Psychologie
Faculteit der Sociale Wetenschappen

Universiteit Leiden

Violation of Homogeneity of Variances: A Comparison Between Welch's *t*-Test and the Permutation Test

Ruchella Kock

Bachelor thesis Psychology
Institute of Psychology
Faculty of Social and Behavioural Sciences – Leiden University
Unit: Methodology and Statistics
Date: 04 – 06 – 2019
Bachelorproject number: 34
Student number: 1815458
First examiner: Julian Karch

Acknowledgments

I want to thank my supervisor Julian Karch, who was always available for my questions. I am deeply indebted to the valuable feedback he has given me. I am grateful to my teammates, Michel Westerik, Faraaz Azizahamad, and Ligaya Breemer, for giving me feedback and moral support. I am thankful to Joel Bartholomew for listening to me talk about this subject for hours and for all the help he provided me. To my family and friends for reading my thesis and motivating me to keep going.

Abstract

A substantial amount of psychological research use statistical tests to compare means. A widely used method to do this is the parametric *t*-test. However, the *t*-test has many assumptions, whereas nonparametric tests such as the permutation test have fewer assumptions. Despite this, nonparametric tests are not widely used in psychological research. Nonetheless, studies have compared the *t*-test and permutation test against each other. Studies that compared the tests were not focused on comparing the tests when there is variance heterogeneity, large sample sizes, and unequal group sizes. In this study, a simulation study was performed to compare the permutation test and Welch's *t*-test in terms of the homogeneity assumption. This study used a range of sample sizes representative of psychological research along with varying group ratios. When there is variance homogeneity, the tests perform equally well. When there is variance heterogeneity and equal sample sizes, the tests performed equally well for the large sample sizes. However, when there is variance heterogeneity and unequal sample sizes, the type I error of the permutation test was much higher or much lower than $\alpha = 0.05$. This is due to the violation of the permutation test's assumption of exchangeability. Welch's *t*-test is not affected by variance heterogeneity. Thus, Welch's *t*-test should be preferred if there is variance heterogeneity and its other assumptions are met. The permutation test is preferred if there is variance homogeneity or if the assumptions of Welch's *t*-test are not met and there variance heterogeneity but equal sample sizes.

Keywords: Welch's *t*-test, permutation test, variance homogeneity, variance heterogeneity

Contents

Introduction

In psychological research, population means are often statistically compared against each other, a widely used method to compare means is the *t*-test (Edgington, 1974; Goodwin & Goodwin, 1985; Skidmore & Thompson, 2010). The first introduced and most commonly used *t*-test is Student's *t*-test (Howell, 2009; Student, 1908). Student's *t*-test has 3 central assumptions, namely, independence, homogeneity of variances, and normality. These assumptions must be met in order to use the test. An alternative to Student's *t*-test, which has fewer assumptions, is Welch's *t*-test (Welch, 1947, 1938). Welch's *t*-test does not assume homogeneity of variances. Student's and Welch's *t*-test, are parametric tests.

An alternative to the parametric *t*-test is the nonparametric permutation test (Fisher, 1937). Nonparametric tests have fewer assumptions than parametric tests. Despite this, the permutation test is used less often than the *t*-test in psychological research (Edgington, 1974; Goodwin & Goodwin, 1985). As these tests should not be used if their assumptions are not met, their assumptions are discussed in the following sections.

Assumptions of the *t*-test

The *t*-test has several assumptions. First, it assumes independent errors. Which means that the residuals should not be able to be predicted above chance. Second, it assumes that the sampling distribution is normal. Another assumption is that there are no outliers. As the means of the groups are compared, an outlier can greatly skew the mean, which can lead to incorrect conclusions. Finally, there is the assumption of homogeneity of variances. Variance (σ^2) refers to the way the scores are distributed around the mean. Homogeneity of variances means that the variances across groups are considered equal. This assumption is important because if the scores in one group were spread differently, compared to the second group before any treatment was given, then these groups are no longer comparable (Salkind, 2010).

Some studies show that the *t*-test is robust against violations of its assumptions (e.g., Sawilowsky & Blair, 1992; Bradley, 1978). It is believed that the t-test is robust against non-normality if the sample size is greater or equal to 30. The t-test is believed to be robust against violation of the assumption of homogeneity if the group sizes are approximately equal. If the assumption of homogeneity is violated, and the group sizes are not equal, Welch's t-test can be used as it does not assume homogenous variances (Howell, 2009; Delacre, Lakens, & Leys, 2017).

Assumptions of the permutation test

There are two kinds of probability models, namely the randomization model and the population model. In the randomization model, the subjects are randomly assigned to a condition. In the population model, subjects are randomly sampled from a population (Ernst, 2004). The name permutation test is often used to refer to both the randomization model and population model because in many cases they are equivalent to each other. The two tests are also referred to as the randomization test and the permutation test, respectively (Nichols & Holmes, 2002).

The randomization and permutation test assume exchangeability, which has different implications for the tests. One implication is the stable unit treatment value assumption (SUTVA) (Rubin, 1980). In an experiment, subjects/units i can be exposed to treatment j. Therefore, Y_{i_j} is the observed effect of unit i in treatment j. In this experiment, each unit is only part of one treatment group at a time. Thus, Y_{i_1} and Y_{i_2} cannot be observed at the same time. Inferences must be made about the value that was not observed. The effect of treatment 1 on unit i should be independent of the effect on other units in any treatment group; otherwise, SUTVA will be violated (Rubin, 1980).

Another implication of exchangeability is that the variances are homogeneous. If the groups have different variances, then the groups are not interchangeable. Thus, variance heterogeneity leads to a violation of exchangeability (Huang, Xu, Calian, & Hsu, 2006).

In the randomization model, most implications of exchangeability are usually fulfilled, because participants are randomly assigned to the groups and should, therefore, be thought of as interchangeable. For the population model, there is no random assignment; therefore, exchangeability cannot be directly assumed. Thus, the population model also assumes that the distributions of the two groups have approximately the same shape (Nichols & Holmes, 2002).

To conclude, there is a subtle difference between the two tests in terms of who the population is. In this study, the randomization model is used. Thus, the assumption of exchangeability is met as long as the variances are equal. The randomization model is chosen because the population model is often not used in psychological studies. Convenient sampling is used instead, which is not possible in the population model (Fife, 2013). Thus, using the randomization model in this thesis is a closer approximation of current psychological research.

Literature review

In this section, existing literature comparing the permutation test and the *t*-test is reviewed. Toothaker (1972) wrote a dissertation on comparing the permutation *t*-test with Student's *t*-test and the Mann Whitney U test. He performed a simulation study using normally distributed data with equal variances and sample sizes ranging from 2 to 5. The study concluded that the permutation *t*-test does not outperform Student's *t*-test and the Mann Whitney U test and the latter two should be preferred when comparing means.

Ludbrook and Dudley (1998) compared the permutation test with the *t*-test and *F*-test in Biomedical Research. They found that researchers in this field often choose an *F*-test or *t*-test instead of a permutation test even if the assumptions are not met. They conclude that exact permutation or randomization tests should be preferred in biomedical research.

Hughes (2010) conducted a simulation study, where she compared the two-sample *t*-test with the two-sample exact permutation test. She used six non-normal distributions, tested at three different significance levels, and the sample sizes ranged from 2 to 6. She concluded that the permutation test should be preferred, especially if power is essential for a study.

Most relevant to this thesis is the simulation study performed by Mendeş and Akkartal (2010). They compared the ANOVA F-test and Welch's t-test with the permutation F-test and the permutation Welch's t-test. They used 3 different distributions, 5 different group sizes ranging from 5 to 15 and 3 different group variances namely, equal variances ($\sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_3^2 = 1$), a small deviation ($\sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_3^2 = 4$) and a larger deviation ($\sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_3^2 = 9$). By comparing these groups, they observed the effects of non-normality and heterogeneity. They concluded that when the assumption of homogeneity and normality is violated, the permutation F-test should be used. When the assumption of normality is violated, but equal variances are assumed, then the permutation Welch's t-test should be used.

There are some gaps in the existing literature. For instance, little attention has been devoted to large sample sizes when comparing the t-test with the permutation test. All reviewed studies used small sample sizes, the largest group size being 15. These sample sizes are not representative of current psychological studies. According to the study from Kühberger, Fritz, and Scherndl (2014), only 14.9% of studies had a sample size of 15 or smaller. Mendeş and Akkartal (2010) looked at the effect of different group sizes. However, the most substantial deviation between groups was 10. Larger deviations between group sizes when comparing the two tests have not been studied. Additionally, only one study compared the two tests when the homogeneity assumption is violated (Mendeş & Akkartal, 2010).

This research aims to fill these gaps, focusing on the comparison between the permutation test and Welch's t-test when there is variance homogeneity or variance heterogeneity. Furthermore, as large sample sizes are not explored in previous research, this study will include small as well as large sample sizes to investigate whether they lead to different conclusions. Unequal group sizes have also not been widely researched. However, it is essential to consider because unequal group sizes can affect the tests, especially for the permutation test (Huang et al., 2006). This study uses equal as well as unequal sample sizes.

The tests are compared in terms of type I and type II errors. All statistical tests may lead to errors. Type I error is when H_0 is rejected when it should not have been. Type II error is when H_0 is not rejected when it should have been. If the type II error decreases, the power of a test increases. The power is the probability that H_0 is rejected when H_1 is true. The type I error of a test is often set to $\alpha = 0.05$, and a power of 0.80 is considered to be high enough (Howell, 2009). In this study, the type I and type II error is calculated using a simulation study.

Welch's t-test is chosen because it provides more reliable type I error rates when the assumption of homogeneity of variance is not met. Compared to Student's t-test, Welch's t-test loses some statistical power. However, the loss of power is minimal. Thus, Welch's t-test is a favorable alternative to Student's t-test (Delacre et al., 2017).

The goal of this thesis is to provide a relevant comparison between the tests, where the results can be applied in current psychological research. To achieve this goal, small and large sample sizes that are often used in psychology were chosen, and the randomization test, which is more common in psychological research, was used.

Research questions and hypothesis

In this section, the hypothesis and research questions of this study are discussed. Welch's t-test does not assume variance homogeneity, but the permutation test does (Boik, 1987). Thus, it may be hypothesized that Welch's t-test performs better than the permutation test when there is variance heterogeneity. However, it is still important to investigate the effects of the tests when there are homogeneous variances, especially whether the type II errors of the tests are similar. Moreover, it is interesting to test whether equal sample sizes affect the performance of the permutation test. According to Huang et al. (2006), if the data is normally distributed, and the sample sizes are equal, the permutation test is not affected by unequal variances.

The research question of this thesis is: How does the permutation test compare to Welch's t-test? The following sub-questions are explored to answer the research question.

- How does the permutation test compare to Welch's t-test under no violation of the assumption of homogeneity of variances?
- What is the effect of sample size, on the performance of the permutation test and Welch's t-test, under violation of the assumption of homogeneity of variances?
- What is the effect of unequal group sizes, on the performance of the permutation test and Welch's t-test, under violation of the assumption of homogeneity of variances?

In the following sections, the design of the simulation, the critical results, the discussion, and conclusion are reported.

Methods

To compare Welch's t-test (in further sections referred to as t-test) and permutation test, a simulation study was conducted using the programming language R (R Core Team, 2018). The type I and type II errors of the t-test and permutation test were estimated and compared against each other. In the following subsections, Welch's t-test and the permutation test are explained. This is followed by the design of the simulation, namely, the chosen sample sizes, effect sizes, means, and standard deviations. Finally, the implementation of the simulation is described.

Description of Welch's *t*-test

Welch's t-test tests the null hypothesis that the means of two groups are equal. Let there be two groups X and Y; the null hypothesis is $H_0 : \overline{X} = \overline{Y}$. Let $\overline{X}$ denote the mean of group X and $\overline{Y}$ denote the mean of group Y. The t statistic is calculated with the following equation:

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y}}} \qquad \text{(Welch (1938))}$$

Let s^2 denote the variance and N the sample size. To calculate the degrees of freedom (df), Welch's t-test uses the Satterthwaite-Welch adjustment:

$$df = \frac{\left(\frac{S_X^2}{N_X} + \frac{S_Y^2}{N_Y}\right)^2}{\frac{\left(\frac{S_x^2}{N_X}\right)^2}{N_X - 1} + \frac{\left(\frac{S_y^2}{N_Y}\right)^2}{N_Y - 1}} \quad \text{(Satterthwaite (1946))}$$

Finally, the t statistic and df are used to get the p-value and test the null hypothesis using the t-distribution.

Description of the permutation test

The steps to perform a permutation test are as follows: suppose there are two groups, X_i and Y_j. $i = 1, ..., n$ and $j = 1, ..., m$ observations for each group. From these groups, the test statistic is calculated. Different test statistics can be calculated, such as Pearson r or the mean difference $\overline{X} - \overline{Y}$. Subsequently, X_i and Y_j are pooled together to form one new group. Re-sample from this group and form two new groups $X*_i$, $Y*_j$. Calculate the test statistic. Repeat this procedure k number of times (e.g., $k = 10000$). This forms a test distribution. The null hypothesis can be tested under this distribution. If the null hypothesis is true, all the possible pairings of the re-sampled groups are equally likely (Anderson & Robinson, 2001; Legendre & Legendre, 1998).

In this study, a linear permutation test was performed, with the test statistic in the following form:

$$T = \sum_{n=1}^{n} c_i z_i \quad \text{(Fay and Shaw (2010))}$$

Let z_i denote a scalar or k x 1 vector, and c_i denotes a scalar. z_i is the group membership. c_i is the score.

Sample size (N)

To simulate data with a normal distribution, a sample size (N), effect size (ES), mean (μ), and standard deviation (σ) are needed. Two different groups were simulated each time. The following strategy was used to choose the sample sizes of both groups. For the first group, sample sizes that are relevant to psychology were chosen with the data provided by Kühberger et al. (2014). They randomly sampled 1000 articles to investigate whether effect size is independent of sample size in psychological research. The sample sizes of the 529 articles that met their criteria, were analyzed in this study, and three were chosen for the simulation. First, a small sample size often used in psychology, namely $N = 10$. Less than 10% of the articles had a sample size that is smaller than 10 (8.9%). Second, $N = 60$, a medium sample size. Almost half of the sample sizes were smaller than or equal to 60 (47.6%). Finally, a large sample size $N = 1000$, with only 10% of the reported sample sizes larger than it. The size of the second group (N_2) varied relative to the size of the first sample. The following percentages for the ratio N_2/N_1 were used: 1, 1.25, 1.5, 1.75, .75, .5, .25. This was motivated by the aim of this research, to investigate equal sample sizes (condition 1) as well as violations with differing degrees of severeness and direction (see Table 1).

Effect size (*ES*)

ES is the standardized mean difference between two groups (Coe, 2002). If there is a strong effect, the *ES* will be large, which means that the probability that the statistical test is significant is also large. Therefore, different effect sizes have different implications. In this thesis *ES* 0.0 and Cohen's three benchmark effect sizes were chosen, namely a small *ES* of 0.2, a medium *ES* of 0.5 and a large *ES* of 0.8 (Cohen, 2013).

Mean (μ) and Standard deviation (σ)

The standard normal distribution was chosen for one group, thus $\mu_1 = 0$ and $\sigma_1 = 1$. The σ_2 was altered to simulate variance homogeneity or heterogeneity, when there is variance homogeneity, the variances of both groups are equal ($\sigma_1^2 = \sigma_2^2$). However, when there is variance heterogeneity, $\sigma_1^2 \neq \sigma_2^2$. Six different deviations were chosen to simulate heterogeneity, σ_2 was either smaller or larger than σ_1 by 25%, 50%, 75% and 300% (see

Table 2). To calculate μ_2, let

$$\overline{\sigma} = \sqrt{\frac{\sigma_1^2 + \sigma_1^2}{2}} \qquad \text{(Bonett (2008))}$$

$$\mu_2 = \overline{\sigma} * ES$$

To conclude, when the two groups were created, the means were compared with both Welch's t-test and the permutation test. Then, the type I and type II error of each test was estimated. When testing for type I error, the ES was 0.0. If the p-value of the t-test or permutation test was smaller than $\alpha = 0.05$, then the test committed a type I error. For type II error, the ES was either 0.2, 0.5 or 0.8. If the p-value of either test was larger than $\alpha = 0.05$, then there was a type II error. After calculating the type I or type II errors, the McNemar test was used to check whether there is a statistically significant difference between Welch's t-test and the permutation test (McCrum-Gardner, 2008).

Implementation

This section describes how the simulation was implemented and performed. Each simulation was repeated 10000 times. The data was simulated using `rnorm()`. Welch's t-test was performed using the `t.test()` formula in R with the argument `var.equal` set to False. The permutation test was performed using the library *perm* (Fay & Shaw, 2010). The Monte Carlo sampling technique was used during the permutation test. Ideally, all permutations are performed in a permutation test. However, with larger sample sizes, the number of permutations becomes very large. Therefore, the Monte Carlo sampling technique should be used. This technique randomly chooses test statistics from the permutation distribution. From this random sample, the p-value for the permutation test can be calculated (Ernst, 2004; Hastings, 1970). The code for the simulation is included on `https://github.com/rushkock/sim_study_thesis/tree/master/src/simulation`.

Table 1

Group sizes used during the simulation

Sample Size	Group Ratios
Small $N = 10$	
Condition 1	$N_1 = 10 : N_2 = 10$
Condition 2a	$N_1 = 10 : N_2 = 8$
Condition 2b	$N_1 = 10 : N_2 = 13$
Condition 3a	$N_1 = 10 : N_2 = 5$
Condition 3b	$N_1 = 10 : N_2 = 15$
Condition 4a	$N_1 = 10 : N_2 = 3$
Condition 4b	$N_1 = 10 : N_2 = 18$
Medium $N = 60$	
Condition 1	$N_1 = 60 : N_2 = 60$
Condition 2a	$N_1 = 60 : N_2 = 45$
Condition 2b	$N_1 = 60 : N_2 = 75$
Condition 3a	$N_1 = 60 : N_2 = 30$
Condition 3b	$N_1 = 60 : N_2 = 90$
Condition 4a	$N_1 = 60 : N_2 = 15$
Condition 4b	$N_1 = 60 : N_2 = 105$
Large $N = 1000$	
Condition 1	$N_1 = 1000 : N_2 = 1000$
Condition 2a	$N_1 = 1000 : N_2 = 750$
Condition 2b	$N_1 = 1000 : N_2 = 1250$
Condition 3a	$N_1 = 1000 : N_2 = 500$
Condition 3b	$N_1 = 1000 : N_2 = 1500$
Condition 4a	$N_1 = 1000 : N_2 = 250$
Condition 4b	$N_1 = 1000 : N_2 = 1750$

Table 2

Standard Deviations used in the simulation

σ_1	σ_2
1.00	1.00
1.00	0.75
1.00	1.25
1.00	0.50
1.00	1.50
1.00	0.25
1.00	1.75
1.00	3.00

Results

The full results can be found in Appendix B, which also includes a digital version. The data analysis was performed using Python (Python Core Team, 2015). The code for the data analysis can be found on https://github.com/rushkock/sim_study_thesis/tree/master/src/features/python_data_analysis. Appendix C contains a few of the plots used to visualize the data, and the URL to find the rest. In this section, the important results are discussed.

When there is no violation of homogeneity of variances, almost no statistically significant differences between the tests were found (Table 3). Both the *t*-test and the permutation test maintained a correct type I error ($\alpha = 0.05(\pm 0.01)$) in almost all conditions. The type I error of the permutation test was significantly better than the *t*-test in 1 out of 84 conditions. In this condition $N_1 = 10$ and $N_2 = 3$, there was an absolute significant difference of 0.0176 between the type I error of the tests ($p < 0.001$). For type II error, a significant difference between the tests was found in 8 out of 84 conditions. All these conditions had a *p*-value of 0.01 or smaller. The type II error of the *t*-test was significantly better than the permutation test in 4 conditions. In these 4 conditions, $N_1 = 10$, but N_2 and ES varied. In the conditions where the permutation test was significantly better, N_1, N_2, and ES varied. However, important to mention is that the significant differences were mostly found for the larger effect sizes ($ES = 0.5$ and 0.8). To conclude, there was much variation between these 4 conditions.

Table 3

Conditions with a statistically significant difference between the permutation test and Welch's t-test when there was variance homogeneity

N_1	N_2	ES	σ_1	σ_2	perm	*t*-test	*p*-value	dif
10	3	0.0	1.0	1.0	0.0481	0.0657	0.000	-0.0176
10	3	0.2	1.0	1.0	0.9430	0.9333	0.009	0.0097

10	10	0.5	1.0	1.0	0.8261	0.8201	0.000	0.0060
10	10	0.8	1.0	1.0	0.6190	0.6122	0.000	0.0068
10	8	0.8	1.0	1.0	0.6670	0.6577	0.000	0.0093
10	5	0.8	1.0	1.0	0.7305	0.7478	0.000	-0.0173
60	30	0.5	1.0	1.0	0.3963	0.4041	0.009	-0.0078
60	15	0.5	1.0	1.0	0.6004	0.6188	0.000	-0.0184
60	15	0.8	1.0	1.0	0.2172	0.2405	0.000	-0.0233

"N_1" and "N_2" are the sizes of the two groups. "ES" is the effect size. An effect size of 0.0 represents a type I error. Effect size 0.2, 0.5, and 0.8 represent type II errors. "σ_1" and "σ_2" are the standard deviations of the two groups. The column perm contains the number of errors for the permutation test. The "t-test" column contains the number of errors for the t-test. The column "p-value" gives the p-value from the McNemar test comparing the permutation test with the t-test. The column "dif" is the difference between errors for the t-test minus the errors of the permutation test. Thus, a negative value indicates that the permutation test outperforms the t-test.

As hypothesized (Section Research questions and hypothesis), when there is variance heterogeneity, the permutation test did not have a type I error of $\alpha = 0.05(\pm 0.01)$ in almost all conditions. This is referred as a failure of the test in further sections. In Table 4, a small overview of the results for the type I error of the small sample sizes is displayed. The results for the remaining sample sizes are qualitatively the same. For the conditions where the standard deviation of group 1 (σ_1) is 3.00, and group 2 (σ_2) is 1.00; the t-test always performs at $\alpha = 0.05(\pm 0.01)$. In contrast, the type I error of the permutation test greatly exceeds $\alpha = 0.05(\pm 0.01)$ when N_1 is smaller than N_2 ($N_1 = 10$ and $N_2 = 13$, 15 or 18). An example of this is seen in Table 4, when $N_1 = 10$ and $N_2 = 13$, the type I error of the permutation test is 0.082. As the difference between N_1 and N_2 gets larger, the type I error of the permutation test also gets further away from $\alpha = 0.05$. An example of this is seen when $N_1 = 10$ and $N_2 = 18$, the type I error of the permutation test is $\alpha = 0.126$.

However, when N_1 is larger than N_2 ($N_1 = 10$ and $N_2 = 8$, 5 or 3), the type I error of the permutation test is a lot smaller than $\alpha = 0.05(\pm 0.01)$. When $N_1 = 10$ and $N_2 = 3$, the type I error of the permutation test was $\alpha = 0.009$ (Table 4). This pattern of failure is consistent for all violations of homogeneity where σ_1 is larger than σ_2 ($\sigma_1 = 1.25, 1.50, 1.75$ or 3.0 and $\sigma_2 = 1.0$).

Table 4

Simulation results for ES 0.0 under violation of homogeneity, where $\sigma_1 = 3.0$ and $\sigma_2 = 1.0$

N_1	N_2	ES	σ_1	σ_2	perm	t-test	p-value	dif
10	10	.0	3.00	1.00	.059	.054	0.000	0.005
10	8	.0	3.00	1.00	.038	.050	0.000	-0.013
10	13	.0	3.00	1.00	.082	.052	0.000	0.031
10	5	.0	3.00	1.00	.023	.050	0.000	-0.027
10	15	.0	3.00	1.00	.106	.054	0.000	0.052
10	3	.0	3.00	1.00	.009	.049	0.000	-0.040
10	18	.0	3.00	1.00	.126	.049	0.000	0.078

See Table 3 for further explanation on column names.

When σ_1 is smaller than σ_2 ($\sigma_1 = 0.25$, 0.50 or 0.75 and $\sigma_2 = 1.0$), the t-test performed once again at $\alpha = 0.05(\pm 0.01)$, but the permutation test did not. The type I error of the permutation test greatly exceeds $\alpha = 0.05(\pm 0.01)$ when N_1 is larger than N_2 ($N_1 = 10$ and $N_2 = 8$, 5 or 3). An example of this is given in table 5, the biggest type I error rate was $\alpha = 0.222$, for the condition where $N_1 = 10$ and $N_2 = 3$. However, when N_1 is smaller than N_2 ($N_1 = 10$ and $N_2 = 13$, 15 or 18), the type I error of the permutation test is a lot smaller than $\alpha = 0.05(\pm 0.01)$. The smallest type I error rate from Table 5 was $\alpha = 0.015$, for the condition where $N_1 = 10$ and $N_2 = 18$. Thus, the permutation test fails in opposite directions when $\sigma_1 < \sigma_2$, compared to when $\sigma_2 < \sigma_1$.

Table 5

Simulation results for ES 0.0 under violation of homogeneity, where $\sigma_1 = 0.25$ *and* $\sigma_2 = 1.0$

N_1	N_2	ES	σ_1	σ_2	perm	t-test	p-value	dif
10	10	.0	0.25	1.00	.059	.052	0.000	0.007
10	8	.0	0.25	1.00	.086	.052	0.000	0.034
10	13	.0	0.25	1.00	.029	.047	0.000	-0.018
10	5	.0	0.25	1.00	.155	.057	0.000	0.098
10	15	.0	0.25	1.00	.022	.048	0.000	-0.025
10	3	.0	0.25	1.00	.222	.065	0.000	0.158
10	18	.0	0.25	1.00	.015	.048	0.000	-0.033

See Table 3 for further explanation on column names.

The permutation test does not have a correct type I error rate when there is variance heterogeneity, regardless of how large the differences in variance are. The failure is stronger when the sample sizes deviate more from each other. Welch's t-test maintains almost the same type I error across all conditions. These findings are consistent across all sample sizes (see Appendix B). Considering that the permutation test does not have a correct type I error rate in the conditions with variance heterogeneity, the type II error of the permutation test is not further explored for these conditions.

Furthermore, as expected, when the sample sizes get larger, less significant differences are found between the tests. In the conditions where the sample size of group 1 is large (N_1 = 1000), there were no significant differences between the two tests for effect size 0.5 and 0.8 (Table B3).

Finally, when the group sizes were equal both the permutation and Welch's t-test maintained a correct type I error rate ($\alpha = 0.05(\pm 0.01)$) in almost all conditions. This was regardless of sample size. In the larger sample sizes ($N_1 = 60$ or 1000), there were almost no statistically significant differences found between the tests when the group sizes were equal. In contrast, for the smaller sample sizes ($N_1 = 10$), there were many statistically significant differences between the tests for both type I and type II errors. Which test had a better type I or type II error depended on the σ_1, this was as follows: the type I error of the t-test was significantly higher than the permutation test for the conditions where $\sigma_1 =$ 3.0 and 0.25. The type II error of the t-test was significantly higher than the permutation test for the conditions where $\sigma_1 = 0.75$, 1.0, and 1.25. The type I error for the permutation test was significantly higher than the type I error of the t-test for $\sigma_1 = 0.75$. The type II error of the permutation test was significantly higher than the t-test for $\sigma_1 = 0.25$, 0.50, 1.75, and 3.0. To conclude, when the sample sizes are large, and the group sizes are equal, the two tests perform equally well.

Discussion

This simulation study compared the permutation test and Welch's t-test. The variances, sample sizes, and group ratios were altered to investigate their effect on the tests. The results suggest that when there is variance homogeneity, both tests perform equally well in terms of type I as well as type II errors. However, as hypothesized, when there is variance heterogeneity, the permutation test did not have a correct type I error rate ($\alpha = 0.05(\pm 0.01)$). This is referred as a failure of the permutation test. The exception to this failure was for the conditions with equal group sizes. Variance heterogeneity does not affect Welch's t-test. This suggests that Welch's t-test should always be chosen because regardless of the variance, it always performs well, whereas the performance of the permutation test depends on the variance and group ratios.

However, as Welch's t-test assumes normality, which the permutation test does not, the permutation test might be beneficial when there is variance homogeneity, but data is not normally distributed. In most conditions, with variance homogeneity, no statistically significant differences were found between the tests, except for 9 conditions. However, there was no pattern between these 9 conditions. This indicates that these differences could be due to false positives of the McNemar test. It can be concluded that both tests perform equally well when there is no violation of homogeneity.

The permutation test does not have a correct type I error rate when there is variance heterogeneity and unequal sample sizes. This is due to the violation of the assumption of exchangeability. Previous research has also reported this failure (Huang et al., 2006; Boik, 1987). In Appendix A, a detailed explanation is given to explain why the permutation test fails. To summarize, a permutation test calculates the mean of two groups. It then re-samples two groups and compares them to the original groups. This is repeated multiple times to form a test distribution. The null hypothesis is then tested under this distribution. To be able to do this, the permutation test assumes exchangeability, that the differences between the groups are not due to extraneous variables such as preexisting differences or measurement errors. However, when there is variance heterogeneity this is not true, there are preexisting differences between the groups. The assumption of the test is violated, so it acts unpredictably. It can be conservative in some situations, liberal in others. The failure of the test depends on the μ and σ of the groups.

However, when the sample sizes are equal, the permutation test is protected against the failure (see Appendix A). With equal sample sizes, the permutation test and t-test should perform equally well, regardless of homogeneity (Appendix B). Consistent with this, the results show that the permutation test had a correct type I error in all conditions with equal sample sizes. However, only when the sample sizes were large ($N_1 = 60$ or 1000), did the tests perform equally well. There were almost no statistically significant differences between the tests for both type I and type II errors in these conditions. In the small sample sizes ($N_1 = 10$) many significant differences were found both for type I and type II errors, this depended on the σ^2. However, no pattern was found in these differences. Given the protection that equal sample sizes offer, if there is variance heterogeneity with non-normality, the permutation test can be chosen over Welch's t-test.

Finally, as the sample size gets larger, fewer differences were found between the tests.

This is to be expected because the larger the sample size, the easier it is for a test to detect a difference. Both tests commit less type I and type II errors. In this case, both tests perform well, and it is harder to find a significant difference between them.

Limitations

The variances are known during this simulation, but in most cases, the true variances are unknown, which makes the suggestion to choose the permutation test when there is variance homogeneity difficult to follow. If the variances are unknown, it may be safer to choose Welch's t-test or make sure the group sizes are equal.

In this study, many conditions were used, and this is a limitation because some conditions become redundant. An example is using both upwards and downwards deviations of group sizes, whereas the group ratio stays the same (Table 1). It also makes data analysis more complicated.

Another limitation is the choice of tests, a nonparametric test that is not affected by a violation of homogeneity may have been a fairer comparison for Welch's t-test. Further research should perform the simulation with a nonparametric test that is not affected by homogeneity, such as the permutation Welch test (Janssen, 1997).

Moreover, the goal of this study was to present relevant results for current psychological research. However, the sample sizes that were chosen to represent current psychological research are from studies more than 10 years ago. Thus, a more recent literature search should have been conducted to choose the sample sizes.

In this study, the randomization model was chosen because it is most often used in psychology. However, in some cases, the population model is used. Future research may perform the simulation under the population model to compare with the randomization model.

Finally, future research should compare the tests in terms of other assumptions such as non-normality. Not many studies investigate the difference between the tests when the assumptions of no outliers and independence are violated.

Conclusion

To conclude, when there is no violation of homogeneity, both tests perform equally well. If there is variance heterogeneity and equal group sizes, both tests perform equally well in the larger sample sizes. When there is variance heterogeneity and unequal group sizes, the permutation test does not have a correct type I error rate ($\alpha = 0.05(\pm 0.01)$). Based on these findings, if there is variance homogeneity, but the other assumptions of the t-test such as normality or independence are not met, the permutation test is recommended. Welch's t-test is recommended if there is variance heterogeneity and the other assumptions are met. If there is variance heterogeneity and the other assumptions of Welch's t-test are not met, the permutation test is recommended given the sample sizes are large and equal.

References

Anderson, M. J., & Robinson, J. (2001). Permutation tests for linear models. *Australian & New Zealand Journal of Statistics*, *43*(1), 75–88. doi: 10.1111/1467-842X.00156

Boik, R. J. (1987). The fisher-pitman permutation test: A non-robust alternative to the normal theory f test when variances are heterogeneous. *British Journal of Mathematical and Statistical Psychology*, *40*(1), 26–42. doi: 10.1111/j.2044-8317.1987.tb00865.x

Bonett, D. G. (2008). Confidence intervals for standardized linear contrasts of means. *Psychological Methods*, *13*(2), 99. doi: 10.1037/1082-989X.13.2.99

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152. doi: 10.1111/j.2044-8317.1978.tb00581.x

Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. Paper presented at Annual Conference of the British Educational Research Association. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.html

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch's t-test instead of student's t-test. *International Review of Social Psychology*, *30*(1), 92–101. doi: 10.5334/irsp.82

Edgington, E. S. (1974). A new tabulation of statistical procedures used in apa journals. *American Psychologist*, *29*(1), 25-26. doi: 10.1037/h0035846

Ernst, M. D. (2004, 11). Permutation methods: A basis for exact inference. *Statist. Sci.*, *19*(4), 676–685. doi: 10.1214/088342304000000396

Fay, M. P., & Shaw, P. A. (2010). Exact and asymptotic weighted logrank tests for interval censored data: The interval R package. *Journal of Statistical Software*, *36*(2), 1–34. Retrieved from http://www.jstatsoft.org/v36/i02/

Fife, D. A. (2013). *The achilles heel of psychology: How convenience sampling affects parameter estimates*. The University of Oklahoma.

Fisher, R. A. (1937). *The design of experiments*. Edinburgh, London: Oliver And Boyd.

Goodwin, L. D., & Goodwin, W. L. (1985). An analysis of statistical techniques used in the journal of educational psychology, 1979-1983. *Educational Psychologist*, *20*(1), 13-21. doi: 10.1207/s15326985ep2001_3

Hastings, W. K. (1970, 04). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97-109. doi: 10.1093/biomet/57.1.97

Howell, D. C. (2009). *Statistical methods for psychology*. Cengage Learning.

Huang, Y., Xu, H., Calian, V., & Hsu, J. C. (2006, 07). To permute or not to permute. *Bioinformatics*, *22*(18), 2244-2248. doi: 10.1093/bioinformatics/btl383

Hughes, M. B. (2010). *Comparison of size/level and power of two-sample t-test to that of two-sample exact permutation test under various population shapes and small sample sizes.* Stephen F. Austin State University.

Janssen, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized behrens-fisher problem. *Statistics Probability Letters*, *36*(1), 9 - 21. doi: 10.1016/S0167-7152(97)00043-6

Kühberger, A., Fritz, A., & Scherndl, T. (2014, 09). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLOS ONE*, *9*(9), 1-8. doi: 10.1371/journal.pone.0105825

Legendre, P., & Legendre, L. F. (1998). Numerical ecology. In (second ed., p. 17-26). Elsevier Science BV.

Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to t and f tests in biomedical research. *The American Statistician*, *52*(2), 127–132. doi: 10.1080/00031305.1998.10480551

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*(1), 38–41. doi: 10.1016/j.bjoms.2007.09.002

Mendeş, M., & Akkartal, E. (2010). Comparison of anova f and welch tests with their respective permutation versions in terms of type i error rates and test power. *Kafkas Univ Vet Fak Derg*, *16*(5), 711–716. doi: 10.9775/kvfd.2009.1507

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, *15*(1), 1–25. doi: 10.1002/hbm.1058

Python Core Team. (2015). Python: A dynamic, open source programming language (3.7 ed.) [Computer software manual]. Retrieved from `https://www.python.org/`

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593. doi: 10.2307/2287653

Salkind, N. J. (2010). *Encyclopedia of research design* (Vol. 1). Sage.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, *2*(6), 110–114. doi: 10.2307/3002019

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and type ii error properties of the t test to departures from population normality. *Psychological* *bulletin*, *111*(2), 352-360. doi: 10.1037/0033-2909.111.2.352

Skidmore, S. T., & Thompson, B. (2010). Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement*, *70*(5), 777-795. doi: 10.1177/0013164410379320

Student. (1908). The probable error of a mean. *Biometrika*, *6*(1), 1–25. doi: doi:10.2307/2331554

Toothaker, L. E. (1972). *An empirical investigation of the permutation t-test as compared to student's t-test and the mann-whitney u-test.* Washington, D.C.: ERIC Clearinghouse.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*(3/4), 350–362. doi: 10.2307/2332010

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, *34(1/2)*(1/2), 28–35. doi: doi:10.2307/2332510

Appendix A

In this appendix, the results are explained with the explanation from Huang et al.(2006). They used a re-sampling with replacement example. Their study can explain the results in this thesis as follows: say a simulation is performed with group X and Y. Both groups have a normal distribution $N(\mu, \sigma^2)$. The means are compared against each other. The null hypothesis is $H_0 : \mu_x = \mu_y$. The test statistic to test this hypothesis can be described with $T = \overline{X} - \overline{Y}$. The true sampling distribution of T is shown with the equation:

$$N(0, \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}) \tag{1}$$

Where m is the number of scores in group X and n is the number of scores of group Y. After re-sampling, the observation can be in group X or Y. The chance of being in group X with σ_x^2 is $\frac{m}{m+n}$. The chance of being in group Y with σ_y^2 is $\frac{n}{m+n}$. Thus, Equation 1 can be written as follows:

$$\frac{\frac{m}{m+n} * \sigma_x^2 + \frac{n}{m+n} * \sigma_y^2}{m} + \frac{\frac{m}{m+n} * \sigma_x^2 + \frac{n}{m+n} * \sigma_y^2}{n} \tag{2}$$

$$\frac{\frac{m}{m+n} * \sigma_x^2}{m} + \frac{\frac{n}{m+n} * \sigma_y^2}{m} + \frac{\frac{m}{m+n} * \sigma_x^2}{n} + \frac{\frac{n}{m+n} * \sigma_y^2}{n} = \tag{3}$$

$$\sigma_x^2(\frac{1}{m+n} + \frac{m/n}{m+n}) + \sigma_y^2(\frac{n/m}{m+n} + \frac{1}{m+n}) = \tag{4}$$

$$(\frac{1}{n})\sigma_x^2 + (\frac{1}{m})\sigma_y^2 = \tag{5}$$

$$\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} \tag{6}$$

After re-sampling, the sampling distribution is thus:

$$N(0, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}) \tag{7}$$

If the variances are equal, then the true null distribution (Equation 1) is the same as the re-sampled distribution (Equation 7). Say $\sigma_x^2 = 1$ and $m = 6$ and $\sigma_y^2 = 1$ and $n = 4$ then the two distributions are the same.

$$N(0, \frac{1}{6} + \frac{1}{4}) == N(0, \frac{1}{4} + \frac{1}{6}) \tag{8}$$

However, if group X and Y had unequal variances ($\sigma_x^2 = 1$ and $\sigma_y^2 = 3$), the distributions are only equal if $m = n$. Say $m = 6$ and $n = 6$.

$$N(0, \frac{1}{6} + \frac{3}{6}) == N(0, \frac{1}{6} + \frac{3}{6}) \tag{9}$$

In the case that the groups had unequal variances and unequal sizes, then the permutation test acts liberal or conservative depending on which variance each group has. The permutation test is liberal when the smaller variance is paired with the largest group size, and the larger variance is paired with the smaller group size. Liberal means that it results in a value much larger than $\alpha = 0.05(\pm 0.01)$. Say $\sigma_x^2 = 1$, $m = 6$ and $\sigma_y^2 = 3$, $n = 4$.

$$N(0, \frac{1}{6} + \frac{3}{4}) \tag{10}$$

If the smaller variance is paired with, the smaller group size then the permutation test is conservative. Conservative being that it results in a value much smaller than $\alpha = 0.05(\pm 0.01)$. Say $\sigma_x^2 = 1$, $m = 4$ and $\sigma_y^2 = 3$, $n = 6$.

$$N(0, \frac{1}{4} + \frac{3}{6}) \tag{11}$$

Applying this knowledge to the findings of this thesis, this also occurs. First, when there is variance heterogeneity, but the group sizes are equal, the permutation test does not fail (Equation 9). Furthermore, when there is variance heterogeneity with unequal sample sizes, the permutation test fails (Equation 10 and Equation 11). If we take the results from Table 4 as an example, where $N_1 = 10$, $\sigma_1 = 3.00$ and $N_2 = 13$, $\sigma_2 = 1.00$ we get a liberal error rate namely $\alpha = 0.082$.

$$N(0, \frac{3^2}{10} + \frac{1^2}{13})$$

The condition where $N_1 = 10$ and $\sigma_1 = 3.00$ $N_2 = 8$ and $\sigma_2 = 1.00$ had a conservative error rate namely $\alpha = 0.038$.

$$N(0, \frac{3^2}{10} + \frac{1^2}{8})$$

However, a liberal or conservative error rate is not correct and indicates a failure of the permutation test. This failure was hypothesized because the assumption of exchangeability is violated when there is variance heterogeneity.

Appendix B

Digital results:

https://github.com/rushkock/sim_study_thesis/tree/master/src/features/csv

Table B1

Simulation results for sample size 10 and its deviations

N_1	N_2	ES	σ_1	σ_2	perm	t-test	p-value	dif
10	10	.0	1.00	1.00	.049	.050	1.000	-0.001
10	8	.0	1.00	1.00	.044	.046	1.000	-0.002
10	13	.0	1.00	1.00	.050	.052	1.000	-0.003
10	5	.0	1.00	1.00	.052	.052	1.000	-0.000
10	15	.0	1.00	1.00	.048	.051	1.000	-0.003
10	3	.0	1.00	1.00	.048	.066	0.000	-0.018
10	18	.0	1.00	1.00	.050	.051	1.000	-0.001
10	10	.2	1.00	1.00	.935	.933	0.493	0.002
10	8	.2	1.00	1.00	.936	.933	1.000	0.003
10	13	.2	1.00	1.00	.931	.928	1.000	0.003
10	5	.2	1.00	1.00	.932	.934	1.000	-0.002
10	15	.2	1.00	1.00	.930	.926	1.000	0.004
10	3	.2	1.00	1.00	.943	.933	0.008	0.010
10	18	.2	1.00	1.00	.924	.922	1.000	0.003
10	10	.5	1.00	1.00	.826	.820	0.000	0.006
10	8	.5	1.00	1.00	.834	.831	1.000	0.004
10	13	.5	1.00	1.00	.800	.795	0.365	0.005
10	5	.5	1.00	1.00	.863	.869	1.000	-0.006
10	15	.5	1.00	1.00	.790	.786	1.000	0.003
10	3	.5	1.00	1.00	.892	.888	1.000	0.005

10	18	.5	1.00	1.00	.770	.771	1.000	-0.002
10	10	.8	1.00	1.00	.619	.612	0.000	0.007
10	8	.8	1.00	1.00	.667	.658	0.000	0.009
10	13	.8	1.00	1.00	.574	.570	1.000	0.004
10	5	.8	1.00	1.00	.730	.748	0.000	-0.017
10	15	.8	1.00	1.00	.543	.542	1.000	0.001
10	3	.8	1.00	1.00	.807	.818	0.340	-0.011
10	18	.8	1.00	1.00	.513	.518	1.000	-0.005
10	10	.0	0.75	1.00	.049	.051	0.002	-0.002
10	8	.0	0.75	1.00	.055	.050	0.000	0.006
10	13	.0	0.75	1.00	.042	.052	0.000	-0.010
10	5	.0	0.75	1.00	.068	.053	0.000	0.015
10	15	.0	0.75	1.00	.037	.047	0.000	-0.010
10	3	.0	0.75	1.00	.076	.066	0.012	0.010
10	18	.0	0.75	1.00	.033	.048	0.000	-0.015
10	10	.2	0.75	1.00	.934	.932	0.117	0.002
10	8	.2	0.75	1.00	.924	.932	0.000	-0.007
10	13	.2	0.75	1.00	.940	.926	0.000	0.013
10	5	.2	0.75	1.00	.913	.933	0.000	-0.020
10	15	.2	0.75	1.00	.941	.926	0.000	0.015
10	3	.2	0.75	1.00	.908	.924	0.000	-0.016
10	18	.2	0.75	1.00	.942	.920	0.000	0.022
10	10	.5	0.75	1.00	.816	.814	1.000	0.001
10	8	.5	0.75	1.00	.832	.845	0.000	-0.013
10	13	.5	0.75	1.00	.816	.792	0.000	0.024
10	5	.5	0.75	1.00	.836	.876	0.000	-0.040
10	15	.5	0.75	1.00	.811	.781	0.000	0.031

10	3	.5	0.75	1.00	.856	.891	0.000	-0.035
10	18	.5	0.75	1.00	.802	.750	0.000	0.052
10	10	.8	0.75	1.00	.620	.615	0.001	0.004
10	8	.8	0.75	1.00	.640	.661	0.000	-0.022
10	13	.8	0.75	1.00	.582	.549	0.000	0.033
10	5	.8	0.75	1.00	.693	.771	0.000	-0.078
10	15	.8	0.75	1.00	.565	.513	0.000	0.052
10	3	.8	0.75	1.00	.763	.837	0.000	-0.073
10	18	.8	0.75	1.00	.546	.476	0.000	0.069
10	10	.0	1.25	1.00	.045	.047	0.153	-0.002
10	8	.0	1.25	1.00	.040	.048	0.000	-0.008
10	13	.0	1.25	1.00	.058	.053	0.001	0.005
10	5	.0	1.25	1.00	.039	.050	0.000	-0.012
10	15	.0	1.25	1.00	.055	.047	0.000	0.007
10	3	.0	1.25	1.00	.030	.056	0.000	-0.027
10	18	.0	1.25	1.00	.061	.050	0.000	0.011
10	10	.2	1.25	1.00	.932	.930	0.046	0.002
10	8	.2	1.25	1.00	.941	.933	0.000	0.008
10	13	.2	1.25	1.00	.923	.929	0.000	-0.006
10	5	.2	1.25	1.00	.947	.936	0.000	0.011
10	15	.2	1.25	1.00	.919	.932	0.000	-0.013
10	3	.2	1.25	1.00	.961	.930	0.000	0.031
10	18	.2	1.25	1.00	.910	.926	0.000	-0.016
10	10	.5	1.25	1.00	.821	.817	0.000	0.005
10	8	.5	1.25	1.00	.852	.834	0.000	0.017
10	13	.5	1.25	1.00	.786	.798	0.000	-0.012
10	5	.5	1.25	1.00	.884	.869	0.000	0.015

10	15	.5	1.25	1.00	.775	.799	0.000	-0.024
10	3	.5	1.25	1.00	.924	.886	0.000	0.038
10	18	.5	1.25	1.00	.753	.790	0.000	-0.037
10	10	.8	1.25	1.00	.622	.616	0.000	0.006
10	8	.8	1.25	1.00	.663	.640	0.000	0.023
10	13	.8	1.25	1.00	.556	.581	0.000	-0.025
10	5	.8	1.25	1.00	.752	.726	0.000	0.025
10	15	.8	1.25	1.00	.516	.554	0.000	-0.038
10	3	.8	1.25	1.00	.857	.814	0.000	0.043
10	18	.8	1.25	1.00	.479	.534	0.000	-0.056
10	10	.0	0.50	1.00	.049	.047	0.074	0.002
10	8	.0	0.50	1.00	.072	.054	0.000	0.017
10	13	.0	0.50	1.00	.033	.048	0.000	-0.014
10	5	.0	0.50	1.00	.106	.057	0.000	0.049
10	15	.0	0.50	1.00	.029	.050	0.000	-0.022
10	3	.0	0.50	1.00	.136	.068	0.000	0.068
10	18	.0	0.50	1.00	.020	.048	0.000	-0.028
10	10	.2	0.50	1.00	.924	.926	0.027	-0.002
10	8	.2	0.50	1.00	.915	.935	0.000	-0.020
10	13	.2	0.50	1.00	.948	.928	0.000	0.020
10	5	.2	0.50	1.00	.877	.933	0.000	-0.057
10	15	.2	0.50	1.00	.950	.923	0.000	0.028
10	3	.2	0.50	1.00	.849	.924	0.000	-0.076
10	18	.2	0.50	1.00	.961	.916	0.000	0.046
10	10	.5	0.50	1.00	.819	.823	0.005	-0.004
10	8	.5	0.50	1.00	.807	.846	0.000	-0.039
10	13	.5	0.50	1.00	.824	.782	0.000	0.042

10	5	.5	0.50	1.00	.796	.892	0.000	-0.096
10	15	.5	0.50	1.00	.826	.753	0.000	0.074
10	3	.5	0.50	1.00	.798	.902	0.000	-0.104
10	18	.5	0.50	1.00	.848	.740	0.000	0.108
10	10	.8	0.50	1.00	.612	.620	0.000	-0.008
10	8	.8	0.50	1.00	.616	.680	0.000	-0.064
10	13	.8	0.50	1.00	.599	.530	0.000	0.069
10	5	.8	0.50	1.00	.661	.801	0.000	-0.140
10	15	.8	0.50	1.00	.591	.482	0.000	0.108
10	3	.8	0.50	1.00	.698	.860	0.000	-0.162
10	18	.8	0.50	1.00	.582	.432	0.000	0.150
10	10	.0	1.50	1.00	.051	.052	1.000	-0.001
10	8	.0	1.50	1.00	.036	.044	0.000	-0.008
10	13	.0	1.50	1.00	.063	.052	0.000	0.011
10	5	.0	1.50	1.00	.031	.048	0.000	-0.017
10	15	.0	1.50	1.00	.066	.049	0.000	0.017
10	3	.0	1.50	1.00	.026	.058	0.000	-0.031
10	18	.0	1.50	1.00	.078	.053	0.000	0.024
10	10	.2	1.50	1.00	.931	.931	1.000	-0.000
10	8	.2	1.50	1.00	.942	.930	0.000	0.013
10	13	.2	1.50	1.00	.918	.931	0.000	-0.013
10	5	.2	1.50	1.00	.957	.941	0.000	0.016
10	15	.2	1.50	1.00	.903	.924	0.000	-0.022
10	3	.2	1.50	1.00	.972	.936	0.000	0.036
10	18	.2	1.50	1.00	.897	.930	0.000	-0.033
10	10	.5	1.50	1.00	.813	.812	1.000	0.001
10	8	.5	1.50	1.00	.856	.832	0.000	0.024

10	13	.5	1.50	1.00	.776	.808	0.000	-0.032
10	5	.5	1.50	1.00	.892	.857	0.000	0.035
10	15	.5	1.50	1.00	.752	.799	0.000	-0.047
10	3	.5	1.50	1.00	.937	.887	0.000	0.050
10	18	.5	1.50	1.00	.725	.799	0.000	-0.073
10	10	.8	1.50	1.00	.619	.616	0.139	0.003
10	8	.8	1.50	1.00	.680	.640	0.000	0.039
10	13	.8	1.50	1.00	.550	.596	0.000	-0.045
10	5	.8	1.50	1.00	.765	.706	0.000	0.059
10	15	.8	1.50	1.00	.497	.567	0.000	-0.070
10	3	.8	1.50	1.00	.862	.786	0.000	0.076
10	18	.8	1.50	1.00	.464	.565	0.000	-0.101
10	10	.0	0.25	1.00	.059	.052	0.000	0.007
10	8	.0	0.25	1.00	.086	.052	0.000	0.034
10	13	.0	0.25	1.00	.029	.047	0.000	-0.018
10	5	.0	0.25	1.00	.155	.057	0.000	0.098
10	15	.0	0.25	1.00	.022	.048	0.000	-0.025
10	3	.0	0.25	1.00	.222	.065	0.000	0.158
10	18	.0	0.25	1.00	.015	.048	0.000	-0.033
10	10	.2	0.25	1.00	.916	.927	0.000	-0.011
10	8	.2	0.25	1.00	.894	.934	0.000	-0.040
10	13	.2	0.25	1.00	.944	.921	0.000	0.023
10	5	.2	0.25	1.00	.824	.939	0.000	-0.115
10	15	.2	0.25	1.00	.959	.920	0.000	0.040
10	3	.2	0.25	1.00	.771	.933	0.000	-0.162
10	18	.2	0.25	1.00	.973	.919	0.000	0.054
10	10	.5	0.25	1.00	.804	.820	0.000	-0.015

10	8	.5	0.25	1.00	.784	.855	0.000	-0.072
10	13	.5	0.25	1.00	.837	.787	0.000	0.050
10	5	.5	0.25	1.00	.728	.895	0.000	-0.167
10	15	.5	0.25	1.00	.847	.751	0.000	0.096
10	3	.5	0.25	1.00	.710	.911	0.000	-0.201
10	18	.5	0.25	1.00	.865	.715	0.000	0.150
10	10	.8	0.25	1.00	.608	.635	0.000	-0.027
10	8	.8	0.25	1.00	.603	.707	0.000	-0.104
10	13	.8	0.25	1.00	.612	.533	0.000	0.079
10	5	.8	0.25	1.00	.586	.817	0.000	-0.230
10	15	.8	0.25	1.00	.620	.477	0.000	0.143
10	3	.8	0.25	1.00	.628	.881	0.000	-0.253
10	18	.8	0.25	1.00	.614	.392	0.000	0.222
10	10	.0	1.75	1.00	.051	.051	1.000	0.000
10	8	.0	1.75	1.00	.038	.048	0.000	-0.009
10	13	.0	1.75	1.00	.060	.046	0.000	0.014
10	5	.0	1.75	1.00	.028	.047	0.000	-0.019
10	15	.0	1.75	1.00	.074	.048	0.000	0.025
10	3	.0	1.75	1.00	.018	.048	0.000	-0.030
10	18	.0	1.75	1.00	.087	.050	0.000	0.037
10	10	.2	1.75	1.00	.932	.932	1.000	-0.000
10	8	.2	1.75	1.00	.950	.936	0.000	0.013
10	13	.2	1.75	1.00	.910	.929	0.000	-0.019
10	5	.2	1.75	1.00	.962	.938	0.000	0.024
10	15	.2	1.75	1.00	.896	.929	0.000	-0.033
10	3	.2	1.75	1.00	.977	.936	0.000	0.041
10	18	.2	1.75	1.00	.872	.921	0.000	-0.049

10	10	.5	1.75	1.00	.820	.820	1.000	-0.001
10	8	.5	1.75	1.00	.849	.820	0.000	0.029
10	13	.5	1.75	1.00	.759	.802	0.000	-0.043
10	5	.5	1.75	1.00	.903	.856	0.000	0.047
10	15	.5	1.75	1.00	.744	.808	0.000	-0.064
10	3	.5	1.75	1.00	.946	.882	0.000	0.065
10	18	.5	1.75	1.00	.718	.810	0.000	-0.092
10	10	.8	1.75	1.00	.610	.614	0.001	-0.005
10	8	.8	1.75	1.00	.679	.632	0.000	0.047
10	13	.8	1.75	1.00	.530	.594	0.000	-0.064
10	5	.8	1.75	1.00	.776	.693	0.000	0.083
10	15	.8	1.75	1.00	.504	.600	0.000	-0.096
10	3	.8	1.75	1.00	.874	.768	0.000	0.106
10	18	.8	1.75	1.00	.445	.582	0.000	-0.136
10	10	.0	3.00	1.00	.059	.054	0.000	0.005
10	8	.0	3.00	1.00	.038	.050	0.000	-0.013
10	13	.0	3.00	1.00	.082	.052	0.000	0.031
10	5	.0	3.00	1.00	.023	.050	0.000	-0.027
10	15	.0	3.00	1.00	.106	.054	0.000	0.052
10	3	.0	3.00	1.00	.009	.049	0.000	-0.040
10	18	.0	3.00	1.00	.126	.049	0.000	0.078
10	10	.2	3.00	1.00	.930	.935	0.000	-0.005
10	8	.2	3.00	1.00	.945	.927	0.000	0.017
10	13	.2	3.00	1.00	.890	.933	0.000	-0.043
10	5	.2	3.00	1.00	.970	.933	0.000	0.037
10	15	.2	3.00	1.00	.868	.928	0.000	-0.060
10	3	.2	3.00	1.00	.987	.934	0.000	0.053

10	18	.2	3.00	1.00	.834	.930	0.000	-0.096
10	10	.5	3.00	1.00	.815	.826	0.000	-0.012
10	8	.5	3.00	1.00	.862	.829	0.000	0.033
10	13	.5	3.00	1.00	.744	.821	0.000	-0.077
10	5	.5	3.00	1.00	.911	.830	0.000	0.081
10	15	.5	3.00	1.00	.704	.814	0.000	-0.110
10	3	.5	3.00	1.00	.963	.861	0.000	0.101
10	18	.5	3.00	1.00	.663	.815	0.000	-0.152
10	10	.8	3.00	1.00	.610	.631	0.000	-0.021
10	8	.8	3.00	1.00	.690	.632	0.000	0.058
10	13	.8	3.00	1.00	.514	.619	0.000	-0.105
10	5	.8	3.00	1.00	.798	.658	0.000	0.139
10	15	.8	3.00	1.00	.471	.620	0.000	-0.149
10	3	.8	3.00	1.00	.908	.705	0.000	0.203
10	18	.8	3.00	1.00	.407	.606	0.000	-0.199

See Table 3 for explanation on column names

Table B2

Simulation results for sample size 60 and its deviations

N_1	N_2	ES	σ_1	σ_2	perm	t_test	p-value	dif
60	60	.0	1.00	1.00	.049	.050	1.000	-0.000
60	45	.0	1.00	1.00	.051	.050	1.000	0.000
60	75	.0	1.00	1.00	.050	.050	1.000	-0.000
60	30	.0	1.00	1.00	.049	.050	1.000	-0.001
60	90	.0	1.00	1.00	.050	.050	1.000	-0.001
60	15	.0	1.00	1.00	.049	.051	1.000	-0.001
60	105	.0	1.00	1.00	.050	.051	1.000	-0.000
60	60	.2	1.00	1.00	.808	.806	0.335	0.001
60	45	.2	1.00	1.00	.826	.825	1.000	0.001
60	75	.2	1.00	1.00	.796	.797	1.000	-0.001
60	30	.2	1.00	1.00	.852	.851	1.000	0.000
60	90	.2	1.00	1.00	.775	.773	1.000	0.002
60	15	.2	1.00	1.00	.896	.899	1.000	-0.004
60	105	.2	1.00	1.00	.764	.765	1.000	-0.001
60	60	.5	1.00	1.00	.225	.224	0.335	0.001
60	45	.5	1.00	1.00	.292	.291	1.000	0.002
60	75	.5	1.00	1.00	.186	.184	1.000	0.001
60	30	.5	1.00	1.00	.396	.404	0.009	-0.008
60	90	.5	1.00	1.00	.153	.152	1.000	0.001
60	15	.5	1.00	1.00	.600	.619	0.000	-0.018
60	105	.5	1.00	1.00	.135	.134	1.000	0.001
60	60	.8	1.00	1.00	.009	.009	1.000	0.000
60	45	.8	1.00	1.00	.019	.020	1.000	-0.001

60	75	.8	1.00	1.00	.004	.004	1.000	0.000
60	30	.8	1.00	1.00	.055	.056	1.000	-0.001
60	90	.8	1.00	1.00	.003	.003	1.000	-0.000
60	15	.8	1.00	1.00	.217	.240	0.000	-0.023
60	105	.8	1.00	1.00	.002	.002	1.000	-0.000
60	60	.0	0.75	1.00	.051	.051	1.000	-0.000
60	45	.0	0.75	1.00	.058	.047	0.000	0.011
60	75	.0	0.75	1.00	.042	.047	0.000	-0.005
60	30	.0	0.75	1.00	.076	.051	0.000	0.025
60	90	.0	0.75	1.00	.040	.053	0.000	-0.012
60	15	.0	0.75	1.00	.090	.042	0.000	0.048
60	105	.0	0.75	1.00	.036	.051	0.000	-0.015
60	60	.2	0.75	1.00	.804	.804	1.000	0.000
60	45	.2	0.75	1.00	.810	.832	0.000	-0.022
60	75	.2	0.75	1.00	.807	.786	0.000	0.021
60	30	.2	0.75	1.00	.826	.870	0.000	-0.044
60	90	.2	0.75	1.00	.801	.765	0.000	0.036
60	15	.2	0.75	1.00	.835	.908	0.000	-0.073
60	105	.2	0.75	1.00	.804	.759	0.000	0.044
60	60	.5	0.75	1.00	.227	.227	1.000	0.001
60	45	.5	0.75	1.00	.280	.309	0.000	-0.029
60	75	.5	0.75	1.00	.183	.166	0.000	0.017
60	30	.5	0.75	1.00	.366	.438	0.000	-0.073
60	90	.5	0.75	1.00	.168	.137	0.000	0.030
60	15	.5	0.75	1.00	.526	.664	0.000	-0.138
60	105	.5	0.75	1.00	.138	.108	0.000	0.030
60	60	.8	0.75	1.00	.009	.009	1.000	0.000

60	45	.8	0.75	1.00	.020	.025	0.000	-0.004
60	75	.8	0.75	1.00	.004	.004	0.992	0.001
60	30	.8	0.75	1.00	.058	.086	0.000	-0.028
60	90	.8	0.75	1.00	.002	.001	1.000	0.001
60	15	.8	0.75	1.00	.192	.319	0.000	-0.127
60	105	.8	0.75	1.00	.001	.001	1.000	0.000
60	60	.0	1.25	1.00	.051	.052	1.000	-0.000
60	45	.0	1.25	1.00	.043	.051	0.000	-0.008
60	75	.0	1.25	1.00	.050	.046	0.000	0.005
60	30	.0	1.25	1.00	.032	.047	0.000	-0.015
60	90	.0	1.25	1.00	.058	.049	0.000	0.010
60	15	.0	1.25	1.00	.024	.051	0.000	-0.027
60	105	.0	1.25	1.00	.068	.051	0.000	0.017
60	60	.2	1.25	1.00	.805	.805	1.000	0.001
60	45	.2	1.25	1.00	.839	.822	0.000	0.017
60	75	.2	1.25	1.00	.776	.788	0.000	-0.012
60	30	.2	1.25	1.00	.884	.850	0.000	0.034
60	90	.2	1.25	1.00	.762	.788	0.000	-0.026
60	15	.2	1.25	1.00	.935	.890	0.000	0.045
60	105	.2	1.25	1.00	.740	.774	0.000	-0.034
60	60	.5	1.25	1.00	.230	.229	1.000	0.001
60	45	.5	1.25	1.00	.294	.271	0.000	0.023
60	75	.5	1.25	1.00	.174	.186	0.000	-0.012
60	30	.5	1.25	1.00	.425	.370	0.000	0.055
60	90	.5	1.25	1.00	.142	.164	0.000	-0.022
60	15	.5	1.25	1.00	.659	.567	0.000	0.092
60	105	.5	1.25	1.00	.127	.156	0.000	-0.028

60	60	.8	1.25	1.00	.008	.008	1.000	0.000
60	45	.8	1.25	1.00	.021	.018	0.000	0.003
60	75	.8	1.25	1.00	.005	.006	0.992	-0.001
60	30	.8	1.25	1.00	.060	.045	0.000	0.015
60	90	.8	1.25	1.00	.003	.004	1.000	-0.001
60	15	.8	1.25	1.00	.248	.184	0.000	0.064
60	105	.8	1.25	1.00	.002	.002	1.000	-0.001
60	60	.0	0.50	1.00	.054	.054	1.000	0.000
60	45	.0	0.50	1.00	.073	.049	0.000	0.024
60	75	.0	0.50	1.00	.038	.056	0.000	-0.017
60	30	.0	0.50	1.00	.112	.053	0.000	0.058
60	90	.0	0.50	1.00	.028	.051	0.000	-0.023
60	15	.0	0.50	1.00	.181	.049	0.000	0.132
60	105	.0	0.50	1.00	.022	.050	0.000	-0.028
60	60	.2	0.50	1.00	.804	.804	1.000	-0.001
60	45	.2	0.50	1.00	.793	.838	0.000	-0.044
60	75	.2	0.50	1.00	.814	.778	0.000	0.036
60	30	.2	0.50	1.00	.777	.873	0.000	-0.096
60	90	.2	0.50	1.00	.827	.755	0.000	0.072
60	15	.2	0.50	1.00	.746	.913	0.000	-0.167
60	105	.2	0.50	1.00	.830	.722	0.000	0.108
60	60	.5	0.50	1.00	.226	.226	1.000	-0.001
60	45	.5	0.50	1.00	.266	.329	0.000	-0.063
60	75	.5	0.50	1.00	.196	.163	0.000	0.033
60	30	.5	0.50	1.00	.337	.495	0.000	-0.158
60	90	.5	0.50	1.00	.169	.112	0.000	0.058
60	15	.5	0.50	1.00	.444	.714	0.000	-0.270

60	105	.5	0.50	1.00	.150	.082	0.000	0.068
60	60	.8	0.50	1.00	.009	.009	1.000	0.000
60	45	.8	0.50	1.00	.018	.030	0.000	-0.012
60	75	.8	0.50	1.00	.004	.002	0.008	0.002
60	30	.8	0.50	1.00	.052	.111	0.000	-0.059
60	90	.8	0.50	1.00	.002	.001	0.196	0.001
60	15	.8	0.50	1.00	.159	.396	0.000	-0.237
60	105	.8	0.50	1.00	.001	.001	1.000	0.000
60	60	.0	1.50	1.00	.047	.047	1.000	-0.000
60	45	.0	1.50	1.00	.037	.049	0.000	-0.012
60	75	.0	1.50	1.00	.057	.048	0.000	0.009
60	30	.0	1.50	1.00	.026	.050	0.000	-0.025
60	90	.0	1.50	1.00	.069	.049	0.000	0.019
60	15	.0	1.50	1.00	.012	.051	0.000	-0.039
60	105	.0	1.50	1.00	.077	.049	0.000	0.029
60	60	.2	1.50	1.00	.807	.807	1.000	0.000
60	45	.2	1.50	1.00	.847	.820	0.000	0.027
60	75	.2	1.50	1.00	.771	.798	0.000	-0.027
60	30	.2	1.50	1.00	.901	.845	0.000	0.057
60	90	.2	1.50	1.00	.739	.784	0.000	-0.044
60	15	.2	1.50	1.00	.956	.880	0.000	0.075
60	105	.2	1.50	1.00	.722	.788	0.000	-0.066
60	60	.5	1.50	1.00	.222	.222	1.000	-0.000
60	45	.5	1.50	1.00	.308	.270	0.000	0.037
60	75	.5	1.50	1.00	.175	.197	0.000	-0.021
60	30	.5	1.50	1.00	.446	.342	0.000	0.104
60	90	.5	1.50	1.00	.150	.187	0.000	-0.037

60	15	.5	1.50	1.00	.701	.506	0.000	0.195
60	105	.5	1.50	1.00	.124	.174	0.000	-0.049
60	60	.8	1.50	1.00	.008	.008	1.000	0.000
60	45	.8	1.50	1.00	.021	.015	0.000	0.005
60	75	.8	1.50	1.00	.005	.006	0.196	-0.001
60	30	.8	1.50	1.00	.062	.034	0.000	0.029
60	90	.8	1.50	1.00	.003	.004	1.000	-0.001
60	15	.8	1.50	1.00	.266	.139	0.000	0.127
60	105	.8	1.50	1.00	.001	.004	0.000	-0.003
60	60	.0	0.25	1.00	.053	.052	0.575	0.001
60	45	.0	0.25	1.00	.084	.048	0.000	0.036
60	75	.0	0.25	1.00	.030	.050	0.000	-0.019
60	30	.0	0.25	1.00	.156	.052	0.000	0.104
60	90	.0	0.25	1.00	.020	.048	0.000	-0.028
60	15	.0	0.25	1.00	.294	.051	0.000	0.242
60	105	.0	0.25	1.00	.013	.050	0.000	-0.037
60	60	.2	0.25	1.00	.804	.808	0.000	-0.004
60	45	.2	0.25	1.00	.778	.845	0.000	-0.067
60	75	.2	0.25	1.00	.831	.777	0.000	0.054
60	30	.2	0.25	1.00	.727	.875	0.000	-0.148
60	90	.2	0.25	1.00	.849	.739	0.000	0.110
60	15	.2	0.25	1.00	.654	.918	0.000	-0.264
60	105	.2	0.25	1.00	.863	.707	0.000	0.156
60	60	.5	0.25	1.00	.225	.229	0.000	-0.003
60	45	.5	0.25	1.00	.252	.340	0.000	-0.088
60	75	.5	0.25	1.00	.201	.151	0.000	0.050
60	30	.5	0.25	1.00	.311	.521	0.000	-0.209

60	90	.5	0.25	1.00	.179	.096	0.000	0.083
60	15	.5	0.25	1.00	.375	.740	0.000	-0.365
60	105	.5	0.25	1.00	.155	.060	0.000	0.094
60	60	.8	0.25	1.00	.008	.008	1.000	-0.000
60	45	.8	0.25	1.00	.021	.038	0.000	-0.017
60	75	.8	0.25	1.00	.003	.002	0.067	0.001
60	30	.8	0.25	1.00	.045	.140	0.000	-0.095
60	90	.8	0.25	1.00	.002	.000	0.196	0.001
60	15	.8	0.25	1.00	.125	.449	0.000	-0.325
60	105	.8	0.25	1.00	.001	.000	1.000	0.001
60	60	.0	1.75	1.00	.050	.050	1.000	0.000
60	45	.0	1.75	1.00	.033	.048	0.000	-0.016
60	75	.0	1.75	1.00	.065	.050	0.000	0.015
60	30	.0	1.75	1.00	.018	.049	0.000	-0.031
60	90	.0	1.75	1.00	.075	.049	0.000	0.026
60	15	.0	1.75	1.00	.009	.052	0.000	-0.043
60	105	.0	1.75	1.00	.083	.046	0.000	0.037
60	60	.2	1.75	1.00	.806	.807	1.000	-0.001
60	45	.2	1.75	1.00	.859	.824	0.000	0.035
60	75	.2	1.75	1.00	.759	.793	0.000	-0.034
60	30	.2	1.75	1.00	.910	.834	0.000	0.076
60	90	.2	1.75	1.00	.732	.793	0.000	-0.060
60	15	.2	1.75	1.00	.971	.875	0.000	0.096
60	105	.2	1.75	1.00	.706	.790	0.000	-0.085
60	60	.5	1.75	1.00	.228	.228	1.000	-0.001
60	45	.5	1.75	1.00	.315	.264	0.000	0.051
60	75	.5	1.75	1.00	.174	.203	0.000	-0.029

60	30	.5	1.75	1.00	.456	.313	0.000	0.143
60	90	.5	1.75	1.00	.143	.193	0.000	-0.050
60	15	.5	1.75	1.00	.734	.474	0.000	0.260
60	105	.5	1.75	1.00	.120	.184	0.000	-0.064
60	60	.8	1.75	1.00	.009	.009	1.000	0.000
60	45	.8	1.75	1.00	.020	.012	0.000	0.008
60	75	.8	1.75	1.00	.006	.007	0.115	-0.001
60	30	.8	1.75	1.00	.062	.028	0.000	0.034
60	90	.8	1.75	1.00	.003	.005	0.000	-0.003
60	15	.8	1.75	1.00	.281	.101	0.000	0.180
60	105	.8	1.75	1.00	.002	.005	0.000	-0.003
60	60	.0	3.00	1.00	.052	.051	0.575	0.001
60	45	.0	3.00	1.00	.030	.054	0.000	-0.024
60	75	.0	3.00	1.00	.076	.050	0.000	0.025
60	30	.0	3.00	1.00	.012	.052	0.000	-0.040
60	90	.0	3.00	1.00	.101	.052	0.000	0.049
60	15	.0	3.00	1.00	.001	.050	0.000	-0.049
60	105	.0	3.00	1.00	.122	.048	0.000	0.075
60	60	.2	3.00	1.00	.815	.818	0.000	-0.003
60	45	.2	3.00	1.00	.870	.812	0.000	0.058
60	75	.2	3.00	1.00	.754	.808	0.000	-0.054
60	30	.2	3.00	1.00	.937	.824	0.000	0.113
60	90	.2	3.00	1.00	.706	.804	0.000	-0.098
60	15	.2	3.00	1.00	.990	.840	0.000	0.151
60	105	.2	3.00	1.00	.677	.810	0.000	-0.133
60	60	.5	3.00	1.00	.224	.227	0.000	-0.003
60	45	.5	3.00	1.00	.321	.239	0.000	0.081

60	75	.5	3.00	1.00	.168	.217	0.000	-0.049
60	30	.5	3.00	1.00	.501	.272	0.000	0.229
60	90	.5	3.00	1.00	.134	.212	0.000	-0.078
60	15	.5	3.00	1.00	.826	.338	0.000	0.488
60	105	.5	3.00	1.00	.111	.202	0.000	-0.091
60	60	.8	3.00	1.00	.008	.008	1.000	-0.000
60	45	.8	3.00	1.00	.021	.012	0.000	0.009
60	75	.8	3.00	1.00	.005	.009	0.000	-0.004
60	30	.8	3.00	1.00	.065	.016	0.000	0.049
60	90	.8	3.00	1.00	.003	.009	0.000	-0.006
60	15	.8	3.00	1.00	.330	.031	0.000	0.298
60	105	.8	3.00	1.00	.002	.006	0.000	-0.004

see Table 3 for explanation on column names

Table B3

Simulation results for sample size 1000 and its deviations

N_1	N_2	ES	σ_1	σ_2	perm	t-test	p-value	dif
1000	1000	.0	1.00	1.00	.049	.049	1.000	0.000
1000	750	.0	1.00	1.00	.048	.048	1.000	0.000
1000	1250	.0	1.00	1.00	.048	.048	1.000	0.000
1000	500	.0	1.00	1.00	.046	.046	1.000	0.000
1000	1500	.0	1.00	1.00	.047	.048	1.000	-0.000
1000	250	.0	1.00	1.00	.052	.052	1.000	-0.000
1000	1750	.0	1.00	1.00	.048	.048	1.000	0.000
1000	1000	.2	1.00	1.00	.006	.006	1.000	0.000
1000	750	.2	1.00	1.00	.013	.013	1.000	0.000
1000	1250	.2	1.00	1.00	.002	.002	1.000	0.000
1000	500	.2	1.00	1.00	.047	.047	1.000	0.000
1000	1500	.2	1.00	1.00	.002	.002	1.000	0.000
1000	250	.2	1.00	1.00	.193	.194	1.000	-0.000
1000	1750	.2	1.00	1.00	.001	.001	1.000	-0.000
1000	1000	.5	1.00	1.00	.000	.000	1.000	0.000
1000	750	.5	1.00	1.00	.000	.000	1.000	0.000
1000	1250	.5	1.00	1.00	.000	.000	1.000	0.000
1000	500	.5	1.00	1.00	.000	.000	1.000	0.000
1000	1500	.5	1.00	1.00	.000	.000	1.000	0.000
1000	250	.5	1.00	1.00	.000	.000	1.000	0.000
1000	1750	.5	1.00	1.00	.000	.000	1.000	0.000
1000	1000	.8	1.00	1.00	.000	.000	1.000	0.000
1000	750	.8	1.00	1.00	.000	.000	1.000	0.000

1000	1250	.8	1.00	1.00	.000	.000	1.000	0.000
1000	500	.8	1.00	1.00	.000	.000	1.000	0.000
1000	1500	.8	1.00	1.00	.000	.000	1.000	0.000
1000	250	.8	1.00	1.00	.000	.000	1.000	0.000
1000	1750	.8	1.00	1.00	.000	.000	1.000	0.000
1000	1000	.0	0.75	1.00	.053	.053	1.000	0.000
1000	750	.0	0.75	1.00	.064	.052	0.000	0.011
1000	1250	.0	0.75	1.00	.046	.053	0.000	-0.007
1000	500	.0	0.75	1.00	.073	.048	0.000	0.025
1000	1500	.0	0.75	1.00	.037	.050	0.000	-0.012
1000	250	.0	0.75	1.00	.099	.050	0.000	0.049
1000	1750	.0	0.75	1.00	.034	.050	0.000	-0.016
1000	1000	.2	0.75	1.00	.006	.006	1.000	0.000
1000	750	.2	0.75	1.00	.016	.019	0.000	-0.003
1000	1250	.2	0.75	1.00	.002	.002	1.000	0.000
1000	500	.2	0.75	1.00	.045	.065	0.000	-0.020
1000	1500	.2	0.75	1.00	.002	.001	1.000	0.001
1000	250	.2	0.75	1.00	.170	.261	0.000	-0.091
1000	1750	.2	0.75	1.00	.001	.001	1.000	0.000
1000	1000	.5	0.75	1.00	.000	.000	1.000	0.000
1000	750	.5	0.75	1.00	.000	.000	1.000	0.000
1000	1250	.5	0.75	1.00	.000	.000	1.000	0.000
1000	500	.5	0.75	1.00	.000	.000	1.000	0.000
1000	1500	.5	0.75	1.00	.000	.000	1.000	0.000
1000	250	.5	0.75	1.00	.000	.000	1.000	0.000
1000	1750	.5	0.75	1.00	.000	.000	1.000	0.000
1000	1000	.8	0.75	1.00	.000	.000	1.000	0.000

1000	750	.8	0.75	1.00	.000	.000	1.000	0.000
1000	1250	.8	0.75	1.00	.000	.000	1.000	0.000
1000	500	.8	0.75	1.00	.000	.000	1.000	0.000
1000	1500	.8	0.75	1.00	.000	.000	1.000	0.000
1000	250	.8	0.75	1.00	.000	.000	1.000	0.000
1000	1750	.8	0.75	1.00	.000	.000	1.000	0.000
1000	1000	.0	1.25	1.00	.050	.050	1.000	0.000
1000	750	.0	1.25	1.00	.043	.049	0.000	-0.006
1000	1250	.0	1.25	1.00	.056	.051	0.000	0.005
1000	500	.0	1.25	1.00	.035	.048	0.000	-0.013
1000	1500	.0	1.25	1.00	.062	.050	0.000	0.012
1000	250	.0	1.25	1.00	.026	.051	0.000	-0.025
1000	1750	.0	1.25	1.00	.066	.053	0.000	0.014
1000	1000	.2	1.25	1.00	.005	.005	1.000	0.000
1000	750	.2	1.25	1.00	.015	.013	0.001	0.002
1000	1250	.2	1.25	1.00	.003	.004	1.000	-0.001
1000	500	.2	1.25	1.00	.048	.034	0.000	0.014
1000	1500	.2	1.25	1.00	.002	.002	1.000	-0.000
1000	250	.2	1.25	1.00	.213	.142	0.000	0.072
1000	1750	.2	1.25	1.00	.001	.001	1.000	-0.000
1000	1000	.5	1.25	1.00	.000	.000	1.000	0.000
1000	750	.5	1.25	1.00	.000	.000	1.000	0.000
1000	1250	.5	1.25	1.00	.000	.000	1.000	0.000
1000	500	.5	1.25	1.00	.000	.000	1.000	0.000
1000	1500	.5	1.25	1.00	.000	.000	1.000	0.000
1000	250	.5	1.25	1.00	.000	.000	1.000	0.000
1000	1750	.5	1.25	1.00	.000	.000	1.000	0.000

1000	1000	.8	1.25	1.00	.000	.000	1.000	0.000
1000	750	.8	1.25	1.00	.000	.000	1.000	0.000
1000	1250	.8	1.25	1.00	.000	.000	1.000	0.000
1000	500	.8	1.25	1.00	.000	.000	1.000	0.000
1000	1500	.8	1.25	1.00	.000	.000	1.000	0.000
1000	250	.8	1.25	1.00	.000	.000	1.000	0.000
1000	1750	.8	1.25	1.00	.000	.000	1.000	0.000
1000	1000	.0	0.50	1.00	.051	.051	1.000	0.000
1000	750	.0	0.50	1.00	.065	.046	0.000	0.019
1000	1250	.0	0.50	1.00	.038	.053	0.000	-0.015
1000	500	.0	0.50	1.00	.109	.050	0.000	0.059
1000	1500	.0	0.50	1.00	.031	.054	0.000	-0.022
1000	250	.0	0.50	1.00	.174	.050	0.000	0.124
1000	1750	.0	0.50	1.00	.020	.048	0.000	-0.028
1000	1000	.2	0.50	1.00	.006	.006	1.000	0.000
1000	750	.2	0.50	1.00	.016	.023	0.000	-0.008
1000	1250	.2	0.50	1.00	.002	.002	1.000	0.001
1000	500	.2	0.50	1.00	.040	.086	0.000	-0.045
1000	1500	.2	0.50	1.00	.002	.001	0.992	0.001
1000	250	.2	0.50	1.00	.142	.320	0.000	-0.179
1000	1750	.2	0.50	1.00	.000	.000	1.000	0.000
1000	1000	.5	0.50	1.00	.000	.000	1.000	0.000
1000	750	.5	0.50	1.00	.000	.000	1.000	0.000
1000	1250	.5	0.50	1.00	.000	.000	1.000	0.000
1000	500	.5	0.50	1.00	.000	.000	1.000	0.000
1000	1500	.5	0.50	1.00	.000	.000	1.000	0.000
1000	250	.5	0.50	1.00	.000	.000	1.000	0.000

1000	1750	.5	0.50	1.00	.000	.000	1.000	0.000
1000	1000	.8	0.50	1.00	.000	.000	1.000	0.000
1000	750	.8	0.50	1.00	.000	.000	1.000	0.000
1000	1250	.8	0.50	1.00	.000	.000	1.000	0.000
1000	500	.8	0.50	1.00	.000	.000	1.000	0.000
1000	1500	.8	0.50	1.00	.000	.000	1.000	0.000
1000	250	.8	0.50	1.00	.000	.000	1.000	0.000
1000	1750	.8	0.50	1.00	.000	.000	1.000	0.000
1000	1000	.0	1.50	1.00	.051	.051	1.000	0.000
1000	750	.0	1.50	1.00	.035	.048	0.000	-0.013
1000	1250	.0	1.50	1.00	.062	.050	0.000	0.012
1000	500	.0	1.50	1.00	.026	.050	0.000	-0.024
1000	1500	.0	1.50	1.00	.072	.051	0.000	0.020
1000	250	.0	1.50	1.00	.012	.054	0.000	-0.042
1000	1750	.0	1.50	1.00	.076	.052	0.000	0.024
1000	1000	.2	1.50	1.00	.008	.008	1.000	0.000
1000	750	.2	1.50	1.00	.015	.011	0.000	0.003
1000	1250	.2	1.50	1.00	.004	.005	0.575	-0.001
1000	500	.2	1.50	1.00	.048	.026	0.000	0.022
1000	1500	.2	1.50	1.00	.002	.003	0.115	-0.001
1000	250	.2	1.50	1.00	.234	.103	0.000	0.132
1000	1750	.2	1.50	1.00	.001	.002	0.992	-0.001
1000	1000	.5	1.50	1.00	.000	.000	1.000	0.000
1000	750	.5	1.50	1.00	.000	.000	1.000	0.000
1000	1250	.5	1.50	1.00	.000	.000	1.000	0.000
1000	500	.5	1.50	1.00	.000	.000	1.000	0.000
1000	1500	.5	1.50	1.00	.000	.000	1.000	0.000

1000	250	.5	1.50	1.00	.000	.000	1.000	0.000
1000	1750	.5	1.50	1.00	.000	.000	1.000	0.000
1000	1000	.8	1.50	1.00	.000	.000	1.000	0.000
1000	750	.8	1.50	1.00	.000	.000	1.000	0.000
1000	1250	.8	1.50	1.00	.000	.000	1.000	0.000
1000	500	.8	1.50	1.00	.000	.000	1.000	0.000
1000	1500	.8	1.50	1.00	.000	.000	1.000	0.000
1000	250	.8	1.50	1.00	.000	.000	1.000	0.000
1000	1750	.8	1.50	1.00	.000	.000	1.000	0.000
1000	1000	.0	0.25	1.00	.050	.050	1.000	0.000
1000	750	.0	0.25	1.00	.088	.052	0.000	0.036
1000	1250	.0	0.25	1.00	.031	.050	0.000	-0.019
1000	500	.0	0.25	1.00	.140	.047	0.000	0.093
1000	1500	.0	0.25	1.00	.020	.051	0.000	-0.031
1000	250	.0	0.25	1.00	.282	.053	0.000	0.229
1000	1750	.0	0.25	1.00	.012	.051	0.000	-0.039
1000	1000	.2	0.25	1.00	.006	.006	1.000	0.000
1000	750	.2	0.25	1.00	.015	.027	0.000	-0.012
1000	1250	.2	0.25	1.00	.003	.001	0.002	0.002
1000	500	.2	0.25	1.00	.040	.105	0.000	-0.066
1000	1500	.2	0.25	1.00	.001	.000	1.000	0.001
1000	250	.2	0.25	1.00	.116	.374	0.000	-0.258
1000	1750	.2	0.25	1.00	.001	.000	1.000	0.000
1000	1000	.5	0.25	1.00	.000	.000	1.000	0.000
1000	750	.5	0.25	1.00	.000	.000	1.000	0.000
1000	1250	.5	0.25	1.00	.000	.000	1.000	0.000
1000	500	.5	0.25	1.00	.000	.000	1.000	0.000

1000	1500	.5	0.25	1.00	.000	.000	1.000	0.000
1000	250	.5	0.25	1.00	.000	.000	1.000	-0.000
1000	1750	.5	0.25	1.00	.000	.000	1.000	0.000
1000	1000	.8	0.25	1.00	.000	.000	1.000	0.000
1000	750	.8	0.25	1.00	.000	.000	1.000	0.000
1000	1250	.8	0.25	1.00	.000	.000	1.000	0.000
1000	500	.8	0.25	1.00	.000	.000	1.000	0.000
1000	1500	.8	0.25	1.00	.000	.000	1.000	0.000
1000	250	.8	0.25	1.00	.000	.000	1.000	0.000
1000	1750	.8	0.25	1.00	.000	.000	1.000	0.000
1000	1000	.0	1.75	1.00	.051	.051	1.000	0.000
1000	750	.0	1.75	1.00	.034	.048	0.000	-0.014
1000	1250	.0	1.75	1.00	.061	.048	0.000	0.013
1000	500	.0	1.75	1.00	.020	.052	0.000	-0.032
1000	1500	.0	1.75	1.00	.072	.044	0.000	0.028
1000	250	.0	1.75	1.00	.006	.050	0.000	-0.044
1000	1750	.0	1.75	1.00	.090	.050	0.000	0.041
1000	1000	.2	1.75	1.00	.006	.006	1.000	0.000
1000	750	.2	1.75	1.00	.015	.010	0.000	0.004
1000	1250	.2	1.75	1.00	.004	.005	0.040	-0.002
1000	500	.2	1.75	1.00	.043	.018	0.000	0.026
1000	1500	.2	1.75	1.00	.002	.004	0.115	-0.001
1000	250	.2	1.75	1.00	.249	.076	0.000	0.173
1000	1750	.2	1.75	1.00	.001	.004	0.000	-0.002
1000	1000	.5	1.75	1.00	.000	.000	1.000	0.000
1000	750	.5	1.75	1.00	.000	.000	1.000	0.000
1000	1250	.5	1.75	1.00	.000	.000	1.000	0.000

1000	500	.5	1.75	1.00	.000	.000	1.000	0.000
1000	1500	.5	1.75	1.00	.000	.000	1.000	0.000
1000	250	.5	1.75	1.00	.000	.000	1.000	0.000
1000	1750	.5	1.75	1.00	.000	.000	1.000	0.000
1000	1000	.8	1.75	1.00	.000	.000	1.000	0.000
1000	750	.8	1.75	1.00	.000	.000	1.000	0.000
1000	1250	.8	1.75	1.00	.000	.000	1.000	0.000
1000	500	.8	1.75	1.00	.000	.000	1.000	0.000
1000	1500	.8	1.75	1.00	.000	.000	1.000	0.000
1000	250	.8	1.75	1.00	.000	.000	1.000	0.000
1000	1750	.8	1.75	1.00	.000	.000	1.000	0.000
1000	1000	.0	3.00	1.00	.054	.054	1.000	0.000
1000	750	.0	3.00	1.00	.026	.051	0.000	-0.025
1000	1250	.0	3.00	1.00	.071	.049	0.000	0.022
1000	500	.0	3.00	1.00	.010	.051	0.000	-0.041
1000	1500	.0	3.00	1.00	.094	.051	0.000	0.043
1000	250	.0	3.00	1.00	.000	.051	0.000	-0.051
1000	1750	.0	3.00	1.00	.112	.047	0.000	0.065
1000	1000	.2	3.00	1.00	.008	.008	1.000	0.000
1000	750	.2	3.00	1.00	.014	.007	0.000	0.007
1000	1250	.2	3.00	1.00	.004	.006	0.003	-0.002
1000	500	.2	3.00	1.00	.047	.012	0.000	0.035
1000	1500	.2	3.00	1.00	.002	.006	0.000	-0.004
1000	250	.2	3.00	1.00	.271	.023	0.000	0.248
1000	1750	.2	3.00	1.00	.001	.004	0.000	-0.003
1000	1000	.5	3.00	1.00	.000	.000	1.000	0.000
1000	750	.5	3.00	1.00	.000	.000	1.000	0.000

1000	1250	.5	3.00	1.00	.000	.000	1.000	0.000
1000	500	.5	3.00	1.00	.000	.000	1.000	0.000
1000	1500	.5	3.00	1.00	.000	.000	1.000	0.000
1000	250	.5	3.00	1.00	.000	.000	1.000	0.000
1000	1750	.5	3.00	1.00	.000	.000	1.000	0.000
1000	1000	.8	3.00	1.00	.000	.000	1.000	0.000
1000	750	.8	3.00	1.00	.000	.000	1.000	0.000
1000	1250	.8	3.00	1.00	.000	.000	1.000	0.000
1000	500	.8	3.00	1.00	.000	.000	1.000	0.000
1000	1500	.8	3.00	1.00	.000	.000	1.000	0.000
1000	250	.8	3.00	1.00	.000	.000	1.000	0.000
1000	1750	.8	3.00	1.00	.000	.000	1.000	0.000

See Table 3 for explanation on column names

Table B4

Conditions with variance homogeneity and equal sample sizes

N_1	N_2	ES	σ_1	σ_2	perm	t-test	p-value	dif
10	10	0.0	1.00	1.00	0.049	0.051	1.000	-0.001
10	10	0.2	1.00	1.00	0.935	0.933	0.493	0.002
10	10	0.5	1.00	1.00	0.826	0.820	0.000	0.006
10	10	0.8	1.00	1.00	0.619	0.612	0.000	0.007
60	60	0.0	1.00	1.00	0.050	0.050	1.000	-0.000
60	60	0.2	1.00	1.00	0.808	0.806	0.345	0.001
60	60	0.5	1.00	1.00	0.226	0.224	0.345	0.001
60	60	0.8	1.00	1.00	0.010	0.009	1.000	0.000
1000	1000	0.0	1.00	1.00	0.049	0.049	1.000	-0.000
1000	1000	0.2	1.00	1.00	0.006	0.006	1.000	-0.000
1000	1000	0.5	1.00	1.00	0.000	0.000	1.000	-0.000
1000	1000	0.8	1.00	1.00	0.000	0.000	1.000	-0.000

See Table 3 for explanation on column names

Appendix C

All figures used in the data analysis can be found on

https://github.com/rushkock/sim_study_thesis/tree/master/reports/figures.

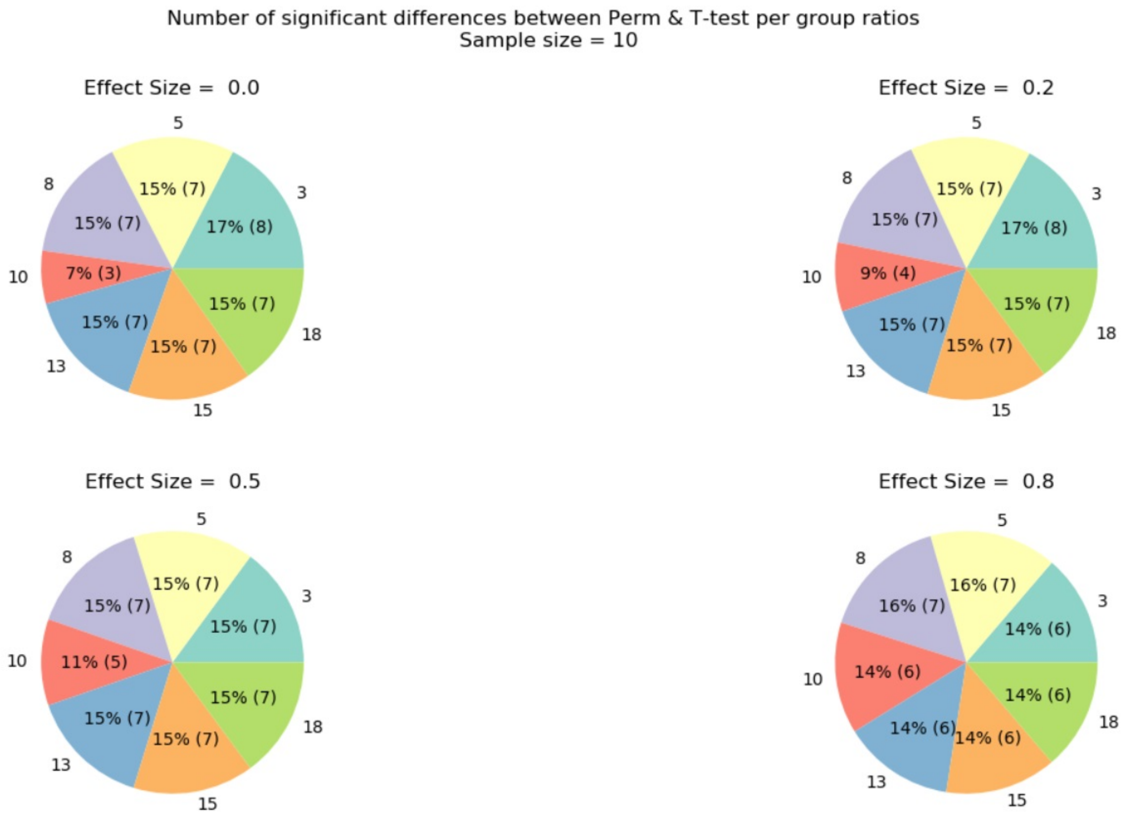

Figure C1. The number of significant differences between the permutation test and t-test for each group ratio.

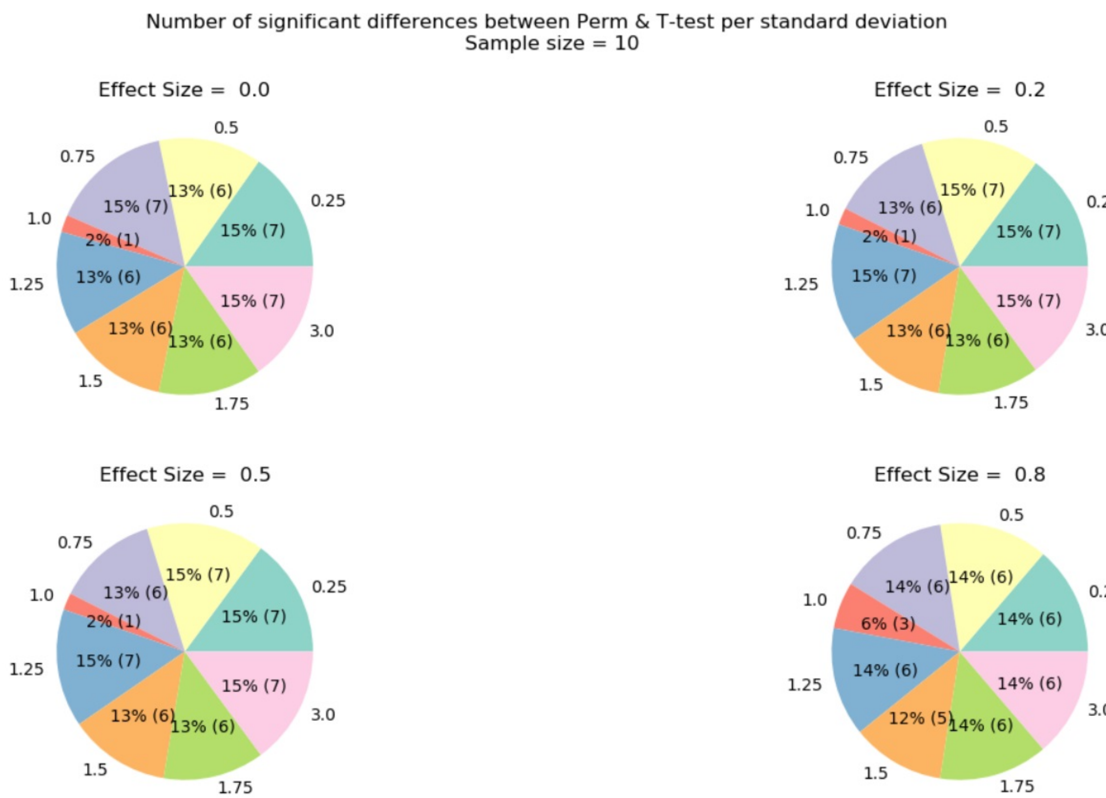

Figure C2. The number of significant differences between the permutation test and *t*-test for each standard deviation.

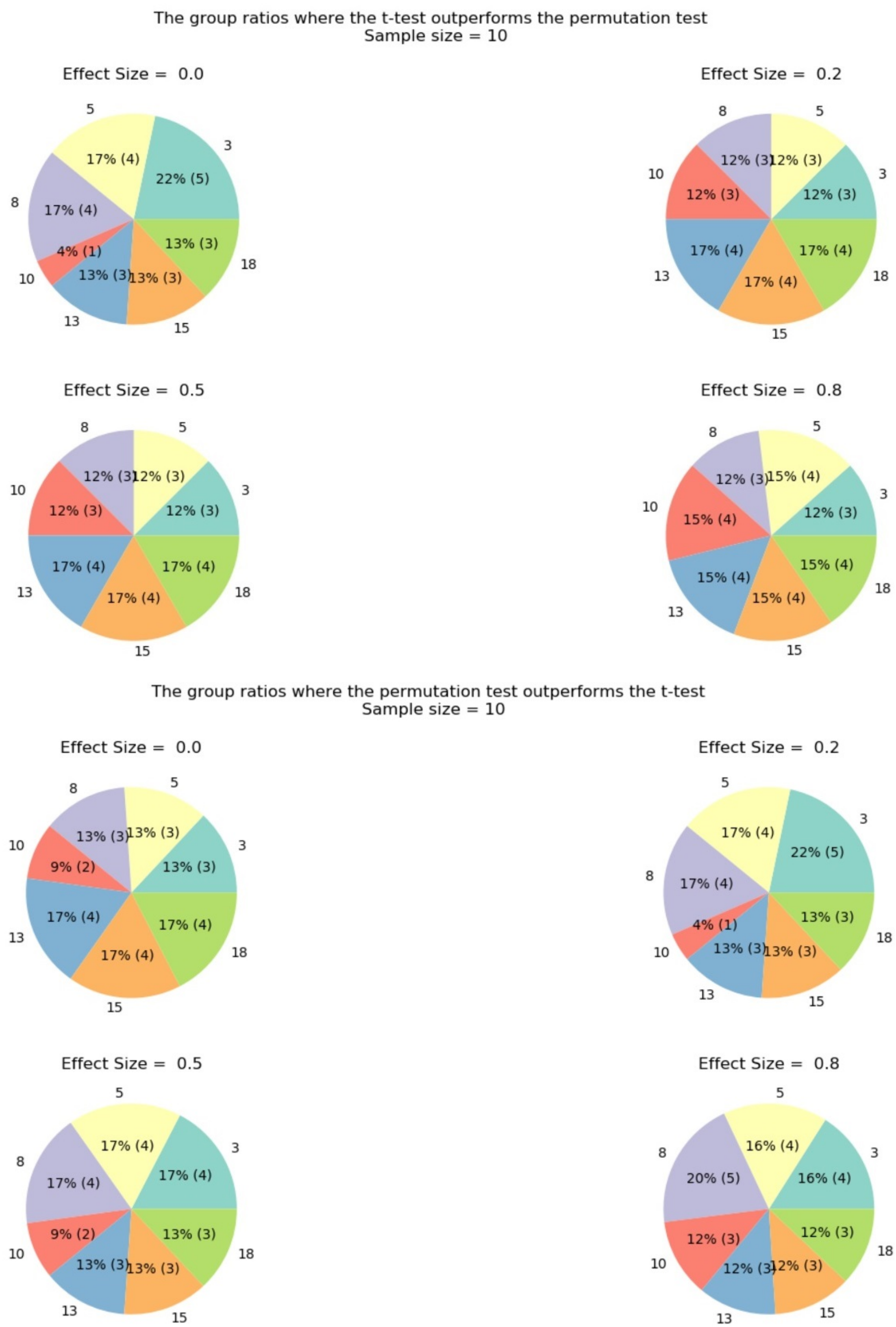

Figure C3. The number of significant differences for the permutation test and *t*-test for each group size.

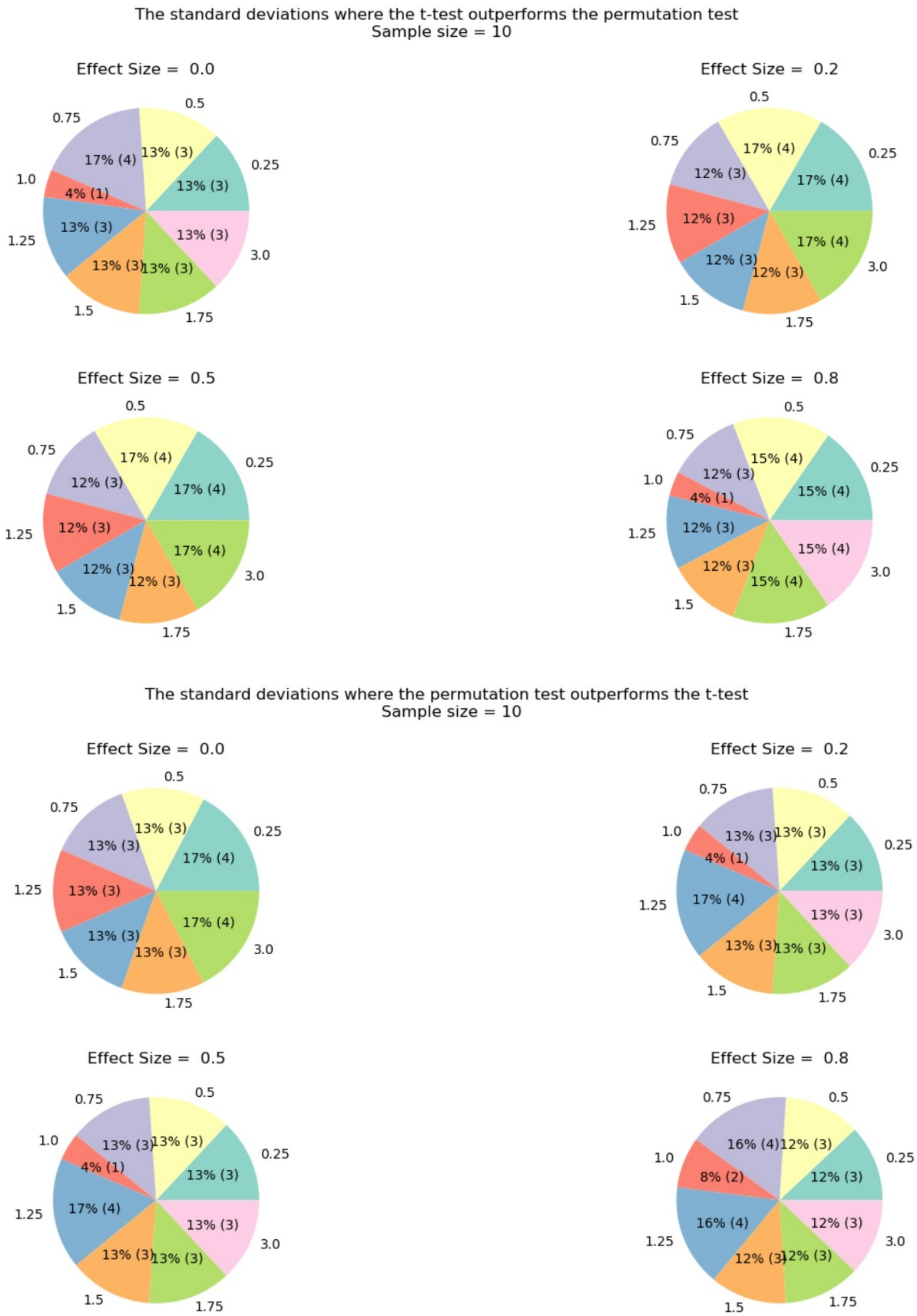

Figure C4. The number of significant differences for the permutation test and *t*-test for each standard deviation.

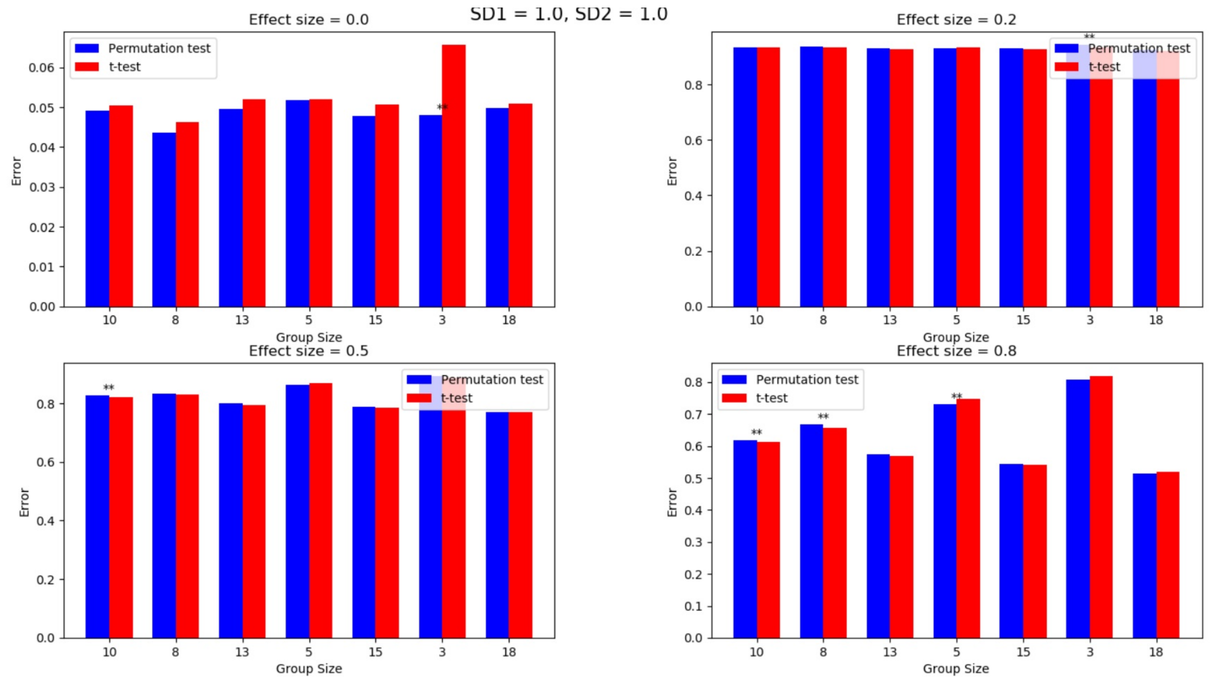

Figure C5. Significant differences between the two tests for the sample size of group 1 = 10 and its deviations visible on the x-axis

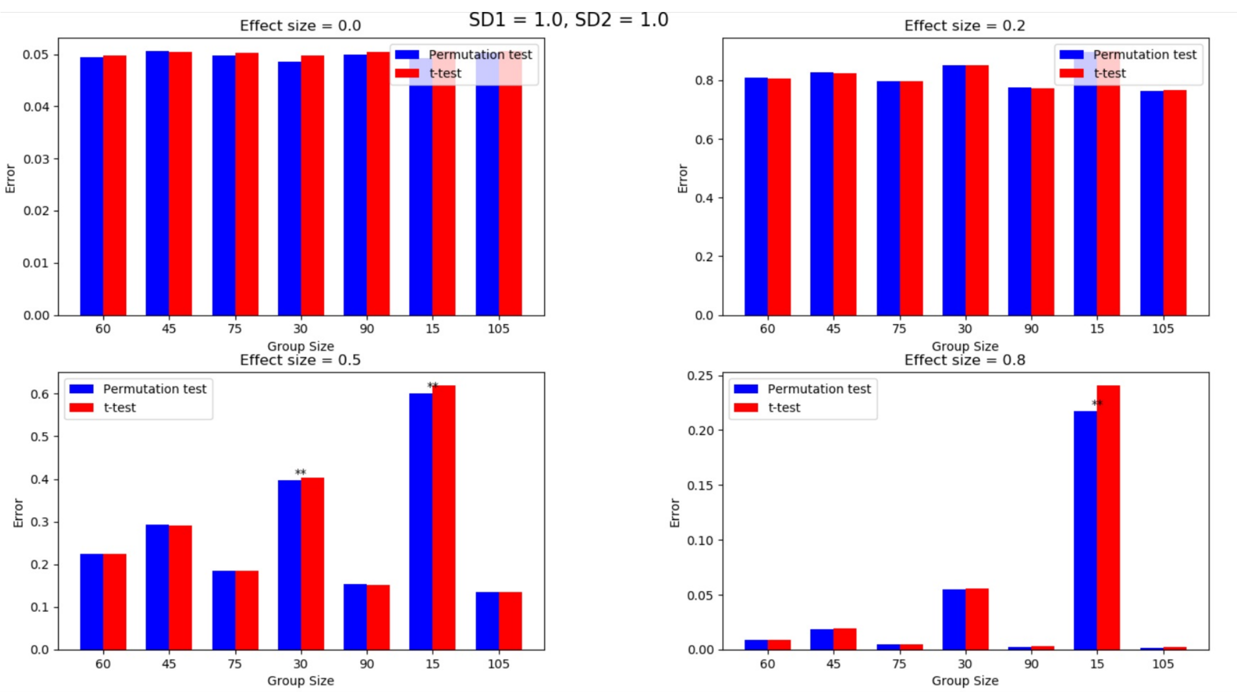

Figure C6. Significant differences between the two tests for the sample size of group 1 = 60 and its deviations visible on the x-axis

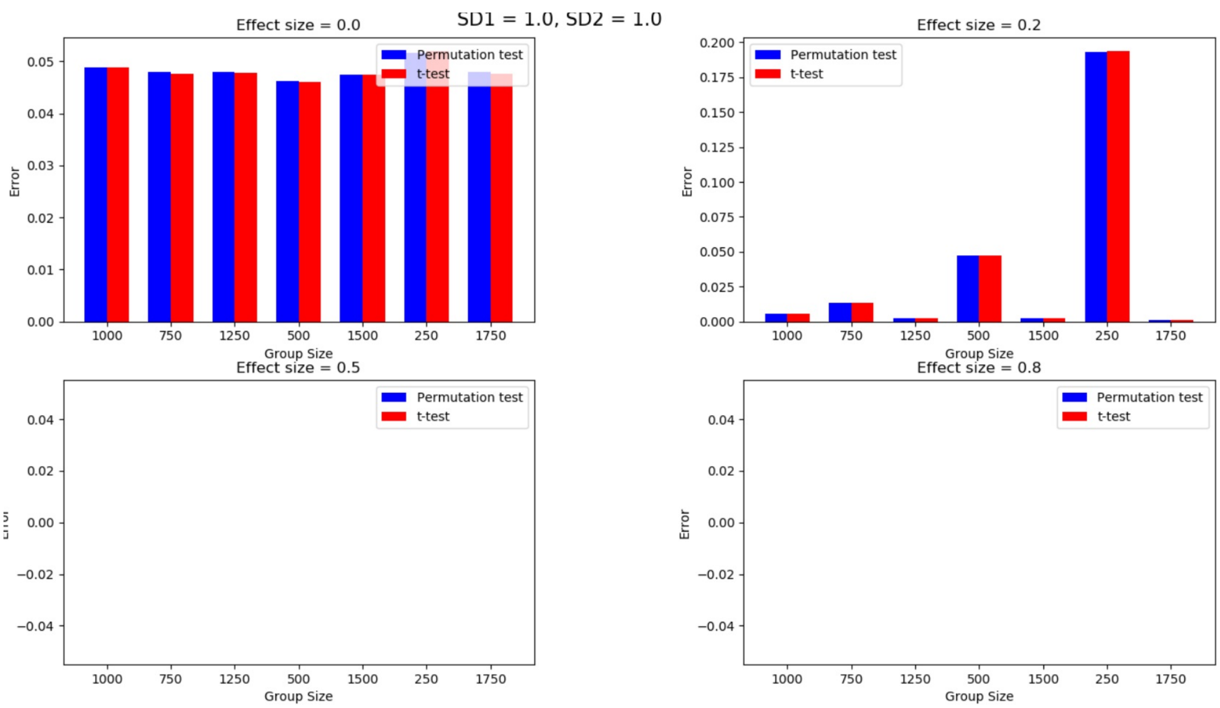

Figure C7. Significant differences between the two tests for the sample size of group 1 = 1000 and its deviations visible on the x-axis