Robust Hotelling's Statistic for Test A Hypothesis of Mean Multivariate Population Based on RMCD

Abdullah A. Ameen **Mohand N. Abdul - Seid**

Department of Mathematics, College of Science
Basra University, Basra, Iraq.

Abstract

Hotelling's statistic T^2 is an important to test a hypothesis about the mean of a multivariate normal population with location and scale parameters (with the proposition that the scale parameter is unknown positive definite matrix) . However, hypothesis test based on this statistic can be adversely affected by outliers. In this paper, an alternative technique is proposed based on a statistic which uses the reweighted minimum covariance determinant (RMCD) estimators instead of the classical mean vector and covariance matrix . A simulation technique has been used as a technique to make a comparison between the classical and the proposed statistic by generating the data that have a contaminated multivariate normal distribution from one side and from two sides . The results have shown that, the proposed robust statistic is almost better than the classical statistic depending on the rates of type I error and the power test.

Keywords: Hotelling's statistic T^2, Minimum Covariance Determinant , Robustness , One- Sample Hypothesis Test.

1. Introduction

It is known that some statistical methods are very sensitive to outliers(1). Thus, in order to reduce the influence of outliers, the robust methods have been proposed(2). In addition, robust estimators of location and scale parameters are played a very important role in statistics. Rousseeuw (3) introduced two estimators, minimum volume ellipsoid (MVE) estimator and minimum covariance determinant (MCD) estimator. The test of hypothesis that based on Hotelling's statistic T^2 can be adversely affected by outliers(3,4) suggested the robust Hotelling's statistic based on minimum covariance determinant (MCD) estimator to test the hypothesis about location parameter of one group. Moreover, Meral Candan and Serpil Aktas(5) implemented, the robust Hotelling's statistic based on minimum volume ellipsoid (MVE) estimator to test the hypothesis about location parameter of one group. Abdullah Ameen (6) suggested the robust Hotelling's statistic based on reweighted minimum covariance determinant (RMCD) estimator to test a hypothesis for equality of two multivariate means .

In the present study, the robust statistic for Hotelling based on RMCD estimator is used to test the hypothesis about location parameter of one group.

2. Hotelling's Statistic

Hotelling's statistic T^2 which has been discussed by Hotelling [7] is the standard tool to test the hypothesis about the mean μ of a multivariate normal population $N_p(\mu,\Sigma)$ where the scale parameter Σ is unknown. The hypothesis $H_0:\underline{\mu}=\underline{\mu}_0$ is rejected at the level α if

$$T^2 = n(\bar{\underline{X}}-\underline{\mu}_0)'S^{-1}(\bar{\underline{X}}-\underline{\mu}_0) > \frac{(n-1)p}{n-p}F(p,n-p;\alpha)$$

Where, $\bar{\underline{X}} = \frac{\sum_{i=1}^{n}\underline{X}_i}{n}$ and $S = \frac{\sum_{i=1}^{n}(\underline{X}_i-\bar{\underline{X}})(\underline{X}_i-\bar{\underline{X}})'}{n-1}$ are the mean vector and covariance matrix of the sample $\underline{X}_1,\underline{X}_2,\ldots,\underline{X}_n$ such that $\underline{X}'_i=(x_{i1},x_{i2},\ldots,x_{ip})$ for each $.\ i=1,2,\ldots,n$

The use of the classical Hotelling's statistic for the data suffer from that, one of the main conditions is not satisfy for testing the above hypothesis ,leads to a wrong decision. On the other hand, if the data have outlier values , the distribution of the test statistic will be effected. Consequently, that will affect the probability of type I and type II error alongside with the difficulty to control the significance level. Therefore, to overcome these difficulties, a robust statistic for Hotelling is suggested based on RMCD estimator.

3. Robust Estimation Methods

There are many robust methods used to estimate the location and scale parameters of multivariate distributions ; including M-estimators that proposed by Maronna, MVE and MCD estimators that proposed by Rousseeuw[3], as mentioned above.

In this article, the RMCD method is adopted to find the robust estimators $\bar{\underline{X}}_w$ and S_w of location and scale parameters μ and Σ respectively, based on the weights function which is defined by Hample[8] as:

$$w(d)=\begin{cases}1 & \text{if}\quad d\le d_0\\ d_0\exp[-\frac{1}{2}(d-d_0)^2/b_2^2] & \text{if}\quad d>d_0\end{cases},$$

such that $d_0=\sqrt{p}+b_1/\sqrt{2}$, $b_1=2$, $b_2=1.25$ and

$$d_j^2(i)=(\underline{X}_i-\bar{\underline{X}}_j)'S_j^{-1}(\underline{X}_i-\bar{\underline{X}}_j)\quad,\forall i=1,2,\ldots,n\ ,\ j=0,1,\ldots$$

(is called Mahalonobis distances) .

Since, the calculation of Mahalonobis distances needs to the initial estimators of $\underline{\mu}$ and Σ . Thus and for that purpose, the MCD method is adopted, by using the following algorithm steps:

(1) Put $k=1$.

(2) Draw a random sample H_k of size $h=p+1$ of different observations.

(3) Calculate the mean vector $\bar{\underline{X}}_k$ and covariance matrix S_k of H_k as:

$$\bar{\underline{X}}_k = \frac{1}{h}\sum_{j \in H_k} \underline{X}_j \quad , \quad S_k = \frac{1}{h}\sum_{j \in H_k} (\underline{X}_j - \bar{\underline{X}}_k)(\underline{X}_j - \bar{\underline{X}}_k)'$$

(4) If $\det(S_k) \le 0$, then go back to step (2).

(5) Calculate Mahalonobis distance $d_k^2(i)$ as:

$$d_k^2(i) = (\underline{X}_i - \bar{\underline{X}}_k)' S_k^{-1} (\underline{X}_i - \bar{\underline{X}}_k) \quad , i = 1,2,\ldots,n$$

(6) Select a set H_{k+1} of size $h = (n+p+1)/2$ observations which have the smallest Mahalonobis distances, and then calculate $\bar{\underline{X}}_{k+1}$ and S_{k+1} as:

$$\bar{\underline{X}}_{k+1} = \frac{1}{h}\sum_{j \in H_{k+1}} \underline{X}_j \quad , \quad S_{k+1} = \frac{1}{h}\sum_{j \in H_{k+1}} (\underline{X}_j - \bar{\underline{X}}_{k+1})(\underline{X}_j - \bar{\underline{X}}_{k+1})'$$

.

(7) If $\det(S_{k+1}) \ne \det(S_k)$, then $\bar{\underline{X}}_k = \bar{\underline{X}}_{k+1}$, $S_k = S_{k+1}$ and go back to step (4), otherwise we get the initial robust estimators for location and scale parameters of multivariate data .

After the calculation of the initial robust estimators, the RMCD approach have been used based on the Hample's weights function to calculate the final robust estimators by applying the following steps:

(1) Calculate the weights w_i for all $i = 1,2,\ldots,n$ as:

$$w_i = w(d_k(i)) = \begin{cases} 1 & \text{if} \quad d_k(i) \le d_0 \\ d_0 \exp[-\frac{1}{2}(d_k(i) - d_0)^2 / b_2^2 \;] & \text{if} \quad d_k(i) > d_0 \end{cases}$$

,

such that $d_0 = \sqrt{p} + b_1 / \sqrt{2}$, $b_1 = 2$ and $b_2 = 1.25$.

(2) Calculate the weighted mean vector $\bar{\underline{X}}_w$ and covariance matrix equation S_w as:

$$\bar{\underline{X}}_w = \frac{\sum_{i=1}^n w_i \underline{X}_i}{\sum_{i=1}^n w_i} \quad , \quad S_w = \frac{1}{\sum_{i=1}^n w_i - \sum_{i=1}^n w_i^2 / \sum_{i=1}^n w_i} \sum_{i=1}^n w_i (\underline{X}_i - \bar{\underline{X}}_w)(\underline{X}_i - \bar{\underline{X}}_w)'$$

(3) If $\left|\bar{\underline{X}}_w - \bar{\underline{X}}_k\right| \le \varepsilon$ to for all $j = 1,2,\ldots,p$ then he stopped, and the only place $\bar{\underline{X}}_k = \bar{\underline{X}}_w$, $S_k = S_w$ and go back to step (1).

4. The Robust Hotelling's Statistic

To derive a robust Hotelling's statistic RT^2 for a sample contains n observations $\{x_1, x_2, ..., x_n\}$ from $N_n(\underline{\mu}, \sigma^2 I)$ where $\underline{\mu} = \underline{1}_n \mu$, the weighted mean and weighted covariance are implemented.

Since $\underline{X} \sim N_n(\underline{\mu}, \sigma^2 I)$, then $\underline{W}' \underline{X} \sim N(\underline{W}'\underline{\mu}, \sigma^2 \underline{W}'\underline{W})$ where $\underline{W}' = (w_1, w_2, \ldots, w_n)$, that is $\sum_{i=1}^n w_i x_i \sim N(\sum_{i=1}^n w_i \mu, \sigma^2 \sum_{i=1}^n w_i^2)$, so the weighed mean is

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \sim N(\mu, k\sigma^2)$$

where,

$$k = \frac{\sum_{i=1}^n w_i^2}{(\sum_{i=1}^n w_i)^2} \quad .$$

then we have,

$$\frac{\bar{x}_w - \mu}{\sqrt{k}\sigma} \sim N(0,1)$$

(1)

Now,

$$\begin{pmatrix} x_1 - \bar{x}_w \\ \vdots \\ x_n - \bar{x}_w \end{pmatrix} =$$

$$\frac{1}{\sum_{i=1}^{n} w_i}\begin{pmatrix} \sum_{i=1}^{n} w_i - w_1 & -w_2 & \cdots & -w_n \\ \vdots & \vdots & \ddots & \vdots \\ -w_1 & -w_2 & \cdots & \sum_{i=1}^{n} w_i - w_n \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and

$$\underline{X}^* = (I - \frac{1}{\sum_{i=1}^{n} w_i}\underline{1}_n \underline{W'})\underline{X}$$

Then,

$$\underline{X}^* \sim N_n(\Sigma\underline{\mu}, \sigma^2 \Sigma\Sigma')$$

where,

$$\Sigma = (I - \frac{1}{\sum_{i=1}^{n} w_i}\underline{1}_n \underline{W'})$$

Also,

$$\sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2 = \underline{X}^{*'} W \underline{X}^* = \underline{X}'(W - \frac{1}{\sum_{i=1}^{n} w_i} W \underline{1}_n \underline{W'})\underline{X}$$

where

$$W = \begin{pmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{pmatrix} = diag(w_i) \quad i = 1,2,\dots,n$$

Here, the matrix $(W - \frac{1}{\sum_{i=1}^{n} w_i} W \underline{1}_n \underline{W'})$ is a symmetric and non-idempotent, but if $w_i = 1$, for all $i = 1,2,\dots,n$, it can be written as:

$$\left(I - \frac{1}{n}\underline{1}\underline{1}'\right)$$, which becomes a symmetric and idempotent (see 2), then

$$\frac{\underline{X}'\left(I - \frac{1}{n}\underline{1}\underline{1}'\right)\underline{X}}{\sigma_2} \sim \chi^2_{(n-1)}.$$

Since, this matrix represents a special case, then

$$\frac{\underline{X}'(W - W\underline{1}_n \underline{W}' / \sum_{i=1}^{n} w_i)\underline{X}}{\sigma^2} = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2}{\sigma^2} \sim \chi^2_{(\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 / \sum_{i=1}^{n} w_i)}$$

(2)

From, (1) , (2) and with independent condition of $\bar{x}_w$ and s_w then(9)

$$RT = \frac{\bar{x}_w - \mu_0}{\sqrt{k}s_w} \sim t_{(\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 / \sum_{i=1}^{n} w_i)},$$

where,

$$s_w^2 = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 / \sum_{i=1}^{n} w_i},$$

Thus,

$$RT^2 = \frac{1}{k}(\bar{x}_w - \mu_0) s_w^{-1} (\bar{x}_w - \mu_0) \sim F_{1,\upsilon}$$

where,

$$\upsilon = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 / \sum_{i=1}^{n} w_i .$$

Generally, for (p≥1) the robust Hotelling's statistic RT^2 can be written as:

$$RT^2 = \frac{1}{k}(\underline{\overline{X}}_w - \underline{\mu}_0)S_w^{-1}(\underline{\overline{X}}_w - \underline{\mu}_0) \sim \frac{p\upsilon}{\upsilon - p + 1}F_{p,\upsilon-p+1}$$

5. Simulation discretisation

In order to make a comparison between a robust and classical statistic, a simulation approach has been adopted through our study. Here, a large number of independent random samples for multivariate distributions has been generated. Additionally, within each distribution, the different relevant cases have been studied with:

(1) The size of sample $n = 10, 20, 30$.

(2) The number of variables $p = 2,3,4$.

(3) The coefficient correlation between variables $\rho = .000001, 0.25, 0.5, 0.75, 0.9$.

The tests of total number in applied in each case 500 experiment , has also been applied all the profiles that we referred to in the following distributions:

(1) Multivariate -normal distribution .
(2) Contaminated multivariate -normal distribution from one side by 20%.
(3) Contaminated multivariate -normal distribution from two sides by 20%.

It was through the application of experimental side this assumed the level of theory incorporeal is $\alpha = .05$ and the purposes of the application of these formations and due to the lack of ready-made programs that can be considered for the application of this aspect of the research has been writing a number of key programs and using the language of subparagraph (Quick Basic) .

6. The conduct of the simulation

Experimental side was applied according to the following:

First: the optimal state of data

We have been through this study application Box-Mueller method as these characterized method of completion speed and accuracy in the results to make them more efficient than others. To generate independent random variables $Z_1, Z_2, ..., Z_p$ have standard normal distribution according to the method Buckeyes must follow the following steps:

(1) Generate two variables U_1 and U_2 such that $U_i \sim U(0,1)$, $i = 1,2$.

(2) Obtain to p of independent variables Z_i where $Z_i \sim N(0,1)$, $i = 1,2,...,p$ by applying the relationships

$$Z_1 = (-2\ln U_1)^{1/2} Sin(2\pi U_2)$$

$$Z_2 = (-2\ln U_1)^{1/2} Cos(2\pi U_2)$$

(3) After obtaining the independent variables $Z_1, Z_2, ..., Z_p$ it can be transferred to the variables $X_1, X_2, ..., X_p$ follow a multivariate -normal average $\underline{\mu}' = (\mu_1, \mu_2, ..., \mu_p)$ and covariance positive definite matrix $\sum_{p \times p}$ using the following relationship:

$$(X_1, X_2, ..., X_p) = (\mu_1, \mu_2, ..., \mu_p) + \Sigma^{1/2}(Z_1, Z_2, ..., Z_p)$$

And so it was getting to watch one of variables $X_1, X_2, ..., X_p$ and to get a sample size *n* of these observations, it is repeated the above number *n* of times.

Second : Get away from the optimal case

This situation has been achieved through the generation of independent random samples for multiple views normal distribution contained outlier values 20% of the sample size.

During this case generate contaminant distribution on the one side to be of type Asymmetric as follows:

(1) Generate a random sample of 80% of the total sample size for multivariate normal-distribution $N_p(\underline{\mu}, \Sigma)$,where

$$\underline{\mu}' = (\mu_1, \mu_2, ..., \mu_p) = (0,0,...,0) \quad ,$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \Lambda & \rho\sigma_1\sigma_p \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \Lambda & \rho\sigma_2\sigma_p \\ M & M & O & M \\ \rho\sigma_1\sigma_p & \rho\sigma_2\sigma_p & \Lambda & \sigma_p^2 \end{pmatrix}$$

Where each is determined Σ depending on the configurations studied.

(2) Generating group outliers views and which constitute 20% of the total sample size to be of normal-distribution parameters Finally, the third case of multiple data pertaining to non-natural similar to its predecessor is that this case is within the variable contamination from two sides by generating a random sample of mixed model (2-10) through the application of the following:

(1) Obtain a random sample of multi-normal distribution accounted 80% of the total sample size

$$\underline{\mu}' = (\mu_1, \mu_2, ..., \mu_p) = (0,0,...,0) \quad , \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots & \rho\sigma_1\sigma_p \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho\sigma_2\sigma_p \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_1\sigma_p & \rho\sigma_2\sigma_p & \cdots & \sigma_p^2 \end{pmatrix}$$

(2) Generate a random sample representing 10% of the total sample size of the community with the distribution of normally $N_p(\underline{\mu} + \underline{\theta}, \Sigma)$ where $\underline{\theta}' = (10, 10, \ldots, 10)$.

(3) generate another random sample, the same percentage, 10% of a community normal distribution $N_p(\underline{\mu} - \underline{\theta}, \Sigma)$.

After obtaining the independent random sample according to each of the four distributions, it is within each variety of formations studied comparison of different methods on the basis of the rates of type I error $\hat{\alpha}$ and the power of the test $\hat{\pi}$ each of them as it is to find $\hat{\alpha}$ the application of the following formula

$$\hat{\alpha} = \frac{L(T)}{R}$$

Where $L(T)$ the number of times statistic rejected T a hypothesis H_0 correct and

R number of times to repeat the experiment

Then determine the appropriate period which is to fall during which the error rates of the way and robust by applying mathematical formula proposed by the Faweett and Salter in 1985 to be in the form $\alpha \pm 2(\alpha(1-\alpha)/R)^{1/2}$

After applying the above formula for $\alpha = .05, R = 500$ result in the following interval (0.03 , 0.07) and that is on the basis of which choose the ways in which her error equations and located within these intervals and therefore the comparison between them on the basis of test force rates of $\hat{\pi}$ each of them and that is to find them using the following formula

$$\hat{\pi} = \frac{K(T)}{R}$$

where $K(T)$ the number of times statistic rejected T a hypothesis H_0 false .

The results:

The tables (1), (2), … , and (12) represent the rates of type I error and power test for the classic and robust Hotelling's statistic.

***:** represent the rates of type I error out of (0.03 , 0.07).

Table(1)

Rates of type I error for classic and robust Hotelling's statistic when the data have multivariate normal distribution, where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ and $\underline{\mu}_0' = (0,0,0,0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.094*	0.044	0.060	0.056	0.050	0.064	0.054
	0.25	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.094*	0.045	0.060	0.056	0.050	0.064	0.054
	0.5	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.094*	0.046	0.060	0.056	0.050	0.064	0.054
	0.75	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.094*	0.044	0.060	0.056	0.050	0.064	0.054
	0.9	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.094*	0.044	0.060	0.056	0.050	0.064	0.054

Table(2)

Rates of power test for classic and robust Hotelling's statistic when the data have multivariate normal distribution, where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ and $\underline{\mu}_0' = (0.5, 0, -0.25, 0)$

Sample size n	10	20	30

The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.192	0.158	0.130	0.436	0.436	0.376	0.636	0.666	0.586
		robust	0.222*	0.192	0.210*	0.438	0.434	0.376	0.642	0.670	0.584
	0.25	classic	0.198	0.224	0.158	0.442	0.560	0.478	0.668	0.783	0.716
		robust	0.226*	0.252	0.238*	0.446	0.564	0.476	0.670	0.790	0.720
	0.5	classic	0.244	0.300	0.206	0.532	0.698	0.668	0.766	0.928	0.910
		robust	0.272*	0.334	0.306*	0.536	0.736	0.668	0.768	0.928	0.912
	0.75	classic	0.398	0.544	0.412	0.798	0.968	0.952	0.938	0.998	0.998
		robust	0.414*	0.570	0.486*	0.802	0.968	0.952	0.940	0.998	0.998
	0.9	classic	0.756	0.936	0.832	0.990	1	1	1	1	1
		robust	0.766*	0.942	0.844*	0.990	1	1	1	1	1

Table(3)

Rates of type I error for classic and robust Hotelling's statistic when the data have multivariate normal distribution, where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9, \sigma_4^2 = 16$ and $\underline{\mu}_0' = (0,0,0,0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.046	0.048	0.038	0.046	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.096*	0.044	0.060	0.058	0.050	0.064	0.054
	0.25	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.096*	0.044	0.060	0.058	0.050	0.064	0.054
	0.5	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.096*	0.044	0.060	0.058	0.050	0.064	0.054
	0.75	classic	0.046	0.048	0.038	0.044	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.096*	0.044	0.060	0.058	0.050	0.064	0.054
	0.9	classic	0.046	0.048	0.038	0.048	0.054	0.054	0.050	0.058	0.054
		robust	0.074*	0.070	0.096*	0.044	0.060	0.058	0.050	0.064	0.054

Table(4)

Rates of power test for classic and robust Hotelling's statistic when the data have multivariate normal distribution, where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9, \sigma_4^2 = 16$ and $\underline{\mu}_0' = (0.5, 0, -0.25, 0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.192	0.126	0.100	0.436	0.346	0.306	0.636	0.538	0.464
		robust	0.222*	0.170	0.182*	0.438	0.350	0.308	0.642	0.548	0.466
	0.25	classic	0.198	0.164	0.124	0.442	0.438	0.364	0.668	0.638	0.552
		robust	0.226*	0.200	0.208*	0.446	0.440	0.366	0.670	0.644	0.560
	0.5	classic	0.244	0.212	0.174	0.532	0.566	0.520	0.766	0.802	0.752
		robust	0.272*	0.240	0.260*	0.536	0.568	0.524	0.768	0.808	0.758
	0.75	classic	0.398	0.386	0.290	0.798	0.866	0.832	0.938	0.972	0.982
		robust	0.414*	0.420	0.368*	0.802	0.868	0.832	0.940	0.972	0.982
	0.9	classic	0.756	0.818	0.706	0.990	1	0.996	1	1	1
		robust	0.766*	0.824	0.728*	0.990	1	0.996	1	1	1

Table(5)

Rates of type I error for classic and robust Hotelling's statistic when the data have contaminated multivariate normal distribution from one side by 20% ,where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ and $\underline{\mu}_0' = (0,0,0,0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.036	0.056	0.052	0.168*	0.138*	0.132*	0.782*	0.438*	0.302*
		robust	0.070	0.046	0.048	0.036	0.066	0.076*	0.046	0.060	0.062
	0.25	classic	0.036	0.058	0.062	0.186*	0.150*	0.128*	0.758*	0.448*	0.298*
		robust	0.070	0.048	0.050	0.038	0.066	0.076*	0.046	0.064	0.066
	0.5	classic	0.034	0.058	0.066	0.180*	0.144*	0.130*	0.734*	0.440*	0.284*
		robust	0.070	0.048	0.062	0.038	0.066	0.078*	0.048	0.058	0.066
	0.75	classic	0.042	0.054	0.060	0.184*	0.142*	0.134*	0.720*	0.424*	0.284*
		robust	0.070	0.050	0.058	0.040	0.070	0.078*	0.052	0.058	0.060
	0.9	classic	0.044	0.050	0.056	0.186*	0.150*	0.134*	0.712*	0.414*	0.282*
		robust	0.070	0.056	0.062	0.038	0.070	0.076*	0.052	0.058	0.068

Table(6)

Rates of power test classic and robust Hotelling's statistic when the data have contaminated multivariate normal distribution from one side by 20% ,where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ and $\underline{\mu}_0' = (0.5,0,-0.25,0)$

The size of sample n	10			20			30		
The number of variables p	2	3	4	2	3	4	2	3	4

The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.114	0.188	0.134	0.372*	0.598*	0.534*	0.758*	0.942*	0.876*
		robust	0.178	0.124	0090	0.338	0.370	0.290*	0.490	0.534	0.442
	0.25	classic	0.136	0.238	0.174	0.434*	0.720*	0.626*	0.822*	0.972*	0.942*
		robust	0.180	0.154	0.118	0.346	0.444	0.374*	0.514	0.668	0.590
	0.5	classic	0.186	0.238	0.248	0.618*	0.858*	0.796*	0.906*	0.990*	0.982*
		robust	0.218	0.212	0.192	0.418	0.584	0.524*	0.644	0.850	0.796
	0.75	classic	0.360	0.572	0.434	0.840*	0.990*	0.984*	0.982*	1*	1*
		robust	0.316	0.384	0.300	0.650	0.880	0.830*	0.878	0.988	0.980
	0.9	classic	0.728	0.946	0.868	0.996*	1*	1*	1*	1*	1*
		robust	0.602	0.810	0.624	0.960	1	1*	1	1	1

Table(7)

Rates of type I error for classic and robust Hotelling's statistic when the data have contaminated multivariate normal distribution from one side by 20% ,where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9, \sigma_4^2 = 16$ and $\underline{\mu}_0' = (0,0,0,0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.034	0.058	0.066	0.184*	0.138*	0.126*	0.730*	0.436*	0.286*
		robust	0.072*	0.048	0.062	0.038	0.068	0.076*	0.048	0.060	0.064
	0.25	classic	0.042	0.050	0.054	0.184*	0.140*	0.140*	0.714*	0.410*	0.268*
		robust	0.072*	0.058	0.050	0.036	0.070	0.076*	0.054	0.060	0.064
	0.5	classic	0.048	0.058	0.048	0.192*	0.140*	0.146*	0.720*	0.418*	0.258*
		robust	0.072*	0.052	0.052	0.036	0.072*	0.078*	0.052	0.062	0.062
	0.75	classic	0.050	0.050	0.038	0.186*	0.158*	0.140*	0.750*	0.422*	0.278*
		robust	0.071*	0.056	0.062	0.036	0.072*	0.078*	0.054	0.060	0.060
	0.9	classic	0.056	0.048	0.044	0.194*	0.166*	0.148*	0.806*	0.426*	0.288*
		robust	0.070	0.050	0.046	0.038	0.070	0.076*	0.050	0.060	0.058

Table(8)

Rates of power test classic and robust Hotelling's statistic when the data have contaminated multivariate normal distribution from one side by 20% ,where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9, \sigma_4^2 = 16$ and $\underline{\mu}_0' = (0.5, 0, -0.25, 0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									

0.05	0.000001	classic	0.052	0.072	0.064	0.174*	0.238	0.200*	0.404*	0.492*	0.406*
		robust	0.178*	0.120	0.088	0.338	0.288	0.242*	0.490	0.428	0.370
	0.25	classic	0.056	0.084	0.062	0.190*	0.214	0.178*	0.408*	0.500*	0.400*
		robust	0.178*	0.142	0.094	0.342	0.348	0.262*	0.516	0.512	0.434
	0.5	classic	0.082	0.074	0.066	0.220*	0.236*	0.194*	0.454*	0.496*	0.396*
		robust	0.214*	0.156	0.116	0.414	0.442*	0.384*	0.640	0.684	0.612
	0.75	classic	0.100	0.082	0.066	0.798*	0.240*	0.222*	0.544*	0.510*	0.472*
		robust	0.306*	0.270	0.224	0.702	0.734*	0.662*	0.876	0.956	0.936
	0.9	classic	0.132	0.110	0.100	0.370*	0.278	0.390*	0.608*	0.732*	0.682*
		robust	0.588	0.608	0.436	0.960	0.976	0.978*	0.998	0.998	0.996

Table(9)

Rates of type I error for classic and robust Hotelling's statistic when the data have contaminated multivariate normal distribution from two sides by 20% ,where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ and $\underline{\mu}_0' = (0,0,0,0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.014*	0.028*	0.018*	0.008*	0.020*	0.032	0.012*	0.032	0.042
		robust	0.034	0.056	0.064	0.042	0.054	0.060	0.040	0.054	0.044
	0.25	classic	0.010*	0.032	0.018*	0.012*	0.018*	0.026*	0.012*	0.034	0.034
		robust	0.046	0.058	0.074*	0.046	0.046	0.050	0.042	0.056	0.042
	0.5	classic	0.010*	0.034	0.026*	0.010*	0.018*	0.026*	0.012*	0.028*	0.032
		robust	0.042	0.062	0.094*	0.046	0.064	0.044	0.040	0.054	0.042
	0.75	classic	0.012*	0.030	0.022*	0.012*	0.018*	0.026*	0.014*	0.024*	0.034
		robust	0.046	0.052	0.088*	0.046	0.064	0.034	0.040	0.052	0.048
	0.9	classic	0.014*	0.038	0.018*	0.014*	0.020*	0.028*	0.016*	0.024*	0.034
		robust	0.054	0.042	0.092*	0.048	0.064	0.042	0.042	0.052	0.040

Table(10)

Rates of power test classic and robust Hotelling's statistic when the data having contaminated multivariate normal distribution from two sides by 20% , where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ and $\underline{\mu}_0' = (0.5, 0, -0.25, 0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									

0.05	0.000001	classic	0.064*	0.110*	0.096*	0.156*	0.354*	0.302	0.270*	0.532	0.516
		robust	0.098	0.138	0.156	0.300	0.366	0.388	0.486	0.534	0.478
	0.25	classic	0.084*	0.166	0.122*	0.202*	0.482*	0.426*	0.372*	0.716	0.656
		robust	0.122	0.240	0.198*	0.338	0.442	0.442	0.532	0.680	0.620
	0.5	classic	0.136*	0.238	0.178*	0.338*	0.692*	0.626*	0.562*	0.902*	0.870
		robust	0.170	0.292	0.300*	0.410	0.636	0.560	0.652	0.858	0.818
	0.75	classic	0.278*	0.492	0.374*	0.700*	0.960*	0.934*	0.878*	0.996*	0.996
		robust	0.328	0.488	0.442	0.648	0.930	0.914	0.886	0.978	0.988
	0.9	classic	0.676*	0.938	0.830*	0.982*	1*	1*	1*	1*	1
		robust	0.662	0.926	0.856*	0.960	1	1	1	1	1

Table(11)

Rates of type I error for classic and robust Hotelling's statistic when the data having contaminated multivariate normal distribution from two sides by 20% ,where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9, \sigma_4^2 = 16$ and $\underline{\mu}_0' = (0,0,0,0)$

The size of sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
0.05	0.000001	classic	0.014*	0.030	0.018*	0.008*	0.020*	0.032	0.012*	0.032	0.042
		robust	0.044	0.060	0.088*	0.048	0.062	0.036	0.040	0.054	0.042
	0.25	classic	0.014*	0.032	0.018*	0.012*	0.018*	0.026*	0.012*	0.034	0.034
		robust	0.052	0.048	0.086*	0.046	0.064	0.044	0.040	0.058	0.038
	0.5	classic	0.010*	0.034	0.026*	0.010*	0.018*	0.026*	0.014*	0.028*	0.034
		robust	0.062	0.056	0.092*	0.046	0.062	0.046	0.042	0.058	0.040
	0.75	classic	0.012*	0.032	0.022*	0.012*	0.018*	0.026*	0.012*	0.024*	0.032
		robust	0.062	0.066	0.076*	0.046	0.060	0.050	0.038	0.058	0.046
	0.9	classic	0.014*	0.038	0.018*	0.014*	0.020*	0.028*	0.016*	0.024*	0.034
		robust	0.068	0.048	0.090*	0.048	0.054	0.060	0.040	0.052	0.046

Table(12)

Rates of power test classic and robust Hotelling's statistic when the data have contaminated multivariate normal distribution from two sides by 20% , where $\underline{\mu}' = (0,0,0,0)$, $\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9, \sigma_4^2 = 16$ and $\underline{\mu}_0' = (0.5, 0, -0.25, 0)$

The size of Sample n			10			20			30		
The number of variables p			2	3	4	2	3	4	2	3	4
The level α	correlation coefficient	Hotelling's statistic									
	0.000001	classic	0.030*	0.050	0.048*	0.060*	0.104*	0.110	0.110*	0.192	0.160

0.05		robust	0.090*	0.078*	0.114*	0.286	0.202	0.150	0.490	0.360	0.266
	0.25	classic	0.038*	0.060	0.056*	0.076*	0.122*	0.118*	0.128*	0.224	0.182
		robust	0.122	0.098*	0.128*	0.308	0.242	0.170	0.530	0.428	0.320
	0.5	classic	0.052*	0.058	0.048*	0.106*	0.128*	0.144*	0.166*	0.276*	0.220
		robust	0.136	0.108*	0.148*	0.358	0.254	0.214	0.644	0.592	0.412
	0.75	classic	0.088*	0.076	0.054*	0.170*	0.182*	0.188*	0.316*	0.330*	0.298
		robust	0.244*	0.162*	0.140*	0.556	0.478	0.306	0.878	0.826	0.614
	0.9	classic	0.136*	0.092	0.098*	0.282*	0.246*	0.324*	0.468*	0.434*	0.552
		robust	0.440*	0.292*	0.216*	0.880	0.644	0.502	0.996	0.904	0.792

Conclusions

There is a close link between the degree of method efficiency and the rate of type I error and the power test of the statistic. Here the results can be summarized as:

(1) If the data have multivariate - normal distribution, then the proposed statistic be very close to classic statistic except in the case of small n and large p.

(2) If the data have multivariate - normal distribution , then the rates of the type I error for classic and robust Hotlling's statistic remain constant after the change of coefficient correlation and the degree of homogenous.

(3) In the case of contaminated multivariate - normal distribution from one side by 20% the best statistic is the robust statistic compared to the classic statistic .

(4) If the data have contaminated multivariate-normal distribution from two sides by 20% , the best statistic is the robust statistic compared to the classic statistic .

References

(1) **Beak R. C. , Cook R. D. (1983)** , "Outliers", Technometrics , Vol.25 , 119-149.

(2) **Hoaglin D. C. , Mosteller F. , Tukey J. W. (1983)**, "Understanding Robust and Exploratory Data Analysis ", John Wiley and Sons, New York .

(3) **Rousseeuw, P. J. (1985)**, "Multivariate Estimation with High Breakdown Point", in: w:

Grossmann, G. Pflug. Vincze, and W. Wertz(Ed), Mathematical Statistics and Applications , Vol.B. Dordrech: Reidel , 283-297.

(4) **Willems G. , Pison , G. Rousseeuw P. J. , & Van Aelst S. (2002)** ,"A robust Hotelling Test", Metrika, 55,125-138 **.**

(5) **Meral C. and Serpil A. (2003)** , " Hotelling's Statistic Based on Minimum Volume Ellipsoid Estimator" , G. U. Journal of Science , 16(4) , 691-695 .

(6) **Abdullah A. A. (2006)** , "The Theory of Linear Models and Multivariate Analysis", John Wiley and Sons, New York.

(7) **Hotelling, H. (1931)** , "The Generalization of Student's Ratio " , Annals of Mathematical Statistics , Vol.2, 360-378 .

(8) **Campbell N. A. (1980)**, "Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation"**,** Appl. Statist. , Vol.29, No.3, pp. 231-237.

(9) **Gabbara S. D. (1994)** , "The Inverse of a Patterned Matrix of a Special Structure ", Basrah Journal of Science, Section A, Vol.12 , 2 ,189-200 .

الخلاصة

إحصاءة هوتيلنك T^2 أداة مهمة لاختبار فرضية معدل مجتمع طبيعي متعدد المتغيرات بمعلمتي موقع وقياس (على فرض أن معلمة القياس غير معلومة وتكون مصفوفة موجبة تماما) . عندما تتضمن البيانات قيماً شاذة ، الأمر الذي يؤثر في التوزيع الاحتمالي لإحصاءة الاختبار . تم اقتراح إحصاءة حصينة لهوتيلنك تعتمد على طريقة محددة مصفوفة التباين المشترك الصغرى الموزونة في إيجاد تقديرات معلمتي الموقع والقياس بدلا من الطريقة الكلاسيكية . قد تم استخدام أسلوب المحاكاة للمقارنة بين الإحصاءة الكلاسيكية و الإحصاءة الحصينة المقترحة ، إذ تم توليد بيانات تتبع التوزيع الطبيعي المتعدد الملوث من جهة واحدة وآخر ملوث من جهتين وقد بينت النتائج إن الإحصاءة الحصينة المقترحة هي الأفضل غالبا مقارنة بالإحصاءة الكلاسيكية وذلك بالاعتماد على نتائج معدلات الخطأ من النوع الأول وقوة الاختبار.