Scientific Research Publishing

Applied Mathematics, 2025, 16(1), 1-14
https://www.scirp.org/journal/am
ISSN Online: 2152-7393
ISSN Print: 2152-7385

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) Selection Rules for Exponential Population Threshold Parameters. *Applied Mathematics*, **16**, 1-14.
https://doi.org/10.4236/am.2025.161001

Received: November 27, 2024
Accepted: January 12, 2025
Published: January 15, 2025

Abstract

This article constructs statistical selection procedures for exponential populations that may differ in only the threshold parameters. The scale parameters of the populations are assumed common and known. The independent samples drawn from the populations are taken to be of the same size. The best population is defined as the one associated with the largest threshold parameter. In case more than one population share the largest threshold, one of these is tagged at random and denoted the best. Two procedures are developed for choosing a subset of the populations having the property that the chosen subset contains the best population with a prescribed probability. One procedure is based on the sample minimum values drawn from the populations, and another is based on the sample means from the populations. An "Indifference Zone" (IZ) selection procedure is also developed based on the sample minimum values. The IZ procedure asserts that the population with the largest test statistic (e.g., the sample minimum) is the best population. With this approach, the sample size is chosen so as to guarantee that the probability of a correct selection is no less than a prescribed probability in the parameter region where the largest threshold is at least a prescribed amount larger than the remaining thresholds. Numerical examples are given, and the computer R-codes for all calculations are given in the Appendices.

Keywords

Weibull Distribution, Probability of Correct Selection, Minimum Statistic Selection Procedure, Means Selection Procedure, Subset Size, Indifference Zone Selection Rule, Least Favorable Configuration

1. Introduction

The Weibull distribution is one of the most widely utilized probability distributions

in applied reliability statistics. Its development spanned from 1922 to 1943, during which three separate groups worked independently toward different objectives. Among these researchers was Waloddi Weibull, whose name has since become synonymous with the distribution (see Rinne [1]). The distribution form that was introduced by Weibull in 1939, depends on three parameters. The cumulative distribution function, the density function and the hazard function or the failure rate of a three-parameter Weibull distribution for $t \geq \gamma$ are given as:

$$F(t \mid \gamma, \eta, \beta) = 1 - \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right] \tag{1.1}$$

$$f(t \mid \gamma, \eta, \beta) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right] \tag{1.2}$$

$$H(t; \gamma, \eta, \beta) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta}, \tag{1.3}$$

where:

- γ is the location parameter, also called the threshold or shift parameter.
- $\eta > 0$ is the scale parameter.
- $\beta > 0$ is the shape parameter.

The mean and the variance of a Weibull random variable are respectively,

$$E(T) = \eta\Gamma\left(\frac{1}{\beta}+1\right) + \gamma \tag{1.4}$$

$$V(T) = \eta^2\left[\Gamma\left(\frac{2}{\beta}+1\right) - \Gamma^2\left(\frac{1}{\beta}+1\right)\right], \tag{1.5}$$

where $\Gamma(z)$ is the gamma function defined as,

$$\Gamma(z) = \int_0^{\infty} t^{z-1}\mathrm{e}^{-t}\mathrm{d}t. \tag{1.6}$$

Weibull percentiles are important in many applications, as they represent the time at which a certain percentage of the population is expected to fail. When γ = 0 the scale parameter η is the 63.2% percentile of the Weibull distribution, *i.e.*, the time at which 63.2% of the population will fail. The percentile function is

$$t_p = \gamma + \eta\left[-\log(1-p)\right]^{1/\beta}. \tag{1.7}$$

The most used Weibull model is the two-parameter model, when the threshold parameter is not present, *i.e.* $\gamma = 0$. So the functions that characterize this form of the distribution are, for $t \geq 0$,

$$F(t \mid \eta, \beta) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \tag{1.8}$$

$$f(t \mid \gamma, \eta, \beta) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \tag{1.9}$$

$$H(t;\gamma,\eta,\beta)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta}. \tag{1.10}$$

The mean in this case is $E(T)=\eta\Gamma\left(\frac{1}{\beta}+1\right)$, and the variance remains the same as in the three-parameter case, since it does not depend on the threshold parameter.

The practical significance of the Weibull distribution lies in its versatility to model failure patterns across various commonly observed shapes. When $0<\beta<1$, the Weibull distribution exhibits a decreasing hazard function. Such a model is applied in electronic testing for components that have a high chance of early failure. For example, testing microchips or circuit boards where early failure often occurs due to minor defects introduced in manufacturing (Meeker, *et al.* [2]). At $\beta=1$, the hazard function remains constant, representing a failure rate that does not change over time. For this value of the shape parameter, Weibull reduces to the exponential distribution, for $t\geq\gamma$,

$$f(t\,|\,\gamma,\eta)=\frac{1}{\eta}\exp\left[-\left(\frac{t-\gamma}{\eta}\right)\right] \tag{1.11}$$

and when $\gamma=0$, for $t\geq 0$ the more familiar form

$$f(t\,|\,\eta)=\frac{1}{\eta}\exp\left(-\frac{t}{\eta}\right). \tag{1.12}$$

This model is useful for items where the probability of failure does not depend on age (memoryless property). Some examples of such items are, light bulbs, electronic resistors, or other items where failure events are random and independent of how long they've been in use.

For $\beta>1$ the hazard function is increasing. When $1<\beta<2$ the failure rate is moderately increasing and such a model is applied to fatigue testing in mechanical engineering. Components subject to moderate stress, like springs, bearings, or light-duty machinery parts, may experience a gradually increasing failure rate as minor wear gradually contributes to eventual failure (Meeker *et al.* [2]).

When $\beta=2$, a special case of the Weibull distribution, known as the Rayleigh distribution, has a linearly increasing failure rate. This model is applied in meteorology, for wind speed modeling. An application might be calculating probabilities of extreme weather events like strong gusts, which are more likely with increasing wind speed.

For $\beta>2$ the distribution has a rapidly increasing failure rate. It models aging or wear-out phase where failure rate accelerates as the item ages, indicating that the likelihood of failure increases significantly over time. This is common for high-stress mechanical systems that are subject to significant wear such as engines, turbines, or heavy equipment (Meeker *et al.* [2]). For the case when β is very large, the Weibull distribution models sudden wear out. It approximates a deterministic lifespan where failure occurs almost predictably after a certain

period. This is useful in setting maintenance schedules for parts that must be replaced after a defined usage time to avoid sudden failure.

The extensive applicability of the Weibull distribution in reliability analysis underscores its importance, often making it the preferred choice over other distributions. A notable example is the study by McDonald, *et al.* [3], which demonstrates the advantages of using the Weibull distribution over the lognormal distribution in analyzing car emission data, highlighting its superior fit and interpretative value in that context.

The goal of this paper is to derive selection rules for the special case of the Weibull distribution when $\beta = 1$. This is the exponential distribution with location and scale parameters (γ, η). In an early paper, An Hsu [4] gives some procedures for Weibull populations. The author constructs optimal selection procedures in order to select a subset of the k populations containing the best population. The procedures control the size of the selected subset and also maximize the minimum probability of a correct selection. The author considers the cases when the shape parameter β is common and known among the k populations and the sample sizes are not necessarily equal. The selection procedures are based on the scale parameter η, where a population is considered best if it has the largest scale parameter. Hence selecting the best population means selecting the population with the largest scale η. So, the main two problems of An Hsu's paper are, to maximize the probability of correct selection and to minimize the subset size.

Two excellent and comprehensive books on ranking and selection procedures are authored by Gupta and Panchapkesan [5] and Gibbons, *et al.* [6].

2. Selection Rules for the Exponential Threshold Model ($\beta = 1$, $\eta = 1$)

Let $\Pi_1, \Pi_2, \cdots, \Pi_k$ be k three-parameter Weibull populations with common, known, shape and scale parameters, ($\beta = 1$, $\eta = 1$) and threshold parameters γ_i, for $i = 1, 2, \cdots, k$. Let T_i be a random variable associated with population Π_i, for $i = 1, 2, \cdots, k$, then T_i follows the distribution

$$F(t \mid \gamma_i) = 1 - \exp\left[-(t - \gamma_i)\right], \quad t \geq \gamma_i. \tag{2.1}$$

Two approaches to selection rules will be developed. The first approach considers subset selection, where the goal is to select a subset of the k populations that contains the best population with a specified probability P^*. The second, involves the indifference zone, which focuses on selecting the single best population among the k, again with a specified probability P^*. In both these scenarios, the best population is the one having the largest(the smallest) threshold parameter, depending on the context. In the exponential model considered here, the threshold parameter is additive in the expression for the mean and for any percentile. Thus, selection for the largest threshold is equivalent to selection for the largest mean or for any of the percentiles.

2.1. Subset Selection for Populations to Contain the One with the Largest Threshold

The subset selection approaches developed herein follow the seminal work by Gupta [7]. Let T_{ij} be n independent identically distributed(iid) random variables from population Π_i, for $i=1,2,\cdots,k$ and $j=1,2,\cdots,n$. Let $\gamma_{[1]} \le \gamma_{[2]} \le \cdots \le \gamma_{[k]}$ be the ordered thresholds, where $\gamma_{[k]}$ corresponds to the best population. Since $t \ge \gamma$, then the minimum order statistic is a typical estimator for γ. So, let $Y_{(i)}$ be the minimum of the sample from the population with threshold parameter $\gamma_{[i]}$,

$$Y_{(i)} = \min_{1\le j\le n}\left(X_{ij}\right), \quad i=1,2,\cdots,k. \tag{2.2}$$

The selection rule is then,

$$R_1: \text{Select } \Pi_i \text{ iff } Y_i \ge \max_{1\le j\le k}\left(Y_j\right) - d, \quad d \ge 0. \tag{2.3}$$

The distribution of the minimum order statistic for independent random variables $X_i, i=1,2,\cdots,n$ following (2.1) can easily be derived.

If $Y=\min\left(X_1, X_2, \cdots, X_n\right)$, then

$$\begin{aligned}
\Pr\left(Y \le y\right) &= 1-\Pr\left(Y > y\right)\\
&= 1-\Pr\left(X_i > y, i=1,2,\cdots,n\right)\\
&= 1-\prod_{i=1}^{n}\Pr\left(X_i > y\right) \quad \text{by independence}\\
&= 1-\prod_{i=1}^{n}\left[1-\Pr\left(X_i \le y\right)\right]\\
&= 1-\prod_{i=1}^{n}\left\{1-\left[1-\mathrm{e}^{-(y-\gamma)}\right]\right\}\\
&= 1-\mathrm{e}^{-n(y-\gamma)}, \quad y \ge \gamma.
\end{aligned} \tag{2.4}$$

From here, the pdf of Y is $f_Y(y) = n\mathrm{e}^{-n(y-\gamma)}$, $y \ge \gamma$. The probability of a (CS) now follows:

$$\begin{aligned}
\Pr(\mathrm{CS}) &= \Pr\left(Y_{(k)} \ge \max_{1\le j\le k-1}\left(Y_{(j)}\right) - d\right)\\
&= \Pr\left(\max_{1\le j\le k-1}\left(Y_{(j)}\right) \le Y_{(k)} + d\right)\\
&= \int_{\gamma_{[k]}}^{\infty}\prod_{j=i}^{k-1}\Pr\left(Y_{(j)} \le y+d\right)\mathrm{d}F_{Y_{(k)}}(y)\\
&= \int_{\gamma_{[k]}}^{\infty}\prod_{j=i}^{k-1}\Pr\left(Y_{(j)} \le y+d\right)\cdot n\cdot \mathrm{e}^{-n\left(y-\gamma_{[k]}\right)}\mathrm{d}y.
\end{aligned}$$

Now, let $x = y-\gamma_{[k]}$, and so $y = x+\gamma_{[k]}$. Then since $\gamma_{[k]} - \gamma_{[j]} \ge 0$, $j=1,2,\cdots,k-1$

$$\begin{aligned}
\Pr(\mathrm{CS}) &= n\int_0^{\infty}\prod_{j=i}^{k-1}\Pr\left(Y_{(j)} - \gamma_{[j]} \le x + \gamma_{[k]} - \gamma_{[j]} + d\right)\mathrm{e}^{-nx}\mathrm{d}x\\
&\ge n\int_0^{\infty}\prod_{j=i}^{k-1}\Pr\left(Y_{(j)} - \gamma_{[j]} \le x + d\right)\mathrm{e}^{-nx}\mathrm{d}x.
\end{aligned} \tag{2.5}$$

The parameter configuration $\gamma_1 = \gamma_2 = \cdots = \gamma_k$ is referred to as the Least Favorable Configuration (LFC) since it yields the minimum value of $\Pr(\mathrm{CS})$. Setting the expression in (2.5) equal to a specified value P^* ($k^{-1} < P^* < 1$), the value of d can now be determined. Now the selection rule R_1 will ensure that $\Pr(\mathrm{CS}) \geq P^*$ no matter the configuration of $\gamma_i's$.

Common and Known Scale Parameter $\eta \neq 1$

We assume that the scale parameter is common for all the populations Π_i for $i = 1, 2, \cdots, k$ and it is known, but it's not 1. Does this change the selection rule for subset selection? Let $T \sim \exp(\eta, \gamma)$. The random variable $\frac{T}{\eta}$ is still exponentially distributed with scale parameter $\eta = 1$,

$$\begin{aligned} \Pr\left(\frac{T}{\eta} \leq t\right) &= \Pr(T \leq \eta t) \\ &= 1 - \exp\left[-\frac{\eta t - \gamma}{\eta}\right] \\ &= 1 - \exp\left[-\left(t - \frac{\gamma}{\eta}\right)\right] \\ &= 1 - \exp\left[-(t - \gamma')\right]. \end{aligned} \tag{2.6}$$

Or otherwise,

$$F\left(\frac{t}{\eta} \mid \gamma\right) = 1 - \exp\left[-(t - \gamma')\right], \quad t \geq \gamma', \quad \gamma' = \frac{\gamma}{\eta}. \tag{2.7}$$

Hence, we get back to the previous case where scale parameter $\eta = 1$ and the same selection rule, R_1, can be applied.

2.2. Selection Based on Sample Means

If $\beta = 1$, then $E(T) = \eta\Gamma(2) + \gamma = \eta + \gamma$ and $V(T) = \eta^2\left[\Gamma(3) - \Gamma^2(2)\right] = \eta^2$. Suppose $\Pi_i, i = 1, 2, \cdots, k$ are k independent populations with Π_i distributed as Weibull $(\gamma_i, \eta, \beta = 1)$, and η is a common value known. Let X_{ij} be an independent random sample from Π_i, $j = 1, 2, \cdots, n$. Our goal is to select a subset of the populations such that the population having the largest γ value is contained in the subset with a specified probability P^* ($k^{-1} < P^* < 1$). Denote the ordered γ-values by

$$\gamma_{[1]} \leq \gamma_{[2]} \leq \cdots \leq \gamma_{[k]}.$$

Since γ is a threshold, as covered in Section 2 it is reasonable to consider a selection rule based on the minimum sample values from the population, such as R_1 given in 0.15.

Since the population means are ordered as the population γ-values, it is also reasonable to consider a selection rule based on the population sample means, $\bar{X}_i$. That is,

$$R_2: \text{Choose } \Pi_i \text{ iff } \bar{X}_i \geq \max_{1 \leq j \leq k}\left(\bar{X}_j\right) - b, \tag{2.8}$$

where $\bar{X}_i = \left(\sum_{j=1}^{n} X_{ij}\right)\Big/ n$ and $b = b\left(k, n, P^*\right)$ is a nonnegative number chosen to satisfy the P^* condition,

$$\min_{\Omega} \Pr(\mathrm{CS}) \geq P^*, \tag{2.9}$$

where $\Omega = \left(\gamma_1, \gamma_2, \cdots, \gamma_k\right)$.

Let $\bar{X}_{(i)}$ be the sample mean drawn from the population possessing the mean value $\eta + \gamma_{[k]}$. Using the Central Limit Theorem, the distribution of $\bar{X}_{(i)}$ is approximately normal with mean $\eta + \gamma_{[k]}$ and variance $V(T)/n = \eta^2/n$. So for large n,

$$\begin{aligned} \Pr(\mathrm{CS}) &= \Pr\left(\bar{X}_{(k)} \geq \max_{1 \leq j \leq k-1} \bar{X}_{(j)} - b\right) \\ &= \Pr\left(Z_k \geq Z_j + \left(\gamma_{[j]} - \gamma_{[k]}\right) - \left(\sqrt{n}/\eta\right)b,\ j = 1, 2, \cdots, k-1\right) \\ &\geq \int_{-\infty}^{\infty} \Phi^{k-1}\left(x + cb\right)\phi(x)\mathrm{d}x, \end{aligned} \tag{2.10}$$

where Z_i, $i = 1, 2, \cdots, k$, are independent standardized normal random variables and $c = \sqrt{n}/\eta$. Thus, $b = b\left(k, n, P^*\right)$ is determined by setting Equation (2.10) equal to P^* and solving for b.

For small values of n, the selection procedure can be based on the sum of the random variables, $n\bar{X}_i$, $i = 1, 2, \cdots, k$, and utilizing the property that a sum of n independent exponential random variables follows the gamma distribution with parameters (n, η) assuming the threshold parameter, γ, is equal to 0 (Casella and Berger [8]).

To assess the sensitivity of the subsets selected using selection rules R_1 and R_2 to the assumption of a common known scale parameter, η, set a bound on possible values of the parameter. That is, say the assessment of the analyst is $L < \eta < U$. Now discretize the interval (L, U) into m discrete values, *i.e.*, $L = v_1 < v_2 < \cdots < v_m = U$. Follow the procedures in Sections (2.1.1) and (2.2) with η assumed to be equal to v_i to obtain the selected subsets. Repeat the process using the remaining values of $\mathbf{v} = (v_1, \cdots, v_m)$. Compare the m selected subsets and assess the sensitivity to the selections based on the uncertainty of the scale parameter. To assess the sensitivity to the assumption that the shape parameter, β, is equal to one is not possible since the LFC has not been determined for values of β not equal to one.

2.3. Application of the Two Selection Rules

In this section, an example for each of the selection rules that were developed previously, rules (2.3) and (2.8) are given. The R-code for these rules is given in the Appendix section.

2.3.1. Application of the Minimum Order Statistics Selection Rule

Using the selection rule for the largest threshold based on minimum order statistics from the k populations, the d-values for different levels of the probability P^* are computed.

The example will be for $k = 10$ populations. From each population, a sample

of size $n = 25$ is drawn, with an equi-spaced configuration for the threshold parameters, *i.e.*, $\gamma_i's = 1 \to 10$, by 1. Simulations are set for $N = 10000$. The d-values obtained are given in Table 1. The R-code for generating the d-values is found in Appendix A. The P^* values based on simulations can be checked using the R-code in Appendix D.

Table 1. P^* and d-values.

P^*	0.75	0.90	0.95	0.975	0.99
d	0.1083	0.1501	0.1785	0.2077	0.2449

Using the R-code in Appendix B random samples of size 25 are generated from 10 exponential populations having γ-values equal to 1 up to 10 with step size 1. Table 2 gives the minimum and mean values of these samples, respectively. The populations selected using rule R_1 are given in Table 4 for the five values of P^* given in Table 1.

Table 2. Minimums and means for each of the k populations.

Population	Minimum	Mean
1	0.0100	0.8842
2	0.1626	2.3289
3	0.4438	3.2933
4	0.3080	4.1323
5	0.0774	4.5331
6	0.0683	5.5062
7	0.2490	6.2474
8	0.5265	9.1137
9	0.2727	6.7970
10	0.3481	9.6343

2.3.2. Application of the Sample Means Rule

For the selection rule based on the sample means, the b-value is calculated using the R-code in Appendix C, which for $k = 10$, $n = 25$ is given in Table 3:

Table 3. P^* and b-values.

P^*	0.75	0.90	0.95	0.975	0.99
b	0.4528	0.5970	0.6836	0.7598	0.8500

The selected populations using R_2 with the example simulated exponential dataset, $\gamma = 1 \to 10$, step size 1, are given in Table 4. The sample means procedure yields subsets with smaller size in most of the P^* values. For this particular simulation R_2 would be preferred over R_1 since the selected subset size is

favorable. Another similarly simulated data set could well result in R_1 being preferred over R_2. Simulation studies, yet to be published, show that the expected subset size of the selected populations is smaller for R_1 than for R_2 with equispaced threshold parameters (or slippage parameter) greater than 0.

Table 4. P^* and the selected populations using rules R_1 and R_2.

P^*	**Min Procedure**, R_1	**Means Procedure**, R_2
0.75	3, 8	10
0.90	3, 8	8, 10
0.95	3, 8, 10	8, 10
0.975	3, 8, 10	8, 10
0.99	3, 4, 8, 10	8, 10

2.4. Indifference Zone Approach

Following the Indifference Zone approach formulated by Bechhofer [9], the population yielding the largest minimum order statistics, $Y_{[k]}$ would be asserted to be the "best" population. The statistical goal is to have the probability of this assertion to be correct with a specified probability, P^*, over the parameter space $\gamma_{[k]} - \gamma_{[k-1]} \geq d^*$, where d^* is a user specified meaningful difference between the best population and all the remainder ones. The minimum of the sample size $n = n\left(k, P^*, d^*\right)$ needs to be determined so that the $\Pr(\mathrm{CS}) \geq P^*$ for the parameter configuration $\gamma_{[k]} - \gamma_{[k-1]} \geq d^*$. Other parameter configurations comprise the "indifference zone" in the sense that the $\Pr(\mathrm{CS})$ is not necessarily applicable.

For this decision rule,

$$\begin{aligned}
\Pr(\mathrm{CS}) &= \Pr\left(Y_{(k)} \geq \max Y_{(j)},\ j = 1, 2, \cdots, k-1\right) \\
&= \Pr\left(Y_{(j)} \leq Y_{(k)},\ j = 1, 2, \cdots, k-1\right) \\
&= \Pr\left(Y_{(j)} - \gamma_{[j]} \leq Y_{(k)} - \gamma_{[k]} + \gamma_{[k]} - \gamma_{[j]},\ j = 1, 2, \cdots, k-1\right) \\
&\geq \Pr\left(Y_{(j)} - \gamma_{[j]} \leq Y_{(k)} - \gamma_{[k]} + d^*,\ j = 1, 2, \cdots, k-1\right) \\
&= n\int_0^{\infty} \left[1 - \mathrm{e}^{-n\left(x + d^*\right)}\right]^{k-1} \mathrm{e}^{-nx} \mathrm{d}x.
\end{aligned} \tag{2.11}$$

The LFC is $\gamma_{[1]} = \cdots = \gamma_{[k-1]} = \gamma_{[k]} - d^*$. The above integral expression is calculated with the R-code found in Appendix 5 and through iteration we can determine the minimum sample size required to meet the $\Pr(\mathrm{CS})$ requirement.

For example, for $k = 5$ populations and requirement $P^* = 0.90$ with $d^* = 0.25$, Table 5 gives evaluations of (2.11). Thus, a sample size of 12 would be the minimum size requirement to meet the experimental goal.

3. Summary and Conclusions

The importance of the Weibull distribution sparked our interest to investigate

Table 5. Sample size n and the P^* requirement.

n	P^*
10	0.8488
11	0.8801
12	0.9053
13	0.9254
14	0.9414

procedures for selecting the best among several Weibull populations. In this paper, the case when the shape parameter is common, known and $\beta = 1$ is considered. Under this consideration, the distribution reduces to the exponential distribution with scale and threshold parameters (η, γ). Further, the scale parameter η is assumed common and known.

Two different approaches, subset selection and indifference zone, are developed. For the subset selection approach, two different procedures, one based on the minimum order statistic and the other based on the sample means are considered. The two selection rules are given by Equations (2.3) and (2.8). Using the R-codes provided in the appendices, there is little difference in the computational requirements for implementation of the selection rules R_1 and R_2.

In the work presented here, the sample sizes for the k populations are equal. As is often the case in practice, samples are not equal. In these cases, Gibbons, *et al.* [6], Section (4), suggest using a generalized average sample size for each of the populations. In particular, they recommend using the square-mean-root, N_0, given by

$$N_0 = \left(\left(\sum_{i=1}^{k} \sqrt{n_i} \right) \Big/ k \right)^2. \tag{3.1}$$

Note that the square root of N_0 is the arithmetic mean of the square root of the actual sample sizes. Gibbons, *et al.* [6] remark that N_0 is always greater than the geometric mean of the actual sample sizes, and always smaller than the arithmetic mean. In practice, N_0 is rarely very different from the arithmetic mean; however (3.1) gives more accurate results most of the time and is therefore preferable to the arithmetic mean.

These theoretical results are illustrated with simulated data to demonstrate the selection procedures. Using the R-codes in Appendices A-C, the values of d and b given the number of populations k, the sample size n and the probability of a correct selection P^*, can be determined. Once these values were determined, selection rules (2.3) and (2.8) determined the subset that contains the best population. The size of the selected subset is a nondecreasing function of $\Pr(\mathrm{CS})$. For our example, the sample means procedure R_2, seems to be superior to the minimum order statistic procedure, R_1, because it selected a subset of smaller size, for almost all the values of P^*. However, this is just one illustrative example. The conditions under which one of the two procedures is guaranteed better, in the

sense of yielding a subset size no larger than the other, have yet to be determined.

For the indifference zone approach formulated by Bechhofer [9], our procedure was based on the minimum order statistics. In this case, the probability of a correct selection is specified over the parameter space determined by $\gamma_{[k]} - \gamma_{[k-1]} \geq d^*$, where d^* is user specified. Here the minimum sample size n is determined so that the probability of a correct selection is at least P^*. This is given by Equation (2.11). An example is provided for this approach. The R-code for its implementation can be found in Appendix D. For a given number of populations $k = 5$, $d^* = 0.25$ and $P^* = 0.90$ a minimum sample size of $n = 12$ is required in order to achieve the experimental goal.

This article naturally leads to further investigations about selection procedures. What are the conditions under which R_1 is preferred to R_2? One can also be interested in developing procedures when the shape parameter is common and known, but not equal to 1, or when the scale parameter is common and unknown and the sample sizes are not necessarily equal.

Conflicts of Interest

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

References

[1] Rinne, H. (2009) The Weibull Distribution, a Handbook. Chapman & Hall/CRC.

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

[3] McDonald, G.C. and Vance, L.C. and Gibbons, D.I. (1995) Some Tests for Dis-criminating Between Lognormal and Weibull Distributions: An Application to Emissions Data. In: Balakrishnan, N., Ed., *Recent Advances in Life-Testing and Reliability*, CRC Press, 475-490. https://doi.org/10.1201/9781003418313

[4] An Hsu, T. (1982) On Some Optimal Selection Procedures for Weibull Populations. *Communications in Statistics—Theory and Methods*, **11**, 2657-2668. https://doi.org/10.1080/03610928208828413

[5] Gupta, S.S. and Panchapakesan, S. (1979) Multiple Decision Procedures. John Wiley & Sons.

[6] Gibbons, J.D., Olkin, I. and Sobel, M. (1999) Selecting and Ordering Populations: A New Statistical Methodology. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971101

[7] Gupta, S.S. (1965) On Some Multiple Decision (Selection and Ranking) Rules. *Technometrics*, **7**, 225-245. https://doi.org/10.2307/1266672

[8] Casella, G. and Berger, R.L. (2002) Statistical Inference. Duxbury.

[9] 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

A. Appendix

```
#exp.sim.10
#set for k = 10 populations; n samples; N simulations
k=10;n=25;N=10000
set.seed(1) #repeat same numbers for each program simulation
Y1<-rep(0,N);Y2<-rep(0,N);Y3<-rep(0,N);Y4<-rep(0,N)
Y5<-rep(0,N);Y6<-rep(0,N);Y7<-rep(0,N);Y8<-rep(0,N)
Y9<-rep(0,N);Y10<-rep(0,N);S<-rep(0,N)
for (i in 1:N){
  X1<-rexp(n);X2<-rexp(n);X3<-rexp(n);X4<-rexp(n);X5<-rexp(n)
  X6<-rexp(n);X7<-rexp(n);X8<-rexp(n);X9<-rexp(n);X10<-rexp(n)
  Y1[i]<-min(X1);Y2[i]<-min(X2);Y3[i]<-min(X3);Y4[i]<-min(X4)
  Y5[i]<-min(X5);Y6[i]<-min(X6);Y7[i]<-min(X7);Y8[i]<-min(X8)
  Y9[i]<-min(X9);Y10[i]<-min(X10)
  S[i]<-max(Y1[i],Y2[i],Y3[i],Y4[i],Y5[i],Y6[i],
            Y7[i],Y8[i],Y9[i],Y10[i])-Y10[i]
}
length(S)
summary(S)
df<-data.frame(Y1,Y2,Y3,Y4,Y5,Y6,Y7,Y8,Y9,Y10,S)
head(df,10)
SS<-round(sort(S),4)
head(SS,20)
CUSUM<-cumsum(SS)
df1<-data.frame(SS,CUSUM)
tail(df1,20)
df2<-data.frame(table(SS))
#df2
df3<-data.frame(df2,cumsum(df2$Freq))
df3
Freq<-df3$Freq
CDF<-df3$cumsum.df2.Freq./N
df4<-data.frame(df2,CDF)
#tail(df4,200)
plot(df4$SS,df4$CDF)
quant<-c(0.75,0.90,0.95,0.975,0.990)
quantile(SS,quant)
```

B. Appendix

```
#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<-0.1083;b<-0.4528}
if (P==0.90){d<-0.1501;b<-0.5970}
if (P==0.95){d<-0.1785;b<-0.6838}
if (P==0.975){d<-0.2077;b<-0.7598}
if (P==0.99){d<-0.2449;b<-0.8500}
set.seed(3) #insure same simulated values on repeat
gamma<-seq(from=1,to=10,by=1)
lambda<-1/gamma
M<-matrix(0,nrow=n,ncol=k)
for (i in 1:k){
  M[,i]<-rexp(n,lambda[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
#####
```

C. Appendix

```
#Aymptotic.Integral
rm(list=ls())
k<-10;n<-25;A<-4.25
int<-function(x){
  ((pnorm(x+A))^(k-1))*dnorm(x)
}
ans<-integrate(int,lower = -Inf,upper = Inf)
ans
ans$value
c<-(sqrt(n))
b<-A/c
b
```

D. Appendix

```
#P(CS)expon.pops.given d
#P* with given d for k exponential populations
#Selection based on min order stats with n samples
n<-25;k<-10;d<-0.1083
exp.sub<-function(x){
  n*((1-exp(-n*(x+d)))^(k-1))*exp(-(n*x))
}
ans<-integrate(exp.sub,lower = 0,upper = Inf)
ans
ans$value
```