Archive of SID

مجلهٔ پژوهشی دانشگاه اصفهان (علوم پایه)
جلد سی و سوم ـ شماره ٤ـ سال۱۳۸۷
صص ١٤-١

مدل‌سازی خطای پاسخ در آمارگیری‌ها به شیوه‌ی مصاحبه‌ی حضوری

روشنک علی‌محمدی* و حمیدرضا نواب‌پور**

***گروه آمار دانشگاه الزهرا**

****گروه آمار دانشگاه علامه طباطبایی**

چکیده

آمارگیری یکی از روش‌های متداول تولید داده‌ها است. کیفیت نتایج حاصل از آمارگیری‌ها به‌دلیل کاربرد آن‌ها در تصمیم‌گیری‌ها، برنامه‌ریزی‌ها و سیاست‌گذاری‌ها دارای اهمیت ویژه‌ای است. برای تعیین میزان کیفیت داده‌ها لازم است که خطای نمونه‌گیری و خطای غیرنمونه‌گیری برآورد شوند. در حالت کلی برآورد خطای غیرنمونه‌گیری که علاوه بر خطای پاسخ شامل مولفه‌های خطای بی‌پاسخی، خطای پوشش و خطای پردازش داده‌ها است، با توجه به مساله کمی‌سازی این مولفه‌ها پیچیده است. خطای پاسخ بخش قابل کمی شدن در خطای غیرنمونه‌گیری است، لذا مدل‌سازی خطای پاسخ بر اساس شیوه‌ی اجرای آمارگیری‌ها دارای اهمیت ویژه‌ای است. در این مقاله ابتدا، با بررسی نحوه‌ی اجرای طرح‌های آماری عمده‌ای که به‌روش مصاحبه‌ی حضوری اجرا می‌شوند، عوامل موثر در ایجاد خطای پاسخ معرفی شده و سپس با ارائه‌ی یک طرح آزمایش، مدل مناسبی برای خطای پاسخ در آمارگیری‌ها به شیوه‌ی مصاحبه‌ی حضوری پیشنهاد شده است. علاوه بر این نحوه‌ی محاسبه‌ی دقت در آمارگیری‌های دارای طرح‌های نمونه‌گیری دومرحله‌ای مورد بررسی قرار گرفته و در این حالت، برآوردگری برای محاسبه‌ی دقت برآوردگرها پیشنهاد شده است. به‌منظور ارائه‌ی کاربردی از مدل پیشنهادی، دقت نتایج حاصل در آمارگیری از طرح هزینه و درآمد خانوار با استفاده از برآوردگر حاصل و با به‌کارگیری داده‌های سال ۱۳۸۲ محاسبه شده است.

واژه‌های کلیدی: کیفیت داده‌ها، خطای نمونه‌گیری، خطای غیرنمونه‌گیری، خطای پاسخ، خطای بی‌پاسخی، خطای پردازش داده‌ها، دقت برآوردگر، واریانس پرسش‌گر، برآوردگر گروه‌های تصادفی، خطای نسبی.

www.SID.ir

Archive of SID

Response Error Modeling in face-to-face Interview Surveys

R. Alimohammadi⁎ and H. Navvabpour⁎⁎

⁎Assistant Professor of Statistics, Alzahra University

⁎⁎Associate Professor of Statistics, Allameh Tabatabaie University

Abstract

Survey implementation is one of the common methods to data collection. Quality of survey results is of prime importance due to their applications in decision making and planning. To determine quality of survey results, it is necessary to estimate both sampling and nonsampling errors. However, estimation of nonsampling error is difficult because of problem of quantifying other parts of nonsampling error that is nonresponse error, Coverage error and data processing error. Response error is a quantifiable part of nonsampling error. Therefore, modeling of the response error has received a particular attention over years. In this paper, we consider some major statistical surveys which are conducted in a face to face interview and sources of error are determined, then a model of response error in face to face surveys has been proposed based on presentation of an experimental design. Furthermore, computation of precision in surveys with two-stage sampling designs is assessed and an estimator of precision is presented in this case. As an application of proposed model, precision of survey estimators of interest is computed based on resulted estimator, for data of household income and expenditure survey for the year 1382.

Keywords: Data quality, Sampling error, Nonsampling error, Response error, Nonresponse error, Coverage error, Data Proccessing error, Precision of estimate, Interviewer variance, Random group estimator, Relative error.

مقدمه

وجود اطلاعات آماری برای هر نوع برنامه‌ریزی، تصمیم‌گیری و سیاست‌گذاری ضروری است. آمارگیری یکی از شیوه‌های تولید اطلاعات است که خود شامل دو روش سرشماری و نمونه‌گیری است. نتایج آمارگیری‌ها وقتی می‌توانند به تصمیم‌گیری‌های موفق منجر شوند که داده‌ها (اطلاعات گردآوری شده) دارای کیفیت بالا باشند. کیفیت داده‌ها دارای چند مولفه است که دقت[1] بخش مهمی از آن‌ها است. دقت رابطه‌ای عکس با واریانس کل در آمارگیری‌ها دارد، بنابراین برای تعیین میزان دقت داده‌ها لازم است واریانس کل برآورد(های) حاصل از آمارگیری‌ها تعیین شود. به منظور بررسی میزان دقت لازم است خطای کل در آمارگیری‌ها مورد نظر قرار گیرد. خطای کل شامل خطای نمونه‌گیری و خطای غیرنمونه‌گیری است.

خطای نمونه‌گیری تنها در نمونه‌گیری‌ها امکان بروز دارد. خطای غیرنمونه‌گیری شامل خطای پوشش، خطای بی‌پاسخی، خطای پاسخ و خطای پردازش داده‌ها است. برای تعیین میزان خطای کل علاوه بر برآورد خطای نمونه‌گیری، نیاز به کمی کردن خطای غیرنمونه‌گیری نیز دارد که در عمل ممکن است به دلیل مشکلات کمی ساختن سایر مولفه‌های خطا پیچیده باشد. خطای پاسخ،

1- Precision

www.SID.ir

Archive of SID

تفاضل مقدار واقعی و مقدار مشاهده شده‌ی صفت مورد نظر است که به آن خطای اندازه‌گیری نیز گفته می‌شود. این خطا بخش قابل کمی شدن در خطای غیرنمونه‌گیری است.

در این مقاله با بررسی نحوه‌ی اجرای طرح‌های آماری عمده‌ای که به شیوه‌ی مصاحبه‌ی حضوری در ایران اجرا می‌شوند، منبع‌های ایجاد خطای پاسخ بررسی شده و سپس بر مبنای یک طرح آزمایش مناسب، مدلی برای خطای پاسخ در آمارگیری‌های حضوری پیشنهاد می‌شود.

مدل‌سازی خطای پاسخ دارای کاربردهای زیادی است که یکی از آن‌ها برآورد میزان دقت در آمارگیری‌ها است. بنا بر اهمیت مدل‌سازی خطای پاسخ، محققان بسیاری به مطالعه در این زمینه پرداخته‌اند. به‌طورکلی دو شیوه‌ی مدل‌سازی برای خطای پاسخ وجود دارد که شامل مدل دفتر سرشماری[1] و مدل تحلیل واریانس هستند. هانسن، هورویتز و برشاد (۱۹۶۱) مدل دفتر سرشماری را ارائه کردند و بعدها فلگی (۱۹۷۴) آن را توسعه داد. این مدل در لسلر و کالس‌بیک (۱۹۹۲) نیز بیان شده است. برای مدل‌سازی خطای پاسخ به‌روش تحلیل واریانس نیز مرجع‌هایی موجود است که در بخش ۲ معرفی شده‌اند.

در این مقاله از روش تحلیل واریانس برای مدل‌سازی خطای پاسخ در آمارگیری‌هایی که داده‌های آن‌ها به شیوه‌ی مصاحبه‌ی حضوری گردآوری می‌شوند، استفاده می‌شود. مروری بر مطالعات انجام شده در زمینه‌ی مدل‌سازی خطای پاسخ در بخش ۲ آمده است. یکی از کاربردهای مدل خطای پاسخ در برآورد واریانس کل (و بنابراین در برآورد دقت) برآوردگر است، لذا برای آمارگیری‌های مورد نظر، اثبات تجزیه واریانس کل به مجموع واریانس‌های نمونه‌گیری و غیرنمونه‌گیری در بخش ۳ ارائه شده است (لاندکوئیست و رتمن (۲۰۰۲) این رابطه را در حالت نمونه‌گیری تصادفی ساده اثبات کردند). هدف از بخش ۴ ارائه‌ی مدلی مناسب برای خطای پاسخ در آمارگیری‌های حضوری است. برای طراحی مدل خطای پاسخ لازم بود که شرایط و نحوه‌ی اجرای این آمارگیری‌ها به‌دقت بررسی شود.

بدین منظور فرایند گردآوری و ثبت داده‌ها در طرح‌های آماری عمده در ایران را مورد مطالعه و بررسی قرار داده و پس از شناسایی عوامل موثر در بروز خطای پاسخ با ارائه‌ی یک طرح آزمایش، مدل خطای پاسخ در آمارگیری‌های حضوری را پیشنهاد کرده‌ایم(علی‌محمدی (۱۳۸۵).

به‌منظور برآورد دقت لازم است واریانس اثرهای تصادفی مدل خطای پاسخ برآورد شوند. در بخش ۵ برآوردگری برای برآورد دقت آماره‌های آمارگیری ارائه شده است. در بخش ۶ به‌منظور ارائه‌ی کاربردی از مدل پیشنهادی، داده‌های طرح هزینه و درآمد خانوار شهر تهران در سال ۱۳۸۲ به‌کارگرفته شد و دقت میانگین سه صفت درآمد کل[2]، هزینه‌ی خالص[3] و هزینه‌ی

1- Census Bureau Model

۲- کل درآمد خانوار در سال مورد نظر، درآمد کل گفته می‌شود.

۳- منظور از هزینه‌ی خالص، هزینه‌های خانوار در آن سال با کسر ارزش دست دوم لوازم خریداری‌شده است.

www.SID.ir

Archive of SID

ناخالص[1] خانوار، با استفاده از برآوردگر ارائه‌شده در بخش ۵ محاسبه شدند.

۲. مروری بر مطالعات انجام شده در مورد مدل‌سازی خطای پاسخ

ماهالانوبیس (۱۹۴۶) برای اولین بار مدل تحلیل واریانس را در مدل‌سازی خطای پاسخ به‌کار برد. کیش (۱۹۶۲) در مدل خود، خطای پاسخ را به مجموع دو مولفه‌ی خطای پرسش‌گر و خطای سایر عوامل تجزیه کرد. هارتلی و رائو (۱۹۷۸) مدلی را ارائه کردند که خطای پرسش‌گر، خطای کدگذار، اثر متقابل پرسش‌گر و مشاهده و اثر متقابل کدگذار و مشاهده را در بر دارد. این مدل به‌صورت زیر است:

$$y = m + B + D + dB + dD \qquad (۱)$$

که در آن مقدار مشاهده‌شده (y) تابعی از مقدار واقعی (m)، خطای پرسش‌گر (B)، خطای کدگذار (D)، اثر متقابل پرسش‌گر و مشاهده (dB) و اثر متقابل کدگذار و مشاهده (dD) است.

هارتلی (۱۹۸۱) مدلی با مولفه‌های خطای پرسش‌گر، خطای کدگذار و خطای پاسخ‌گو را پیشنهاد کرد. بی‌مر و استاکس (۱۹۹۱) مدلی را برای یک تک‌مشاهده، که در آن مقدار برآوردشده تابعی از مقدار واقعی و اریبی اندازه‌گیری است را پیشنهاد کردند.

باسی و فابریس (۱۹۹۷) مدلی شامل مولفه‌های خطای کارشناس مسوول، خطای پرسش‌گر و خطای پاسخ‌گو را ارائه کردند.

بی‌مر و تروین (۱۹۹۷) دو دسته مدل خطای پاسخ را برای داده‌های پیوسته و داده‌های دوحالتی بیان کردند. برای داده‌های پیوسته مدل زیر را در نظر گرفتند:

$$y_i = m_i + d_i \qquad (۲)$$

که در آن y_i مقدار مشاهده‌شده‌ی واحد iام، m_i مقدار واقعی واحد iام و d_i خطایی است که به فرایند اندازه‌گیری بستگی دارد و همان خطای پاسخ (یا خطای اندازه‌گیری) است.

در مورد داده‌های دوحالتی فرض شده است که سوال مورد نظر می‌تواند دارای پاسخ مثبت یا منفی باشد. لذا در صورتی که پاسخ بیان‌شده مثبت باشد، مقدار مشاهده‌شده (y_i) برابر ۱ و در غیر این صورت برابر صفر قرار می‌گیرد. مقدار واقعی پاسخ سوال مورد نظر (m_i) نیز در صورت مثبت بودن آن برابر ۱ و در غیر این صورت برابر صفر قرار خواهد گرفت.

احتمال آن که پاسخ مثبت باشد به صورت: $p = P(m_i = 1)$ تعریف می‌شود و همچنین

$$q_i = P(y_i = 0 \mid m_i = 1)$$

$$f_i = P(y_i = 1 \mid m_i = 0)$$

با در نظر گرفتن مدلی مشابه رابطه‌ی (۲) توزیع خطای پاسخ (d_i) عبارت است از:

۱- هزینه‌ی ناخالص خانوار، کل هزینه‌های خانوار در سال مورد نظر بدون احتساب ارزش دست دوم لوازم خریداری شده است.

www.SID.ir

Archive of SID

جدول توزیع احتمال خطای d_i

y_i	m_i	d_i	$P(y_i \mid m_i)$
۱	۱	۰	$1-q_i$
۰	۱	-۱	q_i
۱	۰	۱	f_i
۰	۰	۰	$1-f_i$

لاندکوئیست و رتمن (۲۰۰۲) مدل خطای پاسخ را در حالت پیوسته بودن داده‌ها (مشابه مدل (۲)) در نظر گرفتند. آن‌ها دو منبع خطای پرسش‌گر و خطای پاسخ‌گو را در مدل منظور کردند. هدف مقاله‌ی آن‌ها برآورد واریانس پرسش‌گر بوده است.

آیهان (۲۰۰۳) مدلی برای خطای پاسخ در آمارگیری‌های با دو مرحله‌ی مصاحبه و مصاحبه‌ی مجدد شامل مولفه‌های خطای کارشناس مسوول، خطای پرسش‌گر و خطای پاسخ‌گو را پیشنهاد کرد. گراوز (۲۰۰۴) به بررسی خطاهای آمارگیری پرداخته است. گراوز و همکاران (۲۰۰۴) ضمن بررسی خطای پاسخ، راهبردهایی را برای کاهش واریانس پرسش‌گر ارائه کرده‌اند.

۳. تجزیه‌ی واریانس کل به واریانس‌های نمونه‌گیری و غیرنمونه‌گیری

از آنجا که اکثر طرح‌های آماری عمده به‌صورت خانواری اجرا می‌شوند و طرح نمونه‌گیری در آن‌ها خوشه‌ای دومرحله‌ای است، لذا برآورد دقت در چنین طرح‌هایی مورد بررسی قرار گرفته است. در این طرح‌ها بلوک‌ها در مناطق شهری و آبادی‌ها در مناطق روستایی حکم خوشه‌ها در مرحله‌ی اول نمونه‌گیری را دارند و خانوارهای داخل هر بلوک، واحدهای نمونه‌گیری مرحله‌ی دوم هستند.

برای توضیح نحوه‌ی به‌کارگیری مدل خطای پاسخ در تجزیه‌ی واریانس کل برآوردگر میانگین، حالت زیر را در نظر می‌گیریم.

فرض کنید یک نمونه‌ی دومرحله‌ای از جامعه‌ای استخراج شود. در مرحله‌ی اول، نمونه‌ای تصادفی به اندازه‌ی n خوشه از جامعه‌ای شامل N خوشه استخراج می‌شود. هر خوشه شامل M واحد است. در مرحله‌ی دوم نمونه‌گیری از هر خوشه m واحد به‌طور تصادفی انتخاب می‌شود. استفاده از بخشی از جامعه برای برآورد پارامتر جامعه، سبب بروز خطای نمونه‌گیری می‌شود.

اگر فرض شود که صفت مورد نظر از تمام افراد نمونه اندازه‌گیری شده باشد، مقدار خطای بی‌پاسخی صفر می‌شود. در صورتی که چارچوب نمونه‌گیری، چارچوبی مناسب و منطبق بر جامعه‌ی هدف باشد، مقدار خطای پوشش نیز صفر می‌شود. همچنین با توجه به پیشرفت فناوری، امکان حذف خطای پردازش داده‌ها نیز وجود دارد، بنابراین خطای پاسخ تنها خطای غیر قابل اجتناب در بین خطاهای غیرنمونه‌گیری است. به‌عبارت دیگر خطای پاسخ برای واحدهایی قابل محاسبه است که سایر خطاهای غیرنمونه‌گیری برای آن‌ها رخ نداده باشد.

برای محاسبه‌ی واریانس کل برآوردگر، لازم است که به دو مرحله‌ی تصادفیدن در آمارگیری‌ها توجه شود:

www.SID.ir

Archive of SID

۱- انتخاب نمونه‌ی تصادفی (بدون جایگذاری) طبق طرح نمونه‌گیری مورد نظر از جامعه و تخصیص آن به پرسش‌گرها (این مرحله با نماد p نشان داده می‌شود).

۲- مدل خطای پاسخ (برای این مرحله از نماد c استفاده می‌شود).

یک مدل خطای پاسخ را به‌صورت زیر در نظر می‌گیریم:

$$y_i = m_i + D_l + R_i \qquad (۳)$$

که در آن مقدار مشاهده‌شده‌ی y_i، $(i = 1,...,n)$، تابعی از مقدار واقعی m_i، خطای پرسش‌گر D_l، $(l = 1,...,L)$ و خطای پاسخ‌گو (R_i) است. فرض می‌شود که اثر پرسش‌گر (D) متغیری تصادفی با میانگین صفر و واریانس s_D^2 و توزیع اثر تصادفی پاسخ‌گو (R) دارای میانگین صفر و واریانس σ_R^2 هستند.در نمونه‌گیری تصادفی دو مرحله‌ای مورد نظر، برای برآورد میانگین مقادیر واقعی جامعه ($\overline{m} = \frac{1}{NM}\sum_{a=1}^{N}\sum_{b=1}^{M} m_{ab}$)، لازم است از برآوردگر $\overline{m}_s = \frac{1}{nm}\sum_{a=1}^{n}\sum_{b=1}^{m} m_{ab}$ که میانگین مقادیر واقعی در نمونه‌ی s است استفاده شود، اما به‌دلیل وجود خطای پاسخ و در دسترس نبودن مقادیر واقعی معمولاً برآوردگر $\overline{y} = \frac{1}{nm}\sum_{a=1}^{n}\sum_{b=1}^{m} y_{ab}$ برای برآورد $\overline{m}$ به‌کار می‌رود.

فرض می‌شود که مقادیر مشاهده‌شده برای پاسخ‌گویان منتسب به پرسش‌گرهای مختلف با هم ناهمبسته‌اند و تنها اندازه‌ی پاسخ‌گویان منتسب به یک پرسش‌گر خاص می‌توانند به هم وابسته باشند.

هر پرسش‌گر با تعدادی از پاسخ‌گویان در خوشه‌های تخصیص‌یافته به وی مصاحبه می‌کند. یعنی پاسخ‌گویان هر خوشه تنها با یک پرسش‌گر مصاحبه می‌شوند و تعداد خوشه‌های منتسب به هر پرسش‌گر برابر با h و n=hL است(به‌عبارت دیگر خوشه‌های نمونه‌ی s به L زیرنمونه‌ی $s_1, s_2, ..., s_L$ تقسیم می‌شود و هر زیرنمونه شامل واحدهای انتخاب شده در h خوشه است که به‌تصادف به یک پرسش‌گر منتسب می‌شوند).

با توجه به دو مرحله‌ی تصادفیدن بالا، واریانس (کل) این برآوردگر از فرمول واریانس شرطی به‌صورت زیر محاسبه می‌شود:

$$Var_{pc}(\bar{y}) = E_p(Var_c(\bar{y} \mid p)) + Var_p(E_c(\bar{y} \mid p)) \qquad (۴)$$

در رابطه‌ی (۴)، E(.|p) و Var(.|p) امید و واریانس به شرط مرحله‌ی ۱ تصادفیدن، یعنی انتخاب نمونه‌ی s و تخصیص آن به L پرسش‌گر را نشان می‌دهد. همچنین اندیسp و c در E و Var به‌ترتیب نشان‌دهنده‌ی امید و واریانس نسبت به مرحله‌ی ۱ و ۲ تصادفیدن است و اندیس pc امید و واریانس را نسبت به هر دو مرحله‌ی ۱ و ۲ نشان می‌دهد.

تحت مدل (۳) امید و واریانس $\bar{y}$ عبارتند از:

$$E_c(\bar{y}) = \overline{m}_s + E(D) + E(R) \qquad (۵)$$

www.SID.ir

Archive of SID

(۶)

$$Var_C(\bar{y}) = \frac{1}{n^2} Var_c\left(\sum_{a=1}^{n}\left(\frac{1}{m}\sum_{b=1}^{m} y_{ab}\right)\right) = \frac{1}{n^2}\left[Var_c\left(\sum_{a=1}^{n}\bar{y}_{am}\right)\right]$$

$$= \frac{1}{n^2}\left(\sum_{a=1}^{n}\left(Var_c(\bar{y}_{am})\right) + \sum_{a\neq a'}\sum Cov_c(\bar{y}_{am}, \bar{y}_{a'm})\right]$$

$$= \frac{1}{n^2}\left[n\left(s_D^2 + \frac{s_R^2}{m}\right) + n(h-1)\left(\frac{m-1}{m}\right)s_D^2\right]$$

$$= \frac{1}{n}\left[s_D^2\left(h\left(1-\frac{1}{m}\right)+\frac{1}{m}\right)+\frac{s_R^2}{m}\right]$$

زیرا

$$Var_C(\bar{y}_{am}) = Var_c\left(\sum_{b=1}^{m}\frac{1}{m} y_{ab}\right)]$$

$$= \frac{1}{m^2}\left(\sum_{b=1}^{m}\left(Var_c(y_{ab})\right) + \sum_{b\neq b'}\sum Cov_c(y_{ab}, y_{ab'})\right]$$

$$= \frac{1}{m^2}\left[m\left(s_D^2 + s_R^2\right) + m(m-1)s_D^2\right]$$

$$= s_D^2 + \frac{s_R^2}{m}$$

از رابطه‌ی (۵) نتیجه می‌شود:

$$Var_p(E_C(\bar{y})) = Var_p(\bar{m}_s) \qquad (۷)$$

رابطه‌ی (۷) نشان‌دهنده‌ی تغییرپذیری میانگین نمونه‌ای مقادیر واقعی است و واریانس نمونه‌گیری را بیان می‌کند. در نمونه‌گیری تصادفی خوشه‌ای دومرحله‌ای مورد نظر، رابطه‌ی (۷) به‌صورت زیر درمی‌آید:

$$Var_p(E_c(\bar{y})) = \left(1-\frac{n}{N}\right)\frac{S_{1m}^2}{n} + \left(1-\frac{m}{M}\right)\frac{S_{2m}^2}{m} \qquad (۸)$$

که در آن S_{1m}^2 واریانس نمونه‌ای بین خوشه‌ها و S_{2m}^2 واریانس نمونه‌ای درون خوشه‌ها را نشان می‌دهد.

برای بررسی واریانس کل برآوردگر میانگین ملاحظه می‌شود که جمله‌ی اول عبارت سمت راست رابطه‌ی (۴)، امید رابطه‌ی (۶) است که با خودش یکسان بوده و واریانس غیرنمونه‌گیری برآوردگر است. جمله‌ی دوم سمت راست رابطه‌ی (۴) همان رابطه‌ی (۸) است و واریانس نمونه‌گیری برآوردگر را در طرح نمونه‌گیری دومرحله‌ای نشان می‌دهد. بنابراین نتیجه می‌شود تحت شرایط بیان‌شده، واریانس کل برآوردگر میانگین برابر با مجموع واریانس نمونه‌گیری و واریانس غیرنمونه‌گیری این برآوردگر است(برای محاسبه‌ی واریانس نمونه‌گیری در آمارگیری‌های پیچیده لازم است که طرح نمونه‌گیری و صورت برآوردگر مورد توجه قرار گیرد. ولتر (۱۹۸۵) مرجع کاملی برای محاسبه‌ی این واریانس است. در عمل با استفاده از نرم‌افزار سودان می‌توان واریانس نمونه‌گیری را در حالت‌های پیچیده برآورد کرد).

۴. مدل پیشنهادی برای خطای پاسخ

خطای پاسخ خطایی است که در زمان گردآوری و ثبت داده‌ها رخ می‌دهد، لذا برای مدل‌سازی مناسب خطای پاسخ باید شرایط اجرای آمارگیری‌ها به‌طور کامل شناسایی شوند، به‌عبارت دیگر لازم است مدل خطای پاسخ با در نظر گرفتن وضعیت اجرای آمارگیری‌ها طراحی شود. در این مقاله با بررسی طرح‌های آماری عمده در ایران، منابع بروز خطای پاسخ با توجه به نحوه‌ی اجرای آمارگیری‌های حضوری تعیین شده و پس از ارائه‌ی یک طرح آزمایش، مدلی کلی برای خطای پاسخ در آمارگیری‌های حضوری پیشنهاد می‌شود.

معمولاً طرح‌های آماری مرکز آمار ایران از طریق معاونت آمار و اطلاعات سازمان مدیریت و برنامه‌ریزی استان‌ها اجرا می‌شوند و به‌همین دلیل تفاوت‌هایی از نظر

www.SID.ir

Archive of SID

نحوه‌ی اجرا بین استان‌ها وجود دارد و بنابراین اثر استان در مدل خطای پاسخ منظور می‌شود. در این طرح‌ها کارشناسان مسوول بر صحت کار پرسش‌گرها و بازبین‌ها نظارت دارند. پرسش‌گرها مصاحبه‌ها را انجام می‌دهند و به‌طرق مختلف از جمله درک نکردن مفهوم سوال، اشتباه در تکمیل کردن پرسشنامه، راهنمایی نادرست پاسخ‌گو و علامت‌گذاری اشتباه می‌توانند سبب بروز خطا شوند.

از آنجا که هر پرسش‌گر با چند پاسخ‌گو مصاحبه می‌کند، اثر پرسش‌گر برای پاسخ‌گویان مربوط به‌خود یکسان فرض می‌شود. پاسخ‌گویان به سوال‌های پرسش‌گر پاسخ می‌دهند و ممکن است به‌دلایل مختلف مانند درک نادرست از سوال، ندانستن پاسخ یا خستگی و بی‌میلی در ارائه‌ی پاسخ صحیح و یا به‌علت فراموشی موجب بروز خطا شوند.

کدگذار وظیفه‌ی کدگذاری پرسشنامه‌ها را به‌عهده دارد و به‌دلیل استفاده از کدهای غلط، ثبت اشتباه و عدم توجه به سازگاری پاسخ‌ها با یکدیگر سبب بروز خطا می‌شود. بازبین کار نظارت بر عملکرد میدانی پرسش‌گرها و چگونگی تکمیل پرسشنامه‌ها را انجام می‌دهد و می‌تواند به‌دلیل کوتاهی در انجام وظایف، عدم تشخیص یا تصحیح موارد نادرست و... منبع بروز خطا باشد. خطای کارشناس مسوول می‌تواند از طریق درک ناکافی از موضوع، آموزش ناکافی و عدم نظارت بر اجرای صحیح طرح باشد.

اثرهای مدل خطای پاسخ بر اساس چگونگی اجرای طرح‌های آماری مشخص شده و با پیشنهاد یک طرح آزمایش، مدل خطای پاسخ در حالت کلی طراحی شده است. در این مدل اثرهای استان، کارشناس مسوول، بازبین، پرسش‌گر و پاسخ‌گو منظور شده‌اند. از آنجا که معمولاً در طرح‌های آماری که توسط معاونت آمار و اطلاعات سازمان مدیریت و برنامه‌ریزی استان‌ها اجرا می‌شوند، کار بازبینی و کدگذاری پرسشنامه‌ها توسط یک گروه خاص انجام می‌شود، اثر کدگذار و بازبین با هم ادغام شده و این اثر در مدل منظور شده است.

پس از تعیین مولفه‌های مدل خطای پاسخ، سوال این است که چه نوع طرحی را برای ارتباط خطای پاسخ با این عوامل انتخاب کنیم؟

در آمارگیری‌ها می‌توان بین سطوح مختلف یک عامل که تحت تاثیر عامل دیگری قرار دارند تفاوت قائل شد. به‌عبارت دیگر بازبین‌های منتسب به یک کارشناس مسوول با کارشناس مسوول دیگر و یا پرسش‌گرهای منتسب به یک بازبین با بازبین دیگر را می‌توان متمایز در نظر گرفت، به‌قسمی که هر کارشناس مسوول تنها با بازبین‌های مختص به‌خود ارتباط داشته باشد، بازبین‌ها تنها بازبینی کار پرسش‌گران مربوط به‌خود را انجام دهند و پرسش‌گران نیز با پاسخ‌گویان منتسب به آن‌ها مصاحبه کنند. اجرای طرحی که ارتباط بین عملگرهای آمارگیری (یعنی کارشناس مسوول، بازبین و پرسش‌گر) به‌صورت بیان‌شده باشد، سبب می‌شود که طرح مذکور تبدیل به طرح آشیانی شود. در این صورت اثر بازبین در اثر کارشناس مسوول، اثر پرسش‌گر در اثر بازبین و اثر پاسخ‌گو در اثر پرسش‌گر و همه‌ی این اثرها در اثر استان آشیانی می‌شوند.

www.SID.ir

Archive of SID

ثابت یا تصادفی بودن اثر عملگرهای آمارگیری در مدل به نحوه‌ی انتخاب آن‌ها بستگی دارد. در صورتی که برای هر یک از این اثرها افراد به‌صورت مجموعه‌ای ثابت در آمارگیری‌ها مشارکت داشته باشند، آن اثر در مدل ثابت و در صورتی که برای هر یک از عملگرهای آمارگیری افراد به‌طور تصادفی از بین واجدین شرایط انتخاب شوند، آن اثر در مدل تصادفی خواهد بود.

معمولاً در طرح‌های آماری مرکز آمار ایران کارشناسان مسوول، بازبین‌ها (و کدگذارها)، در استخدام سازمان مدیریت و برنامه‌ریزی استان‌ها هستند و در اجرای دوره‌های مختلف آمارگیری تغییر نمی‌کنند، لذا اثرهایی ثابت در مدل دارند. به‌دلیل آن‌که پرسش‌گرها از بخش خصوصی هستند و در دوره‌های مختلف آمارگیری افراد متفاوتی به‌کار گرفته می‌شوند، اثر پرسش‌گر در مدل اثری تصادفی است. لازم به‌ذکر است که به‌دلیل انتخاب تصادفی پاسخ‌گویان از جامعه، اثر پاسخ‌گو همواره به‌عنوان اثری تصادفی در مدل خطای پاسخ مطرح است.

به‌طور کلی مدل خطای پاسخ با توجه به اثرهای عمده‌ی آن به‌صورت زیر پیشنهاد می‌شود:

$$Y_{ijkls} = \mu_{ijkls} + A_i + B_{ij} + C_{ijk} + D_{ijkl} + R_{ijkls} \qquad (۹)$$

که در آن μ_{ijkls} مقدار واقعی، Y_{ijkls} مقدار مشاهده شده‌ی پاسخ‌گوی sام در استان iام مربوط به کارشناس مسوول jام و بازبین (و کدگذار) kام که توسط پرسش‌گر lام به‌دست می‌آید، A_i اثر استان iام، B_{ij} اثر کارشناس مسوول jام در استان iام، C_{ijk} اثر بازبین (و کدگذار) kام که زیر نظر کارشناس مسوول jام در استان iام فعالیت می‌کند، D_{ijkl} اثر پرسش‌گر lام که زیر نظر بازبین kام و کارشناس مسوول jام 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 مسوول jام فعالیت می‌کند مصاحبه می‌شود.

در مدل (۹)، اندیس $i = 1,...,۳۰$ مربوط به استان‌ها، اندیس $j = 1,...,J_i$ مربوط به کارشناس مسوول، اندیس $k = 1,...,K_{ij}$ مربوط به بازبین (و کدگذار)، اندیس $l = 1,...,L_{ijk}$ مربوط به پرسش‌گر و اندیس $s = 1,...,S_{ijkl}$ مربوط به پاسخ‌گو است.

با فرض پیروی اثرهای تصادفی پرسش‌گر و پاسخ‌گو از توزیع نرمال، مدل (۹) یک مدل تحلیل واریانس آمیخته‌ی آشیانی است. این مدل در حالتی ارائه شد که در اغلب طرح‌های آماری عمده با آن مواجه هستیم و در آن پرسش‌گرها و پاسخ‌گویان اثرهای تصادفی در مدل دارند. تصادفی بودن اثر پاسخ‌گو اجتناب‌ناپذیر است اما پرسش‌گران در صورتی که در استخدام سازمان آماری ذیربط باشند اثری ثابت در مدل خواهند داشت. در این صورت تعداد اثرهای تصادفی به حداقل خود خواهد رسید و واریانس غیرنمونه‌گیری (اندازه‌گیری) تنها تابعی از واریانس پاسخ‌گویان می‌شود. کاهش واریانس غیرنمونه‌گیری باعث افزایش دقت برآوردهای حاصل از آمارگیری خواهد شد، لذا برای افزایش کیفیت نتایج آمارگیری‌ها پیشنهاد می‌شود که سازمان‌های آماری عملگرهای آمارگیری و از جمله پرسش‌گرها را به

www.SID.ir

Archive of SID

استخدام خود درآورند و با آموزش مناسب آن‌ها موجبات افزایش دقت و درستی داده‌ها را فراهم نمایند.

۵. برآوردگری برای برآورد دقت در آمارگیری‌ها

در این بخش برآوردگری برای برآورد دقت برآوردگر میانگین، در طرح‌های دومرحله‌ای تحت شرایط مورد نظر ارائه می‌شود. بدین منظور، برآوردگر واریانس گروه‌های تصادفی را به‌کار برده و فرض می‌شود هر گروه از پاسخ‌گویان که به‌طور تصادفی به پرسش‌گرها منتسب می‌شوند حکم گروه‌های تصادفی را دارند. برآوردگر واریانس گروه‌های تصادفی به‌صورت زیر تعریف می‌شود:

$$\hat{V}_B = \frac{1}{L(L-1)}\sum_{l=1}^{L}(\bar{y}_l - \bar{y})^2$$

که در آن $\bar{y}_l$ میانگین مقادیر مشاهده‌شده‌ی مربوط به پرسش‌گر lام، L تعداد پرسش‌گرها و $\bar{y}$ میانگین مقادیر مشاهده‌شده در نمونه‌ی مورد نظر است. (روش گروه‌های تصادفی برای برآورد واریانس توسط ولتر (۱۹۸۵)، معرفی شده است.)

طبق فرمول امید شرطی داریم:

$$E_{pc}(\hat{V}_B) = E_p(E_c(\hat{V}_B)) \qquad (۱۰)$$

مدل خطای پاسخ (c) را در رابطه‌ی (۱۰)، مدل (۹) در نظر می‌گیریم.

$$E(\hat{V}_B) = E_pE_c(\hat{V}_B) = E_pE_c\Big(\frac{1}{L(L-1)}\sum_{l=1}^{L}(\bar{y}_l - \bar{y})^2\Big)$$

$$E_c\Big(\sum_{l=1}^{L}(\bar{y}_l - \bar{y})^2\Big) = E_c\Big(\sum_{l=1}^{L}\bar{y}_l^2 - L\bar{y}^2\Big) =$$

$$(L-1)\Big(s_D^2\Big(1-\frac{1}{m}\Big)+\frac{s_D^2}{hm}+\frac{s_R^2}{hm}+\sum_{l=1}^{L}\bar{m}_l^2 - L\bar{m}^2\Big)$$

زیرا

$$E_c(\bar{y}^2) = Var_c(\bar{y}) + (E_c(\bar{y}))^2 = \frac{s_D^2}{L}\Big(1-\frac{1}{m}\Big) + \frac{s_D^2}{mn} + \frac{s_R^2}{mn} + (\bar{m}+a)^2$$

$$E_c(\bar{y}_l^2) = Var_c(\bar{y}_l) + (E_c(\bar{y}_l))^2 = s_D^2\Big(1-\frac{1}{m}\Big) + \frac{s_D^2}{hm} + \frac{s_R^2}{hm} + (\bar{m}+a)^2$$

که در آن a مربوط به جمع اثرهای ثابت مدل (۹) است (یعنی $a = A+B+C$ که در آن A، B و C به‌ترتیب مربوط به اثرهای ثابت استان، کارشناس مسوول و بازبین هستند). بنابراین

$$E_pE_c\Big(\sum_{l=1}^{L}(\bar{y}_l - \bar{y})^2\Big) = (L-1)\Big(s_D^2\Big(1-\frac{1}{m}\Big) + \frac{s_D^2}{hm} + \frac{s_R^2}{hm}\Big) + L\Big(\frac{N-h}{N}\frac{S_{1m}^2}{h} + \frac{M-m}{M}\frac{S_{2m}^2}{mh}\Big) - \frac{N-n}{N}\frac{S_{1m}^2}{n} + \frac{M-m}{M}\frac{S_{2m}^2}{mn}$$

زیرا

$$E_p(\bar{m}_l^2) = Var_p(\bar{m}_l) + (E_p(\bar{m}_l))^2 = Var_p(\bar{m}_l) + (\bar{m}+a)^2$$

$$E_p(\bar{m}^2) = Var_p(\bar{m}) + (E_p(\bar{m}))^2 = Var_p(\bar{m}) + (\bar{m}+a)^2$$

در نتیجه

(۱۱)

$$E_{pc}(\hat{V}_B) = \frac{s_D^2}{L}\Big(1-\frac{1}{m}\Big) + \frac{s_D^2}{mn} + \frac{s_R^2}{mn} + \frac{S_{1m}^2}{n} + \frac{(M-m)S_{2m}^2}{nmM}$$

$$= \frac{s_D^2}{L}\Big(1-\frac{1}{m}\Big) + \frac{s_D^2}{mn} + \frac{s_R^2}{mn} + \frac{N-n}{N}\frac{S_{1m}^2}{n} + \frac{(M-m)S_{2m}^2}{Mnm} + \frac{1}{N}S_{1m}^2$$

$$= Var_{pc}(\bar{y}) + \frac{1}{N}S_{1m}^2$$

www.SID.ir

Archive of SID

که در آن h تعداد بلوک‌های تخصیص یافته به هر پرسش‌گر، m تعداد خانوارهای نمونه در هر بلوک و n تعداد بلوک‌های نمونه از بین N بلوک را نشان می‌دهد.

رابطه (۱۱) نشان می‌دهد که $\hat{V}_B$ یک برآوردگر اریب برای واریانس کل برآوردگر میانگین $(Var_{pc}(\bar{y}))$ است و $\hat{V}_B - \frac{1}{N}S_{1m}^2$ برآوردگری نااریب برای واریانس کل برآوردگر میانگین است. لذا برآوردگر $\hat{V}_B - \frac{1}{N}\hat{S}_{1m}^2 = \hat{V}_B - \frac{1}{N}s_1^2$ برای برآورد واریانس کل آماره‌ی $(\bar{y})$ در آمارگیری‌ها حاصل می‌شود.

٦. کاربرد

یکی از اهداف طراحی مدل خطای پاسخ، برآورد خطای کل و دقت برآوردگرهای آمارگیری است. طرح هزینه و درآمد خانوار از جمله طرح‌های مهم آماری است که هر ساله در کشور اجرا می‌شود.

برای ارائه‌ی کاربردی از برآوردگر و مدل پیشنهادی، از داده‌های طرح هزینه و درآمد خانوار سال ۱۳۸۲ مربوط به شهر تهران استفاده شده و دقت آماره‌های مورد نظر در این آمارگیری برآورد می‌شوند. طرح نمونه‌گیری در آمارگیری از هزینه و درآمد خانوار، خوشه‌ای دومرحله‌ای است. در مرحله‌ی اول، بلوک‌ها و در مرحله‌ی دوم تعدادی از خانوارهای داخل هر بلوک به‌تصادف انتخاب می‌شوند. در این طرح از هر بلوک نمونه، خانوارها به تعداد برابر (۵ خانوار) به‌طور تصادفی تعیین می‌شوند.

در سال ۸۲ تعداد خانوارهای نمونه‌ی این طرح در مناطق شهری کل کشور ۱۰۹۵۹ خانوار و در شهر تهران ۱۱۸۰ خانوار بوده است. فایل داده‌های طرح هزینه و درآمد خانوار شامل کد عملگرهای آمارگیری از جمله کد پرسش‌گر، کد بازبین و کد کارشناس مسوول نیست. برای دستیابی به این اطلاعات لازم بود تا با مراجعه به پرسشنامه‌ها کدهای مربوط استخراج و وارد فایل داده‌ها شوند. به دلیل گستردگی کار و هزینه‌ی زیاد آن امکان اصلاح فایل داده‌ها برای کل کشور وجود نداشت لذا این کار برای شهر تهران انجام شد.برای مجموعه داده‌های مورد نظر، اثر استان، کارشناس مسوول و بازبین قابل آزمون نیستند (زیرا برای این داده‌ها $I = J = K = 1$). لذا در ادامه آماره های آزمون برای سایر اثرهای مدل (۹) ارائه می‌شود. به طور کلی در مدل (۹) می‌توان فرض کرد که $D_{ijkl} \sim N(0, s_D^2)$ و $R_{ijkls} \sim N(0, s_R^2)$ و برای مقادیر واقعی $m_{ijkls} = m + M_{ijkls}$ که $M_{ijkls} \sim N(0, s_M^2)$ باشد. با فرض متعادل بودن داده‌ها در تمامی سطوح (یعنی $J_i = J$ ، $K_{ij} = K$ ، $L_{ijk} = L$ و $S_{ijkl} = S$) که منظور از $L_{ijk} = L$ و $S_{ijkl} = S$، به ترتیب اختصاص تعداد پرسشگران برابر به بازبین ها و تخصیص تعداد برابر پاسخ‌گویان به پرسشگرهاست، می‌توان ثابت کرد:

میانگین توان‌های دوم اثر D برابر

$$MSD = \frac{SSD}{IJK(L-1)} = \frac{S\sum_i\sum_j\sum_k\sum_l(\bar{y}_{ijkl.} - \bar{y}_{ijk..})^2}{IJK(L-1)}$$

حاصل شده است که امید ریاضی آن به‌صورت زیر به دست می‌آید:

$$E(MSD) = s_M^2 + Ss_D^2 + s_R^2$$

www.SID.ir

Archive of SID

$$MSR = \frac{SSR}{IJKL(S-1)} = \frac{\sum_i \sum_j \sum_k \sum_l \sum_s (y_{ijkls} - \bar{y}_{ijkl.})^2}{IJKL(S-1)}$$

میانگین توان‌های دوم اثر R است و امید ریاضی آن به‌صورت زیر محاسبه می‌شود:

$$E(MSR) = s_M^2 + s_R^2$$

آزمون اثر پرسشگر (اثر تصادفی D در مدل (۹)) توسط فرض صفر $H_0 : s_D^2 = 0$ و آماره‌ی آزمون $F_0 = \frac{MSD}{MSR}$ صورت می‌گیرد. این آماره تحت H_0 دارای توزیع F با (IJK(L-1),IJKL(S-1)) درجه آزادی است. در صورتی می‌توان اثر پاسخ‌گو را آزمون کرد که تکرار اندازه‌گیری‌ها قابل انجام باشد، زیرا در غیر این صورت جمله‌ی میانگین توان‌های دوم خطا را نمی‌توان محاسبه کرد.

برای اثبات روابط و نحوه‌ی به دست آوردن آماره‌های آزمون در حالت کلی مدل مورد نظر می‌توان به علی‌محمدی (۱۳۸۵) مراجعه کرد. برای تمامی مجموعه داده‌های مورد نظر، آزمون فرض اثرهای قابل بررسی نشان می‌دهد که اثرها در همه موارد قویاً معنی‌دار هستند.

اکنون به برآورد واریانس کل (دقت) برآوردگرهای مورد نظر از داده‌های طرح هزینه و درآمد خانوار شهر تهران در سال ۱۳۸۲ می‌پردازیم.

بنا بر مطالب ارائه شده در بخش ۵ و با توجه به طرح نمونه‌گیری (خوشه‌ای دومرحله‌ای) در آمارگیری از هزینه و درآمد خانوار، واریانس کل از رابطه‌ی (۱۱) قابل محاسبه است. (با توجه به داده‌ها در رابطه‌ی (۱۱) داریم: $m = ۵$، $L = ۴$، $\bar{y}_l$ میانگین مقادیر صفت مورد نظر واحدهای مربوط به پرسشگر lام، $\bar{y}$ میانگین کل مقادیر مشاهده‌شده‌ی صفت و s_1^2 واریانس بین بلوک‌های نمونه‌ی انتخاب شده است که با استفاده از مقادیر مشاهده شده در این بلوک‌ها محاسبه می‌شود. البته در صورتی که در مدل (۹)، $m_{ijkls} = m + M_{ijkls}$ قرار گیرد، به‌جای s_R^2 در رابطه‌ی (۱۱)، $s_R^2 + s_M^2$ قرار می‌گیرد.)

از آنجا که واریانس (و خطای استاندارد) به واحد اندازه‌گیری بستگی دارد و همچنین برای آمارگیری‌های مختلف و صفات گوناگون قابلیت مقایسه ندارد، برای گزارش دقت برآوردگر معمولاً از درصد خطای نسبی استفاده می‌شود. خطای نسبی از تقسیم نصف طول بازه‌ی اطمینان، بر میانگین صفت مورد نظر حاصل می‌شود. درصد خطای نسبی برای هر یک از صفات، دو بار محاسبه شده است. یک بار با در نظر گرفتن واریانس نمونه‌گیری و بار دوم با توجه به واریانس کل برآوردگر.

با استفاده از داده‌های فایل شهر تهران طرح هزینه و درآمد سال ۸۲، میانگین درآمد کل یک خانوار در شهر تهران ۵۸۱۳۵۳۷۶ ریال در سال ۸۲ برآورد می‌شود. برآورد خطای استاندارد نمونه‌گیری این آماره ۱/۲۷۵×۱۰^۶ و درصد خطای نسبی آن با در نظر گرفتن خطای استاندارد نمونه‌گیری ٤/۲۹ به‌دست می‌آید. خطای استاندارد کل این آماره ۴/۰۸۰×۱۰^۶ برآورد می‌شود و درصد خطای نسبی با در نظر گرفتن خطای استاندارد کل، ۱۳/۷۵۶ می‌شود.

میانگین هزینه‌ی خالص هر خانوار در سال ۸۲ در شهر تهران برابر۶۲۷۲۰۵۸۰ ریال و خطای استاندارد (نمونه‌گیری) آن ۱/۴۲۸×۱۰^۶ برآورد می‌شوند. درصد خطای نسبی این آماره ٤/٤٦ می‌شود. مقدار خطای

www.SID.ir

Archive of SID

استاندارد کل آماره $۷/۰۶۸\times۱۰^{۶}$ شده است و درصد خطای نسبی با استفاده از خطای استاندارد کل برابر ۲۲/۰۹ می‌شود.

میانگین هزینه‌ی ناخالص هر خانوار شهر تهران در سال ۸۲ برابر ٦٤٧٣١٩٠١ ریال با خطای استاندارد نمونه‌گیری $۱/۵۱۰\times۱۰^{۶}$ و درصد خطای نسبی ٤/٥٧ برآورد می‌شود. برآورد خطای استاندارد کل این آماره $۸/۶۹۴\times۱۰^{۶}$ و برآورد درصد خطای نسبی آن با استفاده از خطای استاندارد کل برابر ۲۶/۳۲ می‌شود.

چنان‌که ملاحظه می‌شود درصد خطای نسبی برآوردگر میانگین درآمد کل کمتر از هزینه‌ی خالص و هزینه‌ی ناخالص خانوار است و نتیجه می‌شود که درآمد کل خانوارهای شهر تهران در سال ۱۳۸۲ با دقت بیشتری نسبت به هزینه‌های خالص و ناخالص خانوار اندازه‌گیری شده است.

۷. نتیجه‌گیری

برازش مدلی مناسب به خطای پاسخ دارای اهمیت و کاربردهای فراوانی است که از جمله‌ی این کاربردها در برآورد میزان دقت نتایج حاصل از آمارگیری‌ها است. برای مدل‌سازی خطای پاسخ لازم است که نحوه‌ی اجرای آمارگیری‌ها به‌طور کامل مورد مطالعه و بررسی قرار گیرد و مدل بر مبنای چگونگی اجرا طراحی شود. در این مقاله با بررسی فرایند اجرای طرح‌های آماری عمده در ایران، عوامل موثر در بروز خطای پاسخ تعیین شده و بر مبنای پیشنهاد یک طرح آزمایش، مدلی کلی برای خطای پاسخ در آمارگیری‌های حضوری طراحی شده است. همچنین تثبیت اثر عملگرهای آمارگیری از سوی سازمان‌های آماری، به‌عنوان راهکاری برای افزایش دقت نتایج حاصل از آمارگیری‌ها پیشنهاد شده است.

به‌علاوه برآوردگری برای برآورد دقت نتایج آمارگیری‌ها بر مبنای مدل خطای پاسخ پیشنهادی در طرح‌های دو مرحله‌ای ارائه شده است. به‌عنوان یک کاربرد، از داده‌های طرح هزینه و درآمد خانوار سال ۱۳۸۲ مربوط به شهر تهران استفاده شده و دقت برآوردگرهای مورد نظر در این آمارگیری بر اساس مدل و برآوردگر پیشنهادی محاسبه شدند. نتیجه‌ی محاسبات نشان می‌دهند که درآمد کل خانوارهای شهر تهران در سال ۱۳۸۲ با دقت بیشتری نسبت به هزینه‌های خالص و ناخالص آن‌ها اندازه‌گیری شده است.

منابع

1. H. O., Ayhan, Models of Response error Components in Supervised Interview-Reinterview surveys, Journal of applied statistics, Vol. 3, No 9, 1047-1054; (2003).
2. F., Bassie, & L., Fabbris, Estimation of nonsampling error in interview reinterview supervised in intenpenetrated assignments, Survey measurement and process quality, Wiley, New York; (1997).
3. P., Biemer, and S. L., Stokes, Approaches of modeling of measurement error, in: Measurement error in surveys, wiley, New York; (1991).
4. P., Biemer, and D., Trewin, A Review of measurement error effects on the analysis of survey data, in: Survey measurement and process quality, edited by L. Lyberg, P. Biemer, M. Collins, E. Deleeuw, C. Dippo, N. Schwartz and D. Trewin, Wiley, New York; (1997).
5. I. P., Fellegi, An Improved method of estimating the correlated response variance, Journal of American statistical assoc., 69, 496-

www.SID.ir

Archive of SID

501; (1974).
6. R. M., Groves, Survey Errors and Survey Costs, Wiley, New York; (2004).
7. R. M., Groves, F. J., Fowler, M. P., Couper, J. M., Lepkowski, E., Singer, and R., Tourangeau, Survey Mthodology, Wiley, New York; (2004).
8. M. H., Hansen, W. N., Hurvitz, and M.A., Bershad, Measurement error in censuses and surveys, Bull. Intern. Statist. Inst., 38(2), 359-374; (1961).
9. H. O., Hartley, and J. N. K., Rao, Estimation of nonsampling variance components in sample surveys, in: Current topics in surveys sampling, Eds: D. Krowski, R. Platek and J. N. K. Rao, New York, Academic Press; (1978).
10. H. O., Hartly, Estimation and Design for Nonsampling Errors in Surveys, in: Current Topics in Survey Sampling, 31 – 46, Academic Press, New York; (1981).
11. L., Kish, Studies of interviewer variance for attitudinal variables, Journal of American statistical assoc., 57, 92-115; (1962).
12. J. T., Lessler, and W. D., Kalsbeek, Nonsampling error in surveys, Wiley, New York; (1992).
13. P., Lundquist, and J. H., Wretman, Estimating interviewer variance under a measurement error model for continuous data, Research of Statistics Sweden; (2002).
14. P. C., Mahalanobis, Recent experiments in statistical sampling in the Indian statistical institute, Royal statistical society, 109, 325-378; (1946).
15. SUDAAN, Release 8, Research Triangle Institute (RTI), USA.
16. k., Wolter, Introduction to Variance Estimation, New York, Springer-Verlag; (1985).
۱۷. علی‌محمدی، روشنک (۱۳۸۵). مدل‌بندی خطای پاسخ در آمارگیری‌ها، رساله دکتری، دانشگاه تربیت مدرس.

www.SID.ir