International Journal of Industrial Engineering & Production Management (2014)

January 2014, Volume 24, Number 4
pp. 423-436

http://IJIEPM.iust.ac.ir/

Robust Estimator Detection Outlier Points in First Phase of Multivariate Quality Control Chart with Hierarchical Clustering Technique

M. Bahrami & Gh. Raissi*

Mohammad Ali Bahrami, *Master of Science of Industrial Engineering, Isfahan University of Technology,ma.bahrami@in.iut.ac.ir*
Gholam_Ali Raissi, *Assistant professor of Industrial Engineering, Isfahan University of Technology,raissi@cc.iut.ac.ir*

Keywords

Statistical Process Control,
Outlier points,
Hierarchical clustering technique,
Robust estimator

ABSTRACT

The main objective of performed researches in the field of multivariate statistical process control is to consider the correlation between multiple qualitative attributes for one step of process. In the second phase of the multivariate process control procedure, the rest of process is being studied whether it is under control, using the achieved control limits from the first phase and future observations. So, having found the outlier points of the first phase before the control limits to be computed, it is considered as an important issue. In order to detect these outlier points, Variety of techniques, that the majority of them rely on primary random samples, are proposed. These primary random points can effect on the precision of algorithms and final solution of the problem. In this paper, a robust estimator is issued applying hierarchical clustering technique that is not affected by outlier data in sample or unusual data, rather than the model assumptions and will detect the outlier points in multivariate control charts of the first phase in order to get them removed. Then, the proposed method is evaluated by creating the variety of scenarios from outlier points and the final outcome is compared with the Classical Hoteling and the least determinant covariance estimator. The evaluations represent that the proposed method detects more outlier points in less time rather than the former performed researches.

* **Corresponding author. Mohammad Ali Bahrami**
Email: *ma.bahrami@in.iut.ac.ir*

نشـــریه بین المللـــی مهنـــدسی صنـــایع و مدیـــریت تولیـــد

شمـــاره ۴، جلـــد ۲۴، زمستـــان ۱۳۹۲
صفحـــه ۴۳۶-۴۲۴

نشریه بین المللی مهندسی صنایع و مدیریت تولید دانشگاه علم و صنعت ایران

ISSN: 2008-4870

http://IJIEPM.iust.ac.ir/

برآوردگر باثبات شناسایی نقاط پرت فاز اول نمودارهای کنترل کیفیت چندمتغیره با استفاده از تکنیک خوشه‌بندی سلسله مراتبی

محمدعلی بهرامی* و غلامعلی رئیسی اردلی

کلمات کلیدی

کنترل فرآیند آماری،
نقاط پرت،
خوشه‌بندی سلسله مراتبی،
برآوردگر باثبات.

چکیده:

هدف اصلی تحقیقات صورت گرفته در زمینه‌ی کنترل فرآیند آماری چند متغیره، درنظر گرفتن همبستگی بین چندین مشخصه کیفی برای یک مرحله از فرآیند است. در فاز دوم رویه کنترل فرآیند چندمتغیره با استفاده ازحدود کنترلی بدست آمده از فاز اول و مشاهدات آتی، تحت کنترل بودن ادامه فرآیند بررسی می‌شود، یافتننقاط پرتفاز اول قبل از محاسبه حدود کنترلی برای حصول نتیجه مناسبدارای اهمیت بالاست. تکنیک‌های متفاوتی جهت شناسایی این نقاط انحرافی ارائه شده است که اکثر این الگوریتم‌ها به نمونه تصادفی اولیه وابسته می‌باشد که این نقطه شروع تصادفی می‌تواند بر دقت الگوریتم و جواب نهایی مسئله تأثیرگذار باشد. در این مقاله برآوردگری باثبات با استفاده از تکنیک خوشه‌بندی سلسله مراتبی ارائه می‌شود که تحت تأثیر داده‌های انحرافی در نمونه یا داده‌های نامتعارف نسبت به فرضیات مدل قرار نمی‌گیرد و نقاط پرت موجود در فاز اول نمودارهای کنترلی چندمتغیره را شناسایی و حذف می‌کند. در نهایت با ایجاد سناریوهای مختلف از نقاط پرت و انحرافی، روش پیشنهادی مورد سنجش قرار گرفته و نتیجه کار با روش‌های هتلینگ کلاسیک و برآوردگر حداقل دترمینان کواریانس مقایسه گردیده است. ارزیابی‌ها نشان می‌دهد که روش پیشنهادی نسبت به تحقیقات قبلی انجام شده در این زمینه، با مدت زمان کمتری، نقاط پرت و انحرافی بیشتری را شناسایی می‌کند.

۱. مقدمه

اصولاً یکی از مشکلات عمده‌ای که کشورهای درحال توسعه با آن مواجه می‌گردند، عدم وجود بازار رقابتی سالم و مناسب است. در اینگونه کشورها فرآورده‌های تولید شده به علت عدم اشباع بازار با مانع و مشکل خاصی مواجه نگردیده و غالباًبا هر کیفیتی که تولید شوند به فروش می‌رسند. با توجه به نظر دکتر جوران [۱] "در زمان کمبود اولین چیزی که قربانی می‌شود کیفیت است" و این اصل در کشورهایی که فاقد بازار رقابتی هستند، براحتی ملموس می‌باشد. با گذشت زمان به دلایل مختلف نظیر بروز مشکلات اقتصادی و درک این واقعیت از طرف سازمان‌ها که بهبود کیفیت می تواند با کاهش هزینه‌ها همراه باشد، سبب گردید که به مقوله کیفیت اهمیت داده شود و در تصمیم‌گیری مشتریان جهت ارزیابی محصول و یا خدمت، کیفیت به یک عامل اصلی تبدیل شود. لذا کیفیت یک عامل کلیدی جهت دستیابی به جایگاه رقابتی بهتر محسوب می‌گردد. امروزه اغلب سازمان‌ها تمایل چندانی به انتظار برای پایان یافتن تولید و نمونه‌برداری محصول نهایی ندارند بلکه به این امر اعتقاد پیدا کرده‌اند که لازم است قبل از آن که محصول معیوبی تولید شود فرآیند تولید به گونه‌ای تحت کنترل قرار گیرد که امکان تولید محصول معیوب ایجاد نشود. امروزه نمودارهای کنترلی یکی از بهترین ابزارهای شناخته شده جهت تحت کنترل قرار دادن فرآیندهای مختلف، شناخت انحرافات به وجود آمده و پتانسیل‌های ایجاد بهبود در آن‌ها می-باشد.تمایل زمینه‌های مختلف صنایع تولیدی و سازمان‌های خدماتی برای شناسایی خطاها و انحرافات و تمرکز فراوان این

تاریخ وصول: ۹۰/۶/۹
تاریخ تصویب: ۹۰/۱۲/۱

محمد علی بهرامی، دانشجوی کارشناسی ارشد دانشکده مهندسی صنایع، دانشگاه صنعتی اصفهان، ma.bahrami@in.iut.ac.ir
*نویسنده مسئول مقاله: **دکتر غلامعلی رئیسی اردلی**، دانشیاردانشکده مهندسی،بخش مهندسی صنایع، دانشگاه تربیت مدرس، raissi@cc.iut.ac.ir

ابزار به متن فرآیندها موجب گردید تا روش‌های سنتی نمونه برداری و بازرسی به تدریج جای خود را به انواع نمودارهای کنترلی دهد.گسترش تدریجی اعتقاد مدیران سازمان‌ها بر لزوم پیشگیری قبل از وقوع رخدادها، لزوم کاربرد نمودارهای کنترلی و بررسی روند شاخص‌های فرآیندی را بیش از گذشته نمایان می سازد. رشد تکنولوژی، ارتقاء دانش و خواسته کیفی مشتریان سبب گردیده است که محصولات و فرآیندها عموماً دارای چندین مشخصه کیفی به هم مرتبط باشند. مسائل کنترل کیفیت در صنعت، ممکن است شامل بیش از یک مشخصه کیفی باشد، خصوصاً وقتی این مشخصه‌ها وابسته هستند، باید روش مناسبی برای کنترل همزمان آنها فراهم باشد. نمودارهای کنترل چندمتغیره زیادی تاکنون طراحی شده است. هتلینگ اولین نفری بود که مسئله تحلیل یک مجموعه متغیرهای همبسته را بررسی کرد. با ظهور نمودارهای کنترلی چند متغیره با قابلیت بررسی هم‌زمان چند متغیر برای یک محصول، کاربرد این نمودارها بر نمودارهای کنترل تک متغیره ترجیح داده شد و کاربرد این ابزار قدرتمند به شدت افزایش یافت و موجب ورود این ابزار حتی به سازمان‌های خدماتی نیز گردید[۲]. با توجه به سرمایه‌گذاری‌های سنگین جهت سیستم‌های تولیدی و بخصوص سیستم‌های تولیدی پیشرفته، توقف فرآیند تولید یا خارج از کنترل شدن فرآیند، هزینه‌های سنگینی را بدنبال دارند. از طرف دیگر پیشرفت ابزارهای اندازه‌گیری، دسترسی به داده‌های زمان واقعی در تکرارهای کافی از فرآیند تولید را میسر نموده است. با توجه به اینکه معمولاً طبیعت فرآیندها دارای مشخصات کیفی مرتبط و اثر پذیر از یکدیگرند، لذا روش‌های کنترل فرآیند چندمتغیره بسیار مورد توجه قرار گرفته و موجب دستیابی به اندکی بهبود همراه با کاهش هزینه فراوان در عملکرد فرآیند شده است.روش‌های کنترل‌کیفیت چندمتغیره[1] نه فقط اثرات یک متغیر را نظارت می‌کنند، بلکه ارتباط بین متغیرها را نیز نظارت می‌کنند.این امر بازرسی گروهی از متغیرها را بطور کامل فراهم می‌کند. روش‌های چندمتغیره نسبت به روش‌های تک‌متغیره بسیار حساس به تغییر ارتباط متغیرها هستند.کنترل فرآیند آماری روشی است که برای نظارت بر پروسه به جهت شناسایی عواملی که باعث انحراف یا خروج فرآیند از کنترل می‌شوند، طراحی شده است. برای اینکه یک کالا کیفیت مورد نظر مشتری را داشته باشد، بایستی توسط فرآیندی پایدار و تکرار شدنی تولید شود. به همین منظور، فرآیند بایستی با تغییرات کوچکی حول هدف یا ابعاد اسمی تعریف شده برای خصوصیات محصول، تولید شود. نمودارهای کنترلی یکی از قویترین ابزارهای کنترل فرآیند آماری است.

کنترل آماری فرآیند چند متغیره در طی سال‌های جنگ جهانی و بعد از آن مورد توجه قرار گرفت. رویکرد چند متغیره در کنترل فرآیند آماری ، توسط هتلینگ[2] در سالهای ۱۹۴۷ تا ۱۹۵۱ در مطالعه دقت بمباران یک سایت دشمن مورد بررسی قرار گرفت. هتلینگ آماره T^2 را به عنوان مبنایی جهت نمایش کیفیت کلی بمباران و توزیع آماری آن را جهت برپاسازی آزمون فرض ارائه داد. هتلینگ اولین نفری بود که مسئله تحلیل یک مجموعه متغیرهای همبسته را بررسی کرد. او یک روش کنترل بر اساس مفهومی با عنوان فاصله آماری را توسعه داد. این آماره به احترام وی، آماره T^2 هتلینگ نام گرفت.

در نمودار کنترل T^2 فرض بر این است که داده‌ها دارای توزیع نرمال چند متغیره با میانگین μ و واریانس σ می‌باشد. اگر فرآیند شامل p شاخص کیفی با تعداد m مشاهده n تایی باشد،آماره T^2 بدین صورت تعریف می‌شود:

$$T^2 = n(\bar{X} - \bar{\bar{X}})' S^{-1} (\bar{X} - \bar{\bar{X}}) \qquad (۱)$$

که در این رابطه $\bar{X} = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p]^T$ و $\bar{\bar{X}}$ و S برآوردی از μ و Σ می‌باشد و زمانی که فرآیند در شرایط تحت کنترل به سر می‌برد تخمین زده می‌شوند.در حالتی که $n = 1$ می‌باشد، میسون و یانگ[۳] حدود کنترلی فاز ۱ و فاز ۲ برای نمودار کنترل T^2 را بدین گونه درنظر گرفته اند:

فاز ۱ :

$$UCL = \frac{(m-1)^2}{m} \beta_{[\alpha, p/2, (m-p-1)/2]} \qquad (۲)$$
$$LCL = 0$$

به طوری که $\beta_{[\alpha,p/2,(n-p-1)/2]}$، α امین مقدار توزیع بتا می‌باشد.

فاز۲:

$$UCL = \frac{p(m+1)(m-1)}{m(m-p)} F_{\alpha,p,m-p} \qquad (۳)$$
$$LCL = 0$$

که $\mathcal{F}_{(\alpha;p,n-p)}$، α امین مقدار توزیع $\mathcal{F}_{(p,n-p)}$ می‌باشد.

۲. شرح و بیان مسأله

رسم نمودارهای کنترلی در دو فاز صورت می‌گیرد. در فاز اولهدف دستیابی به یک مجموعه از داده‌های تحت کنترل می‌باشد. داده-های به دست آمده از این مرحله به عنوان مبنایی جهت تعریف

[1] *Multivariate Quality Control*

[2] *Hotteling*

فرآیند تحت کنترل می‌باشد. طبق نظر دانکن[1] [۴]، فاز اول شامل استقرار یک فرآیند تحت کنترل می‌باشد.

در فاز دوم با فرض این که در فاز یک فرآیند تحت کنترل بوده است، تحت کنترل بودن فرآیند برای مشاهدات بعدی مورد بررسی قرار می‌گیرد به طوری که حداقل تغییرپذیری نسبت به فاز یک صورت پذیرد. در حقیقت این مرحله از نمودارهای کنترلی به این منظور استفاده می‌شود که با رسم زیرگروه‌های آینده، وضعیت فرآیند مشخص شود. بنابراین فاز اول رویه نظارت مشخص می کند آیا داده های بدست آمدهدر حالت پایدار (یا تحت کنترل) می باشد یا خیر. در فاز دوم با استفاده از مشاهدات آتی و حدود کنترلی بدست آمده، تحت کنترل بودن ادامه فرآیند بررسی می‌شود. از آنجائیکه در فاز دوم رسم نمودارهای کنترلی، بررسی تحت کنترل بودن ادامه فرآیند با استفاده از مشاهدات آتی، به حدود کنترلی بدست آمده از داده‌های تحت کنترل فاز اول بستگی دارد، لذا وجود داده‌های غیرمعمول، روندها، تغییرات گام، نقاط پرت[2] و دورافتاده در فاز اول می‌تواند یک اثر ناسازگار بر حدود کنترلی فاز دوم داشته باشد، بنابراینیافتن این نقاط غیرمعمول قبل از محاسبه حدود کنترلی بسیار مهم است.

محاسبه حدود کنترلی بر اساس داده‌هایی که از فرآیندهای ناپایدار (یاخارج از کنترل) بدست آمده است منجر به کاهش دقت و کارایی رویه فاز دوممی‌شود.بنابراین مسئله مورد تحقیق در این مقاله شناسایی نقاط پرت و غیرمعمول فاز اول نمودارهای کنترلی چندمتغیره قبل از محاسبه حدود کنترلی فاز دوم می‌باشد. شناسایی نقاط دورافتاده در داده‌های چندمتغیره نسبت به تک-متغیره بسیار مشکل می‌باشد. یک دلیل این است که روش‌های گرافیکی مورد استفاده برای تک‌متغیره اغلب قابل استفاده در ابعاد بالاتر نمی‌باشد. از طرفیوجود خوشه‌ای از نقاط دورافتاده به علت انتقال مکان در یک جهت خاص، وجود چندین خوشه از نقاط دورافتاده در چندین جهت متفاوت، وجود نقاطی با موقعیتیکسان مانند داده‌های مناسب اما با تغییرپذیری بیشتر و یا نقاط دورافتاده به علت انتقال در برخی از اجزای بردار موقعیت، از جمله راه‌های مختلف برای بدست آمدن داده‌هایی از فرآیند خارج از کنترل می‌باشد[۵].

۳. بررسی ادبیات موضوع

آماره باثبات مجموعه‌ای از روش‌های آماری است که تحت تأثیر داده‌های انحرافی در نمونه یا داده‌های نامتعارف نسبت به فرضیات مدل، قرار نمی‌گیرد و در عین حال قابل انطباق با روش‌های متداول آماری باشد. وقتی در میان داده‌ها تعدادی نقطه‌ی نامتعارف وجود داشته باشد، روش‌های پایه وکلاسیک کارایی اندکی از خود نشان می‌دهند. عبارت باثباتی به معنای مقاومت در مقابل اثرات و پیامدهای نقاط غیر معمول و انحرافی می-باشد[۶].تاکنون روش‌های برآورد باثبات متفاوتی جهت شناسایی نقاط پرت و غیرمعمول چند متغیره از جمله در فاز اول نمودارهای کنترلی مطرح شده است که در ادامه مهمترین و کاراترین آنها بررسی می‌شود.

۳-۱. برآوردگر باثبات حداقل حجم بیضیوار[3]

برآوردگر MVE اولین بار (۱۹۸۴) توسط روسئو[4][۷] پیشنهاد شد و به طور وسیعی در نمودارهای کنترلیجهت شناسایی نقاط دورافتاده چندمتغیره مورد مطالعه قرار گرفت. آماره پیشنهادی باثبات برای T^2 بر اساس برآوردگر مینیمم حجم بیضیوار(MVE) با $T^2_{mve,i}$ نشان داده می‌شود و بدین گونه تعریف می‌شود:

$$T^2_{mve,i} = (X_i - \bar{X}_{mve})' S^{-1}_{mve} (X_i - \bar{X}_{mve}) \quad (۴)$$
$$for\ i = 1,2,\dots,m$$

که $\bar{X}_{mve}$ برآوردگر موقعیت و S_{mve} برآوردگر ماتریس واریانس-کواریانس می‌باشد. این برآوردگر به دنبال یافتن بیضیواری با کمترین حجم است که زیرمجموعه‌ای از حداقل h نقطه را پوشش دهد. این زیرمجموعه با سایز h هافست[5] نامگذاری می شود. اغلب h طوری انتخاب شده است که از تعداد نصف داده‌ها بیشتر باشد. برآوردگر موقعیت برابر با مرکز هندسی بیضی و برآوردگر ماتریس کواریانس ضربدر یک مقدار ثابت متعارف، ماتریسی است که خود بیضی را تعریف می‌کند. این مقدار ثابت برای تضمین پایداری می باشد.برای بدست آوردن بیشترین نقطه شکست ممکن، دیویس[6] [۸] و روسئو [۹] نشان دادند که مقدار صحیح $h = \frac{m+p+1}{2}$ باید برای MVE استفاده شود.

۳-۲. برآوردگر باثبات حداقل دترمینان کواریانس[7]

حداقل دترمینان کواریانس یک فرآیند برآوردی با نقطه شکست بالا می‌باشد که اولین بار توسط روسئو [۷] پیشنهاد شد. این الگوریتم به دنبال هافستی می‌باشد که کمترین مقدار دترمینان ماتریس واریانس-کواریانس را داشته باشد. آماره پیشنهادی

[1] *Duncan*
[2] *outlier*
[3] *Minimum Volume Ellipsoid (MVE)*
[4] *Rousseeuw*
[5] *halfsets*
[6] *Davies*
[7] *Minimum Covariance Determinant*

باثبات برای T^2 بر اساس برآوردگر حداقل دترمینان کواریانس (*MCD*) با $T^2_{mcd,i}$ نشان داده می‌شود و بدین گونه تعریف می‌شود:

$$T^2_{mcd,i} = (X_i - X_{mcd})' S^{-1}_{mcd} (X_i - X_{mcd}) \quad for\ i = 1,2,\dots,m \qquad (۵)$$

که X_{mcd} برآوردگرهای موقعیت و S_{mcd} برآوردگر ماتریس کواریانس-واریانس می‌باشد.نتیجه برآوردگر موقعیت، بردار میانگین نمونه از نقاطی است که در هافست موجودند و برآوردگر پراکندگی ماتریس کواریانس-واریانس نمونه نقاط در یک مقدار ثابت مناسبی است. برآوردگر پراکندگی MCD بدین گونه محاسبه می‌شود:

$$S_{MCD} = (c_{\gamma,p} \times b^n_{\gamma,p}) C_{MCD} \qquad (۶)$$

که C_{MCD} ماتریس کواریانس هافست و $c_{\gamma,p}$ یک ضریب ثبات و $b^n_{\gamma,p}$ ضریب تصحیح نمونه متناهی است. ($1-\gamma$) نشان‌دهنده نقطه شکست مجانبی برآوردگر MCD می‌باشد.کروکس[1] و هائزبروئک[2] [۱۰] براساس مطالعات شبیه‌سازی، جدولی پیشنهادی برای $c_{\gamma,p}$ ارائه کرده‌اند. روسئو با همکاری ون دریسن[3] [۱۱] برآوردگر MCD را توسعه دادند و الگوریتمی ترکیبی را به نام FAST-MCD پیشنهاد کردند که بر اساس یک طرح تکرارشونده و برآوردگرهای MCD ساخته شده است. روشFAST-MCD قابلیت کنترل مجموعه داده‌های بزرگ در یک زمان معقول را دارد.

۳-۳. برآوردگر سالیوان[4]- وودال[5] (SW)

سالیوان و وودال [۱۲] دو الگوریتم با دو دیدگاه متفاوت ارائه کردند. آنها نشان دادند که نمودار T^2 با استفاده از ماتریس کواریانس نمونه، در شناسایی انتقال‌هایی در بردار میانگین مؤثر نمی‌باشد، از این رو پیشنهاد کردند که با استفاده از تفاوت بین دو مشاهده متوالی، $v_i = x_{i+1} - x_i$, $i = 1,2,\dots,m-1$ ، ماتریس کواریانس برآورد شود.مقدار S در رابطه(۱) باید با ماتریس کواریانس S_{sd} زیر جابجا شود[۱۳]:

$$S_{sd} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} V_i V_i' \qquad (۷)$$

این تکنیک را با نماد SW1 نشان می‌دهند. با این وجود وارگاس[6] [۱۴] در یک مطالعه مقایسه‌ای نشان داد که این روش در شناسایی تعداد زیادی از نقاط انحرافی مؤثر نمی‌باشد.ایده دوم سالیوان و وودال، که با SW2 نشان می‌دهیم، بدین گونه است با استفاده از $(p+1)$ مشاهده تصادفی، میانگین و ماتریس کواریانس محاسبه می‌شود. سپس برای تمام m مشاهدات، فاصله ماهالانوبیس محاسبه می‌شود. در ادامه، $(p+2)$ مشاهده که کمترین فاصله ماهالانوبیس را دارند، انتخاب می‌شود و این فرآیند با افزودن یک مشاهده تا مقدار ثابتی از m مشاهدات ادامه می‌یابد. این روش نیز در مقابل داده‌هایی با نقاط دورافتاده زیاد، آسیب‌پذیر است و همچنین به نمونه تصادفی اولیه وابسته می‌باشد[۱۵].

۴-۳ الگوریتم ترکیبی حداقل دترمینان کواریانس بازموزون[7] (RMCD)

ویلمز[8] و همکارانش [۱۶] آماره T^2_{RMCD} را پیشنهاد کردند که از جایگزینی برآوردگرهای موقعیت و پراکندگی حداقل دترمینان کواریانس بازموزون با میانگین و واریانس کلاسیک آماره T^2 هتلینگ بدست می‌آید . برآوردگرهای موقعیت و پراکندگی این روش بدین صورت تعریف می‌شوند:

$$\bar{X}_{RMCD} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i} \qquad (۸)$$

$$S_{RMCD} = c_{n,p} d^{n,p}_{\gamma,n} \frac{\sum_{i=1}^n w_i (X_i - \bar{X}_{RMCD})(X_i - \bar{X}_{RMCD})'}{\sum_{i=1}^n w_i} \qquad (۹)$$

که وزن‌ها بر اساس فاصله باثبات مشاهدات $D(X_i)$ زیر بنا می‌شوند:

$$D(X_i) = \sqrt{(X_i - \bar{X}_{MCD})' S^{-1}_{MCD} (X_i - \bar{X}_{MCD})} \qquad (۱۰)$$

وزن‌ها بدین گونه تعریف می‌شود:

$$w_i = \begin{cases} 1 & if D(X_i) \le q_\eta \\ 0 & otherwise \end{cases} \qquad (۱۱)$$

که q_η چندک η ام توزیع خی‌دو با p درجه آزادی می‌باشد. $d^{n,p}_{\gamma,\eta}$ ضریب نمونه متناهی است که توسط پیسون[9] و همکارانش[۱۷] پیشنهاد شده است.

[1] *Croux*
[2] *Haesbroeck*
[3] *Van Driessen*
[4] *Sullivan*
[5] *Woodall*
[6] *Vargas*
[7] *Reweighted Minimum Covariance Determinant*
[8] *Willems*
[9] *Pison*

۴. معرفی روش پیشنهادی

دیدگاه این روش پیشنهادی در راستای شناسایی نقاط پرت فاز اول نمودارهای کنترل کیفیت چندمتغیره ، براساس تکنیک خوشه‌بندی سلسله مراتبی[1] جهت بدست آوردن مجموعه‌ای باثبات از مشاهدات بنا شده است.فرضیات روش بدین صورت است که ابتدا مجموعه$X_{m\times p}$ متشکل از m مشاهده با بعد p و دارای توزیع نرمال چندمتغیره با میانگین μ و واریانس Σ ، در نظر گرفته می‌شود.

در مرحله‌ی اول بروی مجموعه داده $X_{m\times p}$ ، خوشه‌بندی سلسله مراتبی اعمال می‌شود. سپس بزرگترین خوشه از داده‌ها که کمترین ناهمسانی را باهم دارند انتخاب شده و میانگین و واریانس این خوشه به عنوان برآوردگر میانگین و ماتریس واریانس در آماره هتلینگ T^2 قرار داده می‌شود. حال مقدار T^2 برای تمامی m مشاهده محاسبه شده و پس از انتخاب حدود کنترلی، مشاهده‌ای که از حدبالای کنترل تجاوز می‌کند، به عنوان داده پرت و غیرمعمول معرفی می‌شود. در کل ساختار روش پیشنهادی را می‌توان در مراحل زیر خلاصه کرد:

- خوشه‌بندی سلسله مراتبی داده‌های اولیه
- انتخاب مجموعه داده‌های همسان
- محاسبه برآوردگر موقعیت و پراکندگی
- محاسبه T^2 هتلینگ براساس برآوردگر میانگین و ماتریس واریانس پیشنهادی
- انتخاب حدود کنترلی و شناسایی نقاط پرت

با توجه با اینکه روش‌های خوشه‌بندی سلسله مراتبی اطلاعات بیشتر و دقیق‌تری تولید می‌کنند ودر تحلیل داده‌ها با جزئیاتبیشتر کاربرد دارد و همچنین در ابتدا لازم نیست تعداد خوشه‌های مطلوب مشخص شود، لذا در الگوریتم پیشنهادی از تکنیک خوشه‌بندی سلسله مراتبی استفاده شده است.خروجی مرحله اول ماتریسی به نام Z با ابعاد $(m-1)\times 3$ می‌باشد که ستون اول و دوم این ماتریس شامل اندیس خوشه‌هایی است که با یکدیگر در ارتباط هستند. ستون سوم ماتریس برابر با فاصله‌ی گره‌های مشخص شده در ستون ۱ و۲ است. به عبارتی ماتریس Z ، درخت خوشه‌بندی سلسله مراتبی را ایجاد می‌کند.

پس از اینکه خوشه‌بندی اولیه از مجموعه داده‌ها انجام شد، باید از بین این خوشه‌ها، انتخاب داده‌های همسان صورت گیرد. بنابراین ماتریس Z که در مرحله قبل ساخته شده است، به عنوان داده‌های ورودی این مرحله استفاده می‌شود. جهت انتخاب مجموعه داده‌های همسان، پارامتری به نام ضریب ناهمسانی[2] تعریف شده است. ضریب ناهمسانی برای هر اتصال در درخت خوشه‌بندی سلسله مراتبی از مقایسه این اتصال با میانگین طول اتصال‌های دیگر در سطح مشابه، بدست می‌آید. اگر S_i برابر با مجموعه گره‌هایی باشد که با عمق کمتر در زیرشاخه گره $(m+i)$ قرار دارند، مقدار ناهمسانی برای هر گره $(m+i)$ با استفاده از میانگین و انحراف استاندارد فاصله گره‌های موجود در S_i ، محاسبه می‌شود. از اینرو ماتریس Y با ابعاد $(m-1)\times 4$ تشکیل شده است که توضیحات آن در جدول ۱ آمده است.

سطر i ام ماتریس Y، حاوی محاسبات صورت گرفته برای گره $(m+i)$، بدین صورت می‌باشد:

$$\begin{aligned} Y(i,1) &= mean\,[Z(s\_i,3)] \\ Y(i,2) &= std\,[Z(s\_i,3)] \\ Y(i,3) &= length\,[s\_i] \\ Y(i,4) &= \frac{Z(i,3)-Y(i,1)}{Y(i,2)} \end{aligned} \qquad (۱۲)$$

ضریب ناهمسانی برای هر گره، نشان‌دهنده میزان ناهمسانی این گره نسبت به سایر گره‌ها می‌باشد. به عبارتی به هر میزان این ضریب بالاتر باشد، نشان‌دهنده‌ی شباهت کمتر عناصر متصل شده در این گره می‌باشد.

جدول۱: فرم ماتریس Y

ستون	توضیحات
۱	میانگین فاصله اتصال‌های زیرشاخه گره$(m+i)$
۲	انحراف استاندارد فاصله اتصال‌های زیرشاخه گره$(m+i)$
۳	تعداد اتصال‌های زیرشاخه گره$(m+i)$
۴	ضریب ناهمسانی گره$(m+i)$

پس از اینکه ضریب ناهمسانی برای تمام گره‌ها محاسبه شد، گره‌ای که دارای بیشترین مقدار ضریب ناهمسانی می‌باشد، از مجموعه داده‌ها حذف می‌شود. این مجموعه را J می‌نامیم و فرض می‌کنیم دارای c عضو باشد. از آنجائیکه هرگره می‌تواند شامل چندین نقطه از مجموعه $X_{m\times p}$ باشد، بنابراین با حذف گره‌ای که بیشترین مقدار ضریب ناهمسانی را دارد، به عبارتی نقاطی از مجموعه را حذف کرده‌ایم که کمترین شباهت را با سایر نقاط دارند و نقاط باقی‌مانده دارای بیشترین شباهت نسبت بهم می‌باشند. پس از حذف گره با بیشترین ضریب ناهمسانی، نقاط باقی‌مانده را به عنوان داده‌های همسان می‌شناسیم و این مجموعه از نقاط را با Hc نشان می‌دهیم.برآوردگر موقعیت را با $\bar{X}_{Hc}$ و برآوردگر پراکندگی را با S_{Hc} نشان می‌دهیم. جهت برآورد موقعیت و پراکندگی از میانگین و ماتریس کواریانس عناصر موجود در مجموعه Hc استفاده می‌کنیم.

[1] *hierarchical clustering*
[2] *Inconsistency Coefficient*

$$\bar{X}_{Hc} = \frac{1}{m-c}\sum_{X_i \in Hc} X_i$$
$$S_{Hc} = \frac{1}{m-c-1}\sum_{X_i \in Hc}(X_i - \bar{X}_{Hc})(X_i - \bar{X}_{Hc})^t \quad (۱۳)$$

آماره پیشنهادی باثبات برای T^2 بر اساس برآوردگر (Hc) را با $T^2_{Hc,i}$ نشان می‌دهیم که بدین گونه تعریف می‌شود:

$$T^2_{Hc,i} = (X_i - \bar{X}_{Hc})\, S^{-1}_{Hc}(X_i - \bar{X}_{Hc})^t \quad (۱۴)$$
$$for\ i = 1,2,\dots,m$$

پس از محاسبه $T^2_{Hc,i}$ برای تمامی m مشاهده، نقاطی که $T^2_{Hc,i}$ مربوط به آن از حد کنترلی بالا تجاوز می‌کند، به عنوان نقطه پرت و غیرمعمول شناخته می‌شود و جهت محاسبه حدود کنترلی فاز دوم این نقاط از مجموعه مشاهدات باید حذف گردند.

۴-۱ الگوریتم روش پیشنهادی

فرآیند پیاده‌سازی نمودار کنترلی T^2_{Hc} را می‌توان بدین گونه خلاصه کرد:

مرحله۱:

۱. اندازه مشاهدات نمونه m و تعداد متغیرها p و سطح اطمینان $(1-\alpha)$ را مشخص کنید.
۲. در فاصله دوره‌ای مناسب فاز۱، مجموعه داده‌ها $X_{m\times p} = \{X_1, X_2, \dots, X_m\}$ را جمع‌آوری کنید.
۳. هر مشاهده به یک خوشه اختصاص داده شود. فاصله‌ی بین خوشه‌ها برابر با فاصله ماهالانوبیسی بین مشاهدات هر خوشه قرار داده شود.
۴. نزدیکترین جفت از خوشه را یافته و در یک خوشه ادغام - شود. i امین خوشه جدید تشکیل شده را گره $(m+i)$ نام-گذاری شود.
۵. فاصله بین خوشه جدید و خوشه‌های قدیمی محاسبه می-شود. فاصله بین یک خوشه و خوشه دیگر برابر با کوتاهترین فاصله از هر عضو یک خوشه به هر عضو خوشه دیگر است.
۶. گام‌های ۴ و ۵ تکرار شده تا زمانی که تمام مشاهدات در یک خوشه به سایز m خوشه‌بندی شوند.
۷. ماتریس Z با ابعاد $(m-1)\times 3$ تشکیل شود، بدین گونه که ستون اول و دوم این ماتریس شامل اندیس خوشه‌هایی است که با یکدیگر در ارتباط هستند. ستون سوم ماتریس برابر با فاصله‌ی گره‌های مشخص شده در ستون اول و دوم است.

مرحله ۲:

۱. با استفاده از ماتریس Z ، برای گره‌های ساخته شده در خوشه‌بندی سلسله مراتبی مرحله اول ، مقدار ضریب ناهمسانی با استفاده از رابطه (۱۲) محاسبه شود.
۲. ماتریس Y با ابعاد $(m-1)\times 4$ تشکیل شود، بدین گونه که ستون اول و دوم این ماتریس به ترتیب برابر با میانگین و انحراف استاندارد فاصله اتصال‌های زیرشاخه گره$(m+i)$ می‌باشد. ستون سوم ماتریس برابر با تعداد اتصال‌های زیرشاخه گره$(m+i)$ است. در ستون چهارم ماتریس ضریب ناهمسانی گره$(m+i)$ قرار داده شود.
۳. بیشترین مقدار ضریب ناهمسانی در ستون چهارم ماتریس Y را پیدا کرده و گره متناظر با این ضریب را J نامگذاری و تعداد عناصر مجموعه J را در متغیر c ذخیره می‌کنیم.
۴. مجموعه Hc با حذف مجموعه J از مجوعه داده‌های اولیه $X_{m\times p}$ ساخته شود.
۵. $\bar{X}_{Hc}$ و S_{Hc} برابر با میانگین و ماتریس واریانس مجموعه Hc قرار داده شود.
۶. با استفاده از برآوردگرهای $\bar{X}_{Hc}$ و S_{Hc} برای تمام m مشاهدات مقدار T^2_{Hc} را با استفاده از رابطه (۱۴) محاسبهو این نقطه را در نموداری کنترلی به همراه حدود کنترلی رابطه(۲) رسم کنید.
۷. نمودار را از وجود نقطه‌ای خارج از کنترل یا روندی در مشاهدات بررسی کرده و در صورت لزوم فرآیند را اصلاح کنید. مشاهداتی که از حدود کنترلی تجاوز کرده‌اند به عنوان نقاط پرت شناسایی شود.

۵. ارزیابی روش پیشنهادی

در این قسمت عملکرد روش پیشنهادی و توانایی آن در یافتن نقاط پرت مورد بررسی قرار می‌گیرد. از اینرو نخست با استفاده از یک مجموعه داده واقعی و ایجاد تغییراتی بروی آن، نتایج حاصل از روش پیشنهادی گزارش می‌گردد. در ادامه، با استفاده از شبیه-سازی‌های صورت گرفته جهت تولید نقاط پرت در ترکیب‌های مختلف از تعداد متغیرها (p) و مشاهدات(m) و مقایسه نتایج حاصل با تکنیک‌های هتلینگ کلاسیک و حداقل کواریانس دترمینان(MCD)، قابلیت و کارایی روش پیشنهادی مورد ارزیابی قرار می‌گیرد.

۵-۱. مجموعه داده واقعی کوئزنبری[1]

این مجموعه داده شامل ۱۱ شاخص کیفی(متغیر) اندازه‌گیری‌شده از ۳۰ محصول در طول فرآیند تولید یک کارخانه در سال ۲۰۰۱ می‌باشد. از مجموعه داده کوئزنبری، چندین محقق جهت ارزیابی روش پیشنهادیشان استفاده کرده‌اند. در این قسمت، دو متغیر از این مجموعه داده در نظر گرفته شده است که در ستون ۲ و ۳

[1] *Quesenberry*

جدول۲ نشان داده شده است. داده‌های جدول۲ از مرجع [۱۴] گرفته شده است. از این دو متغیر(X_1 و X_2) جهت مقایسه سه روش ایجاد نمودار کنترلی T^2 استفاده شده است. این سه روش عبارتند از: هتلینگ کلاسیک(*Usual*)، حداقل دترمینان کواریانس(*MCD*) و روش پیشنهادی(*HC*). با توجه به مباحث صورت گرفته، در سطح خطای نوع اول ($\alpha = 0.005$)، دو شاخص کیفی و ۳۰ مشاهده حد بالای این سه آماره در جدول۳ نشان داده شده است.

جدول ۲. مجموعه داده‌ها و آماره T^2 بدست آمده براساس تکنیک هتلینگ کلاسیک، حداقل دترمینان کواریانس و روش پیشنهادی

شماره مشاهدات	X_1	X_2	$T^2_{i,usual}$	$T^2_{i,MCD}$	$T^2_{i,Hc}$
۱	۰/۵۶۷	۶۰/۵۵۸	/۸۰۶۶	۰/۵۷۵۰	۰/۹۲۱۰
۲	۰/۵۳۸	۵۶/۳۰۳	۱۲/۹۴۵۴	۲۷/۶۱۲۳	۲۴/۹۵۹۷
۳	۰/۵۳۰	۵۹/۵۲۴	۰/۱۳۷۳	۰/۵۰۸۲	۰/۵۳۳۰
۴	۰/۵۶۲	۶۱/۱۰۲	۱/۸۳۷۵	۲/۰۴۶۹	۲/۶۱۳۷
۵	۰/۴۸۳	۵۹/۸۳۴	۱/۵۶۹۷	۰/۹۵۶۹	۱/۵۰۶۴
۶	۰/۵۲۵	۶۰/۲۲۸	۰/۳۳۰۱	۰/۱۶۳۵	۰/۳۱۳۱
۷	۰/۵۵۶	۶۰/۷۵۶	۰/۹۷۷۲	۰/۹۲۴۵	۱/۲۹۲۵
۸	۰/۵۸۶	۵۹/۸۲۳	۰/۹۰۴۵	۰/۹۳۰۳	۰/۹۲۸۴
۹	۰/۵۴۷	۶۰/۱۵۳	۰/۱۲۶۹	۰/۰۲۹۶	۰/۰۹۴۵
۱۰	۰/۵۳۱	۶۰/۶۴۰	۰/۸۰۰۸	۰/۷۷۶۶	۱/۰۳۳۸
۱۱	۰/۵۸۱	۵۹/۷۸۵	۰/۷۱۹۲	۰/۸۳۱۶	۰/۷۶۷۶
۱۲	۰/۵۸۵	۵۹/۶۷۵	۰/۹۰۹۷	۱/۱۵۹۵	۱/۰۳۳۴
۱۳	۰/۵۴۰	۶۰/۴۸۹	۰/۴۸۳۵	۰/۳۷۹۴	۰/۵۸۵۲
۱۴	۰/۴۵۸	۶۱/۰۶۷	۵/۲۴۱۳	۵/۲۳۶۸	۶/۱۰۱۲
۱۵	۰/۵۵۴	۵۹/۷۸۸	۰/۰۷۳۶	۰/۲۵۰۸	۰/۱۲۱۱
۱۶	۰/۴۶۹	۵۸/۶۴۰	۳/۵۳۳۷	۴/۲۳۲۷	۴/۹۴۸۸
۱۷	۰/۴۷۱	۵۹/۵۷۴	۲/۲۶۹۶	۱/۵۲۹۶	۲/۳۰۳۲
۱۸	۰/۴۵۷	۵۹/۷۱۸	۳,۲۴۴۲	۲/۰۴۵۲	۳/۱۵۱۵
۱۹	۰/۵۶۵	۶۰/۹۰۱	۱/۳۹۸۱	۱/۳۶۲۹	۱/۸۶۷۶
۲۰	۰/۶۶۴	۶۰/۱۸۰	۶/۸۳۲۶	۴/۷۹۸۴	۶/۵۶۸۷
۲۱	۰/۶۰۰	۶۰/۴۹۳	۱/۸۹۷۸	۱/۲۱۹۸	۱/۸۹۸۸
۲۲	۰/۵۸۶	۵۸/۳۷۰	۳/۳۵۶۴	۷/۲۵۸۰	۵/۹۵۲۴
۲۳	۰/۵۶۷	۶۰/۲۱۶	۰/۴۲۷۵	۰/۲۳۹۵	۰/۳۹۰۱
۲۴	۰/۴۹۶	۶۰/۲۱۴	۱/۱۸۳۸	۰/۷۵۱۲	۱/۱۴۶۰
۲۵	۰/۴۸۵	۵۹/۵۰۰	۱/۴۹۶۸	۱/۱۶۱۰	۱/۶۳۱۲
۲۶	۰/۵۷۳	۶۰/۰۵۲	۰/۴۸۴۳	۰/۳۵۹۶	۰/۴۳۹۵
۲۷	۰/۵۲۰	۵۹/۵۰۱	۰/۲۸۹۹	۰/۵۷۶۴	۰/۵۰۹۳
۲۸	۰/۵۵۶	۵۸/۴۷۶	۲/۰۶۳۵	۵/۳۰۲۱	۴/۲۶۵۴
۲۹	۰/۵۳۹	۵۸/۶۶۶	۱/۳۸۶۰	۳/۷۶۴۳	۳/۰۴۳۸
۳۰	۰/۵۵۴	۶۰/۲۳۹	۰/۲۴۰۴	۰/۱۰۳۸	۰/۲۱۸۴

جدول ۳. حدبالای کنترلی متناظر به روش‌های هتلینگ کلاسیک، حداقل دترمینان کواریانس و روش پیشنهادی

نوع تکنیک	حد بالای کنترلی (UCL)
هتلینگ کلاسیک	۱۰/۵۹۶
حداقل دترمینان کواریانس(MCD)	۱۰/۵۹۶
روش پیشنهادی(HC)	۹/۰۹۹

با مقایسه سه آماره بدست آمده با حدکنترلی متناظرشان، مشخص می‌شود که هر سه روش مشاهده شماره ۲ را به عنوان نقطه پرت شناسایی کرده‌اند.

اکنون به طور دلخواه، دو مشاهده پرت به مجموعه داده‌ها وارد می کنیم. مشاهدات ۱۴ و ۲۴ را به ترتیب به مقادیر (۶۵/۲۳۰ ، ۰/۸۸۰) و (۶۶/۰۸۰ ، ۰/۹۸۰) تغییر می‌دهیم. آماره T^2 حاصل از سه روش پس از این تغییرات، متناظر با سه مشاهده ۲،۱۴ و۲۴ در جدول۴ آورده شده است. همان‌طور که مشاهده می‌شود در حالی‌که آماره T^2 کلاسیک فقط مشاهدات ۲ و ۲۴ را به عنوان نقاط پرت شناسایی می‌کند ولی روش پیشنهادی(HC) و MCD سه نقطه ۲، ۱۴ و ۲۴ را به عنوان نقاط پرت شناسایی کرده است. به ترتیب با تغییر مشاهدات۱۸ و ۲۸ به مقادیر (۵۳/۱۸۰ ، ۰/۳۵۰) و (۵۰/۴۷۰ ، ۰/۴۱۰) تعداد مشاهدات پرت مجموعه داده قبلی را به ۵ نقطه افزایش می‌دهیم.

آماره چند متغیره بدست آمده از سه روش به ازای این ۵ مشاهده در جدول ۵ آورده شده است. همان‌طور که مشاهده می‌شود، آماره T^2 کلاسیک فقط مشاهدات ۲۴ و ۲۸ را به عنوان نقاط پرت شناسایی کرده است. روش MCD مشاهدات ۱۴ ، ۱۸، ۲۴ و ۲۸ (تمام نقاط پرت بجز مشاهده ۲) را به عنوان نقطه پرت شناخته است و این در حالی است که روش HC تمام پنج نقطه را به عنوان نقاط انحرافی شناسایی کرده است.

جدول ۴. آماره T^2 بدست آمده براساس تکنیک هتلینگ کلاسیک، حداقل دترمینان کواریانس و روش پیشنهادی با وجود ۳ نقطه انحرافی

شماره مشاهدات	$T^2_{i,usual}$	$T^2_{i,MCD}$	$T^2_{i,Hc}$
۲	۱۱/۱۴۵۳	۲۲/۱۷۸۹	۱۰/۹۰۹۰
۱۴	۹/۰۵۰۹	۵۶/۱۴۸۰	۱۹/۵۱۴۸
۲۴	۱۴/۶۲۵۷	۸۴/۲۴۹۷	۳۱/۵۹۷۲

جدول ۵. آماره T^2 بدست آمده براساس تکنیک هتلینگ کلاسیک، حداقل دترمینان کواریانس و روش پیشنهادی با وجود ۵ نقطه انحرافی

شماره مشاهدات	$T^2_{i,usual}$	$T^2_{i,MCD}$	$T^2_{i,Hc}$
۲	۲/۹۴۱۳	۹/۲۱۳۱	۱۳/۰۷۲۶
۱۴	۷/۴۳۸۶	۵۴/۹۴۲۴	۷۳/۸۴۷۲
۱۸	۵/۴۵۹۹	۶۰/۹۸۶۱	۵۵/۲۱۴۴
۲۴	۱۳/۴۹۸۲	۸۳/۶۷۶۹	۱۱۷/۷۷۱۵
۲۸	۱۵/۵۵۱۴	۱۱۶/۲۴۳۶	۹۱/۲۱۲۱

۵-۲. شبیه سازی و تحلیل نتایج

در این قسمت، عملکرد روش پیشنهادی HC را در مجموعه داده هایی با ابعاد مختلف مورد آزمایش قرار گرفته است و نتایج حاصل با تکنیک هتلینگ کلاسیک و حداقل حجم دترمینان کواریانس مقایسه شده است.

پس از مشخص شدن تعداد شاخص‌های کیفی(p) و تعداد مشاهدات (m)، ابتدا با استفاده از توزیع نرمال چندمتغیره با میانگین بردار صفر μ_0 و کواریانس ماتریس همانی Σ، یک مجموعه داده $m \times p$ ساخته می‌شود.بدون کم شدن از کلیت مسئله، فرض می‌کنیم مجموعه داده تولید شده تحت کنترل می‌باشد[۱۴] .سپس به طور تصادفی k نقطه از m نقطه انتخاب شده و k نقطه دیگر با توزیع نرمال چندمتغیره با میانگین μ_1 و کواریانس Σ جایگزین می‌کنیم. بدین ترتیب k نقطه پرت به مجموعه داده‌ها اضافه کرده‌ایم.میانگین جدید μ_1 ، بر حسب پارامتر عدم مرکزیت(ncp) بدست می‌آید که این پارامتر بدین صورت تعریف شده است:

این پارامتر نشان‌دهنده مقدار انتقال بردار میانگین خارج از کنترل(μ_1) از بردار میانگین تحت کنترل میانگین(μ_0) می‌باشد.به ازای هر m ، p و k مشخص شده، این سناریو ۱۰۰۰ مرتبه شبیه‌سازی شده است. در هر شبیه‌سازی عملکرد روش‌ها را تحت سه شرایط آزمایش کرده‌ایم:

۱. احتمال هشدار صحیح
۲. احتمال هشدار خطا
۳. مدت زمان شناسایی نقاط پرت

با در نظر گرفتن ترکیب‌های مختلف از تعداد شاخص‌های کیفی(p) ، تعداد مشاهدات (m) ، تعداد نقاط پرت(k) و پارامتر عدم مرکزیت (ncp) در مجموع ۱۰۵ سناریوهای متفاوتی ساخته شده است که ترکیب این مقادیر درجدول۶ آمده است.

جدول۶. ترکیب‌های در نظر گرفته از تعداد شاخص‌های کیفی(p)، تعداد مشاهدات (m) ، تعداد نقاط پرت(k) و پارامتر عدم مرکزیت (ncp) جهت ساخت سناریوهای متفاوت

p	m	k	ncp
۲	۳۰	۱،۳،۵،۷	۵،۱۰،۱۵،۲۰،۲۵،۳۰
۳	۳۰ ۵۰ ۱۰۰	۲،۴،۶	۵،۱۵،۲۵
۵	۳۰ ۵۰ ۱۰۰	۲،۵،۱۰	۵،۱۵،۲۵
۱۰	۳۰ ۵۰ ۱۰۰	۵،۱۰،۲۰	۵،۱۵،۲۵

جدول۷ نتایج حاصل از ترکیبی است که ۲ شاخص کیفی با ۳۰ مشاهده در نظر گرفته شده است و فقط یک نقطه پرت در مجموعه داده‌ها ایجاد شده است.

همان‌طور که مشاهده می‌شود هر سه تکنیک در شناسایی نقاط پرت تقریباً دارای عملکرد یکسانی هستند، به جز در حالتی که نقطه پرت از میانگین داده‌ها فاصله زیادی ندارد و ncp آن برابر با ۵ می‌باشد، تکنیک هتلینگ کلاسیک نسبت به دو تکنیک دیگر عملکرد نسبتاً ضعیف‌تری دارد. در این حالت روش پیشنهادی HC بیش از ۹۸ درصد از نقاط پرت را شناسایی کرده است. در تمامی حالت‌های این سناریو متوسط زمان تکنیک HC از دو روش دیگر کمتر بوده است.

جدول۸ احتمال هشدار صحیح سه تکنیک را درحالتی که ۲ شاخص کیفی با ۳۰ مشاهده در نظر گرفته شده است و تعداد نقاط پرت برابر با ۳، ۵ و ۷ می باشد، نشان می‌دهد. با افزایش نقاط پرت به ۳، ۵ و ۷ نقطه، عملکرد تکنیک هتلینگ کلاسیک به شدت افت پیدا می کند. به عنوان مثال در حالتی که ۳ نقطه پرت در مجموعه داده‌ها وجود دارد و پارامتر عدم مرکزیت برابر با ۳۰ می‌باشد، هتلینگ کلاسیک حدود ۵ ٪ از نقاط پرت را شناسایی کرده است و این در حالی است که روش پیشنهادی و تکنیک MCD تمامی نقاط پرت را شناسایی کرده است. از جمله موارد قابل توجه جدول۸ می‌توان به این مورد اشاره کرد که تکنیک HC در حالتی که $ncp = 5$ می‌باشد، از هر دو روش دیگر دارای احتمال هشدار صحیح بالاتری می‌باشد و در بقیه حالت‌ها نسبت به تکنیک MCD یا دارای عملکرد یکسانی بوده و یا دارای اختلاف بسیار کمی در احتمال هشدار صحیح می‌باشد. در جدول۹ که نشان‌دهنده احتمال هشدارهای خطا می‌باشـــد، در حالت $ncp = 5$ ، تکنیک HC نسبت به دو روش دیگر دارای احتمال هشدار خطای بیشتری بوده است.

جدول ۷. احتمال هشدار صحیح، احتمال هشدار خطا و زمان بدست آمده برای ۲ شاخص کیفی، ۳۰ مشاهده و ۱ نقطه پرت

p = ،۲ m = ،۳۰ K = ۱

پارامتر عدم مرکزیت	نوع تکنیک	احتمال هشدار صحیح	احتمال هشدار خطا	زمان (ثانیه)
۵	Usual	۰/۷۶۰۰	۰/۰۰۰۱	
	HC	۰/۹۸۸۰	۰/۰۰۵۰	۰/۰۰۵۳
	MCD	۰/۸۸۸۰	۰/۰۰۰۳	۰/۸۳۷۴
۱۰	Usual	۱/۰۰۰۰	۰/۰۰۰۰	
	HC	۱/۰۰۰۰	۰/۰۰۰۶	۰/۰۰۳۶
	MCD	۱/۰۰۰۰	۰/۰۰۰۳	۰/۴۱۰۰
۱۵	Usual	۱/۰۰۰۰	۰/۰۰۰۰	
	HC	۱/۰۰۰۰	۰/۰۰۰۸	۰/۰۰۳۶
	MCD	۱/۰۰۰۰	۰/۰۰۰۳	۰/۴۰۹۲
۲۰	Usual	۱/۰۰۰۰	۰/۰۰۰۱	
	HC	۱/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۲۹
	MCD	۱/۰۰۰۰	۰/۰۰۰۶	۰/۴۰۹۷
۲۵	Usual	۱/۰۰۰۰	۰/۰۰۰۰	
	HC	۱/۰۰۰۰	۰/۰۰۰۶	۰/۰۰۳۶
	MCD	۱/۰۰۰۰	۰/۰۰۰۵	۰/۴۰۹۷
۳۰	Usual	۱/۰۰۰۰	۰/۰۰۰۱	
	HC	۱/۰۰۰۰	۰/۰۰۰۵	۰/۰۰۳۶
	MCD	۱/۰۰۰۰	۰/۰۰۰۴	۰/۴۱۰۰

به عنوان مثال در حالت $k = 3$ از بین ۱۰۰۰۰ نقطه، احتمال دارد ۵ نقطه را هشدار خطا دهد و به عنوان نقطه پرت شناسایی کند ولی با افزایش ncp نسبت هشدارهای خطای روش پیشنهادی کاهش یافته و در مواردی حتی به صفر نیز رسیده است.در سناریو بعدی، بعد مسئله به $p = 3$ افزایش یافته است. نتایج حاصل برای تعداد مشاهدات $m = 30$ ، $m = 50$ و $m = 100$ در جدول۱۰ آورده شده است.با افزایش تعداد مشاهدات و تعداد نقاط پرت، احتمال هشدار صحیح تمام تکنیک-ها تا حدودی کاهش یافته است ولی نکته قابل ذکر عملکرد بهتر

تکنیک HC در انتقال‌های کوچک ($ncp = 5$) می‌باشد. در اکثر موارد تکنیک هتلینگ کلاسیک نسبت به روش پیشنهادی HC و MCD دارای عملکرد بسیار ضعیفی می‌باشد.

جدول۸. احتمال هشدار صحیح بدست آمده برای ۲ شاخص کیفی، ۳۰ مشاهده و ۳، ۵ و ۷ نقطه پرت

p = ۲، m = ۳۰					
پارامتر عدم مرکزیت	نوع تکنیک	احتمال هشدار صحیح			
		۳K =	۵K =	۷K =	
۵	Usual	۰/۰۵۲۷	۰/۰۰۶۲	۰/۰۰۱۹	
	HC	۰/۹۵۹۷	۰/۸۵۱۴	۰/۶۹۷۶	
	MCD	۰/۸۱۲۰	۰/۶۲۶۴	۰/۳۴۹۱	
۱۰	Usual	۰/۰۷۰۳	۰/۰۰۳۶	۰/۰۰۱۴	
	HC	۰/۹۹۸۷	۰/۹۶۹۶	۰/۹۸۱۶	
	MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	
۱۵	Usual	۰/۰۵۶۰	۰/۰۰۵۶	۰/۰۰۱۶	
	HC	۱/۰۰۰۰	۰/۹۹۷۶	۰/۹۹۷۴	
	MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	
۲۰	Usual	۰/۰۵۵۰	۰/۰۰۳۸	۰/۰۰۲۶	
	HC	۱/۰۰۰۰	۱/۰۰۰۰	۰/۹۹۸۳	
	MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	
۲۵	Usual	۰/۰۶۲۰	۰/۰۰۳۴	۰/۰۰۰۹	
	HC	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	
	MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	
۳۰	Usual	۰/۰۵۶۳	۰/۰۰۳۸	۰/۰۰۱۱	
	HC	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	
	MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰	

تکنیک HC نسبت به روش MCD دارای عملکرد بسیار نزدیک بهم بوده به طوری که با افزایش پارامتر عدم مرکزیت دو تکنیک HC و MCD توانسته است به طور۱۰۰٪ تمامی نقاط پرت را شناسایی کند. جدول۱۱ احتمال هشدار صحیح برای شاخص کیفی $p = 5$ با تعداد مشاهدات ۳۰، ۵۰ و ۱۰۰ را نشان می-دهد. در حالت‌هایی که پارامتر عدم مرکزیت برابر با ۵ می‌باشد و تعداد نقاط انحرافی افزایش می‌یابد، تکنیک هتلینگ کلاسیک کارایی بسیار ضعیفی از خود نشان می‌دهد.

در این حالت تکنیک پیشنهادی HC در مقایسه با تکنیک MCD عملکرد بهتری دارد. در حالت‌هایی که تعداد نقاط پرت برابر با $k = 2$ و $k = 5$ می‌باشد، روش HC و MCD دارای عملکرد بسیار نزدیکی به هم می‌باشند به طوری که با افزایش ncp نسبت هشدار صحیح نزدیک به ۱ و یا برابر با ۱ شده است.در حالتی که تعداد نقاط پرت به ۱۰ نقطه افزایش می یابد و $ncp = 5$ است ، تکنیک HC از تکنیک MCD بهتر عمل کرده است.

جدول۹. احتمال هشدار خطا بدست آمده برای ۲ شاخص کیفی، ۳۰ مشاهده و ۳، ۵ و ۷ نقطه پرت

p = ۲، m = ۳۰					
پارامتر عدم مرکزیت	نوع تکنیک	احتمال هشدار خطا			
		K = ۳	K = ۵	K = ۷	
۵	Usual	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	
	HC	۰/۰۰۰۹	۰/۰۰۰۵	۰/۰۰۰۷	
	MCD	۰/۰۰۰۶	۰/۰۰۰۲	۰/۰۰۰۳	
۱۰	Usual	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۲	
	HC	۰/۰۰۰۱	۰/۰۰۰۶	۰/۰۰۰۲	
	MCD	۰/۰۰۰۱	۰/۰۰۰۰	۰/۰۰۰۰	
۱۵	Usual	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۲	
	HC	۰/۰۰۰۰	۰/۰۰۰۲	۰/۰۰۰۶	
	MCD	۰/۰۰۰۲	۰/۰۰۰۰	۰/۰۰۰۰	
۲۰	Usual	۰/۰۰۰۲	۰/۰۰۰۲	۰/۰۰۰۰	
	HC	۰/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۰۳	
	MCD	۰/۰۰۰۱	۰/۰۰۰۰	۰/۰۰۰۰	
۲۵	Usual	۰/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۰۰	
	HC	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۲	
	MCD	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	
۳۰	Usual	۰/۰۰۰۱	۰/۰۰۰۲	۰/۰۰۰۲	
	HC	۰/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۰۳	
	MCD	۰/۰۰۰۶	۰/۰۰۰۱	۰/۰۰۰۰	

جدول ۱۲ احتمال هشدار صحیح برای شاخص کیفی $p = 10$ با تعداد مشاهدات $m = 30$ ، $m = 50$ و $m = 100$ را نشان می‌دهد. مشابه حالت‌های قبل، در مواردی که $ncp = 5$ می‌باشد، تکنیک HC از سایر تکنیک‌ها بهتر عمل می‌کند. در سایر حالت‌های این سناریو، تکنیک هتلینگ کلاسیک قادر به شناسایی تمام نقاط انحرافی نبوده است و بسیار ضعیف عمل کرده است، به طوری که احتمال هشدار صحیح این تکنیک تقریباً نزدیک به صفر می‌باشد.در اکثر حالت‌ها تکنیک HC تقریباً دارای عملکرد یکسان و یا حتی بهتری نسبت به MCD می‌باشد، به جز در حالت‌های ($m = 30, k = 5$) ، ($m = 50, k = 10$) و ۲، ۵ و ۱۰ نقطه پرت ($m = 100, k = 20$) عملکرد تکنیک MCD به طور قابل ملاحظه‌ای از عملکرد تکنیک HC بهتر می-باشد. جدول ۱۳ مدت زمان رسیدن به جواب دو تکنیک HC و MCD در سناریوهای جدول ۱۲ را نشان می‌دهد. در این حالت مشابه تمامی حالت‌های قبل، تکنیک پیشنهادی HC از میانگین

مدت زمان کمتری در شناسایی نقاط پرت برخوردار می‌باشد.همان‌طور که در جدول ۶ نشان داده شد، در مجموع ۱۰۵ ترکیب مختلف از تعداد شاخص کیفی، تعداد مشاهدات و تعداد نقاط پرت جهت شبیه‌سازی در نظر گرفته شده است.

جدول ۱۰. احتمال هشدار صحیح بدست آمده برای ۳ شاخص کیفی، ۳۰، ۵۰ و ۱۰۰ مشاهده و ۲، ۴ و ۶ نقطه پرت

p = ۳					
تعداد مشاهدات	پارامتر عدم مرکزیت	نوع تکنیک	احتمال هشدار صحیح K = ۲	K = ۴	K = ۶
m = ۳۰	۵	Usual	۰/۱۷۳۰	۰/۰۰۹۸	۰/۰۰۳۰
		HC	۰/۹۵۸۵	۰/۸۵۵۳	۰/۶۶۵۸
		MCD	۰/۸۴۳۵	۰/۷۱۷۸	۰/۵۵۱۲
	۱۵	Usual	۰/۷۲۵۵	۰/۰۰۹۳	۰/۰۰۲۵
		HC	۱/۰۰۰۰	۰/۹۹۷۳	۰/۹۹۶۵
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
	۲۵	Usual	۰/۹۰۰۰	۰/۰۱۱۵	۰/۰۰۲۵
		HC	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
m = ۵۰	۵	Usual	۰/۴۲۵۵	۰/۰۷۷۸	۰/۰۱۴۵
		HC	۰/۹۴۹۰	۰/۸۹۷۳	۰/۷۵۸۳
		MCD	۰/۸۴۷۰	۰/۷۸۴۰	۰/۶۷۲۲
	۱۵	Usual	۱/۰۰۰۰	۰/۱۴۸۸	۰/۰۱۴۳
		HC	۱/۰۰۰۰	۰/۹۹۷۳	۰/۹۹۹۲
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
	۲۵	Usual	۱/۰۰۰۰	۰/۱۴۵۸	۰/۰۱۵۷
		HC	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
m = ۱۰۰	۵	Usual	۰/۶۹۳۰	۰/۳۷۸۸	۰/۱۶۱۲
		HC	۰/۹۳۷۵	۰/۹۰۵۵	۰/۸۳۱۷
		MCD	۰/۸۶۵۰	۰/۸۱۶۰	۰/۷۷۵۲
	۱۵	Usual	۱/۰۰۰۰	۰/۹۹۸۰	۰/۶۲۶۳
		HC	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
	۲۵	Usual	۱/۰۰۰۰	۱/۰۰۰۰	۰/۷۹۷۸
		HC	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰

جدول ۱۴ ارزیابی کلی از عملکرد روش پیشنهادی در مقایسه با تکنیک حداقل دترمینان کواریانس در مجموع ۱۰۵ ترکیب مختلف ارائه می‌دهد. نتایج حاصل از شبیه‌سازی را در سه حالت می‌توان دسته‌بندی کرد. حالت اول نشان دهنده‌ی این می‌باشد که در چند درصد از ۱۰۵ ترکیب در نظر گرفته، روش پیشنهادی بیشتر یا برابر با تکنیک حداقل دترمینان کواریانس، به طور صحیح اعلام هشدار کرده است. حالت دوم نشان‌دهنده‌ی این می‌باشد که چند درصد از ۱۰۵ ترکیب، تکنیک حداقل دترمینان کواریانس به طور صحیح بیشتر از روش پیشنهادی هشدار داده ولی این اختلاف در هشدار حداکثر تا سطح ۰/۰۵ بوده است. حالت سوم نشان می‌دهد که به چه تعداد روش پیشنهادی کمتر از تکنیک حداقل دترمینان کواریانس هشدار صحیح داده و این اختلاف در هشدار بیشتر از ۰/۰۵ بوده است.

جدول ۱۱. احتمال هشدار صحیح بدست آمده برای۵ شاخص کیفی، ۳۰، ۵۰ و ۱۰۰ مشاهده

p = ۵					
تعداد مشاهدات	پارامتر عدم مرکزیت	نوع تکنیک	احتمال هشدار صحیح K = ۲	K = ۴	K = ۶
m = ۳۰	۵	Usual	۰/۰۸۳۰	۰/۰۰۲۶	۰/۰۰۰۶
		HC	۰/۹۲۱۰	۰/۶۹۸۰	۰/۳۷۸۳
		MCD	۰/۸۳۳۵	۰/۵۸۹۶	۰/۰۵۱۶
	۱۵	Usual	۰/۳۵۹۰	۰/۰۰۳۶	۰/۰۰۱۳
		HC	۱/۰۰۰۰	۰/۹۷۴۸	۰/۶۲۷۰
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۰/۰۴۴۱
	۲۵	Usual	۰/۴۰۰۵	۰/۰۰۳۸	۰/۰۰۱۰
		HC	۱/۰۰۰۰	۰/۹۹۵۴	۰/۷۳۹۷
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۰/۰۴۱۰
m = ۵۰	۵	Usual	۰/۲۶۹۵	۰/۰۱۸۰	۰/۰۰۲۲
		HC	۰/۹۲۳۵	۰/۷۷۹۴	۰/۴۹۲۱
		MCD	۰/۷۹۹۵	۰/۷۱۰۲	۰/۳۵۴۹
	۱۵	Usual	۰/۹۹۳۵	۰/۰۲۳۰	۰/۰۰۲۴
		HC	۱/۰۰۰۰	۰/۹۹۶۶	۰/۹۶۱۶
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
	۲۵	Usual	۱/۰۰۰۰	۰/۰۲۰۰	۰/۰۰۲۰
		HC	۱/۰۰۰۰	۱/۰۰۰۰	۰/۹۸۲۵
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
m = ۱۰۰	۵	Usual	۰/۵۵۵۰	۰/۱۴۶۲	۰/۰۱۶۹
		HC	۰/۹۲۱۰	۰/۸۴۶۶	۰/۶۲۹۹
		MCD	۰/۸۱۶۵	۰/۷۳۶۰	۰/۵۹۲۷
	۱۵	Usual	۱/۰۰۰۰	۰/۶۸۸۸	۰/۰۲۲۹
		HC	۱/۰۰۰۰	۱/۰۰۰۰	۰/۹۹۷۴
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
	۲۵	Usual	۱/۰۰۰۰	۰/۸۴۶۸	۰/۰۲۳۶
		HC	۱/۰۰۰۰	۱/۰۰۰۰	۰/۹۹۹۱
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰

همان‌طور که از جدول۱۴ مشاهده می‌شود، روش پیشنهادی توانسته است در ۷۲/۳ درصد از ۱۰۵ ترکیب کل، بیشتر یا برابر با تکنیک حداقل دترمینان کواریانس به طور صحیح هشدار دهد. در ۲۲ درصد از مواقع روش پیشنهادی و تکنیک حداقل دترمینان کواریانس در اعلام هشدار صحیح حداکثر ۰/۰۵ با یکدیگر اختلاف داشته است. در مجموع می‌توان گفت روش پیشنهادی در ۹۴/۳ درصد از مواقع یا عملکرد بهتری نسبت به تکنیک حداقل

دترمینان کواریانس داشته و یا عملکرد نسبتاً مشابهی داشته است.در ۵/۷ درصد از ۱۰۵ ترکیب، روش پیشنهادی کمتر از تکنیک حداقل دترمینان کواریانس هشدار صحیح داده است.

جدول ۱۲. احتمال هشدار صحیح بدست آمده برای۱۰ شاخص کیفی، ۳۰، ۵۰ و ۱۰۰ مشاهده و ۵، ۱۰ و ۲۰ نقطه پرت

p = ۱۰					
تعداد مشاهدات	پارامتر عدم مرکزیت	نوع تکنیک	احتمال هشدار صحیح		
			K = ۵	K = ۱۰	K = ۲۰
m = ۳۰	۵	Usual	۰/۰۰۲۶	۰/۰۰۰۹	۰/۰۰۰۷
		HC	۰/۵۷۱۸	۰/۳۹۵۵	۰/۲۹۷۵
		MCD	۰/۳۷۱۴	۰/۱۷۹۴	۰/۱۴۲۲
	۱۵	Usual	۰/۰۰۲۰	۰/۰۰۱۳	۰/۰۰۰۸
		HC	۰/۷۱۹۴	۰/۴۲۳۱	۰/۲۹۱۶
		MCD	۰/۹۸۳۸	۰/۱۸۶۲	۰/۱۴۰۶
	۲۵	Usual	۰/۰۰۲۴	۰/۰۰۱۷	۰/۰۰۰۸
		HC	۰/۷۳۷۲	۰/۴۲۳۰	۰/۲۷۹۶
		MCD	۱/۰۰۰۰	۰/۱۸۷۳	۰/۱۴۳۷
m = ۵۰	۵	Usual	۰/۰۰۹۶	۰/۰۰۱۴	۰/۰۰۱۲
		HC	۰/۶۷۵۸	۰/۴۴۶۴	۰/۲۶۶۲
		MCD	۰/۶۹۳۲	۰/۱۸۶۸	۰/۰۶۵۵
	۱۵	Usual	۰/۰۱۴۸	۰/۰۰۲۸	۰/۰۰۱۲
		HC	۰/۹۵۹۰	۰/۶۲۹۴	۰/۲۸۶۵
		MCD	۱/۰۰۰۰	۰/۹۹۷۲	۰/۰۶۶۵
	۲۵	Usual	۰/۰۱۷۰	۰/۰۰۲۴	۰/۰۰۱۳
		HC	۰/۹۶۶۸	۰/۶۴۴۴	۰/۲۹۴۰
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۰/۰۶۳۶
m = ۱۰۰	۵	Usual	۰/۰۶۸۲	۰/۰۱۱۴	۰/۰۰۲۰
		HC	۰/۷۹۹۲	۰/۵۷۲۰	۰/۲۹۵۹
		MCD	۰/۶۵۰۰	۰/۴۵۵۲	۰/۰۹۰۲
	۱۵	Usual	۰/۲۷۳۸	۰/۰۱۵۰	۰/۰۰۲۰
		HC	۱/۰۰۰۰	۰/۹۶۹۴	۰/۶۵۹۸
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰
	۲۵	Usual	۰/۳۳۶۸	۰/۰۱۵۹	۰/۰۰۳۰
		HC	۱/۰۰۰۰	۰/۹۸۷۲	۰/۷۰۳۵
		MCD	۱/۰۰۰۰	۱/۰۰۰۰	۱/۰۰۰۰

جدول۱۴. تعداد و درصد سه حالت در نظر گرفته از نتایج شبیه‌سازی

حالت	تعداد	درصد	
$True_{Signal(HC)} \geq True\_Signal(MCD)$	۷۶	٪ ۷۲/۳	٪ ۹۴/۳
$\left[True_{Signal(MCD)} - True\_Signal(HC)\right] < 0.05$	۲۳	٪ ۲۲	
$True_{Signal(MCD)} \gg True\_Signal(HC)$	۶	٪ ۵/۷	
مجموع	۱۰۵		

۶. نتیجه‌گیری

تاکنون آماره‌های باثبات زیادی در راستای شناسایی نقاط پرت مجموعه داده‌های چندمتغیره ارائه شده است، از جمله می‌توان به برآوردگر باثبات حداقل حجم بیضیوار(MVE)، حداقل دترمینان واریانس (MCD)، برآوردگر سالیوان-وودال (SW) و حداقل دترمینان واریانس بازموزون (RMCD) اشاره نمود. همه‌ی روش-ها به دنبال بدست آوردن برآوردی از میانگین و ماتریس کوریانس مجموعه داده‌های چندمتغیره می‌باشند که نسبت به نقاط غیرمعمول ودورافتاده، مقاوم بوده و تحت تأثیر این نقاط قرار نگیرند. دیدگاه روش پیشنهادی در راستای شناسایی نقاط پرت فاز اول نمودارهای کنترل کیفیت چندمتغیره، براساس تکنیک خوشه‌بندی سلسله مراتبی جهت بدست آوردن مجموعه‌ای باثبات از مشاهدات بنا شده است.عملکرد روش پیشنهادی به دو صورت مورد ارزیابی قرار گرفت. در مرحله‌ی اول با استفاده از یک مجموعه داده واقعی و ایجاد تغییراتی بروی آن، روش پیشنهادی با دو تکنیک هتلینگ کلاسیک و حداقل دترمینان کواریانس مقایسه شد. نتایج نشان داد با افزایش تعداد نقاط پرت به ۵ نقطه، فقط روش پیشنهادی هر ۵ مشاهده پرت را شناسایی کرد و این در حالی بود که تکنیک حداقل دترمینان کواریانس ۴ نقطه و تکنیک هتلینگ کلاسیک فقط ۲ مشاهده پرت را شناسایی کردند. در مرحله‌ی دوم با در نظر گرفتن ترکیب‌های مختلفی از تعداد شاخص‌های کیفی، تعداد مشاهدات و تعداد نقاط پرت، مجموعه داده‌های متفاوتی از اعداد تصادفی تولید گردید و تحت سه شرایط احتمال هشدار صحیح، احتمال هشدار خطا و زمان رسیدن به جواب عملکرد سه تکنیک بررسی شد. پیرامون نتایج بدست آمده تحلیل‌هایی ارائه شد و نتایج بدست آمده از الگوریتم پیشنهادی با دو تکنیک دیگر مقایسه شد. نتایج نشان دادند که با افزایش تعداد نقاط پرت کارایی تکنیک هتلینگ کلاسیک به شدت کاهش می‌یابد و حداکثر با وجود یک نقطه پرت این تکنیک قابل اعتماد می‌باشد. برای هر دو برآوردگر پیشنهادی و حداقل دترمینان کواریانس این موضوع صادق است که با افزایش تعداد شاخص‌های کیفی، تعداد نقاط انحرافی که شناسایی می‌شوند، کاهش می‌یابد. در تمامیه ۱۰۵ ترکیب در نظر گرفته شده، میانگین مدت زمان رسیدن به جواب تکنیک HC از تکنیک MCD بسیار کمتر بوده است. در نهایت با یک دید کلی عملکرد روش پیشنهادی و تکنیک حداقل دترمینان کواریانس در ۱۰۵ ترکیب مقایسه گردید که نتایج نشان دهنده‌ی این می‌باشد روش پیشنهادی در بیش از در۷۲ درصد از ۱۰۵ ترکیب کل، بیشتر یا برابر با تکنیک حداقل دترمینان کواریانس به طور صحیح هشدار داده است و در مجموع در بیش از ۹۴ درصد از مواقع یا عملکرد بهتری نسبت به تکنیک حداقل دترمینان کواریانس داشته و یا

عملکرد نسبتاً مشابهی داشته است. در ۶ ترکیب از ۱۰۵ ترکیب، روش پیشنهادی نسبت به تکنیک حداقل دترمینان کواریانس عملکرد نسبتاً ضعیف‌تری داشته است.

مراجع

[1] Juran, J.M., Gr, F.M., "*Quality Planning and Analysis*". New York: McGraw, 1980.

[2] Lorenzen, T.J., Vance, L.C., "*The Economic Design of Control Charts: a Unified Approach.*" Technometrics, Vol. 28, 1986, pp. 3-10.

[3] Mason, R.L., Young, J.C. "*Multivariate Statistical Process control with industrial application.*" Virginia: ASA SIAM, 1946.

[4] Duncan, A.J., "*Quality Control and Industrial Statistics*." Irwin: Homewoo, 1986.

[5] Jensen, W.A., Birch, J.B., Woodall, W.H.,"*High Breakdown Estimation Methods for Phase I Multivariate Control Charts.*" Quality and reliability Engineering International, Vol. 23, 2007, pp. 615–629.

[6] Peña, D., Prieto, F.J., "*Multivariate Outlier Detection and Robust Covariance Matrix Estimation.*" Technometrics, Vol. 43, No. 3, 2001, pp. 286-310.

[7] Rousseeuw, P.J., "*Least Median of Squares Regression*." Journal of the American Statistical Association, Vol. 79, No. 388, 1984, pp. 871–880.

[8] Davies, P.L., "*Asymptotic Behavior of S-Estimators of Multivariate Location Parameters and Dispersion Matrices."* The Annals of Statistics, Vol. 15, 1987, pp. 1269-1292.

[9] Lopuha¨a, H.P., Rousseeuw, P.J., "*Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices*." The Annals of Statistics, Vol. 19, 1991, pp. 229–248.

[10] Croux, C., Haesbroeck, G., "*Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator*",Journal of Multivariate Analysis, Vol. 71, 1999, pp. 161-190.

[11] Rousseeuw, P.J., Driessen, K.V., "*A Fast Algorithm for the Minimum Covariance Determinant Estimator*", Technometrics, Vol. 41, No. 3, 1999, PP. 212-223.

[12] Sullivan, J.H., Woodall, W.H., "*A Comparison of Multivariate Control Charts for Individual Observations.*" Journal of Quality Technology, Vol. 28, No. 4, 1996, pp. 398–408.

[13] Midi, H., Shabbak, A., "*Robust Multivariate Control Charts to Detect Small Shifts in Mean*" Mathematical Problems in Engineering, 2011, pp. 001-019.

[14] Vargas, J.A.N., "*Robust Estimation in Multivariate Control Charts for Individual Observations.*" Journal of Quality Technology, Vol. 35, No. 4, 2003, pp. 367–376.

[15] Stefatos, G. & Ben Hamza, A. "*Fault Detection using Robust Multivariate Control Chart.*" Expert Systems with Applications, Vol. 6, 2003, pp. 5888–5894.

[16] Willems,G., Pison, G., Rousseeuw, P.J., Van Aelst, S.A., "*Robust Hotelling Test*", Metrika, Vol. 55, 2002, pp. 125–138.

[17] Pison, G., Van Alest, S., Willems, G., "*Small Sample Corrections for LTS and MCD*." Metrika, Vol. 55, 2002, pp. 111–123.