مجله‌ی بررسی‌های آمار رسمی ایران
سال۲۲، شماره‌ی ۲، پاییز و زمستان ۱۳۹۰، صص ۱۳۶ – ۱۲۵

مقایسه‌ی روش‌های WGEEs و WEEs برای تعدیل اثر بی‌پاسخی در آمارگیری‌های پانلی

حمیدرضا نواب‌پور* و ناصر فروتن

دانشگاه علامه طباطبایی

چکیده: در اکثر آمارگیری‌ها بی‌پاسخی واحد یک مشکل اساسی است. این مشکل در آمارگیری‌های پانلی نیز رخ می‌دهد. در این آمارگیری‌ها معمولاً با کاهش پاسخ‌گو از دور دوم آمارگیری به بعد مواجه هستیم. مهم‌ترین مشکل ناشی از بی‌پاسخی واحد، اریبی در آماره‌های آمارگیری است. برای برخورد با این مشکل در آمارگیری‌های پانلی می‌توان از روش‌های وزن‌دهی استفاده کرد. در این مقاله با شبیه‌سازی گم‌شدگی در دور دوم آمارگیری، در داده‌های حاصل از آمارگیری پانلی خانواری انگلیس، اثرهای ناشی از گم‌شدگی توسط روش وزن‌دهی روبینز و همکارانش [۶]تعدیل شده و با بکار بردن مدل مورد استفاده در مطالعه، مدل حاشیه‌ای، مقدارهای گم‌شده در دوره‌ی دوم براورد می‌شوند. سپس عمل‌کرد این روش نسبت به روش دیگر تعدیل اریبی که تحت عنوان معادله‌های براورد موزون نام برده می‌شود، مقایسه می‌شود. نتیجه‌های این مطالعه نشان می‌دهد که (برحسب اندازه‌ی نمونه‌ای، همبستگی‌های مختلف بین دو دوره‌ی آمارگیری و نرخ بی‌پاسخی) روش معادله‌های براورد تعمیم‌یافته‌ی موزون روبینز و همکارانش [۶]در همه‌ی حالت‌ها از روش معادله‌های براورد موزون عمل‌کرد بهتری دارد.

واژگان کلیدی: آمارگیری طولی؛ بی‌پاسخی؛ وزن‌دهی؛ معادله‌های براورد موزون؛ معادله‌های براورد تعمیم‌یافته‌ی موزون.

---

* نویسنده‌ی عهده‌دار مکاتبات
دریافت: ۱۳۸۹/۱۱/۳، پذیرش: ۱۳۹۱/۸/۲۳

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

۱- مقدمه

تقریباً در تمامی آمارگیری‌ها با مسئله‌ی بی‌پاسخی واحد و گم‌شدگی داده‌های مربوط مواجه هستیم. بی‌پاسخی در حقیقت به شکست در گرداوری داده از واحد نمونه‌ای گفته می‌شود. بی‌پاسخی به یکی از دو صورت «بی‌پاسخی واحد» و «بی‌پاسخی پرسش (قلم اطلاعاتی)» روی می‌دهد. بی‌پاسخی واحد به عدم موفقیت در گرداوری هر نوع اطلاعاتی از واحد نمونه‌ای و بی‌پاسخی پرسش به عدم موفقیت در گرفتن پاسخ برای برخی از پرسش‌های پرسش‌نامه اطلاق می‌شود. در این مقاله ما اثر بی‌پاسخی واحد را در آمارگیری‌های پانلی مورد بررسی قرار می‌دهیم.

بی‌پاسخی واحد می‌تواند به‌دلیل‌های مختلفی رخ دهد. شاید اولین علت بی‌پاسخی، عدم دسترسی به واحد آمارگیری باشد. دلیل دیگر بی‌پاسخی واحد می‌تواند مربوط به طبیعت آمارگیری باشد. برای مثال افراد با درامد بالا معمولاً از شرکت در آمارگیری‌هایی که به درامد شخصی مربوط می‌شود امتناع می‌ورزند. از طرفی نگرش‌های قومی، مذهبی، سیاسی و ... نیز ممکن است در ایجاد بی‌پاسخی واحد در آمارگیری‌ها مؤثر باشند. بنا بر این وجود بی‌پاسخی در آمارگیری‌ها امری غیر قابل اجتناب است.

از نظر آماری، بی‌پاسخی واحد موجب دو نوع مشکل می‌شود. اول کاهش واحدهای نمونه‌ای (خارج شدن برخی از واحدهای نمونه‌ای از مطالعه) است که موجب کاهش دقت براوردها شده و دومی اریبی در آماره‌های آمارگیری را سبب می‌شود. اگر واحدهای بی‌پاسخ یک زیرمجموعه‌ی تصادفی از نمونه نباشند، یا به‌عبارتی ویژگی‌های آن‌ها به‌طور سیستماتیک با ویژگی‌های پاسخ‌گویان متفاوت باشد، نمونه‌ی پاسخ‌گویان براوردهای اریب برای پارامترهای جمعیت مورد نظر به‌دست می‌دهد.

دو روش برای تعدیل اثرهای بی‌پاسخی وجود دارند، یکی در مرحله‌ی طراحی و دیگری در مرحله‌ی اجرا. در زمان اجرا سعی بر مینیمم کردن اثرهای بی‌پاسخی در مرحله‌ی گرداوری داده‌هاست. این کار می‌تواند با تعریف کردن دقیق اندازه‌ها انجام شود که افزایش نرخ پاسخ را مورد هدف قرار می‌دهد. تعدیل بی‌پاسخی در مرحله‌ی تحلیل داده‌ها با روش‌های وزن‌دهی و جانهی صورت می‌گیرد [۳]. بنا بر این، منظور از تعدیل بی‌پاسخی استفاده از روش‌هایی در مرحله‌ی تحلیل، برای مینیمم کردن اثرهای بی‌پاسخی بر آماره‌های آمارگیری است. در این مقاله از روش دوم برای تعدیل اریبی بی‌پاسخی استفاده می‌شود.

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

ساده‌ترین روش ممکن برای برخورد با مسئله‌ی گم‌شدگی در تحلیل داده‌ها این است که تحلیل مورد نظر را به واحدهایی با داده‌های کامل روی متغیرهای موجود محدود کنیم (این روش بیش‌تر توسط بسته‌های نرم‌افزاری زمانی که بخواهیم تحلیل‌های چندگانه مانند تحلیل‌های توصیفی انجام دهیم به‌کار می‌رود). این روش تحلیل، تحلیل کامل مورد نامیده می‌شود. براوردهای به‌دست آمده از این تحلیل‌ها در صورتی که در اکثر متغیرهای کلیدی آمارگیری به‌طور سیستماتیک بین واحدهای به‌کار رفته در تحلیل و واحدهای کنار گذاشته شده اختلاف وجود داشته باشد، اریب خواهد بود.

غالباً ساختار گم‌شدگی در ماتریس داده‌ها با دو صفت «الگوی گم‌شدگی» و «مکانیزم گم‌شدگی» توصیف می‌شود. الگوی گم‌شدگی به نحوه‌ی به قرار گرفتن نقاط گم‌شده در چیده‌مان ماتریس داده‌ها اشاره دارد. در حقیقت منظور از الگوی گم‌شدگی داده‌ها، پاسخ به این پرسش است که کدام داده‌ها در ماتریس داده‌ها گم‌شده‌اند، در حالی که مکانیزم گم‌شدگی به دنبال پاسخ برای این پرسش است که چرا داده‌ها گم‌شده‌اند و این گم‌شدگی چگونه به متغیرهای موجود در مطالعه ارتباط دارد. روبین در سال ۱۹۷۶ با در نظر گرفتن تابع توزیع برای تابع نشانگر گم‌شدگی، اقدام به فرمول‌بندی مکانیزم‌های گم‌شدگی و تقسیم‌بندی آن‌ها به سه حالت کلی کاملاً تصادفی، تصادفی و غیرتصادفی کرد [۷].

روش‌های برخورد با مقدارهای گم‌شده از اوایل ۱۹۷۰ با پیشرفت تکنولوژی محاسباتی رونق یافت. روش معادله‌های براورد تعمیم‌یافته (Generalized Estimating Equations) برای اولین بار توسط لیانگ و زیگر [۵] و زیگر و لیانگ [۸] به‌عنوان روشی برای تحلیل داده‌های طولی ارایه شد. روبینز و همکارانش [۶] تعدیل اریبی ناشی از بی‌پاسخی واحدها در این نوع مطالعه را با روش وزن‌دهی معادله‌های براورد تعمیم‌یافته‌ی لیانگ و زیگر [۵] انجام دادند که در آن وزن برای هر واحد، به‌صورت یک ماتریس قطری در نظر گرفته می‌شد. همچنین فیتزموریس و همکارانش [۲] نیز نوع دیگر وزن‌دهی برای معادله‌های براورد تعمیم‌یافته‌ی لیانگ و زیگر [۵] را ارایه کردند که در آن وزن برای هر واحد نمونه‌ای تنها به‌صورت یک عدد در نظر گرفته می‌شود.

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

۲- معادله‌های براورد تعمیم‌یافته (GEEs)

فرض کنید در یک مطالعه‌ی طولی N واحد مستقل از هم انتخاب شده‌اند، و واحد iام T_i بار مشاهده شده است. برای هر واحد نمونه‌ای فرض می‌کنیم که متغیر پاسخ به‌صورت $Y_i = (y_{i1}, y_{i2}, \dots, y_{iT_i})^T$ باشد که در آن $i = 1, 2, \dots, N$.

روش معادله‌های براورد تعمیم‌یافته از طریق میانگین و واریانس حاشیه‌ای Y_i تعیین می‌شود. فرض می‌شود که Y_i از طریق تابع ربط h با متغیرهای کمکی X_i همبسته است، یعنی داریم:

که در آن β بردار ضریب‌های رگرسیونی و X_{it} بردار متغیرهای کمکی برای فرد iام در دوره‌ی t می‌باشند. در این صورت معادله‌های براورد تعمیم‌یافته عبارت‌اند از:

$$S(\beta, \alpha, \phi) = \sum_{i=1}^{N} D_i^T V_i^{-1} (Y_i - \mu_i) \qquad (1)$$

که در آن $D_i = \partial \mu_i / \partial \beta$ ماتریس مشتق‌های جزئی و $V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2} \emptyset$ نمایانگر ماتریس پراکندگی کاری Y_i است، که در آن A_i ماتریس قطری با عنصرهای قطری $v(\mu_{it})$ (v تابع واریانس معلوم) است و $R_i(\alpha)$ نیز ماتریس همبستگی کاری (چون اغلب همبستگی واقعی را نمی‌دانیم، لذا اصطلاح همبستگی کاری به آن اطلاق می‌شود) و ϕ پارامتر پراکندگی است. ماتریس همبستگی کاری $R_i(\alpha)$ معمولاً نامعلوم است. این ماتریس به‌صورت عددی و در یک فرایند تکراری براورد می‌شود. پس از محاسبه کردن $\hat{R}_{i(\alpha)}$، ماتریس پراکندگی کاری به‌صورت $\hat{V}_i = A_i^{1/2} \hat{R}_i(\alpha) A_i^{1/2} \hat{\phi}$ براورد می‌شود که در آن $\hat{\phi} = \frac{1}{k-p} \sum_{i=1}^{n} \sum_{j=1}^{t} e_{ij}^2$ و n اندازه‌ی نمونه‌ای، t دوره‌های آمارگیری، k کل مقدارهای اندازه‌گیری‌شده، p تعداد پارامترهای مدل و $e_{ij} \left(= \frac{y_{ij} - \mu_{ij}}{\sqrt{a_{ij}}} \right)$ مانده‌ی پیرسونی هستند. روش GEE با برخی از نرم‌افزارهای موجود مانند SAS، STATA و R قابل اجرا است.

معادله‌ی بالا در صورتی که مکانیزم گم‌شدگی از نوع کاملاً تصادفی باشد، منجر به براوردهای سازگار پارامتر β می‌شود، زیرا در این حالت معادله‌های بالا تنها به متغیرهای

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

کمکی وابسته هستند. ولی در صورتی که مکانیزم گم‌شدگی از نوع تصادفی باشد معادله‌ی بالا برای براورد β دیگر سازگار نیست، زیرا در این حالت معادله‌های بالا به متغیرهای کمکی، هم از طریق واحدهای مشاهده‌شده و هم از طریق واحدهای مشاهده‌نشده وابسته است.

روبینز و همکارانش برای کاهش اریبی و دست‌یابی به براوردهای سازگار، معادله‌های براورد تعمیم‌یافته‌ی موزون (Weighted Generalized Estimating Equations) را پیشنهاد کردند [۶]. این روش با انتساب وزن‌های جدید به پاسخ‌های مشاهده‌شده‌ی آزمودنی‌ها، اثر اریبی ایجاد شده‌ی ناشی از گم‌شدگی را تعدیل می‌کند. در این روش، براورد سازگار پارامتر β از حل معادله‌ی زیر به‌دست می‌آید:

$$\sum_{i=1}^{N} D_i^T V_i^{-1} W_i (Y_i - \mu_i) = 0. \qquad (۲)$$

در معادله‌ی بالا، W_i نشان‌دهنده‌ی وزن اختصاص‌داده شده به واحد iام است که یک ماتریس قطری با عنصرهای قطری $R_{it}w_{it}$ است، بنا بر این برای واحد iام داریم $W_i = \{R_{i1}w_{i1}, R_{i2}w_{i2}, \dots, R_{iT}w_{iT}\}$. از طرفی $R_{it} = 1$، اگر پاسخ واحد iام در زمان t مشاهده‌شده باشد و در غیر این صورت $R_{it} = 0$.

۳- روش معادله‌های براورد موزون (WEEs)

در رگرسیون معمولی وقتی فرض‌های توزیعی درباره‌ی خطا درست باشند، براوردهای ماکسیمم درستنمایی برای داده‌های گم‌شده مورد استفاده قرار می‌گیرند. اگر این فرض‌ها برقرار نباشند این براوردگرها از مطلوبیت کم‌تری برخوردارند. برای دست‌یابی به براوردهای سازگار در این حالت، از معادله‌های براورد استفاده می‌شود. این معادله‌ها عبارت‌اند از:

$$\sum_{i=1}^{N} R_i d_i v_i^{-1} (y_i - \mu_i) = 0 \qquad (۳)$$

که در آن $d_i = \frac{\partial \mu_i}{\partial \beta}$ و R_i نشانگر پاسخ فرد iام است ($R_i = 1$ اگر فرد iام پاسخ دهد، $R_i = 0$ اگر فرد iام پاسخ ندهد) μ_i میانگین پاسخ فرد iام است، که به‌عنوان تابعی از

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

متغیرهای کمکی مدل‌بندی می‌شود. اگر داده‌ها به‌طور کاملاً تصادفی گم‌شده باشند، آن‌گاه این معادله‌ها براورد سازگاری از β فراهم می‌کنند ولی اگر داده‌ها به‌طور تصادفی گم‌شده باشند در این صورت برای تعدیل بی‌پاسخی در معادله‌های بالا از معادله‌های براورد موزون (Weighted Estimating Equations) استفاده می‌شود. این معادله‌ها عبارت‌اند از

$$\sum_{i=1}^{N} R_i w_i d_i v_i^{-1} (y - \mu) = 0 \qquad (۴)$$

که در آن w_i وزن واحد iام است. در این معادله‌ها برای تعدیل بی‌پاسخی می‌توان از روش‌های وزن‌دهی از جمله پساطبقه‌بندی [۴]، وزن‌دهی رگرسیونی تعمیم‌یافته، چنگک‌زنی و کالبیدن [۱] استفاده کرد. در این مقاله چون ما با یک بردار از متغیرهای کمکی که در ارتباط با متغیر پاسخ قرار دارد، روبرو هستیم با توجه به این‌که وزن‌دهی رگرسیونی تعمیم‌یافته از متغیرهای کمکی موجود در مطالعه استفاده‌ی بهینه می‌کنند، از این شیوه‌ی وزن‌دهی در معادله‌های براورد موزون استفاده کرده‌ایم.

۴- مطالعه‌ی شبیه‌سازی

داده‌های استفاده شده در این مقاله از آمارگیری پانلی خانواری انگلیس انتخاب شده است. آمارگیری پانلی خانواری انگلیس از سال ۱۹۹۱ توسط مرکز مطالعه‌های طولی انگلیس همراه با مؤسسه‌ی تحقیق‌های اقتصادی و اجتماعی دانشگاه اسکس هر سال اجرا می‌شود. موضوع اصلی آمارگیری در سه دوره‌ی اول سازمان‌دهی خانوارها، بازار کار، درامد و دارایی، مسکن، معیارهای سلامتی و اجتماعی اقتصادی بوده‌اند. بنا بر این امکان تحقیق در سطح گسترده‌ای از موضوع‌ها برای پژوهش‌گرها وجود دارد. هدف اصلی آمارگیری، یافتن تغییرهای اقتصادی و اجتماعی در سطح فرد و خانوار در انگلیس است. انتخاب مدل، پیش‌گویی و شرح علت تغییرها در ارتباط با رده‌ای از متغیرهای اقتصادی اجتماعی از هدف‌های این آمارگیری است.

با مطالعه‌ی پرسش‌نامه‌های استفاده‌شده در این آمارگیری، پرسش‌نامه‌ای که مربوط به افراد ۱۱ تا ۱۶ سال است، مورد استفاده قرار گرفت. این پرسش‌نامه، از دور چهارم وارد آمارگیری شده است و در تمامی دوره‌های بعدی مورد استفاده قرار گرفته است. در این مقاله از داده‌های دوره‌ی شانزدهم و هفدهم آمارگیری با این پرسش‌نامه استفاده شده است.

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

متغیر پاسخ "سیگاری بودن فرد" در نظر گرفته شده است به‌طوری که اگر فرد در دوره‌ی مورد نظر سیگاری باشد مقدار یک و اگر سیگاری نباشد مقدار صفر می‌گیرد. علاوه بر انتخاب متغیر پاسخ، باید متغیرهای کمکی مناسبی نیز انتخاب شوند. متغیرهای جنسیت، مصرف میوه، نظر فرد در مورد خودش از لحاظ روحی و نظر فرد در مورد خانواده‌اش از لحاظ روحی به‌عنوان متغیرهای کمکی اختیار شدند. با استفاده از نرم‌افزار SAS داده‌های دوره‌ی شانزدهم و هفدهم پرسش‌نامه‌ی مورد نظر را کنار هم قرار دادیم.

برای انجام تحلیل‌های مورد نظر روی داده‌های پانلی، نیازمندیم واحدهای آماری را انتخاب کنیم که علاوه بر این‌که در دوره‌ی شانزدهم آمارگیری حضور داشته‌اند در دوره‌ی هفدهم نیز در آمارگیری مشارکت داشته‌اند. بدین ترتیب جامعه‌ی آماری شامل ۹۴۰ نفر شد. نمونه‌هایی به اندازه‌ی ۵۰، ۱۰۰ و ۲۰۰ به‌صورت تصادفی ساده و بدون جای‌گذاری از این جامعه انتخاب شده و نرخ‌های کاهش پاسخ‌گو به‌صورت تصادفی با نرخ‌های ۰/۱ و ۰/۱۵ در نظر گرفته شدند. همبستگی بین دو دوره نیز در سه مقدار ۰/۴، ۰/۶ و ۰/۸ اختیار شدند. لازم به ذکر است که همبستگی بین داده‌های جامعه در دو دوره ۰/۶۴۲ و کاهش پاسخ‌گو در دور دوم آمارگیری روی می‌دهد.

برای براورد توزیع نمونه‌گیری آماره‌ی تغییر نسبت در دو دوره‌ی آمارگیری، به اندازه‌ی ۱۰۰۰ مرتبه نمونه‌گیری مستقل از هر دوره‌ی آمارگیری انجام شد و براورد تغییر نسبت پاسخ و واریانس آن به‌دست آمد. برای ایجاد بی‌پاسخی دوره با ساختار گم‌شدگی تصادفی از دستور SimulateMissings در نرم‌افزار R استفاده شد.

۱-۴- معیارهای مقایسه

برای مقایسه‌ی دو روش تعدیل اثر بی‌پاسخی از دو معیار براورد قدر مطلق اریبی نسبی و براورد کارایی نسبی مجانبی استفاده کرده‌ایم. در هر دو روش تعدیل اثر بی‌پاسخی مورد نظر، برای محاسبه‌ی براورد قدر مطلق اریبی، ابتدا عمل نمونه‌گیری برای حالت‌های مورد نظر با استفاده از نرم‌افزار R به تعداد ۱۰۰۰ مرتبه اجرا می‌شود، به‌طوری که در هر مرحله نمونه‌ای به اندازه‌ی n و همبستگی ρ از جامعه گرفته می‌شود و تغییر در نسبت‌ها و اریبی آن محاسبه می‌شود. فرض کنید که $\bar{p}_{hd}$ تغییر نسبت در اجرای h ام باشد (تغییر در

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

نسبت بین دو دوره‌ی آمارگیری است) و $h = ۱, ۲, \dots, ۱۰۰۰$. در این صورت براورد اریبی هر یک از این تغییر در نسبت‌های افراد سیگاری در هر اجرا از رابطه‌ی:

$$Bias(\bar{p}_{hd}) = \bar{p}_{hd} - P_d \quad (۵)$$

به‌دست می‌آید. در این رابطه P_d تغییر در نسبت جامعه (اختلاف در متغیر پاسخ جامعه در دو مقطع آمارگیری) که برابر ۰/۰۹ است. اگر $\bar{p}_d$ براورد تغییر در نسبت این ۱۰۰۰ نمونه در دو دوره‌ی آمارگیری باشد، قدر مطلق تفاضل آن با تغییر در نسبت جامعه، برای هر دو روش تعدیل اثر بی‌پاسخی، براورد قدر مطلق اریبی را نتیجه می‌دهد. از تقسیم براورد قدر مطلق اریبی بر P_d براورد قدر مطلق اریبی نسبی به دست می‌آید. این دو معیار به‌ترتیب در زیر تعریف شده‌اند:

$$|Bias(\bar{p}_d)| = \left|\frac{۱}{۱۰۰۰}\sum_{h=۱}^{۱۰۰۰} \bar{p}_{hd} - P_d\right|$$

$$\text{براورد قدر مطلق اریبی نسبی} = \frac{|Bias(\bar{p}_d)|}{P_d} \quad (۶)$$

مقایسه‌ی دو روش تعدیل اثر بی‌پاسخی با معیار براورد کارایی نسبی آن‌ها برای تمام حالت‌های مورد نظر از رابطه‌ی زیر به‌دست می‌آید:

$$\text{براورد کارایی نسبی مجانبی} = \frac{\text{MSE}_{(۱)}(\bar{p}_d)}{\text{MSE}_{(۲)}(\bar{p}_d)} \quad (۷)$$

که در آن $\text{MSE}_{(۱)}(\bar{p}_d)$ براورد میانگین توان دوم خطای روش تعدیل اثر بی‌پاسخی WGEEs و $\text{MSE}_{(۲)}(\bar{p}_d)$ براورد میانگین توان دوم خطای روش تعدیل اثر بی‌پاسخی WEEs است، که از رابطه‌ی زیر به‌دست می‌آیند:

$$\text{MSE}(\bar{p}_d) = \text{var}(\bar{p}_d) + [Bias(\bar{p}_d)]^۲ \quad (۸)$$

که در آن:

$$\text{var}[\bar{p}_d] = \frac{۱}{۱۰۰۰-۱}\sum_{h=۱}^{۱۰۰۰}(\bar{p}_{hd} - \bar{p}_d)^۲ \quad (۹)$$

و

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

$$\bar{p}_d = \frac{۱}{۱۰۰۰}\sum_{h=۱}^{۱۰۰۰} \bar{p}_{hd} \qquad (۱۰)$$

۲-۴- یافته‌های مطالعه‌ی شبیه‌سازی

در جدول‌های ۱ و ۲ به ترتیب مقدار براوردهای به‌دست آمده برای دو معیار تشریح شده، در تمام حالت‌های ممکن آورده شده است. نمودارهای به‌دست آمده از محاسبه‌ی معیار اول نشان می‌دهد که، براورد قدر مطلق اریبی نسبی روش تعدیل اثر بی‌پاسخی WGEEs همواره از روش تعدیل اثر بی‌پاسخی WEEs کم‌تر است. از طرفی در اندازه‌های نمونه‌ای بزرگ با افزایش نرخ بی‌پاسخی تفاوت براورد قدر مطلق اریبی به‌دست آمده بین این دو روش تعدیل اثر بی‌پاسخی، در این اندازه نمونه‌های بزرگ بیش‌تر می‌شود.

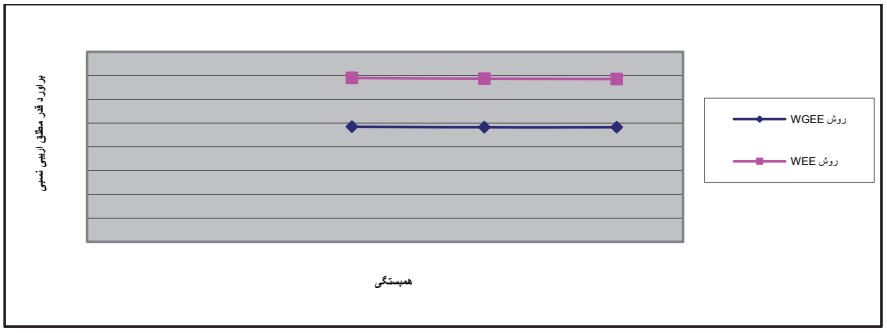
جدول ۱- براورد قدر مطلق اریبی نسبی برای دو روش تعدیل اثر بی‌پاسخی با گم‌شدگی تصادفی

همبستگی	اندازه‌ی نمونه	نرخ بی‌پاسخی	روش WGEEs	روش WEEs
۰/۴	۵۰	۰/۱	۱/۰۰۰۴۷۶	۱/۰۰۴۰۴۶
		۰/۱۵	۰/۹۷۵۴۸۸۶	۱/۲۴۲۳۰۶
	۱۰۰	۰/۱	۰/۹۷۰۲۵۳۲	۱/۳۶۲۹۱۱
		۰/۱۵	۰/۹۵۷۰۴۵۶	۱/۵۵۷۹۳۲
	۲۰۰	۰/۱	۰/۹۶۹۸۹۸۹	۱/۳۸۱۰۵۷
		۰/۱۵	۰/۹۵۱۸۶۹۶	۱/۵۶۱۷۳۹
۰/۶	۵۰	۰/۱	۱/۰۰۱۴۲۸	۱/۰۱۳۵۶۵
		۰/۱۵	۰/۹۶۸۱۱۱۴	۱/۴۳۴۰۶۶
	۱۰۰	۰/۱	۰/۹۶۴۳۰۳۸	۱/۳۷۹۰۹۴
		۰/۱۵	۰/۹۶۵۱۳۶۷	۱/۵۶۴۸۳۳
	۲۰۰	۰/۱	۰/۹۶۵۴۹۳۷	۱/۳۷۴۳۹۴۴
		۰/۱۵	۰/۹۴۶۵۱۵۲	۱/۵۵۸۶۴۶
۰/۸	۵۰	۰/۱	۱/۰۰۰۴۷۶	۱/۰۰۳۰۹۴
		۰/۱۵	۰/۹۸۱۴۳۸	۱/۲۶۲۰۱۰
	۱۰۰	۰/۱	۰/۹۷۸۲۲۵۳	۱/۳۶۱۶۰۳
		۰/۱۵	۰/۹۵۶۵۶۹۶	۱/۵۶۶۱۴۲
	۲۰۰	۰/۱	۰/۹۶۵۷۳۱۶	۱/۳۷۱۳
		۰/۱۵	۰/۹۴۹۳۱۱۴	۱/۵۷۴۷۰۹

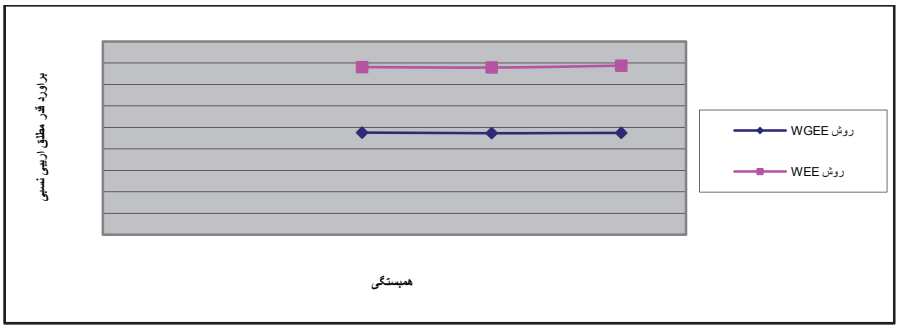
[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

جدول ۲- براورد کارایی نسبی مجانبی برای دو روش تعدیل اثر بی‌پاسخی

همبستگی	کارایی نسبی مجانبی روش تعدیل اثر بی‌پاسخی WGEEs نسبت به روش WEEs					
	n =۵۰		n =۱۰۰		n =۲۰۰	
	Nr=۰/۱۵	Nr=۰/۱	Nr=۰/۱۵	Nr=۰/۱	Nr=۰/۱۵	Nr=۰/۱
۰/۴	۰/۵۸۶۷۵۷۴	۰/۹۹۰۳۰۴	۰/۳۸۳۷۶۱	۰/۵۰۸۸۲۳۳۷	۰/۳۷۷۶۸۷۹	۰/۴۹۷۱۶۶
۰/۶	۰/۴۴۶۲۲۴۳	۰/۹۶۸۳۳۹۸	۰/۳۹۳۱۳۹۷	۰/۴۹۰۲۴۱	۰/۳۷۴۹۵۴۵	۰/۴۹۷۱۸۰
۰/۸	۰/۵۷۵۴۱۶۸	۰/۹۹۳۷۳۵۸	۰/۳۸۲۰۲۲۶	۰/۵۱۸۹۶۱۷	۰/۳۷۰۱۵۶۶	۰/۴۹۷۴۴۷۸



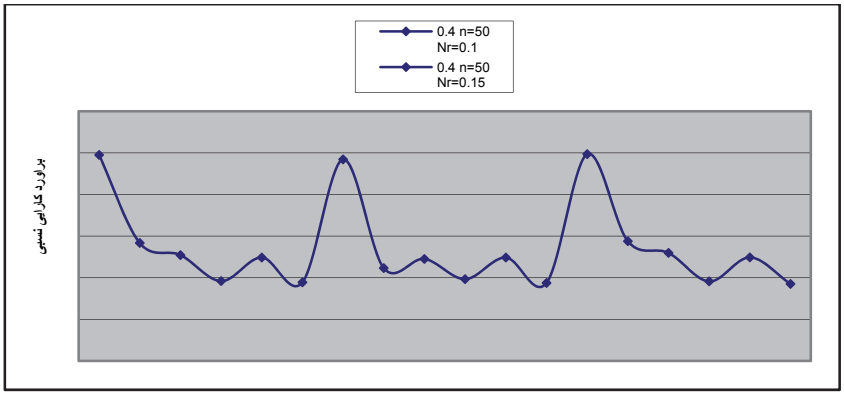
شکل ۱- براورد قدر مطلق اریبی نسبی برای نرخ بی‌پاسخی ۰/۱ و اندازه‌ی نمونه‌ای ۲۰۰



شکل ۲- براورد قدر مطلق اریبی نسبی برای نرخ بی‌پاسخی ۰/۱۵ و اندازه‌ی نمونه‌ای ۲۰۰

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

شکل‌های ۱ و ۲ براورد قدر مطلق اریبی ناشی از دو روش تعدیل اثر بی‌پاسخی را با هم مقایسه می‌کند. همچنین از شکل ۳ مشاهده می‌شود که با این‌که در اندازه‌ی نمونه‌ای کوچک و نرخ بی‌پاسخی کم، کارایی دو روش تعدیل اثر بی‌پاسخی WGEEs و WEEs نزدیک به‌هم هستند ولی زمانی که اندازه‌ی نمونه‌ای افزایش می‌یابد کاراتر بودن روش تعدیل اثر بی‌پاسخی WGEEs نسبت به روش تعدیل اثر بی‌پاسخی WEEs آشکارتر می‌شود.



شکل ۳- براورد کارایی مجانبی روش WGEEs نسبت به روش WEEs

۳-۴- نتیجه‌گیری

با توجه به این‌که همبستگی متغیر سیگاری بودن شخص بین دوره‌ی شانزدهم و هفدهم در داده‌های واقعی بالا است، لذا براساس معیارهای مقایسه‌ی محاسبه‌شده و نتیجه‌های موجود، به نظر می‌رسد که روش تعدیل اثر بی‌پاسخی WGEEs برای تعدیل اثر بی‌پاسخی داده‌های گم‌شده‌ی متغیر پاسخ مورد نظر در آمارگیری پانلی خانواری انگلیس، مفید باشد.

مرجع‌ها

[1] Deville, J.C., Sarndad, C.E. and Sautory, O. (1993). Generalized raking procedures in survey sampling, *Journal of American Statistical Association*, **88**, 1013–1020.

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]

[2] Fitzmaurice, G.M., Molenberg, G. and Lipsitz, L.R. (1995). Regression models for longitudinal binary responses with informative dropouts, *Journal of the Royal statistical society*, **57**, 691-704.

[3] Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2009). *Survey Methodology*, Second Edition, John Wiley & Sons, New York.

[4] Kalton, G. and Flores-Cervantes, I. (2003), Weighting methods, *Journal of Official Statistics*, **19**, 81-97.

[5] Liang, K. and Zeger, S.L. (1986). Longitudinal analysis using generalized linear models, *Biometrika*, **73**, 13-22.

[6] Robins, J., Rotnitzky, A. and Zhao, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data, *Journal of the American Statistical Association*, **95**, 106-121.

[7] Rubin, D.B. (1976). Inference and missing data, *Biometrika*, **63**, 581-592.

[8] Zeger, L.P. and Liang, K.Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics*, **42**, 121-130.

حمیدرضا نواب‌پور
دکترای آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشگاه علامه طباطبایی، دانشکده‌ی اقتصاد، گروه آمار.
رایانشانی: h.navvabpour@srtc.ac.ir

ناصر فروتن
کارشناس ارشد آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشگاه علامه طباطبایی، دانشکده‌ی اقتصاد، گروه آمار.
رایانشانی: naser_frotan@yahoo.com

[Downloaded from ijoss.srtc.ac.ir on 2026-11-23]