Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences

Nonparametric Tests for Homoscedasticity in Randomized Complete Block Designs.

A thesis presented by

Pamela Lizeth Torres Núñez

Submitted to the
School of Engineering and Sciences
in partial fulfillment of the requirements for the degree of

Master of Science

in Manufacturing Systems

Monterrey, Nuevo León, May 31st, 2019

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences

The committee members, hereby, certify that have read the dissertation presented by Pamela Lizeth Torres Núñez and that it is fully adequate in scope and quality as a partial requirement for the degree of Master of Science in Manufacturing Systems.

Dr. Víctor Gustavo Tercero Gómez
Tecnologico de Monterrey
School of Engineering and Sciences
Principal Advisor

Dr. William Jay Conover
Texas Tech Univeristy
Co-Advisor

Dr. Alvaro Eduardo Cordero Franco
Universidad Autónoma de Nuevo León
Committee Member

Dr. Luis Alejandro Benavides Vázquez
Tecnologico de Monterrey
Committee Member

Dr. Rúben Morales Menéndez
School of Engineering and Sciences
Dean of Graduate Studies

Tecnológico de Monterrey
20 MAY 2019
Dirección Nacional de Posgrado
Dr. Rubén Morales Menéndez

Monterrey, Nuevo León, May 31th, 2019

Declaration of Authorship

I, Pamela Lizeth Torres Núñez, declare that this thesis titled "Nonparametric Tests for Homoscedasticity in Randomized Complete Block Designs" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this dissertation is entirely my own work.
- I have acknowledged all main sources of help.
- Where the dissertation is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Pamela Lizeth Torres Núñez
Monterrey, Nuevo León, May 31^{st}. 2019

Dedication

I want to dedicate this work to my parents Mr. Benjamin and Mrs. Josefina, that always have been my force to go ahead in each step that I move on, they are my model, they teach me that every effort will be recognized but, first I need to do all that I can do, and when I cannot, the next thing is looking for the inspiration that I need and continue with the goal that put me there in the first instance. From them I learned the humility to accept when I need help, and recognize that I was wrong.

To my siblings, Benjamin, Norma, Jessica, Carlos and Jorge, they have been an excellent model to me that inspires me.

To my nieces, Salma, Alexa, Sofia, and Mabel, to demonstrate them that all the goals can be possible if they put all the discipline, attitude and if they believe in themselves.

Finally, to these people that have a goal that they know that the only limit to reach them are themselves.

Acknowledgments

Another goal is to achieve in my life, but, this could not be possible if I did not have my support network. For this reason, I would like to express all my gratitude to all the people that always have been on my side in this way.

Thanks to my parents for all their love, patience, and preoccupation that always received from them.

To my siblings, Benjamin, Norma, Jessica, Carlos, and Jorge, for always give me them support, and the confidence to know that if the things do not give me the result that I expect, I will always have my family there to share all my achievements but, also the failures that I could have on the way. For still believing in me.

To my advisers for their guide, for demanding me more of what I thought I could give, for their help in all my doubts.

To my friends whose always demonstrated me that they were here for me if I need it, that shared with me all these long nights of work, stress, laughs and great moments that I will save on my memory and heart.

A special acknowledgment to the Tecnologico de Monterrey and CONACyT for allow me to be part of this program, for all the sources provided and the supports on tuition and living place.

To God, to offer me all that I required to follow my goals, to give me the beautiful people that I have on my side.

Thanks.

Nonparametric Tests for Homoscedasticity in Randomized Complete Block Designs

By

Pamela Lizeth Torres Núñez

Abstract

Variance assessment is a key component in robust design, process improvement, and reliability analysis, among other practical venues. However, most statistical development in nonparametric statistics has been focused on the problem of location changes, whereas the development of tests for homoscedasticity has been scarce, limited, in most cases, to one-factor analysis. By considering a blocking element, in the sense of Friedman, more power can be obtained by extending the approach to be used with linear rank transformations sensitive to scale changes. In this research, a total of 96 different new linear rank tests for homoscedasticity have been created, and their robustness and power evaluated through extensive Monte Carlo simulations. 36 of these tests showed to be either distribution robust or distribution free. 5 approaches of within the remaining tests acted consistently with high power over scale changes, and only 3 (Fligner-Killeen, squared ranks, and Talwar-Gentle, the three are aligned with the overall median) of these tests remained powerful when dealing with scale and location changes. Based on their performance, and easy of use, practitioners and researchers might find the results and recommendations of this work compelling and useful for their practice of data analysis when dealing with nuisance factors in the form of blocks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Research areas in manufacturing are involved in the execution of experiments in a specific process or product to discover or validate the behavior of that process or product when natural or artificial variations are included. As the majority of these variations are controlled (artificial changes), the observed variables can be tested (Montgomery, 2008). In the manufacturing industry, the analysis of these variations is conducted within the framework of *Design and Analysis of Experiments* (DOE), where the study of differences is used as a base to develop robust products and processes, as well as process capability and reliability improvements. Traditionally, this analysis is carried following parametric approaches where the robustness of the ANOVA F test to non-normal data had analyzed location changes practical and reliable. However, tests for scale changes are more sensitive to the normality assumption, and practitioners have to rely more on different statistical approaches, such as the use of nonparametric methods. However, the development of nonparametric procedures to deal with changes in scale has been scarce, and limited in most cases to one factor analysis. A gap in data analysis to address these issues, as is the case of treatment comparisons when nuisance factors are involved in the form of a block, requires further development. This research, is an attempt to fill this gap.

A major intention of applying a DOE in a process is to improve its performance, i.e., to reduce cycle times, cost, defects and variability. Through treatment analysis, experiments are conducted to demonstrate a cause-effect relation between one or more explanatory variable (process inputs), and a response variable (a quality characteristic). However, the presence of nuisance factors might increase the experimental error, intensifying the difficulty of the analysis and sometimes reducing the power to detect treatment changes. To deal with nuisance factors, researchers had created designs like Randomized Complete Block Design (RCBD), where nuisance factors are controlled in the form of blocks. This reduces the experimental error, making the analysis suitable to detect subtle changes between treatments, with high power. Also, as blocks are designed to be heterogeneous between themselves, practitioners gain the ability to generalize their conclusions, as their discoveries hold for the plurality of conditions defined by each block analyzed (see Kutner et al., 2005, for details).

To determine if a variation is acceptable or not, acceptance interval are determined, and tools and equipment are developed to perform a test to a process. Several tests are used to measure the variability, one of the most popular between statisticians is the analysis of variance, which allows practitioners to determine if differences exist (or not) between two or more samples by comparing their means using different treatments. (Angel Gutierrez, 1996).

This method requires few assumptions: (a) The error variables are mutually independent within each block. (b) The error terms are normally distributed. (c) The error terms have constant variance (Kutner et al., 2005). In this research, we are dealing with the problem when the last two assumptions are not accomplished.

First, the assumption that the data come from a normal distribution facilitates the selection of statistical tools such as ANOVA for the processing of the information. This generates errors of the variation in the distributions, comparing them with the normal ones. There are cases in which it is difficult to obtain the distribution, in these situations nonparametric statistical methods are used, that perform the analysis of data in the variables whose distribution is not specific. Nonparametric procedures are applicable in many cases where normal theory procedures cannot be utilized. The nonparametric methods require just the ranks of the observations (Hollander et al., 2013). The disadvantage of these methods is their inaccuracy.

There are several critical factors in the implementation of sampling to properly analyze data; one of them is the distribution function (Conover, 1999). Another factor is the sample size due to the restrictions of many tests to effectively work with a reliable sample science study.

Second, the assumption of constant variance, different processes, and products made in a company may seem the same, but in reality, they are not. Variations in quality characteristics can be very significant. Variables such as time, resistance and weight determine acceptable processes or products when a specification interval is accomplished. The quality of a product is measured by its variation concerning the desired and depends on the existing variation in quality characteristics (Angel Gutierrez, 1996).

Therefore, this study focuses on addressing these problems by developing and evaluating the equality of variances using nonparametric methods when the distribution is unknown. To evaluate the performance several distributions were used. These are normal, double-exponential (Laplace), in the case of the symmetric distributions; the distributions normal-square (chi-squared with one degree of freedom), Double-exponential-square, log-normal, Gamma and Weibull for the skewed distributions.

Each distribution was evaluated over different scenarios that are generated by changing the sample size (2 and 4), the standard deviation, the effect of the treatment, the number of replicates per cell, and the size of the block (2, 4, and 8).

One more variation considered was the statistic to test the hypothesis, four statistics were used over different linear rank transformation as a way to create new statistics; two considering the χ^2 distribution, one of them is which is used in Friedman's test (Conover, 1999), the other from the test of Kruskal Wallis (Hollander et al., 2013) and two for the Snedecor's F

1.1 Motivation

In a manufacturing process of a company, there are three different processes for cable coating that are carried out in three shifts per day. You want to know if there is a difference in the variability of the process. To address this type of situations, we need to use a particular kind of designs as Randomized Complete Block Design (RCBD), in which the shift is a nuisance factor that can be treated as a block to reduce the noise and improve the power; the treatments are the types of coating. This type of applications is what motivates this study, to establish methods where the presence of blocks is detected, and we want to know if there is a significant change in variability between treatments when nuisance factors in the form of block exist.

1.2 Problem statement

This research focuses on the problem of the development nonparametric statistical methods to study the effect of treatments of an experimental study on the variance of a variable of interest when dealing with completely randomized experimental designs with replications. The principal restrictions in these cases are the nonnormality of the residuals.

1.3 Research questions

1. Which tests statistics are distribution robust?
2. Which tests for homoscedasticity are considered robust when there is also a change in the mean?
3. Which tests are considered sensitive for both location and scale changes?
4. In regard to the power within robust test for location:
   (a) Which test provides the most overall power?
   (b) Which test has the biggest power when the distribution is symmetric?
   (c) Which test has the biggest power when the distribution is asymmetric?
   (d) Which test has more power when dealing with the different number of replicates?
   (e) Which test has more power when dealing with relatively different numbers of treatments?
   (f) Which test has more power when dealing with relatively different numbers of blocks?
5. In regard to the power within tests sensitive to scale and location changes:
   (a) Which test provides the most overall power?
   (b) Which test has the biggest power when the distribution is symmetric?

(c) Which test has the biggest power when the distribution is asymmetric?

(d) Which test has more power when dealing with the different number of replicates?

(e) Which test has more power when dealing with relatively different numbers of treatments?

(f) Which test has more power when dealing with relatively different numbers of blocks?

1.4 Research hypotheses

1. At least one test is distribution robust or distribution free, where a test is considered robust if the probability of a type 1 error of 0.05 is not bigger than 0.1.

2. At least one test is robust to changes in location, where a test is considered robust if the probability of a type 1 error of 0.05 is not bigger than 0.1.

3. At least one test is sensitive for both location and scale changes.

4. In regard of the power within robust test for location:

   (a) At least one test has the most overall power.

   (b) At least one test has the most power when the distribution is symmetric.

   (c) At least one test has the most power when the distribution is asymmetric.

   (d) At least one test has the most more power when dealing with different number of replicates.

   (e) At least one test has the most power when dealing with relatively different numbers of treatments.

   (f) At least one test has the most power when dealing with relatively different numbers of blocks.

5. In regard of the power within tests sensitive to scale and location changes:

(a) At least one test has the most overall power.

(b) At least one test has the most power when the distribution is symmetric.

(c) At least one test has the most power when the distribution is asymmetric.

(d) At least one test has the most more power when dealing with different number of replicates.

(e) At least one test has the most power when dealing with relatively different numbers of treatments.

(f) At least one test has the most power when dealing with relatively different numbers of blocks.

1.5 Scope and limitations

The scope of this research is limited to the analysis of linear rank statistics. As analytical solutions are hard to obtain (no known analytical solution is known for limited sample behaviour) the performance of each statistics is limited to Monte Carlo simulations over the following combination factors:

- The number of replicates per cell could be 2,5,10,20,50, and 100.
- The distributions selected were normal, normal-squared, double exponential, double-exponential squared, lognormal, gamma and Weibull.
- The number of treatments compared were 2 and 4.
- The block size could be 2,4, or 8.
- The effect of the treatment varies (this will be shown later).
- Change in standard deviation (this will be shown later).

Due to a large number of scenario analyzed, the number of simulations was limited to 1000.

1.6 Main contributions

New nonparametric statistical tests for homoscedasticity in RCBD were created that are capable of dealing with symmetrical and skewed distributions with competitive power. Performance of each test was analyzed with a different distribution, sample size, replicate size, and block size.

Chapter 2

Background and literature review

2.1 Background

2.1.1 Design and analysis of experiments

An **experiment** can be defined as a test where a practitioner makes deliberate changes into a variable of a process or product to observe and analyze the change reasons in its response. The experimentation is a key factor for the development of new knowledge, processes, and products. One of the main objectives when designing a product or process is to develop a robustness, that is, a product or process that is less affected by external sources of variation. In general, the objectives of an experiment could be:

(a) To determine the variables that have a significant influence on the response.

(b) To determine the necessary levels of a independent variables that could keep the response in a nominal value.

(c) To determine how to modify a set of independent variables to reduce the variability of a response.

(d) To determine a level of the independent variables that could potentially reduce the effect of external and uncontrolled variables.

There are three basic principles in the design of an experiment:

1. *Replicates*. A replicate or an *experimental unit*, are the times that an experiment is carried out under identical conditions. The replicates have

essential properties. First, it makes possible the calculation of the experimental error; which is used to determine if the observed differences are or aren't statistically significant. Second, the use of replicates allows the experimenter to reduce the uncertainty of an estimation.

2. *Randomization.* By randomization, it is understood that the assignment of a treatment to an experimental unit was determined randomly. The use of randomization is a requirement for many statistical procedures. Randomization helps to reduce the minimum the influence of extraneous factors that are not under control. Also, it provides a solid ground when making inferences about cause and effect relationships Kutner et al. (2005).

3. *Blocking.* This technique is used to improve the precision between the variables (factors) of interest. It reduces the variability occasioned by the nuisance factors. The nuisance factors are those who can influence the response, but there is not a specific interest. In this experiment, the experimental material is grouped into homogeneous subgroups called blocks, and separate tests are conducted in each block. Ideal blocks are constructed so that the experimental units are homogeneous within the blocks, but heterogeneous between them. Other benefits of this technique is that the validity of the conclusions obtained is increase with this method. Randomization within blocks provides additional protection against the unknown cause of variability.

Hypothesis test

In statistics, procedures used to make inference of a population from the analysis of data from an experiment are called **hypothesis test** and **confidence intervals**. A statistical hypothesis is an affirmation that is made about the parameters of a distribution. This reflects some supposition about a problem or a situation. For example, if the situation is to prove if two groups of students A and B have the same average grades, this hypothesis can be enunciated as

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

$$H_1 : \mu_1 \neq \mu_2.$$

Table 2.1: An experiment with one factor.

Treatment	**Observations**				**Total**	$\bar{y}_i$
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ddots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a.}$	$\bar{y}_{a.}$
					$y_{..}$	$\bar{y}_{..}$

The $H_0 : \mu_1 = \mu_2$ is known as the *null hypothesis*; in this case, implies that the averages of the group A and B are equal.The $H_1 : \mu_1 \neq \mu_2$ is called the *alternative hypothesis*. To prove a hypothesis, a random sample is collected, then compute a statistic test to reject or not the null hypothesis H_0. There is a set of values of the test statistic that rejects H_0; this set of values is called *rejection region* of the test. There are two error types the might be incurred when applying a hypothesis test:

- *Type I error*. Occurs when the null hypothesis is rejected and the null was true. Here $\alpha = P(\text{reject } H_0 | H_0 \text{ is true})$

- *Type II error*. Occurs when the null hypothesis was not rejected and the null was false. Here $\beta = P(\text{reject } H_0 | H_0 \text{ is false})$ is the probability not to reject the null hypothesis when it is false.

To make a decision whether the null has to be rejected of not is to use the p - value approach. The p-value is minimum significant level to reject reject the null hypothesis with the provided data. The null hypothesis is rejected if the p-value is less than a specific value of α, otherwise it is not rejected.

2.1.2 Analysis of variance: one way

There are experiments that compare more than two treatments in one factor; the procedure to test the equality of means is the Analysis of Variance. This analysis has a wide range of applications in statistics.

Table 2.1 shows the structure of the data in a one factor experiment. In this case, there are a different treatments that will compared. The response variable of each treatment is a random variable.

Models for the data A way to describe the observations of an experiment is with the use of models. There are two models can be used: (1) cells mean model, and (2) factors effect model.

Cells mean model This model can be used when the data comes from observational or experimental studies and is based on completely randomized design.

$$Y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n \end{cases} , \tag{2.1}$$

where Y_{ij} is the response variable in the j-th trial for the i-th factor level or treatment. μ_i are parameters, ε_{ij} are the independent errors $N(0,\sigma^2)$.

Factor Effects Model This model is expressed in terms of the factor effects, and it is an alternative formulation of model (2.1)

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n. \end{cases} , \tag{2.2}$$

where μ is a constant component common to all observation (for example the overall mean of the observations), τ_i is the effect of the ith factor level, ε_{ij} are the independent errors $N(0, \sigma^2)$.

2.1.3 Analysis with fixed effects model

For the analysis of variance (ANOVA) with one factor, the fixed effects model will be utilized. $y_{i.}$ represents the total of observations in the ith treatment. Let $\bar{y}_{i.}$ be the average of the observations of the ith treatment. $y_{..}$ is the grand total of the observations and $\bar{y}..$ represents the grand average of the observations. The expressions for the model are:

$$y_{i.} = \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{i.} = \frac{y_{i}.}{n} \qquad i = 1, 2, ..., a \tag{2.3}$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{..} = \frac{y_{..}}{N}. \tag{2.4}$$

Where $N = an$ it is the total number of observations. The hypothesis are $H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$ versus $H_1 : \mu_i \neq \mu_j$ at least one pair (i, j).

Sum of Squares

The sum of squares are needed to develop the test, the *total sum of squares* (TSS) in equation (2.5) is the sum of the squared differences between the subgroup averages of the observed treatments and the grand average. The sum of squares due to the treatments is shown in equation (2.6). Equation (2.7) shows the sum of squares due to the experimental error. There are $N = an$ observations. Thus, TSS has $N - 1$ degrees of freedom. There are a treatments, SST has $a - 1$ degrees of freedom. Finally, the SSE has $n - 1$ degrees of freedom, where n represents the number of replicates.

$$\text{TSS} = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2, \tag{2.5}$$

$$\text{SST} = n \sum_{i=1}^{a} (\bar{y_{i.}} - \bar{y}_{..})^2, \tag{2.6}$$

$$\text{SSE} = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2. \tag{2.7}$$

Mean Squares

Table 2.2 is called the *analysis of variance table*, in general, these are the steps to develop an ANOVA, the sum of squares are computed with the corresponding equations in this case the equations are (2.5), (2.6), (2.7), for the TSS, SST, and SSE respectively. The degrees of freedom are also computed for each of the sums of squares. To estimate the mean squares is needed to obtain the ratio of each sum of squares divided for its respective degrees of freedom. Once that the mean squares have been calculated, the ratio of the mean squares of the treatments divided by and the Mean squares of the error. Give the F_0 value which will be used to compare it with the statistic to reject or not the null hypothesis. Also can be used to compute the p-value. The ratio is distributed as a F with $(a - 1)$ and $(N - a)$ degrees

Table 2.2: ANOVA in fixed effects model - one factor.

Source	**Degrees of freedom**	**Sum of Squares**	**Mean Square**	$\mathbf{F_0}$
Treatment	$a-1$	SST	$\text{MST} = \frac{\text{SST}}{a-1}$	$F_0 = \frac{\text{MST}}{\text{MSE}}$
Error	$N-a$	SSE	$\text{MSE} = \frac{SSE}{n-1}$	
Total	$N-1$	TSS		

of freedom. Reject H_0 if $F > F_{\alpha,a-1,N-a}$ otherwise do not reject the null hypothesis of the equality of treatments

2.1.4 Model Adequacy

If the ANOVA model is considered for an application where we want to determine if there are or not differences between the means of the treatments, we need to be sure that the model is appropriate for that application. For this, some assumptions need to be satisfied. These assumptions are that the model used, equation (2.1) or (2.2), describes the observations properly, and the errors follow a normal distribution with mean zero and σ^2, where σ^2 is constant but unknown. If these assumptions are satisfied, the ANOVA is appropriate to test the hypothesis that the means of the treatments are equal. Though, in practice, it is common that these assumptions are violated. For this reason, it is convenient to verify the adequacy of the model for the data before inferences based on the model are undertaken. The aptness of the model can be proved with an analysis of the residuals. The residual of the j-th observation in the i-th treatment, and it is defined as

$$e_{ij} = y_{ij} - \hat{y}_{ij}, \tag{2.8}$$

where $\hat{y}_{ij}$ is an estimation of the observation $\hat{y}_{ij}$ and is obtained as

$$\hat{y}_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}). \tag{2.9}$$

The analysis of residuals is required before conducting an ANOVA. If the residuals do not have a clear structure; it means that there is not a major concern, and it implies that the adequacy of the model is accomplished. A graphic analysis of the residual result is helpful to prove the aptness of the model.

Normality assumption

For the analysis of this assumption, normal probability plot residuals can be constructed. If the distribution of the residuals is normal, the plot will show a straight line, whereas a plot that departs substantially from linearity suggests that the distribution of the errors is not normal. Normality plots can also be used to detect outliers. An outlier is a residual that is bigger than any other. These outliers can introduce distortion in the analysis. Sometimes, the cause of an outlier is an measurement error; if this is the case, the observation can be deleted and the practitioner continue with the analysis without this observation.

Independence assumption

Plot the residuals in the collection order helps to detect if there are any correlation between the residuals and time. If a positive trend is shown, it might suggest a positive correlation. This means that the independence assumption has been violated. The key to avoid this is to follow an adequate randomization process.

Constant variance assumption

How was mentioned before, the residuals should not present a structure or a pattern, plot the residuals against the predictor variable is helpful to examine if the variance of the error terms is constant. Sometimes the constant variance of the observations is increasing when the observation increases. If this situation is present, the residuals the plot would show the data shaped like a megaphone. This could be occasioned because the data does not follow a normal distribution. This assumption is called *homoscedasticity*, that means, variance homogeneity (equality); if this assumption is violated, the type I error could be higher than the expected.There are statistics test that has been developed to test the equality of variances; these tests will be studied later.

2.1.5 Randomized Complete Block Design

When an experimenter is conducted, the variability induced by a perturbing factor may cause that the response variable is affected. Sometimes, this factor is the one whose existent is not known and cannot be controlled. In other

Table 2.3: Randomized complete block design Conover (1999).

	Treatment			
Block	1	2	$\cdots$	a
1	X_{11}	X_{21}	$\cdots$	X_{a1}
2	X_{12}	X_{22}	$\cdots$	X_{a2}
3	X_{13}	X_{23}	$\cdots$	X_{a3}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
b	X_{1b}	X_{2b}	$\cdots$	X_{ab}

cases, its existence can be known, but is cannot be controlled. When this kind of factor is known and can be controlled, a technique called *blocking* can be used. This technique is used to eliminate the effect caused by this nuisance factor when performing treatment comparisons. The principal objective of blocking is to make the experimental error as small as possible; to do this each block need to contain all the treatments; this is called complete. This design is known as a *randomized complete block design* (RCBD). This is a widely used design in practice. Suppose that you have a treatments that are going to be compared and b blocks. The RCBD is shown in Table 2.3. There is an observation per treatment in each block; the run order is random within each block.

Statistical Model

The statistical model for the RCBD is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, ...a \\ j = 1, 2, ...b \end{cases} \tag{2.10}$$

where μ is the overall mean, τ_i is the effect of the ith treatment, β_j is the effect of the jth block and ε_{ij} is the error term $N(0, \sigma^2)$.

In this model the effects of the treatments and blocks are considered as fixed. These effects are considered deviations from the overall mean, for this reason:

$$\sum_{i=1}^{a} \tau_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0 \tag{2.11}$$

In practice, it is usually desired to evaluate whether treatment means are equal or not, and the corresponding null and alternative hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$$
$$H_1 : \text{Not all } \mu_i \text{ are equal.}$$

another equivalent hypothesis is to test in terms of the effect of the treatment, the hypotheses are

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \tau_i \neq 0 \text{ at least for one i.}$$

The usual notation used when making calculations is

$$y_{i.} = \sum_{j=1}^{b} y_{ij} \qquad i = 1, 2, ..., a \tag{2.12}$$

$$y_{.j} = \sum_{i=1}^{a} y_{ij} \qquad j = 1, 2, ..., b \tag{2.13}$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1} y_{ij} = \sum_{i=1}^{a} y_{i.} = \sum_{j=1} y_{.j} \tag{2.14}$$

where $y_{i.}$ are the observations in the ith treatment, $y_{.j}$ are the observations in the jth block and $N = an$ it is the total number of observations.

In a similar way $\bar{y}_{i.}$, $\bar{y}_{.j}$, are the averages of the observations in the treatment and block respectively and $\bar{y}_{..}$ is the grand average. Expressed algebraically

$$\bar{y}_{i.} = \frac{y_{i.}}{b}, \quad \bar{y}_{.j} = \frac{y_{.j}}{b}, \quad \bar{y}_{..} = \frac{y_{..}}{N}. \tag{2.15}$$

The sum of squares are needed to develop the test, the *total sum of squares* (TSS) in equation (2.16) is the sum of the squared differences between the subgroup averages of the observed treatments and the grand average. The sum of squares due to the treatments is shown in equation (2.17).The equation (2.18) shows the sum of the squares due to the block. Equation (2.18) shows the sum of squares due to the experimental error. There are N= an

Table 2.4: ANOVA in a randomized complete block design (see Montgomery).

Source	**Sum of squares**	**Degrees of freedom**	**Mean squares**	$\mathbf{F_0}$
Treatments	SST	$a-1$	SST$/(a-1)$	MST/MSE
Blocks	SSB	$b-1$	SSB$/(b-1)$	
Error	SSE	$(a-1)(b-1)$	SSE$/((a-1)(b-1))$	
Total	TSS	$N-1$		

observations. Thus, TSS (2.16) has $N-1$ degrees of freedom. There are a treatments and b blocks, SST and SSB have $a-1$, and $b-1$ degrees of freedom respectively. Finally, the SSE has $(a-1)(b-1)$ degrees of freedom.

$$\text{TSS} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N}, \tag{2.16}$$

$$\text{SST} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}, \tag{2.17}$$

$$\text{SSB} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N}, \tag{2.18}$$

$$\text{SSE} = TSS - SST - SSB. \tag{2.19}$$

This corresponding ANOVA table is shown in Table 2.4

2.2 Literature Review

2.2.1 Previous works

Currently one of the key factors for the success of an industry is to use all the available tools, knowledge and experience that can contribute to the improvement of the company. The DOE is one of the tool for the optimization processes; this method can be defined as the realization of a set of tests for making changes to the variables under control of a process with the objective to observe and identify the reasons for the changes in the study and quantify it. The experiments are more complicated day by day because there are many factors that are subject to control and affect the products and/or processes, hence there are many combinations of these factors that should be tested to obtain valid results.

Several tests have been proposed for the homogeneity of variances, the most frequently used are methods proposed by Bartlett (1937) and Box (1953). In this section, we describe these and other methods to test the homogeneity of variances that have been developed until now.

The hypotheses to test we are interested in is

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2 \tag{2.20}$$

$$H_1 : \text{at least one } \sigma_i^2 \text{ is different.} \tag{2.21}$$

One of the most common tests to evaluate these hypotheses is the Bartlett test. This method includes a statistic whose distribution approximates a χ^2_{a-1} with $a-1$ degrees of freedom when the a are random samples from a normal population (Montgomery, 2008). The test statistic is

$$\chi_0^2 = 2.3026\frac{q}{c}, \tag{2.22}$$

where

$$q = (N-a)log_{10}S_p^2 - \sum_{i=1}^{a}(n_i-1)log_{10}S_i^2,$$

$$c = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a}(n_i-1)^{-1} - (N-a)^{-1}\right),$$

$$S_p^2 = \frac{\sum_{i=1}^{a}(n_i-1)S_i^2}{N-a}.$$

S_i^2 is the sample variance. The null hypothesis is rejected when $\chi_0^2 < \chi^2_{\alpha,\alpha-1}$. This test is sensitive to the normality assumption, for this reason is not recommended its use if normality is not sustained.

Since the Bartlett test is sensitive to normality, other test have been developed. For instance, we have the modified Levene test known as the Brown-Forsythe test. This test is robust to deviations from normality. This test uses the absolute deviations of observations y_{ij} from the treatment median $\tilde{y}_i$. This deviation is defined as

$$d_{ij} = |y_{ij} - \tilde{y}_i| \begin{cases} i = 1, 2, ..., a, \\ j = 1, 2, ..., n_i. \end{cases} \tag{2.23}$$

The modified Levene' test is used to evaluate whether there are differences in the mean of all the treatments. The test statistic is obtained by applying an F-ANOVA over the absolute differences.

Another test for variance was designed by Box (1953) to deal with nonnormal situation and k groups of independently distributed observations. The sample variance of each subgroup is computed. The estimated variance is denoted by s^2, where $s^2 = \sum \phi_t s_t^2 / \Phi$ and Φ is the total number of degrees of freedom. We then take the logarithm of the variance of each subgroup and obtain

$$M_1 = \Phi \ln s^2 - \sum_t \phi_t \ln s_t^2.$$

A χ_k, where k is the number of subgroups, can be used to evaluate if the difference in variance is significant. This statistic rely on large samples. For small samples, a correction is offered by Box. When observations are not normal, the null distribution is adjusted using a function of the kurtosis.

According to Bhandary and Dai (2013), the Box's test has a weak power for the small sized data .

A test for variance in a Randomized Complete Block Design was developed by Gill (1984) to give an exact criterion under the following conditions:

- Error variance is different between treatments.
- Errors are correlated within a block but independent from block to block.

With this assumption the test is executed, ignoring the heterogeneity of variances we found that the tabulated F is too low and hence we get significance. Thus we do not know if this is because there are really significance evidence or the large F is due to the heterogeneity of variances, for this reason, is difficult to draw conclusions about the treatments differences.

Conover et al. (1981) presented a comparative study of test for homogeneity of variances in their research. They provide a list of tests that have a stable Type I error rate when the normality assumption may not be true. Test with modifications of the likelihood ratio, modifications employing an estimate of kurtosis, modification of the F test and modifications of nonparametric tests are presented in this comparison study. Symmetric distribution like normal, and double exponential and asymmetric distributions such as uniform-squared, normal-squared and double-exponential-squared were used. Some of these test, such as Bar, Coch, and Hart, had an uncontrolled risk of

Type I error when the populations are asymmetric. The principal findings are:

- Three tests appear to have the best behavior in terms of robustness and power: Lev1:med, F-K:med χ^2, and F-K: med F.
- Replacing the treatment mean $\bar{X}$ by the treatment median $\tilde{X}$ produced a significant decrease in the Type I error rate in some test.
- The χ^2 and F approximations resulted in almost identical test.
- Poor performance of most of these test when the distributions were asymmetric.
- Test as Talwar-Gentle and FAB: med, never rejected the null hypothesis with a sample size (5,5,5,5).

Bhandary and Dai (2008) developed a method that proves the equality of variance when the data are normal, following a multiple comparison procedure with correction for the global error. Some advantages are: (1) it is easy to perform, (2) there are no stringent requirements about the sample size, and (3) demonstrate good power in comparison with other tests such as Hartley, Levene and Bartlett's tests. This test is sensitive to changes in the distribution.

A test for the equality of variances in RCBD was developed by Bhandary and Dai (2013). This test is useful when there are larger number of treatments and small block size. Their test was compared with other eight test under the normality assumption.

Test for homoscedasticity of variance have been developed using linear ranks such as (Conover et al., 2018) this research compared 66 variations of the test for equal variances and found that three of them were the most powerful. Tests were evaluated with Normal and Laplace, both symmetric distributions; and Normal squared, Laplace squared and Lognormal, skewed distributions.

Chapter 3

Methodology

3.1 Mathematical Framework

To understand the problem under study, it is necessary to define a mathematical model and assumptions for the test we base our proposed test. The test used as a reference to develop this research is the *Friedman test*, which is an extension of the sign test, but using ranks to evaluate more than two treatments (Conover, 1999). This nonparametric test suffers of lack of power when only three treatments are compared, but its power increases when the number of treatments is more than four. This test is preferred to deal with the problem of comparing a different treatments, when a is greater or equal to 4. The design is an RCBD (see Subsection 2.1.5). This method is based on the ranks of the observations within each block. It is considered a two-way ANOVA with ranks with no interaction.

There are b mutually independent blocks, a treatments, and each variables is defined as $(X_{1j}, X_{2j}, ..., X_{bj})$, where $j = 1, 2, ...a$. The observation X_{ij} is in block i and is associated with treatment j.

It is necessary to obtain the ranks for each observation $R(X_{ij})$, the ranking is computed from 1 to b within each block (row) i, this means, each observation is compared with each other, and the rank 1 is assigned to the smallest value within block i, and so on until the a rank, that is for the largest observation. The ranking is computed for all the b blocks. In case of ties, use the average rank. After this sum of the ranks per treatment is calculate and we obtain R_j as follows

$$R_j = \sum_i R(X_{ij}) \qquad\qquad j = 1, 2, ..., b. \tag{3.1}$$

3.1.1 Assumptions

A1. Observations between blocks are independent of each other.

A2. The ranking is made within each block.

3.1.2 Friedman's Test statistic

The statistic used in this test is:

$$T_2 = \frac{(b-1)T_1}{b(a-1) - T_1} \tag{3.2}$$

where T_1 is:

$$T_1 = \frac{(a-1)\left[\sum_{i=1}^{a} R_i^2 - bC_1\right]}{A_1 - C_1}. \tag{3.3}$$

R_j is the sum of the ranks for treatment j and A_1 and C_1 are:

$$A_1 = \sum_{j=1}^{b} \sum_{i=1}^{a} [R(X_{ij}]^2 \tag{3.4}$$

$$C_1 = \frac{ba(a+1)^2}{4} \tag{3.5}$$

3.1.3 Friedman's hypothesis and reject region

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \tag{3.6}$$

$$H_1 : \tau_i \neq 0 \text{ for at least one i.} \tag{3.7}$$

T_2 is approximately distributed as an F variable with $(a-1)$ degrees of freedom in the numerator, and $(b-1)(a-1)$ in the denominator (Conover, 1999).

Table 3.1: Scores.

Test	**Score Function** a_{Ni}
Mood	$(r - \frac{N+1}{2})^2$
FAB	$\frac{N+1}{2} - \lvert r - \frac{N+1}{2} \rvert$
Klotz	$[\Phi^{-1}(\frac{r}{N+1})]^2$
NPL	$\lvert r - \bar{r}_{ij} \rvert$
SR	r^2
TG	r
FK	$\Phi^{-1}[\frac{1}{2} + \frac{r}{2(N+1)}]$

3.2 Methodology

Using the Friedman ranking procedure, several nonparametric methods to test the equality of variance in an RCBD were developed. As we describe in Section 1, nonparametric tests can be created with the use of a rank transformation. In this case, we will use several data alignments and linear rank transformations (scores), as in Conover et al. (1981), to replace the simple rank transformation used by Friedman within each data block.

Proposed steps to evaluate homoscedasticity are presented in the following subsections. A summary of the methodology is show at Subsection 3.2.4.

3.2.1 Select scoring type and alignment

Linear rank test from Conover et al. (1981) were adapted. First, a data a alignment is carried, and then a linear rank transformation, or scoring a_{Ni}, is used, where N is the sum of the individual sample size n_i. This is shown in the tables 3.1 and 3.2 respectively. Table 3.3 shows a summary of the score-alignment combinations that were evaluated in this study. The χ^2 and the F statistics applied over the scores are described later. Next, a brief description of the test developed.

Mood The first test for the equality variance when the normality assumption is not accomplished, and the distribution is not known (nonparametric cases) was present by Mood (1954). In the nonparametric test, the null hypothesis assumes identical distributions and therefore equal means. Therefore Conover (1981) made an adaptation of this test and all the nonparametric test. Instead

Table 3.2: Alignment.

Alignment	**Alignment**
Sample mean	$X_{ij} - \bar{X}_i$
Sample median	X_{ij} - $\tilde{X}_i$
Unadjusted	X_{ij}
Overall mean	$\lvert X_{ij} - \bar{X}_i \rvert$
overall median	$\lvert X_{ij} - \tilde{X}_i \rvert$

Table 3.3: Tests statistics with the combination of scores and alignments.

Test
Mood sample mean
Mood sample median
Mood unadjusted
FAB sample mean
FAB sample median
FAB unadjusted
Klotz sample mean
Klotz sample median
Klotz unadjusted
NPL sample mean
NPL sample median
NPL unadjusted
SR sample mean
SR sample median
SR overall mean
SR overall median
TG sample mean
TG sample median
TG overall mean
TG overall median
FK sample mean
FK sample median
FK overall mean
FK overall median

of letting R_{ij} be the rank of X_{ij} when the equality of means is presented or $(X_{ij} - \mu_i)$ when there are not equal, let R_{ij} be the rank of $(X_{ij} - \tilde{X}_i)$. The X_{ij} observation is replaced by a score of a_N. This test results robust and powerful as some parametric test.

FAB This test was introduced by Freund and Ansari (1957) and then modified by Ansari (1960), similarly to the Mood test, this is based on the scores $(N+1)/2 - |R_{ij} - (N+1)/2|$; in this case apparently is a smallest variances, when the observations are closest to the median, and largest variances when are farthest from the median.

Klotz A normal scores test by Klotz (1962) was introduced; this test uses the normal quantiles. It uses the $R_{ij} = (N+1)$ quantile of the standard normal distribution.

NPL A nonparametric version of Levene's test was develop by Nordstokke and Zumbo (2010) this use the rank transformation by Conover (1981). The score is defined for the absolute of the difference between the overall rank R_{ij}, and the average of the ranks in that r sample is subtracted from each rank.

SR The squared ranks test by Conover (2018), made the same deviation subtracting the X_{ij} to the respective mean or median, and then obtain the squared ranks.

TG Talwar and Gentle Talwar and Gentle (1977) introduce this test, the result was a robust test to some nonnormal distributions.

FK Fligner and Killeen (1976) suggest the T-G test, in this use the normal scores from the positive half of the standard normal distribution.

These are a list of test for variances, that appear to have controlled type I error, and these scores were used to develop this research.

3.2.2 Test statistics

Four statistics were used to analyze the data of this research; two of them are obtained to approximate a χ^2 distribution and the other two for the F distribution. For the χ^s statistics, the Friedman and Kruskal Wallis tests were used as a reference to obtain the χ^2_F and χ^2_K respectively. And these follow a χ^2 distribution with $(a-1)$ degrees of freedom.

χ^2 **statistics**

$$\chi^2_F = \left(\frac{\text{MST}}{Var(x)}\right)\left(\frac{a-1}{a}\right), \tag{3.8}$$

$$\chi^2_K = \frac{b(na-1)(\text{SST})}{\text{TSS}}. \tag{3.9}$$

Where MST is the mean squared of the treatment, var (x) is the sample variance, a is the length of the treatment, b is the length of the block, n is the number of replicates per cell, SST is the sum of squares of the treatment, and TSS is the total sum of squares

In the case of the F statistics, the interaction or the absence of it was taken as a reference to obtain the F_{one} and F_{two}. First, in F_{one} the interaction between factors and blocks was not considered, so this statistic basically an adaptation of the F from a one-way ANOVA. Second, F_{two} considers the interaction from a two way ANOVA. They follow a F distribution: F_{one} with $(a-1)$ degress of freedom in the numerator and $(nba-a)$ in the denominator; and F_{two} with $(a-1)$, $(n-1)(ba)$ degrees of freedom, respectively . The statistics are

F **statistics**

$$F_{two} = \frac{\frac{\text{TSS}-\text{SST}}{(nba-1)-(a-1)}}{\text{SST}}, \tag{3.10}$$

$$F_{one} = \frac{\text{SST}/(a-1)}{(\text{TSS}-\text{SST})/(nba-a)}. \tag{3.11}$$

where a is the length of the treatment, b is the length of the block, n is the number of replicates per cell, SST is the sum of squares of the treatment and TSS is the total sum of squares

3.2.3 Hypothesis

Hypothesis

This statistics test are used to prove the null hypothesis if the variances of the treatments are equal within each block. Hypotheses are:

$$H_0 : \sigma_i = \ldots = \sigma_a = 0, \tag{3.12}$$

$$H_1 : \text{At least one pair is different.} \tag{3.13}$$

3.2.4 Summary of the proposed procedure

To test for homoscedasticity as defined in Subsection 3.2.3, the proposed methods follow:

1. Align your data as in Table 3.2.
2. Rank aligned data within blocks, as done by Friedman.
3. Apply a score function from Table 3.1. to the ranks.
4. Compute a test statistic as shown in Subsection 3.2.2.
5. Calculate the p-value using the corresponding null distribution associated with the test statistic in Section 3.2.2.
6. If the p-value $\leq \alpha$, reject the null hypothesis.

Finally, the `R` codes used to develop the simulations can be found in the following link: https://github.com/ptoram/NP-Test-for-homoscedasticity-with-Blocks

Chapter 4

Simulation Desing and Results

4.1 Simulation Design

With the objective to obtain robust tests, a pre-selection of the tests were developed. Using a normal distribution and with the two test statistics for the χ^2, 1000 simulations for each test were made in the software R. The first part of the analysis consisted in examining which test were robust. This is defined by Conover et al. (1981) as the probability of a type I error less than 0.10 for a 5 percent critical value. The test selected were the only ones with robustness, are shown in the Table 4.1. From the 96 different combinations of alignments, scoring and test statistics, we found robustness in only 9 alignment-scoring combinations with 4 test statistics, for a total of 36 tests. The same analysis was develop for the F statistics, and the results are the equal than for the χ^2.

The simulations were developed with 1000 simulations per scenario; the software used for this was R. Six different size or replicates per cell were used (2,5,10,20,50,100). A total of eight different combinations of main effects τ were evaluated, the first four for the sample size = 2, an the other four for

Table 4.1: Initial assessment. Type I error probabilities over two χ^2 statistics. Robust test are highlighted.

Test	χ^2_k	χ^2_F
Mood Sample Mean	0.121	0.12
Mood Sample Median	0.123	0.119
Mood Unadjusted	0.027	0.034
FAB Sample Mean	0.133	0.127
FAB Sample Median	0.108	0.124
FAB Unadjusted	0.031	0.032
NPL Sample Mean	0.134	0.13
NPL Sample Median	0.137	0.112
NPL Unadjusted	error	error
Klotz Sample Mean	0.133	0.127
Klotz Sample Median	0.134	0.114
Klotz Unadjusted	0.029	0.035
SR Sample Mean	0.108	0.119
SR Sample Median	0.127	0.113
SR Overall Mean	0.05	0.009
SR Overall Median	0.073	0.012
TG Sample Mean	0.123	0.132
TG Sample Median	0.135	0.126
TG Overall Mean	0.025	0.006
TG Overall Median	0.06	0.016
FK Sample Mean	0.127	0.125
FK Sample Median	0.125	0.114
FK Overall Mean	0.048	0.008
FK Overall Median	0.065	0.019

the sample size = 4:

$$\begin{aligned}\tau_1 &= (0,0),\\ \tau_2 &= (-1,1),\\ \tau_3 &= (-2,2),\\ \tau_4 &= (-3,3),\\ \tau_1 &= (0,0,0,0),\\ \tau_2 &= (-1,-1,1,1),\\ \tau_3 &= (-2,-2,2,2),\\ \tau_4 &= (-3,-3,3,3).\end{aligned}$$

Three different size block amounts were considered: two, four, and eight blocks. For simplicity, all of these blocks were simulated with a main effect of zero, as any given effect is nullified by doing a within block ranking as the proposed methods under analysis.

Eight different σ values were used, the first four for two treatments or samples, and the last four for four treatments or samples:

$$\begin{aligned}
\sigma_1 &= (1, 1),\\
\sigma_2 &= (1, 2),\\
\sigma_3 &= (1, 3),\\
\sigma_4 &= (1, 4),\\
\sigma_1 &= (1, 1, 1, 1),\\
\sigma_2 &= (1, 1, 2, 2),\\
\sigma_3 &= (1, 1, 3, 3),\\
\sigma_4 &= (1, 1, 4, 4).
\end{aligned}$$

The distributions tested were the normal distribution with mean 0 and variance 1, the Normal squared distribution from a $N(0, 1)$, double exponential distribution (also known as Laplace distribution), double exponential squared distribution from a Double exponential(0,1), Lognormal(0,1) distribution, a Gamma with a scale parameter of 1, and finally a Weibull with a scale of 1.

For the distributions that require a shape parameter like Gamma and Weibull, four different values of the shape were used: 0.5, 1, 3, and 10. Table 4.1 shows all the score-alignment combination, and the highlighted rows are the robust tests. These were those who got a type I error probability less than 0.1. Only the χ^2 statistics were tested, these are the mentioned in subsection 3.2.2. All the combinations were tested with an alpha of 0.05.

The procedure consisted in generating an scenario with each combination of distribution, replicate, treatment effect, block, standard deviation, and test statistic, calculating the R_{ij} ranks for each test. Then, a score function is obtained and the statistic tests based on the χ^2 and F statistic is computed. After obtaining all results, a selection of the best representation of the results was required, for this a amount of replicates per cell $n = 5$ was selected. This is because with this number of replicates, differences between each test can be appreciate in a better way than if we increase it or decrease. From the simulation when $n > 5$, the powers were to high to easily detect differences between tests. When $n < 5$, power was too low to detect differences. When $n = 5$, differences between tests were high enough to facilitate a conclusion.

4.1.1 Generate a scenario

Different scenarios were generated following the model *Randomized Complete Block Design* (**RCBD**)

$$X_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}, \quad i = 1, \ldots, a; j = 1, \ldots, b, k = 1, \ldots, n. \tag{4.1}$$

To generate the scenario with random observation, we can use the software `R`; the following arguments need to be specified:

- n are the replicates per cell.
- τ are the effect of the i^{th} treatment is defined by the user.
- β are the effect of the j^{th} block.
- Standard deviation, which is defined by the user.
- μ is the mean.
- Distribution, it can be choose for the following list.
  1. *Normal*: Normal distribution.
  2. *Normal2*: Normal-squared distribution from a N(0,1). Basically a χ^2 with one degree of freedom.
  3. *DoubleExp*: Double exponential distribution (also known as Laplace distribution).
  4. *DoubleExp2*: Double exponential squared distribution from a DoubleExp(0,1).
  5. *LogNormal*: Lognormal distribution.
  6. *Gamma*: Gamma distribution.
  7. *Weibull*: Weibull distribution.
- Others parameters that can be required are:
  - *par.location*: location parameter defined by the user usually is 0.
  - *par.scale*: scale parameter defined by the user usually is 1.
  - *par.shape*: shape parameter defined by the user usually is 0.

4.2 Results

An analysis of symmetric distributions as Normal and Double exponential is presented. First, we offer the results for the symmetric distributions, and later for the skewed distributions Normal2, Double Exponential2, Lognormal, Gamma and Weibull. Although, for each distribution two tables are presented, one for the null hypothesis when there is no change in the standard deviation, and other for the average power when the standard deviation has changed.

4.2.1 Symmetric Distributions

Normal distribution null hypothesis

First, Table 4.2 shows the results for the normal distribution when the standard deviation does not have changes.

In the case of two samples without changes in the mean, all tests are robust and the error Type I is under control for all blocks sizes. When there is a change in the mean, the test SR omedian increases but keeps its robustness. When there are four samples, and there is no change in the mean, all the tests are robust; but the tests FK omedian, SR omedian, and TG omedian showed sensitive to a change in the mean in all block sizes.

Normal distribution power

Table 4.3 shows the results for the power with the Normal distribution when the standard deviation changed.

No changes in mean When there are 2 samples all the tests have a good power, the most powerful statistics for different block sizes (2,4, and 8) were the F_{one}, in the case of 2 blocks the test that had the best behavior was the SR omedian, for 4 blocks was FK omedian and for 8 blocks were SR omean and SR omedian. When the number of samples increased to 4, the statistic χ_F tends to be more conservative when the size of the blocks is small. The tests with more power were SR omedian, SR omean and FK omean for 2, 4, and 8 blocks, respectively.

Table 4.2: Performance with the Normal distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are (1 1) and (1 1 1 1) for 2 and 4 treatments respectively.

Normal distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	$\chi^2_{\mathbf{K}}$	F**one**	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
1 1													
0 0	FAB unadj	.005	.056	.048	.046	.006	.043	.071	.056	.011	.041	.057	.035
	FK omean	.004	.063	.049	.044	.010	.062	.067	.054	.003	.046	.075	.030
	FK omedian	.007	.045	.046	.059	.010	.046	.048	.055	.011	.045	.062	.042
	Klotz unadj	.002	.049	.062	.051	.004	.042	.056	.047	.011	.045	.064	.044
	Mood unadj	.005	.054	.046	.036	.005	.067	.076	.056	.010	.051	.067	.046
	SR omean	.006	.044	.066	.053	.004	.047	.054	.043	.011	.046	.060	.051
	SR omedian	.005	.032	.056	.056	.003	.053	.061	.041	.010	.055	.052	.054
	TG omean	.003	.055	.067	.049	.008	.045	.070	.056	.006	.048	.056	.063
	TG omedian	.008	.048	.055	.043	.007	.057	.072	.045	.003	.056	.054	.041
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omedian	.000	.042	.064	.050	.002	.059	.060	.052	.013	.044	.076	.055
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omedian	.000	.066	.070	.062	.004	.062	.097	.064	.012	.075	.064	.076
	TG omean	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	TG omedian	.000	.019	.034	.021	.003	.029	.048	.024	.001	.051	.034	.029
1 1 1 1													
0 0 0 0	FAB unadj	.000	.049	.044	.051	.000	.061	.059	.045	.000	.048	.072	.052
	FK omean	.000	.043	.044	.047	.000	.042	.055	.053	.000	.037	.064	.058
	FK omedian	.000	.053	.059	.044	.000	.049	.060	.040	.000	.040	.062	.045
	Klotz unadj	.000	.031	.059	.042	.000	.055	.052	.053	.000	.044	.061	.058
	Mood unadj	.000	.041	.053	.038	.000	.056	.067	.054	.000	.047	.054	.058
	SR omean	.000	.046	.050	.051	.000	.066	.054	.045	.000	.048	.061	.052
	SR omedian	.000	.029	.065	.038	.000	.046	.062	.053	.000	.047	.049	.062
	TG omean	.000	.043	.054	.051	.000	.046	.057	.046	.000	.046	.047	.056
	TG omedian	.000	.044	.068	.056	.000	.048	.048	.060	.000	.044	.071	.044
-3 -3 3 3	FAB unadj	.000	.016	.027	.020	.000	.026	.037	.019	.000	.024	.028	.024
	FK omean	.000	.022	.024	.023	.000	.021	.023	.027	.000	.023	.028	.025
	FK omedian	.000	.097	.127	.132	.000	.126	.135	.155	.000	.108	.143	.139
	Klotz unadj	.000	.013	.019	.012	.000	.015	.032	.010	.000	.021	.036	.020
	Mood unadj	.000	.019	.026	.014	.000	.022	.035	.017	.000	.031	.034	.016
	SR omean	.000	.022	.030	.017	.000	.031	.035	.027	.000	.034	.038	.028
	SR omedian	.000	.135	.139	.142	.000	.121	.149	.135	.000	.125	.139	.153
	TG omean	.000	.022	.027	.019	.000	.028	.038	.022	.000	.034	.035	.023
	TG omedian	.000	.103	.115	.113	.000	.119	.135	.112	.000	.122	.121	.132

Change in the mean For 2 samples, the tests that remain robust even if there are changes in mean are FK omedian, SR omedian and TG omedian over all statistics; and the most powerful test was the SR omedian over the three different block sizes. The statistics FAB unadj, FK omean, Klotz unadj, Mood unadj, SR omean, had good power when there was no change in mean, but, they lost power when there was a change in mean.

When 4 samples were used, the best statistic was the F_{two}, and the test with better behavior was SR omedian. The test Mood unadj had good power when there was no change in the mean, but, was not sensitive when a change in mean occurred.

Double exponential distribution null hypothesis

The results obtained for the Double exponential distribution for the case, are shown in Table 4.4. For the sample size of two and without a change in the mean, all the tests have control of the error type I in all the cases of the number of blocks. But, the test FK omedian is sensitive to a change in the mean in the blocks size four and eight, and the SR omedian is sensitive in all the number of blocks. If the sample size increases to four and there is any change in the mean all the test are robust. The tests FK omedian, SR omedian, and TG omedian are sensitive to a change in the mean for all the sizes of the block. The statistics χ_F has a conservative behavior when the sample increases.

Double exponential distribution power

In Table 4.5 are shown the results for the power of the Double Exponential Distribution.

No changes in mean When there are two samples all the tests have good power, the most powerful statistics for the different block size (2,4, and 8) is the F_{one}, the test that has the best behavior is FK omedian, for 2, 4, and 8 blocks. When the sample increases to 4, the statistic χ_F tends to be more conservative when the size of the block are 2 and 4; the most potent tests are SR omedian for 2, 4, and FK omedian for 8 blocks.

Table 4.3: Performance with the Normal distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are [(1 2), (1 3), and (1 4)] and [(1 1 2 2), (1 1 3 3), and (1 1 4 4)] for 2 and 4 treatments respectively.

Normal distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
Average Power (1 2), (1 3), (1 4)													
0 0	FAB unadj	.201	.446	.439	.475	.433	.690	.764	.694	.782	.896	.911	.895
	FK omean	.175	.498	.512	.491	.518	.781	.794	.786	.842	.937	.951	.929
	FK omedian	.195	.537	.572	.556	.567	.785	.824	.796	.856	.946	.949	.940
	Klotz unadj	.146	.462	.489	.478	.484	.745	.776	.733	.818	.920	.941	.925
	Mood unadj	.149	.484	.492	.480	.477	.744	.761	.745	.804	.929	.936	.918
	SR omean	.187	.517	.545	.504	.537	.771	.815	.779	.855	.944	.953	.946
	SR omedian	.214	.560	.575	.562	.590	.800	.820	.806	.855	.940	.953	.937
	TG omean	.157	.450	.507	.452	.459	.728	.767	.721	.798	.925	.927	.919
	TG omedian	.195	.507	.527	.533	.522	.743	.775	.759	.812	.924	.937	.926
-3 3	FAB unadj	.001	.020	.015	.021	.006	.022	.027	.022	.013	.042	.058	.041
	FK omean	.005	.027	.032	.030	.009	.060	.071	.054	.035	.144	.161	.131
	FK omedian	.092	.478	.507	.466	.463	.744	.789	.756	.808	.926	.941	.919
	Klotz unadj	.001	.011	.012	.010	.003	.012	.017	.018	.006	.024	.031	.029
	Mood unadj	.002	.010	.014	.011	.000	.015	.022	.018	.005	.041	.048	.031
	SR omean	.003	.035	.053	.039	.015	.082	.092	.078	.050	.197	.223	.189
	SR omedian	.174	.544	.531	.528	.519	.771	.809	.776	.827	.938	.947	.933
	TG omean	.003	.019	.022	.022	.004	.033	.037	.028	.009	.055	.051	.042
	TG omedian	.082	.367	.399	.366	.375	.678	.707	.659	.717	.886	.900	.887
Average Power (1 1 2 2), (1 1 3 3), (1 1 4 4)													
0 0 0 0	FAB unadj	.000	.622	.673	.655	.091	.875	.896	.872	.589	.983	.984	.982
	FK omean	.000	.702	.739	.732	.107	.927	.934	.928	.665	.995	.997	.997
	FK omedian	.000	.724	.747	.740	.135	.929	.935	.919	.687	.991	.994	.993
	Klotz unadj	.000	.696	.729	.699	.040	.915	.933	.920	.656	.995	.995	.992
	Mood unadj	.000	.686	.722	.702	.117	.918	.916	.911	.647	.993	.995	.992
	SR omean	.000	.714	.734	.726	.149	.922	.935	.933	.681	.996	.994	.997
	SR omedian	.000	.725	.756	.745	.186	.910	.933	.930	.692	.994	.994	.996
	TG omean	.000	.654	.701	.654	.102	.889	.903	.886	.600	.986	.991	.984
	TG omedian	.000	.681	.695	.680	.138	.878	.899	.896	.637	.983	.987	.984
-3 -3 3 3	FAB unadj	.000	.039	.054	.044	.000	.065	.074	.069	.000	.134	.146	.131
	FK omean	.000	.069	.075	.077	.000	.125	.147	.125	.000	.288	.316	.275
	FK omedian	.000	.793	.815	.803	.210	.951	.951	.957	.792	.996	.997	.997
	Klotz unadj	.000	.023	.038	.026	.000	.047	.049	.043	.000	.077	.089	.078
	Mood unadj	.000	.039	.042	.034	.000	.053	.056	.053	.000	.101	.115	.091
	SR omean	.000	.078	.103	.090	.000	.153	.181	.148	.000	.348	.357	.328
	SR omedian	.000	.822	.828	.833	.303	.953	.957	.956	.835	.996	.998	.999
	TG omean	.000	.049	.053	.046	.000	.070	.073	.060	.000	.089	.102	.085
	TG omedian	.000	.734	.752	.748	.181	.917	.927	.921	.719	.993	.992	.991

Table 4.4: Performance with the Double exponential distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are (1 1) and (1 1 1 1) for 2 and 4 treatments respectively.

Double exponential distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
1 1													
0 0	FAB unadj	.011	.059	.045	.063	.008	.055	.072	.041	.008	.048	.058	.046
	FK omean	.005	.054	.060	.051	.005	.045	.063	.057	.006	.051	.062	.063
	FK omedian	.003	.056	.056	.038	.006	.044	.054	.047	.003	.051	.061	.045
	Klotz unadj	.006	.042	.057	.039	.006	.048	.056	.057	.006	.053	.046	.071
	Mood unadj	.000	.060	.053	.043	.008	.050	.075	.056	.006	.055	.070	.048
	SR omean	.004	.049	.064	.056	.007	.054	.044	.047	.008	.055	.042	.032
	SR omedian	.004	.049	.048	.045	.003	.047	.063	.048	.007	.062	.068	.047
	TG omean	.007	.050	.065	.054	.005	.045	.056	.040	.013	.041	.049	.054
	TG omedian	.008	.032	.058	.051	.003	.041	.065	.059	.005	.046	.067	.057
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omedian	.000	.083	.079	.082	.018	.101	.092	.076	.021	.088	.106	.087
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.000	.000	.001	.001	.000	.000	.001	.000	.000	.000	.001	.000
	SR omedian	.019	.125	.151	.103	.020	.108	.129	.111	.030	.106	.126	.110
	TG omean	.000	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000
	TG omedian	.000	.058	.076	.044	.010	.075	.088	.059	.008	.061	.064	.066
1 1 1 1													
0 0 0 0	FAB unadj	.000	.041	.058	.055	.000	.052	.067	.056	.000	.044	.067	.062
	FK omean	.000	.049	.052	.059	.000	.059	.051	.046	.000	.055	.062	.063
	FK omedian	.000	.039	.045	.061	.000	.061	.061	.055	.000	.040	.070	.058
	Klotz unadj	.000	.038	.048	.037	.000	.057	.054	.045	.000	.033	.070	.049
	Mood unadj	.000	.033	.048	.046	.000	.049	.067	.058	.000	.051	.064	.050
	SR omean	.000	.058	.051	.040	.000	.049	.053	.058	.000	.041	.058	.047
	SR omedian	.000	.040	.056	.054	.000	.058	.053	.052	.000	.047	.052	.047
	TG omean	.000	.049	.064	.045	.000	.046	.052	.051	.000	.044	.061	.054
	TG omedian	.000	.044	.058	.041	.000	.046	.075	.060	.000	.051	.056	.042
-3 -3 3 3	FAB unadj	.000	.020	.020	.019	.000	.019	.029	.022	.000	.024	.027	.015
	FK omean	.000	.019	.023	.030	.000	.023	.025	.021	.000	.023	.041	.019
	FK omedian	.000	.210	.249	.238	.000	.202	.206	.222	.003	.181	.203	.227
	Klotz unadj	.000	.011	.016	.011	.000	.017	.018	.015	.000	.017	.025	.023
	Mood unadj	.000	.017	.020	.023	.000	.021	.025	.016	.000	.024	.028	.022
	SR omean	.000	.023	.028	.021	.000	.029	.024	.027	.000	.029	.036	.020
	SR omedian	.000	.209	.233	.280	.000	.216	.232	.293	.005	.221	.233	.284
	TG omean	.000	.024	.024	.025	.000	.021	.034	.027	.000	.028	.033	.022
	TG omedian	.000	.212	.242	.240	.001	.189	.222	.240	.002	.177	.223	.227

Change in the mean For the case of 2 samples, the tests that remain robust even if there are changes in mean are FK omedian and SR omedian for almost all the statistics, and the best test is SR omedian for the three different block size. For the other side, the statistics FAB unadj, FK omean, Klotz unadj, Mood unadj, TG omean, have good power when there is no change in mean, but, they are sensitive to the change in mean.

When the sample is 4, the test with better behavior is SR omedian. The test FAB unadj, Klotz unadj, Mood unadj and TG omean to have good power when there is no change in the mean, but, are sensitive to the change in mean.

4.2.2 Skewed Distributions

Normal squared distribution null hypothesis

The Table 4.6 shows the results for the Normal squared distribution when there are no changes in the standard deviation.

When there is no change in the mean, all the test for two and four samples are robust in all the different size of blocks. If there is a change in the mean with a sample size of two FK omedian, SR omedian, and TG omedian are sensitive to a change in the mean in all the size of blocks. If the sample size increases to four, with a change in mean FK omean, SR omean, and TG omean are sensitive when the size of blocks are four and eight, and FK omedian, are sensitive to all size of blocks.

Normal squared distribution power

Table 4.7 shows the results of the power in the Normal squared Distribution when there is a change in the standard deviation.

No changes in mean When there are two samples all the tests have good power, the tests that have the best behavior is TG omedian, FAB unadj and Mood unadj, for 2, 4, and 8 blocks respectively. When the sample increases to 4, the statistic χ_F tends to be more conservative when there are 2 blocks; the most powerful tests are FAB unadj for 2 and 4 blocks and 5 test for 8 blocks.

Change in the mean For the case of 2 samples, the tests that remain robust even if there are changes in mean are FK omedian and SR omedian for almost all the statistics, and the best test is SR omedian for 2 and 4 blocks and TG

Table 4.5: Performance with the Double exponential distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are [(1 2), (1 3), and (1 4)] and [(1 1 2 2), (1 1 3 3), and (1 1 4 4)] for 2 and 4 treatments respectively.

Double exponential distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
Average Power (1 2), (1 3), (1 4)													
0 0	FAB unadj	.136	.319	.329	.346	.291	.538	.614	.564	.613	.791	.830	.801
	FK omean	.086	.333	.372	.352	.326	.599	.639	.596	.658	.844	.867	.853
	FK omedian	.119	.428	.453	.435	.420	.688	.714	.671	.733	.889	.892	.868
	Klotz unadj	.092	.337	.372	.352	.315	.594	.628	.600	.644	.835	.849	.822
	Mood unadj	.094	.356	.343	.348	.303	.584	.631	.585	.654	.827	.846	.826
	SR omean	.097	.370	.387	.352	.348	.614	.649	.622	.678	.845	.868	.848
	SR omedian	.145	.412	.446	.434	.437	.676	.707	.686	.752	.871	.876	.878
	TG omean	.076	.299	.354	.320	.287	.562	.598	.548	.603	.791	.826	.805
	TG omedian	.115	.387	.419	.408	.383	.629	.666	.633	.688	.858	.857	.854
-3 3	FAB unadj	.000	.005	.005	.006	.000	.003	.004	.002	.000	.002	.006	.006
	FK omean	.002	.015	.019	.018	.003	.026	.038	.017	.009	.051	.064	.051
	FK omedian	.049	.441	.469	.435	.402	.658	.702	.679	.719	.880	.881	.872
	Klotz unadj	.000	.002	.002	.002	.000	.002	.003	.001	.000	.002	.002	.001
	Mood unadj	.001	.003	r.005	.002	.000	.001	.004	.002	.000	.001	.003	.002
	SR omean	.001	.031	.028	.030	.005	.038	.054	.047	.018	.077	.112	.078
	SR omedian	.172	.485	.489	.494	.457	.702	.714	.710	.760	.886	.896	.887
	TG omean	.002	.010	.018	.015	.001	.015	.018	.012	.002	.014	.019	.013
	TG omedian	.062	.312	.386	.323	.306	.593	.625	.578	.640	.814	.825	.817
Average Power (1 1 2 2), (1 1 3 3), (1 1 4 4)													
0 0 0 0	FAB unadj	.000	.495	.533	.511	.020	.755	.780	.764	.363	.925	.940	.925
	FK omean	.000	.525	.556	.535	.015	.789	.836	.799	.404	.956	.963	.957
	FK omedian	.000	.568	.607	.586	.036	.828	.842	.835	.491	.962	.969	.973
	Klotz unadj	.000	.493	.537	.527	.006	.796	.818	.794	.388	.957	.961	.951
	Mood unadj	.000	.511	.560	.523	.027	.790	.805	.799	.406	.953	.955	.947
	SR omean	.000	.539	.581	.542	.027	.814	.818	.810	.425	.961	.963	.954
	SR omedian	.000	.574	.619	.587	.054	.827	.851	.840	.509	.960	.966	.963
	TG omean	.000	.461	.486	.488	.020	.741	.763	.750	.330	.928	.932	.934
	TG omedian	.000	.518	.558	.549	.044	.791	.802	.786	.423	.948	.954	.941
-3 -3 3 3	FAB unadj	.000	.026	.030	.025	.000	.026	.041	.026	.000	.037	.040	.032
	FK omean	.000	.050	.068	.058	.000	.076	.093	.069	.000	.145	.152	.129
	FK omedian	.000	.747	.765	.764	.259	.895	.908	.902	.737	.979	.978	.977
	Klotz unadj	.000	.017	.027	.018	.000	.025	.028	.028	.000	.032	.042	.023
	Mood unadj	.000	.023	.026	.023	.000	.028	.030	.023	.000	.034	.040	.026
	SR omean	.000	.059	.077	.068	.000	.094	.103	.095	.000	.182	.203	.166
	SR omedian	.000	.761	.786	.790	.355	.911	.915	.915	.774	.976	.979	.985
	TG omean	.000	.044	.046	.048	.000	.044	.044	.046	.000	.043	.055	.041
	TG omedian	.000	.705	.732	.725	.232	.869	.878	.889	.668	.959	.972	.972

Table 4.6: Performance with the Normal squared distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are (1 1) and (1 1 1 1) for 2 and 4 treatments respectively.

Normal squared distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
1 1													
0 0	FAB unadj	.005	.046	.038	.075	.006	.044	.060	.045	.007	.041	.059	.045
	FK omean	.006	.039	.056	.042	.007	.045	.056	.046	.003	.044	.065	.051
	FK omedian	.001	.045	.042	.064	.006	.033	.062	.056	.010	.053	.066	.058
	Klotz unadj	.005	.035	.065	.048	.003	.054	.066	.047	.006	.055	.068	.047
	Mood unadj	.005	.050	.048	.051	.005	.054	.057	.053	.009	.050	.066	.044
	SR omean	.005	.050	.061	.044	.010	.049	.071	.047	.012	.050	.063	.049
	SR omedian	.006	.033	.056	.057	.007	.046	.067	.037	.008	.047	.069	.050
	TG omean	.005	.037	.076	.052	.011	.052	.067	.044	.012	.057	.054	.059
	TG omedian	.003	.050	.052	.054	.005	.058	.055	.040	.011	.070	.069	.055
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.007	.014	.015	.001	.019	.030	.014	.004	.058	.073	.047
	FK omedian	.004	.201	.212	.203	.111	.311	.348	.312	.291	.525	.555	.546
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.000	.023	.027	.016	.001	.040	.061	.037	.009	.135	.179	.099
	SR omedian	.081	.229	.241	.224	.150	.391	.434	.398	.341	.601	.640	.615
	TG omean	.000	.020	.029	.016	.001	.047	.045	.036	.009	.132	.144	.096
	TG omedian	.011	.142	.214	.172	.097	.348	.371	.329	.309	.588	.608	.599
1 1 1 1													
0 0 0 0	FAB unadj	.000	.042	.065	.056	.000	.047	.055	.047	.000	.049	.065	.056
	FK omean	.000	.049	.059	.065	.000	.050	.063	.062	.000	.060	.053	.054
	FK omedian	.000	.042	.055	.047	.000	.041	.059	.044	.000	.044	.050	.038
	Klotz unadj	.000	.044	.048	.053	.000	.055	.042	.032	.000	.047	.054	.045
	Mood unadj	.000	.036	.057	.053	.000	.046	.065	.042	.000	.047	.066	.056
	SR omean	.000	.047	.060	.066	.000	.048	.049	.044	.000	.052	.062	.058
	SR omedian	.000	.042	.060	.039	.000	.052	.056	.051	.000	.048	.065	.065
	TG omean	.000	.037	.053	.060	.000	.037	.049	.047	.000	.053	.060	.039
	TG omedian	.000	.049	.066	.044	.000	.050	.060	.046	.000	.045	.060	.038
-3 -3 3 3	FAB unadj	.000	.021	.029	.025	.000	.025	.023	.016	.000	.017	.025	.020
	FK omean	.000	.052	.063	.051	.000	.089	.104	.076	.000	.185	.227	.176
	FK omedian	.000	.582	.638	.651	.080	.856	.887	.887	.520	.995	.992	.990
	Klotz unadj	.000	.010	.019	.009	.000	.019	.021	.011	.000	.021	.032	.020
	Mood unadj	.000	.016	.029	.020	.000	.028	.026	.020	.000	.029	.026	.019
	SR omean	.000	.053	.112	.083	.000	.177	.208	.171	.000	.459	.515	.460
	SR omedian	.000	.654	.703	.706	.148	.909	.904	.904	.675	.992	.996	.996
	TG omean	.000	.065	.077	.053	.000	.174	.189	.163	.000	.515	.557	.505
	TG omedian	.000	.718	.750	.716	.090	.931	.959	.940	.726	.998	1.000	.998

omedian for 8 blocks. For the other side, the tests FAB unadj, FK omean, Klotz unadj, Mood unadj, SR omean and TG omean, have good power when there is no change in mean, but, they are sensitive to the change in mean.

When the sample is 4, the test with the better behavior FK omedian, SR omean, SR omedian, TG omean and TG omedian for almost all the statistics except χ^2_F. The test FAB unadj, Klotz unadj and Mood unadj have good power when there is no change in the mean, but, are sensitive to the change in mean.

Double exponential squared distribution null hypothesis

Table 4.8 shows the results for the Double exponential squared distribution when there are no changes in the standard deviation.

When there is no change in the mean, all the test for two and four samples are robust in all the different block sizes. If there is a change in the mean when dealing with two samples, FK omean, SR omean and TG omean are sensitive when the number of blokcs is four or eight. FK omedian, SR omedian, and TG omedian are sensitive for all number of blocks when there is a change in the mean. When the sample increases the tests FK omean, FK omedian, SR omean, SR omedian, TG omean, and TG omedian are sensitive for all block sizes. The statistic chi_F has a conservative behavior in this case when the number of samples increases to four and there is a change in the mean.

Double exponential squared distribution power

Table 4.9 shows the results of the power in the Double exponential squared distribution when there is a change in the standard deviation.

No changes in mean When there are 2 samples all the tests have good power, the tests that have the best behavior is TG omedian, for 2 blocks and TG omean, for 4 and 8 blocks. When the number of samples increases to 4, the statistic χ_F tends to be more conservative when there are 2 blocks; the most powerful tests are TG omean for 2 blocks, FAB unadj for 4 blocks and 8 blocks all the test in F_{one} have excellent power.

Change in the mean For the case of 2 samples the tests that remain robust even if there are changes in mean are FK omedian, SR omedian, TG omean

Table 4.7: Performance with the Normal squared distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are [(1 2), (1 3), and (1 4)] and [(1 1 2 2), (1 1 3 3), and (1 1 4 4)] for 2 and 4 treatments respectively.

Normal squared distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
Average Power (1 2), (1 3), (1 4)													
0 0	FAB unadj	.288	.559	.545	.589	.604	.840	.891	.848	.935	.982	.990	.986
	FK omean	.222	.513	.571	.537	.557	.798	.826	.807	.889	.971	.970	.969
	FK omedian	.231	.545	.564	.552	.574	.772	.797	.783	.847	.936	.948	.947
	Klotz unadj	.200	.472	.505	.500	.531	.806	.841	.815	.909	.980	.990	.987
	Mood unadj	.206	.549	.528	.534	.560	.843	.860	.826	.923	.988	.992	.985
	SR omean	.240	.535	.569	.562	.594	.835	.848	.829	.905	.976	.980	.974
	SR omedian	.254	.553	.563	.546	.571	.790	.797	.795	.849	.936	.945	.934
	TG omean	.243	.533	.590	.560	.606	.840	.863	.831	.918	.973	.983	.979
	TG omedian	.272	.576	.597	.602	.630	.825	.846	.824	.895	.966	.970	.960
-3 3	FAB unadj	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.004	.007	.007	.000	.009	.009	.006	.001	.022	.029	.020
	FK omedian	.004	.074	.088	.079	.012	.095	.114	.101	.037	.145	.177	.154
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.000	.006	.008	.006	.000	.009	.020	.009	.001	.029	.034	.023
	SR omedian	.012	.088	.095	.095	.027	.121	.137	.119	.062	.157	.193	.176
	TG omean	.001	.021	.036	.024	.004	.083	.108	.064	.038	.237	.288	.213
	TG omedian	.002	.043	.057	.048	.008	.063	.084	.062	.017	.084	.098	.089
Average Power (1 1 2 2), (1 1 3 3), (1 1 4 4)													
0 0 0 0	FAB unadj	.000	.812	.841	.827	.195	.987	.990	.985	.881	1.000	1.000	1.000
	FK omean	.000	.746	.756	.758	.157	.954	.954	.953	.735	.998	.998	.997
	FK omedian	.000	.671	.704	.710	.147	.896	.900	.899	.627	.989	.986	.983
	Klotz unadj	.000	.721	.769	.727	.044	.967	.970	.973	.723	1.000	1.000	1.000
	Mood unadj	.000	.793	.836	.809	.151	.986	.988	.986	.869	1.000	1.000	1.000
	SR omean	.000	.776	.816	.796	.221	.964	.974	.967	.794	.999	1.000	.999
	SR omedian	.000	.682	.721	.706	.155	.885	.914	.890	.632	.983	.991	.985
	TG omean	.000	.773	.812	.792	.224	.958	.964	.956	.798	1.000	1.000	.999
	TG omedian	.000	.754	.775	.750	.229	.932	.943	.943	.721	.997	.998	.997
-3 -3 3 3	FAB unadj	.000	.019	.030	.025	.000	.023	.025	.021	.000	.025	.032	.020
	FK omean	.000	.046	.056	.048	.000	.060	.077	.070	.000	.113	.132	.105
	FK omedian	.000	.136	.158	.148	.001	.187	.211	.219	.006	.258	.273	.254
	Klotz unadj	.000	.015	.017	.012	.000	.025	.028	.017	.000	.027	.030	.017
	Mood unadj	.000	.016	.024	.023	.000	.019	.029	.022	.000	.024	.035	.019
	SR omean	.000	.052	.065	.055	.000	.074	.106	.089	.000	.168	.186	.162
	SR omedian	.000	.168	.187	.189	.002	.219	.240	.243	.011	.308	.323	.318
	TG omean	.000	.107	.134	.119	.000	.275	.330	.284	.000	.720	.746	.709
	TG omedian	.000	.181	.219	.216	.001	.289	.304	.304	.026	.414	.436	.432

Table 4.8: Performance with the Double exponential squared distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are (1 1) and (1 1 1 1) for 2 and 4 treatments respectively.

Double exponential squared Distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
1 1													
0 0	FAB unadj	.008	.049	.039	.060	.007	.041	.068	.051	.008	.045	.047	.045
	FK omean	.002	.049	.066	.047	.009	.032	.059	.049	.005	.052	.066	.053
	FK omedian	.005	.042	.049	.056	.009	.052	.054	.041	.011	.047	.063	.050
	Klotz unadj	.003	.046	.055	.060	.008	.049	.056	.047	.010	.057	.072	.045
	Mood unadj	.001	.042	.061	.065	.003	.050	.068	.055	.006	.047	.057	.056
	SR omean	.005	.056	.059	.046	.006	.053	.059	.060	.006	.043	.055	.045
	SR omedian	.002	.044	.063	.054	.005	.068	.068	.043	.011	.054	.062	.042
	TG omean	.005	.045	.062	.049	.003	.055	.075	.049	.013	.050	.052	.067
	TG omedian	.006	.061	.059	.052	.013	.042	.064	.050	.011	.051	.053	.053
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.004	.033	.054	.042	.011	.114	.131	.109	.067	.297	.364	.280
	FK omedian	.021	.239	.237	.254	.171	.377	.450	.391	.381	.641	.662	.629
	Klotz unadj	.000	.000	.001	.000	.000	.000	.000	.001	.000	.000	.000	.000
	Mood unadj	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.007	.066	.100	.069	.032	.193	.217	.179	.163	.471	.542	.442
	SR omedian	.128	.236	.231	.259	.182	.436	.471	.454	.472	.675	.761	.697
	TG omean	.002	.060	.085	.062	.023	.182	.225	.180	.153	.515	.595	.501
	TG omedian	.021	.224	.246	.233	.141	.431	.489	.382	.441	.691	.737	.723
1 1 1 1													
0 0 0 0	FAB unadj	.000	.048	.067	.055	.000	.041	.051	.058	.000	.047	.048	.038
	FK omean	.000	.046	.058	.049	.000	.046	.061	.046	.000	.055	.058	.061
	FK omedian	.000	.043	.056	.044	.000	.041	.051	.047	.000	.057	.058	.055
	Klotz unadj	.000	.049	.053	.045	.000	.035	.060	.045	.000	.048	.052	.050
	Mood unadj	.000	.033	.069	.062	.000	.049	.061	.054	.000	.050	.069	.051
	SR omean	.000	.045	.060	.057	.000	.039	.058	.042	.000	.053	.056	.048
	SR omedian	.000	.051	.064	.040	.000	.048	.069	.052	.000	.051	.059	.050
	TG omean	.000	.041	.045	.039	.000	.051	.053	.047	.000	.051	.065	.055
	TG omedian	.000	.053	.065	.057	.000	.043	.051	.051	.000	.038	.055	.047
-3 -3 3 3	FAB unadj	.000	.021	.020	.028	.000	.029	.026	.020	.000	.024	.042	.028
	FK omean	.000	.125	.156	.151	.000	.324	.380	.312	.001	.725	.744	.702
	FK omedian	.000	.681	.702	.693	.134	.932	.927	.918	.663	.996	.996	.999
	Klotz unadj	.000	.013	.020	.015	.000	.023	.034	.014	.000	.023	.034	.019
	Mood unadj	.000	.020	.031	.020	.000	.024	.025	.014	.000	.026	.026	.024
	SR omean	.000	.221	.257	.223	.000	.534	.581	.515	.013	.912	.944	.921
	SR omedian	.000	.749	.772	.752	.244	.927	.949	.948	.800	.995	.998	.997
	TG omean	.000	.207	.281	.254	.000	.609	.630	.574	.014	.972	.969	.944
	TG omedian	.000	.775	.807	.819	.199	.972	.984	.979	.872	.999	.999	1.000

and TG omedian, for almost all the statistics except the χ_F^2; and the best test is TG omean for all the blocks. For the other side, the tests FAB unadj, Klotz unadj and Mood unadj, have good power when there is no change in the mean, but, they are sensitive to the change in mean.

When the number of samples is 4, the test with the better behavior FK omean, FK omedian, SR omean, SR omedian, TG omean and TG omedian for almost all the statistics except χ_F^2. The test FAB unadj, Klotz unadj and Mood unadj have good power when there is no change in the mean, but, are lack power when there is a change in mean.

Lognormal distribution null hypothesis

Table 4.10 shows the results for the Lognormal distribution when there are no changes in the standard deviation. When there is no change in the mean, all the tests for two and four samples are robust in all the different size of blocks. If there is a change in the mean with a sample size of two the test FK omean is sensitive when the size of block increases to eight. FK omedian, SR omedian, and TG omedian are sensitive in all the size of blocks. With an increase in the sample size to four the tests are sensitive to change in the mean, except FAB unadj, Klotz unadj, and Mood unadj.

Lognormal distribution power

Table 4.11 shows the results of the power in the Lognormal Distribution when there is a change in the standard deviation.

No changes in mean When there are 2 samples all the tests have good power, the tests that have the best behavior is TG omedian, for 2 blocks and TG omean, for 4 and 8 blocks. When the sample increases to 4, the statistic χ_F tends to be more conservative when there are 2 blocks; the most powerful tests are TG omean for 2 blocks, FAB unadj for 4 blocks and 8 blocks all the test in F_{one} have excellent power.

Change in the mean For the case of 2 samples the tests that remain robust even if there are changes in mean are FK omedian, SR omedian, TG omean and TG omedian, for almost all the statistics except the χ_F^2; and the best test is TG omean for all the blocks. For the other side, the tests FAB unadj,

Table 4.9: Performance with the Double exponential squared distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are [(1 2), (1 3), and (1 4)] and [(1 1 2 2), (1 1 3 3), and (1 1 4 4)] for 2 and 4 treatments respectively.

Double exponential squared distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
Average Power (1 2), (1 3), (1 4)													
0 0	FAB unadj	.271	.521	.506	.569	.563	.814	.863	.820	.937	.983	.990	.981
	FK omean	.218	.514	.554	.530	.550	.797	.840	.799	.903	.979	.979	.977
	FK omedian	.251	.546	.573	.559	.570	.794	.806	.789	.862	.951	.951	.952
	Klotz unadj	.171	.414	.419	.415	.439	.765	.771	.735	.847	.966	.972	.961
	Mood unadj	.182	.473	.463	.451	.480	.787	.821	.793	.899	.978	.986	.976
	SR omean	.234	.550	.572	.553	.579	.827	.844	.821	.917	.986	.984	.986
	SR omedian	.261	.542	.569	.554	.576	.790	.809	.787	.861	.949	.953	.949
	TG omean	.261	.572	.599	.602	.634	.870	.881	.851	.942	.986	.994	.989
	TG omedian	.303	.613	.633	.621	.676	.842	.868	.859	.911	.978	.984	.982
-3 3	FAB unadj	.000	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.001	.022	.028	.027	.005	.060	.078	.054	.021	.163	.193	.150
	FK omedian	.007	.098	.105	.096	.029	.109	.123	.104	.038	.135	.148	.133
	Klotz unadj	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.002	.034	.037	.035	.007	.079	.100	.072	.037	.211	.273	.196
	SR omedian	.038	.111	.097	.105	.036	.142	.157	.147	.065	.178	.194	.176
	TG omean	.007	.071	.114	.084	.026	.236	.293	.201	.206	.580	.633	.539
	TG omedian	.007	.070	.084	.069	.026	.114	.136	.130	.055	.168	.186	.166
Average Power (1 1 2 2), (1 1 3 3), (1 1 4 4)													
0 0 0 0	FAB unadj	.000	.801	.815	.815	.159	.986	.992	.985	.882	1.000	1.000	1.000
	FK omean	.000	.757	.789	.767	.165	.965	.965	.970	.811	1.000	1.000	.999
	FK omedian	.000	.704	.741	.730	.154	.917	.934	.920	.686	.994	.994	.994
	Klotz unadj	.000	.619	.677	.661	.020	.933	.948	.945	.532	1.000	1.000	1.000
	Mood unadj	.000	.749	.776	.751	.088	.973	.977	.970	.761	1.000	1.000	1.000
	SR omean	.000	.802	.824	.815	.221	.981	.986	.978	.854	1.000	1.000	1.000
	SR omedian	.000	.706	.743	.727	.191	.920	.929	.921	.695	.991	.992	.994
	TG omean	.000	.830	.848	.840	.281	.985	.988	.983	.887	1.000	1.000	1.000
	TG omedian	.000	.825	.837	.835	.322	.972	.979	.973	.830	.999	1.000	1.000
-3 -3 3 3	FAB unadj	.000	.021	.028	.022	.000	.023	.034	.023	.000	.026	.030	.019
	FK omean	.000	.119	.125	.108	.000	.224	.264	.235	.000	.537	.588	.542
	FK omedian	.000	.271	.309	.285	.008	.408	.414	.444	.064	.568	.599	.598
	Klotz unadj	.000	.014	.020	.014	.000	.022	.024	.021	.000	.020	.030	.019
	Mood unadj	.000	.019	.023	.025	.000	.019	.032	.024	.000	.023	.027	.019
	SR omean	.000	.140	.171	.147	.000	.337	.379	.344	.002	.704	.742	.702
	SR omedian	.000	.303	.350	.357	.020	.463	.488	.495	.114	.640	.670	.673
	TG omean	.000	.278	.334	.306	.000	.666	.708	.661	.035	.981	.982	.974
	TG omedian	.000	.378	.427	.407	.018	.601	.617	.610	.176	.821	.839	.822

Table 4.10: Performance with the Lognormal distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are (1 1) and (1 1 1 1) for 2 and 4 treatments respectively.

Lognormal distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
1 1													
0 0	FAB unadj	.01	.050	.056	.053	.007	.056	.066	.047	.011	.033	.068	.057
	FK omean	.007	.038	.057	.052	.006	.055	.055	.046	.005	.043	.059	.053
	FK omedian	.001	.048	.049	.046	.004	.050	.045	.047	.009	.049	.049	.059
	Klotz unadj	.005	.045	.052	.048	.008	.042	.063	.040	.004	.054	.054	.049
	Mood unadj	.006	.040	.055	.066	.003	.053	.063	.041	.003	.037	.065	.046
	SR omean	.005	.040	.051	.055	.006	.056	.059	.044	.009	.065	.072	.049
	SR omedian	.004	.044	.072	.062	.007	.048	.062	.043	.008	.052	.069	.063
	TG omean	.003	.047	.060	.065	.005	.060	.062	.045	.009	.053	.054	.045
	TG omedian	.002	.040	.043	.057	.010	.052	.053	.048	.005	.050	.066	.058
-3 3	FAB unadj	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.021	.029	.023	.003	.035	.039	.033	.015	.112	.127	.092
	FK omedian	.015	.194	.202	.193	.090	.268	.343	.271	.223	.453	.510	.503
	Klotz unadj	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.002	.023	.036	.038	.006	.065	.083	.050	.023	.166	.216	.145
	SR omedian	.083	.211	.214	.228	.147	.362	.367	.325	.294	.550	.580	.543
	TG omean	.004	.029	.024	.027	.007	.052	.099	.058	.025	.178	.201	.122
	TG omedian	.009	.128	.162	.157	.085	.283	.333	.317	.267	.527	.497	.532
1 1 1 1													
0 0 0 0	FAB unadj	.000	.058	.055	.067	.000	.042	.063	.047	.000	.059	.073	.048
	FK omean	.000	.044	.049	.057	.000	.039	.066	.048	.000	.051	.051	.050
	FK omedian	.000	.039	.050	.049	.000	.043	.048	.035	.000	.053	.064	.052
	Klotz unadj	.000	.036	.043	.038	.000	.037	.057	.047	.000	.059	.059	.052
	Mood unadj	.000	.053	.058	.062	.000	.047	.058	.044	.000	.046	.072	.056
	SR omean	.000	.044	.060	.036	.000	.058	.051	.052	.000	.050	.055	.045
	SR omedian	.000	.043	.053	.034	.000	.041	.062	.047	.000	.045	.049	.035
	TG omean	.000	.037	.058	.040	.000	.049	.051	.049	.000	.049	.073	.060
	TG omedian	.000	.032	.066	.047	.000	.042	.054	.051	.000	.053	.053	.045
-3 -3 3 3	FAB unadj	.000	.021	.031	.020	.000	.029	.038	.027	.000	.023	.025	.019
	FK omean	.000	.067	.094	.082	.000	.126	.147	.150	.000	.323	.365	.318
	FK omedian	.000	.568	.615	.581	.075	.841	.837	.829	.448	.972	.984	.982
	Klotz unadj	.000	.016	.022	.016	.000	.023	.022	.014	.000	.029	.026	.014
	Mood unadj	.000	.014	.020	.020	.000	.030	.021	.027	.000	.023	.032	.026
	SR omean	.000	.109	.125	.101	.000	.240	.261	.230	.001	.581	.667	.589
	SR omedian	.000	.614	.661	.674	.121	.852	.871	.870	.594	.989	.983	.990
	TG omean	.000	.119	.134	.136	.000	.237	.270	.229	.001	.645	.666	.648
	TG omedian	.000	.681	.684	.673	.090	.885	.905	.922	.644	.995	.997	.991

Klotz unadj and Mood unadj, have good power when there is no change in the mean, but, they are sensitive to the change in mean.

When the sample is 4, the tests with the better behavior are FK omean, FK omedian, SR omean, SR omedian, TG omean and TG omedian for almost all the statistics except χ_F^2. The tests FAB unadj, Klotz unadj and Mood unadj have good power when there is no change in the mean, but, are sensitive to the change in mean.

Gamma distribution null hypothesis

Table 4.12 shows the results for the Gamma distribution when there are no changes in the standard deviation.

When there is no change in the mean, all the tests for two and four samples are robust in all the different size of blocks. If there is a change in the mean with a sample size of two the tests FK omedian, SR omedian, and TG omedian are sensitive in all the size of blocks. With an increase in the sample size to four the tests are sensitive to change in the mean, except FAB unadj, Klotz unadj, and Mood unadj. The other tests are sensitives to a change in the mean, and when the size of the block increases.

Gamma distribution power

Table 4.13 shows the results of the power in the Gamma distribution when there is a change in the standard deviation.

No changes in mean When there are 2 samples all the tests have good power, the tests that have the best behavior is SR omedian, FAB unadj and Mood unadj, for 2, 4, and 8 blocks respectively. When the sample increases to 4, the statistic χ_F tends to be more conservative when there are 2 blocks; the most powerful tests are Mood unadj for 2 blocks, Klotz unadj and Mood unadj for 4 and 8 blocks the in F_{one}.

Change in the mean For the case of 2 samples, the tests that remain robust even if there are changes in mean are FK omedian, SR omedian, for almost all the statistics except the χ_F^2; and the best test is SR omean for all the blocks. For the other side, the tests FAB unadj, FK omean, Klotz unadj,

Table 4.11: Performance with the Lognormal distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are [(1 2), (1 3), and (1 4)] and [(1 1 2 2), (1 1 3 3), and (1 1 4 4)] for 2 and 4 treatments respectively.

Lognormal distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
Average Power (1 2), (1 3), (1 4)													
0 0	FAB unadj	.234	.503	.509	.548	.541	.785	.848	.781	.892	.971	.979	.972
	FK omean	.193	.472	.503	.486	.507	.772	.792	.768	.846	.950	.963	.956
	FK omedian	.209	.514	.531	.523	.556	.755	.784	.761	.835	.932	.938	.927
	Klotz unadj	.157	.448	.457	.458	.484	.764	.807	.757	.870	.968	.979	.971
	Mood unadj	.157	.487	.493	.495	.522	.796	.815	.791	.889	.971	.983	.971
	SR omean	.206	.504	.533	.504	.521	.788	.816	.791	.879	.960	.969	.962
	SR omedian	.220	.517	.550	.530	.548	.771	.784	.772	.841	.935	.934	.936
	TG omean	.181	.480	.535	.509	.535	.776	.816	.780	.876	.957	.968	.959
	TG omedian	.245	.533	.583	.559	.601	.806	.817	.801	.865	.953	.951	.954
-3 3	FAB unadj	.000	.001	.000	.001	.000	.000	.001	.000	.000	.000	.000	.009
	FK omean	.000	.008	.012	.009	.000	.012	.018	.012	.002	.025	.035	.037
	FK omedian	.004	.114	.127	.120	.030	.152	.168	.150	.088	.241	.252	.194
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.013
	SR omean	.000	.009	.012	.009	.000	.014	.017	.014	.002	.031	.042	.044
	SR omedian	.024	.137	.149	.135	.052	.167	.208	.197	.107	.272	.299	.290
	TG omean	.001	.028	.044	.028	.006	.079	.117	.061	.045	.240	.299	.204
	TG omedian	.004	.056	.091	.074	.014	.110	.119	.097	.040	.145	.156	.173
Average Power (1 1 2 2), (1 1 3 3), (1 1 4 4)													
0 0 0 0	FAB unadj	.000	.757	.793	.788	.133	.968	.971	.974	.782	1.000	1.000	.999
	FK omean	.000	.700	.742	.718	.113	.930	.944	.933	.684	.995	.997	.996
	FK omedian	.000	.655	.681	.677	.105	.874	.896	.886	.617	.982	.985	.979
	Klotz unadj	.000	.684	.718	.696	.023	.950	.960	.951	.629	.999	1.000	.999
	Mood unadj	.000	.741	.776	.763	.114	.968	.979	.973	.781	1.000	1.000	1.000
	SR omean	.000	.739	.776	.744	.182	.949	.954	.953	.744	.999	.999	.999
	SR omedian	.000	.682	.694	.692	.138	.891	.897	.896	.623	.983	.990	.980
	TG omean	.000	.731	.755	.741	.145	.944	.946	.950	.711	.999	.998	.997
	TG omedian	.000	.710	.742	.720	.186	.913	.923	.929	.682	.994	.992	.994
-3 -3 3 3	FAB unadj	.000	.018	.031	.022	.000	.026	.029	.017	.000	.026	.034	.026
	FK omean	.000	.048	.072	.061	.000	.086	.095	.082	.000	.159	.176	.141
	FK omedian	.000	.171	.193	.200	.001	.198	.225	.248	.004	.288	.291	.312
	Klotz unadj	.000	.014	.022	.015	.000	.022	.028	.018	.000	.023	.031	.018
	Mood unadj	.000	.021	.027	.016	.000	.023	.029	.018	.000	.021	.031	.023
	SR omean	.000	.062	.078	.068	.000	.102	.125	.113	.000	.211	.247	.205
	SR omedian	.000	.210	.227	.229	.002	.229	.270	.263	.015	.321	.325	.343
	TG omean	.000	.114	.165	.139	.000	.334	.383	.319	.001	.757	.782	.735
	TG omedian	.000	.188	.214	.208	.004	.237	.275	.275	.020	.340	.364	.353

Table 4.12: Performance with the Gamma distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are (1 1) and (1 1 1 1) for 2 and 4 treatments respectively.

Gamma distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
1 1													
0 0	FAB unadj	.010	.060	.048	.054	.011	.038	.070	.047	.009	.038	.049	.042
	FK omean	.004	.042	.056	.032	.011	.046	.058	.052	.009	.067	.061	.042
	FK omedian	.007	.049	.046	.052	.004	.052	.064	.046	.010	.044	.066	.051
	Klotz unadj	.003	.052	.056	.039	.005	.052	.059	.047	.008	.052	.064	.045
	Mood unadj	.002	.059	.056	.051	.008	.052	.046	.049	.010	.047	.070	.054
	SR omean	.002	.053	.058	.054	.006	.053	.050	.043	.008	.045	.067	.064
	SR omedian	.007	.046	.060	.048	.005	.067	.065	.047	.008	.042	.060	.055
	TG omean	.005	.041	.061	.052	.010	.044	.047	.062	.009	.043	.059	.045
	TG omedian	.002	.043	.055	.060	.007	.036	.048	.050	.004	.046	.069	.052
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.001	.002	.001	.000	.007	.005	.003	.000	.007	.017	.006
	FK omedian	.003	.149	.164	.141	.058	.219	.254	.212	.144	.391	.418	.408
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.001	.003	.007	.002	.000	.015	.009	.006	.002	.027	.029	.016
	SR omedian	.039	.176	.171	.168	.080	.294	.305	.274	.228	.431	.514	.462
	TG omean	.001	.002	.004	.005	.000	.003	.006	.003	.001	.012	.027	.019
	TG omedian	.002	.101	.139	.085	.043	.230	.242	.202	.146	.418	.414	.391
1 1 1 1													
0 0 0 0	FAB unadj	.000	.043	.052	.049	.000	.049	.055	.059	.000	.055	.055	.059
	FK omean	.000	.025	.059	.053	.000	.043	.050	.057	.000	.058	.064	.038
	FK omedian	.000	.038	.044	.044	.000	.036	.056	.057	.000	.057	.064	.046
	Klotz unadj	.000	.050	.051	.039	.000	.041	.055	.053	.000	.043	.066	.045
	Mood unadj	.000	.040	.056	.055	.000	.050	.066	.052	.000	.041	.074	.042
	SR omean	.000	.049	.057	.057	.000	.044	.055	.048	.000	.053	.054	.051
	SR omedian	.000	.044	.054	.046	.000	.047	.048	.034	.000	.041	.059	.036
	TG omean	.000	.039	.061	.047	.000	.046	.060	.056	.000	.061	.059	.047
	TG omedian	.000	.046	.056	.069	.000	.042	.062	.059	.000	.065	.050	.041
-3 -3 3 3	FAB unadj	.000	.022	.029	.027	.000	.022	.021	.019	.000	.022	.028	.018
	FK omean	.000	.033	.034	.035	.000	.030	.043	.041	.000	.072	.068	.053
	FK omedian	.000	.473	.526	.541	.020	.740	.725	.766	.270	.946	.957	.957
	Klotz unadj	.000	.015	.019	.012	.000	.023	.018	.012	.000	.015	.028	.017
	Mood unadj	.000	.018	.023	.014	.000	.021	.021	.018	.000	.022	.024	.025
	SR omean	.000	.034	.052	.048	.000	.063	.060	.060	.000	.132	.192	.127
	SR omedian	.000	.539	.571	.589	.052	.784	.818	.806	.392	.966	.978	.969
	TG omean	.000	.038	.048	.030	.000	.059	.073	.060	.000	.157	.181	.131
	TG omedian	.000	.548	.575	.557	.032	.840	.839	.819	.367	.985	.980	.991

Mood unadj and SR omean they have good power when there is no change in mean, but, they are sensitive to the change in mean.

When the sample is 4, the test with the better behavior FK omedian, SR omedian, TG omean and TG omedian for almost all the statistics except χ^2_F. SR omedian is the most potent test in all the blocks. The tests FAB unadj, FK omean, Klotz unadj, Mood unadj and SR omedian have good power when there is no change in mean, but, are sensitive to the change in mean.

Weibull distribution null hypothesis

Table 4.15 shows the results for the Weibull distribution when there are no changes in the standard deviation.

When there is no change in the mean, all the tests for two and four samples are robust in all the different size of blocks. If there is a change in the mean with a sample size of two or four the tests FK omedian, SR omedian, and TG omedian are sensitive for all the size of the block.

No changes in mean When there are no changes in mean, and the sample size is 2, the best statistic is the F_{one}; the test FK omean, FK omedian, SR omean, SR omedian and TG omean, have a robust behavior with all the cases of size block. The χ^2_F statistics are conservative for all the size of blocks.

For the sample size 4, Mood unadj, SR omedian, and TG overall

Weibull distribution power

Table 4.16 shows the results of the power in the Weibull distribution when there is a change in the standard deviation.

No changes in mean When there are 2 samples all the tests have good power, the tests that have the best behavior is SR omedian, FAB unadj and Mood unadj, for 2, 4, and 8 blocks respectively. When the sample increases to 4, the statistic χ_F tends to be more conservative when there are 2 blocks; the most powerful tests are Mood unadj for 2 blocks, Klotz unadj for 4 and almost all the test in the 8 blocks case, this is for the F_{one}. the principal idea is

Table 4.13: Performance with the Gamma distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are [(1 2), (1 3), and (1 4)] and [(1 1 2 2), (1 1 3 3), and (1 1 4 4)] for 2 and 4 treatments respectively.

Gamma distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
Average Power (1 2), (1 3), (1 4)													
0 0	FAB unadj	.252	.511	.516	.543	.539	.791	.851	.793	.888	.968	.968	.966
	FK omean	.214	.517	.539	.532	.560	.797	.821	.804	.872	.959	.954	.955
	FK omedian	.214	.547	.566	.551	.571	.781	.806	.785	.847	.943	.951	.934
	Klotz unadj	.178	.501	.509	.508	.520	.798	.823	.787	.897	.973	.976	.973
	Mood unadj	.188	.516	.542	.515	.559	.819	.849	.820	.897	.974	.980	.973
	SR omean	.230	.542	.570	.541	.584	.817	.834	.812	.891	.964	.973	.969
	SR omedian	.243	.555	.589	.541	.582	.802	.802	.790	.841	.940	.945	.936
	TG omean	.190	.519	.562	.535	.538	.790	.817	.787	.867	.955	.958	.957
	TG omedian	.252	.535	.569	.556	.577	.799	.833	.799	.858	.943	.946	.943
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.000	.001	.000	.000	.000	.001	.001	.000	.003	.002	.001
	FK omedian	.000	.132	.140	.128	.053	.237	.253	.215	.179	.395	.424	.387
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.000	.001	.001	.001	.000	.000	.002	.002	.000	.002	.003	.001
	SR omedian	.024	.143	.173	.143	.081	.259	.279	.259	.216	.427	.447	.427
	TG omean	.000	.006	.013	.007	.000	.020	.032	.012	.003	.063	.086	.057
	TG omedian	.000	.052	.085	.049	.021	.131	.162	.110	.075	.234	.252	.238
Average Power (1 1 2 2), (1 1 3 3), (1 1 4 4)													
0 0 0 0	FAB unadj	.000	.745	.779	.756	.153	.956	.969	.964	.763	1.000	.999	.998
	FK omean	.000	.723	.769	.731	.152	.936	.951	.942	.708	.997	.998	.996
	FK omedian	.000	.695	.725	.706	.142	.899	.909	.890	.648	.990	.989	.988
	Klotz unadj	.000	.724	.773	.748	.058	.971	.978	.968	.726	.998	1.000	.999
	Mood unadj	.000	.780	.793	.791	.147	.971	.978	.970	.792	1.000	1.000	1.000
	SR omean	.000	.756	.783	.764	.190	.949	.961	.953	.749	.998	.998	.998
	SR omedian	.000	.701	.727	.714	.169	.899	.921	.907	.658	.986	.992	.988
	TG omean	.000	.734	.753	.743	.155	.927	.949	.939	.709	.997	.997	.996
	TG omedian	.000	.717	.739	.722	.175	.910	.928	.916	.679	.994	.994	.990
-3 -3 3 3	FAB unadj	.000	.023	.029	.019	.000	.027	.033	.023	.000	.025	.030	.020
	FK omean	.000	.029	.048	.038	.000	.045	.048	.047	.000	.049	.057	.040
	FK omedian	.000	.168	.177	.171	.000	.249	.272	.276	.009	.405	.423	.414
	Klotz unadj	.000	.014	.023	.013	.000	.021	.021	.015	.000	.022	.030	.018
	Mood unadj	.000	.022	.029	.015	.000	.027	.033	.020	.000	.020	.030	.020
	SR omean	.000	.037	.046	.037	.000	.041	.049	.044	.000	.045	.057	.039
	SR omedian	.000	.188	.212	.209	.000	.272	.315	.316	.013	.421	.445	.449
	TG omean	.000	.067	.068	.059	.000	.101	.130	.118	.000	.282	.353	.290
	TG omedian	.000	.129	.155	.148	.000	.188	.200	.186	.004	.276	.309	.287

Change in the mean For the case of 2 samples, the tests that remain robust even if there are changes in mean are FK omedian, SR omedian and TG omedian for almost all the statistics except the χ_F^2; and the best test is SR omedian for all the blocks. For the other side, the tests FAB unadj, FK omean, Klotz unadj, Mood unadj, SR omean and TG omean have good power when there is no change in mean, but, they are sensitive to the change in mean.

When the sample is 4, the test with the better behavior FK omedian, SR omedian, TG omean and TG omedian for almost all the statistics except χ_F^2. SR omedian is the most potent test in all the blocks. The tests FAB unadj, FK omean, Klotz unadj, Mood unadj and SR omedian have good power when there is no change in mean, but, are sensitive to the change in mean.

4.3 Discussion

The following list is a summary of findings:

Regarding the robustness of the tests:

1. Table 4.1 shows the tests distribution robust when there is no change in the location and scale.
2. The tests FAB unad, Mood unadj, and Klotz unadj, are robust to change in location. These results were consistent in all evaluated scenarios.
3. The tests FK omedian, SR omedian, and TG omedian are sensitive for both location and scale changes in every scenario.
4. TG omean, FK omean, and SR omean were sensitive to changes in location only in some scenarios.

This behavior under the null distribution is constant over different number of blocks, treatments and test statistic [1].

Regarding the power of the tests:

1. FK omedian, SR omedian, and TG omedian have a consistent high power over all scenarios when changes occurred in location or location and scale.

[1] Technically each combination of score type-alignment-statistic is a different test statistic, but for simplicity in this research, a test statistic is called from each of the statistics in section 3.2.2 and the combination is called "test"

2. FK omean, SR omean, and TG omean showed exceptional levels of power in some scenarios, where TG omean presented the highest power when dealing with lognormal and DoubleExp2 and Normal2 only. However their have poor behavior for Gamma, Weibull and symmetric distributions.

3. In average, the test statistics F_{one} has the high power for all the scenarios.

4. χ_F had the lowest power of all test statistics.

5. Even though F_{one} and F_{two} have high levels of power, some unfortunate permutations of ranks make the $MSE = 0$ and the statistics cannot be calculated. This rarely happens, but it happens.

6. When there is a location change FAB unadj, Mood unadj, and Klotz unadj have almost no power.

7. When there is no change in location, only in scale, all test show high levels of power over test statistics χ_k, F_{one}, and F_{two}. Within this test the ones that shows are relatively higher power where SR omedian and TG omedian.

Figure 4.1 shows a comparison between the power of the four tests statistics, in general, the plot indicates:

- even if the replicate size is small, the F_{one} statistics has better behavior in comparison with the other.

- when the size of replicates increases to 20 the average power of the χ^2_K, F_{one}, and F_{two}, are similar in the three cases.

- the χ^2_F has the lowest power.

In this figure the axis "Y" is the average power and the axis "X" is the number of replicates per cell.

Figure 4.2 is a plot of the nine tests that are compared in this section when there is no change in the location, using the test statistic F_{one}. The "Y" axis represents the average power, and the "X" axis is the n size of replicate. The graphic shows that for all the numbers of replicate the tests have similar behavior.

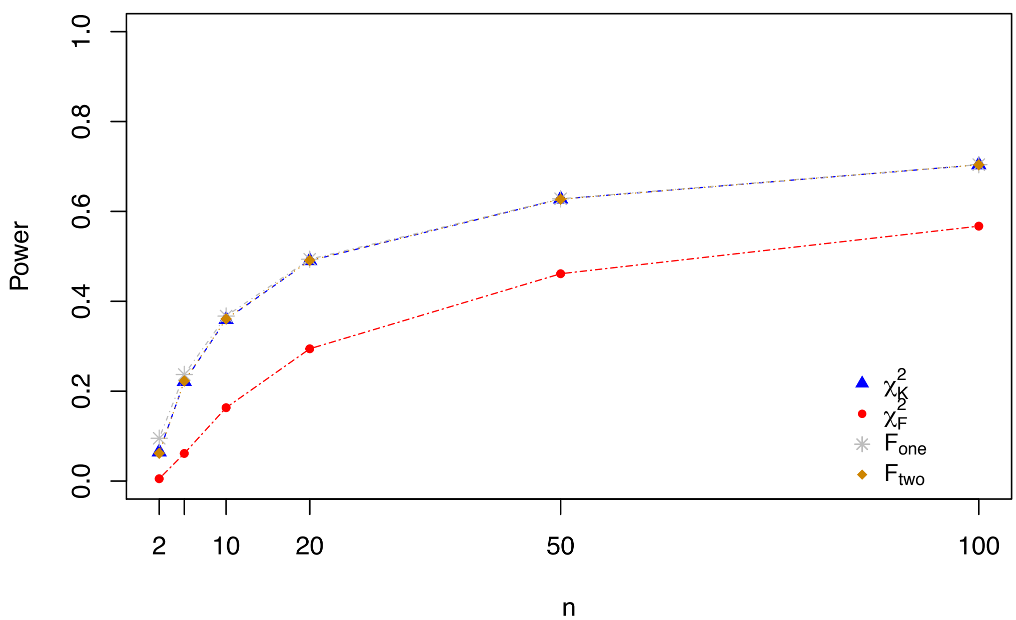

Figure 4.1: A comparison between the average power of the test statistics χ^2_F, χ^2_K, F_{one}, and F_{two} dealing with different number of replicates (2,10,20,50, and 100).

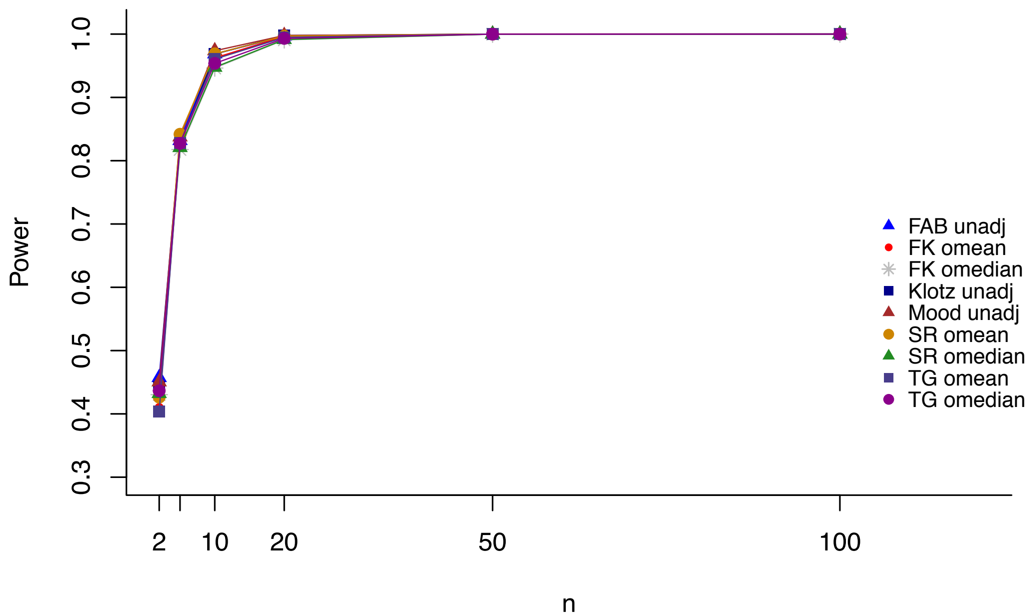

Figure 4.2: Comparison between the average power of the tests for homoscedasticity with the F_{one} test statistic, dealing with different number of replicates (2, 5, 10, 20, and 100) when there are no changes in the location for 2 and 4 treatment and 2,4, and 8 blocks.

In Figure 4.3 we have the same conditions than the previous figure the nine tests are evaluated using the test statistic F_{one}, but in this case, there are changes in the location [(-1 1) (-2 2) (-3 3)]. The three tests with the best behavior are FK omedian, SR omedian, and TG omedian in all the numbers of replicate.

From this findings, a general recommendation can be provided: *If there is no certainty that there won't be a change in the mean between treatments, which is usually the case in practice, FK omedian, SR omedian, and TG omedian are the recommended tests when use with the* F_{one} *test statistic. In the event the* $MSE = 0$ *when evaluating the* F_{one}*, making the test statistic useless, practitioner should change to the* χ^2_K*, since the latter presents high levels of power, and can be always calculated.*

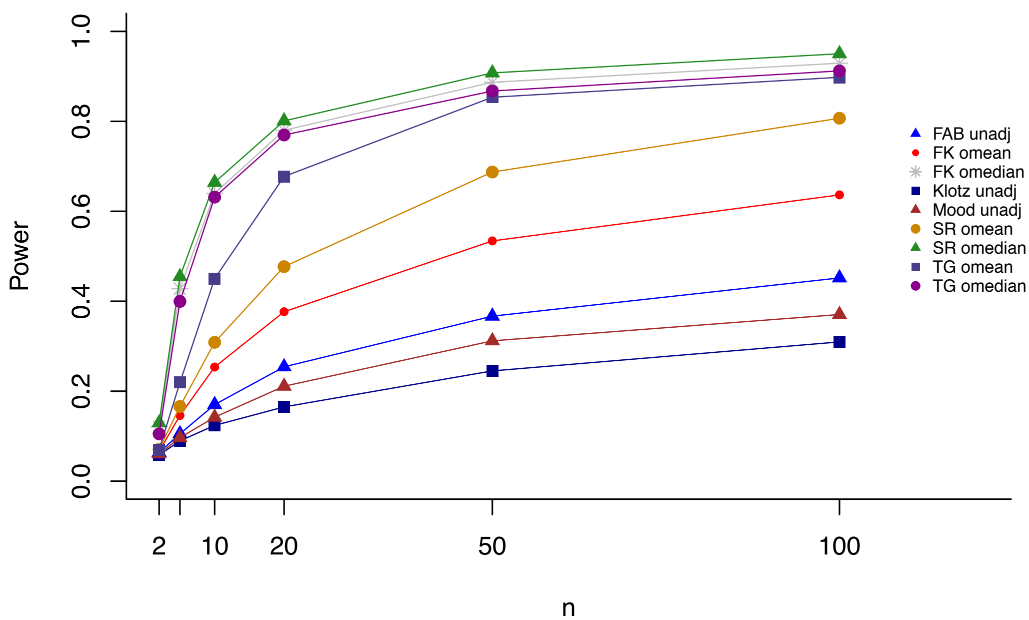

Figure 4.3: Comparison between the average power of the tests for homoscedasticity with the F_{one} test statistic, dealing with different number of replicates (2, 5, 10, 20, and 100) when there are changes in the location [(-1 1), (-2 2), (-3 3)], and [(-1 -1 1 1), (-2 -2 2 2), (-3 -3 3 3)] for 2 and 4 treatment and 2,4, and 8 blocks. Treatment standard deviations are (1 2), (1 3), (1 4)] and [(1 1 2 2), (1 1 3 3), (1 1 4 4)] for 2 and 4 treatments respectively

Table 4.14: Performance with the Weibull distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are (1 1) and (1 1 1 1) for 2 and 4 treatments respectively.

Weibull distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
1 1													
0 0	FAB unadj	.011	.044	.041	.053	.006	.034	.075	.039	.013	.044	.065	.050
	FK omean	.003	.042	.061	.045	.001	.049	.074	.044	.008	.066	.058	.056
	FK omedian	.005	.056	.056	.055	.008	.050	.071	.046	.004	.035	.063	.067
	Klotz unadj	.003	.052	.058	.064	.009	.052	.058	.056	.006	.050	.048	.047
	Mood unadj	.006	.045	.041	.034	.007	.047	.061	.045	.005	.051	.053	.053
	SR omean	.005	.054	.061	.051	.006	.047	.070	.044	.016	.053	.060	.049
	SR omedian	.005	.036	.056	.048	.005	.040	.062	.052	.006	.057	.050	.057
	TG omean	.008	.027	.060	.053	.005	.057	.055	.060	.007	.049	.052	.051
	TG omedian	.005	.062	.049	.050	.007	.055	.070	.056	.012	.043	.077	.054
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.003	.003	.005	.000	.004	.004	.001	.000	.005	.010	.012
	FK omedian	.002	.128	.164	.145	.057	.238	.272	.239	.164	.374	.439	.392
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.000	.005	.005	.008	.002	.007	.012	.005	.000	.019	.040	.017
	SR omedian	.043	.185	.178	.171	.068	.287	.307	.277	.208	.452	.481	.464
	TG omean	.000	.005	.005	.005	.000	.003	.009	.004	.000	.018	.022	.013
	TG omedian	.005	.077	.133	.104	.045	.221	.243	.210	.150	.407	.446	.412
1 1 1 1													
0 0 0 0	FAB unadj	.000	.046	.054	.044	.000	.046	.047	.047	.000	.048	.056	.043
	FK omean	.000	.043	.045	.044	.000	.055	.050	.048	.000	.035	.050	.054
	FK omedian	.000	.042	.054	.040	.000	.060	.045	.056	.000	.047	.048	.063
	Klotz unadj	.000	.036	.046	.047	.000	.041	.062	.046	.000	.050	.047	.035
	Mood unadj	.000	.048	.054	.052	.000	.047	.051	.042	.000	.045	.052	.046
	SR omean	.000	.039	.049	.050	.000	.054	.066	.047	.000	.038	.045	.044
	SR omedian	.000	.038	.054	.053	.000	.052	.056	.072	.000	.052	.054	.044
	TG omean	.000	.042	.059	.051	.000	.042	.070	.043	.000	.050	.051	.058
	TG omedian	.000	.036	.055	.046	.000	.042	.047	.050	.000	.045	.049	.052
-3 -3 3 3	FAB unadj	.000	.019	.024	.024	.000	.018	.027	.025	.000	.034	.029	.018
	FK omean	.000	.034	.030	.030	.000	.044	.052	.038	.000	.072	.065	.051
	FK omedian	.000	.489	.526	.513	.022	.720	.764	.752	.271	.948	.953	.954
	Klotz unadj	.000	.015	.018	.022	.000	.019	.033	.016	.000	.025	.019	.023
	Mood unadj	.000	.015	.042	.023	.000	.021	.032	.017	.000	.025	.033	.021
	SR omean	.000	.032	.052	.041	.000	.061	.084	.066	.000	.160	.165	.150
	SR omedian	.000	.513	.553	.575	.074	.816	.814	.820	.375	.972	.971	.969
	TG omean	.000	.028	.039	.045	.000	.071	.084	.068	.000	.151	.162	.133
	TG omedian	.000	.565	.580	.573	.034	.838	.855	.841	.399	.978	.984	.987

Table 4.15: Weibull distribution - null hypothesis.

Table 4.16: Performance with the Weibull distribution of different test homoscedasticity using statistics χ^2_F, χ^2_K, F_{one}, and F_{two} with 2,4, and 8 blocks when dealing with no changes in the mean of 2 and 4 treatments [(0 0) and (0 0 0 0), respectively], and changes in mean between 2 and 4 treatments [(-3 3) and (-3 -3 3 3), respectively]. All mean changes are measured in standard deviations of the null distribution. Treatment standard deviations are [(1 2), (1 3), and (1 4)] and [(1 1 2 2), (1 1 3 3), and (1 1 4 4)] for 2 and 4 treatments respectively.

Weibull distribution													
Mean	**Test**	**2 Blocks**				**4 Blocks**				**8 Blocks**			
		χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}	χ^2_F	χ^2_K	F_{one}	F_{two}
Average Power (1 2), (1 3), (1 4)													
0 0	FAB unadj	.262	.521	.517	.521	.574	.801	.847	.793	.888	.959	.977	.964
	FK omean	.224	.523	.541	.521	.546	.801	.818	.789	.866	.961	.963	.961
	FK omedian	.230	.543	.573	.555	.571	.785	.795	.790	.840	.938	.952	.934
	Klotz unadj	.179	.492	.512	.506	.526	.807	.831	.816	.896	.969	.974	.969
	Mood unadj	.188	.524	.533	.535	.534	.817	.845	.810	.894	.975	.984	.970
	SR omean	.223	.539	.562	.544	.565	.819	.828	.817	.885	.966	.973	.959
	SR omedian	.239	.540	.587	.541	.591	.797	.823	.803	.838	.937	.942	.945
	TG omean	.209	.484	.560	.541	.538	.778	.825	.791	.863	.947	.960	.957
	TG omedian	.239	.545	.575	.555	.587	.799	.828	.789	.850	.948	.950	.947
-3 3	FAB unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	FK omean	.000	.002	.001	.000	.000	.001	.001	.001	.000	.001	.003	.000
	FK omedian	.000	.121	.143	.125	.047	.225	.248	.212	.185	.385	.420	.385
	Klotz unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Mood unadj	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	SR omean	.000	.001	.001	.001	.000	.002	.002	.001	.000	.002	.003	.001
	SR omedian	.029	.147	.171	.152	.070	.241	.278	.257	.219	.416	.440	.413
	TG omean	.000	.007	.008	.006	.000	.019	.035	.014	.002	.065	.079	.050
	TG omedian	.002	.058	.077	.053	.017	.135	.152	.122	.077	.231	.252	.226
Average Power (1 1 2 2), (1 1 3 3), (1 1 4 4)													
0 0 0 0	FAB unadj	.000	.750	.784	.759	.163	.959	.959	.967	.749	.999	1.000	.999
	FK omean	.000	.725	.758	.733	.147	.934	.946	.939	.707	.997	.997	.997
	FK omedian	.000	.699	.713	.713	.138	.889	.904	.896	.645	.987	.992	.987
	Klotz unadj	.000	.723	.761	.745	.050	.958	.981	.964	.718	.999	1.000	1.000
	Mood unadj	.000	.784	.798	.787	.146	.970	.978	.968	.793	1.000	1.000	1.000
	SR omean	.000	.757	.792	.775	.199	.950	.959	.950	.756	.999	.998	.997
	SR omedian	.000	.708	.725	.709	.177	.898	.911	.900	.662	.992	.990	.987
	TG omean	.000	.726	.743	.743	.179	.929	.945	.940	.711	.997	.996	.997
	TG omedian	.000	.716	.738	.722	.184	.918	.925	.920	.688	.995	.994	.994
-3 -3 3 3	FAB unadj	.000	.020	.032	.027	.000	.021	.027	.020	.000	.022	.028	.019
	FK omean	.000	.039	.043	.039	.000	.042	.052	.038	.000	.042	.054	.043
	FK omedian	.000	.156	.184	.172	.000	.256	.289	.273	.007	.399	.416	.418
	Klotz unadj	.000	.015	.016	.012	.000	.024	.030	.021	.000	.021	.024	.018
	Mood unadj	.000	.018	.032	.021	.000	.026	.028	.018	.000	.028	.029	.015
	SR omean	.000	.032	.044	.038	.000	.040	.055	.043	.000	.055	.051	.039
	SR omedian	.000	.177	.228	.203	.001	.286	.326	.306	.013	.421	.443	.434
	TG omean	.000	.055	.073	.064	.000	.117	.125	.114	.000	.310	.345	.291
	TG omedian	.000	.123	.142	.135	.000	.175	.198	.192	.001	.282	.306	.287

Chapter 5

Example

From (Box et al., 2008) a study to combat the effect of toxic agents was presented. The results of the study are shown in Table 5.1 where the responses are the survival time of four animal groups (treatments) with 3 different toxic substances (blocks).

Statistical model used is :

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk} \begin{cases} i = 1, 2, ...a \\ j = 1, 2, ...b \\ k = 1, 2, ...n \end{cases}$$

We compute the scores for each test as is shown in Table 5.2, with these we proceed to make the respective calculus for the TG overall median test,

Table 5.1: Survival times of the animals in a factorial experiment. Example of agents.

Toxic	Treatment			
	A	B	C	D
I	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
II	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
III	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

Table 5.2: Obtain the individual scores for each test, using the overall median alignment over the residuals of the survival time in the four groups of animals (treatments), and the three different toxic agents (block).

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Block	**Treatment**	y_{ijk}	$\bar{y}_i$	$\bar{y}_{i..} - y_{ijk}$ (e_{ijk})	$\tilde{y}_{i..}$	$(e_{ijk}) - \tilde{y}_{i..}$	**rank**	**SR**	**TG**	**FK**
1	A	0.31	0.62	-0.31	0.01	-0.32	15	225	15	1.56472647
	A	0.45	0.62	-0.17	0.01	-0.18	9	81	9	0.72152228
	A	0.46	0.62	-0.16	0.01	-0.17	7	49	7	0.54139509
	A	0.43	0.62	-0.19	0.01	-0.20	12	144	12	1.0491314
	B	0.82	0.62	0.20	0.01	0.20	12	144	12	1.0491314
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	D	0.33	0.28	0.05	-0.01	0.06	11.5	132.25	11.5	0.98723099

where the scores of the column (10) are represented with an "x" as follows:

$$\begin{aligned} TSS &= \sum_{i=1} \sum_{j=1} \sum_{k=1} x_{ijk}^2 - \frac{x_{...}^2}{N} \\ &= 1006.50 \end{aligned}$$

$$\begin{aligned} SST &= \frac{1}{b} \sum_{i} \sum_{j} x_{i.}^2 - \frac{x_{..}^2}{N} \\ &= 172.125 \end{aligned}$$

The F_{one} obtained is:

$$\begin{aligned} F_{one} &= \frac{\text{SST}/(a-1)}{(\text{TSS} - \text{SST})/(nba - a)} \\ &= \frac{172.125/(4-1)}{(1006.50 - 172.125)/(4 * 3 * 4 - 4)} \\ &= 3.0896 \end{aligned}$$

With an F distribution with $a - 1$ and $(n * b * a - a)$ degrees of freedom. The respective p - value is 0.0366, so H_0 is rejected at a level of significance α = 0.05. We can conclude that there is a difference between the variations in the different treatments. Table 5.3 shows the results for the three tests SR, TG, and FK aligned with the overall median, it can be seen that the TSS corresponding for each test are 309146.6, 1006.50, and 12.51 respectively,

Test	TSS	SST	DF1	DF2	F_{one}	p-value
SR	309146.6	62966.92	3	36	3.7513	0.0174
TG	1006.50	172.125	3	36	3.0896	0.0366
FK	12.5157	2.6068	3	36	3.8584	0.0155

Table 5.3: Summary of results for the tests TG, SR, and FK aligned with the overall median.

them have an effect due to the treatment (SST) of 62966.92, 172.125, and 2.60 for each test. The degrees of freedom for the numerator is 3 (DF1) and for the denominator of 36 (DF2) for all the tests. Finally the corresponding p-value are 0.0174, 0.0366, and 0.0155 these results are consist.

Chapter 6

Conclusions

In the manufacturing industry it is a common practice that the process engineers want to improve their processes to reduce the variability, there are some cases when the presence of a nuisance factor is detected. When this factor can be controlled this type of design is known as RCBD. In these cases, the main objective is to reduce the at the minimum the error within the blocks. The main problem dealing with nonparametric statistical methods is that the actual techniques have poor behavior when different situations are presented such as the nonnormality of the residuals, the presence of changes in the mean, when there is a change in mean and variance at the same time.

The importance to develop a nonparametric statistical test for homoscedasticity in RCB was found those who have the best behavior to deal with the different situations presented.

The linear ranks transformations were used to develop an extension of the Friedman test, and four statistics test were used two that approximate an F distribution and two that approximate a χ^2 under the null distribution.

The experimentation was made with 1000 simulations per scenario, in each combination which was chosen from 6 number of replicates, 2 number of treatments, 3 different block size, 2 symmetric distributions, and 5 skewed distributions, with varying effects in the treatment and change in the standard deviation. In total 96 tests were evaluated for all the combinations.

An initial assessment was made to obtain robust tests when there are no changes in the standard deviation. Also, we evaluated situations where there is a change in the mean but not in the standard deviation.

An interest fact was found in the first selection of the 96 test, the scores-alignments sample mean and sample median (see 3.2 eliminates robustness from the test, while the score-alignments that subtract the overallmean and overall median of the block keeps robustness.

Now, the research questions are retaken with their obtained conclusion:

1. The test statistic with the average higher power in all the scenarios is F_{one}

2. The tests statistics FAB unadj, Mood unadj, and Klotz unadj are robust when there is also a change in the mean.

3. FK omedian, SR omedian, and TG omedian are sensitive for location and scale changes.

4. In regard to the power within robust test for location:

   (a) The tests are robust to change in mean, but their have a poor power level in all the scenarios..

5. In regard to the power within tests sensitive to scale and location changes:

   (a) In average, SR omedian, and TG omedian have higher power levels in almost all the scenarios.

.

Bibliography

Angel Gutierrez, J. C. (1996). The variation and its significance. *Revista Universidad EAFIT*.

Ansari, A. R., Bradley, R. A., et al. (1960). Rank-sum tests for dispersions. *The Annals of Mathematical Statistics*, 31(4):1174–1189.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 160(901):268–282.

Bhandary, M. and Dai, H. (2008). An alternative test for the equality of variances for several populations when the underlying distributions are normal. *Communications in Statistics-Simulation and Computation*, 38(1):109–117.

Bhandary, M. and Dai, H. (2013). An alternative test for the equality of variances for several populations in randomized complete block design. *Statistical Methodology*, 11:22–35.

Box, G. E. (1953). Non-normality and tests on variances. *Biometrika*, 40(3/4):318–335.

Box, G. E., Hunter, J. S., and Hunter, W. G. (2008). *Estadística para investigadores: Diseño, innovación y descubrimiento*. Reverté.

Conover, W. J. (1999). *Practical Nonparametric Statistics*. John Wiley & Sons, 3rd edition.

Conover, W. J., Guerrero-Serrano, A. J., and Tercero-Gómez, V. G. (2018). An update on 'a comparative study of tests for homogeneity of variance'. *Journal of Statistical Computation and Simulation*, 88(8):1454–1469.

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, 23(4):351–361.

Fligner, M. A. and Killeen, T. J. (1976). Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, 71(353):210–213.

Freund, J. E. and Ansari, A. R. (1957). The two-way rank sum test for variances. *Polytechnic Institute*, 34.

Gill, J. (1984). Heterogeneity of variance in randomized block experiments. *Journal of Animal Science*, 59(5):1339–1344.

Hollander, M., Wolfe, D. A., and Chicken, E. (2013). *Nonparametric statistical methods*, volume 751. John Wiley and Sons.

Klotz, J. et al. (1962). Nonparametric tests for scale. *The Annals of Mathematical Statistics*, 33(2):498–512.

Kutner, M., Nachtsheim, C., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill Irwin, 5th edition.

Montgomery, D. (2008). *Design and analysis of experiments*. Limusa Willey.

Mood, A. M. et al. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *The Annals of Mathematical Statistics*, 25(3):514–522.

Nordstokke, D. W. and Zumbo, B. D. (2010). A new nonparametric levene test for equal variances. *Psicologica: International Journal of Methodology and Experimental Psychology*, 31(2):401–430.

Talwar, P. P. and Gentle, J. E. (1977). A robust test for the homogeneity of scales. *Communications in Statistics-Theory and Methods*, 6(4):363–369.