کنترل فرآیند آماری و اثر تحلیلی کشف عامل مشخصه با روش تشخیص فضای ویژه E

عبدالرضا قاسمی[۱]، حامد آقاپناه[۲]

چکیده

کنترل فرآیند آماری مجموعه‌ای قدرتمند از ابزار حل مشکل است که مهم‌ترین هدف آن بهبود و اصلاح کیفیت تولیدات با کاهش تغییرپذیری در فرآیند می‌باشد. امروزه رشد تکنولوژی، باعث دسترسی سریع به داده‌ها و منابع شده است و تولید در حجم گسترده و با فراوانی بالا انجام می‌گیرد. در این فرآیندها همبستگی بین متغیرها مورد انتظار است. بنابراین کنترل فرآیند آماری چند متغیره به عنوان مهم‌ترین ابزار برای افزایش دقت در فرآیند اهمیت پیدا می‌کند. هدف از این مقاله ارائه برخی شیوه‌ها برای تفسیر و شناسایی متغیر های عامل انحراف در حالات خارج از کنترل می‌باشد. مهم‌ترین بخش آن کنترل فرآیند آماری چند متغیره و اثر تحلیلی استفاده از روش‌های شناسایی عامل ویژه است که به تغییرات جهت دار توجه دارد. در این راستا، ماتریس فضای ویژه E را معرفی کرده و به بررسی ویژگی‌های آن در قالب ارائه و اثبات قضایایی می‌پردازیم.

واژه های کلیدی: کنترل فرآیند آماری چند متغیره، تجزیه اجزای اصلی، کوواریانس تعمیم یافته، توزیع ویشارت، ماتریس فضای ویژهE.

۱– مقدمه

کنترل فرآیند آماری(SPC)[۳]، یک شاخه مهم برای کنترل فرآیندتولیدات و خدمات از سال ۱۹۲۰ می‌باشد که هدف آن رسیدن به کیفیت بالاتر در تولید و پایین آوردن هزینه‌ها به وسیله کاهش تولیدات نامرغوب است. یکی از بزرگ‌ترین ابزارها در جهت نیل به این مقصود، نمودار های کنترل فرآیند آماری توسعه یافته والترشوهارت[۴] می‌باشد که معتقد بود "پراکندگی در کیفیت تولیدات از دو منبع نشئت می‌گیرد. نخست عوامل تصادفی که ذاتی فرآیند تولید است و دیگری عوامل غیر تصادفی و قابل تشخیص." شوهارت بر این مبنا نموداری ارائه کرد که تحت کنترل بودن یا نبودن فرآیند آماری را به سادگی نشان می‌داد. از آنجا که طرح‌های کنترل شوهارت تنها زمانی که یک مشخصه کیفی مورد بررسی واقع می‌شود، کاربرد دارد و از آن نمی‌توان برای کنترل همزمان چند مشخصه کیفی فرآیند استفاده نمود، موضوع کنترل فرآیند

[1] دانشجوی کارشناسی ارشد آمار، دانشگاه آزاد واحد مشهد

[2] دانشجوی کارشناسی ارشد مهندسی پزشکی، دانشگاه تربیت مدرس

[3] Statistical process control

[4] Walter A . Shewhart

آماری چند متغیره (MSPC)[1]و روش‌های آن اهمیت خاصی پیدا می‌کند. به طور کلی چهار روش معروف برای کنترل فرآیند آماری چند متغیره وجود دارد که عبارتند از: روش مربع کای[2]، $T^۲$ هتلینگ[3]، میانگین متحرک موزون نمایی چند متغیره (MEWMA)[4] و جمعی تجمعی چند متغیره (MCUSUM)[5] که مهم‌ترین آن، روش $T^۲$ هتلینگ می‌باشد. این روش‌ها همبستگی بین متغیرهای کیفی را در نظر می‌گیرد. در بخش‌های اول و دوم این مقاله به بررسی تحلیلی کنترل فرآیند آماری و کنترل فرآیند آماری چند متغیره و روش‌های آن خواهیم پرداخت. در بخش نهایی عملیات فرآیند، به معرفی فضای ویژه E در قالب مثال و ارائه اثبات قضایا می پردازیم.

۲–کنترل فرآیند آماری تک متغیره(SPC)

کنترل فرآیند آماری به معنای کاربرد اصول و مهارت‌های آماری در تمام مراحل تولیدی، نگهداری و خدمات به منظور ارضای تقاضا به صورت اقتصادی است[۱]. مونتگومری[6] معتقد است کنترل فرآیند آماری در ایجاد ثبات در فرآیند و بهبود کارایی آن از طریق کاهش تغییرات نقش مفیدی ایفا می‌کند[۲] و در این راستا هفت ابزار برای آن تعریف نمود که در صنعت به صورت گسترده مورد توجه قرار گرفت. در برخی منابع از این ابزار با عنوان ابزار عالی هفت‌گانه یاد می‌کنند، که عبارتند از:

۱) هیستوگرام

۲) برگه کنترل

۳)نمودار پارتو

٤) نمودار علت و معلول

٥) نمودار تمرکز نقص‌ها

٦) نمودار پراکندگی

۷) نمودار کنترل .

به اعتقاد اکثر کارشناسان از جمله والتر شوهارت[۱] از میان ابزار های یاد شده، نمودارهای کنترل از مهم‌ترین و پراهمیت‌ترین ابزارها به شمار می‌رود.در سال ۱۹۲٤ شوهارت از آزمایشگاه تلفن بل، اساس نمودارهای کنترل فرآیند را پایه گذاری نمود که برای استفاده از این نمودارها از قبیل نمودار $\bar{X}$ و نمودار R باید نمونه‌هایی در طول زمان جمع آوری نمود و مقادیر آن را بر روی نمودارها رسم کرد. هشدار مبنی بر تحت کنترل نبودن فرآیند زمانی بدست می‌آید که آماره محاسبه شده برای یک نمونه ، خارج از حد و کنترل واقع شود. شکل ۱ نمونه ای از یک نمودار شوهارت کلاسیک را نشان می‌دهد.

این نمودار شامل یک خط مرکزی است که مقدار متوسط مشخصه کیفی را در حالت تحت کنترل نشان می‌دهد، دو خط افقی دیگر حد کنترل بالا و پایین می‌باشند[۳].

---

[1]Multivariate statistical process control

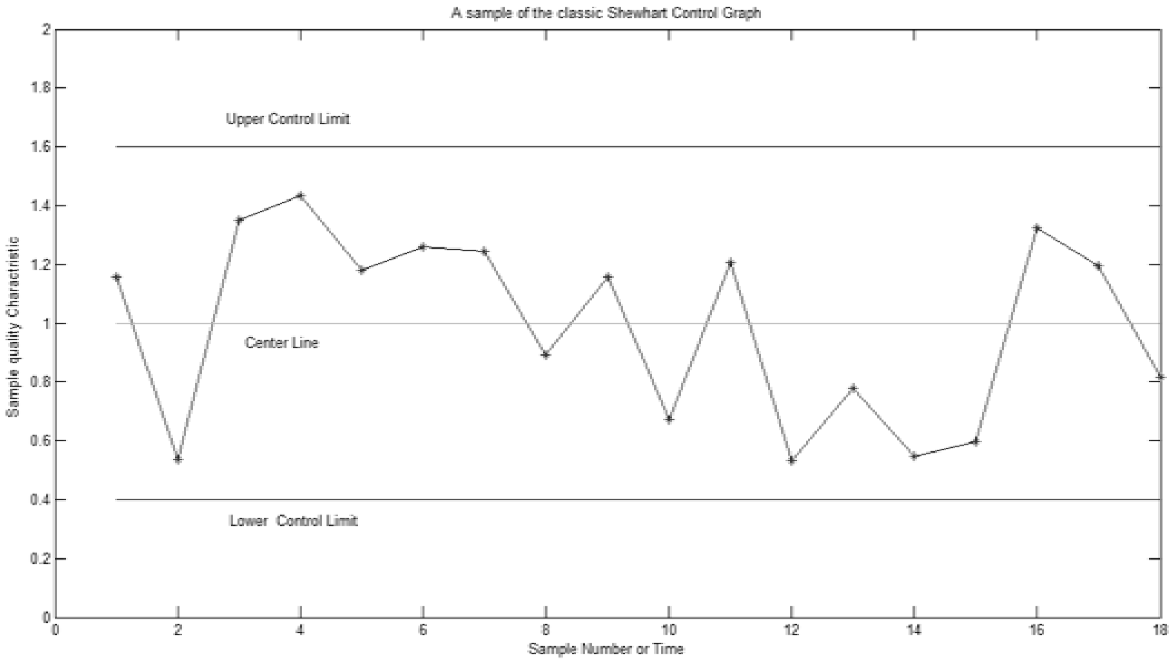
[2]Chi-square

[3]Hotelling's T^2

[4]Multivariate exponentially weighted moving average

[5]Multivariate cumulative sum

[6]Douglas C. Montgomery



شکل ۱- نمونه ای از یک نمودار شوهارت کلاسیک

به طور کلی اگر یک نقطه از ۳۵ نقطه یا دو نقطه از ۱۰۰ نقطه خارج از حدود کنترل واقع شود، می‌توان گفت در فرآیند اشکالی به وجود نیامده است. از طرفی دیگر اگر ۷ نقطه متوالی و یا بیشتر در یک طرف خط مرکزی واقع شوند، فرآیند از کنترل خارج شده است.

۲-۱- ویژگی‌های نمودار کنترل شوهارت

از مزایای نمودار شوهارت می‌توان به دو علت زیر اشاره داشت :

۱) این نمودارها یک روش اثبات شده برای بهبود بهره‌وری محسوب می‌شوند و ابزار موثری در جلوگیری از تولید اقلام معیوبند .

۲) نمودارهای کنترل از تنظیم‌های غیر ضروری جلوگیری و اطلاعاتی در مورد کارایی فرآیند، فراهم می‌سازند[٤].

عمده‌ترین نقطه ضعف نمودار های کنترل مربوط به زمانی است که دو یا چند مشخصه کیفی در یک زمان مشخص با هم در فرآیند اندازه گیری شوند که در این حالت نتایج درستی از آن بدست نمی‌آید[٥].

۳- کنترل فرآیند آماری چند متغیره(MSPC)

امروزه به دلیل پیشرفت سریع و پیچیدگی سیستم تولیدات جدید، اندازه‌های فرآیند می‌توانند بیشتر از قبل در فراوانی بسیار بالایی جمع آوری شوند. از این رو روش‌های آماری چند متغیره اهمیت خاصی پیدا می‌کند. یکی از

پیامدهای اصلی چند متغیره آنست که متغیرهای همبسته باید با هم تحلیل شوند. یک نمونه از این موضوع در صنعت خودرو مصداق پیدا می‌کند.جایی که بین اندازه های مختلف گرفته شده از یک جسم سخت مربوط به یک اتومبیل همبستگی وجود دارد. با اندازه گیری تمام متغیرها به طورهم‌زمان، روش‌های کنترل فرآیند چند متغیره نه تنها می‌تواند اطلاعات را از هر مشخصه استخراج کند،بلکه ساختار همبستگی میان متغیرها را مشخص و نظارت می‌نماید. لازم به ذکر است روش‌های نظارت نمودار کنترل یک متغیره در اینجا دارای اهمیت نیست زیرا متغیرها مستقل از یکدیگر نیستند و اطلاعات همبستگی آن‌هامی‌تواند برای درک رفتار فرآیند بسیار مهم باشد. در این بخش به متغیرها دو مهارت مهم و مقایسه آن‌ها با هم، برای کنترل چندین مشخصه کیفی مرتبط به طور هم‌زمان می‌پردازیم.

۳-۱- آزمون $T^۲$ هتلینگ

فرض کنید $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})'$ به عنوان مشاهده ای به اندازه p، از یک فرآیند در زمان iدر نظر گرفته شود. هنگامی که فرآیند تحت کنترل باشد ، مشخصه X از توزیع نرمال چند متغیره با میانگین برداری μ و ماتریس کوواریانس Σ پیروی می‌کند. اگر از جامعه مورد بررسی، نمونه ای به اندازه‌یnگرفته شود بردار میانگین نمونه $\bar{X}$ (رجوع شود به [۱]) و ماتریس کوواریانس نمونه S برای برآورد پارامترهای μ و Σ مورد استفاده قرار می‌گیرند[۶-۱۳].

$$\bar{X} = (\bar{X}_۱, \bar{X}_۲, \dots, \bar{X}_p)', \qquad (۱)$$

$$S = \frac{۱}{n-۱}\left(\sum (X_i - \bar{X})(X_i - \bar{X})'\right), \qquad (۲)$$

آماره $T^۲$هتلینگ برای یک مشاهده‌ی X_i به صورت زیر است[۶-۱۳]:

$$T^۲ = (X_i - \bar{X})' \times S^{-۱} \times (X_i - \bar{X}), \qquad (۳)$$

در رابطه (۳) هرچه مقدار عددی $T^۲$ بزرگ‌تر باشد، انحراف مشاهده X_i از جامعه تحت کنترل بیشتر است.

می‌توان از ضرب کردن آماره ی $T^۲$ در مقدار $\frac{n(n-p)}{p(n+1)(n-1)}$ ، یک مقدار تبدیل یافته از این آماره را به صورت رابطه (٤) در نظر گرفت.

$$\frac{n(n-p)}{p(n+۱)(n-۱)} \times T^۲ \propto F_{p,n-p}, \qquad (٤)$$

که دارای توزیع فیشر با p و $n-p$ درجه آزادی می‌باشد . چنانچه مقدار تبدیل یافته آماره $T^۲$ هتلینگ بزرگ‌تر از مقدار داده شده جدول F در سطح معنی داری α باشد، فرض صفر که بر پایه تحت کنترل بودن فرآیند است، رد می‌شود. لازم به ذکر است اگر X_i از توزیع نرمال پیروی نکند، مقدار تبدیل یافته آماره $T^۲$ نیز از توزیع F تبعیت نخواهد کرد و از این آماره برای قضاوت در مورد کنترل یا عدم کنترل فرآیند نمی‌توان استفاده کرد. برای تشخیص انحراف، اندازه های چندین مشخصه در یک سیستم اطلاعاتی به وسیله متغیر های تصادفی چندگانه ارائه شده‌اند.

این در حالی است که توزیع هر متغیرتصادفی از قبل مشخص نیستو نمی توان توزیع آن متغیر را نرمال فرض کرد. با این وجود اگر تعداد p متغیر، از هم مستقل باشند و p بزرگ باشد $(p > 30)$، بدون اینکه توزیع هر کدام از p متغیر را در نظر بگیریم، آماره $T^۲$ به نابر قضیه حد مرکزی تقریباً دارای توزیع نرمال است[۷]. با استفاده از مقادیر نمونه آماره $T^۲$، مقادیر میانگین و انحراف استاندارد در جامعه $T^۲$ از میانگین نمونه $\overline{T^2}$ و انحراف معیار نمونه $S_{T^۲}$ بر آورد می‌شوند.حدود کنترل برای شناسایی بی نظمی‌های خارج از کنترل معمولاً به صورت حدود کنترل 3σ در نظر گرفته می‌شود و به شکل رابطه (۵) می‌باشد[۷-۱۵].

$$\bar{T}^{2} - 3S_{T^2}, \bar{T}^{2} + 3S_{T^2}. \qquad (5)$$

عواملی مثل تغییر بردار میانگین در حالت کنترل، انحراف از ساختار کوواریانس در حالت کنترل یا همبستگی متغیر و یا ترکیبی از دو حالت اخیر می‌توانند باعث ایجاد حالات خارج از کنترل شوند.

در حالت تغییر میانگین، حداقل یک متغیر از p متغیر می‌توانند از حالت کنترل خارج شود. اگر چه آزمون T^2 تغییرات میانگین و همبستگی‌های متقابل را نشان می‌دهدولی به همبستگی‌های متقابل بین متغیرها بیشتر از دگرگونی میانگین حساس است زیرا به ساختار همبستگی متغیرها وابسته است.

۳-۲- آزمون χ^2

در این بخش به بیان یکی دیگر از مهارت‌های تحلیل آماری چند متغیره می‌پردازیمکه بر پایه آزمون χ^2 استوار است.در این مقاله آزمون χ^2 و اجرای آن به طور عمده برای ارزیابی عملکرد آزمون T^2 استفاده شده است [۱٤]. اگر pمتغیر برای اندازه گیری داشته باشیم و X_j نشاندهنده‌ی مشاهده متغیرjام $(1 \le j \le p)$ در یک زمان مشخص در نظر گرفته شود، آماره آزمون χ^2 به صورت رابطه (۷) است :

$$\bar{X}_J = \frac{1}{n}\sum_{i=1}^{n} X_{i,J}, \qquad (6)$$

$$\chi^{2} = \sum_{j=1}^{p} \frac{(X_j - \bar{X}_j)^2}{\bar{X}_j}, \qquad (7)$$

$$S = \frac{1}{n-1}\left(\sum (\chi_i - \bar{\chi})(\chi_i - \bar{\chi})'\right), \qquad (8)$$

از آنجا که آماره آزمون بالا فاصله نقاط داده‌ها را از مرکز مجموعه داده‌ها اندازه گیری می‌کند، لذا از آن به عنوان آزمون فاصله χ^2 یاد می‌کنند. چنانچه p بزرگ باشد، آماره χ^2 بر پایه قضیه حد مرکزی تقریباً دارای توزیع نرمال است[۷].

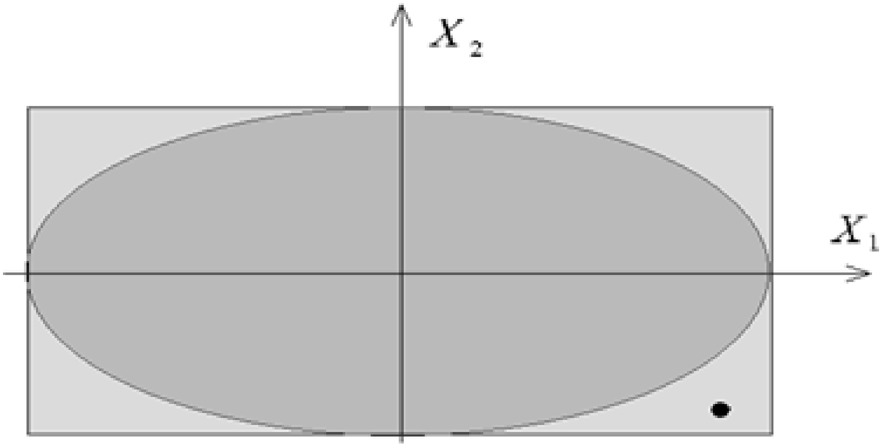
با استفاده از یک نمونه مقادیر آماره χ^2 می‌توان میانگین و انحراف معیار جامعه χ^2 را برآورد نمود. حدود کنترل برای تشخیص بی نظمی‌های خارج از کنترل به صورت رابطه (۹) می‌باشد[۷ و۸].

$$\bar{\chi}^{2} - 3S_{\chi^2}, \bar{\chi}^{2} + 3S_{\chi^2}. \qquad (9)$$

۳-۳- مقایسه آزمون T^2 و χ^2

هردو آماره آزمون χ^2 و T^2 فاصله‌ی یک مشاهده را از بردار میانگین چند متغیره در جامعه اندازه گیری می‌کنند.آماره آزمونT^2 فاصله آماری که در بر گیرنده ماتریس واریانس-کوواریانس است استفاده می‌کند در حالی که آماره آزمونχ^2 فاصله χ^2 را به کارمی‌گیرد که شبیه فاصله اقلیدسی می‌باشد. اما از میانگین (مقدار متوسط) هر متغیر برای اندازه گیری فاصله اقلیدسی روی آن متغیر استفاده می‌کند. در مقایسه با آماره T^2 آماره χ^2 ساختار همبستگی p متغیر را در نظر نمی‌گیرد با این وجود، آزمون T^2 تغییرات میانگین و همبستگی بین متغیرها را تعیین می‌کند در صورتی که آزمون χ^2 تنها متغیر میانگین را روی حداقل یک متغیر مشخص می‌نماید.همچنین این آزمون در تشخیص نقاط خارج از کنترل به کار می‌رود [۱٤] . هنگامی که روی یک مجموعه کوچک از داده‌های کامپیوتری که شامل عوامل نرمال و خارج از کنترل است بخواهیم آزمون انجام دهیم، آزمون χ^2 تمام نقاط خارج از کنترل را مشخص می‌کند و دیگر هشدار های اشتباه روی حالت‌های نرمال ایجاد نمی‌گردد.از آنجا که هر حالت شامل

بسیاری از پیشامدهاست، برای یک مجموعه کوچک از داده‌هاآزمون χ^2 ۷۵٪ از پیامدهای خارج از کنترل را مشخص می‌نماید.از سوی دیگر، در یک مجموعه بزرگ آزمون χ^2 ۶۰ ٪ از حالات خارج از کنترل را شناسایی می‌کند و دیگر هشدار اشتباه روی فرآیندهای نرمال ایجاد نمی‌شود. به شکل ۲ رجوع شود.



□ حدود کنترل برای اطلاعات تک متغیر جدا

○ حدود کنترل برای یک آزمون چند متغیره مانند آزمون χ^2

● یک نمونه به وسیله اطلاعات تک متغیره جدا

شکل ۲ – تفاوت در حدود کنترل برای مقایسه آزمون های تک متغیره و آزمون چند متغیره χ^2

۴– استراتژی‌های تشخیص چند متغیره متعارف

در این بخش روش های تشخیص چند متغیره را بیان می کنیم.

۴–۱– تحلیل مولفه های اصلی [1]

PCA یک تبدیل خطی می‌باشد که داده‌ها را از فضای اولیه به یک فضای ویژه با مؤلفه های مستقل خطی منتقل می‌کند. این فضا در واقع فضایی با محورهای عمود بر هم است که جهت محورهای آن با استفاده از بردارهای ویژه ماتریس کوواریانس نمونه‌های آموزشی که به صورت زیرتعریف می‌شود به دست می‌آید.

$$C = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X_i})(X_i - \overline{X_i})^T \quad (۱۰)$$

که در آن X_i نمونه آموزشی i ام، $\overline{X}_i$ بردار میانگین مجموعه نمونه‌های آموزشی و N تعداد نمونه‌های آموزشی می‌باشد. ماتریس کوواریانس را می‌توان به صورت $C = U\Sigma U^T$ تجزیه کرد . U یک ماتریس واحد می‌باشد که ستون‌های آن از بردارهای ویژه ماتریس کوواریانس تشکیل می‌شود و Σ یک ماتریس قطری است که عناصر روی قطر اصلی آن مقادیرویژه یا مقادیر منفرد ماتریس کوواریانس می‌باشد. اندازه هر مقدار ویژه ماتریس کوواریانس، متناسب با توان موجود در مؤلفه ای از داده‌ها می‌باشد که در راستای بردار ویژه متناظر

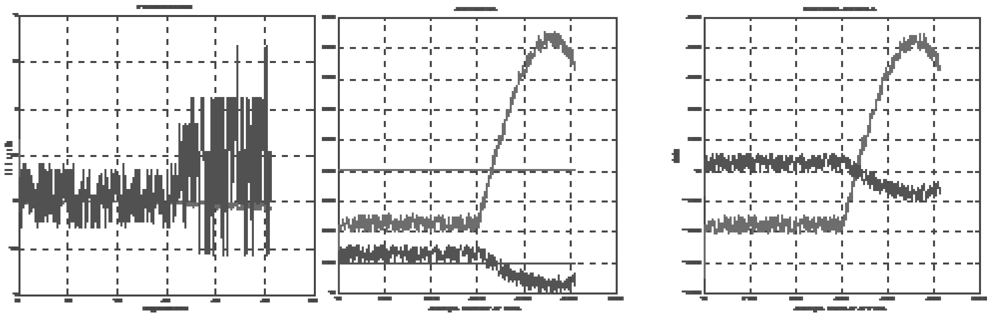
[1] PCA

با آن مقدار ویژه است. بنابراین با تعیین مقادیر ویژه غیر صفر می‌توان بعد مؤثر داده‌ها را بدست آورد. در نهایت با تشکیل یک ماتریس به وسیله تعدادی از سطرهای U^T که متناظر با مقادیرویژه غالب هستند و ضرب کردن آن در ماتریس نمونه‌های آموزشی، بردارهای X_i به یک سیستم جدید با مؤلفه های غیره همبسته در جهت سطرهای U^T منتقل شده و تبدیل به یک بردار تصادفی چند متغیره با ابعاد کمترمی‌شوند. البته توجه داشته باشید که هنوز بیشتر انرژی داده‌های اولیه، در داده‌های موجود در فضای جدید وجود دارد. در واقع تبدیل PCA ضمانت می‌کند که اطلاعاتی از بین نرود. البته توجه داشته باشید که اگر چه در ویژگی‌های به دست آمده از تبدیل PCA انرژی اولیه داده‌ها حفظ می‌شود اما ویژگی‌های حاصل از لحاظ مقدار سیگنال به نویز بهینه نیستند. در واقع PCA بدون در نظر گرفتن نویز از سیگنال استفاده می‌کند و در نتیجه ویژگی‌های حاصل به گونه‌ای نیستند که معیار سیگنال به نویز را بهینه کنند.

در ادامه مثالی از کاربرد طیف انتشار نوری (OES[۱]) که در صنایع نیمه هادی کاربرد دارد ارائه شده است[۱۶]. در شکل ۳، تصویر چپ توزیع داده‌ها و ترسیم میانگین، تصویر میانی فاصله اقلیدسی از میانگین و تصویر راست مؤلفه PCA اولیه دو سیگنال کلاس آماری آورده شده است.

.

[1]Optical Emission Spectroscopy



شکل ۳: (تصویر چپ) توزیع داده‌ها و ترسیم میانگین،(تصویر میانی) فاصله اقلیدسی از میانگین و(تصویر راست) مؤلفهPCA اولیه دو سیگنال کلاس‌ها

۴-۱-۱- PCA و روش‌های T^2

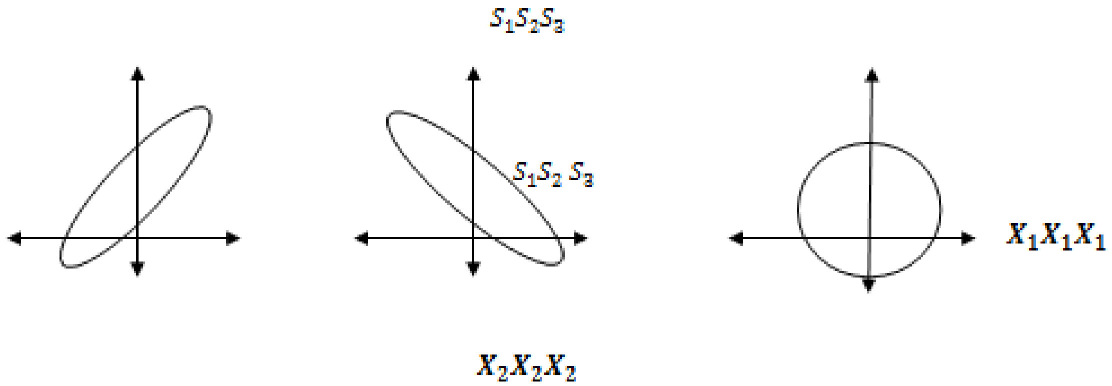
PCA مزایای زیادی را برای داده های متراکم یا فشرده ایجاد می‌کند به عبارت دیگر به جای سروکار داشتن با صدها متغیر، فقط می‌توان با عامل‌های اصلی سروکار داشت. به اعتقاد وایس[1]، T^2 یک اندازه آماری است لذا نمی‌تواند تفاوت‌ها را در حالت جهت دار بر طرف نماید [۱۷]. علاوه بر این آماره T^2 زمانی که متغیرهای فرایند بسیار به هم وابسته‌اند، نمی‌تواند محاسبه گردد زیرا ماتریس کوواریانس نمونه S اغلب معکوس ناپذیر است.در نتیجه از آماره T^2 بر اساس PCA می‌توان برای حذف موضوع وارون پذیری ماتریس کوواریانس S استفاده نمود.

۴-۲- کوواریانس تعمیم یافته

روش کوواریانس تعمیم یافته، تمام اطلاعات در ماتریس داده‌ها را در یک اندازه تعمیم یافته تکی از کوواریانس داده‌ها با گرفتن دترمینان از ماتریس کوواریانس فشرده می‌کند. کوواریانس تعمیم یافته نشان می‌دهد، این مسائل دارای دو بعد هستند. تغییرات در جامعه اول به طور مثبت با ماتریس کوواریانس نمونه S_1 همبسته شده‌اند در حالی که در جوامع دیگر همبستگی منفی S_2 یا هیچ نوع همبستگی S_3 در نظر گرفته شده‌اند. توجه داشته باشید اگر چه کوواریانس تعمیم یافته نمی‌تواند تغییر دورانی در کوواریانس را شناسایی کند اما روش T^2 این چنین تغییر دورانی را به عنوان پیشامد خارج از کنترل در نظر می‌گیرد.

$$S_1 = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} S_2 = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix} S_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}.$$

[1]Wise



شکل ۴. دو ماتریس کوواریانس متفاوت

۴-۳-تشخیص عامل ویژه

در این بخش عامل ویژه ماتریس را به دست می آوریم.

۴-۳-۱- روش تشخیص فضای ویژه

در این بخش روش تازه ای را با به‌کارگیری اصطلاح "تشخیص فضای ویژه " بکار می‌بریم که به تغییرات جهت‌دار در جامعه توجه دارد.باید بدانیم که روش تشخیص فضای ویژه نیازمند تبدیل ویژه از ماتریس کوواریانس است. به دلیل اینکه هر ماتریس کوواریانس نمونه S،حقیقی و متقارن است یک ماتریس حقیقی متعامد V و یک ماتریس قطریγ وجود دارد به قسمی که $S = V\gamma V^T$ و داریم:

$$S = \sum_{i=1}^{p} \lambda_i V_i {V_i}^T = \sum_{i=1}^{p} \left(\sqrt{\lambda_i} V_i\right)\left(\sqrt{\lambda_i} V_i\right)^T, \quad (۹)$$

که در آنpتعداد متغیرها ، λ_i یک مقدار ویژه و یک عنصر قطر اصلی ازγو V_i یک بردار ویژه درV است .به طور خلاصه ماتریس فضای ویژه تبدیلی از ماتریس کوواریانس نمونه است که از نمونه های چند تایی بدست می‌آید و مهارت‌های تشخیص استفاده شده در ماتریس فضای ویژه، جفت آماره های مرتب را محاسبه می‌کند و از این اطلاعات برای تشخیص و شناسایی تغییرات مهم در فرایند استفاده می‌شود.

عامل اصلی یک بردار عمودی از ماتریس فضای نمونه است که به عنوان یک بردار تصادفی بررسی می‌گردد و بازه های اطمینان می‌توانند از توزیع داده شده برقرار شوند.

فرض کنید λ_i به صورت نزولی مانند ترتیبی از مقادیر ویژه تنظیم شده باشد همچنین V_i تقریباً یک بردار ویژه در ارتباط با λ_i باشد مگر آنکه و V_i به صورت منحصر به فرد انتخاب شده باشد. برای انتخاب V_i از یک گشتاور استفاده خواهیم کرد و فرض می‌کنیم مقادیر ویژه قابل تکرار نباشند در این صورت

$$\sqrt{\lambda_i} V_i \quad ; \quad i = 1,2, \dots, p, \quad (۱۰)$$

عاملی اصلی نامیده می‌شود و ماتریسی که شامل تمام عوامل اصلی باشد، ماتریس فضای ویژه $\boldsymbol{E}$ می‌باشد و داریم :

$$S = \sum_{i=1}^{p} \lambda_i V_i {V_i}^T$$
$$= \sum_{i=1}^{p} \left(\sqrt{\lambda_i} V_i\right)\left(\sqrt{\lambda_i} V_i\right)^T$$
$$= (V\gamma^{\frac{1}{2}})(V\gamma^{\frac{1}{2}})^T = EE^T.$$

بنابراین بجای پیگیری تمام عوامل اصلی $\sqrt{\lambda_i} V_i$، $i = 1,2 \dots, p$ به طور مستقل می‌توانیم ماتریس کلی فضای ویژه E را نمایش دهیم یکی از اهداف مهم این مقاله مشخص کردن خصوصیات ماتریس ویژه E در قالب قضایا و اثبات آن می‌باشد .

به منظور برقراری و تعیین توزیع ماتریس **E**، ابتدا باید ویژگی‌های توزیع ماتریس کوواریانس نمونه **S** را مطالعه کنیم توزیع **S** بعد از کنکاش و تحقیق، توزیع ویشاوت نامیده می‌شود که به صورت جمعی از حاصل ضرب‌های مستقل تصادفی عوامل تصادفی نرمال چند متغیره تعریف می‌شود. ما توزیع ویشاوت را به عنوان بسط کلی از توزیع $\boldsymbol{\chi^2}$ در مسائل چند متغیره در نظر خواهیم گرفت.

تعریف ۱: فرض کنید $Z = [Z_1, \dots, Z_n]^T$ که در آن Z_i ها از هم مستقلند و دارای توزیع نرمال p متغیره $N_p(0, \Sigma)$ می‌باشند. اگر $W = Z^T Z$ در این صورت W دارای توزیع ویشارت با n درجه آزادی است. باید بدانیم که W یک ماتریس $p \times p$ مثبت معین است.

قضیه ۱: اگر S دارای توزیع ویشارت $W_p(n, \Sigma)$ و C ماتریس ناویژه با بعد $p \times p$ باشد آنگاه CSC' دارای توزیع $W_p(n, C\Sigma C')$ می باشد .

چون S دارای توزیع $W_p(n, \Sigma)$ است آن را می‌توان به صورت زیر نوشت :

$$S = \sum_{\alpha=1}^{n} Y^{\alpha} (Y^{\alpha})^T,$$

که در آن $Y^{\alpha} = \left(Y_{\alpha 1}, \dots, Y_{\alpha \mathrm{p}}\right)'$ و $\alpha = 1, \dots, n$ بردارهای p بعدی مستقل نرمال با میانگین صفر و ماتریس کوواریانس یکسان Σ هستند . بنابراین $C\Sigma C'$ به صورت زیر توزیع شده اند:

$$\sum_{\alpha=1}^{n} (CY^{\alpha})(CY^{\alpha})' = \sum_{\alpha=1}^{n} (Z^{\alpha})(Z^{\alpha})',$$

$$\alpha = 1, \dots, n; Z^{\alpha} = \left(Z_{\alpha 1}, \dots, Z_{\alpha \mathrm{p}}\right)'.$$

بردارهای p بعدی مستقل نرمال با میانگین صفر و ماتریس کوواریانس CSC' می‌باشد. با توجه به ساختار S و تعریف توزیع ویشارت نتیجه زیر را خواهیم داشت:

$$(n-1)S \sim W_p(n-1, \Sigma).$$

در ارتباط با توزیع ویشارت می‌توان خصوصیات دیگری نیز بیان کرد که در زیر به طور کامل به بیان و اثبات این ویژگی‌ها می‌پردازیم و از آن‌ها برای اثبات قضیه ای مهم استفاده می‌کنیم که برای بدست آوردن توزیع ماتریس فضای ویژه E بکار رفته است.

قضیه ۲ : اگر $S \sim W_p(n, \Sigma)$ و $S = (V\gamma^{\frac{1}{2}})(V\gamma^{\frac{1}{2}})^T = EE^T$ آنگاه توزیع E یک تابع از $E = A^{-1}TU$ خواهد بود که در آن، A ریشه دوم ماتریس معکوس ماتریس کوواریانس جامعه $\Sigma^{-1} = A^T A$ ، U یک ماتریس یکه داده شده و T با تجزیه بارتلت از یک ماتریس توزیع ویشارت در ارتباط است [۱۸].

قبل از اثبات قضیه به صورت خلاصه لازم است توزیع تجزیه بارتلت بررسی گردد.

فرض کنید S دارای توزیع $W_p(n, \Sigma)$ ، $n \geq p$ و $B = \left(B_{ij}\right) i < j ; B_{ij} = 0$ ، ماتریس پایین مثلثی یکتا با عناصر قطری مثبت باشد که در شرط زیر صدق می کند:

$$S = BB',$$

ژاکوبی تبدیل $S \to B$ عبارتست از

$$\left[s = \left(s_{ij}\right), b = \left(b_{ij}\right)\right],$$

$$det\left(\frac{\partial \delta}{\partial b}\right) = 2^p \prod_{i=1}^{p} (b_{ii})^{p+1-i},$$

و تابع چگالی احتمال B نسبت به اندازه لبک db پس از اندکی محاسبات به صورت زیر بدست می آید

$$f_B(b) =$$

$$k 2^p (det\Sigma)^{-\frac{n}{2}} \prod_{i=1}^{p} (b_{ii})^{n-i} \, exp\left\{-\frac{1}{2} tr\Sigma^{-1} bb'\right\},$$

که در آن

$$k^{-1} = \left(\prod_{i=0}^{p-1} \Gamma\left(\frac{n-i}{2}\right)\right) 2^{n\frac{p}{2}} \pi^{p(p-1)/4}.$$

که این مطالب در مرجع شماره [۱۵] اثبات شده است.

حال فرض کنید $T = T_{ij}$ یک ماتریس پایین مثلثی ناویژه باشد (که عناصر قطری آن الزاماً مثبت نیستند). در این صورت

می‌توان نوشت $T = B\theta$ که در آن θ یک ماتریس قطری با عناصر قطری ± 1 است. تابع چگالی T پس از اندکی محاسبات به صورت زیر حاصل می‌شود .

$$f_T(t) = k2^p exp\left(-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{i} {t_{ij}}^2\right)\prod_{i=1}^{p}\left({t_{ij}}^2\right)^{\frac{n-i}{2}},$$

که k همان مقدار قبل است. از این معادله معلوم می‌گردد T_{ij} ها دارای توزیع مستقل و ${T_{ii}}^2$، دارای توزیع مربع کای مرکزی با $n-i+1$ درجه آزادی $(i = 1,2,\dots,p)$ و $T_{ij}\,(i \neq j)$دارای توزیع نرمال با میانگین صفر و واریانس ۱ می‌باشد[۱۵] .

اکنون به اثبات قضیه۲ می‌پردازیم. برای نشان دادن اینکه توزیع ماتریس E به توزیع Tبستگی دارد، باید تغییری در قضیه گوپاونگرا ایجاد کنیم تا بتوانیم هر توزیع ویشارت با پارامترهای (n,Σ)را به توزیع ویشارت با پارامترهای(n,I) تبدیل نماییم . اثبات این ویژگی در مرجع شماره [۱۶] آمده است . با توجه به تجزیه بارتلت و تبدیل Σ به Iداریم :

$$S = TT^T.$$

که T، یک ماتریس پایین مثلثی و به شکل زیر است:

$$T = \begin{bmatrix} t_{11} & & \\ t_{21} & t_{22} & \\ & & \ddots \end{bmatrix},$$

که در آن

$$t_{11}^2 \sim \chi_n^2,$$
$$t_{22}^2 \sim \chi_{n-1}^2,$$
$$;1 \le i \le p t_{ii}^2 \sim \chi_{n-i+1}^2.$$

از دیگر سو t_{ij} که $1 \le j \le i \le p$ به طور مستقل توزیع شده است و $t_{ij}\sim N(0,1)$ لذا داریم:

$$S_{new} = TT^T = ASA^T$$

$$= AEE^TA^T = (AE)(AE)^T,$$

که در آن $S_{new}\sim W_p(n,I)$ در نتیجه داریم

$$AE = TU \;\Rightarrow E = A^{-1}TU,$$

که در آن U، یک ماتریس یکه بوده و توزیع E تابعی از توزیع T است. از این رو خواص جانبی مربوط به توزیع E بسیار به خواص جانبی توزیع T بستگی دارد و T با تجزیه بارتلت از یک ماتریس توزیع ویشارت در ارتباط است.

با مطالعه توزیع مجانبی T، قضیه زیر نشان می‌دهد که واریانس هر عضو T زمانی که تعدادی از نمونه‌ها به بی نهایت میل کند، به صفر میل می‌نماید.

قضیه ۳ : فرض کنید

$$(n-1)S = \left(\sqrt{n-1}T\right)\left(\sqrt{n-1}T\right)^T,$$

و$(n-1)S\sim W_p\left(n-1,I_p\right)$ که در آن $T = t_{ij}$ یک ماتریس پایین مثلثی است که تمام اعضایش دارای توزیع مستقل هستند و $t_{ii} > 0$ و$(n-1)t_{ii}^2\sim\chi_{n-i+1}^2 t_{ii} >$ 0که $1 \le i \le p$، از طرفی $\sqrt{n-1}t_{ij}\sim N(0,1)$ که $1 \le j \le i \le p$، آنگاه $Var\left(t_{ij}\right)$ زمانیکه n برای تمامi و j ها بی نهایت شود به سمت صفر میل می کند.

اثبات : برای اعضای غیر قطریt_{ij} از T، وقتی$i \neq j$ باشد داریم $\sqrt{n-1}t_{ij}\sim N(0,1)$ بنابراین کافی است واریانس t_{ij} محاسبه شود.

$$Var\left(\sqrt{n-1}t_{ij}\right) = (n-1)Var\left(t_{ij}\right) = 1,$$
$$\Rightarrow Var\left(t_{ij}\right) = \frac{1}{n-1},$$
$$\Rightarrow \lim_{n\to\infty} Var\left(t_{ij}\right) = \lim_{n\to\infty}\frac{1}{n-1} = 0.$$

برای اعضای قطری t_{ii}، داریم $(n-1)t_{ii}^2 \sim \chi_{n-i+1}^2$ پس :

$$Var\left((n-1)t_{ii}^2\right) = (n-1)^2Var(t_{ii}^2)$$
$$= 2(n-i+1),$$

$$\Rightarrow Var(t_{ii}^2) = \frac{2(n-i+1)}{(n-1)^2},$$
$$\Rightarrow \lim_{n\to\infty} \frac{2(n-i+1)}{(n-1)^2} = 0.$$

حال باید نشان دهیم وقتی حد $Var(t_{ii})$،$n \longrightarrow \infty$ به صفر تمایل پیدا می‌کند موضع را در قضیه بعدی اثبات می نماییم .

قضیه ٤ : فرض کنید $\lim_{n\to\infty} Var(X^2) = 0$ اگر داشته باشیم $Y = \sqrt{X^2} = X \geq 0$، آنگاه $\lim_{n\to\infty} Var(Y) = 0$

اثبات : چون $Var(X^2)$ به سمت صفر میل می‌کند توزیعش با تابع ثابت در نقطه داده شده aتمایل دارد. به عبارت دیگر، $P(X^2 = a) = 1$ ما جمله اخیر را به وسیله برهان خلف ثابت می‌کنیم. فرض کنیدn بزرگ بوده و X^2 بیش از یک جا، دارای احتمال غیر صفر است، در این صورت واریانس طبق تعریف نمی‌تواند صفر باشد.

X^2	x_1	x_2
$P(X^2 = x)$	p	q

X^4	x_1^2	x_2^2
$P(X^4 = x^2)$	p	q

$$E(X^2) = x_1 p + x_2 q$$
و
$$E(X^4) = x_1^2 p + x_2^2 q$$

واریانس X^2 پس از اندکی محاسبات به صورت زیر است .

$$Var(X^2) = E(X^4) - E^2(X^2) = pq(x_1 - x_2)^2 \neq 0.$$

با توجه به $x_1 > 0$ و $x_2 > 0$ و $x_1 \neq x_2$ و طبق برهان خلف X^2 بیشتر از یک جا احتمال غیر صفر دارد این در حالی است که $Var(X^2) = 0$ می‌باشد. پس فرض خلف باطل و حکم صادق است. حال از این حقیقت استفاده می‌کنیم که ریشه دوم مثبت $Y = \sqrt{X^2}$ پیوسته است چون تابعی ثابت بوده و یکنوا و با ویژگی یک به یک است. بنابراین توزیع Y باید تابع ثابتی مانند $Y = \sqrt{a}$ باشد و داریم:

$$\lim_{n\to\infty} P(Y = \sqrt{a}) = 1 \Rightarrow \lim_{n\to\infty} Var(Y) = 0.$$

در نتیجه با معلوم بودن $\lim_{n\to\infty} Var(t_{ii}^2) = 0$ نتیجه می‌گیریم $\lim_{n\to\infty} Var(t_{ii}) = 0$.

از قضیه بالا نتیجه می‌گیریم که هر عضو از ماتریس فضای ویژه E به مقداری در واریانس خود متمایل استو در حالت کلی می‌توان گفت ماتریس فضای ویژه نمونه E (که E بر اساس ماتریس کوواریانس نمونه S است) به F (ماتریس فضای ویژه جامعه که بر اساس Σ است) میل می‌کند .

۵- بحث و نتیجه گیری

در این مقاله به بررسی و مقایسه کنترل فرآیند آماری (SPC) و کنترل فرایند آماری چند متغیره (MSPC) پرداختیم. تحلیل اجزای اصلی (PCA) به عنوان یکی از استراتژی‌های تشخیص چند متغیره متعارف برای ساده سازی و تفسیر آن‌ها با ارائه مثال مورد بررسی قرار گرفت. در ادامه به معرفی فضای ویژه E و عامل مشخصه پرداختیم و بیان نمودیم عامل اصلی یک بردار عمودی از ماتریس فضای نمونه است که می‌تواند به عنوان یک بردار تصادفی مورد بررسی قرار گیرد. علاوه بر این در محیط‌های داده ای با فراوانی بالا که حداکثر همبستگی میان اندازه گیری‌ها موجود باشد عوامل اصلی تعیین کننده، موجب حذف داده‌ها می‌شوند. از نظر ریاضی اگر λ_i یک مقدار ویژه و یک عنصر قطر اصلی از

ماتریس قطری γ باشد و V_i یک بردار ویژه حقیقی متعامد V در نظر گرفته شود

$$\sqrt{\lambda_i}V_i \qquad ; \ i = 1,2 \dots, p.$$

عامل اصلی نامیده می‌شود و ماتریسی که شامل تمام عوامل اصلی باشد، ماتریس فضای ویژه E است. تحلیل فضای ویژه کاربرد وسیعی دارد. نخست اینکه می تواند برای تشخیص متعارف حالات خارج از کنترل استفاده شود. دیگر آنکه برای تعیین پیشامدهای مهم تعمیم یابد. در نهایت به بررسی توزیع ماتریس فضای ویژه E پرداختیم و ثابت کردیم توزیع E یک تابع از $E = A^{-1}TU$ است که در آن A ریشه دوم ماتریس معکوس ماتریس کوواریانس جامعه $\Sigma^{-1} = A^TA$، U یک ماتریس یکه داده شده و T با تجزیه بارتلت از یک ماتریس دارای توزیع ویشارت در ارتباط است.

منابع

[۱]. فاطمی، د. م. ت. (۱۳۵۷). کنترل کیفیت آماری. دانشگاه صنعتی امیر کبیر.

[۲]. فردوسی، ا. ر. (۱۳۸٤). تشریح و تعمیم طرح‌های کنترل فرایند آماری چند متغیره در صنایع ایران. پایان نامه کارشناسی ارشد، دانشگاه آزاد واحد مشهد.

[۳]. النساء، د. ر. ن. (۱۳۷٦). کنترل کیفیت آماری. د.–س.–. مونتگومری.

[٤]. جری، ن. س. (۱۳٦٦). استنباط آماری چند متغیره. مؤسسه چاپ و نشر آستان قدس رضوی.

[5]. Montgomery, D. C. (2005). Introduction to Statistical Quality Control. New York: John Wily & Sons.

[6]. Mittag, H. J. and Rinne, H., (1993). Statistical Methods of Quality Assurance. Chapman & Hall/CRC.

[7]. Chou, Y. M., Mason, R. L. and Young, J. C., (1999). Power comparisons for a Hotelling's T^2 statistic. Communications in Statistics-Simulation and Computation, vol. 28, 1031-1050.

[8]. Wichern, R. A. J. A. D. W., (1998). Applied Multivariate Statistical Analysis. Upper Saddle River: Prentice Hall.

[9]. Ryan, T. P., (1989). Statistical Methods for Quality Improvement. New York: John Wily & Sons.

[10]. Mason, R. L., Tracy, N. D. and Young, J. C., (1995). Decomposition of T^2 for multivariate control chart interpretation. Journal of Quality Technology, vol. 27, 99-110.

[11]. Mason, R. L., Tracy, N. D. and Young, J. C., (1997). A practical approach for interpreting multivariate T2 control chart signals. Journal of Quality Technology, vol. 29, 396-406.

[12]. Mason, R. L., and Young, J. C., (1999). Improving the sensitivity of the T^2 statistic in multivariate process control. Journal of Quality Technology, vol. 31, 155-165.

[13]. Everitt, B. S., (1979). A Monte Carlo investigation of the robustness of Hotelling's one-and two-sample T^2 tests. Journal of the American Statistical Association, 48-51.

[14]. Mason , R. L., Champ, C. W., Tracy, N. D., Wierda, S. J. and Young, J. C., (1997). Assessment of multivariate process control techniques. Journal of Quality Technology, 290, 140-143.

[15]. Ye, N. and Chen,Q. (2001). An anomaly detection technique based on a chi - square statistic for detecting intrusions into information systems. Quality and Reliability Engineering International, 17, 105-112.

[16]. Coburn, J. and Chen, M., (1980). Optical emission spectroscopy of reactive plasmas: A method for correlating emission intensities to reactive particle density. Journal of Applied Physics, 51, 3134-3136.

[17]. Gallagher, N. B., Wise, B. M., Butler, S. W., White, D. D. and Barna, G. G., (1997). Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: improving robustness through model updating. Venue: Process: Impact of Measurement Selection and Data Treatment on Sensitivity, Safeprocess '97

[18]. Kshirsagar, A., (1959). Bartlett decomposition and Wishart distribution. The Annals of Mathematical Statistics, 30, 239-241.