NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

A NOTE ON SHEPPARD'S CORRECTIONS

By Cecil C. Craig

University of Michigan

As far as the author is aware, H. C. Carver was the first to point out that while the formulae ordinarily given for Sheppard's corrections for central moments are valid for moments computed about the population mean, there are still systematic errors present when they are applied to central moments calculated from any particular grouped frequency distribution [1]. This is due, of course to the fact that the mean of a grouped frequency distribution is in general different from that of the distribution before grouping. For a fixed class interval k, Sheppard's corrections give the average value of a moment about a fixed point of a given order for all the groupings of this class width possible and will fail to do so if the moment in question is calculated for each position of the class limits about a point which varies as the class limits shift. Thus Carver [1] pointed that the commonly used formula (for a continuous variate),

$$\mu_2 = \nu_2 - \frac{k^2}{12}, \tag{1}$$

should, if ν_2 is calculated about the mean of the grouped distribution as it is in practice, be replaced by

$$\mu_2 = \nu_2 - \frac{k^2}{12} + \sigma_M^2 \tag{2}$$

in which σ_M^2 is the variance of the means of grouped distributions over all positions of the class limits with the fixed class width k.

Recently J. A. Pierce [2] gave a method for deriving the required formulae of the type of (2) and gave actual formulae for both moments and seminvariants through the sixth order. It is the purpose of this note to point out that the use of moment generating functions provides a more elegant and concise way of arriving at formulae equivalent to Pierce's though in a somewhat different form. This method can be immediately extended to distributions of two or more variates.

In a previous paper [3] on Sheppard's corrections for a discrete variate, the author made use of the following argument: It is assumed that for a fixed class width k, any point in the scale on which the variate x is plotted is as likely to be chosen as a class limit as any other; choosing a system of class limits for grouping the data is then equivalent to placing at random on the x-axis a scale with division points at intervals of k. Once the system of class limits is chosen any value of x before grouping bears to the class mark, x_i, of the class in which it falls the relation,

$$x_i = x + \epsilon, \tag{3}$$

in which x and ϵ are independent variates. The frequency law governing x, is, of course, that of the population from which it is drawn while ϵ is distributed in a rectangular distribution with the range $\left(-\frac{k}{2}, \frac{k}{2}\right)$ for a continuous variate and $\left(-\frac{m-1}{2m}k, \frac{m-1}{2m}k\right)$ if m consecutive values of a discrete variate are grouped in each class interval. In either case

$$M_{x_i}(\vartheta) = M_x(\vartheta)M_\epsilon(\vartheta) \tag{4}$$

in which $M_{x_i}(\vartheta)$ is the moment generating function of the variate x_i, etc. The expansion of both sides of (4) in powers of ϑ gives the relations between the average values of moments of the grouped distribution over all positions of the scale and the moments of the ungrouped distribution from which Sheppard's corrections are obtained by solving for the moments of the ungrouped distribution. The relations are valid for any fixed point about which the moments are computed; if this fixed point be taken as the mean of the ungrouped distribution the ordinary Sheppard's corrections for central moments result.

But it is quite easy to modify (4) to give the necessary relations in case the moments of each grouped distribution are computed about the mean of that distribution. We have only to write

$$x_i = x_i - \bar{x} + \bar{x} \tag{5}$$

in which $\bar{x}$ is the mean of the grouped distribution for which x_i is one of the class marks. Then

$$\begin{aligned} M_{x_i}(\vartheta) &= M_{x_i-\bar{x},\bar{x}}(\vartheta, \omega)\,|_{\omega=\vartheta} \\ &= M_x(\vartheta)M_\epsilon(\vartheta) \end{aligned} \tag{6}$$

If we write,

$$\lambda_{rs:k_i-\bar{x},\bar{x}} = \bar{\lambda}_{rs},$$

in which $\bar{\lambda}_{rs}$ is the product seminvariant of order rs of moments about the means of the grouped distributions and of such means, the expansion of the logarithm of the second member of (6) gives

$$1 + (\bar{\lambda}_{10} + \bar{\lambda}_{01})\vartheta + (\bar{\lambda}_{20} + 2\bar{\lambda}_{11} + \bar{\lambda}_{02})\frac{\vartheta^2}{2} + (\bar{\lambda}_{10} + \bar{\lambda}_{01})^{(3)}\frac{\vartheta^3}{3!} + \cdots, \tag{7}$$

in which

$$(\bar{\lambda}_{10} + \bar{\lambda}_{01})^{(r)} = \bar{\lambda}_{r0} + r\bar{\lambda}_{r-1,1} + \cdots + \binom{r}{k}\bar{\lambda}_{r-k,k} + \cdots + \bar{\lambda}_{0r}.$$

The expression of the logarithm of the right member is[3]:

$$(8) \qquad \lambda_1\vartheta + \lambda_2\frac{\vartheta^2}{2!} + \lambda_3\frac{\vartheta^3}{3!} + \cdots + \sum_{s=1}^{\infty}(-1)^{s+1}\frac{B_s k^{2s}}{2s}\left(1 - \frac{1}{m^{2s}}\right)\frac{\vartheta^{2s}}{(2s)!},$$

for a discrete variate (the result for a continuous variable is obtained merely by letting $m \to \infty$) in which λ_r is the rth seminvariant of the ungrouped distribution and B_s is the sth Bernoulli number.

We may without loss of generality take the origin for x at the mean of the ungrouped distribution so that $\lambda_1 = 0$. Further it is easy to see that

$$\bar{\lambda}_{1r} = 0, \qquad r = 0, 1, 2, 3, \cdots.$$

Consider

$$E[(x_i - \bar{x})\bar{x}^r] = \bar{\nu}_{1r}$$

For a fixed $\bar{x}$, i.e., for a given grouping, this becomes

$$\bar{x}^r E(x_i - \bar{x}) = 0$$

Then since $\bar{\nu}_{1r}$ is the average of this over all groupings with a given class interval, $\bar{\nu}_{1r} = 0$, and from the expression for $\bar{\lambda}_{1r}$ in terms of the moments $\bar{\nu}_{ij}$ it is obvious that also $\bar{\lambda}_{1r} = 0$.

Then we must also have $\bar{\lambda}_{01} = 0$ as is otherwise obvious and (7) can be rewritten

$$(9) \qquad 1 + (\bar{\lambda}_{20} + \bar{\lambda}_{02})\frac{\vartheta^2}{2} + (\bar{\lambda}_{30} + 3\bar{\lambda}_{21} + \bar{\lambda}_{03})\frac{\vartheta^3}{3!} + \cdots.$$

Now from (8) and (9) by equating coefficients of like powers of ϑ, we get the set of formulae:

$$(10) \qquad \begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= \bar{\lambda}_{20} + \bar{\lambda}_{02} - \left(1 - \frac{1}{m^2}\right)\frac{k^2}{12} \\ \lambda_3 &= \bar{\lambda}_{30} + 3\bar{\lambda}_{21} + \bar{\lambda}_{03} \\ \lambda_4 &= \bar{\lambda}_{40} + 4\bar{\lambda}_{31} + 6\bar{\lambda}_{22} + \bar{\lambda}_{04} + \left(1 - \frac{1}{m_4}\right)\frac{k^4}{120} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

These formulae, however, do not give the sought Sheppard's corrections for seminvariants calculated from grouped distributions of a discrete variate. See below.

Referring to formula (10), p. 58 of the author's paper cited [3], it is easily seen by comparison that the required moment formulae are obtained from the general formula

$$(11) \qquad \mu_n = \sum_{s=0}^{[n/2]}\binom{n}{2s}\alpha_{2s}(\bar{\nu}_{10} + \bar{\nu}_{01})^{(n-2s)},$$

in which α_{2s} is given by formula (9) of this former paper. For $n = 1, 2, 3, 4$ we write down immediately

$$\mu_1 = 0 \qquad (\bar{\nu}_{10} = \bar{\nu}_{01} = 0)$$

$$\mu_2 = \bar{\nu}_{20} + \bar{\nu}_{02} - \left(1 - \frac{1}{m^2}\right)\frac{k^2}{12}$$

(12) $$\mu_3 = \nu_{30} + 3\bar{\nu}_{21} + \bar{\nu}_{03}$$

$$\mu_4 = \nu_{40} + 4\bar{\nu}_{31} + 6\bar{\nu}_{22} + \bar{\nu}_{04} - \left(1 - \frac{1}{m^2}\right)(\bar{\nu}_{20} + \bar{\nu}_{02})\frac{k^2}{2} + \left(1 - \frac{1}{m^2}\right)\left(7 - \frac{3}{m^2}\right)\frac{k^4}{240}.$$

In these formulae, $\bar{\nu}_{r0}$ is, of course, the average value of rth central moments about the means of grouped distributions. From the definition $\bar{\nu}_{rs}(s \neq 0)$ is the average value of the product of the rth central moment of a grouped distribution by the sth power of the mean of the same grouped distribution. Also, it must be noted that in the formulae (10) the $\bar{\lambda}_{rs}$'s there are to be calculated by the usual formulae from the moments, $\bar{\nu}_{ij}$, and are not themselves the average values of like seminvariants calculated from the separate grouped distributions. Thus though the formulae (12) give the sought Sheppard's corrections for moments, the formulae (10) do not do the like for seminvariants in general. However, since in each grouped distribution,

$$\lambda_2 = \nu_2$$

and

$$\lambda_3 = \nu_3$$

we have, taking the expectation or average value over the grouped distributions,

$$E(\lambda_2) = E(\nu_2) = \bar{\nu}_{20} = \bar{\lambda}_{20}$$

and

$$E(\lambda_3) = E(\nu_3) = \bar{\nu}_{30} = \bar{\lambda}_{30},$$

and the first two formulae of (10) do give the Sheppard's corrections for λ_2 and λ_3 calculated from grouped distributions of a discrete variate.

But the case for λ_4 is different. In each grouped distribution,

$$\lambda_4 = \nu_4 - 3\nu_2^2,$$

and if we define l_r by

$$E(\lambda_r) = l_r,$$

we have

$$l_4 = \bar{\nu}_{40} - 3E(\nu_2^2)$$
$$= \bar{\nu}_{40} - 3(\bar{\nu}_{20}^2 + \nu_{2:\nu_2}) = \bar{\lambda}_{40} - 3\nu_{2:\nu_2},$$

if $\nu_{2:\nu_2}$ is the variance of ν_2 in the grouped distributions.

In a similar way one can obtain such formulae for seminvariants as may be required. Through the sixth, the formulae for the Sheppard's corrections for the seminvariants calculated from a grouped distribution of a discrete variate are:

$$
\begin{aligned}
\lambda_2 &= l_2 + \bar{\lambda}_{02} - \left(1 - \frac{1}{m^2}\right)\frac{k^2}{12} \\
\lambda_3 &= l_3 + 3\bar{\lambda}_{21} + \bar{\lambda}_{03} \\
\lambda_4 &= l_4 + 3\nu_{2:\nu_2} + 4\bar{\lambda}_{31} + 6\bar{\lambda}_{22} + \bar{\lambda}_{04} + \left(1 - \frac{1}{m^4}\right)\frac{k_4}{120} \qquad (13) \\
\lambda_5 &= l_5 + 10\nu_{11:\nu_2,\nu_3} + 5\bar{\lambda}_{41} + 10\bar{\lambda}_{32} + 10\bar{\lambda}_{23} + \bar{\lambda}_{05} \\
\lambda_6 &= l_6 + 15\nu_{11:\nu_2,\nu_4} + 10\nu_{2:\nu_2} - 30\nu_{3:\nu_2} - 90\nu_{2:\nu_2}\bar{\nu}_{20} \\
&\quad + 6\bar{\lambda}_{51} + 15\bar{\lambda}_{42} + 20\bar{\lambda}_{33} + 15\bar{\lambda}_{24} + \bar{\lambda}_{06} - \left(1 - \frac{1}{m^6}\right)\frac{k^6}{252}.
\end{aligned}
$$

In these formulae, $\nu_{ij:\nu_r,\nu_s}$ is the ijth central product moment of ν_r and ν_s in the grouped distributions.

To illustrate these formulae numerically and to facilitate comparison with Pierce's results, we will use the example he chose. His ungrouped distribution was:

v	f	v	f	v	f
1	2	4	30	7	1
2	8	5	4	8	1
3	10	6	3	9	1

From this the following three grouped distributions with $k = 3$ can be formed:

(1)		(2)		(3)	
class	f	class	f	class	f
1- 3	20	0- 2	10	−1 [-1]	2
4-	37	3-	44	2-	48
7-	3	6-	5	5-	8
10–12	0	9–11	1	8–10	2

With origin at $\nu = 4$, we have the following table of moment characteristics of these four distributions:

Distribution	ν_1'	$\nu_2 = \bar{\lambda}_2$	$\nu_3 = \bar{\lambda}_3$	ν^4	$\bar{\lambda}^4$	$\delta\nu_1' = \nu_1' - \left(-\frac{10}{60}\right)$
(1)	$\frac{9}{60}$	$\frac{9819}{60^2}$	$-\frac{17442}{60^3}$	$\frac{238849317}{60^4}$	$-\frac{50388966}{60^4}$	$\frac{19}{60}$
(2)	$-\frac{9}{60}$	$\frac{10179}{60^2}$	$\frac{567162}{60^3}$	$\frac{557840277}{60^4}$	$\frac{247004154}{60^4}$	$\frac{1}{60}$
(3)	$-\frac{30}{60}$	$\frac{8820}{60^2}$	$\frac{1317600}{60^3}$	$\frac{528282000}{60^4}$	$\frac{294904800}{60^4}$	$-\frac{20}{60}$
Average	$-\frac{10}{60}$	$\frac{9606}{60^2}$	$\frac{622440}{60^3}$	$\frac{441657198}{60^4}$	$\frac{163839996}{60^4}$	
	μ_1'	$\mu_2 = \lambda_2$	$\mu_3 = \lambda_3$	μ_4	λ_4	
Original Distribution	$-\frac{10}{60}$	$\frac{7460}{60^2}$	$\frac{642400}{60^3}$	$\frac{305034000}{60^4}$	$\frac{138079200}{60^4}$	

From the table,

$$\bar{\nu}_{20} = \bar{\lambda}_{20} = \frac{9606}{60^2}$$

$$\bar{\nu}_{30} = \bar{\lambda}_{30} = \frac{622440}{60^3}$$

$$\bar{\nu}_{40} = \bar{\lambda}_{40} + 3\bar{\lambda}_{20}^2 = \frac{441657198}{60^4}.$$

We further compute:

$$\bar{\nu}_{02} = \frac{\Sigma(\delta\nu_1')^2}{3} = \frac{254}{60^2} = \bar{\lambda}_{02} \qquad \bar{\nu}_{21} = \frac{\Sigma(\nu_2\delta\nu_1')}{3} = \frac{6780}{60^3} = \bar{\lambda}_{21}$$

$$\bar{\nu}_{03} = \frac{-380}{60^3} = \bar{\lambda}_{03} \qquad \bar{\nu}_{30} = -\frac{8705412}{60^4} = \bar{\lambda}_{31}$$

$$\bar{\nu}_{04} = \frac{96774}{60^4} \qquad \bar{\nu}_{22} = \frac{2360946}{60^4}$$

$$\bar{\lambda}_{22} = \bar{\nu}_{22} - \bar{\nu}_{20}\bar{\nu}_{02} = \frac{-72978}{60^4}$$

$$\bar{\lambda}_{04} = \bar{\nu}_{04} - 3\bar{\nu}_{02}^2 = \frac{-96774}{60^4}$$

$$\nu_{2:\nu_2} = \frac{\Sigma \nu_2^2}{3} - \bar{\nu}_{20}^2 = \frac{330948}{60^4}$$

$$l_4 = \bar{\lambda}_{40} - 3\nu_{2:\nu_2} = \bar{\nu}_{40} - 3\bar{\nu}_{20}^2 - 3\nu_{2:\nu_2} = \frac{163839996}{60^4}$$

$$\left(1 - \frac{1}{m^2}\right)\frac{k^2}{12} = \frac{2}{3}$$

$$\left(1 - \frac{1}{m^2}\right)\left(7 - \frac{3}{m^2}\right)\frac{k^4}{240} = 2.$$

With these values one may check the formulae (12) and (13) as far as weight four. For example:

$$\mu_2 = \frac{9606}{60^2} + \frac{254}{60^2} - \frac{2}{3} = \frac{7460}{60^2}$$

$$\lambda_4 = \frac{1}{60^4}(163839996 + 991494 - 34821648 - 437868 - 96774 + 8640000)$$

$$= \frac{138079200}{60^4}.$$

It may appear at first glance that since

$$\bar{\nu}_{rs} = E[\nu_r(\delta\nu_1)^s]$$

and could be expressed by means of the notation, $\nu_{1s:\nu_r,\nu_1}{}'$, the notation in (12) and (13) could be made more uniform. It could be but at the expense of greater complexity in these two sets of results. Moreover, it is convenient that $\bar{\lambda}_{rs}$ is expressible in terms of $\bar{\nu}_{kl}$'s in precisely the same way that product seminvariants are ordinarily expressible in terms of product moments.

Pierce's results differ from the above not only in their mode of derivation but also in the fact that they express $\bar{\nu}_{r0}$'s and l_r's in terms of the characteristics of the ungrouped distribution and moments and seminvariants of moments in the grouped distributions. Thus as they stand they are not formulae for Sheppard's corrections.

Finally it must be remarked that in comparison with the usual formulae for Sheppard's corrections, the formulae (10) and (13) introduce quantities the magnitudes of which are not known in general except that ordinarily they are quite small. It is hoped that results on this point will be forthcoming soon.

REFERENCES

[1] H. C. Carver, "The fundamental nature and proof of Sheppard's adjustments," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 154–163.

[2] J. A. Pierce, "A study of a universe of n finite populations with application to moment-function adjustments for grouped data," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 311–334.

[3] C. C. Craig, "Sheppard's corrections for a discrete variable," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 55–61.