

DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES

ANOVA METHODS FOR THE GROUP MEANS
WITH UNKNOWN VARIANCES

by
Abdullah Fırat ÖZDEMİR

September, 2006
İZMİR

ANOVA METHODS FOR THE GROUP MEANS WITH UNKNOWN VARIANCES

**A Thesis Submitted to the
Graduate School of Natural and Applied Science of Dokuz Eylül
University In Partial Fulfillment of the Requirements for the Degree
of Doctor of Philosophy in Statistics Program**

**by
Abdullah Fırat ÖZDEMİR**

September, 2006

İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “ANOVA METHODS FOR THE GROUP MEANS WITH UNKNOWN VARIANCES” completed by **ABDULLAH FIRAT ÖZDEMİR** under supervision of **PROF. DR. SERDAR KURT** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Serdar KURT

Supervisor

Assist. Prof. Dr. Alirıza FİRUZAN

Committe Member

Assist. Prof. Dr. Cenk ÖZLER

Committe Member

Prof. Dr.Öztaş AYHAN

Jury Member

Assist. Prof. Dr. Ali Kemal ŞEHİRLİOĞLU

Jury Member

Prof.Dr. Cahit HELVACI

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I wish to express my sincere thanks and deep appreciation to my advisor Serdar KURT, for his encouragement and guidance throughout the development of my thesis. He not only gave me endless support and attention which is essential accomplishing this work, but also played an important role in my education.

I would like to acknowledge my dissertation committee members, Alirıza FİRUZAN and Cenk ÖZLER, for their constructive comments and suggestions. I would like to extend special thanks to Tülay SARAÇBAŞI, Sedat ÇAPAR and Engin YILDIZTEPE, their efforts and helpful advices have improved my thesis.

Finally I would like to express my heartfelt thanks and appreciation to my family who has always been extremely supportive of my endeavors and unconditional in their love. My parents, by word and deed, have emphasized the importance of education. They encouraged me to pursue my own interests, willingly offered any advice and would never ever entertain the possibility that I might not be able to accomplish anything I chose to do.

Last, but not least, I wish to express my gratitude to my wife, Bilgen ÖZDEMİR. I am extremely thankful to her for her tremendous patience and support without which this work would be incomplete. Her blessings and love have been a constant source of encouragement and strength.

A.Fırat ÖZDEMİR

ANOVA METHODS FOR THE GROUP MEANS WITH UNKNOWN VARIANCES

ABSTRACT

Analysis of variance (ANOVA) is one of the most powerful tools while investigating the sources of variability in many disciplines like medicine, engineering, agriculture, education, psychology, sociology and biology. In ANOVA, variance of the distributions in which the samples are drawn should be homogeneous to validate the underlying probability distribution of the method and to confine the errors within the desired limits. Violation of this equality of variances assumption is called as heteroscedasticity in literature.

In this study, after describing a general appearance of one-way ANOVA and effects of its inevitable assumptions, the results of heteroscedasticity in one-way fixed effects ANOVA have been examined with a close concern on large sample approximations of treatment and error mean sum of squares and distortion of the distribution of the F ratio. Then, two new and simple approximation procedures which intend to create an easy and applicable alternative under heteroscedasticity and nonnormality have been presented. The purpose of these new approximation procedures is to preserve the actual Type I error rate at a level determined by the researcher and to increase the power as well. Performance of these two new approximation procedures under different experimental patterns have been observed with two separate simulation studies and finally, some recommendations about the preference of these tests and further research topic were given.

Keywords: ANOVA, F test, heteroscedasticity, B^2 test, B_t^2 test, Welch test, Kruskal-Wallis test

VARYANSLARI BİLİNMEYEN GRUPLAR İÇİN VARYANS ANALİZİ YÖNTEMLERİ

ÖZ

Varyans çözümlemesi yöntemi tıp, mühendislik, tarım, eğitim, psikoloji, sosyoloji ve biyoloji gibi birçok alanda değişim kaynaklarının araştırılması sürecinde başvurulan en güçlü araçlardan biridir. Yöntemin dayandığı olasılık dağılımının (F) geçerli olması ve verdiği sonuçların istenilen hata sınırları içinde kalması için her bir deneme ile elde edilen gözlemlerin eşit varyanslı kitlelerden çekilmesi gerekir. Bu varsayımın sağlanmaması literatürde farklı varyanslılık (heteroscedasticity) olarak adlandırılır.

Bu çalışmada tek yönlü varyans analizi ve vazgeçilmez varsayımları hakkında genel açıklamalar yapıldıktan sonra farklı varyanslılığın tek yönlü özel seçimli varyans analizi üzerindeki etkileri, deneme ve hata kare ortalamaları için yapılan büyük örneklem genişliği yaklaşımları ve F oranının dağılımında gözlenen bozulma açısından değerlendirilmiştir. Daha sonra, farklı varyanslılık ve normal dağılımdan uzaklaşma durumunda araştırmacılar için kolay ve uygulanabilir alternatif çözümler olması planlanan iki farklı yaklaşık çözüm önerisi tanıtılmıştır. Bu önerilerin amacı, I. Tip hatanın araştırmacının belirlediği seviyede korunması ile beraber güç değerinde artırılmasıdır. Önerilen bu iki yaklaşık çözüm yönteminin farklı deney düzenlerindeki performansları, yapılan iki farklı benzetim çalışması ile gözlenerek tercihler ve ileri araştırma konuları hakkında önerilerde bulunulmuştur.

Anahtar sözcükler: ANOVA, F testi, farklı varyanslılık, B^2 testi, B_t^2 testi,

Welch testi, Kruskal-Wallis testi

CONTENTS

	Page
THESIS EXAMINATION RESULT FORM	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ÖZ	v
CHAPTER ONE - INTRODUCTION	1
CHAPTER TWO – ONE-WAY FIXED EFFECTS ANOVA AND RESULTS OF HETEROGENEITY OF VARIANCE PROBLEM	6
2.1 Sampling Distribution of Sum of Squares.....	7
2.2 One-Way Analysis of Variance.....	11
2.3 Assumptions and Analysis of Variance.....	13
2.4 Results of Heterogeneity of Variance Problem	16
2.4.1 Distribution of F Ratio.....	16
2.4.2 Discrepancy Between Nominal and Actual Significance Level	17
CHAPTER THREE - DIFFERENT SOLUTION APPROACHES.....	21
3.1 Approximate Tests	21
3.1.1 Cochran Test	22
3.1.2 Welch Test	23
3.1.3 James 2'nd Order Test.....	24
3.1.4 Box Test.....	26
3.1.5 Brown Forsythe Test	27
3.1.6 Wilcox Test.....	27
3.1.7 Krutchkoff K Test	28
3.1.8 Wilcox's 2'nd Test	29
3.1.9 Asiribo and Gurland's Approximate ANOVA Test.....	30
3.1.10 Alexander Govern Test.....	31
3.1.11 Modified Brown Forsythe Test	31
3.1.12 Adjusted Welch Test	32
3.1.13 Smith and Hickling Test	34

3.2 Exact Tests.....	34
3.2.1 Chen-Chen-Ding Two stage ANOVA Test.....	35
3.3 Nonparametric Tests	37
3.3.1 Kruskal-Wallis Test.....	37
3.4 Data Transformations	38
3.5 Weighted Least Square Estimation Method	40
3.6 Robust Statistical Procedures	40
CHAPTER FOUR - NEW AND SIMPLE APPROXIMATION PROCEDURES AND SIMULATION STUDIES	42
4.1 A New and Simple Approximation Procedure and a Numerical Application..	42
4.2 Simulation Study	46
4.3 A New and Simple Approximation Procedure with Trimmed Mean and Winsorized Variance	55
4.3.1 Trimmed Mean B_t^2 Test	55
4.3.2 Simulation Study for Trimmed Mean B_t^2 Test	59
CHAPTER FIVE - CONCLUSION AND FURTHER RESEARCH.....	71
REFERENCES	75

CHAPTER ONE

INTRODUCTION

The appropriate use of any inferential statistical procedure depends on the satisfaction of certain assumptions about samples and in some cases about the population(s) from which samples are drawn. The theory of analysis of variance under the classical assumptions of normality, independence and homoscedasticity of the errors is quite complete. Studies of the robustness of the F-statistic, for example in the one-way layout, have shown that the violation of normality has little effect on inferences about the means (Lindman, 1992; Monthgomery, 1991). However, the violation of independence or of equality of variances can have a serious effect on inferences about the means, especially if the cell sizes are unequal (Bishop, 1976; Brown & Forsythe, 1974; Dijkstra & Werter, 1981; Hartung, Argaç & Makambi, 2002; Scheffe, 1959).

The analysis of variance methodology is concerned with the investigation of the factors likely to contribute significant effects, by suitable choice of experiments. It is a technique by which variations associated with different factors or defined sources may be isolated and estimated. The procedure involves the deviation of total observed variation in data into individual components attributable to various factors and those due to random or chance fluctuation, and performing tests of significance to determine which factors influence the experiment. The methodology was originally developed by Sir Roland A. Fisher who gave it the name of “analysis of variance.” The analysis of variance is the most widely used tool of modern statistics by research workers in the substantive fields of agriculture, biology, psychology, sociology, education, engineering and so forth.

The development of analysis of variance methodology has in turn affected and influenced the types of experimental research being carried out in many fields. For example in quantitative genetics which relies extensively on separating the total variation into environmental and genetic components, many of the concepts are directly linked to the principles and procedures of the analysis of variances.

The use of the term “analysis of variance ” to describe the statistical methodology dealing with the means of a variable across groups of observations is not wholly satisfactory. The term seems to have its origin in the work of Fisher (1918) when he partitioned the total variance of a human attribute into components attributed to heredity, environment and other factors. This partition led him to an equation

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2 \quad (1.1)$$

where σ^2 is the total variance and σ_i^2 's are variances associated with different factors.

The choice of a technique for assessing the statistical significance of any treatments or groups of differences depends on the researcher's assumptions concerning the nature of the population(s) from which the samples were drawn and on the hypothesized model for the observations. If, as is often the case, an observation is envisioned as an additive combination of a constant, a group effect and a unique individual component, the model can be written as

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (1.2)$$

Here, Y_{ij} is an observation on the j 'th unit in the i 'th group on variable Y , μ is the population mean of Y (a constant serving as a general location parameter), τ_i is an effect associated with the i 'th group (the deviation of the i 'th group mean from the overall mean) and ε_{ij} is a component unique to the j 'th unit in group i (the deviation of Y_{ij} from the mean of the i 'th group). When the above model is appropriate and the dependent variable is assumed to be normally distributed in the population, the technique typically used to assess the statistical significance of the group effects (the τ_i) is analysis of variances (ANOVA).

The null hypothesis tested in ANOVA is that the means of all groups are equal, which also implies that all $\tau_i = 0$. Thus the test of the null hypothesis is used to decide if the model in (1.2) provides a significantly more accurate representation of the observations than the simpler model

$$Y_{ij} = \mu + \varepsilon_{ij}. \quad (1.3)$$

The null hypothesis is rejected if the group means differ more than would be expected (at a specified probability level) in random sampling from a single population mean μ and variance σ_ε^2 .

Although ANOVA has been used extensively by researchers for nearly 60 years it is important to recognize that this statistical procedure were derived under certain assumptions concerning the populations from which the samples are drawn. Actually, none of the assumptions underlying the F test are ever fully satisfied by real data. The populations from which the samples are drawn are never exactly normally distributed with precisely equal variance. The relevant question is not whether ANOVA assumptions are met exactly, but rather whether the plausible violations of the assumptions have serious consequences on the validity of probability statements based on the standard assumptions. If the violation of these assumptions were likely to result in a large F, then the F test would be a poor test of the null hypothesis. In this section, these assumptions are briefly outlined and some general comments are made about the consequences of violations. Hereafter, all references to ANOVA are intended to refer to the fixed effects case, where all classifications of interest are included in the design. In one-way ANOVA it is assumed that;

1. Samples are drawn randomly from a normally distributed population with mean μ and equal variance σ_ε^2 .
2. Observations of different subjects are independent of one another.
3. The additive model in equation (1.2) is appropriate.

The first assumption is equivalent to saying that the groups are assumed to be drawn from normal populations with equal means and equal variances. The most important departures from normality are likely to be either nonzero skewness or nonzero kurtosis. Other departures from normality are not likely to be very important. In general, only the kurtosis of the distribution is likely to have any appreciable effect on F. Under the null hypothesis MSE and MSTr are unbiased estimates of σ^2 no matter what the population distribution, their variances depend on the kurtosis of the population. Fortunately, in most cases the effects of nonzero kurtosis can be ignored. Box and Anderson (1955) have shown that when the kurtosis is not zero, an approximate F test can still be used.

The variance of the populations should be constant across groups. Unequality of population variances is called as heteroscedasticity in literature. A more detailed review of the consequences of heteroscedasticity is provided in the next chapter.

The assumptions of randomization and independence are essential in order for the test of statistical significance to make sense at all. Obviously, an answer to the question of whether group means differ more than would be expected in random sampling from a single population is not of interest if the samples are not random. Furthermore, violations of these assumptions alter the sampling distribution of the means under the null hypothesis, which renders the standard F test invalid for assessing statistical significance of group mean differences. Also Scheffe (1959) has demonstrated that nonindependence, in the form of nonzero serial correlations among observations, has severe effects on Type I error rates in ANOVA regardless of sample sizes. The F test yields too many significant results when the serial correlation among observation is positive and too few significant results when it is negative.

Generally probability statement about differences among group means in ANOVA are affected little by nonnormality and, if sample sizes are equal, little by moderate heterogeneity of variances. With unequal sample sizes, ANOVA tends to yield too many significant results when the smaller group is more variable and too

few significant results when the larger group is more variable. The research upon which these general conclusions are based is summarized by Scheffe (1959) and by Glass, Peckham and Sanders (1972).

In second chapter of this study, some general information about one-way analysis of variance and results of heterogeneity of variance problem in one-way analysis of variance will be given. Different solution approaches to heterogeneity of variance problem will be described in chapter three. Presentation of two new and simple approximation procedures and results of simulation studies which were implemented to compare the actual significance level and power of the new approximations, conventional F test and two other alternatives (Welch test, Kruskal-Wallis test) will be given in chapter four. Finally, last chapter consist of some recommendations about the preference of these tests for different types of experimental conditions and some suggestions for further research.

CHAPTER TWO

ONE-WAY FIXED EFFECTS ANOVA AND RESULTS OF HETEROGENEITY OF VARIANCE PROBLEM

The purpose of analysis of variance is to test differences in means for statistical significance. This is accomplished by analyzing the variance of response variable Y_{ij} that is, by partitioning the total variance into the components that is due to true random error (unexplained variation) and the component that is due to differences between means (explained variation).

$$SS_t = SS_w + SS_{bet} \quad (2.1)$$

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 + \sum_{i=1}^k n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \quad (2.2)$$

These variance components are then tested for statistical significance by means of F test. $\forall_i = 1, 2, \dots, k$ and $Y_{ij} \sim NID(\mu_i, \sigma^2)$ $j = 1, 2, \dots, n_i$

$$F = \frac{\sum_{i=1}^k n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 / k - 1}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 / N - k} = \frac{MSTr}{MSE} \sim F_{k-1, N-k} \quad (2.3)$$

and if this calculated F value exceeds $F_{k-1, N-k, 1-\alpha}$ critical F value, we reject the null hypothesis of no differences between means otherwise fail to reject H_0 .

$$\begin{aligned} H_0 : \mu_1 = \mu_2 = \dots = \mu_k \\ H_1 : \text{Not all } \mu' \text{ s are equal} \end{aligned} \quad (2.4)$$

2.1 Sampling Distribution of Sum of Squares

It is known that if $Y \sim N(\mu, \sigma^2)$ then mean of a random sample $\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and

$$\left(\frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \sim \chi_1^2 \quad (2.5)$$

$$(\bar{Y} - \mu)^2 \sim \frac{\sigma^2}{n} \chi_1^2 \quad (2.6)$$

$$n(\bar{Y} - \mu)^2 \sim \sigma^2 \chi_1^2 \quad (2.7)$$

When there are k groups with n observations and the equal variance σ^2 for group one

$$\left(\frac{\bar{Y}_{1.} - \mu}{\frac{\sigma}{\sqrt{n}}} \right) = Z \Rightarrow \left(\frac{\bar{Y}_{1.} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 = Z^2 \sim \chi_1^2 \quad (2.8)$$

for group two

$$\left(\frac{\bar{Y}_{2.} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \sim \chi_1^2 \quad (2.9)$$

and for group k

$$\left(\frac{\bar{Y}_{k.} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \sim \chi_1^2 \quad (2.10)$$

then as a total (by additivity theorem of chi-square)

$$\sum_{i=1}^k \left(\frac{\bar{Y}_{i.} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \sim \chi_k^2 \quad (2.11)$$

can be written. This quantity can be expressed in a different form like

$$\sum_{i=1}^k (\bar{Y}_{i.} - \mu)^2 \sim \frac{\sigma^2}{n} \chi_k^2 \quad (2.12)$$

$$\sum_{i=1}^k n(\bar{Y}_{i.} - \mu)^2 \sim \sigma^2 \chi_k^2 \quad (2.13)$$

When we use estimate $\bar{Y}_{..}$ instead of population mean μ , the distribution of sum of squares due to treatment will be

$$\sum_{i=1}^k n(\bar{Y}_{i.} - \bar{Y}_{..})^2 \sim \sigma^2 \chi_{k-1}^2. \quad (2.14)$$

This quantity is called as between sum of squares or sum of squares due to treatment (SSTr) and since population variances are assumed to be equal, each group (treatment) has an equal weight in this sum. Distribution of mean squares due to treatment can be written as

$$\frac{\sum_{i=1}^k n(\bar{Y}_{i.} - \bar{Y}_{..})^2}{k-1} = MSTr \sim \frac{\sigma^2 \chi_{k-1}^2}{k-1} \quad (2.15)$$

or

$$\frac{MSTr}{\sigma^2} \sim \frac{\chi_{k-1}^2}{k-1} \quad (2.16)$$

Distribution of sum of squares due to error can be derived as follows. If a random variable Y has a normal distribution with mean μ and standard deviation σ , then

$$Z = \frac{Y - \mu}{\sigma} \sim \text{Normal}(0,1) \quad (2.17)$$

The square of such a standard normal variable has a chi-squared (χ^2) distribution with 1 degree of freedom

$$Z^2 = \left(\frac{Y - \mu}{\sigma} \right)^2 \sim \chi_1^2 \quad (2.18)$$

When there are k groups with n observation and the same variance σ^2 for group 1

$$\left(\frac{Y_1 - \mu}{\sigma} \right)^2 + \left(\frac{Y_2 - \mu}{\sigma} \right)^2 + \dots + \left(\frac{Y_{n_1} - \mu}{\sigma} \right)^2 = Z^2 + Z^2 + \dots + Z^2 = \chi_1^2 + \chi_1^2 + \dots + \chi_1^2 = \chi_{n_1}^2 \quad (2.19)$$

$$\sum_{j=1}^{n_1} (Y_{1j} - \mu)^2 \sim \sigma^2 \times \chi_{n_1}^2 \quad (2.20)$$

for group 2

$$\left(\frac{Y_1 - \mu}{\sigma} \right)^2 + \left(\frac{Y_2 - \mu}{\sigma} \right)^2 + \dots + \left(\frac{Y_{n_2} - \mu}{\sigma} \right)^2 = Z^2 + Z^2 + \dots + Z^2 = \chi_1^2 + \chi_1^2 + \dots + \chi_1^2 = \chi_{n_2}^2 \quad (2.21)$$

$$\sum_{j=1}^{n_2} (Y_{2j} - \mu)^2 \sim \sigma^2 \times \chi_{n_2}^2 \quad (2.22)$$

and for group k

$$\left(\frac{Y_1 - \mu}{\sigma}\right)^2 + \left(\frac{Y_2 - \mu}{\sigma}\right)^2 + \dots + \left(\frac{Y_{n_k} - \mu}{\sigma}\right)^2 = Z^2 + Z^2 + \dots + Z^2 = \chi_1^2 + \chi_1^2 + \dots + \chi_1^2 = \chi_{n_k}^2 \quad (2.23)$$

$$\sum_{j=1}^{n_k} (Y_{kj} - \mu)^2 \sim \sigma^2 \times \chi_{n_k}^2 \quad (2.24)$$

since $n_1 + n_2 + \dots + n_k = N$ as a total (by additivity theorem of chi-square)

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2 \sim \sigma^2 \times \chi_N^2 \quad (2.25)$$

When we use i 'th group sample mean $\bar{Y}_i$ instead of population mean μ the distribution of sum of squares due to error will be

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \sim \sigma^2 \times \chi_{N-k}^2. \quad (2.26)$$

This quantity is called as within sum of squares or sum of squares due to error (SSE) and since population variances are assumed to be equal, each group (treatment) has an equal weight in this sum. Distribution of mean squares due to error can be written as

$$\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{N - k} = MSE \sim \frac{\sigma^2 \times \chi_{N-k}^2}{N - k} \quad (2.27)$$

or

$$\frac{MSE}{\sigma^2} \sim \frac{\chi_{N-k}^2}{N - k} \quad (2.28)$$

So, distribution of F ratio is defined as

$$F = \frac{\sum_{i=1}^k n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 / k-1}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 / N-k} = \frac{MSTr}{MSE} = \frac{\sigma^2 \chi_{k-1}^2 / k-1}{\sigma^2 \chi_{N-k}^2 / N-k} \sim F_{k-1, N-k} \quad (2.29)$$

2.2 One-Way Analysis of Variance Model

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, k \\ j = 1, 2, \dots, n_i \end{cases} \quad (2.30)$$

In this model ;

- Y_{ij} represents an individual measurement of the j'th experimental unit on i'th treatment. There are “ $n_1 + n_2 + \dots + n_k = N$ ” values of Y_{ij} .
- μ is a parameter which represents the overall mean of all observations in each treatment
- τ_i is also a parameter represents the value of the i'th treatment effect.
- ε_{ij} is the random error component

This model is called the one-way or single factor analysis of variance model because only one factor is investigated.

This model describes two different situations with respect to the treatment effects. First the k treatments could have been specifically chosen by the experimenter. In this situation hypothesis about the treatment means can be tested and the conclusions will apply only to the factor levels considered in the analysis. The conclusions can not be extended to similar treatments that were not explicitly considered. The model parameters μ, τ_i, σ^2 can also be estimated. This is called the fixed effects model.

Alternatively, the k treatments could be a random sample from a larger population of treatments. In this situation the conclusions can be extended to all treatments in the population, whether they were included in the analysis or not. Here τ_i are random variables, and knowledge about the particular ones investigated is relatively useless. Instead, the hypothesis about the variability of τ_i is tested and variability of τ_i is tried to be estimated. This is called random effects model or components of variance model. In this study only the fixed effects models are to be considered.

One-way analysis of variance model can be written in a different way.

$$Y_{ij} - \mu = (\mu_i - \mu) + (Y_{ij} - \mu_i) \quad (2.31)$$

A graphical representation of the model with three treatments has been given in Figure 2.1 below.

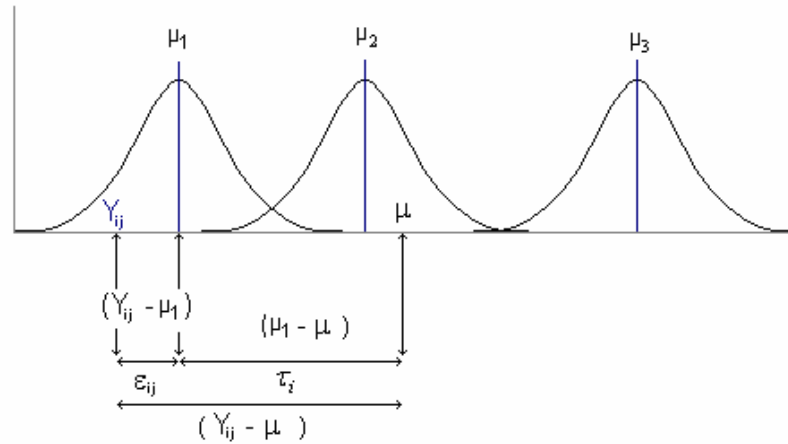


Figure 2.1 Graphical representation of one-way ANOVA model

In the model, $Y_{ij} - \mu$ is called as individual deviation from overall mean. This quantity consist of two kinds of effects. First one is treatment effect and described as

$$\tau_i = \mu_i - \mu \quad (2.32)$$

and the second one is within group variation (unexplained variation) or error and described as

$$\varepsilon_{ij} = Y_{ij} - \mu_i \quad (2.33)$$

Under H_0 all treatment effects are equal to zero and total variation observed in response variable Y_{ij} belongs to within group variation.

2.3 Assumptions and Analysis of Variance

It has been stated that,

“When the formulas and procedures of ANOVA are used merely to summarize properties of data in hand, no assumptions are needed to validate them. On the other hand, when ANOVA is used as a method of statistical inference, for inferring properties of the “population” from which the data in hand were drawn, then certain assumptions, about the “population” and the sampling procedure by means of which the data were obtained, must be fulfilled if the inferences are to be valid ” (Eisenhart 1947).

Investigations of the effects of assumption violation on the ANOVA F test date back more than 70 years, (Pearson 1931), and this body of research therefore has substantial depth and breadth. It is well known that for a given set of data, certain assumptions must be satisfied if the F test is to produce valid results. Briefly, these assumptions are independence of observations, normality of the treatment populations and homogeneity of population variances. The first assumption is primarily a function of the study design and can be satisfied through the use of randomization techniques, while the latter two are functions of the populations under investigation and therefore are generally beyond the control of researchers.

Much of the research on violations of the normality assumption has been consistent in noting the relative insensitivity of the F test to departures from normality. For example, Cochran (1947), in summarizing the earliest investigations of this topic, “observed that nonnormality appeared to have little effect on Type I error performance but could have more serious implications for statistical power”, a point which has been echoed by Wilcox (1995). The subsequent review by Glass et al.(1972) concluded that “skewed populations have very little effect on either the level of significance or the power of the fixed effects model F test and that kurtosis was not much of a consideration either, except when group sizes were small and statistical power was being evaluated.” Harwell, Rubinstein, Hayes and Olds (1992), while noting that nonnormality was not much of a factor in Type I error or power performance, also observed that when group sizes were unequal, skewed distributions were associated with slightly inflated Type I error rates.

Variances of all populations which belong to a specific treatment in the experiment should be the same. This property is known as homogeneity of variance assumption and assures the ratio $MSTr/MSE$ to provide an *exact* test of the hypothesis that the population means are equal. Unequal population variances problem is called as heteroscedasticity in ANOVA and “Behrens-Fisher Problem” for all two sample tests. This is the most critical assumption, the assumption most often violated, and the assumption least often checked. Often an ANOVA is conducted on heteroscedastic data from populations with heterogeneous variances.

The severity of the effect of violating the variance homogeneity assumption is a function of whether group sizes are equal or unequal (whether the design is balanced or unbalanced). The general consensus among the earliest writers on the topic was that the F test is largely robust to variance heterogeneity when group sizes are equal (Boneau, 1960; Norton, 1952). Box (1954) theorized, however, that this is not always true. According to Box, the robustness of the F test for balanced designs is dependent on the spread of the variances; under variance heterogeneity, the F statistics only approximately follows an F distribution, and c , a coefficient of variation of the unequal variances, is a critical determinant of the approximation, where

$$c = \frac{\sqrt{\frac{1}{k} \sum_{i=1}^k (\sigma_i^2 - \bar{\sigma}^2)^2}}{\bar{\sigma}^2} \quad (2.34)$$

σ_i^2 is the variance for the i'th group, and

$$\bar{\sigma}^2 = \frac{1}{k} \sum_{i=1}^k \sigma_i^2 \quad (2.35)$$

is the unweighted mean variance. The parameter, c , ranges in value from 0 to $(k-1)^{1/2}$ and attains a maximum value when $(k-1)$ of the variances are equal to σ^2 and the remaining one is equal to $a\sigma^2$, where a is some constant. Box computed exact Type I error rates for the F test under a variety of conditions of variance heterogeneity when the group sizes were equal and observed that error rates could be seriously inflated; for example, when $k=7$, $n=3$ and $a=10$, the true Type I error rate of the F test is 0.12 when testing at $\alpha = 0.05$.

In commenting on Box's (1954) work, Glass et al. (1972) noted that the inflated Type I error rates reported by Box would rarely be observed in applied research, as they were the result of very small sample sizes and large degrees of variance heterogeneity. However, Glass et al. concluded that "the conventional conclusion that heterogeneous variances are not important when n's are equal seems to have boundary conditions like all other conclusions in this area." Harwell et al. (1992) went even further, cautioning that "equal sample sizes provided little protection against inflated error rates for unequal variances."

The effect of variance heterogeneity may be more serious when group sizes are unequal than when they were equal. When group sizes and variances are positively paired, so that the group with the largest sample size also exhibits the largest degree of variability, the F test will be conservative, that is, the true Type I error rate will be less than α . For negative pairings, in which the group with smallest sample size has the largest degree of variability, the F test will be liberal, that is the true Type I error rate will exceed α (Glass et al., 1972; Harwell et al., 1992).

2.4 Results of Heterogeneity of Variance Problem

2.4.1 Distribution of F Ratio

The numerator and denominator sums of squares in the F ratio $MSTr/MSE$ are distributed as weighted sum of squares of independent normal random variables with weights σ_i^2 which has been shown through equations (2.5) and (2.26). When the variances differ between populations these weights differ from each other and the distributions of mean squares do not follow χ^2 .

By far the best article about the effect of unequal variances on the F test is Box (1954), Rupert and Miller (1986). Box developed the distribution theory for quadratic forms in the case of heteroscedasticity and applied it to the one-way classification. The ratio of mean squares is distributed approximately as $bF(h', h)$ where

$$b = \frac{N-k}{N(k-1)} \frac{\sum_{i=1}^k (N-n_i) \sigma_i^2}{\sum_{i=1}^k (n_i-1) \sigma_i^2} \quad (2.36)$$

$$h' = \frac{\left\{ \sum_{i=1}^k (N-n_i) \sigma_i^2 \right\}^2}{\left\{ \sum_{i=1}^k n_i \sigma_i^2 \right\}^2 + N \sum_{i=1}^k (N-2n_i) \sigma_i^4} \quad (2.37)$$

$$h = \frac{\left\{ \sum_{i=1}^k (n_i-1) \sigma_i^2 \right\}^2}{\left\{ \sum_{i=1}^k (n_i-1) \sigma_i^4 \right\}} \quad (2.38)$$

2.4.2 Discrepancy Between Nominal and Actual Significance Level

Nominal significance level (nominal probability of Type I error) is a concept defined when considering the effects of violation of the assumptions of a statistical test. If an experimenter believes the assumptions underlying a test are met and tests null hypothesis H_0 at the α level (even though the assumptions are not met and he actually has some probability other than α) the nominal significance level is α .

Actual significance level (actual probability of Type I error) is the true probability of a Type I error calculated from a knowledge of the manner in which certain assumptions underlying the statistical tests are violated.

An experimenter may wish to test the omnibus null hypothesis of equality of treatment means at nominal significance level $\alpha = 0.05$ but in reality this level may reach 3 or 4 times this level which is called as actual significance level because of the heterogeneity of variances problem (Wilcox, Charlin & Thompson 1986). Degree of this discrepancy depends mainly on the degree of heterogeneity of population variances and number of replications made with each treatment.

To understand the effect of unequal variances on the F test, it suffices to examine the large sample case where all the n_i are large. The denominator mean sum of squares is converging to its expected value, which is

$$E(MSE) = E \left[\frac{1}{N-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 \right] = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) \sigma_i^2 \quad (2.39)$$

where σ_i^2 is the variance of the observations from the i 'th population. Since $N - k = \sum_{i=1}^k (n_i - 1)$, the expectation of sum of squares due to error is a weighted average of the σ_i^2 and called as $\bar{\sigma}^2$. The expectation of the numerator mean sum of squares under H_0 is

$$\begin{aligned}
& E \left[\frac{1}{k-1} \sum_{i=1}^k n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \right] \\
&= \frac{1}{k-1} \left[\sum_{i=1}^k n_i E(\bar{Y}_{i.} - \mu)^2 - NE(\bar{Y}_{..} - \mu)^2 \right] \\
&= \frac{1}{k-1} \left[\sum_{i=1}^k n_i \frac{\sigma_i^2}{n_i} - N \frac{\sum_{i=1}^k n_i \sigma_i^2}{N^2} \right] \tag{2.40} \\
&= \frac{1}{N(k-1)} \sum_{i=1}^k (N - n_i) \sigma_i^2
\end{aligned}$$

This quantity is a different weighted average of the σ_i^2 and called as $\bar{\sigma}_*^2$

When the n_i are all equal, the two weighted averages agree ($\bar{\sigma}^2 = \bar{\sigma}_*^2$) which means the F ratio is centered near 1 as it should be. But the variance of the numerator is

$$\text{Var}(MSTr) = \frac{2\bar{\sigma}^4}{k-1} \left[1 + \frac{(k-2)}{k(k-1)} \frac{\sum_{i=1}^k (\sigma_i^2 - \bar{\sigma}^2)^2}{\bar{\sigma}^4} \right]. \tag{2.41}$$

When the variances are equal the quantity in brackets in the above formula should be 1, but it obviously exceeds this when the σ_i^2 differ. Thus the actual variance is larger than the theoretical variance (the variance where the σ_i^2 equal) and the upper tail of the distribution of the F ratio has more mass in it than anticipated by the χ_{k-1}^2 distribution. For an observed F ratio the actual significance level is larger than the one calculated from the tables, but numerical studies indicate that the effect is not large. This conclusion is also born out in small samples (Box, 1954, Scheffe, 1959).

When the n_i are unequal, the effects can be more serious. Suppose that the large σ_i^2 happen to be associated with the large n_i . Then in $\bar{\sigma}^2$, the large σ_i^2 receive greater weight, where as in $\bar{\sigma}_*^2$ the small σ_i^2 receive greater weight. The expectation of the numerator mean squares is, therefore, less than the expectation of the denominator, and the center of the distribution of the F ratio is shifted below 1. The actual significance level is less than the one stated from the tables (nominal values).

If the large σ_i^2 are associated with the small n_i , the shift goes in the opposite direction. The actual significance level exceeds its nominal level without too much disparity in the variances. Falsely reporting significant results when the small samples have the larger variances is a serious worry. To balance the experiment is very crucial if it is possible. Then unequal variances and other departures from assumptions have the least effect. As a final point, the extensive study made by Glass et al.(1972) has a table which summaries the consequences of violation of assumptions of the fixed effects analysis of variances. This well prepared table has been given in Table 2.1 here.

Table 2.1 Summary of consequences of violation of assumptions of the fixed effects ANOVA

	Equal n's		Unequal n's	
Type of violation	Effect on α	Effect on Power	Effect on α	Effect on Power
Non independence of errors	Non independence of errors seriously affects both the level of significance and power of the F test regardless whether n's are equal or unequal.			
Nonnormality: Skewness	Skewed populations have very little effect on either the level of significance or the power of the fixed effects model F test			
Kurtosis	Actual α is less than nominal α when populations are leptokurtic ($\beta_2 > 3$). Actual α exceeds nominal α for platykurtic populations. (effects are slight)	Actual power is less than nominal power when populations are platykurtic. Actual power exceeds nominal when populations are leptokurtic. Effects can be substantial for small n	Actual α is less than nominal α when populations are leptokurtic ($\beta_2 > 3$). Actual α exceeds nominal α for platykurtic populations. (effects are slight)	Actual power is less than nominal power when populations are platykurtic. Actual power exceeds nominal when populations are leptokurtic. Effects can be substantial for small n
Heterogeneous variances	Very slight effect on α , which is seldom distorted by more than a few hundredths. Actual α seems always to be slightly increased over the nominal α .	No theoretical power value exists when variances are heterogeneous	α may be seriously affected. Actual α exceeds nominal α when smaller samples are drawn from populations with greater variance; actual α is less than nominal α when smaller samples are drawn from populations with smaller variance.	No theoretical power value exists when variances are heterogeneous
Combined nonnormality and heterogeneous variances	Nonnormality and heterogeneous variances appear to combine additively to affect either level of significance or power. (For example, the depressing effect on α of leptokurtosis (leptokurtic populations) could be expected to be counteracted by the elevating effect on α of having drawn smaller samples from populations with greater variance.			

CHAPTER THREE

DIFFERENT SOLUTION APPROACHES

Over the years, many attempts have been made to find solutions that are robust in both Type I and Type II error rate performance while at the same time having nominal performance when the homogeneity of variance assumption is not violated. There are 6 main approaches proposed to solve the mentioned problem;

- Approximate Tests
- Exact Tests
- Nonparametric Tests
- Data Transformations
- Weighted Least Square Estimation Method
- Robust Statistical Procedures

3.1 Approximate Tests

Approximate tests are the most popular of these several approaches. Under a true null hypothesis, the sampling distribution of these test's test statistics is *approximated* by a well-known sampling distribution like F, t or χ^2 . There are many approximate tests as an alternative for the fixed effect conventional one-way ANOVA F test in the literature and some of them can be listed as follows

- Cochran test (1937)
- Welch test (1951)
- James 2nd order test (1951)
- Box test (1954)
- Brown-Forsythe test (1974)
- A new alternative to the ANOVA by Wilcox R.R (1988)
- Krutchkoff K test (1988)
- Wilcox 's 2nd test (1989)
- Approximate ANOVA F test by Asiribo and Gurland (1990)
- Alexander-Govern test (1994)

- Modified Brown-Forsythe test by Mehrotra (1997)
- Adjusted Welch test by Bökkenhoff and Hartung (1998)
- One-Way ANOVA Under Heteroscedastic Errors by Smith and Hickling (1998)

Let Y_{ij} be the j 'th randomly sampled observation from the i 'th treatment. ($i=1,2,\dots,k$ and $j=1,2,\dots,n_i$). Observations have been assumed to be sampled from normally distributed populations with not necessarily equal population variances. This information is valid for all approximate tests below. And all tests test the null hypothesis of equality of population means.

3.1.1 Cochran Test

The statistics

$$C = \sum_{i=1}^k \omega_i \left(\bar{Y}_i - \sum_{i=1}^k h_i \bar{Y}_i \right)^2 \quad (3.1)$$

where

$$\omega_i = \frac{n_i}{s_i^2} \quad (3.2)$$

s_i^2 is the sample variance for the i 'th group and

$$h_i = \frac{\omega_i}{\sum_{i=1}^k \omega_i} \quad (3.3)$$

was proposed by Cochran (1937), and then modified by James (1951) and Welch (1951). Under H_0 , the Cochran Statistics is distributed approximately as a χ^2 variable with $k-1$ degrees of freedom. Reject H_0 at level α if $C > \chi_{k-1, 1-\alpha}^2$.

3.1.2 Welch Test

The solution given by Welch (1951) is undoubtedly one of the best known approximate procedures. As well, it has served as the model for other approaches to testing for mean equality. The Welch procedure (W) is based on the statistics

$$W = \frac{\sum_{i=1}^k \omega_i \left(\bar{Y}_{i.} - \sum_{i=1}^k h_i \bar{Y}_{i.} \right)^2}{(k-1) + 2(k-2)(k+1)^{-1} \sum_{i=1}^k (n_i - 1)^{-1} (1 - h_i)^2} \quad (3.4)$$

where

$$\omega_i = \frac{n_i}{s_i^2} \quad (3.5)$$

s_i^2 is the sample variance, $\bar{Y}_{i.}$ is the sample mean for the i'th group,

$$h_i = \frac{\omega_i}{\sum_{i=1}^k \omega_i}. \quad (3.6)$$

Welch (1951) proposed to approximate this quantities distribution using an F variable. Under H_0 , the statistics W has an approximate F distribution with k-1 and v_w degrees of freedom, where

$$v_w = \frac{k^2 - 1}{3 \sum_{i=1}^k (n_i - 1)^{-1} (1 - h_i)^2}. \quad (3.7)$$

The hypothesis H_0 is rejected at level α if

$$W > F_{k-1, v_w, 1-\alpha} \quad (3.8)$$

The W test is available in SPSS with 11'th version. As well, a generalized version of this procedure is available in a SAS/IML program developed by Lix and Keselman (1995).

3.1.3 James 2'nd Order Test

At the same time that Welch (1951) was developing his solution, James was working on a similar one. Two statistical procedures were proposed by James (1951); these are referred to as his first and second order procedures. The two are equivalent in terms of the defined test statistics, but differ in the critical value used to assess statistical significance. The former is not widely recommended, as it can not effectively control the rate of Type I error under variance heterogeneity (Brown-Forsythe 1974). The second order procedure (J2) is based on the statistics

$$U = \sum_{i=1}^k d_i (\bar{Y}_i - \bar{Y}_{..}^*)^2 \quad (3.9)$$

where $d_i = \frac{n_i}{\hat{\sigma}_i^2}$, $\bar{Y}_{..}^* = \sum_{i=1}^k \frac{d_i \bar{Y}_i}{D}$ and $D = \sum_{i=1}^k d_i$. $\hat{\sigma}_i^2$ is the sample variance for the i'th group. For any integers t and s, let

$$R_{st} = \sum_{i=1}^k \frac{1}{v_i^s} \left(\frac{d_i}{D} \right)^t \quad (3.10)$$

where $v_i = n_i - 1$ and D is previously defined. Furthermore, let

$$\chi_{2s} = \frac{c^s}{[(k-1)(k+1) \dots (k+2s-3)]} \quad (3.11)$$

where c is the $(1-\alpha)$ percentile of a chi square distribution with k-1 degrees of freedom. The U statistics is compared to the critical value, $h(\alpha)$, where

$$\begin{aligned}
h(\alpha) = & c + \left(\frac{1}{2}\right)(3\chi_4 + \chi_2) \sum_{i=1}^k \frac{\left(1 - \frac{d_i}{D}\right)^2}{v_i} \\
& + \left\{ \left(\frac{1}{16}\right)(3\chi_4 + \chi_2)^2 \left(1 - \frac{(k-3)}{c}\right) \left[\sum_{i=1}^k \frac{\left(1 - \frac{d_i}{D}\right)^2}{v_i} \right]^2 \right. \\
& + \left(\frac{1}{2}\right)(3\chi_4 + \chi_2) \left[(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) \right. \\
& + (2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1) \\
& + \left(\frac{1}{4}\right)(-R_{12}^2 + 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2) \\
& \times (3\chi_4 - 2\chi_2 - 1) \left. \right] \\
& + (R_{23} - 3R_{22} + 3R_{21} - R_{20})(5\chi_6 + 2\chi_4 + \chi_2) \\
& + \left(\frac{3}{16}\right)(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + R_{20})(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2) \\
& + \left(\frac{1}{16}\right)(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2) \\
& \times (9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) \\
& + \left(\frac{1}{4}\right)(-R_{22} + R_{11}^2)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2)
\end{aligned}$$

$$+ \left(\frac{1}{4} \right) (R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2) \} \quad (3.12)$$

and the null hypothesis is rejected if $U > h(\alpha)$.

Applied researchers may feel that one of the main drawbacks to using the J2 procedure is its computational complexity. However, Oshima and Algina (1992a) developed a program written in the SAS language which tests H_0 using this approach.

3.1.4 Box Test

The test has the statistics

$$F' = \frac{\sum_{i=1}^k n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2}{\sum_{i=1}^k \left(1 - \frac{n_i}{N}\right) s_i^2} \quad (3.13)$$

where $\bar{Y}_{..} = \frac{n_i \bar{Y}_{i.}}{N}$ and $N = \sum_{i=1}^k n_i$.

F' is approximately distributed as an F variable with ν'_1 and ν'_2 degrees of freedom where

$$\nu'_1 = \frac{\left[\sum_{i=1}^k (1 - f_i) s_i^2 \right]^2}{\left(\sum_{i=1}^k s_i^2 f_i \right)^2 + \sum_{i=1}^k s_i^4 (1 - 2f_i)} \quad (3.14)$$

$$\nu'_2 = \frac{\left(\sum_{i=1}^k (1 - f_i) s_i^2 \right)}{\frac{\sum_{i=1}^k s_i^4 (1 - f_i)^2}{(n_i - 1)}} \quad (3.15)$$

and $f_i = \frac{n_i}{N}$

3.1.5 Brown Forsythe Test

The procedure presented by Brown and Forsythe is based on a test statistics in which the numerator and denominator have the same expected value under the null hypothesis,

$$B = \frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2}{\sum_{i=1}^k \left(1 - \frac{n_i}{N}\right) s_i^2}. \quad (3.16)$$

This statistics is compared to the critical value $F_{k-1, v, 1-\alpha}$ where

$$v = \frac{\left[\sum_{i=1}^k \left(1 - \frac{n_i}{N}\right) s_i^2 \right]^2}{\sum_{i=1}^k \frac{\left(1 - \frac{n_i}{N}\right)^2 s_i^4}{n_i - 1}} \quad (3.17)$$

While Brown Forsythe procedure may be appealing because of its computational simplicity, its statistical validity has been questioned by Rubin (1983), who suggested a modification for the degrees of freedom $k-1$ that is based on Box's (1954) work. Brown Forsythe procedure is also available in SPSS with 11'th version.

3.1.6 Wilcox Test

The test has the statistics

$$H = \sum_{i=1}^k \frac{(W_i - \bar{W})^2}{\hat{\theta}} \quad (3.18)$$

where

$$\hat{\theta} = \max \left\{ \frac{s_1^2}{n_1}, \frac{s_2^2}{n_2}, \dots, \frac{s_i^2}{n_i} \right\} \quad (3.19)$$

$$\bar{W} = \sum_{i=1}^k \frac{W_i}{k} \quad (3.20)$$

$$W_i = b_i Y_{n_i, i} + \sum_{j=1}^{n_i-1} \frac{(1-b_i) Y_{ji}}{n_i} \quad (3.21)$$

$Y_{n_i, i}$ is the n_i 'th (or last) observation sampled from the i 'th group and

$$b_i = \frac{\left\{ 1 + \left[(n_i - 1)(n_i \hat{\theta} - s_i^2) \right]^{1/2} \right\}}{n_i}. \quad (3.22)$$

If $H > \chi_{k-1, 1-\alpha}$, the $1 - \alpha$ quantile of a chi square distribution with $k-1$ degrees of freedom, reject H_0 .

3.1.7 Krutchkoff K Test

The test has the statistics

$$K = \frac{\sum_{i=1}^k \left[\frac{(\bar{Y}_{i.} - \bar{\bar{Y}})^2 n_i}{s_i^2} \right]}{k-1} \quad (3.23)$$

where

$$\bar{\bar{Y}} = \frac{\left(\sum_{i=1}^k \frac{\bar{Y}_i n_i}{s_i^2} \right)}{\sum_{i=1}^k \frac{n_i}{s_i^2}} \quad (3.24)$$

No attempt has been made to determine the distribution of the K statistics. Critical values of the test has been calculated by using a simulation study and the procedure discussed in Krutchkoff (1985).

3.1.8 Wilcox's 2'nd Test

The test has the statistics

$$H_m = \sum_{i=1}^k D_i (\bar{Y}_i - \bar{Y}_{\cdot})^2 \quad (3.25)$$

where

$$\bar{Y}_i = \frac{Y_{n_i i}}{n_i} + \frac{\sum_{j=1}^{n_i-1} \left(1 - \frac{1}{n_i} \right) Y_{ij}}{n_i + 1} \quad (3.26)$$

$Y_{n_i i}$ is the n_i 'th (or last) observation sampled from the i'th group

$$D_i = \frac{n_i}{s_i^2} \quad (3.27)$$

$$\bar{Y}_{\cdot} = \sum_{i=1}^k \frac{D_i \bar{Y}_i}{W_s} \quad (3.28)$$

$$W_s = \sum_{i=1}^k D_i \quad (3.29)$$

H_m has a chi square distribution with $k-1$ degrees of freedom. Hsiung, Olejnik and Huberty (1994) have been demonstrated that Wilcox 2'nd test for testing the null hypothesis that population means are equal when the population variances differ can result in inflated Type I error rates if sample sizes are unequal and the common population mean differ from 0.

3.1.9 Asiribo and Gurland's Approximate ANOVA Test

Asiribo and Gurland (1990) based their test on

$$F^* = \frac{N - k}{k - 1} \frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2}{\sum_{i=1}^k (n_i - 1) s_i^2}. \quad (3.30)$$

This test gives an approximate solution to the problem of testing equality of means of normal populations in case of heteroscedasticity by making use of the classical ANOVA F test. Under H_0 , the test statistics $F^* / \hat{c}$ is distributed approximately as an F variable with v_1 and v_2 degrees of freedom where

$$\hat{c} = \frac{N - k}{N(k - 1)} \frac{\sum_{i=1}^k (N - n_i) s_i^2}{\sum_{i=1}^k (n_i - 1) s_i^2} \quad (3.31)$$

$$v_1 = \frac{\left[\sum_{i=1}^k \left(1 - \frac{n_i}{N} \right) s_i^2 \right]^2}{\sum_{i=1}^k s_i^4 + \left[\sum_{i=1}^k \frac{n_i s_i^2}{N} \right]^2 - 2 \sum_{i=1}^k \frac{n_i s_i^4}{N}} \quad (3.32)$$

$$v_2 = \frac{\left[\sum_{i=1}^k (n_i - 1) s_i^2 \right]^2}{\sum_{i=1}^k (n_i - 1) s_i^4} \quad (3.33)$$

and H_0 is rejected at level α if $F^* > \hat{c}F_{v_1, v_2, 1-\alpha}$.

3.1.10 Alexander Govern Test

The test has the statistics

$$A = \sum_{i=1}^k Z_i^2 \quad (3.34)$$

where

$$Z_i = c + \frac{(c^3 + 3c)}{b} - \frac{(4c^7 + 33c^5 + 240c^3 + 855c)}{(10b^2 + 8bc^4 + 1000b)} \quad (3.35)$$

$$a = v_i - 0,5; \quad b = 48a^2; \quad c = \left[a \ln \left(1 + \frac{t_i^2}{v_i} \right) \right]^{1/2} \quad (3.36)$$

and t_i is Student t distributed random variable calculated for each treatment. The test statistics is approximately distributed χ^2 with k-1 degrees of freedom. Decision rule that the tests uses is reject the null hypothesis of equal means if A exceeds the 1- α quantile of a chi-square distribution with k-1 degrees of freedom.

3.1.11 Modified Brown Forsythe Test

Mehrotra (1997) developed the following test statistics

$$B^* = \frac{\sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y}_{..})^2}{\sum_{i=1}^k \left(1 - \frac{n_i}{N} \right) s_i^2} \quad (3.37)$$

in an attempt to correct a flaw in the original Brown Forsythe test. The flaw in the Brown Forsythe testing procedure, identified by Mehrotra (1997), is in the specification of the numerator degrees of freedom. Specifically, Brown Forsythe used $k-1$ numerator degrees of freedom whereas Mehrotra (1997) used a Box (1954) approximation to obtain the numerator degrees of freedom, v_1 , where

$$v_1 = \frac{\left[\sum_{i=1}^k \left(1 - \frac{n_i}{N} \right) s_i^2 \right]^2}{\sum_{i=1}^k s_i^4 + \left[\sum_{i=1}^k \frac{n_i s_i^2}{N} \right]^2 - 2 \sum_{i=1}^k \frac{n_i s_i^4}{N}} \quad (3.38)$$

and denominator degrees of freedom

$$v = \frac{\left[\sum_{i=1}^k \left(1 - \frac{n_i}{N} \right) s_i^2 \right]^2}{\sum_{i=1}^k \frac{\left(1 - \frac{n_i}{N} \right)^2 s_i^4}{n_i - 1}}. \quad (3.39)$$

Under H_0 , B^* is distributed approximately as an F variable with v_1 and v degrees of freedom. The null hypothesis H_0 is rejected at level α if $B^* > F_{v_1, v, 1-\alpha}$.

3.1.12 Adjusted Welch Test

It is known that the Welch test becomes too liberal especially with small samples and increasing number of groups. Böckenhoff and Hartung (1998) have examined

the weight of Welch test $\omega_i = \frac{n_i}{s_i^2}$ and offered a new weight

$$\omega_i^* = \frac{n_i}{(\varphi_i s_i^2)} \quad (3.40)$$

where $1 \leq \varphi_i \leq c_i$ and

$$c_i = \frac{(n_i - 1)}{(n_i - 3)} \quad (3.41)$$

This weight defines a new class of Welch type test statistics whose properties can be adjusted accordingly by choosing the control parameter, φ_i , appropriately. The statistics of the test is given by

$$W^* = \frac{\sum_{i=1}^k \omega_i^* \left(\bar{Y}_{i.} - \sum_{i=1}^k h_i^* \bar{Y}_{i.} \right)^2}{(k-1) + 2(k-2)(k+1)^{-1} \sum_{i=1}^k (n_i - 1)^{-1} (1 - h_i^*)^2} \quad (3.42)$$

where

$$h_i^* = \frac{\omega_i^*}{\sum_{i=1}^k \omega_i^*} \quad (3.43)$$

Under H_0 , the adjusted Welch statistic, W^* , is distributed approximately as an F variable with $k-1$ and v_W^* degrees of freedom, with

$$v_W^* = \frac{k^2 - 1}{3 \sum_{i=1}^k (n_i - 1)^{-1} (1 - h_i^*)^2} \quad (3.44)$$

and H_0 is rejected at α level if $W^* > F_{k-1, v_W^*, 1-\alpha}$. When the sample sizes are large,

W^* approaches the Welch test.

3.1.13 Smith and Hickling Test

Smith and Hickling (1998) derived a formula to conduct a one-way ANOVA when the data are heteroscedastic, especially when the sample sizes may be small and unequal. The method they developed is especially for the off-line experiments which are carried out in off-line quality control procedures to establish the appropriate settings at which to run the process. Test statistic of the procedure is

$$Q = \frac{\sum_{1 \leq i < j \leq k} (\mu_i - \mu_j)^2 \gamma_1 \gamma_2 \dots \gamma_{i-1} \gamma_{i+1} \dots \gamma_{j-1} \gamma_{j+1} \dots \gamma_k}{\sum_i \gamma_1 \gamma_2 \dots \gamma_{i-1} \gamma_{i+1} \dots \gamma_k} \quad (3.45)$$

where

$$\gamma_i = \sigma_i^2 / n_i, \quad a = a^* - 1, \quad a^* = \max[k, \hat{a}], \quad \hat{a} = \frac{k\Delta}{2\Delta + k(k+2)} \quad (3.46)$$

and

$$b = \max[\hat{b}, 4 + \varepsilon], \quad \hat{b} = 4 + \frac{2(k-2)^2(k+1)}{(k-1)\Delta - k(k+1)}, \quad \Delta = \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{n_i n_j + n_i + n_j}{(n_i - 1)(n_j - 1)} \quad (3.47)$$

Then the test statistics is distributed as $Q \sim aF_{a,b}$

3.2 Exact Tests

Under a true null hypothesis, the sampling distribution of exact test's test statistics would be given exactly by a well-known sampling distribution like F, t or χ^2 . All tests belongs to this group involves a "two-stage procedures". The first stage consist of randomly sampling some observations to obtain sample variances. Once the sample variances are available, the sample sizes needed to match the power requirements can be determined. The second stage consist of sampling the additional observations and then performing the test of whatever hypothesis happens to be of interest. In many cases, obtaining additional observations is impractical but the

results in this section are still useful when considering whether the available sample sizes are large enough to achieve reasonably high power (Wilcox R.R, 1996). It can be shown that to control power in terms of the means only, a two-stage procedure must be used (Dantzig, 1940). Exact tests offered as an alternative to one-way ANOVA are as follows;

- Bishop-Dudewicz Exact ANOVA with unequal variances (1978).
- Wilcox R.R. A heteroscedastic ANOVA procedure with specified power (1987).
- Chen-Chen-Ding Two stage ANOVA test (2000).

The most actual of these tests is two stage ANOVA by Chen-Chen and Ding (2000)

3.2.1 Chen-Chen-Ding Two stage ANOVA Test

The most actual test of this section constructs H_0 in a different way like

$$H_0 = \frac{\sum_{i=1}^k (\mu_i - \bar{\mu})^2}{k} \leq \delta^2$$

$$H_a = \frac{\sum_{i=1}^k (\mu_i - \bar{\mu})^2}{k} > \delta^2 \quad (3.48)$$

where $\bar{\mu}$ is the average of $\mu_1, \mu_2, \dots, \mu_k$ and $\delta (\geq 0)$: zone of indifference which must be specified in advance by an expert in his experiment . δ can be interpreted as the amount of variation among means about which we are indifferent. Two – stage sampling procedure and test statistics is given as follows;

1. Choose α , P^* and $z > 0$ (z is determined by the required power of the test) and take an initial random sample of size n_0 (at least 2, but 10 or more will give better results) from each of the k populations and define

$$N_i = \max \left\{ n_0 + 1, \left\lceil \frac{s_i^2}{z} + 1 \right\rceil \right\} \quad (3.49)$$

$\left\lceil \frac{s_i^2}{z} + 1 \right\rceil$ denotes the greatest integer less than or equal to $\frac{s_i^2}{z} + 1$ and s_i^2

denotes the usual unbiased estimate of σ_i^2 based on the initial n_0 observations.

2. Take $N_i - n_0$ additional random observations from the i 'th population so we have a total of N_i observations denoted by $Y_{i1}, \dots, Y_{in_0}, \dots, Y_{iN_i}$
3. For each i set the coefficients

$$a_{i1}, \dots, a_{in_0}, \dots, a_{iN_i} = \frac{1 - (N_i - n_0)b_i}{n_0} = a_i \quad (3.50)$$

$$a_{i,n_0+1} = \dots = a_{iN_i} = \frac{1}{N_i} \left[1 + \sqrt{\frac{n_0(N_i z - s_i^2)}{(N_i - n_0)s_i^2}} \right] = b_i \quad (3.51)$$

are then calculated.

4. Compute the weighted mean as

$$\tilde{X}_{i.} = a_i \sum_{j=1}^{n_0} Y_{ij} + b_i \sum_{j=n_0+1}^{N_i} Y_{ij} \quad (3.52)$$

5. Finally compute the test statistic as

$$\tilde{F} = \sum_{i=1}^k \frac{(\tilde{X}_{i.} - \tilde{X}_{..})^2}{z} \quad (3.53)$$

where $\tilde{X}_{..}$ is the arithmetic mean of $\tilde{X}_{1.}, \dots, \tilde{X}_{k.}$ and we reject H_0 at level α if and only if $\tilde{F} > \tilde{F}_\alpha$ where the level α is predetermined.

3.3 Nonparametric Tests

During 1950's and early 1960's a tendency to non-parametric statistics was experienced. Procedures especially the ones based on ranks are appealing to some researchers because they are not dependent on the assumption of normal population distribution. Recent research shows that parametric approximations are generally superior to nonparametric approaches in all but a few cases of nonnormality (Keselman, Rogan & Feir-Walsh, 1977; Tomarken & Serlin, 1986). Some of the nonparametric tests offered as an alternative for one-way ANOVA F test are;

- Kruskal-Wallis test (1952)
- Van Der Waerden Inverse Normal Score test (1952)
- McSweeney-Penfield Normal Score test (1969)
- Rust-Fligner test (1984)

The most well known nonparametric alternative to the ANOVA F test is proposed by Kruskal and Wallis (1952).

3.3.1 Kruskal-Wallis Test

The chief advantage of adopting the Kruskal-Wallis test instead of F test the potential for substantially greater statistical power when the data are nonnormal in shape (Blair & Higgins, 1980). The test has the statistics

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1) \sim \chi_{k-1}^2 \quad (3.54)$$

where

$$n = n_1 + n_2 + \dots + n_k, \quad R_i = \sum_{j=1}^{n_i} R_{ij} \quad (3.55)$$

and H_0 is rejected at α level if $H > \chi_{k-1, 1-\alpha}^2$. The Kruskal-Wallis method is not difficult to use in practise. It is available in most of the statistical software packages.

3.4 Data Transformations

To apply a transformation to the data and proceed with use of the F test is the most traditional way of dealing with variance heterogeneity. Data transformations have two major problems. First, the appropriate transformation should be estimated with a suitable parameter such as log or square root. But one can never prove which transformation is exactly right for the data at hand. Second, the ANOVA is conducted on the transformed data, and there is no easy way to turn back in which one can apply an appropriate interpretation of the results to the original data.

Most of the variance stabilizing transformations are derived from the relation between standard deviation and mean of the response variable. When the standard deviation of response variable Y is a function of its mean such that

$$\sigma_y = f(\mu_y) \quad (3.56)$$

a transformation $h(y)$ to stabilize the variance can be found as follows. A first order Taylor Series approximation of $h(y)$ around μ_y

$$h(y) \approx h(\mu_y) + (y - \mu_y)h'(\mu_y) \quad (3.57)$$

variances of both sides

$$\begin{aligned} \text{Var}[h(y)] &\approx \text{Var}[h'(\mu_y)(y - \mu_y)] \\ \text{Var}[h(y)] &= [h'(\mu_y)]^2 \text{Var}(y) = [h'(\mu_y)]^2 f^2(\mu_y) \\ &= [h'(\mu_y)f(\mu_y)]^2 \end{aligned} \quad (3.58)$$

if

$$h'(\mu_y) = \frac{1}{f(\mu_y)} \quad (3.59)$$

then

$$Var[h(y)] \approx \left[\frac{1}{f(\mu_y)} f(\mu_y) \right]^2 = 1 \quad (3.60)$$

Thus, the variance stabilizing transformation is the $h(y)$ that satisfies the following equation.

$$h'(\mu_y) \propto \frac{1}{f(\mu_y)} \quad (3.61)$$

For example

$$\sigma_y \propto \mu_y^{\frac{1}{2}} \Rightarrow h(y) = y^{\frac{1}{2}}$$

$$\sigma_y \propto \mu_y \Rightarrow h(y) = \ln y$$

$$\sigma_y \propto \mu_y^{\frac{3}{2}} \Rightarrow h(y) = y^{\frac{-1}{2}}$$

$$\sigma_y \propto \mu_y^2 \Rightarrow h(y) = y^{-1}$$

Variance stabilizing transformations are generally used to symmetrize the distribution of response variable as well. When the power of transformation is less than 1 transformations is called as concave and used to symmetrize the right skewed distributions. When the power is greater than one transformations is called as convex and used to symmetrize the left skewed distributions. Generally the effect of the transformation is greater as the ratio $Y_{\max} / Y_{\min}$ gets large and as the power differs from unity. Effect of contracting transformations can be observed at Figure 3.1. For example when a square root transformation is applied to the series 1, 1.5, 2, 2.5, 3,

3.5, 4. which has a range 3, the series becomes 1, 1.22, 1.41, 1.58, 1.73, 1.87, 2 with a range 1.

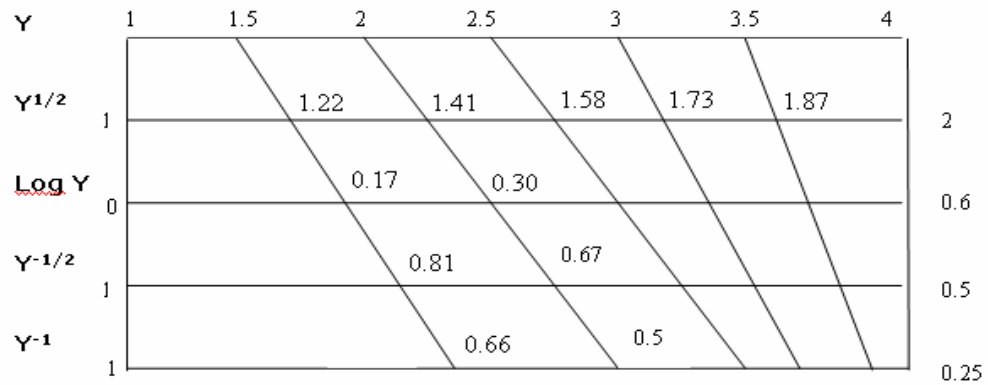


Figure 3.1 Effect of contracting transformations

3.5 Weighted Least Square Estimation Method

If the error terms are normally distributed but the variance of the error terms is not constant, weighted least square estimation method is a standard remedial measure which uses the regression approach to the analysis of variance. The theory behind this method is based on the assumption that weights $\left(\omega_{ij} = \frac{1}{\sigma_i^2} \right)$ are known exactly.

This is almost never the case in real applications. When the weights are especially estimated from small numbers of replicated observations, the result of an analysis can be very badly and unpredictably affected. It is important to remain aware of this potential problem and to only use weighted least squares when the weights can be estimated precisely relative to one another (Carroll, Rupert, 1988).

3.6 Robust Statistical Procedures

The advances and insights achieved during the last half century in statistics provide an opportunity for substantially improving research. Recently developed

methods can provide substantially more power when the standard assumptions of normality and homoscedasticity are violated. They also help deepen our understanding of how groups differ. The theoretical and practical advantages of modern technology have been documented in several books (Hampel, Ronchetti, Rousseeuw & Stahel, 1986; Hoaglin, Mostelier & Tukey, 1983, 1985; Huber, 1981; Tiku, Tan & Balakrishnan 1986; Rousseeuw & Leroy, 1987; Wilcox, 1997a, 2001, 2003) and journal articles.

Many treatises have appeared on the topic of substituting robust measures of location such as 20% trimmed means or M-estimators for the usual least squares estimator (least square means). Indeed, many investigators have demonstrated that one can achieve better control over Type I errors when robust estimators are substituted for least square estimators in a heteroscedastic statistics such as Johanson's (1980) Welch-James type test (Guo & Luh, 2000; Keselman, Kowalchuk & Lix, 1998). There are many studies in which robust measure of location and scale were used. Some of them which have recently been published are here

- A one-way random effects model for trimmed means (Wilcox, 1994)
- Introduction to robust estimation and hypothesis testing (Wilcox, 1997)
- Multiple comparison procedures for trimmed means (Keselman, Lix & Kowalchuk, 1998)
- To trim or not to trim: Tests of location equality under heteroscedasticity and nonnormality (Lix & Keselman, 1998).
- A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under nonnormality and inequality of variances (Luh & Guo, 1999).
- Fundamentals of modern statistical methods: Substantially improving power and accuracy (Wilcox, 2001).
- Repeated measures one-way ANOVA based on a modified one-step M-estimator (Wilcox & Keselman, 2003).

CHAPTER FOUR

NEW AND SIMPLE APPROXIMATION PROCEDURES AND SIMULATION STUDIES

4.1 A New and Simple Approximation Procedure and a Numerical Application

This part of the study includes the presentation of a new and simple approximation procedure which intends to create an easy and applicable alternative for the researchers under heterogeneity of variance problem in one-way fixed effects analysis of variance model. The purpose of this new approximation procedure is to preserve the actual Type I error rate at a level determined by the researcher and to increase the power as well.

Consider k independent randomly sampled groups each measured on a normally distributed random variable (Y). Variances of the populations and sample size of the groups need not to be equal. Each of the k samples of size n_i will have sample mean $\bar{Y}_i$ and each of the means will have a standard error $s_{\bar{Y}_i}$.

$$s_{\bar{Y}_i} = \left[\frac{\sum_{j=1}^{n_i} (Y_j - \bar{Y}_i)^2}{n_i(n_i - 1)} \right]^{1/2} \quad (4.1)$$

For each of the k samples, if we define a weight (ω_i) such that $\sum_{i=1}^k \omega_i = 1$

$$\omega_i = \frac{\frac{1}{s_{\bar{Y}_i}^2}}{\sum_{i=1}^k \left(\frac{1}{s_{\bar{Y}_i}^2} \right)} \quad (4.2)$$

The variance-weighted estimate of the common mean (Y^+) of Y becomes

$$Y^+ = \sum_{i=1}^k \omega_i \bar{Y}_i. \quad (4.3)$$

Finally, for each of the k groups a one sample t statistics is calculated as

$$t_i = \frac{\bar{Y}_i - Y^+}{s_{\bar{Y}_i}} \quad (4.4)$$

Under the usual assumptions, each of the t_i will be distributed as Student's t with $\nu_i = n_i - 1$ degrees of freedom. After this step, this new approximation uses a normalizing transformation on each of the sample t_i values directly in order to transform the each t_i values in to standard normal deviate z_i . Several normalizing transformations for the t statistics have appeared in the literature. One of them is “Accurate Normalizing Transformations of a Student's t Variate” Bailey B.J.R (1980). Bailey offers a locally accurate normalizing transformation which is given here

$$z = \pm \frac{4\nu_i^2 + \frac{5(2z_c^2 + 3)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2 + 9)}{12}} \nu_i^{1/2} \left\{ \ln \left(1 + \frac{t_i^2}{\nu_i} \right) \right\}^{1/2} \quad (4.5)$$

where ν_i =degrees of freedom of t_i variate and z_c is the critical value of related significance level under standard normal distribution. Since χ_{k-1}^2 is simply the mathematical description of the sum of k independent squared standard normal deviates,

$$B^2 = \sum_{i=1}^k z_i^2 = \sum_{i=1}^k \left(\frac{4\nu_i^2 + \frac{5(2z_c^2 + 3)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2 + 9)}{12}} \nu_i^{1/2} \left\{ \ln \left(1 + \frac{\left(\frac{\bar{Y}_i - Y^+}{s_{\bar{Y}_i}} \right)^2}{\nu_i} \right) \right\}^{1/2} \right)^2 \quad (4.6)$$

will be distributed approximately χ^2 with $k - 1$ degrees of freedom. Decision rule that the tests uses is reject the null hypothesis of of equal means if B^2 exceeds the $1 - \alpha$ quantile of a chi-square distribution with $k - 1$ degrees of freedom.

Let us call the first part of the z transformation as coefficient c then the transformation can be written as

$$z = \pm c \left\{ \ln \left(1 + \frac{t_i^2}{\nu_i} \right) \right\}^{1/2} \quad (4.7)$$

where

$$c = \frac{4\nu_i^2 + \frac{5(2z_c^2 + 3)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2 + 9)}{12}} \nu_i^{1/2} \quad (4.8)$$

then

$$B^2 = \sum_{i=1}^k z_i^2 = \sum_{i=1}^k \left(c \left\{ \ln \left(1 + \frac{\left(\frac{\bar{Y}_i - Y^+}{s_{\bar{Y}_i}} \right)^2}{\nu_i} \right) \right\}^{1/2} \right)^2 \quad (4.9)$$

can be written with the help of Table 4.1 which was prepared by the owners of the study.

Table 4.1 Coefficient of c up to 30 degrees of freedom for $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$

df	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$	df	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$	df	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$
1	0.92814	0.88552	0.86478	11	3.24622	3.24453	3.24384	21	4.52985	4.52923	4.52898
2	1.30814	1.28679	1.27752	12	3.39627	3.39480	3.39420	22	4.63882	4.63824	4.63800
3	1.62589	1.61365	1.60852	13	3.54005	3.53874	3.53821	23	4.74530	4.74475	4.74453
4	1.89933	1.89134	1.88803	14	3.67826	3.66710	3.67662	24	4.84944	4.84893	4.84873
5	2.14138	2.13569	2.13335	15	3.81150	3.81046	3.81004	25	4.95140	4.95093	4.95073
6	2.36026	2.35597	2.35420	16	3.94028	3.93934	3.93895	26	5.05131	5.05086	5.05068
7	2.56133	2.55794	2.55656	17	4.06501	4.06414	4.06379	27	5.14929	5.14886	5.14869
8	2.74821	2.74545	2.74432	18	4.18604	4.18525	4.18492	28	5.24544	5.24504	5.24487
9	2.92346	2.92116	2.92022	19	4.30368	4.30295	4.30266	29	5.33986	5.33948	5.33932
10	3.08898	3.08703	3.08624	20	4.41821	4.41754	4.41726	30	5.43265	5.43228	5.43214

A numerical application has been carried out in order to see whether variables t_i , z and B^2 are really distributed as Student t with $v_i = n_i - 1$ degrees of freedom, standard normal and chi square with $k - 1$ degrees of freedom, respectively. New approximation procedure was run 1000 times for $k = 3$ groups with the parameters given in Table 4.2. Then, values which are supposed to follow Student's t with $v_i = n_i - 1 = 5 - 1 = 4$ degrees of freedom (for the first treatment), standard normal and chi square with $k - 1 = 3 - 1 = 2$ degrees of freedom are stored in separate columns by using statistical software MINITAB 13.

Table 4.2 Parameters of the numerical application

i	1	2	3
n_i	5	10	15
σ_i^2	10	6	2

Stored values are then tested with Anderson Darling and Kolmogorov Smirnov goodness of fit tests in order to see the degree of the fit with the specified theoretical sampling distributions at significance level $\alpha = 0.05$ by using probability distribution fitting software BestFit 4.5. Results were given in Tables 4.3, 4.4 and 4.5

Table 4.3 Goodness of fit test results for values which are supposed to follow student's t distribution with 4 degrees of freedom

	Anderson Darling Test	Kolmogorov Smirnov Test
Test value	1.588	0.02286
p value	$0.15 \leq p \leq 0.25$	> 0.25

Table 4.4 Goodness of fit test results for values which are supposed to follow standard normal distribution

	Anderson Darling Test	Kolmogorov Smirnov Test
Test value	1.499	0.02232
p value	$0.15 \leq p \leq 0.25$	> 0.25

Table 4.5 Goodness of fit test results for values which are supposed to follow chi square distribution with 2 degrees of freedom (Distribution of the B^2 test statistics)

	Anderson Darling Test	Kolmogorov Smirnov Test
Test value	0.5622	0.03084
p value	> 0.25	0.1484

4.2 Simulation Study

Performance of the conventional ANOVA F test, Welch test (W), Kruskal-Wallis test (KW), and new approximation B^2 procedures have been examined by means of a simulation study. Simulated actual significance level and power have been obtained using different sample sizes and error variances for $k=3$, $k=6$ and $k=9$ groups by using nominal significance level 0.05. All means were equal to 0 when predicting actual significance level. Two different configurations of mean differences were used when assessing the power of the tests. The first pattern was $\mu_1=2$, $\mu_2=0$, $\mu_3=0$ while the second pattern was $\mu_1=-1$, $\mu_2=0$, $\mu_3=1$ in which the means were equally spaced. Each configuration was executed 10000 times. All data were generated from normal

distribution with parameters which are specified in Table 4.6 and Table 4.7 by using the 13'th version of statistical software MINITAB. Results of the simulations study were given in a series of tables through Table 4.8 to Table 4.14

Table 4.6 Sample sizes and variances of the distributions for $k=3$ and $k=6$

Pattern	i	k = 3			k = 6					
		1	2	3	1	2	3	4	5	6
A ₁	n _i	5	5	5	5	5	5	5	5	5
	σ_i^2	4	4	4	4	4	4	4	4	4
A ₂	n _i	5	5	5	5	5	5	5	5	5
	σ_i^2	2	6	10	2	6	10	2	6	10
B ₁	n _i	10	10	10	10	10	10	10	10	10
	σ_i^2	4	4	4	4	4	4	4	4	4
B ₂	n _i	10	10	10	10	10	10	10	10	10
	σ_i^2	2	6	10	2	6	10	2	6	10
C ₁	n _i	5	10	15	5	10	15	5	10	15
	σ_i^2	4	4	4	4	4	4	4	4	4
C ₂	n _i	5	10	15	5	10	15	5	10	15
	σ_i^2	2	6	10	2	6	10	2	6	10
C ₃	n _i	5	10	15	5	10	15	5	10	15
	σ_i^2	10	6	2	10	6	2	10	6	2
D ₁	n _i	10	20	30	10	20	30	10	20	30
	σ_i^2	4	4	4	4	4	4	4	4	4
D ₂	n _i	10	20	30	10	20	30	10	20	30
	σ_i^2	2	6	10	2	6	10	2	6	10
D ₃	n _i	10	20	30	10	20	30	10	20	30
	σ_i^2	10	6	2	10	6	2	10	6	2

Table 4.7 Sample sizes and variances of the distributions for $k=9$

Pattern		k = 9								
	i	1	2	3	4	5	6	7	8	9
A₁	n_i	5	5	5	5	5	5	5	5	5
	σ_i^2	4	4	4	4	4	4	4	4	4
A₂	n_i	5	5	5	5	5	5	5	5	5
	σ_i^2	2	6	10	2	6	10	2	6	10
B₁	n_i	10	10	10	10	10	10	10	10	10
	σ_i^2	4	4	4	4	4	4	4	4	4
B₂	n_i	10	10	10	10	10	10	10	10	10
	σ_i^2	2	6	10	2	6	10	2	6	10
C₁	n_i	5	10	15	5	10	15	5	10	15
	σ_i^2	4	4	4	4	4	4	4	4	4
C₂	n_i	5	10	15	5	10	15	5	10	15
	σ_i^2	2	6	10	2	6	10	2	6	10
C₃	n_i	5	10	15	5	10	15	5	10	15
	σ_i^2	10	6	2	10	6	2	10	6	2
D₁	n_i	10	20	30	10	20	30	10	20	30
	σ_i^2	4	4	4	4	4	4	4	4	4
D₂	n_i	10	20	30	10	20	30	10	20	30
	σ_i^2	2	6	10	2	6	10	2	6	10
D₃	n_i	10	20	30	10	20	30	10	20	30
	σ_i^2	10	6	2	10	6	2	10	6	2

Table 4.8 Actual significance levels for different configurations of $k = 3$ groups

Pattern	n	σ^2	F	W	KW	B^2
A_1	5,5,5	4,4,4	0.051	0.052	0.044	0.049
A_2	5,5,5	2,6,10	0.059	0.059	0.054	0.049
B_1	10,10,10	4,4,4	0.047	0.050	0.048	0.049
B_2	10,10,10	2,6,10	0.054	0.053	0.053	0.048
C_1	5,10,15	4,4,4	0.068	0.063	0.046	0.050
C_2	5,10,15	2,6,10	0.029	0.055	0.029	0.049
C_3	5,10,15	10,6,2	0.150	0.067	0.082	0.053
D_1	10,20,30	4,4,4	0.069	0.051	0.045	0.051
D_2	10,20,30	2,4,6	0.026	0.049	0.029	0.048
D_3	10,20,30	10,6,2	0.160	0.052	0.088	0.049

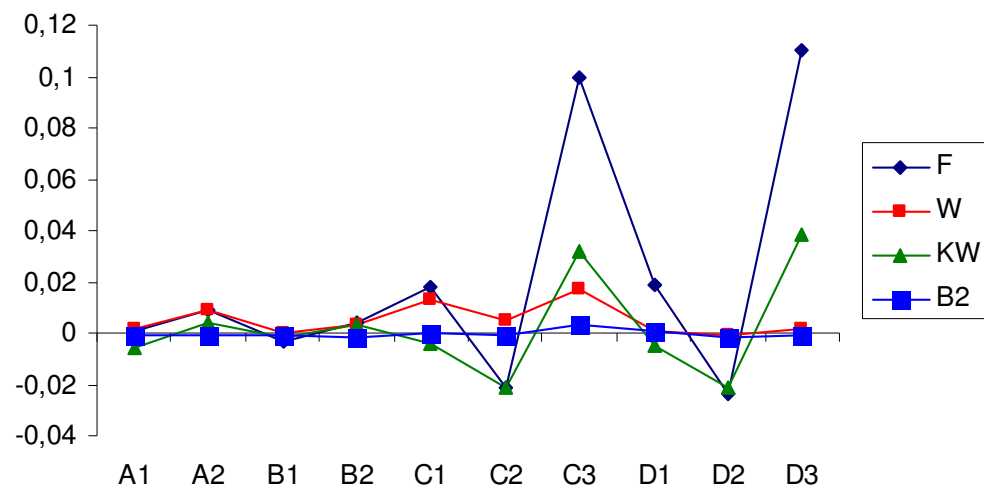
Figure 4.1 Discrepancy between nominal and actual significance level for $k = 3$ groups

Table 4.9 Actual significance levels for different configurations of $k = 6$ groups

Pattern	n	σ^2	F	W	KW	B ²
A ₁	5,5,5,5,5,5	4,4,4,4,4,4	0.051	0.080	0.036	0.051
A ₂	5,5,5,5,5,5	2,6,10,2,6,10	0.064	0.087	0.045	0.051
B ₁	10,10,10,10,10,10	4,4,4,4,4,4	0.049	0.058	0.041	0.048
B ₂	10,10,10,10,10,10	2,6,10,2,6,10	0.059	0.060	0.053	0.049
C ₁	5,10,15,5,10,15	4,4,4,4,4,4	0.061	0.063	0.042	0.054
C ₂	5,10,15,5,10,15	2,6,10,2,6,10	0.025	0.057	0.027	0.054
C ₃	5,10,15,5,10,15	10,6,2,10,6,2	0.180	0.063	0.091	0.055
D ₁	10,20,30,10,20,30	4,4,4,4,4,4	0.061	0.057	0.050	0.048
D ₂	10,20,30,10,20,30	2,6,10,2,6,10	0.026	0.047	0.028	0.051
D ₃	10,20,30,10,20,30	10,6,2,10,6,2	0.160	0.052	0.100	0.049

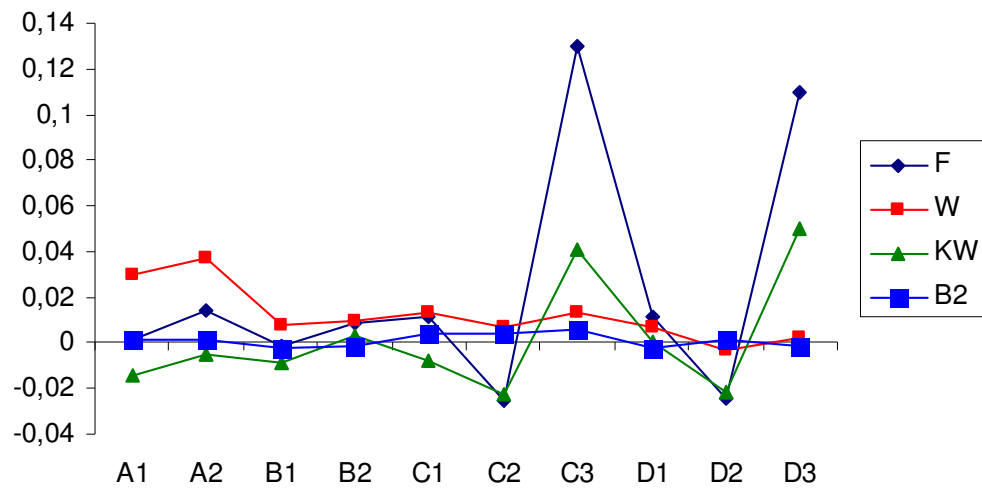
Figure 4.2 Discrepancy between nominal and actual significance level for $k = 6$ groups

Table 4.10 Actual significance levels for different configurations of $k = 9$ groups

Pattern	n	σ^2	F	W	KW	B ²
A ₁	5,5,5,5,5,5,5,5,5	4,4,4,4,4,4,4,4,4	0.049	0.098	0.029	0.048
A ₂	5,5,5,5,5,5,5,5,5	2,6,10,2,6,10,2,6,10	0.060	0.100	0.036	0.046
B ₁	10,10,10,10,10,10,10,10,10	4,4,4,4,4,4,4,4,4	0.049	0.064	0.041	0.047
B ₂	10,10,10,10,10,10,10,10,10	2,6,10,2,6,10,2,6,10	0.067	0.059	0.051	0.044
C ₁	5,10,15,5,10,15,5,10,15	4,4,4,4,4,4,4,4,4	0.055	0.070	0.039	0.046
C ₂	5,10,15,5,10,15,5,10,15	2,6,10,2,6,10,2,6,10	0.023	0.063	0.024	0.045
C ₃	5,10,15,5,10,15,5,10,15	10,6,2,10,6,2,10,6,2	0.200	0.077	0.098	0.046
D ₁	10,20,30,10,20,30,10,20,30	4,4,4,4,4,4,4,4,4	0.051	0.055	0.043	0.047
D ₂	10,20,30,10,20,30,10,20,30	2,6,10,2,6,10,2,6,10	0.027	0.050	0.025	0.049
D ₃	10,20,30,10,20,30,10,20,30	10,6,2,10,6,2,10,6,2	0.200	0.055	0.100	0.046

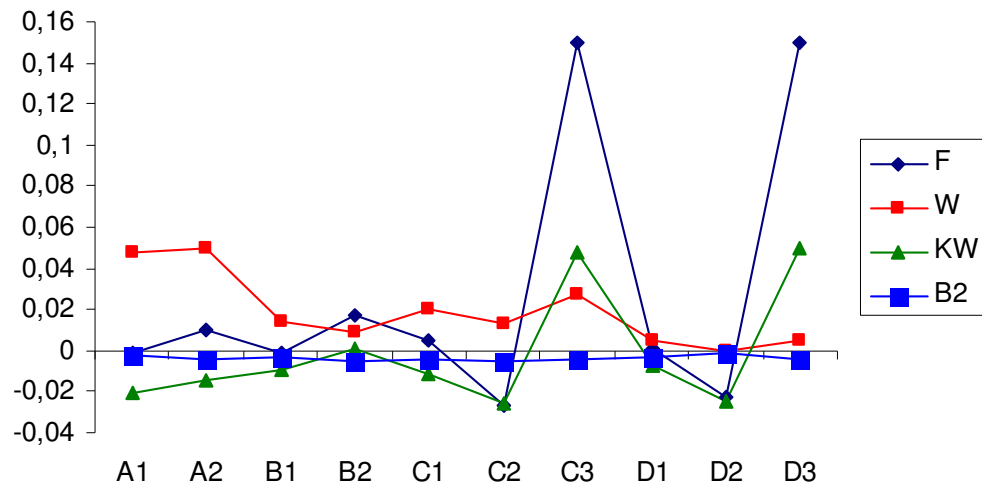
Figure 4.3 Discrepancy between nominal and actual significance level for $k = 9$ groups

Table 4.11 Actual significance levels for greater population variances and different configurations of $k = 3$ groups

Pattern	n	σ_i^2	F	W	KW	B ²
A ₁ *	5,5,5	256,256,256	0.051	0.055	0.044	0.048
A ₂ *	5,5,5	4,100,256	0.081	0.077	0.076	0.055
B ₁ *	10,10,10	256,256,256	0.049	0.053	0.048	0.049
B ₂ *	10,10,10	4,100,256	0.078	0.056	0.070	0.050
C ₁ *	5,10,15	256,256,256	0.069	0.059	0.047	0.052
C ₂ *	5,10,15	4,100,256	0.026	0.055	0.026	0.049
C ₃ *	5,10,15	256,100,4	0.245	0.067	0.128	0.050
D ₁ *	10,20,30	256,256,256	0.070	0.050	0.054	0.048
D ₂ *	10,20,30	4,100,256	0.025	0.049	0.029	0.048
D ₃ *	10,20,30	256,100,4	0.233	0.052	0.142	0.049

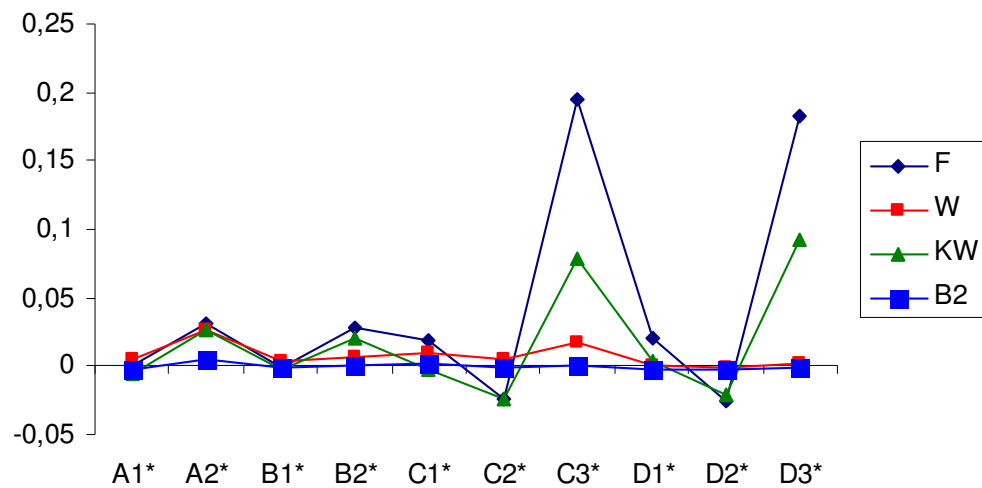


Figure 4.4 Discrepancy between nominal and actual significance level for greater population variances of $k = 3$ groups

Table 4.12 Predicted power values for different configurations of $k = 3$ groups

Pattern	n	σ^2	μ_i	F	W	KW	B^2
A ₁₁	5,5,5	4,4,4	2,0,0	0.27	0.24	0.24	0.24
A ₁₂	5,5,5	4,4,4	-1,0,1	0.21	0.18	0.19	0.18
A ₂₁	5,5,5	2,6,10	2,0,0	0.20	0.25	0.20	0.26
A ₂₂	5,5,5	2,6,10	-1,0,1	0.17	0.17	0.15	0.16
B ₁₁	10,10,10	4,4,4	2,0,0	0.58	0.55	0.53	0.55
B ₁₂	10,10,10	4,4,4	-1,0,1	0.45	0.43	0.42	0.42
B ₂₁	10,10,10	2,6,10	2,0,0	0.42	0.57	0.48	0.57
B ₂₂	10,10,10	2,6,10	-1,0,1	0.31	0.35	0.30	0.35
C ₁₁	5,10,15	4,4,4	2,0,0	0.46	0.35	0.33	0.31
C ₁₂	5,10,15	4,4,4	-1,0,1	0.46	0.34	0.35	0.35
C ₂₁	5,10,15	2,6,10	2,0,0	0.21	0.46	0.21	0.44
C ₂₂	5,10,15	2,6,10	-1,0,1	0.21	0.34	0.19	0.33
C ₃₁	5,10,15	10,6,2	2,0,0	0.44	0.18	0.27	0.16
C ₃₂	5,10,15	10,6,2	-1,0,1	0.46	0.24	0.32	0.26
D ₁₁	10,20,30	4,4,4	2,0,0	0.77	0.68	0.66	0.64
D ₁₂	10,20,30	4,4,4	-1,0,1	0.77	0.68	0.68	0.68
D ₂₁	10,20,30	2,6,10	2,0,0	0.51	0.82	0.57	0.81
D ₂₂	10,20,30	2,6,10	-1,0,1	0.48	0.65	0.45	0.64
D ₃₁	10,20,30	10,6,2	2,0,0	0.66	0.35	0.45	0.32
D ₃₂	10,20,30	10,6,2	-1,0,1	0.71	0.51	0.56	0.51

Table 4.13 Predicted power values for different configurations of $k = 6$ groups

Pattern	n	σ^2	μ_i	F	W	KW	B^2
A ₁₁	5,5,5,5,5,5	4,4,4,4,4,4	2,0,0,2,0,0	0.39	0.35	0.31	0.33
A ₁₂	5,5,5,5,5,5	4,4,4,4,4,4	-1,0,1,-1,0,1	0.29	0.26	0.22	0.25
A ₂₁	5,5,5,5,5,5	2,6,10,2,6,10	2,0,0,2,0,0	0.26	0.34	0.72	0.35
A ₂₂	5,5,5,5,5,5	2,6,10,2,6,10	-1,0,1,-1,0,1	0.22	0.22	0.57	0.20
B ₁₁	10,10,10,10,10,10	4,4,4,4,4,4	2,0,0,2,0,0	0.77	0.72	0.25	0.73
B ₁₂	10,10,10,10,10,10	4,4,4,4,4,4	-1,0,1,-1,0,1	0.62	0.57	0.17	0.58
B ₂₁	10,10,10,10,10,10	2,6,10,2,6,10	2,0,0,2,0,0	0.58	0.75	0.64	0.76
B ₂₂	10,10,10,10,10,10	2,6,10,2,6,10	-1,0,1,-1,0,1	0.42	0.47	0.41	0.54
C ₁₁	5,10,15,5,10,15	4,4,4,4,4,4	2,0,0,2,0,0	0.62	0.49	0.44	0.42
C ₁₂	5,10,15,5,10,15	4,4,4,4,4,4	-1,0,1,-1,0,1	0.62	0.47	0.47	0.48
C ₂₁	5,10,15,5,10,15	2,6,10,2,6,10	2,0,0,2,0,0	0.27	0.62	0.28	0.61
C ₂₂	5,10,15,5,10,15	2,6,10,2,6,10	-1,0,1,-1,0,1	0.27	0.44	0.24	0.45
C ₃₁	5,10,15,5,10,15	10,6,2,10,6,2	2,0,0,2,0,0	0.59	0.25	0.39	0.20
C ₃₂	5,10,15,5,10,15	10,6,2,10,6,2	-1,0,1,-1,0,1	0.62	0.34	0.43	0.34
D ₁₁	10,20,30,10,20,30	4,4,4,4,4,4	2,0,0,2,0,0	0.98	0.84	0.84	0.82
D ₁₂	10,20,30,10,20,30	4,4,4,4,4,4	-1,0,1,-1,0,1	0.92	0.86	0.86	0.86
D ₂₁	10,20,30,10,20,30	2,6,10,2,6,10	2,0,0,2,0,0	0.69	0.95	0.76	0.95
D ₂₂	10,20,30,10,20,30	2,6,10,2,6,10	-1,0,1,-1,0,1	0.65	0.82	0.63	0.83
D ₃₁	10,20,30,10,20,30	10,6,2,10,6,2	2,0,0,2,0,0	0.83	0.49	0.62	0.42
D ₃₂	10,20,30,10,20,30	10,6,2,10,6,2	-1,0,1,-1,0,1	0.87	0.69	0.75	0.69

Table 4.14 Predicted power values for different configurations of $k = 9$ groups

Pattern	n	σ^2	μ_i	F	W	KW	B^2
A ₁₁	5,5,5,5,5,5,5,5,5	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0.48	0.43	0.40	0.41
A ₁₂	5,5,5,5,5,5,5,5,5	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0.36	0.33	0.28	0.30
A ₂₁	5,5,5,5,5,5,5,5,5	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0.32	0.45	0.32	0.43
A ₂₂	5,5,5,5,5,5,5,5,5	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0.25	0.29	0.20	0.24
B ₁₁	10,10,10,10,10,10,10,10,10	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0.88	0.84	0.86	0.84
B ₁₂	10,10,10,10,10,10,10,10,10	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0.75	0.71	0.71	0.70
B ₂₁	10,10,10,10,10,10,10,10,10	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0.71	0.86	0.78	0.87
B ₂₂	10,10,10,10,10,10,10,10,10	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0.53	0.57	0.51	0.58
C ₁₁	5,10,15,5,10,15,5,10,15	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0.72	0.60	0.56	0.53
C ₁₂	5,10,15,5,10,15,5,10,15	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0.73	0.59	0.60	0.59
C ₂₁	5,10,15,5,10,15,5,10,15	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0.33	0.74	0.37	0.73
C ₂₂	5,10,15,5,10,15,5,10,15	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0.33	0.54	0.31	0.56
C ₃₁	5,10,15,5,10,15,5,10,15	10,6,2,10,6,2,10,6,2	2,0,0,2,0,0,2,0,0	0.70	0.32	0.41	0.24
C ₃₂	5,10,15,5,10,15,5,10,15	10,6,2,10,6,2,10,6,2	-1,0,1,-1,0,1,-1,0,1	0.72	0.43	0.51	0.42
D ₁₁	10,20,30,10,20,30,10,20,30	4,4,4,4,4,4,4,4,4	2,0,0,2,0,0,2,0,0	0.98	0.93	0.95	0.92
D ₁₂	10,20,30,10,20,30,10,20,30	4,4,4,4,4,4,4,4,4	-1,0,1,-1,0,1,-1,0,1	0.98	0.94	0.95	0.95
D ₂₁	10,20,30,10,20,30,10,20,30	2,6,10,2,6,10,2,6,10	2,0,0,2,0,0,2,0,0	0.84	0.98	0.89	0.99
D ₂₂	10,20,30,10,20,30,10,20,30	2,6,10,2,6,10,2,6,10	-1,0,1,-1,0,1,-1,0,1	0.80	0.92	0.76	0.93
D ₃₁	10,20,30,10,20,30,10,20,30	10,6,2,10,6,2,10,6,2	2,0,0,2,0,0,2,0,0	0.93	0.61	0.74	0.53
D ₃₂	10,20,30,10,20,30,10,20,30	10,6,2,10,6,2,10,6,2	-1,0,1,-1,0,1,-1,0,1	0.95	0.81	0.86	0.82

4.3 A New and Simple Approximation Procedure with Trimmed Mean and Winsorized Variance

It is well known that the usual group means and variances, which are the basis for all of the parametric procedures, are greatly influenced by the presence of extreme observations in distributions. In particular, the standard error of the usual mean can become seriously inflated when the underlying distribution has heavy tails. Accordingly, adopting a nonrobust measure “can give a distorted view of how the typical individual in one group compares to the typical individual in another, and about accurate probability coverage, controlling the probability of Type I error, and achieving relatively high power”(Wilcox 1995). By substituting robust measures of location and scale for the usual mean and variance, it should be possible to obtain test statistics that are insensitive to the combined effects of variance heterogeneity and nonnormality.

Although a wide range of robust estimators have been proposed in the literature (Gross 1976), the trimmed mean and Winsorized variance are intuitively appealing because of their computational simplicity and good theoretical properties (Wilcox 1995). The standard error of the trimmed mean is less affected by departures from normality than usual mean because extreme observations that is, observations in the tails of the distribution are censored or removed. Furthermore, as Gross (1976) noted, “the Winsorized variance is a consistent estimator of the variance of the corresponding trimmed mean”. In computing the Winsorized variance, the most extreme observations are replaced with less extreme values in the distribution of scores.

4.3.1 Trimmed Mean B_t^2 Test

At first some basic definitions about the test have been given. Consider $Y_1, Y_2, \dots, Y_n$ be a random sample and let $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$ be the order statistics; let $e = [\beta n]$ where $[Y]$ is the greatest integer less than or equal to Y and β is the

trimming size. Let $f = n - 2e$ be the effective sample size. The sample trimmed mean is

$$\bar{Y}_t = \sum_{i=e+1}^{n-e} \frac{Y_{(i)}}{f} \quad (4.10)$$

and the corresponding sample Winsorized mean is

$$\bar{Y}_w = \sum_{i=1}^n \frac{X_i}{n} \quad (4.11)$$

where

$$X_i = \begin{cases} Y_{(e+1)} & Y_i \leq Y_{(e+1)} \\ Y_i & Y_{(e+1)} < Y_i < Y_{(n-e)} \\ Y_{(n-e)} & Y_i \geq Y_{(n-e)} \end{cases} . \quad (4.12)$$

The sample Winsorized variance is

$$S_w^2 = \frac{\sum_{i=1}^n (X_i - \bar{Y}_w)^2}{n-1} \quad (4.13)$$

and the squared standard error of the sample trimmed mean be

$$S_{\bar{Y}_t}^2 = \frac{(n-1)S_w^2}{f(f-1)} \quad (4.14)$$

(Wei-Ming and Jiin-Huang 1999).

B_t^2 test uses trimmed mean and Winsorized variance instead of usual sample arithmetic mean and variance in B^2 test. Consider k independent randomly sampled groups each measured on a random variable (Y). Variances of the populations and sample size of the groups need not to be equal and distribution of the populations

need not to be normal. Each of the k samples of size n_i will have sample trimmed mean $\bar{Y}_{ti.}$ and each of the trimmed means will have a standard error of

$$S_{\bar{Y}_{ti.}}^2 = \frac{(n_i - 1)S_{w_i}^2}{f_i(f_i - 1)} \quad (4.15)$$

where

$$S_{w_i}^2 = \frac{\sum_{i=1}^n (X_i - \bar{Y}_{w_i})^2}{n_i - 1} \quad (4.16)$$

and $\bar{Y}_{w_i}$ is Winsorized mean. For each of the k samples, if we define a weight (ω_i)

such that $\sum_{i=1}^k \omega_i = 1$

$$\omega_i = \frac{1/S_{\bar{Y}_{ti.}}^2}{\sum_{i=1}^k \left(1/S_{\bar{Y}_{ti.}}^2 \right)} \quad (4.17)$$

The variance-weighted estimate of the common mean (Y^+) of Y becomes

$$Y^+ = \sum_{i=1}^k \omega_i \bar{Y}_{ti.} \quad (4.18)$$

Finally, for each of the k groups a one-sample t statistics is proposed and calculated as

$$T_{ti} = \frac{\bar{Y}_{ti.} - Y^+}{S_{\bar{Y}_{ti.}}} \quad (4.19)$$

Under the usual assumptions, each of the T_{ti} will be distributed as Student's t with $\nu_i = f_i - 1$ degrees of freedom. Several of the approximation methods that have appeared in the literature begin with the equivalent of the same derivation. After this step, this new approximation uses a normalizing transformation on each of the sample T_{ti} values directly in order to transform the each T_{ti} values in to standard normal deviate z_i . Several normalizing transformations for the t statistics have appeared in the literature. One of them is "Accurate Normalizing Transformations of a Student's t Variate" Bailey (1980). Bailey offers a locally accurate normalizing transformation which is given as follows

$$z_{ti} = \pm \frac{4\nu_i^2 + \frac{5(2z_c^2 + 3)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2 + 9)}{12}} \nu_i^{1/2} \left\{ \ln \left(1 + \frac{T_{ti}^2}{\nu_i} \right) \right\}^{1/2} \sim N(0,1) \quad (4.20)$$

where $\nu_i = f_i - 1$ degrees of freedom of T_{ti} variate and z_c is the critical value of related significance level under standard normal distribution. Since χ_{k-1}^2 is simply the mathematical description of the sum of k independent squared standard normal deviates,

$$B_t^2 = \sum_{i=1}^k z_{ti}^2 = \sum_{i=1}^k \left(\frac{4\nu_i^2 + \frac{5(2z_c^2 + 3)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2 + 9)}{12}} \nu_i^{1/2} \left\{ \ln \left(1 + \frac{\left(\frac{\bar{Y}_{ti} - Y^+}{S_{\bar{Y}_{ti}}} \right)^2}{\nu_i} \right) \right\}^{1/2} \right)^2 \quad (4.21)$$

will be distributed approximately χ^2 with k-1 degrees of freedom. Decision rule that the test uses is reject the null hypothesis of equal means if B_t^2 exceeds the $1-\alpha$ quantile of a chi-square distribution with k-1 degrees of freedom.

4.3.2 Simulation Study for Trimmed Mean B_t^2 Test

Performance of the conventional F, Kruskal-Wallis(KW), and B_t^2 procedures have been examined by means of a simulation study. Since the Welch test is developed for normally distributed populations, this test had not been included in this simulation study. Simulated actual significance level and power have been obtained using different sample sizes and error variances for $k=4$ and $k=6$ groups by using nominal significance level 0.05. All means were equal to 0 when predicting actual significance level. The first population mean was $\mu_1=2$, $\mu_2=1$ and the other groups were set to be zero when assessing the power of the tests. Each configuration was executed 10000 times by using the 13'th version of statistical software MINITAB.

To ensure a wide range of distribution shapes, the observations that included extreme values for both skewness and kurtosis were produced by g-and-h distributions (Hoaglin et. al. 1985). The parameter g controls the amount and direction of skewness, and the parameter h controls the kurtosis. For example, the case of $g=h=0$ corresponds to a standard normal distribution. It is noted that the tails of the distribution get heavier as h increases. The corresponding skewness and kurtosis of the g-and-h distribution are referred to in Wilcox (1994). To generate data from g-and-h distribution, standard normal variables z were generated and then were converted to

$$\varepsilon = \frac{(\exp(gz)-1)\exp\left(\frac{hz^2}{2}\right)}{g} \quad (4.22)$$

When $g=0$, this last expression was taken to be

$$\varepsilon = z \exp\left(\frac{hz^2}{2}\right). \quad (4.23)$$

It was noted that when comparing means, the g-and-h distribution had a mean of zero when $g = 0$, so multiplying each ε_{ij} by σ_i to obtain unequal variances did not affect the null hypothesis of $H_0 : \mu_{t1} = \mu_{t2} = \dots = \mu_{tk}$. For $g > 0$, the mean of the g-and-h distribution was

$$\mu_{gh} = \frac{\left(\exp \left\{ \frac{g^2}{2(1-h)} \right\} - 1 \right)}{\{g(1-h)^{1/2}\}} \quad (4.24)$$

In this case, μ_{gh} should be subtracted from ε_{ij} before multiplying by σ_i . When dealing with trimmed means, the population trimmed mean μ_{tgh} should be subtracted from ε_{ij} instead, so that again the null hypothesis for equal population trimmed means remains true. The population trimmed mean values for various g-and-h distributions are referred to in Wilcox (1994). Then, for $g > 0$, the ij 'th generated observation for the untrimmed case is

$$Y_{ij} = \mu_i + \sigma_i (\varepsilon - \mu_{gh}) \quad (4.25)$$

and for trimmed case

$$Y_{ij} = \mu_i + \sigma_i (\varepsilon - \mu_{tgh}) \quad (4.26)$$

Values of g and h have been given in Table 4.15

Table 4.15 g and h values which were used in simulation study of trimmed mean B_t^2 test

g	h	μ_{gh}	μ_{tgh}	γ_1 coefficient of skewness	γ_2 coefficient of kurtosis	description
0	0	0	0	0	0	Standard normal
0.5	0.2	0.3782	0.0564	13.16	42895.9	Skew with moderate tail
0.5	0.5	0.8033	0.06	-	-	Skew with heavy tail
1	0	0.6487	0.111	6.2	114	Extereme skew with light tail
1	0.2	0.9707	0.1157	713	5.6×10^{14}	Extereme skew with moderate tail
1	0.5	2.43	0.1231	-	-	Extereme skew with heavy tail

The effect of trimming size was discussed by Wilcox (1994) and a 20% symmetrical trimming was suggested. Experimental patterns used in this study have been given in Table 4.16. Actual significance level which exceeds Bradley's liberal conditions of robustness limits ($0.025 < \alpha < 0.075$ for nominal significance level of 5%) have been written in boldface.

Table 4.16 Experimental patterns used in this simulation study

Pattern	i	k = 4				k = 6					
		1	2	3	4	1	2	3	4	5	6
A₁	n_i	10	15	20	25	10	15	15	20	20	25
	σ_i^2	36	1	1	1	36	1	1	1	1	1
A₂	n_i	10	15	20	25	10	15	15	20	20	25
	σ_i^2	16	9	4	1	16	9	9	4	1	1
B₁	n_i	15	20	25	30	15	20	20	25	25	30
	σ_i^2	36	1	1	1	36	1	1	1	1	1
B₂	n_i	15	20	25	30	15	20	20	25	25	30
	σ_i^2	16	9	4	1	16	9	9	4	1	1

Table 4.17 Actual significance level for $k=4$ $g=0, h=0$ and $g=0.5, h=0.2$

Pattern	Distribution	n	σ_i^2	F	KW	B_i^2
A ₁	$g=0, h=0$	10,15,20,25	36,1,1,1	0.282	0.106	0.055
A ₂	$g=0, h=0$	10,15,20,25	16,9,4,1	0.165	0.097	0.056
B ₁	$g=0, h=0$	15,20,25,30	36,1,1,1	0.206	0.094	0.055
B ₂	$g=0, h=0$	15,20,25,30	16,9,4,1	0.143	0.092	0.058
A ₁	$g=0.5, h=0.2$	10,15,20,25	36,1,1,1	0.276	0.114	0.049
A ₂	$g=0.5, h=0.2$	10,15,20,25	16,9,4,1	0.156	0.102	0.057
B ₁	$g=0.5, h=0.2$	15,20,25,30	36,1,1,1	0.240	0.104	0.053
B ₂	$g=0.5, h=0.2$	15,20,25,30	16,9,4,1	0.128	0.090	0.056

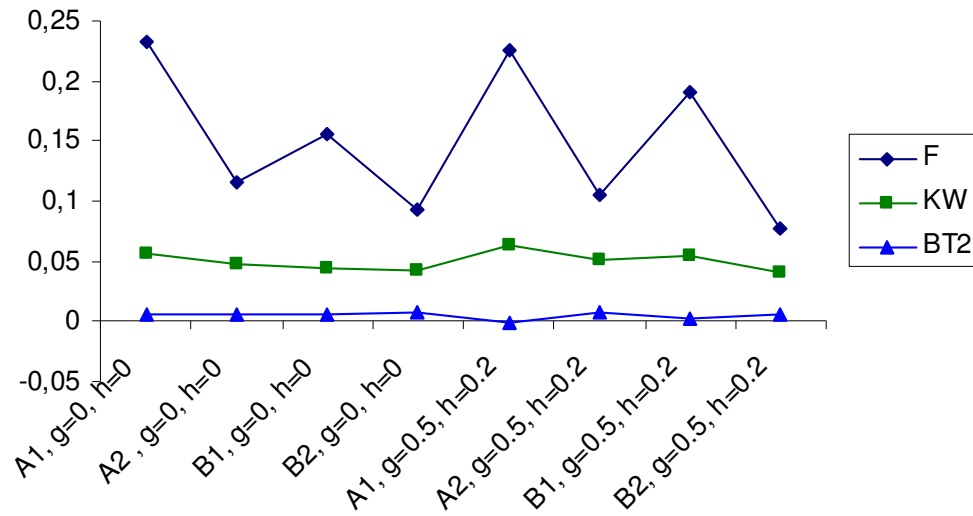
Figure 4.5 Discrepancy between nominal and actual significance level when $g = 0, h = 0$ and $g = 0.5, h = 0.2$ for $k = 4$ groups

Table 4.18 Actual significance level for $k=4$ $g=0.5, h=0.5$ and $g=1, h=0$

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A ₁	$g=0.5, h=0.5$	10,15,20,25	36,1,1,1	0.220	0.104	0.042
A ₂	$g=0.5, h=0.5$	10,15,20,25	16,9,4,1	0.109	0.095	0.044
B ₁	$g=0.5, h=0.5$	15,20,25,30	36,1,1,1	0.179	0.099	0.041
B ₂	$g=0.5, h=0.5$	15,20,25,30	16,9,4,1	0.093	0.088	0.048
A ₁	$g=1, h=0$	10,15,20,25	36,1,1,1	0.305	0.125	0.066
A ₂	$g=1, h=0$	10,15,20,25	16,9,4,1	0.173	0.154	0.090
B ₁	$g=1, h=0$	15,20,25,30	36,1,1,1	0.299	0.136	0.060
B ₂	$g=1, h=0$	15,20,25,30	16,9,4,1	0.151	0.168	0.077

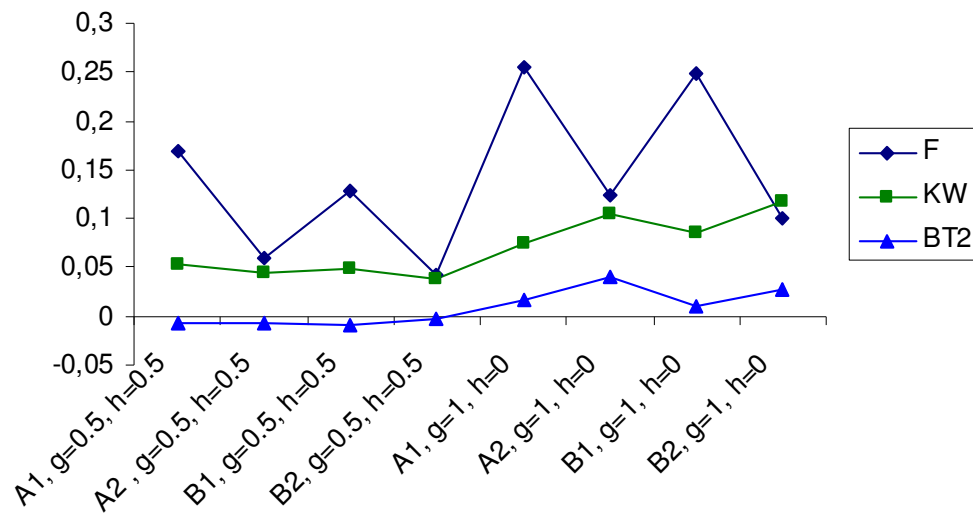
Figure 4.6 Discrepancy between nominal and actual significance level when $g = 0.5, h = 0.5$ and $g = 1, h = 0$ for $k = 4$ groups

Table 4.19 Actual significance level for $k=4$ $g=1, h=0.2$ and $g=1, h=0.5$

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A ₁	$g=1, h=0.2$	10,15,20,25	36,1,1,1	0.264	0.127	0.059
A ₂	$g=1, h=0.2$	10,15,20,25	16,9,4,1	0.135	0.138	0.079
B ₁	$g=1, h=0.2$	15,20,25,30	36,1,1,1	0.245	0.132	0.056
B ₂	$g=1, h=0.2$	15,20,25,30	16,9,4,1	0.121	0.141	0.071
A ₁	$g=1, h=0.5$	10,15,20,25	36,1,1,1	0.198	0.127	0.046
A ₂	$g=1, h=0.5$	10,15,20,25	16,9,4,1	0.100	0.118	0.064
B ₁	$g=1, h=0.5$	15,20,25,30	36,1,1,1	0.161	0.127	0.048
B ₂	$g=1, h=0.5$	15,20,25,30	16,9,4,1	0.084	0.120	0.060

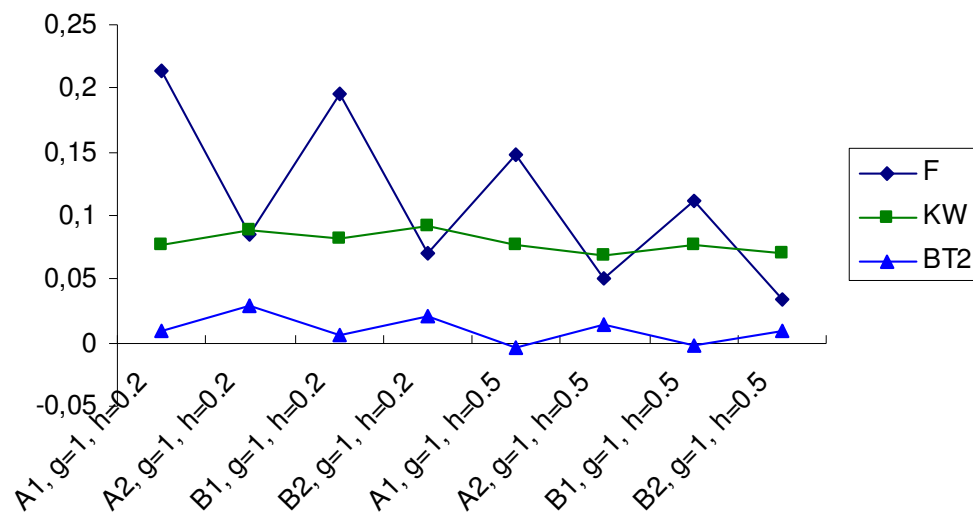
Figure 4.7 Discrepancy between nominal and actual significance level when $g = 1, h = 0.2$ and $g = 1, h = 0.5$ for $k = 4$ groups

Table 4.20 Actual significance level for $k=6$ $g=0, h=0$ and $g=0.5, h=0.2$

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A ₁	$g=0, h=0$	10,15,15,20,20,25	36,1,1,1,1,1	0.276	0.096	0.063
A ₂	$g=0, h=0$	10,15,15,20,20,25	16,9,9,4,1,1	0.167	0.099	0.059
B ₁	$g=0, h=0$	15,20,20,25,25,30	36,1,1,1,1,1	0.224	0.087	0.065
B ₂	$g=0, h=0$	15,20,20,25,25,30	16,9,9,4,1,1	0.145	0.091	0.058
A ₁	$g=0.5, h=0.2$	10,15,15,20,20,25	36,1,1,1,1,1	0.288	0.094	0.053
A ₂	$g=0.5, h=0.2$	10,15,15,20,20,25	16,9,9,4,1,1	0.163	0.099	0.060
B ₁	$g=0.5, h=0.2$	15,20,20,25,25,30	36,1,1,1,1,1	0.256	0.093	0.053
B ₂	$g=0.5, h=0.2$	15,20,20,25,25,30	16,9,9,4,1,1	0.140	0.102	0.053

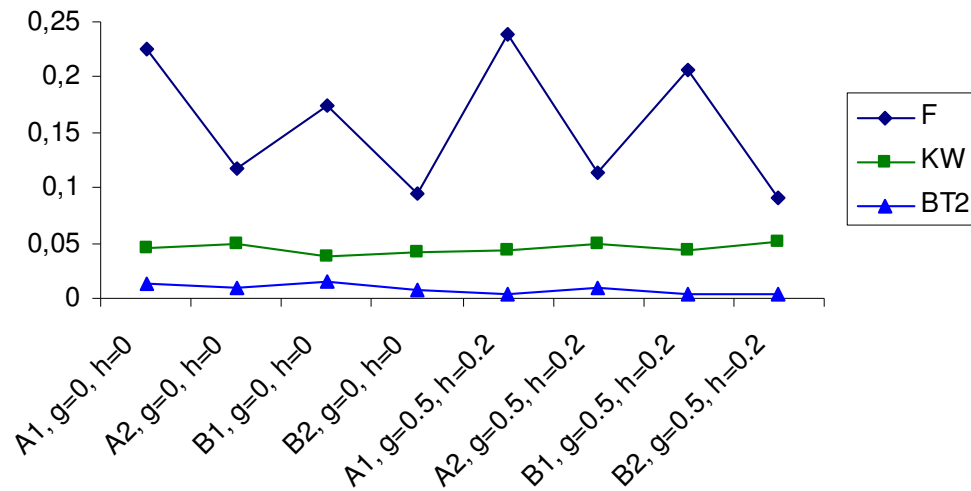
Figure 4.8 Discrepancy between nominal and actual significance level when $g = 0, h = 0$ and $g = 0.5, h = 0.2$ for $k = 6$ groups

Table 4.21 Actual significance level for $k=6$ $g=0.5$, $h=0.5$ and $g=1$, $h=0$

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A ₁	$g=0.5, h=0.5$	10,15,15,20,20,25	36,1,1,1,1,1	0.214	0.091	0.036
A ₂	$g=0.5, h=0.5$	10,15,15,20,20,25	16,9,9,4,1,1	0.115	0.089	0.041
B ₁	$g=0.5, h=0.5$	15,20,20,25,25,30	36,1,1,1,1,1	0.181	0.089	0.039
B ₂	$g=0.5, h=0.5$	15,20,20,25,25,30	16,9,9,4,1,1	0.088	0.088	0.043
A ₁	$g=1, h=0$	10,15,15,20,20,25	36,1,1,1,1,1	0.318	0.116	0.073
A ₂	$g=1, h=0$	10,15,15,20,20,25	16,9,9,4,1,1	0.183	0.161	0.089
B ₁	$g=1, h=0$	15,20,20,25,25,30	36,1,1,1,1,1	0.315	0.121	0.065
B ₂	$g=1, h=0$	15,20,20,25,25,30	16,9,9,4,1,1	0.181	0.173	0.074

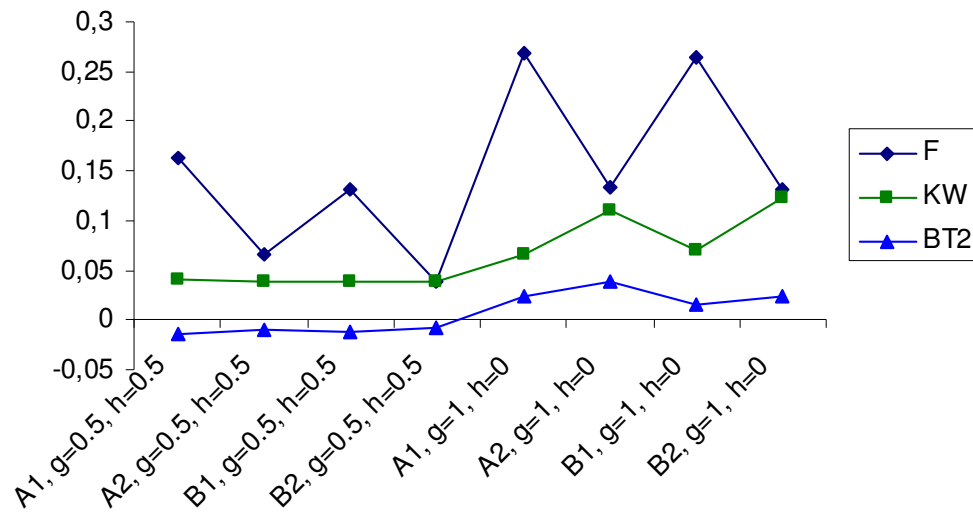
Figure 4.9 Discrepancy between nominal and actual significance level when $g = 0.5, h = 0.5$ and $g = 1, h = 0$ for $k = 6$ groups

Table 4.22 Actual significance level for $k=6$ $g=1, h=0.2$ and $g=1, h=0.5$

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A ₁	$g=1, h=0.2$	10,15,15,20,20,25	36,1,1,1,1,1	0.264	0.108	0.053
A ₂	$g=1, h=0.2$	10,15,15,20,20,25	16,9,9,4,1,1	0.146	0.143	0.061
B ₁	$g=1, h=0.2$	15,20,20,25,25,30	36,1,1,1,1,1	0.250	0.114	0.061
B ₂	$g=1, h=0.2$	15,20,20,25,25,30	16,9,9,4,1,1	0.134	0.153	0.073
A ₁	$g=1, h=0.5$	10,15,15,20,20,25	36,1,1,1,1,1	0.182	0.108	0.043
A ₂	$g=1, h=0.5$	10,15,15,20,20,25	16,9,9,4,1,1	0.101	0.124	0.057
B ₁	$g=1, h=0.5$	15,20,20,25,25,30	36,1,1,1,1,1	0.160	0.117	0.044
B ₂	$g=1, h=0.5$	15,20,20,25,25,30	16,9,9,4,1,1	0.081	0.131	0.060

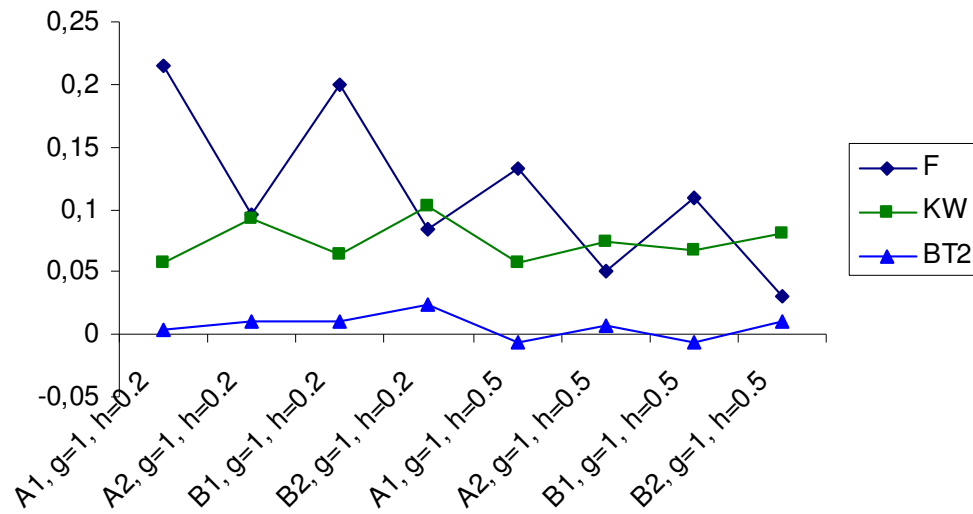
Figure 4.9 Discrepancy between nominal and actual significance level when $g = 1, h = 0.2$ and $g = 1, h = 0.5$ for $k = 6$ groups

Table 4.22 Predicted power values for k=4 g=0, h=0 and g=0.5, h=0.2

Pattern	Distribution	n	σ_i^2	F	KW	B_i^2
A₁	g=0, h=0	10,15,20,25	36,1,1,1	0.534	0.754	0.647
A₂	g=0, h=0	10,15,20,25	16,9,4,1	0.577	0.420	0.192
B₁	g=0, h=0	15,20,25,30	36,1,1,1	0.549	0.846	0.807
B₂	g=0, h=0	15,20,25,30	16,9,4,1	0.662	0.514	0.363
A₁	g=0.5, h=0.2	10,15,20,25	36,1,1,1	0.498	0.626	0.583
A₂	g=0.5, h=0.2	10,15,20,25	16,9,4,1	0.445	0.615	0.161
B₁	g=0.5, h=0.2	15,20,25,30	36,1,1,1	0.530	0.732	0.768
B₂	g=0.5, h=0.2	15,20,25,30	16,9,4,1	0.490	0.364	0.255

Table 4.23 Predicted power values for k=4 g=0.5, h=0.5 and g=1, h=0

Pattern	Distribution	n	σ_i^2	F	KW	B_i^2
A₁	g=0.5, h=0.5	10,15,20,25	36,1,1,1	0.323	0.471	0.487
A₂	g=0.5, h=0.5	10,15,20,25	16,9,4,1	0.221	0.239	0.127
B₁	g=0.5, h=0.5	15,20,25,30	36,1,1,1	0.314	0.568	0.642
B₂	g=0.5, h=0.5	15,20,25,30	16,9,4,1	0.212	0.305	0.209
A₁	g=1, h=0	10,15,20,25	36,1,1,1	0.603	0.728	0.604
A₂	g=1, h=0	10,15,20,25	16,9,4,1	0.496	0.308	0.161
B₁	g=1, h=0	15,20,25,30	36,1,1,1	0.673	0.850	0.759
B₂	g=1, h=0	15,20,25,30	16,9,4,1	0.586	0.398	0.211

Table 4.24 Predicted power values for k=4 g=1, h=0.2 and g=1, h=0.5

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A₁	g=1, h=0.2	10,15,20,25	36,1,1,1	0.454	0.613	0.526
A₂	g=1, h=0.2	10,15,20,25	16,9,4,1	0.324	0.261	0.116
B₁	g=1, h=0.2	15,20,25,30	36,1,1,1	0.492	0.733	0.691
B₂	g=1, h=0.2	15,20,25,30	16,9,4,1	0.356	0.321	0.189
A₁	g=1, h=0.5	10,15,20,25	36,1,1,1	0.319	0.470	0.422
A₂	g=1, h=0.5	10,15,20,25	16,9,4,1	0.230	0.204	0.091
B₁	g=1, h=0.5	15,20,25,30	36,1,1,1	0.279	0.576	0.571
B₂	g=1, h=0.5	15,20,25,30	16,9,4,1	0.206	0.255	0.136

Table 4.25 Predicted power values for k=6 g=0, h=0 and g=0.5, h=0.2

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A₁	g=0, h=0	10,15,15,20,20,25	36,1,1,1,1,1	0.570	0.735	0.596
A₂	g=0, h=0	10,15,15,20,20,25	16,9,9,4,1,1	0.574	0.391	0.211
B₁	g=0, h=0	15,20,20,25,25,30	36,1,1,1,1,1	0.608	0.859	0.778
B₂	g=0, h=0	15,20,20,25,25,30	16,9,9,4,1,1	0.666	0.496	0.288
A₁	g=0.5, h=0.2	10,15,15,20,20,25	36,1,1,1,1,1	0.521	0.586	0.526
A₂	g=0.5, h=0.2	10,15,15,20,20,25	16,9,9,4,1,1	0.423	0.282	0.143
B₁	g=0.5, h=0.2	15,20,20,25,25,30	36,1,1,1,1,1	0.584	0.737	0.708
B₂	g=0.5, h=0.2	15,20,20,25,25,30	16,9,9,4,1,1	0.480	0.360	0.216

Table 4.26 Predicted power values for k=6 g=0.5, h=0.5 and g=1, h=0

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A₁	g=0.5, h=0.5	10,15,15,20,20,25	36,1,1,1,1,1	0.324	0.488	0.429
A₂	g=0.5, h=0.5	10,15,15,20,20,25	16,9,9,4,1,1	0.202	0.229	0.106
B₁	g=0.5, h=0.5	15,20,20,25,25,30	36,1,1,1,1,1	0.320	0.582	0.622
B₂	g=0.5, h=0.5	15,20,20,25,25,30	16,9,9,4,1,1	0.196	0.291	0.163
A₁	g=1, h=0	10,15,15,20,20,25	36,1,1,1,1,1	0.604	0.729	0.547
A₂	g=1, h=0	10,15,15,20,20,25	16,9,9,4,1,1	0.490	0.324	0.139
B₁	g=1, h=0	15,20,20,25,25,30	36,1,1,1,1,1	0.715	0.872	0.742
B₂	g=1, h=0	15,20,20,25,25,30	16,9,9,4,1,1	0.581	0.417	0.192

Table 4.27 Predicted power values for k=6 g=1, h=0.2 and g=1, h=0.5

Pattern	Distribution	n	σ_i^2	F	KW	B_t^2
A₁	g=1, h=0.2	10,15,15,20,20,25	36,1,1,1,1,1	0.457	0.605	0.482
A₂	g=1, h=0.2	10,15,15,20,20,25	16,9,9,4,1,1	0.312	0.262	0.114
B₁	g=1, h=0.2	15,20,20,25,25,30	36,1,1,1,1,1	0.517	0.765	0.658
B₂	g=1, h=0.2	15,20,20,25,25,30	16,9,9,4,1,1	0.337	0.337	0.162
A₁	g=1, h=0.5	10,15,15,20,20,25	36,1,1,1,1,1	0.309	0.447	0.384
A₂	g=1, h=0.5	10,15,15,20,20,25	16,9,9,4,1,1	0.204	0.218	0.105
B₁	g=1, h=0.5	15,20,20,25,25,30	36,1,1,1,1,1	0.291	0.577	0.544
B₂	g=1, h=0.5	15,20,20,25,25,30	16,9,9,4,1,1	0.187	0.267	0.122

CHAPTER FIVE

CONCLUSION AND FURTHER RESEARCH

In this study, heterogeneity of variance problem in one-way fixed effects analysis of variance model have been examined in detail. At first, some general information about one-way analysis of variance and results of heterogeneity of variance problem in one-way analysis of variance have been given. Then, different solution approaches to heterogeneity of variance problem in one-way ANOVA have been described. And finally, presentation of two new and simple approximation procedures and results of two simulation studies which were implemented to compare the actual significance level and power of the new approximations, conventional F test and two other alternatives (Welch test, Kruskal-Wallis test) have been explained.

Two different simulation study have been performed for two newly offered approximation procedures. At the first simulation study, obtained results have been interpreted according to number of treatments, number of replications and population variances. Interpretation criterion was the closeness of the actual significance level and nominal significance level ($\alpha_n = 0.05$). Bradley (1978) has stated that a test can said to be robust to a specific assumption violation if it protects actual significance level between $0.9\alpha_n < \alpha < 1.1\alpha_n$. This criterion is called as Bradley's stringent criterion of robustness and refers to $0.045 < \alpha < 0.055$ for nominal significance level of 0.05.

Power values which directly change with Type I error rates have not been considered while comparing any two tests unless the actual Type I error rates of the tests are approximately equal. Experimental patterns which were used in simulation study have been coded in terms of number of replication, population variances and pairing types. Codes of patterns have been given in Table 5.1

Table 5.1 Codes of patterns used in first simulation study

A_1 A_1^*	Balanced-Small sample- Homoscedasticity	C_2 C_2^*	Unbalanced-Small sample- Heteroscedasticity (Positive pairings)
A_2 A_2^*	Balanced-Small sample- Heteroscedasticity	C_3 C_3^*	Unbalanced-Small sample- Heteroscedasticity (Negative pairings)
B_1 B_1^*	Balanced-Large sample- Homoscedasticity	D_1 D_1^*	Unbalanced-Large sample- Homoscedasticity
B_2 B_2^*	Balanced-Large sample- Heteroscedasticity	D_2 D_2^*	Unbalanced-Large sample- Heteroscedasticity (Positive pairings)
C_1 C_1^*	Unbalanced-Small sample- Homoscedasticity	D_3 D_3^*	Unbalanced-Large sample- Heteroscedasticity (Negative pairings)

According to the Bradley's stringent criterion of robustness, recommended tests can be classified as in Table 5.2.

Table 5.2 Recommended tests with different experimental patterns

		Homoscedasticity	Heteroscedasticity	
Balanced	Small sample	F, B^2	B^2	
	Large sample	F, B^2	KW	
Unbalanced	Small sample	B^2	PP: B^2	NP: B^2
	Large sample	KW, B^2	PP: W, B^2	NP: W, B^2

PP: Positive Pairings
NP: Negative Pairings

New approximation procedure B^2 seems to be suitable for all experimental patterns which have been used in this simulation study.

In patterns like A_1 , A_1^* , B_1 , B_1^* and similar ones, F and B^2 tests are recommended together, but since F test has a greater power value, it should be preferred at first .

In patterns like D_1 , D_1^* and similar ones, KW and B^2 tests are recommended together, but since KW test has a greater power value, it should be preferred at first .

In patterns like B_2 , B_2^* and similar ones, KW and B^2 tests are recommended together, but since B^2 test has a greater power value, it should be preferred at first .

In patterns like D_2 , D_2^* and similar ones, W and B^2 tests are recommended together and since both tests power values are nearly equal, there is no need to assign a priority order.

In patterns like D_3 , D_3^* and similar ones, W and B^2 tests are recommended together, but since W test has a greater power value, it should be preferred at first .

At the second simulation study which was implemented for B_t^2 test, obtained results have also been interpreted according to number of treatments, number of replications, population variances and distribution shapes. Interpretation criterion was the closeness of the actual significance level and nominal significance level ($\alpha_n = 0.05$). Bradley (1978), in his liberal criterion of robustness, has stated that a test can said to be robust to a specific assumption violation if it protects the actual significance level between $0.5\alpha_n < \alpha < 1.5\alpha_n$. This criterion refers to $0.025 < \alpha < 0.075$ for nominal significance level of 0.05.

Actual significance levels of B_t^2 test generally satisfy the Bradley's condition of robustness as can be viewed in a series of tables through Table 4.17 to Table 4.22. F test is the most sensitive of these three tests against violation of variance heterogeneity and normality assumption. Kruskal-Wallis test has not been affected

by assumption violation as much as F test but none of its actual significance levels satisfy the Bradley's liberal condition of robustness.

Predicted power values of both F test and Kruskal-Wallis test are generally higher than B_t^2 test's power values. Since there is a direct positive relation between probability of Type I error and power of the test, this can not be considered as a surprising result.

Among these three tests, use of B_t^2 test is recommended under variance heterogeneity and nonnormality in experimental designs which have been used in this study and similar ones. However, it should be noted that trimmed mean and Winsorized variance should only be adopted if the researcher is interested in testing for treatment effects across groups using a measure of location that more accurately reflects the typical score within a group when working with heavy tailed distributions.

The hypothesis tested when the usual mean is used as an estimate of location is not the same as that tested when the trimmed mean is employed. Consequently, the researcher needs to be clear on the goals of data analysis prior to choosing a particular method of statistical inference and must be clearly communicate these goals to all who will evaluate the results.

In an experimental design study an experimenter is frequently interested in a factor that has a large number of possible levels. If the experimenter randomly selects k of these levels from the population of factor levels, then it is said that the factor is random. A further study on B^2 and B_t^2 tests can be performed in order to make a modification for one-way random effects analysis of variance models.

REFERENCES

- Alexander, R.A., & Govern, D.M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, 19, 91-101.
- Asiribo, O., & Gurland, J. (1990). Coping with variance heterogeneity. *Communications In Statistics: Theory And Methods*, 19, 4029-4048.
- Bailey, B.J.R. (1980). Accurate normalizing transformations of Student's t Variate. *Applied Statistics*, 29, 304-306.
- Bishop, T.A., (1976). Heteroscedastic ANOVA, MANOVA and Multiple Comparisons *The Ohio State University Department of Statistics, Ph.D.Dissertation*
- Bishop, T.A., & Dudewicz E.J. (1978). Exact Analysis of Variance with Unequal Variances: Test Procedures and Tables. *Technometrics*, 20, 419-430
- Blair, R.C., & Higgins, J.J. (1980). A comparison of the power of the Wilcoxon's rank sum statistics to that of student's t statistics under various non-normal distributions. *Journal of Educational Statistics*, 5, 309-335.
- Boneau, C.A.(1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57, 49-64.
- Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. *Annals of Mathematical Statistics*, 25, 290-302.
- Box, G.E.P., & Anderson, S.L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumptions. *Journal of the Royal Statistical Society, Series B*, 17, 1-34.

- Böckenhoff, A., & Hartung, J. (1998). Some corrections of the significance level in meta – analysis. *Biometrical Journal*, 40, 937-947
- Bradley, J.V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.
- Brown, M.B., & Forsythe, A.B. (1974). The small sample behavior of some statistics which test equality of several means . *Technometrics*, 16, 129-132.
- Carrol, R.J.,& Ruppert, D. (1988) *Transformation ans Weighting in Regression*, Chapman&Hall
- Chen, Shun-Yi, Chen, H.J., & Ding, C.G. (2000). An ANOVA Test for the Equivalency of Means Under Unequal Variances. *Computational Statistics & Data Analysis* , 33, 151-175
- Cochran, W.G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of Royal Statistical Society Supp*, 4, 102-118.
- Cochran, W.G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 3, 22-38
- Dantzig, G., (1940). On the non-existence of tests of “Student’s” hypothesis having power functions independent of σ . *Annals of Mathematical Statistics*, Vol.11, 186-192
- Dijkstra, J.B., & Werter, P.S.(1981). Testing the equality of several means when the population variances are unequal. *Communications in Statistics B,Simulation and Computation*, 10, 557-569.
- Eisenhart, C. (1947). Assumptions underlying analysis of variance. *Biometrics*, 3, 1-22.

- Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian law on inheritance. *Trans.R.Soc.,Edin* 52, 399-433.
- Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237-288.
- Gross, A.M. (1976). Confidence interval robustness with long tailed symmetric distributions. *Journal of American Statistical Association*, 71, 409-416.
- Guo, J.H.,& Luh, W.M. (2000). An invertible transformation two sample trimmed t statistics under heterogeneity and nonnormality. *Statistics & Probability Letters*, 49, 1-7
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., & Stahel, W.A. (1986). *Robust statistics: The approached based on influence functions*. Newyork: Wiley
- Hartung, J., Argaç D., & Makambi K.H.(2002). Small sample properties of tests on homogeneity in one-way Anova and meta-analysis. *Statistical Papers*, 43, 197-235.
- Harwell, M.R., Rubinstein, E.N., Hayes, W.S., & Olds, C.C. (1992). Summarizing Monte Carlo results in methodological research: The one and two factor fixed effects ANOVA cases. *Journal of Educational Statistics*, 17, 315-339.
- Hoaglin, D.C., Mosteller, F., & Tukey, J.W. (1983). *Understanding robust and exploratory data analysis*. Newyork: Wiley
- Hoaglin, D.C., Mosteller, F., & Tukey, J.W. (1985). *Exploring data tables, trends and shapes*. Newyork: Wiley

- Hsiung, T.H., Olejnik, S., & Oshima, T.C. (1994). A SAS /IML program for applying the James second order test in two factor fixed effects ANOVA models. *Educational and Psychological Measurement*, 54, 696-698.
- Huber, P. (1981). *Robust statistics*. Newyork: Wiley
- James, G.S.(1951). The Comparison of Several Groups of Observations When the Ratios of the Population Variances are Unknown. *Biometrika*, 67, 85-92.
- Johansen, S. (1980). The Welch-James approximation of the distribution of the residual sum of squares in weighted linear regression. *Biometrika*, 67, 85-92.
- Keselman, H.J., Rogan, J.C., & Feir-Walsh, B.J. (1977). An evaluation of some non-parametric tests for location equality. *British Journal of Mathemetical Statistics and Psychology*, 30, 213-221.
- Keselman, H.J.,Kowalchuk, R.K.,& Lix, L.M. (1998). Robust nonorthogonal analysis revisited: An uptade based on trimmed means. *Psychometrika*, 63, 145-163.
- Kruskal, W.H., & Wallis. W.A. (1952). Use of ranks in one criterion variance analysis. *Journal of American Statistical Association*, 47, 583-621.
- Krutchkoff, R.G. (1988). One-way fixed effects analysis of variance when the error variances may be unequal. *Journal of Statistical Computation and Simulation* , 30, 259-271
- Lindman, H.R. (1992). *Analysis of Variance in Experimental Design*. Springer-Verlag.

- Lix, L.M., & Keselman, H.J. (1995). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. *Psychological Bulletin*, 117, 547-560
- Lix, L.M., & Keselman, H.J. (1998). To trim or not to trim: tests of location equality under heteroscedasticity and non-normality. *Educational and Psychological Measurement*, 58, 409-429.
- Luh, W., & Guo J. (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under non-normal. *British Journal Of Mathematical and Statistical Psychology* , 52, 303-320.
- McSweeney, M., & Penfield, D. (1969). The normal score test for the c-sample problem. *British Journal of Mathematical Statistics and Psychology* , 20, 187-204
- Montgomery, D.C. (1991). *Design and Analysis of Experiments*. Newyork: John Wiley & Sons.Inc.
- Mehrotra, D.V. (1997). Improving the Brown- Forsythe solution to the generalized Behrens-Fisher problem. *Communications in Statistics and Simulation*, 26, 1139-1145
- Norton, D.W. (1952). An empirical investigation of the effects of nonnormality and heterogeneity upon the F test of analysis of variance. Unpublished doctoral dissertation, State University of Iowa.
- Oshima, T.C., & Algina, J. (1992a) A SAS program for testing the hypothesis of the equal means under heteroscedasticity: James's second order test. *Educational and Psychological Measurement*, 52, 117-118

- Pearson, E.S. (1931). The analysis of variance in cases of non-normal variation . *Biometrika*, 23, 114-133.
- Rousseeuw, P.J., & Leroy, A.M. (1987). *Robust regression and outlier detection*. Newyork: Wiley
- Rubin, A.S. (1983). The use of weighted contrasts in analysis of models with heterogeneity of variance. Proceedings of the Business and Economic Statistics Section, *American Statistical Association*, 347-352.
- Rupert, G., & Miller, J.R.(1986). *Beyond ANOVA, basics of applied statistics*. Newyork: John Wiley & Sons. Inc.
- Rust, S.W., & Fligner. M.A. (1984). A modification of the Kruskal Wallis statistics for the generalized Behrens-Fisher problem. *Communications in Statistics Theory and Methods*, 13, 2013-2027.
- Scheffe, H. (1959). *The Analysis of Variance*. Newyork: John Wiley & Sons. Inc.
- Smith, T.A., & Hickling, F. (1998). One-Way analysis of variance under heteroscedastic errors. *Eurostat Journal, Special Edition for Conference Proceedings*.
- Tiku, M.L., Tan, W.Y.,& Balakrishnan, N. (1986) *Robust Inference*. Newyork: Marcel Dekker Inc.
- Tomarken, A.J., & Serlin, R.C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, 99, 90-99.
- Van der Waerden, B.L. (1952). *Order tests for the two sample problem and their power*. *Indagationes Mathematicae*, 14, 453-458.

- Wei-Ming, L., & Jiin-Huarng, G. (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under nonnormality and inequality of variances. *British Journal of Mathematical and Statistical Psychology*, 52, 303-320.
- Welch, B.L. (1951). On the comparison of several mean values. *Biometrika*, 38, 330-336.
- Wilcox, R.R., Charlin, V.L., & Thompson, K.L. (1986). New Monte Carlo results on the robustness of the ANOVA F, W and F* statistics. *Communications in Statistics: Simulation and Computation*, 15, 933-943.
- Wilcox, R.R. (1987). A heteroscedastic ANOVA procedure with specified power. *Journal of Educational Statistics*, 12, 271-281
- Wilcox, R.R.(1988). A new Alternative to the ANOVA F and New Results on James Second Order Method. *British Journal of Mathematical and Statistical Psychology*, 41, 109-117.
- Wilcox, R.R.(1989). Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA models. *Journal of Educational Statistics*, 14, 269-278.
- Wilcox, R.R. (1994). A one-way random effects model for trimmed means. *Psychometrika*, 59, 289-306.
- Wilcox, R.R. (1995). ANOVA: A paradigm for low power and misleading measures of effect sizes? *Review of Educational Research*, 65, 51-77.
- Wilcox, R.R. (1996). *Statistics for the social sciences*. Academic Pres Inc.

- Wilcox, R.R. (1997a). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.
- Wilcox, R.R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*. Newyork: Springer.
- Wilcox, R.R. (2003). *Applying contemporary statistical techniques*. San Diego, CA. Academic Press.
- Wilcox, R.R., & Keselman, H.J. (2003). Repeated measures one-way ANOVA based on a modified one step M estimator. *British Journal of Mathematical and Statistical Psychology*, 56, 15-26