مجله علوم آماری، بهار و تابستان ۱۳۹۱
جلد ۶، شماره ۱، ص ۳۹-۵۴

برآورد واریانس برآوردگر کالیبره مجموع جامعه با مجموع جامعه کمکی نامعلوم

ابراهیم خدایی[1]، سیدروح‌اله شجاعی‌کیاسری[2]
[1] دانشگاه تهران، سازمان سنجش آموزش کشور
[2] بانک مرکزی جمهوری اسلامی ایران

تاریخ دریافت: ۱۳۹۰/۱۰/۳ تاریخ آخرین بازنگری: ۱۳۹۱/۵/۱۶

چکیده : به کارگیری متغیرهای کمکی برای اصلاح برآوردگرها روشی متداول در آمار و به خصوص بررسی‌های نمونه‌ای است. به عنوان مثال می‌توان به برآوردگرهای نسبتی و رگرسیونی در نمونه‌گیری اشاره کرد. با فرض معلوم بودن مجموع جامعه کمکی و برقرار بودن یک‌سری شرایط، برآوردگرهای کالیبره و واریانس آنها را می‌توان با استفاده از برآوردگرهای رگرسیونی تعمیم‌یافته به دست آورد. در این مقاله، با فرض نامعلوم بودن مجموع متغیر کمکی در جامعه، برآوردگر کل و واریانس آن در جامعه هدف با روش رگرسیون تعمیم‌یافته به دست آورده می‌شود. سپس نشان داده می‌شود برآوردگر ارائه شده برای مجموع کل جامعه از نظر کارایی بهتر از برآوردگر هورویتز-تامپسون است. آنگاه نظریه به دست آمده با شبیه‌سازی به روش MCMC در طرح نمونه‌گیری طبقه‌بندی مورد بررسی قرار می‌گیرد.

واژه‌های کلیدی : آ ماره کمکی، برآوردگر رگرسیونی تعمیم‌یافته، برآوردگر کالیبره، نمونه‌گیری طبقه‌بندی، روش‌های مونت کارلوی زنجیر مارکوفی.

آدرس الکترونیک مسئول مقاله: ابراهیم خدایی، khodaie@yahoo.com
کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۲D۰۵

۱ مقدمه

فرض کنید هدف انجام نمونه‌گیری از جامعه‌ای متناهی باشد. شمول عضو kام جامعه در نمونه تصادفی S، یک واقعه تصادفی است که می‌توان آن را بر اساس متغیر تصادفی نشانگر به صورت

$$I_k = \begin{cases} ۱ & k \in S \\ ۰ & k \notin S \end{cases}$$

نشان داد. منظور از $k \in S$ واقعه‌ای تصادفی است که نمونه تصادفی S شامل عضو kام جامعه باشد. به این ترتیب احتمال شمول[۱] عضو kام جامعه در نمونه تصادفی S به صورت $\pi_k = P(I_k = ۱) = P(k \in S)$ تعریف می‌شود. به همین ترتیب احتمال شمول توأم دو عضو k و ℓ در نمونه تصادفی S را به صورت $\pi_{k\ell} = Pr(k\&\ell \in S)$ نشان داده می‌شود. این احتمال‌های شمول با توجه به طرح نمونه‌گیری تعیین می‌شوند. همچنین معکوس احتمال شمول، $d_k = \frac{۱}{\pi_k}$، وزن نمونه‌ای نامیده می‌شود.

براساس نظریه کالیبراسیون، وزن‌های نمونه‌ای با داشتن مجموع کل جامعه کمکی کالیبره می‌شوند. اولین نگاه جدی به کالیبره کردن وزن‌ها به عنوان یک روش برآورد کردن توسط دویل و سارن‌دال (۱۹۹۲) انجام گرفت، که نشان دادند کلاس وسیعی از برآوردگرهای کالیبره با طرح‌های مختلف نمونه‌ای سازگار هستند. به‌علاوه تحت شرایطی به‌صورت مجانبی معادل برآوردگرهای روش رگرسیون تعمیم‌یافته می‌باشند.

فرض اساسی در برآورد کردن پارامتر جامعه (مجموع کل جامعه هدف)، با روش رگرسیون تعمیم‌یافته، معلوم بودن مجموع کل متغیر یا متغیرهای کمکی است. اگر این مجموع‌ها نامعلوم باشند، ممکن است ابتدا با استفاده از یک نمونه مقدماتی از مقادیر جامعه کمکی مقدار مجموع نامعلوم جامعه کمکی را برآورد نموده، سپس با زیر نمونه‌ای از آن، که شامل مقادیر توام متغیر هدف و کمکی است، با روش رگرسیونی مقدار نامعلوم مجموع جامعه هدف را برآورد کرد. سیتر (۱۹۹۷) با این روش که نمونه‌گیری مجدد نام دارد، برآوردگری را ارائه داد و واریانس آن را نیز به چند صورت نشان داد. با استفاده از ایدهٔ باز نمونه‌گیری محققانی چون رائو و سیتر (۱۹۹۵)، سیتر (۱۹۹۷)، بیندر و همکاران (۲۰۰۶) و کیم و همکاران (۲۰۰۶) پارامتر جامعه را با استفاده از اطلاعات کمکی برآوردهای رگرسیونی و نسبتی برآورد نموده و واریانس‌هایی برای آنها ارائه داده‌اند. در این مقاله فرض بر این است که نمونه‌های $S_۱$ و S، به ترتیب، به حجم $n_۱$ و n در دست هستند که $S_۱$ فقط شامل مقادیر

۱ Inclusion Probability

متغیرهای کمکی و S شامل مقادیر توأمی از متغیر هدف و متغیرهای کمکی هستند. برای توجیه استفاده از نمونه S_1 به جای نمونه S در برآورد مجموع نامعلوم جامعه کمکی کمترین فرض این است که حجم S_1 از حجم S بزرگتر است، یعنی $n_1 > n$. با این نقطه‌نظر، برآوردگری برای مجموع کل جامعه هدف و فرمولی برای واریانس مجانبی آن ارائه و با سایر روش‌ها مقایسه می‌شود. برآوردگرهای به‌دست آمده در این مقاله نسبت به برآوردگرهای رگرسیون تعمیم‌یافته با فرض معلوم بودن جامعه کمکی کمتر کارا بوده ولی نسبت به برآوردگر هوررویتز_تامپسون[۲] از کارایی بهتری برخوردار خواهند بود.

در بخش ۲ مقاله نظریه کالیبراسیون و روش رگرسیون تعمیم‌یافته مطرح می‌شود. در بخش ۳ برآوردگر مجموع کل جامعه هدف با فرض نامعلوم بودن مجموع متغیر یا متغیرهای کمکی به همراه واریانس آن ارائه می‌شود. در بخش ۴ صحت نظریه‌های ارائه شده با شبیه‌سازی مورد بررسی قرار می‌گیرد. در بخش پایانی بحث و نتیجه‌گیری ارائه خواهد شد.

۲ کالیبره و برآورد رگرسیون تعمیم‌یافته

شرح کامل برآوردگرهای کالیبره توسط دویل و سارن‌دال (۱۹۹۲) ارائه شده است. در این روش برآوردگر کالیبره $\hat{t}_{yw} = \sum_S w_k y_k$ برای برآورد مجموع کل جامعه، $t_y = \sum_U y_k$، طوری ارائه می‌شود که وزن های w_k با توجه به قید

$$\sum_S w_k x_k = \sum_U x_k$$

به وزن‌های طرح نمونه‌ای، $d_k = \frac{1}{\pi_k}$، تا حد ممکن نزدیک باشند. فرض کنید U جامعه متناهی $\{(\mathbf{x}_k, y_k), k = 1, \ldots, N\}$ و S نمونه‌ای تصادفی به حجم n از جامعه U و π_k احتمال شمول عضو kام جامعه در نمونه است. در این روش یک تابع فاصله دلخواه $G_k(w,d)$، تحت شرایط، به منظور حداقل کردن فاصله وزن های w_k به وزن‌های طرح نمونه‌ای d_k، تعریف می‌شود.

شرط ۱: برای هر مقدار ثابت $d > 0$، تابع $G_k(w,d)$ نامنفی، نسبت به w مشتق‌پذیر، اکیداً محدب، روی بازه $D_k(d)$، که شامل d است، تعریف شده باشد و داشته باشیم $G_k(d,d) = 0$.

[۲] Horitz-Thompson Estimator

شرط ۲: مشتق جزئی $g_k(w,d) = \partial G_k(w,d)/\partial w$ پیوسته باشد و $D_k(d)$ را روی بازهٔ $Im_k(d)$ به صورت یک به یک تصویر کند، یعنی $g_k(w,d)$ تابعی اکیداً صعودی نسبت به w است و $g_k(d,d) = ۰$.

بدیهی است که برای خانواده توابع فاصله‌ای $G_k(w,d)$، که در دو شرط ۱ و ۲ صدق کنند، خانواده‌ای از برآوردگرهای کالیبره پدید می‌آید که هر عضو آن، برآوردگر کالیبره متناظر با یک تابع فاصله از خانواده توابع فاصله $G_k(w,d)$ است. در این میان برآوردگر کالیبره‌ای که توسط تابع فاصله

$$G_k(w,d) = \sum_S \frac{(w_k - d_k)^۲}{d_k q_k}$$

پدید می‌آید، همان برآوردگر رگرسیونی تعمیم‌یافته است.

۱.۲ برآوردگر رگرسیونی تعمیم‌یافته

بنابر تعریف سارن دال و همکاران (۱۹۹۲)، بردار متغیرهای کمکی $\mathbf{x}$ را در نظر بگیرید که مقدار آن برای kامین واحد جامعه $\mathbf{x}_k = (x_{۱k}, \ldots, x_{jk})'$ است. هدف برآورد مقدار نامعلوم کل جامعه، $t_y = \sum_U y_k$، با استفاده از اطلاعات بردار متغیر کمکی $\mathbf{x}$ است. با فرض آن که مقدار کل جامعه کمکی، یعنی بردار $\mathbf{t}_x = \sum_U \mathbf{X}_k$، معلوم است، نمونه S با طرح نمونه‌گیری $p(\cdot)$ از جامعه U استخراج می‌شود، که در آن $p(S)$ احتمال برآمد نمونه S در نمونه‌گیری از جامعه U است. در این مقاله صرفاً طرح‌های نمونه‌گیری اندازه‌پذیر مورد نظر است که احتمال‌های شمول $\pi_k = P(k \in S)$ و $\pi_{k\ell} = P(k\&\ell \in S)$، اکیداً بزرگتر از صفر هستند. به این ترتیب برآوردگر رگرسیونی تعمیم‌یافته برای مقدار کل t_y عبارت است از

$$\hat{t}_{yr} = \hat{t}_{y\pi} + \sum_{j=۱}^{J} \hat{B}_j (t_{xj} - \hat{t}_{x_j\pi}) \tag{۱}$$

که در آن

$$\hat{t}_{y\pi} = \sum_S \frac{y_k}{\pi_k} = \sum_S d_k y_k = \sum_S \breve{y}_k$$

برآوردگر هورویتز-تامپسون $t_y = \sum_U y_k$ و

$$\hat{t}_{x_j\pi} \sum_S \frac{x_{jk}}{\pi_k} = \sum_S d_k x_{jk} = \sum_S \breve{x}_{jk}$$

به‌ازای هر $j = 1, \ldots, J$ و برآوردگر هورویتز–تامپسون مجموع (معلوم) جامعه X_j، $t_{sj} = \sum_U x_{jk}$، و $\hat{B}_j$ها مؤلفه‌های بردار

$$\hat{B} = (\hat{B}_1, \ldots, \hat{B}_J)' = (\sum_S \frac{x_k x'_k}{\sigma_k^2 \pi_k})^{-1} \sum_S \frac{x_k y_k}{\sigma_k^2 \pi_k} \qquad (2)$$

هستند، که در آن σ_k^2 واریانس متغیر تصادفی Y_k با برآمد y_k است. برآوردگر رگرسیونی تعمیم‌یافته که خود عضوی از خانواده برآوردگرهای کالیبره است، در وضعیتی که حجم جامعه و نمونه بزرگ باشند و تحت شرایطی خاص، که در ادامه ذکر خواهد شد، به نوعی نقطه همگرایی همه اعضای خانواده برآوردگرهای کالیبره است (دویل و سارن‌دال، ۱۹۹۲).

فرض کنید حجم جامعه N و به همراه آن حجم نمونه n به تدریج بزرگ شوند و برای هر بردار از مقادیر متغیر کمکی $\mathbf{x}$ شرایط زیر برقرار باشند.

۱) حد $N^{-1}\mathbf{t}_x$ موجود باشد.

۲) $N^{-1}(\hat{\mathbf{t}}_x - \mathbf{t}_x)$ در احتمال به صفر میل کند.

۳) $n^{\frac{1}{2}} N^{-1}(\hat{t}_x - t_x)$ در توزیع به نرمال چند متغیره $\mathcal{N}(\mathbf{o}, \mathbf{A})$ میل کند.

در این صورت برآوردگر کالیبره $\hat{t}_{yw}$، به طور مجانبی با برآوردگر رگرسیونی $\hat{t}_{yr}$ که در رابطهٔ (۱) معرفی شد، هم‌ارز است (به این مفهوم: $N^{-1}(\hat{t}_{yw} - \hat{t}_{yr}) = O_p(n^{-1})$). به عنوان یک نتیجه، دو برآوردگر، واریانس‌های مجانبی (AV)[3] یکسان خواهند داشت. امّا فرم خانوادگی برآوردگر رگرسیونی تعمیم‌یافته به عنوان یک برآوردگر کالیبره عبارت است از $\hat{t}_{yr} = \sum_S \frac{g_{ks}}{\pi_k} y_k$، که در آن $\hat{\mathbf{T}} = \sum_S \frac{\mathbf{x}_k \mathbf{x}'_k}{\sigma_k^2 \pi_k}$ و

$$g_{ks} = 1 + (t_x - \hat{t}_{x\pi})' \hat{T}^{-1} x_k / \sigma_k^2 \qquad (3)$$

۳ برآوردگر رگرسیونی تعمیم‌یافته و کالیبره مجموع جامعه

در این بخش وضعیتی بررسی می‌شود که آماره کمکی مناسبی برای برآورد بهتر مجموع جامعه اصلی Y، t_y، در دست است، امّا مجموع این جامعه، $\mathbf{t}_x$، نامعلوم است. در اینصورت برآورد پارامتر هدف، معمولاً با روش نمونه‌گیری مجدد انجام می‌شود. در اینجا به مسایلی پرداخته می‌شود که عملاً نمونه‌ای مناسب برای متغیرهای کمکی در دست است که برای

---
[3] Asymptotic Variance

برآورد مجموع جامعه کمکی، نسبت به نمونه‌ای که محقق برای برآورد مجموع جامعه هدف، از جامعه کمکی و هدف تهیه می‌کند بهتر است. فرض بر این است که نمونه‌های S_1 و S با حجم‌های n_1 و n در اختیارند که S_1 تنها شامل مقادیر متغیرهای کمکی و S شامل مقادیر توأمی از متغیرهای هدف و کمکی هستند. ضمناً برای توجیه استفاده از نمونه S_1 به جای نمونه S در برآورد مجموع جامعه کمکی، کمترین فرض این است که $n_1 > n$.

تعریف ۱ : برآوردگر رگرسیونی تعمیم‌یافته برای مجموع جامعه اصلی Y، t_y، وقتی که مجموع جامعه کمکی $\mathbf{X}$، $\mathbf{t_X}$، نامعلوم است به صورت

$$\hat{t}_{y\pi 1} = \hat{t}_{y\pi} + (\hat{t}_{x\pi 1} - \hat{t}_{x\pi})'\hat{B} \qquad (۴)$$

است، که در آن $\hat{\mathbf{t}}_{x\pi 1} = \sum_{S_1} \mathbf{x}_k/\pi_{1k}$ برآوردگر هورویتز-تامپسون مجموع جامعه کمکی از روی نمونه S_1 است و $\pi_{1k} = P(k \in S_1)$.

لم ۱ (سارن‌دال و همکاران، ۱۹۹۲): برآوردگر رگرسیونی $\hat{t}_{yr}$ که در رابطه (۱) معرفی شد، توسط خطی‌سازی تیلور[۴] به صورت

$$\hat{t}_{yr0} = \hat{t}_{y\pi} + (\hat{t}_{x\pi 1} - \hat{t}_{x\pi})'B = \sum_U y_k^{\circ} + \sum_S \breve{E}_k$$

تقریب می‌شود، که در آن $\breve{E}_k = E_k/\pi_k$ و $E_k = y_k - y_k^{\circ}$ و $y_k^{\circ} = \mathbf{x}_k'\mathbf{B}$ و $\mathbf{B} = (\sum_U \mathbf{x}_k\mathbf{x}_k'/\sigma_k^2)^{-1}\sum_U \mathbf{x}_k y_k/\sigma_k^2$. برآوردگر $\hat{t}_{yr}$ برای $t = \sum_U y_k$ تقریبا نااریب با واریانس تقریبی $AV(\hat{t}_{yr}) = \sum\sum_U \Delta_{k\ell}\breve{E}_k\breve{E}_\ell$ است، که برآوردگر آن به صورت

$$\hat{V}(\hat{t}_{yr}) = \sum\sum_S \breve{\Delta}_{k\ell}(g_{ks}\breve{e}_{ks})(g_{\ell s}\breve{e}_{\ell s})$$

می‌باشد، که در آن $\breve{e}_{ks} = e_{ks}/\pi_k$ و $e_{ks} = y_k = \hat{y}_k$ در رابطه (۳) داده شده است.

قضیه ۱ : واریانس مجانبی برآوردگر رگرسیونی تعمیم‌یافته (۴) به صورت

$$AV(\hat{t}_{yr1}) = \sum\sum_U \frac{\Delta_{k\ell}}{\pi_k\pi_\ell}E_kE_\ell + B'(\sum\sum_U \frac{\Delta_{1k\ell}}{\pi_{1k}\pi_{1\ell}}x_kx_\ell')B \qquad (۵)$$

است و برآوردگر آن نیز به صورت

$$\hat{A}(\hat{t}_{yr1}) = \sum\sum_S \frac{\Delta_{k\ell}}{\pi_{k\ell}\pi_k\pi_\ell}(g_{ks}e_{ks})(g_{\ell s}e_{\ell s}) + \hat{B}'(\sum\sum_{s\ell} \frac{\Delta_{1k\ell}}{\pi_{1k\ell}\pi_{1k}\pi_{1\ell}}x_kx_\ell')\hat{B} \qquad (۶)$$

---
۴ Taylor linerrization

خواهد بود، که در آن $\pi_{1k} = P(k \in S_1)$، $\pi_{1k\ell} = P(k\&\ell \in S_1)$ و $\hat{\mathbf{t}}_{x\pi 1} = \sum_{S_1} \mathbf{x}_k/\pi_{1k}$ برآوردگر هورویتز-تامپسون مجموع جامعه کمکی از روی نمونه S_1، $\mathbf{B} = (\sum_U \mathbf{x}_k\mathbf{x}'_k/\sigma_k^2)^{-1} \sum_U \mathbf{f}\mathbf{x}_k y_k/\sigma_k^2$ و $\hat{\mathbf{B}}$ نیز در رابطه (۲) تعریف شده است. همچنین $\Delta_{k\ell} = \pi_{k\ell} - \pi_k\pi_\ell$ و $\Delta_{1k\ell} = \pi_{1k\ell} - \pi_{1k}\pi_{1\ell}$ و نیز $E_k = y_k - \mathbf{x}'_k\mathbf{B}$ و $e_{ks} = y_k - \mathbf{x}'_k\hat{\mathbf{B}}$ و g_{ks} نیز که در رابطه (۳) تعریف شده است.

برهان: بنا بر لم ۱، برآوردگر (۴) را می‌توان به طور تقریبی به صورت

$$\hat{t}_{yr1} \simeq \hat{t}_{y\pi} + (\hat{t}_{x\pi 1} - \hat{t}_{x\pi})'B$$

نوشت و واریانس مجانبی آن را به صورت

$$\begin{aligned} AV(\hat{t}_{yr1}) &= Var[\hat{t}_{y\pi} + (\hat{t}_{x\pi 1} - \hat{t}_{x\pi})'B] \\ &= Var[\hat{t}_{y\pi} + (t_x - \hat{t}_{x\pi})'B + \hat{t}'_{x\pi 1}B] \\ &= Var[\hat{t}_{yr0} + \hat{t}'_{x\pi 1}B] \end{aligned}$$

محاسبه کرد، که در آن $\hat{t}_{yr0}$، تقریب خطی‌سازی تیلور $\hat{t}_{yr}$ معرفی شده در رابطه (۱) است. اما با توجه به فرض مجزا بودن نمونه‌های S_1 و S و در نتیجه استقلال آماره‌های متناظرشان می‌توان نوشت

$$AV(\hat{t}_{yr1}) = Var(\hat{t_{yr0}}) + Var(\hat{t}'_{x\pi 1}B) = AV(\hat{t}_{yr}) + B'Var(\hat{t}'_{x\pi 1})B \qquad (7)$$

بنابر لم ۱

$$AV(\hat{t}_{yr}) = \sum\sum_U \frac{\Delta_{k\ell}}{\pi_k\pi_\ell} E_k E_\ell$$

و چون $\hat{\mathbf{t}}_{x\pi 1}$ برآوردگر هورویتز-تامپسون است، رابطه (۷) را می‌توان به صورت

$$AV(\hat{t}_{yr1}) = \sum\sum_U \frac{\Delta_{k\ell}}{\pi_k\pi_\ell} E_k E_\ell + \hat{B}'(\sum\sum_U \frac{\Delta_{1k\ell}}{\pi_{1k}\pi_{1\ell}} x_k x'_\ell)B.$$

ادامه داد. همچنین به منظور برآورد آن با استفاده از لم ۱ داریم

$$\hat{V}(\hat{t}_{yr1}) = \sum\sum_S \frac{\Delta_{k\ell}}{\pi_{k\ell}\pi_k\pi_\ell}(g_{ks}e_{ks})(g_{\ell s}e_{\ell s}) + \hat{B}'(\sum\sum_{S_1} \frac{\Delta_{1k\ell}}{\pi_{1k\ell}\pi_{1k}\pi_{1\ell}} x_k x'_\ell)\hat{B}$$

مثال ۱ : با طرح نمونه‌گیری تصادفی ساده بدون جایگذاری، برآوردگر رگرسیونی تعمیم‌یافته به صورت $\hat{t}_{yr1SI} = \hat{t}_{y\pi SI} + (\hat{t}_{x\pi 1SI} - \hat{t}_{x\pi SI})'\hat{B}$ یا به طور معادل به فرم

$$\hat{t}_{yr1SI} = N(\bar{y}_n + (\bar{x}_{n1} - \bar{x}_n)\hat{B})$$

است. واریانس مجانبی و برآورد آن با توجه به روابط (۵) و (۶) به‌ترتیب به صورت

$$\begin{aligned} V(\hat{t}_{yr1SI}) &= N^2 S_y^2((\frac{1}{n} - \frac{1}{N}) + \rho^2(\frac{1}{n_1} - \frac{1}{n})) \\ \hat{V}(\hat{t}_{yr1SI}) &= N^2 S_y^2((\frac{1}{n} - \frac{1}{N}) + r^2(\frac{1}{n_1} - \frac{1}{n})) \end{aligned} \quad (8)$$

هستند، که نمادها و اثبات (۸) در پیوست الف ارائه شده‌اند.

مثال ۲ : جامعه‌ای متشکل از H طبقه را در نظر بگیرید که هر طبقه شامل N_h واحد، $h = 1, \dots, H$، به صورت (y_k, x_k) است، که در آن x_k مقدار متغیر تک‌بعدی کمکی و y_k مقدار متغیر هدف هستند. با فرض آن که مجموع هر دو جامعه کمکی X و هدف Y نامعلوم هستند با استفاده از برآوردگر (۴)، مجموع جامعه هدف را برآورد می کنیم. در اولین گام برای مشخص شدن احتمال‌های شمول، طرح نمونه‌گیری طبقه‌بندی به روش تصادفی ساده بدون جایگذاری در نظر گرفته می‌شود. در این صورت

$$\hat{t}_{x\pi 1} = \sum_{h=1}^{H} \frac{N_h}{n_{1h}} \sum_{S_{1h}} x_k, \hat{t}_{x\pi} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{S_h} x_k, \hat{t}_{y\pi} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{S_h} y_k$$

که در آن‌ها اندیس h، به منظور منتسب کردن نماد مورد نظر به طبقه hام است. سارن‌دال و همکاران (۱۹۹۲) نشان دادند

$$\hat{B} = \frac{\sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{S_h} x_k y_k - \hat{t}_{x\pi}\hat{t}_{y\pi}}{\sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{S_h} x_k^2 - (\hat{t}_{x\pi})^2}$$

به این ترتیب برآوردگر رگرسیونی تعمیم‌یافته (۴) قابل محاسبه است و برای برآورد واریانس مجانبی آن از روابط

$$Var(\hat{t}_{y\pi}) = Var(\sum_{h-1}^{H} \frac{N_h}{n_h} \sum_{S_h} y_k) = \sum_{h=1}^{H} N_h^2 (\frac{1}{n_h} - \frac{1}{N_h}) S_{yh}^2$$

کمک گرفته می‌شود، که در آن

$$S_{yh}^2 = \frac{1}{N_h - 1}\{\sum_{U_h} y_k^2 - (\sum_{U_h} y_k)^2 / N_h\}$$

نامعلوم است و منجر به نامعلوم بودن $Var(\hat{t}_{y\pi})$ می‌شود. برآوردگر نااریب $Var(\hat{t}_{y\pi})$ با قرار دادن برآوردگر نااریب

$$S^{۲}_{yh} = \frac{۱}{n_h - ۱}\{\sum_{S_h} y^{۲}_k - (\sum_{S_h} y_k)^{۲}/n_h\}$$

به جای $S^{۲}_{yh}$ در فرمول $Var(\hat{t}_{y\pi})$ به دست می‌آید. به طور مشابه $Var(\hat{t}_{x\pi})$ و $Var(\hat{t}_{x\pi۱})$ نیز ارائه و برآورد می‌شوند. به علاوه

$$Cov(\hat{t}_{y\pi}, \hat{t}_{x\pi}) = \sum_{h=۱}^{H} N^{۲}_h(\frac{۱}{n_h} - \frac{۱}{N_h})S_{xyh}$$

که در آن

$$S_{xyh} = \frac{۱}{N_h - ۱}\{\sum_{U_h} x_k y_k - (\sum_{U_h} y_k)(\sum_{U_h} x_k)/N_h\}$$

و به دلیل نامعلوم بودن، توسط برآوردگر نااریب نمونه‌ای آن، یعنی

$$S_{xyh} = \frac{۱}{n_h - ۱}\{\sum_{S_h} x_k y_k - (\sum_{S_h} y_k)(\sum_{S_h} x_k)/n_h\}$$

برآورد می‌شود. از طرفی با توجه به لم ۱ برآوردگر رگرسیونی تعمیم‌یافته (۴) بر مبنای رابطه (۵) به صورت

$$\hat{t}_{yr۱} = \hat{t}_{y\pi} + (\hat{t}_{x\pi۱} - \hat{t}_{x\pi})B$$

خواهد بود. همچنین مجزا بودن دو نمونه $S_۱$ و S، استقلال متغیر تصادفی $\hat{t}_{x\pi۱}$ از دو متغیر تصادفی $\hat{t}_{x\pi}$ و $\hat{t}_{y\pi}$ نتیجه می‌شود. پس

$$AV(\hat{t}_{yr۱}) = Var(\hat{t}_{y\pi}) + B^{۲}(Var(\hat{t}_{x\pi}) + Var(\hat{t}_{x\pi۱})) - ۲BCov(\hat{t}_{y\pi}, \hat{t}_{x\pi})$$

که برآوردگر آن با استفاده از برآوردگرهای به‌دست آورده می‌شود.

تعریف ۲ : برآوردگر کالیبره برای مجموع جامعه اصلی Y، t_y، وقتی که مجموع جامعه کمکی $\mathbf{X}$، $\mathbf{t_X}$، نامعلوم است عبارت است از

$$\hat{t}_{yw۱} = \sum_{S} w_{۱k} y_k$$

که در آن وزن‌های $w_{۱k}$، $k = ۱, \ldots, N$، با مینیمم ساختن فاصله میان آنها با وزن‌های طرح $d_k = \frac{۱}{\pi_k}$ توسط یک تابع فاصله دلخواه $G_k(w, d)$ که در شرایط ۱ و ۲ بیان شده در بخش ۲ صدق کند به همراه قید $\sum_S w_{۱k}\mathbf{x_k} = \hat{\mathbf{t}}_{\mathbf{x}\pi\mathbf{1}}$ که در آن $\hat{\mathbf{t}}_{x\pi۱}$، در تعریف ۱ معرفی شده است.

۴ شبیه‌سازی

برای بررسی صحت نظریه ارائه شده، با استفاده از بسته نرم‌افزاری MATLAB، در محیط برنامه‌نویسیM-file، روش MCMC جامعه‌ای متشکل از پنج طبقه که هر طبقه آن از یک توزیع نرمال دو متغیره به خصوص (قابل مشخص‌سازی) با حجم‌های متفاوت تولید نموده‌ایم. شیوه نمونه‌گیری درون طبقه‌ها تصادفی ساده بدون جایگذاری است. برای نمونه اصلی S حجم نمونه در هر طبقه مشخص شده است، امّا برای نمونه $S_۱$، ابتدا به طور کلی حجم $n_۱$ با شرط $n_۱ < n$ مشخص شده و سپس با استفاده از رابطه

$$n_{۱h} = [[n_h + \frac{N_h}{N}(n_۱ - n)]]$$

حجم زیرنمونه‌های $S_{۱h}$ از طبقه‌ها به دست آمده است. منظور از $[[\cdot]]$ عملگری است که عددی حقیقی را به عنوان ورودی گرفته و نزدیکترین عدد صحیح به آن را به عنوان خروجی باز می‌گرداند. در هر تکرار برنامه شبیه‌سازی، مجموع کل جامعه هدف، $T = \sum_U y_k = \sum_h \sum_{U_h} y_k$، واریانس تجربی جامعه هدف،

$$V(T) = \frac{N^۲}{N-۱}(\sum_U y_k^۲ - (\sum_U y_k)^۲/N)$$

برآوردگرهای هورویتز-تامپسون، رگرسیون تعمیم‌یافته با فرض معلوم بودن مجموع جامعه کمکی و رگرسیونی تعمیم‌یافته با فرض نامعلوم بودن مجموع جامعه کمکی که در رابطه (۴) ارائه شده و برآورد واریانس آنها به ترتیب محاسبه شده است. فرایند تولید جامعه و نمونه‌گیری و محاسبه برآوردگرهای مورد نظر، ۱۰۰۰۰ بار تکرار شده و میانگین مقادیر برآوردگرها به عنوان برآوردی از امید ریاضی آنها محاسبه شده است. در خروجی مخاطره نسبی[۵] برای سه برآوردگر مجموع کل جامعه هدف یعنی برآوردگرهای هورویتز-تامپسون، رگرسیونی تعمیم یافته با فرض معلوم بودن مجموع جامعه کمکی و رگرسیونی تعمیم‌یافته با فرض نامعلوم بودن مجموع جامعه کمکی نسبت به مجموع واقعی جامعه هدف ارائه شده‌اند. به علاوه کخاطره‌های نسبی برآوردهای واریانس دو برآوردگر هورویتز-تامپسون و رگرسیونی تعمیم‌یافته با فرض نامعلوم بودن مجموع جامعه کمکی که در رابطه (۴) ارائه شده نسبت به برآورد واریانس برآوردگر رگرسیونی تعمیم‌یافته با فرض معلوم بودن مجموع جامعه کمکی ارائه شده‌اند.

۵ Relative Risk

برای بررسی کارایی نظریه این برنامه را دو بارتحت دو گروه از اطلاعات ورودی الف و ب که به جز در مقدار ضریب همبستگی دو متغیر کمکی و هدف در مابقی اطلاعات یکسان هستند اجرا شده است. در جداول ۱ و ۲ اطلاعاتی ورودی برنامه از $n_۱ = ۳۰۰$ ارائه شده‌اند. نتایج خروجی نیز در جدول ۳ برای هر دو گروه اطلاعات الف و ب نشان داده شده است.

جدول ۱ اطلاعات ورودی سری الف

	پارامتر							
طبقه	ρ_{xy}	$\sigma_y^۲$	$\sigma_x^۲$	μ_y	μ_x	n_h	$n^*_{۱h}$	N_h
اول	۰/۷۵	۴	۲	۱۰	۵	۵۰	۵۸	۴۰۰
دوم	۰/۶۸	۳	۴	۱۵	۱۰	۵۵	۶۸	۶۰۰
سوم	۰/۷۲	۶	۶	۳۰	۸	۴۲	۴۸	۳۰۰
چهارم	۰/۶۳	۵	۸	۲۵	۱۵	۴۷	۵۵	۳۶۰
پنجم	۰/۷۹	۲	۱۰	۱۸	۲۰	۶۰	۷۱	۵۲۰

جدول ۲ اطلاعات ورودی سری ب

	پارامتر							
طبقه	ρ_{xy}	$\sigma_y^۲$	$\sigma_x^۲$	μ_y	μ_x	n_h	$n^*_{۱h}$	N_h
اول	۰/۲	۴	۲	۱۰	۵	۵۰	۵۸	۴۰۰
دوم	۰/۰۵	۳	۴	۱۵	۱۰	۵۵	۶۸	۶۰۰
سوم	۰/۱	۶	۶	۳۰	۸	۴۲	۴۸	۳۰۰
چهارم	۰/۱۴	۵	۸	۲۵	۱۵	۴۷	۵۵	۳۶۰
پنجم	۰/۰۹	۲	۱۰	۱۸	۲۰	۶۰	۷۱	۵۲۰

جدول ۳ خروجی برنامه برای اطلاعات دریافتی از دو گروه الف و ب

	گروه اطلاعاتی	
کمیت‌ها	الف	ب
t_y	۴/۰۳۶۰۲۵e + ۴	۴/۰۳۶۰۰۳e + ۴
$\hat{t}_{y\pi}$	۴/۰۳۶۰۳۸e + ۴	۴/۰۳۵۹۵۶e + ۴
$\hat{t}_{yr}$	۴/۰۳۶۰۵۳e + ۴	۴/۰۳۵۹۷۹e + ۴
$\hat{t}_{yr۱}$	۴/۰۳۶۰۷۷e + ۴	۴/۰۳۶۰۰۳e + ۴
واریانس کل تجربی	۲/۱۶۵۳۳e + ۱۲	۲/۱۶۵۲۵e + ۱۲
$Var(\hat{t}_{y\pi})$	۶/۰۱۰۶۴۰e + ۴	۶/۰۰۳۰۸۹e + ۴
$A\hat{V}(\hat{t}_{yr})$	۳/۶۵۴۲۷۷e + ۴	۶/۳۰۷۶۷۹e + ۴
مخاطره نسبی	۴/۶۰۷۲۴۰e + ۴	۶/۸۸۹۷۴۱e + ۴
t_y : $\hat{t}_{y\pi}$ مخاطره نسبی	۳/۳۲۶۱۴۵e − ۴	−۰/۰۰۱۱۷e + ۰
t_y : $\hat{t}_{yr}$ مخاطره نسبی	۷/۱۴۴۸۴۴e − ۴	−۶/۰۱۶۶۴e − ۴
t_y : $\hat{t}_{yr}$ مخاطره نسبی	۱۳/۰۲۸۱۸e − ۴	۵/۶۰۴۵۰۱e − ۶
$\hat{V}(\hat{t}_{y\pi})$: $\hat{V}(\hat{t}_{y\pi})$ مخاطره نسبی	۶۴/۴۸۲۳۲e + ۰	−۴/۸۲۸۸۷e + ۰
$A\hat{V}(\hat{t}_{yr})$: $A\hat{V}(\hat{t}_{yr۱}$ مخاطره نسبی	۲۶/۰۷۸۰۱e + ۰	۹/۲۲۷۸۳۹e + ۰

۵ بحث و نتیجه‌گیری

برآوردگر مجموع جامعه هدف Y با استفاده از اطلاعات جامعه کمکی X، که مجموع آن نامعلوم است، ارائه شد. نمونه‌گیری مجدد توسط محققان بسیاری در وضعیت مشابه مورد تحلیل و بررسی قرار گرفته است. اما در عمل گاهی نمونه‌هایی مناسب شامل اطلاعات کمکی، که به هر دلیلی از قبل گردآوری شده‌اند در دست است و دلیلی برای صرف هزینه اضافی برای استفاده از روش نمونه‌گیری مجدد وجود ندارد. تنها کافی است نمونه‌ای جدید از جامعه توام (X,Y) گرفته و با استفاده از برآوردی که برای مجموع نامعلوم جامعه کمکی از روی نمونه گذشته صورت پذیرد، با روش برآورد رگرسیونی، مجموع کل Y برآورد شده است. به علاوه واریانس مجانبی و برآورد واریانس آن نیز ارائه شد. کارایی این برآوردگر در مطالعه‌ای شبیه‌سازی با برآوردگرهای هورویتز_تامپسون و برآوردگر رگرسیونی با فرض معلوم بودن مجموع کل X، مورد مقایسه قرار گرفت.

برای انجام این مقایسه دو گروه اطلاعات برای جامعه در نظر گرفته شد که جز در مقادیر ضریب همبستگی کاملا یکسان بوده‌اند. در نتایج گروه اطلاعاتی الف که ضریب همبستگی X و Y، بیشتر از ۰/۷ در نظر گرفته شده بود، ترتیب $A\hat{V}(\hat{t}_{yr}) < A\hat{V}(\hat{t}_{yr1}) < \hat{V}(\hat{t}_{y\pi})$ ملاحظه شد. یعنی برآوردگر پیشنهادی مطابق انتظار کارایی کمتری نسبت به برآوردگر رگرسیونی وقتی که مجموع کل X معلوم است دارد، اما نسبت به برآوردگر هورویتز_تامپسون که از اطلاعات کمکی استفاده نمی‌کند کارایی بیشتری دارد.

در نتایج گروه اطلاعاتی ب که ضریب همبستگی X و Y، کمتر از ۰/۲ در نظر گرفته شده بود، ترتیب $\hat{V}(\hat{t}_{y\pi}) < A\hat{V}(\hat{t}_{yr}) < A\hat{V}(\hat{t}_{yr1})$ ملاحظه شد، که به دلیل آن است که میزان همبستگی اطلاعات کمکی X با جامعه هدف Y کم می‌باشد و نمی‌توان از اطلاعات X به عنوان متغیر کمکی استفاده نمود.

پیوست الف:

برای اثبات رابطه (۸)، با تعریف نمادهای

$$\bar{x}_{n_1} = \frac{\sum_{S_1} x_k}{n_1}\ ,\ \bar{x}_n = \frac{\sum_S x_k}{n}\ ,\ \bar{y}_n = \frac{\sum_S y_k}{n}\ ,\ S_x^2 = \frac{\sum_U (x_k - \mu_x)^2}{(N-1)}$$

$$S_y^2 = \frac{\sum_U (y_k - \mu_y)^2}{(N-1)}\ , \qquad S_{xy} = \frac{\sum_U (x_k - \mu_x)(y_k - \mu_y)}{(N-1)}$$

$$s_{xy} = \frac{\sum_S (x_k - \bar{x}_n)(y_k - \bar{y}_n)}{(n-1)},\ s_y^2 = \frac{\sum_S (y_k - \bar{y}_n)^2}{(n-1)},\ s_x^2 = \frac{\sum_S (x_k - \bar{x}_n)^2}{(n-1)}$$

$$\rho = \frac{S_{xy}}{(S_x^2 S_y^2)^{\frac{1}{2}}}\ ,\ \hat{B} = \frac{s_{xy}}{s_x^2}$$

و درنظر گرفتن برآوردگر میانگین جامعه به صورت

$$\bar{y}_{r1SI} = \bar{y}_n + (\bar{x}_{n_1} - \bar{x}_n)\hat{B} \qquad (9)$$

داریم

$$\hat{t}_{yr1SI} = N(\bar{y}_n + (\bar{x}_{n_1} - \bar{x}_n)\hat{B}) = N\bar{y}_{r1SI} \qquad (10)$$

بنابراین $Var(\hat{t}_{yr1SI}) = N^2 Var(\bar{y}_{r1SI})$ چون $\hat{B} = \frac{S_{xy}}{S_x^2}$ برای B اریب است، بنابراین برآوردگر (۹) نیز اریب است. با قرار دادن

$$\bar{x}_{n_1} = \mu_x + \varepsilon_1, \quad \bar{x}_n = \mu_x + \varepsilon_2, \quad s_x^2 = S_x^2 + \varepsilon_3, \quad s_{xy} = S_{xy} + \varepsilon_4 \qquad (11)$$

و اینکه $B = \frac{S_{xy}}{S_x^2}$ مقداری ثابت است، داریم

$$\begin{aligned} \bar{y}_{r1SI} &= \bar{y}_n + (\varepsilon_1 - \varepsilon_2)\frac{S_{xy}(1 + \varepsilon_4/S_{xy})}{S_x^2(1 + \varepsilon_3/S_x^2)} \\ &= \bar{y}_n + B(\varepsilon_1 - \varepsilon_2)(1 + \frac{\varepsilon_4}{S_{xy}})(1 - \frac{\varepsilon_3}{S_x^2})^{-1} \end{aligned}$$

از آنجا که $|\frac{\varepsilon_3}{S_x^2}| < 1$

$$\begin{aligned} \bar{y}_{r1SI} &\simeq \bar{y}_n + B(\varepsilon_1 - \varepsilon_2)(1 + \tfrac{\varepsilon_4}{S_{xy}})(1 - \tfrac{\varepsilon_3}{S_x^2} + \ldots) \\ &\simeq \bar{y}_n + B(\varepsilon_1 - \varepsilon_2 + \tfrac{\varepsilon_1\varepsilon_4}{S_{xy}} - \tfrac{\varepsilon_2\varepsilon_4}{S_{xy}} - \tfrac{\varepsilon_1\varepsilon_3}{S_x^2} + \tfrac{\varepsilon_2\varepsilon_3}{S_x^2} - \ldots) \end{aligned}$$

چون $E(\varepsilon_1) = E(\varepsilon_2) = 0$، داریم

$$E(\bar{y}_{r1SI}) \simeq (\bar{y}_n) + B(\frac{E(\varepsilon_1\varepsilon_4)}{S_{xy}} - \frac{E(\varepsilon_2\varepsilon_4)}{S_{xy}} - \frac{E(\varepsilon_1\varepsilon_3)}{S_x^2} + \frac{E(\varepsilon_2\varepsilon_3)}{S_x^2} - \ldots) \qquad (12)$$

امّا با توجه به اینکه نمونه‌های S_1 و S تصادفی ساده مستقل از جامعه‌ای واحد هستند، داریم $E(\bar{y}_n) = \mu_y$. با توجه به روابط (۱۱) داریم

$$E(\varepsilon_1\varepsilon_4) = E[(\bar{x}_{n_1} - \mu_x)(s_{xy} - S_{xy})] = Cov(\bar{x}_{n_1}, s_{xy})$$

به همین ترتیب داریم

$$E(\varepsilon_1\varepsilon_4) = Cov(\bar{x}_n, s_{xy}), \quad E(\varepsilon_1\varepsilon_3) = Cov(\bar{x}_{n_1}, s_x^2), \quad E(\varepsilon_2\varepsilon_3) = Cov(\bar{x}_n, s_x^2)$$

بنابراین

$$E(\bar{y}_{r1SI}) \simeq \mu_y + B\{\frac{Cov(\bar{x}_{n_1}, s_{xy}) - Cov(\bar{x}_n, s_{xy})}{S_{xy}} + \frac{Cov(\bar{x}_{n_1}, s_x^2) - Cov(\bar{x}_n, s_x^2)}{S_x^2}\}$$

جملهٔ دوم در رابطه بالا معرف اریبی تقریبی $\bar{y}_{r1SI}$ است. چون s_{xy} و s_x^2 از نمونه S به دست آمده‌اند، $\bar{x}_{n_1}$ که از نمونه مستقل S_1 محاسبه شده است، مستقل هستند. از طرفی در نمونه تصادفی S، s_{xy} و s_x^2 از $\bar{x}_n$ مستقل هستند. بنابراین در عبارت بالا همه کوواریانس‌ها برابر صفر هستند. در نتیجه $\bar{y}_{r1SI}$ برآوردگر نااریب میانگین جامعه Y است. برای محاسبه واریانس $\bar{y}_{r1SI}$ رابطه (۹) را با توجه به روابط (۱۱) می‌توان به صورت

$$\bar{y}_{r1SI} = \bar{y}_n + (\varepsilon_1 - \varepsilon_2)\hat{B}$$

نوشت. با توجه به نااریبی $\bar{y}_{r1SI}$ داریم

$$V(\bar{y}_{r1SI}) = E[(\bar{y}_{r1sI} - \mu_y)^2] = E[(\bar{y}_n + \hat{B}(\varepsilon_1\varepsilon_2) - \mu_y)^2]$$

در صورتی که N بزرگ باشد، می‌توان قرار داد $\hat{B} \simeq B$. بنابراین

$$\begin{aligned} AV(\bar{y}_{r1SI}) &= E[(\bar{y}_n + B(\varepsilon_1 - \varepsilon_2) - \mu_y)^2] \\ &= E(\bar{y}_n - \mu_y)^2 + B^2E(\varepsilon_1 - \varepsilon_2)^2 + 2BE[(\bar{y}_n - \mu_y)(\varepsilon_1 - \varepsilon_2)] \end{aligned}$$

در نمونه‌گیری تصادفی ساده داریم $E(\bar{y}_n - \mu_y)^2 = V(\bar{y}_n) = (\frac{1}{n} - \frac{1}{N})S_y^2$. از طرفی با توجه به مقادیر ε_1 و ε_2

$$\begin{aligned} E(\varepsilon_1 - \varepsilon_2)^2 &= E(\bar{x}_{n_1} - \mu_x)^2 + E(\bar{x}_n - \mu_x)^2 - 2E(\bar{x}_{n_1} - \mu_x)(\bar{x}_n - \mu_x) \\ &= (\frac{1}{n_1} - \frac{1}{N})S_x^2 + (\frac{1}{n} - \frac{1}{N})S_x^2 - 2E(\bar{x}_{n_1} - \mu_x)(\bar{x}_n - \mu_x) \end{aligned}$$

چون نمونه‌های S_1 و S مستقلند در نتیجه $\bar{x}_{n_1}$ و $\bar{x}_n$ نیز مستقل هستند، بنابراین

$$E(\bar{x}_{n_1} - \mu_x)(\bar{x}_n - \mu_x) = 0 \qquad (۱۳)$$

پس

$$E(\varepsilon_1 - \varepsilon_2)^2 = (\frac{1}{n_1} - \frac{1}{N})S_x^2 + (\frac{1}{n} - \frac{1}{N})S_x^2 = (\frac{1}{n_1} + \frac{1}{n} - \frac{2}{N})S_x^2$$

بنابراین جمله دوم طرف راست (۱۳) برابر $B^2(\frac{1}{n_1}+\frac{1}{n}-\frac{2}{N})S_x^2$ است. در مورد جمله سوم طرف راست (۱۳) داریم

$$E[(\bar{y}_n-\mu_y)(\varepsilon_1-\varepsilon_2)]=E[(\bar{y}_n-\mu_y)(\bar{x}_{n_1}-\mu_x)]-E[(\bar{y}_n-\mu_y)(\bar{x}_n-\mu_x)] \qquad (۱۴)$$

این بار استقلال دو نمونه S_1 و S استقلال $\bar{x}_{n_1}$ و $\bar{y}_n$ را نتیجه می‌دهد و جمله اول طرف راست (۱۴) برابر صفر می‌شود. برای جمله دوم طرف راست (۱۴) داریم

$$E[(\bar{y}_n-\mu_y)(\bar{x}_n-\mu_x)]=Cov(\bar{x}_n,\bar{y}_n)=(\frac{1}{n}-\frac{1}{N})S_{xy}$$

بنابراین جمله سوم طرف راست (۱۴) عبارت خواهد شد از

$$2BE[(\bar{y}_n-\mu_y)(\varepsilon_1-\varepsilon_2)]=-2B(\frac{1}{n}-\frac{1}{N})S_{xy}$$

با منظور کردن مقادیر محاسبه شده در (۱۲)، نتیجه می‌شود

$$\begin{aligned} AV(\bar{y}_{r1SI}) &= (\frac{1}{n}-\frac{1}{N})S_y^2+B^2(\frac{1}{n_1}+\frac{1}{n}-\frac{2}{N})S_x^2-2B(\frac{1}{n}-\frac{1}{N})S_{xy} \\ &= (\frac{1}{n}-\frac{1}{N})S_y^2+\frac{S_{xy}}{S_x^2}(\frac{1}{n_1}-\frac{1}{n}) \\ &= S_y^2[(\frac{1}{n}-\frac{1}{N})+\rho^2(\frac{1}{n_1}-\frac{1}{n})] \end{aligned}$$

که در آن $\rho=\frac{S_{xy}}{(S_x^2S_y^2)^{\frac{1}{2}}}$. بنابراین با استفاده از (۱۰) می‌توان نوشت

$$AV(\hat{t}_{yr1SI})=N^2S_{yy}[(\frac{1}{n}-\frac{1}{N})+\rho^2(\frac{1}{n_1}-\frac{1}{n})]$$

معمولا مقادیر S_{yy} و ρ نامعلومند و از روی مشاهدات نمونه S به ترتیب به‌صورت $s_{xy}=\frac{1}{n-1}\sum_S(x_k-\bar{x}_n)(y_k-\bar{y}_n)$ و $r=\frac{s_{xy}}{(s_x^2s_y^2)^{\frac{1}{2}}}$ برآورد می‌شوند.

تقدیر و تشکر

نویسندگان این مقاله از داوران و سردبیر محترم مجله به خاطر ارائه پیشنهادهای ارزنده تشکر و قدردانی را دارند.

مراجع

Binder, D. A., Babyak, C., Brodeur, M., Hidiroglou, M., and Jocelyn, W. (2000), Variance Estimation for Two-Phase Stratified Sampling, *Journal of The Canadian Statistics*, **28**, 751-764.

Cochran, W. G. (1977), *Sampling Techniques, 3rd edition*, New York: Wiley.

Devile, J. C. and Särndal, C. E., (1992), Calibration Estimators in Survey Sampling, *Jornal of American Statistical Assocciation,* **87**, 376-382.

Horvitz, D. G., and Thompson, D. J. (1952), A Generalization of Sampling Without Replacement from a Finite Universe. *Jornal of American Statistical Assocciation*, **47**, 663-685.

Kim, J. K., Navarro, A. and Fuller, W. A. (2006), Replication Variance Estimation for Two-phase Sampling Stratified Sampling, *Jornal of American Statistical Assocciation*, **101**, 312-320.

Rao, J. N. K. and Sitter, R. R. (1995), Variance Estimation under Two-phase Sampling with Application to Imputation for Missing Data, *Biometrika*, **82** 453-460.

Särndal, C. E., Swensson, B. and Wretman. J. (1992), *Model Assisted Survey Sampling*. Springer, NewYork.

Sitter, R. R. (1997), Variance Estimation for the Regression Estimator in Two-phase Sampling, *Jornal of American Statistical Assocciation*, **92**, 780-787.

Valliant, R., Dorfman, A. H. and Royall, R. M., (2000), *Finite Population Sampling and Inference: A Prediction Approach*, Wiley Series in Probability and Statistics, Survey Methodology Section. Wiley, New York.