

طراحی ابزار و راهکار اجرایی ارزیابی و اندازه‌گیری کیفیت آمارهای ثبتی

مجری:

مریم طالبی

همکاران:

مامک ضیائیان

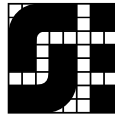
زهره طرانیان

سمانه قادری

راضیه تشخوریان

مشاور:

محمدعلی خطیبیون



پژوهشکده‌ی آمار

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

بهار ۱۳۹۲

به نام خداوند جان و خرد

پیش‌گفتار

آمار و اطلاعات به شیوه‌های گوناگونی تولید می‌شود که می‌توان به انجام طرح‌های آمارگیری (سرشماری و نمونه‌گیری) اشاره کرد. انجام طرح‌های آمارگیری علاوه بر هزینه‌بر و زمان‌بر بودن، دارای خطاهای متعددی نیز می‌باشند که شامل خطاهای نمونه‌گیری (در طرح‌های نمونه‌گیری) و غیر نمونه‌گیری (مانند خطای پوشش، خطای بی‌پاسخی، خطای پردازش و خطای اندازه‌گیری) می‌شوند علاوه بر تولید آمار به روش آمارگیری، آمارهایی نیز وجود دارند که براساس اسناد و مدارک و فرم‌هایی که همه روزه در حین انجام فعالیت‌ها و وظایف جاری هر سازمان تکمیل شده و در آن‌ها اطلاعاتی از فعالیت‌های مربوط به سازمان ثبت می‌شود، تولید می‌شوند. این دسته از آمارها، آمارهای ثبتی نامیده می‌شوند.

در این گزارش پس از معرفی آمارهای ثبتی و تبیین حوزه‌های کاری نظام آمارهای ثبتی و نیز طراحی ابزار اندازه‌گیری کیفیت، به ارزیابی کیفیت ارقام آمار ثبتی تولید شده در برخی از دستگاه‌های اجرایی با استفاده از این ابزار پرداخته شده است و در نهایت راهکارهایی برای ارتقای ابزار مذکور ارائه شده است.

پژوهشکده‌ی آمار با توجه به رسالت خود در زمینه‌ی اجرای طرح‌های پژوهشی، اجرای طرح پژوهشی «طراحی ابزار و راهکار اجرایی ارزیابی و اندازه‌گیری کیفیت آمارهای ثبتی» را در دستور کار خود قرار داد، که گزارش نهایی آن اکنون در دسترس علاقه‌مندان قرار گرفته است. این پژوهش در گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی پژوهشکده‌ی آمار با همکاری خانم مریم طالبی به عنوان مجری طرح به انجام رسیده است، که بدین وسیله از ایشان و سایر همکاران و مشاور طرح تشکر و قدردانی می‌شود.

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

پژوهشکده‌ی آمار

فهرست مطالب

۱ کلیات	۱
۱-۱- مقدمه	۱
۲-۱- مروری بر آمارهای ثبتي و کیفیت آن	۳
۳-۱- آمارهای قابل تولید به روش ثبتي	۴
۴-۱- کیفیت داده	۵
۱-۴-۱- کیفیت آمارهای ثبتي	۱۰
۲ حوزه‌های کاری نظام آمارهای ثبتي، نحوه‌ی ارزیابی هر حوزه و ابعاد کیفیت مرتبط با آنها	۱۱
۱-۲- مقدمه	۱۱
۲-۲- حوزه‌های کاری نظام آمارهای ثبتي و طراحی سئوالات چک لیست بر اساس آنها	۱۲
۳-۲- ابعاد کیفیت آمارهای ثبتي و انطباق آنها با سئوالات چک لیست	۲۳
۴-۲- طیف بندی سئوالات پرسشنامه	۲۹
۱-۴-۲- طیف لیکرت	۳۰
۵-۲- ابزار طراحی شده برای ارزیابی کیفیت آمارهای ثبتي (چک لیست) و شیوه تکمیل آن	۳۱
۱-۵-۲- محاسبه نمره مربوط به هر قلم آمار ثبتي	۳۱
۲-۵-۲- محاسبه نمره مربوط به هر حوزه‌ی نظام آمارهای ثبتي	۳۱
۳ ارزیابی کیفیت اقلام آمار ثبتي تولید شده در دستگاه‌های اجرایی با استفاده از ابزار اندازه‌گیری کیفیت (چک لیست)	۳۳
۱-۳- مقدمه	۳۳
۲-۳- به کارگیری چک لیست طراحی شده در ارزیابی کیفیت اقلام آمار ثبتي منتخب	۳۴
۳-۳- تکمیل چک لیست	۳۵
۴-۳- تحلیل چک لیست ارزیابی کیفیت «تعداد ولادت ثبت شده» و «تعداد فوت ثبت شده» و ارائه‌ی مشکلات موجود در هر حوزه	۳۵
۵-۳- تکمیل چک لیست ارزیابی کیفیت تعداد ازدواج ثبت شده	۳۶
۶-۳- تحلیل چک لیست ارزیابی کیفیت «تعداد ازدواج ثبت شده»	۳۷

۳۹	۴ ارتقای ابزار اندازه‌گیری کیفیت
۳۹	۴-۱- طراحی چک لیست ۲۹ سؤالی
۴۰	۴-۲- ارزیابی کیفیت قلم آمارثبیتی «تعداد سازمان‌های مردم نهاد» به کمک چک لیست نهایی
۴۰	۴-۳- تحلیل چک لیست ارزیابی کیفیت «تعداد سازمان‌های مردم نهاد»
۴۱	۴-۴- پیشنهادات ارائه شده برای بهبود فرایند تولید قلم ثبتی «تعداد سازمان‌های غیر دولتی»
۴۲	۴-۵- نتیجه‌گیری
۴۵	۵ پیوست‌ها
۱۱۹	۶ مرجع‌ها

۱

کلیات

۱-۱- مقدمه

امروزه همگان به نقش آمارها و داده‌ها بر تصمیم‌های اجتماعی و اقتصادی اذعان دارند. چنانچه کم دقتی یا توقف در تولید آمار به وجود آید، بر تمام جنبه‌های زندگی اجتماعی و اقتصادی تأثیر می‌گذارد و عقب ماندگی حاصل از آن چنان عمیق خواهد بود که در زمان‌های بسیار طولانی نیز به راحتی قابل جبران نیست. از این‌رو در کشورهای توسعه یافته، نظام آماری کشورها و کیفیت بانک‌های اطلاعاتی آنها، نه تنها به عنوان یکی از مهم‌ترین شاخص‌های توسعه یافتگی آنان به شمار می‌آید، بلکه زیربنای توسعه نیز قلمداد می‌شود. با توجه به موارد گفته شده مشخص است که آرایه آمارهای صحیح، با کیفیت مناسب و مبتنی بر روش‌های علمی از اهمیت بسزایی در توسعه پایدار کشور برخوردار است. آمار و اطلاعات به شیوه‌های گوناگونی تولید می‌شود که می‌توان به انجام طرح‌های آمارگیری (سرشماری و نمونه‌گیری) اشاره کرد. انجام طرح‌های آمارگیری علاوه بر هزینه‌بر و زمان‌بر بودن، دارای خطاهای متعددی نیز می‌باشند که شامل خطاهای نمونه‌گیری (در طرح‌های نمونه‌گیری) و غیر نمونه‌گیری (مانند خطای پوشش، خطای بی‌پاسخی، خطای پردازش و خطای اندازه‌گیری) می‌شوند. اما زمانی که خلاء آماری وجود دارد، اجرای طرح‌های آمارگیری اجتناب ناپذیر است.

علاوه بر تولید آمار به روش آمارگیری، آمارهایی نیز وجود دارند که براساس اسناد و مدارک و فرم‌هایی که همه روزه در حین انجام فعالیت‌ها و وظایف جاری هر سازمان تکمیل شده و در آنها اطلاعاتی از فعالیت‌های مربوط به سازمان ثبت می‌شود، تولید می‌شوند. در این روش، می‌توان داده‌ها را از طریق ثبت آنها به هنگام رخ دادن عمل، کار یا فعالیت‌های جاری، گردآوری کرد. یک عملی انجام می‌شود، داده‌ای تولید می‌شود و بلافاصله ثبت می‌شود. پس از گردآوری آنها، آمارهایی که حاصل می‌شود، آمار ثبتی نامیده می‌شود. در بیش‌تر سازمان‌های دولتی که می‌توان آنها را مهم‌ترین منبع تولید آمار هر کشور دانست، داده‌های زیادی در هنگام فعالیت‌ها و وظایف جاریشان گردآوری می‌شود.

به عنوان مثال فردی را در نظر بگیرید که برای ساخت یک واحد مسکونی به دنبال مجوز ساخت است. او به شهرداری مراجعه می‌کند و برای دریافت مجوز و پروانه ساخت، فرمی تکمیل می‌کند که حاوی اطلاعاتی در زمینه موقعیت ساختمان، نوع بهره‌برداری (مسکونی و تجاری)، مساحت کل و مساحت زیربنا، تعداد طبقات و تعداد واحد است. پس از تکمیل این فرم، داده‌های ثبتی تولید می‌شود. باید توجه داشت که هدف از تکمیل فرم‌ها جمع‌آوری آمار نیست، بلکه هدف، انجام امور مربوط به ساختمان است. اگر بتوان اطلاعات این فرم را برای کل ساختمان‌ها گردآوری کرد، از جمع‌بندی، دسته‌بندی و پالایش آن‌ها آمارهای ثبتی تولید خواهد شد و می‌توان از آن‌ها گزارش‌های آماری تهیه کرد. به عنوان مثال، تعداد واحدهای مسکونی با متر اژ ۵۰ تا ۷۵ متر مربع. اگر ثبت این‌گونه اطلاعات به‌گونه‌ای صورت گیرد که به محض انجام عملیات داده‌ها ثبت شود، انتقال و ذخیره صورت گیرد، سپس مورد پردازش قرار گرفته و آماده اطلاع رسانی شوند و این کار به صورت به‌هنگام، مکانیزه و منظم صورت گیرد، آمار تولید شده دقت بالایی خواهد داشت. ولی اگر این اطلاع از طریق نمونه‌گیری به دست آید، مسلماً خطا وجود دارد. سرشماری نیز بدون خطا نیست. بنابراین، نظام آماری مؤثر نظامی است که تمام نیازهای آماری را از طریق ثبت اطلاعات یا داده‌ها پوشش دهد و کمتر به نمونه‌گیری و سرشماری متکی باشد. آن‌جایی که آمار ثبتی قادر نباشد آماری ارائه کند، اتکا به روش دیگر گردآوری اطلاعات خواهد بود. هر قدر سیستم آمارهای ثبتی قوی باشد و اطلاعات مورد نیاز را تولید کند، اتکا به روش دیگر کمتر خواهد بود. ایجاد سیستم آمارهای ثبتی کار، می‌تواند منجر به توقف اجرای برخی از طرح‌های آمارگیری شده و در برخی از طرح‌ها موجب کاهش سؤال‌ها خواهد شد.

همان‌گونه که کیفیت تولیدات صنعتی، کشاورزی، بهداشتی و ارائه خدمات در بخش‌های گوناگون دارای اهمیت است، کیفیت داده و آمار و اطلاعات نیز به عنوان نوعی تولید، اهمیت پیدا می‌کند. در طول دو دهه اخیر تحقیقات گسترده‌ای در زمینه کیفیت آمار و اطلاعات صورت گرفته است، این که مفهوم کیفیت آمار چیست و چگونه می‌توان آن را ارزیابی کرد. کیفیت داده را می‌توان به طور کلی چنین تعریف کرد که «خصوصیات داده تا چه اندازه می‌تواند خواسته‌های کیفی مورد انتظار را مرتفع کند». به بیان ساده‌تر داده‌ی با کیفیت، دارای خصوصیات زیر است:

- صحیح؛
- دقیق؛
- نامتناقض؛
- کامل؛
- به‌هنگام؛
- قابلیت یکپارچگی؛
- قابل فهم و خوش تعریف؛ و
- مرتبط و متناسب با دامنه‌ای که برایش تعریف شده است.

با افزایش به‌کارگیری منابع ثبتی در تولید آمارها، کیفیت آمارهای ثبتی نیز مورد توجه قرار گرفته است. اهمیت آمارهای ثبتی در این است که با نظام‌مند سازی منابع ثبت و استفاده از مدارک و مستندات می‌توان در زمانی کوتاه و بدون در نظر گرفتن هزینه اضافی، آمارهای مورد نیاز را تولید نمود. کیفیت آمارهای ثبتی نه تنها به درستی داده ثبت شده بستگی دارد، بلکه به فرایند ثبت، تعاریف و مفاهیم مورد استفاده، فرایند استخراج و پردازش داده‌ها، شیوه انتقال داده و تجمیع آن‌ها نیز وابسته است. از این رو این ویژگی باعث می‌شود که جنبه‌های کیفیت آمارهای ثبتی کمی متفاوت از آمارهای حاصل از طرح‌های آمارگیری باشد.

۱-۲- مروری بر آمارهای ثبتی و کیفیت آن

ادارات، سازمان‌های دولتی، شرکت‌های خصوصی و بسیاری از مؤسسه‌ها برای استفاده خود اطلاعاتی را گردآوری می‌کنند که اگر اداره‌های آماری بتوانند این اطلاعات را دریافت کرده و از آن‌ها استفاده کنند، تقریباً همیشه هزینه کمتری را متحمل خواهند شد. به عنوان مثال اداره‌های دولتی را در نظر بگیرید. این اداره‌ها بخشی به عنوان امور کارکنان دارند که در آن اطلاعات مربوط به نیروی انسانی از جمله جنسیت، مدرک تحصیلی، سابقه کار، نوع همکاری (استخدام رسمی، پیمانکاری، قراردادی) ثبت و بایگانی می‌شود. اگر اداره آمار به تعداد کارکنان بخش دولتی به تفکیک جنسیت، مدرک تحصیلی، سابقه کار و نوع استخدام نیاز داشته باشد، لازم به اجرای سرشماری نیست. کافی است که این اطلاعات را از امور کارکنان تک تک اداره‌های دولتی دریافت و جمع‌بندی کند و آمار کل را به دست آورد. به آمارهایی که از این طریق به دست می‌آیند، آمارهای ثبتی گفته می‌شوند.

آمارهای ثبتی آمارهایی هستند که از پردازش (طبقه‌بندی، مرتب کردن، محاسبات و تلخیص) داده‌های ثبتی حاصل می‌شوند. منظور از داده‌های ثبتی، داده‌هایی است که در حین انجام فعالیت‌های عینی، واقعی و جاری یک سازمان و با استفاده از فرم‌ها، اسناد، مدارک و فایل‌های کامپیوتری جاری در داخل سازمان، ثبت می‌شوند [۲]. تولید آمار به روش ثبتی دارای ویژگی‌هایی است که مهم‌ترین آن‌ها را می‌توان به صورت زیر برشمرد:

- قابلیت ارایه آمار و اطلاعات صحیح و دقیق در کوتاهترین زمان،
- سهولت دسترسی به آمار،
- مستمر بودن تولید آمار،
- هزینه بر نبودن،
- صرفه جویی در منابع،
- بی‌نیازی از اجرای بسیاری از طرح‌های آمارگیری، و
- کاهش سؤالات در برخی دیگر از طرح‌های آمارگیری.

تولید آمار به روش ثبتی به گونه‌ای که نیازها را پوشش داده و بتواند منجر به تولید آمار و اطلاعاتی صحیح، دقیق و با کیفیت بالا شود، نیازمند نظام آمارهای ثبتی است. نظام آمارهای ثبتی نظامی است که داده‌های ثبتی تولید شده در حین عملیات جاری و روزمره یک سازمان را به اطلاعات آماری موردنیاز مصرف‌کنندگان تبدیل کرده و آنها را به نحو مناسب انتقال و ارائه می‌دهد.

موضوع مهم در نظام آمارهای ثبتی، پرهیز از تحمیل فعالیت با عنوان «گردآوری آمار» به فعالیت‌های جاری و روزمره سازمان است. استفاده از فعالیت‌های جاری سازمان برای ثبت و انتقال داده‌ها و تولید آمارهای ثبتی به‌منظور تأمین نیاز کاربران از ویژگی‌های مهم این نظام است. به بیان دیگر سهولت و دقت در انجام وظایف روزمره که منجر به ثبت دقیق اطلاعات می‌شود، همچنین پرهیز از انجام هرگونه فعالیت اضافی، از ویژگی‌هایی است که نظام آمارهای ثبتی بر آن تأکید دارد. به عنوان مثال:

- کسی برای تولید آمار و اطلاعات، فرم کارگزینی را تکمیل نمی‌کند بلکه این کار برای انجام فعالیت‌های مربوط به استخدام است.
- برنامه برای تولید آمار و اطلاعات تکمیل نمی‌شود، بلکه به‌عنوان مجوزی برای بارگیری، ورود، خروج یا انتقال بار است.

- هیچ دانشجویی با هدف اولیه آمار و اطلاعات، فرم ثبت‌نام را تکمیل نمی‌کند، بلکه هدف، ثبت‌نام در دانشگاه است.
 - قبض برق، آب، گاز و تلفن برای تولید آمار و اطلاعات صادر نمی‌شود، بلکه هدف از صدور آن دریافت هزینه‌های مربوط به تولید و انتقال می‌باشد.
 - صدور حکم بازنشستگی در راستای اجرای قانون بازنشستگی انجام می‌شود.
 - هدف اصلی شماره‌گذاری خودرو برای تهیه آمار نیست، بلکه برای شناسایی خودروها و صاحبان آن‌ها انجام می‌شود.
 - هدف از تشکیل پرونده قضایی، رسیدگی به جرائم و تخلفات و شکایات است نه تولید آمار.
 - هدف از تشکیل تعاونی، خدمت‌رسانی و گسترش فعالیت‌های بخش تعاون می‌باشد.
 - هدف از ثبت میزان تولید برق، ارائه و فروش آن به مصرف‌کنندگان می‌باشد.
- این مثال‌ها تنها تعداد معدودی از مثال‌هایی است که می‌توان به آن اشاره کرد. در اجرای هر کدام از فعالیت‌های استخدام، صدور بارنامه و ثبت نام، ده‌ها داده ثبت می‌شود و بسیاری از این داده‌های ثبت شده دارای صحت و دقت بالایی است و می‌توان از آن‌ها برای تولید آمارهایی دقیق و صحیح استفاده کرد.

۱-۳- آمارهای قابل تولید به روش ثبتی

این ادعا که نظام آمارهای ثبتی قادر است کلیه اقلام آماری موردنیاز کاربران را در دستگاه‌های اجرایی ساماندهی کند، ادعایی اشتباه است. ماهیت برخی اقلام چنان است که قابلیت تولید به روش ثبتی را ندارند، از این‌رو نمی‌توان دستگاه‌های اجرایی را از انجام طرح‌های آمارگیری (سرشماری یا نمونه‌گیری) بی‌نیاز دانست. اگر چه سهم اقلام غیرقابل تولید به روش ثبتی در هریک از دستگاه‌های اجرایی، به تناسب مجموعه اقلام، ساختار سازمانی، فرایندهای عملیاتی - اطلاعاتی و محمل‌های قانونی متفاوت است، با این حال نظام آمارهای ثبتی، نقش غیرقابل انکاری در ایجاد چارچوب برای اجرای طرح‌ها، کاهش تعداد طرح‌ها یا سؤال‌های آن‌ها و افزایش دقت طرح‌ها خواهد داشت. حتی می‌توان پس از استقرار نظام آمارهای ثبتی، برای اطمینان از صحت و درستی اطلاعات کلی، اقدام به سرشماری نمود تا اطلاعات قبل از استقرار نظام بطور کامل در دسترس بوده و پس از آن، این اطلاعات را توسط نظام پیاده‌سازی شده به‌روز نگاه داشت.

می‌توان بین داده‌های حاصل از آمارگیری‌ها و آمارهای ثبتی مبتنی بر داده‌های اداری ارتباط ایجاد کرد. دسترسی به آمارهای ثبتی، باعث بهبود آمارگیری‌ها می‌شود و آمارهای ثبتی را نیز می‌توان با استفاده از آمارگیری‌ها ارزیابی کرد و ارتقاء بخشید. اطلاعات کمکی به‌دست آمده از طریق اطلاعات ثبتی می‌تواند خطای نمونه‌گیری در آمارگیری‌ها را کاهش دهد و متقابلاً می‌توان داده‌های حاصل از آمارگیری‌ها را برای بهبود کیفیت آمارهای ثبتی به‌کار برد.

در حال حاضر در بسیاری از مراکز آماری کشورها از جمله مرکز آمار کشور هلند، آمارهای ثبتی و نمونه‌گیری‌های مختلف از جمله نمونه‌گیری‌های عدم اشتغال و خانوار را با هم ادغام و از آنها اطلاعات آماری مورد نظر را تولید می‌نمایند. بعلاوه، با استفاده از آمارهای ثبتی به عنوان یک مبنای وزن‌دهی، می‌توان واریانس نمونه‌گیری و آریبی ناشی از بی‌پاسخی را کاهش داد.

برای رسیدن به یک قلم آمار ثبتی، نیازمند ثبت داده‌های متناظر آن قلم آماری هستیم. به‌عنوان مثال برای دستیابی به آمار «تعداد کارکنان به تفکیک سن و تحصیل»، نیازمند ثبت داده‌های «تاریخ تولد» و «میزان تحصیل» در پروندهٔ استخدامی هر یک از کارکنان هستیم. ممکن است در برخی دستگاه‌های اجرایی داده‌های متناظر با یک قلم آماری ثبت شود. ثبت چنین داده‌هایی از لحاظ هزینه، نیروی انسانی، همخوانی با وظایف، وجود محمول‌های قانونی و دارای اشکال نباشد و در صورت وجود اشکال، امکان رفع آن وجود داشته باشد که در این صورت ایجاد مکانیزم و سازوکار لازم برای تولید چنین اقلامی لازم است. اما در مواردی آمارهای مورد نیاز قابلیت تولید به روش ثبتی را ندارند. به‌عنوان مثال تعداد معنادان به مواد مخدر در سطح کل کشور. لازم به ذکر است که تعریف داده‌ها و آمارهای ثبتی، داده‌ها و آمارهایی را پوشش می‌دهد که قابلیت تولید به روش ثبتی را دارند. برای تشخیص آنها می‌توان به ملاک‌های زیر توجه کرد [۲]:

- برخورداری از تعریف مشخص، روشن و بدون ابهام
- امکان ثبت داده‌ها در پایین‌ترین نقطه^۱
- امکان ثبت داده‌ها به‌نگام وقوع عمل
- داشتن فرایند(گردش کار) منطقی
- امکان ثبت افزایش‌ها و همچنین کاهش‌ها
- وجود افراد مسؤول و پاسخگو

۱-۴- کیفیت داده

همان‌گونه که اشاره شد، برنامه‌ریزی و تصمیم‌گیری نیازمند آمار و اطلاعات صحیح و با کیفیت بالا است. با پیشرفت‌های صنعتی، مدیران مفهوم استراتژیکی از کیفیت را ارائه نمودند: «برآورده نمودن نیازهای کاربران». این مفهوم، ماهیت موضوعی کیفیت را روشن‌تر نموده است. تعاریف دیگری نیز از کیفیت ارائه شده است که عبارت‌اند از:

- بهبود پیوسته فرآیندها، محصولات و خدمات
- مناسبت در بکارگیری
- مجموعه خصوصیات و ویژگی‌های یک محصول و یا خدمتی که قادر به تأمین نیازهای ضمنی و از پیش تعیین شده باشد

^۱ پایین‌ترین نقطه، معمولاً جایی است که عملیات مربوط به آن قلم آماری انجام می‌شود.

در اصطلاح امروز، کیفیت به درجه‌ای گفته می‌شود که کاربر کالا بر اساس معیارهای بد، خوب و عالی و با توجه به کاربرد آن و مقایسه آن با محصولات مشابه، برای آن کالا تعیین می‌کند. گاهی اوقات از واژه کیفیت مفهوم ارزش مثبت استنباط می‌شود و معنی «کیفیت خوب» را می‌رساند. کیفیت یک کالا بر اساس نظرات موجود و بالقوه کاربران و میزان تناسب آن کالا با اهداف مورد نظر کاربران تعیین می‌شود. آمار و اطلاعات نیز به عنوان یک کالا، باید دارای کیفیت بالایی باشند تا بتوان بر اساس آن برنامه‌ریزی‌ها و سیاستگذاری‌ها را انجام داد. مطالب زیادی در مورد کیفیت آمارها و مدیریت آن در سازمان‌های ملی آمار تهیه و تدوین شده است. تعریف‌هایی که از این مفهوم ارائه شده است، بسیار زیاد است اما همه حول این ایده‌اند که مشخصه‌های محصول در دست تهیه، نیازهای اعلام شده یا ضمنی کاربران را تأمین کنند و یا حتی از آن فراتر روند. کیفیت آمارها نیز دارای تعاریف متفاوتی است ولی آن‌چه که در تمام تعاریف مشترک است، آن است که کیفیت داده‌ها را به صورت یک مفهوم چند بعدی و بر اساس ابعاد مختلف داده، مورد بررسی قرار می‌دهند. در این قسمت به طور خلاصه تعاریف گوناگونی که از کیفیت داده ارائه شده است، بیان می‌شود.

در کاربرد آمار و اطلاعات، خصوصاً در حوزه اقتصاد، کیفیت بر حسب بهنگام بودن و درستی آمارهای اصلی تعریف می‌شود. علاوه بر این، قضاوت آمارشناسان اصلی دولت در مورد دقت آمارها (از لحاظ عدم دخالت سیاست در آن)، قابلیت اعتماد اعداد، روش‌شناسی آماری به کار رفته و مناسبت ارقام منتشر شده نیز دارای اهمیت است. آرون‌دل و دی‌پوت^۱ در بررسی‌های خود پیشنهاد می‌کنند که سازمان‌های آماری «کیفیت» را به مؤلفه‌ها یا مشخصه‌هایی تفکیک کنند که بر چند مفهوم کلیدی یعنی درستی^۲، مرتبط بودن^۳، به موقع بودن^۴ و قابلیت دسترسی^۵ تمرکز داشته باشد [۱۴].

نقطه شروع تعریف کنونی کیفیت در اداره آمار اتحادیه اروپا^۶، تعریف ایزو ۸۴۰۲ است که کیفیت را به صورت «مجموعه خصوصیات و ویژگی‌های یک محصول و یا خدمت که قادر به تأمین نیازهای ضمنی و از پیش تعیین شده باشد» تعریف می‌کند. این تعریف بسیار کلی، کاربران و نیازهایشان را در مرکز توجه قرار می‌دهد. با نگاهی از نقطه نظر کاربر به واژه «کیفیت»، اداره آمار اتحادیه اروپا ۷ بعد کیفیت داده را به صورت زیر در نظر گرفت: مناسبت، درستی، بهنگام بودن و به موقع بودن، وضوح^۷ و قابلیت دسترسی، انسجام، مقایسه‌پذیری و کامل بودن^۸.

طبق تعریف سازمان همکاری‌های اقتصادی و توسعه^۹، کیفیت مفهومی چند بعدی است. مهم‌ترین مشخصه‌های کیفیت به نقطه نظر، نیازها و اولویت‌های کاربران وابسته است. این عوامل با توجه به گروه‌های مختلف کاربران، متغیر است. کیفیت بر اساس ابعاد مرتبط بودن، درستی، اعتبارمندی^{۱۰}، بهنگام بودن^{۱۱}، قابلیت دسترسی، تفسیرپذیری^۱ و

^۱ Arondel, and Depoutot

^۲ Accuracy

^۳ Relevance

^۴ Punctuality

^۵ Accessibility

^۶ Eurostat: European Union statistical office

^۷ Clarity

^۸ Completeness

^۹ OECD: Organization for Economic Co-operation and Development

^{۱۰} Credibility

^{۱۱} Timeliness

انسجام^۲ تعریف می‌شود. این تعریف بر اساس فعالیتی که اخیراً توسط چندین سازمان، به خصوص اداره آمار اتحادیه اروپا، صندوق بین‌المللی پول^۳، سازمان همکاری‌های اقتصادی و توسعه و اداره آمار کانادا در حوزه کیفیت داده‌ها انجام شد، ارائه شده است.

کشور کانادا ابعاد کیفیت داده را از جنبه‌های مناسب^۴، درستی، بهنگام بودن، قابلیت دسترسی، انسجام، مقایسه-پذیری^۵ و تفسیرپذیری تعریف کرده است. از نظر کشور هلند کیفیت داده‌ها بر اساس مناسب، درستی، بهنگام بودن، اقتصادی بودن^۶ و بار پاسخگویی^۷ تعریف می‌شود. مناسب، درستی، بهنگام بودن و به موقع بودن، در دسترس بودن، انسجام، مقایسه‌پذیری و کامل بودن ابعادی هستند که کشور انگلستان برای تعریف کیفیت داده‌ها انتخاب کرده است [۵].

کیفیت داده‌ها در صندوق بین‌المللی پول که نقش عمده‌ای در مقایسه آمارهای کشورهای مختلف در حوزه‌های مالی و اقتصادی دارد، نیز دارای اهمیت زیادی است. این صندوق در سال‌های ۱۹۹۶ و ۱۹۹۷ به ترتیب استاندارد تولید داده‌های خاص^۸ و نظام تولید داده‌های عمومی^۹ را معرفی کرد که هر دو راهنمایی‌هایی به کشورها در زمینه ارائه داده‌ها به عموم فراهم می‌آورند. صندوق بین‌المللی پول برای ارزیابی کیفیت داده‌های ارائه شده توسط کشورهای عضو، چارچوب ارزیابی کیفیت داده^{۱۰} را معرفی نمود. این چارچوب که برای ارزیابی جامع کیفیت داده‌های کشورهای عضو استفاده می‌شود، شرایط و ضوابط مؤسسه‌ها، فرایندهای آماری و ویژگی‌های محصولات آماری را پوشش می‌دهد. این چارچوب نگاهی کلی به کیفیت داده دارد و به عنوان دربرگیرنده جنبه‌های گوناگون کیفیت داده در یک مدل سیستماتیک (شامل نهادها، مدیریت، فرایندهای مرکزی، همچنین ویژگی‌های آمارهای خروجی) شناخته می‌شود. چارچوب کیفیت، اداره‌ی سیستم‌های آماری، فرایندهای آماری مرکزی و تولیدات آماری را در نظر می‌گیرد و نظر کاربران درباره محصولات آماری را مورد توجه قرار می‌دهد. بر این اساس درستی و قابلیت اعتماد^{۱۱}، قابلیت استفاده^{۱۲}، قابلیت دسترسی، معتبر بودن روش^{۱۳} و یکپارچگی^{۱۴} به عنوان ابعاد کیفیت داده در نظر گرفته شد.

همان‌گونه که مشاهده می‌شود هر کشور یا سازمانی ابعاد خاصی از کیفیت داده را در نظر گرفته‌است. علت تفاوت در ابعاد کیفیت را می‌توان در دیدگاه و نگرش به کیفیت دانست. به عنوان مثال، صندوق بین‌المللی پول دیدگاه فرایندگرا به داده دارد در حالی که دفتر آمار اتحادیه اروپا، دیدگاه محصول‌گرا دارد. بر این اساس، ملاک‌های ارزیابی کیفیت نیز متفاوت خواهد بود. صندوق بین‌المللی پول ملاک‌های کیفی برای ارزیابی کیفیت داده‌ها دارد. در واقع

^۱ Interpretability

^۲ Coherence

^۳ IMF: International Monetary Fund

^۴ Relevance

^۵ Comparability

^۶ Cost Effective

^۷ Response burden

^۸ SDDS: Special Data Dissemination Standard

^۹ GDDS: General Data Dissemination System

^{۱۰} DQAF: Data Quality Assessment Framework

^{۱۱} Reliability

^{۱۲} Serviceability

^{۱۳} Methodological Soundness

^{۱۴} Integrity

در DQAF ملاک‌های کیفیت به شکل قابل قبولی از نظر بین‌المللی ارائه شده است تا آمارهای ملی در مقایسه با آن‌ها ارزیابی گردد (مانند مشاهده شد، کاملاً مشاهده شد، کاملاً مشاهده نشد، مشاهده نشد). در حالی که در دفتر آمار اتحادیه اروپا، ملاک‌های ارزیابی کیفیت به گونه‌ای تهیه شده‌اند که مقادیر کمی مانند شاخص رضایتمندی کاربران و نرخ‌های پاسخ در مورد کیفیت ارائه می‌دهد. علاوه بر دیدگاه محصول‌گرا و فرایندگرا که برای ارزیابی کیفیت داده ارائه شده‌اند، چارچوب‌های دیگری نیز وجود دارند که در ادامه به معرفی مختصر چند دیدگاه پرداخته می‌شود.

۱- کیفیت بر پایه نظریه کیفیت خدمات و محصولات (دیدگاه داده‌گرا) [۱۵].

این دیدگاه بر اساس ویژگی‌های محصول و خدمات به ارائه تعریف می‌پردازد. بر این اساس در دایره المعارف آمارهای رسمی، ابعاد کیفیت بر پایه این دیدگاه، شامل درستی، بهنگام بودن، انسجام و مقایسه‌پذیری (در طول زمان و مکان)، دسترسی و وضوح می‌شود.

۲- کیفیت بر پایه اصول هستی‌شناسانه (دیدگاه طراحی-گرا) [۱۶].

در این دیدگاه کیفیت داده تولید شده توسط یک سیستم اطلاعاتی به طراحی و عملکرد آن سیستم بستگی دارد. اما استفاده صحیح از داده، خارج از کنترل طراح سیستم است. بنابراین لازم است تعریفی طراحی‌گرا از کیفیت داده ارائه شود به گونه‌ای که استفاده خواسته شده از اطلاعات را منعکس کند. در این جا کیفیت داده نه از طریق داده بلکه از طریق فرایند تولید داده و طراحی سیستم تحلیل می‌شود.

۳- کیفیت بر پایه نظرات شاغلان و شرکت کنندگان [۱۷].

این دیدگاه بر پایه نظرات شاغلان و شرکت کنندگان شکل گرفته است و اعتقاد دارد که مفهوم‌سازی کیفیت داده بیش‌تر به مصرف کنندگان داده‌ها ارتباط دارد تا متخصصان توسعه سیستم‌ها. این دیدگاه به دنبال ارائه چارچوبی از کیفیت داده بر اساس نظر مصرف کنندگان است.

۴- کیفیت بر پایه نظریه نشانه‌شناسی [۱۸].

این دیدگاه بیان می‌کند که کاربران گوناگونی مانند مدیران، کاربران نهایی مختلف و متخصصان توسعه سیستم با سیستم در تعامل هستند و هر یک دیدگاه خاص خود را نسبت به کیفیت داده‌های سیستم دارند. طراحی سیستمی که تمام دیدگاه‌ها را به طور کامل در نظر بگیرد امکان‌پذیر نیست. لذا این دیدگاه به دنبال ارائه چارچوبی بهینه برای بهبود توسعه سیستم بر اساس دیدگاه‌های ذینفعان مختلف است. در استاندارد الزامات و ارزیابی کیفیت محصولات نرم‌افزاری (استاندارد ISO/IEC25012) نیز تعریفی از کیفیت داده ارائه شده است. این مدل کیفیت داده، چارچوبی را برای مشخص کردن نیازهای کیفی و ارزیابی کیفیت داده فراهم می‌کند. این مدل که برای طراحی نرم‌افزارها معرفی شده است، برای کاربردهای زیر مورد استفاده قرار می‌گیرد:

- تعریف ارزیابی نیازمندی‌های کیفی داده در فرایندهای تولید داده، اخذ داده و یکپارچگی آن
- شناسایی معیارهای تضمین کیفیت داده برای مهندسی مجدد، ارزیابی و بهبود داده
- ارزیابی مطابقت داده با قوانین و یا نیازمندی‌ها
- تشخیص خطا و نادرستی داده ناشی از تغییرات اجزای سیستم از قبیل تغییرات کد برنامه، ارتقاء سخت افزار سیستم، تغییر فرآیندهای کاری

در این مدل، کیفیت داده از دو دیدگاه زیر تعریف می‌شود:

- کیفیت ذاتی: آن بعدی از کیفیت داده است که به خود داده مربوط است و توانایی برآوردن نیازمندی‌های کیفی را تحت شرایط تعریف شده دارد. مانند مقادیر دامنه داده، محدودیت، ارتباطات بین مقادیر داده و فراداده.
- کیفیت وابسته به سیستم: بعدی از کیفیت داده است که در یک سیستم کامپیوتری تحت شرایط تعریف شده حاصل می‌شود و وابسته به محیطی است که داده در آن مورد استفاده قرار می‌گیرد. این نوع کیفیت داده توسط توانایی‌های اجزای سیستم کامپیوتری مانند سخت افزار (در دسترس قرار دادن داده برای تأمین دقت داده)، نرم افزار (نرم افزارهای پشتیبان‌گیری برای تأمین قابلیت بازیابی داده) و سایر نرم افزارها (ابزارهای انتقال داده برای حفظ ویژگی انتقال پذیری) تضمین می‌شود.

براساس دو دیدگاه تعریف شده برای کیفیت داده در این استاندارد، ابعاد کیفی در سه دسته زیر و به ۱۵ بعد تقسیم می‌شوند.

- ابعاد کیفی ذاتی داده که عبارتند از: دقت، کامل بودن، سازگاری، معتبر بودن، بروز بودن؛
- ابعاد کیفی داده که وابسته به سیستم هستند و عبارتند از: در دسترس بودن، انتقال پذیری، قابلیت بازیابی.
- هفت ویژگی کیفی وابسته به هر دو دیدگاه هستند. این هفت بعد عبارتند از: قابلیت دستیابی، قابلیت انطباق، محرمانگی، کارایی، دقت، قابلیت ردیابی و قابلیت فهم.

۱-۴-۱- کیفیت آمارهای ثبتی

مفهوم کیفیت داده‌ها و به‌طور ویژه کیفیت داده‌های ثبتی در طرح پژوهشی «تعیین شاخص‌های کیفیت داده‌های ثبتی برای ارزیابی کیفیت داده‌های مذکور» به تفصیل شرح داده شده است. این مفهوم در بخش قبل نیز ارائه شد و دیدگاه‌های متفاوتی که برای ارائه تعریف کیفیت داده وجود دارد، به طور خلاصه معرفی گردید. با توجه به تعریف آمارهای ثبتی و لزوم وجود نظامی برای ساماندهی آمارهای ثبتی [۲]، به نظر می‌رسد که ارائه مفهوم کیفیت درباره این‌گونه آمارها، با رویکرد دیدگاه طراحی‌گرا و نیز استاندارد ISO/IEC25012 مناسب‌تر باشد. چرا که ایجاد نظام آمارهای ثبتی نیازمند طراحی سیستم‌هایی در حوزه‌های شناسایی، ثبت، انتقال، پردازش و ارائه داده‌ها و طراحی نرم‌افزارهای مناسب است. با در نظر گرفتن این رویکرد و ۵ حوزه کاری نظام آمارهای ثبتی، ۱۲ بعد کیفیت مناسب تشخیص داده شد [۹]. لازم به ذکر است که برخی از ابعاد در دو یا چند حوزه مشترک می‌باشند. البته اهمیت این ابعاد یکسان نبوده و برخی دارای اهمیت بیش‌تری می‌باشند. این ابعاد به ترتیب اهمیت عبارتند از [۹]: ۱- درستی، ۲- قابلیت اعتماد، ۳- بهنگام بودن، ۴- کامل بودن، ۵- جاری بودن^۱، ۶- سازگاری^۲، ۷- دقت^۳، ۸- شکل و قالب^۴، ۹- کارایی^۵، ۱۰- کفایت^۶، ۱۱- قابل استفاده بودن و ۱۲- مقایسه‌پذیری. در فصل‌های بعدی به طور کامل به این ابعاد پرداخته شده است.

^۱ Currency

^۲ Consistency

^۳ Precision

^۴ Format

^۵ Efficiency

^۶ Sufficiency

۲

حوزه‌های کاری نظام آمارهای ثبتی، نحوه‌ی ارزیابی هر حوزه و ابعاد کیفیت مرتبط با آنها

۲-۱- مقدمه

نظام آمارهای ثبتی یک نظام جامع است و چنانچه کیفیت در مرحله‌ای از آن با مشکل مواجه شود، مسلماً بر سایر مراحل نیز تأثیرگذار خواهد بود. برای ایجاد نظام آمارهای ثبتی در دستگاه اجرایی باید به پنج حوزه کاری توجه کرد. این حوزه‌ها شامل شناسایی، ثبت، انتقال، پردازش و ارائه داده‌ها هستند [۲]. با توجه به اهمیت هر حوزه، ابزار مورد نظر برای ارزیابی و اندازه‌گیری کیفیت آمارهای ثبتی باید به گونه‌ای طراحی شود که امکان بررسی این موضوع را در هر حوزه داشته باشد. با توجه به نتایج طرح پژوهشی «تعیین شاخص‌های کیفیت داده‌های ثبتی برای ارزیابی کیفیت داده‌های مذکور» واضح است که بهترین راه برای ارزیابی کیفیت یک قلم آماری ثبتی، تهیه فهرستی از سؤال‌هایی است که بتواند تمام موضوعات مهم در هر حوزه را پوشش دهد و با امتیازدهی به هر موضوع، نمره کیفیتی به آن قلم آماری اختصاص داده شود [۹].

در این فصل، ابتدا حوزه‌های کاری آمارهای ثبتی و مفاهیم مهم در هر حوزه معرفی می‌شود. سپس به کمک آن‌ها، فهرست سؤال‌های اصلی در هر حوزه تهیه می‌گردد. این فهرست در واقع ابزار مورد استفاده در ارزیابی کیفیت آمارهای ثبتی است. پس از معرفی سؤال‌های هر حوزه، به ابعاد کیفیت مرتبط با هر سؤال پرداخته می‌شود. در آخر درباره نحوه امتیازدهی به هر سؤال (طیف لیکرت) و دلیل انتخاب این شیوه به عنوان روش امتیازدهی توضیح داده می‌شود.

۲-۲- حوزه‌های کاری نظام آمارهای ثبتی و طراحی سئوالات چک لیست بر

اساس آن‌ها

همان‌گونه که اشاره شد در بحث آمارهای ثبتی ۵ حوزه کاری وجود دارند که عبارتند از: شناسایی، ثبت، انتقال، پردازش و ارایه. ما بر این اعتقادیم که کیفیت داده‌های تولید شده در نظام آمارهای ثبتی به عملکرد این پنج حوزه بر می‌گردد و چنانچه سیستم اطلاعاتی نتواند حداقل یکی از حوزه‌های پنج گانه فوق را به خوبی تحت کنترل قرار دهد، کیفیت داده‌های تولیدی آن سیستم مخدوش خواهد شد و با افزایش تعداد حوزه‌هایی که به خوبی تحت کنترل سیستم اطلاعاتی نیستند میزان خدشه به کیفیت داده‌های تولیدی سیستم نیز افزایش خواهد یافت. به همین دلیل ما برای اندازه‌گیری کیفیت داده‌های ثبتی به این حوزه‌ها پرداخته‌ایم. البته لازم به ذکر است که اهمیت این حوزه‌ها در میزان کیفیت داده‌های ثبتی به یک اندازه نیست. به طوری که حوزه‌ها به ترتیب اهمیت عبارتند از [۹]:

۱. ثبت ۲. شناسایی ۳. انتقال ۴. ارایه ۵. پردازش

الف - شناسایی

برای ایجاد هر نظام اطلاعاتی باید در ابتدا نیازها را شناسایی کرد. این نیازها شامل شناسایی کاربران، اطلاعات و آمار موردنیاز آن‌ها، فرآیندها و سیستم‌های تولیدکننده این آمار، تعاریف و مفاهیم مربوطه، استانداردهای مربوط به ثبت و انتشار چه از نظر زمانی و چه مکانی، روش و واحد اندازه‌گیری، تفکیک‌ها و طبقه‌بندی‌های بین‌المللی می‌شود. در واقع این حوزه کاری به فراداده‌های آماری مرتبط می‌شود.

اصطلاح فراداده نخستین بار توسط جک ای. مایرز و اس. کی. کول جیان در سال ۱۹۶۹ به عنوان نام شرکتی تجاری (آمریکایی) برای ایجاد و توسعه محصولات مربوط به «فراالگوها» به کار رفت و در یک جزوه معرفی محصول در سال ۱۹۷۲ چاپ شد. چند سال بعد، یعنی در سال ۱۹۸۶، علامت تجاری METADATA برای شرکت فراداده ثبت گردید. اصطلاح فراداده بعدها بوسیله متخصصان علوم رایانه، آمار، پایگاه‌های اطلاعاتی، و جامعه کتابداری و اطلاع‌رسانی به شکل‌های مختلف استفاده شد. اما کاربرد مکرر این اصطلاح بیشتر به دهه ۱۹۹۰ و شکل‌گیری شبکه جهانی وب در سال ۱۹۹۳ مربوط می‌شود. فهرست‌ها یا کاتالوگ‌های کتابخانه‌ها مثالی از یک نوع فراداده پرسابقه است که در دهه‌های اخیر برای مدیریت مجموعه‌ها و ابزار کشف منابع به خدمت گرفته شده‌است. از شناخته شده‌ترین قالب‌های فراداده‌ای در دهه ۱۹۹۰ می‌توان به طرح کدگذاری متن، خدمات مکان‌یابی اطلاعات دولتی، تبادل رایانه‌ای اطلاعات موزه‌ها، چارچوب توصیف منبع، توصیف کدگذاری شده آرشیوی و طرح فراداده‌ای هسته دوبلین اشاره کرد. فراداده‌ها محتوا، کیفیت و شرایط داده‌ها را توصیف نموده، به پشتیبانی سیستم‌های اطلاع‌رسانی و به انتقال داده‌ها بین سیستم‌های رایانه‌ای یا اشتراک داده‌ها با سایر سازمان‌ها کمک می‌نمایند. در متون و علوم مختلف، اصطلاحات گوناگونی برای فراداده استفاده می‌شود. از جمله می‌توان به «ابرداده» اشاره کرد. برای فراداده‌ها تعاریف مختلفی وجود دارد که در زیر به چند نمونه آن‌ها اشاره می‌شود:

- فراداده‌ها، داده‌هایی هستند که برای تولید مناسب و استفاده از داده‌هایی که درباره آن‌ها اطلاع‌رسانی می‌شود، لازمند.

- فراداده‌ها، اطلاعاتی در مورد داده‌ها و فرایند تولید و بکارگیری آن‌ها می‌باشند.

- همان‌طور که داده‌ها از طریق سیستم‌های اطلاع‌رسانی مدیریت می‌شوند، فراداده‌ها نیز توسط سیستم‌های فرا اطلاع‌رسانی مدیریت می‌شوند. سیستم فرا اطلاع‌رسانی از یک سو از فراداده‌ها استفاده و از سوی دیگر فراداده‌ها را تولید می‌کند، درباره داده‌ها اطلاع‌رسانی می‌کند و وظایفش را با استفاده از عملیاتی مانند «گردآوری فراداده‌ها»، «پردازش فراداده‌ها»، «ذخیره فراداده‌ها» و «اطلاع‌رسانی فراداده‌ها» انجام می‌دهد.
- فراداده، مکرراً «داده درباره داده» یا «اطلاعات درباره اطلاعات» تعریف شده است. به عبارت دیگر، فراداده، داده‌ای است که منابع اطلاعاتی را توصیف می‌کند. این تعریف گسترده، سطوح گوناگونی از توصیف (ساده تا پیچیده) را در بر می‌گیرد. از جمله: یادداشت توصیفی کوتاهی در مورد یک کتاب، توصیف غیر رسمی که موتورهای کاوش در مورد پیشینه‌های بازیابی شده ارائه می‌دهند، یک پیشینه کتاب‌شناختی چاپی (فهرست‌برگه) یا الکترونیکی، چکیده‌ها و اصطلاحات نمایه‌ای و حتی یک استناد کتاب‌شناختی. اما به طور دقیق‌تر می‌توان فراداده را «داده‌های ساختارمند درباره دیگر داده‌ها» دانست.
- به طور خلاصه فراداده را «داده‌ای برای داده»، «داده‌ای درباره داده» یا «اطلاعات درباره اطلاعات» تعریف می‌کنند.

می‌توان گفت فراداده‌ها در حالت کلی برای تفسیر داده‌ها به کاربر کمک می‌کنند. تعریف کلی و جامع فراداده را می‌توان به صورت زیر بیان نمود:

«فراداده‌های آماری، برای تولید و استفاده مناسب از داده‌های آماری ضروری هستند و تا حدودی فرایندها و ابزارهای به کار رفته در تولید و استفاده از داده‌های آماری را توصیف می‌کند. به بیان خلاصه فراداده‌های آماری، داده‌هایی درباره داده‌های آماری هستند.»

فراداده‌ها در نظام‌های تولید آمار، اعم از نظام‌های ثبتی و طرح‌های آمارگیری به طور دستی یا خودکار تولید می‌شوند. به عبارت دیگر فراداده‌ها بخش جدایی‌ناپذیر این نظام‌ها هستند به گونه‌ای که تولید و اطلاع‌رسانی آمار بدون آن‌ها امکان‌پذیر نیست. در واقع فراداده‌ها ارتباط معنی‌دار بین تولیدکنندگان و کاربران داده‌ها را برقرار می‌سازند. این فراداده‌ها علاوه بر کاربران، مورد نیاز کارشناسان درگیر در مراحل مختلف تولید آمار نیز هستند. گاه نبود فراداده منجر به تصمیم‌گیری‌ها یا ارزیابی‌های اشتباه می‌شود. در مثال‌های زیر اهمیت فراداده به وضوح مشاهده می‌شود.

مثال (۱) یک شبکه خبری معتبر، درآمد کشورهای مختلف در خاورمیانه را اعلام می‌کند و در نهایت نتیجه می‌گیرد «متوسط درآمد در کویت بسیار بیشتر از سایر کشورهای عربی است».

این عبارت به چه معنی است؟ آیا نشان می‌دهد مردم کویت در رفاه بیشتری نسبت به سایر کشورهای عربی هستند؟ این عبارت تا زمانی که بیان نشود از کدام نوع متوسط استفاده شده است، بی معنی است و هیچ چیز خاصی را نشان نمی‌دهد. در واقع باید روش محاسبه میانگین و منابع مورد استفاده در این مورد، بیان شود تا بتوان به درستی قضاوت کرد.

مثال (۲) سال‌ها پیش وقتی که دانشگاه جان هاپکینز آغاز به پذیرش دانشجوی زن کرد، مطلبی غیرمنتظره گزارش نمود: «۳۳/۳۳٪ زنان دانشگاه جان هاپکینز با اعضای هیأت علمی دانشگاه ازدواج کرده‌اند».

ارقام خام موضوع را بر ملا کرد. در آن زمان در دانشگاه مذکور فقط ۳ زن دانشجو وجود داشت و یکی از آن‌ها با یکی از اعضای هیأت علمی دانشگاه ازدواج کرده بود. در واقع اگر تعداد دانشجویان زن در گزارش اعلام شده بود، این نتیجه‌گیری چندان هم غیرمنتظره نبود.

مثال ۳) یک تور مسافرتی برای جذب مشتری به منطقه‌ای، این تبلیغ را به کار برد: «منطقه‌ای خوش آب و هوا با میانگین دمای حدود ۲۳ درجه سانتیگراد». بعد از سفر به این منطقه مسافران متوجه شدند که تغییر دما بین ۰ تا ۵۰ درجه است. اگر مسافران می‌دانستند که تغییرات دما تا این اندازه است، شاید این منطقه را برای مسافرت انتخاب نمی‌کردند.

مثال ۴) برنامه سازمان مسکن آمریکا برای اسکان خانواده‌های ۳/۶ نفره که متوسط افراد خانواده‌های آمریکایی است، تدوین شده است. یعنی خانواده‌هایی که ۳ یا ۴ نفره هستند و نیاز به خانه ۲ اتاق خوابه دارند. بنابراین گزارش، بساز و بفروش‌ها بیشتر شروع به ساخت خانه‌های ۲ اتاق خوابه نمودند. با این شعار که «ما خانه متوسط برای خانواده متوسط می‌سازیم». اما این برنامه دچار مشکل شد.

عدد ۳/۶ گرچه متوسط تعداد افراد یک خانواده آمریکایی است ولی عملاً نسبت به دیگر خانواده‌ها با جمعیت‌های مختلف در اقلیت هستند در حالی که خانواده‌های ۲ نفره و یا ۴ نفره و بیشتر، که اکثریت خانوارها را تشکیل می‌دادند به دست فراموشی سپرده شدند. در نتیجه در بسیاری از مناطق خانه‌های ۲ اتاق خوابه بیشتر از حد لازم است و تعداد خانه‌های بزرگتر و یا کوچکتر به اندازه کافی نیست. وقتی به داده‌های واقعی که ارائه غلطی از آن صورت گرفته است توجه شود دیده می‌شود که خانواده‌های ۳ یا ۴ نفره ۴۵٪ کل را تشکیل می‌دهند، در حالی که ۳۵٪ خانواده‌ها یک نفره یا ۲ نفره و ۲۰٪ خانواده‌ها بیش از ۴ نفرند. اگر در تهیه برنامه سازمان مسکن به فراوانی خانوارها به تفکیک بعد خانوار توجه می‌شد، این مشکل به وجود نمی‌آمد.

مثال ۵) به عنوان مثال عملی دیگر از فراداده فرض کنید در یکی از خانه‌های یک جدول آماری، رقم ۴,۵۰۰,۰۰۰ عنوان شده است. این رقم یک داده خام است. برای مثال شاید دریابید که رقم فوق میانگین درآمد ماهانه مردمی باشد که در طول سال ۱۳۸۶ در شهر تهران زندگی می‌کنند ولی هنوز تردیدهایی وجود دارد. معنی درآمد چیست؟ آیا منظور تنها درآمد حاصل از کار منظم است یا درآمدهای اضافی را هم شامل می‌شود؟ درآمد حاصل از سرمایه‌گذاری، کار، پانسیون است یا خیر؟ چه کسانی یا چه تعدادی در سال ۱۳۸۶ در شهر تهران زندگی می‌کرده‌اند؟ ممکن است برخی افراد زمانی کوتاه از سال را در شهر بوده‌اند و بنابراین چگونه درآمد چنین افرادی با آنهایی که در تمام طول سال در شهر تهران زندگی کرده‌اند مقایسه می‌شود؟ رقم ۴,۵۰۰,۰۰۰ دقیقاً چگونه به دست آمده است؟ آیا از سرشماری بدست آمده یا مبتنی بر یک طرح آمارگیری است یا از اطلاعات ثبتی به دست آمده است؟

مثال‌های بالا نشان دهنده اهمیت فراداده در درک صحیح اعداد، ارقام و آمار منتشر شده برای کاربران است. در بحث ارزیابی کیفیت آمارهای ثبتی، فراداده در حوزه شناسایی و در شناخت وضع موجود اقلام آمارهای ثبتی اهمیت می‌یابد. چرا که در این مرحله، شناخت و آشنایی با فراداده‌های اقلام از قبیل تعاریف و مفاهیم، واحدهای اندازه‌گیری، تفکیک‌ها، استانداردهای مکانی و زمانی و پوشش اقلام (مکانی و زمانی) نقش بسزایی در جلوگیری از بروز تناقض‌های آماری دارد.

در ادامه برای آشنایی بیشتر با فراداده‌های مورد نیاز در حوزه شناسایی نظام آمارهای ثبتی جزئیات بیشتری ارائه می‌شود.

• طراحی سئوال‌ات مرتبط با حوزه شناسایی

اولین مسأله در حوزه شناسایی پوشش قلم مربوطه است. بدین منظور که آیا عنوان انتخابی برای قلم آماری موردنظر تمامی رخدادهای مربوطه را تحت پوشش قرار می‌دهد؟ آیا تمامی افرادی که نسبت به ثبت قلم مورد نظر اقدام می‌کنند، همچنین تمامی کاربران برداشت یکسانی از عنوان آن دارند؟