ScienceDirect

Hotelling's T2

Related terms:

Degrees of Freedom, F Distribution, MANOVA, MATLAB, Sample Mean, Trimmed Mean, Multivariate, Multivariate Data, Univariate

Inferential Statistics V: Multiple and Multivariate Hypothesis Testing

Andrew P. King, Robert J. Eckersley, in Statistics for Biomedical Engineers and Scientists, 2019

8.3.1 Hotelling's T^2 Test

Hotelling's T^2 test is a generalization of the Student's *t*-test to multivariate data. Therefore, it is a parametric test and assumes normally distributed data. More specifically, it assumes that the data were drawn from a *multivariate normal distribution*. A multivariate normal distribution is essentially a generalization of the univariate normal distribution (see Section 4.4.1) to more than one variable.

We will illustrate the test using the case of testing a single sample against expected values (i.e. similar to Section 5.5 for the *t*-test). First, we form our hypotheses. The null hypothesis is that there is no difference between our (multivariate) sample mean and the expected mean. The alternative hypothesis is that there is a difference.

To compute the test statistic for Hotelling's T^2 test, we first compute a T^2 value defined as

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T C^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}), \tag{8.8}$$

where n is the sample size, $\bar{\mathbf{x}}$ is the multivariate sample mean (i.e. a column vector), $\boldsymbol{\mu}$ is the column vector of expected values, and C is the sample *covariance matrix* (see Section 2.3.2).

For example, if the sample $\mathbf{x}$ contained p variables, then we would have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

and

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix},$$

where $\bar{x}_1, \ldots, \bar{x}_p$ are the means of the corresponding elements of **x** across the sample. The definition of the covariance matrix was presented in Eq. (2.3) and is reproduced here for convenience:

$$C = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T. \tag{8.9}$$

Here, the subscript *i* represents the sample index, and *n* is again the sample size.

Looking at Eq. (8.8), we can see that large differences between the sample mean and the expected values will lead to a high value of T^2.

Eq. (8.8) is known as *Hotelling's* T^2 and is named after its inventor Harold Hotelling, who published the mathematical formulation of the test back in 1931 [8]. The test statistic for Hotelling's T^2 test is computed from the T^2 value as follows

$$F = \frac{n-p}{p(n-1)} T^2 \tag{8.10}$$

where *p* is the number of variables, and *n* is the sample size. Importantly, *it can be shown that the test statistic F follows an F-distribution (see Section 8.2.2) with p and* $n - p$ *degrees of freedom*. A high value of *F* indicates a larger difference between the sample data and the expected values (because *F* is a scaled version of T^2). Because we know the distribution of *F*, we can use the critical values from Tables A.8 and A.9 to determine how large *F* should be in order to reject the null hypothesis. When using the tables, we need the degrees of freedom, which, as stated before, are *p* and $n - p$.

Let us illustrate the use of this test with an example. A study is investigating possible links between diet and depression. A group of 10 patients suffering from depression have been recruited, and four daily dietary intake values have been recorded for each patient: *selenium*, *vitamin D*, *omega-3 fatty acids*, and *zinc*. The recommended daily intakes for these are: selenium=35 μg, vitamin D=10 μg, omega-3 = 250 mg and zinc$= 15$ mg. We want to know if the patient cohort have significantly different dietary intakes to the recommended values.

The data for the 10 patients are presented in Table 8.2, which also shows the mean values for each variable. We want to compare these mean values with the expected values from the recommended daily intake figures. Our null hypothesis is that there is no difference between the two, and the alternative hypothesis is that there is a difference. We will work to a 95% degree of confidence.

Table 8.2. Data from the study into the links between diet and depression.

	Daily dietary intake			
Patient no.	***Serenium (μg)***	***Vitamin D (μg)***	***Omega-3 fatty acids (mg)***	***Zinc (mg)***
1	30.21	4.57	243.7	16.39
2	34.8	16.16	274.4	12.21
3	34.54	17.65	259.67	23.08
4	21.02	5.72	239.51	7.83
5	24.56	8.85	254.89	12.36
6	20.59	16.85	235.83	18.09
7	32.49	6.84	251.23	11.72

	Daily dietary intake			
Patient no.	***Serenium (μg)***	***Vitamin D (μg)***	***Omega-3 fatty acids (mg)***	***Zinc (mg)***
8	20.95	15.87	271.98	15.07
9	37.91	12.28	256.98	17.62
10	25.96	9.22	237.61	20.03
Mean:	28.3	11.4	252.58	15.44

Now we will compute our test statistic. First, to compute T^2, Eq. (8.8) needs the sample covariance matrix. Using the data in the table and Eq. (2.3), we can compute this to be

$$C = \begin{bmatrix} 42.3 & 3.03 & 34.86 & 8.04 \\ 3.03 & 24.79 & 35.01 & 10.95 \\ 34.86 & 35.01 & 187.05 & -4.23 \\ 8.04 & 10.95 & -4.23 & 20.44 \end{bmatrix}.$$

We also know the mean of the sample $\bar{\mathbf{x}}$ (this is the bottom row of values in Table 8.2), $\boldsymbol{\mu}$ is the vector of expected values $[35\ 10\ 250\ 15]^T$, and $n = 10$ is the sample size. Therefore, using Eq. (8.8), we compute $T^2 = 19.58$. This value, together with $n = 10$ and $p = 4$ (the number of variables), is used in Eq. (8.10) to compute our test statistic $F = 3.26$. To look up the critical *F* value, we find the degrees of freedom $p = 4$ and $n - p = 6$. From Table A.8 we see that our critical value for 4 and 6 degrees of freedom is 4.534. As 3.26 is not greater than 4.534, we cannot reject the null hypothesis, so we cannot show that there is any significant difference between the dietary intake of the patients and the recommended values.

The Intuition. Hotelling's T^2 Test

To understand Hotelling's T^2 test, we first return to the univariate case. Recall that the Student's *t*-test is based on the *t* statistic, which was defined in Eq. (5.1):

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}},$$

where $\bar{x}$ is the (univariate) sample mean, μ is the expected mean value, *n* is the sample size, and *s* is the sample standard deviation. This computes the difference between the sample mean and the expected mean in terms of the number of standard errors of the mean. Squaring both sides and rearranging results in

$$t^2 = n\left(\bar{x} - \mu\right)\left(s^2\right)^{-1}\left(\bar{x} - \mu\right),$$

where s^2 is the sample variance.

Generalizing this to the multivariate case, the corresponding equation can be shown to be

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T C^{-1}\left(\bar{\mathbf{x}} - \boldsymbol{\mu}\right),$$

where $\bar{\mathbf{x}}$ is now the multivariate sample mean, $\boldsymbol{\mu}$ is the vector of expected values, and *C* is the covariance matrix. This was presented earlier as Eq. (8.8), and the similarity with the univariate version is apparent.

The key to understanding Hotelling's T^2 test is that when we square a *t*-distributed (multivariate) random variable with *p* values and $n - 1$ degrees of freedom, the result is an *F*-distributed random variable with *p* and $n - p$ degrees of freedom. The mathematics underpinning this finding were described by Harold Hotelling in his original 1931 paper [8]. It was also shown that when the null hypothesis of the

test is true (i.e. the sample mean is equal to the expected mean), we have the following approximation:

$$T^2 \approx \frac{p(n-1)}{n-p} F_{p,n-p},$$

where $F_{p,n-p}$ represents the *F*-distribution with *p* and $n - p$ degrees of freedom. Rearranging this formula gives the equation for the *F* statistic from Eq. (8.10):

$$F = \frac{n-p}{p(n-1)} T^2.$$

As for the ANOVA test, the critical values represent values beyond which the area under the *F*-distribution is 5%.

Activity 8.3

A cohort of 20 volunteers has been recruited to take part in an evaluation of a new drug. The investigating team would like to be sure that the cohort have certain blood serum levels that are typical of the population as a whole. The table below shows the constituents measured from the volunteers together with their typical population values.

O8.C

Constituent	Typical level
Calcium	10 mg/dL
Magnesium	2.8 mg/dL
Phosphorous	3.8 mg/dL
Bilirubin	1.0 mg/dL
Albumin	4.5 g/dL

To determine whether the cohort have typical blood serum levels, the team would like to perform a Hotelling's T^2 test. The T^2 statistic has been calculated from the data to be 12.84. What is the result of the test?

Some Multivariate Methods

Rand Wilcox, in Introduction to Robust Estimation and Hypothesis Testing (Third Edition), 2012

6.7.2 Extension of Hotelling's T^2 to Trimmed Means

Hotelling's T^2 test is a classic method for testing

$$H_0 : \mu = \mu_0,$$

where μ represents a vector of *p* population means and μ_0 is a vector of specified constants. The method is readily generalized to making inferences about the marginal trimmed means via the test statistic

$$T^2 = \frac{h(h-p)}{(n-1)p} (\overline{\mathbf{X}}_t - \mu_0)\mathbf{S}^{-1}(\overline{\mathbf{X}}_t - \mu_0)',$$

where **S** is the Winsorized variance–covariance matrix corresponding to the p measures under study and $\overline{\mathbf{X}}_t$ is the vector of marginal trimmed means and h is the number of observations left after trimming. When the null hypothesis is true, T^2 has, approximately, an F distribution with degrees of freedom $\nu_1 = p$ and $\nu_2 = h - p$. That is, reject at the α level if

$$T^2 \geq f,$$

where f is the $1 - \alpha$ quantile of an F distribution with $\nu_1 = p$ and $\nu_2 = h - p$ degrees of freedom.

Some Multivariate Methods

Rand Wilcox, in Introduction to Robust Estimation and Hypothesis Testing (Fourth Edition), 2017

6.7.2 Extension of Hotelling's T^2 to Trimmed Means

Hotelling's T^2 test is a classic method for testing

$$H_0 : \mu = \mu_0,$$

where μ represents a vector of p population means and $\boldsymbol{\mu_0}$ is a vector of specified constants. The method is readily generalized to making inferences about the marginal trimmed means via the test statistic

$$T^2 = \frac{h(h-p)}{(n-1)p}(\bar{\mathbf{X}}_t - \mu_0)\mathbf{S}^{-1}(\bar{\mathbf{X}}_t - \mu_0)',$$

where **S** is the Winsorized covariance matrix corresponding to the p measures under study and $\bar{\mathbf{X}}_t$ is the vector of marginal trimmed means and h is the number of observations left after trimming. (The Winsorized covariance for any two variables was described in Section 5.9.13.) When the null hypothesis is true, T^2 has, approximately, and F distribution with degrees of freedom $\boldsymbol{\nu_1 = p}$ and $\boldsymbol{\nu_2 = h - p}$. That is, reject at the α level if

$$T^2 \geq f,$$

where f is the $\mathbf{1 - \alpha}$ quantile of an F distribution with $\boldsymbol{\nu_1 = p}$ and $\boldsymbol{\nu_2 = h - p}$ degrees of freedom.

Statistical Significance Versus Effect Size

X. Fan, T.R. Konold, in International Encyclopedia of Education (Third Edition), 2010

Multivariate D_M

For multivariate group comparison (e.g., comparison of two groups on multiple outcome variables, as in a Hotelling T^2 test or multivariate analysis of variance (MANOVA)), Mahalanobis distance (D_M) is the multivariate counterpart of d:

$$D_M = \sqrt{(\overline{X}_{G1} - \overline{X}_{G2})' S_{pooled}^{-1} (\overline{X}_{G1} - \overline{X}_{G2})}$$

where $\overline{X}_{G1}$ and $\overline{X}_{G2}$ are the mean vectors of the two groups in the comparison, $(\overline{X}_{G1} - \overline{X}_{G2})'$ is the transposed mean vector difference, and S_{pooled}^{-1} is the inverse of the pooled covariance matrix.

More Regression Methods

Rand Wilcox, in Introduction to Robust Estimation and Hypothesis Testing (Fourth Edition), 2017

Methods Based on the Least Squares Estimator

Now the goal is to test the hypothesis that the regression lines are identical when using least squares regression estimator. Let **V** be the sample covariance matrix based on the d_{kb}^* values, where b_{k1}^* and b_{k2}^* are the least squares estimates based on a bootstrap sample. That is,

$$v_{k\ell} = \frac{1}{B-1}\sum(d_{kb}^* - \bar{d}_k^*)(d_{\ell b}^* - \bar{d}_\ell^*),$$

where $\bar{d}_k^* = \sum d_{kb}/B$. The test statistic is based on a simple modification of Hotelling's T^2 statistic for testing the hypothesis that a multivariate normal distribution has a mean of zero:

$$H = \frac{n(n-2)}{2(n-1)}(d_0, d_1)\mathbf{V}^{-1}(d_0, d_1)'. \tag{11.14}$$

Reject the null hypothesis given by (11.9), at the α level, if *H* is greater than or equal to the $1-\alpha$ quantile of an F distribution with $\nu_1 = 2$ and $\nu_2 = n-2$ degrees of freedom. In the event there are *p* independent variables, proceed in the manner just described, only now $\nu_1 = p+1$.

Now consider the goal of testing the hypotheses given by Eq. (11.13). The estimated squared standard error of d_k is $v_{k+1,k+1}$ and the test statistic is

$$T_k = \frac{d_k}{\sqrt{v_{k+1,k+1}}}.$$

The null distribution is approximated by a Student's t distribution with $n-1$ degrees of freedom. When the sample size is small, say $n = 20$, removing the data associated with leverage points can improve the control over the Type I error probability in some situations when there is heteroscedasticity (Wilcox, 2015d). The method is readily generalized to situations where there are $p > 1$ independent variables.

Volume 4

R. Todeschini, ... P. Gramatica, in Comprehensive Chemometrics, 2009

4.05.7.2 Applicability Domain

The concept of AD concerns the predictive use of QSAR/QSPR models and, then, is closely related to the concept of model validation. In other words, the AD is a concept related to the quality of the QSAR/QSPR model predictions and prevention of the potential misuse of model's results. A key component of the quality prediction is to define when a QSAR/QSPR model is suitable to predict a

property/activity of a new compound, that is, define model's AD.[164,174,176–178,180,181]

A model will yield reliable predictions when model assumptions are fulfilled and unreliable predictions when they are violated. In particular, for QSAR/QSPR models, based on statistical mining techniques, the training set and the model prediction space are the basis for estimation of chemical space where predictions are reliable.

Two basic approaches were proposed to evaluate the AD. The first approach to AD evaluation is the analysis of the training set, which has its grounds in statistics, because the interpolated prediction results are more reliable than extrapolated. Extrapolation is not a problem in principle, because extrapolated results from theoretical well-founded models can often be reliable. However, QSAR/QSPR models are usually based on empirical and limited experimental evidence and/or are only locally valid; therefore, extrapolation always results in higher uncertainty and usually in unreliable predictions.

Different approaches to estimate interpolation regions in a multivariate space were evaluated by Jaworska,[178,179] based on (1) ranges of the descriptor space; (2) distance-based methods, using Euclidean, Manhattan, and Mahalanobis distances, Hotelling T^2 method, and leverage values; and (3) probability density distribution methods based on parametric and nonparametric approaches. Both ranges and distance-based methods were also evaluated in the principal component space.

One of the common tools used to visualize the AD of a QSAR model is the plot of standardized residuals in prediction (r_i) versus leverage values (h_i) for each ith sample. This plot, called Williams plot, allows an immediate and simple graphical detection of both the response outliers (i.e., compounds with standardized residuals in prediction greater than three standard deviation units, $r_i > 3\sigma$) and structurally influential chemicals in a model ($h_i > h^*$), where h^* is a threshold value, usually 2 or 3 times the average leverage value. In effect, when the leverage value of a compound is lower than the critical value h^*, the probability of accordance between predicted and actual values is as high as that for the training set chemicals. Conversely, a high leverage chemical is structurally distant from the other chemicals; thus, it can be considered outside the AD of the model.

Figure 13 shows the Williams plot of a model for polar narcotics in *Pimephales promelas* as an example.[183] Here, chemical 347 is wrongly predicted ($r_i > 3\sigma$); it is a test chemical completely outside the AD of the model, because its leverage value is beyond the vertical leverage threshold line; thus, it is both a response outlier and a high leverage chemical.

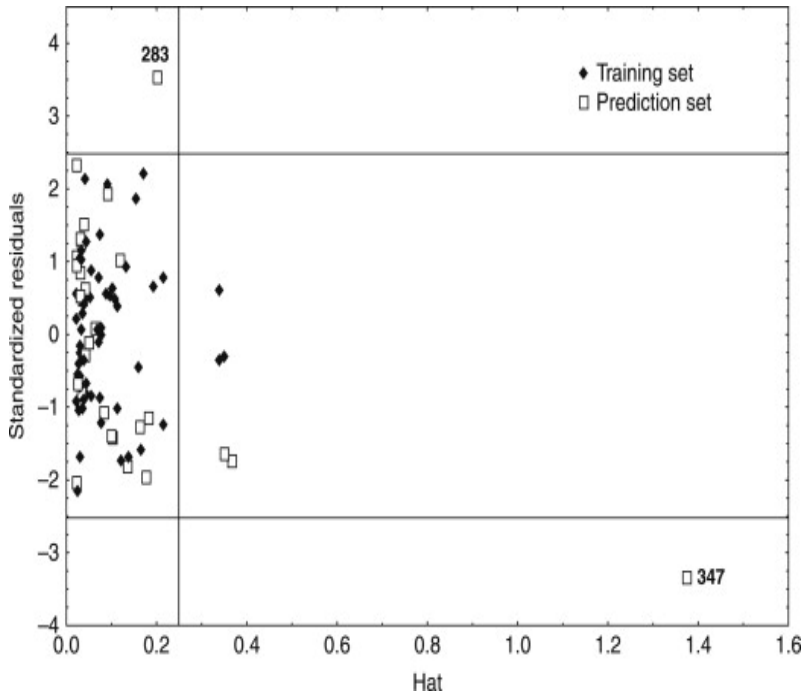

Figure 13. Williams plot for an externally validated model for polar narcotics (leverage cutoff value: 2.5 h^*). Reproduced from Papa, E.; Villa, F.; Gramatica, P. Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in *Pimephales promelas* (Fathead Minnow). *J. Chem. Inf. Model.* **2005**, 45, 1256–1266.

Two other chemicals (squares at 0.35 *h*) slightly exceed the critical leverage value but are close to three chemicals of the training set (rhombus), slightly influential in the model development. The predictions for these test chemicals can be considered as reliable as those of the training chemicals. Chemical 283 is wrongly predicted ($r_i > 3\sigma$), but in this case it belongs to the model AD, being within the cutoff leverage value. Therefore, although the predicted response for chemical 347 should not be accepted because not reliable, prediction for chemical 283 should be.

Another approach to AD evaluation is based on the similarity/diversity, evaluated in the model descriptor space, of the considered compound with respect to those belonging to the training set; in fact, a QSAR/QSPR prediction should be reliable if the compound is – in some way – similar to one or more compounds present in the training set.[184] High similarity is simply another way to use the interpolation ability of the model in place of the extrapolation.

A stepwise procedure was also proposed[177] based on a four stage procedure. General parametric requirements are imposed in the first stage, specifying in the domain only those chemicals that fall in the range of variation of the physicochemical properties of the chemicals in the training set. Such properties (e.g., molecular weight, absorption, water solubility, and volatility) are not usually the driving forces for the studied phenomenon, but they may implicitly affect the measured endpoint, for example, by reducing the bioavailability of chemicals. The second stage defines similarity measures that can be used to quantify the structural similarity between pairs of molecules. Atom-centered fragments are the molecular descriptors used to determine such a similarity. The third stage in defining the domain is based on a mechanistic understanding of the modeled phenomenon. This goal is very difficult to reach because structure and mathematical formalism of the model, computational method used for its derivation, accepted hypotheses, and so forth should be taken into account. The suggested approach is an attempt to reduce the diversity in this matter, where the analysis is focused on functional

groups whose reactivity modulates the studied endpoint and structural fragments used in group contribution models. Finally, the reliability of simulated metabolism (metabolites, pathways, and maps) is taken into account in assessing the reliability of predictions, if metabolic activation of chemicals is a part of the QSAR model.

In any case, regardless of the specific method chosen for AD evaluation, this is always a very important task in order to avoid unreliable predictions and a misuse of the results.

Volume 4

M. Forina, ... P. Oliveri, in Comprehensive Chemometrics, 2009

4.04.4.3.3 Class-modeling techniques

4.04.4.3.3(i) Sensitivity and specificity

A class model is characterized by two parameters: the sensitivity, experimental percent estimate of Correct I decisions (a sample of a class is accepted by the model of the class), that is, of type I errors, and the specificity, experimental percent estimate of Correct II decisions (a sample of other classes is rejected by the model of the class):

	Null hypothesis H_0 TRUE	**Null hypothesis H_0 FALSE**
Statistical decision: Reject H_0	Type I error	Correct II decision
Statistical decision: Do not reject H_0	Correct I decision	Type II error

Therefore, a perfect class model has 100% sensitivity and 100% specificity.

The words sensitivity and specificity were introduced in 1986[57] and are largely used in the literature on CMT. They are not easily accepted by many analytical chemists, because they have different meaning in ordinary analytical chemistry. Moreover, sometimes, people use the same words in classification and prediction, where Correct I decision is “a sample belonging to a class is classified correctly in its own class”. Therefore, generally, one should use the words classification sensitivity, ..., class-modeling specificity. In the following, sensitivity and specificity indicate class-modeling sensitivity and class-modeling specificity.

Specificity and sensitivity are generally evaluated on the final model, with all the objects in the training set, without a measure of the predictive value, which we can be obtained by means of CV, as

1. α% *CV sensitivity*, the percent of the objects in the CV evaluation sets accepted by the models developed with the objects in the training set; with the class boundary determined with α% confidence level;
2. α% *CV specificity*, the percent of the objects of other categories (both of the training and of the CV evaluation sets) rejected by the models developed with the objects in the training set.

α is generally 95%. It can be modified to obtain well-balanced models, with close values of CV sensitivity and specificity.

Generally, CV models are smaller than the model built with all the objects. So, their specificity is larger than that of the final model. Therefore, specificity is the only parameter that CV can overestimate. On the contrary, CV sensitivity is generally worse than the final sensitivity.

As CMTs are important for real problems, especially for multivariate quality control, process control, and fraud detection, class models must be developed taking into account the practical requirements.

In quality control, only sensitivity is important, and it must adequate for the problem: a 95% sensitivity means that about 5% samples are discarded, too much.

Moreover, the statistical parameters used to define the class boundary for sensitivity are the mean of single-variable parameters. The Hotelling T^2 variable is the sum of many t^2 independent variables, the axes of the PCs. In a sample, a single variable can have a very abnormal value, but T^2 can be relatively small because all the other variables are well within the allowed range. So the sample is accepted by the multivariate class model. In the case where the abnormal variable is very important to define the quality, the answer of the multivariate class model is not acceptable. So it is necessary to recognize that multivariate class models can detect abnormal samples that the univariate control cannot detect, owing to the existence of anomalous relationships between the variables, but at the same time the univariate control remains important.

In fraud detection, both sensitivity and specificity are important. When the objective is the protection of a typical food, the users of a model are frequently the members of a consortium. The food produced by the consortium passes an internal quality control, and it is by definition acceptable, so that a sample produced and controlled by the consortium cannot be rejected by the model. In this case, the sensitivity should be close to 100%. Nowadays, models can be obtained with chemical quantities without practical importance to the quality of a food. For example, trace metals have been found useful to detect the origin of an animal or vegetal food. However, the abundance of trace metals in the soil can depend very much on the nonuniformity of lithology and on manuring. So anomalies, which are also important, in these elements cannot be used to decide that the sample must be rejected by the model.

In fraud detection, specificity must be evaluated with objects representative of the real problem. No one tries to sell high quality or expensive foods with a false label. The possible imitations of a typical food generally have a lower quality or cost. To obtain a really useful model for fraud detection, the samples must represent not only the food to be protected but also the possible imitations.

4.04.4.3.3(ii) Drawbacks and positive characteristics of modeling techniques

UNEQ and SIMCA are the two well-known CMT in food chemistry. UNEQ[57] is the name used in chemometrics to indicate the original Hotelling technique,[58] and the class model is built under the hypothesis of a normal multivariate distribution. The class center is a point, the centroid of the class. The boundary of the class model is obtained by the T^2 statistics. The SIMCA model is based on the range of the scores on the significant PCs (the inner space) of the studied category. The center of the model is a parallelepiped in the inner space. The distance from the center has two components, one in the inner space and one in the outer space, the space of noise. The overall distance is evaluated by Fisher statistics to obtain the class model.

Class models can also be obtained by means of the kernel estimate of the multivariate probability density, the potential functions methods.

Recently,[59] a method, known as multivariate range modeling (MRM) has been suggested, where the class center is a parallelepiped (delimited by the range of the variables) in the space of the original variables and in selected directions of high discriminant power.

Chemometricians are also working to modify ANN to be used as CMT.[60] Also classification trees can be transformed in class-modeling methods.

UNEQ is rather sensitive to heavy deviation from the hypothesis of normal multivariate distribution, but relatively insensitive to noisy variables with regard to sensitivity and specificity. SIMCA is presented as a distribution-free technique. However, the objects should fill more or less uniformly the inner space. It is very sensitive to noisy variables with regard to specificity, which decreases with the number of noisy variables. MRM has the advantage of 100% sensitivity of the final model, by definition. Noisy variables increase the apparent specificity of MRM models. The evaluation of CV specificity is very important in this case. To reduce the effect of noisy variables, the range of MRM must be expanded.

In practical food problems, it is very important to obtain specifications, that is, rules that translate in a simple language the class models. Simplicity is a requisite, because food experts usually cannot understand an explicit model based on many variables. Classification trees and MRM seem useful to obtain specifications.

In the case of data set Wines, we obtained the specification (values refer to the autoscaled variables) by means of a classification tree where the LDA canonical functions computed for each combination of two variables were added to the original variables.

If

0.743 × Flavanoids – 0.670 × Color intensity > –0.68

And

–0.393 × Glycerol – 0.920 × Proline < –0.235

And

Fixed acidity > –1.57

Then

BAROLO (probability 98% for authentic Barolo samples)