

A NOTE ON THE USE OF MOOD'S LIKELIHOOD RATIO
TEST FOR ITEM ANALYSES INVOLVING 2×2
TABLES WITH SMALL SAMPLES

IRVING B. WEINER

UNIVERSITY OF MICHIGAN*

Mood's likelihood ratio test is generally considered an unreliable χ^2 approximation in 2×2 contingency tables containing expected cell frequencies less than five. Probability values were computed for 60 such tables as part of an item analysis for two 30-item alternate forms of a measure. The rank orders of the items, from best to worst differentiators, as determined separately by Mood's test and by Fisher's exact test correlated .97 for one form and .96 for the other.

Item analyses are often carried out not to determine the correlations of dichotomously scored items with a total test score, but rather to select a given number of items which best differentiates between high and low scorers. This type of analysis is facilitated by the use of 2×2 contingency tables which compare for each item the score (one or zero) of extreme scorers on the total test. The writer recently administered to groups of 25 and 22 subjects two 30-item alternate forms of a measure which he wanted to condense into a single 30-item form. Because of the restricted sampling, many of the expected cell frequencies in the resulting contingency tables were less than five. Since χ^2 approximations are generally not reliable with small cell frequencies [1], the probability values for the tables were computed directly by Fisher's exact test, and the 30 best items were identified accordingly. The probability values covered a wide range, from .008 to 1.00.

Probability values for the tables were also estimated by Mood's likelihood ratio test ([2], pp. 257-281), which is distributed approximately as χ^2 with $(r - 1)(s - 1)$ degrees of freedom for large samples. Mood states, however, that this large-sample approximation cannot be used without appreciable error when r and s both equal two. To evaluate this possibility with the 2×2 tables involved here, a comparison was made between the rank order of the items (from best to worst) for each of the two forms as determined by Mood's test and as indicated by the exact test. For one form the rank-order correlation (Spearman's rho) achieved between these two methods was .97 and for the other it was .96. Thus, although assumptions for its use were not met, Mood's test gave a very good indication of the *relative* probability values for the items. This may indicate that where ranking

*Now at University of Rochester School of Medicine.

is the goal, as in the item analysis described, the χ^2 approximation by Mood's likelihood ratio test is an adequate statistical tool even with small samples.

REFERENCES

- [1] Cochran, W. G. Some methods of strengthening the common χ^2 tests. *Biometrics*, 1954, 10, 417-451.
- [2] Mood, A. F. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1954.

Manuscript received 6/6/59