ELSEVIER

Journal of Statistical Planning and Inference 93 (2001) 211–223

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Multiple use one-sided hypotheses testing in univariate linear calibration

K. Krishnamoorthy[a], Pandurang M. Kulkarni[b], Thomas Mathew[c, *]

[a] *Department of Mathematics, University of Southwestern Louisiana, Lafayette, LA 70504, USA*
[b] *Department of Mathematics and Statistics, University of South Alabama, Mobile, AL 36688, USA*
[c] *Department of Mathematics and Statistics, University of Maryland, 1000 Hilltop Circle, Baltimore, MD 21250, USA*

Received 28 March 1999; received in revised form 7 February 2000; accepted 16 June 2000

Abstract

Consider a normally distributed response variable, related to an explanatory variable through the simple linear regression model. Data obtained on the response variable, corresponding to known values of the explanatory variable (i.e., *calibration data*), are to be used for testing hypotheses concerning unknown values of the explanatory variable. We consider the problem of testing an unlimited sequence of one-sided hypotheses concerning the explanatory variable, using the corresponding sequence of values of the response variable and the same set of calibration data. This is the situation of multiple use of the calibration data. The tests derived in this context are characterized by two types of uncertainties: one uncertainty associated with the sequence of values of the response variable, and a second uncertainty associated with the calibration data. We derive tests based on a condition that incorporates both of these uncertainties. The solution has practical applications in the decision limit problem. We illustrate our results using an example dealing with the estimation of blood alcohol concentration based on breath estimates of the alcohol concentration. In the example, the problem is to test if the unknown blood alcohol concentration of an individual exceeds a threshold that is safe for driving.

MSC: 62F03; 62J05

Keywords: Calibration curve; Calibration data; Decision limit; Detection limit; Multiple use; Non-central t-distribution

1. Introduction

The problems addressed in this article are in the set up of a linear regression model where normality is assumed for the dependent variable. Thus, let $y_1, y_2, \ldots, y_N$ be N

* Corresponding author.
E-mail address: mathew@math.umbc.edu (T. Mathew).

0378-3758/01/$ - see front matter
PII: S0378-3758(00)00164-6

independent observations on a response variable corresponding to the values $x_1, x_2, \ldots, x_N$, respectively, of an explanatory variable where,

$$y_j \sim \mathrm{N}(\alpha + \beta x_j, \sigma^2), \tag{1.1}$$

$j = 1, 2, \ldots, N$. The x_j's are assumed to be non-random quantities. Now let y_0 be another observation corresponding to an unknown value, say θ, of the explanatory variable, where y_0 is independent of the y_j's in (1.1). Thus, we have

$$y_0 \sim \mathrm{N}(\alpha + \beta\theta, \sigma^2). \tag{1.2}$$

The problem of *calibration*, or *inverse regression*, deals with statistical inference concerning θ. The set of y_j's in (1.1), corresponding to the known values x_j ($j = 1, 2, \ldots, N$), is referred to as the *calibration data*. The relationship between the y_j's and the corresponding x_j's is known as the *calibration curve*. We are thus dealing with the situation where the calibration curve is a straight line.

Most of the available literature on the calibration problem deals with point estimation of θ, or the construction of confidence regions for θ. The purpose of this article is to address some hypothesis testing problems concerning θ. More specifically, we shall consider the following testing problems:

$$\text{(i) } \mathrm{H}_0\colon \theta \geqslant c \text{ vs. } \mathrm{H}_1\colon \theta < c, \qquad \text{(ii) } \mathrm{H}_0\colon \theta \leqslant c \text{ vs. } \mathrm{H}_1\colon \theta > c, \tag{1.3}$$

where c is a known number, specified in advance. Our investigation of this problem was motivated by the following application.

1.1. Estimation of blood alcohol concentration – an example

This example deals with the calibration of breath estimates of blood alcohol concentration, based on the results of a laboratory test. Here, the y_j's are breath estimates of blood alcohol concentrations and the x_j's are the actual concentration of alcohol in the blood, obtained by a laboratory test. The relevant data, given in Section 3, are based on a study conducted at Acadiana Criminalstics Laboratory, New Iberia, Louisiana. It turns out that the model (1.1) provides an adequate fit for the data. We can use models (1.1) and (1.2) in order to estimate an unknown blood alcohol concentration θ, after obtaining the corresponding breath estimate y_0. Such an estimation is preferable to the actual laboratory determination of the blood alcohol concentration, since it is much easier and faster to obtain the breath estimates. In this context, we would also like to test whether an unknown blood alcohol concentration θ is less than or equal to 0.10 or if it exceeds 0.10, since the legal maximum limit of blood alcohol concentration while driving is 0.10% in many states in the USA (in some states it is 0.08%). This is precisely the testing problem (ii) in (1.3).

The hypothesis testing problems in (1.3) will also arise in the context of decision and detection limit problems in bioassays and chemical assays. Suppose we want to determine the concentration of a chemical in a sample where a direct determination of the concentration may be difficult or expensive. On the other hand, it may be easier

to obtain an indirect measurement, for example, spectral measurements. First, several samples having known concentrations (the x_j's) are prepared, and the corresponding indirect measurements (the y_j's) are obtained. Assuming that relationship (1.1) holds, we can use this data for statistical inference concerning an unknown concentration, say θ, in a sample, after obtaining the corresponding indirect measurement y_0. In this context, the testing problems in (1.3) will arise if we want to verify whether the unknown concentration is above a threshold, or if it is below. This is precisely the decision limit problem. We refer to Currie (1988) for a detailed discussion of this problem as it arises in analytical chemistry. The testing problems (1.3) in the context of bioassays are addressed, for example, in Dunne (1995). A good discussion of decision and detection limits, along with environmental applications, can be found in Gibbons (1994, Chapter 5). Other examples and applications are given in a number of books and articles; see the book by Brown (1993) and the article by Osborne (1991) for a review.

For models (1.1) and (1.2), appropriate tests for the hypotheses in (1.3) can be derived under two different situations. The first is when the calibration data are used only once for the purpose of testing hypothesis. In the context of the chemical example mentioned above, this means that the calibration data are collected and used to test hypothesis concerning an unknown concentration of a single sample. A more realistic scenario is one where the same set of calibration data are used repeatedly in order to test hypotheses concerning a sequence of θ-values, one at a time, after observing the corresponding sequence of responses. This is the situation of *multiple use* of the calibration data. In other words, we have a sequence of independent responses y_{0i}, corresponding to a sequence of θ values θ_i, following the model

$$y_{0i} \sim \mathrm{N}(\alpha + \beta\theta_i, \sigma^2), \tag{1.4}$$

similar to (1.2), $i = 1, 2, 3, \ldots$. Correspondingly, we have the following sequence of hypotheses to be tested:

$$\text{(i) } \mathrm{H}_{0i}\text{: } \theta_i \geqslant c_i \text{ vs. } \mathrm{H}_{1i}\text{: } \theta_i < c_i, \qquad \text{(ii) } \mathrm{H}_{0i}\text{: } \theta_i \leqslant c_i \text{ vs. } \mathrm{H}_{1i}\text{: } \theta_i > c_i, \tag{1.5}$$

where the c_i's are known scalars. For instance, in Example 1, we want to verify if for every individual tested, the blood alcohol concentration is no more than 0.10, or no more than 0.08, where the value 0.10 or 0.08 is to be used depending on the legal maximum limit in a state. In other words, we have hypotheses of the type (ii) in (1.5), where the c_i's can take two different values, 0.10 and 0.08. In the decision limit problem, we may be interested in testing if the concentration of a chemical or a pollutant in different samples exceeds a particular safety threshold. Such tests may have to be done a large number of times, as and when a sample is obtained. In other words, we need to test a sequence of hypotheses of the type (i) or (ii) in (1.5). It is also clear that in many applications, the c_i's in (1.5) will have a common value.

Tests that we shall derive for testing the sequence of hypotheses in (1.5) will have two types of uncertainty statements associated with them. One with respect to the sequence of responses y_{0i} in (1.4), and a second uncertainty statement with respect to

the calibration data, i.e., the y_j's in (1.1). This is formally explained in Section 2. The distinction between single use and multiple use of the calibration data are also made for the construction of confidence regions. In fact confidence regions that involve multiple use of the calibration data are derived subject to two types of uncertainty statements, similar to those mentioned above. We refer to the recent article by Mee and Eberhardt (1996) for an excellent discussion on this. The results in Dunne (1995) deal with the set up of single use of the calibration data. In fact, Dunne (1995) provides solutions to the testing problems in (1.3) in the single use situation. Consequently, in this article, we shall study the problem for the case of multiple use of the calibration data.

In the next section, we shall derive the tests. The two uncertainty statements, mentioned earlier, are explicitly stated and tests are derived subject to these. In our derivations, we have assumed that $\beta > 0$ in (1.1). Essentially, this amounts to assuming that the sign of β is known. This is clearly a reasonable assumption since the nature of the dependence of the response variable on the explanatory variable will be known in actual applications. In case $\beta < 0$, we can assume that the slope in (1.1) is $-\beta$, after multiplying y_j by -1. In fact, the tests (i.e., decision limits) derived in Dunne (1995) use the fact that the sign of β is known. In Section 3, we apply our results to the alcohol concentration data, mentioned in example 1. Some concluding remarks appear in Section 4.

2. The test

Among the two hypotheses in (1.5), here we shall consider only the sequence (i). Once we derive tests for testing (i), similar results can be obtained for testing the sequence (ii) in (1.5).

Let $\hat{\alpha}$ and $\hat{\beta}$ denote the least-squares estimators of α and β based on the y_j's in (1.1), and let $\hat{\sigma}^2$ denote the unbiased estimator of σ^2 based on the residual sum of squares. Then

$$\hat{\beta} = \frac{\sum_{j=1}^{N}(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{N}(x_j - \bar{x})^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \text{and}$$

$$\hat{\sigma}^2 = \frac{1}{N-2}\sum_{j=1}^{N}(y_j - \hat{\alpha} - \hat{\beta}x_j)^2, \tag{2.1}$$

where $\bar{y}$ and $\bar{x}$ denote the averages of the y_j's and the x_j's, respectively. As pointed out in the introduction, we assume that $\beta > 0$.

2.1. Motivation of the test statistic

Suppose α, β and σ^2 are known. Then, from (1.2),

$$\hat{\theta} = \frac{y_0 - \alpha}{\beta} \sim \mathrm{N}\left(\theta, \frac{\sigma^2}{\beta^2}\right).$$

Hence, we shall reject H_0: $\theta \geqslant c$ in favor of H_1: $\theta < c$ if $(\hat{\theta} - c)/(\sigma/\beta)$ is small, or equivalently, if $(y_0 - \alpha - \beta c)/\sigma$ is small. In order to arrive at the latter conclusion, we have used the expression for $\hat{\theta}$ (given above), along with our assumption $\beta > 0$. In practice, usually α, β and σ^2 are unknown, and we shall replace them by their estimators. Thus, let

$$T = \frac{y_0 - \hat{\alpha} - \hat{\beta}c}{\hat{\sigma}}.$$

Our test for H_0: $\theta \geqslant c$ will reject H_0 in favor of H_1: $\theta < c$ if T is small. Hence, for testing the sequence of hypotheses (i) in (1.5), consider the test statistic

$$T_i = \frac{y_{0i} - \hat{\alpha} - \hat{\beta}c_i}{\hat{\sigma}}. \tag{2.2}$$

We shall reject H_{0i}: $\theta_i \geqslant c_i$ in favor of H_{1i}: $\theta_i < c_i$ if

$$T_i < k_i, \tag{2.3}$$

$i = 1, 2, 3, \ldots$, where the k_i's are to be determined. Note that the test statistic T_i depends on the response y_{0i} and also on the calibration data y_j, $j = 1, 2, \ldots, N$, through $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$. Furthermore, (2.3) represents a sequence of rejection regions, for $i = 1, 2, 3, \ldots$.

2.2. Criterion for the derivation of the test

The derivation of our test amounts to the computation of the k_i's in (2.3). We shall now explain the condition to be satisfied by the k_i's. Since the calibration data will be used repeatedly for testing a sequence of hypotheses, consider the proportion of H_{0i}'s that are rejected, when they are true, conditionally given the calibration data, i.e., conditionally given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$. In other words, conditionally given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$, consider the proportion of times $T_i < k_i$, under H_{0i}: $\theta_i \geqslant c_i$, $i = 1, 2, 3, \ldots$. The k_i's are to be determined so that with a high probability, the calibration data will guarantee that this proportion is small (i.e., less than a specified bound), for all θ_i's under the corresponding null hypothesis H_{0i}. Note that given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$,

$$T_i \sim \mathrm{N}\left[\frac{\alpha - \hat{\alpha} + \beta\theta_i - \hat{\beta}c_i}{\hat{\sigma}}, \frac{\sigma^2}{\hat{\sigma}^2}\right].$$

Hence, conditionally given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$, $T_i < k_i$ is equivalent to

$$z_{0i} < \left[k_i - \frac{\alpha - \hat{\alpha} + \beta\theta_i - \hat{\beta}c_i}{\hat{\sigma}}\right]\frac{\hat{\sigma}}{\sigma}, \tag{2.4}$$

where $z_{0i} \sim \mathrm{N}(0, 1)$. We shall now give an expression for the proportion of H_{0i} that are rejected (i.e., proportion of times $T_i < k_i$), conditionally given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$. Let

$$\Delta_i = \begin{cases} 1 & \text{if } T_i < k_i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Conditionally given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$, the Δ_i's are independent Bernoulli random variables with success probability, say p_i, given by

$$p_i = P(T_i < k_i | \hat{\alpha}, \hat{\beta}, \hat{\sigma}) = \Phi\left\{k_i \frac{\hat{\sigma}}{\sigma} - \frac{\alpha - \hat{\alpha} + \beta\theta_i - \hat{\beta}c_i}{\sigma}\right\}, \tag{2.6}$$

where Φ denotes the standard normal cdf. In order to arrive at (2.6), we have used the fact that conditionally given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$, $T_i < k_i$ is equivalent to (2.4). If a sequence of n tests are carried out, then the proportion of times $T_i < k_i$ is $(1/n)\sum_{i=1}^{n} \Delta_i$. By the strong law of large numbers, we get

$$\frac{1}{n}\sum_{i=1}^{n} \Delta_i \to \frac{1}{n}\sum_{i=1}^{n} p_i,$$

as n becomes large. In other words, conditionally given $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}$, the proportion of H_{0i} that are rejected is $(1/n)\sum_{i=1}^{n} p_i$, when n is large. We want this proportion to be small, with a high probability, for all θ_i under the corresponding null hypotheses H_{0i}: $\theta_i \geqslant c_i$, $i = 1, 2, 3, \ldots$. That is, the k_i's must satisfy

$$P_{\hat{\alpha},\hat{\beta},\hat{\sigma}}\left[\frac{1}{n}\sum_{i=1}^{n} p_i \leqslant \varepsilon\right] = \gamma, \tag{2.7}$$

for all θ_i under the corresponding null hypotheses H_{0i}: $\theta_i \geqslant c_i$, $i = 1, 2, 3, \ldots, n$, where ε is a specified bound (say, $\varepsilon = 0.05$) and γ is the chosen probability (say, $\gamma = 0.95$). Due to our assumption $\beta > 0$, it follows from the expression for p_i in (2.6) that p_i is maximum under H_{0i}: $\theta_i \geqslant c_i$ when $\theta_i = c_i$. Hence (2.7) holds for all θ_i satisfying $\theta_i \geqslant c_i$ if and only if it holds for $\theta_i = c_i$. Using expression for p_i in (2.6), we thus get the following condition to be satisfied by the k_i's:

$$P_{\hat{\alpha},\hat{\beta},\hat{\sigma}}\left[\frac{1}{n}\sum_{i=1}^{n} \Phi\left\{k_i \frac{\hat{\sigma}}{\sigma} - \frac{\alpha - \hat{\alpha} + (\beta - \hat{\beta})c_i}{\sigma}\right\} \leqslant \varepsilon\right] = \gamma. \tag{2.8}$$

Condition (2.8) can be interpreted as the "size of the tests" being small, i.e., less than or equal to ε, with probability γ. We note that (2.8) must actually hold for every n.

Thus the criterion to be used for the derivation of the test, i.e., the computation of the k_i's, is the condition (2.8), which must hold for all $n \geqslant 1$. As is clear from the derivation of (2.8), this condition reflects the fact that the same calibration data are used a large number of times for testing hypotheses. The fact that the same calibration data are used repeatedly is also the reason why we do not want to consider the unconditional distribution of the T_i's in (2.2).

2.3. Computation of the k_i's

Note that

$$\alpha - \hat{\alpha} + (\beta - \hat{\beta})c_i = (\alpha + \beta\bar{x} - \bar{y}) + (\beta - \hat{\beta})(c_i - \bar{x}).$$

Define

$$z_1 = \left[\sum_{j=1}^{N}(x_j - \bar{x})^2\right]^{1/2}(\beta - \hat{\beta})/\sigma \sim \mathrm{N}(0,1), \quad z_2 = \sqrt{N}(\bar{y} - \alpha - \beta\bar{x})/\sigma \sim \mathrm{N}(0,1),$$

$$u^2 = (N-2)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-2} \quad \text{and} \quad c_{1i} = \frac{(c_i - \bar{x})}{[\sum_{j=1}^{N}(x_j - \bar{x})^2]^{1/2}}. \tag{2.9}$$

The random variables z_1, z_2 and u in (2.9) are independently distributed, and χ^2_r denotes the central chi-square distribution with r df. The k_i that we compute will be a function of c_{1i} given in (2.9). Hence, from now on we shall use the notation $k(c_{1i})$ instead of k_i. Using the quantities in (2.9), (2.8) simplifies to

$$P_{z_1,z_2,u}\left[\frac{1}{n}\sum_{i=1}^{n}\Phi\left\{k(c_{1i})\frac{u}{\sqrt{N-2}} + z_1 c_{1i} + \frac{z_2}{\sqrt{N}}\right\} \leqslant \varepsilon\right] = \gamma. \tag{2.10}$$

It appears difficult to obtain $k(c_{1i})$ explicitly, except in the special case when the c_{1i}'s are all equal. We shall first consider this special case. From the expression for c_{1i}, given in (2.9), it follows that the c_{1i}'s are all equal if and only if the c_i's are all equal. As already pointed out, the c_i's will be equal in many practical applications; see the discussion following (1.5). When the c_i's are all equal, let c denote the common value of the c_i's and let c_0 denote the common value of the c_{1i}'s. Hence, from (2.9),

$$c_0 = \frac{(c - \bar{x})}{[\sum_{j=1}^{N}(x_j - \bar{x})^2]^{1/2}}. \tag{2.11}$$

In this special case, (2.10) simplifies to

$$P_{z_1,z_2,u}\left[\Phi\left\{k(c_0)\frac{u}{\sqrt{N-2}} + z_1 c_0 + \frac{z_2}{\sqrt{N}}\right\} \leqslant \varepsilon\right] = \gamma, \tag{2.12}$$

or, equivalently,

$$P_{z_1,z_2,u}\left[k(c_0)\frac{u}{\sqrt{N-2}} + z_1 c_0 + \frac{z_2}{\sqrt{N}} \leqslant z(\varepsilon)\right] = \gamma, \tag{2.13}$$

where $z(\varepsilon)$ denotes the 100εth percentile of the standard normal distribution. We note that condition (2.12), or equivalently, (2.13), is essentially the condition to be satisfied by one-sided tolerance intervals. Indeed, we can arrive at (2.12) by starting with the expression for a one-sided tolerance interval and using the fact that $\beta > 0$. The solution to the one-sided tolerance interval problem is already known; see Guttman (1970, pp. 87–89). The solution to $k(c_0)$ is given by

$$k(c_0) = \left(\frac{1}{N} + c_0^2\right)^{1/2} t_{1-\gamma}\left[N-2; \frac{z(\varepsilon)}{(1/N + c_0^2)^{1/2}}\right], \tag{2.14}$$

where $t_{1-\gamma}[r;\lambda]$ denotes the $100(1-\gamma)$th percentile of the noncentral t-distribution with df $= r$ and noncentrality parameter λ. Thus, in the special case where the c_i's are all equal to a common value c, $k(c_0)$ in (2.14) satisfies (2.12), where c_0 is given in (2.11).

When the c_{1i}'s are not all equal, it is not possible to obtain $k(c_{1i})$ explicitly. In what follows, we shall explore the following possibility. Let $k(c_{1i})$ satisfy

$$P_{z_1,z_2,u}\left[\Phi\left\{k(c_{1i})\frac{u}{\sqrt{N-2}}+z_1c_{1i}+\frac{z_2}{\sqrt{N}}\right\}\leqslant\varepsilon\right]=\gamma. \tag{2.15}$$

In other words,

$$k(c_{1i})=\left(\frac{1}{N}+c_{1i}^2\right)^{1/2}t_{1-\gamma}\left[N-2;\frac{z(\varepsilon)}{(1/N+c_{1i}^2)^{1/2}}\right]. \tag{2.16}$$

The expression in (2.16) follows since (2.15) is similar to (2.12). We shall numerically investigate whether $k(c_{1i})$ in (2.16) will satisfy (2.10) for various sequences $\{c_{1i}\}$. Note that (2.15) is a tolerance interval condition and is quite different from (2.10). In a recent article on the construction of confidence regions in the calibration problem, Mee and Eberhardt (1996) have indicated that intervals obtained using a tolerance interval condition may satisfy the requirements of multiple use confidence regions. Motivated by this, we shall explore whether $k(c_{1i})$ in (2.16) will satisfy (2.10).

2.4. Numerical results for $k(c_{1i})$ in (2.16)

We have carried out some numerical results to check if $k(c_{1i})$ in (2.16) will satisfy (2.10). For this purpose, we need to simulate the left-hand side (lhs) of (2.10) using $k(c_{1i})$ given in (2.16), for various sequences $\{c_{1i}\}$. Note that when the c_{1i}'s are all equal, i.e., when the c_i's are all equal, such a numerical study is not necessary, since, in this case, $k(c_0)$ satisfies (2.12), where c_0 denotes the common value of the c_{1i}'s. Thus, the numerical results that we report here are only for the case when the c_{1i}'s are unequal. We shall first discuss our choices for the sequence $\{c_{1i}\}$. When the calibration data are collected based on a carefully designed experiment, then the x_j's in (1.1) will cover the range of practical interest. Hence, it is reasonable to assume that the θ_i's in (1.4) belong to the range determined by the x_j's, i.e.,

$$\min(x_j)\leqslant\theta_i\leqslant\max(x_j), \tag{2.17}$$

$i=1,2,3,\ldots$. Consequently, when (2.17) holds, the c_i's in (1.5) should also belong to the range of the x_j's. In other words, when we test the null hypothesis H_{0i}: $\theta_i\geqslant c_i$, we are testing if θ_i is greater than or equal to a specified value in the range of the x_j's. Hence, we shall assume

$$\min(x_j)\leqslant c_i\leqslant\max(x_j), \tag{2.18}$$

$i=1,2,3,\ldots$. From the definition of c_{1i} given in (2.9), (2.18) gives

$$\frac{\min(x_j)-\bar{x}}{[\sum_{j=1}^N(x_j-\bar{x})^2]^{1/2}}\leqslant c_{1i}\leqslant\frac{\max(x_j)-\bar{x}}{[\sum_{j=1}^N(x_j-\bar{x})^2]^{1/2}}, \tag{2.19}$$

$i=1,2,3,\ldots$. In particular, (2.19) implies that

$$-1\leqslant c_{1i}\leqslant 1. \tag{2.20}$$

Table 1

a	γ			
	0.99	0.95	0.90	0.85
(a) Simulated values of (2.22) for $N=20$ and $\varepsilon=0.05$, for different values of γ and a				
2	0.987	0.936	0.852	0.777
1	0.991	0.950	0.889	0.826
0.5	0.993	0.959	0.916	0.857
0.1	0.990	0.952	0.902	0.849
(b) Simulated values of (2.22) for $N=20$ and $\varepsilon=0.01$, for different values of γ and a				
2	0.985	0.923	0.848	0.753
1	0.991	0.938	0.865	0.799
0.5	0.991	0.952	0.885	0.824
0.1	0.992	0.949	0.897	0.857

The actual bounds for c_{1i}, given in (2.19), could be much narrower than $[-1,1]$. In any case, in order to investigate whether $k(c_{1i})$ in (2.16) will satisfy (2.10), it appears reasonable to restrict c_{1i}'s in the interval $[-1,1]$. In our numerical results, we have also considered the wider interval $[-2,2]$.

For various sequences $\{c_{1i}\}$, $i=1,2,\ldots,$ 10,000, we have simulated the following quantity.

$$P_{z_1,z_2,u}\left[\frac{1}{10,000}\sum_{i=1}^{10,000}\Phi\left\{k(c_{1i})\frac{u}{\sqrt{N-2}}+z_1c_{1i}+\frac{z_2}{\sqrt{N}}\right\}\leqslant\varepsilon\right], \tag{2.21}$$

based on 10,000 simulations. The IMSL Fortran function subroutine TNIN is used to compute $k(c_{1i})$. Note that (2.21) is an approximation to the lhs of (2.10). Since (2.10) holds exactly when the c_{1i}'s are all equal, (2.10) is expected to hold approximately when the c_{1i}'s are nearly equal. Thus, it is necessary to investigate whether (2.10) will hold when the c_{1i}'s are as unequal as possible. In any finite sequence $\{c_{1i}\}$, with $c_{1i}\in[-a,a]$, the c_{1i}'s are as unequal as possible, i.e., the variance among them is a maximum, when half of the c_{1i}'s are equal to $-a$ and the remaining half are equal to a. In this case, (2.21) reduces to

$$P_{z_1,z_2,u}\left[\frac{1}{2}\left(\Phi\left\{k(-a)\frac{u}{\sqrt{N-2}}-z_1a+\frac{z_2}{\sqrt{N}}\right\}+\Phi\left\{k(a)\frac{u}{\sqrt{N-2}}+z_1a+\frac{z_2}{\sqrt{N}}\right\}\right)\leqslant\varepsilon\right]. \tag{2.22}$$

The following IMSL subroutines are used in the numerical studies: TNIN to compute $k(c_{1i})$; RNNOA to generate normal variates; and RNCHI to generate chi-square variates. For $N=20$, $\varepsilon=0.05$ and 0.01, $\gamma=0.99,0.95,0.90$ and 0.85, and $a=2,1,0.5$ and 0.1, Table 1 gives the simulated values of (2.22), based on 10,000 simulations. When the sequence $\{c_{1i}\}$, $i=1,2,\ldots,$ 10,000 consists of 10,000 equispaced values in the interval

Table 2
The simulated probabilities (2.21) for $N = 20$ when the sequence $\{c_{1i}\}$ assumes 10,000 equispaced values in the interval $[-a, a]$

$[-a, a]$	γ			
ε	0.99	0.95	0.90	0.85
$[-2, 2]$	0.05	0.989	0.954	0.902	0.852
$[-1, 1]$	0.05	0.992	0.951	0.898	0.851
$[-2, 2]$	0.01	0.991	0.954	0.907	0.846
$[-1, 1]$	0.01	0.993	0.947	0.902	0.851

Table 3
The simulated probabilities (2.21) for $N = 20$ when the sequence $\{c_{1i}\}$ is generated based on N(0, 1), $i = 1, 2, \ldots, 10{,}000$

ε	γ			
	0.99	0.95	0.90	0.85
0.05	0.989	0.950	0.897	0.832
0.01	0.992	0.951	0.899	0.834

$[-2, 2]$ and $[-1, 1]$, the simulated values of (2.21) appear in Table 2. If 10,000 c_{1i}'s are randomly generated according to a standard normal distribution, the simulated values of (2.21) appear in Table 3.

The simulation results indicate that $k(c_{1i})$ in (2.16) will satisfy (2.10) reasonably well at least in situations where the range for c_{1i}, given in (2.19), is a narrow interval. As expected, (2.10) is more satisfactorily met for narrower intervals. As already pointed out, it appears quite reasonable to assume that $-1 \leqslant c_{1i} \leqslant 1$. In practice, the range for c_{1i} is likely to be much narrower, especially when N is somewhat large. This should be clear from (2.19). In terms of meeting the condition (2.10), the results are somewhat unsatisfactory in the set up of Table 1, especially for $\gamma = 0.85$. However, Table 1 represents the 'worst-case scenario', since (2.22) corresponds to the situation where half of the c_{1i}'s are equal to $-a$ and the remaining half are equal to a (for $a = 2, 1, 0.5, 0.1$). In other words, among the sequence of hypotheses being tested, half of the time we are testing if the parameter is greater than or equal to one extreme of the parameter space (i.e., $-a$) and the other half deals with testing if the parameter is greater than or equal to the other extreme (i.e., a). In actual applications, this appears unrealistic. Nevertheless, the purpose of reporting the numerical results in Table 1 is to see of (2.10) (i.e, (2.22)) is met in this extreme situation. For large values of γ, the results are quite satisfactory, especially when a is small. In any case, the numerical results in Tables 1–3 provide evidence regarding the extent to which (2.16) will satisfy (2.10). The overall conclusion is that in practical applications that call for multiple use of the calibration data for testing the hypotheses in (1.5), $k(c_{1i})$ given in (2.16) is quite adequate for meeting the requirement (2.10).

Table 4
Blood alcohol concentrations data, estimates and test statistics

	y_i	x_i	$\hat{x}_i$	T_i	$k_1(c_0)$
1	0.145	0.160	0.149	3.40	2.67
2	0.156	0.170	0.160	4.18	2.67
3	0.181	0.180	0.189	6.00	2.67
4	0.108	0.100	0.112	0.85	2.68
5	0.180	0.170	0.189	6.27	2.67
6	0.112	0.100	0.117	1.19	2.68
7	0.081	0.060	0.086	−1.11	2.69
8	0.104	0.100	0.108	0.53	2.68
9	0.176	0.170	0.185	5.78	2.67
10	0.048	0.056	0.047	−3.56	2.69
11	0.092	0.111	0.093	−0.47	2.68
12	0.144	0.162	0.147	3.38	2.67
13	0.121	0.143	0.124	1.73	2.67
14	0.065	0.079	0.065	−2.42	2.68
15	0.000	0.006	−0.006	−7.05	2.71

2.5. *Testing the hypothesis* 1.5(*ii*)

Tests for the hypotheses (ii) in (1.5) can be similarly derived. The rejection regions now take the form $T_i > k_1(c_{1i})$, where T_i is given in (2.2) and c_{1i} is given in (2.9). When the c_{1i}'s are all equal to c_0, $k_1(c_0)$ is given by

$$k_1(c_0) = \left(\frac{1}{N} + c_0^2\right)^{1/2} t_\gamma \left[N - 2; \frac{z(1-\varepsilon)}{(1/N + c_0^2)^{1/2}}\right], \tag{2.23}$$

similar to (2.14). Also, $k_1(c_{1i})$ has an expression similar to that of $k(c_{1i})$ in (2.16).

3. An example

A study was conducted at Acadiana Criminalstics Laboratory, New Iberia, Louisiana, to compare the breath estimates of blood alcohol concentration with those determined by a laboratory test. A sample of 15 subjects was used. In Table 4, we present the breath estimates (y_j) obtained using Breathalyzer Model 5000 and the results of the laboratory test (x_j). These numbers are percentages of alcohol concentration in blood. Here, we assume that the x_j's are accurately measured, i.e., the measurement errors are small enough to be ignored, so that the x_j's can be assumed to be fixed. A simple linear regression model was fit with y regressed on x based on all 15 pairs of observations. The fitted model is $y = 0.00135 + 0.958x$ and the value of R^2 is 0.93. The normal probability plot of the residuals is reasonably linear and hence the distribution of error terms does not depart from a normal distribution. Since in many states in the USA the legal maximum limit of blood alcohol concentration while driving is 0.1%, we want to test whether $\theta \leqslant 0.10$ or $\theta > 0.10$. In other words, here we shall consider the situation

where the c_i's in (1.5) are all equal and equal to 0.10. In this problem, one may want to control the probability of concluding $\theta > 0.10$ when in fact $\theta \leqslant 0.10$. Thus, we shall consider the following null and alternative hypotheses.

$$H_0: \theta \leqslant 0.10 \quad \text{vs.} \quad H_1: \theta > 0.10. \tag{3.1}$$

We shall carry out the test using $\varepsilon = 0.05$ and $\gamma = 0.95$. From (2.23),

$$k_1(c_0) = \left(\frac{1}{N} + c_0^2\right)^{1/2} t_{0.95}\left[N-2; \frac{z(0.95)}{(\frac{1}{14} + c_0^2)^{1/2}}\right]. \tag{3.2}$$

We reject H_0 for the ith individual if $T_i > k_1(c_0)$.

In Table 4, we give the classical estimates obtained by solving from the fitted model for x, and the values of test statistics T_i. The T_i's are computed using the leave-one-out method; for example to compute the point estimate and the value of the test statistic corresponding to $y_1 = 0.145$ in Table 4, the observation (0.145,0.160) is not used to fit the model. Thus, $N = 14$, in (3.2). Fitting the model by deleting the first observation, we get $\hat{\alpha} = 0.0007$, $\hat{\beta} = 0.97$, $c_0 = -0.077$, $T_i = 3.40$, and $k_1(c_0) = 2.67$. Furthermore, $\hat{x}_1 = y_1 - \hat{\alpha}/\hat{\beta} = (0.145 - 0.0007)/0.97 = 0.149$. This process can be carried out by deleting each observation in Table 4.

We observe from Table 4 that the test results are in agreement with those based on laboratory test except for the 11th and 13th subjects. We also note that for the 15th observation, $x_i = 0.006$ and $\hat{x}_i = -0.006$. This discrepancy may be due to the fact that the corresponding $y_i = 0$ and x_i is very close to zero.

4. Concluding remarks

In this article, we have investigated hypothesis testing in the calibration problem. In the univariate case, we have succeeded in deriving appropriate tests for testing one-sided hypotheses in the situation of multiple use of the calibration data. The results are illustrated using an example. The two-sided testing problem is not considered in this article, since tests in this context can be obtained by inverting the multiple use confidence regions that are available in the literature. One-sided multiple use confidence regions can certainly be developed and can be used for testing the hypotheses in (1.5). The simultaneous tolerance intervals in Mee et al. (1991) can be used to obtain conservative multiple use confidence regions, and hence conservative multiple use tests, for the two-sided testing problem. The one-sided simultaneous tolerance intervals in Odeh and Mee (1990) can similarly be used for obtaining conservative multiple use tests for the one-sided testing problems in (1.5). However, in the present article, our goal has been to investigate whether the tolerance interval condition can be used for carrying out multiple use hypotheses tests. The numerical results in Section 3 show that this is indeed the case. Note that our results are applicable only in the univariate set up and similar results are currently not available in the multivariate case. The problem of testing hypothesis in the multivariate calibration problem is currently under investigation.

Acknowledgements

Mathew's research was supported in part by National Science Foundation Grant DMS 95-30932. Kulkarni's research was supported in part by NASA-JOVE grant. The authors are thankful to Denise Benton for providing useful references related to the example in Section 5 and helping us to get the data, and to Kevin Ardoin for providing the data. They also wish to thank the reviewers for many helpful comments, and for drawing attention to the literature on detection limits.

References

Brown, P.J., 1993. Measurement, Regression and Calibration. Oxford University Press, New York.

Currie, L.A., 1988. Detection in Analytical Chemistry: Importance, Theory and Practice. American Chemical Society, Washington, DC.

Dunne, A., 1995. Decision and detection limits for linear homoscedastic assays. Statist. Med. 14, 1949–1959.

Gibbons, R.D., 1994. Statistical Methods for Groundwater Monitoring. Wiley, New York.

Guttman, I., 1970. Statistical Tolerance Regions, Classical and Bayesian. Charles Griffin & Co. Ltd., London.

Mee, R.W., Eberhardt, K.R., 1996. A Comparison of uncertainty criteria for calibration. Technometrics 38, 221–229.

Mee, R.W., Eberhardt, K.R., Reeve, C.P., 1991. Calibration and simultaneous tolerance intervals for regression. Technometrics 33, 211–219.

Odeh, R.E., Mee, R.W., 1990. One-sided simultaneous tolerance limit for regression. Comm. Statist. Simulation Comput. 19, 663–680.

Osborne, C., 1991. Statistical calibration: a review. Internat. Statist. Rev. 59, 309–336.