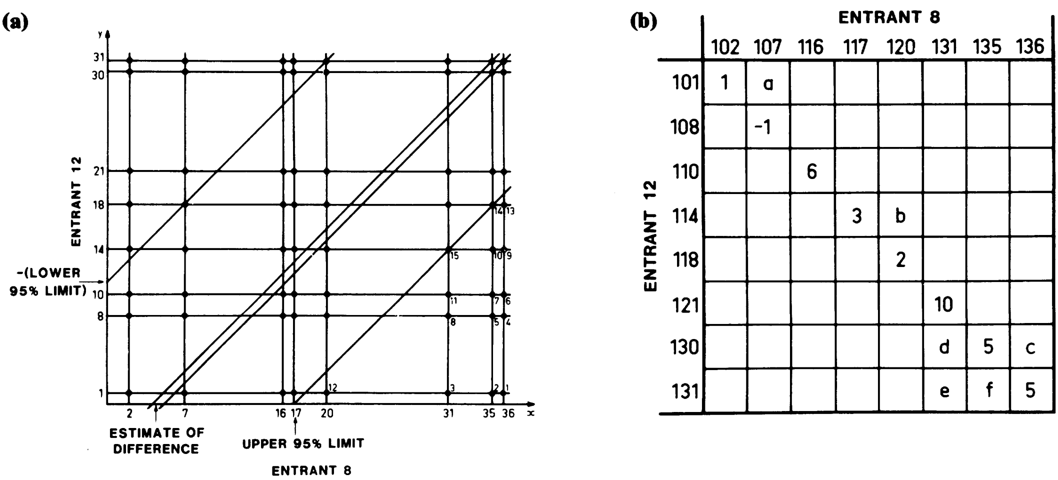

Figure 1. *Determination of the Mann–Whitney–Wilcoxon estimate of location difference.* (a) *Moses' method* (b) *Høyland's method. Data*: *scores by two entrants in Duckpin Bowling* (*Table* 1).

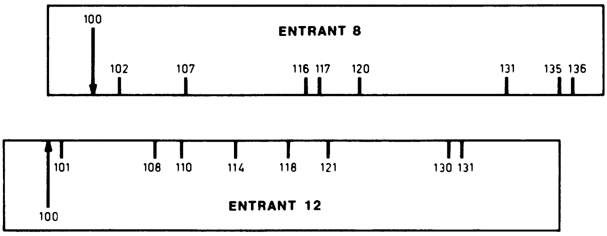

Figure 2. *Sliding papers method for determining Hodges–Lehmann estimate of location shift, and for performing Tukey's quick test. Data: scores by two entrants in Duckpin Bowling* (*Table* 1).

Thus the value 3 in cell (4, 4) is the smallest of the 20 values in the 4×5 rectangle of which it is the bottom left-hand cell, and necessarily less than the values in the four other rectangles whose bottom left-hand cells are (3, 3), (6, 6), (7, 7) and (8, 8). The total number of distinct cells in these rectangles is 32. Hence $3\leqslant 32$nd largest value; similarly, $3\geqslant 29$th smallest value. By similar argument, $6\leqslant 27$th largest value, $6\geqslant 35$th smallest value, and $5\leqslant 29$th largest value, $5\geqslant 32$nd smallest value. We seek the 32nd and 33rd ordered values, which must lie in the range (3, 5), using the above inequalities and the fact that two 5's are already displayed. The only cells with possible values at least three but less than 5 are those labelled a, b, c, d, e and f; these values are 6, 6, 1, 4, 0 and 4 respectively. It is then easy to calculate that 4 is the 32nd ordered value, and hence that 5 is the 33rd. Similar calculations can be used to find the differences corresponding to the 95% confidence limits for θ.

This tabular method was first published by Hoyland (1964), and is also described by Lehmann (1975); after a little practice, it is simple and reasonably efficient to use.

To implement the third method of calculating $\hat{\theta}_w$, mark the values of each sample along separate slips of paper as shown in Fig. 2, together with a common reference mark (100 in Fig. 2). Starting with the X slip completely to the right of the Y slip, move the Y slip gradually to the right, and add one (1) to a mental counter (initiallized at 0) each time a Y value moves past an X value. When the counter reaches 32 note the difference between the reference marks, and similarly when the counter reaches 33: the average of the differences is $\hat{\theta}_w$. Alternatively, obtain the differences from the particular (X_i, Y_j) pairs yielding the 32nd and 33rd counts. The differences required for a 95% confidence interval are similarly determined.

This method (perhaps as *mechanical* as it is graphical) was the one proposed by Hodges & Lehmann (1963) in which the family of Hodges–Lehmann estimators was introduced. It can also be used as a way of approximating the Hodges–Lehmann estimate based on the two-sample normal scores test.

Note that there is a fourth method of computing $\hat{\theta}_w$: in § 2.2 below, Tukey's method of computing the Wilcoxon estimate of the median of a symmetric distribution is described. It can easily be adapted to a method of computing $\hat{\theta}_w$.

Before discarding the slips of paper so carefully prepared above, one may with negligible effort perform Tukey's 'quick, compact, two-sample test to Duckworth's specification' (Tukey A1959‡). The test requires that the smallest and largest observations among $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ belong to different samples. This being the case, suppose that $Y_{(1)}$ is the smallest and $X_{(m)}$ the largest. Count $U=$ number of X's $> Y_{(n)}$, $L=$ number of Y's $< X_{(1)}$, set $T=U+L$, and reject the hypothesis $\theta=0$ at approximately the

‡ Dates prefixed with the letter A indicate references in the auxiliary reference list.

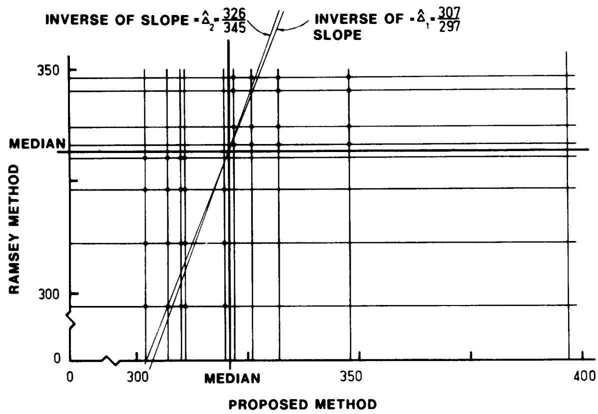

Figure 3. *Bhattacharyya's method for determining the Hodges–Lehmann estimate (based on the Ansari–Bradley test) of scale difference. Data: two methods of determining total serum iron bonding capacity (Table 2).*

5%/1%/0.1% level accordingly as $T \geqslant 7/10/13$. (From Fig. 3, $T = 2 + 1 = 3$.) This is the most compact version of Tukey's procedure, and it relies on the ratio of the sample sizes being no greater than 4:3, and on neither U nor L being zero. Much more discussion of all aspects of this procedure is given by Tukey (A1959), including point and confidence interval estimates for θ. The graphical device is due to Sandelius (1968), who discusses Hodges–Lehmann estimation of θ; the method is also described by Conover (1971). Adaptations of Tukey's test, for which Sandelius' device is also applicable, have been published by Rosenbaum (A1965) and Neave (A1966).

The interval estimates of θ based on the Mann–Whitney–Wilcoxon test also provides a test of hypothesis for θ. However the Mann–Whitney–Wilcoxon test statistic itself, and also Tukey's quick test statistic T may be computed directly (and very simply) using a pair chart; see Quade (1973) and discussion in § 2.1(iii) below. Klotz (1966) used the pair chart to enumerate the distribution of the Mann–Whitney–Wilcoxon test statistic in the presence of ties. Hettmansperger & McKean (1974) present as a teaching aid a graphical method for illustrating the relationship between a hypothesis test and a confidence interval for θ using the Mann–Whitney–Wilcoxon test. It involves plotting the differences $X_i - Y_j$ along the x axis and the distribution of the test statistic up the y axis, and then exhibiting the correspondence between tail probabilities and extreme values of $X_i - Y_j$.

(ii) *Scale difference.* Let (θ_1, σ_1) and (θ_2, σ_2) be pairs of location (median) and scale parameters for the populations underlying the two samples and suppose that

$$G(x) = F\left(\frac{\sigma_2}{\sigma_1}(x - \theta_1) + \theta_2\right) \equiv F((x - \theta)/\Delta)$$

say. The Hodges–Lehmann estimates of Δ (the ratio of scale parameters) based on the Ansari–Bradley test, the Siegel–Tukey test and Sukhatme's test, were studied by Bhattacharyya (1977), who presented a graphical method for their computation. As an example, we illustrate the method for the Ansari–Bradley estimate $\hat{\Delta}$ (see below) for the data on two methods of determining total serum iron bonding capacity (Table 2); the medians θ_1 and θ_2 are unknown.

Plot the data similarly to Fig. 1a, and insert the lines corresponding to the separate sample medians ($\tilde{X}$ and $\tilde{Y}$) as shown in Fig. 3. The (X_i, Y_j) intersections in the top-right-hand and bottom-left-hand corners determined by the medial lines are called

Table 2
Results of two methods of determining total serum iron-bonding capacity without deproteinization

X: Ramsey method ($m=8$)	297,	311,	323,	330,	333,	337,	345,	348		
Y: Proposed method ($n=10$)	302,	307,	310,	311,	320,	322,	326,	332,	348,	397

Source: Table 3, Jung, D. H. & Parekh, A.C. (1970), A semi-micromethod for the determination of serum iron and iron-binding capacity without deproteinization, *Am. J. Clinical Pathol.* **54**, 813–817.
The data are simple random samples drawn from the relevant part of Table 3.

relevant pairs: then $\hat{\Delta}=\text{med}\,\{(X_i-\tilde{X})/(Y_j-\tilde{Y}), \text{ all pairs } (X_i, Y_j)\}$. To compute $\hat{\Delta}$, locate the line which passes through $(\tilde{X}, \tilde{Y})$ such that half the relevant pairs lie between it and the vertical axis, and the other half between it and the horizontal axis. In this example the number of relevant pairs is even (40), so two such lines are determined, with inverse slopes $\hat{\Delta}_1$ and $\hat{\Delta}_2$ as shown. Then $\hat{\Delta}=\frac{1}{2}(\hat{\Delta}_1+\hat{\Delta}_2)$. There are slight modifications if θ_1 and θ_2 are known.

When the scale difference problem can be reduced to the location difference problem (e.g. when the medians are known and the variables necessarily positive) some of the procedures in § 2.1(i) can be applied, as discussed by Hollander & Wolfe (1973, p. 101) and Shorack (1966). Shorack also gives a graphical procedure (based on a method due to P.K. Sen) to obtain point and interval estimates of the ratio of scale parameters when this reduction is not possible. The usual setting is that of estimation of relative potency in which, for example, $X_1, \ldots, X_m$ are responses to doses of a test drug and $Y_1, \ldots, Y_n$ responses to doses of a standard. If we assume that $G(x)=F(\rho x)$ $(\rho>0)$, Shorack's modified version of Sen's estimator is obtained as follows. Set

$$\phi(m, n; d)=\sum_{i=1}^{m}\sum_{j=1}^{n} I[Y_j \leqslant dX_i],$$

where $I[A]$ is the indicator function of the set A; under the null hypothesis that $\rho=1$, $E[\phi]=\frac{1}{2}mn$. Then define

$$\hat{\rho}=\begin{cases}\inf\{d:\phi(m, n; d)=\frac{1}{2}(mn+1)\} & mn \text{ odd},\\ \text{geometric mean }\{d:\phi(m, n:t)=\frac{1}{2}mn\} & mn \text{ even},\end{cases}$$

where geometric mean $(a, b)=\surd(ab)$ for an interval. As a test statistic, the distribution of ϕ is the same as that of the Mann–Whitney–Wilcoxon statistic, hence confidence limits $(\hat{\rho}_L, \hat{\rho}_U)$ for ρ can also be calculated. The graphical computation of $\hat{\rho}, \hat{\rho}_L$ and $\hat{\rho}_U$ is exemplified in Fig. 4, using a well-known data set published by D.J. Finney on the tolerances of cats to Strophanthus A and Strophanthus B. (The data are reproduced here as Table 3, and were also used by Shorack.) Construct the set of intersections (X_i, Y_j) as in

Table 3
Fatal doses of two tinctures of Strophanthus applied to two samples of cats

X: Strophanthus A	1.24	1.44	1.55	1.58	1.71	1.89	2.34
Y: Strophanthus B	1.20	1.47	1.85	2.00	2.20	2.27	2.42

Source: Table 2.1. Finney, D.J. (1964), *Statistical Method in Biological Assay*, 2nd edition, London: Griffin.

Fig. 4

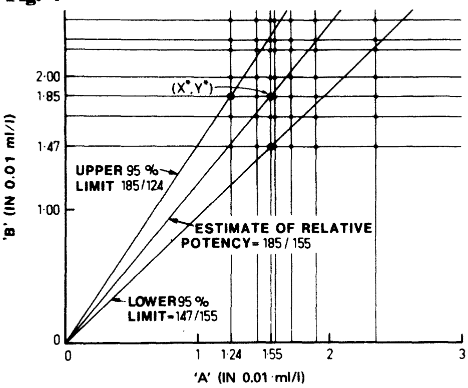

Fig. 5

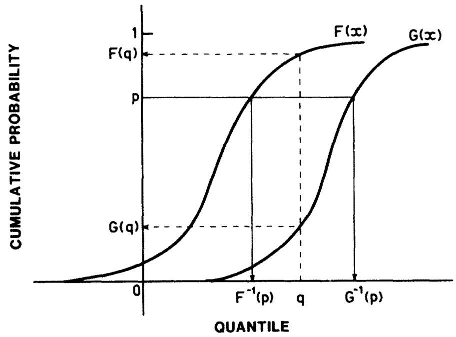

Figure 4. *Shorack's method of determining Sen's estimate of relative potency. Data: two drugs administered to two samples of cats: Strophanthus* A *and Strophanthus* B (*Table* 3).
Figure 5. *Basis of* P–P *and* Q–Q *plots. For two distribution functions* F *and* G, *if* $F(x) \equiv G(x)$ *then* (i) *corresponding to each quantile* q, $F(q)$ *and* $G(q)$ *should be equal, and* (ii) *corresponding to each cumulative probability* p, $F^{-1}(p)$ *and* $G^{-1}(p)$ *should be equal.*

previous examples and find the line through the origin which passes through an intersection (X^*, Y^*) say and which divides the mn intersections into equal groups above and below the line. The slope of the line is $\hat{\rho}$ and is conveniently computed as Y^*/X^*; confidence limits are computed similarly. Had mn been even, the two central lines would have yielded estimates $\hat{\rho}_1$ and $\hat{\rho}_2$, and the overall estimate $\hat{\rho} = \sqrt{(\hat{\rho}_1\hat{\rho}_2)}$.

This procedure is also discussed by Hollander & Wolfe; Shorack also describes a modified estimator for situations in which the data (X_i, Y_i) form matched pairs and hence have a joint distribution $H(x, y) \not\equiv F(x)G(y)$. Several two-sample statistics to detect dispersion difference can be calculated from the pair chart (Quade, 1973) described in § 2.1(iii) below.

(iii) *Location and scale differences, other two-sample comparisons, general difference.* Probably the most powerful and useful graphical methods in nonparametric statistics are those based on comparison of the sample distribution functions. Two methods of comparing sample distribution functions are easily understood by studying Fig. 5 (which is based on a similar figure of Wilk & Gnanadesikan, 1968). If $G(x) \equiv F(x)$, then for any given cumulative probability p, the quantiles $F^{-1}(p)$ and $G^{-1}(p)$ coincide. A comparative plot of sample quantiles $(F_m^{-1}(p), G_n^{-1}(p))$ corresponding to a set of values $p \in [0, 1]$ is termed a Q–Q plot by Wilk & Gnanadesikan. Similarly, for any given quantile q, the cumulative probabilities $F(q)$ and $G(q)$ coincide if F and G are identical. A comparative plot of sample cumulative probabilities $(F_m(q), G_n(q))$ is termed a P–P plot.

The first use of Q–Q plots for comparing two independent samples appears to be in the paper by Lorenz (1905); a detailed analysis of their properties is given by Wilk & Gnanadesikan (1968), Gerson (1975) and Gnanadesikan (1977). Briefly, a roughly linear plot suggests that the underlying random variables have the same distribution, apart from possible location or scale differences: such differences may be estimated using the intercept and slope of the 'line of best fit'. Typically, Q–Q plots are rather more sensitive to differences between F and G in the tails of the distributions than in the centres, for random variables with infinite ranges. This is because the quantile is a rapidly changing function of p where the density is sparse (Wilk & Gnanadesikan, 1968, p. 5; Gerson, 1975).

In practice, to perform a Q–Q plot we select a set of probability levels $0 < p_1 < \ldots < p_k < 1$ and identify the corresponding order statistics, i.e. quantiles, $(X_{[mp_i+1]}, Y_{[np_i+1]})$ for

Fig. 6

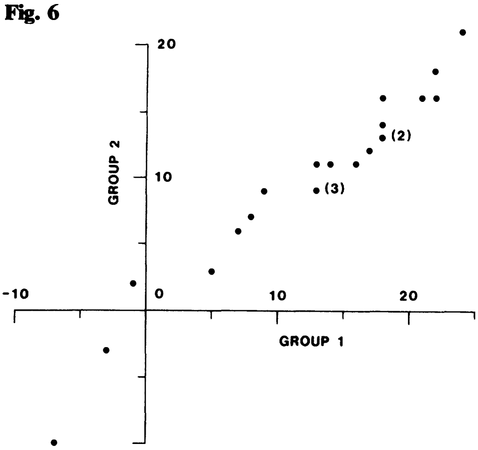

Fig. 7

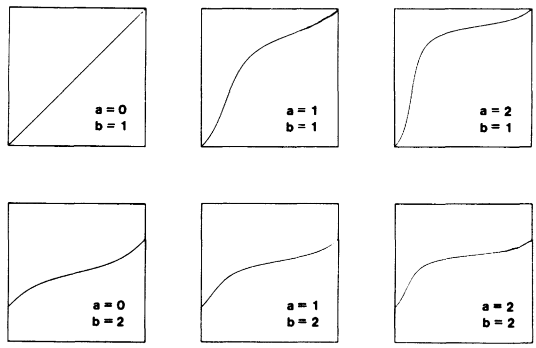

Figure 6. *Quantile–quantile plot. Data: scores by two sets of patients on psychological tests (Table 4).*
Figure 7. *P–P plot, or pair chart, of Cauchy (a, b) versus Cauchy $(0, 1)$, where Cauchy (a, b) has density proportional to $\{1+(x-a)^2/b^2\}^{-1}$.*

$1 \leqslant i \leqslant k$. The situation is simplified if $m=n$, for one then plots (a subset of) the pairs $(X_{(i)}, Y_{(i)})$ for $1 \leqslant i \leqslant n$. Generally speaking, probability plots (P–P or Q–Q) can be influenced considerably by the vagaries of random sampling for sample sizes much less than 30, so any inferences based on plots with substantially fewer data would be very tenuous. To a degree, then, this applies to the next example.

Figure 6 illustrates the use of a Q–Q plot for the data in Table 4, consisting of scores on a psychological test administered to a sample of people with uncontrollable cancer and to another sample with controllable cancer. There is no strong suggestion of departure of the plot from linearity, although a 'best-fitting' straight line through the origin does not quite have slope 1 (compare with the P–P plot of the same data, discussed below).

Plots based on cumulative probabilities have been in use at least as far back as Hazen (A1914) for purposes of assessing goodness of fit (see also Hazen, A1930). Again, detailed discussion of P–P plots is given by Wilk & Gnanadesikan and by Gerson. By way of contrast with Q–Q plots, any departure from the hypothesis $F=G$ will be manifested in the P–P plot as a nonlinear appearance of the plot. Further, because of the constraints that the plot begins at $(0, 0)$ and ends at $(1, 1)$, it is much more sensitive to differences between F and G in their centres rather than in their tails. The way in which certain sorts of differences (i.e. location/scale differences) can manifest themselves in a P–P plot is illustrated in Fig. 7, which is based on similar plots of Quade (1973). Figure 7 shows the

Table 4

FK-scores (with signs changed) on psychological tests administered to two groups of cancer patients

X: Group I	−7	−7	−3	−1	5	7	8	9	13	13	13	13	14
($m=25$)	14	16	17	18	18	18	18	21	22	22	24	25	
Y: Group II	−10	−3	2	3	6	7	9	9	9	9	11	11	11
($n=22$)	12	13	13	14	16	16	16	18	21				

Group I have rapidly progressing (uncontrollable) disease and group II slowly progressing (controllable) disease. High FK-values indicate high defensiveness or tendency to present the appearance of serenity while suffering deep inner distress.

Source: Tables 2 and 3, Blumberg, E.M., West, P.M. & Ellis, P.W. (1954), A possible relationship between psychological factors and human cancer, *Psychosomatic Med.* **16,** 277–286.

asymptotic P–P plot $\{[F(x), G(x)], -\infty < x < \infty\}$ for a standardized Cauchy variable, i.e. density proportional to $(1+x^2)^{-1}$, compared with a variety of unstandardized Cauchy variables, i.e. densities of form proportional to $\{1+(x-a)^2/b^2\}^{-1}$; specifically, a plot of

$$\left\{\left[\frac{1}{\pi}\tan^{-1} x, \frac{1}{\pi}\tan^{-1}((x-a)/b)\right], -\infty < x < \infty\right\}.$$

When the distributions are identical ($a=0$, $b=1$), a straight line results. As one distribution is shifted away from the other ($a=1, b=1$; $a=2, b=1$), the plot takes on a bell-shaped apearance. With a pure change of scale ($a=0, b=2$) the effect on the plot is that it starts and finishes with a jump, while still exhibiting a form of skew-symmetry. If both location and scale differences are present ($a=1, b=2$; $a=2, b=2$) this skew-symmetry is lost.

A P–P plot is created by ordering the combined samples and then, starting at (0, 0), plotting a point $1/m$ to the right or $1/n$ up according as the first member of the sequence is an X or a Y, and continuing in this fashion through the remaining $m+n-1$ members of the sequence. Such a plot for the psychological data in Table 4 is shown in Fig. 8. Note that the large gaps on the right-hand side of the plot are due to between-sample ties: if the plot is positioned at (x_i, y_i) after plotting i points, and m_0 X's are tied with n_0 Y's at this stage of the sequence, the next point is plotted as $(x_i+m_0/m,\ y_i+n_0/n)$. Possible differences between F and G in the centres, which were not well-highlighted in the Q–Q plot, are shown up rather strikingly here (again, with the slightly undersized samples discounting the strength of the inference).

In practice, it is advisable to adjust the two data sets for location and scale differences (e.g. by subtracting the sample median and dividing by the interquartile range, for each sample) before performing the P–P plot, since location/scale differences are manifested satisfactorily in Q–Q plots, and may conceal other effects in P–P plots. For the example used in Figs. 6 and 8, however, the Q–Q plot reveals very little location or scale difference. As a result, the P–P plot of the adjusted samples is essentially the same as that for the unadjusted samples, suggesting that there is some other qualitative difference between (the middles of) the two populations.

Friedman & Rafsky (1979, A1979, 1981) have used minimal spanning trees to produce P–P plots to compare two multivariate samples; they also show that multivariate generalizations of Q–Q plots along these lines are impracticable. Wilk & Gnanadesikan discuss 'hybrid' (P–Q) plots and plots based on other functions of the sample distribution functions and quantiles. Note that P–Q plots have been used by Gnanadesikan & Gupta (A1970) to investigate the goodness-of-fit of a given distribution in terms of the accuracy at various sample quantiles, and by Fowlkes (A1979) to examine a sample for the presence of a mixture of two normal distributions. Associated uses in nonparametric statistics may well exist, but P–Q plots have not been used in this field.

The pair chart is a slight modification of the P–P plot, the difference being that the jump or shift with each new plotting point is the same unit amount (rather than being $1/m$ or $1/n$). Thus if $m=n$, a pair chart is just a P–P plot. A pair chart is more useful in calculating several two-sample test statistics; otherwise there seems little difference between them. The uses of the pair chart, as described by Quade (1973), are both descriptive and computational. We have already seen how certain sorts of departure from $F=G$ may be manifested in a P–P plot (Fig. 7): the same is true for the pair chart.

The pair chart for the cancer data is shown in Fig. 9. The shaded rectangles correspond to between-sample ties in the data, with the perimeters of each rectangle indicating the extremes of possible paths derived from different orderings of the tied observations. If the plot for each of these sections is defined as the diagonal of the rectangle, graphical

Fig. 8

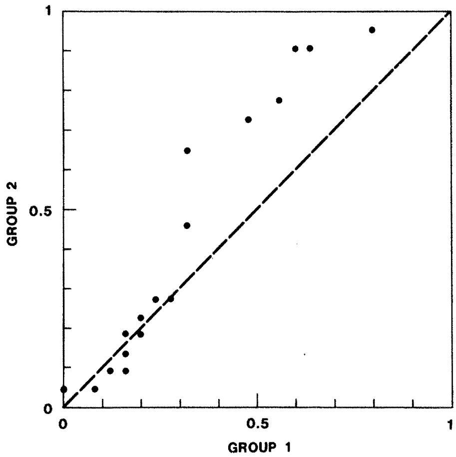

Fig. 9

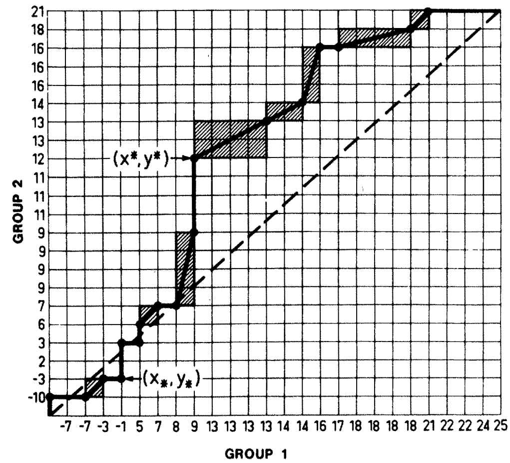

Figure 8. *Probability–probability plot. Data: scores by two sets of parients on psychological tests (Table 4).*
Figure 9. *Pair chart. Data: scores of two sets of patients on psychological tests (Table 4).*

computation of various statistics described below will result in values corresponding to using average ranks for tied observations.

A great variety of two-sample test statistics can be calculated from the pair chart. The Mann–Whitney–Wilcoxon test statistic is just the (number of squares) below the path. In Fig. 9 it is more convenient to calculate the area above the path and subtract it from mn: $mn-(21+21+\dots+\frac{1}{2})=154$. The equivalence of the Mann–Whitney and Wilcoxon forms of the statistic is easy to demonstrate from the chart (Quade, 1973). Tukey's quick test statistic is trivially calculated: in the notation introduced earlier, $L=$ number of steps in initial vertical sequences $=1$, $U=$ number of steps in last horizontal section $=4$, $T=U+L=5$. Among two-sample tests for scale, the test statistics of Ansari & Bradley, Sukhatme, Siegel & Tukey, Mood, and Crouse & Steffens are all shown by Quade to be amenable to graphical computation.

The Wald–Wolfowitz runs test, a test due to Lehmann (see for example Quade) and the Kolmogorov–Smirnov tests are three procedures for testing $F=G$ against more general alternatives. Both the runs test and Lehmann's test are easy to compute, although if ties are present the runs test may be rather unsatisfactory (see Quade for discussion of this point). Several aspects of the Kolmogorov–Smirnov statistics can be studied using the pair chart. Define

$$D^+_{mn}=\sup_{-\infty<x<\infty}\{F_m(x)-G_n(x)\},\quad D^-_{mn}=\inf_{-\infty<x<\infty}\{F_m(x)-G_n(x)\},\quad D_{mn}=\max(D^+,D^-).$$

Hodges (1958) showed that if (x_*, y_*) and (x^*, y^*) are respectively the points of the path furthest below and furthest above the diagonal of the rectangle, then

$$D^+_{mn}=\left|\frac{x_*}{m}-\frac{y_*}{n}\right|,\quad D^-_{mn}=\left|\frac{x^*}{m}-\frac{y^*}{n}\right|.$$

(The points (x^*, y^*) and (x_*, y_*) will always occur at intersections of the mn lines forming the lattice.) From Fig. 9,

$$D^+_{mn}=\left|\frac{4}{25}-\frac{2}{22}\right|=0.009,\quad D^-_{mn}=\left|\frac{8}{25}-\frac{14}{22}\right|=0.316,\quad D_{mn}=0.136.$$

Hodges also used the pair chart to demonstrate that the probability level of an observed value of D^+_{mn}, D^-_{mn} or D_{mn} could be computed in various ways; see Hodges (1958) and Quade (1973) for further details. Quade also discussed the effect of ties on D_{mn}. If m and n are not too large, it is feasible to use the pair chart to construct confidence bands based on the Kolmogorov–Smirnov statistics D_{mn}.

The introduction of the pair chart is attributed by Quade to Drion (1952), who used it to enumerate the probability distribution of D_{mn} when $m = n$, and also to obtain a partial solution to the problem of estimating the probability that one sample distribution function lies entirely above another. Gibbons (1971) describes a recursive method proposed by Hodges (1958) for evaluating the probability level of an observed Kolmogorov–Smirnov statistic. A brief description of the pair chart and its application to computing D_{mn} is given by Daniel (1978).

As a development of Q–Q plots, the concept of a shift (or treatment effect) function for the difference between two populations has been exploited by several authors (Doksum, 1974; Switzer, 1976; Doksum & Sievers, 1976, Doksum, 1977; Nair, 1978) to obtain confidence bands which yield information about a usefully large range of models. If X and Y are any two random variables with continuous distribution functions F and G respectively, then there exists a unique shift function $\Delta(x)$ such that $X + \Delta(X)$ has the same distribution as Y. In fact, $\Delta(x)$ is just the horizontal distance between $F(x)$ and $G(x)$, as shown in Fig. 10, which is based on a similar figure of Doksum (1974), and so satisfies $F(x) = G(x + \Delta(x))$, so that $\Delta(x) = G^{-1}[F(x)] - x$. The Q–Q plot of F_m against G_n is just the sample estimate $G_n^{-1}(F_m)$ of $G^{-1}(F)$, evaluating at (a subset of) the points $X_{(1)}, \ldots, X_{(m)}$. Nair (1978) notes that

(a) $\int \Delta(x)\, dF(x) = E(Y) - E(X)$,
(b) $\Delta(x)|_{x=F^{-1}(1/2)} = \text{med}\,(Y) - \text{med}\,(X)$,
(c) If F and G are failure time distributions, convexity of $\Delta(x) + x$ implies that G has a more slowly increasing failure rate than F.

(At the end of this section there is a discussion of plots for studying the hazard ratio.)

For discussion purposes, it is convenient to think of X as being the response of a control subject, and Y the response of a treated subject, with the Y response tending to be larger if the treatment is beneficial. (The introductory section of Switzer (1976) and Doksum & Sievers (1976) expound this approach clearly and simply.) The following discussion on shift functions is based on the paper by Doksum & Sievers. In this paper, four specific questions are raised: (i) Is the treatment beneficial for all members of the population, i.e. is $\Delta(x) > 0$ for all x? (ii) If not, for which part of the population is it beneficial, i.e. what is the set $\{x : \Delta(x) > 0\}$? (iii) Is a pure location shift model appropriate ($Y = a_0 + X$ in distribution)? (iv) Is a location scale shift model appropriate ($Y = a_0 + b_0 X$ in distribution)? These and other questions can be answered by computing a simultaneous confidence band $[\Delta_*(x), \Delta^*(x)]$ for $\Delta(x)$, that is, a band such that $\text{pr}\{\Delta_*(x) \leq \Delta(x) \leq \Delta^*(x), \text{ all } x\} = 1 - \alpha$, for some prescribed α. Then the above questions can be answered by determining (i) whether $\Delta_*(x) > 0$, all x; (ii) $\{x : \Delta_*(x) > 0\}$; (iii) whether a horizontal line $y = a_0$ fits between $\Delta_*(x)$ and $\Delta^*(x)$; and (iv) whether any straight line $y = a_0 + b_0 x$ fits between $\Delta_*(x)$ and $\Delta^*(x)$.

To obtain the confidence bands, let $T(F_m, G_n)$ be a distribution-free test statistic for the hypothesis $H_0 : F - G$ such that H_0 is accepted if and only if $T \leq T_\alpha$; then the set $\{\text{all functions } \Delta : T(F_m(x), G_n(\Delta(x) + x)) \leq T_\alpha\}$ is a 100 $(1-\alpha)$% confidence set for the shift function Δ. In the particular case where the inequality $T(F_m, G_n) \leq T_\alpha$ is equivalent to $h_*\{F_m(x)\} \leq G_n(x) \leq h^*\{F_m(x)\}$, the confidence set will reduce to a simple confidence

Fig. 10

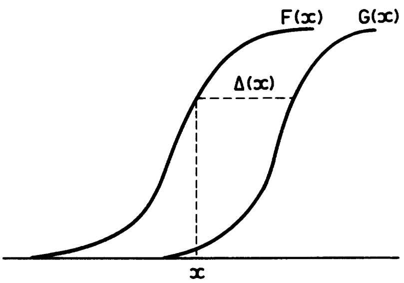

Fig. 11

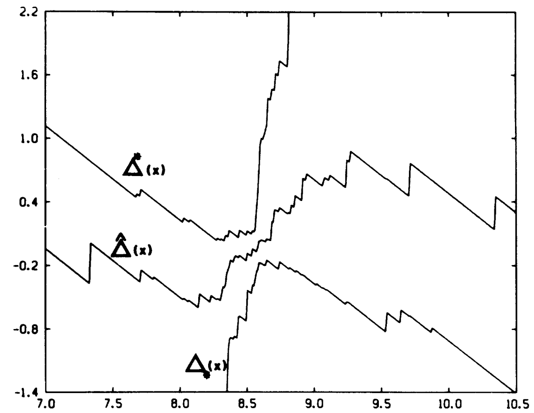

Figure 10. *Definition of shift function* $\Delta(x)=G^{-1}(F(x))-x$.
Figure 11. 80% *confidence band and estimate for shift function* $\Delta(x)$. *Data: measurements of parallax (in seconds of degree) of sun by two different methods (Table* 5).

band. The simplest such case is based on the two-sample Kolmogorov–Smirnov statistic

$$D_{mn}=\sup_{-\infty<x<\infty}|F_m(x)-G_n(x)|,$$

which produces the band

$$\begin{aligned}[\Delta_*(x),\Delta^*(x)]&=[G_n^{-1}\{F_m(x)-T_\alpha\}-x,\ G_n^{-1}\{F_m(x)+T_\alpha\}-x]\\&=[Y_{(j_i)}-x,\ Y_{(k_i)}-x],\quad x\in[X_{(i)},X_{(i+1)})\quad(0\leqslant i\leqslant m),\end{aligned}$$

where $X_{(0)}=-\infty$, $X_{(m+1)}=\infty$; $j_i=\langle ni/m-nT_\alpha\rangle$, the least integer not less than $ni/m-nT_\alpha$; $k_i=[ni/m+nT_\alpha+1]$; and $T_\alpha(mn/(m+n))=K_\alpha(m,n)$, the $100(1-\alpha)\%$ point of the distribution of D_{mn}. This particular band is illustrated (Fig. 11) for some data quoted by S.M. Stigler from an experiment to determine the parallax of the sun based on the 1761 transit of Venus. Stigler presented eight data sets, five based on a comparison of observations at a single observatory with a long list of others, and the other three resulting from pairwise comparisons of seven observatories with nine others. These two groupings determine the two samples displayed in Table 5 and used in Fig. 11. (Clearly there is no notion of a treatment, let alone a beneficial, effect here, but nevertheless it is still of interest to know whether the underlying distributions are similar.) The confidence band is an 80% band for $\Delta(x)$; since a horizontal line of the form $y=0$ fits between $\Delta_*(x)$ and $\Delta^*(x)$, there is no difference manifested between F and G at this level (80%).

Doksum & Sievers also consider other bands based on statistics of the form

$$\sup\{|F_m(x)-G_n(x)|/\psi(H_N(x))\},$$

where $N=m+n$ and $H_N(x)=(m/N)F_m(x)+(n/N)G_n(x)$, for various choices of ψ; the use of such bands allows for emphasis to be placed on particular parts of the range of the population (e.g. small x values, or the centres of the populations), and may result in more efficient estimation. Other aspects of these procedures are discussed by Doksum & Sievers, Switzer, Doksum (1977) and Nair.

Another way of comparing two populations has been developed in connection with the analysis of survival data. Suppose X and Y are failure time random variables from two populations, with respective survivor functions $F^*(t)=\mathrm{pr}(X>t)$ and $G^*(t)=\mathrm{pr}(Y>t)$, and respective cumulative hazard functions $H_F(t)=-\ln F^*(t)$ and $H_G(t)=-\ln G^*(t)$. The associated hazard functions are $h_F(t)=dH_F/dt$, $h_G(t)=dH_G/dt$. Then the hazard ratio $\theta(t)$ is defined as $h_2(t)/h_1(t)$, a quantity which will in general depend on time (t). If $\theta(t)$ is in

Table 5

Data from an experiment (reported by J. Short in 1973) *to determine the parallax of the sun based on the* 1971 *transit of Venus.*

X ($m=95$)					Y ($n=63$)		
8.50	8.65	8.50	8.54	8.54	8.70	8.43	8.63
8.50	8.35	8.80	8.56	8.58	9.66	9.09	10.16
7.33	8.71	8.40	8.54	8.54	8.50	8.50	8.50
8.64	8.31	8.82	8.74	8.94	8.65	8.44	8.31
9.27	8.36	9.02	8.91	9.24	10.33	9.71	10.80
9.06	8.58	10.57	8.40	8.30	8.07	8.07	7.50
9.25	7.80	9.11	8.40	8.33	8.50	8.36	8.12
9.09	7.71	8.66	8.57	8.59	8.60	8.60	8.42
8.50	8.30	8.34	8.69	8.81	9.61	9.11	9.20
8.06	9.71	8.60	8.55	8.56	8.50	8.66	8.16
8.43	8.50	7.99	8.51	8.50	8.35	8.58	8.36
8.44	8.28	8.58	8.57	8.58	10.15	9.54	9.77
8.14	9.87	8.34	8.58	8.58	7.77	8.34	7.52
7.68	8.86	9.64	8.63	8.68	8.23	8.55	7.96
10.34	5.76	8.34	8.56	8.57	7.92	9.03	7.83
8.07	8.44	8.55	8.41	8.33	8.42	10.04	8.62
8.36	8.23	9.54	8.64	8.62	7.75	9.04	7.54
9.71		9.07	8.43	8.37	8.23	8.71	8.28
			8.28	8.03	8.90	10.48	9.32
			8.70	8.85	7.35	8.31	6.96
			8.60	8.74	7.68	8.67	7.47

First sample based on a comparison of observations at a single observatory with a long list of others; second sample based on pairwise comparisons of seven observations with nine others.

Source: Table 4, Stigler, S.M. (1977), Do robust estimators work with real data? (with discussion), *Ann. Statist.* **5,** 1055–1098.

fact a constant (θ), X and Y are said to have proportional hazards: this implies that $H_G(t)=\theta H_F(t)$ and hence that $G^*(t)=\{F^*(t)\}^\theta$.

Methods based on plotting sample cumulative hazard functions were introduced and studied extensively by Nelson (A1969, 1970, 1972), for situations in which censored data may be present or in which several 'modes of failure' are possible. Given two sets of failure data, with one set containing failures by modes not possible for the other, failures due to these modes can be deleted from the appropriate set, and the sets then compared via plots of their estimated cumulative hazard functions.

To examine the hypothesis of proportional hazards, on the basis of independent samples from two populations, the estimated cumulative hazard functions $\hat{H}_{F_m}(t)$ and $\hat{H}_{G_n}(t)$ can be plotted against each other, rather than plotting each separately as a function of t. The graph of $\{\hat{H}_{F_m}(t), \hat{H}_{G_n}(t)\}$ is known as an H–H plot, and was introduced by *R.* Fisher (1977, 1983). Under a model of proportional hazards, the H–H plot will be approximated by the line $y=\theta x$. In general, the asymptotic H–H plot has the property that the slope of the curve at any given value t_0 is the hazard ratio associated with t_0.

As an example, consider the data in Table 6, comparing two groups of measurements of times to death from vaginal cancer of female rats insulted with the carcinogen DMBA. The two groups are distinguished by pretreatment regime. The Kaplan–Meier estimate of the survivor function may be obtained as follows (see e.g. Kalbfleisch & Prentice, 1980): let $t_1<\ldots<t_k$ be the distinct failure times for a sample from a homogeneous population, d_j the number failing at time t_j, m_j the number censored in the interval $[t_j, t_{j+1})$, and $n_j=(m_j+d_j)+\ldots+(m_k+d_k)$: then the estimate of the survivor function at time t is

$$\prod_{j:t_j<t}(1-d_j/n_j).$$

Table 6

Times from insult with carcinogen DMBA *to death by vaginal cancer, for two groups of female rats, the groups differing in pretreatment regime*

X: Group 1	143	164	188	188	190	192	206	209	213	216
($m = 19$)	220	227	230	234	246	265	304	216†	244†	
Y: Group 2	142	156	163	198	205	232	232	233	233	233
($n = 21$)	233	239	240	261	280	280	296	323	204†	
	344†									

† Censored value.
Source: Pike, M.C. (1966), A method of analysis of a certain class of experiments in carcinogenesis, *Biometrics* **22,** 142–161; reproduced as Table 1.1 of Kalbfleisch & Prentice (1980).

The H–H plot of $(-\ln F_m^*(t), -\ln G_n^*(t))$ for the rat data is given in Fig. 12; based on this plot, the hypothesis of proportional hazards seems untenable. Note that the plot could have been effected equally well by plotting $(F_m^*(t), G_n^*(t))$ on log-log paper rotated through 180° (R. Fisher, 1983). Various practical aspects of plotting survivor functions and cumulative hazard functions are discussed by Nelson (1972).

More generally, one may wish to adjust each sample for the values of certain covariates $x_1, \ldots, x_p$ prior to testing for proportional hazards. Thus, θ might be some (unknown) linear function $\sum \beta_i x_i$ of the covariates, but would not depend on t if the proportional hazards model were valid. With this particular form of Cox model (Cox, A1972), the effects of the covariates can be estimated and removed before computing cumulative hazard functions (R. Fisher, 1977; Kalbfleisch & Prentice, 1980). Related to this procedure is the idea of comparing the cumulative hazard functions obtained from a sample before and after inclusion of a particular covariate in the model (Kay, 1977; Kalbfleisch & Prentice, 1980); an alternative approach is to plot the survivor function of the 'generalized residuals' (Cox & Snell, A1968) from the model against the sample survivor function (R. Fisher, 1977). Another graphical approach to this problem has been suggested by Lagakos (1981).

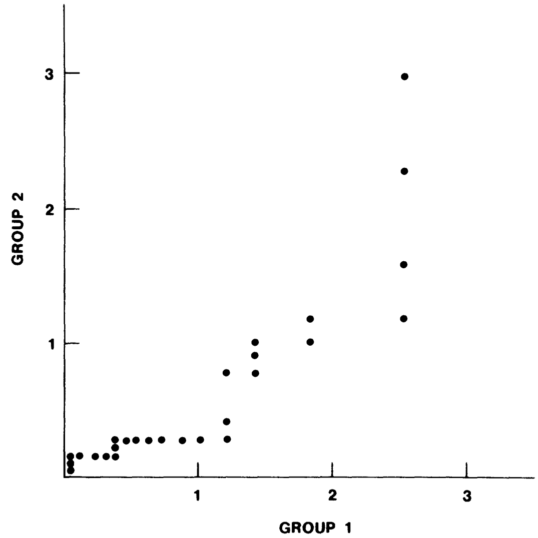

Figure 12. H–H *plot to investigate assumption of proportional hazard rates of two groups of rats. Data: times to death from vaginal cancer in female rats insulated with a carcinogen, the groups being distinguished by pretreatment regime (Table* 6).

Table 7
Masculinities of 11 *Central American countries*

92.5	94.5	97.1	97.5	98.0	98.8	99.3	100.2	101.6	102.7	105.0

Source: United Nations *Demographic Yearbook*, 1967.

Aalen (1978) and Aalen *et al.* (1980) discuss graphical comparison of two sample cumulative functions when the two samples are derived from rather general counting processes not necessarily giving rise to independent survivor times. (For this situation, unbiased estimates of the cumulative hazard functions were derived.)

2.2 *One-sample symmetry procedures*

As in § 2.1, we begin the discussion with specific procedures, and then consider more general ones. By exploiting a natural duality between the two-sample homogeneity problem and the one-sample symmetry problem, a variety of the two-sample procedures discussed in Part I can be adapted for use in this section. (For example, if $X_1, \ldots, X_n$ is a random sample drawn from $F(x)$ then, under the hypothesis (H_0) that the population is symmetric about zero, the 'pseudosample' $-X_1, \ldots, -X_n$ has the same joint distribution as $X_1, \ldots, X_n$. Intuitively, a two-sample test for location shift applied to $(X_1, \ldots, X_n)$ and $(-X_1, \ldots, -X_n)$ could be used to examine the validity of H_0.) Throughout this section we shall assume that $X_1, \ldots, X_n$ is a random sample from a continuous population with distribution function $F(x-\theta)$, where θ is the (known or unknown) population median. The sample order statistics will be denoted by $X_{(1)}, \ldots, X_{(n)}$.

(i) *Centre of symmetry*. Suppose that the underlying population may be assumed symmetric. The Hodges–Lehmann estimate of θ based on the Wilcoxon signed-rank test is $\hat{\theta}_\omega = \text{med}\,\{\frac{1}{2}(X_i + X_j),\ 1 \leqslant i \leqslant j \leqslant n\}$. A similar multiplicity of graphical techniques exists for calculating $\hat{\theta}_\omega$ as exists for the Mann–Whitney–Wilcoxon two-sample shift estimate $\hat{\theta}_w$; see § 2.1(i). To illustrate the most well-known of these, we compute $\hat{\theta}_\omega$ and an associated confidence interval for the data on masculinities of several Central American countries provided in Table 7.

Mark the data along a horizontal axis and construct two sets of parallel lines, at 45° and 135° to the axis, respectively, as shown in Fig. 13. The collection of intersections so

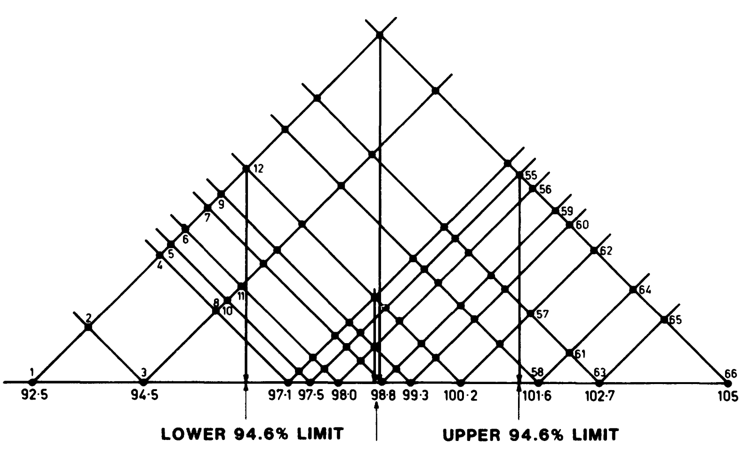

Figure 13. *Tukey's method of determining the Wilcoxon estimate of centre of symmetry. Data: masculinities of Central American states (Table* 7).

formed, when projected down onto the axis, comprise (with the data themselves) the set of Walsh averages or elementary estimates $\{\frac{1}{2}(X_i+X_j), 1 \leq i \leq j \leq n\}$. Thus, by moving a vertical line across from either end, the intersections can be enumerated sequentially and the median intersection and hence median Walsh average determined. In this case, there are 66 such averages, so that the median is the average of the two middle ones (namely $\frac{1}{2}(97.1+100.2)$ and $\frac{1}{2}(92.5+105)$), 98.7. Similarly, if a confidence interval is desired, one determines the critical values (12 and 55, for $n=11$ and a 94.6% confidence interval) from tables of Wilcoxon's statistic, e.g. Noether (1971), and locates the appropriate midranges.

This is one of the best-known graphical procedures in nonparametric statistics, and was attributed to J.W. Tukey by Moses (1965); the earliest reference to it appears to be Moses (1953). Hollander & Wolfe (1973) provide some discussion of the use of the midpoint of the confidence interval as an estimate of θ; Conover (1971) considers the continuity assumption and the effect of ties; Daniel (1978) describes the method and its applicability to problems involving matched pairs.

The other methods of calculating these point and interval estimates of θ are exact one-sample analogues of the two-sample methods in § 2.1 and will not be illustrated here. The analogue of Moses' method is described by Noether (1971), together with a discussion of the effect of ties in the data. Hoyland (1964) describes the analogue of his two-sample tabular procedure; see also Lehmann (1975). The sliding papers method of Hodges & Lehmann (1963) can easily be adapted to the one-sample situation. Hettmansperger & McKean (1974) give a graphical demonstration of the relationship between interval estimation and hypothesis testing, based on the Wilcoxon statistic; compare with the remarks in § 2.1(i). Finally, the Wilcoxon statistic can be calculated from the P plot (N. Fisher, 1981), the one-sample analogue of the pair chart; see § 2.2(ii).

Jaeckel (1969) investigated a generalization of the Wilcoxon estimate, in which different midranges $\frac{1}{2}(X_i+X_j)$ could be assigned different weights; the estimate is then that midrange in the ordered sequence for which the cumulated weights are nearest $\frac{1}{2}$. This estimate is clearly readily computed using graphical procedures of the above type. Other plausible one-sample test statistics which are analogues of two-sample statistics and can easily be calculated from the P plot are discussed in § 2.2(ii).

(ii) *General assessment of symmetry.* The discussion below parallels that in § 2.1(iii) on P–P and Q–Q plots and the shift function. Wilk & Gnanadesikan (1968) described a quantile plot (which we shall term a Q plot) obtained by plotting $X_{(i)}$ against $X_{(n+1-i)}$, $1 \leq i \leq [\frac{1}{2}n]$. If we set $\mathbf{F}(x)=1-F(-x)$ and $\mathbf{F}_n(x)=1-F_n(-x)$, a Q plot corresponds to a Q–Q plot as described in § 2.1(iii), with $G_n(x) \equiv \mathbf{F}_n(x)$. Assuming that the underlying distribution is symmetric about its median θ, the plot should be approximately linear, and reasonably well fitted by the equation $y=2\theta-x$. Because linearity is more easily assessed relative to a horizontal rather than a sloping line, J.W. Tukey suggested plotting $(X_{(n+1-i)}+X_{(i)})$ against $(X_{(n+1-i)}-X_{(i)})$, $1 \leq i \leq [n/2]$, which should result in a roughly constant plot of the form $y=2\theta$ under the symmetry assumption. A third plot suggested by Doksum (see description of 'symmetry function' below) is to plot $X_{(i)}$ against $\frac{1}{2}(X_{(i)}+X_{(n+1-i)})$ for $1 \leq i \leq n$; the plot should have approximate form $y=\theta$ for θ the centre of symmetry.

These three plots are illustrated in Figs. 14a, b, c. Note that the Wilk–Gnanadesikan and Tukey plots have been made for $i=1, \ldots, n$, rather than $i=1, \ldots, [\frac{1}{2}n]$: although these plots have intrinsic symmetry, it is easier to detect departure from linearity if the complete plot is presented. The data used are those in Table 8, the results of 23 determinations of the velocity of light in air, by A.A. Michelson in 1882. It is clear from each of the plots that there is substantial asymmetry in the data, suggesting that the

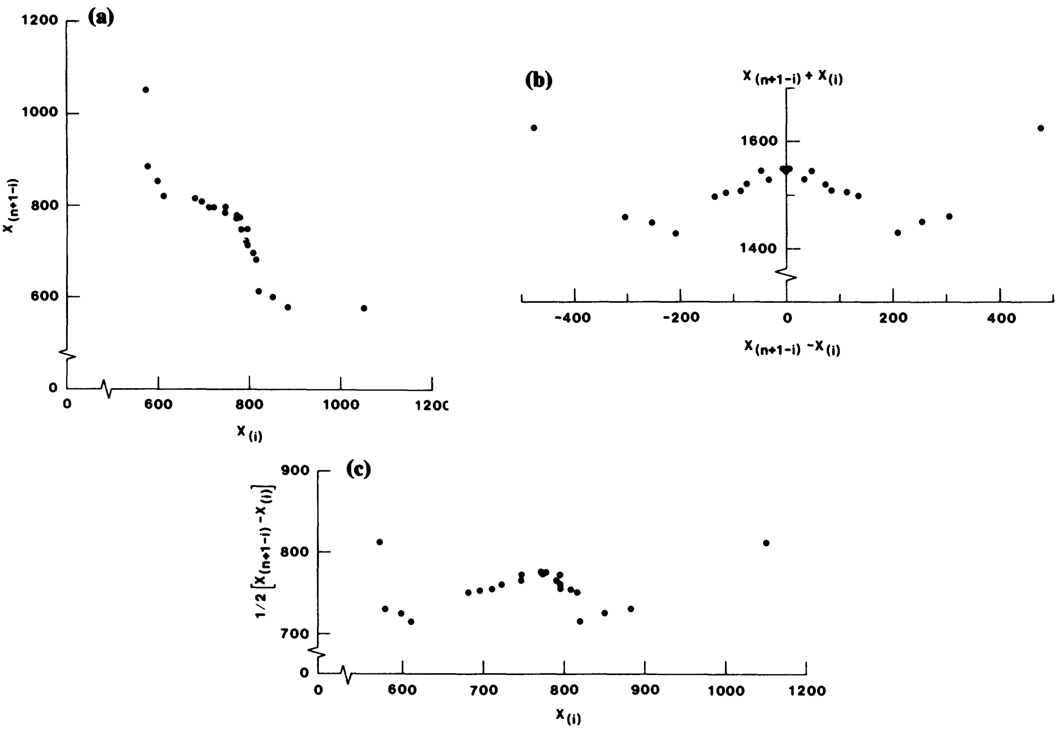

Figure 14. *Quantile plots for assessing symmetry. Data: determinations of the velocity of light (Table 8). (data +299000 give velocity in km/s.)* (a) *Wilk–Gnanadesikan quantile plot.* (b) *Tukey's plot.* (c) *Doksum's plot.*

question of estimating the location of the underlying distribution reduces to one of estimating the median, without the assumption of symmetry.

In the one-sample situation, the analogues of the P–P plot and the pair chart coincide. A one-sample probability plot (hereafter called a P plot) is a P–P plot of $X_1-\theta^*,\ldots,X_n-\theta^*$ against $-(X_1-\theta^*),\ldots,-(X_n-\theta^*)$, that is, of $[F_n(x-\theta^*),\mathbf{F}_n(x-\theta^*)]$ for $-\infty<x<\infty$, where θ^* is the true or estimated median. P plots complement Q plots as data-analytic tools in the same way as P–P plots complement Q–Q plots (N. Fisher, 1981), Q plots being more sensitive to departures for symmetry in the tails than in the middle, and vice versa for P plots. For unimodal distributions, a P plot of data from a symmetric distribution about a point which is not the true median behaves differently from a P plot of data from an asymmetric distribution (about any point, median or otherwise), as can be seen in Fig. 15, and described in more detail by Fisher (1981). If the underlying distribution is symmetric about, say, 0, then the asymptotic form of the P plot, that is, a

Table 8

23 determinations of the velocity of light in air, made during the period 12 *October*–14 *November,* 1882. *The values below,* +299000, *are Michelson's determinations in km/s.*

883	711	578	696	851
816	611	796	573	809
778	599	774	748	723
796	1051	820	748	
682	781	772	797	

Source: Table 7, Stigler, S.M. (1977), Do robust estimators work with real data? (with discussion), *Ann. Statist.* **5,** 1055–1098.

Fig. 15

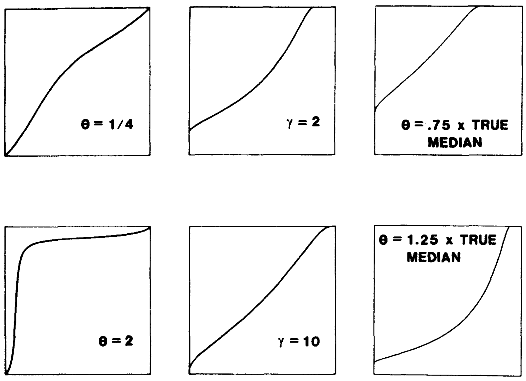

Fig. 16

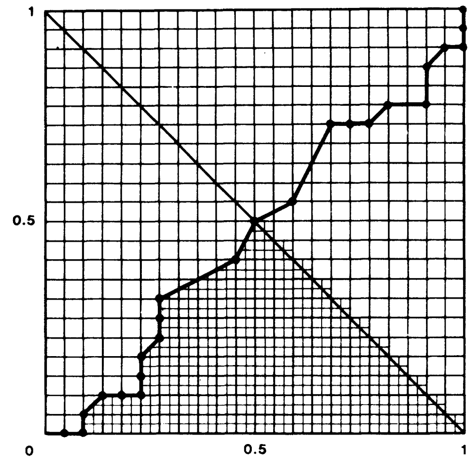

Figure 15. *Asymptotic behaviour of P plots under various departures from symmetry.*

Figure 16. *Computation of Wilcoxon statistic from P plot. Data: effect of group therapy on delinquents* (*Table* 9).

plot of $\{(F(x), \mathbf{F}(x)], -\infty < x < \infty\}$, is a straight line from (0, 0) to (1, 1). If the distribution is symmetric about $\theta \neq 0$, then the asymptotic P plot about the *incorrect* median value 0 will have the characteristic shape shown in the two plots in the first column of Fig. 15. If the distribution is asymmetric and unimodal with median θ_0, the asymptotic P plot $\{(F(x-\theta_0), \mathbf{F}(x-\theta_0)], -\infty < x < \infty\}$ has the characteristic shape shown in the two plots in the second column of Fig. 15. In this latter case, if the incorrect median value $\theta \neq \theta_0$ is used as the postulated centre of symmetry, plots of the type shown in the third column will arise.

The computational uses of the P plot parallel those of the pair chart, albeit in a more limited way. The Wilcoxon signed rank statistic

$$W_n = \sum_{1 \leq i \leq n} I[X_i > 0]\,(\text{rank}\,|X_i|)$$

for testing symmetry about 0 is easily seen to be the shaded area in Fig. 16 (the P plot of the data of Table 9 on the effect of group therapy on delinquents) when one recalls Tukey's alternative representation of the statistic as

$$\sum_{1 \leq i \leq j \leq n} I[X_i + X_j > 0].$$

Here, ties between X_i's and $-X_j$'s are resolved by average rank. The computation of Butler's statistic

$$B_n = \sup_x |F_n(x) - (1 - F_n(-x))| \equiv \sup_x |F_n(x) - \mathbf{F}_n(x)|$$

for the same hypothesis requires identification of the points (x_*, y_*) and (x^*, y^*) on the path farthest below and above the diagonal respectively (Fig. 17). From Figs. 16 and 17,

Table 9

Effect on 22 matched pairs of delinquents of group therapy in terms of emotional and social adjustment, measured by difference in rating between treated and control in each pair.

−1.1	−0.9	−0.6	−0.4	−0.4	−0.2	0.0	0.0	0.0	0.0	0.1	0.3
0.3	0.4	0.4	0.5	0.5	0.6	0.7	0.7	1.0	1.2		

Source: Gerstein, C. (1952), Group therapy with institutionalised juvenile delinquents, *J. Genetic Psychol.* **80,** 35–64.

$W_n = 118\frac{1}{2}$ and

$$B_n = \max\left\{\left|\frac{5}{22}-\frac{2}{22}\right|, \left|\frac{6}{22}-\frac{7}{22}\right|\right\} = \frac{3}{22}.$$

The exact significance level of Butler's statistic can also be computed from the P plot, analogously to the computation of the significance level of the Kolmogorov–Smirnov statistic from the pair chart. This result was conjectured by Fisher (1981) and subsequently proved by R.D. John in a personal communication as follows.

The possible paths must pass through one of the n points on the diagonal $x+y=1$. Altogether there are 2^n such paths, namely 1 to the point (0, 1),

$$\binom{n}{1} \text{ to the point } \left(\frac{1}{n}, \frac{n-1}{n}\right),$$

and generally,

$$\binom{n}{r} \text{ to the point } \left(\frac{r}{n}, \frac{n-r}{n}\right).$$

(The symmetry of the paths $x+y=1$ determines their behaviour on the other side of $x+y=1$ once the behaviour below this line is known.) The probability level can then be calculated by determining the number of such paths which do not encroach in the region outside the lines $y = x \pm B_n$. The same recursive technique described by Quade (1973) for the pair chart can be applied to count these paths.

Further details concerning the Butler, the Kolmogorov–Smirnov and related statistics, plus some suggested analogues of Tukey's quick test and the Wald–Wolfowitz runs test are given by Fisher (1981).

Corresponding to the concept of a shift function $\Delta(x)$ for the difference between two populations, § 2.1(iii), it is possible to define a symmetry function $\Lambda(x)$ for a single population, which measures the way in which a given population departs from being symmetric. This notion was introduced by Doksum (1975) and developed further by Doksum, Fenstad & Aaberge (1977); the symmetry function is defined by $\Lambda(x) = \frac{1}{2}\{1 - \mathbf{F}^{-1}(F_n(x))\}$, with sample estimate $\hat{\Lambda}(x) = \frac{1}{2}\{1 - \mathbf{F}_n^{-1}(F(x))\}$. When the underlying population is symmetric about θ_0, $\Lambda(x) \equiv \theta_0$.

Doksum *et al.* (1977) make the following comments about $\Lambda(x)$: (i) when the population is skew to the right, $\Lambda(x)$ lies wholly on or above θ_0; (ii) the three symmetry plots (Figs. 14a, b, c) are, respectively, plots of $\{x, \mathbf{F}_n^{-1}(F_n(x))\}$, $\{-\mathbf{F}_n^{-1}(F_n(x)) - x, -\mathbf{F}_n^{-1}(F_n(x)) + x\}$, and $\{x, \hat{\Lambda}(x)\}$, (iii) confidence bands for $\Lambda(x)$ can be obtained similarly to those for $\Delta(x)$; (iv) tests for symmetry about θ_0, or more generally, for symmetry, can be performed by determining respectively whether the line $y = \theta_0$, or any horizintal line $y = \theta$, will fit within the confidence band (the latter test being conservative).

To illustrate the use of the symmetry function, we calculate the simplest confidence band given by Doksum *et al.* (1977) which is based on Butler's statistic: using arguments analogous to those in § 2.1(iii), they obtain

$$[\Lambda_*(x), \Lambda^*(x)] = [\tfrac{1}{2}(x + X_{(j_i)}), \tfrac{1}{2}(x + X_{(k_i)})], \quad x \in [X_{(i)}, X_{(i+1)}) \quad (0 \leq i \leq m),$$

where $X_{(0)} = -\infty$, $X_{(n+1)} = \infty$; $j_i = n + 1 - \langle i - nB_\alpha(n)\rangle$, $k_i = n - [i + nB_\alpha(n)]$; and $nB_\alpha(n)$ is the $100(1-\alpha)\%$ point of the distribution of B_n. Figure 18 shows the estimate $\hat{\Lambda}$ and associated 95% confidence band for the data in Table 10, quoted by S.M. Stigler from an earlier experiment by Michelson to estimate the velocity of light in air (100 measurements made in 1879). Stigler quotes the 'true' value 734.5 (+299000) km/s for the velocity of light in air; since the horizontal line $y = 734.5$ fits between the upper and lower bands, this seems to be a reasonable assertion based on the data.

Fig. 17

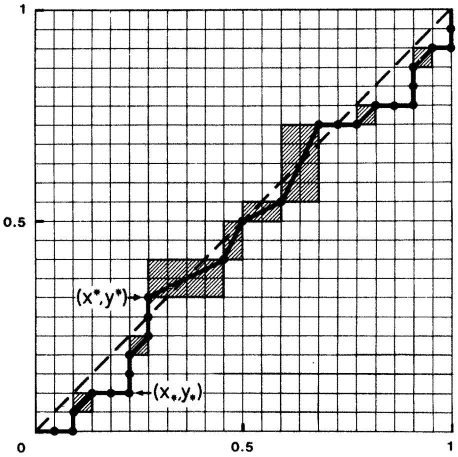

Fig. 18

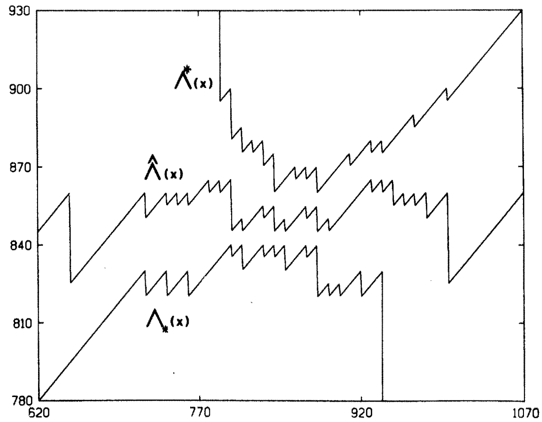

Figure 17. *Computation of Butler's statistic from P plot. Data: effect of group therapy on delinquents (Table 9).*
Figure 18. 95% *confidence band and estimate of symmetry function* $\Lambda(x)$. *Data: determinations of velocity of light (Table* 10). *(data* +299000 *give velocity in km/s.)*

2.3 Association and regression procedures

(i) *Association.* Let $(X_1, Y_1, \ldots, X_n, Y_n)$ be a random sample from some continuous bivariate population with distribution function $F(x, y)$. Denote by R_i the rank of X_i among $X_i, \ldots, X_n$, and by S_i the rank of Y_i among $Y_1, \ldots, Y_n$.

It is convenient to separate the various techniques into those providing representations of association and those which simply facilitate computation of a test statistic. The former group comprises techniques for Kendall's tau and Spearman's rho, and the linked vector method.

Kendall's sample rank correlation coefficient $\hat{\tau}_n$ is defined by

$$\hat{\tau}_n = \sum_{1 \le i < j \le n} \operatorname{sgn}\{(X_i - X_j)(Y_i - Y_j)\} \bigg/ \binom{n}{2},$$

Table 10

100 *determinations of the velocity of light in air, made during the period* 5 *June–*2 *July,* 1879. *The values below,* +299000, *are Michelson's determinations in km/s.*

850	960	880	890	890
740	940	880	810	840
900	960	880	810	780
1070	940	860	820	810
930	880	720	800	760
850	800	720	770	810
950	850	620	760	790
980	880	860	740	810
980	900	970	750	820
880	840	950	760	850
1000	830	880	910	870
980	790	910	920	870
930	810	850	890	810
650	880	870	860	740
760	880	840	880	810
810	830	840	720	940
1000	800	850	840	950
1000	790	840	850	800
960	760	840	850	810
960	800	840	780	870

Source: Table 6, Stigler, S.M. (1977). Do robust estimators work with real data? (with discussion), *Ann. Statist.* **5**, 1055–1098.

where sgn $(u) = -1, 0, 1$ according as $u<0$, $u=0$ or $u>0$. (If there are ties within the X or Y sample, various adjustments can be applied; see Kendall (A1975).) The graphical representation of $\hat{\tau}_n$ is based on its connection with the notion of 'disarray' of one rank ordering of n objects relative to another. The disarray of $Y_1, \ldots, Y_n$ relative to $X_1, \ldots, X_n$ is the smallest number (s say) of *simple* interchanges (interchanges of adjacent Y's) required to bring the Y's into the same rank order as the X's. Note that knowledge of this relationship seems to date back at least as far as Rodrigues (A1839.) As an example, Fig. 19 exhibits this disarray for the data in Table 11, concerning average life expectancy and *per capita* income for nine petroleum-exporting states. In Fig. 19, the data have been replaced by their ranks. Corresponding numbers are then joined by straight lines, and the number of intersections of these lines is just s, the number of simple interchanges needed to convert the second rank order to the first. Then

$$\hat{\tau}_n = 1 - 2s \Big/ \binom{n}{2} = 1 - 2 \times 14 \Big/ \binom{9}{2} = \frac{2}{9}.$$

An early reference to this display of disarray is Symonds (1927), who presented it for two different rankings without calculating s, and then quoted the Pearson product moment correlation of the two rankings (i.e. Spearman's rho). Symonds commented that '... the slope of these lines [in Fig. 19, departures from vertical] indicates the displacement in position and failure to correlate perfectly'. Subsequently, Sandiford (1929) used the graphical display to calculate s, and hence $\hat{\tau}_n$. A more recent discussion of τ as a coefficient of disarray is given by Griffin (1958); Shah (1961) provides a simplification of Griffin's method for dealing with ties.

As noted above, Spearman's rank correlation coefficient for a sample is simply the Pearson sample correlation computed using the ranks of the data. It can be rewritten as

$$\hat{\rho}_n = \left\{ \sum_{i=1}^{n} R_i S_i - n(n+1)^2/4 \right\} \Big/ \{(n^3 - n)/12\};$$

if the data are re-ordered so that the X_i's are in increasing order, and if s_i is the rank of the ith Y in this new ordering of the data, then

$$\hat{\rho}_n = \left\{ \sum_{i=1}^{n} i s_i - n(n+1)^2/4 \right\} \Big/ \{(n^3 - n)/12\}.$$

This can be calculated from a graphical representation (of the relationship between the X ordering and the Y ordering expressed in the same) somewhat akin to the linked vector method of Taguri *et al.* (1976) given below, and to an unpublished representation of $\hat{\rho}_n$ due to T. Yanagawa (personal communication). To do this, form the table

$$\begin{array}{cccc} i & 1 & \cdots & n \\ \tfrac{1}{2}(n+1) - s_i & \tfrac{1}{2}(n+1) - s_1 & \ldots & \tfrac{1}{2}(n+1) - s_n \end{array}$$

and plot $x_i = \sum_j (\tfrac{1}{2}(n+1) - s_j)$, where the sum is over $j = 1, \ldots, i$, against $y_i = i - 1$ for $1 \leqslant i \leqslant n$. (Note that x_i is a simple shift $\tfrac{1}{2}(n+1) - s_i$ from x_{i-1}.) Then construct a step-function through these points, with its jumps at $x_1, \ldots, x_{n-1}$. The *relative rank function* so constructed starts at $x = y = 0$ and finishes at $x = 0$, $y = n-1$, and may lie wholly on one side of the y axis or may take both positive and negative x values. The nett area between the function and the y axis (with area to the left of this axis counted negatively) is the numerator $\hat{\rho}_n$. The denominator is the area between the y axis and the path corresponding to correlation $\hat{\rho}_n = 1$, but may more conveniently be computed as $(n^3 - n)/12$.

Figure 20 shows the relative rank function for the data in Table 12, on the correlation between elapsed time and distance travelled before recapture for 10 tagged tuna. The nett area $= 14.5$, whence $\hat{\rho}_{10} = 14.5/(990/12) = 0.176$.

Fig. 19

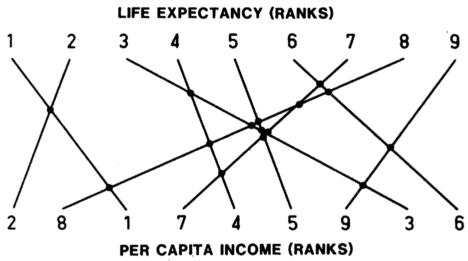

Fig. 20

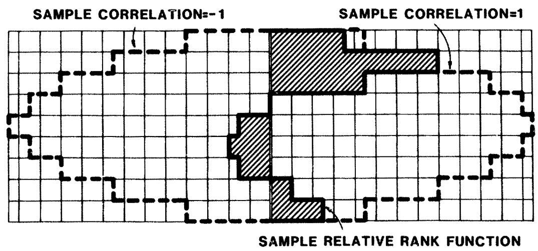

Figure 19. *Calculation of disarray, and hence Kendall's tau. Data: life expectancy and per capita income for some petroleum exporting states (Table 11).*
Figure 20. *Representation of Spearman's rho. Data: distances travelled and times between release and capture for Skipjack tuna (Table 12).*

Table 11

Life expectancies (in years) and per capita income (in US dollars) for 9 petroleum-exporting states

Life expectancy	36.9	42.3	47.5	50.0	50.7	51.6	52.1	52.3	66.4
Per capita income	180	1530	110	1280	430	560	3010	360	1240

Source: Leinhardt, s. & Wasserman, S.S. (1979), Teaching regression: an exploratory approach, *Am. Statistician* **33,** 196–203.

Table 12

Time between date of release and date of recapture (in days), and distance from place of release that capture was effected (in nautical miles), for each of 10 Skipjack tuna

Time	64	67	106	161	164	169	182	192	230	231
Distance	659	744	1616	683	682	678	594	637	1723	1682

Source: Australian tuna caught off Solomon Islands, *Australian Fisheries* **39** (2) p. 17.

Note that this representation of $\hat{\rho}_n$ is not symmetric in its treatment of the variables. However, it is of some relevance to the work of Gordon (A1979a, b) on the identification of data pairs contributing to agreement between ranking and of blocks of data pairs exhibiting agreement, and it provides some insight into the interpretation of the linked vector plots proposed by Taguri *et al.* (1976). The situation considered is one in which interest lies in the comparison of several 'explanatory' variables with a single objective variable. Suppose that n observations are made on a random $(k+1)$ vector $(X_i^{(0)}, \ldots, X_i^{(k)})$ for $1 \leqslant i \leqslant n$. Without loss of generality, suppose that the vectors have been re-ordered so that the values of the objective variable $X^{(0)}$ are increasing: $X_1^{(0)} \leqslant \ldots \leqslant X_n^{(0)}$. Let $R_i^{(j)}$ denote the rank of $X_i^{(j)}$ among the n independent realizations $X_1^{(j)}, \ldots, X_n^{(j)}$ of $X^{(j)}$ for $1 \leqslant j \leqslant k$. We then have the situation shown in Table 13a. Now perform on each column the transformation

$$\alpha_i^{(j)} = \frac{R_i^{(j)} - 1}{n-1}\pi \quad (1 \leqslant i \leqslant n, 1 \leqslant j \leqslant k)$$

to obtain Table 13b. Finally, associate with each quantity $\alpha_i^{(j)}$ the unit vector $\bar{\alpha}_i^{(j)}$ with argument $\alpha_i^{(j)}$. For each variable $X^{(j)}$, plot the vector $\bar{\alpha}_{i-1}^{(j)}, \ldots, \bar{\alpha}_n^{(j)}$ sequentially, with $\bar{\alpha}_i^{(j)}$ starting at the endpoint of $\bar{\alpha}_{i-1}^{(j)}$, and $\bar{\alpha}_1^{(j)}$ starting at the origin (0). The $k+1$ linked-vector paths so formed will finish at a common point (F) at the top of Fig. 21.

Table 13

(*a*) *Rank*, $R_i^{(j)}$, *of observation variable*, $X_i^{(j)}$, *among* n *independent realizations*, $X_1^{(j)}, \ldots, X_n^{(j)}$. (b) *After transformation*

Sample no.	(a) Before transformation: Objective	(a) Objective variable $X^{(1)} \cdots X^{(k)}$			(b) After transformation: Objective	(b) Objective variable $X^{(1)} \cdots X^{(k)}$		
1	1	$R_1^{(1)}$	...	$R_1^{(k)}$	$\alpha_1^{(0)} = 0$	$\alpha_1^{(1)}$	...	$\alpha_1^{(k)}$
2	2	$R_2^{(1)}$	...	$R_2^{(k)}$	$\alpha_2^{(0)} = \pi/(n-1)$	$\alpha_2^{(1)}$	...	$\alpha_2^{(k)}$
.	.				.			
.	.				.			
.	.				.			
n	n	$R_n^{(1)}$	...	$R_n^{(k)}$	$\alpha_n^{(0)} = \pi$	$\alpha_n^{(1)}$	...	$\alpha_n^{(k)}$

As an example, Fig. 21 shows the linked-vector paths for the data in Table 14, on the amounts of four trace elements present at various depths in Antarctic snows. One of the metals (aluminium) has been chosen as the objective variable. The clear indications are that K is highly positively correlated with Al, and Pb highly negatively correlated. Taguri *et al.* (1976) observe that the nett area between any given path and the central vertical line, as a proportion of the largest possible area, is approximately equal to Spearman's rho. The reason for this is clear, in the light of the representation of $\hat{\rho}_n$ presented earlier. Note that in this case,

$$\hat{\rho}_n(\text{Pb, Al}) = -0.66, \quad \hat{\rho}_n(\text{Ag, Al}) = 0.10, \quad \hat{\rho}_n(\text{depth, Al}) = 0.46, \quad \hat{\rho}_n(\text{K, Al}) = 0.87).$$

The sort of information highlighted by the linked-vector plot, especially with larger samples, is the differential association between an objective and an explanatory variable over different ranges of their respective populations. For example, the plot of depth against Al suggests positive association over the lower range of each variable, with a hint of negative association in the upper parts of the ranges. A geochemical problem relating to this occurs in multielement analysis of samples taken sequentially along a drill-core. The samples are supposed to form a homogeneous domain. However, a linked-vector plot of trace element concentrations against distance along the drill-core could detect a change in some element concentrations indicative of heterogeneity in the sampled zone.

Fig. 21

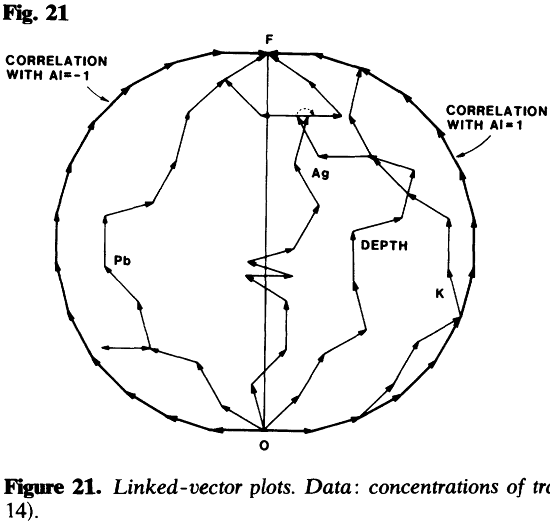

Fig. 22

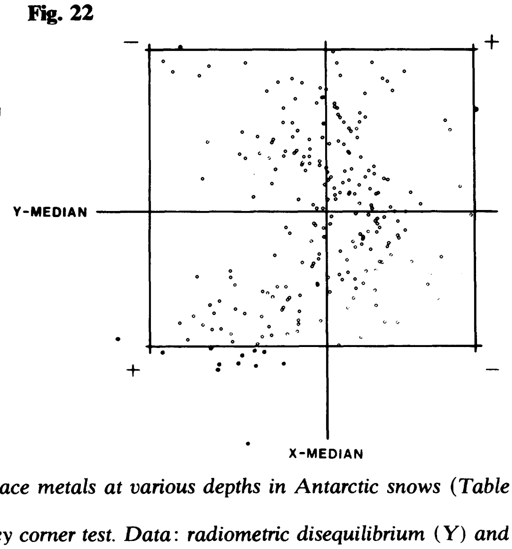

Figure 21. *Linked-vector plots. Data: concentrations of trace metals at various depths in Antarctic snows (Table* 14).

Figure 22. *Scatterplot of data for use with Olmstead–Tukey corner test. Data: radiometric disequilibrium (Y) and depth (X) for uranium samples.*

Table 14

Measured concentrations of four trace elements at various depths in Antarctic snows. 'Depth' has been recorded here as height above lowest average depth

Depth (m)	Potassium (K) (10^{-9} g/g)	Aluminium (Al) (10^{-9} g/g)	Lead (Pb) (10^{-12} g/g)	Silver (Ag) (10^{-12} g/g)
4.86	1.06	1.14	41	6.0
4.46	0.93	1.25	25	6.8
4.06	0.62	0.78	30	7.6
3.66	0.75	0.88	28	4.0
3.25	1.11	0.97	27	6.6
2.77	1.40	1.81	15	2.9
2.41	1.57	1.70	47	22.0
1.99	1.13	1.37	21	14.0
1.56	1.88	3.06	23	7.2
1.16	2.05	3.66	20	6.5
0.77	1.45	1.93	18	6.6
0.38	1.12	1.39	24	2.1
0.00	1.25	0.95	26	0.2

Source: Table 1, Boutron, C. & Lorius, C. (1979), Trace metals in Antartic snows since 1914, *Nature* **277**, 551–554.

A simple and effective way to deal with ties is to use average ranks; however, Taguri *et al.* (1976) have other recommendations for this, and also give further discussion of the interpretation and uses of the method.

Finally, we make brief mention of a suggestion by Bradley (1963) of a mechanical method of obtaining a graphical display of rank association between two variables. Write out each pair of values (X_i, Y_i) on a separate computer card, sort the cards so that the X's increase, and draw a 45° line across an edge of the deck (i.e. not the face of the front or back card), thereby imparting a small mark to the edge of each card. Then re-sort the cards so that the Y's increase. The small marks on the cards are then scattered (unless $|\hat{\rho}_n(X, Y)| = 1$) and give a scatterplot of rank association.

We turn now to consideration of tests of association. The first of these is the Olmstead–Tukey corner test for association in large samples, in situations where it is believed that information about association may be contained in points on the periphery of the data set. Figure 22 is a scatterplot of radiometric disequilibrium against depth for 242 samples of uranium from an Australian one body (data kindly supplied by Dr. B.L. Dickson, CSIRO Division of Mineral Physics). To effect the test, draw in the X and Y medial lines as shown, and label the quadrants so formed as $+, -, +, -$ serially from the top right-hand corner. Then, stating at the top, move vertically down counting points with decreasing y values until it is necessary to cross the X medial line, and attached to this count the sign of the quadrant in which the points lie. In the case, only one point is counted, and receives '−'. Proceed in similar fashion for the bottom, left- and right-hand sides of the data set to obtain the counts, +11, +2 and +1 respectively. The test statistic is $|-1+11+2+1| = 13$, which, if we use the table provided by Olmstead & Tukey (1947), suggests that the hypothesis of no association should be rejected at the 2.5% level.

Olmstead & Tukey discuss handling of ties, and extension of the technique to higher dimensions (joint association of several variables). Mood (1950), Quenouille (1972) and Daniel (1978) also describe the test.

Quenouille (1952, 1972) proposes a variety of graphical 'quick' tests to detect association (monotonic or otherwise) between X and Y: we illustrate two of them. A large-sample procedure for detecting monotone association is illustrated in Fig. 23a. The data, given in Table 15, consist of measurements of psychological test score and reciprocal

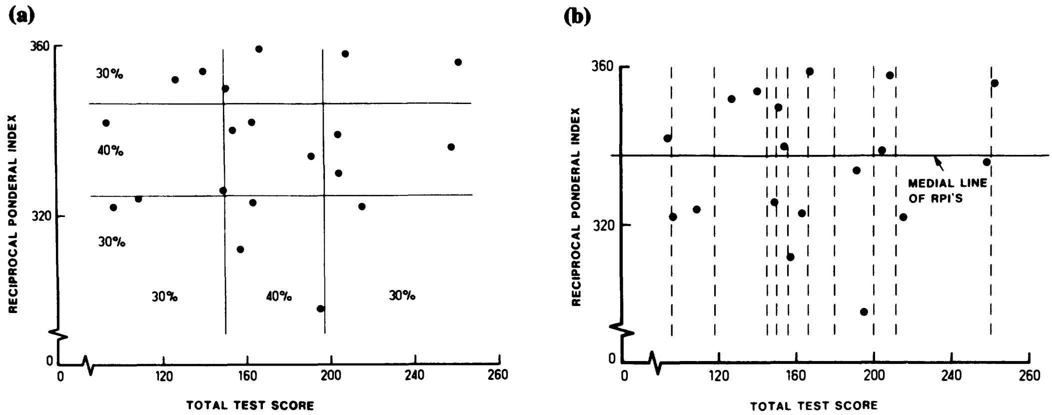

Figure 23. *Quenouille's quick tests for association. Data: psychological test scores and body measurements for first-born of sets of twins (Table* 15). (a) *Test for monotone association.* (b) *Test for general association.*

ponderal index (stature divided by cube root of weight) for the first-born of each of 20 sets of twins. On a scatterplot of the data draw two vertical lines which divide the X's in proportions approximately 3:4:3 and similarly for the Y's. Denote by n_1, n_2, n_3, n_4 the number of points in the four corner cells (labelled cyclically from the top right-hand cell), calculate $N = n_1 - n_2 + n_3 - n_4$ and test it as a normally distributed random variable with mean zero and variance $n_1 + n_2 + n_3 + n_4$; $N = 2 - 1 + 2 - 2 = 1$, which is clearly not an extreme value for a normal random variable with mean 0 and variance 7, although the sample size is probably too small to make the normal approximation reasonable. This test and other similar procedures are also discussed by Shahani (1969).

The second method, to detect general association, is illustrated for the same data set (excluding two points) in Fig. 23b. On the scatterplot draw in the medial line for the Y's. Then proceed from left to right, drawing in vertical lines dividing the data into sets of points with each set comprising a run of consecutive X value points on the same side of the medial line. The number of sets is the critical value: if it is too small, (as assessed from Table 7 of Quenouille (1972)), the hypothesis of independence is rejected. In Fig. 23b, there are 11 such sets, which for a sample size of 18 is significant at the 5% level (critical value = 12).

Table 15

Measurements of Total (T) *of scores on three psychological tests and of reciprocal ponderal index,* RPI = *stature*/(*weight*)$^{1/3}$, *for* 20 *individuals* (*first-born of* 20 *sets of twins*)

Obs no.	T	RPI	Obs. no.	T	RPI	Obs. no.	T	RPI
1	167	359	8	163	323	15	157	312
2*	204	330	9	94	342	16	151	350
3	195	298	10	127	352	17	204	339
4	149	326	11	191	334	18	140	354
5	215	324	12	154	340	19	258	336
6	262	356	13	208	358	20	109	324
7	97	322	14	163	342			

* Not used in Fig. 23.

Source: Clark, P.J., Vandengerg, S.G. & Proctor, C.H. (1961), On the relationship of scores on certain psychological tests with a number of anthropometric characters and birth order in twins, *Human Biol.* **33,** 163–180.

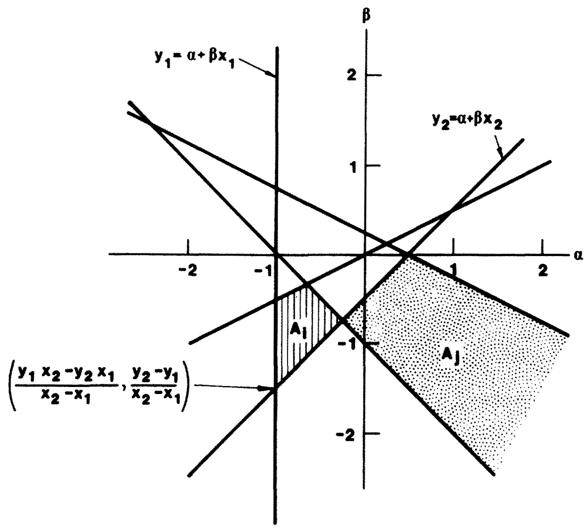

Figure 24. *Parameter space plot, showing lines $y_i = \alpha + \beta x_i$ in $\alpha - \beta$ space, coordinates of intersection of first two lines, and typical regions with differing patterns of signs of residuals. Data: see text and Brown* (1980).

(ii) *Regression.* Let $(x_1, y_1), \ldots, (x_n, y_n)$ be a sample of points to which we wish to fit a model of the form $y = \alpha + \beta x$. Most of the nonparametric methods of estimating α and β can be interpreted in terms of the so-called *parameter space plot,* which is essentially a plot of the n lines $y_i = \alpha + \beta x_i$ $(1 \leq i \leq n)$ in $\alpha - \beta$ space. A detailed review of this topic is given in an unpublished manuscript by D.A. Griffiths and N.I. Fisher ('The parameter space plot in linear regression' (1982), available from the authors upon request); here, we shall confine ourselves to a brief description of the technique. For illustration, consider the parameter space plot given in Fig. 24, based on the data $(x, y) = (-2, 0), (-1, 1), (0, -2), (1, -2), (2, 1)$ (data used by Brown (1980)). The intersection of the lines $y_i = \alpha + \beta x_i$ and $y_j = \alpha + \beta x_j$ in the parameter space plot is the point $[(y_i x_j + y_j x_i)/(x_j - x_i), (y_j - y_i)/(x_j - x_i)]$. The two lines corresponding to the first two data points, and their intersection, are labelled in Fig. 24.

In general, the n lines divide $\alpha - \beta$ space into $N = \frac{1}{2}(n^2 + n + 2)$ regions $A_1, \ldots, A_N$, such that for any (α, β) in some region A, the pattern of signs of the residuals $y_i - \alpha - \beta x_i$ will be the same, with no two regions having the same pattern. Two typical regions A_i and A_j are highlighted in the example. The idea of developing inferential procedures based solely on these patterns dates back at least to Daniels (A1951, 1954), who used them to obtain a test, and associated confidence procedure, for the hypothesis that the regression parameters take specified values. Drummond (1976) and Quade (1979) have studied a general class of statistics which depend on the data only through these patterns of signs. A well-known special case of this is the median regression technique of Brown & Mood (see for example Mood (1950)), for which a different graphical method is given later in this section.

An early use of the parameter space plot was reported by Edgeworth (1923) in a paper on median estimates of regression; subsequent work on this aspect is given by Brown (1980). The fact that the intersections of the lines in $\alpha - \beta$ space are just the points $\{(y_j x_i - y_i x_j)/(x_i - x_j), (y_i - y_j)/(x_i - x_j)\}$ allows easy computation of the Theil–Sen estimator median $\{(y_i - y_j)/(x_i - x_j), 1 \leq i < j \leq n\}$ (see for example Sen, A1968), as these points may be enumerated in increasing order by moving a horizontal line from bottom to top across the plot. Similar estimators of α proposed by Maritz (A1979), for example the median of $\{(y_j x_i - y_i x_j)/(x_i - x_j), 1 \leq i < j \leq n\}$ can be computed by studying the α values in the parameter space plot.

Several papers make use of the parameter space plot to estimate the parameters of the Michaelis–Menton equation, a standard model in the field of enzyme kinetics. The model is itself nonlinear in the parameters but can be transformed and re-parameterised to a linear model. Various graphical methods for estimating the parameters of the linear model are discussed by Eisenthal & Cornish-Bowden (1974), Cornish-Bowden & Eisenthal (1978), Cornish-Bowden, Porter & Trager (A1978), and Cressie & Keightley (1979, 1981).

There are a few other graphical methods for simple linear regression. Shorack's method, see § 2.1(ii), of obtaining an estimate for the ratio of scale parameters is relevant here if a simple model of the form $y = \rho x, \rho > 0$, is being fitted. Hettmansperger & McKean (1974) give a graphical display of the equivalence of a hypothesis test for the slope parameter β in simple linear regression and a confidence interval estimate of β based on Kendall's tau (see for example Hollander & Wolfe 1973, Chapter 9).

We illustrate two techniques described by Daniel (1978), using the data in Table 16 on the relationship between newspaper sales and national income in the United States during 1930–1940. On a scatterplot of the data (see Fig. 25), mark in the medial line of the X's, and in the two groups so formed, identify the X medians (x_1, x_2) and the Y medians (y_1, y_2). Draw a line through the points (x_1, y_1) and (x_2, y_2) as a first approximation, then (using a transparent ruler) adjust the line if necessary so that the median deviation in each group is zero. The resulting line is the Brown–Mood median regression line mentioned above.

A related test of the hypothesis $\beta = \beta_0$ for the model of the form $y = \alpha + \beta x$ can be done using a similar diagram: with the X median drawn on the scatterplot, draw a line parallel to the line $y = \beta x$ which divides the points into two equal groups. (This is equivalent to determining the median $\hat{\alpha}$ of all the deviations $Y_i - \beta_0 X_i$ and drawing the line $y = \hat{\alpha} + \beta_0 x$.) Count the number N of points lying above this line and to the left of the X medial line. Under the hypothesis $\beta = \beta_0$, N is approximately binomially distributed $b(n, \frac{1}{2})$, so that $16(N - n/4)^2/n$ is approximately distributed as χ_1^2 for n not too small.

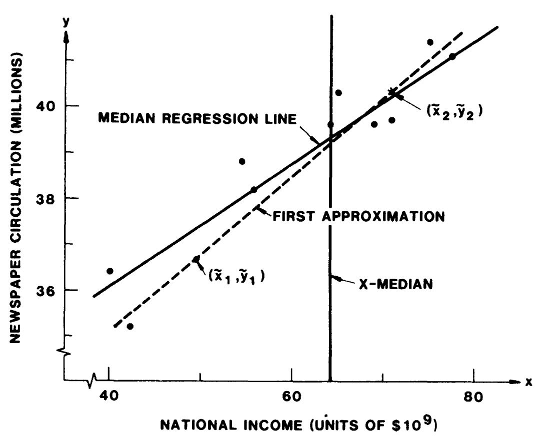

Figure 25. *Brown–Mood median regression line. Data: newspaper circulation and national income in the US,* 1930–1940 (*Table* 16).

Table 16

Newspaper circulation (in millions of copies) and national income (in billions of dollars) in the US during years 1930–1940

	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940
Circulation	39.6	38.8	36.4	35.2	36.7	38.2	40.3	41.4	39.6	39.7	41.1
National income	68.9	54.5	40.0	42.3	49.5	55.7	64.9	70.8	77.5	70.8	77.5

Source: Table 51, Ferber, R. (1949), *Statistical Techniques in Market Research*, New York; McGraw-Hill.

2.4 *Miscellaneous procedures*

In this section, we draw attention to relevant references in the Bibliography, without illustrating the methods.

(i) *k sample procedures.* Doksum (1977) gave an extension of the use of shift functions, § 2.1(iii), to models in which one population could be regarded as a control or reference population with which the others (treatments) were to be compared. This work has been extended by Nair (1978) to situations in which all k populations are treated equally, and to models in which the samples have been randomly censored. Nair has also considered the special case of the differences between populations being just location/scale, making possible more efficient large-sample estimation.

A k-sample analogue of the pair chart, called the k-multiple chart, has been developed by Wakimoto (1981). The chart can be used to obtain a k sample test (the 'area' test) of homogeneity against ordered alternatives, which is nearly equivalent to the usual Terpstra & Jonckheere test, but possibly superior.

Kraft & van Eeden (1968) made use of graphical procedures for the Kruskal–Wallis and Friedman homogeneity tests, with $k=3$ treatments. However, the labour required is considerable, and a certain amount of data reduction is required (even with small samples) before graphical techniques can be employed.

Some k sample analogues of the procedures discussed in §2.1(iv) for comparing survivor functions have been suggested by Kay (1977) and Kalbfleisch & Prentice (1980).

(ii) *Contingency tables.* Insofar as χ^2 tests for contingency tables are regarded as nonparametric procedures, it is appropriate to reference Snee (1974), who used a graphical display of a two-way table to facilitate identification of rows and/or columns contributing the nonhomogeneity (given an overall significant chi-squared statistic), and Boardman (1977), who used a graphical display to exhibit the contribution to the chi-squared statistic of each row-column combination in a two-way table. Modifications of these methods have been published by Cohen (1980).

(iii) *Analysis of covariance.* Quade (1983) has adapted the pair chart, see § 2.1(iii) to create a 'matched pair' chart for use in analysis of covariance.

(iv) *Angular data.* P–P and Q–Q plots, see §2.1(iii), can be adapted to comparison of two independent samples $\theta_1, \ldots, \theta_m$ and $\phi_1, \ldots, \phi_n$ of angular data, provided that each sample is measured modulo its sample mean direction before comparisons are made. If the samples are not too disperse, reasonable assessment of homogeneity is possible. Alternative comparisons can be based on $\cos\theta_1, \ldots, \cos\theta_m$ and $\cos\phi_1, \ldots, \cos\phi_n$ and $\sin\theta_1, \ldots, \sin\theta_m$ with $\sin\phi_1, \ldots, \sin\phi_n$. Similar remarks pertain to one-sample symmetry checks on an angular sample. Griffiths (1981) suggested a simple graphical device for finding the median of a sample of angular data.

(v) *Other multivariate procedures.* Mention has been made in § 2.1(iii) of the multivariate P–P plots due to Friedman & Rafsky (1979), and in §2.3(ii) of higher-dimensional corner tests. The multivariate plots are based on an earlier paper by Friedman & Tukey (A1974): more recently, Tukey & Tukey (1981) have given an extensive discussion of graphical displays for higher-dimensional data, which refers to much of the earlier work on the subject.

Andrews (1972) suggested mapping each multivariate data vector $\bar{x}=(x_1,\ldots,x_k)$ in a random sample of n k-vectors into a function

$$f_{\bar{x}}(t)=x_1\frac{1}{\sqrt{2}}+x_2\sin t+x_2\cos t+x_4\sin 2t+x_5\cos 2t+\ldots$$

on $[-\pi,\pi]$ and then plotting these n functions. More generally, $\bar{x}$ can be mapped into any function $\sum_i x_i\cdot u_i(t)$, where the sum is over $i=1,\ldots,n$, and $u_1,\ldots,u_k$ are orthonormal functions on, say, [0, 1]. Gnanadesikan (1977) describes how Andrews' plots can be used with large samples to study aspects of the shape (e.g. symmetry) and correlation structure of the underlying distribution, by looking at *quantile contour plots*, i.e. plots of quantiles of the n sample values of $f_{\bar{x}}(t)$ at each of a grid of values of t.

An important technique for determining structure in multivariate samples, namely nonmetric scaling (in which only the rank-orders of the 'distances' between individuals in the sample are used), has been reviewed recently by Carroll & Arabie (A1980).

(vi) *Time series.* Quenouille (1972) presented a rapid graphical method of computing a confidence interval for the median of a trend-free time series, and a rapid graphical test for monotone association between two time series.

3 General remarks

It is worth noting that most of the illustrations in this paper were hand-drafted with reasonable ease (although in somewhat less stylized form than that in which they appear here); hence, the methods of §§ 2.1(i), (ii), 2.2(iii) seem particularly useful for teaching purposes, being convenient methods of computation when a computer is not available. The useful correspondence between the two-sample location shift methods in § 2.1(i) and the one-sample centre of symmetry methods in § 2.2(i), § 2.2, serves to highlight the fact that understanding the way a given graphical method works leads to better insight into the behaviour of the statistical procedure it illustrates. The pair chart in § 2.1(iii) is an extremely versatile teaching aid, covering descriptive analysis and a variety of tests; to a lesser extent this also applies to the P plot, § 2.2(ii). For more advanced students, the P–P and Q–Q plots in § 2.1(iii) and the Q plots in § 2.2(ii) can be employed. The computational methods for Kendall's tau and Spearman's rho in §2.3(i), and for simple linear regression in § 2.3(ii) may well be of teaching value. Finally, the paper by Hettmansperger & McKean (1974) which is mentioned several times in the review, illustrates the relationship between hypothesis testing and interval estimation for numerous situations.

From a practical point of view, the most useful methods in § 2.1 are assuredly those discussed in § 2.1(iii): since most of the tests or estimates referred to in §§ 2.1(i), (ii) could be readily calculated on a computer they seem of lesser importance. (The exception of course is Tukey's quick two-sample test, § 2.1(i), for those emergencies on a train or bus when an on-the-spot assessment, or piece of sleight-of-hand, is required.) Similarly, in § 2.2 the methods of § 2.2(ii) offer more practical aid than those in § 2.2(i). In § 2.3(i), the Olmstead–Tukey corner test and some of Quenouille's rapid procedures can be useful at preliminary stages of data analysis when a scatterplot of the data is available; of far greater use, however, is the linked-vector method of Taguri *et al.* (1976). Whilst none of the procedures in § 2.4 has been illustrated, several of them are very powerful data-analytic tools; for example, the k sample comparisons using shift functions, nonmetric scaling and the methods for displaying multivariate data.

It may not be too unreasonable to claim that the range of nonparametric methods for which graphical procedures are available and being used almost defines the range of commonly-used nonparametric methods (except for some methods of analysing designed experiments such as randomized blocks layouts). Part of the reason for the scarcity of examples using nonparametric methods in multivariate analysis and multiple regression is the lack of computer packages necessary to cope with the formidable computations. Perhaps as methods develop which concern themselves more with the association structure of multivariate data, graphical procedures will follow: the few procedures now available provide encouragement for this hope. A specific area of need is that of directional statistics. A class of graphical procedures, analogous to probability plots and shift function plots for the comparison of two samples but of different kind, would be of great value. Other possibly fruitful areas are those of discriminant analysis (in which some nonparametric methodology is available) and time series modelling (in which such methodology exists only in embryo forms).

Acknowledgements

The author first encountered much of the material in lecture cources on Nonparametric Statistics given by D. Quade and P.K. Sen at the University of North Carolina at Chapel Hill, and in a series of seminars given by R. Gnanadesikan while visiting Sydney. These stimulating lectures and seminars are most gratefully acknowledged, as is the constructive criticism provided by the Editor and the referee on an earlier draft of the manuscript. The author also wishes to thank John Pemberton (University of Auckland) and Sue Clancy (CSIRO) for producing the computer-drawn figures, Joanna Salas and Jackie Ryan (CSIRO) for drafting the other figures, and Julie Scholte (CSIRO) and Sue Beskin for typing various drafts of the manuscript. The Bibliography was compiled during a six-month visit to the University of Auckland, supported by their Department of Mathematics and by the DSIR Applied Mathematics Department.

Bibliography

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6,** 701–726. [Graphical comparison of cumulative hazard functions for samples derived from general counting processes.]

Aalen, O.O., Borgan, O., Keiding, N. & Thormann, J. (1980). Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring. *Scand. J. Statist.* **7,** 161–171. [Graphical comparison of cumulative hazard functions for samples derived from general counting processes.]

Andrews, D.F. (1972). Plots of high-dimensional data. *Biometrics* **28,** 125–136. [Graphical representation of high-dimensional data by mapping each multivariate data point $\bar{x}$ into a function $f_x(t)$ of a single variable t (see text).]

Bhattacharyya, H.T. (1977). Nonparametric estimation of ratio of scale parameters. *Am. Statistician* **72,** 459–463. [Hodges–Lehmann estimates of ratio of scale parameters, based on Ansari–Bradley, Siegel–Tukey and Sukhatme tests.]

Billewicz, W.Z. (1955). A simple non-parametric graphical test for the significance of a difference. *Appl. Statist.* **4,** 97–102. ['Simple' multiple median regression.]

Boardman, T.J. (1977). Graphical contributions to the χ^2 statistic for two way tables. *Comm. Statist.* A **6,** 1437–1451. [Graphical display of contributions to the chi-squared statistic for each row-column combination in a two-way table.]

Bradley, J.V. (1963). Rank-order correlation scatter diagrams without plotting points. *Am. Statistician* **17,** 14–15. [Scatter diagram, using a set of computer cards, of sample rank order association between two variables.]

Brown, B.M. (1980). Median estimates in simple linear regression. *Aust. J. Statist.* **22,** 154–165. [Least-absolute-deviation estimate of regression line.]

Cohen, A. (1980). On the graphical display of the significant components in two-way contingency tables. *Comm. Statist.* A **9,** 1025–1041. [Clustering methods for comparing rows or columns, graphical display of cell contributions to chi-squared statistic.]

Conover, W.J. (1971). *Practical Nonparametric Statistics.* New York: Wiley. [Tukey's method for determining Wilcoxon estimate of centre of symmetry, Moses' method for determining Mann–Whitney–Wilcoxon estimate of location difference. (Some discussion of continuity assumption.) Discussion of Tukey's (1959) quick test, and some adaptations and extensions. Discussion of Olmstead–Tukey corner test of association.]

Cornish-Bowden, A. (1975). The use of the Direct Linear Plot for determining initial velocities *Biochem. J.* **149,** 305–312. [Use of the parameter space plot for simple linear regression, to estimate parameters of biochemical reaction process.]

Cornish-Bowden, A. & Eisenthal, R. (1974). Statistical considerations in the estimation of enzyme kinetic parameters by the Direct Linear Plot and other methods. *Biochem. J.* **139,** 721–730. [Use of parameter space plot for simple linear regression, to estimate parameters of Michaelis–Menton model.]

Cornish-Bowden, A. & Eisenthal, R. (1978). Estimation of Michaelis constant and maximum velocity from the direct linear plot. *Biochim. Biophys. Acta* **523,** 268–272. [Use of parameter space plot for simple linear regression, to estimate parameters of Michaelis–Menton model: improved use of direct linear plot.]

Cressie, N.A.C. & Keightley, D.D. (1979). The underlying structure of the direct linear plot with application to the analysis of hormone-receptor interactions. *J. Steroid Biochem.* **11,** 1173–1180. [Use of parameter space plot for simple linear regression, to estimate parameters of Michaelis–Menton model.]

Cressie, N.A.C. & Keightley, D.D. (1981). Analysing data from hormone-receptor assays. *Biometrics* **37,** 235–249. [Use of parameter space plot for simple linear regression, to estimate parameters of Michaelis–Menton model.]

Daniel, W.W. (1978). *Applied Nonparametric Statistics.* Boston: Houghton Mifflin. [Tukey's method for determining Wilcoxon estimate of centre of symmetry. Moses' method for determining Mann–Whitney–Wilcoxon estimate of location difference. Pair chart for comparing two samples. Olmstead–Tukey corner test for association. Graphical methods for Brown–Mood median regression.]

Daniels, H.E. (1954). A distribution-free test for regression parameters. *Ann. Math. Statist.* **25,** 499–513. [Use of the parameter space plot to calculate Daniels' statistic for testing simple linear regression.]

Doksum, K.A. (1974). Empirical probability plots and statistical inference for non-linear models in the two-sample case. *Ann. Statist.* **2,** 267–277. [Confidence bands for shift function $\Delta(x)$ for distribution difference between two populations.]

Doksum, K.A. (1975). Measures of location and asymmetry. *Scand. J. Statist.* **2,** 11–22. [Confidence bands for location parameter (function) $\Lambda(x)$ for distributional departure from symmetry.]

Doksum, K.A. (1977). Some graphical methods in statistics. A review and some extensions. *Statist. Neer.* **31,** 53–68. [Review of earlier work (Doksum, 1974; Doksum & Sievers, 1976; Wilk & Gnanadesikan, 1968) on pointwise and simultaneous confidence bands for shift function $\Delta(x)$ for comparing two populations. Extension of these concepts to comparison of several populations.]

Doksum, K.A., Fenstad, G. & Aaberge, T. (1977). Plots and tests for symmetry. *Biometrika* **64,** 473–487. [Simultaneous confidence bands for supposedly symmetric distribution function. Tests for symmetry.]

Doksum, K.A. & Sievers, G.L. (1976). Plotting with confidence: graphical comparisons of two populations. *Biometrika* **63,** 421–434. [Confidence bands for shift function $\Delta(x)$ for distribution difference between two populations.]

Drion, E.F. (1952). Some distribution-free tests for the difference between two empirical cumulative distribution functions. *Ann. Math. Statist.* **23,** 563–574. [Use of pair chart (A) to enumerate the distribution of two-sample Kolmogorov–Smirnov statistic (equal sample sizes), and (B) to obtain partial solution to problem of estimating the probability that one sample distribution function lies entirely above the other.]

Drummond, D.J. (1976). The cuts procedure for regression based solely on the signs of the residuals. Ph.D. Dissertation, North Carolina State University at Raleigh. [Use of the parameter space plot for simple linear regression.]

Edgeworth, F.Y. (1923). On the use of medians for reducing observations relating to several quantities. *London, Edinburgh and Dublin Phil. Mag. J. Sci.*, Series 6 **46,** 1074–1088. [Least-absolute-deviation estimate (median estimate) of linear relationship.]

Eisenthal, R. & Cornish-Bowden, A. (1974). The Direct Linear Plot. A new graphical procedure for estimating enzyme kinetic parameters. *Biochem. J.* **139,** 715–720. [Use of parameter space plot for simple linear regression, to estimate parameters of Michaelis–Menton model.]

Fienberg, S.E. (1979). Graphical methods in statistics. *Am. Statistician* **33,** 165–178. [Historical review of statistical graphics; discussion of some recent advances.]

Fisher, N.I. (1981). One-sample probability plots. *Aust. J. Statist.* **23,** 352–359. [Uses of P-plot for assessing symmetry and for computing Butler's statistic and the Wilcoxon signed-rank statistic.]

Fisher, R. (1977). Assumptions in the analysis of survival data. Ph.D. thesis, University of Melbourne. [H–H plots and other plots related to comparison of failure time distributions.]

Fisher, R. (1983). A graphical procedure for investigating the hazard ratio. *Biometrics.* To appear. [H–H plot for studying the hazard ratio.]

Friedman, J.H. & Rafsky, L.C. (1979). Fast algorithms for multivariate lining and planning. In *Proceedings of Computer Science and Statistics*: 12*th Annual Conference on the Interface*, Ed. J. Gentleman, pp. 124–136. University of Waterloo. [Multivariate two-sample P–P plots.]

Friedman, J.H. & Rafsky, L.C. (1981). Graphics for the multivariate two-sample problem (with discussion). *J. Am. Statist. Assoc.* **76,** 277–295. [Multivariate two-sample P–P plots, multivariate two-sample runs test, using minimal spanning trees.]

Gerson, M. (1975). The techniques and uses of probability plotting. *Statistician* **24**, 235–257. [Discussion of P–P and Q–Q plots.]
Gibbons, J.D. (1971). *Nonparametric Statistical Inference*. New York: McGraw-Hill. [Moses' method for determining Mann–Whitney–Wilcoxon estimator of location shift. Use of pair chart to evaluate probability level of two-sample Kolmogorov–Smirnov statistic.]
Gnanadesikan, R. (1977). *Methods for Statistical Analysis of Multivariate Data*. New York: Wiley. [Description of P–P and Q–Q plots. Description of Andrews plot and application to large-sample situations.]
Griffin, H.D. (1958). Graphic computation of tau as a coefficient of disarray. *J. Am. Statist. Assoc.* **53,** 441–446. [Computational device for Kendall's tau.]
Griffiths, D.A. (1981). Delivery costs minimised. *Math. Spectrum* **14,** 26. [Graphical determination of the median of an angular sample.]
Hettmansperger, T.P. & McKean, J.W. (1974). A graphical representation for non-parametric inference. *Am. Statistican* **28,** 100–102. [Graphs illustrating relationship between hypothesis testing and interval estimation for (I) one-sample Wilcoxon test (II) two-sample Mann–Whitney–Wilcoxon test (III) test for simple linear regression based on Kendall's tau (Sen's method).]
Hodges, J.L. (1958). The significance probability of the Smirnov two-sample test. *Arkiv. Mat.* **3,** 469–486. [Use of pair chart to calculate two-sample Kolmogorov–Smirnov statistic and its probability level.]
Hodges, J.L. & Lehmann, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist* **34,** 598–611. [Calculation of Mann–Whitney–Wilcoxon estimate of location shift using sliding papers.]
Hollander, M. & Wolfe, D.A. (1973). *Nonparametric Statistical Methods*. New York: Wiley. [Tukey's method for determining Mann–Whitney–Wilcoxon estimate of centre of symmetry; Moses' method for determining Mann–Whitney–Wilcoxon estimate of location difference. Method for determining Shorack's estimate of ratio of scale parameters.]
Hoyland, A. (1964). Numerical evaluation of Hodges–Lehmann estimates. *Det Kongelige Norske Videnskabers Selskabs Forhandlinger* **37,** 42–47. [Tabular method for determining Mann–Whitney–Wilcoxon estimate of centre of symmetry and Mann–Whitney–Wilcoxon estimate of location difference.]
Jaeckel, L.A. (1969). Robust estimates of location. Ph.D. thesis, Department of Statistics, University of California at Berkeley. [General estimate of centre of symmetry, based on midranges.]
Kalbfleisch, J.D. & Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley. [Plots of hazard function and survival function, particularly in context of Cox model for survival data.]
Kay, R. (1977). Proportional hazard regression models and the analysis of censored survival data. *Appl. Statist.* **26,** 227–237. [Graphical method of assessing contribution of a covariate in Cox model for survival data.]
Klotz, J.H. (1966). The Wilcoxon, ties, and the computer. *J. Am. Statist. Assoc.* **61,** 772–787. [Use of pair chart to enumerate distribution of Mann–Whitney–Wilcoxon statistic in presence of ties.]
Kraft, C.H. & van Eeden, C. (1968). *A Non-parametric Introduction to Statistics*. New York: Macmillan. [Procedure to calculate Kruskal–Wallis and Friedman test statistics, for three treatemnts.]
Lagakos, S.W. (1981). The graphical evaluation of explanatory variables in proportional hazard regression models. *Biometrika* **68,** 93–98. [Plots of hazard function to check on explanatory variables in Cox model for survival data.]
Lehmann, E.L. (1975). *Nonparametrics: Statistical Methods based on Ranks*. San Francisco: Holden-Day. [Hoyland's methods for determining Wilcoxon estimate of centre of symmetry and Mann–Whitney–Wilcoxon estimate of location shift.]
Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *J. Am. Statist. Assoc.*, New Series, No. 70, 209–219. [Use of Q–Q plots to compare two populations.]
Mood, A.M. (1950). *Introduction to the Theory of Statistics*. New York: McGraw-Hill. [Olmstead–Tukey corner test; median estimate of regression and associated test.]
Moses, L.E. (1953). Nonparametric methods. *Statistical Inference*, 1st edition, Ed. H.M. Walker and J. Lev, Chapter 18. New York: Holt, Rienhart and Winston. [Tukey's method for determining Wilcoxon estimate of centre of symmetry. Moses' method for determining Mann–Whitney–Wilcoxon estimate of location shift.]
Moses, L.E. (1965). Reply to Query 10: Confidence limits from rank tests. *Technometrics* **7,** 257–260. [Tukey's method for determining Wilcoxon estimate of centre of symmetry; Moses' method for determining Mann–Whitney–Wilcoxon estimate of location shift.]
Nair, V.N. (1978). Graphical comparisons of populations in some non-linear models. Ph.D. thesis. Department of Statistics, University of California at Berkeley. [Confidence bands for shift functions $\Delta_{ij}(X)$ for distribution differences between several populations.]
Nelson, W. (1970). Hazard plotting methods for analysis of life data with different failure modes. *J. Qual. Technol.* **2,** 126–149. [Construction of cumulative hazard function plots; use in adjusting data sets for differing models of failure, before comparison.]
Nelson, W. (1972). Theory and application of hazard plotting for censored failure data. *Technometrics* **14,** 945–966. [Construction of cumulative hazard function plots; comparison of two samples via their sample cumulative hazard functions.]
Noether, G.E. (1971). *Introduction to Statistics* (2nd edition, 1976). Boston: Houghton Mifflin. [Methods for determining Wilcoxon estimate of centre of symmetry and Mann–Whitney–Wilcoxon estimate of location shift. Discussion of effect of ties in the data.]
Olmstead, P.S. & Tukey, J.W. (1947). A corner test for association. *Ann. Math. Statist.* **18,** 495–513. [Method for calculating a nonparametric measure of association which gives considerable weight to any association of peripheral points.]

Porter, W.R. & Trager, W.F. (1977). Improved nonparametric statistical methods for the estimation of Michaelis–Menton kinetic parameters by the direct linear plot. *Biochem. J.* **161,** 293–302. [Use of parameter space plot for simple linear regression, to estimate parameters of the Michaelis–Menton model.]

Quade, D. (1973). The pair chart. *Statist. Neer.* **27,** 29–45. [Review of applications of pair chart for use in two-sample problems: descriptive uses; calculation of Kolmogorov–Smirnov statistic and evaluation of its probability level; calculation of Wald–Wolfowitz runs statistic, Mann–Whitney–Wilcoxon statistic, Sukhatme's statistic for scale difference, Ansari–Bradley scale test, Mood–Steffens squared-rank test, Crouse–Steffens rest for scale difference, Lehmann's test against all alternatives.]

Quade, D. (1979). Regression analysis based on the signs of the residuals. *J. Am. Statist. Assoc.* **74,** 411–423. [Use of the parameter space plot for simple linear regression.]

Quade, D. (1983). Nonparametric analysis of covariance by matching. *Biometrics.* To appear. [Pair chart for nonparametric analysis of covariance.]

Quenouille, M.H. (1952). *Associated Measurements.* London: Butterworth. [Methods of assessing association for random samples of bivariate data.]

Quenouille, M.H. (1972). *Rapid Statistical Calculations,* 2nd edition. London: Griffin. [Methods of assessing association for random samples of bivariate data and for time series. Several methods of estimating a linear regression. Confidence limits for median of trend-free time series.]

Sandelius, M. (1968). A graphical version of Tukey's confidence interval for slippage. *Technometrics* **10,** 193–194. [Calculation of confidence interval based on Tukey's (A1959) test for location shift, using sliding papers.]

Sandiford, P. (1929). *Educational Psychology.* London: Longman, Green. [Graphical representation of rank correlation in terms of disarray.]

Shah, S.M. (1961). A note on Griffin's paper 'Graphic computation of tau as a coefficient of disarray. *J. Am. Statist. Assoc.* **56,** 736. [Simplification of Griffin's method when ties are present in the data.]

Shahani, A.K. (1969). A simple graphical test of association for large samples. *Appl. Statist.* **18,** 185–190. [Method for testing association for bivariate sample based on proportions of data in small corners of plane.]

Shorack, G.R. (1966). Graphical procedures for using distribution-free methods in the estimation of relative potency in dilution (−direct) assay. *Biometrics* **22,** 610–619. [Method for obtaining confidence interval and point estimate of ratio of scale parameters, variables not necessarily non-negative.]

Snee, R.D. (1974). Graphical display of two-way contingency tables. *Am. Statistician* **28,** 9–12. [Representation of two-way table to facilitate identification of those rows and/or columns contributing non-homogeneity (given a significant overall chi-squared statistic).]

Switzer, P. (1976). Confidence procedures for two-sample problems. *Biometrika* **63,** 13–25. [Confidence bands for shift function $\Delta(x)$ for distribution difference between two populations.]

Symonds, P.M. (1927). *Measurement in Secondary Education.* New York: Macmillan. [Graphical representation of correlation in terms of disarray.]

Taguri, M., Hiramatsu, M., Kittaka, T. & Wakimoto, K. (1976). Graphical representation of correlation analysis of ordered data by linked vector pattern. *J. Jap. Statist. Soc.* **6,** 17–25. [Simultaneous display of rank order associations of an objective variable with each of several explanatory variables, using linked vector patterns.]

Tukey, P.A. & Tukey, J.W. (1981). Chapters 10–12 in *Interpreting Multivariate Data,* Ed. V. Barnett, pp. 189–275. Chichester: Wiley. [Graphical methods for examining structure of multivatiate data.]

Wakimoto, K. (1981). k-multiple chart and its application to the test for homogeneity against ordered alternatives. *J. Jap. Statist. Soc.* **11,** 1–7. [k-sample extension of pair chart; area test statistic for testing homogeneity against ordered alternatives.]

Wilk, M.B. & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika* **55,** 1–17. [Application of percent (P–P) plots and quantile (Q–Q) plots to one-sample and two-sample problems.]

Auxiliary references (labelled A in text, e.g. David, A1981)

Carroll, J.D. & Arabie, P. (1980). Multidimensional scaling. *Annual Rev. Psychol.* **31,** 607–649.

Cornish-Bowden, A., Porter, W.R. & Trager, W.F. (1978). Evaluation of distribution-free confidence limits for enzyme kinematic parameters. *J. Theor. Biol.* **74,** 163–175.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc.* B **34,** 187–220.

Cox, D.R. & Snell, E.J. (1968). A general definition of residuals (with discussion). *J. R. Statist. Soc.* B **30,** 248–275.

Daniels, H.E. (1951). The theory of position finding (with discussion). *J. R. Statist. Soc.* B **13,** 186–207.

David, H.A. (1981). *Order Statistics,* 2nd edition, New York: Wiley.

Fowlkes, E.B. (1979). Some methods for studying the mixture of two normal (lognormal) distributions. *J. Am. Statist. Assoc.* **74,** 561–575.

Friedman, J.H. & Rafsky, L.C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7,** 697–717.

Friedman, J.H. & Tukey, J.W. (1974). A projection persuit algorithm for exploratory data analysis. *I.E.E.E. Trans. Computers* C-**23,** 881–890.

Gnanadesikan, M. & Gupta, S.S. (1970). A selection procedure for multivariate normal distributions in terms of generalized variances *Technometrics* **12,** 103–117.
Gordon, A.D. (1979a). A measure of agreement between rankings. *Biometrika* **66,** 7–15.
Gordon, A.D. (1979b). Another measure of the agreement between rankings. *Biometrika* **66,** 327–332.
Hazen, A. (1914), 'Storage to be provided in impounding reservoirs for municipal water supply. *Trans. Am. Soc. Civ. Eng.* **77,** 1539–1669.
Hazen, A. (1930). *Flood Flows: A study of Frequencies and Magnitudes.* New York: Columbia University Press.
Kendall, M.G. (1975). *Rank Correlation Methods,* 4th edition. London: Griffin.
Maritz, J.S. (1979). On Theil's method in distribution-free regression. *Aust. J. Statist.* **21,** 30–35.
Neave, H.R. (1966). A development of Tukey's quick test of location. *J. Am. Statist. Assoc.* **61,** 949–964.
Nelson, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Technol.* **1,** 27–52.
Rodrigues, O. (1839). Notes sur les inversions, ou derangements produits dans les permutations. *Liouville J. Math. Pures Appliquées* **4,** 236–240.
Rosenbaum, S. (1965). On some two-sample nonparametric tests. *J. Am. Statist. Assoc.* **60,** 1111–1126.
Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau, *J. Am. Statist. Assoc.* **63,** 1379–1389.
Tukey, J.W. (1959). A quick, compact, two-sample test to Duckworth's specifications. *Technometrics* **1,** 31–48.
Tukey, J.W. (1977). *Exploratory Data Analysis.* Reading, Mass.: Addison–Wesley.

Résumé

Cette article fait le tour des méthodes graphiques utilisable en statistique nonparamétrique, et donne des exemples d'utilisation de beaucoup de ces méthodes dans les rubriques suivantes: méthodes pour comparer deux populations, méthodes relatif une population, méthodes d'association et régression, méthodes mixtes.

[*Paper received December* 1980, *revised March* 1982]