USE OF BIVARIATE *K*-STATISTICS IN FIXING QUALITY STANDARDS FOR MARKET MILK

By M. V. R. Sastry

Institute of Agricultural Research Statistics

1. Introduction

With a view to reviewing the existing quality standards for market milk data genuine milk samples of individual cows and buffaloes were collected from Punjab, Bihar and W. Bengal under schemes of Indian Council of Agricultural Research. The only rational approach to the problem of fixing tolerance limits for any characteristic of milk is one based on the study of the statistical distribution of that characteristic in genuine samples of milk. Appropriate Pearsonian curves were fitted to the data for each of the characteristics butter-fat, density and solids-not-fat content and tolerance limits at different levels were estimated.[1] However market milk is often not the product of a single animal but a pooled product of several animals. Standards fixed for milk from individual animals need not apply to the 'bulk' milk. In fact standards set for individual animals will be less stringent when applied to 'bulk' milk and thus there is a danger that substandard milk will be accepted more often than the proportion agreed upon. It is thus necessary to develop standards for 'bulk' milk. A method for estimating such standards from data for individual samples using bivariate *K*-statistics is given with illustrations from a part of the data collected under the Council schemes referred to above. Simple but more approximate methods for estimating the tolerance limits have also been discussed.

2. Notations

K-statistics were first introduced by R. A. Fisher in 1928. In his paper[2] properties of *K*-statistics and methods for obtaining the product moments and cumulants were given. The following notations adopted by him will be used in the present paper.

Suppose we have a set of x's $x_1 x_2 x_3 \ldots x_n$ and y's $y_1 y_2 \ldots y_n$.

Symmetrical sums

$$x_1 + x_2 + \ldots + x_n = S_{10}$$
$$y_1 + y_2 + \ldots + y_n = S_{01}$$
$$x_1y_1 + x_2y_2 + x_3y_3 + \ldots + x_ny_n = S_{11}.$$

In general,

$$x_1^p y_1^q + x_2^p y_2^q + \ldots + x_n^p y_n^q = S_{pq}.$$

K-statistics are symmetric functions of the observations and are such that

$$EK_{10} = k_{10}, \quad EK_{01} = k_{01}, \quad EK_{11} = k_{11}, \quad \text{etc.,}$$

where k_{10}, k_{01}, k_{11}, etc., are population cumulants.

e.g.,

$$K_{10} = \frac{S_{10}}{n}, \qquad K_{01} = \frac{S_{01}}{n}$$

$$K_{11} = \frac{1}{n(n-1)} (nS_{11} - S_{10}S_{01}), \text{ etc.}$$

Also

$$E(K_{11} - k_{11})(K_{10} - k_{10}) \text{ is denoted by } k\binom{11}{10}$$

$$E[(K_{ij} - k_{ij})^2 (K_{mn} - k_{mn}) (K_{pq} - k_{pq})] = k\begin{pmatrix} i\,i\,m\,p \\ j\,j\,n\,q \end{pmatrix}.$$

3. Estimation Procedure

Let x and y be two interrelated variates corresponding to milk yield and butter-fat respectively.

Let $x_1x_2x_3 \ldots x_n$ be the milk yields of cows of a herd of size n.

Let $y_1y_2y_3 \ldots y_n$ be the corresponding values of butter-fat content.

The average butter-fat content, $\bar{y}$, for the herd milk under consideration is the total fat content divided by the total milk yield. Hence

$$\bar{y} = \frac{x_1y_1 + x_2y_2 + x_3y_3 + \ldots + x_ny_n}{x_1 + x_2 + x_3 + \ldots + x_n}. \tag{3.1}$$

Equation 3.1 can be written as

$$\bar{y} = \frac{S_{11}}{S_{10}}. \tag{3.2}$$

This can further be expressed in terms of bivariate K-statistics as follows:

$$\bar{y} = \frac{S_{11}}{S_{10}} = K_{01} + \frac{(n-1)}{n} \frac{K_{11}}{K_{10}}. \tag{3.3}$$

Quality standard or tolerance limit for the characteristic butter-fat of the herd milk sample is the deviate of the distribution of $\bar{y}$ at the desired level.

Moments of $\bar{y}$ distribution can be expressed in terms of the moments and product moments of the joint distribution of x and y; assuming such moments exist. From such moments cumulants of $\bar{y}$ distribution can be estimated approximately. Finally following the procedure given by Cornish and Fisher[3] the deviates of the distribution of $\bar{y}$ can be estimated, utilizing the estimates of cumulants of $\bar{y}$.

4. Moments of $\bar{y}$-Distribution

First Raw Moment, μ_1'.—Taking expectations of the equation (3.3) μ_1' can be expressed as

$$\mu_1' = E\bar{y} = EK_{01} + \frac{(n-1)}{n} E \frac{K_{11}}{K_{10}} \tag{4.1}$$

$EK_{01} = k_{01}$ where k_{01} is to be noted as the cumulant.

$\frac{K_{11}}{K_{10}}$ can be written as

$$\frac{k_{11}}{k_{10}} \left(1 + \frac{K_{11} - k_{11}}{k_{11}}\right) \left(1 + \frac{K_{10} - k_{10}}{k_{10}}\right)^{-1} \tag{4.2}$$

where K_{11}, K_{10}, etc., are bivariate K-statistics and k_{11}, k_{10}, etc., are bivariate population cumulants. The expression (4.2) can further be expanded as

$$\frac{k_{11}}{k_{10}} \left(1 + \frac{K_{11} - k_{11}}{k_{11}}\right) \left(1 - \frac{K_{10} - k_{10}}{k_{10}} + \frac{\overline{K_{10} - k_{10}}^2}{k_{10}^2} - \frac{\overline{K_{10} - k_{10}}^3}{k_{10}^3}\right). \tag{4.3}$$

Considering terms up to cubic order in the binomial expansion.

Since

$$E(K_{11} - k_{11}) = 0,$$

$$E(K_{11} - k_{11})\ (K_{10} - k_{10}) = k\begin{pmatrix}11\\10\end{pmatrix}, \text{ etc.,}$$

$$E\left(\frac{K_{11}}{K_{10}}\right)$$

can be obtained as

$$E\left(\frac{K_{11}}{K_{10}}\right) = \frac{k_{11}}{k_{10}}\left\{1 + \frac{k\begin{pmatrix}11\\00\end{pmatrix}}{k_{10}^2} - \frac{k\begin{pmatrix}111\\000\end{pmatrix}}{k_{10}^3} - \frac{k\begin{pmatrix}11\\10\end{pmatrix}}{k_{11}k_{10}} + \frac{k\begin{pmatrix}111\\100\end{pmatrix}}{k_{11}k_{10}^2}\right\} \tag{4.4}$$

considering terms up to

$$0\left(\frac{1}{n^2}\right).$$

Values of

$$k\begin{pmatrix}11\\00\end{pmatrix},\ k\begin{pmatrix}111\\000\end{pmatrix},$$

etc., would be derived in a later section.

Thus

$$\mu_1' = k_{01} + \frac{(n-1)}{n}\frac{k_{11}}{k_{10}}\left\{1 + \frac{k\begin{pmatrix}11\\00\end{pmatrix}}{k_{10}^2} - \frac{k\begin{pmatrix}111\\000\end{pmatrix}}{k_{10}^3} - \frac{k\begin{pmatrix}11\\10\end{pmatrix}}{k_{11}k_{10}} + \frac{k\begin{pmatrix}111\\100\end{pmatrix}}{k_{11}k_{10}^2}\right\}. \tag{4.5}$$

Second Moment.—Squaring both sides of Equation 3.3 and taking expectations we get,—

$$\mu_2' = EK_{01}^2 + \frac{2(n-1)}{n} E\frac{K_{11}}{K_{10}} K_{01} + \frac{(n-1)^2}{n^2} E\left(\frac{K_{11}}{K_{10}}\right)^2$$

$$\left(\frac{K_{11}}{K_{10}}\right)^2$$

can again be written as

$$\left(\frac{k_{11}}{k_{10}}\right)^2\left(1 + \frac{K_{11} - k_{10}}{k_{10}}\right)^2\left(1 + \frac{K_{10} - k_{10}}{k_{10}}\right)^{-2} \tag{4.7}$$

or expanding the factors

$$\left(\frac{k_{11}}{k_{10}}\right)^2 \left\{1 + \frac{2K_{11} - k_{11}}{k_{11}} + \frac{(K_{11} - k_{11})^2}{k_{11}^2}\right\}$$
$$\times \left\{1 - \frac{2K_{10} - k_{10}}{k_{10}} + \frac{3\,(K_{10} - k_{10})^2}{k_{10}^2} - \frac{4\,(K_{10} - k_{10})^3}{k_{10}^3}\right\}.$$

Multiplying and taking expectations we get,

$$E\left(\frac{K_{11}}{K_{10}}\right)^2 = \left(\frac{k_{11}}{k_{10}}\right)^2 \left\{1 + \frac{3k\binom{11}{00}}{k_{10}^2} - \frac{4k\binom{111}{000}}{k_{10}^3}\right.$$
$$\left. - \frac{4k\binom{11}{10}}{k_{11}k_{10}} + \frac{6k\binom{111}{100}}{k_{11}k_{10}^2} + \frac{k\binom{11}{11}}{k_{10}^2} - \frac{2k\binom{111}{110}}{k_{11}^2k_{10}}\right\}. \tag{4.8}$$

Similarly,

$$E\frac{K_{11}}{K_{10}}K_{01} = E(K_{01} - k_{01} + k_{01})\frac{k_{11}}{k_{10}}\left(1 + \frac{K_{11} - k_{11}}{k_{11}}\right)$$
$$\times\left(1 + \frac{K_{10} - k_{10}}{k_{10}}\right)^{-1}$$
$$= \frac{k_{11}}{k_{10}}\left[-\frac{k\binom{01}{10}}{k_{10}} + \frac{k\binom{110}{001}}{k_{10}^2} + \frac{k\binom{10}{11}}{k_{11}}\right.$$
$$- \frac{k\binom{110}{101}}{k_{11}k_{10}} + k_{01}\left\{1 + \frac{k\binom{11}{00}}{k_{10}^2} - \frac{k\binom{111}{000}}{k_{10}^3}\right.$$
$$\left.\left. - \frac{k\binom{11}{10}}{k_{11}k_{10}} + \frac{k\binom{111}{100}}{k_{11}k_{10}^2}\right\}\right] \tag{4.9}$$

and

$$EK_{01}^2 = E(K_{01} - k_{01} + k_{01})^2$$
$$= k\binom{00}{11} + k_{01}^2. \tag{4.10}$$

Adding equations (4.8), (4.9) and (4.10), we get

$$\mu_2' = k_{01}^2 + k\binom{00}{11} + \frac{2(n-1)}{n}\frac{k_{11}}{k_{10}}\left[-\frac{k\binom{01}{10}}{k_{10}} + \frac{k\binom{110}{001}}{k_{10}^2}\right.$$

$$+\frac{k\binom{10}{11}}{k_{11}} - \frac{k\binom{110}{101}}{k_{11}k_{10}} + k_{01}\left\{1 + \frac{k\binom{11}{00}}{k_{10}^2}\right.$$

$$\left.\left.-\frac{k\binom{111}{000}}{k_{10}^3} - \frac{k\binom{11}{10}}{k_{11}k_{10}} + \frac{k\binom{111}{100}}{k_{11}k_{10}^2}\right\}\right]$$

$$+\frac{(n-1)^2}{n^2}\left(\frac{k_{11}}{k_{10}}\right)^2\left\{1 + \frac{3k\binom{11}{00}}{k_{10}^2} - \frac{4k\binom{111}{000}}{k_{10}^3}\right.$$

$$-\frac{4k\binom{11}{01}}{k_{11}k_{10}} + \frac{6k\binom{111}{100}}{k_{11}k_{10}^2} + \frac{k\binom{11}{11}}{k_{11}^2}$$

$$\left.-\frac{2k\binom{111}{110}}{k_{11}^2k_{10}}\right\}. \qquad (4.11)$$

Similarly μ_3' and μ_4' were also obtained.

$$\mu_3' = k\binom{000}{111} + 3k_{01}k\binom{00}{11} + k_{01}^3 + \frac{3(n-1)}{n}\frac{k_{11}}{k_{10}}$$

$$\times\left[k\binom{00}{11} - \frac{k\binom{001}{110}}{k_{10}} + \frac{k\binom{100}{111}}{k_{11}}\right.$$

$$+ k_{01}^2\left\{1 + \frac{k\binom{11}{00}}{k_{10}^2} - \frac{k\binom{111}{000}}{k_{10}^3} - \frac{k\binom{11}{10}}{k_{11}k_{10}}\right.$$

$$\left.+\frac{k\binom{111}{100}}{k_{11}k_{10}^2}\right\} + 2k_{01}\left\{-\frac{k\binom{10}{01}}{k_{10}} + \frac{k\binom{110}{001}}{k_{10}^2}\right.$$

$$\left.\left.+\frac{k\binom{10}{11}}{k_{11}} - \frac{k\binom{110}{101}}{k_{11}k_{10}}\right\}\right]$$

$$+\frac{3(n-1)^2}{n^2}\left(\frac{k_{11}}{k_{10}}\right)^2\Bigg[-\frac{2k\binom{10}{01}}{k_{10}}+\frac{3k\binom{011}{100}}{k_{10}^2}$$

$$+\frac{k\binom{110}{111}}{k_{11}^2}+\frac{2k\binom{10}{11}}{k_{11}}-\frac{4k\binom{101}{110}}{k_{11}k_{10}}$$

$$+k_{01}\Bigg\{1+\frac{3k\binom{11}{00}}{k_{10}^2}-\frac{4k\binom{111}{000}}{k_{10}^3}+\frac{k\binom{11}{11}}{k_{11}^2}$$

$$-\frac{2k\binom{111}{110}}{k_{11}^2k_{10}}-\frac{4k\binom{11}{10}}{k_{11}k_{10}}+\frac{6k\binom{111}{100}}{k_{11}k_{10}^2}\Bigg\}\Bigg]$$

$$+\frac{(n-1)^3}{n^3}\left(\frac{k_{11}}{k_{10}}\right)^3\Bigg[\Bigg[\Bigg\{1+\frac{6k\binom{11}{00}}{k_{10}^2}-\frac{10k\binom{111}{000}}{k_{10}^3}$$

$$+\frac{k\binom{111}{111}}{k_{11}^3}-\frac{9k\binom{11}{10}}{k_{11}k_{10}}+\frac{18k\binom{111}{100}}{k_{11}k_{10}^2}$$

$$+\frac{3k\binom{11}{11}}{k_{11}^2}-\frac{9k\binom{111}{110}}{k_{11}^2k_{10}}\Bigg\}\Bigg]. \qquad (4.12)$$

$$\mu_4'=4k\binom{000}{111}k_{01}+6k_{01}^2k\binom{00}{11}+k_{01}^4+\frac{4(n-1)}{n}\frac{k_{11}}{k_{10}}$$

$$\times\Bigg[k\binom{100}{111}+k_{01}^3\Bigg\{1+\frac{k\binom{11}{00}}{k_{10}^2}-\frac{k\binom{111}{000}}{k_{10}^3}$$

$$-\frac{k\binom{11}{10}}{k_{11}k_{10}}+\frac{k\binom{111}{100}}{k_{11}k_{10}^2}\Bigg\}+3k_{01}^2\Bigg\{-\frac{k\binom{01}{10}}{k_{10}}$$

$$+\frac{k\binom{011}{100}}{k_{10}^2}+\frac{k\binom{10}{11}}{k_{11}}-\frac{k\binom{011}{110}}{k_{11}k_{10}}\Bigg\}$$

$$+3k_{01}\Bigg\{-\frac{k\binom{100}{011}}{k_{10}}+k\binom{00}{11}+\frac{k\binom{100}{111}}{k_{11}}\Bigg\}\Bigg]$$

$$+\frac{6(n-1)^2}{n^2}\left(\frac{k_{11}}{k_{10}}\right)^2\Bigg[k\binom{00}{11}-\frac{2k\binom{100}{011}}{k_{10}}$$

$$+\frac{2k\binom{001}{111}}{k_{11}}+k_{01}^2\Bigg\{1+\frac{3k\binom{11}{00}}{k_{10}^2}$$

$$-\frac{4k\binom{111}{000}}{k_{10}^3}+\frac{k\binom{11}{11}}{k_{11}^2}-\frac{2k\binom{111}{110}}{k_{11}^2k_{10}}$$

$$-\frac{4k\binom{11}{10}}{k_{11}k_{10}}+\frac{6k\binom{111}{100}}{k_{11}k_{10}^2}\Bigg\}+2k_{01}\Bigg\{-\frac{2k\binom{10}{01}}{k_{10}}$$

$$+\frac{3k\binom{110}{001}}{k_{10}^2}+\frac{k\binom{110}{111}}{k_{11}^2}+\frac{2k\binom{10}{11}}{k_{11}}$$

$$-\frac{4k\binom{110}{011}}{k_{11}k_{10}}\Bigg\}\Bigg]$$

$$+\frac{4(n-1)^3}{n^3}\left(\frac{k_{11}}{k_{10}}\right)^3\Bigg[-\frac{3k\binom{10}{01}}{k_{10}}+\frac{6k\binom{110}{001}}{k_{10}^2}$$

$$+\frac{3k\binom{10}{11}}{k_{11}}+\frac{3k\binom{110}{111}}{k_{11}^2}-\frac{9k\binom{110}{101}}{k_{11}k_{10}}$$

$$+k_{01}\Bigg\{1+\frac{6k\binom{11}{00}}{k_{10}^2}-\frac{10k\binom{111}{000}}{k_{10}^3}$$

$$+\frac{k\binom{111}{111}}{k_{11}^3}+\frac{3k\binom{11}{11}}{k_{11}^2}-\frac{9k\binom{111}{110}}{k_{11}^2k_{10}}$$

$$-\frac{9k\binom{11}{10}}{k_{11}k_{10}}+\frac{18k\binom{111}{100}}{k_{11}k_{10}^2}\Bigg\}\Bigg]$$

$$+\frac{(n-1)^4}{n^4}\left(\frac{k_{11}}{k_{10}}\right)^4\Bigg[\Bigg\{1+\frac{6k\binom{11}{11}}{k_{11}^2}+\frac{4k\binom{111}{111}}{k_{11}^3}$$

$$- \frac{16k\binom{11}{10}}{k_{11}k_{10}} - \frac{24k\binom{111}{110}}{k_{10}k_{11}^{2}} + \frac{10k\binom{11}{00}}{k_{10}^{2}} + \frac{40k\binom{111}{001}}{k_{10}^{2}k_{11}} - \frac{20k\binom{111}{000}}{k_{10}^{3}}\Bigg\}\Bigg]. \qquad (4.13)$$

5.1. Methods to express functions such as

$$k\binom{11}{00}, \quad k\binom{111}{000},$$

etc., in terms of the cumulants have been given by R. A. Fisher.[2] Reference may also be made to *Advanced Theory of Statistics*, Vol. I of M. G. Kendall.[4] Some of the values which have been derived for the investigation in hand were tabulated below.

$$k\binom{11}{00} = \frac{k_{20}}{n}, \quad k\binom{10}{01} = \frac{k_{11}}{n}, \quad k\binom{00}{11} = \frac{k_{02}}{n}$$

$$k\binom{111}{000} = \frac{k_{30}}{n^2}, \quad k\binom{101}{010} = \frac{k_{21}}{n^2}, \quad k\binom{011}{101} = \frac{k_{22}}{n^2}$$

$$k\binom{111}{110} = \frac{k_{32}}{n^2} + \frac{k_{30}k_{02}}{n(n-1)} + \frac{2k_{21}k_{11}}{n(n-1)} + \frac{k_{12}k_{20}}{n(n-1)}$$

$$k\binom{110}{111} = \frac{k_{23}}{n^2} + \frac{k_{03}k_{20}}{n(n-1)} + \frac{2k_{12}k_{11}}{n(n-1)} + \frac{k_{21}k_{02}}{n(n-1)}$$

$$k\binom{111}{111} = \frac{k_{33}}{n^2} + \frac{6k_{22}k_{11}}{n(n-1)} + \frac{3k_{20}k_{13}}{n(n-1)} + \frac{3k_{31}k_{02}}{n(n-1)} + \frac{3(n-2)}{n(n-1)^2}k_{21}k_{12} + \frac{(n-2)}{n(n-1)^2}k_{30}k_{03} + \frac{2k_{11}^{2}}{(n-1)^2} + \frac{6k_{11}k_{20}k_{02}}{(n-1)^2}.$$

The last expression has also been derived by M. B. Cook.[5]

It may also be noted that for instance the value of

$$k\binom{111}{110}$$

is same as

$$k\binom{111}{101},$$

etc., since

$$k\binom{111}{110} = E(K_{11} - k_{11})(K_{11} - k_{11})(K_{10} - k_{10})$$

$$= E(K_{11} - k_{11})(K_{10} - k_{10})(K_{11} - k_{11}),$$

and similarly for others.

5.2. Values of the bivariate cumulants can be derived from those of univariate cumulants with the aid of the method of symbolic operation given by M. G. Kendall.[6] These expressions already obtained by M. B. Cook[7] but also verified are as follows:

$$k_{10} = \mu_{10}'$$

$$k_{20} = \mu'_{20} - \mu'^2_{10}$$

$$k_{30} = \mu'_{30} - 3\mu'_{20}\mu'_{10} + 2\mu'^3_{10}$$

$$k_{11} = \mu'_{11} - \mu'_{10}\mu'_{01}$$

$$k_{12} = \mu'_{12} - \mu'_{10}\mu'_{02} - 2\mu'_{11}\mu'_{01} + 2\mu'^2_{01}\mu'_{10}$$

$$k_{13} = \mu'_{13} - \mu'_{03}\mu'_{10} - 3\mu'_{12}\mu'_{01} - 3\mu'_{02}\mu'_{11} + 6\mu'_{02}\mu'_{10}\mu'_{01} + 6\mu'_{11}\mu'^2_{01} - 6\mu'^3_{01}\mu'_{10}.$$

$$k_{22} = \mu'_{22} - 2\mu'_{21}\mu'_{01} + 2\mu'_{20}\mu'^2_{01} - \mu'_{20}\mu'_{02} - 2\mu'_{12}\mu'_{10} - 2\mu'^2_{11} + 8\mu'_{11}\mu'_{10}\mu'_{01} - 6\mu'^2_{10}\mu'^2_{01} + 2\mu'^2_{10}\mu'_{02}$$

$$k_{23} = \mu'_{23} - 2\mu'_{13}\mu'_{10} + 2\mu'_{03}\mu'^2_{10} - \mu'_{03}\mu'_{20} - 3\mu'_{22}\mu'_{01} + 12\mu'_{12}\mu'_{11}\mu'_{01} - 3\mu'_{02}\mu'_{21} + 12\mu'_{02}\mu'_{11}\mu'_{10} - 18\mu'_{02}\mu'_{01}\mu'^2_{10} - 3\mu'_{02}\mu'_{21} + 12\mu'_{02}\mu'_{11}\mu'_{10} - 18\mu'_{02}\mu'_{01}\mu'^2_{10} + 6\mu'_{02}\mu'_{01}\mu'_{20} + 6\mu'_{21}\mu'^2_{01} + 12\mu'_{11}\mu'_{01} - 36\mu'_{11}\mu'^2_{01}\mu'_{10} + 24\mu'^3_{01}\mu'^2_{10} - 6\mu'^3_{01}\mu'_{20}$$

$$k_{33} = \mu'_{33} - 3\mu'_{22}\mu'_{01} + 6\mu'_{31}\mu'^2_{01} - 3\mu'_{31}\mu'_{02} - 6\mu'_{30}\mu'^3_{01} + 6\mu'_{30}\mu'_{01}\mu'_{02} - \mu'_{30}\mu'_{03} - 3\mu'_{23}\mu'_{10} - 9\mu'_{32}\mu'_{11} + 18\mu'_{22}\mu'_{10}\mu'_{01} - 9\mu'_{21}\mu'_{12} + 36\mu'_{21}\mu'_{11}\mu'_{01} - 54\mu'_{21}\mu'_{10}\mu'^2_{01} + 18\mu'_{21}\mu'_{10}\mu'_{02} - 3\mu'_{20}\mu'_{13} + 18\mu'_{21}\mu'_{11}\mu'_{01} - 54\mu'_{20}\mu'_{10}\mu'_{01}\mu'_{02} + 6\mu'_{20}\mu'_{10}\mu'_{03} + 6\mu'_{13}\mu'^2_{10} + 36\mu'_{11}\mu'_{12}\mu'_{10} - 54\mu'_{12}\mu'^2_{10}\mu'_{01} + 12\mu'^3_{11} - 108\mu'^2_{11}\mu'_{10}\mu'_{01} + 216\mu'_{11}\mu'^2_{01}\mu'^2_{01}$$

$$- 54\mu'_{11}\mu'^2_{10}\mu'_{02} - 120\mu'^3_{10}\mu'^3_{01} + 72\mu'^3_{10}\mu'_{01}\mu'_{02}$$
$$- 6\mu'^3_{10}\mu'_{03}.$$

6. With the help of the formulæ derived in Sections 5.1 and 5.2, moments of $\bar{y}$ given by equations (4.5), (4.11), etc., can be expressed in terms of the moments and product moments of the joint distribution (x, y). However cumulants of $\bar{y}$ distribution, required for the Cornish and Fisher's procedure for obtaining the deviates are tedious to be worked out algebraically. The procedure was therefore illustrated with numerical data in the following sections.

6.1. The data pertained to random samples of individual buffaloes were collected in the summer season in the city of Calcutta. x stands for milk yield and y for the butter-fat content.

Moments and product moments of the bivariate distribution (x, y) estimated from 200 samples were as follows:

$\mu'_{10} = 7 \cdot 4150$	$\mu'_{01} = 7 \cdot 0805$
$\mu'_{20} = 56 \cdot 9050$	$\mu'_{02} = 51 \cdot 3615$
$\mu'_{30} = 450 \cdot 9350$	$\mu'_{03} = 381 \cdot 9086$
$\mu'_{40} = 3679 \cdot 1050$	$\mu'_{04} = 2910 \cdot 0041$
$\mu'_{11} = 51 \cdot 3425$	$\mu'_{21} = 385 \cdot 7785$
$\mu'_{12} = 363 \cdot 9004$	$\mu'_{22} = 2673 \cdot 4224$
$\mu'_{13} = 2643 \cdot 1185$	$\mu'_{23} = 18968 \cdot 4241$
$\mu'_{31} = 2998 \cdot 4375$	$\mu'_{32} = 20344 \cdot 9938$
$\mu'_{33} = 141132 \cdot 9408.$	

6.2. Cornish and Fisher give an expression for the deviate ξ for any distribution as an adjustment to the corresponding value of x, the normal deviate for the same level of probability. This adjustment consists of terms with polynomials in x and terms with coefficients a, b, c, d, etc., some constants depending on the cumulants of the distribution under consideration such as $k_1 - m = av^{1/2}$, $k_2 - v = bv$, $k_3 = cv^{3/2}$, $k_4 = dv^2$, $k_5 = ev^{5/2}$, $k_6 = fv^{7/2}$, etc., where k_1, k_2, k_3, etc., are cumulants of the distribution under consideration and m and v are the mean and variance of the selected normal distribution. We may select m and v to be the same as those for the distribution in hand. Hence, $m = k_1$, $v = k_2$ and both a and b become zero.

The expression given by Cornish and Fisher reduces to:

$$\xi - x = \frac{1}{6} c (x^2 - 1) + \frac{1}{24} d (x^3 - 3x) - \frac{1}{36} c^2 (2x^3 - 5x)$$

$$- \frac{1}{24} cd (x^5 - 5x^2 + 2) + \frac{1}{324} c^3 (12x^4 - 53x^2 + 12)$$

$$- \frac{1}{324} d^2 (3x^5 - 24x^3 + 29x)$$

$$+ \frac{1}{288} (c^2 d) (94x^5 - 103x^3 + 107x)$$

$$- \frac{1}{7776} c^4 (25x^5 - 1688x^3 + 11511x).$$

For $p = 0 \cdot 10$, $x = 1 \cdot 28155$ and using the numerical values for the polynomials in x, given by the same authors, the expression can be written as:

$$\xi = 1 \cdot 28155 + 0 \cdot 10706c - 0 \cdot 07249d + 0 \cdot 06106c^2$$
$$+ 0 \cdot 14644cd - 0 \cdot 11629c^3 + 0.07776d^2$$
$$- 0 \cdot 10858c^2 d + 0 \cdot 09585c^4.$$

By substituting the numerical values of c and d, c^2, d^2, etc., values of ξ can be obtained for each case.

For herd size, 5, for instance

$k_1 = 6 \cdot 9547,$

$k_2 = 0 \cdot 2205,$

$k_3 = 0 \cdot 0143,$

$k_4 = 0 \cdot 1862,$

$c = 0 \cdot 1380$

and

$d = 3 \cdot 8312.$

Hence

$$\xi - x = 0 \cdot 00003 - 0 \cdot 00791 + 0 \cdot 11139 - 0 \cdot 00032$$
$$+ 0 \cdot 07742 + 0 \cdot 00116 - 0 \cdot 27772 + 0 \cdot 01477$$

where

$x = 1 \cdot 28155$

for

$$p = 0{\cdot}10.$$

Thus

$$\xi = 1{\cdot}20040.$$

Tolerance limit at 10% level $m - \xi v^{1/2}$ is then

$$6{\cdot}95475 - 0{\cdot}56418 = 6{\cdot}39057.$$

This limit is denoted by T_1.

6·3. As can be seen from the illustrations the calculations involved in the estimation of the tolerance limit are extremely cumbersome. The possibility of using simplified approximations was accordingly explored. In the first instance, the nature of the distribution of $\bar{y}$ for various values of herd sizes was examined. In Table I values of the cumulants of $\bar{y}$ were given for herd sizes 5, 10, 25 and 50.

TABLE I

Cumulants of $\bar{y}$ distribution

n	k_1	k_2	k_3	k_4
5	6.9547	0.2205	0.0143	0.1862
10	6.9392	0.1106	0.0058	0.0486
25	6.9302	0.0435	0.0029	0.0037
50	6.9271	0.0218	0.0010	0.0012

It can be seen from Table I that the third and fourth cumulants are small and decrease repidly as the herd size increases. The normality of $\bar{y}$ distribution was tested by means of Fisher's g_1 and g_2 statistics. This was shown in Table II.

6.4. It is clear from Table II that $\bar{y}$ distribution can be expected to be normal from herd size 25 onwards. In calculating limits or when one is interested in limits for a large herd sample, normal approximation for $\bar{y}$ distribution can be used, thereby neglecting the third and fourth moments of $\bar{y}$ distribution and thus simplifying the method. This limit obtained on the assumption of normality is designated T_2.

TABLE II

n	g_1	s.e. (g_1)	g_2	s.e. (g_2)
5	0·14	0.91	8·6**	2·00
10	0.14	0.68	3·9**	1·33
25	0.11	0.46	0·3	0·90
50	0.03	0.33	0·1	0·16

6.5. From a large number of distributions studied under the Council schemes, it has been observed that the marginal distributions of milk yield and butter-fat are non-normal. For instance, in the example considered in Section 6.1, we find:

$$\mu'_{10} = 7\cdot4150,$$
$$\mu'_{20} = 56\cdot9050,$$
$$\mu'_{30} = 450\cdot9350,$$
$$\mu'_{40} = 3679\cdot1050$$

and

$$\mu'_{01} = 7\cdot0805,$$
$$\mu'_{02} = 51\cdot3613,$$
$$\mu'_{03} = 381\cdot9086,$$
$$\mu'_{04} = 2910\cdot8041$$

to be the first four moments of the distribution of x and y respectively.

For distribution of x,

$$\beta_1 = 0\cdot0310;$$
$$\beta_2 = 2\cdot1137$$

and for the distribution of y,

$$\beta_1 = 0\cdot3976;$$
$$\beta_2 = 2\cdot5533.$$

From the K-criteria, we can observe both the distributions to be Pearson's Type I. Thus the joint distribution of x and y is non-normal.

If the cumulants of order, higher than two, could be ignored (*i.e.*, $k_{r,s} = 0$ for all $r + s > 2$), the four moments of $\bar{y}$ distribution can further be simplified. With the aid of these four moments of $\bar{y}$ distribution, tolerance limits can be worked out in the same way as in the earlier case. These limits were denoted by T_3 and T_4. T_3 is the limit obtained by the aid of Cornish and Fisher's procedure. T_4 is the limit when the third and fourth moments of $\bar{y}$ are neglected.

In Table I limit T_1 is the actual limit calculated with the aid of Cornish and Fisher's procedure, T_2 is the limit when normality assumption is made for the distribution $\bar{y}$. In calculating T_1 and T_2, cumulants $k_{r,s}$ for $r + s > 2$ are not neglected.

6.6. Another approach to simplify the procedure considered was to calculate the tolerance limits on the basis of the assumption of normality for the distribution of simple mean of butter-fat itself as was done previously in an investigation of I.C.A.R.[8] The mean and variance of the distribution of the simple mean of butter-fat are taken as m and σ/S_n where m and σ^2 are the mean and variance of y distribution. Hence the limit $m - \xi\sigma/S_n$ can be easily calculated for any desired level. These limits designated as T_5 were also given in the last column, so as to examine the possibility of adopting this extremely simple procedure for practical purposes, if empirically justified.

6.7. Table III gives the various estimates T_1 to T_5 for the data taken for illustration.

TABLE III

Tolerance limits ($p = 0 \cdot 10$) for butter-fat content of herd milk samples of buffaloes—Calcutta City—Summer

n (Herd size)	T_1	T_2	T_3	T_4	T_5
5	6.41	6.35	6.39	6.36	6.44
10	6.46	6.51	6.54	6.53	6.63
25	6.68	6.66	6.68	6.67	6.80
50	6.76	6.74	6.75	6.74	6.88

We observe from Table III that limits T_1, T_2, T_3, T_4 do not differ much and T_5 to be an overestimate. Thus the use of simple T_5 procedure will lead to more stringent limits being fixed than aimed at.

7. The problem of estimating tolerance limits for mixed milk, where the milk samples are of both buffaloes and cows, is still to be solved. Since a mixture of two non-normal populations is involved the case is more complicated. Solution to this problem is also being investigated.

8. Summary

A procedure to estimate tolerance limits for quality characteristics of herd milk from data on individual samples is given and illustrated with the help of the data collected under the schemes of I.C.A.R. Cumulants of a weighted mean $\bar{y} = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i$ of the butter-fat content weighted with milk yields of different cows were worked out utilising Fisher's K-statistics. Adopting the Fisher and Cornish approach the tolerance limits for $\bar{y}$ corresponding to a prescribed level was then illustrated. The nature of the distribution of $\bar{y}$ was also studied. More approximate, but simple methods, of obtaining the limits were also compared with the limits obtained by the procedure described above.

9. Acknowledgements

The author wishes to express his sincere thanks to Mr. V. N. Amble, Assistant Statistical Adviser, I.C.A.R., and to Dr. V. G. Panse and Dr. P. V. Sukhatme for their keen interest in the work and encouragement given to him.

10. References

1. Amble, V. N. and Jocob, T. — *Indian Journal of Dairy Sciences*, June 1960.
2. Fisher, R. A. .. *Proceedings of London Mathematical Society*, 1928, **30**, Part III, 199–238.
3. Cornish, E. A. and Fisher, R. A. — *Extrait de la Revue l'Institut International de Statistique*, 1937, **4**, 1–14.
4. Kendall, M. G. .. *The Advanced Theory of Statistics*, **1**.
5. Cook, M. B. .. *Biometrika*, 1951, **38**, 368–87.
6. Kendall, M. G. .. *Annals of Eugenics*, 1940, **10**, 392–40.
7. Cook, M. B. .. *Biometrika*, 1951, **38**, 179–95.
8. I.C.A.R. .. *Miscellaneous Bulletin* No. 61, 1948.