

خواص توزیعی آماره T^2 -هتلینگ در کنترل کیفیت چند متغیره

فریده توانگر^۱، محمد حسین علامت ساز^۲ و هوشنگ طالبی^۳

چکیده:

کنترل کیفیت چند متغیره ابزار مهمی در کنترل کیفیت آماری است، که هدف از آن ارائه روشی برای کنترل همزمان چند متغیر همبسته است. یکی از این روشها برای کشف نقاط خارج از کنترل در یک فرایند کنترل کیفیت چندمتغیره، استفاده از آماره T^2 هتلینگ است. خواص توزیعی این آماره به معلوم یا نامعلوم بودن پارامترهای جامعه و همچنین زمان نمونه‌گیری بستگی دارد به طوری که بسته به این شرایط دارای توزیع نمونه‌ای خی-دو، F و یا بتا است. در شرایط خاصی حدود کنترل به دست آمده بر مبنای این توزیعها به یکدیگر نزدیک می‌شوند. در این مقاله به خواص توزیعی و رابطه بین آنها برای شرایط مختلف فرایند پرداخته و کاربرد آنها را در فرایند کنترل آماری با چند گروه داده واقعی تشریح کرده‌ایم. **واژه‌های کلیدی:** حدود کنترل، کنترل کیفیت چند متغیره، نمودار T^2 .

۱ مقدمه

آماره T^2 هتلینگ^۳، آماره‌ای انعطاف‌پذیر در کنترل کیفیت چندمتغیره است که روش کنترل مبتنی بر آن تمام ویژگیهای یک روش کنترل چند متغیره را دارا است. از جمله این ویژگی‌ها می‌توان به سهولت در محاسبات، حساسیت نسبت به تغییرات کوچک ولی مداوم فرایند، کارایی لازم جهت کنترل فرایند در حین تولید و خارج از خط تولید و تفسیر نقاط خارج از کنترل اشاره کرد. کنترل یک فرایند آماری متشکل از دو مرحله است. در مرحله اول، روشهای تشخیص مشاهدات غیرمعمول در دسته‌ای از داده‌های اولیه با استفاده از آماره T^2 بررسی می‌شود. کنار گذاشتن آماری مشاهدات غیرمعمول در مرحله اول، مشابه برخورد با نقاط پرت است. هدف از آزمایش در این مرحله، اطمینان از تحت کنترل بودن فرایند است وقتی n نمونه اولیه انتخاب و آماره‌های $\bar{X}$ و S محاسبه می‌شود. در این مرحله با انتخاب مشاهداتی که در شرایط تحت کنترل تولید شده باشند، حدود کنترل برای مرحله دوم، که مرحله کنترل و نظارت بر تولیدات آتی است، تعیین می‌گردند. در بررسی خواص توزیعی آماره T^2 -هتلینگ فرض می‌شود مشاهدات چندمتغیره، نمونه‌ای تصادفی از جامعه نرمال چندمتغیره با میانگین μ و ماتریس کوواریانس Σ باشند. بر این اساس رفتار هر یک از

^۱دانشگاه آزاد اسلامی واحد شیراز
^۲گروه آمار، دانشگاه اصفهان
^۳Hotelling's T^2 statistic

نامعلوم هستند و یا زمانی که کنترل بردارهای انفرادی یا گروهی مشاهدات نیز مورد نظر باشد به کار برده شود.

۱.۲ کنترل با پارامترهای نامعلوم

در این بخش، ابتدا حالتی که تنها کنترل یک بردار مشاهده‌ای $X' = (x_1, x_2, \dots, x_p)$ مورد نظر باشد مورد مطالعه قرار می‌گیرد. این حالت ممکن است زمانی رخ دهد که نرخ تولید محصول آنقدر کم باشد که نتوان به راحتی نمونه‌های بیشتر از حجم یک، برداشت نمود.

فرض کنید که مشاهدات از جامعه‌ای با توزیع نرمال چندمتغیره با بردار میانگین نامعلوم μ و ماتریس کواریانس نامعلوم Σ باشند. با استفاده از داده‌های اولیه‌ای که نقاط پرت آن کنار گذاشته شده‌اند، برآوردهای $\bar{X}$ و S برای μ و Σ محاسبه می‌شوند. برای رسم یک نمودار کنترل چندمتغیره با استفاده از آماره T^2 برای مشاهده انفرادی X_i و با فرض درستی $H_0: \mu_i = \mu$ توزیع نمونه‌ای آماره به صورت زیر است:

$$T^2 = (X_i - \bar{X})' S^{-1} (X_i - \bar{X}) \quad (1)$$

$$\sim \frac{(n-1)^2}{n} \beta_{(p, n-p-1)}.$$

زیرا در این حالت داریم:

$\sqrt{\frac{n}{n-1}}(X_i - \bar{X}) \sim N_p(\sqrt{\frac{n}{n-1}}(\mu_i - \mu), \Sigma)$ ،
و با تعریف ماتریس تصادفی S_1 مستقل از $(X_i - \bar{X})$ و
 $(n-1)S = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$ به صورت
زیر:

$$(n-1)S = (n-2)S_1 + (X_i - \bar{X})(X_i - \bar{X})',$$

مشاهدات مستقل را می‌توان با توجه به تابع چگالی احتمال که در آن پارامترها معلوم یا نامعلوم هستند دسته‌بندی کرد.

در ادامه این مقاله، با دسته‌بندی توابع توزیع نمونه‌ای متفاوت آماره T^2 برحسب شرایط کاربریشان در کنترل مراحل اول و دوم، آنها را در بخشهای ۲ و ۳ ارائه کرده‌ایم. همچنین در هر بخش توزیع نمونه‌ای آماره T^2 در دو حالت پارامترهای معلوم و نامعلوم به همراه مثالهای عددی با داده‌های واقعی ارائه و موضوع توضیح داده خواهد شد.

۲ کنترل در مرحله اول

در یک فرایند کنترل، مجموعه داده‌های اولیه، به دلیل فراهم ساختن معیاری برای دستیابی به حدود کنترل و تشخیص خارج از کنترل بودن یا نبودن فرایند، بسیار مهم هستند. کم‌دقتی در این مرحله ممکن است هر انحراف کوچک و متفاوت داده‌های آتی را در مقایسه با این داده‌های اولیه خارج از کنترل تلقی نماید. وجود مشاهدات غیرمعمول، میزان تغییرات را افزایش داده و همبستگی بین متغیرها را در هم می‌ریزد. حتی کمترین تعداد مشاهده پرت نیز می‌تواند چنین انحرافات را به دنبال داشته باشد. به همین دلیل است که روشی مناسب برای پاک‌سازی داده‌های اولیه از نقاط پرت ضروری است.

در مرحله اول، به چندین صورت مختلف می‌توان از آماره T^2 در کشف نقاط پرت استفاده نمود. این آماره می‌تواند در موقعیت‌هایی که پارامترهای جامعه معلوم یا

خواهیم داشت:

$$(n-2)S_1 \sim W_p(n-2, \Sigma),$$

و در این صورت بنا به [۱] داریم:

$$T^2 = \frac{(n-1)^2}{n} \beta'_{(p, \frac{n-p-1}{p}, \tau_i^2)},$$

که در آن $\tau_i^2 = (\mu_i - \mu)' \Sigma^{-1} (\mu_i - \mu)$ پارامتر غیرمرکزی است. اما تحت فرض H_0 پارامتر غیرمرکزی برابر صفر شده و توزیع بتای غیرمرکزی $\beta'_{(p, \frac{n-p-1}{p}, \tau_i^2)}$ تبدیل به توزیع مرکزی $\beta_{(p, \frac{n-p-1}{p})}$ می‌شود. این توزیع به تعداد متغیرها، p ، و اندازه نمونه اولیه، n ، بستگی دارد. از این روابط برای تحت کنترل قرار دادن بردارهای انفرادی مشاهدات جمع آوری شده در مرحله اول (که برای محاسبه حدود کنترل به کار برده می‌شوند) استفاده می‌شود.

با خواص توزیعی فوق برای T^2 ، حدود کنترل قابل محاسبه‌اند. حد کنترل پایین^۴ (LCL) و حد کنترل بالا^۵ (UCL) به قرار زیر هستند:

$$LCL = \frac{(n-1)^2}{n} \beta_{(1-\frac{\alpha}{p}, \frac{p}{p}, \frac{n-p-1}{p})},$$

و

$$UCL = \frac{(n-1)^2}{n} \beta_{(\frac{\alpha}{p}, \frac{p}{p}, \frac{n-p-1}{p})},$$

که در آن α ، درصد بالایی توزیع بتا با پارامترهای p و $\frac{(n-p-1)}{p}$ است.

در بسیاری از مواقع، LCL برابر صفر قرار داده می‌شود. زیرا هر انحرافی از میانگین منجر به افزایش آماره T^2

Lower Control Limit^۴
Upper Control Limit^۵

خواهد شد. در این صورت می‌توان حد پایین را نادیده گرفت. در این حالت UCL از رابطه زیر محاسبه می‌شود:

$$UCL = \frac{(n-1)^2}{n} \beta_{(\alpha, \frac{p}{p}, \frac{n-p-1}{p})},$$

مشاهده‌ای با مقدار T^2 بزرگتر از UCL را از میان داده‌های اولیه کنار گذاشته و با بقیه داده‌ها، برآوردهای جدید بردار میانگین و ماتریس کوواریانس محاسبه می‌شوند. این عمل تا زمانی که مجموعه‌ای همگن از مشاهدات، وقتی هیچ کدام خارج از حدود کنترل نباشند به دست آید، تکرار می‌گردد. برآوردهایی که از مجموعه نهایی به دست می‌آیند برای محاسبه حدود کنترل و به کارگیری در مرحله دوم مورد استفاده قرار می‌گیرند [۵].

حالت دوم را برای کنترل میانگین چندین بردار در نظر می‌گیریم. به عبارت دیگر، وقتی k نمونه مستقل به حجم n ($p < n$)، از $N_p(\mu, \Sigma)$ انتخاب شده باشند، فرض صفر متناظر با تحت کنترل بودن این نمونه به شکل $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ است. در این حالت آماره T_i^2 مربوط به بردار i -ام برابر است با:

$$T_i^2 = \frac{kn}{k-1} (\bar{X}_i - \bar{\bar{X}})' S^{-1} (\bar{X}_i - \bar{\bar{X}}),$$

که در آن

$$\bar{\bar{X}} = \frac{1}{(k-1)n} \sum_{l=1}^k \sum_{j=1}^n X_{lj},$$

و

$$S = \frac{1}{kn-1} \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{\bar{X}})(X_{ij} - \bar{\bar{X}})',$$

هستند. بنابراین داریم:

$$T_i^2 \sim (kn-1) \beta'_{(p, kn-p-1, \tau_i^2)}, \quad (2)$$

که در آن

$$\tau_i^2 = \left(\frac{kn}{k-1}\right)(\mu_i - \bar{\mu})' \Sigma^{-1} (\mu_i - \bar{\mu}),$$

پارامتر غیرمرکزی و $\bar{\mu}$ میانگین $\mu_1, \mu_2, \dots, \mu_k$ است. حال ماتریس تصادفی S_1 به صورت زیر تعریف می‌شود:

$$(kn-1)S = (kn-2)S_1 + \frac{kn}{k-1}(\bar{X}_i - \bar{\bar{X}})(\bar{X}_i - \bar{\bar{X}})',$$

که در آن

$$(kn-2)S_1 = \sum_{l=1}^k \sum_{j=1}^n (X_{lj} - \bar{X}_l)(X_{lj} - \bar{X}_l)' + n \sum_{l \neq i}^k (\bar{X}_l - \bar{\bar{X}})(\bar{X}_l - \bar{\bar{X}})',$$

و

$$\sqrt{\frac{kn}{k-1}}(\bar{X}_i - \bar{\bar{X}}) \sim N_p\left(\sqrt{\frac{kn}{k-1}}(\mu_i - \bar{\mu}), \Sigma\right),$$

بنابراین، با توجه به مستقل بودن $(\bar{X}_i - \bar{\bar{X}})(\bar{X}_i - \bar{\bar{X}})'$ و S_1 برای هر i رابطه (۲) به دست می‌آید. با توجه بهخاصیت توزیعی آماره تحت فرض H_0 ، توزیع بتایغیرمرکزی $\beta'_{(p, n-p-1, \tau_i^2)}$ تبدیل به توزیع مرکزی $\beta_{(p, n-p-1)}$ می‌شود [۱]. فرض صفر فوق با رسم مقادیر

$$T_i^2 = (\bar{X}_i - \bar{\bar{X}})' S^{-1} (\bar{X}_i - \bar{\bar{X}})$$

با حد بالایی زیر آزمون می‌شود.

$$UCL = \frac{(kn-1)(k-1)}{kn} \beta_{(\alpha, p, \frac{kn-p-1}{p})},$$

که در آن صدک بالایی یک متغیر

تصادفی از توزیع $\beta_{(p, \frac{kn-p-1}{p})}$ است.

مثال ۱ داده‌های جدول ۱، داده‌های واقعی مرحله اول

یک فرایند صنعتی شیمیایی را نمایش می‌دهند. همچون

بسیاری از فرایندهای شیمیایی، این مثال نیز با اندازه‌گیری

هم‌زمان متغیرهای درصد ناخالصی (X_1)، دما (X_2)و میزان غلظت (X_3) یک ماده خالص همراه است.

آزمونهای مقدماتی حاکی از آن است که داده‌ها از یک

توزیع نرمال سه‌متغیره پیروی می‌کنند.

در این مثال، ۱۴ مشاهده انفرادی از سه متغیر وجود دارد.

بنابراین، $p = 3$ و $n = 14$. بردار میانگین و ماتریس

کواریانس نمونه نیز به صورت زیر هستند:

$$\bar{X}_{14} = \begin{bmatrix} 16/83 \\ 85/19 \\ 43/21 \end{bmatrix},$$

و

$$S_{14} = \begin{bmatrix} 0/365 & -0/022 & 0/100 \\ -0/022 & 1/036 & -0/245 \\ 0/100 & -0/245 & 0/224 \end{bmatrix},$$

ماتریس همبستگی R_{14} به شکل زیر خواهد بود:

$$R_{14} = \begin{bmatrix} 1/000 & -0/035 & 0/348 \\ -0/035 & 1/000 & -0/507 \\ 0/348 & -0/507 & 1/000 \end{bmatrix},$$

اگرچه داده‌های جمع‌آوری شده در مرحله اول بوده و

ممکن است فرایند تحت کنترل نباشد، اما با توجه به

عناصر غیرقطری ماتریس همبستگی، همبسته

بودن دو به دو متغیرها واضح است. بنابراین، استفاده

از نمودار کنترل چندمتغیره ضروری است. برای این

منظور با استفاده از رابطه (۱) مقادیر آماره نموداری T^2

محاسبه و در جدول ۱ ارائه شده است. حد کنترل مربوطه

با استفاده از رابطه (۱) و قرار دادن $\alpha = 0/005$ برابر

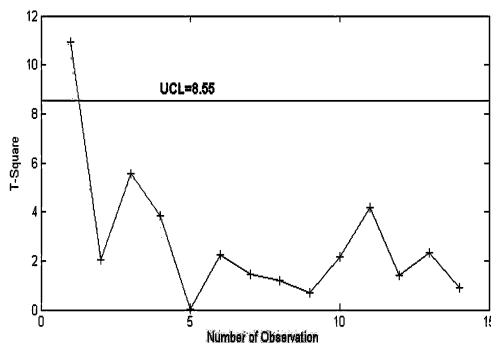
مقدار زیر است:

$$UCL = \frac{(14-1)^2}{14} \times \beta_{(0/005, 1/5, 5)}$$

$$= (12/07)(0/7080) = 8/55.$$

جدول ۱. آماره ها و داده های فرایند صنعتی شیمیایی.

شماره مشاهدات	ناخالصی	دما	غلظت	$T_i^2(n=14)$	$T_i^2(n=13)$
۱	۱۴/۹۲	۸۵/۷۷	۴۲/۲۶	۱۰/۹۳	
۲	۱۶/۹	۸۳/۷۷	۴۳/۴۴	۲/۰۴	۱/۸۴
۳	۱۷/۳۸	۸۴/۴۶	۴۲/۷۴	۵/۵۸	۵/۳۳
۴	۱۶/۹	۸۶/۲۷	۴۳/۶	۳/۸۶	۳/۵۸
۵	۱۶/۹۲	۸۵/۲۳	۴۳/۱۸	۰/۰۴	۰/۲۳
۶	۱۶/۷۱	۸۳/۸۱	۴۳/۷۲	۲/۲۵	۲/۱۷
۷	۱۷/۰۷	۸۶/۰۸	۴۳/۳۳	۱/۴۴	۱/۴۶
۸	۱۶/۹۳	۸۵/۸۵	۴۳/۴۱	۱/۲۱	۱/۰۵
۹	۱۶/۷۱	۸۵/۷۳	۴۳/۲۸	۰/۶۸	۱/۹۱
۱۰	۱۶/۸۸	۸۶/۲۷	۴۲/۵۹	۲/۱۷	۵/۱۶
۱۱	۱۶/۷۳	۸۳/۴۶	۴۴	۴/۱۷	۳/۸۴
۱۲	۱۷/۰۷	۸۵/۸۱	۴۲/۷۸	۱/۴	۱/۶۵
۱۳	۱۷/۶	۸۵/۹۲	۴۳/۱۱	۲/۳۳	۷
۱۴	۱۶/۹	۸۴/۲۳	۴۳/۴۸	۰/۹	۰/۷۷



شکل ۱ نمودار کنترل چندمتغیره T^2 را نشان می‌دهد. در این نمودار مشاهده اول خارج از حدود کنترل قرار گرفته است. برای تشخیص دلیل انحراف، این مشاهده را به طور انفرادی بررسی می‌کنیم و پایین بودن غیرمتعادل سطح ناخالصی آن، علت خطای نمونه‌گیری تشخیص داده می‌شود. به این ترتیب، پس از پی‌بردن به علت انحراف، این مشاهده از نمونه حذف و این بار برآورد پارامترها با $n = 13$ مشاهده محاسبه شده که نتایج زیر به دست آمده است:

$$\bar{X} = \begin{bmatrix} 16/98 \\ 85/14 \\ 43/28 \end{bmatrix},$$

و

$$S_{13} = \begin{bmatrix} 0/068 & 0/076 & -0/055 \\ 0/076 & 1/092 & -0/216 \\ -0/055 & -0/216 & 0/163 \end{bmatrix},$$

و

$$R_{13} = \begin{bmatrix} 1/000 & 0/280 & -0/520 \\ 0/280 & 1/000 & -0/512 \\ -0/520 & -0/512 & 1/000 \end{bmatrix},$$

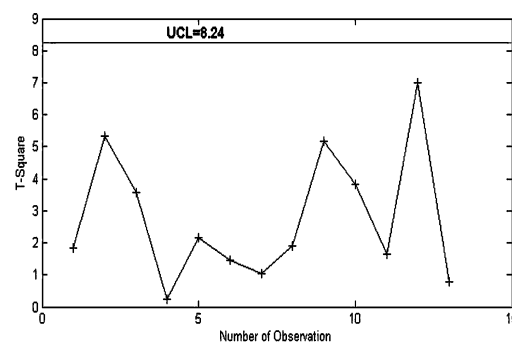
در R_{13} سه متغیر همچنان متقابلاً همبسته‌اند اما علامت همبستگی بین X_1 و X_2 مثبت و بین X_1 و X_3 منفی است. مقادیر T^2 برای ۱۳ مشاهده مجدداً محاسبه شده و در ستون آخر جدول ۱ ارائه شده است. حد بالایی کنترل مربوط به این نمونه، ($n = 13$)، به صورت زیر است:

$$UCL = \frac{(13-1)^2}{13} \times \beta_{(0/005, 1/5, 4/5)} \\ = (11/077)(0/7440) = 8/24$$

باید توجه داشت که با استفاده از توزیع خی-دو، مقدار UCL همواره بزرگتر از مقدار کران بالایی است که با استفاده از توزیع بتا به دست می‌آید. بنابراین احتمال باقی ماندن نقاط پرت در داده‌های اولیه افزایش می‌یابد. در حقیقت، فرض معلوم بودن پارامترها فرضی غیرواقعی است که احتمالاً بر اساس برآورد پارامترها و یا تجارب گذشته بوده و برآوردها به عنوان مقادیر واقعی پارامترها در نظر گرفته شده‌اند. بنابراین، حدود کنترل به دست آمده نمی‌تواند واقعی باشد و لازم است محافظه‌کارانه با آن برخورد شود.

مثال ۲ به منظور توضیح روش فوق و مقایسه آن با حالتی که پارامترها نامعلوم‌اند، فرض کنید که بردار میانگین نمونه و ماتریس کوواریانس آن در مثال ۱، مقادیر واقعی جامعه باشند. با قراردادن $\alpha = 0.005$ ، UCL برابر $12/838 = \chi^2_{(0.005, 3)}$ به دست می‌آید. با مقایسه مقادیر T^2 حاصل، $(n = 14)$ ، در جدول ۱ با این حد بالایی، مشاهده می‌شود که هیچ یک از نقاط به عنوان داده پرت محسوب نشده و بنابراین مشاهده اول نیز حذف نخواهد شد. تفاوت اساسی بین آماره‌های T^2 در (۱) و (۳) به دلیل معلوم و نامعلوم بودن پارامترهای توزیع است که روی توزیع نمونه‌ای و در نتیجه بر مقدار UCL اثر می‌گذارد. جدول ۲ مقادیر UCL حاصل از هر دو توزیع را به ازای چند حجم نمونه نشان می‌دهد. همان‌گونه که مشاهده می‌شود برای مقادیر بزرگ n ، UCL محاسبه شده تحت توزیع بتا به UCL حاصل از توزیع خی-دو نزدیک می‌شود.

در شکل ۲ نمودار با مقادیر جدید T^2 رسم شده و مشاهده می‌شود که هیچ یک از نقاط خارج از محدوده کنترل قرار نگرفته‌اند. به این ترتیب این نمونه ۱۳-تایی برای محاسبه حدود کنترل و نمودار در مرحله دوم مورد استفاده قرار می‌گیرد [۵].



شکل ۲. نمودار کنترل T^2 برای ۱۳ مشاهده باقیمانده

۲.۲ کنترل با پارامترهای معلوم

فرض کنید که مشاهدات در نمونه از توزیع نرمال چندمتغیره با بردار میانگین μ و ماتریس کوواریانس Σ معلوم آمده باشند. بنابراین، برای تک بردار مشاهده‌ای $X' = (x_1, x_2, \dots, x_p)$ آماره T^2 به شکل زیر خواهد بود:

$$T^2 = (X - \mu)' \Sigma^{-1} (X - \mu), \quad (3)$$

که تحت این شرایط دارای توزیع $\chi^2_{(p)}$ است و تنها به p ، تعداد متغیرهای بردار مشاهده‌ای X ، بستگی دارد [۲]. با معلوم بودن سطح α ، حد کنترل بالایی با استفاده از رابطه (۳) برابر $UCL = \chi^2_{(\alpha, p)}$ است که α -امین صدک بالایی توزیع خی-دو با p درجه آزادی است.

جدول ۲. UCL های توزیع بتا و خی-دو با $\alpha = 0.005$

$P = 2$			$P = 6$			$P = 30$		
n	بتا	خی-دو	بتا	خی-دو	بتا	خی-دو	بتا	خی-دو
۴	۲/۲۵	۱۰/۶		۱۸/۵۵		۵۳/۶۷۲		
۱۰	۶/۳۲	۱۰/۶	۸/۵۱	۱۸/۵۵		۵۳/۶۷۲		
۵۰	۹/۷	۱۰/۶	۱۶/۲۵	۱۸/۵۵	۴۰/۱۱	۵۳/۶۷۲		
۱۰۰	۱۰/۱۴	۱۰/۶	۱۷/۳۹	۱۸/۵۵	۴۷/۱۶	۵۳/۶۷۲		
۲۰۰	۱۰/۳۸	۱۰/۶	۱۷/۹۶	۱۸/۵۵	۵۰/۴۶	۵۳/۶۷۲		
۵۰۰	۱۰/۵	۱۰/۶	۱۸/۳۲	۱۸/۵۵	۵۲/۴	۵۳/۶۷۲		

۳ کنترل در مرحله دوم

زیرا در این حالت به دلیل مستقل بودن $X \sim N_p(\mu, \Sigma)$ (که خارج از مجموعه بردارهای مشاهدات در مرحله اول است) از $\bar{X}$ و S محاسبه شده از نمونه اولیه و همچنین با فرض تحت کنترل بودن X با میانگین $E(X) = \mu$ داریم:

$$\sqrt{\frac{n}{n+1}}(X - \bar{X}) \sim N_p(0, \Sigma),$$

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

هدف از کنترل فرایند در این مرحله، تثبیت رابطه بین متغیرها و تضمین باقی ماندن تمام متغیرها در دامنه مطلوب عملکردشان است. در این بخش، ابتدا کنترل فرایند تک مشاهدات برداری را در هر نقطه زمانی با استفاده از آماره T^2 توضیح داده و سپس کنترل فرایند را برای حالتی که میانگین زیرگروهی از مشاهدات در هر نقطه زمانی انتخاب می‌شود، در دو وضعیت پارامترهای معلوم و نامعلوم بسط می‌دهیم.

۱.۳ کنترل با پارامترهای نامعلوم

در نتیجه رابطه (۴) برقرار است. این توزیع به اندازه نمونه داده‌های اولیه، n ، و تعداد متغیرهای توزیع، p ، بستگی دارد [۳].

وقتی که مشاهدات فرایند به جای تنها یک بردار، زیرگروهی شامل m بردار باشد، آماره T^2 برای میانگین i -امین نمونه، $\bar{X}_i$ ، و توزیع آن به صورت زیر است:

$$T_i^2 = (\bar{X}_i - \bar{X})' S^{-1} (\bar{X}_i - \bar{X})$$

$$\sim \frac{(m+n)(n-1)p}{mn(n-p)} F_{(p, n-p)}. \quad (5)$$

زیرا اگر $\bar{X}_i$ بردار میانگین حاصل از m بردار از زیرگروه جدید باشد آن‌گاه مستقل از $\bar{X}$ و S

در یک فرایند پایدار پیوسته که در آن بردار مشاهدات مستقل و از توزیع نرمال چندمتغیره با پارامترهای نامعلوم در نظر گرفته می‌شود، بررسی تحت کنترل بودن فرایند با انتخاب تنها یک بردار مشاهده‌ای $X' = (x_1, x_2, \dots, x_p)$ در هر نقطه زمانی مورد مطالعه قرار می‌گیرد. مقدار T^2 مربوط به بردار جدید مشاهده‌ای و توزیع آن به صورت زیر است:

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

$$\sim \left(\frac{p(n+1)(n-1)}{n(n-p)} \right) F_{(p, n-p)}. \quad (4)$$

محاسبه شده در مرحله اول است. با توجه به این که اینک برای یک زیرگروه جدید خواهیم داشت [۴]:

$$\sqrt{\frac{km}{k+1}}(\bar{X}_i - \bar{X}) \sim N_p(0, \Sigma),$$

و در نتیجه

$$\begin{aligned} & \frac{km}{k+1}(\bar{X}_i - \bar{X})' S_W^{-1}(\bar{X}_i - \bar{X}) \\ & \sim \frac{k(m-1)p}{k(m-1)-p+1} F_{(p, k(m-1)-p+1)}, \end{aligned}$$

و یا

$$\begin{aligned} T^2 &= (\bar{X}_i - \bar{X})' S_W^{-1}(\bar{X}_i - \bar{X}) \\ &\sim \frac{(k+1)(m-1)p}{m(mk-k-p+1)} F_{(p, mk-k-p+1)}. \end{aligned}$$

۲.۳ کنترل با پارامترهای معلوم

اگر پارامترهای توزیع نرمال چندمتغیره جامعه معلوم فرض شوند، توزیع آماره T^2 برای یک بردار مشاهده انفرادی در مرحله دوم $\chi^2_{(p)}$ است و در این صورت حد بالایی کنترل به اندازه نمونه اولیه، m بستگی ندارد. اما تجربه نشان داده است که در عمل چنین موردی به ندرت اتفاق می افتد. با این وجود، کسب اطلاعاتی درباره استفاده و کاربرد حدود توزیع آماره T^2 در این حالت مفید است.

حال اگر مشاهدات فرایند در یک نقطه زمانی خاص به جای تنها یک بردار، زیرگروهی شامل m بردار تصادفی مستقل باشد، میانگین این زیرگروه در محاسبه مقدار آماره نمونه مورد استفاده قرار می گیرد. چنانچه هر یک از بردارهای تصادفی دارای توزیع نرمال p -متغیره،

$$E(\bar{X}_i) = E(\bar{X}) = \mu$$

$$\sqrt{\frac{mn}{m+n}}(\bar{X}_i - \bar{X}) \sim N_p(0, \Sigma),$$

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

در این صورت داریم:

$$\frac{mn}{m+n}(\bar{X}_i - \bar{X})' S^{-1}(\bar{X}_i - \bar{X}) \sim \frac{(n-1)p}{(n-p)} F_{(p, n-p)},$$

و یا

$$\begin{aligned} T_i^2 &= (\bar{X}_i - \bar{X})' S^{-1}(\bar{X}_i - \bar{X}) \\ &\sim \frac{(m+n)(n-1)p}{mn(n-p)} F_{(p, n-p)}. \end{aligned}$$

این توزیع به اندازه نمونه، m اندازه داده های اولیه، n و تعداد متغیرها در بردار مشاهده $\bar{X}_i$ ، p بستگی دارد. در عبارت (۵) به جای S می توان برآورد ترکیبی S_W به صورت زیر را جایگزین کرد

$$S_W = \frac{\sum_{i=1}^k S_i}{k}. \quad (6)$$

بنابراین، در این حالت عبارت S_i برآورد کوواریانس نمونه در i -امین دوره نمونه گیری $i = 1, 2, \dots, k$ است. با این جای گذاری داریم:

$$\begin{aligned} T^2 &= (\bar{X}_i - \bar{X})' S_W^{-1}(\bar{X}_i - \bar{X}) \\ &\sim \frac{(k+1)(m-1)p}{m(mk-k-p+1)} F_{(p, mk-k-p+1)}, \end{aligned} \quad (7)$$

زیرا در این حالت با فرض $E(\bar{X}_i) = E(\bar{X}) = \mu$ برای هر i خواهیم داشت:

$$(m-1)S_i \sim W_p(m-1, \Sigma)$$

و بنابراین

$$k(m-1)S_W = \sum_{i=1}^k (m-1)S_i \sim W_p(k(m-1), \Sigma).$$

همبستگی این داده‌ها به قرار زیر است:

$$\bar{X}' = (۲۵/۰۹۵ \quad ۴۰/۱۸۲ \quad ۷/۱۰۷),$$

$$R = \begin{bmatrix} ۱ & -۰/۰۵۷۹۷ & -۰/۰۶۶۵ \\ -۰/۰۵۷۹۷ & ۱ & -۰/۰۸۰۰ \\ -۰/۰۶۶۵ & -۰/۰۸۰۰ & ۱ \end{bmatrix},$$

$$S = \begin{bmatrix} ۰/۰۲۴۳۳ & -۰/۰۴۵۰ & -۰/۰۰۰۳۲ \\ -۰/۰۴۵۰ & ۰/۰۲۴۷ & -۰/۰۰۰۱۲ \\ -۰/۰۰۰۳۲ & -۰/۰۰۰۱۲ & ۰/۰۰۰۹۷ \end{bmatrix}.$$

مقادیر T^2 با استفاده از برآوردهای فوق و رابطه (۱) محاسبه و در ستون آخر جدول ۳ ارائه شده‌اند. حد کنترل بالایی متناظر با این داده‌ها به صورت زیر محاسبه می‌گردد:

$$UCL = \frac{(10-1)^2}{10} \beta_{(0/05, 4, 10-3-1)} \\ = \frac{81}{10} (0/7040) = 5/7024.$$

با مقایسه مقادیر T^2 و UCL فوق، همان‌گونه که در شکل ۳ نیز مشاهده می‌شود، هیچ یک از بردارهای مشاهده‌ای خارج از ناحیه کنترل قرار نمی‌گیرند. بنابراین از برآوردهای بردار میانگین و کوواریانس این مرحله می‌توان برای اجرای فرایند کنترل در مرحله دوم استفاده نمود. بنابراین داریم:

$$\hat{\mu}' = \bar{X}' = (۲۵/۰۹۵ \quad ۴۰/۱۸۲ \quad ۷/۱۰۷),$$

$$\hat{\Sigma} = S = \begin{bmatrix} ۰/۰۲۴۳۳ & -۰/۰۴۵۰ & -۰/۰۰۰۳۲ \\ -۰/۰۴۵۰ & ۰/۰۲۴۷ & -۰/۰۰۰۱۲ \\ -۰/۰۰۰۳۲ & -۰/۰۰۰۱۲ & ۰/۰۰۰۹۷ \end{bmatrix}.$$

$N_p(\mu, \Sigma)$ باشد آنگاه بردار میانگین m بردار نمونه، $\bar{X}$ دارای توزیع نرمال p -متغیره با بردار میانگین μ و ماتریس کوواریانس $\frac{1}{m}\Sigma$ خواهد بود. در صورتیکه بردارهای تصادفی مستقل از توزیع نرمال چندمتغیره نباشند، با استفاده از قضیه حد مرکزی با بزرگ شدن حجم زیرگروه، m ، توزیع $\bar{X}$ به $N_p(\mu, \frac{1}{m}\Sigma)$ نزدیکتر خواهد شد.

آماره T^2 برای میانگین i -امین زیرگروه، $\bar{X}_i$ ، به صورت زیر است:

$$T_i^2 = (\bar{X}_i - \mu)' \Sigma^{-1} (\bar{X}_i - \mu), \quad (۸)$$

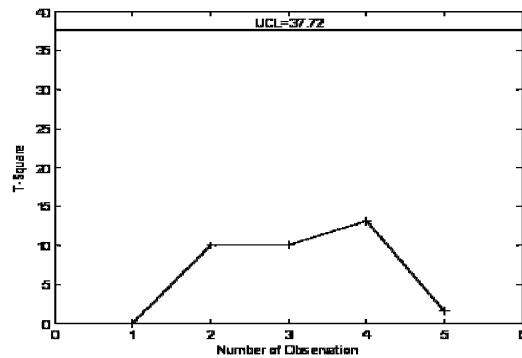
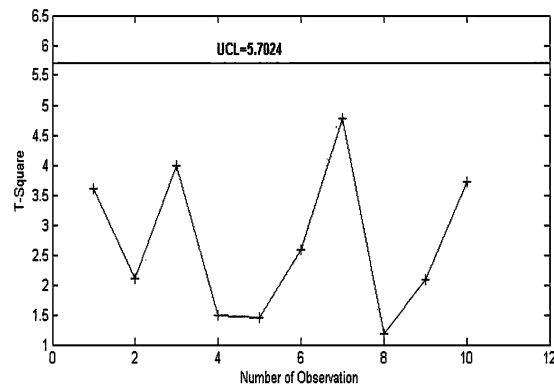
و در نتیجه $UCL = m\chi_{(\alpha, p)}$ خواهد بود.

مثال ۳ داده‌های جدول ۳ اندازه‌گیریهای مربوط به مشاهدات انفرادی ۱۰ قطعه کاسه‌نمد است که در شرکت لاستیک‌سازی نیکوبر شرق سپاهان تولید می‌شود. از این قطعه جهت آببندی استاتیکی و دینامیکی انواع سیالات و گازها استفاده می‌گردد.

جدول ۳. داده‌های کاسه‌نمد و T^2

#م	قطر داخلی	قطر خارجی	ارتفاع	T^2
۱	۲۴/۱۸	۴۰/۳۲	۷/۰۹	۳/۶۱۲۸
۲	۲۵/۲۳	۴۰/۳۲	۷/۱۵	۲/۱۱۲۵
۳	۲۵/۹۲	۳۹/۸۹	۷/۱۲	۳/۹۹۰۶
۴	۲۴/۸۰	۴۰/۲۹	۷/۲	۱/۴۸۸۹
۵	۲۵/۱۸	۴۰/۲۸	۷/۱۷	۱/۴۴۲۴
۶	۲۵/۲۰	۴۰/۲۹	۶/۹۷	۲/۵۹۴۲
۷	۲۵/۱۸	۴۰/۱۳	۶/۹۰	۴/۷۷۷۲
۸	۲۴/۵۲	۴۰/۱۸	۷/۱۲	۱/۱۷۷۶
۹	۲۴/۵۱	۴۰/۱۹	۷/۱۵	۲/۰۷۸۵
۱۰	۲۵/۲۳	۳۹/۹۳	۷/۲	۳/۷۲۵۳

با در نظر گرفتن داده‌های جدول ۳ به عنوان داده‌های اولیه، مرحله اول کنترل این قطعه به صورت زیر انجام می‌گیرد. بردار میانگین، ماتریس کوواریانس و ضرایب

شکل ۴. نمودار T^2 در مرحله دومشکل ۳. نمودار T^2 برای مشاهدات کاسه نمد در مرحله اول

۴ نتیجه گیری

در مواردی که اندازه‌گیری هم‌زمان چند مشخصه کیفی کالایی مورد نظراست، استفاده از نمودارهای کنترل چندمتغیره T^2 ابزار توانمندی جهت تحلیل کیفیت کالای مربوطه می‌باشد. در این حالت ارائه یک نمونه اولیه و تعیین حدود کنترل مبتنی بر برآوردهای دقیق پارامترها بسیار مهم است. در مرحله اول، خصوصاً وقتی زیرگروههایی با مشاهدات انفرادی از متغیرها انجام می‌گیرد، توزیع بتا را باید جهت محاسبه حدود کنترل آماره T^2 به کار برد. استفاده از این توزیع دقیق، بهتر از به کارگیری توزیع تقریبی خی-دو است. همان‌گونه که نشان داده شد با افزایش حجم زیرگروهها حدود کنترل مربوط به توزیع دقیق بتا به حدود کنترل تقریبی خی-دو نزدیک می‌شوند.

جهت کنترل در مرحله دوم پنج مشاهده مستقل انفرادی گرفته می‌شود. این مشاهدات به همراه مقادیر T^2 متناظرشان در جدول ۴ آورده شده‌اند. در مرحله دوم UCL از رابطه (۴) به صورت زیر محاسبه می‌گردد:

$$UCL = \frac{3(11)(9)}{10(7)} F_{(0/05, 3, 7)} = 37/72.$$

با مقایسه حد کنترل بالایی فوق و مقادیر T^2 در جدول ۴ ملاحظه می‌شود که در مرحله دوم نیز هیچ‌گونه شرایط خارج از کنترل را نمایش نمی‌دهد. نمودار T^2 در مرحله دوم در شکل ۴ رسم شده است.

جدول ۴. داده‌های مرحله دوم و مقادیر T^2

ام #	قطر داخلی	قطر خارجی	ارتفاع	T^2
۱	۲۵/۱۲	۴۰/۱۷	۷/۱	۰/۰۱۱۸
۲	۲۵/۱۶	۳۹/۹۹	۶/۸۵	۱۰/۰۳۶۶
۳	۲۵/۶	۴۰/۳۲	۶/۸۴	۱۰/۱۵۶۴
۴	۲۴/۳۵	۴۰/۱۸	۶/۸۳	۱۳/۱۱۹۶
۵	۲۵/۱۱	۴۰/۱۵	۶/۹۹	۱/۵۷۵۰

مراجع

- [۱] توانگر، فریده، طالبی، هوشنگ و علامت ساز، محمدحسین. (۱۳۸۸)، مقایسه نمودار کنترل توأم شوارت و هتلینگ، اندیشه آماری، ۲۷، ۶۴-۵۲.
- [۲] ماردیا، کانتی؛ کنت، جان؛ بی بی، جان (۱۹۷۹). تحلیل چندمتغیره. ترجمه محمد مهدی طباطبائی (۱۳۷۶). مرکز نشر دانشگاهی.
- [3] Mason, R.L.; Young, J.C. (1998). A Hotelling's T^2 Multivariate Statistic for Industrial process Control. ASQ's 52nd Annual Quality Congress Proceedings. pp. 78-85.
- [4] Seber, G.A.F. (1984). Multivariate Observations. Wiley, New York.
- [5] Tracy, N.D.; Young, J.C.; Mason, R.L. (1992). Multivariate Control Chart for Individual Observations. Journal of Quality Technology, 24, pp. 88-95.