

# Detection of a random alteration in a multivariate observation when knowing probable direction

Barry P. KATZ

*Regenstrief Institute and Department of Medicine, Indiana University, Indianapolis, IN, USA*

Morton B. BROWN

*Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA*

Received 27 October 1986

Revised 1 June 1987

*Abstract:* An observation from a multivariate distribution may be subject to perturbation in a known subset of the variables. A likelihood ratio statistic is developed to test whether or not there has been an addition of a random quantity in a prespecified direction to an observation from a multivariate normal distribution. When the variance of this addition is unknown, a secondarily Bayes approach is used to eliminate this variance which acts as a nuisance parameter. The testing procedure is based on a distribution-free tolerance interval.

*Keywords:* Random shift in location, Secondarily Bayes approach, Outlier detection, Quality control

## 1. Introduction

Quality control of naturally occurring products, such as fruit juices, requires that a single multivariate observation be classified as belonging, or not belonging, to a population whose parameters are estimated from previously collected samples of unadulterated juice. Since a single observation is being tested, the central limit theorem is not applicable and a likelihood ratio test procedure based on the assumption of multivariate normality may be very sensitive to the nonnormal distribution that is usually observed in these data.

A unique feature of the quality control of fruit juices is that the adulteration may be deliberate in order to reduce costs. The detection of adulterated juice has been a concern of both importing and exporting countries for many years. The major concern has been with economic fraud; however, in 1985, there were reported adulterations of wines that were hazardous to the consumer. Based on economic considerations, it is possible to predict common methods of adultera-

tion (Vandercook [15]); however, because those altering the juice choose not only the method of adulteration but also the amount, the amount of adulteration may be treated as a random variable.

Since adulteration is easily proven when a constituent found in the test sample does not occur naturally in the juice, it is more common to adulterate using only naturally occurring constituents. A method of adulteration is to dilute the juice with water and then add back constituents that can be cheaply replaced, so that they are restored to the amounts originally in the juice. Special consideration is given to constituents that affect the taste as perceived by the consumer.

Statistical methods currently used to test for adulteration of juice do not consider both a multivariate model for the data and the randomness of the amount of adulteration. The empirical alpha-levels of these tests may differ greatly from the nominal levels and result in too many rejections of pure juice or in a severe loss of power (Brown, Cohen and Volman [5], Brown and Cohen [3]).

The specific problem of testing whether a random quantity with prespecified direction has been added to (or subtracted from) one or more of the components of a single multivariate observation is not addressed directly in the statistical literature. In the two most closely related areas, outlier detection and industrial multivariate quality control, a number of statistical techniques based on principal components have been proposed (Rao [13], Hawkins [7], Jackson and Morris [8], Jackson and Mudholkar [9]). Tests based on statistics involving projections on the principal components with the smallest variance will not be most powerful in a prespecified direction. The positive simplex method of Aitchinson [1] was developed for compositional data, as could exist in this problem, but is not appropriate for testing whether a single observation is an outlier; also, in practice, only a subset of constituents are assayed. Univariate outlier techniques can be adapted, through the use of a data transformation, to incorporate the specified direction of the addition. However, the information contained in the unaltered components is not used. In addition, all of these techniques model the shift in location as constant rather than random.

Formally, the problem can be restated as: Let  $y_1, \dots, y_n$  be  $n$  observations from a  $p$ -variate distribution and let  $x$  be an additional  $p$ -variate observation. The null hypothesis is that  $x$  is from the same distribution as  $y_1, \dots, y_n$ . The alternate hypothesis is that  $x$  is from a distribution whose population mean has moved a random amount in a prespecified direction from the original distribution.

To develop a statistic to test the above hypothesis, we first assume multivariate normality for the underlying population and develop a likelihood ratio statistic (Section 2). The statistic is overparametrized, due to the assumption that the population mean has shifted a random amount. Therefore, a secondarily Bayes solution (Brown [2]) is used to eliminate the variance of the random shift (Section 3). Lastly, to reduce the dependence on normality, a tolerance interval is obtained for the test statistic (Section 4). Since a tolerance interval is distribution free, the validity of the test procedure depends on the rank order of the test statistic relative to a set of test statistics computed from independent samples. Use of a

tolerance interval requires that a large number of samples of unadulterated juices be available; data bases of several hundred samples are currently available (Brown, Katz and Cohen [6]). An example using fruit juice data is presented in Section 5.

## 2. Derivation of a likelihood ratio statistic

Let the  $p$ -variate vector  $x$  represent an observation from the multivariate normal distribution. Let  $r$  be the number of variables being modified where each is independently changed, or the number of independent prespecified directions which represent linear relationships imposed by forms of the adulteration or change. When  $r = p$ , the problem becomes that of discriminating between two populations based on a single observation. When  $r < p$ , it is possible to rotate the parameter space such that the prespecified directions form the first  $r$  axes of the space. Therefore, without loss of generality, the changes are assumed to occur in the first  $r$  components. The null ( $H_0$ ) and alternate ( $H_A$ ) hypotheses, of no change and of change, respectively, in the first  $r$  components only, can be expressed as

$$H_0: x \sim N_p(\mu, \Sigma) \quad \text{and} \quad H_A: x \sim N_p(\mu + \delta, \Sigma + \Sigma^*)$$

where  $N_p(\mu, \Sigma)$  represents a  $p$ -variate normal distribution with mean  $\mu$  and covariance matrix  $\Sigma$ ,

$$\delta' = (\delta_1, \dots, \delta_r, 0, \dots, 0)$$

and  $\Sigma^*$  is a variance-covariance matrix with zeros for all cells corresponding to null entries in  $\delta$ . The likelihood ratio statistic, when all the parameters except  $\Sigma^*$  are estimated, is

$$\lambda = \left[ \frac{|\hat{\Sigma}|}{|\hat{\Sigma} + \Sigma^*|} \right]^{-1/2} \exp \left\{ -\frac{1}{2} \left( (x + \hat{\mu})' \hat{\Sigma}^{-1} (x - \hat{\mu}) - (x - \hat{\mu} - \hat{\delta})' (\hat{\Sigma} + \Sigma^*)^{-1} (x - \hat{\mu} - \hat{\delta}) \right) \right\}. \tag{1}$$

Taking natural logarithms, multiplying by  $-2$  and ignoring the term that does not depend on  $x$  yields

$$L = (x - \hat{\mu})' (\hat{\Sigma}^{-1} - (\hat{\Sigma} + \Sigma^*)^{-1}) (x - \hat{\mu}) + 2\hat{\delta}' (\hat{\Sigma} + \Sigma^*)^{-1} (x - \hat{\mu}) - \hat{\delta}' (\hat{\Sigma} + \Sigma^*)^{-1} \hat{\delta} \tag{2}$$

Since there is only one observation, both  $\delta$  and  $\Sigma^*$  cannot both be estimated. Therefore, it is necessary to treat  $\Sigma^*$  as a nuisance parameter which must be eliminated from the likelihood. Of particular interest for detecting adulterated juice is the special case of a single alteration ( $r = 1$ ).

### 3. A secondarily Bayes approach

We consider the case where there is only one prespecified direction (i.e., one component is altered or there is a single linear relationship governing the  $1 < t < p$  components being changed). Without loss of generality, the random shift is assumed to occur in the first component since an orthogonal rotation can be performed such that  $x_1$  corresponds to the prespecified direction. The parameters  $\delta$  and  $\Sigma^*$  can now be expressed as  $\delta' = (\delta, 0, \dots, 0)$ ,

$$\Sigma^* = \begin{bmatrix} \sigma_{dd} & 0 \cdots 0 \\ 0 & 0 \cdots 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & 0 \cdots 0 \end{bmatrix}.$$

Substituting these quantities into (2) yields

$$\begin{aligned} L = (x - \hat{\mu})' & \left( \frac{\sigma_{dd}}{1 + \hat{\sigma}^{11}\sigma_{dd}} \hat{\Sigma}^1 \cdot \hat{\Sigma}^{1'} \right) (x - \hat{\mu}) + 2\hat{\delta} \left( 1 - \hat{\sigma}^{11} \frac{\sigma_{dd}}{1 + \hat{\sigma}^{11}\sigma_{dd}} \right) (x - \hat{\mu})' \hat{\Sigma}^1 \\ & - \hat{\delta}^2 \hat{\sigma}^{11} \left( 1 - \hat{\sigma}^{11} \frac{\sigma_{dd}}{1 + \hat{\sigma}^{11}\sigma_{dd}} \right) \end{aligned} \quad (3)$$

where

$$\hat{\Sigma}^1 = \text{the first column of } \hat{\Sigma}^{-1}$$

and

$$\hat{\sigma}^{ij} = \text{the element in the } i^{\text{th}} \text{ row and } j^{\text{th}} \text{ column of } \hat{\Sigma}^{-1}.$$

The maximum likelihood estimates,  $\hat{\mu}$ ,  $\hat{\delta}$ , and  $\hat{\Sigma}$ , cannot all be obtained from a single observation. In many applied situations it is reasonable to assume that there is a data base of unaltered observations,  $y_1, \dots, y_N$ , whose mean  $\bar{y}$  is an estimate of  $\mu$  and whose variance-covariance matrix  $\hat{\Sigma}$  is an estimate of  $\Sigma$ . Then,  $\delta$  is estimated by

$$\hat{\delta} = x_1 - \bar{y}_1$$

where  $x_1$  and  $\bar{y}_1$  are the first elements of  $x$  and  $\bar{y}$ , respectively. Although the direction of  $\delta$  is known, the sign may not be; the case of a one-sided test is discussed in Section 4. In the expression for  $L$  (3),  $\sigma_{dd}$  is not estimable from a single observation. We choose to reexpress  $L$  as a function of the ratio  $\sigma_{dd}/\sigma_{11}$  to which a prior distribution can be applied.

If

$$\hat{\Sigma} = DRD \quad (4)$$

where

$D$  = the  $p \times p$  diagonal matrix of standard deviations of  $y$

and

$R$  = The  $p \times p$  sample correlation matrix for  $y$ , then the likelihood ratio

statistic  $L$  can be expressed in terms of the ratio  $\sigma_{dd}/\sigma_{11}$  as

$$L = \frac{((x - \bar{y})' D^{-1} R^1)^2}{r^{11}} - \frac{((x^* - \bar{y}^*)' D^{-1} R^1)^2}{r^{11}(1 + r^{11}(\sigma_{dd}/\sigma_{11}))} \quad (5)$$

where

$x^* = x$  with the first element set to zero,

$y^* = y$  with the first element set to zero,

$R^1 =$  the first column of  $R^{-1}$ ,

and

$r^{11} =$  the first element of  $R^{-1}$ .

(The proof of (5) is given in Katz [10]).

A secondarily Bayes method eliminates nuisance (or secondary) parameters by integrating over a prior distribution for these parameters (Brown [2]). In the likelihood ratio statistic  $L$ , the secondary parameter of interest is  $\nu = \sigma_{dd}/\sigma_{11}$ . Assuming a prior distribution  $p(\nu)$  for the ratio  $\nu$  and computing the posterior distribution yields an estimator of the likelihood ratio statistic  $L^*$  that is not a function of  $\sigma_{dd}/\sigma_{11}$  where

$$\begin{aligned} L^* &= \int_0^\infty \frac{((x - \bar{y})' D^{-1} R^1)^2}{r^{11}} - \frac{((x^* - \bar{y}^*)' D^{-1} R^1)^2}{r^{11}(1 + r^{11}\nu)} p(\nu) d\nu \\ &= \frac{((x - \bar{y})' D^{-1} R^1)^2}{r^{11}} - \frac{((x^* - \bar{y}^*)' D^{-1} R^1)^2}{r^{11}} \int_0^\infty \frac{p(\nu)}{1 + r^{11}\nu} d\nu. \end{aligned} \quad (6)$$

Functional forms of the prior distribution may be chosen that reflect the known information about the ratio and whose integration yields a closed form solution.

### Using a null prior

When there is no information known about a parameter, the null prior distribution is that the logarithm of the variance is uniform; i.e.,  $\ln(\sigma^2)$  is uniform. For the ratio of two variances  $\nu$  this assumption yields an improper prior of the form

$$p(\nu) \propto \frac{1}{\nu}. \quad (7)$$

Since the integral in (6) is not finite for this prior, the statistic  $L^*$  must be expressed using a limit:

$$L^* = \frac{((x - \bar{y})' D^{-1} R^1)^2}{r^{11}} - \frac{((x^* - \bar{y}^*)' D^{-1} R^1)^2}{r^{11}} \lim_{a \rightarrow \infty} \frac{\int_{1/a}^a \frac{1}{\nu(1 + r^{11}\nu)} d\nu}{\int_{1/a}^a \frac{1}{\nu} d\nu} \quad (8)$$

After integrating and taking the limit, the likelihood ratio statistic  $L^*$  reduces to

$$L^* = \frac{((x - \bar{y})' D^{-1} R^1)^2}{r^{11}} - \frac{((x^* - \bar{y}^*)' D^{-1} R^1)^2}{2r^{11}} \quad (9)$$

Equation 9 can be expressed in scalar notation as

$$L^* = \frac{1}{r^{11}} \left( \sum_{i=1}^p r^{ii} (x_i - \bar{y}_i) / d_i \right)^2 - \frac{1}{2r^{11}} \left( \sum_{i=2}^p r^{ii} (x_i - \bar{y}_i) / d_i \right)^2.$$

This statistic differs from the square of the standardized residual from the multiple regression of  $y_1$  on the standardized variables by a factor of one half in the last term.

### Using a conjugate prior

When prior information about the ratio is available, an appropriate conjugate prior distribution for the ratio of two variances may be based on the  $F$ -distribution; i.e.,

$$\nu \sim KF_{m,n} \quad (10)$$

where  $K$  is chosen to reflect prior knowledge about the value of the ratio and  $m$  and  $n$  to indicate the confidence one has in the prior.

In this case the portion of the statistic that depends on the choice of the prior is

$$f = \frac{\left(\frac{m}{n}\right)^{m/2}}{KB \binom{m+n}{2,2}} \int_0^\infty \frac{1}{1+r^{11}\nu} (\nu/K)^{(m/2)-1} \left(1 + \frac{m\nu}{nK}\right)^{-(m+n)/2} d\nu. \quad (11)$$

This can be simplified using the transformation:

$$f = \frac{1}{B(p, q)} \int_1^c \frac{(c-z)^q (z-1)^{p-1}}{z(c-1)^{p+q}} dz$$

where

$$c = \frac{r^{11}Kn}{m}, \quad z = \frac{cm\nu + Kn}{m\nu + Kn}.$$

Equation (12) is an integral of a polynomial in  $z$  that has terms ranging from  $z^{(p+q-2)}$  to  $z^{-1}$ . Therefore the integration can be done in closed form as a sum of integrals. Results of the evaluation of (12) for combinations of  $m$ ,  $n$ , and  $K$  over a range of  $R^2$  values indicate the choice of  $m$  and  $n$  is relatively unimportant when compared to the choice of  $K$  in determining the weight given to the second term of the statistic (Katz [10]).

#### 4. Estimation of the critical value

The procedure to find the critical value for  $L^*$  is designed to be used with any historical data base. The initial step is to estimate  $\mu$  and  $\Sigma$  by  $\bar{y}$  and  $\hat{\Sigma}$ , respectively, from the data. The next step is to construct a sample distribution for  $L^*$ . This can be achieved by deleting each observation  $y_i$  in turn and calculating  $\bar{y}_{(i)}$  and  $\hat{\Sigma}_{(i)}$  from the remaining  $N-1$  observations. Then the statistic  $L_i^*$  can be computed for each observation by replacing  $x$ ,  $\bar{y}$  and  $\hat{\Sigma}$  with  $y_i$ ,  $\bar{y}_{(i)}$ , and  $\hat{\Sigma}_{(i)}$  respectively in the calculation of  $L^*$ . This process results in a collection of  $n$  correlated  $L_i^*$ 's. However, this correlation is of the order  $1/N$  since it is the result of the correlation between  $y_i$  and  $\bar{y}$ . For the large sample sizes for which this procedure is designed, this correlation will be negligible. Therefore, the  $L_i^*$ 's can be used to provide a sample distribution for  $L^*$ .

The final step is to use the sample distribution to choose a critical value. Since the procedure deals with a sample distribution rather than the actual distribution, the probability that the likelihood ratio statistic is greater than some constant must be estimated. A tolerance interval is used to ensure that the  $\alpha$ -level does not exceed the nominal  $\alpha$ -level. The form of the distribution-free tolerance interval (Wilks [16]) is

$$P\{P\{L^* > L_{(j)}^* | H_0\} \leq \alpha\} \geq \gamma \quad (13)$$

where  $L_{(j)}^*$  is the  $j^{\text{th}}$  order statistic of the sample distribution. The value of  $j$  is determined by choosing a  $\gamma$  and an  $\alpha$ ; then the critical value, i.e. the critical rank, is obtained from the charts provided by Murphy [11].

For example, if the historical data set has 500 observations and the values of  $\alpha$  and  $\gamma$  are chosen as 0.05 and 0.99 respectively, then  $j = 485$ , the ninety-seventh percentile of the sample distribution. Any new observation,  $x$ , can be tested by calculating  $L$  using  $\bar{y}$  and  $\hat{\Sigma}$  and then comparing it to  $L_{(485)}^*$ . If it is greater than  $L_{(485)}^*$ , the conclusion is that  $x$  is not drawn from the same distribution as the historical data.

The above procedure is for a two-sided alternative hypothesis that does not specify the sign of  $\delta$ . When a one-sided alternative hypothesis is of interest, this procedure is modified by dividing the historical data set into two groups based on the sign of  $(x_1 - \bar{y}_1)$ . All observations are still used to calculate  $\Sigma$  and  $\mu$ . However, for the group that is in the opposite direction from the alternate hypothesis, the value of the likelihood ratio statistic is set to the minimum  $L^*$  calculated for the other group. The statistic is calculated in the usual manner for the observations in the other group. All values of  $L^*$  are then ranked (including those set to the minimum). The critical value chosen in this manner for the one-sided test will be less than or equal to the critical value obtained for the two-sided test, even though the value of  $j$  remains unchanged. A modification must also be made to the decision rule. For example, if the alternate hypothesis specifies a positive addition, then  $H_0$  would only be rejected for a new observation  $x$  if  $(x_1 - \bar{y}_1) > 0$  and  $L^* > L_{(j)}^*$ .

The derivation of the likelihood ratio statistic is based on the multivariate normal distribution. However, since a distribution-free tolerance interval is used

to determine the critical value, the only necessary assumption about the  $y_i$ 's is that they are identically distributed and the procedure remains valid even when normality is lacking, although the power of the test may be affected.

## 5. Example

The usefulness of this statistic in a real situation was evaluated using an unpublished data set of 572 samples of pure Israeli orange juice collected by government inspectors during eight growing seasons (each growing season runs from the fall of one year to the early summer of the next). Each season is treated, in turn, as a set of new observations while the remaining seven growing seasons are used as the historical data set, producing sample sizes from 480 to 520. That is, observations from a 'new' season are compared against data collected during all the other growing seasons; this reflects the actual quality control situation where data from pure samples in the current season are as yet unavailable when samples of juice are tested for adulteration. Previous analyses of these data have shown that the within-year variance is generally two to three times larger than the between-year variance (Brown & Cohen [4]). Six components from the data set are used in this analysis: ash, potassium, sodium, chloride, acidity and brix. To simulate the case of a random addition, the ash and potassium components were multiplied by a random dilution factor generated from a Gaussian distribution with mean 0.8 and standard deviation 0.04. This is equivalent to adding a random amount of solution containing all components except ash and potassium with an expected value of 20% of the total volume of the juice and a standard deviation of 4%. The likelihood ratio method presented in this paper is implemented using a null prior and a 99% tolerance interval.

The results obtained are compared to the results obtained with two methods previously suggested in the fruit juice literature, a regression method (Rolle and Vandercook [14]), and a chi-square test (Lifshitz, Stepak and Brown [12]).

In the regression method the dependent variable is that constituent whose regression produces the highest  $R^2$  when regressed on the other constituents. Brown and Cohen [3] show that this method produces confidence limits for the residuals that are too narrow and, as a result, too many pure samples are rejected. In the comparison presented in this paper the dependent variable selected by this procedure was ash and the multiple  $R^2$  was 0.675 for the entire data set.

The chi-square test of Lifshitz et al. (1974) is equivalent to using Hotelling's  $T^2$  as a test for an outlier. All six components were used in the test. Due to the large number of observations in the estimate of  $\Sigma$ , it is possible to approximate Hotelling's  $T^2$  by a chi-square test. This test, as may be expected, is very sensitive to the assumption of multivariate normality.

The estimated  $\alpha$ -levels and powers for all three methods are presented in Table 1. The likelihood ratio test is the most conservative and is the only method that does not have any estimates of the  $\alpha$ -level that are significantly greater than the nominal  $\alpha$ -level. In addition, the power estimates show that the likelihood ratio



Table 1  
Results of the analysis of the Israeli fruit juice data

Season	$n$	Likelihood Ratio Method	Chi-square Method	Regression Method
Estimated $\alpha$ -levels				
1976	74	0.041	(0.189)	(0.230)
1977	92	0.054	(0.293)	(0.228)
1978	52	0.019	0.077	0.050
1979	84	0.024	0.048	0.000
1980	66	0.000	0.045	0.000
1981	52	0.115	(0.135)	0.058
1982	75	0.000	0.040	0.000
1983	77	0.000	0.039	0.026
Estimated powers				
1976	74	0.324	(0.378)	(0.203)
1977	92	0.228	(0.457)	(0.402)
1978	52	0.250	0.212	0.154
1979	84	0.131	0.083	0.012
1980	66	0.379	0.212	0.030
1981	52	0.346	(0.173)	0.019
1982	75	0.467	0.267	0.120
1983	77	0.403	0.429	0.104

(·) = the estimated  $\alpha$ -level is significantly greater than 0.05 using a one-sided  $z$ -test ( $\alpha = 0.05$ ).

test is generally more powerful than the other two methods in the cases when the  $\alpha$ -levels are appropriate.

## 6. Summary and conclusions

In this paper a statistical test is developed to detect a random alteration in a prespecified direction to a single multivariate observation. When the direction of the adulteration is not that of the first component  $x_1$  of  $x$ , an orthogonal rotation can be performed so that  $x_1$  conforms to this direction. When the prespecified direction involves  $t$  ( $> 1$ ) variables, the orthogonal rotation will create  $t-1$  variables that are uncorrelated with  $x_1$  after the rotation. These  $t-1$  variables do not affect the value of  $L^*$  and therefore may be eliminated from the computations (Katz [10]). Hence, when  $t = p$  (i.e., all variables are included in the direction), the likelihood ratio statistic is equivalent to Student's  $t$ .

When  $t < p$ , a secondarily Bayes approach is used to obtain a likelihood ratio statistic that is not a function of  $\sigma_{dd}/\sigma_{11}$ . This likelihood ratio statistic incorporates information from all  $p$  components; thus, the entire variance-covariance matrix  $\Sigma$  and not just the variance of the adulteration  $\Sigma^*$  is included. In addition,  $L^*$  is similar to the square of the standardized residual using standardized variables, differing only by a factor in the last term.

A distribution-free tolerance interval is used to obtain the order statistic that represents the critical value. Although multivariate normality is used to define the likelihood ratio statistic, the tolerance interval is distribution-free. Therefore, departures from normality will not affect the significance level of the procedure; however, power may be affected. The use of a tolerance interval requires large sample sizes. For example, no tolerance interval exists for  $\alpha = 0.05$  and  $\gamma = 0.99$  when  $n < 90$ . Although in some situations this may seem like a large data set, historical data sets many times this size are commonly available for fruit juices, as well as in other areas, including medicine.

If the magnitude of the adulteration is assumed to be constant, the likelihood ratio statistic reduces to the square of the standardized residual from the multiple regression of the standardized variables (Katz [10]). If, in addition, the direction of the adulteration is unknown, the likelihood ratio statistic is equivalent to Hotelling's  $T^2$ .

In many situations the direction of expected change is known before the data are collected. For example, the relative costs of the components in citrus fruit juice can be used to determine the most probable direction of adulteration (Vandercook [15]). None of the techniques that are presently used in the quality control of fruit juice incorporate this information. (For a review of the present techniques see, for example, Brown and Cohen [3]). Eventually, techniques may be developed that also assume that the sampling and testing methods are known to the adulterators, thus altering their behavior. We have not considered this situation.

The problem addressed in this paper differs from the issues addressed in related areas such as outlier identification. Consequently, the proposed technique has four main features that are not found together in methods designed for other purposes: (1) information from all of the variates, and not only the components that may have been altered, is utilized; (2) the shift in location is modeled as a random variable to allow for an increase in dispersion; (3) unlike omnibus tests to detect all adulterations, power is focused in a prespecified direction; and (4) as evidenced by the example, the assumption of multivariate normality is not critical.

## References

- [1] J. Aitchison, The statistical analysis of compositional data (with discussion), *Journal of the Royal Statistical Society B* **44** (1982) 139–177.
- [2] M.B. Brown, The two means problem – A secondarily Bayes approach, *Biometrika* **54** (1967) 85–91.
- [3] M.B. Brown and E. Cohen, Discussion of statistical methods for determining purity of citrus juice, *Journal of the Association of Official Analytical Chemists* **66** (1983) 781–788.
- [4] M.B. Brown and E. Cohen, Statistical issues in the quality control of fruit juices. Proceedings of the 18th International Symposium, International Federation of Fruit Juice Producers, Tel Aviv, Israel, March 1984, pp. 123–129.
- [5] M.B. Brown, E. Cohen and L. Volman, Comments on statistical methods for determining the purity of citrus juice, *Flussiges Obst* **48** (1981) 286–298.

- [6] M.B. Brown, B.P. Katz and E. Cohen, Statistical procedures for the identification of adulteration in fruit juices, in: *Adulteration of Fruit Juice Beverages* (1986) in press.
- [7] D.M. Hawkins, The detection of errors in multivariate data using principal components, *Journal of the American Statistical Association* **69** (1974) 340–344.
- [8] J.E. Jackson and R.H. Morris, An application of multivariate quality control to photographic processing, *Journal of the American Statistical Association* **52** (1957) 186–199.
- [9] J.E. Jackson and G.S. Mudholkar, Control procedures for residuals associated with principal component analysis, *Technometrics* **21** (1979) 341–349.
- [10] B.P. Katz, Detection of a random alteration in a multivariate observation based on knowledge of probable direction, Ph.D. dissertation, The University of Michigan, Ann Arbor, Michigan (1984).
- [11] R.B. Murphy, Non-parametric tolerance limits, *Annals of Mathematical Statistics* **19** (1948) 581–589.
- [12] A. Lifshitz, Y. Stepak and M.B. Brown, Method for testing the purity of Israeli citrus juice, *Journal of the Association of Official Analytical Chemists* **57** (1974) 1169–1175.
- [13] C.R. Rao, The use and interpretation of principal component analysis in applied research, *Sankhyá A* **26** (1964) 329–358.
- [14] B.A. Rolle and C.E. Vandercook, Lemon juice composition, 3. Characterization of California-Arizona lemon juice by use of multiple regression analysis, *Journal of the Association of Official Analytical Chemists* **46** (1963) 362–365.
- [15] C.E. Vandercook, Stepwise approach to the determination of orange juice content, Proceedings of the Seventeenth International Symposium, International Federation of Fruit Juice Producers, Tel Aviv, Israel, March 1984.
- [16] S.S. Wilks, Determination of sample sizes for setting tolerance limits, *Annals of Mathematical Statistics* **12** (1941) 91–96.