Computational Statistics and Data Analysis 68 (2013) 352–360

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

One stage multiple comparisons with the average for exponential location parameters under heteroscedasticity

Shu-Fei Wu *

Department of Statistics, Tamkang University, Tamsui, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 3 September 2012
Received in revised form 10 June 2013
Accepted 3 July 2013
Available online 13 July 2013

Keywords:
One-stage procedure
Two-stage procedure

ABSTRACT

Two-stage multiple comparisons with the average for location parameters of two-parameter exponential distributions under heteroscedasticity are proposed by Wu and Wu [Wu, S.F., Wu, C.C., 2005. Two stage multiple comparisons with the average for exponential location parameters under heteroscedasticity. Journal of Statistical Planning and Inference 134, 392–408]. When the additional sample for the second stage may not be available, one-stage procedures including one-sided and two-sided confidence intervals are proposed in this paper. These intervals can be used to identify a subset which includes all no-worse-than-the-average treatments in an experimental design and to identify better-than-the-average, worse-than-the-average and not-much-different-from-the-average products in agriculture, the stock market, pharmaceutical industries. Tables of upper limits of critical values are obtained using the technique given in Lam [Lam, K., 1987. Subset selection of normal populations under heteroscedasticity. In: Proceedings of the Second International Advanced Seminar/Workshop on Inference Procedures Associated with Statistical Ranking and Selection. Sydney, Australia. August 1987. Lam, K., 1988. An improved two-stage selection procedure. Communications in Statistics—Simulation and Computation 17 (3), 995–1006]. An example of comparing four drugs in the treatment of leukemia is given to demonstrate the proposed procedures. The relationship between the one-stage and the two-stage procedures is also elaborated in this paper.

1. Introduction

Bechhofer (1954) and Gupta (1956) are two pioneers in the field of ranking and selection for normal distributions. For exponential distributions, Wu and Chen (1998) proposed single-stage multiple comparison procedures with the average within its own groups by subset selection and simultaneous confidence interval approaches are also discussed under the assumption of common unknown scale parameters when sample sizes are equal. When sample sizes are unequal, Wu and Chen (1997) have some more generalized results. When scale parameters are unknown and unequal, two-stage procedures for comparing several exponential location parameters with their average under heteroscedasticity using the techniques given in Lam (1987, 1988) are proposed by Wu and Wu (2005). There are many applications of exponential distribution in the analysis of reliability and life test experiments. See for example, Johnson et al. (1994), Bain and Engelhardt (1991), Lawless and Singhal (1980) and Zelen (1966). The location parameters of two-parameter exponential distributions are so-called threshold values or "guarantee time" parameters. The two-stage procedure in Wu and Wu (2005) is briefly introduced as follows: consider k (≥ 2) independent exponential populations $\pi_1, \ldots, \pi_k$ and observations from population π_i follow an exponential distribution denoted by $E(\theta_i, \sigma_i)$, $i = 1, \ldots, k$, where $\theta_1, \ldots, \theta_k$ are unknown location parameters and

* Tel.: +886 2 26215656x2876.
E-mail address: 100665@mail.tku.edu.tw.

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2013.07.005

$\sigma_1, \ldots, \sigma_k$ are unknown and possibly unequal scale parameters. Take an initial sample $X_{i1}, \ldots, X_{in_0}$ of size n_0 (≥ 2) from π_i, and let $\tilde{Y}_i = \min(X_{i1}, \ldots, X_{in_0})$ and $S_i = \sum_{j=1}^{n_0}(X_{ij} - \tilde{Y}_i)/(n_0 - 1)$, $i = 1, \ldots, k$. Given a fixed constant $c > 0$ to be chosen to control the width of the confidence intervals for $\theta_i - \bar{\theta}$, the overall sample size N_i in population π_i for the two-stage procedure is given by

$$N_i = \max\left\{n_0, \left[\frac{S_i}{c}\right] + 1\right\}, \quad i = 1, \ldots, k, \tag{1}$$

where $[x]$ denotes the largest integer smaller than or equal to x. When $N_i > n_0$, take $N_i - n_0$ additional observations $X_{i,n_0+1}, \ldots, X_{i,N_i}$ from π_i, then we have a total of N_i observations from π_i and the sample values are denoted by $X_{i1}, \ldots, X_{in_0}, X_{i,n_0+1}, \ldots, X_{i,N_i}$. Let the minimum value of the combined sample be $\tilde{X}_i = \min(\tilde{Y}_i, X_{i,n_0+1}, \ldots, X_{i,N_i})$ when $N_i > n_0$ and $\tilde{X}_i = \tilde{Y}_i$ when $N_i = n_0$. Let $\bar{\tilde{X}} = \sum_{i=1}^{k}\tilde{X}_i/k$.

The upper bound, lower bound and the two-sided simultaneous confidence intervals for $\theta_i - \bar{\theta}$, $i = 1, \ldots, k$ are given by $(-\infty, \tilde{X}_i - \bar{\tilde{X}} + ch_U)$, $(\tilde{X}_i - \bar{\tilde{X}} - ch_L, \infty)$ and $(\tilde{X}_i - \bar{\tilde{X}} - ch_t, \tilde{X}_i - \bar{\tilde{X}} + ch_t)$, where $h_U = h_L = h_t = h_t = (k-1)/kF^{-1}_{2,2n_0-2}(P^{*1/k})$, $i = 1, \ldots, k$.

The two-stage procedures are design-oriented. However, the number of samples required at the second stage can be large due to heterogeneous variances and may make the procedure impracticable. When the additional sample at the second stage may not be available due to the experimental budget shortage or other factors in an experiment, one-stage multiple comparison procedures with the average are proposed in this paper.

The one-stage multiple comparison procedures with the average using Lam's (1987,1988) technique are proposed in Section 2. In Section 3, an example of duration of remission by four drugs used in the treatment of leukemia is used to illustrate the proposed procedures. Each group consists of 20 patients. Each of the data sets has been checked that exponential model is correct by using the G test based on Gini statistic (Gail and Gastwirth, 1978) and the likelihood ratio asymptotic χ^2 test (Lawless, 2003) had shown a significant difference among the scale parameters of four two-parameter exponential distributions. Thus, the one-stage multiple comparison procedures and the two-stage multiple comparison procedures with the average using Lam's (1987,1988) technique for exponential location parameters under heteroscedasticity can be applied and the relationship between the one-stage and the two-stage procedures is also discussed in Section 3. A simulation study of the performance of two procedures for different parameter configurations is given in Section 4. The results show that all simulated coverage rates are higher than the nominal confidence coefficients. Finally, our conclusions are summarized in Section 5.

2. One stage multiple comparisons with the average for exponential location parameters under heteroscedasticity using Lam's (1987,1988) technique

When the additional sample for the second stage may not be available due to the experimental budget shortage or other factors in an experiment, the two-stage multiple comparison procedures with the average proposed by Wu and Wu (2005) cannot be used when k scale parameters are unknown and possibly unequal. Therefore, we proposed one-stage multiple comparison procedures with the average as follows:

Take a one-stage sample $X_{i1}, \ldots, X_{im}$ of size m (≥ 2) from π_i. Let $Y_i = \min(X_{i1}, \ldots, X_{im})$ and $S_i = \sum_{j=1}^{m}(X_{ij} - Y_i)/(m-1)$, and let

$$c^* = \max_{i=1,\ldots,k}\frac{S_i}{m}. \tag{2}$$

We now propose the one-sided and two-sided confidence intervals for $\theta_i - \bar{\theta}$, $i = 1, \ldots, k$, in the following theorem using the one-stage procedure, where $\bar{\theta} = \sum_{i=1}^{k}\theta_i/k$. Let $F^{-1}_{2,2m-2}(P)$ be the 100Pth percentile of F distribution with $(2, 2m-2)df$.

Theorem 1. *For a given $0 < P^* < 1$ and letting $\bar{Y} = \sum_{i=1}^{k}Y_i/k$, we have*

(a) $P(\theta_i - \bar{\theta} \leq Y_i - \bar{Y} + c^*s_U, i = 1, \ldots, k) \geq P^*$ *if* $s_U = (k-1)/kF^{-1}_{2,2m-2}(P^{*1/k})$.

Thus, $(-\infty, Y_i - \bar{Y} + c^*s_U)$ *is a set of upper confidence intervals for* $\theta_i - \bar{\theta}$ *with confidence coefficient* P^*, $i = 1, \ldots, k$.

(b) $P(\theta_i - \bar{\theta} \geq Y_i - \bar{Y} - c^*s_L, i = 1, \ldots, k) \geq P^*$ *if* $s_L = (k-1)/kF^{-1}_{2,2m-2}(P^{*1/k})$.

Thus, $(Y_i - \bar{Y} - c^*s_L, \infty)$ *is a set of lower confidence intervals for* $\theta_i - \bar{\theta}$ *with confidence coefficient* P^*, $i = 1, \ldots, k$.

(c) $P(Y_i - \bar{Y} - c^*s_t \leq \theta_i - \bar{\theta} \leq Y_i - \bar{Y} + c^*s_t, i = 1, \ldots, k) \geq P^*$ *if* $s_t = (k-1)/kF^{-1}_{2,2m-2}(P^{*1/k})$.

Thus, $(Y_i - \bar{Y} \pm c^*s_t)$ *is a set of simultaneous two-sided confidence intervals for* $\theta_i - \bar{\theta}$ *with confidence coefficient* P^*, $i = 1, \ldots, k$.

The techniques given in Lam (1987, 1988) are described in the following lemma:

Lemma 2. *Suppose X and Y are two random variables, a and b are two positive constants, then $[aX \geq bY - d\max(a,b)] \supseteq [X \geq -d, Y \leq d \text{ and } X \geq Y - d]$.*

To prove Theorem 1, we will need the following distributional results (from Roussas (1997)):

(D1) $2(m-1)S_i/\sigma_i, i = 1, \ldots, k$, follows a chi-square distribution with $2m - 2df$.
(D2) $m(Y_i - \theta_i)/\sigma_i = W_i, i = 1, \ldots, k$, is obtained as the standard exponential distribution.
(D3) S_i/σ_i and $m(Y_i - \theta_i)/\sigma_i, i = 1, \ldots, k$, are stochastically independent.
(D4) $m(Y_i - \theta_i)/S_i = W_i^*, i = 1, \ldots, k$, is distributed as an F distribution with $(2, 2m-2)df$.

Proof of Theorem 1. For (a), we have

$$\begin{aligned}
P(\theta_i - \overline{\theta} \leq Y_i - \bar{Y} + c^* s_U, i = 1, \ldots, k)^* &= P\left(Y_i - \theta_i - \sum_{l=1}^{k} \frac{Y_l - \theta_l}{k} \geq -c^* s_U, i = 1, \ldots, k\right) \\
&= P\left(\frac{S_i}{m} W_i^* - \sum_{l=1}^{k} \frac{S_l W_l^*}{mk} \geq -c^* s_U, i = 1, \ldots, k\right) \\
&= P\left(\frac{k-1}{k}\frac{S_i}{m} W_i^* \geq \sum_{l \neq i}^{k} \frac{S_l W_l^*}{mk} - c^* s_U, i = 1, \ldots, k\right) \\
&= P\left(\frac{S_i}{m} W_i^* \geq \sum_{l \neq i}^{k} \frac{S_l W_l^*}{m(k-1)} - \frac{k}{k-1} c^* s_U, i = 1, \ldots, k\right) \\
&\geq P\left(\frac{S_i}{m} W_i^* \geq \max_{l \neq i} W_l^* \sum_{l \neq i}^{k} \frac{S_l}{m(k-1)} \right. \\
&\qquad \left. - \; d \max\left(\frac{S_i}{m}, \sum_{l \neq i}^{k} \frac{S_l}{m(k-1)}\right), i = 1, \ldots, k\right),
\end{aligned}$$

where $d = k/(k-1)s_U$.

The above inequality holds because $c^* \geq S_i/m$ for all i and hence we also have $c^* \geq \sum_{l \neq i}^{k} \frac{S_l}{m(k-1)}$.

$$= E_{S_i, \ldots, S_k} P\left(\frac{S_i}{m} W_i^* \geq \max_{l \neq i} W_l^* \sum_{l \neq i}^{k} \frac{S_l}{m(k-1)} - d \max\left(\frac{S_i}{m}, \sum_{l \neq i}^{k} \frac{S_l}{m(k-1)}\right), i = 1, \ldots, k\right),$$

letting $a = S_i/m$, $b = \sum_{l \neq i}^{k} \frac{S_l}{m(k-1)}$ and applying Lemma 2, we have

$$\begin{aligned}
&\geq E_{S_1, \ldots, S_k} P\left(W_i^* \geq -d, \max_{l \neq i} W_l^* \leq d \text{ and } W_i^* \geq \max_{l \neq i} W_i^* - d, i = 1, \ldots, k\right) \\
&= P\left(\max_{l \neq i} W_l^* \leq d, i = 1, \ldots, k\right) = P(F_{2,2m-2} \leq d)^k = P^*.
\end{aligned}$$

So we have $s_U = (k-1)/kF_{2,2m-2}^{-1}(P^{*1/k})$. The proof is thus obtained.

For (b), we have

$$\begin{aligned}
P(\theta_i - \overline{\theta} \geq Y_i - \bar{Y} - c^* s_L, i = 1, \ldots, k) &= P\left(Y_i - \theta_i - \sum_{l=1}^{k} \frac{Y_l - \theta_l}{k} \leq c^* s_L, i = i, \ldots, k\right) \\
&= P\left(\frac{S_i}{m} W_i^* - \sum_{l=1}^{k} \frac{S_l W_l^*}{mk} \leq c^* s_L, i = 1, \ldots, k\right) \\
&= P\left(\frac{k-1}{k}\frac{S_i}{m} W_i^* - c^* s_L \leq \sum_{l \neq i}^{k} \frac{S_l W_l^*}{mk}, i = 1, \ldots, k\right) \\
&= P\left(\frac{S_i}{m} W_i^* - \frac{k}{k-1} c^* s_L \leq \sum_{l \neq i}^{k} \frac{S_l W_l^*}{m(k-1)}, i = 1, \ldots, k\right) \\
&\geq P\left(\frac{S_i}{m} W_i^* - \frac{k}{k-1} c^* s_L \leq \min_{l \neq i} W_l^* \sum_{l \neq i}^{k} \frac{S_l}{m(k-1)}, i = 1, \ldots, k\right)
\end{aligned}$$

$$\geq P\left(\min_{l\neq i} W_l^* \sum_{l\neq i}^{k} \frac{S_l}{m(k-1)} \geq \frac{S_i}{N_i} W_i^* - d\max\left(\sum_{l\neq i}^{k} \frac{S_l}{m(k-1)}, \frac{S_i}{m}\right), i=1,\ldots,k\right)$$

$$= E_{S_1,\ldots,S_k} P\left(\min_{l\neq i} W_l^* \sum_{l\neq i}^{k} \frac{S_l}{m(k-1)} \geq \frac{S_i}{N_i} W_i^* - d\max\left(\sum_{l\neq i}^{k} \frac{S_l}{m(k-1)}, \frac{S_i}{m}\right), i=1,\ldots,k\right),$$

letting $a = \sum_{l\neq i}^{k} \frac{S_l}{N_l(k-1)}$, $b = \frac{S_i}{N_i}$, $X = \min_{l\neq i} W_i^*$ and $Y = W_i^*$ and applying Lemma 2, we have

$$\geq E_{S_1,\ldots,S_k} P\left(\min_{l\neq i} W_l^* \geq -d, W_i^* \leq d \text{ and } \min_{l\neq i} W_l^* \geq W_i^* - d, i=1,\ldots,k\right)$$
$$= P(W_i^* \leq d, i=1,\ldots,k) = P(F_{2,2m-2} \leq d)^k = P^*.$$

Solving the above equation, then we have $s_L = (k-1)/kF_{2,2m-2}^{-1}(P^{*1/k})$ and the proof is thus obtained.

For (c), combining (a) and (b), we have

$$P(Y_i - \bar{Y} - c^* s_t \leq \theta_i - \bar{\theta} \leq Y_i - \bar{Y} + c^* s_t, i=1,\ldots,k)$$
$$= P(Y_i - \bar{Y} - c^* s_t \leq \theta_i - \bar{\theta} \cap \theta_i - \bar{\theta} \leq Y_i - \bar{Y} + c^* s_t, i=1,\ldots,k)$$
$$\geq P\left(\min_{l\neq i} W_l^* \geq -d, W_i^* \leq d \text{ and } \min_{l\neq i} W_l^* \geq W_i^* - d \cap W_l^* \geq -d, \max_{l\neq i} W_i^* \leq d \text{ and } W_i^* \geq \max_{l\neq i} W_l^* - d, i=1,\ldots,k\right)$$
$$= P\left(W_i^* \leq d \cap \max_{l\neq i} W_l^* \leq d, i=1,\ldots,k\right)$$
$$= P(F_{2,2m-2} \leq d)^k = P^*.$$

Solving the above equation, then we have $s_t = (k-1)/kF_{2,2m-2}^{-1}(P^{*1/k})$ and the proof is thus obtained. □

The approximate critical values $s_U = s_L = s_t$ are listed in Table 1 for $k = 3(1)10(2)18$, $m = 2(1)10(5)40, \infty$ and $P^* = 0.90, 0.95, 0.99$ and 0.995. From Table 1, it can be seen that the approximate critical values $s_U = s_L = s_t$ are decreasing while m is increasing for any given k and P^* or while P^* is decreasing for any given k and m. Let L_1 be the length of the two-sided confidence intervals for $\theta_i - \bar{\theta}$ by the one-stage procedure, then we have $L_1 = 2c^* s_t$. From this equation, we can see that the larger the m, the smaller the value of s_t and the smaller the confidence length of L_1 for any given k and P^*. We can also see that the larger the P^*, the larger the value of s_t and the larger the confidence length of L_1 for any given k and m. Furthermore, we can also see that the larger the k, the larger the value of s_t and the larger the confidence length of L_1 for any given m and P^*. Therefore, the length of simultaneous confidence interval (SCI) is getting wider when we compare more populations for any given m and P^*. The approximate critical values of $h_U = h_L = h_t = s_U = s_L = s_t$ for two-stage procedures under $k = 3(1)10(2)18$ and $P^* = 0.90, 0.95, 0.99, 0.995$ with $n_0 = m = 2(1)10(5)40$ can also be found in Table 1.

Remark. (1) When the unequal scale parameters are known, the unbiased estimator S_i of σ_i is replaced by σ_i throughout Theorem 1 and the statistic W_i^* which is distributed as a F distribution with $(2, 2m-2)df$ is replaced by the statistic W_i which is distributed as a standard exponential distribution from (D3). Therefore, the approximate critical values are $h_U = h_L = h_t = (k-1)/k(-\ln(1-P^{*1/k}))$ when the scale parameters are known.

(2) The proposed one-stage multiple comparison procedure with the control using Lam's (1987, 1988) technique is expected to perform better than using the Bonferroni inequality.

(3) In practice, we may not have the common sample size m. For unequal sample size m_i for the ith population, the procedures in Theorem 1 can be applied by taking $c^* = \max_{i=1,\ldots,k} \frac{S_i}{m_i}$ and replacing m by m_i. The critical values become $s_U = s_L = s_t = (k-1)/kd$, where d is the solution of $\prod_{i=1}^{k} P(F_{2,2m_i-2} \leq d) = P^*$.

Table 1
Approximate critical values of $s_U = s_L = s_t$.

P^*	m	$s_U = s_L = s_t$											
		k											
		3	4	5	6	7	8	9	10	12	14	16	18
0.90	2	18.651	28.100	37.566	47.041	56.520	66.002	75.486	84.972	103.946	122.922	141.900	160.879
	3	5.844	7.803	9.480	10.966	12.311	13.549	14.701	15.782	17.775	19.591	21.269	22.836
	4	4.143	5.345	6.320	7.146	7.869	8.515	9.100	9.638	10.600	11.447	12.209	12.904
	5	3.520	4.471	5.221	5.844	6.378	6.849	7.270	7.651	8.325	8.908	9.425	9.890
	6	3.202	4.031	4.674	5.201	5.649	6.039	6.386	6.698	7.243	7.711	8.122	8.488
	7	3.010	3.768	4.349	4.822	5.220	5.565	5.870	6.143	6.618	7.022	7.375	7.688
	8	2.882	3.593	4.134	4.572	4.939	5.255	5.533	5.782	6.212	6.577	6.893	7.174
	9	2.790	3.469	3.982	4.395	4.740	5.037	5.297	5.529	5.928	6.266	6.558	6.816
	10	2.722	3.376	3.869	4.264	4.593	4.875	5.122	5.341	5.719	6.037	6.311	6.553
	15	2.537	3.127	3.567	3.915	4.203	4.448	4.661	4.849	5.171	5.439	5.669	5.870
	20	2.455	3.018	3.434	3.763	4.033	4.262	4.461	4.636	4.935	5.182	5.394	5.579
	25	2.409	2.956	3.360	3.677	3.938	4.159	4.350	4.518	4.803	5.040	5.242	5.418
	30	2.380	2.917	3.312	3.623	3.878	4.093	4.279	4.442	4.720	4.950	5.145	5.315
	35	2.359	2.890	3.279	3.585	3.835	4.047	4.230	4.390	4.662	4.887	5.078	5.245
	40	2.344	2.870	3.255	3.557	3.805	4.013	4.193	4.352	4.620	4.841	5.029	5.193
	∞	2.244	2.737	3.096	3.376	3.603	3.794	3.959	4.102	4.345	4.544	4.712	4.858
0.95	2	38.659	58.113	77.584	97.063	116.546	136.033	155.522	175.012	213.995	252.980	291.967	330.955
	3	8.907	11.789	14.238	16.398	18.349	20.140	21.805	23.365	26.238	28.853	31.267	33.521
	4	5.785	7.383	8.664	9.744	10.684	11.520	12.276	12.969	14.207	15.295	16.272	17.161
	5	4.724	5.929	6.868	7.641	8.301	8.879	9.394	9.861	10.681	11.390	12.016	12.579
	6	4.201	5.224	6.007	6.642	7.179	7.644	8.055	8.425	9.069	9.618	10.100	10.529
	7	3.892	4.811	5.506	6.065	6.534	6.937	7.292	7.609	8.157	8.623	9.028	9.387
	8	3.689	4.541	5.180	5.691	6.117	6.482	6.801	7.085	7.576	7.989	8.347	8.663
	9	3.545	4.351	4.952	5.430	5.826	6.164	6.460	6.722	7.173	7.551	7.878	8.166
	10	3.439	4.211	4.783	5.237	5.612	5.931	6.210	6.456	6.878	7.232	7.536	7.804
	15	3.155	3.839	4.340	4.732	5.053	5.324	5.559	5.766	6.118	6.409	6.658	6.876
	20	3.032	3.678	4.148	4.514	4.813	5.065	5.282	5.472	5.795	6.061	6.288	6.485
	25	2.963	3.588	4.042	4.394	4.680	4.921	5.128	5.310	5.617	5.870	6.085	6.271
	30	2.919	3.531	3.974	4.317	4.596	4.830	5.031	5.207	5.504	5.748	5.956	6.136
	35	2.888	3.491	3.927	4.264	4.537	4.767	4.964	5.136	5.426	5.665	5.867	6.042
	40	2.865	3.462	3.892	4.225	4.494	4.720	4.914	5.084	5.369	5.604	5.802	5.974
	∞	2.718	3.272	3.668	3.972	4.217	4.421	4.596	4.748	5.002	5.210	5.385	5.536
0.99	2	198.67	298.12	397.60	497.08	596.57	696.06	795.55	895.04	1094.0	1293.0	1492.0	1691.0
	3	21.722	28.444	34.105	39.073	43.544	47.639	51.437	54.993	61.529	67.469	72.949	78.060
	4	11.374	14.307	16.623	18.556	20.228	21.708	23.041	24.259	26.428	28.329	30.031	31.578
	5	8.422	10.404	11.917	13.147	14.188	15.094	15.897	16.621	17.889	18.979	19.939	20.800
	6	7.090	8.670	9.852	10.798	11.588	12.267	12.865	13.399	14.323	15.109	15.794	16.402
	7	6.344	7.707	8.714	9.511	10.171	10.734	11.226	11.663	12.415	13.049	13.597	14.082
	8	5.869	7.099	7.999	8.706	9.287	9.781	10.210	10.590	11.240	11.784	12.253	12.665
	9	5.542	6.682	7.510	8.157	8.686	9.134	9.522	9.864	10.447	10.933	11.350	11.716
	10	5.304	6.379	7.156	7.760	8.252	8.668	9.026	9.342	9.879	10.324	10.705	11.038
	15	4.691	5.604	6.253	6.752	7.154	7.491	7.779	8.031	8.455	8.803	9.099	9.355
	20	4.432	5.279	5.877	6.333	6.700	7.005	7.265	7.492	7.872	8.183	8.445	8.672
	25	4.290	5.101	5.671	6.104	6.451	6.740	6.985	7.199	7.556	7.846	8.092	8.303
	30	4.200	4.988	5.541	5.960	6.295	6.573	6.810	7.015	7.357	7.636	7.870	8.072
	35	4.137	4.910	5.451	5.861	6.188	6.459	6.689	6.889	7.221	7.492	7.719	7.915
	40	4.092	4.854	5.386	5.789	6.110	6.376	6.601	6.797	7.122	7.387	7.609	7.800
	∞	3.800	4.491	4.968	5.327	5.612	5.845	6.043	6.213	6.495	6.722	6.912	7.075
0.995	2	398.67	598.123	797.598	997.081	1196.57	1396.06	1595.55	1795.05	2194.04	2593.03	2992.03	3391.02
	3	31.299	40.887	48.946	56.008	62.360	68.173	73.563	78.608	87.878	96.299	104.067	111.310
	4	14.859	18.624	21.584	24.048	26.174	28.055	29.748	31.292	34.041	36.447	38.600	40.557
	5	10.526	12.947	14.786	16.275	17.532	18.624	19.591	20.461	21.983	23.289	24.438	25.467
	6	8.644	10.522	11.918	13.030	13.956	14.750	15.448	16.069	17.145	18.057	18.850	19.555
	7	7.614	9.206	10.374	11.294	12.052	12.698	13.261	13.760	14.617	15.337	15.960	16.509
	8	6.969	8.389	9.419	10.224	10.884	11.442	11.927	12.354	13.084	13.694	14.218	14.679
	9	6.529	7.834	8.773	9.503	10.098	10.599	11.033	11.414	12.062	12.601	13.063	13.467
	10	6.211	7.433	8.308	8.985	9.535	9.996	10.394	10.743	11.335	11.825	12.243	12.608
	15	5.404	6.424	7.142	7.690	8.130	8.496	8.809	9.081	9.539	9.914	10.231	10.505
	20	5.069	6.007	6.662	7.159	7.556	7.885	8.165	8.408	8.815	9.145	9.424	9.664
	25	4.886	5.779	6.401	6.871	7.246	7.555	7.818	8.045	8.424	8.732	8.991	9.214
	30	4.770	5.637	6.238	6.691	7.051	7.349	7.600	7.818	8.181	8.475	8.721	8.933
	35	4.691	5.539	6.126	6.567	6.918	7.207	7.452	7.663	8.015	8.299	8.537	8.742
	40	4.633	5.467	6.044	6.477	6.821	7.104	7.344	7.550	7.894	8.171	8.404	8.603
	∞	4.264	5.012	5.525	5.907	6.207	6.454	6.661	6.839	7.133	7.368	7.564	7.732

Table 2
The required statistics and critical values.

Statistics	Drug1	Drug2	Drug3	Drug4
Y_i	1.013	2.214	3.071	4.498
S_i	1.238	1.530	3.233	4.075
c^*	0.204			
$Y_i - \bar{Y}$	−1.686	−0.485	0.372	1.799
P^*	0.90	0.95	0.995	
$s_U = s_L = s_t$	3.018	3.678	6.007	

Table 3
The 90%, 95% and 99.5% one-stage one-sided confidence intervals for all drugs compared with the average.

Parameter	$(-\infty, Y_i - \bar{Y} + c^* s_U)$, $(Y_i - \bar{Y} - c^* s_L, \infty)$		
	90%	95%	99.5%
1. $\theta_1 - \bar{\theta}$	$(-\infty, -1.070)$, $(-2.301, \infty)$	$(-\infty, -0.937)$, $(-2.435, \infty)$	$(-\infty, -0.461)$, $(-2.911, \infty)$
2. $\theta_2 - \bar{\theta}$	$(-\infty, 0.131)$, $(-1.100, \infty)$	$(-\infty, 0.264)$, $(-1.234, \infty)$	$(-\infty, 0.740)$, $(-1.710, \infty)$
3. $\theta_3 - \bar{\theta}$	$(-\infty, 0.988)$, $(-0.243, \infty)$	$(-\infty, 1.121)$, $(-0.377, \infty)$	$(-\infty, 1.597)$, $(-0.853, \infty)$
4. $\theta_4 - \bar{\theta}$	$(-\infty, 2.415)$, $(1.184, \infty)$	$(-\infty, 2.548)$, $(1.050, \infty)$	$(-\infty, 3.024)$, $(0.574, \infty)$

Table 4
The 90%, 95% and 99.5% one-stage two-sided confidence intervals for all drugs compared with the average.

Parameter	$(Y_i - \bar{Y} + c^* s_t, Y_i - \bar{Y} + c^* s_t)$		
	90%	95%	99.5%
1. $\theta_1 - \bar{\theta}$	(−2.301, −1.070)	(−2.435, −0.937)	(−2.911, −0.461)
2. $\theta_2 - \bar{\theta}$	(−1.100, 0.131)	(−1.234, 0.264)	(−1.710, 0.740)
3. $\theta_3 - \bar{\theta}$	(−0.243, 0.988)	(−0.377, 1.121)	(−0.853, 1.597)
3. $\theta_4 - \bar{\theta}$	(1.184, 2.415)	(1.050, 2.548)	(0.574, 3.024)

3. Example

Referring to Table 1 of Wu and Wu (2005), the duration of remission achieved by four drugs used in the treatment of leukemia is used to illustrate our proposed one-stage multiple comparison procedures with the control. Four groups of 20 patients each were used and the data of duration of remission by four drugs is given in Table 2. It is shown that each of the four groups of data is exponentially distributed by using the G test based on the Gini statistic (Gail and Gastwirth, 1978) in Wu and Wu (2005). It is also shown that the scale parameters of four two-parameter exponential distributions are significantly different by the likelihood ratio asymptotic χ^2 test (Lawless, 2003). The data-analysis one-stage multiple comparison procedures with the average proposed in Theorem 2 for exponential location parameters under heteroscedasticity can be applied. The required statistics and critical values of $s_U = s_L = s_t$ for $P^* = 0.90$, 0.95 and 0.995 are summarized in Table 2.

Using parts (a) and (b) of Theorem 1, we can obtain the one-stage one-sided confidence bounds with confidence coefficients 0.90, 0.95 and 0.995 given in Table 3. Since the upper confidence bound for $\theta_1 - \bar{\theta}$ is less than zero and the lower confidence bound for $\theta_4 - \bar{\theta}$ is greater than zero, we can conclude that drug 1 is selected in a worse than the average subset and drug 4 is selected in a better than the average subset with the probability of correct selection being at least 0.90, 0.95 and 0.995.

Using parts (c) of Theorem 1, we can obtain the one-stage two-sided confidence bounds with confidence coefficients 0.90, 0.95 and 0.995 given in Table 4. For confidence coefficients 0.90, 0.95 and 0.995, we can conclude that drug 1 is worse than the average; drug 4 is better than the average and drugs 2, 3 are not significantly different from the average.

Let L_1 and L_2 be the lengths of the two-sided confidence intervals for $\theta_i - \bar{\theta}$, $i = 1, \ldots, 4$, by using the one-stage and the two-stage methods, respectively. From Table 4, we can obtain the lengths of one-stage procedures as $L_1 = 1.230$, 1.499 and 2.45 for confidence coefficients 0.90, 0.95 and 0.995, respectively. For the two-stage procedure, take the initial sample size as $n_0 = 20$. From Table 3 of Wu and Wu (2005), the confidence lengths with two-stage procedures are $L_2 = 0.821$, 1.000 and 1.436 for confidence coefficients 0.90, 0.95 and 0.995 which are smaller than L_1 since the required overall sample size $(N_1, N_2, N_3, N_4) = (20, 20, 24, 30)$ for two-stage procedures is greater than $m = 20$ for one-stage procedures. But one-stage procedure is a good remedy for two-stage procedure when the additional sample of second stage is not available due to the experimental budget shortage or other factors in an experiment.

Under the same confidence coefficient P^*, the comparisons between the one-stage procedure and the two-stage procedure are elaborated as follows.

Case 1. If $c^* = c$, then we have $L_1 = L_2$. The one-stage procedure and the two-stage procedure have the same overall sample size, except for a rounding error in sample size by definition (2).

Table 5
The coverage rates of lower, upper and two-sided confidence intervals under structure of scale parameters $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1.0, 1.0, 1.0, 1.0)$.

n_0	P^*	0.9							0.95						
	L_2	Lower		Upper		Two		Ratio	Lower		Upper		Two		Ratio
		One-stage	Two-stage	One-stage	Two-stage	One-stage	Two-stage		One-stage	Two-stage	One-stage	Two-stage	One-stage	Two-stage	
15	1.0	0.989	1.000	1.000	1.000	0.989	1.000	1.000	0.996	1.000	1.000	1.000	0.996	1.000	1.000
	0.8	0.988	0.999	1.000	1.000	0.988	0.999	1.000	0.996	0.999	1.000	1.000	0.996	0.999	1.004
	0.5	0.988	0.990	1.000	1.000	0.988	0.990	1.041	0.995	0.992	1.000	1.000	0.995	0.992	1.137
	0.3	0.985	0.962	1.000	0.999	0.984	0.962	1.438	0.993	0.978	1.000	1.000	0.993	0.978	1.744
	0.1	0.975	0.950	1.000	0.998	0.975	0.950	4.202	0.988	0.975	1.000	0.999	0.988	0.974	5.153
25	1.0	0.983	1.000	1.000	1.000	0.983	1.000	1.000	0.994	1.000	1.000	1.000	0.994	1.000	1.000
	0.8	0.983	1.000	1.000	1.000	0.983	1.000	1.000	0.993	1.000	1.000	1.000	0.993	1.000	1.000
	0.5	0.983	1.000	1.000	1.000	0.983	1.000	1.000	0.993	1.000	1.000	1.000	0.993	1.000	1.000
	0.3	0.983	0.989	1.000	1.000	0.983	0.989	1.012	0.993	0.991	1.000	1.000	0.993	0.991	1.067
	0.1	0.975	0.953	1.000	0.998	0.975	0.953	2.385	0.988	0.976	1.000	1.000	0.988	0.976	2.891
35	1.0	0.980	1.000	1.000	1.000	0.980	1.000	1.000	0.992	1.000	1.000	1.000	0.992	1.000	1.000
	0.8	0.980	1.000	1.000	1.000	0.980	1.000	1.000	0.992	1.000	1.000	1.000	0.992	1.000	1.000
	0.5	0.980	1.000	1.000	1.000	0.980	1.000	1.000	0.992	1.000	1.000	1.000	0.992	1.000	1.000
	0.3	0.980	0.998	1.000	1.000	0.980	0.998	1.000	0.992	0.998	1.000	1.000	0.992	0.998	1.000
	0.1	0.975	0.954	1.000	0.998	0.975	0.954	1.666	0.988	0.977	1.000	1.000	0.988	0.977	2.010

Table 6
The coverage rates of lower, upper and two-sided confidence intervals under structure of scale parameters $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1.0, 2.0, 3.0, 4.0)$.

n_0	P^*	0.9							0.95						
	L_2	Lower		Upper		Two		Ratio	Lower		Upper		Two		Ratio
		One-stage	Two-stage	One-stage	Two-stage	One-stage	Two-stage		One-stage	Two-stage	One-stage	Two-stage	One-stage	Two-stage	
15	1.0	0.982	0.974	1.000	1.000	0.982	0.974	1.265	0.992	0.986	1.000	1.000	0.992	0.986	1.449
	0.8	0.981	0.968	1.000	0.999	0.981	0.968	1.468	0.991	0.983	1.000	1.000	0.991	0.983	1.722
	0.5	0.980	0.961	1.000	0.999	0.980	0.961	2.162	0.991	0.980	1.000	1.000	0.991	0.980	2.613
	0.3	0.979	0.953	1.000	0.998	0.979	0.953	3.512	0.991	0.976	1.000	1.000	0.991	0.976	4.300
	0.1	0.979	0.949	1.000	0.998	0.979	0.949	10.46	0.990	0.974	1.000	0.999	0.990	0.974	12.83
25	1.0	0.982	0.990	1.000	1.000	0.982	0.990	1.016	0.992	0.993	1.000	1.000	0.992	0.993	1.056
	0.8	0.981	0.982	1.000	1.000	0.981	0.982	1.066	0.991	0.989	1.000	1.000	0.991	0.989	1.151
	0.5	0.980	0.968	1.000	0.999	0.980	0.968	1.354	0.991	0.983	1.000	1.000	0.991	0.983	1.565
	0.3	0.979	0.961	1.000	0.999	0.979	0.961	2.042	0.990	0.980	1.000	1.000	0.990	0.980	2.436
	0.1	0.979	0.952	1.000	0.998	0.979	0.952	5.932	0.990	0.976	1.000	1.000	0.990	0.976	7.196
35	1.0	0.981	0.998	1.000	1.000	0.981	0.998	1.000	0.991	0.998	1.000	1.000	0.991	0.998	1.002
	0.8	0.981	0.993	1.000	1.000	0.981	0.993	1.003	0.991	0.994	1.000	1.000	0.991	0.994	1.019
	0.5	0.980	0.975	1.000	1.000	0.980	0.975	1.102	0.991	0.986	1.000	1.000	0.991	0.986	1.210
	0.3	0.979	0.964	1.000	0.999	0.979	0.964	1.506	0.990	0.982	1.000	1.000	0.990	0.982	1.758
	0.1	0.979	0.953	1.000	0.998	0.979	0.953	4.142	0.990	0.977	1.000	1.000	0.990	0.977	5.001

Case 2. If $c^* < c$, then we have $L_1 < L_2$. Then there is no need to draw the second stage sample for the two-stage procedure. Under the same total sample size $m = n_0$, the one-stage procedure has shorter confidence length than the two-stage procedure. Hence, the one-stage procedure is recommended.

Case 3. If $\min_{i=1,\dots,k} \frac{S_i}{m} > c$, then we have $L_1 > L_2$. The overall sample size of the one-stage procedure is smaller than that of the two-stage procedure.

4. Simulation study

A simulation study of the proposed lower, upper and two-sided confidence intervals for $\theta_i - \bar{\theta}, i = 1, \dots, k$ using one-stage and two-stage procedures is investigated based on 500,000 simulation runs in this section. For the given confidence lengths of two stage procedures $L_2 = 1.0, 0.8, 0.5, 0.3, 0.1$ under $k = 4$ and $n_0 = 15$, we can obtain the value of $c = \frac{L_2}{2h_t} = 0.160, 0.128, 0.080, 0.048$ and 0.016 for $P^* = 0.90$ and $c = \frac{L_2}{2h_t} = 0.130, 0.104, 0.065, 0.039$ and 0.013 for $P^* = 0.95$. Then the required overall sample size for two-stage procedure can be determined by the use of Eq. (1). For the plausibility of comparison of two procedures, we take the sample size for each population in the one-stage procedure as $[\sum_{i=1}^{k} N_i/k]$, where $[x]$ stands for the greatest integer less than and equal to x. The coverage rates of the proposed lower, upper and two-sided confidence intervals for $P^* = 0.90, 0.95$, $n_0 = 15, 25, 35$ and the given two-stage confidence length $L_2 = 1.0, 0.8, 0.5, 0.3, 0.1$ are listed in Table 5–Table 6 for various structures of scale parameters $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1.0, 1.0, 1.0, 1.0), (1.0, 2.0, 3.0, 4.0)$, respectively. The sample ratios (denoted as ratio) are defined as the average of the ratios of the required total sample size for the two-stage SCI over the total initial sample size after 500,000 simulation runs and they are listed in Table 5–Table 6 followed by the coverage rates. Based on Table 5–Table 6, the findings are summarized as follows.

(1) All simulated coverage rates are higher than the nominal confidence coefficients for both procedures. It can also be seen that both procedures are quite conservative. But, one cannot deny that these two procedures are practical solutions for the problem of comparing several exponential location parameters with their average.
(2) For a two-stage procedure, the required sample ratio is larger for smaller n_0 with fixed P^* and L_2 and also for larger P^* with fixed n_0 and L_2 under various structures of scale parameters. Furthermore, the required sample ratio is larger for smaller L_2 with fixed n_0 and P^*.
(3) For a two-stage procedure, the coverage rates are higher for larger n_0 with fixed P^* and L_2 and also for larger P^* with fixed n_0 and L_2.
(4) When the sample ratio approaches 1, the overall sample size for a two-stage procedure is approximately equal to the initial sample size n_0 and also equal to the sample size for a one-stage procedure for each population. Under this condition, the one-stage procedure has coverage rates closer to the nominal confidence coefficients than the two-stage procedure. That is, the one-stage procedure is less conservative than the one-stage procedure when the sample ratio is approaching 1. When the sample ratio is greater than 1, the two-stage procedure is less conservative than the one-stage procedure. Under the structure of scale parameters $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1.0, 2.0, 3.0, 4.0)$, all sample ratios are greater than one. For this case, the two-stage procedure has coverage rates closer to the nominal confidence coefficients than the one-stage procedure for most cases. That is, the two-stage procedure is less conservative than the one-stage procedure.
(5) Comparing Table 5 with Table 6, the coverage rates under two different structures of scale parameters are approximately the same for any given n_0, L_2 and P^*. The sample ratios for unequal scale parameters are 1.0–2.489 times of those for structure of scale parameters $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1.0, 1.0, 1.0, 1.0)$ for confidence coefficient $P^* = 0.90$ and 1.002–2.490 times of those for structure of scale parameters $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1.0, 1.0, 1.0, 1.0)$ for confidence coefficient $P^* = 0.95$.

5. Conclusion

The two-stage procedure is a design-oriented procedure. The constant c is chosen to control the confidence length L_2 for $\theta_i - \bar{\theta}, i = 1, \ldots, k$ so that the overall sample size for each population can be determined by the use of Eq. (1). Under the same total sample sizes, the one-stage procedure may be better than, worse than or not much different from the two-stage procedure depending on the various structures of scale parameters and different sample ratios. However, the two-stage multiple comparison procedure with the average requires additional samples at the second stage, which can be large due to heterogeneous variances. The additional sample at the second stage may not be obtained due to time limitation and budgetary reasons in an experiment. In this situation, a one-stage multiple comparison procedure with the average proposed in Theorem 1 should be employed after the data have been collected. The data-analysis one-stage procedure provides a practicable remedy to the two-stage procedure when its required additional sample observations for the second stage are incomplete due to the early termination for some factors in an experiment. The example in Wu and Wu (2005) is given to demonstrate the practical use of one-stage procedure and the results are compared with the results under two-stage procedures. At last, a simulation study is done to investigate the relationship between two procedures for different parameter configurations under a given sample ratio. The results show that all simulated coverage rates are higher than the nominal confidence coefficients for both procedures. Both procedures are the practical solutions of the problem of comparing several exponential location parameters with their average. It is hoping that the possibility of improving one-stage and two-stage procedures can be opened eventually.

Acknowledgments

The author wishes to thank an associate editor and referees for their careful reading and valuable suggestions which made the article more readable and applicable. The author's research was supported by The National Science Council NSC101-2118-M-032-003- in Taiwan, ROC.

References

Bain, L.J., Engelhardt, M., 1991. Statistical Analysis of Reliability and Life Testing Models. Marcel Dekker, New York.
Bechhofer, R.E., 1954. A single sample multiple decision procedure for ranking means of normal populations with known variances. The Annals of Mathematical Statistics 25, 16–39.
Gail, M.H., Gastwirth, J.L., 1978. A scale-free goodness of fit test for the exponential distribution based on the Gini statistic. Journal of the Royal Statistical Society B 40, 350–357.
Gupta, S.S., 1956. On a decision rule for a problem in ranking means. Doctoral dissertation (Mimeograph Series No. 150). Institute of Statistics, University of North Carolina. Chapel Hill, North Carolina.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions. Wiley, New York.
Lam, K., 1987. Subset selection of normal populations under heteroscedasticity. In: Proceedings of the Second International Advanced Seminar/Workshop on Inference Procedures Associated with Statistical Ranking and Selection. Sydney, Australia. August.
Lam, K., 1988. An improved two-stage selection procedure. Communications in Statistics—Simulation and Computation 17 (3), 995–1006.
Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data. Wiley, New York.

Lawless, J.F., Singhal, K., 1980. Analysis of data from life test experiments under an exponential model. Naval Research Logistics Quarterly 27, 323–334.
Roussas, G.G., 1997. A Course in Mathematical Statistics. Academic Press, San Diego.
Wu, S.F., Chen, H.J., 1997. Multiple comparison procedures with the average for exponential location parameters when sample sizes are unequal. International Journal of Information and Management Sciences 8 (4), 41–54.
Wu, S.F., Chen, H.J., 1998. Multiple comparison procedures with the average for exponential location parameters. Computational Statistics and Data Analysis 26, 461–484.
Wu, S.F., Wu, C.C., 2005. Two stage multiple comparisons with the average for exponential location parameters under heteroscedasticity. Journal of Statistical Planning and Inference 134, 392–408.
Zelen, M., 1966. Application of exponential models to problems in cancer research. Journal of the Royal Statistical Society, Series A (General) 129, 368–398.