BIOMETRICS 52, 77–86
March 1996

Exact Confidence Intervals for the Common Mean of Several Normal Populations

Scott M. Jordan

Department of Mathematics, Arkansas Tech University,
Russellville, Arkansas 72801, U.S.A.

and

K. Krishnamoorthy

Department of Statistics, University of Southwestern Louisiana,
Lafayette, Louisiana 70504, U.S.A.

SUMMARY

The problem of interval estimation of the common mean of several normal populations when the variances are unknown and unequal is considered. Some new confidence intervals are proposed. One of them is centered at the well-known Graybill–Deal estimator of the common mean. These new intervals and an existing exact confidence interval are numerically compared with respect to their expected lengths. Based on comparison studies, some recommendations are made regarding the choice of the intervals to be used in applications.

1. Introduction

The problem of combining the results of independent investigations for the purpose of a common objective has received considerable attention in the literature. In particular, it may be necessary to combine the data sets collected by independent agencies to estimate or to test the common parameter of interest. This problem arises, for example, when two or more independent agencies are involved in measuring the effect of a new drug, when several measuring instruments are used to measure the products produced by the same production process to assess the average quality, or when different laboratories are used to measure the amount of toxic waste in a river. Meier (1953) gave an illustrative example in which four experiments are used to estimate the mean percentage of albumin in the plasma protein of normal human subjects. Other examples can be found in the recent papers by Eberhardt, Reeve, and Spiegelman (1989) and Skinner (1991). The former paper gives three examples each of which is concerned with estimating a chemical substance in nonfat milk powder by combining the results of alternative analytical methods; the latter gives examples related to clinical trials and determining the density of nitrogen using different laboratory methods. In these situations, the variances are typically unequal due to the variation in different methods or measuring instruments.

If it is assumed that the samples collected by independent studies are from normal populations, then the problem of interest may be to estimate or test the common mean μ of these populations. If the variances of these populations are assumed to be equal, then there are optimal methods available for combining the results to make inferences on μ. When the variances are unknown and arbitrary, an intensive study has been made on the theory of point estimation of the common mean. A landmark paper in this area is due to Graybill and Deal (1959). For the two-sample case, they showed first that an unbiased estimator has smaller variance than either sample mean provided both sample sizes are greater than 10. Since then there have been numerous articles written by generalizing and extending their result—for example, see Cohen and Sackrowitz (1974), Norwood

Key words: Balanced incomplete block design; Behrens-Fisher problem; Combined estimator; Combined test; Coverage probability; t-interval.

and Hinkelmann (1977), Bhattacharya (1984), and Kubokawa (1987) and the references therein. However, only very limited results were obtained on interval estimation of μ. In the following, we give a brief review of the results on interval estimation of μ available in the literature.

Because the Graybill–Deal estimator $\hat{\mu}_{GD}$ is well known and widely accepted as a point estimator, several authors have proposed approximate confidence intervals for μ that are centered at $\hat{\mu}_{GD}$. Meier (1953) suggested a method for setting an approximate confidence interval for μ centered at $\hat{\mu}_{GD}$. Maric and Graybill (1979) and Pagurova and Gurskii (1979) have developed approximate confidence intervals centered at $\hat{\mu}_{GD}$ using Welch's (1947) approach. Not only are these intervals approximate, but also the tail points needed to compute them are depending on sample data, which makes them somewhat unappealing. Hinkley (1979) also developed an approximate confidence interval using a fiducial method of inference based on the maximum likelihood principle. Eberhardt et al. (1989) suggested approximate methods of combining independent results assuming that the individual estimates are biased and the bounds for the biases are known.

In the context of interblock analysis of a balanced incomplete block design, Brown and Cohen (1974), Bhattacharya and Shah (1978), and Khatri and Shah (1981) have proposed confidence intervals that are centered at some combined estimators based on all the samples. These intervals may be more appropriate in the analysis of balanced incomplete block design where we have some information about the variances, namely, that the intrablock variance is smaller than interblock variance.

It is clear that the distribution of any combined estimator of μ will involve nuisance parameters and so the standard method has serious limitations for the purpose of finding an exact confidence interval. Instead, one can combine independent confidence intervals centered at individual sample means in a reasonable way so that the combined interval has exact prespecified coverage probability. This can be achieved by essentially inverting combined tests for testing μ. This approach is inherent in Fairweather (1972), in which an exact confidence interval is developed based on a linear combination of independent Student's t variables.

In this article, we investigate confidence intervals that can be extracted from some exact combined tests for μ. Although there are several combined tests available (e.g., Cohen and Sackrowitz, 1984; Zhou and Mathew, 1993; and Mathew, Sinha and Zhou, 1993), not all of them are easily invertible to get explicit confidence intervals for μ. Therefore, we restrict our attention to those tests from which confidence intervals for μ can be easily extracted. In the next section, we obtain interval estimates by inverting two combined tests for μ. Interestingly, the test based on a linear combination of independent F-statistics yielded an interval estimate centered at the Graybill–Deal estimator. In passing, we also point out that a conservative test for the well-known Behrens–Fisher problem can be deduced from this interval. These intervals are exact, simple to compute, and easy to use. The necessary percentile points are presented and also a convenient method of approximating them is suggested. Section 3 contains two illustrative examples. In Section 4, for the purpose of comparing these new confidence intervals and the one by Fairweather (1972), we estimate their expected lengths using the Monte Carlo method. Based on comparison studies, some recommendations are made regarding the choice of the intervals to be used in different situations. In Section 5, we make some remarks about the estimators considered in this article and the approach used to derive them.

2. Confidence Intervals for μ

Suppose that we have independent samples from k normal populations with the same mean μ but possibly different unknown variances $\sigma_1^2, \ldots, \sigma_k^2$. Let $\bar{x}_i$ and s_i^2, respectively, denote the mean and variance of the sample of size n_i from the i th population, and let $m_i = n_i - 1$, for $i = 1, \ldots, k$. The problem of interest here is to develop confidence intervals for μ based on all the sample means and variances that are minimal sufficient statistics.

In the following, we first develop confidence intervals for μ based on independent Student's t variables.

2.1 *Interval Estimates Based on Student's t Variables*

Let $t_i = \sqrt{n_i}(\bar{x}_i - \mu)/s_i$, for $i = 1, \ldots, k$, and $M_t = \max_{1 \le i \le k} \{|t_i|\}$. This M_t is used in the literature as a test statistic for testing $H_0 : \mu = \mu_0$ (e.g., see Cohen and Sackrowitz, 1984). If μ is unknown, then solving the inequality $M_t \le c$ for μ, where c is the $100(1-\alpha)$th percentile point of M_t, one can get a confidence interval for μ. This leads to the interval

$$\left(\max_{1 \le i \le k} \{\bar{x}_i - c s_i/\sqrt{n_i}\}, \min_{1 \le i \le k} \{\bar{x}_i + c s_i/\sqrt{n_i}\} \right), \tag{2.1}$$

which has exact coverage probability $1-\alpha$. Note that interval (2.1) is the intersection of all the intervals $(\bar{x}_i - cs_i/\sqrt{n_i},\ \bar{x}_i + cs_i/\sqrt{n_i})$, $i = 1, \ldots, k$. Using the fact that $t_1, \ldots, t_k$ are statistically independent, respectively, with $m_1, \ldots, m_k$ degrees of freedom, the point c can be obtained numerically by solving the equation

$$\prod_{i=1}^{k} \Pr(-c \le t_i \le c) = 1-\alpha$$

for given α. Although this interval is simple to compute, it necessitates new table values for c. To avoid this, one can use the interval

$$\left(\max_{1\le i\le k} \{\bar{x}_i - c_i s_i/\sqrt{n_i}\},\ \min_{1\le i\le k} \{\bar{x}_i + c_i s_i/\sqrt{n_i}\} \right), \tag{2.2}$$

where c_i is chosen such that $\Pr(|t_i| \le c_i) = (1-\alpha)^{1/k}$. Clearly, interval (2.2) also has coverage probability exactly $1-\alpha$. The c_i's can be obtained from the existing tables or the calculators that compute the percentile points of the Student's t distribution.

Remark 2.1. Intervals (2.1) and (2.2) may possibly be empty. However, under the model assumption of common mean, they are more likely nonempty. Also, a simulation study (see Section 3) exhibited that when $\alpha = 0.05$ less than 2% of the generated samples yielded empty intervals and when $\alpha = 0.01$ almost all the generated samples yielded nonempty intervals under different parameter configurations. However, we are unable to prove this analytically.

The interval given in Fairweather (1972) is obtained by solving the inequality

$$|W_t| = \left| \sum_{i=1}^{k} u_i t_i \right| \le b, \tag{2.3}$$

where $b = 100(1-\alpha/2)$ percentile point of W_t and $u_i = (\mathrm{var}(t_i))^{-1}/[\sum_{j=1}^{k} (\mathrm{var}(t_j))^{-1}]$. Inequality (2.3) yields the interval

$$\frac{\sum_{i=1}^{k} u_i \sqrt{n_i}\bar{x}_i/s_i}{\sum_{i=1}^{k} u_i \sqrt{n_i}/s_i} \pm b/\left(\sum_{i=1}^{k} u_i \sqrt{n_i}/s_i \right). \tag{2.4}$$

The percentile points of W_t can be obtained exactly by numerical methods for smaller values of k. For $k = 2$, we give the values of b for $\alpha = 0.05$ in Table 5.

We next develop interval estimates for μ from a linear combination of independent F variables.

2.2 *Interval Estimates Based on Independent F Variables*

Define $F_i = n_i(\bar{x}_i - \mu)^2/s_i^2$, and let $W_f = \sum_{i=1}^{k} w_i F_i$, where w_i's are positive numbers such that $\sum_{i=1}^{k} w_i = 1$. Let $\hat{\mu} = \sum_{i=1}^{k} p_i \bar{x}_i$, where $p_i = (w_i n_i/s_i^2)/[\sum_{j=1}^{k} w_j n_j/s_j^2]$, $i = 1, \ldots, k$. It can be easily checked that $\hat{\mu}$ is the weighted least-squares estimator in the sense that it minimizes $\sum_{i=1}^{k} w_i F_i$ with respect to μ. Therefore, if we can find an a such that

$$W_f = \sum_{i=1}^{k} w_i F_i \le a \tag{2.5}$$

holds with probability $(1-\alpha)$, then the set of values of μ that satisfy inequality (2.5) is an exact confidence interval centered at $\hat{\mu}$. In fact, solving inequality (2.5) for μ, we get the interval estimate

$$\sum_{i=1}^{k} p_i \bar{x}_i \pm \sqrt{a/\left(\sum_{i=1}^{k} w_i n_i/s_i^2\right) - \left[\sum_{i=1}^{k} p_i \bar{x}_i^2 - \left(\sum_{i=1}^{k} p_i \bar{x}_i\right)^2\right]}, \tag{2.6}$$

for μ, which has coverage probability $(1-\alpha)$. Note that, when w_i's are equal to 1, the interval (2.6) is centered at the Graybill–Deal estimator of μ.

Regarding the choice of w_i, one can choose w_i as inversely proportional to the $\text{var}(F_i) = 2m_i^2(m_i-1)/[(m_i-2)^2(m_i-4)]$, $i = 1, \ldots, k$. Again, to satisfy the restriction that $\sum_{i=1}^k w_i = 1$, a reasonable choice of w_i is

$$w_i = (\text{var}(F_i))^{-1} \Big/ \left[\sum_{j=1}^{k} (\text{var}(F_j))^{-1}\right] \tag{2.7}$$

for $i = 1, \ldots, k$.

For $k = 2$, we computed the exact 95th percentile points of W_f with w_i given in (2.7) using IMSL subroutines FDF and QDAG; these are presented in Table 6. For $k \geq 3$, it is difficult to compute the exact percentile points of W_f. However, one can approximate the distribution of W_f by the distribution of $dF_{k,\nu}$, where $F_{k,\nu}$ denotes an F random variable with k and ν degrees of freedom. The unknown positive constants d and ν can be estimated in the traditional way by equating the first two moments of W_f to those of $dF_{k,\nu}$. If $\min\{m_i\} > 4$, then the estimated values are

$$\nu = \frac{4kM_2 - 2(k+2)M_1^2}{kM_2 - (k+2)M_1^2} \quad \text{and} \quad d = (\nu - 2)M_1/\nu, \tag{2.8}$$

where $M_1 = E(W_f) = \sum_{i=1}^k w_i m_i/(m_i - 2)$ and $M_2 = E(W_f)^2 = \sum_{i=1}^k 3w_i^2 m_i^2/[(m_i-2)(m_i-4)] + 2\sum_{i>j} w_i w_j m_i m_j/[(m_i-2)(m_j-2)]$. This approximation not only is simple to use but also gives very satisfactory results. For $k = 2$ and for the choice of w_i in (2.7), both the exact and approximate percentile points are presented in Table 1. We observe from this table that the approximation is good even for small degrees of freedom, except the cases where one of the sample sizes is very small compared to the other. In these cases, the approximation gives percentile points larger than the exact ones and thus leads to conservative intervals. For $k \geq 3$, we compared the percentile points computed using this approximation with those estimated using simulation. In general, they exhibited similar agreement as in the case $k = 2$.

Table 1

Exact and approximate 95th percentile points of W_f when $k = 2$

m_1, m_2	Exact	Approximation
5, 5	5.652	5.765
6, 6	5.022	5.098
8, 8	4.364	4.401
10, 12	3.919	3.946
5, 10	4.635	5.136
5, 20	4.157	4.998
6, 15	4.157	4.565
5, 6	5.313	5.470
6, 7	4.814	4.907
8, 9	4.265	4.304
11, 14	3.793	3.818
5, 15	4.312	5.037
6, 10	4.440	4.679
15, 15	3.631	3.639

Remark 2.2. Interval (2.6) is nonempty provided the sum of the terms under the square root is nonnegative. It can be easily verified that when $k = 2$ this sum is nonnegative if

$$(\bar{x}_1 - \bar{x}_2)^2/[s_1^2/(w_1 n_1) + s_2^2/(w_2 n_2)] \leq a.$$

It is shown in the Appendix that this inequality holds with probability larger than $(1 - \alpha)$. In general, under the model assumption, it is more likely that interval (2.6) is nonempty, which can be argued as follows: For any given data, the first term under the radical sign in (2.6) is positive;

furthermore, as $p_i > 0$ and $\sum_{i=1}^k p_i = 1$, $[\sum_{i=1}^k p_i\bar{x}_i^2 - (\sum_{i=1}^k p_i\bar{x}_i)^2]$ is the variance of the sample means, which is expected to be small under the model assumption that population means are equal. Hence, the failure of (2.6) to provide an nonempty interval is an indication that the population means are not really equal. In any case, the probability that this interval will be empty is no larger than α, because (2.6) has confidence coefficient $(1-\alpha)$. This implies that the test that rejects the null hypothesis of equal population mean if this interval is empty or equivalently

$$\sum_{i=1}^k p_i\bar{x}_i^2 - \left(\sum_{i=1}^k p_i\bar{x}_i\right)^2 > a/\left(\sum_{i=1}^k w_i n_i/s_i^2\right) \tag{2.9}$$

is a conservative test for the well-known Behrens–Fisher problem.

Remark 2.3. If the means are especially different, then in fact a combined interval should not be constructed for that set of samples. In these situations, the interval obtained by using any combined procedure may be dubious. The test in (2.9) helps in making a decision regarding whether to combine the means or not.

3. Illustrative Examples

We now compute 95% confidence intervals for the example given in Meier (1953) and for one of the examples in Eberhardt et al. (1989). Since the raw data are not available, it is assumed that the samples in these examples are from normal populations. Furthermore, for both data sets, (2.6) yielded nonempty intervals. This means that the model assumption of common mean is tenable based on conservative test (2.9).

EXAMPLE 1. (Meier, 1953). The results are the outcomes of four experiments used to estimate the mean percentage of albumin (μ) in the plasma protein of normal human subjects. For ease of reference, they are reproduced in Table 2a. Furthermore, as mentioned in the footnote of Meier's paper, it is possible that experiments with large variances may be biased. If the estimates are known to be biased, then the results of Eberhardt et al. (1989) can be used for the estimation. However, we consider this example for demonstration purposes. Bartlett's test for the equality of the variances showed no significant differences among the variances (p value = 0.14) and, hence, we also give the classical interval based on the pooled estimate of the common σ^2 and Student's t statistic. The computed intervals are given in Table 2b. The critical value b in (2.4) is computed using the approximation suggested by Fairweather (1972).

Table 2a
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 2b
Interval estimates

Intervals	Critical values	Weights	Interval
(2.1)	$c = 3.043$		60.82 ± 1.68
(2.2)	$c_i's = 2.9702, 2.8543, 3.5055, 2.8272$		60.78 ± 1.58
(2.4)	$b \doteq 1.102$	$u_i's = 0.2550, 0.2671, 0.2708, 0.2701$	61.04 ± 1.15
(2.6)	$a \doteq 3.191$	$p_i's = 0.2100, 0.5245, 0.0181, 0.2474$	61.00 ± 1.44
Classical	$t_{46}(.025) = 2.013$		61.05 ± 1.13

This example suggests that the usual classical method is optimal when the variances are not significantly different; however, the results of (2.4) and (2.6), which allow divergence in the population variances, are consistent with that of the classical method.

EXAMPLE 2 (Eberhardt et al. 1989). This example is concerned with the estimation of Selenium in non-fat milk powder by combining the results of four different analytical methods which are given in Table 3a. Eberhardt et al. (1989) gives interval estimates based on Student's-t approximation. Applying Bartlett's test, we found that the variances are significantly different (p-value < 0.00001) and hence the classical method used in Example 1 is not applicable. The computed intervals are given Table 3b.

Table 3a
Selenium in non-fat milk powder

Methods	n_i	Mean	Variance
Atomic absorption spectrometry	8	105.0	85.711
Neutron activation			
Instrumental	12	109.75	20.748
Radiochemical	14	109.5	2.729
Isotope dilution mass spectrometry	8	113.25	33.640

Table 3b
Interval estimates

Intervals	Critical values	Weights	Interval
(2.1)	$c = 3.128$		109.5 ± 1.38
(2.2)	c_i's $= 3.321, 2.970, 2.886, 3.321$		109.5 ± 1.27
(2.4)	$b \doteq 1.118$	$u_i's = 0.2309, 0.2645, 0.2736, 0.2309$	109.7 ± 1.11
(2.6)	$a \doteq 3.341$	$p_i's = 0.0068, 0.0777, 0.8908, 0.0247$	109.6 ± 1.08

For this example, intervals (2.4) and (2.6) are almost identical. Furthermore, both examples suggest that intervals (2.4) and (2.6) are preferable to (2.1) and (2.2) in terms of the length.

We next estimate the expected lengths of these intervals using the Monte Carlo method for the purpose of comparing them over different sample sizes and parameter configurations.

4. Simulation Study

In this section, we estimate the expected lengths of the confidence intervals considered in this article using Monte Carlo simulation. The random variables are generated by IMSL subroutines RNNOR (for normal) and RNCHI (for chi-squared variable). Each simulation consists of 100,000 runs. The simulated expected lengths are presented in Tables 4a and 4b for the following four intervals:

1. The interval in (2.1).
2. The interval in (2.2).
3. The interval (2.4) due to Fairweather (1972).
4. The interval in (2.6) with w_i given in (2.7).

Note that when the sample sizes are equal, intervals 1 and 2 are identical. Also, for fewer than 2% of the generated samples, intervals 1, 2, and 4 became empty. Their expected lengths are estimated by the average lengths of only nonempty intervals.

Table 4a gives estimated lengths of the confidence intervals for equal degrees of freedom. In this case, note that the expected length of an interval at (σ_1^2, σ_2^2) is equal to the expected length at (σ_2^2, σ_1^2). In Table 4b, we present the estimated lengths of the intervals for unequal degrees of freedom. From these tables we observe the following:

1. Interval 2 is, in general, preferable to 1 for practical use. Comparison of interval 2 with interval 4 indicates that 4 has shorter expected lengths than 2 except the situation where the sample sizes are unequal and one of the variances is much larger than the other. However, on an overall basis, interval 4 is preferable to intervals 1 and 2 as an estimate of μ.

2. There is no clearcut winner between intervals 3 and 4. When the sample sizes are equal, interval 3 has shorter expected lengths than 4 except when one of the variances is much greater than the other. On the other hand, when $m_1 > m_2$, interval 4 is expected to have shorter length than 3 if $\sigma_1^2 < \sigma_2^2$ and vice versa if $\sigma_1^2 > \sigma_2^2$. Thus, some knowledge regarding the relationship between the population variances is needed to choose between these two interval estimates.

In general, the width of interval 4 is less sensitive to the difference between the variances compared to interval 3.

Table 4a

Simulated expected widths of 95% intervals

	$m_1 = m_2 = 10$			$m_1 = m_2 = 15$		
(σ_1^2, σ_2^2)	2	3	4	2	3	4
(5,5)	2.54	2.04	2.30	2.02	1.60	1.85
(5,10)	2.94	2.36	2.68	2.31	1.88	2.14
(5,15)	3.07	2.57	2.87	2.42	2.04	2.28
(5,20)	3.13	2.73	2.95	2.48	2.14	2.36
(5,25)	3.18	2.80	3.03	2.51	2.23	2.41
(5,30)	3.22	2.85	2.98	2.54	2.29	2.45
(5,40)	3.24	3.00	3.10	2.56	2.38	2.50
(5,50)	3.22	3.12	3.14	2.58	2.45	2.52
(5,100)	3.32	3.36	3.25	2.61	2.64	2.59
(5,250)	3.33	3.60	3.33	2.63	2.85	2.63
(5,500)	3.36	3.74	3.34	2.63	2.96	2.64
(5,1000)	3.39	3.86	3.36	2.64	3.04	2.65
Critical values	2.626	1.566	4.026	2.483	1.500	3.631

Table 4b

Simulated expected widths of 95% intervals

	$m_1 = 30, m_2 = 10$				$m_1 = 10, m_2 = 30$			
(σ_1^2, σ_2^2)	1	2	3	4	1	2	3	4
(5,5)	1.73	1.68	1.42	1.55	2.03	1.99	1.62	1.83
(5,10)	1.82	1.75	1.58	1.63	2.50	2.50	1.97	2.30
(5,15)	1.84	1.77	1.66	1.66	2.72	2.74	2.18	2.56
(5,20)	1.86	1.78	1.72	1.68	2.84	2.88	2.33	2.72
(5,25)	1.87	1.78	1.76	1.69	2.92	2.97	2.44	2.85
(5,30)	1.87	1.89	1.79	1.69	2.98	3.04	2.53	2.95
(5,40)	1.88	1.79	1.83	1.70	3.04	3.12	2.67	3.07
(5,50)	1.88	1.80	1.87	1.71	3.09	3.17	2.78	3.16
(5,100)	1.89	1.80	1.95	1.71	3.16	3.26	3.08	3.36
(5,250)	1.90	1.80	2.03	1.72	3.19	3.30	3.41	3.49
(5,500)	1.90	1.81	2.08	1.73	3.21	3.33	3.61	3.56
(5,1000)	1.90	1.81	2.11	1.73	3.21	3.33	3.77	3.57
(10,5)	2.22	2.18	1.76	2.01	2.18	2.12	1.83	1.97
(15,5)	2.48	2.47	1.96	2.29	2.24	2.17	1.95	2.03
(20,5)	2.65	2.66	2.12	2.49	2.26	2.19	2.03	2.06
(25,5)	2.75	2.78	2.23	2.63	2.27	2.20	2.08	2.08
(30,5)	2.83	2.87	2.33	2.75	2.28	2.21	2.13	2.09
(40,5)	2.92	2.99	2.47	2.91	2.30	2.22	2.20	2.11
(50,5)	2.98	3.06	2.58	3.01	2.31	2.23	2.24	2.12
(100,5)	3.09	3.21	2.92	3.30	2.32	2.24	2.36	2.14
(250,5)	3.16	3.30	3.30	3.51	2.33	2.24	2.49	2.16
(500,5)	3.17	3.32	3.52	3.58	2.33	2.25	2.55	2.16
(1000,5)	3.18	3.33	3.71	3.64	2.33	2.25	2.60	2.16
Critical values	2.489	—	1.498	3.604	2.521	—	1.515	3.708

We also estimated the expected lengths of these confidence intervals at different sample sizes and parameters configuration. They are not reported here, as they all exhibited a pattern similar to those in Tables 4a and 4b.

Table 5

95th percentile point of W_t[a]

m_1, m_2	5	6	7	8	9	10	11	12
5	1.814							
6	1.766	1.722						
7	1.733	1.692	1.663					
8	1.709	1.669	1.641	1.621				
9	1.690	1.652	1.625	1.605	1.590			
10	1.675	1.638	1.612	1.593	1.578	1.566		
11	1.663	1.627	1.602	1.583	1.568	1.557	1.547	
12	1.653	1.618	1.593	1.574	1.560	1.549	1.540	1.532
13	1.645	1.610	1.585	1.567	1.553	1.542	1.533	1.526
14	1.638	1.603	1.579	1.561	1.547	1.536	1.528	1.520
15	1.632	1.598	1.574	1.556	1.542	1.532	1.523	1.516
20	1.611	1.578	1.555	1.538	1.525	1.515	1.506	1.499
25	1.598	1.567	1.544	1.528	1.515	1.505	1.496	1.490
30	1.590	1.559	1.537	1.521	1.508	1.498	1.490	1.483
40	1.580	1.549	1.528	1.512	1.500	1.490	1.482	1.475
50	1.574	1.544	1.523	1.507	1.495	1.485	1.477	1.471
∞	1.550	1.521	1.501	1.486	1.475	1.466	1.458	1.452
m_1, m_2	13	14	15	20	25	30	40	50
13	1.519							
14	1.514	1.509						
15	1.509	1.504	1.500					
20	1.493	1.488	1.484	1.469				
25	1.484	1.479	1.475	1.460	1.451			
30	1.478	1.473	1.469	1.454	1.446	1.440		
40	1.470	1.465	1.461	1.447	1.438	1.433	1.426	
50	1.465	1.460	1.456	1.442	1.434	1.428	1.422	1.418
∞	1.447	1.442	1.438	1.425	1.417	1.412	1.405	1.401
(∞, ∞)	1.386							

[a] 99th percentile points are available upon request from the authors.

5. Concluding Remarks

Interval (2.4) due to Fairweather is centered at $\hat{\mu}_t = \sum_{i=1}^k (\bar{x}_i\sqrt{n_i}/s_i)/\left(\sum_{i=1}^k \sqrt{n_i}/s_i\right)$ when the sample sizes are equal. In contrast to $\hat{\mu}_{GD}$, the estimator $\hat{\mu}_t$ is the weighted average of independent sample means with weights inversely proportional to the sample standard deviations rather than the sample variances. As a point estimator, $\hat{\mu}_t$ is not known to have any desirable properties except unbiasedness. It can be easily verified that, when σ_i^2's are known, the estimator $\left[\sum_{i=1}^k (\bar{x}_i\sqrt{n_i})/\sigma_i\right] / \left(\sum_{i=1}^k \sqrt{n_i}/\sigma_i\right)$ is not uniformly better than either of the sample means. In view of this fact, it is plausible that, as a point estimator $\hat{\mu}_t$ need not be more efficient than either of the sample means uniformly over the parameter space. Thus, $\hat{\mu}_{GD}$ is preferable to $\hat{\mu}_t$ as a point estimator of μ. However, the performance of the interval (2.4) centered at $\hat{\mu}_t$ is not inferior to that of (2.6) centered at $\hat{\mu}_{GD}$ in terms of the expected length.

In view of Mathew et al. (1993), who proposed several combined tests for testing μ, it may be tempting to invert some of their tests to get interval estimates for μ. But investigation by Jordan and Krishnamoorthy (1994) indicates that none of their tests can be easily inverted to get an interval estimator of μ in an explicit form and the acceptance regions need not always be convex. Besides, they may not yield interval estimates that are uniformly shorter than the ones considered in this article. Nevertheless, they can provide exact interval estimates for any $k \geq 2$ because they are all exact tests. Furthermore, in light of recent work by Zhou and Mathew (1994), the approach used to get (2.6) can be extended to the multivariate case. We are currently investigating this problem and plan to report it separately.

Table 6

95th percentile point of W_f[a]

m_1, m_2	5	6	7	8	9	10	11	12
5	5.652							
6	5.313	5.022						
7	5.068	4.814	4.631					
8	4.887	4.657	4.491	4.364				
9	4.746	4.536	4.383	4.265	4.171			
10	4.635	4.440	4.296	4.185	4.097	4.026		
11	4.545	4.362	4.226	4.119	4.036	3.968	3.912	
12	4.472	4.297	4.167	4.065	3.984	3.919	3.866	3.821
13	4.410	4.243	4.117	4.020	3.941	3.879	3.827	3.783
14	4.357	4.197	4.075	3.980	3.904	3.844	3.793	3.751
15	4.312	4.157	4.039	3.947	3.873	3.813	3.764	3.723
20	4.157	4.021	3.914	3.830	3.763	3.708	3.662	3.624
25	4.067	3.940	3.840	3.761	3.697	3.645	3.602	3.565
30	4.008	3.887	3.792	3.715	3.654	3.604	3.562	3.527
40	3.936	3.822	3.731	3.659	3.600	3.552	3.512	3.479
50	3.893	3.783	3.696	3.625	3.568	3.521	3.483	3.450
∞	3.726	3.632	3.556	3.493	3.442	3.400	3.365	3.336
m_1, m_2	13	14	15	20	25	30	40	50
13	3.746							
14	3.714	3.684						
15	3.688	3.657	3.631					
20	3.592	3.564	3.540	3.455				
25	3.535	3.508	3.485	3.403	3.355			
30	3.497	3.471	3.449	3.370	3.322	3.291		
40	3.450	3.425	3.403	3.327	3.282	3.251	3.213	
50	3.422	3.397	3.376	3.302	3.257	3.228	3.190	3.168
∞	3.311	3.289	3.270	3.202	3.161	3.133	3.099	3.079
(∞, ∞)	2.996							

[a] 99th percentile points are available upon request from the authors.

Acknowledgements

The authors are thankful to the referees and an associate editor for their comments and suggestions. They are grateful to Charles Anderson for his comments, which led to Remark 2.3.

Résumé

Nous examinons le problème de l'estimation par intervalle de la moyenne commune à plusieurs populations Normales lorsque les variances sont inconnues et différentes. Quelques nouveaux intervalles de confiance sont proposés. L'un d'eux est centré sur le bien connu estimateur de Graybille–Deal. Ces nouveaux intervalles et l'intervalle de confiance exact sont comparés relativement à leurs tailles attendues. Quelques recommandations sont faites, à partir d'études comparatives, vis-à-vis du choix des intervalles à retenir dans des applications.

References

Bhattacharya, C. G. (1984). Two inequalities with an application. *Annals of the Institute of Statistical Mathematics* **36,** 129–134.

Bhattacharya, C. G. and Shah, K. R. (1978). Interval estimation of treatment differences in block designs. *Journal of Statistical Computation and Simulation* **6,** 243–255.

Brown, L. D. and Cohen, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. *The Annals of Statistics* **2,** 963–976.

Cohen, A. and Sackrowitz, H. B. (1974). On estimating the common mean of two normal distributions. *The Annals of Statistics* **2,** 1274–1282.

Cohen, A. and Sackrowitz, H. B. (1984). Testing hypotheses about the common mean of normal distributions. *Journal of Statistical Planning and Inference* **9,** 207–227.

Eberhardt, K. R., Reeve, C. P., and Spiegelman, C. H. (1989). A minimax approach to combining means, with practical examples. *Chemometrics and Intelligent Laboratory Systems,* **5,** 129–148.

Fairweather, W. R. (1972). A method of obtaining an exact confidence interval for the common mean of several normal populations. *Applied Statistics* **21,** 229–233.

Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators. *Boimetrics* **15,** 543–550.

Hinkley, D. V. (1979). A Note on the weighted means problems. *Scandinavians Journal of Statistics* **6,** 37–40.

Khatri, C. G. and Shah, K. R. (1981). Interval Estimation of the Common Mean. *Communication in Statistics—Simulation and Computation,* **B10,** 99–107.

Jordan, S. J. and Krishnamoorthy, K. (1994). *Exact confidence intervals for the common mean of several normal populations.* Technical Report 94-3, Department of Statistics, University of Southwestern Louisiana, Lafayette.

Kubokawa, T. (1987). Admissible minimax estimation of a common mean of two normal populations. *The Annals of Statistics* **15,** 1245–1256.

Maric, N. and Graybill, F. A. (1979). Small samples confidence intervals on common mean of two normal distributions with unequal variances. *Communication in Statistics—Theory and Methods* **A8,** 1255–1269.

Mathew, T., Sinha, B. K., and Zhou, L. (1993). Some statistical procedures for combining independent tests. *Journal of the American Statistical Association* **88,** 912–919.

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

Norwood, T. E. and Hinkelmann, K. (1977). Estimating the common mean of several normal populations. *The Annals of Statistics* **5,** 1047–1050.

Pagurova, V. I. and Gurskii, V. V. (1979). A confidence interval for the common mean of several normal distributions. *Theory of Probability and its Applications* **24,** 882–888.

Skinner, J. B. (1991). On combining studies. *Drug Information Journal* **25,** 395–403.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika* **34,** 28–35.

Zhou, L. and Mathew, T. (1993). Combining independent tests in linear models. *Journal of the American Statistical Association* **88,** 650–655.

Zhou, L. and Mathew, T. (1994). Combining independent tests in multivariate linear models. *Journal of Multivariate Analysis* **51,** 265–276.

Received February 1995; *accepted April* 1995.

APPENDIX

Interval (2.6) is nonempty if the sum of the terms under the radical sign is nonnegative or, equivalently,

$$\frac{w_1 n_1 w_2 n_2 (\bar{x}_1 - \bar{x}_2)^2}{(w_1 n_1 s_2^2 + w_2 n_2 s_1^2)} \leq a. \tag{A.1}$$

It can be easily verified that (A.1) holds if and only if

$$\frac{w_1 w_2 X_1}{w_1(1-\beta)X_2/m_2 + w_2 \beta X_3/m_1} \leq a, \tag{A.2}$$

where X_1, X_2, and X_3 are independent chi-squared random variables, respectively, with 1, m_2 and m_1 degrees of freedom and $\beta = (\sigma_1^2/n_1)/(\sigma_1^2/n_1 + \sigma_2^2/n_2)$. Again, (A.2) can be expressed as $w_1 w_2/[w_1(1-\beta)F_{m_2,1} + w_2 \beta F_{m_1,1}] \leq a$, where $F_{m_1,1}$ and $F_{m_2,1}$ are dependent F random variables. Now, using the facts that $F_{a,b}$ is distributed as $1/F_{b,a}$ and $\beta X + (1-\beta)Y \geq \min(X, Y)$, it can be checked that (A.2) holds if

$$\max(w_1 F_{1,m_1}, w_2 F_{1,m_2}) \leq a. \tag{A.3}$$

Because W_f is stochastically larger than both $w_1 F_{1,m_1}$ and $w_2 F_{1,m_2}$ and $\Pr(W_f \leq a) = 1 - \alpha$, inequality (A.3) holds with probability larger than $(1-\alpha)$.