نمودار کنترل ناپارامتری چندمتغیره من-ویتنی

فائقه امیری [۱] و منوچهر خردمندنیا [۲]

چکیده:

بسیاری از نمودارهای کنترل چندمتغیره بر اساس فرض نرمال چندمتغیره استوار هستند. اما واقعیت این است که چنین اطلاعاتی به ندرت وجود دارد و در عمل ممکن است که فرض‌های توزیعی، به طور دقیق برقرار نباشند. در چنین شرایطی، عملکرد نمودار کنترل پارامتری ضعیف و غیرقابل اطمینان است. برای غلبه بر این موضوع، نمودارهای کنترل ناپارامتری یا توزیع آزاد پیشنهاد می‌شود. در این مقاله، بر اساس آزمون چندمتغیره دو نمونه‌ای من‌ویتنی، یک نمودار کنترل چندمتغیره ناپارامتری معرفی نموده عملکرد آن را با نمودار سنتی T^2 هتلینگ بر اساس داده‌های شبیه‌سازی شده مقایسه می‌کنیم و نشان می‌دهیم موقعیت‌هایی وجود دارد که در آن نمودار کنترل ناپارامتری چندمتغیره من-ویتنی دارای عملکرد بهتری در مقایسه با نمودار کنترل T^2 هتلینگ است.

واژه‌های کلیدی: آزمون مکان چندمتغیره، کنترل فرآیند چندمتغیره، نمودارکنترل T^2 هتلینگ.

۱ مقدمه

یکی از مهمترین ابزارهای کنترل فرآیند آماری، نمودار کنترل است. در نمودار کنترل اگر مقادیر آماره رسم شده داخل حدود کنترل قرار گیرد فرآیند تحت کنترل و در غیر این صورت فرآیند خارج از کنترل تلقی می‌شود. از این‌رو بر پایه مقدار آماره و ناحیه تصمیم، یک قاعده تصمیم‌گیری رد یا قبول وجود دارد که این ساختار بسیار شبیه ساختار آزمون فرض آماری است. در برخی مواقع لازم است که دو یا چند متغیر همبسته تواماً بررسی شوند. کنترل این متغیرها به طور مستقل از یکدیگر می‌تواند گمراه کننده باشد. در این مواقع روش‌های متعددی برای کنترل کیفیت وجود دارد که مبحث کنترل کیفیت چندمتغیره را تشکیل می‌دهند. از جمله مزایای نمودار کنترل ناپارامتری می‌توان به سادگی، عدم نیاز به فرض توزیع پارامتری خاص و کارایی بیشتر در کشف تغییرات، هنگامی که توزیع واقعی به طور محسوسی غیرنرمال است، اشاره کرد.

در متون کنترل فرآیند آماری چندین نمودار کنترل ناپارامتری یک و چندمتغیره معرفی شده است. یک مطالعه مروری راجع به نمودارهای یک‌متغیره توسط چاکرابورتی و همکاران [٤] و در سال‌های اخیر توسط چاکرابورتی و گراهام [۳] ارائه شده است. در حالت چندمتغیره نیز تاکنون چند نمودار کنترل ناپارامتری معرفی شده است.

دو مورد از نمودارهای کنترل ناپارامتری چندمتغیره توسط بوون و چاکرابورتی [۲] معرفی گردیده است. هر دو نموداری که آن‌ها معرفی کرده‌اند از نوع شوهارت است. نمودار اول با نام نمودار کنترل چندمتغیره نشانه بر اساس آزمون چندمتغیره نشانه که توسط هتمنزبرگر [۵] معرفی شده استوار است. نمودار دوم نیز با نام نمودار کنترل چندمتغیره ویلکاکسون بر اساس آزمون چندمتغیره ویلکاکسون که آن نیز توسط هتمنزبرگر [۵] معرفی شده استوار است. هتمنزبرگر علاوه بر دو آزمون تک نمونه‌ای نشانه و ویلکاکسون، یک آزمون دو نمونه‌ای چند متغیره با نام آزمون چندمتغیره من‌ویتنی نیز معرفی کرده است. ما در مقاله حاضر بر اساس آزمون دو نمونه‌ای چندمتغیره من‌ویتنی، یک نمودار کنترل چندمتغیره من‌ویتنی معرفی کرده‌ایم.

تفاوت مهم نمودار کنترل در مقاله حاضر با دو نمودار کنترلی که توسط بوون و چاکرابورتی [۲] معرفی گردیده در این است که در آن دو نمودار، پارامترها معلوم فرض می‌شود ولی در مقاله حاضر پارامترها نامعلوم فرض شده و در فاز اول برآورد می‌شوند. این مقاله شامل ۷ بخش است، که در بخش ۲ نمودار T^2هتلینگ و در بخش ۳ آزمون چندمتغیره من‌ویتنی مطرح شده است که این آزمون

[۱] دانشجوی کارشناسی ارشد آمار دانشگاه اصفهان

[۲] استادیار گروه آمار دانشگاه اصفهان

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]

می‌تواند مقدمه‌ای برای معرفی نمودار کنترل چندمتغیره من‌ویتنی در بخش ٤ باشد. در بخش ۵ نیز یک مثال شبیه‌سازی ارائه شده و در نهایت دو نمودار با استفاده از داده‌های شبیه‌سازی شده مورد مقایسه قرار گرفته‌اند.

۲ نمودار T^2 هتلینگ

نمودار کنترل T^2 هتلینگ رایج‌ترین نمودار کنترل کیفیت برای کنترل بردار میانگین در حالتی است که بردار میانگین و ماتریس کوواریانس فرآیند چندمتغیره آماری نامعلوم هستند. برای آشنایی بیشتر با این نمودار به مونتگمری [٦] مراجعه کنید.

فرض کنید $\underline{X} = (X_1, X_2, \ldots, X_p)^T$ برداری از p مشخصه کیفیت است که در شرایط کنترل دارای توزیع $\mathrm{N}_p(\underline{\mu}, \Sigma)$ می‌باشد که در آن $\Sigma, \underline{\mu}$ نامعلوم هستند. همچنین فرض کنید:

$$\underline{\bar{X}} = n_1^{-1} \sum_{r=1}^{n_1} \underline{X}_r,$$
$$\mathrm{S} = (n_1 - 1)^{-1} \sum_{r=1}^{n_1} (\underline{X}_r - \underline{\bar{X}})(\underline{X}_r - \underline{\bar{X}})^T.$$

به ترتیب بردار میانگین و ماتریس کوواریانس یک نمونه تصادفی n_1 تایی در شرایط کنترل در فاز اول هستند. واضح است که $\underline{\bar{X}}$ و S به ترتیب برآوردهای نااریب برای $\underline{\mu}$ و Σ است. اکنون فرض کنید $\underline{\bar{Y}}_t$ بردار میانگین t–امین نمونه n_2 تایی از زمان شروع پیگیری کنترل آماری فرآیند است. به عبارت دیگر $\underline{\bar{Y}}_t$ مربوط به فاز دوم است و لذا مستقل از $\bar{X}$ و S است. با توجه به تعریف متغیر تصادفی T^2 هتلینگ و رابطه آن با F به سهولت می‌توان نشان داد که متغیر تصادفی زیر دارای توزیع T^2 هتلینگ و با افزودن ضریبی دارای توزیع F است:

$$T_t^2 = \frac{n_1 n_2}{n_1 + n_2} (\underline{\bar{Y}}_t - \underline{\bar{X}})^T \mathrm{S}^{-1} (\underline{\bar{Y}}_t - \underline{\bar{X}}) \sim \frac{p(n_1 - 1)}{n_1 - p} F_{p, n_1 - p}.$$

اگر از این آماره کنترل استفاده شود حد بالای کنترل به‌صورت زیر است:

$$\mathrm{UCL} = \frac{p(n_1 - 1)}{n_1 - p} F_{\alpha, p, n_1 - p}, \qquad (۱)$$

که در آن F_{α,p,n_1-p} صدک $100(1-\alpha)$ درصد توزیع F با p و $n_1 - p$ درجه آزادی است.

۳ آزمون چندمتغیره من‌ویتنی

آزمون چندمتغیره من‌ویتنی[۳]، یک آزمون ناپارامتری دو نمونه‌ای می‌باشد که توسط هتمنزبرگر [۵] معرفی شده است. در این بخش به طور اجمالی این آزمون را معرفی می‌کنیم. بردار مکان می‌تواند بردار میانگین‌ها یا میانه‌ها باشد. بنابراین آزمون زیر مورد نظر است:

$$\begin{cases} H_0 : \underline{\theta}_1 = \underline{\theta}_2 \\ H_1 : \underline{\theta}_1 \neq \underline{\theta}_2 \end{cases}$$

که در آن $\underline{\theta}_1$ و $\underline{\theta}_2$ به ترتیب بردار مکان جامعه اول و دوم هستند. آماره این آزمون به صورت زیر است:

$$MW^2 = \underline{W}^T V^{-1} \underline{W}, \qquad (۲)$$

که در آن $\underline{W} = (W_1, W_2, \ldots, W_p)^T$ یک بردار $p \times 1$ است و i–امین مولفه آن، آماره من‌ویتنی یک متغیره مرسوم است که مرکزی شده است. به عبارت دیگر:

$$W_i = \sum_{r=1}^{n_1} R_{ri} - \frac{n_1(N+1)}{2}, \qquad i = 1, \ldots, p \qquad (۳)$$

که n_1 و n_2 به ترتیب اندازه نمونه اول و دوم، $N = n_1 + n_2$ و $\sum_{r=1}^{n_1} R_{ri}$ مجموع رتبه‌های مشاهدات نمونه اول است زمانی که با مشاهدات نمونه دوم ادغام و رتبه‌بندی شده است. مقصود از مرکزی شده این است که آماره مرسوم من‌ویتنی طوری تعدیل شده که امید ریاضی آن صفر است، یعنی داریم $\mathrm{E}(W_i) = 0$.

برای آشنایی با آزمون یک متغیره من‌ویتنی به عنوان مثال به بهبودیان[۱] مراجعه کنید.

ماتریس V یک ماتریس $p \times p$ متقارن است که درایه‌های قطری آن در صورتی که گره وجود نداشته باشد برابر $\frac{n_1 n_2 (N+1)}{12}$ و درایه‌های غیرقطری آن به صورت زیر است:

$$V_{jk} = \left(\frac{n_1 n_2}{N(N-1)}\right) \times \left(\sum_{r=1}^{N} R_{rj} R_{rk} - \frac{1}{4} N (N+1)^2\right). \qquad (٤)$$

با فرض پیوسته بودن متغیر مورد بررسی گره وجود نخواهد داشت. البته در عمل ممکن است مشاهدات گرد شده و صفرهایی ظاهر

[۳] Multivariate Mann Whitney test

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]

شوند.

مشهور است که $\underline{W}$ تحت فرض H_0 به طور مجانبی دارای توزیع $\mathrm{N}_p(\underline{0}, V)$ است که V یک برآورد سازگار برای ماتریس کوواریانس W است و در رابطه ٤ تعریف شده است. بنابراین MW^2 تحت فرض H_0به طور مجانبی دارای توزیع خی‌دو با p درجه آزادی است. لذا فرض H_0 در سطح معنی داری α رد می‌شود اگر $MW^2 > \chi^2_{\alpha,p}$. که در آن $\chi^2_{\alpha,p}$ صدک $100(1-\alpha)$ درصد توزیع خی‌دو با p درجه آزادی است.

٤ نمودار کنترل چندمتغیره من‌ویتنی

فرض کنید $\{\underline{X}_1, \ldots, \underline{X}_{n_1}\}$یک نمونه تصادفی n_1 تایی از بردار $\underline{X}_{p\times 1}$ در فاز اول است و نمونه $\{\underline{Y}_{1t}, \ldots, \underline{Y}_{n_2 t}\}$، t–امین نمونه تصادفی n_2–تایی از $\underline{X}$ در فاز دوم می‌باشد. همچنین فرض کنید که همه p متغیر مورد بررسی از نوع پیوسته‌اند. مقدار آماره کنترل متناظر با t–امین نمونه فاز دوم به صورت زیر است:

$$MW_t^2 = \underline{W}_t^T V_t^{-1} \underline{W}_t \qquad (٥)$$

که در آن $\underline{W}_t = (W_{1t}, W_{2t}, \ldots, W_{pt})^T$. درایه i–ام این بردار یعنی W_{it} آماره مرکزی شده مرسوم یک متغیره من‌ویتنی متناظر با i–امین متغیر و t–امین نمونه است که به صورت زیر تعریف می‌شود:

$$W_{it} = \sum_{r=1}^{n_1} R_{rit} - \frac{n_1(N+1)}{2}, \qquad (٦)$$

که n_1 اندازه نمونه فاز اول، n_2 اندازه زیر نمونه‌های فاز دوم، $N = n_1 + n_2$ و $\sum_{r=1}^{n_1} R_{rit}$ مجموع رتبه‌های نمونه فاز اول، زمانی که با tامین نمونه فاز دوم ادغام و مرتب شده، می‌باشد. V_t یک ماتریس $p\times p$ و معین مثبت است که درایه‌های روی قطر اصلی در صورت نبودن گره برابر $\frac{n_1 n_2(N+1)}{12}$ و درایه سطر j–ام و ستون k–ام آن به صورت زیر می‌باشد:

$$V_{jkt} = \frac{n_1 n_2}{N(N-1)} \times \sum_{r=1}^{N} R_{rjt} R_{rkt} - \frac{1}{4} N(N+1)^2]. \qquad (٧)$$

$\underline{W}_t$ تحت فرض H_0 به طور مجانبی دارای توزیع $N_p(\underline{0}, V_t)$ است. آماره کنترل MW_t^2 به طور مجانبی دارای توزیع خی دو با p درجه آزادی است، یعنی

$$MW_t^2 = \underline{W}_t^T V_t^{-1} \underline{W}_t \sim \chi^2_p.$$

لذا این نمودار کنترل دارای یک حد کنترل بالای $UCL = \chi^2_{\alpha,p}$ و خط مرکزی $CL = p$ می‌باشد.

٥ مثال شبیه‌سازی

با استفاده از نرم‌افزار $MATLAB$ یک نمونه با حجم $n_1 = 10$ از توزیع $\mathrm{N}_3(\underline{\mu}, \Sigma)$ به عنوان داده‌های فاز اول با پارامترهای زیر تولید کردیم

$$\mu_{in} = \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}, \Sigma_{in} = \begin{bmatrix} 1 & -0.4 & -0.4 \\ -0.4 & 1 & -0.4 \\ -0.4 & -0.4 & 1 \end{bmatrix}.$$

ماتریس 10×3 داده‌های تولید شده در زیر ارائه شده است.

$$X = \begin{bmatrix} 1.2648 & 4.0954 & 10.2774 \\ 0.7877 & 5.6327 & 7.6681 \\ 2.1403 & 4.8600 & 8.1273 \\ 0.1505 & 5.9716 & 9.4638 \\ 2.0675 & 3.4371 & 8.6682 \\ 1.3680 & 6.5136 & 7.7463 \\ 0.2968 & 5.2177 & 8.7976 \\ 0.5072 & 4.9073 & 9.6957 \\ 0.0765 & 5.1039 & 10.0138 \\ -0.1467 & 5.0485 & 10.9706 \end{bmatrix}.$$

بردار میانگین و ماتریس کوواریانس مبتنی بر داده‌های فاز اول در زیر ارائه شده است

$$\underline{\bar{X}} = \begin{bmatrix} 0.85126 \\ 5.07878 \\ 9.14288 \end{bmatrix},$$

$$S = \begin{bmatrix} 0.67733 & -0.29023 & -0.52973 \\ -0.29023 & 0.76921 & -0.31385 \\ -0.52973 & -0.31385 & 1.25512 \end{bmatrix}.$$

به عنوان داده‌های فاز دوم تعداد $10 + 10$ نمونه، هر یک به اندازه $n_2 = 30$ از توزیع نرمال سه متغیره تولید کردیم به این ترتیب که ۱۰ نمونه ۳۰–تایی اول از فاز دوم را تحت کنترل (با پارامترهای

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]

μ_{in} و Σ_{in}) تولید کردیم ولی 10 نمونه 30تایی دوم از فاز دوم را در شرایط خارج از کنترل با پارامترهای زیر تولید کردیم.

$$\mu_{out} = \begin{bmatrix} 1 \\ 6 \\ 10 \end{bmatrix}, \quad \Sigma_{out} = \Sigma_{in}.$$

در واقع برای ۱۰ نمونه دوم فاز دوم میانگین متغیر دوم و سوم را به اندازه یک واحد انحراف معیار افزایش دادیم و در متغیر اول تغییری ایجاد نکردیم. در فاز دوم ماتریس‌های $Y_1, Y_2, \ldots, Y_{10}$ حاوی داده‌های تحت کنترل و $Y_{11}, Y_{12}, \ldots, Y_{20}$ حاوی داده‌های خارج از کنترل هستند به طوری که هر یک از Y_t–ها یک ماتریس 30×3 هستند. برای روشن شدن جنبه‌های محاسباتی مقدار آماره من‌ویتنی را برای (X, Y_1) با تفصیل توضیح می‌دهیم. که در آن X ماتریس 10×3 داده‌های فاز اول و Y_1 اولین ماتریس 30×3 داده‌ها از فاز دوم است. جزئیات چگونگی محاسبه آماره کنترل در جدول ملاحظه می‌شود. در این جدول ستون‌های $X_{r11}, X_{r21}, X_{r31}$ حاوی ماتریس $[X^T, Y_1^T]^T$ هستند. به عبارت دیگر ده سطر اول این ستون‌ها حاوی ماتریس X و سی سطر آخر این ستون‌ها حاوی ماتریس Y_1 است.

بر اساس جدول ۱ داریم:

$$W_{11} = \sum_{r=1}^{10} R_{r11} - \frac{10(40+1)}{2} = -37,$$
$$W_{21} = \sum_{r=1}^{10} R_{r21} - \frac{10(40+1)}{2} = 8,$$
$$W_{31} = \sum_{r=1}^{10} R_{r31} - \frac{10(40+1)}{2} = 2.5.$$

توجه کنید به عنوان مثال $\sum_{r=1}^{10} R_{r11} = 168$ مربوط به متغیر اول و جمع رتبه‌های ۱۰ نمونه فاز اول است وقتی که با ۳۰ نمونه دوم ادغام و رتبه‌بندی شده است. به طریق مشابه $\sum_{r=1}^{10} R_{r21} = 213$ و $\sum_{r=1}^{10} R_{r31} = 207.5$ به ترتیب مربوط به متغیرهای دوم و سوم هستند.

به این ترتیب $\underline{W}_1 = (-37 \quad 8 \quad 2.5)^T$ و ماتریس V_1 به صورت زیر است:

$$V_1 = \begin{bmatrix} 1025 & -486.9 & -221.8 \\ -486.9 & 1025 & -511.4 \\ -221.8 & -511.4 & 1024.9 \end{bmatrix}.$$

لذا آماره من‌ویتنی متناظر با نمونه اول فاز دوم برابر

$$MW_1^2 = (-37 \;\; 8 \;\; 2.5)\times \begin{bmatrix} 1025 & -486.9 & -221.8 \\ -486.9 & 1025 & -511.4 \\ -221.8 & -511.4 & 1024.9 \end{bmatrix}^{-1} \begin{pmatrix} -37 \\ 8 \\ 2.5 \end{pmatrix} = 1.7880$$

است. بر اساس داده‌های تولید شده بردار میانگین اولین نمونه در فاز دوم برابر:

$$\underline{\bar{Y}}_1 = (1.0867 \quad 4.93115 \quad 9.11533)^T,$$

است و در نتیجه

$$\underline{\bar{Y}}_1 - \bar{X} = (0.2354 \quad -0.1476 \quad -0.0275)^T.$$

بنابراین مقدار مشاهده شده آماره کنترل T^2 هتلینگ متناظر با اولین نمونه فاز دوم برابر

$$T_1^2 = \frac{10 \times 30}{10+30}(0.2354 \quad -0.1476 \quad -0.0275)\times \begin{bmatrix} 0.6773 & -0.2902 & -0.5298 \\ -0.2902 & 0.7692 & -0.3139 \\ -0.5298 & -0.3139 & 1.2551 \end{bmatrix}^{-1} \begin{bmatrix} 0.2354 \\ -0.1476 \\ -0.0275 \end{bmatrix} = 0.9060$$

است. مقدار آماره کنترل من‌ویتنی و T^2 هتلینگ برای هریک از بیست نمونه فاز دوم در جدول ۲ ارائه شده است.

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]

جدول ۱: چگونگی محاسبه آماره من‌ویتنی برای نمونه اول از فاز دوم

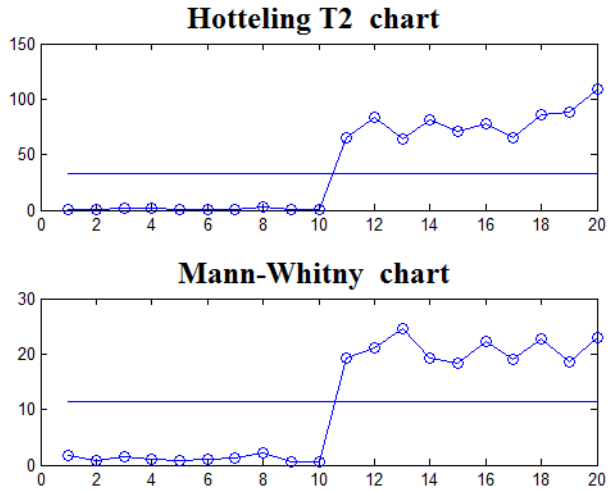
r	X_{r11}	R_{r11}	X_{r21}	R_{r21}	X_{r31}	R_{r31}	R_{r11}^2	$R_{r11}R_{r21}$	$R_{r11}R_{r31}$	R_{r21}^2	$R_{r21}R_{r31}$	R_{r31}^2
1	1.2648	24	4.0954	9	10.2774	33	576	216	792	81	297	1089
2	0.7877	17	5.6327	30	7.6681	4	289	510	68	900	120	16
3	2.1403	38	4.86	17	8.1273	8	1444	646	304	289	136	64
4	0.1505	6	5.9716	34	9.4638	27	36	204	162	1156	918	729
5	2.0675	36	3.4371	3	8.6682	14	1296	108	504	9	42	196
6	1.368	25	6.5136	39	7.7463	5.5	625	975	137.5	1521	214.5	30.25
7	0.2968	7	5.2177	23	8.7976	16	49	161	112	529	368	256
8	0.5072	9	4.9073	18	9.6957	30	81	162	270	324	540	900
9	0.0765	4	5.1039	21	10.0138	32	16	84	128	441	672	1024
10	−0.1467	2	5.0485	19	10.9706	38	4	38	76	361	722	1444
11	2.1785	39	3.7783	5	10.3404	34	1521	195	1326	25	170	1156
12	1.49	27	4.3701	11	8.6668	13	729	297	351	121	143	169
13	1.457	26	6.1682	35	7.3412	2	676	910	52	1225	70	4
14	−1.9343	1	6.4076	38	9.5981	29	1	38	29	1444	1102	841
15	0.9081	19	5.3449	25	10.4425	36	361	475	684	625	900	1296
16	1.9338	34	3.0698	2	10.4177	35	1156	68	1190	4	70	1225
17	0.6184	12	4.7569	15	10.6158	37	144	180	444	225	555	1369
18	1.5543	28	4.7921	16	8.5692	12	784	448	336	256	192	144
19	0.0762	3	5.2736	24	8.6842	15	9	72	45	576	360	225
20	0.9272	21	5.4242	28	8.2077	9	441	588	189	784	252	81
21	2.7491	40	4.6119	13	9.0936	22	1600	520	880	169	286	484
22	1.8225	33	6.3024	37	6.8515	1	1089	1221	33	1369	37	1
23	0.6349	15	4.0429	8	9.7454	31	225	120	465	64	248	961
24	0.619	13	5.7709	31	9.3413	25	169	403	325	961	775	625
25	0.5905	11	5.963	33	8.8173	17	121	363	187	1089	561	289
26	1.1216	22	4.6289	14	11.1658	39	484	308	858	196	546	1521
27	0.6322	14	5.3627	26	8.8893	18	196	364	252	676	468	324
28	1.5905	31	5.9088	32	8.0035	7	961	992	217	1024	224	49
29	0.9051	18	4.584	12	8.9326	19	324	216	342	144	228	361
30	0.9103	20	6.2827	36	7.7463	5.5	400	720	110	1296	198	30.25
31	0.3403	8	5.191	22	9.2319	23	64	176	184	484	506	529
32	0.1152	5	6.7598	40	8.4262	10	25	200	50	1600	400	100
33	1.1323	23	3.8919	6	9.5885	28	529	138	644	36	168	784
34	1.59	30	3.6699	4	9.0361	20	900	120	600	16	80	400
35	1.6514	32	4.0112	7	8.492	11	1024	224	352	49	77	121
36	1.9723	35	4.3043	10	9.0553	21	1225	350	735	100	210	441
37	1.5686	29	5.5272	29	7.3807	3	841	841	87	841	87	9
38	0.7588	16	5.0643	20	9.4419	26	256	320	416	400	520	676
39	2.1098	37	1.2946	1	12.0357	40	1369	37	1480	1	40	1600
40	0.5774	10	5.3764	27	9.3014	24	100	270	240	729	648	576
جمع کل		820		820		820	22140	14278	15657	22140	14150.5	22140

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]

جدول ۲: مقادیر آماره من‌ویتنی و T^2 هتلینگ برای مثال شبیه‌سازی

t	1	2	3	4	5	6	7	8	9	10
MW_t	1.7880	0.8099	1.4722	0.9353	0.9195	0.9565	1.2673	2.3036	0.5049	0.6745
T_t^2	0.9060	0.6224	1.7680	1.3118	1.0345	0.4650	1.0670	3.1452	0.5667	0.4223
t	11	12	13	14	15	16	17	18	19	20
MW_t	19.3208	21.2145	24.5187	19.2602	18.2748	22.2888	19.0011	22.7672	18.5613	22.9799
T_t^2	64.8022	83.7042	63.6599	81.1390	71.0550	77.7763	64.7485	86.5037	88.1414	108.4988

نمودار کنترل T^2 هتلینگ و من‌ویتنی در شکل ۱ مقایسه شده‌اند.



شکل ۱: نمودار کنترل T^2 هتلینگ و من‌ویتنی برای مثال شبیه‌سازی

فرض کنید $\alpha = 0.01$ است. حد کنترل برای نمودار T^2هتلینگ برابر

$$UCLT^2 = \frac{(n_1-1)p}{n_1-p} F_{\alpha,p,n_1-p} = 32.5978, \qquad (۸)$$

است و حد کنترل برای نمودار من‌ویتنی برابر

$$UCLMW = \chi^2_{\alpha,p} = 11.3449, \qquad (۹)$$

است. شکل ۱ نمودار بالایی نمودار کنترل T^2 هتلینگ و نمودار پایینی نمودار من‌ویتنی است. با توجه به این که ده نمونه اول تحت کنترل و ۱۰ نمونه دوم خارج از کنترل هستند، ملاحظه می‌کنیم که در این مثال در هر دو نمودار T^2 هتلینگ و من‌ویتنی خطای نوع اول و دوم وجود ندارد. به عبارت دیگر هر ۱۰ نمونه که در شرایط کنترل تولید شده‌اند توسط نمودارها تحت کنترل نشان داده شده‌اند و هر ۱۰ نمونه که در شرایط خارج کنترل تولید شده‌اند توسط نمودارها خارج کنترل رسم شده‌اند. در بخش بعدی نتیجه ۱۰۰۰۰ تکرار از این مثال ارائه شده است.

٦ مطالعه شبیه‌سازی

در این بخش با استفاده از نرم‌افزار $MATLAB$ داده‌هایی از توزیع نرمال و t چندمتغیره با ۵ درجه آزادی تولید کرده و به مقایسه عملکرد نمودار T^2 هتلینگ و من‌ویتنی در ۱۰۰۰۰ بار تکرار با فرض $\alpha = 0.01$ می‌پردازیم. بنابراین حدود کنترل بالا برای هریک از نمودارها به صورت رابطه ۸ و ۹ است.

الف) توزیع نرمال چندمتغیره.

در این قسمت یک مطالعه شبیه‌سازی با شرایطی همانند مثال بخش قبل، ارائه کرده‌ایم. به طور کلی با شبیه‌سازی که با ۱۰۰۰۰ بار تکرار صورت گرفته است هنگامی که توزیع، نرمال چندمتغیره باشد نمودار T^2 هتلینگ و من‌ویتنی، از بین ۱۰ نمونه خارج از کنترل، به ترتیب به طور متوسط ۹/۳۹۸۱ و ۹/۹۲۴۸ نمونه خارج از کنترل را به درستی تشخیص داده‌اند و احتمال خطای نوع اول تجربی (مبتنی بر شبیه‌سازی فوق) برای نمودارهای T^2 هتلینگ و من‌ویتنی به ترتیب ۰/۰۰۹۴۱ و ۰/۰۰۵۶۹ به دست آمده است. این در حالی است که احتمال خطای نوع اول اسمی در هر دو نمودار $\alpha = 0.01$ می‌باشد. مقصود از احتمال خطای نوع اول تجربی نسبت هشدارهای نادرستی است که در ۱۰۰۰۰ بار تکرار مشاهده شده است. در مورد نمودار من‌ویتنی احتمال خطای نوع اول تجربی با مقدار اسمی 0.01 تفاوت زیاد دارد که دلیل آن تقریبی بودن توزیع خی‌دو برای آماره من‌ویتنی است.

ب) توزیع t چندمتغیره با ۵ درجه آزادی.

در مطالعه شبیه‌سازی دیگری، یک نمونه به اندازه ۱۰ از فاز اول در شرایط کنترل و ۱۰+۱۰ نمونه هر یک به اندازه ۳۰ در فاز دوم از توزیع t چندمتغیره با ۵ درجه آزادی تولید

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]

کردیم که ۱۰ نمونه اول فاز دوم در شرایط تحت کنترل با پارامترهای زیر تولید شده‌اند

$$\mu = \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}, \Sigma_{t_5} = \begin{bmatrix} 0.6 & -0.24 & -0.24 \\ -0.24 & 0.6 & -0.24 \\ -0.24 & -0.24 & 0.6 \end{bmatrix},$$

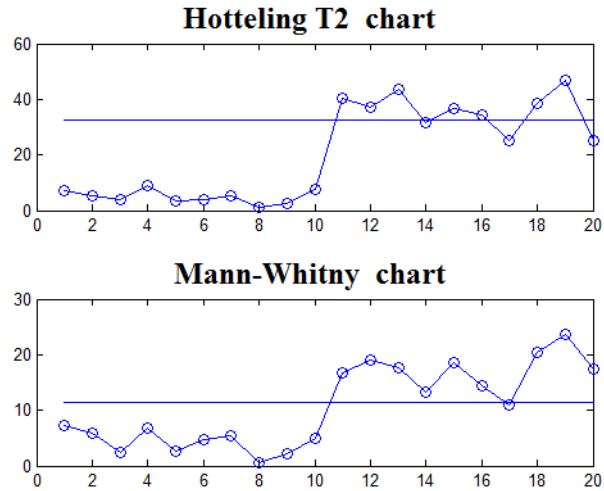
در حالی که ولی ۱۰ نمونه دوم فاز دوم در شرایط خارج از کنترل و با پارامترهای زیر تولید گشته‌اند.

$$\mu = \begin{bmatrix} 1 \\ 6 \\ 10 \end{bmatrix}, \Sigma_{t_5} = \begin{bmatrix} 0.6 & -0.24 & -0.24 \\ -0.24 & 0.6 & -0.24 \\ -0.24 & -0.24 & 0.6 \end{bmatrix},$$

که در آن Σ_{t_5} ماتریس مقیاس می‌باشد. بنابراین ماتریس کوواریانس به صورت زیر است

$$\Sigma = \frac{d}{d-2}\Sigma_{t_5} = \frac{5}{3}\begin{bmatrix} 0.6 & -0.24 & -0.24 \\ -0.24 & 0.6 & -0.24 \\ -0.24 & -0.24 & 0.6 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & -0.4 & -0.4 \\ -0.4 & 1 & -0.4 \\ -0.4 & -0.4 & 1 \end{bmatrix}$$

نتایج به صورت شکل ۲ حاصل شده است.



شکل ۲: نمودار کنترل T^2 هتلینگ و من‌ویتنی برای توزیع t با ۵ درجه آزادی

نتایج در ۱۰۰۰۰ بار تکرار نشان می‌دهد زمانی که توزیع t چندمتغیره با ۵ درجه آزادی باشد نمودارهای T^2 هتلینگ و من‌ویتنی، از بین ۱۰ نمونه خارج از کنترل، به ترتیب به طور متوسط ۷/۹۴۵۳ و ۹/۳۶۶۵ نمونه خارج از کنترل را به درستی تشخیص داده‌اند. احتمال خطای نوع اول تجربی برای نمودار T^2 هتلینگ و من‌ویتنی به ترتیب ۰/۰۱۰۷۳ و ۰/۰۰۵۶۵ است.

پ) اندازه نمونه بزرگ.

در مطالعات شبیه‌سازی فوق اندازه نمونه کوچک بوده است. بنابراین تقریب خی‌دو بودن آماره من‌ویتنی قابل اعتماد نیست. غیرقابل اعتماد بودن تقریب خی‌دو از طریق مقایسه مقدار تجربی خطای نوع اول با مقدار اسمی آن نیز واضح است. در این بخش مطالعه شبیه‌سازی دیگری انجام داده‌ایم که مشابه مطالعه شبیه‌سازی قسمت (ب) است، تنها تفاوت در اندازه نمونه است. در اینجا اندازه نمونه را ۱۰۰ در نظر گرفته‌ایم. ۱۰۰۰۰ بار تکرار حاکی از آن است که مقدار تجربی خطای نوع اول برای آماره من ویتنی در این حالت برابر ۰/۰۰۹ است که به مقدار اسمی ۰/۰۱ نزدیک است. مقادیر ARL_1 هر دو نمودار یکسان و برابر یک به‌دست آمده که نشان می‌دهد که تقریباً عملکرد هر دو نمودار یکسان است. در صورت معلوم نبودن توزیع می‌توان از نمودار من‌ویتنی استفاده کرد.

۷ بحث و نتیجه‌گیری

از طریق شبیه‌سازی صورت گرفته نتیجه می‌شود، موقعیت‌هایی وجود دارند که احتمالاً عملکرد نمودار کنترل من‌ویتنی بهتر از T^2 هتلینگ است حتی زمانی که توزیع نرمال است. زمانی که توزیع غیر نرمال است این عملکرد بهتر، آشکارتر می‌شود. از جمله مزیت‌های نمودار من‌ویتنی نسبت به نمودار سنتی T^2 هتلینگ این است که به فرض توزیعی نیاز ندارد. از جمله معایب نمودار کنترل چندمتغیره من‌ویتنی عدم امکان اجرای آن برای مشاهدات تکی و نیاز به اندازه نمونه بزرگ است.

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]

مراجع

[۱] بهبودیان، ج. (۱۳۸۳). آمار ناپارامتری. انتشارات دانشگاه شیراز.

[2] Boon, J. M. and Chakraborti, S. (2012). Two simple Shewhart-type multivariate nonparametric control charts. *Applied stochastic models in Business and Industry*; **16**(4), 613-623.

[3] Chakraborti S, Graham MA. (2008). Nonparametric control chart. *Encyclopedia of Quality and Reliability*.Wiley: New York, 415-429.

[4] Chakraborti, S., Van Der Laan,P. and Bakir, S. T. (2001). Nonparametric control charts: an overview and som results. *Journal of Quality Technology*; **33**(3), 304-315.

[5] Hettmansperger, T. P .(2006). Multivariate location tests. *Encyclopedia of Statistical Sciences*. Wiley.

[6] Montgomery, D. C. (2001). *Introduction to statistical quality control*, 4th edition. Wiley, New York.

[Downloaded from andisheyeamari.irstat.ir on 2024-10-12]