

روش‌های شناخت کیفیت داده‌های اداری

مهتاب ایرانفر

رئیس گروه آمارگیری‌های صنعت، خدمات و بازرگانی

دفتر روش‌شناسی آماری و طرح‌های نمونه‌گیری مرکز آمار ایران

واژگان کلیدی: نماگرهای کیفیت، داده‌های اداری

مقدمه

برای اندازه‌گیری کیفیت آن دسته از داده‌های اداری که به عنوان منبع ورودی برای فرایند آماری سازمان‌های ملی آمار، استفاده می‌شوند، نماگرهایی با عنوان نماگرهای کیفیت در WP4 (بسته کاری شماره ۴ پروژه BLUE-ETS)^۱ تعریف شده‌اند. این نماگرها قابلیت اندازه‌گیری کیفیت داده‌ها را در یک منبع اداری در قالب پنج بعد کلی بررسی فنی^۲، درستی^۳، کامل بودن^۴، وابستگی به زمان^۵ و جمع‌پذیری^۶ مورد بررسی قرار می‌دهند. این مقاله، برگرفته از اقدامات انجام شده در WP4 است که در آن سعی شده تنها با مرور روش‌های به کارگیری داده‌های اداری، نحوه تعیین کیفیت منابع داده‌های اداری و ابعاد آن، مورد سنجش و ارزیابی قرار گیرد.

روش‌های به کارگیری داده‌های اداری

بسیاری از سازمان‌های ملی آمار^۷ به دنبال افزایش کاربرد داده‌های اداری (یعنی ثبت) برای نیل به اهداف آماری هستند. به منظور استفاده از منابع داده‌های اداری توسط سازمان‌های ملی آمار، پیش‌نیازهایی باید رعایت شوند که به ترتیب اهمیت، عبارتند از:

۱. دسترسی به منابع داده‌های اداری

۲. تطبیق سازمان ملی آمار با مجموعه‌ای از پیش‌شرطها برای امکان استفاده منظم از منابع داده‌های اداری

۳. فراهم بودن روش‌هایی برای ارزیابی قابلیت استفاده آماری (یعنی کیفیت) منابع داده‌های اداری به روش استاندارد

1 workpackage 4 of the BLUE-ETS project (wp4)

BLUE ENTERPRISE AND TRADE STATISTICS (BLUE-ETS) در مورد آمار رسمی کسب و کار است و به طور خاص به یکی از چالش‌های عمده سازمان‌های ملی آمار، کشورهای عضو اتحادیه اروپا و سایر کشورها یا هر سازمان بین‌المللی دیگری که در آمار کارگاهی و تجاری کار می‌کند، پرداخته و هدف آن ارائه اطلاعات کسب و کار آماری با کیفیت بالا به منظور به کارگیری در تحقیقات سیاسی، اقتصادی اجتماعی و حمایت از استراتژی اروپا ۲۰۲۰ است.

2 technical check

3 accuracy

4 completeness

5 Time-related

6 Integrability

7 NSI (National statistical Institute)

برای فراهم کردن امکان استفاده از داده‌های اداری برای اهداف آماری، منابع مربوط به داده‌های اداری باید در دسترس سازمان ملی آمار قرار گیرد. امروزه با توجه به افزایش استفاده از فناوری اطلاعات و ارتباطات در ادارات دولتی (دولت الکترونیک)، این مشکل نباید در بیش‌تر کشورها وجود داشته باشد (سوسیتام، ۲۰۰۲). حداقل این انتظار وجود دارد که برخی از این منابع داده برای سازمان ملی آمار، به‌طور بالقوه مورد توجه باشند (والگرن، ۲۰۰۷).

پیش‌نیاز دوم، موضوع «مقاله بهترین تجربیات» توسط سازمان‌های ملی آمار کشورهای شمال اروپا (کمیسیون اقتصادی ملل متحد برای اروپا «Unece»، ۲۰۰۷) بوده که در آن فهرست کاملی از پیش‌شرط‌های لازم برای توانا کردن سازمان‌های ملی آمار به منظور استفاده گسترده از منابع اداری در تولید آمار را به شرح ذیل ارائه می‌دهد:

(۱) مبنای قانونی برای استفاده از منبع داده‌های اداری،

(۲) درک عمومی و تأیید مزایای استفاده از منابع اداری برای اهداف آماری،

(۳) در دسترس بودن سیستم شناسایی یکپارچه^۸ در میان منابع مختلف استفاده شده،

(۴) وجود سیستم‌های جامع و قابل اعتماد در ادارات دولتی و

(۵) همکاری میان مقامات اداری.

در صورت تطبیق با این شرایط، سازمان ملی آمار قادر خواهد بود به صورت منظم از داده‌های اداری برای تولید آماری استفاده کند و کار بیشتری در این زمینه لازم نیست.

پس از طرح دو پیش‌نیاز، سومین مورد یعنی قابلیت استفاده آماری از منابع داده‌های اداری اهمیت می‌یابد. برای مقابله با نوسان در کیفیت این منابع، ضروری است سازمان ملی آمار به صورت منظم، قابلیت استفاده آماری (کیفیت) این منابع را تعیین کند. این مسئله مهم است، زیرا جمع‌آوری و نگهداری یک منبع داده اداری، فراتر از کنترل سازمان ملی آمار است. نگهدارنده داده‌های اداری است که این جنبه‌ها را مدیریت می‌کند. از این رو اهمیت حیاتی دارد که سازمان ملی آمار دارای رویه‌ای باشد تا بتواند کیفیت داده‌های اداری را به‌طور سریع، ساده و استاندارد تعیین کند. با این حال، هنوز ابزار یا فرایند استاندارد برای چنین منابع داده‌ای در دسترس نیست (برکا و همکاران ۲۰۱۱، داس و همکاران ۲۰۱۰، فراست و همکاران ۲۰۱۰). توسعه چنین فرایندی، تمرکز اصلی WP4 است. نیاز به توسعه رویکردی است که برای طیف وسیعی از منابع داده‌های اداری، عملی، کارآمد و قابل اجرا باشد. برای دستیابی به رویه‌ای برای تعیین کیفیت داده‌های اداری، این دو گام مهم باید برداشته شوند:

(۱) «شناسایی اجزای کیفیت» که کیفیت ورودی منابع داده‌های اداری را تعیین می‌کند.

(۲) «توسعه رویکرد کلی» برای تعیین کیفیت منابع داده‌های اداری.

گام اول، «شناسایی اجزای کیفیت» است که کیفیت ورودی منابع داده‌های اداری را تشکیل می‌دهند. این مسئله موضوعی در مورد کیفیت منابع داده‌های اداری است که در آمار توجه زیادی به آن نشده است. شناسایی اجزای کیفیت ورودی برای داده‌های اداری الزامی است. زیرا این امر سازمان‌های ملی آمار را قادر می‌سازد به سرعت -ترجیحاً بلافاصله پس از دریافت منبع داده- درباره تطابق منبع داده با نیازهایشان تصمیم‌گیری کنند. این امر، خصوصاً برای سازمان‌های ملی آماری که اخیراً شروع به استفاده از منابع داده‌های اداری کرده‌اند، اهمیت می‌یابد تا بتوانند به ریسک وابستگی بالا و بهتری، دست پیدا کنند (داس و همکاران، ۲۰۰۹). گام دوم، توسعه رویکرد کلی برای تعیین کیفیت منابع داده‌های اداری

نیز اهمیت داشته و به اولین گام وابسته است. با این حال، در کنار شناسایی اجزای کیفیت ورودی، رویکرد کاربردی کلی نیز مورد نیاز است تا بتوان قابلیت استفاده آماری یک منبع داده اداری را در فرایند آماری بررسی کرد. رویکرد دوم، نیاز به بررسی تأثیر استفاده از منبع داده اداری بر فرایند تولید و کیفیت محصول نهایی (خروجی) سازمان ملی آمار دارد. با چنین روشی، می‌توان به طور خاص به درک بهتری از مزایای استفاده از داده‌های اداری در فرایند آماری به عنوان یک خروجی جامع و نهایی، دست یافت.

تعیین کیفیت منابع داده‌های اداری

منابع داده‌های اداری می‌توانند برای بسیاری از اهداف سازمان‌های ملی آمار استفاده شوند. به عنوان مثال:

- ورودی برای چارچوب یک آمارگیری نمونه‌ای (مانند ثبت کسب و کار)،
- منبع اطلاعات کمکی، یا
- جایگزین برای داده‌هایی که به طور سنتی از طریق آمارگیری جمع‌آوری شده‌اند (Unece، ۲۰۰۷ و والگرن، ۲۰۰۷).

از این رو کیفیت یک منبع داده اداری، تنها بر پایه کاربرد آن، شکل می‌گیرد. به عنوان مثال، یک منبع ممکن است برای ارائه داده‌ها در مورد متغیر اصلی مورد مطالعه از کیفیت پایینی برخوردار باشد، اما همان منبع می‌تواند برای کاربردی دیگر اطلاعات کمکی مهمی را در برداشته باشد. این وضعیت دشواری است. از یک سو، آمارشناسان قصد دارند کیفیت منبع در حال استفاده را در اسرع وقت (ترجیحاً قبل از به کارگیری) بشناسند، اما از سوی دیگر، استفاده مورد نظر، بر ارزیابی کیفیت منبع توسط آمارشناس، تأثیر می‌گذارد. به همین دلیل، سیستم کلی و مفید برای ارزیابی کیفیت منابع داده‌های اداری نمی‌تواند ناشی از کاربردی خاص و کیفیت آمار حاصل از آن، باشد (لایتیلا و همکاران، ۲۰۱۱). بنابراین، ارزیابی کیفیت منابع داده‌های اداری باید بر موارد ذیل متمرکز باشد:

(۱) اطلاعاتی که در حال حاضر برای منبع در دسترس است.

(۲) اطلاعاتی که حاصل از تجزیه و تحلیل سیستماتیک منبع است.

با این نوع نگاه، طبیعی است که منابع داده‌های اداری به عنوان ورودی‌های یک سیستم تولید، در نظر گرفته شوند. به طور کلی، مواد اولیه را نمی‌توان مستقیماً در فرایند تولید استفاده کرد و ابتدا باید آماده شوند. کیفیت مواد اولیه و میزان آماده‌سازی آن، بر کیفیت فرایند و محصول یا خروجی تأثیرگذار است. از این رو کیفیت مواد اولیه از نظر تولیدکننده اهمیت دارد، در حالی که مصرف‌کننده تنها بر کیفیت محصول نهایی تمرکز دارد. بنابراین، کیفیت منابع داده‌های اداری باید از دو دیدگاه مصرف‌کننده آمار و تولیدکننده آمار بررسی شود. دیدگاه مصرف‌کننده (تمرکز بر کیفیت محصول نهایی یا کیفیت خروجی) دیدگاهی است که به طور سنتی مورد توجه قرار گرفته است (Eurostat^۹، ۲۰۰۹). دیدگاه تولیدکننده به دو مسئله اشاره می‌کند:

(۱) «کیفیت ورودی» - آماده‌سازی ورودی مورد نیاز برای استفاده در فرایند تولید؛ و

(۲) «کیفیت فرایند تولید» - منفعت حاصل از بهره‌وری تولید با استفاده از ورودی (لایتیلا و همکاران، ۲۰۱۱).

برای توسعه سیستم ارزیابی کیفیت منابع داده‌های اداری، هر یک از مفاهیم، باید به مجموعه مؤلفه‌هایی که جنبه‌های متفاوتی از مفاهیم کیفیت را نشان می‌دهند، تقسیم‌بندی شوند.

اولین تحقیق WP4، بر کیفیت شروع فرایند استفاده از منابع داده‌های اداری توسط سازمان‌های ملی آمار (کیفیت ورودی منابع داده‌های اداری) تمرکز داشته است. هدف، شناسایی مؤلفه‌هایی است که در نهایت، کیفیت ورودی منابع اداری را تعیین می‌کنند. این کار، سازمان ملی آمار را قادر می‌سازد تا قابلیت استفاده آماری منابع داده‌های اداری را -در سطح عمومی- قبل از به کارگیری، تعیین کند. این نوع ارزیابی کیفیت گاهی اوقات به عنوان پیش‌بینی^{۱۰} (به جای نتایج واقعی) در نظر گرفته شده است، زیرا تلاش می‌کند تا کیفیت نتایج نهایی را زودتر در فرایند، پیش‌بینی کند. مهم‌ترین شاخص‌های ورودی برای هر نوع منبع، شاخصی است که نشان دهنده کیفیت محصول نهایی (مثبت یا منفی) است. باید دانست، برای یک منبع ورودی جدید، نمی‌توان پیش‌بینی کرد کدامیک از شاخص‌ها قابلیت اطمینان بیشتری دارند، زیرا مشخص نیست منبع جدید چگونه فرایند تولید و کیفیت محصول نهایی را تحت تأثیر قرار می‌دهد. ابزار سنجش کیفیت ورودی داده‌های اداری، باید کارآمد بوده و به صرف زمان و زحمت زیاد برای تعیین کیفیت، نیاز نداشته باشد. زیرا کیفیت ورودی می‌تواند از لحاظ تئوری، هر بار که تحویل جدید منبع انجام می‌شود یا بخشی از منبع دریافت می‌شود، تعیین شود. (داس و همکاران، ۲۰۱۰).

ابعاد کیفیت منابع داده‌های اداری

هنگامی که کیفیت یک منبع داده تعیین می‌شود، لازم است همیشه دو حوزه کیفیت^{۱۱} مورد توجه قرار گیرند: فراداده^{۱۲} و داده^{۱۳} (باتینی و اسکاناچپیکو، ۲۰۰۶)

۱- کیفیت فراداده

بیشتر مواقع، کیفیت فراداده مستقل از داده‌ها مورد مطالعه قرار نمی‌گیرد، اما این روش در اداره آمار هلند به طور موفقیت آمیزی به کار رفته است. (داس و همکاران، ۲۰۰۸ و ۲۰۱۰). این اداره فهرستی برای تعیین اجزای کیفیت در حوزه فراداده، ارائه داده است. مزیت اصلی رویکرد مورد اشاره این است که اجزای کیفیت فراداده یک منبع، می‌تواند مستقل از محتوای آن تعیین شود. در نتیجه لازم نیست، هر بار که داده‌های یک منبع بررسی می‌شوند، فراداده نیز مورد بررسی قرار گیرد (داس و همکاران، ۲۰۱۰). با بهره‌گیری از چک لیست، در مجموع ۳۱ نامگر کیفیت مرتبط با فراداده برای ارزیابی یک منبع، مورد استفاده قرار گرفته است.

۲- کیفیت داده

بعد از تحقیق WP4 در مورد کیفیت ورودی داده‌های اداری، ضرورت محدود کردن برای شناسایی اجزای اصلی نامگرهای کیفیت، به دو دلیل ایجاد شد. دلیل اول کارایی است. با محدود کردن تعداد اجزای مورد مطالعه، تعداد نامگرهای کیفیت در «سیستم تولید» کاهش می‌یابد و در نتیجه زمان کمتری صرف آن می‌شود. دلیل دوم ایجاد تمرکز بر اجزای مورد مطالعه و کاهش رشد غیرقابل کنترل فهرست شاخص‌هاست. در مطالعاتی که هدف آن‌ها تهیه فهرستی از نامگرهای کیفیت است، در نهایت رسیدن به فهرست بزرگی از انواع نامگرهای کیفیت، غیر معمول نیست. (مثال‌های داس ۲۰۱۰ و فراست ۲۰۱۰ را ببینید). باید درک کرد در چنین فهرست‌های مفصلی، همه نامگرهای کیفیت، شاخص‌های واقعی کیفیت نیستند. نامگر کیفیت در واقع جزیی از کیفیت است که قابل اندازه‌گیری است، ضمن این که با یک یا چند روش امکان اندازه‌گیری آن وجود دارد (باتینی و اسکاناچپیکو، ۲۰۰۶). بنابراین باید توجه داشت که فهرست ایجاد شده، فقط به فهرستی شامل روش‌های اندازه‌گیری متفاوت، تبدیل نشود.

10 Ex-ante

11 Quality domains

12 Meta Data

13 Data domain

نتیجه‌گیری

این مقاله، اهمیت پیش‌نیازهای لازم به منظور به کارگیری داده‌های اداری برای نیل به اهداف آماری شامل فراهم ساختن دسترسی به منابع داده‌های اداری، امکان استفاده منظم از این منابع و قابلیت استفاده آماری از آن‌ها، با توجه به افزایش کاربرد داده‌های اداری (ثبت) را نشان داد. همچنین ملاک‌های تعیین کیفیت منابع داده‌های اداری با توجه به کاربرد آن و دیدگاه‌های تولیدکننده و مصرف‌کننده آمار، مشخص شد و دانستیم در ارزیابی کیفیت منبع داده‌ها، توجه به بررسی کیفیت فراداده و کیفیت داده حتی به صورت مستقل از یکدیگر، دارای اهمیت است. هرچند غالباً کیفیت فراداده، مجزا از کیفیت داده بررسی نمی‌شود. شاخص‌های ارزیابی کیفیت معرفی شده توسط WP4 از این ویژگی برخوردار هستند.

توضیحات

- 1- Socitm
- 2- Wallgren, A., Wallgren, B.
- 3- Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., Schwerer, E.
- 4- Daas, P.J.H., Ossen, S.J.L., Tennekes, M. (2010)
- 5- Frost, J-M.
- 6- Daas, P.J.H., Ossen, S.J.L., Vis-Visschers, R.J.W.M., Arends-Toth, J. (2009)
- 7- Laitila, T., Wallgren, A., Wallgren, B.
- 8- Batini, C., Scannapieco, M.
- 9- Daas, P.J.H., Arends-Tóth, J., Schouten, B., Kuijvenhoven, L. (2008)

مرجع

BLUE ENTERPRISE AND TRADE STATISTICS (BLUE-ETS) Deliverable 4.1, Dissemination level: PU, List of quality groups and indicators identified for administrative data sources.