© Statistical Research and Training Center

مجله‌ی پژوهش‌های آماری ایران

J. Statist. Res. Iran **4** (2007): 47-70

دوره‌ی ۴، شماره‌ی ۱، بهار و تابستان ۱۳۸۶، صص ۴۷-۷۰

پژوهشکده‌ی آمار

نمونه‌گیری برینشی با الگوی سرشماری، نمونه‌گیری، حذف

محمد جعفری جوزانی[†,‡,*] و فرشید جمشیدی[‡]

[†] دانشگاه علامه طباطبایی

[‡] پژوهشکده آمار

چکیده. در این مقاله ضمن معرفی انواع روش‌های نمونه‌گیری برینشی، بعضی از دلایل استفاده و کاربرد آن‌ها در آمارگیری‌ها را توضیح می‌دهیم. همچنین روش نمونه‌گیری برینشی با الگوی سرشماری، نمونه‌گیری و حذف را که از آن به‌عنوان روش نمونه‌گیری نوع سوم یاد می‌کنیم مورد مطالعه قرار می‌دهیم. برای این منظور، نخست در براورد پارامتر میانگین (و در نتیجه مجموع) مقادیر صفت Y در جامعه‌ی متناهی Nعضوی مورد نظر، چگونگی ارائه‌ی براوردگرهای مطلوب و خواص آن‌ها را در صورت به‌کارگیری روش نمونه‌گیری برینشی نوع سوم توضیح داده، میزان اریبی و واریانس براوردگرهای ارائه‌شده را به دست می‌آوریم. نکته‌ی قابل توجه در این روش، چگونگی برخورد با قسمت حذف‌شده در تعیین دقت براوردهای به دست آمده است که آن را در قالب چند رویکرد توضیح می‌دهیم. در نهایت، مسئله‌ی تعیین آستانه‌ی برش و اندازه‌ی نمونه‌ی بهینه در روش نمونه‌گیری برینشی نوع سوم را توضیح داده، با یک مطالعه‌ی شبیه‌سازی‌شده نتایج به دست آمده را مورد ارزیابی قرار می‌دهیم.

واژگان کلیدی. آستانه‌ی برش؛ رویکرد حذف؛ رویکرد مدل‌یار؛ نمونه‌گیری برینشی.

---

* نویسنده‌ی عهده‌دار مکاتبات

۱ مقدمه

یک طرح نمونه‌گیری عبارت است از اتخاذ تصمیمی برای انتخاب اعضای جامعه‌ای که قصد داریم در باره‌ی آن اطلاعات آماری به دست آوریم. برای نمونه‌گیری از هر جامعه، روش‌های مختلف نمونه‌گیری وجود دارد که یکی از آن‌ها روش نمونه‌گیری برینشی و در نتیجه استفاده از طرح نمونه‌گیری برینشی است. این روش دارای کاربردهای فراوان است و همان‌طور که ناب (۲۰۰۷-آ) نیز اشاره می‌کند، یکی از دلایل (عملی) استفاده‌ی زیاد از روش‌های نمونه‌گیری برینشی در اداره‌های آمار و مراکز امور اقتصادی و کشاورزی، به‌ویژه در آمارگیری از کارگاه‌های صنعتی (هیلپینن، ۲۰۰۶) و در محاسبه‌ی انواع شاخص‌های اقتصادی، مانند شاخص قیمت مصرف‌کننده و شاخص قیمت تولیدکننده (هان و دیگران، ۱۹۹۹) صرف نظر از مشکلات به وجود آمده برای براوردگرهای ارائه‌شده در آن‌ها (مانند اریبی)، سادگی اجرا و مقرون به صرفه بودن آن است.

در شکل اولیه‌ی این روش، که از آن به‌عنوان نمونه‌گیری برینشی نوع اول تبدیل‌یافته (نمونه‌گیری، حذف) یاد می‌شود، بر خلاف روش‌های معمول نمونه‌گیری که در آن‌ها فرض می‌شود تمام اعضای جامعه، شانس مثبتی برای حضور در نمونه دارند، بخشی از جامعه به‌صورت آگاهانه از مطالعه حذف شده است (قسمت برش داده‌شده) و هیچ شانسی برای حضور در نمونه ندارد. ناب (۲۰۰۷-ب) و سرندال و دیگران (۱۹۹۷) نمونه‌گیری برینشی نوع اول تبدیل‌یافته را به‌صورت زیر تعریف می‌کنند:

> «در روش نمونه‌گیری برینشی، به‌کمک مکانیسم تصادفی، نمونه‌ی nتایی مورد نظر را فقط از بخشی از جامعه که با U_S نشان داده می‌شود گردآوری می‌کنیم، به‌طوری که برای هر $j \in U_S$ داریم $\pi_j > ۰$، در حالی که برای مابقی جامعه (که از آن به قسمت برش داده‌شده یاد می‌شود و با U_E نشان داده می‌شود)، یعنی $j \in U_E = U - U_S$ ، داریم $\pi_j = ۰$ که در آن π_j احتمال گزینش عضو jام جامعه در نمونه است.»

یکی از فرضیات اصلی در به‌کارگیری روش‌های نمونه‌گیری برینشی، آن است که قسمت حذف‌شده در طول بررسی دستخوش تغییرات جدی نشود. اما همان‌طور که استیل و فی (۱۹۹۵) توضیح می‌دهند، هرگاه بخشی از جامعه برای حضور در نمونه هیچ شانسی نداشته باشد و براوردیابی بر اساس داده‌های به دست آمده از قسمت باقی‌مانده و مدلی از آنچه در دست است صورت گیرد، همواره این احتمال وجود دارد که در طول بررسی آن بخش از جامعه که به دلایل مختلف (مثلاً تأثیر کم بر مقدار پارامتر مورد بررسی) از بررسی حذف گردیده است، دستخوش تغییرات زیاد شود و نادیده گرفتن آن منجر به خطاهای جدی در نتایج تحقیقات گردد، هرچند احتمال چنین اتفاقی کم باشد. این مسئله بخشی از خطای آمارگیری در طرح‌های نمونه‌گیری برینشی را نشان می‌دهد که در عمل نیز به آن توجه زیادی نمی‌شود و تا کنون نیز در مورد آن بررسی‌های قابل توجهی صورت نگرفته است. بنا بر این به نظر می‌رسد در عمل، استفاده از

این روش در آمارگیری‌ها به‌خاطر عواملی نظیر کاهش هزینه، سادگی اجرا و مواردی از این قبیل باشد و نه برای افزایش دقت براورد. از دیگر عوامل استفاده از نمونه‌گیری برینشی آن است که گاهی به‌دلیل شرایط اقتصادی، قابلیت دسترسی، امکانات و مواردی از این قبیل، مجبور به انتخاب نمونه در مکان‌هایی هستیم که آمارگیر و سایر شرایط لازم برای آمارگیری فراهم است. در چنین شرایطی استفاده از نمونه‌گیری برینشی اجتناب‌ناپذیر است و علاوه بر هزینه‌ی کم‌تر و انعطاف بیش‌تر، زمان دسترسی به نتایج آمارگیری در آن کم‌تر است.

برای مثال، در مطالعه‌ی میزان بهره‌وری واحدها و بنگاه‌های اقتصادی یا شاخص‌هایی از این قبیل، ممکن است واحدها یا بنگاه‌های اقتصادی‌ای وجود داشته باشند که سهم آن‌ها از مجموع کل بهره‌وری در جامعه، کم یا قابل چشم‌پوشی باشد (مثلاً بنگاه‌های اقتصادی با تعداد کارکنان کم‌تر از ۵ نفر) و در مقابل، بنگاه‌های اقتصادی‌ای وجود داشته باشند که حضور نداشتن آن‌ها در آمارگیری‌ها قابلیت اعتماد نتایجِ به دست آمده را با سؤالات جدی مواجه سازد. در این موارد، اغلب از روش‌های نمونه‌گیری برینشی استفاده می‌شود (مرکز آمار کانادا، ۲۰۰۱). داده‌های کارگاهی عمدتاً داده‌هایی با توزیع بسیار چوله می‌باشند و منطقی است که در آمارگیری از کارگاه‌های صنعتی، بخشی از جامعه به‌طور کامل در آمارگیری لحاظ شود و بخش دیگر به‌علت تأثیرات اندک و قابل چشم‌پوشی در شاخص‌های مورد نظر و با توجه به عوامل اشاره‌شده در بالا از بررسی حذف شود. در روش‌های نمونه‌گیری برینشی و در براورد پارامتر مورد نظر، عموماً دو نوع رویکرد در خصوص مقادیر صفت Y در قسمت حذف‌شده مورد استفاده قرار می‌گیرد که عبارت‌اند از:

- **رویکرد حذف**، که در آن، واحدهایی از جامعه که در ناحیه‌ی برش قرار می‌گیرند (U_E) کلاً از بررسی حذف می‌شوند و مقادیر Y مربوط به آن‌ها هیچ‌گونه نقشی در براورد پارامتر مورد نظر ندارند.
- **رویکرد مدل‌یار**، که در آن پس از تعیین آستانه‌ی برش و تشکیل طبقه‌ی U_E از طریق اطلاعات کمکی و با در نظر گرفتن براوردهای نسبتی و رگرسیونی یا هر نوع براوردگر مناسب دیگر، مجموع (یا میانگین) مقادیر صفت Y برای قسمت حذف‌شده براورد می‌گردد و از آن در براورد پارامتر مورد نظر استفاده می‌شود.

یک مسئله‌ی مهم در خصوص نمونه‌گیری برینشی، ایجاد تغییرات جدید در آن یا براوردگرهای به دست آمده از آن است به‌طوری که بتوان بدون افزایش زیاد در هزینه و از بین بردن سادگی اجرای آن، صحت و دقت نتایجِ به دست آمده را افزایش داد. همان‌طور که رویال (۱۹۷۰) نشان می‌دهد، در جوامع با چولگی زیاد، مانند کارگاه‌های صنعتی، برای اندازه‌ی نمونه‌ی n، حد اقل واریانس برای براوردها زمانی حاصل می‌شود که کارگاه‌های بزرگ یا بسیار بزرگ در نمونه حاضر باشند. این نکته یکی از انگیزه‌های اصلی برای استفاده از روش نمونه‌گیری برینشی با الگوی سرشماری، نمونه‌گیری و حذف است که از آن به‌عنوان روش

نمونه‌گیری برینشی نوع سوم یاد می‌کنیم و مورد مطالعه قرار می‌دهیم.

برای این منظور، در این مقاله نخست انواع روش‌های نمونه‌گیری برینشی را معرفی می‌کنیم و سپس در براورد پارامتر میانگین (و در نتیجه مجموع) مقادیر صفت Y در جامعه‌ی متناهی N-عضوی مورد نظر، چگونگی ارائه‌ی براوردگرهای مطلوب و خواص آن‌ها را در صورت به‌کارگیری روش نمونه‌گیری برینشی نوع سوم توضیح داده، میزان اریبی و واریانس براوردگرهای ارائه‌شده را به دست می‌آوریم. نکته‌ی قابل توجه در این روش، چگونگی برخورد با قسمت حذف‌شده در تعیین دقت براوردهای به دست آمده است، که آن را در قالب چند رویکرد توضیح می‌دهیم. همچنین مسئله‌ی تعیین آستانه‌ی برش و اندازه‌ی نمونه‌ی بهینه را در روش نمونه‌گیری برینشی نوع سوم توضیح داده، با یک مطالعه‌ی شبیه‌سازی‌شده، نتایجِ به دست آمده را مورد ارزیابی قرار می‌دهیم.

۲ انواع روش‌های نمونه‌گیری برینشی

در بیش‌تر مواقع با بررسی میزان تأثیر مقادیر صفت Y روی مقدار Y_+ (مجموع) یا $\bar{Y}$ (میانگین) در جامعه‌ی مورد نظر می‌توان بر حسب مقدار بزرگی Y_iها طبقات مختلفی از واحدهای جامعه تشکیل داد. مثلاً یک گروه می‌تواند شامل بخشی از اعضای جامعه باشد که تأثیر کم و ناچیزی در مقدار پارامتر مورد نظر دارد. گاهی دسترسی به این بخش از جامعه دشوار است یا ممکن است چارچوب مناسبی از آن‌ها در دست نباشد. گروه دیگر می‌تواند شامل بخشی از اعضای جامعه باشد که تأثیر قابل توجهی بر مقدار پارامتر مورد نظر دارند و حضور نداشتن آن‌ها در آمارگیری، اعتبار نتایجِ به دست آمده را با سؤالات جدی مواجه می‌کند. در نمونه‌گیری برینشی، بسته به نحوه‌ی برخورد با چنین گروه‌هایی از اعضای جامعه می‌توان چهار نوع روش نمونه‌گیری برینشی به‌صورت زیر ارائه کرد.

۲،۱ نمونه‌گیری برینشی نوع اول (سرشماری، حذف)

در این روش که از متداول‌ترین روش‌های نمونه‌گیری برینشی موجود است و در اکثر مطالعات اقتصادی، کارگاهی و مواردی از این قبیل دارای کاربرد می‌باشد، پس از تعیین آستانه‌ی برش مناسب، جامعه‌ی N-عضوی $U=\{۱,۲,\ldots,N\}$ به دو طبقه‌ی U_C و U_E تقسیم می‌شود که در آن، U_C بخشی از اعضای جامعه‌ی U را نشان می‌دهد که به‌طور کامل در آمارگیری شرکت داده می‌شوند و برای هر عضو $j\in U_C$ داریم $\pi_j=۱$. همچنین U_E (ناحیه‌ی برش داده‌شده) بخشی از جامعه‌ی U را نشان می‌دهد

که در آمارگیری شرکت داده نمی‌شود و به‌طور کامل از بررسی حذف می‌شود. در واقع برای هر $j \in U_E$ داریم $\pi_j = 0$. واضح است که $U = U_C \cup U_E$. طرح نمونه‌گیری برینشی نوع اول از نوع طرح‌های نااحتمالی است و براوردهای به دست آمده از آن اریب می‌باشند؛ اما با وجود این، هنوز هم در سطح گسترده‌ای در آمارگیری‌ها مورد استفاده قرار می‌گیرد.

۲٫۲ نمونه‌گیری برینشی نوع اول تبدیل‌یافته (نمونه‌گیری، حذف)

در این روش، پس از تقسیم جامعه به دو طبقه (بر حسب بزرگی Y)، فرض می‌شود U_S طبقه‌ای شامل واحدهایی از جامعه با مقادیر بزرگ Y باشد که به دلایل مختلف، امکان سرشماری آن فراهم نبوده یا دشوار است و در نتیجه نمونه‌ای nتایی از آن برای آمارگیری انتخاب می‌شود. همچنین U_E طبقه‌ای از جامعه را نشان می‌دهد که در آن، واحدهای دارای مقادیر کوچک Y قرار دارند که با توجه به رویکرد مورد استفاده، یا به‌طور کامل از بررسی حذف می‌شوند یا مجموع مقادیر یا میانگین مقادیر Y در این طبقه را به‌کمک اطلاعات کمکیِ در دست و با رویکرد مدل‌یار براورد می‌کنند. در این حالت نیز تعیین آستانه‌ی برش مناسب و اندازه‌ی بهینه‌ی نمونه با توجه به دقت براورد مورد نظر از عواملی است که باید مورد توجه قرار گیرد. براوردهای به دست آمده در این نوع روش نمونه‌گیری برینشی، اریب‌اند. در این حالت برای هر $j \in U_E$ داریم $\pi_j = 0$، در حالی که برای هر $j \in U_S$ داریم $\pi_j > 0$.

۲٫۳ روش نمونه‌گیری برینشی نوع دوم (سرشماری، نمونه‌گیری)

یک روش جایگزین برای نمونه‌گیری برینشی نوع اول و حالت تبدیل‌یافته‌ی آن استفاده از روش نمونه‌گیری برینشی نوع دوم (سرشماری، نمونه‌گیری) است که در آن پس از تعیین آستانه‌ی برش مناسب، جامعه‌ی مورد بررسی به دو طبقه‌ی U_C و U_S برش داده می‌شود، که در آن U_C شامل واحدهایی از جامعه است که متناظر با مقادیر بزرگ و غیر قابل چشم‌پوشی Y هستند و به‌طور کامل و با احتمال ۱ در بررسی شرکت می‌کنند؛ یعنی برای هر $j \in U_C$ داریم $\pi_j = 1$، و به آن طبقه‌ی سرشماری گویند. در حالی که طبقه‌ی U_S شامل واحدهای باقی‌مانده‌ی جامعه است که از آن، نمونه‌ای تصادفی برای شرکت در آمارگیری انتخاب می‌شود. در این حالت برای هر $j \in U_S$ داریم $\pi_j > 0$ و به طبقه‌ی U_S طبقه‌ی نمونه‌گیری گویند. این نوع روش نمونه‌گیری برینشی از نوع روش‌های نمونه‌گیری احتمالی طبقه‌بندی‌شده است و براوردهای به دست آمده در این روش، نااریب هستند. برای توضیح بیش‌تر می‌توان به حیدیروگلو (۱۹۸۶) مراجعه کرد.

۲،۴ روش نمونه‌گیری برینشی نوع سوم (سرشماری، نمونه‌گیری، حذف)

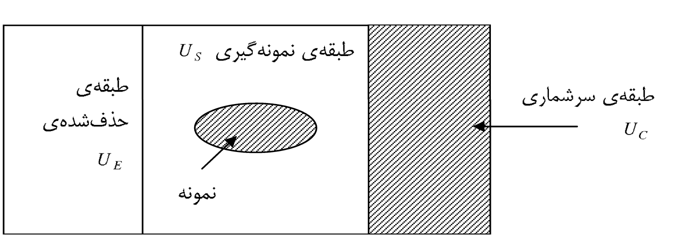
در روش نمونه‌گیری برینشی نوع دوم ممکن است در عمل، واحدهایی در طبقه‌ی U_S موجود باشند که نمونه‌گیری از آن‌ها منجر به صرف وقت و هزینه‌ی زیاد شود. بنا بر این گرچه براوردهای ارائه‌شده برای پارامترهای مورد نظر در این روش عموماً نااریب‌اند، در عمل کارایی کم‌تری دارند. یک روش جایگزین برای این روش که به‌ویژه در بررسی جوامعی با چولگی زیاد کاربرد دارد، استفاده از روش نمونه‌گیری برینشی نوع سوم است که در واقع تلفیقی از روش‌های نمونه‌گیری برینشی نوع اول و نوع دوم است.

در این روش پس از تعیین آستانه‌های برش مناسب، جامعه‌ی مورد بررسی به سه طبقه برش داده می‌شود که در آن، یک طبقه شامل واحدهایی از جامعه با مقادیر بزرگ و غیر قابل چشم‌پوشی Y است که سرشماری می‌شود و آن را با U_C نشان می‌دهیم. طبقه‌ی U_S شامل بخشی از جامعه با مقادیر متوسط Y است که از آن نمونه‌گیری به عمل می‌آید، و نهایتاً طبقه‌ی U_E که از آمارگیری حذف می‌شود شامل واحدهایی از جامعه با مقادیر کوچک و قابل چشم‌پوشی Y است. در واقع در این روش برای هر $j \in U_E$ داریم $\pi_j = 0$، برای هر $j \in U_S$ داریم $\pi_j > 0$، و برای هر $j \in U_C$ داریم $\pi_j = 1$. شکل ۱ ساختار آمارگیری از جامعه در این روش را نشان می‌دهد که در آن، قسمت‌های هاشور خورده نشان‌دهنده‌ی قسمت‌های شرکت داده‌شده در آمارگیری‌اند.

این روش، اولین بار توسط الیسون و الورز (۲۰۰۱) مورد استفاده قرار گرفت. بی و دیگران (۲۰۰۷) بخشی از مبانی نظری آن را بیان کرده، از آن در یک مثال کاربردی استفاده کرده‌اند.

۳ نمونه‌گیری برینشی نوع سوم

در این بخش، ضمن بررسی مبانی نظری نمونه‌گیری برینشی نوع سوم، نحوه‌ی تعیین آستانه‌های برش مناسب و اندازه‌ی نمونه‌ی لازم از طبقه‌ی U_S را نشان می‌دهیم. همچنین نحوه‌ی براورد کردن پارامتر مورد نظر در این نوع روش نمونه‌گیری برینشی را با رویکرد حذف و مدل‌یار مورد بررسی قرار می‌دهیم.



شکل ۱. ساختار آمارگیری از جامعه در نمونه‌گیری برینشی نوع سوم

جدول ۱. نمادگذاری در نمونه‌گیری برینشی نوع سوم

	کل جامعه $(\boldsymbol{U})$	طبقه‌ی سرشماری $(\boldsymbol{U_C})$	طبقه‌ی نمونه‌گیری $(\boldsymbol{U_S})$	طبقه‌ی حذف $(\boldsymbol{U_E})$
تعداد	N	N_C	N_S	N_E
(میانگین , مجموع)	$(Y_+, \bar{Y}_N)$	$(Y_{+C}, \bar{Y}_C)$	$(Y_{+S}, \bar{Y}_S)$	$(Y_{+E}, \bar{Y}_E)$
واریانس	$\sigma_N^۲$	$\sigma_C^۲$	$\sigma_S^۲$	$\sigma_E^۲$

۳٫۱ نمادگذاری

فرض کنید جامعه‌ی N-عضوی $U = \{۱, ۲, \ldots, N\}$ مورد نظر باشد و مقادیر صفت Y در جامعه‌ی U را با $Y_۱, \ldots, Y_N$ نشان دهیم. همچنین فرض کنید $Y_{(۱)} \leqslant Y_{(۲)} \leqslant \cdots \leqslant Y_{(N)}$ مقادیر مرتب‌شده‌ی صفت Y در جامعه‌ی U باشند. نمادگذاری ارائه‌شده در جدول ۱ را برای کل جامعه، طبقه‌ی سرشماری (C:=Census)، طبقه‌ی نمونه‌گیری (S:=Sampling) و طبقه‌ی حذف (E:=Eliminated) در نظر می‌گیریم.

در این نمادگذاری، $U = U_C \cup U_S \cup U_E$ و $N = N_C + N_S + N_E$ و می‌توان نشان داد $\bar{Y}_N$ یک ترکیب خطی محدب از میانگین طبقات سرشماری، نمونه‌گیری و حذف به‌صورت زیر است:

$$\bar{Y}_N = W_C \bar{Y}_C + W_S \bar{Y}_S + W_E \bar{Y}_E, \tag{۱}$$

که در آن $W_E = \frac{N_E}{N}$، $W_S = \frac{N_S}{N}$ و $W_C = \frac{N_C}{N}$. به‌راحتی می‌توان نشان داد:

$$\begin{aligned}\sum_{i=۱}^{N}(Y_i - \bar{Y}_N)^۲ &= \sum_{i\in U_C}(Y_i - \bar{Y}_C)^۲ + \sum_{i\in U_S}(Y_i - \bar{Y}_S)^۲ + \sum_{i\in U_S}(Y_i - \bar{Y}_E)^۲ \\ &\quad + N_C(\bar{Y}_C - \bar{Y}_N)^۲ + N_S(\bar{Y}_S - \bar{Y}_N)^۲ + N_E(\bar{Y}_E - \bar{Y}_N)^۲.\end{aligned}$$

با استفاده از تساوی $\sigma_N^۲ = \frac{۱}{N}\sum_{i=۱}^{N}(Y_i - \bar{Y}_N)^۲ = \frac{N-۱}{N}S_N^۲$ و کمّیت‌های مشابه آن داریم:

$$\begin{aligned}\sigma_N^۲ &= W_C\sigma_C^۲ + W_S\sigma_S^۲ + W_E\sigma_E^۲ + W_C(\bar{Y}_C - \bar{Y}_N)^۲ + W_S(\bar{Y}_S - \bar{Y}_N)^۲ \\ &\quad + W_E(\bar{Y}_E - \bar{Y}_N)^۲,\end{aligned} \tag{۲}$$

$$S_N^۲ = \frac{N_C - ۱}{N - ۱} S_C^۲ + \frac{N_S - ۱}{N - ۱} S_S^۲ + \frac{N_E - ۱}{N - ۱} S_E^۲ + \frac{N_C}{N - ۱} (\bar{Y}_C - \bar{Y}_N)^۲ + \frac{N_S}{N - ۱} (\bar{Y}_S - \bar{Y}_N)^۲ + \frac{N_E}{N - ۱} (\bar{Y}_E - \bar{Y}_N)^۲. \quad (۳)$$

در نمونه‌گیری برینشی نوع سوم، مقادیر به دست آمده از طریق سرشماری طبقه‌ی U_C در معرض تغییرات نمونه‌ای نیستند و تغییرات نمونه‌ای فقط به‌واسطه‌ی انتخاب نمونه‌ی n_Sتایی از طبقه‌ی نمونه‌گیری U_S به وجود می‌آید. همچنین مشاهده می‌شود که اندازه‌ی کل نمونه به‌صورت $n = n_S + N_C$ است. در ادامه با دو رویکرد مبتنی بر حذف و مدل‌یار، طریقه‌ی محاسبه‌ی براوردگر $\bar{Y}_N$ (و در نتیجه Y_+) و خواص آن‌ها را مورد مطالعه قرار می‌دهیم.

۳٫۲ براوردگر برینشی نوع سوم با رویکرد مبتنی بر حذف

در رویکرد مبتنی بر حذف، با فرض آن‌که Y_{+E} در مقایسه با مجموع صفت Y در جامعه (Y_+) بسیار ناچیز و قابل چشم‌پوشی است، براورد $\bar{Y}_N$ فقط به‌کمک مقادیر صفت Y در طبقه‌ی U_C و نمونه‌ی به دست آمده از طبقه‌ی U_S محاسبه می‌شود. پس از استخراج نمونه‌ی n_Sتایی از طبقه‌ی U_S و محاسبه‌ی $\bar{y}_S = \frac{۱}{n_S} \sum_{i=۱}^{n_S} y_i$ براوردگر برینشی نوع سوم با رویکرد حذف برای $\bar{Y}_N$ به‌صورت زیر به دست می‌آید:

$$\hat{\bar{Y}}_{\rm cut} = W_C \bar{Y}_C + W_S \bar{y}_S. \quad (۴)$$

همچنین براوردگر برینشی نوع سوم با رویکرد حذف برای Y_+ نیز عبارت است از $\hat{Y}_{+\rm cut} = N_C \bar{Y}_C + N_S \bar{y}_S$. از آن‌جا که در رویکرد حذف (و دیدگاه مبتنی بر طرح)، مقادیر Y در U_C در معرض هیچ‌گونه تغییرات نمونه‌ای نیستند و به‌طورکامل شمارش می‌شوند، داریم $\text{var}(\bar{Y}_C) = ۰$؛ یعنی $\text{var}(\hat{\bar{Y}}_{\rm cut})$ فقط به واریانس میانگین مقادیر نمونه‌ی تصادفی استخراج‌شده از طبقه‌ی U_S بستگی دارد و داریم:

$$\text{var}(\hat{\bar{Y}}_{\rm cut}) = W_C^۲ \text{var}(\bar{Y}_C) + W_C^۲ \text{var}(\bar{Y}_S) = W_S^۲ (۱ - f_S) \frac{S_S^۲}{n_S}, \quad (۵)$$

که در آن، $f_S = \frac{n_S}{N_S}$ است و یک براورد نااریب برای $\text{var}(\hat{\bar{Y}}_{\rm cut})$ با جایگزینی $S_S^۲$ توسط واریانس نمونه‌ی n_Sتاییِ استخراج‌شده از طبقه‌ی نمونه‌گیری به دست می‌آید. همچنین از آن‌جا که $E(\hat{\bar{Y}}_{\rm cut}) = W_C \bar{Y}_C + W_S \bar{Y}_S$، براوردگر برینشی نوع سوم ارائه‌شده اریب است و میزان اریبی آن عبارت است از $\text{bias}(\hat{\bar{Y}}_{\rm cut}) = W_C \bar{Y}_C + W_S \bar{Y}_S - \bar{Y}_N = -W_E \bar{Y}_E$، که اگر مقادیر Y همواره مثبت فرض شوند، اریبی $\hat{\bar{Y}}_{\rm cut}$ همواره مقداری منفی است. همچنین با فرض آن‌که $\bar{Y}_E$ در مقایسه با $\bar{Y}_N$ ناچیز باشد و W_E، یعنی سهم طبقه‌ی حذف‌شده از کل جامعه، کوچک باشد، می‌توان از مقدار اریبی

چشم‌پوشی کرد. همچنین اریبی نسبی $\hat{\bar{Y}}_{\text{cut}}$ در براورد $\bar{Y}_N$ عبارت است از

$$\frac{E(\hat{\bar{Y}}_{\text{cut}}) - \bar{Y}_N}{\bar{Y}_N} = \frac{-W_E \bar{Y}_E}{\bar{Y}_N} = -\left(۱ - \frac{W_S \bar{Y}_S + W_C \bar{Y}_C}{\bar{Y}_N}\right), \qquad (۶)$$

که با فرض $\bar{Y}_N \simeq W_S \bar{Y}_S + W_C \bar{Y}_C$ مقدار اریبی بسیار ناچیز است. واضح است که اندازه‌ی اریبی، $|\text{bias}(\hat{\bar{Y}}_{\text{cut}})|$، تابعی افزایشی از W_E می‌باشد و با افزایش تعداد واحدهای طبقه‌ی حذف‌شده، میزان اریبی نیز افزایش می‌یابد. با توجه به نتایجِ به دست آمده، میانگین توان دوم خطا (MSE) برای $\hat{\bar{Y}}_{\text{cut}}$ عبارت است از:

$$\text{MSE}(\hat{\bar{Y}}_{\text{cut}}) = W_S^۲ (۱ - f_S) \frac{S_S^۲}{n_S} + W_E^۲ \bar{Y}_E^۲. \qquad (۷)$$

۳٫۳ براوردگر برینشی نوع سوم با رویکرد مدل‌یار

در عمل، هرگاه اطلاعات سرشماری‌های گذشته یا مطالعات مشابه در دست باشد، می‌توان از رویکرد مدل‌یار در ارائه‌ی براوردگرهای دیگری از $\bar{Y}_N$ استفاده کرد، که استفاده‌ی موفقیت‌آمیز از آن می‌تواند حتی منجر به براوردگرهای با اریبی کم‌تر شود. بنا بر این در این بخش با رویکرد مدل‌یار، به‌جای حذف مقادیر Y متناظر با واحدهای جامعه در طبقه‌ی U_E، با استفاده از اطلاعات کمکیِ در دست و اِعمال فرضیات لازم (که در ادامه ذکر می‌شوند) مجموع مقادیر Y در طبقه‌ی U_E را براورد کرده، براورد Y_+ را به‌کمک آن ارائه می‌کنیم. برای این منظور، فرض کنید $\delta = \frac{Y_{+E}}{Y_{+C} + Y_{+S}}$؛ یعنی δ عبارت است از سهم Y_{+E} از Y_+ نسبت به سهم $Y_{+S} + Y_{+C}$ از Y_+. در عمل با استفاده از نظرهای کارشناسی و به‌کمک اطلاعات قبلی و تجربیات گذشته، مقدار δ به‌طور تقریبی معلوم است، که در این صورت از این مقدار پیشنهادی استفاده می‌کنیم و در غیر این صورت به‌کمک مقادیر یک متغیر کمکی X که دارای همبستگی قوی با Y است، δ به‌وسیله‌ی $\tilde{\delta} = \frac{X_{+E}}{X_{+C} + X_{+S}}$ براورد می‌شود. به‌راحتی می‌توان نشان داد $\bar{Y}_E = \frac{\delta}{W_E}(W_C \bar{Y}_C + W_S \bar{Y}_S)$ و در نتیجه $\bar{Y}_N = (۱ + \delta)(W_C \bar{Y}_C + W_S \bar{Y}_S)$. حال با فرض $\tilde{\delta} \simeq \delta$ و با براورد $\bar{Y}_S$ توسط $\bar{y}_S$، براوردگر برینشی نوع سوم مدل‌یار $\bar{Y}_N$ به‌صورت زیر به دست می‌آید:

$$\hat{\bar{Y}}_{\text{cut}} = (۱ + \tilde{\delta})(W_C \bar{Y}_C + W_S \bar{y}_S). \qquad (۸)$$

می‌توان نشان داد $E(\hat{\bar{Y}}_{\text{cut}}) = (۱ + \tilde{\delta})(W_C \bar{Y}_C + W_S \bar{Y}_S)$ و میزان اریبی این براوردگر عبارت است از $\text{bias}(\hat{\bar{Y}}_{\text{cut}}) = (\tilde{\delta} - \delta)(W_C \bar{Y}_C + W_S \bar{Y}_S)$. در این حالت، اندازه‌ی اریبی به مقدار $|\tilde{\delta} - \delta|$

بستگی دارد و هرچه $\tilde{\delta}$ براورد بهتری برای δ باشد، اریبی کم‌تر می‌شود. همچنین

$$\mathrm{var}(\hat{\hat{Y}}_{\mathrm{cut}}) = (1+\tilde{\delta})^2 W_S^2 (1-f_s)\frac{S_S^2}{n_S}, \qquad (9)$$

و یک براورد نااریب برای آن به‌صورت $\mathrm{var}(\hat{\hat{Y}}_{\mathrm{cut}}) = (1+\tilde{\delta})^2 W_S^2(1-f_S)\frac{S_s^2}{n_S}$ به دست می‌آید که در آن $S_s^2 = \frac{1}{n_S-1}\sum_{i=1}^{n_S}(y_i-\bar{y}_S)^2$. همچنین میانگین توان دوم خطا برای براوردگر مدل‌یار ارائه‌شده عبارت است از $\mathrm{MSE}(\hat{\hat{Y}}_{\mathrm{cut}}) = (1+\tilde{\delta})^2 W_S^2(1-f_S)\frac{S_S^2}{n_S} + (\tilde{\delta}-\delta)^2(W_C\bar{Y}_C + W_S\bar{Y}_S)^2$ و با مقایسه‌ی MSE مربوط به براوردگر برینشی نوع سوم مدل‌یار و MSE مربوط به براوردگر متناظر در رویکرد حذف، به‌وضوح دیده می‌شود که MSE در روش مدل‌یار بیش‌تر است (چرا که $\tilde{\delta} > 0$). از طرفی میزان اریبی این براوردگر، همان‌طور که قبلاً نیز اشاره شد، به مقدار $(\tilde{\delta}-\delta)$ بستگی دارد که می‌تواند از اریبی براوردگر برینشی با رویکرد حذف، بیش‌تر یا کم‌تر باشد. در نتیجه براورد موفقیت‌آمیز δ با $\tilde{\delta}$ منجر به ارائه‌ی براوردگرهای تقریباً نااریب برای $\bar{Y}_N$ می‌شود. باید توجه داشت که در این بخش، هیچ‌گونه فرضی مربوط به ناچیز بودن مقدار Y_{+E} نکرده‌ایم، که این مسئله از لحاظ علمی پذیرفتنی‌تر است. همچنین در عمل، بسیاری از آمارشناسان فرض $Y_{+E} \simeq 0$ را چندان نمی‌پسندند و ترجیح می‌دهند از فرض $\delta = \tilde{\delta}$ استفاده کنند.

گاهی با توجه به روند صعودی یا نزولی مقادیر Y در جامعه در دوره‌های زمانی متوالی و با در دست داشتن اطلاعات Y در دوره‌های زمانی t و $t-1$، که t دوره‌ی زمانی حاضر و $t-1$ دوره‌ی زمانی گذشته است، می‌توان فرض کرد برای مقدار معلوم k داریم:

$$\delta = \delta_t = \frac{\sum_{i\in U_E} Y_{i,t}}{\sum_{i\in U_S} Y_{i,t} + \sum_{i\in U_C} Y_{i,t}} = \frac{k\sum_{i\in U_E} Y_{i,t-1}}{\sum_{i\in U_S} Y_{i,t-1} + \sum_{i\in U_C} Y_{i,t-1}} = k\delta_{t-1}.$$

بنا بر این با فرض $X_i = Y_{i,t-1}$ واضح است که $\delta_{t-1} = \tilde{\delta}$، $\delta = \delta_t = k\tilde{\delta}$ و در نتیجه به‌راحتی می‌توان نشان داد $\hat{\hat{Y}}_{\mathrm{cut}} = (1+k\tilde{\delta})(W_C\bar{Y}_C + W_S\bar{y}_S)$ یک براورد منطقی برای $\bar{Y}_N$ با میزان اریبی $\mathrm{bias}(\hat{\hat{Y}}_{\mathrm{cut}}) = (1-k)\tilde{\delta}(W_C\bar{Y}_C + W_S\bar{Y}_S)$ است. همچنین از آن‌جا که $Y_+ = N\bar{Y}_N$ است، داریم $\hat{Y}_{+\mathrm{cut}} = N\hat{\hat{Y}}_{+\mathrm{cut}}$ و می‌توان تمامی نتایج بالا را برای براورد Y_+ در جامعه‌ی مورد بررسی تعمیم داد.

در ارائه‌ی براوردگرهای برینشی نوع سوم با رویکرد حذف و مدل‌یار می‌توان از براوردگرهای نسبی و رگرسیونی نیز استفاده کرد. اما همان‌طور که ناب (۲۰۰۷-آ) نیز اشاره می‌کند، مطالعاتِ به عمل آمده نشان می‌دهد استفاده از رگرسیون خطی با عرض از مبدأ، کمک زیادی برای بهبود نتایجِ به دست آمده نمی‌کند و در عمل، اغلب از براوردگر رگرسیونی گذرنده از مبدأ، و حالت خاص آن، یعنیِ براورد نسبتی استفاده

می‌شود. به همین منظور در ادامه و برای تأکید بیش‌تر، براورد نسبتی $\bar{Y}_N$ (و در نتیجه Y_+) را در رویکرد حذف توضیح می‌دهیم. نتایج در براوردیابی با رویکرد مدل‌یار به‌طور مشابه به دست می‌آیند.

۳٫۴ براورد نسبتی

فرض کنید در جامعه‌ی متناهی N-عضوی U، مقادیر متغیر کمکی X که دارای همبستگی زیاد با صفت مورد نظر Y است، به‌صورت $X_۱, \ldots, X_N$ موجود باشند. همچنین فرض کنید $R = \frac{\bar{Y}_N}{\bar{X}_N}$ نسبت میانگین متغیرهای X و Y در جامعه باشد. گاهی نسبت R به‌تنهایی مورد توجه آمارشناس است. برای مثال، اگر $\bar{Y}_N = \bar{Y}_{N,t}$ میانگین صفت Y در زمان t و $\bar{X}_N = \bar{Y}_{N,t-۱}$ میانگین صفت Y در زمان $t - ۱$ باشد، ممکن است میزان تغییرات صفت مورد بررسی در طول زمان مورد نظر باشد. از آن‌جا که در رویکرد حذف، فرض می‌شود $\bar{Y}_E$ (و در نتیجه $\bar{X}_E$) در مقایسه با $\bar{Y}_N$ ناچیز و قابل صرف نظر کردن است، داریم $R_{\text{cut}} = \frac{W_C\bar{Y}_C + W_S\bar{Y}_S}{W_C\bar{X}_C + W_S\bar{X}_S} \simeq \frac{\bar{Y}_N}{\bar{X}_N} = R$. بنا بر این پس از استخراج نمونه‌ی n_Sتایی از طبقه‌ی U_S می‌توان از

$$\hat{\bar{Y}}_{r,\text{cut}} = \frac{W_C\bar{Y}_C + W_S\bar{y}_S}{W_C\bar{X}_C + W_S\bar{x}_S}\bar{X}_N = r_{\text{cut}}\bar{X}_N$$

به‌عنوان براورد نسبتی برینشی نوع سوم $\bar{Y}_N$ استفاده کرد که در آن $r_{\text{cut}} = f(\bar{x}_S, \bar{y}_S) = \frac{W_C\bar{Y}_C + W_S\bar{y}_S}{W_C\bar{X}_C + W_S\bar{x}_S}$ براوردگر برینشی نوع سوم $R_{\text{cut}} \simeq R$ است و یک تابع غیر خطی از $\bar{x}_S$ و $\bar{y}_S$ می‌باشد. واضح است که $r_{\text{cut}} = R_{\text{cut}} + \left(\frac{W_C\bar{Y}_C + W_S\bar{y}_S}{W_C\bar{X}_C + W_S\bar{x}_S} - R_{\text{cut}}\right)$ و در نتیجه

$$\text{bias}(r_{\text{cut}}) = E\left(\frac{W_C\bar{Y}_C + W_S\bar{y}_S}{W_C\bar{X}_C + W_S\bar{x}_S} - R_{\text{cut}}\right).$$

با فرض آن‌که $R_{\text{cut}} \simeq R$، به نظر می‌رسد می‌توان نتیجه گرفت r_{cut} براوردگری اریب برای R است و این بدان معنی است که $\hat{\bar{Y}}_{r,\text{cut}}$ نیز برای $\bar{Y}_N$ اریب است. با استفاده از تقریب مرتبه‌ی اول (به‌کمک بسط سری تیلور) برای r_{cut} می‌توان نشان داد r_{cut} و در نتیجه $\hat{\bar{Y}}_{r,\text{cut}}$ براوردگرهایی تقریباً نااریب (از مرتبه‌ی اول) هستند. برای این منظور داریم:

$$\partial f_{\bar{X}_S} = \frac{\partial f(\bar{x}_S, \bar{y}_S)}{\partial \bar{x}_S}\bigg|_{(\bar{X}_S, \bar{Y}_S)} = -\frac{W_S(W_C\bar{Y}_C + W_S\bar{Y}_S)}{(W_C\bar{X}_C + W_S\bar{X}_S)^۲},$$
$$\partial f_{\bar{Y}_S} = \frac{\partial f(\bar{x}_S, \bar{y}_S)}{\partial \bar{y}_S}\bigg|_{(\bar{X}_S, \bar{Y}_S)} = -\frac{W_S}{(W_C\bar{X}_C + W_S\bar{X}_S)}.$$

همچنین تقریب مرتبه‌ی اول $f(\bar{x}_S, \bar{y}_S)$ در نقطه‌ی $(\bar{X}_S, \bar{Y}_S)$ عبارت است از

$$f(\bar{x}_S, \bar{y}_S) \simeq f(\bar{X}_S, \bar{Y}_S) + \partial f_{\bar{X}_S}(\bar{x}_S - \bar{X}_S) + \partial f_{\bar{Y}_S}(\bar{y}_S - \bar{Y}_S).$$

بنا بر این داریم:

$$r_{\text{cut}} = R_{\text{cut}} + \frac{W_S(\bar{y}_S - \bar{Y}_S)}{W_C\bar{X}_C + W_S\bar{X}_S} - \frac{W_S(W_C\bar{Y}_C + W_S\bar{Y}_S)(\bar{x}_S - \bar{X}_S)}{(W_C\bar{X}_C + W_S\bar{X}_S)^{۲}}.$$

با توجه به $E(\bar{y}_S) = \bar{Y}_S$ و $E(\bar{x}_S) = \bar{X}_S$، به‌راحتی نتیجه می‌شود $E(r_{\text{cut}}) \simeq R_{\text{cut}}$ و از آن‌جا که $R_{\text{cut}} \simeq R$، با تقریب مرتبه‌ی اول، r_{cut} برای R نااریب است و در نتیجه $\hat{\bar{Y}}_{r,\text{cut}}$ برای $\bar{Y}_N$ نااریب (از مرتبه‌ی اول) است. همچنین

$$\begin{aligned}
\text{var}(r_{\text{cut}}) &\simeq \frac{W_S^{۲}}{(W_C\bar{X}_C + W_S\bar{X}_S)^{۲}} \frac{(۱ - f_S)}{n_S} \\
&\quad \times \left\{ S_{YS}^{۲} + \left(\frac{W_C\bar{Y}_C + W_S\bar{Y}_S}{W_C\bar{X}_C + W_S\bar{X}_S} \right)^{۲} S_{XS}^{۲} - \frac{۲(W_C\bar{Y}_C + W_S\bar{Y}_S)}{W_C\bar{X}_C + W_S\bar{X}_S} S_{XYS} \right\} \\
&= \frac{W_S^{۲}}{(W_C\bar{X}_C + W_S\bar{X}_S)^{۲}} \frac{(۱ - f_S)}{n_S} \left\{ S_{YS}^{۲} + R_{\text{cut}}^{۲} S_{XS}^{۲} - ۲R_{\text{cut}} S_{XYS} \right\} \\
&= \frac{W_S^{۲}}{(W_C\bar{X}_C + W_S\bar{X}_S)^{۲}} \frac{(۱ - f_S)}{n_S} S_{DS}^{۲},
\end{aligned}$$

که در آن $S_{DS}^{۲} = S_{YS}^{۲} + R^{۲} S_{XS}^{۲} - ۲RS_{XYS}$ است. براورد $\text{var}(r_{\text{cut}})$ با جایگزینی $S_{YS}^{۲}$، $S_{XS}^{۲}$ و S_{XYS} با براوردهای آن‌ها به دست می‌آید. بنا بر این، واریانس $\hat{\bar{Y}}_{r,\text{cut}}$ عبارت است از

$$\text{var}(\hat{\bar{Y}}_{r,\text{cut}}) = \frac{N_S^{۲} X_+^{۲}}{(W_C\bar{X}_C + W_S\bar{X}_S)^{۲}} \frac{(۱ - f_S)}{n_S} S_{DS}^{۲},$$

که در آن فرض می‌شود X_+ معلوم است. می‌توان یک براورد تقریبی مرتبه‌ی اول برای $\text{var}(r_{\text{cut}})$ و در نتیجه برای $\text{var}(\hat{\bar{Y}}_{r,\text{cut}}) = \bar{X}_N^{۲} \text{var}(r_{\text{cut}})$ به دست آورد. برای این منظور، با استفاده از خطی‌سازی تیلور در نمونه‌ی n_Sتایی استخراج‌شده قرار می‌دهیم

$$\tilde{Z}_i = \frac{W_S}{W_C\bar{X}_C + W_S\bar{x}_S}(y_i) - \frac{W_S(W_C\bar{Y}_C + W_S\bar{y}_S)}{(W_C\bar{X}_C + W_S\bar{x}_S)^{۲}}(x_i), \qquad i = ۱, \ldots, n_S$$

در این صورت، یک براورد تقریبی واریانس برای r_{cut} عبارت است از

$$\text{var}(r_{\text{cut}}) = \frac{۱}{n_S(n_S - ۱)} \left(\frac{N_S - n_S}{N_S} \right) \sum_{i=۱}^{n_S} (\tilde{Z}_i - \bar{\tilde{Z}})^{۲},$$

که در آن $\bar{\tilde{Z}} = \frac{۱}{n_S}\sum_{i=۱}^{n_S} \tilde{Z}_i$. حال می‌توان نتایجِ به دست آمده را در دو قضیه‌ی زیر بیان کرد:

قضیه‌ی ۱ در نمونه‌گیری برینشی نوع سوم با رویکرد حذف، $\hat{\bar{Y}}_{r,\text{cut}} = \frac{W_C\bar{Y}_C + W_S\bar{y}_S}{W_C\bar{X}_C + W_S\bar{x}_S}\bar{X}_N$ یک براورد نسبتی تقریباً نااریب برای $\bar{Y}_N$ با واریانس تقریبی زیر است:

$$\text{var}(\hat{\bar{Y}}_{r,\text{cut}}) = \frac{W_S^۲\bar{X}_N^۲}{(W_C\bar{X}_C + W_S\bar{X}_S)^۲}\frac{(۱ - f_S)}{n_S}S_{DS}^۲$$

قضیه‌ی ۲ در نمونه‌گیری برینشی نوع سوم با رویکرد حذف، $r_{\text{cut}} = \frac{W_C\bar{Y}_C + W_S\bar{y}_S}{W_C\bar{X}_C + W_S\bar{x}_S}$ یک براورد تقریباً نااریب برای $R = \frac{\bar{Y}_N}{\bar{X}_N}$ با واریانس تقریبی زیر است:

$$\text{var}(r_{\text{cut}}) = \frac{W_S^۲}{(W_C\bar{X}_C + W_S\bar{X}_S)^۲}\frac{(۱ - f_S)}{n_S}S_{DS}^۲$$

۳٫۵ تعیین اندازه‌ی نمونه برای رسیدن به دقت مورد نظر

همان‌طور که اشاره شد، در طرح نمونه‌گیری برینشی نوع سوم، اندازه‌ی نمونه (n) عبارت است از $n = N_C + n_S$ که در آن، N_C اندازه‌ی طبقه‌ی سرشماری، و n_S اندازه‌ی نمونه‌ی اختیار شده از طبقه‌ی نمونه‌گیری است. از آن‌جا که تمام اعضای طبقه‌ی U_C سرشماری شده‌اند و مقدار آن‌ها در معرض هیچ‌گونه تغییرات نمونه‌ای نیست، برای تعیین دقت مورد نظر، نقش اصلی بر عهده‌ی n_S می‌باشد. براوردگرهای برینشی نوع سوم ارائه‌شده برای $\bar{Y}_N$ با رویکرد مبتنی بر حذف یا رویکرد مدل‌یار، همگی براوردگرهای اریب هستند. بنا بر این در تعیین اندازه‌ی نمونه باید به‌جای واریانس از میانگین توان دوم خطای براوردگرها استفاده کرد. فرض کنید آستانه‌های برش بهینه تعیین شده‌اند و مقادیر N_S و N_E و N_C معلوم هستند (هرچند سعی می‌شود براوردگرهایی با مفروضات کم‌تر و پذیرفتنی‌تر نیز ارائه شوند). در این بخش برای دقت مورد نظر، هدف، پیدا کردن اندازه‌ی نمونه‌ی لازم n_S از طبقه‌ی نمونه‌گیری به‌گونه‌ای است که $\text{MSE}(\hat{\bar{Y}}_{\text{cut}}) \leqslant v_۰$ باشد، که در آن، $v_۰$ به‌کمک روش‌های معلوم و با در نظر گرفتن عواملی مانند خطای نسبی، ضریب اطمینان، خطای مطلق و مانند آن تعیین می‌شود.

۳٫۵٫۱ تعیین اندازه‌ی نمونه در رویکرد مبتنی بر حذف

همان‌طور که نشان داده شد، $\text{MSE}(\hat{\bar{Y}}_{\text{cut}}) = W_S^۲(۱ - f_S)\frac{S_S^۲}{n_S} + W_E^۲\bar{Y}_E^۲$ و با استفاده از نامساوی $\text{MSE}(\hat{\bar{Y}}_{\text{cut}}) \leqslant v_۰$، کم‌ترین مقدار اندازه‌ی نمونه‌ی لازم برای دست‌یابی به دقت مورد نظر $v_۰$ در براورد

$\bar{Y}_N$، به‌صورت

$$n \simeq \frac{W_S^2 S_S^2}{\frac{W_S^2}{N_S}(S_S^2) + (v_0 - W_E^2 \bar{Y}_E^2)} + N_C$$

به دست می‌آید. در صورتی که مقدار $\bar{Y}_E$ قابل چشم‌پوشی باشد می‌توان در رابطه‌ی فوق از $\bar{Y}_E \simeq 0$ استفاده کرد. همچنین اگر اطلاعات کمکی قابل اطمینانی از قبیل اطلاعات سرشماری‌های گذشته یا متغیر کمکی X مرتبط با صفت مورد نظر Y در دسترس باشد، می‌توان با جایگذاری $\bar{Y}_E$ توسط $\bar{X}_E$، و S_S^2 توسط S_{XS}^2 اندازه‌ی بهینه‌ی نمونه را به‌طور تقریبی به‌صورت زیر به دست آورد:

$$n \simeq N_C + \frac{W_S^2 S_{XS}^2}{\frac{W_S^2}{N_S}(S_{XS}^2) + (v_0 - W_E^2 \bar{X}_E^2)}.$$

از آن‌جا که در عمل ممکن است تعیین سهم طبقه‌ی سرشماری و طبقه‌ی حذف و طبقه‌ی نمونه‌گیری از کل جامعه، راحت‌تر از تعیین اندازه‌ی آن طبقات باشد و در بسیاری موارد، اندازه‌ی جامعه (N) معلوم است، می‌توان رابطه‌ی به دست آمده برای n را به‌صورت

$$n = N\left(W_C + \frac{W_S^2 S_S^2}{W_S S_S^2 + N(v_0 - W_E^2 \bar{Y}_E^2)}\right)$$

بازنویسی کرد. همچنین در صورت استفاده از براورد نسبتی، با استفاده از مطالب بخش قبل و به‌کارگیری تقریب مرتبه‌ی اول، کم‌ترین اندازه‌ی نمونه‌ی لازم برای ارائه‌ی براوردگر $\hat{\bar{Y}}_{r,\mathrm{cut}}$ با دقت مورد نظر v_0 عبارت است از

$$n = N_C + \frac{W_S^2 S_{DS}^2}{\frac{W_S^2}{N_S}(S_{DS}^2) + \left\{(W_C \bar{X}_C + W_S \bar{X}_S)^2 v_0 - W_E^2 \bar{Y}_E^2\right\}},$$

که مشابه قبل، در صورت ناچیز بودن مقدار $\bar{Y}_E$ می‌توان از آن چشم‌پوشی کرد. در محاسبه‌ی S_{DS}^2 می‌توان با استفاده از یک نمونه‌ی مقدماتی و به‌کمک اطلاعات کمکی X، آن را براورد نمود. شایان ذکر است، شکل بهتر اندازه‌ی نمونه که فقط به سهم طبقات U_C و U_S و U_E از کل جمعیت (و نه اندازه‌ی طبقات) بستگی دارد، مشابه قبل به‌صورت زیر به دست می‌آید:

$$n = N\left(W_C + \frac{W_S^2 S_{DS}^2}{W_S S_{DS}^2 + N\{(W_C \bar{X}_C + W_S \bar{X}_S)^2 v_0 - W_E^2 \bar{Y}_E^2\}}\right).$$

۳٫۵٫۲ تعیین اندازه‌ی نمونه در رویکرد مدل‌یار

در این بخش فقط حالتی را بررسی می‌کنیم که در براورد $\bar{Y}_N$ از اطلاعات کمکی به‌وسیله‌ی $\tilde{\delta}$ که در بخش ۳٫۳ معرفی شد استفاده می‌شود؛ گرچه با انجام عملیات مشابه بخش قبل و اِعمال تغییرات لازم می‌توان اندازه‌ی نمونه‌ی لازم را در حالتی که براورد نسبتی در رویکرد مدل‌یار مورد استفاده قرار گیرد نیز به دست آورد. با توجه به نتایج قبلی، میانگین توان دوم خطای $\hat{\hat{Y}}_{\text{cut}}$ در رویکرد مدل‌یار به‌صورت زیر به دست می‌آید:

$$\text{MSE}(\hat{\hat{Y}}_{\text{cut}}) = (۱+\tilde{\delta})^۲ W_S^۲ (۱-f_S)\frac{S_S^۲}{n_S} + (\tilde{\delta}-\delta)^۲ (W_C\bar{Y}_C + W_S\bar{Y}_S)^۲.$$

اگر بخواهیم براوردگر برینشی ارائه‌شده دارای دقت مورد نظر $v_۰$ باشد، یعنی $\text{MSE}(\hat{\hat{Y}}_{\text{cut}}) \leqslant v_۰$، کم‌ترین اندازه‌ی نمونه‌ی لازم به‌صورت زیر به دست می‌آید:

$$n = N_C + \frac{W_S^۲ S_S^۲}{\frac{W_S^۲}{N_S}(S_S^۲) + \psi_{v_۰}} = N\left(W_C + \frac{W_S^۲ S_S^۲}{W_S S_S^۲ + N\psi_{v_۰}}\right),$$

که در آن،

$$\psi_{v_۰} = \frac{v_۰ - (\tilde{\delta}-\delta)^۲ (W_C\bar{Y}_C + W_S\bar{Y}_S)^۲}{(۱+\tilde{\delta})^۲}.$$

شایان ذکر است، با انجام تغییرات لازم می‌توان فرمول‌های مشابهی برای حالتی که هدف، براورد کردن Y_+ است به دست آورد. برای این منظور، از آن‌جا که $Y_+ = N\bar{Y}$، کافی است در همه‌ی روابطِ به دست آمده $v_۰$ را با $\frac{v_۰}{N^۲}$ جایگزین کرد. همچنین توجه کنید که در صورت وجود رابطه‌ی $\delta = K\tilde{\delta}$ نیز می‌توان فرمول‌های مشابهی به دست آورد، که از تکرار آن‌ها صرف نظر می‌کنیم.

۳٫۶ تعیین آستانه‌ی برش

در این بخش با تعمیم نتایجِ به دست آمده توسط حیدیروگلو (۱۹۸۶) روش‌هایی برای تعیین آستانه‌های برش در نمونه‌گیری برینشی نوعِ سوم ارائه می‌کنیم. برای این منظور، نخست با تعیین آستانه‌ی برش مناسب، طبقه‌ی حذف را تشکیل می‌دهیم. سپس آستانه‌ی برش برای تعیین طبقات سرشماری و نمونه‌گیری را در برای حالتی که اندازه‌ی نمونه از قبل معلوم و ثابت باشد یا دقت $v_۰$ برای براوردها مد نظر باشد، با در نظر گرفتن رویکرد مبتنی بر حذف و رویکرد مدل‌یار بیان می‌کنیم. تمامی مراحل را با معلوم در نظرگرفتن مقادیر صفت Y در جامعه انجام می‌دهیم؛ هرچند در عمل می‌توان از اطلاعات سرشماری‌های گذشته، متغیرهای

کمکی و سایر منابعِ در دست نیز به‌جای مقادیر Y استفاده کرد. پس از مرتب کردن مقادیر Y در جامعه به‌صورت $Y_{(N)} \leqslant \cdots \leqslant Y_{(1)}$ و تشکیل $S_y(l) = \sum_{j=1}^{l} Y_{(j)}$، نخست جامعه را به دو طبقه‌ی U_E و $U_{CS} = U_C \cup U_S$ تقسیم می‌کنیم که در آن، طبقه‌ی U_{CS} شامل $(1-\varepsilon)100$ درصد از اعضای جامعه با مقادیر بزرگ و متوسط Y است و تمام واحدهای جامعه را که برای آن‌ها $S_y(l) \leqslant \varepsilon S_y(N)$ است، در طبقه‌ی U_E قرار می‌دهیم. بنا بر این، واحدهای متناظر با l مقدار کوچک Y در جامعه، از بررسی حذف می‌شوند و در طبقه U_E قرار می‌گیرند هرگاه سهم آن‌ها از مجموع کل مقادیر، یعنی $S_y(N)$، کم‌تر از ε یا مساوی با آن باشد. در عمل، مقدار ε از تجربیات گذشته و با در نظرگرفتن نظرهای کارشناسی تعیین می‌شود، که بیش‌تر مواقع $\varepsilon \in \{0/05, 0/1, 0/15, 0/2\}$ مورد استفاده قرار می‌گیرد.

همان‌طورکه ملاحظه می‌شود، در این حالت، $N_E = l$، و آستانه‌ی برش Y_E^* عبارت است از هر مقدار در فاصله‌ی $\{Y_{(l)}, Y_{(l+1)}\}$. در نتیجه با فرض $Y_{(i)} \neq Y_{(j)}$ برای هر $i \neq j$، طبقه‌ی حذف، شامل فقط l عضوِ جامعه با مقادیر کوچک‌تر از Y_E^* است. اما از آن‌جا که اِعمال فرض $Y_{(i)} \neq Y_{(j)}$ برای هر $i \neq j$ عملاً امکان‌پذیر نیست، ممکن است تعدادی از $Y_{(i)}$ها برابر با Y_E^* باشند که در نتیجه سهم طبقه‌ی U_E از کل جامعه، بیش از مقدارِ از پیش تعیین‌شده‌ی ε می‌شود. در این حالت با یک رویکرد تصادفی می‌توان تعدادی از واحدهای دارای $Y_{(i)}$های برابر با Y_E^* را در طبقه‌ی U_E و بقیه را در طبقه‌ی U_{CS} قرار داد به‌طوری که سهم طبقه‌ی U_E از کل جامعه، برابر با مقدارِ از پیش تعیین‌شده‌ی ε باشد، که البته این کار از کارایی براوردگرهای به دست آمده می‌کاهد. بنا بر این، پس از تعیین آستانه‌ی برش Y_E^*، جامعه به‌صورت $U = U_{CS} \cup U_E$ تقسیم می‌شود که در آن $U_{CS} = U_C \cup U_S$ و $N_{CS} = N_C + N_S = N - l$ است. در ادامه پس از تعیین $N_E = l$ (به‌کمک آستانه‌ی برش Y_E^*)، طبقه‌ی U_{CS} را به دو طبقه‌ی U_S و U_C تقسیم می‌کنیم. بنا بر این، هدف، پیدا کردن آستانه‌ی برش Y_S^* به‌گونه‌ای است که U_{CS} به طبقات U_S و U_C تقسیم شود و براوردگر ارائه‌شده دارای خواص بهینگی باشد. برای این منظور، حالت‌های زیر را در نظر می‌گیریم.

۳٫۶٫۱ تعیین Y_S^* و تقریب آن برای اندازه‌ی نمونه‌ی ثابت با رویکرد حذف

در این بخش فرض کنید اندازه‌ی طبقه‌ی سرشماری، $N_C = m$، و اندازه‌ی نمونه از قبل معلوم و برابر با n باشد. واضح است که $n = m + n_S$ که در آن، U_C شامل m واحد با بزرگ‌ترین مقادیر Y از جامعه‌ی U_{CS} و طبقه‌ی نمونه‌گیری U_S شامل $N_S = N_{CS} - m = N - l - m$ واحد باقی‌مانده‌ی جامعه‌ی U_{CS} است که از آن نمونه‌ای به اندازه‌ی $n_S = n - m$تایی به‌روش تصادفی ساده اختیار می‌شود. اگر آستانه‌ی برش را با Y_S^* نشان دهیم، داریم $Y_S^* \in [Y_{(N-m)}, Y_{(N-m+1)})$ و در نتیجه طبقه‌ی سرشماری

شامل m عضو با مقادیر Y به‌صورت $\{Y_{(N-m+1)},\dots,Y_{(N)}\}$ است؛ حال آن‌که طبقه‌ی نمونه‌گیری عبارت است از $U_S=\{Y_{(l+1)},\dots,Y_{(N-m)}\}$. بنا بر این، هدف، پیدا کردن مقدار m (اندازه‌ی طبقه‌ی سرشماری) و در نتیجه آستانه‌ی برش Y_S^* به‌گونه‌ای است که

$$\begin{aligned}\mathrm{MSE}(\hat{\bar{Y}}_{\mathrm{cut}}(m)) &\leqslant \mathrm{MSE}(\hat{\bar{Y}}_{\mathrm{cut}}(m+1)),\\ \mathrm{MSE}(\hat{\bar{Y}}_{\mathrm{cut}}(m)) &\leqslant \mathrm{MSE}(\hat{\bar{Y}}_{\mathrm{cut}}(m-1)).\end{aligned}\tag{10}$$

با توجه به روابطِ به دست آمده برای $\mathrm{MSE}(\hat{\bar{Y}}_{\mathrm{cut}})$ داریم:

$$\begin{aligned}\mathrm{MSE}(\hat{\bar{Y}}_{\mathrm{cut}}(m+j)) = {}&\left(\frac{N-m-l-j}{N}\right)^2\left(1-\frac{n-m-j}{N-m-l-j}\right)\\ &\times\frac{S_S^2(m+j)}{n-m-j}+\left(\frac{l}{N}\right)^2\bar{Y}_E^2(l),\end{aligned}\tag{11}$$

که در آن $j\in\{-1,0,1\}$. با فرض آن‌که $\bar{Y}_E(l)$ در مقایسه با $\bar{Y}_N$ قابل چشم‌پوشی باشد، شرط‌های (۱۰) معادل شرط‌های زیرند:

$$\begin{aligned}\left(\frac{N-m-l}{N}\right)^2\left(1-\frac{n-m}{N-m-l}\right)\frac{S_S^2(m)}{n-m} &\leqslant \left(\frac{N-m-l-1}{N}\right)^2\\ &\quad\times\left(1-\frac{n-m-1}{N-m-l-1}\right)\frac{S_S^2(m+1)}{n-m-1},\\ \left(\frac{N-m-l}{N}\right)^2\left(1-\frac{n-m}{N-m-l}\right)\frac{S_S^2(m)}{n-m} &\leqslant \left(\frac{N-m-l+1}{N}\right)^2\\ &\quad\times\left(1-\frac{n-m+1}{N-m-l+1}\right)\frac{S_S^2(m-1)}{n-m+1},\end{aligned}\tag{12}$$

که در آن‌ها $S_S^2(m+j)$ واریانس مقادیر صفت Y در طبقه‌ی نمونه‌گیری با $n-m-j$ عضو است و $S_S^2(m+j)=\frac{N-l-m-j}{N-l-m-j-1}\sigma_S^2(m+j)$. همچنین فرض کنید $\bar{Y}_S(m+j)$ میانگین مقادیر صفت Y در طبقه‌ی نمونه‌گیری با $n-m-j$ عضو باشد. با انجام دادن محاسبات لازم، به‌راحتی می‌توان نشان داد که شرط لازم برای برقراری روابط (۱۰) یا به‌طور معادل، روابط (۱۲) عبارت است از

$$(Y_S^*-\bar{Y}_S(m))^2=\frac{N-l-m}{n-m}\sigma_S^2(m)\tag{13}$$

و از آن‌جا که فقط واحدهایی از جامعه با مقادیر بزرگ Y سرشماری می‌شوند، داریم:

$$Y_S^*=\bar{Y}_S(m)+\sqrt{\frac{N-m-l}{N-m}}\sigma_S(m).$$

بنا بر این، واحدهایی از جامعه‌ی U_{CS} با مقادیر $Y > Y_S^*$، در طبقه‌ی سرشماری و بقیه‌ی در طبقه‌ی U_S قرار می‌گیرند.

توجه کنید که رابطه‌ی به دست آمده برای Y_S^* به مقدار میانگین و واریانس طبقه‌ی U_S بستگی دارد که عملاً قبل از تعیین آستانه‌ی برش معلوم نیست. در این حالت می‌توان یک کران تقریبی بالا به‌صورت $Y_S^* < \bar{Y}_{CS} + \sqrt{\frac{N-l}{n}}\sigma_{CS}$ برای آستانه‌ی بهینه‌ی برش به دست آورد که در آن $\bar{Y}_{CS}$ و $\sigma_{CS}^۲$ به‌ترتیب، میانگین و واریانس واحدهای جامعه در طبقه‌ی U_{CS} می‌باشند و $S_{CS}^۲ = \frac{N-l}{N-l-۱}\sigma_{CS}^۲$. در عمل می‌توان پس از تعیین ε و تقسیم جامعه به دو طبقه‌ی U_{CS} و U_E، با استفاده از اطلاعات موجود، از قبیل سرشماری‌ها، مقدار $\bar{Y}_{CS}$ و σ_{CS} را تقریب زد و با جایگذاری آن‌ها در رابطه‌ی بالا آستانه‌ی تقریبی برش را محاسبه کرد. در ادامه حالتی را بررسی می‌کنیم که در آن Y_S^* را برای اندازه‌ی نمونه‌ی ثابت در رویکرد مدل‌یار به دست می‌آوریم.

۳٫۶٫۲ تعیین Y_S^* و تقریب آن برای اندازه‌ی نمونه‌ی ثابت با رویکرد مدل‌یار

در این حالت به‌جای حذف مقادیر Y متناظر با واحدهای جامعه در طبقه‌ی U_E با استفاده از اطلاعات کمکی موجود، مجموع مقادیر Y در طبقه‌ی U_E را براورد کرده، براورد Y_+ یا $\bar{Y}_N$ را به‌کمک آن ارائه می‌کنیم. در این رویکرد، $\hat{\bar{Y}}_{\text{cut}} = (۱ + \tilde{\delta})(W_C\bar{Y}_C + W_S\bar{y}_S)$ و داریم

$$\text{MSE}(\hat{\bar{Y}}_{\text{cut}}) = (۱ + \tilde{\delta})^۲ W_S^۲(۱ - f_S)\frac{S_S^۲}{n_S} + (\tilde{\delta} - \delta)^۲(\bar{Y}_N - W_E\bar{Y}_E)^۲,$$

که پس از تعیین آستانه‌ی برش Y_S^* و معلوم شدن مقدار $N_E = l$، مقادیر $W_E = \frac{l}{N}$ و $\bar{Y}_E$ مشخص می‌شوند و در نتیجه صرف نظر از تعداد اعضای طبقه‌ی سرشماری (m) (و در نتیجه اندازه‌ی نمونه‌ی طبقه‌بندی نمونه‌گیری، $n_S = n - m$)، مقدار $(\tilde{\delta} - \delta)^۲(\bar{Y}_N - W_E\bar{Y}_E)^۲$ ثابت است. از آن‌جا که در این بخش نیز هدف، پیدا کردن مقدار m و در نتیجه آستانه‌ی برش Y_S^* به‌گونه‌ای است که شرط‌های (۱۰) برقرار باشند، به‌راحتی نتیجه می‌شود که شرط‌های (۱۰) در صورت استفاده از رویکرد مدل‌یار به‌صورت زیر است:

$$\begin{aligned}
&(۱ + \tilde{\delta})^۲\left(\frac{N-m-l}{N}\right)^۲\left(۱ - \frac{n-m}{N-m-l}\right)\frac{S_S^۲(m)}{n-m} + (\tilde{\delta} - \delta)^۲(\bar{Y}_N - W_E\bar{Y}_E)^۲ \\
&\leqslant (۱ + \tilde{\delta})^۲\left(\frac{N-m-l-j}{N}\right)\left(۱ - \frac{n-m-j}{N-m-l-j}\right)\frac{S_S^۲(m+j)}{n-m-j} \\
&\quad + (\tilde{\delta} - \delta)^۲(\bar{Y}_N - W_E\bar{Y}_E)^۲,
\end{aligned}$$

که در آن $j \in \{-۱, ۱\}$. مطابق بخش قبل می‌توان نشان داد شرط لازم برای برقراری رابطه‌ی بالا عبارت است از

$$(Y_S^* - \bar{Y}_S(m))^۲ = \frac{N-l-m}{n-m}\sigma_S^۲(m),$$

یعنی $Y_S^* = \bar{Y}_S(m) + \sqrt{\frac{N-l-m}{n-m}}\sigma_S(m)$. همچنین یک کران بالا برای آستانه‌ی برش Y_S^* به‌صورت $Y_S^* < \bar{Y}_{CS} + \sqrt{\frac{N-l}{m}}\sigma_{CS}$ به دست می‌آید. در ادامه حالتی را بررسی می‌کنیم که در آن، هدف، تعیین آستانه‌ی برش به‌گونه‌ای است که براوردهای به دست آمده دارای دقتِ داده‌شده باشند.

۳٫۶٫۳ تعیین Y_S^* و تقریب آن برای دقتِ داده‌شده، با رویکرد مبتنی بر حذف

برای دقتِ داده‌شده‌ی متناظر با $v_۰$، کم‌ترین اندازه‌ی نمونه‌ی لازم برای برآورد کردن $\bar{Y}_N$ توسط $\hat{\bar{Y}}_{\text{cut}}$ عبارت است از

$$n = m + \frac{W_S^۲ S_S^۲}{\frac{W_S^۲}{N_S}S_S^۲ + v_۰ - W_E^۲\bar{Y}_E^۲},$$

که با صرف نظر کردن از مقدار $\bar{Y}_E$ که فرض می‌شود در مقایسه با $\bar{Y}_N$ بسیار ناچیز است، داریم:

$$\begin{aligned} n = n(m) &= m + \frac{(N-l-m)^۲ S_S^۲(m)}{(N-l-m)S_S^۲(m) + N^۲ v_۰} \\ &= (N-l) - \frac{N^۲(N-l-m)v_۰}{N^۲ v_۰ + (N-l-m)S_S^۲(m)}. \end{aligned}$$

همان‌طور که مشاهده می‌شود، مقدار بهینه‌ی $n(m)$ تابعی از m است. در ادامه، مقدار بهینه برای m و در نتیجه آستانه‌ی برش Y_S^* را به‌گونه‌ای به دست می‌آوریم که $n(m)$ نسبت به m کمینه باشد. مشابه حیدیروگلو (۱۹۸۶) برای مقادیر داده‌شده‌ی $v_۰$ و N و l می‌توان نشان داد $n(m)$ به‌عنوان تابعی از m دارای یک مینیمم منحصر به فرد است. واضح است که مقدار بهینه برای تعداد اعضای طبقه‌ی سرشماری، در شرایط $n(m) \leqslant n(m+۱)$ و $n(m) \leqslant n(m-۱)$ صدق می‌کند و یک شرط لازم برای برقراری آن و پیدا کردن m بهینه به‌صورت $(Y_S^* - \bar{Y}_S(m))^۲ = \frac{N-l-m-۱}{(N-l-m)^۲}N^۲ v_۰ + S_S^۲(m)$ به دست می‌آید، یا به عبارت دیگر $Y_S^* = \bar{Y}_S(m) + \sqrt{\frac{N-l-m-۱}{(N-l-m)^۲}N^۲ v_۰ + S_S^۲(m)}$. همچنین یک کران بالا برای آستانه‌ی برش فوق به‌صورت $Y_S^* \leqslant \bar{Y}_{CS} + \sqrt{\frac{N^۲}{N-l}v_۰ + S_{CS}^۲}$ ارائه می‌شود و همان‌طور که ملاحظه می‌شود، کران بالای ارائه‌شده برای Y_S^*، وابسته به m نیست و توصیه می‌شود در عمل با انجام دادن تغییرات لازم (با نظر کارشناس) از آن استفاده شود.

۳٫۶٫۴ تعیین Y_S^* و تقریب آن برای دقتِ داده‌شده با رویکرد مدل‌یار

در این حالت نیز اندازه‌ی بهینه‌ی نمونه برای آن‌که $\mathrm{MSE}(\hat{\bar{Y}}_{\mathrm{cut}})$ کمینه باشد عبارت است از

$$\begin{aligned} n=n(m) &= m+\frac{(N-l-m)^{۲}S_S^{۲}(m)}{(N-l-m)S_S^{۲}(m)+N^{۲}\Psi_{v_۰}} \\ &= (N-l)-\frac{(N-l-m)^{۲}N^{۲}\Psi_{v_۰}}{N^{۲}\Psi_{v_۰}+(N-l-m)S_S^{۲}(m)}, \end{aligned}$$

که در آن $\Psi_{v_۰}=\frac{v_۰+(\tilde{\delta}-\delta)^{۲}(W_C\bar{Y}_C+W_S\bar{Y}_S)^{۲}}{(۱+\tilde{\delta})^{۲}}$. از آن‌جا که $\bar{Y}_N=W_C\bar{Y}_C+W_S\bar{Y}_S+W_E\bar{Y}_E$، داریم $W_C\bar{Y}_C+W_S\bar{Y}_S=\bar{Y}_N-W_E\bar{Y}_E$ و در نتیجه $\Psi_{v_۰}=\frac{v_۰+(\tilde{\delta}-\delta)^{۲}(\bar{Y}_N-W_E\bar{Y}_E)^{۲}}{(۱+\tilde{\delta})^{۲}}$، که می‌توان برای استفاده از آن مقدار، $W_E\bar{Y}_E$ را پس از تعیین آستانه‌ی برش Y_E^* به دست آورد. در نتیجه مقدار بهینه‌ی m برای تعداد اعضای طبقه‌ی سرشماری، در شرایط $n(m)\leqslant n(m+۱)$ و $n(m)\leqslant n(m-۱)$ صدق می‌کند که در نتیجه، یک شرط لازم برای برقراری آن و پیدا کردن مقدار بهینه‌ی m عبارت است از

$$(Y_S^*-\bar{Y}_S(m))^{۲}=\frac{N-l-m-۱}{(N-l-m)^{۲}}(N^{۲}\Psi_{v_۰})+S_S^{۲}(m)$$

یا به عبارت دیگر $Y_S^*=\bar{Y}_S(m)+\sqrt{\frac{N-l-m-۱}{(N-l-m)^{۲}}N^{۲}\Psi_{v_۰}+S_S^{۲}(m)}$. همچنین یک کران بالا برای آستانه‌ی برش فوق عبارت است از

$$Y_S^*\leqslant\bar{Y}_{CS}+\sqrt{\frac{N^{۲}}{N-l}\left(\frac{v_۰+(\tilde{\delta}-\delta)^{۲}(\bar{Y}_N-W_E\bar{Y}_E)^{۲}}{(۱+\tilde{\delta})^{۲}}\right)+S_{CS}^{۲}}.$$

۳٫۷ مطالعه‌ی شبیه‌سازی‌شده

در این بخش با استفاده از روش‌های شبیه‌سازی، نخست جامعه‌ای شامل متغیر اصلی Y و متغیر کمکی X به اندازه‌ی $N=۱۰۰۰$ تولید کرده، سپس با ۱۰۰۰ بار تکرار نمونه‌گیری‌های مورد نیاز و انجام دادن محاسبات لازم، مطالب نظری ارائه‌شده در بخش‌های قبل را روی جامعه‌ی تولید شده اِعمال کرده، با نتایج متناظر در روش نمونه‌گیری برینشی نوع اول که کاربرد آن متداول‌تر است، مورد مقایسه قرار می‌دهیم. جدول ۲ اطلاعات خلاصه‌شده‌ی مربوط به دو صفت Y و X در جامعه‌ی شبیه‌سازی‌شده‌ی مورد نظر را نشان می‌دهد که به‌کمک توزیع چوله‌به‌راست لُگ‌نرمال و با استفاده از الگوهای استاندارد موجود در نرم‌افزار

جدول ۲. اطلاعات خلاصه‌شده‌ی دو صفت Y و X در جامعه‌ی شبیه‌سازی‌شده

متغیر	مینیمم	چارک اول	میانگین	میانه	چارک سوم	ماکسیمم	انحراف معیار	چولگی	جمع
Y	۰٫۱۲۶	۰٫۶۲۴	۱٫۳۲۷	۱٫۰۱۵	۱٫۶۲۴	۱۵٫۲۴	۱٫۱۵۶	۳٫۶۷	۱۳۲۷٫۵۷۷
X	۰	۱٫۹۶۳	۲٫۸۱۴	۲٫۶۶۱	۳٫۳۹۶	۱۷٫۵۸	۱٫۴۱	۲٫۲۶۱	۲۸۴۱٫۱۲۳

جدول ۳. نتایج در صورت استفاده از مقادیر Y برای تعیین طبقات در نمونه‌گیری برینشی نوع اول

روش	N_E	N_S	n_S	$\hat{Y}_+$	bias	a.r.b.	MSE
رویکرد حذف	۲۸۹	۷۱۱	۱۳۵	۱۷۱۴٫۴۷	۳۸۶٫۹	۰٫۲۹	۴۵۱۵٫۵۳
رویکرد مدل‌یار	۲۸۹	۷۱۱	۱۷۵	۱۴۶۵٫۱۸	۱۳۷٫۶	۰٫۳۱	۳۵۰۲٫۷۸

جدول ۴. نتایج در صورت استفاده از مقادیر X برای تعیین طبقات در نمونه‌گیری برینشی نوع اول

روش	N_E	N_S	n_S	$\hat{Y}_+$	bias	a.r.b.	MSE
رویکرد حذف	۲۱۲	۷۸۸	۴۱	۱۶۰۳٫۵۹	۲۷۶٫۰۱	۰٫۲۱	۳۷۵۱۰٫۱۹
رویکرد مدل‌یار	۲۱۲	۷۸۸	۵۰	۱۴۹۲٫۴۴	۱۶۴٫۸۷	۰٫۲۸	۳۰۹۰۸٫۵۰

(a.r.b. = absolute relative bias)

S-PLUS 8 برای تولید داده به دست آمده است.

همان‌طورکه در جدول ۲ مشاهده می‌شود، توزیع دو صفت Y و X در جامعه دارای چولگی مثبت است و محاسبات لازم نشان می‌دهد ضریب همبستگی بین Y و X برابر با ۰٫۸۲ است. بنا بر این می‌توان از روش‌های ارائه‌شده در بخش‌های قبل برای براورد مقدار کل صفت Y در جامعه، یعنی $Y_+ =$ ۱۳۲۷٫۵۷۷ استفاده کرد و امیدوار بود اطلاعات کمکی X در این مورد سودمند باشد.

در ادامه، دو حالت زیر را مورد مطالعه قرار می‌دهیم:

آ) حالتی که در آن، تعیین آستانه‌های برش و تشکیل طبقات مورد استفاده، به‌کمک مقادیر Y انجام می‌شود؛

ب) حالتی که در آن، موارد اشاره‌شده، به‌کمک متغیر کمکی X انجام می‌گیرد.

همچنین در هر مورد، براوردیابی با رویکرد حذف و مدل‌یار را مورد مطالعه قرار می‌دهیم. از آن‌جا که اندازه‌ی نمونه از قبل مشخص نیست، در این بخش فرض می‌کنیم هدف، رسیدن به دقت مورد نظر $v_\circ = (cY_+)^۲$

جدول ۵. نتایج در صورت استفاده از مقادیر Y برای تعیین طبقات و رویکرد حذف در براوردیابی

روش	N_E	N_C	N_S	n_S	$\hat{Y}_+$	bias	a.r.b.	MSE
دقیق	۲۸۹	۲۷	۶۸۴	۵۶	۱۲۸۱٫۴۶	−۴۶٫۱۰۹	۰٫۰۳	۴۶۲۸٫۰۹
شرط لازم	۲۸۹	۲۶	۶۸۵	۵۷	۱۱۵۰٫۶۸	−۱۷۶٫۸۹	۰٫۱۳	۴۴۶۳٫۴۳
کران بالا	۲۸۹	۲۲	۶۸۹	۶۲	۱۲۳۹٫۴۸	−۸۸٫۱۶	۰٫۰۶	۵۱۱۱٫۳۸

جدول ۶. نتایج در صورت استفاده از مقادیر Y برای تعیین طبقات و رویکرد مدل‌یار در براوردیابی

روش	N_E	N_C	N_S	n_S	$\hat{Y}_+$	bias	a.r.b.	MSE
دقیق	۲۸۹	۲۷	۶۸۴	۲۷	۱۴۷۰٫۲۰	۱۴۲٫۶۲	۰٫۱۱	۵۱۶۸٫۸
شرط لازم	۲۸۹	۲۶	۶۸۵	۲۶	۱۴۰۱٫۶۶	۷۴٫۰۸	۰٫۰۵۵	۴۵۹۱٫۹۵
کران بالا	۲۸۹	۲۲	۶۸۹	۲۲	۱۳۸۵٫۹۴	۵۸٫۳۶	۰٫۰۴	۳۴۹۹٫۹۶

با $c = ۰٫۰۵$ در براورد Y_+ است. همچنین فرض می‌کنیم در تعیین آستانه‌ی برش برای تشکیل طبقه‌ی حذف، $\varepsilon = ۰٫۱$ باشد. جدول‌های ۳ و ۴ نتایج به دست آمده از براورد Y_+ به‌کمک روش نمونه‌گیری برینشی نوع اول را در دو حالتِ اشاره‌شده و بر اساس ۱۰۰۰ بار تکرار نمونه‌گیری نشان می‌دهند.

در خصوص روش نمونه‌گیری برینشی نوع سوم، جدول ۵ نتایج به دست آمده از به‌کارگیری سه روش دقیق، شرط لازم و کران بالا روی متغیر Y برای تشکیل طبقات حذف، نمونه‌گیری و سرشماری، تعیین آستانه‌های برش و اندازه‌ی مطلوب نمونه برای رسیدن به دقت مورد نظر را وقتی رویکرد مورد استفاده در براورد Y_+ مبتنی بر حذف باشد، نشان می‌دهد. جدول ۶ اطلاعات مشابه را بر اساس ۱۰۰۰ بار تکرار نمونه‌گیری، وقتی رویکرد مدل‌یار در براورد Y_+ مورد استفاده قرار گیرد، نشان می‌دهد.

جدول ۷ نتایج به دست آمده از به‌کارگیری سه روش معرفی‌شده‌ی دقیق، شرط لازم و کران بالا برای تشکیل طبقات حذف، نمونه‌گیری و سرشماری، تعیین آستانه‌های برش و اندازه‌ی مطلوب نمونه برای رسیدن به دقت مورد نظر را وقتی محاسبات با استفاده از اطلاعات متغیر کمکی X، و رویکرد مورد استفاده در براورد Y_+ مبتنی بر حذف باشد، و بر اساس ۱۰۰۰ بار تکرار نمونه‌گیری نشان می‌دهد. همچنین جدول ۸ اطلاعات مشابه را در صورت استفاده از رویکرد مدل‌یار در براورد Y_+ نشان می‌دهد. واضح است که میزان اریبی و اریبی نسبی براوردهای به دست آمده در روش نمونه‌گیری برینشی نوع سوم در هر دو رویکرد مبتنی بر حذف و مدل‌یار، بسیار کم‌تر از اریبی نتایج متناظر در نمونه‌گیری برینشی نوع اول است.

جدول ۷. نتایج در صورت استفاده از مقادیر X برای تعیین طبقات و رویکرد حذف در براوردیابی

روش	N_E	N_C	N_S	n_S	$\hat{Y}_+$	bias	a.r.b.	MSE
دقیق	۲۱۲	۵	۷۸۳	۳۵	۱۴۷۹٫۱۷	۱۵۱٫۵۹	۰٫۱۱	۱۳۳۹۳٫۲۹
شرط لازم	۲۱۲	۴	۷۸۴	۳۶	۱۳۲۰٫۱۷	$-$۷٫۴۰	۰٫۰۰۷	۸۰۳۳٫۹۵
کران بالا	۲۱۲	۳	۷۸۶	۳۷	۱۳۳۷٫۳۱	۹٫۷۳	۰٫۱۵	۱۱۰۰۱٫۵۹

جدول ۸. نتایج در صورت استفاده از مقادیر X برای تعیین طبقات و رویکرد مدل‌یار در براوردیابی

روش	N_E	N_C	N_S	n_S	$\hat{Y}_+$	bias	a.r.b.	MSE
دقیق	۲۱۲	۵	۷۸۳	۲۹	۱۲۲۶٫۹۴	$-$۱۰۰٫۶۳	۰٫۰۷	۲۳۱۵۹٫۹
شرط لازم	۲۱۲	۴	۷۸۴	۳۰	۱۲۵۳٫۰۴	$-$۷۴٫۵۳	۰٫۰۵	۲۱۵۷۶٫۸
کران بالا	۲۱۲	۳	۷۸۶	۳۱	۱۴۰۰٫۹۶	۷۳٫۳۸	۰٫۰۵	۱۹۳۱۴٫۳

شایان ذکر است که در هر دو روش نمونه‌گیری برینشی مورد مطالعه در این بخش، اندازه‌ی مطلوب نمونه در صورت استفاده از اطلاعات کمکی، به میزان قابل توجهی کاهش می‌یابد. همچنین مقایسه‌ی نتایجِ به دست آمده نشان می‌دهد هرچند گاهی استفاده از متغیر کمکی در تعیین آستانه‌های برش و انجام دادن محاسبات لازم، موجب افزایش اریبی براوردها می‌شود، این افزایش چندان قابل توجه نیست و استفاده‌ی موفقیت‌آمیز از آن در عمل، که معمولاً اطلاعات Y در دسترس نیست، حائز اهمیت است. نکته‌ی قابل توجه دیگر آن است که نتایجِ به دست آمده در صورت استفاده از کران بالا، که محاسبات برای آن ساده‌تر است، دارای اختلاف زیادی با نتایجِ روش دقیق یا روش شرط لازم نیست. بنا بر این در عمل می‌توان محاسبات لازم اولیه را به‌کمک کران بالا انجام داد و سپس با استفاده از نظرهای کارشناسی و انجام تغییرات لازم، طرحِ آمارگیری نهایی را تشکیل داد. شایان ذکر است که در صورت تمایل می‌توان نسخه‌ای از برنامه‌ی مورد استفاده در نرم‌افزار S-PLUS 8 برای انجام دادن محاسبات لازم را از طریق مکاتبه با نویسنده‌ی عهده‌دار مکاتبات تهیه کرد.

سپاس‌گزاری

در پایان از آقای فرشید خان‌زاده و داوران محترم که نظرهای ارزشمندشان منجر به ارائه‌ی بهتر نتایجِ به دست آمده گردید، صمیمانه تشکر و قدردانی می‌شود.

مرجع‌ها

Bee, M.; Benedetti, R.; Espa, G. (2007). A framework for cut-off sampling in business survey design. Working Paper #0709, Department of Economics, University of Trento, http://ideas.repec.org/p/trn/utwpde/0709.html.

Elisson, H.; Elvers, E. (2001). Cut-off sampling and estimation. Proceedings of Statistics Canada Symposium 2001.

Haan, J. De; Opperdoes, E.; Smith C.M. (1999). Item selection in the consumer price index: cut off versus probability sampling. *Survey Methodology* **25**, 31-41.

Hidiroglou, M.A. (1986). The construction of a self representing stratum of large units in survey design. *Amer. Statist.* **40**, 27-31.

Hilpinen, J. (2006). Application of cut-off sampling in enterprise surveys for financial statistics: the case of foreign direct investments. Proceedings of Q 2006. European Conference on Quality in Survey Statistics.

Knaub, J.R. (2007a). Cut off sampling. In *Encyclopedia of Suvey Research Methods*, P.J. Lavrakas, ed., Sage. To appear.

Knaub, J.R. (2007b). Cut off sampling and inference. To appear in *Interstat Journal*.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika* **57**, 377-387.

Särndal, C.E.; Swensson, B.; Wretman, J. (1997). *Model Assisted Survey Sampling*. Springer.

Statistics Canada (2001). Monthly Survey of Manufacturing (MSM). Statistical Data Documentation System, Reference number 2101, Statistics Canada.

Steel, P.; Fey, R.E. (1995). Variance estimation for finite populations with imputed data. In *Proceedings of the Section on Survey Research Methods*, Vol. 1, American Statistical Association, pp. 374-379.

محمد جعفری جوزانی
گروه آمار، دانشکده‌ی اقتصاد،
دانشگاه علامه‌طباطبایی،
خیابان شهید بهشتی، نبش خیابان احمد قصیر،
تهران، ایران.
پیام‌نگار: *mjafari2002@yahoo.com*

فرشید جمشیدی
پژوهشکده‌ی آمار،
شماره‌ی ۱۴۵، خیابان سرتیپ فکوری،
خیابان باباطاهر، خیابان دکتر فاطمی،
تهران، ایران.
پیام‌نگار: *f_jamshidi@srtc.ac.ir*