STATISTICAL REPORT

Assessing uncertainty in knowledge-based systems for data analysis by simulation

by

Jan H. Aarseth and Ivar Heuch

Report no. 30
August 1996

Department of Mathematics
UNIVERSITY OF BERGEN
Bergen, Norway

Department of Mathematics
University of Bergen
N-5007 Bergen
Norway

ISSN 0333-1865

Assessing uncertainty in knowledge-based systems for data analysis by simulation

Jan H. Aarseth and Ivar Heuch

Abstract: There is no consensus on how statistical systems incorporating complex statistical strategies should be evaluated. Only when a suitable standard for evaluation has been established will it be possible to compare the performance of an automated statistical analysis with that of a more conventional approach. This paper describes how the evaluation may be supported by simulation, exemplified by a particular simulation module incorporated into our knowledge-based system Express. In an example involving simple strategies for one-way analysis of variance of 3 samples, a mixed bootstrap strategy, combining the ordinary *F* and the James statistics, is shown to be superior to more traditional procedures.

KEY WORDS: analysis of variance; bootstrap test; expert system; *F* test; James statistic; Levene test; preliminary test; statistical strategy.

Statistical Report No. 30
August 1996

1. INTRODUCTION

Knowledge-based computer systems in statistics may attempt to translate a research goal into a specific data analytic agenda, to select a general statistical approach or to carry out the analysis given the approach (Gale *et al.*, 1993). Systems of the third kind, for statistical data analysis, can incorporate complex strategies for handling particular kinds of problems. A number of statistical packages are available for standard data analysis, with a variety of methods offered at various levels of complexity. Such packages incorporate arithmetic and algebraic expertise but not statistical expertise in the sense of a knowledge-based system (Hand, 1985). In addition to the execution of the actual calculations needed, a knowledge-based system must determine which statistical methods are called for, or at least be able to make reasonable suggestions about choice of methods.

A strategy implemented in a knowledge-based system must be tested to check its properties. In this paper we show how computer intensive methods can be applied for this purpose. Simulation has been used extensively to explore general properties of statistical procedures, and the approach is particularly useful in knowledge-based systems, as statistical strategies typically consist of sequences of rules used repeatedly on the same data set. If the strategy includes, for example, non-independent tests carried out in a certain order, it is difficult to determine the overall performance by theoretical arguments.

Statisticians tend to respond differently in the selection of strategies for solving particular problems (Van den Berg and Visser, 1990; Tung and Schuenemeyer, 1991). The discrepancies are especially marked for statisticians employed in different applied areas. In such situations, simulations comparing alternative strategies can reveal which one is the most informative. For a knowledge-based system handling different kinds of practical problems, an integrated simulation module will facilitate the comparisons. Our system Express (Aarseth and Heuch, 1996a) is equipped with a module of this kind. Considering an example with simple strategies coded in Express for one-way analysis of variance, including a strategy incorporating bootstrap tests, we show how strategies can be compared numerically.

2. UNCERTAINTY IN KNOWLEDGE-BASED SYSTEMS

It is essential in any knowledge-based system to have a measure available of the uncertainty in the conclusions, both for general usage and for evaluation of strategies implemented. In the AI community, several approaches have been used (Bhatnagar and Kanal, 1986). The representation of uncertainty should take into account the fact that conclusions usually rely on incomplete information. Thus essential data may be missing, the information supplied may not represent facts but merely suggestions, or different pieces of information may be contradictory. The handling of information inside the system may add to the uncertainty. Furthermore, it is difficult to combine uncertainty when unreliable information is pooled. Several systems have used the probability concept to treat uncertainty, and Bayesian approaches have become quite popular (Lindley, 1987; Shafer, 1987; Spiegelhalter *et al.*, 1993). In contrast, we adopt the frequentist approach to statistical inference.

Statistical tests or estimation procedures do not lead to exact answers. If several methods are used successively to produce a conclusion, the uncertainty is propagated to the final result. However, measurement of uncertainty is straightforward for many basic methods in statistical analysis. For example, if we want a confidence interval for the mean of a set of normally distributed observations, the accuracy in terms of a cover probability is known. Uncertainty, expressed in this manner by probabilities, is used extensively and is well understood.

This approach is more complicated if we wish to investigate the uncertainty associated with a complete statistical strategy. To identify sources of uncertainty, Hodges (1987) divided statistical activity into three parts: discovery of structure, assessment of variation conditional on structure and execution of techniques. In systems executing selected data analytic techniques, the two latter types of uncertainty are obviously important. Even in our simple example with a confidence interval, however, there is uncertainty both regarding the assumption of normality (structural uncertainty), the cover accuracy (risk given structure) and the possibility of introducing errors or approximations in the calculations (technical uncertainty). All components may affect the final conclusion. In addition, the entire strategy typically consists of many techniques used on the same data set (Hand, 1987). Even if uncertainty can be adequately assessed for each separate technique, it is difficult to combine the results into an overall measure.

3. UNCERTAINTY ASSESSED BY SIMULATION

Monte Carlo simulations are frequently used to estimate statistical measures of performance such as the level and power of hypothesis tests or the cover probability of confidence intervals. We claim that the same basic approach can be useful in evaluating knowledge-based systems. Such systems often aim at providing simple answers to questions raised by the users. Although the strategy itself may be quite complex, simulation results for the final conclusions should thus be easy to interpret. Admittedly, the conclusion cannot always be formulated as a simple categorical response. However, even in such situations simulation can provide essential information about parts of the strategy.

Simulation as a means of dealing with uncertainty has attracted considerable attention in contemporary AI (Paul, 1994; von Rimscha, 1994; Zlatareva and Preece, 1994). Surprisingly, the enthusiasm has not extended to knowledge-based systems for statistical data analysis. There may be several reasons for this. In standard data analysis, simulations typically deal with methods involving a particular statistic. Thus, from the traditional point of view, one works under specific assumptions which do not apply to knowledge-based systems. Also, until recently, complex computer intensive simulations have been very time-consuming.

Other methods for evaluating statistical strategies have been proposed. In his fundamental critique of expert systems, Streitberg (1988) suggested that a system should compete with experts in a selected area. An alternative approach is to let statistical experts suggest procedures for analysing a data set from the literature and then compare their recommendations with those of the computer system. This method was used for evaluating the exploratory data analysis system WAMASTEX (Dorda *et al.*, 1990). To assess the heuristic rules included in this system, simulations were also carried out. Moreover, a simulation technique was used to evaluate a two-sample strategy implemented in an early version of Express (Carlsen and Heuch, 1986). In neither case was the simulation tool part of the knowledge-based system itself. In view of the technical uncertainty associated with implementation of knowledge, it seems preferable to incorporate the simulation unit into the system to obtain information about the overall uncertainty.

A statistical strategy typically consists of three kinds of components, connected by heuristic rules. Parameter estimation forms an important part, informal hypothesis testing another one. Finally, decisions must be made concerning the organization of the data set itself. For example, the system may be forced to decide whether a variable is nominal, ordinal or continuous. Uncertainty associated with decisions made in estimation and testing can in principle be assessed by simulation, provided that data can be generated from relevant probability distributions. If the results from the simulation reveal a substandard performance of a component in the strategy, it is usually easy to carry out the necessary fine-tuning by, for example, modifying threshold values.

It is much harder to assess the uncertainty attached to decisions concerning the structure of the data. The system may ask the user for assistance, but if he is unable to respond, the strategy should proceed with a default rule. The chances of making a wrong decision may be substantial, and it is not normally obvious what the implications are. Thus, simulation should be particularly useful in the testing of such default rules. By fixing particular values during the simulations, it is also possible to study the consequences of an incorrect user response to a question concerning data organization, if, for example, an ordinal variable is specified as continuous. Exploring extreme situations, one may investigate how robust the strategy is. If the final conclusion is very sensitive to misspecification of a particular quantity, the system should warn the user and explain what can happen with an erroneous response.

Thus simulation can be used to assess structural uncertainty, risk given structure and technical uncertainty. Moreover, simulation makes it possible to integrate the uncertainty arising in separate parts of the strategy. As an example, consider a particular step in a strategy at which different test procedures can be applied under various formal assumptions about the distributions involved. A preliminary test will often be made to check these assumptions, and a complex rule must decide which test procedure should be preferred. The usually intractable theoretical problem of assessing the properties of the combined procedure is easily solved as part of the overall simulation for the strategy.

There is a current trend in statistics towards use of Bayesian methods to adjust for structural uncertainty (Draper, 1995). Model selection is regarded as a statistical problem with uncertainty expressed by prior and posterior probabilities, in addition to the inherent uncertainty of parameter estimates. In practice, this often corresponds to averaging over several reasonable models. Our simulation approach is conceptually different and is based on the assumption that an unknown model description exists which is correct for each data set. When the separate rules in the strategy are activated, decisions must sometimes be made about this unknown description, but they are based on the observations. Thus the uncertainty of the model description can be measured by performing simulations. In this manner, the distributions of the possible conclusions can be estimated under relevant competing models.

In particular cases, simulations may indicate that the strategy does not perform very well at all. The results may then provide guidance for correcting the strategy. Thus, if a collection of successive tests turns out to have an inadequate performance, threshold values may be changed, or some tests may be replaced completely by other procedures. In this way, a sequence of simulations, applying slightly different strategies, can form the basis of a learning process. Learning by example is a common method in knowledge acquisition (Winston, 1984), but this process is closer to learning from ones own mistakes. The final result will be an improved implementation of knowledge.

4. SIMULATIONS IN EXPRESS

Express is a tool for constructing rule-based systems for data analysis using already existing statistical software (Aarseth and Heuch, 1996a, 1996b). Statistical strategies may be implemented by specifying chains of rules according to certain conventions. Results produced by standard packages, automatically executed by Express when needed, are extracted from the output and stored in a working memory. This memory also contains information supplied by the user. On the basis of the quantities determined, decisions are made on how to proceed with the analysis. Many intermediate results may be found before the system reaches the main conclusion, and this information is also presented to the user.

The current, most comprehensive version of Express is implemented on a PC in Fortran and assembler under MS-DOS (Aarseth and Heuch, 1996a). A Unix version running under the X Window System was constructed for the purpose of simulation and for testing with a standardized user interface. Because of the modular design, Express could be transferred between systems essentially by changing interface routines (Aarseth and Heuch, 1996b). The X Window version was implemented in Fortran and XVIEW (using C).

When Express executes in standard mode, the user is first asked to indicate which data set should be analysed, before a chaining of rules is activated. Depending on intermediate results, the proper analysis will be performed as defined by the strategy implemented. In simulation mode, a random data set is generated before the chaining of rules begins, and a separate analysis is performed. This cycle is repeated many times, with a particular conclusion for each run. Intermediate and end results can be recorded by Express for subsequent analysis. It is possible to assign fixed values in advance to intermediate quantities considered in the chaining, so that the ordinary inference mechanism is set aside for these quantities only.

In view of the different simulation requirements, a random number generator has not been built into the system, but the simulation module has also been designed to utilize external software. A separate interface to the external generator inserts random numbers into the data storage of Express, where they can be referred to by the rules. In principle any generator may be selected, although appropriate NAG routines (NAG, 1993) were applied for generation of data sets considered in this paper. With this approach, particular care must be taken to ensure independence between data sets generated, by saving and restoring the current seed value.

5. A STRATEGY FOR ONE-WAY ANALYSIS OF VARIANCE

We consider a simple strategy (Fig. 1) for investigating whether location parameters differ between separate samples, without making any prior assumptions about the underlying distributions. This strategy has been implemented in Express (Aarseth and Heuch, 1996a). We emphasize that most statistical strategies in knowledge-based systems will be considerably more complicated, and that this basic example has been selected to illustrate the general use of simulation. The Shapiro-Wilk test (Shapiro and Wilk, 1965) is used first to check for normality in each sample. In case the normality assumption is accepted, the Levene test for equal variances (Levene, 1960) is performed. If one cannot assume that the samples have identical variances, the Brown-Forsythe test (Brown and Forsythe, 1974a) is carried out to decide whether location parameters differ. Otherwise, the standard F test is applied. If there are indications that any sample is non-normal, the non-parametric Kruskal-Wallis test (Kruskal and Wallis, 1952) is executed. Nominal significance levels of the separate tests are set to 0.05 in the current implementation, although much higher levels might be justified for preliminary tests (Bancroft and Han, 1983). In fact, with 3 samples, the effective level of the combined test for normality is $1\text{-}0.95^3=0.143$. In any case, it is clearly difficult to predict the overall performance of the strategy without resorting to simulation.

Despite its simplicity, and despite warnings against reuse of data in successive tests, this strategy probably reflects common practice among many users of statistics. The choice of tests is likely to be influenced by the procedures available in the software used regularly. The tests for normality, standard analysis of variance and the non-parametric test are included in nearly all major statistical packages, whereas the Levene and Brown-Forsythe tests are not always accessible. In our implementation, all test are carried out by BMDP (Brown, 1990), which is started by Express, and the relevant results are extracted from the BMDP output.

To study the properties of the strategy in Fig. 1 as implemented in Express, we generated data from normal, t_4 and exponential distributions in 3 samples, each comprising 20 observations. The range of each exponential distribution was defined so as to give the prescribed mean value. The results, with 6 different combinations of parameter settings, are presented in Table 1. Each simulation included 5000 replicates.

For skewed heteroscedastic samples, the strategy does not perform very well. Thus the probability of concluding that location parameters differ is as high as 22.9% for exponential distributions with identical means and standard deviations 1.0, 2.0 and 3.0, respectively. Table 2 presents a more detailed summary of the simulations in this particular case. Not all intermediate results determined by Express are used to establish a particular final conclusion. For example, the p-values for equal variances, for the ordinary F and for the Brown-Forsythe test are all calculated in the same execution of BMDP. If any such value is needed during the chaining of rules, the remaining two values are found at the same time. For this reason, the Express simulation module reports two different sets of results, as shown in Table 2. The results on the lefthand side are not always referred to in order to reach a final conclusion but may still provide useful information for evaluating components of the strategy.

The counts on the righthand side of Table 2, reflecting replicates for which the chaining of rules passed through the results considered, show that among the 5000 replicates, 4979 led to the conclusion that the observations were non-normal, with subsequent application of the Kruskal-Wallis test. Not surprisingly, this test rejected the hypothesis far too often in the case of scaled exponential distributions with identical means but different variances.

To improve the strategy, the Kruskal-Wallis test might be avoided when variances appear to differ between samples. Because of its robustness (Tan and Tabatabai, 1985), the Brown-Forsythe test might even be applied when non-normality is indicated. However, the Levene variance test may have an unacceptably high probability of type I error under non-normality (Brown and Forsythe, 1974b; Loh, 1987). In addition, the overall check for normality is not very reliable, although an adjustment for multiple testing might be introduced with a higher number of samples. Another option is to introduce robust procedures which are superior in power, e.g. the test proposed by Tan and Tabatabai (1986) in place of the Brown-Forsythe test. We will explore a different approach, incorporating bootstrap methods.

6. STRATEGIES BASED ON BOOTSTRAP METHODS

Bootstrap methods are typically introduced to reduce the uncertainty of risk given structure. As so few assumptions are needed, however, they also seem to hold some promise for reducing structural uncertainty. Hjort (1994) considered the bootstrap and other computer intensive methods in connection with model selection. In this section we evaluate a strategy based on bootstrap tests only.

Fisher and Hall (1990) studied bootstrap methodology in one-way analysis of variance, considering two statistics, the ordinary F

$$T_1 = \frac{n-r}{r-1} \frac{\sum_{i=1}^{r} n_i\left(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}\right)^2}{\sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(X_{ij} - \bar{X}_{i\cdot}\right)^2} \tag{1}$$

and the James statistic (James, 1951)

$$T_2 = \sum_{i=1}^{r} \frac{n_i(n_i-1)\left(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}\right)^2}{\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_{i\cdot}\right)^2}$$

where X_{ij} is observation number j in sample number i, r is the number of samples, $n = \sum_{i=1}^{r} n_i$, $\bar{X}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ and $\bar{X}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ij}$.

As emphasized by Fisher and Hall (1990), the basic principles of bootstrap hypothesis testing differ from those for constructing confidence intervals. First, it is essential that resampling be carried out under the null hypothesis (Hall and Wilson, 1991). In the heteroscedastic case in analysis of variance, this implies that resamples should be drawn from the separate groups of observations, after the transformation $Y_{ij} = X_{ij} - \bar{X}_{i\cdot}$ has been applied. In the homoscedastic case, resampling should be carried out for each group among all original observations, after application of the transformation $\left(X_{ij} - \bar{X}_{i\cdot}\right)/\hat{\sigma}_i$. Resampling ignoring these guidelines lead to inadequate tests with low power. Second, the test statistic should be asymptotically pivotal, i.e. the asymptotic distribution should not depend on unknown parameters; otherwise the level of the test may be affected. Fisher and Hall (1990) showed that T_1 is pivotal in homoscedastic problems while T_2 also has this property in the heteroscedastic case. By means of Edgeworth expansions, they demonstrated that the error in the significance level was greater for T_1 than for T_2 in the case of unequal variances. This result was verified by simulations.

An ideal bootstrap strategy for one-way analysis of variance would be based on T_2 when variances differ and T_1 otherwise. This strategy should be superior to that based on T_2 only, as the test involving T_1 has a higher power in the homoscedastic case. As no prior assumptions can be made about variances, we still need a preliminary test to select the appropriate statistic. We again resort to the Levene test, based on a statistic given by expression (1) with the observations replaced by absolute deviations $Z_{ij} = |X_{ij} - \bar{X}_{i\cdot}|$ from the sample mean. To compensate for the high probability of type I errors in non-normal situations, we first explored a bootstrap version of the original Levene test, resampling from the quantities Z_{ij} in the same way as in the homoscedastic case described above. Table 3 shows the results of simulations for exponential distributions. As before, nominal levels were set to 5%. Obviously, the probability of type I error exceeds by far the nominal level of 5% for the bootstrap Levene test, possibly because of the variance heterogeneity of the quantities Z_{ij} and the lack of independence. In the case of a normal or t_4 distribution, this bootstrap Levene test performs quite well with regard to type I error. However, as indicated by the lower part of Table 3, a strategy based on the preliminary bootstrap Levene test, followed by one-way analysis of variance bootstrap tests involving T_1 or T_2, still gives a reasonable error probability for the final conclusion, relating to location parameters.

We nonetheless decided to replace the variance test with a modified bootstrap version. Exploratory simulations were performed to compare different approaches. Considering absolute deviations $U_{ij} = |X_{ij} - M_i|$ from the sample median M_i, in analogy with a non-bootstrap version of the Levene test introduced by Brown and Forsythe (1974b), led to a more robust but still far from perfect procedure (Table 4). In the extensive study of Conover *et al.* (1981) of about 50 statistics used for testing homogeneity of variances, this particular statistic was among the very few which performed adequately. Another natural modification of the Levene test is to insert the transformed observations U_{ij} into T_2 rather than T_1 in the hope of eliminating the problem of unequal variances. Regardless of statistic used, our simulation results were rather similar, and each bootstrap Levene test had about the same performance as the corresponding non-bootstrap Levene test studied by Brown and Forsythe (1974b). We thus decided to retain the bootstrap version of the test for equal variances, with deviations U_{ij} from the sample median inserted into the T_1 statistic, and with resampling carried out as in the heteroscedastic situation.

Simulation results for the corresponding overall strategy, analogous to those in Table 1, are shown in Table 5. This bootstrap strategy is evidently superior to our first strategy for one-way analysis of variance. The problems involving the level for heteroscedastic exponential distributions have disappeared without any substantial reduction in power.

7. COMPARISON OF STRATEGIES WITH SEPARATE TESTS

The bootstrap strategy for one-way analysis of variance combines the use of the F and James statistics, preceded by a preliminary test. Additional, more extensive simulations were carried out to evaluate this strategy, denoted by S_5, to compare in particular with separate bootstrap tests based on T_1 and T_2 and the ordinary F test. Tables 6-8 show the results for seven different distributions: normal, t_4, exponential, χ_1^2, the S_U Johnson distribution with γ=0, δ=1 (which is symmetric with a very high kurtosis), uniform and a skewed beta distribution. When needed, distributions were scaled to obtain the correct mean and variance.

Tables 6-8 also include results for a slightly modified bootstrap strategy S_{20}, with a formal level for the Levene test of 20% rather than 5%. As the power of the bootstrap Levene test is relatively low for skewed distributions (Table 5), the statistic T_1 will tend to be selected too often for the last test in the strategy S_5. The increase in level for the Levene test is intended to reduce this bias.

The tables indicate that the bootstrap test based on T_1 is very similar to the ordinary F test with regard to significance level and power. This lack of improvement is exactly what can be expected with a bootstrap involving a statistic which is not asymptotically pivotal (Beran, 1988). For the parameter combinations considered, both tests reject the hypothesis too often for heteroscedastic skewed sampling distributions such as the exponential and χ_1^2, although this does not seem to occur with different sample sizes. In contrast, the bootstrap test based on T_2 is rather conservative for distributions with a large kurtosis, skewed or not. Of the two separate bootstrap tests, that based on T_1 has in general a higher power in homoscedastic cases, although any comparison must also take into account the predominantly higher level of this test. The test based on T_2 is superior in heteroscedastic situations, with a pronounced difference especially when sample sizes differ (Table 7).

To a large extent, the strategy S_5 appears to combine the favourable properties of the bootstrap tests based on T_1 and T_2. However, the probability of type I error is still quite high for heteroscedastic skewed distributions. The strategy S_{20} seems to avoid this problem, retaining in most cases the superior power of S_5. Whether the gain in using one of the strategies S_5 and S_{20} is large enough to justify such a procedure, must be decided in each practical application.

8. DISCUSSION

We have compared basic strategies for one-way analysis of variance, adapted to rather general classes of underlying sample distributions. More sophisticated strategies could easily be incorporated into knowledge-based systems, with more far-reaching conclusions about the location parameters. A system of this kind was constructed by Bell *et al.* (1989). Our emphasis has been on illustrating the evaluation of uncertainty through simulation, a purpose best served by simple strategies.

Our initial strategy attempted to combine the general flexibility of a non-parametric method with the superior power of methods based on normal distribution theory. The attempt largely failed, apparently because the Kruskal-Wallis test is not adapted to the null hypothesis prescribing equal location parameters with possibly different degrees of variation. Specific non-parametric tests for such hypotheses are not readily available (Lehmann, 1975; Section 7A). The subsequent strategies consist essentially of a preliminary variance test, followed by a specific test for location parameters. Preliminary tests for selecting statistics have been studied extensively in other models, in particular two-sample problems (Bancroft and Han, 1983; Markowski and Markowski, 1990; Moser and Stevens, 1992). In some situations, a preliminary test with an appropriate significance level can improve the overall performance although the gain is doubtful in other cases. Using a preliminary test before selecting a suitable bootstrap statistic in analysis of variance was suggested by Fisher and Hall (1990).

Efron and Tibshirani (1993) pointed out that bootstrap methodology can increase the degree of automation in statistical practice. This also holds true for more general problems handled by knowledge-based systems. Structural uncertainty is minimized with bootstrap strategies but not eliminated, as confirmed by our simulation results. The structural simplicity of bootstrap methods is another advantage, although this is not always reflected in the implementation. Neither bootstrap test included in our strategy was readily available in any package known to us. Our initial implementation was in S-Plus, but to achieve a higher efficiency in simulations, the routines were rewritten in Fortran calling the NAG library. As the finite sample properties of practical bootstrap methods have often been neglected (Young, 1994), a detailed evaluation of any strategy incorporating such procedures is needed.

We have outlined how simulation may reveal important attributes of strategies. The skepticism among some statisticians towards automated data analysis by knowledge-based systems may be related to the problem of measuring strategy performance. In a review of statistical methods based on computer technology, Hand (1994) emphasized that simulation studies require extensive planning. It is thus essential to recognize underlying distributions which are genuinely important. An initial comparative study of strategies can be more restricted than a subsequent investigation of the strategy thought to be nearly optimal. Even then, our final simulation study of the strategies S_5 and S_{20} included samples drawn from distributions of similar shape only. In view of the large number of possibilities, practical limits must be set to the simulations, depending on conditions likely to arise in applications.

For this reason, a simulation facility should be included in the knowledge-based system itself, enabling the user to investigate strategy performance under conditions relevant to his own practice. Thus, in the study of strategies for one-way analysis of variance, we have only been able to deal with 3 samples. As strategies cannot be evaluated in advance for all conceivable combinations of sample sizes, the possibility of studying finite sample properties in each relevant situation should also be of great value. Moreover, the user may be interested in joint distributions of several statements made while the strategy is executed, or possibly only in certain aspects of the distributions. A system incorporating an interface to a reliable, well-established package of generators is likely to attract most potential users. Such tools should make it easier to construct knowledge-based systems of real practical value. If the user is given an opportunity to study the behaviour of the strategy in a relevant setting, the system designer will not always need to be overly concerned about fine-tuning. On the other hand, complaints about poor strategies may be supported by simulation results.

In practical situations involving hypothesis tests, the simulation module will often be used to determine an estimate of the actual significance level when the observations provide relatively strong support in favour of the alternative hypothesis. If there is no evidence against the null hypothesis, estimates of power will be called for. Ideally the level of the procedure should be adjusted beforehand, but this may be difficult because of the complexity of the strategy and the inaccuracy of the assumptions. Even users who do not wish to take advantage of the implemented strategy in practical data analysis, may be interested in the simulation module for assessing uncertainty.

Constructing a general measure of the uncertainty attached to a particular conclusion, reached at the end of a long chain of rules, is a more difficult problem than evaluating a strategy under reasonable fixed parameter combinations. Data-based simulation, relying on ideas similar to parametric or non-parametric bootstrap, may provide essential information. The mechanism for generating random samples may then depend only on certain aspects of the observations. The objective will differ somewhat from that of ordinary bootstrap, as conclusions will be studied under circumstances related to those observed but still slightly different, an approach similar to model expansion (Draper, 1995). Appropriate distributions for resampling may be determined by methods such as non-parametric density estimation (Silverman, 1986) or adaptation to a flexible family of distributions such as the Johnson system (Aarseth and Heuch, 1996c).

When a strategy is evaluated through simulation, randomly drawn samples should be analysed by the implementation in the knowledge-based system. Only in this way can overall uncertainty be assessed, taking into account approximations, problems of scaling and accuracy of numerical procedures, possibly coded in external software. The simulation module may thus be used by the system designer as an effective debugger. Future work should aim at combining results from simulations with other techniques for evaluation and improvement of strategies. In particular, refinement facilities (Zlatereva and Preece, 1994) which can improve the strategy automatically on the basis of simulation results, represent an interesting challenge.

REFERENCES

Aarseth J.H. and Heuch I. (1996a). User's guide to Express: A tool for building knowledge-based systems for statistical data analysis. *Statistical Report* no. 27. Department of Mathematics, University of Bergen, Bergen.

Aarseth J.H. and Heuch I. (1996b). Interfaces in a knowledge-based statistical system, as exemplified by Express. *Statistical Report* no. 29. Department of Mathematics, University of Bergen, Bergen.

Aarseth J.H. and Heuch I. (1996c). Testing symmetry by a bootstrap based on Johnson distributions. Manuscript.

Bancroft T.A. and Han C.P. (1983). A note on pooling variances. *Journal of the American Statistical Association* **78**, 981-983.

Bell E., Watts P. and Alexander J. (1989). THESEUS: An expert statistical consultant. *American Journal of Mathematical and Management Sciences* **9**, 361-370.

Beran R. (1988). Prepivoting the test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* **83**, 687-697.

Bhatnagar R.K. and Kanal L.N. (1986). Handling uncertain information: a review of numeric and non-numeric methods. In: *Uncertainty in Artificial Intelligence,* Kanal L.N. and Lemmer J.F. (eds.). North-Holland, Amsterdam, 3-26.

Brown M.B. (1990). *BMDP Statistical Software Manual.* University of California Press, Berkeley.

Brown M.B. and Forsythe A.B. (1974a). The small sample behavior of some statistics which test the equality of several means. *Technometrics* **16**, 385-389.

Brown M.B. and Forsythe A.B. (1974b). Robust tests for the equality of variances. *Journal of the American Statistical Association* **69**, 364-367.

Carlsen F. and Heuch I. (1986). Express - An expert system utilizing standard statistical packages. In: *COMPSTAT. Proceedings in Computational Statistics*, de Antoni F., Lauro N. and Rizzi A. (eds.). Physica-Verlag, Heidelberg, 265-270.

Conover W.J., Johnson M.E. and Johnson M.M. (1981). A comparative study of tests for homogeneity of variances, with application to the outer continental shelf bidding data. *Technometrics* **23**, 351-361.

Dorda W., Froeschl K.A and Grossmann W. (1990). WAMASTEX - Heuristic guidance for statistical analysis. In: *COMPSTAT. Proceedings in Computational Statistics*, Momirovic K. and Mildner V. (eds.). Physica-Verlag, Heidelberg, 93-98.

Draper D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society* B **57**, 45-97.

Efron B. and Tibshirani R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Fisher N.I. and Hall P. (1990). On bootstrap hypothesis testing. *Australian Journal of Statistics* **32**, 177-190.

Gale W.A, Hand D.J. and Kelly A.E. (1993). Statistical applications of artificial intelligence. In: *Computational Statistics. Handbook of Statistics*, vol. 9, Rao R. (ed.). North-Holland, Amsterdam, 535-576.

Hall P. and Wilson S.R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics* **47**, 757-762.

Hand D.J. (1985). Statistical expert systems: necessary attributes. *Journal of Applied Statistics*

12, 19-27.

Hand D.J. (1987). The application of expert systems in statistics. In: *Interactions in Artifical Intelligence and Statistical Methods*, Phelps B. (ed.). Gower Technical Press, Aldershot, 3-17.

Hand D.J. (1994). Statistics and computing: the promise and the risk. In: *AI and Computer Power, the Impact on Statistics*, Hand D.J. (ed.). Chapman & Hall, London, 1-8.

Hjorth J.S.U. (1994). *Computer Intensive Statistical Methods*. Chapman & Hall, London.

Hodges J.S. (1987). Uncertainty, policy analysis and statistics. *Statistical Science* **2**, 259-291.

James G.S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika* **38**, 324-329.

Kruskal W.H and Wallis W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47**, 583-621.

Lehmann, E.L. (1975). *Nonparametrics. Statistical Methods Based on Ranks*. Holden-Day, Oakland.

Levene H. (1960). Robust test for equality of variances. In: *Contributions to Probability and Statistics*, Olkin I. (ed.). Stanford University Press, Palo Alto, 278-292.

Lindley D.V. (1987). The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Statistical Science* **2**, 17-24.

Loh W. (1987). Some modifications of Levene's test for variance homogeneity. *Journal of Statistical Computation and Simulation* **28**, 213-226.

Markowski C.A. and Markowski E.P. (1990). Conditions for the effectiveness of a preliminary test of variance. *The American Statistician* **44**, 322-326.

Moser BK. and Stevens GR. (1992). Homogeneity of variance in the two-sample means test. *The American Statistician* **46**, 19-21.

NAG. (1993). *NAG Fortran Library Manual.* Mark 16. NAG, Oxford.

Paul R.J. (1994). AI and simulations. In: *AI and Computer Power, the Impact on Statistics*, Hand D.J. (ed.). Chapman & Hall, London, 9-22.

von Rimscha M. (1994). Simulation of uncertainty: decision support in complex incompletely defined environments. In: *AI and Computer Power, the Impact on Statistics*, Hand D.J. (ed.). Chapman & Hall, London, 37-51.

Shafer G. (1987). Probability judgement in artificial intelligence and expert systems. *Statistical Science* **2**, 3-15.

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

Silverman B.W. (1986). *Density Estimation for Statistical Data Analysis*. Chapman & Hall, London.

Spiegelhalter D.J., Dawid A.P., Lauritzen S.L. and Cowell R.G. (1993). Bayesian analysis in expert systems. *Statistical Science* **8**, 219-283.

Streitberg B. (1988). On the nonexistence of expert systems - critical remarks on artificial intelligence in statistics. *Statistical Software Newsletter* **14**, 55-62.

Tan W.Y. and Tabatabai M.A. (1985). Some robust ANOVA procedures under heteroscedasticity and nonnormality. *Communications in Statistics - Simulation* **14**, 1007-1026.

Tan W.Y. and Tabatabai M.A. (1986). A robust procedure for comparing several means under heteroscedasticity and nonnormality. *Communications in Statistics - Simulation* **15**, 733-745.

Tung S.T.Y. and Schuenemeyer J.H. (1991). An expert system for statistical consulting, *Journal of Applied Statistics* **18**, 35-47.

Van den Berg G.M. and Visser R.A. (1990). Knowledge modelling for statistical consultation systems; two empirical studies. In: *COMPSTAT. Proceedings in Computational Statistics*, Momirovic K. and Mildner V. (eds.). Physica-Verlag, Heidelberg, 75-80.

Winston P. H. (1984). *Artifical Intelligence*. Addison-Wesley, Reading, Mass.

Young G.A. (1994). Bootstrap: more than a stab in the dark? *Statistical Science* **9**, 382-415.

Zlatareva N. and Preece A. (1994). State of the art in automated validation of knowledge-based systems. *Expert Systems with Applications* **7**, 151-167.

Table 1
Probabilities (in percent) of deciding that location parameters differ, estimated by 5000 simulations, for the basic analysis of variance strategy given in Fig. 1, with 3 samples of size 20.

	$\mu_1=0,\ \mu_2=0,\ \mu_3=0$	$\mu_1=0,\ \mu_2=0.5,\ \mu_3=1$	$\mu_1=0,\ \mu_2=1,\ \mu_3=2$
		Normal distribution	
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.1	78.2	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.5	22.0	69.2
		t_4 distribution	
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.3	87.5	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.3	29.1	86.0
		Exponential distribution	
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.8	95.5	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	22.9	7.1	74.1

Table 2

Results of 5000 simulations in Express, for the strategy in Fig. 1, with 3 samples of size 20 from scaled exponential distributions with $\mu_1=0$, $\mu_2=0$, $\mu_3=0$ and $\sigma_1=1$, $\sigma_2=2$, $\sigma_3=3$.

	Among all simulations leading to a conclusion			**Among simulations leading to a conclusion used in the inference**		
	Yes	No	Total	Yes	No	Total
Intermediate conclusions						
Are all samples normally distributed?	21	4979	5000	21	4979	5000
Is p-value for equal variance ≤ 0.05?	18	3	21	18	3	21
Is p-value associated with ordinary F ≤ 0.05?	3	18	21	1	2	3
Is p-value for Brown-Forsythe F ≤ 0.05?	2	19	21	1	17	18
Is p-value for Kruskal-Wallis test ≤ 0.05	1142	3837	4979	1142	3837	4979
Overall conclusion						
Do location parameters differ between samples?				1144	3856	5000

Table 3
Results of 5000 simulations for the bootstrap strategy using a bootstrap version of the original Levene test, with 3 samples of size 20 from a scaled exponential distribution.

	$\mu_1=0,\ \mu_2=0,\ \mu_3=0$	$\mu_1=0,\ \mu_2=0.5,\ \mu_3=1$	$\mu_1=0,\ \mu_2=1,\ \mu_3=2$
	Levene test: Probability (in percent) of deciding that variances differ.		
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	11.0	11.5	10.8
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	66.9	66.3	68.2
	Final conclusion: Probability (in percent) of deciding that location parameters differ.		
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.3	80.9	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	5.3	20.2	79.1

Table 4
Probabilities (in percent) of deciding that variances differ, estimated by 5000 simulations, using the median in the Levene bootstrap test, for 3 samples of size 20 from scaled exponential distributions.

	$\mu_1=0,\ \mu_2=0,\ \mu_3=0$	$\mu_1=0,\ \mu_2=0.5,\ \mu_3=1$	$\mu_1=0,\ \mu_2=1,\ \mu_3=2$
	Levene test: Probability (in percent) of deciding that variances differ		
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	3.1	3.1	2.8
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	44.8	44.8	44.1

Table 5
Probabilities (in percent) of deciding that location parameters differ, estimated by 5000 simulations, for the bootstrap strategy using the sample median in the Levene bootstrap test, for 3 samples of size 20.

	$\mu_1=0,\ \mu_2=0,\ \mu_3=0$	$\mu_1=0,\ \mu_2=0.5,\ \mu_3=1$	$\mu_1=0,\ \mu_2=1,\ \mu_3=2$
		Normal distribution	
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.7	80.0	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	5.4	31.4	84.2
		t_4 distribution	
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.3	87.1	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	4.7	36.9	87.6
		Exponential distribution	
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.5	79.5	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	5.6	21.1	79.4

Table 6

Probabilities (in percent) of deciding that location parameters differ, estimated by 5000 simulations, for the ordinary F test, bootstrap tests based on T_1 and T_2, and bootstrap strategies S_5 and S_{20}, with 3 samples of size 20.

	$\mu_1=0,\ \mu_2=0,\ \mu_3=0$					$\mu_1=0,\ \mu_2=0.5,\ \mu_3=1$					$\mu_1=0,\ \mu_2=1,\ \mu_3=2$				
	F	T_1	T_2	S_5	S_{20}	F	T_1	T_2	S_5	S_{20}	F	T_1	T_2	S_5	S_{20}
						Normal distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.2	5.7	4.4	5.7	4.7	79.6	80.1	75.8	80.0	79.6	100	100	100	100	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.4	6.6	5.1	5.4	4.8	23.1	24.0	31.6	31.4	30.4	72.3	72.2	85.0	84.2	83.8
						t_4 distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.8	5.3	3.9	5.3	5.0	87.1	87.2	83.1	87.1	87.3	100	100	99.9	100	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.0	6.7	4.1	4.7	3.6	28.7	29.9	36.9	36.9	35.0	81.6	82.0	88.9	87.6	88.8
						Exponential distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.3	4.9	3.7	4.5	4.4	79.2	79.6	62.4	79.5	79.6	99.9	99.9	99.2	100	99.9
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.5	7.2	3.7	5.6	4.4	18.0	19.2	22.6	21.1	22.0	78.9	80.9	77.7	79.4	77.4
						χ_1^2 distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	3.8	5.1	3.8	4.5	3.6	80.8	81.0	51.3	81.8	78.4	99.7	99.8	95.9	99.7	99.4
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	7.7	8.6	3.0	6.9	5.4	16.6	19.0	17.8	16.9	17.4	82.6	84.3	65.9	81.8	72.9
						Johnson's S_U distribution; $\gamma=0$, $\delta=1$									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.3	4.9	2.2	4.9	4.6	81.7	82.5	73.9	82.5	81.1	99.5	99.6	98.2	99.6	99.4
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	5.0	5.7	2.6	4.4	3.5	27.8	28.9	31.3	31.3	30.8	76.5	77.5	79.1	79.5	81.0
						Uniform distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.8	5.1	4.3	5.1	5.8	79.5	79.8	76.4	79.8	78.3	100	100	100	100	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.1	6.5	5.2	5.3	5.1	22.2	23.1	31.5	31.6	32.2	71.7	72.5	85.8	85.7	86.1
						Beta distribution; $p=2$, $q=1$									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.1	5.7	4.8	5.8	5.3	79.0	79.4	73.9	79.3	78.2	100	100	100	100	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.5	6.5	4.6	5.1	4.7	23.6	24.3	32.0	31.7	30.8	69.5	70.0	83.9	83.8	84.3

Table 7
Probabilities (in percent) of deciding that location parameters differ, estimated by 5000 simulations, for the ordinary F test, bootstrap tests based on T_1 and T_2, and bootstrap strategies S_5 and S_{20}, with 3 samples of size n_1=10 n_2=20 and n_3=30.

	$\mu_1=0,\ \mu_2=0,\ \mu_3=0$					$\mu_1=0,\ \mu_2=0.5,\ \mu_3=1$					$\mu_1=0\ \mu_2=1\ \mu_3=2$				
	F	T_1	T_2	S_5	S_{20}	F	T_1	T_2	S_5	S_{20}	F	T_1	T_2	S_5	S_{20}
Normal distribution															
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.9	5.2	3.8	5.2	5.2	71.6	72.2	62.2	72.1	70.3	99.9	99.9	99.7	99.9	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	1.9	2.1	2.8	2.9	3.0	9.9	10.4	24.8	24.1	23.8	48.3	49.2	79.3	77.6	79.4
t_4 distribution															
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.1	5.8	3.1	5.8	5.2	80.2	80.3	71.4	80.4	80.0	99.7	100	99.2	100	99.9
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	1.7	2.0	2.4	2.4	2.7	13.2	13.6	28.8	26.2	30.0	48.1	61.2	82.0	78.2	83.1
Exponential distribution															
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.9	5.8	4.8	5.8	5.0	72.9	74.1	53.3	74.0	73.7	99.7	99.7	93.3	99.7	99.5
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	2.8	3.2	2.2	3.5	3.0	6.3	7.2	22.5	17.9	22.3	48.1	50.6	67.9	53.5	64.9
χ^2_1 distribution															
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.3	5.1	4.7	4.7	4.5	76.1	77.2	46.3	77.0	73.6	99.4	99.5	84.1	99.4	98.6
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	3.5	4.0	1.3	3.8	3.1	5.4	6.4	18.8	10.6	18.4	52.4	55.7	57.3	52.2	50.4
Johnson's S_U distribution; γ=0, δ=1															
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.1	5.5	4.1	5.6	4.5	74.9	75.9	64.4	76.0	76.0	99.1	99.2	95.5	99.1	98.8
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	1.8	2.0	1.8	2.3	2.0	13.8	15.0	25.6	22.0	25.5	56.0	57.3	72.9	66.6	72.6
Uniform distribution															
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.1	5.5	4.1	5.6	5.5	71.0	71.4	60.8	71.2	69.6	100	100	99.9	100	100
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	2.4	2.5	3.9	3.9	3.3	9.6	9.9	24.4	24.2	24.5	46.2	47.2	79.9	79.7	81.2
Beta distribution; p=2, q=1															
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.1	5.4	4.3	5.7	5.9	70.4	71.0	57.0	70.5	68.8	100	100	99.0	99.9	99.8
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	2.1	2.6	3.2	3.3	4.7	11.2	11.7	22.5	21.7	23.4	47.7	48.2	77.8	77.2	77.6

Table 8
Probabilities (in percent) of deciding that location parameters differ, estimated by 5000 simulations, for the ordinary F test, bootstrap tests based on T_1 and T_2, and bootstrap strategies S_5 and S_{20}, with 3 samples of size 10.

	$\mu_1=0,\ \mu_2=0,\ \mu_3=0$					$\mu_1=0,\ \mu_2=0.5,\ \mu_3=1$					$\mu_1=0,\ \mu_2=1,\ \mu_3=2$				
	F	T_1	T_2	S_5	S_{20}	F	T_1	T_2	S_5	S_{20}	F	T_1	T_2	S_5	S_{20}
						Normal distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.7	4.9	2.8	4.8	5.0	45.5	46.5	32.9	46.1	43.8	97.1	96.9	91.2	96.8	96.5
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.2	6.5	3.5	5.6	4.1	14.1	14.6	12.9	15.0	13.7	39.9	40.2	44.7	43.3	44.7
						t_4 distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.3	5.7	2.3	5.6	4.9	56.8	57.8	41.8	57.5	55.8	97.7	97.8	92.4	97.8	97.4
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.6	6.9	3.0	5.8	3.8	16.9	17.7	14.2	17.4	15.8	50.5	51.8	53.1	53.2	53.7
						Exponential distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.0	4.3	3.1	4.1	4.1	51.0	52.8	25.8	52.5	50.1	95.3	95.5	73.2	95.4	93.9
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	7.3	7.7	2.8	6.8	5.7	8.3	9.2	6.8	7.4	8.1	37.8	39.6	32.5	37.0	35.5
						χ_1^2 distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.0	4.7	3.2	4.0	3.1	55.1	56.9	20.8	56.7	53.0	93.5	94.3	57.1	93.8	92.1
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	9.2	10.2	2.9	9.0	6.4	8.3	9.4	5.9	7.2	5.6	42.8	46.3	24.5	43.6	32.4
						Johnson's S_U distribution; $\gamma=0,\ \delta=1$									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.7	5.2	1.6	5.1	4.0	56.8	58.1	37.6	58.0	55.1	94.6	94.8	84.6	94.8	94.3
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	5.0	5.4	1.6	4.9	3.8	17.3	18.4	12.7	18.4	17.1	52.2	53.5	46.7	52.7	52.1
						Uniform distribution									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	5.4	5.8	3.4	5.9	5.3	43.9	44.6	30.0	44.2	43.5	98.0	97.9	92.4	97.9	97.5
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.3	6.7	3.7	6.3	5.0	13.4	13.8	12.6	14.3	13.7	36.9	37.3	43.6	42.1	45.5
						Beta distribution; $p=2,\ q=1$									
$\sigma_1=1,\ \sigma_2=1,\ \sigma_3=1$	4.8	5.4	3.3	5.6	5.4	44.9	45.2	28.5	44.8	44.2	97.5	97.5	89.0	97.4	96.7
$\sigma_1=1,\ \sigma_2=2,\ \sigma_3=3$	6.8	7.1	3.4	6.6	5.0	15.5	16.1	12.0	15.5	14.0	38.6	39.3	41.2	41.7	42.5

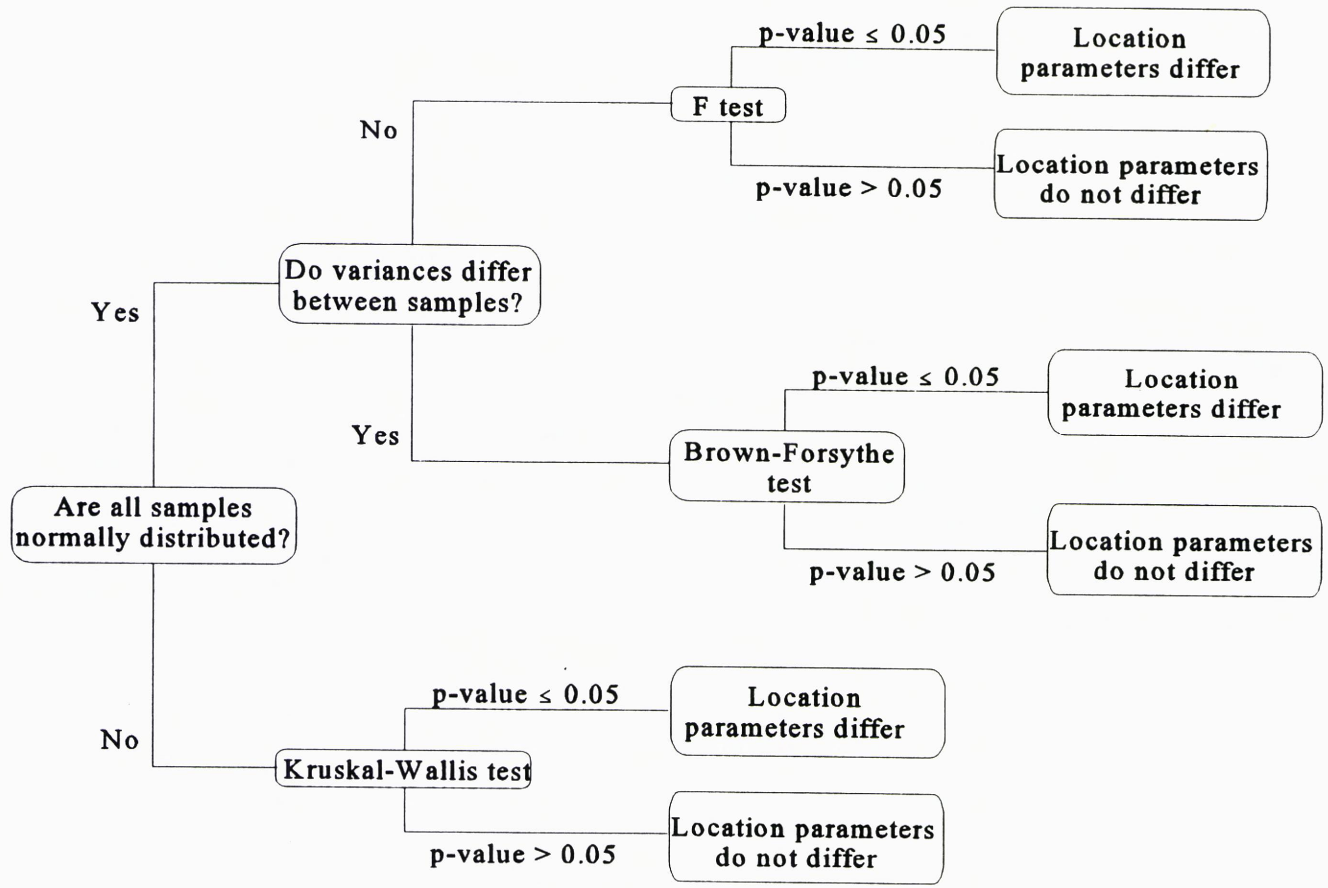

Fig.1 Strategy map for one-way analysis of variance based on standard tests.