CHOOSING THE BEST: RANKING AND SELECTION

Steve Cherry

September 11, 1989

Submitted in Fulfillment of Requirement for
MS-Statistics Writing Project

CHOOSING THE BEST: RANKING AND SELECTION

Identifying the best is a common problem in the plant breeding sciences. Sometimes it is the major goal of the research. Sometimes it is part of the experimental design if, for instance, an investigator needs to choose the best performing varieties part way through an experiment and then conduct a further analysis of these. Statements such as, "Treatment 5 was the best based on sample yields", or, "Treatments followed by * are not significantly different from the best treatment" are common, with the tacit implication that if treatment 5 performed the best it has to be the best.

A common approach to the analysis of data from such experiments is an analysis of variance followed by a multiple comparisons procedure, usually Fisher's Protected Least Significant Difference (LSD). The analysis of variance tests the null hypothesis of equality of means under the assumption of constant variance and normality of the underlying populations. If the null hypothesis is rejected, it is not possible to determine which means differ, nor is it possible to draw justified conclusions about the direction of the differences without additional testing. But, if the identification of the best is the goal, then the researcher starts with the a priori assumption that there are differences among the means. Classical hypothesis testing under this assumption is not proper (Chew 1980, Gibbons et. al. 1977).

The statistical methodology known as ranking and selection (R&S) is specifically designed to deal with problems in which the identification of the best is the major goal of the research.

We reviewed the 1984 and 1988 volumes of Crop Science and found that out of 470 articles at least 65 could have benefited from the use of R&S. None of the articles used R&S techniques. The objective of this paper is to describe R&S and to explain why it is superior to other methods currently being used to identify the best. For the purposes of this paper we will restrict ourselves primarily to the identification of the best one of several treatments (or populations). The methodology has been extended far beyond this special case, however (Gibbons et. al. 1977).

RANKING AND SELECTION

Suppose there are k ($k \geq 2$) populations (e.g. k varieties of wheat) that are known to be different with respect to some parameter of interest and we want to identify the best. Let μ_i (i=1,...,k) represent the parameter of interest. A reasonable approach would be to rank the μ_i's in ascending order represented notationally by $\mu_{[1]} \leq \mu_{[2]} \leq ... \leq \mu_{[k]}$ and we would identify $\mu_{[k]}$ as the true best since it has the highest parameter value. For example, if μ_6 is the mean yield of variety 6 of 10 varieties then $\mu_6 = \mu_{[10]}$ if variety 6 has the highest true mean yield.

The problem of course is that the true parameter values, the μ_i's, are unknown. They are estimated by taking random samples and computing a sample statistic that is a good estimator. If means are the parameters of interest then sample means X_i (i=1,...,k) are the statistic of choice. The population producing the highest sample mean ($X_{[k]}$) is the sample best, and, based on the sample results, that population is selected as the true best. (It is important to keep the distinction between the sample best and the true best in mind. Throughout this paper when we refer to "the best" we will mean the true (unknown) best. If there is any chance of confusion we will be explicit when referring to the sample best and true best). Since the choice is based on the results of a sample, there is a surprisingly large chance that the true best population did not produce the sample best (Becker 1961, Gauch and Zobel 1988).

Bechhofer (1954) was the first to treat this problem rigorously. The probability of correct selection depends on the true differences between the best population and the others, the variability of the populations, and the sample size. When the selection of the one best is the goal then $\mu_{[k]} - \mu_{[k-1]} = \delta$ is the difference of interest. There is a value of δ ($\delta = \delta^*$) that corresponds to an indifference distance. If $\delta \geq \delta^*$ then an investigator wants a high probability of correctly selecting the best. If $\delta < \delta^*$ then mistakenly selecting population k-1 as the best would be a matter of indifference. Thus, there is an indifference zone about $\mu_{[k]}$. Bechhofer's (1954) work was based on an assumption about the relationship of the underlying parameters that he defined as the "least favorable configuration", and his formulation yields conservative values for the probability of correct selection.

Bechhofer's (1954) methods and the concept of the indifference zone were derived within the context of experimental design. An investigator specifies a probability of correct selection (P^*) and an indifference distance (δ^*) and then uses this information to determine the sample size necessary to attain these prespecified levels. The probability of correct selection will be at least P^* when $\mu_{[k]} - \mu_{[k-1]} \geq \delta^*$, and the investigator can be confident at the P^* level that $\mu_{[k]}$ produced the sample best. There are tables available for use in sample size determination (Bechhofer 1954, Gibbons et.al. 1977). The procedure has been shown to be robust to violation of the assumption of normality (Dudewicz and Mishra 1984).

Gupta (1956,1965) extended the theory to subset selection. He considered the problem of selecting a subset C of random size such that the probability that the best population is included in C is greater than or equal to P^*. It is desirable for the number of populations in C to be as small as possible. Assuming that the parameters of interest are means, the decision rule by which populations are chosen to be included in C is based on the values of the sample means X_i (i=1,...,k). The sample means are determined and ranked ($X_{[1]} \leq X_{[2]} \leq ... \leq X_{[k]}$) and population i is chosen to be included in C if $X_i \geq X_{[k]} - d_{(k,P^*,\upsilon)} s\sqrt{2/r}$ where s is the pooled standard deviation, r is the number of replications, and $d_{(k,P^*,\upsilon)}$ is a multivariate t statistic with υ degrees of freedom. Tabled values of $d_{(k,P^*,\upsilon)}$ are easily obtained (e.g. Hochberg and Tamhane 1987). Gupta showed that this procedure would yield a subset C of populations that contained the true best

population with a probability of at least P^*.

Hsu (1981, 1984, 1985) further extended the theory by deriving simultaneous confidence intervals for all distances from the best, a procedure that allows an investigator to construct confidence statements that selected subsets contain the true (unknown) best. Hsu (1985) showed that simultaneous P^* level confidence intervals for $(\mu_i - \max_{(i\neq j)}\mu_j)$ are given by:

$$(L_i,U_i)= (X_i - \max_{(i\neq j)}X_j - h_{(k,P^*,\upsilon)}s\sqrt{2/r},\ \ X_i - \max_{(i\neq j)}X_j + h_{(k,P^*,\upsilon)}s\sqrt{2/r}).$$

(If $\mu_i - \max_{(i\neq j)}\mu_j > 0$, then population i is the best, otherwise it is not the best). Hsu (1985) gives tabled values of $h_{(k,P^*,\upsilon)}$ ($h_{(k,P^*,\upsilon)} \neq d_{(k,P^*,\upsilon)}$, although they are close for large values of k). The ith population is included in C if $U_i \geq 0$. The larger the values of U_i, the "better" the population is. A population with $L_i > 0$ would be declared the best with P^* level of confidence. Hsu (1984) also derived two quantities he called R and S values, that function in a manner analogous to the p-values in classical hypothesis testing. If there are k populations being compared then there will be k-1 R values and R_i (i=1,...,k-1) is the smallest α (α=1 - P^*) level for which the ith population can be rejected as the best. There will be only one S value and it is the smallest α level for which the population corresponding to the highest sample mean $X_{[k]}$ can be selected as the best. The S value is equal to the largest R_i.

Gupta (1956, 1965) and Hsu (1981, 1984) derived their methods within the context of data

analysis. If an investigator chooses the sample best and declares this choice the true best, then their methodology allows the investigator to determine how confident he or she can be that they have indeed chosen the true best population. These confidence levels are determined based on observed differences among the sample statistics. Hsu (1981, 1984, 1985) showed that his method represented a unification of, and was superior to, the methods of Bechhofer and Gupta. Hochberg and Tamhane (1987) showed the connection to Bechhofer's (1954) indifference zone formulation is through the L_i. If population i is considered to be "good" when $\mu_i - \mu_{[k]} \geq -\delta^*$ (where δ^* is the prespecified indifference distance) then all populations with $L_i \geq -\delta^*$ will be good with probability at least P^*. The connection to Gupta's (1956,1965) subset selection procedure is through the U_i. Gupta's procedure allows the assertion that C contains the one best with a probability of at least P^*. Thus, the procedure can be used to eliminate inferior populations, but nothing can be said about the probability that a given population in C is the best. Hsu's (1984, 1985) extension allows for a further assessment of these potentially good populations.

It is interesting to compare the subset selection procedure in R&S with the use of LSD to pick a winner. When LSD is used, the sample means are ranked, and then a series of pairwise comparisons is made. The winners are those populations with the highest sample means that are significantly different from all other sample means, but not from each other. $X_{[k]}$ always denotes a winner. These means are grouped into a subset and are declared to be the winners. Population i is included in this subset when the sample mean $X_i \geq X_{[k]}$ - LSD ($i \neq [k]$). Here LSD = $t_\alpha s\sqrt{2/r}$ (where t_α is the t statistic). The LSD was

designed to control the chances of falsely rejecting the null hypothesis of equal means. Thus, the probabilities and confidence levels associated with the LSD are based on the assumption that the null hypothesis is true. If the goal of the research is to identify the true best then the only error of interest is to declare the sample best the true best when it is not. The LSD can only be used in an indirect way to accomplish this goal and it does not control the error rate of interest. R&S is a superior technique in these types of problems because it gives a guaranteed probability against making the wrong choice (Gupta and Panchapakesan, 1979).

The calculations required to derive the confidence levels and the R and S values can be formidable. Lund (1989) developed an algorithm for balanced designs for choosing the m ($m \geq 1$) best from a set of k populations that incorporates the 3 approaches discussed above. Hsu developed software for choosing the one best from k populations that is available on SAS (see Aubuchon et. al. 1986, and Gupta and Hsu 1984). Hsu's package will work with unbalanced designs.

Some of the earliest discussions of the use of ranking and selection are from the animal breeding literature. Becker (1961, 1962, 1964) and Soller and Putter (1964,1965) discussed the use of ranking and selection in the choice of the best chicken strains and dairy cow breeds. Becker (1961) showed that a sample from the best (he defined the best to be those strains that had the highest egg production) population would not always win; that in fact, poorer chicken strains would win a surprisingly large proportion of the time.

Gauch and Zobel (1988) arrived at the same conclusion in a discussion of choosing the best genotypes from a succession of yield trials. They showed that the selection of winners was more difficult than might be suspected, and that increasing the number of replications did not result in an increased probability of success as great as one would hope. They presented another strategy (the Additive Main effects and Multiplicative Interaction Model) for increasing the precision of the sample estimates that they claimed was more cost effective than simply increasing the number of replications.

There have been only a few applications of R&S procedures in actual experiments. Lorenz et. al. (1982) assessed the performance of water filters. Richard et. al. (1982) evaluated the effectiveness of several different vaccines in the treatment of aspergillosis in domestic turkeys. Ray and Ray (1979) used the methodology to determine the best arrangement of controls for heating elements on kitchen ranges, and Swart et. al. (1988) used R&S to assess the relative susceptibility of six Pinus species to a fungal infection.

Several current texts discuss not only recent theoretical advances in the field, but also applications (Gibbons et al. 1977, Gupta and Panchapakesan 1979, Santner and Tamhane, 1984). Gibbons et. al. (1977) gives a non-technical approach to ranking and selection that is readily accessible to nonstatisticians.

EXAMPLE 1

The data are taken from Petersen (1985), who presented it as an example of using LSD to pick the winner. The data are the results of one yield trial on 12 varieties of wheat with 4 replications are presented in Table 1. We have analyzed the observations using LSD and R&S in an effort to choose the one best from the 12 selections. All analyses were carried out at the .05 level of significance (i.e. $P^*=.95$).

From the LSD analysis, we found that varieties 9 and 3 are not significantly different from one another, but that these two are different from the other 10. The LSD conclusion would be that 9 and 3 "won" (since they have the highest mean yields) and if we want the one best we would choose the sample best (variety 9) for further evaluation. The fact that variety 3 has a mean yield that is different from that of variety 8 at the .05 level tells us that we can be confident that 3 is different from 8, but we cannot put some quantitative level on the confidence we feel. The inference that 3 is "better" than 8 because 3 has a higher sample mean is not an inference we can make directly from the results of the LSD. We conclude that 3 is better than 8 because of our knowledge of wheat variety trials and the indirect evidence of the LSD, but we could be wrong and there is no way to determine how confident we are that we are correct in our conclusion.

The goal is to answer the question, "Given that the 12 varieties are in fact different from

each other with respect to yield, what is the probability that the true best will yield the sample best?". The R&S analysis attacks that goal directly. For the R&S analysis we calculated the rejection p-values (R values), which represent the chance of making an incorrect choice. The rejection p-values were computed assuming the "least favorable configuration" of the underlying population means, and that the data are normally distributed with constant variance. Bechhofer (1954) and Gibbons et. al. (1977, p. 14) present more detailed discussions of the least favorable configuration and its implications.

Variety 9 is the sample best, but, based on R&S analysis we can declare variety 9 to be the true best with only about .27 confidence (1 minus the S value of .73). We can declare one of varieties 9 and 3 to be the true best with about .96 confidence. The rejection p-value of about .73 for variety 3 can be interpreted to mean that we can be about .73 confident that the true best variety produced a sample mean less than or equal to that of variety 3. Similarly, we can be about .04 confident that the true best variety produced a sample mean less than or equal to that of variety 8.

From the 95% confidence intervals in Table 1 we see that only varieties 9 and 3 have positive upper bounds. Thus, they are the only two varieties in contention for being the true best. None of the varieties have positive lower bounds so none can be declared the one best at the .95 level of confidence.

The results here are fairly clearcut. Even if we only used LSD to arrive at a decision, it

is obvious from Table 1 that varieties 9 and 3 form a distinct grouping that we would feel comfortable working with if we wanted the true highest yielding population. The results are not always so clearcut. For example, suppose that we were interested in choosing the lowest yielding variety. Based on the performance of the 12 varieties, we would choose variety 7 to be the true lowest yielding variety since it has the lowest sample mean yield, however, there is a great deal of overlap in the LSD results. Variety 11, for instance, cannot be judged significantly different from the variety with the lowest sample mean yield, and it cannot be judged significantly different from the variety with the third highest sample mean yield. Based on R&S we can judge variety 7 to be the true lowest yielding variety with about .24 confidence, and varieties 5, 1, 10, 11, 6, and 7 constitute a subset of the 12 varieties that contain the true lowest yielding variety with at least .96 confidence (Table 1).

In this example, both LSD and R&S always arrive at the same conclusion. R&S provides a measure of probabilistic uncertainty associated with that choice; no such uncertainty measure can be attached to the LSD choice.

EXAMPLE 2

This data is taken from a recent paper in Crop Science (Foley and Clark 1984). The objectives of this study were "to identify maize accessions of the USDA Plant Introduction (PI) Collection that have superior stalk rind strength...". The authors were not interested

in the actual estimates of stalk rind strength, only in the their relative rankings. Those accessions with the highest values of stalk rind strength were chosen as the best. The study was conducted over 5 years. There were 86 accessions tested the first year, 87 the second, 129 the third and fourth, and 139 the fifth. Ten plants from each accession were planted in 1 x 3m plots in a completely randomized design. The experiment was replicated twice.

The 20 highest ranking entries from the first year's trial are reproduced here (Table 2), along with the results of R&S analysis of the data. PI 262484 is the sample best, but it can be declared the true best with only about .06 confidence. The chance of selecting the one best accession from the 86 just by picking one at random is about 0.012. Growing two replicates of the 86 accessions in a completely randomized design and choosing the accession with the highest sample mean raised the probability of correct selection to approximately 0.06 illustrating Gauch and Zobel's (1988) point about how difficult it can be to select the best with any acceptable level of confidence. Using Lund's (1989) algorithm and the tabled values in Gibbons et. al. (1977) we were able to determine that if the number of replications were increased to 20 the confidence level would be raised to to about 0.49.

In these examples we have emphasized the selection of the one best. In a problem such as the one described here the selection of the one best from some 86 different accessions may not be very interesting. A more pragmatic question might be, "How many accessions do we need to select to be 95% confident that the ten best will be included in the group?".

Lund's (1989) algorithm allows for a determination of the probability of correctly choosing the m ($m \geq 1$) best from up to 100 populations. We determined the value of $d_{(k,P^*,v)}$ that yielded a $P^*=.95$ for the selection of the 10 best from 86 accessions. By multiplying this value (5.28) times the standard error of the mean (.11) we estimated that to be 95% confident of getting the ten best accessions, all accessions with mean stalk rind thickness between 0.15mm and .73mm would have to be included in C. Unfortunately, we do not have the original data, but the overall mean rind thickness of all 86 accessions was .43mm and the minimum was .09mm. Clearly, the majority of the 86 accessions would have to be chosen to be sure of getting the ten best at the prescribed level of confidence.

CONCLUDING REMARKS

R&S methods offer clear advantages over the classical approach when the identification of the best is the goal of an experiment. Although R&S techniques have been available for decades, they have not been extensively utilized. As Hsu (1984) pointed out, use of R&S analyses may be contingent on availability of computer packages. A readily available software package is a practical necessity when analyzing a large number of populations, as in Example 2. Not all R&S calculations require a computer, however. The confidence intervals in Table 1 were easily determined by hand. In any case computer programs are now available. Hsu's (1981, 1984, 1985) multiple comparisons with the best techniques are available on SAS. Lund's (1989) algorithm is available on Version 4.12 of MSUSTAT. (MSUSTAT was developed by R. E. Lund, Montana Agricultural Experiment Station,

Montana State University, Bozeman, Montana 59717). Readers who want additional information are referred to Aubuchon et. al. (1986), Gupta and Hsu (1986), and Lund (1989). Dudewicz (1982) may also be of interest.

The introduction and acceptance of new techniques are often hampered by tradition. Researchers rely on "standard" techniques that are commonly used in their field of interest, and they may be reluctant to switch to some other method. Techniques that have become accepted tend to hang on, sometimes long after they have been superseded by improved methods. For example, of the 470 articles we looked at, 57 used Duncan's multiple range test, although the test has been shown to not control any error rate of interest in multiple comparisons, and is generally no longer recommended by statisticians. We feel that similar historical reasons have led to slow acceptance of R&S methods.

Another reason for the lack of use of this R&S is the failure of statisticians to teach those techniques. Many of the texts used in statistics courses make no reference to R&S. Many scientists have taken such courses and then conducted research believing that statistics is hypothesis testing, and little else. The authors of over 90% of the papers published in the 1984 and 1988 volumes of Crop Science used analysis of variance or linear regression (LSD was used in 144 of the 470 articles; 126 of the 144 articles conducted an analysis of variance and a multiple comparisons analysis). None of the researchers used R&S.

Fortunately, the situation is changing. It is becoming more common to find papers on

R&S in statistical journals. R&S methods in some recent texts (e.g. Dudewicz and Mishra 1988, Hochberg and Tamhane 1987). We hope the trend will continue. If the goal is to choose the best, then start by choosing the best statistical technique to accomplish that task.

Table 1. Results of using LSD and R&S to identify the best of 12 varieties of wheat. Mean yield is in kg/ha.

Variety	Rank	Mean Yield[1]	Rejection P-Values[2]	95% Confidence Intervals for $\mu_i - \max_{(i \neq j)} \mu_j$[3]	Rejection P-Values[4]
9	1	2892a	**	[-415, 665]	.000
3	2	2767a	.733*	[-665, 415]	.000
8	3	2346 b	.041	[-1086, -6]	.018
4	4	2335 b	.036	[-1097, -17]	.021
2	5	2284 b	.020	[-1148, -68]	.037
12	6	2281 b	.019	[-1151, -71]	.038
5	7	2223 bc	.010	[-1209, -129]	.069*
1	8	2205 bc	.008	[-1227, -147]	.083*
10	9	2194 bc	.007	[-1238, -158]	.092*
11	10	2092 bcd	.002	[-1340, -260]	.219*
6	11	1841 d	.000	[-1591, -511]	.757*
7	12	1729 d	.000	[-1703, -623]	**

1-Numbers followed by the same letter are not significantly different from each other at the .05 level of significance. LSD=414.4 kg/ha.

2-Variety 9 (**) can be declared the one best with at least .267 confidence. Variety 9 (**) and 3 (*) are a set of 2 varieties that contain the one best with at least .96 confidence.

3-Confidence intervals calculated from $(X_i - \max_{(i \neq j)} X_j) \pm h_{(k,P^*,\nu)} s\sqrt{2}/r$ with k=12, s=287, r=4 and $h_{(12,.95,33)}=2.658$ based on tabled values in Hsu (1985).

4-Variety 7 (**) can be declared the true lowest yielding variety with at at least .24 confidence. Varieties 7, 6, 11, 10, 1, and 5 are a set of 6 varieties that contain the true lowest yielding variety with at least .96 confidence.

Table 2. Mean rind thickness (mm) of stalks of the 20 best maize PI accessions for one year. Thickness is based on rind strength. Data analyzed by R&S.

PI	(mm)	Rejection P-Values[1]
262484	0.73	**
210404	0.67	.938
257612	0.64	.883
273474	0.63	.858
213713	0.62	.828
270080	0.61	.796
200185	0.60	.761
221835	0.60	.761
186210	0.59	.723
262494	0.58	.683
222618	0.57	.639
167962	0.57	.639
222645	0.56	.594
213702	0.56	.594
274011	0.56	.594
164381	0.55	.548
172331	0.54	.502
163558	0.54	.502
267203	0.54	.502
172330	0.52	.411

1-PI 262484 (**) can be declared the one best of the 86 with at least .062 confidence.

LITERATURE CITED

1. Aubuchon, J. C., S. S. Gupta, and J. C. Hsu. 1986. The RSMCB procedure. In SUGI Supplemental Library User's Guide, Version 5 Edition, Chapter 38, SAS Institute, Inc. Cary, NC.

2. Bechhofer, R. E. 1954. A single-sample multiple decision procedure for ranking means of normal populations with known variances. Annal. Math Stat. 25:16-39.

3. Becker, W. A. 1961. Comparing entries in random sample tests. Poultry Sci. 40:1507-1514.

4. Becker, W. A. 1962. Ranking all-or-none traits in random sample tests. Poultry Sci. 41:1437-1438.

5. Becker, W. A. 1964. Changes in performance of entries in random sample tests. Poultry Sci. 43:716-722.

6. Chew, V. 1980. Testing differences among means: Correct interpretation and some alternatives. HortSci. 15:467-470.

7. Dudewicz, E. J. 1982. Estimation of the probability of correct selection with a computer program. Biom. J. 24:297-307.

8. Dudewicz, E. J., and S. N. Mishra. 1984. The robustness of Bechhofer's normal means selection procedure. In Santner, T. J., and A. C. Tamhane. Design of experiments: Ranking and Selection. Marcel Dekker, Inc., New York. 302pp.

9. Dudewicz, E. J., and S. N. Mishra. 1988. Modern mathematical statistics. J. Wiley and Sons, New York. 838pp.

10. Foley, D.C., and R.L Clark. 1984. Mechanical properties of maize stalks from the plant introduction collections. Crop Sci. 24:1116-1118.

11. Gauch, H. G., Jr., and R. W. Zobel. 1988. Accuracy and selection success in yield trial analyses. Theor. Appl. Genet. 77:473-481.

12. Gibbons, J. D., I. Olkin, and M. Sobel. 1977. Selecting and ordering populations: A new statistical methodology. John Wiley and Sons, New York. 569pp.

13. Gupta, S. S. 1956. On a decision rule for a problem in ranking means. Institute of Statistics. Mimeo Ser. 150. Univ. of North Carolina, Chapel Hill, NC.

14. Gupta, S. S. 1965. On some multiple decision (selection and ranking) rules. Technometrics 7:225-245.

15. Gupta, S. S., and J. C. Hsu. 1984. User's guide to RS-MCB. Tech. Rep., Dept. of Statistics, Ohio State Univ., Columbus, OH.

16. Gupta, S. S., and S. Panchapakesan. 1979. Multiple decision procedures: Theory and methodology of selecting and ranking populations. John Wiley and Sons. New York. 573pp.

17. Hochberg, Y., and A. C. Tamhane. 1987. Multiple comparison procedures. J. Wiley and Sons. New York. 450pp.

18. Hsu, J. C. 1981. Simultaneous confidence intervals for all distances from the "best". Ann. of Stat. 5:1026-1034.

19. Hsu, J. C. 1984. Ranking and selection and multiple comparisons with the best. IN Santner, T. J., and A. C. Tamhane. Design of experiments: Ranking and selection. Marcel Dekker, Inc. New York. 302pp.

20. Hsu, J. C. 1985. A method of unconstrained multiple comparisons with the best. Commun. Stat.-Theor. Meth. 14:2009-2028.

21. Lorenz, R. C., J. C. Hsu, and O. H. Tuovinen. 1982. Performance variability, ranking and selection analysis of membrane filters for enumerating coliform bacteria in river water. J. Am. Waterworks Assoc. 74:429-437.

22. Lund, R. E. 1989. Probabilities and standardized differences for selecting subsets containing the best populations. Unpublished Manuscript.

23. Petersen, R. G. 1985. Design and analysis of experiments. Marcel Dekker, Inc., New York. 429pp.

24. Ray, R. D., and W. D. Ray. 1979. An analysis of domestic cooker control design. Ergonomics 22:1243-1248.

25. Richard, J. L., J. R. Thurston, R. C. Cutlip, A. C. Pier. 1982. Vaccination studies of aspergillosis in turkeys: Subcutaneous inoculation with several vaccine preparations followed by aerosol challenge exposure. J. Am. Vet. Res. 43:488-492.

26. Santner, T. J., and A. C Tamhane. 1984. Design of experiments: Ranking and selection. Marcel Dekker, Inc., New York. 302pp.

27. Soller, M., and J. Putter. 1964. On the probability that the best chicken stock will come out best in a single random sample test. Poultry Sci. 43:1425-1427.

28. Soller, M., and J. Putter. 1965. Probability of correct selection of sires having highest transmitting ability. J. Dairy Sci. 48:747-748.

29. Swart, W. J., M. J. Wingfield, and P. S. Knox-Davies. 1988. Relative susceptibilities to sphaeropsis-sapinea of six Pinus spp. cultivated in South Africa. Europ. J. For. Path. 18:184-189.