

بررسی عملکرد روش‌های تخمین داده‌های از دست رفته‌ی فازی بر شاخص توانایی فرایند دومتغیره‌ی فازی C_{pm}

عرفانه نیکزاد (دانشجوی دکتری)

عاطفه عشوری (کارشناس ارشد)

امیرحسین امیری* (دانشیار)

گروه مهندسی صنایع، دانشگاه شاهد

علی سلماسنیا (استادیار)

گروه مهندسی صنایع، دانشگاه قم

مهندسی صنایع و مدیریت شریف، تابستان ۱۳۹۶ (۱۳۹۶-۷۳-۶۵)
دوری ۱، شماره ۱/۱، ص. ۷۳-۶۵

شاخص‌های توانایی فرایند به منظور ارزیابی توانایی فرایندها در تأمین خواسته‌های مشتریان کاربرد دارد. بسیاری از مشخصه‌های کیفی عملاً دارای عدم قطعیت هستند و به صورت اعداد فازی نشان داده می‌شوند. اغلب در بین مشاهدات داده‌های از دست رفته‌ی وجود دارد که بر عملکرد شاخص‌های توانایی فرایند تأثیر بسیاری دارد. تاکنون تأثیر روش‌های تخمین داده‌های از دست رفته بر عملکرد شاخص‌های توانایی فرایند فازی بررسی نشده است. در تحقیق حاضر دو روش میانگین و رگرسیون به منظور تخمین داده‌های از دست رفته فازی با در نظر گرفتن برش‌های α توسعه داده شده است. سپس عملکرد روش‌های توسعه‌یافته بر نتایج شاخص توانایی فرایند فازی C_{pm} در حضور داده‌های از دست رفته، تحت مقادیر مختلف ضریب همبستگی و درصد داده‌های از دست رفته بررسی می‌شود. نتایج نشان می‌دهد که روش رگرسیون از عملکرد بهتری نسبت به روش میانگین در تخمین این داده‌ها برخوردار است.

واژگان کلیدی: شاخص توانایی فرایند C_{pm} ، داده‌های از دست رفته، مشاهدات چندمتغیره‌ی فازی، نظریه‌ی مجموعه‌های فازی، α - برش‌ها، الگوریتم بهینه‌سازی اجتماع ذرات.

۱. مقدمه

«شاخص توانایی فرایند» ابزار بسیار قدرتمندی است که در دو دهه‌ی اخیر برای تعیین توانایی بالقوه و بالفعل فرایندها در تأمین خواسته‌های مشتریان مورد استفاده قرار می‌گیرند. شناخته شده‌ترین شاخص‌های توانایی فرایند ارائه شده برای فرایندهای تک متغیره‌ی شاخص C_p ، C_{pk} و C_{pm} هستند. کوتزو جانسون^[۱] مروری بر تحقیقات صورت گرفته بین سال‌های ۱۹۹۲ تا ۲۰۰۰ روی شاخص‌های توانایی فرایند انجام دادند. در دنیای واقعی در بسیاری از فرایندها بیش از یک مشخصه‌ی کیفی وجود دارد. بنابراین شاخص‌های توانایی فرایند چندمتغیره برای محاسبه‌ی توانایی فرایندها با بیش از یک مشخصه‌ی کیفی ارائه شد. تام و همکاران^[۲] شاخص‌های توانایی فرایند MC_p و MC_{pm} را برای فرایندهای چندمتغیره ارائه کردند. شهریار و همکاران^[۳] شاخص توانایی فرایند چندمتغیره‌ی C_{pm} را برای تعیین توانایی فرایندهای چندمتغیره

-- هنگامی که مشخصه‌های کیفی دارای توزیع نرمال چندمتغیره هستند -- ارائه کردند. شن^[۴] شاخصی براساس نسبت اقلام نامنطبق را برای محاسبه‌ی توانایی فرایندهای چندمتغیره پیشنهاد داد. ونگ و شن^[۵] شاخص توانایی فرایند چندمتغیره را با استفاده از تکنیک تحلیل مؤلفه‌های اصلی پیشنهاد دادند. شهریار و عبدالله‌زاده^[۶] نیز بر دار توانایی فرایند را برای تعیین توانایی فرایندهای چندمتغیره ارائه دادند. شاخص‌های توانایی فرایند ذکر شده برای محاسبه‌ی توانایی فرایند هنگامی که داده‌ها قطعی هستند ارائه شده است. این در حالی است که غالباً حدود مشخصه‌ی فنی و مشاهدات دارای عدم قطعیت هستند. عوامل خطای موجود در سیستم‌های اندازه‌گیری و تأثیر قضاوت عملگر در ثبت مشاهدات منجر به پدیده‌ی عدم قطعیت در ثبت مشاهدات می‌شود. در این حالت حدود مشخصه‌ی فنی و مشاهدات به صورت اعداد فازی بیان می‌شود. بنابراین شاخص‌های توانایی فرایند را باید به صورت مجموعه‌های فازی محاسبه کرد. تی‌سای و همکاران^[۷] شاخص توانایی فرایند فازی \tilde{C}_p را برای محاسبه‌ی توانایی فرایند هنگامی که مشاهدات فازی

* نویسنده مسئول

تاریخ دریافت: ۱۳۹۳/۹/۲۵، اصلاحیه ۱، ۱۳۹۴/۶/۱، پذیرش ۲۸، ۱۳۹۴/۶/۲۸.

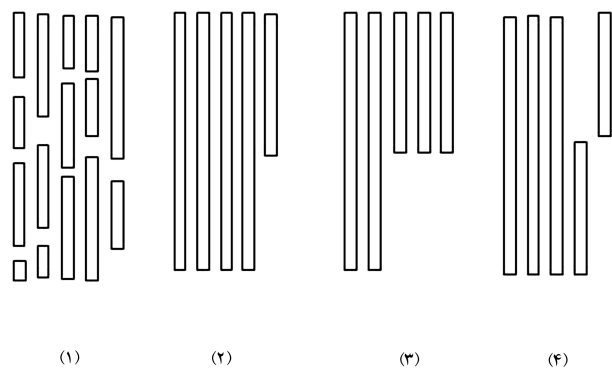
e.nikzad@shahed.ac.ir
atefeh.ashuri@gmail.com
amiri@shahed.ac.ir
a.salmasnia@qom.ac.ir

هستند ارائه کردند. ماشین چپی و پرجمی^[۸] از روش تخمین بالکی برای تخمین فازی شاخص‌های توانایی فرایند تک‌متغیره استفاده کردند. آنان شاخص توانایی فرایند فازی \tilde{C}_{pm} را، هنگامی که حدود مشخصه فنی با اعداد فازی بیان شده، پیشنهاد دادند.^[۹] پرجمی و همکاران^[۱۰] یک فاصله‌ی اطمینان برای شاخص توانایی فرایند فازی \tilde{C}_{pm} در حالتی که حدود مشخصه فنی فازی است، ارائه دادند. چن و همکاران^[۱۱] روشی برای محاسبه‌ی مقدار شاخص توانایی فرایند C_{pm} برای داده‌های فازی ارائه دادند. همچنین عبدالشاه و همکاران^[۱۲] شاخص توانایی فرایند فازی \tilde{C}_{pmk} را در شرایط عدم قطعیت فازی مشاهدات ارائه کردند. موسوی و همکاران^[۱۳] روشی برای محاسبه‌ی شاخص‌های توانایی فرایند چندمتغیره‌ی فازی ارائه دادند.

مسئله‌ی داده‌های از دست رفته به‌عنوان یکی از موضوعات بسیار مهم و کاربردی در تحقیقات سال‌های اخیر معرفی شده است.^[۱۴] این مسئله ممکن است به دلایلی همچون خرابی ماشین‌آلات، هزینه‌ی بالای نمونه‌گیری، خطا در سیستم‌های اندازه‌گیری، از بین رفتن داده‌های تاریخی یا ناخوانا بودن پاره‌یی از اطلاعات ثبت شده توسط اپراتور در بسیاری از کاربردهای عملی اتفاق بیفتد.^[۱۵، ۱۶] به‌عنوان مثال در علوم فتاوری نانو برخی از ترکیبات مولکولی به‌سادگی قابل استقرار در محیط به‌منظور انجام آزمایش نیستند و به‌عنوان داده‌های از دست رفته تلقی می‌شوند.^[۱۷] همچنین در کنترل و پایش فرایندهای دسته‌یی در هر لحظه مانند t ، رفتار آینده‌ی فرایند نامعلوم تلقی می‌شود و به‌عنوان داده‌ی از دست رفته منظور می‌شود.^[۱۸] خرابی در سنسورها و سیستم‌های دیجیتال ثبت‌کننده‌ی اطلاعات نیز در بسیاری از فرایندها منجر به ایجاد داده‌های از دست رفته می‌شود.^[۱۹]

داده‌های از دست رفته به‌دلیل تأثیر زیاد بر برآورد نادرست عملکرد فرایند به‌عنوان یک مشکل مطرح می‌شود. تأثیر داده‌های از دست رفته در تخمین پارامترهای فرایند در شرایطی که بر مبنای داده‌های موجود در خصوص قابلیت و توانایی فرایند تصمیم‌گیری می‌شود دوچندان است. تاکنون تحقیقات معدودی در حوزه‌ی کنترل فرایند آماری در خصوص داده‌های از دست رفته صورت پذیرفته است. محمود و همکاران^[۲۰] عملکرد روش‌های تخمین داده‌های از دست رفته در نمودارهای کنترل چندمتغیره در فاز یک را مورد بررسی قرار دادند. همچنین مبولی و همکاران^[۲۱] اثر داده‌های از دست رفته بر نمودار کنترل میانگین متحرک موزون نمایی را ارزیابی کردند. در تحقیقی دیگر عشوری و امیری^[۲۲] عملکرد روش‌های تخمین داده‌های از دست رفته را در شاخص توانایی فرایند چندمتغیره‌ی نام از طریق مدل خطی تعمیم‌یافته^۱ در جدول تحلیل واریانس ارزیابی کردند. چنان که در شکل ۱ نشان داده شده است، الگوی داده‌های از دست رفته به چهار حالت قابل تعریف است:^[۲۲]

۱. داده‌های از دست رفته در حالت کاملاً تصادفی و بدون ساختار^۲: متداول‌ترین



شکل ۱. الگوی داده‌های از دست رفته.^[۲۲]

حالت وقوع داده‌های از دست رفته است که در آن برداری از مشاهدات وجود دارد؛ برخی از درایه‌های این بردار با استفاده از تولید اعداد تصادفی به‌عنوان داده‌های از دست رفته در نظر گرفته می‌شوند.

۲. داده‌های از دست رفته‌ی تک‌متغیره^۳: در این حالت یک متغیر از بین متغیرهای موجود گم‌شده تلقی می‌شود. در واقع زمانی که بیش از یک مشخصه‌ی کیفی وجود دارد، تنها یکی از مشخصه‌های کیفی برحسب شرایط به‌صورت از دست رفته در نظر گرفته می‌شود.

۳. داده‌های از دست رفته به‌دلیل حضور برخی موانع^۴: در این حالت در برخی از فرایندها از جمله فرایندهای شیمیایی در دمای بالا و شرایط ایزوله، اندازه‌گیری برخی از مشخصه‌های کیفی به‌علت شرایط هزینه‌زا و مجاورت با مناطق پرخطر منطقی به نظر نمی‌رسد و به‌عنوان داده‌ی گم‌شده تلقی می‌شود.

۴. داده‌های از دست رفته به‌علت عدم تطبیق برخی مشاهدات^۵: برخی از مشاهدات قابلیت دسترسی همزمان را ندارند و همواره قسمتی از مشاهدات از دست رفته تلقی می‌شود. در واقع سنسورهایی که به‌منظور ثبت مشاهدات تعبیه شده، در دوره‌های زمانی متفاوتی عمل ثبت مشاهدات را انجام می‌دهند.

با توجه به وجود عدم قطعیت در پایش و ثبت مشخصه‌های کیفی در بسیاری از فرایندها و وجود داده‌های از دست رفته در بسیاری از کاربردهای عملی، ضرورت بررسی عملکرد روش‌های تخمین داده‌های از دست رفته فازی و اثر وجود داده‌های از دست رفته بر عملکرد شاخص‌های توانایی فرایند در فضای عدم قطعیت مشخص است. چنان که در مرور ادبیات اشاره شد، تاکنون اثر وجود داده‌های از دست رفته فازی و روش‌های تخمین داده‌های از دست رفته چندمتغیره‌ی فازی بر عملکرد شاخص‌های توانایی فرایند چندمتغیره‌ی فازی بررسی نشده است. بدین‌منظور در نوشتار حاضر ابتدا دو روش رگرسیون و میانگین فازی به‌منظور تخمین داده‌های از دست رفته‌ی فازی توسعه داده شده است. سپس نتایج شاخص توانایی فرایند میانگین و رگرسیون برآورد می‌شوند، بررسی می‌شود. در این نوشتار برای تخمین مقادیر مشاهدات فازی با روش رگرسیون و همچنین محاسبه‌ی شاخص توانایی فرایند دومتغیره‌ی فازی C_{pm} از توابع بهینه‌سازی استفاده می‌شود. برای حل توابع بهینه‌سازی از الگوریتم اجتماع ذرات (PSO)^۶ استفاده شده است. الگوریتم بهینه‌سازی اجتماع ذرات به‌عنوان یک الگوریتم فراابتکاری مبتنی بر جمعیت در سال ۱۹۹۵ توسط کندی و ابرهارت^[۲۳] ارائه شد. این الگوریتم بر پایه‌ی رفتار اجتماعی گروهی از پرندگان یا حشرات، یا دسته‌یی از ماهیان که به شیوه‌یی بسیار معمولی به جست‌وجوی غذا می‌پردازند، به وجود آمده است. الگوریتم PSO یک تکنیک جست‌وجوی تصادفی با محاسباتی مؤثر، اجرای آسان و نیاز به حافظه‌ی کم‌تر در مقایسه با سایر الگوریتم‌های تکاملی است.

در این تحقیق تحلیل حساسیت روی پارامترهای ضریب همبستگی و درصد داده‌های از دست رفته نیز انجام می‌شود. هدف از تحلیل حساسیت یافتن شرایطی است که شاخص توانایی فرایند کم‌ترین تأثیرپذیری را در حضور داده‌های از دست رفته داشته باشد.

در این نوشتار، در بخش دوم به معرفی شاخص توانایی فرایند C_{pm} ، و سپس محاسبه‌ی شاخص توانایی فرایند فازی \tilde{C}_{pm} هنگامی که مشاهدات فازی هستند پرداخته شده است. در بخش سوم روش‌های تخمین داده‌های از دست رفته در حالت قطعی و فازی توضیح داده شده است. در بخش چهارم رویه‌ی تحقیق و الگوریتم اجتماع ذرات ارائه شده است. عملکرد شاخص توانایی فرایند فازی \tilde{C}_{pm} زمانی که

$$L\hat{P}L_i = \tilde{X}_i - \sqrt{\frac{\chi_{\alpha,k}^2 \cdot \det(\mathbf{S}_i^{-1})}{\det(\mathbf{S}^{-1})}} \quad (7)$$

در شاخص توانایی فرایند C_{pm} به دلیل آن که رابطه‌ی بین شاخص توانایی فرایند و مشاهدات فازی غیرخطی است، بنابراین استفاده از روش‌های معمول برای تعیین تابع درجه‌ی عضویت شاخص توانایی فرایند امکان‌پذیر نیست. در این مقاله تابع درجه‌ی عضویت شاخص توانایی فرایند C_{pm} براساس مفهوم ارائه شده توسط کاو و لیو^[۲۴] محاسبه شده است. برای تعیین تابع درجه‌ی عضویت تخمینی شاخص توانایی فرایند از روش ارائه شده توسط موسوی و همکاران^[۱۳] و در سه گام استفاده می‌شود. در گام اول α - برش‌های مشاهدات براساس رابطه‌ی ۸ تعیین می‌شود. در گام دوم حدود بالا و پایین α - برش‌های شاخص توانایی فرایند با استفاده از توابع بهینه‌سازی محاسبه می‌شود. برای هر سطح α حد پایین شاخص توانایی فرایند با استفاده از کمیته‌سازی این شاخص، هنگامی که مشاهدات بین حد بالا و پایین خود در سطح α است، محاسبه می‌شود. همچنین حد بالای شاخص توانایی فرایند در سطح α با پیشینه‌سازی شاخص توانایی فرایند هنگامی که مشاهدات بین حدود بالا و پایین خود در سطح α هستند تعیین می‌شود. در گام سوم توابع درجه‌ی عضویت شاخص‌های توانایی فرایند براساس α - برش‌ها تعیین می‌شود. حدود پایین و بالای α - برش‌های شاخص توانایی فرایند C_{pm} به ترتیب با استفاده از روابط ۹ و ۱۰ محاسبه می‌شود. همچنین تابع درجه‌ی عضویت شاخص توانایی فرایند C_{pm} به صورت رابطه‌ی ۱۱ تعریف می‌شود:

$$(X_{ij})_{\alpha} = \left[(X_{ij})_{\alpha}^L, (X_{ij})_{\alpha}^U \right] = \min_{x_{ij} \in \mathbf{x}} \left\{ x_{ij} \in \mathbf{x} \mid \mu_{\tilde{X}_{ij}}(x_{ij}) \geq \alpha \right\}, \quad \max_{x_{ij} \in \mathbf{x}} \left\{ x_{ij} \in \mathbf{x} \mid \mu_{\tilde{X}_{ij}}(x_{ij}) \geq \alpha \right\} \quad (8)$$

$$(\hat{C}_{pm})_{\alpha}^L = \min.$$

$$\left[\frac{\prod_{i=1}^k (USL_i - LSL_i)}{\prod_{i=1}^k \left(\left(\tilde{X}_i + \sqrt{\frac{\chi_{k,\alpha}^2 \cdot \det(\mathbf{S}_i^{-1})}{\det(\mathbf{S}^{-1})}} \right) - \left(\tilde{X}_i - \sqrt{\frac{\chi_{k,\alpha}^2 \cdot \det(\mathbf{S}_i^{-1})}{\det(\mathbf{S}^{-1})}} \right) \right)} \right]^{\frac{1}{k}} \quad (9)$$

$$(X_{ij})_{\alpha}^L \leq X_{ij} \leq (X_{ij})_{\alpha}^U$$

$$(\hat{C}_{pm})_{\alpha}^U = \max.$$

$$\left[\frac{\prod_{i=1}^k (USL_i - LSL_i)}{\prod_{i=1}^k \left(\left(\tilde{X}_i + \sqrt{\frac{\chi_{k,\alpha}^2 \cdot \det(\mathbf{S}_i^{-1})}{\det(\mathbf{S}^{-1})}} \right) - \left(\tilde{X}_i - \sqrt{\frac{\chi_{k,\alpha}^2 \cdot \det(\mathbf{S}_i^{-1})}{\det(\mathbf{S}^{-1})}} \right) \right)} \right]^{\frac{1}{k}} \quad (10)$$

$$(X_{ij})_{\alpha}^L \leq X_{ij} \leq (X_{ij})_{\alpha}^U$$

$$\mu_{\hat{C}_{pm}}(\hat{C}_{pm}) = \begin{cases} L(\hat{C}_{pm}), & (\hat{C}_{pm})^L \leq \hat{C}_{pm} \leq (\hat{C}_{pm})^L \\ 1, & (\hat{C}_{pm})^L \leq \hat{C}_{pm} \leq (\hat{C}_{pm})^U \\ R(\hat{C}_{pm}), & (\hat{C}_{pm})^U \leq \hat{C}_{pm} \leq (\hat{C}_{pm})^U \end{cases} \quad (11)$$

$L(\hat{C}_{pm})$ و $R(\hat{C}_{pm})$ به ترتیب توابع چپ و راست تابع درجه‌ی عضویت شاخص توانایی فرایند $\mu_{\hat{C}_{pm}}$ هستند.

داده‌های از دست رفته با روش‌های میانگین و رگرسیون تخمین زده می‌شوند به همراه تحلیل حساسیت روی مقادیر مختلف ضریب همبستگی بین مشخصه‌های کیفی و درصد داده‌های از دست رفته، با استفاده از مثال عددی در بخش پنجم نشان داده شده است. همچنین در پایان نتیجه‌گیری و پیشنهادات برای مطالعات آتی ارائه شده است.

۲. شاخص توانایی فرایند چندمتغیره

یکی از شاخص‌های توانایی فرایند چندمتغیره، شاخص ارائه شده توسط شهریاری و همکاران^[۱] است. شاخص C_{pm} از نسب ناحیه‌ی تولرانس و ناحیه‌ی فرایند تبدیل یافته محاسبه می‌شود:

$$C_{pm} = \left[\frac{\prod_{i=1}^k (USL_i - LSL_i)}{\prod_{i=1}^k (UPL_i - LPL_i)} \right]^{\frac{1}{k}} \quad (1)$$

در این رابطه UPL_i و LPL_i به ترتیب حدود پایین و بالای فرایند هستند که با استفاده از روابط ۲ و ۳ محاسبه می‌شود:

$$LPL_i = U_i - \sqrt{\frac{\chi_{\alpha,k}^2 \cdot \det(\mathbf{\Sigma}_i^{-1})}{\det(\mathbf{\Sigma}^{-1})}} \quad (2)$$

$$UPL_i = U_i + \sqrt{\frac{\chi_{\alpha,k}^2 \cdot \det(\mathbf{\Sigma}_i^{-1})}{\det(\mathbf{\Sigma}^{-1})}} \quad (3)$$

در این روابط U_i میانگین مشخصه‌ی کیفی شماره i و $\mathbf{\Sigma}$ ماتریسی برگرفته از ماتریس واریانس - کوواریانس است که در آن سطر و ستون شماره‌ی i حذف شده است.

۱.۲. شاخص توانایی فرایند چندمتغیره‌ی فازی

در دنیای واقعی همواره مقادیر مشخصه‌های کیفی اعداد قطعی نیستند. هنگامی که مشاهدات برگرفته از یک فرایند غیردقیق است یا مقدار مشاهدات به صورت داده‌های زبانی بیان می‌شود، از اعداد فازی برای بیان مشاهدات استفاده می‌شود که معمولاً در ادبیات موضوع به منظور سادگی به صورت اعداد مثلثی در نظر گرفته می‌شوند. در این مقاله نیز هم‌راستا با ادبیات موضوع از اعداد فازی مثلثی استفاده می‌شود. اگرچه می‌توان روش‌های ارائه شده در این مقاله را به راحتی به اعداد فازی دوزنقه‌یی تعمیم داد. اگر فرض شود \tilde{X}_{ij} یک عدد فازی مثلثی است که مقدار مشاهده‌ی Z_{ij} مشخصه کیفی i ام را نشان می‌دهد، می‌توان آن را به صورت رابطه‌ی ۴ نشان داد:

$$\tilde{X}_{ij} = (X_{ij}^l, X_{ij}^m, X_{ij}^u) \quad (4)$$

شاخص توانایی فرایند C_{pm} برای داده‌های فازی چندمتغیره از رابطه‌ی ۵ محاسبه می‌شود.

$$\hat{C}_{pm} = \left[\frac{\prod_{i=1}^k (USL_i - LSL_i)}{\prod_{i=1}^k (U\hat{P}L_i - L\hat{P}L_i)} \right]^{\frac{1}{k}} \quad (5)$$

مقادیر $L\hat{P}L_i$ و $U\hat{P}L_i$ به ترتیب با استفاده از روابط ۶ و ۷ محاسبه می‌شود:

$$U\hat{P}L_i = \tilde{X}_i + \sqrt{\frac{\chi_{\alpha,k}^2 \cdot \det(\mathbf{S}_i^{-1})}{\det(\mathbf{S}^{-1})}} \quad (6)$$

۳. روش‌های تخمین داده‌های از دست رفته در حالت قطعی

در این بخش دو روش میانگین و رگرسیون به منظور تخمین داده‌های از دست رفته ارائه شده است. هدف از ارائه‌ی این روش‌ها انتخاب بهترین روش به منظور تخمین داده‌های از دست رفته است. اختلاف بین شاخص توانایی فرایند تخمین زده شده در حضور داده‌های از دست رفته، و شاخص توانایی فرایند در حالت عدم حضور داده‌های از دست رفته به عنوان معیاری برای بررسی قابل قبول‌ترین روش در تخمین داده‌های از دست رفته محسوب می‌شود.

۱.۳. روش میانگین

در این روش میانگین مشاهدات موجود به عنوان جایگزینی برای داده‌های از دست رفته تلقی می‌شود. بدین ترتیب اگر مشاهده‌ی X_{ij} گم شده تلقی شود، میانگین مشاهدات موجود در مشخصه‌ی کیفی شماره i (\bar{X}_i) به عنوان برآوردی از داده‌ی از دست رفته استفاده می‌شود.

۲.۳. روش رگرسیون

این روش از معادله‌ی رگرسیون به منظور تخمین داده‌های از دست رفته استفاده می‌کند. اگر (X_1, X_2) برداری از مشاهدات باشند، در آنها X_2 به عنوان متغیر پاسخ و X_1 به عنوان متغیر مستقل در نظر گرفته می‌شود. فرض می‌شود همواره یک مقدار از بردار مشاهدات (X_1, X_2) از دست رفته باشد. در این روش شیب خط رگرسیون ($\hat{\beta}$) با استفاده از رابطه‌ی ۱۲ تخمین زده می‌شود:

$$\hat{\beta} = \frac{Sx_1x_2}{Sx_1x_1} \quad (12)$$

که در آن مقادیر Sx_1x_1 و Sx_1x_2 به ترتیب از طریق روابط ۱۳ و ۱۴ قابل محاسبه است:

$$Sx_1x_2 = \sum_j (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2) \quad (13)$$

$$Sx_1x_1 = \sum_j (X_{1j} - \bar{X}_1)^2 \quad (14)$$

عرض از مبدأ خط رگرسیون ($\hat{\alpha}$) با استفاده از رابطه‌ی ۱۵ برآورد می‌شود:

$$\hat{\alpha} = \bar{X}_2 - \hat{\beta} \bar{X}_1 \quad (15)$$

اگر X_{2j} مشاهده از دست رفته فرض شود، با استفاده از رابطه‌ی ۱۶ قابل تخمین است:

$$\hat{X}_{2j} = \hat{\alpha} + \hat{\beta} X_{1j} \quad (16)$$

۳.۳. روش‌های پیشنهادی تخمین داده‌های از دست رفته‌ی فازی

مقادیر فازی مشاهدات از دست رفته با استفاده از روش میانگین و رگرسیون به صورت فازی تخمین زده می‌شود. هنگامی که مشاهدات اعداد فازی مثلثی‌اند مقدار برآورد شده‌ی داده‌های از دست رفته از رابطه‌ی ۱۷ محاسبه می‌شود:

$$\hat{\tilde{X}} = \left(\frac{\sum_{j=1}^n X_{ij}^l}{n}, \frac{\sum_{j=1}^n X_{ij}^m}{n}, \frac{\sum_{j=1}^n X_{ij}^u}{n} \right) \quad (17)$$

در روش رگرسیون، تابع درجه‌ی عضویت مقدار تخمینی داده‌ی از دست رفته با استفاده از روش ارائه شده برای تعیین تابع درجه‌ی عضویت شاخص‌های توانایی فرایند محاسبه می‌شود. در این روش حدود پایین و بالای α - برش‌های تخمین داده‌ی از دست رفته به ترتیب با استفاده از روابط ۱۸ و ۱۹ تعیین می‌شود:

$$\begin{aligned} (\hat{X}_{1k})^L &= \min . \hat{\alpha} + \hat{\beta} X_{2k} \\ \hat{\beta} &= \frac{\sum_j (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2)}{\sum_j (X_{2j} - \bar{X}_2)^2} \end{aligned}$$

$$\hat{\alpha} = \bar{X}_1 - \hat{\beta} \bar{X}_2 \quad (18)$$

$$\begin{aligned} (\hat{X}_{1k})^U &= \max . \hat{\alpha} + \hat{\beta} X_{2k} \\ \hat{\beta} &= \frac{\sum_j (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2)}{\sum_j (X_{2j} - \bar{X}_2)^2} \end{aligned}$$

$$\hat{\alpha} = \bar{X}_1 - \hat{\beta} \bar{X}_2 \quad (19)$$

α - برش‌های مقدار تخمینی داده‌ی از دست رفته $[(X_{1k})^L, (X_{1k})^U]$ است. روابط ۱۸ و ۱۹ هنگامی کاربرد دارند که مشخصه‌ی کیفی اول داده از دست رفته باشد.

۴. رویه‌ی تحقیق و روش‌های پیشنهادی تخمین داده‌های از دست رفته‌ی فازی

در این بخش ابتدا رویه‌ی تحقیق به منظور بررسی روش‌های پیشنهادی در تخمین داده‌های از دست رفته فازی در شاخص توانایی فرایند دومتغیره‌ی فازی C_{pm} تشریح شده است. سپس در ادامه توسعه‌ی دو روش میانگین و رگرسیون به منظور تخمین داده‌های از دست رفته با رویکرد فازی ارائه شده است.

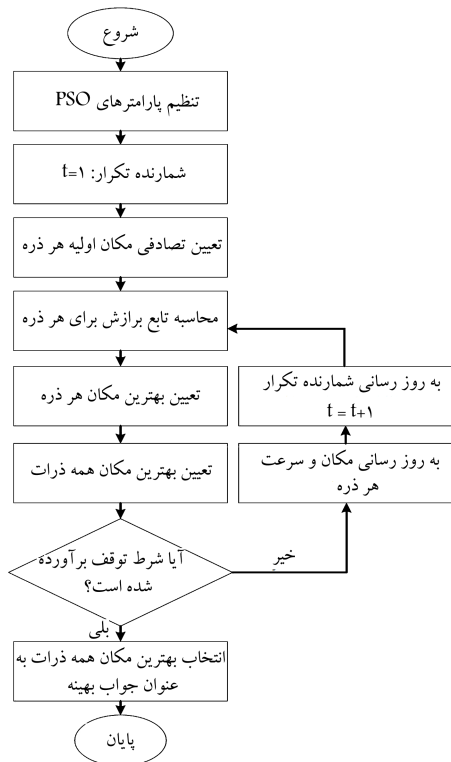
۱.۴. رویه‌ی تحقیق

در این بخش گام‌های تحقیق به منظور بررسی روش‌های پیشنهادی در تخمین مشاهدات از دست رفته و تحلیل پاسخ شاخص توانایی فرایند در حضور داده‌های از دست رفته با در نظر گرفتن ضرایب همبستگی و درصد داده‌های از دست رفته در شکل ۲ ارائه شده است.

هدف از رویه‌ی ارائه شده ارزیابی بهترین شرایطی است که در آن تأثیرپذیری شاخص توانایی فرایند از حضور داده‌های از دست رفته به کم‌ترین حالت خود رسیده (حالت ایده‌آل) و مقدار برآورده شده از توانایی فرایند یک ارزیابی صحیح باشد و در نتیجه ریسک برآورده‌های غلط از توانایی فرایند به علت حضور داده‌های از دست رفته کمیته شود.

۲.۴. الگوریتم اجتماع ذرات (POS)

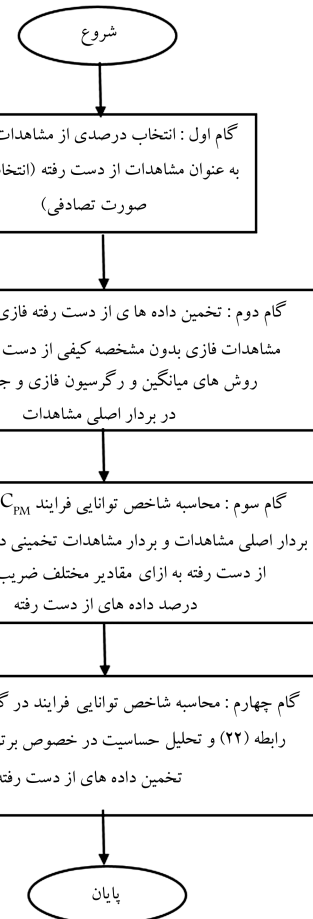
در الگوریتم اجتماع ذرات جهت حرکت ذره (بردار سرعت) در هر تکرار متأثر از سه عامل است: ۱. بهترین خاطره‌ی شخصی (p_i^k) یعنی بهترین مکانی که ذره‌ی i ام



شکل ۳. فرایند الگوریتم بهینه‌سازی اجتماع ذرات.

۵. مثال عددی

بعد از بیان رویه‌ی تحقیق و توسعه‌ی روش‌های تخمین داده‌های از دست رفته در محیط فازی، در این بخش مثال عددی برای بررسی عملکرد روش‌های تخمین داده‌های از دست رفته فازی بر عملکرد شاخص توانایی فرایند دومتغیره فازی ارائه شده است. فرض می‌شود مقادیر مشخصه‌های کیفی اعداد فازی مثلثی برگرفته از یک توزیع نرمال دومتغیره با بردار میانگین و ماتریس واریانس کوواریانس فازی هستند. به منظور ارزیابی اثر ضریب همبستگی در این مقاله فرض می‌شود که ماتریس واریانس - کوواریانس با ماتریس همبستگی برابر است (واریانس مشخصه‌های کیفی برابر ۱ در نظر گرفته شده است).^[۲۰] فرض می‌شود که نمونه‌ی با بیست مشاهده از فرایند نرمال دومتغیره با بردار میانگین [۵ ۸] و ماتریس واریانس کوواریانس فازی مثلثی به تعداد مشاهدات، داده‌ی دارای توزیع یکنواخت در بازه [۰٫۲ ۰٫۷۵] تولید می‌شود و به مشاهدات دارای توزیع نرمال دومتغیره اضافه و از آنها کم می‌شود. هنگامی که درصد داده‌های از دست رفته ۵٪، ۱۵٪ و ۲۵٪ باشند مشاهدات از دست رفته با استفاده از دو روش میانگین و رگرسیون تخمین زده می‌شود و مقادیر به دست آمده جایگزین داده‌های از دست رفته شده و مقادیر شاخص توانایی فرایند محاسبه می‌شود. حدود مشخصه‌ی فنی پایین و بالا برای مشخصه‌های کیفی اول و دوم به ترتیب برابر با [۲٫۳ ۹٫۹] و [۴٫۲ ۱۱٫۷] است. تعداد دفعات شبیه‌سازی ۱۰۰۰ بار است. مقادیر شاخص توانایی فرایند با در نظر گرفتن مقادیر مختلفی از ضریب همبستگی و درصد داده‌های از دست رفته محاسبه شده است. مقدار میانگین شاخص توانایی فرایند فازی C_{pm} در حالت عدم وجود داده‌های از دست رفته به ازای مقادیر مختلف ضریب همبستگی در جدول ۲ نشان داده شده است.



شکل ۲. رویه‌ی تحقیق.

جدول ۱. پارامترهای الگوریتم اجتماع ذرات.

پارامتر	c_1	c_2	w	k	جمعیت ذرات
مقدار	۲	۲	۱	۳۰۰	۳۰

تا تکرار k ام پیدا کرده است؛ ۲. بهترین خاطره‌ی جمعی (p_g^k) یعنی بهترین مکانی که تمام ذرات تا تکرار k ام پیدا کرده‌اند؛ ۳. اینرسی (v_i^k) یعنی تمایل به حرکت در جهت تکرار قبلی. بنابراین، بردارهای سرعت و موقعیت ذره‌ی i ام در تکرار $k+1$ ام به ترتیب با استفاده از روابط ۲۰ و ۲۱ به روزرسانی می‌شود:

$$v_i^{k+1} = wv_i^k + c_1r_1(p_i^k - x_i^k) + c_2r_2(p_g^k - x_i^k) \quad (20)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (21)$$

در روابط ۲۰ و ۲۱، متغیرهای w ، c_1 و c_2 به ترتیب معرف ضریب اینرسی، ضریب یادگیری شخصی و ضریب یادگیری جمعی با مقادیر مثبت هستند. همچنین r_1 و r_2 معرف مقادیر تصادفی با توزیع یکنواخت در بازه [۰، ۱] هستند که مقدار آن‌ها در هر تکرار متفاوت خواهد بود.

مقادیر پارامترها در این مقاله در جدول ۱ نشان داده شده است. فرایند الگوریتم بهینه‌سازی اجتماع ذرات نیز در شکل ۳ نشان داده شده است.

جدول ۲. میانگین شاخص توانایی فرایند فازی \hat{C}_{pm} در حالت عدم وجود داده‌های از دست رفته.

شاخص	ρ
$(1, 0.057, 1, 1422, 1, 3093)$	۰٫۲۵
$(1, 0.075, 1, 1447, 1, 3128)$	۰٫۵
$(1, 0.097, 1, 1478, 1, 3172)$	۰٫۷۵

Paired T for mean method - regression method				
	N	Mean	StDev	SE Mean
mean method	9	0.26363	0.02731	0.00910
regression method	9	0.25859	0.02338	0.00779
Difference	9	0.00504	0.00642	0.00214

95% CI for mean difference: (0.00011, 0.00998)
 T-Test of mean difference = 0 (vs not = 0): T-Value = 2.36 P-Value = 0.046

شکل ۴. آزمون t زوجی برای مقایسه روش‌های تخمین داده‌های از دست رفته.

از دست رفته بهتر است از روش رگرسیون استفاده شود (شکل ۴). در این قسمت برای بررسی اثر نمونه بر عملکرد روش‌های میانگین و رگرسیون مقادیر شاخص توانایی فرایند C_{pm} برای اندازه نمونه ۴۰، ۶۰ و ۱۰۰، هنگامی که ضریب همبستگی برابر ۰٫۲۵ و درصد داده‌های از دست رفته برابر ۱۵٪ باشد، محاسبه و در جدول ۵ نشان داده شده است.

بر اساس نتایج حاصل از جدول ۵، با افزایش اندازه نمونه روش‌های تخمین داده‌های از دست رفته عملکرد بهتری دارند. همچنین روش رگرسیون هنگامی که اندازه نمونه افزایش می‌یابد از عملکرد بهتری نسبت به روش میانگین برخوردار است. برای بررسی اثر میزان فازی بودن مشاهدات بر عملکرد روش‌های تخمین داده‌های از دست رفته اعداد تصادفی از توزیع یکنواخت $[0.25, 0.35]$ به تعداد مشاهدات تولید می‌شوند و با اعداد دارای توزیع نرمال دومتغیره‌ی تولید شده جمع و از آنها کم می‌شود، که این امر سبب تولید اعداد فازی بزرگ‌تر نسبت به مثال قبل می‌شود. مقدار میانگین شاخص توانایی فرایند فازی C_{pm} در حالت عدم وجود داده‌های از دست رفته در این حالت برابر با $(0.8957, 1.1422, 1.5185)$ است. مقادیر میانگین شاخص توانایی فرایند C_{pm} هنگامی که داده‌های از دست رفته به روش میانگین و رگرسیون برآورد می‌شود و میانگین فاصله تا مقدار شاخص در حالت عدم وجود داده از دست رفته برای درصد داده از دست رفته ۱۵٪ و ضریب همبستگی ۰٫۲۵ در جدول ۶ نشان داده شده است.

بر اساس نتایج حاصل از جدول ۶ با افزایش میزان فازی بودن مشاهدات کارایی روش‌های برآورد داده‌های از دست رفته کاهش می‌یابد. به عبارت دیگر فاصله‌ی بین برآورد شاخص توانایی فرایند در حالت وجود و عدم وجود داده‌های از دست رفته افزایش می‌یابد.

۶. نتیجه‌گیری و پیشنهادات آتی

در این نوشتار شاخص توانایی فرایند (C_{pm}) به عنوان معیاری برای سنجش توانایی فرایند در تأمین خواسته‌های مشتریان، در محیط عدم قطعیت فازی و همچنین در حضور داده‌های از دست رفته، مورد ارزیابی قرار گرفت. به منظور تخمین داده‌های از دست رفته‌ی فازی از روش‌های میانگین و رگرسیون فازی استفاده شد. همچنین اثر تخمین داده‌های از دست رفته فازی بر عملکرد شاخص توانایی فرایند (C_{pm}) با در نظر گرفتن مقادیر مختلف درصد داده‌های از دست رفته و ضرایب همبستگی مورد بررسی قرار گرفت. نتایج حاصل از شبیه‌سازی نشان داد که روش رگرسیون نسبت به روش میانگین در تخمین داده‌های از دست رفته، عملکرد بهتری دارد. همچنین با افزایش اندازه نمونه، کارایی روش‌های میانگین و رگرسیون در برآورد داده‌های از دست رفته افزایش می‌یابد. بنابراین فاصله‌ی بین شاخص توانایی فرایند محاسبه شده در حالت وجود داده‌های کامل و وجود داده‌ی از دست رفته کاهش می‌یابد. همچنین با افزایش درجه‌ی فازی بودن مشاهدات، کارایی روش‌های تخمین داده‌های از دست رفته کاهش می‌یابد.

بر اساس نتایج جدول ۲ مقادیر شاخص توانایی فرایند \hat{C}_{pm} با افزایش مقدار ضریب همبستگی افزایش می‌یابد. مقادیر میانگین شاخص توانایی فرایند فازی C_{pm} در حالتی که به ازای برخی از مقادیر مشخصه‌ی کیفی دوم، مشخصه‌ی کیفی اول وابسته به آن موجود نیست، در جدول ۳ نشان داده شده است. در این جدول مقادیر از دست رفته با روش میانگین به ازای مقادیر مختلف برای درصد داده‌های از دست رفته تخمین زده می‌شود. فاصله‌ی متریک بین تخمین شاخص‌های توانایی فرایند در حالت وجود داده‌های از دست رفته و عدم وجود داده‌های از دست رفته از رابطه‌ی ۲۲ تعیین می‌شود: [۲۵]

$$D(\tilde{A}, \tilde{B}) = \left[\int_0^1 (L^{-1}(A) - L^{-1}(B))^2 + \int_0^1 (R^{-1}(A) - R^{-1}(B))^2 \right]^{1/2} \quad (22)$$

که در آن $\tilde{A}(x) = (a, b, c)$ و $\tilde{B}(x) = (d, e, f)$ اعداد فازی مثلثی‌اند. $L^{-1}(A)(L^{-1}(B))$ و $R^{-1}(A)(R^{-1}(B))$ توابع معکوس چپ و راست توابع درجه‌ی عضویت اعداد فازی $\tilde{A}(x)$ و $\tilde{B}(x)$ هستند. $L^{-1}(A)$ و $R^{-1}(A)$ مطابق روابط ۲۳ و ۲۴ تعریف می‌شوند.

$$L^{-1}(A) = a + (b - a)y \quad (23)$$

$$R^{-1}(A) = c + (b - c)y \quad (24)$$

فواصل بین مقدار شاخص توانایی فرایند در حالت عدم وجود داده‌های از دست رفته، و وجود داده‌های از دست رفته با استفاده از رابطه‌ی ۲۲ محاسبه شده و میانگین، کمینه و بیشینه‌ی این فواصل در جدول ۳ نشان داده شده است.

بر اساس نتایج نشان داده شده در جدول ۳ با افزایش درصد داده‌های از دست رفته، فاصله‌ی بین مقادیر شاخص توانایی فرایند در حالت وجود و عدم وجود داده‌های از دست رفته افزایش می‌یابد. بنابراین با افزایش درصد داده‌های از دست رفته کارایی روش میانگین در تخمین داده‌های از دست رفته کاهش می‌یابد.

مقادیر شاخص توانایی فرایند فازی (\hat{C}_{pm})، میانگین، کمینه و بیشینه فاصله‌ی بین شاخص توانایی فرایند در حالت وجود و عدم وجود داده‌های از دست رفته -- به ازای مقادیر مختلف درصد داده‌های از دست رفته، هنگامی که داده‌های از دست رفته با روش رگرسیون تخمین زده می‌شود -- در جدول ۴ نشان داده شده است.

برای مقایسه‌ی روش‌های تخمین داده‌های از دست رفته بر اساس نتایج میانگین فاصله، از آزمون t زوجی استفاده می‌شود زیرا مقادیر ضریب همبستگی و درصد داده‌های از دست رفته باعث وابستگی مقادیر میانگین فاصله حاصل از دو روش می‌شوند. ارزش P^۱ برای این آزمون برابر ۰٫۴۶ در سطح $\alpha = 0.05$ است، بنابراین اختلاف میان دو روش میانگین و رگرسیون معنادار است؛ برای برآورد داده‌های

جدول ۳. شاخص توانایی فرایند فازی \hat{C}_{pm} در صورت وجود داده‌های از دست رفته و تخمین داده‌های از دست رفته با روش میانگین.

فاصله‌ی بین شاخص توانایی فرایند در حالت وجود داده‌های از دست رفته و عدم وجود داده‌های از دست رفته	میانگین مقادیر شاخص توانایی فرایند			درصد داده‌های از دست رفته	ρ
	حداقل	میانگین	حداکثر		
	۱,۱۹۵۵	۰,۲۲۸۰	۰,۰۰۰۹		
۱,۱۹۸۵	۰,۲۳۴۰	۰,۰۰۱۰	(۱,۰۳۵۶, ۱,۱۷۶۶, ۱,۳۴۸۳)	۱۵	
۱,۲۹۵۸	۰,۲۵۹۴	۰,۰۰۱۹	(۱,۰۶۵۷, ۱,۲۱۵۱, ۱,۳۹۷۷)	۲۵	
۱,۳۳۵۳	۰,۲۴۵۲	۰,۰۰۰۸	(۱,۰۱۸۶, ۱,۱۵۸۱, ۱,۳۲۹۲)	۵	۰,۵
۱,۳۸۲۵	۰,۲۵۳۱	۰,۰۰۲۷	(۱,۰۳۷۷, ۱,۱۷۹۶, ۱,۳۵۲۶)	۱۵	
۱,۶۱۰۹	۰,۲۸۰۲	۰,۰۰۲۶	(۱,۰۶۷۸, ۱,۲۱۸۰, ۱,۴۰۱۸)	۲۵	
۱,۴۰۸۲	۰,۲۷۴۲	۰,۰۰۰۸	(۱,۰۲۰۵, ۱,۱۶۱۰, ۱,۳۳۳۴)	۵	۰,۷۵
۱,۴۵۰۴	۰,۲۸۴۸	۰,۰۰۲۲	(۱,۰۴۰۶, ۱,۱۸۳۹, ۱,۳۵۹۴)	۱۵	
۲,۰۱۷۷	۰,۳۱۳۸	۰,۰۰۱۷	(۱,۰۷۰۲, ۱,۲۲۱۵, ۱,۴۰۱۸)	۲۵	

جدول ۴. شاخص توانایی فرایند فازی \hat{C}_{pm} در صورت وجود داده‌های از دست رفته و تخمین داده‌های از دست رفته با روش رگرسیون.

فاصله‌ی بین شاخص توانایی فرایند در حالت وجود داده‌های از دست رفته و عدم وجود داده‌های از دست رفته	میانگین مقادیر شاخص توانایی فرایند			درصد داده‌های از دست رفته	ρ
	حداقل	میانگین	حداکثر		
	۱,۱۹۸۶	۰,۲۲۷۹	۰,۰۰۰۸		
۱,۱۷۰۱	۰,۲۳۰۶	۰,۰۰۱۹	(۱,۰۳۵۶, ۱,۱۶۵۸, ۱,۳۴۴۳)	۱۵	
۱,۲۶۲۴	۰,۲۵۱۱	۰,۰۰۴۴	(۱,۰۳۳۰, ۱,۲۰۰۵, ۱,۳۹۳۴)	۲۵	
۱,۳۳۵۳	۰,۲۴۵۲	۰,۰۰۰۸	(۱,۰۰۸۹, ۱,۱۵۳۸, ۱,۳۲۸۰)	۵	۰,۵
۱,۳۴۱۵	۰,۲۴۷۷	۰,۰۰۴۴	(۱,۰۰۳۶, ۱,۱۶۳۶, ۱,۳۴۴۴)	۱۵	
۱,۶۱۰۹	۰,۲۸۰۲	۰,۰۰۲۶	(۱,۰۱۴۸, ۱,۱۸۹۰, ۱,۳۹۰۱)	۲۵	
۱,۴۰۶۶	۰,۲۷۳۸	۰,۰۰۰۹	(۱,۰۱۵۳, ۱,۱۵۱۹, ۱,۳۲۹۸)	۵	۰,۷۵
۱,۲۷۰۵	۰,۲۷۶۵	۰,۰۰۴۱	(۰,۹۸۸۰, ۱,۱۵۰۲, ۱,۳۴۱۶)	۱۵	
۱,۶۹۰۵	۰,۲۹۴۳	۰,۰۰۵۱	(۰,۹۸۹۱, ۱,۱۶۸۷, ۱,۳۷۹۲)	۲۵	

جدول ۵. بررسی اثر اندازه نمونه بر تخمین شاخص توانایی فرایند \hat{C}_{pm} در حالت وجود داده‌های از دست رفته و تخمین داده از دست رفته به روش میانگین.

روش	n	داده‌های کامل	وجود داده از دست رفته	فاصله
میانگین	۲۰	(۱,۰۰۵۷, ۱,۱۴۲۲, ۱,۳۰۹۳)	(۱,۰۳۵۶, ۱,۱۷۶۶, ۱,۳۴۸۳)	۰,۲۳۴۰
	۴۰	(۱,۹۸۸۴, ۱,۱۱۵۶, ۰,۲۶۸۷)	(۱,۰۲۲۷, ۱,۱۵۵۵, ۱,۳۱۴۴)	۰,۱۶۹۸
	۶۰	(۱,۹۸۶۲, ۱,۱۱۱۷, ۰,۲۶۲۲)	(۱,۰۲۴۲, ۱,۱۵۵۹, ۱,۳۱۲۸)	۰,۱۳۷۵
رگرسیون	۱۰۰	(۱,۹۸۳۷, ۱,۱۰۵۴, ۰,۲۵۱۳)	(۱,۰۲۰۳, ۱,۱۴۷۷, ۱,۲۹۹۶)	۰,۱۱۱۵
	۲۰	(۱,۰۰۵۷, ۱,۱۴۲۲, ۱,۳۰۹۳)	(۱,۰۳۵۶, ۱,۱۶۵۸, ۱,۳۴۴۳)	۰,۲۳۰۶
	۴۰	(۱,۹۸۸۴, ۱,۱۱۵۶, ۰,۲۶۸۷)	(۱,۰۰۶۹, ۱,۱۴۹۰, ۱,۳۱۲۷)	۰,۱۶۶۴
	۶۰	(۱,۹۸۶۲, ۱,۱۱۱۷, ۰,۲۶۲۲)	(۱,۰۱۰۷, ۱,۱۵۰۸, ۱,۳۱۱۹)	۰,۱۳۴۴
	۱۰۰	(۱,۹۸۳۷, ۱,۱۰۵۴, ۰,۲۵۱۳)	(۱,۰۰۸۶, ۱,۱۴۳۸, ۱,۲۹۹۱)	۰,۱۰۸۸

بررسی سایر روش‌های تخمین داده‌های از دست رفته همچون روش بیشینه‌سازی ارزش انتظاری و همچنین بررسی اثر داده‌های از دست رفته بر عملکرد شاخص توانایی فرایند هنگامی که حدود مشخصه‌ی فنی نیز به‌صورت فازی در نظر گرفته شده است به‌عنوان تحقیقات آتی پیشنهاد می‌شود.

جدول ۶. بررسی اثر میزان فازی بودن مشاهدات بر برآورد داده‌های از دست رفته.

روش	میانگین شاخص توانایی فرایند	میانگین فاصله
میانگین	(۰٫۸۸۵۱، ۱٫۱۷۶۶، ۱٫۵۶۳۱)	۰٫۲۵۷۸
رگرسیون	(۰٫۹۲۱۱، ۱٫۱۶۹۱، ۱٫۵۶۳۱)	۰٫۲۵۷۱

پانوشتها

1. general linear model
2. non-structured
3. univariate
4. block-wise
5. file matching
6. particle swarm optimization
7. personal best
8. global best
9. inertia
10. P-value

منابع (References)

1. Kotz, S. and Johnson, N.L. "Process capability indices – a review", *Journal of Quality Technology*, **34**(1), pp. 1-19 (2002).
2. Taam, W., Subbairha, P. and Liddy, J.W. "A note on multivariate capability indices", *Applied Statistics*, **20**(3), pp. 339-351 (1993).
3. Shahriari, H., Hubele, N.F. and Lawrence, F.P. "A multivariate process capability vector", *Proceedings of the 4th Industrial Engineering Research Conference*, Institute of Industrial Engineers, pp. 304-309 (1995).
4. Chen, H. "A multivariate process capability index over a rectangular solid tolerance zone", *Statistica Sinica*, **4**(2), pp. 749-758 (1994).
5. Wang, F.K. and Chen, J.C. "Capability index using principal component analysis", *Quality Engineering*, **11**(1), pp. 21-27 (1998).
6. Shahriari, H. and Abdollahzadeh, M. "A new multivariate process capability vector", *Quality Engineering*, **21**(3), pp. 290-299 (2009).
7. Tsai, C.C. and Chen, C.C. "Making decision to evaluate process capability index C_p with fuzzy numbers", *International Journal of Advanced Manufacturing Technology*, **30**(3-4), pp. 334-339 (2013).
8. Parchami, A. and Mashinchi, M. "Fuzzy estimation for process capability indices", *Information Sciences*, **177**(6), pp. 1452-1462 (2007).
9. Parchami, A. and Mashinchi, M. "Testing the capability of fuzzy processes", *Quality Technology & Quantitative Management*, **6**(2), pp. 125-136 (2009).
10. Parchami, A., Mashinchi, M. and Partovi Nia, V. "A consistent confidence interval for fuzzy capability index", *Applied Computing and Mathematics*, **7**(1), pp. 143-161 (2008).
11. Chen, C.C., Lai, C.M. and Nien, H.Y. "Measuring process capability index C_{pm} with fuzzy data", *Quality & Quantity*, **44**(3), pp. 529-535 (2010).
12. Abdolshah, M., Yusuff, R.M., Hong, S.H., Ismail, Md.Y.B. and Naimi Sadigh, A. "Measuring process capability index C_{pmk} with fuzzy data and compare it with other fuzzy process capability indices", *Expert Systems with Applications*, **38**(6), pp. 6452-6457 (2011).
13. Mousavi, S.M., Nikzad, E. and Amiri, A. "An extension of MPCPI under fuzzy environment", Submitted (2017).
14. Little, R.J.A. and Rubin, D.B., *Statistical Analysis with Missing Data*, 2nd edition, John Wiley & Sons, Inc., New York, NY, 408 p. (2002).
15. Grung, B. and Manne, R. "Missing values in principal component analysis", *Chemometrics and Intelligent Laboratory Systems*, **42**, pp. 125-139 (1998).
16. Schafer, J.L. and Graham, J.W. "Missing data: Our view of the state of the art", *Psychological Methods*, **7**(2), pp. 147-177 (2002).
17. Acharya, N.N. and Black Nembhard, H. "Bayesian algorithms for missing observations in experimental designs for a nanolubrication process", *IIE Transactions*, **41**(11), pp. 969-978 (2009).
18. Flores-Cerrillo, J. and MacGregor, J.F. "Latent variable MPC for trajectory tracking in batch processes", *Journal Process Control*, **15**(6), pp. 651-663 (2005).
19. Lopes, V.V. and Menezes, J.C. "Inferential sensor design in the presence of missing data: A case study", *Chemometrics and Intelligent Laboratory Systems*, **78**(1-2), pp. 1-10 (2005).
20. Mahmoud, M.A., Saleh, N.A. and Madbuly, D.F. "Phase I analysis of individual observation with missing data", *Quality and Reliability Engineering International*, **30**(4), pp. 559-569 (2013).
21. Madbuly, D.F., Maravelakis, P.E. and Mahmoud, M.A. "The effect of methods for handling missing values on the performance of the MEWMA control chart", *Communications in Statistics – Simulation and Computation*, **42**(6), pp. 1437-1454 (2013).
22. Ashuri, A. and Amiri, A. "Evaluating estimation methods of missing data on a multivariate process capability index", *International Journal of Engineering*, **28**(1), pp. 101-110 (2015).

23. Eberhart, R.C. and Kennedy, J. "A new optimizer using particle swarm theory", *Proceedings of the Sixth Symposium Micro Machine and Human Science*, Piscataway, NJ, IEEE Service Center, pp. 39-43 (1995).
24. Kao, C. and Liu, S.T. "Fractional programming approach to fuzzy weighted average", *Fuzzy Sets and Systems*, **120**(3), pp. 435-444 (2001).
25. Cheng, C.H. "A new approach for ranking fuzzy numbers by distance method", *Fuzzy Sets and Systems*, **95**(3), pp. 307-317 (1998).