

برآورد میانگین جامعه در حضور بی‌پاسخی با استفاده از روش نمونه‌گیری دوفازی

مریم تقریبی*، نادر نعمت‌الهی و حمیدرضا نواب‌پور

دانشگاه علامه طباطبائی

چکیده: در این مقاله مسئله‌ی برآورد میانگین جامعه با استفاده از اطلاعات یک متغیر کمکی در حضور بی‌پاسخی مورد بررسی قرار گرفته است. برای این منظور برآوردهای نسبتی، ضربی و رگرسیونی برای برآورد میانگین جامعه با روش نمونه‌گیری دوفازی پیشنهاد شده و ویژگی‌های آن‌ها مورد مطالعه قرار گرفته است. این برآوردها اریب هستند و مقایسه‌ی میانگین توان دوم خطای آن‌ها به صورت نظری مشکل است. بنا بر این با فرض ناچیز بودن اریبی برای اندازه‌ی نمونه‌ی بزرگ و مقایسه‌ی واریانس‌ها، بهترین برآوردها میانگین جامعه تعیین می‌شود. همچنین با انجام مطالعه‌ی شبیه‌سازی، اثر همبستگی‌ها، روی میانگین توان دوم خطای شبیه‌سازی، برای نرخ‌های بی‌پاسخی و اندازه‌های نمونه‌ی مختلف بررسی می‌شود. یافته‌های مطالعه‌ی شبیه‌سازی نشان می‌دهند هرچه همبستگی بین دو متغیر مورد مطالعه و کمکی بالا، نرخ بی‌پاسخی پایین و اندازه‌ی نمونه بزرگ باشد، میانگین توان دوم خطای شبیه‌سازی کم‌تر است.

واژگان کلیدی: برآورد رگرسیونی؛ برآوردگر ضربی؛ برآوردگر نسبتی؛ برآوردگر هم‌گذاشتی؛ متغیر کمکی؛ متغیر مورد مطالعه؛ نمونه‌گیری دوفازی.

۱- مقدمه

در آمارگیری‌های نمونه‌ای اطلاعات برخی واحدها حتی پس از مراجعه‌ی مجدد به آن‌ها به دست نمی‌آیند. در این حالت آماره‌های آمارگیری احتمالاً با اریبی مواجه هستند. این خطا که به خطای بی‌پاسخی معروف است، به علت ناکامی در دستیابی به اطلاعات مفید در

* نویسنده‌ی عهده‌دار مکاتبات

دریافت: ۱۳۹۱/۱۰/۱۰، پذیرش: ۱۳۹۳/۵/۱.

مورد یک یا چند متغیر مورد بررسی برای یک یا چند عضو واجد شرایط در آمارگیری رخ می‌دهد. حتی اگر از طرح‌های آمارگیری دقیقی برای مینیم کردن دیگر منبع‌های بروز خطا استفاده شود، وجود بی‌پاسخی ممکن است تمام نتیجه‌های آمارگیری را نادقیق کند. همچنین برآوردهای به دست آمده از چنین اطلاعات ناکاملی، به‌ویژه هنگامی که ویژگی‌های پاسخ‌گویان از بی‌پاسخ‌ها متفاوت باشند، اریب هستند. در بسیاری از مطالعه‌ها برای آن‌که بتوان با به‌کارگیری روش‌های تعدیل، نرخ بی‌پاسخی را تا حد امکان کاهش داد پاسخ‌گویان و بی‌پاسخ‌ها را در دو گروه متفاوت جای می‌دهند، یک گروه شامل افراد پاسخ‌گو و گروه دیگر شامل افراد بی‌پاسخ است. بر این اساس نرخ بی‌پاسخی که نسبت واحدهای بی‌پاسخ به کل واحدهای واجد شرایط در نمونه است، به‌راحتی برآورد می‌شود. بالا بودن نرخ بی‌پاسخی موجب کوچک شدن اندازه‌ی نمونه‌ای، کاهش دقت و اریبی برآوردها می‌شود. اگر با رعایت تمام موردها در طرح‌های آمارگیری، به‌دلیل‌هایی بی‌پاسخی رخ دهد از تعدیل‌های پس‌آمارگیری که عبارت‌اند از، تعدیل موزون و جانپی برای حداقل کردن اثر خطای بی‌پاسخی برآوردها استفاده می‌شود.

هنسن و هورویتس [۲]، روش مراجعه‌ی مجدد را برای تعدیل اریبی بی‌پاسخی پیشنهاد کردند. این روش بر اساس انتخاب زیرنمونه‌ای از بی‌پاسخ‌ها و تلاش برای دریافت اطلاعات از آن‌ها است. برآوردگر پیش‌نهادی آن‌ها ترکیبی از اطلاعات پاسخ‌گویان و زیرنمونه‌ی انتخابی از بی‌پاسخ‌ها است. همچنین بهتر است برای افزایش دقت برآوردها از اطلاعات کمکی که همان داده‌های ثبتي یا نتایج سرشماری‌های قبلی هستند، استفاده شود. اگرچه ممکن است برآوردهای حاصل در مقایسه با برآوردهای متعارف اریب باشند (که با افزایش اندازه‌ی نمونه اریبی آن‌ها ناچیز خواهد شد)، اما نسبت به برآوردهای متعارف دارای واریانسی کم‌تر هستند. این اطلاعات کمکی هم در طرح نمونه‌گیری و هم در برآورد بهتر پارامترهای مورد نظر مفید هستند. هرچه همبستگی متغیر کمکی با متغیر مورد مطالعه بیشتر باشد، شاهد افزایش دقت و کاهش اریبی برآوردهای حاصل خواهیم بود.

روش نمونه‌گیری دوفازی برای مواردی از جمله تعدیل بی‌پاسخی، نمونه‌گیری از جامعه‌های کم‌یاب، بهبود چارچوب نمونه‌گیری و نیز زمانی که اندازه‌گیری متغیر مورد مطالعه‌ی Y نسبتاً گران بوده و اندازه‌گیری متغیر کمکی همبسته با Y به‌سادگی صورت می‌گیرد، استفاده می‌شود. در روش نمونه‌گیری دوفازی از اطلاعات کمکی به این صورت

استفاده می‌شود که، در فاز اول اطلاعاتی برای متغیر کمکی و در فاز دوم اطلاعاتی برای متغیر مورد مطالعه گردآوری می‌شود. در فاز دوم نمونه‌گیری اغلب بی‌پاسخی وجود دارد. برآورد میانگین متغیر مورد مطالعه، زمانی که در فاز اول میانگین جامعه متغیر کمکی معلوم باشد، توسط رابسون [۱۰] و کوکران [۱] مطالعه شده است و زمانی که میانگین جامعه متغیر کمکی در فاز اول نامعلوم باشد، توسط اکافور و لی [۷]، اکافور [۵] و [۶] و تاباسام و خان [۱۳] و [۱۴] مورد بررسی قرار گرفته است. همچنین رائو [۸] و [۹] و خار و سریواستاوا [۳] و [۴] برآورد میانگین جامعه‌ای متغیر مورد مطالعه را در دو حالت گفته‌شده بررسی کرده‌اند.

علاوه بر آن سینگ و کومار [۱۱] و [۱۲] تعدادی برآوردگر نسبتی، ضربی و رگرسیونی را در نمونه‌گیری دوفازی معرفی کردند و نشان دادند که این برآوردگرها اریب هستند. برای اندازه‌ی نمونه‌ی بزرگ، آن‌ها دو برآوردگر بر اساس تعمیم برآوردگر نسبتی و رگرسیونی پیش‌نهاد کردند که دارای واریانس کم‌تری نسبت به مابقی برآوردگرها هستند. سینک و کومار [۱۱] و [۱۲] در مطالعه‌ی خود از اریبی برآوردگرها چشم‌پوشی کرده و برآوردگرها را به‌وسیله‌ی واریانس آن‌ها مقایسه کرده‌اند. در این مقاله بر اساس یک مطالعه‌ی شبیه‌سازی با در نظر گرفتن میانگین توان دوم خطای برآوردگرها نشان می‌دهیم، این دو برآوردگر کاراتر از مابقی برآوردگرها نبوده و برآوردگرهای کاراتر از آن‌ها را معرفی می‌کنیم.

در بخش بعد، چگونگی استفاده از اطلاعات کمکی به‌منظور افزایش دقت برآوردگرهای حاصل از روش مراجعه‌ی مجدد، با عنوان نمونه‌گیری دوفازی مورد بررسی قرار گرفته است. در بخش سوم تعدادی برآوردگر نسبتی، ضربی و رگرسیونی برای برآورد میانگین جامعه‌ای با روش نمونه‌گیری دوفازی معرفی می‌شود و در بخش چهارم ویژگی‌های آن‌ها مورد مطالعه قرار می‌گیرد. در بخش پنجم، با فرض ناچیز بودن اریبی، واریانس برآوردگرها با هم مقایسه و بهترین برآوردگر میانگین جامعه‌ای تعیین می‌شود. در بخش ششم، با انجام یک مطالعه‌ی شبیه‌سازی، برآورد میانگین توان دوم خطای برآوردگرها تحت شرایط مختلف همبستگی، نرخ بی‌پاسخی و اندازه‌ی نمونه با هم مقایسه می‌شود و در نهایت بهترین شرایط برای برآورد بهتر میانگین جامعه گزارش می‌شود.

۲- نمونه‌گیری دوفازی

اساس روش نمونه‌گیری دوفازی بر انتخاب نمونه‌ای کوچک‌تر از درون نمونه‌ای بزرگ‌تر استوار است. انتخاب نمونه‌ای بزرگ از جامعه را فاز اول و انتخاب نمونه‌ای کوچک‌تر از درون نمونه‌ی بزرگ را فاز دوم می‌نامند. با استفاده از روش نمونه‌گیری دوفازی و به‌کارگیری اطلاعات متغیرهای کمکی که با متغیر مورد مطالعه همبستگی قوی دارند می‌توان دقت برآوردهای به دست آمده از روش مراجعه‌ی مجدد را بهبود بخشید. مهم‌ترین برآوردهایی که از اطلاعات یک متغیر کمکی استفاده می‌کنند عبارت‌اند از برآوردهای نسبتی، ضربی، رگرسیونی و برآوردهای هم‌گذاشتی. برآوردهای حاصل از روش دوفازی نیز حالتی از این برآوردها هستند.

جامعه‌ی $U = \{Y_1, Y_2, \dots, Y_N\}$ را در نظر بگیرید که در آن Y_i ها می‌توانند گسسته یا پیوسته باشند. فرض می‌کنیم با تعدادی از مقادیرهای Y_i بی‌پاسخ در طول آمارگیری مواجه خواهیم شد. به شرط وجود بی‌پاسخی در طول آمارگیری، جامعه به دو زیرجامعه‌ی U_1 و U_2 تقسیم می‌شود. که در آن U_1 نمایان‌گر افرادی است که در مراجعه‌ی اول پاسخ‌گو هستند و U_2 نمایان‌گر افرادی است که در مراجعه‌ی اول پاسخ‌گو نیستند اما با مراجعه‌ی مجدد پاسخ‌گو خواهند بود. در واقع در روش مراجعه‌ی مجدد فرض می‌شود که بی‌پاسخ‌ها با مراجعه‌ی مجدد، پاسخ‌گو خواهند بود. مسئله‌ی برآورد میانگین جامعه‌ای \bar{Y}_U از متغیر مورد مطالعه‌ی Y با استفاده از اطلاعات متغیر کمکی X ، زمانی که میانگین جامعه‌ای \bar{X}_U معلوم باشد، توسط آمارشناسان مختلف مورد بحث و مطالعه قرار گرفته است. در بسیاری از موقعیت‌ها \bar{X}_U نامعلوم است، بنا بر این در غیاب اطلاعات کمکی در فاز اول، نمونه‌ای نسبتاً بزرگ به اندازه‌ی n' به روش نمونه‌گیری تصادفی ساده بدون جای‌گذاری (*SRSWOR*) از جامعه انتخاب و \bar{X}_U با میانگین نمونه $(\bar{x}_{n'})$ برآورد می‌شود. سپس در فاز دوم، نمونه‌ای کوچک‌تر به اندازه‌ی n به همان شیوه‌ی نمونه‌گیری، از n' واحد انتخاب و متغیر مورد مطالعه‌ی Y روی واحدهای آن اندازه‌گیری می‌شود. فرض می‌کنیم که در فاز اول، اطلاعات برای متغیر کمکی X به‌طور کامل موجود است اما در فاز دوم اطلاعات کاملی برای متغیر مورد مطالعه در دست نیست.

به‌علت وجود بی‌پاسخی در فاز دوم برای متغیر مورد مطالعه از روش مراجعه‌ی مجدد برای تعدیل بی‌پاسخی استفاده می‌شود. فرض کنید از n واحد نمونه‌گیری شده، n_1 واحد

پاسخ‌گو و n_2 واحد بی‌پاسخ باشند. آمارشناسان با استفاده از روش مراجعه‌ی مجدد بر اساس نرخ پاسخ‌گویی \sqrt{k} ($\sqrt{k} = n_1/n$) زیرنمونه‌ای به اندازه‌ی n_{2m} ($n_{2m} = n_2/k$) به روش نمونه‌گیری تصادفی ساده بدون جای‌گذاری از n_2 واحد بی‌پاسخ در فاز دوم انتخاب و دوباره به آن‌ها مراجعه کرده و با تمهیداتی متغیرهای مورد نظر را روی آن‌ها اندازه‌گیری می‌کنند. این روش، نمونه‌گیری دوفازی برای تعدیل بی‌پاسخی شناخته می‌شود. در واقع در این روش، نمونه‌گیری در دو فاز انجام می‌گیرد که در فاز اول اطلاعات کامل و در فاز دوم بی‌پاسخی وجود دارد و روش مراجعه‌ی مجدد در فاز دوم برای کاهش نرخ بی‌پاسخی اعمال می‌شود. میانگین متغیرهای کمکی و مورد مطالعه به روش مراجعه‌ی مجدد به صورت زیر برآورد می‌شوند:

$$\bar{y}_n^* = \frac{n_1}{n} \bar{y}_1 + \frac{n_2}{n} \bar{y}_{2m}, \quad \bar{x}_n^* = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_{2m}$$

که در آن، (\bar{x}_1, \bar{y}_1) و $(\bar{x}_{2m}, \bar{y}_{2m})$ به ترتیب میانگین‌های نمونه‌ی n_1 تایی و زیرنمونه‌ی n_{2m} تایی برای متغیرهای (X, Y) هستند. همچنین $\bar{x}_{n'} = \frac{1}{n'} \sum_{i=1}^{n'} x_i$ میانگین متغیر کمکی (بر اساس اطلاعات نمونه‌گیری در فاز اول) است. برای برآورد میانگین جامعه با استفاده از روش نمونه‌گیری دوفازی برآوردگرهای زیادی پیش‌نهاد شده است که در بخش بعد معرفی می‌شوند.

۳- برآوردگرهای پیش‌نهادشده به روش دوفازی

برآوردگرهایی که از اطلاعات یک متغیر کمکی استفاده می‌کنند عبارت‌اند از برآوردگرهای نسبتی، ضربی و رگرسیونی. برآوردگرهای پیش‌نهادشده توسط آمارشناسان در این بخش، حالتی از این برآوردگرها هستند. اگر دسترسی به اطلاعات متغیرهای کمکی به سادگی ممکن نباشد، تحقیق باید در دو فاز انجام شود. در فاز اول اطلاعات برای متغیر کمکی X و در فاز دوم اطلاعات برای متغیر مورد مطالعه‌ی Y جستجو می‌شود. با این پیش‌زمینه، برآوردگرهای نسبتی، ضربی و رگرسیونی به روش نمونه‌گیری دوفازی در حضور بی‌پاسخی به دو دسته به صورت زیر تقسیم شده‌اند.

دسته‌ی اول: این دسته شامل براوردگرهای نسبتی $T_{r\setminus}$ ، ضربی $T_{p\setminus}$ ، رگرسیونی $T_{d\setminus}$ و براوردگرهای هم‌گذاشتی است که عبارت‌اند از $t_d^{(r)}$ ، $t_d^{(p)}$ ، t_e و t_{lrd} و به‌صورت زیر تعریف می‌شوند:

$$(۱) \quad T_{r\setminus} = \frac{\bar{y}_n^*}{\bar{x}_n^*} \bar{x}_{n'}$$

$$(۲) \quad T_{p\setminus} = \frac{\bar{y}_n^*}{\bar{x}_{n'}} \bar{x}_n^*$$

$$(۳) \quad T_{d\setminus} = \bar{y}_n^* + \hat{\beta}_{yx}^* (\bar{x}_{n'} - \bar{x}_n^*)$$

$$(۴) \quad t_d^{(r)} = \bar{y}_n^* \left(\frac{\bar{x}_{n'}}{\bar{x}_n^*} \right) \left(\frac{\bar{x}_{n'}}{\bar{x}_n} \right)$$

$$(۵) \quad t_d^{(p)} = \bar{y}_n^* \left(\frac{\bar{x}_n^*}{\bar{x}_{n'}} \right) \left(\frac{\bar{x}_n}{\bar{x}_{n'}} \right)$$

$$(۶) \quad t_e = \bar{y}_n^* \left(\frac{\bar{x}_n}{\bar{x}_n^*} \right)^{\hat{\beta}_{yx\setminus}/\hat{R}} \left(\frac{\bar{x}_{n'}}{\bar{x}_n} \right)^{\hat{\beta}_{yx}^{**}/\hat{R}}$$

$$(۷) \quad t_{lrd} = \bar{y}_n^* + \hat{\beta}_{yx\setminus} (\bar{x}_n - \bar{x}_n^*) + \hat{\beta}_{yx}^{**} (\bar{x}_{n'} - \bar{x}_n)$$

که در تمام این براوردگرها، $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ میانگین متغیر کمکی X بر اساس n واحد نمونه و $\bar{x}_{n'}$ میانگین متغیر کمکی X بر اساس n' واحد نمونه است. براوردگرهای میانگین جامعه به روش مراجعه‌ی مجدد $(\bar{x}_n^*, \bar{y}_n^*)$ ، در بخش قبل تعریف

شده‌اند. $\hat{\beta}_{yx}^* = \frac{s_{xy}}{s_x^*}$ و $\hat{\beta}_{yx}^{**} = \left(s_{xy}^{**} / s_x^{**\setminus} \right)$ براورد ضریب رگرسیونی $\beta_{yx} = \frac{S_{XY}}{S_X^{\setminus}}$

در جامعه هستند که در آن‌ها s_{xy}^* ، s_{xy}^{**} ، s_x^* و $s_x^{**\setminus}$ به‌صورت زیر تعریف می‌شوند:

$$S_{xy}^* = \frac{1}{n-1} \left(\sum_{u_1} x_i y_i + n_{vm} \sum_{u_{vm}} x_i y_i - n \bar{x}_n^* \bar{y}_n^* \right),$$

$$S_x^{*r} = \frac{1}{n-1} \left(\sum_{u_1} x_i^r + n_{vm} \sum_{u_{vm}} x_i^r - n \bar{x}_n^{*r} \right),$$

$$S_{xy}^{**} = \frac{1}{n-1} \left(\sum_{u_1} x_i y_i + n_{vm} \sum_{u_{vm}} x_i y_i - n \bar{x}_n \bar{y}_n^* \right),$$

$$S_x^{**r} = \frac{1}{n-1} \left(\sum_{u_1} x_i^r + n_{vm} \sum_{u_{vm}} x_i^r - n \bar{x}_n^r \right),$$

$$S_x^r = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_U)^r,$$

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_U)(Y_i - \bar{Y}_U).$$

همچنین

$$\hat{\beta}_{yx^r} = \frac{S_{xy^r}}{S_x^{r*}}, \quad \hat{R} = \frac{\bar{y}_n^*}{\bar{x}_n^*}, \quad \hat{\beta}_{yx} = \frac{S_{xy}^{**}}{S_x^{**r}} = \hat{\beta}_{yx}^{**},$$

$$S_{xy^r} = \frac{1}{(n_{vm}-1)} \sum_{i=1}^{n_{mv}} (x_i - \bar{x}_{vm})(y_i - \bar{y}_{vm}),$$

$$S_x^{r*} = \frac{1}{(n_{vm}-1)} \sum_{i=1}^{n_{mv}} (x_i - \bar{x}_{vm})^r.$$

شایان ذکر است در مطالعاتی که در زمینه‌ی نمونه‌گیری دوفازی توسط آمارشناسان مختلف از جمله سینگ و کومار [۱۱] و [۱۲] انجام شده است، $\hat{\beta}_{yx}^{**}$ به صورت نسبت

s_{xy}^* و $s_x^2 = 1/(n-1) \sum_{i=1}^n (x_i - \bar{x}_n)^2$ معرفی شده است. طبق مطالعاتی که محقق در این زمینه انجام داده است، ادعا می‌شود که اگر از $\hat{\beta}_{yx}^{**} = (s_{xy}^* / s_x^2)$ در برآورد میانگین جامعه‌ای استفاده شود، مقدار به دست آمده عددی بسیار بزرگ، غیر واقعی و متفاوت با میانگین حقیقی جامعه به دست خواهد آمد. در این مقاله برآورد این پارامتر تصحیح شده و به صورت نسبت s_{xy}^{**} و s_x^{**} استفاده می‌شود.

برآوردگرهای T_{r1}, T_{p1}, T_{d1} و T_{r2} اولین بار توسط خار و سریواستاوا [۳] و [۴] پیشنهاد شدند. برآوردگر T_{r1} توسط اکافور و لی [۷] و تاباسام و خان [۱۳] بازنگاری شد. بعدها برآوردگر T_{d1} توسط اکافور و لی [۷] و برآوردگر T_{r2} توسط تاباسام و خان [۱۴] مورد مطالعه مجدد قرار گرفت. همچنین برآوردگر T_{d2} اولین بار توسط خار و سریواستاوا [۴] معرفی شد. برآوردگرهای $t_d^{(p)}, t_d^{(r)}, t_e$ و t_{lrd} نیز توسط سینگ و کومار [۱۱] پیشنهاد شدند که ویژگی‌های آن‌ها را مطالعه کردند. دسته‌ی دوم: این دسته شامل برآوردگرهای نسبتی T_{r2} ، ضربی T_{p2} و رگرسیونی T_{d2} است. این برآوردگرها به صورت زیر تعریف می‌شوند:

$$(۸) \quad T_{r2} = \frac{\bar{y}_n^*}{\bar{x}_n} \bar{x}_{n'}$$

$$(۹) \quad T_{p2} = \frac{\bar{y}_n^*}{\bar{x}_{n'}}$$

$$(۱۰) \quad T_{d2} = \bar{y}_n^* + \hat{\beta}_{yx}^{**} (\bar{x}_{n'} - \bar{x}_n)$$

۴- ویژگی برآوردگرهای پیشنهاد شده

برآوردگرهای (۱) تا (۱۰) ترکیبی از برآوردگرهای نسبتی، ضربی و رگرسیونی هستند. این برآوردگرها به طور کلی اریب هستند. اما اگر اندازه‌ی نمونه به مقدار کافی بزرگ باشد، اریبی این برآوردگرها ناچیز خواهد شد. اریبی تقریبی برآوردگرهای (۱) تا (۱۰) با دومین درجه‌ی تقریب از رابطه‌های زیر به دست می‌آیند:

$$B(T_{r\gamma}) \cong \frac{1}{\bar{X}_U} \left(\frac{1}{n} - \frac{1}{n'} \right) (RS_x^\gamma - S_{xy}) \\ + \frac{W_\gamma(k-1)}{n \bar{X}_U} (RS_{x\gamma}^\gamma - S_{xy\gamma})$$

$$B(T_{p\gamma}) \cong \left(\frac{1}{n} - \frac{1}{n'} \right) \frac{S_{xy}}{\bar{X}_U} + \frac{W_\gamma(k-1)}{n} \frac{S_{xy\gamma}}{\bar{X}_U}$$

$$B(T_{d\gamma}) \cong \circ$$

$$B(t_d^{(r)}) \cong \frac{1}{\bar{X}_U} \left[\left(\frac{1}{n} - \frac{1}{n'} \right) S_x^\gamma (\gamma R - \gamma \beta_{yx}) \right. \\ \left. + \frac{W_\gamma(k-1)}{n} S_{x\gamma}^\gamma (R - \beta_{yx\gamma}) \right]$$

$$B(t_d^{(p)}) \cong \frac{1}{\bar{X}_U} \left[\left(\frac{1}{n} - \frac{1}{n'} \right) S_x^\gamma (R + \gamma \beta_{yx}) \right. \\ \left. + \frac{W_\gamma(k-1)}{n} S_{x\gamma}^\gamma \beta_{yx\gamma} \right]$$

$$B(t_e) \cong \frac{1}{\gamma \bar{X}_U} \left[\left(\frac{1}{n} - \frac{1}{n'} \right) S_x^\gamma \frac{\beta_{yx}}{R} (R - \beta_{yx}) \right. \\ \left. + \frac{W_\gamma(k-1)}{n} S_{x\gamma}^\gamma \frac{\beta_{yx\gamma}}{R} (R - \beta_{yx\gamma}) \right]$$

$$B(t_{lrd}) \cong \circ$$

$$B(T_{r\gamma}) \cong \left(\frac{1}{n} - \frac{1}{n'} \right) \left[\frac{R S_x^\gamma}{\bar{X}_U} - \frac{S_{xy}}{\bar{X}_U} \right]$$

$$B(T_{p\gamma}) \cong \left(\frac{1}{n} - \frac{1}{n'} \right) \frac{S_{xy}}{\bar{X}_U}$$

$$B(T_{d\gamma}) \cong \circ$$

همچنین واریانس براوردگرهای (۱) تا (۱۰) با اولین درجه‌ی تقریب از رابطه‌های زیر به دست می‌آیند:

$$\begin{aligned} \text{Var}(T_{r\gamma}) &\cong \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^r + \left(\frac{1}{n} - \frac{1}{n'}\right) S_r^r \\ &\quad + \frac{W_\gamma(k-1)}{n} S_{r\gamma}^r \end{aligned}$$

$$\begin{aligned} \text{Var}(T_{p\gamma}) &\cong \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^r + \left(\frac{1}{n} - \frac{1}{n'}\right) S_p^r \\ &\quad + \frac{W_\gamma(k-1)}{n} S_{p\gamma}^r \end{aligned}$$

$$\begin{aligned} \text{Var}(T_{d\gamma}) &\cong \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^r + \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^r (1 - \rho_{yx}^r) \\ &\quad + \frac{W_\gamma(k-1)}{n} S_{\beta(\gamma)}^r \end{aligned}$$

$$\begin{aligned} \text{Var}(t_d^{(r)}) &\cong \left\{ \left(\frac{1}{n} - \frac{1}{n'}\right) [S_y^r + \gamma R S_x^r (R - \beta_{yx})] \right. \\ &\quad \left. + \frac{W_\gamma(k-1)}{n} [S_{y\gamma}^r + R S_{x\gamma}^r (R - \gamma \beta_{yx\gamma})] + \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^r \right\} \end{aligned}$$

$$\begin{aligned} \text{Var}(t_d^{(p)}) &\cong \left\{ \left(\frac{1}{n} - \frac{1}{n'}\right) [S_y^r + \gamma R S_x^r (R + \beta_{yx})] \right. \\ &\quad \left. + \frac{W_\gamma(k-1)}{n} [S_{y\gamma}^r + R S_{x\gamma}^r (R + \gamma \beta_{yx\gamma})] + \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^r \right\} \end{aligned}$$

$$Var(t_e) \cong \left[\left(\frac{1}{n'} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{n'} \right) S_y^2 (1 - \rho_{yx}^2) + \frac{W_r(k-1)}{n} S_{y'r}^2 (1 - \rho_{yxr}^2) \right]$$

$$Var(t_{lrd}) = Var(t_e)$$

$$Var(T_{r'}) \cong \left(\frac{1}{n'} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{n'} \right) S_r^2 + \frac{W_r(k-1)}{n} S_{y'r}^2$$

$$Var(T_{P_r'}) \cong \left(\frac{1}{n'} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{n'} \right) S_p^2 + \frac{W_r(k-1)}{n} S_{y'r}^2$$

$$Var(T_{d_r'}) \cong \left(\frac{1}{n'} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{n'} \right) S_y^2 (1 - \rho_{yx}^2) + \frac{W_r(k-1)}{n} S_{y'r}^2$$

که در آن‌ها

$$W_r = \frac{N_r}{N}, \quad R = \frac{\bar{Y}_U}{\bar{X}_U}, \quad \rho_{yx} = \frac{S_{xy}}{S_x S_y}, \quad \beta_{yxr} = \frac{S_{xyr}}{S_{xr}^2},$$

$$\rho_{yxr} = \frac{S_{xyr}}{S_{xr} S_{y'r}}, \quad S_{y'r}^2 = \frac{1}{N_r - 1} \sum_{i=N_r+1}^{N_r+N_r} (Y_i - \bar{Y}_{rU})^2$$

$$S_r^2 = \left(S_y^2 + R^2 S_x^2 - 2R\rho_{yx} S_y S_x \right) = \left[S_y^2 + R S_x^2 (R - 2\rho_{yx}) \right],$$

$$S_{r\gamma}^{\gamma} = \left[S_{y\gamma}^{\gamma} + RS_{x\gamma}^{\gamma} (R - \gamma\beta_{yx\gamma}) \right],$$

$$S_p^{\gamma} = \left(S_y^{\gamma} + R^{\gamma} S_x^{\gamma} + \gamma R \rho_{yx} S_y S_x \right) = \left[S_y^{\gamma} + RS_x^{\gamma} (R + \gamma\beta_{yx}) \right],$$

$$S_{p\gamma}^{\gamma} = \left[S_{y\gamma}^{\gamma} + RS_{x\gamma}^{\gamma} (R + \gamma\beta_{yx\gamma}) \right],$$

$$S_{\beta(\gamma)}^{\gamma} = \left[S_{y\gamma}^{\gamma} + \beta_{yx} S_{x\gamma}^{\gamma} (\beta_{yx} - \gamma\beta_{yx\gamma}) \right],$$

$$S_{xy\gamma} = \frac{1}{N_{\gamma}-1} \sum_{i=N_{\gamma}+1}^N (X_i - \bar{X}_{\gamma U})(Y_i - \bar{Y}_{\gamma U}),$$

$$S_{x\gamma}^{\gamma} = \frac{1}{N_{\gamma}-1} \sum_{i=N_{\gamma}+1}^N (X_i - \bar{X}_{\gamma U})^{\gamma}, \quad S_y^{\gamma} = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y}_U)^{\gamma}.$$

در این مطالعه N_1 نمایانگر واحدهای پاسخگو در جامعه در اولین مراجعه است که در صورت مراجعه به آن‌ها پاسخگو خواهند بود و N_2 نمایانگر واحدهایی است که در اولین مراجعه پاسخگو نیستند و فرض می‌شود که با مراجعه مجدد پاسخگو خواهند بود. همچنین $\bar{X}_{\gamma U}$ و $\bar{Y}_{\gamma U}$ به ترتیب میانگین متغیرهای کمکی و مورد مطالعه از N_2 واحد بی‌پاسخ جامعه هستند و به صورت زیر تعریف می‌شوند:

$$\bar{X}_{\gamma U} = \frac{1}{N_{\gamma}} \sum_{i=N_{\gamma}+1}^{N_1+N_{\gamma}} X_i, \quad \bar{Y}_{\gamma U} = \frac{1}{N_{\gamma}} \sum_{i=N_{\gamma}+1}^{N_1+N_{\gamma}} Y_i$$

شیوه‌ی به دست آوردن اریبی و واریانس برآوردهای (۱) تا (۱۰) در پیوست آمده است (برای مطالعه‌ی بیشتر به [۱۱] مراجعه شود). همان‌طور که در پیوست مشاهده می‌شود واریانس دو برآوردها t_{ind} و t_e با هم برابر است.

۵- مقایسه‌ی واریانس برآوردها

در این مقاله تلاش می‌شود که بهترین برآوردها میانگین جامعه‌ای، با استفاده از روش نمونه‌گیری دوفازی معرفی شود. چون این برآوردها اریب هستند برای به دست آوردن

بهترین برآوردگر باید میانگین توان‌های دوم خطای آن‌ها (MSE) با هم مقایسه شود. اما از آنجا که مقایسه‌ی برآوردگرها با معیار MSE به صورت نظری مشکل است، سینگ و کومار [۱۱] با فرض انتخاب نمونه‌ای بزرگ و ناچیز بودن اریبی این برآوردگرها، واریانس آن‌ها را با هم مقایسه کرده‌اند. نتیجه‌های مقایسه‌ی واریانس برآوردگرهای (۱) تا (۱۰) با یکدیگر، توسط سینگ و کومار [۱۱] در زیر آمده است:

نتیجه‌ی ۱: برآوردگر t_e همواره دقیق‌تر از برآوردگر T_{r1} است. زمانی که $R = \beta_{yx} = \beta_{yx2}$ واریانس این دو برآوردگر یکسان است.

نتیجه‌ی ۲: برآوردگر t_e همواره دارای واریانسی کمتر از برآوردگر T_{p1} است. زمانی که رابطه‌ی $R = -\beta_{yx} = -\beta_{yx2}$ برقرار باشد آن‌گاه هر دو برآوردگر دقت یکسان دارند.

نتیجه‌ی ۳: برآوردگر t_e همواره دقیق‌تر از برآوردگر T_{d1} است. زمانی که رابطه‌ی $\beta_{yx} = \beta_{yx2}$ برقرار باشد، دقت این دو برآوردگر یکسان است.

نتیجه‌ی ۴: برآوردگر t_e همواره دقیق‌تر از برآوردگر T_{r2} است. زمانی که رابطه‌های $R = \beta_{yx}$ و $\rho_{yx2} = 0$ برقرار باشند، آن‌گاه دقت دو برآوردگر یکسان است.

نتیجه‌ی ۵: برآوردگر t_e همواره دقیق‌تر از برآوردگر T_{p2} است. زمانی واریانس این دو برآوردگر یکسان است که رابطه‌های $R = -\beta_{yx}$ و $\rho_{yx2} = 0$ برقرار باشند.

نتیجه‌ی ۶: برآوردگر t_e همواره دقیق‌تر از برآوردگر T_{d2} است. زمانی که رابطه‌ی $\rho_{yx2} = 0$ برقرار باشد، دقت دو برآوردگر یکسان می‌شود.

نتیجه‌ی ۷: برآوردگر t_e همواره دقیق‌تر از برآوردگر $t_d^{(r)}$ است. اگر رابطه‌ی $R = \beta_{yx} / 2 = \beta_{yx2}$ برقرار باشد، آن‌گاه دقت دو برآوردگر یکسان است.

نتیجه‌ی ۸: برآوردگر t_e همواره دقیق‌تر از برآوردگر $t_d^{(p)}$ است، به جز وقتی که رابطه‌ی $R = -\beta_{yx} / 2 = -\beta_{yx2}$ برقرار باشد.

نتیجه‌ی ۹: برآوردگر $t_d^{(r)}$ زمانی دقیق‌تر از برآوردگر T_{r1} است که رابطه‌های

$$R > \frac{3}{4} \beta_{yx} \text{ برای } R \text{ مثبت یا } R < \frac{3}{4} \beta_{yx} \text{ برای } R \text{ منفی برقرار باشند.}$$

نتیجه‌ی ۱۰: برآوردگر $t_d^{(r)}$ زمانی دقیق‌تر از برآوردگر T_{r_2} است که رابطه‌های $\beta_{yx} > \frac{3}{4}R$ و $\beta_{yx^2} > \frac{R}{4}$ برای R مثبت یا رابطه‌های $\beta_{yx} < \frac{3}{4}R$ و $\beta_{yx^2} < \frac{R}{4}$ برای R منفی برقرار باشند.

نتیجه‌ی ۱۱: برآوردگر $t_d^{(p)}$ زمانی دقیق‌تر از برآوردگر T_{p_1} است که رابطه‌ی $\beta_{yx} < -\frac{3}{4}R$ برای R مثبت یا رابطه‌ی $\beta_{yx} > -\frac{3}{4}R$ برای R منفی برقرار باشد.

نتیجه‌ی ۱۲: برآوردگر $t_d^{(p)}$ زمانی دقیق‌تر از برآوردگر T_{p_2} است که رابطه‌های $\beta_{yx} < -\frac{3}{4}R$ و $\beta_{yx^2} < -\frac{R}{4}$ برای R مثبت یا رابطه‌های $\beta_{yx} > -\frac{3}{4}R$ و $\beta_{yx^2} > -\frac{R}{4}$ برای R منفی برقرار باشند.

همان‌طور که در بخش چهارم اشاره شد، دو برآوردگر t_{ld} و t_e که توسط سینگ و کومار [۱۱] معرفی شدند، دارای واریانس یکسانی هستند. از مقایسه‌ی واریانس آن‌ها با واریانس دیگر برآوردگرها با در نظر گرفتن اطلاعات قبلی درمورد جامعه، نتیجه می‌شود که واریانس آن‌ها برابر یا کم‌تر از برآوردگرهای T_{r_1} ، T_{p_1} ، T_{d_1} ، T_{r_2} ، T_{p_2} ، T_{d_2} ، $t_d^{(r)}$ و $t_d^{(p)}$ است. در مقایسه‌ی دیگر برآوردگرها با یکدیگر اطلاعات قبلی بیشتری از جامعه، برای تعیین بهترین برآوردگر موردنیاز است.

۶- مطالعه‌ی شبیه‌سازی

در این بخش برآوردگرهای نسبتی T_{r_1} و T_{r_2} ، برآوردگرهای رگرسیون T_{d_1} و T_{d_2} ، که توسط خار و سریواستاوا [۳]، [۴] معرفی شده‌اند و برآوردگرهای t_e و t_{ld} که توسط سینگ و کومار [۱۱] پیشنهاد شده‌اند با انجام مطالعه‌ی شبیه‌سازی با هم مقایسه می‌شوند. شاخص ارزیابی برآوردگرهای پیشنهادی میانگین توان‌های دوم خطای شبیه‌سازی شده است.

برای این منظور با استفاده از نرم‌افزار R داده‌های مربوط به دو متغیر قد (در مقیاس سانتی‌متر) و وزن (در مقیاس کیلوگرم) از توزیع نرمال دومتغیری با همبستگی معین و

مثبتی تولید می‌شوند که آن‌ها را به‌عنوان مقادیرهای جامعه‌ی مورد مطالعه در نظر می‌گیریم. در این مطالعه‌ی شبیه‌سازی، هدف، تولید جامعه‌ای با جمعیت $N = 100,000$ نفر با دو متغیر X و Y و برآورد پارامترهای توزیع نمونه‌گیری برآوردگرها در یک طرح دوفازی برای تعدیل اثر بی‌پاسخی است. متغیر وزن را به‌عنوان متغیر مورد مطالعه‌ی Y و متغیر قد را به‌عنوان متغیر کمکی X در نظر می‌گیریم.

اگر Y_j و X_j به‌ترتیب نشان‌دهنده‌ی متغیر مورد مطالعه و کمکی برای فرد j ام باشند، با فرض نرمال بودن متغیرهای مورد نظر و وجود همبستگی ρ_{yx} بین متغیرها، از توزیع نرمال دومتغیری زیر برای تولید داده‌های جامعه‌ی آماری مورد نظر استفاده می‌شود.

$$(X, Y) \sim N_2 \left(\mu = \begin{bmatrix} 62 \\ 163 \end{bmatrix}, \Sigma = \begin{bmatrix} 2/65 & \rho_{yx} \times 2/65 \times 3/32 \\ \rho_{yx} \times 2/65 \times 3/32 & 3/32 \end{bmatrix} \right)$$

که در آن ρ_{yx} همبستگی بین متغیرهای X و Y را نشان می‌دهد. میانگین وزن افراد ۶۲ کیلوگرم با انحراف استاندارد ۲/۶۵ کیلوگرم و میانگین قد آن‌ها ۱۶۳ سانتی‌متر با انحراف استاندارد ۳/۳۲ سانتی‌متر در نظر گرفته شده است. با این فرض که برای متغیر مورد مطالعه بی‌پاسخی وجود دارد، با در نظر گرفتن نرخ‌های بی‌پاسخی متفاوت، جامعه را به دو طبقه‌ی پاسخ‌گو و بی‌پاسخ تقسیم می‌کنیم. فرض می‌کنیم که از ابتدا میانگین متغیر کمکی نامعلوم است و با انتخاب نمونه‌ای بزرگ به اندازه‌ی $n' = 10,000$ از جامعه آن را برآورد می‌کنیم.

همچنین در این مطالعه‌ی شبیه‌سازی (با استفاده از $K = 1000$ مرتبه تکرار نمونه‌گیری خودگردان به‌صورت مستقل از هم) تأثیر سه اندازه‌ی نمونه متفاوت $n = 500, 1000, 2000$ ، سه همبستگی متفاوت $\rho_{yx} = 0.55, 0.70, 0.85$ و سه نرخ بی‌پاسخی $0.25, 0.35$ و 0.45 را بر روی برآورد میانگین توان دوم خطای شبیه‌سازی برآوردگرها بررسی می‌کنیم.

میانگین‌ها و واریانس‌های نمونه را به‌ترتیب برای تقریب امید ریاضی و واریانس برآوردگرها به کار می‌بریم و میانگین توان دوم خطای شبیه‌سازی برآوردگرهای T_{r_2}, T_{r_1} ، T_{d_2}, T_{d_1} و $t_e, t_d^{(r)}$ را با هم مقایسه می‌کنیم. فرض کنید t یکی از

براوردهای $T_{r1}, T_{r2}, T_{d1}, T_{d2}, t_d^{(r)}, t_e$ و t_{lrd} باشد. در تکرار i ام نمونه‌گیری مقدار این برآورد را با $t_i, i = 1, 2, \dots, K$ نشان می‌دهیم. آنگاه برآورد میانگین و واریانس توزیع نمونه‌گیری آماره‌ی t عبارت است از:

$$\bar{t} = \frac{1}{K} \sum_{i=1}^K t_i, \quad Var(t) = \frac{1}{K} \sum_{i=1}^K (t_i - \bar{t})^2$$

در واقع \bar{t} و $Var(t)$ برآوردی از میانگین و واریانس توزیع نمونه‌گیری t هستند. همچنین برآورد اریبی برای برآوردگر t به صورت $B(t) = (\bar{t} - \bar{Y}_U)$ است که در آن \bar{Y}_U میانگین واقعی جامعه برای متغیر مورد مطالعه Y است. میانگین توان دوم خطای شبیه‌سازی شده ($SMSE$) در برآوردگر t نیز از رابطه‌ی زیر به دست می‌آید:

$$SMSE(t) = Var(t) + [B(t)]^2 = \frac{1}{K} \sum_{i=1}^K (t_i - \bar{Y}_U)^2$$

برای هر اندازه‌ی نمونه و هر نرخ بی‌پاسخی و ضریب همبستگی بین X و Y برنامه را ۱۰۰۰ مرتبه اجرا کردیم. مقدار $SMSE$ برآوردها در جدول‌های ۱ تا ۳ آمده است. لازم به ذکر است که مقدارهای $SMSE$ در این جدول‌ها تا چهار رقم اعشار گرد شده‌اند.

جدول ۱- میانگین توان دوم خطای شبیه‌سازی برای ضریب همبستگی ۰/۵۵

۰/۲۵			۰/۳۵			۰/۴۵			نرخ بی‌پاسخی
۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰	اندازه‌ی نمونه
۰/۰۰۲۹	۰/۰۰۶۲	۰/۰۱۳۷	۰/۰۰۴۱	۰/۰۰۷۷	۰/۰۱۴۹	۰/۰۰۰۴	۰/۰۰۸۵	۰/۰۱۷	$SMSE(T_{r1})$
۰/۰۰۲۹	۰/۰۰۶۲	۰/۰۱۳۷	۰/۰۰۴۱	۰/۰۰۷۷	۰/۰۱۴۹	۰/۰۰۰۴	۰/۰۰۸۵	۰/۰۱۷	$SMSE(T_{d1})$
۰/۰۰۳۲	۰/۰۰۶۸	۰/۰۱۵	۰/۰۰۴۶	۰/۰۰۸۹	۰/۰۱۶۵	۰/۰۰۰۵	۰/۰۱۰۳	۰/۰۲۰	$SMSE(T_{r2})$
۰/۰۰۳۲	۰/۰۰۶۸	۰/۰۱۵	۰/۰۰۴۶	۰/۰۰۸۹	۰/۰۱۶۵	۰/۰۰۰۵	۰/۰۱۰۳	۰/۰۲۰	$SMSE(T_{d2})$
۰/۰۰۳۳	۰/۰۰۷۷	۰/۰۱۶۲	۰/۰۰۵۸	۰/۰۰۸۵	۰/۰۱۶۹	۰/۰۰۴۴	۰/۰۰۹۸	۰/۰۱۹۵	$SMSE(t_{dr})$
۰/۰۰۳۱	۰/۰۰۶۵	۰/۰۱۴۴	۰/۰۰۴۴	۰/۰۰۸۳	۰/۰۱۶۴	۰/۰۰۴۲	۰/۰۰۹۱	۰/۰۱۸۶	$SMSE(t_{lrd})$
۰/۰۰۳۱	۰/۰۰۶۵	۰/۰۱۴۴	۰/۰۰۴۴	۰/۰۰۸۳	۰/۰۱۶۴	۰/۰۰۴۲	۰/۰۰۹۱	۰/۰۱۸۶	$SMSE(t_e)$

جدول ۲- میانگین توان دوم خطای شبیه‌سازی برای ضریب همبستگی ۰/۷۰

نرخ بی‌پاسخی	۰/۴۵			۰/۳۵			۰/۲۵		
	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰
اندازه‌ی نمونه	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰
$SMSE(T_{r1})$	۰/۰۰۳۲	۰/۰۰۷۷	۰/۰۱۵	۰/۰۰۲۵	۰/۰۰۶۲	۰/۰۱۲۳	۰/۰۰۲۲	۰/۰۰۵۱	۰/۰۱۰۸
$SMSE(T_{d1})$	۰/۰۰۳۲	۰/۰۰۷۷	۰/۰۱۵	۰/۰۰۲۵	۰/۰۰۶۲	۰/۰۱۲۳	۰/۰۰۲۲	۰/۰۰۵۱	۰/۰۱۰۸
$SMSE(T_{r2})$	۰/۰۰۴۴	۰/۰۱۰۳	۰/۰۲۰	۰/۰۰۳۳	۰/۰۰۷۷	۰/۰۱۵۶	۰/۰۰۲۷	۰/۰۰۵۹	۰/۰۱۲۸
$SMSE(T_{d2})$	۰/۰۰۴۴	۰/۰۱۰۳	۰/۰۲۰	۰/۰۰۳۳	۰/۰۰۷۷	۰/۰۱۵۶	۰/۰۰۲۷	۰/۰۰۵۹	۰/۰۱۲۸
$SMSE(t_{dr})$	۰/۰۰۳۲	۰/۰۰۸۹	۰/۰۱۵۹	۰/۰۰۲۸	۰/۰۰۶۵	۰/۰۱۲۳	۰/۰۰۲۸	۰/۰۰۵۱	۰/۰۱۱۲
$SMSE(t_{trd})$	۰/۰۰۳۶	۰/۰۰۸۷	۰/۰۱۶۵	۰/۰۰۲۶	۰/۰۰۶۶	۰/۰۱۳۳	۰/۰۰۲۴	۰/۰۰۵۶	۰/۰۱۱۵
$SMSE(t_e)$	۰/۰۰۳۶	۰/۰۰۸۷	۰/۰۱۶۶	۰/۰۰۲۶	۰/۰۰۶۶	۰/۰۱۳۳	۰/۰۰۲۴	۰/۰۰۵۶	۰/۰۱۱۵

جدول ۳- میانگین توان دوم خطای شبیه‌سازی برای ضریب همبستگی ۰/۸۵

نرخ بی‌پاسخی	۰/۴۵			۰/۳۵			۰/۲۵		
	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰
اندازه‌ی نمونه	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰	۲۰۰۰	۱۰۰۰	۵۰۰
$SMSE(T_{r1})$	۰/۰۰۳	۰/۰۰۵	۰/۰۰۹۸	۰/۰۰۲	۰/۰۰۵	۰/۰۰۹	۰/۰۰۲	۰/۰۰۴	۰/۰۰۸
$SMSE(T_{d1})$	۰/۰۰۳	۰/۰۰۵	۰/۰۰۹۸	۰/۰۰۲	۰/۰۰۵	۰/۰۰۹	۰/۰۰۲	۰/۰۰۴	۰/۰۰۸
$SMSE(T_{r2})$	۰/۰۰۴	۰/۰۰۹	۰/۰۱۶۲	۰/۰۰۳	۰/۰۰۷	۰/۰۱۳۲	۰/۰۰۲	۰/۰۰۵	۰/۰۱۱
$SMSE(T_{d2})$	۰/۰۰۴	۰/۰۰۹	۰/۰۱۶۲	۰/۰۰۳	۰/۰۰۷	۰/۰۱۳۲	۰/۰۰۲	۰/۰۰۵	۰/۰۱۱
$SMSE(t_{dr})$	۰/۰۰۲	۰/۰۰۴	۰/۰۰۸۴	۰/۰۰۲	۰/۰۰۴	۰/۰۰۶۷	۰/۰۰۱	۰/۰۰۴	۰/۰۰۶
$SMSE(t_{trd})$	۰/۰۰۳	۰/۰۰۶	۰/۰۱۱۸	۰/۰۰۲	۰/۰۰۵	۰/۰۱۰۳	۰/۰۰۲	۰/۰۰۴	۰/۰۰۸
$SMSE(t_e)$	۰/۰۰۳	۰/۰۰۶	۰/۰۱۱۸	۰/۰۰۲	۰/۰۰۵	۰/۰۱۰۳	۰/۰۰۲	۰/۰۰۴	۰/۰۰۸

نتایج به دست آمده از جدول‌های ۱، ۲ و ۳:

۱. برای هر نرخ پاسخ‌گویی با افزایش اندازه‌ی نمونه و افزایش میزان همبستگی بین دو برآوردگر، $SMSE$ برآوردگرها کاهش می‌یابد.
۲. برای هر اندازه‌ی نمونه با افزایش نرخ پاسخ‌گویی (کاهش نرخ بی‌پاسخی) و افزایش میزان همبستگی بین متغیر کمکی و مورد مطالعه، اغلب $SMSE$ برآوردگرها کاهش می‌یابد.

۳. برای هر مقدار از ضریب همبستگی بین دو متغیر X و Y با افزایش اندازه‌ی نمونه و کاهش نرخ بی‌پاسخی، اغلب $SMSE$ برآوردهای T_{r1} ، T_{r2} ، T_{d1} ، T_{d2} ، t_{dr} ، t_e و t_{lrd} کاهش می‌یابد.
۴. $SMSE$ دو برآوردهای t_e ، t_{lrd} تقریباً با هم برابر است. همچنین $SMSE$ دو برآوردهای T_{r1} و T_{d1} و نیز $SMSE$ دو برآوردهای T_{r2} و T_{d2} نیز تقریباً با هم برابر هستند.
۵. از بین برآوردهای T_{r1} ، T_{r2} ، T_{d1} ، T_{d2} ، $t_d^{(r)}$ ، t_e و t_{lrd} برای نرخ‌های پاسخ‌گویی، همبستگی و اندازه‌های نمونه‌ی متفاوت، مقدار $SMSE$ برآوردهای $t_d^{(r)}$ برای هر اندازه‌ی نمونه و هر نرخ بی‌پاسخی، برای همبستگی $0/85$ از همه کم‌تر است.
۶. برای ضریب همبستگی متوسط ($0/55$ و $0/70$) مقدار $SMSE$ دو برآوردهای T_{r1} و T_{d1} تقریباً نزدیک به هم است. این دو برآوردهای مابقی برآوردها بهتر عمل می‌کنند و بهتر است برای برآورد میانگین جامعه‌ای از آن‌ها استفاده شود. در حالی که برای ضریب همبستگی بالا ($0/85$) $SMSE$ برآوردهای $t_d^{(r)}$ از مابقی برآوردها کم‌تر است و این برآوردها بهتر از بقیه عمل می‌کند.

۷- نتیجه‌گیری

در این مقاله تلاش شد بهترین برآوردهای میانگین جامعه در حضور بی‌پاسخی به روش نمونه‌گیری دوفازی تعیین شود. نشان داده شد که با فرض ناچیز بودن اریبی، واریانس دو برآوردهای t_e و t_{lrd} (که دارای واریانس یکسان هستند)، برابر، یا کم‌تر از واریانس برآوردهای T_{r1} ، T_{r2} ، T_{d1} ، T_{d2} ، T_{P1} ، T_{P2} ، $t_d^{(r)}$ و $t_d^{(p)}$ است. بنا بر این با فرض ناچیز بودن اریبی، می‌توان این دو برآوردها را به‌عنوان بهترین برآوردهای میانگین جامعه معرفی کرد. اما در عمل ممکن است اندازه‌ی نمونه آن‌قدر بزرگ نباشد که موجب ناچیز بودن اریبی برآوردها شود. بنا بر این با انجام مطالعه‌ی شبیه‌سازی و در نظر گرفتن اریبی برآوردها، شرایط مناسب برای کم‌تر شدن میانگین توان دوم خطای شبیه‌سازی‌شده‌ی برآوردها بررسی شد. برای این منظور $SMSE$ برآوردها برای

اندازه‌های نمونه، نرخ بی‌پاسخی و همبستگی‌های متفاوت برای ۱۰۰۰ نمونه‌ی خودگردان محاسبه شد. مشاهده شد که برای ضریب‌های همبستگی ۰/۵۵ و ۰/۷۰ مقدار $SMSE$ دو برآوردگر T_{r1} و T_{d1} کم‌تر از دیگر برآوردگرها و تقریباً نزدیک به هم است و این برآوردگرها از مابقی برآوردگرها بهتر عمل می‌کنند، در حالی که برای ضریب همبستگی $SMSE$ برآوردگر $t_d^{(r)}$ از مابقی برآوردگرها کم‌تر است و این برآوردگر بهتر از بقیه عمل می‌کند. بنا بر این پیش‌نهاد می‌شود هنگام مواجهه با بی‌پاسخی برای برآورد بهتر میانگین جامعه‌ای به روش نمونه‌گیری دوفازی زمانی که بین متغیر مورد مطالعه و متغیر کمکی میزان همبستگی متوسط است، از برآوردگرهای T_{r1} و T_{d1} برای برآورد بهتر میانگین جامعه‌ای استفاده شود و هنگامی که با میزان بالایی از همبستگی بین متغیرهای مورد مطالعه و کمکی روبرو شدیم از برآوردگر $t_d^{(r)}$ برای برآورد هرچه دقیق‌تر میانگین جامعه استفاده شود.

همچنین مشاهده شد که با افزایش اندازه‌ی نمونه می‌توان دقت برآوردگر میانگین جامعه‌ای را به میزان زیادی بهبود بخشید. با به‌کارگیری جدی روش‌های مناسب در طرح نمونه‌گیری (ایجاد اطمینان، سازگاری بین پرسش‌گر و پاسخ‌گو و ۰۰۰) و افزایش نرخ پاسخگویی نیز می‌توان اغلب، دقت برآوردگرها را افزایش داد. همچنین هرچه متغیر کمکی انتخابی، دارای همبستگی بیش‌تری با متغیر مورد مطالعه باشد نیز دقت برآوردگر افزایش می‌یابد. بنا بر این دقت در انتخاب متغیر کمکی، افزایش اندازه‌ی نمونه به مقدار کافی و کاهش خطاهای غیر نمونه‌گیری به‌ویژه خطای بی‌پاسخی، موجب افزایش دقت برآوردگر میانگین جامعه می‌شود.

مرجع‌ها

- [1] Cochran, W.G. (1977). *Sampling Technique*, John Wiley and Sons, (3rd ed.) ed., New York.
- [2] Hansen, M.H. and Hurwitz, W.N. (1946). The Problem of Non-Response in Sample Surveys, *Journal of the American Statistical Association*, **41**, 517-529.

- [3] Khare, B.B. and Srivastava, S. (1993). Estimation of Population Mean Using Auxiliary Character in the Presence of Non-Response, *Nat. Acad. Sci. Lett (India)*, **16**, 111–114.
- [4] Khare, B.B. and Srivastava, S. (1995). Study of Conventional and Alternative Two-Phase Sampling Ratio, Product and Regression Estimators in the Presence of Non-Response, *Proc. Natl. Acad. Sci. (India)*, **65**, 195–203.
- [5] Okafor, F.C. (2001). Treatment of Non-Response in Successive Sampling, *Journal of the American Statistical LXI*, **26**, 195–204.
- [6] Okafor, F.C. (2005). Sub-Sampling the Non-Respondents in Two-Stage Sampling Over Two Successive Occasions, *Journal of the Indian Statistical Association*, **43**, 33–49.
- [7] Okafor, F.C. and Lee, H. (2000). Double Sampling for Ratio and Regression Estimation With Sub-Sampling the Non-Respondents, *Survey Methodology*, **26**, 183–188.
- [8] Rao, P.S.R.S. (1986). Ratio Estimation with Sub-Sampling the Non-Respondents, *Survey Methodology*, **12**, 217–230.
- [9] Rao, P.S.R.S. (1987). Ratio and Regression Estimates with Sub-Sampling the Non-Respondents, *Paper Presented at A special Contributed Session of the International Statistical Association Meeting*, September 2–16, Tokyo, Japan.
- [10] Robson, D.S. (1957). Application of Multivariate Polykays to the Theory of Unbiased Ratio Type Estimation, *Journal of the American Statistical Association*, **52**, 511–522.
- [11] Singh, H.P. and Kumar, S. (2008). Estimation of Mean in Presence of Non-Response Using Two-Phase Sampling Scheme, *Statistical Papers*, **51**, 559–582.
- [12] Singh, H.P. and Kumar, S. (2010). Improved Estimation of Population Mean Under Two-Phase Sampling with Sub-Sampling

the Non-Respondents,” *Journal of Statistical Planning and Inference*, **140**, 2536–2550.

- [13] Tabasum, R. and Khan, I.A. (2004). Double Sampling for Ratio Estimation with Non-Response, *Journal of the Indian Association Agriculture Statistical*, **58**, 300–306.
- [14] Tabasum, R. and Khan, I.A. (2006). Double Sampling Ratio Estimator for the Population Mean in Presence of Non-Response, *Assam Statistical Review*, **20**, 73–83.

پیوست

برای اثبات اریبی و واریانس برآوردهای ۱ تا ۱۰ که در رابطه‌های ۱۱ تا ۳۰ آمده‌اند از عبارت‌های زیر استفاده می‌شود. برای این منظور ابتدا قرار دهید:

$$\begin{aligned} \bar{y}_n^* &= \bar{Y}_U + (\bar{y}_n^* - \bar{Y}_U) = \bar{Y}_U + \varepsilon_0^* , & \varepsilon_0^* &= \bar{y}_n^* - \bar{Y}_U \\ \bar{x}_n^* &= \bar{X}_U + (\bar{x}_n^* - \bar{X}_U) = \bar{X}_U + \varepsilon_1^* , & \varepsilon_1^* &= \bar{x}_n^* - \bar{X}_U \\ \bar{x}_{n'} &= \bar{X}_U + (\bar{x}_{n'} - \bar{X}_U) = \bar{X}_U + \varepsilon_1' , & \varepsilon_1' &= \bar{x}_{n'} - \bar{X}_U \\ \bar{x}_n &= \bar{X}_U + (\bar{x}_n - \bar{X}_U) = \bar{X}_U + \varepsilon_1 , & \varepsilon_1 &= \bar{x}_n - \bar{X}_U \end{aligned}$$

با استفاده از رابطه‌های بالا برای محاسبه‌ی اریبی و واریانس برآوردهای ۱ تا ۱۰ از لم زیر استفاده می‌کنیم:

لم ۱- با توجه به رابطه‌های بالا داریم

$$E(\varepsilon_0^*) = E(\varepsilon_1^*) = E(\varepsilon_1') = E(\varepsilon_1) = 0 \quad \text{الف-}$$

ب-

$$Var(\varepsilon_0^*) = \left(\frac{1-f}{n}\right) S_y^2 + \frac{W_y(k-1)}{n} S_{y^2}^2$$

$$\text{Var}(\varepsilon_{\gamma}^*) = \left(\frac{1-f}{n}\right) S_x^2 + \frac{W_{\gamma}(k-1)}{n} S_{x\gamma}^2$$

$$\text{Var}(\varepsilon'_{\gamma}) = \left(\frac{1-f'}{n'}\right) S_x^2, \quad f' = \frac{n'}{N}$$

$$\text{Var}(\varepsilon_{\gamma}) = \left(\frac{1-f}{n}\right) S_x^2$$

$$E(\varepsilon_{\gamma}^* \varepsilon_{\gamma}^*) = \left(\frac{1-f}{n}\right) S_{xy} + \frac{W_{\gamma}(k-1)}{n} S_{xy\gamma}$$

$$E(\varepsilon_{\gamma}^* \varepsilon_{\gamma}) = \left(\frac{1-f}{n}\right) S_{xy}$$

$$E(\varepsilon_{\gamma}^* \varepsilon'_{\gamma}) = \left(\frac{1-f'}{n'}\right) S_{xy}$$

$$E(\varepsilon_{\gamma} \varepsilon_{\gamma}) = \left(\frac{1-f}{n}\right) S_x^2$$

$$E(\varepsilon_{\gamma} \varepsilon_{\gamma}^*) = \left(\frac{1-f'}{n'}\right) S_x^2$$

$$E(\varepsilon_{\gamma} \varepsilon'_{\gamma}) = \left(\frac{1-f'}{n'}\right) S_x^2$$

با استفاده از لم ۱ به راحتی می‌توان رابطه‌های (۱۱) تا (۳۰) را به دست آورد.

مریم تقریبی
فوق لیسانس آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه طباطبائی، گروه آمار.
رایانشانی: maryamtagharobi@yahoo.com

نادر نعمت‌الهی
دکتری آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه طباطبائی، گروه آمار.
رایانشانی: na_nemat@yahoo.com

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