

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران

طرح پژوهشی

**اندازه‌گیری و تحلیل شاخص‌های کیفیت داده در فرایند ثبت
پایان‌نامه / رساله‌های دانش‌آموختگان داخل کشور**

مجری

محمدجواد ارشادی

همکار

امیر عزیزی

خرداد ۱۳۹۸

حقوق: پژوهشگاه علوم و فناوری اطلاعات ایران

طرح پژوهشی « اندازه‌گیری و تحلیل شاخص‌های کیفیت داده در فرایند ثبت پایان‌نامه / رساله‌های دانش‌آموختگان داخل کشور » پیرو قرارداد شماره ۳۳/۱۳۹۸ مورخ ۱۳۹۷/۰۲/۰۸ میان پژوهشگاه علوم و فناوری اطلاعات ایران (کارفرما) و آقای محمدجواد ارشادی (به عنوان معجری) اجرا شده است. گزارش حاضر، از مستندات این طرح است.

این گزارش و تمامی حقوق مادی آن بر اساس «قانون حمایت حقوق مؤلفان و مصنفان و هنرمندان، مصوب سال ۱۳۴۸ و اصلاحیه‌های بعدی آن و همچنین آیین‌نامه‌های اجرایی این قانون متعلق به پژوهشگاه علوم و فناوری اطلاعات ایران است و هر گونه استفاده از تمامی یا پاره‌ای از آن، شامل: نقل قول، تکثیر، انتشار، کاربرد نتایج، تکمیل و مانند آنها به صورت چاپی، الکترونیکی یا و سایل دیگر فقط با اجازه کتبی پژوهشگاه امکان‌پذیر است. نقل قول در حد هزار واژه در انتشارات علمی مانند کتاب و مقاله با درج اطلاعات کامل کتاب‌شناختی، نیازی به مجوز پژوهشگاه ندارد.

صحت مندرجات گزارش برعهده معجری طرح پژوهشی است.



به نام خدا

آغاز نیم قرن دوم خدمات ارزشمند ایرانداک به علم، فناوری، و نوآوری کرامی باد

۱۳۴۲ - ۱۳۹۸

گواهی

انجام طرح پژوهشی:

- ◇ عنوان: اندازه گیری و تحلیل شاخص های کیفیت داده در فرایند ثبت پایان نامه / رساله های دانش آموختگان داخل کشور
- ◇ مجری: دکتر محمدجواد ارشادی
- ◇ همکار: دکتر امیر عزیزی
- ◇ شماره قرارداد: ۳۳/۱۳۹۸
- ◇ تاریخ قرارداد: ۱۳۹۷/۲/۸
- ◇ تاریخ شروع: ۱۳۹۷/۲/۸
- ◇ تاریخ پایان: ۱۳۹۸/۲/۲۵
- ◇ ناظر: دکتر سیدتقی اخوان نیایی

در پژوهشگاه علوم و فناوری اطلاعات ایران گواهی می شود. گزارش پایانی این طرح، بر پایه داوری در نشست ۳۳۲ (تاریخ ۱۳۹۸/۳/۱۹) شورای پژوهشی پژوهشگاه به تصویب رسیده است. لازم می دانم مراتب قدردانی خود را به مجری، همکار، و ناظر این طرح تقدیم دارم.

با آرزوی بهروزی
پرویز شهریاری
معاون پژوهش و آموزش

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده در فرایند ثبت پایان‌نامه / رساله‌های دانش‌آموختگان داخل کشور

مدیر طرح پژوهشی: محمدجواد ارشادی، هیئت علمی پژوهشگاه علوم و فناوری اطلاعات ایران

نشانی: تهران، خیابان انقلاب، چهارراه فلسطین، شماره ۱۰۹۰، طبقه ۴، اتاق ۴۰۲

صندوق پستی: ۱۳۱۵۷۷۳۳۱۴ تلفن: ۰۲۱-۶۶۴۹۴۹۸۰، دورنگار: ۰۲۱-۶۶۴۹۴۹۸۰

وبگاه: [www:irandoc.ac.ir](http://www.irandoc.ac.ir)

رایانامه: Ershadi@irandoc.ac.ir

همکار اصلی: امیر عزیزی

چکیده

امروزه سطح کیفیت داده و اطلاعات بعنوان یکی از مهم‌ترین شاخص‌های کلیدی عملکرد در سازمان‌ها محسوب می‌گردد. این دسته از شاخص‌ها به‌ویژه در سازمان‌هایی که برپایه داده و راهکارهای فناوری اطلاعات شکل گرفته‌اند حیاتی‌تر جلوه می‌کنند. پایش و اندازه‌گیری شاخص‌های کیفیت داده‌ها از یک‌سو به شناخت وضعیت موجود عملکرد سازمان در حوزه تولید و پردازش داده کمک می‌کند و از سوی دیگر به تصمیم‌گیری مدیران ارشد بمنظور تعیین اقدامات اثربخش در راستای راهبردهای سازمان کمک شایانی خواهد کرد. در طرح پژوهشی حاضر شاخص‌های کیفیت داده در سامانه ثبت پارساها تدوین شده و این شاخص‌ها در بازه‌های زمانی مناسب اندازه‌گیری و تحلیل شد. بمنظور تحلیل درست و مناسب شاخص‌ها نمودارهای کنترل که برپایه دانش کنترل کیفیت آماری توسعه یافته شده‌اند به کار گرفته شد. جنبه‌های مختلف کیفیت داده مانند دقت، صحت، به‌روز بودن و کامل بودن تعیین شده و اندازه‌گیری شد. نمودارهای چندمتغیره T^2 ، میانگین-دامنه، تعداد نقص و تعداد نقص در واحد بمنظور تحلیل شاخص‌های کیفیت داده مورد استفاده قرار گرفت. سامانه ثبت پایان‌نامه/رساله که به تازگی ارتقاء یافته است در قالب این پژوهش بعنوان مورد مطالعه مورد بررسی قرار گرفت. همچنین در انتها بهبودهایی که از جنبه‌های مختلف کمی و کیفی در این سامانه نسبت به سامانه ثبت قدیم صورت گرفته است معرفی و تحلیل گردید.

کلیدواژه‌ها: کیفیت داده؛ نمودار کنترل؛ سامانه ثبت پارسا؛ قابلیت فرایند

فهرست مطالب

عنوان	شماره صفحه
چکیده.....	۱
فهرست مطالب.....	ب
فصل اول – مقدمه	۱
۱-۱- بیان مسئله/ موضوع پژوهش.....	۲
۱-۲- اهداف پژوهش.....	۴
۱-۳- پرسش ها و فرضیه های پژوهش.....	۴
۱-۴- معرفی پژوهشگاه علوم و فناوری اطلاعات ایران.....	۴
۱-۴- معرفی مرکز اطلاعات و مدارک علمی ایران.....	۶
۱-۴- معرفی سامانه ثبت پایان نامه/ رساله دانش آموختگان کل کشور.....	۹
فصل دوم – مبانی نظری تحقیق	۱۰
۲-۱- مقدمه.....	۱۱
۲-۲- کیفیت داده ها.....	۱۲
۲-۳- مدیریت کیفیت جامع داده ها.....	۱۴
۲-۴- معرفی مدل اصلی کیفیت داده.....	۲۰
۲-۵- ضرورت معرفی شاخص کلیدی عملکرد در ارزیابی کیفیت داده ها.....	۲۳
۲-۶- روش پژوهش.....	۲۶

۲۷	۷-۲ جامعه آماری و حجم نمونه بمنظور تبیین وزن شاخصها
۲۸	۸-۲ نحوه استخراج عوامل و زیرعوامل
۲۹	۹-۲ بررسی روایی و پایایی ابزار جمع آوری اطلاعات
۳۱	۱۰-۲ منطق فازی
۳۲	۱۱-۲ روش امتیازدهی
۳۳	۱۲-۲ روش دیمتل فازی بر مبنای ANP فازی
۳۹	۱۳-۲ فرایند انجام FDANP
۴۰	فصل سوم - استخراج شاخصها و درجه اهمیت آنها
۴۱	۳-۱ مقدمه
۴۱	۳-۲ معرفی تحقیقات انجام شده در خصوص اندازه گیری کیفیت داده
۷۷	۳-۳ توسعه شاخص های کیفیت داده در ارزیابی فرایند ثبت
۸۱	۴-۳ تعیین درجه اهمیت شاخصهای کیفیت داده
۱۲۷	فصل چهارم - اندازه گیری و تحلیل شاخص های کیفیت داده به کمک نمودار کنترل
۱۲۸	۴-۱ مقدمه
۱۳۱	۴-۲ معرفی نمودارهای کنترل
۱۳۱	۴-۲،۱ نمودار کنترل XBar-R
۱۳۲	۴-۲،۲ نمودار کنترل I-MR
۱۳۳	۴-۲،۳ نمودار کنترل EWMA
۱۳۴	۴-۲،۴ نمودار کنترل چندمتغیره T^2
۱۳۵	۴-۲،۵ نمودار تعداد نقص (C)
۱۳۶	۴-۲،۵ نمودار تعداد نقص در واحد (U)
۱۳۷	۴-۳ پیاده سازی نمودارهای کنترل
۱۳۷	۴-۳،۱ فاز اول
۱۴۸	۴-۳،۲ فاز دوم
۱۷۸	۴-۴ مقایسه سامانه های ثبت قدیم و جدید از جنبه های کمی و کیفی

۱۷۹	 ۴-۱ مقایسه کیفی سامانه‌های ثبت قدیم و جدید
۱۸۰	 ۴-۲ مقایسه کمی سامانه‌های ثبت قدیم و جدید
۱۸۰	 ۴-۵ خلاصه‌ای از برخی نتایج کاربردی
۱۸۲	 فهرست منابع
۱۸۳	 منابع لاتین
۱۸۶	 منابع فارسی

فهرست جداول

عنوان	شماره صفحه
جدول ۱-۲ مقایسه تولید محصول و اطلاعات.....	۱۱
جدول ۲-۲ بررسی پایایی ابعاد پرسشنامه عوامل موثر بر کیفیت فراداده های سامانه ثبت پایان نامه در داخل کشور.....	۳۱
جدول ۳-۲ متغیرهای زبانی و اعداد فازی متناظر.....	۳۳
جدول ۱-۳ اطلاعات مورد نیاز برای محاسبه پروتکل متریک.....	۵۶
جدول ۲-۳ ابعاد کیفیت داده	۶۴
جدول ۳-۳ تعاریف انتخاب شده برای بعد رواج.....	۳۴
جدول ۴-۳ ارزیابی معیارهای موجود شاخص رایج.....	۷۶
جدول ۵-۳ شناسنامه شاخص کامل بودن.....	۷۷
جدول ۶-۳ شناسنامه شاخص صحت.....	۷۸
جدول ۵-۳ شناسنامه شاخص به هنگام بودن.....	۷۹
جدول ۶-۳ شناسنامه شاخص سازگاری.....	۸۱
جدول ۷-۳ فراوانی جامعه آماری بر حسب جنسیت.....	۸۲
جدول ۸-۳ فراوانی جامعه آماری بر حسب سن.....	۸۲
جدول ۹-۳ توزیع فراوانی تحصیلات افراد خبره.....	۸۳
جدول ۱۰-۳ توزیع فراوانی سابقه کار افراد نمونه.....	۸۴
جدول ۱۱-۳ عوامل و زیرعوامل موثر بر کیفیت فراداده های سامانه ملی ثبت پایان نامه داخل کشور.....	۸۵
جدول ۱۲-۳ ارزیابی زیرعوامل توسط خبرگان.....	۸۷
جدول ۱۳-۳ جدول نهایی عوامل و زیرعوامل انتخاب شده توسط خبرگان.....	۸۸
جدول ۱۴-۳ متغیرهای زبانی و اعداد فازی متناظر.....	۹۱
جدول ۱۵-۳ ماتریس ارتباط مستقیم ($\tilde{D}$) (میانگین نظر تمام خبرگان).....	۹۲
جدول ۱۶-۳ ماتریس نرمالیزه شده ($\tilde{H}_{ij}$).....	۹۵
جدول ۱۷-۳ ماتریس ارتباط کامل ($\tilde{T}_C$).....	۹۸
جدول ۱۸-۳ ماتریس ارتباط کامل عوامل اصلی ($\tilde{T}_D$).....	۹۹

جدول ۱۹-۳ محاسبه میزان شاخص r و d برای ماتریس $\tilde{T}_C$	۱۰۰
جدول ۲۰-۳ محاسبه شدت و جهت تاثیر عوامل اصلی	۱۰۳
جدول ۲۱-۳ محاسبه شدت و جهت تاثیر زیرعوامل	۱۰۴
جدول ۲۲-۳ ماتریس ارتباط کامل عوامل اصلی (TD) دیفازی شده	۱۰۶
جدول ۲۳-۳ ماتریس ارتباط کامل زیرعوامل (TC) دیفازی شده	۱۰۷
جدول ۲۴-۳ ماتریس نرمال ارتباط کامل عوامل اصلی (T_D^α)	۱۱۲
جدول ۲۵-۳ سوپر ماتریس ناموزون - ماتریس نرمال ارتباط کامل زیرعوامل (T_C^α)	۱۱۴
جدول ۲۶-۳ سوپر ماتریس موزون (W^α)	۱۱۹
جدول ۲۷-۳ سوپر ماتریس موزون همگرا شده	۱۲۲
جدول ۲۸-۳ وزن و رتبه عوامل و زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور	۱۲۴

فهرست شکل ها

عنوان	شماره صفحه
شکل ۱-۲ چرخه ی TDQM.....	۱۶
شکل ۲-۲. معرفی چرخه دمینگ.....	۱۶
شکل ۳-۲. معرفی ابعاد مختلف کیفیت داده.....	۲۲
شکل ۴-۲ تابع عضویت و نمودار عدد فازی مثلثی.....	۳۲
شکل ۵-۲ فرایند انجام FDANP.....	۳۹
شکل ۱-۳ مقایسه ارزیابی ذهنی و عینی.....	۶۸
شکل ۲-۳ کیفیت طراحی در مقابل کیفیت انطباق.....	۷۰
شکل ۱-۴ آزمون تصادفی بودن شاخص کامل بودن.....	۱۳۸
شکل ۲-۴ آزمون تصادفی بودن شاخص صحت.....	۱۳۹
شکل ۳-۴ آزمون تصادفی بودن شاخص دقت.....	۱۴۰
شکل ۴-۴ آزمون تصادفی بودن شاخص به هنگام بودن.....	۱۴۱
شکل ۵-۴ نرمال بودن شاخص کامل بودن.....	۱۴۲
شکل ۶-۴ نرمال بودن شاخص صحت.....	۱۴۳
شکل ۷-۴ نرمال بودن شاخص دقت.....	۱۴۴
شکل ۸-۴ نرمال بودن شاخص به هنگام بودن.....	۱۴۵
شکل ۹-۴ هیستوگرام شاخص کامل بودن.....	۱۴۶
شکل ۱۰-۴ هیستوگرام شاخص صحت.....	۱۴۶
شکل ۱۱-۴ هیستوگرام شاخص دقت.....	۱۴۷
شکل ۱۲-۴ هیستوگرام شاخص به هنگام بودن.....	۱۴۸
شکل ۱۳-۴ نمودار کنترل Xbar-R شاخص کامل بودن.....	۱۴۹

شکل ۴-۱۴ نمودار اصلاح شده $\bar{X}$ -R	۱۵۰
شکل ۴-۱۵ شکل نهایی $\bar{X}$ -R	۱۵۱
شکل ۴-۱۶ نمودار کنترل $\bar{X}$ -R شاخص صحت	۱۵۱
شکل ۴-۱۷ نمودار اصلاح شده $\bar{X}$ -R برای شاخص صحت	۱۵۲
شکل ۴-۱۸ شکل نهایی $\bar{X}$ -R برای شاخص صحت	۱۵۳
شکل ۴-۱۹ نمودار کنترل $\bar{X}$ -R شاخص دقت	۱۵۳
شکل ۴-۲۰ شکل اصلاح شده $\bar{X}$ -R شاخص دقت	۱۵۴
شکل ۴-۲۱ نمودار کنترل $\bar{X}$ -R شاخص به هنگام بودن	۱۵۵
شکل ۴-۲۲ نمودار اصلاح $\bar{X}$ -R شاخص به هنگام بودن	۱۵۶
شکل ۴-۲۳ نمودار نهایی $\bar{X}$ -R شاخص به هنگام بودن	۱۵۶
شکل ۴-۲۴ مشاهدات انفرادی شاخص کامل بودن	۱۵۷
شکل ۴-۲۵ نمودار I -MR با حذف نقاط	۱۵۷
شکل ۴-۲۶ مشاهدات انفرادی شاخص صحت	۱۵۸
شکل ۴-۲۷ نمودار I -MR با حذف نقاط	۱۵۸
شکل ۴-۲۸ مشاهدات انفرادی شاخص دقت	۱۵۹
شکل ۴-۲۹ نمودار I -MR با حذف نقطه برای دقت	۱۵۹
شکل ۴-۳۰ نمودار I -MR با حذف نقاط	۱۶۰
شکل ۴-۳۱ نمودار EWMA شاخص کامل بودن	۱۶۱
شکل ۴-۳۲ نمودار EWMA شاخص کامل بودن پس از حذف نقاط خارج از کنترل	۱۶۱
شکل ۴-۳۳ نمودار EWMA شاخص کامل بودن در شکل نهایی	۱۶۲
شکل ۴-۳۴ نمودار EWMA شاخص صحت	۱۶۳
شکل ۴-۳۵ نمودار EWMA شاخص صحت در شکل نهایی	۱۶۳
شکل ۴-۳۶ نمودار EWMA شاخص دقت	۱۶۴
شکل ۴-۳۷ نمودار EWMA شاخص به هنگام بودن	۱۶۵
شکل ۴-۳۸ نمودار چند متغیره T^2 شاخصهای کامل بودن و به هنگام بودن	۱۶۶
شکل ۴-۳۹ نمودار چند متغیره T^2 شاخص صحت و به هنگام بودن	۱۶۷
شکل ۴-۴۰ نمودار چند متغیره T^2 شاخص دقت و به هنگام بودن	۱۶۸
شکل ۴-۴۱ نمودار چند متغیره T^2 شاخص صحت و دقت	۱۶۹
شکل ۴-۴۲ نمودار چند متغیره T^2 شاخص صحت و کامل بودن	۱۷۰
شکل ۴-۴۳ نمودار چند متغیره T^2 شاخص صحت و کامل بودن	۱۷۱

شکل ۴-۴۴ نمودار چند متغیره T^2 شاخص دقت و کامل بودن.....	۱۷۲
شکل ۴-۴۵ نمودار چند متغیره T^2 شاخص دقت و کامل بودن در حالت نهایی.....	۱۷۳
شکل ۴-۴۶ نمودار اولیه تعداد نقص برای شاخص دقت.....	۱۷۴
شکل ۴-۴۷ نمودار نهایی تعداد نقص برای شاخص دقت.....	۱۷۴
شکل ۴-۴۸ نمودار اولیه تعداد نقص در واحد برای شاخص دقت.....	۱۷۵
شکل ۴-۴۹ نمودار نهایی تعداد نقص در واحد برای شاخص دقت.....	۱۷۶
شکل ۴-۵۰ نمودار اولیه تعداد نقص در سامانه ثبت قدیم برای شاخص دقت.....	۱۷۷
شکل ۴-۵۱ نمودار نهایی تعداد نقص در سامانه ثبت قدیم برای شاخص دقت.....	۱۷۸

فصل اول – مقدمه

۱-۱ بیان مسئله/ موضوع پژوهش

امروزه داده‌ها یکی از منابع اصلی سازمان‌ها هستند، که می‌توانند در تصمیمات و فرایند سازمانی تاثیرگذار باشند و مقدمات بهبود یا رشد سازمان‌ها را بوجود آورند. رشد چشمگیر فن آوری اطلاعات^۱ و ارتباطات در سال‌های اخیر و فراگیر شدن آن، سازمانها را با حجم عظیمی از داده‌ها و اطلاعات روبرو کرده است و روزانه نیز این حجم از داده‌ها رو به افزایش است.

به همین دلیل میتوان گفت که مسئله‌ی کیفیت داده^۲، یکی از مهمترین مسائلی است که سازمانها با آن مواجه هستند. زیرا هرچه کیفیت داده‌ها بهتر باشد در نتیجه‌ی آن تصمیمات بهتر و کاراتری نیز گرفته می‌شود. اگرچه برخی سازمان‌ها به این مبحث نمی‌پردازند و ارزش داده را بدرستی نمی‌دانند و آن را مهم تلقی نمی‌کنند، اما می‌توان گفت که داده‌های با کیفیت نامناسب می‌تواند بر سازمان اثر گذاشته و موجب تاثیرات نامطلوب را بر سازمان فراهم آورد. در این صورت ارزیابی ابعادی که به طور خاص مشخص کننده‌ی کیفیت داده‌ها می‌باشد حائز اهمیت است.

تحولات فناورانه موجب شده است تا سازمان‌ها با حجم داده‌های اطلاعاتی بسیار بیشتری مواجه شوند. اگرچه امروزه اهمیت کیفیت داده‌ها تا حدودی درک شده است، اما هزینه‌های غیر ضروری و چالش‌های زیادی در بسیاری از شرکت‌ها بوجود می‌آید. کیفیت داده‌ها شامل جنبه‌هایی زیادی است که باعث می‌شود عملی‌سازی این مفهوم یک چالش بزرگ باشد. کیفیت داده‌ها شامل ویژگی‌ها یا ابعاد مختلفی است که اهمیت زیادی برای رسیدگی موثر به مسائل مربوطه دارد. کیفیت ضعیف داده‌ها در بسیاری از زمینه‌های تجاری و عملیاتی تاثیر می‌گذارد. یکی از چالش‌های کیفیت داده‌ها تغییر به دلیل داده‌های موجود در پایگاه‌های داده‌ای است که نیاز به هماهنگ‌سازی با کل سامانه‌ها و اجزاء سازمان تا با هدف از پیش تعیین شده متناسب باشد. به طور کلی، کیفیت ضعیف داده‌ها می‌تواند تاثیرات شدیدی بر کل سازمان داشته باشد، که منجر به افزایش هزینه‌ها،

^۱ Information Technology (IT)

^۲ Data Quality (DQ)

نارضایتی مشتری و حتی بر روحیه کارکنان تاثیر می‌گذارد. از این رو، کیفیت داده‌ها فقط مشکل فناوری اطلاعاتی نیست، بلکه با مسائل سازمانی کاملاً مرتبط است.

در کل، چارچوب‌های مختلفی برای کیفیت داده‌ها ایجاد شده است، چارچوب وانگ و استرانگ (۱۹۹۶) که مورد توجه زیادی قرار گرفته بود، بر نیاز مصرف‌کنندگان داده تمرکز دارد. اپپلر و ویتیگ (۲۰۰۰) بررسی چارچوب‌های کیفیت داده را ارائه دادند و نتیجه گرفتند که چارچوب کیفیت داده باید دارای چهار قابلیت شامل ارزیابی، تجزیه و تحلیل، حل مسائل و مبنایی برای اندازه‌گیری و مدیریت پیشگیرانه باشد. با این حال، چارچوب‌های قبلی تا حدودی از دیدگاه عملکرد واقعی کیفیت داده‌ها چالش برانگیز بودند. اقدامات در مورد کیفیت داده باید نیازهای تجاری و انتظارات مشتری را برآورده کند تا پایه محکمی برای اندازه‌گیری کیفیت داده‌ها تشکیل شود.

کیفیت اطلاعات، یک چارچوب امروزی‌تری است که به طرق مختلف در سه دهه گذشته، مورد سنجش قرار گرفته است و یک عامل مهم در تعیین موفقیت سیستم اطلاعاتی، مد نظر قرار گرفته است. یکی از جامع‌ترین چارچوب‌ها برای طبقه‌بندی کیفیت داده‌ها، دوازده بعد کیفیت داده را در چهار بعد داده عمده، طبقه‌بندی می‌نماید. این چهار بعد عبارتند از کیفیت ذاتی، درستی، عینیت و قابلیت باورپذیری داده‌ها.

سامانه ثبت بمنظور گردآوری فراداده‌های پایان‌نامه/ رساله‌های دانشجویان داخل کشور به گونه‌ای طراحی شده است که متشکل از فیلدها یا اقلام داده‌ای مختلفی از جمله نام، نام خانوادگی، شماره دانشجویی، کد ملی، مقطع تحصیلی، رشته دانشگاهی و .. است که به هر رکورد (سابقه) فایل مربوط به پایان‌نامه/ رساله آن دانشجو الصاق گردیده است. مجموعه داده‌های مورد اشاره قرار گرفته یک فراداده در سامانه ثبت خوانده می‌شود. اندازه‌گیری کیفیت فراداده ثبت شده در سامانه ثبت در هر یک از ابعاد چهارگانه کیفیت ذاتی، درستی، عینیت و قابلیت باورپذیری داده‌ها مسئله‌ای است که در پژوهش حاضر مورد توجه قرار گرفته است. رویکرد کنترل فرایند آماری بمنظور اندازه‌گیری و تحلیل شاخص‌های کیفیت داده در فرایند ثبت به کار گرفته خواهد شد.

۲-۱- اهداف پژوهش

هدف کلی تحقیق حاضر عبارتست از اندازه‌گیری و تحلیل آماری فرایند ثبت پایان‌نامه/رساله دانش‌آموختگان داخل کشور در حوزه شاخص‌های کیفیت داده. سایر اهداف به شرح زیر هستند.

- بررسی کلیه روابط موجود در شاخص‌های کیفیت داده از متون علمی و انتخاب متناسب‌ترین رابطه‌ها در اندازه‌گیری قابلیت فرایند در سامانه ثبت پایان‌نامه/رساله دانش‌آموختگان داخل کشور
- تبیین شاخص‌های کیفیت داده در سامانه ثبت پایان‌نامه/رساله دانش‌آموختگان داخل کشور
- تعیین روابط، نحوه محاسبه و روش نمونه‌برداری بمنظور محاسبه شاخص‌های کیفیت داده در سامانه ثبت پایان‌نامه/رساله دانش‌آموختگان داخل کشور
- اندازه‌گیری و مطالعه مقایسه‌ای شاخص‌های کیفیت داده در سامانه‌های ثبت قدیم و جدید

۳-۱- پرسش‌ها/فرضیه‌های پژوهش

پرسش‌های این تحقیق عبارتند از:

- چگونه میتوان فرایند ثبت پایان‌نامه/رساله دانش‌آموختگان داخل کشور را از دیدگاه شاخص‌های کیفیت داده ارزیابی نمود؟ تحلیل‌های آماری مناسب در این رابطه چگونه است؟
- شاخص‌های کیفیت داده در متون علمی کدامند؟ کدامیک در سامانه ثبت پایان‌نامه/رساله دانش‌آموختگان داخل کشور قابلیت پیاده‌سازی دارند؟
- نحوه محاسبه و روش نمونه‌برداری بمنظور محاسبه شاخص‌های کیفیت داده در سامانه ثبت پایان‌نامه/رساله دانش‌آموختگان داخل کشور چگونه است؟
- میزان بهبود شاخص‌های کیفیت داده در سامانه‌های ثبت قدیم و جدید به چه میزان است؟

این تحقیق فرضیه‌ای ندارد.

۱-۴- معرفی پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

به دلیل ضرورت استفاده از پژوهش‌های انجام شده و معتبر برای علاقه‌مندان و دانشجویان و همچنین ضرورت استفاده از اطلاعات معتبر پژوهشگاه علوم و فناوری ایران با داشتن ۴۵ سال سابقه به عنوان سازمان متولی و نهاد اداره‌کننده سامانه ملی مدیریت اطلاعات تحقیقاتی تعیین شده است. پژوهشگاه وظایف خود را با همکاری و همراهی نزدیک ۱۶۰ استاد و کارمند به انجام می‌رساند. دستاوردهای مدیریت دانش و اطلاعات علمی و فنی کشور در سامانه‌ها و پایگاه‌های اطلاعات ارائه می‌شوند. در این میان، سامانه گنج با صدها هزار رکورد از بزرگ‌ترین، دیرین‌ترین و پربازدیدترین پایگاه‌های اطلاعات علمی و فنی کشور است. اطلاعات پایان‌نامه‌ها و رساله‌ها و پیشنهادیه آن‌ها که در ایرانداک به ثبت می‌رسند، بی‌همتا است. تولید ارزش افزوده و اطلاعات پشتیبان و همچنین تحلیل اطلاعات، بخشی دیگر از مدیریت دانش و سامانه‌های پیشینه پژوهش، همانندجو، دروازه اطلاعات علمی ایران، اصطلاح‌نامه، راهنمای پژوهش و پژوهشگران و نمای مقاله‌های ایرانیان در جهان نمونه‌هایی از آن‌ها هستند.

ایرانداک از سال ۱۳۶۱ دوره‌های آموزشی کوتاه‌مدت بسیاری را ارائه می‌کند و از سال ۱۳۹۰ نیز در سه رشته مدیریت فناوری اطلاعات، علم اطلاعات و دانش‌شناسی، و مهندسی فناوری اطلاعات دانشجوی دکترا پذیرفته است که رساله‌های خود را در زمینه‌های گوناگون پژوهشی انجام می‌دهند.

از سال ۱۳۴۸ همکاری‌های پژوهشی و اطلاع‌رسانی با اجرای طرح امانت میان کتابخانه‌ها در ایرانداک آغاز گردید. این طرح هم‌اکنون به نام امین خوانده می‌شود. از سال ۱۳۷۴ نیز طرحی با نام غدیر (عضویت فراگیر کتابخانه‌ها) در این زمینه به انجام می‌رسد. فراهم‌آوری منابع علمی خارجی برای دانشگاه‌ها و پژوهشگاه‌ها از سال ۱۳۸۷ در شورای تأمین منابع علمی در ایرانداک دنبال می‌شود. ایرانداک با نهادهای جامع اطلاع‌رسانی مانند «ایفلا» و «ترمنت» نیز همکاری دارد. سامانه عرضه و تقاضای پژوهش هم‌پیوندی است که ایرانداک برای پشتیبانی از همکاری‌های پژوهشی ساخته است.

سیاست‌گذاری علم و فناوری در کشور به اطلاعاتی نیاز دارد که فراهم‌آوری بخشی از آن را این پژوهشگاه در سامانه ملی اطلاعات تحقیقاتی (سمات ملی) و بخشی را نیز در داش‌بورد مدیران انجام می‌دهد.

دبیرخانه کمیسیون مدیریت اطلاعات علم و فناوری هم که در قلمرو وزارت عتف، سیاست‌گذاری می‌کند در این پژوهشگاه است. پژوهشگاه علوم و فناوری اطلاعات ایران نماینده ایران در شبکه اطلاع‌رسانی آسیا و اقیانوسیه: «اپین» نیز هست که از سوی برنامه اطلاعات برای همه یونسکو هدایت می‌شود.

تولید ارزش افزوده و اطلاعات پشتیبان و همچنین تحلیل اطلاعات، بخشی دیگر از مدیریت دانش و سامانه‌های پیشینه پژوهش، همانندجو، دروازه اطلاعات علمی ایران، اصطلاح‌نامه، راهنمای پژوهش و پژوهشگران، و نمای مقاله‌های ایرانیان در جهان نمونه‌هایی از آن‌ها هستند.

۱-۵- معرفی مرکز اطلاعات و مدارک علمی ایران

ماموریت‌های اصلی مرکز اطلاعات و مدارک علمی ایران فراهم‌آوری، سازمان‌دهی، ذخیره، حفظ، بازیابی، تحلیل و اشاعه اطلاعات و مدارک علمی و فناوری کشور و اطلاعات پشتیبان آن‌ها در سطح ملی و بین‌المللی، مدیریت تامین و دسترسی پژوهشگران کشور به مدارک اطلاعات علم و فناوری خارج از کشور و پشتیبانی و مدیریت سامانه ملی مدیریت اطلاعات علم و فناوری، می‌باشند.

این مرکز دارای سه بخش می‌باشد:

- گروه ارتباطات و فراهم‌آوری
- مدیریت سازماندهی و تحلیل اطلاعات
- مدیریت حفظ و اشاعه اطلاعات

• گروه ارتباطات و فراهم‌آوری

این واحد مرحله ورودی و آغازین ورود منابع اطلاعاتی به مرکز اطلاعات می‌باشد ماموریت‌های اصلی گروه ارتباطات و فراهم‌آوری را «فراهم‌آوری اطلاعات و مدارک علم و فناوری» و «ثبت اطلاعات علم و فناوری کشور» تشکیل می‌دهد.

گروه ارتباطات و فراهم‌آوری اطلاعات از بخشهای زیر تشکیل یافته است:

- تهیه مدرک
- ورود اطلاعات
- کنترل اطلاعات
- دیجیتال سازی (اسکن)
- چاپخانه
- مخزن

• مدیریت سازماندهی و تحلیل اطلاعات

مأموریت مدیریت سازماندهی و تحلیل اطلاعات شامل: تحقیق و بررسی روش‌های سازماندهی و طبقه‌بندی مدارک و اطلاعات، سازماندهی مدارک به منظور پردازش، تحلیل و بازیابی اطلاعات، انتخاب یکی از روش‌های نمایه‌سازی و کنترل واژگان با اصطلاحنامه‌ها می‌باشد.

اهم وظایف:

- بررسی انواع مدارک و نمایه‌سازی آنها که از بخش تهیه مدارک ارسال می‌گردد.
- استفاده از اصطلاحنامه‌ها جهت انتخاب اصطلاحات مرجع و همچنین یکدست نمودن کلیدواژه‌ها.
- تلاش در جهت استانداردسازی نمایه‌سازی و استفاده از دستورالعمل‌های نمایه‌سازی

خدمات مدیریت ورود سازماندهی و تحلیل اطلاعات :

- گردآوری، سازماندهی، ذخیره، حفظ، بازیابی، تحلیل و اشاعه مدارک و اطلاعات علمی و فنی کشور از جمله پایان‌نامه‌های کارشناسی ارشد و رساله‌های دکترای تخصصی دانش‌آموختگان داخل و دانش‌آموختگان ایرانی خارج از کشور، مقاله‌های سمینارها، مشخصات پژوهشگران و متخصصان، طرحهای پژوهشی، اطلاعات مربوط به اختراعات، نرم افزارهای علمی و همچنین اطلاعات پشتیبان آنها از جمله اطلاعات سازمانها و افراد.

- مدیریت دسترسی به مدارک و اطلاعات علمی و فنی بین المللی مورد نیاز پژوهشگران.
- مدیریت دسترسی به مدارک و اطلاعات علمی و فنی گردآوری شده در پژوهشگاه در سطح ملی و جهانی.
- انتشار نمایه‌نامه‌ها، چکیده‌نامه‌ها، آمارنامه‌ها، فرهنگ‌نامه‌ها، و اصطلاح‌نامه‌ها.

• مدیریت حفظ و اشاعه اطلاعات

- سه مأموریت اصلی برای گروه حفظ و اشاعه اطلاعات تعریف شده است:
- نگهداری و حفظ اطلاعات و مدارک علم و فناوری ورودی به پژوهشگاه و سازمان‌دهی شده در پژوهشگاه؛
 - بازیابی و اشاعه اطلاعات و مدارک سازمان‌دهی شده؛
 - ایجاد دسترسی کاربران به اصل مدارک؛
 - اشاعه اطلاعات و تحویل نسخه‌های تکثیرشده مدارک در قالبها و از روشهای ممکن متناسب با زیرساختهای اطلاعاتی و ارتباطی کشور و نیاز کاربران.
- این واحد شامل سه بخش سالن جستجوی حضوری، سفارش مدرک غیرحضوری، گزارش پیشنهاد می باشد.

۱-۶- معرفی سامانه ثبت پایان‌نامه/رساله دانش‌آموختگان کل کشور

مطالعه‌ی موردی پژوهش پیش‌رو سامانه‌ی ثبت پایان‌نامه‌های دانش‌آموختگان داخل کشور می‌باشد. طبق آیین-نامه‌ای که زیر نظر وزارت علوم، تحقیقات و فن‌آوری صادر شد، از همه‌ی دانشگاهها، پژوهشگاهها و ... خواسته شده تا فایل تمامی مدارک را در پژوهشگاه علوم و فن‌آوری اطلاعات ایران (ایرانداک) به ثبت برسانند. لذا برای این فرایند، ایرانداک سامانه ثبت پایان‌نامه/رساله، پروپوزال را در اختیار دانشجویان گذاشته است. این سامانه با هدف جمع‌آوری فراداده‌های پایان‌نامه/رساله دانشجویان داخل کشور طراحی شده است و شامل فیلدها یا اقلام اطلاعاتی مختلفی، نظیر نام، نام خانوادگی، شماره دانشجویی، کد ملی و ... است. و به هر رکورد (سابقه)، فایل مربوط به پایان‌نامه/رساله هر دانشجو قرار گرفته است. داده‌های وارد شده در سامانه، مجموعه‌ای را به وجود می‌آورند که به آن فراداده در سامانه ثبت گویند. بررسی هر کدام از فیلدهای ذکر شده در فراداده و تعیین سطح کیفیت آنها مانند دقت، صحت، نامتناقض بودن، کامل بودن و ... یکی از مهمترین اقدامات در حوزه کیفیت سامانه ثبت محسوب می‌شود. به علاوه در صورتی که در فراداده‌های تکمیل شده توسط دانشجویان تحصیلات تکمیلی نقص یا به عبارتی مشکل کیفی مشاهده گردد عدم انطباق رخ داده بایستی به شکل مناسب درجه‌بندی شده و تصمیم‌گیری مناسب در آن خصوص صورت گیرد.

در این پژوهش فرایند ثبت پایان‌نامه/رساله‌های دانش‌آموختگان کل کشور مورد توجه قرار خواهد گرفت. شاخص‌هایی بمنظور ارزیابی کیفیت داده‌های حاصل از این فرایند توسعه داده خواهد شد و همچنین درجه اهمیت ابعاد کیفیت این داده‌ها در قالب یک مدل تصمیم‌گیری با معیارهای چندگانه تعیین خواهد شد.

فصل دوم –

مبنای نظری تحقیق

۲-۱ مقدمه

داده‌ها یک دارایی اطلاعاتی اساسی هستند و تمرکز اصلی امنیت سیستم‌های اطلاعاتی، وابسته به داده‌ها است. (دهیلون و بکهوس^۱، ۲۰۰۰) از اینرو امروزه کیفیت داده‌ها و فراداده‌هایی که به کار گرفته می‌شوند در فرایند آن کسب و کار تاثیر حائز اهمیتی می‌گذارند و کیفیت داده نقش بسیار مهمی را در بسیاری از سازمان‌ها ایفا می‌کند. داده‌ها در سازمانها نیز منبع حیاتی برای حمایت از فرایند کسب و کار و تصمیم‌گیری مدیریتی، می‌باشند. (خسروانجوم و همکاران^۲، ۲۰۱۱) رشد انبار داده‌ها و دسترسی مستقیم به اطلاعات از منابع مختلف توسط مدیران و کاربران اطلاعات، نیاز و همچنین آگاهی از کیفیت بالای داده‌ها را در سازمانها افزایش داده است. (لی و همکاران^۳، ۲۰۰۲)

همانطور که تولید یک محصول را میتوان به صورت سیستم پردازشی که بر روی مواد خام فعالیت می‌کند تا آنها را به محصولات فیزیکی تبدیل کند، توصیف کرد؛ تولید اطلاعات را نیز میتوان به صورت یک سیستم پردازشی که بر روی داده‌های خام فعالیت میکند تا آنها را به محصولات اطلاعاتی تبدیل کند، تعریف کرد (جدول ۱-۲). (وانگ^۴، ۱۹۹۸)

جدول ۱-۲ مقایسه تولید محصول و اطلاعات (وانگ، ۱۹۹۸)

تولید اطلاعات	تولید محصول	
داده خام	مواد خام	ورودی
سیستم اطلاعاتی	خط تولید	پردازش
محصولات اطلاعاتی	محصولات فیزیکی	خروجی

¹ Dhillon & Backhouse

^۲ Khosroanjom et al.

^۳ Lee et al.

^۴ Wang

با پیشرفت تکنولوژی و فن‌آوری اطلاعات و تغییر سیستم‌های اطلاعاتی از ساختار سلسله‌مراتبی به ساختار مبتنی بر شبکه، با حجم بزرگی از داده‌ها مواجه هستیم. بدین ترتیب با افزایش حجم داده‌ها از نظر اندازه و دامنه در سالهای اخیر مدیریت آنها پیچیده‌تر شده و خطر کاهش کیفیت داده‌ها رو به افزایش است. داده‌های نامرغوب نیز می‌توانند بر قابلیت سیستم و همچنین تضمین عملکرد عملیاتی زیان وارد کنند.^۱ (ردمن، ۱۹۹۸) از جهت دیگر داده‌های نامرغوب می‌توانند اثرات اجتماعی و اقتصادی قابل توجهی بگذارند. (وانگ و استرانگ^۲، ۱۹۹۶) همچنین بر رضایت مشتریان و روحیه کارکنان و اطمینان میان سازمانها تاثیر گذار است. (دهیلون و بکهوس، ۲۰۰۰) در سیستم‌های اطلاعات شبکه‌ای، اگر اطلاعات ورودی و کیفیت فرایندها کنترل نشده باشد، می‌تواند بر روی کیفیت کلی داده‌هایی که در سیستم اطلاعات وجود دارد تاثیر گذاشته و به مرور زمان کیفیت داده‌ها را کاهش دهد.

۲-۲ کیفیت داده‌ها

به علت پیشرفت داده‌های الکترونیکی، کیفیت داده‌ها نقش مهمی در تمامی کسب‌وکارها و برنامه‌های دولتی ایفا می‌کند. کیفیت داده‌ها به عنوان یک مسئله عملکردی و ضروری در ارتباط با فرآیندهای عملیاتی (موسسه داده انبارداری^۳، ۲۰۰۶)، تصمیم‌گیری‌ها (چنگالور اسمیث و همکاران^۴، ۱۹۹۹) و همکاری‌های درون سازمانی (اسکاناپیکو و بتینی^۵، ۲۰۰۲) شناخته شده است. چندین طرح ابتکاری در بخش‌های دولتی و خصوصی راه‌اندازی شده است که در آن کیفیت داده‌ها نقش برجسته‌ای دارد، از جمله این طرح‌ها می‌توان به تصویب قانون کیفیت داده‌ها توسط دولت ایالات متحده در سال ۲۰۰۲ (دفتر مدیریت و بودجه^۶، ۲۰۰۶)

^۱ Redman

^۲ Wang & Strong

^۳ Data Warehousing Institute

^۴ Chengalur-Smith et al.

^۵ Scannapieco & Batini

^۶ Office of Management and Budget

و تصویب چارچوب کیفیت داده‌ها توسط دولت Wales در سال ۲۰۰۴ اشاره کرد، که هدف آن‌ها بهبود کیفیت اطلاعات برای شیوه‌های پزشکی عمومی بود. (DQI^۱ ۲۰۰۴).

با حرکت سیستم‌های اطلاعاتی از ساختار سلسله‌مراتبی / یکپارچه به ساختار مبتنی بر شبکه، مجموعه منابع داده‌ی بالقوه مورد استفاده در سازمان‌ها، از نظر اندازه و دامنه به‌طور چشمگیری افزایش یافت. مسئله کیفیت داده‌ها به دلیل تاثیرپذیری از این تکامل، پیچیده‌تر و بحث‌برانگیزتر شده است. در سیستم‌های اطلاعات شبکه‌ای، فرآیندها در تبادل اطلاعات پیچیده‌ای درگیر شده و اغلب بر اساس ورودی به‌دست‌آمده از منابع خارجی عمل می‌کنند، که این ورودی‌ها نیز اغلب قابل مقایسه نیستند.

همانطور که گفته شد، فاکتورهای بسیاری بر فرایندهای تصمیم‌گیری و نتایج آنها تاثیرگذار است و داده‌های نامرغوب، یکی از آنهاست، که باعث تصمیمات ضعیف می‌شوند. مسئله کیفیت داده^۲، بدلیل وجود حجم بالای داده‌های در دسترس، یک مسئله‌ی مهم در سازمانها است. تحقیقات حاکی از آن است که تصمیم‌گیرندگانی که با DQ آشنایی داشته‌اند از آن بیشتر برای حل در تصمیم‌گیری استفاده کرده و همچنین دقت بالاتری در تصمیمات خود داشته‌اند. (مोगس و همکاران^۳، ۲۰۱۶)

کیفیت داده‌ها خود داده‌ای است، که برای مصرف‌کنندگان داده مناسب است. علاوه بر این ابعاد کیفیت داده‌ها را به عنوان مجموعه‌ای از ویژگی‌ها که نشان دهنده‌ی یک جنبه و یا ساختار DQ است، در نظر گرفته شده است. (وانگ و استرانگ، ۱۹۹۶) مطالعات حاکی از آن است که کیفیت داده‌ها یک مفهوم چند بعدی می‌باشد. (پپینو و همکاران^۴، ۲۰۰۲) از اینرو برای درک بهتر، اندازه‌گیری و بهبود DQ ابعادی مشخص شده است (وزیری و همکاران، ۲۰۱۷). در این تحقیق از ابعاد کیفیت داده که توسط وانگ و استرانگ در سال ۱۹۹۶ عنوان شده است، استفاده شده است. این ابعاد شامل ۱۵ بعد می‌شوند که در ۴ دسته، طبقه‌بندی شده‌اند و این دسته‌بندی به صورت زیر است:

1 Data Quality Information

2 Data Quality (DQ)

3 Moges et al.

4 Pipino et al.

ذاتی

دسترس پذیری

متنی

نمایشگری

ابعاد نمایش سازگاری و کامل بودن، دو بعد مهم در میان ابعاد کیفیت داده‌ها به حساب می‌آیند و از شروط حیاتی برای زیربنای فناوری اطلاعات و منبع کسب و کار بزرگ و همچنین دارایی بسیار مهمی برای دستیابی به مزایای رقابتی بلند مدت است. (کوان و همکاران^۱، ۲۰۱۴)

برای اندازه‌گیری ابعاد کیفیت داده‌ها، ابزارها و یا معیارهایی مشخص شده است، که این معیارها یا به صورت ذهنی^۲ و یا به صورت عینی^۳ هستند. معیارهای ذهنی بر اساس نظرات و تجربیات کاربران داده که می‌توانند مدیران و یا متخصصان داده باشند، مشخص می‌شوند. سنجش ذهنی، معمولاً با استفاده از مصاحبه‌هایی که مربوط به داده‌ها و یا پرسشنامه‌ها است، صورت می‌پذیرد. که منعکس کننده‌ی نیازها و تجارب ذینفعان از جمله گردآوردندگان، نگهبانان و مصرف‌کنندگان داده می‌باشد. از طرف دیگر سنجش عینی بر مبنای فرمول‌های ریاضی است که برای اندازه‌گیری یک مجموعه داده استفاده می‌شود. (وزیری و همکاران، ۲۰۱۷)

۲-۳ مدیریت کیفیت جامع داده‌ها

همانطور که در بخش‌های قبلی نیز عنوان گردید در تضمین کیفیت سامانه‌های اطلاعاتی از منظر فرایندی اقدامات اجرایی لازم در خصوص پیشگیری از وقوع مشکلات کیفی و نیز حل ریشه‌ای عدم انطباق‌ها تعیین خواهد شد. اما از آنجاکه اغلب مشکلات مطرح شده در این حوزه به کیفیت داده‌های موجود در سامانه‌های اطلاعاتی مربوط می‌گردد لذا در این بخش به بررسی در حوزه تضمین کیفیت سامانه‌های اطلاعاتی با تمرکز بر بهبود کیفیت داده‌ها خواهیم پرداخت.

^۱ Kwon et al.

^۲ Subjective

^۳ Objective

از آنجاکه علاوه بر توجه به کیفیت داده و ابعاد آن لازم است مدل‌های مربوط به مدیریت کیفیت داده نیز مورد توجه قرار گیرند در ادامه به معرفی مدل^۱ TDQM در این حوزه خواهیم پرداخت.

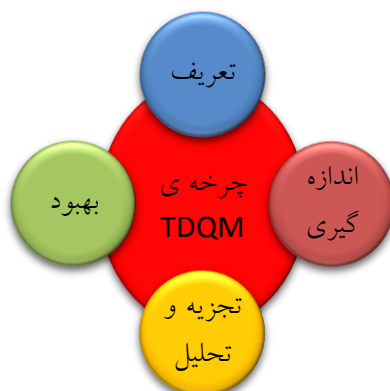
نخستین پژوهش‌ها در حوزه مدیریت کیفیت جامع داده‌ها (TDQM) به دهه ۱۹۸۰ میلادی برمی‌گردد. تحقیقات اولیه در حوزه کیفیت داده‌ها به طور عمده بر توسعه تکنیک‌ها برای پرس و جو از منابع اطلاعاتی متعدد و ایجاد مخازن اطلاعاتی بزرگ متمرکز بود. تحقیق وانگ و مادنیگ (۱۹۸۹) رویکرد سیستماتیک را برای بررسی کیفیت داده‌های مرتبط بکار می‌گیرد. (وانگ و مادنیگ، ۱۹۸۹) این کار بیشتر هنگام تفکیک هویت اطلاعات و یکپارچه سازی اطلاعات از منابع مختلف و با حضور رکوردهایی که تداخل و همپوشانی دارند، شناخته شده و برجسته است. این محققان برای تعیین این که آیا رکوردهای جدا شده واقعا هویت‌های مرتبطی دارند، روش‌هایی را یافتند. این مسئله با واژه‌هایی مانند اتصال رکورد، غیرتکراری کردن رکورد و تطبیق رکورد، شناخته می‌شود. پس از آن، وانگ و مادنیگ (۱۹۹۰) یک مدل پلی‌ژن^۲ (پلی برای متعدد و ژن برای منبع) را توسعه دادند (وانگ و مادنیگ، ۱۹۹۰) تا پردازش برچسبهای منبع داده را در پردازنده پرس و جو زیر نظر بگیرند و به این طریق جوابی برای پرسشهای مرتبط با کیفیت داده‌ها بیابند، پرسش‌هایی مانند: "این داده‌ها از کجا آمده‌اند؟" و "کدام منابع داده‌ای میانجی برای رسیدن به این داده‌ها مورد استفاده قرار گرفته است؟"

تلاش‌های تحقیقاتی اولیه، محققان را برای آغاز تحقیق روشمند در تمام گستره مسائل کیفیت داده‌ها تشویق کرد که به نوبه خود منجر به افتتاح برنامه مدیریت کیفیت داده جامع (TDQM) در اوایل ۱۹۹۰ در دانشگاه MIT شد. تحقیق اولیه در برنامه توسعه چارچوب TDQM بهبود کیفیت مستمر داده‌ها را با دنبال کردن مراحل تعریف، اندازه‌گیری، تجزیه و تحلیل و بهبود حمایت می‌کند (وانگ و مادنیگ، ۱۹۹۲). این مدل، چهارچوب مدیریت کیفیت جامع^۳ (TQM) را برای بهبود کیفیت در حوزه تولید به حوزه داده توسعه می‌دهد. (شکل ۱-۲)

^۱ Total Data Quality Management

^۲ PolyGen

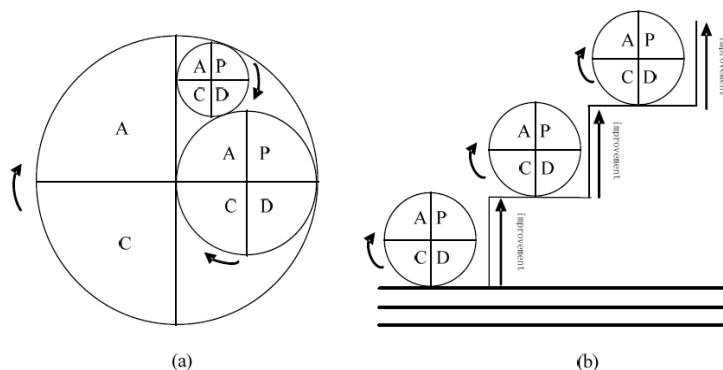
^۳ Total Quality Management



شکل ۱-۲ چرخه ی TDQM (وانگ و همکاران، ۲۰۰۱)

دمینگ در سال ۱۹۸۲ (وانگ و استرانگ، ۱۹۹۶) و جوران و جوفری در سال ۱۹۹۹ از مهمترین پایه‌گذاران و توسعه دهندگان حوزه TQM بودند. بینش کلیدی این بود که اگر چه داده، در واقع، یک محصول تولید شده (یا محصول جانبی) توسط بسیاری از سازمان‌ها است، ولی آن طور که باید با آن رفتار نشده و مورد مطالعه قرار نگرفته است.

مدل TDQM به لحاظ ساختاری و مدیریتی تشابه زیادی به مدل بهبود مستمر دمینگ دارد که در شکل ۲-۲ بصورت شماتیک به آن اشاره شده است.



شکل ۲-۲ معرفی چرخه دمینگ

در هر ۴ مرحله موجود در چرخه دمینگ یعنی: برنامه‌ریزی^۱، اجرا^۲، کنترل^۳ و اقدام^۴ لازم است سازمان اقدامات لازم برای تضمین کیفیت خدمات خود و کاهش خطا در فرآیندهای تولید داده و اطلاعات را انجام دهد. همانگونه که در شکل ۲-۲ مشاهده می‌شود، با کاربرد مداوم چرخه دمینگ بهبود مرحله به مرحله در سامانه یا فرایند مورد کاربرد رخ خواهد داد. و در ادامه به پژوهش‌های صورت گرفته در حوزه نظریه‌ها، روش‌ها، و تکنیک‌ها برای چهار مرحله از چارچوب TDQM توسعه خواهیم پرداخت.

۳-۱-۲ تعریف مراحل

تعریف کیفیت داده‌ها از نقطه نظر مصرف کننده، دستیابی به یک موفقیت بزرگ می باشد. مطابق با این تعریف تناسب بین استفاده و شناسایی ابعاد کیفیت داده‌ها از طریق یک مطالعه پیمایشی چند مرحله‌ای روشمند عملی شد (وانگ و استرانگ، ۱۹۹۶). ابعاد کلیدی کیفیت داده‌ها با استفاده از تحلیل عاملی بر روی بیش از ۱۰۰ نشانه مشخص کیفیت داده‌ها از طریق مطالعه پیمایشی کشف شد. این ابعاد به صورت چهار طبقه کیفیت داده یعنی دسترس پذیری، محتوایی، نمایشگری و ذاتی سازمان‌دهی شده‌اند. در ادامه و در زیربخش بعدی به مرحله بعدی از مدل TDQM یعنی اندازه‌گیری اشاره خواهیم کرد.

۳-۲-۲ اندازه‌گیری

لی و همکاران در سال ۲۰۰۲، ارزیابی و اندازه‌گیری کیفیت داده‌ها که دومین مرحله از رویکرد مدیریت کیفیت جامع است را مشابه شرایط عملیاتی به منظور اندازه‌گیری کیفیت داده‌ها در سازمانها توسعه دادند. این ابزار هر یک از ابعاد را به چهار یا پنج شاخص قابل اندازه‌گیری تقلیل می‌دهد و صورتهای کاربردی

Plan ۱

Do ۲

Check ۳

Act ۴

مناسب برای این آیتم‌ها بکار گرفته می‌شود تا هر یک از ابعاد را ارزیابی نمایند. این ابزار همچنین، این قابلیت را دارد که با نیازهای خاص هر سازمانی منطبق گردد.

۳,۳-۲ تجزیه و تحلیل

نتایج اندازه‌گیری در این مرحله تفسیر می‌گردد. تکنیک‌های تجزیه و تحلیل (لی و همکاران، ۲۰۰۲) تفاوت‌های ادراکی بین ابعاد داده‌ها و نقش‌های از قبل تعیین شده در حوزه مربوط به کیفیت داده را نشان می‌دهد (استرانگ و همکاران، ۱۹۹۷). لی و استرانگ در سال ۲۰۰۳ جمع‌آوری کننده‌های اطلاعات، متولیان داده‌ها، و مصرف کنندگان داده که سه نقش عمده در این حوزه دارا هستند را معرفی نمودند. تجزیه و تحلیل‌های صورت گرفته در این مرحله، ابعادی را که بیشتر نیاز به بهبود دارند را شناسایی می‌کند و به تحلیل ریشه مشکلات کیفیت داده، می‌پردازد.

۳,۴-۲ بهبود

در این مرحله، اقدامات صورت گرفته در راستای تغییر ارزش داده‌ها و هم بمنظور تغییر فرآیندهایی است که داده تولید می‌کنند. روش دوم همان طور که در مطالعات بالو و همکاران^۱ و وانگ و همکاران در سال ۱۹۹۸ مورد بحث قرار داده‌اند موثرتر است. در این روش مراحل به سوی مدیریت اطلاعات به عنوان یک محصول فراهم شده است. علاوه بر این، فناوری‌های قبلی مانند پلی‌ژن و مدل رابطه هویت کیفیت، به عنوان بخشی از فرآیند بهبود مستمر اعمال می‌شود.

علاوه بر توسعه و غنی سازی چارچوب TDQM، برنامه TDQM در جهت یکپارچه سازی حوزه کیفیت داده، تلاش‌های چشمگیری را انجام داده است. همچنین باعث گسترش تأثیر پژوهش و ترویج همکاری دولت با دانشگاه و صنعت از طریق نشریات، سمینارها، دوره های آموزشی و کنفرانس های بین‌المللی سالانه در حوزه کیفیت اطلاعات (ICIQ^۲)، از سال ۱۹۹۶ به بعد شده است. دانشگاه آرکانزاس در شهر

^۱ Ballou et al.

لیتل راک با کمک برنامه TDQM، اولین بار از نوع خود در ایالات متحده، دوره کارشناسی ارشد و دکتری کیفیت داده ایجاد کرده است. این دوره‌ها به منظور پاسخگویی به افزایش تقاضا و نیاز به متخصصان کیفیت داده آموزش دیده در سطح بالا و نیز آماده سازی دانشجویان برای تحقیقات پیشرفته در زمینه کیفیت داده، ایجاد شده است. در بخش صنعت و در حوزه کیفیت داده‌ها، فروشندگان اصلی نرم افزار شروع به پیاده سازی فناوری‌های کیفیت داده در ارائه محصول و خدمات خود کرده‌اند. در بخش دولتی، کیفیت داده به یک جزء مهم در بسیاری از بخش‌های دولت الکترونیک و طرح‌های معماری سازمانی تبدیل شده است. در این میان، تحقیقات کیفیت داده‌ها با چالش‌های جدیدی مواجه است که مربوط به محیط‌های کسب و کار همیشه در حال تغییر، الزامات قانونی، انواع افزایش فرم‌های داده‌ها / رسانه‌ها، و فناوری‌های اینترنت و این که اساساً چگونه اطلاعات تولید، ذخیره، دستکاری، و مصرف می‌شود. تحقیقات کیفیت داده‌ها که از دو دهه پیش شروع شده، وارد عصر جدیدی شده است که در آن تعداد فزاینده‌ای از محققان به طور فعال به منظور افزایش درک از مشکلات کیفیت داده و توسعه راه حل برای مسائل نوظهور در این حوزه تلاش می‌کنند.

وانگ و همکاران چارچوب مدونی بر مبنای الزامات استاندارد ISO 9001 بمنظور دسته‌بندی تحقیقات صورت گرفته روی مبحث کیفیت داده ارائه دادند. چارچوب دسته‌بندی تحقیقات توسط آنها شامل گروهبندی زیر است. (وانگ و همکاران، ۱۹۹۵)

مسئولیت مدیریت

هزینه‌های تضمین و عملیات

تحقیق و توسعه

تولید

توزیع

مدیریت پرسنل

عملکرد قانونی

نلسون و همکاران^۱ (نلسون و همکاران، ۲۰۰۵) عواملی را که به طور همزمان بر کیفیت سیستم و اطلاعات تاثیرگذار هستند را مورد بررسی قرار داده و نتایج را در قالب مطالعه موردی روی سازمان‌هایی که انباشت داده دارند مورد تحلیل قرار دادند. بیاگی و همکاران^۲ در مرکز تحقیقات کشور ایتالیا کنترل کیفیت در سامانه‌های اطلاعاتی تحقیقاتی و عوامل موثر بر آن را مورد بررسی قرار داده‌اند. (بیاگی و همکاران، ۲۰۱۴)

۲-۴ معرفی مدل اصلی کیفیت داده

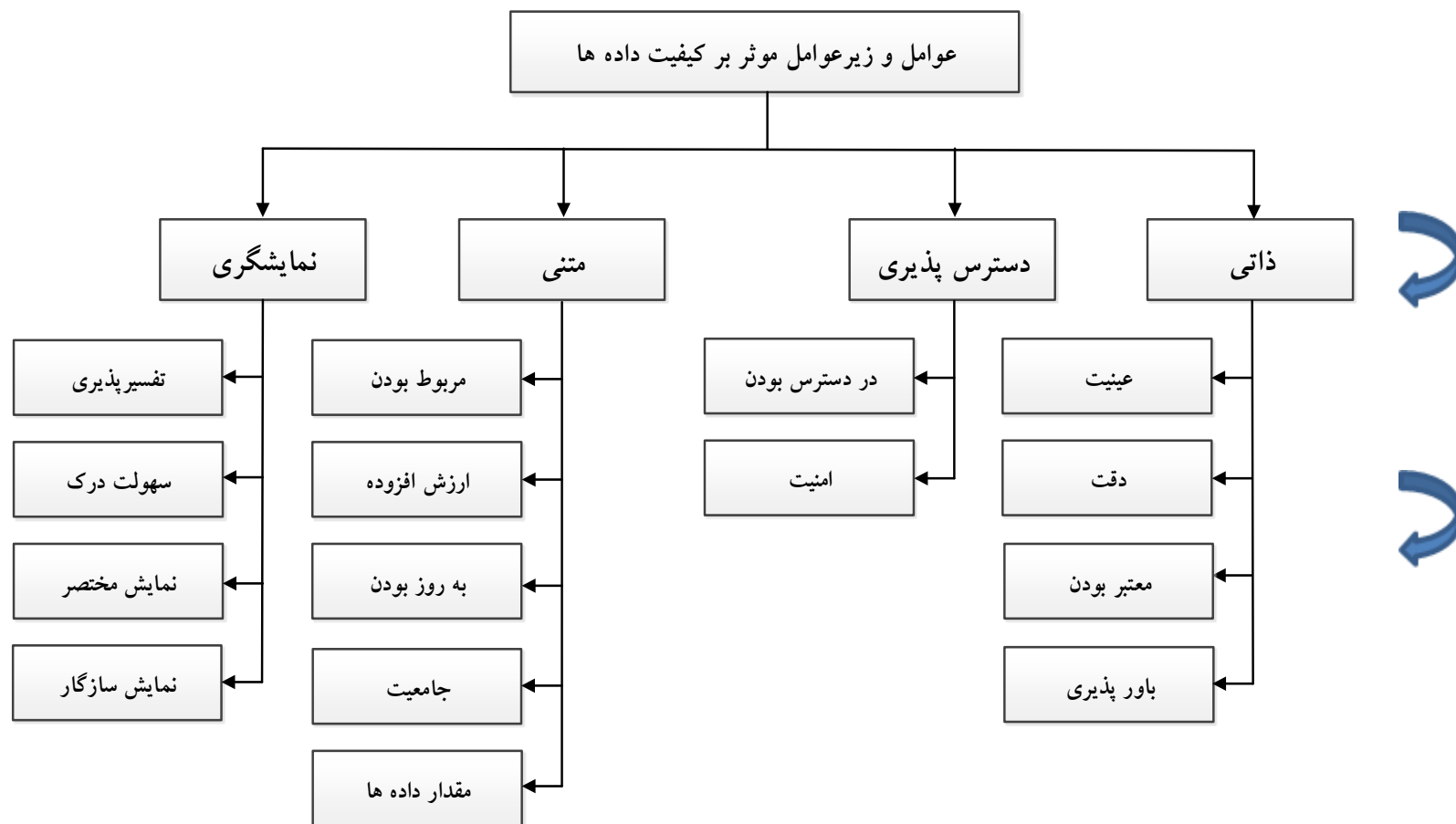
لی (۲۰۰۷)، به مرور کلی مسائل کیفیت داده‌ها پرداخته و مفهوم کیفیت داده‌ها را از دیدگاه داده‌پردازی بررسی می‌کند. طبق تحلیل نظام مند، در می‌یابیم که تعریف مجدد کیفیت داده‌ها برای نیازهای داده‌پردازی کفایت نمی‌کند. طبق نظریه گسترش، مشاوره داده‌پردازی ما راه حل جدیدی برای بهبود کیفیت داده‌ها از مراحل ابتدایی از جمله تحلیل شرایط، طراحی نرم افزار، اجرای نرم افزار و ادغام داده‌ها به واسطه مشاوره داده‌پردازی فراهم می‌کند، برخی شرکت‌های کوچک و متوسط دارای کیفیت داده ضعیف اند یا حتی فاقد سیستم اطلاعات اند که می‌توان راه حل ویژه در اختیار آنها نهاد تا به پروژه‌های داده‌پردازی مبادرت ورزند. با این وجود، هنوز برخی محدودیت‌ها درباره اجرای مشاوره داده‌پردازی وجود دارد. هنوز کار زیادی باید انجام داد. برای نمونه تیم‌های مشاور داده‌پردازی نیاز به کارشناسان متبحر در کسب و کار و داده‌پردازی دارند. کیفیت داده بستگی زیادی به سازماندهی سیستم اطلاعاتی و چگونگی پردازش داده دارد و اندازه‌گیری و بهبود کیفیت داده در سازمانها وظیفه‌ای پیچیده است. از اینرو برای بررسی کیفیت فراداده‌ها، از ابعادی که توسط وانگ و استرانگ در سال ۱۹۹۸ ذکر شده است استفاده شده

^۱ Nelson et al.

^۲ Biagi et al.

است. این ابعاد عبارتند از: دقت، عینیت، باورپذیری، معتبر بودن، مربوط بودن، ارزش افزوده، بروز بودن، جامعیت، مقدار داده‌ها، تفسیرپذیری، سهولت درک، نمایش مختصر، نمایش سازگار، در دسترس بودن و امنیت. که این ابعاد به طور مشخص در ۴ دسته با نام‌های ذاتی، متنی، نمایشگری و دسترس‌پذیری قرار می‌گیرند. که این دسته بندی در شکل ۲-۳ قابل رویت است.

به طور مشخص در این تحقیق سامانه‌ی ملی ثبت پایان‌نامه/ رساله‌های دانش‌آموختگان داخل کشور مورد ارزیابی قرار خواهد گرفت. ارشادی و همکاران (۱۳۹۶) ابعاد پیش‌گفته را در قالب طرح کیفی بمنظور تعیین وضعیت مطلوب اقلام اطلاعاتی سامانه ثبت به کار بردند. در طرح حاضر برپایه این کار، اندازه‌گیری‌ها صورت خواهد پذیرفت.



شکل ۲-۳ معرفی ابعاد مختلف کیفیت داده (لی ۲۰۰۷)

۲-۵ ضرورت معرفی شاخص کلیدی عملکرد در ارزیابی کیفیت داده‌ها

برای اندازه‌گیری ابعاد کیفیت داده‌ها، ابزارها و یا معیارهایی مشخص شده است، که این معیارها یا به صورت ذهنی^۱ و یا به صورت عینی^۲ هستند. معیارهای ذهنی بر اساس نظرات و تجربیات کاربران داده که می‌توانند مدیران و یا متخصصان داده باشند، مشخص می‌شوند. سنجش ذهنی، معمولاً با استفاده از مصاحبه‌هایی که مربوط به داده‌ها و یا پرسشنامه‌ها است، صورت می‌پذیرد. که منعکس کننده‌ی نیازها و تجارب ذینفعان از جمله گردآورندگان، نگهبانان و مصرف‌کنندگان داده می‌باشد. از طرف دیگر سنجش عینی بر مبنای فرمول‌های ریاضی است که برای اندازه‌گیری یک مجموعه داده استفاده می‌شود. (وزیری و همکاران، ۲۰۱۷) از آنجاییکه در کیفیت داده بعنوان یکی از شاخص‌های کلیدی عملکرد (Key Performance Indicators) در سازمان عمل کنند در ادامه به تحقیقاتی خواهیم پرداخت که KPIها در سازمان تحلیل آماری شده‌اند.

شاخص‌ها برای ارزیابی، نظارت و حفظ کنترل کیفیت فراوان است، و بسیاری از آنها به نوعی برای جمع‌بندی داده‌های جمع‌آوری شده مورد استفاده قرار می‌گیرند. تکنیک‌های SPC^۳ به طور خاص برای اندازه‌گیری تغییرات در گذر زمان در یک فرایند طراحی شده است در حالی که تکنیک‌های SQC^۴ یک طبقه کلی‌تر از ابزار آماری برای اندازه‌گیری هستند و بهبود روش‌های عملیاتی را نشان می‌دهند. هدف از کنترل آماری کیفیت محصول، به جای انجام بازرسی ۱۰۰٪، یافتن تعادل اقتصادی بین تلاش‌های انجام شده در بازرنگری کنترل کیفیت و احتمال یافتن محصول معیوب است. بازرسی ۱۰۰٪ از تولید، کار و هزینه فشرده‌ای نیاز دارد. نمودارهای کنترل و ابزارهای کنترل کیفیت آماری مرتبط با استفاده از ایجاد تعادل و پایداری اقتصادی به ما کمک می‌کنند. نمودار کنترل بطور ویژه راهنمای یک مدیر یا سرپرست در تصمیم‌گیری در مورد نحوه مداخله در یک پروسه جمع‌آوری اطلاعات فعال است. بنابراین با صرفه جویی در زمان هشدارهای لازم را ارائه می‌دهد. از سوی دیگر ارزیابی عملکرد یک سازمان به کمک شاخص‌های کلیدی عملکرد (KPI^۵) انتخاب شده، می‌تواند شروع شود. بکمک اطلاعات تولید شده توسط نمودار کنترل و ابزارهای مرتبط، و پیشنهاد چگونگی سیستم‌های نظارت در قالب KPI ممکن است در محیط‌های جمع‌آوری داده‌های بزرگ ساختار یافته و اجرا شوند. (جانز و همکاران، ۲۰۱۳) همچنین کنترل فرآیند آماری یا SPC یک روش برای ترسیم فرآیند است و به سرعت زمانی که یک فرایند خارج از کنترل است را تعیین

^۱ Subjective

^۲ Objective

^۳ Statistical Process Control

^۴ Statistical Quality Control

^۵ Key Performance Indicators

می‌کند و می‌تواند علت یا علل ریشه‌ای مشکلات فرایند یا بررسی کند و یک استراتژی برای اصلاح مشکل تعیین کند. با استفاده از SPC، یک فرآیند را می‌توان از طریق نمونه برداری و از نتایج نمونه‌گیری، تنظیم (ها) را می‌توان به روند قبل از فرایند خارج از کنترل ساخته شده درآورد. در قلب SPC، مفهوم تغییرات وجود دارد. تغییرات می‌تواند در یکی از دو راه 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. KPI نه تنها برای نظارت بر فرآیند تولید و ارائه 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 به جای اینکه از کارایی ضعیف جلوگیری شود، به سمت موفقیت حرکت شود. (گاستا، ۲۰۰۴)

مقاله پژوهشی دیگر به بررسی کاربرد کنترل آماری فرآیند (SPC) روش‌های اندازه‌گیری عملکرد تامین‌کنندگان فروشگاه‌های زنجیره‌ای یا سوپرمارکت‌های ملی می‌پردازد. انتظار می‌رود که استفاده از نمودار کنترل SPC در درک مدیریت خریدار/ تامین‌کننده کمک کند و تاثیر روابط مناسب و عملکرد تامین‌کنندگان در سیستم بازسازی سوپرمارکت را نشان دهد. طراحی، روش شناسی و رویکرد صورت گرفته برای این کار به گونه‌ای بوده که داده‌های عملکرد ۱۲ تامین‌کنندگان سوپرمارکت‌های ملی را طی یک دوره ۷۷ هفته مورد تحلیل قرار دهد. داده‌های کمی با تحلیل اطلاعات کیفی به دست آمده از مدیران تامین‌کنندگان و خریداران سوپرمارکت تامین شده است. یافته‌ها و نتایج این تحلیل نشان می‌دهد که نه آمار توصیفی و نه رویکرد SPC به تنهایی پاسخ کامل به نظارت بر عملکرد تامین‌کنندگان در سوپرمارکت نمیدهد. در عوض یک رویکرد ترکیبی به احتمال زیاد در بهبود روابط خریدار/ تامین‌کننده بهتر خواهد بود استفاده از SPC تسهیل نظارت بر عملکرد را امکان پذیر می‌کند. (مورگان و دوهارست، ۲۰۰۷)

تحقیق دیگر در مورد کارکنان صنعت است که به داده‌ها و مباحث احتمالاتی و تحلیلی مربوط به آنها می‌پردازند. در این تحقیق از داده‌های متنوعی از منابع مختلف از قبیل مطالعات برروی فرهنگ و تخصص کارکنان و تاثیر آنها بر بازتاب‌ها و واکنش‌های آنها استفاده شده است تا در خصوص میزان حساسیت پرسنل با فرهنگ‌ها و تخصص‌های گوناگون نسبت به تحلیل آنها برروی نمودارهای آماری ارزیابی مناسبی صورت دهد. برای این کار فرایند یادگیری پرسنل را تحلیل نموده تاثیر عوامل قومیتی و فرهنگی را در میزان حساسیت پرسنل برنتایج آماری مورد بررسی قرار می‌دهند. نتایج این مقاله نشان می‌دهد چگونه گروه‌های مختلف کارکنان به نمودارهایی که به عنوان بخشی از

کنترل فرایند آماری استفاده می‌شوند، واکنش نشان می‌دهند، به ویژه تغییر در هدف، مشخصات، روند و مقیاس که در نمودارها نشان داده می‌شود، با استفاده از نکته‌گذاری روی نمودارها، سعی شده روشی را شرح دهند که به کارکنان در مورد نمودارها کمک می‌کند و به آنها تصمیمات آگاهانه داده شود. (هیلز و همکاران، ۲۰۰۷)

در مقاله دیگر هدف بهبود فرایند مدیریت پیکربندی در یک شرکت مخابراتی می‌باشد و تمام دارایی‌های اطلاعاتی، پیکربندی و ارتباط بین یکدیگر در این مقاله به خوبی مستند شده است. این مستندسازی‌ها مفید هستند و چند هدف را بررسی می‌کنند، هدف اول ایجاد رابطه واضح بین شاخص‌های کلیدی عملکرد (KPI) با خدمات فناوری اطلاعات است. تغییرات در پیکربندی این دستگاه‌ها، عملکرد سرویس‌های فناوری اطلاعات را بسیار تحت تاثیر قرار می‌دهند. هدف دوم، صحت اطلاعاتی است که توسط فرایندهای تحویل خدمات استفاده می‌شود. بنابراین یک کارمند سرویس خدمت که نیاز به اطلاعات در مورد نحوه استفاده از یک کاربر در یک شعبه برای اتصال به مقر شبکه، مرتبط با مسائل دسترسی به برنامه‌های خاص دارد، اطلاعات دقیق پیکربندی شبکه، مفید خواهد بود. هدف سوم، صحت اطلاعات است که برای ممیزی IT استفاده می‌شود. PT XYZ یک شرکت مخابراتی است که نسبتاً جدید و آگاه از افزایش محیط رقابتی در صنعت ارتباطات راه دور است. PT XYZ عملیات خود را در سال ۲۰۰۶ آغاز کرد. حرکت به سمت تعالی این شرکت با بهبود عملکرد عملیاتی در ارتباط نزدیک است. بنابراین، این یک ضرورت برای شرکت‌ها در عصر جهانی از خدمات مخابراتی یکپارچه است، برای تمرکز بر کیفیت خدمات (QoS)^۱ ارائه شده به مشتریان خود، به منظور زنده ماندن در فضای رقابتی یک کسب و کار مخابراتی می‌باشد. در تحقیق صورت گرفته کیفیت خدمات در قالب KPIها به کمک کنترل کیفیت آماری مورد پایش و بهبود قرار گرفته است. (سوهایری و گل، ۲۰۱۳)

در پژوهش حاضر عملکرد فرایند ثبت پارساها در قالب شاخص‌های کیفیت داده بطور دقیق برپایه KPIها تعیین شده و مورد پایش و اندازه‌گیری قرار می‌گیرند. نمودارهای کنترل که یکی از تکنیک‌های SPC است در این راستا مورد استفاده قرار خواهند گرفت. نمودارهای کنترل C-chart و X-Bar-R-chart متناسب با نوع شاخص و نحوه کاربرد آن استفاده قرار خواهد گرفت.

از سوی دیگر از آنجاکه فرایند ثبت دارای بخش‌های مختلف نظیر ثبت توسط کاربر و کنترل واحد فراهم‌آوری و همچنین دانشگاه می‌باشد شاخص‌های پیشنهادی به گونه‌ای تعیین خواهد شد که هم عملکرد کلی فرایند ثبت مانند

تعداد کل مدارک برگردانده شده و ... را در نظر بگیرد و هم مشخصه‌های کنترلی مانند نواقص مشاهده شده در اقلام اطلاعاتی را مورد توجه قرار دهد.

۲-۶ روش پژوهش

همانگونه که در تحقیق دادک-بارلیکوسکا مشکلات مربوط به نظارت بر کیفیت در شرکت های لهستانی مورد بررسی قرار گرفت، نظارت بر فرایندهای تولید یک عنصر ضروری از خط مشی کیفیت شرکت است. چنین رویکردی با بهبود مستمر شرکت ارتباط تنگاتنگ دارد. لذا امکان استفاده از تکنیک های آماری فرآیند و برآورد کیفیت مورد بحث قرار گرفته است. برای این منظور ایجاد نمودارهای کنترل از نوع $\bar{X} - R$ و همچنین محاسبه قابلیت شاخص فرآیند C_p و C_{pk} پس از تجزیه و تحلیل نمونه مناسب نشان داده شده است که کاربرد آنها اجازه می دهد تا از بسیاری از نقص تولید در حال حاضر در مرحله تولید اجتناب شود. (دادک-بارلیکوسکا، ۲۰۰۵) متدولوژی استفاده از نمودارهای کنترل مانند $\bar{X} - R$ یا C-chart در تحقیق‌های مختلفی بمنظور ارزیابی عملکرد فرایندها در قالب نمودارهای کنترل مورد استفاده قرار گرفته است.

طرح پژوهشی حاضر در دو فاز اصلی زیر صورت خواهد پذیرفت. ذیل هر فاز مراحل اجرای آن فاز و گام‌های مرتبط و همچنین نحوه جمع‌آوری اطلاعات ارائه خواهد شد.

فاز اول) بررسی متون و طراحی شاخص‌های کیفیت داده، نحوه محاسبه و نمونه‌برداری

گام‌های اجرایی در این فاز عبارتند از:

- ۱-۱) مطالعه و جمع‌بندی شاخص‌های موجود کیفیت داده مستخرج از منابع علمی
- ۲-۱) انطباق شاخص‌های استخراج شده از متون علمی با طرح کیفیت فراداده سامانه ثبت (خروجی طرح قبلی مجری)
- ۳-۱) فرمول‌بندی و تدوین روش محاسبه، نمونه‌برداری و تواتر نمونه‌برداری برای شاخص‌های تدوین شده

فاز دوم) اندازه‌گیری و تحلیل شاخص‌های کیفیت داده بکمک نمودارهای کنترل مناسب

این فاز در قالب گام‌های اجرایی زیر صورت خواهد پذیرفت:

۱-۲) نمونه‌برداری از داده‌های سامانه ثبت جدید و قدیم و اندازه‌گیری شاخص‌های کیفیت داده تعیین شده در فاز اول

۲-۲) محاسبه و طراحی نمودار کنترل مناسب بمنظور تحلیل آماری داده‌ها (لازم به توضیح است از برخی نمودارهای میانگین-دامنه، تعداد نقص و تعداد نقص در واحد در این مرحله استفاده خواهد شد)

۳-۲) تحلیل‌های آماری (حالات تحت کنترل و خارج از کنترل) بر روی نمودارهای کنترل طراحی و ترسیم شده قبلی

۴-۲) جمع‌بندی و تحلیل نمودارهای کنترل و ارائه نتایج

۲-۷ جامعه آماری و حجم نمونه بمنظور تبیین وزن شاخص‌های کیفیت داده

منظور از جامعه آماری همان جامعه اصلی است که از آن نمونه‌ای نمایا یا معرف بدست آمده است (ساروخانی، ۱۳۸۸). بنا به تعریفی دیگر، جامعه آماری به کل گروه افراد، وقایع یا چیزهایی اشاره دارد که تحقیقگر می‌خواهد به تحقیق درباره آن بپردازد.

با توجه به تعاریف فوق جامعه آماری تحقیق حاضر، متشکل از خبرگانی است که با مفاهیم و ابعاد کیفیت داده‌ها آشنا بوده و در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) مشغول می‌باشند و از تجربه بالایی برخوردار هستند و می‌توانند در جمع‌آوری داده‌های مربوط محقق را یاری نمایند. همچنین از مدیران ارشد این حوزه‌ها برای تعیین عوامل و زیرعوامل و تناسب مدل استفاده گردیده است. ضمناً پس از تهیه پرسشنامه برای تکمیل آن، پرسشنامه بین مدیران میانی و ارشد و خبرگان حوزه‌های کیفیت توزیع گردیده است. در این پژوهش با بهره‌گیری از نظر تمامی افراد انتخاب شده به عنوان خبرگان از طریق مراجعه مستقیم به جامعه آماری (متخصصان و صاحب نظران) در بخش‌های گوناگون، به جمع‌آوری از اطلاعات از طریق مصاحبه و پرسشنامه اقدام گردید. لازم به ذکر است که به دلیل تخصصی بودن موضوع و محدودیت در شناسایی افراد صاحب نظر در حوزه کیفیت و ابعاد کیفیت، تعداد محدودی صاحب نظر به عنوان خبرگان انتخاب گردیده‌اند.

۲-۱،۲ نمونه آماری

در یک تعریف کلی نمونه عبارتست از مجموعه‌ای از نشانه‌ها که از یک قسمت، گروه یا جامعه‌ای بزرگتر انتخاب می‌شوند. به طوری که مجموعه معرف کیفیات و ویژگی‌های آن قسمت، گروه یا جامعه بزرگتر باشد.

هرگاه بخشی از یک جامعه آماری براساس قضاوت افراد خاص یا خود تحقیق‌گر برای بررسی انتخاب شوند چنین نمونه‌ای را نمونه قضاوتی می‌گویند (خاکی، ۱۳۸۸).

با توجه به روش مورد استفاده در این بخش از تحقیق که تصمیم‌گیری بصورت گروهی می‌باشد لذا از روش نمونه‌گیری استفاده نشده و بایستی از تمام شماری و انتخاب گروه خبره (با توجه به شاخص‌های خبرگی) استفاده شده و از تعداد خبره‌ها بین ۸-۱۲ نفر بکار گرفته خواهد شد. در این تحقیق از ۱۰ نفر خبره استفاده شده است.

۲-۲ شیوه گردآوری داده‌ها

جمع‌آوری اطلاعات در هر تحقیق با توجه به اهداف آن، روش تحقیق و خصوصیات نمونه انتخاب شده صورت می‌گیرد. داده‌های این تحقیق به دو روش کتابخانه‌ای و مطالعات میدانی جمع‌آوری شده است. روش کتابخانه‌ای جهت جمع‌آوری اطلاعات برای مبانی کلی تحقیق مانند تعریف مفاهیم کلیدی و عملیاتی، طرح ضرورت‌ها، بیان کاربرد و شرح اهمیت آنها استفاده می‌گردد. در این پژوهش برای گردآوری داده‌های مورد نیاز از موارد زیر استفاده شده است:

۱- استفاده از مدارک علمی مرتبط داخلی و مراجع خارجی (مطالعات کتابخانه‌ای)

۲- جستجوی اینترنتی و سایت‌های علمی مانند: EMERALD، ELSEVIER و غیره

۳- مصاحبه با خبرگان، کارشناسان و افراد متخصص در حوزه تحقیق

۴- پرسشنامه‌هایی با طیف لیکرت، فرآیند ترکیبی FDANP

۲-۸ نحوه استخراج عوامل و زیرعوامل

به منظور شناسایی عوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان‌نامه داخل کشور، با تنظیم پرسشنامه و توزیع و تکمیل آن‌ها و همچنین مصاحبه حضوری با مدیران مرتبط با موضوع تحقیق، اطلاعات مورد نیاز برای ادامه تحقیق جمع‌آوری شده‌اند. در این تحقیق از دو پرسشنامه استفاده شده است. در قدم اول با استفاده از تطبیق و مقایسه شرایط موجود کشور با استانداردهای روز دنیا و در نظر گرفتن معیارهای استفاده شده در تحقیق‌های مورد مطالعه و از طریق مصاحبه با خبرگان و کارشناسان حوزه کیفیت فراداده‌های سامانه ملی ثبت پایان‌نامه، تعداد ۱۵ زیرعامل از میان زیرعوامل موجود انتخاب و برای غربالگری و مشخص کردن زیرعوامل تأثیرگذار بر کیفیت

فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور در پرسشنامه شماره ۱ قرار گرفت و در بین ۱۰ نفر از خبرگان توزیع شد.

سپس پرسشنامه‌ها جمع‌آوری و با توجه به میانگین وزن‌های نرمالیزه شده بدست آمده در مورد زیرعوامل پرسشنامه در این تحقیق، زیرعوامل نهایی جهت رتبه‌بندی عوامل و زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه شناسایی شدند.

پس از تعیین عوامل و زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور، جهت بدست آوردن میزان تاثیرگذاری و تاثیرپذیری و میزان اهمیت هر یک از این عوامل و زیرعوامل و همچنین رتبه‌بندی آن‌ها از روش ترکیبی FDANP بهره گرفته شده است. در این مرحله پرسشنامه‌ای به منظور براساس مقایسات زوجی تهیه و در اختیار خبرگان قرار گرفت و میزان تاثیرگذاری و تاثیرپذیری هر یک از عوامل و زیرعوامل تعیین گردید و ارتباط میان آن‌ها مشخص گردید و در نهایت رتبه‌بندی عوامل اصلی و زیرعوامل تعیین گردید.

۹-۲ بررسی روایی و پایایی ابزار جمع‌آوری اطلاعات

ابزار سنجش باید از روایی^۱ و پایایی^۲ لازم برخوردار باشد تا محقق بتواند داده‌های متناسب با تحقیق را گردآوری نماید. در این قسمت به بررسی روایی و پایایی ابزار بکار رفته شده در جمع‌آوری اطلاعات پرداخته می‌شود.

۹-۱-۲ بررسی روایی

روایی تعیین می‌کند که ابزار تهیه شده تا چه حد مفهوم خاص مورد نظر را اندازه می‌گیرد. روایی محتوا اطمینان می‌دهد که ابزار مورد نظر به تعداد کافی، پرسشنامه‌های مناسب برای اندازه‌گیری مفهوم مورد سنجش را در بردارد (اوماسکاران، ۱۳۹۱).

تلاش شده است که پرسشنامه‌هایی که بر اساس مبانی نظری طراحی گردیده از روایی لازم برخوردار باشد. در این تحقیق جهت سنجش روایی محتوایی پرسشنامه ابتدا یک پرسشنامه ابتدایی تدوین گردید سپس آن را در اختیار تعدادی از اساتید و صاحب‌نظران قرار داده و پس از اخذ نظرات خبرگان مذکور، اقدامات اصلاحی صورت پذیرفت.

Validity ^۱

Reliability ^۲

خبره فردی است که دارای شرایط زیر باشد:

- داشتن تجربه کافی در حوزه موردنظر
- داشتن آگاهی عمیق در موضوع تحقیق
- در دسترس بودن
- در مشاغل و موقعیت‌های شغلی بالا بودن
- داشتن دانش بیشتر

۹،۲-۲ بررسی پایایی

اعتبار یا اعتماد پذیری ابزار میزان پایایی و سازگاری آن را در اندازه‌گیری یک مفهوم نشان می‌دهد و برای ارزیابی «برازش» ابزار مفید است (اوماسکاران، ۱۳۸۸).

در مورد پرسشنامه بخش اول که داده‌های مورد نیاز را برای سنجش تاثیر گذاری و تاثیر پذیری و همچنین به منظور وزن دهی عوامل و زیرعوامل گردآوری می‌کند از آلفای کرونباخ برای سنجش اعتبار پرسشنامه استفاده گردیده است. یکی از روش‌های محاسبه قابلیت اعتماد استفاده از فرمول کرونباخ است. روش استفاده از ضریب آلفای کرونباخ را برای تعیین پایایی یک پرسشنامه یا آزمون با تاکید بر همبستگی درونی می‌توان استفاده کرد. از آنجا که یک پرسشنامه با تعدادی سوال (مانند طیف ۵ گزینه‌ای لیکرت) مانند یک آزمون است، می‌توان از رابطه زیر مقدار پایایی را به کمک آلفای کرونباخ بدست آورد.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S^2} \right) \quad \text{رابطه ۱-۲}$$

که در آن K تعداد سوالات، S_i^2 واریانس هر سوال، S^2 واریانس کل سوالات است گفته می‌شود. اگر ضریب آلفای کرونباخ بیشتر از ۰,۷ باشد آزمون از پایایی قابل قبولی برخوردار است.

جدول ۲-۲ بررسی پایایی ابعاد پرسشنامه عوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه در داخل کشور

نام بعد	آلفای کرونباخ	تعداد آیتم‌ها
نمایشگری	۰,۷۶	۴
متنی	۰,۷۸	۵
دسترس پذیری	۰,۸۱	۲
ذاتی	۰,۷۲	۴
کل پرسشنامه	۰,۸۲	۱۵

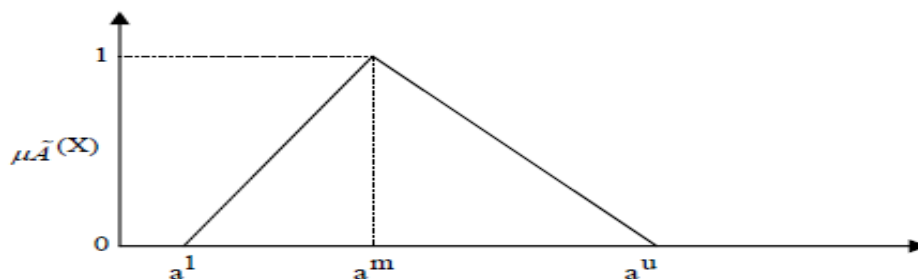
همانگونه که ملاحظه می‌گردد میزان آلفای کرونباخ بدست آمده برای تمامی ابعاد پرسشنامه بالاتر از ۰,۷ می‌باشد که نشان دهنده قابل قبول بودن پایایی پرسشنامه است.

۱۰-۲ منطق فازی

پرفسور لطفی زاده در سال ۱۹۶۵، با معرفی نظریه مجموعه‌های فازی، مقدمات مدل‌سازی اطلاعات نادقیق را فراهم کرد (زاده، ۲۰۰۵). در این تحقیق از اعداد فازی مثلثی استفاده شده است. دلیل استفاده از اعداد مثلثی، محاسبات ساده و قابل فهم آن است و همچنین ثابت شده است که در مسائلی که اطلاعات آن ذهنی و نادقیق است، استفاده از اعداد مثلثی موثر است.

یک عدد فازی مثلثی می‌تواند بصورت سه تایی (al, am, au) نمایش داده شود. تابع عضویت عدد فازی مثلثی $\tilde{A}$ بصورت شکل ذیل است (شوندی، ۱۳۹۶).

$$\mu_{\tilde{A}}(X) = \begin{cases} 0, & x < a^1 \\ (x - a^1) / (a^m - a^1), & a^1 \leq x \leq a^m \\ (a^u - x) / (a^u - a^m), & a^1 \leq x \leq a^m \\ 0, & x > a^u \end{cases}$$



شکل ۲-۴ تابع عضویت و نمودار عدد فازی مثلثی (شوندی، ۱۳۹۶)

۱۱-۲ روش امتیازدهی

جهت مشخص نمودن عوامل ابتدا براساس استانداردهای روز دنیا و در نظر گرفتن معیارهای استفاده شده در تحقیق های مورد مطالعه، فهرستی از عوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور استخراج گردید و در اختیار خبرگان این حوزه در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) قرار داده شد و از ایشان درخواست شد چنانچه عوامل دیگری در این راستا وجود دارد که در فهرست نیامده است به فهرست مذکور اضافه نمایند که نهایتاً ۱۵ زیرعامل به عنوان زیرعوامل اولیه موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) شناسایی گردید.

جهت غربالگری عوامل فوق از شیوه امتیازدهی استفاده شده است به این صورت که از هریک از ۱۰ خبره خواسته شد که میزان اهمیت هر زیرعامل را براساس طیف لیکرت (۷، ۵، ۳، ۱ و ۹) مشخص نماید سپس میانگین نظر خبرگان برای هر زیرعامل محاسبه گردید و زیرعواملی که میانگین آنها پایین تر از عدد ۵ (امتیاز متوسط) بود کنار گذاشته شد. بدین ترتیب ۱۴ زیرعامل از زیرعوامل فوق که بیشترین تاثیر را بر کیفیت فراداده های سامانه ملی ثبت پایان نامه داخل کشور در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) داشته اند شناسایی گردیدند.

۱۲-۲ روش دیمتل فازی بر مبنای ANP فازی (FDANP^۱)

در این تحقیق به منظور وزن دهی و اولویت بندی عوامل و زیرعوامل تاثیرگذار بر کیفیت فراداده های سامانه ملی ثبت پایان نامه داخل کشور در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) از تکنیک دیمتل فازی بر مبنای ANP فازی استفاده شده است. بدین منظور برای مقایسه زیر عوامل اصلی از ۵ عبارت کلامی استفاده شده است که طیف فازی و معادل کلامی آن در جدول ۳-۲ نشان داده شده است.

جدول ۳-۲ متغیرهای زبانی و اعداد فازی متناظر

اعداد فازی مثلثی متناظر	معادل قطعی	متغیر زبانی
۰,۰,۰,۲۵	0	بی تاثیر
۰,۰,۲۵,۰,۵	1	تأثیر خیلی کم
۰,۲۵,۰,۵,۰,۷۵	2	تأثیر کم
۰,۵,۰,۷۵,۱,۰,۰	3	تأثیر زیاد
۰,۷۵,۱,۰,۰,۱,۰,۰	4	تأثیر خیلی زیاد

۱۲,۱-۲ گام نخست - محاسبه ماتریس ارتباط مستقیم (D)

در این گام از پاسخ دهندگان خواسته شد تا میزان تأثیرگذاری معیار A بر معیار J را با استفاده جدول ۳-۲، نشان دهند. برای بررسی معیارها از نظر ۱۰ خبره استفاده شد. در این ماتریس‌ها، $\tilde{x}_{ij} = (l_{ij}, m_{ij}, u_{ij})$ اعداد فازی مثلثی می‌باشند. برای در نظر گرفتن نظر همه خبرگان طبق رابطه ۲-۲، از آن‌ها میانگین حسابی گرفته می‌شود.

$$\tilde{z} = \frac{\tilde{x}^1 \oplus \tilde{x}^2 \oplus \tilde{x}^3 \oplus \dots \oplus \tilde{x}^p}{p} \quad \text{رابطه ۲-۲}$$

در این فرمول p تعداد خبرگان و $\tilde{X}^1, \tilde{X}^2, \dots, \tilde{X}^p$ به ترتیب ماتریس مقایسه زوجی خبره ۱، خبره ۲ و خبره p می‌باشد و $\tilde{Z}$ عدد فازی مثلثی به صورت $\tilde{Z}_{ij} = (l'_{ij}, m'_{ij}, u'_{ij})$ است.

۱۲،۲-۲ گام دوم- نرمال سازی ماتریس ارتباط مستقیم

مطابق با رابطه ۳-۳ ماتریس میانگین را نرمال کرده و آن را ماتریس H می‌نامیم.

برای نرمالایز کردن ماتریس به دست آمده از رابطه های ۳-۲ و ۴-۲ استفاده می‌کنیم.

$$\tilde{H}_{ij} = r \times \tilde{z}_{ij} = (r \times l'_{ij}, r \times m'_{ij}, r \times u'_{ij}) = (l''_{ij}, m''_{ij}, u''_{ij}) \quad \text{رابطه ۳-۲}$$

که ۲ از رابطه زیر به دست می‌آید:

$$r = \min \left\{ \frac{1}{\max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}}, \frac{1}{\max_{1 \leq i \leq n} \sum_{i=1}^n a_{ij}} \right\} \quad \text{رابطه ۴-۳}$$

بنابراین تک تک عناصر جدول ماتریس ارتباط مستقیم را در (r) ضرب می‌کنیم.

۱۲،۳-۲ گام سوم- محاسبه ماتریس ارتباط کامل زیر عامل ها (TC)

بعد از محاسبه ماتریس های فوق، ماتریس روابط کل فازی با توجه به روابط ۵-۲ تا ۸-۲ به دست می‌آید.

$$T = \lim_{k \rightarrow +\infty} (\tilde{H}^1 \oplus \tilde{H}^2 \oplus \dots \oplus \tilde{H}^k) \quad \text{رابطه ۵-۲}$$

که هر درایه آن عدد فازی به صورت است $\tilde{t}_{ij} = (l^t_{ij}, m^t_{ij}, u^t_{ij})$ است و به صورت زیر محاسبه می‌شود:

$$[l^t_{ij}] = H_l \times (I - H_l)^{-1} \quad \text{رابطه ۶-۲}$$

$$[m^t_{ij}] = H_m \times (I - H_m)^{-1} \quad \text{رابطه ۷-۲}$$

$$[u^t_{ij}] = H_u \times (I - H_u)^{-1} \quad \text{رابطه ۸-۲}$$

در این فرمول‌ها H_l ماتریس یکه و H_m و H_u هر کدام ماتریس $n \times n$ هستند که درایه‌های آن را به ترتیب عدد پایین، عدد میانی و عدد بالایی اعداد فازی مثلثی ماتریس H تشکیل می‌دهد.

۱۲،۴-۲ گام چهارم- محاسبه ماتریس ارتباط کامل عوامل اصلی

نخست باید ماتریس T_D را از ماتریس ارتباط کامل زیر عامل‌ها (T_C) استخراج نمود.

هر درایه ماتریس T_D را اگر t_{ij} بدانیم، هر t''_{ij} از میانگین هر T_C^{ij} حاصل می‌گردد.

۱۲،۵-۲ گام پنجم- محاسبه شدت و جهت تأثیر

مطابق با رابطه ۹-۳ و ۱۰-۳ میزان شاخص r_i و C_j را محاسبه می‌نماییم. شاخص r_i بیانگر مجموع سطر i ام و شاخص C_j بیانگر مجموع ستون j ام از ماتریس T_C با توجه به بعد مربوطه می‌باشد. به همین صورت میزان شاخص $\tilde{D}$ و $\tilde{R}$ را محاسبه می‌نماییم. شاخص R_i بیانگر مجموع سطر i ام و شاخص C_j بیانگر مجموع ستون j ام از ماتریس T_D می‌باشد جهت ترسیم و تحلیل نمودار نیاز به ۲ شاخص شدت اثرگذاری و اثرپذیری و جهت تأثیر می‌باشیم که با استفاده از r_i و C_j به دست می‌آیند. برای هر $i=j$ خواهیم داشت:

$$\tilde{D} = (\tilde{D}_i)_{n \times 1} = \left[\sum_{j=1}^n \tilde{T}_{ij} \right]_{n \times 1} \quad \text{رابطه ۹-۳}$$

$$\tilde{R} = (\tilde{R}_i)_{1 \times n} = \left[\sum_{j=1}^n \tilde{T}_{ij} \right]_{1 \times n} \quad \text{رابطه ۱۰-۳}$$

که $\tilde{D}$ و $\tilde{R}$ به ترتیب ماتریس $n \times 1$ و $1 \times n$ هستند.

مرحله بعدی میزان اهمیت شاخص‌ها $(\tilde{D}_i + \tilde{R}_i)$ و رابطه بین معیارها $(\tilde{D}_i - \tilde{R}_i)$ مشخص می‌شود. اگر $\tilde{D}_i - \tilde{R}_i > 0$ باشد معیار مربوطه اثرگذار و اگر $\tilde{D}_i - \tilde{R}_i < 0$ باشد معیار مربوطه اثرپذیر است.

$di+rj$ = شدت اثرگذاری و اثرپذیری (به عبارت دیگر هرچه مقدار $di+rj$ ، عاملی بیشتر باشد، آن عامل تعامل بیشتری با سایر عوامل سیستم دارد).

$di-rj$ = جهت تأثیر گذاری یا تأثیر پذیری (بدین صورت که اگر $di-rj > 0$ باشد معیار مربوطه اثرگذار و اگر $di-rj < 0$ باشد معیار مربوطه اثرپذیر است).

با توجه به مقادیر محاسبه شده در فوق، مقدار شاخص $ri + dj$ و $ri - dj$ را برای معیارها و همچنین شاخص $\tilde{D}_i + \tilde{R}_i$ و $\tilde{D}_i - \tilde{R}_i$ را برای ابعاد بدست می‌آوریم و سپس با استفاده از رابطه ۳-۱۱ فازی زدایی می‌کنیم:

$$B = \frac{(a_1 + a_3 + 2 \times a_2)}{4} \quad \text{رابطه ۳-۱۱}$$

۱۲،۶-۲ گام هشتم- ترسیم نقشه روابط شبکه (NRM)

جهت تعیین نقشه روابط شبکه (NRM)، پس از بدست آوردن ماتریس‌های TC و TD می‌بایست D-R و d-r ماتریس‌های عوامل اصلی و زیر عوامل را محاسبه نماییم. عواملی که D-R بیشتری دارند روی عواملی که D-R کمتری دارند تأثیرگذار می‌باشند و بر عکس عواملی که D-R کمتری دارند از عواملی که D-R بیشتری دارند تأثیر می‌گیرند.

۱۲،۷-۲ گام هفتم- نرمال سازی ماتریس ارتباط کامل ابعاد (T_D^∞)

با توجه به رابطه ۲-۱۲ اقدام به نرمال سازی ماتریس T_D می‌نماییم. به این طریق که مجموع هر سطر از ماتریس T_D را با توجه به بعد مربوطه محاسبه، سپس عنصر هر سطر را بر مجموع عناصر همان سطر تقسیم نموده و در پایان جای سطر و ستون را عوض می‌نماییم.

$$T_D = \begin{bmatrix} t_{11}^{D_{11}} & \dots & t_{1j}^{D_{1j}} & \dots & t_{1m}^{D_{1m}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{i1}^{D_{i1}} & \dots & t_{ij}^{D_{ij}} & \dots & t_{im}^{D_{im}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{m1}^{D_{m1}} & \dots & t_{mj}^{D_{mj}} & \dots & t_{mm}^{D_{mm}} \end{bmatrix} \longrightarrow \begin{aligned} d_1 &= \sum_{j=1}^m t_{1j}^{D_{1j}} \\ d_i &= \sum_{j=1}^m t_{ij}^{D_{ij}}, d_i = \sum_{j=1}^m t_{ij}^{D_{ij}}, i = 1, \dots, m \\ d_m &= \sum_{j=1}^m t_{mj}^{D_{mj}} \end{aligned}$$

$$T_D^\alpha = \begin{bmatrix} t_{11}^{D_{11}} / d_1 & \dots & t_{1j}^{D_{1j}} / d_1 & \dots & t_{1m}^{D_{1m}} / d_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{i1}^{D_{i1}} / d_i & \dots & t_{ij}^{D_{ij}} / d_i & \dots & t_{im}^{D_{im}} / d_i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{m1}^{D_{m1}} / d_m & \dots & t_{mj}^{D_{mj}} / d_m & \dots & t_{mm}^{D_{mm}} / d_m \end{bmatrix} = \begin{bmatrix} t_D^{\alpha 11} & \dots & t_D^{\alpha 1j} & \dots & t_D^{\alpha 1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_D^{\alpha i1} & \dots & t_D^{\alpha ij} & \dots & t_D^{\alpha in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_D^{\alpha n1} & \dots & t_D^{\alpha nj} & \dots & t_D^{\alpha nn} \end{bmatrix} \quad \text{رابطه ۱۲-۳}$$

۱۲، ۸-۲ گام هشتم- نرمال سازی ماتریس ارتباط کامل معیارها (T_C^α) و تشکیل سوپر ماتریس ناموزون
ماتریس TC را با استفاده از روابط زیر نرمال می‌کنیم؛ به این طریق که در این گام مجموع هر سطر T_C^{ij} را با توجه
به بعد مربوطه محاسبه و سپس در هر T_C^{ij} ، هر عنصر بر مجموع عناصر سطر مربوط به خود تقسیم می‌گردد. برای
مثال اگر هر T_C^α را شامل مجموعه‌ای از $T_C^{\alpha ij}$ بدانیم، $T_C^{\alpha 11}$ از نرمال سازی T_C^{11} به دست می‌آید. با ترانسپوز
ماتریس T_C^α ، سوپر ماتریس ناموزون به دست می‌آید.

$$T_C^\alpha = \begin{matrix} & \begin{matrix} D_1 & & D_j & & D_n \\ c_{11} \dots c_{1m_1} & \dots & c_{j1} \dots c_{jm_j} & \dots & c_{n1} \dots c_{nm_n} \end{matrix} \\ \begin{matrix} D_1 \\ c_{11} \\ c_{12} \\ \vdots \\ c_{1m_1} \\ \vdots \\ D_i \\ c_{i1} \\ c_{i2} \\ \vdots \\ c_{im_i} \\ \vdots \\ D_n \\ c_{n1} \\ c_{n2} \\ \vdots \\ c_{nm_n} \end{matrix} & \begin{bmatrix} T_c^{\alpha 11} & \dots & T_c^{\alpha 1j} & \dots & T_c^{\alpha 1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_c^{\alpha i1} & \dots & T_c^{\alpha ij} & \dots & T_c^{\alpha in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_c^{\alpha n1} & \dots & T_c^{\alpha nj} & \dots & T_c^{\alpha nn} \end{bmatrix} \end{matrix} \quad \text{رابطه ۱۳-۲}$$

$$d_{ci}^{11} = \sum_{j=1}^{m_1} t_{cij}^{11}, i = 1, 2, \dots, m_1 \quad \text{رابطه ۱۴-۲}$$

$$T_C^{\alpha 11} = \begin{bmatrix} t_{c11}^{11}/d_{c1}^{11} & \dots & t_{c1j}^{11}/d_{c1}^{11} & \dots & t_{c1m1}^{11}/d_{c1}^{11} \\ \vdots & & \vdots & & \vdots \\ t_{ci1}^{11}/d_{ci}^{11} & \dots & t_{cij}^{11}/d_{ci}^{11} & \dots & t_{cim1}^{11}/d_{ci}^{11} \\ \vdots & & \vdots & & \vdots \\ t_{cm11}^{11}/d_{cm1}^{11} & \dots & t_{cm1j}^{11}/d_{cm1}^{11} & \dots & t_{cm1m1}^{11}/d_{cm1}^{11} \end{bmatrix} \quad \text{رابطه ۲-۱۵}$$

$$= \begin{bmatrix} t_{c11}^{\alpha 11} & \dots & t_{c1j}^{\alpha 11} & \dots & t_{c1m1}^{\alpha 11} \\ \vdots & & \vdots & & \vdots \\ t_{ci1}^{\alpha 11} & \dots & t_{cij}^{\alpha 11} & \dots & t_{cim1}^{\alpha 11} \\ \vdots & & \vdots & & \vdots \\ t_{cm11}^{\alpha 11} & \dots & t_{cm1j}^{\alpha 11} & \dots & t_{cm1m1}^{\alpha 11} \end{bmatrix}$$

۱۲،۹-۲ گام نهم- تشکیل سوپرماتریس موزون

در این مرحله ماتریس T_D^{α} را در ماتریس W (T_C^{α}) ضرب می‌کنیم. به این طریق که هر $t_D^{\alpha ij}$ را در W_{ij} نظیر ضرب می‌کنیم.

۱۲،۱۰-۲ گام دهم- محدود کردن سوپرماتریس موزون

مطابق با رابطه ۲-۱۶، سوپرماتریس موزون را آنقدر به توان (متوالی اعداد فرد) رسانیده تا تمامی اعداد هر سطر همگرا شوند.

$$\lim_{Z \rightarrow \infty} (W^{\alpha l})^Z, \quad \lim_{Z \rightarrow \infty} (W^{\alpha m})^Z, \quad \lim_{Z \rightarrow \infty} (W^{\alpha u})^Z \quad \text{رابطه ۲-۱۶}$$

در ماتریس نهایی همگرا شده اعداد نشان دهنده وزن نسبی زیر عوامل تحقیق می‌باشند که بر اساس آن می‌توان عوامل را با توجه به میزان اهمیتشان اولویت بندی نمود.



شکل ۵-۲ فرایند انجام FDANP

در ادامه و در فصل سوم در ابتدا به مطالعه متون علمی قبلی مرتبط با تعیین و اندازه‌گیری شاخص‌های کیفیت داده خواهیم پرداخت. سپس تبیین شاخص‌های کیفیت داده در سامانه ثبت انجام خواهد شد و در نهایت وزن هر بعد اصلی از شاخص‌های کیفیت داده به کمک FDANP می‌پردازیم.

فصل سوم – استخراج شاخص‌ها و درجه اهمیت آنها در سامانه ثبت

۳-۱ مقدمه

در فصل پیشین به معرفی مبانی و اصول کیفیت داده و مدیریت کیفیت جامع داده و ضرورت پرداختن به آنها در داده‌های سامانه ثبت پارسا پرداخته شد. همچنین باتوجه به یکسان نبودن وزن و درجه اهمیت ابعاد و معیارهای کیفیت داده روشی معرفی گردید که بتوان درجه اهمیت معیارهای مختلف را محاسبه نمود. در این بخش بناداریم شاخص‌هایی را بمنظور ارزیابی کیفیت داده‌ها در فرایند ثبت پارسا توسعه دهیم. بدین‌منظور لازم است در ابتدا تحقیقات پیشین انجام گرفته در خصوص اندازه‌گیری کیفیت داده را مورد مطالعه قرار دهیم و سپس شاخص‌هایی را بمنظور ارزیابی کیفیت فراداده‌های پایان‌نامه/ رساله‌ها تعیین نماییم. همچنین بکمک روش FDANP که در انتهای فصل دوم معرفی شد وزن و درجه اهمیت معیارهای کیفیت داده را تعیین خواهیم کرد.

۳-۲ معرفی تحقیقات انجام شده در خصوص اندازه‌گیری کیفیت داده

همواره اندازه‌گیری کیفیت یک مقوله سخت و پیچیده بوده و به همین دلیل برای مدیران امری دشوار محسوب می‌شود. ویژگی‌های بسیاری از سیستم‌ها وجود دارند که اندازه‌گیری دقیق و قطعی آنها غیرممکن است، کیفیت یکی از این ویژگی‌هاست. همانطور که پیشتر گفته شده کیفیت مقوله‌ای بسیار ذهنی است و افراد مختلف درک متفاوتی از کیفیت خواهند داشت. به عنوان مثال، در مورد فراداده‌ها، ویژگی‌هایی مانند دقت، دسترسی، سازگاری با انتظارات، کامل بودن، قابل فهم بودن یا به موقع بودن وجود دارد که هر ویژگی برای یک معیار خاص مناسب است. بنابراین، این ویژگی‌ها به هیچ عنوان معادل هم نیستند، بلکه جنبه‌های مختلف کیفیت را اندازه‌گیری می‌کنند. گاهی اوقات این معیارها نماینده^۱ ویژگی‌های واقعی یک سیستم هستند.

تحقیقات گذشته نشان می‌دهند بمنظور ارزیابی خودکار کیفیت یک سیستم تعیین شاخص‌ها امری اجتناب ناپذیر است. ارزیابی‌هایی که به صورت دستی توسط انسان‌ها انجام می‌شود امکان دارد همراه با قضاوت‌های شخصی و گاهی همراه با خطا باشند. ارزیابی مستمر یک جزء حیاتی و جدانشدنی برای معیارهای کیفی است. اختصاص منابع برای پوشش هزینه‌های دستی ارزیابی، تنها شاید یکبار امکان پذیر باشد زیرا اگر مستمر و در طول زمان انجام شود هزینه‌های زیادی در بر دارد.

در ادامه در هر زیربخش، بطور جداگانه مهمترین تحقیقاتی که در خصوص اندازه‌گیری شاخص‌های کیفیت صورت گرفته را معرفی خواهیم کرد و سپس براین اساس شاخص‌هایی را بمنظور ارزیابی کیفیت داده‌ها در سامانه ثبت ارائه خواهیم کرد.

۲-۱-۳ معرفی پژوهش (اوچوا^۱ ۲۰۰۸) در خصوص اندازه‌گیری کیفیت فراداده‌ها

چالشی که در خود کارسازی ارزیابی شاخص‌های کیفیت وجود دارد ایجاد تعادل میان مقیاس‌پذیر^۲ بودن و هدفمند بودن^۳ آن ارزیابی است. ارزیابی دستی هدفمند است، اما مقیاس‌پذیر نیست در صورتی که رویکردهای آماری ساده مقیاس‌پذیر هستند، اما معنادار نیستند. ارزیابی کیفی اتوماتیک نیز میتواند معنادار باشد. (اوچوا ۲۰۰۸) محققانی همچون اوچوا و دوال مجموعه غنی از معیارهای قابل اندازه‌گیری کیفیت فراداده را جمع‌آوری کرده‌اند. آنها برای درک بهتر، معیارها را از چند بعد جامعیت، جامعیت دارای وزن، صحت، غنای اطلاعات، قابلیت دسترسی، دسترسی و صحت ذاتی بررسی کرده‌اند که در ادامه به تشریح آنها می‌پردازیم. (اوچوا، خاویر و دوال^۴ ۲۰۰۶)

۲-۱-۳-۱ کامل بودن^۵

معیار کامل بودن با شمارش تعداد فیلدهای تکمیل شده در یک رکورد فراداده، بررسی می‌شود. یک راه برای ساختن یک معیار قابل اندازه‌گیری بر این اساس، برآورد تعداد کل فیلدها و تمام فیلدهایی است که در یک مجموعه می‌باشند. معیار کامل بودن (q_c) به عنوان نسبت تعداد فیلدهای پر شده به کل تعداد فیلدها، به صورت زیر تعریف می‌شود.

$$q_c(record) = \frac{\sum_{i=1}^n [record[field_i] \neq null]}{n} \quad \text{رابطه ۱-۳}$$

در رابطه (۱-۳) q_c به ازای هر رکورد از پایگاه داده شاخص کامل بودن را اندازه‌گیری میکند. $record$ [field_t] نشان‌دهنده فیلد tام از رکورد است و n تعداد فیلدهای مربوط به یک رکورد می‌باشد.

^۱ X Ochoa. Learnometrics ۲۰۰۸.

^۲ Scalable

^۳ Meaningful

^۴ Xavier Ochoa and Erik Duval ۲۰۰۶

^۵ Completeness

۲-۱،۲-۳ کامل بودن دارای وزن^۱

در حالی که کامل بودن یک معیار ساده است این اشکال به وجود می‌آید که کاربران در هنگام جستجو فیلد مورد نظر در تشخیص ارجحیت یکی بر دیگری گمراه می‌شوند بنابراین برای حل این مشکل یک روش جایگزین در نظر گرفته شده که وزن‌های مشخصی را برای هر فیلد در نظر گرفته‌اند. ارزش یک وزن را می‌توان توسط متخصص تعیین کرد. وزن، یک مقدار عددی است که اهمیت نسبی فیلدها بر دیگری نشان می‌دهد. به این صورت که وزن ۱ برای فیلدهای نیمه مهم یا منظم، وزن ۳ برای فیلدهای مهم و وزن ۰ برای فیلدهایی است که باید کاملاً از اندازه‌گیری کنار گذاشته شوند زیرا آنها هیچ اطلاعات مهمی را نشان نمی‌دهند. معیار کامل بودن وزنی می‌تواند جایگزین روش بالا باشد که وزن مشخصی را به هر فیلد اختصاص می‌دهد.

$$q_w(record) = \frac{\sum_{i=1}^n w_i [record[field_i] \neq null]}{\sum_{i=1}^n w_i} \quad \text{رابطه ۲-۳}$$

W_t جایی است که وزن مشخصی از محدوده i را مشخص می‌کند. W_t متناوباً می‌تواند از فرکانس کاربران حاصل شود که هنگام جستجو برای منابع به یک زمینه مشخص تکیه کرد.

۲-۱،۳-۳ صحت^۲

صحت یک رکورد فراداده است که بیانگر درستی مقدار فیلد با توجه به منابع است. اوچوا و دوال پیشنهاد نمودند که صحت می‌تواند به عنوان فاصله معنایی بین اطلاعات داده شده از طریق ثبت فراداده و اطلاعات داده شده از طریق منابع قابل درک باشد. این فاصله معنایی، تفاوت بین اطلاعاتی است که یک کاربر وارد کرده و اطلاعات همان کاربر که می‌توان از منبع، استخراج نمود. فاصله کوتاه‌تر به معنی دقت بالاتر رکورد فراداده است. با این شیوه معیار متری q_a را می‌توان به طریق زیر محاسبه نمود.

$$q_a(record) = 1 - \frac{\sum_{i=1}^n d_i(record[field_i])}{n} \quad \text{رابطه ۳-۳}$$

^۱ Weighted Completeness^۲ Accuracy

مشکلی که در $d(field_t)$ قرار دارد، اندازه‌گیری فاصله از مقدار فیلد $[record[field_t]]$ است. زیرا فیلدهای مختلف نیاز به اندازه‌گیری‌های متفاوت و متناسب با فاصله دارند که برای مقادیر دسته‌ای از یک جدول فاصله از پیش تعریف شده طبق فرمول زیر محاسبه می‌شود،

$$d(field_t) = \begin{cases} 0 & \text{if } field_t = resource_{value_i} \\ 1 & \text{otherwise} \end{cases} \quad \text{رابطه ۳-۴}$$

۳-۴، ۱، ۲، غنای اطلاعات^۱

عبارات و توضیحات موجود به ازای هر رکورد از فراداده به توصیف آن فراداده می‌پردازد. بمنظور ارجاع و توصیف منحصر به فرد یک فراداده لازم است عبارات و توضیحات موجود آن فراداده (اقلام اطلاعاتی) کافی باشد. از دیدگاه کاربر، یک رکورد اطلاعاتی در صورتی با کیفیت است که به اندازه کافی در مورد آنچه که توصیف می‌کند اطمینان کافی را به کاربر منتقل کند. اوچوا و دوال پیشنهاد نمودند که میزان غنای اطلاعات به وسیله اندازه‌گیری میزان اطلاعات منحصر به فرد موجود در فراداده قابل محاسبه خواهد بود. (اوچوا، خاویر و دوال^۲ ۲۰۰۶) آنها این معیار را انطباق با انتظارات نامیدند که به طریق زیر محاسبه می‌شود.

$$q_i(record) = \frac{\sum_{i=1}^n I(resource[field_i])}{n} \quad (۳-۵)$$

در این فرمول که غنای هر رکورد را با شاخص $q_i(record)$ اندازه‌گیری می‌کند تابع I میزان محتوای اطلاعاتی به علاوه سطح منحصر به فرد بودن اطلاعات فیلد I ام را اندازه‌گیری می‌گیرد. در فرمول بالا، یکی از مشکلات در محاسبه شاخص، تعریف تابع I برای انواع متفاوت از فیلدها است. اوچوا و دوال^۳ پیشنهاد استفاده از توابع مختلف برای فیلدهای عددی، فیلدهای واژگان و همچنین فیلدهای متن آزاد را ارائه کردند و دو تابع مختلف را توسعه دادند. برای فیلدهای عددی و واژگان، یک منهای آنتروپی را که می‌تواند با استفاده از تابع زیر بیان شود، تعریف کردند.

$$I(field) = -\log P(resource[field]) \quad (۳-۶)$$

^۱ Richness of Information

^۲ ۲۰۰۶ Xavier Ochoa and Erik Duval

^۳ Ochoa and Duval

که در آن $P (value)$ احتمال آن است که مقدار $value$ (عدد یا واژه) در مجموعه ای از رکوردهای فراداده رخ دهد. برای فیلدهای با متن آزاد نیز فرکانس عبارت- معکوس فرکانس سند^۱ (tf-idf) ارائه شده است که نشان‌دهنده اهمیت نسبی تک کلمات (Single words) نسبت به کل متون است.

$$I(text) = \frac{\sum_{i=1}^n tf(word_i) \cdot \log\left(\frac{m}{df(word_i)}\right)}{n} \quad (۷-۳)$$

تابع tf فرکانس یک کلمه را در فراداده موجود نمایش می‌دهد. به عبارتی کلمه $word_i$ چندبار در رکورد فراداده فعلی رخ داده است.

تابع df بیانگر تعداد کل اسنادی است که کلمه $word_i$ در آنها موجود است. تعداد اسناد بوسیله m و تعداد کلمات مختلف توسط n علامت گذاری می‌شوند.

۳-۱،۵-۲ قابلیت دسترسی^۲

قابلیت دسترسی درجه دسترسی به سوابق یک فراداده را از منظر شناختی^۳ و همچنین فیزیکی و منطقی اندازه‌گیری می‌کند. دسترسی شناختی توضیح می‌دهد که چگونه یک کاربر به سادگی می‌تواند بعد از خواندن سوابق فراداده درک کند که فیلدها در مورد چه چیزی اطلاعات می‌دهند. قابلیت جستجو^۴، نشان خواهد داد که آیا کاربر می‌تواند چیزی را که به دنبال آن است به راحتی پیدا کند یا خیر.

باتوجه به اینکه در برخی متون رسمی (مانند قوانین، آیین نامه ها و ...) درک کاربر از متن ممکن است به راحتی رخ ندهد می‌توان برای اندازه‌گیری این معیار از چندین شاخص قابلیت خواندن^۵ می‌توان استفاده کرد. یکی از آنها،

^۱ Term frequency-inverse document frequency

^۲ Accessibility

^۳ Cognitive

^۴ Searchability

^۵ Readability

تابع Flesch-Kincaid است که در هنگام خواندن یک متن علمی، دشواری درک را اندازه‌گیری می‌کند. نمره سهولت خواندن برای متون انگلیسی را می‌توان با استفاده از عملگر زیر محاسبه کرد که با q_r نمایش داده می‌شود.

$$Q_r(record) = flesch - kincaid(text) \quad (۸-۳)$$

$$= 206.386 - 1.015 \left(\frac{\text{کلمات}}{\text{جملات}} \right) - 84.6 \left(\frac{\text{بخش}}{\text{کلمات}} \right)$$

در رابطه بالا بمنظور محاسبه قابلیت خواندن یک متن لازم است تعداد کلمات، بخشها (Syllables) و جملات آن متن لازم است. اگرچه نمرات توصیف شده در مقیاس بین ۰٫۰ و ۱۰۰٫۰ هستند، اما مقادیر منفی و مقادیر بالای ۱۰۰٫۰ نیز ممکن است در نظر گرفته شود.

۳-۱،۶-۲ در دسترس بودن^۱

این شاخص برخلاف شاخص قبلی به جای دسترس پذیر بودن محتوای رکورد قابلیت دستیابی به منابع اطلاعاتی آن رکورد را مورد ارزیابی قرار می‌دهد. در صورتی که یک رکورد از فراداده در دسترس نباشد، به معنای آن است که URL هایی که آن رکورد را بازیابی می‌کنند که به منابع واقعی آن رکورد اشاره ندارند. اگر منابع یک رکورد قابل بازیابی باشند، به این معنی است که آن رکورد در دسترس است. شاخص در دسترس بودن به کمک فرمول زیر قابل محاسبه خواهد بود.

$$q_{av}(record) = \frac{\sum_{i=1}^n [a_i]}{n} \quad (۹-۳)$$

اگر یک رکورد از طریق آیین URL قابل دسترسی باشد a_i درست (عدد یک) خواهد بود. n تعداد کل URL ها می‌باشد.

^۱ Availability

۳-۱،۷-۲ دقت ذاتی^۱

دقت ذاتی در خصوص فیلدهای متنی به کار برده می‌شود. مشابه با معیار دسترس‌پذیری، این مورد نیز مربوط به قابلیت خواندن می‌باشد. این موضوع به طور مستقیم تحت تاثیر املا درست متن قرار می‌گیرد. میزان دقت ذاتی یک فیلد متنی به کمک شاخص Q_{ip} از رابطه زیر محاسبه می‌شود.

$$q_{ip}(record) = 1 - \frac{m}{n} \quad (۳-۱۰)$$

که در آن m تعداد اشتباهات املایی است و n تعداد کل کلمات است. شاخص دقت ذاتی در یک متن با ۵۰ کلمه و ۱۰ اشتباه املایی ۸۰ درصد می‌شود.

۳-۲،۲ پژوهش رضا وزیری و همکاران (۲۰۱۷) در خصوص اندازه‌گیری کیفیت داده‌ها

محققانی همچون رضا وزیری و همکاران (۲۰۱۷) مطالعاتی را در زمینه اندازه‌گیری کیفیت داده‌ها با در حالتی که معیارها دارای وزن‌های متفاوتی هستند، انجام داده‌اند. کیفیت داده‌ها (DQ) با عنوان "مناسب برای استفاده"^۲ تعریف شده است. (ناندا، ۲۰۱۶) به منظور اندازه‌گیری و بهبود DQ^۳، روش‌های مختلفی تعریف شده است. یک روش DQ مجموعه‌ای از دستورالعمل‌ها و تکنیک‌هایی است که فرایند منطقی را برای اندازه‌گیری و بهبود کیفیت داده‌ها تعریف می‌کند. به منظور ایجاد اندازه‌گیری DQ و بهبود سازمان یافته‌تر، ابعاد DQ تعریف شده است. برخی از مهمترین ابعاد از DQ، دقت، کامل بودن، به موقع بودن و مرتبط بودن می‌باشند. در هر سازمان متناسب با نوع داده و کاربردهای آن شاخص‌های متفاوتی برای اندازه‌گیری کیفیت داده مورد توجه قرار می‌گیرد. در اکثر سازمان‌ها، برخی از داده‌ها بیشتر از سایرین ارجحیت دارند. به عبارت دیگر، برخی داده‌ها وزن بیشتری را دارا می‌باشند. از این رو در اندازه‌گیری شاخص‌های DQ این دسته از داده‌ها بایستی نقش پررنگتری بازی کنند. اکثر معیارهای توسعه یافته تا کنون به وزن داده‌ها توجه نکرده‌اند. وزیری و همکاران، معیارهای جدیدی بر اساس وزن داده‌ها به این منظور ایجاد می‌کنند تا اثربخشی شاخص‌ها را افزایش دهد. مطالعات آنها نشان داد که اثربخشی اندازه‌گیری‌های DQ توسط معیارهای وزنی تا حدود زیادی بیشتر است. در هر شرکت وزن داده‌ها می‌تواند

^۱ Intrinsic Precision

^۲ Suitable for use

^۳ Data quality

متناسب با نقش آن داده در دستیابی شرکت به اهداف کسب و کار خود به شکل متفاوتی تعریف شود. بنابراین، چنین داده‌هایی باید در اندازه‌گیری DQ اهمیت بیشتری داشته باشند. بعنوان مثال در بانک اطلاعاتی مشتریان یک شرکت فیلدهایی مانند آدرس (ستون آدرس) ممکن است به لحاظ اهمیت در سطح بالاتری نسبت به سایر فیلدها (مانند نام، سن و ...) باشد و لذا میتواند با اهمیت جداگانه‌ای در کیفیت داده مورد توجه قرار گیرد. از سوی دیگر مشتریان یک منطقه خاص جغرافیایی ممکن است بمنظور دستیابی سازمان به اهداف استراتژیک خود جایگاه بالاتری نسبت به سایر مناطق داشته باشد پس لازم است داده‌های آن منطقه به طور ویژه‌تری مورد توجه قرار گیرد. در ادامه به معرفی مهمترین شاخص‌های وزن‌دهی شده توسط وزیری و همکاران خواهیم پرداخت.

۳-۲،۲،۱ کامل بودن (وزن‌دهی به ستون‌ها)^۱

این شاخص بدین منظور توسعه داده شده است که بتواند کامل بودن یک پایگاه داده را از منظر کیفیت داده با در نظر گرفتن اهمیت ستونها (فیلدها) ی مختلف مورد ارزیابی قرار دهد. این شاخص میتواند برای کل جدول با فرمول زیر محاسبه شود.

$$\sum_{i=1}^n (cw_i \times cc_i) \quad (11-3)$$

در این فرمول، CW_i درجه اهمیت و CC_i میزان کامل بودن ستون i ام است. n تعداد کل ستونها در پایگاه داده می‌باشد.

۳-۲،۲،۲ کامل بودن وزنی چندتایی‌ها (Tuples)^۲

یک چندتایی (Tuple) در پایگاه داده همان برداری است که هر درایه آن می‌تواند یک یا چند درایه مختلف داشته باشد. بمنظور ارزیابی کیفیت داده در یک چندتایی فرمول وزن‌دهی شده زیر توسط وزیری و همکاران (۲۰۱۷) ارائه شد.

$$\sum_{j=1}^m (tw_j \times tc_j) \quad (12-3)$$

^۱ Weighted column completeness

^۲ Weighted tuple completeness

در رابطه بالا فرض شده است که در یک پایگاه داده m چندتایی (Tuple) مختلف وجود دارد و از نظر سازمان هر چندتایی اهمیت متفاوتی (tw_j) دارد. همچنین میزان کامل بودن چندتایی z با متغیر tc_j مورد ارزیابی قرار می‌گیرد.

در ادامه و در شاخص بعدی می‌توان با ترکیب دو شاخص فوق شاخص کامل‌تری بمنظور ارزیابی کیفیت داده‌ها ارائه نمود.

۳-۲،۲،۳ کامل بودن (وزن‌دهی به هر سلول)^۱

این شاخص روشی برای محاسبه کامل بودن با وزن‌دهی به تک تک سلولهای جدول ارائه می‌کند. هر سلول دارای دو وزن مرتبط با آن است که یکی وزن آن چندتایی (tw_j) است که سلول به آن تعلق دارد و دیگری وزن ستونی است که سلول در آن قرار دارد. با ترکیب این دو وزن فرمول جدید به شکل زیر ارائه خواهد شد.

$$\sum_{i=1}^n \sum_{j=1}^m tw_j \times cw_i \times cell_{ji} \quad (۱۳-۳)$$

بایستی توجه داشت که هنگام ارزیابی سلول بصورت تک‌تک، مقدار $cell_{ji}$ متناسب با کامل یا تهی بودن آن ۱ یا ۰ می‌باشد.

۳-۲،۲،۴ مرتبط بودن ستون‌ها^۲

یکی دیگر از مهمترین مسائلی که باید در نظر گرفته شود، "مرتبط بودن" است. Relevancy به معنی میزان اطلاعات قابل استفاده و مفید برای کار می‌باشد. یکی از معیارهایی که باید برای رعایت آن مورد توجه قرار گیرد تعداد "دسترسی" به داده است. به عبارت دیگر، دامنه‌ها یا ستون‌هایی که بیشترین دسترسی را دارند اهمیت بیشتری دارند و بنابراین در این رویکرد وزن بیشتری به دست می‌آورند. تعداد دسترسی‌ها باید نسبت به دسترسی‌های کل به مجموعه داده‌های مربوطه در نظر گرفته شود. در ادامه مهمترین فرمول‌های شاخص مرتبط بودن ارائه شده است.

^۱ Weighted cell completeness

^۲ Relevancy

$$\text{تعداد دسترسیها به یک پایگاه داده} \\ \text{کل دسترسیها به کل پایگاههای داده} = \text{مرتبط بودن در سطح پایگاه داده} \quad (۱۴-۳)$$

$$\text{تعداد دسترسیها به یک جدول} \\ \text{کل دسترسیها به کل جدولها} = \text{مرتبط بودن در سطح جدول} \quad (۱۵-۳)$$

$$\text{تعداد دسترسیها به یک چندتایی} \\ \text{کل دسترسیها به کل چندتاییها} = \text{مرتبط بودن در سطح چندتایی} \quad (۱۶-۳)$$

$$\text{تعداد دسترسیها به یک ستون} \\ \text{کل دسترسیها به کل ستونها} = \text{مرتبط بودن در سطح ستون} \quad (۱۷-۳)$$

در فرمول‌های ارائه شده بالا میزان دسترسی به پایگاه داده، چندتایی، جدول و ستون در روابط (۱۴-۳) تا (۱۷-۳) مورد ارزیابی قرار می‌گیرد. عبارتی بعنوان مثال در صورتیکه در مجموعه پایگاههای داده تعداد دسترسی‌ها به یک پایگاه بیشتر از سایر پایگاهها باشد شاخص مرتبط بودن برای آن پایگاه داده بایستی عدد بیشتری را نشان دهد.

وزیری و همکاران نشان دادند که نتایج بدست آمده از روابط (۱۶-۳) و (۱۷-۳) میتواند جایگزین مناسبی برای شاخص وزن در روابط (۱۳-۳) و (۱۲-۳) باشد. آنها بمنظور ارائه اهمیت این شاخص مثال زیر را ارائه دادند. مثلاً فرض کنید در یک شرکت فروش محصولات یک مشتری در یک ماه ۵۰ بار تماس بگیرد و مشتری دیگر ۵ بار. پس بایستی در ارزیابی شاخصهای کیفیت داده اهمیت مشتری اول بالاتر از مشتری دوم باشد. موارد این چنینی در شاخص مرتب بودن میتواند مورد توجه قرار گیرد.

فرمول‌های این بخش هنگامی مفید است که سازمان در دست منابع لازم برای بهبود کیفیت همه داده‌ها را در یک زمان نداشته باشد.

۳-۲،۲،۵-صحت^۱

(بتینی و همکاران ۲۰۱۱) صحت را به عنوان نزدیک بودن بین مقدار V و مقدار V' از مجموعه D تعریف می‌کنند، با این فرض که مجموعه D مجموعه‌ای از کلیه مقادیر واقعی و صحیح است. با این تعریف میتوان صحت را در یک چندتایی محاسبه کرد. این کار در رابطه (۱۸-۳) انجام شده است. در فرمول (۱۸-۳)، r_i مقدار

^۱ Accuracy

^۲ Batini et al 2011,

$\bar{a}_i$ از یک چندتایی است که با نماد t نشان داده می‌شود. را نشان می‌دهد و $|t|$ تعداد کل مقادیر موجود در چندتایی است. همچنین شاخص $acc(r_i . D(r_i))$ که همان صحت $\bar{a}_i$ مقدار چندتایی مدنظر است در فرمول (۱۹-۳) تعریف شده است. ، اگر بین مقدار r_i و اعضای مجموعه D انطباق کامل وجود داشته باشد ، فرمول (۱۹-۳) مقدار ۱ را برمی‌گرداند، در غیر این صورت، مقدار برگردانده شده تابع مقدار (NED) را باز می‌گرداند. NED عبارتست از تعداد حداقل درج‌ها، حذف‌ها و جایگزینی برای تبدیل r_i به نزدیکترین مقدار آن در مجموعه D .

صحت بر اساس (NED) :

$$(a) Acc(t) = \frac{\sum_{i=1}^t acc(r_i . D(r_i))}{|t|} \quad (۱۸-۳)$$

$$(b) acc(r_i . D(r_i)) = \left\{ \frac{1}{1 - NED(r_i . D(r_i))} \right\} \quad (۱۹-۳)$$

۲,۲,۶-۳ صحت برای مقادیر عددی

رابطه (۱۹-۳) در مواقعی که یک فیلد شامل کاراکترهاست به خوبی قابل استفاده است. در حالی که در فیلدهای عددی ممکن است این فرمول نتایج مناسبی به همراه نداشته باشد. بعنوان مثال سه عدد ۲۰۰۰ و ۲۰۰۱ و ۱۹۹۹ را در نظر می‌گیریم. واضح است که اختلاف این سه عدد به ترتیب و دو به دو ۱ می‌باشد اما در اختلافی که بین ۱۹۹۹ و ۲۰۰۰ می‌باشد، تغییر عددی بین ۳ رقم و در اختلاف ۲۰۰۰ و ۲۰۰۱ تغییر عددی ۱ رقم می‌باشد. بنابر این وزن تفاضل بین ۱۹۹۹ و ۲۰۰۰ بیشتر از وزن تفاضل بین ۲۰۰۰ و ۲۰۰۱ در نظر گرفته می‌شود.

با در نظر گرفتن توضیحات بالا برای فیلدهای عددی (یا ریاضی) لازم است فرمول دیگری بمنظور تعیین صحت ارائه داد. وزیری و همکاران این رابطه را "فاصله تفاضل ریاضیاتی" (MDD) نامیده و با رابطه (۲۰-۳) آن را ارائه دادند.

در این فرمول، مقدار r_i مقدار $\bar{r}$ از چندتایی است و $D(r_i)$ نزدیکترین مقدار به آن در دامنه D است. به منظور محاسبه تفاوت نسبی بین r_i و $D(r_i)$ تفاضل آنها را به حداکثر دو مقدار تقسیم می‌کنیم. از این رو، تفاضل ۱ بین مقادیر ۹ و ۱۰٪ است، در حالی که تفاوت مشابه بین ۹۹۹ و ۱۰۰۰، ۰٫۱٪ می‌باشد.

$$acc(r_i, D(r_i)) = \left\{ 1 - \frac{1}{\frac{|r_i - D(r_i)|}{\max(r_i, D(r_i))}} \right\} \quad (20-3)$$

۲،۳-۳ پژوهش اوچوا و دووال (۲۰۰۶) در اندازه‌گیری کیفیت فراداده‌های آموزشی

فراداده، نوعی داده ساخته دست انسان است و معیارهای ارزیابی آن توسط روشهای مختلف ایجاد شده است. نتایج نشان می‌دهد که معیارها به ویژگی‌های کیفیت فراداده‌ها بستگی دارند (اوچوا و دووال ۲۰۰۶). در حالیکه مطالعات مختلفی برای تعریف و اندازه‌گیری کیفیت فراداده‌ها فراداده‌ها انجام شده است، روش‌های مقیاس‌پذیر^۱ و موثر محدودی برای ارزیابی کیفیت فراداده‌ها در حال حاضر وجود دارد. آنها تعاریف کیفیت فازی را به معیارهای قابل اجرا (متریک) تبدیل کردند. همچنین برخی از مهمترین شاخص‌های کیفیت داده مانند کامل بودن، دقت، اصالت، سازگاری با انتظارات، انسجام منطقی، به موقع بودن و دسترسی را توسعه دادند.

(اوچوا و دووال ۲۰۰۶)^۲ تحقیقاتی راجع به معیارهای ارزیابی کیفیت داده برای فراداده‌های حوزه آموزش^۳ ارائه داده‌اند. نتایج تحقیقات آنها نشان داد معیارهای کیفیت داده به خوبی می‌تواند کیفیت داده‌های حوزه آموزش را ارزیابی کند. اوچوا و دووال نشان دادند که فازی کردن معیارهای کیفی هرچند به قضاوت انسانی این معیارها کمک می‌کند اما پیچیدگی اجرایی را افزایش می‌دهد. آنها این مفاهیم را به اندازه‌گیری‌های قابل اجرا تبدیل کردند تا بتوانند کیفیت فراداده را به مفهومی تبدیل کنند که اندازه‌گیری شده و مورد استفاده قرار گیرد تا قابلیت‌های اصلی پایگاه‌های داده را مورد ارزیابی قرار دهد. در ادامه به معرفی معیارهای کیفیت داده معرفی شده توسط اوچوا و دووال خواهیم پرداخت.

^۱ Scalable

^۲ 2006.Ochoa, Xavier, and Erik Duval

^۳ Learning Metadata

۳-۱،۳،۲ معیارهای کامل بودن^۱

کامل بودن به این معناست که یک شیء^۲ (محتوای آموزشی) به چه میزان همه رکوردهای مورد نیاز بمنظور ارائه (نمایش) اطلاعات ایده‌آل را در برمی‌گیرد. این نمایش ایده‌آل با توجه به نوع کاربرد و مخاطب داده‌های آموزشی مورد استفاده متفاوت است. که طبق فرمول زیر محاسبه می‌شود. یکی از ابتدایی‌ترین روش‌ها برای ارزیابی سطح کامل بودن یک رکورد از فراداده شمارش تعداد فیلدهایی است که دارای محتوا هستند. (خالی نیستند) در مورد فیلدهای چند متغیره، اگر حداقل یکی از متغیرهای یک فیلد دارای مقدار باشد آن فیلد را کامل در نظر می‌گیریم. در ادامه رابطه ۳-۲۱ فرمول ساده‌ای برای ارزیابی کامل بودن ارائه می‌دهد.

$$Q_{completeness} = \frac{\sum_{i=1}^n p(i)}{N} \quad (3-21)$$

اگر فیلد i دارای یک مقدار غیر صفر باشد، $P(i)$ برابر ۱ است و در غیر این صورت صفر می‌باشد. N تعداد فیلدها است.

متخصصان کیفیت داده اغلب برای هر فیلد از پایگاه داده وزن اختصاص می‌دهند، که اهمیت نسبی آن را نسبت به فیلدهای دیگر نشان می‌دهد. در این حالت فرمول کامل بودن طبق رابطه زیر محاسبه می‌شود.

$$Q_{wcompleteness} = \frac{\sum_{i=1}^N \alpha_i * p(i)}{\sum_{i=1}^n \alpha_i} \quad (3-22)$$

α_i اهمیت نسبی فیلد i ام نسبت به سایر فیلدها است.

مقدار α_i می‌تواند در یک دوره معین از زمان برای فیلد i ام در مقایسه با دوره‌های زمانی دیگر متفاوت باشد. تکیه دارند. همچنین مقدار α_i می‌تواند از یک تحقیق پیمایشی و براساس نظرات خبرگان در خصوص اهمیت فیلدهای مختلف بدست آید.

۲-۳-۲ معیارهای صحت^۱

صحت شاخصی است که نشان می‌دهد به چه میزان مقادیر فراداده "درست"^۲ است. در حقیقت شاخص صحت نشان می‌دهد که یک رکورد چقدر خوب توانسته است یک شیء را توصیف کند. شاخص صحت میتواند در دسته شاخص‌های عینی^۳ یا ذهنی^۴ طبقه‌بندی شود. در حوزه شاخص‌های عینی صحیح یا غلط بودن یک فیلد به راحتی ارزیابی می‌شود. بعنوان مثال حجم یک فایل آموزشی یا زبان آن فایل (مانند انگلیسی، آلمانی یا ...) در فراداده، فیلدهایی هستند که به راحتی پس از ارزیابی و مشاهده خود فایل در قالب شاخص صحت قابل اندازه‌گیری هستند. لذا شاخص صحت برای این دسته از فیلدها میتواند صفر یا یک باشد. بعنوان مثال اگر زبان یک فایل آموزشی انگلیسی بوده باشد و در فراداده مربوط به آن فیلد زبان "آلمانی" درج شده باشد صحت این فیلد صفر لحاظ می‌شود ولی اگر این فیلد "انگلیسی" باشد یک لحاظ می‌شود. اما در حوزه شاخص‌های ذهنی ارزیابی صحت یک رکورد به قضاوت شخص ارزیابی کننده بستگی دارد. به عنوان مثال مناسب بودن عنوان یک تصویر یا فایل آموزشی یا فیلد توصیف یک سند آموزشی کاملاً یک مفهوم ذهنی است و ممکن است ارزیاب‌های مختلف مقادیر متفاوتی به شاخص صحت در این فیلدها تخصیص دهند. به طور کلی صحت می‌تواند فاصله معنایی^۵ بین اطلاعاتی باشد که یک کاربر می‌تواند از یک رکورد فراداده دریافت کند و اطلاعاتی که همان کاربر پس از مشاهده شیء مربوط به آن فراداده دریافت می‌کند. صحت کوتاهترین فاصله، میان این دو مفهوم است که طبق فرمول زیر محاسبه می‌شود.

$$Q_{accuracy} = 1 - \frac{\sqrt{\sum_{i=1}^n d(field_i)^2}}{\sum_{i=1}^n d(field_i)} \quad \left(\sum_{i=1}^n d(field_i) > 0 \text{ با در نظر گرفتن} \right) \quad (23-3)$$

که در این رابطه عبارت $d(field_i)$ میزان صحت تعلق گرفته به فیلد i ام می‌باشد.

دشواری در اینجا اندازه‌گیری فاصله (d) است که برای انواع فایل‌ها و فیلدها متفاوت است برای فیلدهای عددی یا فیلد تاریخ می‌توان آن را به راحتی محاسبه کرد. در ارزیابی ذهنی می‌توان از شرطهای چندگانه برای ارزیابی صحت استفاده کرد. بعنوان مثال شاخص صحت اگر مقدار یک فیلد با مقدار از پیش تعیین شده آن دقیقاً یکسان باشد، صفر، اگر ۱۰ درصد اختلاف معنایی داشته باشند این شاخص ۰/۵ و در غیر اینصورت یک خواهد بود.

^۱ Accuracy Metrics

^۲ Correct

^۳ Objective

^۴ Subjective

^۵ Semantic

برای مقادیر قابل دسته‌بندی^۱ (به عنوان مثال زبان) می‌توان از یک جدول فاصله از پیش ساخته استفاده نمود. به عنوان مثال، فاصله بین اسپانیایی و ایتالیایی را می‌توان ۰٫۲، بین اسپانیایی و ژاپنی را ۱ در نظر گرفت. هر متن می‌تواند به عنوان یک بردار نشان داده شود، جایی که مقدار i ام نشان‌دهنده وجود یا عدم وجود کلمه i در متن است. سپس می‌توانیم هر روش برای محاسبه فاصله بین بردارها، مانند فاصله اقلیدس یا شباهت کوزین استفاده کنیم که در ادامه رابطه آن ارائه شده است.

$$(۲۴-۳) \quad d(\text{textfield1}, \text{textfield2}) = \frac{\sum_{i=1}^N \text{tf } 1\text{vector}_i * \text{tf } 2\text{vector}_i}{\sqrt{\sum_{i=1}^N \text{tf } 1\text{vector}_i^2 \times \sum_{i=1}^N \text{tf } 2\text{vector}_i^2}}$$

جایی که بردار $\text{tf } 1$ و بردار $\text{tf } 2$ ، اگر حاوی کلمه i ام باشد، ۱ خواهد بود در غیر این صورت صفر می‌باشد. N تعداد کل کلمات است. این روش از محاسبه صحت زمانی معتبر می‌باشد که فراداده‌ای به فیلد اضافه شده باشد. به عنوان مثال، فاصله معنایی را می‌توان بین فراداده عنوان یا توضیحات و متن موجود در یک فیلد با استفاده از الگوریتم‌های پیشرفته محاسبه کرد. برای محاسبه، می‌توان از الگوریتم‌های تجزیه و تحلیل پیشرفته کمک گرفت. یکی از مهمترین این ابزارها، تجزیه و تحلیل پنهان معنایی (LSA)^۲ می‌باشد. (لاندائر، فولتز و لاهام،^۳ ۱۹۹۸)

۳-۳،۳،۲ شاخص اصالت^۴

شاخص اصالت، شهرتی را که رکورد فراداده در یک جامعه دارد، پوشش می‌دهد. به عنوان مثال، یک کاربر ممکن است به متادیتای تولید شده توسط یک خبره نسبت به متادیتایی که نرم افزار تولید می‌کند، اعتماد بیشتری داشته باشد حتی اگر در سایر شاخص‌های کیفیت داده وضعیت مطلوبی نداشته باشد. به همین دلیل این شاخص در دسته شاخص‌های ذهنی طبقه‌بندی می‌شود. به منظور تبدیل ارزیابیهای ذهنی کاربران میتوان برداشت‌های آنها را در قالب شاخص‌های قابل اندازه‌گیری تفسیر و در نتیجه محاسبه نمود. به عبارتی تحلیل ثبت وقایع عملکرد

^۱ Categorical

^۲ Latent Semantic Analysis

^۳ Landauer, Foltz, & Laham, 1998

^۴ Provenance Metrics

کاربران (Log) می‌تواند به نوعی به اندازه‌گیری شاخص اصالت منجر شود. در جدول (۳-۱) چند نمونه از این تحلیلها که به شاخصهای کمی منجر شده، نمایش داده شده است.

جدول ۳-۱ اطلاعات مورد نیاز برای محاسبه پروتکل متریک

شاخص کمی اصالت	تحلیل لاگ ^۱
نمره میانگین هر تولیدکننده فراداده	نمره‌ای که کاربر به تولیدکننده فراداده می‌دهد
میانگین نمره اشیاء متعلق به یک تولیدکننده فراداده	نمره‌ای که کاربر به هر رکورد از فراداده می‌دهد
متوسط تعداد دانلودهای انجام شده توسط یک کاربر از یک تولیدکننده فراداده	اشیاء توسط کاربران دانلود میشوند

۳-۴، ۳، ۲ سازگاری^۲

سازگاری شاخصی است که میزان انطباق یک فراداده با استانداردهای از قبل تعیین شده را ارزیابی می‌کند. این استانداردها کاملاً براساس نوع فراداده و هدف طراحی آن میتواند متفاوت باشد. اگر N تعداد فیلدهای مورد تجزیه و تحلیل قرار گرفته باشد فرمول زیر شاخص سازگاری را محاسبه می‌کند.

$$Q_{consistency} = \frac{\sum_{i=1}^n level-of-compliance(field_i.value_i)}{N} \quad (3-29)$$

که در آن Level-of-compliance به ازای هر فیلد میزان انطباق آن فیلد با استانداردها را نمره‌دهی می‌کند.

^۱ Log

^۲ Logical consistency and coherence Metrics

۳-۲,۳,۵ معیارهای به هنگام بودن^۱

به موقع بودن به طور خاص مربوط به درجه‌ای است که یک رکورد از فراداده در بین مخاطبان آن مورد استقبال قرار می‌گیرد. در حقیقت سطح به روز بودن داده توسط این شاخص ارزیابی می‌شود. در واقع با اندازه‌گیری و محاسبه می‌توان فهمید که یک فراداده چگونه با گذشت زمان باز هم می‌تواند مفید و به روز باشد.

$$(30-3) \quad \text{age} = \text{present_year} - \text{publication_year}$$

$$(31-3) \quad \text{frequency - of - use} = \frac{\text{times-retrived}}{\text{total-records-retrived}} (\text{Over a period of 1 year})$$

$$(32-3) \quad Q_{\text{currency}} = \text{age}() * \text{frequency - of - use}()$$

ابتدا در دو رابطه (۳۰-۳) تا (۳۱-۳) به ترتیب دو شاخص عمر فراداده (Age) و همچنین میزان نسبت بازیابی آن فراداده در مقایسه با کل فراداده‌ها (Frequency-of-use) محاسبه شده و سپس در رابطه (۳۲-۳) شاخص به هنگام بودن از ضرب این دو شاخص محاسبه خواهد شد. شاخص frequency-of-use از نسبت تعداد دفعات بازیابی یک رکورد (times-retrived) به کل دفعات بازیابی رکوردها (total-records-retrived) بدست می‌آید.

میزان به هنگام یک رکورد از فراداده آموزشی یا LOM^۲ می‌تواند نشان دهد که فراداده با گذشت زمان چگونه می‌تواند مفید واقع شود. به عنوان مثال، اگر یک LOM در یک دوره ۵ سال پیش ایجاد شود و کاربران هنوز هم بتوانند محتویات این دوره را پیدا کرده و به درستی آن را بارگیری کنند، پس می‌توان این داده‌ها را به هنگام دانست.

^۱ Timeliness Metrics

^۲ Learning Object Metadata

۲,۳,۶-۳ متریک دسترس پذیری

شاخص قابلیت دسترسی درجه ای است که نشان می‌دهد، LOM به چه میزان هم از نظر شناختی و هم فیزیکی / منطقی در دسترس است و همچنین درک اطلاعات را برای کاربر آسان می‌کند. دسترسی شناختی میزان فهم کاربر از اطلاعات موجود در فراداده را ارزیابی می‌کند. خطاهای املایی، انطباق با واژگان نمونه‌هایی از این نوع دسترسی هستند. (گای و همکاران، ۲۰۰۴)^۱ شاخص Flesch (فولتز و همکاران ۱۹۹۸)^۲ شاخصی است که بطور خودکار برای توصیف این دسترس پذیری استفاده می‌شود. این شاخص سطح ثقیل بودن کاراکترها و جملات یک متن را ارزیابی می‌کند.

$$Q_{readability} = \frac{Flesch(description_text)}{100} \quad (3-33)$$

همانطور که (اوچوا و دووال، ۲۰۰۶) نشان دادند میتوان از رابطه (۳-۳۱) برای ارزیابی شاخص دسترس‌پذیری فیزیکی / منطقی استفاده نمود.

۲,۴-۳ پژوهش کرول و همکاران (۲۰۰۷) در اندازه‌گیری کیفیت داده‌های املاک

کرول و همکاران (۲۰۰۷)^۲ مطالعاتی را در زمینه پایگاه داده‌ها که یکی از مهمترین دارایی‌های شرکت‌های بزرگ است انجام داده‌اند. اما معمولاً کیفیت داده‌ها ضعیف است و یا حتی قابل کنترل نیست. آنها چندین معیار کیفیت داده برای نظارت بر پایگاه داده یک سیستم اطلاعاتی در حوزه املاک و زمین‌ها توسعه دادند. هدف اصلی آنها از طراحی معیارهای کیفیت داده‌ها، نظارت بر پایگاه داده‌ها و دادن هشدارهای به موقع در زمان عبور میزان خطاها از یک مقدار آستانه می‌باشد. آنها با توجه به تحقیقات صورت گرفته در مورد تغییرات کیفیت داده‌ها، چندین معیار برای دامنه و نقص‌های ارجاعی استفاده کردند. در ادامه مهم‌ترین شاخص‌های کیفیت داده که توسط کرول و همکاران (۲۰۰۷) ارائه شده است مورد بررسی قرار می‌گیرد. لازم به توضیح است شاخص‌های ارائه شده در سه دسته اصلی دسته‌بندی شده است. در دسته اول در سطح هر پایگاه داده نقص‌ها شناسایی شده و اندازه‌گیری می‌شوند. در دسته دوم با توجه به اینکه صلاحیت کیفیت داده نیازمند صرف هزینه است لذا در این دسته علاوه بر

^۱ Guy et al., 2004

^۲ Foltz et al., 1998

^۳ Crole & el. 2007

تعداد نقص هزینه رفع هر نقص نیز به عنوان یک ضریب در نظر گرفته شده است. از سوی دیگر در دسته سوم شاخص‌های کیفیت داده در کل سیستم معرفی می‌شود. بعبارتی در یک سیستم جمع‌آوری داده پایگاه‌های اطلاعاتی مختلفی مانند اطلاعات مالکین، اطلاعات املاک و مستغلات و همچنین اطلاعات درصد سهم مالکین و ... وجود دارد که ارزیابی کل سیستم نیازمند جمع‌بندی شاخص‌های کیفیت داده در اجزاء مختلف آن می‌باشد. لذا در دسته سوم شاخص‌های کلان‌تر ارائه شده است. در ادامه به معرفی شاخص‌های کیفیت داده در هر دسته می‌پردازیم.

۳-۱، ۲، ۴ شاخص‌هایی برای تعداد نقص‌های شناسایی شده

d^1 تعداد نقص‌های دامنه شناسایی شده در یک پایگاه داده است که می‌تواند با فرمول زیر بیان شود:

$$ND_{dom} = \sum_{i=1}^m d_{dom}(i) \quad (3-34)$$

d^2 ND تعداد نقص‌های دامنه شناسایی شده در یک پایگاه داده است که می‌تواند با فرمول زیر بیان شود:

$$ND_{dom} = \sum_{i=1}^m d_{dom}(i) \quad (3-35)$$

ND کوتاه شده Number of defects به معنی تعداد نقص‌ها می‌باشد و $d_{dom}(i)$ به معنی تعداد نقص‌ها در رکورد i ام در پایگاه داده است. همچنین در فرمول بالا m تعداد رکوردهای مورد بررسی قرار گرفته است.

کوتاه شده Number of defects به معنی تعداد نقص‌ها می‌باشد و $d_{dom}(i)$ به معنی تعداد نقص‌ها در رکورد i ام در پایگاه داده است. همچنین در فرمول بالا m تعداد رکوردهای مورد بررسی قرار گرفته است.

۳-۲، ۴، ۲ تعداد نقص‌های مربوط به سازگاری

این دسته از نقص‌ها به نواقص مربوط به عدم ارتباط مناسب میان پایگاه‌های داده مختلف می‌پردازد. بعنوان مثال ممکن است در پایگاه داده مالکین اطلاعات یک فرد کامل باشد در حالیکه آدرس وی در پایگاه داده آدرسها ناقص باشد. تعداد نقص‌های مربوط به سازگاری که در یک پایگاه داده شناسایی شده است، می‌تواند با فرمول زیر بیان شود:

$$ND_{ref} = \sum_{i=1}^n d_{ref}(j) \quad (3-36)$$

Ref کوتاه شده referential و d_{ref} به معنای نقص های مرجع می باشد

۲,۴,۳-۳ تعداد کل نقص ها^۱

مجموع تعداد دامنه ها و نقص های مرجع شناسایی شده در یک پایگاه داده است که می تواند با فرمول زیر بیان شود:

$$(۳۷-۳) \quad ND_{tot} = ND_{dom} + ND_{ref} = \sum_{i=1}^m d_{dom}(i) + \sum_{j=1}^n d_{ref}(j)$$

۲,۴,۴-۳ تعداد نقصهای پایگاه داده در هر بخش^۲

برابر با تعداد نقص های حوزه شناخته شده در حال سقوط بر روی بخش تست شده و به صورت زیر بیان می شود:

$$(۳۸-۳) \quad NDP_{dom} = \frac{\sum_{i=1}^m d_{dom}(i)}{KREC}$$

NDP_{dom} به معنی تعداد نقص های هر بخش پایگاه و KREC برابر این تعداد تقسیم بر بخش می باشد.

۲,۴,۵-۳ تعداد نقایص سازگاری در هر بخش^۳

برابر با تعداد نقص های مرجع شناسایی شده در حال سقوط بر روی بخش آزمایش شده و به صورت زیر بیان می شود:

$$(۳۹-۳) \quad NDP_{ref} = \frac{\sum_{j=1}^n d_{ref}(j)}{KREC}$$

^۱ Total number of defects

^۲ Number of domain defects per portion

^۳ Number of referential defects per portion

۲,۴,۶-۳ تعداد کل نقص در هر بخش^۱

برابر با مجموع تعداد دامنه‌ها و نقص‌های ارجاعی است که در قسمت آزمایش شده تشخیص داده شده و به صورت زیر بیان می‌شود:

$$(۴۰-۳) \quad NDP_{tot} = NDP_{dom} + NDP_{ref} = \frac{\sum_{i=1}^m d_{dom}(i) + \sum_{j=1}^n (j)}{KREC}$$

۲,۴,۷-۳ درصد سوابق نامعتبر^۲

نسبت تعداد رکوردهایی است که حداقل یک رکورد به تعداد رکوردهای تست شده کشف شده است که بیان شده توسط فرمول زیر می‌باشد:

$$(۴۱-۳) \quad PIR = \frac{NIR}{REC} * 100\%$$

PIR درصد سوابق نامعتبر در نمونه‌های تست شده

NIR تعداد رکوردهای نامعتبر، یعنی تعداد رکوردهایی که حداقل یک نقص دارند را نشان می‌دهد.

REC تعداد کل سوابق تست شده

۲,۴,۸-۳ معیارهای محاسبه هزینه‌های معیوب^۳

۲,۴,۸,۱-۳ هزینه حذف نقص پایگاه داده^۴

کل هزینه حذف تمام نقص‌های پایگاه داده شناسایی شده در پایگاه داده است می‌تواند با فرمول زیر بیان شود:

$$(۴۲-۳) \quad C_{dom} = \sum_{i=1}^m c_{dom}(i) * d_{dom}(i) + \sum_{k=1}^o c_{dom}(k) * f(t_{dom}(k))$$

بخش اول این فرمول مربوط به نقص‌هایی است که باید به طور یکنواخت برداشته شوند و $C_{dom}(i)$ هزینه حذف نقص دامنه i ام را نشان می‌دهد. قسمت دوم به نوبه خود به نوعی نقص می‌پردازد که می‌تواند با استفاده از یکی از

^۱ Total number of defects per portion

^۲ Percentage of invalid records in records tested

^۳ Metrics Calculating Costs of Defect Removal

^۴ Cost of domain defects removal

آنها حذف شود این روش، به طوری که هزینه $C_{dom}(k)$ در مورد وقوع نقص دامنه k ام به تعداد این نقص‌ها می‌پردازد. در رابطه بالا هزینه اصلاح نقص بایستی در زمانی در فرمول مورد محاسبه قرار گیرد که در نمونه بررسی شده نقصی وجود داشته باشد لذا عبارت $f(t_{dom}(k))$ در زمانی که در نمونه k ام نقصی مشاهده شود ۱ و در غیر اینصورت ۰ خواهد بود.

۲-۳، ۴، ۸، ۲-۳ هزینه حذف عبارات^۱

کل هزینه حذف همه نقص‌های سازگاری شناسایی شده در یک پایگاه داده است که میتواند با فرمول زیر بیان شود.

$$C_{ref} = \sum_{j=1}^n C_{ref}(j) * d_{ref}(j) + \sum_{l=1}^p C_{ref}(l) * g(t_{ref}(l)) \quad (۴۳-۳)$$

بخش اول این فرمول مربوط به نقص‌هایی است که باید به صورت یک به یک برداشته شود و $C_{ref}(j)$ نشان‌دهنده هزینه حذف نقص مرجع j است. قسمت دوم به نوبه خود به چنین نقایصی اعمال می‌شود که می‌تواند با استفاده از یک روش حذف شود، به طوری که هزینه‌یابی $C_{ref}(l)$ در رابطه با وقوع نقص l ام است که به تعداد این نقص‌ها نمی‌رسد. لذا عبارت $g(t_{ref}(l))$ در زمانی که در نمونه l ام نقصی مشاهده شود ۱ و در غیر اینصورت ۰ خواهد بود.

۲-۳، ۴، ۸، ۳-۳ هزینه کل حذف نقص^۲

مبلغ هزینه‌های حذف هر دو دامنه و نقص مرجع است که می‌تواند با فرمول زیر بیان شود:

$$C_{tot} = C_{dom} + C_{ref} \quad (۴۴-۳)$$

به همین ترتیب، هزینه حذف نقص دامنه در هر قسمت CP_{dom} را تعریف می‌کنیم. هزینه حذف عبارات معیاری در هر بخش، CP_{ref} و کل هزینه حذف عیوب در هر بخش CP_{tot} .

^۱ Cost of referential defects removal

^۲ Total cost of defects removal

۲,۴,۹,۱-۳ معیارهای خاص برای سیستم کاداستر^۱

۲,۴,۹,۱-۳ تعداد واحدهای ثبت نام معتبر

تعداد واحدهای ثبت شده است که در آنها حداقل یک نقص وجود دارد که می‌تواند توسط $NIRU$ به معنای تعداد رکوردهای ثبت شده نامعتبر مشخص شود.

در صد واحدهای ثبت نام بدون خطا: نسبت تعداد واحدهای ثبت نام است که در آن هیچ عیبی در اشیاء خود به تعداد تمام واحدهای ثبت شده مورد آزمایش قرار نگرفته است، که فرمول زیر را بیان می‌کند.

$$PRU_{errf} = \frac{NRU_{errf}}{RU} * 100\% \quad (۴۵-۳)$$

PRU_{errf} درصدی از واحدهای ثبت قابل قبول

NRU_{errf} رکوردهای ثبت شده نامعتبر

RU تعداد ثبت

این نسبت تعداد واحدهای ثبت است که در آن هیچ نقص بحرانی وجود ندارد، اما باعث می‌شود پردازش داده‌ها به درستی انجام نشود. توسط فرمول زیر بیان می‌شود:

$$PRU_{accg} = \frac{NRU_{crif}}{RU} * 100\% \quad (۴۶-۳)$$

PRU_{accg} درصدی از واحدهای ثبت قابل قبول

NRU_{crif} نقص انتقادی

^۱ Metrics Specific for a Cadastral System

^۲ Number of invalid register units

۲-۵-۳ پژوهش پپینو و همکاران (۲۰۰۲) در خصوص انواع شاخص‌های کیفیت داده

پی پی نو و همکاران^۱ در این مقاله، ارزیابی ذهنی^۲ و عینی^۳ کیفیت داده‌ها را توصیف می‌کنند و سه دسته اصلی برای طبقه‌بندی انواع معیارهای کیفیت داده‌ها ارائه می‌دهند.

جدول ۲-۳ انواع ابعاد کیفیت داده معرفی شده توسط آنها را نمایش می‌دهد و تعریف مربوط به هر یک را ارائه می‌دهد.

جدول ۲-۳ ابعاد کیفیت داده (پپینو و همکاران (۲۰۰۲))

ابعاد	تعاریف
قابلیت دسترسی ^۴	میزان در دسترس بودن و سطح سرعت و بازیابی داده‌ها
مناسب بودن حجم داده ^۵	میزان مناسب بودن حجم داده برای فعالیت مورد نظر
باور پذیری ^۶	میزان سطحی که داده‌ها درست و معتبر در نظر گرفته میشوند
جامعیت ^۷	میزان کفایت و کامل بودن داده‌ها میزانی که داده‌ها با توجه به طول و عرض و دامنه برای حوزه مورد نظر کافی باشند.
ارایه همسان ^۸	سطح موجز بودن داده‌های ارایه شده
ارایه سازگار ^۹	میزان مطابقت داده‌ها با فرم اصلی و ساختار مورد نظر
سهولت محاسبه ^{۱۰}	میزان سهولت استفاده و به کارگیری داده‌ها در نقش‌های مختلف
بدون خطا ^{۱۱}	میزان صحیح و قابل اعتماد بودن داده‌ها
قابلیت تفسیر ^{۱۲}	میزان مناسب بودن و شفاف بودن زبان، علایم، واحدها و تعاریف

^۱ Pipino, Leo L., Yang W. Lee, and Richard Y. Wang.2002

^۲ Subjective evaluation

^۳ Objective evaluation

^۴ accessibility

^۵ Appropriate amount of data

^۶ believability

^۷ completeness

^۸ Concise representation

^۹ Consistent representation

^{۱۰} Ease of manipulation

^{۱۱} Free-of-error

^{۱۲} interpretability

میزان سطح نا اریب بودن و بی طرفانه داده ها	عینیت ^۱
میزان سطح قابلیت کاربرد و مفید بودن داده ها برای حوزه مورد نظر	مرتبط بودن ^۲
میزان اتکا بودن داده از جنبه منبع و محتوا	معتبر بودن ^۳
میزان سطح محدودیت های دسترسی	امنیت ^۴
میزان به روز بودن داده برای فعالیت های مورد نظر	به هنگام بودن ^۵
سطح قابل درک بودن داده ها	درک پذیری ^۶
میزان سودمند بودن و مزایای استفاده از داده	ارزش افزوده ^۷

۳-۱،۵،۲ طبقه‌بندی انواع معیارهای کیفیت داده‌ها

هنگام انجام ارزیابی‌های عینی، شرکت‌ها باید مجموعه‌ای از اصول را دنبال کنند. برای توسعه معیارهای خاص نیازهای آنها سه فرم کاربردی فراگیر: نرخ ساده، عملیات حداقل یا حداکثر و میانگین وزنی طراحی شده است. اصلاح این فرم‌های عملکردی، مانند اضافه کردن پارامترهای حساسیت، میتواند به آسانی قابل تنظیم باشد. اغلب سخت‌ترین بخش تعریف ابعادی است که مربوط به کاربرد خاص شرکت است.

• نرخ ساده

نرخ ساده، نسبت نتایج مطلوب به کل را اندازه می‌گیرد. تجربه نشان می‌دهد که مدیران ترجیح می‌دهند که نسبت نتایج مثبت را نشان دهند، زیرا این فرم برای مقایسه‌های طولی مفید است که نشان‌دهنده روند بهبود مستمر است. بسیاری از معیارهای کیفیت داده‌های سنتی، مانند خطای رایگان، تکمیل و انطباق این فرم را می‌گیرند. ابعاد دیگری که میتواند با استفاده از این فرم ارزیابی شوند عبارتند از نشانه‌های اختصار، ارتباط و سهولت دسترسی و ابعاد خطای نشان‌دهنده صحت اطلاعات. اگر یک واحد داده را در خطا در نظر بگیریم، متریک به عنوان تعداد واحدهای داده در خطا تقسیم بر تعداد کل واحدها محاسبه شده از ۱ تعریف می‌شود.

کامل بودن، درجه‌ای است که یعنی نهادها و ویژگی‌ها از دست نرفته‌اند که شامل سه نوع می‌باشد:

^۱ objectivity

^۲ relevancy

^۳ reputation

^۴ security

^۵ timeliness

^۶ understandability

^۷ Valu-added

۱- کامل بودن طرح

۲- تکمیل ستون

۳- تکمیل جمعیت

میتواند با در نظر گرفتن نسبت تعداد اقلام ناقص به تعداد کل اقلام و تفریق از ۱ محاسبه کرد.

انسجام ابعاد نیز میتواند از چندین دیدگاه، مورد بررسی قرار گیرد

۱- یکسان بودن انطباق داده‌ها در جداول.

یکپارچگی محدودیت مرجع که یک نمونه از این نوع سازگاری است.

• عملکردهای $\max$ یا $\min$

حداقل یا حداکثر عملیات برای رسیدگی به ابعادی که نیاز به اجماع دارند. شاخص‌های کیفیت داده چندگانه (متغیر)، میتواند حداقل یا حداکثر عملیات باشد.

یک مقدار حداقل (یا حداکثر) را از میان مقادیر نرمال شده شاخص‌های کیفیت داده فردی محاسبه می‌کند. اپراتور $\min$ محافظه کارانه است؛ به این معنی که ابعاد ارزشمندی را که بالاتر از ارزش ضعیف‌ترین شاخص کیفیت آن است (به ارزیابی و نرمال سازی شده بین ۰ تا ۱) اختصاص می‌دهد.

اگر تفسیر ضروری باشد حداکثر عملیات مورد استفاده قرار می‌گیرد. متغیرها میتوانند با استفاده از یک نسبت ساده اندازه‌گیری شوند. دو نمونه جالب از ابعادی که میتوان در ساخت اپراتور $\min$ استفاده کنند باورپذیری و مقدار مناسب داده‌ها هستند.

مقدار داده‌هایی که برای اپراتور $\max$ در معیارهای پیچیده تر در استفاده مفید است ابعاد به روز بودن و دسترسی می‌باشد. باورپذیری به این معنی است که میزان داده‌ها در ست و معتبر در نظر گرفته شود. در میان عوامل دیگر، ممکن است ارزیابی فرد از اعتبار منبع داده در مقایسه با استاندارد معمول پذیرفته شده و تجربه قبلی را منعکس کند. هر یک از این متغیرها در یک مقیاس از ۰ تا ۱ رتبه بندی می‌شود و سپس اعتقاد کلی به عنوان حداقل ارزش سه مورد تعیین می‌شود.

حداقل دو نسبت ساده: نسبت تعداد واحدهای داده است به تعداد واحدهای داده مورد نیاز و نسبت تعداد داده‌های داده شده بر واحدهای مورد نیاز برای تعداد واحد‌های ارائه شده می‌باشد.

به‌روز بودن نشان‌دهنده این است که چگونه اطلاعات مربوط به کار مورد استفاده برای آن به‌روز می‌شود.

که به موقع بودن حداکثر از دو اصطلاح^۱ و یک منهای نسبت ارزش به نوسانات به دست می‌آید. در اینجا، ارزش به عنوان زمان تحویل منهای نوسانات زمان ورودی به طول داده‌های زمان اشاره می‌کند

زمان تحویل اشاره به زمانی دارد که داده‌ها به کاربر تحویل داده می‌شوند. زمان ورودی اشاره به هنگامی که داده‌ها توسط سیستم دریافت می‌شود؛ و سن مربوط به سن داده‌های دریافت شده توسط سیستم یک شاخص در ابتدا است که می‌توان به عنوان یک عامل حساسیت استفاده کرد.

✓ میانگین وزنی^۱

برای مورد چند متغیره، جایگزینی برای اپراتور min است

برای اطمینان از ارزیابی نرمال، هر عامل وزن باید بین صفر و یک باشد و عوامل وزن باید به یک اضافه شوند. اگر شرکت بتواند میزان اهمیت هر یک از متغیرها را به اندازه کلی باورپذیری مشخص کند، میانگین وزنی ممکن است یک فرم مناسب برای استفاده باشد.

۳-۲,۵,۱۲ ارزیابی در عمل^۲

برای استفاده از معیارهای ذهنی و عینی برای بهبود کیفیت داده سازمان، نیاز به سه مرحله دارد

- انجام ارزیابی کیفیت ذهنی و عینی
- مقایسه نتایج ارزیابی‌ها، شناسایی اختلافات و تعیین علل ریشه‌ای اختلافات
- تعیین و انجام اقدامات لازم برای بهبود.

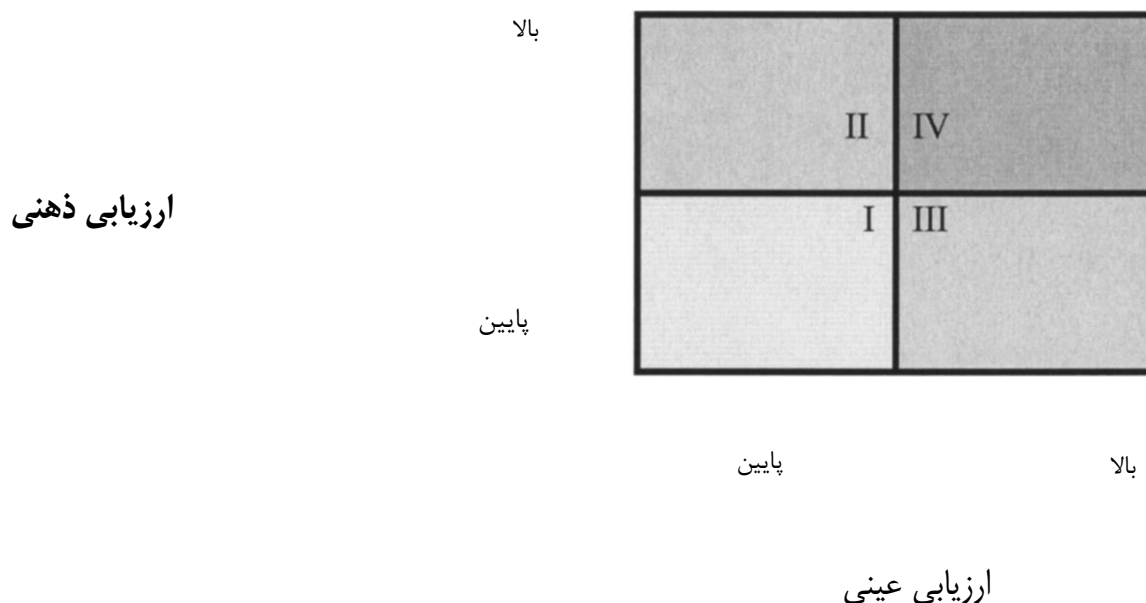
^۱ Weighted Average

^۲ Assessments in Practice

برای شروع تجزیه و تحلیل، ارزیابی ذهنی و عینی یک خاص است بعد مقایسه می‌شود. نتیجه تجزیه و تحلیل به یکی از چهار بخش تقسیم می‌شود اگر تجزیه و تحلیل نشان می‌دهد. این شرکت باید علل ریشه را بررسی و اقدامات اصلاحی انجام دهد. اقدامات اصلاحی برای هر مورد متفاوت خواهد بود.

۳-۲,۵,۳ نتایج ذهنی^۱

ارزیابی‌ها در گروه‌های مختلف نشان می‌دهد که هماهنگی و تکمیل، دو نگرانی اصلی است. هنگامی که این ارزیابی‌ها با ارزیابی عینی داده‌ها مقایسه شد، اندازه‌گیری‌های عینی ارزیابی ذهنی را تأیید می‌کنند. انجام اندازه‌گیری براساس رویکرد ارائه شده در این جا شرکت را قادر می‌سازد تا به طور مداوم نظارت بر هر دو اقدامات عینی و ارزیابی کاربر را داشته باشد و بنابراین برنامه بهبود کیفیت داده خود را نهادینه سازند.



شکل ۳-۱ مقایسه ارزیابی ذهنی و عینی

۲،۶-۳ پژوهش هنریش و همکاران (۲۰۰۹) در خصوص انواع شاخص‌های کیفیت داده

هنریش و همکاران (۲۰۰۹) در قالب یک روش اجرایی معیارهایی را برای ارزیابی کیفیت داده از بعد "رایج بودن" معرفی نمونه و توسعه دادند. در ادامه به برخی از راهکارها و نکاتی که آنها مورد توجه قرار دادند خواهیم پرداخت.

تجربه نشان می‌دهد تعیین مجموعه‌ای از معیارهای فاقد عمومیت یا تک موردی (ad hoc) به تنهایی برای ارزیابی و بهبود کیفیت داده یک راه حل نیست. (پپینو و همکاران ۲۰۰۲) در عوض، ارزیابی کیفیت داده‌ها یک تلاش مستمر است که نیاز به آگاهی از اصول اساسی توسعه دهنده معیارهای کیفیت ذهنی و عینی دارد. در این مقاله، ارزیابی ذهنی و عینی کیفیت داده‌ها، و همچنین نسبت ساده، اپراتورهای حداقل یا حداکثر و فرمول‌های متفاوتی که می‌توانند در ایجاد معیارهای کیفیت داده در عمل کمک کنند، ارائه شده است.

در سال‌های اخیر گسترش استفاده از سیستم‌های انبار داده و سیستم‌های پشتیبانی مدیریت باعث افزایش ارتباطات کیفیت داده‌ها (DQ) در نظریه و عمل شده است. (کاپیلو و همکاران^۱) اهمیت این موضوع به شدت به سود تصمیم‌گیرندگان در جهت "کامل بودن"، "صحیح بودن" و "رایج بودن" داده‌ها مربوط می‌باشد. چنین خصوصیتی به عنوان ابعاد کیفیت شناخته می‌شود (وانگ و همکاران ۱۹۹۷^۲). تاکنون مطالعات بسیاری در مورد هزینه‌ها و مشکلات DQ برای شرکت‌ها انجام شده که نشان‌دهنده ارتباط زیاد کیفیت داده و هزینه‌های مرتبط با آن است. (استرانگ و همکاران ۱۹۹۷)

یک مطالعه توسط PWC نشان داد که تنها ۳۴ درصد از مفسران اطلاعات ادعا می‌کنند که در کیفیت داده‌هایشان "بسیار اطمینان" دارند. (کوپرس. ۲۰۰۴^۳) بر اساس یک نظرسنجی بین‌المللی اخیر در مورد DQ، ۷۵ درصد از همه پاسخ‌دهندگان به دلیل داده‌های نادرست تصمیم اشتباه گرفته‌اند. علاوه بر این، پاسخ‌دهندگان و کارکنان آنها تا ۳۰٪ از وقت کار خود را صرف بررسی صحت داده‌های ارائه شده می‌کنند. (هریس ۲۰۰۶^۴) در مواجهه با چنین مشکلی، شرکت‌های بیشتری خواهان بهبود DQ هستند. با این حال، ضروری است برای تعیین وضعیت کنونی DQ به منظور برنامه‌ریزی مقادیر DQ از یک شیوه اقتصادی استفاده شود. از این رو، روش‌ها و معیارهای تعیین مقدار DQ مورد نیاز است. در بعد رایج بودن DQ چندین تحقیق تجربی نشان می‌دهد که جنبه‌های زمان در

^۱ Cappiello et al. 2003

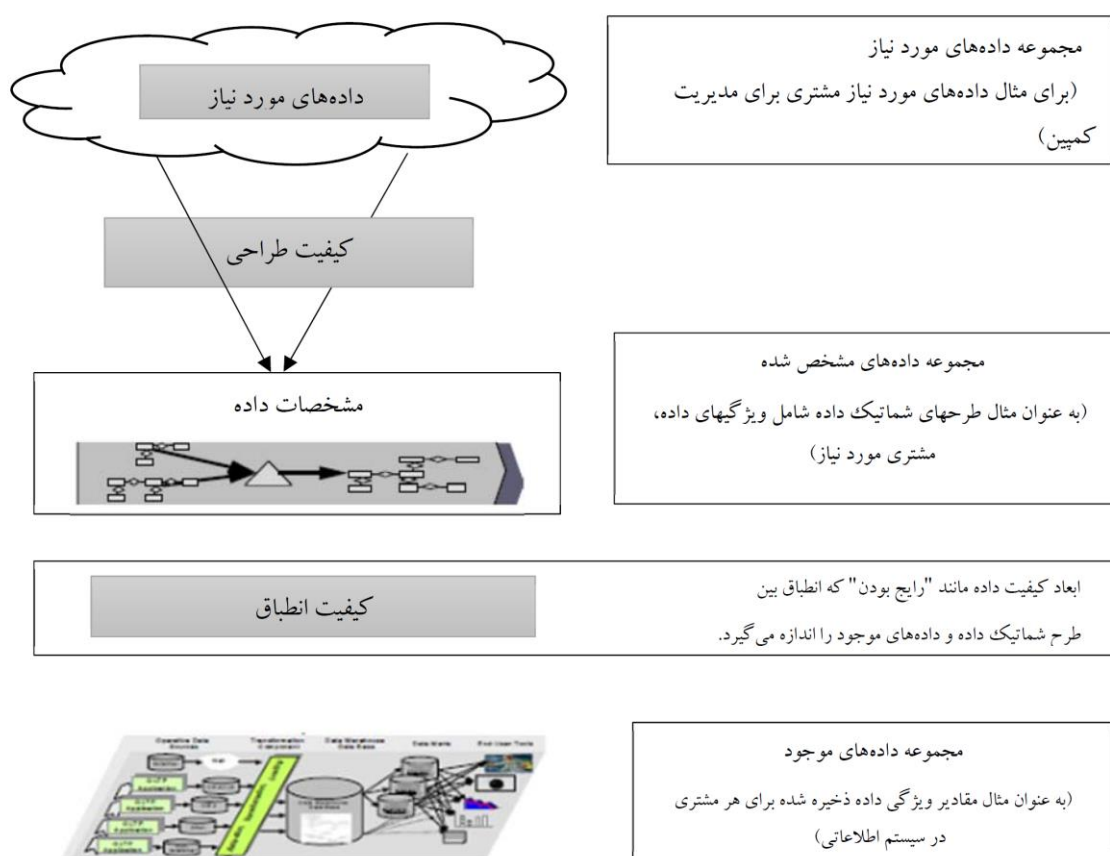
^۲ Wang et al. 199۷

^۳ Price Waterhouse Coopers. 2004

^۴ Harris Interactive 2006

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۷۰

مدیریت DQ بسیار مهم هستند. (یو و وانگ ۲۰۰۷؛ کلاین و کالاهان ۲۰۰۷) لذا در این تحقیق شاخص‌هایی بمنظور توسعه معیار "رایج بودن" معرفی می‌گردند.



شکل (۲-۳) کیفیت طراحی در مقابل کیفیت انطباق^۱

سازمانها بر اساس سطح فعلی کیفیت داده و همچنین با توجه به معیارها و آستانه پذیرش آنها، باید تصمیم بگیرند آیا اقدامات بهبود دهنده (مانند پاک کردن داده‌های کثیف، فراهم کردن داده‌های مناسب از خارج سازمان و

غیره) بایستی صورت گیرد یا خیر. از دیدگاه اقتصادی، تنها اقداماتی که با توجه به هزینه‌ها و مزایا کارآیی لازم را دارند بایستی در این خصوص انتخاب شوند. (بالو و پازار ۱۹۹۵) لذا توسعه معیارهایی برای حمایت از مدیریت اقتصادی کیفیت داده ضروری است. (پی پی نو و همکاران ۲۰۰۲) (کامپانلا ۱۹۹۹؛ هنریش و همکاران ۲۰۰۷)

در ادبیات کیفیت، دو مفهوم اساسی "کیفیت طراحی"^۱ و "کیفیت انطباق"^۲ بیش از سایرین مورد توجه قرار می‌گیرد. (هفلرت و هینریچس ۲۰۰۳)

کیفیت طراحی نشان‌دهنده درجه یا میزان ارتباط بین الزامات کاربران و مشخصات سیستم اطلاعاتی است. (که برای مثال به کمک طرح شماتیک داده مشخص شده است). در مقابل، کیفیت انطباق نشان‌دهنده درجه مطابقت بین مشخصات و سطح تحقق یافته سیستم‌های اطلاعاتی است. (به عنوان مثال کیفیت انطباق بیانگر این است که داده‌های موجود و ذخیره شده در سیستم اطلاعاتی به چه میزان با طرح شماتیک داده منطبق است). شکل (۳-۲) این دو مفهوم را نشان می‌دهد.

۳-۱، ۲، ۱۲ الزامات معیارهای کیفیت داده

هنریش و همکاران نشان‌الزاماتی را معرفی کردند که به منظور ایجاد اطمینان از مناسب بودن شاخص‌های کیفیت داده تعریف شده می‌توانند مورد استفاده قرار بگیرند. در ادامه این الزامات معرفی می‌شوند.

- [قابلیت نرمال سازی^۴] این الزام برای اطمینان از قابل مقایسه بودن مقادیر یک شاخص در حوزه‌ها یا مقاطع زمانی مختلف ضروری است. به عنوان مثال برای مقایسه سطوح مختلف DQ در طول زمان (پی پی نو و همکاران ۲۰۰۲) از معیارهای DQ اغلب یک مقدار بین ۰ (کاملاً بد) و ۱ (کاملاً خوب) تعریف می‌شوند. (پی پی نو و همکاران ۲۰۰۶) و (شانکاران ۲۰۰۷)
- [مقیاس فاصله‌ای^۶] بمنظور پایش شاخص‌های کیفیت داده در طول زمان و ارزیابی اقتصادی از اقدامات انجام گرفته لازم است شاخص‌های تعیین شده از مقیاس فاصله‌ای برخوردار باشند. به این معنا که، تفاوت

^۱ Quality of Design

^۲ Quality of Conformance

^۳ Helfert and Heinrich 2003

^۴ normalization

^۵ Pipino et al. 2002

^۶ Interval scale

بین دو سطح کیفیت داده در یک شاخص باید معنی‌دار باشد. بنابراین، به عنوان مثال، تفاوت ۰,۲ از مقادیر ۰,۷ و ۰,۹ و مقادیر ۰,۴ و ۰,۶ پس از اندازه‌گیری یک شاخص بایستی معنای مشابهی را داشته باشد.

- [تفسیرپذیری^۱] الزام "تفسیر آسان توسط کاربران کسب و کار" توسط برخی نویسندگان بیان شده است. به بیان دیگر، ارزش معیارهای DQ باید به درستی توسط کاربران درک شود. (شانکاران ۲۰۰۷)
- [تجمع‌پذیری^۳] در صورت استفاده از یک مدل داده رابطه‌ای، معیارها باید انعطاف‌پذیری کافی داشته باشند. (شانکاران ۲۰۰۷) به عبارتی، معیارها باید به گونه‌ای تعریف شوند که بتوان آنها را از یک سطح پایین پایگاه داده به سطح بالاتر تعمیم داد و در عین حال مقایسه نمود. بعنوان مثال اگر یک شاخص کیفیت داده در سطح جدول تعریف می‌شود بتوان عدد بدست آمده را در سطح پایگاه داده (که مجموعه‌ای از جداول به هم پیوسته است) تعمیم داد و نتایج آن را مقایسه نمود. همچنین این الزام نشان می‌دهد که بایستی بتوان نتایج بدست آمده از جداول یک پایگاه داده را مقایسه و در صورت لزوم جمع نمود.
- [سازگاری^۴] معیارهای تعیین شده در حوزه کیفیت داده بایستی در زمینه‌ای که مورد استفاده قرار می‌گیرند قابلیت سازگاری داشته باشند. (اون و شانکاران ۲۰۰۷)
- [امکان‌سنجی^۶] (سهولت محاسبه) برای اطمینان از عملی بودن، معیارها باید بر اساس پارامترهای ورودی تعیین شود. هنگام تعریف معیارها، بایستی روش درست تعیین پارامترهای ورودی تعریف شود. اگر تعیین دقیق پارامترهای ورودی امکان‌پذیر نباشد یا هزینه آن زیاد باشد، روش‌های متداول جایگزین (مانند روش‌های آماری) پیشنهاد می‌شود. به عبارتی بایستی بتوان شاخص کیفیت داده را بگونه‌ای تعیین کرد که امکان خودکارسازی اندازه‌گیری آن به راحتی فراهم شود.

همانطور که در مقدمه عنوان شد این شش الزام بمنظور ارزیابی شاخص‌های مختلف کیفیت داده می‌توانند مورد استفاده قرار گیرند.

^۱ Interpretability

^۲ Shankaranarayanan 2007

^۳ Aggregation

^۴ Adaptivity

^۵ Even and Shankaranarayanan 2007

^۶ Feasibility

در ادامه به بررسی ادبیات مربوط به شاخص رایج بودن در مقالات مختلف خواهیم پرداخت. جدول (۳-۳) برخی تعاریف برگرفته شده از مراجع مختلف در خصوص شاخص رایج بودن ارائه می‌کند.

جدول ۳-۳ تعاریف انتخاب شده برای بعد رواج

تعریف شاخص	مرجع
به روز بودن: "منقضی نبودن مقادیر ثبت شده. منقضی بودن یک داده از این جهت که خطایی در آن رخ داده است که با مقدار فعلی (صحیح) متفاوت است."	بالو و پازر ۱۹۸۵
به موقع بودن: "میزانی که عمر داده برای کار فعلی مناسب است."	وانگ و استرانگ ۱۹۹۶
رواج: به میزان به روز بودن داده اطلاق می‌شود. به روز بودن داده به این معناست که داده در گذر زمان و با وجود هر نوع استهلاکی که ممکن است به آن دچار شود کماکان صحیح باقی بماند	ردمان ۱۹۹۶
به موقع بودن: "میزان اطلاعاتی که برای کار در دست به اندازه کافی روزآمد است"	پی پی نو ۲۰۰۲

یکی از مهمترین تعاریف مربوط به رواج توسط بالو و پازر در سال ۱۹۸۵ ارائه شده است. همانگونه که در شکل (۳-۳) مورد بررسی قرار گرفت رواج یکی از شاخص‌های کیفیت انطباق می‌باشد که در این مقاله معرفی گردیده

و توسعه داده شده است. بالو و پازر رواج را بعنوان یکی از ابعاد کیفیت داده اندازه‌گیری نمودند و شاخصی کمی بمنظور ارزیابی این بعد از کیفیت داده توسعه دادند. شاخص رواج میزان به روز بودن یک ویژگی (Attribute) از داده‌های ذخیره شده در یک سیستم اطلاعاتی را به صورت کمی اندازه‌گیری می‌کند. به بیان دیگر این شاخص به این پرسش پاسخ می‌دهد که یک ویژگی از داده‌ها که در زمان ذخیره شدن در سیستم اطلاعاتی صحیح بوده است، آیا در زمان محاسبه شاخص رواج هنوز هم صحت خود را در فضای واقعی حفظ کرده است یا خیر؟

در ادامه چهار رابطه اصلی که توسط هینریچس^۱ (۲۰۰۲)، بالو و همکاران^۲ (۱۹۹۸)، اون و شانکاران^۳ (۲۰۰۷) و هینریچس و همکاران^۴ (۲۰۰۷) در خصوص کمی کردن شاخص رواج توسعه داده شده‌اند ارائه شده است.

هینریچس در سال ۲۰۰۲ شاخص رواج را در قالب فرمول زیر محاسبه کرد.

$$(۴۷-۳) \quad \text{Currency} = \frac{1}{(\text{mean attribute update frequency}) - (\text{age of attribute value}) + 1}$$

در این رابطه پارامتر mean attribute update frequency میزان فرکانس به روز رسانی ویژگی را نشان می‌دهد. (مثلاً ۱۰ بار در سال) از سوی دیگر پارامتر age of attribute value نشانگر عمر یک ویژگی در پایگاه داده است. به بیان دیگر این پارامتر نشان می‌دهد که فاصله زمانی میان ذخیره شدن یک ویژگی در پایگاه داده و زمان اندازه‌گیری شاخص کیفیت داده چقدر بوده است.

بالو و همکاران تعریف شاخص رواج را بصورت زیر ارائه کردند. (بالو و همکاران ۱۹۹۸)

$$(۴۸-۳) \quad \text{Currency} = \max \left[\left(1 - \frac{\text{age of attribute value}}{\text{shelf life}} \right) \right]^s$$

^۱ Hinrichs 2002

^۲ Ballou et al 1998

^۳ Even and Shankaranarayanan 2007

^۴ Heinrich et al 2007

در این رابطه پارامتر *age of attribute value* مشابه رابطه قبلی است و شاخص *shelf life* نشان‌دهنده میزان نو سانات صورت گرفته در یک ویژگی در گذر زمان است. پارامتر *S* نیز توسط خبرگان و با توجه به نوع کاربرد داده تعیین خواهد شد.

اون و شانکاران^۱ در سال ۱۹۹۷ رابطه زیر را بمنظور ارزیابی شاخص رواج توسعه دادند.

$$1 - (age\ of\ attribute\ value / T_{max})^S \quad (49-3)$$

در این رابطه پارامتر *Tmax*، حداکثر دوره عمر یک ویژگی از داده است و شاخص *S* یک عدد حقیقی مثبت است که بر اساس نظر خبرگان و متناسب با کاربرد این رابطه تعیین می‌شود.

شاخص چهارم که توسط هیریچنس در سال ۲۰۰۷ ارائه شد بر اساس نظریه احتمالات شکل گرفته است و میزان احتمال به روز بودن داده را بعنوان شاخص رواج توصیف کرده است. هینریگز در تعیین این شاخص پیش فرض توزیع نمایی را در تعیین فرمول لحاظ کرده است. در رابطه توسعه داده شده که در ادامه ارائه شده است دو پارامتر اصلی بایستی محاسبه شوند. پارامتر اول که با عنوان *decline(A)* معرفی شده است همان نرخ به روز بودن ویژگی *A* از مجموعه داده است که با توجه به نمایی بودن توزیع عمر داده محاسبه می‌شود. همچنین در این رابطه *age(w,A)* میزان فاصله زمانی میان جمع‌آوری ویژگی *A* از داده‌ها و لحظه محاسبه رابطه است.

$$\exp(-decline(A) \cdot age(w,A)) \quad (50-3)$$

در ادامه و در جدول (۴-۳) هر یک از ۴ رابطه اصلی رواج از منظر ۶ ویژگی اصلی که در این بخش برای شاخص‌های کیفیت داده معرفی شده اند مورد مقایسه قرار گرفتند.

جدول (۳-۴) ارزیابی معیارهای موجود شاخص رایج (هینریچس و همکاران^۱)

هینریچس و همکاران ۲۰۰۷	اون و شانکاران ۱۹۹۷	بالو و همکاران ۱۹۹۸	هینریچس ۲۰۰۲	
$\exp(-\text{decline}(A) \cdot \text{age}(w.A))$	$1 - (\text{age of attribute value} / T_{\max})^s$	$\left\{ \max \left[\left(1 - \frac{\text{age of attribute value}}{\text{shelf life}} \right), 0 \right] \right\}^s$	$\frac{1}{(\text{mean attribute update frequency}) + (\text{age of attribute value}) + 1}$	صحت تعریف تکنیک‌ها
کاربرد دارد	کاربرد دارد	کاربرد دارد	کاربرد دارد	قابلیت نرمال سازی
کاربرد دارد	نسبتاً کاربرد دارد	نسبتاً کاربرد دارد (for s = 1)	کاربرد ندارد	مقیاس پذیری
کاربرد دارد	نسبتاً کاربرد دارد	نسبتاً کاربرد دارد (for s = 1)	کاربرد ندارد	تفسیرپذیری
کاربرد دارد	کاربرد دارد	کاربرد ندارد	کاربرد دارد	تجمیع پذیری
نسبتاً کاربرد دارد	کاربرد دارد	نسبتاً کاربرد دارد	نسبتاً کاربرد دارد	سازگاری
کاربرد دارد	نسبتاً کاربرد دارد	کاربرد ندارد	کاربرد دارد	امکان سنجی

^۱ Heinrich, M. Kaiser, and M. Klier

۳-۳ توسعه شاخص‌های کیفیت داده در ارزیابی فرایند ثبت

با توجه به مطالعات انجام گرفته در گذشته، بررسی فرمول‌ها، سوابق، مقالات و با در نظر داشتن تمام ابعاد شاخص‌های کیفیت، حال به شاخص‌های کیفی سامانه ثبت می‌پردازیم تا تمام ابعاد را بررسی و شاخص‌های مهم را برای رسم نمودار کنترل استخراج کنیم.

در ابتدا هر کدام از شاخص‌ها را به تفکیک بررسی کرده و جدول شناسنامه شاخص را جداگانه رسم می‌کنیم.

✓ جامعیت

✓ صحت

✓ به هنگام بودن

✓ سازگاری

۳-۱-۳ جامعیت

معیار متریک جامعیت با تعدادی فیلدهای تکمیل شده در یک رکورد فراداده، بررسی می‌شود. با توجه به این معیار یکی از شاخص‌های اصلی در سامانه ثبت معیار جامعیت یا تمامیت می‌باشد که میزان کامل بودن (پر شده) رکوردها توسط دانشجویان را اندازه‌گیری می‌کند.

جدول (۳-۵) شناسنامه شاخص کامل بودن

شناسنامه شاخص‌های کیفیت داده در سامانه ثبت	
نام شاخص	کد شاخص <i>DQ – Ganj – 01</i>
میزان کامل بودن فیلدها در سامانه ثبت (جامعیت)	
فرمول شاخص	$Completeness(j) = \frac{\sum_{i=1}^n Per_Com(i)}{n}$
دوره اندازه‌گیری شاخص ماه‌یانه	فاصله نمونه برداری

معرفی متغیرها
Completeness(j) = میزان کامل بودن نمونه‌ها (پایان‌نامه‌ها) در نمونه یا دوره ژام
n = تعداد پایان‌نامه‌ها در هربار نمونه‌برداری
Per_Com (i): شاخص کامل بودن پایان‌نامه‌ها در نمونه آم. این شاخص از نسبت تعداد فیلدهای پر شده در پایان‌نامه آم به کل فیلدهای موجود در سامانه ثبت بدست می‌آید.

۳-۲-۳ صحت

صحت یک رکورد فراداده است که بیانگر درستی مقدار فیلد با توجه به منابع است. در سامانه ثبت این فاصله معنایی، نسبت فیلدهای صحیحی است که یک دانشجو وارد کرده به کل فیلدهای پر شده توسط همان شخص. که در واقع این شاخص بر دو اساس درصد^۱ و تعداد^۲ محاسبه می‌شود.

جدول (۳-۶) شناسنامه شاخص صحت

شناسنامه شاخص‌های کیفیت داده در سامانه ثبت	
نام شاخص	کد شاخص
نسبت تعداد فیلدهای صحیح به کل فیلدهای تکمیل شده (صحت)	DQ – Ganj – 02
فرمول شاخص	
$Per - Accuracy(j) = \frac{\sum_{i=1}^n Per_Acc(i)}{n}$	
$Con - Accuracy(j) = \frac{\sum_{i=1}^n Con_Acc(i)}{n}$	
دوره اندازه‌گیری شاخص	فاصله نمونه برداری
ماه‌یانه	

^۱ percentage

^۲ count

معرفی متغیرها

Per_Acc(i) = درصد فیلدهای صحیح پایان نامه آم به کل فیلدهای تکمیل شده توسط دانشجو

Con_Acc(i) = تعداد فیلدهای صحیح پایان نامه آم به کل فیلدهای تکمیل شده توسط دانشجو

n = تعداد پایان‌نامه‌ها در هر بار نمونه‌برداری

۳-۳-۳ به هنگام بودن

به هنگام بودن به طور خاص مربوط به درجه‌ای است که یک رکورد فراداده در میان جامعه خاص در حال حاضر جریان دارد. در واقع با اندازه‌گیری و محاسبه می‌توان فهمید که یک فراداده چگونه با گذشت زمان باز هم می‌تواند مفید و به روز باشد.

نقش این شاخص در سامانه ثبت اندازه‌گیری فاصله زمانی بین دفاع دانشجویان و ثبت پایان‌نامه آنها می‌باشد.

جدول (۳-۶) شناسنامه شاخص به هنگام بودن

شناسنامه شاخص‌های کیفیت داده در سامانه ثبت	
نام شاخص	کد شاخص
فاصله زمانی بین دفاع و ثبت پایان‌نامه (به هنگام بودن)	DQ – Ganj – 03
فرمول شاخص	
$Timeliness(j) = \frac{\sum_{i=1}^n Dis_Tim(i)}{n}$	

دوره اندازه‌گیری شاخص ماهیانه	فاصله نمونه برداری
معرفی متغیرها $Dis-Tim(i)$ = فاصله زمانی بین دفاع دانشجو و ثبت پایان نامه توسط دانشجو در پایان‌نامه آم n = تعداد پایان‌نامه‌ها در هربار نمونه‌برداری	

۳-۴ سازگاری

شاخص سازگاری میزان انطباق یک رکورد فراداده با یک تعریف استاندارد را نشان می‌دهد. در واقع میزان درستی اطلاعات وارد شده توسط دانشجویان را از نظر محتوایی در سامانه می‌سنجد و نسبت آن را به کل فیلدها را محاسبه می‌کند. همان‌طور که در ذیل نوشته شده سازگاری بر دو نوع است.

(۱) درون رابطه‌ای^۱

(۲) برون رابطه‌ای^۲

^۱ inter

^۲ intra

جدول (۳-۶) شناسنامه شاخص سازگاری

شناسنامه شاخص‌های کیفیت داده در سامانه ثبت	
نام شاخص	کد شاخص <i>DQ – Ganj – 04</i>
سازگاری (درون رابطه ای ، برون رابطه ای)	
فرمول شاخص	
$Inter - Consistency(j) = \frac{\sum_{i=1}^n Inter - Con(i)}{n}$ $Intra - Consistency(j) = \frac{\sum_{i=1}^n Intra - Con(i)}{n}$	
دوره اندازه‌گیری شاخص ماه‌بانه	فاصله نمونه برداری
معرفی متغیرها	
Inter-con (i) = تعداد سازگاریهای درون رابطه ای در نمونه (پایان نامه آم) Intra-con (i) = تعداد سازگاریهای برون رابطه ای (پایان نامه آم) n = تعداد پایان‌نامه‌ها در هر بار نمونه‌برداری	

۳-۴ تعیین درجه اهمیت شاخص‌های کیفیت داده

در این بخش بر مبنای مقدمات معرفی شده در انتهای فصل دوم به کمک روش FDANP به تعیین درجه اهمیت ابعاد مختلف کیفیت داده خواهیم پرداخت.

۳-۴-۱ تحلیل توصیفی وضعیت جمعیت شناختی

برای جمع آوری اطلاعات خام (دیدگاه و نظرات جامعه آماری) اقدام به توزیع پرسشنامه بین خبرگان در زمینه مدیریت کیفیت داده نمودیم که تعداد ۱۰ پرسشنامه تکمیل و عودت داده شد. به منظور دستیابی به مشخصات کلی

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۸۲

آزمودنی‌ها، سئوالات قسمت ابتدایی پرسشنامه به مشخصات فردی پاسخگویان اختصاص یافته است. بنابراین در ابتدای فصل حاضر، ابتدا به منظور آشنایی بیشتر با مشخصات عمومی نمونه آماری، جداول توزیع فراوانی و شاخص‌های توصیفی این متغیرها به همراه نمودارهای مربوط ارائه شده است.

۴,۲-۳ جنسیت

همانطور که در جدول ۷-۳ مشاهده می‌شود از ۱۰ نفری که به عنوان خبره انتخاب شده‌اند، ۷ نفر زن و ۳ نفر مرد می‌باشند که در مجموع ۷۰ درصد خبرگان را زنان و ۳۰ درصد خبرگان را مردان تشکیل می‌دهند.

جدول (۷-۳) فراوانی جامعه آماری بر حسب جنسیت

فراوانی تجمعی	درصد	فراوانی	جنسیت
۳۰	۳۰	۳	مرد
۱۰۰,۰	۷۰	۷	زن
۱۰۰,۰	۱۰۰,۰	۱۰	مجموع

۴,۳-۳ سن

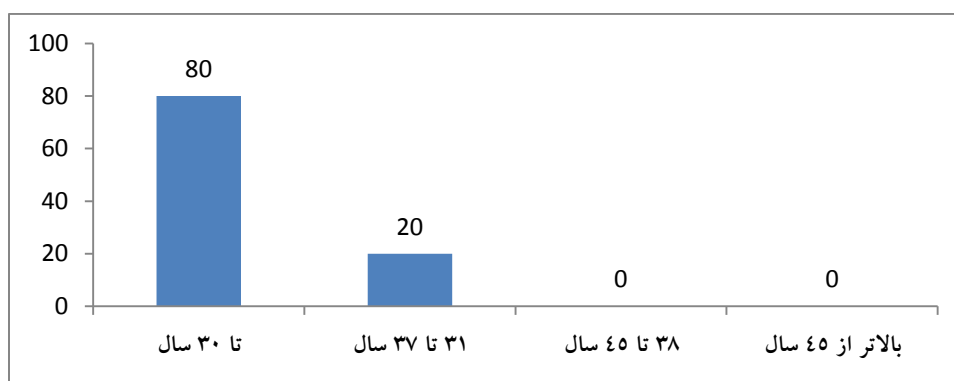
جدول ۸-۳ که خروجی مربوط به سن افراد خبره می‌باشد نشان می‌دهد که توزیع فراوانی گروه‌های مختلف سنی در پاسخ دهنده‌ها با چه وزن و فراوانی می‌باشد. فراوانی گروه سنی تا ۳۰ سال در نمونه مورد بررسی از سایر گروه‌های سنی بیشتر می‌باشد که با فراوانی ۸ و درصد فراوانی ۸۰ بیشترین درصد را به خود اختصاص داده‌اند. همچنین گروه سنی ۳۱ تا ۳۷ سال فراوانی ۲ و درصد فراوانی ۲۰ را به خود اختصاص داده‌اند.

جدول ۸-۳ فراوانی جامعه آماری بر حسب سن

فراوانی تجمعی	درصد فراوانی	فراوانی	بازه سنی
۸۰	۸۰	۸	تا ۳۰ سال

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۸۳

۳۷-۳۱	۲	۲۰	۱۰۰
۴۵-۳۸	۰	۰	۱۰۰
بالاتر از ۴۵	۰	۰	۱۰۰
مجموع	۱۰	۱۰۰	



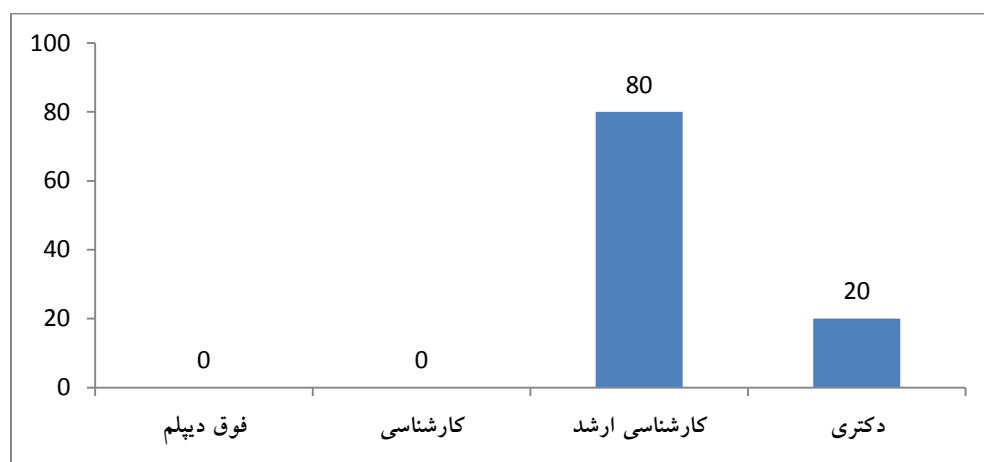
نمودار ۳-۱ توزیع فراوانی افراد خبره بر حسب سن افراد

۴-۴-۳ تحصیلات

مشاهده جدول زیر نشان می‌دهد که از مجموع ۱۰ نفر افراد خبره ۸ نفر کارشناس ارشد می‌باشند که ۸۰ درصد از حجم نمونه را به خود اختصاص داده است. همچنین ۲ نفر از افراد خبره دارای مدرک دکتری بوده که ۲۰ درصد از حجم نمونه را به خود اختصاص داده‌اند.

جدول ۳-۹ توزیع فراوانی تحصیلات افراد خبره

تحصیلات	فراوانی	درصد فراوانی	فراوانی تجمعی
فوق دیپلم	۰	۰	۰
کارشناسی	۰	۰	۰
کارشناسی ارشد	۸	۸۰	۸۰
دکتری	۲	۲۰	۱۰۰
مجموع	۱۰	۱۰۰	



نمودار ۲-۳ توزیع فراوانی تحصیلات افراد خبره

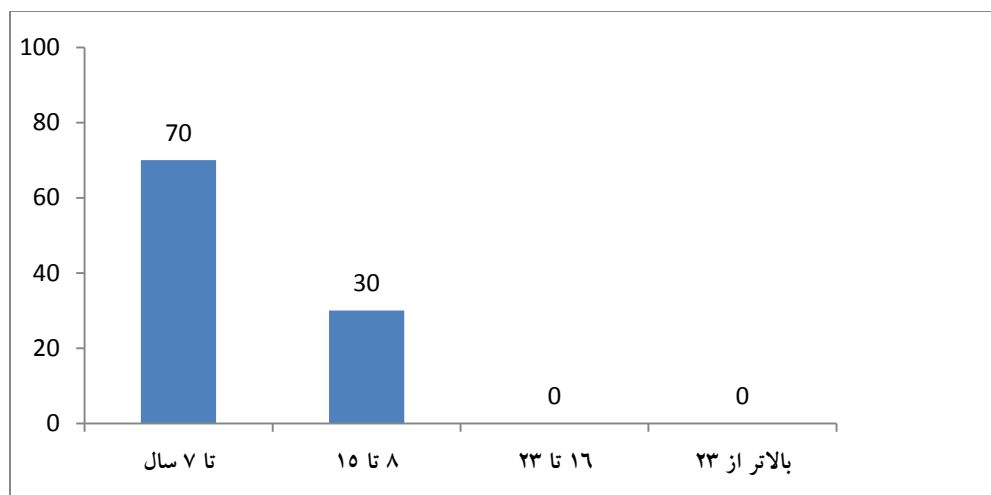
۴-۵-۳ سابقه کار افراد خبره

سابقه کار افراد مورد مطالعه در اینجا به تفکیک در قالب جدول زیر ارائه شده است. مشاهده می‌شود که سابقه کاری تا ۷ سال با فراوانی ۷ و درصد فراوانی ۷۰ بالاترین میزان فراوانی را در میان افراد نمونه به خود اختصاص داده است. همچنین سابقه کاری ۸ تا ۱۵ سال نیز درصد فراوانی ۳۰ درصد را به خود اختصاص داده است.

جدول ۱۰-۳ توزیع فراوانی سابقه کار افراد نمونه

سابقه کاری	فراوانی	درصد فراوانی	فراوانی تجمعی
تا ۷ سال	۷	۷۰	۷۰
۸ تا ۱۵ سال	۳	۳۰	۱۰۰
۱۶ تا ۲۳ سال	۰	۰	۱۰۰
بالاتر از ۲۳	۰	۰	۱۰۰
مجموع	۱۰	۱۰۰	

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۸۵



نمودار ۳-۳ پراکندگی سابقه کار افراد خبره

۳-۶، ۴ وضعیت توصیفی مقیاس‌های پرسشنامه

با توجه به ادبیات موضوع و پیشینه موضوع تحقیق، عوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور و همچنین زیرعوامل مربوطه به شرح جدول ۳-۱۱ می‌باشد.

جدول ۳-۱۱ عوامل و زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور

ردیف	ارکان (عامل)	علائم اختصاری	زیرعامل	علائم اختصاری
۱	عامل نمایشگری	C1	تفسیرپذیری	SC11
			سهولت درک	SC12
			نمایش مختصر	SC13
			نمایش سازگار	SC14
۲	عامل متنی	C2	مربوط بودن	SC21
			ارزش افزوده	SC22

SC23	به روز بودن			
SC24	جامعیت			
SC25	مقدار داده ها			
SC31	در دسترس بودن	C3	عامل دسترس پذیری	۳
SC32	امنیت			
SC41	عینیت	C4	عامل ذاتی	۴
SC42	دقت			
SC43	معتبر بودن			
SC44	باورپذیری			

برای شناسایی میزان اهمیت زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور، از پرسشنامه ای شامل ۵ مقیاس در قالب ۱۵ سوال نظر خبرگان جامعه آماری را در خصوص تعیین میزان اهمیت زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور مورد بررسی قرار داده ایم.

۴,۷-۳ ارزیابی اولیه پرسشنامه ها با استفاده از روش امتیاز دهی

بعد از گردآوری پرسشنامه ها از آنجا که پاسخ ها در پرسشنامه به صورت طیف کیفی فازی پنج گزینه ای از بسیار مهم تا بی اهمیت (اعداد فازی مثلثی) قید شده بود و تعداد ۱۰ نفر به پرسش ها پاسخ داده بودند، لازم بود که در ابتدا میانگین حسابی هر کدام از زیرعوامل محاسبه گردد. پس از این کار زیرعواملی که میانگین آن ها از عدد ۵ کمتر بود، حذف گردیدند. بعد از روش امتیازدهی تعداد ۱ زیرعامل حذف گردید و مدل اولیه تحقیق با ۱۴ زیرعامل تشکیل گردید.

مرحله اول: محاسبه میانگین پاسخ خبرگان

جدول ۳-۱۲ ارزیابی زیرعوامل توسط خبرگان

	خیلی زیاد	زیاد	متوسط	کم	خیلی کم	
میانگین	۹	۷	۵	۳	۱	زیرعوامل
۷,۸	۵	۴	۱	۰	۰	تفسیرپذیری
۷,۸	۵	۴	۱	۰	۰	سهولت درک
۷,۴	۳	۶	۱	۰	۰	نمایش مختصر
۷,۲	۶	۲	۰	۱	۱	نمایش سازگار
۷,۴	۴	۴	۲	۰	۰	مربوط بودن
۷,۴	۵	۲	۳	۰	۰	ارزش افزوده
۷,۴	۵	۲	۳	۰	۰	به روز بودن
۷,۲	۴	۳	۳	۰	۰	جامعیت
۴,۶	۰	۰	۸	۲	۰	مقدار داده ها
۷	۰	۱۰	۰	۰	۰	در دسترس بودن
۷,۸	۵	۴	۱	۰	۰	امنیت
۹	۱۰	۰	۰	۰	۰	عینیت
۸,۴	۸	۱	۱	۰	۰	دقت
۷,۴	۵	۳	۱	۱	۰	معتبر بودن
۷,۸	۵	۴	۱	۰	۰	باورپذیری

همانگونه که ملاحظه می‌گردد تعداد ۱ زیرعامل از زیرعوامل موثر بر کیفیت قرارداد های سامانه ملی ثبت پایان نامه داخل کشور توسط خبرگان نامناسب تشخیص داده شدند (مقدار میانگین زیر ۵). به همین دلیل این زیرعامل حذف گردیده و زیرعوامل نهایی انتخاب شده از دید خبرگان مطابق جدول ۳-۱۳ به شرح زیر تعیین گردید.

جدول ۳-۱۳ جدول نهایی عوامل و زیرعوامل انتخاب شده توسط خبرگان

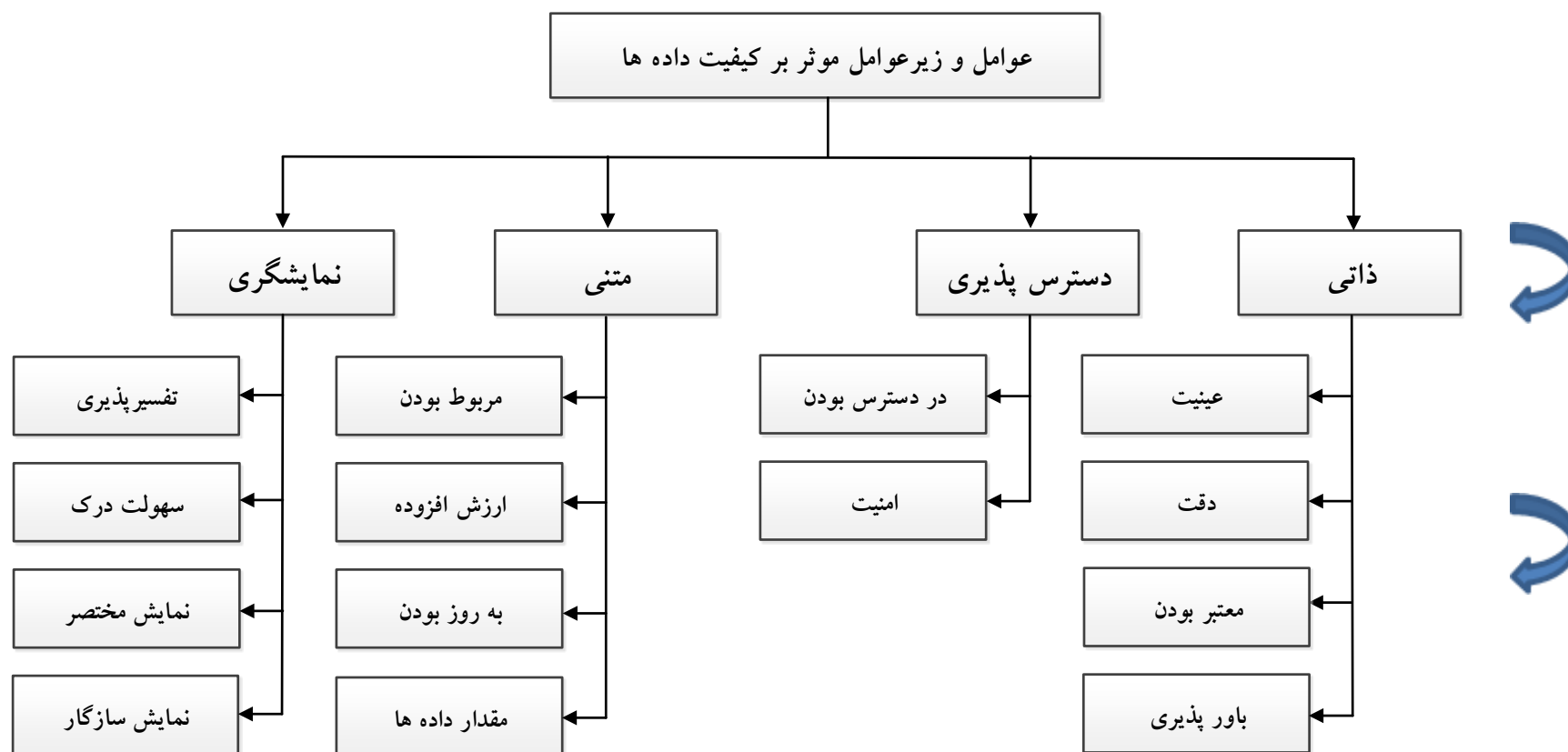
ردیف	ارکان(عامل)	علائم اختصاری	زیرعامل	علائم اختصاری
۱	عامل نمایشگری	C1	تفسیر پذیری	SC11
			سهولت درک	SC12
			نمایش مختصر	SC13
			نمایش سازگار	SC14
۲	عامل متنی	C2	مربوط بودن	SC21
			ارزش افزوده	SC22
			به روز بودن	SC23
			جامعیت	SC24
۳	عامل دسترس پذیری	C3	در دسترس بودن	SC31
			امنیت	SC32
۴	عامل ذاتی	C4	عینیت	SC41
			دقت	SC42
			معتبر بودن	SC43

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۸۹

SC44	باورپذیری			
------	-----------	--	--	--

۴-۸-۳ مدل شبکه‌ای نهایی تحقیق

با توجه به مراحل تعیین اهمیت و شدت تاثیر عوامل و زیرعوامل بر هم توسط خبرگان، مدل نهایی تحقیق به شرح نمودار زیر می‌باشد.



نمودار ۳-۴ مدل شبکه‌ای نهایی تحقیق

۴-۹-۳ فرآیند تکنیک دیمتل فازی بر مبنای ANP فازی

در این تحقیق به منظور وزن دهی و رتبه بندی عوامل و زیرعوامل موثر بر کیفیت فراداده های سامانه ملی ثبت پایان نامه داخل کشور از تکنیک دیمتل فازی بر مبنای ANP فازی استفاده شده است. بدین منظور برای مقایسه زیرعوامل از ۵ عبارت کلامی استفاده شده است که طیف فازی و معادل کلامی آن در جدول ۳-۱۴ نشان داده شده است.

جدول ۳-۱۴ متغیرهای زبانی و اعداد فازی متناظر (زو و همکاران، ۲۰۱۱)

متغیر زبانی	معادل قطعی	اعداد فازی مثلثی متناظر
بی تأثیر	0	25)0.0,0,(0
تأثیر خیلی کم	1	25,0.5).(0,0
تأثیر کم	2	75)25,0.5,0,(0
تأثیر زیاد	3	75,1).5,0,(0
تأثیر خیلی زیاد	4	75,1,1).(0

۴-۹-۱-۳ گام نخست: محاسبه ماتریس ارتباط مستقیم (D)

در این گام از پاسخ دهندگان خواسته شد تا میزان تأثیرگذاری زیرعامل $\bar{A}$ بر زیرعامل $\bar{J}$ را با استفاده جدول ۳-۱۴ نشان دهند. برای بررسی زیرعوامل از نظر ۱۰ خبره استفاده شد. در این ماتریس‌ها، $\tilde{x}_{ij} = (l_{ij}, m_{ij}, u_{ij})$ اعداد فازی مثلثی می‌باشند. برای در نظر گرفتن نظر همه خبرگان طبق رابطه (۳-۵۱)، از آن‌ها میانگین حسابی گرفته می‌شود.

$$\bar{\tilde{z}} = \frac{\tilde{x}^1 \oplus \tilde{x}^2 \oplus \tilde{x}^3 \oplus \dots \oplus \tilde{x}^p}{p} \quad ۵۱-۳$$

در این رابطه p تعداد خبرگان و $\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^p$ به ترتیب ماتریس مقایسه زوجی خبره ۱، خبره ۲ و خبره p می‌باشد و $\bar{\tilde{z}}$ عدد فازی مثلثی به صورت $\bar{\tilde{z}}_{ij} = (l'_{ij}, m'_{ij}, u'_{ij})$ است. میانگین نظرات خبرگان تحقیق در جدول ۳-۱۵، قابل مشاهده است.

جدول ۳-۱۵ ماتریس ارتباط مستقیم ($\tilde{D}$) (میانگین نظر تمام خبرگان)

	SC11	SC12	SC13	SC14	SC21	SC22	SC23	SC24
SC11	(0,0,0.25)	(0.7,0.95,1)	(0.45,0.7,0.85)	(0.3,0.55,0.8)	(0.4,0.65,0.9)	(0.3,0.5,0.7)	(0.2,0.45,0.65)	(0.45,0.7,0.9)
SC12	(0.5,0.75,0.95)	(0,0,0.25)	(0.35,0.55,0.8)	(0.15,0.4,0.65)	(0.35,0.6,0.8)	(0.05,0.3,0.55)	(0.15,0.25,0.45)	(0.2,0.4,0.65)
SC13	(0.45,0.7,0.9)	(0.6,0.85,0.95)	(0,0,0.25)	(0.3,0.55,0.75)	(0.4,0.65,0.85)	(0.05,0.25,0.5)	(0,0.1,0.35)	(0.35,0.6,0.75)
SC14	(0.35,0.6,0.8)	(0.5,0.75,0.9)	(0.3,0.5,0.7)	(0,0,0.25)	(0.5,0.75,0.95)	(0.15,0.35,0.6)	(0.15,0.3,0.55)	(0.35,0.6,0.8)
SC21	(0.35,0.6,0.85)	(0.4,0.65,0.8)	(0.25,0.45,0.7)	(0.4,0.65,0.9)	(0,0,0.25)	(0.25,0.5,0.75)	(0.3,0.55,0.75)	(0.55,0.8,0.95)
SC22	(0.1,0.35,0.6)	(0.15,0.35,0.6)	(0.15,0.35,0.6)	(0.2,0.4,0.65)	(0.15,0.4,0.65)	(0,0,0.25)	(0.3,0.55,0.8)	(0.3,0.55,0.75)
SC23	(0.15,0.4,0.65)	(0.3,0.55,0.75)	(0.2,0.4,0.65)	(0.1,0.3,0.55)	(0.1,0.35,0.6)	(0.3,0.55,0.75)	(0,0,0.25)	(0.35,0.6,0.75)
SC24	(0.5,0.75,0.95)	(0.6,0.85,1)	(0.3,0.5,0.75)	(0.4,0.65,0.85)	(0.25,0.5,0.75)	(0.3,0.5,0.7)	(0.3,0.55,0.8)	(0,0,0.25)
SC31	(0.25,0.45,0.7)	(0.3,0.5,0.7)	(0.1,0.3,0.55)	(0.15,0.4,0.65)	(0.1,0.25,0.5)	(0.15,0.35,0.6)	(0.3,0.5,0.75)	(0.25,0.45,0.7)
SC32	(0.05,0.15,0.4)	(0.05,0.2,0.45)	(0.2,0.35,0.6)	(0.15,0.3,0.55)	(0.25,0.45,0.7)	(0.1,0.25,0.5)	(0.35,0.55,0.75)	(0.1,0.3,0.55)
SC41	(0.35,0.6,0.85)	(0.45,0.7,0.85)	(0.15,0.4,0.65)	(0.25,0.5,0.75)	(0.45,0.7,0.9)	(0.15,0.3,0.55)	(0.4,0.65,0.85)	(0.25,0.5,0.75)
SC42	(0.45,0.7,0.9)	(0.5,0.75,0.95)	(0.2,0.4,0.65)	(0.4,0.65,0.9)	(0.35,0.6,0.85)	(0.3,0.5,0.75)	(0.35,0.55,0.75)	(0.3,0.5,0.7)
SC43	(0.25,0.45,0.7)	(0.35,0.55,0.75)	(0.15,0.35,0.6)	(0.25,0.5,0.75)	(0.5,0.75,0.9)	(0.3,0.5,0.7)	(0.3,0.55,0.8)	(0.3,0.55,0.75)
SC44	(0.35,0.6,0.85)	(0.5,0.75,0.9)	(0.25,0.5,0.75)	(0.4,0.65,0.8)	(0.45,0.7,0.85)	(0.25,0.5,0.7)	(0.2,0.45,0.7)	(0.25,0.5,0.7)

ادامه جدول ۳-۱۵ ماتریس ارتباط مستقیم ($\tilde{D}$) (میانگین نظر تمام خبرگان)

	SC31	SC32	SC41	SC42	SC43	SC44
SC11	(0.2,0.35,0.55)	(0.15,0.3,0.55)	(0.55,0.8,0.95)	(0.4,0.65,0.85)	(0.15,0.4,0.65)	(0.45,0.7,0.85)
SC12	(0.2,0.4,0.65)	(0.05,0.15,0.4)	(0.45,0.7,0.9)	(0.25,0.45,0.65)	(0.25,0.5,0.7)	(0.5,0.75,0.85)
SC13	(0.25,0.5,0.7)	(0.1,0.25,0.5)	(0.1,0.3,0.55)	(0.2,0.45,0.7)	(0.05,0.3,0.55)	(0.25,0.5,0.75)
SC14	(0,0.2,0.45)	(0.1,0.25,0.5)	(0.25,0.5,0.75)	(0.35,0.6,0.8)	(0.35,0.6,0.8)	(0.45,0.7,0.8)
SC21	(0.1,0.25,0.5)	(0.15,0.35,0.6)	(0.35,0.6,0.85)	(0.45,0.7,0.95)	(0.4,0.65,0.85)	(0.45,0.7,0.9)
SC22	(0.1,0.3,0.55)	(0.1,0.25,0.5)	(0.25,0.5,0.75)	(0.35,0.55,0.8)	(0.3,0.5,0.7)	(0.2,0.45,0.7)
SC23	(0.2,0.4,0.65)	(0.3,0.5,0.7)	(0.2,0.45,0.7)	(0.15,0.4,0.65)	(0.25,0.5,0.75)	(0.15,0.4,0.6)
SC24	(0.35,0.55,0.75)	(0.25,0.45,0.7)	(0.55,0.8,0.95)	(0.15,0.4,0.65)	(0.25,0.5,0.75)	(0.4,0.65,0.8)
SC31	(0,0,0.25)	(0.6,0.85,0.95)	(0.3,0.5,0.75)	(0.2,0.4,0.65)	(0.2,0.4,0.6)	(0.15,0.35,0.6)
SC32	(0.35,0.6,0.8)	(0,0,0.25)	(0.1,0.3,0.55)	(0.1,0.3,0.55)	(0.45,0.7,0.9)	(0.1,0.2,0.45)
SC41	(0.2,0.4,0.65)	(0.1,0.3,0.55)	(0,0,0.25)	(0.45,0.7,0.9)	(0.35,0.6,0.8)	(0.45,0.7,0.85)
SC42	(0.15,0.3,0.55)	(0.2,0.4,0.65)	(0.25,0.45,0.7)	(0,0,0.25)	(0.35,0.55,0.75)	(0.4,0.6,0.8)
SC43	(0.05,0.2,0.45)	(0.3,0.55,0.75)	(0.35,0.6,0.8)	(0.4,0.65,0.85)	(0,0,0.25)	(0.35,0.6,0.8)
SC44	(0.1,0.3,0.55)	(0,0.15,0.4)	(0.45,0.7,0.85)	(0.4,0.65,0.85)	(0.35,0.6,0.85)	(0,0,0.25)

۳-۹،۴ گام دوم: نرمال سازی ماتریس ارتباط مستقیم

برای نرمالیزه کردن ماتریس به دست آمده از رابطه‌های ذیل استفاده می‌کنیم.

$$\tilde{H}_{ij} = r \times \tilde{z}_{ij} = (r \times l'_{ij}, r \times m'_{ij}, r \times u'_{ij}) = (l''_{ij}, m''_{ij}, u''_{ij}) \quad ۳-۵۲$$

که ۲ از رابطه زیر به دست می‌آید:

$$r = \min \left\{ \frac{1}{\max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}}, \frac{1}{\max_{1 \leq i \leq n} \sum_{i=1}^n a_{ij}} \right\} = 0.09216$$

پس از بدست آوردن مقدار ۲ تمامی عناصر جدول ۳-۱۵ را در مقدار ۲ ضرب می‌نماییم.

نتایج حاصل از این گام مطابق جدول ۳-۱۶ ارائه شده است.

جدول ۳-۱۶ ماتریس نرمالیزه شده ($\tilde{H}_{ij}$)

	SC11	SC12	SC13	SC14	SC21	SC22	SC23	SC24
SC11	(0,0,0.023)	(0.0645,0.0876,0.0922)	(0.0415,0.0645,0.0783)	(0.0276,0.0507,0.0737)	(0.0369,0.0599,0.0829)	(0.0276,0.0461,0.0645)	(0.0184,0.0415,0.0599)	(0.0415,0.0645,0.0829)
SC12	(0.0461,0.0691,0.0876)	(0,0,0.023)	(0.0323,0.0507,0.0737)	(0.0138,0.0369,0.0599)	(0.0323,0.0553,0.0737)	(0.0046,0.0276,0.0507)	(0.0138,0.023,0.0415)	(0.0184,0.0369,0.0599)
SC13	(0.0415,0.0645,0.0829)	(0.0553,0.0783,0.0876)	(0,0,0.023)	(0.0276,0.0507,0.0691)	(0.0369,0.0599,0.0783)	(0.0046,0.023,0.0461)	(0,0.0092,0.0323)	(0.0323,0.0553,0.0691)
SC14	(0.0323,0.0553,0.0737)	(0.0461,0.0691,0.0829)	(0.0276,0.0461,0.0645)	(0,0,0.023)	(0.0461,0.0691,0.0876)	(0.0138,0.0323,0.0553)	(0.0138,0.0276,0.0507)	(0.0323,0.0553,0.0737)
SC21	(0.0323,0.0553,0.0783)	(0.0369,0.0599,0.0737)	(0.023,0.0415,0.0645)	(0.0369,0.0599,0.0829)	(0,0,0.023)	(0.023,0.0461,0.0691)	(0.0276,0.0507,0.0691)	(0.0507,0.0737,0.0876)
SC22	(0.0092,0.0323,0.0553)	(0.0138,0.0323,0.0553)	(0.0138,0.0323,0.0553)	(0.0184,0.0369,0.0599)	(0.0138,0.0369,0.0599)	(0,0,0.023)	(0.0276,0.0507,0.0737)	(0.0276,0.0507,0.0691)
SC23	(0.0138,0.0369,0.0599)	(0.0276,0.0507,0.0691)	(0.0184,0.0369,0.0599)	(0.0092,0.0276,0.0507)	(0.0092,0.0323,0.0553)	(0.0276,0.0507,0.0691)	(0,0,0.023)	(0.0323,0.0553,0.0691)
SC24	(0.0461,0.0691,0.0876)	(0.0553,0.0783,0.0922)	(0.0276,0.0461,0.0691)	(0.0369,0.0599,0.0783)	(0.023,0.0461,0.0691)	(0.0276,0.0461,0.0645)	(0.0276,0.0507,0.0737)	(0,0,0.023)
SC31	(0.023,0.0415,0.0645)	(0.0276,0.0461,0.0645)	(0.0092,0.0276,0.0507)	(0.0138,0.0369,0.0599)	(0.0092,0.023,0.0461)	(0.0138,0.0323,0.0553)	(0.0276,0.0461,0.0691)	(0.023,0.0415,0.0645)
SC32	(0.0046,0.0138,0.0369)	(0.0046,0.0184,0.0415)	(0.0184,0.0323,0.0553)	(0.0138,0.0276,0.0507)	(0.023,0.0415,0.0645)	(0.0092,0.023,0.0461)	(0.0323,0.0507,0.0691)	(0.0092,0.0276,0.0507)
SC41	(0.0323,0.0553,0.0783)	(0.0415,0.0645,0.0783)	(0.0138,0.0369,0.0599)	(0.023,0.0461,0.0691)	(0.0415,0.0645,0.0829)	(0.0138,0.0276,0.0507)	(0.0369,0.0599,0.0783)	(0.023,0.0461,0.0691)
SC42	(0.0415,0.0645,0.0829)	(0.0461,0.0691,0.0876)	(0.0184,0.0369,0.0599)	(0.0369,0.0599,0.0829)	(0.0323,0.0553,0.0783)	(0.0276,0.0461,0.0691)	(0.0323,0.0507,0.0691)	(0.0276,0.0461,0.0645)
SC43	(0.023,0.0415,0.0645)	(0.0323,0.0507,0.0691)	(0.0138,0.0323,0.0553)	(0.023,0.0461,0.0691)	(0.0461,0.0691,0.0829)	(0.0276,0.0461,0.0645)	(0.0276,0.0507,0.0737)	(0.0276,0.0507,0.0691)
SC44	(0.0323,0.0553,0.0783)	(0.0461,0.0691,0.0829)	(0.023,0.0461,0.0691)	(0.0369,0.0599,0.0737)	(0.0415,0.0645,0.0783)	(0.023,0.0461,0.0645)	(0.0184,0.0415,0.0645)	(0.023,0.0461,0.0645)

ادامه جدول ۳-۱۶ ماتریس نرمالیزه شده $(\tilde{H}_{ij})$

	SC31	SC32	SC41	SC42	SC43	SC44
SC11	(0.0184,0.0323,0.0507)	(0.0138,0.0276,0.0507)	(0.0507,0.0737,0.0876)	(0.0369,0.0599,0.0783)	(0.0138,0.0369,0.0599)	(0.0415,0.0645,0.0783)
SC12	(0.0184,0.0369,0.0599)	(0.0046,0.0138,0.0369)	(0.0415,0.0645,0.0829)	(0.023,0.0415,0.0599)	(0.023,0.0461,0.0645)	(0.0461,0.0691,0.0783)
SC13	(0.023,0.0461,0.0645)	(0.0092,0.023,0.0461)	(0.0092,0.0276,0.0507)	(0.0184,0.0415,0.0645)	(0.0046,0.0276,0.0507)	(0.023,0.0461,0.0691)
SC14	(0,0.0184,0.0415)	(0.0092,0.023,0.0461)	(0.023,0.0461,0.0691)	(0.0323,0.0553,0.0737)	(0.0323,0.0553,0.0737)	(0.0415,0.0645,0.0737)
SC21	(0.0092,0.023,0.0461)	(0.0138,0.0323,0.0553)	(0.0323,0.0553,0.0783)	(0.0415,0.0645,0.0876)	(0.0369,0.0599,0.0783)	(0.0415,0.0645,0.0829)
SC22	(0.0092,0.0276,0.0507)	(0.0092,0.023,0.0461)	(0.023,0.0461,0.0691)	(0.0323,0.0507,0.0737)	(0.0276,0.0461,0.0645)	(0.0184,0.0415,0.0645)
SC23	(0.0184,0.0369,0.0599)	(0.0276,0.0461,0.0645)	(0.0184,0.0415,0.0645)	(0.0138,0.0369,0.0599)	(0.023,0.0461,0.0691)	(0.0138,0.0369,0.0553)
SC24	(0.0323,0.0507,0.0691)	(0.023,0.0415,0.0645)	(0.0507,0.0737,0.0876)	(0.0138,0.0369,0.0599)	(0.023,0.0461,0.0691)	(0.0369,0.0599,0.0737)
SC31	(0,0,0.023)	(0.0553,0.0783,0.0876)	(0.0276,0.0461,0.0691)	(0.0184,0.0369,0.0599)	(0.0184,0.0369,0.0553)	(0.0138,0.0323,0.0553)
SC32	(0.0323,0.0553,0.0737)	(0,0,0.023)	(0.0092,0.0276,0.0507)	(0.0092,0.0276,0.0507)	(0.0415,0.0645,0.0829)	(0.0092,0.0184,0.0415)
SC41	(0.0184,0.0369,0.0599)	(0.0092,0.0276,0.0507)	(0,0,0.023)	(0.0415,0.0645,0.0829)	(0.0323,0.0553,0.0737)	(0.0415,0.0645,0.0783)
SC42	(0.0138,0.0276,0.0507)	(0.0184,0.0369,0.0599)	(0.023,0.0415,0.0645)	(0,0,0.023)	(0.0323,0.0507,0.0691)	(0.0369,0.0553,0.0737)
SC43	(0.0046,0.0184,0.0415)	(0.0276,0.0507,0.0691)	(0.0323,0.0553,0.0737)	(0.0369,0.0599,0.0783)	(0,0,0.023)	(0.0323,0.0553,0.0737)
SC44	(0.0092,0.0276,0.0507)	(0,0.0138,0.0369)	(0.0415,0.0645,0.0783)	(0.0369,0.0599,0.0783)	(0.0323,0.0553,0.0783)	(0,0,0.023)

۳-۳، ۹، ۴ گام سوم: محاسبه ماتریس ارتباط کامل زیرعوامل ($\tilde{T}_C$)

بعد از محاسبه ماتریس‌های فوق، ماتریس روابط کل فازی با توجه به رابطه‌های ذیل به دست می‌آید.

$$\tilde{T}_C = \lim_{k \rightarrow +\infty} (\tilde{H}^1 \oplus \tilde{H}^2 \oplus \dots \oplus \tilde{H}^k) \quad ۵۳-۳$$

که هر درایه آن عدد فازی به صورت $\tilde{t}_{ij} = (l_{ij}^t, m_{ij}^t, u_{ij}^t)$ است و به صورت زیر محاسبه می‌شود:

$$[l_{ij}^t] = H_l \times (I - H_l)^{-1} \quad ۵۴-۳$$

$$[m_{ij}^t] = H_m \times (I - H_m)^{-1} \quad ۵۵-۳$$

$$[u_{ij}^t] = H_u \times (I - H_u)^{-1} \quad ۵۶-۳$$

در این رابطه‌ها l ماتریس یک‌ه و H_l ، H_m و H_u هر کدام ماتریس $n \times n$ هستند که درایه‌های آن را به ترتیب عدد پایین، عدد میانی و عدد بالایی اعداد فازی مثلثی ماتریس H تشکیل می‌دهد.

جدول ۳-۱۷ ماتریس ارتباط کامل ($\tilde{T}_c$)

	SC11	SC12	SC13	SC14	SC21	SC22	SC23	SC24
SC11	(0.0209,0.0914,0.6364)	(0.0877,0.1875,0.7282)	(0.0549,0.1339,0.6142)	(0.0434,0.1303,0.6502)	(0.056,0.1492,0.6962)	(0.0386,0.1106,0.5705)	(0.0326,0.1127,0.5964)	(0.0583,0.1479,0.6648)
SC12	(0.0599,0.1398,0.6312)	(0.0205,0.088,0.5946)	(0.0426,0.1086,0.5521)	(0.0266,0.1036,0.5755)	(0.0472,0.1292,0.6217)	(0.0142,0.0822,0.5031)	(0.0246,0.0832,0.5216)	(0.0325,0.1079,0.5812)
SC13	(0.0547,0.1337,0.6146)	(0.0714,0.1579,0.6423)	(0.0109,0.0586,0.493)	(0.0383,0.1137,0.5721)	(0.0495,0.13,0.6129)	(0.0127,0.0756,0.4884)	(0.0098,0.0674,0.5014)	(0.0441,0.1218,0.5774)
SC14	(0.0486,0.1332,0.6412)	(0.0662,0.1589,0.6746)	(0.0395,0.1089,0.5636)	(0.0142,0.0732,0.5614)	(0.0619,0.1476,0.6572)	(0.024,0.0909,0.5265)	(0.0256,0.0918,0.5501)	(0.047,0.13,0.6154)
SC21	(0.0504,0.1403,0.6967)	(0.0603,0.1591,0.7203)	(0.0365,0.1106,0.6089)	(0.0515,0.1364,0.6667)	(0.0194,0.0901,0.6483)	(0.0346,0.1097,0.5824)	(0.041,0.12,0.6132)	(0.0662,0.1539,0.6772)
SC22	(0.0204,0.096,0.5772)	(0.028,0.107,0.6006)	(0.0215,0.0833,0.5139)	(0.0274,0.0944,0.5527)	(0.025,0.1018,0.5846)	(0.0074,0.0489,0.4572)	(0.0358,0.1021,0.5318)	(0.0373,0.1112,0.5668)
SC23	(0.0245,0.102,0.5833)	(0.0406,0.1255,0.615)	(0.0262,0.0895,0.5202)	(0.0182,0.0876,0.546)	(0.0202,0.0995,0.5825)	(0.0339,0.0984,0.5028)	(0.0088,0.0557,0.4855)	(0.0414,0.1172,0.5688)
SC24	(0.0634,0.1535,0.7037)	(0.0776,0.1766,0.7354)	(0.0413,0.116,0.6123)	(0.0508,0.1365,0.6607)	(0.0421,0.1349,0.6908)	(0.0383,0.1096,0.5767)	(0.0411,0.1206,0.6161)	(0.0176,0.0857,0.6154)
SC31	(0.0338,0.1032,0.5819)	(0.0414,0.1182,0.6051)	(0.0182,0.0793,0.5073)	(0.0231,0.0935,0.5493)	(0.0214,0.089,0.5689)	(0.0212,0.0794,0.4856)	(0.0371,0.0984,0.5253)	(0.0332,0.102,0.5594)
SC32	(0.0135,0.0667,0.5037)	(0.0163,0.0798,0.5287)	(0.0244,0.0735,0.4644)	(0.021,0.0751,0.4913)	(0.0317,0.0939,0.5323)	(0.0155,0.0631,0.4338)	(0.039,0.0934,0.4792)	(0.0184,0.0794,0.4968)
SC41	(0.0484,0.1355,0.671)	(0.062,0.1577,0.6974)	(0.0263,0.1026,0.582)	(0.0367,0.1194,0.6296)	(0.0573,0.1457,0.6786)	(0.0247,0.0894,0.5442)	(0.0486,0.1245,0.5988)	(0.0387,0.1246,0.636)
SC42	(0.0575,0.1419,0.6728)	(0.067,0.1597,0.7033)	(0.0315,0.1017,0.5805)	(0.0501,0.1302,0.6401)	(0.0494,0.1359,0.6726)	(0.0381,0.1048,0.5592)	(0.0444,0.1145,0.5883)	(0.0436,0.1231,0.6299)
SC43	(0.0384,0.1187,0.639)	(0.0515,0.1401,0.6688)	(0.0256,0.0952,0.5612)	(0.0361,0.1162,0.6118)	(0.0608,0.1464,0.6599)	(0.0373,0.1036,0.5414)	(0.0396,0.1142,0.5788)	(0.0422,0.1254,0.618)
SC44	(0.0489,0.1365,0.6639)	(0.0667,0.1628,0.6941)	(0.0352,0.1114,0.5839)	(0.0499,0.1325,0.6269)	(0.058,0.1468,0.6673)	(0.0331,0.1061,0.5504)	(0.0307,0.1076,0.5791)	(0.0388,0.1253,0.6248)

۳-۴،۹،۴ گام چهارم: محاسبه ماتریس ارتباط کامل عوامل اصلی $\tilde{T}_D$

جدول ۳-۱۸ ماتریس ارتباط کامل عوامل اصلی ($\tilde{T}_D$)

ذاتی	دسترس پذیری	متنی	نمایشگری	
(0.0438,0.1266,0.6161)	(0.0199,0.0767,0.4817)	(0.0362,0.1111,0.5803)	(0.0438,0.1201,0.6091)	نمایشگری
(0.0416,0.1236,0.6237)	(0.0254,0.0834,0.4945)	(0.0319,0.1037,0.5813)	(0.0399,0.1196,0.6196)	متنی
(0.0281,0.0928,0.5354)	(0.0278,0.0713,0.4301)	(0.0272,0.0873,0.5102)	(0.024,0.0862,0.529)	دسترس پذیری
(0.0419,0.1209,0.6307)	(0.0209,0.0809,0.5019)	(0.0428,0.1211,0.608)	(0.0457,0.1289,0.6391)	ذاتی

۳-۴،۹،۵ گام پنجم: محاسبه شدت و جهت تأثیر

مطابق با رابطه ۳-۵۷ و ۴-۵۸ میزان شاخص d_i و r_j را محاسبه می‌نماییم. شاخص d_i بیانگر مجموع سطر i ام و شاخص r_j بیانگر مجموع ستون j ام از ماتریس T_C با توجه به بعد مربوطه می‌باشد (جدول ۳-۱۸). به همین صورت میزان شاخص $\tilde{D}_i$ و $\tilde{R}_j$ را محاسبه می‌نماییم. شاخص D_i بیانگر مجموع سطر i ام و شاخص R_j بیانگر مجموع ستون j ام از ماتریس T_D می‌باشد (جدول ۳-۱۸). جهت ترسیم و تحلیل نمودار نیاز به ۲ شاخص شدت اثرگذاری و اثرپذیری و جهت تأثیر می‌باشیم که با استفاده از d_i و r_j به دست می‌آیند. برای هر $i=j$ خواهیم داشت:

$$\tilde{D} = (\tilde{D}_i)_{n \times 1} = [\sum_{j=1}^n \tilde{T}_{ij}]_{n \times 1} \quad ۳-۵۷$$

$$\tilde{R} = (\tilde{R}_j)_{1 \times n} = [\sum_{i=1}^n \tilde{T}_{ij}]_{1 \times n} \quad ۳-۵۸$$

که $\tilde{D}$ و $\tilde{R}$ به ترتیب ماتریس $n \times 1$ و $1 \times n$ هستند.

مرحله بعدی میزان اهمیت عوامل ($\tilde{D}_i + \tilde{R}_i$) و رابطه بین عوامل ($\tilde{D}_i - \tilde{R}_i$) مشخص می‌شود. اگر $\tilde{D}_i - \tilde{R}_i > 0$ باشد عامل مربوطه اثرگذار و اگر $\tilde{D}_i - \tilde{R}_i < 0$ باشد عامل مربوطه اثرپذیر است.

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۰۰

$di+rj$ = شدت اثرگذاری و اثرپذیری (به عبارت دیگر هرچه مقدار $di+rj$ زیرعاملی بیشتر باشد، آن زیرعامل تعامل بیشتری با سایر زیرعوامل سیستم دارد).

$di-rj$ = جهت تأثیرگذاری یا تأثیرپذیری (بدین صورت که اگر $di-rj > 0$ باشد زیرعامل مربوطه اثرگذار و اگر $di-rj < 0$ باشد زیرعامل مربوطه اثرپذیر است).

جدول ۳-۱۹ محاسبه میزان شاخص r و d برای ماتریس $\tilde{T}_C$

	SC11	SC12	SC13	SC14	D
SC 11	(0.0209,0.0914,0.6364)	(0.0877,0.1875,0.7282)	(0.0549,0.1339,0.6142)	(0.0434,0.1303,0.6502)	(0.2069,0.5431,2.629)
SC 12	(0.0599,0.1398,0.6312)	(0.0205,0.088,0.5946)	(0.0426,0.1086,0.5521)	(0.0266,0.1036,0.5755)	(0.1496,0.44,2.3534)
SC 13	(0.0547,0.1337,0.6146)	(0.0714,0.1579,0.6423)	(0.0109,0.0586,0.493)	(0.0383,0.1137,0.5721)	(0.1753,0.4639,2.322)
SC 14	(0.0486,0.1332,0.6412)	(0.0662,0.1589,0.6746)	(0.0395,0.1089,0.5636)	(0.0142,0.0732,0.5614)	(0.1685,0.4742,2.4408)
r	(0.1841,0.4981,2.5234)	(0.2458,0.5923,2.6397)	(0.1479,0.41,2.2229)	(0.1225,0.4208,2.3592)	

ادامه جدول ۳-۱۹ محاسبه میزان شاخص r و d برای ماتریس $\tilde{T}_C$

	SC21	SC22	SC23	SC24	D
SC 21	(0.0194,0.0901,0.6483)	(0.0346,0.1097,0.5824)	(0.041,0.12,0.6132)	(0.0662,0.1539,0.6772)	(0.1612,0.4737,2.5211)
SC 22	(0.025,0.1018,0.5846)	(0.0074,0.0489,0.4572)	(0.0358,0.1021,0.5318)	(0.0373,0.1112,0.5668)	(0.1055,0.364,2.1404)
SC 23	(0.0202,0.0995,0.5825)	(0.0339,0.0984,0.5028)	(0.0088,0.0557,0.4855)	(0.0414,0.1172,0.5688)	(0.1043,0.3708,2.1396)
SC 24	(0.0421,0.1349,0.6908)	(0.0383,0.1096,0.5767)	(0.0411,0.1206,0.6161)	(0.0176,0.0857,0.6154)	(0.1391,0.4508,2.499)
r	(0.1067,0.4263,2.5062)	(0.1142,0.3666,2.1191)	(0.1267,0.3984,2.2466)	(0.1625,0.468,2.4282)	

ادامه جدول ۳-۱۹ محاسبه میزان شاخص r و d برای ماتریس $\tilde{T}_C$

	SC31	SC32	D
SC31	(0.0074,0.0437,0.425)	(0.0607,0.1166,0.4819)	(0.0681,0.1603,0.9069)
SC32	(0.0363,0.0878,0.4319)	(0.0066,0.0371,0.3816)	(0.0429,0.1249,0.8135)
r	(0.0437,0.1315,0.8569)	(0.0673,0.1537,0.8635)	

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۰۲

ادامه جدول ۳-۱۹ محاسبه میزان شاخص r و d برای ماتریس $\tilde{T}_C$

	SC41	SC42	SC43	SC44	D
SC41	(0.0175,0.0837,0.6135)	(0.056,0.1405,0.6581)	(0.0459,0.1296,0.6352)	(0.0582,0.1464,0.6515)	(0.1776,0.5002,2.5583)
SC42	(0.0407,0.1222,0.6514)	(0.0164,0.0782,0.5995)	(0.0464,0.124,0.6293)	(0.0545,0.1365,0.6454)	(0.158,0.4609,2.5256)
SC43	(0.0474,0.1322,0.6428)	(0.0508,0.1331,0.6358)	(0.0146,0.0748,0.57)	(0.0484,0.1341,0.6289)	(0.1612,0.4742,2.4775)
SC44	(0.0574,0.145,0.6585)	(0.0521,0.1373,0.647)	(0.0457,0.1296,0.6319)	(0.0189,0.087,0.5924)	(0.1741,0.4989,2.5298)
r	(0.163,0.4831,2.5662)	(0.1753,0.4891,2.5404)	(0.1526,0.458,2.4664)	(0.18,0.504,2.5182)	

ادامه جدول ۳-۱۹ محاسبه میزان شاخص r و d برای ماتریس $\tilde{T}_C$

	SC21	SC22	SC23	SC24	D
SC21	(0.0194,0.0901,0.6483)	(0.0346,0.1097,0.5824)	(0.041,0.12,0.6132)	(0.0662,0.1539,0.6772)	(0.1612,0.4737,2.5211)
SC22	(0.025,0.1018,0.5846)	(0.0074,0.0489,0.4572)	(0.0358,0.1021,0.5318)	(0.0373,0.1112,0.5668)	(0.1055,0.364,2.1404)
SC23	(0.0202,0.0995,0.5825)	(0.0339,0.0984,0.5028)	(0.0088,0.0557,0.4855)	(0.0414,0.1172,0.5688)	(0.1043,0.3708,2.1396)
SC24	(0.0421,0.1349,0.6908)	(0.0383,0.1096,0.5767)	(0.0411,0.1206,0.6161)	(0.0176,0.0857,0.6154)	(0.1391,0.4508,2.499)
r	(0.1067,0.4263,2.5062)	(0.1142,0.3666,2.1191)	(0.1267,0.3984,2.2466)	(0.1625,0.468,2.4282)	

با توجه به مقادیر محاسبه شده در فوق، مقدار شاخص $di + rj$ و $di - rj$ را برای زیرعوامل و همچنین شاخص $\tilde{D}_i + \tilde{R}_i$ و $\tilde{D}_i - \tilde{R}_i$ را برای عوامل اصلی بدست می آوریم و سپس با استفاده از رابطه ۴-۹ فازی زدایی می کنیم:

$$B = \frac{(a_1 + a_3 + 2 \times a_2)}{4}$$

۵۹-۳

جدول ۲۰-۳ محاسبه شدت و جهت تاثیر عوامل اصلی

وضعیت عامل	(Di- Ri)def	(Di+Ri)def	Di-Ri	Di+Ri	Ri	Di	
تاثیر پذیر	-0.04	1.6899	(-0.0097,-0.0203,- 0.1096)	(0.2971,0.8893,4.684)	(0.1534,0.4548,2.3968)	(0.1437,0.4345,2.2872)	نمایشگری
تاثیر گذار	0.0136	1.6457	(0.0007,0.0071,0.0393)	(0.2769,0.8535,4.5989)	(0.1381,0.4232,2.2798)	(0.1388,0.4303,2.3191)	متنی
تاثیر گذار	0.0401	1.3535	(0.0131,0.0253,0.0965)	(0.2011,0.6499,3.9129)	(0.094,0.3123,1.9082)	(0.1071,0.3376,2.0047)	دسترس پذیری
تاثیر پذیر	- 0.0136	1.7309	(-0.0041,-0.0121,- 0.0262)	(0.3067,0.9157,4.7856)	(0.1554,0.4639,2.4059)	(0.1513,0.4518,2.3797)	ذاتی

جدول ۳-۲۱ محاسبه شدت و جهت تاثیر زبرعوامل

	Di	Ri	Di+Ri	Di-Ri	(Di+Ri)def	(Di-Ri)def	وضعیت زیر عامل
SC11	(0.2069,0.5431,2.629)	(0.1841,0.4981,2.5234)	(0.391,1.0412,5.1524)	(0.0228,0.045,0.1056)	1.9065	0.0546	تاثیرگذار
SC12	(0.1496,0.44,2.3534)	(0.2458,0.5923,2.6397)	(0.3954,1.0323,4.9931)	(-0.0962,-0.1523,-0.2863)	1.8633	-0.1718	تاثیرپذیر
SC13	(0.1753,0.4639,2.322)	(0.1479,0.41,2.2229)	(0.3232,0.8739,4.5449)	(0.0274,0.0539,0.0991)	1.654	0.0586	تاثیرگذار
SC14	(0.1685,0.4742,2.4408)	(0.1225,0.4208,2.3592)	(0.291,0.895,4.8)	(0.046,0.0534,0.0816)	1.7203	0.0586	تاثیرگذار
SC21	(0.1612,0.4737,2.5211)	(0.1067,0.4263,2.5062)	(0.2679,0.9,5.0273)	(0.0545,0.0474,0.0149)	1.7738	0.0411	تاثیرگذار
SC22	(0.1055,0.364,2.1404)	(0.1142,0.3666,2.1191)	(0.2197,0.7306,4.2595)	(-0.0087,-0.0026,0.0213)	1.4851	0.0019	تاثیرگذار
SC23	(0.1043,0.3708,2.1396)	(0.1267,0.3984,2.2466)	(0.231,0.7692,4.3862)	(-0.0224,-0.0276,-0.107)	1.5389	-0.0462	تاثیرپذیر
SC24	(0.1391,0.4508,2.499)	(0.1625,0.468,2.4282)	(0.3016,0.9188,4.9272)	(-0.0234,-0.0172,0.0708)	1.7666	0.0033	تاثیرگذار
SC31	(0.0681,0.1603,0.9069)	(0.0437,0.1315,0.8569)	(0.1118,0.2918,1.7638)	(0.0244,0.0288,0.05)	0.6148	0.033	تاثیرگذار
SC32	(0.0429,0.1249,0.8135)	(0.0673,0.1537,0.8635)	(0.1102,0.2786,1.677)	(-0.0244,-0.0288,-0.05)	0.5861	-0.033	تاثیرپذیر
SC41	(0.1776,0.5002,2.5583)	(0.163,0.4831,2.5662)	(0.3406,0.9833,5.1245)	(0.0146,0.0171,-0.0079)	1.8579	0.0102	تاثیرگذار
SC42	(0.158,0.4609,2.5256)	(0.1753,0.4891,2.5404)	(0.3333,0.95,5.066)	(-0.0173,-0.0282,-0.0148)	1.8248	-0.0221	تاثیرپذیر

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۰۵

SC43	(0.1612,0.4742,2.4775)	(0.1526,0.458,2.4664)	(0.3138,0.9322,4.9439)	(0.0086,0.0162,0.0111)	1.7805	0.013	تاثیرگذار
SC44	(0.1741,0.4989,2.5298)	(0.18,0.504,2.5182)	(0.3541,1.0029,5.048)	(-0.0059,- 0.0051,0.0116)	1.852	-0.0011	تاثیرپذیر

۳-۶،۹،۴ گام ششم: ترسیم نقشه روابط شبکه (NRM)

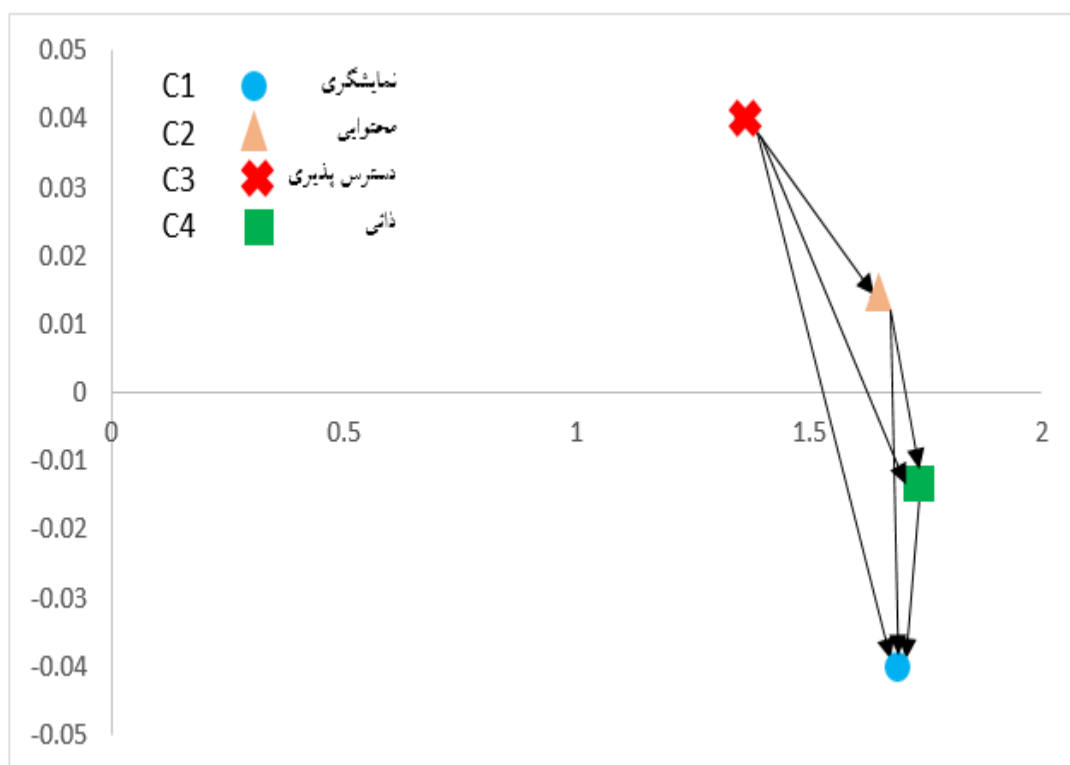
جهت تعیین نقشه روابط شبکه (NRM)، پس از بدست آوردن ماتریس‌های TC و TD می‌بایست D-R و d-r ماتریس‌های عوامل اصلی و زیرعوامل را محاسبه نماییم. عاملی که D-R بیشتری دارند روی عاملی که D-R کمتری دارند تاثیرگذار می‌باشند و بر عکس عاملی که D-R کمتری دارند از عاملی که D-R بیشتری دارند تاثیر می‌پذیرند.

جدول ۳-۲۲ ماتریس ارتباط کامل عوامل اصلی (TD) دیفازی شده

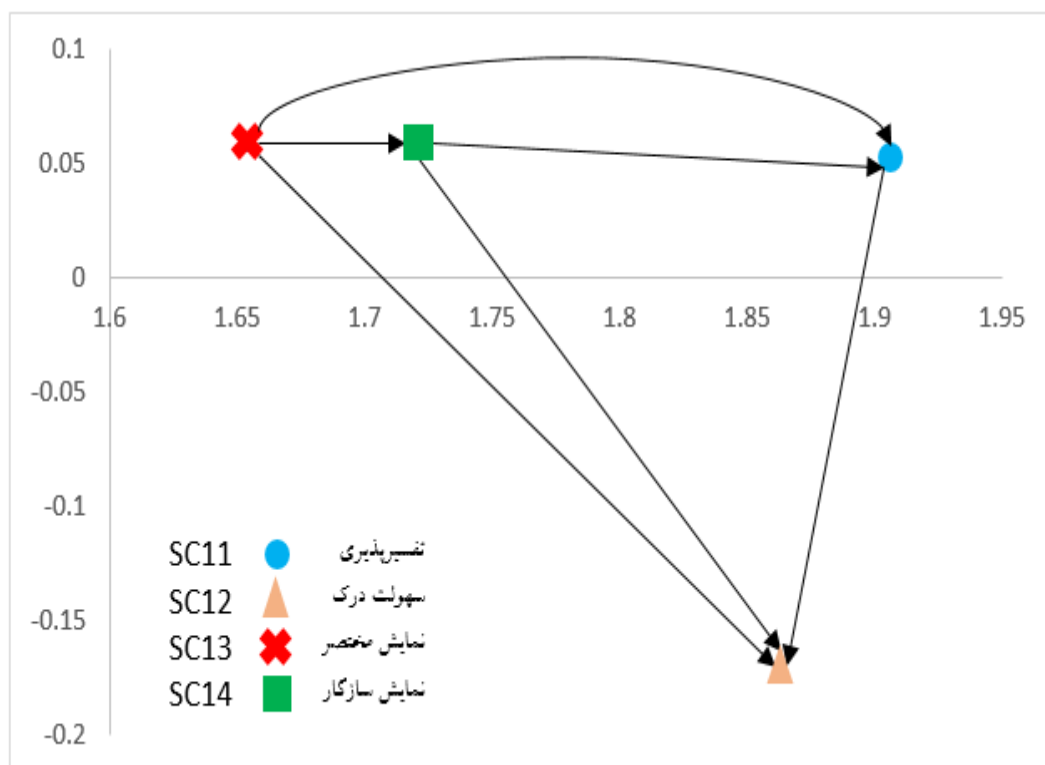
ذاتی	دسترس پذیری	متنی	نمایشگری	
نمایشگری	0.2283	0.1638	0.2097	0.2233
متنی	0.2281	0.1717	0.2052	0.2247
دسترس پذیری	0.1873	0.1501	0.178	0.1814
ذاتی	0.2286	0.1712	0.2233	0.2357

جدول ۳-۲۳ ماتریس ارتباط کامل زیرعوامل (TC) دیفازی شده

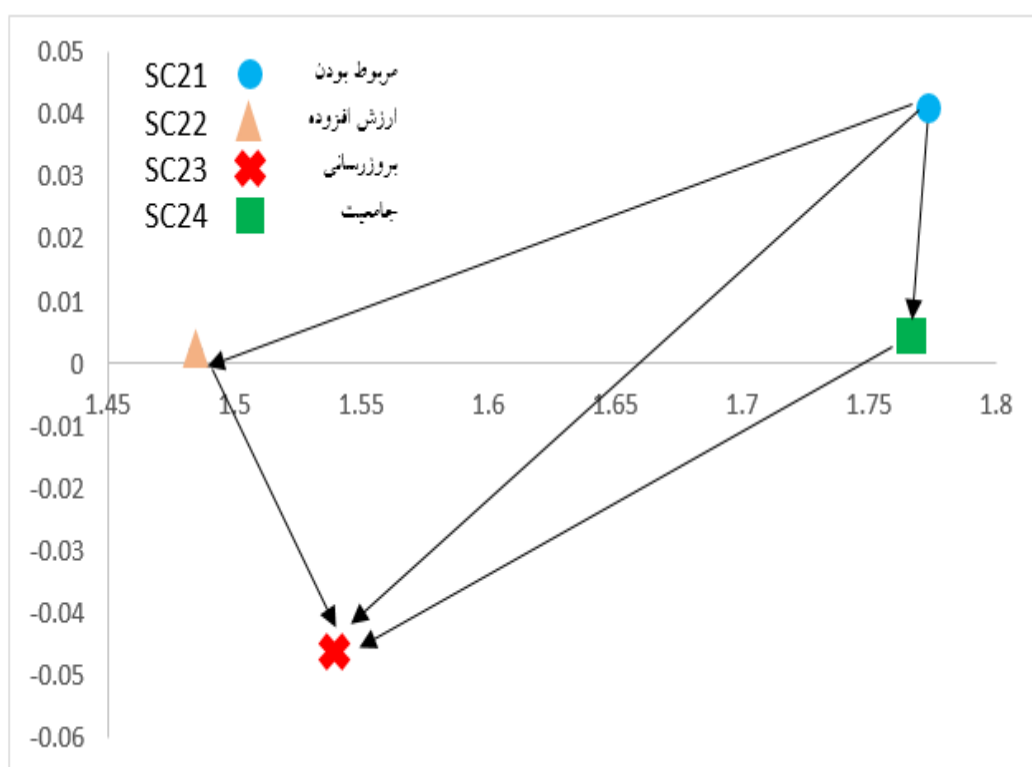
	SC11	SC12	SC13	SC14	SC21	SC22	SC23	SC24	SC31	SC32	SC41	SC42	SC43	SC44
SC11	0.21	0.2977	0.2342	0.2386	0.2627	0.2076	0.2136	0.2547	0.1824	0.1759	0.2701	0.2526	0.2268	0.2597
SC12	0.2427	0.1978	0.2029	0.2023	0.2318	0.1704	0.1782	0.2074	0.1675	0.1464	0.2375	0.2128	0.2099	0.2375
SC13	0.2342	0.2574	0.1553	0.2094	0.2306	0.1631	0.1615	0.2163	0.1709	0.1505	0.2012	0.2071	0.1893	0.2148
SC14	0.239	0.2647	0.2052	0.1805	0.2536	0.1831	0.1898	0.2306	0.1566	0.1596	0.2305	0.2327	0.2272	0.2423
SC21	0.2569	0.2747	0.2166	0.2477	0.212	0.2091	0.2235	0.2628	0.1751	0.1804	0.256	0.2586	0.2479	0.2609
SC22	0.1974	0.2106	0.1755	0.1922	0.2033	0.1406	0.1929	0.2066	0.1514	0.1471	0.2094	0.2104	0.201	0.2027
SC23	0.203	0.2267	0.1814	0.1848	0.2004	0.1834	0.1514	0.2111	0.1621	0.168	0.2069	0.1977	0.2026	0.1988
SC24	0.2685	0.2916	0.2214	0.2461	0.2507	0.2086	0.2246	0.2011	0.1995	0.1896	0.2712	0.2335	0.2365	0.2558
SC31	0.2055	0.2207	0.1711	0.1899	0.1921	0.1664	0.1898	0.1992	0.13	0.1939	0.2098	0.1965	0.1929	0.194
SC32	0.1626	0.1761	0.159	0.1656	0.188	0.1439	0.1762	0.1685	0.1609	0.1156	0.1729	0.1696	0.1986	0.1641
SC41	0.2476	0.2687	0.2034	0.2263	0.2568	0.187	0.2241	0.231	0.18	0.1701	0.1996	0.2488	0.2351	0.2506
SC42	0.2535	0.2724	0.2038	0.2376	0.2485	0.2017	0.2154	0.2299	0.1718	0.1773	0.2341	0.1931	0.2309	0.2432
SC43	0.2287	0.2501	0.1943	0.2201	0.2534	0.1965	0.2117	0.2277	0.1596	0.1843	0.2386	0.2382	0.1836	0.2364
SC44	0.2464	0.2716	0.2105	0.2355	0.2547	0.1989	0.2063	0.2285	0.1697	0.1562	0.2515	0.2434	0.2342	0.1963



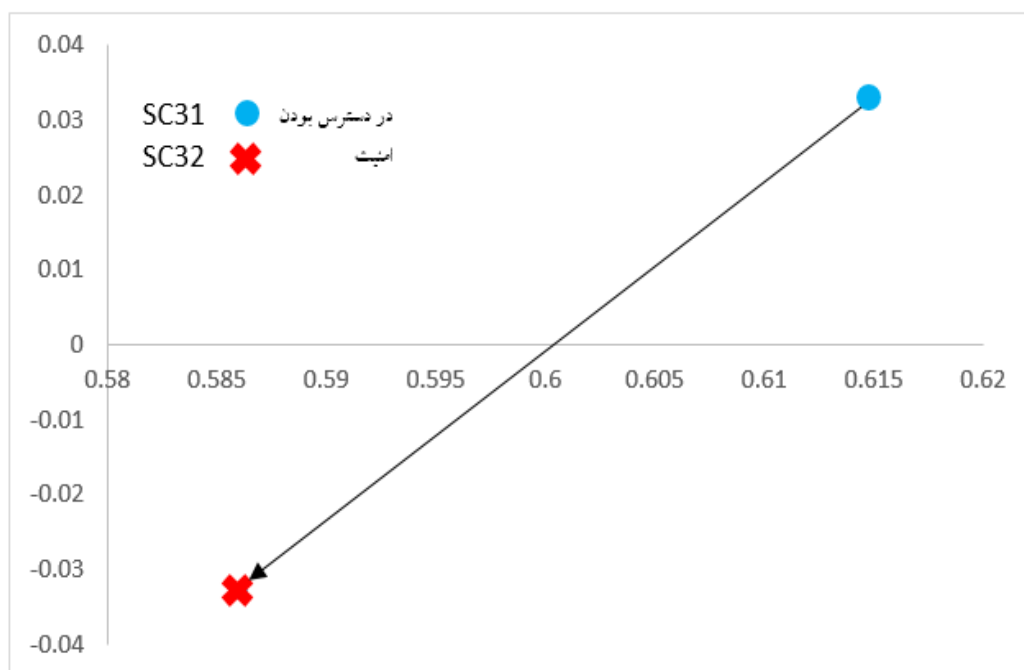
نمودار ۳-۴ روابط درون شبکه مربوط به عوامل اصلی



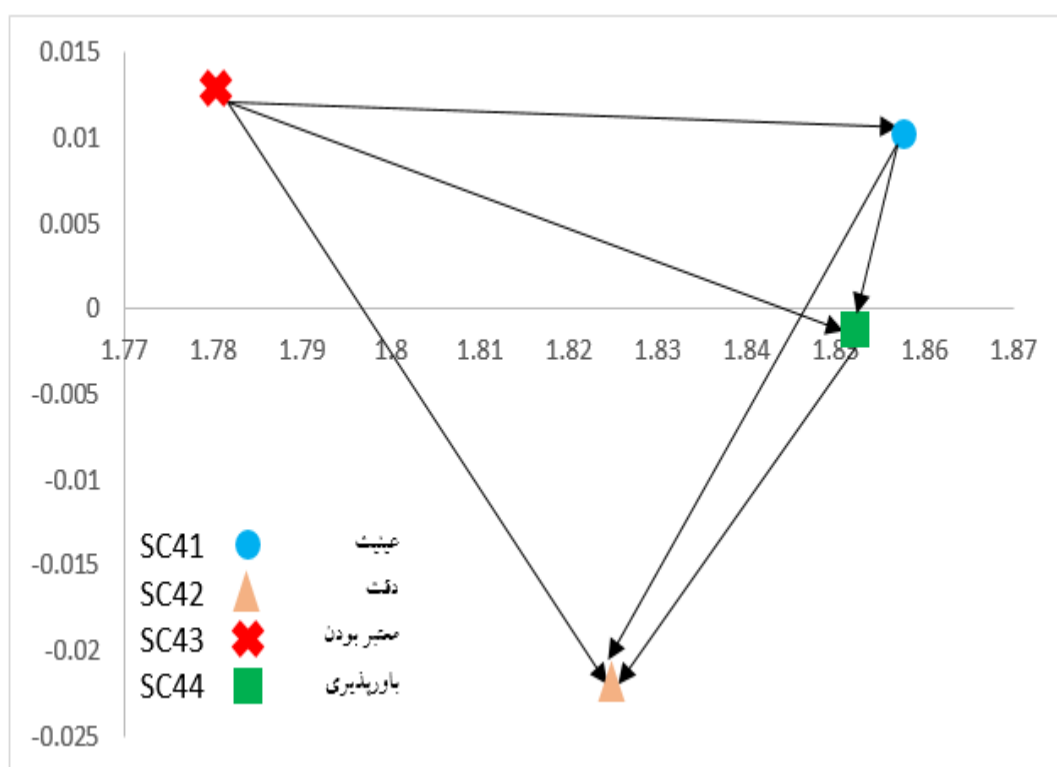
نمودار ۳-۵ روابط درون شبکه مربوط به عامل نمایشگری



نمودار ۳-۶ روابط درون شبکه مربوط به عامل متنی

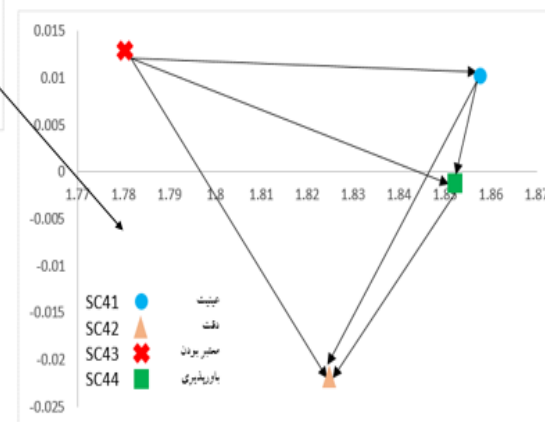
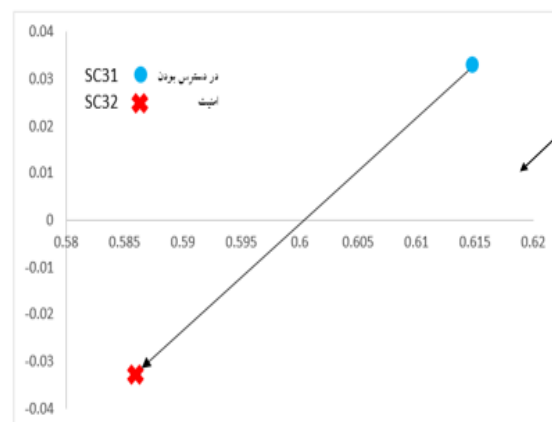
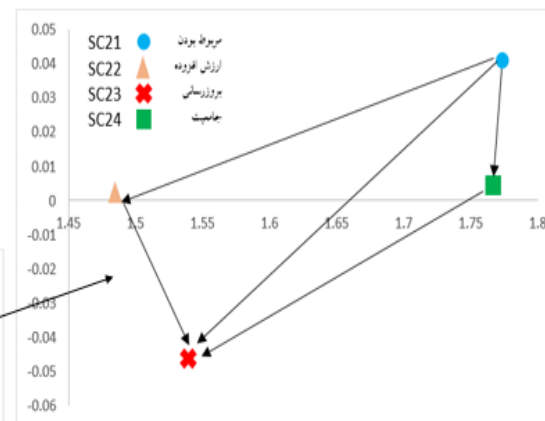
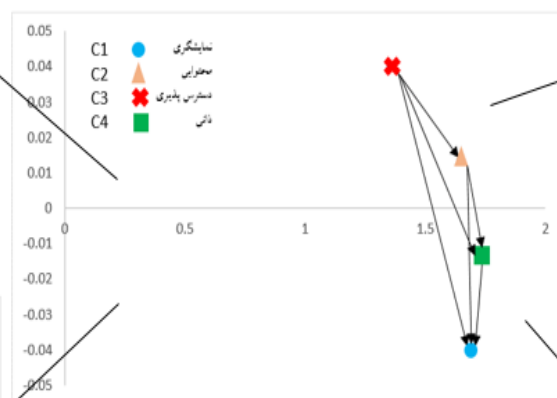
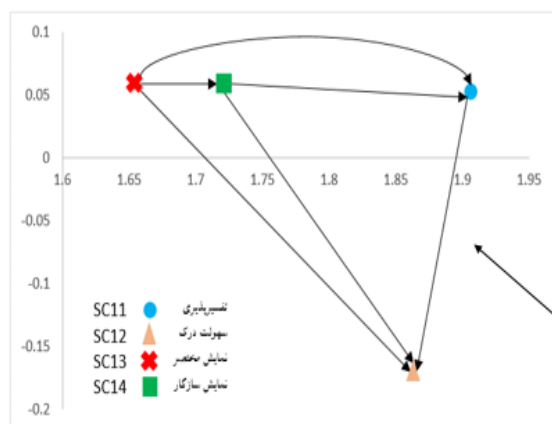


نمودار ۳-۷ روابط درون شبکه مربوط به عامل دسترس پذیری



نمودار ۳-۸ روابط درون شبکه مربوط به عامل ذاتی

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۱۱



نمودار ۳-۹ روابط درون شبکه کلی

۳-۹،۴ گام هفتم: نرمال سازی ماتریس ارتباط کامل عوامل اصلی (T_D^α)

با توجه به رابطه ۳-۶۰ اقدام به نرمال سازی ماتریس T_D می‌نماییم. به این طریق که مجموع هر سطر از ماتریس T_D (جدول ۳-۲۴) را با توجه به بعد مربوطه محاسبه، سپس عنصر هر سطر را بر مجموع عناصر همان سطر تقسیم نموده و در پایان جای سطر و ستون را عوض می‌نماییم.

$$T_D = \begin{bmatrix} t_{11}^{D_{11}} & \dots & t_{1j}^{D_{1j}} & \dots & t_{1m}^{D_{1m}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{i1}^{D_{i1}} & \dots & t_{ij}^{D_{ij}} & \dots & t_{im}^{D_{im}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{m1}^{D_{m1}} & \dots & t_{mj}^{D_{mj}} & \dots & t_{mm}^{D_{mm}} \end{bmatrix} \longrightarrow \begin{aligned} d_1 &= \sum_{j=1}^m t_{1j}^{D_{1j}} \\ d_i &= \sum_{j=1}^m t_{ij}^{D_{ij}}, d_i = \sum_{j=1}^m t_{ij}^{D_{ij}}, i = 1, \dots, m \\ d_m &= \sum_{j=1}^m t_{mj}^{D_{mj}} \end{aligned}$$

$$T_D^\alpha = \begin{bmatrix} t_{11}^{D_{11}} / d_1 & \dots & t_{1j}^{D_{1j}} / d_1 & \dots & t_{1m}^{D_{1m}} / d_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{i1}^{D_{i1}} / d_i & \dots & t_{ij}^{D_{ij}} / d_i & \dots & t_{im}^{D_{im}} / d_i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{m1}^{D_{m1}} / d_m & \dots & t_{mj}^{D_{mj}} / d_m & \dots & t_{mm}^{D_{mm}} / d_m \end{bmatrix} = \begin{bmatrix} t_D^{\alpha 11} & \dots & t_D^{\alpha 1j} & \dots & t_D^{\alpha 1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_D^{\alpha i1} & \dots & t_D^{\alpha ij} & \dots & t_D^{\alpha in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_D^{\alpha n1} & \dots & t_D^{\alpha nj} & \dots & t_D^{\alpha nn} \end{bmatrix} \quad ۳-۶۰$$

جدول ۳-۲۴ ماتریس نرمال ارتباط کامل عوامل اصلی (T_D^α)

	نمایشگری	متنی	دسترس پذیری	
نمایشگری	(0.3048,0.2764,0.2663)	(0.2875,0.2779,0.2672)	(0.2241,0.2553,0.2639)	(0.302,0.2853,0.26
متنی	(0.2519,0.2557,0.2537)	(0.2298,0.241,0.2507)	(0.254,0.2586,0.2545)	(0.2829,0.268,0.25
دسترس پذیری	(0.1385,0.1765,0.2106)	(0.183,0.1938,0.2132)	(0.2596,0.2112,0.2145)	(0.1381,0.1791,0.21
ذاتی	(0.3048,0.2914,0.2694)	(0.2997,0.2872,0.2689)	(0.2624,0.2749,0.2671)	(0.2769,0.2676,0.2

۳-۸،۹،۴ گام هشتم: نرمال سازی ماتریس ارتباط کامل زیرعوامل T_C^α و تشکیل سوپر ماتریس ناموزون

ماتریس T_C (جدول ۳-۲۳) را با استفاده از روابط زیر نرمال می‌کنیم؛ به این طریق که در این گام مجموع هر سطر T_C^{ij} را با توجه به بعد مربوطه محاسبه و سپس در هر T_C^{ij} ، هر عنصر بر مجموع عناصر سطر مربوط به خود تقسیم می‌گردد. برای مثال اگر هر T_C^α (جدول ۳-۳۴) را شامل مجموعه‌ای از $T_C^{\alpha ij}$ بدانیم، $T_C^{\alpha 11}$ از نرمال سازی T_C^{11} به دست می‌آید. با ترانسپوز ماتریس T_C^α ، سوپر ماتریس ناموزون به شرح جدول ۳-۲۵ می‌باشد.

$$T_C^\alpha = \begin{matrix} & \begin{matrix} D_1 & & D_j & & D_n \\ c_{11} \dots c_{1m_1} & \dots & c_{j1} \dots c_{jm_j} & \dots & c_{n1} \dots c_{nm_n} \end{matrix} \\ \begin{matrix} D_1 \\ c_{11} \\ c_{12} \\ \vdots \\ c_{1m_1} \\ D_j \\ c_{j1} \\ c_{j2} \\ \vdots \\ c_{jm_j} \\ D_n \\ c_{n1} \\ c_{n2} \\ \vdots \\ c_{nm_n} \end{matrix} & \begin{bmatrix} T_c^{\alpha 11} & \dots & T_c^{\alpha 1j} & \dots & T_c^{\alpha 1n} \\ \vdots & & \vdots & & \vdots \\ T_c^{\alpha i1} & \dots & T_c^{\alpha ij} & \dots & T_c^{\alpha in} \\ \vdots & & \vdots & & \vdots \\ T_c^{\alpha n1} & \dots & T_c^{\alpha nj} & \dots & T_c^{\alpha nn} \end{bmatrix} \end{matrix} \quad ۶۱-۳$$

$$d_{ci}^{11} = \sum_{j=1}^{m_1} t_{cij}^{11}, i = 1, 2, \dots, m_1 \quad ۶۲-۳$$

۶۳-۳

$$T_C^{\alpha 11} = \begin{bmatrix} t_{c11}^{11}/d_{c1}^{11} & \dots & t_{c1j}^{11}/d_{c1}^{11} & \dots & t_{c1m_1}^{11}/d_{c1}^{11} \\ \vdots & & \vdots & & \vdots \\ t_{ci1}^{11}/d_{ci}^{11} & \dots & t_{cij}^{11}/d_{ci}^{11} & \dots & t_{cim_1}^{11}/d_{ci}^{11} \\ \vdots & & \vdots & & \vdots \\ t_{cm11}^{11}/d_{cm1}^{11} & \dots & t_{cm1j}^{11}/d_{cm1}^{11} & \dots & t_{cm1m_1}^{11}/d_{cm1}^{11} \end{bmatrix}$$

$$= \begin{bmatrix} t_{c11}^{\alpha 11} & \dots & t_{c1j}^{\alpha 11} & \dots & t_{c1m_1}^{\alpha 11} \\ \vdots & & \vdots & & \vdots \\ t_{ci1}^{\alpha 11} & \dots & t_{cij}^{\alpha 11} & \dots & t_{cim_1}^{\alpha 11} \\ \vdots & & \vdots & & \vdots \\ t_{cm11}^{\alpha 11} & \dots & t_{cm1j}^{\alpha 11} & \dots & t_{cm1m_1}^{\alpha 11} \end{bmatrix}$$

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۱۴

جدول ۳-۲۵ سوپر ماتریس ناموزون - ماتریس نرمال ارتباط کامل زیرعوامل ($\mathcal{T}^c$)

	SC11	SC12	SC13	SC14	SC21	SC22	SC23	SC24
SC 11	(0.101,0.1683,0.2421)	(0.4004,0.3177,0.2682)	(0.312,0.2882,0.2647)	(0.2884,0.2809,0.2627)	(0.2536,0.2568,0.2587)	(0.2097,0.2522,0.2572)	(0.2237,0.2521,0.2576)	(0.272,0.2635,0.2595)
SC 12	(0.4239,0.3452,0.277)	(0.137,0.2,0.2527)	(0.4073,0.3404,0.2766)	(0.3929,0.3351,0.2764)	(0.3035,0.2912,0.2675)	(0.2878,0.2811,0.2676)	(0.3708,0.3102,0.2716)	(0.3329,0.3031,0.2712)
SC 13	(0.2653,0.2465,0.2336)	(0.2848,0.2468,0.2346)	(0.0622,0.1263,0.2123)	(0.2344,0.2296,0.2309)	(0.1837,0.2024,0.2261)	(0.221,0.2188,0.229)	(0.2393,0.2212,0.2297)	(0.1772,0.1991,0.2258)
SC 14	(0.2098,0.2399,0.2473)	(0.1778,0.2355,0.2445)	(0.2185,0.2451,0.2464)	(0.0843,0.1544,0.23)	(0.2592,0.2496,0.2476)	(0.2816,0.248,0.2463)	(0.1662,0.2165,0.2411)	(0.2179,0.2343,0.2436)
SC 21	(0.3019,0.2867,0.2754)	(0.3983,0.321,0.2791)	(0.4264,0.3293,0.2811)	(0.3905,0.3207,0.2798)	(0.1203,0.1902,0.2571)	(0.237,0.2797,0.2731)	(0.1937,0.2683,0.2722)	(0.3027,0.2992,0.2764)
SC 22	(0.2081,0.2125,0.2257)	(0.1198,0.2042,0.2258)	(0.1094,0.1915,0.224)	(0.1514,0.1975,0.2241)	(0.2146,0.2316,0.231)	(0.0701,0.1343,0.2136)	(0.325,0.2654,0.235)	(0.2753,0.2431,0.2308)
SC 23	(0.1757,0.2166,0.2359)	(0.2076,0.2067,0.2342)	(0.0844,0.1707,0.23)	(0.1615,0.1994,0.2342)	(0.2543,0.2533,0.2432)	(0.3393,0.2805,0.2485)	(0.0844,0.1502,0.2269)	(0.2955,0.2675,0.2465)
SC 24	(0.3143,0.2842,0.263)	(0.2743,0.2681,0.2609)	(0.3798,0.3085,0.2649)	(0.2965,0.2824,0.262)	(0.4107,0.3249,0.2686)	(0.3536,0.3055,0.2648)	(0.3969,0.3161,0.2658)	(0.1265,0.1901,0.2463)
SC 31	(0.5556,0.5212,0.5028)	(0.6809,0.579,0.5138)	(0.6534,0.5731,0.5117)	(0.34,0.4915,0.5003)	(0.4506,0.4794,0.4984)	(0.4934,0.5177,0.5045)	(0.4249,0.4776,0.4999)	(0.5619,0.5256,0.5044)

SC 32	(0.4444,0.4788, 0.4972)	(0.3191,0.421,0. 4862)	(0.3466,0.4269, 0.4883)	(0.66,0.5085,0.4 997)	(0.5494,0.5206, 0.5016)	(0.5066,0.4823, 0.4955)	(0.5751,0.5224, 0.5001)	(0.4381,0.4744, 0.4956)
SC 41	(0.3209,0.2776, 0.259)	(0.2957,0.2707, 0.2595)	(0.2222,0.2386, 0.2521)	(0.2089,0.2381, 0.2538)	(0.2305,0.2452, 0.2539)	(0.2369,0.2537, 0.2555)	(0.2668,0.2572, 0.2561)	(0.3515,0.2867, 0.2603)
SC 42	(0.251,0.2481,0. 2512)	(0.1968,0.2228, 0.2462)	(0.2815,0.2577, 0.2527)	(0.2468,0.2467, 0.2509)	(0.2606,0.2518, 0.2524)	(0.2935,0.2565, 0.2528)	(0.2131,0.2381, 0.2495)	(0.1621,0.2181, 0.2461)
SC 43	(0.1416,0.2065, 0.2393)	(0.1851,0.2236, 0.2419)	(0.1472,0.2213, 0.2414)	(0.2379,0.2408, 0.2452)	(0.2351,0.24,0.2 438)	(0.2595,0.2428, 0.2436)	(0.2933,0.2556, 0.2479)	(0.1987,0.228,0. 2437)
SC 44	(0.2864,0.2677, 0.2505)	(0.3223,0.2829, 0.2523)	(0.3491,0.2824, 0.2538)	(0.3063,0.2743, 0.2501)	(0.2738,0.2631, 0.2499)	(0.21,0.247,0.24 81)	(0.2268,0.2491, 0.2466)	(0.2877,0.2672, 0.2499)

ادامه جدول ۳-۲۵ سوپر ماتریس ناموزون - ماتریس نرمال ارتباط کامل زیرعوامل (T^{α})

	SC31	SC32	SC41	SC42	SC43	SC44
SC1 1	(0.2901,0.2618,0.259 4)	(0.1795,0.226,0.2534)	(0.2791,0.263,0.2601)	(0.279,0.266,0.2591)	(0.2533,0.2524,0.257 6)	(0.2436,0.2513,0.258 4)
SC1 2	(0.3554,0.2998,0.269 7)	(0.2168,0.2704,0.265 9)	(0.3576,0.3061,0.270 3)	(0.3251,0.2993,0.270 8)	(0.3397,0.298,0.2696)	(0.3323,0.2997,0.270 2)
SC1 3	(0.1562,0.2012,0.226 1)	(0.3245,0.2491,0.233 6)	(0.1517,0.1991,0.225 6)	(0.1528,0.1906,0.223 6)	(0.1689,0.2025,0.226 2)	(0.1754,0.2051,0.227 3)
SC1 4	(0.1983,0.2372,0.244 8)	(0.2793,0.2545,0.247 1)	(0.2116,0.2318,0.244)	(0.2431,0.244,0.2465)	(0.2381,0.2471,0.246 6)	(0.2486,0.2439,0.244)
SC2 1	(0.1895,0.2413,0.265 9)	(0.3031,0.2847,0.274 1)	(0.3385,0.3009,0.276 1)	(0.2815,0.2841,0.274 5)	(0.338,0.299,0.2752)	(0.3611,0.3022,0.275 6)
SC2 2	(0.1878,0.2153,0.227)	(0.1482,0.1913,0.223 4)	(0.1459,0.1846,0.221 4)	(0.2171,0.2191,0.228 2)	(0.2073,0.2116,0.225 8)	(0.2061,0.2184,0.227 3)
SC2 3	(0.3286,0.2668,0.245 6)	(0.3728,0.2832,0.246 7)	(0.2871,0.2571,0.243 7)	(0.253,0.2394,0.2401)	(0.2201,0.2333,0.241 4)	(0.1912,0.2215,0.239 1)
SC2 4	(0.2941,0.2766,0.261 5)	(0.1759,0.2408,0.255 8)	(0.2286,0.2573,0.258 8)	(0.2484,0.2574,0.257 1)	(0.2346,0.2561,0.257 7)	(0.2416,0.2579,0.258)
SC3 1	(0.1087,0.2726,0.468 6)	(0.8462,0.703,0.5309)	(0.5959,0.5271,0.506 3)	(0.4553,0.4788,0.498 3)	(0.2741,0.416,0.4895)	(0.6745,0.5477,0.508 8)

SC3 2	(0.8913,0.7274,0.531 4)	(0.1538,0.297,0.4691)	(0.4041,0.4729,0.493 7)	(0.5447,0.5212,0.501 7)	(0.7259,0.584,0.5105)	(0.3255,0.4523,0.491 2)
SC4 1	(0.3167,0.2723,0.258 9)	(0.1764,0.2343,0.252 3)	(0.0985,0.1673,0.239 8)	(0.2576,0.2651,0.257 9)	(0.294,0.2788,0.2595)	(0.3297,0.2906,0.260 3)
SC4 2	(0.2335,0.2426,0.250 3)	(0.1735,0.2285,0.248)	(0.3153,0.2809,0.257 2)	(0.1038,0.1697,0.237 4)	(0.3151,0.2807,0.256 6)	(0.2993,0.2752,0.255 8)
SC4 3	(0.2392,0.2428,0.243 5)	(0.4717,0.3254,0.257 3)	(0.2584,0.2591,0.248 3)	(0.2937,0.269,0.2492)	(0.0906,0.1577,0.230 1)	(0.2625,0.2598,0.249 8)
SC4 4	(0.2106,0.2423,0.247 3)	(0.1784,0.2118,0.242 4)	(0.3277,0.2927,0.254 7)	(0.3449,0.2962,0.255 5)	(0.3002,0.2828,0.253 8)	(0.1086,0.1744,0.234 2)

۳-۹،۹،۴ گام نهم: تشکیل سوپر ماتریس موزون (W^α)

در این مرحله ماتریس T_D^α (جدول ۳-۲۴) را در ماتریس W (T_C^α) (جدول ۳-۲۵) ضرب می‌کنیم. به این طریق که هر $t_D^{\alpha ij}$ را در W_{ij} نظیر ضرب می‌کنیم (جدول ۳-۲۶).

۳-۹،۹،۱۰ گام دهم: محدود کردن سوپر ماتریس موزون

مطابق با رابطه ۴-۱۴، سوپر ماتریس موزون را آنقدر به توان (متوالی اعداد فرد) رسانیده تا تمامی اعداد هر سطر همگرا شوند.

$$\lim_{Z \rightarrow \infty} (W^{\alpha l})^Z, \quad \lim_{Z \rightarrow \infty} (W^{\alpha m})^Z, \quad \lim_{Z \rightarrow \infty} (W^{\alpha u})^Z \quad ۳-۶۴$$

در ماتریس نهایی همگرا شده اعداد نشان دهنده‌ی وزن نسبی زیرعوامل تحقیق می‌باشند، که بر اساس آن می‌توان عوامل اصلی را با توجه به میزان اهمیتشان اولویت بندی نمود (جدول ۳-۲۶).

جدول ۳-۲۶ سوپر ماتریس موزون (W^α)

	SC11	SC12	SC13	SC14	SC21	SC22
SC11	(0.0308,0.0465,0.0645)	(0.122,0.0878,0.0714)	(0.0951,0.0797,0.0705)	(0.0879,0.0776,0.07)	(0.0729,0.0714,0.0691)	(0.0603,0.0701,0.0687)
SC12	(0.1292,0.0954,0.0738)	(0.0418,0.0553,0.0673)	(0.1241,0.0941,0.0737)	(0.1198,0.0926,0.0736)	(0.0873,0.0809,0.0715)	(0.0827,0.0781,0.0715)
SC13	(0.0809,0.0681,0.0622)	(0.0868,0.0682,0.0625)	(0.019,0.0349,0.0565)	(0.0714,0.0635,0.0615)	(0.0528,0.0562,0.0604)	(0.0635,0.0608,0.0612)
SC14	(0.0639,0.0663,0.0659)	(0.0542,0.0651,0.0651)	(0.0666,0.0677,0.0656)	(0.0257,0.0427,0.0612)	(0.0745,0.0694,0.0662)	(0.081,0.0689,0.0658)
SC21	(0.076,0.0733,0.0699)	(0.1003,0.0821,0.0708)	(0.1074,0.0842,0.0713)	(0.0984,0.082,0.071)	(0.0276,0.0458,0.0645)	(0.0545,0.0674,0.0685)
SC22	(0.0524,0.0543,0.0573)	(0.0302,0.0522,0.0573)	(0.0276,0.049,0.0568)	(0.0381,0.0505,0.0569)	(0.0493,0.0558,0.0579)	(0.0161,0.0324,0.0535)
SC23	(0.0443,0.0554,0.0598)	(0.0523,0.0529,0.0594)	(0.0213,0.0436,0.0584)	(0.0407,0.051,0.0594)	(0.0584,0.061,0.061)	(0.078,0.0676,0.0623)
SC24	(0.0792,0.0727,0.0667)	(0.0691,0.0686,0.0662)	(0.0957,0.0789,0.0672)	(0.0747,0.0722,0.0665)	(0.0944,0.0783,0.0673)	(0.0813,0.0736,0.0664)
SC31	(0.077,0.092,0.1059)	(0.0943,0.1022,0.1082)	(0.0905,0.1012,0.1078)	(0.0471,0.0867,0.1054)	(0.0825,0.0929,0.1063)	(0.0903,0.1003,0.1076)
SC32	(0.0615,0.0845,0.1047)	(0.0442,0.0743,0.1024)	(0.048,0.0753,0.1028)	(0.0914,0.0898,0.1052)	(0.1005,0.1009,0.1069)	(0.0927,0.0935,0.1056)
SC41	(0.0978,0.0809,0.0698)	(0.0901,0.0789,0.0699)	(0.0677,0.0695,0.0679)	(0.0637,0.0694,0.0684)	(0.0691,0.0704,0.0683)	(0.071,0.0729,0.0687)
SC42	(0.0765,0.0723,0.0677)	(0.06,0.0649,0.0663)	(0.0858,0.0751,0.0681)	(0.0752,0.0719,0.0676)	(0.0781,0.0723,0.0679)	(0.088,0.0737,0.068)
SC43	(0.0432,0.0602,0.0645)	(0.0564,0.0652,0.0652)	(0.0449,0.0645,0.065)	(0.0725,0.0702,0.0661)	(0.0705,0.0689,0.0656)	(0.0778,0.0697,0.0655)
SC44	(0.0873,0.078,0.0675)	(0.0982,0.0824,0.068)	(0.1064,0.0823,0.0684)	(0.0934,0.0799,0.0674)	(0.0821,0.0756,0.0672)	(0.0629,0.0709,0.0667)

ادامه جدول ۳-۲۶ سوپر ماتریس موزون (W^{α})

	SC31	SC32	SC41	SC42	SC43	SC44
SC1 1	(0.065,0.0668,0.068 5)	(0.0402,0.0577,0.06 69)	(0.0843,0.075,0.069 9)	(0.0843,0.0759,0.06 96)	(0.0765,0.072,0.069 2)	(0.0736,0.0717,0.06 94)
SC1 2	(0.0796,0.0765,0.07 12)	(0.0486,0.069,0.070 2)	(0.108,0.0873,0.072 6)	(0.0982,0.0854,0.07 27)	(0.1026,0.085,0.072 4)	(0.1004,0.0855,0.07 26)
SC1 3	(0.035,0.0514,0.059 7)	(0.0727,0.0636,0.06 16)	(0.0458,0.0568,0.06 06)	(0.0461,0.0544,0.06 01)	(0.051,0.0578,0.060 8)	(0.053,0.0585,0.061 1)
SC1 4	(0.0444,0.0606,0.06 46)	(0.0626,0.065,0.065 2)	(0.0639,0.0661,0.06 55)	(0.0734,0.0696,0.06 62)	(0.0719,0.0705,0.06 62)	(0.0751,0.0696,0.06 55)
SC2 1	(0.0481,0.0624,0.06 77)	(0.077,0.0736,0.069 8)	(0.0958,0.0806,0.07 05)	(0.0796,0.0761,0.07 01)	(0.0956,0.0801,0.07 03)	(0.1022,0.081,0.070 4)
SC2 2	(0.0477,0.0557,0.05 78)	(0.0376,0.0495,0.05 69)	(0.0413,0.0495,0.05 66)	(0.0614,0.0587,0.05 83)	(0.0586,0.0567,0.05 77)	(0.0583,0.0585,0.05 81)
SC2 3	(0.0835,0.069,0.062 5)	(0.0947,0.0732,0.06 28)	(0.0812,0.0689,0.06 23)	(0.0716,0.0642,0.06 13)	(0.0623,0.0625,0.06 17)	(0.0541,0.0594,0.06 11)
SC2 4	(0.0747,0.0715,0.06 66)	(0.0447,0.0623,0.06 51)	(0.0647,0.069,0.066 1)	(0.0703,0.069,0.065 7)	(0.0664,0.0686,0.06 58)	(0.0683,0.0691,0.06 59)
SC3 1	(0.0282,0.0576,0.10 05)	(0.2197,0.1485,0.11 39)	(0.0823,0.0944,0.10 68)	(0.0629,0.0858,0.10 51)	(0.0379,0.0745,0.10 32)	(0.0931,0.0981,0.10 73)

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۲۳

SC3 2	(0.2314,0.1536,0.114)	(0.0399,0.0627,0.106)	(0.0558,0.0847,0.1041)	(0.0752,0.0933,0.1058)	(0.1002,0.1046,0.1077)	(0.045,0.081,0.1036)
SC4 1	(0.0831,0.0749,0.0692)	(0.0463,0.0644,0.0674)	(0.0273,0.0448,0.0635)	(0.0713,0.0709,0.0683)	(0.0814,0.0746,0.0688)	(0.0913,0.0778,0.069)
SC4 2	(0.0613,0.0667,0.0669)	(0.0455,0.0628,0.0662)	(0.0873,0.0752,0.0682)	(0.0287,0.0454,0.0629)	(0.0873,0.0751,0.068)	(0.0829,0.0736,0.0678)
SC4 3	(0.0628,0.0667,0.065)	(0.1238,0.0895,0.0687)	(0.0716,0.0693,0.0658)	(0.0813,0.072,0.066)	(0.0251,0.0422,0.061)	(0.0727,0.0695,0.0662)
SC4 4	(0.0553,0.0666,0.0661)	(0.0468,0.0582,0.0647)	(0.0907,0.0783,0.0675)	(0.0955,0.0793,0.0677)	(0.0831,0.0757,0.0673)	(0.0301,0.0467,0.0621)

ماتریس همگرا شده پس از دیفازی کردن بصورت جدول ۳-۲۷ می‌باشد.

جدول ۳-۲۷ سوپر ماتریس موزون همگرا شده

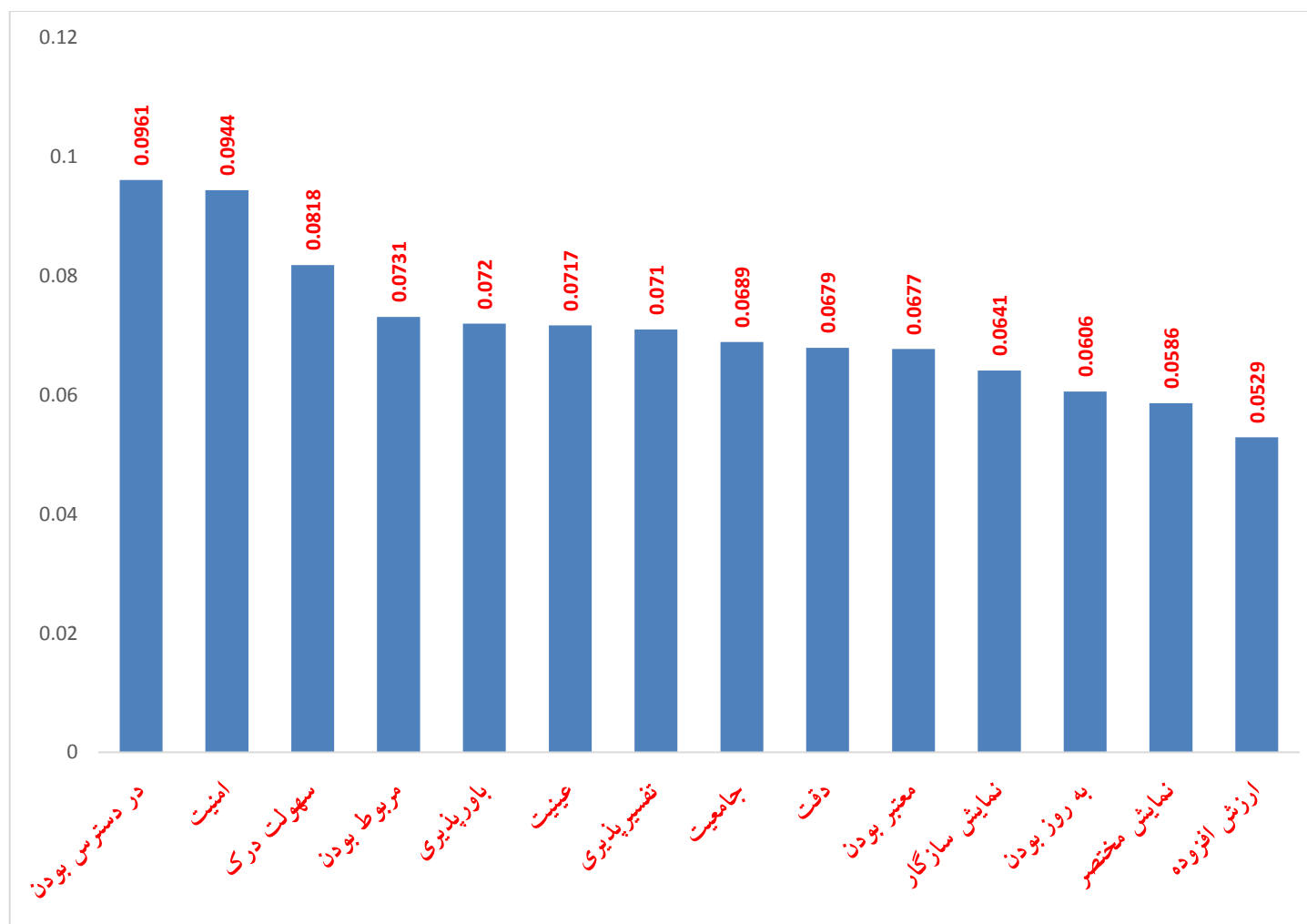
[illegible]

[illegible]

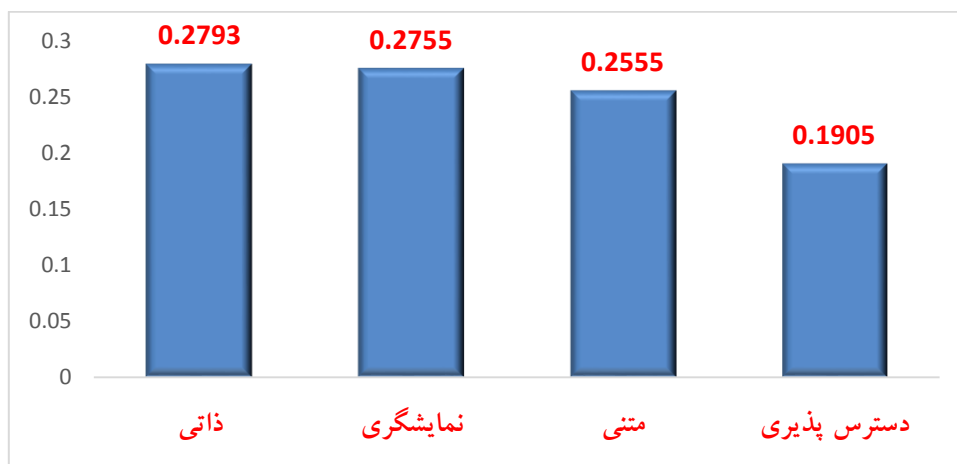
۳-۵ رتبه بندی عوامل و زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور

جدول ۳-۲۸ وزن و رتبه عوامل و زیرعوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور

عوامل اصلی	زیرعوامل	علائم اختصاری	وزن نسبی زیرعوامل و اولویت آنها	وزن نهایی زیر عوامل و اولویت آن ها	
نمایشگری (0.2755)	تفسیرپذیری	SC11	0.257713	2	7
	سهولت درک	SC12	0.296915	1	3
	نمایش مختصر	SC13	0.212704	4	13
	نمایش سازگار	SC14	0.232668	3	11
متنی (0.2555)	مربوط بودن	SC21	0.286106	1	4
	ارزش افزوده	SC22	0.207045	4	14
	به روز بودن	SC23	0.237182	3	12
	جامعیت	SC24	0.269667	2	8
دسترس پذیری (0.1905)	در دسترس بودن	SC31	0.504462	1	1
	امنیت	SC32	0.495538	2	2
	عینیت	SC41	0.256713	2	6
	دقت	SC42	0.243108	3	9
ذاتی (0.2793)	معتبر بودن	SC43	0.242392	4	10
	باورپذیری	SC44	0.257787	1	5



نمودار ۳-۱۰ رتبه بندی نهایی زیرعوامل موثر بر موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور



نمودار ۳-۱۱ رتبه بندی نهایی عوامل موثر بر کیفیت فراداده‌های سامانه ملی ثبت پایان نامه داخل کشور

فصل چهارم – اندازه‌گیری و تحلیل شاخص‌های کیفیت داده به کمک نمودار کنترل

۴-۱ مقدمه

نمودار کنترل (Control Chart) ابزاری آماری برای تشخیص علل نوسانات مشاهده شده در یک فرآیند است. نمودار کنترل تصویری از ثبات یا عدم ثبات فرآیند را عرضه می‌کند. هر فرآیند دارای نوساناتی است. بعضی از نوسانات از عللی غیر معمول ریشه می‌گیرند، که آنها را علل خاص می‌نامند. برخی دیگر از نوسانات مشاهده شده در فرآیند از عللی حاصل می‌گردند که جزو ذات فرایند می‌باشند و علل عام خوانده می‌شوند. نمودار کنترل تفاوت بین این نوع نوسانات را ظاهر می‌سازد.

یکی از اهداف نمودارهای کنترل دستیابی به ثبات در فرآیند و حفظ آن می‌باشد. ثبات در فرآیند بر مبنای شرایطی تعیین می‌گردد که طی آن فرآیند درجه بالایی از یکنواختی عملکرد در گذشته را نشان داده و انتظار آن می‌رود که این یکنواختی در آینده نیز حفظ شود. مشخصه این یکنواختی در عملکرد، قرار گرفتن داده‌ها در حدود کنترل است. حدود کنترل در فاصله مثبت و منفی سه برابر انحراف معیار ($\pm 3\sigma$)، مثبت و منفی سه سیگما) جامعه آماری مورد بررسی از خط مرکزی واقع شده‌اند. حدود کنترل نمایانگر محدوده تغییرات و یا نوساناتی است که انتظار می‌رود یک فرآیند تحت کنترل آماری از خود نشان دهد. این نمودار شامل یک خط مرکز^۱ است که مقدار متوسط مشخصه کیفی را در حالت تحت کنترل و یا به عبارتی در حالتی که فقط خطاهای تحت کنترل قرار دارند نشان می‌دهد. دوخط افقی دیگر که حد کنترل پایین^۲ و حد کنترل بالا^۳ نامیده می‌شوند در نمودار نشان داده شده‌اند. (داگلاس مونتگومری^۴، ۲۰۰۷)

فرض کنید مشخصه کیفی مورد نظر به وسیله آماره W اندازه‌گیری می‌شود. همچنین فرض کنید میانگین و انحراف معیار W به ترتیب برابر μ_W و σ_W باشد. در این صورت، خط مرکز، حد کنترل بالا و حد کنترل پایین برابر با روابط زیر است.

$$UCL = \mu_W + k\sigma_W \quad (۴-۱)$$

$$CL = \mu_W \quad (۴-۲)$$

$$LCL = \mu_W - k\sigma_W \quad (۴-۳)$$

در رابطه فوق k فاصله حدود کنترل از خط مرکز را بر حسب واحد انحراف معیار نشان می‌دهد.

^۱ Center line

^۲ Lower Control Limit

^۳ Upper Control Limit

^۴ Montgomery, D.C., 2009

انحراف معیار پارامتری است که شکل پراکندگی توزیع متغیر را بیان می‌نماید. در یک جامعه آماری نرمال مقدار انحراف معیار از رابطه زیر محاسبه می‌شود.

$$\sigma = \frac{\sum \sqrt{(X - \bar{X})^2}}{n} \quad (4-4)$$

در این رابطه:

$$\bar{X} = \text{میانگین جامعه آماری}$$

$$n = \text{تعداد اعضای جامعه آماری}$$

حدود کنترلی فرآیند به وسیله UCL و LCL مشخص شود که به کمک توزیع‌های آماری حدود کنترل $\mu \pm 3\sigma$ به تعیین و تفکیک علل عام و علل خاص انحرافات می‌پردازد.

اگر میانگین دامنه را از رابطه (۴-۵) و میانگین کلی را از رابطه (۴-۶) محاسبه کنیم آنگاه شکل نهایی حدود بالایی و پایینی کنترل برای دامنه و میانگین به ترتیب از روابط (۴-۷) و (۴-۸) بدست خواهد آمد.

$$\bar{R} = \frac{\sum_{i=1}^g R_i}{g} \quad (4-5)$$

$$\bar{\bar{X}} = \frac{\sum_{i=1}^g \bar{X}_i}{g} \quad (4-6)$$

$$LCL_R = \bar{R} - 3\sigma_R \quad \text{و} \quad UCL_R = \bar{R} + 3\sigma_R \quad (4-7)$$

$$LCL_{\bar{X}} = \bar{\bar{X}} - 3\sigma_{\bar{X}} \quad \text{و} \quad UCL_{\bar{X}} = \bar{\bar{X}} + 3\sigma_{\bar{X}} \quad (۴-۸)$$

نمودارهای کنترل بطور کلی شامل دو دسته زیر هستند.

الف) نمودارهای کنترل داده‌های وصفی (Attributes)

ب) نمودارهای کنترل داده‌های کمی (Variables)

در نمودارهای کنترل داده‌های وصفی داده‌های مربوط به دفعات مشاهده یا رخداد یک حالت خاص نظیر رد/قبول، بله/خیر یا وجود/عدم وجود یک مشخصه خاص مورد بررسی قرار می‌گیرد.

در نمودارهای کنترل داده‌های متغیر مقادیر یک متغیر مانند ابعاد یک محصول، سیکل زمانی ارائه یک خدمات، درجه حرارت کوره و ... بطور پیوسته مورد بررسی قرار می‌گیرد. منظور از پیوسته بودن مقادیر این است که متغیر مورد اندازه‌گیری بتواند هر مقداری را در دامنه تغییرات خود، بدون التزام به گسسته بودن آن مقدار کسب کند.

نمودارهای متفاوتی برای کنترل هر دو نوع داده‌های ذکر شده مورد استفاده قرار می‌گیرند، ولی در فعالیتهای بهبود فرآیندها عمدتاً نمودارهای کنترل مربوط به داده‌های متغیر مورد استفاده قرار می‌گیرند.

در این فصل بناداریم به کمک نمودارهای کنترل، کیفیت فراداده‌های پایان‌نامه/رساله‌های دانش‌آموختگان کل کشور را برپایه شاخصها و روابطی که در فصل سوم استخراج شده‌اند، پایش آماری کنیم. نمودارهای میانگین-دامنه، میانگین متحرک موزون نمایی (EWMA)، نمودار I-MR و همینطور نمودار چندمتغیره T^2 نمودارهایی هستند که در این بخش مورد استفاده قرار گرفته‌اند.

در ادامه این فصل ابتدا در بخش ۴-۲ به معرفی نمودارهای کنترل مورد استفاده قرار گرفته خواهیم پرداخت و سپس نتایج استفاده از این نمودارها را در بخش ۴-۳ ارائه خواهیم نمود. چگونگی پیاده‌سازی را در همین بخش در قالب فازهای یک و دو ارائه خواهیم کرد.

۴-۲ معرفی نمودارهای کنترل

۴-۱-۲ نمودار کنترل XBar-R

برای یک مشخصه کیفی قابل اندازه‌گیری متداول است که میانگین و تغییرپذیری به طور هم زمان پایش شوند. میانگین از این نظر مهم است که نشان می‌دهد آیا فرآیند در حول مقدار هدف عمل می‌کند یا خیر و تغییرپذیری از این لحاظ مهم است که نشان می‌دهد آیا فرآیند یکسان عمل می‌کند یا خیر.

نمودار $\bar{X}$ برای پایش متوسط کیفیت محصول و نمودار دامنه غالباً برای پایش انحراف معیار استفاده می‌شود.

حدود کنترل برای نمودار $\bar{X}$ به شرح زیر هستند.

$$(۴-۹) \quad UCL = \bar{\bar{X}} + A_2 \bar{R}$$

$$(۴-۱۰) \quad CL = \bar{\bar{X}}$$

$$(۴-۱۱) \quad LCL = \bar{\bar{X}} - A_2 \bar{R}$$

A_2 : عدد ثابتی است که برای اندازه‌های مختلف نمونه بک کمک جدول تعیین می‌شود.

$\bar{\bar{X}}$: میانگین نمونه می‌باشد

حدود کنترل برای نمودار R بصورت زیر تعیین می‌شوند.

$$(۴-۱۲) \quad UCL = D_4 \bar{R}$$

$$(۴-۱۳) \quad CL = \bar{R}$$

$$(۴-۱۴) \quad LCL = D_3 \bar{R}$$

ثابت‌های D_3 و D_4 برای مقادیر مختلف n بک کمک جدول تعیین می‌شود. R دامنه نمونه می‌باشد.

۴-۲-۲ نمودار کنترل I-MR

محدوده حرکت فردی یا I-MR، یک نوع کنترل است که معمولاً برای داده‌های پیوسته استفاده می‌شود (نوعی داده‌های ارجاع). این نمودار در ابتدا توسط والتر شوهارت توسعه یافته و از آن پس نمودارهای کنترل گاهی اوقات نیز به عنوان نمودار شوهارت نامیده می‌شود. همانطور که می‌دانیم، در I-MR دو نمودار داریم: یکی داده‌های فردی (نمودار انفرادی یا I chart) و نمودار دیگر دارای نقاط داده است که مقادیر محدوده‌ای بین نقاط مختلف متوالی (نمودار محدوده حرکت یا نمودار (MR)) می‌باشد. این دو نمودار با هم اطلاعات مفیدی را ارائه می‌کنند و استفاده از یکی از آنها به تنهایی ممکن است اطلاعات کامل در مورد رفتار فرآیند را نشان ندهد. داده‌های نمودار I-MR را نمیتوان به صورت زیرگروه طبقه بندی کرد زیرا در نمودار R-Chart داده‌ها به عنوان اطلاعات فردی و محدوده نقاط برای تفسیر صحیح پردازش اطلاعات مهم هستند. (موکوندام و همکاران^۱ ۲۰۱۳)

در این بخش جهت آشنایی بهتر و دقیق‌تر به طور جداگانه هریک از نمودارهای I-MR را بررسی می‌کنیم.

• نمودار فردی

نمودار فردی نشان‌دهنده داده‌های فردی در طی یک دوره 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 تولید شده باید پی در پی باشند. اندازه نمونه در نمودار انفرادی ۱ و روابط بشرح زیر می‌باشد.

$$(۴-۱۵) \quad UCL = \bar{x} + 3 \frac{\overline{MR}}{d_2}$$

$$(۴-۱۶) \quad CL = \bar{x}$$

$$(۴-۱۷) \quad LCL = \bar{x} - 3 \frac{\overline{MR}}{d_2}$$

در روابط بالا MR محدوده حرکتی^۲ می‌باشد و $\bar{x}$ میانگین را نشان می‌دهد. d_2 عدد ثابتی است که از جدول پیوست استخراج می‌شود.

^۱ K. Mukundam.al.2013

^۲ Moving Range

• نمودار محدوده حرکت

نمودار محدوده حرکت همانطور که از نام آن مشخص است، نموداری است که توسط مقادیر حاصل از داده‌های متوالی مرتب شده بر اساس زمان ترسیم می‌شود. هر نقطه در محدوده جابجایی به صورت $X_n - X_{n-1}$ محاسبه می‌شود. (قبادی و نقندریان، ۲۰۱۲)

$$MR_i = |X_i - X_{i-1}| \quad (۴-۱۸)$$

X_i : پارامتر کیفیت محصول در زیر گروه i

برخی از مراقبت‌ها باید در تفسیر الگوها، در نمودار محدوده حرکتی انجام شود. محدوده حرکتی همبستگی دارند و این همبستگی اغلب ممکن است الگوی چرخه را در نمودار تغییر دهد. بعضی از محققان توصیه می‌کنند که نمودار، محدوده جابجایی نداشته باشند (ریگدان و همکاران، ۱۹۹۴).^۱

۲،۳-۴ نمودار کنترل EWMA

یک نمودار کنترل مفید برای شناسایی تغییرات کوچک و یا کار با داده‌های غیر نرمال می‌تواند جایگزین مناسبی برای نمودار کنترل شوهارت باشد. در نمودار EWMA به هر مشاهده وزنی تخصیص داده می‌شود به طوری که مشاهده اخیر وزن بیشتری نسبت به مشاهدات قبل می‌گیرد. وزن تخصیص داده شده به وسیله پارامتر λ که مقداری بین ۰/۰۵ و ۰/۲۵ دارد کنترل می‌شود. همچنین یک ضریب L پهنای حدود کنترل را کنترل می‌کند. (داگلاس مونتگومری، ۲۰۰۷).^۲

حدود کنترل نمودار EWMA از روابط زیر محاسبه می‌شود.

$$UCL = \mu + L\sigma\sqrt{\frac{\lambda}{2-\lambda}}[1 - (1 - \lambda)^2] \quad (۴-۱۹)$$

$$(۴-۲۰)$$

$$CL = \mu$$

$$(۴-۲۱) \quad LCL = \mu - L\sigma \sqrt{\frac{\lambda}{2-\lambda}} [1 - (1 - \lambda)^2]$$

۲،۴-۴ نمودار چندمتغیره T^2

شرایط و موقعیت‌های بسیاری وجود دارد که نظارت یا کنترل همزمان دو یا چند ویژگی مرتبط با کیفیت ضروری است. در فرآیندهای چند متغیره که نیازمند کنترل دو یا چند مشخصه کیفی به طور همزمان هستند، باید دو هدف عمده محقق گردد. هدف اول تشخیص وضعیت‌های خارج از کنترل و هدف دوم شناسایی مشخصه‌های کیفی منحرف شده، در زمان وقوع یک وضعیت خارج از کنترل می‌باشد. (شهریاری و حافظی، ۱۳۸۸) در مبحث کنترل فرایند آماری، برای پایش یک فرایند که دارای بیش از یک متغیر بوده متغیرهای آن مستقل هستند میتوان از نمودارهای کنترل چند متغیره مانند T^2 استفاده کرد. به عنوان مثال، فرض کنید که یک حامل دارای دو قطر داخلی x_1 و یک قطر بیرونی x_2 است که با استفاده از آن، سودمندی قطعه را تعیین می‌کند. فرض کنید x_1 و x_2 دارای توزیع نرمال مستقل هستند. از آنجا که هر دو ویژگی کیفیت اندازه‌گیری می‌شوند، میتوان از نمودار کنترل T^2 استفاده کرد. یکی از ویژگی‌های این نمودار کنترل حفظ توالی زمان داده‌ها به طوری که می‌توان الگوهای تصادفی و الگوهای غیر تصادفی را مورد بررسی قرار داد. علاوه بر این، مزیت دیگری نیز دارد که "حالت" فرآیند با یک مقدار آماری مشخص می‌کند. فرض کنید که توزیع احتمالی مشترک از ویژگی‌های کیفی p توزیع نرمال متغیر p است. برای محاسبه نمودار کنترلی نیاز به محاسبه میانگین نمونه برای هر یک از مشخصات کیفی p از نمونه‌ای از اندازه n داریم. این مجموعه وسیله $p \times$ ۱ نمایش داده می‌شود.

$$(۴-۲۲) \quad \bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

آمار آزمون بر روی نمودار کنترل مربع چی برای هر نمونه طراحی شده است

$$(۴-۲۳) \quad X_0^2 = n(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)$$

جایی که $\hat{\mu} = [\mu_1, \mu_2, \dots, \mu_p]$ بردار ابزار کنترل در هر کدام از ویژگی‌های کیفیت است و Σ ماتریس کوواریانس است. حد بالای نمودار کنترل طبق رابطه زیر می‌باشد.

$$(۴-۲۴) \quad UCL = X_{\alpha.p}^2$$

از آنجایی که $p \times p$ ماتریس کوواریانس نمونه S به صورت زیر است:

$$(۴-۲۵) \quad S = \begin{bmatrix} \bar{s}_1^2 & \bar{s}_{12} & \bar{s}_{13} & \cdots & \bar{s}_{1p} \\ \vdots & \bar{s}_2^2 & \bar{s}_{23} & \cdots & \bar{s}_{2p} \\ \vdots & \vdots & \bar{s}_3^2 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \bar{s}_p^2 \end{bmatrix}$$

ماتریس بالا میانگین نمونه ماتریس کوواریانس S یک برآورد بی طرفی از Σ است زمانی که فرایند تحت کنترل می‌باشد. حال فرض می‌کنیم که S از رابطه بالا برای برآورد Σ استفاده می‌شود و متغیر X به عنوان مقدار کنترل درون میانگین در نظر گرفته می‌شود. اگر μ با $\bar{x}$ و Σ با S در رابطه (۴-۲۳) جایگزین کنیم، بنابراین آماره آزمون T^2 به صورت رابطه زیر می‌باشد.

$$(۴-۲۶) \quad T^2 = n(\bar{x} - \bar{\bar{x}})S^{-1}(\bar{x} - \bar{\bar{x}})$$

رابطه بالا یک آزمون کنترل غیر مستقیم است. بدین معنی که توانایی آن برای تشخیص و تغییر بردار تنها برای اندازه بردار کاربرد دارد و برای جهت کارایی ندارد.

این نمودار برای کنترل تعداد نامنطبقها در هر محصول به گونه ای که C تعداد نامنطبقها در هر واحد است. اندازه نمونه در این نمودار یک واحد بازرسی شده است. در صورتیکه در یک دوره بازرسی متوسط تعداد نقص را $\bar{C}$ بنامیم، خط مرکزی و حدود بالا و پایین این نمودار بصورت زیر قابل محاسبه است.

$$(۴-۲۷) \quad CL = \bar{C}$$

$$(۴-۲۸) \quad UCL = \bar{C} + 3\sqrt{\bar{C}}$$

$$(۴-۲۹) \quad LCL = \bar{C} - 3\sqrt{\bar{C}}$$

۲,۶-۴ نمودار تعداد نقص در واحد (U)

این نمودار برای کنترل تعداد نامنطبقها در واحد هر محصول به گونه ای که u تعداد نامنطبقها به ازای هر واحد است.. اندازه نمونه در این نمودار میتواند متغیر باشد. خط مرکزی و حدود بالا و پایین این نمودار بصورت زیر قابل محاسبه است.

$$(۴-۳۰) \quad CL = \bar{U}$$

$$(۴-۳۱) \quad UCL = \bar{U} + 3\sqrt{\frac{\bar{U}}{n}}$$

$$(۴-۳۲) \quad LCL = \bar{U} - 3\sqrt{\frac{\bar{U}}{n}}$$

در ادامه این بخش به چگونگی پیاده‌سازی انواع نمودارهای کنترل معرفی شده در این بخش خواهیم پرداخت.

۳-۴ پیاده‌سازی نمودارهای کنترل

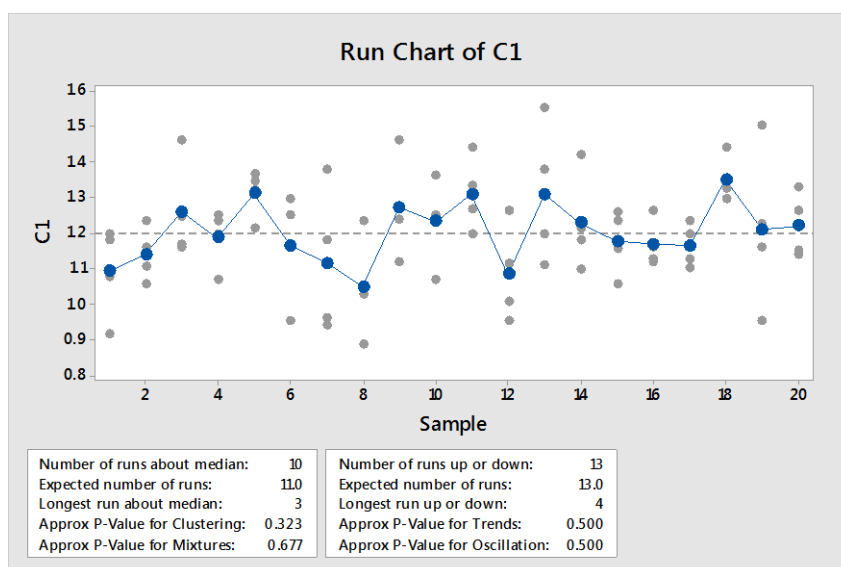
در این بخش پس از شناخت تعاریف و لزوم استفاده از نمودارهای کنترل و نحوه به کارگیری آن در داده‌های به دست آمده از سامانه ثبت به محاسبه شاخص‌های معرفی شده در فصل سوم می‌پردازیم و پس از محاسبه اندازه نمونه، اطلاعات دریافتی از پایان‌نامه دانشجویان را با استفاده از نرم افزار مینی تب^۱ پایش می‌کنیم.

این بخش شامل دو فاز می‌باشد که در هر فاز تست نرمال بودن داده‌ها، هیستوگرام داده‌ها و نمودارهای I-MR, RUN chart و نمودارهای کنترلی EWMA, Xbar-R را برای هر شاخص اندازه‌گیری کرده‌ایم، سپس در فاز بعد با حذف نقاط خارج از کنترل بار دیگر آزمون‌های فوق را اندازه‌گیری می‌کنیم و نتایج را بررسی خواهیم کرد.

۳،۱-۴ فاز اول

قبل از اینکه نمودارهای کنترل را رسم کنیم لازم است روندهای غیر تصادفی کلی را از داده‌ها برطرف کنیم، که این کار توسط RUN Chart در نرم افزار مینی تب انجام می‌شود که در آن چهار فرض گرایش‌ها نوسان‌ها، دسته بندی‌ها و ترکیب‌ها به صورت دو به دو بررسی می‌شوند. در این آزمون فرض صفر برابر این است که هیچ روند غیر تصادفی در فرآیند دیده نمی‌شود و فرض یک برابر با وجود حداقل یکی از چهار روند غیر تصادفی ذکر شده است و در صورتی که P-value بزرگتر از α باشد فرض صفر قبول می‌باشد. (پیش فرض برنامه $\alpha = 0.05$ است)

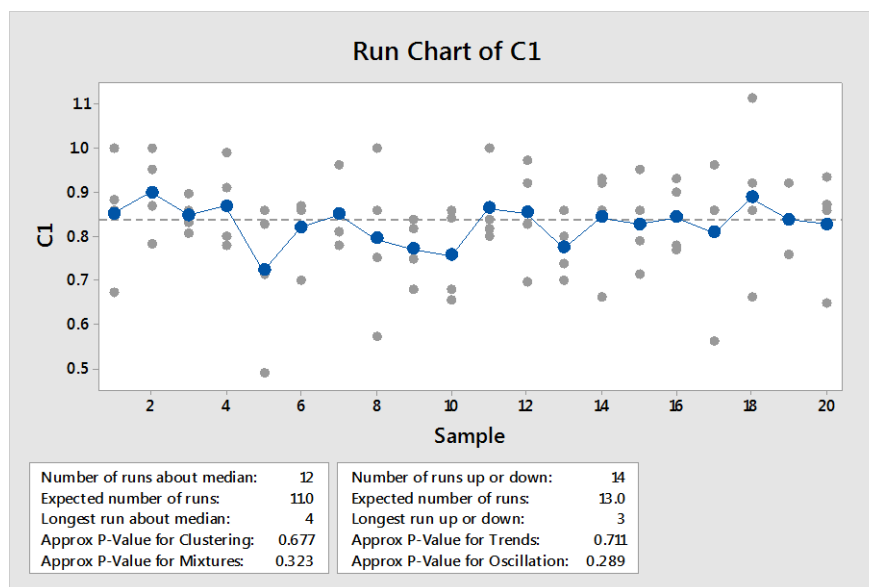
• آزمون تصادفی بودن شاخص کامل بودن



شکل (۴-۱). آزمون تصادفی بودن شاخص کامل بودن

همانطور که در نمودار بالا مشاهده می‌کنیم، مقدار P-value بزرگتر از ۰/۰۵ می‌باشد. پس داده‌ها برای شاخص کامل بودن تصادفی می‌باشند.

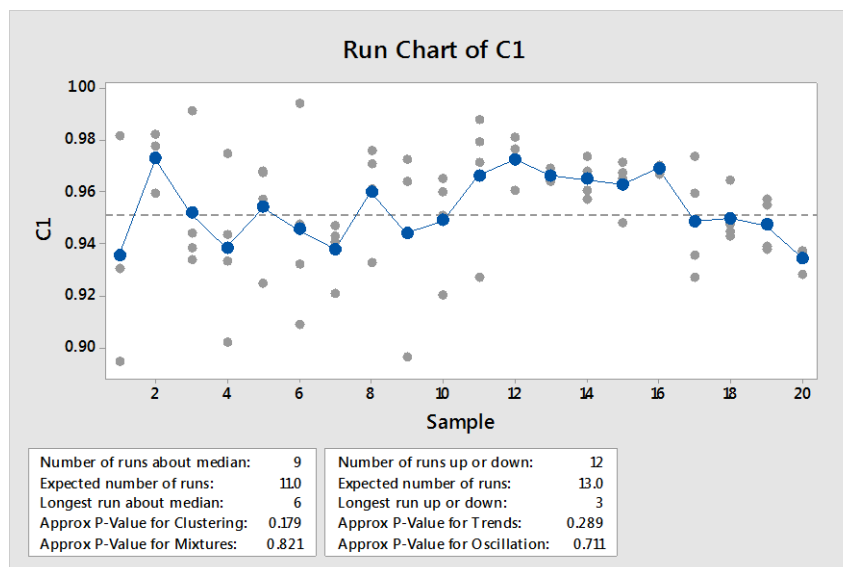
• آزمون تصادفی بودن شاخص صحت



شکل (۴-۲). آزمون تصادفی بودن شاخص صحت

همانطور که در نمودار بالا مشاهده میکنیم، مقدار P-value بزرگتر از ۰/۰۵ می‌باشد. پس داده‌ها برای شاخص صحت نیز تصادفی می‌باشند.

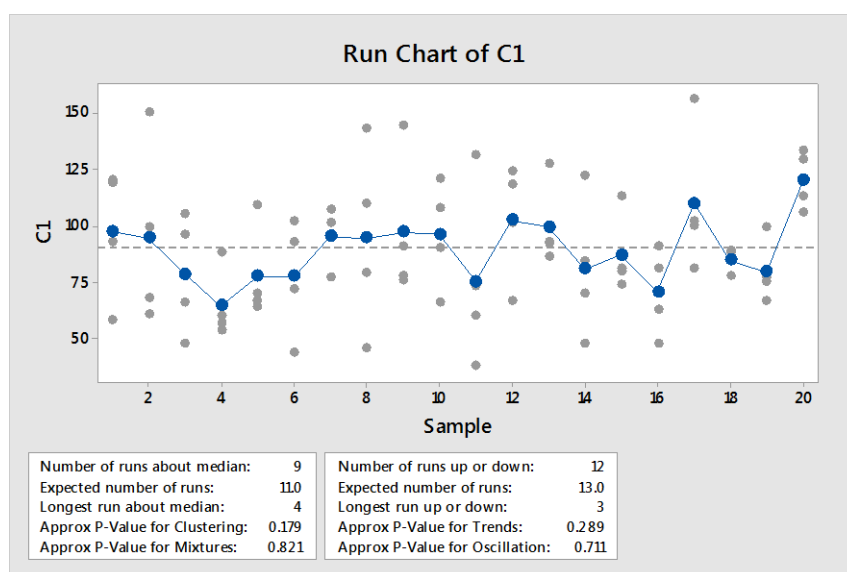
• آزمون تصادفی بودن شاخص دقت



شکل (۳-۴). آزمون تصادفی بودن شاخص دقت

همانطور که در نمودار بالا مشاهده میکنیم، مقدار P-value برای شاخص دقت مقدار ۰/۲ می‌باشد که بزرگتر از ۰/۰۵ است، پس داده‌ها برای شاخص دقت نیز وضعیت مطلوبی دارند.

• آزمون تصادفی بودن شاخص به هنگام بودن



شکل (۴-۴). آزمون تصادفی بودن شاخص به هنگام بودن

همانطور که در چهار نمودار بالا مشاهده می‌کنیم، آزمون تصادفی بودن داده‌ها برای هر چهار شاخص بزرگتر از ۰/۰۵ می‌باشد. پس تصادفی بودن داده‌ها صحت دارد و درست است. حال که از تصادفی بودن داده‌ها اطمینان پیدا کردیم آزمون نرمال بودن را بر اساس آزمون‌های نیکویی برازش برای تک تک شاخص‌ها انجام می‌دهیم.

۴-۱,۱,۳ آزمون نرمال بودن داده‌ها

با توجه به اینکه می‌خواهیم از آزمون‌های کنترلی استفاده کنیم پس ابتدا باید نرمال بودن داده‌های جمع‌آوری شده را بررسی کنیم. بر این اساس با استفاده از نرم افزار مینی تب آزمون نرمال بودن داده‌ها را بر اساس آزمون کلوموگروف-اسمیرنوف^۱ می‌سنجیم. این آزمون از ساده‌ترین و مناسب‌ترین آزمون‌های ناپارامتری آزمون نیکویی برازش (میزان انطباق) برای تجزیه و تحلیل داده‌های اسمی در یک گروه می‌باشد که امتحان می‌کند آیا مشاهدات می‌تواند بطور معقولانه از توزیع معینی نشأت بگیرد یا خیر.

مفروضه های آزمون k-S عبارتند از:

✓ پیوسته بودن ابعاد و یا ستون‌های اساسی به کار رفته در آن

✓ اسمی بودن مقیاس طبقه‌بندی داده‌ها

هنگام بررسی نرمال بودن داده‌ها ما فرض صفر مبتنی بر اینکه توزیع داده‌ها نرمال است را در سطح خطای ۵٪ تست می‌کنیم. بنابراین اگر آماره آزمون بزرگتر مساوی ۰/۰۵ بدست آید، در این صورت دلیلی برای رد فرض صفر مبتنی بر اینکه داده نرمال است، وجود نخواهد داشت. به عبارت دیگر توزیع داده‌ها نرمال خواهد بود. برای آزمون نرمالیت فرض‌های آماری به صورت زیر تنظیم می‌شود. (تد هنری^۲، ۲۰۰۲)

✓ H_0 : توزیع داده‌های مربوط به هر یک از متغیرها نرمال است.

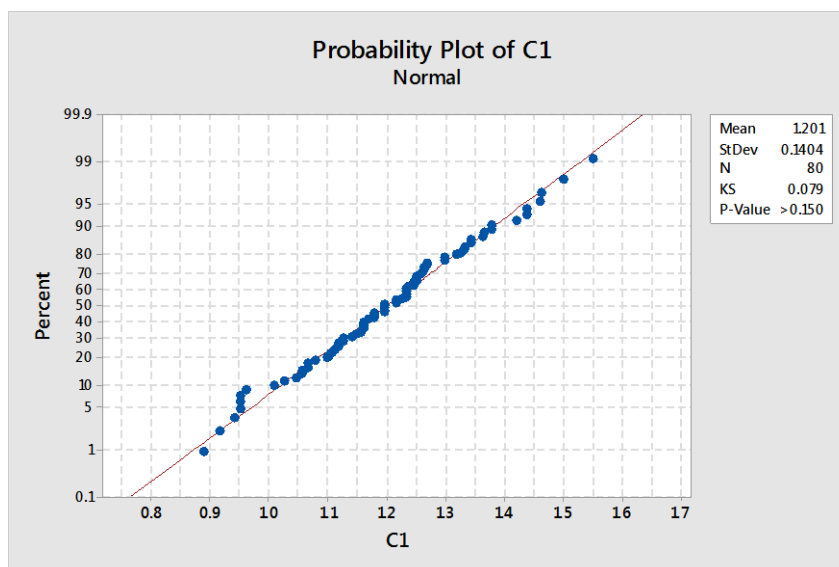
✓ H_1 : توزیع داده‌های مربوط به هر یک از متغیرها نرمال نیست.

• آزمون نرمال بودن داده‌ها برای شاخص کامل بودن

^۱ kolmogorove _ smirnov

^۲ Thode, H. C. (2002)

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۴۳

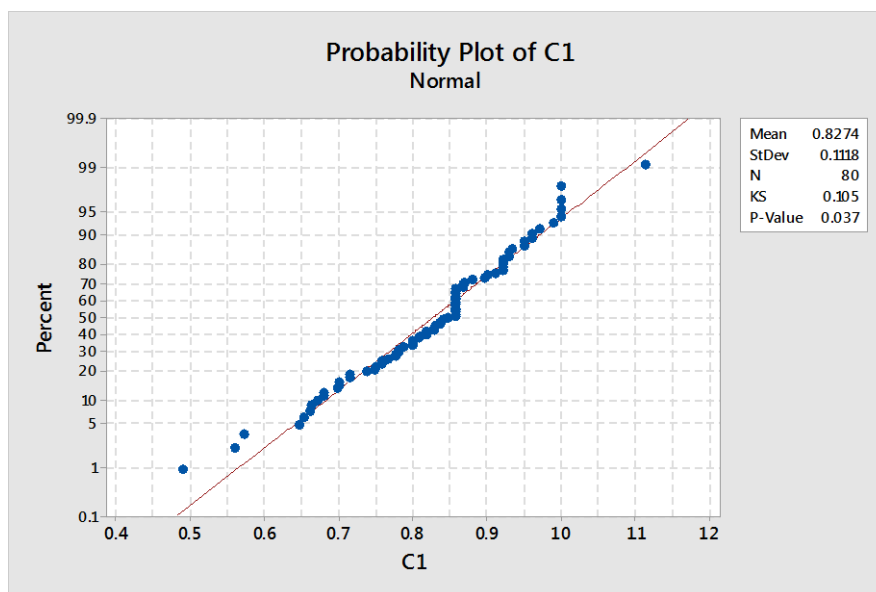


شکل (۴-۵). نرمال بودن شاخص کامل بودن

همانطور که در نمودار بالا مشاهده می‌کنیم، اکثر داده‌ها بر روی خط نرمال تجمع دارند و این امر حاکی از نرمال بودن داده‌ها می‌باشد. همچنین

مقدار P-value بزرگتر از ۰/۰۵ می‌باشد که قطعیت نرمال بودن را اثبات می‌کند. پس داده‌های شاخص کامل بودن نرمال هستند.

• آزمون نرمال بودن داده‌ها برای شاخص صحت

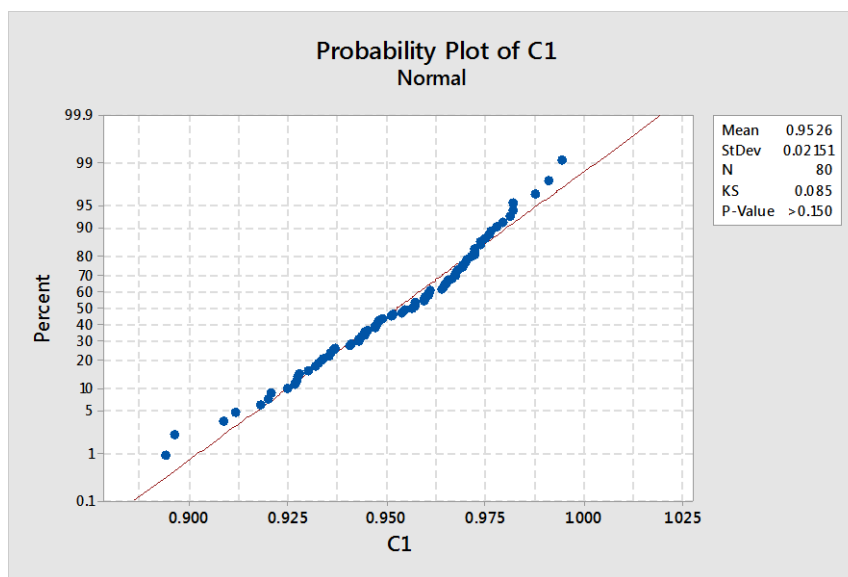


شکل (۴-۶). نرمال بودن شاخص صحت

همانطور که در نمودار بالا مشاهده می‌کنیم، اکثر داده‌ها بر روی خط نرمال تجمع دارند و تنها ۴ داده از خط نرمال کمی با فاصله قرار دارند که در محاسبه فرمول شاخص صحت به مقدار ۰/۴۸ و ۰/۵۵ و ۱/۰ و ۱/۱۱ نتیجه شده‌اند. این نقاط اگرچه داده‌های این شاخص را کمی از نرمال بودن دور کرده‌اند، اما مقدار P-value که عدد ۰/۰۳ را نشان می‌دهد حاکی از این است که با اختلاف ۰/۰۲ داده‌های این شاخص نیز به سمت نرمال بودن می‌روند.

• آزمون نرمال بودن داده‌ها برای شاخص دقت

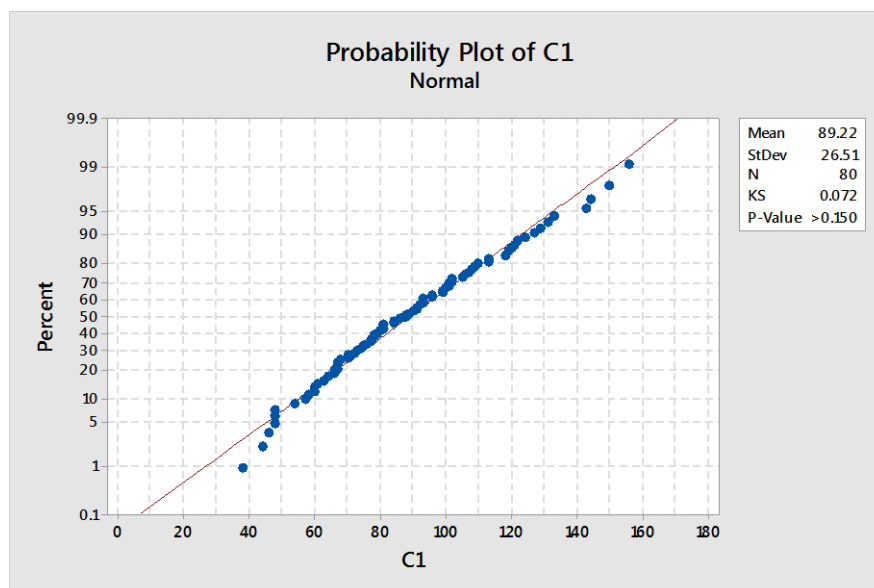
اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۴۵



شکل (۴-۷). نرمال بودن شاخص دقت

همانطور که در نمودار بالا مشاهده می‌کنیم، اکثر داده‌ها بر روی خط نرمال تجمع دارند و این امر حاکی از نرمال بودن داده‌ها می‌باشد. همچنین مقدار $P\text{-value}$ بزرگتر از ۰/۰۵ می‌باشد که قطعیت نرمال بودن را اثبات می‌کند. پس داده‌های شاخص دقت نیز نرمال هستند.

• آزمون نرمال بودن داده‌ها برای شاخص به هنگام بودن



شکل (۴-۸). نرمال بودن شاخص به هنگام بودن

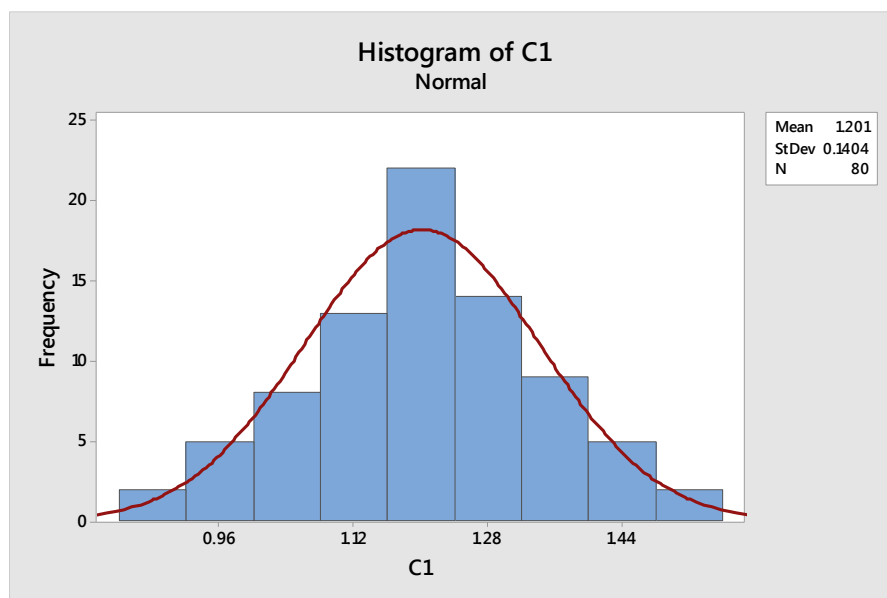
همانگونه که مشاهده کردیم p-value تمامی شاخص‌ها با توجه به آزمون کلموگروف مقداری بیشتر از ۰/۰۵ دارند که شاخص به هنگام بودن نیز از این امر تبعیت می‌کند. پس تمامی شاخص‌ها نرمال می‌باشند.

حال که از نرمال بودن شاخص‌ها اطمینان پیدا کردیم هیستوگرام داده‌ها را برای درک بهتر رسم می‌کنیم.

۴-۳،۱،۲ ترسیم هیستوگرام برای داده‌ها

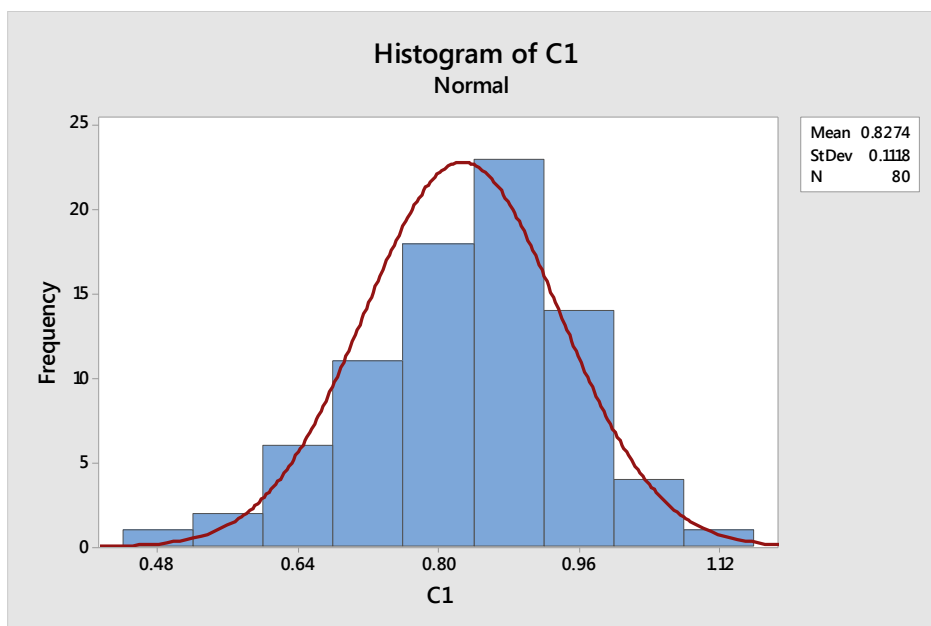
• رسم هیستوگرام شاخص کامل بودن

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۴۷



شکل (۴-۹). هیستوگرام شاخص کامل بودن

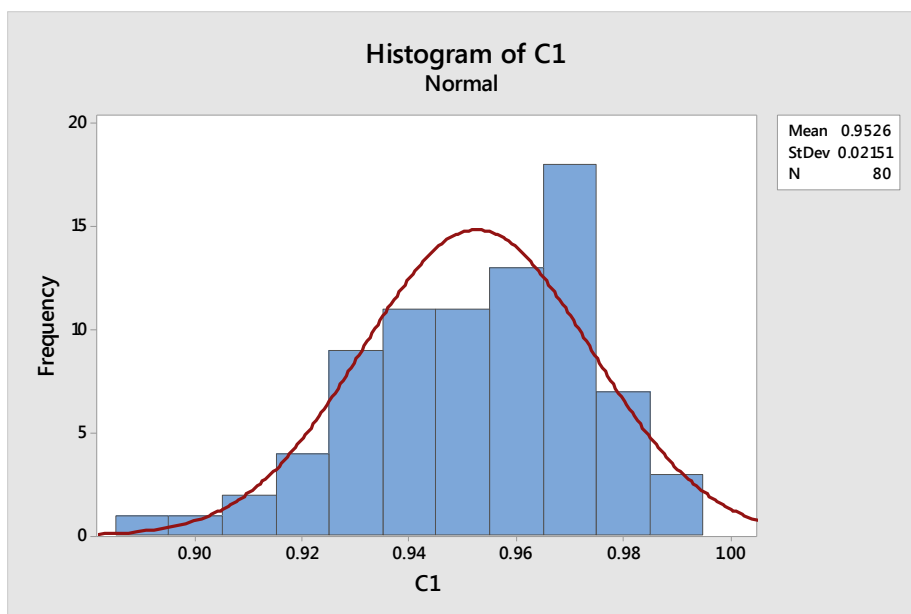
• رسم هیستوگرام شاخص صحت



اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۴۸

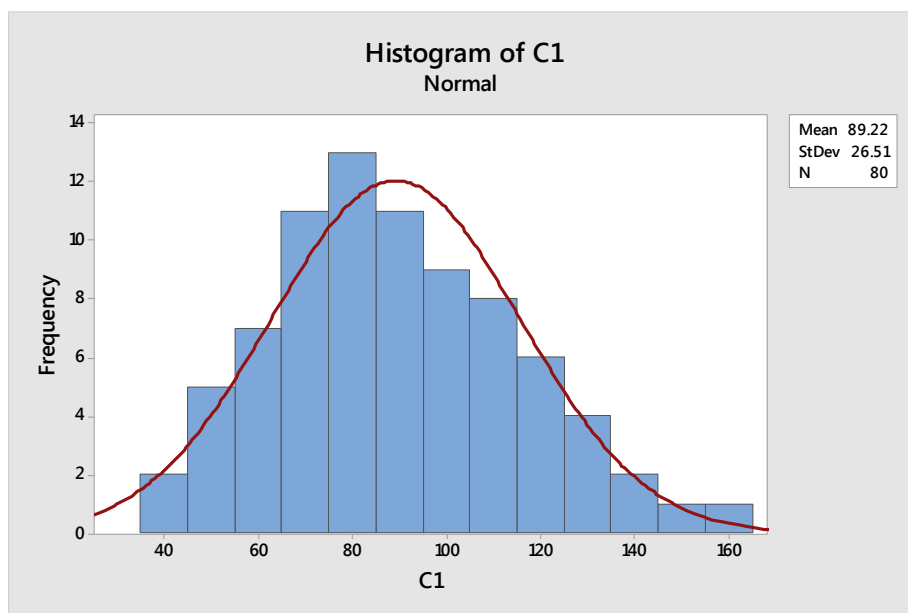
شکل (۴-۱۰). هیستوگرام شاخص صحت

• رسم هیستوگرام شاخص دقت



شکل (۴-۱۱). هیستوگرام شاخص دقت

• رسم هیستوگرام شاخص به هنگام بودن



شکل (۴-۱۲). هیستوگرام شاخص به هنگام بودن

همانطور که در هیستوگرام‌های بالا مشاهده کردیم، هیستوگرام تمامی شاخص‌ها نشان‌دهنده حالت زنگوله‌ای هستند که گویای نرمال بودن داده‌های گردآوری شده از پایان نامه دانشجویان در سیستم سامانه ثبت می‌باشد. با توجه به آزمون‌ها و نمودارهای رسم شده با در نظر داشتن تصادفی بودن و نرمال بودن داده‌های شاخص‌های تعریف شده وارد فاز دوم پروژه می‌شویم و برای تفسیر و تجزیه و تحلیل اطلاعات گردآوری شده نمودارهای کنترل را رسم می‌کنیم.

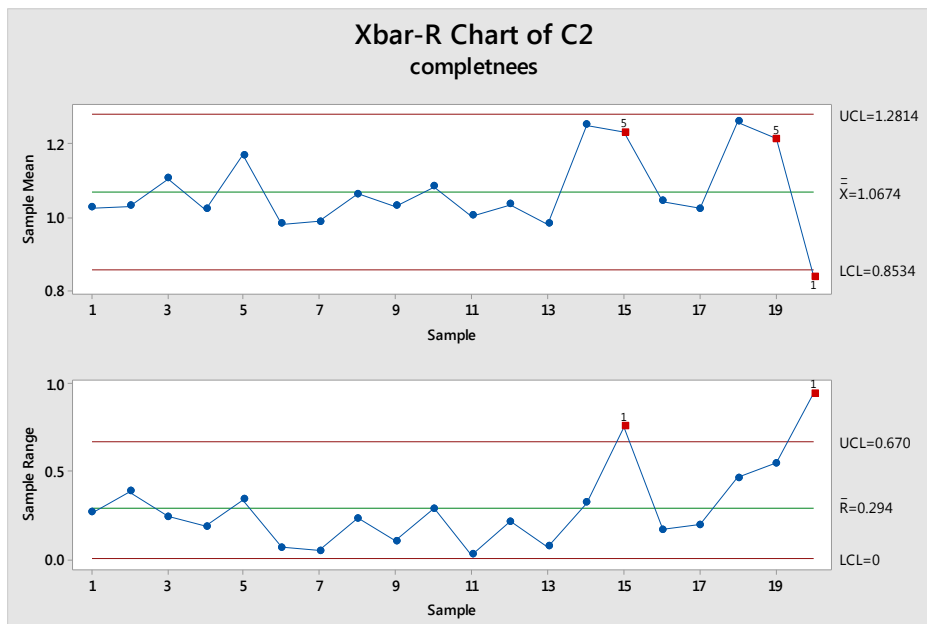
۳،۲-۴ فاز دوم

با توجه به به نمودارهای رسم شده در فاز یک در این مرحله نمودارهای کنترلی را رسم خواهیم کرد و طی مراحل با حذف نقاط خارج از حدود کنترل روند شاخص‌ها را می‌سنجیم و تغییرات را ثبت می‌کنیم. در ادامه روند تغییرات را در نمودارهای کنترل مشاهده خواهیم کرد.

۴-۱،۲،۳ نمودار میانگین-دامنه

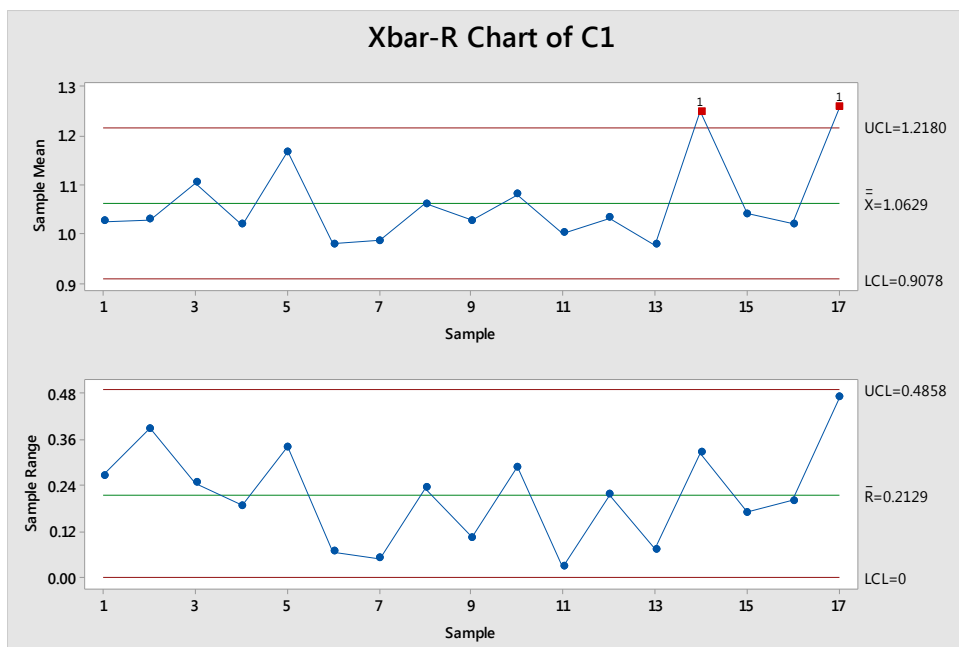
همانطور که در بخش قبل مطرح کردیم نمودار کنترل $\bar{X}$ و R نمودار کنترلی است که هم مقدار میانگین و هم دامنه تغییرات را برای یک سری داده‌های معین ارائه می‌کند. این نوع نمودار، متداولترین نوع نمودار کنترل مورد استفاده برای داده‌های پیوسته است. بدیهی است که در این نوع نمودار، $\bar{X}$ عمدتاً هر گونه تغییر در مقادیر میانگین شاخص را نشان می‌دهد و R نیز هر گونه تغییر در پراکندگی را نمایان می‌سازد. این نمودار به ویژه از این نظر حائز اهمیت است که تغییرات در مقادیر میانگین و پراکندگی را همزمان و توأم با هم نشان می‌دهد. بنابراین ما نیز برای هر شاخص نمودار $\bar{X}$ - R را رسم کرده‌ایم تا هر گونه تغییر در وضعیت میانگین و دامنه شاخص مورد نظر را در حدود کنترلی بررسی کنیم.

• شاخص کامل بودن



شکل (۴-۱۳). نمودار کنترل $\bar{X}$ - R شاخص کامل بودن

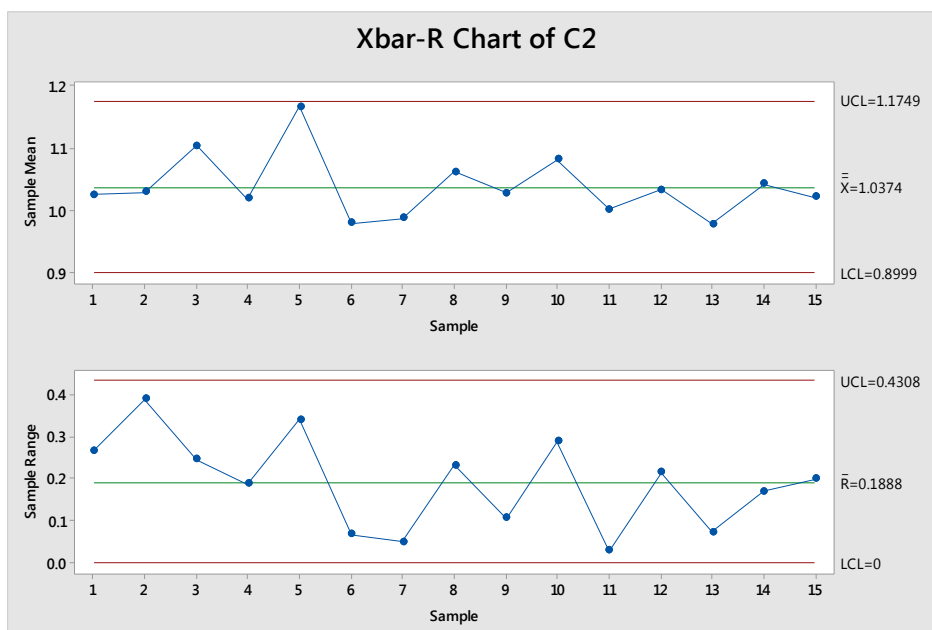
نمودار کنترلی $\bar{X}$ -R برای شاخص کامل بودن، نشان‌دهنده سه نقطه خارج از کنترل بر اساس آزمون ۵ می‌باشد. این نقاط در زیر گروه‌های ۱۵، ۱۹ و ۲۰ قرار دارند. که بر اساس نمودار دامنه هم زیر گروه ۱۵ و ۲۰ طبق آزمون ۱ جزو نقاط خارج از کنترل قرار دارند. طبق نکات ذکر شده در بخش قبلی ابتدا با حذف نقاط خارج از کنترل در دامنه وضعیت زیر گروه‌ها را بررسی می‌کنیم، سپس نقاط را در نمودار میانگین بررسی کرده و نقاط خارج از کنترل را حذف می‌کنیم. در ادامه با حذف نقاط زیر گروه ۱۵ و ۲۰ بار دیگر نمودار کنترلی $\bar{X}$ -R را برای این شاخص رسم می‌کنیم.



شکل (۴-۱۴). نمودار اصلاح شده $\bar{X}$ -R

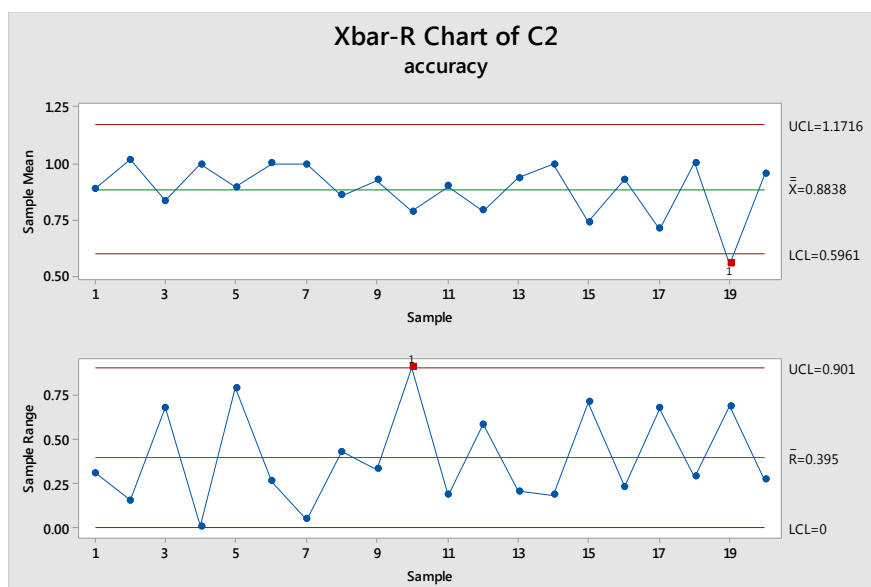
همانطور که در نمودار بالا مشاهده می‌کنیم با حذف نقاط ۲۰ نمودار دامنه تحت کنترل قرار گرفته اما با حذف نقاط زیر گروه ۱۵ باز هم نقاطی خارج از کنترل مشاهده می‌شد پس بار دیگر نقاط زیر گروه ۱۴ را حذف کردیم و تمام نقاط دامنه تحت کنترل قرار گرفته‌اند. در ادامه به حذف نقاط خارج از کنترل میانگین می‌پردازیم. نتیجه در نمودار ۴-۱۵ قابل مشاهده است.

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۵۲



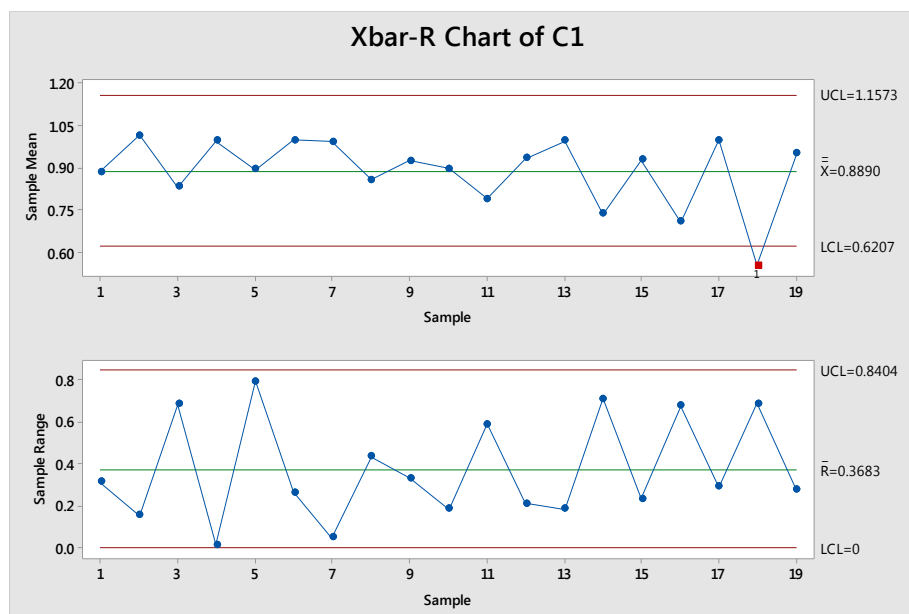
شکل (۴-۱۵). شکل نهایی $Xbar-R$

• شاخص صحت



شکل (۴-۱۶). نمودار کنترل $Xbar-R$ شاخص صحت

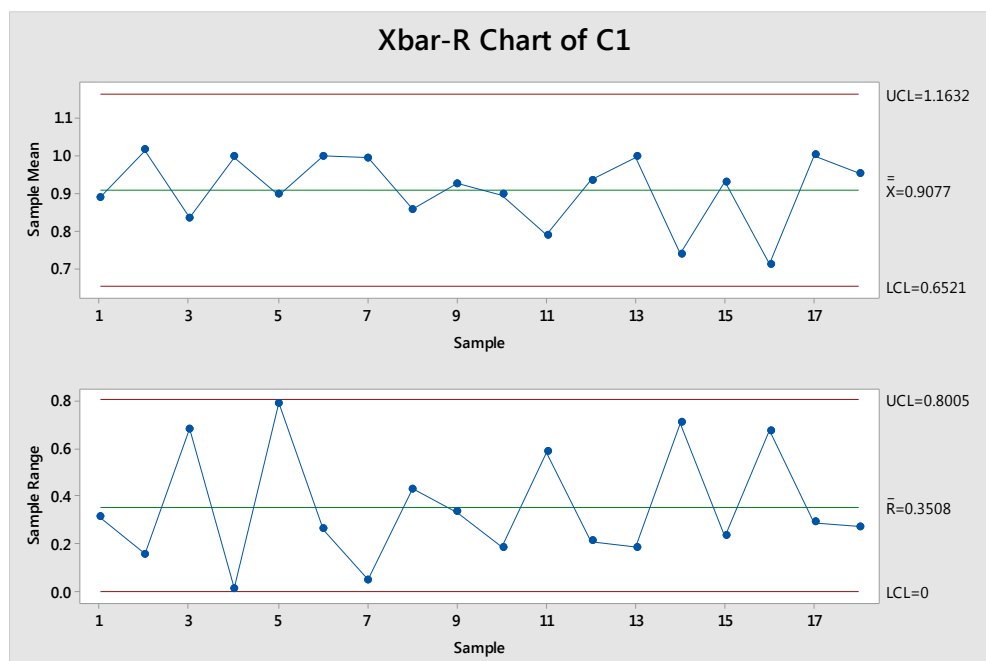
همانطور که در نمودار بالا مشاهده می‌کنیم در حدود تغییرات میانگین یک نقطه خارج از کنترل و در نمودار دامنه نیز یک نقطه خارج از کنترل بر اساس آزمون ۱ وجود دارد. پس ابتدا با حذف نقطه خارج از کنترل در نمودار دامنه بار دیگر نمودار $\bar{X}$ -R را برای شاخص صحت رسم می‌کنیم و سپس نقاط خارج از کنترل را در نمودار میانگین حذف می‌کنیم و روند تغییرات را بررسی خواهیم کرد.



شکل (۴-۱۷). نمودار اصلاح شده $\bar{X}$ -R برای شاخص صحت

در نمودار بالا با حذف نقاط زیر گروه ۱۰ نمودار دامنه کاملاً تحت کنترل قرار دارد و در ادامه با حذف نقاط زیر گروه ۱۸ روند را بررسی خواهیم کرد.

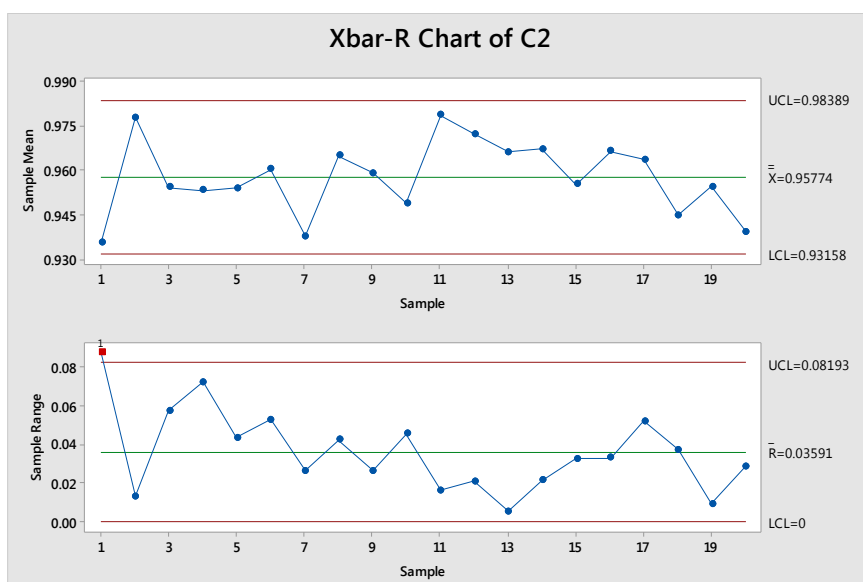
اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۵۴



شکل (۴-۱۸). شکل نهایی $Xbar-R$ برای شاخص صحت

در شکل بالا با حذف نقاط زیر گروه ۱۰ در دامنه و ۱۸ در میانگین همه نقاط در نمودار $Xbar-R$ تحت کنترل قرار گرفته‌اند.

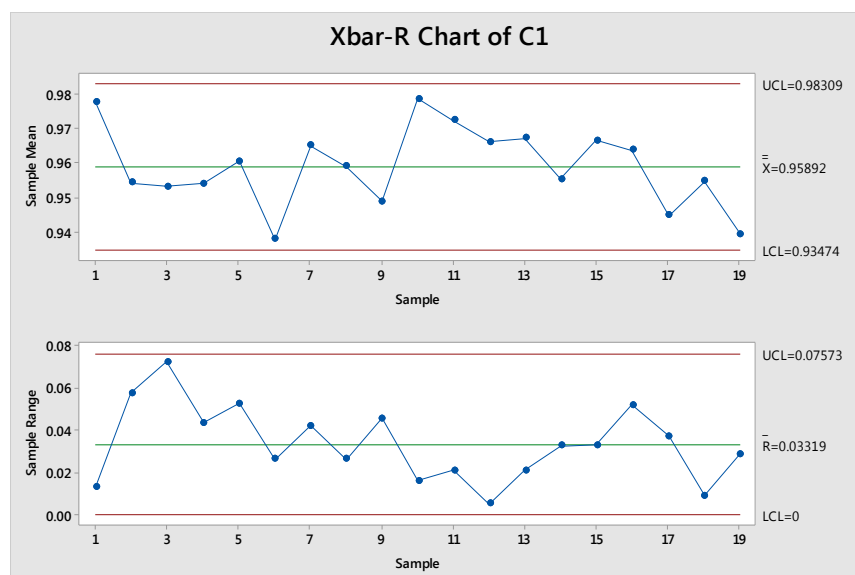
• شاخص دقت



شکل (۴-۱۹). نمودار کنترل $Xbar-R$ شاخص دقت

اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۵۵

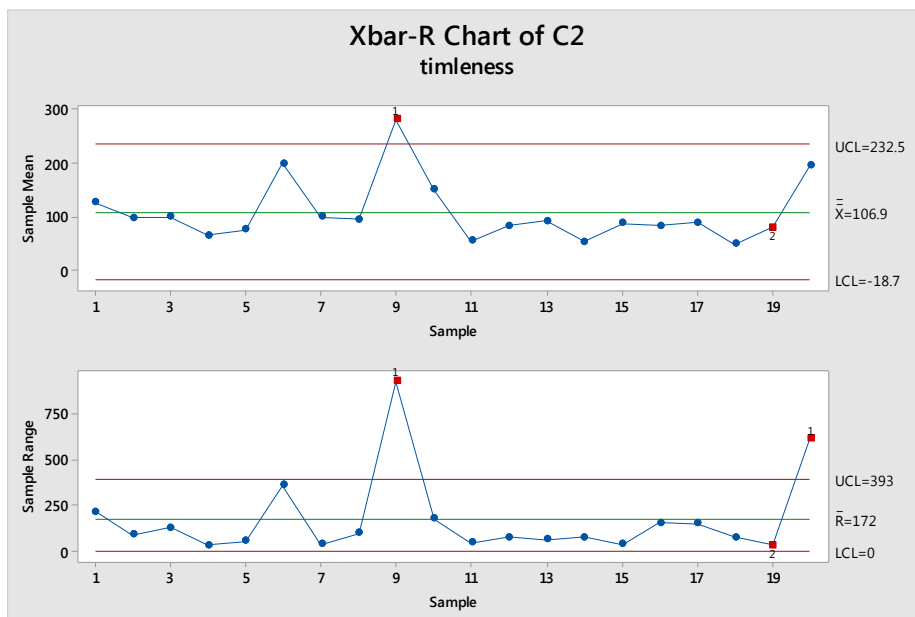
شاخص دقت تنها یک نقطه خارج از حدود کنترلی دارد که در نمودار تغییرات دامنه بر اساس آزمون فرض ۱ مشاهده می‌شود، پس با حذف این نقطه تغییرات را بررسی می‌کنیم.



شکل (۲۰-۴) شکل اصلاح شده $Xbar-R$ شاخص دقت

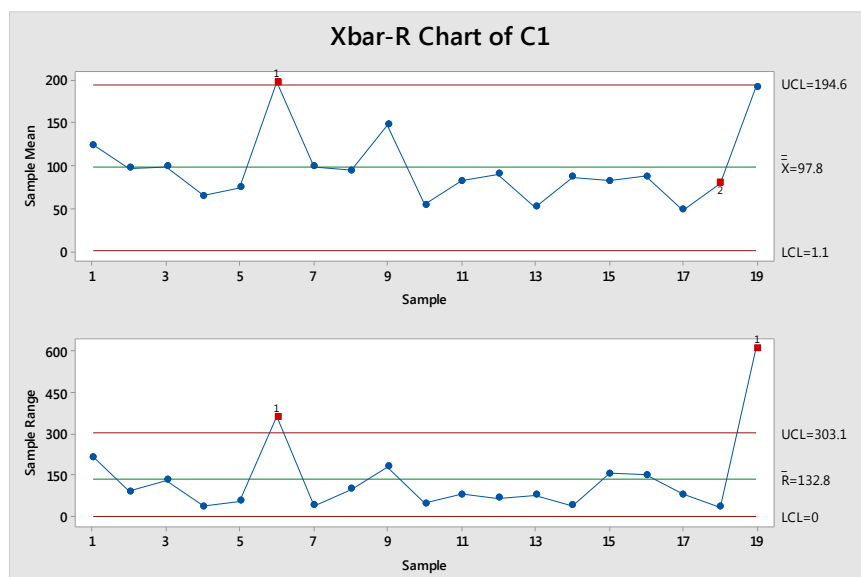
در شکل بالا با حذف نقاط زیر گروه ۱ نمودار $Xbar-R$ شاخص دقت تحت کنترل می‌باشد. پس در این شاخص نقاطی خارج از کنترل وجود ندارد.

• شاخص به هنگام بودن

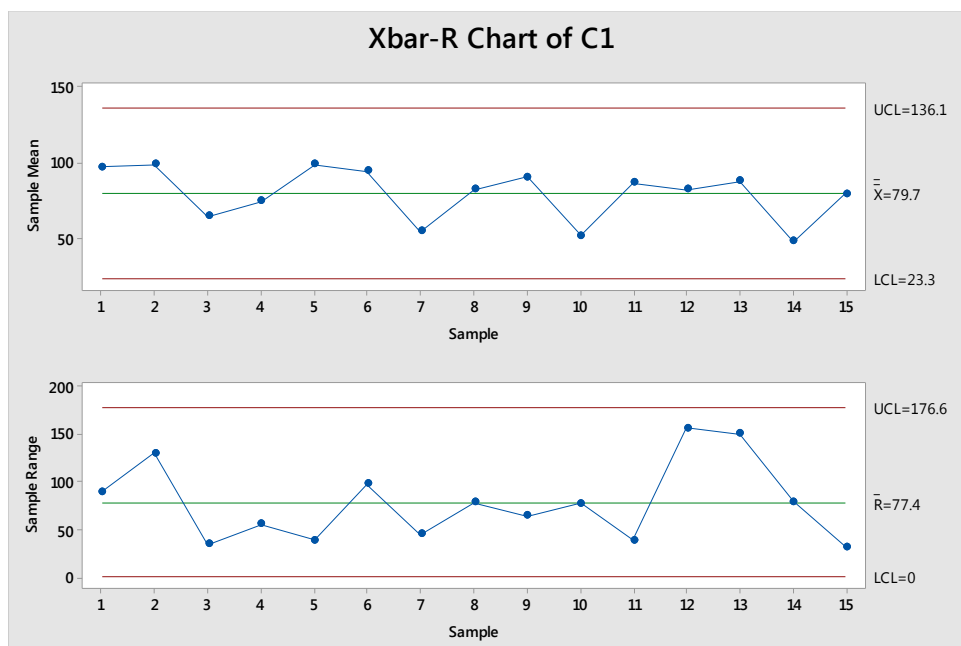


شکل (۴-۲۱). نمودار کنترل $Xbar-R$ شاخص به هنگام بودن

نمودار کنترل شاخص به هنگام بودن نشان می‌دهد که شش نقطه روی خط مرزی LCL قرار دارند که تقریباً داخل حدود کنترلی هستند اما در زیر گروه ۹ یک نقطه هم در نمودار میانگین و هم در نمودار دامنه با اختلافی زیادی خارج از حدود کنترل است که باید حذف گردد. نقطه ۱۹ نیز در دو نمودار مشترکاً خارج از کنترل می‌باشد و نقطه ۲۰ در نمودار دامنه بالاتر از خط UCL قرار دارد. در ادامه با حذف نقاط خارج از کنترل بار دیگر نمودار کنترلی را رسم می‌کنیم. نمودار ۴-۲۳ حالت نهایی را نمایش می‌دهد.



شکل (۴-۲۲). نمودار اصلاح $Xbar-R$ شاخص به هنگام بودن

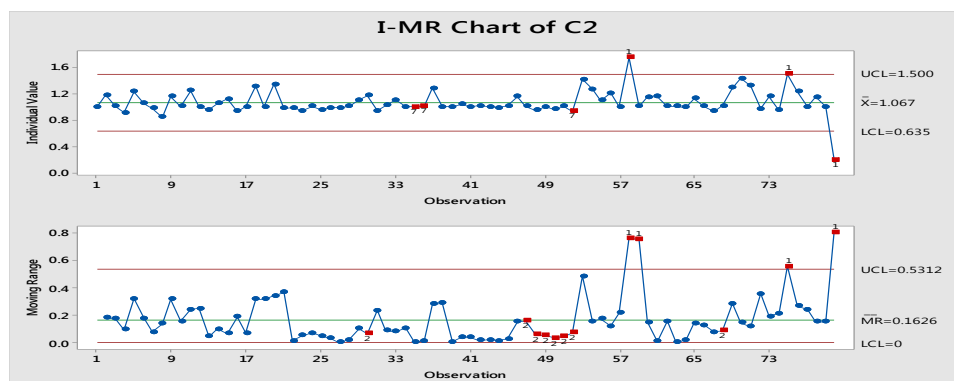


شکل (۴-۲۳). نمودار نهایی $Xbar-R$ شاخص به هنگام بودن

۴-۲،۳ نمودار I-MR

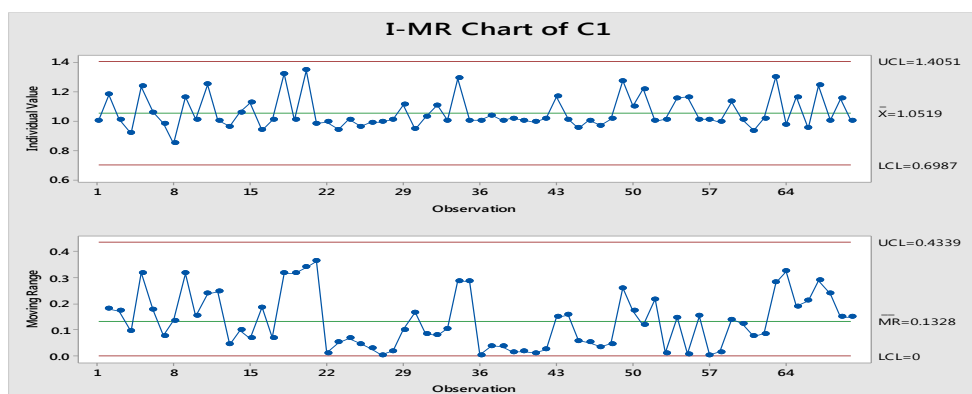
همانطور که در بخش ۳ بیان کردیم، برای اینکه دید دقیق‌تری نسبت به اطلاعات شاخص‌ها پیدا کنیم، از نمودار *I-MR* استفاده می‌کنیم. این نمودار شامل دو بخش می‌باشد در بخش بالایی (*Individual value*) نقاط به صورت تک به تک با یکدیگر مقایسه می‌شوند و بخش پایینی از نمودار (*Moving range*) بیانگر اختلاف بین دو نقطه متوالی می‌باشد که این اختلاف بیانگر فاصله منطقی میان دو نقطه است. در ادامه نمودار *I-MR* هر شاخص را رسم و بررسی می‌کنیم.

● شاخص کامل بودن



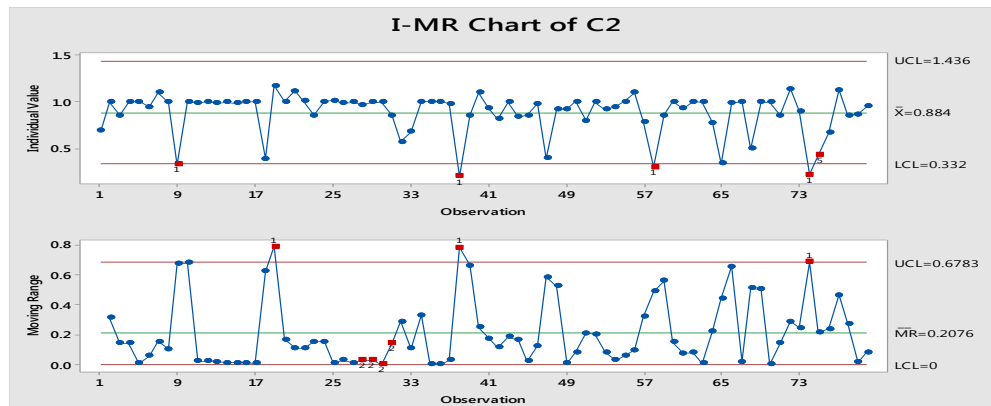
شکل (۴-۲۴). مشاهدات انفرادی شاخص کامل بودن

پس از بررسی دقیق و کارشناسی نمودار، نقاطی از زیر گروه‌ها که عامل خارج از کنترل شدن نمودار هستند، نقاط ۳۵، ۳۶، ۵۲، ۵۸، ۷۵ و ۸۰ خواهند بود، که لازم است پس از بررسی علت حذف شوند. با حذف نقاط ذکر شده نمودار اصلاح شده زیر را مشاهده می‌کنیم.



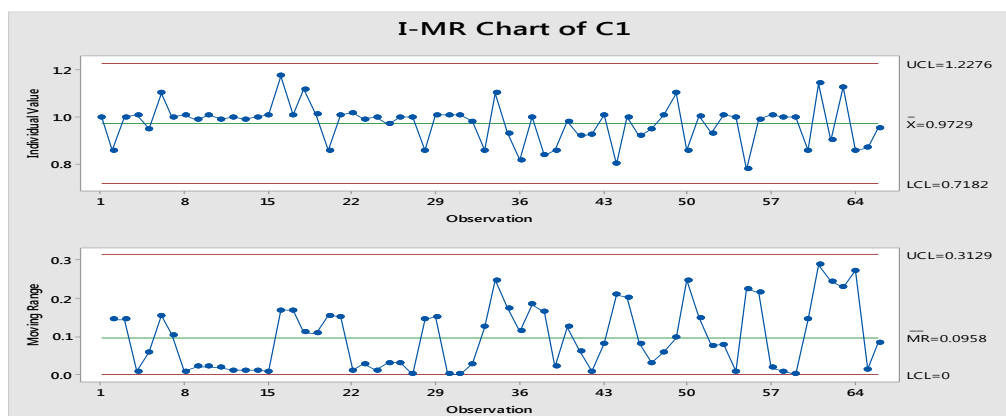
شکل (۴-۲۵). نمودار *I-MR* با حذف نقاط

● شاخص صحت



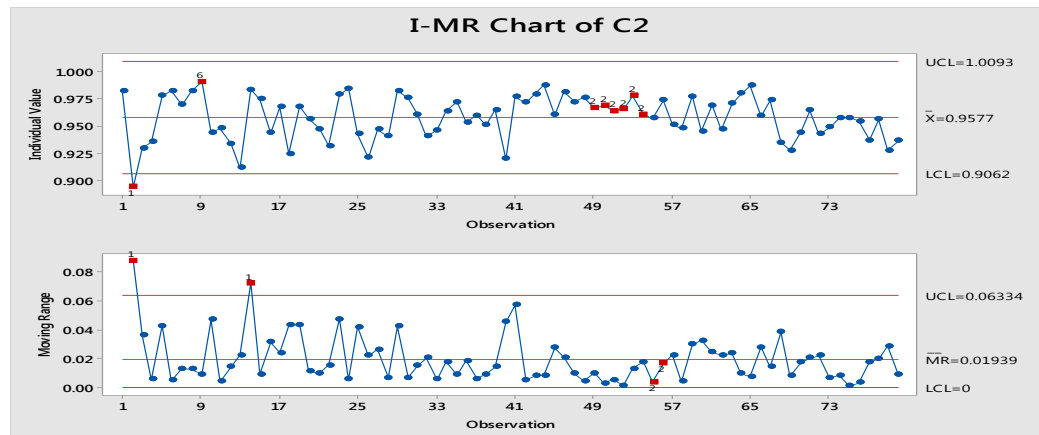
شکل (۴-۲۶). مشاهدات انفرادی شاخص صحت

همانطور که در نمودار بالا مشاهده می‌کنیم نقاط ۹، ۳۹، ۵۸، ۷۴ و ۷۵ نقاط خارج از کنترل می‌باشند. با حذف نقاط ذکر شده که حذف پایان نامه‌ها به صورت تک به تک بوده است تنها دو نقطه خارج از کنترل می‌باشند که در نمودار انفرادی خودنمایی می‌کنند. با حذف آنها نمودار انفرادی را بار دیگر رسم کرده‌ایم. در نمودار بعد می‌بینیم که با حذف نقاط ۳۰ و ۳۱ طبق آزمون ۱ نقاط خارج از حد کنترل نمودار به نقطه ۱ منتقل شدند، پس با حذف نقطه ۱، نقطه ۵۰ در نمودار محدوده حرکتی خارج از حد کنترل قرار می‌گیرد.



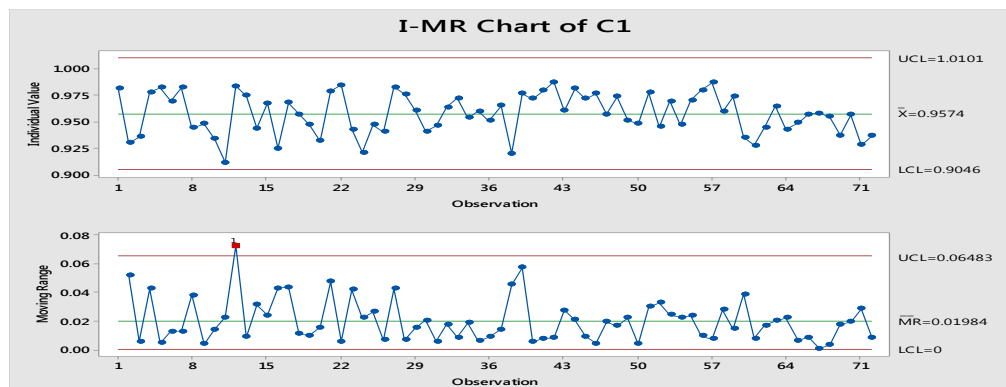
شکل (۴-۲۷). نمودار *I-MR* با حذف نقاط

● شاخص دقت



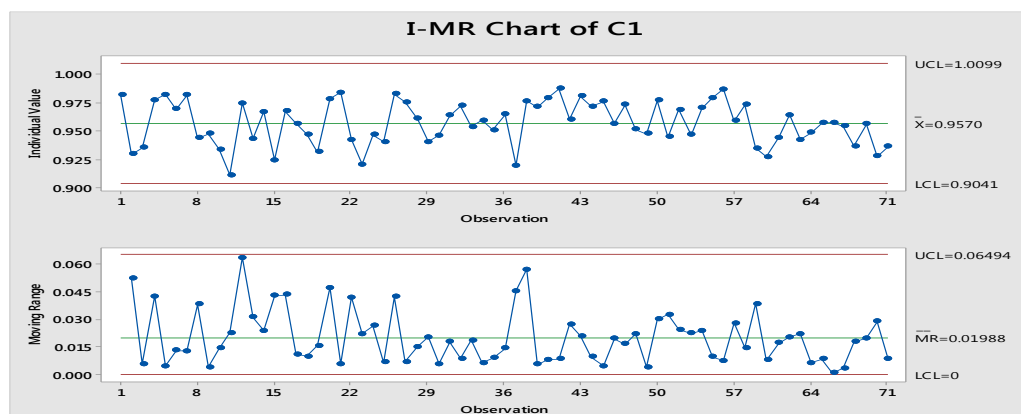
شکل (۴-۲۸). مشاهدات انفرادی شاخص دقت

نمودار بالا نشان دهنده ۹ نقطه خارج از کنترل می‌باشد. بنابراین با حذف آن نمودار را با دیگر رسم می‌کنیم. با حذف نقاط ذکر شده از هر زیر گروه بهتر توانستیم تاثیر هر پایان نامه را در هر زیر گروه بر نمودار مشاهده کنیم. حال میتوان گفت که کدام یک از پایان‌نامه ها در هر زیر گروه وضعیتی خارج از کنترل داشتند که بر کل زیر گروه تاثیر داشته است. طبق نمودار پایان نامه دوم از زیر گروه یک در شاخص دقت طبق آزمون ۱ یک نقطه پایین تر از حدود سه سیگما قرار دارد که نشان دهنده شاخص دقت پایینی است.



شکل (۴-۲۹). نمودار *I-MR* با حذف نقطه برای دقت

نقاط ۴۷ الی ۵۲ که در زیر گروه ۱۲ و ۱۳ قرار دارند. طبق آزمون ۲ که بیانگر نه نقطه متوالی در یک طرف حد کنترل قرار گرفته‌اند آزمون را رد می‌کند. با حذف آن‌ها نمودار را بررسی می‌کنیم.



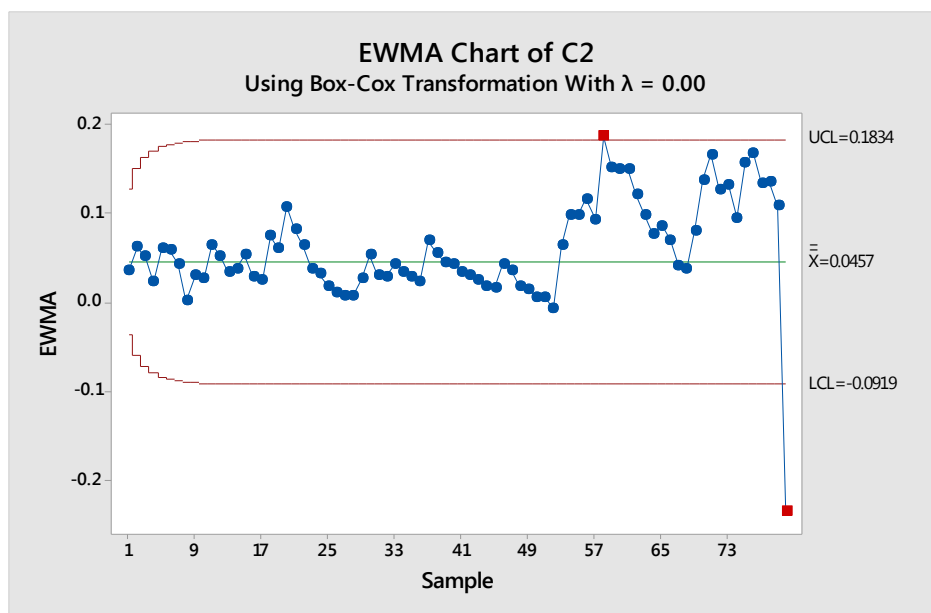
شکل (۴-۳۰). نمودار $I-MR$ با حذف نقاط

در انتهای مراحل مختلف پیاده‌سازی نمودارهای کنترل و همراه با حذف نقاط خارج از کنترل، زمینه 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.

۴-۲،۳-۳ نمودار EWMA

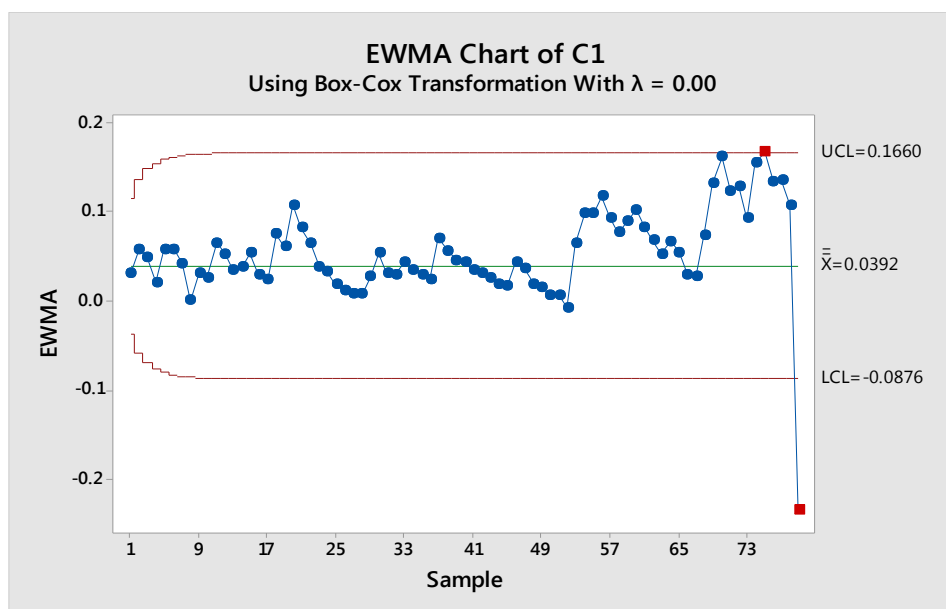
از آنجا که نمودارهای کنترل تجمعی مانند $EWMA$ و $CUSUM$ تغییر وضعیت نقاط را پیش از نمودار $Xbar-R$ تشخیص داده و به اصطلاح صنعت تولید نقطه معیوب را بسیار سریع‌تر از نمودار کنترل شوهارت نشان می‌دهند بر آن شدیم که با قرار دادن اطلاعات پایان‌نامه‌ها در نمودار کنترلی $EWMA$ داده‌ها را از نظر کیفی دقیق‌تر و جزئی‌تر بررسی کنیم، تا پایان‌نامه‌هایی که از نظر کلیه شاخص‌ها مورد تایید نمی‌باشند بهتر شناسایی شوند و منبع ایرادات در سامانه مشخص شود. طبق تعریفاتی که در فصل قبل ارائه دادیم نمودار $EWMA$ را میتوان نمودار جایگزین مناسبی برای نمودار $XBar-R$ دانست. در ادامه نمودار $EWMA$ را برای شاخص‌های مختلف کیفیت داده رسم خواهیم کرد.

- شاخص کامل بودن



شکل (۴-۳۱). نمودار $EWMA$ شاخص کامل بودن

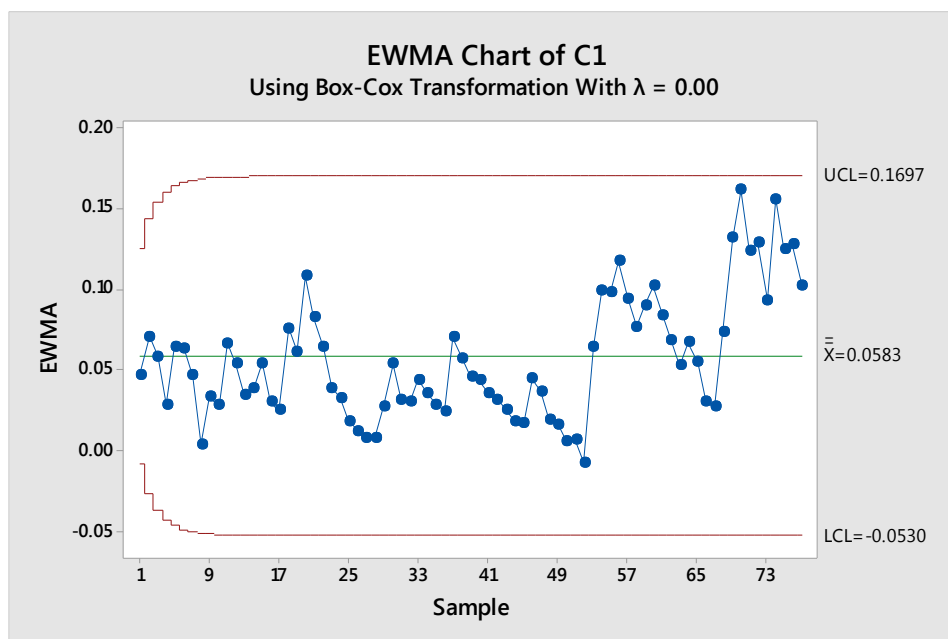
نقطه ای که در نمودار بالا اول از همه از نمودار کنترل خارج شده نقطه ۵۸ می‌باشد که با حذف آن روند تغییرات را بررسی می‌کنیم.



شکل (۴-۳۲). نمودار $EWMA$ شاخص کامل بودن پس از حذف نقاط خارج از کنترل

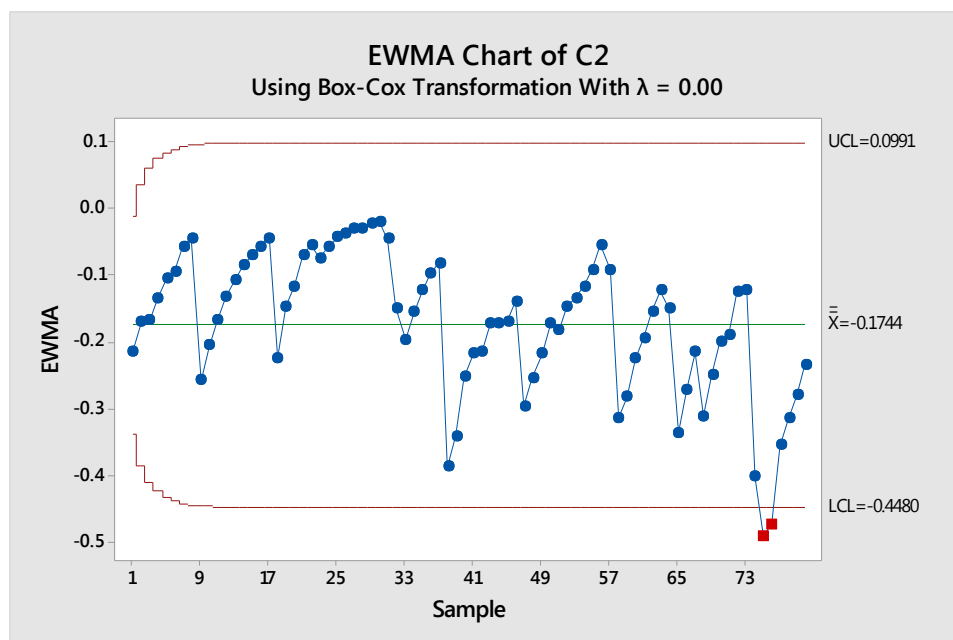
اندازه‌گیری و تحلیل شاخص‌های کیفیت داده ... □ ۱۶۳

در نمودار بالا نقطه ۷۵ و ۷۹ نقاط دیگری هستند که از حدود کنترلی خارج می‌باشند.



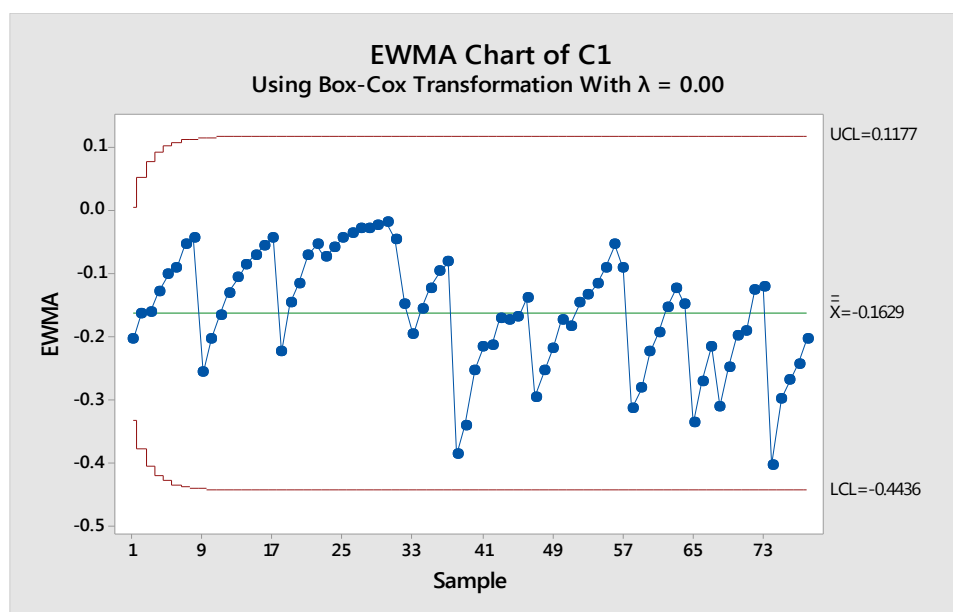
شکل (۴-۳۳). نمودار $EWMA$ شاخص کامل بودن در شکل نهایی

• شاخص صحت



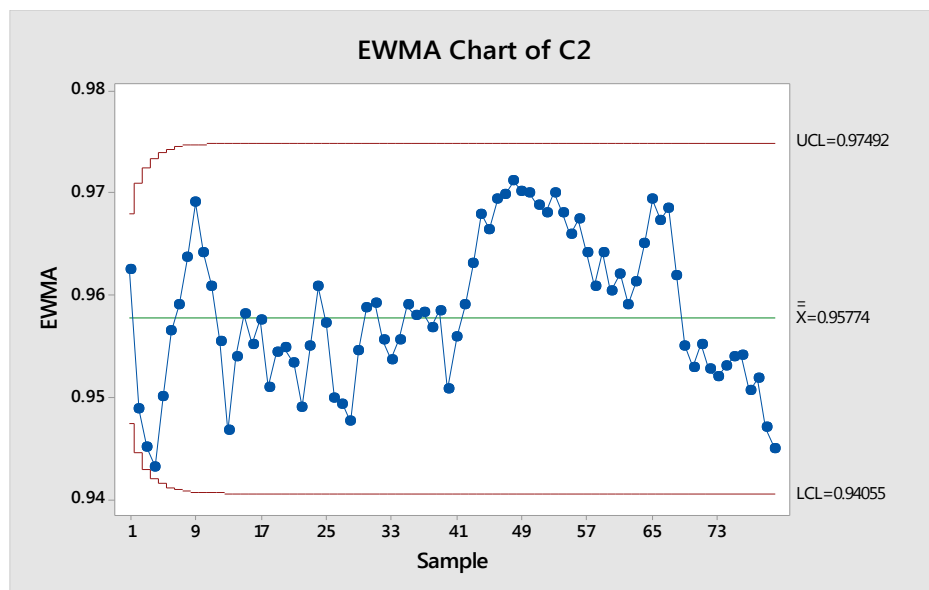
شکل (۴-۳۴). نمودار $EWMA$ شاخص صحت

نمودار بالا نشان می‌دهد اولین نقطه‌ای که از حدود کنترل خارج شده ۷۵ بوده و سپس ۷۶ که با حذف آنها مجدداً نمودار را بررسی می‌کنیم. در ادامه و در نمودار ۴-۳۵ شکل نهایی نمودار قابل مشاهده است.



شکل (۴-۳۵). نمودار $EWMA$ شاخص صحت در شکل نهایی

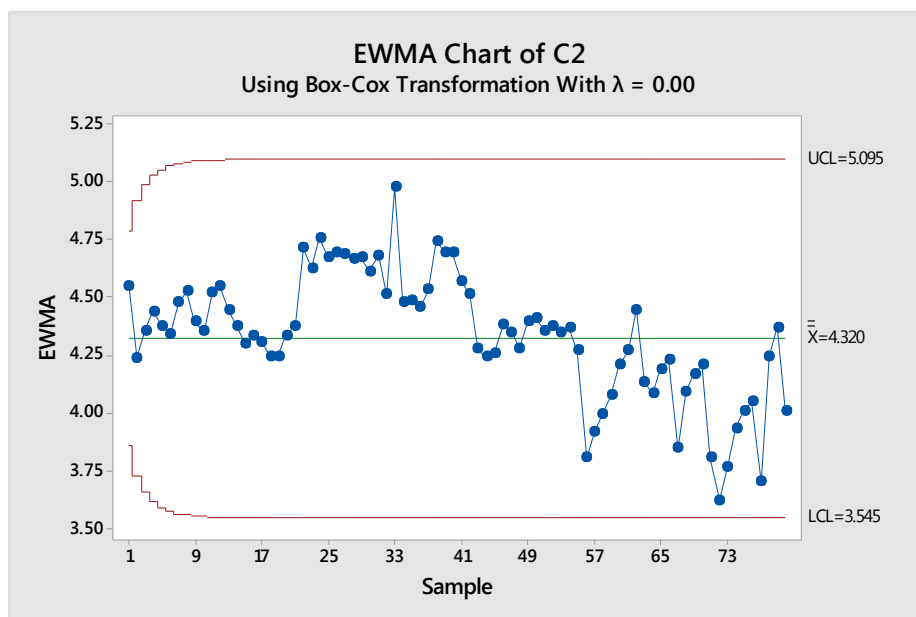
● شاخص دقت



شکل (۴-۳۶). نمودار $EWMA$ شاخص دقت

طبق نمودار بالا هیچ کدام از نقاط شاخص دقت بر اساس حدودکنترلی نمودار $EWMA$ خارج از حدود نیستند. لذا وضعیت ارائه شده در نمودار ۴-۳۷ شکل نهایی نمودار برای شاخص دقت است.

● شاخص به هنگام بودن



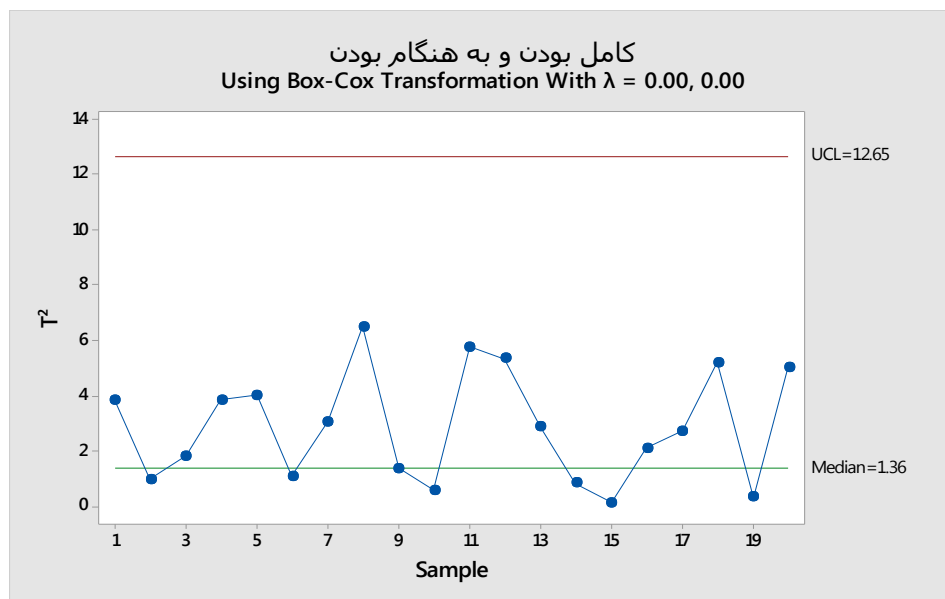
شکل (۴-۳۷). نمودار $EWMA$ شاخص به هنگام بودن

طبق نمودار $EWMA$ شاخص به هنگام بودن، نقطه‌ای خارج از کنترل وجود ندارد به همین دلیل نمودار ۴-۳۷ شکل نهایی نمودار $EWMA$ خواهد بود.

۴-۲، ۴-۳، ۴-۴ نمودار کنترل چند متغیره T^2

با شناخت نقاط مورد هر شاخص و رسم نمودار کنترل تک برای هر شاخص، اگر بخواهیم تاثیرات هر شاخص بر دیگری را مورد بررسی قرار دهیم ضروری است از نمودار کنترل چند متغیره T^2 استفاده کنیم. با رسم نمودار T^2 میتوان دریافت که هر کدام از شاخص‌های ذکر شده دقت، کامل بودن، صحت و به هنگام بودن که هر کدام برای خود شاخص‌های مستقلی هستند چه تاثیراتی برهم دارند. در ادامه برای نمونه نمودارهای چند متغیره برای اثر متقابل این شاخصها مورد بررسی قرار میگیرند.

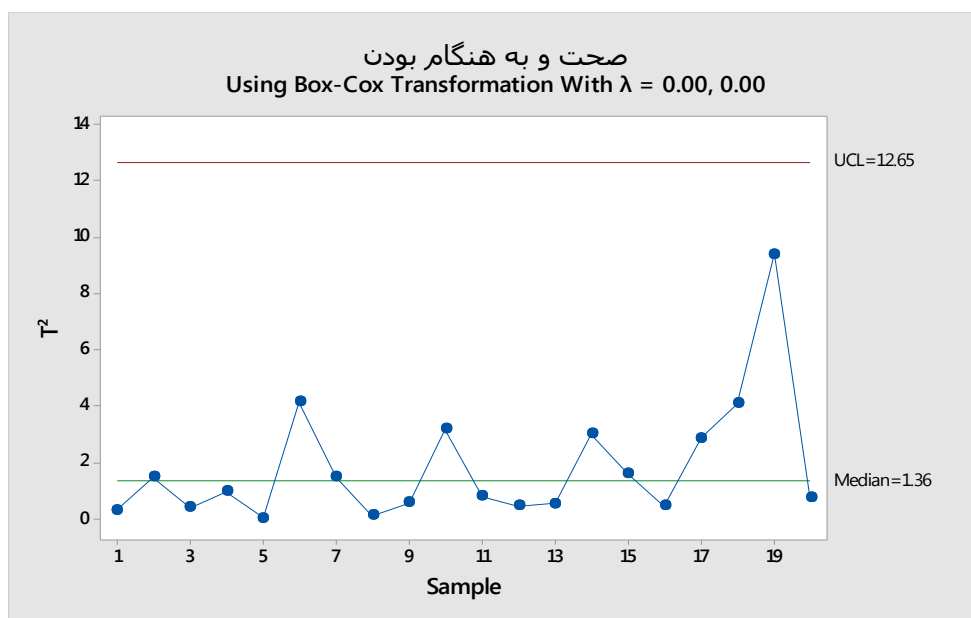
- شاخص کامل بودن و به هنگام بودن



شکل (۴-۳۸). نمودار چند متغیره T^2 شاخصهای کامل بودن و به هنگام بودن

نمودار بالا نشان می‌دهد که دو شاخص کامل بودن و به هنگام بودن طبق نمودار چند متغیره T^2 هیچ نقطه اشتراکی خارج از حد کنترل ندارند. پس میتوان نتیجه گرفت که دو شاخص در عین مستقل بودن تاثیری هم بر دیگری نخواهند داشت و نوسانات یکی بر دیگری خللی ایجاد نمی‌کند.

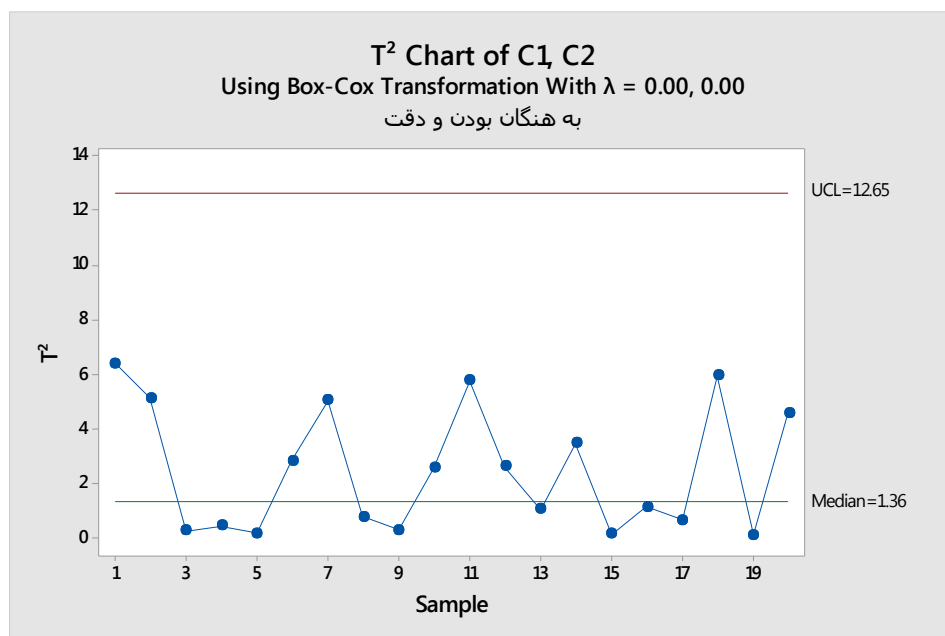
• شاخص صحت و به هنگام بودن



شکل (۴-۳۹). نمودار چند متغیره T^2 شاخص صحت و به هنگام بودن

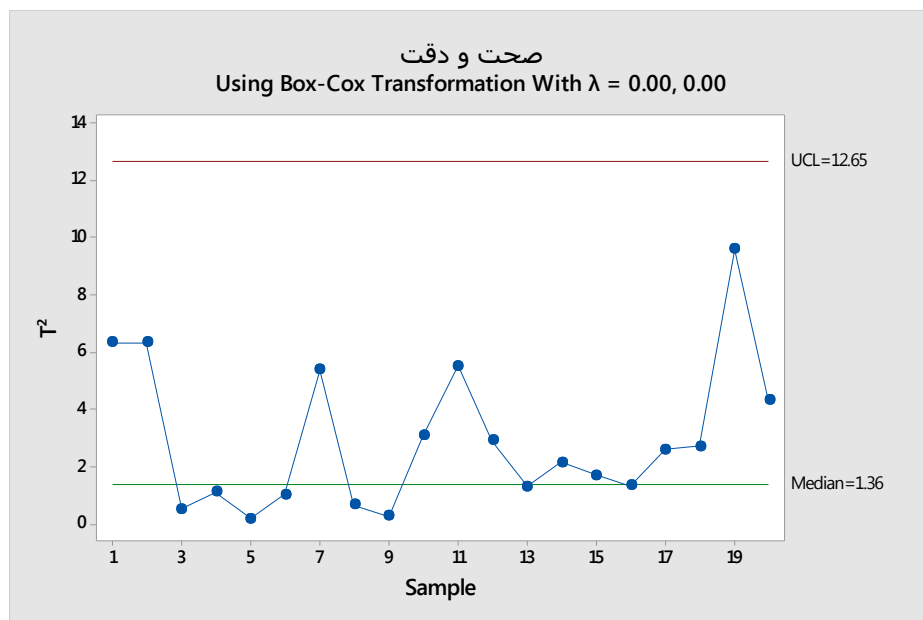
دو شاخص صحت و به هنگام بودن نیز نقطه مشترکی که خارج از حد کنترل قرار داشته باشند ندارند و تقریباً همه نقاط روی خط تعادل قرار گرفتند. تنها نقطه‌ای که نزدیک به خط UCL می‌باشد زیر گروه ۱۹ می‌باشد.

• شاخص دقت و به هنگام بودن

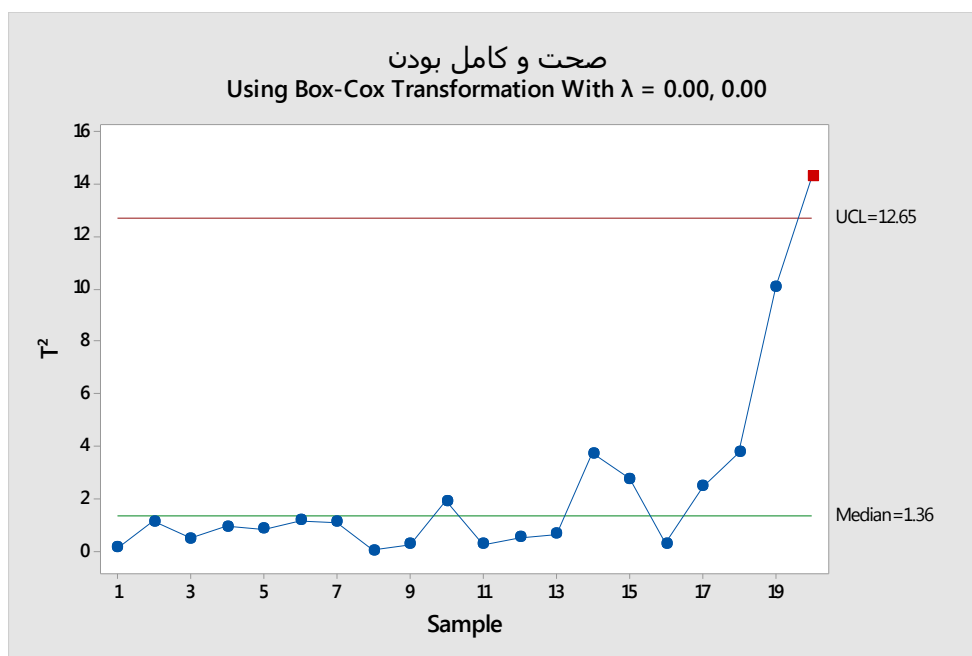


شکل (۴-۴۰). نمودار چند متغیره T^2 شاخص دقت و به هنگام بودن

نمودار دو متغیره دو شاخص دقت و به هنگام بودن نیز نقطه‌ای خارج از حد کنترلی نشان نمی‌دهد و تقریباً تمام نقاط نزدیک به خط تعادلی قرار گرفته‌اند. این نمودار حاکی از این امر است که شاخص دقت و به هنگام بودن در عین مستقل بودن نیز تاثیر روند دیگری نخواهند داشت.

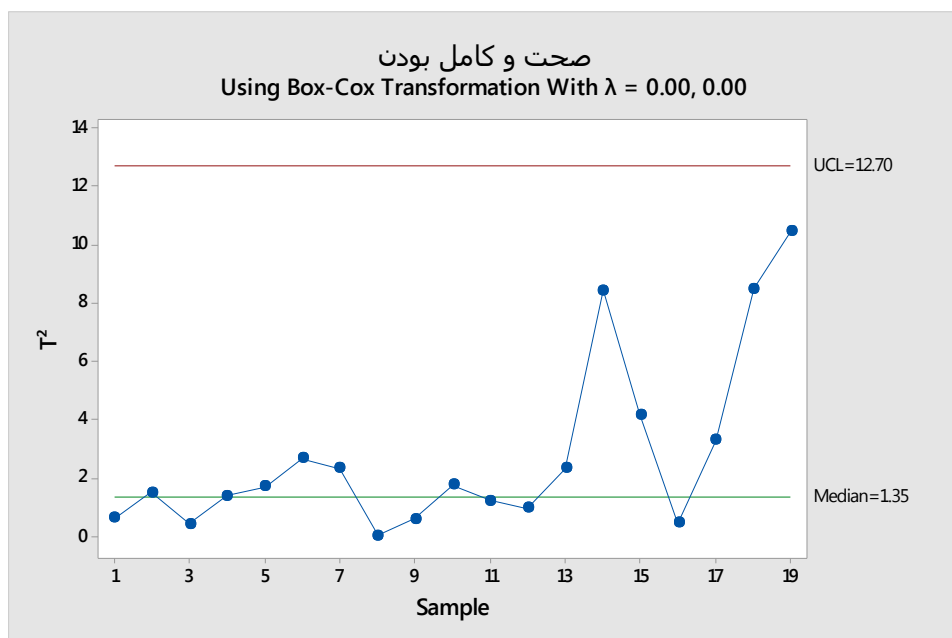


شکل (۴-۴۱). نمودار چند متغیره T^2 شاخص صحت و دقت

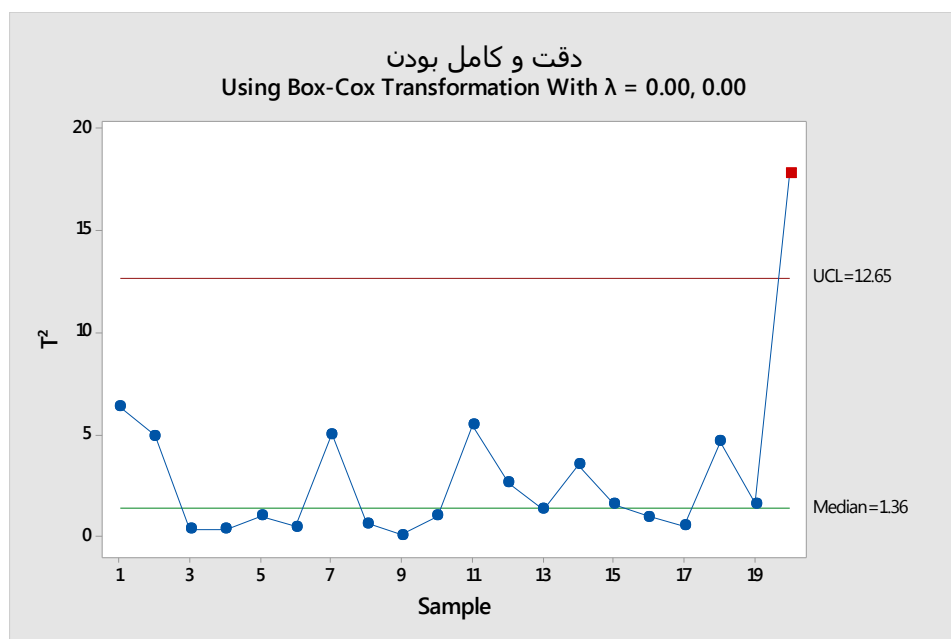


شکل (۴-۴۲). نمودار چند متغیره T^2 شاخص صحت و کامل بودن

نمودار دو متغیره دو شاخص صحت و کامل بودن، نشان‌دهنده یک نقطه خارج از حد کنترل در زیر گروه ۲۰ می‌باشد. از این حالت خارج از کنترل میتوان نتیجه گرفت که دو شاخص مذکور اگرچه مستقل هستند اما بر روی هم تاثیر گذارند و نوسانات یکی بر دیگری نیز تغییراتی را پدید می‌آورد. در هر دو شاخص زیر گروه ۲۰ که شامل پایان‌نامه‌های ۸۰-۷۶ می‌باشد فیلدهای پر نشده بسیار و عدم صحت اطلاعات وارد شده در سامانه مشاهده می‌شود، که با پیگیری و مشاهدات مجموعه پایان‌نامه‌ها متوجه شدیم که ثبت اطلاعات متعلق دانشجویان یک دانشگاه بوده است. با پیگیری‌های صورت گرفته و استعلام از مسئول دانشگاه متوجه شدیم که عدم صحت اطلاعات وارده و تکمیل نبودن فیلدها ناشی از انتقال غلط و آموزش ناکافی به دانشجویان بوده است. در نمودار زیر با حذف زیر گروه ۲۰ بار دیگر نتایج را بررسی می‌کنیم.

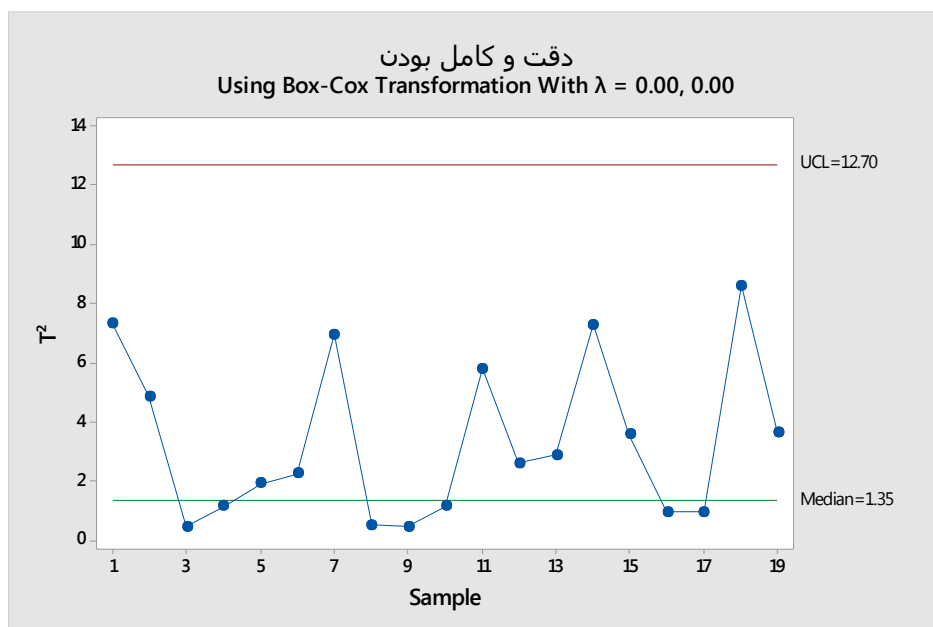


شکل (۴-۴۳). نمودار چند متغیره T^2 شاخص صحت و کامل بودن



شکل (۴-۴۴). نمودار چند متغیره T^2 شاخص دقت و کامل بودن

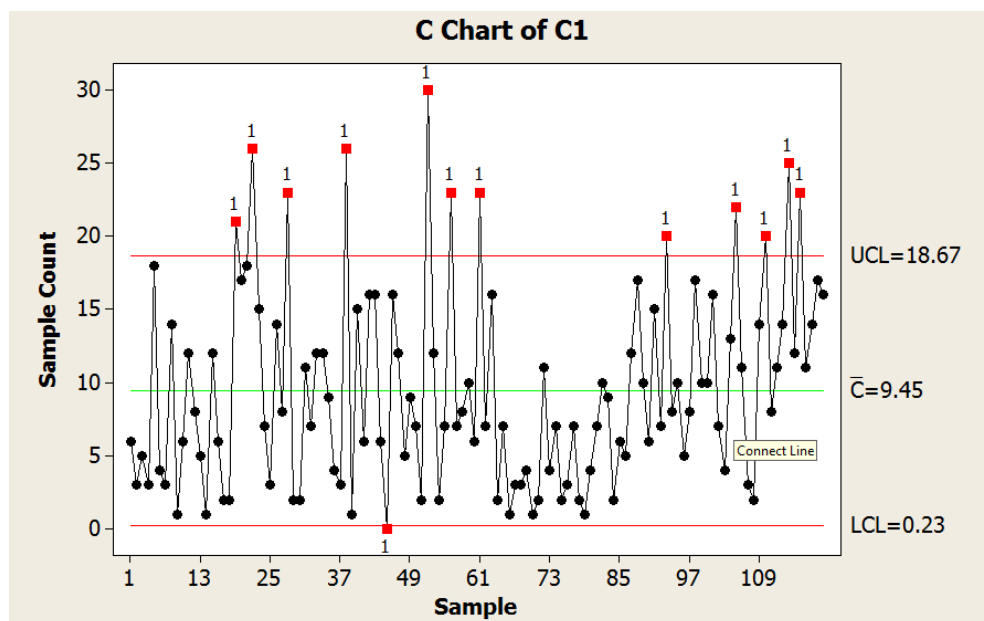
همانطور که در نمودار بالا مشاهده می‌کنیم دو شاخص دقت و کامل بودن باز هم در زیر گروه ۲۰ خارج از حد کنترل قرار دارند که متعلق به دانشگاه مذکور می‌باشد. پس میتوان نتیجه گرفت که این دو شاخص بر دیگری تاثیر دارند و نوسانات یکی بر دیگری تغییر ایجاد می‌کند. در شکل ۴-۴۵ حالت نهایی ای نمودار نشان داده شده است.



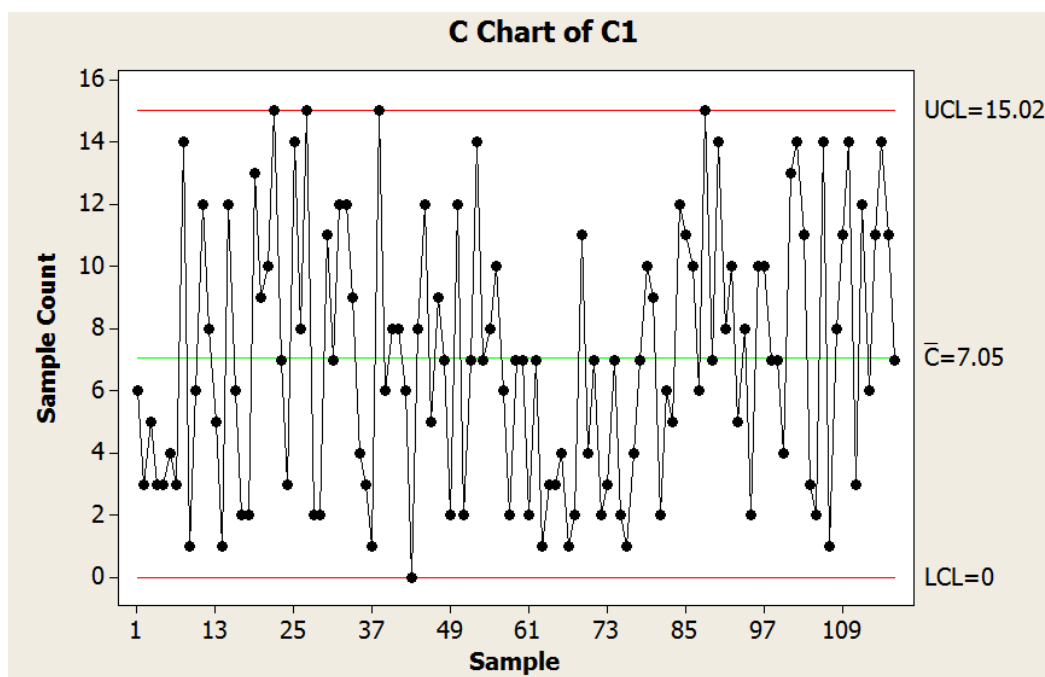
شکل (۴-۴۵). نمودار چند متغیره T^2 شاخص دقت و کامل بودن در حالت نهایی

۴-۵، ۲، ۳ نمودار کنترل C

در این بخش نمودار کنترل تعداد نقص (C) بمنظور پایش شاخص دقت مورد استفاده قرار گرفته است. در این راستا شاخص دقت با تعریف تعداد نقص در دو فیلد اصلی پایان نامه/ رساله یعنی عنوان و چکیده مورد ارزیابی قرار گرفته است. به عبارتی مجموع تعداد مشکلات ویرایشی مشاهده شده در دو فیلد اصلی پایان نامه/ رساله یعنی عنوان و چکیده بعنوان شاخص دقت در نظر گرفته شد. نتایج در نمودار ۴-۴۶ قابل مشاهده است.



شکل (۴-۴۶). نمودار اولیه تعداد نقص برای شاخص دقت

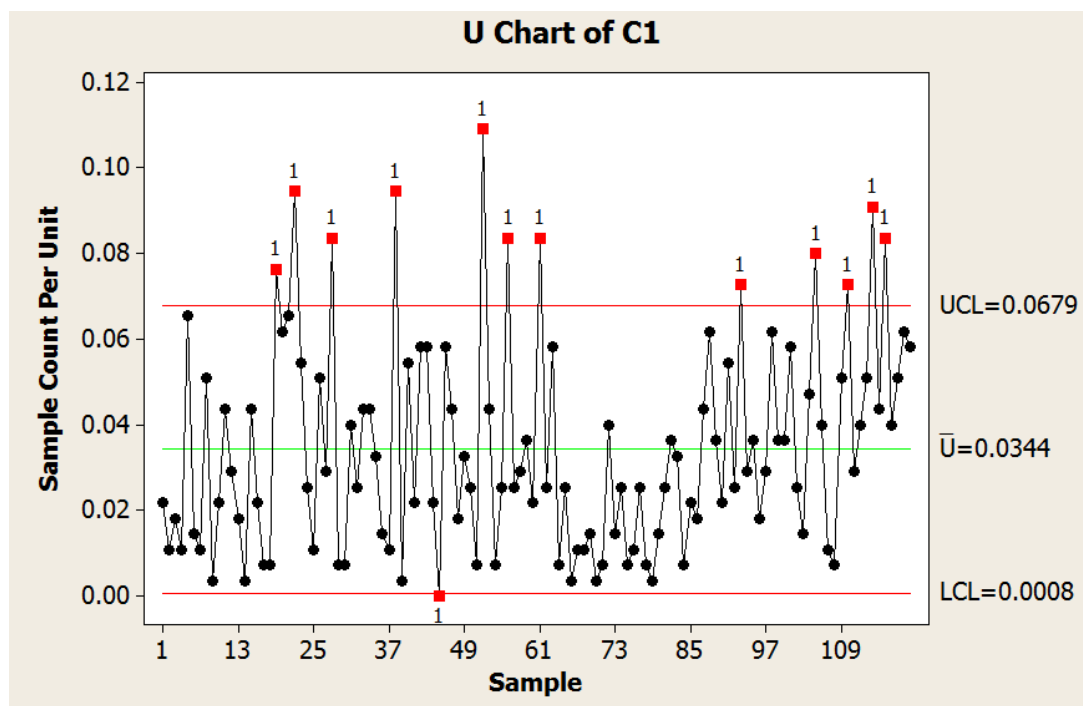


شکل (۴-۴۷). نمودار نهایی تعداد نقص برای شاخص دقت

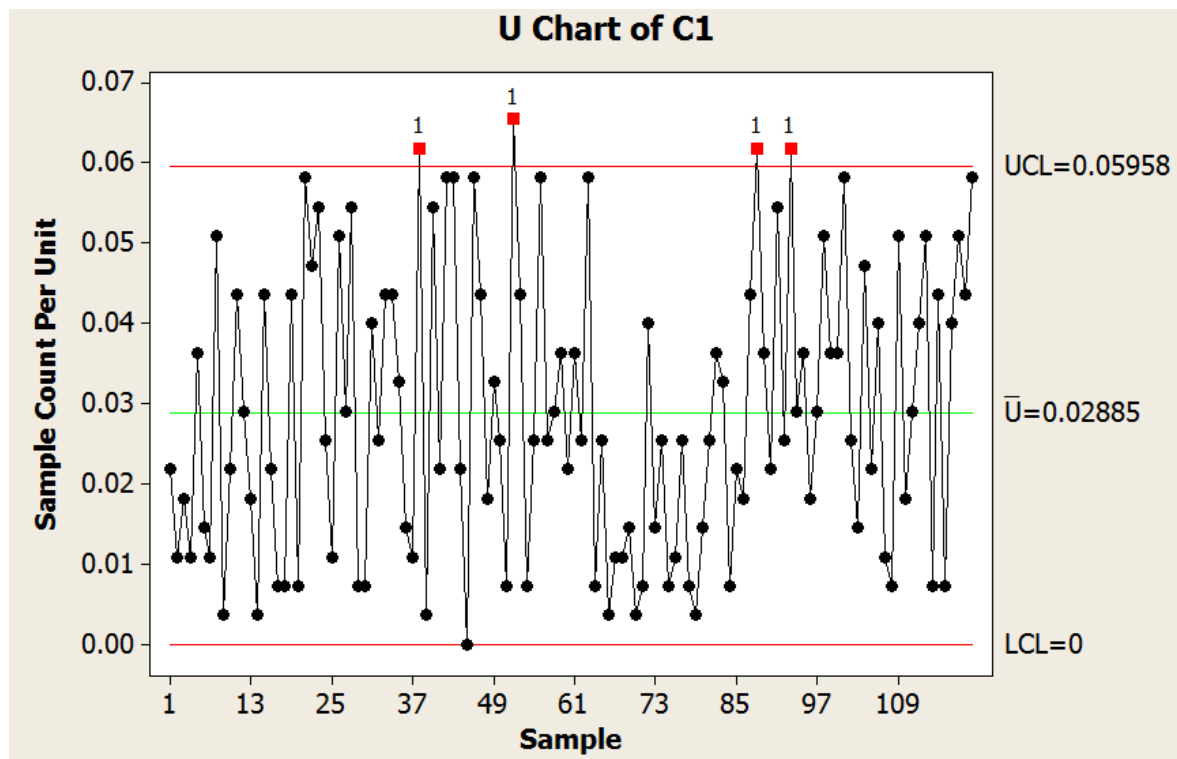
پس از ریشه‌یابی و رفع نواقص مربوط به داده‌ها و حذف نقاط خارج از کنترل نمودار اصلاح شده در شکل ۴-۴۷ قابل مشاهده خواهد بود.

۴-۳،۲،۶ نمودار کنترل U

از آنجا که نمودار تعداد نقص ارائه شده در بخش قبلی مربوط به پایان نامه‌ها و رساله‌های مختلف است و از سوی دیگر تعداد کلمات فیلدهای عنوان و چکیده مربوط به هر پایان نامه / رساله متفاوت است لذا از نمودار تعداد نقص در واحد نیز میتوان استفاده نمود. به عبارتی تعداد نقصهای مشاهده شده در فیلدهای عنوان و چکیده به متوسط تعداد کلمات (۲۷۵) تقسیم میشوند. نتایج در ادامه قابل مشاهده است.



شکل (۴-۴۸). نمودار اولیه تعداد نقص در واحد برای شاخص دقت

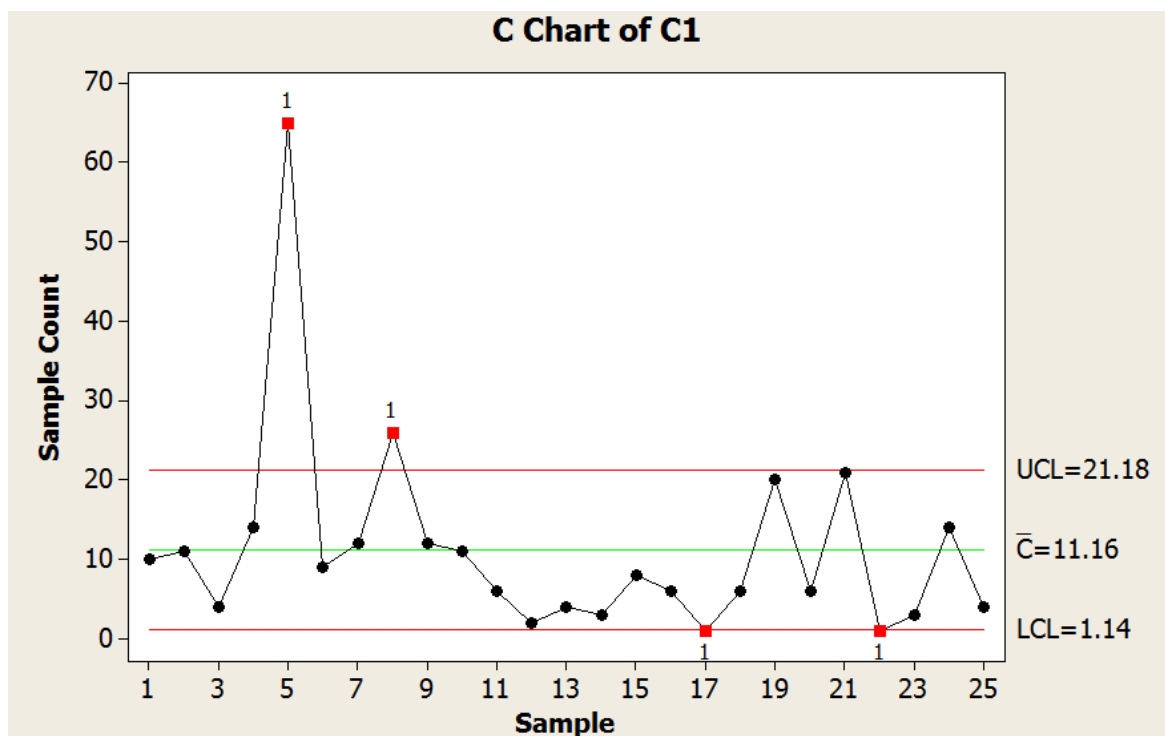


شکل (۴-۴۹). نمودار نهایی تعداد نقص در واحد برای شاخص دقت

پس از ریشه یابی و رفع نواقص مربوط به داده ها و حذف نقاط خارج از کنترل نمودار اصلاح شده در شکل ۴-۴۹ قابل مشاهده خواهد بود.

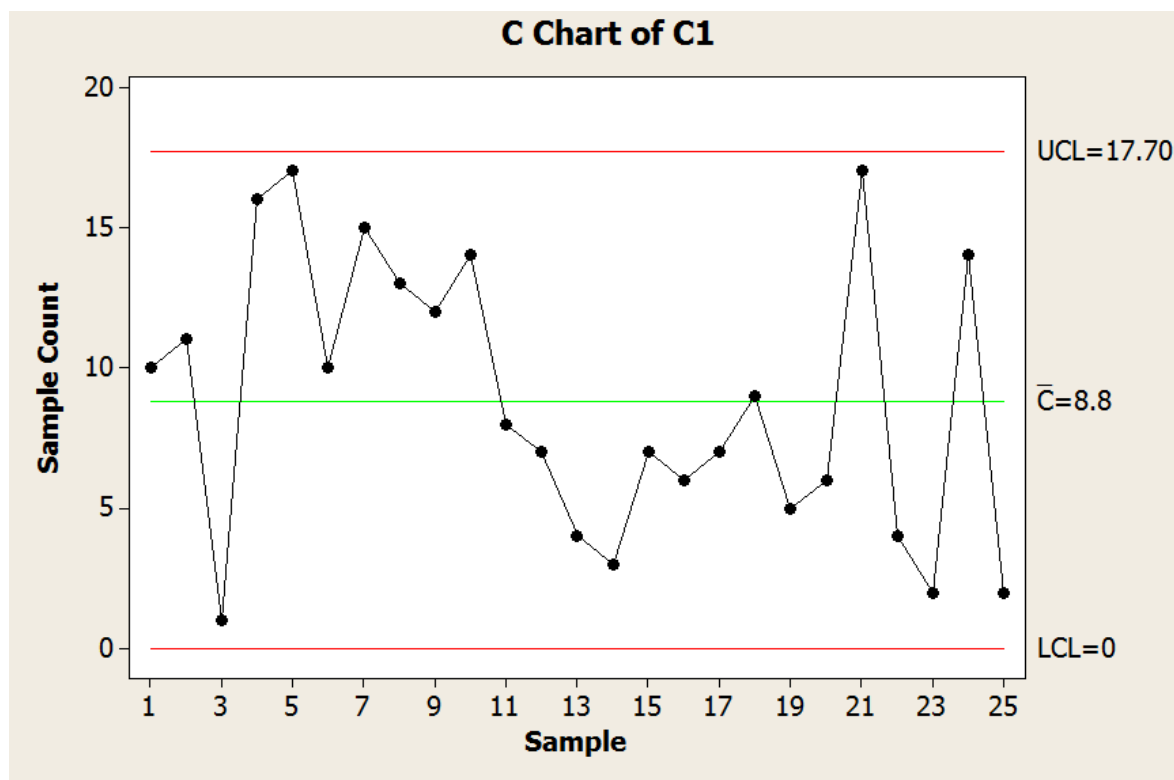
۴-۲،۷،۳ نمودار کنترل C برای سامانه ثبت قدیم

در این بخش نمودار کنترل تعداد نقص (C) بمنظور پایش شاخص دقت در سامانه ثبت قدیم مورد استفاده قرار گرفته است. در این راستا شاخص دقت با تعریف تعداد نقص در دو فیلد اصلی پایان نامه/ رساله یعنی عنوان و چکیده مورد ارزیابی قرار گرفته است. به عبارتی مجموع تعداد مشکلات ویرایشی مشاهده شده در دو فیلد اصلی پایان نامه/ رساله یعنی عنوان و چکیده بعنوان شاخص دقت در نظر گرفته شد. نتایج در نمودار ۴-۵۰ قابل مشاهده است.



شکل (۴-۵۰). نمودار اولیه تعداد نقص در سامانه ثبت قدیم برای شاخص دقت

پس از ریشه‌یابی و رفع نواقص مربوط به داده‌ها و حذف نقاط خارج از کنترل نمودار اصلاح شده در شکل ۴-۵۱ قابل مشاهده خواهد بود.



شکل (۴-۵۱). نمودار نهایی تعداد نقص در سامانه ثبت قدیم برای شاخص دقت

۴-۴ مقایسه سامانه‌های ثبت قدیم و جدید از جنبه‌های کمی و کیفی

سامانه‌های گردآوری و ثبت، سازمان‌دهی، و اشاعه اطلاعات پایان‌نامه‌ها/رساله‌های دانش‌آموختگان داخل کشور یکی از مهم‌ترین سامانه‌های اطلاعاتی تحقیقاتی هستند که رسالت اصلی آن حفظ و اشاعه پایان‌نامه‌ها و رساله‌های داخل کشور بوده و روزانه بازدیدکنندگان فراوانی دارد. از آنجاکه پایان‌نامه‌ها و رساله‌های اشاعه داده شده در سامانه گنج مبنای پژوهش خیل کثیری از پژوهشگران در مقطع کارشناسی ارشد و دکتری قرار خواهد گرفت لذا ارتقاء کیفیت این سامانه تاثیری انکارناپذیر در افزایش کیفیت پژوهش‌های صورت گرفته از سوی پژوهشگران خواهد داشت. بمنظور ارتقاء کیفیت سامانه ثبت که مقدمه و بخش بنیادین مجموعه سامانه‌های ایرانداک در حفظ و اشاعه محسوب می‌شود از سال ۱۳۹۴ پروژه بهبود این سامانه آغاز شد و از سال ۱۳۹۶ مورد استفاده دانشجویان قرار گرفت. در ادامه این گزارش به مقایسه اجمالی این دو سامانه از جنبه‌های کمی و کیفی خواهیم پرداخت.

۴-۱-۴ مقایسه کیفی سامانه‌های ثبت قدیم و جدید

بصورت تیتروار بهبودهای صورت گرفته در سامانه ثبت جدید در مقایسه با سامانه ثبت قدیم از جنبه‌های مدیریت کیفیت به شرح زیر است.

- اضافه شدن تایید دانشگاه بعد از تایید ایرانداک در فرایند ثبت پارساها بمنظور افزایش قابلیت اطمینان مدارک اشاعه داده شده و بهبود کیفیت داده‌های نهایی
- استفاده از فهرست‌های آماده کرکره‌ای بمنظور افزایش خطانپذیری سازی در ثبت داده‌ها توسط دانشجویان. بعنوان مثال برخلاف سامانه قدیم در این سامانه نام دانشگاه، نوع دانشگاه و ... از یک فهرست از پیش طراحی شده انتخاب می‌شود که این کار از یکسو امکان اشتباه در ایجاد داده را حذف می‌کند و از سوی دیگر امکان گزارش‌گیری در مدیریت علم و فناوری در ایرانداک را افزایش می‌دهد.
- امکان ارائه بازخورد مناسب به دانشجو در یک فضای متناسب که اجرای اصلاح مناسب داده و رفع اشکال را برای دانشجو فراهم می‌کند.
- اطلاع‌رسانی بکمک پیامک و ایمیل به افزایش سرعت در اجرای فرایند با حفظ کیفیت خروجی کمک شایانی خواهد کرد.
- احراز هویت دانشجو و اساتید راهنما و مشاور به کمک کد ملی فرایند ثبت را در حوزه نام و نام خانوادگی خطانپذیر خواهد کرد و کیفیت نتایج جستجو و گزارش‌گیری نهایی را افزایش خواهد داد.
- در برخی از فیلدها مانند زبان، ارائه امکان انتخاب توسط دانشجو بصورت تیک گزینه‌ها امکان خطا در فرایند را به شدت کاهش می‌دهد.
- ارائه دستورالعمل‌های مناسب در هنگام ثبت نام در سیستم ثبت جدید توسط دانشجویان به ارتقاء کیفیت خروجی کمک شایانی خواهد نمود.
- ارائه نکته‌های مختلف به دانشجو در حین تکمیل هر فیلد در قالب *tool tip* امکان بروز خطا در فرایند ثبت را به مراتب کاهش خواهد داد.
- عدم اجازه به کاربر برای رفتن به مرحله بعد یا نهایی کردن ثبت مدرک در صورتیکه فیلدهای مهم و اجباری تکمیل نشده‌اند. لازم به توضیح است که در این صورت کیفیت داده از بعد کامل بودن در خروجی نهایی می‌تواند تا حد زیادی افزایش یابد. همچنین امکان افزایش کیفیت جستجو را برای کاربران فراهم خواهد ساخت.

۴-۲-۴ مقایسه کمی سامانه‌های ثبت قدیم و جدید

از آنجا که پروژه حاضر در زمانی انجام شده است که از آغاز به کار سامانه ثبت جدید مدتی گذشته بود لذا بخش زیادی از داده‌های تحلیل شده در این گزارش (همه داده‌ها بجز داده‌های زیربخش ۳،۲،۷ از فصل ۴) به سامانه ثبت جدید مرتبط می‌شود. تنها تحلیل کمی صورت گرفته در رابطه با سامانه ثبت قدیم در زیربخش ۳،۲،۷-۴ ارائه شده است. لذا برای مقایسه کمی سامانه‌های ثبت قدیم و جدید پایه و مبنای مقایسه داده‌های این زیربخش خواهد بود. همانگونه که در زیربخش ۳،۲،۷-۴ مشاهده شد پس از به ثبات رسیدن نمودار متوسط تعداد نقصهای مشاهده شده در مجموع فیلدهای عنوان و چکیده برابر ۸،۸ است. از سوی دیگر در تحلیل ارائه شده در زیربخش ۳،۲،۵-۴ میانگین تعداد نقص در سامانه ثبت جدید در مجموع فیلدهای عنوان و چکیده برابر ۷،۰۵ است. لذا بصورت کمی میتوان عنوان کرد که بطور متوسط ۱،۷۵ در شاخص متوسط تعداد نقص در سامانه ثبت بهبود ایجاد شده است. که با تقسیم به عدد ۸،۸ معادل ۱۹،۸ درصد بهبود در این سامانه مشاهده می‌شود. بدیهی است در سایر جنبه‌ها همانگونه که در زیربخش قبلی ارائه شد بهبودهای مختلفی صورت گرفته است که بصورت کمی قابل ارائه نخواهد بود.

۴-۵ خلاصه ای از برخی نتایج کاربردی

در این پژوهش به منظور پایش کیفیت داده‌ها از نمودار کنترل استفاده شده است. نمودار کنترل به علت پویا بودن، مبتنی بر زمان بودن و پایش روند نقطه‌ها یکی از بهترین ابزارها برای پایش کیفیت در دنیا شناخته می‌شود. در این پژوهش، از نمودارهای کنترل متفاوتی بمنظور ارزیابی کیفیت داده‌های پژوهشی سامانه ثبت استفاده شده است. همچنین شاخص‌هایی همچون کامل بودن، صحت، دقت و به هنگام بودن در این فرایند مورد بررسی قرار گرفت. پیشنهادات زیر بمنظور بهبود کیفیت داده‌های پژوهشی پس از پایش مستمر فرایند ارائه گردید.

- ارائه بازخورد به موقع به دانشجویان در خصوص مشکلات کیفی داده‌های ارسال شده در سامانه گنج بمنظور رفع نواقص و ایجاد بهبود در داده‌ها
- فرهنگ‌سازی مناسب برای دانشجویان بکمک ارائه توضیحات کامل در سامانه گنج توسط کارشناسان کنترل کیفیت داده در فیلد بازخورد
- کاهش دوباره کاری در فرایند ثبت پایان نامه/رساله به کمک ارائه آموزش‌های تصویری ویا تدوین دستورالعملهای لازم
- ارائه پیغام‌های اطلاع‌رسانی (Notification) مناسب در زمان تکمیل فیلدها توسط دانشجویان تحصیلات تکمیلی

فهرست منابع

Ballou D, Wang R, Pazer H, Tayi GK. Modeling information manufacturing systems to determine information product quality. *Management Science*. 1998 Apr;44(4):462-84.

Carlo, B., Daniele, B., Federico, C., & Simone, G. (2011). A data quality methodology for heterogeneous data. *International Journal of Database Management Systems*, 3(1), 60-79.

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

Confessore, Giuseppe, Stefano Giordani, and Giuseppe Stecca. "A distributed simulation model for inventory management in a supply chain." *Processes and Foundations for Virtual Organizations*. Springer, Boston, MA, 2004. 423-430.

Data Warehousing Institute. 2006. Data quality and the bottom line: Achieving business success through a commitment to high quality data. <http://www.dw-institute.com/>

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

DQI. 2004. Data quality initiative framework. Project report. [www.wales.nhs.uk/sites/documents/319/DQI Framwork Update Letter 160604.pdf](http://www.wales.nhs.uk/sites/documents/319/DQI_Framwork_Update_Letter_160604.pdf)

Dudek-Burlikowska, M. "Quality estimation of process with usage control charts type XR and quality capability of process Cp, Cpk." *Journal of Materials Processing Technology* 162 (2005): 736-743.

Gasta, Walt. "Statistical process control using key process indicators for vacuum devices." *Vacuum Electronics Conference, 2004. IVEC 2004. Fifth IEEE International*. IEEE, 2004.

Heinrich, Bernd, and Mathias Klier. "Datenqualitätsmetriken für ein ökonomisch orientiertes Qualitätsmanagement." *Daten-und Informationsqualität*. Vieweg+ Teubner, 2011. 49-67.

Heinrich, Bernd, Mathias Klier, and Marcus Kaiser. "A procedure to develop metrics for currency and its application in CRM." *Journal of Data and Information Quality (JDIQ)* 1.1 (2009): 5.

Hoyles, Celia, et al. "Attributing meanings to representations of data: The case of statistical process control." *Mathematical Thinking and Learning* 9.4 (2007): 331-360.

Jans, Matt, Robyn Sirkis, and David Morgan. "Managing Data Quality Indicators with Paradata based Statistical Quality Control Tools: the Keys to Survey Performance." *Improving surveys with paradata* (2013): 191-229.

Jones-Farmer, L. Allison, Jeremy D. Ezell, and Benjamin T. Hazen. "Applying control chart methods to enhance data quality." *Technometrics* 56.1 (2014): 29-41.

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

Król, Dariusz, et al. "Investigation of application specific metrics to data quality assessment." *International conference on business information systems*. Springer, Berlin, Heidelberg, 2007.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387-394.

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

Moges, H. T., Van Vlasselaer, V., Lemahieu, W., & Baesens, B. (2016). Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes—An exploratory study. *Decision Support Systems*, 83, 32-46.

Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons

Morgan, Chris, and Adam Dewhurst. "Using SPC to measure a national supermarket chain's suppliers' performance." *International Journal of Operations & Production Management* 27.8 (2007): 874-900.

Mukundam, K., Varma, D. R., Deshpande, G. R., Dahanukar, V., & Roy, A. K. (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.

Nanda, V. (2016). *Quality Management System Handbook for Product Development Companies*. CRC Press. p. 352. ISBN 9781420025309.

Nelson, R. Ryan, Peter A. Todd, and Barbara H. Wixom. 2005. Antecedents of information and system quality: an empirical examination within the context of data warehousing. *Journal of management information systems*. 21(4), 199-235

Ochoa X., *Learnometrics: Metrics for Learning Objects*. PhD thesis, Katholieke Universiteit Leuven, 2008.

Ochoa, Xavier, and Erik Duval. "Quality metrics for learning object metadata." *EdMedia: World Conference on Educational Media and Technology*. Association for the Advancement of Computing in Education (AACE), 2006.

Office of Management and Budget. 2006. Information quality guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by agencies. <http://www.whitehouse.gov/omb/fedreg/reproducible.html>.

Pipino, L., Lee, Y., and Wang, R. 2002. Data quality assessment. Commun. ACM 45, 4.

Pipino, Leo L., Yang W. Lee, and Richard Y. Wang. "Data quality assessment." Communications of the ACM 45.4 (2002): 211-218.

Ghobadi, S., & Noghondarian, K. (2012). Monitoring of fuzzy quality profiles via fuzzy I-MR control chart. Arch. Sci, 65(7), 119-136.

Richard W., Stuart E. Madnick. 1990. A Polygen Model for Heterogeneous Database Systems: The Source Tagging Perspective, Proceedings of the 16th International Conference on Very Large Data Bases, 519-538.

Richard Y. Wang, Diane M. Strong. 1996. Beyond accuracy: what data quality means to data consumers, Journal of Management Information Systems, 12 (4), 5-33.

Scannapieco, M., Perinci, B., and Pierce, E. 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.

Setijono, Djoko, and Jens J. Dahlgaard. "Customer value as a key performance indicator (KPI) and a key improvement indicator (KII)." Measuring Business Excellence 11.2 (2007): 44-61.

Strong, Diane M., Yang W. Lee, and Richard Y. Wang. "Data quality in context." Communications of the ACM 40.5 (1997): 103-110.

Suhairi, Kasman, and Ford Lumban Gaol. "The measurement of optimization performance of managed service division with ITIL framework using statistical process control." Journal of Networks 8.3 (2013): 518-529.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. Communications of the ACM, 41(2), 79-82

Rigdon, S. E., Cruthis, E. N., & Champ, C. W. (1994). Design strategies for individuals and moving range control charts. Journal of Quality Technology, 26(4), 274-287.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. Communications of the ACM, 40(5), 103-110.

Thode, H. C. (2002). Testing for normality. CRC press.

Vaziri, R., Mohsenzadeh, M., & Habibi, J. (2017). Measuring data quality with weighted metrics. Total Quality Management & Business Excellence, 1-13

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, 41(2), 58-65.

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

منابع فارسی

ارشادی، محمدجواد؛ نصیری، جلال‌الدین؛ شیرانی، فرهاد. (۱۳۹۶)، طراحی مدل کیفیت فراداده در سامانه ثبت پایان‌نامه‌ها/ رساله‌های دانش‌آموختگان داخل و خارج کشور. تهران، پژوهشگاه علوم و فناوری اطلاعات ایران.

حافظی سعید و شهریاری حمید. "تفسیر نمودار کنترل چند متغیره T^2 بر اساس تجزیه زوجی آماره-23".

31.

خاکی، غلامرضا. ۱۳۸۸. روش تحقیق با رویکردی به پایان نامه نویسی. انتشارات بازتاب، فوژان. سکاران، اوما، ترجمه، صائبی، م، شیرازی، م، ۱۳۸۸. روش های تحقیق در مدیریت، چاپ هفتم انتشارات موسسه عالی آموزش و پژوهش مدیریت و برنامه ریزی، تهران.

شوندی، حسن. ۱۳۹۶. نظریه مجموعه های فازی و کاربرد آن در مهندسی صنایع و مدیریت. تهران: گسترش علوم پایه

Abstract

Today, data quality and information quality are the key indicators of performance in organizations. These categories of indicators are more crucial in organizations that are based on and built on IT strategies. Monitoring and measuring data quality indicators, on the one hand, will help identify the status of the organization's performance in the area of production and processing of data, and, on the other hand, it will help senior decision makers to determine effective measures in line with the organization's strategies. In the present research, data quality indicators have been developed in the registration system and these dissertations have been measured and analyzed in appropriate time intervals. In order to properly analyze the indicators, control charts developed on the basis of statistical quality control statistics were developed. Different aspects of data quality such as accuracy, precision, up-to-dateness and completeness were determined and measured. T^2 multivariate control charts, mean-R, number of defects, and number of defects per unit were used to analyze the data quality indices. The newly developed thesis / dissertation registration system was studied as a case study in this study. Finally, improvements have been introduced and analyzed in different aspects of quantitative aspects of this system compared to the old registration system.

Keywords: Metadata; Metadata; Data quality, Statistical Process Control; IranDoc; Research information system



Iranian Research Institute for Information Science and Technology

Research Project

**Measurement and analysis of data quality
indicators in the process of registration of
theses / dissertations of graduates in the
country**

By:

Mohammad Javad Ershadi

Co-Author:

Dr. Amir Azizi

June 2019