گزیده‌مطالب آماری
سال ۱۹، شماره‌ی ۱، بهار و تابستان ۱۳۸۷، صص ۲۵-۱۵

براورد دقت و مقایسه‌ی شیوه‌های اطلاع‌گیری از خانوارها

روشنک علی‌محمدی

دانشگاه الزهرا (س)

چکیده. طرح هزینه و درامد خانوار یکی از مهم‌ترین طرح‌های آماری در کشور است. داده‌های این طرح کاربردهای فراوانی در برنامه‌ریزی‌ها و تصمیم‌گیری‌های اقتصادی دارد. از جمله کاربردهای عمده‌ی آن در بررسی الگوی مصرف خانوارها، مطالعه‌ی روند تغییرات مصرف کالاها و خدمات، ارزیابی سیاست‌های اقتصادی و استفاده در حساب‌های ملی و منطقه‌ای است. بدیهی است که به‌دلیل استفاده‌های متعدد این طرح در برنامه‌ریزی‌ها و تصمیم‌گیری‌های دولت، کیفیت داده‌های آن از اهمیت زیادی برخوردار است. دقت، یکی از مهم‌ترین مؤلفه‌ها در کیفیت داده‌های حاصل از آمارگیری‌ها است. برای تعیین میزان دقت داده‌های حاصل لازم است که واریانس نمونه‌گیری و واریانس غیرنمونه‌گیری در آمارگیری‌ها براورد شود.

به‌دلیل تنوع اطلاعاتی که در این طرح گردآوری می‌شود، براوردگرهای متعددی قابل تعریف‌اند. در این مقاله میانگین هزینه‌ی خالص، هزینه‌ی ناخالص و درامد کل خانوار در نظر گرفته می‌شود. علاوه بر این واریانس نمونه‌گیری و واریانس غیرنمونه‌گیری، با به‌کارگیری مدل خطای پاسخ، برای داده‌های حاصل از دو روش اطلاع‌گیری از خانوارها محاسبه شده و دقت این روش‌های اطلاع‌گیری مقایسه می‌شود.

۱- مقدمه

اطلاعات آماری در تمامی برنامه‌ریزی‌ها از جمله در زمینه‌های اقتصادی کاربرد فراوان دارند. بدیهی است برنامه‌ریزی‌ها و تصمیم‌گیری‌های موفّق به اطلاعات آماری با کیفیت بالا نیاز دارند. یکی از طرح‌های آماری که کاربرد زیادی در زمینه‌های اقتصادی دارد، طرح

واژگان کلیدی: خطای غیرنمونه‌گیری؛ خطای نمونه‌گیری؛ دقت آمارگیری؛ نمونه‌گیری خوشه‌ای دومرحله‌ای.

دریافت: ۱۳۸۷/۶/۱۸، پذیرش: ۱۳۸۸/۲/۴

هزینه و درامد خانوار است که هر ساله توسط مرکز آمار ایران اجرا می‌شود. از کاربردهـای اساسی داده‌های این آمارگیری می‌توان بررسی الگوی مصرف خانوارها، تعیین سهم اقـلام در سبد هزینه‌های خوراکی، غیرخوراکی و کل، مطالعه‌ی روند تغییـرات مـصرف کالاهـا و خدمات، ارزیابی سیاست‌های اقتصادی دولت، محاسـبه‌ی توزیـع درامـد و ضـریب جینـی (GINI) که در تعیین عادلانه بودن توزیع درامد به‌کار می‌رود، و تعیـین خـط فقـر را نـام برد. به‌علاوه داده‌های این طرح در بخش حساب‌های ملّی شامل محاسبه‌ی مصرف نهایی بخـش خـصوصی در محاسـبه‌ی GDP (محـصول ناخـالص داخلـی) از روش هزینـه، محاسـبه‌ی ارزش افـزوده‌ی بخـش غیرمتـشکل (Informal Sector) کلیـه‌ی رشـته فعالیت‌های اقتـصادی، محاسـبه‌ی ارزش افـزوده‌ی رشـته فعالیـت‌هـایی کـه محاسـبه‌ی ستانده‌ی آن‌ها از روش مستقیم امکان‌پـذیر نیـست، محاسـبه‌ی سـهم خـانوار از خـدمات واسطه‌گری‌های مالی، کنترل نتایج حاصل از طرح‌های آمارگیری ویژه‌ی حساب‌های ملّـی و محاسبه‌ی شاخص بهای کالاها و خدمات مصرفی، مورد استفاده قرار می‌گیرد.

با توجه به اهمیت این طرح و کاربرد داده‌های حاصل از آن در زمینـه‌هـای اقتـصادی، این مقاله به بررسـی میـزان دقـت داده‌هـای ایـن طـرح از طریـق محاسـبه‌ی واریـانس نمونه‌گیری و غیرنمونه‌گیری چنـد آمـاره‌ی اصـلی در دو روش اطـلاع‌گیـری از خانوارهـا می‌پردازد. براوردگرهای مورد نظر عبارت از میانگین هزینه‌ی خالص، هزینـه‌ی ناخـالص و درامد کل خانوارها است (شایان ذکـر اسـت کـه انتخـاب براوردگـر میـانگین درامـد کـل علی‌رغم ترجیح معیارهای مرکزی دیگر بر میانگین برای این صفت، تنها به‌منظور بررسـی و مقایسه‌ی دقت میانگین برخی صفات اصلی در طرح هزینه و درامد خانوار است).

به‌طور کلی اگر فرایند ثبت داده‌ها بدون اشکال باشد، روش ثبتی می‌تواند روش دقیقی برای تولید آمار باشد. شیوه‌ی معمول برای گردآوری داده‌ها در طرح هزینه و درامد خـانوار مصاحبه‌ی حضوری و تکمیل پرسش‌نامه است. در روش جاری با مراجعه‌ی پرسـش‌گـران به خانوارها در پایان هر ماه پرسش‌نامه‌ی مزبور تکمیل می‌شود. با توجه به تعـدد سـؤالات پرسش‌نامه و خطای حافظه، داده‌های حاصل از این روش می‌تواند با خطا همراه باشد.

در شیوه‌ی ثبتی، با ارائه‌ی پرسش‌نامه‌ی طرح هزینه و درامد به خانوار نمونه، پاسخ‌گـو یا فردی باسواد از خانوار یا همسایه به ثبت اطلاعات خواسته شده و تکمیـل پرسـش‌نامـه

می‌پردازد. به‌عبارت دیگر با توجه به هزینه‌ها و درامدهای حاصل در طول دوره‌ی آمارگیری، خانوار به شیوه‌ی ثبت اطلاعات، پرسش‌نامه‌ی طرح را تکمیل می‌کند، بنا بر این انتظار می‌رود که داده‌ها به‌صورت دقیق‌تری گردآوری شوند. این مقاله با محاسبه‌ی واریانس کل چند براوردگر مهم، به مقایسه‌ی میزان دقت دو شیوه‌ی آمارگیری (روش ثبتی و مصاحبه‌ی حضوری) می‌پردازد. داده‌های مورد استفاده مربوط به یک طرح تحقیقاتی است که مرکز آمار ایران اجرا کرده است [۴].

دقت همواره به‌عنوان یک مؤلفه‌ی مهم در کیفیت داده‌های آماری مورد توجه بوده است و رابطه‌ی عکس با واریانس کل در آن آمارگیری دارد، لذا به‌منظور تعیین میزان دقت علاوه بر براورد و گزارش واریانس نمونه‌گیری لازم است که واریانس غیرنمونه‌گیری نیز محاسبه و گزارش شود.

براورد واریانس نمونه‌گیری با توجه به طرح نمونه‌گیری و شکل براوردگر انجام می‌شود. بخش اعظم نظریه‌ی نمونه‌گیری کلاسیک در مورد طرح‌های نمونه‌گیری ساده و براوردگرهای خطی است، اما در عمل سایر حالت‌ها نیز به وفور رخ می‌دهند. براورد واریانس در ارتباط با براوردگرها و طرح‌های نمونه‌گیری پایه‌ای در مرجع‌های [۵]، [۶]، [۹] و [۱۰] بحث شده است. لوی [۷] به بررسی برخی روش‌های براورد واریانس پرداخته است. ولتر [۱۱] انواع روش‌های براورد واریانس در آمارگیری‌های پیچیده را بیان کرده است.

از طرفی خطای غیرنمونه‌گیری شامل مؤلفه‌های خطای پوشش، خطای بی‌پاسخی، خطای پاسخ و خطای پردازش داده‌ها است. در بخش بعد توضیحاتی در این مورد ارائه می‌شود.

۲- واریانس کل آمارگیری

واریانس کل در آمارگیری‌ها تابعی از واریانس‌های نمونه‌گیری و غیرنمونه‌گیری است. (نشان داده شده که تحت فرضیاتی برای طرح نمونه‌گیری خوشه‌ای دومرحله‌ای، واریانس کل برابر با مجموع واریانس‌های نمونه‌گیری و غیرنمونه‌گیری است [۳]).

۱-۲- خطای غیرنمونه‌گیری

خطای غیرنمونه‌گیری یک بخش اصلی در خطای کل یک آمارگیری است. این خطا دارای مؤلفه‌های مختلفی است که به نوع آمارگیری، هدف، زمینه و نحوه‌ی اجرای آن بستگی دارد. اما به‌طور کلی مؤلفه‌های آن عبارت‌اند از خطای پوشش، خطای بی‌پاسخی، خطای پاسخ و خطای پردازش داده‌ها.

خطای پوشش به میزان انطباق بین جامعه‌ی هدف و چارچوب بستگی دارد. هر قدر این انطباق بیش‌تر باشد خطای پوشش کاهش می‌یابد. این خطا می‌تواند ناشی از کم‌پوششی یا بیش‌پوششی باشد. کم‌پوششی وقتی روی می‌دهد که برخی از واحدهای جامعه در چارچوب مورد استفاده قرار نداشته باشند. بیش‌پوششی به‌دلیل وجود واحدهای تکراری در چارچوب یا در اثر وجود واحدهایی که متعلق به جامعه‌ی هدف نیستند و در چارچوب وجود دارند، رخ می‌دهد. منبع بروز خطای پوشش، چارچوب مورد استفاده است. روش‌های اندازه‌گیری خطای پوشش استفاده از منابعی غیر از آمارگیری جاری است؛ به‌عنوان مثال مقایسه‌ی براوردها از منابع مستقل یا مقایسه‌ی تک‌تک افراد از دو فهرست مجزا.

خطای بی‌پاسخی هنگامی رخ می‌دهد که پاسخ مورد نظر به هر دلیلی از برخی واحدها دریافت نشود. این خطا دارای دو نوع است: بی‌پاسخی واحد و بی‌پاسخی پرسش. بی‌پاسخی واحد در صورتی رخ می‌دهد که یک یا چند واحد به کلی به پرسش‌ها پاسخ ندهند. بی‌پاسخی پرسش به مواردی گفته می‌شود که یک یا چند پرسش پرسش‌نامه برای برخی از واحدها بدون پاسخ باشند.

خطای پاسخ عبارت است از تفاضل مقدار واقعی (نامعلوم) و مقدار مشاهده‌شده‌ی صفت مورد نظر. این خطا بر حسب نوع آمارگیری می‌تواند دارای مؤلفه‌های مختلفی باشد. علی‌محمدی [۳] به بررسی انواع خطاهای آمارگیری پرداخته و مدلی برای خطای پاسخ طراحی کرده است.

خطای پردازش داده‌ها تمام خطاهایی که پس از اجرای آمارگیری و از مرحله‌ی ورود اطلاعات تا انتشار جدول‌های نهایی امکان بروز دارند را شامل می‌شود. مؤلفه‌های این

خطا عبارت‌اند از خطـای ورود اطلاعـات، خطـای کدگـذاری، خطـای ویـرایش و خطـای جانهی. خطای موجود در برنامه‌های کامپیوتری نیز جزء این دسته از خطاها است.

۲-۲- خطای نمونه‌گیری

این خطا در اثر انتخاب یک نمونه و تعمیم نتایج آن به جامعه رخ می‌دهد. در سرشماری‌ها مقدار خطای نمونه‌گیری صفر است. در صورتی که هدف، براورد واریـانس یـک براوردگـر خطی (از مشاهدات) در یک طرح نمونه‌گیری ساده باشـد، ایـن عمـل بـه سـهولت انجـام می‌شود، اما در صورتی که طرح نمونه‌گیری پیچیده و یـا براوردگـر مـورد نظـر غیرخطـی باشد، روش عمل پیچیده‌تر خواهد شد. به‌طور کلی تعیین براوردگر واریانس یک آمـاره بـه نـوع طـرح نمونـه‌گیـری و شـکل آمـاره بـستگی دارد. روش‌هـای بـراورد واریـانس در آمارگیری‌های پیچیده عبارت‌اند از روش گروه‌های تـصادفی، روش تکـرار مکـرر متعـادل، روش جک‌نایف و روش تابع واریانس تعمیم‌یافته. علـی‌محمـدی [۲] بـه بررسـی خطـای نمونه‌گیری برخی براوردگرها در طرح‌های آماری (که تـا آن زمـان محاسـبه نـشده بـود)، پرداخت.

۳- کاربرد

یکی از شیوه‌های ارزیابی سطح زندگی افراد یک جامعـه دسترسـی بـه الگوهـای مختلـف هزینه و درامد خانوارها با ویژگی‌های اقتصادی و اجتماعی متفاوت است. داده‌های آمـاری مورد نیاز این نوع مطالعات معمولاً با اجرای طرح هزینه و درامد خانوار در کشورها فـراهم می‌شود. به‌علاوه نتایج این طرح برای برنامه‌ریـزی‌هـا و سیاسـت‌گـزاری‌هـای اقتـصادی، اجتماعی و فرهنگی در کشور مورد استفاده قرار می‌گیرد. بـا توجـه بـه اهمیـت روزافـزون نتایج این طرح، لازم است که کاربران از میزان دقت داده‌های آن مطلع شوند.

روش معمول برای گردآوری داده‌ها در طرح هزینه و درامد خانوار، مصاحبه‌ی حضوری است. جامعه‌ی آماری این طرح شامل تمام خانوارهای معمولی سـاکن در نقـاط شـهری و روستایی کشور است. واحد آماری در این طـرح یـک خـانوار سـاکن در نقـاط شـهری یـا

روستایی است. اطلاعات لازم با مراجعه و مصاحبه‌ی مستقیم پرسش‌گران با خانوارهای نمونه و تکمیل پرسش‌نامه‌های مربوط در پایان هر ماه انجام می‌شود.

روش گردآوری داده‌ها در طرح هزینه و درامد خانوار مصاحبه‌ی حضوری است. در این روش امکان بروز خطاهایی از جانب پرسش‌گر و پاسخ‌گو وجود دارد که در نهایت منجر به افزایش خطاهای غیرنمونه‌گیری داده‌های طرح می‌شود. بخش مهمی از این خطاها مربوط به « خطای حافظه » است که به نظر می‌رسد یکی از مهم‌ترین منابع این خطا وجود فاصله‌ی زمانی بین وقوع هزینه و زمان ثبت آن و فراموشی پاسخ‌گو باشد، در این صورت انتظار می‌رود که با حذف این فاصله و ثبت اطلاعات توسط خانوار در زمان وقوع هزینه، خطای مزبور کاهش یابد. داده‌های مورد استفاده مربوط به دو شیوه‌ی اطلاع‌گیری از هزینه و درامد خانوار است که در بهمن و اسفند سال ۱۳۷۷ از ۵۰۰ خانوار نمونه‌ی شهر تهران اجرا شد. در ماه بهمن، اطلاعات توسط خانوار ثبت شد و در ماه اسفند همان اطلاعات با مراجعه‌ی پرسش‌گران به همان ۵۰۰ خانوار نمونه گردآوری شدند. در هر دو روش زمان آماری اقلام مورد پرسش، یکسان و پوشش طرح ثابت بوده است.

در گردآوری اطلاعات به‌روش ثبتی که در بهمن ماه انجام شد، مأموران آمارگیر دو هفته قبل از اجرای آمارگیری به محل سکونت خانوارهای نمونه مراجعه و ضمن آشنایی با سرپرست و اطلاع از تعداد افراد خانوار، فرد مطلعی را برای تکمیل فرم‌ها تعیین کرده و شیوه‌ی تکمیل فرم‌ها را به او آموزش دادند. وظیفه‌ی مأموران آمارگیر در هر هفته علاوه بر تحویل دادن فرم‌های جدید به خانوارهای نمونه و دریافت فرم‌های تکمیل‌شده، بازبینی و کنترل نهایی فرم‌ها و رفع ابهام احتمالی به‌ویژه در تکمیل فرم‌های مربوط به درامد بوده است.

روش نمونه‌گیری طرح هزینه و درامد خانوار، خوشه‌ای دومرحله‌ای است. مرحله‌ی اول مربوط به انتخاب بلوک‌ها و مرحله‌ی دوم مربوط به انتخاب خانوارهای نمونه در بلوک‌های انتخاب‌شده است. بلوک‌های نمونه با استفاده از فایل چارچوب و به‌روش سیستماتیک PPS (احتمال انتخاب متناسب با تعداد خانوار) انتخاب می‌شوند.

پرسش‌نامه‌ی این طرح سؤالات متعددی دارد. این سؤالات مربوط به ریز اقلام هزینه و درامدهای گوناگون خانوار است. در این بررسی چند براوردگر یعنی میانگین هزینه‌ی خالص، هزینه‌ی ناخالص و درامد کل خانوار، مورد نظر قرار گرفته است.

خطای استاندارد برای هر یک از براوردگرها در هر دو شیوه‌ی گردآوری داده‌ها محاسبه شده است. خطای استاندارد هر براوردگر معیاری برای تعیین میزان واریانس است.

حال امکان بروز انواع خطای غیرنمونه‌گیری برای داده‌های مورد نظر بررسی می‌شود.

خطای پردازش داده‌ها با توجه به استخراج دستی نتایج طرح مورد نظر برای تأمین اطلاعات لازم که دقت کافی در استخراج و ثبت آن اعمال شده، قابل اغماض است. از طرفی با توجه به نوع اجرای این طرح و هماهنگی‌های لازم، می‌توان میزان خطای پوشش و بی‌پاسخی را ناچیز فرض کرد.

به‌دلیل ماهیت خطای پاسخ، بروز آن در آمارگیری‌ها غیر قابل اجتناب است. یکی از کاربردهای مدل خطای پاسخ در آمارگیری‌ها، براورد دقت در آمارگیری‌ها است. مدل ارائه‌شده برای خطای پاسخ در مرجع [۳] عبارت است از

$$Y_{ijkls} = \mu_{ijkls} + A_i + B_{ij} + C_{ijk} + D_{ijkl} + R_{ijkls} \ , \qquad (۱)$$

که در آن μ_{ijkls} مقدار واقعی، Y_{ijkls} مقدار مشاهده‌شده‌ی پاسخ‌گوی sام در استان iام مربوط به کارشناس مسئول j ام و بازبین (و کدگذار) kام که توسط پرسش‌گر lام به‌دست می‌آید، A_i اثر استان iام، B_{ij} اثر کارشناس مسئول j ام در استان iام، C_{ijk} اثر بازبین (و کدگذار) kام که زیر نظر کارشناس مسئول j ام در استان iام فعالیت می‌کند، D_{ijkl} اثر پرسش‌گر lام که زیر نظر بازبین kام و کارشناس مسئول j ام در استان iام پرسش‌گری می‌کند، R_{ijkls} اثر پاسخ‌گوی sام در استان iام است که با lامین پرسش‌گر که زیر نظر بازبین kام و کارشناس مسئول j ام فعالیت می‌کند مصاحبه می‌شود.

در مدل (۱) عوامل مؤثر در ایجاد این خطا، اثرهای پرسش‌گر، پاسخ‌گو، کدگذار، بازبین، کارشناس مسئول و شیوه‌ی اجرای طرح در استان در نظر گرفته شده است. با توجه به نحوه‌ی اجرای آمارگیری‌ها در ایران، اثر پرسش‌گر و پاسخ‌گو همواره اثرهایی تصادفی هستند.

در این بررسی، داده‌های طرح هزینه و درامد خانوار در شهر تهران مورد نظر بوده است، لذا با توجه به این مسئله و روش اجرای این طرح، عوامل مؤثر در بروز خطای پاسخ شامل اثرهای پرسش‌گر، پاسخ‌گو و و بازبین (وکدگذار) است. از طرف دیگر تنها عوامل مؤثر در ایجاد واریانس، اثرهای تصادفی پرسش‌گرها و پاسخ‌گویان است. با توجه به روابط ارائه‌شده در مرجع [۳]، واریانس غیرنمونه‌گیری تابعی از مؤلفه‌های واریانس مدل خطای پاسخ است. براوردیابی این مؤلفه‌ها در مرجع [۱] مورد بررسی قرار گرفته است.

علاوه بر این، محاسبه‌ی واریانس نمونه‌گیری با توجه به طرح نمونه‌گیری مربوط یعنی خوشه‌ای دومرحله‌ای انجام شده است.

نتیجه‌ی محاسبات واریانس نمونه‌گیری و واریانس کل برای براوردگرهای مورد نظر در جدول ۱ آمده است.

جدول ۱– واریانس نمونه‌گیری و واریانس کل محاسبه‌شده برای براوردگرها

		روش ثبتی	روش جاری
هزینه‌ی خالص خانوار	خطای استاندارد نمونه‌گیری	۶۹۷۲۸۰	۸۸۲۰۴۳
	خطای استاندارد کل	۷۱۰۵۸۴	۹۰۶۲۵۰
هزینه‌ی ناخالص خانوار	خطای استاندارد نمونه گیری	۶۴۱۷۹۴	۸۹۵۶۰۰
	خطای استاندارد کل	۶۵۶۳۵۳	۹۲۶۰۸۸
درامد کل	خطای استاندارد نمونه گیری	۷۷۲۰۱۰	۹۱۲۷۹۷
	خطای استاندارد کل	۷۷۴۰۸۰	۹۳۴۷۹۱

چنان‌که در جدول ۱ در مورد میانگین هزینه‌ی خالص خانوار ملاحظه می‌شود، دقت آمارگیری در شیوه‌ی مصاحبه‌ی حضوری (روش معمول در گردآوری داده‌ها) کمتر از روش ثبتی است. برای حصول اطمینان از معنی‌دار بودن تفاوت بین نتایج حاصل از دو روش، آزمون مقایسه‌ی واریانس دو شیوه‌ی گردآوری داده‌ها (با فرض عدم استقلال) برای این صفت با استفاده از آزمون پیتمن انجام شد [۸]. این آزمون در سطح ۵ درصد معنی‌دار است و کمتر بودن واریانس نتایج حاصل از روش ثبتی را نسبت به روش جاری قویاً تأیید می‌کند.

در صورتی که کارایی شیوه‌ی ثبتی به روش جاری را نسبت خطاهای استاندارد براوردهای حاصل از روش جاری به روش ثبتی تعریف کنیم، ملاحظه می‌شود که برای میانگین درامد کل، شیوه‌ی ثبتی ۲۷ درصد کاراتر است.

برای براوردگر میانگین هزینه‌ی ناخالص خانوار با توجه به جدول ۱، میزان خطای استاندارد در روش جاری کمتر از شیوه‌ی ثبتی به دست می‌آید. آزمون مقایسه‌ی واریانس‌های دو شیوه‌ی گردآوری داده‌ها در حالت وابستگی مجموعه داده‌های حاصل از دو روش در سطح ۵ درصد قویاً معنی‌دار است و بیانگر واریانس کمتر نتایج حاصل از روش ثبتی است. محاسبه‌ی کارایی نشان می‌دهد که براورد میانگین هزینه‌ی ناخالص حاصل از شیوه‌ی ثبتی ۴۱ درصد کاراتر از روش جاری است.

همچنین برای میانگین هزینه‌ی خالص خانوار، با توجه به جدول ۱ میزان خطای استاندارد (کل) در شیوه‌ی ثبتی کمتر از روش جاری حاصل می‌شود. آزمون مقایسه‌ی واریانس‌های دو شیوه‌ی گردآوری داده‌ها در سطح ۵ درصد معنی‌دار است و دلالت بر دقت بیش‌تر نتایج حاصل از روش ثبتی نسبت به روش جاری دارد.

محاسبه‌ی کارایی نتیجه می‌دهد که براورد حاصل از شیوه‌ی ثبتی ۲۱ درصد کاراتر از روش جاری در این مورد است.

مقایسه‌ی واریانس کل حاصل در دو روش گردآوری داده‌ها نشان می‌دهد که در تمامی موارد، روش ثبتی به‌طور قابل ملاحظه‌ای دارای واریانس کمتر از شیوه‌ی جاری است. به‌عبارت دیگر این محاسبات حکایت از دقت بیش‌تر روش ثبت اطلاعات توسط خانوار دارد تا روش مصاحبه‌ی حضوری. لذا میزان خطای استاندارد برای همه‌ی براوردگرهای مورد بررسی نشان می‌دهد که شیوه‌ی ثبتی بازه‌ی اطمینان دقیق‌تری را با طول کوتاه‌تر نسبت به‌روش جاری ارائه می‌دهد.

۴- نتیجه‌گیری

طرح هزینه و درامد خانوار به‌عنوان یک طرح آماری پراهمیت هر ساله توسط مرکز آمار ایران اجرا می‌شود. توجه به کیفیت داده‌های این طرح به‌دلیل کاربردهای فراوان آن، ضروری می‌نماید. دقت یک اصل مهم در کیفیت داده‌ها است.

در این مقاله میزان دقت چند آماره‌ی اصلی در طرح هزینه و درامد خانوار بر اساس اطلاعات حاصل از شیوه‌های ثبت توسط خانوار و مصاحبه‌ی حضوری محاسبه شده است. مقایسه‌ی نتایج حاصل نشان می‌دهد که برای تمام آماره‌های مورد نظر، روش ثبتی دارای دقت بیش‌تری نسبت به روش مصاحبه‌ی حضوری است. لذا گردآوری داده‌ها به‌شیوه‌ی ثبت اطلاعات توسط خانوار می‌تواند میزان دقت داده‌ها را افزایش دهد و به‌همین دلیل به‌کارگیری شیوه‌ی ثبتی برای اطلاع‌گیری از خانوارها توصیه می‌شود.

مرجع‌ها

[۱] علی‌محمدی، روشنک (۱۳۸۶). آماره‌های آزمون و براورد مؤلفه‌های واریانس مدل خطای پاسخ در آمارگیری‌ها، مجله‌ی علوم دانشگاه شهید چمران، شماره ۱۸، صص ۳۵–۴۵.

[۲] علی‌محمدی، روشنک (۱۳۸۳). براورد خطای نمونه‌گیری در آمارگیری‌های پیچیده، هفتمین کنفرانس آمار ایران، دانشگاه علامه طباطبایی، تهران.

[۳] علی‌محمدی، روشنک (۱۳۸۴). مدل بندی خطای پاسخ در آمارگیری‌ها، رساله‌ی دکترای آمار، دانشگاه تربیت مدرس، تهران.

[۴] مرکز آمار ایران (۱۳۷۸). مقایسه‌ی دو روش اطلاع‌گیری از هزینه و درامد خانوار. گزارش طرح تحقیقاتی، مدیریت پژوهش‌های آماری، مرکز آمار ایران، تهران.

[5] Cochran, W.G. (1977). *Sampling Techniques.* 3rd ed., Wiley, New York.

[6] Deming, W.E. (1660). *Sample Design in Business Research*. Wiley, New York.

[7] Levy, P.S. (1991). *Sampling of Populations: Methods and Applications*. Wiley, New York.

[8] Pitman, E.J.G. (1939). A note on normal correlation. *Biometrika*, **31**, 9-12.

[9] Raj, D. (1968). *Sampling Theory*, Mc-Grow Hill, New York.

[10] Sukhatme, P.V.; Sukhatme, B.V. (1970). *Sampling Theory of Surveys with Applications*. Ames: Iowa State University Press.

[11] Wolter, k. (2007). *Introduction to Variance Estimation*. Springer-Verlag, New York.

روشنک علی‌محمدی
استادیار آمار
تهران، ده ونک، دانشگاه الزهرا، ساختمان خوارزمی، گروه ریاضی.
پیام‌نگار: alimohammadim@yahoo.com