Statistical Research and Training Center
©(2005) J. Stat. Res. Iran **2** 187-200

مجله‌ی پژوهش‌های آماری ایران
دوره‌ی ۱، شماره‌ی ۲، پاییز و زمستان ۱۳۸۳، صص ۱۰۹-۱۲۲
پژوهشکده‌ی آمار

طرح نمونه‌گیری حذف سطر و ستون ۱+ و کارایی آن

محمد صالحی مرزیجرانی[†] و محمدامین جمال‌زاده[‡,*]

[†] دانشگاه صنعتی اصفهان

[‡] پژوهشکده‌ی آمار

در این مقاله طرح نمونه‌گیری «حذف سطر و ستون ۱+» را معرفی می‌کنیم. این طرح نمونه‌گیری یک حالت عمومی از طرح نمونه‌گیری «مربع لاتین ساده-k» است که توسط بورکوفسکی (۲۰۰۳) معرفی شده است. این طرح نمونه‌گیری برای جوامعی که به‌صورت آرایه‌ی مستطیلی قابل استفاده است. برعکس، طرح نمونه نمونه‌گیری «مربعی لاتین ساده-k» فقط برای آرایه‌های مربعی قابل استفاده است. در این مقاله شرط لازم و کافی برای اینکه طرح نمونه‌گیری «حذف سطر و ستون ۱+» از طرح نمونه‌گیری تصادفی ساده بدون جایگذاری کاراتر باشد، ارائه می‌شود. با توجه به‌شرط ارائه شده، مشخص می‌شود چنانچه جامعه دارای روند افقی یا عمودی باشد، این طرح نمونه‌گیری کاراتر است. با مطالعه روی یک جامعه از صدف‌های کم‌یاب آب شیرین که دارای روند عمودی بسیار ملایمی هستند، چگونگی تغییرات کارایی مورد بحث قرار می‌گیرد.

واژگان کلیدی. طرح نمونه‌گیری مربع لاتین؛ برآوردگر هورویتز-تامپسون؛ ضریب خودهمبستگی؛ تحلیل داده‌های فضایی.

* نویسنده‌ی عهده‌دار مکاتبات.

[DOI: 10.18869/acadpub.jsri.1.2.109]

۱ مقدمه

در مطالعات زیست‌شناسی، جامعه‌شناسی، زمین‌شناسی و کشاورزی غالباً مکان جغرافیایی جامعه به تعدادی چهارضلعی افراز می‌شود. بر اساس یک طرح نمونه‌گیری، نمونه‌ای از چهارضلعی‌ها انتخاب و با استفاده از آن نمونه، پارامترهای جامعه برآورد می‌شوند. در چنین جوامعی معمولاً مقدار مورد مطالعه در چهارضلعی‌ها، دارای خودهمبستگی فضایی (Spatial Autocorrelation) مثبت هستند. استفاده از طرح نمونه‌گیری که واحدهای نمونه پوشش خوبی از ناحیه‌ی جامعه را ارائه دهد، موجب افزایش دقت برآوردگر پارامترهای جامعه خواهد شد (شرودر و همکاران، ۱۹۹۳). پوشش خوب از ناحیه‌ی جامعه توسط نمونه، یعنی واحدهای نمونه به‌خوبی در تمام ناحیه‌ی جامعه پخش شوند.

هیاک (۱۹۵۹) تحت مدلی با خودهمبستگی فضایی مثبت ثابت کرد که نمونه‌گیری سیستماتیک، طرح نمونه‌گیری بهینه برای جوامع یک‌بعدی است. در عین حال، نمونه‌گیری سیستماتیک و طبقه‌ای با یک نمونه از هر طبقه، دو نمونه‌گیری سنتی هستند که پوششی مناسب از ناحیه‌ی جامعه ارائه می‌دهند (مک‌کنزی و همکاران، ۱۹۹۱)، امّا برآوردگر نااریب واریانس برای این دو طرح نمونه‌گیری وجود ندارد. طرح نمونه‌گیری مربع لاتین ساده، طرح نمونه‌گیری دیگری است که پوششی مناسب از ناحیه‌ی جامعه را به ما ارائه می‌دهد که برای آن نیز برآوردگر نااریب واریانس وجود ندارد. از دیدگاه عملی، عدم وجود یک برآوردگر مناسب واریانس، نقطه‌ی ضعف مهم یک طرح نمونه‌گیری محسوب می‌شود. مانهلند و بورکوفسکی (۱۹۹۶) طرح نمونه‌گیری مربع لاتین +۱ را معرفی کردند. پس از انتخاب نمونه‌ی مربع لاتین ساده، آن‌ها یک نمونه‌ی اضافی انتخاب کردند تا بتوانند یک برآوردگر نااریب واریانس ارائه دهند.

دو مشکل مهم دیگر در مورد نمونه‌گیری مربع لاتین ساده وجود دارد:

۱) جامعه باید شکل مربعی داشته باشد.

۲) اندازه‌ی نمونه باید به اندازه‌ی یک ضلع جامعه باشد.

صالحی (۲۰۰۲) با معرفی نمونه‌گیری سیستماتیک مربع لاتین ساده شکل جامعه را به هر چندضلعی که در آن می‌توان مربع‌هایی را بتوان محاط کرد، گسترش داد. در این طرح اندازه‌ی نمونه می‌تواند بیش از اندازه‌ی یک ضلع جامعه باشد. بورکوفسکی (۲۰۰۳) طرح نمونه‌گیری مربع لاتین ساده $\pm k$ را معرفی کرد که در آن اندازه‌ی نمونه می‌تواند هر مقداری بزرگ‌تر از اندازه‌ی یک ضلع جامعه باشد. لاوری و بل‌هاوس (۱۹۹۲) خواص بهینه‌ی طرح نمونه‌گیری مربع لاتین ساده را تحت یک مدل همبستگی فضایی بررسی کردند. صالحی (۲۰۰۴، ۲۰۰۲) خواص بهینه‌ی طرح نمونه‌گیری سیستماتیک مربع لاتین ساده را تحت همان مدل همبستگی فضایی بررسی کردند.

امّا هنوز یک سؤال باقی می‌ماند. اگر بخواهیم نمونه‌ای به اندازه‌ی کم‌تر از طول یک ضلع جامعه

[DOI: 10.18869/acadpub.jsri.1.2.109]

بگیریم چه طرح نمونه‌گیری پیشنهاد می‌شود؟ بورکوفسکی (۲۰۰۳) طرح نمونه‌گیری را تحت نام «k-مربع لاتین ساده» SLSS-k ارائه داد. امّا برای طرح بالا برآورد واریانس ارائه نشده است. عملاً این طرح قابل استفاده نخواهد بود چون نمی‌توان دقت برآوردگر آن را مشخص کرد. برای حل این مشکل و گسترش طرح SLSS-k از جامعه‌ای مستطیلی، در بخش ۲ طرح نمونه‌گیری حذف سطر و ستون ۱+ را معرفی می‌کنیم. در بخش ۳ برآوردگر، واریانس آن و برآورد واریانس را ارائه می‌دهیم. در بخش ۴ کارایی طرح نمونه‌گیری ارزیابی می‌شود. در بخش ۵ نتیجه‌گیری صورت می‌گیرد.

۲ طرح نمونه‌گیری حذف سطر و ستون (۱+ $RCES$)

فرض کنید که جامعه به یک آرایه‌ی مستطیلی به اندازه‌ی $R \times C$ واحد افراز و واحدها به‌صورت $\{1, 2, \ldots RC\}$ اندیس‌گذاری شوند. R تعداد سطرها و C تعداد ستون‌ها را نشان می‌دهند. در نظر داریم یک نمونه به اندازه‌ی $n + 1$ انتخاب کنیم که $n \leqslant \min\{R, C\}$.

یک روش اجرای نمونه‌گیری این است که یک واحد از جامعه به‌طور تصادفی انتخاب، سپس سطر و ستون مربوط به آن از جامعه حذف شود. دومین واحد از واحدهای باقی‌مانده انتخاب شود و سطر و ستون مربوط به دومین واحد نمونه نیز از جامعه حذف شود. این فرایند را ادامه می‌دهیم تا n واحد انتخاب شود. این مجموعه‌ی نمونه را نمونه‌ی حذف سطر و ستون RCES (Row Column Elimination Sampling) می‌نامیم. در این طرح نمونه‌گیری بعضی از احتمالات شمول توأم (Joint Inclusion Probabilities) صفر می‌باشند. بنا بر این، برآوردگر ناااریب واریانس وجود ندارد. یک واحد دیگر از واحدهای انتخاب نشده به تصادف انتخاب می‌کنیم. این طرح نمونه‌گیری را طرح نمونه‌گیری حذف سطر و ستون ۱+ می‌نامیم. اگر تعداد سطرها و ستون‌ها با یکدیگر برابر و مساوی R باشند (جامعه به شکل مربع) و اندازه‌ی نمونه هم $R + 1$ باشد، آن‌گاه طرح نمونه‌گیری حذف سطر و ستون ۱+ همان طرح نمونه‌گیری مربع لاتین ساده ۱+ می‌شود.

۳ برآوردگر، واریانس و برآوردگر واریانس

تعداد راه‌هایی که اوّلین نمونه می‌تواند انتخاب شود RC است. بعد از حذف سطر و ستون مربوط $(R-1)(C-1)$ واحد باقی می‌ماند. بنا بر این، دومین واحد نمونه را به $(R-1)(C-1)$ روش

[DOI: 10.18869/acadpub.jsri.1.2.109]

می‌توان انتخاب کرد. به‌طور کلّی kامین واحد نمونه می‌تواند به $(R-k)(C-k)$ روش انتخاب شود. در نتیجه تعداد مجموعه‌های ممکن به اندازه‌ی n، برابر $\prod_{k=0}^{n-1}(R-k)(C-k)$ و تعداد نمونه‌های نامرتب ممکن برابر است با:

$$\frac{\prod_{k=0}^{n-1}(R-k)(C-k)}{n!}.$$

از آن‌جا که آن نمونه‌ی اضافی (۱+) می‌تواند به $(RC-n)$ روش انتخاب شود، تعداد مجموعه‌های نمونه‌ی ممکن نامرتب برابر است با:

$$\frac{\prod_{k=1}^{n-1}(R-k)(C-k)}{n!}(RC-n). \tag{1}$$

در پیوست نشان می‌دهیم که احتمال شمول واحد j برابر $\pi_j = \frac{n+1}{RC}$ است. با استفاده از برآوردگر هورویتز-تامپسون (۱۹۵۲)، یک برآوردگر نااریب برای میانگین جامعه (μ) برابر است با:

$$\begin{aligned}\hat{\mu} &= \frac{1}{RC}\sum_{i=1}^{n+1}\frac{y_i}{\pi_i}\\ &= \frac{1}{n+1}\sum_{i=1}^{n+1} y_i = \bar{y},\end{aligned}$$

که در آن y_i مقدار مورد مطالعه‌ی واحد iام است.

برای به دست آوردن واریانس و برآورد نااریب واریانس لازم است که احتمال شمول توأم π_{ij} را مشخص کنیم. اگر واحد i و j در یک سطر یا ستون قرار داشته باشند، هر دو در نمونه هستند اگر و تنها اگر یکی در مرحله‌ی انتخاب RCES و دیگری به‌عنوان آن نمونه‌ی آخر (۱+) انتخاب شود. تعداد راه‌هایی که پیشامد بالا رخ می‌دهد برابر است با:

$$2\left\{\frac{\prod_{k=1}^{n-1}(M-k)(N-k)}{(n-1)!}\right\},$$

که با تقسیم آن بر (۱) احتمال شمول توأم برای این حالت به‌صورت زیر خواهد بود:

$$\pi_{1ij} = \frac{2n}{RC(RC-n)}.$$

اگر واحدهای i و j در سطر و ستونی متفاوت قرار گیرند، آن‌گاه

i) هر دو واحد در RCES، یا

(ii) یکی از آن‌ها در RCES و دیگری آن نمونه‌ی آخر (۱+) باشد.

از آن‌جا که

$$\frac{\prod_{k=۲}^{n-۱}(R-k)(C-k)(RC-n)}{(n-۲)!}$$

تعداد نمونه‌های ممکن است که در آن دو واحد i و j در RCES انتخاب می‌شوند، احتمال حالت (i) برابر $\frac{n^۲-n}{RC(RC-R-C+۱)}$ است.

تعداد نمونه‌های ممکن در حالت (ii) برابر است با:

$$۲\left\{\frac{\prod_{k=۱}^{n-۱}(R-k)(C-k)}{(n-۱)!}-\frac{\prod_{k=۲}^{n}(R-k)(C-k)}{(n-۲)!}\right\}.$$

بنا بر این، با تقسیم آن بر (۱) احتمال حالت (ii) برابر است با:

$$\frac{۲n(RC-R-C-n+۲)}{RC(RC-R-C+۱)(RC-n)}.$$

در نتیجه احتمال شمول توأم برای واحدهایی که در سطرها و ستون‌های متفاوت قرار دارند برابر مجموع احتمال برای دو حالت مختلف است.

$$\pi_{۲ij}=\frac{n(n-۱)(RC-n)+۲n(RC-R-C-n+۲)}{RC(RC-R-C+۱)(RC-n)}.$$

برای واحد i دو مجموعه تعریف می‌کنیم. مجموعه‌ی واحدهایی که در سطر یا ستون واحد i قرار دارند را با $R_۱^i$ نمایش می‌دهیم. مجموعه‌ی واحدهایی که در سطر و ستون‌های متفاوت با واحد iام قرار دارند را با $R_۲^i$ نمایش می‌دهیم.

بنا بر این، واریانس $\bar{y}$ به‌صورت زیر خواهد بود:

$$\begin{aligned}
\operatorname{var}(\bar{y}) &= \frac{۱}{(RC)^۲}\Bigg\{\sum_{i=۱}^{RC}(\pi_i-\pi_i^۲)\frac{y_i^۲}{\pi_i^۲}+\sum_{\substack{i=۱\\ i\neq j}}^{RC}\sum_{j=۱}^{RC}(\pi_{ij}-\pi_i\pi_j)\frac{y_i}{\pi_i}\cdot\frac{y_j}{\pi_j}\Bigg\} \qquad (۲)\\
&= \frac{۱}{(RC)^۲}\Bigg\{\frac{RC-n-۱}{n+۱}\sum_{i=۱}^{RC}y_i^۲\\
&\quad +\sum_{h=۱}^{۲}\sum_{i=۱}^{RC}\sum_{j\in R_h^i}\left(\pi_{hij}-\frac{(n+۱)^۲}{(RC)^۲}\right)\frac{y_iy_j(RC)^۲}{(n+۱)^۲}\Bigg\}.
\end{aligned}$$

یک برآورد نااریب واریانس به‌صورت زیر خواهد بود:

$$\widehat{\text{var}}(\bar{y}) = \frac{1}{(RC)^2}\left\{\sum_{i=1}^{n+1}\frac{1-\pi_i}{\pi_i^2}\ y_i^2 + \sum_{\substack{i=1\\ i\neq j}}^{n+1}\sum_{j=1}^{n+1}\left(\frac{1}{\pi_i\pi_j}-\frac{1}{\pi_{ij}}\right)y_iy_j\right\} \tag{3}$$

$$= \frac{1}{(RC)^2}\left\{\frac{(RC-n-1)RC}{(n+1)^2}\ \sum_{i=1}^{n+1}y_i^2 + \sum_{h=1}^{2}\sum_{i=1}^{n+1}\sum_{j\in S_h^i}\left(\frac{(RC)^2}{(n+1)^2}-\frac{1}{\pi_{hij}}\right)y_iy_j\right\},$$

که در آن مجموعه‌ی S_h^i مشابه R_h^i (برای $h = 1, 2$) برای مجموعه‌ی نمونه‌ای به‌جای جامعه تعریف می‌شود.

۴ ارزیابی طرح نمونه‌گیری $RCES$ + ۱

۴٫۱ ارزیابی تحلیلی

برای ارزیابی یک طرح نمونه‌گیری، کارایی به‌صورت زیر تعریف می‌شود:

$$\text{eff}(\hat{\mu}) = \frac{\text{var}(\bar{y_s})}{\text{var}(\hat{\mu})},$$

که در آن $\text{var}(\bar{y_s})$ واریانس میانگین نمونه برای طرح نمونه‌گیری تصادفی ساده با حجم معادل طرح نمونه‌گیری تحت بررسی و $\hat{\mu}$ برآوردگر نااریب میانگین جامعه برای طرح نمونه‌گیری تحت بررسی است. اگر $\text{eff}(\hat{\mu}) > 1$ باشد، طرح نمونه‌گیری تحت بررسی کاراتر از طرح نمونه‌گیری تصادفی ساده است.

لم ۱ (تامپسون، [۲۰۰۲]) واریانس نمونه‌ی تصادفی به اندازه‌ی $n + 1$ از جامعه‌ای به اندازه‌ی RC برای

[DOI: 10.18869/acadpub.jsri.1.2.109]

نمونه‌گیری ساده‌ی تصافی را می‌توان به‌صورت زیر نوشت:

$$\begin{aligned}\mathrm{var}(\bar{y_s}) &= \left(۱ - \frac{n+۱}{RC}\right)\frac{s^۲}{n+۱}\\ &= \frac{۱}{(RC)^۲}\sum_{i=۱}^{RC}\left(\frac{RC}{n+۱} - ۱\right)y_i^۲\\ &\quad + \sum_{\substack{i=۱\\ i\neq j}}^{RC}\sum_{j=۱}^{RC}\left\{\frac{(n+۱)n}{RC(RC-۱)} - \frac{(n+۱)^۲}{(RC)^۲}\right\}\cdot\frac{y_iy_j(RC)^۲}{(n+۱)^۲}.\end{aligned} \tag{۴}$$

برهان. با استفاده از واریانس برآوردگر هورتیز-تامپسون و جایگذاری احتمالهای شمول رابطه‌ی (۴) به دست می‌آید.

قضیه‌ی ۱ طرح نمونه‌گیری حذف سطر و ستون از طرح نمونه‌گیری تصادفی ساده کاراتر است، اگر و فقط اگر

$$\sum_{i=۱}^{RC} y_i\bar{\omega}_i^c < \sum_{i=۱}^{RC} y_i\bar{\omega}_i, \tag{۵}$$

که در آن $\bar{\omega}_i$ میانگین واحدهایی است که در سطر و ستون واحد iام قرار دارند و $\bar{\omega}_i^c$ میانگین واحدهایی است که در سطر و ستون‌های واحد iام قرار ندارند.

برهان. از رابطه‌های (۲) و (۴) داریم:

$$\begin{aligned}\mathrm{var}(\bar{y}_s) - \mathrm{var}(\bar{y}) = \frac{۱}{(n+۱)^۲}\Bigg[\sum_{i=۱}^{R} y_i \sum_{j\in R_۱^i}\left\{\frac{(n+۱)n}{RC(RC-۱)} - \frac{۲n}{RC(RC-n)}\right\}y_i\\ + \sum_{i=۱} y_i \sum_{j\in R_۲^i}\left\{\frac{(n+۱)n}{RC(RC-۱)} - \frac{n(n-۱)}{RC(RC-R-C+۱)}\right.\\ \left. - \frac{۲n(RC-R-C-n+۲)}{RC(RC-R-C+۱)(RC-n)}\right\}y_j\Bigg]\end{aligned}$$

$$= \frac{1}{(n+1)^2}\left[\sum_{i=1}^{RC} y_i \frac{n}{RC}\left\{\frac{(n-1)(RC-n-2)}{(RC-1)(RC-n)}\right\}\right.$$
$$\times (R+C-2)\frac{1}{R+C-2}\sum_{j\in R_1^i} y_j$$
$$+\sum_{i=1}^{RC} y_i\left\{\frac{(n+1)n(RC-R-C+1)}{RC(RC-1)} - \frac{n(n-1)}{RC}\right.$$
$$\left.\left. - \frac{2n(RC-R-C-n+2)}{RC(RC-n)}\right\}\frac{1}{RC-R-C+1}\sum_{j\in R_2^i} y_j\right].$$

با تعریف $\bar{\omega}_i^c = \frac{1}{RC-R-C+1}\sum_{j\in r_2^i} y_i$ و $\bar{\omega}_i = \frac{1}{R+C-2}\sum_{j\in R_1^i} y_i$ داریم:

$$\mathrm{var}(\bar{y}_s) - \mathrm{var}(\bar{y}) = \frac{1}{(n+1)^2}\left\{\frac{n(n-1)(RC-n-2)(R+C-2)}{RC(RC-1)(RC-n)}\sum_{i=1}^{RC} y_i\bar{\omega}_i\right\}$$
$$+\left\{\frac{n(n+1)(RC-R-C+1)}{RC(RC-1)} - \frac{n(n+1)}{RC}\right.$$
$$\left. - \frac{2n(RC-2)}{RC(RC-2)}\right\}\sum_{i=1}^{RC} y_i\bar{\omega}_i^c,$$

پس از محاسبه‌ی جبری داریم:

$$\mathrm{var}(\bar{y}_s) - \mathrm{var}(\bar{y}) = \frac{1}{(n+1)^2}\left[\frac{n(n-1)(RC-n-2)(R+C-2)}{RC(RC-1)(RC-n)}\right.$$
$$\left.\times\left\{\sum y_i\bar{\omega}_i - \sum y_i\bar{\omega}_i^c\right\}\right].$$

بنا بر این،

$$\mathrm{var}(\bar{y}_s) - \mathrm{var}(\bar{y}) > 0 \iff \sum y_i\bar{\omega}_i > \sum y_i\bar{\omega}_i^c,$$

یا

$$\mathrm{eff}(\bar{y}) > 1 \iff \sum y_i\bar{\omega}_i > \sum y_i\bar{\omega}_i^c.$$

از آن‌جا که

$$\frac{1}{RC}\sum_{i=1}^{RC}\bar{\omega}_i = \frac{1}{RC}\sum_{i=1}^{RC}\bar{\omega}_i^c = \mu,$$

رابطه‌ی (۵) برقرار است، اگر و تنها اگر کواریانس $\bar{\omega}_i$ها و y_iها بزرگ‌تر از کواریانس $\bar{\omega}_i^c$ها و y_iها باشد. اگر مقادیر جامعه دارای خودهمبستگی مثبت باشند یا دارای روندی افقی یا عمودی باشند، انتظار می‌رود

[DOI: 10.18869/acadpub.jsri.1.2.109]

که رابطه‌ی خطی مقادیر هر واحد y_i، با بقیه‌ی مقادیر واحدهای سطر یا ستونش از رابطه‌ی خطی آن با مقادیر واحدهایی که در سطر و ستون‌های دیگر قرار دارند، بیش‌تر باشد. بنا بر این، طرح نمونه‌گیری «حذف سطر و ستون +۱» طرح کارایی برای جامعه‌هایی با روند افقی یا عمودی است.

۴٫۲ ارزیابی عددی

برای مشاهده‌ی چگونگی رفتار کارایی طرح نمونه‌گیری «حذف سطر و ستون +۱»، واریانس و کارایی این طرح نمونه‌گیری را برای جامعه‌ای نادر از صدف‌های آب شیرین رودخانه‌ی کاکاپون ویرجینیای غربی برای نمونه‌هایی با اندازه‌های مختلف محاسبه می‌کنیم، (صالحی و اسمیت، ۲۰۰۵). این جامعه دارای روند ضعیف عمودی است. قسمتی از رودخانه‌ی کاکاپون به ۸۰۰ واحد افراز می‌شود (شکل ۱). جامعه دارای ۴۰ سطر و ۲۰ ستون است. اعداد نشان‌دهنده‌ی تعداد صدف‌های موجود در هر واحد هستند.

در جدول ۱ به‌ترتیب واریانس نمونه‌گیری تصادفی ساده، نمونه‌گیری «حذف سطر و ستون +۱» و کارایی آن برای نمونه‌هایی با اندازه‌های ۲ تا ۲۱ ارائه شده است.

جدول ۱. جدول کارایی طرح $RCES1+$ در مقایسه با طرح نمونه‌گیری تصادفی ساده بدون جایگذاری

n	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
$\mathrm{var}(\bar{y}_s)$	۰٫۲۲۶	۰٫۱۵۰	۰٫۱۱۲	۰٫۰۹۰	۰٫۰۷۵	۰٫۰۶۴	۰٫۰۵۶	۰٫۰۵۰	۰٫۰۴۵	۰٫۰۴۱
$\mathrm{var}(\bar{y})$	۰٫۱۵۰	۰٫۱۱۱	۰٫۰۸۹	۰٫۰۷۳	۰٫۰۶۲	۰٫۰۵۴	۰٫۰۴۹	۰٫۰۴۳	۰٫۰۳۸	۰٫۰۳۵
$\mathrm{eff}(\bar{y})$	۱٫۵۰۸	۱٫۳۴۷	۱٫۲۷۰	۱٫۲۲۶	۱٫۱۹۹	۱٫۱۸۱	۱٫۱۷۱	۱٫۱۶۳	۱٫۱۵۹	۱٫۱۵۷

n	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱
$\mathrm{var}(\bar{y}_s)$	۰٫۰۳۷	۰٫۰۳۴	۰٫۰۳۲	۰٫۰۳۰	۰٫۰۲۸	۰٫۰۲۶	۰٫۰۲۶	۰٫۰۲۳	۰٫۰۲۲	۰٫۰۲۱
$\mathrm{var}(\bar{y})$	۰٫۰۳۲	۰٫۰۳۰	۰٫۰۲۷	۰٫۰۲۵	۰٫۰۲۴	۰٫۰۲۲	۰٫۰۲۱	۰٫۰۲۰	۰٫۰۱۹	۰٫۰۱۸
$\mathrm{eff}(\bar{y})$	۱٫۱۵۶	۱٫۱۵۷	۱٫۱۵۹	۱٫۱۶۱	۱٫۱۶۵	۱٫۱۶۹	۱٫۱۷۳	۱٫۱۷۸	۱٫۱۸۳	۱٫۱۸۹

افزایش کارایی از ۱۶ درصد تا ۵۱ درصد تغییر پیدا می‌کند. مطالعات مقدماتی ما نشان دادند که با افزایش اندازه‌ی نمونه، کارایی نیز افزایش پیدا می‌کند ولی همان‌طور که در این جامعه مشاهده می‌شود، این جامعه از چنین قاعده‌ای پیروی نمی‌کند.

[DOI: 10.18869/acadpub.jsri.1.2.109]

۵ نتیجه‌گیری

طرح نمونه‌گیری حذف سطر و ستون ۱+ طرح نمونه‌گیری است که می‌توان برای هر جامعه از آرایه‌ی مستطیلی به کار برد، ضمن اینکه این طرح دارای برآورد واریانس ناااریب است. از سوی دیگر، این طرح نمونه‌گیری برای حالتی که اندازه‌ی نمونه کوچک‌تر از مینیمم تعداد سطرها و ستون‌ها به‌علاوه‌ی یک باشد طراحی شده است. این طرح نمونه‌گیری اجازه نمی‌دهد که نمونه‌های تکراری در سطرها و ستون‌ها به‌جز یک مورد گرفته شود؛ بنا بر این، پوشش خوبی از ناحیه‌ی جامعه ارائه خواهد داد. مطالعه‌ی ما نشانگر این است که وقتی خودهمبستگی بین واحدها وجود دارد ــ بدین معنی که واحدهای همسایه به‌گونه‌ای دارای مقادیر مشابه هستند ــ طرح نمونه‌گیری حذف سطر و ستون ۱+ طرح کارایی است.

بورکوفسکی (۱۹۹۹) طرح نمونه‌گیری مربع لاتین ۱+ سازوار را معرفی کرد و نشان داد که نمونه‌گیری مربع لاتین ساده ۱+ نمونه‌گیری مناسبی برای انتخاب نمونه‌ی اوّلیه برای طرح نمونه‌گیری سازوار خوشه‌ای است. می‌توان پیش‌بینی کرد که طرح نمونه‌گیری حذف سطر و ستون ۱+ نیز طرح نمونه‌گیری مناسبی برای انتخاب نمونه‌ی اوّلیه‌ی طرح نمونه‌گیری سازوار خوشه‌ای باشد، (تامپسون، ۱۹۹۰). صالحی (۲۰۰۶) طرح نمونه‌گیری حذف سطر و ستون ۱+ سازوار خوشه‌ای را معرفی و خواص آن را مورد بررسی قرار داد.

ما انتظار داریم که طرح نمونه‌گیری حذف سطر و ستون ۱+ تقریباً شبیه طرح نمونه‌گیری SLSS-k عمل کند که امتیاز داشتن برآورد نااریب واریانس و امکان اعمال آن بر جامعه‌های مستطیلی وجود دارد.

اگر آرایه‌ی جامعه، مربعی و اندازه‌ی نمونه هم برابر «تعداد سطرها و ستون‌ها» ۱+ باشد، آن‌گاه طرح نمونه‌گیری حذف سطر و ستون ۱+ همان طرح مربع لاتین ۱+ خواهد شد که در این حالت کارایی طرح بیش‌ترین مقدار را خواهد داشت. همان‌طورکه بورکوفسکی (۲۰۰۳) اشاره کرد، طرح نمونه‌گیری SLSS$\pm k$ برای نمونه‌گیری جوامعی با روند قطری کارایی بیش‌تری دارد. این مطلب در مورد طرح RCES+۱ نیز صادق است. طبیعی است که طرح‌های نمونه‌گیری فوق‌الذکر دارای کارایی بالاتری برای جوامعی با روند افقی یا عمودی نسبت به جوامعی با روند قطری باشند.

مرجع‌ها

Borkowski, J.J. (1999). Network inclusion probabilities and Horvitz–Thompson estimation for adaptive simple latin square sampling. *Environ. Ecol. Statist.* **6**, 291-311.

Borkowski, J.J. (2003). Simple latin square sampling $\pm k$ design. *Comm. Statist. Theory Methods* **32**, 215-237.

[DOI: 10.18869/acadpub.jsri.1.2.109]

Hájek, J. (1959). Optimum strategy and other problems in probability sampling. *Časopis Pěst. Mat.* **84**, 387-423.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47**, 663-685.

Lawry, K.A. and Bellhouse, D.R. (1992). Relative efficiency of certain randomization procdures in an $n \times n$ array when spatial correlation is present. *J. Statist. Plann. Inference* **32**, 385-399.

McKenzie, N.L., Robinson, A.C., and Belbin, L. (1991). Biogeographic survey of the nullarbor district, Australia. In *Nature Conservation: Cost Effective Biological Survey and Data Analysis*, C.R. Margules and M.P. Austin, eds. pp. 109-126, CSIRO, Australia.

Munholland, P.L. and Borkowski, J.J. (1996). Latin square sampling +1 design: a spatial design using quadrats. *Biometrics* **52**, 125-136.

Salehi, M.M. (2002). Systematic simple latin square sampling (+1) design and its optimality. *J. Propagations Probab. Statist.* **2**, 191-200.

Salehi, M.M. (2004). Optimal sampling design under a spatial correlation model. *J. Statist. Plann. Inference* **118**, 8-19.

Salehi, M.M. (2006). Adaptive cluster row and column elimination sampling design +1. *Comm. Statist. Theory and Method.* **35**, No. 2. to appear.

Salehi, M.M. and Smith, D. (2005). Two-stage sequential Sampling design: a neighborhood-free adaptive sampling Procedure, *J. Agric. Biol. Environ. Stat.* **10**, No. 1, 84-103.

Schreuder,. H.T., Gregoire, T.G., and Wood, G.B. (1993). *Sampling Methods for Multiresource Forest Inventory.* Wiley, New York.

Thompson, S.K. (1990). Adaptive cluster sampling. *J. Amer. Statist. Assoc.* **85**, 1050-1059.

Thompson, S.K. (2002). Sampling, 2nd ed. Wiley, New York.

پیوست

فرض کنید A پیشامد این باشد که RCES به اندازه‌ی n شامل واحد iام باشد. پس از حذف سطر و ستون واحد iام، $(n-۱)$ نمونه از $(R-۱)(C-۱)$ واحد باقی‌مانده به‌روش حذف سطر و ستون انتخاب می‌شوند. آن واحد اضافی (۱+) به $(RC-n)$ روش می‌تواند انتخاب شود. بنا بر این،

$$|A| = \frac{\{\prod_{k=۱}^{n-۱}(R-k)(C-k)\}(RC-n)}{(n-۱)!}$$

به‌طوری که $|\cdot|$ نشان دهنده‌ی کاردینالیتی است. با تقسیم $|A|$ بر (۱) داریم:

$$\Pr(A) = \frac{n}{RC}.$$

فرض کنید B پیشامد این که واحد i به‌عنوان واحد اضافی ۱+ انتخاب شود، باشد. بنا بر این،

$$\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|A^c)\Pr(A^c).$$

از آن‌جا که $\Pr(B|A) = ۰$ و $\Pr(B|A^c) = \frac{۱}{RC-n}$

$$\begin{aligned}\Pr(B) &= \Pr(B|A^c)\Pr(A^c)\\ &= \Pr(B|A^c)(۱-\Pr(A))\\ &= \frac{۱}{RC-n}\left(۱-\frac{n}{RC}\right) = \frac{۱}{RC}\end{aligned}$$

پیشامد این که RCES+۱ شامل واحد iام باشد $A \cup B$ است.

$$\begin{aligned}\Pr(A\cup B) &= \Pr(A) + \Pr(B)\\ &= \frac{n+۱}{RC}.\end{aligned}$$

[DOI: 10.18869/acadpub.jsri.1.2.109]

۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰
۰	۰	۱	۰	۰	۰	۱	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰
۱	۰	۱	۱	۱	۰	۱	۰	۰	۱	۱	۲	۱	۰	۰	۰	۰	۰	۰	۰
۱	۰	۰	۰	۱	۰	۰	۰	۱	۰	۰	۲	۳	۰	۱	۰	۰	۰	۰	۰
۱	۰	۰	۰	۱	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
۰	۰	۰	۱	۰	۰	۰	۰	۰	۱	۱	۰	۰	۰	۰	۱	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۲	۰	۰	۱	۰	۰	۰	۱	۱
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۲	۰	۰	۱	۱	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۱	۰	۰	۰	۰	۱
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۳	۰	۰	۱	۱	۱	۰	۲	۰	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۲	۱	۰	۱	۰	۰	۲	۲	۱	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۱	۲	۱	۰	۱
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۱	۰	۰	۰	۱	۱	۱	۱	۲
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۱	۱	۰	۰	۰	۱	۰	۱	۲
۰	۰	۰	۰	۰	۰	۰	۰	۰	۲	۰	۱	۱	۲	۰	۰	۱	۰	۱	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۱	۰	۱
۰	۰	۰	۰	۰	۰	۰	۰	۱	۱	۱	۰	۰	۰	۱	۰	۰	۰	۱	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۱	۲	۰
۰	۰	۰	۰	۰	۰	۰	۱	۰	۳	۱	۲	۱	۰	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۱	۰	۲	۱	۰	۰	۲	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۰	۰	۲	۰	۱	۰	۱	۰	۰	۱	۰	۰	۱
۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۲	۰	۰	۰	۰	۰	۱	۴	۱
۰	۰	۰	۰	۰	۰	۰	۲	۲	۱	۰	۱	۰	۰	۰	۰	۰	۰	۰	۱
۰	۰	۰	۰	۰	۰	۰	۱	۰	۱	۰	۰	۰	۱	۱	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۰	۱	۱	۱	۰	۰	۳	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۱	۱	۳	۱	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۱	۰	۱	۰	۱	۱	۱	۳	۲	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۱	۰	۰	۱	۰	۰	۱	۰	۱	۰	۰	۰	۰	۱	۰	۰
۰	۰	۰	۰	۰	۰	۱	۰	۲	۱	۱	۰	۱	۰	۱	۱	۱	۰	۰	۰
۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۱	۰	۰	۱	۰	۰	۰	۰	۰
۰	۰	۰	۱	۰	۰	۱	۲	۰	۲	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰
۰	۰	۰	۰	۰	۱	۵	۱	۱	۰	۱	۱	۰	۰	۰	۰	۰	۰	۱	۱
۰	۰	۰	۰	۰	۰	۱	۰	۰	۱	۴	۱	۰	۱	۱	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۴	۰	۱	۲	۱	۱	۰	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۱	۲	۰	۳	۱	۰	۰	۰	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۱	۱	۰	۰	۱	۰	۰	۲	۳	۰	۰	۰	۰	۰
۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۴	۰	۲	۲	۰	۰	۰	۰	۱	۰
۰	۰	۰	۱	۰	۰	۰	۳	۰	۱	۲	۱	۲	۰	۰	۰	۰	۱	۱	۰
۰	۰	۰	۰	۰	۰	۰	۰	۱	۱	۰	۱	۰	۰	۰	۰	۰	۰	۱	۰

شکل ۱. تعداد صدف‌های آب شیرین که از افراز قسمتی از رودخانه به هشت‌صد چهارضلعی به دست آمده است.

[DOI: 10.18869/acadpub.jsri.1.2.109]

دریافت: ۱۳ مهر ۱۳۸۳
آخرین اصلاح: ۶ شهریور ۱۳۸۴

محمد صالحی مرزیجرانی
گروه آمار، دانشکده‌ی علوم ریاضی،
دانشگاه صنعتی اصفهان،
اصفهان، ایران.
پیام‌نگار: $salehi\_m@cc.iut.ac.ir$

محمدامین جمال‌زاده
پژوهشکده‌ی آمار،
خیابان فاطمی، خیابان باباطاهر، خیابان سرتیپ فکوری،
تهران، ایران.
پیام‌نگار: $amin\_jamalzadeh@srtc.ac.ir$