معرفی یک برآوردگرجدید برای برآورد میانگین جامعه با وجودخطاهای اندازه‌گیری و عدم پاسخ

لیدر نوایی؛ دانشگاه پیام نور، ایران
رسول ایماز؛ پیام نور تبریز

دریافت ۹۵/۱۰/۱۳ پذیرش ۹۶/۱۰/۱۷

چکیده

بر اساس نظریهٔ نمونه‌گیری کلاسیک، خطاهایی که به‌طور عمده در برآوردها بررسی شده‌اند، خطاهای نمونه‌گیری است. اگر چه، اغلب خطاهای غیر نمونه‌گیری در خواص برآوردگرها مؤثرتر از خطاهای نمونه‌گیری هستند. این امر را پژوهش‌گران در طول دو دههٔ اخیر تأیید کرده است، خصوصاً در رابطه با خطاهای عدم‌پاسخ که یکی از اساسی‌ترین خطاهای غیرنمونه‌گیری است. بیش‌تر پژوهش‌های انجام گرفته، خطاهای غیرنمونه‌گیری را به‌صورت جداگانه بررسی کرده‌اند، هر چند معمولاً در حالت واقعی دو یا چند خطای غیرنمونه‌گیری به‌صورت هم‌زمان رخ می‌دهند. در این مقاله دو نوع از خطاهای غیرنمونه‌گیری در مرحلهٔ برآورد ملاحظه می‌شود: خطای اندازه‌گیری و خطای عدم‌پاسخ. یک برآوردگر جدید برای برآورد میانگین جامعه با وجود خطاهای اندازه‌گیری و عدم‌پاسخ پیشنهاد شده است. از لحاظ نظری و تجربی نشان داده شده است که برآوردگر پیشنهاد شده از برآوردگر نااریب هنسن و هورویتس و دیگر برآوردگرهای موجود کارایی بیش‌تری دارد.

واژه‌های کلیدی: میانگین جامعه، برآوردگر نمایی، میانگین مربعات خطا، خطای اندازه‌گیری، عدم پاسخ، مراجعهٔ مجدد.

مقدمه

در نظریهٔ نمونه‌گیری دو نوع خطای غیرنمونه‌گیری وجود دارد: خطای اندازه‌گیری و خطای عدم‌پاسخ. خطای اندازه‌گیری (پاسخ) وقتی روی می‌دهد که مقدار گزارش شده با مقدار واقعی تفاوت دارد. خطای عدم پاسخ وقتی روی می‌دهد که پژوهش‌گر نمی‌تواند اطلاعات یک یا بیش‌تر از یک واحد بررسی شده را گردآوری کند. خطای اندازه‌گیری ممکن است ناشی از بیش‌گزارش، کم‌گزارش، ضعف حافظهٔ پاسخ‌گویان و مانند آن هنگام جمع‌آوری داده‌ها باشد. مسئلهٔ عدم پاسخ وقتی روی می‌دهد که پژوهش‌گر با توجه به دلایلی مانند عدم دسترسی به واحدهای جامعه، خودداری از جواب دادن پرسش‌نامه، فقدان اطلاعات و غیره موفق به جمع‌آوری اطلاعاتِ برخی از واحدها نمی‌شود.

مسئلهٔ برآورد میانگین جامعه با وجود خطاهای اندازه‌گیری و عدم‌پاسخ به‌صورت جداگانه را تعدادی پژوهش‌گر بررسی کرده‌اند. شالب [12] برآوردگر نسبتی کلاسیک میانگین جامعه را با وجود خطای اندازه‌گیری بررسی کرد. مانیشا و سینگ [10] اثر خطای اندازه‌گیری را روی یک برآوردگر جدید بررسی کردند که یک ترکیب خطی از نسبت و میانگین هر واحد از برآوردگرها بود. علاوه براین، مسئلهٔ خطای اندازه‌گیری را فولر [4]، سینگ و کارپ [14]، کومار و دیگران [7]، شوکلا و دیگران [13] بررسی کرده‌اند. هنسن- هورویتس [5] برای اولین بار مسئلهٔ عدم پاسخ را بررسی کردند و روش مراجعهٔ مجدد را برای گردآوری اطلاعات از زیرنمونه‌ای از واحدهای بی‌پاسخ پیشنهاد کردند.

*نویسنده مسئول leadernavaei@yahoo.com

کجران [3] بعضی برآوردگرهای جدید میانگین جامعه را با استفاده از اطلاعاتِی که متغیر کمکی تنها برای موقعیت‌هایی که برخی از عدم‌پاسخ‌ها در بررسی روی می‌دهند، پیشنهاد کرد. هم‌چنین مسئله عدم‌پاسخ را رائو [11]، خاری و سریواستاوا [6] ، سینگ و کومار [15]، کومار و سونیل [8] و مانند آن را بررسی کرده‌اند.

در عمل، پژوهش‌گر اغلب با موقعیت‌هایی مواجه می‌شود که خطای اندازه‌گیری و عدم پاسخ هنگام گردآوری اطلاعات به‌طور هم‌زمان رخ می‌دهند. مسئلهٔ برآورد میانگین جامعه با وجود خطاهای اندازه‌گیری و عدم‌پاسخ به‌صورت هم‌زمان به‌وسیلهٔ عظیم و حنیف [10] و کومار و دیگران [9] بررسی شده است. هدف این مقاله پیشنهاد یک برآوردگر جدید میانگین جامعه با استفاده از اطلاعاتی است که متغیر کمکی با وجود خطای اندازه‌گیری و عدم‌پاسخ است. یک برآوردگر جدید برای میانگین جامعه معرفی شده است و پژوهشی تطبیقی با برآوردگرهای هنسن- هورویتس [5]، کجران [2] و سینگ و کومار [15] انجام گرفته است.

نمادگذاری

فرض کنید که Y و X به‌ترتیب متغیر بررسی شده و متغیر کمکی باشند. هم‌چنین فرض کنید که جامعه از دو طبقه تشکیل شده است: N_1 واحد پاسخگو و N_2 واحد بی‌پاسخ. یک نمونهٔ تصادفی ساده به حجم n از جامعه‌ای به حجم N به‌روش نمونه‌گیری تصادفی ساده بدون جای‌گذاری انتخاب شده است. فرض کنید n_1 تا پاسخ دهند و n_2 تا پاسخ ندهند $(n_1 + n_2 = n)$. با استفاده از روش مراجعهٔ مجدد یک زیرنمونه به حجم m_2 از n_2 واحد بی‌پاسخ انتخاب می‌شود. فرض کنید که (x_i^*, y_i^*) مقادیر مشاهده شده و (X_i^*, Y_i^*) مقادیر واقعی دو صفت (x, y) برای iامین واحد نمونه $(i = 1,2, \dots, n)$ باشند. فرض کنید که خطاهای اندازه‌گیری (۱) باشند:

$$u_i^* = y_i^* - Y_i^* \text{ و } v_i^* = x_i^* - X_i^* \tag{۱}$$

که ذاتاً تصادفی و ناهمبسته با میانگین صفر و واریانس S_U^2 و S_V^2 به‌ترتیب برای طبقهٔ پاسخگوی جامعه هستند. فرض کنید که $S_{U(2)}^2$ و $S_{V(2)}^2$ به‌ترتیب واریانس خطاهای اندازه‌گیری متغیرهای Y و X طبقهٔ بی‌پاسخ جامعه باشند. فرض کنید که S_Y^2 و S_X^2 به‌ترتیب واریانس‌های جامعه‌ای Y و X برای طبقهٔ پاسخگو و $S_{Y(2)}^2$ و $S_{X(2)}^2$ به‌ترتیب واریانس‌های جامعه‌ای Y و X طبقهٔ بی‌پاسخ جامعه باشند. فرض کنید که $\bar{Y}$ و $\bar{X}$ به‌ترتیب میانگین‌های جامعه‌ای و $\bar{y}$ و $\bar{x}$ به‌ترتیب میانگین‌های نمونه‌ای Y و X باشند. فرض کنید که ρ_{XY} ضریب هم‌بستگی بین متغیرهای Y و X طبقهٔ پاسخگو و $\rho_{XY(2)}$ ضریب هم‌بستگی بین متغیرهای Y و X طبقهٔ بی‌پاسخ جامعه باشند. فرض کنید که C_Y و C_X به‌ترتیب ضریب تغییر برای متغیرهای Y و X طبقهٔ پاسخگو و $C_{Y(2)}$ و $C_{X(2)}$ به‌ترتیب ضریب تغییر برای متغیرهای Y و X طبقهٔ بی‌پاسخ جامعه باشند. علاوه بر این فرض کنید که میانگین جامعه‌ای متغیر بررسی شده $(\bar{Y})$ نامعلوم و میانگین جامعه‌ای متغیر کمکی $(\bar{X})$ معلوم است.

برآوردگر پیشنهاد شده

هنگامی که خطای اندازه‌گیری و عدم پاسخ به‌طور هم‌زمان رخ می‌دهند، برآوردگر نسبتی-نمایی

$$t_p = \bar{y}^* \left(\frac{\bar{x}^{*\prime}}{\bar{X}}\right)\left[\alpha\, exp\left(\frac{\bar{X}-\bar{x}^*}{\bar{X}+\bar{x}^*}\right) + (1-\alpha) exp\left(\frac{\bar{x}^{*\prime}-\bar{X}}{\bar{x}^{*\prime}+\bar{X}}\right)\right] \tag{۲}$$

برای برآورد میانگین جامعه پیشنهاد می‌شود، به‌طوری‌که α یک عدد ثابت منتخب مناسب است، و

$$\bar{x}^{*\prime} = \frac{N\bar{X} - n\bar{x}^*}{N-n}\ . \tag{۳}$$

برای به‌دست آوردن میانگین مربع خطای برآوردگر پیشنهاد شده، تعدادی علائم دیگر را معرفی می‌کنیم. فرض کنید که

$$\omega_Y^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i^* - \bar{Y}), \qquad \omega_U^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i^* \tag{۴}$$

$$\omega_X^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i^* - \bar{X}), \qquad \omega_V^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i^* . \tag{۵}$$

از جمع بستن روابط (۴) پس از ساده کردن، داریم:

$$\bar{y}^* = \bar{Y} + \frac{1}{\sqrt{n}}\,(\omega_Y^* + \omega_U^*) \tag{۶}$$

به‌طور مشابه، از جمع بستن روابط (۵) پس از ساده کردن، داریم:

$$\bar{x}^* = \bar{X} + \frac{1}{\sqrt{n}}\,(\omega_X^* + \omega_V^*). \tag{۷}$$

علاوه بر این، با فرض

$$\begin{cases} \Delta_Y^* = \bar{y}^* - \bar{Y} = \frac{1}{\sqrt{n}}\,(\omega_Y^* + \omega_U^*) \\ \Delta_X^* = \bar{x}^* - \bar{X} = \frac{1}{\sqrt{n}}\,(\omega_X^* + \omega_V^*) \end{cases} \tag{۸}$$

داریم:

$$E(\Delta_Y^*) = E(\Delta_X^*) = 0 \tag{۹}$$

زیرا $\bar{y}^*$ برآوردگر نااریب $\bar{Y}$ و $\bar{x}^*$ برآوردگر نااریب $\bar{X}$ است و

$$\begin{cases} E({\Delta_Y^*}^2) = Var(\bar{y}^*) = \lambda_1\bar{Y}^2\left(C_Y^2 + \frac{S_U^2}{\bar{Y}^2}\right) + \lambda_2\bar{Y}^2\left(C_{Y(2)}^2 + \frac{S_{U(2)}^2}{\bar{Y}^2}\right) \\ E({\Delta_X^*}^2) = Var(\bar{x}^*) = \lambda_1\bar{X}^2\left(C_X^2 + \frac{S_V^2}{\bar{X}^2}\right) + \lambda_2\bar{X}^2\left(C_{X(2)}^2 + \frac{S_{V(2)}^2}{\bar{X}^2}\right) \\ E(\Delta_Y^*\Delta_X^*) = Cov(\bar{y}^*, \bar{x}^*) = \bar{Y}\bar{X}\{\lambda_1\rho_{XY}C_YC_X + \lambda_2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\}. \end{cases} \tag{۱۰}$$

از جای‌گذاری (۳) در (۲) برآوردگر پیشنهاد شده را می‌توان بدین‌صورت نوشت:

$$t_p = \bar{y}^*\left(\frac{N\bar{X} - n\bar{x}^*}{(N-n)\bar{X}}\right)\left[\alpha\, exp\left(\frac{\bar{X} - \bar{x}^*}{\bar{X} + \bar{x}^*}\right) + (1-\alpha)exp\left(\frac{-n(\bar{x}^* - \bar{X})}{(2N-n)\bar{X} - n\bar{x}^*}\right)\right] \tag{۱۱}$$

حال از جای‌گذاری (۸) در (۱۱) داریم:

$$t_p = (\bar{Y} + \Delta_Y^*)\left(\frac{(N-n)\bar{X} - n\Delta_X^*}{(N-n)\bar{X}}\right)\left[\alpha\, exp\left(\left(\frac{-\Delta_X^*}{2\bar{X}}\right)\frac{1}{1 + \frac{\Delta_X^*}{2\bar{X}}}\right)\right.$$
$$\left. + (1-\alpha)exp\left(\left(\frac{-n\Delta_X^*}{2(N-n)\bar{X}}\right)\frac{1}{1 - \frac{n\Delta_X^*}{2(N-n)\bar{X}}}\right)\right]$$

با در نظر گرفتن دو جملهٔ اول بسط $exp(.)$ داریم:

$$t_p=(\bar{Y}+\Delta_Y^*)\left(1-\frac{n\Delta_X^*}{(N-n)\bar{X}}\right)\left[\alpha\left\{1+\left(\left(\frac{-\Delta_X^*}{2\bar{X}}\right)\frac{1}{1+\frac{\Delta_X^*}{2\bar{X}}}\right)\right\}\right.$$
$$\left.+(1-\alpha)\left\{1+\left(\left(\frac{-n\Delta_X^*}{2(N-n)\bar{X}}\right)\frac{1}{1-\frac{n\Delta_X^*}{2(N-n)\bar{X}}}\right)\right\}\right]$$

با فرض این‌که $1>\left|\frac{\Delta_X^*}{2\bar{X}}\right|$ و $1>\left|\frac{n\Delta_X^*}{2(N-n)\bar{X}}\right|$ می‌توانیم عبارت‌های $\left(\frac{1}{1+\frac{\Delta_X^*}{2\bar{X}}}\right)$ و $\left(\frac{1}{1-\frac{n\Delta_X^*}{2(N-n)\bar{X}}}\right)$ را بسط دهیم. از این رو، با در نظر گرفتن جمله‌های تا مرتبهٔ دوم داریم:

$$t_p\cong(\bar{Y}+\Delta_Y^*)\left(1-\frac{n\Delta_X^*}{(N-n)\bar{X}}\right)\left[\alpha\left\{1+\left(\left(\frac{-\Delta_X^*}{2\bar{X}}\right)\left\{1-\frac{\Delta_X^*}{2\bar{X}}\right\}\right)+\frac{1}{2}\left(\left(\frac{-\Delta_X^*}{2\bar{X}}\right)\left\{1-\frac{\Delta_X^*}{2\bar{X}}\right\}\right)^2\right\}\right.$$
$$+(1-\alpha)\left\{1+\left(\left(\frac{-n\Delta_X^*}{2(N-n)\bar{X}}\right)\left\{1+\frac{n\Delta_X^*}{2(N-n)\bar{X}}\right\}\right)\right.$$
$$\left.\left.+\frac{1}{2}\left(\left(\frac{-n\Delta_X^*}{2(N-n)\bar{X}}\right)\left\{1+\frac{n\Delta_X^*}{2(N-n)\bar{X}}\right\}\right)^2\right\}\right]$$

پس از ساده کردن و نادیده گرفتن جملات با مرتبه‌های بالاتر از دو، داریم:

$$t_p\cong\bar{Y}+\Delta_Y^*-\frac{1}{2}\left\{\frac{3n}{N-n}+\alpha\left(1-\frac{n}{N-n}\right)\right\}\frac{\bar{Y}}{\bar{X}}\Delta_X^*-\frac{1}{2\bar{X}}\left\{\frac{3n}{N-n}+\alpha\left(1-\frac{n}{N-n}\right)\right\}\Delta_Y^*\Delta_X^*$$
$$+\left\{\frac{n\bar{Y}}{2(N-n)\bar{X}}\left\{\alpha+(1-\alpha)\left(\frac{n}{N-n}\right)\right\}\right.$$
$$\left.-\frac{\bar{Y}}{8\bar{X}^2}\left\{3\alpha+(1-\alpha)\left(\frac{n}{N-n}\right)^2\right\}\right\}{\Delta_X^*}^2$$

حال با فرض $\alpha_\circ=\frac{3n}{N-n}+\left(1-\frac{n}{N-n}\right)\alpha$ داریم:

$$t_p-\bar{Y}\cong\Delta_Y^*-\frac{1}{2}\alpha_\circ\frac{\bar{Y}}{\bar{X}}\Delta_X^*-\frac{1}{2\bar{X}}\alpha_\circ\Delta_Y^*\Delta_X^*$$
$$+\left[\frac{n\bar{Y}}{2(N-n)\bar{X}}\left\{\alpha+(1-\alpha)\left(\frac{n}{N-n}\right)\right\}-\frac{\bar{Y}}{8\bar{X}^2}\left\{3\alpha+(1-\alpha)\left(\frac{n}{N-n}\right)^2\right\}\right]{\Delta_X^*}^2$$

با مجذور کردن دو طرف رابطهٔ اخیر و نادیده گرفتن جملات با مرتبه‌های بالاتر از دو، داریم:

$$\left(t_p-\bar{Y}\right)^2={\Delta_Y^*}^2+\frac{1}{4}\alpha_\circ^2\frac{\bar{Y}^2}{\bar{X}^2}{\Delta_X^*}^2-\alpha_\circ\frac{\bar{Y}}{\bar{X}}\Delta_Y^*\Delta_X^*$$

بنابراین،

$$MSE(t_p)\cong E\left(t_p-\bar{Y}\right)^2=E\left({\Delta_Y^*}^2\right)+\frac{1}{4}\alpha_\circ^2\frac{\bar{Y}^2}{\bar{X}^2}E\left({\Delta_X^*}^2\right)-\alpha_\circ\frac{\bar{Y}}{\bar{X}}E(\Delta_Y^*\Delta_X^*)$$

از جای‌گذاری (۱۰) در آخرین تساوی داریم:

$$\begin{aligned}MSE(t_p) &= \lambda_1\bar{Y}^2\left(C_Y^2+\frac{1}{4}\alpha_\circ^2C_X^2-\alpha_\circ\rho_{XY}C_YC_X\right)\\ &+\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+\frac{1}{4}\alpha_\circ^2C_{X(2)}^2-\alpha_\circ\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)\\ &+\lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+\frac{1}{4}\alpha_\circ^2\frac{S_V^2}{\bar{X}^2}\right]+\\ &\lambda_2\bar{Y}^2\left[\frac{S_{U(2)}^2}{\bar{Y}^2}+\frac{1}{4}\alpha_\circ^2\frac{S_{V(2)}^2}{\bar{X}^2}\right]\end{aligned} \tag{۱۲}$$

با مشتق‌گیری از (۱۲) نسبت به $\alpha_\circ$ و برابر صفر قرار دادن آن، مقدار بهینهٔ $\alpha_\circ$ برابر است با

$$\alpha_\circ^* = 2\frac{\lambda_1\rho_{XY}C_YC_X+\lambda_2\rho_{XY(2)}C_{Y(2)}C_{X(2)}}{\lambda_1\left(\frac{S_X^2+S_V^2}{\bar{X}^2}\right)+\lambda_2\left(\frac{S_{X(2)}^2+S_{V(2)}^2}{\bar{X}^2}\right)}. \tag{۱۳}$$

مقایسهٔ کارایی برآوردگر پیشنهاد شده با برخی از برآوردگرهای موجود

دراین بخش، شرایطی را به‌دست می‌آوریم که تحت این شرایط برآوردگر پیشنهاد شده (t_p) از برآوردگر نااریب هنسن - هورویتس [5]، برآوردگر کجران [3] و برآوردگر سینگ و کومار [15] کاراتراست.

۱. مقایسهٔ کارایی برآوردگر پیشنهاد شده با برآوردگر هانسن و هرویتز

هانسن و هرویتز [5] هنگامی که عدم‌پاسخ رخ می‌دهد، برآوردگر نااریب

$$\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_{m_2} \tag{۱۴}$$

را پیشنهاد کردند، که واریانس آن برابر

$$Var(\bar{y}^*) = \lambda_1S_Y^2 + \lambda_2S_{Y(2)}^2$$

است. اگر خطای اندازه‌گیری در نظر گرفته شود، واریانس برآوردگر $\bar{y}^*$برابر است با:

$$Var^*(\bar{y}^*) = \lambda_1S_Y^2 + \lambda_2S_{Y(2)}^2 + \lambda_1S_U^2 + \lambda_2S_{U(2)}^2\,. \tag{۱۵}$$

برآوردگر t_p از برآوردگر $\bar{y}^*$ کاراتر است اگر

$$Var^*(\bar{y}^*) - MSE\,(t_p) \geq 0. \tag{۱۶}$$

از جای‌گذاری (۱۲) و (۱۵) در (۱۶) داریم:

$$\begin{aligned}\lambda_1S_Y^2+\lambda_2S_{Y(2)}^2+\lambda_1S_U^2+\lambda_2S_{U(2)}^2-\lambda_1\bar{Y}^2\left(C_Y^2+\frac{1}{4}\alpha_\circ^2C_X^2-\alpha_\circ\rho_{XY}C_YC_X\right)\\ -\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+\frac{1}{4}\alpha_\circ^2C_{X(2)}^2-\alpha_\circ\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)\\ -\lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+\frac{1}{4}\alpha_\circ^2\frac{S_V^2}{\bar{X}^2}\right]-\lambda_2\bar{Y}^2\left[\frac{S_{U(2)}^2}{\bar{Y}^2}+\frac{1}{4}\alpha_\circ^2\frac{S_{V(2)}^2}{\bar{X}^2}\right]\geq 0\end{aligned}$$

یا اگر

$$\lambda_1\bar{Y}^2\left(\frac{\alpha_\circ^2}{4}\frac{S_X^2+S_V^2}{\bar{X}^2}-\alpha_\circ\rho_{XY}C_YC_X\right)+\lambda_2\bar{Y}^2\left(\frac{\alpha_\circ^2}{4}\frac{S_{X(2)}^2+S_{V(2)}^2}{\bar{X}^2}-\alpha_\circ\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)\leq 0. \tag{۱۷}$$

نامساوی (۱۷) برقرار است اگر

$$\begin{cases}\frac{1}{4}\alpha_\circ^2\frac{S_X^2+S_V^2}{\bar{X}^2}-\alpha_\circ\rho_{XY}C_YC_X\le 0 \\ \\ \frac{1}{4}\alpha_\circ^2\frac{S_{X(2)}^2+S_{V(2)}^2}{\bar{X}^2}-\alpha_\circ\rho_{XY(2)}C_{Y(2)}C_{X(2)}\le 0\ .\end{cases} \quad \text{و} \qquad (۱۸)$$

بنابراین از (۱۸) نتیجه می‌گیریم که برآوردگر t_p از برآوردگر $\bar{y}^*$ کاراتر است اگر

$$\begin{cases}\rho_{XY}\ge\frac{1}{4}\alpha_\circ\frac{C_X}{C_Y}\left(1+\frac{S_V^2}{S_X^2}\right) \\ \\ \rho_{XY(2)}\ge\frac{1}{4}\alpha_\circ\frac{C_{X(2)}}{C_{Y(2)}}\left(1+\frac{S_{V(2)}^2}{S_{X(2)}^2}\right).\end{cases} \quad \text{و} \qquad (۱۹)$$

۲. مقایسهٔ کارایی برآوردگر پیشنهاد شده با برآوردگر کجران

کجران[3] هنگامی که عدم‌پاسخ رخ می‌دهد، برآوردگر نسبتی

$$t_C=\bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right) \qquad (۲۰)$$

را پیشنهاد کرد که میانگین مربع خطای آن برابر

$$MSE(t_C)\cong\lambda_1\bar{Y}^2(C_Y^2+C_X^2-2\rho_{XY}C_YC_X)+\lambda_2\bar{Y}^2(C_{Y(2)}^2+C_{X(2)}^2-2\rho_{XY(2)}C_{Y(2)}C_{X(2)})$$

است. اگر خطای اندازه‌گیری در نظر گرفته شود، میانگین مربع خطای برآوردگر t_C برابر است با:

$$MSE^*(t_C)\cong\lambda_1\bar{Y}^2(C_Y^2+C_X^2-2\rho_{XY}C_YC_X)+\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+C_{X(2)}^2-2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)$$
$$+\ \lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+\frac{S_V^2}{\bar{X}^2}\right]+\lambda_2\bar{Y}^2\left[\frac{S_{U(2)}^2}{\bar{Y}^2}+\frac{S_{V(2)}^2}{\bar{X}^2}\right]. \qquad (۲۱)$$

برآوردگر t_p از برآوردگر t_C کاراتر است اگر

$$MSE^*(t_C)-MSE\left(t_p\right)\ge 0\ . \qquad (۲۲)$$

از جای‌گذاری (۱۶) و (۲۱) در (۲۲) داریم:

$$\lambda_1\bar{Y}^2(C_Y^2+C_X^2-2\rho_{XY}C_YC_X)+\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+C_{X(2)}^2-2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)+\lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+\frac{S_V^2}{\bar{X}^2}\right]$$
$$+\lambda_2\bar{Y}^2\left[\frac{S_{U(2)}^2}{\bar{Y}^2}+\frac{S_{V(2)}^2}{\bar{X}^2}\right]-\lambda_1\bar{Y}^2\left(C_Y^2+\frac{1}{4}\alpha_\circ^2C_X^2-\alpha_\circ\rho_{XY}C_YC_X\right)$$
$$-\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+\frac{1}{4}\alpha_\circ^2C_{X(2)}^2-\alpha_\circ\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)-\lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+\frac{1}{4}\alpha_\circ^2\frac{S_V^2}{\bar{X}^2}\right]$$
$$-\lambda_2\bar{Y}^2\left[\frac{S_{U(2)}^2}{\bar{Y}^2}+\frac{1}{4}\alpha_\circ^2\frac{S_{V(2)}^2}{\bar{X}^2}\right]\ge\ 0$$

یا اگر

$$\lambda_1\bar{Y}^2\left[\left\{1-\frac{1}{4}\alpha_\circ^2\right\}\frac{S_X^2+S_V^2}{\bar{X}^2}-2\rho_{XY}C_YC_X\left\{1-\frac{1}{2}\alpha_\circ\right\}\right]+\lambda_2\bar{Y}^2\left[\left\{1-\frac{1}{4}\alpha_\circ^2\right\}\frac{S_{X(2)}^2+S_{V(2)}^2}{\bar{X}^2}-\right.$$
$$2\rho XY2CY2CX21-12\alpha^\circ\ge 0\ . \qquad (۲۳)$$

نامساوی (۲۳) برقرار است اگر

$$\begin{cases} \left\{1-\frac{1}{4}\alpha_{\circ}^{2}\right\}\frac{S_X^2+S_V^2}{\bar{X}^2}-2\rho_{XY}C_YC_X\left\{1-\frac{1}{2}\alpha_{\circ}\right\}\geq 0 \\ \text{و} \\ \left\{1-\frac{1}{4}\alpha_{\circ}^{2}\right\}\frac{S_{X(2)}^2+S_{V(2)}^2}{\bar{X}^2}-2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\left\{1-\frac{1}{2}\alpha_{\circ}\right\}\geq 0\ . \end{cases} \tag{۲۴}$$

بنابراین از (۲۴) نتیجه می‌گیریم که برآوردگر t_p از برآوردگر t_C کاراتر است اگر

$$\begin{cases} \rho_{XY}\leq\frac{1}{2}\left\{1+\frac{1}{2}\alpha_{\circ}\right\}\frac{C_X}{C_Y}\left(1+\frac{S_V^2}{S_X^2}\right) \\ \text{و} \\ \rho_{XY(2)}\leq\frac{1}{2}\left\{1+\frac{1}{2}\alpha_{\circ}\right\}\frac{C_{X(2)}}{C_{Y(2)}}\left(1+\frac{S_{V(2)}^2}{S_{X(2)}^2}\right). \end{cases} \tag{۲۵}$$

۳. مقایسهٔ کارایی برآوردگر پیشنهاد شده بابرآوردگر سینگ و کومار

سینگ و کومار [15] برآوردگر زنجیروار نسبتی

$$t_{SK}=\bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right)\left(\frac{\bar{X}}{\bar{x}}\right) \tag{۲۶}$$

را برای برآورد میانگین پیشنهاد کردند، به‌طوری‌که میانگین مربع خطای t_{SK} علیرغم وجود عدم‌پاسخ و بدون خطای اندازه‌گیری برابر

$$MSE(t_{SK})\cong\lambda_1\bar{Y}^2(C_Y^2+4C_X^2-4\rho_{XY}C_YC_X)+\lambda_2\bar{Y}^2(C_{Y(2)}^2+C_{X(2)}^2-2\rho_{XY(2)}C_{Y(2)}C_{X(2)})$$

است. اگر خطای اندازه‌گیری در نظر گرفته شود، میانگین مربع خطای t_{SK}برابر

$$\begin{aligned} &MSE^*(t_{SK})\cong \\ &\lambda_1\bar{Y}^2(C_Y^2+4C_X^2-4\rho_{XY}C_YC_X)+\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+C_{X(2)}^2-2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)+ \\ &\lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+4\frac{S_V^2}{\bar{X}^2}\right])+\lambda_2\bar{Y}^2\left[\frac{S_{U(2)}^2}{\bar{Y}^2}+\frac{S_{V(2)}^2}{\bar{X}^2}\right]. \end{aligned} \tag{۲۷}$$

است.

برآوردگر t_p از برآوردگر t_{SK}کاراتر است اگر

$$MSE^*(t_{SK})-MSE(t_p)\geq 0\ . \tag{۲۸}$$

با استفاده از (۱۶) و (۲۷) در (۲۸) داریم:

$$\begin{aligned} &\lambda_1\bar{Y}^2(C_Y^2+4C_X^2-4\rho_{XY}C_YC_X)+\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+C_{X(2)}^2-2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right) \\ &\quad+\lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+4\frac{S_V^2}{\bar{X}^2}\right]+\lambda_2\bar{Y}^2\left[\frac{S_{U(2)}^2}{\bar{Y}^2}+\frac{S_{V(2)}^2}{\bar{X}^2}\right] \\ &\quad-\lambda_1\bar{Y}^2\left(C_Y^2+\frac{1}{4}\alpha_{\circ}^2C_X^2-\alpha_{\circ}\rho_{XY}C_YC_X\right) \\ &\quad-\lambda_2\bar{Y}^2\left(C_{Y(2)}^2+\frac{1}{4}\alpha_{\circ}^2C_{X(2)}^2-\alpha_{\circ}\rho_{XY(2)}C_{Y(2)}C_{X(2)}\right)-\lambda_1\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+\frac{1}{4}\alpha_{\circ}^2\frac{S_V^2}{\bar{X}^2}\right] \\ &\quad-\lambda_2\bar{Y}^2\left[\frac{S_U^2}{\bar{Y}^2}+\frac{1}{4}\alpha_{\circ}^2\frac{S_V^2}{\bar{X}^2}\right]\geq 0 \end{aligned}$$

یا اگر

$$\begin{aligned} &\lambda_1\bar{Y}^2\left[4\left\{1-\frac{1}{16}\alpha_{\circ}^2\right\}\frac{S_X^2+S_V^2}{\bar{X}^2}-4\rho_{XY}C_YC_X\left\{1-\frac{1}{4}\alpha_{\circ}\right\}\right]+\lambda_2\bar{Y}^2\left[\left\{1-\frac{1}{4}\alpha_{\circ}^2\right\}\frac{S_{X(2)}^2+S_{V(2)}^2}{\bar{X}^2}-\right. \\ &\left.2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\left\{1-\frac{1}{2}\alpha_{\circ}\right\}\right]\geq 0\ . \end{aligned} \tag{۲۹}$$

نامساوی (۲۹) برقرار است اگر

$$\begin{cases} 4\left\{1-\frac{1}{16}\alpha_{\circ}^{2}\right\}\frac{S_X^2+S_V^2}{\bar{X}^2}-4\rho_{XY}C_Y C_X\left\{1-\frac{1}{4}\alpha_{\circ}\right\}\geq 0 \\ \text{و} \\ \left\{1-\frac{1}{4}\alpha_{\circ}^{2}\right\}\frac{S_{X(2)}^2+S_{V(2)}^2}{\bar{X}^2}-2\rho_{XY(2)}C_{Y(2)}C_{X(2)}\left\{1-\frac{1}{2}\alpha_{\circ}\right\}\geq 0\ . \end{cases} \tag{۳۰}$$

بنابراین از (۳۰) نتیجه می‌گیریم که برآوردگر t_p از برآوردگر t_{SK} کاراتر است اگر

$$\begin{cases} \rho_{XY}\leq\left\{1+\frac{1}{4}\alpha_{\circ}\right\}\frac{C_X}{C_Y}\left(1+\frac{S_V^2}{S_X^2}\right) \\ \text{و} \\ \rho_{XY(2)}\leq\frac{1}{2}\left\{1+\frac{1}{2}\alpha_{\circ}\right\}\frac{C_{X(2)}}{C_{Y(2)}}\left(1+\frac{S_{V(2)}^2}{S_{X(2)}^2}\right). \end{cases} \tag{۳۱}$$

مقایسهٔ عدد یو شبیه‌سازی

برای مقایسهٔ عددی برآوردگر پیشنهاد شده t_p^* (برآوردگر پیشنهاد شده به‌ازای $\alpha_{\circ}^*$) با برآوردگر هنسن- هورویتس $(\bar{y}^*)$، برآوردگرکجران (t_C) و برآوردگر سینگوکومار(t_{SK}) یک جامعهٔ فرضی و شبیه‌سازی‌شده را در نظر می‌گیریم. هنسن- هورویتس، کجران، سینگوکومار هنگام پیشنهاد برآوردگرهایشان خطای اندازه‌گیری را در نظر نگرفته‌اند. برای مقایسهٔ برآوردگر پیشنهاد شده، MSE برآوردگرها محاسبه شده است. ۶ جامعه شبیه‌سازی شده‌اند به‌طوری‌که ضریب هم‌بستگی آن‌ها پایین، بالا، مثبت و منفی است. فرض شده است که حجم هر جامعه ۵۰۰۰، حجم طبقهٔ پاسخ‌گو۳۷۵۰ و حجم نمونه ۵۰۰ و میزان عدم‌پاسخ ۰/۳ است. متغیرکمکی واقعی مربوط به جامعهٔ$X\sim N(10,2)$ و متغیرکمکی اندازه‌گیری شده، $x=X+N(10,2)$ فرض شده است، پس$V=x-X$. متغیربررسی شده با مدل خطی $Y=2+bX+N(0,1)$ شبیه‌سازی شده است. برای کنترل هم‌بستگی بین متغیر بررسی شده و متغیر کمکی، مقدار b تغییر کرده است. مقادیر مفروض b عبارتند از: ۰/۱، ۰/۳، ۰/۵، ۰/۱-، ۰/۳- و ۰/۵-. همانند متغیر کمکی، متغیر بررسی شده اندازه‌گیری شده با $y=Y+N(0,1)$ تعمیم داده شده است، از این رو، $U=y-Y$.پارامترهای جامعه در جدول ۱ خلاصه شده است.

جدول ۱. پارامترهای جامعه

پارامترها	جامعه ۱	جامعه ۲	جامعه ۳	جامعه ۴	جامعه ۵	جامعه ۶
$\bar{X}$	۳/۰۰۶۷۱۳	۵/۰۱۱۱۷۵	۶/۹۷۸۸۳۸۴	۱/۰۰۷۳۹۱	-۰/۹۷۲۸	-۲/۹۵۵۶۵
$\bar{Y}$	۱۰/۰۶۶۶۸	۱۰/۰۱۲۴۳	۹/۹۸۱۷۱۶	۹/۹۵۳۷۲۲	۹/۹۵۰۷۶۴	۹/۹۴۷۱۱۴
S_Y^2	۱/۰۵۷۳۲۲	۱/۳۸۴۶۳۵	۲/۰۸۶۰۲۴	۱/۰۳۱۰۷۸	۱/۳۳۱۷۳۶	۱/۹۶۸۲۷۷
S_X^2	۴/۰۲۱۱۵	۴/۰۶۵۱۸۷	۴/۱۶۰۳۳۲	۴/۰۶۲۹۶۹	۴/۰۶۲۷۲۵	۴/۰۲۹۰۱۱
S_U^2	۰/۹۸۴۷۲۹	۰/۹۵۳۵۷۳	۰/۹۹۰۷۶۷	۱/۰۴۴۵۸۴	۱/۰۰۳۷۱۴	۰/۹۹۹۶۶۷
S_V^2	۰/۹۹۲۰۹۴	۰/۹۹۱۹۹۹	۰/۹۷۷۴۸۶	۱/۰۴۰۳۳۱	۱/۰۱۲۲۷۱	۱/۰۳۸۵۲۱
ρ_{YX}	۰/۱۶۳۲۰۴	۰/۵۲۱۶۳۵	۰/۷۱۲۸۳۴	-۰/۱۸۲۷۵	-۰/۵۰۱۷۵	-۰/۷۱۵۵۹
$S_{Y(2)}^2$	۰/۹۸۹۲۰۶	۱/۳۲۳۸۱۳	۲/۱۰۳۰۵۱	۱/۰۲۲۱۹۸	۱/۳۱۷۷۵۴	۲/۰۱۵۷۹۷
$S_{X(2)}^2$	۴/۰۳۰۲۰۷	۳/۹۷۸۴۹۱	۴/۰۴۲۹۹۴	۳/۹۸۲۵۵۷	۴/۱۷۸۵۹۷	۳/۹۳۷۳۳
$S_{U(2)}^2$	۱/۰۳۷۳۵۵	۰/۹۳۲۵۷۷	۰/۹۵۵۶۹۴	۱/۰۱۵۹۷۵	۱/۰۱۹۹۷۵	۰/۹۳۴۰۶۹
$S_{V(2)}^2$	۰/۹۹۶۱۳۹	۰/۹۶۴۷۸۶	۰/۹۲۵۲۰۳	۰/۹۷۰۸۴	۱/۰۴۴۴۲	۰/۹۷۹۲۵۷
$\rho_{YX(2)}$	۰/۱۲۳۷۵۸	۰/۵۲۲۸۶۹	۰/۷۰۳۷۶۶	-۰/۱۶۷۰۵	-۰/۵۱۷۹۲	-۰/۷۱۸۷۳

جدول ۲، میانگین مربع خطا (MSE) برآوردگرها و درصد کارایی نسبی (PRE) برآوردگرکجران، برآوردگر سینگ و کومار و برآوردگر پیشنهاد شده نسبت به برآوردگر نااریب هانسن و هورویتس را در بر می‌گیرد. جدول ۲ نشان می‌دهد که میانگین مربع خطای برآوردگر پیشنهاد شده در مقایسه با میانگین مربع خطای برآوردگرهای دیگر کم‌تر است. با توجه به درصد کارایی نسبی داده شده در جدول ۲، می‌توان نتیجه گرفت که در هر شش جامعه برآوردگر پیشنهاد شده از سه برآوردگر دیگر بهتر است.

جدول ۲. میانگین مربع خطا و درصد کارایی نسبی برآوردگرها با وجود خطای اندازه‌گیری و عدم‌پاسخ

پارامترها	جامعه ۱	جامعه ۲	جامعه ۳	جامعه ۴	جامعه ۵	جامعه ۶
	۰/۰۰۴۸۹۲	۰/۰۰۵۵۶۳	۰/۰۰۷۳۷۳	۰/۰۰۴۹۵۹	۰/۰۰۵۶۰۶	۰/۰۰۷۱۱۲
$MSE^*(t_C)$	۰/۰۰۵۵۱۵	۰/۰۰۵۶۳۵	۰/۰۰۶۳۴۴	۰/۰۰۵۲۶۱	۰/۰۰۵۱۷۰	۰/۰۰۵۳۰۰
$MSE^*(t_{SK})$	۰/۰۰۷۵۶۸	۰/۰۱۰۲۴۶	۰/۰۱۴۶۱۹	۰/۰۰۵۶۷۹	۰/۰۰۵۰۲۲	۰/۰۰۵۵۶۱
$MSE^*(t_p)$	۰/۰۰۴۸۴۴	۰/۰۰۴۸۴۳	۰/۰۰۵۳۲۰	۰/۰۰۴۸۹۶	۰/۰۰۴۹۵۴	۰/۰۰۵۱۷۰
$PRE(t_C)$	۸۹	۹۹	۱۱۶	۹۴	۱۰۸	۱۲۴
$PRE(t_{SK})$	۶۵	۵۴	۵۰	۸۷	۱۱۲	۱۲۸
$PRE(t_p)$	۱۰۱	۱۱۵	۱۳۹	۱۰۱	۱۱۳	۱۳۸

منابع

1. Azeem M., Hanif M., "On estimation of population mean in the presence of measurement error and nonresponse", Pak. J. Statist, 31(5) (2015) 657-670.
2. Cochran W. G., "Errors of measurement in statistics", Technometrics, 10(4) (1968) 637-666.
3. Cochran W. G., "Sampling Techniques (3rd ed.)", John Wiley and Sons, New York (1977).
4. Fuller W. A., "Measurement Error Models", Wiley, New York (1987).
5. Hansen M. H., Hurwitz W. N., "The problem of non-response in sample surveys", Journal of the American Statistical Association, 41(236) (1946) 517-529.
6. Khare B. B., Srivastava S., "Estimation of population mean using auxiliary character inpresence of non-response", The National Academy of Sciences, Letters, India,16 (1993) 111-114.
7. Kumar M., Singh R., Singh A.K., Smarandache, F., "Some ratio type estimators under measurement errors", World Applid Sciences Journal, 14 (2) (2011) 272-276.
8. Kumar S., Bhougal S., "Estimation of the population mean in presence of non-response", Communications of the Korean Statistical Society, 18 (4) (2011) 537-548.
9. Kumar S., Bhougal S., Nataraja N. S., Viswanathaiah M., "Estimation of population mean in the presence of non-response and measurement error,Revista Colombiana de Estadística", 38 (1) (2015) 145-161.
10. Manisha, Singh R., "An estimation of population mean in the presence of measurement errors", Journal of the Indian Society of Agricultural Statistics, 54(1) (2001) 13-18.

11. Rao P. S. R. S., "Ratio estimation with subsampling the nonrespondents", Survey Methodology, 12(2) (1986) 217-230.
12. Shalabh G., "Ratio method of estimation in the presence of measurement errors", Journal of Indian Society of Agricultural Statistics, 50(2) (1997)150-155.
13. Shukla D., Pathak S., Thakur N. S., "An estimator for mean estimation in presence of measurement error", Research and Reviews: A Journal of Statistics,1(1) (2012) 1-8.
14. Singh H., Karpe N., "Estimation of mean, ratio and product using auxiliary information in the presence of measurement errors in sample surveys", Journal of Statistical Theory and Practice, 4 (2010) 111-136.
15. Singh H. P., Kumar S., "Estimation of mean in presence of non-response using two phase sampling scheme", Statistical Papers, 50 (2008) 559-582.