Applied Mathematics, 2025, 16(12), 877-883
https://www.scirp.org/journal/am
ISSN Online: 2152-7393
ISSN Print: 2152-7385

A Note Comparing Two Subset Selection Procedures for the Threshold Parameters of Two Exponential Populations

Gary C. McDonald, Jezerca Hodaj

Department of Mathematics and Statistics, Oakland University, Rochester, MI, USA
Email: mcdonald@oakland.edu, jhodaj@oakland.edu

How to cite this paper: McDonald, G.C. and Hodaj, J. (2025) A Note Comparing Two Subset Selection Procedures for the Threshold Parameters of Two Exponential Populations. *Applied Mathematics*, **16**, 877-883.
https://doi.org/10.4236/am.2025.1612046

Received: November 16, 2025
Accepted: December 19, 2025
Published: December 22, 2025

Abstract

This note builds on two earlier articles by the authors on the topic of subset selection procedures for threshold parameters characterizing two exponential populations. In this special case of two populations, exact computational methods are developed for computing the constants required to implement the selection rules based on the minimum order statistics and on the sample mean statistics. The operating characteristics of these selection rules are then compared for slippage configurations and several choices of sample sizes and the minimum probability of a correct selection. The computer R-code for these calculations is given in the Appendix.

Keywords

Minimum Statistic Selection Procedure, Means Selection Procedure, Probability of Correct and Incorrect Selections, Expected Subset Size

1. Introduction

McDonald and Hodaj [1] investigated two subset selection rules for $k(\geq 2)$ exponential populations that possibly differ in their threshold parameters. The scale parameters of the populations were assumed to be known and equal. Without loss of generality, the scale parameters can be taken to be one as dividing the data by the known common scale parameter leads to a data set possessing a unity scale parameter for all populations. Let X_{ij} be the j^{th} independent sample value drawn from the i^{th} population, denoted by Π_i, $i=1,\cdots,k$, $j=1,\cdots,n$. Let γ_i denote the threshold parameter of the i^{th} population. The cumulative distribution function (cdf) and the probability density function (pdf) for the sample value X_{ij} is given, respectively, by

$$F(x;\gamma_i) = 1 - \exp\left[-(x-\gamma_i)\right], \quad x \ge \gamma_i, \quad |\gamma_i| < \infty, \tag{1.1}$$

and

$$f(x;\gamma_i) = \exp\left[-(x-\gamma_i)\right], \quad x \ge \gamma_i, \quad |\gamma_i| < \infty. \tag{1.2}$$

Subset selection and ranking problems have a long history in the statistical decision literature; see for example Bechhofer [2], Gupta [3], Gupta and Panchapakesan [4], and Gibbons *et al.* [5], among others. These works develop general multiple decision procedures for selecting the best population under various distributional and sampling assumptions. In two recent papers, McDonald and Hodaj [1] [6] specialize these ideas to exponential populations with possibly different threshold parameters. They develop the operating characteristics of several subset selection rules under both exact and approximate computation requirements. The present note focuses on the special case, where exact analytic calculations are feasible, and compares in detail two natural selection rules based on the sample minimum and the sample mean.

The goal of the subset selection procedure is to choose a subset of the k populations such that the probability of the population with the largest threshold parameter is included in the subset with a probability no less than a user specified value P^* ($1/k < P^* < 1$) no matter what the underlying configuration of the threshold parameters. That is, the probability of a Correct Selection (CS) is greater than or equal to P^*,

$$\Pr(\text{CS}) \ge P^*. \tag{1.3}$$

Two subset selection procedures are considered by McDonald and Hodaj [1]. The first, denoted by R_1, is based on the sample minimum statistic from each population. Let $Y_i, i = 1,\cdots,k$ denote the minimum order statistic from the sample values $X_{ij}, j = 1,\cdots,n$. The expected value of Y_i is $\gamma_i + (1/n)$, and the variance is $(1/n^2)$. The selection rule is given by

$$R_1: \text{Select } \Pi_i \text{ iff } Y_i \ge \max\left(Y_j, j = 1,2,\cdots,k\right) - d, \quad d \ge 0. \tag{1.4}$$

The d-value is chosen to be as small as possible while still preserving the P^* condition (1.3).

A second selection procedure is based on the sample means. This is motivated by the expected value of the sample mean for the i^{th} population being $1+\gamma_i$, and the variance $1/n$. The selection rule is similar to that based on the sample minimum values. Let $\overline{X_i}$ denote the sample mean of X_{ij}, $j = 1,\cdots,n$, for $i = 1,\cdots,k$. Then

$$R_2: \text{Select } \Pi_i \text{ iff } \bar{X}_i \ge \max\left(\bar{X}_j, j = 1,2,\cdots,k\right) - b, \quad b \ge 0. \tag{1.5}$$

As with d, the b-value is chosen to be as small as possible while still maintaining the inequality (1.3).

The articles by McDonald and Hodaj [1] and [6] address computational methods to determine the values of b and d to satisfy the P^* condition and operating characteristics (OCs) of the selection rules R_1 and R_2. The OCs include the Pr(CS), the probability of an incorrect selection (Pr(ICS)) and the expected size of the selected subset (ESS) under several parametric configurations (slippage and equi-spaced) of the threshold parameters. Computational methods include simulations for arbitrary $\left(k, n, P^*\right)$, calculations based on the Central Limit Theorem (CLT) for R_2, exact calculations for R_2 with $k=2$. Section 2 provides exact computational methods for the selection rule R_1 for the d-value and the OCs.

2. Special Case for *k* = 2 Populations for Computing the OCs for *R*1 and *R*2

Let $Y_i, i=1,2$ be the minimum values of the samples from Π_i respectively; and $\gamma_i, i=1,2$ be the corresponding threshold parameters for the two populations. The ordered values of the $\gamma's$ are denoted by $\gamma_{[1]} \le \gamma_{[2]}$, and $Y_{(i)}$ denotes the minimum sample value that emanates from the population associated with $\gamma_{[i]}$. For a sample of size n, the cumulative density function (cdf) and probability density function (pdf) of the minimum value are given respectively by (see, e.g., McDonald and Hodaj [1])

$$\Pr\left(Y \le y\right) = 1 - \exp\left[-n\left(y-\gamma\right)\right], \quad y \ge \gamma, \tag{2.1}$$

and

$$f_Y\left(y\right) = n \cdot \exp\left[-n\left(y-\gamma\right)\right], \quad y \ge \gamma. \tag{2.2}$$

For the calculation of the OCs for R_1, set $\gamma_{[1]} = 0$ and $\gamma_{[2]} = \delta \ge 0$. Denote the Pr(CS) for selection rule R_1 by Pr(CS1) and for an incorrect selection by Pr(ICS1), and the expected subset size by ESS1. Then

$$\begin{aligned} \Pr\left(\text{CS1}\right) &= \Pr\left(Y_{(2)} \ge Y_{(1)} - d\right) \\ &= n \cdot \exp\left(n\delta\right) \int_{\delta}^{\infty} 1 - \exp\left[-n\left(y+d\right)\right] \cdot \exp\left(-ny\right) \mathrm{d}y \\ &= 1 - \frac{1}{2}\exp\left[-n\left(d+\delta\right)\right]. \end{aligned} \tag{2.3}$$

The Pr(CS1) is minimized when $\delta = 0$, so for a given value of P^*,

$$d = -\log\left[2\left(1-P^*\right)\right] / n. \tag{2.4}$$

Following the same line of derivation for Pr(CS1) with $\delta \ge d$, gives

$$\begin{aligned} \Pr\left(\text{ICS1}\right) &= \Pr\left(Y_{(1)} \ge Y_{(2)} - d\right) \\ &= \frac{1}{2}\exp\left[-n\left(\delta - d\right)\right]. \end{aligned} \tag{2.5}$$

And with $d \ge \delta$, then

$$\Pr\left(\text{CS1}\right) = 1 - \frac{1}{2}\exp\left[-n\left(d-\delta\right)\right], \tag{2.6}$$

and

$$\text{ESS1} = \Pr\left(\text{CS1}\right) + \Pr\left(\text{ICS1}\right). \tag{2.7}$$

The OCs described in this Section for $k=2$ and the selection rule R_1 based on the minimum order statistics, and the corresponding OCs for R_2 based on the sample means, are codified in the R-code of Appendix A. Table 1 provides the output of this code for $n=25$, $P^*=0.95$, and a slippage value for γ_2 denoted by δ. Calculations are given to 5 dp. The constants for the selection rules R_1 and R_2 are $d=0.0921$ and $b=0.46523$. The entries in Table 1 for R_2 are given in Section 2, Table 1 of McDonald and Hodaj [6], and are repeated here for comparison to the corresponding values for R_1.

Table 1. OCs for rules R_1 (red) and R_2 (blue): $k=2$, $n=25$, $P^*=0.95$, $\gamma_{[2]}-\gamma_{[1]}=\delta$.

δ=	0	0.05	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55
Pr (CS)	0.95	0.98567	0.99882	0.99966	0.9999	0.99997	0.99999	1	1	1	1
Pr (CS)	0.95	0.96574	0.98519	0.99066	0.99428	0.99659	0.99803	0.99889	0.99939	0.99968	0.99983
Pr (ICS)	0.95	0.82548	0.11759	0.03369	0.00965	0.00277	0.00079	0.00023	0.00007	0.00002	0.00001
Pr (ICS)	0.95	0.92896	0.86747	0.82581	0.77666	0.72045	0.65815	0.5912	0.52148	0.45109	0.38221
ESS	1.9	1.81116	1.11641	1.03335	1.00956	1.00274	1.00078	1.00022	1.00006	1.00002	1.00001
ESS	1.9	1.8947	1.85266	1.81648	1.77094	1.71704	1.65618	1.59009	1.52087	1.45077	1.38204

The probabilities of correct and incorrect selection for R_2 are denoted, respectively, by Pr(CS2) and Pr(ICS2), and the expected subset size by ESS2. Since the exponential distribution has finite variance, the CLT applies to the sample mean. As noted in the earlier publication [6], by the CLT (see Navidi [7], Section 4.11), the distribution of the sample mean, $\overline{X}_i$, is approximately normal with mean and variance equal to γ_i+1 and $1/n$ respectively. In McDonald and Hodaj [6], CLT-based probabilities were found to be numerically very close to those obtained from the exact gamma distribution for moderate sample sizes supporting the use of the normal approximation here. Then using the CLT, after noting that $\overline{X}_{(2)} \geq \overline{X}_{(2)} - b$, and that a linear combination of normal random variables follows a normal distribution, it follows that

$$\Pr(\text{CS2}) = \Pr\left(\overline{X}_{(2)} \geq \overline{X}_{(1)} - b\right) = \Phi\left[(b+\delta)\big/\sqrt{2/n}\right]. \tag{2.8}$$

Since the Pr(CS2) is minimized when $\delta=0$, the value of b is given by

$$b = \sqrt{2/n}\cdot\Phi^{-1}\left(P^*\right). \tag{2.9}$$

The expressions for Pr(ICS2) and ESS2 follow in a similar manner.

$$\Pr(\text{ICS2}) = \Phi\left[(b-\delta)\big/\sqrt{2/n}\right], \tag{2.10}$$

and

$$\text{ESS2} = \Pr(\text{CS2}) + \Pr(\text{ICS2}). \tag{2.11}$$

Here the functions $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the cumulative distribution function (cdf) and the inverse cdf for a standardized normal random variable.

The OCs for R_1 are substantially better than those for R_2: higher Pr(CS), lower Pr(ICS), and lower ESS.

Table 2 and Table 3 examine the OCs of the two selection rules, as done in Table 1, for different choices of sample sizes and the P^* values. Table 2 uses a sample size of 10 with P^* equal to 0.90 and selection constants $d = 0.16094$ and $b = 0.57313$. Table 3 uses a sample size of 50 with P^* equal to 0.99 and selection constants $d = 0.07824$ and $b = 0.46527$. The OC results in these two additional cases are qualitatively the same as those given in Table 1. The selection rule R_1, compared to R_2, displays consistently higher Pr(CS), lower Pr(ICS), and lower ESS.

Table 2. OCs for rules R_1 (red) and R_2 (blue): $k = 2$, $n = 10$, $P^* = 0.90$, $\gamma_{[2]} - \gamma_{[1]} = \delta$.

$\delta =$	0	0.05	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55
Pr (CS)	0.90	0.93935	0.97769	0.98647	0.99179	0.99502	0.99698	0.99817	0.99889	0.99933	0.99959
Pr (CS)	0.90	0.91824	0.94706	0.95807	0.96716	0.97455	0.9805	0.98522	0.98992	0.99179	0.99399
Pr (ICS)	0.90	0.83513	0.55183	0.33834	0.20521	0.12447	0.07549	0.04579	0.02777	0.01684	0.01022
Pr (ICS)	0.90	0.87895	0.82796	0.79795	0.76502	0.72931	0.69108	0.65067	0.60847	0.56494	0.52062
ESS	1.8	1.77447	1.52952	1.3248	1.197	1.11949	1.07247	1.04396	1.02666	1.01617	1.00981
ESS	1.8	1.79719	1.77502	1.75603	1.73217	1.70386	1.67158	1.63589	1.59739	1.55674	1.51461

Table 3. OCs for rules R_1 (red) and R_2 (blue): $k = 2$, $n = 50$, $P^* = 0.99$, $\gamma_{[2]} - \gamma_{[1]} = \delta$.

$\delta =$	0	0.05	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55
Pr (CS)	0.99	0.99918	0.99999	1	1	1	1	1	1	1	1
Pr (CS)	0.99	0.99501	0.99895	0.99956	0.99983	0.99993	0.99998	0.99999	1	1	1
Pr (ICS)	0.99	0.87818	0.01383	0.00113	0.00009	0.00001	0	0	0	0	0
Pr (ICS)	0.99	0.98107	0.94253	0.90764	0.85911	0.7957	0.71781	0.62792	0.53043	0.43107	0.33591
ESS	1.98	1.87735	1.01382	1.00113	1.00009	1.00001	1	1	1	1	1
ESS	1.98	1.97608	1.94148	1.9072	1.85894	1.79563	1.71779	1.62791	1.53043	1.43107	1.33591

3. Conclusions

The results herein obtained further reinforce the conclusion that the subset selection rule based on the sample minimum values outperforms that based on the sample means when the CLT is applicable. Both selection rules are based on unbiased estimators of the exponential threshold parameter. The variance of the sample minimum value is n^{-2}, while that of the sample mean is n^{-1} making the min value a more efficient estimator of the threshold parameter. When applying subset selection within the context herein described, the expected number of populations selected using the minimum values will be no greater than that using the mean values while still preserving the required P^* condition.

From a practical standpoint, such subset selection rules are directly relevant in reliability and life-testing studies where experimenters must select a design, component, or treatment with the largest threshold (or guaranteed minimum lifetime) among several candidates. Using the rule based on sample minima allows practitioners to achieve a desired probability of correct selection with an expected subset size no greater than that obtained with a selection rule based on sample means. This can translate into fewer units being carried forward for further testing or field deployment while still maintaining prespecified reliability guarantees; see Meeker *et al.* [8] for a discussion of applications of exponential and related lifetime models.

The sum of independent exponential random variables follows a gamma distribution which may be used with small samples when the use of the CLT might not be appropriate. This approach was investigated by McDonald and Hodaj [6], and results using the gamma distribution were shown to be comparable to those using the CLT.

This note focuses on analytic expressions for the operating characteristics of two subset selection rules with $k = 2$ permitting quantitative comparisons of the rules with computer evaluations. Further challenges and avenues for future work arise in several areas. Might there be a mathematical proof that Pr(CS1) is never less than that of the rule based on sample means? And can similar results be derived for greater than two populations? More broadly, the exponential distribution applies in applications where the hazard function is constant. This suggests that the population of units is not wearing out over time. The exponential distribution is a special case of the Weibull distribution which is another direction in which to expand these analyses.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] McDonald, G.C. and Hodaj, J. (2025) Selection Rules for Exponential Population Threshold Parameters. *Applied Mathematics*, **16**, 1-14. https://doi.org/10.4236/am.2025.161001

[2] 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. https://doi.org/10.1214/aoms/1177728845

[3] Gupta, S.S. (1965) On Some Multiple Decision (Selection and Ranking) Rules. *Technometrics*, **7**, 225-245. https://doi.org/10.1080/00401706.1965.10490251

[4] Gupta, S.S. and Panchapakesan, S. (1979) Multiple Decision Procedures. Wiley.

[5] Gibbons, J.D., Olkin, I. and Sobel, M. (1977) Selecting and Ordering Populations: A New Statistical Methodology. Wiley.

[6] McDonald, G.C. and Hodaj, J. (2025) Operating Characteristics of Subset Selection Rules for Exponential Population Threshold Parameters. *Applied Mathematics*, **16**, 441-460.

https://doi.org/10.4236/am.2025.165024

[7] Navidi, W.N. (2024) Statistics for Engineers and Scientists. 6th Edition, McGraw Hill.

[8] Meeker, W.Q., Escobar, L.A. and Pascual, F.G. (2022) Statistical Methods for Reliability Data. 2nd Edition, Wiley.

Appendix

```
#OCs for R1 and R2, k=2
#R1 values end in 1; R2 end in 2
#Set n, delta, and P
n<-50; delta<-c(0,0.05,0.15,0.20,0.25,0.30,0.35,0.40,0.45,0.50,0.55); P<-0.99
len<-length(delta)
PCS1<-rep(0,len);PCS2<-rep(0,len);PICS1<-rep(0,len)
PICS2<-rep(0,len)
d<--log(2*(1-P))/n
b<-sqrt(2/n)*qnorm(P)
params<-c(d,b)
params
for (i in 1:len){
PCS1[i]<-1-0.5*exp(-n*delta[i]-n*d)
PCS2[i]<-pnorm((b+delta[i])/sqrt(2/n))
if(delta[i]>=d){PICS1[i]<-(0.5)*exp(-n*(delta[i]-d))}
if(d>delta[i]){PICS1[i]<-1-0.5*exp(-n*(d-delta[i]))}
PICS2[i]<-pnorm((b-delta[i])/sqrt(2/n))
}
ESS1<-PCS1+PICS1
ESS2<-PCS2+PICS2
df1<-data.frame(n,d,b,P,delta,PCS1,PCS2,PICS1,PICS2,ESS1,ESS2)
df1<-round(df1,4)
df1
length(delta)
length(PCS1)
```