



Asymptotic Distribution of Some Statistics in Multivariate Inference Based on Taylor Series Expansion

Sakineh Dehghan ,

Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran.

Corresponding author: Sakineh Dehghan sa_dehghan@sbu.ac.ir

Received: 5/9/2022 Revised: 22/12/2022 Accepted and Published Online: 27/12/2022.

Introduction

The exact distribution of many applicable statistics could not be accessible in various statistical inference problems. To deal with such an issue in the large sample problem, an approach is to obtain the asymptotic distribution. The central limit theorem describes the asymptotic behavior of statistics defined as arithmetic averages. However, most statistics are not restricted to this form but are asymptotically equivalent to arithmetic averages. In this article, we have expressed the asymptotic distribution of multivariate statistics class approximated by averages based on the Taylor expansion. Then, the asymptotic distribution of an empirical Mahalanobis depth-based statistic is obtained, and the statistic is applied to test the scale difference between two multivariate distributions.

Material and Methods

Given a multivariate distribution, a depth function measures the centrality of a multivariate point by a number while satisfying certain properties regarding invariance, monotonicity, convexity, and continuity. Accordingly, several notions of multivariate depth have been proposed in the literature, and several depth-based tests have been offered for testing different types of hypotheses in multivariate nonparametric settings. This paper presents an overview of the depth function concept, precisely the Mahalanobis depth. Then, we proposed a depth-based test statistic and derived its asymptotic distribution based on the Taylor expansion. The methodology has been verified using a simulation study and a real data set. Calculations have been

performed in R software, and R codes are available by request to the author.

Results and Discussion

In this paper, the asymptotic distribution of the mean of empirical Mahalanobis depth of a sample of random vectors has been achieved. The following statistic has been employed to test the scale difference between two multivariate distributions. Then, simulation studies are carried out to explore the behavior of the asymptotic distribution of the test statistic. Our empirical investigations show that the asymptotic null distribution is substantially accurate for the sample size $n = ۲۰۰$. Finally, to illustrate the method developed in this paper, we have applied the statistics on a well-known data set to test the scale difference between three species of Iris. The results declare that the scale differences are significant and coincide with previous studies results.

Conclusion

Indeed, most statistics that arise in practice are average-like, and the statistician familiar with the asymptotic theory of averages can handle most statistical asymptotic problems. This paper presents the asymptotic distribution of a broad class of multivariate statistics approximated by averages. Accordingly, a depth-based statistic has been proposed to test scale differences between two multivariate distributions, and its asymptotic distribution has been derived. The procedure has been completed by approximating the moments of the statistic. The numerical results confirm the accuracy of the asymptotic distribution.

Keywords: Multivariate scale test, Taylor series expansion, Asymptotic distribution, Mahalanobis depth, Central limit theorem.

Mathematics Subject Classification (2010): 60F05, 62H10.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [\(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

توزیع مجانبی برخی آمارها در استنباط چندمتغیره بر اساس بسط سری تیلور

سکینه دهقان

گروه آمار، دانشکده علوم ریاضی، دانشگاه شهید بهشتی

نویسنده مسئول : سکینه دهقان، sa_dehghan@sbu.ac.ir

تاریخ دریافت: ۱۴۰۱/۶/۱۴ تاریخ بازنگری: ۱۴۰۱/۱۰/۱ تاریخ پذیرش و انتشار: ۱۴۰۱/۱۰/۶

چکیده: توزیع دقیق بسیاری از آمارهای پرکاربرد در مسائل مختلف استنباط آماری، قابل دستیابی نیست. یک رویکرد جایگزین در حالت بزرگ نمونه‌ای بدست آوردن توزیع مجانبی است. در این مقاله، توزیع مجانبی یک رده خاص از آمارهای چندمتغیره را که بر اساس بسط سری تیلور دارای نمایش تقریبی به صورت میانگین بردارهای مستقل هستند، بیان می‌کنیم. در ادامه، توزیع مجانبی یک آمار مبتنی بر تابع ژرفای ماکسول-بولتزمن تجربی حاصل می‌شود و آماره برای آزمون اختلاف مقیاس بین دو توزیع چندمتغیره به کار می‌رود. مطالعات شبیه‌سازی به منظور بررسی رفتار توزیع مجانبی آماره آزمون انجام می‌شود و همچنین آماره پیشنهادی برای تحلیل یک مجموعه داده واقعی به کار می‌رود.

واژه‌های کلیدی: آزمون مقیاس چندمتغیره، بسط سری تیلور، توزیع مجانبی، ژرفای ماکسول-بولتزمن، قضیه حد مرکزی

کد موضوع بندی ریاضی (۲۰۱۰): 62H10 ، 60F05.

۱ مقدمه

یک مسئله اساسی در استنباط آماری، بدست آوردن توزیع دقیق یا مجانبی یک آماره بر اساس یک نمونه تصادفی است. معمولاً در استنباط آماری به دو منظور آمارها به کار می‌روند. در بحث برآوردیابی، به منظور برآورد پارامترهای جامعه از آمارها استفاده می‌شود و برای بررسی میزان دقت برآوردگر به توزیع آن نیاز خواهیم داشت. همچنین هرگاه با یک فرض آماری برای پارامترهای جامعه روبرو هستیم، به یک آماره آزمون نیاز خواهیم داشت که به منظور تعیین



©نویسندگان). ناشر انجمن آمار ایران است.

این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

ناحیه بحرانی و محاسبه توان آزمون باید توزیع آماره آزمون را تعیین کرد. در بسیاری از موارد، توزیع دقیق آماره‌ها قابل دستیابی نیست، که در این صورت در تلاش خواهیم بود، توزیع مجانبی آنها را به عنوان تقریب بزرگ نمونه‌ای بدست آوریم.

تئوری بزرگ نمونه‌ای برای آماره‌هایی که به صورت میانگین متغیرهای تصادفی تعریف می‌شوند بر قانون ضعیف اعداد بزرگ و قضیه حد مرکزی استوار است. اگر $\bar{X}$ میانگین نمونه‌ای n متغیر تصادفی مستقل و هم‌توزیع $X_1, \dots, X_n$ با میانگین $\mu = E(X_i)$ باشد، قانون ضعیف اعداد بزرگ تضمین می‌کند که $\bar{X}$ در احتمال به اندازه کافی به μ نزدیک است. همچنین قضیه حد مرکزی بیان می‌کند هرگاه $n \rightarrow \infty$ و $\sigma^2 = Var(X_1)$ متناهی باشد آنگاه

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq t\right) = \Phi(t)$$

که در آن $\Phi(t)$ تابع توزیع تجمعی متغیر تصادفی نرمال است. واضح است قضیه حد مرکزی و قانون اعداد بزرگ برای میانگین حسابی متغیرهای تصادفی کاربرد دارند، در حالی که بسیاری از آمارها از جمله گشتاورهای مرکزی به صورت میانگین حسابی توابعی دلخواه از متغیرهای تصادفی و برآوردگرهایی از پارامتر نامعلوم که بر اساس متغیرهای تصادفی تعریف شده است، بیان می‌شوند. به بیان دقیق‌تر، با اینکه فرض مستقل و هم‌توزیع بودن $X_1, \dots, X_n$ همچنان برقرار است، در تعریف این آمارها به دلیل اینکه جمع‌وندها مبتنی بر آمارهای دیگر هستند، فرض استقلال بین جمع‌وندها وجود ندارد و بدست آوردن توزیع مجانبی آماره‌هایی از این نوع سراسر نخواهد بود. در این مقاله ابتدا در حالت چندمتغیره، یک کلاس خاص از آمارها بیان شده و نشان داده می‌شود که چگونه این آمارها بر اساس بسط تیلور، به صورت تقریبی نمایشی به صورت میانگین بردارهای تصادفی دارند. سپس قضیه‌های مرتبط برای بدست آوردن توزیع مجانبی این آمارها بیان می‌شوند. در ادامه به عنوان یک مثال توزیع مجانبی میانگین ژرفای ماهالانوبیس^۱ یک نمونه از بردارهای تصادفی حاصل می‌شود و کاربرد این آماره در آزمون اختلاف مقیاس بین دو توزیع چندمتغیره بیان می‌شود. توابع ژرفا در بسیاری از مباحث آمار چندمتغیره از جمله برآوردیابی، آزمون فرضیه، رده‌بندی، خوشه‌بندی، کشف نقاط دورافتاده به عنوان یک ابزار به‌کار می‌روند. به‌ویژه در مبحث آزمون فرضیه، آماره‌های چندمتغیره بسیاری بر اساس این توابع تعریف شده‌اند. **لی و لیو (۲۰۰۴)** یک نمودار بر اساس توابع ژرفا معرفی کردند و بر پایه آن، دو رده از آماره‌های آزمون برای آزمون اختلاف مکان بین دو توزیع چندمتغیره معرفی کردند. در ادامه **لیو و سینگ (۲۰۰۶)**، بر اساس تابع ژرفا به رده‌بندی داده‌های چندمتغیره پرداختند و یک رده از آماره‌های آزمون رتبه‌ای برای اختلاف مقیاس بین دو توزیع چندمتغیره ارائه دادند. برای مسئله مکان چندمتغیره یک‌نمونه‌ای و چندنمونه‌ای نیز بر اساس تابع ژرفا رده‌هایی از آماره‌های آزمون توسط **دهقان و فریدروحانی (۲۰۲۱)**، **دهقان و فریدروحانی (۲۰۲۲)** پیشنهاد شد. بر اساس تابع دورافتادگی ژرفا-مبنا **دهقان و فریدروحانی (۱۴۰۰)** روشی برای شناسایی نقاط دورافتاده چندمتغیره بیان کردند.

¹Mahalanobis Depth

در بخش ۲ توزیع مجانبی یک رده خاص از آماره‌های چندمتغیره بیان می‌شود. در بخش ۳ مفهوم ژرفای داده و تعریف تابع ژرفا مرور می‌شوند، سپس تابع ژرفای ماهالانوبیس تعریف می‌شود. در بخش ۴ به عنوان یک کاربرد از قضیه مطرح شده در بخش ۲، توزیع مجانبی مجموع ژرفای ماهالانوبیس نمونه تصادفی اثبات می‌شود. در بخش ۵ با روش‌های عددی نتایج بدست آمده در بخش ۴ مورد بررسی قرار خواهد گرفت. سرانجام در بخش ۶، آماره معرفی شده برای تحلیل داده‌های واقعی به کار می‌رود.

۲ توزیع مجانبی میانگین توابعی از بردارهای تصادفی و برآوردگر پارامترها

فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی مستقل و هم‌توزیع از بردار تصادفی p -متغیره X با میانگین متناهی μ باشد. همچنین فرض کنید $\bar{X}$ میانگین نمونه‌ای، f تابعی دلخواه از R^p به R و $f'(x)$ بردار سطری $\partial f(x)/\partial x$ باشد. بر اساس قانون اعداد بزرگ چندمتغیره واضح است $\bar{X}$ به میانگین جامعه μ در احتمال میل می‌کند، بنابراین می‌توان بسط تیلور $f(\bar{X})$ حول μ را به صورت

$$f(\bar{X}) \approx f(\mu) + f'(\mu)(\bar{X} - \mu) = n^{-1} \sum_{i=1}^n \{f(\mu) + f'(\mu)(X_i - \mu)\}, \quad (۱)$$

نوشت. بنابراین $\bar{W} = n^{-1} \sum_{i=1}^n W_i$ که در آن $W_i = f(\mu) + f'(\mu)(X_i - \mu)$ بنابراین در اینجا تابع دلخواهی از $\bar{X}$ به صورت میانگین متغیرهای تصادفی مستقل و هم‌توزیع $W_1, \dots, W_n$ تعریف شد. حال برای بدست آوردن توزیع مجانبی $f(\bar{X})$ می‌توان قضیه حد مرکزی را به کار برد.

فرض کنید $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_p)$ تابعی از $X_1, \dots, X_n$ و برآوردگر بردار پارامتر $\eta = (\eta_1, \dots, \eta_p)$ باشد. رابطه (۱) ابزاری برای بدست آوردن توزیع مجانبی آماره‌هایی به صورت $n^{-1} \sum_{i=1}^n f(X_i, \hat{\eta})$ که میانگین توابعی از بردارهای تصادفی و برآوردگر $\hat{\eta}$ هستند، نیز فراهم آورده است که در ادامه به آن پرداخته می‌شود. فرض کنید بتوان توابع $h_j(X_i)$ ، $j = 1, \dots, p$ ، $i = 1, \dots, n$ را پیدا کرد که در شرایط

$$E\{h_j(X_i)\} = 0, \quad \text{Var}\{h_j(X_i)\} < \infty,$$

صدق کنند و هر مولفه بردار تصادفی $\hat{\eta} = (\hat{\eta}_1, \dots, \hat{\eta}_p)$ نمایشی به صورت

$$\hat{\eta}_j - \eta_j = \frac{1}{n} \sum_{i=1}^n h_j(X_i) + R_{nj}, \quad j = 1, \dots, p \quad (۲)$$

داشته باشد، که برای $n \rightarrow \infty$ ، $\sqrt{n}R_{nj} = o_p(1)$ ، حال فرض کنید $T_n = n^{-1} \sum_{i=1}^n f(X_i, \hat{\eta})$ که

۴۸ توزیع مجانبی برخی آمارها در استنباط چندمتغیره

مشتقات جزئی تابع f در همسایگی η وجود دارند و در η پیوسته است. **بووس و استفانسکی (۲۰۱۳)** ثابت کردند اگر رابطه (۲) برای مولفه‌های $\hat{\eta}$ برقرار باشد، آنگاه

$$\frac{1}{n} \sum_{i=1}^n f(\mathbf{X}_i, \hat{\eta}) - E(f(\mathbf{X}_i, \eta)) = \frac{1}{n} \sum_{i=1}^n H(\mathbf{X}_i) + R_n \quad (۳)$$

که $\sqrt{n}R_n = o_p(1)$ و $H(\mathbf{X}_i) = f(\mathbf{X}_i, \eta) - E(f(\mathbf{X}_i, \eta)) + [E(f'(\mathbf{X}_i, \eta))] h(\mathbf{X}_i)$ و برای $\hat{\eta}^*$ بین $\hat{\eta}$ و η

$$\begin{aligned} R_n = & \left\{ \frac{1}{n} \sum_{i=1}^n f'(\mathbf{X}_i, \eta) - E(f'(\mathbf{X}_i, \eta)) \right\} (\hat{\eta} - \eta) \\ & + E(f'(\mathbf{X}_i, \eta)) \left\{ \hat{\eta} - \eta - \frac{1}{n} \sum_{i=1}^n h(\mathbf{X}_i) \right\} \\ & + \frac{1}{n} (\hat{\eta} - \eta)^T \left\{ \frac{1}{n} \sum_{i=1}^n f''(\mathbf{X}_i, \hat{\eta}^*) \right\} (\hat{\eta} - \eta). \end{aligned}$$

واضح است از رابطه (۳) توزیع نرمال مجانبی آماره T_n بر اساس قضیه حد مرکزی حاصل می‌شود. بنابراین به طور کلی چنانچه بتوان آماره‌ای را به صورت رابطه (۳) نوشت، آنگاه توزیع نرمال مجانبی آن حاصل خواهد شد. به عنوان مثال، **لیو و همکاران (۲۰۲۲)** برای بدست آوردن توزیع مجانبی برآوردگر پارامتر مدل پیشنهادی این قضیه را در قضیه ۲ مقاله‌شان به کار برده‌اند. به عنوان مثال‌های دیگر از به کارگیری از این قضیه، می‌توان به قضایای حدی مطرح شده در **ژانگ (۲۰۱۹)** اشاره کرد.

۳ تابع ژرفای ماهالانوبیس

فرض کنید اندازه احتمال P بر فضای $\mathbb{R}^p$ داده شده و F تابع توزیع متناظر با آن باشد. ژرفای نقطه $\mathbf{x} \in \mathbb{R}^p$ ، اندازه‌ای از میزان ژرف بودن یا دورافتاده بودن این نقطه نسبت به توزیع F است. هر تابع $D(\mathbf{x}, F)$ که از مرکز توزیع به سمت بیرون در تمامی جهات، نقاط $\mathbf{x}$ متعلق به $\mathbb{R}^p$ را مرتب کند، تابع ژرفای متناظر با F نامیده می‌شود (**سرفلینگ، ۲۰۰۶**). بنابراین بر اساس تابع ژرفای مفروض، نقاط p متغیره به نحوی رتبه‌بندی می‌شوند که ژرفای بزرگ‌تر نشان‌دهنده مرکزی‌تر بودن نقاط و ژرفای کوچک‌تر نشان‌دهنده دورافتاده‌تر بودن آنها است. تحت تابع ژرفای مفروض، نقطه یا میانگین نقاط با بیشترین مقدار ژرفا، مرکز را تشکیل می‌دهند. تابع ژرفای نمونه‌ای متناظر با $D(\cdot, F)$ با جایگذاری تابع توزیع تجربی F_n به جای تابع توزیع F به صورت $D(\cdot, F_n)$ حاصل می‌شود. با توجه به اینکه در روش‌های ناپارامتری یک متغیره، رتبه مشاهدات نقش بسیار تعیین کننده‌ای در استنباط‌های آماری دارد،

با استفاده از رتبه‌بندی حاصل شده بر اساس تابع ژرفا بسیاری از این روش‌ها به حالت چندمتغیره تعمیم داده شده‌اند.

تعریف ۱. زو و سرفلینگ (۲۰۰۰) فرض کنید X یک بردار تصادفی p بعدی بر فضای احتمال $(\Omega, \mathcal{F}, P)$ و F تابع توزیع متناظر با اندازه احتمال P باشد. تابع $D(\cdot, F) : \mathbb{R}^p \rightarrow \mathbb{R}$ ، تابع ژرفای آماری نامیده می‌شود، هرگاه در ویژگی‌های زیر صدق کند.

- $P1$: ناوردایی آفین^۱: $D(\mathbf{x}, F)$ مستقل از دستگاه مختصات باشد. یعنی برای هر بردار $\mathbf{x}$ در $\mathbb{R}^p$ ، هر ماتریس A نامفرد $p \times p$ و هر بردار p بعدی $\mathbf{b}$ ، $D(A\mathbf{x} + \mathbf{b}, F_{A\mathbf{x} + \mathbf{b}}) = D(\mathbf{x}, F_X)$.
- $P2$: ماکسیم شدن در مرکز: اگر توزیع F نسبت به نقطه θ متقارن باشد (منظور یکی از مفاهیم تقارن چندمتغیره شامل تقارن کروی، بیضوی، مرکزی و یا زاویه‌ای است)، در این صورت $D(\mathbf{x}, F)$ در θ ماکسیم شود.
- $P3$: به صفر رسیدن در بینهایت: اگر $\|\mathbf{x}\| \rightarrow \infty$ که $\|x'\| = \sqrt{x'x}$ ، آنگاه $D(\mathbf{x}, F) \rightarrow 0$.
- $P4$: یکنوایی نسبت به ژرف‌ترین نقطه: اگر برای هر $\mathbf{x}$ در $\mathbb{R}^p$ ، $D(\theta, F) \geq D(\mathbf{x}, F)$ ، آنگاه برای هر α در $[\cdot, 1]$ ، $D(\theta + \alpha(\mathbf{x} - \theta), F) \geq D(\mathbf{x}, F)$ که در آن θ ژرف‌ترین نقطه است.

تا به حال توابع ژرفاهای مختلفی توسط محققین پیشنهاد شده است، که در ادامه تابع ژرفای ماهالانوبیس **لیو و سینگ (۲۰۰۶)** معرفی می‌شود. فرض کنید μ_F و Σ_F به ترتیب بردار میانگین و ماتریس کوواریانس توزیع F باشند. فاصله ماهالانوبیس بین $\mathbf{x}$ و μ_F به صورت

$$d(\mathbf{x}, \mu_F) = (\mathbf{x} - \mu_F)^T \Sigma_F^{-1} (\mathbf{x} - \mu_F)$$

تعریف می‌شود. بر اساس این فاصله، ژرفای ماهالانوبیس نقطه $\mathbf{x} \in \mathbb{R}^p$ نسبت به توزیع F به صورت

$$MD(\mathbf{x}, F) = \frac{1}{1 + d(\mathbf{x}, \mu_F)} = \frac{1}{1 + (\mathbf{x} - \mu_F)^T \Sigma_F^{-1} (\mathbf{x} - \mu_F)}. \quad (۴)$$

تعریف می‌شود. برای به دست آوردن نسخه نمونه‌ای ژرفای ماهالانوبیس نقطه $\mathbf{x} \in \mathbb{R}^p$ ، کافی است به جای μ_F و Σ_F برآورد نمونه‌ای آنها جایگزین شود، که به طور معمول به ترتیب از بردار میانگین نمونه‌ای و ماتریس کوواریانس نمونه‌ای استفاده می‌شود و با $MD(\mathbf{x}, F_n)$ نشان داده می‌شود.

¹Affine invariance

۵۰ توزیع مجانبی برخی آمارها در استنباط چندمتغیره

۴ توزیع مجانبی میانگین ژرفای ماهالانوبیس تجربی

قضیه ۱: فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی از بردار تصادفی پیوسته X با گشتاورهای مرتبه چهارم متناهی باشد. تعریف کنید $\theta = (\mu^\top, V_\Sigma^\top)^\top$, $h(X_i, \theta) = ((X_i - \mu)^\top, V_A^\top)^\top$ و

$$Q(X_i, \theta) = M(X_i, F) - E(M(X_i, F)) + E\left(\frac{\partial M(X_i, F)}{\partial \theta}\right)^\top h(X_i, \theta)$$

که در آن $A = (X_i - \mu)(X_i - \mu)^\top - \Sigma$ و $V_A = \text{vec}(A)$ از زیر هم قرار دادن ستون‌های ماتریس A از چپ به راست حاصل شده است. آنگاه برای $V = \text{Var}(Q(X_i, \theta))$

$$W_n = \sqrt{\frac{n}{V}} \left(\frac{1}{n} \sum_{i=1}^n MD(X_i, F_n) - E(MD(X_1, F)) \right) \xrightarrow{d} N(0, 1). \quad (5)$$

برهان: فرض کنید $\theta = (\mu^\top, V_\Sigma^\top)^\top$ و $\hat{\theta}_n = (\bar{X}^\top, V_{S_n}^\top)^\top$ که در آن $\bar{X}$ و S_n به ترتیب بردار میانگین نمونه‌ای و ماتریس کواریانس نمونه‌ای هستند. رابطه $(\bar{X} - \mu) = o_p(1)$ ، به راحتی از نامساوی چیبی-چف نتیجه می‌شود. بنابراین از لم اسلاتسکی (سرفلینگ، ۱۹۸۰) می‌توان رابطه $\sqrt{n}(\bar{X} - \mu)(\bar{X} - \mu)^\top = o_p(1)$ را نتیجه گرفت.

$$\begin{aligned} S_n - \Sigma &= \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)(X_i - \mu)^\top - \Sigma \right\} - (\bar{X} - \mu)(\bar{X} - \mu)^\top \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ (X_i - \mu)(X_i - \mu)^\top - \Sigma \right\} + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

تعریف کنید $h(X_i, \theta) = ((X_i - \mu)^\top, V_A^\top)^\top$ که در آن $A = (X_i - \mu)(X_i - \mu)^\top - \Sigma$. بنابراین

$$\hat{\theta}_n - \theta = \frac{1}{n} \sum_{i=1}^n h(X_i, \theta) + R_n, \quad (6)$$

که در آن $E(h(X_i, \theta)) = 0$ ، $\text{Cov}(h(X_i, \theta)) = V$ متناهی و $\sqrt{n}R_n = o_p(1)$ است. بنابراین هر مولفه برآوردگر $\hat{\theta}_n$ نمایش تقریبی به صورت میانگین بردارهای تصادفی مستقل دارد.

در ادامه برای فهم بهتر روند اثبات، $MD(x, F)$ و $MD(x, F_n)$ به ترتیب با $MD(x, \theta)$ و $MD(x, \hat{\theta}_n)$

نمایش داده می‌شوند. بر اساس بسط تیلور حول θ ,

$$MD(\mathbf{x}, \hat{\theta}_n) = MD(\mathbf{x}, \theta) + \left(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} \right)^\top (\hat{\theta}_n - \theta) + \frac{1}{2} (\hat{\theta}_n - \theta)^\top \frac{\partial M(\mathbf{X}_i, \theta^*)}{\partial \theta^* \partial \theta^{*\top}} (\hat{\theta}_n - \theta), \quad (7)$$

که در آن $\theta^* = (\mu^{*\top}, \mathbf{V}_\Sigma^{*\top})^\top$ بین θ و $\hat{\theta}_n$ است. بنابراین برای $i = 1, \dots, n$ و $j, k = 1, \dots, p$

$$\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} = M^\top(\mathbf{X}_i, \theta) (\mathbf{G}_i^\top, K_{i,11}, K_{i,12}, \dots, K_{i,pp}), \quad (8)$$

که در آن $\mathbf{G}_i = 2(\mathbf{X}_i - \mu)^\top \Sigma^{-1}(\mathbf{X}_i - \mu)$, $K_{i,jk} = (\mathbf{X}_i - \mu)^\top \Sigma^{-1} \mathbf{B}_{jk} \Sigma^{-1}(\mathbf{X}_i - \mu)$ و $\mathbf{B}_{jk}$ یک ماتریس $p \times p$ است که درایه‌های (j, k) و (k, j) آن ۱ و مابقی صفر هستند. $E(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta})$ متناهی است. بنابراین با به‌کارگیری قانون ضعیف اعداد بزرگ رابطه

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} \right)^\top (\hat{\theta}_n - \theta) &= \left[\frac{1}{n} \sum_{i=1}^n \frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} - E\left(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} \right) \right]^\top (\hat{\theta}_n - \theta) \\ &+ E\left(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} \right)^\top \left[(\hat{\theta}_n - \theta) - \frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{X}_i, \theta) \right] \\ &+ E\left(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} \right)^\top \left(\frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{X}_i, \theta) \right) \\ &= E\left(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta} \right)^\top \left(\frac{1}{n} \sum_{i=1}^n \mathbf{h}(\mathbf{X}_i, \theta) \right) + o_p(1/\sqrt{n}). \end{aligned} \quad (9)$$

نتیجه می‌شود. بنابر قانون ضعیف اعداد بزرگ برای $\partial M(\mathbf{X}_i, \theta)/\partial \theta$ برای $\sum_{i=1}^n n^{-1}$ و از (۶) دو عبارت اول سمت راست معادله اول $o_p(n^{-1/2})$ هستند. همچنین اگر $\Sigma^* = (\sigma_{ij}^*)_{p \times p}$ ، رابطه

$$\frac{\partial M(\mathbf{X}_i, \theta^*)}{\partial \theta^* \partial \theta^{*\top}} = M^\top(\mathbf{X}_i, \theta) \begin{bmatrix} \mathbf{U}_i^{(1)} & \mathbf{U}_i^{(\top)T} \\ \mathbf{U}_i^{(\top)} & \mathbf{U}_i^{(\top)} \end{bmatrix},$$

برقرار است، که در آن

$$\mathbf{U}_i^{(1)} = -2\Sigma^{*-1} + 4M(\mathbf{X}_i, \theta^*) \Sigma^{*-1} (\mathbf{X}_i + \mu^*) (\mathbf{X}_i + \mu^*)^\top \Sigma^{*-1},$$

۵۲ توزیع مجانبی برخی آمارها در استنباط چندمتغیره

و $\mathbf{U}_i^{(\mathfrak{r})} = (\mathbf{S}_{i,(\mathfrak{I},\mathfrak{I})}^\top, \mathbf{S}_{i,(\mathfrak{I},\mathfrak{r})}^\top, \dots, \mathbf{S}_{i,(p,p)}^\top)^\top$ که برای $j, k = \mathfrak{I}, \dots, p$,

$$\begin{aligned} \mathbf{S}_{i,(j,k)} &= \mathfrak{r}(\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{kj} \boldsymbol{\Sigma}^{*- \mathfrak{I}} - \mathfrak{r}M(\mathbf{X}_i, \boldsymbol{\theta}^*) \\ &\times [(\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{kj} \boldsymbol{\Sigma}^{*- \mathfrak{I}} (\mathbf{X}_i + \boldsymbol{\mu}^*)] (\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}}, \end{aligned}$$

و $\mathbf{U}_i^{(\mathfrak{r})} = \left(\frac{\partial M(\mathbf{X}_i, \boldsymbol{\theta}^*)}{\partial \sigma_{ij}^* \partial \sigma_{lf}^*} \right)_{p^\top \times p^\top}$ به طوری که برای $j, k, l, f = \mathfrak{I}, \dots, p$

$$\begin{aligned} \frac{\partial M(\mathbf{X}_i, \boldsymbol{\theta}^*)}{\partial \sigma_{ij}^* \partial \sigma_{lf}^*} &= -(\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \left(\boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{lf} \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{jk} \boldsymbol{\Sigma}^{*- \mathfrak{I}} + \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{jk} \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{lf} \boldsymbol{\Sigma}^{*- \mathfrak{I}} \right) (\mathbf{X}_i + \boldsymbol{\mu}^*) \\ &+ \mathfrak{r}M(\mathbf{X}_i, \boldsymbol{\theta}^*) \left((\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{lf} \boldsymbol{\Sigma}^{*- \mathfrak{I}} (\mathbf{X}_i + \boldsymbol{\mu}^*) \right) \left((\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{jk} \boldsymbol{\Sigma}^{*- \mathfrak{I}} (\mathbf{X}_i + \boldsymbol{\mu}^*) \right). \end{aligned}$$

بنابراین

$$\begin{aligned} \left| \frac{\partial M(\mathbf{X}_i, \boldsymbol{\theta}^*)}{\partial \sigma_{ij}^* \partial \sigma_{lf}^*} \right| &\leq |(\mathbf{X}_i + \boldsymbol{\mu}^*)^\top (\boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{lf} \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{jk} \boldsymbol{\Sigma}^{*- \mathfrak{I}} + \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{jk} \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{lf} \boldsymbol{\Sigma}^{*- \mathfrak{I}}) (\mathbf{X}_i + \boldsymbol{\mu}^*)| \\ &+ \mathfrak{r}M(\mathbf{X}_i, \boldsymbol{\theta}^*) |((\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{lf} \boldsymbol{\Sigma}^{*- \mathfrak{I}} (\mathbf{X}_i + \boldsymbol{\mu}^*)) ((\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} \mathbf{B}_{jk} \boldsymbol{\Sigma}^{*- \mathfrak{I}} (\mathbf{X}_i + \boldsymbol{\mu}^*))| \\ &\leq \mathfrak{r} \lambda_{jk} \lambda_{lf} \left\{ |(\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} (\mathbf{X}_i + \boldsymbol{\mu}^*)| + M(\mathbf{X}_i, \boldsymbol{\theta}^*) ((\mathbf{X}_i + \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*- \mathfrak{I}} (\mathbf{X}_i + \boldsymbol{\mu}^*))^\mathfrak{r} \right\}, \end{aligned}$$

که λ_{ab} بزرگترین ویژه مقدار ماتریس $\boldsymbol{\Sigma}^{*- \mathfrak{I}/\mathfrak{r}} \mathbf{B}_{ab} \boldsymbol{\Sigma}^{*- \mathfrak{I}/\mathfrak{r}}$ است. بنابراین بر اساس [سرفلینگ \(۱۹۸۰\)](#)

$$n^{- \mathfrak{I}} \sum_{i=\mathfrak{I}}^n \partial M(\mathbf{X}_i, \boldsymbol{\theta}^*) / \partial \boldsymbol{\theta}^* \partial \boldsymbol{\theta}^{*\top} = O_p(\mathfrak{I})$$

نتیجه می شود. این رابطه همراه با رابطه (۶)، رابطه

$$\frac{\mathfrak{I}}{\mathfrak{r}} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^\top \frac{\mathfrak{I}}{n} \sum_{i=\mathfrak{I}}^n \frac{\partial M(\mathbf{X}_i, \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^* \partial \boldsymbol{\theta}^{*\top}} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = o_p\left(\frac{\mathfrak{I}}{\sqrt{n}}\right), \quad (۱۰)$$

را نتیجه می دهد. معادله (۷) و نتایج (۹) و (۱۰) بر

$$\frac{\mathfrak{I}}{n} \sum_{i=\mathfrak{I}}^n M(\mathbf{X}_i, \hat{\boldsymbol{\theta}}) = \frac{\mathfrak{I}}{n} \sum_{i=\mathfrak{I}}^n M(\mathbf{X}_i, \boldsymbol{\theta}) + \mathbf{E} \left(\frac{\partial M(\mathbf{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \left(\frac{\mathfrak{I}}{n} \sum_{i=\mathfrak{I}}^n h(\mathbf{X}_i, \boldsymbol{\theta}) \right) + o_p\left(\frac{\mathfrak{I}}{\sqrt{n}}\right)$$

دلالت می‌کنند. بنابراین

$$\frac{1}{n} \sum_{i=1}^n MD(\mathbf{X}_i, F_n) - E(M(\mathbf{X}_i, \theta)) = \frac{1}{n} \sum_{i=1}^n Q(\mathbf{X}_i, \theta) + o_p(1/\sqrt{n})$$

که در آن

$$Q(\mathbf{X}_i, \theta) = M(\mathbf{X}_i, \theta) - E(M(\mathbf{X}_i, \theta)) + E\left(\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta}\right)^T \mathbf{h}(\mathbf{X}_i, \theta).$$

بنابراین، چنانچه $V = \text{var}(Q(\mathbf{X}_i, \theta))$ آنگاه از قضیه اسلاتسکی (سرفلینگ، ۱۹۸۰) و قضیه حد مرکزی نتیجه مورد نظر حاصل می‌شود. توزیع مجانبی بدست آمده برای آماره W_n این امکان را فراهم می‌کند که بر اساس این آماره به انجام آزمون فرض برای اختلاف مقیاس بین دو توزیع چندمتغیره پرداخت. فرض کنید نمونه‌های تصادفی $\mathbf{X}_1, \dots, \mathbf{X}_n$ و $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ به ترتیب از توزیع‌های F و G در R^p به دست آمده‌اند که توزیع‌های F و G به جز در مقیاس در سایر ویژگی‌ها یکسان‌اند. هدف انجام آزمون فرضیه‌های

$$\begin{cases} H_0 : & \text{دو توزیع یکسان هستند} \\ H_a : & \text{دو توزیع اختلاف مقیاس دارند} \end{cases}$$

است. اگر بین F و G اختلاف مقیاس وجود داشته باشد، به طور مثال G دارای مقیاس بزرگتری از F باشد، در نمونه ترکیب شده با احتمال بیشتری $\mathbf{X}_i$ ‌ها در اطراف مرکز و $\mathbf{Y}_i$ ‌ها در موقعیت دورافتاده‌تری نسبت به مرکز قرار می‌گیرند. از آن‌جا که ژرفای داده‌ها، موقعیت از مرکز به بیرون داده‌ها را نشان می‌دهد، برای نشان دادن این اختلاف مقیاس بسیار مفید خواهد بود. ابتدا عناصر نمونه ترکیب شده و $Z_{n+m} = \{\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{Y}_1, \dots, \mathbf{Y}_m\}$ تشکیل می‌شود. حال ژرفای $\mathbf{X}_1, \dots, \mathbf{X}_n$ نسبت به نمونه ترکیب شده Z_{n+m} محاسبه می‌شود. اگر H_a درست باشد، آنگاه $\mathbf{X}_i$ ‌ها نسبت به $\mathbf{Y}_i$ ‌ها دورافتاده‌تر یا مرکزی‌تر هستند، پس نسبت به $\mathbf{Y}_i$ ‌ها دارای ژرفای کوچکتر یا بزرگتری هستند. بنابراین آماره W_n بر اساس میانگین ژرفای $\mathbf{X}_1, \dots, \mathbf{X}_n$ نسبت به نمونه ترکیب شده Z_{n+m} محاسبه می‌شود و مقادیر بزرگ و یا کوچک آن نشان‌دهنده رد فرض صفر هستند. به بیان دقیق‌تر، هرگاه $|nW_n/(n+m)|$ بزرگتر از $z_{1-\frac{\alpha}{2}}$ باشد، فرض H_0 در سطح α رد می‌شود.

۵ مطالعه شبیه سازی

در این بخش، مطالعات شبیه سازی به منظور بررسی همگرایی

$$W_n = \sqrt{\frac{n}{V}} \left(\frac{1}{n} \sum_{i=1}^n MD(\mathbf{X}_i, F_n) - E(MD(\mathbf{X}_1, F)) \right) \xrightarrow{d} N(0, 1), \quad (11)$$

که در قضیه ۱ بخش ۴ اثبات شد و همچنین نحوه به کارگیری آن ارائه می شود. ابتدا نشان می دهیم تابع ژرفای ماهالانوبیس ناوردای آفین است. فرض کنید $\mu = E(\mathbf{X}_i)$ و $\Sigma = Cov(\mathbf{X}_i)$. تعریف کنید

$$\mathbf{Z}_i = \Sigma^{-\frac{1}{2}} (\mathbf{X}_i - \mu), \quad i = 1, \dots, n. \quad (12)$$

واضح است $E(\mathbf{Z}_i) = 0$ و $Cov(\mathbf{Z}_i) = \mathbf{I}_p$ که یک ماتریس همبستگی $p \times p$ است. حال رابطه

$$\begin{aligned} MD(\mathbf{X}_i, F_n) &= \frac{1}{1 + (\mathbf{X}_i - \bar{\mathbf{X}})^T \mathbf{S}_n^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})} \\ &= \frac{1}{1 + \left(\Sigma^{\frac{1}{2}} \mathbf{Z}_i + \mu - \Sigma^{\frac{1}{2}} \bar{\mathbf{Z}} - \mu \right)^T \mathbf{S}_n^{-1} \left(\Sigma^{\frac{1}{2}} (\mathbf{X}_i - \mu) - \Sigma^{\frac{1}{2}} (\bar{\mathbf{X}} - \mu) \right)} \\ &= \frac{1}{1 + (\mathbf{Z}_i - \bar{\mathbf{Z}})^T \mathbf{S}_Z^{-1} (\mathbf{Z}_i - \bar{\mathbf{Z}})} \end{aligned}$$

حاصل می شود که نشان می دهد مقدار ژرفای ماهالانوبیس تحت تبدیل (۱۲) تغییر نمی کند و ناوردای آفین است. بنابراین می توان توزیع مجانبی مجموع $MD(\mathbf{X}_i, F_n)$ را تحت فرض $\mu = 0$ و $\Sigma = \mathbf{I}_p$ بدست آورد و به سایر حالات تعمیم داد.

فرض کنید $\mathbf{X}_1, \dots, \mathbf{X}_n$ یک نمونه تصادفی از توزیع نرمال دومتغیره استاندارد است. تابع $h(\mathbf{X}_i, \theta)$ در رابطه (۲) به صورت $h(\mathbf{X}_i, \theta) = \left(\mathbf{X}_i^T, \mathbf{G}_{i1}^T, \dots, \mathbf{G}_{ip}^T \right)^T$ حاصل می شود، که در آن برای $j = 1, \dots, p$ و $\mathbf{G}_{ij} = (g_{ij1}, \dots, g_{ijp})^T$

$$g_{ijk} = \begin{cases} X_{ij}^2 - 1 & j = k \\ X_{ij} X_{ik} & j \neq k, \end{cases}$$

و $E(h(\mathbf{X}_i, \theta)) = 0$. همچنین مولفه‌های عبارت $\frac{\partial M(\mathbf{X}_i, \theta)}{\partial \theta}$ در رابطه (۸) به صورت

$$M(\mathbf{X}_i, \theta) = (\mathbf{I} + \mathbf{X}_i^T \mathbf{X}_i)^{-1}, \quad \mathbf{G}_i = -\mathbf{X}_i, \quad K_{i,jk} = \begin{cases} X_{ij}^2 & j = k \\ 2X_{ij}X_{ik} & j \neq k \end{cases}$$

تعریف می‌شوند. بنابراین بر اساس روابط بالا، برای بررسی توزیع مجانبی W_n ، مقادیر $E(M(\mathbf{X}_i, F))$ و $Var(Q(\mathbf{X}_i, \theta)) = E(Q(\mathbf{X}_i, \theta)^2)$ با استفاده از روش‌های عددی تقریب زده می‌شوند. برای حل انتگرال‌ها از تابع adaptIntegrate بسته cubature در نرم‌افزار R استفاده شده است. نتایج برای مقادیر مختلف p در جدول ۱ آمده است.

جدول ۱. مقادیر حاصل از روش‌های عددی		
V	$E(M(\mathbf{X}_i, \theta))$	p
۰/۰۱۹۴	۰/۴۶۱۵	۲
۰/۰۱۳۴	۰/۳۴۴۳	۳
۰/۰۰۸۶	۰/۲۶۹۳	۴
۰/۰۰۵۴	۰/۲۱۸۶	۵
۰/۰۰۳۴	۰/۱۸۲۷	۶
۰/۰۰۲۲	۰/۱۵۶۳	۷
۰/۰۰۱۵	۰/۱۳۶۲	۸
۰/۰۰۰۹	۰/۱۲۰۵	۹
۰/۰۰۰۷	۰/۱۰۸۰	۱۰

نمونه‌های تصادفی با اندازه‌های ۱۰۰، ۱۵۰ و ۲۰۰ از توزیع $N_p(\mu, \Sigma)$ ، تولید می‌شوند. برای p دو مقدار ۲ و ۵ در نظر گرفته می‌شود. برای بررسی عدم وابستگی توزیع مجانبی W_n به مقادیر μ و Σ برای μ دو بردار 0 و $\mathbf{I}_p$ که به ترتیب نشان‌دهنده بردارهای p -بعدی با همه مولفه‌ها برابر با صفر و یک هستند و همچنین برای ماتریس کوواریانس دو ماتریس $\mathbf{I}_p$ و $\Sigma^* = 0.5\mathbf{I}_p + 0.5\mathbf{J}_p$ که $\mathbf{J}_p$ نشان‌دهنده ماتریسی $p \times p$ است که همه مولفه‌های آن برابر با ۱ هستند، در نظر گرفته می‌شوند.

در هر حالت میانگین ژرفای ماهالانوبیس نمونه‌ای حاصل شده و بر اساس نتایج جدول ۱، استاندارد می‌شود. این فرایند ۱۰۰۰ بار تکرار می‌شود، سپس با استفاده از آزمون شاپیرو-ویلک و کولموگروف-اسمیرنوف، نرمال بودن این مقادیر استاندارد شده، آزمون می‌شود. نتایج برای ۱۰۰، ۱۵۰ و ۲۰۰ در جدول ۲ ارائه شده‌اند. همان‌گونه که مشاهده می‌شود برای $n = ۱۰۰$ ، در بیشتر حالات فرض نرمال بودن مشاهدات تایید نمی‌شود، برای $n = ۱۵۰$ ، فرض نرمال بودن مشاهدات هرگاه $p = ۲$ تایید می‌شود و برای $p = ۵$ در بیشتر حالات پذیرفته نمی‌شود. برای $n = ۲۰۰$ در همه حالات و بر اساس هر دو آماره شاپیرو-ویلک و کولموگروف-اسمیرنوف فرض نرمال بودن داده‌های تولید شده از W_n تایید می‌شود. بنابراین می‌توان $n = ۲۰۰$ را مقدار مناسبی برای به‌کارگیری توزیع مجانبی در نظر

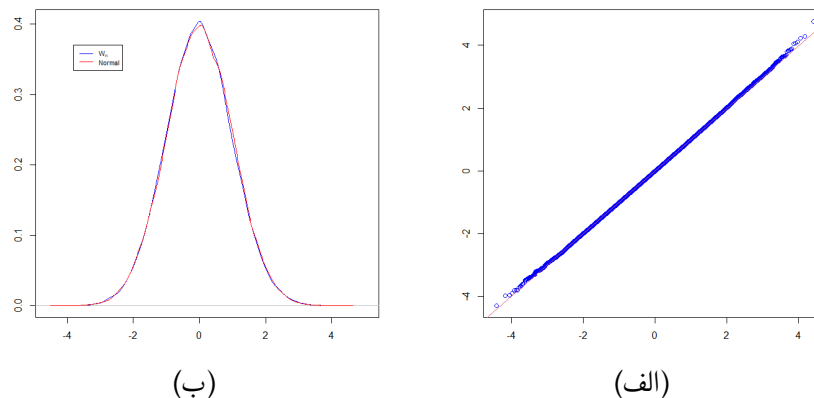
جدول ۲. مقدار آماره آزمون و p -مقدار آزمون‌های نرمال بودن داده‌های تولید شده از W_n

کولموگروف-اسمیرنوف		شاپیرو-ویلک		Σ^*	μ	p	n
مقدار $-p$	آماره آزمون	مقدار $-p$	آماره آزمون				
۰/۰۰۳۲	۰/۰۲۵۴	۰/۰۰۵۶	۰/۹۹۹۰	I_2	0	۲	۱۰۰
۰/۰۰۸۴۱	۰/۰۱۷۸	۰/۰۰۹۵	۰/۹۹۹۱	Σ^*	0		
۰/۲۷۳۲	۰/۰۱۴۱	۰/۰۳۷۸	۰/۹۹۹۳	I_2	1 ₂		
۰/۱۳۱۴	۰/۰۱۶۵	۰/۰۰۰۲	۰/۹۹۸۶	Σ^*	1 ₂		
$۲/۲ \times ۱۰^{-۱۶}$	۰/۰۶۴۶	$۱/۲۴۱ \times ۱۰^{-۹}$	۰/۹۹۶۵	I_2	0	۵	
$۲/۲ \times ۱۰^{-۱۶}$	۰/۰۷۳۸	$۶/۳۱۸ \times ۱۰^{-۱۱}$	۰/۹۹۵۷	Σ^*	0		
$۲/۲ \times ۱۰^{-۱۶}$	۰/۰۶۵۶	$۱/۳۹۶ \times ۱۰^{-۶}$	۰/۹۹۶۵	I_2	۱ ₂		
$۲/۲ \times ۱۰^{-۱۶}$	۰/۰۶۸۶	$۴/۳۹۷ \times ۱۰^{-۱۰}$	۰/۹۹۶۲	Σ^*	۱ ₂		
۰/۳۳۰۵	۰/۰۱۳۴	۰/۱۴۸۴	۰/۹۹۹۵	I_2	0	۲	۱۵۰
۰/۴۷۸۴	۰/۰۱۱۹	۰/۰۱۱۲	۰/۹۹۹۱	Σ^*	0		
۰/۱۰۰۰۳	۰/۰۱۷۳	۰/۰۶۶۲	۰/۹۹۹۸	I_2	1 ₂		
۰/۸۲۳۲	۰/۰۰۸۹	۰/۶۵۴۳	۰/۹۹۹۷	Σ^*	1 ₂		
۰/۰۱۵۱	۰/۰۲۲۱	۰/۲۶۹۳	۰/۹۹۹۵	I_2	0	۵	
۰/۰۸۷۲	۰/۰۱۷۷	۰/۰۰۰۶	۰/۹۹۸۷	Σ^*	0		
۰/۰۱۶۵	۰/۰۲۱۹	۰/۰۸۸۹	۰/۹۹۹۴	I_2	۱ ₂		
۰/۰۰۵۲	۰/۰۲۴۴	۰/۰۶۷۱	۰/۹۹۹۴	Σ^*	۱ ₂		
۰/۵۰۰۴	۰/۰۳۷۰	۰/۴۸۳۱	۰/۹۹۸۴	I_2	0	۲	۲۰۰
۰/۴۰۰۵	۰/۰۴۰۰	۰/۳۷۷۰	۰/۹۹۸۲	Σ^*	0		
۰/۶۰۹۹	۰/۰۳۴۰	۰/۳۸۵۶	۰/۹۹۸۲	I_2	1 ₂		
۰/۴۳۲۴	۰/۰۳۹۰	۰/۶۶۱۱	۰/۹۹۸۷	Σ^*	1 ₂		
۰/۲۲۰۲	۰/۰۲۱۰	۰/۱۵۸۰	۰/۹۹۹۵	I_2	0	۵	
۰/۶۴۴۰	۰/۰۱۴۸	۰/۱۶۷۱	۰/۹۹۹۵	Σ^*	0		
۰/۵۱۲۰	۰/۰۱۶۴	۰/۲۴۲۴	۰/۹۹۹۵	I_2	۱ ₂		
۰/۲۴۹۲	۰/۰۴۰۴	۰/۱۲۵۴	۰/۹۹۹۴	Σ^*	۱ ₂		

گرفت. البته با افزایش بعد مشاهدات ممکن است به تعداد نمونه بیشتری نیاز باشد. در حالت توزیع نرمال استاندارد دومتغیره، نمودار تابع چگالی و چندک-چندک برای $n = ۲۰۰$ در شکل ۱ رسم شده است که همچنان نرمال بودن مقادیر شبیه‌سازی شده از متغیر W_n تایید می‌شود.

۶ کاربرد

چهار متغیر طول کاسبرگ، عرض کاسبرگ، طول گلبرگ و پهنای گلبرگ بر حسب سانتی‌متر برای ۵۰ گل از سه گونه گل زنبق شامل setosa، versicolor و virginica ثبت شده است. این مجموعه داده در بسته datasets در نرم افزار R تحت عنوان iris قابل دسترسی است. تحلیل‌های مختلفی بر روی این داده‌ها توسط محققین انجام شده است که از آن جمله می‌توان به پوار و شیرک (۲۰۲۲) اشاره کرد، که اختلاف مقیاس بین گونه‌ها را مورد بررسی قرار داده‌اند. برای آزمون اختلاف مقیاس این سه گونه ابتدا میانگین هر گونه محاسبه می‌شود و مشاهدات بر اساس آن مرکزی می‌شوند تا دارای مکان یکسان باشند. از نمودار پراکندگی متغیرهای مرکزی شده در شکل ۲، به نظر می‌رسد



شکل ۱. نمودار الف-چندک-چندک و ب- تابع چگالی تجربی متغیر W_n .

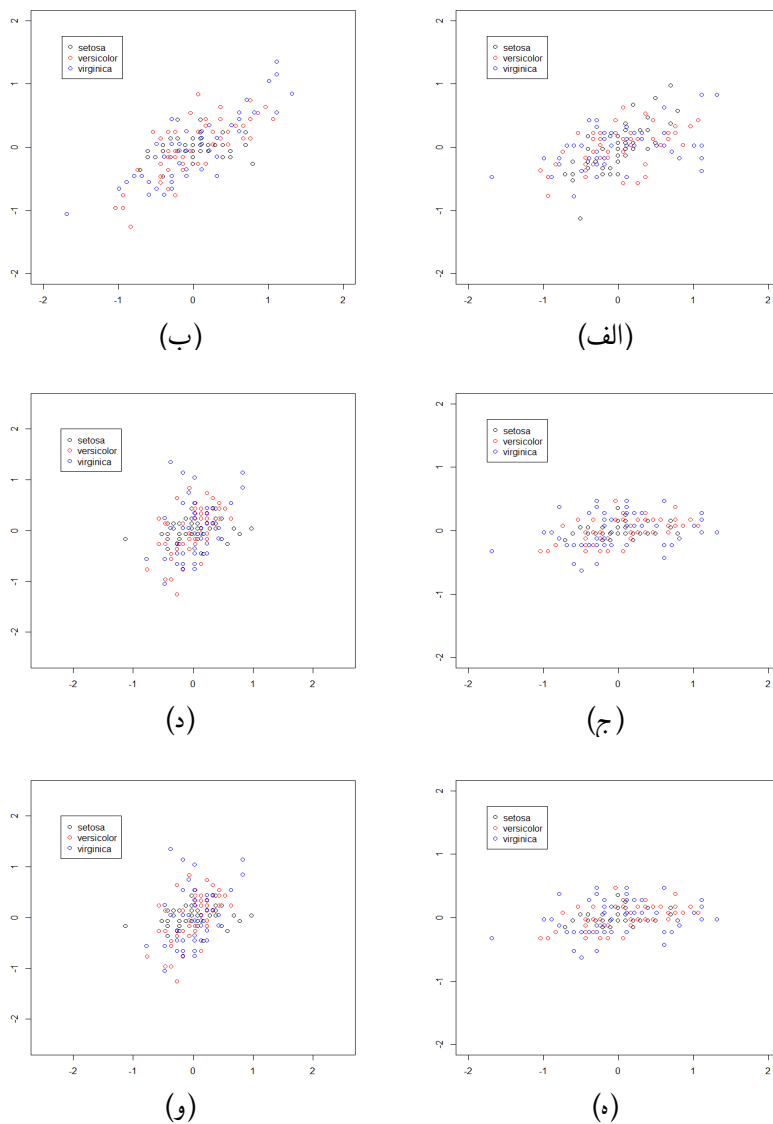
گونه‌های مختلف دارای اختلاف مقیاس هستند. حال آماره پیشنهادی را برای مقایسه گونه‌های مختلف به کار می‌بریم. نتایج در جدول ۳ ارائه شده است. p -مقدارهای بدست آمده نشان می‌دهند که در سطح ۱٪ اختلاف مقیاس بین گونه‌ها معنی‌دار است. لازم به ذکر است، پوار و شیریک (۲۰۲۲) نیز نشان دادند اختلاف مقیاس بین گونه‌ها معنی‌دار است.

جدول ۳. مقدار آماره آزمون و p -مقدار آماره پیشنهادی برای داده‌های گونه‌های گل زنبق

آزمون	مقدار آماره W_n	p -مقدار
(setosa, versicolor)	۴۱۰۹۸	۳۹۵×۱۰^{-۵}
(setosa, virginica)	۷۰۷۶۳	$۱/۴۸ \times ۱۰^{-۱۲}$
(versicolor, virginica)	۲۱۰۶۵	$۳/۵۲ \times ۱۰^{-۴}$

۷ بحث و نتیجه‌گیری

در این مقاله ابتدا قضیه‌ای که بر اساس بسط تیلور راه‌حلی برای بدست آوردن توزیع مجانبی رده خاصی از آماره‌ها فراهم می‌کند، بیان شد. سپس، توزیع مجانبی میانگین ژرفای ماهالانوبیس یک نمونه از بردارهای تصادفی اثبات شد. بر اساس توزیع مجانبی بدست آمده، نشان داده شد که می‌توان یک آماره برای آزمون اختلاف مقیاس بین دو توزیع چندمتغیره تعریف کرد. در ادامه با استفاده از روش‌های عددی میانگین و واریانس این آماره تقریب زده شد و دقت توزیع مجانبی ارزیابی شد. در پایان آماره معرفی شده برای تحلیل اختلاف مقیاس بین سه گونه از گل زنبق به کار رفت.



شکل ۲. نمودار پراکنده‌ی طول و عرض کاسبرگ، طول و پهنای گلبرگ ۵۰ گل از گونه‌های setosa، versicolor و virginica ، الف- (طول کاسبرگ، عرض کاسبرگ)، ب- (طول کاسبرگ، طول گلبرگ)، ج- (طول کاسبرگ، پهنای گلبرگ)، د- (عرض کاسبرگ، طول گلبرگ)، ه- (عرض کاسبرگ، پهنای گلبرگ)، و- (طول گلبرگ، پهنای گلبرگ).

تقدیر و تشکر

نویسنده از سردبیر، داوران و ویراستار محترم مجله که نظرات ارزشمندشان باعث بهبود مطالب ارائه شده در این مقاله گردید، کمال تشکر و قدردانی را دارد.

مراجع

دهقان، س. و فریدروحانی، م. (۱۴۰۰)، شناسائی نقاط دورافتاده چندمتغیره بر اساس تابع دورافتادگی ژرفا-مبنا، مجله علوم آماری، ۱۵(۲)، ۴۴۳-۴۶۲.

Boos, D. D. and Stefanski, L. A. (2013), *Essential Statistical Inference: Theory and Methods*, Vol. 120, Berlin: Springer Science & Business Media.

Dehghan, S. and Faridrohani, M. R. (2019), Affine Invariant Depth-Based Tests for the Multivariate One-Sample Location Problem. *Test*, **28**(3), 671-693.

Dehghan, S. and Faridrohani, M. R. (2021), Nonparametric Depth-Based Tests for the Multivariate Location Problem, *Australian & New Zealand Journal of Statistics*, **63**(2), 309-330.

Dehghan, S., Faridrohani, M. R., and Barzegar, Z. (2022), Testing for Diagonal Symmetry Based on Center-Outward Ranking, *Statistical Papers*, 1-29.

Li, J. and Liu R. Y. (2004), New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth, *Statistical Science*, **19**(4), 686-696.

Liu, R. Y. and Singh, K. (1993), A Quality Index Based on Data Depth and Multivariate Rank Tests, *Journal of the American Statistical Association*, **88**(421), 252-260.

Liu R. Y. and Singh K. (2006), Rank Tests for Multivariate Scale Difference Based on Data Depth, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, **72**(17).

Liu, S., Zhang, Y., Golm, G. T., & Yang, S. (2022), Robust Analyses for Longitudinal Clinical Trials with Missing and Non-Normal Continuous Outcomes, arXiv preprint arXiv, 2203.10561.

Pawar, S. D., & Shirke, D. T. (2022), Data Depth-Based Nonparametric Tests for Multivariate Scales, *Journal of Statistical Theory and Practice*, **16**(1), 1-21.

- Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, Hoboken: John Wiley & Sons.
- Serfling, R. (2006), Depth Functions in Nonparametric Multivariate Inference, *DI-MACS Series in Discrete Mathematics and Theoretical Computer Science*, **72**, 1-16.
- Tukey, J.W. (1975), Mathematics and the Picturing of Data. *In Proceedings of the International Congress of Mathematicians, Vancouver*, **2**, 523–531.
- Zhang, Di. (2019), Inference on Win Ratio for Clustered Semi-Competing Risk Data, *PhD Thesis, University of Pittsburgh*.
- Zuo Y, and Serfling R. (2000), General Notions of Statistical Depth Function, *Annals of Statistics*, **28**(2), 461-482.