Research Article
Ziraat Mühendiliği (372), 92-100
DOI: 10.33724/zm.948879

Which Test is More Reliable for The Testing Statistical Significance of Canonical Correlation Coefficients?

Kanonik Korelasyon Katsayılarının İstatistiksel Önemliliğini Test Etmek için Hangi Test Daha Güvenilirdir?

Yasin ALTAY[1]
Soner YİĞİT[2]*

[1]Eskisehir Osmangazi University Department of Animal Science, Biometry and Genetics Unit, Eskisehir, Turkey
0000-0003-4049-8301

[2]Canakkale Onsekiz Mart University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, Canakkale, Turkey
0000-0002-2899-2246

***Corresponding author**: soneryigit@comu.edu.tr

Application Date : 7 June 2021
Acceptance Date : 19 June 2021

ABSTRACT

In this study, Wilks' Λ (W), Hotelling-Lawley Trace (H) and Pillai's Trace (P) tests which are used in testing of statistically significance for canonical correlation coefficients were compared in terms of actual type I error rate. As a result of 10000 simulation experiments conducted, when samples were taken from multivariate distributions which are normal and deviate slightly or moderately from normality, the W test was conservative in terms of protecting actual type I error rate in all cases. However, when there is excessively deviate from normality, actual type I error rates for the W test exceeded the upper limit of Bradley's criterion (4.50-5.50%) almost in all cases. On the other hand, the H test and P test generally obtained actual type I error rates which were outside Bradley limits.

Keyword: Wilks' Λ, Hotelling-Lawley Trace, Pillai's Trace, type I error rate, Monte Carlo simulation

ÖZET

Bu çalışmada, kanonik korelasyon katsayılarının istatistiksel olarak önemlilik testinde kullanılan Wilks' Λ (W), Hotelling-Lawley Trace (H) ve Pillai's Trace (P) testleri gerçek tip I hata oranı açısından karşılaştırılmıştır. Yapılan 10000 simülasyon

deneyi sonucunda, normal olan ve normallikten hafif veya orta derecede sapan çok değişkenli dağılımlardan örnekler alındığında, W testi gerçek tip I hata oranını tüm durumlarda koruma açısından muhafazakar olmuştur. Ancak normallikten aşırı derecede sapma olduğunda, W testi için gerçek tip I hata oranları hemen hemen tüm durumlarda Bradley kriterinin üst sınırını (%4,50-5,50) aşmıştır. H testi ve P testi ise genel olarak Bradley sınırlarının dışında kalan gerçek tip I hata oranları elde etmiştir.

Anahtar Kelimeler: Wilks' Λ, Hotelling-Lawley Trace, Pillai's Trace, I. tip hata oranı, Monte Carlo simülasyonu

INTRODUCTION

Canonical correlation analysis is a statistical technique used to examine the linear relationship between two multivariate datasets (Hotelling, 1936; Carroll, 1968; Anderson, 1984; Yanai & Takane, 1992; Ferreira & Purcell, 2009; Andrew *et al.,* 2013). Canonical correlation analysis derives linear combinations between two sets of variables to maximize the correlation coefficient between them (Hotelling, 1951; Gauch & Wentworth, 1976; Baggaley, 1981; Anderson, 1999).

$$U_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{mp}X_p \qquad (1)$$

$$V_m = b_{m1}Y_1 + b_{m2}Y_2 + \dots + b_{mq}Y_q \qquad (2)$$

U_m and V_m are linear combinations of X and Y, respectively in Equation (1) and (2). Those linear combinations are known as new variables or canonical variates (Takane *et al.*, 2006; Tang & Ferreira, 2012). The correlations between corresponding pairs of canonical variates are called canonical correlations, C (Kerlinger & Pedhazur, 1973; Thompson, 1984; Meloun & Militky, 2011). Canonical correlation aims to estimate $a_{11}, a_{12},\dots,a_{1p}$ and $b_{11}, b_{12},\dots, b_{1p}$ such that C is maximum (Stewart & Love, 1968; Sharma, 1996; Van De Velden & Bijmolt, 2006). The first step in evaluating canonical correlations is to determine whether they are statistically significant or not. The null (H_0) and alternative (H_a) hypotheses for assessing the statistical significance of the canonical correlations are:

$$H_0\!: C_{mi} = 0,\ \forall_i \in \{1,\dots,p\},\ p \le q$$

$$H_a\!: \exists_i\!: C_{mi} \ne 0,\ i \in \{1,\dots,p\},\ p \le q$$

The null hypothesis (H_0), which states all the canonical correlations are equal to zero, implies that the correlation among X and Y variables is equal to zero. Rejection of the null hypothesis, that is acceptance of the alternative hypothesis (H_a), means that at least the first canonical correlation coefficient is statistically significant or is not equal to zero (Sharma, 1996). Several test statistics were developed for testing these hypotheses (Knapp, 1978). However, in this study, Wilks' Λ (W), Hotelling-Lawley Trace (H) and Pillai's Trace (P) tests which are the most popular in practice were considered.

The main purpose of this study is to determine the performances of Wilks' Λ (W), Hotelling-Lawley Trace (H) and Pillai's Trace (P) tests under different experimental conditions such as sample size, number of variables, distribution shape and correlation structures.

MATERIAL AND METHODS

In this study, random numbers generated by the Monte Carlo simulation technique were used (Waller, 2016). Random numbers were generated using the monte1 function of the fungible package in the R (R Core Team, 2019). The monte1 function simulates multivariate normal and non-normal data using methods that are developed by Fleishman (1978) and Vale and Maurelli (1983). All experimental situations which were considered in this study are given in Table 1. Type I error rate was used to compare Wilks' Λ (W), Hotelling-Lawley Trace (H) and Pillai's Trace (P) tests in terms of performances. A nominal significance level *(α)* was determined as 5.00% for all experimental cases. Bradley (1978) has reported that the actual type I error rate of a robust test should be between 4.50% and 5.50% when testing at the 5.00% level. In this work, Bradley's conservative criterion was taken into account as a measure of robustness

Table 1. All experimental conditions considered in the study

The correlations of the population (ρ_{XX} and ρ_{YY})	ρ_{XX}=0.3, ρ_{YY}=0.3; ρ_{XX}=0.5, ρ_{YY}=0.5, ρ_{XX}=0.7, ρ_{YY}=0.7, ρ_{XX}=0.3, ρ_{YY}=0.5
Multivariate distribution shapes	N(0,1), t(10), t(5), β(5,10), β(10,5) and $\chi^2(3)$
Number of variables for each dataset (p=q)	2, 4, 6, 8, 10 and 20
n/(p+q)	2, 5, 10 and 20

In order to compare the mentioned tests in terms of their performances, the following steps were followed:

1- The correlations of the population (ρ_{XX} and ρ_{YY}) for both X and Y datasets were determined.
2- n/(p+q) random numbers were generated for each dataset from multivariate distributions that have correlations specified.
3- H_0 hypothesis was tested for all the test statistics.
4- The previous three steps were repeated 10000 times for each experimental condition.
5- Number of H_0 rejected were determined for each test.
6- Actual type I error rate was calculated by dividing the number of H_0 rejected by the number of simulations.

Statistical Significance Tests for the Canonical Correlations

In this section, statistical tests used to evaluate the statistical significance of the null (H_0) and alternative (H_a) hypotheses are introduced.

Wilks' Λ Test Statistic

The statistic was developed by Wilks (1932).

$$\Lambda = \prod_{i=1}^{p} (1 - C_i^2) \quad (3)$$

$$F_W = \left(\frac{1-\Lambda^{1/t}}{\Lambda^{1/t}}\right)\left[\frac{wt-(pq/2)+1}{pq}\right] \quad (4)$$

F_W is approximately F distributed, where $w = N - (p+q+3)/2$, $t = \sqrt{(p^2q^2-4)/(p^2+q^2-5)}$.

The degrees of freedom are pq and $wt - (pq/2) + 1$.

The distribution is exact if min(p,q)≤2 (Rao, 1973).

Hotelling-Lawley Test Statistic

The test statistic was improved by Lawley (1938) and Hotelling (1951).

$$T_{HL} = \sum_{i=1}^{p}\left(\frac{C_i^2}{1-C_i^2}\right) \quad (5)$$

When n>0,

$$F_{HL} = \left(\frac{T_{HL}}{c}\right)\left[\frac{4+(pq+2)/(b-1)}{pq}\right] \quad (6)$$

F_{HL} is approximately F distributed with pq and $4+(pq+2)/(b-1)$ degrees of freedom, where $b = (p+2n)(q+2n)/2(2n+1)(n-1)$, $c = [2+(pq+2)/(b-1)]/2n$, $s = min(p,q)$, $m = (|p-q|-1)/2$ and $n = (N-q-p-2)/2$ (Pillai, 1955).

When n≤0,

$$F_{HL} = \left(\frac{T_{HL}}{c}\right)\left[\frac{4+(pq+2)/(b-1)}{pq}\right] \quad (7)$$

F_{HL} is approximately F distributed with $s(2m+s+1)$ and $2(sn+1)$ degrees of freedom.

Pillai's Trace Test Statistic

This statistic was defined by Pillai (1955).

$$T_P = \sum_{i=1}^{p} C_i^2 \qquad (8)$$

$$F_P = \left(\frac{2n+s+1}{2m+s+1}\right)\left(\frac{T_P}{s-T_P}\right) \qquad (9)$$

F_p is approximately F distributed with $s(2m+s+1)$ and $s(2n+s+1)$ degrees of freedom.

RESULTS

Table 2 shows the actual type I error rates when $\rho_{XX} = \rho_{YY} = 0.3$ which is weak correlations of population. When samples were drawn from a multivariate normal distribution, while Bradley's criterion was met in all conditions for the W test, it could not have been met in more than half of all experimental conditions for the H and P tests. Similar results were obtained when samples were drawn from multivariate t(10), β(5,10) and β(10,5) distributions which deviate slightly or moderately from normality. When samples were taken from multivariate

Table 2. Actual type I error rates when $\rho_{XX} = \rho_{YY} = 0.3$

		N(0,1)			t(10)			t(5)			β(5,10)			β(10,5)			χ^2(3)		
p=q	n/(p+q)	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P
2	2	4.89	**5.72**	**2.25**	5.03	**5.69**	**2.12**	5.25	**6.13**	**2.15**	4.82	**5.63**	**2.35**	5.25	**6.01**	**2.36**	**5.64**	**6.49**	**2.48**
	5	5.27	**5.97**	**4.41**	5.30	**5.76**	4.55	**6.23**	**6.80**	5.42	5.14	**6.05**	4.59	5.08	**5.70**	**4.20**	5.47	**6.14**	4.54
	10	5.18	5.49	4.74	4.90	5.24	4.64	**5.81**	**6.06**	5.48	4.77	5.09	**4.46**	4.93	5.19	4.60	**5.57**	**5.85**	5.15
	20	5.04	5.20	4.88	5.25	5.35	5.05	**5.77**	**5.96**	**5.59**	4.74	4.85	**4.46**	4.95	**5.89**	5.50	**5.65**	**5.82**	5.50
4	2	5.34	**8.22**	**2.40**	5.29	**7.95**	**2.65**	**5.70**	**8.70**	**2.86**	5.16	**7.97**	**2.45**	4.97	**7.66**	**2.23**	**6.26**	**9.24**	**2.82**
	5	4.60	**5.64**	**3.68**	5.07	**6.09**	**3.93**	**5.87**	**6.99**	4.58	5.25	**6.42**	**4.13**	5.20	**6.03**	**4.10**	**6.19**	**7.44**	5.00
	10	5.18	**5.72**	4.68	4.70	5.20	**4.20**	**6.06**	**6.50**	5.49	4.96	5.34	**4.44**	5.23	**5.59**	4.72	**5.52**	**6.12**	4.99
	20	4.98	5.16	4.79	5.17	5.39	4.94	**5.72**	**5.99**	5.50	4.97	5.16	4.85	5.18	5.44	4.97	5.48	**5.61**	5.20
6	2	5.05	**8.26**	**2.26**	5.10	**8.52**	**2.58**	**6.05**	**9.77**	**2.88**	4.95	**8.88**	**2.57**	5.32	**8.57**	**2.61**	**6.12**	**10.27**	**3.10**
	5	4.88	**6.08**	**3.88**	4.76	**5.84**	**3.87**	**6.21**	**7.58**	4.96	4.93	**6.12**	**3.93**	4.63	**5.59**	**3.87**	**6.38**	**7.89**	5.06
	10	4.96	5.50	**4.46**	4.79	5.31	**4.24**	**5.84**	**6.30**	5.19	4.99	5.35	**4.47**	4.90	5.50	**4.43**	5.22	**5.82**	4.68
	20	4.98	5.20	4.76	4.76	4.98	4.52	**5.92**	**6.23**	**5.56**	4.78	4.95	4.60	5.06	5.15	4.82	5.39	**5.62**	5.19
8	2	4.93	**8.87**	**2.62**	5.25	**8.97**	**2.79**	**5.73**	**9.83**	**2.91**	5.11	**8.83**	**2.54**	5.46	**9.14**	**2.70**	**6.69**	**11.13**	**3.33**
	5	5.26	**6.57**	**4.25**	4.96	**5.95**	**3.94**	**6.30**	**7.51**	5.14	5.40	**6.51**	**4.49**	4.85	**5.92**	**3.90**	**6.34**	**7.96**	5.01
	10	5.24	**5.95**	4.76	5.22	**5.83**	4.73	**5.69**	**6.40**	5.12	4.81	5.34	**4.35**	4.94	5.46	**4.49**	**5.82**	**6.39**	5.25
	20	5.13	5.41	4.89	5.43	**5.66**	5.30	5.20	5.42	4.91	4.73	4.99	4.50	4.70	5.14	**4.41**	**5.74**	**5.97**	**5.55**
10	2	4.87	**8.60**	**2.56**	5.19	**9.46**	**2.59**	**6.18**	**10.58**	**3.42**	4.93	**8.61**	**2.62**	4.91	**8.98**	**2.71**	**7.20**	**11.84**	**3.93**
	5	5.03	**6.29**	**3.95**	5.19	**6.31**	**4.30**	**5.71**	**6.96**	4.60	4.99	**6.48**	**3.95**	5.23	**6.16**	**4.21**	**6.02**	**7.32**	4.89
	10	4.93	5.43	4.54	5.22	**5.77**	4.84	**5.65**	**6.21**	5.06	5.10	**5.75**	4.70	5.15	**5.76**	4.62	**5.93**	**6.49**	5.38
	20	4.61	4.80	**4.41**	4.68	4.84	**4.47**	5.39	**5.67**	5.14	5.13	5.40	4.85	4.77	5.02	4.52	5.20	5.45	4.88
20	2	4.91	**9.25**	**2.66**	5.23	**9.42**	**2.89**	**5.95**	**10.30**	**3.40**	5.01	**9.28**	**2.66**	4.73	**8.76**	**2.68**	**7.75**	**13.50**	**4.48**
	5	5.15	**6.95**	4.67	4.92	**5.87**	**3.96**	5.30	**6.64**	**4.27**	5.18	**6.07**	**4.22**	5.38	**6.53**	4.50	**6.63**	**8.07**	5.48
	10	4.98	**5.51**	**4.47**	5.06	**5.54**	4.64	**5.60**	**5.98**	4.94	5.27	**5.84**	4.68	5.02	**5.52**	4.50	**6.22**	**6.79**	**5.61**
	20	5.37	**5.63**	5.23	5.33	5.48	5.21	5.37	**5.69**	5.15	4.86	5.11	4.65	5.05	5.26	4.76	5.39	**5.64**	5.12

t(5), which is a symmetric and heavy-tailed, and multivariate χ^2 (3), which is extremely skewed and heavy-tailed, actual type I error rates for W and H tests were not between 4.5% and 5.5% reported by Bradley in almost all cases. However, under the same conditions, actual type I error rates for the P test fallen into between 4.5% and 5.5% in most cases.

Table 3 shows the actual type I error rates when ρ_{XX} = ρ_{YY} = 0.5 which is moderate correlations of population. When samples were drawn from multivariate N(0,1), t(10), β(5,10) and β(10,5) distributions, regardless of the experimental conditions, actual type I error rates for the W test met Bradley's conservative criterion. However, when samples were drawn from multivariate t(5) and χ^2(3), actual type I error rates for the W test did not meet Bradley's criterion in also all cases. Actual type I error rates for the H test were generally outside Bradley limits.

Table 3. Actual type I error rates when ρ_{XX}= ρ_{YY} = 0.5

		N(0,1)			t(10)			t(5)			β(5,10)			β(10,5)			χ^2(3)		
p=q	n/(p+q)	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P
2	2	5.22	**5.83**	**2.24**	5.11	**5.71**	**2.04**	5.36	**6.11**	**2.28**	5.07	**5.72**	**2.35**	4.92	**5.53**	**1.90**	5.42	**6.22**	**2.22**
	5	4.81	**5.54**	**4.27**	4.88	5.42	**4.09**	**5.85**	**6.48**	4.92	5.16	**5.69**	**4.31**	5.00	5.43	**4.23**	**5.63**	**6.13**	4.72
	10	5.13	5.44	4.76	4.95	5.27	4.57	**6.25**	**6.66**	**5.75**	4.85	5.13	**4.48**	4.85	5.16	**4.49**	**5.65**	**6.03**	5.32
	20	4.81	4.95	4.72	4.95	5.03	4.72	5.25	5.41	5.14	4.90	5.01	4.67	4.92	5.11	4.78	4.98	5.16	4.89
4	2	4.78	**7.40**	**2.25**	5.50	**8.17**	**2.48**	**5.83**	**8.64**	**2.86**	4.94	**7.60**	**2.31**	4.70	**7.08**	**2.39**	**6.45**	**9.56**	**2.90**
	5	5.15	**6.25**	**4.11**	5.31	**6.25**	**4.27**	**5.92**	**7.06**	4.73	4.77	**5.67**	**3.80**	4.90	**5.70**	**3.85**	**6.33**	**7.44**	4.94
	10	5.43	**5.93**	4.86	5.18	**5.59**	4.79	**6.01**	**6.56**	5.48	5.01	5.40	4.54	4.96	5.43	**4.48**	**5.81**	**6.43**	5.12
	20	4.97	5.28	4.74	5.20	5.35	4.99	**6.07**	**6.27**	**5.79**	4.89	5.17	4.64	4.71	4.90	4.51	**6.16**	**6.36**	**5.96**
6	2	4.97	**8.59**	**2.42**	4.99	**8.26**	**2.28**	**5.65**	**9.69**	**2.95**	5.06	**8.07**	**2.55**	5.08	**8.28**	**2.38**	**6.81**	**10.80**	**3.20**
	5	5.10	**6.43**	**4.30**	5.08	**6.14**	**4.21**	**6.26**	**7.71**	5.08	4.92	**6.08**	**3.85**	5.20	**6.09**	**4.09**	**6.45**	**7.96**	5.15
	10	5.11	**5.73**	4.64	5.15	**5.55**	4.64	**5.88**	**6.43**	5.25	5.00	5.42	4.50	5.34	**5.78**	4.78	**5.75**	**6.42**	5.27
	20	4.89	5.11	4.62	4.99	5.29	4.74	**5.66**	**5.85**	5.47	5.17	5.36	4.99	5.04	5.29	4.92	**5.56**	**5.80**	5.29
8	2	5.15	**8.48**	**2.79**	5.42	**9.16**	**2.70**	**6.18**	**10.30**	**2.88**	5.24	**9.15**	**2.69**	4.80	**8.69**	**2.63**	**7.41**	**12.33**	**4.02**
	5	4.97	**6.08**	**3.97**	5.25	**5.64**	**3.67**	**6.36**	**7.92**	5.21	4.82	**5.95**	**3.98**	5.03	**6.00**	**3.98**	**6.94**	**8.54**	**5.53**
	10	5.17	**5.56**	4.64	5.12	**5.64**	4.71	**6.04**	**6.73**	5.48	4.95	**5.62**	4.59	5.02	5.50	**4.48**	**6.34**	**7.04**	**5.76**
	20	4.74	5.02	**4.45**	4.94	5.21	4.72	**5.65**	**5.99**	5.37	4.93	5.21	4.65	4.77	5.11	4.54	**5.73**	**5.97**	5.49
10	2	4.78	**8.51**	**2.36**	5.10	**8.91**	**2.47**	**6.27**	**10.74**	**3.16**	4.97	**8.90**	**2.68**	5.01	**8.65**	**2.81**	**7.03**	**12.17**	**3.67**
	5	5.31	**6.61**	**4.31**	5.11	**6.26**	**4.17**	**6.19**	**7.93**	4.85	5.20	**6.73**	**4.49**	5.03	**5.91**	**4.19**	**7.08**	**8.70**	**5.76**
	10	4.98	**5.52**	**4.49**	5.14	**5.51**	4.67	**6.46**	**7.09**	**5.80**	5.36	**5.82**	4.90	4.79	5.24	**4.32**	**6.28**	**6.83**	**5.68**
	20	4.61	4.81	**4.34**	4.76	4.97	4.58	**5.74**	**6.16**	**5.53**	5.10	5.35	4.81	4.86	5.07	4.58	**6.03**	**6.34**	**5.71**
20	2	4.88	**8.87**	**2.79**	4.91	**8.45**	**2.68**	**6.72**	**11.43**	**3.76**	5.35	**9.32**	**3.06**	5.22	**9.41**	**2.74**	**9.78**	**16.01**	5.35
	5	5.17	**6.25**	**4.28**	5.06	**6.27**	**3.94**	**6.72**	**8.37**	5.46	4.83	**6.03**	**3.99**	4.82	**5.87**	**3.94**	**8.39**	**10.28**	**6.81**
	10	4.51	5.06	**4.21**	5.04	5.48	4.64	**6.49**	**7.17**	**5.84**	4.74	5.25	**4.36**	5.27	**5.72**	4.88	**7.01**	**7.74**	**6.52**
	20	4.71	5.08	4.53	4.96	5.17	4.66	**6.36**	**6.60**	**5.97**	4.99	5.17	4.76	4.83	5.04	4.60	**5.88**	**6.25**	**5.61**

This situation became much clearer when skewness and kurtosis of the distributions increased. Although P test was very successful compared to W and H tests in terms of protecting actual type I error rates between 4.5% and 5.5% when skewness and kurtosis of the distributions (t(5) and $\chi^2(3)$) increased, it was negatively affected by the increase of ρ_{XX} and ρ_{YY}.

Table 4 shows the actual type I error rates when $\rho_{XX} = \rho_{YY} = 0.7$ which is strong correlations of population. Unless samples were taken from multivariate t(5) and $\chi^2(3)$, actual type I error rates for the W test satisfied Bradley's

Table 4. Actual type I error rates when $\rho_{XX} = \rho_{YY} = 0.7$

		N(0,1)			t(10)			t(5)			β(5,10)			β(10,5)			$\chi^2(3)$		
p=q	n/(p+q)	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P
2	2	5.15	**5.98**	**2.01**	5.07	**5.86**	**2.42**	5.20	**5.84**	**1.93**	4.93	**5.74**	**2.14**	5.14	**5.89**	**2.26**	**6.03**	**6.52**	**2.49**
	5	5.12	**5.75**	**4.31**	4.91	5.50	**4.02**	**6.01**	**6.48**	4.84	4.99	**5.52**	**4.28**	4.61	5.20	**3.91**	**5.84**	**6.51**	4.84
	10	4.91	5.23	4.64	5.21	**5.80**	5.15	**5.65**	**5.96**	5.33	4.82	5.08	4.52	5.07	5.46	4.64	5.49	**5.70**	5.20
	20	5.24	5.39	5.12	5.02	5.25	4.82	5.34	5.42	5.14	4.95	5.04	4.88	5.08	5.22	4.87	5.46	**5.62**	5.30
4	2	4.63	**7.20**	**2.13**	4.85	**7.33**	**2.24**	**5.73**	**8.34**	**2.69**	4.89	**7.64**	**2.30**	4.72	**7.26**	**2.20**	**6.73**	**9.82**	**3.09**
	5	5.11	**6.20**	**4.20**	5.02	**6.14**	**3.95**	**5.97**	**7.14**	4.73	4.76	**5.64**	**3.79**	4.50	**5.63**	**3.51**	**6.52**	**7.79**	5.25
	10	4.70	5.21	**4.35**	4.95	5.41	**4.39**	**6.09**	**6.71**	5.49	4.83	5.27	**4.25**	4.59	5.16	**4.09**	**6.09**	**6.57**	5.49
	20	5.17	5.45	4.91	4.99	5.18	4.73	**5.73**	**6.02**	5.44	5.07	5.21	4.87	5.21	5.34	5.02	**6.04**	**6.34**	**5.74**
6	2	4.69	**8.05**	**2.42**	4.93	**8.21**	**2.54**	**5.96**	**9.76**	**3.12**	4.85	**7.66**	**2.39**	5.12	**8.22**	**2.48**	**7.10**	**11.43**	**3.73**
	5	5.25	**6.34**	**4.03**	5.11	**6.74**	**4.47**	**6.72**	**7.90**	5.28	5.33	**6.44**	**4.26**	4.74	**5.82**	**3.67**	**7.52**	**9.07**	**5.83**
	10	5.04	**5.53**	4.50	4.68	5.30	**4.24**	**6.55**	**7.23**	**5.85**	4.75	5.29	**4.46**	4.90	5.37	**4.46**	**6.23**	**6.86**	**5.62**
	20	4.91	5.28	4.77	5.21	5.48	4.93	**6.04**	**6.24**	**5.75**	5.11	5.41	4.82	4.93	5.22	4.60	**5.55**	**5.82**	5.37
8	2	4.84	**8.64**	**2.57**	5.35	**8.88**	**2.72**	**6.36**	**11.15**	**3.45**	5.30	**9.35**	**2.83**	5.00	**8.41**	**2.62**	**8.80**	**13.83**	**4.17**
	5	4.93	**6.06**	**3.86**	5.36	**6.46**	**4.40**	**6.39**	**7.83**	5.15	5.07	**6.29**	**3.97**	4.86	**6.04**	**3.81**	**7.34**	**8.96**	**5.84**
	10	4.86	5.46	**4.35**	5.19	**5.87**	4.67	**6.64**	**7.38**	**5.99**	4.91	5.39	**4.44**	4.71	5.18	**4.16**	**6.65**	**7.24**	**5.99**
	20	4.83	5.12	4.60	5.03	5.34	4.78	**6.33**	**6.63**	**6.05**	4.74	5.04	4.52	4.90	5.17	4.64	**6.36**	**6.72**	**6.06**
10	2	4.75	**9.45**	**2.82**	5.00	**9.04**	**2.57**	**6.67**	**11.58**	**3.72**	5.11	**9.24**	**2.53**	4.53	**8.97**	**2.33**	**8.72**	**14.03**	4.70
	5	4.91	**6.06**	**4.00**	5.27	**6.56**	**4.28**	**6.94**	**8.48**	**5.66**	5.24	**6.67**	4.63	5.24	**6.39**	**4.41**	**7.27**	**9.01**	**5.86**
	10	4.77	5.24	**4.38**	4.95	5.42	**4.49**	**6.96**	**7.67**	**6.49**	4.73	5.23	**4.37**	4.92	5.36	4.52	**7.30**	**8.06**	**6.65**
	20	5.03	5.26	4.81	4.80	5.05	4.60	**6.18**	**6.48**	**5.88**	5.08	5.30	4.86	5.35	**5.83**	5.26	**6.08**	**6.47**	**5.68**
20	2	4.77	**9.06**	**2.73**	5.13	**9.58**	**2.56**	**7.90**	**13.68**	**4.41**	5.20	**9.34**	**2.84**	5.23	**9.04**	**3.03**	**12.52**	**20.53**	**6.82**
	5	5.21	**6.31**	**4.37**	5.16	**6.34**	**4.05**	**8.02**	**9.76**	**6.45**	4.77	**5.96**	**3.81**	5.33	**6.23**	**4.38**	**9.88**	**11.92**	**7.94**
	10	4.88	5.43	4.52	4.60	5.12	**4.12**	**7.30**	**8.14**	**6.60**	5.47	**6.05**	4.93	5.02	5.45	4.54	**7.99**	**8.76**	**7.24**
	20	4.61	4.79	**4.34**	5.16	5.39	4.93	**6.34**	**6.74**	**6.06**	4.77	5.06	4.55	5.26	**5.51**	5.02	**6.75**	**7.02**	**6.46**

conservative criterion in all cases. The H test and P test generally obtained actual type I error rates which were outside Bradley limits. When samples were drawn from multivariate t(5) and χ^2 (3) distributions, the P test was more successful compared to the other tests. However, this success became negligible when $\rho_{XX} = \rho_{YY} = 0.7$. Table 5 shows the actual type I error rates when $\rho_{XX} = 0.3$ and $\rho_{YY} = 0.5$. While actual type I error rates for the W test were between Bradley's limits when samples were taken from multivariate N(0,1), t(10), β(5,10) and β(10,5) in all cases, actual type I error rates for the H and P tests were not generally between Bradley's limits. While the P test was generally between 4.5% and 5.5% when the shape of distributions changed (t(5) and χ^2(3)), W and H tests were generally not between 4.5% and 5.5%. Results obtained under these conditions were similar to those $\rho_{XX} = \rho_{YY} = 0.3$.

Table 5. Actual type I error rates when $\rho_{XX} = 0.3$ and $\rho_{YY} = 0.5$

		N(0,1)			t(10)			t(5)			β(5,10)			β(10,5)			χ^2(3)		
p=q	n/(p+q)	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P	W	H	P
2	2	4.81	**5.52**	**2.13**	5.24	**5.98**	**2.27**	**5.73**	**6.39**	**2.13**	4.79	5.44	**2.09**	5.15	**6.00**	**1.98**	**5.68**	**6.45**	**2.28**
	5	5.24	**5.72**	**4.34**	5.16	**5.71**	**4.21**	**6.07**	**6.71**	5.11	4.98	**5.64**	**4.24**	5.06	**5.74**	**4.24**	**5.58**	**6.22**	4.67
	10	5.02	5.33	4.72	5.28	**5.64**	4.84	**5.59**	**5.98**	5.18	5.03	5.34	4.57	4.98	5.22	4.61	5.01	5.37	4.79
	20	4.93	4.97	4.77	5.43	**5.54**	5.25	**5.54**	**5.65**	5.39	5.30	5.37	5.20	5.32	5.45	5.17	5.41	**5.57**	5.27
4	2	5.20	**7.98**	**2.58**	5.10	**7.48**	**2.37**	**5.66**	**8.76**	**2.59**	5.17	**7.95**	**2.66**	5.16	**7.59**	**2.40**	**6.22**	**8.95**	**2.96**
	5	4.96	**6.10**	**4.21**	4.66	**5.79**	**3.79**	**5.81**	**6.92**	4.68	5.19	**6.13**	**4.14**	4.91	**5.71**	**4.07**	**6.32**	**7.47**	4.84
	10	4.92	5.38	**4.42**	4.81	5.33	**4.32**	**5.96**	**6.49**	5.16	4.57	5.19	**4.20**	4.94	5.39	**4.48**	**5.57**	**6.08**	5.00
	20	4.93	5.15	4.66	4.88	5.08	4.60	**5.70**	**5.97**	5.50	4.81	4.95	4.58	5.44	**5.75**	5.18	5.08	5.36	4.78
6	2	5.25	**8.94**	**2.50**	5.04	**8.45**	**2.51**	**5.68**	**9.45**	**2.81**	4.98	**8.51**	**2.68**	5.13	**8.25**	**2.59**	**6.19**	**10.50**	**3.11**
	5	4.71	**5.88**	**3.60**	5.28	**6.51**	**4.10**	**6.37**	**7.59**	5.17	5.03	**6.06**	**4.10**	5.35	**6.19**	**4.34**	**6.58**	**8.00**	5.06
	10	5.16	**5.62**	4.72	4.98	5.48	4.58	**5.69**	**6.54**	4.96	4.79	5.28	**4.32**	5.28	**5.95**	4.90	**6.22**	**6.92**	**5.61**
	20	4.73	5.02	4.50	5.07	5.34	4.74	**5.69**	**6.04**	5.47	4.92	5.16	4.78	4.81	5.05	4.59	**5.70**	**5.89**	5.38
8	2	5.08	**8.72**	**2.61**	4.80	**8.68**	**2.48**	**6.15**	**10.31**	**3.30**	5.17	**8.75**	**2.50**	5.30	**8.73**	**2.72**	**6.72**	**11.29**	**3.58**
	5	5.18	**6.35**	**4.16**	5.19	**6.25**	**4.04**	**6.17**	**7.62**	5.07	5.18	**6.38**	**4.11**	5.01	**6.06**	**3.99**	**6.66**	**8.45**	5.37
	10	4.93	5.45	**4.44**	4.95	**5.59**	4.58	**6.08**	**6.67**	5.41	5.04	5.39	**4.46**	5.06	**5.53**	4.65	**6.05**	**6.68**	5.35
	20	4.79	5.06	4.64	4.81	5.09	4.61	**5.76**	**6.08**	5.45	5.14	5.43	4.92	4.98	5.19	4.73	5.15	5.39	4.89
10	2	4.95	**9.23**	**2.57**	4.88	**8.80**	**2.34**	**6.50**	**11.07**	**3.48**	5.15	**8.99**	**2.60**	5.13	**9.44**	**2.82**	**7.12**	**12.15**	**3.65**
	5	4.91	**6.05**	**4.09**	4.91	**6.03**	**3.94**	**6.07**	**7.42**	4.89	5.09	**6.16**	**4.03**	5.12	**6.38**	**4.04**	**7.04**	**8.59**	**5.52**
	10	5.25	**5.72**	4.70	4.87	5.33	**4.39**	**6.22**	**6.82**	**5.63**	4.81	5.37	**4.35**	5.15	**5.67**	4.69	**6.12**	**6.80**	5.42
	20	5.11	5.27	4.82	5.10	5.36	4.92	5.26	5.47	5.03	4.95	5.35	4.79	5.41	**5.63**	5.21	5.09	5.42	4.91
20	2	5.07	**9.43**	**2.72**	4.87	**8.71**	**2.61**	**6.51**	**10.92**	**3.62**	4.91	**9.12**	**2.62**	5.28	**9.35**	**3.36**	**8.49**	**14.05**	4.55
	5	4.90	**5.94**	**4.02**	4.99	**6.11**	**4.19**	**6.51**	**7.91**	5.26	4.98	**6.07**	**4.12**	4.89	**6.13**	**3.98**	**7.80**	**9.33**	**6.30**
	10	4.96	5.42	4.53	4.99	5.41	**4.43**	**5.90**	**6.37**	5.42	4.88	5.31	4.54	4.75	5.23	**4.29**	**6.56**	**7.17**	**5.86**
	20	4.72	4.93	**4.45**	4.92	5.30	4.63	5.47	**5.81**	5.23	5.24	5.39	5.02	4.87	5.21	4.62	5.43	**5.75**	5.09

CONCLUSIONS

In this study, Wilks' Λ (W), Hotelling-Lawley Trace (H) and Pillai's Trace (P) tests which are widely used in practice were compared with regards to their performances. When samples were drawn from multivariate distributions which are normal and deviate slightly or moderately from normality, the W test was conservative in all cases. However, when samples were taken from multivariate distributions which excessively deviate from normality, actual type I error rates for the W test exceeded the upper limit of Bradley's criterion almost in all cases. Regardless of distribution shape, sample size, ρ, number of variables for each dataset, actual type I error rates for the H test were above upper limit of Bradley's conservative criterion. P test generally was not more successful compared to the W test. However, when samples were drawn from multivariate distributions which excessively deviate from normality, the P test was more successful compared to the other tests in terms of protecting type I error rate.

In this simulation study, 576 experimental cases were examined for each test. In 411 cases (71.35% of all cases), the W test obtained actual type I error rates which were within Bradley limits. Actual type I error rates which were outside Bradley limits for the W test were obtained in multivariate distributions which excessively deviate from normality. Because the type I error rates for the H test were within the Bradley limits in only 165 (28.65% of all cases) experimental cases, it was generally unsuccessful. This situation was clearer, when samples were drawn from multivariate distributions which excessively deviate from normality. In 245 cases (42.53% of the all cases), the P test were conservative. Most of these conditions were in the multivariate distributions which excessively deviate from normality.

As a result, when samples were taken from multivariate distributions which were normal or slightly or moderately deviated from normality, the W test was certainly robust. The P test was more successful than other tests in multivariate normal distributions that deviated excessively from normality. Regardless of experimental conditions, the H test was not generally robust.

ACKNOWLEDGEMENTS

In this study, the rules of research and publication ethics were followed. Also the authors declares that this study has not been published in any scientific meeting and congress before.

AUTHOR CONTRIBUTION

The design and planning of the study was done by SY and YA. The simulations in the study were made by SY. The writing of the article was carried out by YA and SY. Also, authors have approved the submitted version.

REFERENCES

Anderson, T. W. (1984). An Introduction to Multivariate Statistical Analysis (2nd edition). John Wiley and Sons.

Anderson, T. W. (1999). Asymptotic theory for canonical correlation analysis. Journal of Multivariate Analysis, 70(1), 1-29.

Andrew, G., Arora, R., Bilmes, J. & Livescu, K. (2013). Deep canonical correlation analysis. In International conference on machine learning (pp. 1247-1255).

Baggaley, A. R. (1981). Multivariate analysis: an introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Bradley, J. V. (1978). Robustness?. British Journal of Mathematical and Statistical Psychology, 31(2), 144-152.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. Proceedings of the 76th Annual Convention of the Psychological Association, 3, 227–228.

Ferreira, M. A. & Purcell, S. M. (2009). A multivariate test of association. Bioinformatics, 25(1), 132-133.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. Psychometrika, 43, 521-532.

Gauch, H. G. & Wentworth, T. R. (1976). Canonical correlation analysis as an ordination technique. Vegetatio, 33(1), 17-22.

Hotelling, H. (1936). Relations between two sets of variates. Biometrika, 28(3/4):321–377, 1936.

Hotelling, H. (1951) “A generalized T-test and measure of multivariate dispersion,” in Proceedings of the Second Berkely Symposium on Mathematics and Statistics, pp. 23–41, Berkeley, CA, USA, August 1951.

Kerlinger, F. N. & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York, NY:Holt Rinehart & Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance-testing system. Psychological Bulletin, 85(2), 410.

Lawley, D. N. (1938). A generalization of Fisher’s z test, Biometrika, 30(1-2),180–187.

Meloun, M. & Militky, J. (2011). Statistical data analysis: A practical guide. Woodhead Publishing, Limited.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. The Annals of Mathematical Statistics, 26(1), 117-121.

R Core Team. (2019). R: A language and environment for statistical computing. Ankara, Turkey: R Foundation for Statistical Computing. URL http://www.R-project.org/

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. 2nd ed. New York: John Wiley & Sons.

Sharma, S. (1996). Applied Multivariate Techniques: Canonical Corelation, 391-418. John Willey and Sons Inc., USA.

Stewart, D. & Love, W. (1968). A general canonical correlation index. Psychological bulletin, 70(3p1), 160.

Takane, Y., Yanai, H. & Hwang, H. (2006). An improved method for generalized constrained canonical correlation analysis. Computational statistics & data analysis, 50(1), 221-241.

Tang, C. S. & Ferreira, M. A. (2012). A gene-based test of association using canonical correlation analysis. Bioinformatics, 28(6), 845-850.

Thompson, B. (1984). Canonical correlation analysis uses and interpretations. Newbury Park, CA: Sage.

Vale, D. C. & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. Psychometrika, 48, 465-471

Van De Velden, M. & Bijmolt, T. H. (2006). Generalized canonical correlation analysis of matrices with missing rows: a simulation study. Psychometrika, 71(2), 323-331.

Waller, N. G. (2016). Fungible correlation matrices: A method for generating nonsingular, singular, and improper correlation matrices for Monte Carlo research. Multivariate behavioral research, 51(4), 554-568.

Wilks, S. S. (1932). Certain generalizations made in the analysis of variance, Biometrica, 24(3-4), 471–494.

Yanai, H. & Takane, Y. (1992). Canonical correlation analysis with linear constraints. Linear algebra and its applications, 176, 75-89.