VARIANCES OF VARIANCE COMPONENTS: III. THIRD MOMENTS IN A BALANCED SINGLE CLASSIFICATION[1]

By John W. Tukey

Princeton University

1. Summary. The methods used in earlier papers of this series [2], [3] are extended from variances to third moments, and applied to the third moment about the mean (= third cumulant) of the usual estimate of the between variance component in a balanced single classification. The result is moderately complex, but manageable.

2. Introduction. The technique of this paper grows directly out of those used for variances in the earlier papers in this series [2], [3]. We assume familiarity with the terminology and notations used there.

We begin by discussing the third moment of the variance of a sample as an illustration of problems and technique, and then pass on directly to the main problem. We need the multiplication formulas (see Wishart [4])

$$k_2^2 = k_{22} + \frac{1}{n} k_4 + \frac{2}{n-1} k_{22},$$

$$k_2^3 = k_{222} + \frac{3(n+3)}{n(n-1)} k_{24} + \frac{6n+2}{(n-1)^2} k_{222} + \frac{1}{n^2} k_6 + \frac{4(n-2)}{n(n-1)^2} k_{33}.$$

$$(k_2')^3 = k_{222}' + 3\,\frac{N+3}{N(N-1)} k_{24}' + \frac{6N+2}{(N-1)^2} k_{222}' + \frac{1}{N^2} k_6' + 4\,\frac{(N-2)}{(N-1)^2} k_{33}',$$

$$k_2' k_{22}' = k_{222}' + \frac{2}{N} k_{24}' + \frac{4}{N-1} k_{222}' - \frac{2}{N(N-1)} k_{33}',$$

$$k_2' k_4' = k_{24}' + \frac{1}{N} k_6' + \frac{8}{N-1} k_{24}' + \frac{6}{N-1} k_{33}'$$

3. The variance. We can now proceed to write down the third moment about the mean (that is the *third cumulant*) of the variance of a sample drawn from a finite population. There are various ways to do this, but we shall begin with one resembling the first method we used [1] to get the variance of the variances (we refer to using primes for population quantities for three sections):

$$\begin{aligned} \text{ave}\,\{k_2 - k_2'\}^3 &= \text{ave}\,\{k_2^3 - 3k_2^2 k_2' + 3k_2(k_2')^2 - (k_2')^3\} \\ &= \text{ave}\,\{k_2^3\} - 3k_2'\,\text{ave}\,\{k_2^2\} + 2(k_2')^3 \end{aligned}$$

Received April 6, 1955.

[1] Prepared in connection with research sponsored by the Office of Naval Research and based on part of Memorandum Report 45, "Finite Sampling Simplified," Statistical Research Group, Princeton University, which was written while the author was a Fellow of the John Simon Guggenheim Foundation.

$$= \operatorname{ave}\left\{k_{222} + \frac{3(n+3)}{n(n-1)} k_{24} + \frac{6n+2}{(n-1)^2} k_{222} + \frac{1}{n^2} k_6 + \frac{4(n-2)}{n(n-1)^2} k_{33}\right\}$$

$$- 3k_2' \operatorname{ave}\left\{k_{22} + \frac{1}{n} k_4 + \frac{2}{n-1} k_{22}\right\} + 2k_2'^3$$

$$= k_{222}' + \frac{3(n+3)}{n(n-1)} k_{24}' + \frac{6n+2}{(n-1)^2} k_{222}' + \frac{1}{n^2} k_6' + \frac{4(n-2)}{n(n-1)^2} k_{33}'$$

$$- 3k_2'k_{22}' - \frac{3}{n} k_2'k_4' - \frac{6}{n-1} k_2'k_{22}' + 2(k_2')^3$$

$$= k_{222}' + \frac{3(n+3)}{n(n-1)} k_{24}' + \frac{6n+2}{(n-1)^2} k_{222}' + \frac{1}{n^2} k_6' + \frac{4(n-2)}{n(n-1)^2} k_{33}'$$

$$- 3k_{222}' - \frac{6}{N} k_{24}' - \frac{12}{N-1} k_{222}' + \frac{6}{N(N-1)} k_{33}' - \frac{3}{n} k_{24}'$$

$$- \frac{3}{nN} k_6' - \frac{24}{(N-1)n} k_{24}' - \frac{18}{(N-1)n} k_{33}' - \frac{6}{n-1} k_{222}' - \frac{12}{(n-1)N} k_{24}'$$

$$- \frac{24}{(n-1)(N-1)} k_{222}' + \frac{12}{(n-1)N(N-1)} k_{33}' + 2k_{222}'$$

$$+ 6\frac{N+3}{N(N-1)} k_{24}' + 2\frac{6N+2}{(N-1)^2} k_{222}' + \frac{2}{N^2} k_6' + \frac{8(N-2)}{N(N-1)^2} k_{33}'.$$

This may be written as

$$\left\{\frac{1}{n^2} - \frac{3}{nN} + \frac{2}{N^2}\right\} k_6' + \left\{\frac{4}{n(n-1)} - \frac{18}{(n-1)N} + \frac{14}{N(N-1)}\right\} k_{33}'$$

$$- \left\{\frac{4}{n(n-1)^2} - \frac{12}{(n-1)N(N-1)} + \frac{8}{N(N-1)^2}\right\} k_{33}'$$

$$+ \left\{\frac{8}{(n-1)^2} - \frac{24}{(n-1)(N-1)} + \frac{16}{(N-1)^2}\right\} k_{222}'$$

$$+ \left\{\frac{12}{n(n-1)} - \frac{12}{(n-1)N} - \frac{24}{(N-1)n} + \frac{24}{N(N-1)}\right\} k_{24}',$$

which vanishes for $n = N$, just as it should.

Now it is even clearer for the third cumulant (which might perhaps be called the *skewmulant*) than it was for the variance, that direct calculation would be long, tedious, boring, and, because of its length, likely to be erroneous for any component in an analysis of variance situation. We must go to more ingenious methods.

4. Structure. We notice that the third cumulant of any algebraic function of a random sample takes the form

$$\text{ave}\,\{u^4\} - 3\,\text{ave}\,\{u\}\,\text{ave}\,\{u^2\} + 2(\text{ave}\,\{u\})^3 = \alpha(n) - 3\delta\beta(n)\gamma(n) + 2\delta^4,$$

where $\alpha(n)$, $\beta(n)$, and $\gamma(n)$ depend on n alone, and δ is independent of n and N, all four being expressable linearly in polykays. The only appearance of N comes from the multiplication formulas involved in

$$\beta(n)\cdot\gamma(n)$$

and

$$\delta\cdot\delta\cdot\delta.$$

The final result has the form

$$A(n) + \sum B(n)C(N) + D(N),$$

where the functions of N come from

$$-3\,\text{ave}\,\{u\}\,\text{ave}\,\{u^2\} + 2\{\text{ave}\,\{u\}\}^3,$$

which we may suppose already known.

We can again proceed by finding these two terms first, and then using special populations to determine the coefficients in $A(n)$.

5. The variance-second method. We have

$$\begin{aligned}
\text{ave}\,\{k_2\} &= k_2',\\
\text{ave}\,\{k_2^2\} &= (\text{ave}\,\{k_2\})^2 + \text{var}\,\{k_2\}\\
&= (k_2')^2 + \left(\frac{1}{n} - \frac{1}{N}\right)k_4' + \left(\frac{2}{n-1} - \frac{2}{N-1}\right)k_{22}'\\
&= \left\{1 + \frac{1}{n-1}\right\}k_{22}' + \frac{1}{n}k_4',
\end{aligned}$$

then $\sum B(n)C(N) + D(N)$ must come from

$$-\frac{6}{n-1}k_{22}'k_2' - \frac{3}{n}k_4'k_2' + 2(k_2')^3 - 3k_{22}'k_2',$$

whence

$$\begin{aligned}
\sum B(n)C(N) = &-\frac{6}{n-1}\left\{\frac{2}{N}k_{24}' + \frac{4}{N-1}k_{222}' - \frac{2}{N(N-1)}k_{33}'\right\}\\
&-\frac{3}{n}\left\{\frac{1}{N}k_6' + \frac{8}{N-1}k_{24}' + \frac{6}{N-1}k_{33}'\right\},
\end{aligned}$$

$$\begin{aligned}
D(N) = 2\left\{3\frac{N+3}{N(N-1)}k_{24}' + \frac{6N+2}{(N-1)^2}k_{222}' + \frac{1}{N^2}k_6' + 4\frac{N-2}{N(N-1)^2}k_{33}'\right\}\\
-3\left\{\frac{2}{N}k_{24}' + \frac{4}{N-1}k_{222}' - \frac{2}{N(N-1)}k_{33}'\right\},
\end{aligned}$$

thus, the part of the third cumulant of the usual estimate of k_2 which depends on the population size, N, is

$$\left\{\frac{2}{N^2} - \frac{3}{nN}\right\} k_6'$$

$$+ \left\{\frac{14}{N(N-1)} - \frac{18}{n(N+1)} - \frac{8}{N(N-1)^2} + \frac{12}{(n-1)N(N-1)}\right\} k_{33}'$$

$$\left\{\frac{24}{N(N-1)} - \frac{12}{N(n-1)} - \frac{24}{n(N-1)}\right\} k_{24}'$$

$$+ \left\{\frac{16}{(N-1)^2} - \frac{24}{(n-1)(N-1)}\right\} k_{222}'.$$

When $n = N$ this reduces to the negative of

$$(*) \quad \frac{1}{n^2} k_6' + \left\{\frac{4}{n(n-1)} - \frac{4}{n(n-1)^2}\right\} k_{33}' + \frac{12}{n(n-1)} k_{24}' + \frac{8}{(n-1)^2} k_{222}',$$

and since the k_2 estimate is constant when $n = N$ the additional terms not involving N must be those just given in (*).

6. The between component in a balanced single classification. We now come back to the balanced single classification and drop the primes. Hence the column contribution is drawn from n $k_1, k_{11}, \cdots$, and the error contribution is drawn from $N, K_1, K_{11}, \cdots$. We are going to find the third cumulant of the between component in sampling from an arbitrary finite population.

When we express this third cumulant multilinearly in the k's and K's, there may be terms involving

(1) $k_6, k_{24}, k_{33}, k_{222}$,

(2) $k_4K_2, k_{22}K_2, k_3K_3, K_2K_4, k_2K_{22}$,

(3) $K_6, K_{24}, K_{33}, K_{222}$,

and no others (because of homogeneity and invariance under translation). The discussion just given for the variance of a sample applies with minor changes to the case where the errors are constant. This determines the coefficient of the terms involving the monomials in (1) to be the same as for the third cumulant of the variance, with c replacing n.

We go next to the terms involving the monomials in (3). Suppose first that the column contributions are constant. If we take a minimal unit population for the individual contributions, the between component is constant. Thus, the desired third cumulant, and everything it could involve except K_6, vanishes. Thus the coefficient of K_6 is itself zero.

In order to deal with the next coefficients, we can use a population with two non-zero values for the individual contributions. But if both values are alike, we cannot distinguish between the coefficients of K_{33} and K_{24}. So we use a population of rc values 1, t, 0, $\cdots$, 0, for which

$$K_{33} = \frac{2t^3}{rc(rc-1)}, \qquad K_{24} = \frac{t^2 + t^4}{rc(rc-1)}.$$

There are two cases to consider

CASE 1. (Probability $(r - 1)/(rc - 1)$.) One $x = 1$ and another $x = t$ in the same column; others, zero.

CASE 2. (Probability $r(c - 1)/(rc - 1)$.) One $x = 1$ and another $x = t$ in different columns; others, zero.

The corresponding analyses of variance are

CASE 1

	DF	SS	MS	CMS
Between	$c - 1$	$\frac{(t+1)^2(c-1)}{rc}$	$\frac{(t+1)^2}{rc}$	$\frac{2t}{rc(r-1)}$
Within	$c(r - 1)$	$t^2 + 1 - \frac{(t+1)^2}{r}$	$\frac{t^2+1}{rc} - \frac{2t}{rc(r-1)}$	$\frac{t^2+1}{rc} - \frac{2t}{rc(r-1)}$

CASE 2

	DF	SS	MS	CMS
Between	$c - 1$	$\frac{(c-1)(t^2+1)}{rc} - \frac{2t}{rc}$	$\frac{t^2+1}{rc} - \frac{2t}{rc(c-1)}$	$\frac{-2t}{r^2c(c-1)}$
Within	$c(r - 1)$	$\frac{(r-1)(t^2+1)}{r}$	$\frac{t^2+1}{rc}$	$\frac{t^2+1}{rc}$

Thus the mean of the between component vanishes, as it should, and the third cumulant reduces to the third moment about zero, which is

$$\frac{r-1}{rc-1}\left\{\frac{2}{rc}\frac{t}{r-1}\right\}^3 + \frac{r(c-1)}{rc-1}\left\{-\frac{2}{rc}\frac{t}{r(c-1)}\right\}^3$$

$$= \frac{8t^3}{r^3c^3(rc-1)}\left\{\frac{1}{(r-1)^2} - \frac{1}{r^2(c-1)^2}\right\}.$$

This has t^3 as a factor, and does not involve $t^2 + t^4$, hence the coefficient of K_{24} is zero, while that of K_{33} is

$$\frac{4}{r^2c^2}\left\{\frac{1}{(r-1)^2} - \frac{1}{r^2(c-1)^2}\right\},$$

which, as we might have suspected, is small when the two sets of degrees of freedom are of nearly the same size.

We can deal with the coefficient of K_{222}, and certain of the other terms, by resorting to normal theory. In this case, the two mean squares become independent and their third cumulants are

$$8(K_2 + rk_2)^3/(c - 1)^2, \qquad 8K_2^3/c^2(r - 1)^2,$$

so that the third cumulant of the between component is

$$\frac{8}{r^3}\left\{\frac{(K_2 + rk_2)^3}{(c-1)^2} - \frac{K_2^3}{c^2(r-1)^2}\right\} = \frac{8}{r^3}\left(\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right)K_2^3$$

$$+ \frac{3r}{(c-1)^2}K_2^2k_2 + \frac{3r^2}{(c-1)^2}K_2k_2^2 + \frac{r^3}{(c-1)^2}k_2^3\Bigg\}$$

$$= \frac{8}{r^3}\left(\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right)K_{222} + \frac{24}{r^2(c-1)^2}K_{22}k_2$$

$$+ \frac{24}{r(c-1)^2}K_2k_{22} + \frac{8}{(c-1)}k_{222}.\Bigg\}$$

Thus, we have determined the coefficients of K_{222}, $K_{22}k_2$, and K_2k_{22}, and have confirmed the part of the coefficient of k_{222} which is independent of N.

There remain the coefficients of k_4K_2, k_3K_3, and k_2K_4. These we shall seek by taking a minimal unit population for the column contributions and a minimal population with one nonzero value equal to t for the error. We have

$$k_4 = 1/c, \qquad k_3 = 1/c, \qquad k_2 = 1/c,$$
$$K_4 = t^4/rc, \qquad K_3 = t^3/rc, \qquad K_2 = t^2/rc,$$

so that

$$k_4K_2 = \frac{t^2}{rc^2}, \qquad k_3K_3 = \frac{t^3}{rc^2}, \qquad k_2K_4 = \frac{t^4}{rc^2}.$$

We have the usual two cases to deal with.

CASE 3. (Probability $1/c$). One $x = t + 1$, others in column $= 1$, others zero.

CASE 4. (Probability $(c - 1)/c$). One column of x's $= 1$, one other $x = t$, all others zero.

Corresponding analyses of variance are

CASE 3

	DF	SS	MS	CMS
Between	$c - 1$	$\frac{(t + r)^2(c - 1)}{rc}$	$\frac{(t + r)^2}{rc}$	$\frac{1}{c} + \frac{2t}{rc}$
Within	$c(r - 1)$	$t^2\frac{r - 1}{r}$	$\frac{t^2}{rc}$	$\frac{t^2}{rc}$

CASE 4

	DF	SS	MS	CMS
Between	$c - 1$	$\left(r + \frac{t^2}{r}\right)\frac{c - 1}{c} - \frac{2t}{c}$	$\frac{t^2 + r^2}{rc} - \frac{2t}{c(c - 1)}$	$\frac{1}{c} - \frac{2t}{rc(c - 1)}$
Within	$c(r - 1)$	$t^2\left(\frac{r - 1}{r}\right)$	$\frac{t^2}{rc}$	$\frac{t^2}{rc}$

The mean between component is $1/c$, as it should be, and its third moment about the mean is

$$\frac{1}{c}\left(\frac{2t}{rc}\right)^3 + \frac{c-1}{c}\left(-\frac{2t}{rc(c-1)}\right)^3 = \frac{8t^3}{r^3c^3}\left\{1 - \frac{1}{(c-1)^2}\right\}.$$

The terms in t^2 and t^4 are conspicuous in their absence, and hence the coefficients of k_4K_2 and k_2K_4 vanish. The coefficient of k_3K_3 is

$$\frac{4}{r^2c^2}\left\{1 - \frac{1}{(c-1)^2}\right\}.$$

We can now reassemble all our coefficients and write down the third cumulant of the between component in a balanced single classification with c categories and r units per category. The result is

$$\begin{aligned}
&\left\{\frac{1}{c^2} - \frac{3}{cn} + \frac{2}{n^2}\right\} k_6 + \left\{\frac{4}{c(c-1)} - \frac{18}{(c-1)n} + \frac{14}{n(n-1)}\right\} k_{33} \\
&\quad - \left\{\frac{4}{c(c-1)^2} - \frac{12}{(c-1)n(n-1)} + \frac{8}{n(n-1)^2}\right\} k_{33} \\
&\quad + \left\{\frac{12}{c(c-1)} - \frac{12}{(c-1)n} - \frac{24}{c(n-1)} + \frac{24}{n(n-1)}\right\} k_{24} \\
&\quad + \left\{\frac{8}{(c-1)^2} - \frac{24}{(c-1)(n-1)} + \frac{16}{(n-1)^2}\right\} k_{222} \\
&\quad + \frac{24}{r(c-1)^2} k_{22}K_2 + \frac{4}{r^2c^2}\left\{1 - \frac{1}{(c-1)^2}\right\} k_3K_3 + \frac{24}{r^2(c-1)^2} k_2K_{22} \\
&\quad + \frac{4}{r^2c^2}\left\{\frac{1}{(r-1)^2} - \frac{1}{r^2(c-1)^2}\right\} K_{33} + \frac{8}{r^3}\left\{\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right\} K_{222},
\end{aligned}$$

for populations of n column contributions and N error contributions.

For infinite populations this reduces slightly, and becomes

$$\begin{aligned}
&\frac{1}{c^2} k_6 + \frac{4(c-2)}{c(c-1)^2} k_{33} + \frac{12}{c(c-1)} k_{24} + \frac{8}{(c-1)^2} k_{222} + \frac{24}{r(c-1)^2} k_{22}K_2 \\
&\quad + \frac{4}{r^2c^2}\left\{1 - \frac{1}{(c-1)^2}\right\} k_3K_3 + \frac{24}{r^2(c-1)^3} k_2K_{22} \\
&\quad + \frac{4}{r^2c^2}\left\{\frac{1}{(r-1)^2} - \frac{1}{r^2(c-1)^2}\right\} K_{33} + \frac{8}{r^3}\left\{\frac{1}{(c-1)^2} - \frac{1}{c^2(r-1)^2}\right\} K_{222}.
\end{aligned}$$

REFERENCES

[1] JOHN W. TUKEY, "Some sampling simplified," *J. Amer. Statist. Assn.*, Vol. 45 (1950), 501–519.

[2] JOHN W. TUKEY, "Variances of variance components: I. Balanced designs," *Ann. Math. Stat.*, Vol. 27 (1956), 722–736.

[3] JOHN W. TUKEY, "Variances of variance components: II. Unbalanced single classifications." *Ann. Math. Stat*, Vol. 28 (1957), 43–56.

[4] JOHN WISHART, "Moment coefficients of the k-statistics in samples from a finite population," *Biometrika*, Vol. 39, pp. 1–13.