

IMPROVING AGRICULTURAL EDUCATION RESEARCH

J. Dale Oliver
Vocational Education

Virginia Polytechnic Institute
and State University

In recent years, there has been an increase in agricultural education research. The rate of increase was accelerated by the appropriation of funds specifically designated for research through the Vocational Education Act of 1963 and subsequent federal legislation.

The growth in agricultural education research has been accompanied by an increase in the use of statistical techniques, with both positive and negative results. From a positive standpoint, more complex problems are being investigated, the information produced is becoming more meaningful and the efficiency of the research is increasing. The negative effects primarily are that statistical techniques are being used in cases where the assumptions have not been met and there is generally a failure to distinguish between statistical significance and practical importance.

The purpose of this article is to identify some statistical problems which exist in agricultural education research. Hopefully, the suggested alternatives for dealing with the problems will result in improvement in future research efforts. Specifically, the following two problems will be addressed: (1) Is the Type I error rate affected when non-independent multiple statistical tests are made? and (2) Does statistical significance automatically mean that the results are of practical importance?

Multiple Statistical Tests and the Error Rate

Discussion of the Problem

Quite frequently, data are available from two independent samples. The researcher is interested in testing the null hypothesis that the means of the populations from which the samples were drawn are equal. The Type I error rate (the probability of rejecting a true null hypothesis) or the significance level is simply α .

Suppose the case is different and data are available from three independent samples. Interest again is in testing the null hypothesis that the population means are equal. A simple approach, it would appear, is to run a series of t-tests for the differences among the means, using the means in all possible combinations of two. This would involve three tests (the means of samples 1 vs. 2, 1 vs. 3, and 2 vs. 3).

If all possible comparisons are run, the error rate used for each comparison would be α and this is known as the "comparison-

wise error rate" (Hinkle, Wiersma, and Jurs, 1979, p. 270). In the example, three comparisons would be run at the pre-determined alpha level.

Is the error rate equal to α for the set of three comparisons? To answer this, Hinkle *et. al.* (1979) utilized the concept of the "experiment-wise error rate (α_E)" which "is the probability of making a Type I error for the set of all possible comparisons" (p. 270). The experiment-wise error rate is determined as $\alpha_E = 1 - (1 - \alpha)^c$, where α = comparison-wise error rate and c = number of comparisons. In the three-group example, using $\alpha = .05$, $\alpha_E = .14$. This means that the probability of making at least one Type I error in the comparisons of the three means is not .05 but is .14. If one uses $\alpha = .05$ and makes all possible comparisons with a four-group design, $\alpha_E = .27$; if the design includes five groups, $\alpha_E = .40$; and for six groups, $\alpha_E = .54$.

To deal with this problem, R. A. Fisher developed the technique of analysis of variance (ANOVA), which permits one to simultaneously test the equality of all means while maintaining the error rate at the α -level for the set of comparisons. When a significant F -ratio is found, one still faces the problem of determining the combination or combinations of means which differ significantly. The procedures for doing this are known as "post hoc multiple comparison tests" (Hinkle *et. al.*, 1979, p. 268).

One alternative in making the post hoc comparisons is to make all pair-wise comparisons and test the comparisons at a smaller α than the desired α_E -level. The formula for calculating α_E can be used to determine α . In the three-group example, α would be set at .017 in order to have $\alpha_E = .05$. Therefore, each of the three comparisons would be tested at $\alpha = .017$ in order to give an experiment-wise error rate of .05. Hinkle *et. al.* (1979) noted that this procedure is extremely conservative and indicated that a less conservative approach is to apply either the Tukey or Newman-Keuls post hoc multiple comparison test for equal group sample sizes and the Scheffé method for unequal group sample sizes.

Illustration of the Problem from Research

In the three-group example, the problem of using t -tests in the multiple comparisons arose because of overlapping information with respect to the independent variable. The respective tests were not independent and thus α_E was not equal to α .

Cases involving multiple comparisons when there is overlapping information with respect to the independent variable are difficult to find in the literature. The problem does arise, however, in studies where there is overlapping information with respect to the dependent variables. Such studies relate to competencies, attitudes, beliefs, ratings of teaching, and others. They generally involve one independent variable (which categorizes by position, occupation, age, etc.) and a large number of depen-

dent variables (such as competencies, beliefs, etc.). Each respondent rates all dependent variables on some sort of scale. The responses are normally regarded as correlated data because the response a person gives to one item is not independent of the responses to the other items. Such is the case unless a correlation matrix reveals there is no correlation between the items (dependent variables).

The correlation between items indicates the existence of overlapping information. When multiple t-tests are made on all possible sets of the means of such data, α_E is not equal to α but must be computed by the formula $\alpha_E = 1 - (1 - \alpha)^C$.

Some reports of research involving the comparison of multiple dependent variables will now be examined. From the nature of the studies, one would expect the data to be correlated and no evidence was presented to the contrary.

The first study to be considered was conducted by Williams (1979) and focused on the benefits students received from their vocational agriculture occupational experience program. The independent variable classified the respondents as Chapter Farmer Degree recipients and State Farmer Degree recipients. The questionnaire used included 40 statements of benefits from the program. Thus, the study involved 40 dependent variables. With regard to significance tests, the researcher reported the following (Williams, 1979):

The mean value and standard deviation were computed for each benefit item, and the t-test was used to test for significant differences between chapter farmers and state farmers. (p. 34)

This implies that 40 separate t-tests were run. If $\alpha = .05$, then $\alpha_E = .87$. This means that the probability of committing at least one Type I error over the set of 40 comparisons is not .05 but is estimated to be .87.

Another example of multiple comparisons is a study by Carter and Crawford (1977), which dealt with the professional competencies needed by beginning teacher educators in agricultural education. The independent variable categorized the respondents into three groups. A questionnaire containing 114 professional competencies was used. Thus, the study involved one independent and 114 dependent variables. The researchers reported the following data analysis: "Analysis of variance scores were computed for each of the competencies for a comparison among subgroups" (Carter and Crawford, 1977, p. 20). Since ANOVA was used, did this not take care of the problem? No--there were merely 114 F-tests rather than 343 t-tests. ANOVA was used because the independent variable involved three groups.

What would α_E be for the 114 comparisons? Using $\alpha = .05$, it would be about .997. The probability of at least one Type I error over the set of comparisons is almost 1.0!

Recommended Solutions to the Problem

The solution sometimes proposed is to run an ANOVA for the total set of variables. This would be inappropriate, however, since ANOVA is only used where there is one dependent variable. The appropriate technique to consider is multivariate analysis of variance (MANOVA). This technique extends the concepts of ANOVA to include two or more dependent variables and one or more independent variables in a single analysis. The drawback to using MANOVA is that the results would be somewhat difficult to interpret since, in the illustrations, a large number of comparisons would be involved.

Before using MANOVA in the studies illustrated previously, it would be desirable to reduce the number of dependent variables under consideration. There is a statistical procedure known as factor analysis which is useful for grouping the items into factors based on underlying constructs. In many instances, it provides the data-reducing capacity needed in multiple comparison studies and should be used prior to MANOVA if there is a sufficient number of cases relative to the number of variables. Cattell (1952) has suggested the following ratios of cases to variables:

While one might reasonably aim to define the principal factors operating in a certain realm on as few as twenty variables and eighty persons, it is desirable for other purposes, e.g., factor estimation and individual prediction, to have about fifty variables and four or five hundred subjects as a minimum. (p. 390)

It would appear that a simple solution to the problem of multiple statistical tests is to determine and use an α -level that will give the desired α_E . With a large number of comparisons, α becomes extremely small, and it may be difficult to determine the value to use. As was noted previously, such a procedure also tends to be very conservative and may result in no significant comparisons even where the overall statistical test is significant.

Whatever one decides to do, it is important to clearly indicate the error rate used. For the benefit of the users of the research, it would seem that α_E should be presented and explained if it differs from α .

Statistical Significance vs. Practical Importance

Discussion of the Problem

A review of several studies in agricultural education indicates that researchers often equate statistical significance and practical importance. This observation is based on the fact that the interpretation of data is dependent largely on the results of significance tests.

Is there a problem in equating statistical significance and practical importance? The answer is yes if one considers the importance of sample size in determining statistical significance. As Hays (1963) noted: "Virtually any study can be made to show significant results if one uses enough subjects, regardless of how nonsensical the content may be" (p. 326). Hinkle *et. al.* (1979) provided a more explicit statement on sample size:

The observant reader may have noted that a sufficiently large sample size will lead to the rejection of any null hypothesis based upon a fixed difference between the hypothesized parameter and the observed sample statistic. On the other hand, a researcher may not reject a null hypothesis even though there is a seemingly large difference between the parameter and the corresponding statistic. It is possible that the sample size is then too small. (p. 192)

Since statistical significance depends so heavily upon sample size, it may or may not indicate practical significance. There is a need, then, for additional information to aid in determining practical importance.

Illustration of the Problem from Research

Examples of studies where there is a tendency to equate statistical significance with practical importance are not difficult to find. Such was the case in a Master of Science study by Kittrell (1978) which determined the relationship between selected variables and the morale of vocational agriculture teachers. There were four findings in the study and each was supported by the results of a statistical test. These findings were followed by four conclusions which related to the results of the statistical tests. It appears that the interpretation of the data was based entirely on the results of the significance tests with no additional analysis.

The magnitude of this problem was illustrated by McNamara and Gill (1978). These individuals did a reanalysis of the research studies presented in the first eight issues of *The Journal of Vocational Education Research*. They found ten inquiries that employed a univariate analysis of variance as the basic statistical model for testing hypotheses. They noted that in almost all cases there was an avoidance of any reference to practical significance.

Recommended Solutions to the Problem

Hays (1963) expressed concern about the need to define the strength of a statistical association. He noted, in general, "that the strength is reflected by the extent to which knowing X reduces uncertainty about Y" (Hays, 1963, p. 325). He used an index which is called ω^2 (Greek omega, squared) and noted that it is sometimes called "the proportion of variance in Y accounted for by X" (Hays, 1963, p. 325). For a univariate analysis of

variance with one independent variable, Hays (1963) indicated that a reasonable estimate of ω^2 is given by:

$$\text{est. } \omega^2 = \frac{\text{SS between} - (k - 1) \text{MS within}}{\text{SS total} + \text{MS within}}$$

where,

SS between = sum of squares between groups
SS total = total sum of squares
MS within = mean square within groups
k = number of levels of the independent variable

The index ω^2 takes into account the sample size and can assume values ranging from zero to unity.

In a discussion of substantive significance, Gold (1969) notes that "the use of a correlation coefficient calls direct attention to the degree of association" (p. 45). If the correlation coefficient is squared, it shows the proportion of variance in Y that can be associated with the variance in X. Thus ω^2 is, as Hays (1963) noted, almost identical to r^2 .

In their discussion of practical significance, McNamara and Gill (1978) indicated that the practical assessment often begins by asking the question: "How much of the variance in a criterion measure can be accounted for by a prediction measure?" (p. 29). They chose to use the index ω^2 in answering the question. They illustrated the use of ω^2 in an ANOVA study and found, for example, a case where the relationship between the independent and dependent variables was highly significant ($\alpha < .01$) statistically while it was estimated that less than two percent of the variance in the dependent variable was associated with the variance in the independent variable (McNamara and Gill, 1968, p. 37). How large should the percentage be to indicate practical significance? Although there is no general answer, Cohen (1977) has offered the following definitions of effect size for the behavioral sciences:

Small effect size: $r^2 = .01$
Medium effect size: $r^2 = .09$
Large effect size: $r^2 = .25$ (pp. 79-80)

Statistical significance does not automatically indicate that the results are of practical importance. It is recommended that ω^2 be used in estimating the strength of the statistical association. The index should be interpreted so that, as its value increases, the practical importance of the results increases.

Summary and Conclusions

The recent growth in agricultural education research has been accompanied by an increase in the use of statistical techniques. Some important statistical problems have arisen, two of which were

addressed in this article. The proposed solutions to these problems appear to be reasonable. Their implementation should not unduly constrain researchers in carrying out their inquiries.

It is strongly recommended that efforts be made to solve these problems. In so doing, the credibility of research in agricultural education will be greatly increased and our reputation in the research community will be enhanced.

References

- Carter, R. I., and Crawford, H. R. "What Professional Competencies are Needed by Beginning Teacher Educators in Agricultural Education?" *The Journal of the American Association of Teacher Educators in Agriculture*, 1977, XVIII (2), 19-24.
- Cattell, R. B. *Factor Analysis*. New York: Harper and Brothers, 1952.
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press, 1977.
- Gold, D. "Statistical Tests and Substantive Significance." *The American Sociologist*, 1969, 4 (1), 42-46.
- Hays, W. L. *Statistics for Psychologists*. New York: Holt, Rinehart, and Winston, 1963.
- Hinkle, D. D., Wiersma, W., and Jurs, S. G. *Applied Statistics for the Behavioral Sciences*. Chicago: Rand McNally, 1979.
- Kittrell, D. "Relationships Between Selected Teacher and Job Setting Variables and Morale of Vocational Agriculture Teachers in Mississippi." *The Journal of the American Association of Teacher Educators in Agriculture*, 1978, XIX (2), 17-21.
- McNamara, J. F., and Gill, D. H. "Practical Significance in Vocational Education Research." *The Journal of Vocational Education Research*, 1978, III (3), 27-48.
- William, D. L. "Benefits Received from Supervised Occupational Experience Programs as Perceived by Students." *The Journal of the American Association of Teacher Educators in Agriculture*, 1979, XX (2), 33-40.