Journal of Statistical Sciences, Spring and Summer, 2023
Vol. 17, No. 1, pp 201-218
DOI: 10.52547/jss.17.1.11

A Permutation Test for Multiple Correlation Coefficient in High Dimensional Normal Data

Najarzadeh, D.[1,2]
[1]Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran.
[2]Research Department of Computational Algorithms and Mathematical Models, University of Tabriz, Tabriz, Iran.
Corresponding author: D. Najarzadeh, d_najarzadeh@tabrizu.ac.ir

Received: 24/2/2023 **Revised:** 23/5/2023 **Accepted and Published Online:** 25/5/2023.

Introduction

The population multiple correlation coefficient (PMCC) measures the correlation between a given one-dimensional random variable X_1 and multi-dimensional random vector $\mathbf{X}_2$. This measure is the maximum correlation between X_1 and any linear combination of components $\mathbf{X}_2$. Testing the null hypothesis of zero PMCC is the most favourable approach to investigate the existence or non-existence of PMCC between X_1 and $\mathbf{X}_2$. The classical procedures for testing this hypothesis in high-dimensional data settings are invalid since the sample covariance matrix inverse or sample precision matrix is undefined. To cope with this problem, a simple test is constructed for testing zero PMCC in high-dimensional normal data. A small simulation study was carried out to evaluate the performance of the proposed test in both high-dimensional and low-dimensional normal data sets. Finally, the proposed test is applied to mice tumour volumes data.

Material and Methods

The proposed test includes the following two steps: In the first step, using a plug-in estimator of the sample precision matrix, a simple test statistic, called plug-in test statistic, was derived which uses the EQUALs method (Wang and Jiang, ۲۰۲۰) to estimate the precision matrix. In the second step, a permutation test is constructed using the proposed plug-in statistic to test the null hypothesis of zero PMCC. Theoretical investigations demonstrate that the proposed testing procedure is a level α test.

Results and Discussion

According to the simulation results, across all configurations of simulation parameters, we observe that the proposed test ϕ_{NRP} and classical exact test ϕ_{low} (in low-dimensional data case) have good control of Type I error, with the false positive rate close to the pre-assigned significance level $\alpha = 0/05$. Across all configurations of the simulation parameters, the power of ϕ_{NRP} is more significant than that for ϕ_{low} in low-dimensional data. For moderately high-dimensional data settings, the power of ϕ_{low} is dramatically smaller than that for ϕ_{NRP}. For high-dimensional data settings, ϕ_{low} becomes practically infeasible due to the singularity of the sample covariance matrix. The empirical powers of ϕ_{NRP} for different covariance structures reveal that the proposed test has a high performance in detecting even small deviations from the null hypothesis.

Conclusion

A simple permutation test for testing zero PMCC in high-dimensional normal data is proposed herein. Some theoretical investigations and simulation studies demonstrate that the proposed test has reasonable control of Type I error apart from how the data dimension p varies with the sample size n. Simulation results showed that the proposed test has higher power than the exact classical test in low-dimensional normal data settings, and it has a high performance in detecting even slight deviations from the null hypothesis in high-dimensional normal data. The proposed method is straightforward to implement and has a broad scope of applications. We finally demonstrated the usefulness of our procedure by applying our methodology to actual tumour volume data and obtaining likely results.

Keywords: Multiple correlation coefficient, High-dimensional normal data, Sparse precision matrix, Plug-in estimator, Permutation test.

Mathematics Subject Classification (2010): 62H15 60F05.

مجله علوم آماری، بهار و تابستان ۱۴۰۲
جلد ۱۷، شماره ۱، ص ۲۰۱ – ۲۱۸
DOI: 10.52547/jss.17.1.11
مقاله پژوهشی

آزمون جایگشتی برای ضریب همبستگی چندگانه در داده‌های نرمال بعد بالا

داریوش نجارزاده[1,2]

[1]گروه آمار، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تبریز
[2]گروه پژوهشی الگوریتم‌های محاسباتی و مدل‌های ریاضی، دانشگاه تبریز

نویسنده مسئول : داریوش نجارزاده، d_najarzadeh@tabrizu.ac.ir

تاریخ دریافت: ۱۴۰۱/۱۲/۵ تاریخ بازنگری: ۱۴۰۲/۳/۲ تاریخ پذیرش و انتشار: ۱۴۰۲/۳/۵

چکیده: در تحلیل رگرسیون چندگانه، ضریب همبستگی چندگانه جامعه به شکل گسترده‌ای برای اندازه‌گیری میزان همبستگی بین یک متغیر با یک مجموعه از متغیرهای تصادفی به کار می‌رود. به منظور ارزیابی وجود یا عدم وجود این نوع از همبستگی، آزمون صفر بودن مورد استفاده است. در داده‌های بعد بالا، به دلیل تکین بودن ماتریس کواریانس نمونه، روش‌های کلاسیک موجود برای آزمون این فرض همگی غیر قابل استفاده هستند. در این مقاله، به منظور آزمون صفر بودن این ضریب، آماره آزمونی بر پایه برآوردگر جایگذاری وارون ماتریس کوواریانس نمونه معرفی و سپس به کمک آماره آزمون پیشنهادی، یک آزمون جایگشتی پیشنهاد شده است. مطالعه شبیه‌سازی برای ارزیابی آزمون پیشنهادی هم در داده‌های بالا و هم در داده‌های بعد پایین انجام شده است. در نهایت، کاربردی از آزمون پیشنهادی بر روی داده‌های اندازه‌های تومور موش‌ها ارائه شده است.

واژه‌های کلیدی: ضریب همبستگی چندگانه، داده‌های نرمال بعد بالا، ماتریس کوواریانس تکین، برآوردگر جایگذاری، آزمون جایگشتی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62H15، 60F05.

۱ مقدمه

ضریب همبستگی چندگانه جامعه (PMCC) و معادل نمونه‌ای آن (SMCC) به عنوان ملاکی برای سنجش همبستگی یک متغیر تصادفی تک بعدی X_1 و بردار تصادفی $(p-1)$-بعدی $\boldsymbol{X}_2$ شناخته می‌شود. در اینجا PMCC با $0 \leq \rho \leq 1$ و SMCC با $0 \leq R \leq 1$ تعریف شده است. این معیار به صورت ماکسیمم همبستگی خطی بین X_1 و هر ترکیب خطی از مؤلفه‌های $\boldsymbol{X}_2$ تعریف می‌شود (میورهد، ۲۰۰۵).

مسئله آزمون فرض $H_0 : \rho = 0$ در مقابل $H_1 : \rho > 0$ به عنوان روش شناخته شده برای بررسی وجود یا عدم وجود ضریب همبستگی چندگانه بین X_1 و $\boldsymbol{X}_2$ است. در ادامه، در چهارچوب داده‌های کلاسیک بعد پایین، به برخی از مهم‌ترین پژوهش‌های انجام شده مربوط به این مسئله آزمون اشاره شده است. آزمون نسبت درست‌نمایی برای فرض صفر توسط گوپتا (۱۹۹۷) ارائه شد. آزمون صفر بودن ضریب همبستگی چندگانه در داده‌های چندمتغیره نرمال ناقص[1] به وسیله پرووس (۱۹۸۷) مورد بررسی قرارگرفت. به منظور مطالعه بیشتر در مورد آزمون‌های کلاسیک برای این فرض به کتاب میورهد (۲۰۰۵) مراجعه گردد. تحت مدل مؤلفه‌های مستقل[2] در داده‌های غیر نرمال و وقتی نسبت بعد داده‌ها p به اندازه نمونه n کوچکتر از یک و مقداری نزدیک به آن است، بازه اطمینانی برای PMCC توسط ژنگ و همکاران (۲۰۱۴) معرفی شد. با استفاده از ایده کاهش بعد تصویرسازی تصادفی، آزمونی برای ضریب همبستگی چندگانه در داده‌های بعد بالا توسط نجارزاده (۲۰۲۰) پیشنهاد شد. تعمیمی از این آزمون بر مبنای انتخاب بهینه بعد داده‌های تصویرشده توسط نجارزاده (۲۰۲۲) ارائه شد.

در این مقاله، آزمونی ساده برای فرض صفر بودن ضریب همبستگی چندگانه جامعه در چهارچوب داده‌های نرمال بعد بالا در قالب گام‌های زیر پیشنهاد شده است. در گام اول، با استفاده از برآوردگر جایگذاری[3] از ماتریس دقت نمونه‌ای، آماره آزمون ساده که آنرا آماره آزمون جایگذاری می‌نامیم، معرفی می‌شود. در داده‌های بعد بالا، که امکان محاسبه ماتریس دقت نمونه‌ای ممکن نیست، مسئله برآورد ماتریس دقت جامعه کاری بسیار چالش برانگیز است که بحث در مورد جزئیات این کار از چهارچوب مبحث این مقاله خارج است. به طور خلاصه مهمترین روش‌های برآورد ماتریس دقت در داده‌های بعد بالا عبارتند از CLIME (کای و همکاران، ۲۰۱۱)، glasso (فریدمن و همکاران، ۲۰۰۸)، BigQuic (شه و همکاران، ۲۰۱۳)، SCIO (لیو و لوو، ۲۰۱۵)، و EQUALs (وانگ و جیانگ، ۲۰۲۰). اخیراً وانگ و جیانگ (۲۰۲۰) نشان دادند که روش EQUALs عملکرد بهتری نسبت به سایر روش‌ها به لحاظ کارایی محاسباتی و دقت برآورد دارد. از این رو در این مقاله، از این روش به منظور برآورد ماتریس دقت در داده‌های بالا استفاده شده است. در گام دوم به کمک آماره آزمون جایگذاری پیشنهادی، آزمون جایگشتی برای فرض صفر ارائه می‌شود. به کمک بررسی نظری، در این مقاله ثابت می‌شود که آزمون به دست آمده یک آزمون سطح α است. در بخش ۲.۲ جزئیات بیشتر این آزمون ارائه خواهد شد.

ضریب همبستگی چندگانه و همچنین آزمون جایگشتی برای فرض صفر بودن این ضریب در بخش ۲ ارائه خواهد

[1] Incomplete multivariate normal data
[2] Independent component model
[3] Plug-in estimator

شد. در بخش ۳، به کمک داده‌های شبیه‌سازی شده، عملکرد آزمون پیشنهادی در مجموعه داده‌های بعد بالا و پایین مورد ارزیابی قرار خواهد گرفت. کاربردی از آزمون پیشنهادی بر مجموعه داده‌های واقعی اندازه‌های تومور موش‌ها در بخش ۴ ارائه خواهد شد. در نهایت، بحث و نتیجه‌گیری در مورد روش پیشنهادی در بخش ۵ معرفی می‌شود.

۲ مبانی نظری

در این بخش، ابتدا ضریب همبستگی چندگانه و آزمون کلاسیک صفر بودن این ضریب به طور مختصر معرفی می‌شود. سپس، در چهارچوب داده‌های بعد بالا، آزمون جایگشتی پیشنهادی ارائه می‌شود.

۲.۱ ضریب همبستگی چندگانه

نمونه تصادفی $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ از یک توزیع p-متغیره با ماتریس کوواریانس ناتکین $\boldsymbol{\Sigma}_X$ را در نظر بگیرید و فرض کنید $\hat{\boldsymbol{\Sigma}}_X$ ماتریس کوواریانس نمونه‌ای حاصل از این نمونه است. فرض کنید

$$\boldsymbol{X}_i = \left[X_{1i} \mid \boldsymbol{X}_{2i}^T\right]^T, \quad i = 1, \ldots, n,$$

افرازی از $\boldsymbol{X}_i = [X_{1i} \ \ldots \ X_{pi}]^T$ باشد. با در نظر گرفتن این افراز، افرازی متناظر بر ماتریس کوواریانس جامعه $\boldsymbol{\Sigma}_X$ و ماتریس کوواریانس نمونه‌ای $\hat{\boldsymbol{\Sigma}}_X$ به ترتیب به صورت

$$\hat{\boldsymbol{\Sigma}}_X = \left[\begin{array}{c|c} \hat{\sigma}_{11} & \hat{\boldsymbol{\sigma}}_{12}^T \\ \hline \hat{\boldsymbol{\sigma}}_{12} & \hat{\boldsymbol{\Sigma}}_{22} \end{array}\right] \quad \text{و} \quad \boldsymbol{\Sigma}_X = \left[\begin{array}{c|c} \sigma_{11} & {\boldsymbol{\sigma}_{12}}^T \\ \hline \boldsymbol{\sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{array}\right] \tag{1}$$

اعمال می‌شود، که در آن $\hat{\sigma}_{11}$، $\hat{\boldsymbol{\sigma}}_{12}$ و $\hat{\boldsymbol{\Sigma}}_{22}$ برآوردهای نمونه‌ای به ترتیب از $\sigma_{11} = \mathrm{Var}(X_{1i})$، $\boldsymbol{\sigma}_{12} = \mathrm{Cov}(\boldsymbol{X}_{2i}, X_{1i})$ و $\boldsymbol{\Sigma}_{22} = \mathrm{Cov}(\boldsymbol{X}_{2i})$ هستند. با این توضیحات، توان دوم PMCC و SMCC بین اولین مؤلفه از $\boldsymbol{X}_i$ (X_{1i}) و $(p-1)$ مؤلفه دیگر $\boldsymbol{X}_i$ $(\boldsymbol{X}_{2i})$ به ترتیب به صورت

$$R^2 = \frac{\hat{\boldsymbol{\sigma}}_{12}^T \hat{\boldsymbol{\Sigma}}_{22}^{-1} \hat{\boldsymbol{\sigma}}_{12}}{\hat{\sigma}_{11}} \quad \text{و} \quad \rho^2 = \frac{\boldsymbol{\sigma}_{12}^T \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\sigma}_{12}}{\sigma_{11}} \tag{2}$$

تعریف می‌شوند (میورهد، ۲۰۰۵). ذکر این نکته در اینجا ضروری است که در داده‌های با ابعاد بالا، به دلیل کم-رتبه بودن[1] ماتریس کوواریانس نمونه، $\hat{\boldsymbol{\Sigma}}_{22}$ تکین خواهد بود. به این دلیل، R^2 در (۲) به عنوان آماره آزمون برای فرض صفر بودن PMCC غیر قابل محاسبه خواهد بود.

[1] Rank deficiency

در ادامه، آزمون فرض صفر بودن ضریب همبستگی چندگانه در داده‌های بعد پایین معرفی می‌شود. فرض کنید $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$ یک نمونه تصادفی از توزیع نرمال p-متغیره $(p < n)$ با بردار میانگین $\boldsymbol{\mu}_X$ و ماتریس کوواریانس $\boldsymbol{\Sigma}_X$، $N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$، است. میورهد (۲۰۰۵) ثابت کرده است که آماره

$$T_{low} = T_{low}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n) := \frac{R^2}{1 - R^2} \frac{n-p}{p-1}$$

دارای توزیع F غیرمرکزی $F_{p-1,n-p}(\delta_X)$ با درجات آزادی $p-1$ و $n-p$ و پارامتر غیرمرکزی

$$\delta_X = \frac{n\boldsymbol{\sigma}_{12}^T \boldsymbol{\Sigma}_{22}^{-1} \hat{\boldsymbol{\Sigma}}_{22} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\sigma}_{12}}{\sigma_{11} - \boldsymbol{\sigma}_{12}^T \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\sigma}_{12}}$$

است. تحت فرض صفر $H_0 : \rho = 0$ همواره $\boldsymbol{\sigma}_{12}$ برابر بردار صفر $\boldsymbol{\sigma}_{12} = 0_{p-1}$ است که در نتیجه آن $\delta_X = 0$ خواهد بود. پس، تحت فرض صفر توزیع F آماره T_{low} یک توزیع مرکزی خواهد بود. از این رو، بر اساس T_{low}، تابع آزمون دقیق سطح α برای آزمودن $H_0 : \rho = 0$ در برابر $H_1 : \rho > 0$ به صورت

$$\phi_{low}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n) = I(T_{low} \geq F_{p-1,n-p;1-\alpha}), \tag{۳}$$

خواهد بود، که در آن $I(.)$ تابع نشانگر و $F_{p-1,n-p;1-\alpha}$ صدک $100(1-\alpha)$ام توزیع مرکزی F با درجات آزادی $p-1$ و $n-p$ است.

۲.۲ آزمون جایگشتی

در بخش ۲.۱ ملاحظه شد آزمون دقیق (۳) برای فرض صفر بودن PMCC در چهارچوب داده‌های نرمال بعد پایین وجود دارد. همانطور که گفته شد، در داده‌های بعد بالا به دلیل تکین بودن $\hat{\boldsymbol{\Sigma}}_{22}$، آماره آزمون T_{low} غیر قابل استفاده است و در نتیجه $\phi(T_{low})$ نمی‌تواند به عنوان یک آزمون برای H_0 استفاده شود. برای حل این مشکل، پیشنهاد می‌شود از آماره آزمون جایگذاری حاصل از جایگذاری برآوردی تنک[1] از ماتریس دقت $\boldsymbol{\Sigma}_{22}^{-1}$ (پوراحمدی، ۲۰۱۳) به جای $\hat{\boldsymbol{\Sigma}}_{22}^{-1}$ در (۲) به شرح زیر استفاده شود. فرض کنید

$$\boldsymbol{X}_i = [X_{1i} \ \dots \ X_{pi}]^T = \left[X_{1i} \mid \boldsymbol{X}_{2i}^T\right]^T, i = 1, \dots, n, \tag{۴}$$

یک نمونه تصادفی از $N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ است. تعریف می‌کنیم $\boldsymbol{X}^{(1)} = [X_{11}, \dots, X_{1n}]$ و $\boldsymbol{X}^{(2)} = [\boldsymbol{X}_{21}, \dots, \boldsymbol{X}_{2n}]$. گیریم $\hat{\boldsymbol{\Sigma}}_{EQUALs}^{-1}$ برآوردی از $\boldsymbol{\Sigma}_{22}^{-1}$ به دست آمده با روش برآورد ماتریس دقت

[1] Sparse estimate

تنک EQUALs ارائه شده توسط وانگ و جیانگ (۲۰۲۰) است. دلیل اینکه در این بخش از روش EQUALs به منظور برآورد $\boldsymbol{\Sigma}_{۲۲}^{-۱}$ استفاده شده این است که بر اساس نتایج گزارش شده توسط وانگ و جیانگ (۲۰۲۰)، روش EQUALs دارای خطای برآورد کمتر و سرعت محاسباتی به مراتب بالاتر نسبت به روش‌های مشابه جدید مخصوصاً در حالت p بزرگ دارد. پس از استفاده از این برآوردگر، آماره آزمون جایگذاری با استفاده از $\hat{\boldsymbol{\Sigma}}_{EQUALs}^{-۱}$ به جای $\hat{\boldsymbol{\Sigma}}_{۲۲}$ به شرح

$$T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = \frac{\hat{\boldsymbol{\sigma}}_{۱۲}^{T} \hat{\boldsymbol{\Sigma}}_{EQUALs}^{-۱} \hat{\boldsymbol{\sigma}}_{۱۲}}{\hat{\sigma}_{۱۱}}, \tag{۵}$$

است، که در آن $\hat{\sigma}_{۱۱} = \frac{۱}{n-۱}\boldsymbol{X}^{(۱)}(\boldsymbol{I}_n - \frac{۱}{n}\boldsymbol{J}_n)\boldsymbol{X}^{(۱)T}$ و $\hat{\boldsymbol{\sigma}}_{۱۲} = \frac{۱}{n-۱}\boldsymbol{X}^{(۲)}(\boldsymbol{I}_n - \frac{۱}{n}\boldsymbol{J}_n)\boldsymbol{X}^{(۱)T}$ و $\boldsymbol{I}_n$ و $\boldsymbol{J}_n$ به ترتیب ماتریس قطری همانی و ماتریس با درایه‌های یک هستند. برای ساخت آزمون جایگشتی برای آزمودن فرض $H_۰$ بر اساس T، در حالی که ستون‌های مشاهدات $\boldsymbol{X}^{(۲)}$ ثابت نگه داشته شده است، جایگشت‌هایی از ستون‌های $\boldsymbol{X}^{(۱)}$ تولید و به ازای هر جایگشت مقدار آماره (۵) محاسبه می‌شود. برای این کار فرض کنید

$$\boldsymbol{\pi}_j = (\pi_۱^{(j)}, \ldots, \pi_n^{(j)}), \quad j = ۱, \ldots, n!$$

تمامی جایگشت‌های ممکن از $(۱, \ldots, n)$ هستند. آنگاه ستون‌های

$$\boldsymbol{\pi}_j(\boldsymbol{X}^{(۱)}) := [X_{۱\pi_۱^{(j)}}, \ldots, X_{۱\pi_n^{(j)}}], \quad j = ۱, \ldots, n!$$

جایگشت‌های مختلفی از ستون‌های $\boldsymbol{X}^{(۱)} = [X_{۱۱}, \ldots, X_{۱n}]$ هستند. بنابراین، آماره آزمون جایگذاری برای jامین جایگشت به صورت

$$T_j(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) \equiv T(\boldsymbol{\pi}_j(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) = \frac{\hat{\boldsymbol{\sigma}}_{۱۲}^{(j)T} \hat{\boldsymbol{\Sigma}}_{EQUALs}^{-۱} \hat{\boldsymbol{\sigma}}_{۱۲}^{(j)}}{\hat{\sigma}_{۱۱}}, \tag{۶}$$

خواهد بود که در آن $\hat{\boldsymbol{\sigma}}_{۱۲}^{(j)}$ معادل $\hat{\boldsymbol{\sigma}}_{۱۲}$ بر اساس $\boldsymbol{\pi}_j(\boldsymbol{X}^{(۱)})$ است. توجه شود که جابجایی ستون‌های $\boldsymbol{X}^{(۱)}$ تاثیری در محاسبه $\hat{\boldsymbol{\Sigma}}_{EQUALs}^{-۱}$ ندارد.

قضیه ۱. با فرض نرمال بودن داده‌ها، تحت فرض $H_۰$، آزمون جایگشتی دقیق سطح α برای آزمون $H_۰$ در برابر $H_۱$ وجود دارد.

برهان: گیریم $\boldsymbol{P}_n = \{\boldsymbol{\pi}_۱, \ldots, \boldsymbol{\pi}_{n!}\}$ مجموعه تمامی جایگشت‌های $\boldsymbol{\pi}$ از $(۱, \ldots, n)$ هستند. مقادیر $T(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)})$ در (۶) به ازای تمامی $\boldsymbol{\pi} \in \boldsymbol{P}_n$ را محاسبه و آماره‌های ترتیبی نظیر آنها را با $T_{(۱)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) \leq \ldots \leq T_{(n!)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$ نشان می‌دهیم. حال به ازای یک مقدار معلوم $۰ <$

$\alpha < ۱$، قرار می‌دهیم $k = n! - [\alpha n!]$ که در آن $[\alpha n!]$ بزرگترین عدد صحیح کوچکتر مساوی $\alpha n!$ است. گیریم

$$M^+(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = \sum_{\boldsymbol{\pi} \in \boldsymbol{P}_n} I\Big(T(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) > T_{(k)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})\Big),$$
$$M^{\circ}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = \sum_{\boldsymbol{\pi} \in \boldsymbol{P}_n} I\Big(T(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) = T_{(k)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})\Big).$$

فرض کنید

$$\eta(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = \frac{\alpha n! - M^+(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})}{M^{\circ}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})}.$$

حال آزمون تصادفی را به صورت

$$\phi(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = \begin{cases} ۱ & T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) > T_{(k)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}), \\ \eta(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) & T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = T_{(k)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}), \\ \circ & T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) < T_{(k)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}). \end{cases} \qquad (۷)$$

تعریف می‌کنیم. توجه شود که به ازای هر $\boldsymbol{\pi} \in \boldsymbol{P}_n$، به شرط $(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$،

$$T_{(j)}(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) = T_{(j)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}), \quad j = ۱, ۲, \ldots, n!.$$

علاوه بر این، $\boldsymbol{P}_n$ یک مجموعه بسته تحت ترکیب است، یعنی، اگر $\boldsymbol{\pi}_۱, \boldsymbol{\pi}_۲ \in \boldsymbol{P}_n$ آنگاه ترکیب $\boldsymbol{\pi}_۱ o \boldsymbol{\pi}_۲$ نیز عضوی از $\boldsymbol{P}_n$ است. بنابراین،

$$M^+(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) = M^{\circ}(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) = M^+(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}).$$

در نتیجه $\eta(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) = \eta(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$. بنابراین به شرط $(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$، داریم:

$$\sum_{\pi \in P_n} \phi(\pi(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) = M^+(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) + \eta(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) M^{\circ}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$$
$$= \alpha n!. \qquad (۸)$$

حال از آنجا که به شرط $(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$، مؤلفه‌های $(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)})$ و $(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$ دارای توزیع یکسانی

هستند، با گرفتن امید ریاضی از طرفین رابطه (۸) نسبت به توزیع $(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$، داریم

$$E\big[\sum_{\pi\in P_n}\phi(\pi(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)})\big] = \sum_{\pi\in P_n} E\big[\phi(\pi(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)})\big] = \alpha\, n!.$$

بنابراین، $E\left[\phi(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})\right] = \alpha$. به عبارت دیگر، آزمون جایگشتی (۷) دارای اندازه آزمون دقیق α است. توجه شود که تابع آزمون ϕ در معادله (۷) متفاوت از یک آزمون معمولی است چرا که مقدار بحرانی $T_{(k)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$ در (۷) خود یک متغیر تصادفی است (هوفدینگ، ۱۹۵۲). علاوه بر این، اگر $\alpha n!$ یک عدد صحیح نباشد، تابع آزمون ϕ احتمالا یک آزمون تصادفی است. در مقابل، اگر نخواهیم که یک آزمون تصادفی داشته باشیم، همانگونه که در اینجا نیز چنین است، یک آزمون جایگشتی غیر تصادفی[۱] (NRP) و البته اندکی محافظه‌کار با ناحیه رد $T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) > T_{(k)}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$ یک آزمون سطح α خواهد بود. به طور معادل، آزمون NRP فرض $H_۰$ را رد می‌کند هرگاه پی-مقدار

$$\hat{p}_{NRP}(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = \frac{۱}{n!}\big[۱ + \sum_{\boldsymbol{\pi}\in P_n\setminus\{(۱,\ldots,n)\}} I(T(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) \geq T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}))\big] \qquad (۹)$$

کوچکتر از α باشد. همانطور که رومانو و وولف (۲۰۰۵) نشان داده است، تحت فرض صفر و فرض نرمال بودن، به ازای هر $۰ \leq u \leq ۱$، با استفاده از تبدیل انتگرالی احتمال داریم:

$$P(\hat{p}_{NRP} \leq u) = P(F_{\boldsymbol{\pi}}(T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})) \geq ۱ - u) = u,$$

که در آن متغیر تصادفی یکنواخت $F_{\boldsymbol{\pi}}(T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}))$ تابع توزیع متغیر $T(\boldsymbol{\pi}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)})$ در $T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)})$ است. بنابراین، آزمون جایگشتی غیر تصادفی با ناحیه رد $\hat{p}_{NRP} \leq \alpha$ یک آزمون سطح α است. از آنجا که $\boldsymbol{P}_n = \{\boldsymbol{\pi}_۱, \ldots, \boldsymbol{\pi}_{n!}\}$ ممکن است مجموعه‌ای بزرگ باشد، که معمولاً نیز چنین است، می‌توان به یک تقریب تصادفی به منظور ساخت یک آزمون تصادفی بر اساس نمونه‌های با جایگذاری یا بدون جایگذاری π از $\boldsymbol{P}_n$ متوسل شد. برای این کار، فرض کنید $\boldsymbol{\pi}_{(۱)}, \ldots, \boldsymbol{\pi}_{(R-۱)}$ با $R \leq n!$ نمونه‌های تصادفی با جایگذاری از $\boldsymbol{P}_n$ باشند. در این حالت، تقریبی از پی-مقدار (۹) آزمون NRP به صورت

$$\tilde{p}_{NRP}(R; \boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}) = \frac{۱}{R}\big[۱ + \sum_{j=۱}^{R-۱} I(T(\boldsymbol{\pi}_{(j)}(\boldsymbol{X}^{(۱)}), \boldsymbol{X}^{(۲)}) \geq T(\boldsymbol{X}^{(۱)}, \boldsymbol{X}^{(۲)}))\big] \qquad (۱۰)$$

[1] Non-randomized permutation test

خواهد بود. بنابراین، تابع آزمون NRP به صورت

$$\phi_{NRP}(R;\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}) = I(\tilde{p}_{NRP}(R;\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}) \leq \alpha) \tag{11}$$

حاصل می‌شود. تحت فرض‌های صفر و نرمال بودن داده‌ها، لم یک از رومانو و وولف (۲۰۰۵) نتیجه می‌دهد ϕ_{NRP} در (۱۱) یک آزمون سطح α به ازای هر انتخاب از R است. به هر حال، برای مقدار بزرگ از R، کمیت‌های $\hat{p}_{NRP}$ و $\tilde{p}_{NRP}$ به یکدیگر نزدیک هستند. به بیان دقیق‌تر، به شرط $(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$، بنابه هم‌توزیعی و استقلال $(\boldsymbol{\pi}_{(i)}(\boldsymbol{X}^{(1)}),\boldsymbol{X}^{(2)})$ و $(\boldsymbol{\pi}_{(j)}(\boldsymbol{X}^{(1)}),\boldsymbol{X}^{(2)})$ برای $i \neq j$ دلخواه، داریم

$$E_{\boldsymbol{\pi}}[\tilde{p}_{NRP}(R;\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})] = \hat{p}_{NRP}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$$

و وقتی $R \to n!$،

$$Var_{\boldsymbol{\pi}}[\tilde{p}_{NRP}(R;\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})] = \frac{\hat{p}_{NRP}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})(1-\hat{p}_{NRP}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)}))}{R} \to 0.$$

بنابراین، به شرط $(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$، مقدار $\tilde{p}_{NRP}(R;\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$ در احتمال به $\hat{p}_{NRP}(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$ همگرا خواهد بود وقتی که $R \to n!$.

الگوریتم ۱. NRP الگوریتم اجرای آزمون:

ورودی‌های الگوریتم: $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$، R، α

گام ۱: قرار دهید $\boldsymbol{X}^{(1)} = [X_{11},\ldots,X_{1n}]$ و $\boldsymbol{X}^{(2)} = [\boldsymbol{X}_{21},\ldots,\boldsymbol{X}_{2n}]$.

گام ۲: $T = T(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$ را از رابطه (۵) محاسبه کنید.

گام ۳: برای $\boldsymbol{X}^{(2)}$ ثابت، جایگشت تصادفی $\boldsymbol{X}^{(1)}_{per}$ از ستون‌های $\boldsymbol{X}^{(1)}$ را بدست آورید.

گام ۴: مقدار $T(\boldsymbol{X}^{(1)}_{per},\boldsymbol{X}^{(2)})$ را به وسیله رابطه (۵) محاسبه کنید.

گام ۵: گام‌های سوم و چهارم را $(R-1)$ بار تکرار و مقادیر $T_1,\ldots,T_{R-1}$ را بدست آورید.

گام ۶: مقدار $\tilde{p}_{NRP}(R;\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)})$ را از رابطه (۱۰) را محاسبه کنید.

گام ۷: فرض H_0 را رد کنید اگر $\tilde{p}_{NRP} \leq \alpha$؛ در غیر اینصورت H_0 را بپذیرید.

برای محاسبه برآورد EQUALs از $\boldsymbol{\Sigma}_{22}^{-1}$ در گام‌های ۲ و ۴ الگوریتم فوق، از بسته نرم‌افزار R با عنوان EQUAL[1] معرفی شده توسط وانگ و جیانگ (۲۰۲۰) استفاده شده است. کدهای لازم برای اجرای الگوریتم فوق توسط نویسنده مقاله در صورت درخواست قابل ارائه است.

[1] https://github.com/cescwang85/EQUAL

۳ مطالعه شبیه‌سازی

برای ارزیابی عملکرد آزمون جایگشتی ϕ_{NRP} در (۱۱) در چهارچوب داده‌های با ابعاد بالا و همچنین مقایسه آزمون پیشنهادی با آزمون دقیق کلاسیک ϕ_{low} در (۳) در داده‌های نرمال بعد پایین، مطالعه شبیه‌سازی انجام شده است. ملاک ارزیابی و مقایسه این آزمون‌ها اندازه تجربی آزمون و توان تجربی آنها است. در این بخش فرض‌های زیر در نظر گرفته شده است: سطح معنی‌داری اسمی α برابر ۰/۰۵، تعداد جایگشت‌ها برابر $R =$ ۱۰۰ و تعداد تکرارهای آزمون M برابر ۵۰۰۰ لحاظ شده است. همچنین، بعد داده‌های شبیه‌سازی شده p از آن $\{۵, ۲۵, ۴۵, ۹۵, ۱۴۵, ۱۰۰, ۱۵۰, ۲۰۰, ۲۵۰, ۳۰۰\}$ و اندازه نمونه n عضوی $\{۵۰, ۱۰۰, ۱۵۰\}$ است. در کل شبیه‌سازی، بردار میانگین به صورت $\boldsymbol{\mu}_X = \circ_p$ در نظر گرفته شده است. به منظور اطمینان از کارایی آزمون پیشنهادی روی انتخاب‌های متنوع از ماتریس کوواریانس $\boldsymbol{\Sigma}_X$، این ماتریس به صورت یک ماتریس قطری با عناصر قطری برابر یک و عناصر غیر قطری با $i \neq j$ به صورت یکی از حالت‌های زیر تعریف شده است:

الف. ساختار مرکب[۱] $\boldsymbol{\Sigma}_\tau^{CSS}$ با درایه غیر قطری (i,j)ام برابر τ؛

ب. ساختار خودهمبستگی[۲] $\boldsymbol{\Sigma}_\tau^{AS}$ با درایه غیر قطری (i,j)ام برابر $\tau^{|i-j|}$؛

ج. ساختار آمیخته[۳] $\boldsymbol{\Sigma}_\tau^{MS}$ با درایه غیر قطری (i,j)ام برابر $۰/۵\tau^{|i-j|} + ۰/۵\tau$؛

در اینجا، فرض می‌شود $\tau \in \{۰, ۰/۳۵, ۰/۷۵\}$. هدف از این کار بررسی تأثیر همبستگی τ بر توان آزمون پیشنهادی است. توجه شود که $\tau = ۰$ اگر و تنها اگر $\boldsymbol{\Sigma}_\tau^{CSS} = \boldsymbol{\Sigma}_\tau^{AS} = \boldsymbol{\Sigma}_\tau^{MS} = \boldsymbol{I}_p$ که معادل با صفر بودن ضریب همبستگی چندگانه PMCC تحت هریک از ساختارهای کوواریانس معرفی شده است. به عبارت دیگر، در این شبیه‌سازی‌ها، به ازای $\tau = ۰$، نمونه‌های تصادفی شبیه‌سازی‌شده تحت فرض $H_۰$ و در غیر اینصورت تحت فرض مقابل هستند. علاوه بر این، مقادیر بزرگ τ نشانگر انحراف بیشتر از فرض $H_۰$ خواهند بود. نسبت‌های رد آزمون‌های مد نظر تحت ساختارهای کوواریانس $\boldsymbol{\Sigma}_\tau^{CSS}$، $\boldsymbol{\Sigma}_\tau^{AS}$ و $\boldsymbol{\Sigma}_\tau^{MS}$ بر اساس $M = ۵۰۰۰$ تکرار نمونه‌گیری محاسبه و در جداول ۱ تا ۳ آمده است.

ابتدا نسبت‌های رد متناظر با ستون‌های مربوط به $\tau = ۰$ را در نظر بگیرید. این نسبت‌ها به ازای $\tau = ۰$ در حقیقت نشانگر اندازه تجربی آزمون هستند. از آنجا که سطح معنی‌داری اسمی در اینجا $\alpha = ۰/۰۵$ است، خطای تقریب مونت‌کارلو به روش تجربی برابر $۱/۹۶\sqrt{\frac{۰/۰۵\times۰/۹۵}{۵۰۰۰}} = ۰/۰۰۶$ خواهد بود. این به این معناست که خطاهای نوع اول تجربی آزمون باید مقادیری بین ۰/۰۴۴ و ۰/۰۵۶ هستند. به ازای تمامی ترکیبات ممکن از مقادیر (n,p)، τ و $\boldsymbol{\Sigma}_X$، به وضوح مشاهده می‌شود که آزمون‌های ϕ_{NRP} و ϕ_{low} دارای خطاهای نوع اول نزدیک و کنترل‌شده در سطح $\alpha = ۰/۰۵$ هستند. توان تجربی آزمون به شدت به مقادیر نویز τ و همینطور ساختار کوواریانس وابسته است. از آنجا که سیگنال بزرگتر یه به عبارتی انحراف بزرگتر از فرض صفر قابلیت شناسایی بالاتری دارد، با ثابت نگه داشتن سایر پارامترها، افزایش قدرت سیگنال τ منجر به افزایش توان آزمون خواهد شد. همانطور که

---
[1] Compound structure
[2] Autocorrelation structure
[3] Mixed structure

جدول ۱. اندازه و توان تجربی آزمون‌های ϕ_{NRP} و ϕ_{low} برای ساختار کوواریانس Σ_{τ}^{CSS}

$\tau = ۰/۷۵$		$\tau = ۰/۳۵$		$\tau = ۰$		
ϕ_{low}	ϕ_{NRP}	ϕ_{low}	ϕ_{NRP}	ϕ_{low}	ϕ_{NRP}	(n, p)
۱	۱	۰/۸۴۸	۰/۹۱۴	۰/۰۵۳	۰/۰۴۷	(۵۰, ۵)
۰/۹۹۹	۱	۰/۵۱۳	۰/۹۸۸	۰/۰۵۰	۰/۰۴۹	(۵۰, ۲۵)
۰/۵۴۱	۱	۰/۱۲۱	۰/۹۹۴	۰/۰۵۲	۰/۰۵۲	(۵۰, ۴۵)
—	۱	—	۰/۹۹۵	—	۰/۰۵۵	(۵۰, ۱۰۰)
—	۱	—	۰/۹۹۴	—	۰/۰۵۲	(۵۰, ۱۵۰)
—	۱	—	۰/۹۹۶	—	۰/۰۵۱	(۵۰, ۲۰۰)
۱	۱	۰/۹۹۴	۰/۹۹۵	۰/۰۵۰	۰/۰۴۸	(۱۰۰, ۵)
۱	۱	۰/۸۵۷	۱	۰/۰۴۶	۰/۰۴۹	(۱۰۰, ۴۵)
۰/۵۳۷	۱	۰/۱۲۰	۱	۰/۰۴۵	۰/۰۴۸	(۱۰۰, ۹۵)
—	۱	—	۱	—	۰/۰۴۸	(۱۰۰, ۱۵۰)
—	۱	—	۱	—	۰/۰۵۱	(۱۰۰, ۲۰۰)
—	۱	—	۱	—	۰/۰۴۹	(۱۰۰, ۲۵۰)
۱	۱	۱	۱	۰/۰۴۷	۰/۰۴۷	(۱۵۰, ۵)
۱	۱	۰/۸۰۲	۱	۰/۰۴۶	۰/۰۴۹	(۱۵۰, ۹۵)
۰/۵۳۲	۰/۹۵۷	۰/۱۱۹	۱	۰/۰۵۰	۰/۰۴۷	(۱۵۰, ۱۴۵)
—	۱	—	۱	—	۰/۰۴۸	(۱۵۰, ۲۰۰)
—	۱	—	۱	—	۰/۰۵۱	(۱۵۰, ۲۵۰)
—	۱	—	۱	—	۰/۰۴۹	(۱۰۰, ۳۰۰)

جدول ۲. اندازه و توان تجربی آزمون‌های ϕ_{NRP} و ϕ_{low} برای ساختار کوواریانس Σ_{τ}^{AS}

$\tau = ۰/۷۵$		$\tau = ۰/۳۵$		$\tau = ۰$		
ϕ_{low}	ϕ_{NRP}	ϕ_{low}	ϕ_{NRP}	ϕ_{low}	ϕ_{NRP}	(n, p)
۱	۱	۰/۴۷۶	۰/۴۸۲	۰/۰۵۳	۰/۰۴۷	(۵۰, ۵)
۰/۹۴۷	۰/۹۸۹	۰/۱۴۱	۰/۱۹۲	۰/۰۵۰	۰/۰۴۹	(۵۰, ۲۵)
۰/۲۶۰	۰/۹۳۸	۰/۰۶۷	۰/۱۵۲	۰/۰۵۲	۰/۰۵۲	(۵۰, ۴۵)
—	۰/۷۲۰	—	۰/۱۰۸	—	۰/۰۵۵	(۵۰, ۱۰۰)
—	۰/۵۵۸	—	۰/۰۹۶	—	۰/۰۵۲	(۵۰, ۱۵۰)
—	۰/۴۷۴	—	۰/۰۸۹	—	۰/۰۵۱	(۵۰, ۲۰۰)
۱	۱	۰/۸۲۷	۰/۸۲۰	۰/۰۵۰	۰/۰۴۸	(۱۰۰, ۵)
۱	۱	۰/۲۳۴	۰/۳۲۱	۰/۰۴۶	۰/۰۴۹	(۱۰۰, ۴۵)
۰/۲۵۸	۰/۹۹۷	۰/۰۶۶	۰/۲۰۴	۰/۰۴۵	۰/۰۴۸	(۱۰۰, ۹۵)
—	۰/۹۸۲	—	۰/۱۵۴	—	۰/۰۴۸	(۱۰۰, ۱۵۰)
—	۰/۹۴۷	—	۰/۱۴۹	—	۰/۰۴۹	(۱۰۰, ۲۰۰)
—	۰/۹۲۶	—	۰/۱۳۷	—	۰/۰۵۰	(۱۰۰, ۲۵۰)
۱	۱	۰/۹۵۶	۰/۹۵۹	۰/۰۴۷	۰/۰۴۷	(۱۵۰, ۵)
۰/۹۹۸	۱	۰/۲۰۵	۰/۳۵۹	۰/۰۴۶	۰/۰۴۹	(۱۵۰, ۹۵)
۰/۲۴۸	۱	۰/۰۶۴	۰/۲۸۰	۰/۰۵۰	۰/۰۴۷	(۱۵۰, ۱۴۵)
—	۱	—	۰/۲۱۸	—	۰/۰۴۶	(۱۵۰, ۲۰۰)
—	۱	—	۰/۲۰۱	—	۰/۰۴۷	(۱۵۰, ۲۵۰)
—	۰/۹۹۴	—	۰/۱۹۵	—	۰/۰۴۹	(۱۰۰, ۳۰۰)

مشاهده می‌شود، به ازای همه ترکیبات (n, p)، τ و Σ_X، توان آزمون ϕ_{NRP} بزرگتر از آزمون ϕ_{low} است. علاوه بر این، آزمون ϕ_{NRP} به ازای نویزهای کوچک یا همان انحراف‌های کوچک از فرض صفر دارای توان بالاتر نسبت

جدول ۳. اندازه و توان تجربی آزمون‌های ϕ_{NRP} و ϕ_{low} برای ساختار کوواریانس $\boldsymbol{\Sigma}_\tau^{MS}$

	$\tau = ۰$		$\tau = ۰/۳۵$		$\tau = ۰/۷۵$	
(n, p)	ϕ_{NRP}	ϕ_{low}	ϕ_{NRP}	ϕ_{low}	ϕ_{NRP}	ϕ_{low}
(۵۰, ۵)	۰/۰۴۷	۰/۰۵۳	۰/۶۵۵	۰/۵۸۶	۱	۱
(۵۰, ۲۵)	۰/۰۴۹	۰/۰۵۰	۰/۷۶۱	۰/۲۳۶	۱	۰/۹۷۴
(۵۰, ۴۵)	۰/۰۵۲	۰/۰۵۲	۰/۸۰۹	۰/۰۸۱	۰/۹۹۹	۰/۳۰۴
(۵۰, ۱۰۰)	۰/۰۵۵	—	۰/۸۳۱	—	۰/۹۹۹	—
(۵۰, ۱۵۰)	۰/۰۵۲	—	۰/۸۳۶	—	۰/۹۹۷	—
(۵۰, ۲۰۰)	۰/۰۵۱	—	۰/۸۴۴	—	۰/۹۹۹	—
(۱۰۰, ۵)	۰/۰۴۸	۰/۰۵۰	۰/۹۲۴	۰/۹۱۱	۱	۱
(۱۰۰, ۴۵)	۰/۰۴۹	۰/۰۴۶	۰/۹۷۹	۰/۴۴۷	۱	۱
(۱۰۰, ۹۵)	۰/۰۴۸	۰/۰۴۵	۰/۹۸۸	۰/۰۸۱	۱	۰/۳۰۳
(۱۰۰, ۱۵۰)	۰/۰۴۸	—	۰/۹۸۷	—	۱	—
(۱۰۰, ۲۰۰)	۰/۰۴۹	—	۰/۹۹۰	—	۱	—
(۱۰۰, ۲۵۰)	۰/۰۴۹	—	۱	—	۱	—
(۱۵۰, ۵)	۰/۰۴۷	۰/۰۴۷	۰/۹۹۰	۰/۹۸۸	۱	۱
(۱۵۰, ۹۵)	۰/۰۴۹	۰/۰۴۶	۱	۰/۴۰۳	۱	۱
(۱۵۰, ۱۴۵)	۰/۰۴۷	۰/۰۵۰	۱	۰/۰۸۴	۱	۰/۲۸۸
(۱۵۰, ۲۰۰)	۰/۰۴۶	—	۰/۹۹۲	—	۱	—
(۱۵۰, ۲۵۰)	۰/۰۴۸	—	۱	—	۱	—
(۱۰۰, ۳۰۰)	۰/۰۵۰	—	۱	—	۱	—

به آزمون ϕ_{low} است. به ازای $p = n - ۵$، برای داده‌های با بعد نسبتا بالا[۱]، یعنی، حالتی که $p < n$ و $\frac{p}{n} \to ۱$، توان آزمون ϕ_{low} به شدت از توان آزمون ϕ_{NRP} کوچکتر است. برای داده‌های با بعد بالا[۲]، یعنی حالتی که در آن $n < p$، آزمون ϕ_{low} به دلیل تکین بودن ماتریس کوواریانس نمونه‌ای غیر قابل استفاده است. برای این نوع از داده‌ها، توان‌های تجربی آزمون ϕ_{NRP}، به ازای $\tau = ۰/۳۵$، برای ساختارهای کوواریانس $\boldsymbol{\Sigma}_\tau^{CSS}$ و $\boldsymbol{\Sigma}_\tau^{MS}$، نشان می‌دهد که آزمون پیشنهادی حتی برای انحرافات کوچکتر نیز دارای توان بالاتری است. نتایج شبیه‌سازی برای ساختار کوواریانس $\boldsymbol{\Sigma}_\tau^{AS}$ در جدول ۲ نشان می‌دهد که توان آزمون پیشنهادی با افزایش بعد داده‌ها کاهش می‌یابد. چرا که در این حالت ماتریس $\boldsymbol{\Sigma}_\tau^{AS}$ به ماتریس همانی $\boldsymbol{I}_p$ همگراست. این نتیجه می‌دهد که ضریب همبستگی ρ برای ساختار $\boldsymbol{\Sigma}_\tau^{AS}$ وقتی $p \to \infty$ به صفر همگرا خواهد بود. برای ساختار کوواریانس $\boldsymbol{\Sigma}_\tau^{AS}$، شناسایی انحرافات از فرض صفر برای هر روش آزمون فرض از جمله آزمون پیشنهادی وقتی $p \to \infty$ کار آسانی نیست.

۴ تحلیل اندازه‌های تومور موش‌ها

تان و همکاران (۲۰۰۵) مؤثر بودن داروی جدید ضدسرطان ایرونوتیکن[۳] (**CPT-11**) در مقابل بافت‌های سرطانی

[1]Moderately high-dimensional data
[2]High-dimensional data
[3]Irinotecan anticancer drug

نیروبلاستوما[1] تحت دو رژیم درمانی را بررسی نمودند. موش‌های سرطانی در گروه اول به میزان ۰/۴ میلی‌گرم بر کیلوگرم **CPT-11** و موش‌های سرطانی در گروه دوم به میزان ۰/۲۶ میلی‌گرم بر کیلوگرم **CPT-11** دریافت نمودند. در این مطالعه از موش‌های با نژاد یکسان استفاده شده است. اندازه‌های تومور بر حسب سانتی‌متر مکعب یک بار در هفته آغازین، یعنی هفته ۰ و یک بار در هفته به مدت ۱۲ هفته، ثبت شده‌اند. داده‌های گم شده به دلیل مرگ ۶ موش به خاطر مسمومیت یا رشد چهار برابری اندازه تومور در طول مطالعه مشاهده شده است. تان و همکاران (۲۰۰۵) و لیانگ و همکاران (۲۰۰۹) از این مجموعه داده به عنوان کاربردی بر روش‌های آزمون پیشنهادی خود برای فرض نرمال بودن استفاده نمودند. برای نشان دادن کاربردی از آزمون پیشنهادی، دو زیر مجموعه از این داده‌ها (جدول ۵) در نظر گرفته شده است: مجموعه اول شامل مشاهدات مربوط به هشت موش با شماره‌های ۲، . . .، ۸ و ۱۱ برای هفته‌های ۰ تا ۱۰ ($p =$ ۱۱ هفته)، مجموعه دوم از داده‌ها شامل $n =$ ۵ موش با شماره‌های ۲، ۳، ۴، ۷ و ۱۱ برای هفته‌های ۰ تا ۱۱ ($p =$ ۱۲ هفته) است.

جدول ۴. اندازه‌های تومور موش‌ها

		هفته											
مجموعه داده	موش	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
اول	۲	۱/۱۱	۱/۵۴	۰/۸۱	۰/۹۳	۱/۳۷	۱/۰۰	۰/۹۸	۱/۷۰	۰/۷۷	۰/۵۱	۱/۸۸	
	۳	۰/۹۶	۰/۹۹	۰/۹۹	۰/۵۳	۰/۷۲	۰/۳۶	۰/۱۷	۰/۶۳	۰/۴۵	۰/۴۸	۲/۱۴	
	۴	۰/۶۶	۰/۶۰	۰/۴۹	۰/۷۸	۱/۴۰	۱/۳۳	۱/۱۱	۱/۳۱	۰/۶۹	۰/۹۸	۳/۴۴	
	۵	۲/۰۸	۲/۱۵	۱/۸۷	۰/۸۳	۰/۷۸	۰/۲۶	۰/۱۶	۰/۶۴	۰/۳۰	۰/۵۱	۲/۰۲	
	۶	۱/۰۹	۱/۰۴	۰/۷۶	۰/۷۷	۰/۷۲	۰/۲۹	۰/۲۰	۰/۲۹	۰/۶۸	۰/۵۵	۰/۵۵	
	۷	۰/۷۴	۰/۹۳	۰/۸۳	۰/۵۹	۰/۶۰	۰/۴۱	۰/۲۲	۰/۳۱	۰/۱۵	۰/۱۸	۱/۰۳	
	۸	۰/۹۴	۱/۱۲	۱/۶۷	۲/۶۹	۳/۵۱	۲/۷۷	۲/۳۶	۲/۸۹	۲/۴۹	۲/۶۴	۳/۹۶	
	۱۱	۱/۲۴	۱/۳۲	۱/۶۳	۲/۴۳	۳/۰۰	۲/۰۴	۱/۰۸	۱/۰۷	۰/۳۹	۰/۸۸	۱/۸۶	
دوم	۲	۱/۱۱	۱/۵۴	۰/۸۱	۰/۹۳	۱/۳۷	۱/۰۰	۰/۹۸	۱/۷۰	۰/۷۷	۰/۵۱	۱/۸۸	۴/۹۷
	۳	۰/۹۶	۰/۹۹	۰/۹۹	۰/۵۳	۰/۷۲	۰/۳۶	۰/۱۷	۰/۶۳	۰/۴۵	۰/۴۸	۲/۱۴	۲/۸۹
	۴	۰/۶۶	۰/۶۰	۰/۴۹	۰/۷۸	۱/۴۰	۱/۳۳	۱/۱۱	۱/۳۱	۰/۶۹	۰/۹۸	۳/۴۴	۳/۰۸
	۷	۰/۷۴	۰/۹۳	۰/۸۳	۰/۵۹	۰/۶۰	۰/۴۱	۰/۲۲	۰/۳۱	۰/۱۵	۰/۱۸	۱/۰۳	۸/۲۱
	۱۱	۱/۲۴	۱/۳۲	۱/۶۳	۲/۴۳	۳/۰۰	۲/۰۴	۱/۰۸	۱/۰۷	۰/۳۹	۰/۸۸	۱/۸۶	۵/۹۰

لیانگ و همکاران (۲۰۰۹) نشان دادند فرض نرمال چندمتغیّره بودن هر دو مجموعه از داده‌ها را نمی‌توان در سطح معنی داری ۰/۰۵ رد کرد. در این بخش آزمون صفر بودن ضریب همبستگی چندگانه بین اندازه تومور در هفته ابتدایی ۰ و اندازه‌های تومور در هفته‌های ۱ تا ۱۰ در مجموعه داده اول و همینطور اندازه‌های تومور در هفته‌های ۱ تا ۱۱ در مجموعه داده دوم مورد بررسی قرار می‌گیرد. به ازای $R =$ ۱۰۰ و $\alpha =$ ۰/۰۵، برای مجموعه داده اول، پی‌مقدار برابر ۰/۲۴۰ و برای مجموعه داده دوم برابر ۰/۳۸۰ است. از آنجا این مقادیر از $\alpha =$ ۰/۰۵ بزرگتر هستند، نمی‌توان فرض صفر بودن ضریب همبستگی چندگانه در دو گروه را رد کرد. با توجه به اینکه فرض نرمال بودن داده‌ها در دو گروه مورد تایید است، از نتایج حاصل از آزمون‌ها می‌توان چنین استنباط نمود که اندازه تومور در هفته ابتدایی ۰ مستقل از هر ترکیب خطی از اندازه تومور در هفته‌های ۱ تا ۱۰ در مجموعه داده اول و همینطور اندازه‌های تومور

[1] Neuroblastoma in xenografts

در هفته‌های ۱ تا ۱۱ در مجموعه داده دوم است. به عبارت دیگر، اندازه تومور در هفته ابتدایی ۰ به شکل همزمان بر اندازه‌های تومور در هفته‌های ۱ تا ۱۰ در مجموعه داده اول و همینطور اندازه‌های تومور در هفته‌های ۱ تا ۱۱ در مجموعه داده دوم مؤثر نیست. این امر را می‌توان نشانه‌ای از موثر بودن مداخله دارویی **CPT-11** دانست، چرا که در صورت بی اثر بودن دارو و حفظ شرایط درمانی موجود انتظار بر این بود که اندازه‌های تومورها رفته رفته رشد نموده و ضریب همبستگی چندگانه قوی در گروه‌های اول و دوم مشاهده شود.

فرض کنید مجموعه داده‌های بعد پایین سوم و چهارم را به صورت اندازه‌های تومور موش‌ها در هفته‌های ۰ تا ۲ در مجموعه داده‌های اول و دوم هستند. در اینجا مجموعه داده‌های سوم و چهارم را زیر مجموعه‌ای از داده‌های اول و دوم انتخاب شده‌اند. این بار به منظور نمایش کاربردی دیگر از آزمون پیشنهاد شده برای مجموعه داده‌های بعد پایین، آزمون صفر بودن ضریب همبستگی چندگانه بین اندازه‌های تومور در هفته آغازین ۰ و اندازه‌های تومور در هفته‌های ۱ و ۲ در این دو مجموعه داده مد نظر است. پی‌مقدار آزمون پیشنهادی ϕ_{NRP} برای مجموعه داده سوم برابر ۰/۰۰۱۴ و برای مجموعه داده چهارم برابر ۰/۰۱۶۹ به دست می‌آید. بنابراین با استفاده از ϕ_{NRP} فرض صفر بودن ضریب همبستگی چندگانه در دو مجموعه داده رد می‌شود. به بیان نه چندان دقیق، می‌توان گفت تاثیر معناداری اثر دارو در تنها دو هفته بعد اعمال آن در اندازه‌های تومور را نمی‌توان مشاهده کرد. همچنین پی مقدار آزمون دقیق کلاسیک ϕ_{low} برای مجموعه داده سوم برابر ۰/۰۴۶ و برای مجموعه داده چهارم برابر ۰/۰۶۶ به دست می‌آید. از این رو، فرض صفر بودن ضریب همبستگی چندگانه در مجموعه داده سوم رد و در مجموعه داده چهارم پذیرفته می‌شود. البته باید این نکته را در نظر داشته باشیم که برای داده‌های با بعد نسبتاً بالا، این آزمون دارای توان خیلی کوچک است و از این رو این نتایج را نمی‌توان چندان معتبر دانست. جداول ۱ تا ۳ برای p نزدیک به n را مشاهده کنید.

بحث و نتیجه‌گیری

ضریب همبستگی چندگانه جامعه به طور گسترده در تحلیل همبستگی چندگانه و رگرسیون برای توصیف همبستگی بین یک متغیر و مجموعه‌ای از متغیرها استفاده می‌شود. برای ارزیابی وجود این نوع همبستگی، آزمون فرض صفر بودن PMCC را می‌توان به عنوان راه حل در نظر گرفت. در داده‌های کلاسیک با ابعاد پایین، آزمون دقیقی برای این فرض وجود دارد. در داده‌های با ابعاد بالا، آزمون دقیق کلاسیک به دلیل تکین بودن ماتریس کوواریانس نمونه دیگر قابل استفاده نیست. با توجه به کاربرد گسترده PMCC در داده‌های با ابعاد بالا، معرفی یک روش جدید برای آزمون فرض صفر بودن PMCC مورد نیاز است. به منظور آزمون این فرض، روش‌هایی توسط نجارزاده (۲۰۲۰، ۲۰۲۲) پیشنهاد شده است. در این مقالات، برای آزمون صفر بودن ضریب همبستگی چندگانه از ایده تصویرسازی تصادفی بدون و با انتخاب بهینه بعد تصویرسازی استفاده شده است. در ایده تصویرسازی داده‌های بعد بالا بر بعد پایین، همیشه مشکل اصلی انتخاب ماتریس تصویرساز است که همواره بحثی چالش برانگیز است. با این انگیزه، آزمونی جایگشتی برای فرض صفر بودن PMCC در داده‌های نرمال با ابعاد بالا معرفی شد. به کمک مطالعه شبیه‌سازی، جدا از اینکه رابطه بین بعد p با اندازه نمونه n چگونه است، نشان داده شده است که آزمون پیشنهادی کنترل خوبی

بر خطای نوع اول دارد. نتایج شبیه‌سازی همچنان نشان داد که آزمون پیشنهادی نسبت به آزمون دقیق کلاسیک در تنظیمات داده‌های نرمال با بعد پایین، توان بالاتری دارد و در تشخیص انحرافات کوچک از فرض صفر در داده‌های نرمال با ابعاد بالا، موفق عمل می‌کند.

تقدیر و تشکر

نویسنده مقاله از دو داور محترم، سردبیر و ویراستار مجله به دلیل تلاش ایشان در راستای ارتقای کیفی مقاله و برطرف نمودن ایرادات احتمالی کمال تشکر و قدردانی را دارد.

مراجع

Cai, T., Liu, W. and Luo, X. (2011), A constrained L1 Minimization Approach to Sparse Precision Matrix Estimation, *Journal of the American Statistical Association*, **106**(494), 594-607.

Friedman, J., Hastie, T. and Tibshirani, R. (2008), Sparse Inverse Covariance Estimation with the Graphical Lasso, *Biostatistics*, **9**(3), 432-441.

Gupta, S. D. (1977), Tests on Multiple Correlation Coefficient and Multiple Partial Correlation Coefficient, *Journal of Multivariate Analysis*, **7**(1), 82-88.

Hoeffding, W. (1952), The Large-sample Power of Tests Based on Permutations of Observations, *The Annals of Mathematical Statistics*, **23**(2), 169-192.

Hsieh, C. J., Sustik, M. A., Dhillon, I. S., Ravikumar, P. K., and Poldrack, R. (2013), Big & quic: Sparse Inverse Covariance Estimation for a Million Variables, *In Advances in Neural Information Processing Systems*, 3165-3173.

Liang, J., Tang, M. L. and Chan, P. S. (2009), A Generalized Shapiro-wilk W Statistic for Testing High-dimensional Normality, *Computational Statistics & Data Analysis*, **53**(11), 3883-3891.

Liu, W. and Luo, X. (2015), Fast and Adaptive Sparse Precision Matrix Estimation in High Dimensions, *Journal of Multivariate Analysis*, **135**, 153-162.

Muirhead, R. (2005), Aspects of Multivariate Statistical Theory, John Wiley and Sons, New York.

Najarzadeh, D. (2020), A Simple Test for Zero Multiple Correlation Coefficient in High-dimensional Normal Data using Random Projection, *Computational Statistics & Data Analysis*, **148**, 106955.

Najarzadeh, D. (2022), An Optimal Projection Test for Zero Multiple Correlation Coefficient in High-dimensional Normal Data, *Communications in Statistics-Theory & Methods*, **51**(4), 1011-1028.

Pourahmadi, M. (2013), High-dimensional Covariance Estimation: with High-dimensional Data, John Wiley and Sons, New York.

Provost, S. B. (1987), Testing for the Nullity of the Multiple Correlation Coefficient with Incomplete Multivariate Data, *In Advances in the Statistical Sciences: Foundations of Statistical Inference*, 149-161.

Romano, J. P. and Wolf, M. (2005), Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing, *Journal of the American Statistical Association*, **100**(469), 94-108.

Tan, M., Fang, H. B., Tian, G. L. and Wei, G. (2005), Testing Multivariate Normality in Incomplete Data of Small Sample Size, *Journal of Multivariate Analysis*, **93**(1), 164-179.

Wang, C. and Jiang, B. (2019), EQUAL: An Efficient ADMM Algorithm for High Dimensional Precision Matrix Estimation via Penalized Quadratic Loss, R package version 1.2.

Wang, C. and Jiang, B. (2020), An Efficient Admm Algorithm for High Dimensional Precision Matrix Estimation via Penalized Quadratic Loss, *Computational Statistics & Data Analysis*, **142**, 106812.

Zheng, S., Jiang, D., Bai, Z. and He, X. (2014), Inference on Multiple Correlation Coefficients with Moderately High Dimensional Data, *Biometrika*, **101**(3), 748-754.