CHAPTER 4

Sample cumulants

4.1 Introduction

We now address problems of a more inferential flavour, namely problems concerning point estimation or, better, interval estimation of population cumulants based on observed data. Moment and cumulant estimators have a lengthy history in the statistical literature going back to the work of K. Pearson, Chuprov, Student, Fisher and others in the early part of this century. In recent years, the volume of work on such topics has declined, partly because of concerns about robustness, sensitivity to outliers, misrecorded or miscoded values and so on. Estimates of the higher-order cumulants are sensitive to such errors in the data and, for some purposes, this sensitivity may be considered undesirable. In addition, the variance of such an estimate depends on cumulants up to twice the order of the estimand and these are even more difficult to estimate accurately. If there is insufficient data it may not be possible to estimate such cumulants at all. We are confronted immediately and forcibly with what Mosteller & Tukey (1977, Chapter 1) call 'the misty staircase'. That is to say that, to assess the variability of a primary statistic, we compute a secondary statistic, typically more variable than the primary statistic. It is then possible to estimate the variability of the secondary statistic by computing a third statistic and so ad infinitum. Fortunately, we usually stop long before this extreme stage.

Section 4.2 is concerned with simple random samples from an infinite population. The emphasis is on the various symmetric functions, i.e. functions of the data that are unaffected by re-ordering the data values. The functions having the most pleasing statistical properties are the k-statistics and, to a lesser extent, the generalized k-statistics. The k-statistics are unbiased estimates of ordinary cumulants and generalized k-statistics are estimates of generalized cumulants, including moments as a special case.

Section 4.3 is concerned with simple random samples from a finite population. The idea here is to construct statistics whose expectation under simple random sampling is just the value of the statistic computed in the whole population. The generalized k-statistics have this property but it turns out to be more convenient to work with linear combinations called 'polykays'. These are unbiased estimates of population polykays or, in the infinite population case, unbiased estimates of cumulant products. Again, they are most conveniently indexed by set partitions.

The remaining sections are concerned with estimates of cumulants and cumulant products in the presence of identifiable systematic structure, typically in the mean value. A common example is to estimate the second- and higher-order cumulants when the data are divided into k groups differing only in mean value. A second example is to estimate the cumulants of the error distribution based on residuals after linear regression. The difficulty here is that neither the raw data nor the observed residuals are identically distributed, so there is no compelling reason for restricting attention to symmetric functions in the usual sense. Indeed, better estimates can be obtained by using functions that are not symmetric: see Exercise 4.22.

4.2 k-statistics

4.2.1 *Definitions and notation*

Let $Y_1, \ldots, Y_n$ be independent and identically distributed p-dimensional random variables where Y_i has components $Y_i^1, \ldots, Y_i^p$. The cumulants and generalized cumulants of Y_i are written κ^r, $\kappa^{r,s}$, $\kappa^{r,s,t}$, $\kappa^{r,st}$ and so on. No subscripts are necessary because of the assumption that the observations are identically distributed. For each generalized cumulant, κ, with appropriate superscripts, there is a unique polynomial symmetric function, denoted by k with matching superscripts, such that k is an unbiased estimate of κ. The lower-order k-statistics are very familiar, though perhaps by different names. Thus, for example, the simplest k-statistic

$$k^r = n^{-1}\sum_i Y_i^r = \bar{Y}^r \tag{4.1}$$

is just the sample mean, an unbiased estimate of κ^r. Also,

$$\begin{aligned} k^{r,s} &= \sum_i (Y_i^r - \bar{Y}^r)(Y_i^s - \bar{Y}^s)/(n-1) \\ &= n^{-1}\phi^{ij} Y_i^r Y_j^s, \end{aligned} \tag{4.2}$$

where $\phi^{ii} = 1$ and $\phi^{ij} = -1/(n-1)$ for $i \neq j$, is just the usual sample covariance matrix. It is well known that $k^{r,s}$ is an unbiased estimate of $\kappa^{r,s}$.

We need not restrict attention to ordinary cumulants alone. A straightforward calculation shows that

$$k^{rs} = n^{-1}\sum_i Y_i^r Y_i^s \tag{4.3}$$

is an unbiased estimate of κ^{rs} and that

$$k^{r,st} = n^{-1}\sum_{ij} \phi^{ij} Y_i^r Y_j^s Y_j^t \tag{4.4}$$

is an unbiased estimate of $\kappa^{r,st}$.

The four statistics (4.1) to (4.4) are all examples of k-statistics. Following the terminology of Chapter 3, we refer to (4.1) and (4.2) as ordinary k-statistics and to (4.3) and (4.4) as generalized k-statistics. It is important at the outset to emphasize that while

$$\kappa^{rs} \equiv \kappa^{r,s} + \kappa^r \kappa^s,$$

the corresponding expression with κ replaced by k is false. In fact, we may deduce from (4.1) to (4.3) that

$$nk^{rs} = (n-1)k^{r,s} + nk^r k^s.$$

Equivalently, we may write

$$\begin{aligned} k^{rs} - k^{r,s} &= n^{-1}\sum_{ij} (\delta^{ij} - \phi^{ij}) Y_i^r Y_j^s \\ &= \sum\nolimits^{\#} Y_i^r Y_j^s / n^{(2)}, \end{aligned}$$

where $n^{(2)} = n(n-1)$, which is an unbiased estimate of the product $\kappa^r \kappa^s$, and is not the same as $k^r k^s$. The symbol, $\sum^{\#}$, which occurs frequently in the calculations that follow, denotes summation over unequal values of the indices, $i, j, \ldots$.

4.2.2 *Some general formulae for k-statistics*

It follows from the definition of moments that

$$\begin{aligned} k^r &= n^{-1}\delta^i Y_i^r \\ k^{rs} &= n^{-1}\delta^{ij} Y_i^r Y_j^s \\ k^{rst} &= n^{-1}\delta^{ijk} Y_i^r Y_j^s Y_k^t \end{aligned}$$

and so on, where $\delta^{ijk} = 1$ if $i = j = k$ and zero otherwise, are unbiased estimates of the moments κ^r, κ^{rs}, κ^{rst} and so on. To construct unbiased estimates of the ordinary cumulants, we write

$$\begin{aligned} k^{r,s} &= n^{-1}\phi^{ij} Y_i^r Y_j^s \\ k^{r,s,t} &= n^{-1}\phi^{ijk} Y_i^r Y_j^s Y_k^t \\ k^{r,s,t,u} &= n^{-1}\phi^{ijkl} Y_i^r Y_j^s Y_k^t Y_l^u \end{aligned} \tag{4.5}$$

and so on, and aim to choose the coefficients ϕ to satisfy the criterion of unbiasedness. The k-statistics are required to be symmetric in two different senses. First, they are required to be symmetric functions in the sense that they are unaffected by permuting the n observations $Y_1, \ldots, Y_n$. As a consequence, for any permutation $\pi_1, \ldots, \pi_n$ of the first n integers, it follows that

$$\phi^{ijkl} = \phi^{\pi_i \pi_j \pi_k \pi_l}.$$

This means, for example, that $\phi^{iijj} = \phi^{1122}$ but it does not follow from the above criterion that ϕ^{1221} is the same as ϕ^{1122}. In fact, however, it turns out that the coefficients ϕ are symmetric under index permutation. This follows not from the requirement that the ks be symmetric functions, but from the requirement that the k-statistics, like the corresponding cumulants, be symmetric under index permutation. In Section 4.3.2, symmetric functions will be introduced for which the coefficients are not symmetric in this second sense. It follows that ϕ^{ijk} can take on at most three distinct values depending on whether $i = j = k$, $i = j \neq k$ or all three indices are distinct. Similarly, ϕ^{ijkl} can take on at most five distinct values, namely ϕ^{1111}, ϕ^{1112}, ϕ^{1122}, ϕ^{1123} and ϕ^{1234}, corresponding to the five partitions of the *number* 4.

On taking expectations in (4.5), we find that the following identities must be satisfied by the coefficients, ϕ.

$$\begin{aligned} \kappa^{r,s} &= n^{-1}\phi^{ij}\left(\kappa^{r,s}\delta_{ij} + \kappa^r\kappa^s\delta_i\delta_j\right) \\ \kappa^{r,s,t} &= n^{-1}\phi^{ijk}\left(\kappa^{r,s,t}\delta_{ijk} + \kappa^r\kappa^{s,t}\delta_i\delta_{jk}[3] + \kappa^r\kappa^s\kappa^t\delta_i\delta_j\delta_k\right) \\ \kappa^{r,s,t,u} &= n^{-1}\phi^{ijkl}\left(\kappa^{r,s,t,u}\delta_{ijkl} + \kappa^r\kappa^{s,t,u}\delta_i\delta_{jkl}[4] + \kappa^{r,s}\kappa^{t,u}\delta_{ij}\delta_{kl}[3]\right. \\ &\qquad \left. + \kappa^r\kappa^s\kappa^{t,u}\delta_i\delta_j\delta_{kl}[6] + \kappa^r\kappa^s\kappa^t\kappa^u\delta_i\delta_j\delta_k\delta_l\right). \end{aligned}$$

Thus we must have

$$\begin{array}{llll} \phi^{ij}\delta_{ij} = n, & \phi^{ij}\delta_i = 0, & & \\ \phi^{ijk}\delta_{ijk} = n, & \phi^{ijk}\delta_{ij} = 0, & \phi^{ijk}\delta_i = 0, & \\ \phi^{ijkl}\delta_{ijkl} = n, & \phi^{ijkl}\delta_{ijk} = 0, & \phi^{ijkl}\delta_{ij} = 0, & \phi^{ijkl}\delta_i = 0, \end{array} \tag{4.6}$$

and so on. From these formulae, we find that, for i, j, k, l all distinct,

$$\begin{aligned} \phi^{ii} &= \phi^{iii} = \phi^{iiii} = 1 \\ \phi^{ij} &= \phi^{iij} = \phi^{iiij} = \phi^{iijj} = -1/(n-1) \\ \phi^{ijk} &= \phi^{iijk} = 2/\{(n-1)(n-2)\} \\ \phi^{ijkl} &= -6/\{(n-1)(n-2)(n-3)\}. \end{aligned}$$

One can then show by induction that the general expression for the coefficients ϕ is

$$(-1)^{\nu-1}\Big/\binom{n-1}{\nu-1} = (-1)^{\nu-1}(\nu-1)!/(n-1)^{(\nu-1)} \tag{4.7}$$

where $\nu \le n$ is the number of distinct indices, and

$$(n-1)^{(\nu-1)} = (n-1)(n-2)\cdots(n-\nu+1).$$

There are no unbiased estimates for cumulants of order greater than n.

Many of the pleasant statistical properties of k-statistics stem from the orthogonality of the ϕ-arrays and the δ-arrays as shown in (4.6). More generally, if υ_1, υ_2 are sets of indices, we may write $\langle \phi(\upsilon_1), \delta(\upsilon_2)\rangle$ for the sum over those indices in $\upsilon_1 \cap \upsilon_2$. This notation gives

$$\langle \phi(\upsilon_1),\, \delta(\upsilon_2)\rangle = \begin{cases} 0, & \text{if } \upsilon_2 \subset \upsilon_1; \\ n, & \text{if } \upsilon_2 = \upsilon_1; \\ \delta(\upsilon_2 - \upsilon_1), & \text{if } \upsilon_1 \subset \upsilon_2, \end{cases} \tag{4.8}$$

no simplification being possible otherwise. In the above expression, the symbol $\subset$ is to be interpreted as meaning 'proper subset of'.

For an alternative derivation of expression (4.7) for ϕ, we may proceed as follows. Unbiased estimates of the moments are given by

$$k^{rs} = n^{-1}\sum Y_i^r Y_i^s, \quad k^{rst} = n^{-1}\sum Y_i^r Y_i^s Y_i^t$$

and so on. Unbiased estimates of products of moments are given by the so-called symmetric means

$$\begin{aligned} k^{(r)(s)} &= \textstyle\sum^{\#} Y_i^r Y_j^s / n^{(2)}, & k^{(r)(st)} &= \textstyle\sum^{\#} Y_i^r Y_j^s Y_j^t / n^{(2)}, \\ k^{(rs)(tu)} &= \textstyle\sum^{\#} Y_i^r Y_i^s Y_j^t Y_j^u / n^{(2)} & k^{(r)(stu)} &= \textstyle\sum^{\#} Y_i^r Y_j^s Y_j^t Y_j^u / n^{(2)} \\ k^{(r)(s)(t)} &= \textstyle\sum^{\#} Y_i^r Y_j^s Y_k^t / n^{(3)}, & k^{(r)(st)(u)} &= \textstyle\sum^{\#} Y_i^r Y_j^s Y_j^t Y_k^u / n^{(3)} \end{aligned}$$

and so on, with summation extending over unequal subscripts. It is a straightforward exercise to verify that $E(k^{(r)(st)(u)}) = \kappa^r \kappa^{st} \kappa^u$ and similarly for the remaining statistics listed above. All linear combinations of these statistics are unbiased for the corresponding parameter. Thus

$$k^{rst} - k^{(r)(st)}[3] + 2k^{(r)(s)(t)} \tag{4.9}$$

is an unbiased estimate of $\kappa^{rst} - \kappa^r\kappa^{st}[3] + 2\kappa^r\kappa^s\kappa^t = \kappa^{r,s,t}$. By expressing (4.9) as a symmetric cubic polynomial with coefficients ϕ^{ijk}, it can be seen that the coefficients must satisfy (4.7), the numerator coming from the Möbius function and the denominator from the above sums over unequal subscripts.

A similar argument applies to higher-order k-statistics.

4.2.3 *Joint cumulants of ordinary k-statistics*

Before giving a general expression for the joint cumulants of k-statistics, it is best to examine a few simple cases. Consider first the covariance of k^r and k^s, which may be written

$$\operatorname{cov}(k^r, k^s) = n^{-2}\phi^i\phi^j\kappa_{i,j}^{r,s}$$

where $\kappa_{i,j}^{r,s}$ is the covariance of Y_i^r and Y_j^s. This covariance is zero unless $i = j$, in which case we may write $\kappa_{i,j}^{r,s} = \kappa^{r,s}\delta_{ij}$ because the random variables are assumed to be identically distributed. Thus

$$\operatorname{cov}(k^r, k^s) = n^{-2}\phi^i\phi^j\delta_{ij}\kappa^{r,s} = \kappa^{r,s}/n.$$

Similarly, for the covariance of $k^{r,s}$ and k^t, we may write

$$\mathrm{cov}(k^{r,s}, k^t) = n^{-2}\phi^{ij}\phi^k\kappa^{rs,t}_{ij,k}.$$

On expansion of the generalized cumulant using (3.3), and after taking independence into account, we find

$$\mathrm{cov}(k^{r,s}, k^t) = n^{-2}\phi^{ij}\phi^k\left(\kappa^{r,s,t}\delta_{ijk} + \kappa^s\kappa^{r,t}\delta_j\delta_{ik}[2]\right).$$

Application of (4.8) gives $\mathrm{cov}(k^{r,s}, k^t) = n^{-1}\kappa^{r,s,t}$. Similarly, the covariance of two sample covariances is

$$\begin{aligned}\mathrm{cov}(k^{r,s}, k^{t,u}) &= n^{-2}\phi^{ij}\phi^{kl}\kappa^{rs,tu}_{ij,kl}\\ &= n^{-2}\phi^{ij}\phi^{kl}\big\{\kappa^{r,s,t,u}\delta_{ijkl} + \kappa^r\kappa^{s,t,u}\delta_i\delta_{jkl}[4]\\ &\qquad + \kappa^{r,t}\kappa^{s,u}\delta_{ik}\delta_{jl}[2] + \kappa^r\kappa^t\kappa^{s,u}\delta_i\delta_l\delta_{jk}[4]\big\}\\ &= n^{-1}\kappa^{r,s,t,u} + \kappa^{r,t}\kappa^{s,u}[2]\textstyle\sum_{ij}\phi^{ij}\phi^{ij}/n^2\\ &= n^{-1}\kappa^{r,s,t,u} + \kappa^{r,t}\kappa^{s,u}[2]/(n-1).\end{aligned}$$

In the case of third and higher joint cumulants of the ks, it is convenient to introduce a new but obvious notation. For the joint cumulant of $k^{r,s}$, $k^{t,u}$, and $k^{v,w}$ we write

$$\kappa_k(r,s|t,u|v,w) = n^{-3}\phi^{ij}\phi^{kl}\phi^{mn}\kappa^{rs,tu,vw}_{ij,kl,mn}.$$

In the expansion for the 3,6 cumulant, all partitions having a unit part can be dropped because of (4.8). The third-order joint cumulant then reduces to

$$\begin{aligned}\kappa_k(r,s|t,u|v,w) = n^{-3}\phi^{ij}\phi^{kl}\phi^{mn}\big(&\kappa^{r,s,t,u,v,w}\delta_{ijklmn}\\ &+ \kappa^{r,s,t,v}\kappa^{u,w}\delta_{ijkm}\delta_{ln}[12] + \kappa^{r,s,t}\kappa^{u,v,w}\delta_{ijk}\delta_{lmn}[6]\\ &+ \kappa^{r,t,v}\kappa^{s,u,w}\delta_{ikm}\delta_{jln}[4] + \kappa^{r,t}\kappa^{s,v}\kappa^{u,w}\delta_{ik}\delta_{jm}\delta_{ln}[8]\big),\end{aligned}$$

where the third term has zero contribution on account of orthogonality. To simplify this expression further, we need to evaluate the coefficients of the cumulant products. For example, the coefficient of the final term above may be written as

$$n^{-3}\sum_{ijk}\phi^{ij}\phi^{ik}\phi^{jk}$$

and this sum, known as pattern function, has the value $(n-1)^{-2}$. See Table 4.1 under the pattern coded 12/13/23. On evaluating the remaining coefficients, we find

$$\begin{aligned}\kappa_k(r,s|t,u|v,w) &= \kappa^{r,s,t,u,v,w}/n^2 + \kappa^{r,s,t,v}\kappa^{u,w}[12]/\{n(n-1)\}\\ &\quad + \kappa^{r,t,v}\kappa^{s,u,w}[4](n-2)/\{n(n-1)^2\} + \kappa^{r,t}\kappa^{s,v}\kappa^{u,w}[8]/(n-1)^2.\end{aligned}$$

In the univariate case where k^r and $k^{r,s}$ are commonly written as $\bar{Y}$ and s^2, we may deduce from the cumulants listed above that

$$\begin{aligned}&\mathrm{var}(\bar{Y}) = \kappa_2/n, \qquad \mathrm{cov}(\bar{Y}, s^2) = \kappa_3/n,\\ &\mathrm{var}(s^2) = \kappa_4/n + 2\kappa_2^2/(n-1)\end{aligned}$$

and the third cumulant of s^2 is

$$\begin{aligned}\kappa_3(s^2) = \kappa_6/n^2 + 12\kappa_4\kappa_2/\{n(n-1)\} &+ 4(n-2)\kappa_3^2/\{n(n-1)^2\}\\ &+ 8\kappa_2^3/(n-1)^2.\end{aligned}$$

More generally, the joint cumulant of several k-statistics can be represented by a partition, say Υ^*. On the right of the expression for $\kappa_k(\Upsilon^*)$ appear cumulant products corresponding to the partitions Υ complementary to Υ^*, multiplied by a coefficient that depends on n and on the intersection matrix $\Upsilon^* \cap \Upsilon$. This coefficient is zero for all partitions Υ having a unit block and also for certain other partitions that satisfy the first condition in (4.8), possibly after simplification by the third condition in (4.8). For example, if $\Upsilon^* = \{ij|kl|mn\}$ and $\Upsilon = \{ijk|lmn\}$, then the coefficient

$$\phi^{ij}\phi^{kl}\phi^{mn}\delta_{ijk}\delta_{klm} = \phi^{kl}\phi^{mn}\delta_k\delta_{lmn} = 0$$

is zero even though no block of Υ is a subset of a block of Υ^*. In general, the coefficient of the complementary partition $\Upsilon = \{\upsilon_1, \ldots, \upsilon_\nu\}$ may be written as

$$n^{-\alpha}\langle\phi(\Upsilon^*), \delta(\Upsilon)\rangle = \sum \phi(\upsilon_1^*)\cdots\phi(\upsilon_\alpha^*)\,\delta(\upsilon_1)\cdots\delta(\upsilon_\nu)$$

with summation over all indices. With this notation, the joint cumulant of several k-statistics may be written as

$$\kappa_k(\Upsilon^*) = \sum_{\Upsilon\vee\Upsilon^*=1} n^{-\alpha}\langle\phi(\Upsilon^*), \delta(\Upsilon)\rangle\,\kappa(\upsilon_1)\cdots\kappa(\upsilon_\nu). \tag{4.10}$$

The main difficulty in using this formula lies in computing the coefficients $\langle\phi(\Upsilon^*), \delta(\Upsilon)\rangle$.

4.2.4 *Pattern matrices and pattern functions*

We now examine various ways of expressing and computing the coefficients $\langle\phi(\Upsilon^*), \delta(\Upsilon)\rangle$, also called *pattern functions*, that arise in (4.10). Evidently the coefficient is a function of n that depends on the intersection matrix $\Upsilon^* \cap \Upsilon$. In fact, since the value of ϕ depends only on the number of distinct indices and not on the number of repetitions of any index, it follows that $\langle\phi(\Upsilon^*), \delta(\Upsilon)\rangle$ must depend only on the pattern of non-zero values in $\Upsilon^* \cap \Upsilon$ and not on the actual intersection numbers. The so-called *pattern matrix* is determined only up to separate independent permutations of the rows and columns.

To take a few simple examples, suppose that $\Upsilon^* = ijk|lmn$, $\Upsilon_1 = ijl|kmn$ and $\Upsilon_2 = il|jkmn$. The intersection matrices are

$$\Upsilon^* \cap \Upsilon_1 = \begin{matrix} 2 & 1 \\ 1 & 2 \end{matrix} \quad \text{and} \quad \Upsilon^* \cap \Upsilon_2 = \begin{matrix} 1 & 2 \\ 1 & 2 \end{matrix},$$

which are different, but the pattern matrices are identical. The pattern functions $\langle\phi(\Upsilon^*), \delta(\Upsilon)\rangle$, written explicitly as sums of coefficients, are

$$\sum \phi^{iij}\phi^{ijj} \quad \text{and} \quad \sum \phi^{ijj}\phi^{ijj},$$

both of which reduce to $\sum(\phi^{ij})^2 = n^2/(n-1)$. Similarly, if we take $\Upsilon^* = ij|kl|mn$ and $\Upsilon = ik|jm|ln$, the pattern matrix is

$$\Upsilon^* \cap \Upsilon = \begin{matrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{matrix} \tag{4.11}$$

corresponding to the algebraic expression $\sum \phi^{ij}\phi^{ik}\phi^{jk}$, as if the columns of the pattern matrix were labelled i, j and k. This pattern function appears in the final term of the expression for $\kappa_k(r, s|t, u|v, w)$ above and takes the value $n^3/(n-1)^2$: see Exercise 4.6.

Because of the orthogonality of the ϕ-arrays and the δ-arrays of coefficients, many pattern functions are identically zero and it is helpful to identify these at the outset. Evidently, from (4.10), if $\Upsilon^* \vee \Upsilon < 1$, the pattern function is zero. Moreover, additional vanishing pattern functions

Table 4.1 *Some useful non-zero pattern functions*

Pattern	*Pattern function*
1/1/···/1	n
	Two rows
12/12	$n^2/(n-1)$
123/123	$n^3/(n-1)^{(2)}$
1234/1234	$n^3(n+1)/(n-1)^{(3)}$
12345/12345	$n^4(n+5)/(n-1)^{(4)}$
12356/123456	$n^3(n+1)(n^2+15n-4)/(n-1)^{(5)}$
	Three rows
12/12/12	$n^2(n-2)/(n-1)^2$
12/13/23, 123/12/13	$n^3/(n-1)^2$
123/123/12	$n^3(n-3)/\{(n-1)^2(n-2)\}$
123/123/123	$\dfrac{n^3(n^2-6n+10)}{(n-1)^2(n-2)^2}$
123/124/34, 1234/123/34	$\dfrac{n^4}{(n-1)^2(n-2)}$
123/124/134	$\dfrac{n^4(n-3)}{(n-1)^2(n-2)^2}$
123/24/1234	$\dfrac{n^4(n-4)}{(n-1)^2(n-2)^2}$
1234/1234/12	$\dfrac{n^3(n^2-4n-1)}{(n-1)^2(n-2)(n-3)}$
1234/1234/123	$\dfrac{n^3(n^3-18n^2+17n+2)}{(n-1)^2(n-2)^2(n-3)}$
1234/1234/1234	$\dfrac{n^3(n^4-12n^3+51n^2-74n-18)}{(n-1)^2(n-2)^2(n-3)^2}$

can be identified using (4.8), but this test must, in general, be applied iteratively. For example, if $\Upsilon^* = ij|kl|m$ and $\Upsilon = ijk|lm$, we find that

$$\phi^{ij}\phi^{kl}\phi^{m}\delta_{ijk}\delta_{lm} = \phi^{kl}\phi^{m}\delta_{k}\delta_{lm} = 0$$

on applying (4.8) twice. The pattern matrix is

$$\begin{matrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{matrix}$$

and the corresponding pattern function is zero because the columns of this matrix can be partitioned into two blocks that connect only through one row. The pattern matrix for $\Upsilon = ikm|jl$ cannot be partitioned in this way and the corresponding pattern function is $n^2/(n-1)$. Rows containing a single entry may be deleted since $\phi^i = 1$. Table 4.1 gives a list of some of the more useful pattern functions.

Table 4.1 *Non-zero pattern functions (continued)*

Pattern	*Pattern function*
	Three rows (contd.)
1234/125/345, 12345/215/435	$n^5/(n-1)^2(n-2)^2$
1234/1235/45, 12345/1235/34	$\dfrac{n^4(n+1)}{(n-1)^2(n-2)(n-3)}$
1234/1235/145	$\dfrac{n^4(n^2-4n-1)}{(n-1)^2(n-2)^2(n-3)}$
1234/1235/1245	$\dfrac{n^4(n^3-9n^2+19n+5)}{(n-1)^2(n-2)^2(n-3)^2}$
12345/1234/125	$\dfrac{n^4(n^2-5n-2)}{(n-1)^2(n-2)^2(n-3)}$
12345/1245/12	$\dfrac{n^4(n^2-4n-9)}{(n-1)(n-1)^{(4)}}$
1234/1256/3456	$\dfrac{n^4(n+1)(n^2-5n+2)}{(n-1)^2(n-1)^2(n-3)^2}$
12345/1236/456	$\dfrac{n^4(n^3-9n^2+19n+5)}{(n-1)^2(n-2)^2(n-3)^2}$
	Four rows
12/12/12/12	$n^2(n^2-3n+3)/(n-1)^3$
13/13/12/12	$n^3/(n-1)^2$
13/23/12/12, 123/23/12/12	$n^3(n-2)/(n-1)^3$
123/13/23/12	$n^3(n-3)/(n-1)^3$
123/123/12/12	$\dfrac{n^3(n^2-4n+5)}{(n-1)^3(n-2)}$
123/123/12/13	$\dfrac{n^3(n^2-5n+7)}{(n-1)^3(n-2)}$
123/123/123/12	$\dfrac{n^3(n-3)(n^2-4n+6)}{(n-1)^3(n-2)^2}$
123/123/123/123	$\dfrac{n^3(n^4-9n^3+33n^2-60n+48)}{(n-1)^3(n-2)^3}$

To conserve space, the patterns in Table 4.1 are coded numerically. For example, the pattern (4.11) is coded under the section marked *Three rows* as 12/13/23 and takes the value $n^3/(n-1)^2$. Patterns that can be derived from (4.11) by permuting rows or columns are not listed explicitly.

4.3 Related symmetric functions

4.3.1 *Generalized k-statistics*

Generalized k-statistics are the sample versions of the generalized cumulants, so that, for example, $k^{r,st}$ is a symmetric function and is an unbiased estimate of $\kappa^{r,st}$. A fairly simple extension of the argument used in Section 4.2.2 shows that

$$k^{r,st} = n^{-1} \sum \phi^{ij} Y_i^r Y_j^s Y_j^t \tag{4.12}$$

is the required symmetric function. Similarly,

$$k^{rs,tu} = n^{-1} \sum \phi^{ij} Y_i^r Y_i^s Y_j^t Y_j^u \tag{4.13}$$

$$k^{r,s,tu} = n^{-1} \sum \phi^{ijk} Y_i^r Y_j^s Y_k^t Y_k^u \tag{4.14}$$

and so on. Note that the coefficients ϕ^{ij} in (4.12) and (4.13) could have been replaced by ϕ^{ijj} and ϕ^{iijj}, while that in (4.14) could have been written ϕ^{ijkk}, matching the partition corresponding to the required k-statistic.

To verify that these are indeed the appropriate estimators of the generalized cumulants, we observe first that the generalized k-statistics are symmetric functions of $Y_1, \ldots, Y_n$. Also, on taking expectations, we have that

$$E\{k^{r,st}\} = n^{-1}\phi^{ij}\{\kappa^{r,st}\delta_{ij} + \kappa^r\kappa^{st}\delta_i\delta_j\} = \kappa^{r,st}.$$

Similarly for $k^{rs,tu}$ and $k^{r,s,tu}$. More generally, if we define $Y_i^{rs} = Y_i^r Y_i^s$, it is immediately evident that (4.13) is an unbiased estimate of $\operatorname{cov}(Y_i^{rs}, Y_i^{tu})$. Evidently, the sample moments are special cases of generalized k-statistics, for we may write

$$k^{rst} = n^{-1} \sum_i \phi^i Y_i^r Y_i^s Y_i^t,$$

which is the same as the expression given in Section 4.2.2. In fact, unbiased estimates exist for all generalized cumulants of order (α, β), provided only that $\alpha \le n$. In particular, unbiased estimates exist for all moments of all orders but only for ordinary cumulants whose order does not exceed n.

The importance of generalized k-statistics stems from the following properties:

(i) The generalized k-statistics are linearly independent (but not functionally independent).

(ii) Every polynomial symmetric function can be expressed uniquely as a *linear* combination of generalized k-statistics.

(iii) Any polynomial symmetric function whose expectation is independent of n can be expressed as a *linear* combination of generalized k-statistics with coefficients independent of n.

These properties are not sufficient to identify the generalized k-statistics uniquely. In fact, as will be shown in the sections that follow, there are alternative systems of symmetric functions that are in some ways more convenient than generalized k-statistics. All such systems are invertible linear functions of generalized k-statistics with coefficients independent of the sample size, and the three properties listed above are preserved under such transformations.

In view of the results derived earlier, particularly in Section 3.8, the proofs of the above assertions are fairly elementary. Only broad outline proofs are provided here. To establish (i), we suppose that there exists a linear combination of generalized k-statistics that is identically zero for all Y and show that this assumption leads to a contradiction. The expectation of such a linear combination is the same linear combination of generalized cumulants, which, by assumption, must be zero for all distributions. But this is known to be impossible because the generalized cumulants are linearly independent. Hence (i) follows.

To prove (ii), we note first that if a linear combination exists, it must be unique because the generalized k-statistics are linearly independent. For the remainder of (ii), we need to show that there are enough generalized k-statistics to span the space of polynomial symmetric functions. Without loss of generality, we may restrict attention to homogeneous symmetric functions of degree one in each of the variables $Y^1, \ldots, Y^p$. Any such polynomial may be written in the form

$$a^{i_1 i_2 \cdots i_p} Y_{i_1}^1 \cdots Y_{i_p}^p.$$

This is one of the few examples in this book of an array of coefficients that is not symmetric under index permutation. The following discussion is given for $p = 4$ but generalizes in an obvious way to arbitrary p. For $p = 4$, symmetry implies that the array a^{ijkl} can have at most $B_4 = 15$ distinct values, namely a^{1111}, $a^{1112}[4]$, $a^{1122}[3]$, $a^{1123}[6]$ and a^{1234}. To see that this is so, we note the quartic $\sum a^{ijkl} Y_i^r Y_j^s Y_k^t Y_l^u$ must be invariant under permutation of $Y_1, \ldots, Y_n$. Hence, for any $n \times n$ permutation matrix π_r^i, we must have

$$a^{ijkl} = \pi_r^i \pi_s^j \pi_t^k \pi_u^l a^{rstu} = a^{\pi_i \pi_j \pi_k \pi_l},$$

where $\pi_1, \ldots, \pi_n$ is a permutation of the first n integers. Hence, if i,j,k,l are distinct integers, then

$$a^{iiij} = a^{1112} \neq a^{2111}$$
$$a^{iijk} = a^{1123} \neq a^{1233}$$

and so on. It follows that there are exactly 15 linearly independent symmetric functions of degree one in each of four distinct variables. Any convenient basis of 15 linearly independent symmetric functions, each of degree one in the four variables, is adequate to span the required space. For example, one possibility would be to set each of the distinct as to unity in turn, the remainder being kept at zero. However, the generalized cumulants provide an alternative and more convenient basis. Of course, if the four variables were not distinct, as, for example, in univariate problems, the number of linearly independent symmetric functions of total degree four and of specified degree in each of the component variables would be reduced. In the univariate case, there are five linearly independent symmetric functions of degree four, one for each of the partitions of the *number* four.

The argument just given holds for homogeneous symmetric functions of degree one in an arbitrary number of *distinct* random variables. Even when the number of distinct variables is less than the degree of the polynomial, it is often convenient in the algebra to sustain the fiction that there are as many distinct variables as the total degree of the polynomial. Effectively, we replicate an existing variable and the algebra treats the replicate as a distinct variable. This device of algebraic pseudo-replication often simplifies the algebra and avoids the need to consider numerous special cases.

The final assertion (iii) follows from completeness of the set of generalized k-statistics together with the fact that generalized k-statistics are unbiased estimates of the corresponding cumulant.

4.3.2 *Symmetric means and poly*

The symmetric means have previously been introduced in Section 4.2.2 as the unique polynomial symmetric functions that are unbiased estimates of products of moments. Symmetric means, also called power products in the combinatorial literature dealing with the univariate case (Dressel, 1940), are most conveniently indexed in the multivariate case by a partition of a set of indices. Of course, this must be done in such a way that the notation does not give rise to confusion with generalized k-statistics, which are indexed in a similar manner. Our proposal here is to write $k^{(rs)(tuv)}$ for the estimate of the moment product $\kappa^{rs}\kappa^{tuv}$. The bracketing of superscripts is intended to suggest that some kind of product is involved. In a sense, the letter k is redundant or may be inferred from the context, and we might well write $(rs)(tuv)$ or $\langle (rs)(tuv) \rangle$ corresponding more closely to the conventions used in the univariate case where $k^{(11)(111)}$ would typically be written as 23 or $\langle 2\,3 \rangle$ (MacMahon, 1915; Dressel, 1940; Tukey, 1950, 1956a). In this chapter, we use the notations $k^{(rs)(tuv)}$ and $(rs)(tuv)$ interchangeably: the latter notation has the advantage of greater legibility.

The expressions for the symmetric means are rather simple. For instance, it is easily verified that

$$(rs)(tuv) \equiv k^{(rs)(tuv)} = \sum_{i \neq j} Y_i^r Y_i^s Y_j^t Y_j^u Y_j^v / n^{(2)}$$

and

$$(rs)(tu)(v) \equiv k^{(rs)(tu)(v)} = \sum_{i \neq j \neq k} Y_i^r Y_i^s Y_j^t Y_j^u Y_k^v / n^{(3)},$$

where the sum extends over distinct subscripts and the divisor is just the number of terms in the sum. The extension to arbitrary partitions is immediate and need not be stated explicitly. Additional symmetric means are listed in Section 4.2.2.

The polykays are the unique polynomial symmetric functions that are unbiased estimates of cumulant products. It is natural therefore to write $k^{(r,s)(t,u,v)}$ or $(r,s)(t,u,v)$ to denote that unique symmetric function whose expectation is the cumulant product $\kappa^{r,s}\kappa^{t,u,v}$. Again, the bracketing suggests multiplication, and the commas indicate that cumulant products rather than moment products are involved. We first give a few examples of polykays and then show how the three systems of symmetric functions, generalized k-statistics, symmetric means and polykays are related.

It was shown in Section 4.2.2 that

$$k^{(r)(s)} = \sum \phi^{i|j} Y_i^r Y_j^s / n^{(2)}$$

is an unbiased estimate of the product $\kappa^r\kappa^s$, where $\phi^{i|j} = 1$ if $i \neq j$ and zero otherwise. By extension,

$$k^{(r)(s,t)} = \sum \phi^{i|jk} Y_i^r Y_j^s Y_k^t / n^{(2)},$$

with suitably chosen coefficients $\phi^{i|jk}$, is an unbiased estimate of $\kappa^r\kappa^{s,t}$. The required coefficients are

$$\phi^{i|jk} = \begin{cases} 0 & \text{if } i = j \text{ or } i = k \\ 1 & \text{if } j = k \neq i \\ -1/(n-2) & \text{otherwise,} \end{cases}$$

as can be seen by writing $k^{(r)(s,t)} = k^{(r)(st)} - k^{(r)(s)(t)}$ in the form

$$\sum\nolimits^{\#} Y_i^r Y_j^s Y_j^t / n^{(2)} - \sum\nolimits^{\#} Y_i^r Y_j^s Y_k^t / n^{(3)}.$$

In addition, we may write

$$k^{(r,s)(t,u)} = \sum \phi^{ij|kl} Y_i^r Y_j^s Y_k^t Y_l^u / n^{(2)}$$

for the unbiased estimate of the product $\kappa^{r,s}\kappa^{t,u}$, where

$$\phi^{ij|kl} = \begin{cases} 0 & i \text{ or } j = k \text{ or } l \\ 1 & i = j \text{ and } k = l \neq i \\ -1/(n-2) & i = j,\ k \neq l \neq i \text{ or reverse} \\ 1/\{(n-2)(n-3)\} & \text{all distinct.} \end{cases}$$

Also,

$$k^{(r)(s)(t,u)} = \sum \phi^{i|j|kl} Y_i^r Y_j^s Y_k^t Y_l^u / n^{(3)}$$

is an unbiased estimate of the product $\kappa^r\kappa^s\kappa^{t,u}$, where the coefficients are given by

$$\phi^{i|j|kl} = \begin{cases} 0 & \text{same value occurs in different blocks} \\ 1 & k = l,\ i \neq j \neq k \\ -1/(n-3) & \text{all indices distinct.} \end{cases}$$

Evidently, the coefficients for the polykays are more complicated than those for ordinary k-statistics, symmetric means or generalized k-statistics. These complications affect the algebra but do not necessarily have much bearing on computational difficulty. Neither the formulae given above nor the corresponding ones for generalized k-statistics are suitable as a basis for computation. Computational questions are discussed in Section 4.5.

Since the generalized k-statistics, symmetric means and polykays are three classes of symmetric functions indexed in the same manner, it is hardly surprising to learn that any one of the three classes can be expressed as a linear function of any other. In fact, all of the necessary formulae have been given in Section 3.6.2: we need only make the obvious associations of symmetric means with moment products, polykays with cumulant products and generalized k-statistics with generalized cumulants. The following are a few simple examples of symmetric means expressed in terms of polykays.

$$\begin{aligned}(rs)(t) &= (r,s)(t) + (r)(s)(t) = \{(r,s) + (r)(s)\}(t)\\ (rs)(tu) &= (r,s)(t,u) + (r,s)(t)(u) + (r)(s)(t,u) + (r)(s)(t)(u)\\ &= \{(r,s) + (r)(s)\}\{(t,u) + (t)(u)\}\\ (rs)(tuv) &= \{(r,s) + (r)(s)\}\{(t,u,v) + (t)(u,v)[3] + (t)(u)(v)\}.\end{aligned}$$

Of course, $(rs) = (r,s) + (r)(s)$ and similarly

$$(tuv) = (t,u,v) + (t)(u,v)[3] + (t)(u)(v)$$

but this does not imply that $k^{(rs)(tuv)}$ is the same as $k^{(rs)}k^{(tuv)}$. The above multiplication formulae for the indices are purely symbolic and the multiplication must be performed first before the interpretation is made in terms of polykays. Thus $(rs)(tuv)$ is expressible as the sum of the 10 polykays whose indices are sub-partitions of $rs|tuv$.

The corresponding expressions for polykays in terms of symmetric means may be written symbolically as

$$\begin{aligned}(r,s)(t) &= \{(rs) - (r)(s)\}(t)\\ (r,s)(t,u) &= \{(rs) - (r)(s)\}\{(tu) - (t)(u)\}\\ (r,s)(t,u,v) &= \{(rs) - (r)(s)\}\{(tuv) - (t)(uv)[3] + 2(t)(u)(v)\}.\end{aligned} \tag{4.15}$$

Again, it is intended that the indices should be multiplied algebraically before the interpretation is made in terms of symmetric means. Thus, $(r,s)(t,u,v)$ is a linear combination of the 10 symmetric means whose indices are sub-partitions of $rs|tuv$. The coefficients in this linear combination are values of the Möbius function for the partition lattice.

From the identity connecting generalized cumulants with products of ordinary cumulants, it follows that we may express generalized k-statistics in terms of polykays using (3.3) and conversely for polykays in terms of generalized cumulants using (3.18). By way of illustration, we find using (3.3) that

$$k^{rs,tu} = k^{(r,s,t,u)} + k^{(r)(s,t,u)}[4] + k^{(r,t)(s,u)}[2] + k^{(r)(t)(s,u)}[4],$$

where $k^{(r,s,t,u)} \equiv k^{r,s,t,u}$ and the sum extends over all partitions complementary to $rs|tu$. The inverse expression giving polykays in terms of generalized cumulants is a little more complicated but fortunately it is seldom needed. Application of (3.18) gives, after some arithmetic,

$$6k^{(r,s)(t,u)} = k^{r,stu}[4] - k^{rs,tu}[3] - 2k^{r,s,tu}[2] + k^{r,t,su}[4] - k^{r,s,t,u}.$$

This identity can be read off the sixth row of the 15×15 matrix given at the end of Section 3.6.2. The remaining expressions involving four indices are

$$\begin{aligned}6k^{(r)(s,t,u)} &= -k^{r,stu}[4] + k^{rs,tu}[3] + 2k^{rs,t,u}[3] - k^{r,s,tu}[3] - 2k^{r,s,t,u}\\ 6k^{(r)(s)(t,u)} &= -k^{r,stu}[2] + 2k^{rst,u}[2] - 2k^{rs,tu} + k^{rt,su}[2]\\ &\qquad + 2k^{r,s,tu} - k^{r,t,su}[4] - k^{rs,t,u} + k^{r,s,t,u}\\ 6k^{(r)(s)(t)(u)} &= 6k^{rstu} - 2k^{r,stu}[4] - k^{rs,tu}[3] + k^{r,s,tu}[6] - k^{r,s,t,u}.\end{aligned}$$

These examples emphasize that the relationship between the polykays and the generalized k-statistics is linear, invertible and that the coefficients are independent of the sample size. Because of linearity, unbiasedness of one set automatically implies unbiasedness of the other.

4.4 Derived scalars

To each of the derived scalars discussed in Section 2.8 there corresponds a sample scalar in which the κs are replaced by ks. Denote by $k_{i,j}$ the matrix inverse of $k^{i,j}$ and write

$$\begin{aligned}
p\bar{r}_{13}^2 &= k^{r,s,t}k^{u,v,w}k_{r,s}k_{t,u}k_{v,w}\\
p\bar{r}_{23}^2 &= k^{r,s,t}k^{u,v,w}k_{r,u}k_{s,v}k_{t,w}\\
p\bar{r}_4 &= k^{r,s,t,u}k_{r,s}k_{t,u}
\end{aligned}$$

for the sample versions of $p\bar{\rho}_{13}^2$, $p\bar{\rho}_{23}^2$ and $p\bar{\rho}_4$ defined by (2.14)–(2.16). Although these three statistics are in a sense, the obvious estimators of the invariant parameters, they are not unbiased because, for example, $k_{r,s}$ is not unbiased for $\kappa_{r,s}$.

By their construction, the three statistics listed above are invariant under affine transformation of the components of X. Hence their expectations and joint cumulants must also be expressible in terms of invariants. To obtain such expressions, it is convenient to expand the matrix inverse $k_{r,s}$ in an asymptotic expansion about $\kappa_{r,s}$. If we write

$$k^{r,s} = \kappa^{r,s} + \epsilon^{r,s}$$

it follows that $\epsilon^{r,s} = O_p(n^{-1/2})$ in the sense that $n^{1/2}\epsilon^{r,s}$ has a non-degenerate limiting distribution for large n. The matrix inverse may therefore be expanded as

$$k_{r,s} = \kappa_{r,s} - \epsilon_{r,s} + \epsilon_{r,i}\epsilon_{s,j}\kappa^{i,j} - \epsilon_{r,i}\epsilon_{j,k}\epsilon_{l,s}\kappa^{i,j}\kappa^{k,l} + \cdots$$

where $\epsilon_{r,s} = \kappa_{r,i}\kappa_{s,j}\epsilon^{i,j}$ is not the matrix inverse of $\epsilon^{r,s}$. The advantage of working with this expansion is that it involves only $\epsilon^{r,s}$ whose cumulants, apart from the first, are the same as those of $k^{r,s}$. In the case of the scalar $p\bar{r}_4$, we may write

$$\begin{aligned}
p\bar{r}_4 = k^{r,s,t,u}&(\kappa_{r,s} - \epsilon_{r,s} + \epsilon_{r,i}\epsilon_{s,j}\kappa^{i,j} - \cdots)\\
&\times(\kappa_{t,u} - \epsilon_{t,u} + \epsilon_{t,i}\epsilon_{u,j}\kappa^{i,j} - \cdots).
\end{aligned}$$

On taking expectation and including terms up to order $O(n^{-1})$ only, we find using the identity (3.3) that

$$\begin{aligned}
E(p\bar{r}_4) = p\bar{\rho}_4 &- 2\kappa_k(r,s,t,t|r,s) + \kappa_k(r,s,t,u)\kappa_k(r,s|t,u)\\
&+ 2\kappa_k(r,s,t,t)\kappa_k(r,u|s,u),
\end{aligned} \tag{4.16}$$

where, for example, $\kappa_k(r,s,t,t|r,s)$ is a convenient shorthand notation for the scalar

$$\kappa_{r,v}\kappa_{s,w}\kappa_{t,u}\,\mathrm{cov}(k^{r,s,t,u}, k^{v,w})$$

and $\kappa_k(r,s,t,u) = \kappa^{r,s,t,u}$ by construction. Simplification of this particular scalar gives

$$\begin{aligned}
\kappa_{r,v}\kappa_{s,w}\kappa_{t,u}\{&\kappa^{r,s,t,u,v,w}/n + (\kappa^{r,s,t,v}\kappa^{u,w}[6] + \kappa^{r,t,u,v}\kappa^{s,w}[2])/(n-1)\\
&+ (\kappa^{r,s,v}\kappa^{t,u,w}[4] + \kappa^{r,t,w}\kappa^{s,u,v}[2])/(n-1)\}\\
&= p\bar{\rho}_6/n + (6p\bar{\rho}_4 + 2p^2\bar{\rho}_4 + 4p\bar{\rho}_{13}^2 + 2p\bar{\rho}_{23}^2)/(n-1).
\end{aligned}$$

After simplification of the remaining terms in (4.16), we are left with

$$\begin{aligned}
E(\bar{r}_4) = \bar{\rho}_4(1 - 8/n - 2p/n) &- 2\bar{\rho}_6/n - 8\bar{\rho}_{13}^2/n - 4\bar{\rho}_{23}^2/n\\
&+ 2\bar{\rho}_{14}^2/n + \bar{\rho}_{24}^2/n + O(n^{-2}),
\end{aligned}$$

where $p\bar{\rho}_{14}^2$ and $p\bar{\rho}_{24}^2$ are defined in Section 2.8.

Similar expressions may be found for higher-order cumulants, though such expressions tend to be rather lengthy especially when carried out to second order. To first order we have, for example, that

$$\begin{aligned}\operatorname{var}(p\bar{r}_4) = \kappa_k(r,r,s,s|t,t,u,u) - 4\kappa_k(r,r,s,t)\kappa_k(u,u,v,v|s,t)\\ + O(n^{-2}),\end{aligned}$$

both terms being of order $O(n^{-1})$. This variance can be expressed directly in terms of invariants but the expression is rather lengthy and complicated and involves invariants of a higher degree than those so far discussed.

In the case of jointly normal random variables, the second term above vanishes and the first reduces to

$$\operatorname{var}(p\bar{r}_4) = (8p^2 + 16p)/n + O(n^{-2}). \tag{4.17}$$

In fact, it can be shown (Exercise 4.10), that the limiting distribution of $n^{1/2}p\bar{r}_4$ is normal with zero mean and variance $8p^2 + 16p$. The normal limit is hardly surprising because all the k-statistics are asymptotically normal: the joint cumulants behave in the same way as the cumulants of a straightforward average of independent random variables.

In the case of the quadratic scalar, $\bar{r}_{13}^2$, we may make the following expansion

$$\begin{aligned}p\bar{r}_{13}^2 = k^{r,s,t}k^{u,v,w}&(k_{r,s} - \epsilon_{r,s} + \epsilon_{r,i}\epsilon_{s,j}\kappa^{i,j} - \cdots)\\ &\times(k_{t,u} - \epsilon_{t,u} + \epsilon_{t,i}\epsilon_{u,j}\kappa^{i,j} - \cdots)\\ &\times(k_{v,w} - \epsilon_{v,w} + \epsilon_{v,i}\epsilon_{w,j}\kappa^{i,j} - \cdots)\end{aligned}$$

On taking expectation, and including terms up to order $O(n^{-1})$, we find

$$\begin{aligned}E(p\bar{r}_{13}^2) = p\bar{\rho}_{13}^2 &+ \kappa_k(r,r,s|s,t,t) - 2\kappa_k(r,s,t)\kappa_k(t,u,u|r,s)\\ &- 2\kappa_k(u,v,v)\kappa_k(r,s,u|r,s) - 2\kappa_k(r,r,s)\kappa_k(u,t,t|s,u)\\ &+ 2\kappa_k(r,r,t)\kappa_k(t,v,w)\kappa_k(v,u|u,w)\\ &+ \kappa_k(r,r,t)\kappa_k(u,u,w)\kappa_k(t,s|s,w) + O(n^{-2}).\end{aligned}$$

Again, the above formula may be expressed directly in terms of invariants. For example, the second term may be written as

$$p(\bar{\rho}_6 + p\bar{\rho}_4 + 8\bar{\rho}_4 + 5\bar{\rho}_{13}^2 + 4\bar{\rho}_{23}^2 + 2p + 4)/n + O(n^{-2}).$$

In the case of normal random variables, only the second term contributes, giving

$$E(p\bar{r}_{13}^2) = 2p(p+2)/n + O(n^{-2}).$$

In fact, it may be shown (Exercise 4.11), that for large n,

$$\frac{(n-1)(n-2)p\bar{r}_{13}^2}{2n(p+2)} \quad\sim\quad \chi_p^2$$

under the assumption of normality.

Similar calculations for $\bar{r}_{23}^2$ give

$$\frac{(n-1)(n-2)p\bar{r}_{23}^2}{6n} \quad\sim\quad \chi^2_{p(p+1)(p+2)/6}$$

for large n under the assumption of normality (Exercise 4.11).

Finally, it is worth pointing out that, although $\bar{r}_{13}^2$, $\bar{r}_{23}^2$ and $\bar{r}_4$ are the most commonly used scalars for detecting multivariate non-normality, they are not the only candidates for this purpose. Other scalars that have an equal claim to be called the sample versions of $p\bar{\rho}_{13}^2$ and $p\bar{\rho}_{23}^2$ include

$$k^{(r,s,t)(u,v,w)}k_{r,s}k_{t,u}k_{v,w} \quad \text{and} \quad k^{(r,s,t)(u,v,w)}k_{r,u}k_{s,v}k_{t,w}.$$

Another class of invariant scalars that has considerable geometrical appeal despite its incompleteness, may be defined by examining the directional standardized skewness and kurtosis and choosing the directions corresponding to maxima and minima. In two dimensions, if $\bar{\rho}_{13}^2 = 0$, there are three directions of equal maximum skewness separated by $2\pi/3$, and zero skewness in the orthognal direction. On the other hand, if $4\bar{\rho}_{23}^2 = 3\bar{\rho}_{13}^2$ there is one direction of maximum skewness and zero skewness in the orthogonal direction. More generally, a complete picture of the directional skewness involves a combination of the above: see Exercises 2.36 and 2.37. In the case of the directional kurtosis, it is necessary to distinguish between maxima and minima. Machado (1976, 1983) gives approximate percentage points of the distribution under normality of the sample versions of the maximized directional skewness and kurtosis and the minimized directional kurtosis.

4.5 Computation

4.5.1 *Practical issues*

Practical considerations suggest strongly that we are seldom likely to require k-statistics or polykays of degree more than about four. Without further major assumptions, this is enough to enable the statistician to make approximate interval estimates for the first two cumulants only and to make approximate point estimates for cumulants of orders three and four. As a general rule, the higher-order k-statistics tend to have large sampling variability and consequently, large quantities of data are required to obtain estimates that are sufficiently precise to be useful. By way of example, under the optimistically favourable assumption that the data are approximately normally distributed with unit variance, approximately 40 000 observations are required to estimate the fourth cumulant accurately to one decimal place. In the case of observations distributed approximately as Poisson with unit mean, the corresponding sample size is just over half a million. More realistically, to estimate κ_4 accurately to the nearest whole number, we require 400 observations if circumstances are favourable, and more than 5000 if they are only a little less favourable. These rough calculations give some idea of the kind of precision achievable in practice.

In addition to the statistical considerations just mentioned, it should be pointed out that the number of k-statistics and, more emphatically, the number of polykays, grows rapidly with the number of variables and with the order of k-statistic or polykay considered. For example, the number of distinct k-statistics of total order k in up to p variables is $\binom{k+p-1}{k}$, while, if we include those of order less than k, the number becomes $\binom{k+p}{k}$. The latter formula includes the sample size itself as the k-statistic of order zero. On the other hand, the number of polykays of total order exactly k in up to q variables is $p_q(k)$, where, for example, $p_1(k)$ is the number of partitions of the number k and

$$p_2(k) = \sum_{\substack{a+b=k\\ a,b\geq 0}} p(a,b),$$

where $p(a,b)$ is the number of distinct partitions of a set containing a objects of type A and b objects of type B. For the purposes of this discussion, objects of the same type are assumed to be indistinguishable. In an obvious notation,

$$p_3(k) = \sum_{\substack{a+b+c=k\\ a,b,c\geq 0}} p(a,b,c)$$

Table 4.2 *Numbers of k-statistics and polykays of various orders*

Number of distinct k-statistics

	Number of variables			
Order	1	2	3	4
1	1	2	3	4
2	1	3	6	10
3	1	4	10	20
4	1	5	15	35
5	1	6	21	56
k	1	$k+1$	$\binom{k+2}{2}$	$\binom{k+3}{3}$

Number of distinct polykays

1	1	2	3	4
2	2	6	12	20
3	3	14	38	80
4	5	33	117	305
5	7	70	336	1072
k	$p_1(k)$	$p_2(k)$	$p_3(k)$	$p_4(k)$

involves a sum over various tri-partite partition numbers of three integers. Some values for these totals are given in Table 4.2.

Partly for the reasons just given, but mainly to preserve the sanity of author and reader alike, we discuss computation only for the case $k \leq 4$.

4.5.2 *From power sums to symmetric means*

As a first step in the calculation, we compute the following $\binom{p+4}{4}$ 'power sums' and 'power products': $k^0 = 1$,

$$k^r = n^{-1}\sum_i Y_i^r \qquad k^{rs} = n^{-1}\sum_i Y_i^r Y_i^s$$
$$k^{rst} = n^{-1}\sum_i Y_i^r Y_i^s Y_i^t \qquad k^{rstu} = n^{-1}\sum_i Y_i^r Y_i^s Y_i^t Y_i^u.$$

One simple way of organizing these calculations for a computer is to augment the data by adding the dummy variable $Y_i^0 = 1$ and by computing k^{rstu} for $0 \leq r \leq \cdots \leq u$. This can be accomplished in a single pass through the data. The three-index, two-index and one-index quantities can be extracted as k^{0rst}, k^{00rs} and k^{000r} as required.

The above quantities are special cases of symmetric means. All subsequent $k$statistics and polykays whose total degree does not exceed 4 may be derived from these symmetric functions. No further passes through the data matrix are required. The remaining symmetric means are $k^{(r)} = k^r$,

$$\begin{aligned} k^{(r)(s)} &= \{nk^r k^s - k^{rs}\}/(n-1) \\ k^{(r)(s)(t)} &= \{n^2 k^r k^s k^t - nk^r k^{st}[3] + 2k^{rst}\}/(n-1)^{(2)} \\ k^{(r)(s)(t)(u)} &= \{n^3 k^r k^s k^t k^u - n^2 k^r k^s k^{tu}[6] + 2nk^r k^{stu}[4] \\ &\qquad + nk^{rs}k^{tu}[3] - 6k^{rstu}\}/(n-1)^{(3)}. \end{aligned} \tag{4.18}$$

Included, essentially as special cases of the above, are the following symmetric means:

$$\begin{aligned} k^{(r)(st)} &= \left\{nk^r k^{st} - k^{rst}\right\}/(n-1) \\ k^{(rs)(tu)} &= \left\{nk^{rs} k^{tu} - k^{rstu}\right\}/(n-1) \\ k^{(r)(s)(tu)} &= \left\{n^2 k^r k^s k^{tu} - nk^r k^{stu} - nk^s k^{rtu} \right. \\ &\qquad \left. - nk^{rs} k^{tu} + 2k^{rstu}\right\}/(n-1)^{(2)}. \end{aligned} \tag{4.19}$$

Of course, if the computations are organized as suggested in the previous paragraph, then the above formulae may all be regarded as special cases of expression (4.18) for $k^{(r)(s)(t)(u)}$. By way of example, direct substitution gives

$$\begin{aligned} k^{(0)(r)(s)(t)} &= \left(n^3 k^r k^s k^t - 3n^2 k^r k^s k^t - n^2 k^r k^{st}[3] \right. \\ &\quad \left. + 2nk^{rst} + 2nk^r k^{st}[3] + nk^r k^{st}[3] - 6k^{rst}\right)/(n-1)^{(3)} \\ &= k^{(r)(s)(t)}. \end{aligned}$$

For $p = 2$, there are 55 such terms, all linearly independent but functionally dependent on the 15 basic power sums.

4.5.3 *From symmetric means to poly*

In going from symmetric means to polykays, the expressions are all linear and the coefficients are integers independent of the sample size. These properties lead to simple recognizable formulae. Some examples are as follows:

$$\begin{aligned} k^{(r,s)} &= k^{(rs)} - k^{(r)(s)} \\ k^{(r)(s,t)} &= k^{(r)(st)} - k^{(r)(s)(t)} \\ k^{(r,s,t)} &= k^{(rst)} - k^{(r)(st)}[3] + 2k^{(r)(s)(t)} \\ k^{(r,s)(t,u)} &= k^{(rs)(tu)} - k^{(r)(s)(tu)} - k^{(rs)(t)(u)} + k^{(r)(s)(t)(u)}. \end{aligned}$$

More generally, the Möbius coefficients that occur in the above formulae may be obtained using the symbolic index multiplication formulae (4.15). For example, in the final expression above, the indices and the coefficients are given by the expression

$$\{(rs) - (r)(s)\}\{(tu) - (t)(u)\}.$$

Similarly, in the expression for $k^{(r)(s,t,u)}$, the indices and the coefficients are given by

$$(r)\{(stu) - (s)(tu)[3] + 2(s)(t)(u)\}.$$

The foregoing two-stage operation, from power sums to symmetric means to polykays, produces the symmetric means as an undesired by-product of the computation. For most statistical purposes, it is the polykays that are most useful and, conceivably, there is some advantage to be gained in going from the power sums to the polykays directly. The machinery required to do this will now be described. First, we require the 15×15 upper triangular matrix $\mathbf{M} = m(\Upsilon_i, \Upsilon_j)$ whose elements are the values of the Möbius function for the lattice of partitions of four items. In addition, we require the vector $\mathbf{K}$ whose 15 elements are products of power sums.

In practice, since $\mathbf{M}$ is rather sparse, it should be possible to avoid constructing the matrix explicitly. To keep the exposition as simple as possible, however, we suppose here that $\mathbf{M}$ and $\mathbf{K}$ are constructed explicitly as shown in Table 4.3.

Table 4.3 *The Möbius matrix and the vector of power sums used in* (4.20) *to compute the polykays*

M															K
1	−1	−1	−1	−1	−1	−1	−1	2	2	2	2	2	2	−6	k^{rstu}
	1							−1	−1		−1			2	$k^{rst}k^u$
		1						−1		−1		−1		2	$k^{rsu}k^t$
			1						−1	−1			−1	2	$k^{rtu}k^s$
				1							−1	−1	−1	2	$k^{stu}k^r$
					1			−1					−1	1	$k^{rs}k^{tu}$
						1			−1			−1		1	$k^{rt}k^{su}$
							1			−1	−1			1	$k^{ru}k^{st}$
								1						−1	$k^{rs}k^t k^u$
									1					−1	$k^{rt}k^s k^u$
										1				−1	$k^{ru}k^s k^t$
											1			−1	$k^{st}k^r k^u$
												1		−1	$k^{su}k^r k^t$
													1	−1	$k^{tu}k^r k^s$
														1	$k^r k^s k^t k^u$

Evidently, (4.18) and (4.19) amount to a statement that the vector of symmetric means is given by the matrix product

$$\mathbf{D}_1\mathbf{M}^T\mathbf{D}_2\mathbf{K}$$

where

$$\mathbf{D}_1 = \operatorname{diag}\{1/n^{(\nu_j)}\}, \quad \mathbf{D}_2 = \operatorname{diag}\{n^{\nu_j}\}$$

and ν_j is the number of blocks in the jth partition. The polykays are obtained by Möbius inversion of the symmetric means, giving the vector

$$\mathbf{P} = \mathbf{M}\mathbf{D}_1\mathbf{M}^T\mathbf{D}_2\mathbf{K}. \tag{4.20}$$

By this device, the computation of polykays is reduced to a simple linear operation on matrices and vectors. Despite this algebraic simplicity, it is necessary to take care in the calculation to avoid rounding error, particularly where n is large or where some components have a mean value that is large.

The inverse relationship, giving $\mathbf{K}$ in terms of $\mathbf{P}$, though not of much interest for computation, involves a matrix having a remarkably elegant form. The (i, j) element of the inverse matrix is n raised to the power $|\Upsilon_i \vee \Upsilon_j| - |\Upsilon_i|$, where $|\Upsilon|$ is the number of blocks in the partition. See Exercise 4.18. It follows that the inverse matrix has a unit entry where $\mathbf{M}$ is non-zero and negative powers of n elsewhere.

4.6 Application to sampling

4.6.1 *Simple random sampling*

In a population containing N units or individuals, there are $\binom{N}{n}$ distinct subsets of size n. The subsets are distinct with regard to their labels though not necessarily in their values. A sample of size n chosen in such a way that each of the distinct subsets occurs with equal probability, namely $\binom{N}{n}^{-1}$, is said to be a simple random sample of size n taken without replacement from the population. In particular, each unit in the population occurs in such a sample with probability

n/N; each distinct pair occurs in the sample with probability $n(n-1)/\{N(N-1)\}$, and similarly for triplets and so on.

Suppose that on the ith unit in the sample, a p-dimensional variable $Y_i = Y_i^1, \ldots, Y_i^p$, $(i = 1, \ldots, n)$ is measured. Most commonly, $p = 1$, but it is more convenient here to keep the notation as general as possible. Let k^r be the rth component of the sample average and let K^r be the corresponding population average. Each unit in the population occurs in exactly $\binom{N-1}{n-1}$ of the $\binom{N}{n}$ distinct samples. Thus, if we denote by ave(k^r), the average value of the sample mean, averaged over all possible samples, we have that

$$\begin{aligned}\text{ave}(k^r) &= \binom{N}{n}^{-1} n^{-1} \sum_{\text{all subsets}} Y_1^r + \cdots + Y_n^r \\ &= \binom{N}{n}^{-1} n^{-1} \binom{N-1}{n-1} \{Y_1^r + \cdots + Y_N^r\} \\ &= N^{-1}\{Y_1^r + \cdots + Y_N^r\} = K^r.\end{aligned}$$

In other words, ave(k^r), averaged over all possible samples, is just the same function computed for the whole population of N units, here denoted by K^r.

The same argument immediately gives $\text{ave}(k^{rs}) = K^{rs}$, where

$$K^{rs} = N^{-1}\{Y_1^r Y_1^s + \cdots + Y_N^r Y_N^s\}$$

is a population average of products. It follows by direct analogy that $\text{ave}(k^{rst}) = K^{rst}$ and so on.

In order to show that symmetric means have the same property, namely

$$\begin{aligned}\text{ave}\{k^{(r)(s)}\} &= \sum\nolimits^{\#} Y_i^r Y_j^s / \{N(N-1)\} = K^{(r)(s)} \\ \text{ave}\{k^{(r)(st)}\} &= \sum\nolimits^{\#} Y_i^r Y_j^s Y_j^t / \{N(N-1)\} = K^{(r)(st)},\end{aligned}$$

where summation runs from 1 to N over unequal indices, we need only replace each occurrence of 'unit' in the previous argument with 'pair of distinct units' and make appropriate cosmetic changes in the formulae. For example, each pair of distinct units in the population occurs in $\binom{N-2}{n-2}$ of the $\binom{N}{n}$ distinct samples. It follows that the average value of the symmetric mean, $k^{(r)(s)}$ is

$$\begin{aligned}\text{ave}\{k^{(r)(s)}\} &= \binom{N}{n}^{-1} \frac{1}{n(n-1)} \sum_{\text{all subsets}} Y_1^r Y_2^s + \cdots + Y_{n-1}^r Y_n^s \\ &= \binom{N}{n}^{-1} \frac{1}{n(n-1)} \binom{N-2}{n-2} \sum_{\substack{1 \le i,j \le N \\ i \ne j}} Y_i^r Y_j^s \\ &= K^{(r)(s)}.\end{aligned}$$

This argument easily extends to any symmetric mean and hence to any polykay. The conclusion can therefore be summarized as follows.

Under simple random sampling, the average value of any polykay or symmetric mean, averaged over all $\binom{N}{n}$ possible samples, is just the same function computed in the population of N values.

Tukey (1950), carefully avoiding the more conventional terminology of unbiasedness, refers to this property as 'inheritance on the average'. To see that this property does not hold for arbitrary symmetric functions, see Exercise 4.13.

The main advantage in the present context of working with polykays, rather than with the less numerous power sums and products, is that variances and higher-order cumulants of computed statistics are more aesthetically appealing when expressed linearly in terms of population polykays.

4.6.2 *Joint cumulants of k-statistics*

In this section, no attempt will be made at achieving total generality. Instead, we concentrate mainly on the joint cumulants likely to be of most use in applications, usually where the degree does not exceed four. First, we give a few simple derivations of the more useful formulae. Since $\text{ave}(k^r) = K^r$, it follows that the covariance of k^r and k^s is

$$\text{ave}\{k^r k^s - K^r K^s\},$$

averaged over all possible samples. From (4.18) it follows that

$$\begin{aligned} k^r k^s &= k^{(r)(s)} + k^{(r,s)}/n \\ K^r K^s &= K^{(r)(s)} + K^{(r,s)}/N. \end{aligned}$$

Hence, since $\text{ave}(k^{(r)(s)}) = K^{(r)(s)}$, we have

$$\text{cov}(k^r, k^s) = K^{r,s}\left(\frac{1}{n} - \frac{1}{N}\right),$$

a well known result easily derived in other ways.

Similarly, the covariance of k^r and $k^{s,t}$ may be written as

$$\text{ave}\{k^r k^{s,t} - K^r K^{s,t}\}.$$

From the multiplication formula

$$\begin{aligned} k^r k^{s,t} &= k^{(r)(s,t)} + k^{r,s,t}/n \\ K^r K^{s,t} &= K^{(r)(s,t)} + K^{r,s,t}/N, \end{aligned}$$

it follows that

$$\text{cov}(k^r, k^{s,t}) = K^{r,s,t}\left(\frac{1}{n} - \frac{1}{N}\right).$$

In the case of the covariance of two sample variances or covariances, we use the multiplication formula

$$k^{r,s}k^{t,u} = k^{(r,s)(t,u)} + k^{r,s,t,u}/n + k^{(r,t)(s,u)}[2]/(n-1),$$

together with an identical expression for a product of Ks. This gives

$$\begin{aligned} \text{cov}(k^{r,s}, k^{t,u}) = K^{r,s,t,u}&\left(\frac{1}{n} - \frac{1}{N}\right) \\ &+ K^{(r,t)(s,u)}[2]\left(\frac{1}{n-1} - \frac{1}{N-1}\right), \end{aligned} \tag{4.21}$$

which should be compared with the corresponding infinite population expression

$$\kappa_k(r, s|t, u) = \kappa^{r,s,t,u}/n + \kappa^{r,t}\kappa^{s,u}[2]/(n-1).$$

Note that if (4.21) were expressed in terms of products of population k-statistics such as $K^{r,t}K^{s,u}[2]$, it would be necessary to introduce the additional product $K^{r,s}K^{t,u}$ thereby involving a partition not satisfying the connectivity condition in (3.3).

The key to the derivation of formulae such as those given above, is evidently to express multiple products of k-statistics and polykays as a *linear* combination of polykays. Formulae for multiple products of ordinary k-statistics are easy to write down, particularly with the help of the expressions in Exercises 4.5 and 4.7 for the joint cumulants of k-statistics. The following example helps to illustrate the method.

Consider, in an infinite population, the mean value of the product of three covariances. We find

$$\begin{aligned} E(k^{r,s}k^{t,u}k^{v,w}) = & \\ & \kappa^{r,s}\kappa^{t,u}\kappa^{v,w} + \kappa^{r,s}\kappa_k(t,u|v,w)[3] + \kappa_k(r,s|t,u|v,w) \\ =& \kappa^{r,s}\kappa^{t,u}\kappa^{v,w} + \kappa^{r,s}\{\kappa^{t,u,v,w}/n + \kappa^{t,v}\kappa^{u,w}[2]/(n-1)\}[3] \\ & + \kappa^{r,s,t,u,v}/n^2 + \kappa^{r,t}\kappa^{s,u,v,w}[12]/\{n(n-1)\} \\ & + \kappa^{r,t,v}\kappa^{s,u,w}[4](n-2)/\{n(n-1)^2\} \\ & + \kappa^{r,t}\kappa^{s,v}\kappa^{u,w}[8]/(n-1)^2. \end{aligned}$$

Evidently, the combination

$$\begin{aligned} & k^{(r,s)(t,u)(v,w)} + k^{(r,s)(t,u,v,w)}[3]/n + k^{(r,s)(t,v)(u,w)}[6]/(n-1) \\ & + k^{r,s,t,u,v,w}/n^2 + k^{(r,t)(s,u,v,w)}[12]/\{n(n-1)\} \\ & + k^{(r,t,v)(s,u,w)}[4](n-2)/\{n(n-1)^2\} + k^{(r,t)(s,v)(u,w)}[8]/(n-1)^2 \end{aligned} \tag{4.22}$$

is an unbiased estimate of $E(k^{r,s}k^{t,u}k^{v,w})$. It follows immediately by linear independence that (4.22) is identical to the product $k^{r,s}k^{t,u}k^{v,w}$, giving the required multiplication formula. Although it is certainly possible to write down general formulae for the product of two arbitrary polykays, such expressions tend to be rather complicated. For most purposes, a multiplication table is more useful: see Table 4.4, which gives complete products up to fourth degree and selected products up to sixth degree. In the univariate case, more extensive tables and general formulae are given by Wishart (1952), Dwyer & Tracy (1964) and Tracy (1968).

By definition, the third-order joint cumulant of $k^{r,s}$, $k^{t,u}$ and $k^{v,w}$ is equal to

$$\text{ave}(k^{r,s}k^{t,u}k^{v,w}) - K^{r,s}\,\text{ave}(k^{t,u}k^{v,w})[3] + 2K^{r,s}K^{t,u}K^{v,w}.$$

On substituting

$$\text{ave}(k^{t,u}k^{v,w}) = K^{(t,u)(v,w)} + K^{t,u,v,w}/n + K^{(t,v)(u,w)}[2]/(n-1)$$

and simplifying using (4.22), we find that the third cumulant may be written in the form

$$\begin{aligned} & \alpha_1\left(K^{(r,s)(t,u,v,w)} - K^{(r,s)}K^{(t,u,v,w)}\right)[3] \\ & + \beta_1\left(K^{(r,s)(t,v)(u,w)} - K^{(r,s)}K^{(t,v)(u,w)}\right)[6] \\ & + (n^{-2} - N^{-2})K^{r,s,t,u,v,w} + \gamma_1 K^{(r,t)(s,u,v,w)}[12] \\ & + \left(\frac{n-2}{n(n-1)^2} - \frac{N-2}{N(N-1)^2}\right)K^{(r,t,v)(s,u,w)}[4] \\ & + \left((n-1)^{-2} - (N-1)^{-2}\right)K^{(r,t)(s,v)(u,w)}[8]. \end{aligned} \tag{4.23}$$

The coefficients in this formula are given by

$$\begin{aligned} \alpha_1 &= n^{-1} - N^{-1}, & \alpha_2 &= \alpha_1 - N^{-1}, \\ \beta_1 &= (n-1)^{-1} - (N-1)^{-1}, & \beta_2 &= \beta_1 - (N-1)^{-1}, \\ \gamma_1 &= \{n(n-1)\}^{-1} - \{N(N-1)\}^{-1}, & \gamma_2 &= \gamma_1 - \{N(N-1)\}^{-1}. \end{aligned}$$

Table 4.4 *Some multiplication formulae for polykays*

Product	*Linear expression in polykays*
$k^r k^s$	$k^{(r)(s)} + k^{(r,s)}/n$
$k^r k^{s,t}$	$k^{(r)(s,t)} + k^{r,s,t}/n$
$k^r k^{(s)(t)}$	$k^{(r)(s)(t)} + \{k^{(s)(r,t)} + k^{(t)(r,s)}\}/n$
$k^r k^s k^t$	$k^{(r)(s)(t)} + k^{(r)(s,t)}[3]/n + k^{r,s,t}/n^2$
$k^r k^s k^t k^u$	$k^{(r)(s)(t)(u)} + k^{(r)(s)(t,u)}[6]/n + k^{(r,s)(t,u)}[3]/n^2$ $+ k^{(r)(s,t,u)}[4]/n^2 + k^{r,s,t,u}/n^3$
$k^r k^s k^{t,u}$	$k^{(r)(s)(t,u)} + k^{(r,s)(t,u)}/n + k^{(r)(s,t,u)}[2]/n$ $+ k^{r,s,t,u}/n^2$
$k^r k^{s,t,u}$	$k^{(r)(s,t,u)} + k^{r,s,t,u}/n$
$k^r k^s k^{(t)(u)}$	$k^{(r)(s)(t)(u)} + k^{(r)(t)(s,u)}[4]/n + k^{(t)(u)(r,s)}/n$ $+ k^{(r,t)(s,u)}[2]/n^2 + k^{(t)(r,s,u)}[2]/n^2$
$k^r k^{(s)(t)(u)}$	$k^{(r)(s)(t)(u)} + k^{(r,s)(t)(u)}[3]/n$
$k^r k^{(s)(t,u)}$	$k^{(r)(s)(t,u)} + k^{(r,s)(t,u)}/n + k^{(s)(r,t,u)}/n$
$k^{r,s} k^{(t)(u)}$	$k^{(r,s)(t)(u)} + k^{(t)(r,s,u)}[2]/n - k^{(r,t)(s,u)}[2]/\{n(n-1)\}$
$k^{(r)(s)} k^{(t)(u)}$	$k^{(r)(s)(t)(u)} + k^{(r,t)(s)(u)}[4]/n$ $+ k^{(r,t)(s,u)}[2]/\{n(n-1)\}$
$k^{r,s} k^{t,u}$	$k^{(r,s)(t,u)} + k^{r,s,t,u}/n$ $+ \{k^{(r,t)(s,u)} + k^{(r,u)(s,t)}\}/(n-1)$
$k^{r,s} k^{t,u} k^{v,w}$	$k^{(r,s)(t,u)(v,w)} + k^{(r,s)(t,u,v,w)}[3]/n$ $+ k^{(r,s)(t,v)(u,w)}[6]/(n-1) + k^{r,s,t,u,v,w}/n^2$ $+ k^{(r,t)(s,u,v,w)}[12]/\{n(n-1)\} + k^{(r,t)(s,v)(u,w)}[8]/(n-1)^2$ $+ k^{(r,t,v)(s,u,w)}[4](n-2)/\{n(n-1)^2\}$
$k^{r,s} k^{t,u,v,w}$	$k^{(r,s)(t,u,v,w)} + k^{r,s,t,u,v,w}/n$ $+ k^{(r,t)(s,u,v,w)}[8]/(n-1) + k^{(r,t,u)(s,v,w)}[6]/(n-1)$
$k^{r,s} k^{(t,u)(v,w)}$	$k^{(r,s)(t,u)(v,w)} + k^{(r,t)(s,u)(v,w)}[4]/(n-1)$ $+ k^{(t,u)(r,s,v,w)}[2]/n - k^{(r,t,u)(s,v,w)}[2]/\{n(n-1)\}$

On replacing all products by population polykays, we find the following gratifyingly simple expression involving only connecting partitions, namely

$$\begin{aligned}\mathrm{cum}(k^{r,s}, k^{t,u}, k^{v,w}) = {} & K^{r,s,t,u,v,w}\alpha_1\alpha_2 + K^{(r,t)(s,u,v,w)}[12]\alpha_1\beta_2 \\ & + K^{(r,t,v)(s,u,w)}[4]\{\alpha_1\beta_2 - \beta_1\gamma_2\} \\ & - K^{(r,s,t)(u,v,w)}[6]\alpha_1/(N-1) \\ & + K^{(r,t)(s,v)(u,w)}[8]\beta_1\beta_2. \end{aligned} \tag{4.24}$$

The simplicity of this formula, together with the corresponding one (4.21) for the covariance of $k^{r,s}$ and $k^{t,u}$, should be sufficient to justify the emphasis on polykays.

Wishart (1952), dealing with the univariate case, does not make this final step from (4.23) to

(4.24). As a result, his formulae do not bring out fully the simplicity afforded by expressing the results in terms of polykays alone.

In this final step, it was necessary to find a linear expression in polykays for the product $K^{(r,s)}K^{(t,v)(u,w)}$, which appears in (4.23). The required multiplication formula is given in Table 4.4. Notice that as $N \to \infty$, the fourth term on the right vanishes and the remaining terms converge to the corresponding expression in the infinite population cumulant $\kappa_k(r, s|t, u|v, w)$.

Table 4.5 *Joint cumulants of k-statistics for finite populations*

k-statistics	*Joint cumulant*
k^r, k^s	$\alpha_1 K^{r,s}$
$k^r, k^{s,t}$	$\alpha_1 K^{r,s,t}$
k^r, k^s, k^t	$\alpha_1\alpha_2 K^{r,s,t}$
k^r, k^s, k^t, k^u	$\alpha_1\bigl(n^{-2} - 6\alpha_1/N\bigr)K^{r,s,t,u} - 2\alpha_1^2 K^{(r,s)(t,u)}[3]/(N-1)$
$k^r, k^{s,t,u}$	$\alpha_1 K^{r,s,t,u}$
$k^{r,s}, k^{t,u}$	$\alpha_1 K^{r,s,t,u} + \beta_1 K^{(r,t)(s,u)}[2]$
$k^r, k^s, k^{t,u}$	$\alpha_1\alpha_2 K^{r,s,t,u} - \alpha_1 K^{(r,t)(s,u)}[2]/(N-1)$
$k^{r,s}, k^{t,u,v}$	$\alpha_1 K^{r,s,t,u,v} + \beta_1 K^{(r,t,u)(s,v)}[6]$
$k^r, k^{s,t}, k^{u,v,w}$	$\alpha_1\alpha_2 K^{r,s,t,u,v,w} + \alpha_1\beta_2 K^{(r,s,u)(t,v,w)}[6]$
	$+ \alpha_1\beta_2 K^{(r,s,u,v)(t,w)}[6] - \alpha_1 K^{(r,s)(t,u,v,w)}[2]/(N-1)$
	$- \alpha_1\{K^{(r,u)(s,t,v,w)}[3] + K^{(r,u,v)(s,t,w)}[3]\}/(N-1)$
$k^{r,s}, k^{t,u}, k^{v,w}$	Equation (4.24)
$k^{r,s,t}, k^{u,v,w}$	$\alpha_1 K^{r,s,t,u,v,w} + \beta_1 K^{(r,u)(s,t,v,w)}[9]$
	$+ \beta_1 K^{(r,s,u)(t,v,w)}[9]$
	$+ \{n/(n-1)^{(2)} - N/(N-1)^{(2)}\}K^{(r,u)(s,v)(t,w)}[6]$

Table 4.5 gives a selection of joint cumulants of k-statistics for finite populations, including all combinations up to fourth order and selected combinations up to sixth order.

There is, naturally, a fundamental contradiction involved in using formulae such as those in Table 4.5. The purpose of sampling is presumably to learn something about the population of values, perhaps the average value or the variance or range of values, and the first two k-statistics are useful as point estimators. However, in order to set confidence limits, it becomes necessary to know the population k-statistics – something we had hoped simple random sampling would help us avoid computing. The usual procedure is to substitute the estimated k-statistic for the population parameter and to hope that any errors so induced are negligible. Such procedures, while not entirely satisfactory, are perfectly sensible and are easily justified in large samples. However, it would be useful to have a rule-of-thumb to know roughly what extra allowance for sampling variability might be necessary in small to medium samples.

Note that if the objective is to test some fully specified hypothesis, no such contradiction arises. In this case, all population k-statistics are specified by the hypothesis and the problem is to test whether the sample k-statistics are in reasonable agreement with the known theoretical values.

4.7 k-statistics based on least squares residuals

4.7.1 *Multivariate linear regression*

In the previous sections where we dealt with independent and identically distributed observations, it was appropriate to require summary statistics to be invariant under the permutation group. The rationale for this requirement is that the joint distribution is unaffected by permuting the observations and therefore any derived statistics should be similarly unaffected. This argument leads directly to consideration of the symmetric functions. Of course, if departures from the model are suspected, it is essential to look beyond the symmetric functions in order to derive a suitable test statistic. For example, suspected serial correlation might lead the statistician to examine sample autocorrelations: suspected dependence on an auxiliary variable might lead to an examination of specific linear combinations of the response variable. Neither of these statistics is symmetric in the sense understood in Section 4.2. In this section, we consider the case where the mean value of the response is known to depend linearly on given explanatory variables. We must then abandon the notion of symmetric functions as understood in Section 4.2 and look for statistics that are invariant in a different, but more appropriate, sense.

In previous sections we wrote Y_i^r for the rth component of the ith independent observation. Subscripts were used to identify the observations, more out of convenience and aesthetic considerations, than out of principle or necessity. Only permutation transformations were considered and the k-statistics are invariant under this group. Thus, there is no compelling argument for using superscripts rather than subscripts. In this section, however, we consider arbitrary linear transformations and it is essential to recognize that Y and its cumulants are contravariant both with respect to transformation of the individual observations and with respect to the components. In other words, both indices must appear as superscripts.

Suppose then, that the rth component of the ith response variable has expectation

$$E(Y^{r;i}) = \omega^{\alpha;i}\kappa^r_\alpha,$$

where $\omega^{\alpha;i}$ is an $n \times q$ array of known constants and κ^r_α is a $q \times p$ array of unknown parameters to be estimated. When we revert to matrix notation we write

$$E(\mathbf{Y}) = \mathbf{X}\beta,$$

this being the more common notation in the literature on linear models. Here, however, we choose to write κ^r_α for the array of regression coefficients in order to emphasize the connection with cumulants and k-statistics. In addition to the above assumption regarding the mean value, we assume that there are available known arrays $\omega^{i,j}$, $\omega^{i,j,k}$ and so on, such that

$$\begin{aligned}\operatorname{cov}(Y^{r;i}, Y^{s;j}) &= \kappa^{r,s}\omega^{i,j},\\ \operatorname{cum}(Y^{r;i}, Y^{s;j}, Y^{t;k}) &= \kappa^{r,s,t}\omega^{i,j,k}\end{aligned}$$

and so on for the higher-order cumulants. The simplest non-trivial example having this structure arises when the ith observation is the sum of m_i independent and identically distributed, but unrecorded, random variables. In this case, the ω-arrays of order two and higher take the value m_i on the main diagonal and zero elsewhere.

The notation carries with it the implication that any nonsingular linear transformation

$$\bar{Y}^{r;i} = a^i_j Y^{r;j} \tag{4.25}$$

applied to Y has a corresponding effect on the arrays $\omega^{\alpha;i}$, $\omega^{i,j}$, $\omega^{i,j,k}, \ldots$, namely

$$\bar{\omega}^{\alpha;i} = a^i_j\omega^{\alpha;j}, \quad \bar{\omega}^{i,j} = a^i_k a^j_l \omega^{k,l}, \quad \bar{\omega}^{i,j,k} = a^i_l a^j_m a^k_n \omega^{l,m,n}$$

and so on. In other words, the ω-arrays are assumed to behave as contravariant tensors under the action of the general linear group (4.25).

In what follows, it is assumed that the estimation problem is entirely invariant under the group of transformations (4.25). In some ways, this may seem a very natural assumption because, for example,

$$E(\bar{Y}^{r;i}) = \bar{\omega}^{\alpha;i}\kappa^r_\alpha, \qquad \operatorname{cov}(\bar{Y}^{r;i}, \bar{Y}^{s;j}) = \bar{\omega}^{i,j}\kappa^{r,s}$$

and so on, so that the definition of the cumulants is certainly invariant under the group. However, it is important to emphasize the consequences of the assumption that *all* linear transformations of the observations are to be treated on an equal footing and that no particular scale has special status. We are forced by this assumption to abandon the concepts of interaction and replication, familiar in the analysis of variance, because they are not invariant under the group (4.25). See Exercise 4.22. This is not to be construed as criticism of the notions of interaction and replication, nor is it a criticism of our formulation in terms of the general linear group rather than some smaller group. Instead, it points to the limitations of our specification and emphasizes that, in any given application, the appropriateness of the criteria used here must be considered carefully. In simple cases, it may be possible to find a better formulation in terms of a group more carefully tailored to the problem.

Often it is appropriate to take $\omega^{i,j} = \delta^{ij}$, $\omega^{i,j,k} = \delta^{ijk}$ and so on, but, since no particular significance attaches to the diagonal arrays, we shall suppose instead that $\omega^{i,j}$, $\omega^{i,j,k}, \ldots$ are arbitrary arrays and that $\omega^{i,j}$ has full rank.

Our aim, then, is to construct unbiased estimates of κ^r_α, $\kappa^{r,s}$, $\kappa^{r,s,t}, \ldots$ that are invariant under transformations (4.25). Note that it was necessary to introduce the arrays $\omega^{\alpha;i}$, $\omega^{i,j}, \ldots$ as tensors in order that the definition of the cumulants be unaffected by linear transformation. Otherwise, if the cumulants were not invariant, it would not be sensible to require invariant estimates.

Let $\omega_{i,j}$ be the matrix inverse of $\omega^{i,j}$. In matrix notation, $\omega_{i,j}$ is written as $\mathbf{W}$. By assumption, $\omega^{i,j}$ has full rank, but even in the rank deficient case, the choice of inverse does not matter (see Exercise 2.24). Now define

$$\omega^\alpha_i = \omega_{i,j}\omega^{\alpha;j}, \qquad \omega_{i,j,k} = \omega_{i,l}\omega_{j,m}\omega_{k,n}\omega^{l,m,n}$$

and so on, lowering the indices in the usual way. In the calculations that follow, it is convenient to have a concise notation for the $q \times q$ array

$$\lambda^{\alpha,\beta} = \omega^{\alpha;i}\omega^{\beta;j}\omega_{i,j} = \omega^\alpha_i\omega^\beta_j\omega^{i,j},$$

and the $q \times q \times q$ array

$$\lambda^{\alpha,\beta,\gamma} = \omega^\alpha_i\omega^\beta_j\omega^\gamma_k\omega^{i,j,k}.$$

In matrix notation, $\lambda^{\alpha,\beta}$ is usually written as $\mathbf{X}^T\mathbf{W}\mathbf{X}$, and $\lambda_{\alpha,\beta}$ as $(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$, but there is no convenient matrix notation for $\lambda^{\alpha,\beta,\gamma}$.

A straightforward calculation shows that

$$E(\omega^\alpha_i Y^{r;i}) = \lambda^{\alpha,\beta}\kappa^r_\beta$$

and hence, by matrix inversion, that

$$k^r_\alpha = \lambda_{\alpha,\beta}\omega^\beta_i Y^{r;i} \tag{4.26}$$

is an unbiased estimate of κ^r_α, invariant under linear transformations (4.25). In matrix notation, (4.26) becomes

$$\hat{\beta} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y},$$

which is instantly recognized as the weighted least squares estimate of the regression coefficients. Note that the equally weighted estimate is not invariant under linear transformation.

The notation used in (4.26) is chosen to emphasize that the weighted least squares estimate is in fact a k-statistic of order one. Although it is in a sense the obvious invariant estimate of κ^r_α, the k-statistic, k^r_α is not unique except in rather special cases. See Exercise 4.19.

The second- and third-order cumulants of k^r_α are

$$\text{cov}(k^r_\alpha, k^s_\beta) = \lambda_{\alpha,\beta}\kappa^{r,s}$$
$$\text{cum}(k^r_\alpha, k^s_\beta, k^t_\gamma) = \lambda_{\alpha,\beta,\gamma}\kappa^{r,s,t}$$

where the indices of $\lambda^{\alpha,\beta,\gamma}$ have been lowered in the usual way by multiplying by $\lambda_{\alpha,\beta}$ three times. Similar expressions hold for the higher-order cumulants.

In the case of quadratic expressions, we write

$$E(\omega_{i,j}Y^{r;i}Y^{s;j}) = n\kappa^{r,s} + \lambda^{\alpha,\beta}\kappa^r_\alpha\kappa^s_\beta$$
$$E(\omega^\alpha_i\omega^\beta_j Y^{r;i}Y^{s;j}) = \lambda^{\alpha,\beta}\kappa^{r,s} + \lambda^{\alpha,\gamma}\lambda^{\beta,\delta}\kappa^r_\gamma\kappa^s_\delta.$$

These equations determine the k-statistic $k^{r,s}$ and the polykays $k^{(r)(s)}_{(\alpha)(\beta)}$, which are unbiased estimates of the products $\kappa^r_\alpha\kappa^s_\beta$. There are as many equations as there are k-statistics and polykays. We find that

$$(n-q)k^{r,s} = \{\omega_{i,j} - \lambda_{\alpha,\beta}\omega^\alpha_i\omega^\beta_j\}Y^{r;i}Y^{s;j},$$

which is the usual weighted residual sum of squares and products matrix, and

$$k^{(r)(s)}_{(\alpha)(\beta)} = k^r_\alpha k^s_\beta - \lambda_{\alpha,\beta}k^{r,s}.$$

These k-statistics and polykays are not unique except in very special cases: see Exercise 4.19. Among the estimates that are invariant under the general linear group, it appears that the ks do not have minimum variance unless the observations are normally distributed and, even then, additional conditions are required for exact optimality. See Exercise 4.23. Under a smaller group such as the permutation group, which admits a greater number of invariant estimates, they are neither unique nor do they have minimum variance among invariant estimates. Conditions under which $k^{r,s}$ has minimum variance in the univariate case are given by Plackett (1960, p. 40) and Atiqullah (1962). See also Exercises 4.22–4.26 for a more general discussion relating to higher-order ks.

When dealing with higher-order k-statistics, it is more convenient to work with the unstandardized residuals

$$R^{r;i} = Y^{r;i} - \omega^{\alpha;i}k^r_\alpha = \{\delta^i_j - \omega^{\alpha;i}\lambda_{\alpha,\beta}\omega^\beta_j\}Y^{r;j}$$
$$= \rho^i_j Y^{r;j}.$$

In matrix notation, this becomes

$$\mathbf{R} = (\mathbf{I} - \mathbf{H})\mathbf{Y}, \quad \text{where} \quad \mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W},$$

the projection matrix producing fitted values, is not symmetrical unless $\mathbf{W}$ is a multiple of $\mathbf{I}$. Working with residuals is entirely analogous to working with central moments, rather than moments about the origin, in the single sample case. Note, however, that polykays having a unit part are not estimable from the residuals because the distribution of $R^{r;i}$ does not depend on κ^r_α.

The least squares residuals have cumulants

$$E(R^{r;i}) = 0,$$
$$\text{cov}(R^{r;i}, R^{s;j}) = \rho^i_k\rho^j_l\omega^{k,l}\kappa^{r,s} = \rho^{i,j}\kappa^{r,s}$$
$$\text{cum}(R^{r;i}, R^{s;j}, R^{t;k}) = \rho^i_l\rho^j_m\rho^k_n\omega^{l,m,n}\kappa^{r,s,t} = \rho^{i,j,k}\kappa^{r,s,t}$$

and so on. Effectively, the arrays $\omega^{i,j}$, $\omega^{i,j,k}, \ldots$ are transformed to $\rho^{i,j}$, $\rho^{i,j,k}, \ldots$ by projection on to the residual space and its associated direct products. Note that $\omega_{i,j}$ is a generalized inverse of $\rho^{i,j}$, and hence we may write

$$\rho_{i,j} = \omega_{i,k}\omega_{j,l}\rho^{k,l}, \qquad \rho_{i,j,k} = \omega_{i,l}\omega_{j,m}\omega_{k,n}\rho^{l,m,n}$$

and so on, in agreement with the definitions given previously. Evidently, $\rho_{i,j}$ is the Moore-Penrose inverse of $\rho^{i,j}$. In the discussion that follows, we may use $\rho_{i,j}$ and $\omega_{i,j}$ interchangeably to lower indices, both being generalized inverses of $\rho^{i,j}$. Thus, the invariant unbiased estimates of the second- and third-order cumulants are

$$\begin{aligned} n_2 k^{r,s} &= \rho_{i,j} R^{r;i} R^{s;j} = \omega_{i,j} R^{r;i} R^{s;j} \\ n_3 k^{r,s,t} &= \rho_{i,j,k} R^{r;i} R^{s;j} R^{t;k} = \omega_{i,j,k} R^{r;i} R^{s;j} R^{t;k} \end{aligned} \tag{4.27}$$

where $n_2 = n - q$, $n_3 = \omega_{i,j,k}\omega^{l,m,n}\rho^i_l\rho^j_m\rho^k_n$, which reduces to $\sum_{ij}(\rho^{i,j})^3$ when $\omega^{i,j} = \delta^{ij}$ and $\omega^{i,j,k} = \delta^{i,j,k}$. In matrix notation, $n_2 k^{r,s}$ is just $\mathbf{R}^T\mathbf{W}\mathbf{R}$ or equivalently, $\mathbf{R}^T\mathbf{W}\mathbf{Y}$, but there is no simple matrix notation for $n_3 k^{r,s,t}$, the residual sum of cubes and products array.

In the case of cumulants of degree four, we write

$$\begin{aligned} E(\omega_{i,j,k,l} R^{r;i} R^{s;j} R^{t;k} R^{u;l}) &= n_4 \kappa^{r,s,t,u} + n_{22}\kappa^{r,s}\kappa^{t,u}[3] \\ E(\omega_{i,j}\omega_{k,l} R^{r;i} R^{s;j} R^{t;k} R^{u;l}) &= n_{22}\kappa^{r,s,t,u} + n_2^2\kappa^{r,s}\kappa^{t,u} \\ &\quad + n_2\kappa^{r,t}\kappa^{s,u}[2], \end{aligned}$$

where the coefficients are defined by

$$\begin{aligned} n_4 &= \rho^{i,j,k,l}\rho_{i,j,k,l} = \omega_{i,j,k,l}\rho^{i,j,k,l} \\ n_{22} &= \rho^{i,j,k,l}\rho_{i,j}\rho_{k,l} = \omega_{i,j,k,l}\rho^{i,j}\rho^{k,l}. \end{aligned}$$

In the particular case where $\omega^{i,j} = \delta^{ij}$, the coefficients reduce to $n_4 = \sum_{ij}(\rho^{i,j})^4$ and $n_{22} = \sum_i(\rho^{i,i})^2$.

Matrix inversion gives the following invariant unbiased estimates of $\kappa^{r,s,t,u}$ and $\kappa^{r,s}\kappa^{t,u}$.

$$\begin{aligned} \Delta_1 k^{r,s,t,u} &= \{n_2(n_2+2)\omega_{i,j,k,l} - n_{22}\omega_{i,j}\omega_{k,l}[3]\} R^{r;i} R^{s;j} R^{t;k} R^{u;l} \\ \Delta_2 k^{(r,s)(t,u)} &= \{-n_2(n_2-1)n_{22}\omega_{i,j,k,l} + (n_2(n_2+1)n_4 - 2n_{22}^2)\omega_{i,j}\omega_{k,l} \\ &\quad + (n_{22}^2 - n_2 n_4)\omega_{i,k}\omega_{j,l}[2]\} R^{r;i} R^{s;j} R^{t;k} R^{u;l} \end{aligned}$$

where

$$\Delta_1 = n_2 n_4(n_2 - 1) + 3(n_2 n_4 - n_{22}^2) \quad \text{and} \quad \Delta_2 = n_2(n_2-1)\Delta_1.$$

The final expression above simplifies to some extent when $r = s = t = u$, as is always the case in univariate regression problems. In particular, the final two terms can be combined and $n_2(n_2 - 1)$ may be extracted as a factor.

Note that, in the case where the observations are independent with identical cumulants of order two and higher, these k-statistics and polykays are symmetric functions of the least squares residuals, even though the residuals are neither independent nor identically distributed. It is this feature that is the source of non-uniqueness of k-statistics and polykays in regression problems.

4.7.2 *Univariate linear regression*

Some condensation of notation is inevitable in the univariate case, but here we make only the minimum of adjustment from the notation of the previous section. However, we do assume that the observations are independent with cumulants

$$\begin{aligned} \operatorname{cov}(Y^i, Y^j) &= \kappa_2 \delta^{ij} \\ \operatorname{cum}(Y^i, Y^j, Y^k) &= \kappa_3 \delta^{ijk} \end{aligned}$$

and so on. It follows that the projection matrix producing fitted values, $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, unusually for a projection matrix, is symmetrical and $\kappa_2\mathbf{H}$ is also the covariance matrix of the fitted values. The residual projection matrix is also symmetrical with elements $\rho_{i,j} = \delta_{ij} - h_{ij}$ and it is immaterial whether we use subscripts or superscripts.

If we write

$$S_r = \sum (R^i)^r$$

for the rth power sum of the residuals, it follows that

$$\begin{aligned} k_2 &= S_2/n_2 \\ k_3 &= S_3/n_3 \\ k_4 &= \{n_2(n_2+2)S_4 - 3n_{22}S_2^2\}/\Delta_1 \\ k_{22} &= \{n_4 S_2^2 - n_{22}S_4\}/\Delta_1 \end{aligned}$$

where

$$n_2 = n - q, \quad n_r = \sum_{ij} (\rho_{i,j})^r \quad \text{and} \quad n_{22} = \sum_i (\rho_{i,i})^2,$$

and Δ_1 is given at the end of the previous section. These are the unbiased estimates of residual cumulants and products of residual cumulants up to fourth order. In other words, $E(k_4) = \kappa_4$ and $E(k_{22}) = \kappa_2^2$.

Little simplification of the above formulae seems to be possible in general. However, in the case of designs that are quadratically balanced, the residuals have identical marginal variances, though they are not usually exchangeable, and $h_{ii} = q/n$, $\rho_{i,i} = 1 - q/n$. It follows then that $n_{22} = n(1-q/n)^2$. In addition, if n is large, we may wish to avoid the extensive calculations involved in computing n_3 and n_4. The following approximations are often satisfactory if n is large and if the design is not grossly unbalanced:

$$n_3 \simeq n(1-q/n)^3, \quad n_4 \simeq n(1-q/n)^4, \tag{4.28}$$

the errors of approximation being typically of order $O(n^{-1})$ and $O(n^{-2})$ respectively. Evidently, the above approximation is a lower bound for n_4, but not for n_3 The upper bounds, $n_3, n_4 \le n-q$, are attained if $n-q$ of the observations are uninformative for the regression parameters. In practice, n_3 and n_4 are typically closer to the lower limits than to the upper limits. It follows then that

$$n_2 n_4 - n_{22}^2 \simeq -q n_4$$

and hence that

$$\Delta_1 \simeq n_4(n_2(n_2-1) - 3q).$$

The above approximations are often adequate even if the design is not quadratically balanced and even if n is not particularly large. For example, in simple linear regression, with 20 observations and with a single quantitative covariate whose values are equally spaced, the error incurred in using the approximations is less than 2%. In less favourable circumstances, other approximations might be preferred. For example, for $r \geq 3$,

$$n_r \simeq \sum (\rho_{i,i})^r, \tag{4.29}$$

which reduces to the earlier approximation if the design is quadratically balanced. On the basis of limited numerical comparisons, it appears that, on balance, (4.29) is usually more accurate than (4.28), at least when q is small relative to n.

For further details, including more efficient estimates of κ_3 and κ_4, see McCullagh & Pregibon (1987) and also, Exercise 4.22.

4.8 Bibliographic notes

The estimation of means and variances has been a concern of statisticians from the earliest days, and consequently the literature is very extensive. No attempt will be made here to survey the early work concerned with sample moments and the joint moments of sample moments.

The idea of simplifying calculations by estimating cumulants rather than moments is usually attributed to Fisher (1929), although similar ideas were put forward earlier by Thiele (1897). Thiele's book was published in English in 1903 and the work was cited by Karl Pearson. It is curious that, in 1929, Fisher was apparently unaware of Thiele's earlier contribution.

Following the publication of Fisher's landmark paper, subsequent work by Wishart (Fisher & Wishart, 1931) and Kendall (1940a,b,c), seems directed at demonstrating the correctness of Fisher's combinatorial method for evaluating pattern functions. Fisher's algorithm must have appeared mystifying at the time: it has weathered the passage of half a century extraordinarily well.

Cook (1951) gives extensive tables of the joint cumulants of multivariate k-statistics. Kaplan (1952), using a form of tensor notation, gives the same formulae, but more succinctly. For further discussion, including explicit formulae in the univariate and bivariate case, see Kendall & Stuart (1977, Chapters 12,13).

The connection with finite population sampling was made by Tukey (1950, 1956a), who also bears responsibility for the term *polykay*: see Kendall & Stuart (1977, Section 12.22), where the less informative term l-statistic is used. Calculations similar to Tukey's were previously given by Dressel (1940), though without any suggestion that the results might be useful for sampling theory calculations. See also Irwin & Kendall (1943–45) For later developments, see Dwyer & Tracy (1964) and Tracy (1968).

The reader who finds the notation used in this chapter cumbersome or impenetrable should be warned that notational problems get much worse for symmetric functions derived from two-way or higher-order arrays. In a one-way layout, we talk of the *between groups variance* and the *within groups variance*. A similar decomposition exists for the higher-order statistics, so that we may talk of the *between groups skewness*, the *within groups skewness* and also the *cross skewness*. These are the three k-statistics of the third degree. In the case of a two-way design with one observation per cell, there are three variance terms and five third-order k-statistics, one of which is Tukey's (1949) non-additivity statistic. Such statistics are called *bi-kays*. There are of course, the corresponding *bi-polykays*, also derived from a two-way array, not to mention *poly-bi-kays* and *poly-poly-kays*, which are much worse. Any reader who is interested in such matters should read Hooke (1956a,b) or Tukey (1956b). For those intrepid souls whose appetites are whetted by this fare, we heartily recommend Speed (1986a,b).

Anscombe (1961) discusses the computation of and the applications of k-statistics in detecting departures from the usual linear model assumptions. There has been much subsequent work in this vein. See, for example, Bickel (1978), Hinkley (1985) or McCullagh & Pregibon (1987) for some recent developments.

4.9 Further results and exercises 4

4.1 Show that

$$k^{r,s} = n^{-1}\sum_{ij}\phi^{ij}Y_i^rY_j^s$$

is the usual sample covariance matrix based on n independent and identically distributed observations. Show that $k^{r,s}$ is unbiased for $\kappa^{r,s}$.

4.2 By applying the criteria of unbiasedness and symmetry to $k^{r,s,t}$ and $k^{r,s,t,u}$, derive the expression (4.7) for the coefficients ϕ^{ijk} and ϕ^{ijkl}.

4.3 Find the mean value of the following expressions:

$$\sum_i Y_i^rY_i^sY_i^t, \qquad \sum_{i\neq j} Y_i^rY_j^sY_j^t, \quad \text{and} \quad \sum_{i\neq j\neq k} Y_i^rY_j^sY_k^t.$$

Hence find the symmetric functions that are unbiased estimates of κ^{rst}, $\kappa^r\kappa^{st}$ and $\kappa^r\kappa^s\kappa^t$. By combining these estimates in the appropriate way, find the symmetric functions that are unbiased estimates of $\kappa^{r,s,t}$, $\kappa^{r,st}$ and $\kappa^r\kappa^{s,t}$.

4.4 Show that the expressions for ordinary k-statistics in terms of power sums are

$$\begin{aligned}
k^{r,s} &= \{k^{rs} - k^rk^s\}n/(n-1)\\
k^{r,s,t} &= \bigl\{k^{rst} - k^rk^{st}[3] + 2k^rk^sk^t\bigr\}n^2/(n-1)^{(2)}\\
k^{r,s,t,u} &= \bigl\{(n+1)k^{rstu} - (n+1)k^rk^{stu}[4] - (n-1)k^{rs}k^{tu}[3]\\
&\quad + 2nk^rk^sk^{tu}[6] - 6nk^rk^sk^tk^u\bigr\}n^2/(n-1)^{(3)},
\end{aligned}$$

Kaplan (1952).

4.5 Show that

$$\sum_{ij}(\phi^{ij})^2 = n^2/(n-1).$$

Hence derive the following joint cumulants of k-statistics:

$$\begin{aligned}
\kappa_k(r|s) &= \kappa^{r,s}/n, & \kappa_k(r|s,t) &= \kappa^{r,s,t}/n,\\
\kappa_k(r|s|t) &= \kappa^{r,s,t}/n^2, & \kappa_k(r|s,t,u) &= \kappa^{r,s,t,u}/n,\\
\kappa_k(r,s|t,u) &= \kappa^{r,s,t,u}/n + \kappa^{r,t}\kappa^{s,u}[2]/(n-1), &&\\
\kappa_k(r|s|t,u) &= \kappa^{r,s,t,u}/n^2, & \kappa_k(r|s,t,u,v) &= \kappa^{r,s,t,u,v}/n,\\
\kappa_k(r,s|t,u,v) &= \kappa^{r,s,t,u,v}/n + \kappa^{r,t,u}\kappa^{s,v}[6]/(n-1), &&\\
\kappa_k(r|s|t,u,v) &= \kappa^{r,s,t,u,v}/n^2, &&\\
\kappa_k(r|s,t|u,v) &= \kappa^{r,s,t,u,v}/n^2 + \kappa^{r,s,u}\kappa^{t,v}[4]/\{n(n-1)\} &&\\
\kappa_k(r|s|t|u,v) &= \kappa^{r,s,t,u,v}/n^3 &&
\end{aligned}$$

Show explicitly that the term $\kappa^{r,t,u}\kappa^{s,v}$ does not appear in the third cumulant, $\kappa_k(r|s|t,u,v)$.

4.6 By considering separately the five distinct index patterns, show that

$$\begin{aligned}
\sum_{ijk}\phi^{ij}\phi^{ik}\phi^{jk} &= n + \frac{3n(n-1)}{(n-1)^2} - \frac{n(n-1)(n-2)}{(n-1)^3}\\
&= n^3/(n-1)^2.
\end{aligned}$$

4.7 Using the pattern functions given in Table 4.1, derive the following joint cumulants:

$$\begin{aligned}
\kappa_k(r,s|t,u,v,w) &= \kappa^{r,s,t,u,v,w}/n + \kappa^{r,t}\kappa^{s,u,v,w}[8]/(n-1)\\
&\quad + \kappa^{r,t,u}\kappa^{s,v,w}[6]/(n-1),\\
\kappa_k(r,s|t,u|v,w) &= \kappa^{r,s,t,u,v,w}/n^2 + \kappa^{r,t}\kappa^{s,u,v,w}[12]/\{n(n-1)\}\\
&\quad + \kappa^{r,t,v}\kappa^{s,u,w}[4](n-2)/\{n(n-1)^2\}\\
&\quad + \kappa^{r,t}\kappa^{s,v}\kappa^{u,w}[8]/(n-1)^2,\\
\kappa_k(r,s,t|u,v,w) &= \kappa^{r,s,t,u,v,w}/n + \kappa^{r,u}\kappa^{s,t,v,w}[9]/(n-1)\\
&\quad + \kappa^{r,s,u}\kappa^{t,v,w}[9]/(n-1)\\
&\quad + \kappa^{r,u}\kappa^{s,v}\kappa^{t,w}[6]n/(n-1)^{(2)}
\end{aligned}$$

Kaplan (1952). Compare these formulae with those given in Table 4.5 for finite populations.

4.8 Show that joint cumulants of ordinary k-statistics in which the partition contains a unit block, e.g. $\kappa_k(r|s,t|u,v,w)$, can be found from the expression in which the unit block is deleted, by dividing by n and adding the extra index at all possible positions. Hence, using the expression given above for $\kappa_k(s,t|u,v,w)$, show that

$$\begin{aligned}
\kappa_k(r|s,t|u,v,w) &= \kappa^{r,s,t,u,v,w}/n^2 + \kappa^{r,s,u,v}\kappa^{t,w}[6]/\{n(n-1)\}\\
&\quad + \kappa^{r,s,u}\kappa^{t,v,w}[6]/\{n(n-1)\}.
\end{aligned}$$

4.9 Show that as $n \to \infty$, every array of k-statistics of fixed order has a limiting joint normal distribution when suitably standardized. Hence show in particular, that, for large n,

$$\begin{aligned}
n^{1/2}(k^r - \kappa^r) &\sim N_p(0, \kappa^{r,s})\\
n^{1/2}(k^{r,s} - \kappa^{r,s}) &\sim N_{p^2}(0, \tau^{rs,tu}),
\end{aligned}$$

where $\tau^{rs,tu} = \kappa^{r,s,t,u} + \kappa^{r,t}\kappa^{s,u}[2]$ has rank $p(p+1)/2$.

4.10 Under the assumption that the data have a joint normal distribution, show that $n^{1/2}k^{r,s,t,u}$ has a limiting covariance matrix given by

$$n\,\mathrm{cov}(k^{i,j,k,l}, k^{r,s,t,u}) \;\to\; \kappa^{i,r}\kappa^{j,s}\kappa^{k,t}\kappa^{l,u}[4!].$$

Hence show that $n^{1/2}p\bar{r}_4$ has a limiting normal distribution with mean zero and variance $8p^2+16p$.

4.11 Under the assumptions of the previous exercise, show that

$$n\ \mathrm{cov}(k^{i,j,k}, k^{r,s,t}) \;\to\; \kappa^{i,r}\kappa^{j,s}\kappa^{k,t}[3!]$$

as $n \to \infty$. Hence show that, for large n,

$$\begin{aligned}
\frac{np\bar{r}_{13}^2}{2(p+2)} &\sim \chi^2_p\\
\frac{np\bar{r}_{23}^2}{6} &\sim \chi^2_{p(p+1)(p+2)/6}.
\end{aligned}$$

4.12 By considering the symmetric group, i.e. the group comprising all $n \times n$ permutation matrices, acting on $Y_1, \ldots, Y_n$, show that every invariant polynomial function of degree k is expressible as a *linear* combination of polykays of degree k.

4.13 Show that

$$\sum_{i,j=1}^{n} Y_i^r Y_j^s = nk^{rs} + n(n-1)k^{(r)(s)}.$$

Hence deduce that, under simple random sampling, the average value over all samples of $n^{-2}\sum_{i,j=1}^{n} Y_i^r Y_j^s$ is

$$n^{-1}K^{rs} + (1-n^{-1})K^{(r)(s)} = n^{-1}K^{r,s} + K^{(r)(s)}$$

while the same function calculated in the population is

$$N^{-2}\sum_{i,j=1}^{N} Y_i^r Y_j^s = N^{-1}K^{r,s} + K^{(r)(s)}.$$

4.14 Derive the following multiplication formulae for k-statistics and polykays

$$\begin{aligned}
k^r k^s &= k^{(r)(s)} + k^{(r,s)}/n \\
k^r k^{s,t} &= k^{(r)(s,t)} + k^{r,s,t}/n \\
k^r k^{(s)(t)} &= k^{(r)(s)(t)} + \{k^{(s)(r,t)} + k^{(t)(r,s)}\}/n \\
k^r k^s k^t &= k^{(r)(s)(t)} + k^{(r)(s,t)}[3]/n + k^{r,s,t}/n^2
\end{aligned}$$

4.15 Derive the following multiplication formulae for k-statistics and polykays

$$\begin{aligned}
k^{r,s}k^{t,u} &= k^{(r,s)(t,u)} + k^{r,s,t,u}/n \\
&\quad + \{k^{(r,t)(s,u)} + k^{(r,u)(s,t)}\}/(n-1) \\
k^{r,s}k^{(t,u)(v,w)} &= k^{(r,s)(t,u)(v,w)} + k^{(r,t)(s,u)(v,w)}[4]/(n-1) \\
&\quad + k^{(t,u)(r,s,v,w)}[2]/n - k^{(r,t,u)(s,v,w)}[2]/\{n(n-1)\}
\end{aligned}$$

4.16 Using the multiplication formulae given in the previous two exercises, derive the finite population joint cumulants listed in Table 4.3.

4.17 By considering the sample covariance matrix of the observed variables and their pairwise products, show that

$$k^{rs,tu} - k^{rs,i}k^{tu,j}k_{i,j},$$

regarded as a $p^2 \times p^2$ symmetric matrix, is non-negative definite. Hence deduce that

$$k^{r,s,t,u} - k^{r,s,i}k^{t,u,j}k_{i,j} + \{k^{(r,t)(s,u)} + k^{(r,u)(s,t)}\}n/(n-2)$$

is also non-negative definite when regarded as a matrix whose rows are indexed by (r,s) and columns by (t,u). Hence derive lower bounds for $\bar{r}_4 - \bar{r}_{13}^2$ and $\bar{r}_4 - \bar{r}_{23}^2$.

4.18 By expanding the product of several power sums along the lines suggested in Section 4.6.2 for products of k-statistics, show that

$$\mathbf{K} = \mathbf{HP}$$

in the notation of (4.20), where $\mathbf{P}$ is a vector of polykays, $\mathbf{K}$ is a vector of power sum products and $\mathbf{H}$ has components

$$h_{ij} = n^l \qquad (l = |\Upsilon_i \vee \Upsilon_j| - |\Upsilon_i|),$$

where $|\Upsilon|$ denotes the number of blocks of the partition. Show also that $l \le 0$ and $l = 0$ if and only if Υ_j is a sub-partition of Υ_i. Hence prove that as $n \to \infty$, the components of $\mathbf{P}$ are the Möbius transform of the components of $\mathbf{K}$.

4.19 Show, in the notation of Section 4.7.1, that if the scalar $n_{22} = \rho_{i,j,k,l}\rho^{i,j}\rho^{k,l}$ is non-zero, then

$$n_{22}^{-1}\rho_{i,j,k,l}\rho^{k,l}R^{r;i}R^{s;j}$$

is an invariant unbiased estimate of $\kappa^{r,s}$. Under what circumstances is the above estimate the same as $k^{r,s}$?

4.20 Show that the covariance of $k^{r,s}$ and $k^{t,u}$, as defined in (4.27), is

$$\text{cov}(k^{r,s}, k^{t,u}) = n_{22}\kappa^{r,s,t,u}/n_2^2 + \kappa^{r,t}\kappa^{s,u}[2]/n_2,$$

which reduces to

$$\kappa^{r,s,t,u}/n + \kappa^{r,t}\kappa^{s,u}[2]/(n-q)$$

for quadratically balanced designs. Hence show that, in the univariate notation of Section 4.7.2,

$$n_{22}k_4/n_2^2 + 2k_{22}/n_2$$

is an unbiased estimate of $\text{var}(k_2)$.

4.21 Show that, for the ordinary linear regression problem with an intercept and one dependent variable, x, that

$$n_{22} = (n-1)(n-3)/n + \sum(x_i - \bar{x})^4 \Big/ \Big(\sum(x_i - \bar{x})^2\Big)^2$$

and that, for equally spaced x-values, this reduces to

$$n_{22} = (n-2)^2/n + 4/(5n) + O(n^{-2})$$

in reasonable agreement with the approximations of Section 4.7.2. Show more generally, that the discrepancy between n_{22} and the approximation $(n-2)^2/n$, is a function of the standardized fourth cuumulant of the x-values. Find this function explicitly.

4.22 In the notation of Section 4.7.2, show that, when third- and higher-order cumulants are neglected, the cubes of the least squares residuals have covariance matrix $\text{cov}(R^iR^jR^k,\, R^lR^mR^n)$ given by

$$\kappa_2^3\{\rho^{i,j}\rho^{k,l}\rho^{m,n}[9] + \rho^{i,l}\rho^{j,m}\rho^{k,n}[6]\},$$

here taken to be of order $n^3 \times n^3$. Show that, if $\nu = n - p$ is the rank of $\rho^{i,j}$, then

$$w_{ijk,lmn} = \rho_{i,l}\rho_{j,m}\rho_{k,n}[6]/36 - \rho_{i,j}\rho_{k,l}\rho_{m,n}[9]/\{18(\nu+4)\}$$

is the Moore-Penrose inverse matrix. Hence prove that

$$l_3 = \frac{k_3 - 3k_2\bar{R}\,n(n-p)/\{n_3(n-p+4)\}}{1 - 3\sum_{ij}\rho_{i,j}\rho_{i,i}\rho_{j,j}/\{n_3(n-p+4)\}}$$

where $n\bar{R} = \sum\rho_{i,i}R^i$, is unbiased for κ_3 and, under normality, has minimum variance among homogeneous cubic forms in the residuals, (McCullagh & Pregibon, 1987).

4.23 Deduce from the previous exercise that if the vector having components $\rho_{i,i}$ lies in the column space of the model matrix $\mathbf{X}$, then $l_3 \equiv k_3$. More generally, prove that if the constant vector lies in the column space of $\mathbf{X}$, then

$$n^{1/2}(l_3 - k_3) = O_p(n^{-1})$$

for large n under suitably mild limiting conditions on $\mathbf{X}$. Hence, deduce that k_3 is nearly optimal under normality.

4.24 Suppose $n = 4$, $p = 1$, $y = (1.2, 0.5, 1.3, 2.7)$, $x = (0, 1, 2, 3)$. Using the results in the previous two exercises, show that $k_2 = 0.5700$, $k_3 = 0.8389$, with variance $6.59\sigma^6$ under normality, and $l_3 = 0.8390$ with variance $4.55\sigma^6$ under normality. Compare with Anscombe (1961, p.15) and Pukelsheim (1980, p.110). Show also that $k_{22} = 0.6916$ and $k_4 = 0.7547$.

4.25 Repeat the calculations of Exercise 4.22, but now for l_4 and l_{22}, the optimal unbiased estimates of the fourth cumulant and the square of the second cumulant. Show that l_4 and l_{22} are linear combinations of

$$S_2^2 \quad \text{and} \quad S_4 - 6S_2 \sum \rho_i^i R_i^2/(n_2 + 6).$$

Deduce that, in the quadratically balanced case, k_4 and k_{22} are the optimal unbiased estimates. [Hint: It might be helpful to consult McCullagh & Pregibon (1987) to find the inverse of the $n^4 \times n^4$ covariance matrix of the quartic functions of the residuals.] Note that, in the case of the fourth cumulant, this calculation is different from Pukelsheim (1980, Section 4), who initially assumes κ_2 to be known and finally replaces κ_2 by k_2.

4.26 Let $\mathbf{Y}$ have components Y^i satisfying

$$E(Y^i) = \mu^i = \omega^{\alpha;i}\beta_\alpha$$

or, in matrix notation, $E(\mathbf{Y}) = \mathbf{X}\beta$, where $\mathbf{X}$ is $n \times q$ of rank q. Let $\omega^{i,j}$, $\omega^{i,j,k}, \ldots$ be given tensors such that $Y^i - \mu^i$ has cumulants $\kappa_2\omega^{i,j}$, $\kappa_3\omega^{i,j,k}$ and so on. The first-order interaction matrix, $\mathbf{X}^*$, is obtained by appending to $\mathbf{X}$, $q(q+1)/2$ additional columns having elements in the ith row given by $\omega^{i;r}\omega^{i;s}$ for $1 \le r \le s \le q$. Let $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$ and let $\mathbf{H}^*$, defined analogously, have rank $q^* < n$. Show that

$$k_2 = \mathbf{Y}^T\mathbf{W}(\mathbf{I} - \mathbf{H})\mathbf{Y}/(n - q)$$
$$k_2^* = \mathbf{Y}^T\mathbf{W}(\mathbf{I} - \mathbf{H}^*)\mathbf{Y}/(n - q^*)$$

are both

(i) unbiased for κ_2, (ii) invariant under the symmetric group (applied simultaneously to the rows of $\mathbf{Y}$ and $\mathbf{X}$),

(iii) invariant under the general linear group applied to the columns of $\mathbf{X}$ (i.e. such that the column space of $\mathbf{X}$ is preserved).

Show also that k_2 is invariant under the general linear group (4.25) applied to the rows of $\mathbf{X}$ and $\mathbf{Y}$, but that k_2^* is not so invariant.

4.27 Justify the claims made in Section 4.7.1 that interaction and replication are not invariant under the general linear group (4.25). Show that these concepts are preserved under the permutation group.

4.28 In the notation of Section 4.7.2, let

$$F_i = h_{ij}Y^j, \quad R_i = \rho_{ij}Y^j = (\delta_{ij} - h_{ij})Y^j$$

be the fitted value and the residual respectively. Define the derived statistics

$$T_1 = \sum R_j F_j^2 \quad \text{and} \quad T_2 = \sum R_j^2 F_j.$$

Show that

$$E(T_1) = \kappa_3(n_2 - 2n_{22} + n_3)$$
$$E(T_2) = \kappa_3(n_2 - n_3).$$

Show also, under the usual normal theory assumptions, that conditionally on the fitted values,

$$\text{var}(T_1) = \kappa_2 \sum_{ij} \rho_{i,j} F_i^2 F_j^2$$
$$\text{var}(T_2) = 2\kappa_2^2 \sum \rho_{i,j}^2 F_i F_j.$$

4.29 A population of size $N = N_0 + N_1$ comprises N_1 unit values and N_0 zero values. Show that the population Ks are

$$
\begin{aligned}
K_1 &= N_1/N,\\
K_2 &= N_0 N_1/N^{(2)},\\
K_3 &= N_0 N_1 (N_0 - N_1)/N^{(3)},\\
K_4 &= N_0 N_1 (N(N+1) - 6N_0 N_1)/N^{(4)}
\end{aligned}
$$

Hence derive the first four cumulants of the central hypergeometric distribution (Barton, David & Fix, 1960).

4.30 In a population consisting of the first N natural numbers, show that the population k-statistics are

$$
\begin{aligned}
K_1 &= (N+1)/2,\\
K_2 &= N(N+1)/12,\\
K_3 &= 0,\\
K_4 &= -N^2(N+1)^2/120.
\end{aligned}
$$

(Barton, David & Fix, 1960).