Anomalies in the analysis of calibrated data

D.R. Jensen and D.E. Ramirez*

Department of Mathematics, University of Virginia, Charlottesville, VA, USA

(Received 30 November 2006; final version received 18 October 2007)

This study examines the effects of calibration errors on model assumptions and data-analytic tools in direct calibration assays. These effects encompass induced dependencies, inflated variances, and heteroscedasticity among the calibrated measurements, whose distributions arise as mixtures. These anomalies adversely affect conventional inferences, including the inconsistency of sample means; the underestimation of measurement variance; and the distributions of sample means, sample variances, and student’s t as mixtures. Inferences in comparative experiments remain largely intact, although error mean squares continue to underestimate the measurement variances. These anomalies are masked in practice, as conventional diagnostics cannot discern the irregularities induced through calibration. Case studies illustrate the principal issues.

Keywords: direct calibration; induced dependencies; non-standard distributions; diagnostics; case studies

1991 Mathematics Subject Classifications: Primary: 62J05; Secondary: 62E10

1. Introduction

Calibrated measurements, intrinsic to the sciences and engineering, are inherently subject to errors of calibration. These errors induce dependencies in the violation of a basic tenet in much of applied statistics, namely, that observations should be uncorrelated if not independent. These issues traditionally have been overlooked by both scientists and statisticians, despite a century of emerging methodologies for the analysis of experimental data. Not only are many parametric and non-parametric procedures at risk under such violations, but so also are conventional diagnostics for checking critical features of a model. We return to these subsequently.

To fix ideas, observed responses \(\{Z_0, Z_1, \ldots, Z_n\}\) are often adjusted to \(Z_0\) as standard, giving differences \(\{Y_i = (Z_i - Z_0); \ 1 \leq i \leq n\}\) as the objects of interest to the investigator. Moreover, if \(\{Z_0, Z_1, \ldots, Z_n\}\) are mutually uncorrelated having variances \(\text{var}(Z_0) = \sigma_0^2, \text{var}(Z_i) = \sigma^2; \ 1 \leq i \leq n\) then \(\{Y_1, \ldots, Y_n\}\) are equicorrelated with the parameter \(\rho = \sigma_0^2/(\sigma^2 + \sigma_0^2)\) having variances \(\text{var}(Y_i) = \sigma^2 + \sigma_0^2; \ 1 \leq i \leq n\) that are inflated in comparison with unadjusted values.

Linearly calibrated instruments are pervasive. Some unintended consequences, to be examined here, include (1) the structure of induced dependencies, heteroscedasticity, and other departures

*Corresponding author. Email: der@virginia.edu
from conventional model assumptions; (2) the inflation of measurement variances in comparison with intended values; and (3) effects of calibration on conventional inferences for location, scale, and model diagnostics. We first examine moments, then effects of calibration on the actual measurement distributions themselves. We focus here on direct calibration assays to be identified subsequently. An outline follows.

Section 2 gives notation and technical support. Section 3 reconsiders the calibration process with reference to the irregularities induced through calibration errors. Section 4 addresses the impact of these irregularities on conventional inferences, including (1) inferences regarding the mean and variance in a single sample and (2) the analysis of one–way experimental data, including tests on means and variances. The latter remain largely intact, although measurement variances continue to be underestimated. Section 5 re-examines the ability of conventional diagnostics to uncover the violations induced through calibration. Section 6 undertakes a case study to illustrate essential findings. Section 7 offers a brief summary and cautionary note. Some peripheral matters continue to be underestimated. Section 5 re-examines the ability of conventional diagnostics to uncover the violations induced through calibration. Section 6 undertakes a case study to illustrate essential findings. Section 7 offers a brief summary and cautionary note. Some peripheral matters are deferred to an Appendix.

2. Preliminaries

2.1. Notation

Designate $\mathbb{R}^n$ as an Euclidean $n$-space, $\mathbb{R}_+^n$ as its positive orthant, $\mathbb{S}_n$ as the real symmetric $(n \times n)$ matrices, and $\mathbb{S}_n^+$ and $\mathbb{S}_n^-$ as their positive definite and positive semidefinite varieties.

Arrays appear in bold type, including the transpose $A'$ and inverse $A^{-1}$ of $A$, the unit vector $1_n = [1, \ldots, 1] \in \mathbb{R}^n$, the identity matrix $I_n$, a block-diagonal matrix diag $(A_1, \ldots, A_k)$, and $B_n = (I_n - n^{-1}1_n1_n')$. Following Loewner [1], matrices $(A, B)$ in $\mathbb{S}_n$ are said to be ordered as $A \geq_L B$ for $A - B \in \mathbb{S}_n^+$, with $A >_L B$ whenever $A - B \in \mathbb{S}_n^+$. Moreover, $C(n)$ comprises the convex sets in $\mathbb{R}^n$ symmetric under reflection through $0 \in \mathbb{R}^n$. Operators $E(Y)$ and $V(Y)$ designate the expected vector and dispersion matrix for $Y \in \mathbb{R}^n$; with var$(Y)$ as the variance on $\mathbb{R}^1$. We further require $\{\mu_r(Z) = E(Z^r); r = 1, 2\}$ as moments about 0 $\in \mathbb{R}^1$, identifying $\kappa_2 = \mu_2(\beta_1)$ in terms of a linear estimator $\beta_1$ to be encountered subsequently. The comparative concentration of probability measures on $\mathbb{R}^n$ may be gauged on defining the measure $\mu(\cdot)$ to be more peaked about $0 \in \mathbb{R}^n$ than $\nu(\cdot)$, if and only if $\mu(A) \geq \nu(A)$ for every set $A \in C(n)$, as in [2]. Specifically, the peakedness ordering for scale mixtures of Gaussian measures on $\mathbb{R}^n$ is tantamount to the stochastic ordering of their mixing distributions, as demonstrated in the Appendix.

2.2. Special distributions

Here pdf and cdf refer to probability density and cumulative distribution functions—for $Y \in \mathbb{R}^n$, $\mathcal{L}(Y)$ designates its law of distribution and $G(y)$ its cdf—and iid refers to independent and identically distributed random variables. Distributions of note on $\mathbb{R}^1$ include the Gaussian law $N_1(\mu, \sigma^2)$, with parameters $(\mu, \sigma^2)$; non-central versions of student’s $t(v, \lambda)$, and $t^2(v, \lambda)$, chi-squared $\chi^2(v, \lambda)$, and Snedecor–Fisher $F(v_1, v_2, \lambda)$ distributions, with $(v, v_1, v_2)$ as degrees of freedom and $\lambda$ as a non-centrality parameter; and $G_0(\alpha, \beta)$ as the gamma distribution on $\mathbb{R}_+^1$ having parameters $(\alpha, \beta)$. In particular, $g_T(t; v, \lambda)$, $g_T^2(u; v, \lambda)$, $g_F(u; v_1, v_2, \lambda)$, and $g_0(u; \alpha, \beta)$ designate the densities corresponding to $t(v, \lambda), t^2(v, \lambda), F(v_1, v_2, \lambda)$, and $G_0(\alpha, \beta)$, respectively.

Further, $N_n(\theta, \Sigma)$ designates the Gaussian law on $\mathbb{R}^n$ and $g_n(y; \theta, \Sigma)$ its pdf, having location–scale parameters $(\theta, \Sigma)$. Ensembles on $\mathbb{R}^n$, and mixtures over these, include the translation–scale mixtures

$$f_1(y; \theta, \Sigma, G_1) = \int_{-\infty}^{\infty} g_n(y; \theta(t), \Sigma(t))dG_1(t) \quad (1)$$
and purely scale mixtures when $\theta = 0$. Non-standard distributions for quadratic forms proceed conditionally on letting $\mathcal{L}(U|w)$ have the scaled gamma density $g_0(u; \alpha, w\beta) = (w\beta)^{-\alpha} u^{\alpha-1} e^{-u/w\beta} / \Gamma(\alpha)$, then compounding these as

$$f_3(u; \alpha, \beta, G_3) = \frac{u^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty w^{-\alpha} e^{-x/w\beta} dG_3(w)$$

with $G_3(w)$ as a cdf on $\mathbb{R}_+^1$.

Subsequent developments have links to exchangeable sequences. Consider $\{Z_0, Z_1, Z_2, \ldots \}$ such that $Z_0 \in \mathbb{R}^k$ is independent of $\{Z_1, Z_2, \ldots \}$, whereas $\{Z_1, Z_2, \ldots \}$ are iid on $\mathbb{R}^\infty$: Further, let $\psi : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$, define $Y_i = \psi(Z_0, Z_i); i = 1, 2, \ldots$, recall from DeFinetti’s theorem that the sequence $\{Y_1, Y_2, \ldots \}$ is now exchangeable on $\mathbb{R}^\infty$, and infer that the joint distributions projected onto $\mathbb{R}^n$ are invariant under permutations. In short, $\{Y_1, \ldots, Y_n\}$ are distributed identically under second moments, having common values for the parameters $(\mu, \sigma^2, \rho)$.

3. Calibration

3.1. Essentials

Instruments are calibrated using outputs at successive inputs to determine a calibration curve, new readings are assigned values on the scale of measurements using the calibrated device, and periodic checks against a standard determine when recalibration is required. In this study, we utilize direct assays in which instrumental readings $\{X_i; 1 \leq i \leq n_0\}$ during calibration relate to observed measurements $\{U_i; 1 \leq i \leq n_0\}$ through $U_i = \beta_0 + \beta_1 X_i + \varepsilon_i; 1 \leq i \leq n_0$. For example, the octane rating ($U$) in the production of gasoline relates linearly to the percent of purity ($X$) in a specimen to be assayed. Octane numbers require expensive and time-consuming dynamic laboratory testing, whereas the percent purity is readily determined. Once calibrated, the octane number of a given specimen is determined vicariously from its percent purity. On the other hand, indirect assays proceed on reversing the roles of $U_i$ and $X_i$ during calibration. Models for calibration and their analyses have been debated by several authors; for a summary and early references, see [3]. Problems with moments and consistency remain to be resolved in indirect assays, but the technical issues between the two types differ mainly in detail. It is noteworthy that research has yet to address the principal issues undertaken here, namely, the irregularities in models and supporting analyses attributable to calibration.

3.2. Error analysis

Now, consider the calibrating model $\{U_i = \beta_0 + \beta_1 X_i + \varepsilon_i; 1 \leq i \leq n_0\}$ under Gauss–Markov assumptions, such that $\{\text{var}(U_i) = \sigma_0^2; 1 \leq i \leq n_0\}$ and $(\hat{\beta}_0, \hat{\beta}_1)$ are least-squares estimators determining the empirical calibration line. Under Gaussian calibration, the error terms $\{\varepsilon_i; 1 \leq i \leq n_0\}$ comprise iid $N_1(0, \sigma_0^2)$ random variables. Subsequent readings $\{Z_1, \ldots, Z_n\}$, taken independently of $\{U_1, \ldots, U_{n_0}\}$, are then projected as the calibrated measurements $\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \leq i \leq n\}$. In practice $\{Z_1, \ldots, Z_n\}$ often will have been discarded as redundant, or will have been converted directly without record, so that $\{Y_1, \ldots, Y_n\}$ remain to be analysed and interpreted. If we now suppose that $Z' = [Z_1, \ldots, Z_n]$ have means $\mu' = [\mu_1, \ldots, \mu_n]$ and second moments $V(Z) = \Sigma = [\sigma_{ij}]$, independently of $(\hat{\beta}_0, \hat{\beta}_1)$, then the conditional moments of $\mathcal{L}(Y_1, \ldots, Y_n|\hat{\beta}_1)$ are found directly as follows.
Lemma 1 Suppose that \( \{Z_1, \ldots, Z_n\} \) have means \( \{\mu_1, \ldots, \mu_n\} \) and second moments \( V(Z) = [\sigma_{ij}] \), independently of \((\hat{\beta}_0, \hat{\beta}_1)\), and let \( Y_i = \beta_0 + \beta_1 Z_i; 1 \leq i \leq n \). Then

(i) \( E(Y_i | \hat{\beta}_1) = E(\hat{\beta}_0) + \hat{\beta}_1 \mu_i \),

(ii) \( \text{var}(Y_i | \hat{\beta}_1) = \hat{\beta}_1^2 \sigma_{ii} + \text{var}(\hat{\beta}_0) \hat{\beta}_1 \), and

(iii) \( \text{cov}(Y_i, Y_j | \hat{\beta}_1) = \hat{\beta}_1^2 \sigma_{ij} + \text{var}(\hat{\beta}_0) \hat{\beta}_1 \).

If instead \((\hat{\beta}_0, \hat{\beta}_1)\) were known, then \( Y_i = \beta_0 + \beta_1 Z_i; 1 \leq i \leq n \) would be recovered without an error, in which case \( E(Y_i) = \beta_0 + \beta_1 \mu_i \) and \( \text{var}(Y_i) = \beta_1^2 \text{var}(Z_i) \). This ideal case serves as a reference against which a recovery subject to calibration errors may be gauged.

Expressions simplify when neither \( E(\hat{\beta}_0 | \hat{\beta}_1) \) nor \( \text{var}(\hat{\beta}_0 | \hat{\beta}_1) \) depends on \( \hat{\beta}_1 \), so that \( \text{cov}(Y_i, Y_j | \hat{\beta}_1) = \beta_1^2 \sigma_{ij} + \text{var}(\hat{\beta}_0) \hat{\beta}_1 \), for example. This clearly holds under Gaussian calibration, provided that the calibrating readings \( \{X_1, \ldots, X_{n_0}\} \) have been centred to \( \{(X_1 - \bar{X}), \ldots, (X_{n_0} - \bar{X})\} \). This incurs no loss in generality, as subsequent readings \( \{Z_1, \ldots, Z_n\} \) may be shifted by \( \bar{X} \) units before projecting onto the scale of measurements. It then follows that \( \hat{\beta}_0 = \bar{U} \); \( \text{var}(\hat{\beta}_0) = \sigma_0^2 = \hat{\sigma}_0^2/n_0 \); and \( \text{var}(\hat{\beta}_1) = \sigma_1^2 = \hat{\sigma}_1^2/S_{XX} \), where \( S_{xi} = \sum_{i=1}^{n_0} (X_i - \bar{X})^2 \); so that \( (\hat{\beta}_0, \hat{\beta}_1) \) are uncorrelated and thus independent under Gaussian calibration errors. We henceforth take the initial calibration to have been centred.

We next consider the conditional and unconditional properties of \( \{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \leq i \leq n\} \) for the general case that \( E(Z) = \mu_Z \in \mathbb{R}^n \) and \( V(Z) = \Sigma \in \mathbb{S}_+^n \), to be specialized subsequently. Clearly, the conditional means and dispersion parameters are \( E(Y_i | \hat{\beta}_1) = \beta_0 1_n + \hat{\beta}_1 \mu_Z = \mu_Y(\hat{\beta}_1) \), say, and \( V(Y_i | \hat{\beta}_1) = \hat{\beta}_1^2 \Sigma + \sigma_0^2 1_n 1_n^t = \Xi(\hat{\beta}_1) \). Moreover, for the case that \( \mathcal{L}(Z) = N_n(\mu_Z, \Sigma) \) in addition to Gaussian calibration errors, the conditional distribution of \( Y \) is \( \mathcal{L}(Y | \hat{\beta}_1) = N_n(\mu_Y(\hat{\beta}_1), \Xi(\hat{\beta}_1)) \). Basic unconditional properties follow next.

Theorem 1 Consider the calibrated measurements \( \{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \leq i \leq n\} \) projected from readings \( \{Z_1, \ldots, Z_n\} \) obtained independently of \((\hat{\beta}_0, \hat{\beta}_1)\); let \( Y = \hat{\beta}_0 I_n + \hat{\beta}_1 Z \), such that \( E(Z) = \mu_Z \in \mathbb{R}^n \) and \( V(Z) = \Sigma \in \mathbb{S}_+^n \); and let \( \sigma_0^2 = \text{var}(\hat{\beta}_0) \) and \( \sigma_1^2 = \text{var}(\hat{\beta}_1) \). Then the unconditional moments \( E(Y) = \mu_Y \) and \( V(Y) = \Xi \) of \( \mathcal{L}(Y) \) are given by

(i) \( \mu_Y = \beta_0 1_n + \beta_1 \mu_Z \), and

(ii) \( \Xi = \kappa_2 \Sigma + \sigma_0^2 1_n 1_n^t + \sigma_1^2 \mu_Z \mu_Z^t \), with \( \kappa_2 = \mu_2(\hat{\beta}_1) = \sigma_1^2 + \beta_1^2 \).

(iii) Moreover, if \( \mathcal{L}(Z) = N_n(\mu_Z, \Sigma) \) in addition to Gaussian calibration errors, then the unconditional joint density of the elements of \( Y \) is the translation–scale mixture

\[
 f_i(y; \mu_Y, \Xi, G_1) = \int_{-\infty}^{\infty} g_n(y; \mu(t), \Xi(t)) dG_1(t)
\]

as in Equation (1), with \( \mu(t) = \beta_0 1_n + t \mu_Z \), \( \Xi(t) = t^2 \Sigma + \sigma_0^2 1_n 1_n^t \), and with mixing distribution \( G_1(t) = N_1(\beta_1, \sigma_1^2) \).

Proof Conclusion (i) follows directly through deconditioning. Conclusion (ii) follows using \( \text{var}(Y_i) = E_{\hat{\beta}_1}[\text{var}(Y_i | \hat{\beta}_1)] + \text{var}_{\hat{\beta}_1}(E(Y_i | \hat{\beta}_1)) \) for variances and

\[
 \text{cov}(Y_i, Y_j) = E_{\hat{\beta}_1}[\text{cov}(Y_i, Y_j | \hat{\beta}_1)] + \text{cov}_{\hat{\beta}_1}(E(Y_i | \hat{\beta}_1), E(Y_j | \hat{\beta}_1))
\]

for covariances. Conclusion (iii) follows since \( Y \) is a linear function of \((Z, \hat{\beta}_0)\) with \( \hat{\beta}_1 \) fixed, so that \( \mathcal{L}(Y | \hat{\beta}_1) = N_n(\mu_Y(\hat{\beta}_1), \Xi(\hat{\beta}_1)) \), as noted, and then mixing over the distribution of the conditioning variable.

The above-mentioned results are basic. We next specialize them as appropriate for specific experimental settings encountered routinely in practice.
4. Topics in inference

Induced dependencies and other model irregularities violate the tenets of conventional data analysis as noted, specifically, in estimation and hypothesis testing. We focus on normal-theory procedures, as the independence typically required by non-parametric competitors is conspicuously absent. The following sections specialize earlier findings, as they apply in a single sample, and in one-way comparative experiments.

4.1. Single sample

Consider \( \{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \leq i \leq n\} \) such that the elements of \( Z = [Z_1, \ldots, Z_n]' \) satisfy \( E(Z_i) = \mu_Z; 1 \leq i \leq n \) and \( V(Z) = \sigma_Z^2 I_n \). If in addition \( \{Z_1, \ldots, Z_n\} \) are iid, then \( \mathcal{L}(Y_1, \ldots, Y_n) \) is exchangeable, as noted earlier. We are concerned not only with properties of the joint distribution \( \mathcal{L}(Y) \), but also of \((\bar{Y}, S_Y^2, t_0)\) as the sample mean, the sample variance, and Student’s statistic \( t_0 = n^{1/2}(\bar{Y} - \mu_Y^0)/S_Y \), as well as the ordinary residuals \( \{R_i = (Y_i - \bar{Y}); 1 \leq i \leq n\} \). Let \( \bar{Y} \) be calibrated measurements from \( \{Z_1, \ldots, Z_n\} \), such that \( E(Z) = \mu_Z I_n \) and \( V(Z) = \sigma_Z^2 I_n \) independently of \((\hat{\beta}_0, \hat{\beta}_1)\), and consider the sample quantities \((\bar{Y}_n, S_Y^2)\), together with the ordinary residuals \( \{R_i = (Y_i - \bar{Y}); 1 \leq i \leq n\} \).

**Theorem 2** Let \( \{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \leq i \leq n\} \) be calibrated measurements from \( \{Z_1, \ldots, Z_n\} \), such that \( E(Z) = \mu_Z I_n \) and \( V(Z) = \sigma_Z^2 I_n \) independently of \((\hat{\beta}_0, \hat{\beta}_1)\), and consider the sample quantities \((\bar{Y}_n, S_Y^2)\), together with the ordinary residuals \( \{R_i = (Y_i - \bar{Y}); 1 \leq i \leq n\} \).

- (i) \( \bar{Y}_n \) is unbiased but inconsistent for estimating \( \mu_Y = \beta_0 + \beta_1 \mu_Z \).
- (ii) \( E(S_Y^2) = \kappa_2 \sigma_Z^2 = \sigma_Z^2 (\sigma_0^2 + \beta_1^2) \), and
- (iii) \( E(R_i) = 0; 1 \leq i \leq n \).

**Proof** The unbiasedness of \( \bar{Y}_n \) follows routinely, and its variance from \( \text{var}(n^{-1} 1_n Y) = \Xi_n \) with

\[
\Xi_n = n^{-2} 1_n' [\kappa_2 \sigma_Z^2 I_n + (\sigma_0^2 + \sigma_1^2 \mu_Z^2) 1_n 1_n'] 1_n \\
= n^{-1} \kappa_2 \sigma_Z^2 + (\sigma_0^2 + \sigma_1^2 \mu_Z^2).
\]

Since \( \lim_{n \to \infty} \text{var}(\bar{Y}_n) = (\sigma_0^2 + \sigma_1^2 \mu_Z^2) > 0 \), it follows that its limit distribution is non-degenerate at \( \mu_Y \), so that \( \bar{Y}_n \) is consistent neither in probability, nor in mean square, nor almost surely, as asserted. Conclusion (ii) follows from evaluating the expected value of the quadratic form \((n - 1) S_Y^2 = Y' B_n Y\) as \(E[((n - 1) S_Y^2) = \text{tr}(B_n V(Y)) + \mu_Y' B_n \mu_Y \). Details are

\[
E[\{(n - 1) S_Y^2 \} = \text{tr} \left( B_n [k_2 \sigma_Z^2 I_n + (\sigma_0^2 + \sigma_1^2 \mu_Z^2) 1_n 1_n'] B_n \right) + \mu_Y' B_n \mu_Y \\
= (n - 1) k_2 \sigma_Z^2
\]

where \( \mu_Y' B_n \mu_Y = (\beta_0 + \beta_1 \mu_Z)^2 1_n' B_n 1_n = 0 \), since \( B_n \) is idempotent and \( B_n 1_n = 0 \). Conclusion (iii) follows directly, to complete our proof. \( \blacksquare \)
The following consequences may be noted.

- Conclusion (i) appears to dash the usual expectation that lengths of \((1 - \alpha)\) confidence intervals for \(\mu_Y\) will decrease at the rate \(O(n^{-1/2})\).
- The sample variance \(S_Y^2\) underestimates the actual variance \(\sigma_Y^2\). The bias is \(E(S_Y^2) - \sigma_Y^2 = -(\sigma_0^2 + \sigma_1^2 \mu_Z^2)\).
- This bias increases with decreasing precision in estimating the calibration line, and with increasing \(|\mu_Z|\) and thus \(|\mu_Y|\).
- On the other hand, the expectation \(E(S_Y^2) = \kappa_2 \sigma_Z^2\), with \(\kappa_2 = \mu_2(\hat{\beta}_1)\), may be compared with the ideal variance, \(\text{var}(Y_i) = \beta_1^2 \sigma_Z^2\), attained under linear calibration with known parameters \((\beta_0, \beta_1)\).

We have seen how unconditional moments of calibrated measurements depend on those of the conditioning variable \(\hat{\beta}_1\). It remains to examine effects of the fitted calibration line on unconditional distributions, including those of various sample statistics. Recall from Theorem 2 and its proof that \(E(\hat{Y}) = \mu_Y = \beta_0 + \beta_1 \mu_Z\) and \(\text{var}(\hat{Y}) = n^{-1} k_2 \sigma_Z^2 + (\sigma_0^2 + \sigma_1^2 \mu_Z^2)\). To invoke expression (1) and its special case at \(\theta = 0\), under Gaussian assumptions, we have \(G_1(\hat{\beta}_1) = N_1(\beta_1, \sigma_1^2)\), together with \(G_2(\hat{\beta}_1^2; \lambda)\), such that \(L(\hat{\beta}_1^2 / \sigma_1^2) = \chi^2(1, \lambda)\) with \(\lambda = \beta_1^2 / \sigma_1^2\). Basic unconditional distributions follow next as mixtures.

**Theorem 3** Let \(\{Y_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i; 1 \leq i \leq n\}\) be calibrated measurements; suppose that \(\{Z_1, \ldots, Z_n\}\) are iid \(N_1(\mu_2, \sigma_2^2)\) independently of \((\hat{\beta}_0, \hat{\beta}_1)\) under Gaussian calibration; and consider the sample quantities \((\hat{Y}, S_Y^2, t_0^2)\), together with the ordinary residuals \(\{R_i = (Y_i - \hat{Y}); 1 \leq i \leq n\}\), where \(t_0^2 = n(\hat{Y} - \mu_Y^0)^2 / S_Y^2\) for testing \(H_0 : \mu_Y = \mu_Y^0\) against \(H_1 : \mu_Y \neq \mu_Y^0\). Then the unconditional properties are as follows.

(i) The unconditional density of \(L(\hat{Y})\) is the translation–scale mixture

\[
f_1(u; \mu_Y, \Sigma_n, G_1) = \int_0^\infty g_1(u; \mu(t), \Sigma_n(t))dG_1(t)
\]

with mixing distribution \(G_1(\hat{\beta}_1) = N_1(\beta_1, \sigma_1^2)\) as in Equation (1) for the case \(n = 1\), where \(\mu(t) = \beta_0 + t \mu_Z\) and \(\Sigma_n(t) = (t^2 \sigma_Z^2 / n + \sigma_0^2)\), and \(\Sigma_n\) is defined in Equation (4).

(ii) The joint density of residuals \(R = [R_1, \ldots, R_n]^T\) is given by

\[
f_n(r; 0, \sigma_Z^2 B_n, G_2) = \int_0^\infty g_n(r; 0, t \sigma_Z^2 B_n) dG_2(t)
\]

with mixing distribution \(G_2(\hat{\beta}_1^2; \lambda) = \chi^2(1, \lambda)\) and \(\lambda = \beta_1^2 / \sigma_1^2\).

(iii) The joint distribution \(L(R_1, \ldots, R_n)\) increases in peakedness about \(0 \in \mathbb{R}^n\) with decreasing \(\lambda\).

(iv) \(L(vS_Y^2 / \sigma_Z^2)\) has the density \(f_0(u; v/2, 2, G_2)\) as in Equation (2), with \(v = n - 1\), mixing distribution \(G_2(\hat{\beta}_1^2; \lambda) = \chi^2(1, \lambda)\) and \(\lambda = \beta_1^2 / \sigma_1^2\), and \(E(S_Y^2) = \sigma_Z^2(\sigma_1^2 + \beta_1^2)\).

(v) The distribution \(L(vS_Y^2 / \sigma_Z^2)\) increases stochastically with \(\lambda = \beta_1^2 / \sigma_1^2\).

(vi) The unconditional density of \(t_0^2 = n(\hat{Y} - \mu_Y^0)^2 / S_Y^2\) is given as the mixture

\[
g(u; v, \delta, G_2) = \int_0^\infty g_T\left(u; v, \frac{\delta}{t}\right) dG_2(t)
\]

with mixing distribution \(G_2(\hat{\beta}_1^2; \lambda)\), where \(v = n - 1, \delta = (\mu_Y - \mu_Y^0)^2 / \sigma_Z^2\), and \(\lambda = \beta_1^2 / \sigma_1^2\).

(vii) The unconditional cdf of \(L(t_0^2)\) increases stochastically with increasing \(\delta = (\mu_Y - \mu_Y^0)^2 / \sigma_Z^2\) for fixed \(\lambda = \beta_1^2 / \sigma_1^2\), and for fixed \(\delta\), it decreases stochastically with increasing \(\lambda\).
Proof. The conditional distribution of note is $\mathcal{L}(\tilde{Y}|\tilde{\beta}_1) = N_1(\mu(\tilde{\beta}_1), \Sigma_n(\tilde{\beta}_1))$, where $\mu(\tilde{\beta}_1) = \beta_0 + \hat{\beta}_1 \mu_Z$ and $\Sigma_n(\tilde{\beta}_1) = \left(\hat{\beta}_1^2 \sigma_Z^2 / n + \sigma_Y^2\right)$. Its unconditional density thus is $f_1(u; \mu_Y, \Sigma_n, G_1)$ from Equation (1), to give conclusion (i) with mixing distribution as asserted. Now, observe that $R = B_n Y$ and $B_n \Sigma(\tilde{\beta}_1) B_n = B_n (\hat{\beta}_1^2 \sigma_Z^2 I_n + \sigma_Y^2 I_n) B_n = \hat{\beta}_1^2 \sigma_Z^2 B_n$ since $B_n$ is idempotent and $B_n I_n = 0$. We infer conditionally that $\mathcal{L}(R|\tilde{\beta}_1) = N_n(0, \hat{\beta}_1^2 \sigma_Z^2 B_n)$, since $B_n \mu(\tilde{\beta}_1) = (\beta_0 + \hat{\beta}_1 \mu_Z) B_n I_n = 0$. The unconditional distribution is the scale mixture as in conclusion (ii), with $G_2(\beta_1^2, \lambda) = \chi^2(1, \lambda)$ as the mixing distribution over $[0, \infty]$. Conclusion (iii) follows from Lemma A1 of the Appendix, as the mixing distribution $\mathcal{L}(\hat{\beta}_1^2 / \sigma_Y^2) = \chi^2(1, \lambda)$ increases stochastically with $\lambda = \beta_1^2 / \sigma_Y^2$. Now, observe that $(n-1)S_Y^2 = R' R$, so that $\mathcal{L}(R' R / \hat{\beta}_1^2 | \beta_1^2) = \chi^2(n-1, 0)$. It follows that $\mathcal{L}(n-1)S_Y^2 / \sigma_Y^2$ is a central chiquared variate scaled by $\beta_1^2$. On identifying $(n-1)S_Y^2 / \sigma_Y^2$ with $U$ and $\beta_1^2$ with $w$ in developments leading to Equation (2), we thus establish conclusion (iv) on specializing from gamma to chi-squared distributions. Conclusion (v) follows directly from (iii) since the set $A_1 = \{ R \in \mathbb{R}^n : R' R \leq t \}$ is convex and symmetric in $\mathbb{R}^n$, whereas $P(A_1) = P((n-1)S_Y^2 \leq t \sigma_Y^2)$.

To see conclusion (vi), recall the affine invariance of $t^2 = n(\bar{Y} - \mu_Y)^2 / S_Y^2$ for testing $H_0 : \mu_Y = \mu_Y^0$ under $\{Y_i \to a + b Z_i ; 1 \leq i \leq n\}$, namely, $t^2 = n[(\bar{Y} - \mu_Y) + (\mu_Y - \mu_Y^0)]^2 / S_Y^2 = n[(a + b\bar{Z}) - (a + b\mu_Y)] + b(\mu_Y - \mu_Y^0)^2 / b^2 S_Y^2 = n[(\bar{Z} - \mu_Z) + (\mu_Y - \mu_Y^0)^2 / b^2 S_Y^2]$. Clearly $\mathcal{L}(t^2) = \mathcal{L}(v, \delta)$, with $v = n - 1$ and $\delta = (\mu_Y - \mu_Y^0)^2 / b^2 S_Y^2$, independently of $a$. We next apply these facts conditionally, given $(\hat{\beta}_0, \hat{\beta}_1)$, on replacing $(a, b)$ with $(\hat{\beta}_0, \hat{\beta}_1)$, to infer that $\mathcal{L}(t^2_0 | \hat{\beta}_0, \hat{\beta}_1) = \mathcal{L}(t^2(v, \delta(\hat{\beta}_1)))$, where $\delta(\hat{\beta}_1) = (\mu_Y - \mu_Y^0)^2 / \hat{\beta}_1^2 \sigma_Y^2$, independently of $\hat{\beta}_0$. It follows that the unconditional distribution of $t^2_0$ has the mixture density $g(u; v, \delta, G_2) = \int_0^\infty g_T(v; u, \delta) dG_2(u)$ with mixing distribution $G_2(\hat{\beta}_1^2, \lambda)$ as asserted, where $\delta = (\mu_Y - \mu_Y^0)^2 / \hat{\beta}_1^2 \sigma_Y^2$ and $\lambda = \hat{\beta}_1^2 / \sigma_Y^2$. Therefore, the test for $H_0 : \mu_Y = \mu_Y^0$ against $H_1 : \mu_Y \neq \mu_Y^0$ rejects at level $\alpha$ for $t^2_0 > c_2^2$, moreover, the conditional cdf $\mathcal{L}(t^2_0 | \hat{\beta}_1)$ increases stochastically with $\delta(\hat{\beta}_1)$, pointwise for each fixed $\hat{\beta}_1$ from standard properties of non-central $t^2$ distributions. It follows that the unconditional cdf increases stochastically with increasing $\delta$ under mixing. That the unconditional cdf decreases stochastically with increasing $\lambda$ with $\delta$ held fixed, follows unconditionally through mixing as in the proof for conclusion (v), to complete our proof.

Note from conclusion (iv) that $E(S_Y^2) = \sigma_Y^2(\beta_1^2 + \beta_1^2)$. This may be compared with the ideal case $\text{var}(Y_i) = \beta_1^2 \sigma_Z^2$, where $(\hat{\beta}_0, \hat{\beta}_1)$ are known. It is instructive to re-examine the unconditional properties of this section if we retain the homogeneity of variances of $\{Z_1, \ldots, Z_n\}$, but assume instead that means are not, i.e., that $E(Z) = \mu = [\mu_1, \ldots, \mu_n]$. Therefore, we see that $\{\text{var}(Y_i) = \kappa \sigma_Z^2 + \sigma_Y^2 + \beta_1^2 ; 1 \leq i \leq n\}$. We have the curious finding that calibrated measurements in a single sample will have heterogeneous variances under heterogeneous means, despite the homogeneity of variances of $\{Z_1, \ldots, Z_n\}$.

4.2. One-way experiments

Clearly $\{Y_1, \ldots, Y_n\}$ and $\{Z_1, \ldots, Z_n\}$ have the same experimental structure, hence a one-way experiment comprising $k$ samples of sizes $\{n_1, \ldots, n_k\}$, with $n_1 + \cdots + n_k = n$ in keeping with conventional notation, partition $Z' = [Z_1, \ldots, Z_n]$ as $[Z_1', \ldots, Z_k']$ such that $|Z_i| = [Z_i, \ldots, Z_{i+k}]$; $1 \leq i \leq k$; similarly for $Y' = [Y_1, \ldots, Y_k']$, with $|Y_i| = [Y_{i+1}, \ldots, Y_{i+k}]$; $1 \leq i \leq k$; and suppose that $\{E(Z_{ij}) = \mu_i ; 1 \leq j \leq n_i\}$ and $\{\text{var}(Z_{ij}) = \omega^2_i ; 1 \leq j \leq n_i\}$. Accordingly, take $\mu_Z = [\mu_1 I_{n_1}, \ldots, \mu_k I_{n_k}]$ and $V(Z) = \text{diag}(\omega^2_1 I_{n_1}, \ldots, \omega^2_k I_{n_k}) = D(\omega^2_1, \ldots, \omega^2_k)$, say. Specializing from Section 3, we have the conditional moments $E(Y'(\hat{\beta}_1)) = \mu_Y(\hat{\beta}_1) = \beta_0 I_n + \hat{\beta}_1 [\mu_1 I_{n_1}, \ldots, \mu_k I_{n_k}]$ and $V(Y' \hat{\beta}_1) = \Sigma(\hat{\beta}_1) = \hat{\beta}_1^2 \text{diag}(\omega^2_1 I_{n_1}, \ldots, \omega^2_k I_{n_k}) + \sigma_Y^2 I_n$, together with $\mathcal{L}(Y' \hat{\beta}_1) = N_n(\mu_Y(\hat{\beta}_1), \Sigma(\hat{\beta}_1))$ under Gaussian errors. Moreover, unconditional moments
are \( E(Y) = \mu_Y = \beta_0 1_n + \beta_1 [\mu_1 1_{n'}, \ldots, \mu_k 1_{n'}] \) and \( V(Y) = \Xi = k_2 \text{diag}(\omega_1^2 I_{n1}, \ldots, \omega_k^2 I_{nk}) + \sigma_0^2 I_n 1_n' + \sigma_i^2 M \), where \( M = [M_{ij}] = \mu_Z \mu_Z' \) has the block structure \( M_{ij} = \mu_i \mu_j 1_n 1_n' \). In particular, for typical calibrated measurements in sample \( i \) of the \( k \) samples, the conditional and unconditional means are \( E(Y_{ij} | \hat{\beta}_1) = \beta_0 + \beta_1 \mu_i; 1 \leq j \leq n_i \) and \( E(Y_{ij}) = \beta_0 + \beta_1 \mu_j; 1 \leq j \leq n_j \), whereas the corresponding variances are \( \text{var}(Y_{ij} | \hat{\beta}_1) = \beta_1^2 \omega_i^2 + \sigma_i^2; 1 \leq j \leq n_i \) and \( \text{var}(Y_{ij}) = \kappa_2 \omega_i^2 + \sigma_0^2 + \sigma_i^2 \mu_j^2; 1 \leq j \leq n_j \). We are concerned with the dual issues of homogeneity of means, and of the homogeneity of variances, across the \( k \) samples. Clearly the induced irregularities are artifacts of the calibration process, rather than consequences of the experimental structure itself. It remains to determine unintended effects of calibration on conventional comparisons among the means and the variances.

In contrast to conventional one-way experiments, where homoscedasticity can be checked regardless of heterogeneity among the \( k \) population means, under calibration we see that homoscedasticity of the unconditional variances is possible only in unusual circumstances. Specifically, homoscedasticity holds unconditionally if and only if, for every pair \( (\omega_i^2, \mu_j^2) \), that \((\omega_i^2 - \omega_j^2) = c(\mu_j^2 - \mu_j^2) \) with \( c = \sigma_i^2/\kappa_2 \).

Now, consider the transformations \( T_1(Y), T_2(Y), \) and \( T_3(Y) \) such that \( T_1(Y) = \bar{Y} = [\bar{Y}_1, \ldots, \bar{Y}_k]' \) comprise the \( k \) sample means; \( T_2(Y) = R' = [R_1', \ldots, R_k']' \) consists of the ordinary within-sample residuals, with \( R_i = B_{n_i} Y_i \) and \( B_{n_i} = (I_{n_i} - n_i^{-1} 1_{n_i} 1_{n_i}') \); and \( T_3(Y) = [\nu_1 S_1^2, \ldots, \nu_k S_k^2]' \) are the residual sums of squares, i.e., \( \nu_i S_i^2 = R_i R_i' = Y_i' B_{n_i} Y_i; 1 \leq i \leq k \), with \( \nu_i = n_i - 1 \). Basic properties may be summarized as follows.

**Theorem 4** Consider the calibrated measurements \( Y' = [Y_1', \ldots, Y_k'] \) corresponding to \( Z' = [Z_1, \ldots, Z_k] \) such that \( E(Z) = [\mu_1 I_{n1}', \ldots, \mu_k I_{nk}'] \) and \( V(Z) = \text{diag}(\omega_1^2 I_{n1}, \ldots, \omega_k^2 I_{nk}) \), and let \( T_1(Y) = \bar{Y} = [\bar{Y}_1, \ldots, \bar{Y}_k]' \), \( T_2(Y) = [R_1', \ldots, R_k'] \), and \( T_3(Y) = [\nu_1 S_1^2, \ldots, \nu_k S_k^2]' \), with \( \nu_i = n_i - 1; 1 \leq i \leq k \). Moreover, a Gaussian model asserts that \( (Z_i - \mu_i)/\omega_i; 1 \leq j \leq n_i, 1 \leq i \leq k \) are iid \( N(0, 1) \) random variates independently of \((\hat{\beta}_0, \hat{\beta}_1)\) under Gaussian calibration.

(i) The conditional and unconditional moments of \( T_1(Y) = \bar{Y} \) are given by \( E(\bar{Y} | \hat{\beta}_1) = \theta(\hat{\beta}_1) = \beta_0 1_k + \beta_1 \mu \), \( E(\bar{Y}) = \Xi = \Xi_1(\hat{\beta}_1) = \beta_1^2 \text{diag}(\omega_1^2/n_1, \ldots, \omega_k^2/n_k) + \sigma_0^2 1_k 1_k' \), and \( V(\bar{Y}) = \Xi = \Xi_1(\hat{\beta}_1) = \beta_1^2 \text{diag}(\omega_1^2/n_1, \ldots, \omega_k^2/n_k) + \sigma_0^2 1_k 1_k' + \sigma_i^2 \mu_\mu \), where \( \mu_\mu = [\mu_1, \ldots, \mu_k] \).

(ii) Under Gaussian models, the unconditional density of \( \mathcal{L}(Y) \) is the translation–scale mixture

\[
    f_k(u; \theta, \Xi_1, G_1) = \int_{-\infty}^{\infty} g_k(u; \theta(t), \Xi_1(t)) dG_1(t) 
\]

with mixing distribution \( G_1(\hat{\beta}_1) = N_1(\beta_1, \sigma_1^2) \) as in Equation (1), where \( \theta(t) = \beta_0 1_k + t \mu \) and \( \Xi_1(t) = t^2 \text{diag}(\omega_1^2/n_1, \ldots, \omega_k^2/n_k) + \sigma_0^2 1_k 1_k' \).

(iii) The conditional and unconditional moments of the residuals are \( E(R | \hat{\beta}_1) = E(R) = 0 \), \( V(R | \hat{\beta}_1) = \Xi_2(\hat{\beta}_1) = \beta_1^2 \text{diag}(\omega_1^2 B_{n1}, \ldots, \omega_k^2 B_{nk}) \), and

\[
    V(R) = \Xi_2 = \kappa_2 \text{diag}(\omega_1^2 B_{n1}, \ldots, \omega_k^2 B_{nk}).
\]

(iv) Under Gaussian models, the joint density of residuals \( R = [R_1', \ldots, R_k'] \) is given by \( f_2(r; \theta, \Xi_2, G_2) \) as in Equation (6), with mixing distribution \( G_2(\hat{\beta}_1; \lambda) \) and \( \lambda = \beta_1^2/\sigma_1^2 \).

(v) Under Gaussian models, the joint density of elements of \( [\nu_1 S_1^2/\omega_1^2, \ldots, \nu_k S_k^2/\omega_k^2] \) is given by

\[
    f(u; \nu_1, \ldots, \nu_k) = \int_{0}^{\infty} \prod_{i=1}^{k} g_0(\frac{u_i}{\nu_i^2}, \frac{2w}{\nu_i}) dG_2(w)
\]

with \( \nu_i = n_i - 1 \) and \( g_0(u; \alpha, w) = (w \beta)^{-u} u^{u-1} e^{-u/w \beta} / \Gamma(\alpha) \), having the mixing distribution \( G_2(\hat{\beta}_1; \lambda) \) with \( \lambda = \beta_1^2/\sigma_1^2 \).
Proof Arguments follow step by step as in the proofs given in Section 4.1. The details differ, but proceed similarly on noting that \( \bar{Y} = \text{diag}(n_{1}^{-1}1_{n_{1}}, \ldots, n_{k}^{-1}1_{n_{k}})Y = L_{2}Y \), say, whereas \( R = \text{diag}(B_{n_{1}}, \ldots, B_{n_{k}})Y = L_{2}Y \). The conditional and unconditional moments follow directly as linear functions, together with the idempotencies of \( \{B_{n_{i}}; 1 \leq i \leq k\} \) and the annihilations achieved through \( \{B_{n_{i}}1_{n_{i}}; 1 \leq i \leq k\} \). A Gaussian model for \( \mathcal{L}(Z) \) and Gaussian errors during calibration give conditional Gaussian laws for \( \mathcal{L}(\bar{Y} | \hat{\beta}_{1}) \) and \( \mathcal{L}(R | \hat{\beta}_{1}) \), whereas the unconditional distributions are mixtures as in Section 2.2. Moreover, \( \{S_{1}^{2}, \ldots, S_{k}^{2}\} \), are conditionally independent given \( \hat{\beta}_{1} \). As in the proof for Theorem 3(iv), the marginal density of \( \mathcal{L}(v_{i}S_{i}^{2}/\omega_{i}^{2} | \hat{\beta}_{1}) \) is the scaled chi-squared density \( g_{0}(u_{i} ; v_{i}/2, 2\omega) \) as defined preceding Equation (2), with \( w = \hat{\beta}_{1} \). Their unconditional density now follows on mixing as in Section 2.2, as asserted in conclusion (v), to complete our proof.

It is essential to examine effects of calibration in comparing variances across the \( k \) groups, typically based on \( \{S_{1}^{2}, \ldots, S_{k}^{2}\} \). In the ideal case where \( \{\beta_{0}, \beta_{1}\} \) are known, we would have \( \{\text{var}(Y_{ij}) = \beta_{1}^{2}\omega_{i}^{2} ; 1 \leq i \leq k\} \), so that homoscedasticity across groups for measurements \( \{Y_{ij}\} \) would be tantamount to that for \( \{Z_{ij}\} \). Under calibration errors, however, Theorem 2(ii) shows that \( S_{i}^{2} \) underestimates \( \text{var}(Y_{ij}) = \kappa_{2}\omega_{i}^{2} + \alpha_{0}^{2} + \sigma_{i}^{2}\mu_{i}^{2} \), the amount of bias, \( B_{i} = -\left(\sigma_{i}^{2} + \sigma_{i}^{2}\omega_{i}^{2}\right) \), being an artifact of the calibration process itself. Accordingly, it is germane to examine homogeneity among the expected values \( \{\kappa_{2}\omega_{1}^{2}, \ldots, \kappa_{2}\omega_{k}^{2}\} \) of \( \{S_{1}^{2}, \ldots, S_{k}^{2}\} \). To these ends, let \( T_{4}(S_{1}^{2}, \ldots, S_{k}^{2}) \) be any scale-invariant statistic based on the sample variances from the measurements \( \{Y_{ij} ; 1 \leq j \leq n_{i}, 1 \leq i \leq k\} \). The following result is basic.

**Theorem 5** Let \( \{S_{1}^{2}, \ldots, S_{k}^{2}\} \) be within-sample variances from the calibrated measurements \( \{Y_{ij} ; 1 \leq j \leq n_{i}, 1 \leq i \leq k\} \) in a one-way experiment; let \( T_{4}(S_{1}^{2}, \ldots, S_{k}^{2}) \) be any scale-invariant statistic; and consider a Gaussian model where \( \{Z_{ij} - \mu_{i}\}/\omega_{i} ; 1 \leq j \leq n_{i}, 1 \leq i \leq k\} \) are iid \( N_{1}(0, 1) \) random variables independently of \( \{\hat{\beta}_{0}, \hat{\beta}_{1}\} \) under Gaussian calibration errors. Then the distribution of \( T_{4}(S_{1}^{2}, \ldots, S_{k}^{2}) \) is identical to its normal-theory form, independently of \( \{\hat{\beta}_{0}, \hat{\beta}_{1}\} \) and the empirical calibration line.

Proof The proof for Theorem 4(v) asserts \( f(u; v_{1}, \ldots, v_{k}) = \prod_{i=1}^{k} g_{0}(u_{i} ; v_{i}/2, 2) \) as the conditional density for \( \mathcal{L}(v_{1}S_{1}^{2}/\hat{\beta}_{1}, \ldots, v_{k}S_{k}^{2}/\hat{\beta}_{1}, \sigma_{i}^{2}) | \hat{\beta}_{1} \), with \( v_{i} = n_{i} - 1 \) and \( g_{0}(u; \alpha, \beta) = u^{{\alpha} - 1}e^{-u/\beta}/\beta^{\alpha} \Gamma(\alpha) \), so that \( \{S_{1}^{2}, \ldots, S_{k}^{2}\} \) are conditionally independent given \( \hat{\beta}_{1} \). But since \( T_{4}(S_{1}^{2}, \ldots, S_{k}^{2}) \) is scale-invariant, \( \mathcal{L}(T_{4}(S_{1}^{2}, \ldots, S_{k}^{2}) | \hat{\beta}_{1}) = \mathcal{L}(T_{4}(S_{1}^{2}, \ldots, S_{k}^{2}) \text{unconditionally}, to complete our proof.

It deserves note that meaningful comparisons among variances are necessarily scale-invariant. Moreover, it is seen that procedures based on \( \{S_{1}^{2}/\omega_{1}^{2}, \ldots, S_{k}^{2}/\omega_{k}^{2}\} \) support tests for conditional hypotheses that \( \{\text{var}(Y_{ij} | \hat{\beta}_{1}) = \beta_{1}^{2}\omega_{i}^{2} + \sigma_{0}^{2} ; 1 \leq i \leq k\} \), or equivalently, \( H_{0} : \omega_{1}^{2} = \omega_{2}^{2} = \cdots = \omega_{k}^{2} \), to be tested against alternatives as appropriate. Theorem 5 applies in the case of both null and non-null distributions of invariant test statistics. Tests in common usage include

- modifications of Bartlett’s [4] likelihood ratio test,
- Cochran’s [5] test based on \( S_{\max}^{2}/\{S_{1}^{2} + \cdots + S_{k}^{2}\} \),
- Hartley’s [6] F-max test based on the maximal ratio \( \max\{S_{j}^{2}/S_{j}^{2}\} \), and

To examine effects of calibration errors on the one-way analysis of variance for comparing means, we proceed conditionally given \( \hat{\beta}_{1} \), first assuming that \( \{\text{var}(Y_{ij} | \hat{\beta}_{1}) = \beta_{1}^{2}\omega_{i}^{2} + \sigma_{0}^{2} ; 1 \leq j \leq n_{i}, 1 \leq i \leq k\} \), so that \( V(Y | \hat{\beta}_{1}) = \beta_{1}^{2}\omega_{i}^{2}1_{n_{i}} + \sigma_{0}^{2}1_{n_{i}}1_{n_{i}}' = \Xi(\beta_{1}) \) in the notation of Section 3.2. We are concerned with comparative inferences regarding elements of \( \mu(\beta_{1}) = \beta_{1}[\mu_{1}, \ldots, \mu_{k}]' \)
from \( E(Y) = \beta_0 1_n + \beta_1 [\mu_1 1_n', \ldots, \mu_k 1_n'] \). Recall that \( I_n = A_0 + A_1 + A_2 \) partitions \( Y' I_n Y = Y'A_0 Y + Y'A_1 Y + Y'A_2 Y \) such that \( Y'A_0 Y = n \bar{Y}^2 \), with \( \bar{Y} \) as the grand mean and \( A_0 = n^{-1} 1_n 1_n' \), \( Y'A_1 Y = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2 \); and \( Y'A_2 Y = \sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2 \). To validate the Fisher–Cochran theorem conditionally requires that \( \{ A_i \hat{\Xi}(\hat{\beta}) A_j = 0; i \neq j \} \). Moreover, scale parameters associated with the quadratic forms are found as \( \{ \kappa_i^2 A_i = A_i \hat{\Xi}(\hat{\beta}_i) A_i; i = 1, 2, 3 \} \), whereas non-centrality parameters derive from expected mean squares. This program of study is carried out next in support of the following.

**Theorem 6** Let \( \{ Y_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Z_{ij}; 1 \leq j \leq n_i, 1 \leq i \leq k \} \) be calibrated measurements in a one-way experiment such that \( \{(Z_{ij} - \mu_i)/\omega; 1 \leq j \leq n_i, 1 \leq i \leq k \} \) are iid \( N(0, 1) \) random variables independently of \( (\hat{\beta}_0, \hat{\beta}_1) \) under Gaussian calibration errors.

(i) The analysis of variance test for equality of elements of \( \mu(\beta_1) = \beta_1[\mu_1, \ldots, \mu_k]' \), pertaining to the group measurement means, is identical in level and power to its normal-theory form.

(ii) Support tests, based on linear contrasts among the group means, are identical in level and power to their normal-theory forms.

**Proof** To validate the Fisher–Cochran theorem conditionally, observe \( \{ A_i \hat{\Xi}(\hat{\beta}_i) A_0 = 0; i = 1, 2 \} \), since \( A_i \hat{\Xi}(\hat{\beta}_i) A_0 = A_i (\hat{\beta}_i^2 \omega^2 1_n + \sigma_0^2 1_n 1_n') A_0 \) and \( \{ A_i A_0 = 0; i = 1, 2 \} \) Similarly \( \{ A_i \hat{\Xi}(\hat{\beta}_1) A_j = 0; (i, j) = 1, 2, i \neq j \} \) since \( A_i \hat{\Xi}(\hat{\beta}_1) A_j = A_i (\hat{\beta}_1^2 \omega^2 1_n + \sigma_0^2 1_n 1_n') A_j \) and \( \{ A_i A_j = 0; (i, j) = 0, 1, 2, i \neq j \} \) from standard properties of the one-way classification. Scale parameters, as determined from \( \{ \kappa_i^2 A_i = A_i \hat{\Xi}(\hat{\beta}_i) A_i; i = 1, 2 \} \) are found to be equal—namely \( \{ \kappa_i^2 A_i = A_i (\hat{\beta}_i^2 \omega^2 1_n + \sigma_0^2 1_n 1_n') A_i = \hat{\beta}_i^2 \omega^2 A_i; i = 1, 2 \} \) from idempotency together with the annihilation \( \{ A_i 1_n = 0; i = 1, 2 \} \), so that \( \{ \kappa_i^2 = \kappa^2; i = 1, 2 \} \). Finally, the noncentrality parameters and degrees of freedom associated with \( \{ Y'A_i Y; i = 1, 2 \} \) are determined from their expected mean squares. These are \( \{ E(Y'A_i Y|\hat{\beta}_1) = \text{tr}(A_i \hat{\Xi}(\hat{\beta}_1) + [\mu(\hat{\beta}_1)] A_i^2 \mu(\hat{\beta}_1); i = 1, 2 \} \). It follows directly that

\[
E(Y'A_1 Y|\hat{\beta}_1) = \text{tr}(A_1 \hat{\Xi}(\hat{\beta}_1) + [\mu(\hat{\beta}_1)] A_1^2 \mu(\hat{\beta}_1) = (k - 1) \kappa^2 + \hat{\beta}_1^2 \sum_{i=1}^k n_i (\mu_i - \mu)^2
\]

with \( \bar{\mu} = \sum_{i=1}^k n_i \mu_i/n \). Similarly \( E(Y'A_2 Y|\hat{\beta}_1) = \text{tr}(A_2 \hat{\Xi}(\hat{\beta}_1) + [\mu(\hat{\beta}_1)] A_2^2 \mu(\hat{\beta}_1) = (n - k) \kappa^2 \)

since \( [\mu(\hat{\beta}_1)] A_2^2 \mu(\hat{\beta}_1) = \hat{\beta}_1^2 \sum_{i=1}^k n_i (\mu_i - \mu)^2 / \hat{\beta}_1^2 \omega^2 = \sum_{i=1}^k n_i (\mu_i - \mu)^2 / \omega^2 \). Thus the conditional and unconditional distributions are identical, i.e., \( \mathcal{L}(F|\hat{\beta}_1) = \mathcal{L}(F) = (k - 1, n - k, \lambda) \), with \( \lambda = \sum_{i=1}^k n_i (\mu_i - \mu)^2 / \omega^2 \).

**5. Diagnostics**

**5.1. Objectives**

Calibration errors exact profound disturbances, both in models and in data-analytic procedures, as shown. Myriad calibrated data sets have been analysed to date, supported of late by an evolving battery of diagnostic tools. On these grounds, it is tempting to dismiss the present study as academic: in fact, these issues long since would have surfaced in practice, to be addressed accordingly. At issue is the capacity of known diagnostics to uncover calibration-induced irregularities as documented here. We now address these concerns with regard to induced correlations,
non-normality, mixture distributions having excessive tails, and possible outliers. For definiteness, we return to the case of a single sample as in Section 4.1.

5.2. Correlation

Neither the conditional \( (\sigma_0^2(\beta_1^2\sigma_2^2 + \sigma_0^2)) \) nor the unconditional \( (\sigma_0^2 + \sigma_1^2\mu_2^2)/(\kappa_2\sigma_2^2 + \sigma_2^2 + \sigma_1^2\mu_2^2) \) correlations need to be negligible. Tests for correlation entail dispersion matrices \( V(Y) = \tau^2 \Xi(\omega) \), for which \( \Xi(\omega) = (I_n + \omega A) \) with \( A \) fixed and \( \Xi(\omega) \in S_+^n \). Specializing gives \( \tau^2 \Xi(\omega) \) as \( \Sigma(\rho) \) under the equicorrelation models encountered here. Tests of note are due to Durbin and Watson [8–10], Anderson and Anderson [11], Theil [12], and others, all based on versions of von Neumann’s [13] ratio \( U = BR/BR^2 \), with \( R \) as the observed residuals and with \( B(n \times n) \) fixed; for further details, see [14]. However, here the unconditional distributions are all identical to their normal-theory forms as if \( \mathcal{L}(R) = N_n(0, \sigma^2 B_n) \). This is seen from the proof for Theorem 3(ii), where \( \mathcal{L}(R|\hat{\beta}_1) = N_n(0, \hat{\beta}_1^2\sigma_2^2 B_n) \) together with the scale invariance of \( U = BR/BR^2 \), assuring that \( \mathcal{L}(R|\hat{\beta}_1) = \mathcal{L}(R|\hat{\beta}_1) \) unconditionally. All such diagnostics for correlative dependencies are totally blind, both to the conditional \( V(Y|\hat{\beta}_1) = \beta_1^2\sigma_2^2 I_n + \sigma_0^2 I_n I'_n \) and unconditional \( V(Y) = \kappa_2\sigma_2^2 I_n + (\sigma_0^2 + \sigma_1^2\mu_2^2) I_n I'_n \) dispersion structures. In short, demonstrated calibration-induced correlations cannot be discerned through conventional diagnostic tools.

5.3. Non-normality

Diagnostics for normality encompass both graphical and hypothesis testing procedures. Graphs include plots of ordered residuals against their normal-theory expectations. Common usage includes the scaled residuals \( \{R_i/S_Y; 1 \leq i \leq n\} \) or the Studentized residuals \( \{W_i R_i/S_Y; i = 1, 2, \ldots, n\} \), standardized so that \( \text{var}(W_i R_i) = \sigma_Y^2 \); see Sections 2.12 and 5.7 of ref. [15], for example. In calibrated data, these residual plots are indistinguishable from those for the conventional Gaussian model \( N_n(\mu I_n, \sigma^2 I_n) \), whatever be the joint mixture density at Equation (1) for the calibrated measurements. This follows since \( \mathcal{L}(R/(R^1/2)|\hat{\beta}_1) = \mathcal{L}(R/(R^1/2)|\hat{\beta}_1) \) from scale invariance, the latter as a scaled singular multivariate Student’s \( t \) distribution having \( v = n - 1 \) degrees of freedom, depending on neither \( \hat{\beta}_1 \) nor \( \sigma_Y^2 \).

Tests for normality include the regression tests of Shapiro and Wilk [16], known to be powerful against a wide range of alternatives, including skewed or distributions having short or very long tails, even in small samples; see [17], for example. These tests utilize statistics \( W = (\sum_{i=1}^n w_i Y_{[i]}^2)/(n - 1)S_Y^2 \), where \( \{Y_{[1]} \leq Y_{[2]} \leq \cdots \leq Y_{[n]}\} \) are the ordered values of \( \{Y_1, \ldots, Y_n\} \), and \( \{w_1, \ldots, w_n\} \) are fixed weights. Such tests would appear promising for detecting the non-standard mixture distributions of calibrated measurements, where

\[
W = \frac{(\sum_{i=1}^n w_i Y_{[i]}^2)}{(n - 1)S_Y^2} = \frac{\hat{\beta}_0 \sum_{i=1}^n w_i + \hat{\beta}_1 \sum_{i=1}^n w_i Z_{[i]}^2}{(n - 1)S_Y^2}.
\]

However, since \( \sum_{i=1}^n w_i = 0 \) for these tests, together with the identity \( S_Y^2 = \hat{\beta}_1^2 S_Z^2 \), it follows that \( W = (\sum_{i=1}^n w_i Z_{[i]}^2)/(n - 1)S_Y^2 \). Then \( \mathcal{L}(W|\hat{\beta}_1) = \mathcal{L}(W) \) holds unconditionally from cancellation. Briefly, all such regression tests fail to distinguish between Gaussian distributions, and the Gaussian mixtures of type (1). With regard to further variations on regression tests, as in [18], similar arguments show that none is able to distinguish between Gaussian distributions and their mixtures from calibrated measurements. Given the sample moments \( \{m_r = \sum_{i=1}^n (Y_i - \overline{Y})^r; r = 2, 3, 4\} \), tests based on the moment ratios \( \{b_1 = m_3^2/m_2^3, b_2 = m_4/m_2^2\} \) are useful against skewed alternatives or distributions having excessive or short tails [18]. It is readily shown that these
ratios are precisely those obtainable from \{Z_1, \ldots, Z_n\}, so that their null distributions are identical to those for which \(L(Y) = N_n(\mu_1, \sigma^2 I_n)\), whatever be the joint mixture distribution as in Equation (1). On the other hand, the preceding tests do offer a clear check on normality of the distribution of \{Z_1, \ldots, Z_n\}, on which the mixtures (1) are predicated.

Briefly, conventional Gaussian diagnostics are bereft of any capacity to distinguish between Gaussian errors, and Gaussian mixtures of type (1). Thus radical calibration-induced departures from Gaussian models cannot be discerned through routine screening using any of these diagnostics.

5.4. Outliers

Commonly used diagnostics for a shift in location or scale at observation \(Y_i\) include the studentized residuals \(t_i = R_i/S_Y \sqrt{(1 - 1/n)}\) and the R-Student deletion diagnostic \(Rt_i = \frac{R_i}{S_{-i}} \sqrt{(1 - 1/n)}\), where \(S_{-i}\) is the sample standard deviation found on deleting \(Y_i\) from \{\(Y_1, \ldots, Y_n\}\}. As mixture distributions may have heavy tails, and since conventional diagnostics for normality have failed, it is natural to ask whether outlier diagnostics might be sensitive to observations from mixtures of type (1). If so, then evidence for apparent outliers in calibrated data instead might be attributable to the calibration process itself. However, these diagnostics are all scale-invariant functions of the observed residuals \{\(R_1, \ldots, R_n\}\), so that they are indistinguishable from the statistics derived from the standard Gaussian model \(N_n(\mu_1, \sigma^2 I_n)\). Briefly, conventional outlier diagnostics cannot distinguish between Gaussian errors, and between heavy-tailed mixtures as in Equation (1), even if a shift in location or scale has occurred at observation \(Y_i\).

Section 5 has re-examined whether conventional diagnostics can detect calibration-induced anomalies, including correlations, non-normality, distributions having excessive tails, and possible outliers. Even radical departures from conventional assumptions cannot be discerned through routine screening using any of the above-mentioned diagnostics. In summary, the present study cannot be dismissed as merely academic, as evidence for anomalies traceable to calibration could not have surfaced in practice through a battery of diagnostic tools as it has evolved to date.

6. Case studies

We apply the results of Section 4.1 to a numerical data set under the assumptions of Theorem 3. Table 1 gives the percent of purity (\(X\)) and the octane number (\(U\)) from a sample of \(n = 11\) different gasoline production runs. Percent purity is determined readily, whereas octane numbers require expensive and time-consuming dynamic laboratory tests; hence the need for calibration.

The least-squares fit for \(U = \beta_0 + \beta_1 (X - \bar{X}) + \varepsilon\) has \(n = 11, \hat{\beta}_0 = 87.2818, \hat{\sigma}_0 = 0.1846, \hat{\beta}_1 = 1.8546, \hat{\sigma}_1 = 0.5837\). Suppose that subsequent determinations of percent purity satisfy \(\{L(Z_i) = N_1(0, 1); 1 \leq i \leq n\}\), so that calibrated measurements are recovered as \(\{Y_i = \hat{\beta}_0 + \beta_1 Z_i; 1 \leq i \leq n\}\) in units of octane number. Then the distribution of \(\bar{Y}\) is the mixture of a normal distribution \(N_1(\mu(t), \Xi(t))\), with \(\mu(t) = \beta_0 + t \mu_Z\) and \(\Xi(t) = t^2 \sigma_Z^2/n + \sigma_0^2\), having the density \(g_1(u; \mu(t), \Xi(t))\), and a mixing distribution \(N_1(\beta_1, \sigma^2)\) having the density \(dG_1(t)\).

For convenience, we write this as \(L(\bar{Y}) = N_1(\beta_0 + t \mu_Z, t^2 \sigma_Z^2/n + \sigma_0^2) \Lambda, N_1(\beta_1, \sigma^2)\), where \(\Lambda\),

<table>
<thead>
<tr>
<th>(X)</th>
<th>99.8</th>
<th>99.7</th>
<th>99.6</th>
<th>99.5</th>
<th>99.4</th>
<th>99.3</th>
<th>99.2</th>
<th>99.1</th>
<th>99.0</th>
<th>98.9</th>
<th>98.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>(U)</td>
<td>88.6</td>
<td>86.4</td>
<td>87.2</td>
<td>88.4</td>
<td>87.2</td>
<td>86.8</td>
<td>86.1</td>
<td>87.3</td>
<td>86.4</td>
<td>86.6</td>
<td>87.1</td>
</tr>
</tbody>
</table>
designates the mixing operation. Accordingly, the density of $\bar{Y}$ is

$$f_1(u) = \int_{-\infty}^{\infty} g_1(u; \beta_0 + t\mu_Z, \frac{t^2\sigma_Z^2}{n} + \sigma_0^2) \, dG_1(t)$$

$$= \int_{-\infty}^{\infty} \frac{\exp(-[u - (\beta_0 + t\mu_Z)]^2/(2(t^2\sigma_Z^2/n + \sigma_0^2))} {\sqrt{2\pi(t^2\sigma_Z^2/n + \sigma_0^2)}} \frac{\exp(-(t - \beta_1)^2/2\sigma_1^2)} {\sqrt{2\pi\sigma_1^2}} \, dt, \tag{9}$$

a function of the parameters $\Omega = \{n, \beta_0, \sigma_0, \mu_Z, \sigma_Z, \beta_1, \sigma_1\}$, with skewness 0.1464 × 10⁻⁶ and kurtosis 3.855, and with conditional mean $E(\bar{Y}) = 87.2818$, given the empirical calibration.

Using Equation (9), we compute the 95% probability region for $\bar{Y}$ as (86.037, 88.526) compared with $\hat{\beta}_1 \pm 1.96|\hat{\beta}|\sigma_Z/\sqrt{11} = (86.184, 88.376)$ if $\bar{Y}$ were normal. This latter interval is actually a 92.2% probability region. In addition, the density $f_1(u)$ is bell-shaped but is not normal. Table 2 gives its moments (mean, variance) and moment ratios (skewness ($\gamma$), kurtosis ($\kappa$)) for selected values of the parameters $\Omega$.

The scaled sample variance is a mixture of a gamma distribution, $G_0(\cdot, \cdot)$, with mixing distribution $dG_2(w)$ as a non-central chi-squared distribution, to give $\mathcal{L}(\rho(w)) = G_0((n - 1)/2, 2\rho^2)\Lambda_1\chi_2^2(\rho = \beta_1^2/\sigma_1^2)$ with $E(S_\rho^2/\sigma_1^2) = 1 + \lambda$, which for the octane data is the conditional value $E(S_\rho^2/\sigma_1^2) = 11.095$, so that $E(S_\rho^2) = 3.780$. The mixture distribution has the density

$$f_0(u) = \frac{u^{v/2 - 1}} {2^{v/2}\Gamma(v/2)} \int_0^\infty w^{-v/2}e^{-w/2}dG_2(w)$$

$$= \frac{u^{v/2 - 1}} {2^{v/2}\Gamma(v/2)} \int_0^\infty w^{-v/2}e^{-w/2} \left[ \frac{e^{-\lambda/2-w/2}} {2^{1/2}} \sum_{j=0}^{\infty} \left( \frac{\lambda} {4} \right)^j \frac{u^{j-1/2}} {j\Gamma(1/2+j)} \right] \, dw \tag{10}$$

with $v = (n - 1)$.

If $Y$ were normal with $\hat{\beta}_1$ a constant, then a 95% probability region for $S_\rho^2$ could be found from $P(\chi^2(10; 0.025) < (n - 1)\beta_1^2/\hat{\beta}_1^2\sigma_1^2 < \chi^2(10; 0.975))$ or equivalently $P(1.1167 < S_\rho^2 < 7.0449)$, which actually is a 74% probability region when variation in $\hat{\beta}_1$ is taken into account. The correct probability region is found by numerically integrating Equation (10) to get

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta_0$</th>
<th>$\sigma_0$</th>
<th>$\mu_Z$</th>
<th>$\sigma_Z$</th>
<th>$\beta_1$</th>
<th>$\sigma_1$</th>
<th>$E(\bar{T})$</th>
<th>$var(\bar{T})$</th>
<th>$\gamma$</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.0000</td>
<td>2.2000</td>
<td>0.1839</td>
<td>3.2851</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.0000</td>
<td>2.1000</td>
<td>0.0986</td>
<td>3.1463</td>
</tr>
<tr>
<td>20</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5000</td>
<td>2.1000</td>
<td>0.0986</td>
<td>3.1463</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3.0000</td>
<td>2.1000</td>
<td>0.0986</td>
<td>3.1463</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.0000</td>
<td>1.3500</td>
<td>0.1913</td>
<td>3.3539</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.0000</td>
<td>5.1000</td>
<td>0.0260</td>
<td>3.0248</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.5000</td>
<td>1.3500</td>
<td>0.0956</td>
<td>3.1070</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3.0000</td>
<td>5.1000</td>
<td>0.0521</td>
<td>3.0940</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>2.0000</td>
<td>2.0250</td>
<td>0.0260</td>
<td>3.0373</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2.0000</td>
<td>2.4000</td>
<td>0.3327</td>
<td>3.5417</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1.5000</td>
<td>2.0625</td>
<td>0.0506</td>
<td>3.1463</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3.0000</td>
<td>2.2500</td>
<td>0.1778</td>
<td>3.1452</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>2.0000</td>
<td>1.3125</td>
<td>0.0499</td>
<td>3.0267</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2.0000</td>
<td>5.2500</td>
<td>0.0998</td>
<td>3.3614</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2.0000</td>
<td>6.2500</td>
<td>0.6144</td>
<td>5.5559</td>
</tr>
</tbody>
</table>
\[ P[10.8 < (n-1)S^2 / \sigma^2 < 336.5] = 0.95 = P[0.3680 < S^2_Y < 11.46]. \]

We find that using the first 20 terms in the infinite sum is adequate.

For the density of \( t_3^2 = n(\bar{Y} - \mu_Y^0)^2 / S^2_Y \), set

\[
\delta = \frac{(\mu_Y - \mu_Y^0)^2}{\sigma_Y^2 \sigma_Z^2}, \quad \delta(\beta_i) = \frac{\delta}{\beta_i^2 / \sigma_i^2}.
\]

Its density is found on mixing the non-central \( t^2(v, \delta / t) \) over a non-central chi-squared as the mixing distribution, which we write as \( L(t_3^2) = t^2(v, \delta / t) \Lambda_\lambda^1(\lambda = \beta_1^2 / \sigma_1^2) \). The density is given by

\[
g(u; v, \delta, \lambda) = \int_0^\infty \frac{1}{v} \sum_{j=0}^\infty \left( \frac{\delta / t}{2} \right)^j B((1 + 2j)/2, v/2)(1 + u/v)^{1/2 + u/v/2 + j} \frac{e^{-\lambda / 2 - t/2}}{2^{1/2}}
\times \sum_{k=0}^\infty \frac{(\lambda / 4)^k}{k!\Gamma(1 + 2k)/2} \, dr. \tag{11}
\]

For the first sum, we use the first \( N_1 = 15 \) terms, and for the second sum the first \( N_2 = 30 \) terms.

This distribution is useful for computing the power of the \( t^2 \) test. For example, with \( n = 11 \) so \( v = 10 \), the 95% critical value is 4.9646 with \( \delta = 0 \). Table 3 gives the power of the test for \( \delta = \{0, 1, 4, 9\} \) and \( \lambda = \{1, 4, 9\} \).

For the octane data, the power of the test of \( H_0 : \mu_Y = \mu_Y^0 \), with \( (\mu_Y - \mu_Y^0)^2 = 1 \), has \( v = 10, \ \delta = (\mu_Y - \mu_Y^0)^2 / \sigma_Y^2 \sigma_1^2 = [(1)(0.5837)]^2 = 2.9351, \ \lambda = \beta_1^2 / \sigma_1^2 = (1.8546/0.5837)^2 = 10.0953 \). The power of the test is 90%.

An equivalent form for \( L(t_3^2) \) is based on \( t_0 \) as \( L(t_0) = t(v, \delta_0 / s) \Lambda_s(\lambda = \beta_1^2 / \sigma_1^2) \), mixing over a shifted half-normal distribution. Its density is

\[
f_{t_0}(u) = \int_0^\infty \frac{e^{-((\delta_0 / s)^2)/2}}{\sqrt{\pi v}(\Gamma(v/2)} 
\times \sum_{j=0}^\infty \frac{\Gamma((v + j + 1)/2)}{j!\Gamma((v + 1)/2)} \left( \frac{\sqrt{2u\delta_0 / s}}{\sqrt{v + u^2}} \right)^j \frac{d\sqrt{G_2(s)}}{2^{1/2}}. \tag{12}
\]

having the non-centrality parameter \( \delta_0 = \sqrt{\delta} \) and

\[
d\sqrt{G_2(s)} = \frac{e^{-(s-\lambda_0)^2/2} + e^{-(s-\lambda_0)^2/2}}{2\pi} \, ds
\]

with the non-centrality parameter \( \lambda_0 = \sqrt{\lambda} \). This series has faster convergence and we used \( N = 20 \) terms in the preceding power calculations with non-centrality parameters \( \delta_0 = \sqrt{\delta} \) and

<table>
<thead>
<tr>
<th>( \lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1 )</td>
</tr>
<tr>
<td>( 0 )</td>
</tr>
<tr>
<td>( 1 )</td>
</tr>
<tr>
<td>( 4 )</td>
</tr>
<tr>
<td>( 9 )</td>
</tr>
</tbody>
</table>
\[ \lambda_0 = \sqrt{\lambda}. \] Computations reported here were executed by the second author using the Maple software package.

7. Conclusions

In summary, the widespread and necessary use of calibration may have devastating effects, even on elementary data-analytic procedures pertaining to location and scale parameters. It is unfortunate that these difficulties cannot be flagged by the ever expanding use of available diagnostic tools. It is thus incumbent on knowledgeable users of statistical methodology, and the statistical consultants advising them, to assess the extent of these difficulties as they might impact the analysis and interpretation of data in a particular experimental setting. Let the user be forewarned. Fortunately, comparisons among means and among variances, in the context of comparative one-way experiments, are largely unaffected by the use of calibrated instruments when subject to errors of calibration, provided that the results are interpreted accordingly.

References


Appendix

It is germane to examine the comparative concentration of probability measures on \( \mathbb{R}^n \). Following Sherman [2], the measure \( \mu(\cdot) \) is said to be more peaked about \( 0 \in \mathbb{R}^n \) than \( \nu(\cdot) \) if and only if \( \mu(A) \geq \nu(A) \) for every set \( A \) in the class \( \mathcal{C}(n) \) comprising the convex sets in \( \mathbb{R}^n \) symmetric under reflection through \( 0 \in \mathbb{R}^n \). For scale mixtures of Gaussian measures on \( \mathbb{R}^n \), their peakedness ordering is tantamount to the stochastic ordering of their mixing distributions. Details follow.
Lemma A1  Let $\mathbb{G}_n(\theta, \Xi, G_1)$ and $\mathbb{G}_n(\theta, \Xi, G_2)$ be Gaussian mixtures on $\mathbb{R}_n$ of type (2) having mixing distributions $G_1(\cdot)$ and $G_2(\cdot)$ on $\mathbb{R}_1$. Then $\mathbb{G}_n(\theta, \Xi, G_1)$ is more peaked about $\theta \in \mathbb{R}^n$ than $\mathbb{G}_n(\theta, \Xi, G_2)$ if and only if $G_1(t) \leq G_2(t)$ for every $t > 0$.

Proof  The ordering $G_1(t) \leq G_2(t)$; i.e., that $G_1(\cdot)$ is stochastically larger than $G_2(\cdot)$, holds if and only if there are increasing functions $\psi_1(\cdot), \psi_2(\cdot)$ ordered pointwise as $\psi_1(t) \geq \psi_2(t)$, together with a random variable $U$, such that $G_1(t) = P(\psi_1(U) \leq t)$ and $G_2(t) = P(\psi_2(U) \leq t)$; see, for example, [19, Lemma 1, p. 84]. Accordingly, we provisionally write

$$ \mu(A) = \int_A f(x; \theta, \Xi, G_1) \, dx $$
$$ \nu(A) = \int_A f(x; \theta, \Xi, G_2) \, dx $$

and their difference as

$$ \mu(A) - \nu(A) = \int_0^\infty \int_A [g(x; \theta, \Xi/\psi_1(t)) - g(x; \theta, \Xi/\psi_2(t))] \, dx \, dG(t). $$

Given that $G_1(t) \leq G_2(t)$, so that $\psi_1(t) \geq \psi_2(t)$, the ordering $\int_A [g(x; \theta, \Xi/\psi_1(t)) - g(x; \theta, \Xi/\psi_2(t))] \, dx \geq 0$ follows point wise for each fixed $t \in \mathbb{R}_1$ from [20, Corollary 3], since $\Xi/\psi_2(t) \succeq \Xi/\psi_1(t)$ uniformly in $t$. That $[\mu(A) - \nu(A)] \geq 0$ now follows directly. Conversely, suppose that $\mu(A) \geq \nu(A)$. We now apply the converse to Anderson’s [20, Corollary 3], as proved in [21], to infer that $\psi_1(t) \geq \psi_2(t)$ for each $t > 0$, thus establishing the necessity of the condition $G_1(t) \leq G_2(t)$, to complete our proof. ■