

برآورد واریانس برآوردگر میانگین جامعه‌ای در نمونه‌گیری دوفازی در حضور بی‌پاسخی واحد

مرضیه نزل‌آبادی و حمیدرضا نواب‌پور*

دانشگاه علامه طباطبائی

چکیده: در آمارگیری‌های نمونه‌ای همواره با دو نوع خطا مواجه هستیم. خطای نمونه‌گیری و خطای غیرنمونه‌گیری. یکی از عمومی‌ترین خطاهای غیرنمونه‌گیری، بی‌پاسخی^۱ است. این خطا زمانی اتفاق می‌افتد که برخی از واحدهای نمونه‌ای مشاهده نشوند و یا مشاهده شوند ولی به برخی از پرسش‌ها پاسخ ندهند. پیش‌گیری کامل از این خطا امکان‌پذیر نیست ولی می‌توان آن را به‌طور قابل ملاحظه‌ای کاهش داد. بی‌پاسخی باعث اریبی و تورم واریانس برآوردگرها و در نتیجه موجب کم‌دقتی آماره‌های آمارگیری می‌شود. دو نوع بی‌پاسخی وجود دارد. بی‌پاسخی واحد آماری و بی‌پاسخی قلم اطلاعاتی. روش‌های متفاوتی برای تعدیل اثر هر یک از آن‌ها معرفی شده است. وزن‌دهی و جانهی از جمله روش‌های تعدیل اثر بی‌پاسخی پس از گردآوری داده‌ها هستند. وزن‌دهی برای تعدیل اثر بی‌پاسخی واحد آماری و جانهی نیز برای تعدیل اثر بی‌پاسخی قلم اطلاعاتی به‌کار می‌روند. طرح‌های متعددی وجود دارند که می‌توان با آن‌ها نمونه را انتخاب کرد. طرح نمونه‌گیری مورد نظر در این مقاله طرح نمونه‌گیری دوفازی^۲ طبقه‌بندی است که در آن بی‌پاسخی رخ داده است. هدف، برآورد واریانس میانگین نمونه‌ای در این طرح نمونه‌گیری با استفاده از روش برآورد واریانس جک‌نایفی^۳ است. همچنین در قالب یک مطالعه‌ی شبیه‌سازی برآوردگرهای واریانس جک‌نایفی برای دو برآوردگر (سوم و چهارم) از چهار برآوردگر معرفی‌شده، محاسبه می‌شوند و از نظر کارایی نسبی و ضریب تغییرات با هم مقایسه می‌شوند. مجموعه داده‌های مورد استفاده در این مطالعه‌ی شبیه‌سازی، داده‌های حاصل از طرح آمارگیری هزینه و درآمد خانوارهای شهری در سال ۱۳۹۴ مرکز آمار ایران است.

* نویسنده‌ی عهده‌دار مکاتبات

دریافت: ۱۳۹۷/۵/۱۲، پذیرش: ۱۳۹۸/۴/۵.

واژگان کلیدی: بی پاسخی؛ جانهی؛ برآورد واریانس جک نایفی؛ نمونه گیری دوفازی؛ وزن دهی.

۱- مقدمه

بی پاسخی یکی از عمومی ترین خطاهای غیرنمونه گیری است که در اغلب آمارگیری ها رخ می دهد. پیشگیری کامل از این خطا امکان پذیر نیست اما می توان آن را به طور قابل ملاحظه ای کاهش داد. دو نوع بی پاسخی وجود دارد. بی پاسخی واحد آماری و بی پاسخی قلم اطلاعاتی. اثرهای بی پاسخی اریبی و تورم واریانس برآوردگرها هستند. با افزایش اندازه نمونه می توان مسئله ی تورم واریانس برآوردگرها را برطرف کرد اما این کار به کاهش اریبی ناشی از بی پاسخی کمکی نخواهد کرد. از جمله روش هایی که برای تعدیل اثر بی پاسخی می توان نام برد، روش های وزن دهی، تماس مجدد و جانهی هستند. روش های وزن دهی و تماس مجدد برای تعدیل اثر بی پاسخی واحد آماری و روش های جانهی برای تعدیل اثر بی پاسخی قلم اطلاعاتی مورد استفاده قرار می گیرند [۹].

نمونه گیری دوفازی توسط نیمین [۱۰] معرفی شد. او برای دستیابی به پاسخ گویی بیشینه استفاده از متغیر کمکی را پیش نهاد داد. هم چنین نظریه ی به کارگیری طبقه بندی در فاز دوم را مطرح کرد. طرح نمونه گیری دوفازی طرحی است که در آن انتخاب نمونه در دو مرحله صورت می گیرد. ابتدا یک نمونه ی نسبتاً بزرگ از جامعه به عنوان نمونه ی فاز اول انتخاب می شود. در این مرحله متغیر کمکی مشاهده می شود. سپس این نمونه به عنوان جامعه در نظر گرفته شده و از آن یک نمونه ی کوچک تر به عنوان نمونه ی فاز دوم گزینش می شود. در این مرحله متغیر مورد مطالعه اندازه گیری می شود. نمونه گیری دوفازی زمانی مفید است که هزینه ی مشاهده ی متغیر کمکی ارزان تر از هزینه ی مشاهده ی متغیر مورد مطالعه باشد. برای اولین بار هنسین و هورویتس [۵] زیرنمونه گیری از بی پاسخ ها و استفاده از نمونه گیری دوفازی را برای طبقه بندی و به منظور برآورد میانگین و یا مجموع کل جامعه ای پیش نهاد دادند. یکی از روش های نمونه گیری که در آن از روش تماس مجدد برای تعدیل بی پاسخی استفاده می شود، روش نمونه گیری دوفازی است. روش نمونه گیری دوفازی در حضور بی پاسخی برای متغیر مورد مطالعه بر اساس انتخاب نمونه ای کوچک تر از درون نمونه ای بزرگ تر از جامعه است. هنگامی که برای متغیر مورد مطالعه بی پاسخی وجود دارد و اطلاعاتی برای متغیر کمکی در دسترس نیست، می توان از ترکیب

دو روش دوفازی و تماس مجدد استفاده کرد و دقت برآوردگرها را بهبود بخشید. مقاله‌های زیادی از جمله [۱]، [۲] و [۳] در ارتباط با تعدیل اثر بی‌پاسخی نگارش شده‌اند. یکی از مهم‌ترین معیارها در بیان میزان دقت یک برآوردگر یا طرح نمونه‌گیری استفاده از واریانس برآوردگر است. دقت برآوردگر واریانس برای آمارشناسان آمارگیری همواره مهم و مورد توجه بوده است. روش‌های تکراری برای برآورد واریانس ابزاری مفید برای این مسئله هستند. ولتر [۱۳] مروری جامع بر موضوع برآورد واریانس تکراری ارائه داد. برآورد واریانس برآوردگر میانگین جامعه‌ای به روش جک‌نایفی در نمونه‌گیری دوفازی زمانی که نمونه‌ی فاز دوم طبقه‌بندی شده باشد، توسط کیم و همکاران [۷] مورد بررسی قرار گرفت. روش جک‌نایف توسط کنویی [۱۱] در ارتباط با کاهش اریبی برآوردگرهای ناخطی معرفی شد. امکان استفاده از این روش با هدف برآورد واریانس توسط توکی [۱۲] ارائه شد. کیم و یو [۸] یک برآوردگر واریانس تعدیل‌یافته‌ی اریب را معرفی کردند که روش کیم و همکاران [۷] را گسترش می‌داد. روش ارائه‌شده توسط آن‌ها حتی زمانی که نرخ نمونه‌گیری فاز اول قابل اغماض نباشد، قابل استفاده است. هم‌چنین این روش زمانی که نمونه‌ی فاز دوم با استفاده از نمونه‌گیری پواسونی با احتمال نابرابر درون هر طبقه انتخاب شود، کاربرد دارد. در این مقاله برآورد واریانس تکراری برآوردگر میانگین جامعه‌ای در حضور بی‌پاسخی در نمونه‌گیری دوفازی ارائه می‌شود. این روش توسط جاوید و همکاران [۶] معرفی شد. برای برآورد واریانس در این حالت از روش ارائه‌شده توسط کیم و یو [۸] استفاده خواهد شد.

در بخش ۲ برآوردگرهای میانگین جامعه‌ای که با استفاده از دو روش برای برخورد با بی‌پاسخی در نظر گرفته شده‌اند، معرفی می‌شوند. در بخش ۳ واریانس این برآوردگرها به روش جک‌نایفی محاسبه خواهد شد و در بخش ۴ یک مطالعه‌ی شبیه‌سازی ارائه شده است.

۲- برآورد میانگین جامعه‌ای در حضور بی‌پاسخی

۲-۱- برآوردهای میانگین جامعه‌ای در حضور بی‌پاسخی در فاز اول

در این قسمت برآوردهای دوفازی طبقه‌بندی برای برآورد میانگین جامعه‌ای متغیر مورد مطالعه زمانی که در فاز اول با بی‌پاسخی (در مورد متغیر کمکی) مواجه شویم در دو حالت متفاوت زیر ارائه می‌شوند.

- نادیده گرفتن واحدهای بی‌پاسخ، و
- تماس مجدد با واحدهای بی‌پاسخ

۲-۱-۱- برآورد با نادیده گرفتن واحدهای بی‌پاسخ

فرض کنید نمونه‌ی فاز اول به اندازه‌ی n به روش تصادفی ساده‌ی بدون جای‌گذاری از یک جامعه‌ی متناهی به اندازه‌ی N گزینش شده است. این نمونه‌ی گزینش شده به دو گروه پاسخ‌گو و بی‌پاسخ به ترتیب با اندازه‌های نمونه‌ای n_r و $n_{\bar{r}}$ تقسیم می‌شود. یعنی بعد از گزینش نمونه و انجام آمارگیری $n_{\bar{r}}$ واحد بی‌پاسخ هستند. با استفاده از اطلاعات کمکی به دست آمده از واحدهای پاسخ‌گوی فاز اول، واحدها درون H طبقه، طبقه‌بندی می‌شوند. فرض کنید n_{rh} تعداد واحدهای نمونه‌ای پاسخ‌گوی فاز اول در طبقه‌ی h ام و A_{rh1} مجموعه اندیس‌های نمونه‌ی پاسخ‌گوی فاز اول در طبقه‌ی h ام باشند. حال یک نمونه به اندازه‌ی n'_r به روش تصادفی طبقه‌بندی با اندازه‌ی نمونه‌ای n'_{rh} در طبقه‌ی h ام به عنوان نمونه‌ی فاز دوم انتخاب می‌شود که $n'_r = \sum_{h=1}^H n'_{rh}$ و $\frac{n'_{rh}}{n_{rh}}$ نرخ نمونه‌گیری برای طبقه‌ی h ام است. هر چند که استنباط بر اساس داده‌های کامل (نادیده گرفتن واحدهای بی‌پاسخ) باعث اریبی برآوردها و به طور کلی موجب کاهش اعتبار استنباط‌ها می‌شود اما می‌توان این اثر را با استفاده از روش‌های وزن‌دهی تعدیل کرد. باید توجه داشت که متغیر مورد مطالعه، y ، در نمونه‌ی فاز دوم مشاهده می‌شود، بنا بر این میانگین جامعه‌ای $\bar{y}$ به صورت زیر برآورد می‌شود.

$$(۱) \quad \bar{y}_{ts1} = \frac{1}{n_r} \sum_{h=1}^H \sum_{i \in A_{rh1}} \frac{n_{rh}}{n'_{rh}} y_{rhi}$$

که در آن A_{rh2} مجموعه اندیس‌های واحدهای نمونه‌ای پاسخ‌گو در فاز دوم آمارگیری در طبقه h ام است. واریانس این برآوردگر به صورت زیر می‌باشد.

$$(2) \quad V(\bar{y}_{ts1}) = E \left[n_r^{-1} (1 - f_1) \sum_{h=1}^H w_h (\bar{y}_{rh1} - \bar{y}_1)^2 + \sum_{h=1}^H (n_{rh}' - n_{rh}^{-1} f_1) w_h^2 s_{rh1}^2 \right]$$

که در آن $\bar{y}_1 = \sum_{i=1}^{n_r} \frac{y_{ri}}{n_r}$ میانگین نمونه‌ی فاز اول، $\bar{y}_{rh1} = n_{rh}^{-1} \sum_{i \in A_{rh1}} y_{ri}$ ، $f_1 = \frac{n_{rh}}{n_r}$ و $w_h = \frac{n_{rh}}{n_r}$ هستند. برآوردگر سازگار واریانس $\bar{y}_{ts1}$ را می‌توان با جای‌گذاری $\bar{y}_{rh1}$ و s_{rh1}^2 به ترتیب با برآوردهای آن‌ها یعنی $\bar{y}_{rh2} = n_{rh}'^{-1} \sum_{i \in A_{rh2}} y_{ri}$ و $s_{rh2}^2 = (n_{rh}' - 1)^{-1} \sum_{i \in A_{rh2}} (y_{ri} - \bar{y}_{rh2})^2$ در (۲) به دست آورد. بنا بر این برآوردگر سازگار واریانس $\bar{y}_{ts1}$ به صورت زیر خواهد بود.

$$(3) \quad \hat{V}(\bar{y}_{ts1}) = n_r^{-1} (1 - f_1) \sum_{h=1}^H w_h (\bar{y}_{rh2} - \bar{y}_2)^2 + \sum_{h=1}^H (n_{rh}' - n_{rh}^{-1} f_1) w_h^2 s_{rh2}^2$$

که در آن $\bar{y}_2 = \sum_{h=1}^H w_h \bar{y}_{rh2}$

۲-۱-۲- برآورد از طریق تماس مجدد با واحدهای بی‌پاسخ

در این قسمت برآوردگر میانگین جامعه‌ای با استفاده از نظریه‌ی هنس و هورویتس [۵] برای تماس مجدد با واحدهای بی‌پاسخ ارائه می‌شود. مشابه حالت قبل از یک جامعه‌ی متناهی دارای N عضو، نمونه‌ای بزرگ به اندازه‌ی n به روش تصادفی ساده‌ی بدون جای‌گذاری گزینش می‌شود که n_r واحد پاسخ‌گو و n_f واحد بی‌پاسخ هستند. از بین n_f واحد بی‌پاسخ، زیرنمونه‌ای به روش تصادفی ساده‌ی گزینش شده و با آن‌ها تماس مجدد برقرار می‌شود. این تعداد گزینش شده با نماد n_{f*} نشان داده می‌شود. بعد از به دست آوردن اطلاعات کمکی از این زیر نمونه و نمونه‌ی پاسخ‌گوی اولیه، واحدها درون H طبقه، طبقه‌بندی می‌شوند. نمادهای n_{rh} و A_{rh1} مانند قبل تعریف می‌شوند. همچنین n_{f*h} تعداد واحدهای نمونه‌ای پاسخ‌گوی تماس مجدد در هر طبقه در فاز اول است. برای انجام نمونه‌گیری فاز دوم، نمونه‌ای به اندازه‌ی n_1' به روش تصادفی طبقه‌بندی از واحدهای پاسخ‌گوی اولیه انتخاب می‌شود. اندازه‌ی نمونه‌ای هر طبقه با نماد n_{1h}' نشان داده

می‌شود. هم‌چنین نمونه‌ای دیگر به اندازه‌ی n'_h با همان روش نمونه‌گیری از واحدهای پاسخ‌گوی دوم (از طریق تماس مجدد)، گزینش می‌شود. در این حالت اندازه‌ی نمونه‌ای هر طبقه با n'_{rh} نشان داده می‌شود به‌طوری‌که $n'_1 = \sum_{h=1}^H n'_{1h}$ ، $n'_2 = \sum_{h=1}^H n'_{2h}$ و $\frac{n'_{1h}}{n_{1h}}$ نرخ نمونه‌گیری برای هر طبقه‌ی پاسخ‌گوی فاز در فاز دوم است. هم‌چنین $\frac{n'_{2h}}{n_{2h}}$ نرخ نمونه‌گیری برای هر طبقه‌ی پاسخ‌گوی تماس مجدد فاز اول در فاز دوم است. همان‌طور که در قبل گفته شد، متغیر مورد مطالعه‌ی y در نمونه‌ی فاز دوم مشاهده می‌شود. میانگین جامعه‌ای متغیر y به‌صورت زیر برآورد می‌شود:

$$(4) \quad \bar{y}_{ts2} = \frac{n_r}{n} \left[\frac{1}{n_r} \sum_{h=1}^H \sum_{i \in A_{rh2}} \frac{n_{rh}}{n'_{1h}} y_{rhi} \right] + \frac{n_{\bar{r}}}{n} \left[\frac{1}{n_{\bar{r}}} \sum_{h=1}^H \sum_{i \in A_{\bar{r}h2}} \frac{n_{\bar{r}h}}{n'_{2h}} y_{\bar{r}hi} \right]$$

که در آن A_{rh2} مانند قبل تعریف می‌شود و $A_{\bar{r}h2}$ مجموعه‌ی اندیس‌های نمونه‌ی فاز دوم در طبقه‌ی h ام که با آن‌ها تماس مجدد برقرار شده است، می‌باشد. برآورد واریانس $\bar{y}_{ts2}$ عبارت است از:

$$(5) \quad \hat{V}(\bar{y}_{ts2}) = w_1^2 \left(n_r^{-1} (1 - f_1) \sum_{h=1}^H w_h (\bar{y}_{rh2} - \bar{y}_{r2})^2 \right) + \sum_{h=1}^H (n_{1h}^{-1} - n_{rh}^{-1} f_1) w_h^2 s_{rh}^2 + w_2^2 \left(n_{\bar{r}}^{-1} (1 - f_2) \sum_{h=1}^H w_h^* (\bar{y}_{\bar{r}h2} - \bar{y}_{\bar{r}2})^2 \right) + \sum_{h=1}^H (n_{2h}^{-1} - n_{\bar{r}h}^{-1} f_2) w_h^{*2} s_{\bar{r}h}^2$$

که در آن $w_1 = \frac{n_r}{n}$ و $w_2 = \frac{n_{\bar{r}}}{n}$ ، $f_1 = \frac{n_r}{N}$ ، $w_h^* = \frac{n_{\bar{r}h}}{n_{\bar{r}}}$ و $w_h = \frac{n_{rh}}{n_r}$.

۲-۲- برآورد میانگین جامعه‌ای در حضور بی‌پاسخی در فاز دوم

در این قسمت برآوردهای دوفازی طبقه‌بندی برای برآورد میانگین جامعه‌ای متغیر مورد مطالعه زمانی که در فاز دوم با بی‌پاسخی مواجه شویم در دو حالت متفاوت زیر ارائه می‌شوند.

- نادیده گرفتن واحدهای بی‌پاسخ، و
- تماس مجدد با واحدهای بی‌پاسخ

۲-۱-۲- برآورد با نادیده گرفتن واحدهای بی‌پاسخ

ابتدا از یک جامعه‌ی متناهی به اندازه‌ی N ، نمونه‌ای به اندازه‌ی n به روش تصادفی ساده‌ی بدون جای‌گذاری به‌عنوان نمونه‌ی فاز اول گزینش می‌شود. سپس با استفاده از اطلاعات به‌دست آمده از فاز اول این نمونه به H طبقه (برای انجام نمونه‌گیری فاز دوم) طبقه‌بندی می‌شود. تعداد واحدهای طبقه‌ی h ام را با n_h نشان می‌دهیم و نماد A_{h1} نشان‌دهنده‌ی مجموعه‌ی اندیس‌های نمونه‌ی فاز اول در طبقه‌ی h ام است. حال نمونه‌ای به اندازه‌ی n' به روش تصادفی طبقه‌بندی به‌عنوان نمونه‌ی فاز دوم گزینش می‌شود و اندازه‌ی نمونه‌ای در هر طبقه با n'_h نشان داده می‌شود که n'_{rh} تای آن‌ها پاسخ‌گو و $n'_{\bar{r}h}$ تای آن‌ها بی‌پاسخ می‌باشند به‌طوری‌که $n' = \sum n'_h = \sum (n'_{rh} + n'_{\bar{r}h})$. نرخ نمونه‌گیری برای هر طبقه به‌صورت $\frac{n'_{rh}}{n'_{rh}}$ تعریف می‌شود. متغیر مورد مطالعه‌ی y در نمونه‌ی فاز دوم مشاهده می‌شود. بنا بر این میانگین جامعه‌ای y به‌صورت زیر برآورد می‌شود:

$$(۶) \quad \bar{y}_{ts3} = \frac{1}{n} \sum_{h=1}^H \sum_{i \in A_{rh2}} \frac{n_h}{n'_{rh}} y_{hi}$$

که در آن A_{rh2} مجموعه اندیس‌های نمونه‌ی پاسخ‌گوی فاز دوم در طبقه‌ی h ام است. واریانس $\bar{y}_{ts3}$ به‌صورت زیر می‌باشد:

$$(۷) \quad V(\bar{y}_{ts3}) = E \left[n^{-1} (1 - f_1^*) \sum_{h=1}^H w_h^{\circ} (\bar{y}_{rh1} - \bar{y}_1)^2 + \sum_{h=1}^H (n'^{-1}_{rh} - n_h^{-1}) f_1^* w_h^{\circ} s_{rh1}^2 \right]$$

که در آن $w_h^{\circ} = \frac{n_h}{n}$ ، $f_1^* = \frac{n}{N}$ و $\bar{y}_1$ میانگین نمونه‌ی فاز اول هستند.

برآوردگر سازگار واریانس $\bar{y}_{ts3}$ را می‌توان به ترتیب با جای‌گذاری $\bar{y}_{rh1}$ و s_{rh1}^2 با برآوردگرهای آن‌ها یعنی

$$s_{rh2}^2 = (n'_{rh} - 1)^{-1} \sum_{i \in A_{rh2}} (y_{ri} - \bar{y}_{rh2})^2 \quad \text{و} \quad \bar{y}_{rh2} = n'^{-1}_{rh} \sum_{i \in A_{rh2}} y_{ri}$$

در (۷) به‌دست آورد. بنا بر این برآوردگر سازگار واریانس برآوردگر (۶) عبارت است از:

$$(۸) \quad \hat{V}(\bar{y}_{ts3}) = n^{-1} (1 - f_1^*) \sum_{h=1}^H w_h^{\circ} (\bar{y}_{rh2} - \bar{y}_2)^2 + \sum_{h=1}^H (n'^{-1}_{rh} - n_h^{-1}) f_1^* w_h^{\circ} s_{rh2}^2$$

که در آن $\bar{y}_2 = \sum_{h=1}^H w_h^{\circ} \bar{y}_{rh2}$.

۲-۲-۲- برآورد از طریق تماس مجدد با واحدهای بی پاسخ

حالتی را در نظر بگیرید که نمونه‌ی فاز اول به اندازه‌ی n به روش تصادفی ساده‌ی بدون جای‌گذاری از یک جامعه‌ی متناهی به اندازه‌ی N گزینش شده است. با استفاده از اطلاعات کمکی به دست آمده از این نمونه واحدها درون H طبقه (برای انجام نمونه‌گیری فاز دوم) طبقه‌بندی می‌شوند. تعداد واحدهای طبقه‌ی h ام با n_h نشان داده می‌شود. نماد A_{h1} نیز نشان‌دهنده‌ی مجموعه‌ی اندیس‌های واحدهای نمونه‌ی فاز اول در طبقه‌ی h ام است. اکنون یک نمونه‌ی تصادفی طبقه‌بندی به اندازه‌ی n' با اندازه‌ی نمونه‌ای n'_h در طبقه‌ی h به عنوان نمونه‌ی فاز دوم انتخاب می‌شود. از بین نمونه‌ی گزینش شده در طبقه‌ی h ام، n'_{rh} واحد، پاسخ‌گو و $n'_{\bar{r}h}$ واحد، بی‌پاسخ هستند به طوری که $n' = \sum_{h=1}^H n'_h = \sum_{h=1}^H (n'_{rh} + n'_{\bar{r}h})$. با توجه به این که بی‌پاسخی درون طبقه‌ها روی می‌دهد، در هر طبقه که بی‌پاسخی واحد دارد برای جبران اثر بی‌پاسخی روش تماس مجدد به کار می‌رود. به این ترتیب که درصدی از واحدهای بی‌پاسخ در هر طبقه به صورت تصادفی انتخاب می‌شوند و با آن‌ها تماس مجدد برقرار شده و فرض می‌شود که از تمام واحدهای نمونه‌ای پاسخ دریافت می‌شود. اندازه‌ی نمونه‌ای که در این روش در طبقه‌ی h ام گزینش می‌شود را با نماد $n'^*_{\bar{r}h}$ نشان می‌دهیم. نرخ نمونه‌گیری برای واحدهای پاسخ‌گوی طبقه‌ی h ام با $\frac{n'_{rh}}{n_h}$ و نرخ نمونه‌گیری برای واحدهای پاسخ‌گوی تماس مجدد در این طبقه با $\frac{n'^*_{\bar{r}h}}{n_h}$ نشان داده می‌شود. متغیر مورد مطالعه‌ی $\mathcal{Y}$ در نمونه‌فاز دوم مشاهده می‌شود. در این حالت، میانگین جامعه‌ای متغیر مورد مطالعه به صورت زیر برآورد می‌شود:

$$(9) \quad \bar{y}_{ts\mathcal{F}} = \frac{1}{n} \sum_{h=1}^H n_h \left(\frac{n'_{rh}}{n'_h} \bar{y}_{rh\mathcal{F}} + \frac{n'^*_{\bar{r}h}}{n'_h} \bar{y}_{\bar{r}h\mathcal{F}} \right)$$

که در آن $A_{\bar{r}^*h\mathcal{F}}$ و $A_{rh\mathcal{F}}$ ، $\bar{y}_{\bar{r}^*h\mathcal{F}} = \frac{1}{n'^*_{\bar{r}h}} \sum_{i \in A_{\bar{r}^*h\mathcal{F}}} y_{\bar{r}^*i}$ ، $\bar{y}_{rh\mathcal{F}} = \frac{1}{n'_{rh}} \sum_{i \in A_{rh\mathcal{F}}} y_{ri}$ که در آن $\mathcal{F}$ نیز مانند قبل تعریف می‌شوند. برآورد واریانس $\bar{y}_{ts\mathcal{F}}$ به صورت زیر می‌باشد:

$$\begin{aligned}\hat{V}(\bar{y}_{ts\psi}) &= \sum_{h=1}^H \left[\left((1 - w_{rh}^{\psi}) n_h^{-1} + (w_{rh}^{\psi} n_{rh}'^{-1} - n_h^{-1} f_1^*) \right) \right] w_h^{\psi} s_{rh\psi}^{\psi} \\ &\quad + \sum_{h=1}^H \left[\left((1 - w_{\bar{r}h}^{\psi}) n_h^{-1} + (w_{\bar{r}h}^{\psi} n_{\bar{r}h}'^{-1} - n_h^{-1} f_1^*) \right) \right] w_h^{\psi} s_{\bar{r}h\psi}^{\psi} \\ &\quad + n^{-1} (1 - f_1^*) \sum_{h=1}^H w_h^{\psi} \left[(\bar{y}_{rh\psi} - \bar{y}_{\psi})^2 + (\bar{y}_{\bar{r}h\psi} - \bar{y}_{\psi})^2 \right]\end{aligned}$$

که در آن $w_{rh}^{\psi} = \frac{n_{rh}'}{n_h}$ و $w_{\bar{r}h}^{\psi} = \frac{n_{\bar{r}h}'}{n_h}$ ، $w_h^{\psi} = \frac{n_h}{n}$ ، $\bar{y}_{\psi} = \sum_{h=1}^H w_h^{\psi} \bar{y}_{rh\psi}$ ، $f_1^* = \frac{n}{N}$ از آنجایی که دقت یک برآوردگر از مهم‌ترین معیارهای سنجش آن محسوب می‌شود در بخش بعد برآوردگرهای واریانس با دقت بالاتر معرفی می‌شوند.

۳- برآورد واریانس برآوردگرهای ارائه‌شده به روش جک‌نایفی

یکی از مهم‌ترین معیارها در بیان میزان دقت یک برآوردگر یا طرح نمونه‌گیری، واریانس برآوردگر مورد نظر است. در حالت کلی واریانس‌ها نامعلوم‌اند اما باید از داده‌های آماری برآورد شوند. تشخیص این‌که واریانس یک آماره تابعی از هر دو شکل آماره و ماهیت طرح نمونه‌گیری است، مهم می‌باشد. یکی از خطاهای رایج آمارشناسان اعتقاد به این است که شاید فرمول‌های نمونه‌گیری تصادفی ساده برای برآورد واریانس بدون در نظر گرفتن طرح یا برآوردگر به‌کار گرفته‌شده، قابل استفاده باشد. این اعتقاد نادرست است، زیرا برآوردگر واریانس باید هر دو مولفه‌ی برآوردگر و طرح نمونه‌گیری را در نظر بگیرد. دقت رابطه‌ای عکس با واریانس دارد. از میان طرح‌ها و برآوردگرهای مختلف، سعی می‌شود طرح و برآوردگری با بیش‌ترین دقت انتخاب شود. اما افزون بر ماهیت طرح یا برآوردگر، روش محاسبه‌ی واریانس نیز در کیفیت برآوردگر واریانس مؤثر است. روش‌های برآورد واریانس برآوردگرهای ناخطی عبارت‌اند از: روش خطی‌سازی تیلور، روش گروه‌های تصادفی و روش‌های تکراری.

روش‌های تکراری شامل سه روش، نیم‌نمونه‌ی متعادل، جک‌نایفی و خودگردان هستند. روش جک‌نایفی به‌عنوان یک روش مفید برآورد واریانس و هم‌چنین کاهش اریبی برآوردگرها پیش‌نهاد شده است. روش جک‌نایف به این‌صورت است که یک نمونه به اندازه‌ی n ، به روش تصادفی ساده‌ی با جای‌گذاری گزینش می‌شود. فرض کنید $\hat{\theta}$

برآوردگر پارامتر θ در نمونه‌ی n تایی است. فرض کنید، $\hat{\theta}_j$ برآوردگر پارامتر θ وقتی که از مشاهده‌ی j ام در برآورد استفاده نشود (برآورد پارامتر θ در زیرنمونه‌ی j ام)، باشد. بنا بر این برآورد واریانس به روش جک‌نایفی با حذف یک مشاهده در نمونه‌گیری تصادفی ساده‌ی با جای‌گذاری (در ایجاد هر زیرنمونه) به صورت زیر است [۱۳]:

$$\hat{V}_J(\hat{\theta}) = \frac{n-1}{n} \sum_{j=1}^n (\hat{\theta}_j - \hat{\theta})^2$$

واریانس جک‌نایفی برآوردگرهای معرفی شده در هر یک از چهار حالت بیان شده در قبل به صورت زیر است.

برآوردگر واریانس جک‌نایفی برآوردگر (۱) عبارت است از:

$$\hat{V}_{J1} = \sum_{k \in A_1} c_k \left(\bar{y}_{ts1}^{(k)} - \bar{y}_{ts1} \right)^2$$

که در آن $\bar{y}_{ts1}^{(k)} = \frac{1}{n_r^*} \sum_{h=1}^H n_{rh}^* \bar{y}_{rh1}^{(k)}$ و k نشان‌دهنده‌ی واحد حذف شده در تکرارهای جک‌نایف است. نماد A_1 مجموعه‌ی اندیس‌های واحدهای نمونه‌ی فاز اول و c_k عامل مربوط به k امین تکرار است که با استفاده از روش تکراری تعیین می‌شود. هم‌چنین

$$\frac{1}{n_r^*} n_{rh}^* = \begin{cases} (n_r - 1)^{-1} (n_{rh} - 1) & \text{اگر } k \in A_{rh1} \\ (n_r - 1)^{-1} n_{rh} & \text{اگر } k \notin A_{rh1} \end{cases}$$

و

$$\bar{y}_{rh2}^{(k)} = \begin{cases} (n'_{rh} - 1)^{-1} (n'_{rh} \bar{y}_{rh2} - y_{rk}) & \text{اگر } k \in A_{rh2} \\ \bar{y}_{rh2} & \text{اگر } k \notin A_{rh2} \end{cases}$$

برآوردگر واریانس جک‌نایفی برآوردگر (۲) عبارت است از:

$$\hat{V}_{J2} = \sum_{k \in A_1} c_k \left(\bar{y}_{ts2}^{(k)} - \bar{y}_{ts2} \right)^2$$

که در آن $\bar{y}_{ts\gamma}^{(k)} = \frac{w_{\gamma}}{n_r^*} \sum_{h=\gamma}^H n_{rh}^* \bar{y}_{rh\gamma}^{(k)} + \frac{w_{\gamma}}{n_{\bar{r}^*}^*} \sum_{h=\gamma}^H n_{\bar{r}^*h}^* \bar{y}_{\bar{r}^*h\gamma}^{(k)}$ هم چنین

$$\frac{\gamma}{n_r^*} n_{rh}^* = \begin{cases} (n_r - \gamma)^{-\gamma} (n_{rh} - \gamma) & k \in A_{rh\gamma} \text{ اگر} \\ (n_r - \gamma)^{-\gamma} n_{rh} & k \notin A_{rh\gamma} \text{ اگر} \end{cases}$$

و

$$\frac{\gamma}{n_{\bar{r}^*}^*} n_{\bar{r}^*h}^* = \begin{cases} (n_{\bar{r}^*} - \gamma)^{-\gamma} (n_{\bar{r}^*h} - \gamma) & k \in A_{\bar{r}^*h\gamma} \text{ اگر} \\ (n_{\bar{r}^*} - \gamma)^{-\gamma} n_{\bar{r}^*h} & k \notin A_{\bar{r}^*h\gamma} \text{ اگر} \end{cases}$$

و

$$\bar{y}_{rh\gamma}^{(k)} = \begin{cases} (n'_{\gamma h} - \gamma)^{-\gamma} (n'_{\gamma h} \bar{y}_{rh\gamma} - y_{rk}) & k \in A_{rh\gamma} \text{ اگر} \\ \bar{y}_{rh\gamma} & k \notin A_{rh\gamma} \text{ اگر} \end{cases}$$

و

$$\bar{y}_{\bar{r}^*h\gamma}^{(k)} = \begin{cases} (n'_{\gamma h} - \gamma)^{-\gamma} (n'_{\gamma h} \bar{y}_{\bar{r}^*h\gamma} - y_{\bar{r}^*k}) & k \in A_{\bar{r}^*h\gamma} \text{ اگر} \\ \bar{y}_{\bar{r}^*h\gamma} & k \notin A_{\bar{r}^*h\gamma} \text{ اگر} \end{cases}$$

به‌طور مشابه برآوردگر واریانس جک‌نایفی برآوردگر (۳) به‌صورت زیر است:

$$\hat{V}_{J\gamma} \sum_{k \in A_{\gamma}} c_k \left(\bar{y}_{ts\gamma}^{(k)} - \bar{y}_{ts\gamma} \right)^2$$

که در آن $\bar{y}_{ts\gamma}^{(k)} = \frac{\gamma}{n^*} \sum_{h=\gamma}^H n_h^* \bar{y}_{h\gamma}^{(k)}$ هم چنین

$$\frac{\gamma}{n^*} n_h^* = \begin{cases} (n - \gamma)^{-\gamma} (n_h - \gamma) & k \in A_{h\gamma} \text{ اگر} \\ (n - \gamma)^{-\gamma} n_h & k \notin A_{h\gamma} \text{ اگر} \end{cases}$$

و

$$\bar{y}_{h2}^{(k)} = \begin{cases} (n'_{rh} - 1)^{-1} (n_h - 1) & \text{اگر } k \in A_{h1} \\ (n - 1)^{-1} n_h & \text{اگر } k \notin A_{h1} \end{cases}$$

به همین ترتیب براوردگر واریانس جک‌نایفی براوردگر (۴) را می‌توان به صورت زیر نوشت:

$$\hat{V}_{J4} = \sum_{k \in A_1} c_k \left(\bar{y}_{ts4}^{(k)} - \bar{y}_{ts4} \right)^2$$

که در آن $\bar{y}_{ts4}^{(k)} = \frac{1}{n^*} \sum_{h=1}^H n_h^* \left(w_{rh}^* \bar{y}_{rh2}^{(k)} + w_{\bar{r}h}^* \bar{y}_{\bar{r}^*h2}^{(k)} \right)$ هم چنین

$$\frac{1}{n^*} n_h^* = \begin{cases} (n - 1)^{-1} (n_h - 1) & \text{اگر } k \in A_{h1} \\ (n - 1)^{-1} n_h & \text{اگر } k \notin A_{h1} \end{cases}$$

و

$$\bar{y}_{rh2}^{(k)} = \begin{cases} (n'_{rh} - 1)^{-1} (n'_{rh} \bar{y}_{rh2} - y_{rk}) & \text{اگر } k \in A_{rh2} \\ \bar{y}_{rh2} & \text{اگر } k \notin A_{rh2} \end{cases}$$

و

$$\bar{y}_{\bar{r}^*h2}^{(k)} = \begin{cases} (n'_{\bar{r}^*h} - 1)^{-1} (n'_{\bar{r}^*h} \bar{y}_{\bar{r}^*h2} - y_{\bar{r}^*k}) & \text{اگر } k \in A_{\bar{r}^*h2} \\ \bar{y}_{\bar{r}^*h2} & \text{اگر } k \notin A_{\bar{r}^*h2} \end{cases}$$

$$w_{rh}^* = (n'_h - 1)^{-1} (n'_{rh} - 1) \text{ و } w_{\bar{r}h}^* = (n'_h - 1)^{-1} (n'_{\bar{r}h} - 1)$$

۴- مطالعه‌ی شبیه‌سازی

در این قسمت با انجام یک مطالعه‌ی شبیه‌سازی دو براوردگر از چهار براوردگر معرفی شده در بخش ۲ (براوردهای سوم و چهارم) و واریانس‌های آن‌ها به روش جک‌نایفی که در بخش ۳ معرفی شدند، محاسبه می‌شوند. این کار به این دلیل انجام می‌شود که متغیر مورد

مطالعه در فاز دوم مشاهده می‌شود و بی‌پاسخی زمانی اهمیت ویژه خواهد یافت که در مورد متغیر مورد مطالعه اتفاق بیافتد. از آنجایی که این حالت در مورد متغیرهای سوم و چهارم روی می‌دهد از این برآوردگرها استفاده شده است. بی‌پاسخی در این مطالعه بر اساس ساختار گم‌شدگی تصادفی ساخته می‌شود. نوع بی‌پاسخی، بی‌پاسخی واحد آماری است. مجموعه داده‌های استفاده شده در این مطالعه شبیه‌سازی، داده‌های حاصل از طرح آمارگیری هزینه و درآمد خانوارهای شهری در سال ۱۳۹۴ است که توسط مرکز آمار ایران اجرا شده است [۴]. به‌طور خلاصه این شبیه‌سازی با اجرای گام‌های زیر انجام می‌شود:

- ۱- انتخاب نمونه‌ای به اندازه‌ی ۵۰۰۰ به روش نمونه‌گیری تصادفی ساده‌ی بدون جای‌گذاری از جامعه‌ی ۱۸۸۷۱ عضوی،
 - ۲- انتخاب نمونه‌های ۲۰۰، ۳۰۰ و ۵۰۰ تایی به عنوان نمونه‌ی فاز دوم،
 - ۳- ایجاد بی‌پاسخی تصادفی با نرخ‌های ۱۰، ۱۵ و ۳۰ درصد روی هزینه‌ی خالص با فرض اینکه بی‌پاسخی در هزینه وابسته به درآمد (متغیر کمکی) افراد است، و
 - ۴- تکرار گام‌های ۲ و ۳ به تعداد ۵۰۰۰ بار.
- فرض کنید در مورد متغیر کمکی اطلاعات کامل در دسترس باشد. بر اساس این متغیر، هزینه‌ی خالص در پنج رده طبقه‌بندی می‌شود و سپس در تمام رده‌ها بی‌پاسخی ایجاد می‌شود ولی به این صورت که در رده‌های با درآمد بالا و کم میزان بی‌پاسخی بیش‌تر از رده‌های دیگر است. به این ترتیب فرض گم‌شدگی تصادفی برقرار می‌شود.
- زمانی که بی‌پاسخی در فاز دوم وجود دارد، برای محاسبه‌ی برآوردگر معرفی‌شده در قسمت ۱-۱-۲ ابتدا یک نمونه با اندازه‌ی نمونه‌ای ۵۰۰۰ به روش تصادفی ساده‌ی بدون جای‌گذاری گزینش می‌شود. این واحدها طبقه‌بندی می‌شوند. فرض کنید ۳۰ درصد بی‌پاسخی در نمونه‌ی فاز دوم وجود دارد. بنا بر این با نادیده گرفتن واحدهای بی‌پاسخ، یک نمونه‌ی طبقه‌بندی فاز دوم به اندازه‌ی n'_{rh} از طبقه‌ی h ام گزینش می‌شود. برای حالت دوم، برآوردگر با استفاده از تماس مجدد با واحدهای بی‌پاسخ فاز دوم و با گزینش یک نمونه‌ی تصادفی طبقه‌بندی به اندازه‌ی n'_{r^*h} از $n' = \sum_{h=1}^H n'_{rh} + n'_{r^*h}$ که از واحدهای پاسخ‌گو و واحدهای تماس مجدد در هر طبقه تشکیل شده، محاسبه می‌شود.

برآوردهای معرفی شده در این مقاله بر اساس معیارهای برآورد کارایی نسبی و برآورد ضریب تغییرات در سه اندازه‌ی نمونه‌ای و سه درصد بی‌پاسخی مقایسه می‌شوند. این معیارها به صورت زیر تعریف می‌شوند.

برآورد کارایی نسبی به صورت زیر تعریف می‌شود:

$$\widehat{RE}(\varphi, \psi) = \frac{\overline{MSE}_{\psi}}{\overline{MSE}_{\varphi}}$$

که در آن $\overline{MSE} = \frac{1}{I} \sum_{i=1}^I (\hat{T}_i - \bar{T})^2$ برآورد میانگین توان‌های دوم خطا، $\hat{T}_i$ میانگین محاسبه شده در هر یک از تکرارهای مونته‌کارلویی، $\bar{T}$ میانگین میانگین‌های محاسبه شده در تکرارهای مونته‌کارلویی و منظور از (φ, ψ) برآوردگر سوم و چهارم هستند. برآورد ضریب تغییرات برای برآوردگر میانگین به صورت زیر تعریف می‌شود:

$$\widehat{CV}_m = \frac{\sqrt{\hat{V}(\hat{T})}}{\bar{T}}$$

که در آن $\hat{V}(\hat{T})$ برآورد واریانس مقدارهای برآورد شده در تکرارهای مونته‌کارلویی است. برآورد ضریب تغییرات برای برآوردگر واریانس به صورت زیر تعریف می‌شود:

$$\widehat{CV}_v = \frac{\sqrt{\frac{1}{I} \sum_{i=1}^I (V_{J_i}(\hat{T}) - \overline{MSE})^2}}{\overline{MSE}}$$

که در آن $V_{J_i}(\cdot)$ واریانس جک‌نایفی در هر بار تکرار و I تعداد تکرارهای مونته‌کارلویی هستند.

این مطالعه‌ی شبیه‌سازی با استفاده از نرم افزار R برنامه‌نویسی شده است. نتیجه‌های آن در جدول‌های زیر آورده شده است. واحدها بر حسب میلیون ریال هستند.

جدول ۱- برآوردها و معیارهای مقایسه در سه اندازه‌ی نمونه‌ای و درصدهای بی‌پاسخی

اندازه‌ی نمونه‌ای	درصد بی‌پاسخی	میانگین جامعه‌ای	برآورد میانگین		MSE		CV	
			سوم	چهارم	سوم	چهارم	سوم	چهارم
۲۰۰	۱۰	۲۲۱,۵۵۵	۲۲۲,۰۷	۲۲۱,۸۰۸	۹۷,۶۰۰۱	۸۸,۸۵۵۶	۴,۴۴۶۹	۴,۲۴۳۸
۲۰۰	۱۵	۲۲۱,۵۵۵	۲۲۱,۷۲۳	۲۲۱,۶۹۷	۱۰۴,۹۹۷	۹۲,۴۷۶۹	۴,۶۱۴۹	۴,۳۳۰۴
۲۰۰	۳۰	۲۲۱,۵۵۵	۲۲۲,۳۷۵	۲۲۱,۸۹۷	۱۲۷,۶۱۸	۸۸,۴۰۷۴	۵,۰۸۷۷	۴,۲۳۸۸
۳۰۰	۱۰	۲۲۱,۵۵۵	۲۲۱,۸۸۴	۲۲۱,۹۲۶	۶۵,۳۴۷۹	۵۷,۲۶۷۷	۳,۶۴۳۴	۳,۴۰۸۵
۳۰۰	۱۵	۲۲۱,۵۵۵	۲۲۲,۰۱۵	۲۲۱,۶۱۵	۷۱,۳۶۲۹	۶۱,۳۷۱۶	۳,۸۰۳۳	۳,۵۲۸۷
۳۰۰	۳۰	۲۲۱,۵۵۵	۲۲۲,۰۰۴	۲۲۱,۸۹۸	۸۴۰,۵۵۹	۵۹,۵۸۶	۴,۱۳۰۲	۳,۴۷۴۹
۵۰۰	۱۰	۲۲۱,۵۵۵	۲۲۲,۳۸۳	۲۲۱,۷۸۴	۳۸,۲۸۷۹	۳۵,۷۸۴۱	۲,۷۸۶۱	۲,۶۹۶
۵۰۰	۱۵	۲۲۱,۵۵۵	۲۲۱,۸۵۱	۲۲۱,۶۹۱	۴۳,۲۳۲۷	۳۵,۴۸۳۶	۲,۹۶۱۶	۲,۶۸۲۷
۵۰۰	۳۰	۲۲۱,۵۵۵	۲۲۲,۲۰۵	۲۲۱,۷۳۱	۵۲,۷۲۹۴	۳۶,۲۶۷۵	۳,۲۷۰۷	۲,۷۱۲۵

جدول ۲- مقایسه‌ی دو برآوردگر با استفاده از برآورد کارایی نسبی

اندازه‌ی نمونه‌ای	درصد بی‌پاسخی	RE
۲۰۰	۱۵	۱,۱۳۵۳
۲۰۰	۳۰	۱,۴۴۳۵
۳۰۰	۱۰	۱,۱۴۱
۳۰۰	۱۵	۱,۱۶۲۸
۳۰۰	۳۰	۱,۴۱۰۶
۵۰۰	۱۰	۱,۰۶۹۹
۵۰۰	۱۵	۱,۲۱۸۳
۵۰۰	۳۰	۱,۴۵۳۹

جدول ۳- واریانس‌های جک‌نایفی و ضریب تغییرات در سه اندازه‌ی نمونه‌ای و درصد بی‌پاسخی

اندازه‌ی نمونه‌ای	درصد بی‌پاسخی	برآورد واریانس جک‌نایفی		CV	
		سوم	چهارم	سوم	چهارم
۲۰۰	۱۰	۷۳,۸۲۲	۸۵,۹۹۴	۶۵,۴۱۸۷	۶۰,۵۰۹۱
۲۰۰	۱۵	۷۸,۸	۹۱,۸۲۴	۶۶,۹۷۱۳	۵۷,۹۸۲۸
۲۰۰	۳۰	۹۸,۸۸۵	۱۰۹,۱۳۷	۷۱,۳۴۶۵	۶۵,۴۷۹۳
۳۰۰	۱۰	۴۹,۱۸۲	۵۷,۲۲	۵۴,۴۲۵۱	۴۸,۲۶۰۲
۳۰۰	۱۵	۵۲,۵۴	۶۱,۶۶۴	۵۵,۴۵۶۹	۴۸,۲۱۹۱
۳۰۰	۳۰	۶۶,۲۶۵	۷۲,۹۳۱	۵۹,۹۳۸۴	۵۵,۶۹۰۹
۵۰۰	۱۰	۳۰,۸۶۴	۳۴,۶۹۲	۴۳,۳۹۸۹	۳۷,۴۲۷۶
۵۰۰	۱۵	۳۱,۴۴۵	۳۷,۵۳	۴۵,۵۱۴	۳۷,۶۸۷
۵۰۰	۳۰	۴۰,۲۴۲	۴۴,۶	۴۸,۷۴۷۴	۴۴,۷۸۴۶

همان‌طور که در جدول ۱ مشاهده می‌شود، با افزایش درصد بی‌پاسخی، برآوردهای میانگین توان‌های دوم خطا و ضریب تغییرات برای برآوردها سوم افزایش می‌یابند. در مورد برآوردها چهارم برای اندازه‌های نمونه‌ای ۲۰۰ و ۳۰۰ با افزایش درصد بی‌پاسخی معیارها ابتدا افزایش و سپس کاهش می‌یابند اما برای اندازه‌ی نمونه‌ای ۵۰۰ ابتدا کاهش و سپس افزایش می‌یابند. با افزایش اندازه‌ی نمونه‌ای نیز هر دو معیار مربوط به هر دو برآوردها، کاهش می‌یابند. میانگین‌های برآوردها نیز به مقدار میانگین جامعه‌ای نزدیک هستند. به‌منظور مقایسه‌ی بین برآوردها برآورد کارایی نسبی محاسبه شده است. همان‌طور که در جدول ۲ مشاهده می‌شود، تمام مقدارهای به‌دست آمده بزرگ‌تر از ۱ هستند. به این معنی که برآوردها چهارم کارایی بیشتری نسبت به برآوردها سوم دارد. نتیجه‌های به دست آمده در جدول ۳ نیز به این صورت است که با افزایش درصد بی‌پاسخی برآوردهای واریانس جک‌نایفی و ضریب تغییرات مربوط به هر دو برآوردها، افزایش می‌یابند و با افزایش اندازه‌ی نمونه‌ای هر دو مورد کاهش می‌یابند. با توجه به فرمول ضریب تغییرات برای واریانس، این احتمال وجود دارد که در حالت‌هایی که واریانس کمتر است CV بزرگ‌تر شود. به همین دلیل در جدول ۳، آن‌جایی که واریانس جک‌نایفی برای برآوردها سوم کوچک‌تر از برآوردها چهارم است برآورد ضریب تغییرات برآوردها سوم بزرگ‌تر از برآوردها چهارم است. با مقایسه‌ی نظیر به نظیر معیار محاسبه‌شده در این جدول، مشاهده می‌شود که مقدارهای به‌دست آمده برای برآوردها چهارم کوچک‌تر هستند. لذا با توجه به مشاهداتی که از هر سه جدول حاصل شد، نتیجه می‌گیریم که برآوردها چهارم که از طریق تماس مجدد با واحدهای بی‌پاسخ محاسبه شده است از برآوردها سوم که بر اساس داده‌های کامل (نادیده گرفتن واحدهای بی‌پاسخ) به‌دست آمده، بهتر عمل می‌کند.

۵- بحث و نتیجه‌گیری

زمانی که اندازه‌گیری متغیر پاسخ زمان‌بر و هزینه‌بر باشد و متغیر کمکی در دسترس نباشد می‌توان روش نمونه‌گیری دوفازی را به کار برد که از اندازه نمونه‌ی کوچک‌تری استفاده می‌کند. نمونه‌گیری دوفازی مانند تمام آمارگیری‌ها با مسئله‌ی بی‌پاسخی مواجه است. این

مسئله دقت آماره‌های آمارگیری را تحت تأثیر قرار می‌دهد. این مقاله تلاش کرده است تا با تعدیل وزن‌های نمونه‌گیری سهم این اثرها را کاهش دهد. در یک کاربرد با استفاده از داده‌های آمارگیری هزینه و درآمد خانوار، اثر بی‌پاسخی واحد (اریبی و تورم واریانس) در داده‌های این آمارگیری با تغییر وزن‌های نمونه‌گیری تعدیل شد. این مطالعه حالتی را در نظر گرفته که در آن بی‌پاسخی واحد در فاز دوم روی داده است. در این حالت با تغییر وزن‌های آمارگیری در فاز دوم اثر بی‌پاسخی واحد تعدیل شد.

توضیحات

1. Non-response
2. Two Phase Sampling
3. Jackknife

مرجع‌ها

- [۱] امینی فارسانی، زهرا و نواب‌پور، حمیدرضا (۱۳۹۰). کاهش اریبی بی‌پاسخی توسط برآوردگرهای کالبیده در مطالعه‌های پانلی. مجله‌ی بررسی‌های آمار رسمی ایران، دوره‌ی ۲۲، شماره‌ی ۱، ۴۵-۶۲.
- [۲] تقربی، مریم؛ نعمت‌الهی، نادر و نواب‌پور، حمیدرضا (۱۳۹۲). برآورد میانگین جامعه در حضور بی‌پاسخی با استفاده از روش نمونه‌گیری دوفازی. مجله‌ی بررسی‌های آمار رسمی ایران، دوره‌ی ۲۴، شماره‌ی ۱، ۱۹-۴۱.
- [۳] خطیبی نوری، شهرزاد و نواب‌پور، حمیدرضا (۱۳۹۲). مقایسه‌ی تجربی عمل‌کرد و رویکرد یک‌پارچه خطی‌سازی در برآورد واریانس پس از جانمایی با چند روش دیگر. مجله‌ی پژوهش‌های آماری ایران، دوره‌ی ۱۰، شماره‌ی ۲، ۱۲۵-۱۴۶.
- [۴] مرکز آمار ایران (۱۳۹۴). آمارگیری از هزینه و درآمد خانوارهای شهری سال ۱۳۹۴. مرکز آمار ایران، تهران.
- [5] Hansen, M.H. and Hurwitz, W.N. (1946). The Problem of Non-response in Sample Surveys. JASA, **41**, 517-529.
- [6] Javed, M., Ahmad, Z. and Hanif, M. (2014). Replication Variance Estimation under Two-phase Sampling in the Presence of Non-

response. STATISTICA, anno LXXIV, n. 3.

- [7] Kim, J.K., Navarro, A. and Fuller, W.A. (2006). Replication Variance Estimation for Two-phase Stratified Sampling. Journal of the American Statistical Association, **101**, 312-320.
- [8] Kim, J.K. and Yu, C.L. (2011). Replication Variance Estimation under Two Phase Sampling. Statistic Sinica, **13**, 67-74.
- [9] Lohr, S.L. (2010). Sampling Design and Analysis. 2nd Edition, Duxbury Press, Pacific, Grove, California.
- [10] Neyman, J. (1938). Contribution to the Theory of Sampling Human Populations. Journal of the American Statistical Association, **33**, 101-116.
- [11] Quenouille, M.H. (1956). Notes on Bias in Estimation, Biometrika, **43**, 353-360.
- [12] Tukey, J. (1958). Bias and Confidence in Not Quite Large Samples. Ann. Math. Statist., **29**, 614.
- [13] Wolter, K. (2007). Introduction to Variance Estimation. second edition, Springer, New York.

مرضیه نزل آبادی

فوق لیسانس آمار

تهران، خیابان شهید بهشتی، خیابان خالد استانیولی (وزرا)، کوچه ی هفتم، شماره ی ۱۴، دانشکده ی علوم ریاضی و رایانه ی دانشگاه علامه طباطبائی.

رایانشانی: marzieh.nazlabadi@yahoo.com

حمیدرضا نوابپور

دانشیار آمار

تهران، خیابان شهید بهشتی، خیابان خالد استانیولی (وزرا)، کوچه ی هفتم، شماره ی ۱۴، دانشکده ی علوم ریاضی و رایانه ی دانشگاه علامه طباطبائی.

رایانشانی: hnnavvabpour@yahoo.com