# ON A CLASS OF DECISION PROCEDURES FOR RANKING NORMAL POPULATIONS ACCORDING TO THEIR VARIANCES

BY M. N. KAPUR

*Indian Council of Agricultural Research, New Delhi*

GIVEN a finite set of normal populations Seal[1] has given a class of decision procedures for choosing the smallest group of populations which includes the population with the greatest mean. In this paper a similar problem for the variances has been considered.

Let $\pi_i = N(\mu_i, \sigma_i)$, $i = 0, 1, 2, \cdots k$ be $k + 1$ normal populations with means $\mu_i$ and variances $\sigma_i^2$. It is desired to divide these $k + 1$ populations into two groups to be called groups 1 and 2 such that group 1 contains the population with the largest variance on the basis of one random sample drawn from each of the $k + 1$ populations. Let $\bar{x}_i$ and $s_i^2$ be the sample mean and variance respectively from the population $\pi_i$. Define the following class of procedures D:—"Exclude the population $\pi_0$ from group 1 if

$$\sum_{i=1}^{k} \frac{c_i s_{(i)}^2}{s_0^2} > F_a^{c_1, c_2, \ldots, c_k},$$

where $c_i$'s are constants such that

$$c_i \geqslant 0, \sum_{i=1}^{k} c_i = 1$$

and include $\pi_0$ in group 1 otherwise. $s_{(1)}^2 > s_{(2)}^2 \ldots, > s_{(k)}^2$ are ranked $s_i^2$ and $F_a^{c_1, c_2, \ldots, c_k}$ denotes the upper $100\,\alpha\%$ point in the probability density function (p.d.f.) of

$$F(c_1, c_2, \ldots c_k) = \sum_{i=1}^{k} \frac{c_i s_{(i)}^2}{s_0^2}$$

under the assumption of equality of $\sigma_i$'s, $i = 0, 1, 2, \ldots, k$. Proceed as such for each of the populations $\pi_i$ so that it plays the part of $\pi_0$". It is shown that this class of decision rules possess the following desirable properties (similar to those proved by Seal for the problem of means).

(*a*) The property of unbiasedness, *i.e.*, the probability of excluding any population not with the largest variance from group 1 is not less than the probability of excluding the population with the largest variance.

(*b*) The property of gradation, *i.e.*, corresponding to any $\gamma$ $(0 < \gamma < 1)$ there exists a constant $\sigma^2_{0\gamma}$ such that the chance of retaining a population with variance $\sigma_0^2$ in group 1 is greater or less than $\gamma$ according as $\sigma_0^2$ is greater or less than $\sigma^2_{0\gamma}$.

As a particular case if among the $k + 1$ populations, all populations except one are equal in their variances, the procedure '*d*' in class D, defined by

$$c_i = \frac{1}{k}, \; i = 1, 2, \ldots, k$$

is optimum in the sense that it (i) maximises the probability of retaining in group 1 the population with the unequal variance if this is larger than the common variance of the remaining $k$ populations, and (ii) maximises the probability of not retaining the population with the unequal variance in group 1 if this is smaller than the common variance of the remaining $k$ populations.

Proofs follow closely on the lines adopted by Seal[1] for the case of grouping of means except when the optimum properties of '*d*' are to be established. To prove these properties the following lemmas will be first established.

LEMMA 1.—Let

$$F\left(\frac{s_1^2}{\sigma_1^2}, \frac{s_2^2}{\sigma_2^2}, \ldots, \frac{s_k^2}{\sigma_k^2}\right)$$

be the joint cumulative distribution function (c.d.f.) of $k$ variables

$$s_i^2, \; i = 1, 2, \ldots, k$$

and

$$T(u_1^2, u_2^2, \ldots, u_k^2)$$

be a real valued function of

$$u_i^2, \; i = 1, 2, \ldots, k$$

such that

$$T(u_1^2\alpha_1^2, u_2^2\alpha_2^2, \ldots, u_k^2\alpha_k^2) \geqslant T(u_1^2, u_2^2, \ldots, u_k^2)$$

where

$$(\alpha_1^2, \alpha_2^2, \ldots, \alpha_k^2)$$

is a set of real numbers such that

$$\alpha_i^2 \geqslant 1, \; i = 1, 2, \ldots, k$$

and if for an arbitary constant $c$,

$$P_r [T(s_1^2, s_2^2, \ldots s_k^2) > c/\sigma_1^2, \ldots \sigma_k^2]$$

denotes the probability of

$$T(s_1^2, s_2^2, \ldots s_k^2) > c,$$

then

$$P_r [T(s_1^2, s_2^2, \ldots s_k^2) > c/\alpha_1^2\sigma_1^2, \ldots, \alpha_k^2\sigma_k^2]$$
$$\geqslant P_r [T(s_1^2, s_2^2, \ldots s_k^2) > c/\sigma_1^2, \ldots, \sigma_k^2].$$

*Proof.*—

$$P_r [T(s_1^2, s_2^2, \ldots, s_k^2) > c/(\alpha_1^2\sigma_1^2, \ldots, \alpha_k^2\sigma_k^2)]$$
$$= P_r [T(s_1^2\alpha_1^2, \ldots, s_k^2\alpha_k^2) > c/(\sigma_1^2, \ldots, \sigma_k^2)]$$
$$\geqslant P_r [T(s_1^2, s_2^2, \ldots, s_k^2) > c/(\sigma_1^2 \ldots, \sigma_k^2)].$$

Since

$$T(s_1^2\alpha_1^2, s_2^2\alpha_2^2, \cdots, s_k^2\alpha_k^2) \geqslant T(s_1^2, s_2^2, \ldots, s_k^2).$$

1·1 Cor. 1.—If

$$T(u_1^2\alpha_1^2, \ldots, u_k^2\alpha_k^2) > T(u_1^2, \ldots, u_k^2)$$

when $\alpha_i^2 \geqslant 1$ and $\alpha_i^2 > 1$ for at least one $i, 1 < i < k$ and if the c.d.f. of

$$T(s_1^2, \ldots, s_k^2)$$

assigns a positive measure to every non-degenerate interval then

$$P_r [T(s_1^2, \ldots, s_k^2) > c]$$

where $c$ is an arbitary constant, is an increasing function for each

$$\sigma_i^2, \; i = 1, 2, \ldots, k.$$

2. LEMMA 2.—

$$P_r\left[\sum_{i=1}^{k} \frac{c_i s_{(i)}^2}{s_0^2} > F_a^{c_1, c_2, \ldots, c_k}\right]$$

is an increasing function for each

$$\frac{\sigma_i^2}{\sigma_0^2},\ i = 1, 2, \ldots, k.$$

*Proof.*—Let

$$s_i'^2 = \frac{s_i^2}{\sigma_0^2},\ i = 0, 1, 2, \ldots, k$$

then

$$P_r\left[\sum_{i=1}^{k} \frac{c_i s_{(i)}^2}{s_0^2} > F_a^{c_1, c_2, \ldots, c_k}\right]$$

$$= P_r\left[\sum_{i=1}^{k} c_i s_{(i)}'^2 > s_0'^2 F_a^{c_1, c_2, \ldots, c_k}\right].$$

For fixed $s_0'^2$

$$P_r\left[\sum_{i=1}^{k} c_i s_{(i)}'^2 > s_0'^2 F_a^{c_1, c_2, \ldots, c_k}\right]$$

is an increasing function of each $\lambda_i = \sigma_i^2/\sigma_0^2$ by cor. 1. Since the distribution of $s_0'^2$ does not involve $\lambda_i$, it is now obvious that the (unconditional) value of

$$P_r\left[\sum_{i=1}^{k} \frac{c_i s_{(i)}^2}{s_0^2} > F_a^{c_1, c_2, \ldots, c_k}\right]$$

is an increasing function for $\lambda_i$'s.

3. *Property of unbiasedness.*—The probability of excluding any population not with the largest variance from group 1 is never less than the probability of excluding the population with the largest variance from group 1.

Let

$$\sigma_{[0]}^2 > \sigma_{[1]}^2 \geqslant \sigma_{[2]}^2 \geqslant \ldots \geqslant \sigma_{[k]}^2$$

be the ranked variances of the given $k+1$ populations. Probability of excluding the population with the largest variance $\sigma_{[0]}^2$ is defined as

$$P_r\left[\sum_{i=1}^{k} c_i s_{(i)}^2 > s_0^2 F_\alpha^{c_1, c_2, \ldots, c_k}\right]$$

and will depend on

$$P_{c_1, c_2, \ldots, c_k}\left(\frac{\sigma_{[1]}^2}{\sigma_{[0]}^2}, \frac{\sigma_{[2]}^2}{\sigma_{[0]}^2}, \ldots, \frac{\sigma_{[k]}^2}{\sigma_{[0]}^2}\right). \tag{i}$$

Similarly the probability of excluding the population with variance $\sigma_{[i]}^2$ other than the largest variance $\sigma_{[0]}^2$, will depend on:

$$P_{c_1, c_2, \ldots, c_k}\left(\frac{\sigma_{[0]}^2}{\sigma_{[i]}^2}, \frac{\sigma_{[1]}^2}{\sigma_{[i]}^2}, \frac{\sigma_{[2]}^2}{\sigma_{[i]}^2}, \ldots, \frac{\sigma_{[k]}^2}{\sigma_{[i]}^2}\right) \tag{ii}$$

Comparing (i) and (ii) we notice that

$$\frac{\sigma_{[0]}^2}{\sigma_{[i]}^2} \geqslant \frac{\sigma_{[j]}^2}{\sigma_{[0]}^2}, \quad \frac{\sigma_{[0]}^2}{\sigma_{[i]}^2} \geqslant \frac{\sigma_{[i]}^2}{\sigma_{[0]}^2}, \quad i = 1, 2, \ldots k, \; j = 1, 2, \ldots, k$$

$$i \neq j.$$

This enables us to make one to one correspondence between $k$ arguments of $P_{c_1, c_2, \ldots, c_k}$, such that no argument in (ii) is less than the corresponding argument in (i). Hence from the monotonic behaviour of $P_{c_1, c_2, \ldots, c_k}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ with regard to $\lambda_i$ (Lemma 2) it follows that the probability of excluding any population not with the largest variance from group 1 is not less than the probability of excluding the population with the largest variance. This proves property (*a*) of the class 'D'.

As no argument in (i) is greater than that in $P_{c_1, \ldots, c_k}(1, 1, 1, \ldots, 1)$, it further shows that the probability of excluding the population with the largest variance from group 1 will not exceed the significance level $\alpha$; *i.e.*, $\alpha$ will be the least upper bound of the probability of incorrect choice.

*Property of gradation.*—From Lemma 2 it follows that

$$P_r\left(\sum_{1}^{k} \frac{c_i s_{(i)}^2}{s_0^2} > F_\alpha^{c_1, c_2, \ldots, c_k}\right)$$

is a decreasing function of $\sigma_0^2$. It is shown as follows that this probability approaches 0 when $\sigma_0^2 \to \infty$ and 1 when $\sigma_0^2 \to 0$. Because of the monotonic behaviour of this probability it is enough to show that the property holds for

$$\sigma_i^2 = \sigma^2, \; i = 1, 2, \ldots, k.$$

It can be easily seen that p.d.f. of ($\sum_1^k c_i s^2_{(i)}/s_0^2$, denote by $u$) is a function of $(u\,\sigma_0^2/\sigma^2)$, therefore

$$\underset{\sigma_0^2 \to 0}{Lt}\ P_r\,[u > F_\alpha^{c_1, c_2, \ldots, c_k}]$$

$$= \underset{\sigma_0^2 \to 0}{Lt} \int_{u = F_\alpha^{c_1, c_2, \ldots, c_k}}^{\infty} f\left(u\,\frac{\sigma_0^2}{\sigma^2}\right) du\,\frac{\sigma_0^2}{\sigma^2}$$

$$= \underset{\sigma_0^2 \to 0}{Lt} \int_{u = F_\alpha^{c_1, \ldots, c_k} \times \sigma_0^2/\sigma_1^2}^{\infty} f(u)\,du = \int_0^{\infty} f(u)\,du = 1,$$

being the total probability.

$$\underset{\sigma_0^2 \to \infty}{Lt}\ P_r\,[u > F_\alpha^{c_1, c_2, \ldots, c_k}]$$

$$= \underset{\sigma_0^2 \to \infty}{Lt} \left[1 - \int_0^{F_\alpha^{c_1, \ldots, c_k}} f\left(u\,\frac{\sigma_0^2}{\sigma^2}\right) du\,\frac{\sigma_0^2}{\sigma^2}\right]$$

$$= \underset{\sigma_0^2 \to \infty}{Lt}\ [1 - \int_0^{F_\alpha^{c_1, \ldots, c_k}\sigma_0^2/\sigma^2} f(u)\,du] = 0.$$

Moreover it is obvious that $P_r\,[u > F_\alpha^{c_1, \ldots, c_k}]$ is a continuous function of $\sigma_0^2$. Hence a value $\sigma^2_{0\gamma}$ of $\sigma_0^2$ exists for which this probability is exactly equal to $\gamma$. This value $\sigma^2_{0\gamma}$ will depend upon $\sigma_1^2, \ldots, \sigma_k^2$ and $c_1, c_2, \ldots, c_k$ besides $\gamma$, and if $\sigma_1^2, \ldots, \sigma_k^2$ increase $\delta$ times this value will also increase $\delta$ times. In this situation we, therefore, find that

$$P_r\left[\sum_{i=1}^{k} \frac{c_i s_{(i)}^2}{s_0^2} > F_\alpha^{c_1, \ldots, c_k}\right] \gtreqless \gamma$$

according as $\sigma_0^2 \lesseqgtr \sigma^2_{0\gamma}$.

This property is designated as the property of gradation.

4. The optimum properties of the procedure '$d$' are proved as follows:

Let $\sigma_i = \sigma$, $i = 1, 2, \ldots, k$, then the probability of including $\pi_0$ in group 1 by the procedures $D$ is readily written down from Traux[2] as:

$$(k)!\,\frac{\Gamma\frac{(k+1)(n-1)}{2}}{\left[\Gamma\left(\frac{n-1}{2}\right)\right]^{k+1}}$$

$$\times \int \cdots \int_B \frac{[u_{(1)} \ldots u_{(k)}]^{(n-3)/2}}{\left(\frac{1}{\lambda^2}\right)^{k(n-1)/2}\left[\lambda^2 \sum_1^k u_{(\alpha)}+1\right]^{(k+1)(n-1)/2}}$$

$$\times \prod_1^k du_{(i)} \qquad \text{(iii)}$$

where

$$u_{(i)} = \frac{s_{(i)}^2}{s_0^2}, \quad B = \left[\sum_1^k c_i u_{(i)} < F_\alpha^{c_1, \ldots, c_k}\right]$$

and

$$F_\alpha^{c_1, c_2, \ldots, c_k}$$

is given by

$$K!\,\frac{\Gamma\frac{(k+1)(n-1)}{2}}{\left[\Gamma\left(\frac{n-1}{2}\right)\right]^{k+1}}$$

$$\times \int \cdots \int_B \frac{[u_{(1)} \ldots u_{(k)}]^{(n-3)/2}}{\left[\sum_1^k u_{(\alpha)}+1\right]^{(k+1)(n-1)/2}} \prod_1^k du_{(i)} = 1 - \alpha \qquad \text{(iv)}$$

It can be seen that the ratio of the integrand in (iii) to that in (iv) is a decreasing function of $\sum_1^k u_{(i)}$ if $\lambda^2 > 1$. It follows (on applying Neyman's Lemma) that (iii) would be maximum if the integration is over

$$A: \left(\sum_1^k u_{(i)} < F_\alpha^{1, 1, \ldots, 1}\right),$$

where $F_\alpha^{1, 1, \ldots, 1}$ is given by (iv). This proves that '$d$' maximises the probability of excluding from group 1 the population with the variance $\sigma_0^2$ if this is larger than the common variance $\sigma^2$ of the remaining $k$ populations. Similarly as the ratio of the integrand in

(iii) to that in (iv) is an increasing function of $\sum_{1}^{k} u_{(i)}$ if $\lambda^2 < 1$, it can be shown that '$d$' maximises the probability of not retaining the population with variance $\sigma_0^2$ if this is smaller than the common variance $\sigma^2$ of the remaining $k$ populations.


## SUMMARY

For the problem of dividing a set of populations on the basis of their variances a class of procedures is suggested following the analogous case of dividing the set of means as considered by Seal. It is shown that this procedure possesses similar properties as the one suggested by Seal. Under certain restrictions it is further shown that a particular member of this class is optimum as it maximises the probability of rightly classifying a given population.



## ACKNOWLEDGEMENT

I am thankful to Dr. A. R. Roy for his help in preparing this paper.


## REFERENCES


1. Seal, K. C. .. "On a class of decision procedures for ranking means of normal populations," *Annals of Math. Stat.*, 1955, **26**.
2. Truax, D. R. .. "An optimum slippage test for the variances of $k$ normal distributions," *Ibid.*, 1953, **24**.