

پارادیتا و کاربرد آن در بهبود نتایج آمارگیری‌ها

لیدا کلهری (مجری)

هایده حجت‌الاسلامی

امیر الفت



پژوهشکده‌ی آمار

گروه پژوهشی طرح‌های فنی و روش‌های آمار

پاییز ۱۳۹۵

پیش‌گفتار

پارادیتا اطلاعاتی در خصوص فرایند گردآوری داده‌ها در آمارگیری‌ها هستند که از طریق ابزار گردآوری داده‌ها به دست می‌آیند. این اطلاعات، به‌عنوان مثال می‌توانند برای تعیین مواردی که در ایجاد ارتباط یا همکاری با واحدها با مشکلاتی مواجه بوده و مشخص کردن واحدهایی از نمونه که به پیگیری‌های بیشتر یا راهبردی متفاوت برای مراجعه نیاز دارند، مورد استفاده قرار گیرند. گاهی پارادیتا هم برای پاسخ‌گویی و هم برای بی‌پاسخ‌ها در دسترس است که در این صورت می‌توان از آن برای مدل‌بندی بی‌پاسخی و کسب اطلاعاتی برای کاهش بی‌پاسخی، که مشکلی در حال رشد در آمارگیری‌ها است، استفاده کرد. این داده‌ها همچنین برای ارزیابی عملکرد آمارگیران مفیدند. از مزایای دیگر استفاده از پارادیتا بهبود کیفیت داده‌های آمارگیری (به‌عنوان مثال با انجام تعدیل برای اریبی بی‌پاسخی‌ها و ارزیابی خطای اندازه‌گیری در برآوردهای آمارگیری) است.

پارادیتا قابلیت گردآوری در هر نوع آمارگیری را دارد، گرچه مقدار، نوع و سطح جزئیاتی که با پارادیتا به دست می‌آید بسته به این که داده‌ها به‌صورت دستی یا به‌صورت خودکار با استفاده از نرم‌افزارهای کامپیوتری ثبت شده باشند، متفاوت است. در سال‌های اخیر برای ارزیابی کیفیت داده‌های گردآوری‌شده در بررسی‌های روش‌شناختی آمارگیری به‌طور فزاینده‌ای از پارادیتا استفاده شده است.

بر این اساس پژوهشکده‌ی آمار انجام طرح «پارادیتا و کاربرد آن در بهبود نتایج آمارگیری‌ها» را در دستور کار خود قرار داده است. در این طرح انواع پارادیتا و روش‌های گردآوری آن‌ها مورد مطالعه قرار خواهد گرفت. همچنین، بر اساس روش‌های مختلف گردآوری داده‌ها به گردآوری پارادیتا در پرسش‌نامه‌ی آمارگیری‌ها به‌ویژه آمارگیری‌های مرکز آمار ایران پرداخته خواهد شد.

گروه پژوهشی طرح‌های فنی و روش‌های آمار

پژوهشکده‌ی آمار

چکیده

محققان و روش‌شناسان آمارگیری به دنبال روش‌های جدید و نوآورانه برای ارزیابی کیفیت داده‌های حاصل از آمارگیری‌ها هستند. پارادیتا، نوعی داده کمکی است که از منابع داخلی آمارگیری به دست می‌آید و استفاده از آن روش‌هایی برای ارزیابی کیفیت داده‌ها در اختیار قرار می‌دهد. در این مطالعه متداول‌ترین انواع پارادیتا شامل داده‌های ثبت اطلاعات تماس، مشخصات آمارگیر، تعامل و اقدامات متقابل در بدو آمارگیری، دلایل عدم همکاری، مسیرهای ممیزی و فایل‌های خلاصه وقایع، نظارت صامت، ارزیابی آمارگیر از کیفیت داده‌ها، ارزیابی پاسخ‌گو از کیفیت پرسشنامه و اطلاعات تماس با پاسخگو معرفی و به کاربردهای آن اشاره شده است. پس از آن تجربه کشورهای کانادا و امریکا در خصوص استفاده از پارادیتا مورد بررسی قرار گرفته است. مفاهیم گردآوری داده و پارادیتا و مواردی از دستاوردهای مطالعات قبلی پارادیتا در خصوص بهینه‌سازی هزینه‌های آمارگیری در آمار کانادا مرور شده است. همچنین مثال‌هایی در رابطه با مداخلات عملیات میدانی با استفاده از ثبت اطلاعات تماس، پایش ترکیب نمونه و مداخلات با استفاده از مشاهده آمارگیر، استفاده از پارادیتا به منظور تعدیل وزن بی‌پاسخی با هدف کاهش اریبی برآوردها، استفاده از اطلاعات پارادیتای ضربه کلید در آمارگیری‌های اینترنتی به منظور هدایت پاسخگویان به عنوان کاربردی از پارادیتا در طرح‌های آمارگیری امریکا ارائه شده است. در پایان پیشنهاداتی برای استفاده از پارادیتا در طرح‌های آمارگیری مرکز آمار ایران ارائه شده است.

فهرست مطالب

فصل اول:

- ۱-آشنایی با مفهوم پارادیتا و کاربردهای آن..... ۱
- ۱-۱- مقدمه ۱-۱
- ۲-۱- پارادیتا و برخی کاربردهای آن ۲
- ۱-۲-۱- روش‌شناسی آمارگیری ۳
- ۲-۲-۱- مدیریت گردآوری داده ۴
- ۳-۲-۱- تعدیل خطای آمارگیری ۴
- ۳-۱- مرور کلی بر انواع پارادیتا ۵
- ۱-۳-۱- مشخصات آمارگیر ۵
- ۲-۳-۱- داده‌های ثبت اطلاعات تماس ۶
- ۳-۳-۱- مشاهدات آمارگیر درباره ناحیه و محل سکونت ۷
- ۴-۳-۱- تعامل و اقدامات متقابل در بدو آمارگیری ۸
- ۵-۳-۱- مسیرهای ممیزی ۹
- ۶-۳-۱- ضبط صدا ۱۰
- ۷-۳-۱- سایر انواع پارادیتا ۱۲
- ۴-۱- پارادیتا و کاربرد آن‌ها بعد از آمارگیری ۱۳
- ۵-۱- پارادیتا در چشم‌انداز کنترل فرایند ۱۷
- ۶-۱- پارادیتای مورد استفاده در پایش و مدیریت آمارگیری‌های در حال انجام ۱۸

فصل دوم:

- مروری بر برخی کاربردهای پارادیتا در کانادا و آمریکا ۲۱
- ۱-۲-۱- مقدمه ۲۱
- ۲-۲- استفاده از پارادیتا در آمار کانادا ۲۲
- ۱-۲-۲- گردآوری داده و پارادیتا در آمار کانادا ۲۲
- ۲-۲-۲- نکاتی که از فرایند گردآوری داده فرا گرفته شده است ۲۴
- ۳-۲-۲- مطالعات در مورد بهره‌وری آمارگیری و هزینه‌ها ۲۷
- ۴-۲-۲- ارزیابی روش نوین گردآوری داده ۳۰
- ۳-۲- بهینه‌سازی حجم کار CATI برای به حداقل رساندن هزینه گردآوری داده ۳۱
- ۱-۳-۲- روش‌شناسی ۳۲
- ۲-۳-۲- مدل‌ها برای پیش‌گویی احتمال یک تماس مولد ۳۴
- ۳-۳-۲- تعداد بهینه تماس‌ها در برش زمان ۳۵
- ۴-۳-۲- اعتبارسنجی مدل ۳۷
- ۵-۳-۲- صرفه جویی در هزینه ۳۹
- ۴-۲- کاربردی از پارادیتا در طرح‌های آمارگیری آمریکا ۴۰
- ۵-۲- بحث و نتیجه‌گیری ۴۷

فصل سوم:

- پیشنهاد استفاده از پارادیتا در مرکز آمار ایران ۴۹
- مرجع‌ها ۶۳
- واژه‌نامه ۷۳

فهرست جدول‌ها

- جدول ۱-۱- خلاصه رده‌بندی انواع پارادیتا برای روش‌های مختلف آمارگیری ۱۲
- جدول ۱-۲- توزیع زمان تجمعی سیستم و زمان پرداخت در دوره‌های ده روزه طرح آمارگیری
نیروی کار و درآمد پویا ۲۰۰۷ ۲۸
- جدول ۲-۲- خلاصه‌ای از پارادیتای آمارگیری نیروی کار و درآمد پویا به تفکیک دفتر منطقه‌ای ۳۶
- جدول ۳-۲- برآورد پارامترهای مدل‌های رگرسیون خطی و لجستیک با استفاده از تعداد متوسط تماس‌های تجمعی
به ازای واحد نمونه‌گیری شده به عنوان پیش‌گویی کننده ۳۷
- جدول ۴-۲- درصد ARD برای مدل‌های رگرسیون خطی و لجستیک به تفکیک دفاتر منطقه‌ای ۳۸
- جدول ۵-۲- مقایسه بین تعداد واقعی و شبیه‌سازی شده پرسشنامه‌های تکمیل شده و هزینه‌ها (زمان به
دقیقه) برای مدل‌های رگرسیون خطی و لجستیک دفاتر منطقه‌ای ۳۸
- جدول ۶-۲- درصد کاهش تعداد تماس‌ها و هزینه (زمان صرف شده) با استفاده از برنامه زمان‌بندی بهینه با مدل‌های
رگرسیون خطی و لجستیک برای تماس‌ها و هزینه واقعی ۳۹

فهرست شکل‌ها

- شکل ۱-۱- انواع داده‌های آمارگیری و داده‌های کمکی ۲
- شکل ۱-۲- مؤلفه‌های خطای کلی آمارگیری و پارادیتا برای ارزیابی آنها ۱۵
- شکل ۲-۱- فرایند تولید و هزینه در طول چرخه آمارگیری برای طرح آمارگیری از نیروی کار و درآمد پویا از ژانویه تا آپریل ۲۰۰۷ ۲۸
- شکل ۲-۲- نتیجه مداخله نظرسنجی ملی رشد خانواده : تعادل مجموعه داده‌ها ۴۱

آشنایی با مفهوم پارادیتا و کاربردهای آن

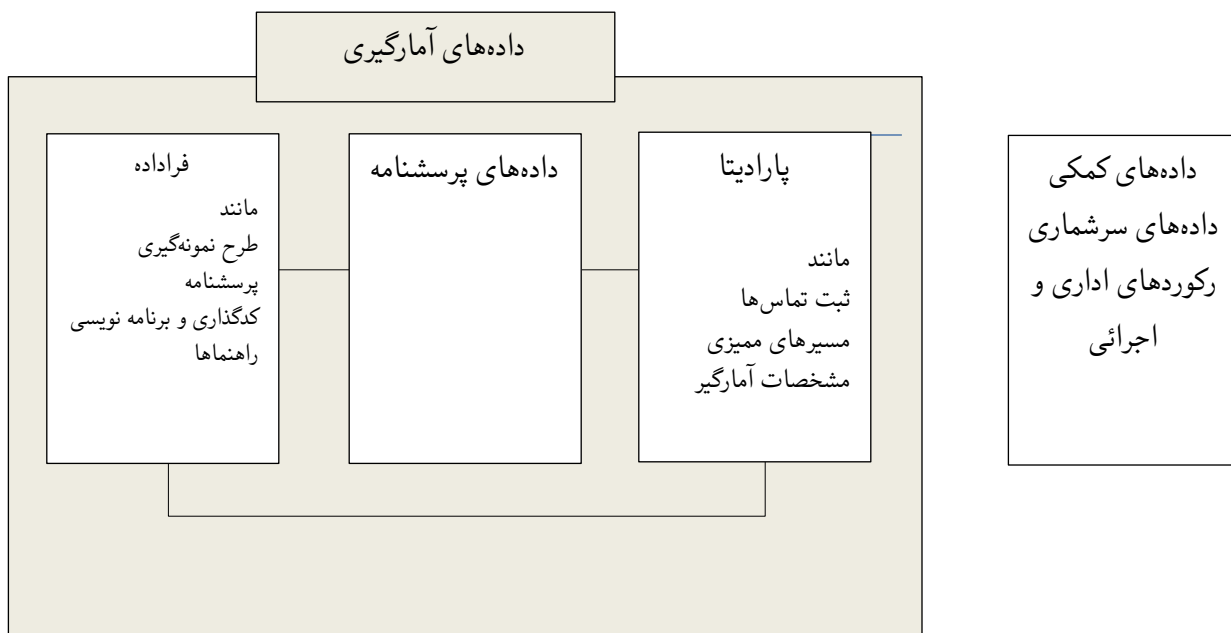
۱-۱- مقدمه

کیفیت داده‌ها در آمارگیری‌ها و روش‌های بهبود کیفیت از موضوعات مهمی است که پرداختن به آن از اهمیت بالایی برخوردار است. از این رو آمارشناسان همواره به دنبال روش‌ها و ابزارهایی هستند که با استفاده از آن‌ها بتوانند خطاهای آمارگیری را کاهش داد و در کیفیت نتایج بهبود ایجاد کرد. پارادیتا از ابزارهایی است که برای مطالعه این خطاها و همچنین هزینه‌ی آمارگیری‌ها به کار می‌رود. استفاده از پارادیتا برای شناسایی خطاهایی که به دلیل تصمیم‌های گرفته شده در طول فرایند تولید داده‌ها رخ می‌دهند، مفید است. به عنوان مثال، در مرحله فهرست برداری ممکن است مأمور آمارگیری بر اساس ارزیابی و تصمیمات شخصی خود برای ثبت یک واحد در فرم فهرست برداری عمل کند و به اشتباه برخی از واحدها را فهرست کند یا از فهرست کردن آن‌ها صرف‌نظر نماید.

اطلاعات مربوط به فرایند گردآوری داده‌ها از سال‌های دور حتی پیش از مطرح شدن واژه‌ای به نام پارادیتا، اغلب مورد استفاده آمارشناسان بوده است اما در سال‌های اخیر به‌ویژه با آغاز استفاده از روش‌های جدید در اجرای آمارگیری‌ها استفاده از آن رشد چشم‌گیری داشته است. کوپر (۱۹۹۸) اولین شخصی بود که اصطلاح «پارادیتا» را در حوزه روش‌شناسی آمارگیری معرفی کرد. او ابتدا این اصطلاح را برای اشاره به داده‌های فرایندهای خودکار، ارائه کرد. از آن به بعد این اصطلاح به همه انواع داده‌های مرتبط با فرایند گردآوری داده‌ها مانند مکالمه ضبط شده آمارگیری، مدت زمان مصاحبه و مشخصات آمارگیر و داده‌های ضربه کلید^۱ توسعه یافت. داده‌های ضربه کلید، داده‌های مربوط به ورود اطلاعات از طریق تایپ با صفحه کلید را شامل می‌شود. به عنوان مثال این‌که در فرایند پاسخ‌گویی به

یک سؤال چند مرتبه جواب تغییر داده شده است. مشاهدات آمارگیر و اطلاعات حاصل از پرسشنامه آمارگیر که فرایند گردآوری داده‌ها را توصیف نمی‌کند، به عنوان پارادیتا در نظر گرفته نمی‌شوند. ولی مشاهدات آمارگیر که طی عملیات میدانی و مواردی که در این خصوص در پرسشنامه ویژه‌ای درج می‌کند، پارادیتا محسوب می‌شود. اصطلاح پارادیتا برای داده‌های حاصل از پرسشنامه که صرفاً مرتبط با اهداف آمارگیری است، به کار گرفته نمی‌شود. به طور کلی می‌توان داده‌ها را به رده‌های زیر تقسیم کرد:

۱. داده‌های پرسشنامه آمارگیری
۲. فراداده‌ی آمارگیری: اطلاعاتی درباره خود داده‌ها، مانند: طرح نمونه‌گیری، پرسشنامه و طبقه‌بندی‌ها
۳. پارادیتای آمارگیری: اطلاعاتی درباره فرایند آمارگیری، مانند: ثبت تماس‌ها، مدت زمان گردآوری اطلاعات هر واحد و نحوه برخورد پاسخگویان
۴. داده‌های کمکی: داده‌های موجود از سایر منابع اطلاعاتی، مانند: اطلاع متغیرهای کمکی موجود در چارچوب نمونه‌گیری، مشخصات حوزه‌های سرشماری و داده‌های اداری



شکل ۱-۱ انواع داده‌های آمارگیری و داده‌های کمکی

۱-۲- پارادیتا و برخی کاربردهای آن

کروتر (۲۰۱۳) بیان می‌کند پارادیتا، داده‌های بیشتری است که در طول فرایند تهیه نتایج آمارگیری‌ها می‌توان گردآوری کرد. او همچنین پارادیتا را نوعی داده کمکی معرفی می‌کند که از منابع داخلی آمارگیری به دست می‌آید. گردآوری پارادیتای مربوط به آمارگیری، موضوع جدیدی نیست اما دامنه و جزئیات پارادیتایی که گردآوری می‌شود در سال‌های اخیر به طور اساسی افزایش یافته است، که البته بیشتر به دلیل کامپیوتری شدن تمام فرایند گردآوری داده در آمارگیری است و تنها به بخش گردآوری داده‌های پرسشنامه محدود نمی‌شود. به طور کلی آگاهی در مورد فواید