The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 35–45

REJECTING OUTLIERS BY MAXIMUM NORMED RESIDUAL[1]

By Wilhelmine Stefansky

University of California, Berkeley

1. Introduction. The maximum normed residual (MNR) has been proposed as a test statistic in connection with the problem of rejecting outliers. An outlying observation is one that does not fit in with the pattern of the remaining observations. Outliers may be mistakes, or else accurate but unexpected observations which could shed new light on the phenomenon under study. On the other hand, it is possible that an outlier is simply a manifestation of the inherent variability of the data. It is of interest, therefore, to test whether or not a given outlier comes from a population different from the one hypothesized.

A thorough discussion of the special case of detecting outliers in a single sample from a normal population is given by F. E. Grubbs in [5]. Several tests are discussed, among them tests based on the MNR, and tables of critical values are included. For unreplicated factorial designs C. Daniel [3] proposed a statistic equivalent to the MNR. T. S. Ferguson [4] has proved that the maximum Studentized residual possesses the optimum property of being admissible within all invariant procedures. For designs with residuals having a common variance, the MNR is equivalent to the maximum Studentized residual, hence also has this optimum property. Nevertheless, except for the case of a single sample from a normal population, critical values of the MNR are not available.

In 1936 E. S. Pearson and C. Chandra Sekar [7] noticed that for the case of a single sample from a normal distribution, critical values for statistics equivalent to the MNR can be calculated quite easily from tables of the t-distribution, provided the level α is not too large. In order to determine if α is not too large, it is necessary to know M_2, the largest value which the second largest among the absolute values of the normed residuals can take on.

In Section 4 the results of E. S. Pearson and C. Chandra Sekar [7] are extended to designs with the property that the residuals have a common variance. Applications are given in Section 5.

The main problem in extending the results of [7] is the calculation of M_2. In Section 3 a more general problem is dealt with, namely the calculation of M_k for arbitrary designs, where M_k denotes the largest value which the kth largest among the absolute values of the residuals can have. A method for calculating M_k is developed and this method yields an explicit expression for M_2 for designs having residuals with a common variance.

Designs with the property that the residuals have a common variance include all ordinary factorial designs, where the different levels of each factor are replicated

Received April 30, 1970.

[1] This paper is based on the author's doctoral dissertation and was prepared with the partial support of a National Science Foundation Grant GP 15283.