

مقایسه‌ی روش‌های $WEEs$ و $WGEEs$ 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 برآورد می‌شوند. سپس عمل‌کرد این روش نسبت به روش دیگر تعدیل اریبی که تحت عنوان معادله‌های برآورد 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 برآورد تعمیم‌یافته‌ی موزون روبینز و همکارانش [۶] در همه‌ی حالت‌ها از روش معادله‌های برآورد موزون عمل‌کرد بهتری دارد.

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

* نویسنده‌ی عهده‌دار مکاتبات

دریافت: ۱۳۸۹/۱۱/۳، پذیرش: ۱۳۹۱/۸/۲۳

۱- مقدمه

تقریباً در تمامی آمارگیری‌ها با مسئله‌ی بی‌پاسخی واحد و گم‌شدگی داده‌های مربوط مواجه هستیم. بی‌پاسخی در حقیقت به شکست در گردآوری 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 گردآوری 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 به دلیل‌های 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 درآمد بالا معمولاً از شرکت در آمارگیری‌هایی که به درآمد 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 برآوردها شده و دومی اریبی در آماره‌های آمارگیری را سبب می‌شود. اگر واحدهای بی‌پاسخ یک زیرمجموعه‌ی تصادفی از نمونه نباشند، یا به عبارتی ویژگی‌های آن‌ها به‌طور سیستماتیک با ویژگی‌های پاسخ‌گویان متفاوت باشد، نمونه‌ی پاسخ‌گویان برآوردهای اریب برای پارامترهای جمعیت مورد نظر به دست می‌دهد.

دو روش برای تعدیل اثرهای بی‌پاسخی وجود دارند، یکی در مرحله‌ی طراحی و دیگری در مرحله‌ی اجرا. در زمان اجرا سعی بر مینیم کردن اثرهای بی‌پاسخی در مرحله‌ی گردآوری 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 جانمایی صورت می‌گیرد [۳]. بنا بر این، منظور از تعدیل بی‌پاسخی استفاده از روش‌هایی در مرحله‌ی تحلیل، برای مینیم کردن اثرهای بی‌پاسخی بر آماره‌های آمارگیری است. در این مقاله از روش دوم برای تعدیل اریبی بی‌پاسخی استفاده می‌شود.

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. برآوردهای 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 به‌قرار 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 برآورد تعمیم‌یافته (Generalized Estimating Equations) برای اولین بار توسط لیانگ و زیگر [۵] و زیگر و لیانگ [۸] به‌عنوان روشی برای تحلیل داده‌های طولی ارائه شد. روبینز و همکارانش [۶] تعدیل اریبی ناشی از بی‌پاسخی واحدها در این نوع مطالعه را با روش وزن‌دهی معادله‌های برآورد 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 برآورد تعمیم‌یافته‌ی لیانگ و زیگر [۵] را ارائه کردند که در آن وزن برای هر واحد نمونه‌ای تنها به‌صورت یک عدد در نظر گرفته می‌شود.

۲- معادله‌های برآورد تعمیم‌یافته (GEEs)

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

روش معادله‌های برآورد 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 بردار ضریب‌های رگرسیونی و بردار متغیرهای کمکی برای فرد i ام در دوره‌ی t می‌باشند. در این صورت معادله‌های برآورد تعمیم‌یافته عبارت‌اند از:

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

که در آن $D_i = \partial \mu_i / \partial \beta$ ماتریس مشتق‌های جزئی و $V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2} \Phi$ نمایانگر ماتریس پراکندگی کاری Y_i است، که در آن A_i ماتریس قطری با عنصرهای قطری $v(\mu_{it})$ 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 به صورت عددی و در یک فرایند تکراری برآورد می‌شود. پس از

محاسبه کردن $\hat{R}_i(\alpha)$ ، ماتریس پراکندگی کاری به صورت $\hat{V}_i = A_i^{1/2} \hat{R}_i(\alpha) A_i^{1/2} \hat{\Phi}$ برآورد 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 اندازه‌گیری شده، p تعداد پارامترهای مدل و $e_{ij} = \left(\frac{y_{ij} - \mu_{ij}}{\sqrt{a_{ij}}} \right)$ مانده‌ی پی‌یرسونی 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 برآوردهای سازگار پارامتر β می‌شود، زیرا در این حالت معادله‌های بالا تنها به متغیرهای

کمکی وابسته هستند. ولی در صورتی که مکانیزم گم‌شدگی از نوع تصادفی باشد معادله‌ی بالا برای برآورد 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 برآوردهای سازگار، معادله‌های برآورد 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، برآورد سازگار پارامتر β از حل معادله‌ی زیر به دست می‌آید:

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

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

۳- روش معادله‌های برآورد موزون (WEEs)

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

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

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

متغیرهای کمکی مدل بندی می شود. اگر داده ها به طور کاملاً تصادفی گم شده باشند، آنگاه این معادله ها برآورد سازگاری از β فراهم می کنند ولی اگر داده ها به طور تصادفی گم شده باشند در این صورت برای تعدیل بی پاسخی در معادله های بالا از معادله های برآورد موزون (Weighted Estimating Equations) استفاده می شود. این معادله ها عبارت اند از

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

که در آن w_i وزن واحد i ام است. در این معادله ها برای تعدیل بی پاسخی می توان از روش های وزن دهی از جمله پساطبقه بندی [۴]، وزن دهی رگرسیونی تعمیم یافته، چنگک زنی و کالبدن [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 این که وزن دهی رگرسیونی تعمیم یافته از متغیرهای کمکی موجود در مطالعه استفاده می کنند، از این شیوهی وزن دهی در معادله های برآورد موزون استفاده کرده ایم.

۴- مطالعه ی شبیه سازی

داده های استفاده شده در این مقاله از آمارگیری پانلی خانواری انگلیس انتخاب شده است. آمارگیری پانلی خانواری انگلیس از سال ۱۹۹۱ توسط مرکز مطالعه های طولی انگلیس همراه با مؤسسه ی تحقیق های اقتصادی و اجتماعی دانشگاه اسکس هر سال اجرا می شود. موضوع اصلی آمارگیری در سه دوره ی اول سازمان دهی خانوارها، بازار کار، درآمد و دارایی، مسکن، معیارهای سلامتی و اجتماعی اقتصادی بوده اند. بنا بر این امکان تحقیق در سطح گسترده ای از موضوع ها برای پژوهش گر ها وجود دارد. هدف اصلی آمارگیری، یافتن تغییرهای اقتصادی و اجتماعی در سطح فرد و خانوار در انگلیس است. انتخاب مدل، پیش گویی و شرح علت تغییرها در ارتباط با رده ای از متغیرهای اقتصادی اجتماعی از هدف های این آمارگیری است.

با مطالعه ی پرسش نامه های استفاده شده در این آمارگیری، پرسش نامه ای که مربوط به افراد ۱۱ تا ۱۶ سال است، مورد استفاده قرار گرفت. این پرسش نامه، از دور چهارم وارد آمارگیری شده است و در تمامی دوره های بعدی مورد استفاده قرار گرفته است. در این مقاله از داده های دوره ی شانزدهم و هفدهم آمارگیری با این پرسش نامه استفاده شده است.

متغیر پاسخ "سیگاری بودن فرد" در نظر گرفته شده است به طوری 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 به عنوان 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 این که 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 به صورت تصادفی ساده و بدون جای‌گذاری از این جامعه انتخاب شده و نرخ‌های کاهش پاسخ‌گو به صورت تصادفی با نرخ‌های ۱/۰ و ۱۵/۰ در نظر گرفته شدند. همبستگی بین دو دوره نیز در سه مقدار ۴/۰، ۶/۰ و ۸/۰ اختیار شدند. لازم به ذکر است که همبستگی بین داده‌های جامعه در دو دوره ۶۴۲/۰ و کاهش پاسخ‌گو در دور دوم آمارگیری روی می‌دهد.

برای برآورد توزیع نمونه‌گیری آماری تغییر نسبت در دو دوره‌ی آمارگیری، به اندازه‌ی ۱۰۰۰ مرتبه نمونه‌گیری مستقل از هر دوره‌ی آمارگیری انجام شد و برآورد تغییر نسبت پاسخ و واریانس آن به دست 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 برآورد قدر مطلق اربیبی نسبی و برآورد کارایی نسبی مجانبی استفاده کرده‌ایم. در هر دو روش تعدیل اثر بی‌پاسخی مورد نظر، برای محاسبه‌ی برآورد قدر مطلق اربیبی، ابتدا عمل نمونه‌گیری برای حالت‌های مورد نظر با استفاده از نرم‌افزار R به تعداد ۱۰۰۰ مرتبه اجرا می‌شود، به طوری که در هر مرحله نمونه‌ای به اندازه‌ی n و همبستگی ρ از جامعه گرفته می‌شود و تغییر در نسبت‌ها و اربیبی آن محاسبه می‌شود. فرض کنید که \bar{p}_{ha} تغییر نسبت در اجرای h ام باشد (تغییر در

نسبت بین دو دوره‌ی آمارگیری است) و $h = 1, 2, \dots, 1000$. در این صورت برآورد اریبی هر یک از این تغییر در نسبت‌های افراد سیگاری در هر اجرا از رابطه‌ی:

$$(5) \quad Bias(\bar{p}_{hd}) = \bar{p}_{hd} - P_d$$

به دست می‌آید. در این رابطه P_d تغییر در نسبت جامعه (اختلاف در متغیر پاسخ جامعه در دو مقطع آمارگیری) که برابر $0/09$ است. اگر \bar{p}_d برآورد تغییر در نسبت این 1000 نمونه در دو دوره‌ی آمارگیری باشد، قدر مطلق تفاضل آن با تغییر در نسبت جامعه، برای هر دو روش تعدیل اثر بی‌پاسخی، برآورد قدر مطلق اریبی را نتیجه می‌دهد. از تقسیم برآورد قدر مطلق اریبی بر P_d برآورد قدر مطلق اریبی نسبی به دست می‌آید. این دو معیار به ترتیب در زیر تعریف شده‌اند:

$$|Bias(\bar{p}_d)| = \left| \frac{1}{1000} \sum_{h=1}^{1000} \bar{p}_{hd} - P_d \right|$$

$$(6) \quad \text{برآورد قدر مطلق اریبی نسبی} = \frac{|Bias(\bar{p}_d)|}{P_d}$$

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

$$(7) \quad \text{برآورد کارایی نسبی مجانبی} = \frac{MSE_{(1)}(\bar{p}_d)}{MSE_{(2)}(\bar{p}_d)}$$

که در آن $MSE_{(1)}(\bar{p}_d)$ برآورد میانگین توان دوم خطای روش تعدیل اثر بی‌پاسخی WGEES و $MSE_{(2)}(\bar{p}_d)$ برآورد میانگین توان دوم خطای روش تعدیل اثر بی‌پاسخی WEES است، که از رابطه‌ی زیر به دست می‌آیند:

$$(8) \quad MSE(\bar{p}_d) = \text{var}(\bar{p}_d) + [Bias(\bar{p}_d)]^2$$

که در آن:

$$(9) \quad \text{var}[\bar{p}_d] = \frac{1}{1000-1} \sum_{h=1}^{1000} (\bar{p}_{hd} - \bar{p}_d)^2$$

و

$$(۱۰) \quad \bar{p}_d = \frac{1}{1000} \sum_{h=1}^{1000} \bar{p}_{hd}$$

۲-۴- یافته‌های مطالعه‌ی شبیه‌سازی

در جدول‌های ۱ و ۲ به ترتیب مقدار برآوردهای به‌دست آمده برای دو معیار تشریح شده، در تمام حالت‌های ممکن آورده شده است. نمودارهای به‌دست آمده از محاسبه‌ی معیار اول نشان می‌دهد که، برآورد 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 برآورد قدر مطلق اریبی به‌دست آمده بین این دو روش تعدیل اثر بی‌پاسخی، در این اندازه نمونه‌های بزرگ بیش‌تر می‌شود.

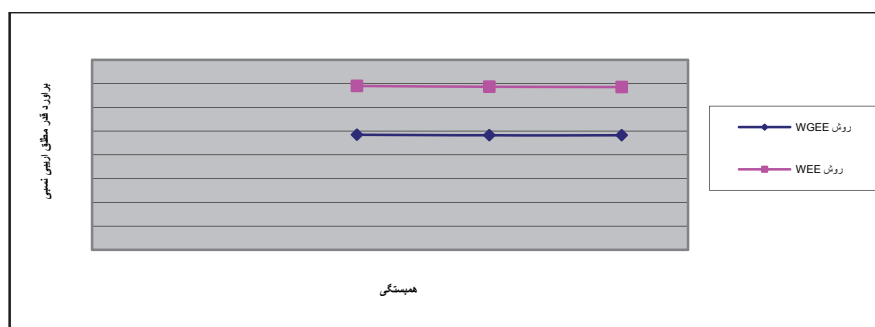
جدول ۱- برآورد قدر مطلق اریبی نسبی برای دو روش تعدیل اثر بی‌پاسخی با گم‌شدگی تصادفی

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

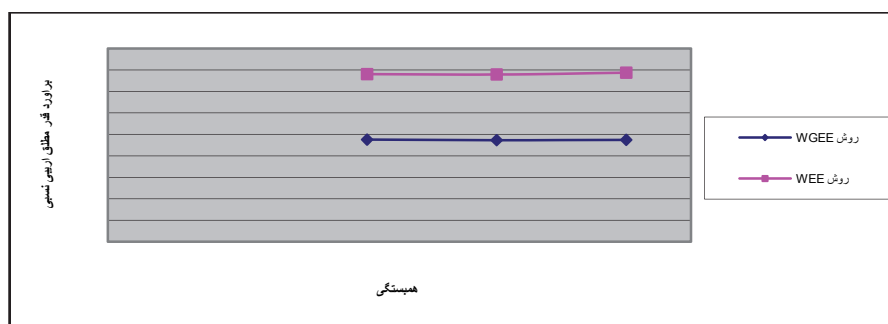
جدول ۲- برآورد کارایی نسبی مجانبی برای دو روش تعدیل اثر بی‌پاسخی

کارایی نسبی مجانبی روش تعدیل اثر بی‌پاسخی WGEs نسبت به روش WEEs					
$n = 200$		$n = 100$		$n = 50$	
$Nr=0/1$	$Nr=0/15$	$Nr=0/1$	$Nr=0/15$	$Nr=0/1$	$Nr=0/15$
۰/۴۹۷۱۶۶	۰/۳۷۷۶۸۷۹	۰/۵۰۸۸۲۳۳۷	۰/۳۸۳۷۶۱	۰/۹۹۰۳۰۴	۰/۵۸۶۷۵۷۴
۰/۴۹۷۱۸۰	۰/۳۷۴۹۵۴۵	۰/۴۹۰۲۴۱	۰/۳۹۳۱۳۹۷	۰/۹۶۸۳۳۹۸	۰/۴۴۶۲۲۴۳
۰/۴۹۷۴۴۷۸	۰/۳۷۰۱۵۶۶	۰/۵۱۸۹۶۱۷	۰/۳۸۲۰۲۲۶	۰/۹۹۳۷۳۵۸	۰/۵۷۵۴۱۶۸

همبستگی

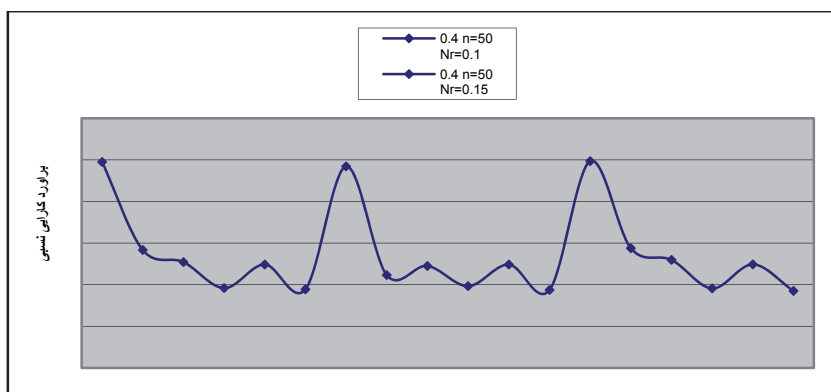


شکل ۱- برآورد قدر مطلق اریبی نسبی برای نرخ بی‌پاسخی ۱٪ و اندازه‌ی نمونه‌ای ۲۰۰



شکل ۲- برآورد قدر مطلق اریبی نسبی برای نرخ بی‌پاسخی ۱۵٪ و اندازه‌ی نمونه‌ای ۲۰۰

شکل‌های ۱ و ۲ برآورد 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 به هم 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- برآورد 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.

- [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، دانشکده اقتصاد، گروه آمار.

رایانشانی: h.navvabpour@src.ac.ir

ناصر فروتن

کارشناس ارشد آمار

تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشگاه علامه طباطبایی، دانشکده اقتصاد، گروه آمار.

رایانشانی: naser_frotan@yahoo.com