

استانداردهای نحوه نمایش و گزارش دهی خطاهای آمارگیری

محمد رضا اناری

مرکز آمار ایران

چکیده. استانداردها و کاربست‌هایی برای نمایش خطاهای آمارگیری در گزارش‌ها و نشریه‌های آماری توسط کشورها و سازمان‌های بین‌المللی معرفی و توصیه شده است. بنای این استانداردها اصول بنیادی آمارهای رسمی سازمان ملل متحد هستند. در این اصول اهمیت ویژه‌ای به پیشگیری از تعبیر و استفاده نادرست از آمارها توسط کاربران و صراحت و شفافیت در زمینه گزارش‌دهی کاستی‌ها و مزایای آمار منتشر شده، داده شده است. ارائه خطاهای آمارگیری، در راستای افزایش قابلیت اعتماد به آمار ارائه شده انجام می‌شود. نمایش خطاهای آمارگیری همراه با برآوردها، کاربران را در تصمیم‌گیری بهتر درباره دقت و صحت برآوردها یاری می‌دهد و باعث اعتماد بیش‌تر کاربران به مراکز آماری می‌شود. در این مقاله ضمن معرفی این کاربست‌ها، مزایا و کاستی‌های آن‌ها نیز بیان شده است. سپس گزارش‌دهی و نمایش خطاهای آمارگیری در نشریه‌های مربوط به ۲ طرح آمارگیری مرکز آمار ایران بررسی شده است. در پایان نیز الگوی استاندارد برای نمایش و گزارش‌دهی خطاهای آمارگیری در نشریه‌های آماری مرکز آمار ایران ارائه شده است.

۱- مقدمه

اصول بنیادی آمارهای رسمی توسط اداره آمار سازمان ملل متحد پیشنهاد و در سال ۱۹۹۳ توسط کمیسیون آمار این سازمان تصویب شد (کمیسیون آمار سازمان ملل متحد

۱۹۹۴). در این اصول، به ویژه اصول ۱ تا ۴ اهمیت ویژه‌ای به پیشگیری از تعبیر و استفاده نادرست از آمارها توسط کاربران و صراحت و شفافیت در زمینه گزارش‌دهی کاستی‌ها و مزایای آمار منتشر شده، داده شده است. ارائه و نمایش خطاهای آمارگیری، در راستای افزایش قابلیت اعتماد به آمار ارائه شده انجام می‌شود. اطلاعاتی که همراه با جدول‌ها و نشریه‌های آماری منتشر می‌شوند از جمله فراداده‌های آماری هستند. در حوزه طرح‌های آمارگیری، فراداده‌ها طیف گسترده‌ای از اسناد و گزارش‌ها مانند گزارش‌های کیفیت، تعاریف و مفاهیم و طبقه‌بندی‌ها، روش‌های، طرح نمونه‌گیری، الگوریتم‌های ادیت، جدول‌های نهایی، نمایش خطا و... را شامل می‌شوند. ایزو- سازمان استانداردسازی جهانی- فراداده‌ها را به صورت زیر تعریف کرده است:

فراداده‌ها، اطلاعاتی درباره داده‌ها و فرایند تولید و استفاده از داده‌ها هستند.

در واقع فراداده‌ها ارتباط معنی‌دار بین تولیدکنندگان و کاربران داده‌ها را برقرار می‌سازند. فراداده‌هایی که داده‌های آماری اعم از داده‌های خرد و کلان و فرایندها و ابزارهای دخیل در تولید و کاربرد داده‌های آماری را توصیف می‌نمایند به فراداده‌های آماری موسومند. به بیان دیگر، فراداده‌های آماری داده‌هایی هستند که به توصیف جنبه‌های مختلف کیفی داده‌های آماری می‌پردازند. از جمله:

محتوا: تعاریف ارقام، جمعیت‌ها، متغیرها و...

دروستی: تشریح و توصیف انواع خطاها از جمله خطای برآورد، خطای نمونه‌گیری و خطاهای غیر نمونه‌گیری،

دسترسی: بیان این که چه داده‌های آماری در دسترس هستند، در کجا قرار دارند و چگونه می‌توان به آن‌ها دست یافت.

خطاهای آمارگیری (خطاهای نمونه‌گیری و غیر نمونه‌گیری) نیز جزئی از فراداده‌های آماری، فراداده‌های مربوط به دقت آمار، محسوب می‌شوند. نمایش خطاهای آمارگیری همراه با برآوردها، کاربران را در تصمیم‌گیری بهتر درباره دقت و صحت برآورد یاری می‌نماید و باعث اعتماد بیش‌تر کاربران به مراکز آماری می‌شود.

استانداردهای مختلفی در زمینه نمایش و گزارش‌دهی فراداده‌های آماری تهیه شده

که در ادامه به مهم‌ترین آن‌ها که در تهیه استانداردهای پیشنهادی مورد استفاده قرار گرفته‌اند، اشاره می‌شود.

دفتر سرشماری امریکا، استانداردهای حداقل درباره اطلاعات همراه با آمار حاصل از سرشماری‌ها و نمونه‌گیری‌ها در گزارش‌ها و نشریه‌های آماری را در ۱۳ بند به‌طور کلی مشخص کرده است (دفتر سرشماری ۲۰۰۶).

استانداردهایی برای تهیه و ارائه فراداده‌های آماری توسط صندوق بین‌المللی پول در چارچوب نظام انتشار داده‌های عمومی و استاندارد انتشار داده‌های خاص برای اطلاع‌رسانی از طریق شبکه جهانی طراحی شده است (صندوق بین‌المللی پول ۲۰۰۳). سازمان همکاری‌های اقتصادی و توسعه نیز، راهنمایی را برای نمایش و گزارش‌دهی داده‌ها و فراداده‌های آماری تدوین کرده است (سازمان همکاری‌های اقتصادی و توسعه ۲۰۰۶). در این راهنما علاوه بر بیان مزیت‌ها و عیب‌های کاربردهای مورد استفاده در کشورها، استانداردهای محتوایی فراداده‌های آماری در زمینه‌های مختلف مانند شاخص قیمت‌ها، نقل قول، تجدید نظر و سری‌های زمانی نیز مشخص و توصیه شده‌اند. کمیسیون آمار سازمان ملل متحد نیز در سال ۱۹۹۵ الگوهای استاندارد را برای مدل‌بندی داده‌ها و فراداده‌ها در اینترنت مشخص کرده است (کمیسیون آمار سازمان ملل متحد ۱۹۹۵ و ۲۰۰۰).

علاوه بر این، بیش‌تر مراکز آماری و مؤسسه‌های بین‌المللی، دارای دستورالعمل و راهنمای انتشار هستند. این دستورالعمل‌ها شامل نکات فنی در زمینه‌های مختلف مانند استانداردهای نمایش داده‌ها، طرح جدول‌بندی، نمودارها، قلم‌ها و... هستند که در نشریه‌های چاپی و الکترونیکی می‌توانند مورد استفاده قرار گیرند. این راهنماها عموماً به مسائلی مانند نمایش وقفه‌های سری زمانی، تجدید نظر، نحوه نمایش خطاهای نمونه‌گیری و غیر نمونه‌گیری، نقل قول و... می‌پردازند. بنا بر این، مراکز آماری باید به‌منظور حصول اطمینان از سازگاری در ارائه و نمایش داده‌ها در نشریه‌های چاپی و الکترونیکی، راهنمای انتشار را برای کارشناسان تهیه‌کننده گزارش‌ها و نشریه‌ها و مولفان داخل سازمان تهیه نمایند. پس از این مقدمه و در بخش دوم به درستی و خطای آماری پرداخته شده و توصیه‌ها و کاربردهای ارائه و نمایش اطلاعات در زمینه انواع

خطاهایی که منشأ آن‌ها فرایندها و روش‌های مورد استفاده در گردآوری اطلاعات است، به‌همراه مزایا و کاستی‌های هر یک از کاربردها شرح داده می‌شود. در بخش سوم الگوی استاندارد برای نمایش و گزارش‌دهی خطاهای آمارگیری ارائه می‌شود و در بخش آخر نحوه نمایش و گزارش‌دهی خطای کل آمارگیری در ۲ نشریه مربوط به طرح آمارگیری نیروی کار و هزینه و درآمد خانوار بررسی شده است.

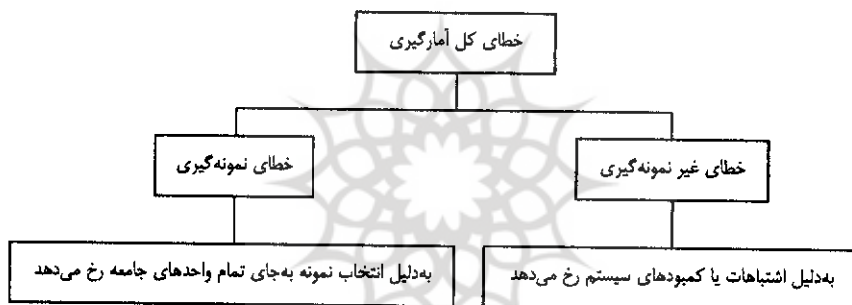
۲- درستی و خطای کل آمارگیری

هدف مراکز آماری تهیه و انتشار آمار مناسب و مرتبط با نیازهای کاربران با توجه به کیفیت است. کیفیت عمدتاً به‌صورت «مناسبت در به کارگیری» تعریف می‌شود (اداره آمار کانادا ۲۰۰۲) و شامل ابعادی در زمینه چارچوب‌های کیفیت است که در سال‌های گذشته توسط سازمان‌های ملی و بین‌المللی توسعه یافته است. سازمان استانداردسازی بین‌المللی، کیفیت را به‌صورت «مجموعه صفات و ویژگی‌های یک محصول یا خدمت که قادر به تأمین نیازهای از پیش تعیین شده و ضمنی کاربران باشند» تعریف کرده است (ایزو ۸۴۰۲، سال ۱۹۸۶). هر چارچوب کیفیتی شامل تعدادی از ابعاد است که تعریف کیفیت را پوشش می‌دهد. درستی، بعد مشترک تمام چارچوب‌های کیفیت است و به نزدیکی مقادیر محاسبه شده (ارائه شده) و مقادیر واقعی (نامعلوم) گفته می‌شود. یک جنبه از درستی عبارت است از نزدیکی مقادیر اولیه یا مقدماتی منتشر شده با مقادیر نهایی. نکته اصلی، در نزدیکی برآوردهای مقدماتی به برآوردهای بعدی است. در این زمینه در نظر گرفتن منابع و دلایل تجدید نظر، بسیار مفید است. تجدید نظر به هرگونه تغییر در مقدار آمار منتشر شده برای عموم گفته می‌شود. با وجود این، نبود تجدید نظر، لزوماً به‌معنای آمار درست نیست.

مراکز آماری همواره به‌دنبال تهیه آمارهایی با حدود خطای تا حد امکان کوچک و قابل قبول برای کاربران هستند. حدود خطا از سوی آمارشناسان با مفهیمی چون خطای استاندارد، اریبی، ضریب تغییرات و بازه اطمینان بیان می‌شود. در طرح‌های نمونه‌گیری، درستی به‌صورت فاصله بین مقدار برآورد شده و مقدار واقعی پارامتر در جامعه تعریف شده است. ارزیابی درستی یک برآورد با تحلیل خطای کل برآورد سر و کار دارد. کیش (۱۹۶۵)

درستی را به صورت وارون خطای کل تعریف کرده است. خطای کل نیز شامل آریبی و واریانس است.

کیفیت یک برآورد، رابطه تنگاتنگی با درستی برآورد دارد و درستی نیز تابعی از خطای کل آمارگیری است. بنا بر این درستی با تمام مراحل یک آمارگیری از مرحله طراحی تا انتشار مرتبط است و تنها به خطای نمونه‌گیری محدود نمی‌شود. خطای کل یک آمارگیری دارای مولفه‌هایی از خطا است که یکی ناشی از انتخاب نمونه به‌جای سرشماری تمام جامعه است که خطای نمونه‌گیری نامیده می‌شود و دیگری مربوط به خطایی است که از مرحله گردآوری تا انتشار داده‌ها رخ می‌دهد و خطای غیر نمونه‌گیری نامیده می‌شود. خطای نمونه‌گیری به‌عنوان حد پایین خطای کل آمارگیری در نظر گرفته می‌شود. نمودار ۱ این خطاها را نشان می‌دهد.



شکل ۱- مولفه‌های خطای کل آمارگیری

۱-۲- خطاهای نمونه‌گیری

خطاهای نمونه‌گیری به دلیل عدم شمارش تمام واحدهای جامعه هدف، در آمارگیری نمونه‌ای رخ می‌دهد. هنگامی که از نمونه‌گیری‌های احتمالاتی استفاده می‌شود، بسته به نوع طرح نمونه‌گیری، فرمول‌های متفاوتی برای برآورد خطای نمونه‌گیری یک برآوردگر خاص وجود دارد.

از طرف دیگر برآوردها با دو نوع از خطاها مرتبطند: آریبی و خطاهای متغیر. آریبی به خطاهایی گفته می‌شود که بر مقدار مورد انتظار برآوردهای آمارگیری مؤثرند و به صورت

تفاوت آن با مقدار واقعی پارامتر تعریف می‌شود. از طرف دیگر، خطاهای متغیر بر پراکندگی برآوردهای آمارگیری در تکرارهای ممکن فرایند انتخاب نمونه، مؤثر است. در زمینه خطاهای نمونه‌گیری، اریبی می‌تواند با اتخاذ و انتخاب روش‌های نمونه‌گیری، اندازه نمونه و روش‌های برآورد مناسب، مینیمم شود. بنا بر این پراکندگی باید به‌عنوان جنبه اصلی توزیع خطای نمونه‌گیری در نظر گرفته شود. پارامتر اصلی توصیف‌کننده این پراکندگی، خطای استاندارد، یعنی انحراف استاندارد توزیع خطای نمونه‌گیری است.

اندازه خطای نمونه‌گیری نسبت به خطاهای غیر نمونه‌گیری وقتی که برآوردهای تفصیلی‌تر یا تفکیک شده‌ای تهیه شوند، افزایش می‌یابد. محاسبه و ارائه خطای نمونه‌گیری برای تضمین استفاده مناسب از داده‌های آماری در تحلیل‌های کاربران با سطوح مختلف از آگاهی آماری، لازم است. در دسترس بودن چنین اطلاعاتی ارزیابی و انجام اصلاحات در طرح‌ها و روش‌های آماری را آسان می‌کند.

در انتشار داده‌های دارای خطای نمونه‌گیری زیاد، باید مراقبت‌های خاصی صورت پذیرد و تعادل مناسبی در تصمیم‌گیری برای انتشار چنین اطلاعاتی برای عموم برقرار شود یا در بیان آن‌ها به کاربرانی که اطلاع اندکی درباره آمار دارند، محدودیت‌هایی اعمال شود. تهیه راهبرد برای برآورد و انتشار چنین خطاهایی، برای مراکز آماری الزامی است. در حال حاضر، بسیاری از مراکز آماری با فرض آن‌که کاربران مسئول استفاده مناسب و آگاهانه از داده‌ها هستند برآوردهای دارای خطای نمونه‌گیری بزرگ را منتشر می‌کنند. این مراکز آماری مسئولیت استفاده مناسب و آگاهانه از داده‌ها را نمی‌پذیرند. با وجود این، مراکز آماری باید به ارائه هشدارها و اخطارهای مناسب در زمینه استفاده از آن‌ها متعهد باشند (راهنمای انتشار اداره آمار استرالیا).

مطالعه جامع اداره مدیریت و برنامه‌ریزی امریکا درباره کاربردهای گزارش خطا، نه تنها در بین مراکز آماری بلکه در داخل یک مرکز آماری برای آمارگیری‌های مختلف، تفاوت‌ها را بیش‌تر نمایان ساخته است. اتخاذ کاربردهای مختلف حتی برای یک آمارگیری که نتایج آن از طریق رسانه‌های اطلاع‌رسانی متفاوت، منتشر شده است به‌طور گسترده‌ای مخاطبان متفاوت با درجه گوناگونی از تخصص آماری و نیاز به اطلاعات تفصیلی درباره خطای نمونه‌گیری را منعکس می‌نماید. رئوس کلی توصیه‌های

کاربست‌های نمایش مناسب که در مطالعه امریکا مشخص شده تا حد زیادی به نوع رسانه انتشار آمار وابسته است.

۱-۱-۲- شکل‌های نمایش خطای نمونه‌گیری

در مطالعه اداره مدیریت و بودجه امریکا دو روش گسترده نمایش اطلاعات درباره درستی برآورد، شناسایی شده است. این دو روش عبارتند از:

الف) گزارش‌دهی برآورد مستقیم خطا از طریق نمایش برآورد خطای نمونه‌گیری برای تمام آماره‌های مورد استفاده در یک گزارش آماری؛

در این حالت، برای کاربران در یک زمان آماره‌ها و خطای نمونه‌گیری مربوط به آن‌ها فراهم می‌شود. محدودیت این روش، افزایش حجم نشریه است. دلیل دیگر برای عدم توصیه این است که این رویکرد، این برداشت را ایجاد می‌کند که فقط خطاهای نمونه‌گیری منبع خطا هستند یا منبع اصلی خطا به شمار می‌روند و تفسیر خطاهای نمونه‌گیری برای تمام آماره‌ها به یک اندازه مهم است.

ب) استفاده از برآوردهای غیر مستقیم خطاهای نمونه‌گیری.

گزینه‌های این روش عبارتند از:

- ارائه روش‌های غیر مستقیم برای برآورد خطاهای نمونه‌گیری مانند تابع واریانس تعمیم‌یافته به کاربران که آن‌ها را قادر نماید خطای نمونه‌گیری تقریبی را برای برآوردهای منتشر شده، محاسبه نمایند؛

عیب اصلی این روش در گزارش اداره مدیریت و بودجه به این صورت نقل شده است که روش مورد استفاده برای محاسبه خطاهای نمونه‌گیری تقریبی ممکن است مقادیری به درستی برآوردهای مستقیم، فراهم ننماید.

- ارائه خطاهای نمونه‌گیری برای منتخبی از برآوردهای کلیدی یا اصلی؛

مزیت این روش عبارت است از این که برآوردهای مستقیم خطاهای نمونه‌گیری

همراه با داده‌هاست، گرچه به تعداد اندکی از آمارها محدود است. در این حالت کاربران نیاز دارند تا خطاها را برای آماره‌هایی که خطاهای نمونه‌گیری به‌طور مستقیم برای آن‌ها تهیه نشده است، محاسبه نمایند.

اداره آمار استرالیا استفاده از برخی نمادها (مثلاً ستاره) را برای برجسته کردن آماره‌هایی که خطاهای نمونه‌گیری (یا خطاهای غیر نمونه‌گیری) بالایی دارند، توصیه نموده است. علاوه بر این، این اداره ارائه داده‌های تکمیلی و روش‌هایی برای تقریب زدن خطاهای معیار آماره‌های به دست آمده از قبیل برآوردهای تغییرات و نسبت‌های برآوردهای مختلف برای یک دوره، و تغییرات و نسبت‌ها در طول دوره‌های مختلف را توصیه می‌کند.

صرف نظر از روش مورد استفاده برای نمایش خطاهای نمونه‌گیری باید در نظر داشت که این اطلاعات ضمیمه نتایج آمارگیری هستند و بنا بر این نباید آن‌ها را به هم بریزد یا پیچیده و مبهم کند. به عبارت دیگر این عمل باید با هدف برجسته کردن قابلیت اعتماد به آماره‌ها به خدمت گرفته شود (اداره آمار سازمان ملل متحد سال ۱۹۹۳).

۲-۱-۲- کاربست‌های توصیه‌شده برای نمایش و گزارش خطای نمونه‌گیری

در راستای توجه به شفافیت و کمک به تضمین استفاده مناسب از داده‌ها، آمار حاصل از تمام آمارگیری‌های نمونه‌ای باید با اطلاعاتی درباره خطاهای نمونه‌گیری همراه باشد. چنین اطلاعاتی باید برای تمام رسانه‌های انتشاراتی، بانک‌های اطلاعاتی برخط، وب‌گاه‌ها و سایر محصولات الکترونیکی و نشریه‌های چاپی تهیه شود. این اطلاعات باید به‌صورت عبارت‌های غیر فنی قابل فهم برای کاربران غیر متخصص بیان شود. بنا بر این شیوه نمایش و میزان جزئیات تهیه شده، باید نیازهای خاص گروه‌های خاص کاربران را تأمین نماید (اداره آمار سازمان ملل متحد ۱۹۹۳). اطلاعات لازم شامل تهیه اطلاعات زیر همراه با داده‌ها یا نکات فنی واضح پیوند داده شده می‌باشد (اداره مدیریت و بودجه ۲۰۰۱):

- هشدار به کاربران درباره این واقعیت که داده‌ها از نمونه‌های تصادفی یا غیر تصادفی به دست آمده‌اند؛

- خطاهای نمونه‌گیری باید به‌عنوان یکی از منابع خطا به کاربران شناسانده شوند. باید برای کاربران از طریق تعریف خطای نمونه‌گیری و تفسیر آن، این خطا توضیح داده شود. مثلاً، به کاربران در ارتباط با غیر قابل اعتماد بودن آمارهایی که خطای نمونه‌گیری بالایی دارند، هشدارهایی داده شود؛
- خطای نمونه‌گیری باید در قالب خطای کل آماری بیان شود. در این زمینه کاربران باید از این واقعیت که خطاهای نمونه‌گیری فقط یکی از اجزای خطای کل آماری هستند و اغلب خیلی چشمگیر نیستند، آگاه شوند؛
- اگر از آزمون‌های آماری در گزارش استفاده می‌شود، سطح معنی‌داری آزمون‌های آماری باید به‌طور واضح بیان شود؛
- خطای نمونه‌گیری برای برآوردهای اصلی باید از طریق جدولی در نشریه یا از طریق پیوندهای مناسب اینترنتی در دسترس کاربران قرار گیرد. همچنین از برخی از نمادها، مستقیماً در کنار برآوردهایی که خطای نمونه‌گیری (یا غیر نمونه‌گیری بالایی) دارند، استفاده شود.

خطای نمونه‌گیری می‌تواند به یکی از روش‌های مختلف زیر بیان شود:

- مقدار مطلق، خطای استاندارد؛
- مقدار نسبی، خطای استاندارد تقسیم بر برآورد؛
- به شکل احتمال یا بازه اطمینان.

اولویت بین شکل مطلق و نسبی، به ماهیت برآورد بستگی دارد. البته برای بعضی از برآوردهای بزرگ نمونه‌ای، مقدار خطای نسبی و مطلق یکسان است. در این حالت، نمایش خطاهای استاندارد نسبی برای خوانندگان مقرون به صرفه‌تر به نظر می‌رسد. در عین حال گاهی اوقات شرح مقادیر مطلق خطای استاندارد به‌ویژه برای نسبت‌ها، درصدها و نرخ‌ها آسان‌تر است. شایان ذکر است عدم ابهام، زمانی که خطای استاندارد را بر حسب درصد بیان می‌کنیم ضروری به نظر می‌رسد. یعنی در این حالت باید تفاوت بین خطای مطلق و نسبی بیان شود.

برای تضمین سازگاری در انتشار خطای نمونه‌گیری در بین تمام خروجی‌ها و

محصولات منتشر شده یک سازمان، برخی مراکز آماری استفاده از مجموعه عبارت‌های استاندارد در نشریه‌ها را اجباری کرده‌اند. هر جا که برخی ملاحظات از لحاظ نمودن اطلاعات مفصل جلوگیری نماید، باید مرجع‌ها یا ابزار ارتباط به گزارش‌های مفصل فنی‌تر یا راهنمایی کاربران تهیه شود. چنین اطلاعاتی باید کاربران متخصص را قادر سازد تا داده‌ها را تحلیل یا جدول‌های جدید را تهیه نمایند و بنا بر این باید:

- روش‌های خاص مورد استفاده برای محاسبه خطای نمونه‌گیری در آن‌ها مشخص شود؛
- محاسبات خطای نمونه‌گیری (جدول‌ها) برای انواع مختلف برآوردها (مثلاً نرخ‌ها، درصد‌ها، نسبت‌ها، تغییرات، میانگین‌ها و میانه‌ها) برای تعدادی از متغیرها و تفکیک‌های مختلف آن‌ها تهیه شود؛
- شامل ارزیابی‌هایی از روش‌های مورد استفاده برای برآورد خطای نمونه‌گیری باشد.

۳-۱-۲- اصول نمایش خطای نمونه‌گیری

لازم است اصول خاصی برای انتخاب حالت مناسب نمایش خطای نمونه‌گیری، رعایت شود. در ادامه به چند نمونه از آن‌ها اشاره می‌شود:

الف) خطای نمونه‌گیری باید در چارچوب خطای کل آمارگیری مورد بررسی قرار گیرد. در واقع کاربر باید بداند که خطای نمونه‌گیری فقط یکی از اجزای خطای کل آمارگیری است؛

به‌عنوان مثال، خطای کل آمارگیری در یک طرح آمارگیری در کشور نپال به‌صورت زیر معرفی شده است:

برآوردهای موجود در این گزارش از یک نمونه حدوداً ۶۰۰۰ تایی از زنان در نپال به دست آمده است. البته اگر آمارگیری با یک نمونه متفاوت از زنان تکرار شود نتایج برآوردها و همچنین مقدار برآورد خطای نمونه‌گیری تغییر خواهد کرد. چون خطای نمونه‌گیری، تغییرات ناشی از محدود کردن آمارگیری به نمونه‌ای از جامعه را اندازه می‌گیرد. مزیت

مهم نمونه‌گیری احتمالاتی به آن است که می‌توان برآورد خطای نمونه‌گیری را از نتایج نمونه‌ای که واقعاً انتخاب شده است، به دست آورد. البته واضح است که خطاهای غیر نمونه‌گیری از قبیل اشتباه در اجرای طرح نمونه‌گیری، اشتباه در تکمیل پرسشنامه، جواب‌های حاصل از عدم درک صحیح سوال‌ها توسط پاسخگویان یا خطای حافظه و خطا در ثبت اطلاعات، با وجود این که مطمئناً وجود داشته‌اند به حساب نیامده‌اند. به همین دلیل است که برآورد خطای نمونه‌گیری باید به‌عنوان حد پایین خطای کل برآورد، تفسیر شود. خطای نمونه‌گیری استفاده شده در این گزارش، خطای استاندارد است. خطای استاندارد مفهوم زیر را دارد:

اگر از خطاهای غیر نمونه‌گیری چشم‌پوشی کنیم، آن‌گاه در ۲ نمونه از ۳ نمونه مقدار واقعی در محدوده یک خطای استاندارد از مقدار برآورد قرار خواهد گرفت و در ۱۹ نمونه از ۲۰ نمونه مقدار واقعی در محدوده دو خطای استاندارد از مقدار برآورد قرار خواهد گرفت. برای آماره‌های مهم در گزارش، برآورد خطای استاندارد به‌شکل پانویس و به‌وسیله یک یا چند ستاره نشان داده شده است.

(ب) حالت نمایش و میزان جزئیات باید متناسب با نیازهای گروه‌های مختلف کاربران باشد؛

(پ) نوع نمایش باید مناسب و مرتبط با نیازهای مشخص طبقه خاصی از کاربران باشد و به استفاده و تفسیر درست از اطلاعات کمک کند؛

(ت) باید به‌خاطر داشت که با وجود اهمیت خطاهای نمونه‌گیری، این خطاها در بعد دوم از اهمیت در تفسیر نتایج آمارگیری قرار دارند. بنا بر این دانستن این نکته حائز اهمیت است که اطلاعات در مورد خطاهای نمونه‌گیری مفهوم نتایج واقعی آمارگیری را مختل نمی‌کند. زیرا هدف اصلی در تهیه این اطلاعات روشن شدن میزان قابلیت اعتماد به نتایج واقعی است نه تحت تأثیر قرار دادن نتایج؛

(ث) درستی این موضوع که چه اطلاعاتی درباره خطای نمونه‌گیری مورد نیاز است بستگی به کاربردهای خاص این اطلاعات دارد. برای بسیاری مقاصد مقادیر تقریبی یا الگوهای کلی کافی هستند. همچنین مطابق بندهای (پ) و (ت) بهتر است مقادیر

تقریبی که کاربردی ترند به جای مقادیر دقیق که محاسبه آن‌ها مشکل است، تهیه شوند؛

ج) برای هر طبقه از کاربران، خلاصه کردن اطلاعات و تهیه توضیحات کافی برای تسهیل درک مفهوم درست نتایج، حائز اهمیت است. کاربران به‌طور مستقیم در طراحی و اجرای آمارگیری شرکت نمی‌کنند و نمی‌توان انتظار داشت جزئیات و خصوصیات طرز اجرای آمارگیری برای آن‌ها آشنا باشد.

برای کاربران عمومی، کاربردی‌ترین حالت، ارائه تمام برآوردهای مهم گزارش همراه با خطای نمونه‌گیری است به‌ویژه جایی که ممکن است خطا روی نتیجه نهایی حاصل از آمارگیری تأثیر داشته باشد. اما آن‌چه برای کاربران راحت‌تر به‌نظر می‌رسد شرح نتایج به صورت فاصله اطمینان است مخصوصاً زمانی که نتیجه آماری معناداری به‌وجود می‌آید.

ولی از آن‌جا که توافق همه‌جانبه در انتخاب بازه اطمینان مناسب وجود ندارد، باید اصول زیر مورد توجه قرار گیرد:

الف) تعیین این که چه بازه اطمینانی مورد استفاده قرار می‌گیرد؛
ب) تعیین یک سطح اطمینان واحد برای تمام مراحل آمارگیری.

آن‌چه بیش از همه مورد استفاده قرار می‌گیرد سطح اطمینان ۹۵ درصد است به عبارت دیگر (خطای استاندارد) ± 2 برآورد. با وجود این، مثال‌های فراوانی وجود دارد که (خطای استاندارد) \pm برآورد به‌عنوان بازه در نظر گرفته شده است.

در بسیاری مواقع کافی است که اطلاعات تقریبی درباره میزان خطای استاندارد ارائه شود. این در حالتی خواهد بود که مقادیر خطای استاندارد (یا خطای نسبی) مشابهی برای تعدادی از برآوردها وجود داشته باشد، به‌طوری که مقدار میانگین کافی باشد. به‌طور مشابه هنگامی که خطاهای نمونه‌گیری کم‌اهمیت باشد، مقادیر تقریبی کفایت می‌نمایند. در چنین مواقعی یک عبارت ساده مانند «خطای نسبی این برآوردها بین ۳ تا ۵ درصد است» می‌تواند در متن، پانویس‌ها یا جدول‌های متنی گنجانده شود.

اطلاعات نسبتاً تفصیلی‌تری می‌تواند با نشان دادن دامنه مختلفی از مقادیر خطای استاندارد یا خطای نسبی به‌وسیله نمادهای مختلف تهیه شود. برای مثال:

- خطای استاندارد نسبی کمتر از ۵ درصد است مگر این که مشخص شود؛
- خطای نسبی ۵ تا ۱۵ درصد به‌وسیله * نشان داده می‌شود؛
- خطای نسبی بیش از ۱۵ درصد با قرار دادن برآوردها در داخل پرانتز مشخص می‌شود؛

یک نوع ساده از این الگو در برخی گزارش‌های آمارگیری به کار رفته است و به‌منظور صرفه‌جویی در فضا و بهبود قابلیت خواندن متن یا جدول‌های خلاصه در این گزارش‌ها عموماً تعداد مشاهدات نمونه‌ای که برآورد بر اساس آن‌ها به دست آمده است نشان داده نمی‌شود. برای نشان دادن دامنه اندازه نمونه (مستقیماً به‌جای خطاهای نسبی) برای خانه‌ها یا جدول‌های متنی از نمادهای زیر استفاده می‌شود:

- اندازه نمونه (فراوانی هر خانه جدول) بیش از ۵۰ است مگر این که نشان داده شود؛
- در صورتی که فراوانی هر خانه جدول ۲۰ تا ۵۰ باشد، برآورد در داخل پرانتز قرار گرفته است؛
- در صورتی که فراوانی هر خانه جدول کمتر از ۲۰ باشد، برآورد متوقف شده (ارائه نشده) و با * جایگزین شده است.

شایان ذکر است که توقف ارائه داده در برخی خانه‌های جدول به‌دلیل آن است که خطای نمونه‌گیری مربوط بسیار بزرگ است (یعنی فراوانی نمونه در هر خانه جدول بسیار کوچک است). به‌طور کلی این امر یک کاربست مناسب نیست. توقف ارائه مقادیر در تک تک خانه‌ها از تلفیق رده‌های جدول‌ها برای کاربران جلوگیری می‌نماید. علاوه بر این، نتایجی که از نظر آماری، به‌دلیل خطای نمونه‌گیری بزرگ معنی‌دار نیستند، هنوز با مفهوم هستند.

۲-۲- خطاهای غیر نمونه‌گیری

خطاهای غیر نمونه‌گیری، خطاهایی در داده‌های منتشر شده هستند که نمی‌توان آن‌ها را به نوسانات نمونه‌گیری نسبت داد. خطاهای غیر نمونه‌گیری می‌توانند ناشی از تغییرات

وسیع در منابع از جمله پوشش ناقص چارچوب، کاستی‌های دستورالعمل یا راهنمای گردآوری داده‌ها یا پرسشنامه، مشکلات سیستم پردازش و مشکلات در رسیدن به نرخ پاسخ قابل قبول مربوط به هر سوال باشند.

در پیش‌نویس راهنمای سازمان ملل متحد درباره آمارگیری‌های خانواری در کشورهای در حال توسعه و گذر و گزارش استاندارد کیفیت استاندارد (اداره آمار اتحادیه اروپا ۲۰۰۳) از تفکیک جدول ۱ برای خطاهای غیر نمونه‌گیری استفاده شده است.

جدول ۱- ابعاد و طبقه‌بندی خطاهای غیر نمونه‌گیری در اداره آمار اتحادیه اروپایی

خطاهای غیر نمونه‌گیری

<p>خطاهای پوشش - حاصل از عدم تطابق بین جامعه چارچوب و جامعه هدف و شامل خطاهای کم‌پوششی، بیش‌پوششی، تکرار یا لحاظ شدن چندباره واحد آماری و اطلاعات کمکی نادرست در چارچوب (فعالیت، اندازه طبقه و آدرس و...) است.</p> <p>خطاهای بی‌پاسخی - حاصل از عدم موفقیت در به دست آوردن اطلاعات از واحدهای نمونه‌گیری در آمارگیری است. این خطاها، بی‌پاسخی کلی (واحد) وقتی که هیچ اطلاعی از واحد نمونه‌گیری به دست نیاید و بی‌پاسخی جزئی (قلم اطلاعاتی) که اطلاعات درباره برخی اما نه تمام مجموعه متغیرها از واحد نمونه‌گیری به دست نمی‌آید را شامل می‌شود.</p>	<p>خطاهای غیر مشاهده‌ای</p>
<p>خطاهای اندازه‌گیری - در طی گردآوری اطلاعات رخ می‌دهد و منجر به ثبت مقادیری متفاوت از مقادیر حقیقی متغیرها می‌شود. چنین خطاهایی ناشی از نقص ابزار آمارگیری (فرم، پرسشنامه یا مقیاس اندازه‌گیری) است که منجر به ثبت مقادیر نادرست می‌شود یا ناشی از ارائه عمدی یا غیر عمدی اطلاعات نادرست توسط پاسخگویان، تأثیر آمارگیران بر پاسخ‌های ارائه‌شده توسط پاسخگویان، است.</p> <p>خطاهای پردازش - منشأ این خطاها فرایندهای مورد استفاده پس از دریافت داده‌ها است و خطاهای کدگذاری، ورود اطلاعات، ادیت و جانپی را شامل می‌شود.</p> <p>خطاهای ناشی از فرض‌های مدل - ناشی از مدل‌های آماری مورد استفاده در مرحله برآورد است. در این‌جا خطاها ناشی از عدم انتخاب مدل مناسب، گردآوری اطلاعات مربوط به آن و برآورد پارامترهای مدل هستند.</p>	<p>خطاهای مشاهده‌ای</p>

تنوع زیاد خطاهای غیر نمونه‌گیری، تهیه اطلاعات درباره وجود ملاک‌های کمی مربوط به آن‌ها را در گزارش‌های آماری مشکل می‌سازد. به همین دلیل در اطلاعات

منتشر شده توسط مراکز آماری و سازمان‌های بین‌المللی عموماً اطلاعات مفصلی درباره خطاهای نمونه‌گیری ارائه شده و تنها به صورت کلی با عباراتی به وجود خطای غیر نمونه‌گیری پرداخته شده است. این امر می‌تواند این تفکر گمراه‌کننده را منتقل نماید که خطاهای نمونه‌گیری خیلی مهم‌تر از خطاهای غیر نمونه‌گیری هستند. راهنمای سازمان ملل درباره آمارگیری‌های خانواری در کشورهای در حال توسعه، مسئله نبود روش‌های استاندارد برای برآورد پارامترهای مولفه‌های مختلف خطاهای غیر نمونه‌گیری را مطرح کرده است. این مرجع اشاره می‌نماید که فرهنگ تشخیص اهمیت چنین خطاهایی و توجه یکسان به این منابع به اندازه خطاهای نمونه‌گیری در گزارش‌های آماری وجود ندارد.

با وجود این، برخی مراکز آماری و سازمان‌های بین‌المللی به‌ویژه اداره آمار اتحادیه اروپایی توصیه‌هایی را درباره نمایش و ارائه خطاهای غیر نمونه‌گیری و در برخی موارد چگونگی کمی کردن چنین معیارهایی فراهم کرده‌اند. اگرچه معیارها و ملاک‌های کمی برای برخی خطاهای غیر نمونه‌گیری به آسانی در دسترس قرار نمی‌گیرند اما مواردی در این زمینه مانند نرخ‌های بی‌پاسخی یا پوشش وجود دارند.

نبود ملاک‌های کمی برای تمام خطاهای غیر نمونه‌گیری، وظیفه تخصصی در سازمان‌های آماری برای ارائه حداقل برخی اطلاعات در مورد خطاهای غیر نمونه‌گیری را تغییر نمی‌دهد. بسته به ماهیت آمارگیری، اطلاعات کیفی بیانگر منابع بالقوه خطا به کاربران در تفسیر و استفاده از اطلاعات منتشر شده کمک خواهد نمود. کاربست‌هایی برای نمایش و گزارش خطاهای غیر نمونه‌گیری در زیر توصیه شده است:

- گزارش‌دهی خطاهای نمونه‌گیری برای تمام آمارگیری‌ها، تمام خروجی‌های آماری منتشر شده توسط مراکز آماری ملی و بین‌المللی باید همراه با اطلاعاتی درباره خطاهای غیر نمونه‌گیری باشد؛
- چنین اطلاعاتی باید برای تمام رسانه‌های انتشاراتی، بانک‌های اطلاعاتی برخط، وب‌گاه‌ها و سایر محصولات فنی الکترونیکی، نشریه‌های چاپی و خبرنامه‌ها تهیه شود. همچنین بسیار مهم است که این اطلاعات با واژه‌ها و عبارات‌های غیر فنی

کارآمدی که توسط کاربران غیر متخصص قابل فهم باشد، بیان شود. چنین اطلاعاتی باید همراه با اطلاعات منتشر شده باشند یا حداقل به‌طور خیلی واضح به نکات فنی ارتباط داده شود؛

- تا آن‌جا که ممکن است باید ملاک‌های کمی خطاهای غیر نمونه‌گیری تهیه شود. با وجود این، به خاطر مشکلات کمی کردن برخی خطاهای غیر نمونه‌گیری، لازم است که ترکیبی از اطلاعات کمی و کیفی توسط مراکز آماری منتشر شوند. مرحله آخر، تهیه اطلاعاتی است که کاربران غیر متخصص را قادر سازد تا به‌وضوح نقاط قوت و ضعف و محدودیت‌های موجود در اطلاعات را درک نمایند. به‌ویژه اطلاعات درباره خطاهای غیر نمونه‌گیری باید به‌وضوح به این صورت به کاربران انتقال یابد که چنین خطاهایی نسبت به خطای نمونه‌گیری، تک تک یا کلی ممکن است تأثیر بیشتری بر قابلیت اعتماد به داده‌ها داشته باشند و به آسانی در دسترس بودن اندازه‌های کمی خطاهای نمونه‌گیری لزوماً بیانگر اهمیت نسبی آن‌ها نیست.

حال، چه اطلاعاتی باید درباره خطاهای نمونه‌گیری و غیر نمونه‌گیری گزارش شود. توصیه‌های ایده‌آل برای مراکز آماری ملی و بین‌المللی انتشار اطلاعات درباره تمام خطاهای مشاهده‌ای و غیر مشاهده‌ای خلاصه شده در بالا است. گزینه مناسب دوم، اتخاذ یک روش عملی‌تر است که مراکز آماری را به استفاده از بحث‌های تخصصی و دانش تفصیلی‌تر درباره داده‌ها، برای شناسایی و تعیین زیرمجموعه‌ای از خطاهای غیر نمونه‌گیری کلیدی که تأثیر چشمگیری بر قابلیت اعتماد داده‌ها دارند، مجاب نماید. یک عامل مهم، توسعه فرهنگ ارزیابی جدی نتایج طرح‌های آماری به‌منظور مستندسازی و انتشار نقاط قوت و ضعف کلیدی است. مثال‌های مناسب از گزارش‌های استاندارد اداره آمار اتحادیه اروپایی برای ارزیابی کیفیت در آمار در جدول ۲ آمده است.

جدول ۲- مثال‌هایی از نماگرهای گزارش‌دهی خطاهای غیر نمونه‌گیری

نوع خطای غیر نمونه‌گیری	نماگر گزارش‌دهی
خطاهای پوشش	<ul style="list-style-type: none"> اطلاعاتی درباره چارچوب، دوره مرجع، عملیات بهنگام‌سازی و بازنگری کیفیت نقاط قوت و ضعف مشخص پوشش و ارزیابی کیفی از آریبی ممکن نوع و اندازه خطاهای پوشش (مثل نرخ‌های پوشش)
خطاهای بی‌پاسخی	<ul style="list-style-type: none"> نرخ‌های بی‌پاسخی روش‌های جانپی مورد استفاده (در صورت استفاده) نشانه‌هایی از اثر بی‌پاسخی‌های باقی‌مانده بر داده‌های منتشر شده، آریبی و ... نشانه‌هایی از علل بی‌پاسخی اطلاعاتی درباره عملیات و تشویق‌ها برای مینیمم کردن بی‌پاسخی
خطاهای اندازه‌گیری	<ul style="list-style-type: none"> علایم و نشانه‌هایی درباره علل خطاهای اندازه‌گیری در خروجی‌های منتشر شده و آثار آن‌ها (مثلاً موضوعاتی درباره نواقص دستورالعمل‌ها و ابزار جمع‌آوری) کاربست‌های مورد استفاده برای مینیمم کردن خطاهای اندازه‌گیری در مراحل طراحی و آزمایش هر اندازه کمی (مثل میانگین و واریانس خطای اندازه‌گیری به‌تفکیک متغیرهای مورد علاقه)
خطاهای پردازش	<ul style="list-style-type: none"> رتوس کلی فرایندهای مورد استفاده برای مینیمم کردن خطاهای پردازش (مثلاً آموزش کارکنان، روش ادیت مورد استفاده، استفاده از فرایندهای خودکار، فناوری اطلاعات و ...) نشانه‌هایی از خطاهای باقی‌مانده و تأثیر آن‌ها بر روی آماره‌ها، مثل آریبی
خطای فرض‌های مدل	<ul style="list-style-type: none"> مدل‌های مورد استفاده در تولید آماره و فرض‌هایی که این آماره بر آن‌ها متکی هستند شواهدی درباره اعتبار فرض‌ها عباراتی درباره درستی و داده‌های اضافی مورد استفاده در مدل برآورد نشانه‌هایی درباره باقی‌مانده آریبی (در نظر گرفته نشده) و ... که می‌توانند بر آماره مؤثر باشند

۳- الگوی استاندارد برای نمایش و گزارش‌دهی خطاهای آمارگیری

در این بخش با استفاده از تجربیات کشورها و سازمان‌های نام برده شده در بخش‌های

قبلی، نحوه عملیاتی کردن استانداردهای نمایش و گزارش دهی خطاهای آمارگیری بیان می‌شود. این الگو می‌تواند به‌عنوان یک استاندارد داخلی در تدوین نشریه‌های مربوط به طرح‌های آمارگیری مرکز آمار ایران مورد استفاده قرار گیرد. بدیهی است استفاده از این استاندارد مستلزم محاسبه خطاهای نمونه‌گیری و غیر نمونه‌گیری است. نکته مهمی که در تدوین این استاندارد لحاظ شده آن است که ارائه خطاهای نمونه‌گیری و غیر نمونه‌گیری به‌همراه برآوردها، حجم نشریه‌ها را قطور و زمان انتشار و اطلاع‌رسانی را افزایش می‌دهد. اما بهتر است اطلاعات تفصیلی درباره خطاهای آمارگیری در قالب گزارش‌های فنی یا کیفیت تهیه شده و بنا بر درخواست در اختیار کاربران قرار گیرد. همچنین خلاصه‌ای از این گزارش‌ها در بخش جدیدی با عنوان «کیفیت آمارگیری» به نشریه‌های نتایج طرح‌های آمارگیری اضافه شود طوری که تعداد صفحه‌های مربوط به آن باید از ۳ صفحه بیش‌تر نشود. همچنین قالب مورد استفاده برای این الگو، الگوی استاندارد مؤسسه استاندارد و تحقیقات صنعتی ایران (استاندارد ۵) است که از استانداردهای ایزو تبعیت می‌کند. شایان ذکر است که از بخش تعاریف و مفاهیم و مراجع الزامی در استاندارد بالا صرف نظر شده است.

- هدف

اطلاع‌رسانی نقاط قوت و ضعف در تولید آمار، جلوگیری از برداشت نادرست و استفاده نابجا از آمار منتشر شده و تسهیل قضاوت کاربران در زمینه صحت و دقت داده‌ها و دسترسی آن‌ها به منابع و گزارش‌های فنی و آشنایی کاربران با ملاک‌ها و معیارهای کیفیت آمارگیری.

- الگوهای استاندارد

در بخش «کیفیت آمارگیری» که در آن درباره خطاها و کیفیت طرح آمارگیری گزارش می‌شود، استانداردهای حداقل عبارتند از:

الف) مشخص شود که نمونه‌گیری احتمالاتی (تصادفی) یا نااحتمالاتی است؛

ب) ماهیت خطای کل آمارگیری اعم از خطاهای غیر نمونه‌گیری و نمونه‌گیری با عبارت‌های ساده و قابل فهم برای تمام کاربران بیان شود؛

یک الگوی مناسب به صورت زیر است:

برآوردهای طرح آمارگیری «عنوان طرح» از یک نمونه حدوداً «**» تایی از «واحدهای آماری» به دست آمده است. این نمونه‌ها به صورت تصادفی در قالب طرح نمونه‌گیری بخش «شماره بخش» انتخاب شده‌اند. البته اگر آمارگیری با یک نمونه متفاوت از «واحدهای آماری» تکرار شود نتایج برآورد و همچنین مقدار برآورد خطای نمونه‌گیری تغییر خواهد کرد. خطای نمونه‌گیری، تغییرات ناشی از محدود کردن آمارگیری به نمونه‌ای از جامعه را اندازه می‌گیرد. مزیت مهم نمونه‌گیری تصادفی آن است که می‌توان برآورد خطای نمونه‌گیری را از نتایج نمونه‌ای که واقعاً انتخاب شده است، به دست آورد. البته واضح است که خطاهای غیر نمونه‌گیری از قبیل اشتباه در اجرای طرح نمونه‌گیری، اشتباه در تکمیل پرسشنامه، جواب‌های حاصل از عدم درک صحیح سوال‌ها توسط پاسخگویان یا خطای حافظه، خطا در ثبت اطلاعات، امتناع از پاسخگویی و اشتباه محاسباتی، با وجود این که مطمئناً وجود داشته‌اند، ناچیز در نظر گرفته شده‌اند. خطاهای غیر نمونه‌گیری ناشی از اشتباهات آمارگیر، پاسخگو و ابزار و سیستم‌های نرم‌افزاری مورد استفاده در طراحی، گردآوری و پردازش داده‌ها است و می‌تواند در تمام مراحل آمارگیری بروز کند. این خطاها غیر قابل پیش‌بینی هستند. تمام برآوردها در معرض خطای نمونه‌گیری و غیر نمونه‌گیری هستند. خطای کل آمارگیری نیز ترکیبی از خطاهای نمونه‌گیری و غیر نمونه‌گیری است. لازم است که هر دو نوع خطا در سطح معقولی کنترل شوند.

پ) در صورتی که هر یک از مولفه‌های خطای کل آمارگیری محاسبه نشود، دلایل عدم محاسبه و ارائه آن ذکر شود؛

ت) در صورتی که خطاهای نمونه‌گیری محاسبه می‌شود، عبارت‌های زیر نیز در بخش «کیفیت آمارگیری» پس از توصیف خطاهای نمونه‌گیری و غیر نمونه‌گیری گنجانده شود:

- ملاک خطای نمونه‌گیری مورد استفاده در این گزارش، « خطای استاندارد » است. « خطای استاندارد » مفهوم زیر را دارد:

اگر از خطاهای غیر نمونه‌گیری صرف نظر و نمونه‌گیری نیز تکرار شود، آن گاه در ۲ تکرار از ۳ تکرار نمونه‌گیری مقدار واقعی در محدوده یک خطای استاندارد از مقدار برآورد قرار خواهد گرفت و در ۱۹ تکرار از ۲۰ تکرار نمونه‌گیری مقدار واقعی در محدوده دو خطای استاندارد از مقدار برآورد قرار خواهد گرفت.

گرچه در این طرح آمارگیری تنها خطاهای نمونه‌گیری محاسبه شده است، این امر به این مفهوم نیست که خطاهای نمونه‌گیری مهم‌تر از خطاهای غیر نمونه‌گیری هستند. به همین دلیل، برآورد خطای نمونه‌گیری باید به‌عنوان مینیمم مقدار خطای کل آمارگیری تفسیر شود.

همچنین روش محاسبه خطای نمونه‌گیری نیز به‌طور مختصر شرح داده شود.

ث) در صورتی که خطاهای نمونه‌گیری برای تک تک برآوردها در نشریه‌ای به‌صورت جداگانه محاسبه شده است، اطلاعات مربوط به نحوه دسترسی به آن در بخش « کیفیت آمارگیری » ذکر شود. الگوی پیشنهادی به‌صورت زیر است:

اطلاعات تفصیلی مربوط به خطای نمونه‌گیری برآوردها در نشریه یا گزارش فنی با نام « ... » و « در درگاه ملی آمار از طریق آدرس ... » قابل دسترسی است.

ج) برای گزارش دهی خطای نمونه‌گیری در بسیاری موارد مانند طرح‌های آمارگیری مربوط به قیمت که به دلیل اعمال کنترل‌های لازم توسط دولت، نوسانات زیادی در آن مشاهده نمی‌شود، کافی است که اطلاعات تقریبی درباره میزان خطای استاندارد ارائه شود. هنگامی که خطاهای نمونه‌گیری برای تمام برآوردها کم‌اهمیت باشند، مقادیر تقریبی یا ارائه متوسط آن کفایت می‌نماید. در چنین مواقعی کافی است که یک عبارت ساده مانند « خطای نسبی این برآوردها بین ۳ تا ۵ درصد است » باید در متن بخش « کیفیت آمارگیری » گنجانده شود؛

چ) در نشریه‌هایی که تعداد جدول‌های انتشاراتی آن‌ها زیاد است و نمی‌توان خطای نمونه‌گیری را برای تک تک خانه‌های جدول‌ها به‌روشن مستقیم برآورد کرد، یا این که

محاسبه به‌روش مستقیم طولانی و زمان‌بر است، با استفاده از روش‌های غیر مستقیم برآورد واریانس مانند توابع واریانس تعمیم‌یافته یا اثر یا ضریب‌های طرح، جدول‌های کمکی مربوط به ضریب‌ها و فرمول محاسبه در اختیار کاربران قرار داده شود تا در صورت نیاز کاربران بتوانند خطای نمونه‌گیری را محاسبه نمایند. همچنین باید مثالی عددی از نحوه محاسبه خطای نمونه‌گیری به آن‌ها ارائه شود؛

(ح) در صورتی که خطاهای غیر نمونه‌گیری محاسبه نمی‌شوند، باید برخی فعالیت‌ها که با هدف کاهش خطاهای غیر نمونه‌گیری انجام می‌شوند، ذکر شوند. الگوی مناسب در بخش «کیفیت آمارگیری» و در صورت به‌کارگیری آن‌ها، به‌صورت زیر است:

برای کاهش خطاهای غیر نمونه‌گیری مانند خطاهایی که توسط پاسخگویان، آمارگیران و کدگذاران رخ می‌دهد، از روش‌های «کنترل کیفیت» و «بازبینی» استفاده شده است. برای جبران بی‌پاسخی کلی و جزئی، روش تعدیل وزنی در محاسبه برآورد، به کار گرفته شده است.

(خ) در صورت محاسبه شاخص‌ها و نماگرهای مربوط به خطاهای غیر نمونه‌گیری، خلاصه‌ای از یافته‌ها در بخش «کیفیت آمارگیری» آورده شود. یک الگوی مناسب به‌صورت زیر است:

نرخ تحقق نمونه در طرح آمارگیری «عنوان طرح» و در کل کشور «**» درصد است. نرخ پاسخ در بین «واحدهای آماری» واجد شرایط برای پاسخگویی، «**» درصد است. نرخ بی‌پاسخی کلی برابر «**» درصد و مهم‌ترین دلیل بی‌پاسخی در این طرح، «...» با «**» درصد است. بالاترین نرخ بی‌پاسخی جزئی مربوط به قلم اطلاعاتی «...» با «**» درصد و کمترین آن مربوط به قلم اطلاعاتی «...» با «**» است. متوسط نرخ بی‌پاسخی جزئی در این آمارگیری برابر «**» درصد است.

(د) در صورتی که خطاهای غیر نمونه‌گیری برای تک تک برآوردها در نشریه‌ای به‌صورت جداگانه محاسبه شده و گزارش‌های استاندارد کیفیت آمارگیری نیز تهیه شده باشد، باید اطلاعات مربوط به نحوه دسترسی به آن‌ها در بخش «کیفیت آمارگیری» ذکر شود. الگوی پیشنهادی به‌صورت زیر است:

اطلاعات تفصیلی مربوط به خطاهای غیر نمونه‌گیری و کیفیت طرح آمارگیری در نشریه یا گزارش فنی با نام «...» و «در درگاه ملی آمار از طریق آدرس...» قابل دسترسی است.

استانداردهای تشویقی در بخش «کیفیت آمارگیری» عبارتند از:

- ارائه جدول‌ها و نمودارها برای توصیف بهتر نرخ‌های پاسخ، تحقق نمونه و بی‌پاسخی جزئی و کلی و تفکیک آن‌ها بر حسب مناطق جغرافیایی،
- توصیف میزان پوشش چارچوب نمونه‌گیری و خطاهای مربوط به آن و روش بهنگام‌سازی چارچوب، و
- توصیف مختصر روش‌های مورد استفاده برای بالا بردن نرخ پاسخ آمارگیری و ابزارهای کنترل کیفیت مورد استفاده به‌ویژه در مرحله پردازش داده‌ها.

در جدول‌های انتشاراتی استانداردهای حداقل عبارتند از:

الف) اطلاعات نسبتاً مفصل‌تری می‌تواند با نشان دادن دامنه مختلفی از مقادیر خطای استاندارد یا خطای نسبی به‌وسیله نمادهای مختلف در جدول‌ها تهیه شود. برای این منظور باید از نمادهایی استفاده شود که با نمادها و نشانه‌های استاندارد مورد استفاده در نشریه متفاوت باشد.

برای مثال:

- خطای نسبی کمتر از ۱۰ درصد است مگر این که مشخص شود؛
- خطای نسبی ۱۰ تا ۲۰ درصد با نماد ● نشان داده می‌شود؛
- خطای نسبی بیش از ۲۰ درصد با نماد ●● مشخص می‌شود.

ب) برای جدول‌هایی که خطای نسبی تمام خانه‌های آن‌ها اهمیت زیادی نداشته باشند و یا نوسان چندانی ندارند، یک عبارت ساده مانند «خطای نسبی این برآوردها بین ۳ تا ۵ درصد است» باید در پی‌نوشت جدول‌ها گنجانده شود؛

پ) در صورتی که خطای نمونه‌گیری محاسبه نمی‌شود، اندازه نمونه (مستقیماً به‌جای خطاهای نمونه‌گیری) در جدول‌ها با استفاده از نمادهای زیر نشان داده شود:

- اندازه نمونه (فراوانی هر خانه جدول) بیش از ۵۰ است مگر این که نشان داده شود؛
- در صورتی که فراوانی هر خانه جدول ۲۰ تا ۵۰ باشد، برآورد در داخل پرنتر قرار می‌گیرد؛
- در صورتی که فراوانی هر خانه جدول کمتر از ۲۰ باشد، برآورد در داخل (()) قرار داده می‌شود.

ت) در صورتی که از بازه اطمینان برای گزارش‌دهی خطای نمونه‌گیری ارقام اصلی طرح آماری استفاده شود، باید سطح اطمینان ۹۵ درصد در نظر گرفته شود. به عبارت دیگر، برآورد ± 2 (خطای استاندارد)، در جدول‌ها نمایش داده شود؛

ث) روش‌های جانپی و تعدیل وزنی، دو روش مورد استفاده برای جبران اثر بی‌پاسخی پس از گردآوری داده‌ها هستند. در صورتی که از روش‌های جانپی و یا روش‌های تعدیل وزنی برای جبران اثر بی‌پاسخی در برآوردها استفاده نمی‌شود، برآوردهایی که نرخ پاسخ آن‌ها کمتر از ۸۵ درصد است باید با استفاده از نماد خاصی در جدول‌ها مشخص شوند. این نماد و مکان آن باید به گونه‌ای انتخاب شود که با سایر نمادهای استاندارد مورد استفاده در نشریه که معنی خاصی دارند، تداخل نداشته باشند. در پی‌نوشت جدول نیز باید توضیحات لازم در زمینه عدم دخالت مشاهده‌های گمشده در برآورد داده شود. در صورت ارائه برآورد مجموع در چنین مواردی لازم است که در پی‌نوشت جدول ذکر شود:

برآورد بدون جانپی مشاهدات گمشده (بی‌پاسخ) انجام شده است. به همین دلیل، برآورد ارائه شده در معرض کم‌برآوردی برای کل جامعه است زیرا در این حالت مقدار صفر برای مشاهدات بی‌پاسخ در نظر گرفته می‌شود.

در صورت ارائه برآورد نسبی در چنین مواردی لازم است که در پی‌نوشت جدول ذکر شود:

برآورد بدون جانپی مشاهدات بی‌پاسخ انجام شده است. در این حالت برای برآورد، نسبت مشاهده شده در هر خانه جدول، برای مشاهدات بی‌پاسخ در آن خانه در نظر گرفته

می‌شود که ممکن است منجر به ناسازگاری برآوردها در بین جدول‌ها شود.

۴- نمایش خطا در نشریه‌های آماری مرکز آمار ایران

در این بخش به بررسی نحوه نمایش خطاها در نشریه‌های چاپی و الکترونیکی مرکز آمار ایران پرداخته می‌شود. بررسی نشان می‌دهد که تمامی نشریات مربوط به طرح‌های آمارگیری مرکز آمار ایران، از بدو تأسیس تا سال ۱۳۸۴ عاری از هرگونه توصیف و تشریح خطاهای غیر نمونه‌گیری بوده است. یکی از دلایل این امر، زمان‌بر بودن محاسبه این نوع خطاهاست. تا قبل از سال ۱۳۷۳ تنها به ارائه فرمول‌های برآورد واریانس (خطای نمونه‌گیری) اکتفا می‌شد. مشکل اصلی در این زمینه آن است که راهبرد مناسبی در خصوص نحوه گزارش‌دهی خطاهای آمارگیری تا سال ۱۳۸۴ اتخاذ نشده بود. علاوه بر این، به دلیل عدم استفاده از نرم‌افزارهای آماری، محاسبه خطاهای نمونه‌گیری به‌سختی انجام می‌شده است. مشکل اساسی در محاسبه خطاهای نمونه‌گیری چند منظوره بودن طرح‌های آمارگیری و پیچیدگی مراحل انتخاب واحدهای آماری در بیش‌تر طرح‌های آمارگیری مرکز آمار ایران است. با توجه به این که بیش‌تر طرح‌های آمارگیری مرکز آمار ایران، طرح‌های آمارگیری پیچیده محسوب می‌شوند لذا محاسبه واریانس آماره‌های ناخطی مثل نسبت‌ها پیچیده است. این امر سبب می‌شود که از روش‌های تکراری برای محاسبه مستقیم واریانس مانند خطی‌سازی تیلور، گروه‌های تصادفی، جک‌نایف و نیم نمونه متعادل و از روش‌های غیر مستقیم مانند تابع واریانس تعمیم‌یافته برای تعمیم آن استفاده شود.

نشریه‌های الکترونیکی نیز ناخواسته از نشریه‌های چاپی الگو می‌پذیرند و مطالبی که در نشریه‌های چاپی گزارش شده، به صورت فایل‌های متنی یا در قالب فایل‌های تصویری مانند PDF از طریق درگاه ملی آمار ارائه می‌شوند.

خوشبختانه در سال‌های اخیر تلاش‌های زیادی برای محاسبه و ارزیابی خطاها، اعم از خطاهای نمونه‌گیری و غیر نمونه‌گیری صورت پذیرفته است که در ادامه به بررسی آن‌ها در دو طرح آمارگیری مهم مرکز آمار ایران می‌پردازیم:

- طرح آمارگیری از هزینه و درآمد خانوار

طرح آمارگیری از هزینه و درآمد خانوار، جامع‌ترین طرح نمونه‌گیری مرکز آمار ایران محسوب شده و به صورت سالانه اجرا می‌شود و در آن در حدود ۷۰۰ قلم اطلاعاتی از خانوارها پرسیده شده می‌شود. این طرح، اولین طرحی است که در آن خطای نسبی میانگین برای برخی اقلام، متوسط هزینه‌های خوراکی و سالانه، هزینه‌های غیر خوراکی و متوسط درآمد سالانه محاسبه و برای ستون کل در جدول‌های مربوط منتشر شده است. محاسبه خطای نسبی در این طرح آمارگیری از سال ۱۳۷۳ آغاز شده است. مانند دیگر نشریه‌های طرح‌های آمارگیری، درباره خطاهای غیر نمونه‌گیری و منابع آن مطلبی ارائه نشده است و به کاربران القا می‌نماید که تنها خطای نمونه‌گیری برای اقلام محاسبه شده مهم است.

- طرح آمارگیری نیروی کار

طرح آمارگیری نیروی کار بر اساس آخرین استانداردها و توصیه‌های بین‌المللی، از سال ۱۳۸۳ به صورت فصلی اجرا و در آن از نمونه‌گیری چرخشی استفاده می‌شود. این طرح اولین طرح آمارگیری است که در آن از خطاهای غیر نمونه‌گیری، خطای بی‌پاسخی جزئی و کلی به تفکیک علل آن محاسبه شده و در نشریه طرح در بخشی با عنوان «وضعیت علل عدم تکمیل پرسشنامه‌ها» به آن پرداخته شده است. در بخش مذکور تنها به خطاهای غیر نمونه‌گیری پرداخته شده و مطلبی در مورد خطاهای نمونه‌گیری بیان نشده است و به کاربران القا می‌نماید که تنها خطای غیر نمونه‌گیری آن هم علل عدم تکمیل پرسشنامه‌ها اهمیت بسیاری دارد. البته محاسبه واریانس آماره‌ها در این طرح بسیار پیچیده است و پروژه‌های تحقیقاتی در این خصوص در حال انجام است.

مرجع‌ها

- [1] Australian Bureau of Statistics (2004). Australian Bureau of Statistics Publishing Manual, ABS, Canberra, unpublished.
- [2] Eurostat(2002), "Standard Quality Report", discussion item 4, Assessment of Quality in Statistics, fifth meeting, Luxemburg.

- [3] International Monetary Fund (2003), "SDDS Overview", Special Data Dissemination Standards, IMF, Washington DC, Available from <http://www.dsbm.imf.org/Applications/web/overview/> [Accessed 16 August 2006].
- [4] International Organization for Standardization/International Electrotechnical Commission, n.d., ISO/IEC 11179, Information Technology-Metadata Registries(MDR), published on the Joint Technical Committee 1 (JTC1) Information Technology website, Available at <http://metadata-stds.org/> [Accessed 7 August 2006].
- [5] Kish, L.(1965), Survey Sampling ,Wiley, New York.
- [6] National Center for Education Statistics (2003). NCES Statistical Standards, National Center for Education Statistics, U.S Department of Education.
- [7] Organization for Economic Cooperation and Development (2006). Data and Metadata Reporting and presentation Handbook, OECD, Paris Available at <http://www.oecd.org/Long abstract>.
- [8] Statistics Canada (2000). Policy on Informing Users of Data Quality and Methodology, StatisticsCanada, Ottawa, available from <http://www.statcan.ca/english/about/policy/infousers.htm#standards>.
- [9] United Nations Statistical Commission and United Nations Economic Commission for Europe (1995), Guideline for Modeling Statistical Data and Metadata, Conference of European Statistician Statistical Standards and Studies-, United Nations, Geneva.
- [10] United Nations Statistical Commission and United Nations Economic Commission for Europe (2000), Guideline for Statistical Metadata on the Internet, Conference of European Statistician Statistical Standards and Studies-No. 52, United Nations, Geneva, Available at <http://www.unecec.org/stats/publicatiopns/metadata.pdf> [Accessed 16 August 2006].
- [11] United Nations Statistical Commission (1994), Fundamental Principles of Official Statistics, UNSD, New York, Available from <http://unstats.un.org/unsd/goodprac/bpabout.asp> [accessed 16 August 2006].
- [12] United Nations Statistical Division (1964). Recommendations for the Preparation of Sample Survey Reports (Provisional Issue), Statistical Papers, Series C, No. 1 Rev. 2, United Nations, New York.
- [13] United Nations Statistical Division (1982). Non-sampling Errors in Household Surveys: Sources, Assessment and Control, United Nations, New York.
- [14] United Nations Statistical Division (1993). Sampling Errors in Household Surveys, UN Department for Economic and Social Information and Policy Analysis, Statistical Division, New York. Available at <http://unstats.un.org/unsd/pubs/gesgrid.asp?ID=316> [Accessed 16 August 2006].
- [15] US Office of Management and Budget (2001). Measuring and Reporting Sources of Error in Surveys, Statistical Policy Working Paper 31, Statistical

Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, Washington DC, available from http://www.fcs.m.gov/01papers/SPWP31_final.pdf.

[16] US Census Bureau (2006). Census Bureau Standard: Minimal Information to Accompany Any Report of Survey or Census Data. Available from www.census.gov/quality/S12-0_v1.2_Minimal_Info.pdf.

[۱۷] مؤسسه استاندارد و تحقیقات صنعتی ایران، استاندارد ملی ایران شماره ۵۰، مقررات مربوط به ساختار و شیوه نگارش.



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



ثرويشگاه علوم انسانی و مطالعات فرهنگي
پرتال جامع علوم انسانی