Journal of Statistics & Applied Sciences, Issue -10
ISSN 2718-0999

*Research Article*

Received: date:02.09.2024
Accepted: date:24.10.2024
Published: date:31.06.2024

# Examination of Statistical Methods Used in Group Comparisons in Visual Analogue Scale (VAS) Data in Terms of Type I Error Rate and Power of Test: Monte Carlo simulation study for VAS data

**Özgür Koşkan[1], Malik Ergin[2*]**

[1]Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science, Isparta, Türkiye; ozgurkoskan@isparta.edu.tr
[2]Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science, Isparta, Türkiye; malikergin@isparta.edu.tr
Orcid: 0000-0002-5089-6250[1] Orcid: 0000-0003-1810-6754[2]
*Correspondence: malikergin@isparta.edu.tr

**Abstract:** In this study, how the application changes with the size of the scale, type I error rate and power of test values were examined in different sample sizes. The material of this study is the random numbers generated according to different sample sizes and different group means of standard deviation out of populations that hold multinomial distribution. In the study, permutation test, F-test and Kruskal-Wallis (KW) test were examined using combinations with different sample sizes (n = 5, 10, 20, 30, 50) and different p values (probability of occurrence of events which are 0.25, 0.50, 0.75) first interms of Type I error rate and then power of the test using different group averages (Δ =1.0, 1.5, 2.0). As a result of the simulations, it is seen that with small sample sizes, Kruskal-Wallis test was unable to maintain Type I error rate at 0.05 level. In the evaluation of such Likert-type data, it can be stated that, permutation test, one of the distribution free tests, is more practical than other tests in terms of maintaining the Type I error rate at 0.05 level and high power of test values. The permutation test is satisfactory in terms of both the type I error rate and the power of the test. Also, permutation test is a distribution-free test. Therefore, can be used without prerequisites. In almost every combination (π, Δ, k, n), permutation test had similar or superior type 1 error and power values than the F and KW tests. It was observed that compared to the 5 visual analogue scale, when the data is measured in 10 and 20 visual analogue scale, the power of test values decreased. In other words, if more than 5 visual analogue scale are made on the visual analogue scale data, the results would negatively be affected.

**Keywords:** quantitized data; likert-type data; permutation test; monte carlo simulation; power of test values

## 1. Introduction

Visual analogue scale (VAS) data are the data sets obtained from scoring verbal predictions. The most widely used visual analogue scales are 5 and 10-point scales however larger scales are also often used too. The points can vary from no pain, little bit pain, medium pain, much pain, and very much pain. For evaluation, analysis is conducted by attributing numerical values to these expressions.

The technique's apparent simplicity and adaptability to a wide range of research settings have made it an appealing measurement option. VAS is preferred by researchers as it is simple, quick, easily understood by untrained staff and subjects, and allows the use of numerical values suitable for statistical analysis. Patient satisfaction is one of the most important evaluations and it will be even more important in the future. The scale needs to be practical, objective, and applicable for the measurement. In the literature, the distribution of the obtained data is reported to show a discrete uniform distribution [1-3].

**Citation :** Ö. Koşkan and M. Ergin, "Examination of Statistical Methods Used in Group Comparisons in Visual Analogue Scale (VAS) Data in Terms of Type I Error Rate and Power of Test: Monte Carlo Simulation Study for VAS Data," *Journal of Statistics and Applied Sciences*, no. 10, pp. 95–104, Dec. 2024, doi:10.52693/jsas.1542318.

The discrete uniform distribution of data, such as VAS values, usually has the assumption that each value will be selected from the sample with equal probability in a given draw. The probability mass function of the discrete uniform distribution is given by P(X=x)=1/(n+1),x=0.1,….n, which means that X can take any integer value between 0 and n with equal probability. The mean and variance of the distribution are n/2 and (n(n+2))/12, respectively. To generate a random number from the discrete uniform distribution, a random number (R) can be drawn from the U(0,1) distribution, S=(n +1)R can be calculated, and the integer part of S can be drawn from the discrete uniform distribution [4]. In the literature, the power of ANOVA and PermANOVA tests are generally tested with symmetrical distributions such as normal distribution, t-distribution, and curved distributions such as Beta, Gamma and Chi-square, and both type I error and power of test values are tested. As it is known, all of these distributions are continuous distributions.

In fact, in the studies it was assumed that the distribution of VAS values approached normal [5-8]. However, likert type variables such as VAS values are discrete variables and show a discrete distribution. Likert types such as having 5 or 10 scale values are data sets with discrete uniform distribution. This study, which compares the methods used in group comparisons such as PermANOVA, Kruskall_Wallis, on the discrete uniform distribution obtained in different ways by changing the parameters, will shed light on the literature. Because, in practice, these tests are still widely used on data obtained in likert type, such as VAS values.

There are many studies showing that PermANOVA and Kruskal-Wallis tests are distribution-independent tests and are also used in ordinal data [9-12]. It is thought that choosing the distribution as discrete uniform and examining these tests in terms of both type I error and power of the test will provide an innovative approach to the literature.

Based on these principles, VAS data show a discrete uniform distribution. As it is known, F, Kruskal-Wallis (KW), and permutation tests are commonly used in group comparisons. If the data meets the prerequisites of parametric tests, the F test is used, if not, the KW test is used. The permutation test is a resampling method and does not require any prerequisites. With the development of cheap and fast computers, the permutation tests now tend to be used more often as fast computers are needed to compute these powerful tests whose calculations are easy. One of the advantages of the resampling method is that there is not much need for too much knowledge in mathematical and statistical formulas. Enough knowledge to ensure an understanding of the concepts and methods is sufficient for the application of the resampling approach [13]. While the expected results cannot be obtained using traditional methods, they can be obtained using the resampling method [14]. In addition, due to the limitations of traditional parametric and non-parametric significance tests, permutation testing is considered an important alternative [15].

The purpose of this study was to examine how type I error rate and power of test values change in different sample sizes when statistical methods comparing groups are used for visual analogue scale data.

**2. Materials and Methods**

The material of this study consisted of random numbers in three groups, generated using Microsoft Power Station Developer Studio and IMSL Library in terms of different sample sizes and different group means of standard deviation out of populations that hold discrete uniform distribution. Permutation, F, and Kruskal-Wallis (KW) tests were examined in terms of type I error rate and power of the test by using combinations with different sample sizes (n = 5, 10, 20, 30, 50), $\pi$ values (probability of occurrence of events which were 0.25, 0.50, 0.75) and group means ($\Delta$ =1.0, 1.5, 2.0).In the study, 100.000 simulations were performed for each sample size (n) and scale (k). Among these, the distribution of random numbers was showed below with an example for n=50 and k=5, 10, and 20. These distributions represent scenarios that could be observed. The three distributions given in the figures are presented as representative in Figure 2. Each sample was drawn completely randomly from the population 100.000 times. Consequently, due to this random sampling, the distribution of VAS values will have varying frequencies in each sample. Similarly, by adjusting the distribution's parameter 'p' up and down, the mean and standard deviation of VAS values decrease and increase. This allows for the creation of situations that are closer to reality.

Permutation Test: The permutation test is a resampling method that does not require any prerequisites. The permutation test is also called the distribution-free test, which does not require the normality assumption for the distribution of data. The permutation test can be an alternative to the F test, especially in studies with small sample sizes [16]. In the study, after the F test table value calculated for 3 groups was found, the F value was calculated again by resampling which means grouping the data of these three groups again randomly. When the F value was equal to or greater than the first calculated F value, it was counted. The total counted F value was divided to resampling iterated 10000 times to find the percentage ratio of the F value. When the percentage ratio of the F value was either equal to or smaller than 0.05, the $H_0$ hypothesis is rejected by the permutation test. F test (ANOVA): As it is known, the variance analysis technique is the most widely used statistical method to examine whether the difference between the means of two or more independent groups happens coincidentally or not. Let formulize ($i$=1,….,$n_j$) and ($j$=1,….,$k$) and $\sum nk$=N to show the $X_{ij}$ as $i^{th}$ observation in the $J^{th}$ treatment group. It is assumed that the $X_{ij}$ here is distributed independently and normally with $\mu_j$ and $\sigma^2$ parameters. In this case, $\mu_j$ and $\sigma^2$ are the best linear prediction estimators. The data obtained at the end of the study can be identified with the $Y_{ij}=\mu+\alpha_i+e_{ij}$ model. In this mathematical model; μ; is the overall mean, that is the population mean, $\alpha_i$; is the effect of $i_{th}$ treatment $e_{ij}$; is the error term. It is assumed that all $e_{ij}$ are independent of each other with normal distribution, and $\sigma_j^2$ is pooled variance. For this purpose, to examine whether there is a difference between the aforementioned group means or not, the following control and alternative hypothesis need to be tested as:

$H_0$: $\mu_1=\mu_2=\ldots..=\mu_k$

$H_1$: "the difference between the means of at least two groups is significant" and calculated as:

$$\textbf{\textit{Total Sum of Square}} = \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\bar{X})^2 \quad (1)$$

$$\textbf{\textit{Between Sum of Square}} = \sum_{i=1}^{k} n_i * (\bar{X}_i-\bar{X})^2 \quad (2)$$

$$\textbf{\textit{Within Sum of Square}} = \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\bar{X})^2 \quad (3)$$

$F=\frac{\sum_{i=1}^{k} n_i*(\bar{X}_i-\bar{X})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\bar{X})^2}$ that calculated value (k-1) and (N-k) is tested with F table value with degrees of freedom to accept or reject the $H_0$ hypothesis.

Kruskal-Wallis Test

It is well known that Kruskal-Wallis test is the most widely used statistical technique to examine whether the difference between the median of two or more independent groups stems from coincidence or not. The statistics applied for the Kruskal-Wallis test is H-type statistics. To apply this kind of statistics, the units of the experiment are supposed to be randomized through the population, observations are independent of each other, and the group rank means be compared using the rank values of these data. The following control and alternative hypothesis need to be tested.

$H_0$: $M_1=M_2=\ldots..=M_k$

$H_1$: "the difference between the median of at least two groups is significant".

$$H=\frac{12}{N(N+1)} * \sum_{j=1}^{k}\left(\frac{(\sum R_j)^2}{n_j}\right) - 3(N+1) \quad (4)$$

The calculated H statistics (k-1) value is checked using the table value of chi-square with degrees of freedom. Later, it is determined whether the difference between the rank means of the groups, in other words, medians is significant or not [9].

Power of test and type I error rate

Two types of error occur in the controlling hypothesis when the control hypothesis is decided by being tested with the alternative hypothesis. At the end of the controls, when the correct control hypothesis is rejected, this error is called the type I error. The second type error is the one that occurs in case the alternative hypothesis is actually true and when the control hypothesis is accepted. We obtain the power of test by subtracting this error from 1 [20]. At the end of 100 000 simulations performed in the study, type I error and power of test values were found.

**3. Results**

Type I error rate values for permutation, F, and KW tests are provided in Table 1. When Table 1 is examined, as it is observed in all scales (k) when the sample size (n) was 5, KW test was unable to keep the type I error rate at 0.05 level as determined at the beginning and had lower values. This case also continued relatively less when (n) was 10. However, when the sample size was larger than 20, the type I error rate determined at the beginning was maintained. The changes in the distribution of the probability of event occurrence (π) did not affect the type I error. In small sample sizes, while the F test maintained the type I error at 0.05 level, KW test did not maintain the type I error at 0.05 level and had values lower than 0.05. When the probability of event occurrence (π) was 0.25, the scale (k) was 5 and the sample size (n) was 5, the type I error rate for F, KW, and permutation test were 0.051, 0.042, and 0.050, respectively. This shows that the type I error rate at the beginning could not be maintained as 0.05 in KW test. A similar situation was observed when the sample size was 10; where type I error rate values were 0.054, 0.043, and 0.051 for F, KW, and permutation test respectively. Other cases also gave similar results for other π values. For sample size 20 and over, the type I error rate was maintained at 0.05 for F, KW, and permutation test.

**Table 1.** Type I error rate values for F, KW, and permutation tests.

| π | k | n | F | KW | Permutation |
|---|---|---|---|---|---|
| 0.25 | 5 | 5 | 0.051 | 0.042 | 0.050 |
| | | 10 | 0.052 | 0.046 | 0.049 |
| | | 20 | 0.051 | 0.049 | 0.051 |
| | | 30 | 0.050 | 0.046 | 0.052 |
| | | 50 | 0.052 | 0.050 | 0.050 |
| | 10 | 5 | 0.054 | 0.043 | 0.051 |
| | | 10 | 0.052 | 0.046 | 0.050 |
| | | 20 | 0.051 | 0.048 | 0.050 |
| | | 30 | 0.051 | 0.049 | 0.049 |
| | | 50 | 0.050 | 0.050 | 0.051 |
| | 20 | 5 | 0.056 | 0.043 | 0.052 |
| | | 10 | 0.053 | 0.047 | 0.051 |
| | | 20 | 0.051 | 0.048 | 0.050 |
| | | 30 | 0.051 | 0.049 | 0.050 |
| | | 50 | 0.051 | 0.049 | 0.050 |
| 0.50 | 5 | 5 | 0.051 | 0.041 | 0.052 |
| | | 10 | 0.053 | 0.047 | 0.049 |
| | | 20 | 0.050 | 0.048 | 0.051 |
| | | 30 | 0.051 | 0.049 | 0.050 |
| | | 50 | 0.053 | 0.050 | 0.050 |
| | 10 | 5 | 0.054 | 0.041 | 0.052 |
| | | 10 | 0.051 | 0.046 | 0.051 |
| | | 20 | 0.050 | 0.048 | 0.049 |
| | | 30 | 0.049 | 0.049 | 0.051 |
| | | 50 | 0.050 | 0.049 | 0.050 |
| | 20 | 5 | 0.055 | 0.042 | 0.054 |
| | | 10 | 0.052 | 0.047 | 0.052 |
| | | 20 | 0.051 | 0.049 | 0.050 |
| | | 30 | 0.051 | 0.049 | 0.051 |
| | | 50 | 0.051 | 0.049 | 0.050 |

| | | | | | |
|---|---|---|---|---|---|
| 0.75 | 5 | 5 | 0.050 | 0.041 | 0.049 |
| | | 10 | 0.051 | 0.046 | 0.050 |
| | | 20 | 0.052 | 0.048 | 0.051 |
| | | 30 | 0.050 | 0.049 | 0.051 |
| | | 50 | 0.053 | 0.049 | 0.051 |
| | 10 | 5 | 0.055 | 0.042 | 0.052 |
| | | 10 | 0.051 | 0.045 | 0.052 |
| | | 20 | 0.052 | 0.049 | 0.051 |
| | | 30 | 0.051 | 0.050 | 0.049 |
| | | 50 | 0.051 | 0.050 | 0.050 |
| | 20 | 5 | 0.055 | 0.042 | 0.053 |
| | | 10 | 0.052 | 0.046 | 0.051 |
| | | 20 | 0.050 | 0.047 | 0.049 |
| | | 30 | 0.051 | 0.049 | 0.050 |
| | | 50 | 0.050 | 0.049 | 0.050 |

Power of test values for F, KW, and permutation tests when Δ=1, Δ=1.5, and Δ=2 are provided in Table 2, Table 3, and Table 4, respectively. When Table 2, Table 3, and Table 4 are examined it is observed that the power of test was affected by changes in $\pi$ value. When one standard deviation difference (Δ) was created among group means (Table 2), and the power of test is examined, it is observed that resampling methods such as the permutation test and F test had higher values compared to the Kruskal Wallis test especially when the sample size was larger than 20. When looking at Tables 2, 3 and 4, in general, it can be said that compared to normally distributed data, VAS values showing discrete uniform distribution had lower power of test values in all three tests. In addition, it is shown that when $\pi$ value increased the difference between tests decreased (Table 3). However, it is also seen that the permutation test, known as an independent test, had greater power of test values than KW test in all cases (Table 2, 3, 4). When the difference among the group means was (Δ) 1.5 standard deviation, permutation, and F tests continued to be relatively more powerful in small sample sizes. However, this apparent situation decreased as the sample size increased in the groups.

When the difference among the group means was (Δ) 2 standard deviation, only in the cases where the sample size was 5 and 10, the permutation and F tests had higher power of the test values than KW test (Table 4). When the sample size increased, the difference between tests in terms of power of test decreased and when the sample size was 30, the power of test values of tests become similar. But it is observed that in all cases when evaluating VAS data, permutation, and F test had higher power of test values than KW test (Table 2, 3, 4). As can be seen in Tables 2, 3, and 4, where the power of test values are given, the power of all three tests relatively decreased as the $\pi$ value increased. When the overall results are examined, it can be inferred that KW test had lower power of test values in the analysis of sensory data, in this study VAS data. However, especially in small sample sizes, the permutation and F tests had higher power of test values. It is seen that whatever the distribution is, as the difference among the group means increases, the power of all three tests increases. As it is known, the larger the sample size gets, the higher the power values become.

Consequently, we can say that in visual analogue scale type data analysis, the resampling method, which is also known as distribution independent tests, is a suitable test. It should be favored instead of KW test, especially in visual analogue scale data such as VAS. After an overall examination of this study, it is concluded that in almost every case, the permutation method, which is one of the resampling methods can be favored for visual analogue scale type data in terms of both type I error rate and power of test values.

**Table 2.** Empirical Test Power for Samples from Multinomial Distributions (Δ=1)

| **Δ** | **p** | **k** | **n** | **F** | **KW** | **Permutation** |
|---|---|---|---|---|---|---|
| | | 5 | 5 | 0.136 | 0.110 | 0,129 |
| | | | 10 | 0.542 | 0.433 | 0.524 |
| | | | 20 | 0.971 | 0.946 | 0.968 |
| | | | 30 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1.0 | 0.25 | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.070 | 0.057 | 0.066 |
| | | | 10 | 0.151 | 0.141 | 0.148 |
| | | | 20 | 0.600 | 0.520 | 0.061 |
| | | | 30 | 0.960 | 0.920 | 0.957 |
| | | | 50 | 0.999 | 0.999 | 0.999 |
| | | 20 | 5 | 0.059 | 0.045 | 0.053 |
| | | | 10 | 0.076 | 0.069 | 0.074 |
| | | | 20 | 0.171 | 0.158 | 0.174 |
| | | | 30 | 0.363 | 0.329 | 0.375 |
| | | | 50 | 0.824 | 0.716 | 0.827 |
| | 0.50 | 5 | 5 | 0.148 | 0.121 | 0,151 |
| | | | 10 | 0.697 | 0.645 | 0.669 |
| | | | 20 | 0.994 | 0.983 | 0.994 |
| | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.085 | 0.068 | 0.076 |
| | | | 10 | 0.249 | 0.217 | 0.236 |
| | | | 20 | 0.500 | 0.450 | 0.485 |
| | | | 30 | 0.930 | 0.882 | 0.923 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 20 | 5 | 0.063 | 0.049 | 0.056 |
| | | | 10 | 0.088 | 0.082 | 0.088 |
| | | | 20 | 0.216 | 0.189 | 0.220 |
| | | | 30 | 0.166 | 0.330 | 0.485 |
| | | | 50 | 0.881 | 0.834 | 0.881 |
| | 0.75 | 5 | 5 | 0.151 | 0.120 | 0.141 |
| | | | 10 | 0.434 | 0.423 | 0.430 |
| | | | 20 | 0.994 | 0.983 | 0.993 |
| | | | 30 | 0.990 | 0.990 | 0.999 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.077 | 0.059 | 0.067 |
| | | | 10 | 0.173 | 0.160 | 0.179 |
| | | | 20 | 0.550 | 0.490 | 0.540 |
| | | | 30 | 0.880 | 0.830 | 0.879 |
| | | | 50 | 0.999 | 0.999 | 0.999 |
| | | 20 | 5 | 0.060 | 0.047 | 0.054 |
| | | | 10 | 0.078 | 0.070 | 0.074 |
| | | | 20 | 0.181 | 0.163 | 0.181 |
| | | | 30 | 0.343 | 0.311 | 0.340 |
| | | | 50 | 0.824 | 0.765 | 0.823 |

**Table 3.** Test Power for F, KW, and Permutation Tests with Standard Deviation Difference (Δ = 1.5)

| Δ | π | k | n | F | KW | Permutation |
|---|---|---|---|---|---|---|
| | 0.25 | 5 | 5 | 0.485 | 0.409 | 0.480 |
| | | | 10 | 0.870 | 0.775 | 0.867 |
| | | | 20 | 1.000 | 1.000 | 1.000 |
| | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.123 | 0.099 | 0.107 |
| | | | 10 | 0.278 | 0.254 | 0.275 |
| | | | 20 | 0.892 | 0.803 | 0.892 |
| | | | 30 | 0.999 | 0.997 | 0.999 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | | 5 | 0.063 | 0.049 | 0.061 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1.5 | | 20 | 10 | 0.105 | 0.095 | 0.102 |
| | | | 20 | 0.331 | 0.296 | 0.319 |
| | | | 30 | 0.631 | 0.585 | 0.627 |
| | | | 50 | 0.995 | 0.981 | 0.995 |
| | 0.50 | 5 | 5 | 0.248 | 0.212 | 0.220 |
| | | | 10 | 0.698 | 0.648 | 0.676 |
| | | | 20 | 1.000 | 1.000 | 1.000 |
| | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.106 | 0.083 | 0.099 |
| | | | 10 | 0.373 | 0.300 | 0.366 |
| | | | 20 | 0.898 | 0.852 | 0.900 |
| | | | 30 | 0.999 | 0.999 | 0.999 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 20 | 5 | 0.068 | 0.052 | 0.063 |
| | | | 10 | 0.127 | 0.145 | 0.125 |
| | | | 20 | 0.423 | 0.366 | 0.425 |
| | | | 30 | 0.840 | 0.738 | 0.833 |
| | | | 50 | 0.991 | 0.983 | 0.990 |
| | 0.75 | 5 | 5 | 0.151 | 0.121 | 0.160 |
| | | | 10 | 0.781 | 0.757 | 0.777 |
| | | | 20 | 1.000 | 1.000 | 1.000 |
| | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.077 | 0.059 | 0.078 |
| | | | 10 | 0.325 | 0.268 | 0.322 |
| | | | 20 | 0.919 | 0.869 | 0.918 |
| | | | 30 | 0.999 | 0.994 | 0.999 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 20 | 5 | 0.065 | 0.050 | 0.061 |
| | | | 10 | 0.105 | 0.093 | 0.105 |
| | | | 20 | 0.351 | 0.298 | 0.351 |
| | | | 30 | 0.652 | 0.606 | 0.651 |
| | | | 50 | 0.985 | 0.966 | 0.985 |

**Table 4.** Test Power for F, KW, and Permutation Tests with Standard Deviation Difference (Δ = 2)

| Δ | π | k | n | F | KW | Permutation |
|---|---|---|---|---|---|---|
| | 0.25 | 5 | 5 | 0.488 | 0.410 | 0.483 |
| | | | 10 | 0.959 | 0.873 | 0.954 |
| | | | 20 | 1.000 | 1.000 | 1.000 |
| | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.128 | 0.103 | 0.123 |
| | | | 10 | 0.464 | 0.399 | 0.460 |
| | | | 20 | 0.985 | 0.954 | 0.983 |
| | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 20 | 5 | 0.073 | 0.058 | 0.065 |
| | | | 10 | 0.155 | 0.134 | 0.151 |
| | | | 20 | 0.535 | 0.484 | 0.531 |
| | | | 30 | 0.894 | 0.845 | 0.889 |
| | | | 50 | 0.999 | 0.999 | 1.000 |
| | | 5 | 5 | 0.443 | 0.354 | 0.418 |
| | | | 10 | 0.998 | 0.999 | 0.997 |
| | | | 20 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2.0 | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | 0.50 | 10 | 5 | 0.179 | 0.145 | 0.167 |
| | | | 10 | 0.672 | 0.595 | 0.664 |
| | | | 20 | 0.992 | 0.980 | 0.992 |
| | | | 30 | 0.999 | 0.999 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 20 | 5 | 0.082 | 0.064 | 0.080 |
| | | | 10 | 0.198 | 0.179 | 0.194 |
| | | | 20 | 0.699 | 0.601 | 0.696 |
| | | | 30 | 0.980 | 0.937 | 0.980 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | 0.75 | 5 | 5 | 0.520 | 0.424 | 0.499 |
| | | | 10 | 0.970 | 0.940 | 0.971 |
| | | | 20 | 1.000 | 1.000 | 1.000 |
| | | | 30 | 1.000 | 1.000 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 10 | 5 | 0.150 | 0.117 | 0.138 |
| | | | 10 | 0.527 | 0.403 | 0.520 |
| | | | 20 | 0.994 | 0.983 | 0.993 |
| | | | 30 | 1.000 | 0.999 | 1.000 |
| | | | 50 | 1.000 | 1.000 | 1.000 |
| | | 20 | 5 | 0.076 | 0.059 | 0.071 |
| | | | 10 | 0.157 | 0.134 | 0.159 |
| | | | 20 | 0.572 | 0.489 | 0.566 |
| | | | 30 | 0.883 | 0.854 | 0.885 |
| | | | 50 | 1.000 | 0.999 | 1.000 |

**4. Discussion**

In the analysis of sensory data, for every case, permutation and F tests maintained the type I error rate at 0.05 level [21,22]. The study by Routledge showed similar results in terms type I error rate [23]. The study showed that KW test was not able to maintain the type I error rate at 0.05 level. It can be said that in the analysis of sensory data, each of the three tests had low power of test values. It is considered that this situation stems from the abrupt distribution of the data. Hence it is seen that in the studies in which other kinds of continuous distribution other than normal distributions are applied, the power values take higher values in similar simulations. Still, when the differences between the group mean in terms of standard deviation increase, realized the power of test values increase.

Similar results to this can also be observed in the studies by Başpınar and Gürbüz, Koşkan and Gürbüz, Mendeş and Tekindal, and Weber [18,22,24,25]. It is also stated in many research that the resampling method should be favored in cases where the sample size is small and the distribution is not normal [18,26-30].

As is seen in Table 1, the type I error rate of the permutation and F tests holds its determined value. However, in small sample sizes, the KW test cannot maintain the determined type I error value.

Keskin and Mendeş compared variance analysis and some approach tests (Marascuilo, James's second degree and Alexander - Govern Tests) in terms of power values that occur empirically in the samplings from populations with exponential distribution (1.00) [31]. In their study, the number of groups was 4 and 5, and the number of observations both equal and different in each group was between 3 and 100. The differences among the group means were 0.5, 1.0, 1.5, and 2 standard deviations (Δ). The power values of the tests considered for each combination were obtained out of 100000 simulation tests. Their results showed that the power of tests changed regarding the number of observations, whether the number of observations within the groups was equal or not, and the difference among the group means.

In a study reported by Vidoni which used 10 000 simulations, where the probability of occurrence was $\pi$=0.3, 0.5, and 0.7, the sample size was 5 and 10, it was concluded that different models should be used for discrete data when the sample size was 10 [32]. This agrees with our results. In order to have more reliable results in the evaluation of sensory data, it is essential to acquire the data with the help of an expert before the evaluation step and to work with more experts as much as possible. The problem remains that it is hard to find enough expert ratings. If this problem is eliminated, acquiring especially small-scale figures such as 5-point likert scales shows more reliable results. Another result is that for evaluating sensory data, resampling tests are alternatives for KW and F tests. Heller and Venkatraman, Ludbrook and Dudley, Good, Balasubramani et al., Pesarin and Salmaso, Koşkan and Gürbüz, Figueiredo reported that resampling approaches that are also named as distribution-free tests can be favored especially in small sample sizes instead of F test [18,26-30].

5. Conclusions

As a result of this study, it can be inferred that when comparing groups in visual analogue scale data, permutation test should be used. The permutation test is satisfactory in terms of both the type I error rate and the power of the test. Also, permutation test is a distribution-free test. Therefore, can be used without prerequisites.

In almost every combination ($\pi$, $\Delta$, k, n), permutation test had similar or superior type 1 error and power values than the F and KW tests. It was observed that compared to the 5 visual analogue scale, when the data is measured in 10 and 20 visual analogue scale, the power of test values decreased. In other words, if more than 5 visual analogue scale are made on the visual analogue scale data, the results would negatively be affected.

**Author Contributions:** Conceptualization, Ö.K.; methodology, Ö.K.; software, Ö.K.; validation, Ö.K., M.E.; formal analysis, Ö.K.; investigation, Ö.K., M.E.; resources,Ö.K.; data curation, Ö.K.; writing—original draft preparation, Ö.K.; writing—review and editing, M.E., Ö.K.; visualization, Ö.K.; supervision, Ö.K.; project administration, Ö.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

[1] D. L. Clason and T. J. Dormody, "Analyzing data measured by individual likert-type items," *J. Agric. Educ.*, vol. 35 no.4, pp. 31-35, 1994.
[2] K. A. A. Small, "Discrete choice model for ordered alternatives," *Econometrica*, vol. 55 no. 2, pp. 409-424, 1987.
[3] M. A. Garcia-Perez, "Are the steps on likert scales equidistant? Responses on visual analog scales allow estimating their distances," *Educ. Psychol. Meas.*, 1-32, 2023.
[4] S. Sinharay, Discrete Probability Distributions, McGaw Group Pty, 2010.
[5] O. Kavuncu, İstatistik Teorisi ve Teorik Dağılımlar, T.C. Ziraat Bank Publishing, 1995.
[6] F. X. Lesage, S. Berjot, and F. Deschamps, "Clinical stress assessment using a visual analogue scale," *Occupational medicine*, vol. 62 no. 8, pp. 600-605, 2012.
[7] M. Kanda, M. Matsuhashi, N. Sawamoto, T. Oga, T. Mima, T. Nagamine and H. Shibasaki, "Cortical potentials related to assessment of pain intensity with visual analogue scale (VAS)," *Clinical Neurophysiology*, vol. 113 no. 7, pp. 1013-1024, 2002.
[8] A. Reich, M. Heisig, N. Q. Phan, K. Taneda, K. Takamori, S. Takeuchi and J. C. Szepietowski, "Visual analogue scale: evaluation of the instrument for the assessment of pruritus," *Acta Dermato Venereologica*, vol. 92 no. 5, pp. 497, 2012.
[9] F. Dexter and D. H. Chestnut, "Analysis of statistical tests to compare visual analog scale measurements among groups," *Anesthesiology*, vol. 82, pp. 896-902, 1995.
[10] M. Kliger, S. Stahl, M. Haddad, E. Suzan, R. Adler and E. Eisenberg, "Measuring the intensity of chronic pain: are the visual analogue scale and the verbal rating scale interchangeable?" *Pain Practice*, vol. 15, no. 6, pp. 538-547, 2015.
[11] M. J. Hjermstad, P. M. Fayers, D. F. Haugen, A. Caraceni, G. W. Hanks, J. H. Loge, R. Fainsinger, N. Aass and S. Kaasa, "Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review," *Journal of Pain and Symptom Management*, vol. 41 no. 6, pp. 1073-1093, 2011.
[12] H. Van Laerhoven, H. J. Van der Zaag-Loonen and B. H. Derkx, "A comparison of likert scale and visual analogue scales as response options in children's questionnaires," *Acta paediatrica*, vol. 93 no. 6, pp. 830-835, 2004.
[13] J. Bielewicz, B. Daniluk, and P. Kamieniak, "VAS and NRS, same or different? Are visual analog scale values and numerical rating scale equally viable tools for assessing patients after microdiscectomy?" Pain Res Manag, (vol,page), 2022.
[14] A. Butler, P. Rothery and D. Roy, "Minitab macros for resampling methods,". *Teach. Stat.*, vol. 25, no. 1, pp. 22-25, 2003.
[15] E. Stark and M. Abeles, "Applying resampling methods to neurophysiological data," *J Neurosci Methods*, vol. 145, pp. 133–14, 2005. doi: 10.1016/j.jneumeth.2004.12.005
[16] D. S. Collingridge, "A primer on quantitized data analysis and permutation testing," *J Mix Methods Res*, vol. 7 no. 1, pp. 81-97, 2012.

[17] A. Figueiredo, "Resampling methods in ANOVA for data from the von Mises-Fisher distribution," *Commun. Stat. Simul. Comput*, pp. 1-15, 2021.
[18] Ö. Koşkan and F. Gürbüz, "Comparison of F test and resampling approach for type I error rate and test power by simulation method," *Journal of Agricultural Sciences,* vol. 15, no. 1, pp. 105-111, 2009.
[19] D. J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, Chapman and Hall/CRC, 1987.
[20] T. Kesici and Z. Kocabaş, Biyoistatistik, Ankara University Press, 1998.
[21] D. J. K. Mewhort, "A comparison of the randomization test with the F test when error is skewed," *Behav. Res. Methods*, vol. 37 no.3, pp. 426-435, 2005.
[22] M. Weber, Robustness and Power of the t, Permutation T and Wilcoxon Tests(Doctoral thesis, University of Wayne State University, Detroit, Michigan*), 2006*. Retrieved from https://www.proquest.com/docview/304968464?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses
[23] R. D. Routledge, "P-values from permutation and F tests," *Comput Stat Data Anal*, vol. 24, pp. 379-386, 1997.
[24] E. Başpınar, and F. Gürbüz, "The power of the test in the samples of various sample sizes were taken from the binary combinations of the normal, beta, gamma, and weibull distributions," *Journal of Agricultural Sciences,* vol. 6, no. 1, pp. 116-127, 2000.
[25] M. Mendeş and B. Tekindal, "Determining the appropriate sample size in testing the difference between the means in normal and non-normal populations," *The Journal of the Industrial Arts Education Faculty of Gazi University,* vol. 111, pp. 25–38, 2002.
[26] G. Heller and E. S. Venkatraman, "Resampling procedures to compare survival distributions in the presence of right-censored data," *Biometrics*, vol. 52, no. 4, pp. 1204-1213, 1996. doi: 10.2307/2532836.
[27] J. Ludbrook and H. Dudley, "Why permutation tests are superior to t and F tests in biomedical research," *Am. Stat.*, vol. 52, no. 2, pp. 127-132, 1998. doi: 10.1080/00031305.1998.10480551.
[28] P. Good, Permutation Tests: Practical Guide to Resampling Methods for Testing Hypotheses, Springer Verlag, New York, USA, 2000. doi: 10.1007/978-1-4757-2346-5.
[29] G. K. Balasubramani, S. R. Wisniewski, H. Zhang, and H. F. Eng, "Development of an efficient SAS macro to perform permutation test for two independent samples," *Comput. Methods Programs Biomed.*, vol. 79, pp. 179-187, 2005. doi: 10.1016/j.cmpb.2005.03.010.
[30] F. Pesarin and L. Salmaso, "Permutation test for univariate and multivariate ordered categorical data,"*Austrian J. Sta.t*, vol. 35, no. 2, pp. 315-324, 2006. doi: 10.17713/ajs.v35i2&3.378.
[31] S. Keskin and M. Mendeş, "Comparison of the power of the test of one-way ANOVA method and some approximation tests for the samples drawn from the exponential distributed population," *J. Agric. Sci.*, vol. 8, no. 4, pp. 293-299, 2002.
[32] P. Vidoni, "Improved prediction limits for continuous and discrete observations in generalised linear models," *Biometrica*, vol. 88, no. 3, pp. 881-887, 2001.