The Annals of Mathematical Statistics
1972, Vol. 43, No. 3, 823-831

LOCALLY MOST POWERFUL RANK TESTS FOR THE TWO-SAMPLE PROBLEM WITH CENSORED DATA[1]

BY RICHARD A. JOHNSON AND K. G. MEHROTRA

University of Wisconsin

1. Introduction. Suppose that two samples of size m and n, respectively, are placed on life test. Here the observations are the time to failure, and they are naturally ordered. In some statistical applications it may be necessary and can be desirable, not to observe all of the failure times but to reach a decision on the basis of a part of a sample.

Incomplete data situations arise naturally in many fields. For example, they occur in experiments where the measuring device may fail to record very large or very small values. In medical experiments they can occur when it is not possible to wait until all the observations are available. Further, in life tests with costly electronic equipment, units which have not failed can be used in the future.

In this paper, the procedure of Rao, Savage and Sobel (1960) has been further studied for the case when only the smallest r failure times, in the combined sample, are observed. From this procedure, we obtain l.m.p.r. tests for both location and scale alternatives. The results can be extended directly to include left censoring when the first r^* observations are not available or when both first r_1 and last r_2 observations are not available.

It is also shown how to obtain the asymptotic distribution of the test statistics from the extension of the Chernoff-Savage theorem by Pyke and Shorack (1968) or by the method of Dupač and Hájek (1969).

We use the following notation throughout. Let $X_1, \cdots, X_m$ be a random sample of size m from a population with cumulative distribution function (cdf) $F(x)$ and $Y_1, \cdots, Y_n$ a random sample of size n from a population with cdf $G(x)$. The ordered observations in the combined sample of size $N = m + n$ are denoted by $W_1 \leqq W_2 \leqq \cdots \leqq W_N$. Only the first r (fixed) order statistics are observed and these are identified as coming from the first or the second population by the vector $\mathbf{Z} = (Z_1, \cdots, Z_r)$ where $Z_i = 1$ if the ith ordered observation, W_i, is from $F(x)$ and $Z_i = 0$ if W_i is from $G(x)$; $i = 1, 2, \cdots, r$. Let us denote the density function (pdf) corresponding to $F(x)$ by $f(x)$ and that corresponding to $G(x)$ by $g(x)$. Further, set,

$$m_r = \sum_{i=1}^{r} Z_i ,$$

so that m_r denotes the random number of observations from the first sample among the r observations. Set $n_r = r - m_r$. Then Rao, Savage and Sobel (1960) have proved that

[1] This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-69-1789.