ON POOLING DATA I: ESTIMATION OF THE MEAN

J. S. MEHTA AND R. SRINIVASAN

(Received Sept. 10, 1969; revised March 23, 1971)

1. Summary

Methods are developed for estimating the mean of a normal population utilising the information available from two samples by properly modifying the continuous weight function approach. The resulting estimators are shown to be better in some cases than the ones obtained using the test estimation procedure due to Bancroft [2], [3] and Berkson [4]. The method of comparison is based on the concept of premium and protection introduced by Anscombe [1].

2. Introduction

The following problem is of frequent occurrence in statistical practice. If we have a random sample of n measurements, $x_1, x_2, \cdots, x_n$, from a certain normal population $N(\mu, \sigma^2)$ made by an experienced observer, and a second similar sample of size m, $y_1, y_2, \cdots, y_m$, taken and recorded by a less experienced observer, how shall we estimate μ by utilizing the information contained in both the samples? The procedure that is often followed in the above as well as several similar situations is to estimate the mean by pooling the estimates based on the separate samples. In the sequel we shall refer to this as the AP procedure (always pool). Thus according to the AP procedure, μ is estimated by

$$A=\frac{n\bar{x}+m\bar{y}}{n+m} \tag{2.1}$$

where

$$\bar{x}=\frac{1}{n}\sum_1^n x_i\ , \tag{2.2}$$

and

$$\bar{y}=\frac{1}{m}\sum_1^m y_i\ . \tag{2.3}$$

But, considering that the second observer is less experienced, it is obviously more realistic to take into account in estimating μ the possibility that his sample comes from a 'similar' but different normal population. For example, it could well be that $y_1, y_2, \cdots, y_m$ is a random sample from $N(\mu+a\sigma, \sigma^2)$ where a is an unknown constant, or that it is a sample from $N(\mu, \kappa\sigma^2)$ where κ is unknown. If we allow for this possibility then the following methods for estimating μ should also be considered as possible alternatives to the AP procedures. (1) *The TE procedure.* Use an appropriate statistic to test whether the two samples come from the same population, and use this test as a criterion in making the decision as to whether to pool the two means or not. If the test discriminates properly between cases where pooling should or should not be made, the preliminary test of significance criterion will utilize the extra m observations from the second sample whenever permissible. This procedure is due to Bancroft ([2], [3]) and Berkson [4]; we shall refer to it in this paper as the TE procedure (test estimation). (2) *The CW procedure.* Always pool the two sample means $\bar{x}$ and $\bar{y}$ using continuous weights as opposed to discrete weights used in the AP procedure. According to this approach, the suggested estimator of μ is of the form $\phi(R)\bar{x}+[1-\phi(R)]\bar{y}$ where R is an appropriate test statistic for deciding whether the two samples come from the same population or not, and $\phi(R)$ is a continuous function of R. This procedure was suggested by Huntsberger [7], and has been successfully used by Mehta [9] and Mehta and Gurland [10] in different situations. We will call this the CW procedure (continuous weight).

It is clear that even though A, as given by (2.1), is the best unbiased estimator of μ if the two samples happen to come from the same population, the estimators given by the TE or CW procedure have the advantage of guarding against possible loss in efficiency when, in fact, the populations are different. In this paper we shall study the relative merits of these two procedures as compared to the AP procedure. An appropriate criterion for comparison of the two procedures can be based on the concept of premium and protection introduced by Anscombe [1]. For illustration, consider the case when $x_1, x_2, \cdots, x_n$ come from $N(\mu, \sigma^2)$ and $y_1, y_2, \cdots, y_m$ is a sample from $N(\mu+a\sigma, \sigma^2)$ where a is unknown. Let U be an estimator of μ based on either the TE or the CW procedure, and let A be the always pool estimator given by (2.1). Since A is the best unbiased estimator of μ if a happens to be zero (the null case), we suffer a loss by using U when $a=0$. A convenient measure of this loss is given by the premium (in analogy with a fire insurance policy) of U defined by

$$\text{Premium } (U)=\left.\frac{\text{MSE }(U)-\text{MSE }(A)}{\text{MSE }(A)}\right|_{a=0} \tag{2.4}$$

where MSE stands for mean squared error. On the other hand, we experience a gain in the use of U over that of A when a is different from zero. This gain can be measured by the protection of U defined by

$$\text{(2.5)} \qquad \text{Protection } (U) = \frac{\text{MSE } (A) - \text{MSE } (U)}{\text{MSE } (A)}\bigg|_{a\neq 0} .$$

It is now clear how two estimators T and C, given respectively by the TE and CW procedure, can be compared with respect to A. Choose the variables involved in T and C so that they have a common preassigned premium. For this fixed premium base the choice on the behavior of the protections afforded by them as functions of the nuisance parameter. We shall follow this technique throughout this paper.

In Section 3 we consider the case where the first population is $N(\mu, \sigma^2)$ and the second is $N(\mu+a\sigma, \sigma^2)$ where σ^2 is known. Section 4 deals with the same situation with the added assumption that σ^2 is unknown. Finally, in Section 5 we present the case where both populations have the same mean but different unknown variances. In each case we have fixed the premium at .05 and studied the protection afforded for various values of the nuisance parameter and the sample sizes.

3.

In this section we will assume that $x_1, x_2, \cdots, x_n$ is a sample from $N(\mu, \sigma^2)$ and $y_1, y_2, \cdots, y_m$ is a sample from $N(\mu+a\sigma, \sigma^2)$. It is required to estimate μ when σ^2 is known while a is unknown.

The estimator T_1 derived using the TE procedure is given by [8]

$$\text{(3.1)} \qquad T_1 = \begin{cases} \bar{x} & \text{if } |z| \geqq \xi_\alpha \sigma_z \\ A = \dfrac{n\bar{x}+m\bar{y}}{n+m} & \text{if } |z| < \xi_\alpha \sigma_z \end{cases}$$

where

$$\text{(3.2)} \qquad z = \bar{x} - \bar{y} , \qquad \sigma_z^2 = \sigma^2(1/n+1/m)$$

and ξ_α is the solution of $1-\Phi(\xi_\alpha)=\alpha/2$ where Φ denotes the cumulative distribution function of the standardized normal variate.

Letting $\delta = -a\sqrt{(nm/n+m)}$ and ϕ denote the density of the standard normal variate, it is seen that

$$\text{(3.3)} \quad \text{E } (T_1) = \mu + \sqrt{\frac{m}{n(n+m)}}\,\sigma\{\delta[\Phi(\delta+\xi_\alpha) - \Phi(\delta-\xi_\alpha)] + \phi(\delta+\xi_\alpha) - \phi(\delta-\xi_\alpha)\}$$

and

$$\text{(3.4)} \qquad \text{MSE}\,(T_1)=\frac{\sigma^2}{n}+\frac{\sigma^2 m}{n(n+m)}\left\{\sigma^2\int_{\delta-\xi_\alpha}^{\delta+\xi_\alpha}\phi(u)du-\int_{\delta-\xi_\alpha}^{\delta+\xi_\alpha}u^2\phi(u)du\right\}\,.$$

We can now compute the premium and protection of T_1 using (3.4) for various values of ξ_α. It is seen that the premium equals 0.05 when $\xi_\alpha=2.678$. For this value of ξ_α the protection of T_1 is computed for various values of a and given in Table 1. We have chosen $n=8$ and $m=6$.

Table 1. Protection afforded by T_1 when Premium $(T_1)=0.05$, $n=8$ and $m=6$

a	Protection (T_1) $\xi_\alpha=2.6780$	a	Protection (T_1) $\xi_\alpha=2.6780$
± 0.02	−0.050	± 1.20	0.177
± 0.04	−0.050	± 1.40	0.327
± 0.06	−0.051	± 1.60	0.482
± 0.08	−0.051	± 1.80	0.621
± 0.20	−0.058	± 2.00	0.732
± 0.40	−0.071	± 4.00	0.959
± 0.60	−0.066	± 6.00	0.981
± 0.80	−0.027	± 8.00	0.989
± 1.00	0.055	±10.0	0.993

We now turn to the derivation of the estimator C_1 given by the CW procedure. Let us write our estimator in the following initial form

$$\text{(3.5)} \qquad U=\alpha\bar{x}+(1-\alpha)\bar{y}$$

where $\bar{x}$ and $\bar{y}$ are given by (2.2) and (2.3) respectively, and α is to be suitably selected. We determine α by requiring that it minimizes the MSE of C which is given by

$$\text{(3.6)} \qquad \text{MSE}\,(C)=\sigma^2(\alpha^2/n)+(1-\alpha)^2(1/m+a^2)\,.$$

This is minimized with respect to α when

$$\text{(3.7)} \qquad \alpha=(1/m+a^2)/(1/n+1/m+a^2)$$

which is a function of a^2—an unknown quantity. In order to use C in any given situation a^2 must be estimated.

Now $(\bar{y}-\bar{x})^2\sim\sigma^2(1/n+1/m)\chi^2_{1,a^2}$ which is a weighted non-central chi-square with non-centrality parameter a^2. Since

$$\text{(3.8)} \qquad \text{E}\,[(\bar{y}-\bar{x})^2]=\sigma^2(1/n+1/m)(1+a^2)$$

it seems reasonable to estimate a^2 by

$$\hat{a}^2=(d/\sigma^2)(\bar{y}-\bar{x})^2+c \tag{3.9}$$

where c and d are constants and will eventually be determined so as to give suitable properties to the final estimator C_1. If we substitute in (3.7) the estimate of a^2 given by (3.9), and insert the resulting expression for α in (3.4) we obtain C in the form

$$C=\bar{x}+\frac{m\sigma^2(\bar{y}-\bar{x})}{\sigma^2[(n+m)+cnm]+dnm(\bar{y}-\bar{x})^2}\ . \tag{3.10}$$

To facilitate derivations of the first two moments of our estimator we have slightly modified C without any loss of generality and obtain the final estimator C_1 according to the CW procedure as

$$C_1=\bar{x}+\frac{m\sigma^2(\bar{y}-\bar{x})}{L+M(\bar{y}-\bar{x})^2} \tag{3.11}$$

where $L=c\sigma^2(n+m)$; $M=dnm$. W can obtain the following expressions for E (C_1) and MSE (C_1) the details of which are omitted.

$$\text{E}\,(C_1)=\mu+\frac{\sigma\sqrt{nm^3}}{M\sqrt{2\pi(n+m)}}\,J_2\left(\frac{L}{M},\,\frac{nm}{25^2(n+m)},\,a\sigma\right) \tag{3.12}$$

$$\begin{aligned}\text{MSE}\,(C_1)=&\frac{2\sigma^2d^2(n+m)-\sigma^2m(1+4d)}{2d^2n(n+m)}+\frac{\sqrt{m}\,\sigma^2}{d^2\sqrt{8n\pi(n+m)}}\\ &\cdot\left\{\frac{(1+4d)[c(n+m)+dnma^2]+d(n+m)}{dn(n+m)}\right.\\ &\cdot J_1\left(\frac{c(n+m)}{dnm},\,\frac{nm}{2(n+m)},\,a\right)+\frac{a(1+4d)}{n}\,\frac{d}{da}\\ &\left.\cdot J_1\left(\frac{c(n+m)}{dnm},\,\frac{nm}{2(n+m)},\,a\right)\right\}\ .\end{aligned} \tag{3.13}$$

The functions J_1 and J_2 are defined by:

$$J_1(\alpha,\,\beta,\,\gamma)=\int_{-\infty}^{\infty}\frac{1}{\alpha+t^2}e^{-\beta(t-\gamma)^2}dt$$

$$J_2(\alpha,\,\beta,\,\gamma)=\int_{-\infty}^{\infty}\frac{t}{\alpha+t^2}e^{-\beta(t-\gamma)^2}dt\ .$$

Using the auxiliary integrals

$$\bar{J}_n(\alpha,\,\beta)=\int_{\sqrt{\beta}}^{\infty}\frac{e^{-\alpha v^2}}{v^{2n}}dv \qquad (n=0,\,1,\,2,\cdots)$$

it can be shown that

$$J_1(\alpha, \beta, \gamma)=2e^{\beta(\alpha-\gamma^2)}\sqrt{\pi}\sum_{n=0}^{\infty}\frac{(\beta\gamma)^{2n}}{n-1}\bar{J}_n(\alpha, \beta) .$$

The evaluation of J_1 can now be completed by noting that

$$\bar{J}_n(\alpha, \beta)=\frac{e^{-\alpha\beta}}{(2n-1)\beta^{n-1/2}}-\frac{2\alpha}{2n-1}\bar{J}_{n-1}(\alpha, \beta) \qquad (n=1, 2, \cdots)$$

and that $\bar{J}_0(\alpha, \beta)$ is essentially the standard normal integral. In order to evaluate J_2, we simply use the relation:

$$J_2(\alpha, \beta, \gamma)=\frac{1}{2\beta}\frac{d}{d\gamma}J_1(\alpha, \beta, \gamma)+\gamma J_1(\alpha, \beta, \gamma) .$$

Since the integrand in (3.12) is an odd function of t when $a=0$, we see that C_1 is unbiased for μ in the null case, i.e., when $a=0$. It can also be verified that MSE (C_1) is symmetric about the origin with respect to a.

Using (3.13) we can now easily calculate the premium and protection of C_1. As in the case of T_1, we shall fix the premium of C_1 at 0.05. For a given m and n, there are several possible choices of c and d values which give a premium of 0.05. The behavior of the pro-

Table 2. Comparison of the protection afforded by C_1 for several values of c and d giving Premium $(C_1)=0.05$ when $n=8$ and $m=6$

a	Protection (C_1)					
	$c=0.7000$ $d=0.1600$	$c=0.8000$ $d=0.1570$	$c=0.9000$ $d=0.1375$	$c=1.000$ $d=0.1100$	$c=1.1000$ $d=0.0800$	$c=1.2000$ $d=0.0484$
± 0.02	−0.050	−0.050	−0.050	−0.050	−0.050	−0.050
± 0.04	−0.050	−0.049	−0.049	−0.049	−0.049	−0.049
± 0.06	−0.050	−0.048	−0.048	−0.047	−0.046	−0.046
± 0.08	−0.049	−0.047	−0.046	−0.044	−0.044	−0.042
± 0.20	−0.043	−0.031	−0.023	−0.017	−0.012	−0.008
± 0.40	−0.011	0.024	0.045	0.061	0.074	0.085
± 0.60	0.057	0.106	0.135	0.155	0.171	0.183
± 0.80	0.153	0.205	0.233	0.251	0.263	0.269
± 1.00	0.265	0.312	0.334	0.344	0.348	0.344
± 1.20	0.379	0.416	0.430	0.432	0.427	0.411
± 1.40	0.484	0.513	0.519	0.513	0.499	0.471
± 1.60	0.577	0.597	0.598	0.586	0.565	0.527
± 1.80	0.655	0.669	0.655	0.650	0.624	0.578
± 2.00	0.719	0.728	0.722	0.705	0.676	0.625
± 4.00	0.944	0.943	0.944	0.934	0.922	0.890
± 6.00	0.978	0.978	0.978	0.976	0.972	0.960
± 8.00	0.988	0.988	0.988	0.988	0.986	0.982
±10.00	0.993	0.993	0.993	0.992	0.992	0.990

tection of C_1 for various such solutions is presented in Table 2 for $m=8$ and $n=6$. The results of Table 2 indicate that by increasing c we can obtain increased protection for C_1 in an a-interval around the origin. In so doing, however, we pay a price by decreasing the protection obtained for values of a lying outside this interval. The choice of the solution thus depends to a large extent on the discretion of the statistican and any idea he might have regarding the possible range for the nuisance parameter a.

With the help of Tables 1 and 2 we can now compare the protection afforded by T_1 with that of C_1 for various values of c and d. Confining attention to the case when $c=0.8000$ and $d=0.1570$, we note that C_1 affords much better protection than T_1 for values of a in the interval $(-2, 2)$, i.e., when the second population has a mean value lying within 2σ limits of the first one. This is obviously the case of utmost interest to us since deviations by any greater amount are much less likely to occur in practice. It is also clear that, for values of a larger in absolute value than 2, the protection afforded by C_1, even though less than that of T_1, is not significantly different from it. We could even safely assume that they are equal. Thus we can conclude that when a lies in the interval $(-1, 1)$ the analysis comes out overwhelmingly in favour

Table 3. Comparison of the protection afforded by C_1 for several values of m and d giving Premium $(C_1)=0.05$ when $n=12$ and $c=0.9000$

a	Protection (C_1)			
	$m=4$ $d=0.2551$	$m=6$ $d=0.1828$	$m=8$ $d=0.1483$	$m=10$ $d=0.1276$
± 0.02	−0.050	−0.050	−0.050	−0.049
± 0.04	−0.049	−0.049	−0.048	−0.048
± 0.06	−0.049	−0.048	−0.047	−0.046
± 0.08	−0.047	−0.046	−0.044	−0.042
± 0.20	−0.034	−0.024	−0.015	−0.006
± 0.40	0.010	0.045	0.072	0.093
± 0.60	0.080	0.142	0.185	0.213
± 0.80	0.168	0.253	0.305	0.337
± 1.00	0.265	0.365	0.422	0.455
± 1.20	0.362	0.471	0.528	0.561
± 1.40	0.454	0.564	0.619	0.650
± 1.60	0.537	0.643	0.694	0.723
± 1.80	0.608	0.708	0.754	0.780
± 2.00	0.669	0.760	0.802	0.824
± 4.00	0.913	0.946	0.960	0.966
± 6.00	0.962	0.978	0.984	0.987
± 8.00	0.979	0.988	0.991	0.993
±10.00	0.986	0.992	0.994	0.996

of C_1 (note that Protection $(T_1) \leqq 0$ in this interval), that when $1 \leqq a \leqq 2$ we should still favour C_1 over T_1 (Protection (C_1) is almost twice Protection (T_1) in this interval), and that for all other values of a the two protections are practically the same.

Note that even though the above analysis is based on a common premium level of 0.05 there is no reason to believe that the basic conclusions would be any different if the premium were fixed at 0.01 say. The same comment holds good regarding the sizes of the samples, n and m. However, to gain some idea about the behaviour of C_1 for varying sample sizes, we have calculated the protection afforded by C_1 for various values of m, when $n=12$, Premium $(C_1)=0.05$ and c is fixed at 0.9000. The results of this calculation, presented in Table 3, indicate that the protection of C_1 can be uniformly increased as m increases.

4.

In this section we consider the same basic situation as in Section 3 except that we relax the assumption made there that σ^2 is known. Thus now we are interested in estimating μ when σ^2 is unknown.

For this case the estimator T_2 derived using the TE prcedure is given by [6].

$$(4.1) \qquad T_2 = \begin{cases} \bar{x} & \text{if } |t| \geqq t_\alpha \\ \dfrac{n\bar{x}+m\bar{y}}{n+m} & \text{if } |t| < t_\alpha \end{cases}$$

where t_α is the $(1-\alpha/2)100\%$ point of the t variate with $n+m-2$ d.f.

It can be shown [6] that for $n+m$ even and >4 we have

$$(4.2) \qquad \mathrm{E}\,(T_2) = a\sqrt{\frac{m}{n(n+m)}}\sqrt{\frac{t_\alpha^2}{t_\alpha^2+n+m-2}} \cdot \exp\{-(a^2/2)[nm/n+m][n+m-2/t_\alpha^2+n+m-2]\} \cdot \left\{\sum_{i=0}^{(n+m-4)/2}\left(\frac{n+m-2}{2t_\alpha^2}\right)^i \frac{\mu'_{2i+1}}{i!}\right\}$$

and

$$(4.3) \qquad \mathrm{MSE}\,(T_2) = \frac{\sigma^2}{n}[1+f(a)]$$

where

$$(4.4) \qquad f(a) = \frac{n}{n+m}\sqrt{\frac{t_\alpha^2}{t_\alpha^2+n+m-2}} \exp\left\{\frac{-(a^2/2)[nm/n+m][n+m-2]}{t_\alpha^2+n+m-2}\right\} \cdot \left\{\sum_{i=0}^{(n+m-4)/2}\left(\frac{n+m-2}{2t_\alpha^2}\right)^i \frac{1}{i!}\left[2a\sqrt{\frac{nm}{n+m}}\mu'_{2i+1} - \mu'_{2i+2}\right]\right\}$$

where

$$\mu'_k=\sum_{j=0}^{k}\binom{k}{j}\sigma_z^j\mu_z^{k-j}\mu_j^0 \tag{4.5}$$

$$\sigma_z^2=\frac{t_\alpha^2}{t_\alpha^2+n+m-2}$$

$$\mu_z=a\sqrt{\frac{nm}{n+m}}\frac{t_\alpha^2}{t_\alpha^2+n+m-2}$$

and μ_j^0 is the jth moment of a standard normal variate.

The premium and protection of T_2 with respect to A can now be calculated using the above formulas. As in the previous section we have fixed the premium at 0.05 and chosen $n=8$ and $m=6$. For these values we have computed Protection (T_2) at different a values and the results are given in Table 4.

Table 4. Protection afforded by T_2 when Premium $(T_2)=0.05$, $n=8$ and $m=6$

a	Protection (T_2) $t_\alpha=3.056$	a	Protection (T_2) $t_\alpha=3.056$
± 0.02	−0.050	± 1.20	0.137
± 0.04	−0.050	± 1.40	0.250
± 0.06	−0.050	± 1.60	0.378
± 0.08	−0.051	± 1.80	0.507
± 0.20	−0.054	± 2.00	0.625
± 0.40	−0.057	± 4.00	0.958
± 0.60	−0.047	± 6.00	0.981
± 0.80	−0.014	± 8.00	0.989
± 1.00	0.047	±10.00	0.993

We now turn to the derivation of the estimator C_2 following the CW procedure. It is easily seen that C_2 is of the same form as C_1 given by (3.11) except that σ^2 is unknown now and has to be estimated. We use s^2 to estimate σ^2 where

$$s^2=\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-2} \tag{4.6}$$

and

$$(n-1)s_1^2=\sum_1^n(x_i-\bar{x})^2\ ,$$

$$(m-1)s_2^2=\sum_1^m(y_i-\bar{y})^2\ .$$

C_2 can now be given as

$$C_2=\bar{x}+\frac{ms^2(\bar{y}-\bar{x})}{cs^2(m+n)+dmn(\bar{y}-\bar{x})^2}\,.$$

The following expressions can be obtained:

$$\text{(4.8)}\qquad \mathrm{E}\,(C_2)=\mu+ma\sigma(n+m-2)^{(n+m)/2}\exp\left\{\frac{-mna^2}{2(m+n)}\right\}I_1$$

$$\text{(4.9)}\qquad \mathrm{MSE}\,(C_2)=\frac{\sigma^2}{n}+\sigma^2(n+m-2)^{(n+m)/2}\exp\left\{\frac{-mna^2}{2(m+n)}\right\}$$

$$\cdot\left\{\frac{m(n+m)}{2nc}\,[I_2-(n+m-2)I_3]+\left[\frac{2a^2m^2I_1}{n+m}-\frac{2mI_2}{n}\right]\right\}$$

where

$$I_1=J\left(\frac{n+m}{2},\frac{3}{2}\right)$$

$$I_2=J\left(\frac{n+m}{2},\frac{3}{2}\right)+\frac{mna^2}{m+n}J\left(\frac{n+m}{2},\frac{5}{2}\right)$$

$$I_3=J\left(\frac{n+m+2}{2},\frac{3}{2}\right)+\frac{mna^2}{m+n}J\left(\frac{m+n+2}{2},\frac{5}{2}\right)$$

Table 5. Comparison of the protection afforded by C_2 for several values of c and d giving Premium $(C_2)=0.05$ when $n=8$ and $m=6$

a	Protection (C_2)					
	$c=0.7000$ $d=0.1184$	$c=0.800$ $d=0.1284$	$c=0.900$ $d=0.1142$	$c=1.000$ $d=0.0922$	$c=1.000$ $d=0.0670$	$c=1.2000$ $d=0.0404$
± 0.02	−0.050	−0.050	−0.050	−0.050	−0.050	−0.050
± 0.04	−0.050	−0.049	−0.049	−0.049	−0.048	−0.048
± 0.06	−0.050	−0.048	−0.048	−0.047	−0.046	−0.046
± 0.08	−0.050	−0.047	−0.046	−0.045	−0.044	−0.043
± 0.20	−0.050	−0.034	−0.025	−0.018	−0.013	−0.008
± 0.40	−0.034	0.014	0.039	0.057	0.072	0.084
± 0.60	0.016	0.089	0.124	0.148	0.166	0.180
± 0.80	0.100	0.182	0.218	0.240	0.256	0.265
± 1.00	0.204	0.284	0.314	0.330	0.338	0.338
± 1.20	0.314	0.385	0.447	0.415	0.414	0.403
± 1.40	0.419	0.480	0.494	0.494	0.484	0.461
± 1.60	0.515	0.565	0.572	0.565	0.548	0.514
± 1.80	0.598	0.638	0.647	0.629	0.605	0.564
± 2.00	0.667	0.700	0.698	0.683	0.657	0.609
± 4.00	0.933	0.937	0.934	0.926	0.910	0.874
± 6.00	0.975	0.976	0.975	0.973	0.968	0.952
± 8.00	0.987	0.988	0.987	0.986	0.984	0.978
±10.00	0.992	0.992	0.992	0.992	0.991	0.988

$$J(k,l)=\int_0^\infty \frac{\exp\left\{\dfrac{mna^2}{4d(m+n)^2t+2(m+n)}\right\}dt}{[2c(m+n)t+m+n-2]^k[2d(m+n)t+1]^l}\ .$$

It is easily seen from the definition of $J(k,l)$ that C_2 is unbiased for μ in the null case, and that MSE (C_2) is symmetric about the origin with respect to a.

In Table 5 we have presented the protection of C_2 with respect to A for $m=8$, $n=6$ and Premium $(C_2)=0.05$ corresponding to several possible solutions in terms of c and d. Table 6 gives the protection of C_2 for various values of m, when $n=12$, Premium $(C_2)=0.05$ and c is fixed at 0.9000. These tables are analogous respectively to Tables 2 and 3 of the previous section. The comparison of T_2 and C_2 and the behavior of C_2 are similar to the corresponding aspects of T_1 and C_1 given in Section 3.

Table 6. Comparison of the protection afforded by C_2 for several values of giving Premium $(C_2)=0.05$ when $c=0.8$

a	Protection (C_2)			
	$m=4$ $d=0.1827$	$m=6$ $d=0.1492$	$m=8$ $d=0.1492$	$m=10$ $d=0.1296$
± 0.02	−0.050	−0.050	−0.049	−0.049
± 0.04	−0.049	−0.049	−0.049	−0.049
± 0.06	−0.049	−0.048	−0.048	−0.047
± 0.08	−0.048	−0.047	−0.046	−0.046
± 0.20	−0.037	−0.031	−0.026	−0.023
± 0.40	−0.002	0.026	0.043	0.054
± 0.60	−0.066	0.115	0.147	0.167
± 0.80	0.150	0.224	0.268	0.295
± 1.00	0.247	0.339	0.392	0.423
± 1.20	0.346	0.450	0.505	0.538
± 1.40	0.440	0.548	0.603	0.634
± 1.60	0.526	0.631	0.682	0.712
± 1.80	0.600	0.699	0.746	0.772
± 2.00	0.662	0.753	0.795	0.819
± 4.00	0.912	0.945	0.959	0.966
± 6.00	0.962	0.977	0.984	0.987
± 8.00	0.979	0.988	0.991	0.993
±10.00	0.987	0.992	0.994	0.996

5.

In this section we assume that $x_1, x_2, \cdots, x_n$ are from $N(\mu, \sigma^2)$, and $y_1, y_2, \cdots, y_m$ are from $N(\mu, k\sigma^2)$ where σ^2 and $k\ (\geqq 1)$ are unknown. The problem is to estimate μ.

The TE procedure estimator T_3 of μ is given by

$$T_3=\begin{cases} \bar{x} & \text{if } F\geqq F_\alpha \\ \dfrac{n\bar{x}+m\bar{y}}{n+m} & \text{if } F<F_\alpha \end{cases} \tag{5.1}$$

where $F=s_2^2/s_1^2$ with s_1^2 and s_2^2 given by (4.6), and F_α is the $(1-\alpha)100\%$ point of an F variate with $m-1$ and $n-1$ degrees of freedom. The following identities can be easily proved:

$$\mathrm{E}\,(T_3)=\mu \tag{5.2}$$

$$\mathrm{MSE}\,(T_3)=\sigma^2\left[\frac{n+km}{(n+m)^2}P(F<F_\alpha)+\frac{1}{n}P(F\geqq F_\alpha)\right]. \tag{5.3}$$

The probabilities in the above expression can be expressed in terms of Incomplete Beta Functions. It is seen that when $F_\alpha=3.5$, T_3 attains the premium of 0.05 when $n=8$ and $m=6$. For these choices of F_α, n and m we have obtained the protection afforded by T_3 with respect to A for $k=2(1)10$. These are given in Table 7. Note that the null case corresponds to $k=1$.

Table 7. Protections afforded by T_3 and C_3 when Premium (T_3)=Premium $(C_3)=0.05$ $(n=8,\ m=6)$

k	Protection (T_3) $(F_\alpha=3.5)$	Protection (C_3) $(c=1.32,\ d=1.326)$
2	−0.054	0.069
3	0.024	0.195
4	0.127	0.298
5	0.228	0.381
6	0.316	0.447
7	0.392	0.502
8	0.456	0.547
9	0.509	0.585
10	0.555	0.618

The derivation of the CW procedure estimator C_3 is similar to that of C_1 and C_2. It is of the form [9]

$$C_3=\frac{(c+F)\bar{x}+d\bar{y}}{(c+d+F)} \tag{5.4}$$

where $F=s_2^2/s_1^2$. The derivation of the first two moments of C_3 is rather straightforward, and we get

$$\mathrm{E}\,(C_3)=\mu \tag{5.5}$$

and

$$\text{(5.6)}\qquad \text{MSE}\,(C_3)=\sigma^2\frac{\Gamma((n+m-2)/2)}{\Gamma((n-1)/2)\Gamma((m-1)/2)}(m-1)^{(m-1)/2}[k(n-1)]^{(n-1)/2}\cdot\left[\frac{D_1}{n}+\frac{kd^2D_2}{m}\right]$$

where

$$\text{(5.7)}\qquad D_1=\int_0^\infty\left(\frac{c+x}{c+d+x}\right)^2 x^{(m-3)/2}[k(n-1)+(m-1)x]^{-(n+m-2)/2}dx$$

and

$$\text{(5.8)}\qquad D_2=\int_0^\infty\frac{x^{(m-3)/2}}{(c+d+x)^2}[k(n-1)+(m-1)x]^{-(n+m-2)/2}dx\ .$$

Using the above formulas we can now compute the premium and protection of our estimator C_3 as functions of k. We have chosen $c=1.32$ and $d=1.326$ so that the premium is 0.05, and the protection is nearly optimum. Sample sizes n and m are respectively 8 and 6 as in the case of T_3. For this choice of the constants, Protection (C_3) is given in Table 7 for $k=2(1)10$.

The conclusions to be drawn from the table are obvious. The protection afforded by our estimator C_3 against always pooling is uniformly higher than that of T_3.

6. Concluding remarks

In all the cases considered in this paper we have shown that the estimators (C_1, C_2 and C_3), constructed by us by properly modifying the continuous weight function approach, are much better, when a is not too small, than the ones (T_1, T_2 and T_3) obtained via the test estimation approach in the sense of buying higher protections with respect to the always pool estimator A.

The extension of these ideas to the estimation of the variance is considered in a forthcoming publication.

Acknowledgements

The authors are grateful to the referee for many valuable suggestions which helped improve an earlier version of this paper. All the numerical work involved in this paper was performed at the Temple University Computer Center.

TEMPLE UNIVERSITY, PHILADELPHIA

REFERENCES

[1] Anscombe, F. J. (1960). Rejection of outliers, *Technometrics*, **2**, 123-166.

[2] Bancroft, T. A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance, *Biometrics*, **20**, 427-442.

[3] Bancroft, T. A. (1965). Inferences for incompletely specified models in physical sciences, *Bulletin of the International Statistical Institute: Proceedings of the 35th Session*, Belgrade.

[4] Berkson, J. (1942). Tests of significance considered as evidence, *J. Amer. Statist. Assn.*, **37**, 325-335.

[5] Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Tables of Integrals, Series, and Products*, Academic Press, New York.

[6] Han, C. P. and Bancroft, T. A. (1968). On pooling means when variance is unknown, *J. Amer. Statist. Assn.*, **63**, 1333-1342.

[7] Huntsberger, D. V. (1955). A generalization of a preliminary testing procedure for pooling data, *Ann. Math. Statist.*, **26**, 734-743.

[8] Kale, B. K. and Bancroft, T. A. (1967). Inference for incompletely specified models involving normal approximations to discrete data, *Biometrics*, **23**, 335-348.

[9] Mehta, J. S. (1968). *Preliminary Testing and Continuous Weight Functions as an Approach to Problems in Statistical Inference*, Ph.D. Thesis, University of Wisconsin.

[10] Mehta, J. S. and Gurland, J. (1969). Combinations of unbiased estimators which consider inequality of unknown variance, *J. Amer. Statist. Assn.*, **64**, 1042-1055.