

بررسی تأثیر دو روش جایگزینی و تعدیل وزنی بر برآوردهای طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی

مرجان نورینی

مرکز آمار ایران

چکیده: در این مقاله، ضمن معرفی دو روش جایگزینی و تعدیل وزنی به عنوان روش‌های تعدیل بی‌پاسخی و محدود به مقایسه تأثیر آن‌ها بر برآورد واریانس برآوردگر میانگین در طرح هزینه و درآمد خانوارهای شهری و روستایی پرداخته می‌شود.

واژه‌های کلیدی: بی‌پاسخی، وزن دهی، جایگزینی

۱ مقدمه

عدم توانایی در به دست آوردن اطلاعات مورد نیاز در آمارگیری‌ها «بی‌پاسخی» نام دارد که می‌تواند یکی از منابع مهم خطای آمارگیری محسوب شود و ناشی از عدم مشاهده و بازتاب تلاشی ناموفق در به دست آوردن اطلاعات از یک واحد واجد شرایط آمارگیری در جامعه است. مقوله بی‌پاسخی را می‌توان به بی‌پاسخی واحد و بی‌پاسخی قلم^۱ تقسیم کرد. زمانی که یکی از واحدهای آماری به هر دلیلی در عملیات آمارگیری مشارکت نمی‌کند با بی‌پاسخی واحد مواجه می‌شویم. هنگامی که واحد آماری در عملیات آمارگیری مشارکت می‌کند اما در مرحله‌ی تحلیل اطلاعات، داده‌های بعضی از اقلام اطلاعاتی مربوط به این واحد آماری، بنا به دلایلی مانند امتناع پاسخگو از پاسخ دادن به برخی پرسش‌ها، قادر نبودن وی در پاسخ دادن به بعضی پرسش‌ها، موفق نبودن پرسشگر در پرسیدن پرسش، ثبت نادرست پاسخ یا ناسازگار بودن پاسخ در مرحله‌ی ویرایش داده‌ها، در دسترس نیست که در این حالت، بی‌پاسخی مربوط به اقلام آماری بوده و آن را بی‌پاسخی قلم یا بی‌پاسخی سؤال می‌نامند. به دلیل دشواری اندازه‌گیری مقدار خطای بی‌پاسخی، بیش‌تر تلاش‌ها به حداقل کردن رویداد این خطا هدایت شده است. بعد از این مقدمه و در بخش ۲ راهکارهای کاهش بی‌پاسخی شرح داده شده است. بخش ۳، ضمن معرفی فرآیند جایگزینی و وزن‌دهی بکار رفته در طرح آمارگیری از هزینه و درآمد خانوار، به مقایسه تأثیر آن‌ها روی برآورد واریانس برآوردگر

^۱ Item Nonresponse

میانگین با استفاده از نتایج این طرح طی سال‌های ۱۳۸۷ و ۱۳۸۸ می‌پردازد و نهایتاً در بخش آخر نتیجه‌گیری از مطالب گفته شده بیان می‌شود.

۲ راهکارهای کاهش بی‌پاسخی

بطور کلی دو راهبرد برای مقابله با بی‌پاسخی وجود دارد: کاهش و تعدیل. کاهش بی‌پاسخی به معنی به کارگیری راهکارهایی قبل و یا در حین گردآوری داده‌ها برای افزایش احتمال مشارکت واحدهای نمونه در آمارگیری است به گونه‌ای که از بروز بی‌پاسخی واحد و بی‌پاسخی قلم تا جایی که ممکن است کاسته شود. تعدیل بی‌پاسخی به معنی استفاده از روش‌هایی در مرحله‌ی تحلیل، برای مینیمم کردن اثرات بی‌پاسخی بر برآوردهای آماری است. با وجود تلاش‌های زیاد در زمینه‌ی کاهش بی‌پاسخی، رخداد آن امری غیر قابل اجتناب است. بنا بر این طراحان نمونه‌گیری اغلب تعدیل‌هایی را برای جبران بی‌پاسخی ارائه می‌دهند. سه روش زیر از روش‌های تعدیل بی‌پاسخی واحد است.

- ۱) تعدیل حجم نمونه با در نظر گرفتن حجم نمونه بالاتر از آن چه مورد نیاز است، صورت گیرد تا از این طریق بتوان بی‌پاسخی‌های مورد انتظار را نیز به حساب آورد.
- ۲) جایگزینی؛ فرایند جایگزینی خانوارهای بی‌پاسخ با خانوارهای دیگری که در نمونه نیستند که یا به صورت تصادفی انتخاب می‌شوند و یا به صورتی انتخاب می‌شوند که با پاسخگوی اولیه در یک یا چند مشخصه شباهت داشته باشند. [۲]
- ۳) روش‌های وزن‌دهی؛ به این ترتیب که در آن پاسخگویان را به گروه‌های جامع متقابلاً ناسازگار تقسیم بندی نموده و آن‌ها را رده‌های وزن‌دهی نام‌گذاری می‌کنند. سپس به هر رده یا به هر عضو هر رده، وزنی اختصاص داده می‌شود. [۴]

۳ بررسی و مقایسه دو روش جایگزینی و تعدیل وزنی

در این بخش ابتدا وضعیت جاری مرکز آمار ایران در استفاده از روش جایگزینی و تعدیل وزنی بی‌پاسخی مورد بررسی قرار می‌گیرد. سپس برآوردگر میانگین و خطای استاندارد آن با استفاده از نتایج طرح هزینه و درآمد خانوارهای شهری و روستایی سال‌های ۱۳۸۷ و ۱۳۸۸ تحت دو روش محاسبه و با هم مقایسه می‌شوند.

۱.۳ فرآیند جایگزینی در طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی

طرح آمارگیری از هزینه و درآمد خانوار با هدف برآورد میانگین هزینه خوراک، هزینه غیر خوراک، هزینه کل و درآمد یک خانوار در سطح مناطق شهری و روستایی هر یک از استان‌ها و کل کشور هر ساله توسط مرکز آمار ایران اجرا می‌شود. جامعه‌ی هدف در این طرح شامل همه‌ی خانوارهای معمولی و گروهی ساکن در مناطق شهری/روستایی کشور است. روش نمونه‌گیری این طرح، خوشه‌ای سه مرحله‌ای با طبقه‌بندی است. در طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی در موارد ذیل استفاده از خانوار جایگزین مجاز تشخیص داده شده است.

- (۱) خانوار در طول فصل آمارگیری به دلایلی از قبیل زیارت، مسافرت و مانند آن موقتاً در محل سکونت خود نبوده و خانوار غایب تلقی شود.
- (۲) مکان خالی از سکنه باشد.
- (۳) مکان از بین رفته، در دست ساخت یا تخریب باشد.
- (۴) مکان اقامتگاه معمولی خانوار نباشد. به عنوان مثال کارگاه یا اقامتگاه دوم خانوار باشد.
- (۵) مکان محل اقامت خانوار مؤسسه‌ای یا غیر ساکن باشد (خانوارهای مؤسسه‌ای یا غیر ساکن از پوشش طرح خارج هستند).
- (۶) آدرس مورد نظر پیدا نشده باشد.
- (۷) سایر(دلایلی غیر از دلایل بالا، به عنوان مثال دسترسی به خانوارهای نمونه‌ی اصلی به دلایلی مانند خراب شدن قسمتی از واحد نمونه‌گیری اولیه یا ترک دائمی همه‌ی خانوارها یا بخشی از آن‌ها امکان پذیر نباشد).

در صورت استفاده از خانوار جایگزین، اطلاعات این خانوار به عنوان اطلاعات خانوار نمونه در مجموعه‌ی داده‌ها ثبت می‌شود. در مواردی غیر از موارد هفت گانه بالا به عنوان مثال در مواردی مانند «عدم همکاری خانوار»، «ناتوانی خانوار در پاسخگویی به سؤالات به دلیل کهولت سن، ناتوانی ذهنی یا بی‌سوادی تمام اعضای خانوار» و یا «نبودن امکان دسترسی به مکان به دلایلی مانند شرایط خاص آب و هوایی، صعب‌العبور بودن مسیر یا مسائل امنیتی تا پایان فصل آمارگیری» به دلیل این که ممکن است الگوی هزینه و درآمد این نوع خانوارها متفاوت از سایر خانوارها باشد، پرسش نامه تکمیل نشده باقی می‌ماند و از خانوار جایگزین برای آن‌ها استفاده نمی‌شود.

در طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی سال ۱۳۸۷ انتخاب خانوارهای جایگزین همراه با خانوارهای نمونه‌ی اصلی در مرحله‌ی انتخاب نمونه،

صورت گرفت. به همین منظور در مرحله‌ی انتخاب خانوارهای نمونه، ۷ خانوار به روش نمونه‌گیری سیستماتیک دایره‌ای انتخاب شدند و سپس با استفاده از روش نمونه‌گیری تصادفی ساده (بدون جایگذاری) دو خانوار از بین هفت خانوار نمونه به عنوان خانوار جایگزین انتخاب شدند. در سال ۱۳۸۸، ۱۰ خانوار نمونه اصلی به روش سیستماتیک دایره‌ای انتخاب شد و از بین آن‌ها ۵ خانوار به روش تصادفی ساده (بدون جایگذاری) انتخاب و به عنوان خانوار جایگزین در نظر گرفته شد [۱].

۲.۳ فرایند وزن‌دهی در طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی

فرایند وزن‌دهی طرح هزینه و درآمد خانوار در سال ۱۳۸۷ شامل دو مرحله است. ابتدا وزن پایه اعمال می‌شود که عبارت از عکس احتمال قرار گرفتن واحدهای آماری در نمونه‌ای با اندازه‌ی تعیین شده است که بر اساس روش نمونه‌گیری محاسبه می‌شود. در این طرح، برای تمام خانوارهای واحد نمونه‌گیری اولیه، وزن پایه یکسان است. دومین مرحله‌ی وزن‌دهی برای جبران بی‌پاسخی انجام می‌شود. بدین ترتیب که وزن پایه‌ی محاسبه‌شده بر اساس عکس مقدار نرخ پاسخ تعدیل می‌شود. از سال ۱۳۸۸ به بعد تعدیل دیگری برای عدم پوشش نیز به فرایند وزن دهی اضافه شده است به این ترتیب که با توجه به اختلاف تعداد خانوار چارچوب و جامعه‌ی هدف، برای جبران کم پوششی چارچوب، وزن به دست آمده در مرحله‌ی دوم بر اساس پیش بینی تعداد خانوارها تعدیل می‌شود [۱].

۳.۳ محاسبه خطای استاندارد برآوردگر میانگین در طرح هزینه و درآمد خانوار تحت دو روش جایگزینی و تعدیل وزن و مقایسه‌ی آن‌ها

در این بخش با استفاده از نتایج طرح آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی در سال‌های ۱۳۸۷ و ۱۳۸۸، برآورد میانگین و خطای معیار آن برای چهار صفت اصلی این طرح شامل هزینه‌ی کل، هزینه خوراک، هزینه غیر خوراک و درآمد تحت دو روش جایگزینی و تعدیل وزنی با استفاده از نرم افزار SAS محاسبه شده است. به این صورت که ابتدا با استفاده از فایل اصلی یعنی فایلی که خانوارهای بی‌پاسخ در آن جایگزین شده‌اند برآوردگر میانگین و خطای معیار آن برای صفات مذکور محاسبه شد. سپس با جایگزین‌ها همانند داده بی‌پاسخ رفتار شده یعنی از مجموعه داده حذف شدند و سپس برآوردها با استفاده از تعدیل وزنی واحدهای تکمیل شده برآورد شد. واریانس برآوردگر میانگین در هر دو روش با استفاده فرمول زیر [۳] محاسبه شده

است. برای محاسبه خطای معیار از جذر واریانس استفاده شده است.

$$Var(\hat{Y}) = \sum_h \frac{n_h(1-f_h)}{n_h-1} \sum_i (e_{hi} - \bar{e}_{h..})^2$$

$$e_{hi} = (\sum_j w_{hij}(y_{hij} - \hat{Y}))/w_{h..} \quad \bar{e}_{h..} = (\sum_i e_{hi})/n_h$$

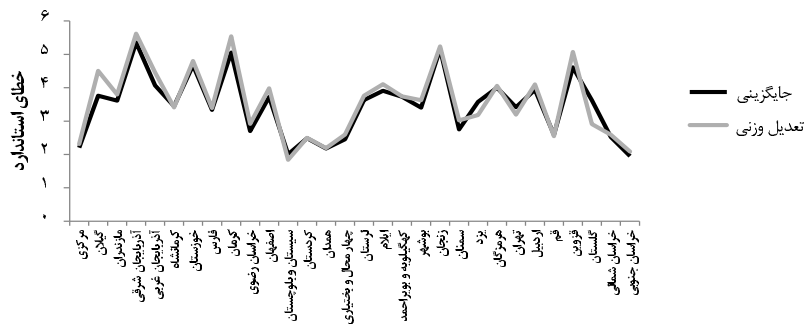
در روابط بالا i اندیس psu ، j اندیس عناصر، h اندیس طبقه و w_{hij} وزن هر واحد در i امین psu در h امین طبقه است. نتایج برای چهار صفت در جداول ۱ و ۲ و برای هزینه‌ی کل در نمودارهای ۴-۱ ارائه شده است.

جدول ۱: برآورد میانگین و خطای استاندارد میانگین صفات مورد نظر تحت دو روش بر اساس نتایج شهری سال ۱۳۸۷ در سطح کل کشور (مقادیر بر حسب میلیون ریال)

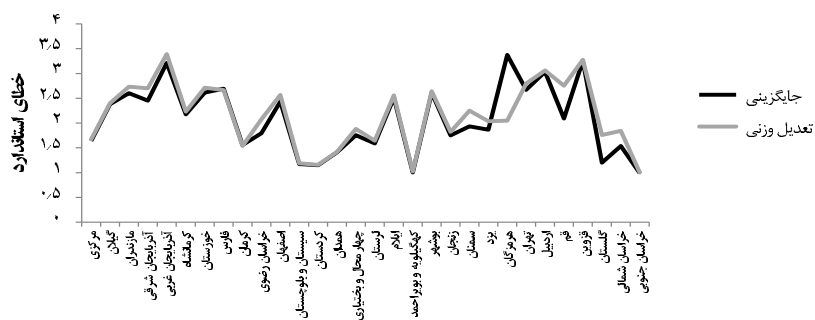
| صفت مورد نظر | روش جایگزینی | | روش تعدیل وزنی | |
|------------------|------------------------|------------------------|------------------------|------------------------|
| | میانگین خطای استاندارد | میانگین خطای استاندارد | میانگین خطای استاندارد | میانگین خطای استاندارد |
| هزینه‌ی کل | ۹۷/۷۲ | ۱/۱۶ | ۹۸/۳۹ | ۱/۱۶ |
| درآمد | ۸۹/۰۹ | ۱/۲۲ | ۸۹/۴۱ | ۱/۲۱ |
| هزینه‌ی خوراک | ۲۰/۲۸ | ۰/۱۸ | ۲۰/۲۵ | ۰/۱۷ |
| هزینه‌ی غیرخوراک | ۷۷/۵۲ | ۱/۰۸ | ۷۸/۲۲ | ۱/۰۹ |

جدول ۲: برآورد میانگین و خطای استاندارد میانگین صفات مورد نظر تحت دو روش بر اساس نتایج روستایی سال ۱۳۸۸ در سطح کل کشور (مقادیر بر حسب میلیون ریال)

| صفت مورد نظر | روش جایگزینی | | روش تعدیل وزنی | |
|------------------|------------------------|------------------------|------------------------|------------------------|
| | میانگین خطای استاندارد | میانگین خطای استاندارد | میانگین خطای استاندارد | میانگین خطای استاندارد |
| هزینه‌ی کل | ۶۰/۵۲ | ۰/۶۲ | ۶۰/۸۵ | ۰/۶۶ |
| درآمد | ۵۲/۴۴ | ۰/۵۳ | ۵۲/۶۹ | ۰/۵۵ |
| هزینه‌ی خوراک | ۲۱/۴۳ | ۰/۲۱ | ۲۱/۵۵ | ۰/۲۲ |
| هزینه‌ی غیرخوراک | ۳۹/۱۴ | ۰/۴۸ | ۳۹/۳۴ | ۰/۵۱ |



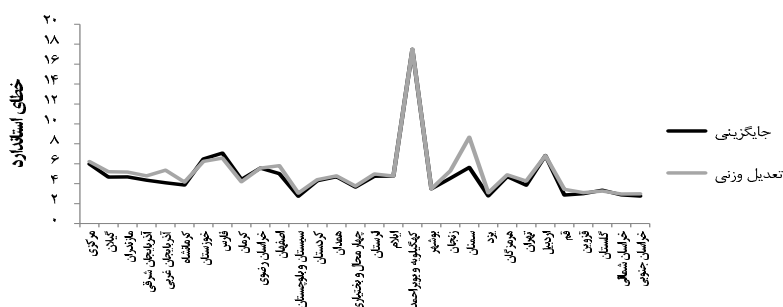
شکل ۱: برآورد خطای استاندارد میانگین هزینه‌ی کل تحت دو روش جایگزینی و تعدیل وزنی در نقاط شهری سال ۱۳۸۷ (مقادیر بر حسب میلیون ریال)



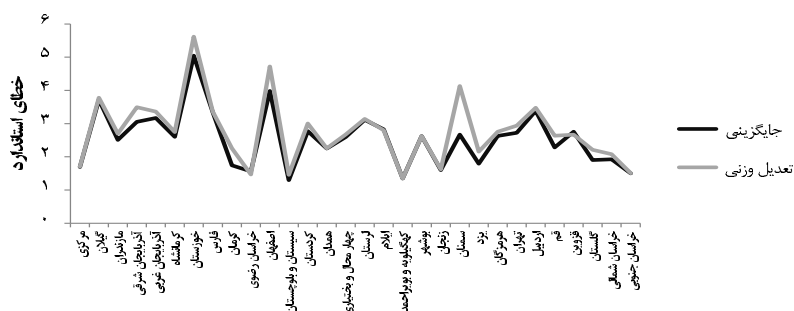
شکل ۲: برآورد خطای استاندارد میانگین هزینه‌ی کل تحت دو روش جایگزینی و تعدیل وزنی در نقاط روستایی سال ۱۳۸۷ (مقادیر بر حسب میلیون ریال)

بحث و نتیجه‌گیری

بی‌پاسخی جزء اجتناب ناپذیر هر آمارگیری است. قبل از انجام هر بررسی بایستی با به‌کارگیری روش‌های پیشگیری از بی‌پاسخی در حد امکان بی‌پاسخی را به حداقل رساند. بهر حال بی‌پاسخی در بیشتر آمارگیری‌ها روی می‌دهد و در تحلیل‌ها بایستی روش‌های تعدیل را برای مقابله با اریبی ناشی از بی‌پاسخی استفاده کرد. در این مقاله دو روش تعدیل بی‌پاسخی از جمله جایگزینی و تعدیل وزنی مورد بررسی قرار گرفت.



شکل ۳: برآورد خطای استاندارد میانگین هزینه‌ی کل تحت دو روش جایگزینی و تعدیل وزنی در نقاط شهری سال ۱۳۸۸ (مقادیر بر حسب میلیون ریال)



شکل ۴: برآورد خطای استاندارد میانگین هزینه‌ی کل تحت دو روش جایگزینی و تعدیل وزنی در نقاط روستایی سال ۱۳۸۸ (مقادیر بر حسب میلیون ریال)

با استفاده نتایج طرح هزینه و درآمد خانوار در سال‌های ۱۳۸۷ و ۱۳۸۸ تحت نرم افزار SAS، برآوردگر میانگین و خطای استاندارد آن برای چهار صفت هزینه‌ی کل، هزینه‌ی خوراک، هزینه‌ی غیر خوراک و درآمد خانوارها در سطح کل کشور و هر یک از استان‌ها با استفاده از روش ورما تحت هر دو روش جایگزینی و تعدیل وزنی محاسبه شد. روش برآورد واریانس تقریبی پیشنهاد شده توسط ورما که برآورد واریانس براساس مجموع‌های خوشه‌های نهایی یا براساس مقایسه بین تکرارها نیز عنوان می‌شود

یک روش تقریبی بر آورد داریانس برای محاسبه واریانس در آمارگیری‌های پیچیده است. نتایج در سطح نقاط شهری و روستایی کل کشور برای هر چهار صفت در جداول ۱ و ۲ و در سطح نقاط شهری و روستایی هر یک از استان‌ها تنها برای متغیر هزینه‌ی کل در نمودارهای ۱ تا ۴ آورده شده است. نتایج سه متغیر دیگر در سطح استان‌ها مشابه هزینه‌ی کل است که در اینجا آورده نشده است. همانطور که نتایج نشان می‌دهد خطای استاندارد تحت هر دو روش در سطح کل کشور تقریباً یکسان است. اما نمودارهای ۴-۱ نشان می‌دهد که در بعضی استان‌ها، روش جایگزینی اندکی دقیقتر از روش تعدیل وزنی عمل کرده و در بعضی دیگر روش تعدیل وزنی دارای خطای معیار کمتری است. اما در اکثر استان‌ها این دو روش تقریباً یکسان عمل می‌کنند. با این وجود در استفاده از روش جایگزینی توجه به این نکته مهم است که در صورت استفاده از روش جایگزینی ممکن است آگاهی آمارگیر از امکان دسترسی به واحد جایگزین از میزان تلاش‌های وی برای به دست آوردن اطلاعات از واحدهای نمونه‌گیری اولیه، بکاهد. در این صورت اگر دسترسی به واحدهای جایگزین نیز امکان پذیر نباشد، نرخ پاسخ در نمونه‌ی اصلی، کمتر از حد مورد انتظار خواهد بود. بنا بر این لازم است در آمارگیری‌ها حداکثر تلاش به عمل آید تا قبل از اینکه اجازه استفاده از خانوار جایگزین صادر شود، از هر یک از واحدهای نمونه‌ی اصلی تا جایی که امکان دارد اطلاعات مورد نظر گرفته شود. همچنین استفاده از آن ممکن است عملیات میدانی آمارگیری (اگر جایگزین‌ها بعد از تلاش زیادی که برای همکاری واحد نمونه اولیه می‌شود، انتخاب شوند) را طولانی کند.

مراجع

دستورالعمل طرح نمونه‌گیری آمارگیری از هزینه و درآمد خانوارهای شهری و روستایی سال ۱۳۸۷ و سال ۱۳۸۸، دفتر روش شناسی آماری و طرح‌های نمونه‌گیری، مرکز آمار ایران.

- Chapman, D.W. (1983), The impact of Substitution on Survey estimate, In *Incomplete Data in Sample Survey vol. 2*, Theory and Bibliographies, W.G. Madow, I. Olkin, and D.B. Rubin, eds. New York: Academic Press.
- Verma, V.,(2002), *Sampling Methods*, Training Handbook, Revised, Siap, Tokyo.
- Yansaneh, I. S.,(2003), Construction and Use of Sample Weights,*Handbook on Designing of Household Sample Surveys: Practical Guidelines*, Technical Reports.