مقایسه‌ی برآوردگرهای انحراف معیار، میانگین برد متحرک و میانه‌ی برد متحرک در برآورد انحراف معیار فرآیند با استفاده از داده‌های خود همبسته

مهدی کلانتری[۱] و سید محمد تقی فاطمی قمی[۲*]

[۱] مربی آموزشیار دانشگاه پیام نور مرکز ابهر

[۲] استاد دانشکده‌ی مهندسی صنایع- دانشگاه صنعتی امیرکبیر

(تاریخ دریافت ۸۲/۷/۶ ، تاریخ دریافت روایت اصلاح شده ۸۸/۹/۳، تاریخ تصویب ۸۹/۱/۱۷)

چکیده

استقلال مشاهده‌های فرضی اساسی در طراحی یک نمودار X برای مشاهده‌های انفرادی است. ولی متأسفانه در بعضی از فرآیندهای تولیدی فرض وجود داده‌های مستقل حتی به طور تقریبی نیز برقرار نیست و داده‌ها از خود نوعی خود همبستگی نشان می‌دهند. در این مقاله با شبیه‌سازی یک فرآیند AR(1) نشان خواهیم داد که وجود خودهمبستگی در داده‌ها، تأثیر عمیقی بر نمودار X می‌گذارد. علاوه بر این با مقایسه‌ی برآوردگرهای انحراف معیار، میانگین برد متحرک و میانه‌ی برد متحرک، بهترین آن‌ها برای برآورد انحراف معیار فرآیند با استفاده از داده‌های خود همبسته، معرفی شده و در نهایت دو برآوردگر جدید انحراف معیار فرآیند با استفاده از داده‌های خود همبسته ارائه می‌شود.

واژه‌های کلیدی: نمودارX، داده‌های مستقل، خودهمبستگی، فرآیند AR(1)، نرمال چند متغیره، UMVUE، آماره‌ی بسنده‌ی مینیمال کامل، برآوردگر نااریب

مقدمه

یک فرض اساسی در طراحی نمودار کنترل شوهارت این است که مشاهده‌های به دست آمده از یک فرآیند تولید، باید از نظر آماری مستقل باشند. باید توجه داشت که شرط استقلال، شرطی بسیار ایده‌آل است و در عمل به ندرت برقرار می‌شود. زیرا با توسعه‌ی روز افزون فناوری اندازه‌گیری و ساخت وسایل و تجهیزات پیشرفته‌ی اندازه‌گیری و حسگرهای کامپیوتری در دهه‌های اخیر، امروزه می‌توان مشخصه‌های کیفی اکثر محصولات تولیدی را برای بهبود کیفیت آن‌ها اندازه‌گیری کرد. بدین ترتیب می‌توان فاصله‌ی بین نمونه‌گیری‌های متوالی را کاهش داد و محصولات بیشتری را کنترل کرد. ولی به تجربه ثابت شده است که اگر فاصله‌ی بین نمونه‌گیری‌ها کاهش یابد، مشاهده‌های دیگر از نظر آماری مستقل نخواهند بود و از خود، نوعی خودهمبستگی نشان می‌دهند. متأسفانه در بعضی از فرآیندهای تولیدی، فرض وجود داده‌های مستقل حتی به طور تقریبی نیز برقرار نیست. به عنوان مثال، می‌توان به فرآیندهای شیمیایی که در آن‌ها مشخصه‌های محصول یا فرآیند از همبستگی زیادی بهره‌مند هستند و به طور متوالی نیز اندازه‌گیری می‌شوند و یا به روش‌های آزمایش و بازرسی خودکار که در آن‌ها هر مشخصه‌ی کیفی محصول در فواصل زمانی معین اندازه‌گیری می‌شود، اشاره کرد. بنابراین فرض پایه‌ای طراحی نمودار کنترل نقض می‌شود. بدین ترتیب با برقرار نبودن این فرض، با سؤال اساسی زیر مواجه می‌شویم:

آیا استفاده از نمودارهای کنترل شوهارت مرسوم برای مشاهده‌هایی که مستقل نیستند، تأثیر منفی روی استنباط‌های ما در مورد فرآیند خواهد گذاشت؟

در این مقاله، ابتدا با شبیه‌سازی یک فرآیند AR(1) نشان می‌دهیم که وجود همبستگی بین مشاهده‌ها، چگونه می‌تواند اثر نامطلوبی روی عملکرد نمودار X داشته باشد و سبب شود که زنگ خطرهای اشتباهی افزایش یابد. سپس برای کاهش اثرات سوء همبستگی در مشاهده‌ها، استفاده از دو برآوردگر جدید پیشنهاد خواهد شد.

نمودار X

در بسیاری از موقعیت‌های صنعتی، برای کنترل فرآیند از نمونه‌ای به اندازه‌ی یک استفاده می‌شود (نظیر بسیاری از فرآیندهای شیمیایی) و یا به عبارت دیگر نمونه فقط شامل یک محصول است. در چنین مواردی نمونه‌های

* نویسنده مسئول : تلفن : ۶۴۵۴۲۷۸۴ , فاکس : ۶۴۵۴۲۷۸۴ ، Email: fatemi@aut.ac.ir

تکی در فواصل زمانی معین از فرآیند گرفته شده و سپس برای کنترل فرآیند با استفاده از این نمونه‌های تکی یا مشاهده‌های انفرادی[1]، از نمودار X استفاده می‌شود. از این نمودار اغلب در فرآیندهای شیمیایی که دما یا غلظت فرآیند اندازه‌گیری می‌شوند و یا صنایعی که از یک سیستم نظارت و اندازه‌گیری خودکار بهره می‌برند، استفاده می‌شود. علاوه بر این، نمودار X در صنایعی که سرعت خط تولید کند بوده و هزینه‌ی اندازه‌گیریِ مشخصه‌ی کیفی محصول بالا است، مفید واقع می‌شود. حدود کنترل این نمودار با استفاده از رابطه‌ی (۱) محاسبه می‌شوند.

$$\begin{aligned} UCL &= \mu + 3\sigma_p \\ CL &= \mu \\ LCL &= \mu - 3\sigma_p \end{aligned} \tag{۱}$$

در این رابطه μ و σ_p به ترتیب نشان دهنده‌ی میانگین و انحراف معیار فرآیند بوده و در عمل مجهول هستند و باید آن‌ها را با استفاده از مشاهده‌هایی که در زمان تحت کنترل بودن فرآیند تهیه شده‌اند، برآورد کرد.

مدل شوهارت

فرض کنید X_t نشان دهنده‌ی یک مشخصه‌ی کیفی در زمان t باشد. در کنترل فرآیند آماری اغلب فرض می‌شود که X_t توزیع نرمال با میانگین μ و انحراف معیار σ_p دارد. حال اگر X_tها مستقل باشند، می‌توان مدل زیر را برای مشاهده‌ها به کار برد:

$$X_t = \mu + \varepsilon_t \qquad t = 1,2,... \tag{۲}$$

در این رابطه $\{\varepsilon_t\} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$. این مدل را اغلب مدل شوهارت[2] فرآیند می‌نامند. توجه کنید که در این مدل باید داشته باشیم:

$$\sigma_p^2 = Var(X_t) = Var(\varepsilon_t) = \sigma_\varepsilon^2 \tag{۳}$$

برای محاسبه‌ی حدود کنترل نمودار X، نیاز به برآوردی از مقادیر μ و σ_p داریم. بدین منظور فرض می‌کنیم $X_1, X_2, ..., X_n$ نمونه‌هایی هستند از جامعه‌ای نرمال با میانگین μ و انحراف معیار σ_p که بر حسب زمانِ نمونه‌گیری مرتب شده‌اند. μ را اغلب با استفاده از میانگین حسابی نمونه‌ها یعنی $\overline{X}$ برآورد می‌کنند. ولی مسئله‌ی اصلی و مهم، برآورد σ_p است که به روش‌های متعددی می‌توان آن را برآورد کرد. بسته به اینکه از چه روشی برای برآورد σ_p استفاده کنیم، حدود کنترل نمودار X نیز متفاوت خواهند بود. در بخش‌های بعدی سه برآوردگر برای برآورد σ_p با استفاده از داده‌های مستقل، معرفی می‌شود.

برآوردگر انحراف معیار

یکی از روش‌های معروف و پرکاربرد برای برآورد انحراف معیار فرآیند، استفاده از انحراف معیار نمونه است که به این ترتیب تعریف می‌شود:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \tag{۴}$$

می‌دانیم که S یک برآوردگر ناااریب برای σ_p نیست. یک برآوردگر نااریب برای σ_p بر اساس انحراف معیار نمونه، برآوردگر $\hat{\sigma}_1$ است که به این شکل تعریف می‌شود:

$$\hat{\sigma}_1 = \frac{S}{c} \tag{۵}$$

و به برآوردگر انحراف معیار معروف است. c عددی است که بستگی به تعداد نمونه دارد و به این ترتیب تعریف می‌شود:

$$c = \sqrt{\frac{2}{n-1}}\,\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \tag{۶}$$

مقادیر مختلف c به ازای تعداد نمونه‌های متفاوت را می‌توان در مرجع [۱] یافت. از آنجا که $\hat{\sigma}_1$ تابعی از آماره‌ی بسنده‌ی مینیمال کاملِ S^2 بوده و نااریب نیز هست، بنا به قضیه‌ی لهمن- شفه[3] [۲] یک برآوردگر نااریب با کمترین واریانس به طور یکنواخت[4] برای σ_p خواهد بود.

برآوردگر میانگین برد متحرک

بسیاری از نویسندگان مثل دانکن[5] [۳]، نلسون[6] [۴]، وادوورس، استفنس و گادفری[7] [۵] استفاده از برآوردگر دیگری را برای برآورد انحراف معیار فرآیند پیشنهاد می‌کنند که به برآوردگر **میانگین برد متحرک**[8] معروف است و به این ترتیب است:

$$\hat{\sigma}_2 = \frac{\sqrt{\pi}}{2(n-1)}\sum_{i=2}^{n} MR_i = \frac{\sqrt{\pi}}{2}\overline{MR} \tag{۷}$$

در این رابطه MR_i که برد متحرک[۹] نامیده می‌شود، به این شکل تعریف می‌شود:

$$MR_i = |X_i - X_{i-1}| \qquad i = 2,..., n \tag{۸}$$

$\hat{\sigma}_2$ نیز برآوردگری نااریب برای σ_p است [۳]. تجربه نشان داده است که استفاده از این برآوردگر، هم در کنترل کیفیت یک متغیره و هم در کنترل کیفیت چند متغیره، بسیار مفید خواهد بود. شایان ذکر است که در اکثر کتاب‌ها و مقالات مربوط به کنترل کیفیت آماری از این برآوردگر برای طراحی یک نمودار X استفاده می‌شود.

برآوردگر میانه‌ی برد متحرک

کلیفورد[۱۰] [۶] برآوردگر نااریب دیگری را معرفی کرده است که به جای میانگین MR_iها از میانه‌ی MR_iها استفاده می‌کند و به این صورت است:

$$\hat{\sigma}_3 = 1.047\ \tilde{M} \tag{۹}$$

در این رابطه $\tilde{M}$ نشان دهنده‌ی میانه‌ی MR_iها است. این برآوردگر به برآوردگر میانه‌ی برد متحرک معروف است.

مدل AR(1)

همه‌ی فرآیندهای تولید تحت تأثیر عناصر و نیروهای خاص خود قرار دارند و اگر فاصله‌ی بین نمونه‌ها نسبت به این نیروها کاهش یابد، آن گاه بین مشاهده‌هایی که در طول زمان از فرآیند مورد نظر تهیه می‌شوند، همبستگی وجود خواهد داشت. به طور کلی اگر به هر دلیلی بین مشاهده‌ها همبستگی وجود داشته باشد، مدل شوهارت اعتبار خود را از دست خواهد داد. در مدل شوهارت فرض می‌شود که $\{X_t - \mu\} \sim IID\left(0, \sigma_p^2\right)$، یعنی سری زمانی $\{X_t - \mu\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با میانگین صفر و واریانس σ_p^2 است. در کنترل کیفیت آماری، اغلب فرآیندهایی که همبستگی داشته و در عمل با آن‌ها مواجه می‌شویم، یک فرآیند AR(1) هستند. بنابراین مدل (۱۰) در اکثر کاربردها مفید خواهد بود.

$$\{X_t - \mu\} \sim AR(1) \tag{۱۰}$$

یا به عبارت دیگر خواهیم داشت:

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t \tag{۱۱}$$

در این رابطه ϕ ضریب اتورگرسیو نامیده می‌شود و $|\phi|<1$. همچنین $\{\varepsilon_t\} \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$. توجه کنید که به ازای $\phi = 0$ همان مدل شوهارت نتیجه می‌شود. میانگین، واریانس، تابع اتوکوواریانس و تابع خود همبستگی فرآیند عبارتند از [۷]:

$$E(X_t) = \mu \tag{۱۲}$$

$$Var(X_t) = \sigma_p^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2} \tag{۱۳}$$

$$\gamma_X(k) := Cov(X_{t+k}, X_t) = \frac{\phi^k \sigma_\varepsilon^2}{1-\phi^2} \qquad k = 0,1,... \tag{۱۴}$$

$$\rho_X(k) := Corr(X_{t+k}, X_t) = \phi^k \qquad k = 0,1,... \tag{۱۵}$$

تأثیر خود همبستگی بر نمودار X

فرض کنید فرآیندی از مدل (۱۰) پیروی کند، ولی هنگام طراحی یک نمودار کنترل از وجود همبستگی بین مشاهده‌ها بی‌اطلاع باشیم و یا نادیده بگیریم. در این صورت، به اشتباه نمودار کنترلی به کار خواهیم برد که برای مشاهده‌های مستقل مناسب است. اگر نمودار X را به کار بریم حدود کنترل به صورت $\mu \pm 3\sigma_p$ محاسبه خواهند شد.

برای نشان دادن تأثیر به کارگیری نمودار X برای مشاهده‌های همبسته، از شبیه‌سازی استفاده کرده‌ایم. برای این منظور ۱۰۰ مشاهده از یک فرآیند AR(1) با میانگین صفر، به ازای مقادیر مختلف $\phi = -0.9,-0.6,-0.3,0,0.3,0.6,0.9$، تولید شده و این عمل ۱۰۰۰۰ مرتبه تکرار شد. در این شبیه‌سازی σ_ε^2 برابر یک فرض شده است. نتایج این شبیه‌سازی در جدول (۱) ارائه شده است.

در ستون اول جدول (۱) مقادیر مختلف ϕ درج شده است. ستون دوم نیز نشان دهنده‌ی انحراف معیار فرآیند (σ_p) است که از رابطه‌ی $\frac{1}{\sqrt{1-\phi^2}}$ به ازای مقادیر مختلف ϕ به دست آمده است. در ستون سوم، انحراف معیار فرآیند شبیه‌سازی شده با استفاده از برآوردگر انحراف معیار به ازای مقادیر مختلف ϕ برآورد شده است. این مقادیر تفاوت آشکاری با مقادیر دقیق انحراف معیار فرآیند ندارند، به عبارت دیگر $\hat{\sigma}_1$ در برآورد σ_p عملکردی

به نسبت موفق داشته است. این موضوع اتفاقی نیست. برای روشن شدن موضوع به قضیه‌ی زیر توجه کنید.

قضیه‌ی ۱: فرض کنید $X_1, X_2, \dots, X_n$ نمونه‌هایی از یک فرآیند تصادفی با میانگین μ و واریانس σ_p^2 و تابع اتوکوواریانس $\gamma_X(\cdot)$ باشند. در این صورت داریم:

$$E\left(S^2\right) = \sigma_p^2 - \frac{2}{n-1}\sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\gamma_X(k) \qquad (۱۶)$$

اثبات: بنا به تعریف S^2 داریم:

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = \sum_{i=1}^{n}\left(X_i-\mu+\mu-\bar{X}\right)^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2 + n(\bar{X}-\mu)^2 - 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)\\ &= \sum_{i=1}^{n}(X_i-\mu)^2 + n(\bar{X}-\mu)^2 - 2(\bar{X}-\mu)(n\bar{X}-n\mu)\\ &= \sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2 \qquad (۱۷)\end{aligned}$$

اگر از دو طرف تساوی امید ریاضی بگیریم خواهیم داشت:

$$\begin{aligned}(n-1)E\left(S^2\right) &= \sum_{i=1}^{n}E\left(X_i-\mu\right)^2 - nE\left(\bar{X}-\mu\right)^2\\ &= \sum_{i=1}^{n}Var\left(X_i\right) - nVar\left(\bar{X}\right)\\ &= n\sigma_p^2 - n\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(X_i, X_j\right)\right]\\ &= n\sigma_p^2 - n\left[\frac{1}{n^2}\left(\sum_{i=1}^{n}\gamma_X(0) + 2\sum_{k=1}^{n}(n-k)\gamma_X(k)\right)\right]\\ &= n\sigma_p^2 - \sigma_p^2 - \frac{2}{n}\sum_{k=1}^{n}(n-k)\gamma_X(k)\\ &= (n-1)\sigma_p^2 - 2\sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\gamma_X(k) \qquad (۱۸)\end{aligned}$$

در نتیجه داریم:

$$E\left(S^2\right) = \sigma_p^2 - \frac{2}{n-1}\sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\gamma_X(k)$$

از قضیه‌ی ۱ می‌توان این نتایج را گرفت.

نتیجه‌ی ۱: اگر $X_1, X_2, \dots, X_n$ مشاهده‌هایی مستقل باشند، آن گاه به ازای هر k، $\gamma_X(k)=0$ در نتیجه S^2 برآوردگری نااریب برای واریانس فرآیند خواهد بود.

نتیجه‌ی ۲: اگر تابع $\gamma_X(\cdot)$ این خاصیت را داشته باشد که با افزایش تأخیر[11] (k) به سرعت کاهش یابد، S^2 به طور مجانبی نااریب برای واریانس فرآیند خواهد بود.

نتیجه‌ی ۳: اگر سری $\sum_{k=1}^{\infty}\gamma_X(k)$ همگرا باشد، S^2 نااریب مجانبی برای σ_p^2 خواهد بود.

توجه داشته باشید که تابع $\gamma_X(\cdot)$ فرآیند AR(1) در شرایط نتایج ۲ و ۳ صدق می‌کند. بنابراین اگر $\{X_t-\mu\}\sim AR(1)$ آن گاه S^2 برآوردگری نااریب مجانبی برای σ_p^2 خواهد بود. به همین دلیل $\hat{\sigma}_1$ در برآورد σ_p عملکردی به نسبت موفق داشته است.

ستون چهارم جدول (۱) مقادیر برآورد شده‌ی انحراف معیار فرآیند را با استفاده از برآوردگر میانگین برد متحرک، نشان می‌دهد. با مقایسه‌ی این ستون و ستون دوم جدول (۱)، در می‌یابیم که $\hat{\sigma}_2$ در برآورد σ_p موفق نبوده است. زیرا به ازای ϕهای منفی، σ_p بیش برآورد[12] شده است و به ازای ϕهای مثبت، زیر برآورد[13]. این موضوع اتفاقی نیست و به کمک قضیه‌ی ۲ می‌توان به درستی این موضوع اطمینان حاصل کرد.

قضیه‌ی ۲: فرض کنید $\{X_t\}$ فرآیندی تصادفی با میانگین μ و تابع اتوکوواریانس $\gamma_X(\cdot)$ باشد. اگر:

$\begin{pmatrix} X_i \\ X_{i-1}\end{pmatrix} \sim N_2\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \begin{pmatrix}\sigma_p^2 & \gamma_X(1)\\ \gamma_X(1) & \sigma_p^2\end{pmatrix}\right)$ آن گاه $E\left(\hat{\sigma}_2\right) = \sigma_p\sqrt{1-\rho_X(1)}$.

جدول۱: نتایج شبیه‌سازی فرآیند AR(1).

ϕ	σ_p	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_3$
-0.9	2.294157	2.260499	4.023418	2.989129
-0.6	1.25	1.251751	2.013725	1.457981
-0.3	1.048285	1.050467	1.522056	1.096396
0	1	1.000753	1.274876	0.91731
0.3	1.048285	1.044383	1.116973	0.801959
0.6	1.25	1.22882	1.005788	0.722342
0.9	2.294157	2.060237	0.923472	0.663253

اثبات: می‌دانیم که اگر $\underset{\sim}{a}$ برداری $p\times 1$ از مقادیر ثابت و $\underset{\sim}{Y}$ برداری $p\times 1$ از متغیرهای تصادفی باشد و $\underset{\sim}{Y} \sim N_p\left(\underset{\sim}{\mu},\Sigma\right)$ آن گاه $\underset{\sim}{a}'\underset{\sim}{Y} \sim N\left(\underset{\sim}{a}'\underset{\sim}{\mu},\underset{\sim}{a}'\Sigma\underset{\sim}{a}\right)$ [۸]. قرار می‌دهیم $\underset{\sim}{a}=\begin{pmatrix}1\\-1\end{pmatrix}$ و $\underset{\sim}{Y}=\begin{pmatrix}X_i\\X_{i-1}\end{pmatrix}$ در نتیجه خواهیم داشت:

$$X_i - X_{i-1} \sim N\left(0,2\left(\sigma_p^2-\gamma_X(1)\right)\right) \quad (۱۹)$$

و واضح است که:

$$\frac{X_i - X_{i-1}}{\sqrt{2\left(\sigma_p^2-\gamma_X(1)\right)}} \sim N(0,1) \quad (۲۰)$$

از طرفی می‌دانیم که اگر $Z \sim N(0,1)$ آن گاه:

$$E(|Z|)=\sqrt{\frac{2}{\pi}} \quad (۲۱)$$

بنابراین

$$\begin{aligned}&E\left(|X_i-X_{i-1}|\right)=E(MR_i)\\&=\frac{2}{\sqrt{\pi}}\sqrt{\sigma_p^2-\gamma_X(1)}=\frac{2\sigma_p}{\sqrt{\pi}}\sqrt{1-\rho_X(1)}\end{aligned} \quad (۲۲)$$

در نتیجه داریم:

$$E(\hat{\sigma}_2)=\frac{\sqrt{\pi}}{2(n-1)}\sum_{i=2}^{n}E(MR_i)=\sigma_p\sqrt{1-\rho_X(1)} \quad (۲۳)$$

با استفاده از رابطه‌ی (۱۵) داریم $\rho_X(1)=\phi$ بنابراین $E(\hat{\sigma}_2)=\sigma_p\sqrt{1-\phi}$. اگر $\phi<0$ آن گاه $\sqrt{1-\phi}>1$ در نتیجه $E(\hat{\sigma}_2)>\sigma_p$. به عبارت دیگر در حالتی که $\phi<0$ برآوردگر میانگین برد متحرک، انحراف معیار فرآیند را بیش برآورد می‌کند و اگر $\phi>0$ آن گاه $\sqrt{1-\phi}<1$ در نتیجه $E(\hat{\sigma}_2)<\sigma_p$. به عبارت دیگر در حالتی که $\phi>0$ برآوردگر میانگین برد متحرک، انحراف معیار فرآیند را زیر برآورد می‌کند.

ستون پنجم جدول (۱) مقادیر برآورد شده‌ی انحراف معیار فرآیند را با استفاده از برآوردگر میانه‌ی برد متحرک نشان می‌دهد. وضعیت این ستون تا حدودی شبیه ستون چهارم است. یعنی اینکه σ_p به ازای ϕهای منفی بیش برآورد شده است و به ازای ϕهای مثبت زیر برآورد.

مقایسه‌ی برآوردگرهای $\hat{\sigma}_1$، $\hat{\sigma}_2$ و $\hat{\sigma}_3$

همان طور که در بخش قبل اشاره شد، برآوردگر $\hat{\sigma}_1$ در برآورد σ_p عملکردی به نسبت موفق داشته است. ولی $\hat{\sigma}_2$ و $\hat{\sigma}_3$ چنین نیستند. استفاده از روش میانگین برد متحرک برای برآورد انحراف معیار فرآیند، کاربرد بسیار گسترده‌ای در کنترل کیفیت آماری دارد. ولی وقتی که بین داده‌ها خودهمبستگی وجود داشته باشد، این روش بسیار ناکارا است، زیرا همان طور که نشان داده شد، اگر $\phi>0$، به دلیل زیر برآورد شدن σ_p حدود کنترل به یکدیگر خیلی نزدیک می‌شوند و در نتیجه تمایل نمودار کنترل برای ایجاد زنگ خطرهای اشتباهی افزایش می‌یابد. در حالتی هم که $\phi<0$ است، به دلیل بیش برآورد شدن σ_p، حدود کنترل از همدیگر دور می‌شوند و در نتیجه حساسیت نمودار کنترل نسبت به پی بردن به وجود انحراف‌ها با دلیل کاهش می‌یابد. این نکته را باید به یاد داشت که در اکثر کاربردها با حالت $\phi>0$ مواجه می‌شویم.

در بخش بعد برای کاهش اثرات سوء همبستگی در مشاهده‌ها، دو برآوردگر جدید پیشنهاد خواهد شد.

دو برآوردگر جدید

در قضیه‌ی ۱ ثابت کردیم که:

$$E\left(S^2\right)=\sigma_p^2-\frac{2}{n-1}\sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\gamma_X(k)$$

یا به عبارت دیگر:

$$E\left(S^2\right)=\sigma_p^2\left(1-\frac{2}{n-1}\sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\rho_X(k)\right) \quad (۲۴)$$

تعریف می‌کنیم:

$$a^2=1-\frac{2}{n-1}\sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\rho_X(k) \quad (۲۵)$$

در نتیجه خواهیم داشت:

$$E\left(\frac{S^2}{a^2}\right)=\sigma_p^2 \quad (۲۶)$$

یعنی $\frac{S^2}{a^2}$ برآوردگری ناریب برای σ_p^2 است. بنابراین یک برآوردگر جدید برای برآورد σ_p به صورت:

$$\hat{\sigma}_4=\frac{S}{a} \quad (۲۷)$$

تعریف می‌کنیم. مزیت این برآوردگر نسبت به برآوردگر $\hat{\sigma}_1$ این است که بر خلاف $\hat{\sigma}_1$، خودهمبستگی در داده‌ها در نظر گرفته شده است.

بنا به رابطه‌ی (۲۳) واضح است که:

$$E\left(\frac{\hat{\sigma}_2}{\sqrt{1-\rho_X(1)}}\right)=\sigma_p \tag{۲۸}$$

و یا

$$E\left(\frac{1}{2}\sqrt{\frac{\pi}{1-\rho_X(1)}}\,\overline{MR}\right)=\sigma_p \tag{۲۹}$$

بنابراین یک برآوردگر ناریب جدید برای برآورد σ_p را به این ترتیب تعریف می‌کنیم:

$$\hat{\sigma}_5=\frac{1}{2}\sqrt{\frac{\pi}{1-\rho_X(1)}}\,\overline{MR} \tag{۳۰}$$

مزیت این برآوردگر نسبت به برآوردگر $\hat{\sigma}_2$ این است که بر خلاف $\hat{\sigma}_2$، خودهمبستگی در داده‌ها در نظر گرفته شده است. توجه داشته باشید که برای استفاده از برآوردگرهای $\hat{\sigma}_4$ و $\hat{\sigma}_5$ نیاز به مقادیر تابع خودهمبستگی ($\rho_X(\cdot)$) داریم. مقادیر تابع $\rho_X(\cdot)$ در عمل مجهول است و باید آن‌ها را با استفاده از داده‌ها برآورد کرد. برای برآورد آن می‌توان از تابع خودهمبستگی نمونه‌ای که به این شکل است، استفاده کرد [۷].

$$\hat{\rho}(k)=\frac{\sum_{t=1}^{n-k}(x_{t+k}-\bar{x})(x_t-\bar{x})}{\sum_{t=1}^{n}(x_t-\bar{x})^2}\qquad k=1,2,\ldots,n \tag{۳۱}$$

نتیجه‌گیری

در بخش ۸ نشان دادیم که وجود همبستگی بین مشاهده‌ها، اثر نامطلوبی بر عملکرد نمودار X دارد. هریس و راس[۱۴] [۹] نیز تأثیر خودهمبستگی بر عملکرد نمودارهای کنترل را بررسی کرده‌اند. آن‌ها دریافتند که خودهمبستگی، تأثیر زیادی بر توانایی نمودارهای کنترل برای کشف تغییری در سطح میانگین فرآیند، ندارد. ولی وجود خودهمبستگی حتی به مقدار کم، عملکرد نمودار در حالت تحت کنترل را تحت تأثیر قرار می‌دهد. آن‌ها نیز افزایش قابل توجه زنگ خطرهای اشتباهی در حالت تحت کنترل را گزارش کردند. بنابراین وجود تعداد زیادی مشاهده‌ی خارج از حدود سه سیگمای نمودار X می‌تواند وجود خودهمبستگی در داده‌ها را تأیید کند. به عبارت دیگر توصیه می‌شود، اگر با مشاهده‌های خارج از کنترل زیادی مواجه شدیم، به طور حتم استقلال مشاهده‌ها را آزمون کنیم. در انتهای فصل اول مرجع [۷] چندین روش برای آزمون مستقل بودن داده‌ها ارائه شده است.

به طور کلی استفاده از نمودارهای کنترل شوهارت مرسوم برای مشاهده‌های خودهمبسته، دو پیامد منفی به دنبال خواهد داشت:

۱- افزایش زنگ خطرهای اشتباهی؛ یعنی تعداد دفعاتی که محصولات مرغوب به اشتباه جزو محصولات نامرغوب به حساب آورده می‌شود، افزایش خواهد یافت و این خود سبب می‌شود که کارخانه‌ی سازنده متحمل هزینه‌های اضافی و سنگینی شود.

۲- کاهش توانایی نمودار کنترل نسبت به کشف مشاهده‌های خارج از کنترل؛ یعنی محصولاتی نامرغوب تولید می‌شوند، ولی نمودار کنترل قادر به شناسایی همه‌ی این محصولات نیست؛ در نتیجه برخی از محصولات نامرغوب به اشتباه جزو محصولات مرغوب به حساب آورده می‌شود و این خود سبب نارضایتی مشتری و از بین رفتن اعتبار کارخانه‌ی سازنده خواهد شد.

برای کاهش اثرات سوء همبستگی بین مشاهده‌ها پیشنهاد می‌شود که از دو برآوردگر جدید $\hat{\sigma}_4$ و $\hat{\sigma}_5$ و یا برآوردگر $\hat{\sigma}_1$ استفاده شود. بررسی ویژگی‌های آماری و مقایسه‌ی این برآوردگرها، نیاز به تحقیق دارد. ایده‌ی اصلی برای به دست آوردن برآوردگرهای جدید $\hat{\sigma}_4$ و $\hat{\sigma}_5$، قضایای ۱ و ۲ بودند؛ ولی به نظر می‌رسد قضیه‌ی مشابهی در مورد برآوردگر $\hat{\sigma}_3$ وجود ندارد و نیاز به تحقیق دارد.

مراجع

1- Montgomery, D. C. (1997). *Introduction to Statistical Quality Control*. 3rd Ed., John Wiley & Sons, New York.

2- Casella, G. and Berger, R. L. (1990). *Statistical Inference*. Duxbury Press, California.

3- Duncan, A. J. (1986). *Quality Control and Industrial Statistics*. 5th Ed., Richard D. Irvin, Homewood, IL.

4- Nelson, L. S. (1975)."Use of the ranges to estimate variability." *Journal of Quality Technology*, Vol. 7, PP. 46-48.

5- Wadsworth, H. M., Jr Stephens, K. S. and Godfrey, A. B. (1986). *Modern Methods for Quality Control and Improvement*, John Wiley & Sons, New York.

6- Clifford, P. C. (1959)."Control charts without calculations." *Industrial Quality Control*, Vol. 15,PP. 2-6.

7- Brockwell, P. J. and Davis, A. R. (1996). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York.

8- Rencher, A. C. (2000). *Linear Models in Statistics*. John Wiley & Sons, New York.

9- Harris, T. J. and Ross, W. M. (1991)."Statistical process control for correlated observations." *Canadian Journal of Chemical Engineering*, Vol. 69, PP. 48-57.

واژه‌های انگلیسی به ترتیب استفاده در متن

1-Individual Observations
2- Shewhart Model
3- Lehmann-Scheffe
4- Uniform Minimum Variance Unbiased Estimator (UMVUE)
5- Duncan
6- Nelson
7- Wadsworth, Stephens and Godfrey
8- Mean Moving Range
9- Moving Range
10- Clifford
11- Lag
12- Over Estimate
13- Under Estimate
14- Harris and Ross