

# پارادیتا چیست و چگونه می‌توان با استفاده از آن خطاهای

## آمارگیری را کاهش داد؟



لیدا کلهری، هیئده حجت‌الاسلامی و امیر الفت

### \* مقدمه

پارادیتا اطلاعاتی در خصوص فرایند گردآوری داده‌ها در آمارگیری‌ها هستند که از طریق ابزار گردآوری داده‌ها (پرسش‌نامه‌ی کاغذی، تبلت یا ...) به دست می‌آیند. این اطلاعات می‌توانند با هدف مدیریت کارآتر آمارگیری یا بهبود کیفیت برآوردها، در مراحل مختلف فرایند آمارگیری از جمله تعیین واحدهای نمونه‌ای که در ایجاد ارتباط یا همکاری با آنها مشکلاتی رخ می‌دهد و مشخص کردن واحدهایی از نمونه که به پیگیری‌های بیش‌تر یا راهبردی متفاوت برای مراجعه نیاز دارند، مورد استفاده قرار گیرند. برخی از انواع پارادیتا هم برای واحدهای پاسخ‌گو و هم برای واحدهای بی‌پاسخ در دسترس است که در این صورت می‌توان از آنها برای مدل‌بندی بی‌پاسخی و کسب اطلاعاتی برای کاهش بی‌پاسخی نیز استفاده کرد. همچنین استفاده از پارادیتا می‌تواند در تعدیل آریبی بی‌پاسخی و ارزیابی خطای اندازه‌گیری مفید باشد. این داده‌ها همچنین برای ارزیابی عملکرد آمارگیران قابل استفاده هستند. در سال‌های

اخیر برای ارزیابی کیفیت داده‌های گردآوری‌شده در بررسی‌های روش‌شناختی آمارگیری به‌طور فزاینده‌ای از پارادیتا استفاده شده است. در این مقاله به ارائه‌ی مفهوم پارادیتا و انواع آن می‌پردازیم و علاوه بر آن به برخی از کاربردهای پارادیتا اشاره خواهیم کرد.

### \* تعریف و انواع پارادیتا

کیفیت داده‌ها و نتایج به‌دست آمده از آمارگیری‌ها و روش‌های بهبود کیفیت از موضوعات پرچالشی هستند که پرداختن به آنها از اهمیت بالایی برخوردار است. از این رو آمارشناسان همواره به دنبال روش‌ها و ابزارهایی هستند که با استفاده از آنها بتوان خطاهای آمارگیری را کاهش داد و کیفیت نتایج را افزایش داد. پارادیتا از ابزارهایی است که برای کشف منابع خطا در آمارگیری‌ها به کار می‌رود. کوپر (۱۹۹۸) اولین شخصی بود که اصطلاح «پارادیتا» را در حوزه‌ی روش‌شناسی آمارگیری معرفی کرد. او ابتدا این اصطلاح را برای اشاره به داده‌های فرایندهای خودکار ارائه کرد. از آن به

بعد، این اصطلاح به همه‌ی انواع داده‌های مرتبط با فرایند گردآوری داده‌ها مانند مکالمه‌ی ضبط‌شده‌ی آمارگیر با پاسخگو، مدت زمان مصاحبه و مشخصات آمارگیر توسعه یافت. کروتز (۲۰۱۳) بیان می‌کند پارادیتا، داده‌های بیش‌تری است که در طول فرایند اجرای آمارگیری‌ها می‌توان گردآوری کرد. او همچنین پارادیتا را نوعی داده‌ی کمکی معرفی می‌کند که از منابع داخلی آمارگیری به‌دست می‌آید. گردآوری پارادیتای مربوط به آمارگیری، موضوع جدیدی نیست اما دامنه و جزئیات پارادیتایی که گردآوری می‌شود، در سال‌های اخیر به‌طور اساسی افزایش یافته است، که البته بیش‌تر به دلیل کامپیوتری شدن فرایندهای مختلف در آمارگیری‌ها است و تنها به بخش گردآوری داده‌های پرسشنامه محدود نمی‌شود. به‌طور کلی، آگاهی در مورد فواید بالقوه استفاده از پارادیتا برای از میان برداشتن برخی چالش‌ها از قبیل، کاهش نرخ پاسخگویی، افزایش مخاطره آریبی ناشی از بی‌پاسخی و خطای اندازه‌گیری و هزینه‌های فزاینده‌ی گردآوری داده، در میان تولیدکنندگان و استفاده‌کنندگان آمارها رو به افزایش است.

لیدا کلهری دکترای آمار و عضو هیئت علمی پژوهشکده‌ی آمار است.

هیئده حجت‌الاسلامی کارشناس دفتر صنعت، معدن و زیربنایی مرکز آمار ایران است.

امیر الفت کارشناس بیمه‌های درمان بیمه‌ی البرز است.

در این بخش، برخی از انواع پارادیتا شامل مشخصات آمارگیرها، اطلاعات ثبت شده از تماس با واحدها، مشاهدات آمارگیرها درباره‌ی ناحیه و مشخصات واحدهای نمونه، تعامل و اقدامات متقابل در شروع آمارگیری، مسیرهای ممیزی و صداهای ضبط شده معرفی می‌شوند، البته باید توجه داشت که امکان گردآوری تمام انواع پارادیتا در همه‌ی آمارگیری‌ها وجود ندارد، و نوع آمارگیری در تعیین پارادیتای قابل حصول، مؤثر است. به‌عنوان مثال، مشاهدات آمارگیر درباره‌ی ناحیه و واحدهای نمونه، مربوط به آمارگیری‌هایی است که با مراجعه‌ی حضوری آمارگیر انجام می‌شوند و در آمارگیری‌های تلفنی یا اینترنتی امکان دستیابی به این پارادیتا فراهم نیست. مطالب این بخش برای آمارگیری‌های خانواری ارائه شده‌اند و اغلب به راحتی قابل تعمیم به انواع دیگر آمارگیری‌ها هستند.

## ۱- مشخصات آمارگیرها

در آمارگیری‌هایی که اطلاعات توسط آمارگیر جمع‌آوری می‌شود، آمارگیر نقش مهمی در افزایش میزان پاسخ به پرسش‌های آمارگیری و کاهش خطای اندازه‌گیری دارد. همچنین بخشی از تفاوت بین پاسخ ارائه‌شده توسط پاسخ‌گو و آنچه پژوهشگر یا کاربر اصلی نتایج واقعاً به دنبال آن است، می‌تواند در اثر خصوصیات فردی آمارگیر ایجاد شود. مهارت آمارگیران در مدیریت مصاحبه با یکدیگر تفاوت دارد، از طرف دیگر مهارت آمارگیران در درک مفهوم سؤالات و انتقال صحیح آن به پاسخ‌گویان نیز یکسان نیست. علاوه بر آن، برخی از ویژگی‌های فردی آمارگیر مانند جنس، سن، نژاد یا زبان در آمارگیری‌هایی که سؤالات حساس دارند، در کسب اطلاعات در برخی مناطق با فرهنگ یا زبان خاص اثر دارد. بنا بر این کسب برخی اطلاعات در مورد ویژگی‌های آمارگیر و تحلیل آن به فهم

بهبتر عوامل مؤثر بر این تغییرات در بین آمارگیران کمک می‌کند، و می‌تواند برای بهبود فرایند جذب، آموزش و نظارت بر آمارگیران مفید باشد.

## ۲- اطلاعات ثبت شده از تماس با واحدها

اطلاعات درباره‌ی مراجعه یا تماس با خانوار می‌تواند نشان‌دهنده‌ی تلاش انجام‌شده توسط آمارگیر برای تکمیل پرسشنامه‌ی واحد نمونه‌ای باشد. این اطلاعات شامل زمان مراجعه، تعداد مراجعات، نتیجه‌ی مراجعات و نام و نام خانوادگی یا کد شناسایی آمارگیر یا آمارگیرانی است که به خانوار مراجعه کرده‌اند.

هدف اصلی از ثبت اطلاعات تماس، رصد کردن نحوه تصمیم‌گیری‌های آمارگیر در عملیات میدانی و تأثیر آن بر میزان مشارکت افراد است که برای مدیریت آمارگیری‌ها با هدف افزایش نرخ پاسخ‌گویی مورد استفاده قرار می‌گیرد. به‌عنوان مثال، نشان می‌دهد که در چه زمان‌هایی از روز افراد تمایل بیشتری به پاسخ‌گویی دارند، که ممکن است با توجه به نوع آمارگیری و پاسخ‌گویان هدف، متفاوت باشد. ثبت اطلاعات تماس برای تمام واحدهای نمونه، اعم از پاسخگو و بی‌پاسخ، امکان‌پذیر است. اطلاعات مربوط به واحدهای بی‌پاسخ برای تحلیل بی‌پاسخی مفیدند، بنا بر این اطلاعات آن‌ها به اندازه‌ی اطلاعات پاسخگویان ارزشمند است.

## ۳- مشاهدات آمارگیر درباره‌ی ناحیه، محل سکونت و مشخصات واحدهای نمونه

مشاهدات ناحیه برای ارزیابی تأثیر شرایط ناحیه یا محله بر پیامدهای مهم وضعیت اجتماعی و اقتصادی گردآوری می‌شوند. در دسترس بودن شواهد برای تمام واحدهای نمونه، فرصت را برای کشف این‌که آیا موارد بی‌پاسخ به صورت معنی‌داری با موارد

پاسخگو متفاوت هستند، به وجود می‌آورد. برای مؤثر بودن این کار، مشاهدات باید قویاً با احتمال پاسخگویی واحد نمونه و متغیرهای مورد نظر آمارگیری مرتبط باشند.

مشاهدات واحد مسکونی معمولاً شامل ارزیابی کیفیت واحد مسکونی در مقایسه با سایر واحدها در یک محله است. به‌عنوان مثال این‌که، واحد مسکونی در یک ساختمان چند واحدی قرار دارد، آیا موانع دسترسی وجود دارد و این‌که شرایط آن ایمن است یا خیر. شرایط واحد مسکونی در مقایسه با سایر واحدها در منطقه، به طور قابل توجهی با نرخ غیبت خانوار و امتناع از پاسخگویی مرتبط است (دورانت و دیگران، ۲۰۰۹).

مشاهدات آمارگیر از اعضای ساکن در واحدهای مسکونی نمونه، به دو رده تقسیم می‌شوند: مشاهدات در مورد ویژگی‌های جمعیتی افراد و مشاهداتی که نماینده‌ی متغیرهای کلیدی آمارگیری هستند. خصوصیات جمعیت‌شناسی که در آمارگیری‌ها ثبت می‌شوند، اطلاعاتی در خصوص افراد و زیرمجموعه‌هایی که حضورشان در نمونه اهمیت دارد فراهم می‌کنند. به‌عنوان مثال، گروه سنی، نژاد یا زبان افراد می‌تواند بر پاسخگویی آن‌ها اثرگذار باشد. باید توجه داشت که شناسایی شواهدی که به صورت تنگاتنگ با شانس پاسخگویی و متغیرهای مورد نظر آمارگیری مرتبط باشند و همچنین اندازه‌گیری آن‌ها حائز اهمیت است.

## ۴- تعامل و اقدامات متقابل در شروع آمارگیری

گام بسیار مهم در جلب همکاری اعضای نمونه، تعامل متقابل بین آمارگیر و عضو نمونه‌ی کاندید پاسخگویی است. تلاش‌هایی برای گردآوری اطلاعات خلاصه‌ای از این تعامل‌ها در برخی آمارگیری‌ها صورت گرفته است. این اطلاعات، برای ایجاد تمایز بین

مشخصه‌های واحدهای نمونه‌ای که در آمارگیری شرکت کرده‌اند و نمونه‌هایی که در نهایت در آمارگیری شرکت نکرده‌اند و برای پیش‌گویی این‌که کدام مخاطب از بین واحدهای نمونه شرکت در آمارگیری را رد خواهد کرد، به‌کار گرفته می‌شود. به‌عنوان مثال در کشور انگلستان در تعدادی از مطالعات، داده‌هایی درباره‌ی تعامل آمارگیر با واحد نمونه گردآوری می‌شود. آمارگیران در این آمارگیری‌ها ارزیابی خود درباره‌ی میزان مقاومت اولیه‌ی مخاطبین در همکاری را تحت عناوینی مانند هیچ، خیلی کم، متوسط و قوی ارائه می‌کنند.

#### ۵- مسیرهای ممیزی

فایل‌های مسیرهای ممیزی حاوی اطلاعاتی هستند که با استفاده از آن‌ها فرایند گردآوری داده‌ها قابل تمیز دادن است. در حال حاضر، در برخی از آمارگیری‌های بزرگ از سیستم‌های نرم‌افزاری مصاحبه‌ی رایانه‌یار (CAI) استفاده می‌شود؛ که این سیستم در کنار ثبت داده‌های پرسشنامه، دامنه‌ی گسترده‌ای پارادیتا هم تولید می‌کند. این فایل‌ها طیف وسیعی از اطلاعات را راجع به تمام اقلام نمونه‌ای فراهم می‌کنند که می‌تواند ابزارهای تحقیق، عملکرد آمارگیر و کیفیت داده را ارزیابی کند. به‌عنوان مثال، اگر در یک حوزه‌ی آمارگیری بیش‌تر از حد معمول از دکمه‌ی Help استفاده شود، یعنی مشکلی در رابطه با آن سؤال یا آموزش آن وجود دارد.

#### ۶- صداهای ضبط شده

سیستم‌های نرم‌افزاری پیشرفته CAI همچنین امکان ضبط صدا در طول مصاحبه را فراهم می‌کنند که به آن مصاحبه‌ی ضبط‌شده‌ی رایانه‌یار (CARI) اطلاق می‌شود. با رضایت مخاطبین، می‌توان کل مصاحبه، بخشی از مصاحبه یا یک بخش تصادفی از مصاحبه را ضبط کرد. در آمارگیری‌های تلفنی، سرپرستان

آمارگیری می‌توانند به صورت زنده مصاحبه آمارگیر را گوش کنند (نظارت صامت). CARI این فرصت را برای مدیران و روش‌شناسان آمارگیری فراهم می‌کند که به تعامل بین آمارگیر و مخاطب گوش دهند و اطلاعات مورد نیاز برای ارزیابی ابزارها، آمارگیر و کیفیت داده‌ها را به‌دست آورند.

#### \* کاربردهای پارادیتا

پارادیتا می‌تواند توسط روش‌شناسان آمارگیری در بحث قابلیت اطمینان، پاسخ‌گو بودن به سؤال‌ها و قابل تعمیم بودن دانش راه و روش آمارگیری استفاده شود. هدف کلی استفاده از این دانش، کمینه‌کردن خطای آمارگیری است. تاکنون تمرکز اصلی پژوهش‌های روش‌شناسی که از پارادیتا بهره می‌برند، بر روی خطای بی‌پاسخی بوده است. برای مثال، تاریخچه‌ی طولانی از تحلیل ثبت اطلاعات مربوط به تماس‌ها برای دستیابی به دانش الگوی مناسب تماس در آمارگیری‌های چهره به چهره وجود دارد. روش‌شناسان آمارگیری همچنین به ارزش استفاده از پارادیتا برای مطالعه‌ی اریبی ناشی از بی‌پاسخی پی‌برده‌اند. به‌ویژه پارادیتایی که برای همه‌ی موارد نمونه‌گیری شده در دسترس باشد و با مشارکت پاسخ‌گویان و متغیرهای مورد نظر آمارگیری، همبستگی داشته باشد. پارادیتای حاصل از ثبت اطلاعات تماس‌ها که معمولاً شامل زمان تماس (روز و زمان) و نیز نتیجه‌ی تماس (عدم برقراری تماس، امتناع از پاسخگویی، غیر واجد شرایط بودن واحد نمونه، انجام مصاحبه، تعیین وقت مصاحبه و غیره) هستند، اغلب برای ارزیابی و کاهش خطای بی‌پاسخی مورد استفاده قرار می‌گیرند. داده‌های ثبت اطلاعات تماس برای پاسخ‌گویان و غیر پاسخ‌گویان در یک آمارگیری موجود هستند و در نتیجه، کاندیدهای اولیه برای مطالعه‌ی اریبی

بی‌پاسخی به شمار می‌روند (نیکولاس ۲۰۱۱).

روش‌شناسان آمارگیری، از پارادیتا برای کاوش در مورد منابع ایجاد خطای اندازه‌گیری نیز استفاده می‌کنند (نیکولاس ۲۰۱۱). برای مثال، مدت زمان تکمیل پرسشنامه، فایل مسیرهای ممیزی و ضبط صداها می‌تواند مشکلات پاسخ‌گو در هنگام پاسخ به سؤالات را مشخص کند. پارادیتا همچنین برای آزمودن اثر شخصیت آمارگیر در اندازه‌گیری، مورد استفاده قرار می‌گیرد. علاوه بر خطای بی‌پاسخی و خطای اندازه‌گیری در صورت امکان این پتانسیل وجود دارد که از پارادیتا برای ارزیابی سایر منابع خطا از جمله خطای پردازش نیز استفاده شود.

پارادیتا علاوه بر آن که برای کاهش خطای آمارگیری از طریق درک بهتر دلایل بروز خطاها و بهبود مدیریت در گردآوری داده‌ها قابل استفاده است، بعد از گردآوری داده‌ها برای تعدیل خطای برآوردهای آمارگیری نیز به کار می‌آید.

تعدیل بی‌پاسخی برای داده‌های حاصل از اغلب آمارگیری‌هایی که در مقیاس بزرگ اجرا می‌شوند، انجام می‌شود. به‌طور کلی، وزن‌های نمونه‌گیری اغلب بر اساس اطلاعات حاصل از داده‌های کمکی، تعدیل می‌شوند. به‌عنوان مثال، اطلاعات چارچوب نمونه‌گیری، داده‌های سرشماری یا منابع جمعیت‌شناسی معتبر برای تعدیل اختلاف مشخصه‌های جمعیت‌شناسی بین جمعیت مورد نظر در کل جامعه و نمونه مورد استفاده قرار می‌گیرند. داده‌های آمارگیری اغلب برای تعدیل اثرات بی‌پاسخی، وزن‌دهی مجدد می‌شوند. با این وجود اگر در تعدیل وزن برای بی‌پاسخی، مشخصه‌های جمعیت‌شناسی به کار گرفته شده برای تعدیل وزن با متغیرهای مورد نظر در آمارگیری و میزان تمایل به پاسخگویی مرتبط نباشند، تعدیل وزن اریبی ناشی از بی‌پاسخی

را کاهش نمی‌دهد.

با توجه به این محدودیت، در حال حاضر توجه آمارشناسان به سمت استفاده از پارادیتا برای تعدیل خطای بی‌پاسخی جلب شده است، اما چالشی که هم‌اکنون روش‌شناسان آماری با آن رو به رو هستند، شناسایی اقلامی از پارادیتا است که با تمایل به مشارکت و متغیرهای مورد نظر آماری، همبستگی داشته باشد. این امر ممکن است نیازمند شناسایی دامنه‌ای از اقلام پارادیتای مورد استفاده برای تعدیل‌های بی‌پاسخی باشد و در نظر گرفتن یک تعدیل برای همه‌ی برآوردها امکان‌پذیر نباشد.

به‌کارگیری پارادیتا در تعدیل خطای اندازه‌گیری، کم‌تر از کاربرد آن در تعدیل خطای بی‌پاسخی است. با این حال، شواهدی مبنی بر این‌که پارادیتا در سطح اقلام می‌تواند اطلاعات مفیدی را در مورد فرایند پرسش و پاسخ فراهم کند، وجود دارد. این امید وجود دارد که پارادیتای در سطح قلم به همان خوبی که در مدل‌بندی آماری برای تحلیل داده‌ها مورد استفاده قرار گرفته است بتواند برای تعدیل خطای اندازه‌گیری هم استفاده شود. برای دستیابی به این امر و درک بهتر ارتباط بین پارادیتا و خطای اندازه‌گیری لازم است مطالعات بیشتری انجام شود.

### \* نتیجه‌گیری

پارادیتا پتانسیل تبدیل شدن به متغیر کمکی قدرتمند برای تعدیل بی‌پاسخی و کنترل سایر خطاهای آماری را دارد. دخیل بودن آمارگیران در فرایند گردآوری پارادیتا، خود یک چالش اضافی ایجاد می‌کند. آمارگیران اغلب برای گردآوری پارادیتا آموزش ندیده‌اند و کار اولیه‌ی آن‌ها کاملاً متفاوت و تقریباً در تقابل با کار گردآوری پارادیتا است. درحالی‌که برای گردآوری داده در آماری، از آماری می‌خواهیم اجازه ندهد نظرات و مشاهدات

شخصی خود بر مصاحبه تأثیر بگذارد، در گردآوری پارادیتا، گاهی به دنبال قضاوت آماری هستیم. گردآوری داده‌های مشاهده‌ای باید برای آماری ساده‌سازی شود. همچنین این نوع سنجش‌ها باید با همان دقتی که در مورد اقلام گردآوری داده‌های اولیه به کار گرفته می‌شود، انجام پذیرد. توجه به این نکته دارای اهمیت است که تکمیل اطلاعات اضافی علاوه بر اقلام مورد نیاز در پرسشنامه، می‌تواند حجم کار آماری را به‌صورت چشم‌گیری افزایش و در نتیجه کیفیت داده‌های اصلی را کاهش دهد. بنا بر این، باید دقت کرد که در هر آماری، مواردی از پارادیتا که از اهمیت بیش‌تری برخوردارند و گردآوری آن‌ها ضرورت دارد به صورت دقیق تعیین شوند. در آماری‌های رایانه‌ای که پارادیتا به‌صورت اتوماتیک گردآوری می‌شود، ضمن افزایش کیفیت آن، حجم کار آماری یا پاسخ‌گو تغییر نمی‌کند و گردآوری پارادیتا اثری بر کیفیت داده‌های آماری ندارد. استفاده از تبلت در آماری‌های اخیر نیز زمینه‌ی استفاده بیش‌تر از پارادیتا را فراهم کرده است. البته برخی از اقلام موجود در پرسشنامه‌ها مانند تاریخ تکمیل فرم یا علل عدم همکاری در حال حاضر در پرسشنامه‌ها وجود دارند. همچنین توجه کردن به سؤالاتی که ادیت یا جانهای بیش‌تری برای آن‌ها انجام می‌شود نیز می‌تواند به‌عنوان سرنخی برای تشخیص وجود کاستی‌ها یا سؤالات حساس در پرسشنامه باشد که در طراحی‌های بعدی باید بیش‌تر مورد توجه قرار گیرد. پارادیتای به‌دست آمده از آماری‌ها بهتر است به‌صورت مجزا از داده‌ها ذخیره شود ولی امکان برقراری پیوند بین آن‌ها و مجموعه داده‌ها وجود داشته باشد. همچنین گردآوری اطلاعات در مورد حجم کار آماری و استفاده از آن در بهینه‌سازی شرح وظایف آن‌ها نیز می‌تواند بر میزان پاسخ‌گویی افراد از

طریق ایجاد شرایط مناسب برای آماری مؤثر باشد.

### توضیحات

<sup>۱</sup> Audiot Trials

<sup>۲</sup> Computer Assisted Interview

<sup>۳</sup> Computer Assisted Recorded Interview

<sup>۴</sup> Silent Monitoring

<sup>۵</sup> Item-Level

### منابع

- کلهری، لیدا؛ حجت‌الاسلامی، هایده؛ الفت، امیر؛ (۱۳۹۵). پارادیتا و کاربرد آن در بهبود نتایج آماری‌ها، پژوهشکده‌ی آمار.

- Couper, M. (1998): Measuring survey quality in a CASIC environment. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

- Durrant, G., D'Arrigo, J., and Steele, F.(2009). Using field process data to predict best times of contact conditioning on household and interviewer influences. Technical report, Southampton Statistical Science Research Institute.

- Kreuter, F. (2013). Improving Surveys with Paradata: Analytic Uses of Process Information, Wiley Series in Survey Methodology

- Nicolaas, G., (2011). ESRC National Centre for Research Methods Review Paper, Survey Paradata: A Review. National Centre for Research Methods. NCRM/017.