فصلنامه‌ی گزیده‌مطالب آماری
سال ۱۶، شماره‌ی ۲، تابستان ۱۳۸۴، صص ۱۳-۱

معرفی چند براوردگر
در آمارگیری‌های نمونه‌ای چندچارچوبی

ماریا رهبر

مرکز آمار ایران

چکیـــده. در ایـن مقالـه چنـدین براوردگـر بـرای آمـاره‌ی مجمـوع یـک صـفت در آمارگیری‌های مبتنی بر دو چارچوب، معرفی و مقایسه شده‌اند. نتـایج حاصـل از حالـت دوچارچوبی، قابل تعمیم به حالت چندچارچوبی است.

۱- مقدمه

چارچوب آماری یا فهرست واحدهای آمارگیری، یکی از مهم‌ترین ارکان اجرای یک طـرح آمارگیری نمونه‌ای است. در صورت وجود مشکلاتی نظیر عـدم پوشـش کامـل واحـدهای جامعه‌ی آماری و وجود واحدهای آمارگیری تکراری، خطای براورد پارامترها افـزایش پیـدا می‌کند. یکی از راه‌های مقابله با خطاهای چارچوبی، استفاده از دو یا چند چارچوب آمـاری مناسب است. موارد زیر از جمله‌ی حالت‌هایی هستند که اسـتفاده از دو یـا چنـد چـارچوب آماری در آن‌ها توصیه می‌شود:

- گاهی ممکن است یک چارچوب آماری، تمامی واحدهای جامعه‌ی مورد مطالعـه را پوشش ندهد، اما امکان دست‌یابی به پوشش کامل، بـا تلفیقـی از دو یـا چنـد چارچوب غیرکامل، وجود داشته باشد. در چنین حالتی می‌توان بـه‌منظـور تـأمین پوشش کامل، به‌طور همزمان از چند چارچوب استفاده کرد.

واژگان کلیدی: آمارگیری‌های نمونه‌ای چندچارچوبی؛ چارچوب ناحیه‌ای؛ چارچوب فهرستی.

- گاهی نیز ممکن است یک چارچوب آماری، پوشش مناسب را فراهم کند، اما چارچوب ناقص دیگری موجود باشد که هزینه‌ی آمارگیری از آن، نسبت به چارچوب کامل، کمتر باشد. بدین ترتیب می‌توان با هزینه‌ای ثابت، اندازه‌ی نمونه را افزایش داد.
- تغییرپذیری بسیار زیاد صفت یا صفات مورد بررسی در جامعه‌ی مورد مطالعه، در صورتی که از نمونه‌گیری طبقه‌بندی استفاده نشود، منجر به چولگی شدید می‌گردد. لذا چنان‌چه بتوان فهرستی از واحدهای آمارگیری با صفت یا صفات مورد نظر را تهیه کرد، با تلفیق دو چارچوب ناحیه‌ای (area frame) و فهرستی (list frame) می‌توان کارایی براوردگر را بالاتر برد.
- معمولاً چارچوب‌های فهرستی با گذشت زمان، به‌سرعت تغییر می‌کنند و منبعی برای بروز خطاهای غیر نمونه‌گیری می‌شوند، اما چارچوب‌های ناحیه‌ای از پایداری بیش‌تری برخوردارند. لذا با تلفیق یک چارچوب فهرستی و یک چارچوب ناحیه‌ای می‌توان از مزایای هر دو بهره برد.

شاید بتوان گفت برای اوّلین بار، در سال ۱۹۴۹ دو چارچوب ناحیه‌ای و فهرستی همزمان در آمارگیری از فروشگاه‌های خرده‌فروشی، توسط دفتر سرشماری آمریکا مورد استفاده قرار گرفت. بعدها نحوه‌ی استفاده از این روش توسط هارتلی [۳] بهبود یافت و به براورد پارامترها به‌روش خاصی که اساساً خود وی بنیان‌گذار آن بود، توجه خاصی مبذول شد.

اخیراً استفاده از آمارگیری‌های دوچارچوبی، که یکی از چارچوب‌ها را فهرست شماره‌ی تلفن خانوارها و دیگری را یک چارچوب ناحیه‌ای تشکیل می‌دهد، در برخی کشورها مرسوم شده است. به‌علاوه، استفاده از آمارگیری‌های دوچارچوبی، جایگاه خاصی را در بسیاری از طرح‌های آمارگیری کشاورزی به دست آورده و اخیراً سازمان خواربار و کشاورزی ملل متحد (فائو) [۱] استفاده از آن را در طرح‌های آمارگیری کشاورزی توصیه کرده است.

در این مقاله به معرفی و مقایسه‌ی چند براوردگر برای مجموع صفت پرداخته می‌شود. برای معرفی این براوردگرها دو حالت مختلف در نظر گرفته می‌شود، ولی پیش از

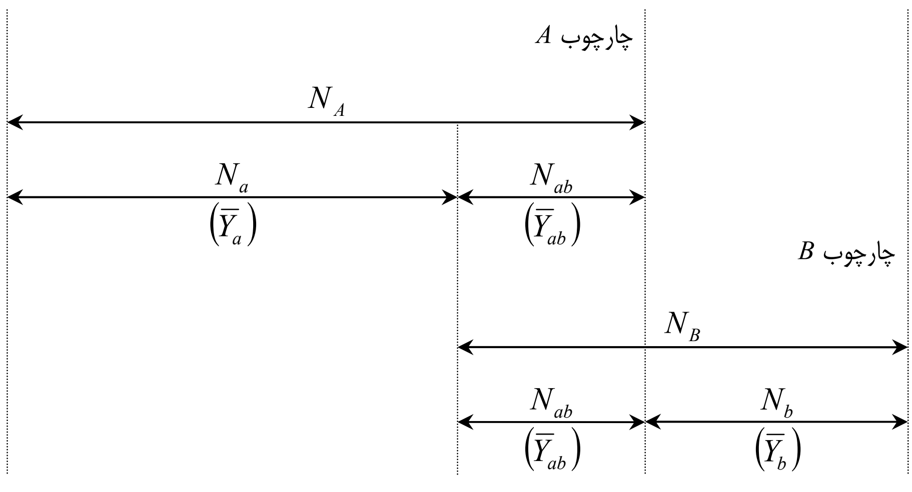
آن، اصول کلّی و نمادهای مورد استفاده توضیح داده می‌شود.

۲- اصول کلّی و نمادها

ابتدا فارغ از در نظر گرفتن حالتی خاص، اصول و نمادهای مورد استفاده در این مقاله توضیح داده می‌شود. دو چارچوب A و B را در نظر بگیرید که از هر یک، نمونه‌ای استخراج شده است. برای این دو چارچوب، فرض‌های زیر را در نظر می‌گیریم:

- هر واحد آماری جامعه‌ی مورد مطالعه، دست‌کم به یکی از دو چارچوب تعلّق دارد؛
- امکان تعیین تعلّق یا عدم تعلّق هر واحد نمونه به یک چارچوب وجود دارد.

با در نظر گرفتن دو فرض فوق می‌توان واحدهای نمونه را به سه حوزه تقسیم کرد: حوزه‌ی a که شامل واحدهای آماری چارچوب A است و نه B، حوزه‌ی b که شامل واحدهای واحد آماری چارچوب B است و نه A، و حوزه‌ی ab که شامل واحدهای مشترک در هر دو چارچوب A و B است. جدول ۱ نمادهای به کار گرفته‌شده در این مقاله را نشان می‌دهد، که در آن، n'_{ab} و n''_{ab} اندازه‌ی زیرنمونه‌هایی را نشان می‌دهند که هر



شکل ۱ــ نمادهای به کار گرفته‌شده برای طرح‌های آمارگیری دوچارچوبی

جدول ۱ـ نمادهای به کار گرفته‌شده برای طرح‌های آمارگیری دوچارچوبی

	چارچوب		حوزه		
	A	***B***	***a***	***b***	***ab***
اندازه‌ی جامعه	N_A	N_B	N_a	N_b	N_{ab}
اندازه‌ی نمونه	n_A	n_B	n_a	n_b	n'_{ab} و n''_{ab}
مجموع در جامعه	Y_A	Y_B	Y_a	Y_b	Y_{ab}
مجموع در نمونه	y_A	y_B	N_a	y_b	y'_{ab} و y''_{ab}
میانگین جامعه	$\overline{Y}_A$	$\overline{Y}_B$	$\overline{Y}_a$	$\overline{Y}_b$	$\overline{Y}_{ab}$
میانگین نمونه	$\overline{y}_A$	$\overline{y}_B$	$\overline{y}_a$	$\overline{y}_b$	$\overline{y}'_{ab}$ و $\overline{y}''_{ab}$
هزینه‌ی نمونه‌گیری هر واحد آماری	C_A	C_B			

یک به‌ترتیب در نمونه‌ی حاصل از چارچوب A و چارچوب B کـه درون حـوزه‌ی مـشترک ab هستند، قرار می‌گیرند. میانگین متناظر با این زیرنمونه‌ها با $\overline{y}'_{ab}$ و $\overline{y}''_{ab}$ نمـایش داده می‌شود. نمادهای جدول ۱ را می‌توان به‌صورت شکل ۱ نیز نمایش داد.

تغییـر متغیـر زیـر را بـرای هـر یـک از دو چـارچوب A و B در نظـر مـی‌گیـریم. در چارچوب A،

$$u_i = \begin{cases} y_i, & \text{اگر } i \text{امین واحد، متعلّق به حوزه‌ی } a \text{ باشد} \\ py_i, & \text{اگر } i \text{امین واحد، متعلّق به حوزه‌ی } ab \text{ باشد} \end{cases}$$

و در چارچوب B،

$$u_i = \begin{cases} y_i, & \text{اگر } i \text{امین واحد، متعلّق به حوزه‌ی } b \text{ باشد} \\ qy_i, & \text{اگر } i \text{امین واحد، متعلّق به حوزه‌ی } ab \text{ باشد} \end{cases}$$

p و q دو عدد ثابت هستند به‌طوری که $p+q=1$. با تغییر متغیر فوق، دو چارچوب A و B به دو طبقه‌ی جدا از هم، هر یک با اندازه‌های N_A و N_B تبدیل می‌شـوند و بـه ایـن ترتیب، مجموع y_iها (Y) در جامعه‌ای با تعداد $N = N_a + N_b + N_{ab}$، برابر اسـت بـا مجموع u_iها (U) در جامعه‌ای با تعداد $N^* = N_a + N_b + 2N_{ab}$؛ یعنی،

$$Y = Y_a + Y_{ab} + Y_b = Y_a + pY_{ab} + qY_{ab} + Y_b = U\,. \qquad (۱)$$

استفاده از تغییر متغیر فوق، ابتکار جالبی بود که ابتدا هارتلی آن را ارائه کرد و سپس توسط دیگران مورد استفاده قرار گرفت. چنان‌که قبلاً اشاره شد، برای براورد مجموع، دو حالت در نظر گرفته شده است: حالت اوّل، وقتی که اندازه‌ی حوزه‌ها معلوم است، و حالت دوم، وقتی که اندازه‌ی حوزه‌ها نامعلوم است.

۳- براورد مجموع در حالتی که N_a و N_b و N_{ab} معلوم هستند

در این حالت، چنان‌چه هر یک از حوزه‌های (a)، (ab) و نیز (b) و (ab) را به‌عنوان طبقات پسین (post strata) در نظر بگیریم، براورد مجموع با استفاده از روش طبقه‌بندی پسین و نیز در نظر گرفتن شیوه‌ی نمونه‌گیری تصادفی ساده به‌صورت زیر خواهد بود:

$$\hat{Y}_H = N_a\bar{y}_a + N_{ab}\left(p\bar{y}'_{ab} + q\bar{y}''_{ab}\right) + N_b\bar{y}_b\,, \qquad (۲)$$

که واریانس آن عبارت است از:

$$\operatorname{var}\left(\hat{Y}_H\right) = N_a^2\operatorname{var}\left(\bar{y}_a\right) + N_{ab}^2p^2\operatorname{var}\left(\bar{y}'_{ab}\right) + N_{ab}^2q^2\operatorname{var}\left(\bar{y}''_{ab}\right) + N_b^2\operatorname{var}\left(\bar{y}_b\right).$$

با صرف‌نظر کردن از FPC (ضریب تصحیح جامعه‌ی متناهی) و نیز آگاهی از این نکته که واریانس براوردگر مجموع در طبقه‌بندی پسین، با واریانس براوردگر مجموع در نمونه‌گیری طبقه‌بندی با تخصیص متناسب تقریباً یکسان است، خواهیم داشت:

$$\operatorname{var}\left(\hat{Y}_H\right) = \frac{N_A^2}{n_A}\left\{\sigma_a^2(1-\alpha) + p^2\sigma_{ab}^2\alpha\right\} + \frac{N_B^2}{n_B}\left\{\sigma_b^2(1-\beta) + q^2\sigma_{ab}^2\beta\right\}, \qquad (۳)$$

که در آن، σ_a^2 و σ_b^2 و σ_{ab}^2، به‌ترتیب واریانس جامعه برای حوزه‌های a و b و ab هستند، و نیز

$$\beta = \frac{N_{ab}}{N_B} \quad \text{و} \quad \alpha = \frac{N_{ab}}{N_A}$$

به‌ترتیب نشان‌دهنده‌ی نسبت اندازه‌ی حوزه‌ی مشترک به اندازه‌ی چارچوب A و B می‌باشند.

با در نظر گرفتن عامل هزینه به‌صورت تابع خطی ساده‌ی

$$c = c_A n_A + c_B n_B , \qquad (۴)$$

رابطه‌ی (۳) که تابعی از p و n_A و n_B است، با توجه به محدودیت (۴) مینیمم می‌شود. با استفاده از ضرایب نامعین لاگرانژ برای پیدا کردن مقدار بهینه‌ی p، معادله‌ی زیر به دست می‌آید:

$$\frac{c_A p^2}{c_B q^2} = \frac{\sigma_a^2(1-\alpha) + p^2 \sigma_{ab}^2 \alpha}{\sigma_b^2(1-\beta) + q^2 \sigma_{ab}^2 \beta} \qquad (۵)$$

و به همین ترتیب، کسر نمونه‌گیری بهینه عبارت است از:

$$\frac{n_A}{N_A} = c\left\{\frac{\left(\sigma_a^2(1-\alpha) + p^2 \sigma_{ab}^2 \alpha\right)}{c_A}\right\}^{\frac{1}{2}},$$
$$\frac{n_B}{N_B} = c\left\{\frac{\left(\sigma_b^2(1-\beta) + q^2 \sigma_{ab}^2 \beta\right)}{c_B}\right\}^{\frac{1}{2}}. \qquad (۶)$$

لوند [۴] برای ارائه‌ی براورد مجموع، طرفین رابطه‌ی (۳) را نسبت به p و q مشتق‌گیری کرد و روابط زیر را به دست آورد:

$$q_{opt} = \frac{\beta n_B}{\alpha n_A + \beta n_B} \quad \text{و} \quad p_{opt} = \frac{\alpha n_A}{\alpha n_A + \beta n_B} \qquad (۷)$$

به این ترتیب، ملاحظه می‌شود که مقدار بهینه‌ی p، عبارت است از نسبت امید ریاضی n'_{ab} به امید ریاضی $n'_{ab} + n''_{ab}$. اساس استدلال‌های هارتلی، بدون توجه به تصادفی بودن توزیع n_A و n_B درون حوزه‌هاست؛ در حالی که لوند، تصادفی بودن n'_{ab} و n''_{ab} را اساس کار خود قرار داده است. رابطه‌ی (۳) را می‌توان با استفاده از قضیه‌ی شرطی کردن، به‌صورت زیر نوشت:

$$\mathrm{var}(\hat{Y}) = E\left\{\mathrm{var}\left(\hat{Y} \middle| n'_{ab}, n''_{ab}\right)\right\} + \mathrm{var}\left\{E\left(\hat{Y} \middle| n'_{ab}, n''_{ab}\right)\right\}, \qquad (۸)$$

که شرط $\left(\cdot \middle| n'_{ab}, n''_{ab}\right)$، بیان‌کننده‌ی تعداد نمونه‌ی واقعی به‌دست‌آمده در حوزه‌ی ab

است. $\hat{Y}$ براوردگری نااریب برای Y است؛ لذا مقدار جمله‌ی دوم طرف راست رابطه، معادل صفر است. مجدداً با صرف‌نظر کردن از FPC خواهیم داشت:

$$\operatorname{var}\left(\hat{Y}\middle|n'_{ab},n''_{ab}\right)=\frac{N_a^2}{n_a}\sigma_a^2+p^2\frac{N_{ab}^2}{n'_{ab}}\sigma_{ab}^2+(1-p)^2\frac{N_{ab}^2}{n''_{ab}}\sigma_{ab}^2+\frac{N_b^2}{n_b}\sigma_b^2. \tag{۹}$$

با مینیمم کردن رابطه‌ی (۹) نسبت به p، مقدار بهینه‌ی p عبارت است از:

$$p_{opt}=\frac{n'_{ab}}{n'_{ab}+n''_{ab}}. \tag{۱۰}$$

به این ترتیب، ملاحظه می‌شود که مقدار بهینه‌ی p، کسری است مبتنی بر نمونه‌های انتخاب‌شده از دو چارچوب A و B که در حوزه‌ی مشترک ab قرار گرفته‌اند. لوند با جایگذاری p_{opt} به‌دست‌آمده از رابطه‌ی (۱۰) در رابطه‌ی (۲)، براوردگر زیر را برای مجموع ارائه کرد:

$$\hat{Y}_L=N_a\bar{y}_a+N_{ab}\bar{y}_{ab}+N_b\bar{y}_b, \tag{۱۱}$$

که در آن،

$$\bar{y}_{ab}=\frac{n'_{ab}\bar{y}'_{ab}+n''_{ab}\bar{y}''_{ab}}{n'_{ab}+n''_{ab}}.$$

با کمی دقت، ملاحظه می‌شود که براوردگر لوند، همان براوردگر هارتلی است که در آن به‌جای p و q، مقدار بهینه‌ی آن‌ها قرار گرفته است. واریانس براوردگر یاد شده عبارت است از

$$\operatorname{var}\left(\hat{Y}_L\right)=\frac{N_A^2}{n_A}(1-\alpha)\sigma_a^2+\frac{N_AN_B\alpha\beta}{\alpha n_A+\beta n_B}\sigma_{ab}^2+\frac{N_B^2}{n_B}(1-\beta)\sigma_b^2. \tag{۱۲}$$

از مقایسه‌ی روابط (۳) و (۱۲) نتیجه می‌شود که همواره $\operatorname{var}\left(\hat{Y}_H\right)\geq\operatorname{var}\left(\hat{Y}_L\right)$. لذا براوردگر لوند، کاراتر از براوردگر هارتلی است.

۴- براورد مجموع در حالتی که N_a و N_b و N_{ab} نامعلوم باشند

در حالتی که N_a و N_b و N_{ab} نامعلوم باشند، هـارتلی [۳] از فرمـول‌هـای معمـول در نمونه‌گیری طبقه‌بندی‌شده برای متغیر U در هر یک از طبقات استفاده کرده است. بنـا بـر این، براوردگر $U = Y$ عبارت است از

$$\hat{Y}_H = \frac{N_A}{n_A}\{y_a + py'_{ab}\} + \frac{N_B}{n_B}\{y_b + qy''_{ab}\}, \qquad (۱۳)$$

که واریانس آن برابر است با

$$\begin{aligned} \operatorname{var}(\hat{Y}_H) = \frac{N_A^2}{n_A}\left\{(1-\alpha)\sigma_a^2 + p^2\sigma_{ab}^2\alpha + \alpha(1-\alpha)\left(\bar{Y}_a - p\bar{Y}_{ab}\right)^2\right\} \\ + \frac{N_B^2}{n_B}\left\{(1-\beta)\sigma_B^2 + q^2\sigma_{ab}^2\beta + \beta(1-\beta)\left(\bar{Y}_b - q\bar{Y}_{ab}\right)^2\right\} \end{aligned} \qquad (۱۴)$$

اما لوند، براوردگر (۱۱) را به‌عنوان نقطه‌ی شـروع بـرای حـالتی کـه N_a و N_b و N_{ab} نامعلوم‌اند، در نظر گرفته است. از آن‌جا که اندازه‌ی حوزه‌ها نـامعلوم اسـت، بـا اسـتفاده از اندازه‌ی نمونه و مقادیر معلوم N_A و N_B، اندازه‌ی آن‌ها براورد می‌شود. لذا از

$$N_A\left(\frac{n_a}{n_A}\right) \quad \text{و} \quad N_B\left(\frac{n_b}{n_B}\right)$$

به‌عنوان براوردگرهایی نااریب برای N_a و N_b، و از

$$N_A\left(\frac{n'_{ab}}{n_A}\right) \quad \text{و} \quad N_B\left(\frac{n''_{ab}}{n_B}\right)$$

به‌عنوان براوردگرهایی نااریب برای N_{ab} استفاده می‌شـود. بـا اختـصاص وزن‌هـای w و $1-w$ به هر یک از دو براوردگر اخیر و جایگذاری براوردگرهـای فـوق در رابطـه‌ی (۱۱)، براوردگر نااریب برای مجموع به دست می‌آید:

$$\hat{Y}_L = \frac{N_A}{n_A}n_a\bar{y}_a + \left[\frac{N_A}{n_A}n'_{ab}w + \frac{N_B}{n_B}n''_{ab}(1-w)\right]\bar{y}_{ab} + \frac{N_B}{n_B}n_b\bar{y}_b, \qquad (۱۵)$$

که در آن، $\bar{y}_{ab}$ میانگین همه‌ی واحدهای انتخاب‌شده از حوزه‌ی مشترک است. واریـانس براوردگر فوق عبارت است از

$$
\begin{aligned}
\operatorname{var}\left(\hat{Y}_L\right) = {} & \frac{N_A^2}{n_A}(1-\alpha)\sigma_a^2 + \frac{N_A N_B \alpha\beta}{\alpha n_A + \beta n_B}\sigma_{ab}^2 \\
& + \frac{N_B^2}{n_B}(1-\beta)\sigma_B^2 + \frac{N_A^2(1-\alpha)\alpha}{n_A}\left(\bar{Y}_a - w\bar{Y}_{ab}\right)^2 \\
& + \frac{N_B^2(1-\beta)\beta}{n_B}\left[\bar{Y}_b - (1-w)\bar{Y}_{ab}\right]^2
\end{aligned}
\qquad (۱۶)
$$

دو جمله‌ی آخر سمت راست، نشان‌دهنده‌ی افزایش واریانس در حالتی می‌باشـند کـه N_a و N_b و N_{ab} نامعلوم‌اند.

رابطه‌ی (۱۶) به‌عنوان تابعی از w و n_A و N_B بایـد نـسبت بـه تـابع هزینـه‌ی (۴) مینیمم شود. به این ترتیب،

$$
w_{opt} = \frac{\dfrac{N_A(1-\alpha)}{n_A}\bar{Y}_a + \dfrac{N_B(1-\beta)}{n_B}\left(\bar{Y}_{ab} - \bar{Y}_b\right)}{\left[\dfrac{N_A(1-\alpha)}{n_A} + \dfrac{N_B(1-\beta)}{n_B}\right]\bar{Y}_{ab}}, \qquad (۱۷)
$$

اما چون عملاً N_a و N_b و N_{ab} نامعلوم‌اند، بایـد از بـراورد w_{opt} اسـتفاده شـود؛ یعنـی به‌جای N_a، N_b، N_{ab}، $\bar{Y}_a$، $\bar{Y}_b$، و $\bar{Y}_{ab}$ از براوردهـای‌شـان اسـتفاده مـی‌شـود. بـدین ترتیب،

$$
\hat{w} = \frac{\dfrac{N_A n_a}{n_A^2}\bar{y}_a + \dfrac{N_B n_b}{n_B^2}\left(\bar{y}_{ab} - \bar{y}_b\right)}{\left(\dfrac{N_A n_A}{n_A^2} + \dfrac{N_B n_B}{n_B^2}\right)\bar{y}_{ab}}. \qquad (۱۸)
$$

براوردگر ارائه‌شده توسط لوند از لحاظ محاسباتی ساده‌تر است. همچنین می‌توان ثابت کرد که کارایی براوردگر (۱۵)، همواره بزرگ‌تر از کارایی براوردگـر پیـشنهاد شـده توسـط

هارتلی، یا برابر با آن است.

حال به براوردگر دیگری برای مجموع می‌پردازیم کـه توسـط فـولر و بورمایـستر [۲] پیشنهاد شده است. به‌منظور براورد مجموع، ابتدا برای N_{ab} یک براوردگر ارائه شده است. براوردگری که هارتلی برای N_{ab} پیشنهاد کرد به‌صورت زیر است:

$$\hat{N}_{ab}=\frac{pn'_{ab}N_A}{n_A}+\frac{qn''_{ab}N_B}{n_B} \qquad (۱۹)$$

با واریانس

$$\operatorname{var}\left(\hat{N}_{ab,H}\right)=p^2 f_A^{-2}\operatorname{var}\left(n'_{ab}\right)+q^2 f_B^{-2}\operatorname{var}\left(n''_{ab}\right), \qquad (۲۰)$$

که f_A و f_B به‌ترتیب، کسرهای نمونه‌گیری در چارچوب‌هـای A و B هـستند، و n'_{ab} و n''_{ab}، متغیرهای تصادفی با توزیع فوق هندسی. از آن‌جا که

$$n'_{ab}\sim HG\left(N_A,N_{ab},n_A\right),\qquad \operatorname{var}\left(n'_{ab}\right)=\frac{\left(N_A-n_A\right)}{\left(N_A-1\right)N_A^2}N_{ab}N_a n_A,$$

$$n''_{ab}\sim HG\left(N_B,N_{ab},n_B\right),\qquad \operatorname{var}\left(n''_{ab}\right)=\frac{\left(N_B-n_B\right)}{\left(N_B-1\right)N_B^2}N_{ab}N_b n_B,$$

داریم:

$$\begin{aligned}\operatorname{var}\left(\hat{N}_{ab,H}\right)=&\ p^2 f_A^{-2}\left[\frac{\left(N_A-n_A\right)N_{ab}N_a n_A}{\left(N_A-1\right)N_A^2}\right]\\&+q^2 f_B^{-2}\left[\frac{\left(N_B-n_B\right)N_{ab}N_b n_B}{\left(N_B-1\right)N_B^2}\right].\end{aligned} \qquad (۲۱)$$

با مشتق‌گیری طرفین رابطه‌ی (۲۱) نسبت به p و ساده کردن آن، مقدار بهینه‌ی زیر برای p به دست می‌آید:

$$p_{opt}=\frac{n_A N_b g_B}{n_A N_B g_B+n_B N_a g_A}, \qquad (۲۲)$$

که در آن

$$g_B = \frac{N_B - n_B}{N_B - 1} \quad \text{و} \quad g_A = \frac{N_A - n_A}{N_A - 1}$$

با جایگذاری p_{opt} در رابطه‌ی (۲۰)، واریانس بهینه‌ی $\hat{N}_{ab}$ عبارت است از:

$$\operatorname{var}_{opt}\left(\hat{N}_{ab,H}\right) = \frac{N_{ab} N_a N_b g_A g_B}{n_A N_b g_B + n_B N_a g_A}. \tag{۲۳}$$

با جایگذاری رابطه‌ی (۲۲) در (۱۹)، معادله‌ی درجه دومی از $\hat{N}_{ab}$ به‌صورت زیر به دست می‌آید:

$$\begin{aligned}&[n_A g_B + n_B g_A]\hat{N}^{2}_{ab,s} + [n'_{ab} g_B + n''_{ab} g_A] N_A N_B \\ &-[n_A N_B g_B + n_B N_A g_A + n'_{ab} N_A g_B + n''_{ab} N_B g_A]\hat{N}_{ab,s} = 0,\end{aligned} \tag{۲۴}$$

که ریشه‌های آن همواره حقیقی‌اند و بزرگ‌ترین ریشه‌ی آن نیز همواره بزرگ‌تر از مینیمم N_A و N_B یا برابر با آن است. کوچک‌ترین ریشه نیز همواره در فاصله‌ی بین صفر و مینیمم N_A و N_B قرار می‌گیرد. بنا بر این، کوچک‌ترین ریشه را به‌عنوان براورد N_{ab} در نظر می‌گیریم. شایان ذکر است که براوردگر (۱۹) بر خلاف براوردگر اخیر، همواره در دامنه‌ی مقادیر قابل قبول برای N_{ab} قرار نمی‌گیرد. خواص نمونه‌ی بزرگ برای $\hat{N}_{ab}$ در قضیه‌ی زیر ارائه شده است.

فولر و بورمایستر [۲] نشان دادند که دنباله‌ی براوردگرهای $\hat{N}_{ab,s}$، تعریف‌شده در رابطه‌ی (۲۴)، در روابط زیر صدق می‌کند:

$$\operatorname{var}\left(\hat{N}_{ab,s}\right) = \frac{N_{ab} N_a N_b g_A g_B}{n_A N_b g_B + n_B N_a g_A} + o(1),$$

$$E\left(\hat{N}_{ab,s}\right) = N_{ab} + o\left(\frac{1}{n}\right).$$

بنا بر این، واریانس $\hat{N}_{ab,s}$ همان واریانس حدی براوردگر خطی (۱۹) است که در آن از وزن‌های بهینه استفاده شده است. منظور از $o(1)$ وقتی که $x \to \infty$، عبارت است از تابعی از x مانند $g(x)$ با شرط

$$\lim_{x\to 1}\frac{g(x)}{x-1}=0.$$

حال به توضیح براوردگر برای مجموع جامعه می‌پردازیم. می‌دانیم که

$$E\left(\bar{y}_a \middle| n_a, n'_{ab}, n''_{ab}, n_b\right)=\bar{Y}_a .$$

به‌شرط استفاده از $\hat{N}_{ab,s}$ به‌جای N_{ab}، براوردگر زیر برای مجموع در نظر گرفته می‌شود:

$$\hat{Y}_s=\left(N_A-\hat{N}_{ab.s}\right)\bar{y}_a+\hat{N}_{ab,s}\bar{y}_{ab,s}+\left(N_B-\hat{N}_{ab,s}\right)\bar{y}_b , \qquad (۲۵)$$

که در آن

$$\bar{y}_{ab,s}=w\bar{y}'_{ab}+(1-w)\bar{y}''_{ab}$$

و

$$w=\frac{n'_{ab}\left(1-f_B\right)}{n'_{ab}\left(1-f_B\right)+n''_{ab}\left(1-f_A\right)}.$$

w به‌گونه‌ای انتخاب شده است که واریانس $\bar{y}_{ab,s}$ را مینیمم کند.

$$\begin{aligned}\operatorname{var}\left(\hat{Y}_s\right)=&\ N_a\left(f_A^{-1}-1\right)S_a^2+\left[(1-f_B)f_A+(1-f_A)f_B\right]^{-1}(1-f_A)(1-f_B)N_{ab}S_{ab}^2\\&+N_b\left(f_B^{-1}-1\right)S_b^2+\left(\bar{Y}_{ab}-\bar{Y}_a-\bar{Y}_b\right)^2\frac{N_{ab}N_aN_bg_Ag_B}{n_AN_bg_B+n_BN_ag_A}+o(1)\end{aligned}$$

و

$$E\left(\hat{Y}_s\right)=Y+o\left(\frac{1}{N}\right),$$

که در آن

$$S_a^2=\frac{1}{N_a-1}\sum_{i=1}^{N_a}\left(y_{ai}-\bar{y}_a\right)^2,$$

$$S_b^2=\frac{1}{N_b-1}\sum_{i=1}^{N_b}\left(y_{bi}-\bar{y}_b\right)^2,$$

$$S_{ab}^{2} = \frac{1}{N_{ab} - 1}\sum_{i=1}^{N_{ab}} (y_{abi} - \bar{y}_{ab})^{2}.$$

اگر $\hat{w}$ به‌دست‌آمده در رابطه‌ی (۱۸) که لوند به‌عنوان براوردگری برای w ی بهینـه ارائـه کرده بود، به‌جای $\hat{Y}_s$ در رابطه‌ی (۲۵) قـرار داده شـود، براوردگـری مـشابه $\hat{Y}_s$ بـه دسـت می‌آید، با این تفاوت که اریبی آن از رتبه‌ی $o(\frac{1}{n})$ است؛ در حالی‌که اریبی براوردگر لونـد، $o(1)$ بود. به‌علاوه، براوردگر $\bar{Y}_{ab}$ که در $\hat{Y}_s$ به کـار رفتـه اسـت، در نمونـه‌گیـری بـدون جایگذاری، کاراتر از آن براوردگر $\bar{Y}_{ab}$ می‌باشد که در $\hat{Y}_L$ (رابطـه‌ی (۱۵)) مـورد اسـتفاده قرار گرفته است. همچنین فرم خطی رابطه‌ی (۲۵) در نمونه‌گیـری‌هـای بـزرگ، از لحـاظ محاسباتی ساده‌تر است.

بدین ترتیب در حالتی که اندازه‌ی حوزه‌ها نـامعلوم باشـد، اسـتفاده از براوردگـر لونـد نسبت به براوردگر هارتلی برتری دارد، و به نوبه‌ی خود، براوردگـر ارائـه‌شـده توسـط فـولر و بورمایستر بنا به دلایل ارائه‌شده در بالا، نسبت به براوردگر لوند، ارجح است.

مرجع‌ها

[۱] Food and Agriculture Organization of the United Nations (1998). *Multiple Frame Agricultural Surveys*. Food and Agriculture Organization of the United Nations, Rome.

[۲] Fuller, W.A; Burmeister, L.F. (1972). Estimators for samples selected from two overlapping frames. *Proc. Social Stat. Sect*. American Statistical Association. 245-249.

[۳] Hartley, H.O. (1962). Multiple Frame Surveys. *Proc. Social Stat. Sect*. American Statistical Association, 203-206.

[۴] Lund, R.E. (1968). Estimates in Multiple Surveys. *Proc. Social Stat. Sect*. American Statistical Association, 282-288.