

مجله علوم آماری، بهار و تابستان ۱۳۹۴

جلد ۹، شماره ۱، ص ۷۷-۱۰۰

روشی برای برآورد خطای اندازه‌گیری درآمد

روشنک علی‌اکبری صبا، علی‌رضا زاهدیان، مرضیه اربابی
پژوهشکده آمار

تاریخ دریافت: ۱۳۹۳/۵/۲۷ تاریخ آخرین بازنگری: ۱۳۹۳/۱۲/۲۱

چکیده: یکی از اهداف اصلی آمارگیری هزینه و درآمد خانوار، برآورد سالانه میانگین درآمد خانوارها در کشور است. از این رو با توجه به اهمیت درستی اطلاعات گردآوری‌شده و با در نظر گرفتن دلایلی که گاهی منجر به بروز خطا در اندازه‌گیری درآمد خانوارها می‌شود، در این مقاله بر اساس نتایج آمارگیری از هزینه و درآمد خانوار در سال ۱۳۹۰، به برآورد خطای اندازه‌گیری درآمد و تعدیل مقدار درآمد اظهارشده خانوارهای نمونه با استفاده از روش‌های مبتنی بر مدل پرداخته شده است.

واژه‌های کلیدی: خطای اندازه‌گیری، آمارگیری هزینه و درآمد خانوار، مدل رگرسیونی.

۱ مقدمه

در هر آمارگیری خطاهای مختلفی ممکن است رخ دهد که میزان درستی نتایج آمارگیری را تحت تأثیر قرار می‌دهند. اغلب اوقات، از بین بردن کامل این خطاها در عمل میسر نیست، اما می‌توان آن‌ها را کنترل کرد یا برای کاهش تأثیر آن‌ها بر درستی

آدرس الکترونیک مسئول مقاله: مرضیه اربابی، marzieh.arbabi@gmail.com
کد موضوع‌بندی ریاضی (۲۰۱۰): ۶۲D۰۵، ۶۲J۰۵

نتایج آمارگیری‌ها از روش‌هایی برای تعدیل یا اصلاح برآوردها استفاده کرد. یکی از عمده‌ترین خطاهایی که ممکن است در هر آمارگیری رخ دهد، خطای اندازه‌گیری است. خطای اندازه‌گیری، تفاوت مقدار به‌دست آمده یا همان مقدار مشاهده شده با مقدار واقعی (متغیر مورد بررسی) است که با انجام اندازه‌گیری بر روی واحد نمونه رخ می‌دهد (گروز و همکاران، ۲۰۰۴).

یکی از مهم‌ترین طرح‌های آمارگیری خانواری که در بسیاری از کشورهای جهان به منظور کسب اطلاعات اقتصادی خانوارها و استفاده از اطلاعات حاصل برای برنامه‌ریزی‌های اقتصادی صورت می‌گیرد، طرح آمارگیری از هزینه و درآمد خانوار است. با توجه به نقش برآوردهای خاصی از این آمارگیری‌ها در مطالعه‌های اقتصادی، برنامه‌ریزی‌ها و سیاست‌گذاری‌ها، میزان درستی اطلاعات و نتایج حاصل از آن‌ها از اهمیت ویژه‌ای برخوردار است.

بر اساس نتایج طرح آمارگیری از هزینه و درآمد خانوار در ایران در سال‌های اخیر، متوسط درآمد خانوارها در کشور اغلب کمتر از متوسط هزینه‌های آن‌ها است که این موضوع می‌تواند ناشی از خطای اندازه‌گیری درآمد باشد. این در حالی است که بر اساس بررسی‌های انجام شده در مورد میانگین درآمد و میانگین هزینه‌های خانوارها در سایر کشورها که از آمارگیری‌های مشابهی به دست آمده‌اند، متوسط درآمد خانوارها در اغلب این کشورها بیش از متوسط هزینه‌های آن‌ها است (زاهدیان و همکاران، ۱۳۹۲).

به عنوان مثال، در کشور آلمان، میانگین درآمد ماهانه خانوار در سال ۱۹۹۸ به اندازه ۶۰۳ یورو، در سال ۲۰۰۳ به اندازه ۷۰۸ یورو و در سال ۲۰۰۸ به اندازه ۷۲۰ یورو از میانگین هزینه ماهانه خانوار بیشتر بود. در کشور استرالیا، میانگین درآمد هفتگی خانوار در سال‌های ۱۹۹۹ و ۲۰۰۹ به ترتیب به اندازه ۱۸۰ دلار و ۴۴۹ دلار از میانگین هزینه هفتگی خانوار بیشتر بود. همچنین در مالزی، میانگین درآمد ماهانه خانوار در سال ۲۰۰۴ به اندازه ۲۹۷ رینگیت و در سال ۲۰۰۹ به اندازه ۱۸۳۵ رینگیت از میانگین هزینه ماهانه خانوار بیشتر بود (زاهدیان و همکاران، ۱۳۹۲).

به‌طور کلی، میزان درستی اطلاعات مربوط به هزینه و درآمد اظهارشده توسط پاسخ‌گویان از جمله مواردی است که در بسیاری از مطالعات به‌ویژه مطالعات

مربوط به آمارگیری هزینه و درآمد خانوار مورد توجه قرار می گیرد و از آنجا که اجتناب از خطا امری غیرممکن است، محاسبه مقدار خطای اندازه گیری موجود در داده ها یا برآورد آن نظر بسیاری از محققین را به خود جلب کرده است. بر اساس مطالعات انجام شده در خصوص خطای اندازه گیری، منابع بروز و روش های برآورد آن، روش های زیر را می توان برای برآورد این خطا به کار گرفت (زاهدیان و همکاران، ۱۳۹۲):

- مقایسه با داده های ثبتی؛
- مقایسه با اطلاعات آمارگیری های دیگر؛
- مقایسه با اطلاعات دوره قبل در آمارگیری های پانلی؛
- آمارگیری مجدد؛
- تکرار بخشی از آمارگیری؛
- استفاده از سؤال های کنترلی؛
- بررسی نظرات مأموران آمارگیری؛
- استفاده از روش های مبتنی بر مدل.

کشورهای مختلف بسته به امکانات، شرایط و اطلاعات موجود از روش های مختلفی برای برآورد خطای اندازه گیری استفاده می کنند. در ادامه به چند مورد از بررسی های انجام شده به اختصار اشاره می شود.

باند و همکاران (۱۹۹۰) از یک مجموعه داده منحصر به فرد استفاده کردند که این مجموعه، داده های دو سال متوالی آمارگیری های جمعیت جاری آمریکا و مدارک دریافتی های تأمین اجتماعی گزارش شده توسط کارفرماها را به هم مرتبط می کرد. داده ها برای دریافتی های گزارش شده هر مستخدم توسط خود او و دریافتی های گزارش شده توسط کارفرما در هر زمان موجود بود. آن ها فرض کردند مدارک صورت پرداخت تأمین اجتماعی منبع نسبتاً دقیقی از دریافتی های درست است. نتایج این بررسی نشان داد که خطای گزارش شده در دریافتی های آمارگیری جمعیت جاری از یک سال به سال دیگر همبستگی مثبتی دارد. همچنین ضریب خودهمبستگی مقدار خطای اندازه گیری برای مردان برابر با $0/4$ و برای زنان برابر با

۸۰ روشی برای برآورد خطای اندازه‌گیری درآمد

۰/۱ بود. همچنین آنها بر اساس مقایسه داده‌های مطالعه اعتبارسنجی^۱ (PSID) و آمارگیری جمعیت جاری، خطای اندازه‌گیری درآمد در آمارگیری جمعیت جاری را ۲۴ درصد برآورد کردند.

گراندین و میچاد (۱۹۹۴) بر اساس اطلاعات اظهارنامه مالیاتی مربوط به مزایای بیکاری کشور کانادا و با در نظر گرفتن این‌که حدود ۱۲ تا ۱۹ درصد از گزارش‌های آمارگیری دارای خطا هستند، به این نتیجه دست یافتند که متوسط مقدار خطای درآمد مقداری بین ۱۰ تا ۱۳ درصد است.

همچنین مور و همکاران (۱۹۹۷) نشان دادند در آمریکا خانوارها کمک‌های دریافت شده از دولت در رابطه با خانوارهای کم درآمد را ۱۳ درصد کمتر اعلام می‌کنند.

در آفریقای جنوبی، آگرو و همکاران (۲۰۰۷) در مطالعه پویایی درآمد بر اساس داده‌های پانلی درآمد در سال‌های ۱۹۹۳ تا ۱۹۹۸، به این نتیجه دست یافتند که میزان خطای اندازه‌گیری درآمد تا ۶۰ درصد است.

نری و زیزا (۲۰۱۰) در بررسی نتایج طرح‌های نمونه‌ای درآمد در ایتالیا از داده‌های آمارگیری درآمد و دارایی خانوارهای ایتالیا در سال ۲۰۰۶ و مدل لجستیک برای بررسی ویژگی‌های واحدهای کم‌گو استفاده کردند و نشان دادند که برآورد کم‌گویی در متوسط درآمد خانوارها برابر ۴۰۰۰ یورو یعنی ۱۲ درصد است.

هارتس و همکاران (۲۰۱۲) نیز با استفاده از ساختار هزینه و درآمد کارکنان مزد و حقوق‌بگیر خطای اندازه‌گیری درآمد خویش‌فرماها در آمارگیری از هزینه‌های مصرف‌کنندگان را حدود ۲۵ درصد برآورد کردند.

لازم به توضیح است که هر یک از روش‌های مطرح شده برای برآورد خطای اندازه‌گیری، بسته به نوع صفت‌های مورد آمارگیری، هزینه و امکان دسترسی به منابع اطلاعاتی بیرونی، دارای مزایا و معایب خاص خود هستند. از آنجا که از نظر هزینه و دسترسی به اطلاعات مورد نیاز و همچنین حفظ شرایط برآورد، روش‌های مبتنی بر مدل از قابلیت کاربرد بیشتری برخوردار هستند، در این مقاله برای برآورد خطای اندازه‌گیری درآمد از روش‌های مبتنی بر مدل استفاده شده است (زاهدیان و

^۱ Panel Study of Income Dynamics

همکاران، ۱۳۹۲).

روش‌های مبتنی بر مدل یکی از روش‌های برآورد خطای اندازه‌گیری است که با استفاده از داده‌های آمارگیری‌های نمونه‌ای یا داده‌های منابع اطلاعاتی دیگر به کار می‌رود. مدل رگرسیونی یکی از پرکاربردترین مدل‌هایی است که برای بررسی خطای اندازه‌گیری یا تأثیر آن بر نتایج سایر بررسی‌های آماری مورد استفاده قرار می‌گیرد. در این مقاله برای بررسی خطای اندازه‌گیری درآمد خانوارها در آمارگیری هزینه و درآمد خانوارهای شهری و روستایی کشور در سال ۱۳۹۰، از یک مدل رگرسیونی استفاده و درآمد خانوارها تعدیل شده است. اختلاف بین درآمد تعدیل شده و درآمد اظهار شده خانوارها به‌عنوان شاخصی برای خطای اندازه‌گیری (نه لزوماً مقدار دقیق خطا) در نظر گرفته شده است. از این رو در بخش ۲ ابتدا به نکاتی در خصوص روش‌شناسی آمارگیری از هزینه و درآمد خانوارها اشاره می‌شود. سپس بر اساس داده‌های این آمارگیری یک مدل رگرسیونی برای برآورد خطای اندازه‌گیری درآمد خانوارها برآزش داده می‌شود که شیوه برآزش مدل، ویژگی‌ها و نتایج حاصل از آن در بخش ۳ به تفصیل توضیح داده می‌شود. پس از آن، در بخش ۴ استنباطی در مورد خطای اندازه‌گیری درآمد با داده‌های نمونه‌های چرخشی ارائه می‌شود و در بخش ۵ به تعدیل اجزای درآمد خانوارها در سال ۱۳۹۰ پرداخته می‌شود. بخش ۶ نیز به بحث و نتیجه‌گیری اختصاص داده شده است.

۲ آمارگیری هزینه و درآمد خانوار

آمارگیری هزینه و درآمد خانوار یک آمارگیری نمونه‌ای سالانه است که از سال ۱۳۴۲ در مناطق روستایی و از سال ۱۳۴۷ در مناطق شهری کشور به اجرا درآمده است. از سال ۱۳۵۳ علاوه بر هزینه‌های خانوار، در این آمارگیری اطلاعات درآمد خانوارها نیز گردآوری شده است. هدف کلی این طرح آمارگیری در سال ۱۳۹۰، برآورد سالانه میانگین هزینه‌های کل، خوراک، غیر خوراک و درآمد خانوارها و نیز برآورد تغییرات سالانه این میانگین‌ها در سطح مناطق شهری و روستایی کل کشور و برآورد سالانه میانگین هزینه‌های کل، خوراک، غیر خوراک و درآمد خانوارها

در سطح نقاط شهری و روستایی هر یک از استان‌ها است (دستورالعمل طرح نمونه‌گیری آمارگیری هزینه و درآمد خانوار شهری/روستایی - سال ۱۳۹۰).

جامعه هدف این آمارگیری، شامل همه خانوارهای معمولی ساکن و گروهی سال ۱۳۹۰ در مناطق شهری و روستایی کشور است. خانوارهای معمولی غیر ساکن و مؤسسه‌ای از شمول این طرح مستثنی هستند. جامعه آمارگیری طرح نیز با حذف تعدادی از بلوک/آبادی‌ها در مناطق نظامی و انتظامی، صعب‌العبور و ناامن و آبادی‌های دارای ۵ خانوار یا کمتر از جامعه هدف به دست آمده است (دستورالعمل طرح نمونه‌گیری آمارگیری هزینه و درآمد خانوار شهری/روستایی - سال ۱۳۹۰).

اطلاعات مورد نیاز از طریق مصاحبه رو در رو با خانوارهای نمونه و تکمیل پرسشنامه گردآوری شده است. برآوردهای اصلی این طرح، در سطح نقاط شهری و روستایی کل کشور یا هر یک از استان‌ها محاسبه و ارائه می‌شود. حداکثر خطای نسبی پذیرفته‌شده در این طرح برای برآورد میانگین متغیرهای مورد نظر در اهداف طرح، برابر ۱۰ درصد است (چکیده نتایج طرح آمارگیری هزینه و درآمد خانوارهای شهری و روستایی سال ۱۳۹۰).

نمونه‌های این طرح طی سه مرحله و با استفاده از روش‌های نمونه‌گیری احتمالی انتخاب شده‌اند. با توجه به لزوم ارائه برآورد تغییرات در طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی از سال ۱۳۸۹، استفاده از نمونه‌گیری چرخشی در این طرح مورد نیاز بود. نمونه‌گیری چرخشی با ثابت نگه داشتن بخشی از واحدهای نمونه بین دو دوره آمارگیری و تغییر بقیه واحدها، به بهترین نحو، امکان برآورد سطوح و تغییرات را فراهم می‌کند. مرحله‌ی اول و دوم نمونه‌گیری در آمارگیری از هزینه و درآمد خانوار سال ۱۳۹۰ برای انتخاب واحدهای نمونه‌گیری اولیه (شبه‌حوزه‌ها) و ثانویه (قطعه‌ها) و ساخت گروه‌های چرخش صورت گرفته است. هر گروه چرخش شامل ۵ خانوار است. در مرحله آخر نمونه‌گیری، برای تعیین خانوارهای نمونه در هر سال داخل هر یک از گروه‌های چرخش، یک خانوار (به‌عنوان خانوار نمونه اصلی) به روش تصادفی ساده انتخاب شده است.

الگوی چرخش انتخاب‌شده برای این طرح، الگوی (۰)-۵ است، که در آن طی پنج سال از هر خانوار نمونه، حداکثر ۵ بار آمارگیری به عمل می‌آید. به این معنا که

خانوار، بسته به گروه چرخشی که به آن تعلق دارد یک یا چند (حداکثر ۵) سال در نمونه قرار می گیرد و پس از آن برای همیشه از نمونه خارج می شود. در این طرح پس از گردآوری داده‌ها، برای سهولت در محاسبه‌ی برآورد پارامترهای مورد نظر، از جمله میانگین و نسبت، واحدهای نمونه وزن دهی می شوند. فرمول‌های وزن دهی و برآورد در دستورالعمل طرح نمونه‌گیری ارائه شده است.

۳ برآزش مدل

در این بخش برای برآورد خطای اندازه‌گیری درآمد در آمارگیری هزینه و درآمد خانوارها از یک مدل رگرسیونی خطی استفاده می شود. در این مدل، برآورد یک متغیر پاسخ بر اساس داده‌های مجموعه‌ای از متغیر مستقلی صورت می گیرد که رابطه‌ی خطی نسبتاً قوی با متغیر پاسخ دارند. الگوی کلی مدل رگرسیون خطی با یک متغیر پاسخ به صورت

$$Y = X\beta + \epsilon$$

است، که در آن $Y = (y_1, \dots, y_n)'$ ، $\beta = (\beta_0, \dots, \beta_r)'$ ، $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ و

$$X = \begin{pmatrix} 1 & \dots & 1 \\ x_{11} & \dots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \dots & x_{nn} \end{pmatrix}$$

و n تعداد کل مشاهدات است. فرض می شود مشاهده‌های مختلف ناهمبسته و دارای توزیع نرمال با میانگین صفر و واریانس ثابت σ^2 هستند. با فرض کامل بودن رتبه X ، برآورد کمترین توان‌های دوم β از رابطه $\hat{\beta} = (X'X)^{-1}X'Y$ به دست می آید.

در ادامه، با استفاده از یک مدل رگرسیونی خطی، مقدار درآمد مورد انتظار خانوارها بر اساس مشخصات محل سکونت، تسهیلات و لوازم عمده زندگی، ویژگی‌های اعضا و سایر متغیرهای مرتبط با هزینه و درآمد خانوارها برآورد می شود. سپس با در نظر گرفتن اختلاف درآمد مورد انتظار و درآمد اظهارشده

خانوارها به‌عنوان خطای اندازه‌گیری، درآمد خانوارها با روشی که در ادامه توضیح داده خواهد شد، تعدیل می‌شود تا تأثیر این خطا بر داده‌های درآمد و برآوردهای محاسبه‌شده جبران شود. برای برآزش مدل از نرم‌افزار SAS استفاده شده است.

این بررسی با پذیرفتن فرض اولیه وجود خطای ۲۰ درصدی در اندازه‌گیری درآمد خانوارها انجام شده است. بدین معنی که درآمد واقعی خانوارها پیش از برآزش مدل، ۲۰ درصد بیش از درآمد اظهارشده خانوارها در نظر گرفته شده است. با توجه به نتایج حاصل از بررسی داده‌های خانوارهای نمونه یکسان در سال‌های متوالی، که در بخش بعد ارائه شده است، متوسط خطای اندازه‌گیری اولیه خانوارها حداقل ۲۰ درصد درآمد اظهارشده آنها است، از این رو پذیرفتن این فرض منطقی به نظر می‌رسد. پس از مطالعه اطلاعات آمارگیری هزینه و درآمد خانوار و همبستگی اقلام اطلاعاتی مختلف با درآمد خانوارها، مواردی چون تعداد اعضای خانوار، نسبت هزینه غیر خوراک به هزینه کل خانوار (منظور از هزینه خانوار در این بررسی، هزینه ناخالص خانوار است)، نسبت درآمد واقعی به هزینه کل خانوار، سطح سرانه زیربنای محل سکونت خانوار، مدرک تحصیلی سرپرست خانوار، اجاری بودن یا نبودن محل سکونت خانوار، لگاریتم مبلغ سرانه وام دریافت‌شده توسط خانوار (در صورت دریافت وام در سال گذشته) و لگاریتم طبیعی هزینه کل خانوار به‌عنوان متغیرهای مستقل مدل رگرسیون در نظر گرفته شدند.

با در نظر گرفتن لگاریتم درآمد سرانه خانوارها به‌عنوان متغیر پاسخ و متغیرهای مستقل در نظر گرفته شده، یک مدل رگرسیون خطی با عرض از مبدأ به داده‌ها برآزش داده شد. از آنجا که اعتبار نتایج حاصل از مدل‌های رگرسیونی به انتخاب متغیرهای مناسب و برقرار بودن فرض‌های پایه‌ای بستگی دارد، پس از برآزش مدل اولیه، درستی این فرض‌ها بررسی شد. یکی از شاخص‌های بررسی هم‌خطی متغیرهای مستقل، که در صورت وجود همبستگی قوی بین این متغیرها رخ می‌دهد، شاخص عامل تورم واریانس^۲ (VIF) (جانسون و ویچرن، ۱۹۸۲). این شاخص نشانگر میزان افزایش واریانس برآورد پارامترها به دلیل وجود هم‌خطی بین متغیرها است. VIF مقداری بزرگتر از یک اختیار می‌کند و هر چه رابطه بین متغیرها بیشتر باشد

^۲ Variance Inflation Factor

مقدار آن افزایش می یابد. در این بررسی، بر اساس مطالعه‌ی اولیه مدل رگرسیونی، بزرگترین مقدار VIF در نقاط شهری برابر با ۱/۸۹ و در نقاط روستایی برابر با ۱/۷۹ است که نشانگر نبود هم خطی معنی دار بین متغیرهای تحت بررسی است.

با توجه به این که فرض نرمال بودن توزیع خطای مدل یکی از فرض‌های اساسی مدل رگرسیون است و خطای مورد بررسی دارای توزیع غیر نرمال است، برای تبدیل توزیع مقادیرهای خطای مدل به توزیع نرمال، در این بررسی از لگاریتم درامد سرانه (به جای درامد سرانه) به عنوان متغیر پاسخ استفاده شد. بدین منظور پیش از برازش مدل، خانوارهای دارای درامد منفی شامل ۱۰ خانوار از ۱۸۷۲۷ خانوار شهری و ۲۷ خانوار از ۱۹۷۸۶ خانوار روستایی کنار گذاشته شدند.

با استفاده از نمودار مقدار خطاهای مدل در برابر مقادیرهای برآورد شده متغیر پاسخ و نمودارهای توزیع‌های احتمال و چندک‌های خطاها در مقابل توزیع نرمال به تفکیک نقاط شهری و روستایی فرض‌های استقلال مقادیر خطا و تبعیت آن‌ها از توزیع نرمال تأیید و پذیرفته شدند.

با اطمینان از درستی فرض‌های پایه‌ای، مدل برازش داده شده پذیرفته شد. ضریب تعیین تصحیح شده مدل، که با در نظر گرفتن وزن نمونه‌گیری به تفکیک نقاط شهری و روستایی کشور در جدول ۱ ارائه شده است، نشان می دهد نسبتی از تغییرات کل متغیر پاسخ که توسط متغیرهای مستقل بیان می شود نسبت قابل توجهی است.

جدول ۱: ضریب تعیین تصحیح شده و ریشه میانگین کمترین توان‌های دوم در مدل رگرسیونی

نقاط روستایی	نقاط شهری	برآورد
۰/۷۹	۰/۸۹	ضریب تعیین تصحیح شده
۰/۲۸	۰/۲۱	ریشه میانگین کمترین توان‌های دوم

ضریب تعیین تصحیح شده برای مدل دارای عرض از مبدأ از رابطه

$$R_{adj}^2 = 1 - \frac{(n-1)(1-R^2)}{n-p}$$

به دست می‌آید، که در آن

$$R^2 = 1 - \frac{SS_{Error}}{SS_{Total}}$$

ضریب تعیین مدل رگرسیونی، SS_{Error} مجموع توان دوم مقادیر خطا، SS_{Total} مجموع کل توان دوم مقادیر متغیر پاسخ، n تعداد مشاهده‌ها و p تعداد پارامترهای مدل است (جانسون و ویچرن، ۱۹۸۲).

ضرایب رگرسیونی متغیرهای مدل با در نظر گرفتن وزن نمونه‌گیری به تفکیک نقاط شهری و روستایی کشور در جدول‌های ۲ و ۳ ارائه شده‌اند.

جدول ۲: برآورد ضرایب‌های رگرسیونی و p -مقدار آن‌ها در مدل نقاط شهری

۱۳۹۰

نام متغیر	سطح متغیر کیفی	برآورد	خطای استاندارد	p -مقدار
عرض از مبدأ	-	۱/۶۵۴	۰/۱۷۱	< ۰/۰۰۱
تعداد اعضای خانوار	-	-۰/۲۲۸	۰/۰۰۳	< ۰/۰۰۱
نسبت هزینه‌ی غیر خوراک به هزینه‌ی کل خانوار	-	۰/۲۲۳	۰/۰۲۳	< ۰/۰۰۱
نسبت درآمد واقعی به هزینه‌ی کل خانوار	-	۰/۶۱۸	۰/۰۱۸	< ۰/۰۰۱
سطح سرانه‌ی زیربنای محل سکونت خانوار	-	۰/۰۰۴	۰/۰۰۰	< ۰/۰۰۱
مدرک تحصیلی سرپرست خانوار	کمتر از لیسانس لیسانس یا کمتر از فوق‌لیسانس فوق‌لیسانس و بالاتر	۰/۰۳۴ ۰/۰۶۳ ۰/۱۱۱	۰/۰۲۵ ۰/۰۲۶ ۰/۰۲۹	۰/۱۷۱ ۰/۰۱۴ < ۰/۰۰۱
اجاری بودن یا نبودن محل سکونت خانوار	دوره‌های علوم دینی ملک غیر اجاری ملک اجاری	۰/۰۶۲ ۰/۰۰۰ ۰/۰۰۰	۰/۰۰۰ ۰/۰۰۰ ۰/۰۰۰	< ۰/۰۰۱ < ۰/۰۰۱ < ۰/۰۰۱
لگاریتم مبلغ سرانه‌ی وام دریافت‌شده توسط خانوار	-	۰/۰۰۲	۰/۰۰۰	< ۰/۰۰۱
لگاریتم طبیعی هزینه‌ی کل خانوار	-	۰/۸۲۹	۰/۰۰۹	< ۰/۰۰۱

در مدل برآزش داده‌شده برای خانوارهای شهری دو متغیر مدرک تحصیلی سرپرست خانوار و اجاری بودن یا نبودن محل سکونت خانوار متغیر کیفی محسوب می‌شوند که با در نظر گرفتن سطح معنی‌داری ۵ درصد هم برای نقاط شهری و هم برای نقاط روستایی، هر دو تأثیری معنی‌دار بر درآمد سرانه خانوارها دارند. برای تحلیل بهتر نتایج مدل، هر یک از سطوح متغیرهای کیفی به‌عنوان یک کلاس در نظر گرفته شده‌اند تا امکان مقایسه‌ی ضرایب‌های رگرسیونی برای سطوح مختلف این

متغیرها و شدت تأثیر آن‌ها بر متغیر پاسخ فراهم شود. از این رو ضریب‌های رگرسیونی آخرین رده از متغیرهای کیفی، که به‌عنوان سطح معیار در نظر گرفته شده‌اند، برابر صفر است. شایان ذکر است برای مقایسه میزان تأثیر متغیرهای کیفی بر متغیر پاسخ، اختلاف بین کوچکترین و بزرگترین مقدار ضریب‌ها برای سطوح هر یک از این متغیرها نشانگر میزان تأثیر آن متغیر کیفی بر متغیر پاسخ است. همان‌گونه که انتظار می‌رفت تعداد اعضای خانوار و لگاریتم مبلغ سرانه وام دریافت‌شده توسط خانوار تأثیری منفی بر لگاریتم درآمد سرانه خانوار دارند.

جدول ۳: برآورد ضریب‌های رگرسیونی و p -مقدار آن‌ها در مدل نقاط روستایی ۱۳۹۰

نام متغیر	برآورد	خطای استاندارد	p -مقدار
عرض از مبدأ	۲/۸۷۷	۰/۲۴۵	< ۰/۰۰۱
تعداد اعضای خانوار	-۰/۱۹۱	۰/۰۰۴	< ۰/۰۰۱
نسبت درآمد واقعی به هزینه کل خانوار	۰/۵۲۵	۰/۰۳۵	< ۰/۰۰۱
نسبت زیربنای محل سکونت خانوار	۰/۰۰۵	۰/۰۰۰	< ۰/۰۰۱
لگاریتم طبیعی هزینه کل خانوار	۰/۷۷۴	۰/۰۱۲	< ۰/۰۰۱

جدول ۳ برآورد ضریب‌های رگرسیونی را در مدل نقاط روستایی ۱۳۹۰ نشان می‌دهد. متغیرهای نسبت هزینه غیر خوراک به هزینه کل، مدرک تحصیلی سرپرست خانوار، اجاری بودن یا نبودن محل سکونت و مبلغ سرانه وام دریافت‌شده توسط خانوار در مدل نقاط روستایی حضور ندارند، زیرا در بررسی‌های اولیه، این متغیرها به دلیل معنی دار نبودن ضرایب، از مجموعه متغیرهای مستقل مدل نقاط روستایی حذف شدند.

پس از برآزش مدل، درآمد مورد انتظار خانوارها و مقدار سرانه آن‌ها در هر دو سطح نقاط شهری و روستایی کشور محاسبه شدند. درآمد تعدیل‌شده هر خانوار به‌عنوان بزرگترین مقدار بین درآمد اظهارشده خانوار و درآمد مورد انتظار برآوردشده توسط مدل برای آن خانوار در نظر گرفته شد. درآمد تعدیل‌شده سرانه خانوارها نیز به‌صورت مشابه تعریف شد. اختلاف بین درآمد تعدیل‌شده و درآمد (اظهارشده) خانوارها نیز به‌عنوان شاخصی برای کم‌گویی درآمد خانوار در نظر گرفته شد. در هر

سطح، خانوارها بر اساس متغیرهای مختلف دهک‌بندی و برآوردهای توصیفی مورد نظر محاسبه شدند. نتایج در جدول‌های ۴، ۵ و ۶ ارائه شده است. لازم به ذکر است که اندک اختلاف میانگین هزینه و درآمد اظهارشده خانوارها در این جدول‌ها با نتایج آمارگیری از هزینه و درآمد خانوار شهری و روستایی سال ۱۳۹۰ ناشی از حذف خانوارهایی است که دارای درآمد منفی بوده‌اند. همچنین پس از برازش مدل، دو خانوار نمونه در نقاط روستایی و یک خانوار نمونه در نقاط شهری به‌عنوان داده پرت شناسایی و از مجموعه‌ی داده‌ها کنار گذاشته شدند. همچنین واژه دهک، که در این مقاله برای زیربخش‌های مختلف جامعه به‌کار رفته است، شامل گروهی از خانوارها است که متغیر مورد بررسی آن‌ها بین دهک‌های آماری آن متغیر قرار گرفته است.

جدول ۴: توزیع متغیرهای هزینه و درآمد خانوارها به تفکیک نقاط شهری و روستایی در سال ۱۳۹۰

نام متغیر	دهک آماری	خانوارهای شهری		خانوارهای روستایی	
		برآورد	خطای استاندارد	برآورد	خطای استاندارد
درآمد اظهارشده	۱	۵۰,۰۹۶,۸۱۸	۶۵۴,۷۴۱	۲۵,۳۷۶,۸۶۸	۳۵۳,۴۱۲
	۲	۶۸,۶۳۷,۷۱۶	۶۰۲,۳۴۳	۳۷,۲۸۴,۶۹۶	۳۲۶,۹۷۳
	۳	۸۳,۶۷۰,۷۷۸	۷۱۹,۰۹۰	۴۷,۶۴۱,۹۲۷	۳۰۵,۴۶۴
	۴	۹۷,۳۹۵,۳۲۲	۷۰۳,۹۵۲	۵۷,۱۷۰,۷۳۵	۳۵۰,۴۰۱
	۵	۱۱۱,۶۰۵,۶۷۸	۷۲۶,۱۳۸	۶۷,۲۰۵,۰۳۷	۴۸۵,۵۷۶
	۶	۱۲۷,۶۰۴,۰۴۷	۹۴۸,۰۴۴	۷۸,۴۵۶,۵۹۴	۵۱۸,۶۰۳
	۷	۱۴۷,۹۶۴,۵۳۵	۱,۰۵۵,۴۸۴	۹۲,۰۸۵,۴۹۸	۶۸۳,۳۷۳
	۸	۱۷۷,۱۳۶,۶۷۷	۱,۷۶۵,۸۸۹	۱۱۰,۷۶۳,۴۹۹	۶۹۳,۳۳۹
	۹	۲۲۹,۶۲۰,۹۳۹	۲,۳۶۹,۷۱۷	۱۴۳,۸۶۹,۵۲۱	۱,۴۳۵,۲۳۵
درآمد تعدیل‌شده	۱	۶۵,۱۹۷,۳۶۶	۶۶۳,۳۲۷	۳۴,۸۷۳,۱۰۸	۴۸۲,۱۴۳
	۲	۸۵,۴۵۶,۸۱۰	۶۲۲,۴۴۰	۵۱,۰۵۸,۹۵۵	۴۶۲,۹۴۰
	۳	۱۰۲,۴۵۵,۰۰۸	۷۲۱,۰۱۰	۶۲,۴۲۸,۷۲۵	۴۳۷,۳۳۰
	۴	۱۱۹,۱۸۶,۷۶۵	۸۵۹,۲۱۷	۷۲,۳۴۸,۱۱۴	۴۰۹,۲۰۸
	۵	۱۳۴,۲۱۷,۶۵۵	۸۵۰,۳۷۴	۸۲,۴۳۳,۶۷۶	۴۳۴,۱۵۷
	۶	۱۵۲,۸۱۰,۰۲۳	۹۷۳,۵۰۵	۹۳,۰۸۴,۱۶۱	۴۹۷,۵۳۳
	۷	۱۷۵,۹۷۳,۰۱۴	۱,۳۳۸,۶۸۹	۱۰۶,۶۸۰,۵۵۹	۵۵۷,۵۳۲
	۸	۲۰۸,۱۴۴,۱۵۱	۱,۷۸۴,۱۵۸	۱۲۴,۷۷۲,۵۹۹	۸۲۴,۸۹۴
	۹	۲۶۵,۶۶۳,۴۹۵	۳,۱۱۱,۹۴۳	۱۵۷,۹۹۰,۲۹۷	۱,۴۵۸,۳۵۲

میانگین هزینه، درآمد اظهارشده و درآمد تعدیل‌شده خانوارها در نقاط شهری و روستایی به تفکیک دهک‌های درآمد تعدیل‌شده به ترتیب در شکل‌های ۱ و ۲ ارائه شده است.

جدول ۵: برآورد شاخص‌های موردنظر پس از برآزش مدل به داده‌ها در نقاط شهری

سال ۱۳۹۰ به تفکیک دهک درآمد تعدیل شده خانوارها (بر حسب هزار ریال)

دهک درآمد	تعداد خانوارهای نمونه	میانگین درآمد	میانگین هزینه تعدیل شده	میانگین درآمد نمونه	میانگین کم‌گویی
۱	۲,۱۶۶	۳۶,۴۲۵	۴۴,۶۲۶	۴۵,۵۲۳	۹,۰۹۸
۲	۲,۰۴۱	۶۱,۹۷۱	۶۸,۸۷۶	۷۶,۰۷۷	۱۴,۱۰۶
۳	۲,۰۲۶	۷۷,۶۲۰	۸۴,۰۹۷	۹۳,۹۳۷	۱۶,۳۱۷
۴	۱,۹۴۸	۹۱,۷۱۳	۹۷,۶۳۱	۱۱۰,۸۸۸	۱۹,۱۷۵
۵	۱,۸۶۶	۱۰۵,۶۳۹	۱۱۰,۸۶۲	۱۲۶,۵۶۵	۲۰,۹۲۶
۶	۱,۸۶۲	۱۱۹,۶۶۰	۱۲۷,۳۰۶	۱۴۳,۲۶۸	۲۳,۶۰۸
۷	۱,۸۵۲	۱۳۷,۵۹۹	۱۴۵,۹۱۷	۱۶۳,۴۸۳	۲۵,۸۸۴
۸	۱,۸۱۳	۱۶۲,۰۱۶	۱۶۸,۹۶۴	۱۹۱,۵۰۹	۲۹,۴۹۲
۹	۱,۶۶۳	۱۹۸,۸۸۴	۲۰۳,۲۹۸	۲۳۳,۶۶۷	۳۴,۷۸۳
۱۰	۱,۴۷۹	۳۱۱,۹۵۴	۲۹۶,۰۲۳	۴۴۸,۴۸۲	۱۳۶,۵۲۸
کل	۱۸,۸۱۶	۱۳۰,۳۲۸	۱۳۴,۷۴۱	۱۶۳,۳۱۱	۳۲,۹۸۳

جدول ۶: برآورد شاخص‌های موردنظر پس از برآزش مدل به داده‌ها در نقاط

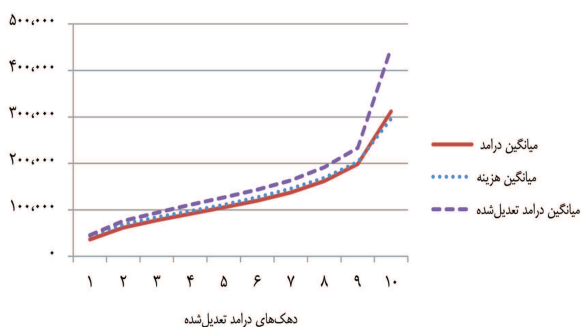
روستایی سال ۱۳۹۰ به تفکیک دهک درآمد تعدیل شده خانوارها (بر حسب هزار

ریال)

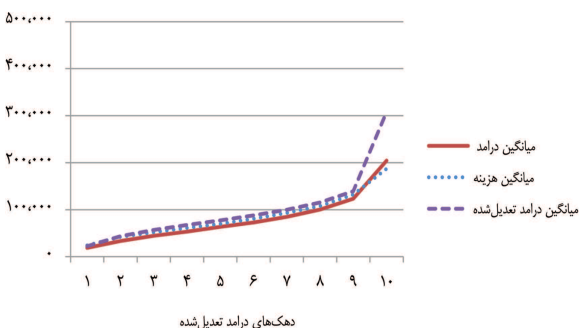
دهک درآمد	تعداد خانوارهای نمونه	میانگین درآمد	میانگین هزینه تعدیل شده	میانگین درآمد نمونه	میانگین کم‌گویی
۱	۲,۰۱۶	۱۸,۵۳۰	۲۴,۰۳۴	۲۲,۲۵۹	۳,۷۲۹
۲	۱,۹۱۸	۳۳,۵۱۴	۴۲,۰۳۲	۴۳,۶۲۱	۱۰,۱۰۷
۳	۱,۸۷۳	۴۴,۵۰۸	۵۲,۰۱۸	۵۶,۸۲۴	۱۲,۳۱۶
۴	۱,۹۶۷	۵۳,۳۶۵	۶۱,۴۵۹	۶۷,۴۴۹	۱۴,۰۸۴
۵	۱,۹۸۳	۶۳,۵۲۷	۶۹,۸۴۰	۷۷,۳۵۷	۱۳,۸۳۰
۶	۱,۹۷۵	۷۲,۸۸۲	۸۰,۶۳۴	۸۷,۷۷۰	۱۴,۸۸۸
۷	۲,۰۵۶	۸۴,۷۹۰	۹۳,۰۷۷	۹۹,۷۴۶	۱۴,۹۵۷
۸	۲,۰۱۷	۱۰۰,۱۰۵	۱۰۷,۸۱۶	۱۱۵,۲۴۰	۱۵,۱۳۵
۹	۲,۰۶۲	۱۲۳,۰۹۵	۱۳۱,۹۱۴	۱۳۹,۱۹۸	۱۶,۱۰۲
۱۰	۱,۸۹۰	۲۰۴,۵۶۹	۱۸۷,۰۰۷	۳۰۷,۸۹۵	۱۰۳,۳۲۶
کل	۱۹,۷۵۷	۷۹,۸۶۹	۸۴,۹۶۶	۱۰۱,۷۰۵	۲۱,۸۳۶

۹۰..... روشی برای برآورد خطای اندازه‌گیری درآمد

همان‌گونه که شکل‌های ۱ و ۲ نشان می‌دهند با تعدیل درآمد، میانگین درآمدها کمی بیش از میانگین هزینه‌ها در دهک‌های مختلف خواهد بود، اما در دهک آخر، روش تعدیل به کار رفته میانگین درآمد را به شدت افزایش می‌دهد که این امر به دلیل چوله بودن توزیع متغیرهای هزینه و درآمد رخ می‌دهد.



شکل ۱: میانگین هزینه، درآمد اظهار شده و تعدیل شده خانوارها در نقاط شهری به تفکیک دهک‌های درآمد تعدیل شده در سال ۱۳۹۰



شکل ۲: میانگین هزینه، درآمد اظهار شده و تعدیل شده خانوارها در نقاط روستایی به تفکیک دهک‌های درآمد تعدیل شده در سال ۱۳۹۰

ضریب همبستگی پیرسون بین درآمد اظهار شده و درآمد تعدیل شده خانوارها در نقاط شهری و روستایی کشور به تفکیک دهک‌های درآمد تعدیل شده در جدول ۷ و به تفکیک دهک‌های درآمد تعدیل شده سرانه خانوارها در جدول ۸ ارائه شده است. مقدار بیشتر ضریب همبستگی‌ها در جدول ۷ نشان می‌دهد با از بین بردن اثر

جدول ۷: ضریب همبستگی بین درآمد اظهار شده و تعدیل شده خانوارها به تفکیک دهک درآمد تعدیل شده خانوارها

دهک درآمد تعدیل شده	نقاط شهری	نقاط روستایی
۱	۰/۸۶	۰/۸۴
۲	۰/۴۵	۰/۴۵
۳	۰/۳۶	۰/۳۵
۴	۰/۳۳	۰/۲۷
۵	۰/۲۷	۰/۲۷
۶	۰/۲۸	۰/۲۸
۷	۰/۲۹	۰/۳۰
۸	۰/۳۵	۰/۲۹
۹	۰/۴۴	۰/۳۹
۱۰	۰/۲۷	۰/۴۳
کل	۰/۲۶	۰/۳۷

جدول ۸: ضریب همبستگی بین درآمد اظهار شده و تعدیل شده سرانه خانوارها به تفکیک دهک درآمد تعدیل شده سرانه خانوارها

دهک درآمد تعدیل شده سرانه	نقاط شهری	نقاط روستایی
۱	۰/۹۶	۰/۹۵
۲	۰/۹۵	۰/۹۵
۳	۰/۹۶	۰/۹۴
۴	۰/۹۵	۰/۹۴
۵	۰/۹۳	۰/۹۳
۶	۰/۹۳	۰/۹۴
۷	۰/۹۳	۰/۹۳
۸	۰/۹۱	۰/۹۱
۹	۰/۹۱	۰/۹۱
۱۰	۰/۲۸	۰/۴۳
کل	۰/۲۶	۰/۳۷

۹۲ روشی برای برآورد خطای اندازه‌گیری درآمد

تعداد اعضای خانوار بر دهک بندی خانوارها، رابطه خطی قوی تری بین درآمد اظهار شده و درآمد تعدیل شده خانوارها در دهک‌های سرانه درآمد تعدیل شده مشاهده می‌شود. در هر دو جدول ۷ و ۸ ضریب همبستگی بین درآمد اظهار شده و درآمد تعدیل شده خانوارهای دهک اول بیشترین مقدار را به خود اختصاص داده است.

۴ تحلیل خطای اندازه‌گیری درآمد با داده‌های نمونه‌های چرخشی یکسان

اداره سرشماری آمریکا در گزارش کیفیت آمارگیری درآمد و مشارکت در برنامه (SIPP) که یک آمارگیری پانلی است، وجود واحدهای نمونه یکسان در دوره‌های متوالی آمارگیرهای پانلی (یا چرخشی) را منبعی برای بروز انواعی از خطاهای اندازه‌گیری بر شمرده است که این خطاها در سایر آمارگیری‌ها رخ نمی‌دهند. به‌عنوان مثال می‌توان به شرطی شدن پاسخ‌گویان یا تغییر رفتار آن‌ها در طول زمان اشاره کرد که منجر به بروز خطای اندازه‌گیری در این آمارگیری‌ها می‌شود. با این وجود، ویژگی آمارگیرهای پانلی یا چرخشی فرصتی برای تشخیص برخی خطاهای اندازه‌گیری فراهم می‌آورد که این خطاها در آمارگیری‌های مقطعی قابل تشخیص نیستند. این امر از آنجایی ناشی می‌شود که خطاهای اندازه‌گیری می‌توانند اثری جدی بر داده‌های واحدهای نمونه یکسان در طول زمان بگذارند که اغلب قابل بررسی است.

همان‌گونه که در بخش دوم اشاره شد آمارگیری هزینه و درآمد خانوار شهری و روستایی ایران از سال ۱۳۸۹ با طراحی جدید به صورت یک آمارگیری چرخشی اجرا شده است. ماهیت چرخشی آمارگیری از هزینه و درآمد خانوار امکان بررسی خطای اندازه‌گیری را برای خانوارهای نمونه یکسان در دو سال ۱۳۸۹ و ۱۳۹۰ فراهم می‌آورد. از آنجایی که برای مراجعه به واحدهای نمونه این آمارگیری، از اطلاعات محل سکونت خانوارها (بر اساس چارچوب سرشماری سال ۱۳۸۵) استفاده شده است، در مواردی ممکن است به دلیل جابه‌جایی، خانوار ساکن در مکان واحد نمونه یکسان در سال‌های متوالی یکسان نباشد. بر اساس راهنمای

آمارگیری، در چنین مواردی از خانوار وارد شده به مکان، آمارگیری به عمل می آید. در این بررسی برای به دست آوردن برآوردهای دقیق تر، واحدهای نمونه یکسان در سال های ۱۳۸۹ و ۱۳۹۰ به واحدهایی از نمونه محدود شده اند که خانوار آمارگیری شده در سال ۱۳۹۰ در مکان آن واحد، همان خانوار آمارگیری شده در سال ۱۳۸۹ در آن مکان است. بدین ترتیب تعداد خانوارهای نمونه یکسان در سال های ۱۳۸۹ و ۱۳۹۰ (با درآمد مثبت) در نقاط شهری ۸۸۳۷ و در نقاط روستایی ۱۱۱۶۶ خانوار بود. برای بررسی خطای اندازه گیری درآمد خانوارهای نمونه یکسان، وزن نمونه گیری این خانوارها به گونه ای تعدیل شد که مجموع وزن های این خانوارها با تعداد خانوارهای پیش بینی شده سال ۱۳۹۰ برابر باشد. با فرض این که افزایش درآمد خانوارها از سال ۱۳۸۹ تا سال ۱۳۹۰ حداقل به اندازه افزایش هزینه های خانوارها در این یک سال بوده است، درآمد خانوار در سال ۱۳۹۰ از رابطه

$$D_{a-90} = \max(D_{1-90}, D_{90})$$

برآورد شد، که در آن

$$D_{1-90} = D_{89} + (GH_{90} - GH_{89})$$

و D_{90} و D_{89} درآمد اظهار شده خانوار به ترتیب در سال های ۱۳۸۹ و ۱۳۹۰ و GH_{89} و GH_{90} هزینه کل خانوار به ترتیب در سال های ۱۳۸۹ و ۱۳۹۰ است. نسبت درآمد برآورد شده به درآمد اظهار شده خانوارهای نمونه چرخشی یکسان از رابطه

$$P_{90} = \frac{D_{a-90}}{D_{90}}$$

به دست می آید، که معیاری برای تعیین میزان خطای اندازه گیری درآمد برای خانوارهای نمونه چرخشی یکسان است. جدول ۹ میانگین این نسبت را به تفکیک دهک های درآمد اظهار شده (وزنی) خانوارها در نقاط شهری و روستایی کشور نشان می دهد.

همان طور که ملاحظه می شود، مقدار میانگین نسبت درآمد برآورد شده به درآمد اظهار شده برای خانوارهای نمونه تکراری در نقاط شهری ۱/۲۴ و در نقاط روستایی

۹۴ روشی برای برآورد خطای اندازه‌گیری درآمد

جدول ۹: میانگین نسبت درآمد برآوردشده خانوارها به تفکیک دهک درآمد اظهارشده

دهک درآمد اظهار شده	نقاط شهری	نقاط روستایی
۱	۱/۵۸	۱/۲۹
۲	۱/۳۱	۱/۲۸
۳	۱/۲۴	۱/۲۵
۴	۱/۲۲	۱/۳۰
۵	۱/۱۵	۱/۳۲
۶	۱/۲۷	۱/۲۶
۷	۱/۲۰	۱/۳۱
۸	۱/۲۰	۱/۲۶
۹	۱/۱۱	۱/۳۵
۱۰	۱/۰۸	۱/۵۲
کل	۱/۲۴	۱/۳۱

۱/۳۱ است. به عبارت دیگر درآمد واقعی خانوارهای شهری به‌طور متوسط ۲۴ درصد و درآمد واقعی خانوارهای روستایی به‌طور متوسط ۳۱ درصد بیش از درآمد اظهارشده آن خانوارها است. از این رو انتظار می‌رود متغیر درآمد اظهارشده خانوارها به‌طور متوسط دارای خطای اندازه‌گیری بیش از ۲۰ درصد باشد.

مقدار بالای میانگین نسبت درآمد برآوردشده به درآمد اظهارشده خانوارها به‌ویژه برای دهک‌های پایین جامعه، که قدرت پس‌انداز بالایی ندارند، نشان می‌دهد خطای اندازه‌گیری درآمد حتی برای دهک‌های پایین خطای قابل توجهی است.

مرز دهک‌های توزیع درآمد اظهارشده و درآمد برآوردشده خانوارها در سال ۱۳۹۰ به‌همراه نسبت درآمد برآوردشده به درآمد اظهارشده خانوارهای نمونه تکراری در این سال به تفکیک نقاط شهری و روستایی به ترتیب در جدول ۱۰ ارائه شده است.

همان‌طور که ملاحظه می‌شود درآمد برآوردشده تقریباً نیمی از خانوارها به‌طور قابل توجهی بیش از درآمد اظهارشده آن‌ها است.

جدول ۱۰: توزیع متغیرهای درآمد برای خانوارهای نمونه چرخشی یکسان در نقاط شهری و روستایی در سال ۱۳۹۰

نام متغیر	دهک	نقاط شهری		نقاط روستایی	
		خطای استاندارد	برآورد	خطای استاندارد	برآورد
درآمد اظهارشده	۱	۴۸,۹۷۴,۳۴۰	۲۵,۴۱۰,۲۶۶	۵۲۴,۸۳۵	۲۵,۴۱۰,۲۶۶
	۲	۶۷,۴۷۱,۶۷۵	۳۷,۵۳۹,۲۲۷	۴۴۶,۱۵۹	۳۷,۵۳۹,۲۲۷
	۳	۸۲,۱۵۳,۵۹۳	۴۷,۷۱۹,۱۸۸	۴۲۶,۹۵۵	۴۷,۷۱۹,۱۸۸
	۴	۹۶,۲۳۶,۰۱۵	۵۷,۰۵۸,۵۳۶	۵۰۵,۲۲۰	۵۷,۰۵۸,۵۳۶
	۵	۱۱۱,۲۹۲,۴۹۱	۶۷,۱۹۸,۶۶۷	۶۱۹,۵۶۵	۶۷,۱۹۸,۶۶۷
	۶	۱۲۷,۸۷۵,۷۵۱	۷۸,۱۴۸,۷۵۰	۷۰۸,۵۰۷	۷۸,۱۴۸,۷۵۰
	۷	۱۴۸,۲۸۰,۸۳۴	۹۱,۷۴۹,۷۰۱	۸۷۰,۴۹۳	۹۱,۷۴۹,۷۰۱
	۸	۱۷۵,۴۶۶,۰۲۶	۱۱۱,۰۲۰,۷۰۰	۸۹۶,۴۰۱	۱۱۱,۰۲۰,۷۰۰
	۹	۲۲۶,۹۷۴,۲۴۰	۱۴۶,۷۲۱,۵۱۹	۲,۰۷۴,۶۵۱	۱۴۶,۷۲۱,۵۱۹
درآمد برآوردشده	۱	۵۸,۶۰۴,۶۱۰	۳۲,۲۸۴,۱۹۶	۴۹۳,۶۲۷	۳۲,۲۸۴,۱۹۶
	۲	۷۷,۴۹۵,۵۷۶	۴۶,۱۳۰,۶۲۹	۴۹۴,۲۸۷	۴۶,۱۳۰,۶۲۹
	۳	۹۴,۱۸۰,۵۱۶	۵۶,۹۲۰,۳۲۰	۵۲۵,۹۱۲	۵۶,۹۲۰,۳۲۰
	۴	۱۱۰,۱۹۵,۳۹۱	۶۷,۶۰۳,۳۶۵	۵۸۷,۵۳۴	۶۷,۶۰۳,۳۶۵
	۵	۱۲۷,۴۶۷,۹۵۵	۷۸,۹۷۵,۷۴۹	۷۰۴,۰۹۶	۷۸,۹۷۵,۷۴۹
	۶	۱۴۸,۳۶۹,۶۷۷	۹۱,۷۳۶,۷۴۳	۸۷۱,۰۴۷	۹۱,۷۳۶,۷۴۳
	۷	۱۷۲,۶۷۰,۲۵۵	۱۰۸,۶۷۲,۶۹۵	۱,۰۸۵,۵۰۳	۱۰۸,۶۷۲,۶۹۵
	۸	۲۰۸,۳۰۳,۸۹۵	۱۳۲,۳۱۹,۸۸۲	۱,۴۷۹,۷۷۵	۱۳۲,۳۱۹,۸۸۲
	۹	۲۶۷,۷۰۹,۰۱۵	۱۸۰,۱۹۸,۶۳۱	۲,۲۹۳,۴۶۲	۱۸۰,۱۹۸,۶۳۱

۵. تعدیل اجزای درآمد

درآمد خانوارها در طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی از سه بخش تشکیل می شود:

بخش ۱- درآمد پولی اعضای شاغل خانوار از مشاغل مزد و حقوق بگیری،

بخش ۲- درآمد پولی اعضای شاغل خانوار از مشاغل غیر مزد و حقوق بگیری (آزاد)،

بخش ۳- درآمدهای متفرقه خانوار در ۱۲ ماه گذشته.

علاوه بر این در بخش هزینه های خانوار، مصرف کالاهای خوراکی یا غیر خوراکی خانوار که روش تهیه آنها در برابر خدمت عمومی، تعاونی یا خصوصی یا از محل کسب کشاورزی یا غیر کشاورزی یا رایگان (نه از خانوار دیگر) باشد درآمد غیر پولی خانوار محسوب می شود. بدین ترتیب اجزای درآمد اظهارشده خانوارها عبات هستند از

- درآمد خالص مزد و حقوق بگیری بخش دولتی (پولی)

- درآمد خالص مزد و حقوق‌بگیری بخش دولتی (غیر پولی)
- درآمد خالص مزد و حقوق‌بگیری بخش خصوصی (پولی)
- درآمد خالص مزد و حقوق‌بگیری بخش خصوصی (غیر پولی)
- درآمد خالص مزد و حقوق‌بگیری بخش تعاونی (پولی)
- درآمد خالص مزد و حقوق‌بگیری بخش تعاونی (غیر پولی)
- درآمد از مشاغل آزاد کشاورزی (پولی)
- درآمد از مشاغل آزاد کشاورزی (غیر پولی)
- درآمد از مشاغل آزاد غیر کشاورزی (پولی)
- درآمد از مشاغل آزاد غیر کشاورزی (غیر پولی)
- درآمدهای متفرقه (پولی - شامل یارانه)
- درآمدهای متفرقه (غیر پولی)

هر یک از اجزای درآمد بسته به مورد ممکن است برای یک خانوار دارای اطلاع یا سفید باشد. درآمد اظهارشده‌ی خانوارها از جمع درآمد خانوارها در هر یک از اجزای آن به دست می‌آید و خطای اندازه‌گیری می‌تواند در اندازه‌گیری هر یک از اجزای درآمد رخ دهد، از این رو با تعدیل درآمد لازم است مقدار هر یک از اجزای درآمد نیز تعدیل شود. از آنجایی که انتظار می‌رود هنگام تکمیل پرسشنامه‌ها در آمارگیری از هزینه و درآمد خانوارها تحقیق کافی برای شناسایی و ثبت منبع درآمد از جمله درآمدهای متفرقه‌ی خانوارها (شامل دریافت یارانه نقدی) صورت گرفته باشد در این بررسی فرض می‌شود با تعدیل درآمد، منبع (یا منبع‌های) ثبت شده درآمد تغییر نمی‌کند. به عبارت دیگر در صورتی که مقدار درآمد خانوار در یکی از اجزای آن صفر باشد روش به کار رفته برای تعدیل درآمد، مقدار درآمد از این منبع را تغییر نمی‌دهد. با این فرض برای تعدیل هر یک از اجزای درآمد خانوارها در سال ۱۳۹۰، ابتدا نرخ تغییر درآمد هر خانوار پس از تعدیل درآمد به صورت

$$r_{90} = \frac{D_{adj}^{90}}{D^{90}}$$

محاسبه شد، که در آن D_{adj}^{90} درآمد تعدیل شده خانوار و D^{90} درآمد اظهارشده خانوار در سال ۱۳۹۰ است. سپس مقدار هر یک از اجزای درآمد خانوار در این نرخ

ضرب شد و مقدار تعدیل شده اجزای درآمد به دست آمد. بدین ترتیب اجزای درآمد به گونه‌ای تعدیل شدند که جمع مقادیر آنها همان درآمد تعدیل شده خانوارها در سال ۱۳۹۰ باشد. جدول ۱۱ میانگین اجزای درآمد اظهار شده و درآمد تعدیل شده خانوارها را به تفکیک نقاط شهری و روستایی کشور در سال ۱۳۹۰ نشان می‌دهد. بر اساس اطلاعات این جدول، کمترین مقدار تغییر نسبی حاصل از تعدیل درآمد در نقاط شهری به درآمد غیر پولی از حقوق‌بگیری بخش خصوصی و در نقاط روستایی به درآمد غیر پولی از حقوق‌بگیری بخش عمومی اختصاص دارد. همچنین درآمد پولی از مشاغل آزاد کشاورزی در نقاط شهری و درآمدهای متفرقه در نقاط روستایی بیشترین مقدار تغییر نسبی حاصل از تعدیل درآمد را به خود اختصاص داده است.

۶ بحث و نتیجه‌گیری

پس از تعدیل درآمد به روشی که اشاره شد میانگین درآمد سالانه خانوارهای شهری به‌طور تقریبی ۲۵ درصد و میانگین درآمد سالانه خانوارهای روستایی به‌طور تقریبی ۲۷ درصد افزایش می‌یابد.

نتایج تحقیق نشان داد نسبت درآمد تعدیل شده به درآمد اظهار شده، برای خانوارهای دهک‌های پایین بیشتر از خانوارهای دهک‌های بالا است. این نسبت برای خانوارهای روستایی بیشتر از خانوارهای شهری است.

کمترین مقدار تغییر نسبی حاصل از تعدیل درآمد در نقاط شهری به درآمد غیر پولی از حقوق‌بگیری بخش خصوصی و در نقاط روستایی به درآمد غیر پولی از حقوق‌بگیری بخش عمومی اختصاص دارد. همچنین درآمد پولی از مشاغل آزاد کشاورزی در نقاط شهری و درآمدهای متفرقه در نقاط روستایی بیشترین مقدار تغییر نسبی حاصل از تعدیل درآمد را به خود اختصاص داده است.

لازم به ذکر است که بر اساس نتایج حاصل از این بررسی، خطای اندازه‌گیری درآمد در ایران در مقایسه با سایر کشورها مقدار قابل قبولی است با این حال برای کاهش آن همچنان باید فعالیت‌های لازم انجام شود. بدیهی است نمی‌توان انتظار داشت که مقدار خطای اندازه‌گیری درآمد به صفر برسد.

جدول ۱: میانگین اجزای درآمد اظهارشده و تعدیل‌شده خانوارها به تفکیک نقاط شهری و روستایی سال ۱۳۹۰

نوع	منبع درآمد	نقاط شهری			نقاط روستایی		
		درآمد اظهارشده	درآمد تعدیل‌شده	درآمد اظهارشده	درآمد تعدیل‌شده	درآمد تعدیل‌شده	
حقوق‌گیری	بخش عمومی	۱۳,۱۳۵,۷۵۱	۱۵,۵۴۴,۶۳۱	۳,۸۷۱,۶۷۵	۴,۳۰۲,۸۶۲	۱۵,۹۲۷,۲۵۳	
	پولی	۱۴۰,۰۱۰	۱۶۶,۸۰۷	۷۶,۵۹۲	۸۹,۴۸۹	۱۰,۹۹۷,۴۱۵	
	بخش تعاونی	۱۹,۸۷۸,۵۱۰	۲۳,۸۷۸,۸۳۶	۱۵,۶۶۷,۶۷۷	۱۸,۲۷۲,۰۱۸	۱,۰۳۴,۸۳۳	
	بخش خصوصی	۲,۷۹۵,۸۲۱	۳,۲۸۶,۰۰۷	۸۴۱,۲۴۷	۹۰۷,۵۹۹	۲۴,۶۲۰,۸۴۶	
	غیر پولی	۲۵,۴۸۷	۳۲,۳۷۲	۱۲,۵۲۸	۱۴,۰۴۵	۲۴,۶۲۰,۸۴۶	
بخش تعاونی	۲,۳۲۹,۹۰۵	۲,۷۰۱,۲۸۴	۹۳۴,۷۹۸	۱,۰۳۴,۸۳۳	۲۴,۶۲۰,۸۴۶		
بخش خصوصی	۳۸,۳۰۵,۴۸۵	۴۵,۶۰۹,۹۳۸	۲۱,۴۰۴,۵۱۶	۲۴,۶۲۰,۸۴۶	۲۴,۶۲۰,۸۴۶		
کل		۱,۳۵۷,۰۲۱	۳,۳۹۲,۷۲۹	۱۱,۷۴۶,۳۲۱	۱۵,۹۲۷,۲۵۳	۱۵,۹۲۷,۲۵۳	
مشاغل آزاد	کشاورزی	۱۸,۳۴۳,۵۷۸	۲۴,۲۶۸,۸۶۰	۸,۸۸۵,۷۲۵	۱۰,۹۹۷,۴۱۵	۱۰,۹۹۷,۴۱۵	
	غیرکشاورزی	۵۲,۳۲۲	۶۶,۱۵۶	۱,۲۸۳,۵۴۱	۱,۴۷۴,۳۰۱	۱,۴۷۴,۳۰۱	
	کشاورزی	۳۲۱,۶۷۵	۳۸۴,۷۶۸	۳۲۰,۸۷۳	۳۶۷,۲۳۵	۳۶۷,۲۳۵	
کل		۲۰,۰۷۴,۵۹۶	۲۸,۱۱۲,۵۱۳	۲۲,۲۳۶,۴۶۰	۲۸,۷۶۶,۲۰۳	۲۸,۷۶۶,۲۰۳	
متفرقه	کشاورزی	۴۳,۰۱۶,۲۵۶	۵۴,۹۶۴,۵۸۲	۲۵,۷۰۶,۸۹۸	۳۵,۳۲۳,۸۸۱	۳۵,۳۲۳,۸۸۱	
	غیرکشاورزی	۲۸,۹۳۱,۵۹۵	۳۴,۶۲۴,۳۰۱	۱۰,۵۲۰,۸۱۲	۱۲,۹۹۳,۹۳۳	۱۲,۹۹۳,۹۳۳	
	کشاورزی	۷۱,۹۴۷,۸۵۱	۸۹,۵۸۸,۸۸۴	۳۶,۲۲۷,۷۱۰	۴۸,۳۱۷,۸۱۴	۴۸,۳۱۷,۸۱۴	
کل		۱۳۰,۳۲۷,۹۳۲	۱۶۳,۳۱۱,۳۳۴	۷۹,۸۶۸,۶۸۵	۱۰۱,۷۰۴,۸۶۳	۱۰۱,۷۰۴,۸۶۳	
درآمد کل							

از سوی دیگر، بر اساس ویژگی‌های آمارگیری از هزینه و درآمد خانوار در ایران، انتظار می‌رود خطای اندازه‌گیری هزینه‌ها کمتر از خطای اندازه‌گیری درآمد باشد. زیرا در طرح آمارگیری هزینه و درآمد خانوار، هزینه‌ها به صورت جزئی‌تر پرسش می‌شوند. بنابراین می‌توان انتظار داشت خطای اندازه‌گیری هزینه کل خانوارها کمتر از ۲۵ درصد است.

در پایان، با توجه به این که در حال حاضر بسیاری از شاخص‌های اقتصادی مانند ضریب جینی، خط فقر و سهم دهک‌ها از درآمد کل کشور بر اساس هزینه‌ی خانوارها محاسبه می‌شوند، پیشنهاد می‌شود این شاخص‌ها بر اساس درآمد تعدیل شده‌ی خانوارها محاسبه شوند تا نتایج درست‌تری به دست آیند.

تقدیر و تشکر

نویسندگان مقاله از داوران محترم نشریه، که با رهنمودهای ارزنده خود باعث بهبود مقاله شده‌اند، کمال تشکر و قدردانی را دارند.

مرجع‌ها

جانسون، ر. آ. و ویچرن، د. د. (۱۹۸۲)، تحلیل آماری چندمتغیره کاربردی، نیرومند، ح. ع. (مترجم)، دانشگاه فردوسی مشهد.

چکیده‌ی نتایج طرح آمارگیری هزینه و درآمد خانوارهای شهری و روستایی سال ۱۳۹۰، مرکز آمار ایران، تهران.

دستورالعمل طرح نمونه‌گیری آمارگیری هزینه و درآمد خانوار شهری / روستایی - سال ۱۳۹۰، مرکز آمار ایران، تهران.

زاهدیان، ع. ر.، علی اکبری صبا، ر. مصطفوی، ن. و اربابی، م. (۱۳۹۲)، برآورد خطای اندازه‌گیری درآمد در آمارگیری هزینه و درآمد خانوار، مرکز آمار ایران، پژوهشکده آمار، تهران.

۱۰۰ روشی برای برآورد خطای اندازه‌گیری درآمد

گروز، ر. م.، فولر، ف. ج.، کوپر، م. پ.، لپکوسکی، ج. م.، سینگر، ا. و تورانگیو، ر. (۲۰۰۴)، روش‌شناسی آمارگیری، صالحی، م. و جمال‌زاده، م. ا. (مترجمان)، پژوهشکده آمار، تهران.

Aguero, J., Carter, M. R. and May, J. (2007), Poverty and Inequality in the First Decade of South Africa's Democracy: What can be Learnt from Panel Data from KwaZulu-Natal?, *Journal of African Economies*, **16**, 782-812.

Bound, J., Brown, C., Duncan, G. and Rodgers, W. (1990), *Measurement Error in Cross- Sectional and Longitudinal Labor Market Surveys: Validation Study Evidence*, In Hartog, J., G. Ridder, and J. Theeuwes (Eds.) *Panel Data and Labor Market Studies*, Amsterdam.

Grondin, C. and Michaud, S. (1994), Data Quality of Income Data Using Computer- Assisted Interview: The Experience of the Canadian Survey of Labour and Income Dynamics, Proceedings of the Survey Research Methods Section, *American Statistical Association*, 830-835.

Hurts, E., Li, G. and Pugsley, B. (2012), Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self- Employed, *The Review of Economics and Statistics*, **96**, 19-33.

Moore, J., Stinson, L. and Welniak, J. E. J. (1997), Income Measurement Error in Surveys: A Review, *Journal of Official Statistics*, **16**, 331-362.

Neri, A. and Zizza, R. (2010), Income Reporting Behaviour in Sample Surveys, Temi di Discussion (Working Paper), Available at: www.bancaditalia.it/pubblicazioni/econo/temidi/td10/td777_10/en_td.777_10/en_tema.777.pdf.