نمونه‌گیری از نمونه طبقه‌ای به روش تصادفی طبقه‌ای

علی فتح‌خانی، جواد قاسمیان
گروه آمار، دانشگاه دامغان

چکیده

جامعه‌ای با N عنصر را در نظر می‌گیریم. فرض کنیم بتوان یک نمونه تصادفی با طبقه‌بندی به حجم n واحد از این جامعه گرفت ($n \leq N$). اگربخواهیم بعد از گذشت مدتی خاص دوباره برآوردهایی برای پارامترهای جامعه (به شرط آنکه عناصر جامعه تغییر نکرده باشند) به‌دست آوریم، روش مطرح امروزی این است که دوباره به جامعه مراجعه کرده و نمونه‌ای دیگر از همان جامعه بگیریم. می‌دانیم که در نمونه‌گیری طبقه‌ای چارچوب در دسترس است. حال اگر فهرست عناصری که در نمونه اولیه به حجم n انتخاب شده‌اند را به عنوان چارچوب نمونه‌گیری جدیدی در نظر بگیریم، به عبارت دیگر اگر عناصر منتخب در نمونه n عضوی را به عنوان عناصر جامعه جدید به حجم n تلقی کنیم، می‌توان با طبقه‌بندی کردن این عناصر به روش‌های موجود، نمونه‌ای جدید به حجم n^* ($n^* \leq n$) به روش تصادفی با طبقه‌بندی از این جامعه جدید استخراج کرد. نشان می‌دهیم که میانگین وزنی میانگین‌های طبقات، برآوردی نااریب برای میانگین جامعه اصلی است، که شاید بعد از مدتی دچار تغییر اندازه‌ها شده باشد. یکی از مزایای این روش صرفه‌جویی در وقت و هزینه است که برای طرح‌های نمونه‌گیری اداره‌ها و سازمان‌ها معرفی می‌شود.

واژه‌های کلیدی: نمونه‌گیری تصادفی ساده با طبقه‌بندی، نمونه‌گیری بدون جایگذاری، نمونه‌گیری طبقه‌ای در حالت تخصیص متناسب.

۱ مقدمه

در بحث نمونه‌گیری تصادفی ساده بدون جایگذاری، یکی از قضایای مهم و کاربردی قضیه زیر است.

قضیه ۱.۱. *زیر نمونه تصادفی ساده بدون جایگذاری به حجم n^* از نمونه تصادفی ساده بدون جایگذاری به حجم n ($n^* \leq n$)، خود نمونه‌ای تصادفی ساده بدون جایگذاری به حجم n^* از جامعه اصلی به حجم N است.*

مفهوم قضیه بالا این است که براساس نمونه‌ای به حجم n^*، می‌توان به برآوردهایی (نااریب) برای پارامترهای مجهول جامعه اصلی دست یافت. از این قضیه می‌توان جهت تعمیم ایده به روش نمونه‌گیری طبقه‌ای استفاده کرد. در واقع می‌خواهیم ببینیم آیا گرفتن نمونه‌های تصادفی طبقه‌ای مکرر منجر به نتیجه‌ای مشابه می‌شود یا خیر؟ هدف این مقاله بسط چنین ایده‌ای است.

جامعه‌ای به حجم N در نظر گرفته و با روش نمونه‌گیری

تصادفی با طبقه‌بندی، نمونه‌ای به حجم $n\,(n \leq N)$ از آن استخراج می‌کنیم. مطابق روال معمول این نوع نمونه‌گیری، اگر $\bar{Y}_{\mathrm{N}}$ میانگین مجهول جامعه باشد، می‌توان از برآوردگر $\bar{Y}_{st} = \sum\limits_{h=۱}^{L} w_h \bar{y}_h$ برای تخمین آن استفاده کرد. بدیهی است مشخصه‌های (اندازه‌های) عناصر این جامعه (که شامل متغیر مورد نظر ما یعنی Y نیز می‌شود) بعد از گذشت زمان، تغییر می‌کنند. یکی از روش‌های معمول برای سنجش میزان تغییر پارامترهای این جامعه رجوع به جامعه اصلی و انجام یك نمونه‌گیری جدید است. شاید به دلایلی از جمله کمبود زمان، هزینه و دلایل دیگر نخواهیم یا نتوانیم با مراجعه به جامعه اصلی نمونه‌گیری انجام دهیم. با توجه به ایده مذکور می‌توانیم عناصر نمونه اخذ شده در مرحله اول را به عنوان عناصر جامعه‌ای جدید در نظر بگیریم. دقت شود که منظور عناصر جامعه و نه اندازه است. پیشنهاد این است که این n عنصر را با کمك روش‌های ساخت طبقات به K طبقه تقسیم کرده وسپس نمونه‌ای جدید به روش طبقه‌ای به حجم $n^*\,(n^* \leq n)$ از آن بگیریم. نشان می‌دهیم که میانگین وزنی میانگین‌های طبقات جدید برآوردی نااریب برای میانگین جامعه تحول یافته اصلی است. شرطی که باید به آن توجه داشت این است که نمونه اولیه به روش طبقه‌ای با تخصیص متناسب گرفته شده باشد. نکته قابل توضیح دیگر این است که عناصر جامعه اصلی نباید تغییر کرده باشند و حجم نمونه اولیه به اندازه‌ای باشد که بتوان نمونه دوم را از آن اخذ کرد.

۲ نمونه‌گیری تصادفی با طبقه‌بندی (نمونه اول)

در طرح نمونه‌گیری طبقه‌ای، چارچوب نمونه‌گیری از اهمیت بالایی برخوردار است. فرض کنیم جامعه‌ای N عنصری را برای به‌دست آوردن برآوردهایی در مورد پارامترهای مجهول آن در نظر بگیریم. از این جامعه، n عنصر به تصادف با اصول روش طبقه‌ای جهت عضویت در نمونه انتخاب می‌شوند. بعد از انتخاب عناصر، اطلاعات مورد نیاز راجع به ویژگی مورد مطالعه برای هر عنصر نمونه را به‌دست می‌آوریم. تحلیل‌های آماری به منظور یافتن برآوردهای نااریب برای پارامترهای مجهول جامعه براساس اندازه همین ویژگی‌ها، انجام می‌شوند.

۱.۲ تعاریف و نمادها

در این بررسی واحدهای جامعه را با $Y_۱, Y_۲, \ldots, Y_N$ نشان می‌دهیم که همان اندازه‌های مجهول جامعه هستند. در نمونه‌گیری تصادفی ساده با طبقه‌بندی، ابتدا جامعه به حجم N را به L طبقه تقسیم می‌کنیم که این طبقه‌ها یا در جامعه وجود دارند (مانند تعداد کلاس‌ها در یک مدرسه) یا با روش‌های ساخت طبقه ایجاد می‌شوند. در هر صورت حجم این طبقه‌ها را $N_۱, N_۲, \ldots, N_L$ در نظر می‌گیریم. این طبقه‌ها متداخل نبوده و هر واحد جامعه به یک و تنها یک طبقه تعلق دارد. در واقع طبقه‌ها افرازی از جامعه به وجود می‌آورند. پس $N = \sum\limits_{h=۱}^{L} N_h$. این طبقه‌ها باید به گونه‌ای ساخته شوند که همگنی درون هر طبقه زیاد باشد. برای به‌دست آوردن برآوردهای دقیق باید مقادیر $N_h\,(h = ۱, ۲, \ldots, L)$ را بدانیم. وقتی طبقه‌ها و حجم آن‌ها مشخص شد، از هر طبقه نمونه‌ای به روش تصادفی

ساده بدون جایگذاری انتخاب می‌کنیم. انتخاب نمونه در هر طبقه مستقل از سایر طبقه‌هاست. حجم نمونه گرفته شده از هر طبقه را به ترتیب با $n_1,n_2,\ldots,n_L$ نشان می‌دهیم. در این صورت $n=\sum_{h=1}^{L} n_h$، معرف حجم نمونه از کل جامعه است. به این نمونه، نمونه تصادفی با طبقه‌بندی می‌گویند. تعداد طبقه‌ها را با L نشان می‌دهیم. زیرنویس h، معرف شماره طبقه است که از ۱ تا L تغییر می‌کند. زیرنویس i برای تعیین شماره واحدها در داخل هر طبقه به کار می‌رود. N تعداد کل افراد جامعه و $\bar{Y}_N=\frac{1}{N}\sum_{i=1}^{N} Y_i$، میانگین جامعه است. نمادهای زیر برای طبقه hام، $h=1,2,\ldots,L$ تعریف می‌شوند.

- تعداد کل واحدها در طبقه hام (N_h)،
- تعداد واحدهای نمونه از طبقهhام (n_h)،
- مقدار صفت واحد iام در طبقه hام $(Y_{h_i} i=1,2,\ldots,N_h)$،
- مقدار صفت واحدiام در نمونه گرفته شده از طبقه hام $(y_{h_i} \quad i=1,2,\ldots,n_h)$،
- نسبت تعداد واحدهای طبقه hام به تعداد واحدهای کل جامعه یا وزن طبقه hام $(w_h=\frac{N_h}{N})$،
- کسر نمونه‌گیری برای طبقه hام $\left(f_h=\frac{n_h}{N_h}\right)$،
- میانگین طبقه hام $\left(\bar{Y}_h=\frac{1}{N_h}\sum_{i=1}^{N_h} Y_{h_i}\right)$،
- میانگین نمونه طبقه hام $\left(\bar{y}_h=\frac{1}{n_h}\sum_{i=1}^{n_h} y_{h_i}\right)$.

میانگین وزنی میانگین‌های نمونه‌ای طبقه‌ها را میانگین نمونه با طبقه‌بندی نامیده و آن‌را با $\bar{Y}_{st}$ نشان می‌دهیم.

$$\begin{aligned}\bar{Y}_{st} &= \frac{1}{N}\sum_{h=1}^{L} N_h\bar{y}_h\\ &= \sum_{h=1}^{L} w_h\bar{y}_h.\end{aligned}$$

توجه داریم که این نمونه با نمونه تصادفی به حجم n از کل جامعه، تفاوت دارد.

قضیه ۱.۲. *اگر $\bar{Y}_h$ برآوردکننده ناااریب میانگین طبقه hام $(h=1,2,\ldots,L)$ باشد، آنگاه $\bar{Y}_{st}$ برآوردکننده نااریب میانگین کل جامعه یعنی $\bar{Y}_N$ است، به عبارتی داریم،*

$$E\left(\bar{Y}_{st}\right)=\bar{Y}_N.$$

برهان: رجوع شود به [۲].

۲.۲ بیان مسئله

فرض کنیم نمونه‌ای تصادفی به حجم n از جامعه‌ای به حجمN با روش طبقه‌ای گرفته‌ایم. اگر فهرست عناصری که نمونه n عنصری فوق را تشکیل می‌دهند، به عنوان چارچوب نمونه‌گیری جدید در نظر بگیریم یا به عبارتی اگر عناصر این نمونه را به عنوان عناصر جامعه‌ای جدید تلقی کنیم، می‌توان این جامعه جدید را براساس قواعد ساخت طبقه به K طبقه تقسیم کرد. حال از این جامعه، نمونه‌ای به حجم n^* به روش طبقه‌ای به‌دست می‌آوریم. نشان خواهیم داد که میانگین وزنی میانگین‌های نمونه‌ای طبقه‌ها (از این نمونه نیز) برآوردی نااریب برای میانگین اندازه ویژگی عناصر جامعه اصلی است. این اندازه‌ها احتمالا بعد از گذشت زمان دچار تغییر شده و بر این

اساس میانگین اندازه‌ها نیز تغییر کرده است.

۳ نمونه‌گیری تصادفی با طبقه‌بندی از نمونه اول

برای اجرای نمونه‌گیری طبقه‌ای از نمونه اول، ابتدا به لیست نمونه‌گیری جدید مراجعه و شروع به طبقه‌بندی کردن این جامعه جدید می‌کنیم، شاید جامعه جدید از قبل به صورت طبیعی حالت طبقه‌ای داشته باشد (در مثال ۱.۳ به این مورد اشاره شده است)، سپس با توجه به نوع نمونه‌گیری طبقه‌ای (متناسب، نیمن، اپتیمم و ...)، نمونه‌ای به حجم n^* از آن می‌گیریم. بعد از انجام تحلیل‌های آماری براساس داده‌های حاصل از نمونه به حجم n^*، نشان خواهیم داد که می‌توان برآوردهایی نااریب برای پارامترهای مجهول جامعه اصلی، که احتمالا اندازه ویژگی عناصر آن بعد گذشت زمان دچار تغییر شده است، به‌دست آورد.

۱.۳ تعاریف و نمادها

فرض کنیم همه واحدهای اخذ شده از طبقات در مرحله اول یعنی $y_{h_1}, y_{h_2}, \dots, y_{h_{n_h}}$ $(h = 1, 2, \dots, L)$ که در مجموع n واحد هستند را روی هم ریخته و مجدداً اندیس گذاری کنیم. این نمونه‌ی اندیس گذاری شده را با $y_1, y_2, \dots, y_n$ نشان می‌دهیم. توجه شود که این اندازه‌ها متعلق به واحدهایی از جامعه هستند که آن‌ها را، مجدداً طبقه‌بندی می‌کنیم.

در این مرحله می‌توان با نمونه‌گیری مجدد، هم برآوردی برای اندازه‌های تغییر یافته ویژگی قبلی عناصر و هم برای اندازه‌های ویژگی جدیدی که قصد مطالعه آن‌را داریم و قبلا این ویژگی مورد بررسی قرار نگرفته، به‌دست آورد. اندازه تغییر یافته ویژگی مورد مطالعه یا اندازه ویژگی جدید را با $Y_1^*, Y_2^*, \dots, Y_n^*$ نشان می‌دهیم. برای به‌دست آوردن برآوردهایی از پارامترهای مجهول این جامعه جدید روش نمونه‌گیری تصادفی طبقه‌ای را انتخاب می‌کنیم. جامعه را به K طبقه به حجم‌های $M_1, M_2, \dots, M_K$ تقسیم می‌کنیم. به طوری که $n = \sum_{h=1}^{K} M_h$.

طبقه‌ها باید به گونه‌ای باشند که همگنی درون طبقه‌ها زیاد باشد. از هر طبقه نمونه‌ای تصادفی بدون جایگذاری انتخاب می‌کنیم. فرض کنیم حجم نمونه‌ها در طبقه‌ها به ترتیب $m_1, m_2, \dots, m_K$ باشد، به طوری که $n^* = \sum_{h=1}^{K} m_h$. n^* حجم کل نمونه در نمونه‌گیری این مرحله (مرحله دوم) است.

از نماد Y° برای اندازه جدید ویژگی مورد مطالعه (خواه جدید، خواه اندازه تغییر یافته ویژگی قبلی) استفاده می‌کنیم. بنابراین نمادهای زیر معرفی می‌شوند $(h = 1, 2, \dots, L)$،

- اندازه جدید ویژگی مورد مطالعه جامعه اصلی $(Y_1^\circ, Y_2^\circ, \dots, Y_N^\circ)$،
- میانگین جامعه اصلی $\left(\bar{Y}_N^\circ = \frac{1}{N}\sum_{i=1}^{N} Y_i^\circ\right)$،
- مقدار صفت واحد iام در طبقه hام $(Y_{h_i}^\circ i = 1, 2, \dots, N_h)$،
- مقدار صفت واحد iام در نمونه گرفته شده از طبقه hام $(y_{h_i}^\circ \quad i = 1, 2, \dots, n_h)$،
- میانگین طبقه hام $\left(\bar{Y}_h^\circ = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{h_i}^\circ\right)$،
- میانگین نمونه طبقه hام $\left(\bar{y}_h^\circ = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{h_i}^\circ\right)$.

در واقع می‌توان چنین تصور کرد که نمونه $Y_1^*, Y_2^*, \dots, Y_n^*$ نمونه‌ای طبقه‌ای از جامعه $Y_1^\circ, Y_2^\circ, \dots, Y_N^\circ$ است و ما برای افزایش دقت با در نظر گرفتن این نمونه به عنوان جامعه‌ای جدید از آن نمونه‌ای طبقه‌ای گرفته و برآوردیابی می‌کنیم. بعد از طبقه‌بندی این نمونه $(Y_1^*, Y_2^*, \dots, Y_n^*)$ نمادگذاری زیر را داریم $(h = 1, 2, \dots, K)$،

- میانگین جامعه جدید(مجهول) $\left(\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^*\right)$،
- تعداد کل واحدها در طبقه hام (M_h)،
- تعداد واحدهای نمونه از طبقه hام (m_h)،
- مقدار صفت واحد iام در طبقه hام $(Y_{h_i}^{**} i = 1, 2, \dots, M_h)$،
- مقدار صفت واحد iام در نمونه گرفته شده از طبقه hام $(y_{h_i}^{**} \quad i = 1, 2, \dots, m_h)$،
- نسبت تعداد واحدهای طبقه hام به تعداد واحدهای کل جامعه یا وزن طبقه hام $\left(w_h^{**} = \frac{M_h}{n}\right)$
- میانگین طبقه hام $\left(\bar{Y}_h^{**} = \frac{1}{M_h}\sum_{i=1}^{M_h} Y_{h_i}^{**}\right)$،
- میانگین نمونه طبقه hام $\left(\bar{y}_h^{**} = \frac{1}{m_h}\sum_{i=1}^{m_h} y_{h_i}^{**}\right)$.

میانگین وزنی میانگین‌های نمونه‌ای طبقه‌ها را میانگین نمونه با طبقه‌بندی از جامعه جدید نامیده و آن‌را با $\bar{Y}_{st}^*$ نشان می‌دهیم.

$$\begin{aligned}\bar{Y}_{st}^* &= \frac{1}{n}\sum_{h=1}^{K} M_h \bar{y}_h^{**} \\ &= \sum_{h=1}^{K} w_h^{**} \bar{y}_h^{**}.\end{aligned}$$

با توجه به تغییر در اندازه ویژگی عناصر جامعه اصلی و نمونه گرفته شده از آن باید به این نکته توجه کنیم که اندازه میانگین وزنی میانگین‌های نمونه‌ای طبقه‌ها از جامعه اصلی نیز تغییر کرده و داریم:

$$\begin{aligned}\bar{Y}_{st}^\circ &= \frac{1}{N}\sum_{h=1}^{L} N_h \bar{y}_h^\circ \\ &= \sum_{h=1}^{L} w_h \bar{y}_h^\circ.\end{aligned}$$

برای تشکیل نمونه به حجم n^* باید واحدهای نمونه‌های گرفته شده از طبقات یعنی $y_{h_1}^{**}, y_{h_2}^{**}, \dots, y_{h_{m_h}}^{**}$ را روی هم ریخته و نمادگذاری جدیدی روی این واحدها انجام دهیم. این n^* واحد را با $y_1^*, y_2^*, \dots, y_{n^*}^*$ نشان می‌دهیم. $(h = 1, 2, \dots, K)$ [۵، ۶، ۷].

۲.۳ معرفی برآوردگر

می‌دانیم اگر از جامعه‌ای به حجم N به روش طبقه‌ای، نمونه‌ای به حجم n بگیریم و $n_1, n_2, \dots, n_L$ حجم نمونه از طبقه‌ها به ترتیب به حجم $N_1, N_2, \dots, N_L$ باشند. در این صورت تعداد کل نمونه‌های ممکن برابر است با،

$$A = \prod_{h=1}^{L} \binom{N_h}{n_h}.$$

بنابراین به تعداد A حالت می‌توان نمونه به حجم n به‌دست آورد. فرض کنیم که این A نمونه را داشته باشیم و آن‌ها را از ۱ تا A شماره گذاری کنیم، در هنگام انتخاب نمونه به تصادف یکی از این A نمونه برای ما رخ می‌دهد. پیشامد رخ دادن یکی از این نمونه‌ها کاملا تصادفی بوده و احتمال رخ دادن هر یک از آن‌ها $\frac{1}{A}$ است. در این صورت

فقط با یک نمونه، مثلا نمونه jام ($j = 1, 2, \dots, A$) مواجه هستیم. با انتخاب نمونه jام فرض کنیم، میانگین اندازه‌های تغییر یافته برابر $\bar{Y}_n$ باشد. در این صورت میانگین وزنی میانگین‌های نمونه‌ای طبقه‌ها از جامعه جدید ($\bar{Y}^*_{st}$) برآوردگری نااریب برای $\bar{Y}_n$ است. به عبارتی دیگر $\bar{Y}^*_{st}$ برآوردگری نااریب برای $\bar{Y}_n$ است به شرط آنکه j از قبل تثبیت شده باشد. یعنی،

$$E\left(\bar{Y}^*_{st} | j\right) = \bar{Y}_n. \qquad (1)$$

ادعا می‌کنیم که $\bar{Y}^*_{st}$ برآوردی نااریب برای میانگین جامعه اصلی به حجم N یعنی $\bar{Y}^\circ_N$ است.

۳.۳ خاصیت نااریبی

برای یک نمونه تصادفی ساده به حجم n^* از جامعه‌ای به حجم n (نمونه دوم)، تعداد نمونه‌های ممکن برابر است با،

$$S = \prod_{h=1}^{K} \begin{pmatrix} M_h \\ m_h \end{pmatrix}.$$

در این قسمت لازم است که یک تعریف بنیادین از نمونه‌گیری تصادفی ساده با طبقه‌بندی با حالت تخصیص متناسب را بیان کنیم.

تعریف ۱.۳. *اگر در نمونه‌گیری تصادفی ساده با طبقه‌بندی، حجم نمونه‌های انتخاب شده از طبقه‌ها متناسب با حجم طبقه‌ها باشند، نمونه‌گیری را با* ***تخصیص متناسب***[1] *می‌نامند. در واقع در این نوع نمونه‌گیری داریم* ($h = 1, 2, \dots, L$)،

$$\frac{n_h}{n} = \frac{N_h}{N} = w_h.$$

با توجه به تعریف بالا نتیجه می‌گیریم که:

$$\frac{1}{n} = \frac{N_h}{N} \times \frac{1}{n_h} \quad \Rightarrow \quad \frac{1}{n} = w_h \times \frac{1}{n_h}. \qquad (2)$$

از طرفی،

$$\begin{aligned} \bar{Y}^\circ_{st} &= \frac{1}{N} \sum_{h=1}^{L} N_h \bar{y}^\circ_h \\ &= \sum_{h=1}^{L} w_h \bar{y}^\circ_h, \end{aligned} \qquad (3)$$

با توجه به روابط (۲)، (۳) و تعریف $\bar{y}^\circ_h$ داریم،

$$\begin{aligned} \bar{Y}^\circ_{st} &= \sum_{h=1}^{L} w_h \times \frac{1}{n_h} \sum_{i=1}^{n_h} y^\circ_{h_i} \\ &= \frac{1}{n} \sum_{h=1}^{L} \sum_{i=1}^{n_h} y^\circ_{h_i} \\ &= \frac{1}{n} \sum_{i=1}^{n} Y^*_i \\ &= \bar{Y}_n. \end{aligned} \qquad (4)$$

براساس قضیه ۱.۲ رابطه زیر را داریم،

$$E\left(\bar{Y}^\circ_{st}\right) = \bar{Y}^\circ_N. \qquad (5)$$

با توجه به رابطه‌های (۴) و (۵) داریم،

$$E\left(\bar{Y}_n\right) = \bar{Y}^\circ_N, \qquad (6)$$

[1] Proportional allocation

و از روابط (۱) و (۶) نتیجه می‌شود که،

$$E\left(E\left(\bar{Y}^{*}_{st}|J=j\right)\right)=\bar{Y}^{\circ}_{N},\quad j=۱,۲,\ldots,A.\quad (۷)$$

برای رسیدن به نتیجه مطلوب لازم است به یک گزاره مهم در احتمال اشاره کنیم،

$$E\left(E\left(X|Y\right)\right)=E\left(X\right).$$

با توجه به گزاره فوق و روابط (۱) تا (۷) به هدف اصلی خود یعنی،

$$E\left(\bar{Y}^{*}_{st}\right)=\bar{Y}^{\circ}_{N},$$

می رسیم [۱، ۳، ۴].

مثال ۱.۳. *فرض کنیم پژوهشگری علاقه‌مند به اندازه‌گیری ویژگی‌های عناصر جامعه‌ای مانند جامعه دانش‌آموزان مشغول به تحصیل در مدرسه‌ای با ۲۰۸ دانش‌آموز است. این پژوهشگر برای رسیدن به هدف خود از روش برآورد نمونه‌ای بهره می‌گیرد. او با مراجعه به دفتر آموزش مدرسه و تهیه لیست تمام کلاس‌های دایر و فهرست کردن اسامی دانش‌آموزان، چارچوب نمونه‌گیری را تشکیل می‌دهد(در ابتدای سال تحصیلی).*

فرض کنیم این مدرسه داری ۱۱ کلاس در ۳ طبقه به ترتیب در طبقه همکف ۳ کلاس، در طبقه اول و دوم هر کدام ۴ کلاس باشد. تعداد دانش‌آموزان کلاس‌های طبقه همکف به ترتیب ۱۷، ۱۴ و ۲۱ دانش‌آموز، در طبقه اول به تعداد ۱۸، ۲۲، ۱۹ و ۱۶ دانش‌آموز و در طبقه دوم به تعداد ۲۳، ۲۶، ۱۷ و ۱۵ دانش‌آموز است.

با توجه به اینکه ساختار جامعه مورد نظر طبقه‌ای است، پژوهشگر تصمیم می‌گیرد که از روش نمونه‌گیری طبقه‌ای برای یافتن برآوردهای خود استفاده کند. بنابراین هر کلاس این مدرسه را به عنوان یک طبقه در نظر گرفته و نمونه‌برداری خود را به روش تصادفی با طبقه‌بندی و با تخصیص متناسب انجام می‌دهد. کلاس با تعداد ۱۷ نفر طبقه اول جامعه را می‌سازد، به همین ترتیب کلاس با تعداد ۱۴ طبقه دوم،... و کلاس با تعداد ۱۵ نفر طبقه یازدهم را تشکیل می‌دهد.

هدف نخست پژوهشگر، برآورد میانگین وزن دانش‌آموزان مدرسه است که با مراجعه به هر کلاس و انتخاب تعدادی از دانش‌آموزان به صورت تصادفی ساده بدون جایگذاری به عنوان عناصر نمونه و محاسبه وزن دانش‌آموزان منتخب در نمونه، اندازه مربوط به وزن هر دانش‌آموز را ثبت می‌کند. از هر کلاس متناسب با حجم آن، نمونه‌برداری انجام شده است.

ذکر این نکته ضروری است که با توجه به تخصیص متناسب، حجم هر نمونه از طبقات به ترتیب ۷، ۵، ۸، ۷، ۸، ۷، ۶، ۹، ۱۰، ۷، ۶ است. یعنی در مجموع برابر است با،

$$n=\sum_{h=۱}^{۱۱}n_h=۸۰.$$

فرض کنیم که هیچ کدام از دانش‌آموزان مدرسه غایب نباشند، پس پژوهشگر با اندازه‌گیری وزن این ۸۰ دانش‌آموز و ثبت آن‌ها و استفاده از تعریف $\bar{Y}_{st}$ میانگین وزن دانش‌آموزان مدرسه را برابر ۵۷/۸ کیلوگرم برآورد می‌کند.

حال فرض می‌کنیم که این پژوهشگر بعد از گذشت مدتی (مثلا چند ماه) از برآورد وزن دانش‌آموزان، علاقه‌مند به مطالعه میزان تغییر وزن دانش‌آموزان باشد و بخواهد برآوردی نیز برای میزان فشار خون دانش‌آموزان این مدرسه به‌دست آورد. با توجه به روش بیان شده در این

مقاله و معرفی برآوردگری نااریب، پژوهشگر نیازی به مراجعه مجدد به جامعه ندارد. در این مثال پژوهشگر می‌تواند واحدهای نمونه ۸۰ عضوی را به عنوان عناصر جامعه جدید در نظر گرفته و با طبقه‌بندی مجدد آن‌ها به روش تصادفی طبقه‌ای، مجدداً نمونه‌ای از داخل این جامعه اختیار کند. لازم به توضیح است که با توجه به این که مدرسه ساختمانی سه طبقه‌ای بود، می‌توان نمونه‌های هر طبقه ساختمانی را به عنوان یک طبقه ساختار جدید در نظر گرفت.

نمونه‌گیری تصادفی طبقه‌ای از این جامعه ۸۰ عضوی به هر روش دلخواه می‌تواند انجام گیرد، مانند تخصیص متناسب، تخصیص نیمن، تخصیص اپتیمم و ... . جامعه جدید دارای سه طبقه است، طبقه اول همان طبقه همکف ساختمان مدرسه با ۲۰ (۷ + ۵ + ۸) عضو، طبقه دوم همان طبقه اول مدرسه با ۲۸ (۶ + ۷ + ۸ + ۷) عضو و طبقه سوم همان طبقه دوم مدرسه با ۳۲ (۶ + ۷ + ۱۰ + ۹) عضو است. پژوهشگر برای راحتی کار دوباره روش نمونه‌گیری تصادفی با طبقه‌بندی و با حالت تخصیص متناسب را در نظر می‌گیرد. هدف پژوهشگر در این مرحله اندازه‌گیری وزن‌های جدید و فشار خون دانش‌آموزان این مدرسه است.

با توجه به تخصیص متناسب، حجم نمونه از طبقه اول به تعداد ۸ دانش‌آموز، از طبقه دوم به تعداد ۱۰ دانش‌آموز و از طبقه سوم به تعداد ۱۲ دانش‌آموز است، یعنی حجم کل نمونه در این حالت ۳۰ است. با مراجعه به این عناصر و اندازه‌گیری وزن آن‌ها و ثبت فشار خونشان، پژوهشگر می‌تواند به برآوردهای دلخواه خود دست یابد. فرض می‌کنیم که پژوهشگر میانگین وزن دانش‌آموزان را این بار ۵۹/۴ کیلوگرم و میانگین فشار خون دانش‌آموزان مدرسه را ۷ روی ۱۲ برآورد می‌کند. این برآوردهای جدید حاکی از این است که در این مدت اخیر، میانگین وزن دانش‌آموزان کمی بالاتر رفته و فشار خون دانش‌آموزان مدرسه تقریبا نرمال است. در واقع میانگین به‌دست آمده از نمونه اخیر در مورد وزن ۵۹/۴ کیلوگرم می‌تواند به عنوان برآورد میانگین وزن دانش‌آموزان در کل مدرسه در نظر گرفته شود. این مطلب در مورد فشار خون نیز صادق است. حسن این کار در این است که جهت برآورد وزن یا متغیر جدیدی مانند فشار خون رجوع به جامعه اصلی نیاز نیست و کافی است به عناصر اخذ شده در نمونه‌گیری‌های قبلی، مراجعه کنیم.

۴ بحث و نتیجه‌گیری

در سازمان‌ها و اداره‌ها وقت و هزینه دو معیار از معیارهای مهم برای انجام پروژه‌های مختلف هستند. با توجه به ایده بیان شده در این مقاله، سازمان‌ها می‌توانند تا حد امکان در وقت و هزینه خود، صرفه‌جویی کنند. واضح است که نمونه‌گیری از جامعه‌ای به حجم n وقت و هزینه کمتری نسبت به نمونه‌گیری از جامعه‌ای بزرگتر به حجم N $(n \leq N)$ می‌طلبد (جامعه به حجم n همان نمونه n عضوی از جامعه به حجم N مورد بحث است). بنابراین سازمان‌ها می‌توانند به جای مراجعه دوباره به عناصر جامعه اصلی، عناصر نمونه به حجم n را به عنوان عناصر جامعه جدید در نظر گرفته و نمونه‌گیری خود را براساس همین n عنصر، انجام دهند. بدیهی است که دقت برآورد نمونه‌ای، از جامعه به حجم n، نسبت به دقت برآورد نمونه‌ای، از جامعه به حجم N کمتر (یا در حالت خاص مساوی) است.

مراجع

[1] Christofides, T. C. (2005). Randomized response in stratified sampling. Journal of statistical planning and inference, 128(1), 303-310.

[2] De Paschal, J. N. (1961). Unbiased ratio estimators in stratified sampling. Journal of the American Statistical Association, 56(293), 70-87.

[3] Ericson, W. A. (1965). Optimum stratified sampling using prior information. Journal of the American Statistical Association, 60(311), 750-771.

[4] Etoré, P. and Jourdain, B. (2010). Adaptive optimal allocation in stratified sampling methods. Methodology and Computing in Applied Probability, 12(3), 335-360.

[5] Imbens, G. W. and Lancaster, T. (1996). Efficient estimation and stratified sampling. Journal of Econometrics, 74(2), 289-318.

[6] Singh, H. P. and Vishwakarma, G. K. (2008). A family of estimators of population mean using auxiliary information in stratified sampling. Communications in Statistics –Theory and Methods, 37(7), 1038-1050.

[7] Tong, C. (2006). Refinement strategies for stratified sampling methods. Reliability Engineering and System Safety, 91(10), 1257-1265.