A COMPARISON STUDY ON SELECTED TESTS ON HOMOGENEITY OF VARIANCES

SALIZAH SAMAD

MASTER OF SCIENCE (APPLIED STATISTICS)
UNIVERSITI PUTRA MALAYSIA
JULY 2007

A COMPARISON STUDY ON SELECTED TESTS ON HOMOGENEITY OF VARIANCES

BY

SALIZAH SAMAD

A PROJECT SUBMITTED TO

FACULTY OF SCIENCE

UNIVERSITI PUTRA MALAYSIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR

MASTER OF SCIENCE (APPLIED STATISTICS)

JULY 2007

CERTIFICATION OF SUPERVISOR

This project paper entitled

A COMPARISON STUDY ON SELECTED TESTS ON HOMOGENEITY OF VARIANCES

was submitted by

SALIZAH SAMAD
(GS 15132)

In partial fulfillment of the requirement for
Master of Science (Applied Statistics).

Certified by :

(PUAN FAUZIAH MAAROF)
Project Supervisor
July 2007

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
UNIVERSITI PUTRA MALAYSIA
SERDANG, SELANGOR

ACKNOWLEDGEMENT

In the name of Allah, the Most Beneficent, Most Merciful

First and foremost, I would like to express my sincere appreciation to my supervisor, Puan Fauziah Maarof, for her inspiring guidance, advice and cooperation which enable this project to materialize.

My gratitude is also extended to the other lecturers of the Mathematics Department of UPM, my fellow coursemates and friends, who have made the duration of my studies a colourful and fruitful experience.

I am also grateful to my sponsor, MARA, who gave me the opportunity to widen my horizons in Applied Statistics.

Last but not least, I acknowledge the support of my mother and family members for their love and encouragement. Truly, the acquisition of knowledge is a lifelong process and along with every hardship there is also joy.

ABSTRACT

The main purpose of this project is to choose the best test among the three selected tests on homogeneity of variances for small sample sizes from Normal distributions. The three tests considered in this study are Bartlett's, Hartley's and Breusch-Pagan tests which are based on sample variances.

SAS (9.1) and S-Plus programs are used in the generation of random samples, computing the test statistic and determination of the estimated type I error and power of the tests. Comparisons are carried out among the tests, based on the estimated type I error and the estimated power, for number of samples of five and ten. In this study, the relationship between the estimated power of tests and number of unequal sample variances is also investigated.

Project findings indicate that Breusch-Pagan is the best, since it portrays the most powerful test compared to Hartley's and Bartlett's. However, power increases most rapidly for Hartley's test. Furthermore, the power of tests increases when total number of observations and number of unequal sample variances increase. The recommendation is that, the Breusch-Pagan is the best among the three tests for small sample sizes from Normal distributions.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

The assumption of homogeneity of variances, also known as homoscedasticity , is one of the critical assumptions underlying most parametric statistical procedures such as analysis of variance, ANOVA.

Testing for homogeneity of variances is important for at least two purposes:

1. The homogeneity of variances reflects upon the quality of the variable of interest, especially for quality improvement.
2. In making an appropriate comparison of group means, we often make the assumption that the within-group variation is constant (approximately similar) , thus requiring a diagnosis of homogeneity of variances.

Three tests on homogeneity of population variances considered in this study are Bartlett's, Hartley's and Breusch–Pagan tests that are based on sample variances and non-complexity. Simulation was conducted to compute the type I error and power of tests in order to perform a statistical comparison among those tests.

1.2 Objective

The objectives of this project are :

1. Compute and compare the probability of type I error and the power of tests among Bartlett's, Hartley's and Breusch-Pagan tests.

2. Find the best test among those tests on homogeneity of variances considered in this project, for small sample sizes from the normal distribution.

1.3 Project Organization

This project paper consists of several chapters, all of which explains the relevant topics required to report the study. The following describes the organization of the project paper :

CHAPTER 1 - Contains a brief introduction to the main features in this project. It presents the objectives of the project.

CHAPTER 2 - Gives an introduction and the literature review of the selected tests on homogeneity of variances.

Describes the size and power of tests and presents the factors that affect the power.

CHAPTER 3 - Explains the methodology used in this project, which is the selected tests on homogeneity of variances (Bartlett's, Hartley's and Breusch-Pagan tests) and simulation methods.

CHAPTER 4 - Focuses on the results and the comparison study of the tests. It discusses the conclusions acquired from the project and recommendation that can be derived from the simulation results.

CHAPTER 5 - Gives summary and suggestions for further research.

CHAPTER 2

TESTS ON HOMOGENEITY OF VARIANCES

2.1 Introduction

Several tests of homogeneity of variances have been proposed in the literature between 1930's to 1970's (Bartlett, 1937 ; Cochran, 1941, 1951 ; Hartley, 1950 ; Box, 1953 ; Levene, 1960 ; Breusch and Pagan, 1979). Natural selection has left us with only a few that are presented in current textbooks. Many authors claim that a test on homogeneity of variances is a prerequisite to analysis of variance and other parametric statistical procedures. The tests of homogeneity of population variances considered in this study are Bartlett's, Hartley's and Breusch-Pagan tests.

2.2 Literature Review

Hartley's Test (Hartley,1950) is the simplest test of homogeneity of variances, which uses the ratio of the largest to the smallest sample variances.

Bartlett's test is one of the most often presented in textbooks and taught in introductory courses because of its ease of computation.

The test statistic involves a comparison of the separate within-group sums-of-squares to the pooled within-group sum-of-squares.

Bartlett's test is known to be powerful if the sampled populations are normal, but badly affected by non-normality (Box 1953, Zar 1999). Pierre and Daniel (1972) who did a simulation study on several tests of homogeneity of variances found that one of the best overall methods is Bartlett's Test.

Breusch-Pagan Test (Breusch and Pagan, 1979) is a large-sample test, which assumes that the error terms of the regression model are independent and normally distributed. The variance of the error term is assumed to be related to the level of X in the following way :

$$\log_e \sigma_i^2 = \gamma_0 + \gamma_1 X_i$$

2.3 Size and Power of tests

The quantities used in this comparison study are the probability of type I error and power of tests of homogeneity of variances. It is important to understand the two quantities in order to choose the best performance among the three tests.

The size of a test, often called the significance level, is the probability of type I error and is usually denoted by α. A type I error occurs when the null hypothesis is rejected when it is true. This test size is denoted by alpha, α . It is the extent of the risk of making wrong conclusions. A 0.05 level means that we are taking a risk of being wrong five times per 100 trials.

The power of a statistical test is the probability that it will correctly lead to the rejection of a false null hypothesis (Greene, 2000). The statistical power is the ability of a test to detect an effect, if the effect actually exists (High, 2000). Power is denoted by ($1 - \beta$), where β is the probability of type II error, that is, the probability of failing to reject the null hypothesis when it is false. (Refer to Table 3.1.1). Notationally,

$$\alpha = P(\text{type I error}) = P(\text{Reject } H_O / H_O \text{ true})$$

and $$\beta = P(\text{type II error}) = P(\text{fail to reject } H_O / H_O \text{ false}) .$$

Hence,

$$\text{Power} = 1 - \beta = 1 - P(\text{fail to reject } H_O / H_O \text{ false})$$

$$= P(\text{reject } H_O / H_O \text{ false}) .$$

Thus, an experimenter would want to use the most powerful test, which is one with maximum power. Table 2.3.1 show the events leading to the occurances of each of the error types and their probabilities.

Table 2.3.1 : Size and Power of tests.

	Do not reject H_O	Reject H_O
H_O is true	Correct Decision $(1 - \alpha)$ = confidence level	Type I error α = size of tests = significance level = P(Type I error)
H_O is false	Type II error β = P(Type II error)	Correct Decision $(1 - \beta)$ = Power of tests

2.3.1 Factors Affecting Power.

The power of a hypothesis test is affected by several factors.

1. **Sample size.**

   Increasing sample size makes the hypothesis test more sensitive, that is more likely to reject the null hypothesis when it is false. Thus, increase the power of the test.

2. **Size of the difference between population means.**

The size of the difference between population means is an important factor in determining power. Naturally, the more the means differ from each other, the easier it is to detect the difference. Hence, the smaller the probability of type II error (β) , the larger the power.

3. **Significance level.**

The higher the significance level, the higher the power of the test. If we increase the significance level, we reduce the region of acceptance. As a result, we are more likely to reject the null hypothesis. This means that it is less likely to accept the null hypothesis when it is false, that is, less likely to make a type II error. Hence, the power of the test is increased.

4. **Distribution of a population.**

One other factor that may affect the power is normality of the distribution. Deviations from the assumption of normality usually lower the power of a test.

5. **Type of statistical procedure.**

The type of statistical procedure used may also affect the power of a test. Some of the distribution-free tests are less powerful than other tests when the distribution is normal but more powerful when the distribution is highly-skewed. One-tailed tests are more powerful than two-tailed tests as long as the effect is in the expected direction. Otherwise, their power is zero.

6. **Choice of an experimental design.**

The choice of an experimental design can have a profound effect on power. Within-subject designs are usually much more powerful than between-subject designs.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter details the procedures conducted in this study, performed on uncontaminated data set. The procedures carried out are presented in the following schematic diagrams.

Figure 3.1.1 : Schematic Diagram of SAS and S-Plus procedures to determine the estimated type I error.

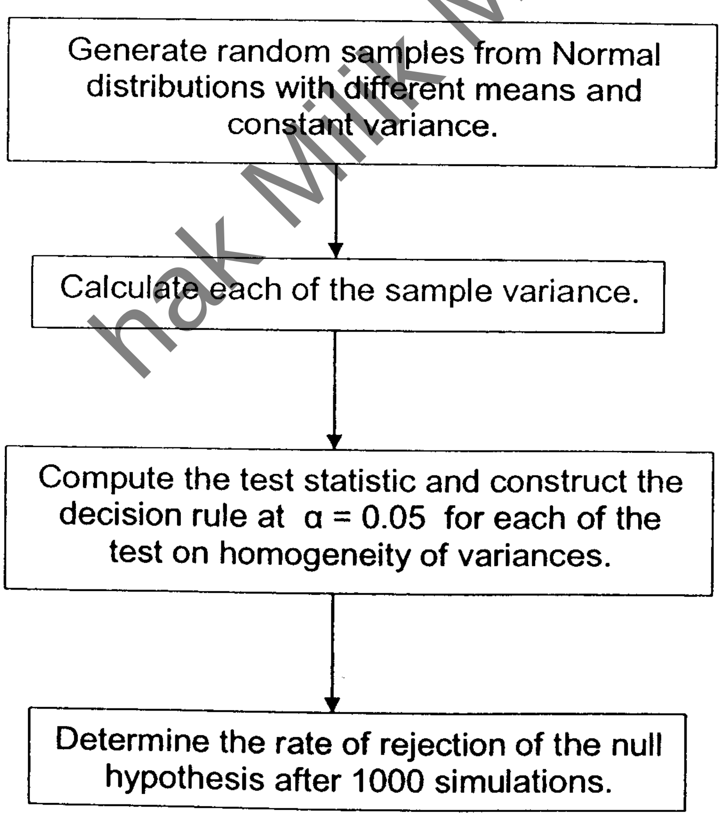
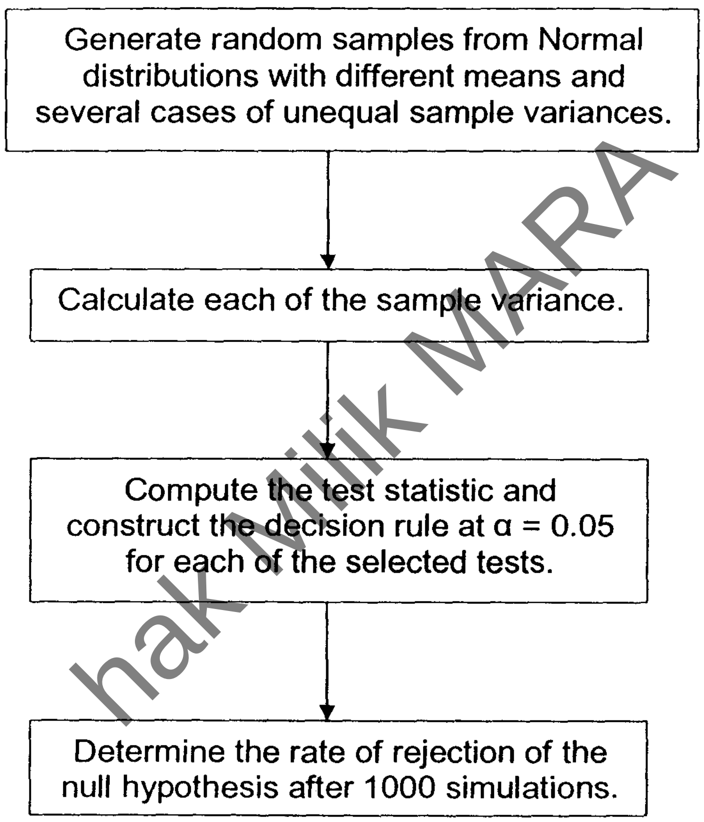


Figure 3.1.2 : Schematic Diagram of SAS and S-Plus procedures to determine the estimated power of tests on homogeneity of variances.



3.2 Selected Tests on Homogeneity of Variances.

The set of hypotheses to be tested is :

H o: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2$; a = number of random samples.

H 1: not all the variances are equal.

The critical values are determined at α = 0.05 .

3.2.1 Hartley's Test (Hartley, 1950) .

A : Equal Sample Sizes Across Factor Levels (all n i = n).

1. Calculate the sample variances s_i^2 for each of the 'a' factor levels (samples).

2. Compute **Hartley's Fmax test statistic** :

$$F_{\max} = \frac{largest\ s_i^2}{smallest\ s_i^2} = \frac{s_{\max}^2}{s_{\min}^2}$$

3. Find the critical value F(α, a, n -1) from F-max table.

4. If $F_{\max} = \dfrac{s_{\max}^2}{s_{\min}^2}$ > F(α, a, n -1) , **reject Ho** and conclude that the variances are not all equal.

3.2 Selected Tests on Homogeneity of Variances.

The set of hypotheses to be tested is :

H o: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2$; a = number of random samples.

H 1: not all the variances are equal.

The critical values are determined at α = 0.05 .

3.2.1 Hartley's Test (Hartley, 1950) .

A : Equal Sample Sizes Across Factor Levels (all $n_i = n$).

1. Calculate the sample variances s_i^2 for each of the 'a' factor levels (samples).
2. Compute **Hartley's Fmax test statistic** :

$$F_{\max} = \frac{largest\ s_i^2}{smallest\ s_i^2} = \frac{s_{\max}^2}{s_{\min}^2}$$

3. Find the critical value F(α, a, n -1) from F-max table.
4. If $F_{\max} = \dfrac{s_{\max}^2}{s_{\min}^2}$ > F(α, a, n -1) , **reject Ho** and

   conclude that the variances are not all equal.

B : Unequal Sample Sizes Across Factor Levels.

- When the number of observations (n_i 's) for each factor level are not all equal, but the n_i's are relatively close, the largest of the sample sizes n_{max} may be substituted for n when determining the critical value $F(\alpha, a, n_{max} - 1)$.
- This procedure leads to a slight positive bias in the test. That is, it will reject H_O more frequently than should be the case.

3.2.2 Bartlett's Test (Douglas C. Montgomery ; 2000)

1. **Test statistic:** $\chi_0^2 = 2.3026\frac{q}{c}$

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_i^2 = sample variance of the i th. Population.

$$s_p^2 = \frac{\sum_{i=1}^{a}(n_i - 1)s_i^2}{(N-a)}$$

2. **Reject Ho if** $\chi_0^2 > \chi_{(\alpha,a-1)}^2$

3.2.3 Breusch-Pagan Test (Kutner, Nachtsheim, Netter ; 2004)

It is assumed that σ_i^2 is related to X_i as follows :

$$[\log_e \sigma_i^2 = \gamma_0 + \gamma_1 X_i]$$

1. Test : $Ho : \gamma_1 = 0$ **(=> σ_i^2 is constant)**

$H_1 : \gamma_1 \neq 0$

2. Test statistic : $X_{BP}^2 = \dfrac{SSR^* / 2}{\left(SSE / n\right)^2}$;

$SSR^* \Rightarrow$ Regressed e_i^2 vs X_i

3. Reject Ho if $X_{BP}^2 > \chi^2_{(1-\alpha,\nu)}$

3.3 Simulation Methods

- The computer programs were written in SAS (9.1) and S-Plus to carry out the 1000 independent simulations for the selected tests of homogeneity of variances. The programs were designed to compute the three test statistics described in section 3.2 and compared with their respective critical values at α = 0.05 . Random data were generated from normal distributions with different means and several cases of equality of variances.

- In the study of probability of **type I error**, data were generated in such a way that the null hypothesis was true (that is, the population variances are equal). The rate of type I error was computed as the proportion of the simulations whereby the null hypothesis was rejected at the 5 % significance level.

$$P(\text{type I error}) = P(\text{reject } H_O / H_O \text{ is true}) = \alpha .$$

We would like a test to have rate of rejection close to 0.05 since the alpha level used is 0.05 .

- In the **power** study, data were simulated in such a way that the null hypothesis was false (that is, not all population variances are equal). Power was computed as the proportion of the simulations whereby the null hypothesis was rejected at the 5 % significance level.

$$\text{Power of test} = 1 - \beta = P(\text{reject Ho / Ho is false}) .$$

Hence, we would like a test to have high power.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Estimated type I error and power of tests

The observed rate of rejection after 1000 simulations are reported in Table 4.1.1 , Table 4.1.2 and Table 4.1.3 . Samples are generated from the normal distribution, with different means and several cases of sample variances.

Table 4.1.1: The estimated type I error for Hartley's, Bartlett's and Breusch-Pagan tests .

case	trt	obs	Normal (μ different, $\sigma_i^2 = 1$) Rate of rejection (estimated type I error)		
			Hartley	Bartlett	B.Pagan
1	5	5	0.053	0.054	0.042
2	5	10	0.046	0.034	0.052
3	5	15	0.03	0.028	0.034
4	5	20	0.051	0.046	0.049
5	10	5	0.064	0.06	0.059
6	10	10	0.038	0.023	0.034
7	10	15	0.03	0.034	0.041
8	10	20	0.031	0.039	0.05

Table 4.1.1 contains the simulated estimated type I error for Hartley's, Bartlett's and Breusch-Pagan tests with equal sample variances , which is $\sigma_i^2 = 1$. The number of treatments or samples is five and ten with number of observations of 5, 10, 15 and 20 respectively.

Table 4.1.2: The estimated power of the three tests when number of samples equals to five.

case	trt	obs	Normal (different μ_i)	Rate of rejection (Est. Power)		
				Hartley	Bartlett	B.Pagan
1	5	5	$\sigma_1^2 \neq \sigma_2^2$	0.393	0.572	0.796
			$\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$	0.577	0.741	0.871
			$\sigma_1^2 \neq .. \neq \sigma_4^2$	0.64	0.673	0.746
2	5	10	$\sigma_1^2 \neq \sigma_2^2$	0.888	0.939	0.984
			$\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$	0.979	0.99	0.995
			$\sigma_1^2 \neq .. \neq \sigma_4^2$	0.99	0.988	0.988
3	5	15	$\sigma_1^2 \neq \sigma_2^2$	0.983	0.996	1
			$\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$	0.999	1	1
			$\sigma_1^2 \neq .. \neq \sigma_4^2$	1	1	1
4	5	20	$\sigma_1^2 \neq \sigma_2^2$	0.996	1	1
			$\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$	1	1	1
			$\sigma_1^2 \neq .. \neq \sigma_4^2$	1	1	1

Table 4.1.2 presents the simulated estimated power for the three selected tests. The number of samples for each case is **five** and the number of observations are 5, 10, 15 and 20 respectively with several cases of unequal sample variances.

Table 4.1.3 : The estimated power of the three tests when number of samples equals to ten.

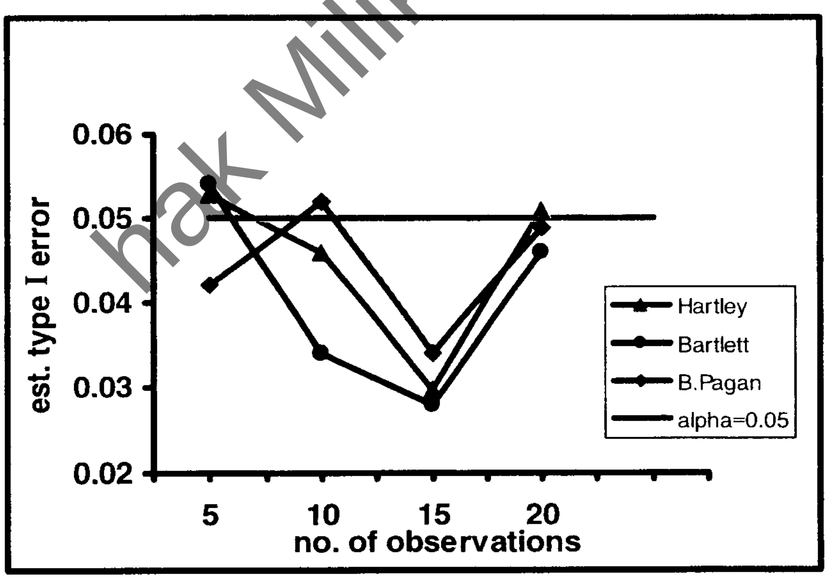
case	trt	obs	Normal (diff. μ_i)	Rate of rejection (Est. Power)		
				Hartley	Bartlett	B.Pagan
1	10	5	$\sigma_1^2 \neq \sigma_2^2$	0.435	0.665	0.87
			$\sigma_1^2 \neq .. \neq \sigma_5^2$	0.921	0.994	1
			$\sigma_1^2 \neq .. \neq \sigma_8^2$	0.97	0.995	0.998
2	10	10	$\sigma_1^2 \neq \sigma_2^2$	0.886	0.97	0.996
			$\sigma_1^2 \neq .. \neq \sigma_5^2$	1	1	1
			$\sigma_1^2 \neq .. \neq \sigma_8^2$	1	1	1
3	10	15	$\sigma_1^2 \neq \sigma_2^2$	0.99	0.997	1
			$\sigma_1^2 \neq .. \neq \sigma_5^2$	1	1	1
			$\sigma_1^2 \neq .. \neq \sigma_8^2$	1	1	1
4	10	20	$\sigma_1^2 \neq \sigma_2^2$	0.999	1	1
			$\sigma_1^2 \neq .. \neq \sigma_5^2$	1	1	1
			$\sigma_1^2 \neq .. \neq \sigma_8^2$	1	1	1

Table 4.1.3 provides the simulated results of estimated power for the three selected tests. The number of samples for each case is **ten** and the number of observations are 5, 10, 15 and 20 respectively with several cases of unequal sample variances.

4.2 Discussion on Summarized Results

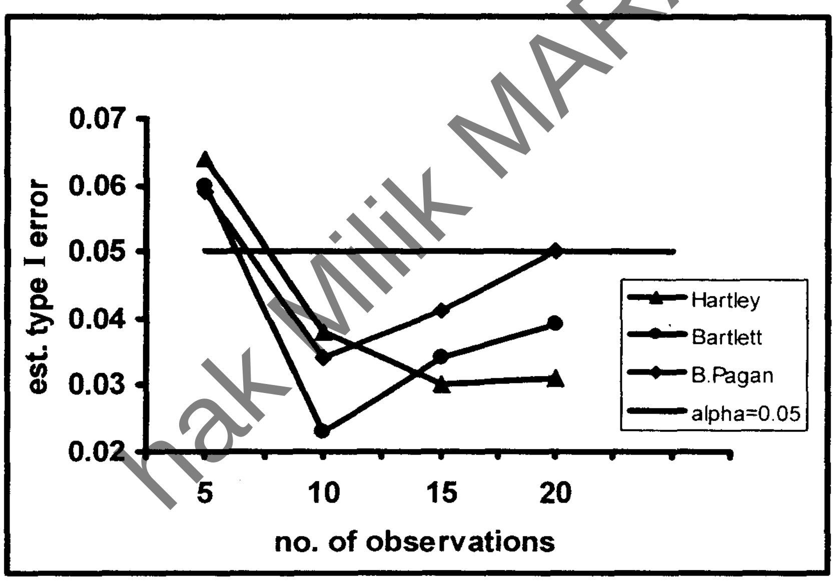
All the tabulated results are summarized graphically in the following figures.

Figure 4.2.1: Plot of estimated type I error vs. number of observations. (For number of samples = 5)



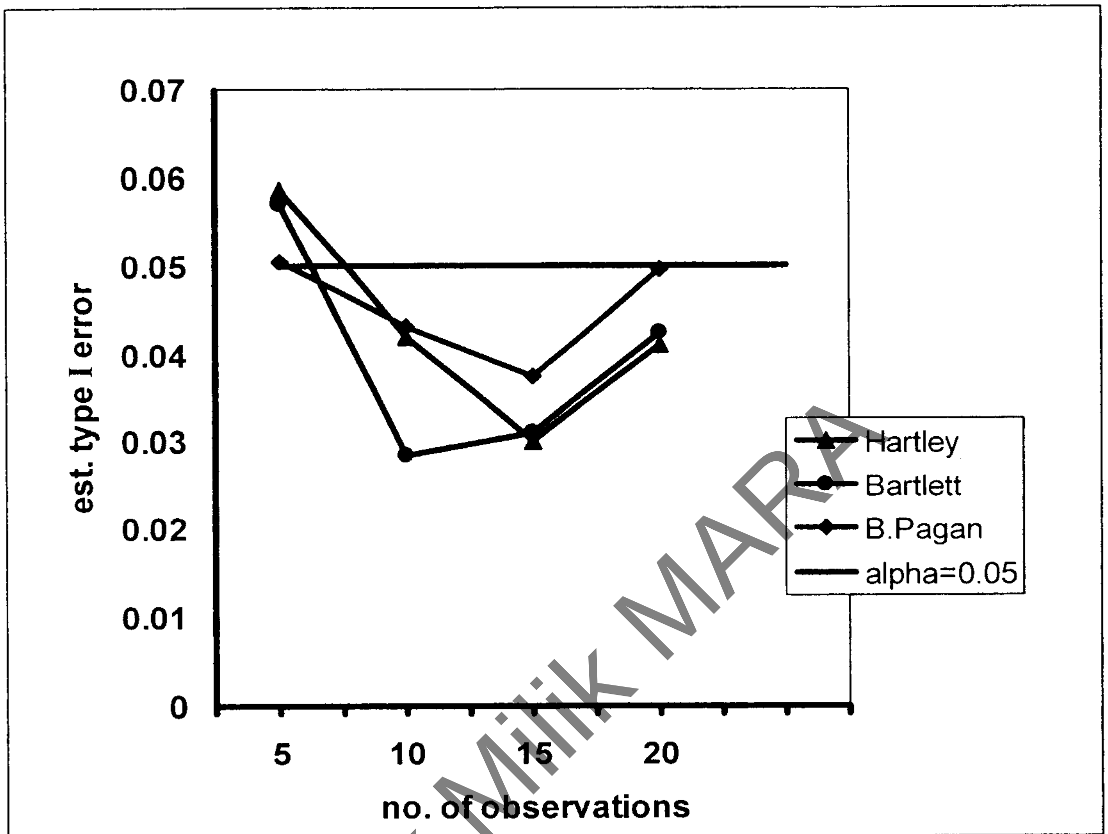
The plots in Figure 4.2.1 shows that Hartley's and Bartlett's tests better than Breusch-Pagan test for sample of five with five observations for each sample. However, as the total number of observations increases, Breusch-Pagan exhibits the best test since the points have the shortest distance from the horizontal line α = 0.05.

Figure 4.2.2 : Plot of estimated type I error vs. number of observations. (For number of samples = 10)



Plot in Figure 4.2.2 portrays clearly that Breusch-Pagan is the best test, as the number of observations increases for number of samples of ten. The plot also displays that Hartley's test is less appropriate for large total number of observations.

Figure 4.2.3 : Plot of estimated type I error vs. number of observations.
(Increase both number of samples and observations)



By taking the average values of the estimated type I error in Table 4.1.1 for both samples of five and ten, we obtained the plot in Figure 4.2.3. The plot confirms that Breusch-Pagan is the best test compared to the other two tests since the points fall closely around the horizontal line $\alpha = 0.05$.

Figure 4.2.4 : Plot of estimated power vs. number of observations.
(Number of samples = 5)

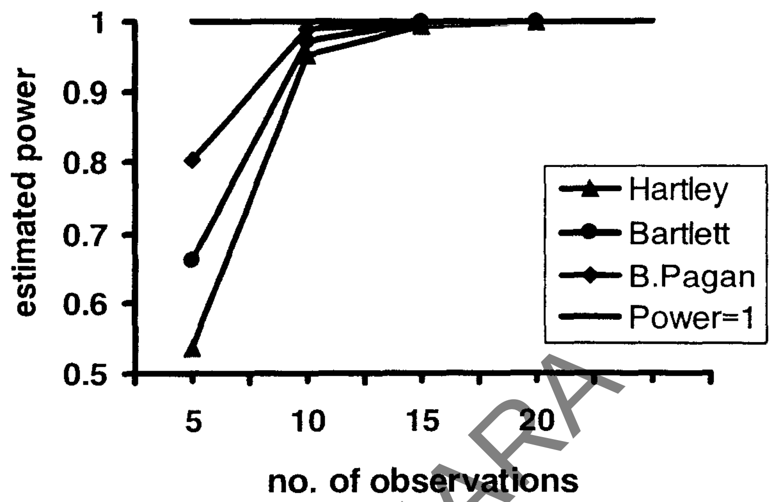
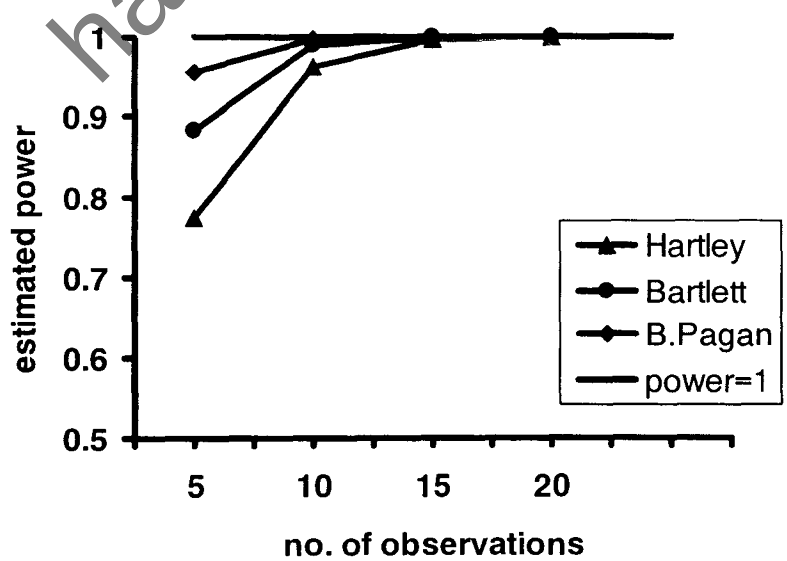
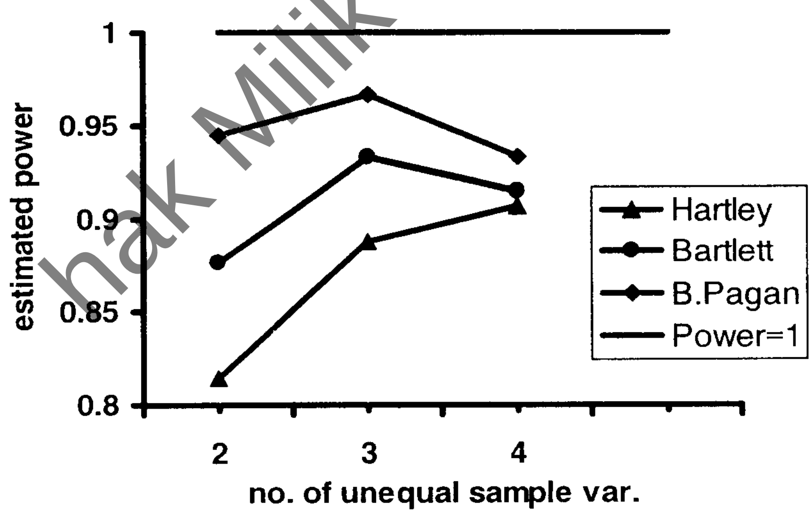


Figure 4.2.5 : Plot of estimated power vs. number of observations.
(Number of samples = 10)



Both plots in Figure 4.2.4 and Figure 4.2.5 show the same pattern of estimated power versus number of observations for number of samples of five and ten. Breusch-Pagan test appears to be more powerful than the other two tests since the points on both plots are closer to the horizontal line at power = 1. On the other hand, it is also observed that the power increase most rapidly for Hartley's test. Moreover, both plots illustrate a great improvement of the power of tests when the total number of observations is increased.

Figure 4.2.6 : Plot of estimated power vs. number of unequal sample variances. (Number of samples = 5)



The plot of estimated power versus number of unequal sample variances for number of samples of **five** in Figure 4.2.6 also indicates that the best test is Breusch-Pagan. It also points out that the estimated power decreases for Bartlett's and Breusch-Pagan when the number of unequal sample variances is four. This is so because when all the four samples, out of five, have unequal variances, it means that all samples have unequal variances. It is also noticeable that the power of test is still increasing for Hartley's.

Figure 4.2.7 : Plot of estimated power vs. number of unequal sample variances. (Number of samples = 10)

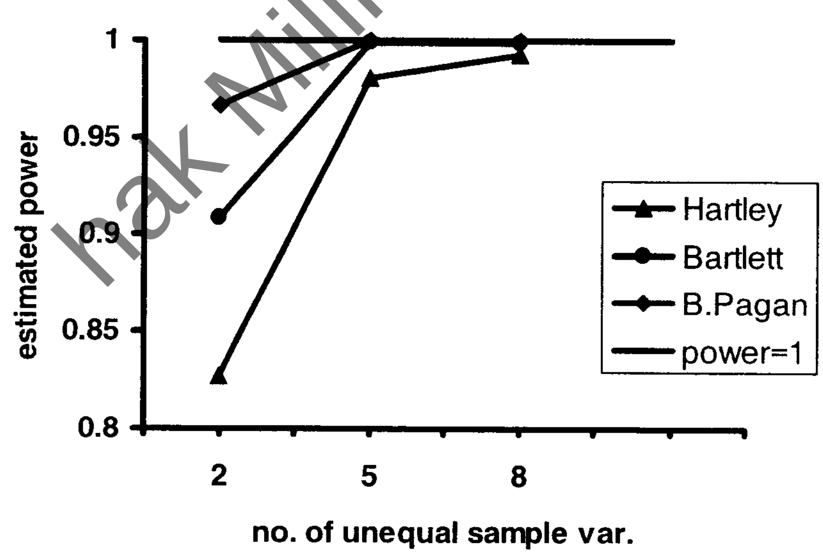


Figure 4.2.7 contains the plot of estimated power versus number of unequal sample variances for number of samples of **ten**. This plot illustrates the most powerful test is Breusch-Pagan since it has nearest points nearest to the horizontal line , power = 1. In addition, it can be seen from the plot that the points for Bartlett's and Breusch-Pagan tests almost reach the maximum power = 1 when five out of ten samples have unequal variances. Furthermore, the plot also reveals that the power of tests increases when the number of unequal variances increases.

4.3 Conclusions and Recommendation.

The conclusions that can be drawn from the figures in section 4.2 are as follows :

1. [Refer to Figure 4.2.1 and Figure 4.2.2]

   Breusch-Pagan is the best test since the points on the plots (that represent the estimated type I error values) exhibit the shortest distance from the horizontal line $\alpha = 0.05$, as the number of observations increases for each sample.

2. [Refer to Figure 4.2.3]

   If both number of samples and number of observations increase, the plot confirms that Breusch-Pagan is the best test compared to

Hartley's and Bartlett's. (The points represent the average values of the estimated type I error)

3 [Refer to Figure 4.2.4 and Figure 4.2.5]

In general, the power of tests improves greatly when the number of observations increases. Among the three tests, Breusch-Pagan test appears to have more power than the other two tests since the points close to the horizontal line at power = 1. On the other hand, if one observe the estimated power values and the plots, it can be concluded that the power increase most rapidly for Hartley's test.

[Refer to Figure 4.2.5 and Figure 4.2.7]

Furthermore, we can suggest that Breusch-Pagan and Bartlett's tests could be used for large number of observations.

5 [Refer to Figure 4.2.6 and Figure 4.2.7]

Besides that, the plots also reveal that the power of tests increases when the number of unequal sample variances increases.

6. Overall comparisons between the estimated type I error and the power of tests suggest that Breusch- Pagan is the best test.

Therefore, in the light of these conclusions, it is recommended that Breusch-Pagan test is the best among the three tests when data are Normal and sample sizes are small.

CHAPTER 5

SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

5.1 Summary

The project findings may be summarized as follows :

- From the study of estimated type I error and power of tests, Breusch-Pagan is suggested as the best test compared to Bartlett's and Hartley's tests for number of samples of five and ten.

- Generally, the power of tests improves greatly when the number of observations increases. Among the three selected tests, Breusch-Pagan appears to be the most powerful test as the total number of observations increases.

- When number of unequal sample variances increases, so does the power of tests.

5.2 Suggestions for Further Research

The following are suggestions for further research related to this project :

1. In future comparison studies, include other tests of homogeneity of variances which are not based on sample variances only, or other robust tests such as Modified Levene's test.

2. Compare the performance of the tests using samples from several other symmetrical or skewed distributions for instance Cauchy and Exponential distributions, and consider unequal sample sizes across factor levels.

REFERENCES

1. Anderson, V. L., and R. A. MacLean (1974). *Design of Experiments: A Realistic Approach*. Dekker, New York.

2. Bartlett, M. S. (1937). Properties of sufficieny and statistical tests. *Proc. Roy. Soc. Ser. A,* Vol. 160 , pp. 268-282.

3. Breusch, T. and Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, Vol. 47 , pp. 1287-1294.

4. Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrica*, Vol. 40 , pp. 318-315.

5. Conover, W. J., M. E. Johnson, and M. M. Johnson (1981). A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data, Vol. 23, pp. 351-361.

6. Douglas C. Montgomery (2000). *Design And Analysis Of Experiments,* Fifth Edition : John Wiley.

7. Engineering Statistics Handbook.
http://www.itl.nist.gov/div898/hanbook/eda/section3/eda357.htm

8. Greene, William H. (2000). *Econometric Analysis*, 4th ed. : Prentice Hall.

9. Hartley, H. O. (1950). The maximum F-ratio as a short-cut test for homogeneity of variance. *Biometrica,* Vol. 37 , pp. 308-312.

10. High, Robin. (2000).
http://cc.uoregon.edu/cnews/summer2000/statpower.html

11. Hun Myoung Park (2003). Understanding the Statistical Power of a Test.
http://www.indiana.edu/˜ statmath/stat/all/power/power.html

12. Kutner, Nachtsheim, Netter (2004). *Applied Linear Regression Models,* (Fourth Edition) : Mc Graw Hill.

13. Levene, H. (1960). Robust tests for equality of variance. In *Contributions to Probability and Statistics.* Z. Olkin, ed. Stanford University Press, Palo Alto, CA. pp: 278-292.

14. Pierre Legendre and Daniel Borcard (1972). Statistical comparison of univariate tests of homogeneity of variances.*Journal of Statistical Computation and Simulation.* http://biol10.biol.umontreal.ca/BIO2042/MS_THV.PDF.

15. SAS Macro that tests homogeneity of variance in ANOVA models. http://www.sas.com/techsup/download/stat/homovar.sas

16. Statistics Tutorial : Power of a Hypothesis Test. http://davidmlane.com/hyperstat/A108717.html

17. Zar, J. H (1999). *Biostatistical analysis* (Fourth Edition) .Upper Sadle River, N.J. : Prentice Hall.

APPENDIX

PROGRAMS

PROGRAM 1

(SAS 9.1 – Estimated type 1 error)

```
OPTION NODATE NONUMBER;
DATA SimBartHart(keep=rateb rateh);
rejectb=0;rejecth=0;
DO rep=1 to 1000;
  SQ = 0; SLOG=0;Smax=0;Smin=234;mu=0;sigma=1;
  /* Do for each row, i=1,...,5 */
  DO I=1 to 10;
     SUMY=0; SSQY=0;
     ARRAY y{10} y1-y10;
      ARRAY SSQ{10} SSQ1-SSQ10;
     /* given row i, Do for each column */
     DO J=1 to 10;
       y{j}=mu+sigma*rannor(99999);
        SUMY=SUMY+y{j}; SSQY=SSQY+y{j}*y{j};

     /* End for loop j*/
     END;
     SSQ{I}=(1/9)*(SSQY-((SUMY)**2)/10);
     SQ=SQ+SSQ{I};
      /*OUTPUT;*/
     SLOG=SLOG+9*LOG10(SSQ{I});
     if (SSQ{I}>Smax) then Smax=SSQ{I};
     if (SSQ{I}<Smin) then Smin=SSQ{I};
     /*OUTPUT;*/
     mu=mu+2;
  /* End for loop i */
  END;
  SSQP=(9/90)*SQ;
  Q=90*LOG10(SSQP)-SLOG;
  C=281/270;
  Bart=2.3026*(Q/C);
  Hart=Smax/Smin;
  /*OUTPUT;*/
  If(Bart>16.92)then rejectb=rejectb+1;
  If(Hart>9.91)then rejecth=rejecth+1;
End;
rateb=rejectb/1000;
rateh=rejecth/1000;

PROC PRINT;
var rateb rateh;

RUN;
```

PROGRAM 2

(SAS 9.1 - Estimated power of test)

```
OPTION NODATE NONUMBER;
DATA SimBartHart(keep=rateb rateh);
rejectb=0;rejecth=0;
DO rep=1 to 1000;
  SSQ = 0; SLOG=0;Smax=0;Smin=234;mu=0;sigma=1;
  /* Do for each row, i=1,...,5 */
  DO I=1 to 10;
     SUMY=0; SSQY=0;
      If (1<i<=2) then sigma=2;
      If (2<i<=3) then sigma=3;
      If (3<i<=4) then sigma=4;
      If (i>4) then sigma=5;
     ARRAY y{*} y1-y10;
     /* given row i, Do for each column */
     DO J=1 to 10;
       y(j)=mu+sigma*rannor(99999);
        SUMY=SUMY+y(j); SSQY=SSQY+y(j)*y(j);
     /* End for loop j*/
     END;
     SSQI=(1/9)*(SSQY-((SUMY)**2)/10);
     /*OUTPUT;*/
     SSQ=SSQ+SSQI;
     SLOG=SLOG+9*LOG10(SSQI);
     if (SSQI>Smax) then Smax=SSQI;
     if (SSQI<Smin) then Smin=SSQI;
     /*OUTPUT;*/
     mu=mu+2;
  /* End for loop i */
  END;
  SSQP=(9/90)*SSQ;
  Q=90*LOG10(SSQP)-SLOG;
  C=281/270;
  Bart=2.3026*(Q/C);
  Hart=Smax/Smin;
  /*OUTPUT;*/
  If(Bart>16.92)then rejectb=rejectb+1;
  If(Hart>9.91)then rejecth=rejecth+1;
End;
rateb=rejectb/1000;
rateh=rejecth/1000;
PROC PRINT;
var rateb rateh;
RUN;
```

PROGRAM 3

(S-PLUS - Estimated type 1 error)

```
> simHomVar3 <- function(rep)
{
   rejectBart <- 0
   rejectHart <- 0
   rejectX2 <- 0
   for(m in 1:rep) {
      #treatment = 10
      #nobs = 5
      x <- rep(seq(15, 60, 5), rep(5, 10))
      y <- rep(0, length(x))
      mu <- seq(0, 18, 2)
      sigma <- rep(1, 10)
      start <- 1
      end <- 5
      i <- 1
      for(level in unique(x)) {
         set.seed(start * m)
         y[start:end] <- rnorm(5, mean = mu[i], sd =
         sigma[i])
         i <- i + 1
         start <- end + 1
         end <- end + 5
      }
      data <- data.frame(x, y)
      #Bartlet
      a <- length(unique(x))
      S2 <- rep(0, a)
      medlevel <- rep(0, a)
      i <- 1
      for(level in unique(x)) {
         S2[i] <- var(data$y[data$x == level])
         medlevel[i] <- median(data$y[data$x == level])
         i <- i + 1
      }
      Sp2 <- (4/40) * sum(S2)
      q <- 40 * log10(Sp2) - 4 * sum(log10(S2))
      c <- 131/120
      Bart <- (2.3026 * q)/c
      #Hartley
      Hart <- max(S2)/min(S2)
      #Breusch - Pagan
```

```
      regr1 <- lm(y ~ x, data)
      SSE <- sum(residuals(regr1)^2)
      res2 <- residuals(regr1)^2
      newdata2 <- data.frame(res2, x)
      regr2 <- lm(res2 ~ x, newdata2)
      SSRstar <- sum((fitted(regr2) - mean(res2))^2)
      X2 <- (SSRstar/2)/((SSE/length(x))^2)
      rejectBart <- rejectBart + ifelse(Bart > 16.92,
      1, 0)
      rejectHart <- rejectHart + ifelse(Hart > 44.6, 1,
      0)
      rejectX2 <- rejectX2 + ifelse(X2 > 3.84, 1, 0)
   }
   rateBart <- rejectBart/rep
   rateHart <- rejectHart/rep
   rateX2 <- rejectX2/rep
   return(rateBart, rateHart, rateX2)


}
> result3 <- simHomVar3(1000)
> result3
$rateBart:
[1] 0.06

$rateHart:
[1] 0.064

$rateX2:
[1] 0.059
```

PROGRAM 4

(S-PLUS - Estimated power of tests)

```
> simHomVar23b <- function(rep)
{
   rejectBart <- 0
   rejectHart <- 0
   rejectX2 <- 0
   for(m in 1:rep) {
      #treatment = 10
      #nobs = 5
      x <- rep(seq(15, 60, 5), rep(5, 10))
      y <- rep(0, length(x))
      mu <- seq(0, 18, 2)
      #here sigma1=6, sigma2=5, sigma3=4, sigma4=3,
      sigma5=2, others=1
      sigma <- c(6, 5, 4, 3, 2, rep(1, 5))
      start <- 1
      end <- 5
      i <- 1
      for(level in unique(x)) {
         set.seed(start * m)
         y[start:end] <- rnorm(5, mean = mu[i], sd =
         sigma[i])
         i <- i + 1
         start <- end + 1
         end <- end + 5
      }
      data <- data.frame(x, y)
      #Bartlet
      a <- length(unique(x))
      S2 <- rep(0, a)
      medlevel <- rep(0, a)
      i <- 1
      for(level in unique(x)) {
         S2[i] <- var(data$y[data$x == level])
         medlevel[i] <- median(data$y[data$x == level])
         i <- i + 1
      }
      Sp2 <- (4/40) * sum(S2)
      q <- 40 * log10(Sp2) - 4 * sum(log10(S2))
      c <- 131/120
      Bart <- (2.3026 * q)/c
```

```
    #Hartley
    Hart <- max(S2)/min(S2)

    #Breusch - Pagan
    regr1 <- lm(y ~ x, data)
    SSE <- sum(residuals(regr1)^2)
    res2 <- residuals(regr1)^2
    newdata2 <- data.frame(res2, x)
    regr2 <- lm(res2 ~ x, newdata2)
    SSRstar <- sum((fitted(regr2) - mean(res2))^2)
    X2 <- (SSRstar/2)/((SSE/length(x))^2)
    rejectBart <- rejectBart + ifelse(Bart > 16.92,
    1, 0)
    rejectHart <- rejectHart + ifelse(Hart > 44.6, 1,
    0)
    rejectX2 <- rejectX2 + ifelse(X2 > 3.84, 1, 0)
  }
  rateBart <- rejectBart/rep
  rateHart <- rejectHart/rep
  rateX2 <- rejectX2/rep
  return(rateBart, rateHart, rateX2)
}
> result23b <- simHomVar23b(1000)
> result23b
$rateBart:
[1] 0.994

$rateHart:
[1] 0.921

$rateX2:
[1] 1
```