SIGNIFICANCE LEVELS FOR A k-SAMPLE SLIPPAGE TEST[1]

FREDERICK MOSTELLER AND JOHN W. TUKEY

Harvard University and Princeton University

1. Summary. Mosteller has recently [1, 1948] proposed a k-sample slippage test and has given percentage points for selected n, k and r for the case of k equal samples of size n. When the samples are of unequal size, exact significance levels can be calculated very quickly from

$$P_r = \frac{\sum n_i^{(r)}}{N^{(r)}} \text{ where } x^{(r)} = x(x - 1) \cdots (x - r + 1),$$

by the method explained in section 3 below.

The significance values for k equal samples of $n \geq 10$ are very well approximated by

$$P_r = \frac{1}{k^{r-1}} e^{-r(r-1)(k-1)/2N}$$

where $N = kn$.

A convenient rough approximation for unequal samples may be given in terms of k^*, an "effective" number of samples, which is given by

$$k^* = \frac{(\sum_i n_i)^2}{\sum n_i^2},$$

the one-sided significance level will then be approximately given by

$$P_r = (k^*)^{-(r-1)}.$$

This approximation can be easily applied with the aid of Table 1. Thus, for example, with four samples of sizes 7, 5, 5, 2, we have

$$k^* = \frac{(7 + 5 + 5 + 2)^2}{49 + 25 + 25 + 4} = \frac{361}{103} = 3.50,$$

whence from the table $r = 3$ lies at a one-sided level approximately between 5% and 10%, $r = 4$ approximately between 1% and 2.5%, $r = 5$ between 0.5% and 1%, $r = 6$ near 0.2%, and so on. Direct calculation yields 5.7%, 1.2%, 0.2% and 0.03%. The approximation is, in this example, quite satisfactory for moderate significance levels and conservative for more extreme significance levels.

2. Derivation. The statistic considered by Mosteller is the number of cases in one sample greater than all cases in all the $k - 1$ other samples. We derive its distribution briefly.

Since the statistic depends only on the order of the $n_1 + n_2 + \cdots + n_k = N$

[1] Prepared in connection with research sponsored by the Office of Naval Research.

values, we can consider the actual values taken on to be fixed, and consider their allotment to the various samples. Assuming all of them to come from a single continuous distribution, we may consider these fixed values to be all distinct, and any way of allotting them to labelled places in the various samples as equally likely.

Consider the r largest values. They can all be allotted to places in the i-th sample in $n_i^{(r)} = n_i(n_i - 1) \cdots (n_i - r + 1)$ ways, and to arbitrary places in $N^{(r)}$ ways. Thus they will be allotted to some single sample in the fraction

$$P_r = \frac{\sum_i n_i^{(r)}}{N^{(r)}}$$

of all cases. This is clearly the probability that Mosteller's statistic is r or more.

TABLE 1

Approximate critical values of k^* for various levels of significance

One-sided level	10%	5%	2.5%	1%	0.5%	0.2%	0.1%
Two-sided level	20%	10%	5%	2%	1%	0.4%	0.2%
$r = 2$	10.0	20.0	40.0	100.0	200.0	500.0	1000.0
$r = 3$	3.2	4.5	6.3	10.0	14.1	22.4	31.6
$r = 4$		2.7	3.4	4.6	5.8	7.9	23.0
$r = 5$			2.5	3.2	3.8	4.7	5.6
$r = 6$				2.5	2.9	3.5	4.0
$r = 7$						2.8	3.2
$r = 8$							2.6

3. Unequal samples—an exact computation. Our practical problem is to compute P_r for small values of r and a fixed set of n_i. If we recognize the numerators as the unnormalized factorial moments of the distribution of sample sizes, we see that the computation goes smoothly according to the scheme shown in Table 2 (where the columns of multipliers $n - 1, n - 2, n - 3$, etc. may be partially covered for convenience during the computation.): For example: $132 = 11(12)$, $1320 = 10(132)$, $\cdots 42 = 6(7)$. The numbers in the last line of Table 2 give successively the percentages $100\, P_1$, $100\, P_2$, $\cdots$. Of course $P_1 = 1$ because some sample must have the largest value. It is clear that exact computation for any reasonable set of n_i is quite easy.

4. Equal samples—an approximation. In the case of k equal samples, we have

$$P_r = \frac{kn^{(r)}}{N^{(r)}}.$$

Let us try to approximate to $n^{(r)}$ by expansion in powers. We have

$n^{(r)} = n(n-1)\cdots(n-r+1) = n^r(1-1/n)(1-2/n)\cdots(1-(r-1)/n),$
so that

$$\log n^{(r)} = r\log n + \sum_{x=1}^{r-1}\log(1-x/n)$$
$$= r\log n - \sum_{x=1}^{r-1}(x/n + x^2/2n^2 + x^3/3n^3\cdots)$$
$$= r\log n - r(r-1)/2n - r(r-1)(2r-1)/12n^2 + 0(n^{-3}),$$

TABLE 2

Sample Computation

for $\{n_i\}$ = (12, 11, 11, 11, 10, 10, 10, 10, 9, 9, 7, 4)

$n-2$	$n-1$	n	f	nf	$n^{(2)}f$	$n^{(3)}f$
10	11	12	1	12	132	1320
9	10	11	3	33	330	2970
8	9	10	4	40	360	2880
7	8	9	2	18	144	1008
5	6	7	1	7	42	210
2	3	4	1	4	12	24
$N-2$	$N-1$	N	Sums	114	1020	8412
112	113	114	$N^{(r)}$	114	12882	1442784
			P_r	100%	7.9%	0.58%

and hence

$$\log P_r = \log k + \log n^{(r)} - \log N^{(r)} = \log k + \log n^{(r)} - \log(nk)^{(r)}$$
$$= \log k + r\log n - r(r-1)/2n - r(r-1)(2r-1)/12n^2$$
$$- r\log nk + r(r-1)/2nk + r(r-1)(2r-1)/12n^2k^2 + 0(n^{-3})$$
$$= -(r-1)\log k - \frac{r(r-1)}{2n}\left(1 - \frac{1}{k} + \frac{2r-1}{6n} - \frac{2r-1}{6nk^2} + 0(n^{-2})\right).$$

We get the following three approximations:

(1) $$P_r \doteq \frac{1}{k^{r-1}};$$

and noting that $\dfrac{1-1/k}{n} = \dfrac{k-1}{kn} = \dfrac{k-1}{N}$,

(2) $$P_r \doteq k^{-(r-1)}\, e^{(-r(r-1)(k-1))/2N} = \frac{1}{[ke^{(r(k-1))/2N}]^{r-1}};$$

and finally

$$P_r \doteq k^{-(r-1)} e^{(-r(r-1)(k-1)/2N)(1+(2r-1)/6n)}. \tag{3}$$

5. Comparison of results. The results obtained with various equal sample approximations will be compared with the exact values for several cases. The effective number of samples, k^*, used with (1), (2), and (3), is computed from

$$k^* = \frac{(\sum n_i)^2}{\sum n_i^2},$$

a formula which is often an easy and effective way to allow for different sizes of samples.

TABLE 3

Comparison of Approximations

Sizes of Samples	N	k	r	P_r in exact	(1)	(2)	(3)	(4)
10, 10, 10, 10	40	4.00	2	23.08	25.00	23.19	23.13	≤25.00
			3	4.85	6.25	4.99	4.80	≤6.25
7, 5, 5, 2	19	3.50+	2	24.56	28.53	25.01	24.82	≤28.53
			3	5.67	8.14	5.48	5.18	≤8.76
12, 11, 11, 11								
10, 10, 10, 10	114	11.46	2	7.92	8.73	7.96	7.96	≤8.73
9, 9, 7, 4			3	0.58	0.76	0.58	0.56	≤0.78

A fourth approximation, which always gives a conservative estimate of the significance of the result is obtained by replacing $n^{(r)}$ by n^r throughout, this gives

$$P_r = \frac{\sum n_i^r}{N^r}, \tag{4}$$

which is equivalent to approximation (1) when the samples are of equal size, or when $r = 2$.

The results are shown in Table 3.

Thus it seems clear that either (1) or (4) are good enough for rough work. The choice will depend on which formula one prefers to remember. The amount of work is about the same for either method. When something better is required the exact method of section 3 seems appropriate. Indeed some may prefer it to any approximation.

REFERENCE

[1] FREDERICK MOSTELLER, "A k-sample slippage test for an extreme population," *Annals of Math. Stat.*, Vol. 19 (1948), p. 58–65.