آمارگیریهای نمونه ای چند چارچوبی

مرجان نورینی[۱] محمد صالحی[۲]

چکیده

چارچوب آماری سنگ بنای یک فرآیند نمونه گیری است. اما گاه چارچوبی که به تنهایی بتواند کلیه واحدهای جامعه مورد مطالعه را پوشش دهد در دسترس نمی باشد. از سوی دیگر ممکن است امکان دستیابی به چارچوبی کامل وجود داشته اما هزینه آمارگیری از آن بالا باشد. در این مواقع در آمارگیریها می توان با استفاده همزمان از چندچارچوب (که لزوماً کامل نبوده و هزینه آمارگیری از آنها کمتر است) کارایی را افزایش و هزینه آمارگیری را کاهش داد. چنین آمارگیریهایی تحت عنوان آمارگیری های چند چارچوبی بکار می روند. مرکز آمار کانادا و ایالات متحده پیشرو در اجرای این طرح می باشند. اهمیت کاربردی این موضوع ما را بر آن داشت تا مقالهٔ حاضر را که مروری بر آمارگیری های نمونه ای چند چارچوبی است، ارائه دهیم.

واژه های کلیدی: چارچوب چندگانه[۳]، بررسی های پیچیده[۴]، برآورد جک نایف واریانس[۵]، برآورد خطی واریانس[۶]، روشهای بازنمونه گیری[۷]

۱. مقدمه

چارچوب آماری یا فهرست واحدهای آمارگیری، اساس و مبنای یک طرح آمارگیری نمونه ای را تشکیل می دهد. گاه ممکن است چارچوبی که کلیه ی واحدهای جامعه ی مورد مطالعه را پوشش دهد در دسترس نباشد، اما امکان دستیابی به پوشش کامل، با تلفیقی از دو یا چند چارچوب فراهم شود. در چنین حالتی به منظور دسترسی به پوشش مناسب از دو چارچوب یا بیشتر بطور همزمان استفاده می شود. گاهی نیز ممکن است یک چارچوب، پوشش کامل را برای جامعه ی مورد مطالعه فراهم کند، اما چارچوب ناقص دیگری موجود باشد که هزینهٔ آمارگیری از چارچوب کامل باشد. در این شرایط به دلیل پایین تر بودن هزینهٔ آمارگیری از این چارچوب، می توان با هزینه ای مشخص و ثابت، از دو چارجوب استفاده کرده و حجم نمونه را بزرگتر و واریانس برآورد را کاهش داد. گاهی نیز ممکن است یک چارچوب فهرستی کامل در دسترس باشد اما عملاً با گذشت زمانی نسبتاً طولانی بدلیل بروز تغییرات فراوان در آن، منبعی برای بروز خطاهای غیر نمونه گیری شود. از آنجا که یک فهرست ناحیه ای، چارچوبی کم هزینه، کمتر در معرض تغییرات و سهل الوصول می باشد ترکیب آن با یک چارچوب از اعضای جامعه که احتمالاً ناقص باشد می تواند نتایج مفیدی را حاصل نماید.

شاید بتوان گفت اولین شالوده آمارگیریهای چند چارچوبی در سال (۱۹۴۹) با آمارگیری از فروشگاه های خرده فروشی که مجری آن دفتر سرشماری آمریکا بود گذاشته شد. سپس هارتلی[۸] [۱] در سال (۱۹۶۲) نظریه مقدماتی چارچوبهای چندگانه را توسعه داد. او با فرض اینکه اجتماع چارچوبها، جامعه را پوشش می دهد، واحدهای جامعه را به زیر مجموعه های دو به دو ناسازگار شامل اجتماعها و اشتراکهای چارچوبهای مختلف تقسیم کرد. بعد از وی لاند[۹] [۲] در سال (۱۹۶۸)،

۱- مرکز آمار ایران

۲-دانشکدهٔ علوم ریاضی، دانشگاه اصفهان

Multiple Frame -۳

Complex Survey -۴

Jackknife Variance Estimation -۵

Linearization Variance Estimation -۶

Resampling Methods -۷

Hartley -۸

Lund-۹

فولر و بورمیستر[۱] [۳] در سال (۱۹۷۲)، و گل[۲][۴] در سال (۱۹۷۵)، فوردوبوسکر[۳][۵] در سال (۱۹۷۶)، بانکیر[۴][۶] در سال (۱۹۸۶)، اسکینر[۵][۷] در سال (۱۹۹۱)، اسکینر و رائو[۶][۸] در سال (۱۹۹۶) و کالتون و اندرسون[۷][۹] در سال (۱۹۸۶) مقالاتی را در این زمینه ارائه نموده اند.

اخیراً استفاده از نمونه گیریهای دو چارچوبی، که یکی از چارچوبها را فهرست شماره تلفن افراد و دیگری را یک چارچوب ناحیه ای تشکیل می دهد، در برخی کشورها مرسوم شده است. در ایالات متحده، بررسی های اولیه جهت استفاده از دو چارچوب در طرحهای نمونه گیری کشاورزی، از اواسط دهۀ ۱۹۵۰ و جایگزینی نمونه گیری مبتنی بر این روش در بسیاری از نمونه گیریهای کشاورزی صورت گرفت. بدین ترتیب می توان مشاهده کرد که استفاده توأم از دو چارچوب در طرحهای آماری کشاورزی از سالها پیش در ایالات متحده باب شده اشت. لیکن به دلیل نیاز به تکنولوژی پیشرفته برای ایجاد یک چارچوب ناحیه ای، تاکنون تعداد اندکی از کشورها از این روش بهره جسته اند. سازمان خواربار و کشاورزی وابسته به سازمان ملل متحدFAO [۸] به دلیل مزایایی که نمونه گیریهای دو چارچوبی به همراه دارد، برای سرشماری های آتی کشاورزی، استفاده از این روش را به سایر کشورها توصیه کرده است. البته استفاده از طرح دو چارچوبی در طرحهای آماری کشاورزی مستلزم وجود چارچوب ناحیه ای و فهرستی است. برای این منظور، تهیه یک چارچوب ناحیه ای کشاورزی مستلزم صرف هزینه بسیار و نیروی کارشناسی بسیار متبحر و تکنولوژی پیشرفته در زمینه سیستم های ماهواره ای و کامپیوتری است.

طرحهای دو چارچوبی در حالتی که هر دو چارچوب ناقصند نیز بکار می روند. آمارگیری وسیع ملی کودکان و جوانان زیر نظر اداره آمار کانادا برای تهیه اطلاعاتی در رابطه با مطالعه صفات و کارآزمودگی کودکان کانادایی از سه چارچوب ناقص مجزای متداخل استفاده می کند.

اخیراً در ایالات متحده در آمارگیری های مربوط به افراد بی خانمان از طرحهای چند چارچوبی استفاده می شود. مرکز آمار ایران نیز تحقیقاتی را در زمینه کاربرد و استفاده از طرح چارچوب چندگانه در آمارگیری های مختلف آغاز کرده است.

۲. برآوردگرهای مجموع جامعه

برای سادگی فرض کنید دو چارچوب A و B موجود است که هر دو ناقص و دارای واحدهای مشترک با یکدیگر هستند بطوریکه مجموع آنها رویهم، کل جامعه مورد مطالعه را پوشش می دهد. از چارچوبهای A و B ، سه (۱-۲۲) حوزه[۹] دو به دو ناسازگار به دست می آید.

حوزه *a* : شامل واحدهایی است که فقط در چارچوب A می باشند. $a=A\bigcap B^c$

حوزه *b* : شامل واحدهایی است که فقط در چارچوب B می باشند. $b=A^c\bigcap B$

حوزه *ab* : شامل واحدهایی است که در هر دو چارچوب می باشند. $ab=A\bigcap B$

(c نشان دهندۀ مکمل مجموعه می باشد)

N_A و N_B تعداد واحدها در چارچوبهای A و B ، N_a و N_b و N_{ab} تعداد واحدهای موجود در حوزه های a و b و ab می باشد.

دو نمونه مستقل s_A و s_B بر اساس طرحهای نمونه گیری احتمالی $p_B(s_A)$ و $p_B(s_B)$ به اندازه n_A و n_B از دو چارچوب فوق گرفته می شود. بطوریکه احتمال شمول نمونه حاصل از چارچوب A عبارتست از: $\pi_i^A=p\{i\in s_A\}$ و احتمال شمول نمونه حاصل از چارچوب B برابر است با: $\pi_i^B=p\{i\in s_B\}$.

در مجموع $\tilde{n}_A$ واحد اولیه نمونه گیری[۱۰] از چارچوب A و $\tilde{n}_B$ واحد اولیه نمونه گیری از چارچوب B انتخاب می شود بطوریکه $\tilde{n}=\tilde{n}_A+\tilde{n}_B$ واحد اولیه نمونه گیری داریم.

بعضی از این واحدها در نمونه ممکن است مشابه باشند یعنی به هر دو نمونه $s_A\bigcap s_B$ تعلق داشته باشند.

بر اساس نمونه های مستقل فوق، $\tilde{n}_a$ و $\tilde{n}_{ab}^A$، تعداد واحد های اولیه نمونه گیری حاصل از چارچوب A می باشند که به ترتیب در زیر مجموعه های a و ab قرار دارند. به همین ترتیب $\tilde{n}_b$ و $\tilde{n}_{ab}^B$ نیز تعداد واحدهای اولیه نمونه گیری حاصل از چارچوب B می باشند که به ترتیب در زیر مجموعه های a و ab هستند. با فرض اینکه Y_a ، Y_b ، Y_{ab} ، μ_a ، μ_b و μ_{ab} به ترتیب مجموع و میانگین جامعه در حوزه های a و b و ab باشند، داریم:

$$Y=Y_a+Y_{ab}+Y_b \qquad (۱)$$

۱- Fuller and Burmeister
۲- Vogel
۳- Ford and Boseker
۴- Bankier
۵- Skinner
۶- Rao
۷- Kalton and Anderson
۸- Food and Agriculture Organization

۹- domain
۱۰- Primery sampling unit (psu)

هدف برآورد Y است.

چندین برآوردگر نقطه ای تحت عنوان برآوردگر چارچوب دوگانه برای برآورد Y پیشنهاد شده که همگی آنها به فرم $\hat{Y}=\hat{Y}_a+\hat{Y}_{ab}+\hat{Y}_b$ می باشند. هر یک از این برآوردگرها بسته به اینکه اطلاعات حاصل از دو نمونه برای برآورد Y چگونه با هم ترکیب می شوند با هم تفاوتهایی دارند.

مسئله برآورد بطور خاص به اطلاع و آگاهی ما از N_A، N_B، N_a، N_{ab} و N_b بستگی دارد که بطور کلی سه حالت عمده وجود دارد:

۱) N_A، N_B و N_{ab} معلوم

۲) N_A و N_B معلوم اما N_{ab} نامعلوم

۳) N_A، N_B و N_{ab} نامعلوم

برای سادگی در این مقاله روی حالت ۲ متمرکز شده و برآوردگرها را در این حالت ارائه می دهیم.

فرض می کنیم w_i^A و w_i^B وزنهای هایک [۱۰] برای طرحهای نمونه گیری مورد استفاده در چارچوبهای A و B باشند:

$$w_i^A = N_A\,[\pi_i^A \sum_{j\in s_A}(\frac{1}{\pi_j^A})]^{-1}$$

$$w_i^B = N_B\,[\pi_i^B \sum_{j\in s_B}(\frac{1}{\pi_j^B})]^{-1} \qquad (۲)$$

دو متغیر نشانگر زیر را برای هر یک از چارچوبهای A و B در نظر می گیریم:

$$\delta_i^A = \begin{cases} 1 & \text{واحد } I \text{ متعلق به چارچوب } A \text{ باشد} \\ 0 & \text{در غیر این صورت} \end{cases}$$

$$\delta_i^B = \begin{cases} 1 & \text{واحد } I \text{ متعلق به چارچوب } B \text{ باشد} \\ 0 & \text{در غیر این صورت} \end{cases}$$

در نتیجه برآوردگر مجموع در سه حوزه a و b و ab به قرار زیر می باشند.

$$\hat{Y}_a^A = \sum_{i\in s_A} w_i^A\,(1-\delta_i^B)\,y_i\,,\quad \hat{Y}_{ab}^A = \sum_{i\in s_A} w_i^A\,\delta_i^B\,y_i$$

$$\hat{Y}_b^B = \sum_{i\in s_B} w_i^B\,(1-\delta_i^A)\,y_i\,,\quad \hat{Y}_{ab}^B = \sum_{i\in s_B} w_i^B\,\delta_i^A\,y_i$$

برآوردگر اندازه ی هر یک از حوزه ها نیز به صورتی مشابه با قرار دادن $y_i=1$ در تعاریف $\hat{Y}_a^A$ و $\hat{Y}_{ab}^A$ و $\hat{Y}_b^B$ و $\hat{Y}_{ab}^B$ بدست می آیند.

برای سادگی مقایسهٔ برآوردگرهای آتی، تعاریف زیر را در نظرمی گیریم:

$$\Delta Y=\hat{Y}_{ab}^A-\hat{Y}_{ab}^B$$

$$\Delta N=\hat{N}_{ab}^A-\hat{N}_{ab}^B \qquad (۳)$$

$$\hat{Y}_{ab}(\theta)=\theta\,\hat{Y}_{ab}^A+(1-\theta)\,\hat{Y}_{ab}^B$$

$$\hat{N}_{ab}(\theta)=\theta\,\hat{N}_{ab}^A+(1-\theta)\,\hat{N}_{ab}^B$$

$$\hat{Y}(\theta)=\hat{Y}_a^A+\hat{Y}_b^B+\hat{Y}_{ab}(\theta)$$

که مقدار بهینهٔ θ واریانس $\hat{Y}_{ab}(\theta)$ را می نیمم می کند.

فرض می کنیم که D، ماتریس واریانس-کواریانس بردار $\Delta=[\Delta Y \quad \Delta N]^T$ (T نشان دهندهٔ ترانهاد بردار می باشد) و نیز تعریف می کنیم $c=cov\{\Delta,\,\hat{Y}(0)\}$

۱.۲ برآوردگر هارتلی و فولر – بورمیستر

هارتلی [۱] در سال (۱۹۶۲) و فولر و بورمیستر [۳] در سال (۱۹۷۲) برآوردگرهای دو چارچوبی زیر را برای برآورد مجموع پیشنهاد کردند:

$$\hat{Y}_H=\hat{Y}(0)+\theta_H\,\Delta Y$$

$$\hat{Y}_{FB}=\hat{Y}(0)+\beta^T\,\Delta \qquad (۴)$$

که در صورت ناویژه بودن ماتریس D مقادیر بهینهٔ $\theta_H=-c_1/D_{11}$ و $\beta=-D^{-1}c$ به ترتیب واریانس $\tilde{Y}_H$ و $\tilde{Y}_{FB}$ را می نیمم می کند. اما در عمل چون واریانس- کواریانس ها در c و D مجهول بوده و باید از روی نمونه برآورد شوند، به همین خاطر هر دوی $\hat{\theta}_H=-\hat{c}_1/\hat{D}_{11}$ و $\hat{\beta}=-\hat{D}^{-1}\hat{c}$ به $\hat{V}(\Delta Y)$ بستگی داشته و برآوردگرهای $\tilde{Y}_H(\hat{\theta}_H)$ و $\tilde{Y}_{FB}(\hat{\beta}_{FB})$ بطور کلی توابع خطی از y نخواهند بود و وزنها برای هر متغیر پاسخ باید بطور جداگانه محاسبه شوند. برای مثال اگر $\tilde{Y}_{H1}$ و $\tilde{Y}_{H2}$ و $\tilde{Y}_{H3}$ برآوردگرهای هارتلی برای تعداد کل بیمارانی که دچار تنگی نفس در گروههای سنی ۱۶-۰ و ۴۵-۱۷ و ۴۵ به بالا باشد آن گاه $\tilde{Y}_{H1}+\tilde{Y}_{H2}+\tilde{Y}_{H3}$ لزوماً مساوی برآورد هارتلی برای تعداد کل بیماران دچار تنگی نفس در جامعه نخواهد بود. در نتیجه فولر و بورمیستر بر آن شدند برآوردگری ارائه دهند که از یک مجموعه وزنهای یکسان برای همه متغیرهای پاسخ استفاده کند. آنها ابتدا حالت خاص انتخاب نمونه های تصادفی ساده از هر دو چارچوب را بررسی کردند. در این حالت برآوردگر زیر را پیشنهاد کردند:

$$\tilde{Y}_{FB,srs}=(N_A-\hat{N}_{ab,srs})\,\bar{y}_{a,srs}+\hat{N}_{ab,srs}\,\bar{y}_{ab,srs}+(N_B-\hat{N}_{ab,srs})\,\bar{y}_{b,srs} \qquad (۵)$$

بطوریکه :

$$\bar{y}_{ab,srs}=w\,\bar{y}_{ab}^A+(1-w)\,\bar{y}_{ab}^B$$

$$w=\frac{\tilde{n}_{ab}^A(1-f_B)}{\tilde{n}_{ab}^A(1-f_B)+\tilde{n}_{ab}^B(1-f_A)}$$

و $\hat{N}_{ab,srs}$ کوچکترین ریشه معادله درجه دو بفرم زیر است:

$$\tilde{n}_A+\tilde{n}_B)x^2-(\tilde{n}_A N_B N_A+\tilde{n}_{ab}^A N_A+\tilde{n}_{ab}^A N_A+\tilde{n}_{ab}^B N_B)x$$
$$+(\tilde{n}_{ab}^A+\tilde{n}_{ab}^B)N_A N_B=0 \qquad (۶)$$

فولر و بورمیستر[۳] در سال (۱۹۷۲) نشان دادند که $\hat{N}_{ab,srs}$ بعنوان یک برآوردگر حداکثر درستنمایی از N_{ab} بوده و نیز اسکینر[۷] در سال (۱۹۹۱) از $\hat{Y}_{FB,srs}$ تفسیری به صورت یک برآوردگر حداکثر درستنمایی داشت. بنابر این خواص مجانبی بهینه داشته و چون از وزنهای یکسانی برای هر متغیر پاسخ استفاده می کند به $\hat{Y}_{FB}$ ارجحیت دارد. اما این برآوردگر مستقیماً با طرحهای نمونه ای پیچیده بکار نمی رود. چون عموماً به صورتی ناسازگار با این نوع نمونه گیریها طراحی می شود. به همین دلیل آنها تصمیم گرفتند این برآوردگر را طوری اصلاح کنند که به برآوردگرهای سازگار با طرحهای پیچیده نائل آیند. آنها برآورد زیر را به دست آورند:

$$\tilde{Y}_{PML}(\theta)=\frac{N_A-\hat{N}_{ab}^{PML}(\theta)}{\hat{N}_a^A}\hat{Y}_a^A+\hat{N}_{ab}^{PML}(\theta)\frac{\hat{Y}_{ab}(\theta)}{\hat{N}_{ab}(\theta)}$$
$$+\frac{N_A-\hat{N}_{ab}^{PML}(\theta)}{\hat{N}_a^B}\tilde{Y}_b^B \qquad (۷)$$

بطوریکه $\hat{N}_{ab}^{PML}(\theta)$ تابعی از $\hat{N}_{ab}^A$ و $\hat{N}_{ab}^B$ و θ بوده و کوچکترین ریشه معادله درجه دو به فرم زیر است:

$$[\frac{\theta}{N_B}+\frac{(1-\theta)}{N_A}]x^2-[1+\theta\frac{\hat{N}_{ab}^A}{N_B}+(1-\theta)\frac{\hat{N}_{ab}^B}{N_A}]x$$
$$+\ \hat{N}_{ab}(\theta)=0 \qquad (۸)$$

اسکینر و رائو [۸] در سال (۱۹۹۶) انتخاب $\theta=\theta_P$ را برای می نیمم کردن واریانس مجانبی $\hat{N}_{ab}^{PML}(\theta)$ با

$$\theta_P=\frac{N_a N_B V(\hat{N}_{ab}^B)}{N_a N_B V(\hat{N}_{ab}^B)+N_b N_A V(\hat{N}_{ab}^A)} \qquad (۹)$$

پیشنهاد کردند.

همچنین آنها با خطی کردن ریشه های معادله درجه دو به صورت $\hat{N}_{ab}^{PML}(\theta_P)\approx\hat{N}_{ab}(\phi)$ راهی ساده تر برای محاسبهٔ برآوردگر شبه حداکثر درستنمایی ارائه دادند. مقدار بهینهٔ ϕ در این حالت عبارتست از:

$$\phi=\frac{V(\hat{N}_{ab}^B)}{V(\hat{N}_{ab}^A)+V(\hat{N}_{ab}^B)} \qquad (۱۰)$$

در عمل N_a و N_b و واریانس ها در روابط (۹) و (۱۰) مجهول بوده و باید از روی داده ها برآورد شوند در نتیجه برآوردگر حاصل عبارتست از: $\hat{Y}_{PML}(\hat{\theta}_P)$

۲.۲ برآوردگرهای چارچوب منفرد

بانکیر [۶] در سال (۱۹۸۶)، کالتون و اندرسون [۹] در سال (۱۹۸۶) و اسکینر [۷] در سال (۱۹۹۱) برای برآورد مجموع جامعه با در نظر گرفتن مشاهدات بگونه ای که آنها از یک چارچوب منفرد با وزنهای اصلاح شده برای مشاهدات در حوزه متداخل ab نمونه گیری شده اند، برآوردی ارائه دادند. وزنهای اصلاح شده برای برآوردگرهای تک چارچوبی اسکینر و کالتون و اندرسون به تشخیص واحدهای مشابه در نمونه ها نیازی ندارد. این وزنها عبارتند از:

$$w_i=\begin{cases}(\pi_i^A+\pi_i^B)^{-1} & i\in ab\\ (\pi_i^A)^{-1} & i\in a\\ (\pi_i^B)^{-1} & i\in b\end{cases}$$

آن گاه برآوردگر زیر را پیشنهاد کردند:

$$Y_{SF}=\sum_{i\in s_A}w_i y_i+\sum_{i\in s_B}w_i y_i \qquad (۱۱)$$

همانطور که ملاحظه می شود این برآوردگر از هیچ گونه اطلاعات کمکی در ارتباط با N_A و N_B استفاده نمی کند. از آنجایی که استفاده از اطلاعات کمکی باعث کاهش واریانس و افزایش دقت برآوردگر می شود (طبق نتایجی که بانکیر [۶] در سال (۱۹۸۶) به دست آورد) اسکینر و رائو [۸] در سال (۱۹۹۶) دو روش برای تصحیح این برآوردگر ارائه دادند که عبارتند از:

۱) روش تعدیل نسبتی [۱]

۲) روش رگرسیونی

اسکینر و رائو پیرو اثبات قضیه (۱) اسکینر [۷] در سال (۱۹۹۱) نشان دادند که فرآیند تعدیل به برآوردگر تجربی زیر همگراست:

$$\hat{Y}_{SFrake}=\frac{N_A-\hat{N}_{ab}^{rake}}{\hat{N}_a}\hat{Y}_a+\frac{N_B-\hat{N}_{ab}^{rake}}{\hat{N}_b}\hat{Y}_b$$
$$+\frac{\hat{N}_{ab}^{rake}}{\hat{N}_{abS}}\hat{Y}_{abS} \qquad (۱۲)$$

که در آن:

$$\hat{Y}_{abS}=\sum_{s_A}w_i\delta_i^B y_i+\sum_{s_B}w_i\delta_i^A y_i$$
$$\hat{N}_{abS}=\sum_{s_A}w_i\delta_i^B+\sum_{s_B}w_i\delta_i^A$$

و $\hat{N}_{ab}^{rake}$ کوچکترین ریشه معادله درجه دو بفرم زیر می باشد:

$$\hat{N}_{abS}x^2-[\hat{N}_{abS}(N_A+N_B)+\hat{N}_{aS}^A\hat{N}_{bS}^B]x$$
$$+\ \hat{N}_{abS}N_A N_B=0 \qquad (۱۳)$$

۱- Raking Ratio

تصحیح با N_A و N_B معلوم از طریق روش رگرسیونی برآورگر را نتیجه می دهد:

$$\hat{Y}S_{Freg} = \hat{Y}_S + \hat{\beta}^T [N_A - N_S^A,\ N_B - \hat{N}_S^B] \qquad (۱۴)$$

بطوریکه مقدار بهینهٔ $\hat{\beta}$ عبارتست از:

$$\hat{\beta}_S^T = Cov\{[\hat{N}_{AS}/V(\hat{N}_{AS}), \hat{N}_{BS}), \hat{N}_{BS}/V(\hat{N}_{BS})], \hat{Y}_S\}$$

تحت طرحهای نمونه گیری خودوزن، در حالتی که $\theta_S = w_i^B/(w_i^A + w_i^B)$ برای همه واحدها مقدار ثابتی است آن گاه $\hat{Y}_{abS} = \hat{Y}_{ab}(\theta_S)$ و $\hat{N}_{abS} = \hat{N}_{ab}(\theta_S)$ است و

$$N_A - \hat{N}_{AS} = (1-\theta_S)\Delta N\ ,\ N_B - \hat{N}_{BS} = -\theta_S \Delta N$$

آن گاه برآوردگر SF و $SFreg$ در این حالت به صورت زیر خواهند بود:

$$\hat{Y}_{SF} = \hat{Y}(0) + \theta_S \Delta Y = \hat{Y}(\theta_S)$$
$$\hat{Y}_{SFreg} = \hat{Y}_S + \beta_{S2}\Delta Y$$

که مقدار بهینه β_{S2} عبارتست از:

$$\beta_{S2} = -(c_2 + \theta_S D_{12})/D_{22}$$

در این حالت نیز با خطی کردن ریشه معادله (۱۳) به صورت $\hat{N}_{ab}^{rake} \approx \hat{N}_{ab}(\psi)$ می توان برآوردگر نسبتی تعدیل شده را به فرم ساده تری محاسبه کرد. مقدار بهینه ψ در این حالت عبارت از $\psi = N_b(\theta_S N_a + N_{ab})/(N_A N_A - N_{ab}^2)$ است و در صورتی که N_a/N و N_b/N هر دو نسبتاً کوچک باشند (حدود ۰/۱ یا ۰/۲) آن گاه $\psi \approx N_b/(N_a + N_b)$

۳. برآورد واریانس

اسکینر و رائو [۸] در سال (۱۹۹۶) با استفاده از بسط سری تیلور برآوردگری خطی برای برآورد واریانس برآوردگرهای چارچوب دوگانه ارائه دادند در این بخش ابتدا برآورد واریانس به روش جک نایف برای برآوردگرهای چارچوب دوگانه را ارائه داده و سپس نشان می دهیم که با برآوردگر خطی واریانس بطور مجانبی هم ارز می باشد.

نخستین بار کوئنولی [۱۱]جک نایف را به عنوان روشی بر اساس حذف هر بار یک مشاهده از مجموعه داده های اولیه و محاسبه مجدد برآوردگر با استفاده از بقیهٔ داده ها برای کاهش اریبی برآوردگر ضریب همبستگی پیاپی[۱] معرفی نمود. در مقاله ای در سال (۱۹۵۶) این روش را تعمیم داده و خواص کلی کاهش اریبی آن را در یک جامعه متناهی بیان کرد. سپس توکی[۲] [۱۲] پیشنهاد کرد که این روش علاوه بر کاهش اریبی، می تواند برای برآورد واریانس نیز استفاده شود. به نظر می رسد که استفاده از جک نایف در برآورد واریانس جوامع متناهی اولین بار توسط دوربین[۳] [۱۳] که استفاده از آن را در برآورد نسبتی مورد مطالعه قرار داد، بررسی شد. کتاب جامعی توسط میلر[۴][۱۴] در سال (۱۹۷۴) و در رساله ای از افرون[۵] [۱۵] در سال (۱۹۸۲) داده شده است.

فرض کنید چارچوب های A و B به ترتیب دارای H و L طبقه باشند بطوریکه طبقات h و l هر یک از آنها به ترتیب شامل N_h^A و N_l^B واحد بوده و در مجموعه $\hat{N}_h^A$ و $\hat{N}_l^B$ واحد اولیه نمونه گیری (psu) دارند که به ترتیب $\tilde{n}_h^A$ و $\tilde{n}_l^B$ واحد از آنها نمونه گیری می شود. بطوریکه $\tilde{n}_A = \sum_{h=1}^{H} \tilde{n}_h^A$ و $\tilde{n}_B = \sum_{l=1}^{L} \tilde{n}_l^B$ کل نمونه ای است که به ترتیب از چارچوبهای A و B انتخاب می شود.

حال فرض می کنیم که نمونه هایی به اندازه (≥ 2) $\tilde{n}_h^A$ و $\tilde{n}_l^B$ به ترتیب با احتمالات شمول متناسب با اندازه، $\tilde{\pi}_{hi}^A = \tilde{n}_h^A\, p_{hi}^A$ و $\tilde{\pi}_{lj}^B = \tilde{n}_l^B\, p_{lj}^B$ از چارچوبهای A و B انتخاب می شود. بطوریکه p_{hi}^A و p_{lj}^B به ترتیب برابر با احتمال انتخاب واحد i ام و j ام در طبقات h و l متناسب با اندازه psu ها بوده و $\sum_i p_{hi}^A = 1$ و $\sum_j p_{lj}^B = 1$ ($p_{hi}^A = u_i/\sum_A u_i$ و $p_{lj}^B = u_j/\sum_B u_j$ که u_i اندازه تقریبی I امین psu می باشد)

با در نظر گرفتن دو بردار A و B به صورت :

$$A = (Y_a, Y_{ab}, N_{ab}, N_A)^T \quad \text{و} \quad B = (Y_b, Y_{ab}, N_{ab}, N_B)^T$$

مجموع جامعه عبارت خواهد بوداز:

$$Y = N\bar{Y} = N_g(\bar{A}, \bar{B}) = N_\tau$$

بطوریکه $\bar{A} = A/N_A$ و $\bar{B} = B/N_B$ آن گاه $\bar{Y}$ به صورت زیر قابل بیان است :

$$g(\bar{A}, \bar{B}) = \bar{Y} = (N_A/N)(\bar{A}_1 + \theta \bar{A}_2) + (N_B/N)(\bar{B}_1 + (1-\theta)\bar{B}_2)$$

که θ مقدار ثابت می باشد. میانگین های $\bar{A}$ و $\bar{B}$ به ترتیب با $\hat{\bar{A}}$ و $\hat{\bar{B}}$ برآورد شده و $\hat{\tau} = g(\hat{\bar{A}}, \hat{\bar{B}})$ نیز برآوردی از τ می باشد.

۱- Serial Correlation Cofficient

۲- Tukey
۳- Durbin
۴- Miller
۵- Efron

برآوردگرهای چارچوب دوگانه در این حالت می توانند به صورت تابعی از $\hat{\tau}$ بیان شوند. به عنوان مثال $\tilde{Y}_H(\theta_H) = N\, g(\hat{\bar{A}}, \hat{\bar{B}})$. برآوردگر PML نیز به ازای θ ثابت، می توانند به صورتی مشابه به عنوان تابعی از $\hat{\bar{A}}$ و $\hat{\bar{B}}$ و با توجه به اینکه $\hat{N}_{ab}^{PML}(\theta)$ تابعی از $\hat{\bar{A}}_۴$ و $\hat{\bar{B}}_۳$ است، و سپس با استفاده از رابطه (۷) بیان شود.

با فرض اینکه $g(\hat{\bar{A}}, \hat{\bar{B}})$ تابعی مشتق پذیر و دارای مشتقات اول و دوم پیوسته و کراندار در یک همسایگی از $(\bar{A}, \bar{B})$ باشد، برآوردگر خطی واریانس $\hat{\tau}$ عبارست از:

$$v_L(\hat{\tau}) = g_A^T(\hat{\bar{A}}, \hat{\bar{B}})\, S^A\, g_A(\hat{\bar{A}}, \hat{\bar{B}}) + g_B^T(\hat{\bar{A}}, \hat{\bar{B}})\, S^B\, g_B(\hat{\bar{A}}, \hat{\bar{B}}) \qquad (۱۵)$$

که $g_A(\bar{A}, \bar{B})$ و $g_B(\bar{A}, \bar{B})$ به ترتیب بردارهای ۴ بعدی از مشتق های اول $g(\circ)$ نسبت به مؤلفه های $\hat{\bar{A}}$ و $\hat{\bar{B}}$ بوده و S^A و S^B به ترتیب برابر با ماتریس واریانس - کواریانس نمونه ای بردارهای $\hat{\bar{A}}$ و $\hat{\bar{B}}$ می باشند. برای محاسبه برآورد جک نایف واریانس $\hat{\tau}$ فرض می کنیم $\hat{\tau}_{hi}^A$ برآوردگری به فرم $\hat{\tau}$ باشد که بعد از حذف مشاهدات i امین psu نمونه طبقه h چارچوب A به دست آمده و $\hat{\tau}_{(hi)}^A = g(\hat{\bar{A}}, \hat{\bar{B}}_{(lj)})$ که $\hat{\bar{B}}_{(lj)}$ برآورد $\bar{B}$ محاسبه شده بعد از حذف j امین psu نمونه طبقه l چارچوب B است. برآورد جک نایف واریانس $\hat{\tau}$ با توجه به استقلال نمونه ها عبارتست از :

$$v_J(\hat{\tau}) = \sum_{h=۱}^{H} \frac{\tilde{n}_h^A - ۱}{\tilde{n}_h^A} \sum_{i=۱}^{\tilde{n}_h^A} (\hat{\tau}_{(hi)}^A - \hat{\tau})^۲ + \sum_{l=۱}^{L} \frac{\tilde{n}_l^B - ۱}{\tilde{n}_l^B} \sum_{j=۱}^{\tilde{n}_l^B} (\hat{\tau}_{(lj)}^B - \hat{\tau})^۲ \qquad (۱۶)$$

رائو و لهر[۱۶] طی قضیه ای ثابت کردند که $v_J(\hat{\tau})$ و $v_L(\hat{\tau})$ بطور مجانبی هم ارزند بعبارت دیگر :

۱) $v_L(\hat{\tau}) = V(\hat{\tau}) + o_p(\tilde{n}^{-۱})$

۲) $v_J(\hat{\tau}) = v_L(\hat{\tau}) + o_p(\tilde{n}^{-۱})$

جک نایف با توابع هموار قابل کاربرد است. در بین برآوردگرهای دو چارچوبی، برآوردگرهای SF و $SFrake$ بصورت توابع هموارای از میانگین جوامع قابل بیان هستند. دیگر برآوردگرها نیز در حالتی که پارامترهای θ_p و θ_H و β_{FB} و $\beta_{S۲}$ ثابت بوده و از روی داده ها برآورد نشوند، نیز توابع هموارای از میانگین جوامع خواهند بود. بنابر این قضیه فوق بیان می دارد که برآوردگر جک نایف واریانس برای شکل بهینه هر برآوردگر سازگار است.

٤. نتایج شبیه سازی

در این بخش با استفاده از شبیه سازی به مطالعه خواص تجربی برآوردگرهای واریانس، می پردازیم.

فرض می شود که جامعه نامتناهی بوده و طرح نمونه گیری برای هر چارچوب یک طبقه دارد. یک نمونه خوشه ای دو مرحله ای با $\tilde{n}_A$ خوشه و m عنصر از هر خوشه به عنوان نمونه حاصل از چارچوب A $(n_A = \tilde{n}_A.m)$ و یک نمونه تصادفی ساده با n_B مشاهده به عنوان نمونه حاصل از چارچوب B تولید می شود. همانطور که گفته شد فرض بر این است که جامعه نامتناهی بوده و $\frac{N_a}{N}$ و $\frac{N_b}{N}$ را به ترتیب با γ_a و γ_b(معلومند) جایگزین می کنیم. نمونه حاصل از چارچوب A شامل مقادیر

$$\{(y_{ij}, m_{ai}), i = ۱, \cdots, n_A, j = ۱, \cdots, m\}$$

می باشد که مقادیر m_{ai} عبارت از تعداد عناصر نمونه گیری شده از I امین خوشه که متعلق به حوزه a است و y_{ij} نیز مقدار مرتبط با j امین عنصر در i امین خوشه می باشد. نمونه حاصل از چارچوب B شامل

$$\{(y_j, n_b), j = ۱, \cdots, n_B\}$$

است که n_b عبارت از تعداد نمونه ای متعلق به حوزه b است و y_j نیز مقدار مرتبط با j امین عنصر است. برای به دست آوردن نمونه حاصل از چارچوب A ، ابتدا m_{ai} را از یک توزیع بتا – دو جمله ای بصورت زیر تولید می کنیم:

۱) $\tilde{\gamma}_{ai}$ نسبن مورد انتظار مشاهداتی از خوشه ای i ام که در حوزه a قرار دارد، از توزیع بتا با پارامترهای $a_۱$ و $a_۲ = a_۱(۱ - \gamma_a - \gamma_b)/\gamma_a$ تولید می شوند بطوریکه:

$$E(\tilde{\gamma}_{ai}) = \frac{\gamma_a}{۱ - \gamma_b}$$

$$V(\tilde{\gamma}_{ai}) = \frac{\gamma_a^۲ (۱ - \gamma_a - \gamma_b)}{\{[a_۱(۱ - \gamma_b) + \gamma_a)](۱ - \gamma_b)^۲\}}$$

بنابر این مقدار کوچکتر $a_۱$ تغییر پذیری $\tilde{\gamma}_{ai}$ را از یک خوشه به خوشهٔ دیگر افزایش می دهد.

۲) به ازای مقداری از $\tilde{\gamma}_{ai}$، m_{ai} را از یک توزیع دو جمله ای با پارامترهای m و $\tilde{\gamma}_{ai}$ تولید می کنیم.

۳) مراحل (۱) و (۲) را مستقلاً تکرار کرده تا $\{m_{ai}, i = ۱, \cdots, \tilde{n}_A\}$ به دست آید.

۱- Lohr

به ازای مقدار از m_{ai} مقادیر y_{ij} از مدل خطای لانه ای[۱] زیر تولید می شوند:

$$y_{ij} = \mu_a + a_{ai} + \varepsilon ij \qquad j = ۱,\cdots,m_{ai}\}$$
$$y_{ij} = \mu_{ab} + a_{abi} + \varepsilon ij \qquad j = m_{ai}+۱,\cdots,m$$

از مقادیر ۹= μ_a و ۱۰ = μ_{ab} و ۱۱= μ_b برای میانگین جامعه در حوزه های a و ab و b استفاده می کنیم. ε_{ij} ها متغیرهای تصادفی مستقل و هم توزیع (iid) با $[\varepsilon_{ij} \sim N[۰,\sigma^۲(۱-\rho)]$ به ازای ۱= $\sigma^۲$ و ۰/۱ = ρ و نیز مستقل از $(\alpha_{ai},\alpha_{abi},m_{ai})$ هستند. $(\alpha_{ai},\alpha_{abi})$نیز از یک توزیع نرمال دو متغیره با میانگین **0** و ماتریس کواریانس ۲×۲ با عناصر غیر قطری $\rho\sigma^۲$ و عناصر غیرقطری $\rho\delta\sigma^۲$ به ازای ۰/۵ = δ به دست می آیند. این مدل، یک همبستگی یکسان ρ را درون حوزه های a و ab و یک همبستگی یکسان $\rho\delta$ را بین حوزه ها تضمین می کند.

برای به دست آوردن نمونه حاصل از چارچوب B ابتدا n_b را از یک توزیع دو جمله ای با پارامترهای n_B و $\gamma = \frac{\gamma_b}{۱-\gamma_a}$ تولید می کنیم. به ازای مقداری از n_B مقادیر نمونه ای $\{y_j, j=۱,\cdots,n_B\}$ از چارچوب B را از مدل

$$y_j = \mu_b + \delta_j \qquad j=۱,\cdots,n_b$$
$$y_j = \mu_{ab} + \delta_j \qquad j = n_b+۱,\cdots,n_B$$

تولید می کنیم بطوریکه δ_jها متغیرهای تصادفی مستقل و هم توزیع با $\delta_j \sim N(۰,\sigma^۲)$ می باشند.

به ازای مقادیر مختلف $\tilde{n}_A$ ، n_B ، $a_۱$، γ_a و γ_b ۱۰۰۰۰ = R بار این شبیه سازی را تکرار می کنیم. در هر تکرار برآوردگرهای $\hat{Y}_H\,(\hat{\theta}_H)$ ، $\hat{Y}_{FB}\,(\hat{\beta}_{FB})$ و $\hat{Y}_{PML}\,(\hat{\theta}_P)$ و برآوردگرهای تک چارچوبی $\hat{Y}_{SF}$ ، $\hat{Y}_{SFrake}$ و $\hat{Y}_{SFreg}\,(\hat{\beta}_{S۲})$ را با استفاده از مقادیر بهینهٔ $\hat{\theta}_H$ و $\hat{\beta}_{FB}$ و $\hat{\theta}_P$ و $\hat{\beta}_{S۲}$ محاسبه می کنیم. سپس میانگین توان دوم خطای تجربی هر برآوردگر $(EMSE)$ را به صورت زیر محاسبه می کنیم:

$$EMSE = \frac{۱}{R}\sum_{r=۱}^{R}(\hat{Y}_r - Y)^۲$$

که $\hat{Y}_r$ عبارت از مقدار $\hat{Y}$ برای r امین تکرار شبیه سازی و R تعداد تکرارهای شبیه سازی می باشد. لازم به ذکر است که تمام محاسبات با استفاده از نرم افزار *SPLUS 2000* صورت گرفته است. چون جامعه را نامتناهی فرض کردیم و $N \to \infty$ ، بجای مجموع Y با میانگین جامعه Y/N و بجای برآورد مجموع $\hat{Y}$ با برآورد میانگین جامعه $\hat{Y}/N$ کار می کنیم. برای مثال:

$$\frac{Y}{N} = \gamma_a\mu_a + \gamma_b\mu_b + \gamma_{ab}\mu_{ab} \qquad (\gamma_{ab} = ۱-\gamma_a-\gamma_b)$$

و برآوردگر میانگین جامعه در حوزه ها عبارتند از:

$$\frac{\hat{Y}_a}{N} = \frac{۱-\gamma_b}{n_A}\sum_{s_a} y_{ij} \qquad \frac{\hat{Y}_{ab}^A}{N} = \frac{۱-\gamma_b}{n_A}\sum_{s_{ab}^A} y_{ij}$$

$$\frac{\hat{Y}_b}{N} = \frac{۱-\gamma_a}{n_B}\sum_{s_b} y_j \qquad \frac{\hat{Y}_{ab}^B}{N} = \frac{۱-\gamma_a}{n_B}\sum_{s_{ab}^B} y_j$$

(لازم به ذکر است که برآوردگرهای فوق تحت طرح نمونه گیری خودوزن تعریف شده اند)

$EMSE$ خود منجر به خطای مونت کارلو شده و واریانس مونت کارلوی آن بصورت زیر برآورد می شود:

$$s^۲(EMSE) = \frac{۱}{R-۱} = \sum_{r=۱}^{R}(\hat{Z}_r - \bar{\hat{Z}})^۲$$

که $\hat{Z} = (\hat{Y}_r - Y)^۲$ و $\bar{\hat{Z}} = \sum_r \hat{Z}_r / R$ می باشد با گرفتن جذر از رابطه فوق، خطای معیار مونت کارلو $s(EMSE)$ را می توان محاسبه کرد. نتایج شبیه سازی برای میانگین توان دوم خطای تجربی $EMSE$ و خطای استاندارد مونت کارلوی نظیرش $s(EMSE)$ تحت طرحهای نمونه گیری فوق به ازای ۱۰= $\tilde{n}_A$ ، ۳۰= m ، ۱۰۰= n_B و ۲۰= $\tilde{n}_A$، ۳۰= m ، ۲۰۰= n_B و ۱= $a_۱$ و نیز مقادیر معین γ_a و γ_b در انتها آورده شده است. در مقایسه با برآوردگرهای دو چارچوبی، برآوردگر چارچوب منفرد افزایش قابل ملاحظه ای را در $EMSE$ وقتی $\gamma_a \le \gamma_b$ یا n_A خیلی بزرگتر از n_B باشد، نشان می دهد. برای هر شش برآوردگر چارچوب دو گانه، دو برآوردگر واریانس خطی و جک نایف، را محاسبه می کنیم. برای محاسبهٔ برآورد جک نایف واریانس، در صورت وابسته بودن پارامترهای $\hat{\theta}_H$، $\hat{\beta}_{FB}$، $\hat{\theta}_P$ و $\hat{\beta}_{S۲}$ به ماتریسهای واریانس – کواریانس نمونه، $S_{(hi)}^A$ و $S_{(hi)}^B$ را با استفاده از یک جک نایف جداگانه درون هر تکرار جک نایف محاسبه کرده و سپس برآورد جک نایف واریانس برآوردگرها را با استفاده از رابطهٔ (۱۷) به دست می آوریم. نایج شبیه سازی نشان می دهد که اریبی نسبی برآوردگر خطی واریانس در صورت افزایش حجم نموه میل به کاهش دارد. به هر حال، در نمونه های کوچکتر، برآوردگر خطی واریانس، واریانس را کم برآورد می کند در حالی که جک نایف آن را بیش برآورد می کند. همچنین نتایج حاصل از خطای معیار نسبی حاکی از آن است که برآوردگر جک نایف واریانس از پایداری کمتری نسبت به برآورد خطی برخوردار است. بعنوان مثال وقتی ۱۰= $\tilde{n}_A$ خطای معیار نسبی

۱- Nested Error Model

برآوردگر جک نایف واریانس دو برابر خطای معیار نسبی برآورگر خطی است. با این وجود با افزایش حجم نمونه این پایداری بهبود می یابد. با استفاده از فاصله اطمینان $\hat{\tau} \pm ۱/۹۶\sqrt{\hat{V}(\hat{\tau})}$ احتمال پوشایی تجربی جک نایف بیشتر از خطی است. با استفاده از توزیع t با $\widetilde{n}_A$ - ۱ درجه آزادی بجای ۱/۹۶، جک نایف احتمال پوشایی نزدیک ۰/۹۵ دارد در حالی که روش خطی به وضوح احتمال پوشایی کمتری دارد.

۵. نتیجه گیری

برآوردگرهای مجموع در بررسی های دو چارچوبی را بطور خلاصه مطرح و مقایسه کردیم. این برآوردگرها قابل تعمیم به بررسی هایی با بیش از دو چارچوب نیز می باشند. برآوردگر جک نایف واریانس را بطور نظری توجیه کرده با استفاده از شبیه سازی نشان دادیم که دارای اریبی کوچکتری از برآوردگر خطی واریانس است. همچنین برآوردگر جک نایف به وضوح دارای احتمال پوشایی بالاتری از برآوردگر خطی است خصوصاً وقتی از صدک t استفاده می شود احتمال پوشانی جک نایف خیلی نزدیک ۰/۹۵ است.

جک نایف معمولاً با توابع غیر خطی نظیر نسبت مجموع دو جامعه بکار می رود. در این حالت مشتقات جزئی که در محاسبهٔ برآورد خطی چنین مقادیر غیر خطی استفاده می شود، در بررسی های دو چارچوبی پیچیده تر از تک چارچوبی است. در صورتی که با استفاده از جک نایف می توان از این محاسبات پیچیده اجتناب کرد. برآوردگرهای جک نایف اخیراً در اداره آمار کانادا برای محاسبه برآورد واریانس در آمارگیری وسیع ملی کودکان و جوانان[۱] استفاده می شود. روشهای بازنمونه گیری دیگر برآورد واریانس نظیر تکرار نامتعادل پاسخ[۲] *(BRR)* و بوتسترپ[۳] نیز می توانند به موازات جک نایف در این گونه طرحها مورد استفاده قرار گیرند. حسن این روش ها این است که برخلاف جک نایف با توابع غیر هموار نظیر میانه نیز قابل کاربرد هستند.

برآوردگر	*EMSE*	اریبی نسبی		احتمال پوشایی نرمال		احتمال پوشایی *t*	
		L	*J*	*L*	*J*	*L*	*J*
H	۷/۱۶	-۱۹/۹۴ (۰/۷۸)	۱۵/۴۵ (۱/۹۰)	۰/۸۸۷	۰/۹۱۹	۰/۹۲۳	۰/۹۴۳
FB	۶/۸۳	-۱۲/۷۵ (۰/۷۷)	۱۰/۴۶ (۱/۶۴)	۰/۹۰۱	۰/۹۲۱	۰/۹۳۱	۰/۹۴۶
PML	۶/۵۸	-۹/۹۱ (۰/۸۲)	۱۱/۴۶ (۱/۸۷)	۰/۹۱۲	۰/۹۲۴	۰/۹۳۹	۰/۹۵۰
PMLlin	۶/۸۳	-۱۳/۲۵ (۰/۸۰)	۱۲/۲۱ (۱/۸۱)	۰/۸۹۹	۰/۹۱۹	۰/۹۳۰	۰/۹۴۴
SF	۱۱/۲۶	۴/۱۷ (۰/۴۷)	۴/۱۷ (۰/۴۷)	۰/۹۴۸	۰/۹۴۸	۰/۹۷۲	۰/۹۷۲
SFreg	۶/۹۷	-۱۳/۷۳ (۰/۸۲)	۱۱/۴۲ (۱/۷۸)	۰/۸۹۷	۰/۹۱۸	۰/۹۲۸	۰/۹۴۳
SFrake	۶/۵۵	-۰/۱۱ (۱/۰۸)	۴/۸۵ (۱/۱۹)	۰/۹۱۸	۰/۹۲۳	۰/۹۴۴	۰/۹۵۰
H	۳/۵۸	-۱۵/۹۷ (۰/۴۱)	۱/۸۰ (۰/۷۴)	۰/۹۰۶	۰/۹۲۸	۰/۹۲۳	۰/۹۴۴
FB	۳/۵۲	-۱۲/۶۹ (۰/۴۰)	-۰/۶۳ (۰/۶۷)	۰/۹۱۳	۰/۹۲۸	۰/۹۳۰	۰/۹۴۳
PML	۳/۴۴	-۱۱/۲۶ (۰/۴۲)	-۱/۱۰ (۰/۷۴)	۰/۹۱۹	۰/۹۳۰	۰/۹۳۵	۰/۹۴۵
PMLlin	۳/۵۲	-۱۳/۱۰ (۰/۴۲)	-۰/۰۵ (۰/۷۲)	۰/۹۱۱	۰/۹۲۸	۰/۹۲۸	۰/۹۴۳
SF	۵/۸۳	-۰/۶۹ (۰/۲۳)	-۰/۶۹ (۰/۲۳)	۰/۹۴۴	۰/۹۴۴	۰/۹۵۸	۰/۹۵۸
SFreg	۳/۵۷	-۱۲/۹۲ (۰/۴۲)	۰/۰۸ (۰/۷۲)	۰/۹۱۳	۰/۹۲۸	۰/۹۲۸	۰/۹۴۲
SFrake	۳/۴۸	-۵/۳۳ (۰/۵۳)	-۳/۰۷ (۰/۵۵)	۰/۹۲۵	۰/۹۲۹	۰/۹۴۰	۰/۹۴۴

۱- National Longitudinal Survey of Children and Youth
۲- Balanced Repeated Replication
۳- Bootstrap

لازم به توجه است که در این جدول نتایج تنها به ازای ۱=a_1 و ۰/۱ = γ_aو ۰/۲ = γ_b ارائه شده است. در بخش بالایی، ۱۰= $\tilde{n}_A$ و ۱۰۰= n_B و در بخش پائینی ۲۰= $\tilde{n}_A$ و ۲۰۰ = n_B می باشد. $EMSE$ عبارت از میانگین مربعات خطای مونت کارلو برای ۱۰۰۰۰ بار تکرار شبیه سازی است. اریبی نسبی برای روش های خطی (L) و جک نایف (J) بصورت $(EV\text{-}EMSE)EMSE$ ۱۰۰ محاسبه شده است، بطوریکه EV عبارت از میانگین ۱۰۰۰۰ برآورد واریانس برای آن روش است. احتمال پوشایی هر روش با استفاده از $[SE\ (Estimate)$ ۱/۹۶ + $Estimate]$ و همچنین یک صدک t با ۱- $\tilde{n}_A$ درجه آزادی محاسبه شده است. خطای معیار نسبی برای هر روش (= انحراف معیار۱۰۰۰۰ برآورد واریانس / $\sqrt{EMSE}$) در پرانتز در زیر اریبی نسبی داده شده است. همچنین شبیه سازی را به ازای مقادیر مختلف a_1، (ρ,δ) و (n_B و $\tilde{n}_A$) بررسی کرده ایم. و نتایج شبیه نتایج بالا است.

مراجع

[1] Hartley, H. O., 1962. Multiple Frame Surveys, in Proceedings of the Social Statistics Section*, American Statistical Association*. 203-206.

[2] Lund, R, E., 1968. Estimators in Multiple Frame Surveys, in *Priccedings of the Social Statistics Section, American Statistical Association*. 282-288.

[3] Fuller, W. A., & Burmeister, L. F., 1972. Estimators for Samples Selected From two Overlapping Farmes *Proccedings of the Social Statistics Section, American Statistical Association.* 245-249.

[4]Vogel, Frederic A., 1975. Survey With Overlapping Frame-Problems in Application, *Proccedings of the Social Statistics Section, American Associaton*. 694-699.

[5] Ford, B. L., & Bosecker, R. R., 1979. Multiple Frame Estimation With Stratified Overlap Domain *Proccedings of the Social Statistics Section, American Statistical Association*. 219-224.

[6] Bankier, M. D., 1986. Estimators Based on Several Stratified Samples with Applications to Multiple Frame Surveys, *Journal of the American Statistical Association,* 81, 1074-1079.

[7] Skinner, C. J., 1991, On the Efficiency of Raking Ration Estimation for Multiple Frame Surveys'' *Journal of the American Statistical Association,* 86. 779-784.

[8] Skinner, C. J., & Rao, J. N. K., 1996. Estimation in Dual Frame Survey with Complex Designs, *Journal of the American Statistical Association*, 91. 349-356.

[9] Kalton, G., & Anderson, T. W., 1986.''Sampling Rare Populations'' *Journal of the Royal Statistcal Society*, Ser.A, 149. 65-82.

[10] Sarndal, C-E., Swensson, B., & Wretman, J. 1992., Model Assisted Survey Sampling, New York: Speinger-Verlag.

[11] Quenoulle, M. H., 1956, Notes on Bias in Estimation, *Biometrika*, 43. 353-360.

[12] Tukey, J. W., 1958. Bias and Confidence in Note Quite Large Samples, *Annals of Mathematical Statistics*, 29. 614.

[13] Durbin, J., 1959, *A Note on the Application of Quenouille's Method of Bias Reduction To the Estimation of Ratios*, *Biometrika*, 46. 477-480.

[14] Miller, R. G. Jr, 1974, The Jackknife-A Review , *Biometrika*, 61, 1-15.

[15] Efron, B., 1982. The Jackknife, *The Bootstrap, and The other Resampling Plans*, Philadelphia: Society for Industrial and Applied Mathematics.

[16] Rao, J. N. K., & Lohr, Sharon, 2000. Inference from Dual Frame Survey, *Journal of The American Statistical Association*, 95. 271-280.