

نمونه‌گیری

روشها و کاربردها

پل اس. لهوی

استنلی لمی شو

مترجم

گیتی مختاری امیرمجدی



پژوهشکده آمار

لوی، پل	Levy, Paul S.
نمونه‌گیری: روشها و کاربردها / پل اس. لهوی، استنلی لمی شو؛ مترجم گیتی مختاری امیرمجدی	
-- تهران: مرکز آمار ایران، پژوهشکده آمار، ۱۳۸۱.	
۵۸۵ ص. : جدول، نمودار.	
فهرست‌نویسی بر اساس اطلاعات فیفا.	ISBN 964-365-150-9 ریال : ۳۰۰۰۰
عنوان اصلی: Sampling of Populations: Methods and Applications.	
کتاب‌نامه.	
نمایه	
۱. جمعیت - روشهای آماری. ۲. آمارگیری نمونه‌ای. الف. لمشو، استنلی Lemeshow, Stanley	
ب. مختاری امیرمجدی، گیتی، ۱۳۳۳- ، مترجم. ج. مرکز آمار ایران، پژوهشکده آمار. د. عنوان.	
ان ۹/۴۹/۴۹ HB8۴۹/۴۹/۴۹	۳۰۴/۶۰۱۵۱۹۵۲
۱۳۸۱	
کتابخانه ملی ایران	۴۰۶۴۷-۸۱ م

- مدیریت تولید : گروه پژوهشی طرحهای فنی و روشهای آماری
- ویراستار علمی و ادبی : دکتر علی عمیدی
- حروف‌نگاری، نمونه‌خوانی، صفحه‌بندی، صفحه‌آرایی : محبوبه کاظمی
- ویراستار هنری : فرشید خان‌زاده
- طراحی جلد : نیما دانش‌پرور
- مدیر فنی : علی اصغر حائری مهریزی
- امور فنی و چاپ : مؤسسه انتشارات ستایش

© ۱۳۸۱ پژوهشکده آمار

شماره ۵۲، خیابان شهید فکوری، خیابان باباطاهر، خیابان دکتر فاطمی
تهران ۱۴۱۳۷۱۷۹۱۱، ایران



URL: <http://www.src.ac.ir>

e-mail: src@src.ac.ir

تلفن: ۸۹۵۹۰۲۹ دورنگار: ۸۰۰۷۹۸۹

همه حقوق این اثر برای پژوهشکده آمار محفوظ است. هیچ بخشی از این کتاب را نمی‌توان بدون اجازه کتبی از ناشرش تکثیر یا به هر شکلی و با هر وسیله‌ای ذخیره کرد. استفاده یا تقلید از طرح جلد، ممنوع است.

حروف‌نگاری شده با قلم‌های فارسی لوتوس و تیترو میترا، و قلم لاتین Times New Roman.

چاپ و صحافی شده در ایران.

چاپ یکم

شمارگان: ۶۰۰

پیشنهاد برای نحوه نقل مطلب، جدول یا نمودار از این کتاب، به صورت زیر است:

مختاری امیرمجدی، گیتی (۱۳۸۱). نمونه‌گیری: روشها و کاربردها. لهوی، پل اس.؛ لمی شو، استنلی. ترجمه از انگلیسی به فارسی. تهران: پژوهشکده آمار.

شابک ۹۶۴ - ۳۶۵ - ۱۵۰ - ۹

ISBN 964-365-150-9

بها: سی و پنج هزار ریال

فصل ۸

نمونه‌گیری خوشه‌ای:

مقدمه و بررسی اجمالی

فنون نمونه‌گیری که در فصلهای قبل مورد بحث قرار گرفتند همه مستلزم چارچوبهای نمونه‌گیری هستند که واحدهای شمارش تکی (یا واحدهای فهرست‌برداری) را فهرست می‌کنند. ولی گاهی اوقات، بخصوص در آمارگیریهای نمونه‌ای از جوامع انسانی، تدوین چارچوبهایی برای نمونه‌گیری که همه واحدهای شمارش را برای تمام جامعه فهرست کنند عملی و شاید اصلاً امکان‌پذیر نیست. از سوی دیگر، غالباً می‌توان چارچوبهایی برای نمونه‌گیری ساخت که گروهها یا خوشه‌هایی از واحدهای شمارش را شناسایی کنند بدون این که واحدهای شمارش تکی را صریحاً فهرست نمایند. می‌توان از روی این قبیل چارچوبها، با گرفتن نمونه‌ای از خوشه‌ها، به دست آوردن فهرستی از واحدهای شمارش، فقط برای آن خوشه‌هایی که در نمونه انتخاب شده‌اند، و سپس انتخاب نمونه‌ای از واحدهای شمارش، نمونه‌گیری را انجام داد. این طرحهای نمونه به نام *نمونه‌های خوشه‌ای موسوم‌اند* و در عمل کاربرد وسیعی دارند. در این فصل، برخی مفاهیم پایه‌ای نمونه‌گیری خوشه‌ای را معرفی می‌کنیم و در سه فصل بعدی به بحث در مورد برخی فنون خاص نمونه‌گیری خوشه‌ای می‌پردازیم که در سطح گسترده‌ای به کار می‌روند.

برای نشان دادن فرایند نمونه‌گیری خوشه‌ای و مقابله آن با طرحهای مبتنی بر نمونه‌گیری مستقیم از واحدهای شمارش، مثالی ساده را در نظر می‌گیریم.

مثال تشریحی: فرض کنید می‌خواهیم به منظور بررسی استفاده از خدمات بهداشتی در میان اهالی یک شهر با اندازه متوسط، نمونه‌ای از خانوارهای آن شهر انتخاب کنیم. اگر راهنمای بهنگام شهر که تمام خانوارهای شهر را فهرست می‌کند در دسترس باشد می‌توان آن را به عنوان چارچوب نمونه‌گیری برای انتخاب نمونه به کار برد. ولی اگر چنین راهنمایی یا هیچ گونه فهرست دیگری از خانوارها موجود نباشد، ساختن چنین چارچوبی برای نمونه‌گیری، با توجه به نفر - ساعت لازم، بسیار پرهزینه خواهد بود.

ولی ساختن فهرستی از بلوکهای شهر نسبتاً آسان است. برای بیشتر شهرها، نقشه‌هایی که هر بلوک شهری در آن مشخص و شماره‌گذاری شده است از طریق دفتر سرشماری امریکا در اختیار قرار می‌گیرد. این فهرست بلوکهای شهری را می‌توان به عنوان چارچوب نمونه‌گیری به کار برد. هر بلوک شهری را می‌توان به عنوان خوشه‌ای از خانوارها در نظر گرفت و هر خانوار در شهر با یک بلوک خاص همراه خواهد بود. نمونه خانوارها را می‌توان ابتدا با گرفتن نمونه‌ای از بلوکها و سپس با فهرست‌برداری از یکایک خانوارها در داخل هر بلوکی که در نمونه انتخاب شده است تعیین کرد. نمونه نهایی خانوارها را می‌توان از این فهرست به دست آمده از خانوارها انتخاب نمود. توجه کنید که لازم است فهرست‌برداری فقط از خانوارهای بلوکهایی به عمل آید که در نمونه بلوکهای انتخاب شده قرار دارند.

□

۱.۸ نمونه‌گیری خوشه‌ای چیست؟

اصطلاح خوشه، وقتی در روش‌شناسی آمارگیری نمونه‌ای به کار می‌رود، می‌تواند به صورت هر واحد نمونه‌گیری که بتوان یک یا چند واحد فهرست‌برداری را با آن همراه کرد تعریف شود. واحد می‌تواند دارای ماهیت جغرافیایی، زمانی، یا مکانی باشد. چند مثال از خوشه‌هایی که ممکن است در عمل پیش بیاید در جدول ۱.۸ نشان داده شده‌اند.

برای توضیح بیشتر مثال دوم خوشه را که در جدول نشان داده شده است در نظر می‌گیریم. اگر بخواهیم نسبت همه بیماران بستری شده در بیمارستان را که مرده آنها از بیمارستان خارج شده است در یک ایالت خاص، طی یک سال خاص، برآورد کنیم می‌توانیم ابتدا تمام بخشهای آن ایالت را فهرست کنیم و از این فهرست، نمونه‌ای از بخشها بگیریم. سپس می‌توانیم برای هر یک از بخشهای انتخاب شده در نمونه، کلیه بیمارستانهای موجود در آن بخش را فهرست کنیم و از بیمارستانهای تکی،

نمونه‌ای بگیریم. بالاخره، برای هر بیمارستانی که در نمونه انتخاب شده است کل تعداد اشخاصی را که در طول سال پذیرفته شده‌اند و کل افرادی را که مرده ترخیص شده‌اند به دست می‌آوریم. از روی این داده‌ها می‌توان نسبت کسانی را که مرده ترخیص شده‌اند برای سراسر ایالت برآورد کرد. برای آخرین مثالی که در جدول ۱۸ ارائه شده است، می‌توان فهرستی شامل ۵۲ هفته‌تقویمی تهیه و نمونه‌ای از هفته‌ها را از روی فهرست انتخاب کرد. برای هر یک از هفته‌های انتخاب شده در نمونه، می‌توان نمونه‌ای از روزها را انتخاب کرد و اندازه‌گیری اُژن را در هر روز نمونه انجام داد.

جدول ۱۸ برخی مثالهای کاربردی برای خوشه‌ها

خوشه	واحد فهرست‌برداری	واحد اولیه	کاربرد
بلوک شهری	خانوار	شخص	برآورد کردن کل ساکنان شهر که فشار خون دارند
بخش	بیمارستان	بیمار	برآورد کردن نسبت ترخیص‌شدگان مرده در یک ایالت خاص
مدرسه	کلاس	دانش‌آموز	برآورد کردن میانگین دستاوردهای تحصیلی در میان دانش‌آموزان یک ناحیه تحصیلی
بسته سرنگ	سرنگ تکی	سرنگ تکی	برآورد کردن نسبت تمام سرنگهای معیوب
کشو بایگانی	پوشه تکی	حساب	برآورد کردن صورت حسابهایی که به موقع پرداخت نشده‌اند
صفحه متن	سطر متن	واژه	برآورد کردن کل تعداد واژه‌های موجود در یک کتاب
هفته	روز	روز	برآورد کردن کل روزهایی که اُژن در حداکثر سطح خود و بالاتر از یک سطح تعیین شده است

رابطه بین خوشه‌ها و واحدهای فهرست‌برداری بسیار شبیه به رابطه بین واحدهای فهرست‌برداری و واحدهای اولیه است. یک خوشه دارای واحدهای فهرست‌برداری همراه با خود است، درست به همان ترتیبی که یک واحد فهرست‌برداری دارای واحدهای اولیه همراه با خود است. به طوری که بعداً خواهیم دید، نمونه‌گیری خوشه‌ای نوع سلسله‌مراتبی نمونه‌گیری است که در آن، واحدهای اولیه غالباً حداقل دو مرحله با نمونه‌گیری اولیه خوشه‌ها فاصله دارند.

حال که مفهوم خوشه‌ها را تعریف کردیم می‌توانیم نمونه‌گیری خوشه‌ای را در سطحی وسیع مانند هر برنامه نمونه‌گیری دیگری که از چارچوب شامل خوشه‌هایی از واحدهای فهرست‌برداری استفاده می‌کند تعریف کنیم. تعریف بالا گسترده‌تر از آن است که در برخی متون دیگر ارائه شده است از این نظر که شامل نمونه‌گیری است که در بیش از یک مرحله اجرا می‌شود. (یعنی نمونه‌گیری چندمرحله‌ای). برخی پژوهشگران اصطلاح نمونه‌گیری خوشه‌ای را به آن طرح‌های نمونه‌گیری محدود می‌کنند که در آن خوشه‌ها توسط نوعی برنامه نمونه‌گیری انتخاب می‌شوند و سپس هر واحد شمارش در داخل هر خوشه نمونه‌گیری می‌شود. در کتاب حاضر، ما به این قبیل طرح‌ها به عنوان نمونه‌گیری خوشه‌ای یک مرحله‌ای اشاره می‌کنیم.

با در نظر گرفتن این تعریف گسترده‌تر از نمونه‌گیری خوشه‌ای، برخی جنبه‌های مهم نمونه‌گیری خوشه‌ای را در زیر فهرست می‌کنیم.

۱. فرایندی که طی آن نمونه‌ای از واحدهای فهرست‌برداری انتخاب می‌شود ممکن است گام به گام باشد. برای مثال، اگر بلوکهای شهری خوشه‌ها باشند و خانوارها واحدهای فهرست‌برداری را تشکیل دهند انتخاب خانوارهای نمونه ممکن است در دو گام اجرا شود. گام اول ممکن است مستلزم انتخاب نمونه بلوکها و گام دوم می‌تواند مستلزم انتخاب خانوارهای نمونه در داخل هر یک از بلوکهای انتخاب شده در گام اول باشد. در اصطلاح‌شناسی نمونه‌گیری، این گامها را مرحله می‌نامند و برنامه‌های نمونه‌گیری غالباً بر حسب تعداد مرحله‌های به کار رفته مشخص می‌شوند. برای مثال، نمونه خوشه‌ای یک مرحله‌ای نمونه‌ای است که طی آن نمونه‌گیری فقط در یک گام اجرا می‌شود. یعنی همین که نمونه خوشه‌ها انتخاب شد، همه واحدهای فهرست‌برداری در داخل هر یک از خوشه‌های انتخاب شده در نمونه منظور می‌شوند.

در بسیاری از آمارگیریهایی که مناطق وسیع جغرافیایی را پوشش می‌دهند، غالباً نمونه‌گیری در چندین مرحله انجام می‌شود. برای مثال، آمارگیری مصونیت از کودکان دبستانی در یک ایالت ممکن است مستلزم پنج مرحله باشد. اول، یک نمونه از بخشهای داخل ایالت

می‌گیریم. دوم، یک نمونه از شهرها یا سایر تقسیمات شهری کوچکتر در داخل هر یک از بخشهایی که در مرحله اول انتخاب شده‌اند می‌گیریم. سوم، یک نمونه از نواحی آموزشی در داخل شهرهایی که در مرحله دوم انتخاب شده‌اند می‌گیریم. چهارم، یک نمونه از مدرسه‌ها در داخل نواحی آموزشی که در مرحله سوم انتخاب شده‌اند می‌گیریم. پنجم، یک نمونه از کلاسهای درس در داخل هر یک از مدرسه‌هایی که در مرحله چهارم انتخاب شده‌اند می‌گیریم. و بالاخره یکایک کودکان را در کلاسهای درسی که در مرحله پنجم انتخاب شده‌اند آمارگیری می‌کنیم. در این مثال، کودکان واحدهای اولیه محسوب می‌شوند و کلاسهای درس واحدهای فهرست‌برداری هستند. باید پنج مرحله نمونه‌گیری اجرا شود که مستلزم چهار نوع خوشه‌اند: بخشها، شهرها، نواحی آموزشی، و مدرسه‌ها. در طرحهای نمونه‌ای که شامل دو یا چند مرحله‌اند، خوشه‌های مورد استفاده در مرحله اول نمونه‌گیری عموماً واحدهای نمونه‌گیری اولیه نامیده می‌شوند که به صورت خلاصه شده PSU^۱ به آن اشاره می‌شود.

۲. خوشه‌ها را می‌توان با انواع گوناگونی از فنون نمونه‌گیری انتخاب کرد. برای مثال می‌توانیم نمونه‌ای از خوشه‌ها را با نمونه‌گیری تصادفی ساده یا با نمونه‌گیری سیستماتیک انتخاب کنیم. می‌توانیم خوشه‌ها را در طبقاتی گروه‌بندی کنیم و یک نمونه تصادفی طبقه‌بندی شده از خوشه‌ها بگیریم.

هرگاه خوشه‌ها با نمونه‌گیری تصادفی ساده انتخاب شده باشند معمولاً برای توصیف طرح نمونه‌گیری از عبارت نمونه‌گیری خوشه‌ای ساده استفاده می‌شود. به طور دقیقتر برای رسته‌بندی طرحهای نمونه‌ای که در آن نمونه‌گیری تنها در یک مرحله اجرا می‌شود و خوشه‌ها با نمونه‌گیری تصادفی ساده انتخاب می‌شوند اصطلاح نمونه‌گیری خوشه‌ای یک مرحله‌ای ساده به کار می‌رود. به همین ترتیب، برای توصیف طرحهای نمونه‌گیری که در آن خوشه‌ها در مرحله اول با نمونه‌گیری تصادفی ساده انتخاب می‌شوند، و واحدهای فهرست‌برداری در مرحله دوم به طور مستقل در داخل هر خوشه نمونه باز هم با نمونه‌گیری تصادفی ساده انتخاب می‌شوند، و کسر واحدهای فهرست‌برداری که در مرحله دوم انتخاب شده برای هر خوشه نمونه یکسان است، از اصطلاح نمونه‌گیری خوشه‌ای دو مرحله‌ای ساده استفاده می‌شود. نمونه‌گیری خوشه‌ای یک مرحله‌ای ساده در فصل ۹ و نمونه‌گیری خوشه‌ای دو مرحله‌ای ساده در فصل ۱۰ مورد بحث قرار گرفته‌اند.

^۱ Primary Sampling Units

نوع دیگری از نمونه‌گیری خوشه‌ای که غالباً در عمل مورد استفاده قرار می‌گیرد نمونه‌گیری با احتمال متناسب با اندازه یا نمونه‌گیری PPS^۱ نامیده می‌شود. این نوع نمونه‌گیری که در فصل ۱۱ مورد بحث قرار می‌گیرد، نمونه‌گیری است که در آن خوشه‌ها با نمونه‌گیری تصادفی ساده انتخاب نمی‌شوند.

۳. در این فرایند ممکن است بیش از یک چارچوب نمونه‌گیری به کار رود. برای نشان دادن این وضعیت، به مثال بالا برمی‌گردیم که در آن یک آمارگیری مصونیت با یک شیوه پنج‌مرحله‌ای اجرا می‌شود. چارچوب نمونه‌گیری در مرحله اول، فهرست تمام بخشهای داخل ایالت است. چارچوب نمونه‌گیری در مرحله دوم، فهرست شهرها یا سایر تقسیمات کوچکتر شهری در داخل هر یک از بخشهای انتخاب شده در مرحله اول است. چارچوب نمونه‌گیری در مرحله سوم، فهرست تمام نواحی آموزشی است که در داخل شهرها یا تقسیمات کوچکتر شهری انتخاب شده در مرحله دوم واقع شده‌اند. چارچوب نمونه‌گیری در مرحله چهارم نمونه‌گیری، فهرست مدرسه‌ها در داخل هر یک از نواحی آموزشی مرحله سوم است. بالاخره، چارچوب نمونه‌گیری مورد استفاده در مرحله پنجم، فهرست کلاسهای درس در داخل مدرسه‌هایی است که در مرحله چهارم نمونه‌گیری شده‌اند.

۴. پس از اولین مرحله نمونه‌گیری، چارچوب نمونه‌گیری فقط برای خوشه‌هایی که در نمونه انتخاب شده‌اند تهیه می‌شود. به محض این که خوشه‌های نمونه مرحله اول انتخاب شدند، فهرست‌برداری از واحدهای نمونه‌گیری مرحله دوم فقط برای خوشه‌های نمونه انجام می‌شود. به همین ترتیب، اگر نمونه‌گیری در بیش از دو مرحله اجرا می‌شود واحدهای نمونه‌گیری هر یک از مراحل بعدی فقط در واحدهای نمونه‌گیری که در مرحله قبلی انتخاب شده‌اند فهرست می‌شوند. چون فهرست‌برداری از خوشه‌ها یا واحدهای فهرست‌برداری غالباً یکی از عملیات میدانی پرهزینه است، معمولاً هزینه‌های فهرست‌برداری برای طرحهای نمونه‌گیری خوشه‌ای به مراتب کمتر از طرحهای دیگر خواهد بود.

۲.۸ چرا نمونه‌گیری خوشه‌ای در سطحی گسترده به کار می‌رود؟

مهمترین دو دلیلی که نمونه‌گیری خوشه‌ای در سطحی چنین گسترده در عمل مورد استفاده قرار می‌گیرد، به خصوص در آمارگیریهای نمونه‌ای از جوامع انسانی و در آمارگیریهایی که مناطق جغرافیایی وسیعی را پوشش می‌دهند/مکان‌پذیر بودن و اقتصادی بودن آن است.

^۱ Probability Proportional to Size

نمونه‌گیری خوشه‌ای غالباً تنها روش نمونه‌گیری امکان‌پذیر است، زیرا تنها چارچوبهای نمونه‌گیری که برای جوامع هدف به راحتی در دسترس قرار می‌گیرند فهرستهای خوشه‌ها هستند. این حالت بخصوص در مورد آمارگیریهای جوامع انسانی که در آن از خانوار به عنوان واحد فهرست‌برداری استفاده می‌شود مصداق دارد. تهیه فهرستی از خانوارها برای یک جامعه آماری بزرگ (مانند آمریکا، یک ایالت، یا حتی یک شهر) فقط به منظور اجرای یک آمارگیری با توجه به زمان و منابع لازم، تقریباً هرگز امکان‌پذیر نیست. ولی فهرست بلوکها یا سایر واحدهای جغرافیایی را می‌توان نسبتاً به آسانی تهیه کرد و آنها را به عنوان چارچوب نمونه‌گیری به کار برد.

نمونه‌گیری خوشه‌ای غالباً باصرفه‌ترین شکل نمونه‌گیری است. نه تنها هزینه‌های فهرست‌برداری تقریباً همیشه برای نمونه‌گیری خوشه‌ای از همه کمتر است، بلکه هزینه‌های رفت و آمد هم غالباً از همه کمتر است. برای مثال، اگر خوشه یک واحد جغرافیایی از قبیل یک ناحیه سرشماری باشد، در آن صورت، به محض این که خانوارها در داخل نواحی سرشماری انتخاب شدند هزینه‌های رفت و آمد در داخل نواحی نمونه از خانواری به خانوار دیگر نسبتاً اندک خواهد بود. به این ترتیب، یک نمونه خوشه‌ای از خانوارها در داخل چند ناحیه سرشماری نمونه، مستلزم رفت و آمدی به مراتب کمتر از یک نمونه تصادفی ساده با همان تعداد خانوار است که در تعداد بسیار بیشتری از نواحی سرشماری پراکنده هستند.

نمونه‌گیری خوشه‌ای می‌تواند در آمارگیریهای نهادها از قبیل بیمارستانها سودمند واقع شود. وارد کردن یک نهاد در یک بررسی غالباً بسیار پرخرج و وقت‌گیر است. برای مثال، کسب اجازه از مدیر یک بیمارستان برای بیرون کشیدن نمونه‌ای از پرونده‌های بیماران بیمارستان ممکن است مستلزم تلاش بسیار زیاد، مقداری تخصص در روابط عمومی، و گاهی حتی اعمال نفوذ فراوان باشد. بنابراین، به محض این که دسترسی به پرونده‌های بیمارستان فراهم شد غالباً این ارزش را دارد که به جای چند پرونده تعداد زیادی از پرونده‌ها نمونه‌گیری شوند.

برای توضیح بیشتر در مورد این که نمونه‌گیری خوشه‌ای چگونه می‌تواند هزینه‌های یک آمارگیری نمونه‌ای را پایین بیاورد به بررسی مثال زیر می‌پردازیم.

مثال تشریحی: فرض کنید پنج شهرک مسکونی برای شهروندان سالخورده در سراسر یک ایالت خاص به طور پراکنده وجود دارند و هر یک از این شهرکها دارای ۲۰ آپارتمان است. فرض کنید قرار است نمونه‌ای از ۱۰ آپارتمان انتخاب شود تا کل تعداد شهروندان سالخورده در این شهرکها که نیاز به خدمات پرستار متخصص دارند برآورد شود. علاوه بر این، فرض کنیم که خانوارها واحدهای

فهرست‌برداری هستند و هزینه‌های میدانی عمده آمارگیری مربوط به کارهای فهرست‌برداری و مصاحبه است.

برای فهرست کردن همه آپارتمانها در یک شهرک مسکونی، فرض می‌کنیم که یکی از اعضای کارکنان میدانی ناچار است به شهرک برود و نامهای خانوادگی را از روی جعبه‌های پستی با سرعت یادداشت کند. فرض کنید رفتن به شهرک ۰/۵ ساعت و فهرست کردن هر خانوار ۳ دقیقه وقت بگیرد، یا فهرست‌برداری از تمام خانوارها در شهرک (دقیقه $60 = 3 \times 20$) وقت لازم داشته باشد. پس فرض می‌کنیم که برای فهرست کردن سراسر یک شهرک ۱/۵ نفر ساعت (۱ ساعت فهرست‌برداری به اضافه ۰/۵ ساعت سفر) و برای فهرست کردن همه شهرکهای پنج‌گانه ۷/۵ ساعت (۱/۵ ساعت برای هر شهرک ضربدر ۵ شهرک) وقت لازم باشد.

فرض کنید مصاحبه با هر خانوار انتخاب شده در نمونه ۱۵ دقیقه (۰/۲۵ ساعت) وقت بگیرد. پس هزینه مصاحبه از لحاظ نفر - ساعت در یک شهرک خاص برابر است با ۰/۲۵ برابر تعداد خانوارهای انتخاب شده در آن شهرک به اضافه هزینه سفر به شهرک (۰/۵ نفر - ساعت برای هر سفر).

حالا بیایید دو طرح نمونه‌گیری را با توجه به هزینه‌های فهرست‌برداری و مصاحبه برای نمونه‌ای از ۱۰ خانوار مقایسه کنیم. طرح اول، نمونه‌گیری تصادفی ساده است و فرض می‌کنیم خانوارهایی از چهار شهرک در نمونه انتخاب شده‌اند. طرح دوم، یک نمونه خوشه‌ای دو مرحله‌ای ساده از ۱۰ خانوار است یعنی پنج خانوار از هر یک از دو شهرک. هزینه‌های فهرست‌برداری و مصاحبه برای هر طرح در جدول ۲.۸ نشان داده شده است.

جدول ۲.۸ مقایسه هزینه‌ها در دو طرح نمونه‌گیری

طرح		هزینه‌ها
نمونه‌گیری خوشه‌ای دو مرحله‌ای ساده	نمونه‌گیری تصادفی ساده	
۱/۵ برای هر شهرک	۱/۵ برای هر شهرک	هزینه‌های فهرست‌برداری
$2 \times \text{شهرک} = 3/0$	$5 \times \text{شهرک} = 7/5$	
سفر به ۲ شهرک + مصاحبه	سفر به ۴ شهرک + مصاحبه	هزینه‌های مصاحبه
$2 \times 0/5 + 10 \times 0/25 = 3/5$	$4 \times 0/5 + 10 \times 0/25 = 4/5$	برحسب نفر - ساعت
۶/۵	۱۲/۰	کل هزینه‌های میدانی
		برحسب نفر - ساعت

از محاسبات جدول ۲.۸ پی می‌بریم که حتی در مورد این نمونه کوچک نیز هزینه‌های میدانی مبتنی بر نمونه‌گیری خوشه‌ای، می‌تواند به مراتب کمتر از هزینه‌های میدانی نمونه‌گیری تصادفی ساده باشد. صرفه‌جوییها بخصوص در هزینه‌های فهرست‌برداری و مصاحبه است.



۳۸ عیب نمونه‌گیری خوشه‌ای: خطاهای معیار زیاد

اکنون که نشان دادیم نمونه‌گیری خوشه‌ای چقدر می‌تواند مقرون به صرفه باشد و غالباً چقدر امکان‌پذیر است، باید خاطر نشان کنیم که خطاهای معیار برآوردهای به دست آمده از طرحهای نمونه‌گیری خوشه‌ای در مقایسه با خطاهای معیار به دست آمده از نمونه‌های دیگری با همین تعداد واحدهای فهرست‌برداری که از طرحهای نمونه‌گیری دیگری انتخاب شده‌اند غالباً بسیار زیاد است. علت این وضعیت آن است که واحدهای فهرست‌برداری در داخل یک خوشه از لحاظ بسیاری از مشخصه‌ها غالباً همگن‌اند. برای مثال، خانوارهای ساکن در یک بلوک غالباً از لحاظ وضعیت اجتماعی - اقتصادی، قومی، و سایر متغیرها بسیار شبیه‌اند. به علت همگنی میان واحدهای فهرست‌برداری در داخل یک خوشه، انتخاب بیش از یک خانوار در داخل همان خوشه، به صورتی که در نمونه‌گیری خوشه‌ای اجرا می‌شود، به تعبیری، زیاده‌روی محسوب می‌شود. اثر این زیاده‌روی در خطاهای معیار بسیار زیاد برآوردها آشکار می‌شود که غالباً در نمونه‌گیری خوشه‌ای به چشم می‌خورد. اگر قرار باشد بین نمونه‌گیری خوشه‌ای و هر طرح دیگری صرفاً براساس هزینه و امکان‌پذیر بودن، یکی را انتخاب کنیم، نمونه‌گیری خوشه‌ای به صورتی اجتناب‌ناپذیر طرح نمونه‌گیری منتخب خواهد بود. از سوی دیگر، اگر قرار باشد طرحی را صرفاً براساس قابل اعتماد بودن برآوردها انتخاب کنیم، در آن صورت نمونه‌گیری خوشه‌ای هرگز نمی‌تواند طرح منتخب باشد. ولی چون امکان گرفتن نمونه‌ای بزرگتر با هزینه ثابت با نمونه‌گیری خوشه‌ای فراهم است، دقت نتایج به مراتب به بیش از آن خواهد رسید که در سایر روشها می‌رسد. معمولاً در انتخاب بین نمونه‌گیری خوشه‌ای و سایر گزینه‌ها، از ملاکی استفاده می‌کنیم که قابل اعتماد بودن و هزینه را مشترکاً در نظر بگیرد. در واقع، عموماً آن طرح نمونه‌گیری را انتخاب می‌کنیم که کمترین خطای معیار ممکن را با هزینه‌ای مشخص به دست آورد یا، برعکس، آن طرح نمونه‌گیری را انتخاب می‌کنیم که با کمترین هزینه، برآوردهایی را با خطای معیار مشخص نتیجه دهد.

۴.۸ چگونگی بررسی نمونه‌گیری خوشه‌ای در این کتاب

به دنبال بررسی مجمل نمونه‌گیری خوشه‌ای که در این فصل ارائه شد، در سه فصل بعدی به شرح و بسط طرح‌های مشخص نمونه‌گیری می‌پردازیم و دربارهٔ طرح‌های نمونه‌گیری بسیار ساده تا طرح‌های نمونه‌گیری بسیار پیچیده بحث می‌کنیم. ساده‌ترین نوع نمونه‌گیری خوشه‌ای، نمونه‌گیری خوشه‌ای یک‌مرحله‌ای است که در آن خوشه‌ها با نمونه‌گیری تصادفی ساده انتخاب می‌شوند و از این نوع نمونه‌گیری خوشه‌ای در فصل بعد (فصل ۹) بحث شده است. دو فصل پس از آن (فصل‌های ۱۰ و ۱۱) هر دو به نمونه‌گیری خوشه‌ای دو مرحله‌ای می‌پردازند. در نمونه‌گیری خوشه‌ای دو مرحله‌ای، تهیهٔ برآوردهای مشخصه‌های جامعه و برآوردهای خطاهای معیار آن دسته از آماره‌هایی که برای برآورد مشخصه‌های جامعه به کار می‌روند اگر تعداد واحدهای فهرست‌برداری در هر خوشه در جامعه یکسان باشد نسبتاً آسان است. نمونه‌گیری خوشه‌ای دو مرحله‌ای برای این وضعیت در فصل ۱۰ بررسی شده است. نمونه‌گیری خوشه‌ای دو مرحله‌ای در پیچیده‌ترین شکل خود هنگامی روی می‌دهد که تعداد واحدهای فهرست‌برداری در هر خوشه از جامعه یکسان نباشد. متأسفانه، این حالت در بسیاری از وضعیتهای کاربردی اتفاق می‌افتد و احساس می‌کنیم که باید در این کتاب به آن بپردازیم، گرچه روش مزبور مستلزم استفاده از فرمولهایی است که به مراتب خسته‌کننده‌تر از آنهایی هستند که در سایر بخشهای این کتاب ارائه شده‌اند. این نوع نمونه‌گیری خوشه‌ای را در فصل ۱۱ شرح و بسط می‌دهیم. طی ۲۵ سال گذشته، پیشرفتهایی قابل ملاحظه در توسعهٔ روشها و نرم افزارهای آماری برای برآورد خطاهای معیار برآوردهای حاصل از حتی پیچیده‌ترین طرح‌های نمونه‌گیری خوشه‌ای به دست آمده است. همان طور که در فصلهای قبل انجام شد، نشان خواهیم داد که چگونه دو نرم افزار موجود فعلی یعنی SUDAAN و STATA را می‌توان برای برآورد کردن خطاهای معیار برآورد مشخصه‌های جامعه تحت طرح‌های نمونه‌گیری خوشه‌ای پیچیده به کار برد. در فصل ۱۲، برخی از روشهای آماری را که کاربرد وسیعتری دارند برای به دست آوردن برآوردهای این خطاهای معیار به صورتی کاملتر مورد بحث قرار خواهیم داد. رفتار ما با نمونه‌گیری خوشه‌ای بسیار شبیه به رفتار ما با سایر طرح‌های نمونه‌گیری است. به خصوص بحثهایی را به چگونگی انتخاب نمونه‌ها، چگونگی به دست آوردن برآوردها، طبیعت توزیعهای نمونه‌گیری این برآوردها، و روشهای برآورد کردن خطاهای معیار این برآوردها اختصاص می‌دهیم.

۵.۸ خلاصه

در این فصل، بررسی مجملی از نمونه‌گیری خوشه‌ای ارائه دادیم. مفهوم یک خوشه از واحدهای فهرست‌برداری یا شمارش را معرفی و نمونه‌گیری خوشه‌ای را مانند هر طرح نمونه‌گیری دیگری تعریف کردیم که از یک چارچوب نمونه‌گیری شامل خوشه‌هایی از واحدهای فهرست‌برداری استفاده می‌کند. مثالهایی برای خوشه‌ها ارائه شد و مشخصه‌هایی از نمونه‌گیری خوشه‌ای مورد بحث قرار گرفت از قبیل ماهیت گام به گام دنباله‌ای و این واقعیت که واحدهای فهرست‌برداری لازم است فقط از خوشه‌های نمونه فهرست‌برداری شوند و نه از تمام خوشه‌ها. به مزایای نمونه‌گیری خوشه‌ای (یعنی، امکان‌پذیر بودن و اقتصادی بودن) و نیز به عیب عمده آن (یعنی، خطاهای معیار بسیار زیاد برآوردها) اشاره کردیم. ولی به علت امکانات بالقوه این طرح برای داشتن مشاهداتی بیشتر از آنچه که با استفاده از سایر طرحها با همان بودجه امکان‌پذیر است غالباً این امکان وجود دارد که خطاهای معیار تا آن حد کاهش یابد که نمونه‌گیری خوشه‌ای روش منتخب باشد.

تمرین

- ۱.۸ در نمونه‌ای با ۱۰۰ عنصر، کدام یک از طرحهای زیر احتمال دارد برآوردهایی با بیشترین خطاهای معیار به دست دهد؟
- الف. نمونه‌گیری سیستماتیک
- ب. نمونه‌گیری خوشه‌ای
- پ. نمونه‌گیری تصادفی ساده
- ت. نمونه‌گیری تصادفی طبقه‌بندی شده
- ۲.۸ اگر کسی بخواهد از این صفحه در مورد واژه‌هایی که غلط نوشته شده‌اند نمونه بگیرد، خوشه منطقی کدام خواهد بود؟
- ۳.۸ اگر کسی بخواهد در سرتاسر این کتاب در مورد واژه‌هایی که غلط نوشته شده‌اند براساس یک طرح نمونه‌گیری خوشه‌ای سه‌مرحله‌ای نمونه بگیرد، واحدهای نمونه‌گیری اولیه چه می‌توانند باشند؟ واحدهای نمونه‌گیری ثانویه، واحدهای فهرست‌برداری، و واحدهای مقدماتی کدام خواهند بود؟

کتابشناسی

Cluster sampling is discussed in all of the texts on sampling referenced in previous chapters. References to specific methods and applications of cluster sampling are given in Chapters 9, 10, 11, and 14.