Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

Comparison of Some Robust Wilks' Statistics for the One-Way Multivariate Analysis of Variance (MANOVA)

Abdullah A. Ameen[1] Osama H. Abbas[2]
[1] Department of Mathematics, College of Science,
University of Basrah, Basrah, Iraq,
[2] General Directorate of Education in Basrah.

[1]E-mail: dr-abd64@yahoo.com [2] E-mail: osama.stat82@gmail.com

Recived : 23\12\2018 Revised : 13 \ 1 \ 2019 Accepted : 12 \ 3 \ 2019

Available online : 16 /4/2019

Abstract:

The classical Wilks' statistic is mostly used to test hypotheses in the one-way multivariate analysis of variance (MANOVA), which is highly sensitive to the effects of outliers. The non-robustness of the test statistics based on normal theory has led many authors to examine various options. In this paper, we presented a robust version of the Wilks' statistic and constructed its approximate distribution. A comparison was made between the proposed statistics and some Wilks' statistics. The Monte Carlo studies are used to obtain performance assessment of test statistics in different data sets. Moreover, the results of the type I error rate and the power of test were considered as statistical tools to compare test statistics. The study reveals that, under normally distributed, the type I error rates for the classical and the proposed Wilks' statistics are close to the true significance levels, and the power of the test statistics are so close. In addition, in the case of contaminated distribution, the proposed statistic is the best.

Keywords: One-Way Multivariate Analysis of Variance, Outliers, Rank Order, Robustness, Minimum Covariance Determinant Estimator, Wilks' Statistic.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019

ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

1. Introduction

One-way MANOVA deals with testing the null hypothesis H_0 of equal mean vectors of multivariate normal groups. To formalize the hypothesis, let us assume that there are many independent random groups, say $k \geq 2$ groups, for every sample there are n_i multivariate normal observations $\boldsymbol{y}_{ij}, i = 1,2, \dots, k, j = 1,2, \dots, n_i$ of p dimension with mean vector $\boldsymbol{\mu}_i$ and equal covariance matrix Σ. Then, the null and alternative hypotheses can be written as follows:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_k$$

$$H_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \text{ for at least one } \boldsymbol{i} \neq \boldsymbol{j}.$$

Many of statistics used for testing H_0, one of the most widely used is Wilks' statistic Λ which is defined as (see Rencher, (2002) [9]):

$$\Lambda = \frac{|W|}{|W+B|} \qquad \dots (1)$$

where B and W are the" between" and within" of $p \times p$ matrices, respectively, have formulas:

$$B = \sum_{i=1}^{k} n_i (\overline{\boldsymbol{y}}_{i\cdot} - \overline{\boldsymbol{y}}_{\cdot\cdot})(\overline{\boldsymbol{y}}_{i\cdot} - \overline{\boldsymbol{y}}_{\cdot\cdot})^t \qquad \dots (2)$$

$$W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_{i\cdot})(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_{i\cdot})^t \qquad \dots (3)$$

where

$$\overline{\boldsymbol{y}}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{y}_{ij}, \overline{\boldsymbol{y}}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \boldsymbol{y}_{ij}, n = \sum_{i=1}^{k} n_i$$

The hypothesis H_0 is reject if $\Lambda \leq \Lambda_{\alpha,p,v_W,v_B}$ where $\Lambda_{\alpha,p,v_W,v_B}$ is the true critical values for Wilks' statistic with significance level α and degrees of freedom p, $v_W = n - k$ and $v_B = k - 1$.

Assuming that all groups originate from the multivariate normal distribution, many classical statistics are extremely sensitive to the influence of outliers (see [2]). Several statistics have been presented which are robust against possible outliers in the data. In 1985, Nath and Pavur [8] presented an alternative statistic for the one-way MANOVA depend on the rank order of the data. In the one-group, Hotelling's statistic is the basic tool for inference about the mean of a multivariate normal distribution. Willems et al. (2002) [14] introduced a robust Hotelling's statistic depend on the minimum covariance determinant (MCD) estimator. Candan and Aktas (2003) [7] proposed another robust Hotelling's statistic upon minimum volume ellipsoid (MVE) estimator. In 2010, Todorov and Filzmoser [12] introduced a robust Wilks' statistic for the one-way MANOVA depend on MCD estimator. Van Aelst and Willems (2011) [13] used S and MM-estimators to construct a robust Wilks' statistic for testing the hypotheses in the one-way MANOVA.

The effect of outliers on the Wilks' statistic will be explained in the simulation study in section 5. Therefore, we introduced another alternative robust Wilks' statistic to the classical Wilks' statistic and has approximation differs from those suggested by Todorov and Filzmoser. The MCD estimator that proposed by Rousseeuw in (1985) [10] is a highly robust estimator of location and scatter, for this purpose it is used. To increase efficiency while retaining high robustness, one can apply reweighted MCD estimator (RMCD) which is summarized in section 2. The robust Wilks' statistic is reviewed in section 3. In section 4, we construct the proposed approximation and examine its accuracy. A simulation study is used to evaluate the proposed statistical performance and to compare the different test statistics in different distribution cases in terms of significance level, the power of the test and robustness. Section 5 describes the simulation study and its results.

2. Minimum Covariance Determent (MCD) Estimator

Rousseeuw's MCD estimator (1985) looks for a subset of h observations with the lowest determinant of the sample covariance matrix, where the subset size h is selected between half and the full size of sample. The mean observations of the subset h represent the MCD location estimate T and a multiple of its covariance matrix is the MCD scatter estimate C. The effective algorithm for calculating the MCD estimates is found in most known statistical software packages such as R, $S\text{—}Plus$, SAS and $Matlab$. To increase the efficiency of the MCD estimator, a reweighted version is used. Several methods have been proposed to estimate the common covariance matrix. The method which was introduced by He and Fung (2000) [5] for S estimates and by Hubert and Van Driessen (2004) [6] for MCD estimates is used. In this method, the observations $\boldsymbol{y}_{ij}$ are centered and pooled as a single sample $Z = \boldsymbol{z}_{ij}$ to estimate the covariance matrix. First, it starts by computing the location estimates $\boldsymbol{t}_i, i = 1,2, \dots, k$ for each group as the RMCD location estimates. These group means are swept from the original observations for centralized observations

$$\boldsymbol{z}_{ij} = \boldsymbol{y}_{ij} - \boldsymbol{t}_i$$

Second, the common covariance matrix C is estimated as the RMCD covariance matrix of

the centered observations Z. Finally, the location estimate $\boldsymbol{\delta}$ of Z is used to adjust the group means $\boldsymbol{m}_i = \boldsymbol{\delta} + \boldsymbol{t}_i, i = 1, 2, \ldots, k$.
In order to increase efficiency while retaining high robustness, one can apply the RMCD estimators. By using the final obtained estimates $\boldsymbol{m}_i^*$ and C^* we can calculate the Mahalanobis distances as:

$$MD(\boldsymbol{y}_{ij}) = \sqrt{(\boldsymbol{y}_{ij} - \boldsymbol{m}_i^*)^t C^{*\,-1}(\boldsymbol{y}_{ij} - \boldsymbol{m}_i^*)},$$

from these distances, we can assign a weight w_{ij} based on appropriate weight function for each observation $\boldsymbol{y}_{ij}$.

3. The Robust Wilks' Statistic

Assuming that all groups arise from the multivariate normal distribution, the classical Wilks' statistic is very sensitive to the influence of outliers. Therefore, Nath and Pavur [8] are presented the robust Wilks' statistic Λ_{rank} depends on the ranks of the observations. Also, Todorov and Filzmoser [12] are introduced an alternative proposal for the Wilks' statistic based on RMCD estimator defined as:

$$\Lambda_R = \frac{|W_R|}{|W_R + B_R|} \quad \ldots (4)$$

where B_R and W_R are the weighted "between" and "within" matrices, respectively, given by:

$$B_R = \sum_{i=1}^{k} w_{i\cdot}(\bar{\boldsymbol{y}}_{w_{i\cdot}} - \bar{\boldsymbol{y}}_{w_{\cdot\cdot}})(\bar{\boldsymbol{y}}_{w_{i\cdot}} - \bar{\boldsymbol{y}}_{w_{\cdot\cdot}})^t \ldots (5)$$

$$W_R = \sum_{i=1}^{k}\sum_{j=1}^{n_i} w_{ij}(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{w_{i\cdot}})(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{w_{i\cdot}})^t \ldots (6)$$

where

$$w_{i\cdot} = \sum_{j=1}^{n_i} w_{ij} \ , \qquad \bar{\boldsymbol{y}}_{w_{i\cdot}} = \frac{1}{w_{i\cdot}}\sum_{j=1}^{n_i} w_{ij}\,\boldsymbol{y}_{ij} \ ,$$

$$\bar{\boldsymbol{y}}_{w_{\cdot\cdot}} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{w_{ij}\,\boldsymbol{y}_{ij}}{w} \ , \quad w = \sum_{i=1}^{k} w_{i\cdot} \ ,$$

and the weight w_{ij} for each observation $\boldsymbol{y}_{ij}$ computed by the Huber weight function defined as

$$w_{ij} = \begin{cases} 1 & , \ MD(\boldsymbol{y}_{ij}) \le \sqrt{\chi^2_{p,0.975}} \\ 0 & , \ \text{otherwise.} \end{cases}$$

4. The proposed approximation distribution of Wilks' statistic

The distribution of classical Wilks' statistic Λ, which was considered by Anderson (1958) [1] as a ratio of two Wishart distributions, is very complicated. Therefore, Bartlett introduced a good approximation of the Wilks' statistic given by (see [9]):

$$-\left(v_E - \tfrac{1}{2}(p - v_H + 1)\right) ln\,\Lambda \approx \chi^2_{p\,v_H} \ \ldots (7)$$

Todorov and Filzmoser are assumed for Λ_R the following approximation:

$$L_R = -\,ln\,\Lambda_R \approx d\chi^2_q \quad \ldots (8)$$

where the multiplication factor d and the degrees of freedom q of the χ^2 distribution as

$$d = \frac{E(L_R)}{q} \ , \qquad q = \frac{2E(L_R)^2}{Var(L_R)}.$$

The mean $E[L_R]$ and variance $Var[L_R]$ of the approximation L_R are not possible to obtain analytically. So, they are determined by simulation after repeated m times as:

$$ave[L_R] = \frac{1}{m}\sum_{i=1}^{m} L_R^{(i)} \ ,$$

$$var[L_R] = \frac{1}{m-1}\sum_{i=1}^{m}\left(L_R^{(i)} - ave[L_R]\right)^2$$

The estimated parameters d and q will be reused to analyze data with the same dimension and number of groups.
To perform the robust Wilks' statistic that proposed by Todorov and Filzmoser, it will take a lot of time during simulations to find d and q for approximate distribution. Therefore, in this the present study, the same robust Wilks' statistic, which proposed by Todorov and Filzmoser is used but with different weight function, denoted as the modified robust Wilks' statistic Λ_{MR}. The novel of the study is that, construct another an approximate distribution for this statistic.
The matrices B and W in (2) and (3) can be written as formulas:

$$B = Y^t(A_n - \tfrac{1}{n}J_n)Y \ , \quad W = Y^t(I_n - A_n)Y$$

where Y is the data matrix, $J_n = 1_n 1_n^t$ and $A_n = diag(\frac{1}{n_i}J_{n_i})$ is a block diagonal matrix with $k \times k$ blocks of size $n_i \times n_i$, $i = 1, 2, \ldots, k$. Also, the degrees of freedom $v_W = trace(I_n - A_n) = n - k$ and $v_B = trace(A_n - \frac{1}{n}J_n) = k - 1$.
Analogously we can be written the matrices B_R and W_R in (5) and (6) as formulas:

$$B_R = Y^t(Q_n - P_n)Y \ , \ W_R = Y^t(W_n - Q_n)Y,$$

where $Q_n = diag(Q_{ii})$, $Q_{ii} = \left[\frac{1}{w_i}w_{ij}w_{ih}\right]$ is a block diagonal matrix with $k \times k$ blocks of size $n_i \times n_i$, $P_{ii} = \left[\frac{1}{w}w_{ij}w_{ih}\right]$ is a block

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019

ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

matrix with $k \times k$ blocks of size $n_i \times n_i$, and $W_n = diag(w_{ii})$,$w_n = diag(w_{ij})$,$i = 1, 2, \dots, k$, j and $h = 1,2,\dots,n_i$.

The null hypothesis H_0 is reject if $\Lambda_{MR} \leq \Lambda_{\alpha,p,v_{W_R},v_{B_R}}$ where $\Lambda_{\alpha,p,v_{W_R},v_{B_R}}$ is the exact critical values for Wilks' statistic with significance level α and degrees of freedom p, v_{W_R} and $v_{B_R}$where

$$v_{W_R} = trace(W_n - Q_n) = w - \sum_{i=1}^{k} \frac{v_i}{w_i} \ ,$$

$$v_{W_R} = trace(Q_n - P_n) = \sum_{i=1}^{k} \frac{v_i}{w_i} - \frac{\sum_{i=1}^{k} v_i}{w}$$

and $v_i = \sum_{j=1}^{n_i} w_{ij}^2$. The weight w_{ij} for each observation y_{ij} computed by Hampel weight function (see Campbell, (1980) [3]) as:

$$w_{ij} = \begin{cases} 1 & , \ MD(y_{ij}) \leq d_0 \\ d/\,MD(y_{ij}) \ , & \text{otherwise,} \end{cases}$$

where

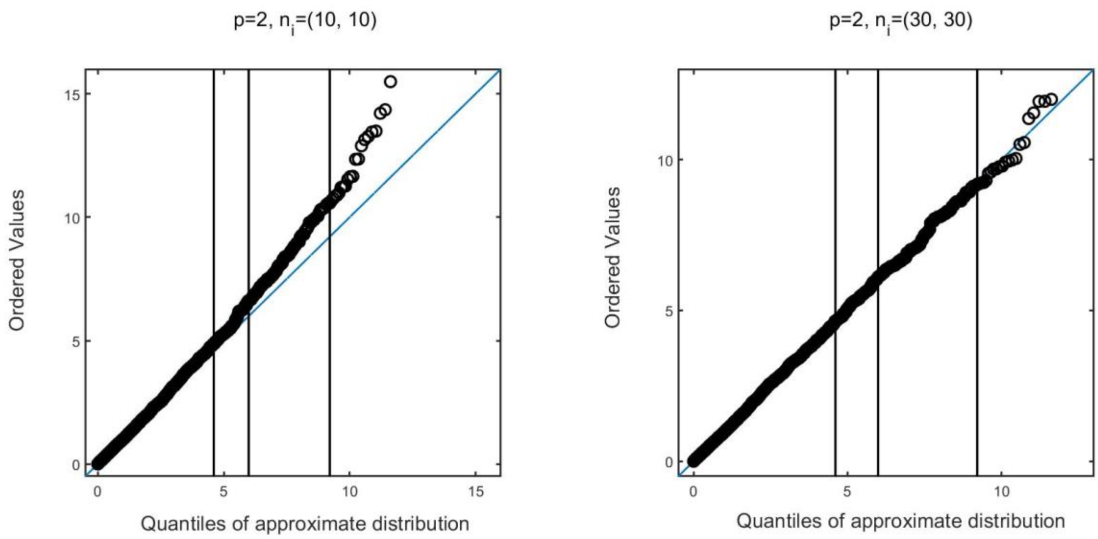
$$d = d_0 exp\left(-\frac{1}{2}\left(\frac{MD(y_{ij}) - d_0}{b_2}\right)^2\right),$$

$$d_0 = \sqrt{p} + \frac{b_1}{\sqrt{2}}, \ b_1 = 2, \ b_2 = 1.25.$$

Similarly, to the χ^2 approximation of the classical Wilks' statistic Λ, we can assume for Λ_{MR} the following approximation:

$$-\left(v_{W_R} - \frac{1}{2}(p - v_{B_R} + 1)\right) ln\,\Lambda_{MR} \approx \chi^2_{pv_{B_R}} \ \dots (9)$$

Now we will investigate the accuracy of the approximation in (9) by computing the robust Wilks' statistic Λ_{MR} for $m = 3000$ samples from the standard normal distribution and several values of the dimension p, the number of groups k and the sample sizes n_i, $i = 1,2,\dots,k$. The distribution of these m statistics will be compared to the approximate distribution in (9) by QQ-plots, some of them are shown in Figure 1. The usual cutoff values of a test, 95%, 97.5%, and 99% are shown in these plots of vertical lines. One can see from these plots that the approximation is accurate for lower and higher dimensions, large and small sample sizes, and for equal and unequal groups sizes.



Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

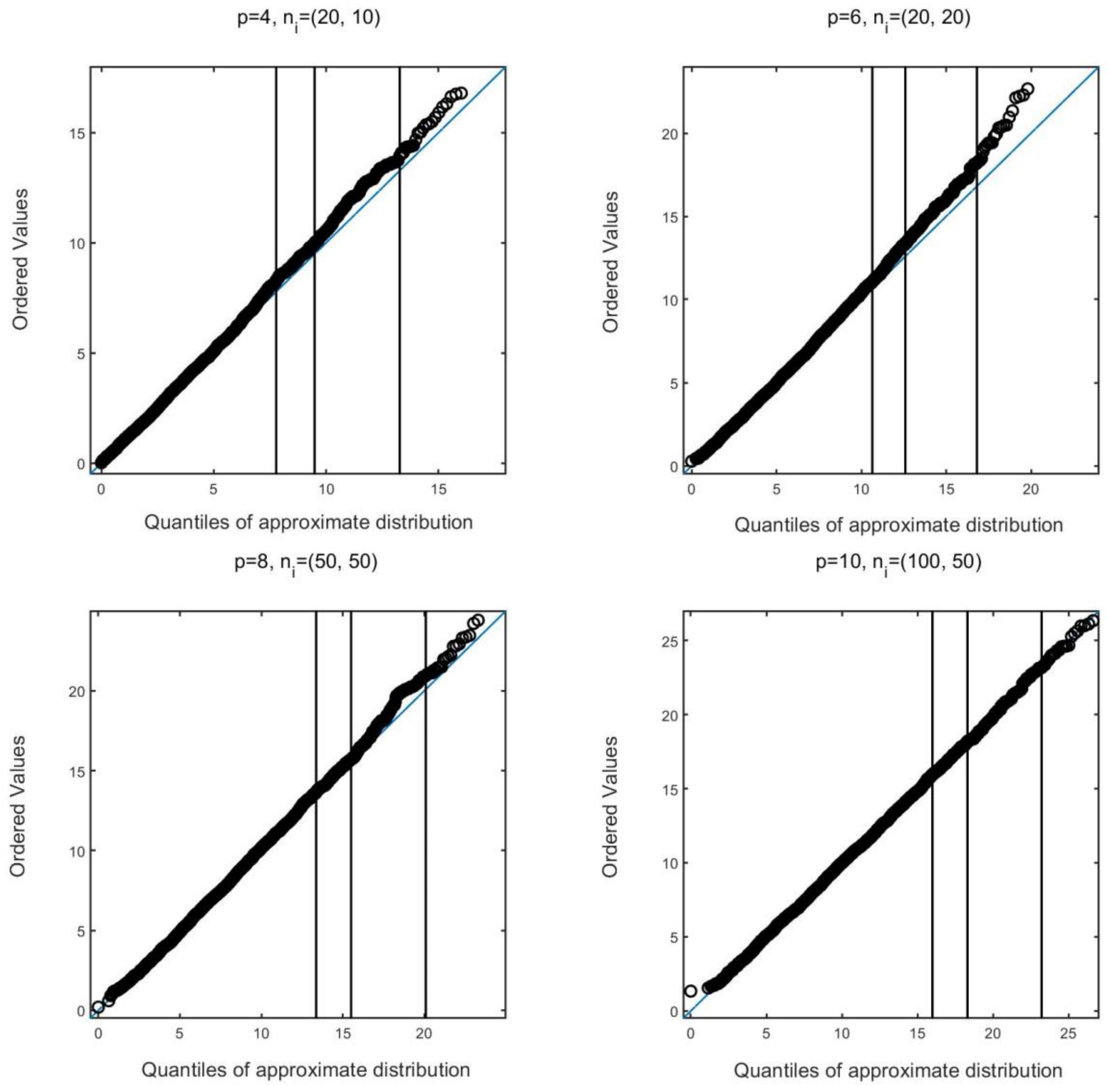


Figure 1: QQ-plots for the modified robust Wilks' statistics Λ_{MR} in the case of three groups and several dimension values for p and $n = \sum_{i=1}^{k} n_i$.

5. Monte Carlo Simulation

Monte Carlo study is a good method to assess the statistical performance for the test statistics. The evaluation of the performance of the test statistics includes two measures the type I error rate and the power of the test. In addition, we will investigate the robust statistics behavior in the existence of outliers and compare the results with the classic Wilks' statistic and the Wilks' statistic based on rank orders of data. To study the type I error rate and the power of test of the robust statistics, let us consider number of groups $k = \{2,3\}$, several dimension $p = \{2, 4, 6, 8, 10\}$, and sample sizes $n_i, i = 1,2, \ldots, k$. The selected sample sizes are shown in Table 1.

Table 1: Selected sample sizes for the simulation study

Two groups (n_1, n_2)	Three groups (n_1, n_2, n_3)
$(10, 10)$	$(10, 10, 10)$
$(20, 10)$	$(20, 10, 10)$
$(20, 20)$	$(20, 20, 20)$
$(30, 20)$	$(30, 20, 10)$
$(30, 30)$	$(30, 30, 10)$
$(50, 20)$	$(50, 20, 10)$
$(50, 50)$	$(50, 50, 20)$
$(100, 50)$	$(100, 50, 30)$

5.1 Significance level

To compare the type I error rates $\hat{\alpha}$ for the test statistics, we generate the observations from the multivariate normal distribution

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019

ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

$\boldsymbol{y}_{ij} \sim N_p(\boldsymbol{0}, I)$ under the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k$. The classical Wilks' statistic Λ and the robust Wilks' statistic Λ_{rank} based on the ranks are compared to the Bartlett' χ^2 approximation given in (7), the robust Wilks' statistic Λ_R is compared to the approximation given in (8) and the proposed Wilks' statistic Λ_{MR} is compared to the approximate distribution given in (9). This is repeated $m = 3000$ times and then calculate $\hat{\alpha} = L(T)/m$ (where $L(T)$ is the number of times of rejected the test statistic when the hypothesis is true) for the test statistics above. The values $\hat{\alpha}$ are taken as an estimate of the true significance level when the simulated critical values are above the true significance level. The true significance level $= 0.01$, 0.05, and 0.10 with the number of times $m = 3000$, and from the standard error formula of Salter and Fawcett (1989) [11] $\alpha \pm 2\sqrt{\alpha(1-\alpha)/R}$ gives the standard deviation interval about the nominal level as $(0.089, 0.111)$, $(0.042, 0.058)$, and $(0.006, 0.014)$ respectively. In Table 2, the results of the type I error rates $\hat{\alpha}$ are shown for two groups. It is clear that $\hat{\alpha}$ of the test statistics are very close to the nominal value α (true significance level). We will use the p-value plots proposed by Davidson and McKinnon (1998) [4], which gives a more complete picture of how the test statistics follow the approximate distribution under the null hypothesis in the simulated samples. Figures 2 and 3 show p-value plots of test statistics in three groups $k = 3$ of the multivariate normal distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. It is seen that the test statistics Λ, Λ_{rank}, and Λ_{MR} are close to the 45° line, and the robust Wilks' statistic Λ_R is considerably below the 45° line for small sample sizes.

5.2 Power of test

To compare the power of the test $\hat{\pi}$ for the test statistics we will generate data samples $Y_i \sim N_p(\boldsymbol{\mu}_i, I)$ under an alternative hypothesis $H_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for at least one i ≠ j. Also, we will use the same cases of dimensions p, number of groups k, and sample sizes $n_i, i = 1, 2, \ldots, k$ but each sample has a different mean $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})^t$. The means of dimensions $p = 2, 4, 6, 8, 10$ for the groups $i = 1, 2, 3$ are selected as:

$\boldsymbol{\mu}_1 = (0, 0, \ldots, 0)^t$, $\boldsymbol{\mu}_2 = (0, 0.5, \ldots, 0)^t$, $\boldsymbol{\mu}_3 = (0, 0, 0.5, \ldots, 0)^t, \ldots$ $\boldsymbol{\mu}_k = (0, 0, \ldots, 0.5, 0)^t$

The power of the test statistics were compared by the resulting size-power curves under alternative hypothesis, as proposed by Davidson and MacKinnon (1998) [4].

The results for the three groups are shown in Figures 4 and 5. It is clearly seen that the size-power curves for the classical statistic Λ, the rank-transformed Wilks' statistic Λ_{rank}, and the proposed statistic Λ_{MR} are close while the robust Wilks' statistic Λ_R by Todorov and Filzmoser is less.

5.3 Robustness comparisons

Now we will investigate the robustness for the proposed test statistic in the one-way MANOVA. Therefore, we will generate data samples under the null and alternative, and we will contaminate them by adding outliers. The same cases of dimensions p, number of groups k, and sample sizes $n_i, i = 1, 2, \ldots, k$ will be used.

5.3.1 Significance level

Under the hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k$, the data will be generated from the following contamination model:

$\boldsymbol{y}_{ij} \sim (1-\varepsilon) N_p(\boldsymbol{0}, I) + \varepsilon\, N_p(\boldsymbol{\mu}^*, cI)$, where $\varepsilon = 0.1$, $\boldsymbol{\mu}^* = v\sqrt{\chi^2_{p,0.001}1^t_p}$, $v = 5$, and $c = 0.0625$. The p-value plots of the test statistics for three groups are shown in Figures 6 and 7. In these Figures, the p-value plots (actual size) based on the test statistics Λ_{MR}, is so close to the 45° line compared to the same of the test statistic Λ_R, while the classical statistic Λ and the rank transformed statistic Λ_{rank} are very bad for all the different cases of dimension p and sample sizes.

5.3.2 Power of test

Under the alternative hypothesis $H_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for at least one $\boldsymbol{i} \neq \boldsymbol{j}$, the data samples will be generated from the following contamination model:

$$Y_i \sim (1-\varepsilon) N_p(\boldsymbol{\mu}_i, I) + \varepsilon\, N_p(\boldsymbol{\mu}^*, cI),$$

where $\boldsymbol{\mu}_i$ is the same mean groups value as in section (5.1), $\varepsilon, \boldsymbol{\mu}^*$, and c is take the same values as in section (5.3.1). The Figures 8 and 9 show the size-power curves of test statistics. It is clearly seen that the proposed robust Wilks' statistic Λ_{MR} is the best compared to the other statistics for all investigated cases of dimension p and sample sizes.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019

ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

Table 2: Levels of significance of test statistics Λ, Λ_{rank}, Λ_R, and Λ_{MR} for two groups $k = 2$ of multivariate normal distribution, several values of the dimension p and the sample size $n = n_1 + n_2$.

Dimension p	Sample Size n_1 n_2	Statistic	Significance Level		
			0.01	0.05	0.10
2	10 10	Λ	0.010	0.052	0.105
		Λ_{rank}	0.012	0.051	0.107
		Λ_R	0.012	0.039	0.075
		Λ_{MR}	0.014	0.061	0.117
	30 30	Λ	0.009	0.048	0.093
		Λ_{rank}	0.012	0.048	0.098
		Λ_R	0.012	0.049	0.092
		Λ_{MR}	0.010	0.051	0.098
4	20 10	Λ	0.013	0.057	0.109
		Λ_{rank}	0.015	0.059	0.107
		Λ_R	0.017	0.046	0.088
		Λ_{MR}	0.015	0.062	0.116
	50 20	Λ	0.010	0.048	0.097
		Λ_{rank}	0.011	0.049	0.103
		Λ_R	0.012	0.047	0.094
		Λ_{MR}	0.011	0.051	0.100
6	20 20	Λ	0.013	0.054	0.107
		Λ_{rank}	0.015	0.052	0.105
		Λ_R	0.013	0.052	0.085
		Λ_{MR}	0.013	0.057	0.114
	50 50	Λ	0.008	0.043	0.097
		Λ_{rank}	0.010	0.041	0.090
		Λ_R	0.012	0.047	0.094
		Λ_{MR}	0.009	0.045	0.099
8	20 20	Λ	0.010	0.055	0.107
		Λ_{rank}	0.014	0.058	0.113
		Λ_R	0.016	0.053	0.094
		Λ_{MR}	0.020	0.069	0.123
	50 50	Λ	0.009	0.047	0.097
		Λ_{rank}	0.010	0.051	0.099
		Λ_R	0.012	0.048	0.098
		Λ_{MR}	0.009	0.049	0.099
10	30 30	Λ	0.010	0.045	0.100
		Λ_{rank}	0.015	0.052	0.101
		Λ_R	0.011	0.054	0.091
		Λ_{MR}	0.010	0.046	0.104
	100 50	Λ	0.011	0.051	0.102
		Λ_{rank}	0.011	0.053	0.102
		Λ_R	0.013	0.047	0.097
		Λ_{MR}	0.012	0.053	0.102

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

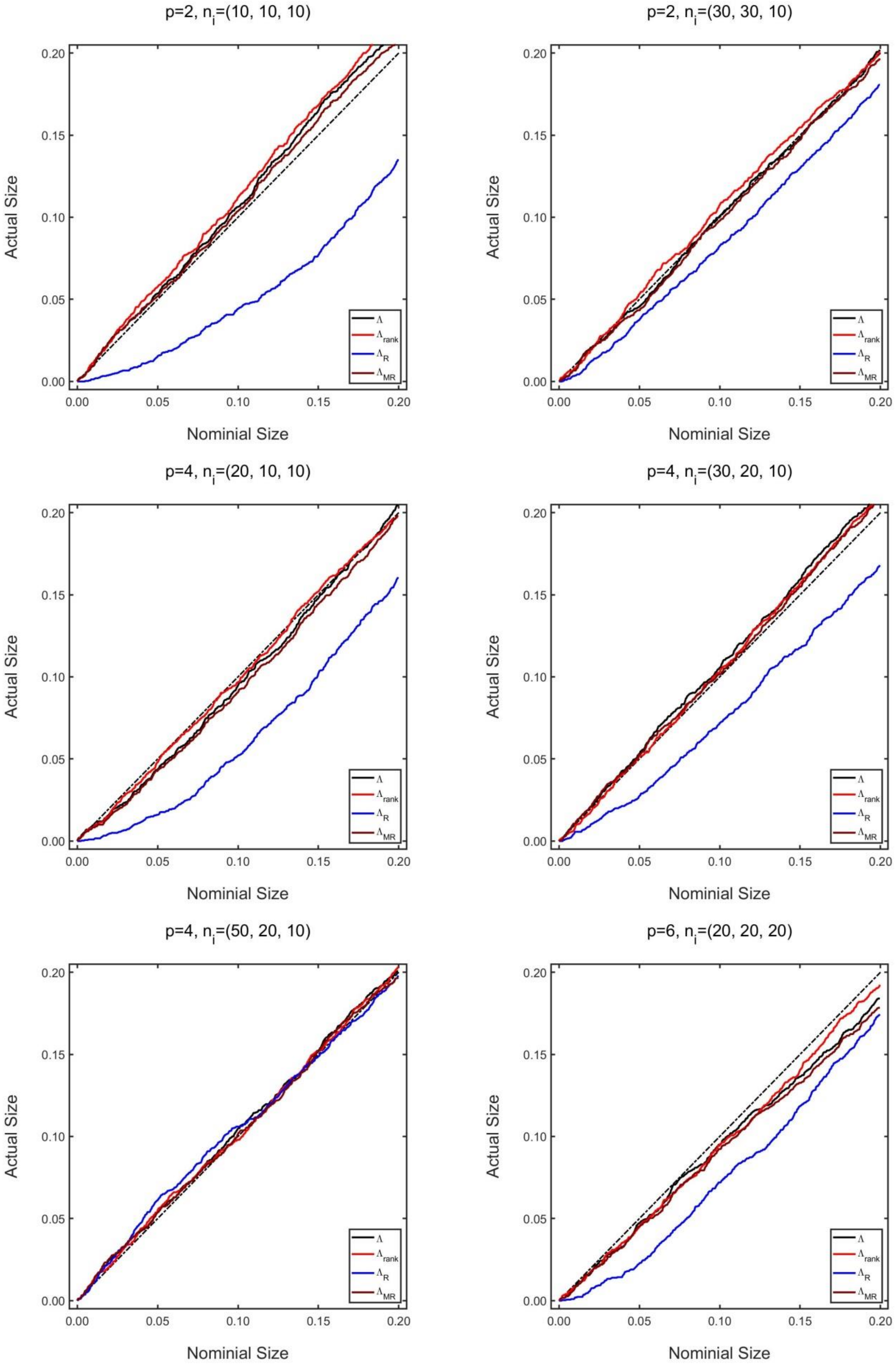


Figure 2: P-value plots for test statistics Λ (black line), Λ_{rank} (red line), Λ_R (blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate normal distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019

ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

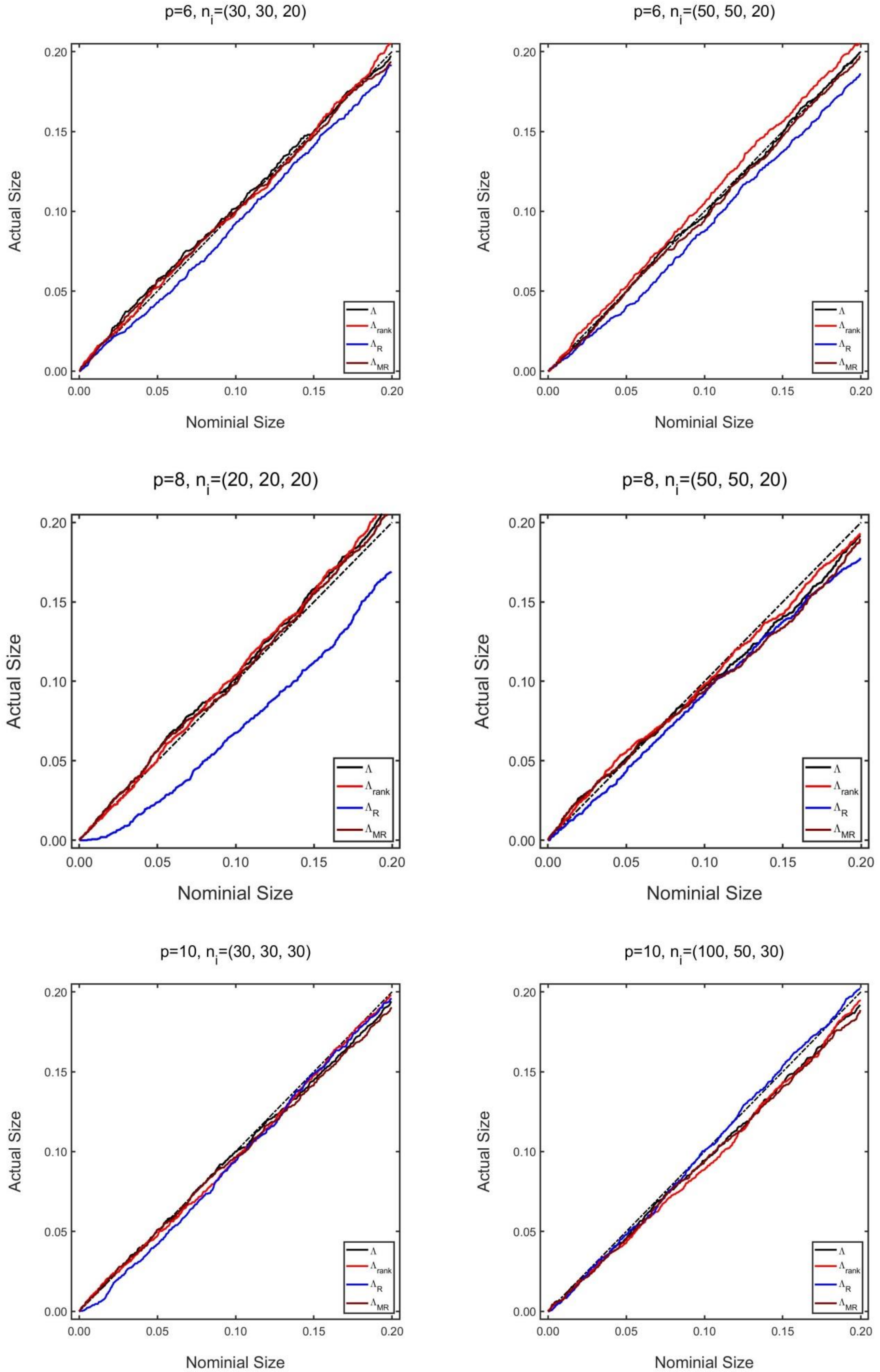


Figure 3: P-value plots for test statistics Λ (black line), Λ_{rank} (red line), Λ_R (blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate normal distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

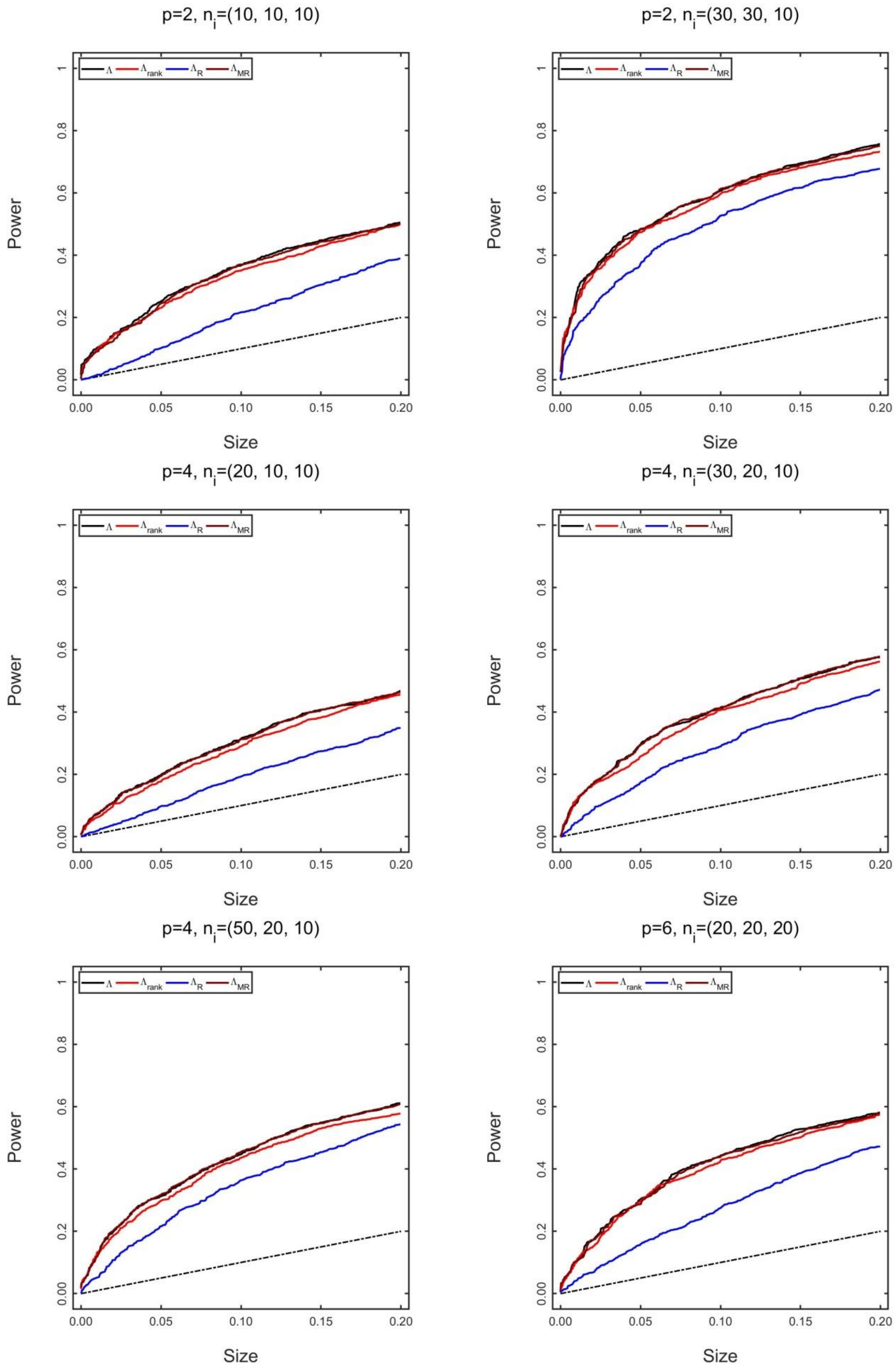


Figure 4: Size-power curves for test statistics Λ (black line), Λ_{rank} (red line), Λ_R (blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate normal distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

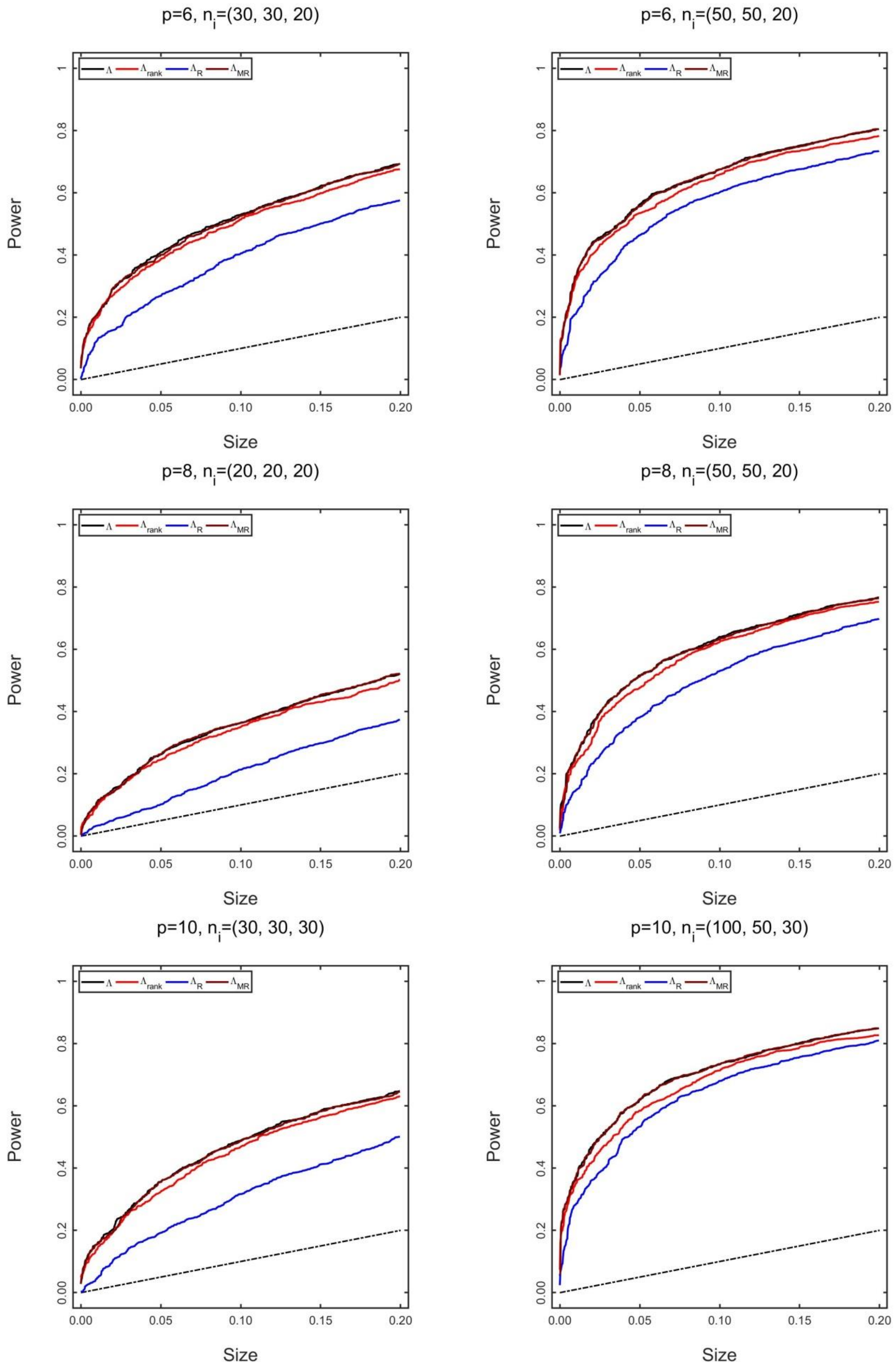


Figure 5: Size-power curves for test statistics Λ (black line), Λ_{rank} (red line), Λ_R (blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate normal distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

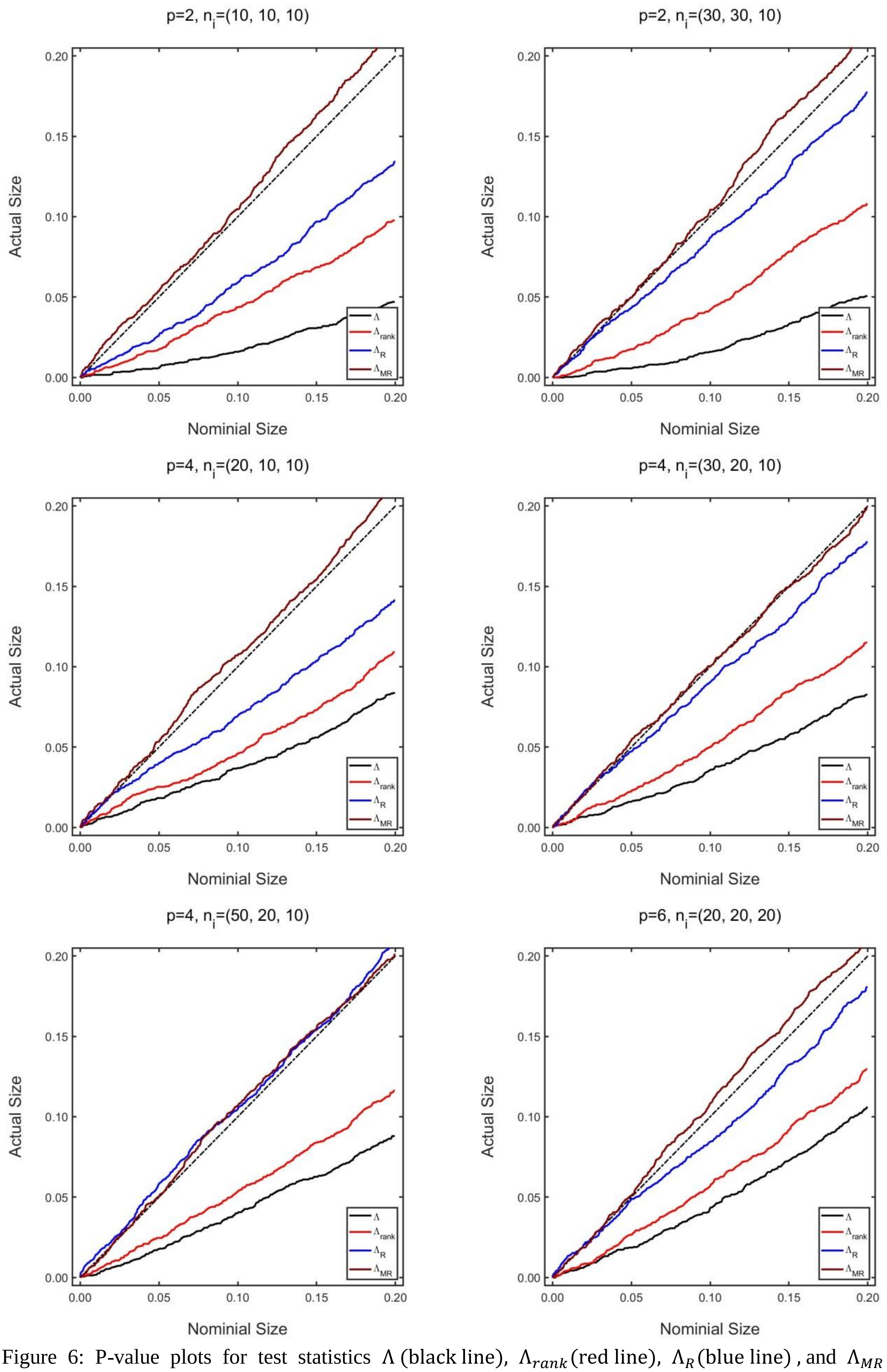


Figure 6: P-value plots for test statistics Λ (black line), Λ_{rank} (red line), Λ_R (blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate contaminated distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

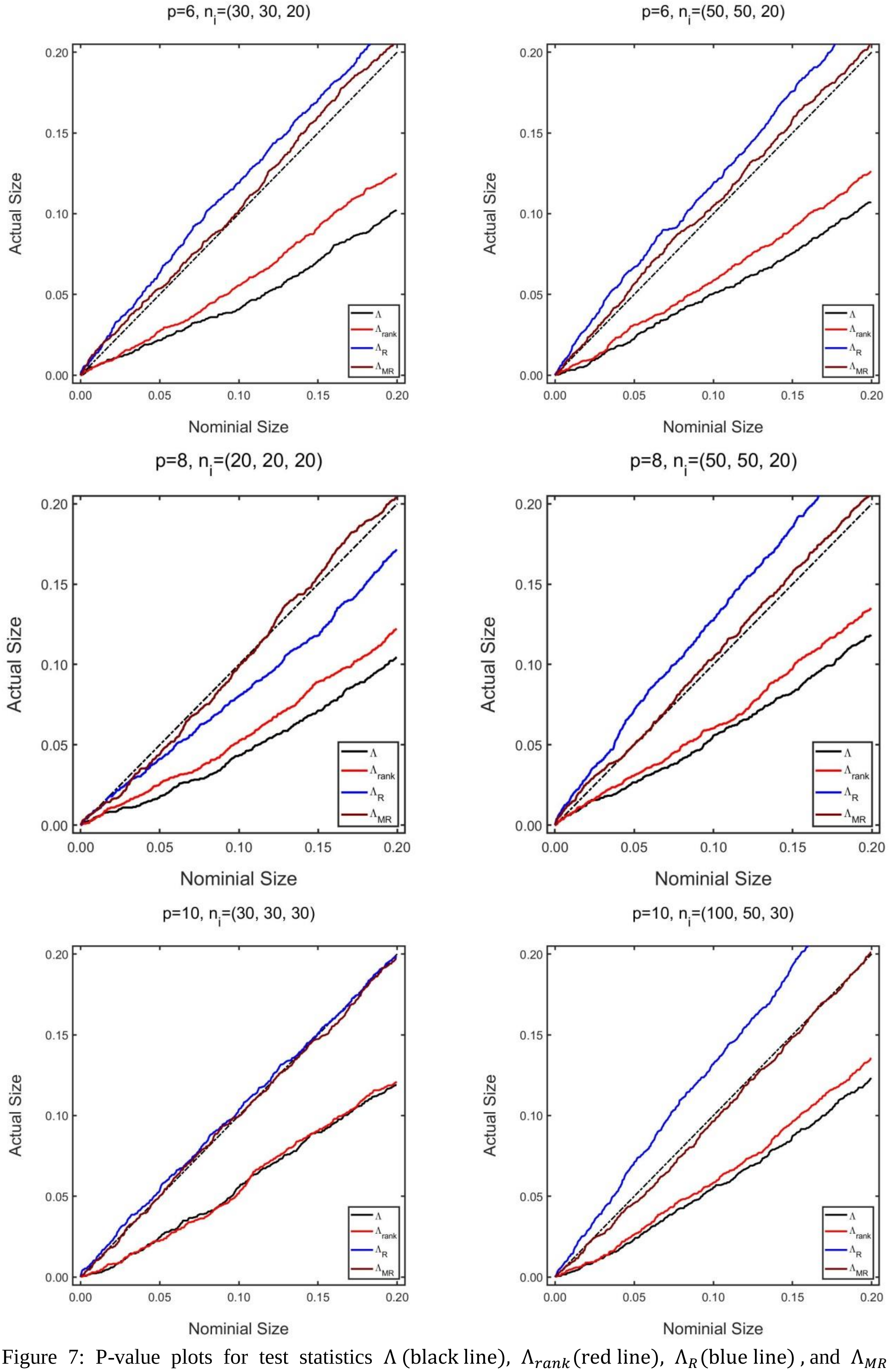


Figure 7: P-value plots for test statistics Λ (black line), Λ_{rank} (red line), Λ_R (blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate contaminated distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

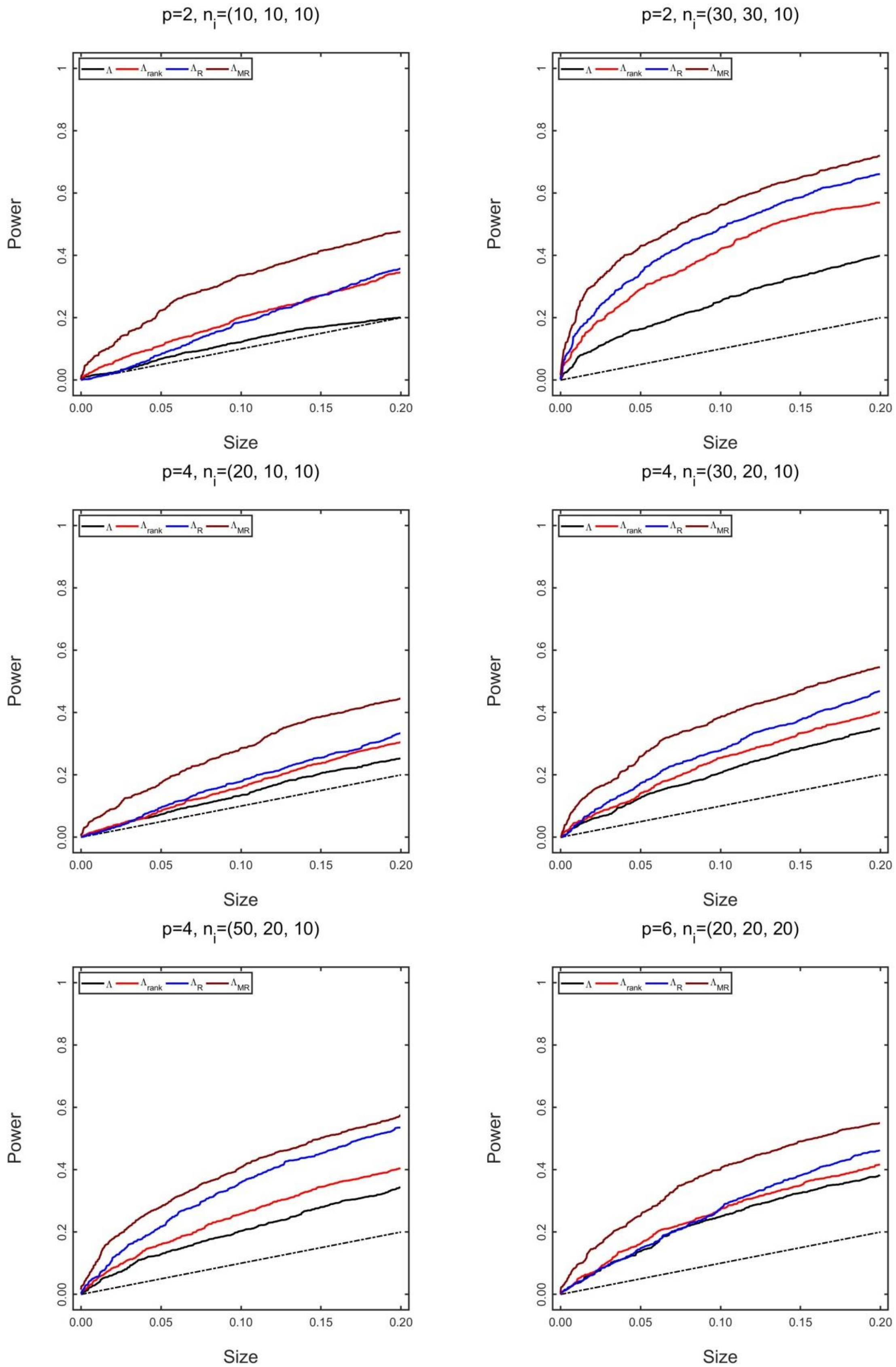


Figure 8: Size-power curves for test statistics Λ (black line), Λ_{rank} (red line), Λ_R (blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate contaminated distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.2 Year 2019

ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

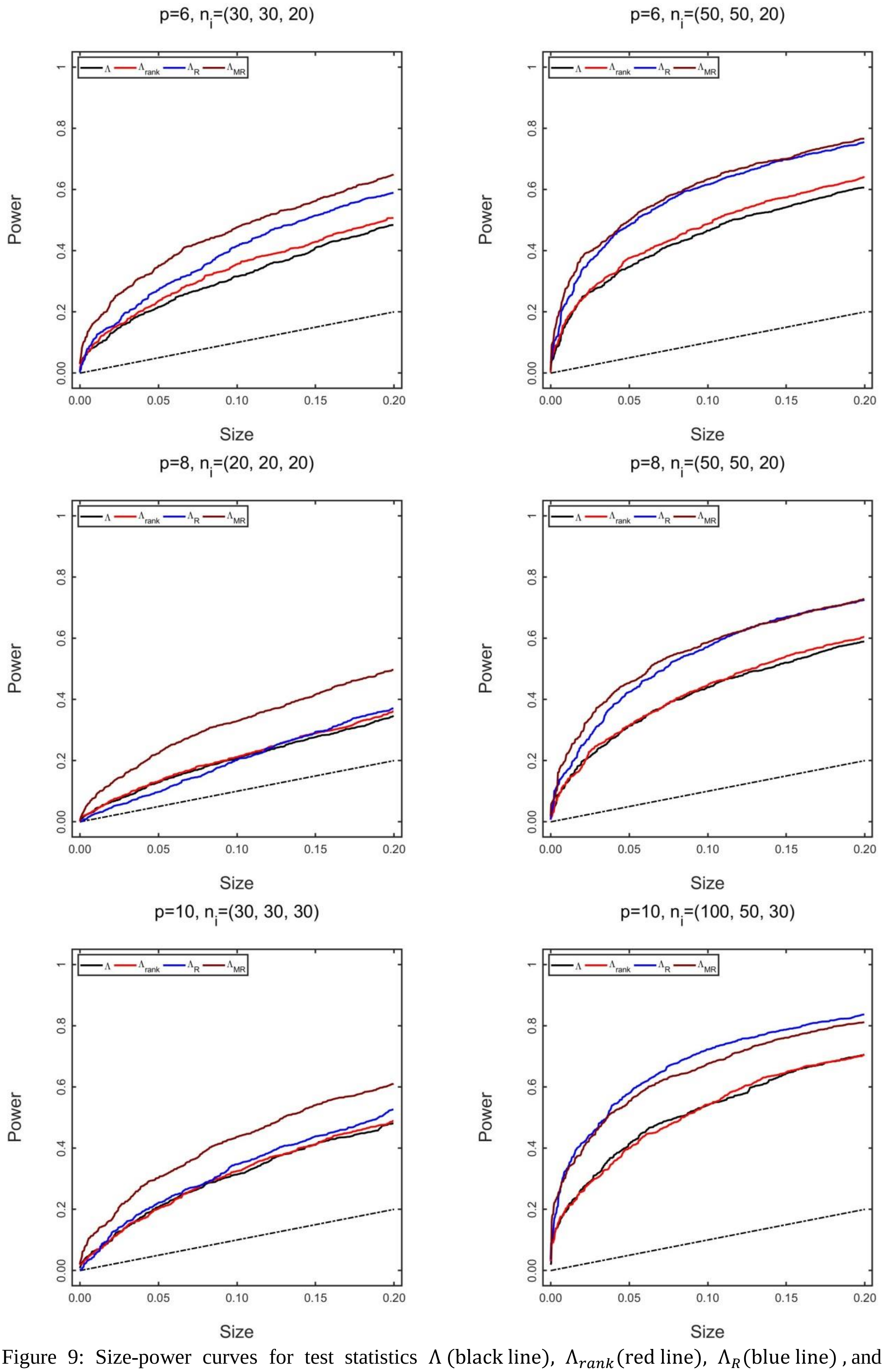


Figure 9: Size-power curves for test statistics Λ (black line), Λ_{rank}(red line), Λ_R(blue line) , and Λ_{MR} (dark red line) for three groups $k = 3$ of multivariate contaminated distribution, several dimensions p and the sample size $n = \sum_{i=1}^{k} n_i$. The 45 line is given too.

6. Conclusions

In this study, we presented a robust version of the Wilks' statistic and constructed its approximate distribution. The results show that the p-value plots and size-power curves for the proposed robust statistic are close to the classical and the rank transformed Wilks' statistics in case of normal distribution for the data set, while in case of contaminated distribution the proposed robust statistic is the best. Also, the results show the advantage of the proposed robust statistic over the robust Wilks' statistic of Todorov, and Filzmoser especially with small sample sizes.

References

[1] Anderson, T. (1958). "An Introduction to Multivariate Statistical Analysis", John Wiley and Sons, New York.
[2] Beak, R. C., Cook, R. D. (1983). "Outliers", Technometrics, 25, 119-149.
[3] Campbell, N. A. (1980). "Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation", Applied Statistics, 29, 231-237.
[4] Davidson, R., McKinnon, J., 1998. Graphical methods for investigating the size and power of hypothesis tests, The Manchester School of Economic & Social Studies, 66 (1), pp. 1-26.
[5] He, X. and Fung, W. K. (2000). "High breakdown estimation for multiple populations with applications to discriminant analysis". J. Multivariate Anal, 72, 151–162.
[6] Hubert, M. and Van Driessen, K. (2004). "Fast and robust discriminant analysis", Comput. Statist. Data Anal, 5, 301-320.
[7] Meral Candan, and Serpil Aktas (2003). "Hotelling's Statistic Based on Minimum Volume Ellipsoid Estimator" , G. U. Journal of Science , 16(4) , 691-695.
[8] Nath, R., and Pavur, R. (1985). "A new statistic in the one-way multivariate analysis of variance", Computational Statistics and Data Analysis, 2(4), 297-315.
[9] Rencher, A. C. (2002). "Methods of Multivariate Analysis", Second Edition, Brigham Young University, John Wiley and Sons, New York.
[10] Rousseeuw, P. J. (1985). "Multivariate Estimation with High Breakdown Point", In Mathematical Statistics and Applications, Vol. B, 283-297.
[11] Salter, K.C., and Fawcett, R.F. (1989), "A robust and powerful rank test of treatment effects in balanced incomplete block designs", Communications in Partial Differential Equations, 14(4), 807-828.
[12] Todorov, V., and Filzmoser, P. (2010). "Robust Statistic for the One-way MANOVA", Computational Statistics and Data Analysis, 54(1), 37-48.
[13] Van Aelst, S. and Willems, G. (2011). "Robust and Effcient One-way MANOVA Tests", Journal of the American Statistical Association, 106-494.
[14] Willems, G., Pison, G., Rousseeuw, P. J., and Van Aelst, S. (2002). "A robust Hotelling Test", Metrika, 55,125-138.

ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

مقارنة بين بعض احصاءات ويلكس الحصينة في تحليل التباين المتعدد المتغيرات في اتجاه واحد

عبدالله عبدالقادر أمين[1] أسامة هادي عباس[2]

[1] قسم الرياضيات - كلية العلوم - جامعة البصرة / البصرة – العراق
[2] المديرية العامة لتربية البصرة / البصرة – العراق

المستخلص :

تستخدم احصاءة ويلكس الكلاسيكية في الغالب لاختبار الفرضيات في تحليل التباين متعدد المتغيرات في اتجاه واحد، حيث تكون شديدة الحساسية ازاء تأثير القيم الشاذة. عدم حصانة الاختبارات الاحصائية المبنية على النظرية الطبيعية قاد العديد من الباحثين للبحث عن بدائل. في هذا البحث، قدمنا نسخة حصينة من أحصاءة ويلكس وبنينا توزيعها التقريبي. تم تنفيذ المقارنة بين الاحصاءة المقترحة وبين بعض أحصاءات ويلكس. دراسة مونت كارلو استخدمت لتقييم أداء إحصاءات الاختبار في مختلف حالات مجموعة البيانات. أضافة الى ذلك، تم اعتبار نتائج معدل الخطأ من النوع الأول وقوة الاختبار بمثابة ادوات إحصائية للمقارنة بين احصاءات الاختبار. أظهرت الدراسة في التوزيع الطبيعي أن معدلات الخطأ من النوع الأول لإحصائيات ويلكس الكلاسيكية والحصينة قريبة من مستويات المعنوية، وقوة الاختبار لإحصاءات ويلكس متقارب جدا. بالإضافة إلى ذلك، في حالة التوزيع الملوث، فإن الإحصائية المقترحة هي الأفضل.