Selçuk J. Appl. Math.
Vol. 7. No.2. pp. 81-90, 2006

Selçuk Journal of Applied Mathematics

One Way Fixed Effect Analysis Of Variance Under Variance Heterogeneity And A Solution Proposal

A.Fırat Özdemir and Serdar Kurt

Dokuz Eylul University, Faculty of Arts and Sciences, Department of Statistics, Kaynaklar Campus, 35160 Buca / İzmir Turkey
e-mail: firat_ozdemir@deu.edu.tr, serdar_kurt@deu.edu.tr

Summary. In the first part of this study, the results of heteroscedasticity in one way fixed effect 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. Second part includes the presentation of new and simple approximation procedure which intends to create an easy and applicable alternative. 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. Third part of the study consist of a simulation study which was implemented to compare the actual significance level and power of the new approximation, conventional F test and two other alternatives (Welch Test, Kruskal-Wallis Test). Finally, some recommendations about the preference of these tests for different types of experimental conditions were given.

Key words: Heteroscedasticity, B^2 test, F test, Welch Test, Kruskal-Wallis Test

1. Introduction

The purpose of analysis of variance is to test the population mean differences for statistical significance. This is accomplished by analyzing the variance of the response variable of the experiment into two parts. First one is due to random variation and second one is due to differences between population means. This relationship is shown at Eq.(1)

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

These two components are then used in a test procedure called F test which has a test statistics as in Eq.(2). Under the assumption $\forall_i = 1, 2, ..., k$

$$Y_{ij} \sim NID\left(\mu_i, \sigma^2\right) j = 1, 2, ..., n_i$$

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

In ANOVA, variance of the distributions in which the samples are drawn should be the same 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 the case of heteroscedasticity, the distribution of response variable Y will be $\forall_i = 1, 2, ..., k \quad Y_{ij} \sim NID\left(\mu_i, \sigma_i^2\right) j = 1, 2, ..., n_i$ when the other two basic assumptions of analysis of variance hold. There are two main results of heteroscedasticity; distortion of the distribution of the F ratio and discrepancy between nominal and actual significance level.

1.1. Distribution of F Ratio

The numerator and denominator sums of squares of F ratio $\frac{MSTr}{MSE}$ are distributed as weighted sum of squares of independent normal random variables with weights σ_i^2. When the variances differ between populations these weights are unequal and the distributions are not chi-square. By far the best article about the effect of unequal variances on the F test is Box (1954a) (G Rupert and Jr. 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

$$(3) \qquad b = \frac{N-k}{N(k-1)} \frac{\sum_{i=1}^{k} (N - n_i)\sigma_i^2}{\sum_{i=1}^{k} (n_i - 1)\sigma_i^2}$$

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

$$h = \frac{\left\{ \sum_{i=1}^{k} (n_i - 1)\, \sigma_i^2 \right\}^2}{\left\{ \sum_{i=1}^{k} (n_i - 1)\, \sigma_i^4 \right\}} \tag{5}$$

1.2. Discrepancy between nominal and actual significance level

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 et al 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. (G Rupert and Jr. Miller, 1986). The denominator mean sum of squares is converging to its expected value, which is

$$E\left[\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2\right] = \frac{1}{N-k}\sum_{i=1}^{k}(n_i-1)\sigma_i^2 \tag{6}$$

where$\sigma_i^2$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 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

$$\begin{aligned}
&E\left[\tfrac{1}{k-1}\sum_{i=1}^{k} n_i\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2\right] \\
&= \tfrac{1}{k-1}\left[\sum_{i=1}^{k} n_i E\left(\bar{Y}_{i.}-\mu\right)^2 - NE\left(\bar{Y}_{..}-\mu\right)^2\right] \\
&= \tfrac{1}{k-1}\left[\sum_{i=1}^{k} n_i \tfrac{\sigma_i^2}{n_i} - N\tfrac{\sum_{i=1}^{k} n_i\sigma_i^2}{N^2}\right] \\
&= \tfrac{1}{N(k-1)}\sum_{i=1}^{k}(N-n_i)\sigma_i^2
\end{aligned} \tag{7}$$

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. In this case the variance of the numerator should be controlled in order to see the effect of variance heterogeneity.

$$(8) \qquad Var(MSTr) = \frac{2\bar{\sigma}^4}{k-1}\left[1 + \frac{(k-2)}{k(k-1)}\frac{\sum_{i=1}^{k}\left(\sigma_i^2 - \bar{\sigma}^2\right)^2}{\bar{\sigma}^4}\right]$$

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,1954a, 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 then 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.

2. 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 and Robust Statistical Procedures.

2.1. A New And Simple Approximation

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 $S_{\bar{Y}_i}$

(9) $$S_{\bar{Y}_i} = \left[\frac{\sum\limits_{j=1}^{n_i} \left(Y_j - \bar{Y}_i\right)^2}{n_i\left(n_i - 1\right)} \right]^{1/2}$$

For each of the k samples, if we define a weight $\omega_i$such that $\sum\limits_{i=1}^{k} \omega_i = 1$

(10) $$\omega_i = \frac{\frac{1}{S^2_{\bar{Y}_i}}}{\sum\limits_{i=1}^{k}\left(\frac{1}{S^2_{\bar{Y}_i}}\right)}$$

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

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

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

(12) $$t_i = \frac{\bar{Y}_i - Y^+}{S_{\bar{Y}_i}}$$

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. 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_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 as follows

(13) $$z_i = \pm \frac{4\nu_i^2 + \frac{5\left(2z_c^2+3\right)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2+9)}{12}} \nu_i^{1/2} \left\{ log\left(1 + \frac{t_i^2}{\nu_i}\right)\right\}^{1/2}$$

(14)

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

will be approximately distributed χ^2_{k-1}. Decision rule that the tests uses is reject the null hypothesis of of equal means if B^2 exceeds the 1-α quantile of a chi-square distribution with k-1 degrees of freedom. Let us call the first part of the z_i transformation as coefficient c, then ;

$$z_i = \pm c \left\{ log\left(1 + \frac{t_i^2}{\nu_i}\right)\right\}^{1/2} \tag{15}$$

where

$$c = \frac{4\nu_i^2 + \frac{5\left(2z_c^2+3\right)}{24}}{4\nu_i^2 + \nu_i + \frac{(4z_c^2+9)}{12}} \nu_i^{1/2} \tag{16}$$

then

$$B^2 = \sum_{i=1}^{k} z_i^2 = \sum_{i=1}^{k} \left(c\left\{ log\left(1 + \frac{\left(\frac{\bar{Y}_i - Y^+}{S_{\bar{Y}_i}}\right)^2}{\nu_i}\right)\right\}^{1/2}\right)^2 \tag{17}$$

can be written with the help of a table of c's (Table1).

3. Simulation Study

Performance of the F, Welch (W), Kruskal-Wallis(KW), Alexender-Govern(AG) and new approximation (B^2) procedures have been examined by means of a simulation study. Simulated actual significance level and and power of the test 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 assesing the power of the tests. The first pattern was μ_1 = 2, μ_2 = 0, μ_3 = 0 while the second pattern was μ_1 = -1, μ_2 = 0, μ_3 = 1 in which the means were equally spaced. Each configuration was executed 10000 times. All data were generated from normal distribution by using the 13'th version of statistical software MINITAB. Different experimental desings used in simulation study and all results were given in table section at the end of the paper.

4. Conclusions

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 (α_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. Power values for k=9 groups could not be given because of size problem. Recommended tests for different experimental designs were given at Table 9.

References

1. R.A. Alexender and D.M Govern, A new and Simpler Approximation for ANOVA Under Variance Heterogeneity, Journal of Educational Statistics 19, 91-101 (1994).

2. B.J.R. Bailey, Accurate Normalizing Transformations of Student's t Variate, Applied Statistics 29, 304-306 (1980).

3. G.E.P. Box, Some theorems on quadratic forms applied in the study of analysis of variance problems. Annals of Mathematical Statistics 25, 290-302 (1954a).

4. J.V. Bradley, Robustness? Brithish Journal of Mathematical and Statistical Psychology 31, 144-152. (1978).

5. J. Hartung, D. Argaç and K.H. Makambi, Small Sample Properties of Tests on Homogeneity in One-Way Anova and Meta-Analysis, Statistical Papers 43, 197-235. (2002).

6. W.H. Kruskal and W.A. Wallis Use of ranks in one criterion variance analysis. JASA 47, 583-621. (1952).

7. G.Rupert and J.R. Miller, Beyond ANOVA, basics of applied statistics, John Wiley & Sons. Inc Newyork (1986).

8. H. Scheffe, The Analysis of Variance, John Wiley & Sons.Inc. Newyork (1959).

9. B.L. Welch, On the comparison of several mean values, Biometrika 38, 330-336. (1951).

10. R.R. Wilcox, V.L. Charlin and K.L. Thompson, New Monte Carlo results on the robutness of the ANOVA F, W and F* statistics, Communications in Statistics: Simulation and Computation 15, 933-943. (1986).

Tables

df	1	2	3	4	5	6	7	8	9	10
α= 0.05	0.88552	1.28679	1.61365	1.89134	2.13569	2.35597	2.55794	2.74545	2.92116	3.08703
α= 0.01	0.92814	1.30814	1.62589	1.89933	2.14138	2.36026	2.56133	2.74821	2.92346	3.08898

Table 1: Coefficient of c's up to 10 degrees of freedom for α= 0.05 and α= 0.01

		K = 3			K = 6						K = 9								
Pattern	i	1	2	3	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9
A_1	n_i	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	σ_i^2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
A_2	n_i	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	σ_i^2	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10
B_1	n_i	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	σ_i^2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
B_2	n_i	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	σ_i^2	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10
C_1	n_i	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15
	σ_i^2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
C_2	n_i	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15
	σ_i^2	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10
C_3	n_i	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15	5	10	15
	σ_i^2	10	6	2	10	6	2	10	6	2	10	6	2	10	6	2	10	6	2
D_1	n_i	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	σ_i^2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
D_2	n_i	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	σ_i^2	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10	2	6	10
D_3	n_i	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30	10	20	30
	σ_i^2	10	6	2	10	6	2	10	6	2	10	6	2	10	6	2	10	6	2

Table 2: Sample sizes and variances of the distributions for k=3, k=6 and k=9

Pattern	n	σ^2	F	W	KW	AG	B^2
A_1	5,5,5	4,4,4	0.051	0.052	0.044	0.039	0.049
A_2	5,5,5	2,6,10	0.059	0.059	0.054	0.044	0.049
B_1	10,10,10	4,4,4	0.047	0.050	0.048	0.049	0.049
B_2	10,10,10	2,6,10	0.054	0.053	0.053	0.047	0.048
C_1	5,10,15	4,4,4	0.068	0.063	0.046	0.055	0.050
C_2	5,10,15	2,6,10	0.029	0.055	0.029	0.048	0.049
C_3	5,10,15	10,6,2	0.150	0.067	0.082	0.057	0.053
D_1	10,20,30	4,4,4	0.069	0.051	0.045	0.051	0.051
D_2	10,20,30	2,4,6	0.026	0.049	0.029	0.050	0.048
D_3	10,20,30	10,6,2	0.160	0.052	0.088	0.045	0.049

Table 3 : Actual significance sevels for different experimental designs of k = 3 groups

Pattern	n	σ^2	F	W	KW	AG	B^2
A_1	5,5,5,5,5,5	4,4,4,4,4,4	0.051	0.080	0.036	0.055	0.051
A_2	5,5,5,5,5,5	2,6,10,2,6,10	0.064	0.087	0.045	0.057	0.051
B_1	10,10,10,10,10,10	4,4,4,4,4,4	0.049	0.058	0.041	0.05	0.048
B_2	10,10,10,10,10,10	2,6,10,2,6,10	0.059	0.060	0.053	0.047	0.049
C_1	5,10,15,5,10,15	4,4,4,4,4,4	0.061	0.063	0.042	0.059	0.054
C_2	5,10,15,5,10,15	2,6,10,2,6,10	0.025	0.057	0.027	0.054	0.054
C_3	5,10,15,5,10,15	10,6,2,10,6,2	0.180	0.063	0.091	0.058	0.056
D_1	10,20,30,10,20,30	4,4,4,4,4,4	0.061	0.057	0.050	0.049	0.048
D_2	10,20,30,10,20,30	2,6,10,2,6,10	0.026	0.047	0.028	0.047	0.051
D_3	10,20,30,10,20,30	10,6,2,10,6,2	0.160	0.052	0.100	0.051	0.049

Table 4 : Actual significance levels for different experimental designs of k = 6 groups

Pattern	n	σ^2	F	W	KW	AG	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	0.060	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	0.057	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	0.049	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	0.049	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	0.060	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	0.058	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	0.056	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	0.053	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	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	0.047	0.046

Table 5 : Actual significance levels for different experimental designs of k= 9 groups

Pattern	n	σ^2	μ_i	F	W	KW	AG	B^2
A_{11}	5,5,5	4,4,4	2,0,0	0.27	0.24	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	0.18
A_{21}	5,5,5	2,6,10	2,0,0	0.20	0.25	0.20	0.25	0.26
A_{22}	5,5,5	2,6,10	-1,0,1	0.17	0.17	0.15	0.15	0.16
B_{11}	10,10,10	4,4,4	2,0,0	0.58	0.55	0.53	0.54	0.55
B_{12}	10,10,10	4,4,4	-1,0,1	0.45	0.43	0.42	0.42	0.42
B_{21}	10,10,10	2,6,10	2,0,0	0.42	0.57	0.48	0.57	0.57
B_{22}	10,10,10	2,6,10	-1,0,1	0.31	0.35	0.30	0.35	0.35
C_{11}	5,10,15	4,4,4	2,0,0	0.46	0.35	0.33	0.32	0.31
C_{12}	5,10,15	4,4,4	-1,0,1	0.46	0.34	0.35	0.36	0.35
C_{21}	5,10,15	2,6,10	2,0,0	0.21	0.46	0.21	0.44	0.44
C_{22}	5,10,15	2,6,10	-1,0,1	0.21	0.34	0.19	0.33	0.33
C_{31}	5,10,15	10,6,2	2,0,0	0.44	0.18	0.27	0.16	0.16
C_{32}	5,10,15	10,6,2	-1,0,1	0.46	0.24	0.32	0.26	0.26
D_{11}	10,20,30	4,4,4	2,0,0	0.77	0.68	0.66	0.63	0.64
D_{12}	10,20,30	4,4,4	-1,0,1	0.77	0.68	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	0.81
D_{22}	10,20,30	2,6,10	-1,0,1	0.48	0.65	0.45	0.64	0.64
D_{31}	10,20,30	10,6,2	2,0,0	0.66	0.35	0.45	0.32	0.32
D_{32}	10,20,30	10,6,2	-1,0,1	0.71	0.51	0.56	0.52	0.51

Table 6 : Predicted power values for different experimental designs of k=3 groups

Pattern	n	σ^2	μ_i	F	W	KW	AG	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	0.32	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	0.29	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	0.42	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	0.24	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	0.84	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	0.70	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	0.86	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	0.57	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	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	0.47	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	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	0.44	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	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	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	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	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	0.94	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	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	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	0.64	0.69

Table 7 : Predicted power values for different experimental designs of k=6 groups

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

Table 8 : Codes of designs used in this simulation study

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

PP: Positive Pairings
NP: Negative Pairings

Table 9: Recommended tests with different experimental designs