Bulletin of the Psychonomic Society
1985, 23 (6), 513-516

Expected mean squares in psychological statistics: A brief history

JOHN GAITO and PETER SHERMER
York University, Downsview, Ontario, Canada

Statistical models and expected mean squares [E(MS)] are important concepts that facilitate the extensive use of analysis of variance designs. These concepts were developed in the basic statistics area from 1939 through the 1950s. They were introduced into psychological statistics during the late 1950s and have been useful in attacking some statistical problems. Also, they simplify the teaching of ANOVA designs.

Rucci and Tweney (1980) indicated that analysis of variance techniques (ANOVA) in psychology developed in three stages: (1) an initial, expository phase lasting until World War II; (2) a wartime interregnum in which the use of ANOVA declined, and (3) a postwar resurgence. During this last period, there has been increased use of ANOVA procedures in psychological research, as indicated in a survey of APA journals by Edgington (1974) for the period 1948 through 1972. One interesting event that probably was instrumental in the resurgence during Stage 3 was the development and use of the concept of expected value of mean squares [E(MS)].

Lovie (1979) discussed the emergence of ANOVA in experimental psychology research during the period 1934-1945 (Stages 1 and 2 of Rucci & Tweney, 1980). He indicated the slowness of its acceptance by psychologists and the difficulties involved in understanding these statistical procedures. The availability of the concept of E(MS) at that time would have reduced these difficulties.

As one might expect, the E(MS) concept was introduced early in the basic statistics area, in 1939, and became quite prominent in the 1950s. Then the concept found its way into psychology during the late 1950s and early 1960s. During the 1960s and later, the E(MS) concept has become common in psychological statistics literature. In this article, we will provide a brief history of the development and use of the E(MS) concept in the literature of two broad areas, basic statistics and psychological statistics, during the period from 1939 to 1960. The paper will be devoted to two aspects: the development of the E(MS) concept in basic statistics and in psychological statistics and its use in psychology.[1]

The first author would like to express his gratitude to Joseph Albert Greenwood for introducing him to mathematical models and expected mean squares early in his career as a psychologist. Also we want to thank Joseph Greenwood, Leslie Atkinson, and Caroline Davis for reading and commenting on an earlier version of this paper. The authors' mailing address is: Department of Psychology, York University, 4700 Keele Street, Downsview, Ontario, Canada M3J 1P3.

DEVELOPMENT IN BASIC STATISTICS LITERATURE[2]

Papers

Scheffé (1956a) credits a number of writers with reports using the equivalent of E(MS) within a type of components of variance analysis. The individuals cited by Scheffé were Airy in 1861, Chauvenet in 1863, Fisher in 1918 and 1925, Jackson in 1939, and Tippett in 1929 and 1931. But the earliest writer to use expected values of mean squares in systematic fashion was Daniels (1939). He provided an example from the textile industry in England involving the E(MS) for three factors and differentiated between random and systematic (fixed) sources in his concern with components of variance in an experimental analysis.

The first complete treatment of a components of variance analysis was given by Crump (1946). He began with a mathematical model for a two-dimensional design. From this model, he stated the assumptions of normality, independence, and homogeneity of errors; then he differentiated between random and fixed sources; finally, he obtained the average value of each mean square and estimated each component involved in the mean square. Crump's (1946) paper led to a components of variance report in 1946 by Satterthwaite, who provided details of an approximate distribution for estimates of variance components.

Two important papers followed. In his analysis of the assumptions involved in ANOVA models, Eisenhart (1947) made extensive use of mathematical models and E(MS) ideas. Two years later, an unpublished memorandum report was issued by the Statistical Research Group at Princeton University; Tukey (1949) indicated the mathematical models and E(MS) for ANOVA designs under many different conditions or assumptions.

A large number of papers were produced in the 1950s. In 1951, Crump provided a review of components of variance analyses. He discussed the estimation of variance components involved in the E(MS) for a number of models. Another unpublished paper which influenced the writings of a number of psychologists was that by Greenwood (1952, 1956)[3]. In this paper, he provided the mathematical model and E(MS) for the four models with a three-factor design (3, 2, 1, and 0 random variables). At the end of his article, Greenwood showed how the components in each E(MS) could be determined (i.e., an estimation of the components of variation). This latter aspect was important in the later development of various types of intraclass correlation formulations (e.g., Gaito, 1958a; Haggard, 1958) which assumed importance in psychology for the estimation of experimental effects.

Two important papers appeared in 1955. One was a technical report by Harter and Lum (1955) from the Wright Air Development Center of the United States Air Force. These individuals were concerned with the mathematical models and E(MS) for a number of partially hierarchial designs. These designs are used frequently in psychology studies when repeated measurements are involved. Wilk and Kempthorne (1955) contributed a paper concerned with mathematical derivations within various ANOVA models and used mathematical models and E(MS) extensively. This paper was a frequently cited one for a number of years.

Three papers in 1956 were important ones. In one, Scheffé (1956a) presented a review of various ANOVA models with associated E(MS). His chief interest was in the mathematical models—the formulation, motivation, and statistical influence involved. Another paper (Scheffé, 1956b) was concerned with the mixed model in ANOVA, that is, one random and one fixed effect. In a third paper, Cornfield and Tukey (1956) discussed the E(MS) for factorial designs. They stressed the importance of E(MS) as a basis for significance tests or confidence statements. They presented the general E(MS), with suppressor coefficients for some components (1-x/X), and the three specific cases for a two-factor design. They also presented the general E(MS) case for a three-factor design.

There were a number of important papers during the middle and late 1950s. Schultz (1955) provided a set of rules that simplified the process of determining E(MS) for various designs. His rules were similar to those of Greenwood (1952), but extended to additional ANOVA designs. In 1957, Wilk and Kempthorne were concerned with the model and E(MS) for a Latin square design. One main conclusion was that the error term might be overestimated and the component of variation associated with the treatment effect might be underestimated.

By 1960, mathematical models and E(MS) were common in basic statistics. Thus, in ending this section, it is appropriate to cite the 1960 review of ANOVA models by Plackett in which he outlined the effect of assumptions on E(MS) for different models and discussed the problems which may arise in estimating components of variation.

Books

For many years, Snedecor (e.g., 1946) had used what was essentially E(MS) ideas to illustrate various points throughout his book. However, he did not use them in a systematic fashion for any ANOVA designs.

The earliest books to use E(MS) in a systematic way were those by Mood (1950), Anderson and Bancroft (1952), Kempthorne (1952), and Federer (1955). These books made extensive use of E(MS) for a number of ANOVA designs. An excellent book by Scheffé was published in 1959. This book presented a geometric interpretation of ANOVA designs. The E(MS) concept was employed extensively. The books by Anderson and Bancroft (1952), Kempthorne (1952), Federer (1955), and Scheffe (1959) have influenced greatly the thinking of individuals in basic statistics and in psychological statistics.

DEVELOPMENT IN PSYCHOLOGY LITERATURE

Papers

Although the systematic E(MS) concept first appeared with Daniels in 1939 and was prominent in basic statistics literature during the 1940s and early 1950s, it was not until the late 1950s that these ideas appeared in psychological literature. The earliest paper, by Julian Stanley, was published in *Psychological Reports* in 1956. He provided the general E(MS) case for a three-factor design and indicated how the fixed, random, and mixed models could be obtained for this case.

Gaito had been influenced greatly by the Greenwood paper (1952). He recognized that the E(MS) concept clarified the various ANOVA designs, and, in 1958, he began to use E(MS) relative to a number of statistical problems. Also he presented the general E(MS) case of Cornfield and Tukey and the resulting specific cases for a three-factor design (1960). His papers over a few years brought to the attention of psychologists the E(MS) concept within a number of designs or problems [e.g., magnitude estimation (1958a), the Latin square design (1958b), and counterbalancing problems (1958c)].

Danford, Hughes, and McNee (1960) also found E(MS) of value in discussing a repeated measurements design of interest in many psychology studies, namely, a partially nested or partially hierarchical design.

In an excellent paper in 1960, Green and Tukey considered problems that arise in complex ANOVA designs. The main design they discussed was a repeated measurements analysis. They used E(MS) and components of variation to aid in the understanding of experimental results.

Books

The usual approach to ANOVA in psychological statistics books either was "cookbook" in nature or was an attempt to provide an intuitive understanding of the procedures involved. The earliest treatment of ideas similar to E(MS) within psychological statistics textbooks was by Lindquist (1953) and McNemar (1955). Both of these books had some E(MS)-type examples, but did not provide them in an effective and systematic fashion characteristic of some basic statistics books at that time (e.g., Anderson & Bancroft, 1952; Kempthorne, 1952).

In the last chapter of his book, Lindquist (1953) provided a discussion of E(MS) with a number of designs under the rubric of "Estimation of variance components in reliability studies." It is unfortunate that these ideas were not incorporated into the earlier chapters with each design. If so, readers probably would have been more apt to have read and understood E(MS) notions, and the excellent book would have been more important. The ideas in a separate chapter probably were overlooked by most readers.

Likewise, McNemar (1955) headed in the direction of E(MS) treatment in his chapter on complex ANOVA designs with an attempt to determine components present in the variance estimates (mean squares). He provided the equivalent of E(MS) for a number of models of two- and three-factor designs.

The earliest psychology book to use E(MS) in a systematic fashion was by Haggard (1958). His book provided an excellent treatment of intraclass correlation formulations for the determination of magnitude of experimental effects. It is unfortunate that this book has been overlooked by many psychologists. Very few papers on magnitude estimation which were published in psychological journals show an awareness of this excellent book.

A book by Ferguson in 1959 utilized E(MS) ideas with ANOVA designs. In his treatment of the one- and two-factor designs, he employed E(MS) extensively to provide an understanding of what was involved in each variance estimate and the basis for the use of specific error items in each F test.

During the 1960s, the use of mathematical models and E(MS) in an efficient fashion became commonplace for most books on ANOVA (e.g., Kirk, 1968).

USE OF E(MS) IN PSYCHOLOGY

In Evaluating Specific Problems

Practical significance. A major use of E(MS) in psychology has been relative to theoretical and practical problems. Probably its most valuable application has been in the estimation of magnitude of specific effects via an intraclass-correlation-type procedure. When a result of statistical significance occurs, some researchers have become concerned with practical significance as well: Is the magnitude of the effect sufficient to support the conclusion that it is large enough to be of practical importance? Haggard's excellent book in 1958 described the value of intraclass correlation procedures in handling this problem.

The E(MS) of ANOVA designs with a number of components of variation lead easily to an intraclass correlation formulation which is of the following nature: $\sigma_x^2/\sigma_x^2 + \sigma_e^2$, where σ_x^2 is the component associated with the variable of concern (X), and σ_e^2 is the error associated with X. This formulation indicates the proportion of variation contributed by X.

Gaito's coefficient of utility (1958a) was essentially a type of intraclass correlation similar to one of Haggard's formulations. More recently, some individuals have used ω^2 (Hays, 1963) when fixed effects are present. However, ω^2 can be considered as an intraclass-correlation-type measure, for it is defined in a somewhat similar fashion. Different versions of ω^2 have been discussed by a number of individuals (e.g., Dodd & Schultz, 1973; Fleiss, 1969; Vaughan & Corballis, 1969). In these cases, E(MS) has been useful.

Reliability aspects. Intraclass correlation can provide a type of reliability statement. Since the 1940s, psychologists and educators have been using an intraclass-correlation-type measurement to determine reliability. If we consider σ_1^2 as reliable variance and σ_2^2 as error variance, then the intraclass correlation indicates the proportion of reliable variance present, which is the basic definition of reliability. This measure has been used (directly or indirectly) to estimate the reliability of a test (e.g., Burt, 1955; Guilford, 1954; Hoyt, 1941) and of an experiment (Alexander, 1947; Gaito, 1959).

In psychological testing, variance due to subjects is considered as reliable variance, usually the only reliable variance of concern. In this case the estimation is $(MS_s - MS_e)/MS_s$ or, in terms of E(MS), as indicated above, where a coefficient appears before σ_s^2(i.e., $p\sigma_s^2/p\sigma_s^2 + \sigma_e^2$); p is the number of items used in determining reliability. Notice that, the greater the number of such items used, the greater the reliability. Going from the case where the coefficient of σ_s^2 is 1 to that in which it is p is the same as using the Spearman-Brown formula to estimate a test p times as long as the one for which reliability has been determined.

Other issues. The E(MS) concept has been of value in other cases also. Gaito employed E(MS) extensively in dealing with a number of issues, for example, F test bias within a Latin square design (1958b); problems of counterbalancing in simple or complex ANOVA designs (Gaito, 1958c, 1961).

In Teaching ANOVA

There are numerous methods of teaching ANOVA procedures, ranging from a rigorous mathematical treatment to a simplified "cookbook" approach. The obvious advantage of the former is the mathematical sophistication attained by those who are competent to master the material. Unfortunately, many students in psychology are unable to comprehend the mathematics and become lost in the symbols, definitions, and derivations. On the other hand, the simplicity of the "cookbook" approach is a definite advantage. Yet this approach may leave the student with little understanding of what he is doing. An appropriate teaching method should incorporate the simplicity of the "cookbook" approach while conveying the major aspects of the mathematical rationale behind the ANOVA procedures. The E(MS) concept can provide this combination.

In the experience of the first author in teaching ANOVA designs to upper-level undergraduates and to graduate students at a number of institutions, he has discovered that the topic of E(MS) is valuable to emphasize so as to facilitate the development of understanding for the students. If extensive coverage of E(MS) (both general and specific) and mathematical models are provided prior to discussing any specific ANOVA design, the student is capable of operating with greater understanding and is more alert to the attacking of novel designs. He finds that most students have great success and are quite excited about the independence and accuracy of their efforts. After these experiences, the student is ready to develop the mathematical model and E(MS) for any design with a minimum of assistance from the instructor.

One can use the general E(MS) case as an example of the teaching utility of E(MS) ideas. The general E(MS) case with suppressor coefficients [(1−x)/X] is extremely useful in order to understand the fixed, mixed, and random effects models. With a specific design, one proceeds from the general E(MS) case to the specific cases of fixed effects (fixed model—Model I), one random effect (mixed model), or two random effects (random effects model—Model II). These E(MS) aspects simplify the process for the instructor. Because of this simplification, the student can become more independent in the handling of novel ANOVA designs.

Other teachers of psychological statistics also have found E(MS) concepts of value.

CONCLUSIONS

The resurgence of ANOVA use following World War II probably would have occurred without extensive E(MS) employment. However, these ideas provided the researcher with aids that allowed a quick intuitive understanding of, and an easy interpretation of, the results of the analysis. Without this concept, the understanding of ANOVA results for many psychologists probably would have remained at the pre-E(MS) "cookbook" or "near-cookbook" level.

REFERENCES

ALEXANDER, H. W. (1947). The estimation of reliability when several trials are available. *Psychometrika*, **12**, 79-99.

ANDERSON, R. L., & BANCROFT, T. A. (1952). *Statistical theory in research*. New York: McGraw-Hill.

BURT, C. (1955). Test reliability by analysis of variance. *British Journal of Statistical Psychology*, **8**, 103-118.

CORNFIELD, J., & TUKEY, J. W. (1956). Average values of mean squares in factorials. *Annals of mathematical statistics*, **27**, 907-949.

CRUMP, S. L. (1946). The estimation of variance components in analysis of variance. *Biometrics Bulletin*, **2**, 7-11.

CRUMP, S. L. (1951). The present status of variance component analysis. *Biometrics*, **7**, 1-16.

DANFORD, M. B., HUGHES, H. M., & MCNEE, R. C. (1960). On the analysis of repeated-measurements experiments. *Biometrics*, **4**, 547-565.

DANIELS, H. E. (1939). The estimation of components of variance. *Royal Society Journal*, (Suppl. 6), 186-197.

DODD, D. H., & SCHULTZ, R. F., JR. (1973). Computational procedures for estimating magnitude of effect for some analysis of variance procedures. *Psychological Bulletin*, **79**, 391-395.

EDGINGTON, E. S. (1974). A new tabulation of statistical procedures used in APA journals. *American Psychologist*, **29**, 25-26.

EISENHART, C. (1947). The assumption underlying the analysis of variance. *Biometrics*, **3**, 1-21.

FEDERER, W. T. (1955). *Experimental design-theory and application*. New York: Macmillan.

FERGUSON, G. A. (1959). *Statistical analysis in psychology and education*. New York: McGraw-Hill.

FLEISS, J. L. (1969). Estimating the magnitude of experimental effects. *Psychological Bulletin*, **72**, 273-276.

GAITO, J. (1958a). The Bolles-Messick coefficient of utility. *Psychological Reports*, **4**, 595-598.

GAITO, J. (1958b). The single Latin square in psychological research. *Psychometrika*, **23**, 369-378.

GAITO, J. (1958c). Statistical dangers involved in counterbalancing. *Psychological Reports*, **4**, 463-468.

GAITO, J. (1959). Nonparametric methods in psychological research. *Psychological Reports*, **5**, 115-125.

GAITO, J. (1960). Expected mean squares in analysis of variance techniques. *Psychological Reports*, **7**, 3-10.

GAITO, J. (1961). Repeated measurements designs and counterbalancing. *Psychological Bulletin*, **58**, 46-54.

GREEN, B. F., JR., & TUKEY, J. (1960). Complex analyses of variance: General problems. *Psychometrika*, **25**, 127-152.

GREENWOOD, J. A. (1952, 1956). Analysis of variance and components of variance: Factorial experiments. Unpublished paper, USN Bureau of Aeronautics, Washington, DC.

GUILFORD, J. P. (1954). *Psychometric Methods*. New York: McGraw-Hill.

HAGGARD, E. A. (1958). *Intraclass correlation and the analysis of variance*. New York: Dryden.

HARTER, H. L., & LUM, M. D. (1955). *Partially hierarchial models in the analysis of variance* (Report 55-33). Wright Air Development Center.

HAYS, W. L. (1963). *Statistics for psychologists*. New York: Holt Rinehart & Winston.

HOYT, C. (1941). Test reliability obtained by analysis of variance. *Psychometrika*, **6**, 253-260.

KEMPTHORNE, O. (1952). *The design and analysis of experiments*. New York: Wiley.

KIRK, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.

LINDQUIST, E. F. (1953). *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin.

LOVIE, A. D. (1979). The analysis of variance in experimental psychology. *British Journal of Mathematical & Statistical Psychology*, **32**, 151-178.

MCNEMAR, O. (1955). *Psychological statistics*. New York: Wiley.

MOOD, A. M. (1950). *Introduction to the theory of statistics*. New York: McGraw-Hill.

PLACKETT, R. L. (1960). Models in the analysis of variance. *Journal of the Royal Statistical Society*, **22**, 195-217.

RUCCI, A. J., & TWENEY, R. D. (1980). Analysis of variance and the "Second Discipline" of scientific psychology: A historical account. *Psychological Bulletin*, **87**, 166-184.

SATTERTHWAITE, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2**, 110-114.

SCHEFFÉ, H. (1956a). Alternative models for the analysis of variance. *Annals of Mathematical Statistics*, **27**, 251-271.

SCHEFFÉ, H. (1956b). A mixed model for the analysis of variance. *Annals of Mathematical Statistics*, **27**, 23-36.

SCHEFFÉ, H. (1959). *Analysis of Variance*. New York: Wiley.

SCHULTZ, E. F., JR. (1955). Rules of thumb for determining expectations of mean squares in analysis of variance. *Biometrics*, **11**, 123-135.

SNEDECOR, G. W. (1946). *Statistical methods*. Ames, IA: Iowa State College Press.

STANLEY, J. C. (1956). Fixed, random, and mixed models in the analysis of variance as special cases of finite model III. *Psychological Reports*, **2**, 369.

TUKEY, J. W. (1949). Interaction in a row-by-column design (Memorandum Report No. 18). Princeton, NJ: Princeton University, Statistical Research Group.

VAUGHAN, G. M., & CORBALIS, M. C. (1969). Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin*, **72**, 204-213.

WILK, M. B., & KEMPTHORNE, O. (1955). Fixed, mixed, and random models. *Journal of the American Statistical Association*, **50**, 1144-1167.

WILK, M. B., & KEMPTHORNE, O. (1957). Non-additivities in a Latin square design. *Journal of American Statistical Association*, **52**, 218-236.

NOTES

1. Because of its widespread use, E(MS) is described in most advanced statistical books. Anyone not familiar with E(MS) should consult Kirk (1968).

2. A few papers and books have not been included. The ones included are the most frequently cited ones.

3. Copies of the 1956 revision can be obtained by writing to the first author of this paper.

(Manuscript received for publication June 21, 1985.)