No. 9

Statistical Methods to Measure The Consensus of Experts Opinions in Delphi Forecasts and Assessments

January 1977

Hajime Eto

University of Tsukuba

PRÉCIS

Consensus among experts is here shown to be interpretable as a leading indicator or rather an antecedent factor of a national consensus and therefore as an intersubjective probability of choosing among prospective policies when experts are properly selected. The condition under which this interpretation is valid is shown to hold in several Delphi surveys. Hence a logical foundation is obtained to fill two gaps: a gap between subjective opinions and objective technology breakthroughs, and the other gap between distributed opinions and a single event. Thereupon statistical methods are proposed to measure the consensus among experts in terms of distribution matchings of their opinions. Experiences in technological and social forecasts and assessments are reported.

1. INTRODUCTION

The Delphi method is a prevalent tool for technology forecastings. Its procedure consists of [4, 11, 13] (i) asking experts when a technological breakthrough is expected to be accomplished, (ii) feeding them back the information of distribution of their opinion, and (iii) asking the experts again the same questions for their possible revision of opinions. This iterative procedure has indeed yielded a satisfactory consensus in most Delphi surveys thus far [9, 12, 19, 20, 21, 26]. A doubt remains unresolved as to what the consensus means for technological forecasts. This is exactly why the proposed program for further development in Delphi [14] failed to mention the quantitative analysis of the implication of a consensus. There are two gaps to be filled before the quantitative analysis. A gap exists between the subjectivity of opinions and the objectivity of the technology development process. This gap raises a doubt whether the analysis of subjective opinions yields any useful information about the objective process. The other gap exists between distributed opinions and the uniqueness of a single event. This latter gap raises another doubt whether the analysis of distributions yields any useful information about a single event. These doubts will be resolved in Chapter 2. Psychological experiments with students as subjects show that guesses on the foregotten facts converge to the truth [5, 6, 7], but they disclose a third gap between eligible experts and randomly selected students as well as one between the facts and the prospects on the future because the truth is undefinable with the prospects. This paper takes a different course from this experimentalism.

2. THE CONSENSUS AS A LEADING FACTOR

The consensus in Delphi can be viewed as a leading indicator or rather a leading or antecedent factor of technology planning for two reasons. One reason is that a technology planning is expected to be based on the result of the Delphi consensus. The other reason is that most of the experts who join Delphi surveys are members of some technology planning board. Especially in the Japanese governmental Delphi surveys in which hundreds of experts took part, the influential experts in the important institutes and agencies were almost exhausted. This fact makes Delphi influential in technology planning. For these reasons the experts opinions can be viewed as the most influential factor when the past data are almost inapplicable. Nowadays the planned technology is important in the whole technology, and a leading factor of technology planning is also that of the whole technology. Accordingly the subjective opinions play a key role in the objective process of technology development and thus carry objective meanings.

The consensus between groups is measurable in terms of the degree of opinion distribution matching wherein many statistical concepts are meaningful. The averaged opinion itself may be uninterpretable as an opinion, but the historical fact that the concept of average has been associated with the equality and fairness (e.g. an average score of subjects in examination) induces planners to consider the average opinion as the most reasonable and persuasive opinion. This is explained by noting that the mean is the gravity center to which opinions tend to drift. Hence the distance between the average opinions of two groups may be assumed to monotonically correspond to the confidence level or subjective probability of many people (intersubjective probability) that the groups are agreeable. (X monotonically corresponds to Y when $x_1 > x_2 \rightarrow y_1 = f(x_1) > y_2 = f(x_2)$). Medical doctors expect averagedly the accomplishment of automatic diagnosis to be much later than electronic engineers do. This difference indicates the difficulty in its social acceptance. The relative frequencies at the cells of mode and median may also be assumed to monotonically correspond to the intersubjective probability that the within-group consensus is attainable at the cells. Even regarding the opinion of electronic engineers about automatic diagnosis the frequency at the cell of mode and median (eventually both fall at the same cell) is quite low, which indicates that engineers are hardly expected to attain the concordant prospect thereabout in a short time. Analogously the variance may be assumed to measure the within-group discord. The variance of opinions about the solar energy utilization is large in any group, which indicates the absense of its effective approach. The difference among the mean, mode and median may be assumed to indicate the within-group discord.

Statistical inferences within the population-sample frame are sometimes inapplicable to Delphi which is postulated as a census to exhaust all the eligible experts whose opinions seldom vary. It may deserve noting that the opinion of judges in a court is postulated not to be sample in a constitutional decision rule. Though experts may change their opinions through rounds, the opinions in each round have their own meaning and may not be samples. Experts may change their opinions after a survey, but their new opinions are taken into no account in a technology planning process without a counter scientific evidence (e.g. the evidence about the noxiousness of a chemical material in question.) A judge may change his opinion after his case is over, but his new opinion has legally no effect on his former judgement without a counter-physical evidence. Hence the opinions in Delphi are claimed to be the finite population unless our problem is to criticize the Delphi opinions.

Hereafter groups may be by occupation, major field, educational background, political stand, etc. Grouping by stand may be done by the cluster analysis.

3. MEASURES FOR SINGLE-ISSUE CONSENSUS BETWEEN GROUPS

3.1 The rationale of linearization of opinion space

It is assumed in sequel that each expert is to choose in a questionnaire exactly one alternative among mutually disjoint alternatives. (e.g., 1970', 1980', 1990' as time of a technological breakthrough) Let natural numbers 1, 2, ..., m be assigned to alternatives whose total number equals m so that the mean and variance of opinions are definable. This assignment is justifiable as follows. Each alternative is recognized as a reasonable opinion with the equal raison dêtre, and accordingly the various alternatives are usually, in course of debate, considered to be linearly arranged from "left wing" to "right wing" with the equal intervals.

3.2 Correlation ratio as the consensus between the representatives

The average opinion is interpretable as the representative opinion, and its fractional part is interpretable as the tendency for a compromize. For example the average 2.4 and 1.6 fall in the same alternative 2, but the former tends to compromize with alternative 3 while the latter with alternative 1.

The standardized distance between the average opinions of two groups is measured by the correlation ratio $d_1 = \sqrt{d_1^2}$ such that

$$d_1^2 = N_1 N_2(\mu_1 - \mu_2)^2/\{(N_1 + N_2)^2 S^2\}$$

and the degree of consensus between groups c_1 is $c_1 = 1 - d_1$ where N_i denotes the number of the members in group i, μ_i does the average opinion and S^2 does the total variance. As is seen, d_1 and c_1 lie in [0, 1], i.e., c_1 is also standardized. This enable the degree of consensus on an issue to be compared with that on another for the given pair of groups. Note that this measure does not (at least explicitly) presume the population-sample frame. The majority opinion (mode) and the moderate opinion (median) may also be used to define: $d_1'^2 = (\mu_1' - \mu_2')^2/m^2$ where μ_i' denotes the majority opinion or the moderate opinion in group i and m does the number of the alternatives. The value of d_1' vanishes in many issues and may not be used for ranking the issue.

3.3 Separability and inseparability as the possibility of mediation

When the opinions of two groups are separated as in Fig. 1, the groups may be unable to sympathize each other. Therefore a slight discord may cause a hostile conflict between the groups. Otherwise the penetrators (the shaded area in Fig. 2) serve as mediators for the groups to understand each other, avoiding a hostile conflict. This is the case when each group (e.g., a local community, trade union or professional union) is so firmly organized that the shaded part do not split from but stay in their respective groups to play a role of mediators. The ratio of the shaded area to the whole in Fig. 2 measures the inseparability or the penetrability and corresponds monotonically to the intersubjective probability that a conflict be avoided. This measure which also presumes no population behind samples works better than the foregoing measure in §3.2 when the two groups are opposite like the workers and the executives in social forecastings.

3.4 Minority adjusted distribution matching

It is often desirable to consider the majority and the minority opinion with the equal weights. This is particularly important in a technology assessment. Long before the danger of the agricultural chemicals and inorganic mercury was proved by scientists, the minorities of consumers and fishermen had pointed out the danger which was persistently denied by scientists. Meanwhile no discord was found among the majorities of consumers, fishermen and scientists because the majorities of consumers and fishermen believed in the statement of scientists.

Even if the majorities of groups are fairly concordant each other, it is still possible that the minorities of groups have an opinion gap which may indicate a future conflict. Conversely, even if the majorities of groups are fairly discordant, the opinions of the groups are hardly expected to repulse each other in the future if the minorities (i.e., the distribution tails) of groups are accordant. Existing methods for measuring the distribution matching either belittle the minority or use the order statistics which presumes the continuous or nearly-continuous distribution under the population-sample frame [2, 17, 25, 27] while Delphi yields strongly discrete distributions with a few cells and stands out of the population-sample frame. Since the alternatives in the questionnaires range from two (yes-no question) to five (e.g., very important, important, medium, not very important, unimportant) in most issues, the order statistics makes little sense.

The minority adjusted difference d_3 between two histograms may be measured by the following procedure: (i) those alternatives at which the sum of real frequencies of the two groups is less than six are combined into the adjacent alternative so that the sum of real frequencies exceeds five at every alternative, (ii) let f_{ij} denote the observed relative frequency (e.g., permillage) of group i at alternative j, and define $e_j = (f_{1j} - f_{2j})/(f_{1j} + f_{2j})$ as a standardized difference which lies in $[-1, 1]$, (iii) define $d_3 = \sum_{j=1}^{m} e_j^2/m$ where m denotes the number of the alternatives at all of which the sum of real frequencies exceeds five. Obviously d_3 lies in $[0, 1]$ with $d_3 = 1$ only as in Fig. 1 and $d_3 = 0$ only in the complete matching. Hence $c_3 = 1 - d_3$ serves as a correlation coefficient of two histograms.

Define c_3' as follows,

$$c_3' = \sum_{(F_{11}, F_{12}, \ldots, F_{1m}) \varepsilon R} \binom{f_{11}+f_{21}}{F_{11}}\binom{f_{12}+f_{22}}{F_{12}}\cdots\binom{f_{1m}+f_{2m}}{F_{1m}}\Big/\binom{2N}{N}$$

where $R = \{(F_{11}, F_{12}, \ldots, F_{1m}): \sum_{j=1}^{m} E_j^2 > \sum^{m} e_j^2\}$, E_j is a random variable corresponding to e_j, and F_{ij} varies under the constraints $F_{1j} + F_{2j} = f_{1j} + f_{2j}$ for all j satisfying $\sum_{j=1}^{m} F_{1j} = \sum_{j=1}^{m} F_{2j} = \sum_{j=1}^{m} f_{1j} = \sum_{j=1}^{m} f_{2j} = N$ (= 100 in case of percentage). The value of c_3' is interpretable as the confidence level that each group is just a random division of the same group of size 2N. This presumes no unobserved population because the population here is the set of all possible divisions of the union of the observed two groups. So $(f_{1j} + f_{2j})/2$ is not an estimate but the true value of the relative frequency of population. The values of d_3 (d_3') and c_3 (c_3') measure respectively the dissimilarity and similarity of two histograms primarily in regard to the width of dispersion. The difference of scores of two groups in a discriminant analysis on percentage distributions [16] may serve the same as c_3.

When the population of infinitely large size is meaningfully presumable, the chi-square or the F-distribution are applicable, and $\chi^2 = \sum_{j=1}^{m} \{ (f_{1j} - f_{2j})^2 / 2(f_{1j} + f_{2j}) \}$ with m-1 degrees of freedom or $F = \chi^2/(m-1)$ with m-1 and infinite degrees of freedom is used. This method yields a good approximation even when the number of experts in both groups is as small as 25 if the procedure (i) is applied [8, p197].

When the alternatives are numerous, the correlation between f_{1j} and f_{2j} for $j = 1, \ldots, m$ is obtainable in the ordinary sense and measures the degree of consensus with the minority adjusted. The Kolmogorov-Smirnov statistics is definable on a discrete distribution as follows.

$$d_k = \max(d_+, d_-), \; d_+ = \max_J \sum_{j=1}^{J} (f_{1j} - f_{2j}), \; d_- = \min_J \sum_{j=1}^{J} (f_{1j} - f_{2j})$$

with $J \leqq m$. The confidence level of the identity hypothesis of two distributions is obtainable under the population presumption by the Conover's method for discontinuous case [3] unless the distributions are too discrete. Let $d_k' = d_+ + d_-$ then d_k' measures the difference more discriminatorily than d_k though d_k' fails to provide informations of the population. Since d_k and d_k' belittle the minority cells, some standardization to make the distribution of the union of two groups uniform is needed for the minority adjustment.

3.5 Evaluation of the three measures

The above-described three measures were applied to both a technology forecasting and a social forecasting, the latter of which was done by the Ministry of Labour [19] with leaders of workers unions, executives of important firms and neutral specialists in labor problems as panelists. As was expected, the three measures revealed more interpretable informations about issues of controversy (e.g., pollution, computerization in medical or educational fields, income policy and redistribution etc.) than about issues of less controversy, but they revealed some useful information about a few of issues of the latter kind (e.g., telecommunications). Throughout the applications to both forecastings the correlation ratio (§3.2) and the minority adjusted distribution matching (§3.4) were found complementary in the sense that they provided some different informations. Each of the three measures was used to rank the issues with respect to the consensus between the workers and the executives in the social forecasting. The rank given by the correlation ratio was the most reasonable and appealing to our intuition in this experience.

The three measures were together applied to each pair of groups regarding each issue. For most issues the distance given by each measure formed the group-relation simplexes [1] of expected configuration that the workers stood the farthest from the executives. Also for most issues, the neutrals stood much closer to the executives than to the workers in the simplexes formed in terms of the correlation ratio and stood in the nearly same distance from both in the simplexes in terms of the other two measures.

In conclusion each measure alone was not enough for the analysis.

Each pair of groups was ranked according to the degree of consensus given by each of the three measures for every issue. These ranks were analyzed by the multi-issue rank correlations described in

Chapter 4 to give another evaluation of the three measures.

4. MULTI-ISSUE RANK ANALYSIS

4.1 Multi-issue rankings

For a given issue and a given measure, each pair of groups was ranked according to the degree of consensus (Table 1). Issues were collected into several fields (e.g., pollution, agricultural technology etc.). Issues in the same field were considered here not to be independent samples from the unobserved population but to form a multi-issue population itself or a system of issues. The major purpose of this analysis was to see if or not the crucial pair of groups (e.g., workers and executives in a social matter, engineers and medical doctors in a pollution problem) is persistently discordant each other on the issues in a specified field.

4.2 Multi-issue rank correlation regarding the relative ranks

The ranks are accordant regarding the relative ranks when an item steadily precedes (or is preceded by) another in their ranks on two ranking criteria. Let $\gamma_{i,ab,kk'} = 1$ if $c_{i,a,k} < (>) c_{i,b,k}$ and $c_{i,a,k'} < (>) c_{i,b,k'}$, and $\gamma_{i,ab,kk'} = -1$ otherwise where $c_{i,x,y}$ denotes the degree of consensus of pair x of groups on criterion y about issue i. Let $\gamma_{ab,kk'} = \sum_{i=1}^{n} \gamma_{i,ab,kk'}/n$ where n denotes the number of issues with the tied cases excluded. Clearly $\gamma_{ab,kk'}$ lies in $[-1, 1]$ and serves as a correlation coefficient. The null hypothesis could be H: issues and criteria are independent and pairs are indifferent. Under H, $\gamma_{ab,kk'}$ would be symmetrically distributed around $E\{\gamma_{ab,kk'}\} = 0$ with $Var\{\gamma_{ab,kk'}\} = 1/n$. Its density is generated by $(x^{-1/n} + x^{1/n})^n$. The distribution of $\gamma_{ab,kk'}\, n^{1/2}$ could be approximated by N(0, 1) for $n \geqq 10$ under H. The continuity correction of unity [15] may improve the approximation. This could provide the confidence level against H if the population could be conceived. Since H is compound, its rejection is ambiguously interpreted. The averaging method for tied cases lessens the variance with the mean unchanged [15].

Let $\gamma_{i,kk'} = \sum_a \sum_{b<a} \gamma_{i,ab,kk'}/p$ and $\gamma_{kk'} = \sum_{i=1}^{n} \gamma_{i,kk'}/n$ where p denotes the number of pairs of groups. Clearly $\gamma_{kk'}$ lies in [-1, 1] serving as a correlation coefficient. Under H, $\gamma_{kk'}$ would be symmetrically distributed around $E\{\gamma_{kk'}\} = 0$ with $Var\{\gamma_{kk'}\} = Var\{\gamma_{i,kk'}\}/n = 2(2p + 5)/\{9np(p - 1)\}$. Its generating function for p = 3 is $(x^{-1/n} + 2x^{-2/3n} + 2x^{2/3n} + x^{1/n})^n$. It is $(\sum_{s=-p}^{p} a_{s,p}\, x^{s/np})^n$ for $p > 3$ with $a_{s,p}$ obtained by the Kendall's asymptotic equation $a_{s,p} = a_{s-p,p-1} + a_{s-p+2,p-1} + \dots + a_{s+p,p-1}$ although p - 1 may not denote the number of pairs. Under H, the distribution of $3\gamma_{kk'}\{np(p - 1)\}^{1/2}/\{2(2p + 5)\}^{1/2}$ could be approximated by N(0, 1) for $n \geqq 10$ with some room for improvement by the continuity correction of unity. This rank correlation is the average of the Kendall's rank correlation across issues.

4.3 Multi-issue rank correlation regarding the absolute ranks

The ranks are accordant when an item occupies the same rank on two ranking criteria. Let $w_{i,a,k}$ be the rank value assumed by pair a on issue i on criterion k, and define $s_{i,a,kk'} = (w_{i,a,k} - w_{i,a,k'})^2$ and $s_{i,kk'} = \{6 \sum_a s_{i,a,kk'} - p(p + 1)(p - 1)\}/\{p(p + 1)(p-1)\}$ where p is the number of pairs of groups. Clearly $s_{i,kk'}$ lies in [-1, 1]. Under H, $s_{i,kk'}$ is symmetrically distributed around $E\{s_{i,kk'}\} = 0$ with $Var\{s_{i,kk'}\} = 1/(p - 1)$. Further define $s_{kk'} = \sum_{i=1}^{n} s_{i,kk'}/n$

where n denotes the number of issues with the tied cases excluded to which the averaging method may be applied to lessen the variance. Clearly $s_{kk'}$ lies in $[-1, 1]$, serving as a correlation coefficient. Under H, it would be symmetrically distributed around $E\{s_{kk'}\} = 0$ with $\text{Var}\{s_{kk'}\} = 1/\{n(p-1)\}$. Its generating function for p = 3 is $(x^{-1/n} + 2x^{-1/2n} + 2x^{1/2n} + x^{1/n})^n$. Under H, the distribution of $s_{kk'}\{n(p-1)\}^{1/2}$ could be approximated by N(0, 1) for $n \geqq 10$ with some room for improvement by the continuity correction of unity. This is the average of the Spearman's rank correlation [18] across issues.

4.4 Multi-issue rank correlation regarding the alliance and isolation

The closest pair tends to get allied and the group standing the furthest from the others tends to rebel against the whole. Let the pair and the group assume value one in the respective cases and be summed over issues. If the population is presumable, the hypothesis H would imply that their distributions are subject to B(n, 1/p) and B(n, 1/g) respectively where n, p and g denote the number of issues, pairs and groups respectively.

4.5 Multi-issue rank test

A pair of groups is sometimes hypothesized to occupy the specified rank. For example in a social forecasting, workers and executives are hypothesized to be the most distant. The values 1 and -1 are assigned according as the realized rank is above or below the hypothetical one, then the argument analogous to §4.2 holds. The value difference from the hypothetical rank is squared, then the argument like in §4.3 holds. The Page's method [24] is also applicable.

4.6 Results of multi-issue rank correlation analysis

No particular correlation was found among the three ranks each given by the three measures for pairs of groups on issues in the specified fields (Table 1). This result has a three-fold meaning as a result of the conventional single-issue rank correlation analysis has a two-fold meaning. Three-fold interpretation: (i) measures are not highly correlated, (ii) issues are not highly correlated in the specified field, i.e., there is no persistent discordance between particular groups on a series of problems, (iii) pairs of groups are not highly correlated, i.e., there is no particular opposition among groups. Which of these possible conclusions is correct was undecidable, but any of the three conclusions seemed desirable.

Under the hypothesis H these correlations could be standardized in terms of the probability and be compared each other. But in our experience they assumed nearly the same value and were not clearly compared.

5. VECTOR CORRELATION ANALYSIS

5.1 Opinion distribution as a vector

In the stimuli-responses frame in psychology, an issue can be considered to stimulate each group whose response is expressed in its opinion distribution as a sample. A discrete distribution is regarded as a vector whose dimension equals the number of alternatives in the questionnaire, and a vector correlation on distributions of two issues with groups as samples is regarded as a measure of distribution matching. Let f_{ijk} denote the opinion frequency on issue i at alternative j by group k ($j = 1, 2, \ldots, m$, $k = 1, 2, \ldots, K$) and let the m-vector correlation be measured between issue i and issue i' with K samples. The vector correlation by Hayashi [10] was used. The purpose of this analysis was to relate an issue with another.

5.2 Results of vector correlation analysis and discussions

The vector correlation analysis was applied to several fields in the technology forecasting with a remarkable success only in the pollution-proof technology. In this interesting result the similarity between issues were found interpretable as the similarity of pollution sources from an administrative point of view, not as the similarity of pollution materials from a physical or chemical point of view. More specifically the following pairs were found similar, problems in air-ports and sewage disposal plants (problems in public facilities); problems of autos and biological pollutedness (problems of moving objects); and problems of drainage systems and construction noises (problems of multi-source in the urban area).

The reason of unsuccesses was attributable to the unsharpness of discrimination by the vector correlation. Many pairs assumed the nearly same value of the correlation coefficient which was quite near unity. Ogawara [23] pointed out that the Hayashi's correlation tends to be near unity and established the relation

$$0 \leqq \rho_o \leqq \lambda_1 \leqq \rho_H \leqq 1$$

where ρ_o, λ_1 and ρ_H denote respectively the square root of the Ogawara's vector correlation [22], the well known canonical correlation and the Hayashi's vector correlation. Hence ρ_H may be recommendable only when the consensus is poor, which was not the case in the technology forecasting.

6. CONCLUSION

Introducing statistical concepts and applying statistical methods to opinion distributions of Delphi surveys were found to be meaningful and to disclose some useful information of a consensus for assessing technology or policies.

7. ACKNOWLEDGEMENT

The author thanks professor K. Matsui at Rikkyo University and Mr. T. Hama at the Ministry of Labour for their making the Delphi data accessible to him. He also thanks Mr. Hama, professor S. Kawasaki at Kobe Commerce University, professor K. Takeuchi and professor S. Moriguchi at the University of Tokyo, professor T. Inakoshi at Rissho University and Dr. S. Shimada at Hitachi Central Research Laboratory for their criticism on the earlier version of the paper.

8. REFERENCES

1. Bloxom, B., The Simplex in Pair Comparisons, Psychometrika, 37 (1972), 119-136
2. Chatterjee, S. K., Rank Procedures for Some Two-Population Multivariate Extended Classification Problems, J. Multivariate Analysis, 3 (1973), 26-56
3. Conover, W. J., A Kolmogorov Goodness of Fit Test for Discontinuous Distribution, J. Amer. Stat. Ass., 67 (1972), 591-596
4. Dalkey, N. C., Delphi, RAND P3704, 1967
5. ——————, Experiments in Group Prediction, RAND P3820, 1968
6. ——————, The Delphi Method: An Experimental Study of Group Opinion, RAND RM5888PR, 1969
7. ——————, B. Brown and S. W. Cochran, The Delphi Method IV: Effect of Percentile Feedback and Feed-in of Relevant Facts, RAND RM6118PR, 1970

8. Gibbons, J., NONPARAMETRIC STATISTICAL INFERENCE, McGraw Hill, 1971

9. Gordon, T. J. and O. Helmer, Report on a Long Range Forecasting Study, RAND P2982, 1964

10. Hayashi, C., I. Higuchi and T. Komazawa, JOHO SHORI TO TOKEI SURI (Data Processing and Statistical Mathematics), Sangyo Tosho, 1970

11. Helmer, O., Convergence of Expert Consensus Through Feedback, RAND P2973, 1964

12. ———— , The Use of the Delphi Technique in Problems of Educational Innovations, RAND P3499, 1966

13. ———— , Analysis of the Future; The Delphi method, RAND P3558, 1967

14. ———— , Systematic Use of Expert Opinions, RAND P3721, 1967

15. Kendall, M. G., RANK CORRELATION METHODS, New York, Haffner Pub. Co., 1955 (2nd ed.)

16. Komazawa, T. Informal Discussion on his Modified Qualification Theory, 1974

17. Linhart, H., Approximate test for m rankings, Biometrika, 47 (1960), 476-480

18. Lyerly, S. B., The Average Spearman Rank Correlation Coefficient, Psychometrika, 17 (1952), 421-428

19. Ministry of Labour of Japan, A Forecast of Workers Life in 1980's, 1973

20. Miyazaki Prefecture, Delphi Surveys on Goals of Development Planning, 1975

21. NHK (Japan Broadcasting Corporation), Social Forecasting of 1980's, 1974

22. Ogawara, M., On the Correlation between Vector Random Variables with Different Dimensions, Sugaku, 1, 216, 1947

23. ―――――― , A Note on the Vector Correlations, Science Report of Tokyo Woman's Christian College, 32, 386-392, 1975

24. Page, E. B., Ordered Hypothesis for Multiple Treatments: A Significance Test for Linear Ranks, J. Amer. Stat. Ass., 58 (1963), 216-230

25. Schucany, W. R. and W. H. Frawley, A Rank Test for Two Group Concordance, Psychometrika, 38 (1973), 249-258

26. Science and Technology Agency of Japan, Science & Technology Developments upto A. D. 2000, JATES, 1971.

27. Thompson, W. A. Jr. and Willke, T. A., On an Extreme Rank Sum Test for Outliers, Biometrika, 50 (1963), 375-383

issue	measure	pair of groups			
		$g_1\ g_2$	$g_1\ g_2$	.....	$g_{G-1}\ g_G$
1	1	1	2		6
	2	6	3		2
	3	4	1		5
⋮		⋮	⋮		⋮
i	1	:	:		:
	2	:	:		:
	3	:	:		:
⋮		⋮	⋮		⋮
n	1	:	:		:
	2	:	:		:
	3	:	:		:

TABLE 1. Multi-issue Ranking

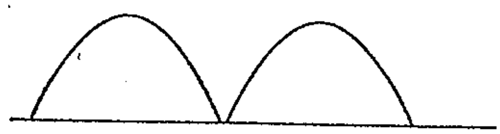

FIGURE 1 Separated destributions

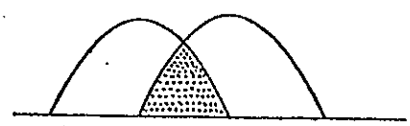

FIGURE 2 Penetrating distributions