مجله‌ی بررسی‌های آمار رسمی ایران
سال۲۴، شماره‌ی ۱، بهار و تابستان ۱۳۹۲، صص ۱۹–۴۱

# براورد میانگین جامعه در حضور بی‌پاسخی با استفاده از روش نمونه‌گیری دوفازی

مریم تقربی[*]، نادر نعمت‌الهی و حمیدرضا نواب‌پور

دانشگاه علامه طباطبایی

**چکیده:** در این مقاله مسئله‌ی براورد میانگین جامعه با استفاده از اطلاعات یک متغیر کمکی در حضور بی‌پاسخی مورد بررسی قرار گرفته است. برای این منظور براوردگرهای نسبتی، ضربی و رگرسیونی برای براورد میانگین جامعه با روش نمونه‌گیری دوفازی پیش‌نهاد شده و ویژگی‌های آن‌ها مورد مطالعه قرار گرفته است. این براوردگرها اریب هستند و مقایسه‌ی میانگین توان دوم خطای آن‌ها به‌صورت نظری مشکل است. بنا بر این با فرض ناچیز بودن اریبی برای اندازه‌ی نمونه‌ی بزرگ و مقایسه‌ی واریانس‌ها، بهترین براوردگر 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: براوردگر رگرسیونی؛ براوردگر ضربی؛ براوردگر نسبتی؛ براوردگر 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، به‌علت ناکامی در دست‌یابی به اطلاعات مفید در

* نویسنده‌ی عهده‌دار مکاتبات
دریافت: ۱۳۹۱/۱۰/۱۰، پذیرش: ۱۳۹۳/۵/۱.

مورد یك یا چند متغیر مورد بررسی برای یك یا چند عضو واجد شرایط در آمارگیری رخ می‌دهد. حتی اگر از طرح‌های آمارگیری دقیقی برای مینیمم کردن دیگر منبع‌های بروز خطا استفاده شود، وجود بی‌پاسخی ممکن است تمام نتیجه‌های آمارگیری را نادقیق کند. همچنین براوردهای 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، یك 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 براورد می‌شود. بالا بودن نرخ بی‌پاسخی موجب کوچک شدن اندازه‌ی نمونه‌ای، کاهش دقت و اریبی براوردگرها می‌شود. اگر با رعایت تمام موردها در طرح‌های آمارگیری، به‌دلیل‌هایی بی‌پاسخی رخ دهد از تعدیل‌های پساآمارگیری که عبارت‌اند از، تعدیل موزون و جانهی برای حداقل کردن اثر خطای بی‌پاسخی براوردگرها 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. براوردگر پیش‌نهادی آن‌ها ترکیبی از اطلاعات پاسخ‌گویان و زیرنمونه‌ی انتخابی از بی‌پاسخ‌ها است. همچنین بهتر است برای افزایش دقت براوردگرها از اطلاعات کمکی که همان داده‌های ثبتی یا نتایج سرشماری‌های قبلی هستند، استفاده شود. اگرچه ممکن است براوردگرهای حاصل در مقایسه با براوردگرهای متعارف اریب باشند (که با افزایش اندازه‌ی نمونه اریبی آن‌ها ناچیز خواهد شد)، اما نسبت به براوردگرهای متعارف دارای واریانسی کم‌تر هستند. این اطلاعات کمکی هم در طرح نمونه‌گیری و هم در براورد بهتر پارامترهای مورد نظر مفید هستند. هرچه همبستگی متغیر کمکی با متغیر مورد مطالعه بیش‌تر باشد، شاهد افزایش دقت و کاهش اریبی براوردهای 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 گرداوری می‌شود. در فاز دوم نمونه‌گیری اغلب بی‌پاسخی وجود دارد. براورد 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] براورد میانگین جامعه‌ای متغیر مورد مطالعه را در دو حالت گفته‌شده بررسی کرده‌اند.

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

در بخش بعد، چگونگی استفاده از اطلاعات کمکی به‌منظور افزایش دقت براوردگرهای حاصل از روش مراجعه‌ی مجدد، با عنوان نمونه‌گیری دوفازی مورد بررسی قرار گرفته است. در بخش سوم تعدادی براوردگر نسبتی، ضربی و رگرسیونی برای براورد 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 براوردگرها با هم مقایسه و بهترین براوردگر میانگین جامعه‌ای تعیین می‌شود. در بخش ششم، با انجام یک مطالعه‌ی شبیه‌سازی، براورد میانگین توان دوم خطای براوردگرها تحت شرایط مختلف همبستگی، نرخ بی‌پاسخی و اندازه‌ی نمونه با هم مقایسه می‌شود و در نهایت بهترین شرایط برای براورد 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 براوردگرهای به دست آمده از روش مراجعه‌ی مجدد را بهبود بخشید. مهم‌ترین براوردگرهایی که از اطلاعات یک متغیر کمکی استفاده می‌کنند عبارت‌اند از براوردگرهای نسبتی، ضربی، رگرسیونی و براوردگرهای هم‌گذاشتی. براوردگرهای حاصل از روش دوفازی نیز حالتی از این براوردگرها هستند.

جامعه‌ی $U=\{Y_1,Y_2,\ldots,Y_N\}$ را در نظر بگیرید که در آن $Y_i$ها می‌توانند گسسته یا پیوسته باشند. فرض می‌کنیم با تعدادی از مقدارهای $Y_i$ی 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 براورد 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 $(\bar{x}_{n'})$ براورد 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 نمونه‌گیری‌شده، $n_1$ واحد

پاسخ‌گو و $n_{۲}$ واحد بی‌پاسخ باشند. آمارشناسان با استفاده از روش مراجعه‌ی مجدد بر اساس نرخ پاسخ‌گویی $۱/k$ ($۱/k = n_{۱}/n$) زیرنمونه‌ای به اندازه‌ی $n_{۲m}$ ($n_{۲m} = n_{۲}/k$) به روش نمونه‌گیری تصادفی ساده بدون جای‌گذاری از $n_{۲}$ 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 به‌صورت زیر براورد می‌شوند:

$$\bar{y}_n^* = \frac{n_{۱}}{n}\bar{y}_{۱} + \frac{n_{۲}}{n}\bar{y}_{۲m}, \qquad \bar{x}_n^* = \frac{n_{۱}}{n}\bar{x}_{۱} + \frac{n_{۲}}{n}\bar{x}_{۲m}$$

که در آن، $(\bar{x}_{۱}, \bar{y}_{۱})$ و $(\bar{x}_{۲m}, \bar{y}_{۲m})$ به‌ترتیب میانگین‌های نمونه‌ی $n_{۱}$ تایی و زیرنمونه‌ی $n_{۲m}$ تایی برای متغیرهای $(X, Y)$ هستند. همچنین $\bar{x}_{n'} = \frac{۱}{n'}\sum_{i=۱}^{n'} x_i$ میانگین متغیر کمکی (بر اساس اطلاعات نمونه‌گیری در فاز اول) است.

برای براورد میانگین جامعه با استفاده از روش نمونه‌گیری دوفازی براوردگرهای زیادی پیش‌نهاد شده است که در بخش بعد معرفی می‌شوند.

## ۳- براوردگرهای پیش‌نهادشده به‌روش دوفازی

براوردگرهایی که از اطلاعات یک متغیر کمکی استفاده می‌کنند عبارت‌اند از براوردگرهای نسبتی، ضربی و رگرسیونی. براوردگرهای پیش‌نهادشده توسط آمارشناسان در این بخش، حالتی از این براوردگرها هستند. اگر دسترسی به اطلاعات متغیرهای کمکی به‌سادگی 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، براوردگرهای نسبتی، ضربی و رگرسیونی به روش نمونه‌گیری دوفازی در حضور بی‌پاسخی به دو دسته به‌صورت زیر تقسیم شده‌اند.

**دسته‌ی اول:** این دسته شامل براوردگرهای نسبتی $T_{r1}$، ضربی $T_{p1}$، رگرسیونی $T_{d1}$ و براوردگرهای هم‌گذاشتی است که عبارت‌اند از $t_d^{(r)}$، $t_d^{(p)}$ ، $t_e$ و $t_{lrd}$ و به‌صورت زیر تعریف می‌شوند:

$$T_{r1} = \frac{\bar{y}_n^*}{\bar{x}_n^*}\bar{x}_{n'} \quad (۱)$$

$$T_{p1} = \frac{\bar{y}_n^*}{\bar{x}_{n'}}\bar{x}_n^* \quad (۲)$$

$$T_{d1} = \bar{y}_n^* + \hat{\beta}_{yx}^*\left(\bar{x}_{n'} - \bar{x}_n^*\right) \quad (۳)$$

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

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

$$t_e = \bar{y}_n^*\left(\frac{\bar{x}_n}{\bar{x}_n^*}\right)^{\hat{\beta}_{yx2}/\hat{R}}\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^{\hat{\beta}_{yx}^{**}/\hat{R}} \quad (۶)$$

$$t_{lrd} = \bar{y}_n^* + \hat{\beta}_{yx2}\left(\bar{x}_n - \bar{x}_n^*\right) + \hat{\beta}_{yx}^{**}\left(\bar{x}'_n - \bar{x}_n\right) \quad (۷)$$

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. $\hat{\beta}_{yx}^* = \frac{s_{xy}^*}{s_x^{*2}}$ و $\hat{\beta}_{yx}^{**} = \left(s_{xy}^{**} / s_x^{**2}\right)$ براورد ضریب رگرسیونی $\beta_{yx} = \frac{S_{XY}}{S_X^2}$ در جامعه هستند که در آن‌ها $s_{xy}^*$، $s_{xy}^{**}$، $s_x^{*2}$ و $s_x^{**2}$ به‌صورت زیر تعریف می‌شوند:

$$s_{xy}^{*}=\frac{۱}{n-۱}\left(\sum_{u_۱}x_i y_i+n_{۲m}\sum_{u_{۲m}}x_i y_i-n\bar{x}_n^{*}\,\bar{y}_n^{*}\right),$$

$$s_x^{*۲}=\frac{۱}{n-۱}\left(\sum_{u_۱}x_i^۲+n_{۲m}\sum_{u_{۲m}}x_i^۲-n\,\bar{x}_n^{*۲}\right),$$

$$s_{xy}^{**}=\frac{۱}{n-۱}\left(\sum_{u_۱}x_i y_i+n_{۲m}\sum_{u_{۲m}}x_i y_i-n\bar{x}_n\,\bar{y}_n^{*}\right),$$

$$s_x^{**۲}=\frac{۱}{n-۱}\left(\sum_{u_۱}x_i^۲+n_{۲m}\sum_{u_{۲m}}x_i^۲-n\,\bar{x}_n^۲\right),$$

$$S_x^۲=\frac{۱}{N-۱}\sum_{i=۱}^{N}\left(X_i-\bar{X}_U\right)^۲,$$

$$S_{xy}=\frac{۱}{N-۱}\sum_{i=۱}^{N}\left(X_i-\bar{X}_U\right)\left(Y_i-\bar{Y}_U\right).$$

همچنین

$$\hat{\beta}_{yx۲}=\frac{s_{xy۲}}{s_{x۲}^۲},\ \hat{R}=\frac{\bar{y}_n^{*}}{\bar{x}_n^{*}},\ \ \hat{\beta}_{yx}=\frac{s_{xy}^{**}}{s_x^{**۲}}=\hat{\beta}_{yx}^{**},$$

$$s_{xy۲}=\frac{۱}{\left(n_{۲m}-۱\right)}\sum_{i=۱}^{n_{m۲}}\left(x_i-\bar{x}_{۲m}\right)\left(y_i-\bar{y}_{۲m}\right),$$

$$s_{x۲}^۲=\frac{۱}{\left(n_{۲m}-۱\right)}\sum_{i=۱}^{n_{m۲}}\left(x_i-\bar{x}_{۲m}\right)^۲.$$

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

$s_{xy}^{*}$ و $s_x^{۲} = ۱/(n-۱)\sum_{i=۱}^{n}(x_i - \overline{x}_n)^{۲}$ معرفی شده است. طبق مطالعاتی که محقق در این زمینه انجام داده است، ادعا می‌شود که اگر از $\hat{\beta}_{yx}^{**} = \left(s_{xy}^{*} / s_x^{۲}\right)$ در براورد 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 براورد این پارامتر تصحیح شده و به‌صورت نسبت $s_{xy}^{**}$ و $s_x^{**۲}$ استفاده می‌شود.

براوردگرهای $T_{r۱}$، $T_{p۱}$، $T_{d۱}$ و $T_{r۲}$ اولین بار توسط خار و سریواستاوا [۳] و [۴] پیش‌نهاد شدند. براوردگر $T_{r۱}$ توسط اکافور و لی [۷] و تاباسام و خان [۱۳] بازنگری شد. بعدها براوردگر $T_{d۱}$ توسط اکافور و لی [۷] و براوردگر $T_{r۲}$ توسط تاباسام و خان [۱۴] مورد مطالعه مجدد قرار گرفت. همچنین براوردگر $T_{d۲}$ اولین بار توسط خار و سریواستاوا [۴] معرفی شد. براوردگرهای $t_d^{(r)}$، $t_d^{(p)}$ ، $t_e$ و $t_{lrd}$ نیز توسط سینگ و کومار [۱۱] پیش‌نهاد شدند که ویژگی‌های آن‌ها را مطالعه کردند.

**دسته‌ی دوم:** این دسته شامل براوردگرهای نسبتی $T_{r۲}$، ضربی $T_{p۲}$ و رگرسیونی $T_{d۲}$ است. این براوردگرها به‌صورت زیر تعریف می‌شوند:

$$T_{r۲} = \frac{\overline{y}_n^{*}}{\overline{x}_n}\overline{x}_{n'} \qquad (۸)$$

$$T_{p۲} = \frac{\overline{y}_n^{*}}{\overline{x}_{n'}}\overline{x}_n \qquad (۹)$$

$$T_{d۲} = \overline{y}_n^{*} + \hat{\beta}_{yx}^{**}\left(\overline{x}_{n'} - \overline{x}_n\right) \qquad (۱۰)$$

## ۴- ویژگی براوردگرهای پیش‌نهادشده

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

$$B\left(T_{r۱}\right) \cong \frac{۱}{\bar{X}_U}\left(\frac{۱}{n}-\frac{۱}{n'}\right)\left(RS_x^۲ - S_{xy}\right) + \frac{W_۲\left(k-۱\right)}{n\,\bar{X}_U}\left(RS_{x۲}^۲ - S_{xy۲}\right)$$

$$B\left(T_{p۱}\right) \cong \left(\frac{۱}{n}-\frac{۱}{n'}\right)\frac{S_{xy}}{\bar{X}_U} + \frac{W_۲\left(k-۱\right)}{n}\frac{S_{xy۲}}{\bar{X}_U}$$

$$B(T_{d۱}) \cong ۰$$

$$B\left(t_d^{(r)}\right) \cong \frac{۱}{\bar{X}_U}\left[\left(\frac{۱}{n}-\frac{۱}{n'}\right)S_x^۲\left(۳R - ۲\beta_{yx}\right) + \frac{W_۲\left(k-۱\right)}{n}S_{x۲}^۲\left(R-\beta_{yx۲}\right)\right]$$

$$B\left(t_d^{(p)}\right) \cong \frac{۱}{\bar{X}_U}\left[\left(\frac{۱}{n}-\frac{۱}{n'}\right)S_x^۲\left(R + ۲\beta_{yx}\right) + \frac{W_۲\left(k-۱\right)}{n}S_{x۲}^۲\,\beta_{yx۲}\right]$$

$$B\left(t_e\right) \cong \frac{۱}{۲\bar{X}_U}\left[\left(\frac{۱}{n}-\frac{۱}{n'}\right)S_x^۲\,\frac{\beta_{yx}}{R}\left(R-\beta_{yx}\right) + \frac{W_۲\left(k-۱\right)}{n}S_{x۲}^۲\,\frac{\beta_{yx۲}}{R}\left(R-\beta_{yx۲}\right)\right]$$

$$B(t_{lrd}) \cong ۰$$

$$B\left(T_{r۲}\right) \cong \left(\frac{۱}{n}-\frac{۱}{n'}\right)\left[\frac{R\,S_x^۲}{\bar{X}_U} - \frac{S_{xy}}{\bar{X}_U}\right]$$

$$B\left(T_{p۲}\right) \cong \left(\frac{۱}{n}-\frac{۱}{n'}\right)\frac{S_{xy}}{\bar{X}_U}$$

$$B(T_{d۲}) \cong ۰$$

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

$$Var\left(T_{r۱}\right) \cong \left(\frac{۱}{n'}-\frac{۱}{N}\right)S_y^۲+\left(\frac{۱}{n}-\frac{۱}{n'}\right)S_r^۲ + \frac{W_۲\left(k-۱\right)}{n}S_{r۲}^۲$$

$$Var\left(T_{p۱}\right) \cong \left(\frac{۱}{n'}-\frac{۱}{N}\right)S_y^۲+\left(\frac{۱}{n}-\frac{۱}{n'}\right)S_p^۲ + \frac{W_۲\left(k-۱\right)}{n}S_{p۲}^۲$$

$$Var\left(T_{d۱}\right) \cong \left(\frac{۱}{n'}-\frac{۱}{N}\right)S_y^۲+\left(\frac{۱}{n}-\frac{۱}{n'}\right)S_y^۲\left(۱-\rho_{yx}^۲\right) + \frac{W_۲\left(k-۱\right)}{n}S_{\beta(۲)}^۲$$

$$Var\left(t_d^{(r)}\right) \cong \left\{\left(\frac{۱}{n}-\frac{۱}{n'}\right)\left[S_y^۲+۴RS_x^۲\left(R-\beta_{yx}\right)\right] + \frac{W_۲\left(k-۱\right)}{n}\left[S_{y۲}^۲+RS_{x۲}^۲\left(R-۲\beta_{yx۲}\right)\right]+\left(\frac{۱}{n'}-\frac{۱}{N}\right)S_y^۲\right\}$$

$$Var\left(t_d^{(p)}\right) \cong \left\{\left(\frac{۱}{n}-\frac{۱}{n'}\right)\left[S_y^۲+۴RS_x^۲\left(R+\beta_{yx}\right)\right] + \frac{W_۲\left(k-۱\right)}{n}\left[S_{y۲}^۲+RS_{x۲}^۲\left(R+۲\beta_{yx۲}\right)\right]+\left(\frac{۱}{n'}-\frac{۱}{N}\right)S_y^۲\right\}$$

$$Var\left(t_e\right) \cong \left[\left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{n}-\frac{1}{n'}\right)S_y^2\left(1-\rho_{yx}^2\right)\right.$$
$$\left.+\frac{W_2\left(k-1\right)}{n}S_{y2}^2\left(1-\rho_{yx2}^2\right)\right]$$

$$Var\left(t_{lrd}\right)=Var\left(t_e\right)$$

$$Var\left(T_{r2}\right) \cong \left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{n}-\frac{1}{n'}\right)S_r^2$$
$$+\frac{W_2\left(k-1\right)}{n}S_{y2}^2$$

$$Var\left(T_{P2}\right) \cong \left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{n}-\frac{1}{n'}\right)S_p^2$$
$$+\frac{W_2\left(k-1\right)}{n}S_{y2}^2$$

$$Var\left(T_{d2}\right) \cong \left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{n}-\frac{1}{n'}\right)S_y^2\left(1-\rho_{yx}^2\right)$$
$$+\frac{W_2\left(k-1\right)}{n}S_{y2}^2$$

که در آن‌ها

$$W_2=\frac{N_2}{N},\ \ R=\frac{\bar{Y}_U}{\bar{X}_U}\ ,\ \ \rho_{yx}=\frac{S_{xy}}{S_x\,S_y},\ \ \beta_{yx2}=\frac{S_{xy2}}{S_{x2}^2},$$

$$\rho_{yx2}=\frac{S_{xy2}}{S_{x2}\,S_{y2}}\ \ ,\ \ S_{y2}^2=\frac{1}{N_2-1}\sum_{i=N_1+1}^{N_1+N_2}\left(Y_i-\bar{Y}_{2U}\right)^2$$

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

$$S_{r۲}^{۲} = \left[S_{y۲}^{۲} + RS_{x۲}^{۲}\left(R - ۲\beta_{yx۲}\right)\right],$$

$$S_{p}^{۲} = \left(S_{y}^{۲} + R^{۲}S_{x}^{۲} + ۲R\rho_{yx}S_{y}S_{x}\right) = \left[S_{y}^{۲} + RS_{x}^{۲}\left(R + ۲\beta_{yx}\right)\right],$$

$$S_{p۲}^{۲} = \left[S_{y۲}^{۲} + RS_{x۲}^{۲}\left(R + ۲\beta_{yx۲}\right)\right],$$

$$S_{\beta(۲)}^{۲} = \left[S_{y۲}^{۲} + \beta_{yx}S_{x۲}^{۲}\left(\beta_{yx} - ۲\beta_{yx۲}\right)\right],$$

$$S_{xy۲} = \frac{۱}{N_{۲} - ۱}\sum_{i=N_{۱}+۱}^{N}\left(X_i - \bar{X}_{۲U}\right)\left(Y_i - \bar{Y}_{۲U}\right),$$

$$S_{x۲}^{۲} = \frac{۱}{N_{۲} - ۱}\sum_{i=N_{۱}+۱}^{N}\left(X_i - \bar{X}_{۲U}\right)^{۲},\ S_{y}^{۲} = \frac{۱}{N - ۱}\sum_{i=۱}^{N}\left(Y_i - \bar{Y}_{U}\right)^{۲}.$$

در این مطالعه $N_{۱}$ نمایانگر واحدهای پاسخ‌گو در جامعه در اولین مراجعه است که در صورت مراجعه به آن‌ها پاسخ‌گو خواهند بود و $N_{۲}$ نمایانگر واحدهایی است که در اولین مراجعه پاسخ‌گو نیستند و فرض می‌شود که با مراجعه مجدد پاسخ‌گو خواهند بود. همچنین $\bar{X}_{۲U}$ و $\bar{Y}_{۲U}$ به‌ترتیب میانگین متغیرهای کمکی و مورد مطالعه از $N_{۲}$ واحد بی‌پاسخ جامعه هستند و به‌صورت زیر تعریف می‌شوند:

$$\bar{X}_{۲U} = \frac{۱}{N_{۲}}\sum_{i=N_{۱}+۱}^{N_{۱}+N_{۲}}X_i \quad , \qquad \bar{Y}_{۲U} = \frac{۱}{N_{۲}}\sum_{i=N_{۱}+۱}^{N_{۱}+N_{۲}}Y_i$$

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

## ۵- مقایسه‌ی واریانس براوردگرها

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

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

**نتیجه‌ی ۱**: براوردگر $t_e$ همواره دقیق‌تر از براوردگر $T_{r۱}$ است. زمانی که $R = \beta_{yx} = \beta_{yx۲}$ واریانس این دو براوردگر یکسان است.

**نتیجه‌ی ۲**: براوردگر $t_e$ همواره دارای واریانسی کم‌تر از براوردگر $T_{P۱}$ است. زمانی که رابطه‌ی $R = -\beta_{yx} = -\beta_{yx۲}$ برقرار باشد آن‌گاه هر دو براوردگر دقت یکسان دارند.

**نتیجه‌ی ۳**: براوردگر $t_e$ همواره دقیق‌تر از براوردگر $T_{d۱}$ است. زمانی که رابطه‌ی $\beta_{yx} = \beta_{yx۲}$ برقرار باشد، دقت این دو براوردگر یکسان است.

**نتیجه‌ی ۴**: براوردگر $t_e$ همواره دقیق‌تر از براوردگر $T_{r۲}$ است. زمانی که رابطه‌های $R = \beta_{yx}$ و $\rho_{yx۲} = ۰$ برقرار باشند، آن‌گاه دقت دو براوردگر یکسان است.

**نتیجه‌ی ۵**: براوردگر $t_e$ همواره دقیق‌تر از براوردگر $T_{P۲}$ است. زمانی واریانس این دو براوردگر یکسان است که رابطه‌های $R = -\beta_{yx}$ و $\rho_{yx۲} = ۰$ برقرار باشند.

**نتیجه‌ی ۶**: براوردگر $t_e$ همواره دقیق‌تر از براوردگر $T_{d۲}$ است. زمانی که رابطه‌ی $\rho_{yx۲} = ۰$ برقرار باشد، دقت دو براوردگر یکسان می‌شود.

**نتیجه‌ی ۷**: براوردگر $t_e$ همواره دقیق‌تر از براوردگر $t_d^{(r)}$ است. اگر رابطه‌ی $R = \beta_{yx} /۲ = \beta_{yx۲}$ برقرار باشد، آن‌گاه دقت دو براوردگر یکسان است.

**نتیجه‌ی ۸**: براوردگر $t_e$ همواره دقیق‌تر از براوردگر $t_d^{(p)}$ است، به جز وقتی که رابطه‌ی $R = -\beta_{yx} /۲ = -\beta_{yx۲}$ برقرار باشد.

**نتیجه‌ی ۹**: براوردگر $t_d^{(r)}$ زمانی دقیق‌تر از براوردگر $T_{r۱}$ است که رابطه‌های $\beta_{yx} > \frac{۳}{۲}R$ برای $R$ مثبت یا $\beta_{yx} < \frac{۳}{۲}R$ برای $R$ منفی برقرار باشند.

**نتیجـه‌ی ۱۰**: براوردگـر $t_d^{(r)}$ زمـانی دقیـق‌تر از براوردگـر $T_{r۲}$ اسـت کـه رابطـه‌های $\beta_{yx} > \frac{۳}{۲}R$ و $\beta_{yx۲} > \frac{R}{۲}$ بـرای $R$ مثبـت یـا رابطـه‌های $\beta_{yx} < \frac{۳}{۲}R$ و $\beta_{yx۲} < \frac{R}{۲}$ برای $R$ منفی برقرار باشند.

**نتیجـه‌ی ۱۱**: براوردگـر $t_d^{(p)}$ زمـانی دقیـق‌تر از براوردگـر $T_{P۱}$ اسـت کـه رابطـه‌ی $\beta_{yx} < -\frac{۳}{۲}R$ برای $R$ مثبت یا رابطه‌ی $\beta_{yx} > -\frac{۳}{۲}R$ برای $R$ منفی برقرار باشد.

**نتیجـه‌ی ۱۲**: براوردگـر $t_d^{(p)}$ زمـانی دقیـق‌تر از براوردگـر $T_{P۲}$ اسـت کـه رابطـه‌های $\beta_{yx} < -\frac{۳}{۲}R$ و $\beta_{yx۲} < -\frac{R}{۲}$ بـرای $R$ مثبـت یـا رابطـه‌های $\beta_{yx} > -\frac{۳}{۲}R$ و $\beta_{yx۲} > -\frac{R}{۲}$ برای $R$ منفی برقرار باشند.

همان‌طور که در بخش چهارم اشاره شد، دو براوردگر $t_{lrd}$ و $t_e$ که توسط سینگ و کومار [۱۱] معرفـی شـدند، دارای واریـانس یکسـانی هسـتند. از مقایسـه‌ی واریـانس آن‌هـا بـا واریانس دیگر براوردگرها با در نظر گرفتن اطلاعات قبلی درمورد جامعـه، نتیجه می‌شـود که واریانس آن‌ها برابـر یـا کـم‌تـر از براوردگرهـای $T_{r۱}$، $T_{P۱}$، $T_{d۱}$، $T_{r۲}$، $T_{P۲}$، $T_{d۲}$، $t_d^{(r)}$ و $t_d^{(p)}$ است. در مقایسه‌ی دیگر براوردگرها با یکدیگر اطلاعـات قبلـی بیش‌تـری از جامعه، برای تعیین بهترین براوردگر موردنیاز است.

## ۶- مطالعه‌ی شبیه‌سازی

در این بخـش براوردگرهـای نسـبتی $T_{r۱}$ و $T_{r۲}$، براوردگرهـای رگرسـیونی $T_{d۱}$ و $T_{d۲}$، کـه توسط خـار و سریواسـتاوا [۳]، [۴] معرفـی شـده‌اند و براوردگرهـای $t_d^{(r)}$، $t_e$ و $t_{lrd}$ کـه توسط سینگ و کومار [۱۱] پیش‌نهاد شده‌اند با انجام مطالعه‌ی شبیه‌سازی با هـم مقایسـه می‌شـوند. شـاخص ارزیـابی براوردگرهـای پیش‌نهـادی میـانگین توان‌هـای دوم خطـای شبیه‌سازی‌شده است.

برای این منظور با اسـتفاده از نرم‌افـزار R داده‌هـای مربـوط بـه دو متغیـر قـد (در مقیـاس سانتی‌متر) و وزن (در مقیاس کیلوگرم) از توزیع نرمـال دومتغیـری بـا همبسـتگی معـین و

مثبتی تولید می‌شوند که آن‌ها را به‌عنوان مقدارهای جامعه‌ی مورد مطالعه در نظر می‌گیریم.
در این مطالعه‌ی شبیه‌سازی، هدف، تولید جامعه‌ای با جمعیت $N = ۱۰۰,۰۰۰$ نفر با دو متغیر $X$ و $Y$ و برآورد پارامترهای توزیع نمونه‌گیری براوردگرها 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 $\rho_{yx}$ بین متغیرها، از توزیع نرمال دومتغیری زیر برای تولید داده‌های جامعه‌ی آماری مورد نظر استفاده می‌شود.

$$\left(X, Y\right) \sim N_{۲}\left(\mu = \begin{bmatrix} ۶۲ \\ ۱۶۳ \end{bmatrix}, \Sigma = \begin{bmatrix} ۲/۶۵ & \rho_{yx} \times ۲/۶۵ \times ۳/۳۲ \\ \rho_{yx} \times ۲/۶۵ \times ۳/۳۲ & ۳/۳۲ \end{bmatrix}\right)$$

که در آن $\rho_{yx}$ همبستگی بین متغیرهای $Y$ و $X$ 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 $n' = ۱۰,۰۰۰$ از جامعه آن را براورد می‌کنیم.

همچنین در این مطالعه‌ی شبیه‌سازی (با استفاده از $K = ۱۰۰۰$ مرتبه تکرار نمونه‌گیری خودگردان به‌صورت مستقل از هم) تأثیر سه اندازه‌ی نمونه متفاوت $n = ۵۰۰, ۱۰۰۰, ۲۰۰۰$، سه همبستگی متفاوت $\rho_{yx} = ۰/۵۵, ۰/۷۰, ۰/۸۵$ و سه نرخ بی‌پاسخی ۰/۲۵، ۰/۳۵ و ۰/۴۵ را بر روی براورد میانگین توان دوم خطای شبیه‌سازی براوردگرها بررسی می‌کنیم.

میانگین‌ها و واریانس‌های نمونه را به‌ترتیب برای تقریب امید ریاضی و واریانس براوردگرها به کار می‌بریم و میانگین توان دوم خطای شبیه‌سازی براوردگرهای $T_{r۱}$، $T_{r۲}$، $T_{d۱}$، $T_{d۲}$، $t_d^{(r)}$، $t_e$ و $t_{lrd}$ را با هم مقایسه می‌کنیم. فرض کنید $t$ یکی از

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

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

در واقع $\bar{t}$ و $Var(t)$ براوردی از میانگین و واریانس توزیع نمونه‌گیری $t$ هستند. همچنین براورد اریبی برای براوردگر $t$ به‌صورت $B(t)=\left(\bar{t}-\bar{Y}_U\right)$ است که در آن $\bar{Y}_U$ میانگین واقعی جامعه برای متغیر مورد مطالعه‌ی $Y$ است. میانگین توان دوم خطای شبیه‌سازی‌شده ($SMSE$) در براوردگر $t$ نیز از رابطه‌ی زیر به دست می‌آید:

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

برای هر اندازه‌ی نمونه و هر نرخ بی‌پاسخی و ضریب همبستگی بین $X$ و $Y$ برنامه را ۱۰۰۰ مرتبه اجرا کردیم. مقدار $SMSE$ براوردها 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

| نرخ بی‌پاسخی | ۰/۴۵ | | | ۰/۳۵ | | | ۰/۲۵ | | |
|---|---|---|---|---|---|---|---|---|---|
| اندازه‌ی نمونه | ۵۰۰ | ۱۰۰۰ | ۲۰۰۰ | ۵۰۰ | ۱۰۰۰ | ۲۰۰۰ | ۵۰۰ | ۱۰۰۰ | ۲۰۰۰ |
| $SMSE(T_{r1})$ | ۰/۰۱۷ | ۰/۰۰۸۵ | ۰/۰۰۴ | ۰/۰۱۴۹ | ۰/۰۰۷۷ | ۰/۰۰۴۱ | ۰/۰۱۳۷ | ۰/۰۰۶۲ | ۰/۰۰۲۹ |
| $SMSE(T_{d1})$ | ۰/۰۱۷ | ۰/۰۰۸۵ | ۰/۰۰۴ | ۰/۰۱۴۹ | ۰/۰۰۷۷ | ۰/۰۰۴۱ | ۰/۰۱۳۷ | ۰/۰۰۶۲ | ۰/۰۰۲۹ |
| $SMSE(T_{r2})$ | ۰/۰۲۰ | ۰/۰۱۰۳ | ۰/۰۰۵ | ۰/۰۱۶۵ | ۰/۰۰۸۹ | ۰/۰۰۴۶ | ۰/۰۱۵ | ۰/۰۰۶۸ | ۰/۰۰۳۲ |
| $SMSE(T_{d2})$ | ۰/۰۲۰ | ۰/۰۱۰۳ | ۰/۰۰۵ | ۰/۰۱۶۵ | ۰/۰۰۸۹ | ۰/۰۰۴۶ | ۰/۰۱۵ | ۰/۰۰۶۸ | ۰/۰۰۳۲ |
| $SMSE(t_{dr})$ | ۰/۰۱۹۵ | ۰/۰۰۹۸ | ۰/۰۰۴۴ | ۰/۰۱۶۹ | ۰/۰۰۸۵ | ۰/۰۰۵۸ | ۰/۰۱۶۲ | ۰/۰۰۷۷ | ۰/۰۰۳۳ |
| $SMSE(t_{lrd})$ | ۰/۰۱۸۶ | ۰/۰۰۹۱ | ۰/۰۰۴۲ | ۰/۰۱۶۴ | ۰/۰۰۸۳ | ۰/۰۰۴۴ | ۰/۰۱۴۴ | ۰/۰۰۶۵ | ۰/۰۰۳۱ |
| $SMSE(t_e)$ | ۰/۰۱۸۶ | ۰/۰۰۹۱ | ۰/۰۰۴۲ | ۰/۰۱۶۴ | ۰/۰۰۸۳ | ۰/۰۰۴۴ | ۰/۰۱۴۴ | ۰/۰۰۶۵ | ۰/۰۰۳۱ |

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

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

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

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

## 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 براوردگر، $SMSE$ براوردگرها 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 $SMSE$ براوردگرها کاهش می‌یابد.

۳. برای هر مقدار از ضریب همبستگی بین دو متغیر $Y$ و $X$ با افزایش اندازه‌ی نمونه و کاهش نرخ بی‌پاسخی، اغلب *SMSE* براوردگرهای $T_{r1}$، $T_{r2}$، $T_{d1}$، $T_{d2}$، $t_{dr}$، $t_e$ و $t_{lrd}$ کاهش می‌یابد.

٤. *SMSE* دو براوردگر $t_e$, $t_{lrd}$ تقریباً با هم برابر است. همچنین *SMSE* دو براوردگر $T_{r1}$ و $T_{d1}$ و نیز *SMSE* دو براوردگر $T_{r2}$ و $T_{d2}$ نیز تقریباً با هم برابر هستند.

۵. از بین براوردگرهای $T_{r1}$، $T_{r2}$، $T_{d1}$، $T_{d2}$، $t_d^{(r)}$، $t_e$ و $t_{lrd}$ برای نرخ‌های پاسخ‌گویی، همبستگی و اندازه‌های نمونه‌ی متفاوت، مقدار *SMSE* براوردگر $t_d^{(r)}$ برای هر اندازه‌ی نمونه و هر نرخ بی‌پاسخی، برای همبستگی ۰/۸۵ از همه کم‌تر است.

۶. برای ضریب همبستگی متوسط (۰/۵۵و ۰/۷۰) مقدار *SMSE* دو براوردگر $T_{r1}$ و $T_{d1}$ تقریباً نزدیک به هم است. این دو براوردگر از مابقی براوردگرها بهتر عمل می‌کنند و بهتر است برای براورد میانگین جامعه‌ای از آن‌ها استفاده شود. در حالی که برای ضریب همبستگی بالا (۰/۸۵) *SMSE* براوردگر $t_d^{(r)}$ از مابقی براوردگرها کم‌تر است و این براوردگر بهتر از بقیه عمل می‌کند.

## ۷- نتیجه‌گیری

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

اندازه‌های نمونه، نرخ بی‌پاسخی و همبستگی‌های متفاوت برای ۱۰۰۰ نمونه‌ی خودگردان محاسبه شد. مشاهده شد که برای ضریب‌های همبستگی ۰/۵۵و 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 ۰/۸۵، $SMSE$ 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 ...) 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 برآوردگرهای ۱ تا ۱۰ که در رابطه‌های ۱۱ تا ۳۰ آمده‌اند از عبارت‌های زیر استفاده می‌شود. برای این منظور ابتدا قرار دهید:

$$\bar{y}_n^* = \bar{Y}_U + \left(\bar{y}_n^* - \bar{Y}_U\right) = \bar{Y}_U + \varepsilon_\circ^* , \qquad \varepsilon_\circ^* = \bar{y}_n^* - \bar{Y}_U$$

$$\bar{x}_n^* = \bar{X}_U + \left(\bar{x}_n^* - \bar{X}_U\right) = \bar{X}_U + \varepsilon_1^* , \qquad \varepsilon_1^* = \bar{x}_n^* - \bar{X}_U$$

$$\bar{x}_{n'} = \bar{X}_U + \left(\bar{x}_{n'} - \bar{X}_U\right) = \bar{X}_U + \varepsilon_1' , \qquad \varepsilon_1' = \bar{x}_{n'} - \bar{X}_U$$

$$\bar{x}_n = \bar{X}_U + \left(\bar{x}_n - \bar{X}_U\right) = \bar{X}_U + \varepsilon_1 , \qquad \varepsilon_1 = \bar{x}_n - \bar{X}_U$$

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

**لم ۱-** با توجه به رابطه‌های بالا داریم

الف-
$$E\left(\varepsilon_\circ^*\right) = E\left(\varepsilon_1^*\right) = E\left(\varepsilon_1'\right) = E\left(\varepsilon_1\right) = 0$$

ب-

$$Var\left(\varepsilon_\circ^*\right) = \left(\frac{1-f}{n}\right) S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2$$

$$Var\left(\varepsilon_{۱}^{*}\right)=\left(\frac{۱-f}{n}\right)S_{x}^{۲}+\frac{W_{۲}(k-۱)}{n}S_{x^{۲}}^{۲}$$

$$Var\left(\varepsilon_{۱}'\right)=\left(\frac{۱-f'}{n'}\right)S_{x}^{۲},\qquad f'=\frac{n'}{N}$$

$$Var\left(\varepsilon_{۱}\right)=\left(\frac{۱-f}{n}\right)S_{x}^{۲}$$

$$E\left(\varepsilon_{۰}^{*}\varepsilon_{۱}^{*}\right)=\left(\frac{۱-f}{n}\right)S_{xy}+\frac{W_{۲}(k-۱)}{n}S_{xy^{۲}}$$

$$E\left(\varepsilon_{۰}^{*}\varepsilon_{۱}\right)=\left(\frac{۱-f}{n}\right)S_{xy}$$

$$E\left(\varepsilon_{۰}^{*}\varepsilon_{۱}'\right)=\left(\frac{۱-f'}{n'}\right)S_{xy}$$

$$E\left(\varepsilon_{۱}^{*}\varepsilon_{۱}\right)=\left(\frac{۱-f}{n}\right)S_{x}^{۲}$$

$$E\left(\varepsilon_{۱}\ \varepsilon_{۱}^{*}\right)=\left(\frac{۱-f'}{n'}\right)S_{x}^{۲}$$

$$E\left(\varepsilon_{۱}^{*}\varepsilon_{۱}'\right)=\left(\frac{۱-f'}{n'}\right)S_{x}^{۲}$$

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

**مریم تقربی**
فوق لیسانس آمار
تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه‌طباطبایی، گروه آمار.
رایانشانی: maryamtagharobi@yahoo.com

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

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