Monitoring Indicators of Research Data using I-MR Control Charts

Aynaz Ashtarian Esfahani
MA in Industrial Engineering; Systems Management and Productivity; Islamic Azad University; Science and Research Branch; Tehran, Iran Email: a_ashtarian@yahoo.com

Mohammad Javad Ershadi*
PhD in Industrial Engineering; Information Technology Department; Assistant Professor; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran;
Email: Ershadi@irandoc.ac.ir

Amir Azizi
PhD in Industrial Engineering Department; Assistant Professor; Islamic Azad University; Science and Research Branch; Tehran, Iran Email: Azizi@srbiau.ac.ir

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 35 | No. 4 | pp. 953-978
Summer 2020

Received: 22, Jul. 2019 **Accepted: 10, Mar. 2020**

Abstract: Key performance indicators are a useful tool in the field of operations management and help senior managers in the organization continually to improve their processes. In research information systems, data is considered as a valuable resource throughout the research process and much of the added value created in this process depends on the quality of the research data. Data quality control, on the other hand, is one of the most important activities to ensure that the quality of an organization's output is optimal, and is nowadays achieved through a powerful tool called control chart. In this paper, indicators such as accuracy, completeness and precision have been developed to monitor the performance of research information systems. The dissertation / thesis data of postgraduate students throughout the country that are disseminated in the GANJ System are monitored using the I-MR control chart. The results of continuous monitoring of data quality in different indices shows that in some of the newly integrated systems, there is a great deal of fluctuation in some indicators such as lack of familiarity with the system and information recording and lack of adequate training in some indices such as accuracy and precision. Lastly, providing timely feedback to students on data quality issues, appropriate notifications when completing fields for graduate students, and providing visual training and formulating necessary guidelines introduced as corrective actions to improve data quality and production process capability.

Keywords: Data Quality, Control Chart, Key Performance Indicators, I-MR Chart, GANJ System

* Corresponding Author

پایش شاخص‌های کیفیت داده‌های پژوهشی به‌کمک نمودار کنترل مشاهده‌های انفرادی با دامنهٔ متحرک

آیناز اشتریان اصفهانی
کارشناسی ارشد مهندسی صنایع؛ گروه مهندسی صنایع؛ دانشکده فنی‌مهندسی؛ دانشگاه آزاد اسلامی؛ واحد علوم و تحقیقات؛ تهران، ایران a_ashtarian@yahoo.com

محمدجواد ارشادی
دکتری مهندسی صنایع؛ استادیار؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران؛ پدیدآور رابط Ershadi@irandoc.ac.ir

امیر عزیزی
دکتری مهندسی صنایع؛ استادیار؛ گروه مهندسی صنایع؛ دانشکده فنی‌مهندسی؛ دانشگاه آزاد اسلامی؛ واحد علوم و تحقیقات؛ تهران، ایران؛ Azizi@srbiau.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۳۹۸/۰۴/۳۱ پذیرش: ۱۳۹۸/۱۲/۲۰ مقاله برای اصلاح به مدت ۸۵ روز نزد پدیدآوران بوده است.

چکیده: شاخص‌های کلیدی عملکرد ابزاری سودمند در حوزهٔ مدیریت عملیات است و به مدیران ارشد سازمان در بهبود مستمر فرایندها و واحدهای مختلف اجرایی یاری می‌رساند. در سامانه‌های اطلاعاتی پژوهشی، داده به‌عنوان منبعی ارزشمند در سرتاسر فرایند پژوهش به حساب می‌آید و بخش زیادی از ارزش افزودهٔ ایجادشده در این فرایند وابسته به کیفیت داده‌های پژوهشی است. از سوی دیگر، کنترل کیفیت داده‌ها یکی از مهم‌ترین فعالیت‌ها در راستای اطمینان از مطلوب‌بودن سطح کیفیت خروجی سازمان است که امروزه به‌کمک یک ابزار قدرتمند به نام نمودار کنترل انجام می‌شود. در این مقاله به‌منظور پایش عملکرد سامانه‌های اطلاعاتی تحقیقاتی شاخص‌هایی مانند صحت، کامل‌بودن و دقت توسعه داده شده‌اند. داده‌های پایان‌نامه/رساله، (پارسا)های دانشجویان تحصیلات تکمیلی کل کشور که در سامانهٔ «گنج» اشاعه داده می‌شوند، به کمک نمودار کنترل I-MR پایش شده‌اند. نتایج پایش مستمر کیفیت داده در شاخص‌های مختلف نشان می‌دهد که در برخی داتشگاه‌ها که به‌تازگی به سامانهٔ ثبت پیوسته‌اند، به‌دلیل عدم آشنایی با روند سامانه

فصلنامه | علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۵ | شمارهٔ ٤ | ص‌ص ۹۵۳-۹۷۸
تابستان ۱۳۹۹

و ثبت اطلاعات و نبود آموزش کافی در برخی شاخص‌ها مانند دقت و صحت، نوسانات زیادی وجود دارد و نیاز است اقدامات اصلاحی مناسبی انجام شود. در انتها، به ارائهٔ بازخورد به موقع به دانشجویان در خصوص مشکلات کیفی داده‌ها، پیغام‌های اطلاع‌رسانی (Notifications) مناسب در زمان تکمیل فیلدها توسط دانشجویان تحصیلات تکمیلی و ارائهٔ آموزش‌های تصویری و تدوین دستورالعمل‌های لازم به‌عنوان اقدامات اصلاحی در خصوص ارتقای کیفیت داده‌ها و قابلیت فرایندهای تولید داده تأکید شده است.

کلیدواژه‌ها: کیفیت داده، نمودار کنترل، شاخص‌های کلیدی عملکرد، نمودار I-MR، سامانهٔ گنج

۱. مقدمه

مدیران سازمان برای اندازه‌گیری عملکرد و پایش وضعیت کیفیت خروجی‌های سازمان از سنجه‌های متفاوتی استفاده می‌کنند. در یک نگاه کلی، سنجه‌های عملکردی را می‌توان به سه گروه اصلی شامل شاخص‌های کلیدی، نتیجه‌های عملکرد و شاخص‌های کلیدی عملکرد[1] دسته‌بندی کرد (Arabi, Heidarpour & Khoshgftar 2017). شاخص‌های کلیدی عملکرد بر خلاف نتیجه‌های عملکرد به فعالیت‌های اصلی سازمانی توجه کرده و اثربخشی اقدامات و فرایندها را مورد توجه قرار می‌دهند؛ در حالی که نتیجه‌های عملکرد، خروجی‌های کلان سازمان، مانند سود و میزان فروش را اندازه می‌گیرند. به‌منظور ارزیابی درست شاخص‌های کلیدی لازم است در ارزیابی‌ها از دیدگاه مشتریان (یا کاربران سازمان) به‌عنوان یک ذینفع کلیدی سازمان استفاده شود (Ershadi, Niaki & Sadeghee 2019). امروزه، پایش آماری شاخص‌ها جایگاه ویژه‌ای در برنامه‌ریزی‌های انجام‌شده در سطح مدیریت عملیات سازمان دارد. هدف از پایش آماری کیفیت، یافتن تعادل اقتصادی میان تلاش‌های انجام‌شده در بازنگری کنترل کیفیت و احتمال یافتن خروجی‌های نامطلوب است. بازرسی ۱۰۰ درصد یا به بیان دیگر، بازرسی از تمام نمونه‌های تولیدشده در فرایند، به کار و هزینهٔ زیادی نیاز دارد. نمودارهای کنترل و ابزارهای کنترل کیفیت آماری با استفاده از ایجاد تعادل و پایداری اقتصادی به ما کمک می‌کنند تا نمونه‌برداری در بهینه‌ترین شکل ممکن انجام شود و کارایی و اثربخشی مناسبی داشته باشد. نمودار کنترل در صورتی که به‌درستی پیاده‌سازی شود، می‌تواند راهنمای هر مدیر یا سرپرست در تصمیم‌گیری در

1. key performance indicator (KPI)

خصوص چگونگی ورود به یک فرایند جمع‌آوری اطلاعات باشد. از این رو، این ابزار آماری می‌تواند با صرفه‌جویی در زمان، هشدارهای لازم را به موقع ارائه کند. از سوی دیگر، ارزیابی عملکرد هر سازمان به کمک تعیین و تبیین شاخص‌های کلیدی عملکرد شروع می‌شود. در مرحلهٔ بعد، چگونگی نظارت در قالب KPIها توسط نمودار کنترل و ابزارهای مرتبط با آن بر پایهٔ اطلاعات تولیدشده در حین فرایند اجرا تعیین خواهد شد. این موضوع در فرایندهایی با کلان‌داده‌های ساختاریافته اهمیت ویژه‌ای می‌یابد (Jans, Sirkis & Morgan 2013). همچنین، کنترل فرایند آماری[1] یک روش نوین برای پایش فرایند است و به‌سرعت زمانی را که یک فرایند خارج از کنترل است، تعیین می‌کند و می‌تواند علت یا علل ریشه‌ای مشکلات فرایند را بررسی نماید و یک استراتژی برای بهبود ارائه دهد. در SPC از نمودارهای مختلف کنترل با توجه به ویژگی داده‌ها و پیچیدگی اجرا استفاده خواهد شد. انتخاب بهترین نمودار بستگی به نوع داده و اطلاعات و امکانات مدیریت و حتی بودجهٔ سازمان دارد. KPI نه‌تنها برای نظارت بر فرایند تولید و ارائهٔ محصول کاربرد دارد، بلکه به مشتری اجازه می‌دهد که عملکرد آن محصول را قضاوت کند. در نتیجه، KPI یک ابزار ضروری است و به ما کمک می‌کند تا به‌جای این‌که از کارایی ضعیف جلوگیری شود، به سمت موفقیت حرکت کنیم (Gasta 2004).

در سازمان‌های داده‌محور، باتوجه به حجم بالای دادهٔ تولید و پردازش‌شده، توسعهٔ KPI در کیفیت داده‌های تولیدشده به‌عنوان یک استراتژی کلیدی به حساب می‌آید. از این رو، امروزه شاخص‌های کیفیت داده و بومی‌سازی آن‌ها در سازمان به‌عنوان یک فعالیت حیاتی مورد توجه مدیران ارشد است که پژوهش‌های مختلفی نیز بر روی این موضوع انجام پذیرفته است. به‌عنوان نمونه، واحد مدیریت ارتباط با مشتریان یکی از مهم‌ترین بخش‌های سازمان به حساب می‌آید که داده‌های باارزشی در آن تولید و مدیریت می‌شود. امروزه، شاخص‌های کلیدی عملکرد بر روی کیفیت این داده‌ها در پژوهش‌های گوناگون مورد توجه است (Vaziri, Mohsenzadeh & Habibi 2019). شاخص‌هایی مانند صحت، دقت، کامل‌بودن، اعتبار و دسترس‌پذیری نمونه‌هایی از KPIها هستند که در صورت تعریف مناسب در سازمان می‌توانند به پایش عملکرد سازمان کمک شایانی کنند (Masayna et al. 2007). سیستم اندازه‌گیری کیفیت داده مقادیر کیفیت داده‌ها را در نقاط اندازه‌گیری

1. statistical process control (SPC)

بـا تواتـر خاصـی اندازه‌گیـری می‌کنـد. همچنیـن، شـاخص‌های عملکـرد کلیـدی کیفیـت داده، ابعـاد کیفیـت داده را در قالـب روابـط قابـل ارزیابـی توصیـف می‌کنـد (Otto & Österle 2015).[۱]

در سـامانه‌های اطلاعاتـی پژوهشـی داده‌هـای پژوهشـی به‌عنـوان یـک عنصـر کلیـدی بـه حسـاب می‌آیـد و پایـش کیفیـت ایـن دسـته از داده‌هـا در سـازمان‌های عهـده‌دار مسـئولیت و پاسـخگو، یکـی از مهم‌تریـن مأموریت‌هـای اصلـی اسـت. از جملـه مهم‌تریـن سـامانه‌های اطلاعاتـی پژوهشـی، سـامانهٔ «گنـج» اسـت. ایـن سـامانه به‌منظـور گـردآوری همـهٔ داده‌هـای پایان‌نامـه/ رسـاله‌های دانشـجویان داخـل کشـور به‌گونـه‌ای طراحـی شـده اسـت کـه دارای فیلدهـا یـا اقـلام داده‌ای مختلفـی شـامل نـام، نـام خانوادگـی، شـمارهٔ دانشـجویی، کـد ملـی، مقطـع تحصیلـی، و رشـتهٔ دانشـگاهی اسـت. در ایـن سـامانه بـه هـر رکـورد (سـابقه) فایـل مربـوط بـه پایان‌نامـه/ رسـالهٔ دانشـجو ضمیمـه شـده اسـت. مجموعـهٔ داده‌هـای مـورد اشـاره یـک فـراداده خوانـده می‌شـود. پایـش آمـاری کیفیـت فـرادادهٔ اشـاعه داده‌شـده در سـامانهٔ «گنـج» در قالـب طراحـی شـاخص‌های کلیـدی عملکـرد به‌عنـوان مسـئله‌ای اسـت کـه در ایـن پژوهـش مـورد توجـه قـرار گرفتـه اسـت.

در ایـن پژوهـش و در ادامـهٔ پژوهش‌هـای قبلـی در خصـوص کاربـرد رویکـرد کنتـرل فراینـد آمـاری به‌منظـور اندازه‌گیـری و تحلیـل شـاخص‌های کلیـدی عملکـرد، بـرای نخسـتین‌بار شـاخص‌های عملکـرد در حـوزهٔ کیفیـت داده بـه کمـک نمـودار کنتـرل پایـش خواهـد شـد. هـدف نهایـی ایـن پژوهـش ایجـاد بهبـود در کیفیـت داده‌هـای پژوهشـی در سـامانهٔ «گنـج» و ارائـهٔ راهکارهـای بهبـود کیفیـت در ایـن سـامانه اسـت. بـر پایـهٔ ایـن هـدف، پرسـش‌های ایـن پژوهـش به‌طـور مشـخص شـامل مـوارد زیـر اسـت:

- ◇ شـاخص‌های کیفیـت داده در متـون علمـی کدام‌انـد؟ و کدام‌یـک در سـامانهٔ ثبـت پایان‌نامـه/ رسـالهٔ دانش‌آموختـگان داخـل کشـور قابلیـت پیاده‌سـازی دارنـد؟
- ◇ چگونـه می‌تـوان فراینـد ثبـت پایان‌نامـه/ رسـالهٔ دانش‌آموختـگان داخـل کشـور را از دیـدگاه شـاخص‌های کیفیـت داده و بـر پایـهٔ نمـودار کنتـرل ارزیابـی کـرد؟

سـاختار مقالـهٔ حاضـر دارای بخش‌هـای زیـر اسـت: در بخـش دوم، بـه پیشـینهٔ مطالعـات صورت‌گرفتـه در حـوزهٔ کنتـرل فراینـد آمـاری و پایـش کیفیـت داده می‌پردازیـم و سـپس،

---

۱. بـرای مطالعـهٔ بیشـتر در مـورد چارچوب‌هـای ارزیابـی سـامانه‌های اطلاعاتـی بـه مقـالات Chapman & Kihn (2009) و Chen et al. (2010) مراجعـه شـود.

در بخش سوم، چارچوب پژوهش و گام‌های آن معرفی می‌شود. در ادامه و در بخش چهارم، نتایج را ارائه خواهیم کرد. و در نهایت، در بخش پنجم، به برخی نتایج کلیدی و پیشنهادهایی برای مطالعهٔ بیشتر اشاره خواهیم کرد.

۲. پیشینهٔ پژوهش

برای اندازه‌گیری ابعاد کیفیت داده‌ها، ابزارها و یا معیارهای مختلفی وجود دارد که به‌طور کلی، به‌صورت ذهنی[1] و یا عینی[2] دسته‌بندی می‌شوند. معیارهای ذهنی بر اساس نظرات و تجربیات کاربران داده که می‌توانند مدیران، متخصصان داده و یا سایر کاربران باشند، مشخص می‌شوند. «وزیری، محسن‌زاده و حبیبی» عنوان کردند که سنجش ذهنی، معمولاً با استفاده از مصاحبه‌هایی صورت می‌پذیرد که منعکس‌کنندهٔ نیازها و تجارب ذی‌نفعان از جمله گردآورندگان، نگهبانان و مصرف‌کنندگان داده باشد (Vaziri, Mohsenzadeh & Habibi 2019). از طرف دیگر، سنجش عینی بر پایهٔ فرمول‌های ریاضی است که برای اندازه‌گیری کیفیت یک مجموعه داده استفاده می‌شود (Shahbazi et al. 2019). ارزیابی و اطمینان از اعتبار مشتریان و کاربرانی که در سنجش و ارزیابی کیفیت تأثیرگذار هستند، برای سازمان‌ها کاری کلیدی و حیاتی است و لازم است به‌شکل مناسب انجام شود (Ershadi and Omidzadeh 2018). نمودار کنترل، ابزاری آماری و قدرتمند در حوزهٔ اندازه‌گیری و بهبود کیفیت یک مشخصه از محصول یا فرایند است که امروزه گسترهٔ کاربرد آن روزبه‌روز افزایش می‌یابد. در ادامهٔ این بخش، در زیربخش ۲-۱ پیشینهٔ مطالعات در حوزهٔ کاربرد کنترل فرایند آماری در پایش شاخص‌های کلیدی عملکرد معرفی می‌شود. سپس، در زیربخش ۲-۲ باتوجه به این‌که در پژوهش جاری پایش صورت‌گرفته در حوزهٔ کیفیت داده است، بنابراین، پژوهش‌هایی که به پایش کیفیت داده پرداخته‌اند، معرفی خواهد شد. و در نهایت، در زیربخش ۲-۳ پیشینهٔ نمودار کنترل استفاده‌شده برای پایش کیفیت داده ارائه می‌شود.

۲-۱. پایش شاخص‌های کلیدی عملکرد به کمک روش کنترل فرایند آماری

ارزیابی عملکرد به کمک مدل‌های رایج مانند مدل‌های تعالی سازمان، یک استراتژی کلیدی به حساب می‌آید که ارتباطی تنگاتنگ با میزان دانش سازمان در سطح

1. subjective 2. objective

داده و اطلاعات و حتی فرهنگ سازمانی دارد (Ershadi and Dehdazzi 2019). برای ارزیابی صحیح عملکرد سازمان، تدوین شاخص‌های کلیدی عملکرد (KPI) و پایش مستمر آن اجتناب‌ناپذیر است. در حوزهٔ پایش آماری KPI به کمک روش کنترل فرایند آماری SPC در صنایع و حوزه‌های مختلف پژوهش‌های گوناگونی انجام شده است؛ برای نمونه، اندازه‌گیری عملکرد تأمین‌کنندگان فروشگاه‌های زنجیره‌ای (2007) Morgan & Dewhurst، اندازه‌گیری تأثیر عوامل قومیتی و فرهنگی بر عملکرد کارکنان (2007) .Hoyles et al و همچنین، پایش KPI در خدمات فناوری اطلاعات شرکت‌های مخابراتی Suhairi & Gaol (۲۰۱۳) که در همهٔ آن‌ها از SPC استفاده شده است. از سوی دیگر، پایش KPI به کمک SPC در حوزهٔ داده‌ها سابقه و نمونهٔ چندانی ندارد. Jones-Farmer, Ezell & Hazen (2014) از جمله نویسندگانی هستند که در حوزهٔ پایش کیفیت داده به کمک نمودار کنترل مطالعه کرده‌اند. در درون مفهوم SPC، مبحث تغییرات[1] وجود دارد. تغییرات می‌تواند به دو شکل رخ دهد: نوسانات و یا تغییرات ذاتی در روند و یا تغییرات ناشی از برخی از مشکلات یا رخداد در روندهای دیگر. SPC از بسیاری از نمودارهای مختلف کنترل استفاده می‌کند. انتخاب بهترین نمودار بستگی به نوع خاصی از اطلاعات دارد که باید تحلیل شود. با توجه به این‌که در این پژوهش شاخص‌های کلیدی عملکرد کیفیت داده به کمک نمودار کنترل مورد پایش قرار می‌گیرد، در ادامه، به پیشینهٔ کیفیت داده و اهمیت آن می‌پردازیم.

۲-۲. کیفیت داده

داده در سازمان یک منبع حیاتی برای حمایت از فرایند کسب‌وکار و تصمیم‌گیری مدیریتی به حساب می‌آید (Khosroanjom et al. 2011؛ Recker et al. 2019). رشد انبار داده‌ها و دسترسی مستقیم به اطلاعات از منابع مختلف توسط مدیران و کاربران اطلاعات نیاز به آگاهی از کیفیت داده‌ها را در سازمان‌ها افزایش داده است (Azeroual and Schöpfel 2019؛ Lee et al. 2002). همان‌گونه که تولید یک محصول را می‌توان به‌صورت سیستم پردازشی توصیف کرد که بر روی مواد خام فعالیت می‌کند تا آن‌ها را به محصولات فیزیکی تبدیل کند، تولید اطلاعات را نیز می‌توان به‌صورت یک سیستم پردازشی تعریف کرد که بر روی داده‌های خام فعالیت می‌کند تا آن‌ها را به محصولات اطلاعاتی تبدیل کند (Wang 1998).

1. variation

داده‌های نامرغوب می‌توانند بر قابلیت سیستم و همچنین، تضمین عملکرد عملیاتی زیان وارد کنند (Redma 1998). از جهت دیگر داده‌های نامرغوب می‌توانند اثرهای اجتماعی و اقتصادی قابل توجهی داشته باشند (Wang & Strong 1996). همچنین، کیفیت پایین داده بر رضایت مشتریان و روحیهٔ کارکنان تأثیرگذار است (Dhillon & Backhouse 2000). از این رو، کیفیت داده‌ها به‌عنوان یک مسئلهٔ عملکردی و ضروری در ارتباط با فرایندهای عملیاتی، تصمیم‌گیری‌ها (Chengalur-Smith, Ballou & Pazer 1999) و همکاری‌های درون‌سازمانی (Batini and Scannapieco. 2006) شناخته شده است.

پژوهش‌های گوناگونی نشان می‌دهد که کیفیت داده‌ها یک مفهوم چندبعدی است (Pipino, Lee & Wang 2002). از این رو، برای درک بهتر اندازه‌گیری و بهبود کیفیت داده ابعاد مختلفی مشخص شده است (Vaziri, Mohsenzadeh & Habibi 2019). در این پژوهش ابعاد کیفیت داده Wang & Strong (1996) مورد استفاده قرار گرفته است. این ابعاد شامل ۱۵ بعد است که در ۴ دسته، طبقه‌بندی شده‌اند و این دسته‌بندی شامل ذاتی، دسترس‌پذیری، زمینه‌ای، و نمایشگری است. کیفیت داده، همچون هر مشخصهٔ کیفی یک محصول لازم است در سازمان مورد پایش قرار گیرد (Jones-Farmer, Ezell & Hazen 2014) و برای این کار لازم است KPI مناسب برای این حوزه تدوین گردد. در ادامه، به پیشینهٔ نمودار کنترل مورد استفاده در این پژوهش اشاره خواهیم کرد.

۲-۳. پایش کیفیت داده به کمک نمودار کنترل

یکی از اهداف نمودارهای کنترل دستیابی به ثبات در فرایند و حفظ آن است. ثبات در فرایند بر مبنای شرایطی تعیین می‌گردد که طی آن فرایند درجهٔ بالایی از یکنواختی عملکرد در گذشته را نشان داده و انتظار آن می‌رود که این یکنواختی در آینده نیز حفظ شود. مشخصهٔ این یکنواختی در عملکرد، قرار گرفتن داده‌ها در حدود کنترل است (Montgomery 2009). در ادبیات نمودار کنترل انواع مختلفی از نمودارها معرفی شده‌اند که در دو دستهٔ انفرادی و گروهی قابل تقسیم‌بندی هستند. در دستهٔ انفرادی هر نمونه به‌صورت تکی مورد ارزیابی و قضاوت قرار می‌گیرد و در دستهٔ گروهی چند نمونه به‌صورت هم‌زمان بررسی می‌شود. در این مقاله جهت حفظ ثبات و پایداری داده‌های پژوهشی از نمودار کنترل استفاده شده و با توجه به این‌که داده‌ها به‌صورت تکی توسط پژوهشگر تولید می‌شود، نمودار کنترل انفرادی انتخاب شده است. یکی از پرکاربردترین

این دسته از نمودارها، نمودار کنترل I-MR[1] است.

«بینسزیک، زیمسال و سمولینسکی» در پژوهشی از نمودار کنترلی I-MR به‌عنوان ابزاری برای ارزیابی پایداری فرایند فناوری ذوب و ریخته‌گری و استفاده از خاکستر آهن کم‌کربن برای دیسک‌های ترمز استفاده کردند. آن‌ها تصمیم گرفتند که برای تجزیه و تحلیل نتیجه‌گیری از پارامترهای مورد بررسی پژوهش خود و بررسی میزان کربن در ذرات، اندازه‌گیری سختی و غیره، از نمودارهای I-MR استفاده کنند (Binczyk, Szymszal & Smoliński 2007). نمودارهای کنترل I-MR در علم کنترل کیفیت آماری در دستهٔ نمودارهای متغیر[2] قرار دارد. این دسته از نمودارها برای تحلیل شاخص‌هایی به‌کار می‌روند که بتوان آن‌ها را در قالب عدد و رقم ارزیابی کرد. دستهٔ دیگری نیز در تحلیل آماری به کمک نمودار کنترل با نام نمودارهای کنترل وصفی[3] وجود دارند که در حالتی می‌توان از این دسته بهره برد که شاخص قابلیت کمی‌سازی نداشته باشد.

همان‌طورکه پیش از این اشاره شد، در این مقاله با توجه به این‌که هر داده به‌عنوان یک هویت مستقل در جریان تولید و اشاعهٔ داده‌ها در سامانه‌های اطلاعاتی تحقیقاتی نقش به‌سزایی دارد، به کمک نمودار I-MR به پایش کیفیت داده خواهیم پرداخت. لذا در ادامه، به نحوهٔ پایش آماری به کمک این نمودار اشاره می‌شود.

۲-٤. نمودار I-MR

دامنهٔ متحرک فردی یا I-MR، یک نوع نمودار کنترل است که معمولاً برای داده‌های تکی استفاده می‌شود. این نمودار در ابتدا توسط «والتر شوهارت»[4] توسعه یافت و از آن پس نمودارهای کنترل به‌عنوان نمودار «شوهارت» نامیده می‌شود. به‌صورت کلی، در I-MR دو دسته نمودار متفاوت مورد استفاده قرار می‌گیرد. دستهٔ اول شامل نمودار داده‌های فردی (نمودار انفرادی یا I chart) است و دستهٔ دوم دامنهٔ متحرک را کنترل می‌کند. دامنهٔ متحرک فاصله میان دو نقطهٔ پشت سر هم یا به بیان دیگر دو نمونهٔ آماری پشت سر هم است. این دو نمودار با هم اطلاعات مفیدی را ارائه می‌کنند و استفاده از یکی از آن‌ها به‌تنهایی ممکن است اطلاعات درست در مورد رفتار فرایند را نشان ندهد. داده‌های نمودار I-MR را نمی‌توان به‌صورت زیرگروه طبقه‌بندی کرد، زیرا ممکن است از منابع

1. individual moving range
2. variable
3. attribute
4. Walter Shewhart

مختلف و در شرایط تولید متفاوت فراهم آمده باشند و گروه‌بندی آن‌ها نتایج آماری را خدشه‌دار کند (Mukundam et al. 2013). از سوی دیگر، نرمال بودن داده‌ها پیش‌شرط اصلی استفاده از نمودارهای کنترل است. از آنجا که پایهٔ اصلی تحلیل‌های کمی صورت‌گرفته در نمودار کنترل توزیع آماری داده‌های استفاده‌شده در این ابزار است، ارزیابی چگونگی این توزیع پیشنیاز اصلی در استفاده از آن است. بیشتر نمودارهای کنترل طراحی‌شده بر پایهٔ نرمال بودن داده‌ها توسعه یافته‌اند و پیش از استفاده می‌توان تناسب داده‌ها با این توزیع را ارزیابی کرد. این موضوع در بخش چهارم تشریح خواهد شد.

۲-۵. تفسیر نمودارهای I-MR

در این بخش جهت آشنایی بهتر و دقیق‌تر، هر یک از نمودارهای I-MR را به‌طور جداگانه بررسی می‌کنیم.

۲-۵-۱. نمودار انفرادی

نمودار فردی نشان‌دهندهٔ داده‌های فردی در یک دورهٔ زمانی است. این نمودار برای تشخیص روند و تغییراتی که در این روند وجود دارد، مفید است و به ما در یافتن علت خاص در نمودار کمک می‌کند. لازم به ذکر است که داده‌های تولیدشده باید پی‌درپی باشند. اندازهٔ نمونه در نمودار انفرادی یک نمونه است و روابط به شرح زیر است:

$$UCL = \bar{x} + 3\frac{\overline{MR}}{d_2} \quad (۱)$$

$$CL = \bar{x} \quad (۲)$$

$$LCL = \bar{x} - 3\frac{\overline{MR}}{d_2} \quad (۳)$$

در روابط بالا MR دامنهٔ متحرک[1] است و $\bar{x}$ میانگین را نشان می‌دهد. d_2 عدد ثابتی است که در منابع کنترل کیفیت آماری بر پایهٔ جداول از پیش تعیین‌شده مقداردهی می‌شود.

۲-۵-۲. نمودار دامنهٔ متحرک

نمودار دامنهٔ متحرک، همان‌طور که از نام آن مشخص است، نموداری است که

1. moving range

توسط مقادیر حاصل از داده‌های متوالی مرتب‌شده بر اساس زمان ترسیم می‌شود. هر نقطه در محدودهٔ جابه‌جایی به‌صورت Xn - Xn-1 محاسبه می‌شود (Montgomery 2009).

$$MR_i = |X_i - X_{i-1}| \qquad (۴)$$

x_i: پارامتر کیفیت محصول در زیر گروه i

برخی نکات باید در تفسیر الگوهای مشاهده‌شده در نمودار محدودهٔ حرکتی انجام شود. نقاط در محدودهٔ حرکتی همبستگی دارند و این همبستگی اغلب ممکن است الگوی مشاهده‌شده را در نمودار تغییر دهد (همان).

۳. روش پژوهش

همان‌گونه که در مقدمهٔ این پژوهش اشاره شد، پایش شاخص‌های کیفیت داده‌های پژوهشی به کمک نمودار کنترل مناسب یکی از مهم‌ترین اهداف این پژوهش است. از سوی دیگر، بر پایهٔ مطالعهٔ پیشینهٔ پژوهش، در حوزه‌های مختلف صنعت و خدمات نمودار کنترل برای پایش عملکرد کیفیت استفاده شده است، اما در حوزهٔ کیفیت داده تاکنون پژوهشی انجام نشده است. از این رو، در این پژوهش، در سامانهٔ ثبت پایان‌نامه/رسالهٔ دانش‌آموختگان کل کشور به پایش کیفیت داده خواهیم پرداخت. از سوی دیگر، از آنجا که موضوع کیفیت داده و تعیین شاخص‌ها در این حوزهٔ پژوهشی، جدید به حساب می‌آید، لازم است قبل از پیاده‌سازی نمودار کنترل، ابتدا شاخص‌های مناسبی برای پایش تعیین نمود. به بیان دیگر، قبل از هرگونه کنترلی باید آنچه قرار است کنترل شود به‌درستی تعیین و تبیین شده و از نرمال بودن داده‌ها اطمینان پیدا کرد. بر پایهٔ این مقدمه، می‌توان پژوهش جاری را در سه مرحلهٔ اصلی زیر انجام داد (شکل ۱).

گام اول

تعیین و تبیین آیتم‌های ارزیابی مانند شاخص‌ها و روش محاسبه آنها

- بررسی پیشینه کیفیت داده و تعیین شاخص‌های کیفیت داده
- ارزیابی روایی شاخص‌های کیفیت داده
- تعیین قلمرو اجرایی ارزیابی

گام دوم

جمع‌آوری داده‌ها برپایه چارچوب اجرایی نمودارهای کنترل

- انتخاب نمودارکنترل مناسب از بین نمودارهای کنترل موجود در پیشینه
- جمع‌آوری داده‌ها

گام سوم

تحلیل داده‌ها به‌کمک نمودار کنترل I-MR

- اندازه‌گیری و تحلیل شاخص‌ها برپایه روابط طراحی‌شده
- پیاده‌سازی نمودار کنترل
- تحلیل داده‌ها به‌کمک قواعد نمودار کنترل

شکل ۱. مراحل پژوهش

هدف از این پژوهش ارائه یک چارچوب ارزیابی برای کیفیت داده‌های ثبت‌شده در سامانهٔ ثبت است. آیتم‌های ارزیابی، اقلام اطلاعاتی ثبت‌شده توسط کاربران سامانهٔ ثبت و به بیان دیگر فراداده‌های «پارسا»های ثبت‌شده در این سامانه است. اقلام اصلی این فراداده‌ها عبارت‌اند از نام، نام خانوادگی، اساتید راهنما و مشاور و اطلاعات کتابشناختی مدرک ثبت‌شده مانند عنوان، چکیده، مقطع تحصیلی، نام دانشگاه، دانشکده و گروه تحصیلی. از آنجا که این اقلام اطلاعاتی در بازیابی اطلاعات در سامانهٔ «گنج» نقشی کلیدی دارند، کیفیت آن‌ها باید در سطح مطلوبی باشد و از این رو، به‌عنوان آیتم‌های ارزیابی انتخاب شدند. همان‌گونه که در شکل ۱، قابل مشاهده است، مراحل انجام این پژوهش، در راستای ارائهٔ یک چارچوب ارزیابی شامل سه گام اصلی است. در گام اول، بر پایهٔ شاخص‌های کیفیت داده که در پیشینهٔ پژوهش به آن‌ها اشاره شده، به تعیین شاخص‌های ارزیابی کیفیت داده خواهیم پرداخت. در حقیقت، خروجی این گام شاخص‌های ارزیابی کیفیت داده و رابطه‌های ریاضی آن‌ها خواهد بود. ارزیابی روایی شاخص‌ها بر پایهٔ نظر پنج نفر از خبرگان انجام شده است. خبرگانی که در ارزیابی روایی این شاخص‌ها همکاری داشتند، دارای تجربهٔ کمینه پنج سال در حوزهٔ فناوری اطلاعات و آشنایی کافی با علم داده و مدیریت اطلاعات علمی بودند. تحصیلات همهٔ خبرگان، دکتری تخصصی بوده و در مورد چگونگی محاسبه و نوع رابطه‌های ریاضی به‌کاررفته (ارزیابی به‌صورت

نسبی یا مطلق) روایی شاخص‌ها را تأیید کردند. قلمرو ارزیابی شاخص‌های کیفیت داده، اطلاعات ثبت‌شده توسط دانش‌آموختگان تحصیلات تکمیلی در سامانهٔ ثبت در سال ۹۷ بوده است. در گام دوم، داده‌های مورد نیاز از سامانهٔ ثبت را گردآوری کرده و فرایند کنترل کیفیت داده را آغاز می‌کنیم. روش ارزیابی نمودار کنترل خواهد بود. در بخش ۳-۲ اهمیت و نقش این روش در ارزیابی کیفیت اشاره شد. در نهایت، در گام سوم پس از اطمینان از نرمال بودن داده‌های هر شاخص با استفاده از نرم‌افزار «مینی‌تب»[1] نمودار I و MR را رسم کرده و حدود کنترلی را مشخص می‌کنیم. با شناسایی نقاط خارج از حدود کنترلی فرایند حذف نقاط را مرحله‌به‌مرحله اجرا می‌کنیم تا زمانی که نمودار تحت کنترل قرار گیرد.

٤. یافته‌های پژوهش

بر پایهٔ سه گام اصلی معرفی‌شده در بخش چهارم، در این بخش به نتایج هر گام به‌تفکیک اشاره می‌شود. به بیان دیگر، در بخش ۴-۱ بر پایهٔ گام اول پژوهش شاخص‌های کیفیت داده به‌منظور ارزیابی کیفیت فراداده‌های «پارسا»های ثبت‌شده در سامانهٔ ثبت توسعه داده خواهند شد. در بخش ۴-۲ بر پایهٔ گام دوم پژوهش به چگونگی کنترل کیفیت داده‌ها در سامانهٔ ثبت خواهیم پرداخت و در نهایت، در بخش ۴-۳ نتایج تحلیل داده‌های جمع‌آوری‌شده در گام سوم را ارائه خواهیم کرد.

٤-۱. تعیین و تبیین شاخص‌های کیفیت داده

همان‌گونه که در مقدمه بیان کردیم در گام اول، هدف، تعیین و تبیین شاخص‌های پژوهش است. در ادامه، به تعاریف و بررسی روابط سه شاخص تعیین‌شده می‌پردازیم. نتایج تعیین شاخص‌های کیفیت داده در ادامه ارائه شده است.

٤-۱-۱. کامل بودن

معیار کامل‌بودن با شمارش تعداد فیلدهای تکمیل‌شده در یک رکورد فراداده بررسی می‌شود. کامل بودن به این معناست که یک شیء[2] (محتوای پژوهشی) به چه میزان همهٔ رکوردهای مورد نیاز را به‌منظور ارائهٔ (نمایش) اطلاعات کامل دربرمی‌گیرد (Ochoa & Duval 2006). این نمایش ایده‌آل با توجه به نوع کاربرد و مخاطب داده‌های

1. Minitab
2. object

مورد استفاده متفاوت است. در ادامه، رابطهٔ (۵) فرمول ساده‌ای برای ارزیابی کامل بودن داده‌های پژوهشی تولیدشده در سامانهٔ ثبت «پارسا» را ارائه می‌دهد.

$$\boldsymbol{Completeness(j) = \frac{\sum_{i=1}^{n} Per_Com(i)}{n}} \qquad (۵)$$

Completeness (j) = میزان کامل بودن نمونه‌ها (پایان‌نامه‌ها) در نمونه یا دورهٔ jام

n = تعداد پایان‌نامه‌ها در هر بار نمونه‌برداری

Per _Com (i): شاخص کامل بودن پایان‌نامه‌ها در نمونهٔ iام. این شاخص از نسبت تعداد فیلدهای پرشده در پایان‌نامهٔ iام به کل فیلدهای موجود در سامانهٔ «گنج» به‌دست می‌آید.

٤-١-٢. صحت

صحت یک رکورد فراداده شاخصی است که بیانگر درستی مقدار فیلد با توجه به منابع اطلاعاتی موجود است. «اوچوا و دوال» پیشنهاد کردند که صحت می‌تواند به‌عنوان فاصلهٔ معنایی بین اطلاعات داده‌شده از طریق ثبت فراداده و اطلاعات داده‌شده از طریق منابع معتبر باشد (Ochoa & Duval 2006). فاصلهٔ کوتاه‌تر به‌معنای دقت بالاتر رکورد فراداده است (Batini et al. 2009). در ادامه، بر پایهٔ فرمول‌های ارائه‌شده در برخی منابع و همچنین، نشست‌های برگزارشده رابطهٔ زیر برای ارزیابی شاخص صحت در «پارسا»های ثبت‌شده ارائه شد.

$$Per - Accuracy(j) = \frac{\sum_{i=1}^{n} Per_Acc(i)}{n} \qquad (۶)$$

$$Con - Accuracy(j) = \frac{\sum_{i=1}^{n} Con_Acc(i)}{n} \qquad (۷)$$

Per _ Acc(i) = درصد فیلدهای صحیح پایان‌نامهٔ iام به کل فیلدهای تکمیل‌شده توسط دانشجو

Con _ Acc(i) = تعداد فیلدهای صحیح پایان‌نامهٔ iام به کل فیلدهای تکمیل‌شده توسط دانشجو

n = تعداد پایان‌نامه‌ها در هر بار نمونه‌برداری

به بیان دیگر، به‌منظور ارزیابی شاخص صحت در کنترل کیفیت فراداده‌های سامانهٔ ثبت «پارسا»، هم از شاخص نسبت (Per _ Acc) و هم از شاخص تعداد (Con _ Acc) برای ارزیابی استفاده خواهد شد.

٤-١-٣. دقت

دقت ذاتی در فیلدهای متنی (مانند چکیده، عنوان و ...) به‌کار برده می‌شود. مشابه با

معیار دسترس‌پذیری، این مورد نیز مربوط به قابلیت خواندن است. این موضوع به‌طور مستقیم تحت تأثیر املای درست متن قرار می‌گیرد (Ochoa & Duval 2006). میزان دقت ذاتی یک فیلد متنی به‌کمک شاخص q_{ip} از رابطهٔ زیر محاسبه می‌شود.

$$q_{ip}(record) = 1 - \frac{m}{n} \qquad (۸)$$

در رابطهٔ (۸)، m تعداد اشتباهات املایی است و n تعداد کل کلمات است. برای مثال، شاخص دقت ذاتی در یک متن با ۵۰ کلمه و ۱۰ اشتباه املایی ۸۰ درصد است.

٤-۲. اعتبارسنجی شاخص‌ها

در این گام، پس از طراحی شاخص‌های کیفیت داده، خبرگان معرفی‌شده در بخش سوم به اعتبارسنجی شاخص‌های طراحی‌شده پرداختند. معیارهای استفاده‌شده برای این کار عبارت بودند از مرتبط بودن[1] شاخص با هدف پژوهش، سهولت درک، در دسترس بودن داده‌ها، اثربخشی، کاربردپذیری و عینیت (نقش نداشتن قضاوت شخصی در ارزیابی‌ها) (De Mendonça Lima, Aguiar & Storpirtis 2019). از آنجا که شاخص‌های طراحی‌شده ویژگی‌های اشاره‌شده را داشتند، روایی آن‌ها توسط خبرگان مورد تأیید قرار گرفت.

٤-۲-۱. اجرای کنترل کیفیت داده

همان‌گونه که در بخش‌های پیشین مطرح گردید، کارشناسان واحد ثبت و فراهم‌آوری اطلاعات، فراداده‌های «پارسا» را در همهٔ مدارک ثبت‌شده در سامانهٔ ثبت مورد بررسی قرار می‌دهند. گام‌های این کار به‌صورت زیر است:

- ورود به سامانهٔ ثبت: کارشناس اداره ثبت و فراهم‌آوری اطلاعات پس از ورود به سامانهٔ ثبت با درج عنوان sabt.irandoc.ac.ir مراحل کنترل پایان‌نامه/ رساله را آغاز می‌کند؛
- انتخاب پایان‌نامه/ رساله به‌منظور کنترل کیفیت فراداده: پس از کلیک روی تب عملیات و سپس، انتخاب گزینهٔ «پذیرش ایرانداک» مراحل کنترل کیفیت آغاز می‌شود؛
- کنترل اطلاعات پایان‌نامه/ رساله: در این مرحله اطلاعات پایان‌نامه/ رساله شامل عنوان فارسی و انگلیسی، گروه تحصیلی، رشته و گرایش، زبان اصلی، تاریخ دفاع، استاد راهنمای اول و دوم و نیز استاد مشاور مورد بررسی قرار می‌گیرد؛

1. relevancy

- کنترل محتوای پایان‌نامه/ رساله: در این بخش محتوای پایان‌نامه/ رساله و میزان همخوانی آن با فایل‌های Pdf و Word بر پایهٔ طرح کیفیت فرادادهٔ طراحی‌شده مورد ارزیابی و کنترل قرار می‌گیرد؛
- بازبینی نهایی در خصوص پایان‌نامه/ رساله: در این بخش پس از بررسی اطلاعات ثبت‌شده در سیستم و بر پایهٔ طرح کیفی فراداده‌ها، کارشناس مربوطه تصمیم به رد یا پذیرش مدرک خواهد گرفت.

٤-٢-٢. پیاده‌سازی نمودار I-MR

پیش از رسم نمودار کنترلی I-MR ضروری است که از نرمال بودن داده‌ها اطمینان پیدا کنیم. بنابراین، در بخش ۴ نتایج سنجش نرمال بودن داده‌ها و پیاده‌سازی نمودار کنترل ارائه خواهد شد.

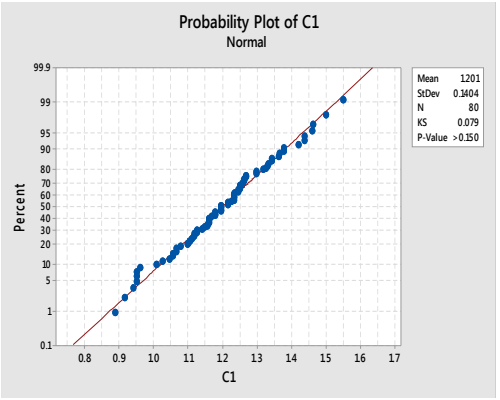
٤-٣. نتایج ارزیابی نرمال بودن داده‌ها و پیاده‌سازی نمودار

در این بخش بر پایهٔ گام سوم از مراحل پژوهش (ارائه‌شده در شکل ۱)، ابتدا نرمال بودن داده‌ها مورد ارزیابی قرار گرفته و نتایج پیاده‌سازی نمودارهای کنترل ارائه خواهد شد.

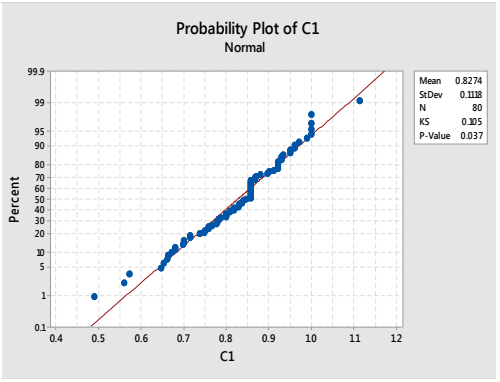
٤-٣-١. ارزیابی توزیع نرمال داده‌ها

پیش از رسم نمودار کنترل، نرمال بودن داده‌ها را بر اساس آزمون «کلوموگروف-اسمیرنوف»[1] می‌سنجیم. هنگام بررسی نرمال بودن داده‌ها، فرض صفر را بر این پایه که توزیع داده‌ها نرمال است، در سطح خطای ۵ درصد تست می‌کنیم (Montgomery 2009).

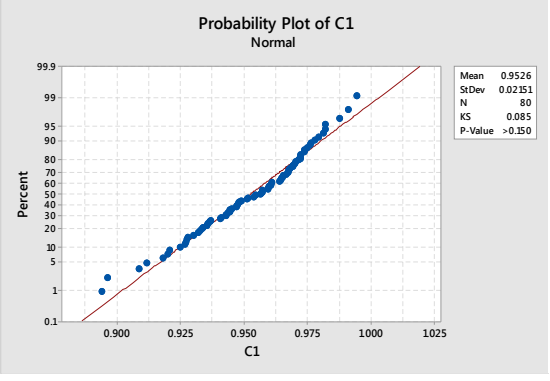
1. Kolmogorove _ Smirnov



نمودار ۱. شاخص کامل بودن



نمودار ۲. شاخص صحت



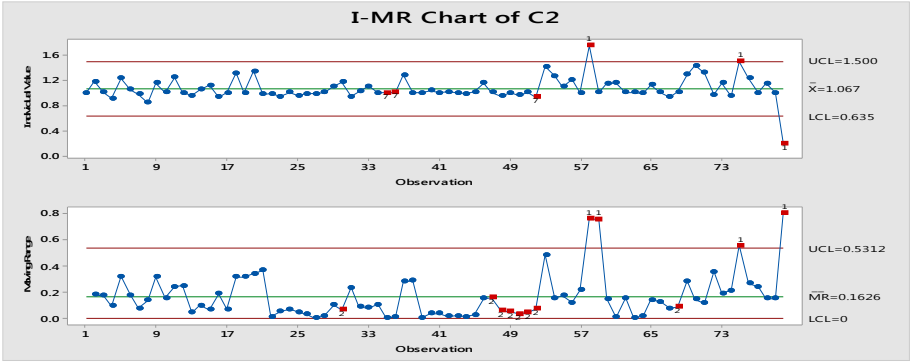
نمودار ۳. شاخص دقت

همان‌گونه که در نمودارهای بالا (۱ تا ۳) مشاهده می‌شود، p-value تمامی شاخص‌ها مقداری بیشتر از ۰/۰۵ دارند، که نشان‌دهندهٔ نرمال بودن دو شاخص کامل بودن و دقت است. اما در شاخص صحت با وجود این‌که بیشینهٔ داده‌ها بر روی خط نرمال قرار دارند و تنها ۴ داده از خط نرمال با فاصلهٔ کمی قرار دارند (مقادیر ۰/۴۸ و ۰/۵۵ و ۱/۰ و ۱/۱۱)، نتیجهٔ محاسبه P-value مقدار ۰/۰۳ بوده است. این عدد نشان می‌دهد که تنها با اختلاف ۰/۰۲ داده‌های این شاخص نرمال نیست، اما به‌دلیل ضرورت نرمال بودن داده‌ها هنگام استفاده از نمودار کنترل می‌توان با روش‌های نرمال‌سازی داده‌ها، اطلاعات این شاخص را نیز نرمال کرد.

٤-۳-۲. اجرای نمودار کنترل

همان‌طور که در بخش دوم بیان شد، از آنجا که هر داده مربوط به «پارسا» توسط دانشجوی تحصیلات تکمیلی و به‌صورت مستقل در سامانه ثبت می‌شود، از نمودار I-MR به‌منظور پایش و ارزیابی کیفیت استفاده خواهد شد. این نمودار شامل دو بخش است. در بخش بالایی نمودار (Individual value) نقاط به‌صورت تک‌به‌تک با یکدیگر مقایسه می‌شوند و بخش پایینی نمودار (Moving range) بیانگر اختلاف بین دو نقطهٔ متوالی است که این اختلاف بیانگر فاصله (رابطهٔ ۴) میان دو نقطه است. در ادامه، نمودار I-MR هر شاخص را رسم و بررسی می‌کنیم.

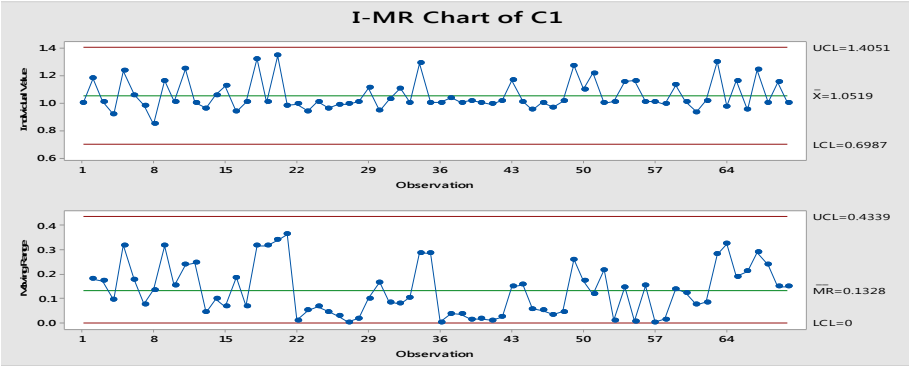
- **شاخص کامل بودن**



نمودار ٤. مشاهدات انفرادی شاخص کامل بودن

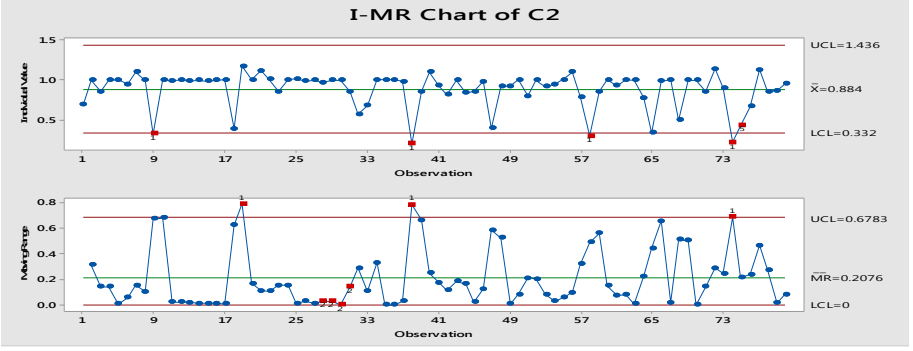
پس از بررسی دقیق و کارشناسی نمودار، نقاطی از زیرگروه‌ها که عامل خارج از

کنترل بودن نمودار هستند، نقاط ۳۵، ۳۶، ۵۲، ۵۸، ۷۵، و ۸۰ خواهند بود که لازم است پس از بررسی علت حذف شوند. با حذف نقاط ذکرشده، نمودار اصلاح‌شدهٔ زیر را مشاهده می‌کنیم (نمودار ۵).



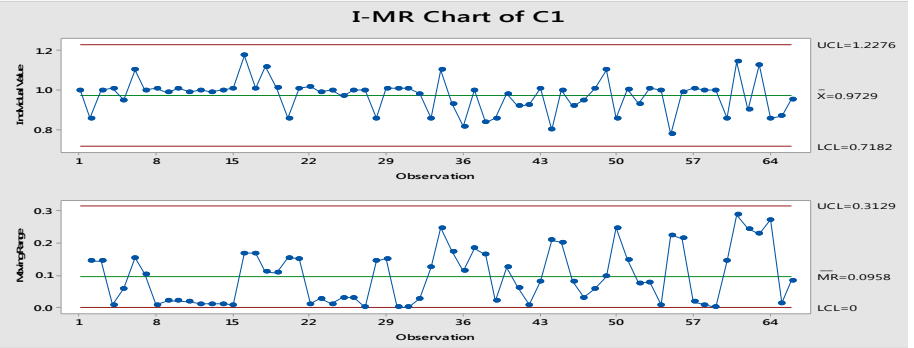
نمودار ۵. نمودار I-MR با حذف نقاط

◆ **شاخص صحت**



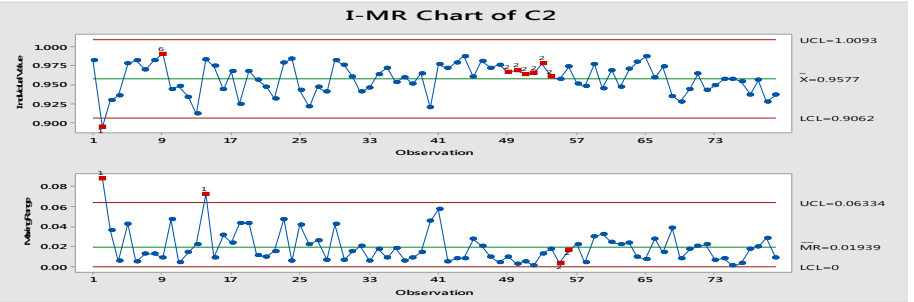
نمودار ٦. مشاهدات انفرادی شاخص صحت

همان‌طور که در نمودار بالا مشاهده می‌کنیم، نقاط ۹، ۳۹، ۵۸، ۷۴، و ۷۵ نقاط خارج از کنترل هستند. با حذف نقاط گفته‌شده تنها دو نقطه خارج از کنترل هستند که در نمودار انفرادی مشاهده می‌شوند. با حذف آن‌ها نمودار انفرادی را بار دیگر رسم کرده‌ایم. نمودار نهایی را در ادامه می‌بینیم.



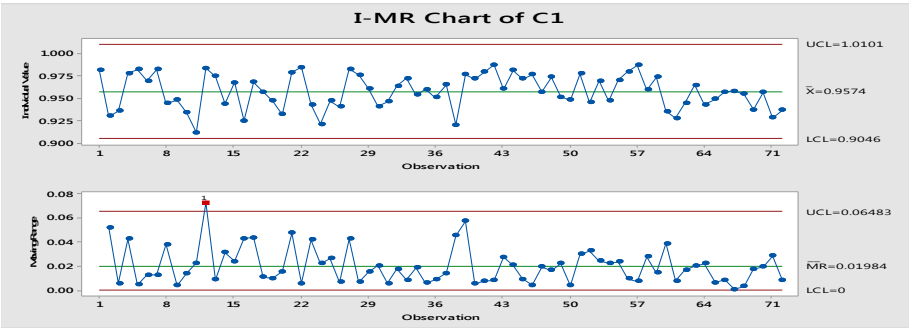
نمودار ۷. نمودار I-MR با حذف نقاط

◆ **شاخص دقت**

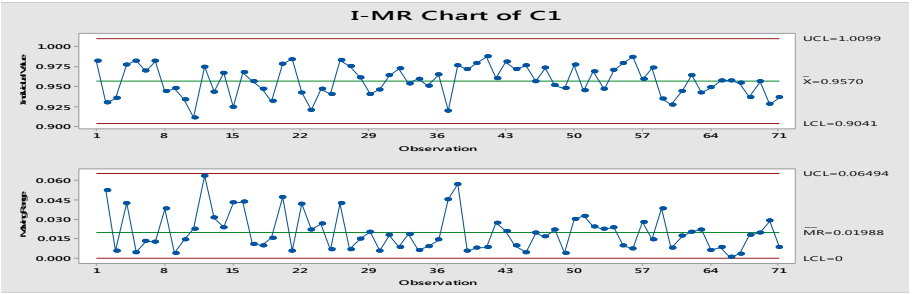


نمودار ۸. مشاهدات انفرادی شاخص دقت

نمودار بالا نشان‌دهندهٔ ۹ نقطهٔ خارج از کنترل است. بنابراین، با حذف آن‌ها نمودار را بار دیگر رسم می‌کنیم. حال، می‌توان گفت کدام‌یک از پایان‌نامه‌ها وضعیتی خارج از کنترل داشتند. بر پایهٔ نمودار (۹) پایان‌نامهٔ دهم در فرایند کنترل کیفیت در شاخص دقت طبق آزمون ۱، وضعیت خارج از کنترل را نمایش می‌دهد که لازم است دوباره بررسی و نمودار نهایی تهیه شود (نمودار ۱۰).



نمودار ۹. نمودار I-MR با حذف نقطهٔ خارج از کنترل



نمودار ۱۰. نمودار I-MR با حذف نقاط

در انتهای مراحل مختلف پیاده‌سازی نمودارهای کنترل و همراه با حذف نقاط خارج از کنترل، زمینهٔ مناسب برای به ثبات رسیدن فرایندهای ثبت داده‌ها فراهم گردید. همچنین، در طول فرایند کنترل کیفیت، مشکلات کیفی داده‌ها شناسایی و به کاربران مربوطه بازخوردهای لازم ارائه شد.

جهت افزایش شاخص کامل بودن بهتر است که پیغام (Error) در همان لحظهٔ ورود اطلاعات توسط دانشجویان و فیلدهای خالی ظاهر شود و برای بالا بردن شاخص صحت در سامانه، اطلاع‌رسانی مناسبی قبل از ثبت نام دانشجویان در سایت انجام شود، به‌طوری که هر نفر قبل از ثبت نام مؤظف به خواندن متن اطلاع‌رسانی باشد.

در بخش بعدی برخی نتایج کلیدی حاصل‌شده در حین پایش کیفیت داده‌ها و نیز پیشنهاداتی برای مطالعات آتی ارائه خواهد گردید.

۵. بحث و نتیجه‌گیری

در این پژوهش به‌منظور پایش کیفیت داده‌ها از نمودار کنترل استفاده شده است. نمودار کنترل به‌علت پویا بودن، مبتنی بر زمان بودن و پایش روند نقطه‌ها یکی از بهترین ابزارها برای پایش کیفیت در دنیا شناخته می‌شود. در این مقاله با توجه به مستقل بودن هر داده نسبت به دیگری از نمودار کنترل I-MR به‌منظور ارزیابی کیفیت داده‌های پژوهشی سامانهٔ «گنج» استفاده شده است. همچنین، شاخص‌هایی همچون کامل بودن، صحت، و دقت در این فرایند مورد بررسی قرار گرفت. پیشنهادات زیر به‌منظور بهبود کیفیت داده‌های پژوهشی پس از پایش مستمر فرایند ارائه شد.

- ارائهٔ بازخورد به‌موقع به دانشجویان در خصوص مشکلات کیفی داده‌های ارسال‌شده در سامانهٔ «گنج» به‌منظور رفع نواقص و ایجاد بهبود در داده‌ها؛
- فرهنگ‌سازی مناسب برای دانشجویان به‌کمک ارائهٔ توضیحات کامل در سامانهٔ «گنج» توسط کارشناسان کنترل کیفیت داده در فیلد بازخورد؛
- کاهش دوباره‌کاری در فرایند ثبت پایان‌نامه/رساله به‌کمک ارائهٔ آموزش‌های تصویری و با تدوین دستورالعمل‌های لازم؛
- ارائهٔ پیغام‌های اطلاع‌رسانی[1] مناسب در زمان تکمیل فیلدها توسط دانشجویان تحصیلات تکمیلی افزون بر موارد بالا در تفکیک هر دسته از شاخص‌ها پیشنهادات زیر ارائه می‌شود؛
- در مورد شاخص کامل‌بودن داده‌ها، از آنجا که جای برخی از اقلام اطلاعاتی مانند چکیده در برخی مدارک ثبت‌شده خالی است، بهبود طراحی سامانه در راستای جلوگیری از خالی بودن فیلد و اجازه ندادن به کاربر برای رفتن به مرحلهٔ بعد در صورت مشاهدهٔ چنین مشکلی یکی از مهم‌ترین راهکارهاست؛
- در شاخص صحت، از آنجا که امکان صحت‌سنجی اطلاعات اساتید راهنما و مشاور و حتی دانشجوی ثبت‌کننده وجود ندارد، پیشنهاد می‌شود به‌منظور بهبود روند مقادیر این شخص استفاده از کد ملی افراد و یا تعیین و طراحی شمارهٔ پژوهشگر صورت پذیرد. با این کار در صورت مشابهت اسامی امکان‌سنجی اعتبار پژوهشگر به‌راحتی صورت می‌پذیرد؛

1. notification

- در شاخص دقت، افزایش توانمندی در پرسنل نمایه‌ساز، که در مرحلهٔ بعدی کنترل تکمیلی را در این حوزه انجام می‌دهند، می‌توانند به بهبود شاخص کمک کند. همچنین، طراحی سیستم‌های انگیزشی حقوق و دستمزد به‌منظور افزایش دقت کارکنان در کنترل این شاخص می‌تواند به بهبود سیستم کمک شایانی نماید.

در مطالعات آتی پیشنهاد می‌شود اقدامات زیر مورد توجه قرار گیرد.

- از پروفایل‌های خطی یا غیرخطی که ابزاری برای پایش داده‌ها هستند در حوزهٔ کیفیت داده استفاده شود و همچنین، می‌توان از سایر نمودارهای کنترل مانند P و C در بعضی مواقع استفاده کرد که به بهبود مستمر کیفیت داده کمک می‌کنند.
- همچنین، می‌توان با وزن‌دهی شاخص‌های کیفیت داده به کمک تکنیک‌های تصمیم‌گیری اهمیت نسبی آن‌ها را سنجید.
- از آنجا که برخی شاخص‌های کیفیت داده با یکدیگر همبستگی آماری دارند، استفاده از نمودارهای چندمتغیره به‌منظور بهبود کیفیت داده در سامانه‌های اطلاعاتی می‌توانند مورد استفاده قرار گیرد.

References

Arabi, Pourya, Mehdi Zafar Heidarpour, & Ali Khoshgftar. 2017. Accelerating the achievement of critical success factors and key performance indicators, with an emphasis on the fundamental transformation document in the education research area of Mashhad 6th District. In the International Conference on Modern Approaches in Humanities. Stockholm.

Azeroual, O. & J. Schöpfel. 2019. Quality Issues of CRIS data: An exploratory investigation with universities from twelve countries. *Publications* 7 (1): 14.

Batini, C., C. Cappiello, C. Francalanci, & A. Maurino. 2009. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)* 41 (3): 16.

Batini, C. and M. Scannapieco. 2006. *Data Quality: Concepts, Methodologies and Techniques*. Berlin Heidelberg: Springer Verlag.

Binczyk, F., J. Szymszal, & A. Smoliński. 2007. IX-MR control chart as a tool in assessment of the cast iron properties stability. *Archives of Foundry Engineering* 7 (3): 25-28.

Chapman, C. S., & L. A. Kihn. 2009. Information system integration, enabling control and performance. *Accounting, organizations and society* 34 (2): 151-169.

Chen, L., W. Jian, Y. B. Li, Y. F. Zhai, & M. Rao. 2010. Research of performance monitoring for spatial information system. In 2010 International Conference on Future Information Technology and Management Engineering (Vol. 2, pp. 104-106). IEEE. Changzhou, China.

Chengalur-Smith, I. N., D. P. Ballou, & H. L. Pazer. 1999. The impact of data quality information on decision making: an exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering* 11 (6): 853-864.

De Mendonça Lima, T., P. M. Aguiar, & S. Storpirtis. 2019. Development and validation of key performance indicators for medication management services provided for outpatients. *Research in*

Social and Administrative Pharmacy 15 (9): 1080-1087.

Dhillon, G., & J. Backhouse. 2000. Technical opinion: Information system security management in the new millennium. *Communications of the ACM* 43 (7): 125-128.

Ershadi, M. J., & R. E. Dehdazzi. 2019. Investigating the role of strategic thinking in establishing organizational excellence model. *The TQM Journal* 31 (4): 620-640.

Ershadi, M. J., S . T. A. Niaki, & R. Sadeghee. 2019. Evaluation and improvement of service quality in information technology department of a detergent production company using the SERVQUAL approach. *International Journal of Services and Operations Management* 34 (2): 228-240.

Ershadi, M. J., & D. Omidzadeh. 2018. Customer validation using hybrid logistic regression and credit scoring model: A case study. *Calitatea* 19 (167): 59-62.

Gasta, W. 2004. Statistical process control using key process indicators for vacuum devices. In Vacuum Electronics Conference, 2004. IVEC 2004. Fifth IEEE International (p.109). IEEE. Monterey, CA, USA.

Hoyles, C., A. Bakker, P. Kent, & R. Noss. 2007. Attributing meanings to representations of data: The case of statistical process control. *Mathematical Thinking and Learning* 9 (4): 331-360

Jans, M., R. Sirkis, & D. Morgan. 2013. Managing data quality indicators with paradata based statistical quality control tools: the keys to survey performance. Improving Surveys with Paradata. Analytic Uses of Process Information, edited by Frauke Kreuter, 191-229. New Jersey: John Wiley & Sons.

Jones-Farmer, L. A., J. D. Ezell, & B. T. Hazen. 2014. Applying control chart methods to enhance data quality. *Technometrics* 56 (1): 29-41.

Khosroanjom, D., M. Ahmadzade, A. Niknafs, & R. K. Mavi. 2011. Using fuzzy AHP for evaluating the dimensions of data quality. *International Journal of Business Information Systems* 8 (3): 269-285.

Lee, Y. W., D. M. Strong, B. K. Kahn, & R. Y. Wang. 2002. AIMQ: a methodology for information quality assessment. *Information & management* 40 (2): 133-146.

Masayna, V., A. Koronios, J. Gao, & M. Gendron. 2007. Data quality and KPIs: a link to be established'. In The 2nd World Congress on Engineering Asset Management (EAM) and The 4th International Conference on Condition Monitoring (pp. 1376-1386). Harrogate, UK.

Montgomery, D. C. 2009. *Statistical quality control* (Vol. 7). New York: Wiley.

Morgan, C., & A. Dewhurst. 2007. Using SPC to measure a national supermarket chain's suppliers' performance. *International Journal of Operations & Production Management* 27 (8): 874-900.

Mukundam, K., D. R. Varma, G. R. Deshpande, V. Dahanukar, & A. K. Roy. 2013. I-MR control chart: A tool for judging the health of the current manufacturing process of an API and for setting the trial control limits in phase I of the process improvement. *Organic Process Research & Development* 17 (8): 1002-1009.

Ochoa, X. & E. Duval. 2006. Quality Metrics for Learning Object Metadata. In E. Pearson & P. Bohman (Eds.), *Proceedings of ED-MEDIA 2006--World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 1004-1011). Orlando, FL USA: Association for the Advancement of Computing in Education (AACE). Retrieved from Lirias.kuleuven.be/retrieve/6650 (accessed June 14, 2020).

Otto, Boris & Hubert Österle. 2015. *Corporate Data Quality: Prerequisite for Successful Business Models.* Berlin: e-publi.

Pipino, L. L., Y. W. Lee, & R. Y. Wang. 2002. Data quality assessment. *Communications of the ACM* 45 (4): 211-218.

Recker, J., M. Indulska, P. Green, A. Burton-Jones, & R. Weber. 2019. Information systems as representations: A review of the theory and evidence. *Journal of the Association for Information Systems* 20 (6): 5.

Redman, T. C. 1998. The impact of poor data quality on the typical enterprise. *Communications of the ACM* 41 (2): 79-83.

Scannapieco, M., B. Perinci, and E. Pierce. 2002. IP-UML: Towards a Methodology for Quality Improvement based on the IP-MAP Framework. In Proceedings of the 7th International Conference on Information Quality (ICIQ). Boston.

Shahbazi, M., A. Farajpahlou, F. Osareh, & A. Rahimi. 2019. Development of a scale for data quality assessment in automated library systems. *Library & Information Science Research* 41 (1): 78-84.

Suhairi, K., & F. L. Gaol. 2013. The measurement of optimization performance of managed service division with ITIL framework using statistical process control. *Journal of Networks* 8 (3): 518-529.

Vaziri, R., M. Mohsenzadeh, & J. Habibi. 2019. Measuring data quality with weighted metrics. *Total Quality Management & Business Excellence*708-720 :(6 ,5) 30 .

Wang, R. Y. 1998. A product perspective on total data quality management. *Communications of the ACM* 41 (2): 58-66.

_____, & D. M. Strong. 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems* 12 (4): 5-33.

آیناز اشتریان اصفهانی

متولـد ۱۳۷۲ دارای مـدرک کارشناسـی ارشـد در رشـتهٔ مهندسـی صنایـع از دانشـگاه آزاد واحـد علـوم و تحقیقـات تهـران اسـت.

کنتـرل کیفیـت آمـاری به‌ویـژه در حـوزهٔ داده و اطلاعـات از جملـه علایـق پژوهشـی وی اسـت.

محمدجواد ارشادی

دانش‌آموختـه دکتـری تخصصـی از دانشـگاه علـم و صنعـت در رشـتهٔ مهندسـی صنایـع اسـت. ایشـان هم‌اکنـون اسـتادیار گـروه پژوهشـی مدیریـت فنـاوری اطلاعـات پژوهشـگاه علـوم و فنـاری اطلاعـات ایـران (ایرانـداک) اسـت.

کنتـرل کیفیـت آمـاری، مدیریـت کیفیـت جامـع، بازمهندسـی فراینده‌ـای کسـب‌وکار، بهینه‌سـازی، الگوریتم‌هـای فراابتـکاری، تجزیـه و تحلیـل سیسـتم‌ها و داده‌کاوی ازجملـه علایـق پژوهشـی وی اسـت.

امیر عزیزی

دارای مدرک دکتری مهندسی صنایع است. ایشان هم‌اکنون استادیار گروه تخصصی مهندسی صنایع دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران است.

برنامه‌ریزی و کنترل در تولید ناب، مدیریت کیفیت جامع، مدیریت پروژه، بهره‌وری جامع، نت بهره‌ور فراگیر و همچنین، مدل‌سازی آماری، استنتاج فازی، شبکه‌های عصبی مصنوعی، نروفازی و تصمیم‌گیری چندمعیاره و شبیه‌سازی گسسته و سیستم‌های پویا از جمله علایق پژوهشی وی است.