

GENERALIZED h -STATISTICS AND OTHER SYMMETRIC FUNCTIONS¹

BY D. S. TRACY AND B. C. GUPTA²

University of Windsor

Dwyer's (1937) h -statistic is extended to the generalized h -statistic $h_{p_1 \dots p_u}$ such that $E(h_{p_1 \dots p_u}) = \mu_{p_1} \dots \mu_{p_u}$, similar to the extension of Fisher's k -statistic to the generalized k -statistic $k_{p_1 \dots p_u}$ requiring $E(k_{p_1 \dots p_u}) = \kappa_{p_1} \dots \kappa_{p_u}$. The h -statistics follow simpler multiplication rules than for k -statistics and involve smaller coefficients. Generalized h -statistics are studied in terms of symmetric means, unrestricted sums, and ordered partitions, and their relationships with generalized k -statistics are established. The statistics are useful in obtaining approximate forms for sampling distributions when parent population is not completely known.

1. Introduction. Fisher (1928) introduced the k -statistic k_p as unbiased estimate of population cumulant κ_p . When the population is finite of size N , $E_N(k_p) = K_p$, where E_N denotes expectation over the finite population and K_p is the K -parameter of the finite population (being the same function of the elements in the population as k_p is of the sample elements). The k -statistics were later extended to generalized k -statistics in [11] or polykays in [10], which serve as unbiased estimates of products of cumulants. Methods expressing their powers and products as linear combinations of the same were developed in [4], [11] in order to obtain their sampling moments and product moments. These help in finding approximate forms for sampling distributions when the parent population is not completely specified ([8] Chapter 12).

Kendall (1942) surmised that the k -statistics were the only seminvariants which follow simple multiplication rules allowing the computation of the numerical coefficient and the pattern function involved. In this paper, we consider the h -statistic h_p , proposed by Dwyer (1937, page 26) as the unbiased estimate of the population central moment μ_p , $p > 1$, with $E(h_1) = \mu_1'$. One has $E_N(h_p) = H_p$, the H -parameter of the finite population. We extend the h -statistic to generalized h -statistics, which provide unbiased estimates of products of moments. Methods for expressing their powers and products as linear combinations of the same can be similarly developed, and, in some ways, are easier than for the case of k -statistics. Approximate forms for sampling distributions may then be obtained. The h -statistics, which are also seminvariant,

Received September 1972; revised September 1973.

¹ Research supported by National Research Council of Canada, Grant A-3111.

² Now at Instituto de Matemática, Universidade Federal do Rio de Janeiro.

AMS 1970 subject classifications. 62.10, 62.20.

Key words and phrases. Moment, cumulant, seminvariant, h -statistic, k -statistic, generalized h -statistic, polykay, generalized k -statistic, unbiased estimation, finite population, combinatorial coefficient, symmetric mean, power sum, augmented monomial symmetric function, power product sum, unrestricted sum, partition, ordered partition.

follow even simpler rules and involve fewer terms and smaller coefficients than the corresponding k -statistics.

2. h -Statistic. One may express the central moment μ_p in terms of ordinary moments as

$$(2.1) \quad \mu_p = \sum_{r=0}^p (-1)^r \binom{p}{r} \mu'_{p-r} \mu_1'^r.$$

The sample symmetric mean $\langle p_1 \cdots p_u \rangle$ ([10]), which is $[p_1 \cdots p_u]/n^{(u)}$, where $[p_1 \cdots p_u]$ is the power product sum or augmented monomial symmetric function ([2]), has the property that

$$E\langle p_1 \cdots p_u \rangle = \mu'_{p_1} \cdots \mu'_{p_u},$$

and

$$E_N\langle p_1 \cdots p_u \rangle = \langle p_1 \cdots p_u \rangle_N$$

where $\langle p_1 \cdots p_u \rangle_N$ denotes the finite population symmetric mean. Thus the statistic h_p is expressible as

$$(2.2) \quad h_p = \sum_{r=0}^p (-1)^r \binom{p}{r} \langle p-r, 1^r \rangle$$

and one has $E_N(h_p) = H_p$, $E(h_p) = \mu_p$ (with $E(h_1) = \mu_1'$).

Associated with partition $P = p_1^{\pi_1} \cdots p_s^{\pi_s}$, $p_1 > \cdots > p_s$, of integer p , we use the combinatorial coefficient $C(P) = p!/(p_1!)^{\pi_1} \cdots (p_s!)^{\pi_s} \pi_1! \cdots \pi_s!$ ([4]), and the number of parts $\lambda = \sum \pi_i$. We can then write (2.1) and (2.2) as

$$(2.3) \quad \mu_p = \sum (-1)^{\lambda-1} \beta C(P) \mu'_{p_1}{}^{\pi_1} \cdots \mu'_{p_s}{}^{\pi_s}$$

$$(2.4) \quad h_p = \sum (-1)^{\lambda-1} \beta C(P) \langle p_1^{\pi_1} \cdots p_s^{\pi_s} \rangle$$

where the summation extends over all partitions P of integer p and

$$\begin{aligned} \beta &= p-1, & \text{if } p_1 &= 1 \\ &= 1, & \text{if } p_1 > 1, \quad p_2 &= 1 \\ &= 0, & \text{otherwise.} \end{aligned}$$

It may be recalled that

$$\begin{aligned} \kappa_p &= \sum (-1)^{\lambda-1} (\lambda-1)! C(P) \mu'_{p_1}{}^{\pi_1} \cdots \mu'_{p_s}{}^{\pi_s}, \\ k_p &= \sum (-1)^{\lambda-1} (\lambda-1)! C(P) \langle p_1^{\pi_1} \cdots p_s^{\pi_s} \rangle. \end{aligned}$$

For example,

$$\begin{aligned} h_1 &= k_1 = \langle 1 \rangle \\ h_2 &= k_2 = \langle 2 \rangle - \langle 11 \rangle \\ h_3 &= k_3 = \langle 3 \rangle - 3\langle 21 \rangle + 2\langle 111 \rangle \\ h_4 &= \langle 4 \rangle - 4\langle 31 \rangle + 6\langle 211 \rangle - 3\langle 1111 \rangle \\ k_4 &= \langle 4 \rangle - 4\langle 31 \rangle - 3\langle 22 \rangle + 12\langle 211 \rangle - 6\langle 1111 \rangle. \end{aligned}$$

Expressions for h_p and k_p are the same linear combinations of symmetric means as μ_p and κ_p are of the products of corresponding ordinary moments.

Given the expression of k_p in terms of symmetric means, that for h_p may be obtained by omitting terms with more than one non-unit part and dividing the coefficients of the rest by $(\lambda - 1)!$, except the last one (coefficient of $\langle 1^p \rangle$), which is to be divided by $(p - 2)!$.

The statistic h_p above is uniquely defined, since if $h_{p'}$ is another statistic for which $E(h_{p'}) = \mu_p$, then $E(h_p - h_{p'}) = 0$. By an argument similar to that for k -statistics ([8], page 278), this would imply a relationship between the moments, which, in general, is not possible. Also $h_p, p > 1$, is seminvariant, i.e., $h_p(x) = h_p(x - c), p > 1$.

3. Generalized h -statistic. It follows from (2.1) that

$$(3.1) \quad \mu_{p_1} \cdots \mu_{p_v} = \sum (-1)^{\sum r_i} \binom{p_1}{r_1} \cdots \binom{p_v}{r_v} \mu'_{p_1-r_1} \cdots \mu'_{p_v-r_v} \mu_1^{\sum r_i}.$$

If the generalized h -statistic $h_{p_1 \cdots p_v}$ is defined such that

$$E(h_{p_1 \cdots p_v}) = \mu_{p_1} \cdots \mu_{p_v},^3$$

then

$$(3.2) \quad h_{p_1 \cdots p_v} = \sum (-1)^{\sum r_i} \binom{p_1}{r_1} \cdots \binom{p_v}{r_v} \langle p_1 - r_1, \cdots, p_v - r_v, 1^{\sum r_i} \rangle$$

and

$$E_N(h_{p_1 \cdots p_v}) = H_{p_1 \cdots p_v},$$

where

$$(3.3) \quad H_{p_1 \cdots p_v} = \sum (-1)^{\sum r_i} \binom{p_1}{r_1} \cdots \binom{p_v}{r_v} \langle p_1 - r_1, \cdots, p_v - r_v, 1^{\sum r_i} \rangle_N.$$

Alternatively, one can use Tukey's \circ -multiplication ([9]),

$$(3.4) \quad \langle ab \rangle \circ \langle cd \rangle = \langle abcd \rangle$$

for symmetric means and define, using (2.4),

$$(3.5) \quad h_{p_1 \cdots p_v} = h_{p_1} \circ \cdots \circ h_{p_v} = \sum (-1)^{\sum (\lambda_i - 1)} \prod_{i=1}^v \beta_i C(P_I) \langle P_I \rangle$$

where P_I is the specified partition of $P = p_1 \cdots p_v$ obtained by partitioning one or more of the p_i 's, and the summation is over all partitions P_I .

Like generalized k -statistics, we may also define generalized h -statistics through ordered partitions ([1]). The expression for ordinary moment $\mu_{p'}$ in terms of central moments μ_m and the mean μ_1' exhibits only those partitions of p which have at most one non-unit part, i.e.,

$$(3.6) \quad \mu_{p'} = \sum_{j=0}^p \binom{p}{j} \mu_{p-j} \mu_1'^j$$

unlike the expression for $\mu_{p'}$ in terms of cumulants

$$(3.7) \quad \mu_{p'} = \sum_{s=0}^p \sum_P C(P) \kappa_{p_1}^{\pi_1} \cdots \kappa_{p_s}^{\pi_s}$$

which exhibits all possible partitions $P = p_1^{\pi_1} \cdots p_s^{\pi_s}$ of $p = \sum_1^s p_i \pi_i$. For multipartite number $p = 11 \cdots 1$, the combinatorial coefficients $\binom{p}{j}$ and $C(P)$ in the above expressions become unity, and whereas all multipartitions of p

³ Should some p_i be 1, the corresponding μ_{p_i} is μ_1' .

appear in the expression of μ_p' in terms of multivariate cumulants, only those with at most one non-unit part appear in the expression of μ_p' in terms of central moments. Thus

$$(3.8) \quad \mu'_{1111} = \mu_{1111} + (\mu_{1110} \mu'_{0001} + \dots) + (\mu_{1100} \mu'_{0010} \mu'_{0001} + \dots) + \mu'_{1000} \mu'_{0100} \mu'_{0010} \mu'_{0001},$$

and the multipartitions $\begin{smallmatrix} 1100 \\ 0011, \end{smallmatrix} \begin{smallmatrix} 1010 \\ 0101, \end{smallmatrix} \begin{smallmatrix} 1001 \\ 0110 \end{smallmatrix}$ do not show up here, while they do when μ'_{1111} is expressed in terms of κ 's ([8], page 319).

Taking unbiased estimates, the expression of symmetric means in terms of h -statistics ignores partitions with more than one non-unit part, whereas symmetric means expressed in terms of k -statistics exhibit all possible partitions. Using the notation of ordered partitions ([1]), and $\langle \rangle$, $()$ to denote symmetric means and h -statistics,

$$\langle 1111 \rangle = (1111) + (1112) + (1121) + (1211) + (2111) + (1123) + (1213) + (1231) + (2113) + (2131) + (2311) + (1234),$$

whereas the expression in terms of k -statistics would also exhibit 1122, 1212, 1221 ([1]). Using Tukey's \circ -multiplication in [10] for symmetric means and h -statistics, as in (3.4) and (3.5), we see that for ordered partition α of weight m ,

$$(3.9) \quad \langle \alpha \rangle = \sum (\beta_\alpha)$$

where summation is over all ordered subpartitions β_α of α which do not have more non-unit parts than α , and (β_α) denotes the generalized h -statistic corresponding to the ordered subpartition β_α . (In the case of generalized k -statistics, the summation is over all ordered subpartitions.)

We now proceed to consider vectors of symmetric means, power sums, generalized h -statistics, ordered partitions, etc., of a given weight, with a view to establishing some general relationships between them.

If α denotes the vector of all ordered partitions of weight m , with components α^i arranged in descending order, and $\langle \alpha \rangle$ and (α) denote the vectors of symmetric means and of the generalized h -statistics of m th degree, then

$$(3.10) \quad \langle \alpha \rangle = \Delta(\alpha)$$

where Δ is a nonsingular upper triangular matrix with elements

$$\begin{aligned} \delta_{ij} &= 1, & \text{if } \alpha^j \text{ does not have more non-unit parts than} \\ & & \alpha^i \text{ and } \alpha^j \leq \alpha^i \\ &= 0, & \text{otherwise.} \end{aligned}$$

Thus

$$(3.11) \quad (\alpha) = \Delta^{-1} \langle \alpha \rangle$$

expresses generalized h -statistics in terms of symmetric means. The expressions are simpler than the corresponding ones for generalized k -statistics.

If $\{\alpha\}$ denotes power sum ([3]) or unrestricted sum ([1]) and $[\alpha]$ the augmented monomial symmetric function, (for $\alpha = 1123$, $\{\alpha\} = \sum_{i,j,k}^n x_i^2 x_j x_k$, $[\alpha] = \sum_{i \neq j \neq k}^n x_i^2 x_j x_k$), then ([1])

$$(3.12) \quad [\alpha] = \Lambda' \{\alpha\}$$

where Λ is a nonsingular upper triangular matrix with elements

$$(3.13) \quad \begin{aligned} \lambda_{ij} &= 1, & \text{if } \alpha^j \leq \alpha^i \\ &= 0, & \text{otherwise.} \end{aligned}$$

Also in [1], $\langle \alpha \rangle = N^{-1}[\alpha]$, where N is a diagonal matrix with $n_{ii} = n^{\phi(\alpha^i)}$, $\phi(\alpha^i)$ being the number of parts in α^i . Hence, from (3.11),

$$(3.14) \quad (\alpha) = \Delta^{-1} N^{-1} \Lambda' \{\alpha\}$$

which expresses generalized h -statistics in terms of unrestricted sums.

4. Generalized h -statistics and generalized k -statistics. If $(\alpha)^*$ denotes the vector of generalized k -statistics of m th degree, then

$$(4.1) \quad \langle \alpha \rangle = \Lambda(\alpha)^*$$

with Λ as defined in (3.13). Then, comparing (3.10) and (4.1),

$$\Delta(\alpha) = \Lambda(\alpha)^*$$

and

$$(4.2) \quad (\alpha) = \Delta^{-1} \Lambda(\alpha)^* ,$$

$$(4.3) \quad (\alpha)^* = \Lambda^{-1} \Delta(\alpha)$$

express generalized h -statistics and generalized k -statistics in terms of each other.

A table expressing these relationships for weight 12 is provided. The diagonal terms are all unity 1, and a generalized h -statistic is obtained in terms of generalized k -statistics by reading across the appropriate row up to 1, whereas a generalized k -statistic is obtained in terms of generalized h -statistics by reading vertically downwards the corresponding column up to 1. Thus,

$$\begin{aligned} h_{732} &= k_{732} + 21k_{532^2} + 35k_{43^22} + 105k_{3^22^3} , \\ k_{732} &= h_{732} - 21k_{532^2} - 35h_{43^22} + 210h_{3^22^3} . \end{aligned}$$

Results for smaller weights can be obtained by dropping the appropriate subscript. Thus,

$$h_{73} = k_{73} + 21k_{532} + 35k_{43^2} + 105k_{3^22^2} .$$

5. Conclusion. Products of generalized h -statistics are obtained as for generalized k -statistics by using ordered partitions ([6]) or by developing a combinatorial method, paralleling that in [4], [9]. These are then useful in obtaining their sampling moments. Work with generalized h -statistics, in many ways, is found to be simpler than with generalized k -statistics.

TABLE 1. Continued

	k_{632}	k_{642}	k_{651}	k_{62}	k_{7221}	k_{732}	k_{741}	k_{75}	k_{822}	k_{831}	k_{84}	k_{921}	k_{93}	$k_{10,2}$	$k_{11,1}$	k_{12}
h_{25}	.	-90	.	900	-630	.	1890	.	.	22680	.	-1247400
h_{32^2+1}	.	30	-300	-600	.	.	-630	-2100	560	-630	-630	-7560	.	-37800	302400	3326400
h_{3^2+3}	30	30	100	100	.	210	.	-2100	560	560	-1680	560	-7560	-92400	-92400	1247400
h_{3321}	.	-10	.	100	420	.	-1890	.	560	.	.	-92400
h_{34}	.	75	.	-900	-1890	-92400
h_{42^2+1}	.	.	150	-900	-35	.	315	.	420	420	-1890	2520	.	-20900	.	1247400
h_{43^2+1}	.	-15	-10	300	-35	-35	350	350	-35	.	560	2520	2520	4200	.	-831600
h_{432}	.	-15	-15	225	-35	.	525	.	.	3150	.	-311850
h_{4^2+31}	-35	-35	.	.	.	11550	.
h_{43}	-35	11550
h_{52^2+1}	.	.	30	.	-21	.	63	420	-56	.	168	756	.	.	-41580	-498960
h_{522}	.	.	-10	-10	.	-21	-21	420	-56	-56	168	756	756	5040	.	.
h_{53^2+1}	.	.	-15	-15	.	.	-21	.	.	-56	.	-126	.	.	9240	.
h_{5421}	-35	.	.	-56	-126	.	.	13860	.
h_{543}	-21	.	.	-56	-126	.	.	.	55440
h_{552}	-21	.	.	-56	-126	-126	.	.	16632
h_{552}	-21	.	.	-56	-126	-126	.	.	16632
h_{562}	.	-3	.	60	.	.	.	-21	-28	.	84	-84	.	.	.	-83160
h_{62^2}	.	.	-4	-28	-28	.	-84	.	.	9240	.
h_{6321}	1	.	.	-20	-84	.	.	18480
h_{642}	.	1	.	-30	-28	.	-84	-210	.	27720
h_{651}	.	.	1	-462	-426
h_{62}	20	30	.	1	1980	15840
h_{72^2+1}	.	.	.	1	1	1	-3	-10	.	.	.	-36
h_{732}	.	.	.	1	3	10	1	-36	-120	.	.
h_{741}	.	.	.	1	3	10	1	-330	.
h_{75}	.	.	.	1	3	10	1	-792
h_{82^2}	.	.	.	1	3	10	1	-45	.	2970
h_{831}	.	.	.	1	3	10	1	-165	-495
h_{84}	.	28	3	1	1	1
h_{921}	.	.	.	36	36	-55	.
h_{93}	84	.	.	.	36	120	330	1	.	.	-220
$h_{10,2}$.	210	.	.	990	7920	330	.	45	1	.	-66
$h_{11,1}$.	.	462	792	1485	165	495	55	.	.	1	.
h_{12}	9240	13860	.	462	.	7920	330	792	1485	165	495	55	220	66	.	1

REFERENCES

- [1] CARNEY, EDWARD J. (1968). Relationship of generalized polykays to unrestricted sums for balanced complete finite populations. *Ann. Math. Statist.* **39** 643-656.
- [2] DAVID, F. N. and KENDALL, M. G. (1949). Tables of symmetric function-I. *Biometrika* **36** 431-449.
- [3] DWYER, PAUL S. (1937). Moments of any rational integral isobaric sample moment function. *Ann. Math. Statist.* **8** 21-65.
- [4] DWYER, P. S. and TRACY, D. S. (1964). A combinatorial method for products of two polykays with some general formulae. *Ann. Math. Statist.* **35** 1174-1185.
- [5] FISHER, R. A. (1928). Moments and product moments of sampling distributions. *Proc. London Math. Soc.* (2) **30** 199-238.
- [6] GUPTA, B. C. (1972). Ordered partitions and unbiased estimates of central moments. *Symmetric Functions in Statistics—Proceedings of a Symposium*, (D. S. Tracy, ed.). Windsor. 131-141.
- [7] KENDALL, M. G. (1942). On seminvariant statistics. *Ann. Eugenics* **11** 300-305.
- [8] KENDALL, MAURICE G. and STUART, ALAN (1963). *The Advanced Theory of Statistics* **1** (2nd ed.). Hafner, New York.
- [9] TRACY, DERRICK S. (1968). Some rules for a combinatorial method for multiple products of generalized k -statistics. *Ann. Math. Statist.* **39** 983-998.
- [10] TUKEY, JOHN W. (1956). Keeping moment-like sampling computations simple. *Ann. Math. Statist.* **27** 37-54.
- [11] WISHART, JOHN (1952). Moment coefficients of the k -statistics in samples from a finite population. *Biometrika* **39** 1-13.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WINDSOR
WINDSOR 11, ONTARIO, CANADA

INSTITUTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
RIO DE JANEIRO, BRAZIL