Skillings-Mack Statistic: Computer-Intensive Methods

Patchanok Srisuradetchai*[1], Nantapath Trakultraipruk[1]

[1]Department of Mathematics and Statistics, Faculty of Science and
Technology, Thammasat University, Pathum-Thani 12121, Thailand
*spatchan@tu.ac.th

Abstract. The Skillings-Mack statistic is the appropriate statistic when missing observations occur randomly in a block design. In this paper, the exact randomization test and the Monte Carlo method are applied for the Skilling-Mack test statistic. We found that when there is more missing data, the chi-squared distribution is worse for being a reference distribution to calculate critical values even for significance levels near .10. Furthermore, for some designs with two missing values it is impossible to reject the null hypothesis when using the chi-squared distribution. Also, we have developed the R package named Skillings.Mack to calculate the Skillings-Mack statistic as well as produce p-values based on the Monte Carlo method. Our R package is very useful when there are many ties and/or small designs are conducted. Finally, we present tables of critical values for block designs with two missing observations.

Keywords: block designs, Monte Carlo method, missing values

1 Introduction

When the experimental units are not uniform, but can be grouped into subsets of homogeneous units, randomized complete block designs (RCBDs) are widely used. One of the assumptions in the analysis of variance (ANOVA) is that error variables are normally distributed. This assumption, however, is sometimes doubtful, so the ANOVA should be used with care. A distribution-free method is alternative when the normality assumption is violated. For RCBDs, the model is:

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij},$$

where y_{ij} is the response of the j th treatment in the i th block, μ is a baseline mean or median, β_i is the i th block effect, and τ_j is the j th treatment effect. The continuous random error variables, ε_{ij} , are assumed to be independent and identically distributed. Suppose, there are k treatments, n blocks, and n_{ij} equals 1 if y_{ij} is not missing and 0 if y_{ij} is missing. Also, a null hypothesis of interest is $H_0 : \tau_1 = \tau_2 = \ldots = \tau_k$.

Friedman [1] proposed the nonparametric test statistic for a RCBD. Where $n_{ij} = 1$ for all i and j . Durbin [2] developed a Friedman-type test for a balanced incomplete block design (BIBD), where $n_{ij} = 1$ for specific i and j . Benard and Van Elteren [3] proposed their statistic for block designs with an unstructured set of missing cells and with any number of observations, i.e. $n_{ij} \geq 0$. The Benard-Van Elteren test statistic is the generalization of the method of m rankings as used in the Friedman rank test and Durbin rank test, but the form of the Benard-Van Elteren test statistic is not easy to use in practice. Brunden and Mohberg [4] developed a test statistic based on the Benard-Van Elteren test for the cases with at least one observation per cell, $n_{ij} > 0$. The Brunden-Mohberg statistic has a quadratic form, which can be calculated easily for a small number of treatments, i.e. $k \leq 3$. Brunden and Mohberg [4] showed that the Benard-Van Elteren test statistic can be computed by multiplication and using an inverse matrix or a generalized inverse.

Later, Skillings and Mack [5] proposed a test statistic for block designs with n_{ij} equals to either 0 or 1. The Skillings-Mack statistic is also in the quadratic form of the Brunden-Mohberg statistic and can be used over a wide range of cases. For example, in a BIBD, researchers intentionally have unobserved data with a special structure. The Skillings-Mack statistic is equivalent to the statistic in the Friedman rank test for a RCBD and the Durbin test for a BIBD [5].

There are a few papers discussing the practical implementation of the Skillings-Mack statistic. Chatfield and Mander [6] introduced the Skillings-Mack test in the Stata®, and Cunningham [7] introduced the use of SAS/IML® in SAS version 9.1 for calculating the Skillings-Mack test statistic. Besides Skillings-Mack's paper in 1981, no journal article was found that discussed the distribution of the Skillings-Mack *T* statistic in relation to the number of missing values. Skillings and Mack [5] used the Monte Carlo method for estimating the distribution of their *T* statistic for designs having a single missing observation for those designs considered to be large designs (i.e. for the cases $k = 3, n \geq 10;\ k = 4, n \geq 6;\ k = 5, n \geq 5;\ k = 6, n \geq 3$). For small designs, Skillings-Mack [5] determined the exact distribution of statistic *T*. Because the Skillings-Mack statistic is discrete, it is not always possible to find critical values corresponding to exactly $\alpha = .01, .05, \text{ and } .10$. Skillings and Mack [5] recommended that when critical values are not obtained from the exact distribution of *T*, those from the Monte Carlo approach are preferable to those based on using the chi-squared approximation.

Because the Durbin rank test statistic is a special case of the Skillings-Mack statistic, many studies of the distribution of the Durbin rank statistic also are applicable when discussing the Skillings-Mack statistic. In a simulation study, Bi Jian [8] showed that the chi-squared approximation should be avoided for the Durbin rank test statistic unless the BIBD is very large. For $k = 2$ and $n \leq 15$, they recommended using the randomization test. Fawcett and Salter [9] studied the distribution of the Durbin rank test statistic using a Monte Carlo approach, and then concluded, "we found that asymptotic, χ^2 , distributions do not provide adequate approximations in BIBD's… ". As we have seen, the previous studies showed that using a chi-squared approximation in small BIBDs is not appropriate.

In practice, two missing values can easily occur in a RCBD. There are three distinct situations when two missing values occur: (1) they are absent in the same block; (2) they are absent for the same treatment; (3) they are absent for different treatments appearing in different blocks. For each situation, we present a table of critical values of the Skillings-Mack statistic under the null hypothesis. The adequacy of an assumption chi-squared approximation as the number of missing values increases is not known. Based on our study of the Skillings-Mack statistic, readers will gain an understanding of the impact when the number of missing values is greater than one. We also use R 3.2.2 [10] software to develop a package called *Skillings.Mack* [11]. The *Ski.Mack* function in the *Skillings.Mack* package will calculate the Skillings-Mack statistic, p-values estimated by the Monte Carlo approach, and approximate p-values based on a chi-squared distribution. Like Stata® [6] and SAS/IML® provided by Cunningham [7], if there are ties, the *Ski.Mack* function will use their average rank to shuffle with ranks from data with no ties within blocks when simulating under the null hypothesis to preserve the missing-data structure.

2 Notations and Equations

The Skillings-Mack test has appeared in several books but using different notations. The notations used in this paper are the following:

k	the number of treatments in a design
n	the number of blocks in a design
k_i	the number of treatments present in the i th block, $i = 1, 2, ..., n$
y_{ij}	the response of the j th treatment in the i th block, $j = 1, 2, ..., k$
r_{ij}	the rank assigned to y_{ij}

and $n_{ij} = 1$ if y_{ij} is not missing, and $n_{ij} = 0$ otherwise.

The response variable y_{ij} is assumed to be independent and has a common probability density function $f_{ij}(y) = f(y - \beta_i - \tau_j)$, where β_i is the block effect and τ_j is the treatment effect. If there are no missing observations in a RCBD (all $n_{ij} = 1$), it is well known that $E_0(r_{ij}) = (k+1)/2$ and $\text{var}_0(r_{ij}) = (k^2 - 1)/12$ under $H_0 : \tau_1 = \tau_2 = ... = \tau_k$. With missing values, the expectation and variance of r_{ij} are

$$E_0\left(r_{ij}\right)=\begin{cases}\dfrac{k_i+1}{2}, & \text{if } n_{ij}=1\\ 0\quad, & \text{otherwise}\end{cases}$$

and

$$\operatorname{var}_0(r_{ij})=\begin{cases}\dfrac{k_i^2-1}{12}=\dfrac{(k_i-1)(k_i+1)}{12}, & \text{if } n_{ij}=1\\ 0\quad, & \text{otherwise}\end{cases}$$

respectively [5]. Furthermore,

$$\operatorname{cov}_0(r_{ij},r_{i'j'})=\begin{cases}-\dfrac{k_i+1}{12}, & \text{if } i=i', j\neq j', n_{ij}=1 \text{ and } n_{i'j'}=1\\ 0\quad, & \text{otherwise}\end{cases}$$

Let's consider the value of $T_j=\sum_{i=1}^{n}\left[r_{ij}-\left(k_i+1\right)/2\right]$ where the summation is taken over the blocks. For the j th treatment, under null hypothesis $H_0:\tau_1=\tau_2=\ldots=\tau_k$, we have

$$\operatorname{var}_0(T_j)=\sum_{i=1}^{n}\frac{k_i^2-1}{12}n_{ij},\quad j=1,2,...,k$$

and

$$\operatorname{cov}_0(T_j,T_{j'})=-\sum_{i=1}^{n}\frac{k_i+1}{12}n_{ij}n_{ij'},\quad \leq j\neq j'\leq k\ ,$$

where n_{ij} equals 1 if the j th treatment presents in the i th block and equals 0 otherwise. Skillings and Mack [5] adjusted the difference $r_{ij}-\left(k_i+1\right)/2$ by weight $w_i=\left[12/(k_i+1)\right]^{1/2}$. Thus, they used an adjusted treatment sum A_j for the j th treatment, instead of T_j, where A_j is defined as

$$A_j=\sum_{i=1}^{n}w_i\left[r_{ij}-\frac{k_i+1}{2}\right]. \tag{1}$$

Using the weight w_i simplifies the variance and the covariance. For the j th treatment, under $H_0:\tau_1=\tau_2=\ldots=\tau_k$,

$$\operatorname{var}_0(A_j)=\sum_{i=1}^{n}(k_i-1)n_{ij}=\sigma_{jj},\quad j=1,2,...,k$$

and

$$\operatorname{cov}_0(A_j,A_{j'})=-\sum_{i=1}^{n}n_{ij}n_{ij'}=\sigma_{jj'},\quad 1\leq j\neq j'\leq k\ .$$

When calculating the Skillings-Mack statistic, any block having only one observation must be removed. Next, the observations are ranked within each block; if tie occurs, the average rank is used. For a missing value, y_{ij}, in the i th block the value of $\left(k_i+1\right)/2$ will be assigned to be the rank of that missing [5]. Then, A_j, the adjusted treatment sum for the j th treatment, can be calculated. The vector of k adjusted sums is $\boldsymbol{A}'=(A_1,A_2,...,A_k)$. Note that $\boldsymbol{A}'\mathbf{1}=0$, $E_0(\boldsymbol{A})=\mathbf{0}$, and the variance-covariance matrix for $\boldsymbol{A}$ is

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \cdots & \sigma_{kk} \end{bmatrix}. \tag{2}$$

The matrix $\sum$ is not full-rank because $\boldsymbol{A}'\mathbf{1} = 0$. The Skillings-Mack test statistic is defined as

$$T = \boldsymbol{A}' \sum^{-} \boldsymbol{A}, \tag{3}$$

where $\sum^{-}$ is a symmetric generalized inverse of $\sum$. The value of T is invariant to the choice of the generalized inverse $\sum^{-}$ [5], i.e., if $\sum_1^{-1}$ and $\sum_2^{-1}$ are generalized inverses of $\sum$, then $\boldsymbol{A}' \sum_1^{-1} \boldsymbol{A} = \boldsymbol{A}' \sum_2^{-1} \boldsymbol{A}$. Skillings and Mack [5] showed that as the number of blocks increases, $T = \boldsymbol{A}' \sum^{-} \boldsymbol{A}$ has a limiting chi-squared distribution with the degrees of freedom p equal to the rank of $\sum^{-}$. In particular, if all treatment pairs occur together in at least one block, then T has a limiting chi-squared distribution with $p = k-1$. The hypothesis $H_0 : \tau_1 = \tau_2 = \ldots = \tau_k$ will be rejected if $T > \chi^2(p)$.

3 Randomization test for Skillings-Mack Statistic

The randomization test is a well-known non-parametric test in which the distribution of the test statistic is obtained by calculating the test statistic for all possible data rearrangements or data shuffle. In other words, given that the null hypothesis is true, the sampling distribution is generated. In an experimental study, we need to find all possible treatment assignments for randomization tests. The p-value from the randomization test is the probability that the test statistic would be at least as extreme as observed when the null hypothesis is true. The earliest discussions of randomization tests were presented by Fisher [12] and Pitman [13]. Although being more computationally intensive than standard statistical tests, randomization tests are widely used when researchers are uncomfortable making assumptions about the distribution of a test statistic. Kennedy [14] summarized the main advantages of randomization tests. First, a researcher does not need to use a traditional test statistic that has an assumed distribution under the null hypothesis. Second, it is not necessary to rely on normally distributed errors. Third, a reference distribution of a test statistic which is usually derived on the asymptotic theory is suspect when used with a small sample size, but when using the randomization test, there is no need to be worried whether or not the sample size is large enough. The randomization test would produce the exact p-value.

For example, consider a RCBD with $k = 3$ (treatments A, B, and C), and $n = 3$. Suppose there are 2 missing values. The first missing value occurred for treatment B in the 2nd block. The second occurred for treatment A in the 3rd block. Therefore, there are $3!2!2! = 24$ possible data sets as shown in Fig. 1; the "NA" means a missing observation.

1st Data Set

Block		
1	2	3
A	NA	NA
B	A	B
C	C	C

2nd Data Set

Block		
1	2	3
A	NA	NA
B	A	C
C	C	B

3rd Data Set

Block		
1	2	3
A	NA	NA
B	C	B
C	A	C

4th Data Set

Block		
1	2	3
A	NA	NA
B	C	C
C	A	B

5th Data Set

Block		
1	2	3
B	NA	NA
A	A	B
C	C	C

6th Data Set

Block		
1	2	3
B	NA	NA
A	A	C
C	C	B

7th Data Set

Block		
1	2	3
B	NA	NA
A	C	B
C	A	C

...

24th Data Set

Block		
1	2	3
C	NA	NA
B	C	C
A	A	B

Fig. 1. All possible data sets for RCBDs with $k = 3$, $n = 3$, and 2 missing observations

For another example, consider the Skillings-Mack test in a RCBD with $k = 4$, and $n = 5$. For 2 missing observations within the same block, there are $4!^4 \times 2! = 663{,}552$ possible rank assignments. To determine the exact distribution of the Skillings-Mack statistic, it must calculate for all possible 663,552 rank assignments. Suppose the observed data have the ranks shown in Table 1 and using the *Ski.Mack* function, the value of the Skillings-Mack statistic is 7.566 with an approximate p-value of .0599 calculated from the chi-squared distribution with 3 degrees of freedom. The exact distribution of the Skillings-Mack statistic is shown in Fig. 2, and the exact p-value is .041. The null hypothesis would not be rejected at $\alpha = .05$ based on the chi-squared p-value which disagree with the conclusion based on the exact p-value.

Table 1. An example of rank for a RCBD with $k = 4$, $n = 5$ and 2 missing values within the 5th block.

Treatment	Block				
	1	2	3	4	5
A	1	1	1	2	1
B	2	2	4	1	2
C	3	3	3	3	NA
D	4	4	2	4	NA

Table 2. Two cases of ranks for a RCBD with $k = 4$, $n = 3$, and 2 missing values within the same block that generate the largest Skillings-Mack statistic T of 6.979.

	Block		
Treatment	1	2	3
A	1	1	1
B	2	2	NA
C	3	3	NA
D	4	4	2

	Block		
Treatment	1	2	3
A	4	4	2
B	3	3	NA
C	2	2	NA
D	1	1	1

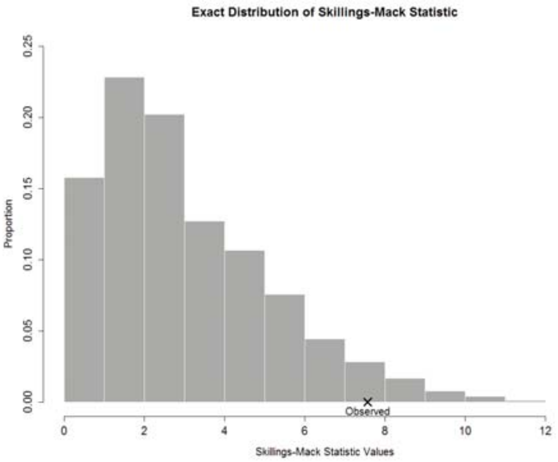

Fig. 2. Exact distribution of Skillings-Mack statistic for a RCBD with $k = 4, n = 5$, and 2 missing within the same block

For some designs, it may be impossible to reject the null hypothesis. For example, in a RCBD with $k = 4, n = 3$ and 2 missing values within the same block, no matter how strong the evidence might be, the null hypothesis can never be rejected for any $\alpha < .05$ if the chi-squared approximation is used. The exact distribution of the Skillings-Mack statistic for this RCBD is shown in Fig. 3. There are $4!^2 \times 2 = 1{,}152$ possible ways to assign rank, and Table 2 contains the two cases that produce the largest Skilling-Mack statistic of 6.979 . This observed statistic is still less than 7.8147 , the 95th percentile of $\chi^2(3)$, thus H_0 cannot be rejected using a chi-squared approximation to the $\alpha = .05$ critical value. Fig. 3 shows that the 95th percentile of $\chi^2(3)$ is larger than all possible Skillings-Mack statistics.

Even though the randomization test has many advantages, the computational time increases rapidly as the number of blocks or treatments increases. For example, for a RCBD with $n = 7, k = 4,$ and 2 missing values within the same block, there are $4!^6 \times 2! = 382{,}205{,}952$ possible rank assignments, hence a randomization test based on the exact randomization distribution is not always possible or practical to implement.

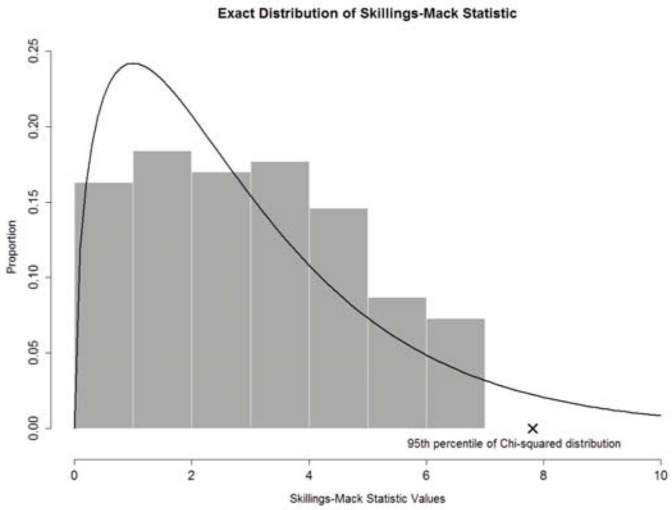

Fig. 3. Exact distribution of Skillings-Mack statistic for a RCBD with $k = 4, n = 3$, and 2 missing values within the same block and a superimposed chi-squared distribution with 3 degrees of freedom

4. Monte Carlo Methods for the Skillings-Mack Statistic

“A definition of a Monte-Carlo method would be one that involves deliberate use of random numbers in a calculation that has the structure of a stochastic” [15]. Before the Monte Carlo method was introduced, simulations were employed in deterministic problems and applied statistical sampling to estimate uncertainties in the simulations. Physicist Enrico Fermi was an early user of what was later called a “Monte Carlo simulation” [16]. Fermi used Monte Carlo simulations to study the behavior of large numbers of neutrons. Ulam and Metropolis [17] also used the Monte Carlo technique to work on similar problems in nuclear physics. The name “Monte Carlo” was chosen by Von Neumann, a colleague of Ulam and Metropolis [18]. Hartley [19] said “… The fact that any distributional problem, whatever its analytic difficulties, can be solved by Monte Carlo Methods has undoubtedly a profound effect on the orientation of mathematical research…”. His advice was that rather than relying on the asymptotic distribution, one should simulate the distribution of the statistic with a computer, then compare this with the asymptotic distribution. Researchers can then decide if the asymptotic distribution is good enough.

North et al. [20] gave three reasons why we should Monte Carlo methods. First, test statistics do not always have a known asymptotic distribution. Second, for those cases when an asymptotic distribution exists, it should not be applied to smaller sample sizes. Third, calculating the exact sampling distribution may be too time-consuming. In contrast, an approximate p-value can be obtained by the Monte Carlo method without depending on an asymptotic distribution or exhaustive computation. Broman and Caffo [21] summarized how to calculate a simulation-based p-value. Let t be the observed statistic that follows some distribution, f . Let Y be a random variable following the distribution f . The parameter $p = \Pr(Y \geq t)$ is what needs to be estimated. Let $y_1, y_2, \ldots, y_m$ be an independent draw from f obtained by computer simulation and $r = \#\{i \mid y_i \geq t\}$. Then the conventional estimator of p is $\hat{p} = r/m$. This is applicable to the Skillings-Mack statistic because the null hypothesis will be rejected if t is “large” enough so that $\hat{p}$ is less than the specified significance level α .

Table 3. Uniform random numbers for the first sample.

	Block						
Treatment	1	2	3	4	5	6	7
A	0.603	0.116	0.370	0.899	0.547	0.316	0.500
B	0.352	0.366	0.041	0.605	0.055	0.529	0.602
C	0.639	0.980	NA	NA	0.781	0.188	0.578
D	0.265	0.953	0.233	0.517	0.847	0.485	NA
E	0.969	0.676	0.663	0.650	0.700	0.634	NA

Table 4. Ranks of Uniform random numbers for the first sample.

	Block						
Treatment	1	2	3	4	5	6	7
A	3 → 3.5	1	3	4	2	2	1
B	2	2	1	2	1	4	3
C	4 → 3.5	5	NA	NA	4	1	2
D	1	4	2	1	5	3	NA
E	5	3	4	3	3	5	NA

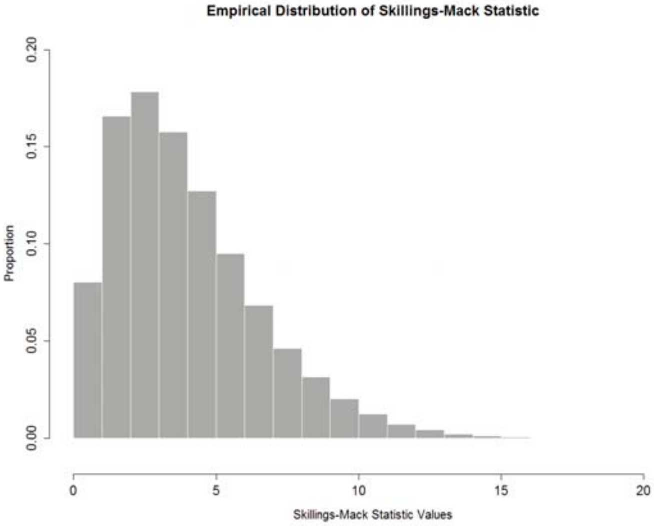

Fig. 4. Monte Carlo simulations of Skillings-Mack statistic for a RCBD with 4 missing, $k = 5, n = 7,$ and 2 ties

An example of the Monte Carlo simulation is conducted for a RCBD with $k = 5, n = 7$, and 4 missing observations: 2 missing values from treatments D and E in the 7th block, and the other 2 missing values from treatment C in the 3rd and 4th blocks. Suppose also that, in the observed data there were ties for treatments A and C in the first block. Both have the average rank of 3.5. Under the null hypothesis of equality of treatment effects, a large number of simulated data sets were generated which is $B = 200,000$ in this example. The first data set was generated by first using uniform random numbers which are shown in Table 3. In this table, generation of random numbers is not applied to the missing data to preserve the missing-data structure. The uniform random numbers are then ranked within each block, and to preserve the tie-data structure, ranks of 3 and 4 in the first block are changed to 3.5 as seen in Table 4. For this sample, the Skillings-Mack test statistic is calculated using the *Ski.Mack* function. With 200,000 sets of uniform random numbers, 200,000 Skillings-Mack's statistic values are calculated, and its empirical distribution is presented in Fig. 4.

4 Comparisons

Comparisons of the p-values based on the exact distribution of the Skillings-Mack statistic, the Monte Carlo method, and the chi-squared approximation produced by *Ski.Mack()* are shown in Table 5. We use the exact distribution to find the critical values corresponding to the upper tail probabilities closet to .10 and .01, and then use those critical values to find the upper tail probabilities from the Monte Carlo method and the chi-squared approximation. From Table 5, given a particular statistic value the upper tail probability from the chi-squared approximation is at least larger than those from other 2 methods. This indicates that for a specified significance level α , the α -level critical value from the chi-squared distribution is greater than the actual α -level critical value obtained from the exact distribution. We might think that the chi-squared distribution tends to give critical values that are too large. Thus, the actual significance level will be smaller than the nominal significance level α . Skillings and Mack [5] suggested that one should not use the chi-squared distribution to obtain critical values for significance levels near .01 or less. However, from Table 5 for RCBDs with 2 missing observations, it is found that every situation for $n = 3$ using the chi-squared distribution for calculating critical values should be avoided even for high significance levels near .10. Therefore, as the number of missing observations increases in small RCBDs, the greater the concern when using a chi-squared approximation. We strongly recommend using the Monte-Carlo whenever possible. As in Table 5, the estimated p-values based on the Monte Carlo method are considerably closer to the exact p-values. The Monte Carlo method with the default $B = 10,000$ replications is capable of providing accurate and precise estimates of p-values for the designs presented in Table 5. For larger designs, we recommend that users should increase the number of replications in the Monte Carlo method (*B* in the *Ski.Mack* function).

5 Small Sample Tables

In this section, extended tables of critical values the Skillings-Mack statistic are presented. Skillings and Mack [5] gave the table of critical values for certain RCBDs having a single missing observation. In our paper, however, the small sample tables are presented for certain RCBDs that have 2 missing observations. Tables are presented for three cases: (1) two missing values in the same block shown in Table 6; (2) two missing values for the same treatment shown in Table 7; and (3) two missing values occurring in different treatments and blocks shown in Table 8. These tables contain critical values for upper tail probabilities closest to .01, .05 and .10. The exact randomization distribution was used to obtain critical values for small designs, i.e. $k=3, n\leq 9$; $k=4, n\leq 5$; $k=5, n\leq 4$. The Monte Carlo method was used to estimate critical values for large designs which are denoted with asterisks. In those tables, B=200,000 simulations were used in *Ski.Mack()* for estimating critical values. Thus, we are confident that the estimated critical values are very close to the exact values even for large designs.

6 Summary and Conclusion

The major contributions of this paper are to present tables of critical values for block designs with 2 missing values and to provide an R program which provides p-values based on the Monte Carlo method for block designs. The nonparametric Skillings-Mack test is a procedure that is not as well-known as it should be. The primary reason has been the lack of a program to implement the method especially the free software like R; the *Ski.Mack()* function directly addresses this issue.

Above all, we have tried to point out that the chi-squared approximations are not sufficient for the Skillings-Mack test in small designs. With two missing values in small designs, the estimated critical values obtained by the chi-squared distribution tend to be too large not only for significance levels near .01 but also for .10. From these results, it appears that as the number of missing values increases in a design, the chi-squared approximation becomes less accurate in estimating critical values. In comparison, with only B=10,000 in the Monte Carlo simulation, the estimated p-values are very close to the exact p-values.

With two missing observations in block designs, researchers can now easily use the tables of critical values provided in this paper. For other cases such as balanced and partially balanced incomplete block designs, the *Ski.Mack()* function in the *Skillings.Mack* package is capable of estimating the p-values based on the Monte Carlo method and the chi-squared distribution. For large designs, the number of replications in the Monte Carlo simulation should be increased from the default, B=10,000.

References

1. Friedman, M.: The use of ranks to the avoid the assumption of normality Implicit in the analysis of variance. Journal of the American Statistical Association. 32, 675-701 (1937)
2. Durbin, J.: Incomplete blocks in Ranking Experiment. British Journal of Statistical Psychology. 4, 85-90 (1951)
3. Benard, A. and Van Elteren, P.: A generalization of the method of m rankings. In: Koninklijke Akademie van Wetenschappen. International Mathematical Congress Amsterdam, Series A. 56, 358-369 (1954)
4. Brunden, M.N. and Mohberg, N.R.: The Benard-Van Elteren statistic and nonparametric computation. Communications in Statistic-Simulation and Computation. 5, 155-162 (1976)
5. Skillings, J.H. and Mack, G.A.: On the use of a Friedman-type statistic in balanced and unbalanced block designs. Journal of the Royal Statistical Society. 23, 171-177 (1981)
6. Chatfield, M. and Mander, A.: The Skillings-Mack test (Friedman test when there are missing data). The Stata journal. 9, 299-305 (2009)
7. Cunningham, M.: A nonparametric method to assess treatment effects for unbalanced designs using SAS/IML. SAS Global Forum (2010)
8. Bi, J.: Computer-intensive methods for sensory data analysis, exemplified by Durbins rank test. Food Quality and Preference. 20, 195-202 (2009)
9. Fawcett, R.F. and Salter, K.C.: Distributional studies and the computer: An analysis of Durbin's rank test. The American Statistician. 41, 81-83 (1987)
10. R Core Team. R: A language and environmental for statistical computing (2015)
11. Srisuradetchai, P.: Skillings.Mack: The Skillings.Mack test statistic for block designs with missing observations. R package version 1.10 (2015), `http://CRAN.R-project.org/package=Skillings.Mack`
12. Fisher, R.A.: The design of experiments. 9th ed. Macmillan Pub Co (1971)
13. Pitman, E.J.G.: Significance tests which may be applied to samples from any populations. Technometrics. 4, 119-130 (1937)
14. Kennedy, P.E.: Randomization tests in econometrics. Journal of Business & Economic Statistics. 13, 85-94 (1995)

15. Kalos, M.H. and Whitlock, P.A.: Monte Carlo methods. 2nd ed. Darmstadt, Wiley-VCH Verlag GmbH & Co (2008)
16. Hubbard, D.W.: How to measure anything: Finding the Value of “Intangibles” in Business. Hoboken. John Wiley & Sons Inc (2010)
17. Metropolis, N. and Ulam, S.: The Monte Carlo method. Journal of the American Statistical Association. 44, 335-341 (1949)
18. Metropolis, N.: The beginning of the Monte Carlo method. Los Alamos Science, 125 (1987)
19. Hartley, H.O.: The impact of computers on statistics. In: D.B. Owens, ed. 1976. On the history of statistics and probability, pp. 421-442. Marcel Dekker, New York (1976)
20. North, B.V., Curtis, D. and Sham, P.C.: A note on the calculation of empirical P values from Monte Carlo procedures. American journal of human genetics. 71, 439-441 (2002)
21. Broman, K.W., Caffo, B.S.: Simulation-based P values: response to North et al.. American journal of human genetics. 72, 496 (2003)

Table 5. Comparisons of p-values.

Statistic value	Exact^a	MC^b	χ^{2c}	Statistic value	Exact^a	MC^b	χ^{2c}
1 missing observation							
k = 4, *n* = 3				*k* = 4, *n* = 4			
5.936	.1007	.0987	.1147	5.920	.1013	.1015	.1155
7.400	.0104	.0111	.0602	8.876	.0098	.0084	.0310
k = 5, *n* = 3				*k* = 5, *n* = 4			
7.167	.1001	.0975	.1273	7.439	.0999	.1060	.1144
9.589	.0101	.0095	.0480	10.717	.0100	.0098	.0299
Statistic value	Exact^d	MC^b	χ^{2c}	Statistic value	Exact^d	MC^b	χ^{2c}
2 missing observations within the same block							
k = 4, *n* = 3				*k* = 4, *n* = 4			
5.600	.1076	.1091	.1328	5.928	.1098	.1125	.1151
6.959	.0104	.0100	.0732	8.392	.0100	.0102	.0385
k = 5, *n* = 3				*k* = 5, *n* = 4			
6.992	.0996	.0982	.1362	7.379	.1098	.1030	.1172
9.092	.0094	.0102	.0588	10.485	.0100	.0104	.0330
2 missing observations occurring in the same treatment							
k = 4, *n* = 3				*k* = 4, *n* = 4			
5.753	.0972	.0980	.1242	5.900	.1000	.1050	.1166
6.940	.0139	.0155	.0738	8.557	.0098	.0102	.0358
k = 5, *n* = 3				*k* = 5, *n* = 4			
7.039	.1031	.1021	.1338	7.306	.1001	.0979	.1206
9.158	.0101	.0093	.0572	10.477	.0100	.0108	.0331
2 missing observations occurring in different treatments and blocks							
k = 4, *n* = 3				*k* = 4, *n* = 4			
5.767	.0926	.0965	.1235	5.902	.0995	.0961	.1165
6.840	.0139	.0126	.0772	8.601	.0102	.0086	.0351
k = 5, *n* = 3				*k* = 5, *n* = 4			
7.056	.0997	.1012	.1329	7.347	.1000	.0997	.1186
9.066	.0101	.0110	.0595	10.491	.0100	.0089	.0329

[a] Exact p-value from Table 3 in Skillings and Mack [5]

[b] Approximate p-value produced by *Ski.Mack()* using a default *B* = 10,000

[c] Chi-squared approximation produced by *Ski.Mack()*

[d] Exact p-value produced by a randomization test

Table 6. Critical values from the null distribution of T for two missing observations in the same block (for each critical value t_α, the parenthesized value is $\alpha = P_0(T \geq t_\alpha)$).

***k* = 4**

n = 3	*n* = 4	*n* = 5	*n* = 6*	*n* = 7*	*n* = 8*	*n* = 9*	*n* = 10*	*n* = 11*
5.600(.1076)	5.928(.1098)	6.055(.1000)	6.140(.1004)	6.093(.0999)	6.147(.1003)	6.150(.0999)	6.177(.0999)	6.157(.1001)
6.379(.0382)	6.792(.0506)	7.285(.0499)	7.335(.0500)	7.438(.0507)	7.455(.0503)	7.563(.0501)	7.548(.0502)	7.605(.0500)
6.959(.0104)	8.392(.0100)	9.231(.0101)	9.811(.0100)	10.054(.0101)	10.328(.0101)	10.426(.0100)	10.483(.0100)	10.626(.0100)

***k* = 5**

n = 3	*n* = 4	*n* = 5*	*n* = 6*	*n* = 7*	*n* = 8*	*n* = 9*	*n* = 10*	*n* = 11*
6.992(.0996)	7.379(.0999)	7.465(.1000)	7.549(.1000)	7.596(.1001)	7.617(.1000)	7.637(.1000)	7.660(.1000)	7.664(.1000)
7.892(.0499)	8.457(.0502)	8.778(.0501)	8.944(.0500)	9.017(.0501)	9.096(.0500)	9.125(.0500)	9.193(.0500)	9.222(.0500)
9.092(.0094)	10.485(.0100)	11.157(.0100)	11.610(.0100)	11.915(.0100)	12.128(.0100)	12.219(.0100)	12.424(.0100)	12.518(.0100)

***k* = 6**

n = 3*	*n* = 4*	*n* = 5*	*n* = 6*	*n* = 7*	*n* = 8*	*n* = 9*	*n* = 10*	*n* = 11*
8.368(.1000)	8.736(.1000)	8.890(.1000)	8.955(.1000)	9.014(.1000)	9.027(.1000)	9.061(.1000)	9.100(.1000)	9.083(.1000)
9.375(.0500)	9.967(.0500)	10.263(.0500)	10.442(.0500)	10.578(.0500)	10.622(.0500)	10.686(.0500)	10.739(.0500)	10.740(.0500)
10.925(.0100)	12.253(.0100)	12.920(.0100)	13.368(.0100)	13.685(.0100)	13.930(.0100)	13.988(.0100)	14.097(.0100)	14.162(.0100)

*These critical values were obtained by the Monte Carlo method.

Table 7. Critical values from the null distribution of T for two missing observations for the same treatment (for each critical value t_α, the parenthesized value is $\alpha = P_0(T \geq t_\alpha)$).

***k* = 3**

n = 3	*n* = 4	*n* = 5	*n* = 6	*n* = 7
3.847(.2500)	4.979(.0972)	4.615(.0995)	4.875(.0968)	4.674(.1007)
3.980(.0833)	5.786(.0417)	5.764(.0486)	6.000(.0505)	5.875(.0483)
-	5.971(.0139)	7.753(.0069)	7.973(.0120)	8.072(.0089)

n = 8	*n* = 9	*n* = 10*	*n* = 11*
4.364(.0987)	4.697(.1012)	4.713(.0988)	4.517(.0985)
5.893(.0502)	5.635(.0494)	6.027(.0519)	6.000(.0498)
8.247(.0110)	8.297(.0102)	8.694(.0100)	8.517(.0099)

***k* = 4**

n = 3	*n* = 4	*n* = 5	*n* = 6*	*n* = 7*
5.753(.0972)	5.900(.1000)	6.159(.0971)	6.130(.0999)	6.055(.0999)
5.867(.0417)	7.010(.0509)	7.215(.0499)	7.345(.0499)	7.411(.0498)
6.940(.0139)	8.557(.0098)	9.270(.0099)	9.794(.0099)	10.066(.0100)

n = 8*	*n* = 9*	*n* = 10*	*n* = 11*
6.107(.1001)	6.143(.1001)	6.148(.1000)	6.184(.1000)
7.533(.0498)	7.523(.0500)	7.563(.0500)	7.604(.0501)
10.253(.0100)	10.380(.0100)	10.532(.0100)	10.652(.0100)

***k* = 5**

n = 3	*n* = 4	*n* = 5*	*n* = 6*	*n* = 7*
7.039(.1031)	7.306(.1001)	7.467(.1000)	7.540(.1000)	7.590(.1000)
7.866(.0497)	8.433(.0501)	8.753(.0500)	8.910(.0500)	9.028(.0500)
9.158(.0101)	10.477(.0100)	11.151(.0100)	11.631(.0100)	11.911(.0100)

n = 8*	*n* = 9*	*n* = 10*	*n* = 11*
7.580(.1000)	7.638(.1000)	7.656(.1000)	7.660(.1000)
9.077(.0500)	9.160(.0500)	9.185(.0500)	9.196(.0500)
12.101(.0100)	12.279(.0100)	12.416(.0100)	12.479(.0100)

***k* = 6**

n = 3*	*n* = 4*	*n* = 5*	*n* = 6*	*n* = 7*
8.260(.0999)	8.686(.1000)	8.871(.1000)	8.936(.1000)	8.998(.1000)
9.260(.0501)	9.924(.0500)	10.235(.0500)	10.423(.0500)	10.549(.0500)
10.968(.0100)	12.193(.0100)	12.884(.0100)	13.300(.0100)	13.653(.0100)

n = 8*	*n* = 9*	*n* = 10*	*n* = 11*
9.033(.1000)	9.049(.1000)	9.078(.1000)	9.083(.1000)
10.614(.0500)	10.658(.0500)	10.723(.0500)	10.748(.0500)
13.839(.0100)	13.962(.0100)	14.171(.0100)	14.211(.0100)

*These critical values were obtained by the Monte Carlo method.

Table 8. Critical values from the null distribution of T for two missing observations that are in different treatments and different blocks (for each critical value t_α, the parenthesized value is $\alpha = P_0(T \ge t_\alpha)$).

***k* = 3**

n = 3	*n* = 4	*n* = 5	*n* = 6	*n* = 7
3.857(.1667)	4.250(.1250)	4.332(.1134)	4.671(.1049)	4.540(.0930)
-	5.694(.0417)	6.057(.0486)	5.661(.0505)	5.801(.0522)
-	5.833(.0278)	7.679(.0069)	7.973(.0104)	8.484(.0101)

n = 8	*n* = 9	*n* = 10*	*n* = 11*
4.863(.0999)	4.614(.1008)	4.429(.1018)	4.791(.0992)
5.673(.0524)	6.073(.0503)	5.755(.0493)	5.779(.0491)
8.348(.0100)	8.714(.0101)	8.371(.0099)	8.700(.0097)

***k* = 4**

n = 3	*n* = 4	*n* = 5	*n* = 6*	*n* = 7*
5.767(.0926)	5.902(.0995)	6.091(.0998)	6.148(.1001)	6.142(.1000)
6.117(.0463)	6.891(.0498)	7.214(.0500)	7.293(.0500)	7.431(.0500)
6.840(.0139)	8.601(.0102)	9.252(.0100)	9.833(.0100)	10.108(.0100)

n = 8*	*n* = 9*	*n* = 10*	*n* = 11*
6.146(.1000)	6.198(.1000)	6.135(.1000)	6.209(.1000)
7.546(.0500)	7.617(.0500)	7.562(.0500)	7.640(.0500)
10.296(.0100)	10.430(.0100)	10.487(.0100)	10.625(.0100)

***k* = 5**

n = 3	*n* = 4	*n* = 5*	*n* = 6*	*n* = 7*
7.056(.0997)	7.347(.1000)	7.505(.1000)	7.551(.1000)	7.580(.1000)
7.908(.0498)	8.460(.0500)	8.776(.0500)	8.930(.0500)	9.036(.0500)
9.066(.0101)	10.491(.0100)	11.183(.0100)	11.592(.0100)	11.921(.0100)

n = 8*	*n* = 9*	*n* = 10*	*n* = 11*
7.596(.1000)	7.639(.1000)	7.651(.1000)	7.663(.1000)
9.081(.0500)	9.160(.0500)	9.185(.0500)	9.196(.0500)
12.101(.0100)	12.311(.0100)	12.378(.0100)	12.479(.0100)

***k* = 6**

n = 3*	*n* = 4*	*n* = 5*	*n* = 6*	*n* = 7*
8.438(.1000)	8.726(.1000)	8.871(.1000)	8.969(.1000)	9.001(.1000)
9.397(.0500)	9.956(.0500)	10.259(.0500)	10.493(.0500)	10.543(.0500)
10.926(.0100)	12.242(.0100)	12.922(.0100)	13.408(.0100)	13.658(.0100)

n = 8*	*n* = 9*	*n* = 10*	*n* = 11*
9.020(.1000)	9.070(.1000)	9.070(.1000)	9.087(.1000)
10.607(.0500)	10.697(.0500)	10.732(.0500)	10.749(.0500)
13.858(.0100)	14.052(.0100)	14.094(.0100)	14.212(.0100)

*These critical values were obtained by the Monte Carlo method.