MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 42 (1): 32-41 (February 2023) https://mjs.um.edu.my

# TWO-SAMPLE TEST FOR RANDOMLY CENSORED DATA

Ayushee[1a], Narinder Kumar[2a]* and Manish Goyal[3b]

**Abstract:** In this study, a nonparametric test was proposed for the two-sample scale problem, when sample observations are randomly right censored. The proposed test was based on the extremes of observations as an extension of the widely used Gehan's test for the two-sample problem. Critical values were obtained through simulations of various lifetime distributions at various sample sizes. Power performance for the proposed test was investigated considering various distributions. Upon comparing with the Gehan's test, it was found that the proposed test has more statistical power and efficiency for some special cases. An empirical experiment with a real-life data set was also presented.

***Keywords:*** *Nonparametric test, two-sample scale, right censored, critical values*

## 1. Introduction

In statistical analysis, nonparametric approaches do not require any assumptions regarding the distribution of a population. Two-sample nonparametric tests are employed to compare the distribution of two samples. Two-sample scale problems arise when the analyzer is interested in determining whether the populations follow the same distribution, or when there is a difference in their scale parameters. This issue has numerous applications in the field of Agriculture, Engineering, Business, Trade, Industries, and Medicine. For the two-sample scale problem, nonparametric tests have been proposed by Mood (1954), Sukhatme (1957), Kössler (1994), Kössler & Kumar (2010), and Goyal & Kumar (2020).

In real life, there are certain situations, where we do not have the complete information about the data, there involves the role of censoring, these cases are of much practical use. We say that an observation is censored when we do not observe it completely. Censored observations can be statistically-treated in various forms, ranging from parametric to nonparametric approaches. Several nonparametric tests are also available for censored data. Kaplan & Meier's (1958) method is marked as a great finding in the field of survival analysis, especially from the perspective of nonparametric approaches. This impelled the advancement of existing nonparametric approaches in the presence of censored data. Some two-sample nonparametric tests with censored data are discussed hereafter.

**Authors information:**
[a]Department of Statistics, Panjab University, Sector 14, Chandigarh, INDIA. Email: ayusheechoudhary98@gmail.com[1]; nkumar@pu.ac.in[2]
[b]Department of Statistics, Post Graduate Government College, Sector 11, Chandigarh, INDIA. Email: manishgoyal33@gmail.com[3]
*Corresponding Author: nkumar@pu.ac.in

In the context of industrial life-testing, Halperin (1960) considered a special case of that by Wilcoxon (1945), which involved statistics for comparing two samples in the presence of type-I censoring. A rank order theory for the two-sample problem was developed by Rao, Savage & Sobel (1960) when the data were censored. To arrive at an early decision, a sequential modification to Wilcoxon's test was proposed by Alling (1963). Furthermore, for comparing two samples in the presence of random censoring, Gehan (1965) proposed a generalized form of Wilcoxon's test, conditioned on the observational pattern. Efron (1967) proposed a two-sample problem with censored data as an extension of Gehan's method. Mantel (1967) proposed an approach to simplify both the method of computation and determination of the permutation distribution of Gehan's statistic. Lee, Desu & Gehan (1975) presented a Monte-Carlo study on a series of two-sample tests with or without censoring. For an in-depth literature review, one can refer to the monographs: Survival analysis by Miller (2011) and Lifetime Data: Statistical Models and Methods by Deshpande & Purohit (2015).

The statistical problem we have considered in this study mostly arises in the field of medicine, wherein we compare two treatments for their effects on patients' health and life, where the observations under study are the lifetimes of patients. A common problem in clinical trials arises when the data is not observed completely, or we have partial information about it; we consider such an observation to be censored. We considered random censoring, as it is mostly used in clinical trials due to the failure to follow-up or termination of the study.

Received: December 23, 2021
Accepted: February 15, 2022
Published: February 28, 2023



DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

The proposed distribution-free two-sample test is based on the extension of Gehan's test statistics, whereby for each individual, the observation is either time to censoring or time to failure. From the point when the study was initiated, an observation was noted as time to failure if the patient was found to be dead or relapsed before a pre-fixed time T. Moreover, it was noted as time to censored if the patient was alive until time T, or in remission at time T. In other words, for two different treatments if $n_1$ ($n_2$) is the total patients that participated in the study, out of which $r_1$ ($r_2$) are censored at time T, then $n_1$-$r_1$ ($n_2$-$r_2$) individuals have failed. Two-sample scale problem involves comparing the survival of these $n_1$ and $n_2$ patients. The objective of the study is to propose a new test that has more efficiency and power with respect to Gehan's test for the two-sample scale problem.

The remainder of this paper is structured as follows. Section 2 defines a newly proposed test statistic. The mean and variance of the test statistic are evaluated in Section 3. Critical points of the test statistic at various sample sizes and percentage censoring are given in Section 4, along with a comparison of critical points of Gehan's test statistic. The asymptotic relative efficiency of the test statistic is derived in Section 5, and a real-life data example for the statistic is illustrated in Section 6. The statistical power of the proposed test is given in comparison to the statistical power of Gehan's test at various sample sizes and percentage censoring in Section 7.

## 2. The Proposed Test Statistic

Let us suppose that we have two samples, $X$ and $Y$, with $n_1, n_2$ individuals, randomly allocated to two treatments, $A$ and $B$, respectively. Suppose that an experiment was conducted for a fixed time $T$, and all the individuals were followed up. If $x_i, y_j$ represents the time to failure and $x_i', y_j'$ represents the time to censoring for all $(i = 1, 2, \dots, n_1$ and $j = 1, 2, \dots, n_2)$, we have the following observations:

$$\left.\begin{matrix} x_1', x_2', \dots, x_{r_1}', \\ x_{r_1+1}, x_{r_1+2}, \dots, x_{n_1}, \end{matrix}\right\} \begin{matrix} r_1 \text{ censored} \\ n_1 - r_1 \text{ failures} \end{matrix} \quad \text{treatment } A,$$

$$\left.\begin{matrix} y_1', y_2', \dots, y_{r_2}', \\ y_{r_2+1}, y_{r_2+2}, \dots, y_{n_2}, \end{matrix}\right\} \begin{matrix} r_2 \text{ censored} \\ n_2 - r_2 \text{ failures} \end{matrix} \quad \text{treatment } B.$$

Furthermore, the cumulative distribution functions of time to failure $x_i, y_j$ are $F_1(x), F_2(y)$ and that of the time to censoring $x_i', y_j'$ are $G_1(x), G_2(y)$.

The null hypothesis is:

$H_0: F_1(t) = F_2(t), \quad (t \le T)$ (Treatments $A$ and $B$ are equally effective),

against the alternative

$H_1: F_1(t) = F_2(\theta t), \quad (t \le T \ \ \& \ \ \theta \ne 1)$

(Treatments $A$ and $B$ are significantly different.)

In the proposed test, we have taken a sub-sample of size two from each sample, and compared their extremes to derive more information from the samples. Let $(x_1, x_2)$ and $(y_1, y_2)$ be the uncensored sub-samples chosen from samples $X$ and $Y$ respectively. If the maximum of the sub-sample $(x_1, x_2)$ from the random sample $X$ treated with $A$ is greater than the maximum of the subsample $(y_1, y_2)$ from the random sample $Y$ treated with $B$, then we assign 1 to the kernel $U_{ij}$. Otherwise, we assign $-1$ to the kernel $U_{ij}$.

If in the subsample from sample $X$, one observation is censored and the other is uncensored, with the censored value greater than the uncensored value, and observations in the subsample from sample $Y$ are uncensored. Here, if the maximum of the subsample of two observations $(x_1', x_2)$ from sample $X$ is greater than the maximum of the subsample of any two observations $(y_1, y_2)$ from sample $Y$, we assign 1 to the kernel $U_{ij}$. A similar procedure is done for the opposite case, i.e., when one censored and one uncensored observation comes in the sub-sample from sample $Y$, and both uncensored come in the subsample from sample $X$, we assign $-1$ to the kernel $U_{ij}$. For remainder of the cases, we assign zero to the kernel $U_{ij}$. In mathematical terms, the kernel for the proposed test is:

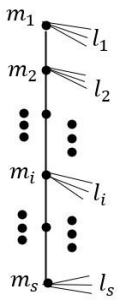

DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

$$U_{ij} = \begin{cases} 1, & \text{if } Max(x_{i_1}, x_{i_2}) > Max(y_{j_1}, y_{j_2}) \text{ or } Max(x'_{i_1}, x_{i_2}) \geq Max(y_{j_1}, y_{j_2}) \\ -1, & \text{if } Max(x_{i_1}, x_{i_2}) < Max(y_{j_1}, y_{j_2}) \text{ or } Max(x_{i_1}, x_{i_2}) \leq Max(y'_{j_1}, y_{j_2}) \\ 0, & \text{elsewhere,} \end{cases} \quad (1)$$

where $i_1 \neq i_2$ in $(1, 2, \ldots, n_1)$ and $j_1 \neq j_2$ in $(1, 2, \ldots, n_2)$ Define the statistic $V = \sum_{i,j} U_{ij}$, where the sum is extended over all $n_1$, $n_2$ combinations. There will be a contribution to statistic $V$ for all possible comparisons where both the patients have failed and, in all comparisons, where a patient who is censored has more survival than one who has failed earlier.

## 3. The Mean and Variance of Test Statistic

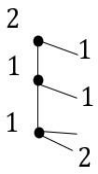
We have considered the same observational pattern described by Gehan (1965), i.e., if we have $n_1, n_2$ observations that can be settled in the following observational pattern:




when we rank the data, then $m_i$'s are the total uncensored observations at the $i^{th}$ rank with dissimilar values, and $l_i$'s are the total randomly right censored observations with values larger than the observational value at the $i^{th}$ rank, but should be smaller than the observational value at the $(i+1)^{th}$ rank.

The dots at the upright line represent the ordered ranks of the dissimilar values of time to failure observations, and these fall at $s$ dissimilar failure dots. Any observation that has either censored or failed can be characterized in the manner of the above pattern. Prior to the first failure, any censored observation will be counted as $l_1$ with $m_1 = 0$. Generally, these observations do not yield any difference between $A$ and $B$ treatments. Therefore, we omit these observations. As our calculation is restricted to the defined observational pattern, the omission of these observations does not have any consequence on mean and variance.

For example, if we have the following sample of survival times of patients (in months): $6, 8, 10+, 11, 11+, 13, 14+, 15+$ (the $+$ sign represents a censored observation at that particular point), the observational pattern will be:




Suppose, $H_0$ is true, i.e., the survival of patients in both the treated groups is same. We consider the conditional mean denoted by $E(V|P, H_0)$ and variance by $var(V|P, H_0)$ of $V$ under $H_0$, where $P$ is the observational pattern. The expectation was considered over the possible number of samples $\binom{n_1 + n_2}{n_1}$ that are equally likely and follow same defined observational pattern. Due to symmetry, we can easily observe:

$$E(V|P, H_0) = E\left(\sum_{i,j} U_{ij} \,\middle|P, H_0\right) = 0. \quad (2)$$

The variance of $V$ under $H_0$ is restricted to the defined pattern $P$, and can be defined as:

$$var(V|P, H_0) = E\left(\sum_{i,j} U_{ij} - E\left(\sum_{i,j} U_{ij}\right)\middle| P, H_0\right)^2. \quad (3)$$

From eq. (2), we know that $E\left(\sum_{i,j} U_{ij} \,\middle|P, H_0\right) = 0$. Thus, eq. (3) becomes:

$$var(V|P, H_0) = E\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} U_{ij}^2 + \sum_{i\neq i'=1}^{n_1}\sum_{j=1}^{n_2} U_{ij}U_{i'j} + \sum_{i=1}^{n_1}\sum_{j\neq j'=1}^{n_2} U_{ij}U_{ij'} + \sum_{i\neq i'=1}^{n_1}\sum_{j\neq j'=1}^{n_2} U_{ij}U_{i'j'}\middle| P, H_0\right). \quad (4)$$


DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

After evaluating each term of eq. (4), we obtain:

$$var(V|P,H_0)=\frac{24\binom{n_1+n_2-4}{n_1-2}}{\binom{n_1+n_2}{n_1}}K_1+\left\{\frac{\binom{n_1+n_2-6}{n_1-4}}{\binom{n_1+n_2}{n_1}}+\frac{\binom{n_1+n_2-6}{n_2-4}}{\binom{n_1+n_2}{n_1}}\right\}K_2. \quad (5)$$

In eq. (5), the coefficient of $K_1$ is the proportion of times a specific pair of observations $(i,j)$ turns up in samples $X$ and $Y$. Similarly, the coefficient of $K_2$ is the proportion of times a specific pair $(i,i'\text{and } i\neq i')$ turns up in any one of the samples with observation $j$ from the other sample. Here

$$K_1=\sum_{i=1}^{s}\left[\binom{m_i}{1}\binom{M_{i-1}}{3}+\binom{m_i}{1}\binom{n_1+n_2-M_i-L_{i-1}}{3}+\binom{l_i}{1}\binom{M_i}{3}\right.$$
$$\left.+\binom{m_i}{1}\binom{M_{i-1}}{1}\binom{n_1+n_2-M_i-L_{i-1}}{2}+\binom{m_i}{1}\binom{M_{i-1}}{2}\binom{n_1+n_2-M_i-L_{i-1}}{1}\right], \quad (6)$$

where the first and second terms in eq. (6) represent the total number of ways of pairing any failed observation at $i^{th}$ rank with any three observations at a lesser rank and any three observations of a rank greater than $i$ respectively. The third term represents the total number of ways of pairing a censored observation immediately after the $i^{th}$ rank, with any three that have failed earlier. The fourth term shows the total number of ways of pairing any failed observation with one of rank lesser than $i$ and two other observations with a rank greater than $i$. The last term represents the total number of ways of pairing any failed observation with any two observations of a rank lesser than $i$ and one other observation with a rank greater than $i$. Similarly,

$$K_2=\sum_{i=1}^{s}\left[120\binom{m_i}{1}\binom{M_{i-1}}{5}+56\binom{m_i}{1}\binom{n_1+n_2-M_i-L_{i-1}}{5}+120\binom{l_i}{1}\binom{M_i}{5}\right.$$
$$+56\binom{m_i}{1}\binom{M_{i-1}}{1}\binom{n_1+n_2-M_i-L_{i-1}}{4}-72\binom{m_i}{1}\binom{M_{i-1}}{4}\binom{n_1+n_2-M_i-L_{i-1}}{1}$$
$$\left.+56\binom{m_i}{1}\binom{M_{i-1}}{2}\binom{n_1+n_2-M_i-L_{i-1}}{3}+24\binom{m_i}{1}\binom{M_{i-1}}{3}\binom{n_1+n_2-M_i-L_{i-1}}{2}\right] \quad (7)$$

The first and second terms within square brackets in eq. (7), represent the total number of ways of pairing any failed observation at the $i^{th}$ rank with any five observations of a lesser rank and any five of a rank greater than $i$ respectively. The third term shows the total number of ways of pairing any censored observation immediately after the $i^{th}$ rank with any five that have failed earlier. The Fourth term shows the total number of ways of pairing any failed observation with a rank lesser than $i$ and four other observations with a rank greater than $i$. The fifth term shows the total number of ways of pairing any failed observation with any four observations of a rank lesser than $i$ and one other observation with a rank greater than $i$. The sixth term shows the total number of ways of pairing any failed observation with any two observations of a rank lesser than $i$ and three other observations with a rank greater than $i$. The last term represents the total number of ways of pairing any failed observation with any three observations of a rank lesser than $i$ and two other observations with a rank greater than $i$, and

$$M_j=\sum_{i=1}^{j}m_i, \qquad M_0=0,$$

$$L_j=\sum_{i=1}^{j}l_i\ , \qquad L_0=0.$$

$m_i$ and $l_i$ are in their original meanings, as defined initially in this Section.

## 4. Critical Points

In hypothesis testing, we determine whether sufficient evidence exists from the sample to accept or reject $H_0$. Critical points are essentially the cut-off values such that if the calculated test statistic value comes out to be greater than the cut-off value, we reject $H_0$; otherwise, we do not reject $H_0$. These values are specific for a test statistic that depends on the type of test and the level of significance $\alpha$. Using this concept, critical points are found using a


DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

simulation study for both the tests $V$ (proposed) and $G$ (by Gehan 1965). We consider three different lifetime distributions (Lindley, Exponential and Weibull) for generating time to failure observations and Exponential distribution for generating time to censored observations. Two samples, each of size $n$, are generated from these distributions, and the standardized test statistic value can be found using the following formula:

$$Z = \frac{V - E(V|P, H_0)}{\sqrt{var(V|P, H_0)}} \tag{8}$$

where $E(V|P, H_0)$ and $var(V|P, H_0)$ are given in eqs. (2) and (5). We then find $Z_\alpha$ such that $P(Z > Z_\alpha) = 0.025$. This process is simulated 10,000 times, and the critical points are found as the average of $Z_\alpha$-values. Critical points are given in Tables $1-3$ for various censoring percentages $(pcens)$ and sample sizes $(n_1 = n_2 = n)$ for each distribution.

**Table 1.** Critical points of the proposed test, when the time to failure distribution is Lindley and time to censoring distribution is Exponential

| Test | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ |
|---|---|---|---|---|---|---|---|---|
| $n$ \ $pcens$ | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.5 |
| 10 | 2.91112 | 3.35186 | 1.97061 | 3.31972 | 1.47104 | 3.24155 | 1.21353 | 3.07304 |
| 15 | 2.89617 | 3.80541 | 1.96009 | 3.91377 | 1.55739 | 3.86303 | 1.38231 | 3.72432 |
| 20 | 2.84070 | 4.27193 | 2.04971 | 4.46372 | 1.67048 | 4.40178 | 1.51094 | 4.23448 |
| 25 | 2.92921 | 4.74102 | 2.12339 | 4.92879 | 1.76357 | 4.88847 | 1.63321 | 4.73802 |
| 30 | 3.01728 | 5.10966 | 2.24109 | 5.39651 | 1.89059 | 5.35712 | 1.77043 | 5.16320 |

**Table 2.** Critical points of the proposed test, when both time to failure and time to censoring distributions are Exponential

| Test | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ |
|---|---|---|---|---|---|---|---|---|
| $n$ \ $pcens$ | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.5 |
| 10 | 2.92186 | 3.38056 | 1.91547 | 3.32851 | 1.43057 | 3.24729 | 1.21353 | 3.06260 |
| 15 | 2.90303 | 3.91751 | 1.95929 | 3.93480 | 1.53092 | 3.88975 | 1.33974 | 3.70345 |
| 20 | 2.88096 | 4.40868 | 2.04536 | 4.52102 | 1.64325 | 4.41977 | 1.46696 | 4.22876 |
| 25 | 2.93422 | 4.96588 | 2.14505 | 5.00451 | 1.75391 | 4.87901 | 1.60609 | 4.70970 |
| 30 | 3.02035 | 5.34874 | 2.25501 | 5.50358 | 1.86846 | 5.35288 | 1.73120 | 5.11489 |

**Table 3.** Critical points of the proposed test, when the time to failure distribution is Weibull and time to censoring distribution is Exponential

| Test | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ |
|---|---|---|---|---|---|---|---|---|
| $n$ \ $pcens$ | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.5 |
| 10 | 2.93616 | 3.44826 | 1.91492 | 3.33473 | 1.38603 | 3.15828 | 1.13679 | 2.90428 |
| 15 | 2.94675 | 4.09885 | 1.98431 | 4.013749 | 1.50179 | 3.81971 | 1.27410 | 3.52024 |
| 20 | 2.98755 | 4.72489 | 2.07529 | 4.69799 | 1.63558 | 4.44922 | 1.43581 | 4.07413 |
| 25 | 3.03894 | 5.30250 | 2.21491 | 5.23519 | 1.76892 | 4.95976 | 1.55849 | 4.51801 |
| 30 | 3.16017 | 5.86804 | 2.31089 | 5.76923 | 1.90151 | 5.42111 | 1.68737 | 4.98331 |

## 5. Asymptotic Relative Efficiency

Herein, we find the asymptotic relative efficiency $(ARE)$ of the proposed test statistic $V$ relative to the usual $F-$test, assuming Exponential lifetime distribution. Let us suppose that the time to failure probability density function of a patient who is receiving treatment $A$ is given as

$$f_1(x) = \phi \exp(-\phi x),$$

and that for a patient who is receiving treatment $B$ is given as


DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

$$f_2(y) = \theta\phi \exp(-\theta\phi y).$$

Our interest is to test the hypothesis:

$$H: F_1(t) = F_2(\theta t) \qquad (t \leq T).$$

Thus, our null hypothesis would be, $H_0: \theta = 1$. This type of test would be relevant in situations when we are interested in determining whether there is any constant proportion $(\theta)$ of failure times of the patients who are receiving treatment $B$ to that of who are receiving treatment $A$.

Another competent test for the above hypothesis is to consider $\bar{t}_1/\bar{t}_2$, following $F$- distribution with degrees of freedom $(2(n_1 - r_1), 2(n_2 - r_2))$, where

$$\bar{t}_1 = \frac{\left(\sum_{i=1}^{r_1} x_i' + \sum_{i=r_1+1}^{n_1} x_i\right)}{(n_1 - r_1)}, \quad \bar{t}_2 = \frac{\left(\sum_{i=1}^{r_2} y_i' + \sum_{i=r_2+1}^{n_2} y_i\right)}{(n_2 - r_2)}.$$

We aim to find the $ARE$ of the proposed test relative to $F$ test in the situation, when all the individuals enter study at time zero and the experiment is stopped at time $T$. The $ARE$ of the proposed $V$ test relative to $F$ test is given by,

$$ARE_{VF} = \lim_{n\to\infty} \frac{\left(\frac{\partial E(n^{-2}V)}{\partial\theta}\Big|_{\theta=1}\right)^2}{(nvar(n^{-2}V|H_0))} \times \frac{(nvar(z|H_0))}{\left(\frac{\partial E(z)}{\partial\theta}\Big|_{\theta=1}\right)^2}. \qquad (9)$$

For $F$ test, it is appropriate to transform the $F$ statistic to $z = \frac{1}{2}\log(F)$, where $z$ is following the Normal distribution asymptotically with $var \cong \frac{1}{2}\left(\frac{1}{2(n_1 - r_1)} + \frac{1}{2(n_2 - r_2)}\right)$ and

$$E(z) = E_s E(z|s),$$

$$var(z|H_0) = E_s var(z|H_0, s) + var_s E(z|H_0, s).$$

Herein, the observational pattern is defined by the total sample size $(2n)$ and total failure observations preceding time $T$. The expectations and variances of $z$ are calculated under the conditional pattern, where $s = 2n - r_1 - r_2$ is fixed, and then allow variation in $s$. The calculations are asymptotic as $n, s \to \infty$. Under $H_0$, $s$ follows a Binomial distribution with mean $2n\left(1 - e^{-\phi T}\right)$. Thus,

$$E(z) \cong \frac{1}{2}\log\theta \quad \text{and} \quad \frac{\partial E(z)}{\partial\theta}\Big|_{\theta=1} = \frac{1}{2}, \qquad (10)$$

$$var(z|H_0) \cong \frac{1}{2n(1 - e^{-\phi T})}. \qquad (11)$$

For the proposed test, we have $V = \sum_{i,j} U_{ij}$, as defined earlier, thus:

$$E(V) = n^2 \left\{\Pr\left(Max\left(X_{i_1}, X_{i_2}\right) > Max\left(Y_{j_1}, Y_{j_2}\right)\right) + \Pr\left(Max\left(X_{i_1}', X_{i_2}\right) \geq Max\left(Y_{j_1}, Y_{j_2}\right)\right)\right.$$
$$\left. - \Pr\left(Max\left(X_{i_1}, X_{i_2}\right) < Max\left(Y_{j_1}, Y_{j_2}\right)\right) - \Pr\left(Max\left(X_{i_1}, X_{i_2}\right) \leq Max\left(Y_{j_1}', Y_{j_2}\right)\right)\right\} \qquad (12)$$

where random variables $X$ and $X'$ denote the time to failure and time to censoring of the patients who are receiving treatment $A$, and determined by the probability density function $f_1(x)$. Similarly, random variables $Y$ and $Y'$ denote the time to failure and time to censoring of the patients who are receiving treatment $B$, and determined by probability density function $f_2(y)$. Herein, $X' \equiv Y' \equiv T$ and according to the assumed lifetime distribution, for a patient receiving treatment $A$ and $B$, the probability of being censored at $T$ is $e^{-T\phi}$ and $e^{-T\theta\phi}$ respectively. We now obtain these probabilities as follows:

$$\Pr\left(Max\left(X_{i_1}, X_{i_2}\right) > Max\left(Y_{j_1}, Y_{j_2}\right)\right) + \Pr\left(Max\left(X_{i_1}', X_{i_2}\right) \geq Max\left(Y_{j_1}, Y_{j_2}\right)\right)$$


DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

$$= 2\theta\phi \int_0^T \left(1-\left(1-e^{-\phi u}\right)^2\right)\left(1-e^{-\theta\phi u}\right)e^{-\theta\phi u}du$$

$$= \theta\left\{\frac{4}{(1+\theta)}\left(1-e^{-(1+\theta)\phi T}\right)-\frac{4}{(1+2\theta)}\left(1-e^{-(1+2\theta)\phi T}\right)-\frac{2}{(2+\theta)}\left(1-e^{-(2+\theta)\phi T}\right)\right.$$
$$\left.-\frac{1}{(1+\theta)}\left(1-e^{-2(1+\theta)\phi T}\right)\right\}, \qquad (13)$$

and

$$\Pr\left(Max\left(X_{i_1},X_{i_2}\right)<Max\left(Y_{j_1},Y_{j_2}\right)\right)+\Pr\left(Max\left(X_{i_1},X_{i_2}\right)\le Max\left(Y'_{j_1},Y_{j_2}\right)\right)$$

$$= 2\theta\phi\int_0^T\left(1-e^{-\phi u}\right)^2\left(1-e^{-\theta\phi u}\right)e^{-\theta\phi u}du$$

$$= \theta\left\{\frac{2}{\theta}\left(1-e^{-\theta\phi T}\right)-\frac{1}{\theta}\left(1-e^{-2\theta\phi T}\right)+\frac{2}{(2+\theta)}\left(1-e^{-(2+\theta)\phi T}\right)-\frac{1}{(1+\theta)}\left(1-e^{-2(1+\theta)\phi T}\right)\right.$$
$$\left.-\frac{4}{(1+\theta)}\left(1-e^{-(1+\theta)\phi T}\right)+\frac{4}{(1+2\theta)}\left(1-e^{-(1+2\theta)\phi T}\right)\right\}. \qquad (14)$$

Substituting eqs. (13) and (14) in eq. (12), we arrive at

$$E(n^{-2}V) = \theta\left\{\frac{8}{(1+\theta)}\left(1-e^{-(1+\theta)\phi T}\right)-\frac{8}{(1+2\theta)}\left(1-e^{-(1+2\theta)\phi T}\right)-\frac{1}{\theta}\left(1-e^{-2\theta\phi T}\right)\right.$$
$$\left.-\frac{4}{(2+\theta)}\left(1-e^{-(2+\theta)\phi T}\right)+\frac{2}{(1+\theta)}\left(1-e^{-2(1+\theta)\phi T}\right)-\frac{2}{\theta}\left(1-e^{-\theta\phi T}\right)\right\}.$$

Thus,

$$\left.\frac{\partial E(n^{-2}V)}{\partial\theta}\right|_{\theta=1} = \left\{\frac{8}{9}-2\phi Te^{-\phi T}+6\phi Te^{-2\phi T}-\frac{8}{9}e^{-3\phi T}-\frac{20}{3}\phi Te^{-3\phi T}+\frac{4}{3}\phi Te^{-4\phi T}\right\}. \qquad (15)$$

Now,

$$var(n^{-2}V|H_0) = n^{-4}E\left\{\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}U_{ij}-E\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}U_{ij}\right)\middle|H_0\right)\right\}^2$$

$$\Rightarrow \qquad var(n^{-2}V|H_0) = n^{-4}E\left\{\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}U_{ij}\middle|H_0\right)\right\}^2. \qquad (16)$$

Since, $E\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}U_{ij}\right)=0$ and $E\left(\sum_{i\neq i'=1}^{n_1}\sum_{j}^{n_2}U_{ij}U_{i'j}\right)=E\left(\sum_{i=1}^{n_1}\sum_{j\neq j'=1}^{n_2}U_{ij}U_{ij'}\right)$. Also, $E\left(\sum_{i\neq i'=1}^{n_1}\sum_{j\neq j'=1}^{n_2}U_{ij}U_{i'j'}\right)=0$, since $U_{ij}$ *and* $U_{i'j'}$ are independent of each other and have expectation zero. Further evaluating each term of eq. (16), we get

$$var(n^{-2}V|H_0) = n^{-1}\left\{\frac{2}{3}\left(1-e^{-\phi T}\right)^3+2e^{-\phi T}\left(1-e^{-\phi T}\right)\right\}. \qquad (17)$$

Substituting eqs. (10), (11), (15) and (17) in eq. (9), we get the $ARE$ of $V$ to $F$ as,

$$ARE_{VF} = \frac{\left(\frac{8}{9}-2\phi Te^{-\phi T}+6\phi Te^{-2\phi T}-\frac{8}{9}e^{-3\phi T}-\frac{20}{3}\phi Te^{-3\phi T}+\frac{4}{3}\phi Te^{-4\phi T}\right)^2}{\frac{1}{3}(1-e^{-\phi T})^4+e^{-\phi T}(1-e^{-\phi T})^2}. \qquad (18)$$

Similarly, we can determine the ARE of the proposed $V$ test to $G$ (Gehan's test) as,

$$ARE_{VG} = \frac{\left(\frac{8}{9}-2\phi Te^{-\phi T}+6\phi Te^{-2\phi T}-\frac{8}{9}e^{-3\phi T}-\frac{20}{3}\phi Te^{-3\phi T}+\frac{4}{3}\phi Te^{-4\phi T}\right)^2\left(\frac{4}{3}\left(1-e^{-\phi T}\right)+4e^{-\phi T}\right)}{(1-e^{-2\phi T})^2\left(\frac{1}{3}(1-e^{-\phi T})^2+e^{-\phi T}\right)}, \qquad (19)$$


DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

where

$$\phi T = \frac{total\ study\ time}{average\ failure\ time\ on\ treatment\ A}.$$

Using eqs. (18) and (19), the $ARE s$ of the proposed test $V$ with respect to (w.r.t.) $F$ test and Gehan's test $G$ for various values of $\phi T$ are shown in Table 4.

Table 4. $ARE$ of the $V$ test w.r.t. $F$ and $G$ (Gehan) tests for various values of $\phi T$

| $\phi T$ | $\rightarrow 0$ | 1 | 2 | 3 | 4 | 5 | $\rightarrow \infty$ |
|---|---|---|---|---|---|---|---|
| $ARE_{VF}$ | 1.778 | 1.879 | 0.982 | 1.262 | 1.720 | 2.045 | 2.370 |
| $ARE_{VG}$ | 1.778 | 2.013 | 1.179 | 1.607 | 2.253 | 2.709 | 3.160 |

When the value of $\phi T$ is greater than two, the $ARE$ for our test with respect to both $F$ test and Gehan's test increases with $\phi T$. The value of $ARE$ with respect to Gehan's test is always greater than 1, i.e., our test performs better than Gehan's test for all $\phi T$ considered here. Moreover, the proposed test performs better than $F$ test for all considered values, except at $\phi T = 2$.

## 6. Real-Life Example

We worked on a real-life example derived from Stablein & Koutrouvelis (1985), which is based on a trial of patients who suffered from locally unrestricted gastric cancer and were treated with chemotherapy and chemotherapy with radiotherapy. This data offers the survival time (in days) for the 45 patients on each treatment.

To check the distribution of data, we applied a Kolmogorov-Smirnov test and found that this data set follows Exponential distribution. Our aim is to determine if there is a significant difference in the survival times of the patients treated with chemotherapy and chemotherapy with radiotherapy. The critical value of the proposed test statistic for sample size $(45, 45)$ in the case of Exponential distribution is found to be $= 4.9453$ using the procedure described in Section 4. The standardized test statistic $(z)$ for this data is $= 8.4554$. Since the calculated test statistic value turns out to be greater than its critical value, the null hypothesis of no significant difference is rejected. It is concluded that there is a significant difference in the survival time of the patients treated with chemotherapy and chemotherapy with radiotherapy.

## 7. Power Comparison of The Proposed Test

The statistical power of a test is defined as the probability that the test rejects the null hypothesis when it is true. Using the critical values shown in Section 4, the statistical power of the proposed test and Gehan's test was computed through a Monte-Carlo simulation. Data were simulated from three lifetime distributions viz., Lindley, Exponential and Weibull 10,000 times, with a scale parameter of the second sample as $\theta = 2, 3, \text{and } 4$. The statistical power of the proposed test $V$ and Gehan's test $G$ is shown in the Tables 5-7 with same sample sizes and censoring percentages that we have considered for computing the critical points.

Table 5. Statistical power of the $V$ and $G$ tests, when the time to failure distribution is Lindley and time to censoring distribution is Exponential

| | Test | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $pcens$ | | | |
| $n$ | $\theta$ | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.5 |
| 10 | 2 | 0.395 | 0.298 | 0.255 | 0.290 | 0.203 | 0.254 | 0.141 | 0.207 |
| | 3 | 0.611 | 0.569 | 0.404 | 0.483 | 0.301 | 0.441 | 0.204 | 0.362 |
| | 4 | 0.693 | 0.717 | 0.474 | 0.639 | 0.372 | 0.549 | 0.263 | 0.509 |
| 15 | 2 | 0.414 | 0.379 | 0.396 | 0.375 | 0.360 | 0.314 | 0.235 | 0.277 |
| | 3 | 0.698 | 0.642 | 0.595 | 0.620 | 0.494 | 0.588 | 0.382 | 0.526 |
| | 4 | 0.767 | 0.849 | 0.730 | 0.782 | 0.603 | 0.738 | 0.424 | 0.632 |
| 20 | 2 | 0.570 | 0.442 | 0.498 | 0.434 | 0.406 | 0.404 | 0.307 | 0.374 |
| | 3 | 0.843 | 0.772 | 0.771 | 0.745 | 0.671 | 0.698 | 0.451 | 0.622 |
| | 4 | 0.926 | 0.911 | 0.879 | 0.878 | 0.750 | 0.818 | 0.571 | 0.793 |
| 25 | 2 | 0.681 | 0.511 | 0.638 | 0.509 | 0.493 | 0.501 | 0.374 | 0.424 |
| | 3 | 0.935 | 0.832 | 0.909 | 0.822 | 0.776 | 0.812 | 0.599 | 0.736 |
| | 4 | 0.980 | 0.934 | 0.963 | 0.928 | 0.875 | 0.934 | 0.665 | 0.857 |
| 30 | 2 | 0.807 | 0.579 | 0.727 | 0.559 | 0.605 | 0.555 | 0.405 | 0.504 |
| | 3 | 0.982 | 0.888 | 0.965 | 0.863 | 0.878 | 0.844 | 0.651 | 0.794 |
| | 4 | 0.995 | 0.970 | 0.988 | 0.952 | 0.944 | 0.934 | 0.792 | 0.916 |


DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

Table 6. Statistical power of the $V$ and $G$ tests, when both time to failure and time to censoring distributions are Exponential

| | Test | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $pcens$ | | | | |
| $n$ | $\theta$ | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.5 |
| | 2 | 0.203 | 0.214 | 0.180 | 0.179 | 0.143 | 0.158 | 0.129 | 0.156 |
| 10 | 3 | 0.313 | 0.372 | 0.281 | 0.341 | 0.211 | 0.290 | 0.190 | 0.263 |
| | 4 | 0.409 | 0.554 | 0.350 | 0.471 | 0.283 | 0.434 | 0.227 | 0.397 |
| | 2 | 0.302 | 0.235 | 0.255 | 0.218 | 0.202 | 0.211 | 0.163 | 0.208 |
| 15 | 3 | 0.503 | 0.472 | 0.450 | 0.431 | 0.391 | 0.425 | 0.215 | 0.379 |
| | 4 | 0.630 | 0.615 | 0.581 | 0.575 | 0.461 | 0.554 | 0.333 | 0.490 |
| | 2 | 0.381 | 0.304 | 0.338 | 0.279 | 0.274 | 0.286 | 0.214 | 0.224 |
| 20 | 3 | 0.716 | 0.575 | 0.603 | 0.498 | 0.493 | 0.489 | 0.346 | 0.465 |
| | 4 | 0.830 | 0.764 | 0.777 | 0.697 | 0.634 | 0.654 | 0.416 | 0.599 |
| | 2 | 0.501 | 0.357 | 0.453 | 0.329 | 0.336 | 0.285 | 0.229 | 0.272 |
| 25 | 3 | 0.825 | 0.641 | 0.740 | 0.637 | 0.643 | 0.592 | 0.426 | 0.530 |
| | 4 | 0.921 | 0.826 | 0.883 | 0.806 | 0.766 | 0.757 | 0.552 | 0.678 |
| | 2 | 0.613 | 0.388 | 0.509 | 0.372 | 0.434 | 0.367 | 0.274 | 0.316 |
| 30 | 3 | 0.904 | 0.749 | 0.834 | 0.672 | 0.709 | 0.648 | 0.496 | 0.620 |
| | 4 | 0.981 | 0.878 | 0.952 | 0.835 | 0.835 | 0.827 | 0.646 | 0.752 |

Table 7. Statistical power of the $V$ and $G$ tests, when the time to failure distribution is Weibull and time to censoring distribution is Exponential

| | Test | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ | $V$ | $G$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $pcens$ | | | | |
| $n$ | $\theta$ | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.5 |
| | 2 | 0.325 | 0.360 | 0.278 | 0.295 | 0.238 | 0.251 | 0.168 | 0.233 |
| 10 | 3 | 0.477 | 0.613 | 0.421 | 0.576 | 0.371 | 0.484 | 0.263 | 0.416 |
| | 4 | 0.538 | 0.798 | 0.499 | 0.681 | 0.427 | 0.617 | 0.319 | 0.550 |
| | 2 | 0.478 | 0.455 | 0.436 | 0.428 | 0.393 | 0.363 | 0.267 | 0.330 |
| 15 | 3 | 0.723 | 0.769 | 0.685 | 0.725 | 0.582 | 0.665 | 0.439 | 0.567 |
| | 4 | 0.807 | 0.916 | 0.769 | 0.879 | 0.674 | 0.808 | 0.506 | 0.734 |
| | 2 | 0.645 | 0.514 | 0.599 | 0.506 | 0.499 | 0.498 | 0.351 | 0.430 |
| 20 | 3 | 0.897 | 0.859 | 0.862 | 0.836 | 0.752 | 0.801 | 0.572 | 0.723 |
| | 4 | 0.957 | 0.954 | 0.929 | 0.920 | 0.839 | 0.906 | 0.665 | 0.849 |
| | 2 | 0.786 | 0.625 | 0.717 | 0.614 | 0.612 | 0.598 | 0.447 | 0.532 |
| 25 | 3 | 0.975 | 0.917 | 0.942 | 0.910 | 0.875 | 0.869 | 0.707 | 0.816 |
| | 4 | 0.995 | 0.977 | 0.979 | 0.972 | 0.953 | 0.945 | 0.801 | 0.921 |
| | 2 | 0.864 | 0.701 | 0.816 | 0.695 | 0.699 | 0.667 | 0.537 | 0.601 |
| 30 | 3 | 0.992 | 0.942 | 0.980 | 0.938 | 0.932 | 0.910 | 0.803 | 0.873 |
| | 4 | 0.999 | 0.986 | 0.995 | 0.983 | 0.971 | 0.970 | 0.882 | 0.955 |

According to the Tables 5-7, we observe the following about the statistical power of the tests:

a. The power of both tests increases with an increase in the sample size $(n)$ and scale parameter $(\theta)$. However, the power decreases with an increase in the censoring percentage $(pcens)$.
b. In the case of Lindley and Weibull distribution, from Tables 5 and 7, we observe that for the smaller censoring percentage, the proposed test performs better than Gehan's test for all scale parameters considered here when the sample size is $\geq 20$.
c. In the case of Exponential distribution, from Table 6, we observe that for the smaller censoring percentage, the proposed test performs better than Gehan's test for all scale parameters considered here when the sample size is $\geq 15$.

## 6. References

Alling D. (1963). Early decision in the Wilcoxon two-sample test. Journal of the American Statistical Association 58: 713-720.

Deshpande JV. & Purohit SG. (2015). Lifetime Data: Statistical Models and Methods (2nd edition), Singapore: World Scientific Publishing Co. Pvt. Ltd.

Efron B. (1967). The two-sample problem with censored data. Berkeley Symposium on Mathematical Statistics and Probability 5.4: 831-853.



DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)

Gehan EA. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52: 202-223.

Goyal M. & Kumar N. (2020). Two new classes of nonparametric tests for scale parameters. Journal of Statistical Computation and Simulation 90: 3093-3105.

Halperin M. (1960). Extension of the Wilcoxon-Mann-Whitney test to samples censored at the same fixed point. Journal of the American Statistical Association 55: 125-138.

Kaplan EL. & Meier P. (1958). Non-Parametric estimation from incomplete observations. Journal of the American Statistical Association 53: 457-481.

Kössler W. (1994). Restrictive adaptive tests for the treatment of the two-sample scale problem. Computational Statistics & Data Analysis 18: 513–524.

Kössler W. & Kumar N. (2010). An adaptive test for the two-sample scale problem based on U-statistics. Communications in Statistics - Simulation and Computation 39: 1785-1802.

Lee ET., Desu MM. & Gehan EA. (1975). A Monte Carlo study of the power of some two-sample tests. Biometrika 62: 425-432.

Mantel N. (1967). Ranking procedures for arbitrarily restricted observation. Biometrics 23: 65-78.

Miller RG. (2011). Survival Analysis (2nd edition), New York, USA: John Wiley & Sons.

Mood AM. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. Annals of Mathematical Statistics 25: 514–522.

Rao UVR., Savage IR. & Sobel M. (1960). Contributions to the theory of rank order statistics: the two-sample censored case. Annals of Mathematical Statistics 31: 415-426.

Stablein DM. & Koutrouvelis IA. (1985). A two-sample test sensitive to crossing hazards in uncensored and singly censored data. Biometrics 41: 643-652.

Sukhatme BV. (1957). On certain two-sample nonparametric tests for variances. Annals of Mathematical Statistics 28: 188–194.

Wilcoxon F. (1945). Individual comparisons by ranking methods. Biometrics 1: 80-83.



DOI:https//doi.org/10.22452/mjs.vol42no1.4
Malaysian Journal of Science 42(1): 32-41 (February 2023)