Applied Mathematics, 2025, 16(5), 441-460
https://www.scirp.org/journal/am
ISSN Online: 2152-7393
ISSN Print: 2152-7385

Operating Characteristics of Subset Selection Rules for Exponential Population Threshold Parameters

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) Operating Characteristics of Subset Selection Rules for Exponential Population Threshold Parameters. *Applied Mathematics*, **16**, 441-460.
https://doi.org/10.4236/am.2025.165024

Received: April 25, 2025
Accepted: May 25, 2025
Published: May 28, 2025

Abstract

This article provides the operating characteristics (OCs) of two subset selection rules for exponential populations having a common known scale parameter and possibly differing threshold parameters. One selection rule is based on the minimum sample values and the other is based on the mean (or sum) of the sample values. The random samples drawn from the populations are of equal size. The goal of the selection rules is to choose a subset of the populations such that the population possessing the largest threshold parameter (the "best" population) is contained in the subset with a probability no less than a user prescribed value P^*. A correct selection occurs if the best population is contained in the selected subset. The OCs are the probability of a CS and the expected size of the selected subset. The OCs are calculated and compared for several formulations of the selection rules and for two threshold parameter configurations—slippage, and equi-spaced. The computer R-codes for all calculations are given in the Appendices.

Keywords

Minimum Statistic Selection Procedure, Means Selection Procedure, Probability of Correct Selection, Probability of Incorrect Selection, Expected Subset Size, Slippage Configuration, Equi-Spaced Configuration

1. Introduction

The Weibull distribution is one of the most widely employed models in reliability and survival analysis due to its flexibility in modeling various hazard rate behaviors. It is frequently used to characterize the lifetime distributions of components and systems in engineering applications, including mechanical parts, electronic devices, and structural materials. Beyond engineering, it is also extensively applied in biomedical and epidemiological studies to model time-to-event data, such as the latency period of diseases or time to failure in biological systems. See, for example, Lawless [1] and Nelson [2].

The exponential distribution rises as a special case of the Weibull distribution when β, the shape parameter, equals 1. The exponential distribution, characterized by a constant failure rate, is especially useful for modeling electronic components and systems with memoryless lifetimes. In contrast, the Weibull distribution with the shape parameter, allows it to model increasing, decreasing, or constant failure rates, making it suitable for a wide range of mechanical, structural, and industrial applications. This adaptability makes the Weibull distribution a cornerstone in life data analysis, failure prediction, and maintenance scheduling. Both distributions support parameter estimation, hazard rate modeling, and reliability function derivation, providing critical insights into product life cycles, risk assessment, and quality control. (Meeker, *et al.* [3])

Two authoritative resources on subset selection procedures are the comprehensive works by Gupta and Panchapakesan [4], and Gibbons *et al.* [5]. In particular, subset selection methods for populations following the exponential distribution have been extensively studied in the literature. For instance, Ng [6] presents procedures for identifying desirable exponential populations under both known and unknown scale parameter scenarios. The definition of a "good" population in this context follows the criteria outlined in Lam [7].

In this Section, the exponential threshold model considered by McDonald and Hodaj [8] is further assessed within the framework of subset selection rules. Specifically, the performance of their selection rules R_1 and R_2 will be compared. Let π_i, $i=1,\cdots,k$, be k (≥2) independent populations with random draws from π_i following an exponential probability distribution with scale parameter equal to 1 and threshold parameter equal to Υ_i. Without loss of generality, the scale parameter can be any known value. If not equal to 1, then simply divide all the sample values by that common known scale parameter and proceed with the modified sample as coming from populations with unit scale parameter. Let X_{ij}, $j=1,\cdots,n$, denote an independent random sample of size n from the i^{th} population. Let $Y_i=\min\left(X_{ij}, j=1,\cdots,n\right)$, and $\bar{X}_i$ equal the sample mean of $X_{ij}, j=1,\cdots,n$.

The goal of the subset selection rules is to select a subset of the k populations so as to include the "best" population, *i.e.*, the population that is associated with the largest threshold parameter, Υ, with a user prescribed probability (P^*) no less than $1/k$. That is, the probability of a Correct Selection (CS) is at least equal to a user specified value of P^* no matter what the underlying configuration of the population threshold parameters may be. If two or more populations possess the largest threshold parameter, one of these is tagged at random and denoted as the "best".

The two subset selection rules considered are:

R_1: Select π_i iff $Y_i \ge \max\left(Y_j, j=1,\cdots,k\right)-d$, $d\ge 0$, and (1.1)

R_2: Select π_i iff $\bar{X}_i \geq \max\left(\bar{X}_j, j=1,\cdots,k\right)-b$, $b \geq 0$. (1.2)

The nonnegative constants, *d* and *b*, are chosen to satisfy the *P** condition, *i.e.*,

$$\min \Pr\left(\text{CS}\right) \geq P*, \tag{1.3}$$

where the minimum is taken over all possible configurations of $\Upsilon_i, i=1,\cdots,k$. Computational methods for these constants are given in [8].

2. Special Case for *k* = 2 Populations and Large *n* for Computing *b* in R2

We first consider the case of two populations and a large sample size. This will give a basis for comparing the accuracy of a relevant simulation methodology to be introduced in Section 3. By the Central Limit Theorem, the distribution of the sample mean, $\bar{X}_i$, is approximately normal with mean and variance Υ_i+1 and $1/n$, respectively. Now consider selection rule R_2 given in (1.2). Let $\bar{X}_{(i)}$ denote the sample mean drawn from the population associated with $\Upsilon_{[i]}$, where $\Upsilon_{[1]}=1 \leq \Upsilon_{[2]}=1+\delta$, $\delta \geq 0$. Then as *n* grows large, and noting $\bar{X}_{(2)} \geq \bar{X}_{(2)}-b$, it follows that Pr(CS) approaches

$$\begin{aligned}\Pr\left(\text{CS}\right) &= \Pr\left(\bar{X}_{(2)} \geq \bar{X}_{(1)}-b\right)\\ &= \Pr\left(\bar{X}_{(1)}-\bar{X}_{(2)} \leq b\right)\\ &= \Phi\left[\left(b+\delta\right)/\text{sqrt}\left(2/n\right)\right],\end{aligned} \tag{1.4}$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of a normal variable with mean 0 and variance 1. The move from the second to the third equality in (1.4) results from the normality of a linear combination of independent normal variates, e.g., $\bar{X}_{(1)}-\bar{X}_{(2)}$, Navidi [[9], Chapt. 4]. The mean and variance of $\bar{X}_{(1)}-\bar{X}_{(2)}$ are $-\delta$ and $2/n$, respectively. Since $\delta \geq 0$, it follows from (1.4) that $\Pr\left(\text{CS}\right) \geq \Phi\left[b/\text{sqrt}\left(2/n\right)\right]$. Thus, for *k* = 2 and a given value of *n* and *P**, the *b*-value is given by

$$b = \text{sqrt}\left(2/n\right)*\Phi^{-1}\left(P*\right), \tag{1.5}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the cdf $\Phi(\cdot)$.

An incorrect selection (ICS) occurs when the population associated with $\Upsilon_{[1]}$ is included in the selected subset. Thus, following the derivation of (1.4)

$$\begin{aligned}\Pr\left(\text{ICS}\right) &= \Pr\left(\bar{X}_{(1)} \geq \bar{X}_{(2)}-b\right)\\ &= \Pr\left(\bar{X}_{(2)}-\bar{X}_{(1)} \leq b\right)\\ &= \Phi\left[\left(b-\delta\right)/\text{sqrt}\left(2/n\right)\right].\end{aligned} \tag{1.6}$$

Let $S_{[i]}=1$ if the population associated with $\Upsilon_{[i]}$ is included in the selected subset, otherwise $S_{[i]}=0$, $i=1,2$. Then the expected size of the selected subset (ESS) is

$$\begin{aligned}\text{ESS} &= E\left(S_{[1]}\right)+E\left(S_{[2]}\right)\\ &= \Pr\left(\text{ICS}\right)+\Pr\left(\text{CS}\right).\end{aligned} \tag{1.7}$$

Appendix A provides R-code for calculating the Operating Characteristics (OCs), *i.e.*, Pr(CS), Pr(ICS), and ESS, for given values of $k = 2$, n, P^*, and b. Table 1 provides such output for a reasonably large value of $n = 25$. Gupta and Panchapakesan [[4], Sec. 11.2] and Gibbons *et al.* [[5], Sec. 3.2] address the expected subset size properties within the context of criteria for evaluating the performance of a subset selection procedure.

Table 1. R_2 operating characteristics for $k = 2$, $n = 25$, $P^* = 0.95$, and $b = 0.46523$.

δ	Pr (CS)	Pr (ICS)	ESS
0	0.95	0.95	1.90
0.1	0.97716	0.90170	1.87886
0.2	0.99066	0.82581	1.81648
0.3	0.99659	0.72045	1.71704
0.4	0.99889	0.59120	1.59009
0.5	0.99968	0.45109	1.45077
0.6	0.99992	0.31687	1.31679
0.7	0.99998	0.20326	1.20324
0.8	1.00000	0.11829	1.11829
0.9	1.00000	0.06213	1.06213
1.0	1.00000	0.02933	1.02933

3. OCs of R_1 and R_2 for Arbitrary k, n, and P^*: Slippage Configuration

Now consider the case of $k\,(\geq 2)$ populations with a random sample of size n drawn from each of the populations. Each of the populations follow an exponential distribution with scale parameter equal to 1 and threshold parameter equal to 0 for k-1 populations and equal to $\delta\,(\geq 0)$ for the remaining population. Using subset selection rule R_2 (1.2) a subset of the populations is chosen to contain the "best" population, *i.e.*, the one associated with the threshold value δ, with a probability no less than a prescribed P^* $(1/k < P^* < 1)$.

The R-code in **Appendix B** is used to simulate the operating characteristics Pr(CS), Pr(ICS), and ESS for specified values of k, n, P^*, δ and N. The process of generating a data set as specified is repeated N times. For each of the repeats, the subset selection is made, and the populations chosen given a score of 1. The averages of the population scores estimate the probabilities that the individual populations are chosen by R_2. From these probability estimates the OCs are then computed, the ESS being the sum of the estimated selection probabilities of the individual populations.

To assess the accuracy of the simulation approach, the values similar to those given in Table 1 are calculated and displayed in Table 2. Note the b-value is determined by the simulation of 200,000 draws and differs very slightly from those given in Section 2. The entries in Table 2 are quite close to those in Table 1. For example, the ESS entry in Table 2 for $\delta = 0.5$ is 0.145% less than the corresponding entry from Table 1. Overall, the mean absolute percentage difference for ESS Table 2 entries compared to those from Table 1 is 0.115%. Thus, for $k = 2$ the simulation results are in very close agreement with the exact values obtained from the Central Limit Theorem presented in Section 2.

Table 2. R_2 OCs for $k = 2$, $n = 25$, $P^* = 0.95$, $b = 0.46550$, and $N = 200{,}000$.

δ	Pr (CS)	Pr (ICS)	ESS
0	0.94999	0.95025	1.90024
0.1	0.97723	0.90343	1.88066
0.2	0.99038	0.82858	1.81896
0.3	0.99612	0.72367	1.71979
0.4	0.99864	0.59281	1.59144
0.5	0.99958	0.44909	1.44867
0.6	0.99986	0.31404	1.31390
0.7	0.99996	0.20040	1.20036
0.8	1.00000	0.11680	1.11680
0.9	1.00000	0.06171	1.06171
1.0	1.00000	0.02935	1.02935

Appendix C provides an R-code to compute the operating characteristics of selection rule R_1 (1.1) and the appropriate constant $d = d(k, n, P^*)$. This code is structured very similar to that of **Appendix B** with the exception that the simulation of the sample means statistics are replaced by the sample minimum value statistics. A comparison of the OCs of the two selection procedures is given in Table 3 (to four dp) and plotted in Figure 1 for $k = 10$, $n = 25$, $P^* = 0.95$. These values are based on 200,000 simulations for a slippage configuration of exponential populations with common scale parameter equal to 1 and nine populations with threshold parameters equal to 0 and one population, the "best", with threshold parameter $\delta = 0(0.1)1$. While the Pr(CS) for the two selection procedures are very close, the ESS for selection rule R_1 based on the sample minimum values is substantially less than that for R_2 based on the sample means.

Table 3. OCs for rules R_1 (red) and R_2 (blue): $k = 10$, $n = 25$, $P^* = 0.95$, $N = 200{,}000$, slippage configuration.

$\delta =$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Pr (CS)	0.9499	0.9960	0.9997	1	1	1	1	1	1	1	1
Pr (CS)	0.9500	0.9801	0.9930	0.9976	0.9993	0.9998	0.9999	1	1	1	1
ESS	9.5022	9.0118	3.6245	1.2189	1.0180	1.0013	1.0001	1	1	1	1
ESS	9.4934	9.4575	9.3269	9.0513	8.5699	7.8283	6.8168	5.6195	4.3817	3.2654	2.3802

4. OCs of R_1 and R_2 for Arbitrary *k*, *n*, and *P**: Equi-Spaced Configuration

A second parametric configuration for OC comparisons is the equi-spaced configuration. In this particular setup for $k = 10$ populations, the threshold parameters are fixed at $\Upsilon_i = (i-1)\cdot\delta$, $i = 1,\cdots,k$. Thus, the difference between any two adjacently ordered population threshold parameters is δ. R-codes in **Appendix D** and **Appendix E** provide the OCs for selection rules R_2 and R_1, respectively. Table 4 provides the output for $k = 10$, $n = 25$, $P^* = 0.95$ for $N = 200{,}000$ simulations. The quantity $\Pr(\pi_i)$ is the estimated probability of selecting the i^{th} population. Population π_{10} is the "best" and its selection probability is denoted by Pr(CS). The ESS is the sum of the ten estimated selection probabilities. As in the case with the slippage configuration, the ESS for R_1 is substantially less than that for R_2 for all positive values of δ.

Table 4. OCs for rules R_1 (red) and R_2 (blue): $k = 10$, $n = 25$, $P^* = 0.95$, $N = 200{,}000$, equi-spaced configuration.

$\delta =$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Pr (π_1)	0.9505	0.0000	0.0000	0.0000	0.0000	0	0	0	0	0	0
Pr (π_1)	0.9490	0.1294	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_2)	0.9500	0.0000	0.0000	0.0000	0.0000	0	0	0	0	0	0
Pr (π_2)	0.9498	0.2258	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_3)	0.9502	0.0000	0.0000	0.0000	0.0000	0	0	0	0	0	0
Pr (π_3)	0.9487	0.3594	0.0041	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_4)	0.9506	0.0000	0.0000	0.0000	0.0000	0	0	0	0	0	0
Pr (π_4)	0.9496	0.5118	0.0247	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_5)	0.9498	0.0001	0.0000	0.0000	0.0000	0	0	0	0	0	0
Pr (π_5)	0.9498	0.6671	0.1060	0.0022	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_6)	0.9502	0.0020	0.0000	0.0000	0.0000	0	0	0	0	0	0
Pr (π_6)	0.9496	0.7961	0.3074	0.0323	0.0009	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_7)	0.9503	0.0237	0.0000	0.0000	0.0000	0	0	0	0	0	0
Pr (π_7)	0.9490	0.8899	0.5999	0.2211	0.0373	0.0027	0.0001	0.0000	0.0000	0.0000	0.0000
Pr (π_8)	0.9504	0.2881	0.0021	0.0000	0.0000	0	0	0	0	0	0
Pr (π_8)	0.9488	0.9486	0.8451	0.6326	0.3620	0.1468	0.0418	0.0083	0.0012	0.0002	0.0000
Pr (π_9)	0.9504	0.9312	0.2967	0.0244	0.0020	0.0001	0	0	0	0	0
Pr (π_9)	0.9493	0.9810	0.9643	0.9290	0.8691	0.7786	0.6587	0.5189	0.3768	0.2521	0.1529
Pr (CS)	0.9499	0.9996	1.0000	1.0000	1.0000	1.0000	1	1	1	1	1
Pr (CS)	0.9500	0.9963	0.9991	0.9997	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ESS	9.5022	2.2448	1.2987	1.0244	1.0020	1.0001	1	1	1	1	1
ESS	9.4935	6.5053	3.8512	2.8169	2.2692	1.9281	1.7006	1.5272	1.3780	1.2522	1.1529

The ESS values for the two parameter configurations are displayed in **Figure 1** (see **Appendix F** and **Appendix G**). The values for the slippage (equi-spaced) configuration are given in the left (right) side. Clearly the selection procedure R_1 outperforms R_2 with respect to these metrics. This is somewhat explained by the moments of the sample minimum value and the sample mean given in **Table 5**. Since both R_1 and R_2 can be expressed in terms of unbiased estimators of ϒ, the sample minimum is more efficient as its variance is a factor of n^{-1} times that of the sample mean.

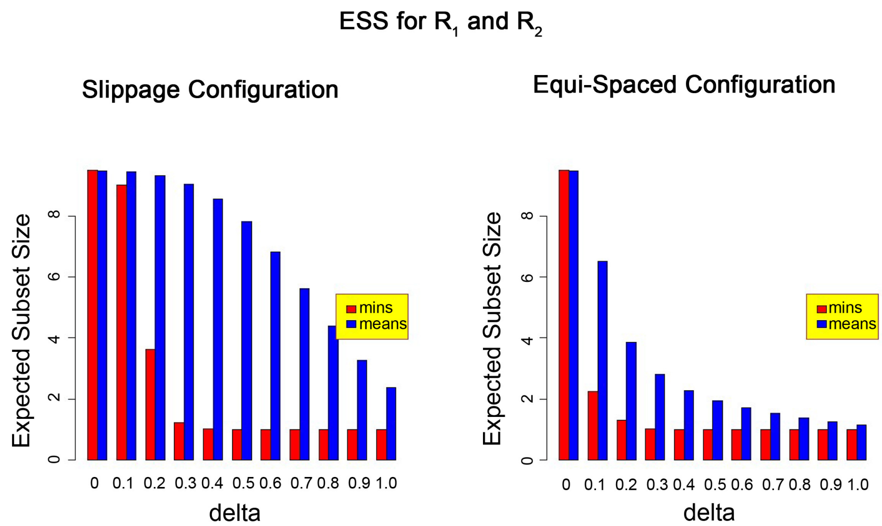

Figure 1. Expected subset sizes with data from **Table 3** and **Table 4**.

Table 5. Moments of sample statistics underlying selection rules R_1 and R_2.

Statistic	Sample Minimum	Sample Mean
Expected Value =	$\Upsilon + (1/n)$	$\Upsilon + 1$
Variance =	$1/n^2$	$1/n$

McDonald and Hodaj [8] generate a data set for $k = 10$ exponential populations with a random sample of size 25 from each population. This data set is produced using the R-code in their **Appendix B**. For each of the population draws, the minimum value and sample means are calculated and rules R_1 and R_2 applied to select subsets to contain the "best" with $P^* = 0.75$, 0.90, 0.95, 0.975, and 0.99. The findings are reported in their **Table 4**, and show the means procedure R_2 chooses fewer populations for four of the P^* values and an equal number for one value of P^*. These findings seem somewhat at odds with what has been reported in this article. However, the data set leading to these findings in [8] is based on 10 exponential populations all with threshold parameters equal to 0 and rate parameters (=1/scale) equal to $1/i$, $i = 1(1)10$. The expected values of the ten populations are 1(1)10. Thus, the selection rules were applied to exponential populations differing in expected values but not threshold values, and those results are not meaningful for the model under consideration in this article. The five lines in the R-code from **Appendix B** [8] beginning with "gamma < −seq (from = 1, to =10, by = 1)" and ending with "M[,i] < −rexp (n,lambda[i])" should be replaced by the four lines beginning with "gamma < −seq (from = 1, to =2.8, by = 0.2)" and ending with "M[,i] < −rexp(n,1)+gamma[i]" from **Appendix H** given here. This issue will now be further addressed.

Using the R-code in **Appendix H** random samples of size 25 are generated from 10 exponential populations having ϒ-values equal to 1(0.2)2.8. Table 6 gives the minimum and mean values of these samples. Table 7 gives the constants required to implement the two selection rules for five values of P^*, and Table 8 indicates which of the ten populations are selected by R_1 and R_2 for each of the five P^* values. For $P^* = 0.75$ each of the selection rules only select the "best" population, *i.e.*, π_{10}. For all values of P^*, R_1 only selects the "best" population, whereas R_2 progressively chooses two or three populations as P^* increases from 0.90 to 0.99. For this one set of data R_1 outperforms R_2 insofar as choosing an equal or smaller number of populations for any of the P^* values investigated. This aligns with the earlier results herein obtained with respect to expected subset sizes, and displayed in Table 3 and Table 4.

Table 6. Minimum and means for each sample of 25 from the 10 populations.

π_i	1	2	3	4	5	6	7	8	9	10
Min	1.0182	1.2084	1.4353	1.6025	1.8048	2.0039	2.2168	2.4234	2.6066	2.8829
Mean	2.0802	1.8510	2.1934	2.4002	2.9885	3.1033	3.0788	3.4006	3.5907	4.0738

Table 7. The *d*- and *b*-values used for R_1 and R_2, respectively.

P^*	0.75	0.90	0.95	0.975	0.99
***d*-value, R_1**	0.10878	0.14995	0.17914	0.20744	0.24319
***b*-value, R_2**	0.48159	0.62742	0.71436	0.79057	0.88100

Table 8. Selected populations using rules R_1 and R_2.

P^*	0.75	0.90	0.95	0.975	0.99
R_1	10	10	10	10	10
R_2	10	9, 10	8, 9, 10	8, 9, 10	8, 9, 10

Table 3 and Table 4 provide results for moderately large values of $k = 10$ and $n = 25$. Table 9 and Table 10 are very similar to those of 3 and 4 only using "small" values for the number of populations and common sample sizes, $k = 5$ and $n = 10$. Results for Table 9 and Table 10 are obtained from the R-codes in (**Appendices B-E**). The conclusions from these computations are very similar to those derived earlier. The selection rules based on the sample minimums, R_1, are notably better than those based on the sample means, R_2, with respect to the Pr(CS) and ESS.

Table 9. OCs for rules R_1 (red) and R_2 (blue): $k = 5$, $n = 10$, $P^* = 0.95$, $N = 200{,}000$, slippage configuration.

$\delta =$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Pr (CS)	0.9500	0.9810	0.9930	0.9974	0.9990	0.9997	0.9999	1	1	1	1
Pr (CS)	0.9500	0.9700	0.9822	0.9896	0.9941	0.9966	0.9981	0.9990	0.9995	0.9997	0.9999
ESS	4.7508	4.6962	4.4797	3.8597	2.3896	1.5164	1.1908	1.0698	1.0259	1.0096	1.0036
ESS	4.7526	4.7418	4.7065	4.6436	4.5483	4.4111	4.2272	3.989	3.7079	3.3821	3.0321

Table 10. OCs for rules R_1 (red) and R_2 (blue): $k = 5$, $n = 10$, $P^* = 0.95$, $N = 200{,}000$, equi-spaced configuration.

$\delta =$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Pr (π_1)	0.9506	0.2930	0.0062	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_1)	0.9504	0.8375	0.5765	0.2596	0.0708	0.0122	0.0015	0.0001	0.0000	0.0000	0.0000
Pr (π_2)	0.9498	0.6552	0.0448	0.0024	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pr (π_2)	0.9499	0.8930	0.7559	0.5302	0.2861	0.1153	0.0349	0.0085	0.0017	0.0003	0.0000
Pr (π_3)	0.9502	0.8690	0.3383	0.0469	0.0065	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000
Pr (π_3)	0.9506	0.9358	0.8863	0.7937	0.6547	0.4844	0.3145	0.1789	0.0888	0.0391	0.0152
Pr (π_4)	0.9503	0.9581	0.9024	0.7397	0.3540	0.1299	0.0480	0.0176	0.0066	0.0024	0.0010
Pr (π_4)	0.9499	0.9638	0.9594	0.9431	0.9165	0.8794	0.8300	0.7665	0.6908	0.6054	0.5134
Pr (π_5)	0.9500	0.9926	0.9978	0.9993	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
Pr (π_5)	0.9500	0.9832	0.9932	0.9969	0.9984	0.9992	0.9996	0.9998	0.9999	0.9999	1.0000
ESS	4.7508	3.7680	2.2896	1.7883	1.3603	1.1308	1.0481	1.0176	1.0066	1.0024	1.0010
ESS	4.7508	4.6133	4.1713	3.5234	2.9265	2.4906	2.1804	1.9538	1.7812	1.6448	1.5286

5. OCs of R_2 Using the Gamma Distribution: Slippage Configuration

Throughout this article, when using the selection rule R_2 based on the sample mean values, the Central Limit Theorem or computer simulation was invoked to calculate the implementation constant, b, and evaluate the OCs of the procedure. In this Section, the exact distribution of the sum of independent exponential random variables is used and the resultant OCs compared to those in Table 2 and Table 3.

As before, the setup is k populations each having an independent random sample of size n, X_{ij}, $i = 1, \cdots, k$, $j = 1, \cdots, n$, and X_{ij} following an exponential distribution with unknown threshold parameter Υ_i and a common known scale parameter η. Without loss of generality, η is assumed to be 1. Denote the sum of the sample values from the i^{th} population, π_i, by S_i. The i^{th} sample mean is then simply S_i/n. The random variables $X_{ij} - \Upsilon_i$ follow an exponential distribution with a zero threshold value and unit scale parameter. Then

$$n\left(\bar{X}_i - \Upsilon_i\right) = \sum_{j=1}^{n}\left(X_{ij} - \Upsilon_i\right) \quad (5.1)$$

follows a gamma distribution (scale version) with probability density

$$g(x \mid n, \eta = 1) = x^{n-1} \exp(-x) / (n-1)!, \ x \geq 0. \quad (5.2)$$

The subset selection rule R_2 (1.2) can now be rewritten as

R_3: Select π_i iff $S_i \geq \max(S_j, j = 1, \cdots, k) - nb$, $b \geq 0$. (5.3)

Following the setup from Section 3 for slippage configurations, all but one of the populations have a zero threshold parameter, and the "best" population has a threshold parameter $\delta \geq 0$. The OCs can be derived as in earlier sections

$$\Pr(\text{CS}) = \int_0^\infty G^{k-1}\left[x + n(b+\delta)\right] \cdot g(x) \mathrm{d}x, \quad (5.4)$$

where $g(\cdot)$ is the probability density (5.2) and $G(\cdot)$ is the corresponding cumulative distribution function. The value b is chosen so as to satisfy the P^* condition (1.3). Continuing,

$$\Pr(\text{ICS}) = \int_0^\infty G^{k-2}(x + nb) \cdot G\left[x + n(b-\delta)\right] \cdot g(x) \mathrm{d}x, \quad (5.5)$$

and

$$\text{ESS} = \Pr(\text{CS}) + (k-1)\Pr(\text{ICS}) \quad (5.6)$$

The OCs for R_3 using sample sums with associated gamma distributions can be calculated using the R-code given in **Appendix I** with input for k, n, b, and δ. R-codes given in **Appendix C** and **Appendix D** can be used to calculate the b-value for selection rule R_2 (1.2) for slippage configurations and equi-spaced configurations respectively.

Three methods have been presented for implementing the selection rule R_2. The first is a simulation approach using the R-code in **Appendix B**. The second is employing the Central Limit Theorem (CLT) and treating the sample means as normally distributed random variables as in the R-code of **Appendix J**. The final approach uses the distribution of the sample sums as gamma random variables as in **Appendix I**. The OCs for these approaches are given in **Table 11**. The entries for the simulation approach (blue) agree with those given in **Table 3**. While the results for the three approaches are somewhat close, the results for simulation (blue) and gamma (purple) approaches are in very close agreement. The ESS based on the CLT yields slightly lower values than the other two entries.

6. Summary and Conclusions

For $k = 2$ and n sufficiently large so that the sample means follow, approximately, a normal distribution, the exact OCs for R_2 can be calculated. Results so obtained are in very close agreement with results based on simulations, thus providing support for the simulation approach for practical applications. The OCs for R_1 are substantially better than those for R_2: higher Pr(CS) and lower ESS for $\delta > 0$. Comparisons herein made of the OCs for R_1 with those for R_2, $k = 25$ and $n = 10$, strongly favor R_1 for both slippage and equi-spaced configurations. The same conclusion followed when similar analyses were done with $k = 5$ and $n = 10$. While

Table 11. OCs for R_2 based on simulation (blue), CLT (black), and gamma distribution (purple): $k = 10$, $n = 25$, $P^* = 0.95$, $N = 200{,}000$ with slippage configuration.

$\delta =$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Pr (CS)	0.95	0.9801	0.993	0.9976	0.9993	0.9998	0.9999	1	1	1	1
Pr (ICS)	8.5434	8.4747	8.3340	8.0537	7.5706	6.8285	5.8168	4.6195	3.3817	2.2654	1.3802
ESS	9.4934	9.4575	9.3269	9.0513	8.5699	7.8283	6.8168	5.6195	4.3817	3.2654	2.3802
Pr (CS)	0.95	0.9802	0.9932	0.998	0.9995	0.9999	1	1	1	1	1
Pr (ICS)	8.5499	8.4781	8.3115	7.9806	7.4141	6.5741	5.4894	4.2648	3.0518	1.995	1.1836
ESS	9.4998	9.4583	9.3047	8.9785	8.4136	7.574	6.4894	5.2648	4.0518	2.995	2.1836
Pr (CS)	0.95	0.9796	0.9925	0.9975	0.9992	0.9998	0.9999	1	1	1	1
Pr (ICS)	8.55	8.4841	8.3417	8.0647	7.5832	6.8413	5.8353	4.6405	3.4016	2.28	1.3913
ESS	9.5	9.4637	9.3343	9.0622	8.5824	7.8411	6.8352	5.6405	4.4016	3.28	2.3913

an earlier result reported in [8] seems to be somewhat at odds with that reported here, it is, in fact, not. Further clarification of the finding in [8] is herein given and shown not to be based on the exponential models herein considered.

Using the fact that sums of exponential random variables follow a gamma probability distribution permits exact calculations of the OCs for the selection rule R_1. A comparison of these exact results with those based on simulations and those based on the normal distribution shows the three approaches yield quite comparable estimates. Overall, the selection rule based on the sample minimums, R_1, has superior OCs over those based on the sample means, R_2, in the cases herein examined.

Conflicts of Interest

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

References

[1] Lawless, J.F. (2002) Statistical Models and Methods for Lifetime Data. Wiley. https://doi.org/10.1002/9781118033005

[2] Nelson, W. (2004) Applied Life Data Analysis. Wiley.

[3] Meeker, W.Q., Escobar, L.A. and Fascual, F.G. (2022) Statistical Methods for Reliability Data. Wiley.

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

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

[6] Ng, C.K. (2013) Procedures for Selecting Good Exponential Populations. *Communications in Statistics-Simulation and Computation*, **42**, 1681-1692. https://doi.org/10.1080/03610918.2012.674598

[7] Lam, K. (1968) A New Procedure for Selecting Good Populations. *Biometrika*, **73**, 201-206. https://doi.org/10.1093/biomet/73.1.201

[8] 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

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

Appendix A

```
#Exact probability calculations
#CS=Correct Selection;ICS=Incorrect Selection;ESS=Expected Subset Size
#special case of slippage for R2 with k=2
#exponential populations with scale parameter = 1
rm(list=ls())
#specify sample size so that sample mean is approx. normal
#specify the slippage, delta
k=2;n=25;delta=0
#P is the P*-value for the min prob of correct selection
P=c(0.75,0.90,0.95,0.975,0.99)
b<-(sqrt(2/n))*qnorm(P)
df<-data.frame(P,b)
v<-(b+delta)/sqrt(2/n)
w<-(b-delta)/sqrt(2/n)
CS<-pnorm(v)
ICS<-pnorm(w)
ESS<-CS+ICS
df1<-data.frame(df,CS,ICS,ESS)
message("k = ",k,", n = ",n,", delta = ",delta)
round(df1,5)
```

Appendix B

```
#Matrix_means_SC
#Simulation of exponential distributions matrix format
#Selection Rule R2 (means) with slippage configuration (SC)
rm(list=ls())
set.seed(17)
#Input values of P* as P; number of simulations as N
#k=number of pops;n=sample size per pop
#Use delta=0 to obtain b-values for R2 given as quantiles
#at end of program and enter them in the 5 if statements below
k<-10;n<-25; delta<-0.3; P<-0.95
N<-200000; T<-rep(0,k); W<-rep(-1,N); avg<-rep(-1,k)
mean.exp<-rep(0,k); x<-rep(-1,n*k)
M<-matrix(x,ncol=k,nrow=n)
if (P==0.75){b<-0.4816}
if (P==0.90){b<-0.6274}
if (P==0.95){b<-0.7144}
if (P==0.975){b<-0.7906}
if (P==0.99){b<-0.8810}
for (j in 1:N){
for (i in 1:k){M[,i]<-rexp(n,rate=1)}
M
for (i in 1:k){mean.exp[i]<-mean(M[,i])}
mean.exp
S<-rep(0,k)
M[,k]<-M[,k]+delta
for (i in 1:k){avg[i]<-mean(M[,i])}
avg.max<-max(avg)
diff<-avg.max-avg
W[j]<-diff[k]
for (i in 1:k){
if (diff[i]<=b){S[i]<-1}
}
T<-T+S
}
message("k= ",k," n= ",n," delta= ",delta," P*= ",P,
" b= ",b," N= ",N)
CS<-T[k]/N
CS<-round(CS,4)
ICS<-sum(T[1:k-1])/N
ICS<-round(ICS,4)
ESS<-sum(T)/N
ESS<-round(ESS,4)
message("ICS =",ICS," ,Pr(CS) =",CS," ,ESS =",ESS)
#For use with delta=0 to determine b-values
length(W)
quan<-c(0.75,0.90,0.95,0.975,0.99)
round(quantile(W,quan),5)
```

Appendix C

```
#Matrix_mins_SC
#Simulation of exponential distributions matrix format
#Selection Rule R1 (mins) with slippage configuration
rm(list=ls())
set.seed(17)
#Input values of P* as P; number of simulations as N
#k=number of pops;n=sample size per pop
#Use delta=0 to obtain d-values for R1 given as quantiles
#at end of program and enter them in the 5 if statements below
k<-10;n<-25; delta<-0.3; P<-0.95
N<-200000; T<-rep(0,k); W<-rep(-1,N); mini<-rep(-1,k)
mini.exp<-rep(0,k); x<-rep(-1,n*k)
M<-matrix(x,ncol=k,nrow=n)
if (P==0.75){d<-0.1088}
if (P==0.90){d<-0.1500}
if (P==0.95){d<-0.1791}
if (P==0.975){d<-0.2074}
if (P==0.99){d<-0.2432}
for (j in 1:N){
for (i in 1:k){M[,i]<-rexp(n,rate=1)}
M
for (i in 1:k){mini.exp[i]<-min(M[,i])}
mini.exp
S<-rep(0,k)
M[,k]<-M[,k]+delta
for (i in 1:k){mini[i]<-min(M[,i])}
mini.max<-max(mini)
diff<-mini.max-mini
W[j]<-diff[k]
for (i in 1:k){
if (diff[i]<=d){S[i]<-1}
}
T<-T+S
}
message("k= ",k," n= ",n," delta= ",delta," P*= ",P,
" d= ",d," N= ",N)
CS<-T[k]/N
CS<-round(CS,4)
ICS<-sum(T[1:k-1])/N
ICS<-round(ICS,4)
ESS<-sum(T)/N
ESS<-round(ESS,4)
message("ICS =",ICS," ,Pr(CS) =",CS," ,ESS =",ESS)
#For use with delta=0 to determine d-values
length(W)
quan<-c(0.75,0.90,0.95,0.975,0.99)
round(quantile(W,quan),5)
```

Appendix D

```
#Matrix_means_ES
#Simulation of exponential distributions matrix format
#Selection Rule R2 (means) with equi-spaced configuration
rm(list=ls())
set.seed(17)
#Input values of P* as P; number of simulations as N
#k=number of pops;n=sample size per pop
#Use delta=0 to obtain b-values for R2
k<-10;n<-25; delta<-0.1; P<-0.95
N<-200000; T<-rep(0,k); W<-rep(-1,N); avg<-rep(-1,k)
mean.exp<-rep(0,k); x<-rep(-1,n*k)
M<-matrix(x,ncol=k,nrow=n)
#Enter proper b-values for k,n,P*,delta=0
if (P==0.75){b<-0.4816}
if (P==0.90){b<-0.6274}
if (P==0.95){b<-0.7144}
if (P==0.975){b<-0.7906}
if (P==0.99){b<-0.8810}
for (j in 1:N){
for (i in 1:k){M[,i]<-rexp(n,rate=1)}
M
for (i in 1:k){mean.exp[i]<-mean(M[,i])}
mean.exp
S<-rep(0,k)
for (i in 1:k){M[,i]<-M[,i]+(i-1)*delta}
for (i in 1:k){avg[i]<-mean(M[,i])}
avg.max<-max(avg)
diff<-avg.max-avg
W[j]<-diff[k]
for (i in 1:k){
    if (diff[i]<=b){S[i]<-1}
    }
T<-T+S
}
round(T/N,4)
ESS<-sum(T)/N
round(ESS,4)
#For use with delta=0 to determine b-values
length(W)
quan<-c(0.75,0.90,0.95,0.975,0.99)
round(quantile(W,quan),4)
```

Appendix E

```
#Simulation of exponential distributions matrix format
#Selection Rule R1 (mins) with equi-spaced configuration
rm(list=ls())
set.seed(17)
#Input values of P* as P; number of simulations as N
k<-10;n<-25; delta<-0.1; P<-0.95
N<-200000; T<-rep(0,k); W<-rep(-1,N); mini<-rep(-1,k)
min.exp<-rep(0,k); x<-rep(-1,n*k)
M<-matrix(x,ncol=k,nrow=n)
if (P==0.75){d<-0.1088}
if (P==0.90){d<-0.1500}
if (P==0.95){d<-0.1791}
if (P==0.975){d<-0.2074}
if (P==0.99){d<-0.2432}
for (j in 1:N){
for (i in 1:k){M[,i]<-rexp(n,rate=1)}
M
for (i in 1:k){min.exp[i]<-min(M[,i])}
min.exp
S<-rep(0,k)
for (i in 1:k){M[,i]<-M[,i]+(i-1)*delta}
for (i in 1:k){mini[i]<-min(M[,i])}
mini.max<-max(mini)
diff<-mini.max-mini
W[j]<-diff[k]
for (i in 1:k){
    if (diff[i]<=d){S[i]<-1}
    }
T<-T+S
}
round(T/N,4)
ESS<-sum(T)/N
round(ESS,4)
#For use with delta=0 to determine d-values
length(W)
quan<-c(0.75,0.90,0.95,0.975,0.99)
round(quantile(W,quan),4)
```

Appendix F

```
#plot of ESS for R1 and R2 slippage configuration
#k=25; n=10; P*=0.95
rm(list=ls())
delta<-c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0)
ESS.R2<-c(9.4934,9.4575,9.3269,9.0513,8.5699,7.8283,6.8168,5.6195,4.3817,3.2654,2.3802)
ESS.R1<-c(9.5022,9.0118,3.6245,1.2189,1.0180,1.0013,1.0001,1,1,1,1)
df<-data.frame(delta,ESS.R1,ESS.R2)
df
data=matrix(c(ESS.R1,ESS.R2),ncol=11,byrow=TRUE)
colnames(data)=c('0','0.1','0.2','0.3','0.4','0.5','0.6','0.7','0.8','0.9','1.0')
rownames(data)=c('ESS.mins','ESS.means')
#data
final=as.table(data)
final
barplot(final,beside=TRUE,col=c("red","blue"),xlab="delta",ylab="Expected Subset Size")
#main="ESS for R1 and R2\n Slippage Configuration\n k=10, n=25, P*=0.95,
#N=200,000",ylim=c(0,10))
legend("right",box.col="brown",bg="yellow",legend=c("mins","means"),fill=c("red","blue"))
```

Appendix G

```
#plot of ESS for R1 and R2 equi-spaced configuration
#k=25; n=10; P*=0.95
rm(list=ls())
delta<-c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0)
ESS.R2<-c(9.4935,6.5053,3.8512,2.8169,2.2692,1.9281,1.7006,1.5272,1.3870,1.2522,1.1529)
ESS.R1<-c(9.5022,2.2448,1.2987,1.0244,1.0020,1.0001,1.0000,1.0000,1.0000,1.0000,1.0000)
df<-data.frame(delta,ESS.R1,ESS.R2)
df
data=matrix(c(ESS.R1,ESS.R2),ncol=11,byrow=TRUE)
colnames(data)=c('0','0.1','0.2','0.3','0.4','0.5','0.6','0.7','0.8','0.9','1.0')
rownames(data)=c('ESS.mins','ESS.means')
#data
final=as.table(data)
final
barplot(final,beside=TRUE,col=c("red","blue"),xlab="delta",ylab="Expected Subset Size")
#main="ESS for R1 and R2\n Equi-Spaced Configuration\n k=10, n=25, P*=o.95,
#N=200,000, set.seed(17)",ylim=c(0,10))
legend("right",box.col="brown",bg="yellow",legend=c("mins","means"),fill=c("red","blue"))
```

Appendix H

```
#exp.sim.subset
#Subset selection for exponential distributions
#differening in threshold parameters
#k populations with samples of size n
#d and b must be determined for a given P*=P
rm(list=ls())
#Input the value of P* as P along with k and n
k<-10; n<-25; P<-0.75
if (P==0.75){d<-0.1088;b<-0.4816}
if (P==0.90){d<-0.1500;b<-0.6274}
if (P==0.95){d<-0.1791;b<-0.7144}
if (P==0.975){d<-0.2074;b<-0.7906}
if (P==0.99){d<-0.2432;b<-0.8810}
set.seed(15) #insure same simulated values on repeat
gamma<-seq(from=1,to=2.8,by=0.2)
M<-matrix(0,nrow=n,ncol=k)
for (i in 1:k){
M[,i]<-rexp(n,1)+gamma[i]
}
M
#####
y<-rep(0,k)
for (i in 1:k){
y[i]<-min(M[,i])
}
print(y)
max.y<-max(y)
s<-rep(0,k)
for (i in 1:k){
if (y[i]>=max.y-d){s[i]<-1}
}
print(y-max.y)
print(s)
#####
z<-rep(0,k)
for (i in 1:k){
z[i]<-mean(M[,i])
}
print(z)
max.z<-max(z)
t<-rep(0,k)
for (i in 1:k){
if (z[i]>=max.z-b){t[i]<-1}
}
print(z-max.z)
print(t)
#####
w<-seq(1:k)
df<-data.frame(w,y,z)
colnames(df)<-c("populations","minimums","means")
round(df,4)
#####
df1<-data.frame(s,t)
colnames(df1)<-c("minsel","meansel")
message("k = ",k," n = ",n," P* = ",P)
print('The selected populations denoted by 1')
df1
#####
```

Appendix I

```
#gamma.int.k
#OCs for R1 using gamma distribution for sum of exp rv's
#input k, n, b, and delta
rm(list=ls())
k=10; n=25; b=0.71559; delta=1
fun1<-function(x){
((pgamma(x+n*b+n*delta,shape=n,scale=1))^(k-1))*dgamma(x,shape=n,scale=1)
}
PrCS<-integrate(fun1,lower=0,upper=Inf)
PrCS
fun2<-function(x){
((pgamma(x+n*b,shape=n,scale=1))^(k-2))*
pgamma(x+n*b-n*delta,shape=n,scale=1)*
dgamma(x,shape=n,scale=1)
}
#Pr1 = Pr(choosing pop1) =,...,= Pr(choosing pop(k-1))
Pr1<-integrate(fun2,lower=0,upper=Inf)
Pr1
PrICS<-(k-1)*Pr1$value
ESS<-PrCS$value+PrICS
df<-data.frame(delta,PrCS$value,PrICS,ESS)
df<-round(df,4)
df
```

Appendix J

```
#P(CS) for k populations with slippage configuration
#Assuming sample means are normally distributed: CLT
rm(list=ls())
#input model values
#see "Selection Rules for Exponential Population Threshold
#Parameters, Sections 2.2 and 2.3, Means Rule R2
#Input value of P* as P
k<-10; delta<-0; n<-25; P<-0.95
if (P==0.75){b<-0.4528}
if (P==0.90){b<-0.5970}
if (P==0.95){b<-0.6836}
if (P==0.975){b<-0.7598}
if (P==0.99){b<-0.8500}
c<-(sqrt(n))*(b+delta)
int<-function(x){
((pnorm(x+c))^(k-1))*dnorm(x)
}
PCS<-integrate(int,lower = -Inf,upper = Inf)
message("k = ",k,", n = ",n,", delta = ",delta,", P* = ",P,
", b = ",b)
PCS
round(PCS$value,4)
u<-(sqrt(n))*b
v<-(sqrt(n))*(b-delta)
int1<-function(x){
((pnorm(x+u))^(k-2))*(pnorm(x+v))*dnorm(x)
}
P1<-integrate(int1,lower = -Inf,upper = Inf)
P1
round(P1$value,4)
#PICS is the probability of an incorrect selection
PICS<-(k-1)*P1$value
round(PICS,4)
#ESS is the expected subset size
ESS<-PCS$value+(k-1)*P1$value
message("The expected subset size is ",round(ESS,4))
```