Biometrika (1968), 55, 1, p. 1
With 23 text-figures
Printed in Great Britain

Probability plotting methods for the analysis of data

By M. B. Wilk and R. Gnanadesikan
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

Summary

This paper describes and discusses graphical techniques, based on the primitive empirical cumulative distribution function and on quantile (Q–Q) plots, percent (P–P) plots and hybrids of these, which are useful in assessing a one-dimensional sample, either from original data or resulting from analysis. Areas of application include: the comparison of samples; the comparison of distributions; the presentation of results on sensitivities of statistical methods; the analysis of collections of contrasts and of collections of sample variances; the assessment of multivariate contrasts; and the structuring of analysis of variance mean squares. Many of the objectives and techniques are illustrated by examples.

1. Introduction

This paper reviews a variety of old and new statistical techniques based on the cumulative distribution function and its ramifications. Included in the coverage are applications, for various situations and purposes, of quantile probability plots (Q–Q plots), percentage probability plots (P–P plots) and extensions and hybrids of these. The general viewpoint is that of analysis of data by statistical methods that are suggestive and constructive rather than formal procedures to be applied in the light of a tightly specified mathematical model. The technological background is taken to be current capacities in data collection and high-speed computing systems, including graphical display facilities.

It is very often useful in statistical data analysis to examine and to present a body of data as though it may have originated as a one-dimensional sample, i.e. data which one wishes to treat for purposes of analysis, as an unstructured array. Sometimes this is applicable to 'original' data; even more often such a viewpoint is useful with 'derived' data, e.g. residuals from a model fitted to the data.

The empirical cumulative distribution function and probability plotting methods have a key role in the statistical treatment of one-dimensional samples, being of relevance for summarization and palatable description as well as for exposure and inference.

2. The empirical cumulative distribution function (e.c.d.f.) as a statistical method

Extensive attention has been given to the empirical cumulative distribution function; see, for example, the bibliography of Barton & Mallows (1965). The present discussion is concerned, however, with descriptive uses of the e.c.d.f. rather than formal tests.

The discussion in this section is addressed directly to the case of a one-dimensional sample, an example of which is replicate observations on tensile strength of specimens of a given metal. However, many phases of the analysis of complex situations involve the appreciation of one-dimensional characterizations. For example, with data on the energies in a spoken word associated with a cross-classification according to time and frequency, it will often be of interest, in one treatment of the data, to ignore the associated structure and to regard the observations as though they were a one-dimensional sample. Such data are in fact used in this way in an example to be presented later.

For one-dimensional samples, the empirical cumulative distribution function (e.c.d.f.), i.e. a plot of the ith ordered value as ordinate against $(i-\frac{1}{2})/n$ as abscissa, provides an exhaustive representation of the data, under the following broad assumptions: (i) that the order of the observations is immaterial; (ii) that there is no classification of the observations, based on extraneous considerations, which one wishes to employ; and (iii) if the sample is non-random, then appropriate weights are specified. [J. W. Tukey points out that this is really the empirical inverse of the c.d.f. and suggests the term 'empirical representing function' for it.]

The use of the e.c.d.f. does not depend on any assumption of a parametric distributional specification. It may usefully describe data even when random sampling is not involved. Furthermore, the e.c.d.f. has additional advantages, including:

(i) It is invariant under monotone transformation, in the sense of quantiles (but not, of course, in appearance).

(ii) It lends itself to graphical representation.

(iii) The complexity of the graph is essentially independent of the number of observations.

(iv) It can be used directly and valuably in connexion with censored samples.

(v) It is a robust carrier of information on location, spread and shape, and an effective indicator of peculiarities.

(vi) It lends itself very well to condensation and to interpolation and smoothing.

(vii) It does not involve the 'grouping' difficulties that arise in using a histogram.

(viii) It is directly associated with the body of statistical methods which may be referred to as probability plotting procedures, a discussion of which constitutes a major part of the sequel.

The above discussion is not meant to imply that all techniques and concepts can or should be viewed in terms of the e.c.d.f. For instance, the direct merging of two e.c.d.f.s poses some possibly awkward operational questions. Also, it is not at all clear that, in terms of graphing and conceptualization, the e.c.d.f. has a productive multivariate generalization.

3. THE USE OF THE e.c.d.f. IN STATISTICAL DESCRIPTION

Some useful features of the e.c.d.f. are now illustrated by two examples. Table 1 summarizes the computer print-out of ordered log energies associated with a specific utterance of a given word by a particular talker. These data, and much of the other illustrative data used in this paper, are taken from unpublished Bell Telephone Laboratories memoranda, this one by Becker *et al.* Though the individual energies are in fact associated with identified time and frequency bands, for the present purpose this cross-classified identification is ignored and the data regarded as a one-dimensional sample.

Summary statistics in the form of mean, standard deviation, and standardized third and fourth moments are also given in Table 1. Another statistical presentation of the entire body of these data is provided by Fig. 1, which is the e.c.d.f.

Clearly Fig. 1 is a more palatable form of data presentation than is the original print-out. But even more important, it constitutes not only a summary but also an analysis in the sense that various properties of the data may be studied in Fig. 1 which are not as directly available in the print-out. Thus, Fig. 1 indicates the data to be reasonably smooth while clearly displaying the degree and location of quantization. There is no indication of wild observations but it is apparent from Fig. 1 that the data are bimodal, with a major mode at the lower end and a minor mode at about the 85 % point. The complexity of Fig. 1 is in no way increased because of the large number of observations involved and the plot provides a robust basis for choosing arithmetic summary statistics by indicating skewness, bimodality, outliers, etc., in addition to giving location and scale information.

Table 1. *Talker identifications data (log energies)*

[Unpublished report of Becker, Gnanadesikan, Mathews, Pinkham, Pruzansky, & Wilk; Word 8, Utterance 1, Speaker 1.]

1·3863$^{(56)}$, 1·7918$^{(69)}$, 2·0794$^{(62)}$, 2·1972$^{(2)}$, 2·3026$^{(43)}$, 2·3979$^{(3)}$, 2·4849$^{(17)}$, 2·5649, 2·6391$^{(15)}$, 2·7726$^{(9)}$, 2·8332$^{(2)}$, 2·8904$^{(5)}$, 2·9957$^{(5)}$, 3·0445$^{(2)}$, 3·0910$^{(5)}$, 3·2581$^{(5)}$, 3·2958, 3·3322$^{(2)}$, 3·4657$^{(2)}$

3·5264	3·5553	3·5835	3·5835	3·6889	3·7377	3·7377	3·8712	3·9890
4·0943	4·1271	4·1271	4·1589	4·1589	4·2195	4·2195	4·3041	4·3041
4·3041	4·4308	4·4308	4·4543	4·4543	4·4773	4·5433	4·6052	4·6250
4·6634	4·7536	4·7707	4·8363	4·8520	4·8828	4·8828	4·9127	4·9698
4·9972	5·0876	5·0999	5·0999	5·1358	5·1818	5·2679	5·3471	5·3753
5·3845	5·4027	5·4293	5·4381	5·4553	5·5134	5·5294	5·5607	5·5683
5·5683	5·5835	5·6490	5·6560	5·6904	5·7494	5·7746	5·7869	5·7869
5·7991	5·7991	5·8111	5·8171	5·8348	5·8464	5·8579	5·8693	5·8749
5·8749	5·8805	5·9402	5·9506	5·9506	5·9610	5·9610	5·9661	5·9661
5·9814	5·9915	6·0014	6·0210	6·0259	6·0544	6·0544	6·0776	6·0776
6·1048	6·1137	6·1269	6·1485	6·1527	6·1527	6·1527	6·1654	6·2383
6·2383	6·2461	6·2461	6·2500	6·2577	6·2577	6·2653	6·2766	6·2916
6·3063	6·3244	6·3456	6·3630	6·3902	6·3902	6·4135	6·4167	6·4552
6·4615	6·4677	6·4677	6·5338	6·5338	6·5453	6·5568	6·5624	6·5681
6·5875	6·5875	6·6039	6·6771	6·6896	6·6970	6·7093	6·7117	6·7452
6·7616	6·7776	6·7845	6·8024	6·8373	6·8773	6·9527	6·9754	7·0934
7·1357	7·1515	7·2471	7·2752	7·3185	7·3291	7·3727	7·4933	7·5110
7·5380	7·5380							

Mean	3·3248	β_1	0·636
St. dev.	1·8690	β_2	2·090

The computer print-out listed all values in increasing order and began with 56 values all equal to 1·3863. To save space the first part of the table has been coded in an obvious way.

Other modes of statistical summary might, of course, be employed. The e.c.d.f. is not claimed to be better, in every way, in all cases, than alternate modes such as a simple tabulation, or a frequency table, or a histogram, or a list of a few key quantiles, or a set of moment-statistics. But for general use, especially in a preliminary mode, the e.c.d.f. by-passes several ambiguities (e.g. of grouping) and combines convenience of condensation with substantial exposure power.

A second example, involving 1000 observations, is given in Fig. 2, which is the e.c.d.f. for the data on signal strength in log volts in a study of mobile radio transmission. (These data were supplied by Freeny, Pinkham & Wilk). While these data were in fact a space series, it is also of direct interest to treat them without this auxiliary identification. The e.c.d.f. of Fig. 2 is nicely behaved and quick indications of location and scale are available. It is evident that a good condensation for this e.c.d.f. is given by the quantiles corresponding to about $50p$ values [0(1) 10(2) 20(5) 80(2) 90(1) 100] so chosen that linear interpolation is adequate elsewhere.

The e.c.d.f. in Fig. 2 indicates that the distribution is somewhat skew but shows no outliers. In this application, it is the behaviour at the lower end of the distribution which is of particular interest. A more sensitive description for studying asymmetry in the mobile radio data is provided by Fig. 3 which is a plot of the upper half of the ordered data against the lower, i.e. a plot of y_{1000} versus y_1, y_{999} versus y_2, etc. Symmetry would be indicated by a

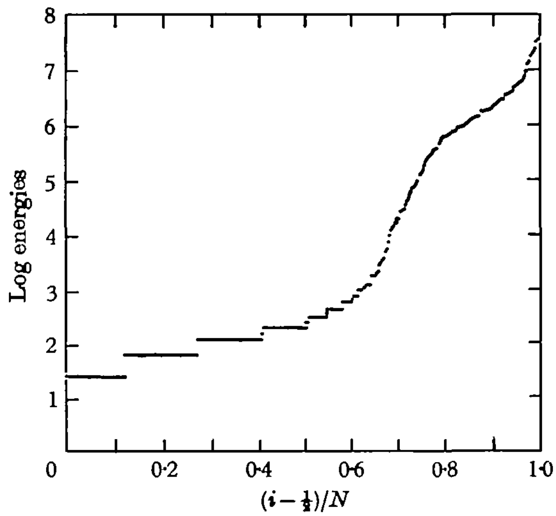

Fig. 1. e.c.d.f. for data of Table 1.

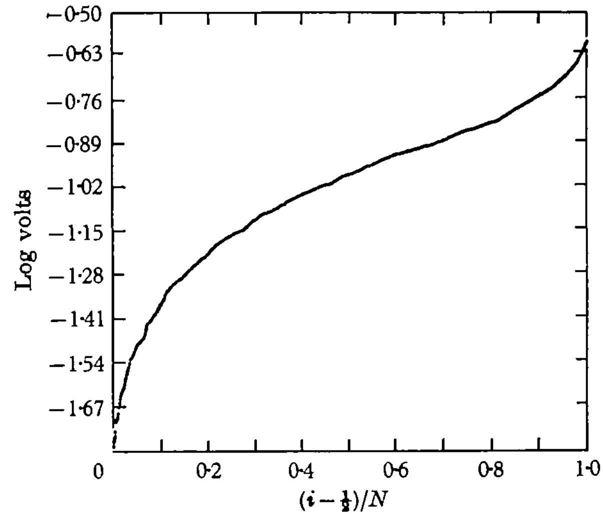

Fig. 2. e.c.d.f. for mobile radio data.

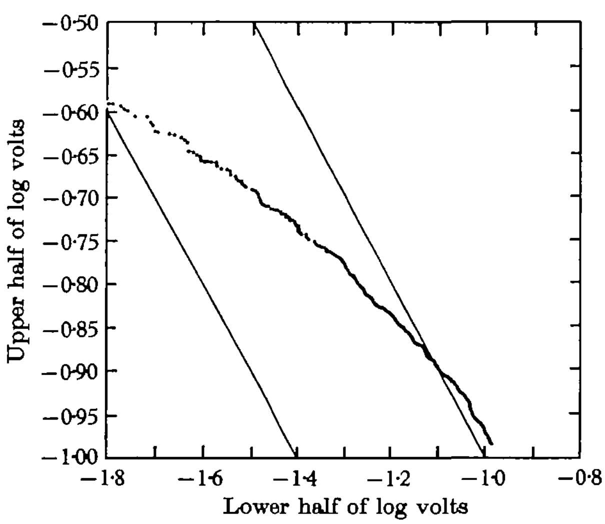

Fig. 3. Symmetry check for e.c.d.f. of Fig. 2.

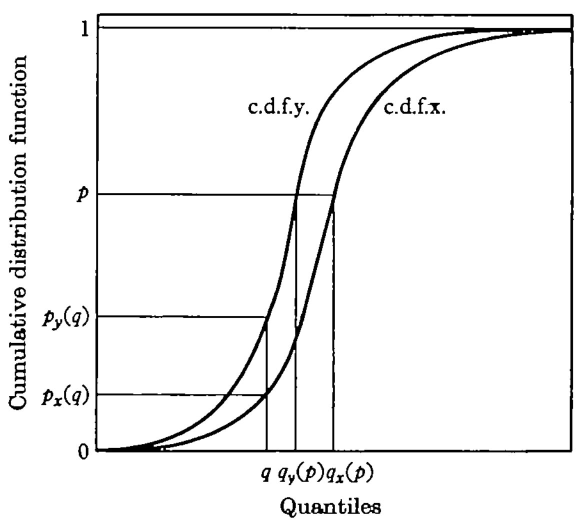

Fig. 4. Illustration for P–P and Q–Q plots.

straight line configuration with negative unit slope (J. W. Tukey has suggested the alternative scheme of plotting the sums, $y_{n-i+1}+y_i$, against the differences, $y_{n-i+1}-y_i$, which would yield a horizontal configuration for a symmetric distribution). In Fig. 3, two straight lines of negative unit slope are shown to aid the assessment. The contra-indication in the plot reflects a negative skewness, i.e. the lower percentage points are farther from the median than are the upper.

4. Comparison of two one-dimensional samples

4·1. *General comments*

For purposes of present discussion, it is convenient not to distinguish between finite and infinite samples, i.e. between empirical and theoretical c.d.f.'s. After all, either an empirical or a theoretical c.d.f. may be characterized as closely as one may desire by a set of its quantiles.

The discussion of this section is applicable to the comparison of two e.c.d.f.'s (i.e. of two finite samples), to the comparison of an e.c.d.f. with a theoretical c.d.f. (possibly standardized) and also to the comparison of two theoretical c.d.f.'s.

Two basic kinds of probability plots are discussed, namely quantile versus quantile plots (Q–Q plots) and percent versus percent plots (P–P plots). These and various hybrids and modifications may be defined in terms of the c.d.f.'s of the two samples.

Figure 4 shows two c.d.f.'s and illustrates the definition of P–P and Q–Q plots. Corresponding to any ordinate value p there are two quantile values $q_x(p)$ and $q_y(p)$. A Q–Q plot of the x and y samples is then just a scatter plot of $q_y(p)$ versus $q_x(p)$ for various p. Corresponding to any abscissa value q, there are two c.d.f. values $p_x(q)$ and $p_y(q)$. A P–P plot of the x and y samples is then just a scatter plot of $p_y(q)$ versus $p_x(q)$ for various q. Various modifications and mixtures of these basic forms are possible and examples of some will be given.

In one special case P–P and Q–Q plots are in fact identical, namely when the two variables are both uniform on $(0, 1)$. This may occur, for instance, when both the x and y variables are probability integral transforms.

4·2. *Quantile plots* (Q–Q *plots*)

If x and y are identically distributed variables, then the plot of x-quantiles versus y-quantiles will of course be a straight line configuration with slope 1, pointed towards the origin. An elementary property of Q–Q plots is that if y is a linear function of x then the corresponding Q–Q plot will still be linear but with possibly changed location and slope. It is this linear invariance property which has made the use of Q–Q plots appealing and valuable, mainly because linearity seems to be a geometric configuration which humans are able to perceive most easily. Moreover, departures from linearity are sensitively appreciated.

For the case in which the variables have long tails (unlimited range) the Q–Q plot tends to emphasize the comparative structure in the tails and to blur the distinctions in the 'middle' where the densities are high. The reason for this is that the quantile is a rapidly changing function of p when the density is sparse (in the tails) and a slowly changing function of p where the density is high (in the middle).

An example of a quantile plot for comparing two distributions is given in Fig. 5 which shows the quantiles of the double exponential distribution plotted against those of the normal (taken from unpublished work of Blake & Fowlkes). This plot shows a smooth nonlinearity everywhere, indicating general departure of the double exponential from the normal. The 'thicker tails' of the double exponential are shown by the slope of the curve exceeding 1, increasingly, at large absolute quantile values. The greater concentration at the centre of the double exponential relative to the normal is shown by the slope being less than 1 in the central region.

This Q–Q plot is representative of others in providing a very sensitive indicator of distributional discrepancies and provides a useful basis for examining the adequacy of composite hypotheses, where the unspecified parameters are location and/or scale.

Figure 6 presents the use of a Q–Q plot as one way of gauging the adequacy of the approximation of the distribution of a standardized log χ^2_{10} by a standard normal variable. This is

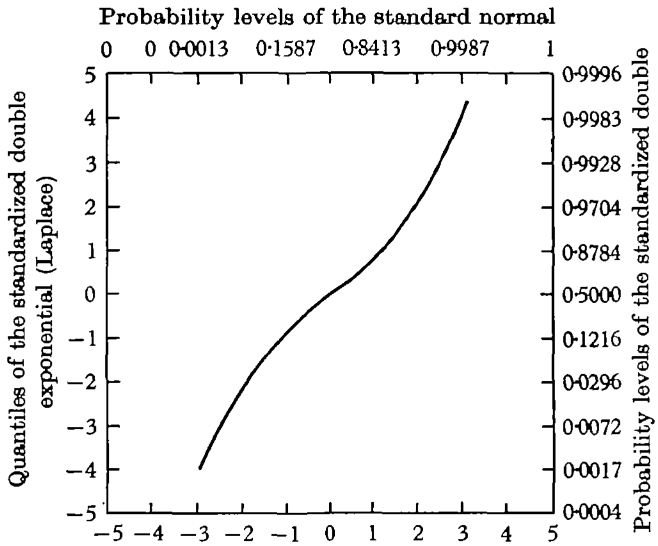

Fig. 5. Q–Q plot of Laplace *vs.* normal.

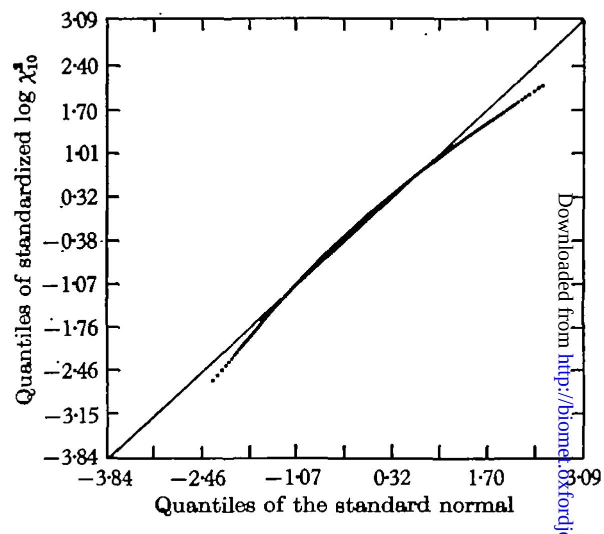

Fig. 6. Q–Q plot of standardized log $\chi^2_{(10)}$ *vs.* normal.

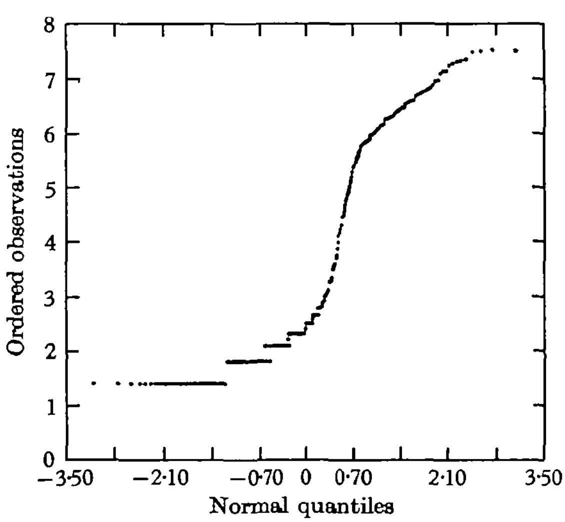

Fig. 7. Normal Q–Q plot for data of Table 1.

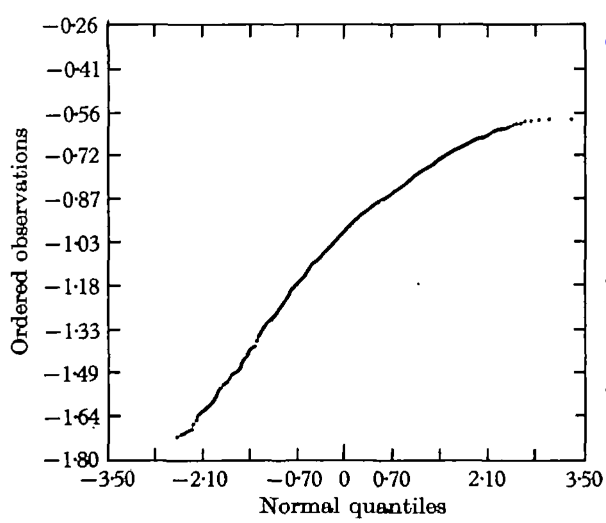

Fig. 8. Normal Q–Q plot of log volts for mobile radio data.

taken from unpublished work of M. Gnanadesikan. The comparison of the configuration of the plotted points with the 45° line indicates the systematic error in the tails, shows a small but definite asymmetry in the approximation and suggests the region over which the quantile approximation may be regarded as adequate. Other graphical procedures for evaluating this approximation are discussed below.

Figure 7 presents a Q–Q plot for the talker energy data against standard normal quantiles. The data are obviously very non-normal.

Figure 8 shows a normal Q–Q plot for the mobile radio data on the log scale. The departures from normality are clearly evident, the actual data having shorter tails than the normal as well as being somewhat asymmetric; Fig. 9 shows a Q–Q plot of this mobile radio data on

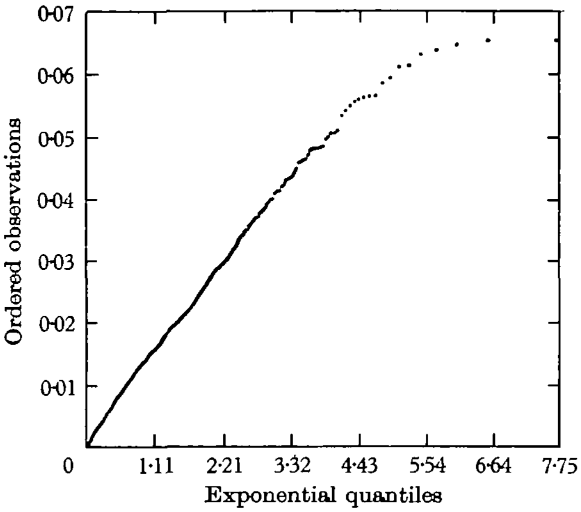

Fig. 9. Exponential Q–Q plot of (volts)^2 for mobile radio data.

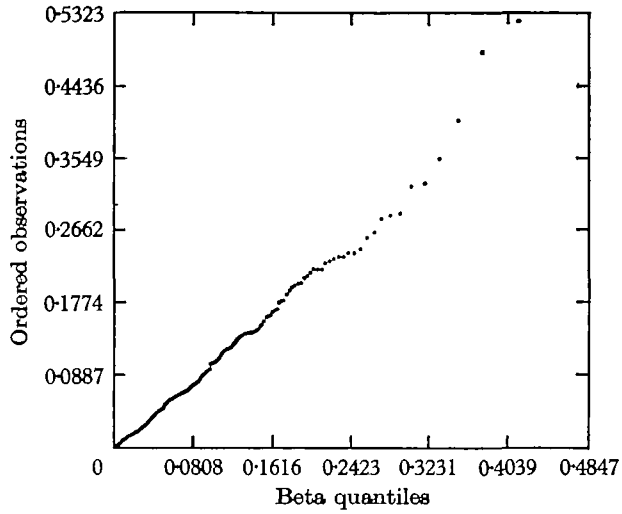

Fig. 10. Beta Q–Q plot of Monte Carlo values of eigenvalue function.

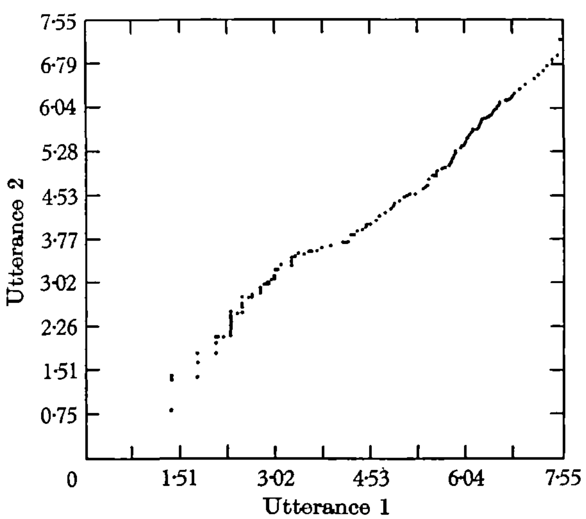

Fig. 11. Q–Q plot of e.c.d.f.'s of utterances 1 and 2 of talker 1.

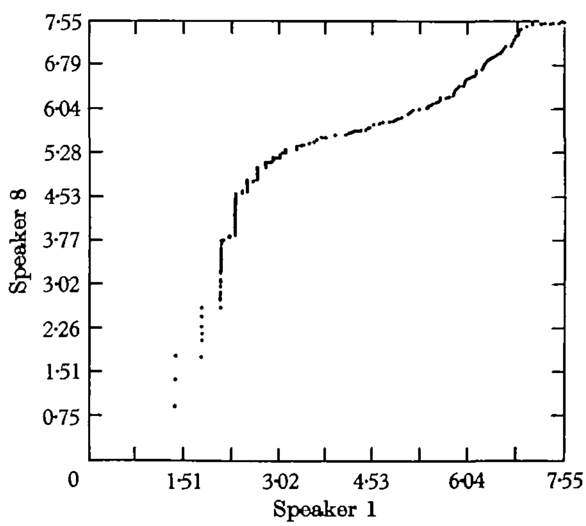

Fig. 12. Q–Q plot of e.c.d.f.'s of same word by talkers 1 and 8.

a (volts)^2 scale against the quantiles of an exponential distribution (i.e., gamma with shape parameter equal to 1). Though some theoretical considerations suggest that such a distribution might hold, it is apparent that the data do not so conform, at least in the tails.

Another example is shown in Fig. 10 which is a Q–Q plot of order statistics of a sample of observations on a function of eigenvalues against quantiles of a fitted beta distribution. The sample values are derived from a random sampling experiment concerned with the relative sensitivities of some multivariate tests (Gnanadesikan *et al.* 1965). The plot indicates that, with the beta parameters estimated from the data by maximum likelihood (Gnanadesikan, Pinkham & Hughes, 1967), the fitted distribution is reasonably good, except in the extreme right tail.

Further applications of Q–Q plots are shown in Figs. 11 and 12. These deal with the talker energy data and each shows a plot of the empirical quantiles of one sample against those of another. Figure 11 involves two utterances of the same word by the same talker. Figure 12 involves utterances of the same word by two different talkers. In comparison of these, it is evident that the first is markedly more regular and linear than is the second. Clearly, these plots are relevant to the problem of discrimination amongst talkers.

4·3. *Percent plots* (P–P *plots*)

If x and y are identically distributed variables, then a P–P plot for x and y will be a straight line configuration oriented from $(0, 0)$ to $(1, 1)$. However, unlike the Q–Q plot, the P–P plot will not remain linear in the event that either of x and y is subject to a linear transformation.

Though this limits the general usefulness of P–P plots, they do, nevertheless, have some areas of value. In particular, the P–P plot will usually be especially sensitive to discrepancies in the middle of a distribution rather than in the tails, for reasons comparable to those in the case of Q–Q plots. (Other techniques also exist for sensitive discrimination in the middle.) Moreover, the basic idea of P–P plots can be applied to the multivariate situation, while the Q–Q plot does not seem to have any meaningful multivariate analogue. Techniques for using, generalizing, and interpreting the general P–P plot formulation require more investigation, including estimation procedures and robustness properties in preliminary data standardization methods.

An example of the use of a P–P plot is given in Fig. 13 which shows the results for the normal approximation to the distribution of $\log \chi^2_{10}$, corresponding to the Q–Q plot of Fig. 6. One can see on this scale of percentages that the distributions differ not only in the tails but also in the middle.

A P–P plot, corresponding to the Q–Q plot of Fig. 12, is given in Fig. 14 for the data on the energies from two talkers. This plot shows a marked departure from the 45° line, though in these data the energy level may vary from talker to talker due to artifactual effects. Such a shift of location would destroy the usefulness of the P–P plot but not of the Q–Q plot.

A particular case of the use of P–P plots is in the presentation of results on the power of a statistical test as a function of the so-called significance level. Indeed, the power function is just the c.d.f. of the achieved significance level, p, which is of course a random variable. Thus the abscissa is the p value of the null distribution while the ordinate is the p value of the non-null distribution. An example of such a use is shown in Fig. 15 which gives the comparative empirical powers of 9 tests for normality, in samples of size 15 for the χ^2_2 alternative (taken from an unpublished report by Shapiro, Wilk & Chen).

4·4. *Extensions and hybrids of* P–P *and* Q–Q *plots*

Some immediate extensions of P–P and Q–Q plots may be mentioned. A Q–Q plot has been described as a plot of $q_y(p)$ versus $q_x(p)$. Clearly one may have this as a special case of a plot of $g(q_x, q_y)$ versus $h(q_x, q_y)$, and the concept of P–P plots may be extended in the same way.

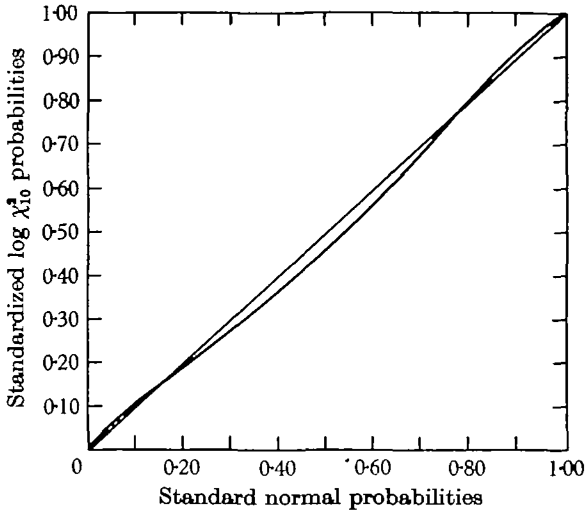

Fig. 13. P–P plot of standardized log χ^2_{10} *vs.* normal.

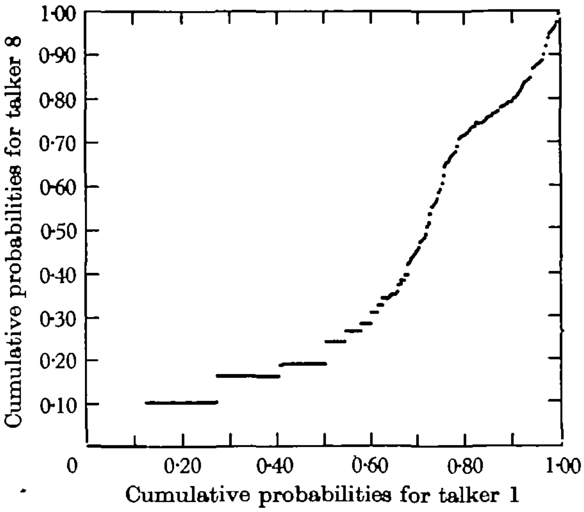

Fig. 14. P–P plot of e.c.d.f.'s. of same word by talkers 1 and 8.

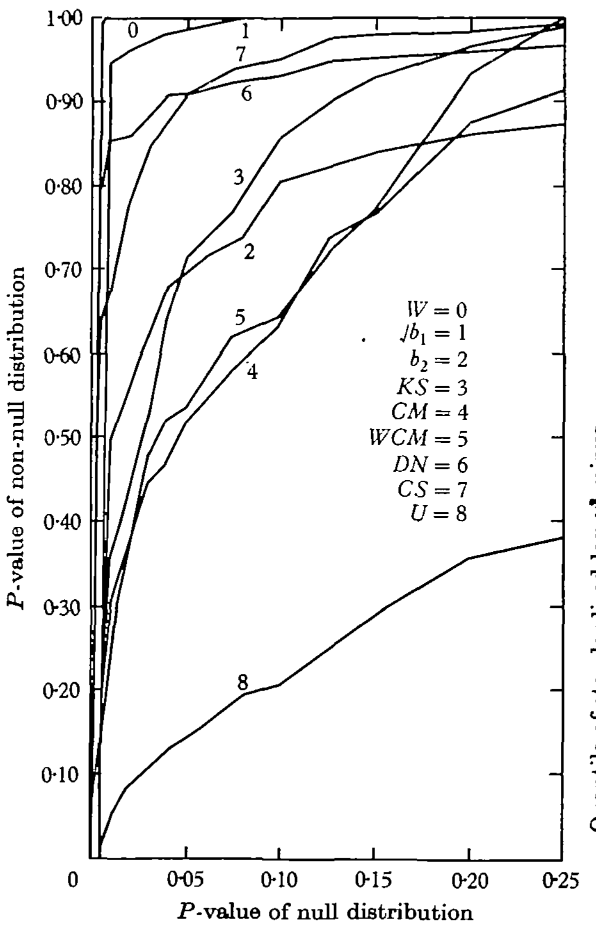

Fig. 15. P–P plot for empirical power comparisons of several tests for normality (alternative is χ^2_2, sample size 15).

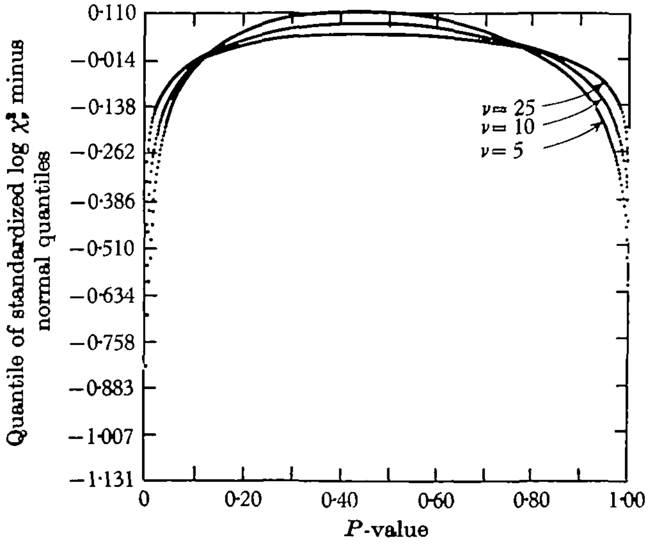

Fig. 16. Difference in Q values of standardized χ^2_ν and normal *vs.* P.

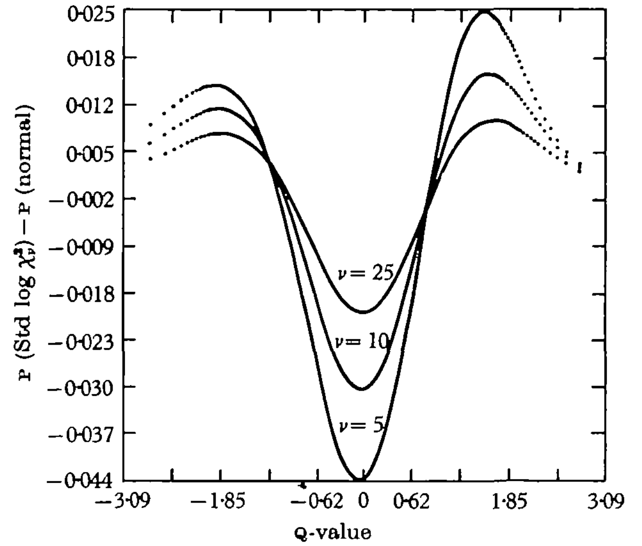

Fig. 17. Difference in P values of standardized χ^2_ν and normal *vs.* Q.

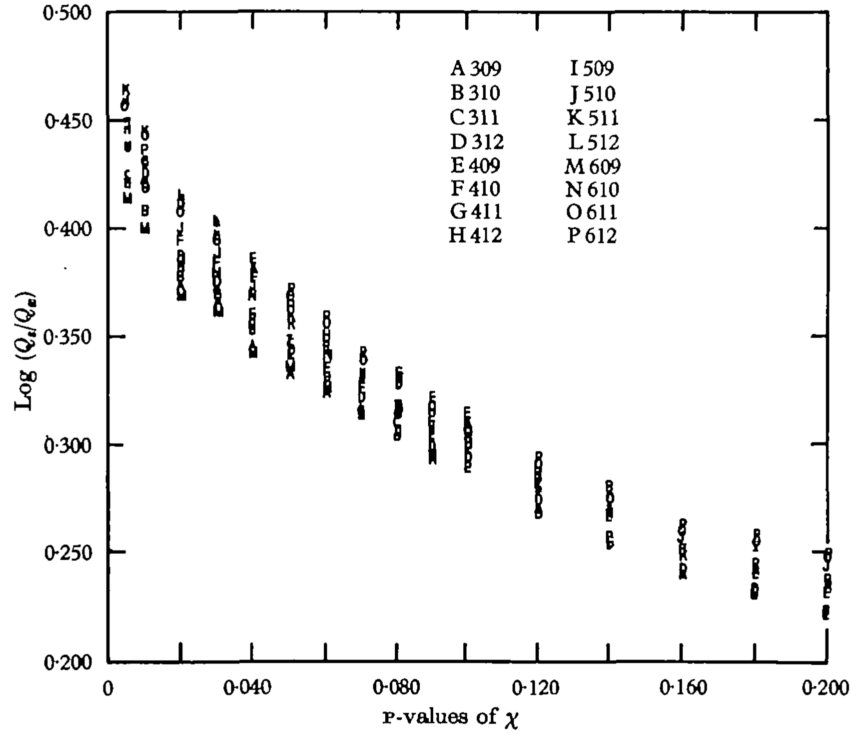

Fig. 18. Db gain of simulated signal *vs.* P value of original.

It is also useful in some circumstances to employ hybrids of P–P and Q–Q plots, for instance to plot $g\{q_x(p), q_y(p)\}$ versus p, or more generally $g\{q_x(p), q_y(p)\}$ versus $h(p)$. Similarly, one may consider plotting $g\{p_x(q), p_y(q)\}$ versus $h(q)$.

Examples of such hybrid plots are given in Figs. 16 and 17 which summarize the adequacy of a normal approximation to the distribution of log χ^2, for various degrees of freedom (as shown in unpublished work by M. Gnanadesikan). These plots show, respectively, the difference $q_y(p) - q_x(p)$ versus p and the difference $p_y(q) - p_x(q)$ versus q, where y refers to the standardized log χ^2 and x to the standard normal. From both plots, it is seen that the approximation improves as the number of degrees of freedom of the χ^2 increases.

Another example of a hybrid plot is given in Fig. 18 which derives from the mobile radio data. One object of the analysis was to evaluate the comparative effectiveness of antennae systems with different spacings by means of simulation. Thus, Fig. 18 shows the db gain (i.e. the logarithm of the ratio of quantiles of the simulated and observed signals) plotted against the p value of the original signal, for 16 different antennae configurations. The object is to produce a large db gain for small values of p. Accordingly the upper points in Fig. 18 represent better systems. For instance, when the p value of the original signal is 0·005 or 0·01, the db gain is highest for the combining scheme K.

Other modifications may be appropriate to particular circumstances. Thus, for instance, the data underlying a P–P plot, such as Fig. 15, might be plotted in terms of equivalent normal deviate, or instead of the difference $p_y(q) - p_x(q)$ in Fig. 17, one might plot the difference of two angle transforms.

4·5. *Various derivatives of probability plots*

The probability plot can serve as a valuable stimulus for a variety of statistical procedures. For example, the well-known Kolmogorov–Smirnov statistic may be viewed, after the appropriate probability transform of the sample, as the maximum deviation from the 45° line on a uniform P–P plot (which for the uniform distribution on (0, 1) is equivalent to a Q–Q or P–Q plot). As another example, one may consider the regression of order statistics on expected values of standard order statistics in a Q–Q plot to generate test procedures for composite distributional hypotheses, as developed by Shapiro & Wilk for the normal, uniform and exponential distributions using complete or censored samples (Shapiro & Wilk, 1965).

Various graphical transforms stimulated by probability plotting have been advanced by Tukey in unpublished as well as published work; some examples are the procedures labelled FUNOP, FUNOR, FUNOM described by Tukey (1962).

5. Probability plots as informal aids to inference

The procedures of probability plotting may also be usefully employed in connexion with the complex objectives generally associated with the analysis of variance. It is typical of analysis of variance situations that one wishes to ask many questions of the same body of data, and moreover that one wishes to provide a climate such that unanticipated characteristics may be spotted. Examples of non-obvious, interesting indications concerning the structure of the data are the presence of possible real effects, existence of outliers, distributional peculiarities, etc. The real applied value of the analysis of variance table is as a summary of patterns of variability. It provides a collection of mean squares each associated with an identifiable facet of the experiment and it will often be true that appropriate subsets of these are meaningfully comparable.

It is of value to provide procedures which use statistical models as backgrounds to aid in the simultaneous comparisons of the mean squares, but without the need to commit oneself on a narrow specification of objectives. The present authors have termed such procedures as *internal comparisons* methods (Wilk & Gnanadesikan, 1961, 1964; Wilk, Gnanadesikan, & Lauh, 1966). Some probability plotting techniques which have been proposed as internal comparisons methods for the analysis of variance will be discussed in the sequel. These procedures provide a statistical measure to aid in the assessment of relative magnitudes, which may be non-intuitive especially in large collections. Moreover, the procedures can provide insight into various inadequacies of the statistical models and are not overly influenced by some data-independent aspects, such as a need to prechoose an error term.

The essence of a probability plot as an aid to informal inference is to plot the ordered 'sample' values against some *representative* values from a presumed null standard distribution. For example, from a uniresponse 2^n experiment, one calculates a new 'sample' of 2^n-1 main effects and interactions. The ordered values of these 2^n-1 derived quantities may be plotted against representative values from a standard normal distribution. Under certain null statistical conditions, this will give rise to a linear configuration, passing through the origin, whose slope provides an indication of the underlying error standard deviation. The presence of real effects, the existence of distributional peculiarities, of outliers, and of heterogeneities of variance result in distortions of the linear configuration of the plot. In addition, the graphical summary provided by the plot focuses attention on the large effects and groupings among them, if any, and moreover does this in a simple and palatable fashion.

Choice among representative values in probability plotting involves both practical convenience and conceptual insight. Two conceptual categories may be distinguished: (1) the representative values are 'corresponding quantiles' of the reference distribution. The configurations of plots of the ordered sample values against the quantiles of the reference distribution corresponding to any of the fractions, $(i-\frac{1}{2})/n$, $i/(n+1)$, etc., will not be very different except for very small sample sizes; (2) the representative values are determined by the expected values of the standard order statistics from the reference distribution. Typically, where both quantile and expected value plotting positions may be used, the resulting configurations will be similar.

In most of the simple cases, it is far more convenient, computationally and conceptually, to employ quantiles rather than expected values. In more complex cases, however, such as the case of independent unequal statistical components, which arises in analysis of variance with differing degrees of freedom, the notion of an expected value can be well defined, conceptually, while it is quite unclear as to what meaning, if any, may be given to the notion of a quantile. Similar comments are applicable to the cases of non-independent equal or unequal components.

6. Specific internal comparisons probability plotting techniques

6·1. *Classification of orthogonal analysis of variance*

Table 2 shows a classification of orthogonal analysis of variance situations and indicates the probability plotting techniques available. The classification distinguishes uni-response from multi-response situations, and for each of these gives three categories of decomposition of the treatment structure, namely, all single degrees of freedom, all multiple but equal degrees of freedom and the general mixed degrees of freedom case. Also given in Table 2 is a summary of the available internal comparisons methods, with references, appropriate to this categorization.

Table 2. *A classification of orthogonal analysis of variance situations*

Decomposition of treatment structure	Response structure	
	Univariate	Multivariate
All 1 d.f.	I	IV
All ν d.f.	II	V
Mixed d.f.	III	VI

Cell I. Full-normal and/or half-normal probability plotting. Note half-normal on absolute contrasts is equivalent to gamma plotting of squared contrasts with shape parameters $\eta = \frac{1}{2}$ (Daniel, 1959; Wilk, Gnanadesikan & Huyett, 1962*a*; Wilk, Gnanadesikan & Freeny, 1963).

Cell II. Gamma plotting with $\eta = \frac{1}{2}\nu$ (Wilk, Gnanadesikan & Huyett, 1962*a*, 1963).

Cell III. Generalized probability plotting using conditional expected values (Wilk, Gnanadesikan & Lauh, 1966).

Cell IV. Gamma plotting of generalized squared distances using estimated shape parameter (Wilk & Gnanadesikan, 1961, 1964; Wilk, Gnanadesikan & Huyett, 1962*b*).

Cell V. Work in progress.

Cell VI. Open.

6·2. *The univariate single degree of freedom case*

Figures 19*a–d* give examples appropriate to Cell I of Table 2. The data, taken from the book edited by Davies (1956), concern the dependence of quality of a certain dyestuff on various factors, in a half-replicate of a 2^5 experiment. Figures 19*a* and *b* are full-normal while Figs. 19*c* and *d* are half-normal Q–Q plots. Thus, in particular, the full-normal plot gives the ordered effects plotted against the $(i-\frac{1}{2})/n$ quantiles of the standard normal. The half-normal plot, shows the ordered absolute effects plotted against the $(i-\frac{1}{2})/n$ quantiles of the half-normal distribution. Figures 19*b* and *d* are obtained after dropping the largest effect, which is the *D* main effect.

It will be seen from Figs. 19*a* and 19*c* that the *D* main effect does not appear to belong with the remaining statistical configuration. However, from Figs. 19*b* and 19*d* one sees that the remaining 14 contrasts do seem to behave like a reasonable error configuration.

The reason for doing half-normal plotting is that the sign of the contrast is, from the null viewpoint, irrelevant and one may obtain a more stable and focused display. One reason for having also the full-normal plot is the possible interest in the actual signs of the individual or groups of contrasts when in fact some may be exhibited as being non-null. Another reason is to aid in exhibiting possible distributional peculiarities, when some of the contrasts reflect real experimental effects, which may coincidentally be concealed when the distribution is 'folded'. While the latter contingency may arise only rarely, it would appear the better part of data analytic caution to include both full- and half-normal plots routinely in analysis of 2^n experiments. The use of half-normal plotting in two-level factorial experiments was suggested by Daniel (1959), and independently considered by Bennett & Tukey, and Hamaker.

6·3. *The univariate equal degrees of freedom case*

The Q–Q plot in Fig. 20 illustrates the application of gamma plotting in the case of a collection of mean squares all with the same number of degrees of freedom. The example is from Hald (1952) and the data concern percentage concentration of calcium carbonate in a mixing experiment. The ordered 52 within cell variances, each based on 4 degrees of freedom, are shown plotted against the $(i-\frac{1}{2})/n$ quantities of the standard gamma distribution with

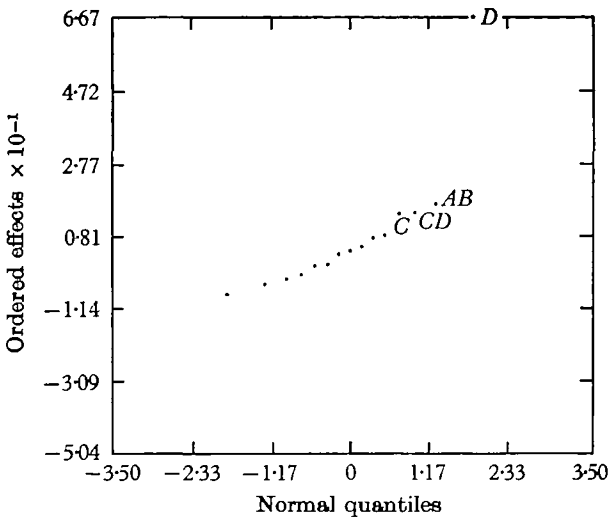

Fig. 19*a*. Normal Q–Q plot of 15 effects.

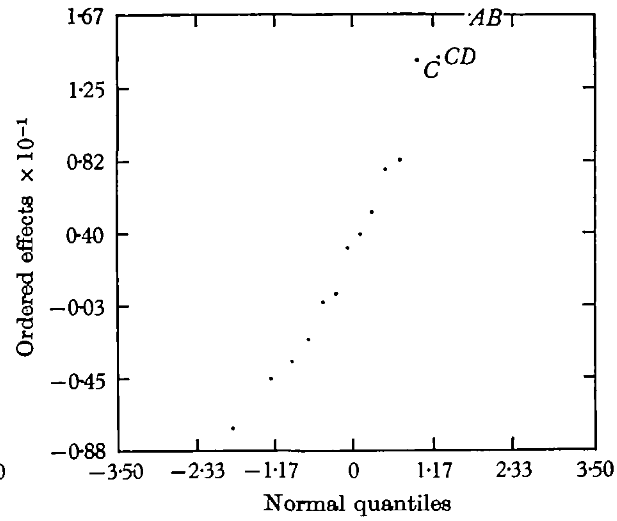

Fig. 19*b*. Normal Q–Q plot of 14 effects.

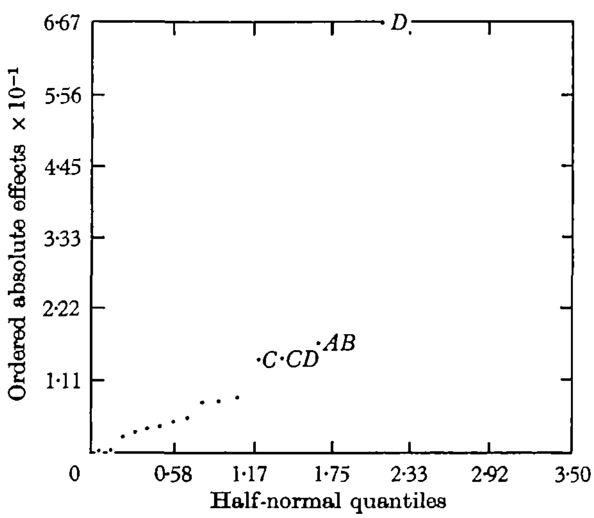

Fig. 19*c*. Half-normal Q–Q plot of 15 effects.

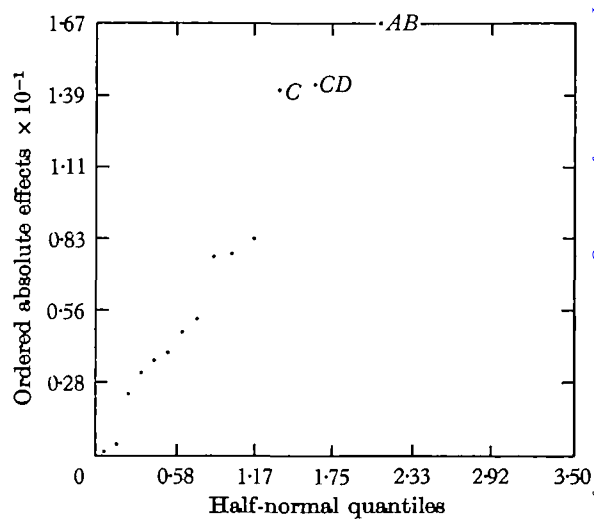

Fig. 19*d*. Half-normal Q–Q plot of 14 effects.

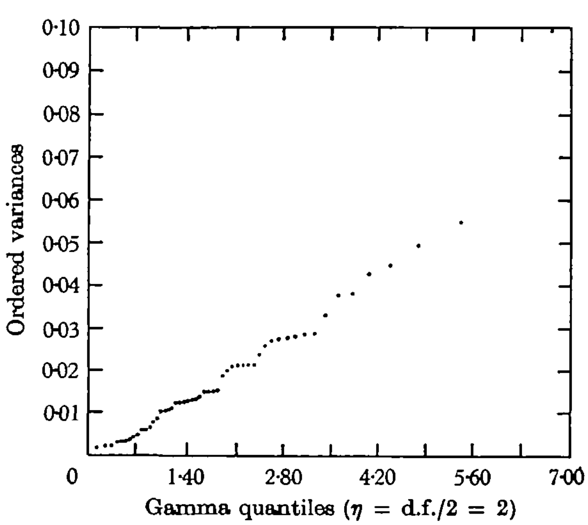

Fig. 20. Gamma Q–Q plot of 52 within-cell variances.

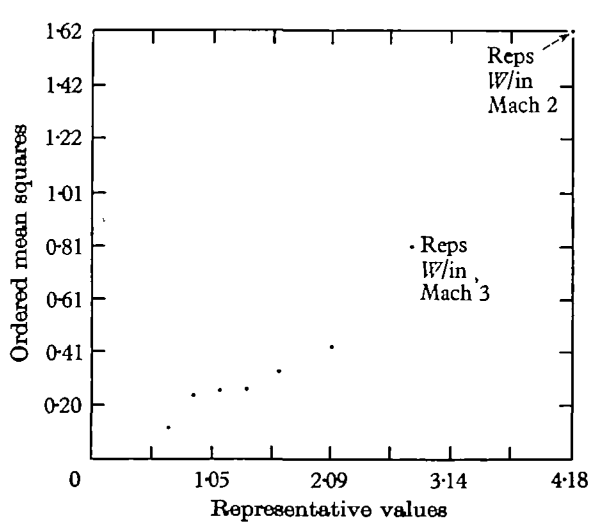

Fig. 21. Generalized probability plot for 8 mean squares in a metal fatigue study.

shape-parameter equal to 2. It will be seen that, with the exception of the largest variance, the remainder form a coherent sample. With one simple-to-interpret plot one is able both to pinpoint one aberrant value and also to gain assurance on the reasonable statistical behaviour of the remaining points.

6·4. *The univariate mixed degrees of freedom case*

The general analysis of variance case involves mean squares with possibly unequal degrees of freedom. The difficulty in applying probability plotting in this circumstance is that even under the null model of no real effects, and assumptions of χ^2 distributions and common error variance, the mean squares do not all have the same distribution because their degrees of freedom differ. To handle this problem, the present authors have proposed a technique of generalized probability plotting which consists of plotting the ordered analysis of variance mean squares against representative values defined as expected values of appropriately conditioned order statistics of standardized mean squares. An application of this technique is provided by Fig. 21 which concerns a subset of 8 mean squares from an analysis of variance of some transformed metal fatigue data (supplied by Torrey, Gohn & Wilk). These mean squares have degrees of freedom ranging between 2 and 12. The plot indicates that, relative to the configuration of the others, the largest mean square in the collection does appear overly large and worthy of further examination.

6·5. *The multivariate single degree of freedom case*

Techniques have been proposed in Wilk & Gnanadesikan (1964) for internal comparisons amongst a collection of single degree of freedom contrast vectors in the multi-response case. The technique is based on gamma plotting of positive semi-definite quadratic forms in the elements of the contrast vectors, using a shape parameter estimated from a collection of the 'smaller' of these squared distances. Figures 22*a* and *b* illustrate the use of this method in analysis of a half-replicate of a 2^9 experiment on factors affecting quality of television pictures (data supplied by Gabbe & Wilk). The response was eight-dimensional. Figure 24*a* is a plot for all 1, 2 and 3-factor interactions. It will be seen from Fig. 22*a* that there is a strong suggestion of two intersecting linear configurations. This does indeed correspond to the fact that in this experiment factors A, B, C and D were whole-plot factors, while the remaining five were split-plot factors. Figure 22*b* is a replot omitting all 14 effects involving A, B, C and D. The resulting configuration gives strong indication of a number of real effects against a background of a stable error configuration.

6·6. *Residuals in regression analysis*

In regression analysis, it has been traditional to base inferences on the mean squared error, i.e. the sum of squares of residuals from the fitted model divided by the degrees of freedom. The use of only this summary precludes the employment of one of the most sensitive and informative tools in regression studies, namely plots of various kinds of the individual residuals. Such plots may be used to check on the adequacy of the model, on the appropriateness of independent variables, on the existence of outliers, the relevance of extraneous variables and on distributional peculiarities. One kind of such plots, by no means the most important, is a probability plot.

Figures 23*a* and *b* give, respectively, full-normal and half-normal plots of 745 residuals from a 9-parameter non-linear model for the spatial distribution of high energy protons as measured by Telstar I (data supplied by Gabbe, Wilk & Brown). These are residuals based on the square root of observations which are believed to be approximately Poisson distributed. These plots indicate that the residuals behave reasonably like a sample from a normal distribution with mean zero and standard deviation approximately 0·2. This compares reasonably with the predicted approximate standard deviation of 0·16 from the Poisson assumption. The plots provide assurances regarding outliers as well.

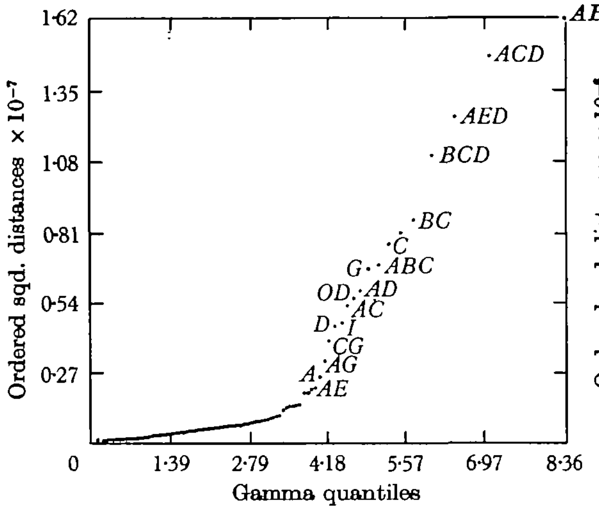

Fig. 22a. Multivariate gamma Q–Q plot of 129 effects in Videotelephone data.

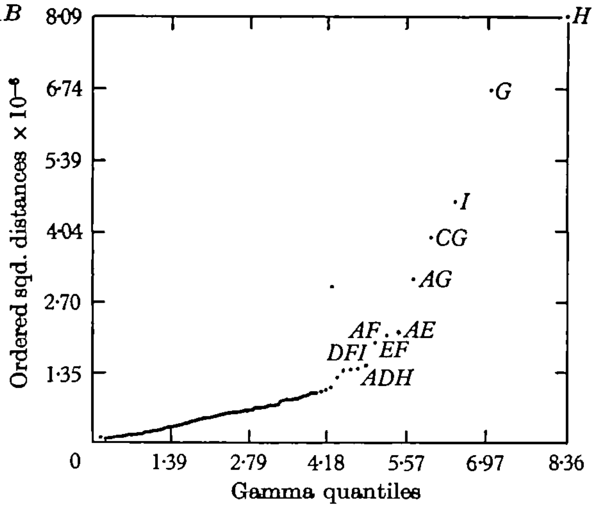

Fig. 22b. Multivariate gamma Q–Q plot of 115 effects in Videotelephone data.

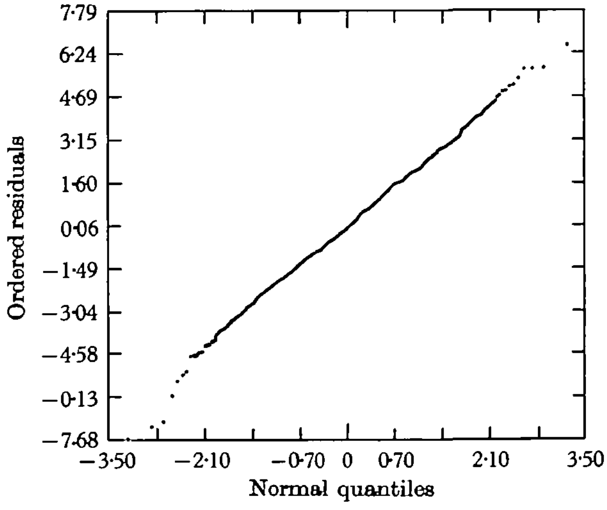

Fig. 23a. Normal Q–Q plot of residuals in space radiation study.

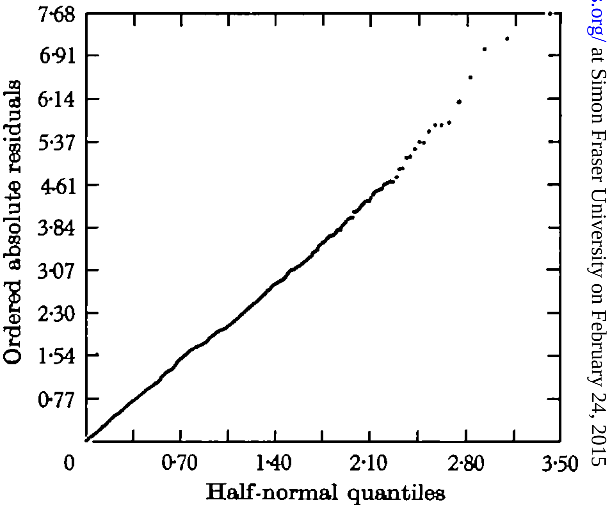

Fig. 23b. Half-normal Q–Q plot of residuals in space radiation study.

6·7. *Other applications*

Amongst other applications of probability plotting in the analysis of variance framework one might mention the cases of collections of regression coefficients and of residuals in multiway tables. Typically these involve dealing with order statistics of correlated or singular random variables.

7. Concluding remarks

The graphical methods which have been exemplified in the present paper are informal tools for the statistical analysis of data and may be employed for describing and summarizing as well as for uncovering and understanding the structure underlying a body of data. While using statistical assumptions and models to generate the graphical display or for analysing a given body of data, the methods do not, however, imply a commitment to interpretation independent of a judgement of the adequacy of the assumptions. Thus these techniques possess many of the desirable attributes of productive data analytic methodology as discussed by Tukey & Wilk (1966).

It may be remarked in closing that most of the procedures discussed in the earlier sections can be and have been implemented for routine availability in data analysis on a high speed computer in conjunction with microfilm graphical output.

The authors wish to thank Miss Elizabeth Lauh for her assistance in organizing the presentation of the examples and D. R. Cox, J. D. Gabbe, C. L. Mallows, H. O. Pollak and J. W. Tukey for their comments on an earlier draft.

References

Barton, D. E. & Mallows, C. L. (1965). Some aspects of the random sequence. *Ann. Math. Statist.* **36**, 236–60.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1**, 311–42.

David, H. A., Hartley, H. O. & Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika* **41**, 482–93.

Davies, O. L. (Editor). (1956). *The Design and Analysis of Industrial Experiments*. Edinburgh: Oliver and Boyd.

Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika* **48**, 41–55.

Gnanadesikan, R., Lauh, E., Snyder, M. & Yao, Y. (1965). Efficiency comparisons of certain multivariate analysis of variance tests procedures. Abstract 28, *Ann. Math. Statist.* **36**, 356–7.

Gnanadesikan, R., Pinkham, R. S. & Hughes, L. P. (1967). Maximum likelihood estimation of the parameters of the beta distribution from smallest order statistics. *Technometrics* **9**, 607–20.

Hald, A. (1952). *Statistical Theory with Engineering Applications*. New York: Wiley.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611.

Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Statist.* **33**, 1–67.

Tukey, J. W. & Wilk, M. B. (1966). Data analysis and statistics: an expository overview. *AFIPS Conference Proceedings* **29**, 695–709.

Wilk, M. B. & Gnanadesikan, R. (1961). Graphical analysis of multi-response experimental data using ordered distances. *Proc. Nat. Acad. Sci. U.S.A.* **47**, 1209–12.

Wilk, M. B. & Gnanadesikan, R. (1964). Graphical methods for internal comparisons in multiresponse experiments. *Ann. Math. Statist.* **35**, 613–31.

Wilk, M. B., Gnanadesikan, R. & Freeny, A. E. (1963). Estimation of error variance from smallest ordered contrasts. *J. Am. Statist. Ass.* **58**, 152–60.

Wilk, M. B., Gnanadesikan, R. & Huyett, M. J. (1962*a*). Probability plots for the gamma distribution. *Technometrics* **4**, 1–20.

Wilk, M. B., Gnanadesikan, R. & Huyett, M. J. (1962*b*). Estimation of parameters of the gamma distribution using order statistics. *Biometrika* **49**, 525–45.

Wilk, M. B., Gnanadesikan, R. & Huyett, M. J. (1963). Separate maximum likelihood estimation of scale or shape parameters of the gamma distribution using order statistics. *Biometrika* **50**, 217–21.

Wilk, M. B., Gnanadesikan, R. & Lauh, E. (1966). Scale parameter estimation from the order statistics of unequal gamma components. *Ann. Math. Statist.* **37**, 152–76.

[*Received May* 1967. *Revised October* 1967]