Two Nonparametric Tests for Testing the Randomness of Samples Drawn from Finite Populations

MITCHELL O. LOCKS, University of Oklahoma, Norman

ABSTRACT

Assume all of the elements of a given population with N items can be ranked according to a certain numerical characteristic with each element being assigned a rank from 1 to N. If each item is assigned its rank in the population, the mean rank for the population is $\frac{N+1}{2}$ and the variance of the population ranks (σ^2_r) is $\frac{N^2-1}{12} - C$.[1] If a sample of K items is randomly selected from this population, the square of the standard error of the mean rank ($\sigma^2_{\bar{r}}$) for that sample is:

$$\sigma^2_{\bar{r}} = \frac{\sigma^2_r}{K}\left\{1-\frac{K}{N}\right\} = \frac{\frac{N^2-1}{12}-C}{K}\left\{1-\frac{K}{N}\right\}$$

In addition, even though the ranks themselves for a rectangular distribution, the mean ranks for random samples of size K selected from this population form a distribution which is approximately normal.

THE EFFECT OF TIES ON THE COMPUTATION OF THE VARIANCE OF RANKS.

When two or more items in the population have identical values of the numerical characteristic according to which the population is being ranked, both of these tied items are assigned the same rank. The usual practice (which will be followed in this paper) is to compute this rank as the average of the ranks the items would have had had there been no ties.

This procedure does not change the total of the population ranks and therefore does not affect the average population rank of $\frac{N+1}{2}$. However, it does reduce the variance of the population ranks to an amount somewhat less than $\frac{N^2-1}{12}$.

The following formula gives an exact calculation of the difference between the sum of squares without ties, and the sum of squares with a tie of T items:

$$\frac{T^3-T}{12}.$$

Since the sum of squares is always smaller by this amount than it would be without ties because of a tie of T items, the variance of ranks

[1] The correction factor, C, in this formula is a function of the size and the number of ties in the population ranks. The computation of C is discussed below.

for a population of N items is smaller by $\frac{T^3 - T}{12N}$ because of the tie. However each separate tie in the population ranks itself reduces the variance of ranks by this amount. Therefore, the correction factor C in the computation of the variance of ranks is the sum of these terms, or

$$C = \frac{\Sigma\,(T^3 - T)}{12N} = \frac{1}{12N}\Sigma\,(T^3 - T)$$

TEST OF THE MEAN OF THE POPULATION RANKS OF THE ITEMS IN THE SAMPLE.

The objective of this test is to determine whether the arithmetic mean of the population ranks of the items in the sample differs significantly from the arithmetic mean of the population ranks of $\frac{N+1}{2}$. For this purpose, it is necessary only to calculate how many standard errors the mean rank of the sample differs from $\frac{N+1}{2}$. In order to do so, we calculate the unit standard deviate of

$$\frac{\bar{r} - \frac{N+1}{2}}{\sigma_{\bar{r}}}$$

where $\bar{r}$ is the mean of the population ranks of the items in the sample. The probability of a difference that large between $\bar{r}$ and $\frac{N+1}{2}$ may be calculated with the aid of a table of areas under the normal curve.

TEST OF THE VARIATION OF THE POPULATION RANKS OF THE ITEMS IN THE SAMPLE.

The other nonparametric test that may be used from these ranked data is one to determine whether the variation of the population ranks of the items in the sample is comparable to the variation of the ranks for all the items in the population. Since the exact variance of the population ranks for all of the items in the population is known, the Chi-square distribution may be used for this purpose.

Since the mean ranks for samples of size K (K greater than 1) are approximately normally distributed, the Chi-square distribution defined below with K − 1 degrees of freedom may be used:

$$\text{Chi-square} = \frac{\dfrac{(K-1) \times \Sigma\,(r - \bar{r})^2}{(K-1)^2}\left\{1 - \dfrac{K}{N}\right\}}{\left\{\dfrac{N^2-1}{12} - C\right\}\left\{1 - \dfrac{K}{N}\right\}} = \frac{K}{K-1} \times \frac{\Sigma\,(r - \bar{r})^2}{\dfrac{N^2-1}{12} - C}.$$

A critical value of Chi-square may be established with the aid of a table of Chi-square for K − 1 degrees of freedom.