Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 134 (2005) 568–582
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Generalized inferences on the common mean of several normal populations

Shu-Hui Lin[a,*], Jack C. Lee[b]

[a] *National Taichung Institute of Technology, 129 Sanmin Road, Sec. 3, Taichung 404, Taiwan*
[b] *Institute of Statistics and Graduate Institute of Finance, National Chiao Tung University, Hsinchu 300, Taiwan*

Received 24 November 2003; accepted 18 February 2004
Available online 21 July 2004

Abstract

The hypothesis testing and interval estimation are considered for the common mean of several normal populations when the variances are unknown and possibly unequal. A new generalized pivotal is proposed based on the best linear unbiased estimator of the common mean and the generalized inference. An exact confidence interval for the common mean is also derived. The generalized confidence interval is illustrated with two numerical examples. The merits of the proposed method are numerically compared with those of the existing methods with respect to their expected lengths, coverage probabilities and powers under different scenarios.
© 2004 Published by Elsevier B.V.

MSC: 62 F03; 62 F25; 62 E25

Keywords: Coverage probability; Expected length; Generalized pivotal quantity; Generalized test variable; Graybill–Deal estimator; Power

1. Introduction

Estimating the common mean of several normal populations with unknown and possibly unequal variances is one of the oldest and most interesting problems in statistical inference. This problem arises, for example, when two or more independent agencies are involved in measuring the effect of a new drug, while utilizing several measuring instruments to measure the products produced by the same production process to estimate the average

* Corresponding author. Tel: +886-4-2219-6700; fax: +886-4-2219-6141.
E-mail address: suelin@ntit.edu.tw (S.-H. Lin).

0378-3758/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.jspi.2004.02.018

quality, or when different laboratories are employed to measure the amount of toxic waste in a river. If it is assumed that the samples collected by independent studies are from normal populations with a common mean but possibly with different variances, then the problem of interest may be to estimate or construct a confidence interval for the common mean μ of these populations. If the variances of these populations are assumed to be equal, then there are optimal methods available to make inferences on μ. However, when the variances are unknown and unequal, it is clear that the distribution of any combined estimators of μ will involve nuisance parameters. Consequentially, the standard method has serious limitations for the purpose of finding an exact confidence interval. Thus, intensive studies have been made over the last four decades from both classical and decision theoretic points of view.

In the literature, Meier (1953), Maric and Graybill (1979), Pagurova and Gurskii (1979), Sinha (1985), and Eberhardt et al. (1989) provided approximate confidence intervals for μ, centered at the well-known Graybill and Deal (1959) estimator $\hat{\mu}_{\rm GD}$ of μ, $\hat{\mu}_{\rm GD} = \sum_{i=1}^{I} n_i \bar{x}_i / s_i^2 / \sum_{i=1}^{I} n_i / s_i^2$, where $\bar{x}_i$, s_i^2 are sample means and unbiased sample variances for the ith population, $i = 1, \ldots, I$; Fairweather (1972) and Jordan and Krishnamoorthy (1996) provided exact confidence intervals for μ based on inverting weighted linear combinations of the Student's t statistics and the Fisher–Snedecor's F statistics, respectively. In general, there is no clear-cut winner between these two intervals. Fairweather's intervals are shorter than Jordan and Krishnamoorthy's when the variance ratios are small; otherwise Jordan and Krishnamoorthy's interval is narrower than Fairweather's. Therefore, some knowledge regarding the relationship between the population variances is needed to choose between these two intervals estimates. However, it should be noted that the method considered by Jordan and Krishnamoorthy (1996) does not always produce nonempty intervals. Yu et al. (1999) considered several confidence intervals that are obtained based on pivots and combinations of appropriately defined p-values. Based on simulation studies, they recommended the methods byFisher (1932), Fairweather (1972) and Jordan and Krishnamoorthy (1996) for different scenarios. The methods considered by Yu et al. (1999), however, do not always produce nonempty confidence intervals except Fairweather's method (1972). A recent work by Krishnamoorthy and Lu Yong (2003) provided a procedure based on inverting weighted linear combinations of the generalized pivotal quantities, which is similar in spirit to ours, whereas the pivotal quantity derived in this paper is based on the best unbiased estimator of μ. Both works are based on the concepts of generalized p-values and generalized confidence interval, but with different pivotal quantities.

In this paper, we intend to provide a method that is readily applicable for both hypothesis testing and interval estimation of the common mean μ. Our approach is based on the concepts of generalized p-values and generalized confidence intervals. The notions of generalized p-values and generalized confidence intervals were proposed by Tsui and Weerahandi (1989) and Weerahandi (1993) and since then these ideas have been applied to solve many statistical problems, for examples, Lin and Lee (2003) have provided exact tests in simple growth curve models and one-way ANOVA model, Lee and Lin (2004) have constructed generalized confidence intervals for the ratio of means of two normal populations, etc. The methods are exact in the sense that the tests and the confidence intervals developed are based on exact probability statements rather than on asymptotic approximations. This means that the inferences based on the generalized p-values can be made with any desired accuracy, provided that the assumed parametric model and/or other assumptions are correct. Based on the comparison studies, the expected lengths of the new confidence intervals, coverage probabilities and power performances are compared with classical method and the methods proposed by Fairweather (1972), Jordan and Krishnamoorthy (1996) and Krishnamoorthy and Lu Yong (2003). The numerical results in Sections 4 and 5 also show that our method performs better than the existing methods.

This article is organized as follows. The theory of generalized p-values and generalized confidence interval will be briefly introduced in Section 2. Our procedures for hypothesis testing and constructing the generalized confidence intervals about the common mean μ are presented in Section 3. Three existing procedures including those proposed by Fairweather (1972), Jordan and Krishnamoorthy (1996) and Krishnamoorthy and Lu Yong (2003) will be briefly addressed in Section 3. We apply these results to two sets of data, and compare our procedure with the classical method and the other methods with respect to their expected lengths in Section 4. Three simulation studies are presented in Section 5 to compare the expected lengths, the coverage probabilities and power performances of these methods in different combinations of sample sizes and variances.

2. Generalized p-values and generalized confidence intervals

The concept of generalized p-value was first introduced by Tsui and Weerahandi (1989) to deal with the statistical testing problem in which nuisance parameters are present and it is difficult or impossible to obtain a nontrivial test with a fixed level of significance. The setup is as follows. Let $\boldsymbol{X}$ be a random quantity having a density function $f(\boldsymbol{X} \mid \zeta)$, where $\zeta = (\theta, \boldsymbol{\eta})$ is a vector of unknown parameters, θ is the parameter of interest, and $\boldsymbol{\eta}$ is a vector of nuisance parameters. Suppose we are interested in testing

$$\mathrm{H}_0 : \theta \leqslant \theta_0 \quad \text{versus} \quad \mathrm{H}_1 : \theta > \theta_0, \tag{2.1}$$

where θ_0 is a specified value.

Let $\boldsymbol{x}$ denote the observed value of $\boldsymbol{X}$ and consider the generalized test variable $T(\boldsymbol{X}; \boldsymbol{x}, \zeta)$, which depends on the observed value $\boldsymbol{x}$ and the parameters ζ, and satisfies the following requirements:

(i) For fixed $\boldsymbol{x}$ and $\zeta = (\theta_0, \boldsymbol{\eta})$, the distribution of $T(\boldsymbol{X}; \boldsymbol{x}, \zeta)$ is independent of the nuisance parameters $\boldsymbol{\eta}$.
(ii) $t_{\text{obs}} = T(\boldsymbol{x}; \boldsymbol{x}, \zeta)$ does not depend on unknown parameters. (2.2)
(iii) For fixed $\boldsymbol{x}$ and $\boldsymbol{\eta}$, $P(T(\boldsymbol{X}; \boldsymbol{x}, \zeta) \geqslant t)$ is either stochastically increasing or decreasing in θ for any given t.

Under the above conditions, if $T(\boldsymbol{X}; \boldsymbol{x}, \zeta)$ is stochastically increasing in θ, then the generalized p-values for testing the hypothesis in (2.1) can be defined as

$$p = \sup_{\theta \leqslant \theta_0} P\{T(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta}) \geqslant t\} = P\{T(\boldsymbol{X}; \boldsymbol{x}, \theta_0, \boldsymbol{\eta}) \geqslant t\}, \tag{2.3}$$

where $t = T(\boldsymbol{x}; \boldsymbol{x}, \theta_0, \boldsymbol{\eta})$.

In the same setup, suppose $T_1(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta})$ satisfies the following conditions:

(i) The distribution of $T_1(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta})$ does not depend on any unknown parameters. (2.4)
(ii) The observed value of $T_1(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta})$ is free of the nuisance parameters.

Then, we say $T_1(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta})$ is a generalized pivotal quantity. If t_1 and t_2 are such that

$$P\{t_1 \leqslant T_1(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta}) \leqslant t_2\} = 1 - \alpha, \tag{2.5}$$

then, $\{\theta : t_1 \leqslant T_1(\boldsymbol{x}; \boldsymbol{x}, \theta, \boldsymbol{\eta}) \leqslant t_2\}$ is a $100(1-\alpha)\%$ generalized confidence interval for θ. For example, if the value of $T_1(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta})$ at $\boldsymbol{X} = \boldsymbol{x}$ is θ, then $\{T_1(\boldsymbol{x}; \alpha/2), T_1(\boldsymbol{x}; 1-\alpha/2)\}$ is a $(1-\alpha)$ confidence interval for θ, where $T_1(\boldsymbol{x}; \gamma)$ stands for the γth quantile of $T_1(\boldsymbol{X}; \boldsymbol{x}, \theta, \boldsymbol{\eta})$.

For further details and for several applications based on the generalized p-value, we refer to the book by Weerahandi (1995).

3. Inferences for $\boldsymbol{\mu}$

Suppose we have $I\,(I \geqslant 2)$ independent samples $(X_{i1}, X_{i2}, \ldots, X_{in_i})$ from normal populations with a common mean μ and possibly unequal variances σ_i^2, $i = 1, \ldots, I$. For the ith population, let $\bar{X}_i = 1/n_i \sum_{j=1}^{n_i} X_{ij}$ and $S_i^2 = 1/(n_i - 1) \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$ be the sample mean and sample variance, then $t_i = (\sqrt{n_i}(\bar{X}_i - \mu)/S_i)$ follows the student t distribution with $n_i - 1$ degrees of freedom and $F_i = (n_i(\bar{X}_i - \mu)^2/S_i^2)$ follows the Fisher–Snedecor's F distribution with 1 and $n_i - 1$ degrees of freedom. In this section, we will first provide a confidence interval of μ based on a generalized pivotal quantity and then briefly review three other exact confidence intervals of μ by Fairweather (1972), Jordan and Krishnamoorthy (1996) and Krishnamoorthy and Lu Yong (2003), respectively.

3.1. Solutions based on the generalized pivotal quantity and generalized test variable

Suppose we have independent samples from I normal populations with the common mean μ and possibly unequal variances σ_i^2, $i = 1, \ldots, I$. We are interested in developing a confidence interval for the common mean, μ, based on the sufficient statistics $\bar{X}_i$ and S_i^2. It is noted that $\bar{X}_i$ and S_i^2 are mutually independent with

$$\bar{X}_i \sim \mathrm{N}\left(\mu, \frac{\sigma_i^2}{n_i}\right), \quad U_i = \frac{(n_i - 1)S_i^2}{\sigma_i^2} = \frac{V_i}{\sigma_i^2} \sim \chi^2_{n_i - 1}, \quad i = 1, \ldots, I. \tag{3.1}$$

It is known that if the variances σ_i^2's are known, the best linear unbiased estimator for μ is

$$\hat{\mu} = \frac{\sum_{i=1}^{I} n_i \bar{X}_i / \sigma_i^2}{\sum_{i=1}^{I} n_i / \sigma_i^2}, \tag{3.2}$$

with $\hat{\mu} \sim \mathrm{N}(\mu, 1/\sum_{i=1}^{I} n_i/\sigma_i^2)$ and thus $\sqrt{\sum_{i=1}^{I} n_i/\sigma_i^2}(\hat{\mu} - \mu) = Z \sim \mathrm{N}(0, 1)$.

If the variance σ_i^2 for the ith population is unknown, the generalized variable for estimating σ_i^2 can be expressed as

$$R_i = \frac{\sigma_i^2}{(n_i-1)S_i^2}(n_i-1)s_i^2 = \frac{v_i}{U_i}, \quad i = 1, \ldots, I, \tag{3.3}$$

where s_i^2 and v_i denote the observed values of S_i^2 and V_i, respectively. Since the observed value of R_i is σ_i^2, the parameter of interest, then an exact $(1-\alpha)$ confidence interval for σ_i^2 can be obtained as $\{R_i(\alpha/2), R_i(1-\alpha/2)\}$, where $R_i(\gamma)$ stands for the γth quantile of R_i for $i = 1, \ldots, I$. The result is the same as the traditional confidence interval for σ_i^2 constructed by using chi-square distribution.

Let $\bar{\boldsymbol{X}} = (\bar{X}_1, \ldots, \bar{X}_I)$ and $\boldsymbol{V} = (V_1, \ldots, V_I)$ with the corresponding observed values $\bar{\boldsymbol{x}}$ and $\boldsymbol{v}$. We then define a generalized pivotal quantity for estimating the common mean μ through the best linear unbiased estimator of μ in (3.2) and (3.3) by

$$\begin{aligned} T(\bar{\boldsymbol{X}}, \boldsymbol{V}; \bar{\boldsymbol{x}}, \boldsymbol{v}) &= \frac{\sum_{i=1}^I \frac{n_i \bar{x}_i}{\sigma_i^2} \frac{V_i}{v_i}}{\sum_{i=1}^I \frac{n_i}{\sigma_i^2} \frac{V_i}{v_i}} - \frac{\sqrt{\sum_{i=1}^I \frac{n_i}{\sigma_i^2}}(\hat{\mu}-\mu)}{\sqrt{\sum_{i=1}^I \frac{n_i}{\sigma_i^2} \frac{V_i}{v_i}}} \\ &= \frac{\sum_{i=1}^I \frac{n_i U_i}{v_i} \bar{x}_i}{\sum_{i=1}^I \frac{n_i U_i}{v_i}} - \frac{Z}{\sqrt{\sum_{i=1}^I \frac{n_i U_i}{v_i}}} \\ &= \frac{\sum_{i=1}^I \frac{n_i U_i}{v_i} \bar{x}_i - Z\sqrt{\sum_{i=1}^I \frac{n_i U_i}{v_i}}}{\sum_{j=1}^I \frac{n_j U_j}{v_j}}, \end{aligned} \tag{3.4}$$

where $\bar{x}_i$ and v_i are the observed values of $\bar{X}_i$ and V_i, respectively. Because T satisfies the two conditions in (2.4) and the observed value $T(\bar{\boldsymbol{x}}, \boldsymbol{v}; \bar{\boldsymbol{x}}, \boldsymbol{v})$ of T is μ, T is indeed a generalized pivotal quantity. Therefore, we can construct a generalized confidence interval for the common mean μ based on T. Let $T(\bar{\boldsymbol{x}}, \boldsymbol{v}; \gamma)$ stand for the γth quantile of $T(\bar{\boldsymbol{X}}, \boldsymbol{V}; \bar{\boldsymbol{x}}, \boldsymbol{v})$, the exact $(1-\alpha)$ confidence interval for μ is

$$\{T(\bar{\boldsymbol{x}}, \boldsymbol{v}; \alpha/2), T(\bar{\boldsymbol{x}}, \boldsymbol{v}; 1-\alpha/2)\}. \tag{3.5}$$

Note that the distribution of $T(\bar{\boldsymbol{X}}, \boldsymbol{V}; \bar{\boldsymbol{x}}, \boldsymbol{v})$ does not depend on any unknown parameters and the observed value $T(\bar{\boldsymbol{x}}, \boldsymbol{v}; \bar{\boldsymbol{x}}, \boldsymbol{v})$ of $T(\bar{\boldsymbol{X}}, \boldsymbol{V}; \bar{\boldsymbol{x}}, \boldsymbol{v})$ is μ which is free of the nuisance parameters, σ_i^2. Hence, we can utilize Monte Carlo method to find the confidence limits in (3.5).

We next consider the problem of testing the following hypothesis concerning the common mean μ,

$$\mathrm{H}_0: \mu \leqslant \mu_0 \quad \text{versus} \quad \mathrm{H}_1: \mu > \mu_0. \tag{3.6}$$

The p-value for testing this hypothesis can be deduced directly from the generalized pivotal quantity defined by (3.4). The properties of a generalized pivotal quantity are basically the same as the first two properties of a generalized test variable and usually one can be deduced from the other. The property (ii) of (2.2) can be achieved if we define a potential test variable T_2 by setting $T_2 = T - \mu$. Then the observed value of T_2 is t_2, which is zero. From (3.4), the generalized test variable T_2 can be represented as

$$T_2 = \frac{\sum_{i=1}^{I} \frac{n_i U_i}{v_i} \bar{x}_i - Z\sqrt{\sum_{i=1}^{I} \frac{n_i U_i}{v_i}}}{\sum_{j=1}^{I} \frac{n_j U_j}{v_j}} - \mu. \tag{3.7}$$

Since each of the random variables $U_1, \ldots, U_I$ and Z are free of unknown parameters, it is clear that $P\{T_2 \leqslant t_2; \mu\} = P\{T \leqslant \mu\}$ is an increasing function of μ. This means that T_2 satisfies property (iii) of (2.2) and thus T_2 is a generalized test variable. Because T_2 is stochastically decreasing in μ, the generalized p-value for testing (3.6) is

$$\begin{aligned} p &= P\{T_2 < t_2 \mid \mu = \mu_0\} \\ &= P\{T < \mu_0\}. \end{aligned} \tag{3.8}$$

If $\mu \neq \mu_0$, the power function of tests based on the generalized p-value is to apply (3.7) by utilizing

$$Z = \sqrt{\sum_{i=1}^{I} n_i/\sigma_i^2}(\hat{\mu} - \mu_0) \sim \mathrm{N}\left(\sqrt{\sum_{i=1}^{I} n_i/\sigma_i^2}(\mu - \mu_0), 1\right). \tag{3.9}$$

It is noted that the p-values and the power of the test can be obtained in a similar manner. For example, for testing the null hypothesis of the form

$$\mathrm{H}_0 : \mu = \mu_0 \quad \text{versus} \quad \mathrm{H}_1 : \mu \neq \mu_0, \tag{3.10}$$

the p-value is

$$p = 2 * \min\{P\{T < \mu_0\}, P\{T > \mu_0\}\}, \tag{3.11}$$

where T is defined in (3.4) and H_0 can be rejected when $p < \alpha$.

3.2. Solutions based on combined tests

We will briefly introduce three exact combined tests in the literature which will be utilized to compare with our procedure in numerical examples.

3.2.1. Solutions based on linear combinations of t distributions

Fairweather (1972) suggested using W_t, a weighted linear combination of the Student's t_i statistics, with weights inversely proportional to variances $\mathrm{Var}(t_i)$, to construct confidence interval for the common mean μ, where

$$W_t = \sum_{i=1}^{I} w_i t_i, \quad w_i = \frac{[\mathrm{Var}(t_i)]^{-1}}{\sum_{i=1}^{I}[\mathrm{Var}(t_i)]^{-1}} = \frac{(n_i - 3)/(n_i - 1)}{\sum_{i=1}^{I} (n_i - 3)/(n_i - 1)}. \tag{3.12}$$

It is noted that $\min n_i > 3$ to ensure that $\mathrm{Var}(t_i)$ exists for all $i = 1, \ldots, I$ and $\sum_{i=1}^{I} w_i = 1$.

If $b_{\alpha/2}$ denotes the cut-off point of the distribution of W_t, such that for given $\alpha \in (0, 1)$,

$$1-\alpha = P(|W_t| \leqslant b_{\alpha/2}), \tag{3.13}$$

then the $100(1-\alpha)\%$ confidence interval for μ is obtained as

$$\left[\frac{\sum_{i=1}^{I}\sqrt{n_i}\,w_i\bar{x}_i/s_i}{\sum_{i=1}^{I}\sqrt{n_i}\,w_i/s_i} - \frac{b_{\alpha/2}}{\sum_{i=1}^{I}\sqrt{n_i}\,w_i/s_i}, \right. \left. \frac{\sum_{i=1}^{I}\sqrt{n_i}\,w_i\bar{x}_i/s_i}{\sum_{i=1}^{I}\sqrt{n_i}\,w_i/s_i} + \frac{b_{\alpha/2}}{\sum_{i=1}^{I}\sqrt{n_i}\,w_i/s_i} \right]. \tag{3.14}$$

Determination of the cut-off point $b_{\alpha/2}$ is not easy in practice, and approximation may be necessary. Under the additional requirement of min $n_i > 5$, Fairweather (1972) noted that $b_{\alpha/2}$ can be approximated by $ct_{1-\alpha/2}(v)$, where $t_{1-\alpha/2}(v)$ is the $(1-\alpha/2)$th quantile of the student t distribution with v degrees of freedom and

$$v = 4 + \frac{1}{\sum_{i=1}^{I} w_i^2/(n_i-5)}, \quad c = \sqrt{\frac{v-2}{v\sum_{i=1}^{I}(n_i-3)/(n_i-1)}}. \tag{3.15}$$

3.2.2. Solutions based on linear combinations of F distributions

Jordan and Krishnamoorthy (1996) suggested using $W_{\rm f}$, a weighted linear combination of the F_i statistics, with weights inversely proportional to variance $\mathrm{Var}(F_i)$ to construct the exact interval for the common mean μ, where

$$W_{\rm f} = \sum_{i=1}^{I} w_i^* F_i, \quad w_i^* = \frac{[(n_i-3)^2(n_i-5)]/[(n_i-1)^2(n_i-2)]}{\sum_{i=1}^{I}[(n_i-3)^2(n_i-5)]/[(n_i-1)^2(n_i-2)]}. \tag{3.16}$$

It is noted that min $n_i > 5$ to ensure that $\mathrm{Var}(F_i)$ exists for all $i = 1, \ldots, I$ and $\sum_{i=1}^{I} w_i^* = 1$.

If a_α denotes the cut-off point of the distribution of $W_{\rm f}$, such that for given $\alpha \in (0, 1)$,

$$1-\alpha = P(W_{\rm f} \leqslant a_\alpha), \tag{3.17}$$

then the $100(1-\alpha)\%$ confidence interval for μ is obtained as

$$\left[\sum_{i=1}^{I} p_i\bar{x}_i - \Delta, \ \sum_{i=1}^{I} p_i\bar{x}_i + \Delta \right], \tag{3.18}$$

where

$$p_i = \frac{w_i^* n_i/s_i^2}{\sum_{j=1}^{I} w_j^* n_j/s_j^2},$$

$$\Delta = \sqrt{\frac{a_\alpha}{\Sigma_{i=1}^{I} w_i^* n_i/s_i^2} - \left\{ \sum_{i=1}^{I} p_i\bar{x}_i^2 - \left(\sum_{i=1}^{I} p_i\bar{x}_i \right)^2 \right\}}. \tag{3.19}$$

Jordan and Krishnamoorthy (1996) suggested to approximate a_α by $\mathrm{d}F_{1-\alpha}(I,u)$, where $F_{1-\alpha}(I,u)$ is the $(1-\alpha)$th quantile of the F-distribution with I and u degrees of freedom and

$$u=\frac{4IM_2-2(I+2)M_1^2}{IM_2-(I+2)M_1^2},\quad d=\frac{u-2}{u}M_1 \tag{3.20}$$

with

$$M_1=E(W_{\mathrm{f}})=\sum_{i=1}^{I}\frac{w_i^*(n_i-1)}{n_i-3} \tag{3.21}$$

and

$$M_2=E(W_{\mathrm{f}})^2=3\sum_{i=1}^{I}\frac{w_i^{*2}(n_i-1)^2}{(n_i-3)(n_i-5)}+2\sum_{i>j}\frac{w_i^*w_j^*(n_i-1)(n_j-1)}{(n_i-3)(n_j-3)}. \tag{3.22}$$

As noted by Jordan and Krishnamoorthy (1996), Δ in (3.19) could be undefined, so the interval of (3.18) might be empty.

3.2.3. Solutions based on linear combinations of generalized pivot variables

Krishnamoorthy and Lu Yong (2003) suggested using W_T, a weighted linear combination of the generalized pivot variables T_i, with weights inversely proportional to variance $\mathrm{Var}(\bar{X}_i)$ to construct the exact interval for the common mean μ,

$$\begin{aligned}W_T&=\sum_{i=1}^{I}w_i^\circ T_i=\sum_{i=1}^{I}w_i^\circ\left[\bar{x}_i-\frac{\sqrt{n_i}(\bar{x}_i-\mu)}{\sigma_i}\sqrt{\frac{v_i}{n_iU_i}}\right]\\&=\frac{\sum_{i=1}^{I}\frac{n_iU_i}{v_i}\left(\bar{x}_i-Z\sqrt{\frac{v_i}{n_iUi}}\right)}{\sum_{j=1}^{I}\frac{n_jU_j}{v_j}}=\frac{\sum_{i=1}^{I}\frac{n_iU_i}{v_i}\bar{x}_i-Z\sum_{i=1}^{I}\sqrt{\frac{n_iUi}{v_i}}}{\sum_{j=1}^{I}\frac{n_jU_j}{v_j}},\end{aligned} \tag{3.23}$$

where $Z\sim \mathrm{N}(0,1)$, $U_i=((n_i-1)S_i/\sigma_i^2)=V_i/\sigma_i^2\sim\chi^2_{n_i-1}$, v_i and $\bar{x}_i$ are the observed values of V_i and $\bar{X}_i$, respectively. The weight $w_i^\circ=n_i/\sigma_i^2/\sum_{j=1}^{I}n_j/\sigma_j^2$ is taken at $V_i=v_i$ and the observed value of W_T is μ.

Comparing (3.4) with (3.23), we note that the expected length of (3.4) is shorter than that of (3.23) for the reason that $\sqrt{\sum_{i=1}^{I}(n_iU_i/v_i)}$ is less than $\sum_{i=1}^{I}\sqrt{(n_iU_i/v_i})$ and both expected lengths are identical only when one population is involved. Moreover, if U_i is replaced by its expectation n_i-1, then the confidence intervals of the common mean μ constructed by (3.4) and (3.23) become

$$\left\{\frac{\sum_{i=1}^{I}n_i\bar{x}_i/s_i^2}{\sum_{i=1}^{I}n_i/s_i^2}-Z_{1-\alpha/2}\sqrt{\sum_{i=1}^{I}\frac{n_i}{s_i^2}},\ \frac{\sum_{i=1}^{I}n_i\bar{x}_i/s_i^2}{\sum_{i=1}^{I}n_i/s_i^2}+Z_{1-\alpha/2}\sqrt{\sum_{i=1}^{I}\frac{n_i}{s_i^2}}\right\} \tag{3.24}$$

and

$$\left\{ \frac{\sum_{i=1}^{I} n_i \bar{x}_i / s_i^2}{\sum_{i=1}^{I} n_i / s_i^2} - Z_{1-\alpha/2} \sum_{i=1}^{I} \sqrt{\frac{n_i}{s_i^2}},\ \frac{\sum_{i=1}^{I} n_i \bar{x}_i / s_i^2}{\sum_{i=1}^{I} n_i / s_i^2} + Z_{1-\alpha/2} \sum_{i=1}^{I} \sqrt{\frac{n_i}{s_i^2}} \right\}, \tag{3.25}$$

respectively, where $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution. From (3.24) and (3.25), we notice that both intervals are centered at the well-known Graybill and Deal estimator with the confidence length of (3.24) shorter than that of (3.25).

4. Illustrative examples

Two examples are given to illustrate our proposed method for setting limits on the common mean of several normal populations. The first example with mild heteroscedasticity is excerpted from Meier (1953) in which four experiments are used to estimate the mean percentage of albumin in the plasma protein of normal human subjects. The second example with serious heteroscedasticity can be found in the recent papers by Eberhardt et al. (1989), Skinner (1991), Jordan and Krishnamoorthy (1996) and Krishnamoorthy and Lu Yong (2003), among others. For demonstration purposes, we will provide the results of Fairweather (1972), Jordan and Krishnamoorthy (1996), Krishnamoorthy and Lu Yong (2003) and the classical procedure with assumption of identical variance to make a comparison.

4.1. *Example 1*

The data reported by Meier (1953) and analyzed in Jordan and Krishnamoorthy (1996) are about the percentage of albumin in plasma protein in human subjects. For ease of reference, the data based on four independent experiments are reproduced in Table 1. It is assumed that the samples are from normal populations. Five confidence intervals, given in Table 2, include our proposed interval, Fairweather (1972), Jordan and Krishnamoorthy (1996), Krishnamoorthy and Lu Yong (2003) and the classical interval which is based on the pooled estimate of identical variance and the Student's *t* statistic.

The results in Table 2 suggests that when population variances are not significantly different, four intervals except Jordan and Krishnamoorthy (1996) are comparable with each other. The interval based on new generalized pivotal quantity turns out to be optimal in the sense of having the shortest observed width. It may also be noted that the interval

Table 1
Percentage of albumin in plasma protein

Experiment	n_i	Mean	Variance
A	12	62.3	12.986
B	15	60.3	7.840
C	7	59.5	33.433
D	16	61.5	18.513

Table 2
Interval estimates for μ

Methods	Interval	Width
Classical	(59.92, 62.19)	2.27
Fairweather (1972)	(59.90, 62.19)	2.29
Jordan and Krishnamoorthy (1996)	(59.56, 62.44)	2.88
Krishnamoorthy and Lu Yong (2003)	(59.79, 62.23)	2.44
Generalized interval in (3.5)	(59.92, 62.10)	2.18

Table 3
Selenium in non-fat milk powder

Methods	n_i	Mean	Variance
Atomic absorption spectrometry	8	105.00	85.711
Neutron activation:			
(1) Instrumental	12	109.75	20.748
(2) Radiochemical	14	109.50	2.729
Isotope dilution mass spectrometry	8	113.25	33.640

Table 4
Interval estimates for μ

Methods	Interval	Width
Classical	(107.75, 111.11)	3.36
Fairweather (1972)	(108.53, 110.77)	2.24
Jordan and Krishnamoorthy (1996)	(108.45, 110.67)	2.22
Krishnamoorthy and Lu Yong (2003)	(108.67, 110.53)	1.86
Generalized interval in (3.5)	(108.75, 110.51)	1.76

derived by classical method performs quite well when the population variances are only slightly different.

4.2. Example 2

The data for the second example are taken from Eberhardt et al. (1989) who reported the data on selenium in non-fat milk power by combining the results of four independent measurement methods. The data in Table 3 show that serious non-homogeneity is present. For demonstration purposes, we will also compare our interval with the intervals by Fairweather (1972), Jordan and Krishnamoorthy (1996), Krishnamoorthy and Lu Yong (2003) and the classical method. The confidence intervals and confidence widths are given in Table 4.

For this example, intervals based on Fairweather (1972), Jordan and Krishnamoorthy (1996) are similar to each other and both intervals are wider than those of ours and Krishnamoorthy and Lu Yong (2003). The interval based on our new generalized pivotal

Table 5
Expected lengths of 95% confidence intervals for the five methods

σ_1^2	σ_2^2	$n_1=10,\ n_2=10$					$n_1=15,\ n_2=15$				
		(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
5	5	2.067	2.125	2.533	2.266	2.038	1.660	1.682	1.967	1.748	1.641
5	10	2.540	2.493	2.935	2.626	2.385	2.026	1.968	2.287	2.027	1.911
5	15	2.915	2.708	3.135	2.782	2.539	2.333	2.132	2.424	2.161	2.031
5	20	3.264	2.846	3.235	2.880	2.657	2.605	2.249	2.495	2.214	2.106
5	30	3.844	3.031	3.359	2.988	2.790	3.084	2.397	2.596	2.293	2.199
5	40	4.360	3.175	3.433	3.024	2.828	3.498	2.496	2.659	2.316	2.244
5	50	4.819	3.246	3.468	3.046	2.888	3.874	2.563	2.693	2.361	2.285
σ_1^2	σ_2^2	$n_1=10,\ n_2=30$					$n_1=30,\ n_2=10$				
		(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
5	5	1.422	1.482	1.612	1.498	1.418	1.420	1.481	1.610	1.499	1.417
5	10	1.889	1.838	2.106	1.916	1.788	1.580	1.641	1.683	1.581	1.515
5	15	2.257	2.061	2.416	2.159	2.011	1.722	1.726	1.713	1.620	1.562
5	20	2.574	2.224	2.634	2.317	2.165	1.854	1.781	1.729	1.629	1.586
5	30	3.122	2.446	2.928	2.532	2.357	2.090	1.850	1.741	1.644	1.608
5	40	3.578	2.614	3.124	2.639	2.468	2.306	1.898	1.747	1.643	1.621
5	50	3.983	2.728	3.253	2.732	2.582	2.490	1.926	1.757	1.674	1.624

Table 6
Comparison of 95% coverage probabilities among the five methods

σ_1^2	σ_2^2	$n_1=10,\ n_2=10$					$n_1=15,\ n_2=15$				
		(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
5	5	0.951	0.948	0.964	0.959	0.946	0.950	0.947	0.958	0.956	0.949
5	10	0.951	0.949	0.961	0.959	0.949	0.946	0.949	0.962	0.949	0.946
5	15	0.947	0.951	0.965	0.951	0.946	0.945	0.947	0.963	0.955	0.949
5	20	0.942	0.952	0.963	0.951	0.947	0.945	0.950	0.959	0.955	0.949
5	30	0.939	0.953	0.966	0.955	0.949	0.945	0.951	0.962	0.957	0.952
5	40	0.945	0.950	0.962	0.957	0.948	0.945	0.953	0.966	0.947	0.945
5	50	0.942	0.949	0.964	0.952	0.947	0.942	0.951	0.963	0.957	0.953

σ_1^2	σ_2^2	$n_1=10,\ n_2=30$					$n_1=30,\ n_2=10$				
		(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
5	5	0.949	0.951	0.958	0.946	0.945	0.948	0.949	0.957	0.944	0.944
5	10	0.948	0.947	0.958	0.953	0.952	0.950	0.952	0.956	0.951	0.945
5	15	0.949	0.950	0.957	0.954	0.949	0.946	0.951	0.958	0.947	0.946
5	20	0.949	0.949	0.959	0.951	0.944	0.941	0.950	0.957	0.952	0.950
5	30	0.948	0.951	0.961	0.957	0.950	0.934	0.948	0.956	0.951	0.950
5	40	0.951	0.951	0.961	0.953	0.945	0.940	0.955	0.961	0.953	0.947
5	50	0.949	0.953	0.961	0.953	0.946	0.934	0.947	0.958	0.951	0.950

Table 7
Powers of the tests for testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ ($I = 2$ and $\alpha = 0.05$)

σ_2^2/σ_1^2	Tests	$n_1 = 9,\ n_2 = 9$					$n_1 = 15,\ n_2 = 10$				
		μ					μ				
		0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
1	(1)	0.13	0.36	0.67	0.89	0.98	0.16	0.49	0.82	0.97	1.00
	(2)	0.13	0.34	0.64	0.87	0.97	0.16	0.47	0.80	0.96	1.00
	(3)	0.08	0.23	0.48	0.74	0.92	0.11	0.35	0.68	0.91	0.99
	(4)	0.11	0.29	0.59	0.83	0.96	0.13	0.42	0.77	0.95	0.99
	(5)	0.14	0.37	0.67	0.89	0.98	0.17	0.48	0.82	0.97	1.00
4	(1)	0.08	0.18	0.34	0.53	0.71	0.11	0.26	0.50	0.73	0.90
	(2)	0.09	0.21	0.41	0.64	0.82	0.12	0.32	0.59	0.83	0.95
	(3)	0.07	0.15	0.31	0.52	0.72	0.09	0.27	0.55	0.81	0.95
	(4)	0.09	0.20	0.41	0.62	0.83	0.12	0.31	0.63	0.86	0.96
	(5)	0.11	0.25	0.49	0.69	0.88	0.14	0.34	0.68	0.88	0.97
9	(1)	0.08	0.13	0.20	0.31	0.44	0.08	0.17	0.31	0.48	0.66
	(2)	0.08	0.17	0.34	0.54	0.72	0.10	0.27	0.52	0.76	0.91
	(3)	0.06	0.13	0.27	0.46	0.66	0.08	0.25	0.52	0.78	0.93
	(4)	0.08	0.18	0.36	0.60	0.75	0.12	0.32	0.59	0.83	0.95
	(5)	0.09	0.22	0.41	0.65	0.80	0.13	0.34	0.62	0.85	0.96
16	(1)	0.07	0.10	0.15	0.21	0.30	0.09	0.13	0.22	0.33	0.46
	(2)	0.08	0.16	0.30	0.49	0.66	0.10	0.24	0.47	0.71	0.88
	(3)	0.05	0.13	0.26	0.44	0.64	0.08	0.24	0.51	0.77	0.92
	(4)	0.08	0.19	0.36	0.57	0.75	0.12	0.29	0.57	0.82	0.95
	(5)	0.10	0.22	0.39	0.60	0.78	0.13	0.34	0.62	0.85	0.96

quantity turns out to be the best in the sense of having the shortest observed width. Besides, the four methods allowing heterogeneity are much better than the classical method, which is calculated basing on the identical variance assumption. Thus the procedures based on an identical variance assumption would be dangerous to use when the population variances are seriously non-homogeneous.

5. Simulation studies

Some simulation studies to compare the expected lengths, 95% coverage probabilities and powers of the tests of five procedures for the common mean μ have been carried out for $I=2$, $(n_1, n_2)=(10,\ 10)$, $(15,\ 15)$, $(30, 10)$ and $(10, 30)$, respectively, for comparing the expected lengths and coverage probabilities and $(n_1, n_2)=(9, 9)$, $(n_1, n_2)=(15, 10)$ for various values of the ratio of σ_2^2/σ_1^2, with 10,000 replicates for each combination. The results appear in Tables 5, 6 and 7 . The simulated studies presented in these tables correspond to

(1) Classical method.
(2) Fairweather (1972).
(3) Jordan and Krishnamoorthy (1996).
(4) Krishnamoorthy and Lu Yong (2003).
(5) The generalized interval in (3.5).

From these tables, we find that the classical method is quite robust when the population variances are identical or with slight non-homogeneity, but, as expected, its performance grows worse as the degree of heteroscedasticity increases. For overall comparisons, the four methods allowing heteroscedasticity are much better than the method obtained by classical approach. However, it is clear from these tables that our proposed method, derived through the best linear unbiased estimator of μ, the concept of generalized p value and generalized confidence interval, is better than any of the existing methods in the senses of having the shortest expected lengths and highest powers.

Acknowledgements

We gratefully acknowledge the editor and the referee for their valuable comments and suggestions of the paper.

References

Eberhardt, K.R., Reeve, C.P., Spiegelman, C.H., 1989. A minimax approach to combining means with practical examples. Chemometr. Intell. Lab. Systems 5, 129–148.

Fairweather, W.R., 1972. A method of obtaining an exact confidence interval for the common mean of several normal populations. Appl. Statist. 21, 229–233.

Fisher, R.A., 1932. Statistical Methods for Research Workers. 4th Edition. Oliver & Boyd, London.

Graybill, F.A., Deal, R.B., 1959. Combining unbiased estimators. Biometrics 15, 543–550.

Jordan, S.M., Krishnamoorthy, K., 1996. Exact confidence intervals for the common mean of several normal populations. Biometrics 52, 77–86.

Krishnamoorthy, K., Lu Yong, Lu., 2003. Inferences on the common mean of several normal populations based on the generalized variable method. Biometrics 59, 237–247.

Lee, J.C., Lin, S.H., 2004. Generalized confidence intervals for the ratio of means of two normal populations. J. Statist. Plann. Inference 123, 49–60.

Lin, S.H., Lee, J.C., 2003. Exact tests in simple growth curve models and one-way ANOVA with equicorrelation error structure. J. Multivariate Anal. 84, 351–368.

Maric, N., Graybill, F.A., 1979. Small samples confidence intervals on common mean of two normal distributions with unequal variances. Comm. Statist.—Theory Methods A 8, 1255–1269.

Meier, P., 1953. Variance of a weighted mean. Biometrics 9, 59–73.

Pagurova, V.I., Gurskii, V.V., 1979. A confidence interval for the common mean of several normal distributions. Theory Probab. Appl. 88, 882–888.

Sinha, B.K., 1985. Unbiased estimation of the variance of the Graybill-deal estimator of the common mean of several normal populations. Can. J. Statist. 13, 243–247.

Skinner, J.B., 1991. On combining studies. Drug Inform. J. 25, 395–403.

Tsui, K., Weerahandi, S., 1989. Generalized *p*-values in significance testing of hypotheses in the presence of nuisance parameters. J. Amer. Statist. Assoc. 84, 602–607.

Weerahandi, S., 1993. Generalized confidence intervals. J. Amer. Statist. Assoc. 88, 899–905.

Weerahandi, S., 1995. Exact Statistical Methods for Data Analysis. Springer, New York.

Yu, P.L.H., Sun, Y., Sinha, B.K., 1999. On exact confidence intervals for the common mean of several normal populations. J. Statist. Plann. Inference 81, 263–277.