NORMAL MEANS REVISITED

by

John Deely
Purdue University

and

Wes Johnson
University of California
Davis

Technical Report #97-01

Department of Statistics
Purdue University
West Lafayette, IN USA

January 1997

◇

1

Normal Means Revisited

John Deely and Wes Johnson

Purdue University, West Lafayette, IN, USA
University of California, Davis, CA, USA

Abstract.

The problem of ranking and selecting normal means as originally studied by Shanti Gupta is approached herein from a robust Bayesian perspective. This model uses the usual hierarchical Bayesian setup but does not require complete specifications of the hyperpriors. Instead, elicited prior information about the population of means as a group is used to specify a quantile class for the hyperpriors. Two criteria are suggested for ranking and selecting and provide insight not only to which population is best, but in addition give quantitative methods for deciding how much better one population is than another. Using these criteria, minimum and maximum values are calculated for the derived quantile class. Relative sizes of these evaluations and the distance between the max and min give insight as to the quality of the data and the sufficiency of the sample size. These concepts are illustrated with a numerical example.

Keywords and phrases:

normal means, hierarchical Bayesian, ranking and selection, robust Bayesian, quantile class, predictive distribution.

1.1 Introduction

Shanti Gupta began a career in Ranking and Selection Methods when he wrote his thesis "On a Decision Rule for a Problem in Ranking Means" in 1956. His new approach to the selection problem was to derive procedures which selected a subset of "random" size in such a way that the probability of obtaining the "best" population was greater than a specified level. This formulation is to be contrasted to that initiated by Bechhofer (1954) in which a fixed size subset

(generally size one) was selected. The basic problem with this new random approach involved deriving selection procedures which had certain desirable properties; namely the selected subset should include the "best" population with reasonably high probability and the size of the selected subset should be as small as possible while still assuring the probability requirement. Thus began the quest to find an optimal selection procedure "t" which minimized $E[S|\underline{\vartheta}, t]$ subject to satisfying $P(CS|\underline{\vartheta}, t) \geq P^*$ where $\underline{\vartheta}$ is the k vector of unknown means, "t" denotes the particular selection procedure used and P^* is a pre-specified number close to unity.

It turns out that there is no one optimal solution for this problem. In fact it has no solution in that context because for one configuration of the parameter space one procedure does better than another and vice versa for another configuration. See for example Paulson (1952) and Seal (1955). This impossibility thus lead to an enormous amount of work which dealt with various formulations and models and their correspondent procedures. The fact that this area of research has flourished so much is due in no small measure to the influence and encouragement Shanti provided. He was always at the forefront encouraging development in a wide variety of areas directions. In that regard it should be noted that both of the computations above were made initially in the frequentist sense (as though the problem facing the practitioner were going to be encountered over and over again under exactly the same experimental conditions) since modern Bayesian theory was just beginning to surface at that time. A classic paper by Dunnett (1960) was the first paper to deal with selection problems using a Bayesian flavor. But Shanti's perspective was so broad that all approaches were supported. In fact at a very early stage, 1965, he even encouraged his first Ph.D. student (the first author) to write a thesis on *Empirical Bayes Multiple Selection Procedures.* Since that time a number of Bayesian and empirical Bayesian papers mainly from a decision theoretic point of view have appeared (see for example Goel and Rubin (1977)). A recent paper by Berger and Deely (1988 does develop a more practical approach to this problem.

But in spite of all this work, both frequentist and Bayesian, on the normal means problem there has not been a full scale adoption and application of them in the practical world. Practitioners are still using the old fashioned but more importantly inadequate AOV type analysis of data much of which really requires a procedure to rank the means. Of course the computer facilities for these methods are well developed compared to the frequentist or Bayesian ranking procedures. In addition frequentist ranking procedures require statement about the parameter space that may not always be practically realizable and whereas the Bayesian formulation is more practical, there is still the problem in that approach with the prior or lack of it for die hard frequentists.

The notion of exchangeability amongst the means and the hierarchical Bayesian formulation adds another dimension to the problem. Since the equipments,

suppliers or processes generating the means to be compared are assumed to be somewhat similar, we treat the population means as exchangeable, an assumption which is conveniently modeled through a HB setup. Thus the main feature of this approach is that it facilitates in a much more practical way the use of the type of prior information that practitioners are likely to have available; namely, information about the **group** of means as opposed to information about specific **individual** members of the group.

The first phase of this approach is contained in a paper by Berger and Deely. With the advent of the MCMC methods with emphasis on the Gibbs sampler a more general approach than taken in that paper is now computationally feasible. Even so the HB approach uses either a non-informative or an elicited subjective hyperprior both of which may not be consistent with the type of prior information readily available about the group of means being studied. Specifically, it may be that the available prior information does not allow complete specification of the hyperpriors but on the other hand should not be ignored as in the non-informative case. Our approach to the hyperpriors assumes that they are determined **only** up to a **family** of distributions which depend on the available prior information. With this type of model and consequent analysis we believe we are providing the practitioner with tools effectively use the type of prior information that might be available in many situations; namely information about the populations as a group as opposed to information about individual populations.

Thus specification of a prior on the unknown population means using the HB model can easily incorporate this kind of information can be modelled using Hierarchical Bayes. The prior is a mixture of conditional distributions with the mixing distribution determined by what is known about the group of populations. Hence for the normal means problem we can think of the individual means coming from a normal distribution with mean β and variance τ^2 which are distributed according to some hyperprior $h(\beta, \tau^2)$ where h is determined by the prior information. The robust Bayesian approach we take in this paper is to relax the requirement that h has to be completely specified to an assumption that the type of prior information available allows specification of h to below to the Quantile class of distributions.

Specifically, we assume this family is the quantile class, (cf. Lavine (1991)) where the elicited prior information specifies particular quantiles but nothing more about the hyperpriors. Using a technique from Lavine (1991), we then compute maximum and minimum values for the above criteria where these extrema are taken over the family of hyperpriors specified by the prior information. The utility of the prior information is assessed by computing the **difference** between the maximum and the minimum of the probabilities thus obtained; a small difference indicating a useful inference whereas large differences would be meaningless and thus not very useful. The effect of the sample size and the observed test data on inferences is assessed by consideration of the **magnitudes**

of the criterion probabilities; for example, values of these probabilities near unity would indicate that the sample size and test data were very effective in determining a "best" member of the entire group, whereas small values would indicate ineffective test data. It is to be emphasized that a major implication of our treatment of the hyperpriors is that the prior distribution on the means implied through the HB model need not be completely specified. It is this latter feature which utilizes whatever partial prior information is available and in this sense guarantees the robustness of the suggested ranking procedures over all priors satisfying such information.

Finally, it should be noted that another feature of the Bayesian approach is the fact that the ranking criteria not only gives the ranks of the means but in addition can be used to determine **how much better** one population is than another. Further amplification of this feature is made in **Section 2** where we define and discuss two specific criteria to be used in the ranking process. In **Section 3** the details of the robust procedures are developed while numerical examples illustrating our methods are given in **Section 4**

1.2 Selection Criteria

The suggested ranking criteria is based on Bayesian concepts arising from the posterior distribution. We focus on two concepts here:
Criterion 1: the posterior probability that any one of the means is larger than all of the others by an amount "b", i.e. compute for each $i = 1, 2, \ldots, k$ the quantity

$$P_i = P(\vartheta_i \geq \vartheta_j + b \text{ for ALL } j \neq i | \text{ data})$$

where b is a non-negative specified constant;
Criterion 2: the predictive distribution that any one of these populations will have a larger observation than all the others by an amount "c", i.e. compute for each $i = 1, 2, \ldots, k$ the quantity

$$PR_i = P(Y_i \geq Y_j + c \text{ for ALL } j \neq i | \text{ data})$$

where Y_i is a new observation from the ith population and c is a non-negative specified constant.

Closer examination of the proposed ranking criteria reveals how they can be used to make quantitative rankings among the k individuals in the particular group being studied. Firstly, by computing each P_i and PR_i for $i = 1, 2, \ldots, k$ we can compare each member to all of the individuals remining in the **total** group; from these computations **subgroups** for further comparisons may be suggested. That is, suppose P_i and PR_i are very close to unity for a particular 'i'; this would indicate we have conclusively found the best amongst all k members. However when this is not the case it may be that a subgroup of just a few members may have their sum of P_i or PR_i very large in which case

we would then be interested in comparisons amongst that subgroup only. It is easily seen that such iterations might eventually lead to simply a comparison of just two members of the original group of k. In general, we allow for the possibility of finding subgroups and making comparisons within those subgroups which includes ultimately all possible pairwise comparisons. There are clearly many possibilities each of which can be computed as required without effecting the validity of any other calculation. In addition a salient feature of these computations is that the **degree of how much better** one member is than any other in the particular group being compared, be it one other or many, can be assessed by varying the quantities "b" and "c" in the formulas. This process lends itself to a type of "OC" analysis by plotting P_i against $\boldsymbol{b}$ and PR_i against $\boldsymbol{c}$. Note that Criterion 1 reduces to the Bayesian Probability of Correct Selection (PCS) when $b = 0$ but the fact that we allow b to take on positive values is an important improvement over the usual PCS criterion. Specific examples of these concepts will be given in Section 4.

The proposed criteria above are not new. The first criterion has been discussed extensively in Berger and Deely (1988) for the problem of ranking normal means. The second criterion has been treated in a general context by Geisser (1971) and more specifically in Geisser and Johnson (1994).

1.3 Model and Computations

Let X_j, the sample mean for the jth population based on n_j observations, be normally distributed with unknown mean θ_j and known variance σ_j^2/n_j. The prior on $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ will be described by a two stage process as in the usual hierarchical Bayesian model. For the first stage let $\theta_1, \ldots, \theta_k$ be conditionally i.i.d. with a normal distribution with mean β and variance τ^2. For the second stage we assume only that β and τ^2 are independent with distributions h_1 and h_2 which are known to belong to a specific quantile class. This class is determined by the prior information available about the unknown means $\theta_1, \ldots, \theta_k$. Further discussion on this point will be made in the next section.

We can now proceed with the computational forms for Criteria 1 and 2. Firstly it will be helpful to adopt the following notation. Let

$$A_i = \{\vartheta_i \geq \vartheta_j + b \text{ for all } j \neq i\} \text{ and } B_i = \{Y_i \geq Y_j + c \text{ for all } j \neq i\}.$$

From our model it follows that conditional upon β and τ^2 the posterior pdf of θ_j is a normal distribution denoted by $\pi(\vartheta_j|x_j, \beta, \tau^2)$ with mean $\text{mpos}_j = \alpha_j x_j + (1 - \alpha_j)\beta$ and variance $\text{vpos}_j = \alpha_j \sigma_j^2/n_j$ where $\alpha_j = \tau^2\{\tau^2 + \sigma_j^2/n_j\}^{-1}$. We can then write the conditional joint posterior pdf of $\underline{\theta}$ as

$$\pi(\underline{\vartheta}|\underline{x}, \beta, \tau^2) = \prod_{j=1}^{k} \pi(\vartheta_j|x_j, \beta, \tau^2).$$

It will be convenient to denote the univariate normal cdf and pdf with mean μ and variance v by $G(\bullet|\mu, v)$ and $g(\bullet|\mu, v)$ respectively. In addition we require the following form for the posterior distribution of β and τ^2 denoted by $h(\beta, \tau^2|\underline{x})$:

$$h(\beta, \tau^2|\underline{x}) = \frac{f(\underline{x}|\beta, \tau^2) \cdot h(\beta, \tau^2)}{\int f(\underline{x}|\beta, \tau^2) \cdot h(\beta, \tau^2) d\beta d\tau^2} \tag{1.1}$$

where $h(\beta, \tau^2)$ denotes the hyperprior on β and τ^2. We will not assume a specific form for $h(\beta, \tau^2)$ but we will be able to use prior information about the group of populations to locate $h(\beta, \tau^2)$ in the quantile class. Let $\boldsymbol{H}$ denote the so elicited class of hyperpriors. For the purposes of this paper we will assume that the form of the prior information can be interpreted as follows: $h(\beta, \tau^2) = h_1(\beta) \cdot h_2(\tau^2)$ and that for β and τ^2 respectively we are able to ascertain three regions R_1, R_2, R_3 and S_1, S_2, S_3 with respective prior probabilities p_1, p_2, p_3 and q_1, q_2, q_3. Thus

$$\boldsymbol{H} = \{h\colon \int_{R_j} h_1(\beta) d\beta = p_j \text{ and } \int_{S_j} h_2(\tau^2) d\tau^2 = q_j \text{ for } j = 1, 2, 3\}$$

Criterion 1: Using the above notation, we can then write

$$P_i = P(A_i|\underline{x}) = \int P(A_i|\underline{x}, \beta, \tau^2) h(\beta, \tau^2|\underline{x}) d\beta d\tau^2 \tag{1.2}$$

where

$$P(A_i|\underline{x}, \beta, \tau^2) = \int P(A_i|\underline{x}, \beta, \tau^2, \vartheta_i) g(\vartheta_i|\text{mpos}_i, \text{vpos}_i) d\vartheta_i \tag{1.3}$$

$$= \int_0^\infty \left\{ \prod_{j \neq i} G(\vartheta_i - b|\text{mpos}_j, \text{vpos}_j) \right\} g(\vartheta_i|\text{mpos}_i, \text{vpos}_i) d\vartheta_i.$$

Criterion 2: For this criterion we firstly note that the predictive distribution of Y_j given β and τ^2 is normal with mean mpos_j and variance $u_j = \sigma_j^2 + \text{vpos}_j$. Thus we can write

$$PR_{ij} = P(B_i|\underline{x}) = \int P(B_i|\underline{x}, \beta, \tau^2) h(\beta, \tau^2)|\underline{x}) d\beta d\tau^2 \tag{1.4}$$

where

$$P(B_i|\underline{x}, \beta, \tau^2) = \int P(B_i|\underline{x}, \beta, \tau^2, y_i) g(y_i|\text{mpos}_i, u_i) dy_i \tag{1.5}$$

$$= \int_0^\infty \left\{ \prod_{j \neq i} G(y_i - b|\text{mpos}_j, u_j) \right\} g(y_i|\text{mpos}_i, u_i) dy_i.$$

Letting RC denote "Ranking Criterion" and using (1.1), we can then write both (1.2) and (1.4) as a function of the hyperprior h as

$$RC(h) = \frac{\int P(\beta, \tau^2) \cdot L(\beta, \tau^2) \cdot h(\beta, \tau^2) d\beta d\tau^2}{\int L(\beta, \tau^2) \cdot h(\beta, \tau^2) d\beta d\tau} = \frac{N(h)}{D(h)} \tag{1.6}$$

where $P(\beta, \tau^2)$ is given by (1.3) and (1.5) respectively for Criteria 1 and 2 and $L(\beta, \tau^2)$ is the second stage likelihood function given by

$$L(\beta, \tau^2) = f(\underline{x}|\beta, \tau^2) = \int \prod_{j=1}^{k} g(x_j|\vartheta_j, \sigma_j^2/n_j) \cdot g(\vartheta_j|\beta, \tau^2) d\vartheta_j \quad (1.7)$$

$$= \prod_{j=1}^{k} g(x_j|\beta, (\sigma_j^2/n_j) + \tau^2)$$

Our goal is to find maximum and minimum values for $RC(h)$ over the family $\boldsymbol{H}$. To accomplish this, we use the form (1.6) and invoke the Linearization Principle made popular recently by Lavine (1991) which allows us to write:

$$\max_{h \in H} \frac{N(h)}{D(h)} = \bar{a} \text{ iff } \bar{a} = \min\{a\colon \max_h [N(h) - a \cdot D(h)] \leq 0\} \quad (1.8)$$

$$\min_{h \in H} \frac{N(h)}{D(h)} = \underline{a} \text{ iff } \underline{a} = \min\{a\colon \max_h [N(h) - a \cdot D(h)] \leq 0\} \quad (1.9)$$

The value of this principle can be appreciated by observing the fact that for a given "a" the quantities $\max\{N(h) - aD(h)]\colon h \in \boldsymbol{H}\}$ and $\min\{[N(h) - aD(h)]\colon h \in \boldsymbol{H}\}$ are easily computed. Specifically,

$$\begin{aligned} \max_h [N(h) - a \cdot D(h)] &= \max_h \int \{[P(\beta, \tau^2) - a] L(\beta, \tau^2)\} h(\beta, \tau^2) d\beta d\tau^2 \\ &= \sum_{i=1}^{3} \sum_{j=1}^{3} \overline{PL}_{ij}(a) p_i q_j \end{aligned}$$

where

$$\overline{PL}_{ij}(a) = \max\{[P(\beta, \tau^2) - a] L(\beta, \tau^2)\colon (\beta, \tau^2) \in R_i \cap S_j\}.$$

A similar calculation is used to obtain

$$\underline{PL}_{ij}(a) = \min\{[P(\beta, \tau^2) - a] L(\beta, \tau^2)\colon (\beta, \tau^2) \in R_i \cap S_j\}.$$

By noting the forms of the functions P (in either (1.3) or (1.5)) and L in (1.7), it can be seen that for any given "a", values of $\overline{PL}_{ij}(a)$ and $\underline{PL}_{ij}(a)$ are easily obtained numerically. This in turn leads to the computation of the desired values $\bar{a}$ and $\underline{a}$. The numerical example in Section 4 illustrates these computations.

1.4 Numerical Example

Consider the example given in Moore and McCabe (1993) (p. 756 Ex. 10.18) concerning a study of the effects of exercise on physiological and psychological

variables. For illustrative purposes here we focus on the psychological data which is summarized in Table 1. The Treatment consisted of a planned exercise program, the Control group were average type people while the Sedentary group were chosen by their inactivity. High scores indicate more depressed than low scores and here we are interested in computing which group is most (least) depressed and by how much. Thus if we think of θ_j as the unknown mean depression of the jth group, then we can use either Criteria described in Section 1 to understand the effect of exercise (or lack thereof) on depression.

Table 1

Group	n	sample mean	sample stdev
Treatment (T)	10	51.9	6.42
Control (C)	5	57.4	10.46
Joggers (J)	11	49.7	6.27
Sedentary (S)	10	58.2	9.49

Firstly we want to consider the value of prior information about the group as opposed to information about each population. Of course the type of prior information available in any particular situation can be very different. Here we are simply indicating just one of many different elicitation scenarios. In any case we do assume that the available prior information will be specific enough to indicate a particular quantile class which will then be used for the hyperpriors. For purposes of illustration consider eliciting answers to the following question.

1. What is the smallest interval which contains all of the unknown means θ_i's? Ans. (35,70)

2. What is the smallest interval containing the average of the θ_i's? Ans. (45,60)

3. How confident are you that the distance between the maximum θ_i and the minimum θ_i is at least 10% of the range given in (1), i.e. the difference between the maximum and the minimum depression scores will be at least 3.5 units on the scale of measurements to be used. Ans. 90%

For Questions (1) and (2) we assume that we can place a probability of at least 0.95 on the answers. Using this elicited information we can obtain regions in the (β, τ^2) space with their respective probabilities which then specifies the particular quantile class to be used for the hyperpriors on β and τ^2. Thus from (2) we have that $P\{45 < \beta < 60\} = .95$ since β is the prior mean of the θ_i's. This in gives the three intervals R_1, R_2, R_3 on β as (0,45), (45,60) and (60,∞) with their respective probabilities of 0.025, 0.95, 0.025.
and with respective prior probabilities p_1, p_2, p_3 and q_1, q_2, q_3. Thus

From (1) we can write $\boldsymbol{P}\{\theta_{[k]}| - \theta_{[1]} < 35\} = .95$ and from (3) we have $\boldsymbol{P}\{\theta_{[k]} - \theta_{[1]} \geq 3.5\} = .9$. These two expressions can be used to obtain probability intervals for τ^2 by noting that

$$\boldsymbol{P}\{\theta_{[k]} - \theta_{[1]} < c\} = \boldsymbol{P}\{T - S < c/\tau\}$$

where T and S are the maximum and minimum respectively of k standard normal rv's. Using $k = 4$ in Figure 1 with (1) and (3) gives the following results:

$$0.95 = \boldsymbol{P}\{35/\tau \geq 3.65\} = \boldsymbol{P}\{\tau^2 < 100\}$$

and

$$0.9 = \boldsymbol{P}\{3.5/\tau < 1.1\} = \boldsymbol{P}\{\tau^2 \geq 10\}.$$

This in turn gives three intervals S_1, S_2, S_3 on τ^2 as (0,10), (10,100) and (100,∞) with their respective probabilities of 0.1, 0.85, 0.05.

Putting the β and τ^2 intervals together gives nine regions with their correspondent probabilities. Thus we have determined which quantile class describes the hyperpriors for the elicited information. Table 2 indicates symbolically the regions and their respective probabilities.

Table 2

β	$R_3 = (60, \infty)$	0.0025	0.0213	0.0013
	$R_2 = (45, 60)$	0.0950	0.8075	0.0475
	$R_1 = (0.45)$	0.0025	0.0213	0.0013
		$S_1 = (0, 10)$	$S_2 = (10, 100)$	$S_3 = (100, \infty)$
			τ^2	

For this configuration the solutions to (1.8) and (1.9) when using Criterion 1 yield the (min, max) interval as (0.4575 0.6919). An indication of the sensitivity to these values can be seen by changing the probabilities to

.0025	.0025	.0025
.0025	.98	.0025
.0025	.0025	.0025

We obtain (0.489, 0.702). If we keep the original probabilities, but tighten up the regions to make the middle interval on τ^2 to be (30, 40), then the (computation gives (0.490, 0.697). Using this configuration and the increased probabilities we obtain (0.5054 0.6968). Changing the middle interval for τ^2 to (100, 500) and the new probabilities produces (.5249, .6294). Thus it can be seen that the type and size of the prior information has an effect but there is a general agreement amongst the resulting computations. In particular when comparing all four populations it is not overwhelmingly true that the Sedentary group

suffers the most depression. This is not surprising when looking at the values for the Control group. However when comparing the two populations, Sedentary and Joggers, there is overwhelming evidence that the Sedentary group are much more depressed than the Joggers. This data is reported in Table 3.

Table 3

		Criterion 1			Criterion 2	
	b	min	max	c	min	max
	0	0.827	0.993	0	0.744	0.953
sample	1	0.640	0.984	1	0.574	0.929
sizes	2	0.406	0.966	2	0.487	0.898
10,11	5	0.101	0.829	5	0.244	0.746
	0	1.000	1.000	0	1.000	1.000
sample	5	0.978	0.999	5	0.937	0.985
sizes	8	0.304	0.656	8	0.362	0.612
100,110	10	0.016	0.083	10	0.060	0.164

In addition it can be seen in Table 3 just how much more depressed the Sedentary group is. For example, when using Criterion 1 we can say that there is at least one unit difference with posterior probability between 0.640 and 0.984. Another aspect of these computations is illustrated by noting the effect of the sample size. The lower part of Table 3 has been calculated assuming the data for these two populations had been based on sample sizes of 100 and 110. If this had been the case then the data would have indicated there was a five unit difference with probability between 0.978 and 0.999.

Many other comparisons and calculations are suggested by the above. The computations are easily performed and a wide variety of inferences possible. Our purpose here through these brief illustrations has been to indicate how this methodology can be easily applied and how the results obtained give valuable new insight into the normal means ranking and selection problem.

References

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25**, 16–39.

Berger, J. O. and Deely, J. J. (1988). A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology. *J. Amer. Statist. Assoc.*, **83**, 364–373.

Dunnett, C. W. (1960). On selecting the largest of k normal population means (with discussion). *J. Roy. Statist. Soc., Ser. B* **22**, 1–40.

Geisser, S. (1971). The inferential use of predictive distributions. *In Foundations of Statistical Inference*, V. P. Godambe and D. A. Sprott (Eds.) Holt, Rinehart and Winston, Toronto.

Geisser,S. and Johnson, W. O. (1996). Sample size considerations in multivariate normal classification. *In Bayesian Analysis of Statistics and Econometrics*, Eds. D. A. Berry, K. M. Chaloner and J. W. Geweke, Wiley, New York.

Goel, P. K., and Rubin, H. (1977). On selecting a subset containing the best population — A Bayesian approach. *Ann. Statist.* **5**, 969–983.

Lavine, M. (1991). An approach to robust Bayesian analysis for multidimensional parameter spaces. *J. Amer. Statist. Assoc.* **86**, 400–403.

Moore, D. S. and McCabe,G. P. (1993). *Introduction to the Practice of Statistics* (second Edition) W. H. Freeman and Company, New York.

Paulson, E. (1952). On the comparison of several experimental categories with a control. *Ann. Math. Statist.* **23**, 239–246.

Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. *Ann. Math. Statist.* **26**, 387–398.