

نمونه‌گیری

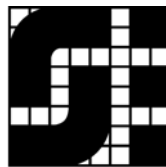
روشها و کاربردها

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 ۱۵

نمونه‌گیری تلفنی

رابرت جی. کاسادی و جیمز ام. لپکوفسکی^۱

۱.۱۵ نظری اجمالی

استفاده از تلفن برای جمع‌آوری داده‌های آمارگیری مستلزم انتخاب نمونه‌هایی از شماره تلفن‌ها به منظور تعیین واحدهای نمونه‌گیری برای مصاحبه است. فنون نمونه‌گیری پایه که برای این انتخاب به کار گرفته می‌شوند همانهایی هستند که برای بسیاری دیگر از مسایل طراحی نمونه مورد استفاده قرار می‌گیرند. ولی چارچوبهای مورد استفاده برای انتخاب نمونه تلفنی دارای چندین خصیصه منحصر به خود هستند که بسط طرحهای نمونه‌ای ویژه‌ای برای آمارگیریهای تلفنی را برانگیخته‌اند.

چارچوبهای موجود از کشوری به کشور دیگر و از یک نوع واحد به نوعی واحد دیگر (مانند خانوار در برابر کارگاه) فرق می‌کنند. ما بحث خود را در اینجا به چارچوبها و طرحهای نمونه‌گیری برای خانوارهای تلفن‌دار در آمریکا محدود می‌کنیم. طرحها و چارچوبها در کشورهای دیگر یا برای واحدهای دیگری غیر از خانوارها خصیصه‌هایی مشابه آنچه در اینجا بحث شده است خواهند داشت. ولی، همان‌طور که می‌توان انتظار داشت، جنبه‌های ویژه طرحهای سایر کشورها یا سایر واحدها را می‌توان برای بهبود کارایی و سایر خواص عملیات آمارگیری تا حدودی تعدیل کرد.

^۱ Robert J. Casady and James M. Lepkowski

نمونه‌گیری تلفنی از خانوارها باید با بررسی دقیق خصیصه‌های چارچوبهای موجود شروع شود. روشهایی برای انتخاب نمونه ابداع شده‌اند تا به خصیصه‌های ویژه هر چارچوب بپردازند. در بقیه این بخش به بحث در مورد چارچوبها و طرحهای اساسی که برای آنها به کار گرفته شده‌اند می‌پردازیم. در بخش ۲.۱۵ خواص طرحهای جایگزین را با جزئیات بیشتر بررسی می‌کنیم و کاراییهای نسبی چندین طرح را برای تهیه رهنمودهایی درباره انتخاب طرح مناسب یک مسئله داده شده مقایسه می‌کنیم. برآورد کردن با روشهای اصلی نمونه‌گیری تلفنی در بخش ۳.۱۵ مورد بررسی قرار گرفته است. فصل حاضر در بخش ۴.۱۵ با مقایسه طرحها براساس هزینه، واریانس، اجرا، و ملاحظات مربوط به اریبی به پایان می‌رسد.

۱.۱.۱۵ جامعه خانوار تلفنی

یکی از دلایل عمده‌ای که آمارگیریهای تلفنی برای اولین بار در آمارگیریهای خانوار مورد استفاده قرار گرفتند کاهش هزینه‌های جمع‌آوری داده‌ها بود. تماس تلفنی به مراتب کم‌هزینه‌تر از تماس رودررو است، و غالباً هم برای پژوهشگر و هم برای پاسخگو راحت‌تر است. از تلفن از سالهای دهه ۱۹۳۰ در آمریکا برای آمارگیریهای خانوار استفاده می‌شده است و حتی در همان زمان نیز معلوم شده بود که این آمارگیریها به زیرمجموعه‌ای از همه خانوارها محدود می‌شوند. این روش با افزایش نسبت خانوارهای دارای تلفن در دهه ۱۹۶۰ در میان سازمانهای آمارگیری بازرگانی متداول شد.

با وجود این که معلوم شده بود آمارگیریهای تلفنی به خانوارهای دارای تلفن محدود می‌شوند، نتایج آمارگیریهای تلفنی غالباً به صورتی ارائه می‌شدند، و هنوز هم می‌شوند، که گویی به سراسر جامعه همه خانوارها ارتباط دارند. اساس کاربست مزبور برپایه این فرض استوار است که بین خانوارهای با تلفن و بدون تلفن از لحاظ نوع مشخصه مورد مطالعه هیچ تفاوتی نیست. با این که ممکن است فرض مزبور درست باشد، مهم است که تا حد امکان مورد آزمون قرار بگیرد. آزمون تفاوتها می‌تواند به مشخصه‌هایی محدود شود که با مشخصه‌های مورد اندازه‌گیری در آمارگیری تنها همبستگی جزئی دارند. برای مثال، چون سن و جنس با نگرشها در ارتباطاند، پژوهشگران آمارگیری از نگرشهای اقتصادی ممکن است به همین رضایت دهند که ساکنان خانوارهای تلفن‌دار و بدون تلفن دارای توزیع سنی و جنسی مشابه‌اند.

تقریباً ۵ درصد خانوارهای آمریکا تلفن ندارند و درصد اشخاصی که در خانوارهای بدون تلفن زندگی می‌کنند تقریباً ۴ درصد است. با این که نرخ کلی عدم پوشش کوچک است و می‌تواند برای برخی پژوهشگران اطمینان‌بخش باشد، عدم پوشش، برحسب تعدادی از مشخصه‌ها که ممکن است با متغیرهای مورد اندازه‌گیری در یک آمارگیری وابسته باشند تا حدودی قابل ملاحظه متغیر است. برای

مثال، خانوارهای بدون تلفن گرایش به داشتن افراد جوانتر و پرتحرکتر دارند. گرایش دارند که با نسبتهای زیاد در مناطق روستایی نواحی جنوبی آمریکا و شهرهای مرکزی مستقر باشند. نرخهای عدم پوشش در بعضی زیرجامعهها ممکن است تا ۱۵٪ یا بیشتر افزایش یابد و این سطحی است که از نظر کسانی که باید از روی داده‌های آمارگیری برآوردهایی برای زیرگروههای کوچک جامعه آمریکا تهیه کنند غیر قابل قبول تلقی می‌شود.

مشخصه‌های خانوارهای بدون تلفن در مطالعاتی متعدد مورد بررسی قرار گرفته‌اند (برای مثال نگاه کنید به تورنبری و ماسی [۱۶]). خانوارهای بدون تلفن گرایش به داشتن نرخهای بیکاری بالاتر، نرخهای تدخین بالاتر، و نرخهای بالاتر تجربه قربانی جرایم را دارند. به این ترتیب، آمارگیریهای تلفنی می‌توانند برای اشتغال، بهداشت، یا مشخصه‌های اجتماعی، برآوردهای اریب تولید کنند.

برخی پژوهشگران تلاش کرده‌اند نمونه‌های تلفنی خانوار را از طریق شیوه‌های استاندارد وزن‌دار کردن آمارگیری تعدیل کنند. عدم پوشش در آمارگیریها غالباً با استفاده از وزنهای پس - طبقه‌بندی تعدیل می‌شود. در مورد نمونه‌های تلفنی، این تعدیلها ممکن است حقیقتاً قسمتی از اریبی عدم پوشش را جبران کنند. درباره استفاده از پس - طبقه‌بندی، یا تعدیل‌های کنترلی جامعه برای این منظور بعداً بحث خواهد شد.

۲.۱.۱۵ سیستمهای تلفن

در آمریکا، شماره تلفن‌ها برحسب نواحی جغرافیایی گروه‌بندی شده‌اند. شماره تلفنهای آمریکا از سه قسمت تشکیل شده‌اند: یک کد ناحیه‌ای سه رقمی، یک پیش شماره سه رقمی (یا کد اداره مرکزی)، و یک پس شماره چهار رقمی. کد ناحیه‌ای و پیش شماره به عنوان بخشی از یک سیستم بین‌المللی جا افتاده‌اند که در سراسر آمریکا، کانادا، مکزیک و منطقه کارائیب گسترش یافته است. البته این شمارهها در سراسر این ناحیه جغرافیایی تحت پوشش سیستم تلفنی مزبور، به طور تصادفی اختصاص نیافته‌اند. کدهای ناحیه‌ای به گونه‌ای به نواحی جغرافیایی مشخص اختصاص داده شده‌اند که حتی‌الامکان از مرزهای ایالتی تجاوز نمی‌کنند. ولی به هیچ صورت دیگری با مرزهای سیاسی از قبیل بخشها، شهرها، یا سایر تقسیمات شهری کوچکتر متناظر نیستند. تا این اواخر (به علت «پوشش» کدهای ناحیه‌ای) تناظری یک به یک بین کد ناحیه‌ای و ناحیه جغرافیایی برقرار بوده است.

پیش شمارهها در داخل کدهای ناحیه‌ای عموماً دارای تعریف جغرافیایی نیستند. علاوه بر این، یک پیش شماره ممکن است در بسیاری از کدهای ناحیه‌ای به کار رود. پیش شمارهها در داخل یک کد ناحیه‌ای در نواحی جغرافیایی موسوم به مراکز تلفن گروه‌بندی شده‌اند. مراکز تلفن (یا به عبارت مشخص‌تر از لحاظ اصطلاح‌شناسی سیستم تلفن، جایگاهها) نواحی جغرافیایی هستند که

برای مقاصد نگهداری و تأمین خدمات تلفن تعریف شده‌اند. مثلاً مرکز تلفن دیترویت در داخل کد ناحیه‌ای ۳۱۳ یک ناحیه جغرافیایی است که به طور تقریبی شامل شهر دیترویت و چند ناحیه اطراف آن است. مرکز تلفن دیترویت به بیش از ۱۲۵ پیش شماره مختلف اختصاص داده شده است.

خانوارها و مشاغل متقاضی خدمات تلفن در داخل ناحیه جغرافیایی که به عنوان مرکز تلفن دیترویت تعریف شده است باید شماره تلفنی دریافت کنند که پیش شماره آن از این مرکز تلفن خدمات‌رسانی شود. هیچ مرکز تلفن دیگری در داخل همین کد ناحیه‌ای نمی‌تواند از پیش شماره‌های تعیین شده برای مرکز تلفن دیترویت استفاده کند.

در داخل محدوده مرکز تلفن دیترویت از لحاظ پیش شماره‌ها کمی تفکیک جغرافیایی بیشتر وجود دارد. به یک خانوار که در هر نقطه‌ای در داخل محدوده این مرکز تلفن واقع شده باشد شماره تلفنی با هر پیش شماره که در مرکز موجود باشد اختصاص داده می‌شود. برخی مراکز تلفن که پیش شماره‌های زیادی دارند به تلفنخانه‌هایی تقسیم می‌شوند که هر یک مسئولیت یک بخش فرعی در داخل ناحیه تحت پوشش مرکز تلفن را به عهده دارد. به این ترتیب، در این مراکز تلفن، گروه‌بندی جغرافیایی اندکی، با توجه به پیش شماره‌ها در داخل مرکز تلفن وجود دارد. ولی این طرز تفکیک جغرافیایی پیش شماره‌ها در داخل یک مرکز تلفن نه یک قاعده، بلکه استثناست.

اکثریت مراکز تلفنی آمریکا دارای تعداد کمی از شماره تلفنهای تخصیص یافته هستند. در نتیجه، به بیشتر مراکز تلفن تنها یک پیش شماره اختصاص یافته است. تا این اواخر، مراکز تلفن مربوط به نواحی بودند که محدوده آنها توسط کمیسیونهای خدمات عمومی تعیین می‌شد و شرکتها می‌توانستند حقوق انحصاری برای تأمین خدمات تلفن به دست آورند. نیازهای خدماتی چنان‌اند که مساحت تحت پوشش یک مرکز تلفن محدود است. با وجود این، تراکم جمعیت در هر مرکز تلفن می‌تواند بسیار متفاوت باشد. به این ترتیب، برخی مراکز تلفن مشتریان بسیار کمی دارند و شماره تلفن برای ارائه به همه مشتریان با یک پیش شماره خاص به قدر کافی موجود است. برخی دیگر از مراکز تلفن مشتریان بسیار زیادی دارند و پیش شماره‌های متعددی را در اختیار گرفته‌اند.

پس شماره‌ها در مجموعه‌های ۱۰۰۰۰ تایی گروه‌بندی شده‌اند. این شماره‌ها نوعاً توسط کارکنان سرویس تلفن براساس الگوهای واگذاری موجود، در اختیار مشتریان قرار می‌گیرند. به نظر نمی‌رسد هیچ سیستم استاندارد تأمین درخواست مشتریان تازه برای شماره تلفنهای خانگی با یک پیش شماره خاص وجود داشته باشد.

بررسیهای تجربی، الگوهایی را برای تخصیص پس شماره به پیش شماره‌ها حداقل در مواردی که سراسر سیستم مورد بررسی بوده است ابداع نموده‌اند. در مراکز تلفن دارای پیش شماره‌های متعدد و

مشتریان زیاد به نظر می‌رسد که پیش شماره‌ها و پس شماره‌ها به طور اتفاقی ارائه شده‌اند. در داخل انبوه شماره‌ها، هیچ خوشه‌بندی آشکاری از شماره تلفنهای تخصیص یافته وجود ندارد. ولی در مراکز تلفنی که دارای تنها یک پیش شماره و تعداد مشتریان محدودند، پس شماره‌ها برای کاهش هزینه‌های تخصیص، از قدیم به صورت خوشه‌هایی خاص ارائه شده‌اند. تجهیزات تلفنی الکترومکانیکی قدیمتر این امکان را برای شرکتهای کوچکتر فراهم می‌ساختند که همه شماره‌ها را در یک «انبوه» ۱۰۰۰ تایی از شماره‌های متوالی اختصاص دهند به طوری که همه پس شماره‌های چهار رقمی با یک رقم شروع می‌شدند. شرکتی که جمعیت کمی را تحت پوشش می‌گرفت فقط باید یک انبوه ۱۰۰۰ تایی یا تعداد محدودی از آنها را خریداری می‌کرد تا تعداد مناسبی از شماره تلفن‌ها را برای مشتریان خود در اختیار داشته باشد. به این ترتیب، شماره تلفنهای موجود در مراکز تلفنی که دارای چند پیش شماره تکی هستند، در سطح انبوه ۱۰۰۰ تایی خوشه‌بندی شده‌اند. بررسیهای تجربی نشان داده‌اند که این «خوشه‌بندی» شماره‌های واگذار شده با کاهش اندازه انبوه افزایش می‌یابد. مثلاً، خوشه‌بندی انبوه‌هایی که با دو رقم اول پس شماره یا «در انبوه ۱۰۰ تایی» تعریف می‌شوند، فشرده‌تر است. بسیاری از روشهای نمونه‌گیری تلفنی که متعاقباً مورد بحث قرار می‌گیرند از این تخصیص خوشه‌بندی شده پیش شماره‌ها برای افزایش کارایی در شناسایی شماره تلفنهای واگذار شده به واحدهای مسکونی سود جسته‌اند.

۳.۱.۱۵ چارچوبهای نمونه‌گیری

چهار نوع مشکل چارچوب در نمونه‌گیریهای تلفنی وجود دارند. موارد فهرست شده در چارچوب که عناصری از جامعه محسوب نمی‌شوند به عنوان خالیها مورد اشاره قرار می‌گیرند، در حالی که عناصری از جامعه که فهرست‌بندی متناظری برای آنها وجود ندارد عناصر پوشش داده نشده نامیده می‌شوند. موارد فهرست شده در چارچوب که به عناصر چندگانه‌ای در جامعه اشاره دارند خوشه نامیده می‌شوند و عناصری از جامعه که دو یا چند بار در چارچوب، فهرست شده‌اند به نام فهرست‌برداریهایی تکراری خوانده می‌شوند.

هر یک از این نارساییها می‌تواند به آریبی در برآوردهای آمارگیری یا ناکارآمدی عملیات آمارگیری منجر شود. آمارشناسان نمونه‌گیری، شیوه‌هایی را برای انتخاب کردن ابداع کرده‌اند که آریبی ناشی از این نارساییها را کاهش می‌دهند یا از میان می‌برند. این آمارشناسان در یافتن شیوه‌هایی برای انتخاب کردن به گونه‌ای که این نارساییها کاهش یابد نیز مؤثر بوده‌اند.

برای نمونه‌گیری تلفنی از سه چارچوب اصلی استفاده می‌شود: شماره تلفن، کتابچه‌های راهنمای تلفن، و فهرستهای بازرگانی. چارچوب شماره تلفن‌ها را می‌توان از طریق دو منبع اولیه ایجاد کرد که

هر یک فهرستی از کدهای ناحیه‌ای و ترکیبهای از پیش شماره‌ها را با پس شماره‌هایی که به صورت تصادفی تهیه شده‌اند ترکیب می‌کند. ترکیبهای کد ناحیه‌ای و پیش شماره‌ها را می‌توان برای بررسیهای محلی از بررسی کتابچه‌های راهنمای تلفنهای محلی به دست آورد که معمولاً شامل فهرستهای روزآمد از پیش شماره‌های مربوط به مراکز تلفنی موجود در کتابچه راهنما هستند. برای آمارگیری‌هایی که نواحی جغرافیایی وسیعتری را پوشش می‌دهند می‌توان ترکیبهای کد ناحیه‌ای و پیش شماره‌ها را در آمریکا از یک مؤسسه بازرگانی به نام شرکت تحقیقاتی بل کور^۱ (BCR) تهیه کرد.

چارچوب BCR هر ماه بهنگام می‌شود و شامل همه ترکیبهای کد ناحیه‌ای و پیش شماره برای آمریکا، کانادا، مکزیک و منطقه کارائیب است. چارچوب BCR عملاً پوشش کاملی از خانوارهای دارای تلفن را در اختیار قرار می‌دهد ولی این عیب را دارد که تعداد قابل توجهی از موارد فهرست‌برداری شده آن خالی است. در واقع، کمتر از ۲۵٪ شماره‌های ارائه شده به واحدهای مسکونی اختصاص دارند. به همین علت، استفاده از طرحی مبتنی بر تهیه یک پس شماره تصادفی همراه با کد ناحیه‌ای و پیش شماره‌ای که از روی چارچوب BCR انتخاب شده باشد از نظر عملیاتی ناکارآمد است. روشهای دیگری براساس چارچوب BCR (به شرح زیر) بسط داده شده‌اند تا از خوشه‌بندی ذاتی شماره تلفنهای خانگی در انبوه‌های ۱۰۰۰ تایی برای کاهش نسبت خالیهای حاصل از این چارچوب استفاده شود.

چارچوب BCR عیب دیگری دارد که در دو چارچوب دیگر نیز مشترک است. خانوارهایی که از بیش از یک شماره تلفن برای مقاصد مسکونی استفاده می‌کنند چندین بار در چارچوب وارد می‌شوند. روشهای نمونه‌گیری احتمالاتی مستلزم آن است که تعداد شماره تلفنهای موجود در خانوار به دست آید و برای تهیه وزنی جبران‌کننده در برآورد کردن مورد استفاده قرار گیرد.

چارچوب دوم، یا کتابچه راهنمای تلفن، در سطح گسترده‌ای به عنوان چارچوب برای بررسیهای محلی مورد استفاده قرار گرفته است. کتابچه‌های راهنما فهرستی از شماره تلفن‌ها را در اختیار قرار می‌دهند که تهیه آن برای یک محله هزینه‌چندانی ندارد. روشهای نمونه‌گیری ساده از روی فهرست (مانند نمونه‌گیری سیستماتیک) را می‌توان برای انتخاب سریع ولی نه الزاماً ساده نمونه‌ها مورد استفاده قرار داد. جفت و جور کردن چارچوبهای کتابچه‌ای برای نواحی جغرافیایی وسیعتر مشکل است، زیرا هر سال بیش از ۵۰۰۰ کتابچه راهنمای تلفن در سراسر آمریکا منتشر می‌شود. متداول شدن آنها به عنوان چارچوب نمونه‌گیری به دلیل هزینه و راحتی کار و نسبت کم موارد فهرست‌برداری خالی در

^۱ Bell Core Research, Inc., (BCR)

مقایسه با چارچوب شماره تلفن است: تقریباً ۱۰ تا ۱۵ درصد از موارد فهرست شده در راهنمای تلفنهای خانگی آمریکا، دیگر خانگی نیستند.

از سوی دیگر، چارچوبهای کتابچه‌ای دچار عدم پوشش جمعیت خانوارهای دارای تلفن در سطحی قابل ملاحظه به علت شماره‌های فهرست نشده و تغییرات در وضعیت تلفنی خانوارها هستند. درصد خانوارهای تلفن‌داری که در کتابچه‌های راهنمای تلفن آمریکا دیده نمی‌شوند از ۳۵ درصد بیشتر است و از نسبتهای کم (۱۰ درصد) در حومه شهرها و نواحی روستایی تا نسبتهای بیشتر از ۶۰ درصد در برخی مکانهای شهری متغیر است. طراحان آمارگیریها نمی‌توانند این نسبتهای زیاد فهرست‌نشده‌ها یا موارد فهرست شده قدیمی را بپذیرند و استفاده از روشهای شماره‌گیری ارقام تصادفی یا سایر روشهایی را که پوشش بیشتری در اختیار قرار می‌دهند برمی‌گزینند.

علاوه بر این، سطوح مداخل تکراری در کتابچه‌های تلفن بیشتر از چارچوب BCR است، زیرا مشترکین می‌توانند مداخل اضافی خریداری کنند. مثلاً، زن و شوهری که با دو نام خانوادگی مختلف در یک آدرس به سر می‌برند می‌توانند با پرداخت دستمزدی ناچیز تحت نام خانوادگی هر دو نفر در کتابچه راهنمای تلفن درج شوند. تکرار موجود در فهرست باعث افزایش احتمال انتخاب شدن آن خانوار تلفن‌دار می‌شود که از نظر نمونه‌گیری احتمالاتی باید از طریق تعیین وزنی برای خانوار جبران شود.

چارچوب نوع سوم توسط شرکتهای تجاری در پرونده‌های الکترونیکی براساس کتابچه‌های راهنمای اطلاعات تلفن جمع‌آوری شده در سراسر کشور، تهیه می‌شود. درایه‌های مندرج در کتابچه راهنما (نام، آدرس، و شماره تلفن) یا وارد رایانه می‌شوند یا در صورت وجود چارچوب الکترونیکی به آن افزوده می‌شوند. فهرستهای مبتنی بر راهنماهای تجاری با سایر فهرستها تکمیل می‌شوند که بزرگترین آنها فهرستهای ثبتي اتومبیلهاست که از تقریباً ۳۰ ایالتی که این قبیل داده‌ها را برای عموم منتشر می‌کنند به دست می‌آیند. پرونده ترکیبی پردازش می‌شود تا برای مقاصد پستی به هر درایه یک کد پستی تخصیص داده شود. چندین شرکت از وجود شماره تلفن در این قبیل پرونده‌ها برای ایجاد راهنماهای تلفن در سطح کشور و انتخاب و فروش نمونه‌هایی از شماره تلفن‌ها از روی آنها بهره برده‌اند. چارچوبهای تجاری درگیر نسبت کمی از اقلام خالی (تقریباً ۱۰ تا ۱۵ درصد، یعنی برابر با کتابچه‌های راهنما) و عدم توانایی در پوشش شماره‌های فهرست نشده و نیز وجود اقلام تکراری‌اند.

۲.۱۵ طرح‌های نمونه‌ای تلفنی

آمارشناسان نمونه‌گیری روشهایی را برای نمونه‌گیری ابداع کرده‌اند که سعی در بهره‌گیری از توانمندیهای هر یک از این سه چارچوب دارند. روشهای مزبور غالباً به صورت یکی از این سه نوع،

رده‌بندی می‌شوند: روشهای نمونه‌گیری با چارچوب فهرست ساده که برای کتابچه‌های راهنمای تلفن مناسب‌اند، روشهای شماره‌گیری ارقام تصادفی مبتنی بر چارچوب شماره تلفن، و روشهایی با کمک فهرست مبتنی بر کتابچه‌های راهنمای تلفن یا فهرستهای تجاری که نمونه‌هایی تولید می‌کنند که شامل شماره تلفنهای فهرست نشده نیز هستند. در اینجا درباره روشهای نمونه‌گیری مورد استفاده برای کتابچه‌های راهنما بحث نمی‌کنیم. این روشها اصولاً کاربرد روشهایی هستند که قبلاً در فصلهای پیشین کتاب حاضر مورد بحث قرار گرفته‌اند. در عوض به بررسی شماره‌گیری ارقام تصادفی و روشهای با کمک فهرست می‌پردازیم، زیرا کاربردهایی تازه از روشهایی محسوب می‌شوند که در جاهای دیگر این کتاب توصیف شده‌اند.

۱.۲.۱۵ طرحهای نمونه‌ای با استفاده از چارچوب BCR

همان‌گونه که قبلاً توصیف شد، ساده‌ترین و مستقیم‌ترین رهیافت برای استفاده از چارچوب BCR، انتخاب تصادفی شماره تلفنها از روی چارچوب، تلفن کردن به شماره‌های انتخاب شده، و اجرای مصاحبه در صورت رسیدن به یک خانوار است. شماره‌ها انتخاب و گرفته می‌شوند تا به اندازه نمونه مطلوب خانوارهای موجود در آن حوزه، مثلاً n ، برسیم. به طوری که پیش از این اشاره شد تنها در حدود ۲۵٪ شماره تلفنهای نمونه به خانوارها تخصیص یافته‌اند. بنابراین، شماره تلفنهای مورد نیاز، مثلاً n' ، بسیار بیشتر از n خواهد بود. تعداد مورد انتظار شماره تلفنهای مورد نیاز n/p است که در آن p نسبت شماره تلفنهایی است که به واحدهای مسکونی خانوار اختصاص یافته‌اند. پس برای جبران موارد فهرست‌برداری شده که واجد شرایط نیستند باید نمونه شماره تلفنهای انتخاب شده از چارچوب BCR تقریباً چهار برابر نمونه مطلوب متشکل از n خانوار تلفن‌دار باشد.

به طور کلی، تعیین وضعیت یک شماره تلفن مستلزم صرف هزینه است به خصوص در مورد شماره تلفنهایی که در اختیار خانوارها نیستند. غالباً باید برای تعیین وضعیت، یک شماره تلفن چندین بار شماره‌گیری شود. چون برای هر نوع نتیجه حاصل از شماره‌گیری باید شیوه‌ها مشخص شوند، استفاده از BCR (یا هر فهرستی که دارای نسبت زیادی از اقلام خالی است) هزینه‌های اداری و عملیاتی آمارگیری تلفنی را به شدت افزایش خواهد داد. (ن. ک. لیکوفسکی [۱۰]).

خلاصه کردن هزینه‌های عملیاتی استفاده از هر چارچوب، برای مقایسه طرحهای گوناگون، مفید خواهد بود. هزینه‌های استفاده از BCR را می‌توان در یک مدل هزینه‌ای ساده به صورت زیر خلاصه کرد. فرض کنید C_0 هزینه تعیین وضعیت شماره‌ای باشد که به خانواری در محدوده مورد نظر (خانواری که تلفن دارد) واگذار نشده باشد؛ C_1 نشان‌دهنده هزینه تعیین وضعیت شماره‌ای در اختیار

خانوار در محدوده مورد نظر باشد؛ و C_p نشان‌دهنده هزینه اجرای مصاحبه برای آمارگیری باشد. پس کل هزینه آمارگیری براساس چارچوب BCR از این فرمول به دست می‌آید

$$C = n(C_1 + C_p) + (n' - n)C.$$

با منظور کردن نسبت شماره تلفنهایی که در محدوده قرار دارند، کل هزینه مورد انتظار برای یک آمارگیری با شماره‌گیری ارقام به صورت تصادفی ساده از فرمول زیر به دست می‌آید

$$E(C) = n((C_1 + C_p) + C_1(1 - p)/p)$$

با در نظر گرفتن p در همسایگی $0/25$ ، مؤلفه هزینه مورد انتظار ناشی از تلفن زنده‌های بی‌حاصل [یعنی $nC_1(1 - p)/p$]، نسبت قابل ملاحظه‌ای از کل هزینه مورد انتظار C را تشکیل خواهد داد. انگیزه همه طرح‌های تلفنی که در بخشهای زیر توصیف شده‌اند اشتیاق به کاهش نسبت هزینه‌های ناشی از تلفن زنده‌های بی‌حاصل بوده است.

۱.۱.۲.۱۵ طرح میتوفسکی - واکسبرگ

طرح رقم تصادفی دومرحله‌ای که میتوفسکی [۱۱] پیشنهاد کرد و توسط واکسبرگ [۱۸] شرح و بسط کاملتری داده شد در چنان سطح گسترده‌ای در آمارگیریهای تلفنی به کار برده شده است که تقریباً با آمارگیریهای تلفنی با شماره‌گیری ارقام تصادفی مترادف شده است. روش مذکور از این واقعیت بهره می‌برد که شماره تلفنهای واگذار شده به خانوارهای ساکن، گرایش به خوشه‌بندی در انبوهه‌هایی از شماره تلفنهای متوالی دارند. در حالی که در مجموع فقط ۲۵ درصد از شماره‌های موجود در چارچوب BCR در اختیار خانوارها هستند، در میان انبوهه‌های ۱۰۰ تایی از شماره‌های متوالی که حداقل یک شماره تلفن آن مربوط به یک خانوار باشد ۶۰ درصد شماره تلفنهای خانگی هستند. واضح است که اگر بتوان انبوهه‌های صدتایی با یک یا چند شماره تلفن خانگی را شناسایی کرد و نیز اگر نمونه‌گیری به همین انبوهه‌ها محدود باشد، در آن صورت، نسبت تلفن زنده‌های بی‌حاصل به طور قابل ملاحظه‌ای کاهش خواهد یافت و هزینه‌های عملیاتی در مجموع پایین خواهد آمد.

فن میتوفسکی - واکسبرگ با گروه‌بندی شماره‌های موجود در چارچوب BCR در مجموعه‌های ۱۰۰ تایی از شماره‌های متوالی با استفاده از کد ناحیه‌ای، پیش شماره سه رقمی و دو رقم اول پس شماره برای تعیین انبوهه (یا انبوهه ۱۰۰ تایی) آغاز می‌شود. در مرحله اول، انبوهه‌های ۱۰۰ تایی به صورت تصادفی با جایگذاری انتخاب می‌شوند. یک شماره تلفن در داخل انبوهه به طور تصادفی انتخاب و شماره‌گیری می‌شود. اگر معلوم شود که شماره انتخاب شده خانگی است انبوهه مزبور برای نمونه‌گیری مرحله دوم حفظ می‌شود. این فرایند ادامه می‌یابد تا نمونه تعیین شده متشکل از m انبوهه

۱۰۰ تایی به دست آید. شماره تلفن‌ها در داخل هر انبوهه ۱۰۰ تایی حفظ شده، به صورت تصادفی بدون جایگذاری انتخاب می‌شوند تا مجموع k شماره تلفن خانگی (شامل اولین شماره‌ای که برای حفظ انبوهه به کار رفته است) شناسایی شوند. وضعیت خانگی بودن یک شماره تلفن فقط با گرفتن آن شماره تعیین می‌شود.

به این ترتیب، در فن میتوفسکی - واکسبرگ از طرح دو مرحله‌ای استفاده می‌شود که در آن، انبوهه‌های ۱۰۰ تایی با احتمال متناسب با تعداد شماره تلفن‌های خانگی در مرحله اول و نمونه‌ای با اندازه ثابت در مرحله دوم از خانوارهای ساکن این اماکن انتخاب می‌شوند. به این ترتیب، نمونه $n = mk$ خانوار ساکن با احتمال برابر (ولی نامعلوم) انتخاب می‌شود. کارایی فن میتوفسکی - واکسبرگ از این واقعیت ناشی می‌شود که شماره تلفن‌های واجد شرایط در نسبت تقریباً کوچکی از انبوهه‌های ۱۰۰ تایی متمرکز شده‌اند.

برای تعیین هزینه استفاده از این طرح دیگر نمونه‌ای برای چارچوب BCR، فرض می‌کنیم t نسبت انبوهه‌های ۱۰۰ تایی بدون شماره تلفن‌های واجد شرایط باشد. در آن صورت، کل تعداد مورد انتظار شماره تلفن‌ها در سراسر مراحل اول و دوم به صورت $n(1-t(k-1)/k)/p$ است و کل هزینه عملیاتی مورد انتظار به صورت زیر خواهد بود:

$$E(C) = n((C_v + C_f) + C_s(1-p-t(k-1)/k)/p)$$

واضح است که هم تعداد تلفن زدن‌های مورد انتظار و هم هزینه مورد انتظار با افزایش k کاهش می‌یابد. t در سطح کل کشور در همسایگی ۰/۶۵ است، بنابراین حتی مقادیر نه چندان زیاد k می‌تواند به صرفه‌جویی‌های قابل ملاحظه در هزینه‌ها بینجامد.

با این که فن میتوفسکی - واکسبرگ روشی ظریف برای بهبود کارایی آمارگیری تلفنی است، در عمل مشکلاتی وجود دارند. واضحترین آنها این است که برخی از انبوهه‌های صدتایی ممکن است کمتر از مقدار مورد نیاز k خانوار واجد شرایط داشته باشند که در این حالت به اجبار باید به همه شماره‌های موجود در انبوهه تلفن شود. در این صورت وزن دادن برای جبران آن ضروری خواهد بود. مشکل دوم آن است که همیشه این امکان وجود ندارد که وضعیت واجد شرایط بودن یک شماره انتخابی با دقت تعیین شود. در مرحله اول، چنین حالتی می‌تواند به حذف یا گنجاندن نادرست انبوهه‌هایی ۱۰۰ تایی منجر شود. در مرحله دوم ممکن است در پایان دوره آمارگیری هنوز تکلیف بعضی شماره‌ها روشن نشده باشد به طوری که کمتر از k خانوار واجد شرایط در انبوهه شناسایی شده باشند. بالاخره این که مشکل پیچیده‌تر، موضوع همبستگی درون انبوهه‌ای است که با جزئیات بیشتر در بخش ۳.۱۵، هنگام بررسی برآورد کردن مورد بحث قرار خواهد گرفت.

۲.۱.۲.۱۵ طرح پوتوف

طرح پیشنهادی پوتوف [۱۲] شبیه طرح میتوفسکی - واکسبرگ است با این تفاوت که واجد شرایط بودنش به ردهٔ وسیعتری از شماره تلفن‌ها گسترش یافته است. پوتوف پیشنهاد کرده است که به جای تعیین واجد شرایط بودن یک انبوههٔ ۱۰۰ تایی بر مبنای این که آیا شمارهٔ تلفن به واحد مسکونی اختصاص دارد یا نه، بهتر است واجد شرایط بودن وضعیت آن شماره تلفن، پس از شماره‌گیری تعریف شود. او شماره تلفنهایی را که تحت برآمدهای گوناگون واجد شرایط بوده‌اند شماره‌های خوش‌یمن نامیده است. شماره‌های خوش‌یمن نوعاً نه تنها شماره تلفنهای خانگی خانوارها را در برمی‌گیرند بلکه شماره‌هایی را هم شامل می‌شوند که بدون جواب زنگ می‌خورند و وضعیت خانگی بودن آنها پس از شماره‌گیری نیز نامعلوم است. این تعریف گسترده‌تر، از مقدار غربالگری مورد نیاز برای مرحلهٔ اول و مقدار جایگذاری مورد نیاز در مرحلهٔ دوم می‌کاهد. پوتوف [۱۲]، [۱۳] همچنین پیشنهاد کرده است که در مرحلهٔ اول در هر انبوههٔ ۱۰۰ تایی، $c \geq 2$ شماره تلفن انتخاب شود. در این صورت، نمونه‌گیری در مرحلهٔ دوم به تعداد شماره‌های خوش‌یمن مشاهده شده در مرحلهٔ اول بستگی خواهد داشت. از جزئیات روش پوتوف در این کتاب به تفصیل بحث نخواهد شد.

طرح نمونه‌گیری پوتوف نمونه‌ای از شماره‌های واجد شرایط با احتمال برابر به دست می‌دهد. جایگذاری فقط برای تعداد کمی از پیش شماره‌های ناحیه‌ای مورد نیاز است و ابهامهای مربوط به وضعیت شماره تلفنهای شماره‌گیری شده در مرحلهٔ اول را کاهش می‌دهد. همچنین، وقتی c ، تعداد شماره تلفنهای انتخاب شده به ازای هر انبوههٔ ۱۰۰ تایی در مرحلهٔ اول زیاد می‌شود، شانسهای به دست آوردن انبوهه‌ای که در مرحلهٔ دوم خالی شود کاهش پیدا می‌کند. اجرای طرح پوتوف برای تعیین اندازهٔ نمونه‌ای مناسب، مستلزم شناختی دربارهٔ نسبت شماره‌های خوش‌یمنی است که واقعاً واجد شرایطاند. ساختار اجرایی این روش، پیچیده‌تر و نیازهای آموزشی آن برای کارکنان عملیاتی به نسبت طرح میتوفسکی - واکسبرگ بیشتر است.

۲.۲.۱۵ طرح‌های نمونه‌ای که از شماره تلفنهای خانگی منتشر شده استفاده می‌کنند

همان طور که در بالا توصیف شد ۸۵ تا ۹۰ درصد شماره تلفنهای موجود در فهرستهای تجاری به خانوارهای ساکن در اماکن مسکونی مربوط می‌شوند. انتخاب سیستماتیک سر راست شماره تلفن (انتخاب تصادفی فقط در صورتی امکان‌پذیر است که فهرست به ترتیب شماره‌گذاری شده باشد) از روی چنین فهرستی به مراتب کارآمدتر از طرحهایی خواهد بود که برای نمونه‌گیری از روی فهرست BCR مورد استفاده قرار می‌گیرند. متأسفانه فهرست نوعاً مبتنی بر راهنمای تلفن، به علت منسوخ بودن شماره‌های فهرست شده و فهرست نشده فقط شامل تقریباً ۷۰ درصد خانوارهای دارای تلفن خانگی

است. مقایسه شماره تلفنهای منتشر شده و منتشر نشده خانوارهای تلفن‌دار نشان می‌دهد که اگر خانوارهایی که شماره تلفنهای آنها منتشر نشده است از چارچوب نمونه‌گیری حذف شوند اریبی قابل ملاحظه‌ای نتیجه خواهد شد (برانر و برانر [۲]). طرحهایی که در این بخش مورد بحث قرار گرفتند می‌کوشند تا از کارایی ذاتی نمونه‌گیری بر مبنای راهنمای تلفن بهره‌مند شوند و در عین حال پوشش طرح را گسترش دهند تا کل جمعیت دارای تلفن خانگی را شامل شوند.

۱.۲.۲.۱۵ شماره‌گیری رقم به اضافه یک

شماره‌گیری رقم به اضافه یک، شیوه‌ای با استفاده از راهنمای تلفن است که در آن نمونه‌ای از شماره تلفن‌ها از روی دفترچه راهنما انتخاب و به پس شماره تلفن عدد صحیح یک اضافه می‌شود. مثلاً به پس شماره هر شماره تلفنی که از راهنمای تلفن انتخاب شود در این شماره‌گیری به پس شماره «یک» اضافه می‌شود. به این ترتیب اگر شماره انتخابی ۰۰۲۱-۹۳۶-۷۳۴ باشد جایگزین آن، شماره ۰۰۲۲-۹۳۶-۷۳۴ خواهد بود. نمونه شماره تلفن‌ها که به این ترتیب به دست می‌آید به طور کلی شامل شماره‌های فهرست شده و فهرست نشده، هر دو، خواهد بود. علاوه بر این، در مقایسه با طرح شماره‌گیری رقم تصادفی ساده نسبت بیشتری از شماره‌های نتیجه‌بخش به دست خواهد داد.

متأسفانه این شیوه دارای چند مشکل نظری و عملیاتی است. بررسیهای تجربی مزبور نشان داده‌اند که انتخاب خانوارهای دارای تلفن خانگی در روش شماره‌گیری به اضافه یک دارای اریبی قابل توجهی است. به علاوه احتمالهای انتخاب شدن شماره‌های موجود در جامعه هدف، نابرابر و نامعلوم‌اند. در واقع، احتمال انتخاب شدن برخی شماره‌های فهرست نشده ممکن است صفر باشد مگر این که شماره‌های فهرست نشده به صورت یکنواخت با شماره‌های فهرست شده آمیخته باشند. تعمیمهای طرح مذکور که در آن رقم آخر (دو یا چند رقم آخر) با d رقم دیگر که به طور تصادفی به دست آمده‌اند جایگزین می‌شوند پیشنهاد شده است.

طرحی که در ارتباط نزدیک با طرح بالاست براساس بازه‌های نیمه باز شماره تلفن‌ها توسط فرانکل و فرانکل [۵] پیشنهاد شده است. در راهنماهایی که به ترتیب شماره تنظیم شده‌اند، خوشه، شامل یک شماره تلفن فهرست شده همراه با همه شماره‌های دیگر تا شماره فهرست شده بعدی ولی بدون منظور نمودن آن تعریف شده است. نمونه‌ای از خوشه‌ها از روی راهنمای تلفن، صرفاً با انتخاب یک نمونه تصادفی ساده از شماره تلفنهای موجود در راهنما، گرفته می‌شود. سپس همه شماره‌های موجود در داخل خوشه انتخابی شماره‌گیری می‌شوند. این روش به احتمالهای معلوم غیرصفر، برای انتخاب شدن همه خانوارهای تلفنی تحقق می‌بخشد. ولی تغییرپذیری بالقوه زیاد اندازه خوشه می‌تواند مشکلات عملیاتی دشوار را مطرح سازد. اندازه نمونه می‌تواند نسبت به یک سطح هدف به شدت تغییر کند. به

علاوه، این روش در معرض مشکلات برآورد کردن قرار دارد چرا که اندازه نمونه و خوشه هر دو متغیرهای تصادفی‌اند.

۲.۲.۲.۱۵ نمونه‌گیری دومرحله‌ای

یک طرح نمونه‌گیری دومرحله‌ای با استفاده از فهرست راهنمای تلفن توسط سودمن [۱۵] پیشنهاد شده است. در این شیوه که در اصل توسط استاک [۱۴] پیشنهاد شده بود از انبوه‌های ۱۰۰۰ تایی شماره تلفن‌ها (که با شش رقم اول شماره تلفن ۱۰ رقمی مشخص می‌شوند) به عنوان واحد نمونه‌گیری مرحله اول استفاده می‌شود. انتخاب انبوه‌های ۱۰۰۰ تایی شبیه انتخاب مرحله اول در روش میتوفسکی - واکسبرگ است با این تفاوت که برای انتخاب نمونه مرحله اول از راهنمای شماره تلفن‌های فهرست شده استفاده می‌شود. به این ترتیب، احتمال انتخاب شدن در مرحله اول متناسب با تعداد شماره تلفن‌های فهرست شده در انبوه ۱۰۰۰ تایی است. در مرحله دوم، انتخاب شماره تلفن‌ها ادامه می‌یابد تا تعداد ثابتی از شماره تلفن‌های فهرست شده که از پیش تعیین شده است انتخاب شود. سعی بر آن است که خانوارهای دارای شماره تلفن‌های فهرست شده و فهرست نشده هر دو مورد مصاحبه قرار بگیرند. باید توجه داشت که احتمال انتخاب شدن شماره تلفن‌های فهرست نشده در انبوه‌های ۱۰۰۰ تایی که فاقد هر نوع شماره فهرست شده‌اند صفر است ولی نسبت خانوارهای تلفن‌دار در این قبیل انبوه‌های ۱۰۰۰ تایی کوچک است و این در بیشتر موارد مشکلی جدی محسوب نمی‌شود. نگران کننده‌تر، این واقعیت است که تعیین وضعیت فهرست شدن، به خود - گزارش پاسخگو بستگی دارد که می‌تواند خطا باشد. یک «راهنمای شماره تلفن‌های به ترتیب معکوس» که در بسیاری از نواحی کلانشهری موجود است می‌تواند این منشأ خطا را از میان بردارد.

شیوه سودمن برخلاف روش میتوفسکی - واکسبرگ، خوشه‌هایی با اندازه نابرابر برای خانوارهای تلفن‌دار تولید خواهد کرد (هر چند خوشه‌هایی که برای خانوارهای تلفن‌دار فهرست شده تولید می‌کند دارای اندازه برابرند). معمولاً تغییرپذیری اندازه خوشه چندان زیاد نیست. همچنین، وجود خوشه‌های خالی محتمل است ولی با انبوه‌های ۱۰۰۰ تایی به جای ۱۰۰ تایی در روش میتوفسکی - واکسبرگ این امر چندان جای نگرانی ندارد.

۳.۲.۱۵ طرح‌هایی که از چارچوب BCR و شماره تلفن‌های منتشر شده استفاده می‌کنند

باید توجه داشت که طرح‌های مورد بحث در بخش ۱.۲.۱۵ تنها به چارچوب BCR نیاز دارند در حالی که طرح‌های مورد بحث در بخش ۲.۲.۱۵ تنها به فهرستی منتشر شده از شماره تلفن‌های خانگی نیازمندند. طرح‌هایی که در بخش حاضر مورد بحث قرار می‌گیرند به هر دو نیاز دارند. ایده اصلی

زیربنایی این طرحها آن است که خواص پوششی مطلوب چارچوب BCR، با کارایی نسبتاً زیاد نمونه‌گیری از چارچوب متشکل از شماره تلفنهای فهرست شده، پیوند داده شود.

۱.۳.۲.۱۵ طرحهایی با چارچوب دوگان

یک نمونه شماره‌گیری رقم تصادفی متشکل از n_B خانوار تلفنی از روی چارچوب BCR انتخاب می‌شود و همزمان نمونه‌ای متشکل از n_D خانوار تلفنی از چارچوب فهرست راهنمای تلفن انتخاب می‌شود. با فرض این که n'_B و n'_D به ترتیب تعداد شماره تلفنهای مورد نیاز از هر چارچوب برای به دست آوردن اندازه نمونه مطلوب باشند، هزینه طرح با چارچوب دوگان از فرمول زیر به دست می‌آید

$$C = (n_B + n_D)(C_1 + C_r) + C_2(n'_B + n'_D - n_B - n_D)$$

هزینه مورد انتظار برای آمارگیری با چارچوب دوگان به صورت زیر است

$$E(C) = n(C_1 + C_r + C_2(\lambda(1-p_B)/p_B + (1-\lambda)(1-p_D)/p_D))$$

که در آن $n = n_B + n_D$ کل اندازه نمونه، $\lambda = n_B/n$ نسبت کل نمونه تعیین شده برای چارچوب BCR، p_B نسبت شماره تلفنها در چارچوب BCR که به خانوارهای دارای تلفن خانگی واگذار شده‌اند، و p_D نسبت شماره تلفنها در چارچوب راهنمای تلفن که به خانوارهای دارای تلفن خانگی واگذار شده‌اند. چون p_B در همسایگی 0.25 و p_D معمولاً در حدود 0.85 است، هزینه مورد انتظار (برای کل اندازه نمونه‌ای ثابت n) با کاهش λ کم خواهد شد.

برای ترکیب داده‌های حاصل از دو چارچوب به منظور برآورد کردن، راههای ممکن متعددی وجود دارند. به طور کلی، برآوردهای چارچوب دوگان پیچیده‌تر از برآوردهای مربوط به طرحهای مورد بحث پیشین‌اند. گروز و لیکوفسکی [۶] بحث مفصلی درباره مسئله برآورد کردن از چارچوب دوگان و مسئله انتساب نمونه به دو چارچوب برای تحقق کمترین هزینه به ازای یک واریانس توصیف شده ارائه کرده‌اند.

برای اجرای روش چارچوب دوگان، وضعیت راهنمای تلفنی (یعنی فهرست شده یا فهرست نشده) هر خانواری که تلفن خانگی دارد باید روی نمونه BCR معلوم باشد. برای اجتناب از به کار گرفتن گزارشهای بالقوه غیر قابل اعتماد پاسخگویان درباره وضعیت در فهرست بودن آنها، می‌توان شماره تلفنهای انتخاب شده از روی چارچوب BCR را با فهرست راهنمای تلفن در هنگام انتخاب نمونه جور کرد. اگر چارچوب راهنمای تلفن، آدرس شماره‌های فهرست شده را داشته باشد این امکان

وجود خواهد داشت که پیشاپیش نامه‌هایی به منظور بهبود نرخ پاسخگویی ارسال شود. به طور کلی، طرح چارچوب دوگان مستلزم عملیات اجرایی نسبتاً پیچیده‌تری است. نیاز به جور کردن نمونه BCR با چارچوب راهنمای تلفن برای تعیین شماره‌های واجد شرایط انتخاب از هر دو چارچوب، ارسال نامه‌های پیشاپیش، و استفاده از برآوردگر پیچیده‌تر باعث افزایش هزینه‌ها می‌شوند. مزایای بیشتر بودن نرخ پاسخگویی به مراتب بیشتر از جبران کردن هزینه‌های ارسال نامه‌های پیشاپیش است.

۲.۳.۲.۱۵ طبقه‌بندی مبتنی بر راهنمای تلفن

از فهرست راهنمای تلفن می‌توان به منظور طبقه‌بندی چارچوب BCR برای بهبود کارایی نمونه‌گیری از روی چارچوب BCR استفاده کرد. در یک کاربرد نوعی، از فهرست راهنمای تلفن برای شناسایی انبوهه‌های ۱۰۰ تایی در چارچوب BCR که دارای یک یا چند شماره تلفن فهرست شده در راهنمای تلفن هستند استفاده می‌شود. سپس چارچوب BCR به دو طبقه افراز می‌شود: یک طبقه شامل همه شماره تلفن‌ها در انبوهه‌های ۱۰۰ تایی است که یک یا چند شماره تلفن فهرست شده دارند و طبقه دیگر شامل همه شماره تلفن‌های دیگر است. به طبقه اول به عنوان طبقه با چگالی بالا اشاره می‌شود در حالی که طبقه دوم را طبقه باقیمانده می‌نامند. سپس نمونه‌های شماره‌گیری با رقم تصادفی ساده از هر طبقه انتخاب می‌شود به طوری که از طبقه با چگالی بالا، نمونه‌ای به مراتب بزرگتر انتخاب می‌شود زیرا تقریباً همه شماره تلفن‌های خانگی در این طبقه قرار دارند.

در طرح طبقه‌بندی شده سعی بر این است که از همان جنبه خوشه‌بندی چارچوب شماره تلفن‌ها در طرح میتوفسکی - واکسبرگ استفاده شود. شماره تلفن‌های خانوارهای دارای تلفن خانگی، اعم از فهرست شده و فهرست نشده، گرایش به خوشه‌بندی در انبوهه‌های ۱۰۰ تایی دارند. اگر انبوهه‌های محتوی این قبیل شماره تلفن‌ها را بتوان به میزان بیشتری شناسایی و نمونه‌گیری کرد، آنگاه کارایی نمونه‌گیری می‌تواند افزایش بسیار زیادی پیدا کند. کاسادی و لپکوفسکی [۳] پی برده‌اند که در سطح کل کشور، نسبت شماره تلفن‌های موجود در چارچوب BCR که به طبقه با چگالی بالا اختصاص می‌یابند تقریباً ۰/۳۸ خواهد بود و بیش از ۹۷٪ شماره‌های واگذار شده به خانوارهای دارای تلفن خانگی را شامل خواهد شد. آنها نشان داده‌اند که نسبت شماره تلفن‌های تخصیص یافته به خانوارها در طبقه با چگالی بالا تقریباً ۰/۵۵ است در حالی که نسبت شماره تلفن‌های تخصیص یافته به خانوارها در طبقه باقیمانده فقط حدود ۰/۰۲ است.

با فرض این که یک نمونه شماره‌گیری ارقام تصادفی متشکل از n_1 خانوار تلفن‌دار از طبقه با چگالی بالا و یک نمونه متشکل از n_2 خانوار تلفن‌دار از طبقه باقیمانده انتخاب شده باشد، هزینه طرح طبقه‌بندی از فرمول زیر به دست می‌آید

$$C = (n_1 + n_2)(C_1 + C_2) + C_3(n'_1 + n'_2 - n_1 - n_2)$$

که در آن n'_1 و n'_2 تعداد شماره تلفنهای مربوط به هر طبقه برای دستیابی به اندازه‌های نمونه‌ای مطلوب است. هزینه مورد انتظار نمونه طبقه‌بندی شده چنین است

$$E(C) = n(C_1 + C_2 + C_3(\gamma(1-p_1)/p_1 + (1-\gamma)(1-p_2)/p_2))$$

که در آن n ، کل اندازه نمونه، $\gamma = n_1/n$ ، نسبت کل نمونه تخصیص یافته به طبقه با چگالی بالا، p_1 ، نسبت شماره تلفنهای واگذار شده به خانوارهای دارای تلفن خانگی در طبقه با چگالی بالا، و p_2 نسبت شماره تلفنهای واگذار شده به خانوارهای دارای تلفن خانگی در طبقه باقیمانده است. چون p_1 تقریباً $0/55$ و p_2 معمولاً در حدود $0/02$ است، هزینه مورد انتظار (برای n ، کل اندازه ثابت نمونه) با افزایش γ کاهش خواهد یافت. انتساب نمونه به طبقات برای به حداقل رساندن هزینه مربوط به یک واریانس ثابت (یا به حداقل رساندن واریانس برای یک هزینه ثابت) توسط کاسادی و لیکوفسکی [۳] به تفصیل مورد بحث قرار گرفته است.

احتمالهای انتخاب شدن در داخل طبقات، معلوم، مثبت، و برابرند، بنابراین، برآورد کردن مجموعهای جامعه‌ای سراسر است. برآورد کردن میانگین جامعه‌ای در سطح طبقه نیز سراسر است، ولی برآورد کردن میانگین کل جامعه مستلزم آن است که کل جامعه دارای تلفن خانگی، برای هر طبقه برآورد شود و یک برآوردگر نسبتی برای برآورد میانگین جامعه‌ای مورد استفاده قرار گیرد. بحث مفصلتر درباره برآورد میانگینها و واریانسها در بخش ۳.۱۵ ارائه خواهد شد.

تحت مدل هزینه‌ای نسبتاً ساده‌ای که در بالا ارائه شد، طرح مزبور به طرز مناسب با طرح میتوفسکی - واکسبرگ مطابقت می‌کند. در کاربست، ثابت شده است که طبقه‌بندی مبتنی بر راهنمای تلفن در نمونه‌گیری شماره‌گیری رقم تصادفی ساده در داخل طبقات از جهت اجرا و مدیریت دارای برتری است. این طرح با دو هزینه همراه است که در مدل ساده وجود ندارند: هزینه خود فهرست تجاری و هزینه طبقه‌بندی چارچوب BCR. هزینه فهرست تجاری برحسب فروشنده فرق می‌کند ولی برای عملیات آمارگیری بزرگ مداوم، مؤلفه هزینه‌ای نسبتاً ناچیزی محسوب می‌شود. این هزینه‌ها هر دو ثابت‌اند و می‌توانند در چندین بررسی تقسیط شوند تا تأثیر آنها بر هر مطالعه تکی تا حد بسیار زیادی کاهش یابد.

۳.۳.۲.۱۵ برش دادن براساس راهنمای تلفن

یک مورد خاص در روش طبقه‌بندی شده که در آخرین بخش مورد بحث قرار گرفت هیچ نمونه‌ای را به طبقه باقیمانده منتسب نمی‌کند؛ یعنی این که چارچوب BCR با برداشتن طبقه باقیمانده بریده می‌شود. نرخ دستیابی به هدف که به شدت افزایش یافته است همراه با سایر مزایای طرح طبقه‌بندی مبتنی بر راهنمای تلفن، جذابیت بسیار زیادی به این رهیافت بخشیده‌اند. از معایب آشکار آن این است که وقتی چارچوب بریده شده باشد همه جامعه هدف پوشش داده نمی‌شود. در مثالی که در بالا ارائه شد، تقریباً ۲٪ خانوارهای تلفن‌دار پوشش داده نخواهند شد. ولی، شواهد تجربی (بریک و همکاران [۱]، کونور و هیرینگا [۲]) نشان می‌دهند که جامعه خارج از حوزه از نظر بسیاری از متغیرها بسیار شبیه به جامعه هدف است و اریبی ناشی از برش دادن بسیار اندک خواهد بود. همان طور که قبلاً اشاره شد، تقریباً ۵٪ جامعه خانوارهای آمریکا جزو جامعه تلفن‌دار نیستند و هرگونه اریبی اضافی ناشی از برش دادن چارچوب BCR احتمالاً ناچیز خواهد بود.

۳.۱۵ برآورد کردن

در محاسبه برآوردها از روی نمونه‌ها باید جنبه‌های احتمالاتی این طرحها نیز در نظر گرفته شوند. در اینجا اصول اساسی این قبیل برآورد کردن به طور مختصر برای میانگینها (و به طور ضمنی برای نسبتها) و واریانسهای نمونه‌گیری آنها توصیف شده‌اند. به علاوه، در بعضی موارد، پس طبقه‌بندی، یا تعدیل کنترل جامعه، در داده‌های آمارگیری تلفنی به کار رفته است تا در تعدیل نمونه خانوارهای تلفن‌دار با توزیع همه خانوارها تلاش شود.

۱.۳.۱۵ برآورد میانگینها

برای طرح شماره‌گیری رقم تصادفی ساده، فرض کنید \bar{y}_{RDD} میانگین ساده n مشاهده از Y ، متغیر خانوار باشد. به همین ترتیب فرض کنید \bar{y}_{MW} میانگین ساده mk مشاهده تحت طرح میتوفسکی - واکسبرگ باشد. هم \bar{y}_{RDD} و هم \bar{y}_{MW} برای میانگین جامعه‌ای \bar{Y} از لحاظ طرح نارایب‌اند. به علاوه،

$$\text{var}(\bar{y}_{MW}) \cong \frac{\sigma^2}{mk} (1 + \rho(k-1)) \quad \text{و} \quad \text{var}(\bar{y}_{RDD}) = \frac{\sigma^2}{n}$$

که در آن σ^2 واریانس جامعه‌ای و ρ همبستگی درون انبوهه ۱۰۰ تایی برای متغیر Y است.

برآورد کردن میانگین جامعه برای طرح‌های طبقه‌بندی شده مبتنی بر راهنمای تلفن قدری پیچیده‌تر است. نمونه‌گیری در داخل طبقه، شماره‌گیری رقم تصادفی است. پس \bar{y}_h (میانگین ساده n_h مشاهده از طبقه h^n) برای میانگین طبقه‌ای \bar{Y}_h نارایب است. نتیجه آن که

$$y'_i = \sum_{h=1}^H N_h \left(\frac{n_h}{n'_h} \right) \bar{y}_h$$

در انبوهش جامعه‌ای مقادیر y برای خانوارهای تلفن‌دار تقریباً نارایب است و

$$N'_i = \sum_{h=1}^H N_h \left(\frac{n_h}{n'_h} \right)$$

برای کل تعداد خانوارهای تلفن‌دار، مثلاً N'_i ، تقریباً نارایب است. به این ترتیب، برآوردگر نسبتی $\bar{y}_{strat} = (y'_i / N'_i)$ برای میانگین جامعه‌ای تقریباً نارایب است و

$$\text{var}(\bar{y}_{strat}) \cong \sum_{h=1}^H \frac{z_h^2 \sigma_h^2 (1 + (1 - p_h) \lambda_h)}{n_h}$$

که در آن p_h نسبت شماره تلفن‌های واگذار شده به خانوارهای دارای تلفن خانگی در طبقه h ام است، z_h نسبت جمعیت خانوارهای تلفن‌دار منظور شده در طبقه h ام است، و $\lambda_h = (\bar{Y}_h - \bar{Y})^2 / \sigma_h^2$. چندین نکته آماری دیگر هست که باید هنگام استفاده از طرح‌های تلفنی به یاد داشت:

۱. به طور کلی، برای برآورد کردن میانگین‌های زیررده‌ای به برآوردگرهای نسبتی نیاز است که عبارت‌های واریانسی نسبتاً ساده بالا در مورد آنها کاربرد ندارند.
۲. این طرح‌ها نمونه‌هایی از خانوارها را ارائه می‌دهند، نه اشخاص را. اگر اشخاص در درون خانوارها انتخاب شوند، آن‌گاه وزن‌دار کردن اضافی و برآوردگرهای پیچیده‌تری مورد نیازند.
۳. برای داشتن برآوردگرهای نارایب باید وزن‌های مربوط به خانوارهای دارای چند تلفن تعدیل شود تا احتمال بیشتر برای انتخاب شدن آنها به حساب آید.
۴. برآوردگرهای بالا برای دستیابی به اندازه نمونه‌ای ثابت بر استفاده از شماره‌گیری رقم تصادفی مبتنی هستند. این مستلزم آن است که وضعیت همه شماره تلفن‌های انتخاب شده تعیین شده باشد، که این نیز به نوبه خود مستلزم دفترداری دقیق و نظارت از نزدیک است. چون برای هر انبوهه ۱۰۰ تایی که نگهداری می‌شود اندازه‌های نمونه‌ای ثابت مورد نیاز است، روش میتوفسکی - واکسبرگ پیچیده‌تر می‌شود و به همین دلیل نیاز به کنترل شدیدتر نیز اهمیتی بیشتر دارد.

۲.۳.۱۵ برآورد کردن واریانس نمونه‌گیری

به منظور برآورد $Var(\bar{y}_{rdd})$ ، فرض کنید y_i مقدار متغیر y برای i امین خانوار انتخابی باشد. برآوردگر نارایب برای $Var(\bar{y}_{rdd})$ از فرمول زیر به دست می‌آید

$$\hat{Var}(\bar{y}_{rdd}) = \frac{\hat{\sigma}^2}{n}$$

که در آن،

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_{rdd})^2}{n-1}$$

برای نمونه‌گیری میتوفسکی - واکسبرگ فرض می‌کنیم y_{ij} مقدار متغیر y برای j امین خانوار انتخاب شده در i امین انبوهه ۱۰۰ تایی نگهداری شده باشد. برآوردگر نارایب برای $Var(\bar{y}_{mw})$ از فرمول زیر به دست می‌آید

$$\hat{Var}(\bar{y}_{mw}) = \frac{1}{m} \frac{\sum_{i=1}^m (\bar{y}_i - \bar{y}_{mw})^2}{m-1}$$

که در آن،

$$\bar{y}_i = \frac{\sum_{j=1}^k y_{ij}}{k}$$

برای طرح طبقه‌بندی شده فرض می‌کنیم y_{hi} مقدار متغیر y برای i امین خانوار انتخاب شده در h امین طبقه باشد. کاربرد فن خطی‌سازی در مورد برآوردگر نسبتی \bar{y}_{strat} ، برآوردگر واریانس زیر را به دست می‌دهد

$$\hat{Var}(\bar{y}_{strat}) = \sum_{h=1}^H \frac{\hat{z}_h^2 \hat{\sigma}_h^2 (1 + (1 - \hat{p}_h) \hat{\lambda}_h)}{n_h}$$

که در آن،

$$\hat{p}_h = \frac{n_h}{n'_h} \quad \hat{z}_h = \frac{N_h \hat{p}_h}{N'_t} \quad \hat{\sigma}_h^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1} \quad \hat{\lambda}_h = (\bar{y}_h - \bar{y}_{strat})^2 / \hat{\sigma}_h^2$$

اگرچه نتایج به تفصیل ارائه نشده‌اند، فن خطی‌سازی می‌تواند در به دست آوردن برآوردگرهایی برای واریانس برآوردگرهای نسبتی مورد نیاز مربوط به میانگینهای زیررده نیز مورد استفاده قرار بگیرد.

۳.۳.۱۵ پس طبقه‌بندی

در نظریه نمونه‌گیری سنتی، پس طبقه‌بندی هنگامی مطرح می‌شود که متغیرهایی که باید برای ایجاد طبقات مورد استفاده قرار گیرند در زمان انتخاب کردن در دسترس نباشند. یعنی ممکن است علاقه‌مند به افراز جامعه به G پس طبقه، با استفاده از متغیرهای جمع‌آوری شده در طی آمارگیری باشند. مانند نمونه‌گیری طبقه‌بندی شده با انتساب متناسب، بهبود در دقت با تعدیل مناسب برای برآورد کردن واریانس امکان‌پذیر است. پس طبقه‌بندی مستلزم آن است که پس طبقه‌ها برای هر عنصر نمونه معلوم باشد و وزنهای پس طبقه، مثلاً W_g ، برای هر پس طبقه موجود باشند. وزنهای پس طبقه‌ای باید از یک منبع خارجی از قبیل یک سرشماری، پیش‌بینیهای سرشماری، یا سوابق اداری به دست آمده باشند. مثلاً، پس طبقه‌های مبتنی بر سن و جنسیت در صورتی می‌توانند برای پاسخگویان ایجاد شوند که فراوانیها یا نسبتهای W_g جامعه‌ای مناسب را برای گروههای سنی و جنسی بتوان در جامعه پیدا کرد.

در نمونه‌گیری تلفنی غالباً پس طبقه‌بندی نه تنها با جمعیت ساکن در خانوارهای تلفن‌دار بلکه با جمعیت ساکن در همه خانوارها تعدیل می‌شود. این حالت از پس طبقه‌بندی به منظور به دست آوردن برآوردهایی به کار برده می‌شود که به مفهومی خاص با توزیع جمعیت، نه تنها در خانوارهای تلفن‌دار، بلکه در همه خانوارها تعدیل شده باشند.

گامهای پس طبقه‌بندی را می‌توان به صورت زیر خلاصه کرد:

۱. تفکیک نمونه در G پس طبقه براساس مشاهده برخی مشخصه‌ها.
۲. به دست آوردن وزنهای نمونه‌ای W_g برای جامعه، نوعاً از روی یک منبع خارجی از قبیل یک آمارگیری بزرگتر، یک سرشماری، پیش‌بینیهای سرشماری، یا سوابق اداری.
۳. محاسبه میانگینهای \bar{y}_g برای مشخصه موردنظر، جداگانه برای هر پس طبقه، و محاسبه میانگین کل $\bar{y}_{ps} = \sum_{g=1}^G W_g \bar{y}_g$.
۴. برای برآوردهای واریانسی، استفاده از

$$\hat{Var}(\bar{y}_{ps}) \cong \frac{1}{n} \left[\sum_{g=1}^G W_g s_g^2 + \sum_{g=1}^G W_g (1 - W_g) \frac{s_g^2}{N_g} \right]$$

یا

$$\hat{Var}(\bar{y}_{ps}) \cong \frac{1}{n} \sum_{g=1}^G W_g s_g^2 \left[1 + \frac{1 - W_g}{N_g} \right]$$

که در آن، s_g^2 برآورد عنصر در داخل پس طبقه، و N_g اندازه جامعه برای پس طبقه g است. شکل برآورد s_g^2 به طرح نمونه‌ای بستگی خواهد داشت.

تقریباً در همه وضعیتهای عملی، \bar{y}_{ps} ، برآورد پس طبقه‌بندی، واریانسهای کوچکتر از واریانس برآورد میانگین بدون پس طبقه‌بندی خواهد داشت.

از پس طبقه‌بندی غالباً به عنوان *تعدیل کنترل جامعه* نیز نام برده می‌شود. وزنه‌های پس طبقه‌بندی شده برای هر عنصر بسط داده می‌شوند. برآوردهای موزون که با استفاده از وزنه‌های پس طبقه‌بندی شده محاسبه می‌شوند نسبت به توزیع بیرونی که به صورت W_g نشان داده می‌شود «تعدیل» می‌شوند. در مورد طرح شماره‌گیری رقم تصادفی، اثرهای این تعدیل را در صورتی می‌توان با وضوح بیشتر مشاهده کرد که برآورد میانگین پس طبقه‌بندی شده را مجدداً به شکل زیر بیان کنیم. فرض می‌کنیم r معرف تعداد پاسخگویان در نمونه و r_g معرف تعداد پاسخگویان در پس طبقه g ام باشد. علاوه بر این، فرض می‌کنیم y_{gi} نشانگر مقدار مشخصه Y برای i امین پاسخگو در g امین پس طبقه است. در این صورت می‌توان میانگین پس طبقه را برحسب عنصر وزنها، W_{gi} ، به صورت زیر نوشت:

$$\begin{aligned}\bar{y}_{ps} &= \sum_{g=1}^G W_g \bar{y}_g = \frac{\sum_{g=1}^G N_g \bar{y}_g}{N} = \frac{\sum_{g=1}^G \left(\frac{r}{N}\right) \left(\frac{N_g}{r_g}\right) \sum_{i=1}^{r_g} y_{gi}}{\sum_{g=1}^G \left(\frac{r}{N}\right) N_g} \\ &= \frac{\sum_{g=1}^G \sum_{i=1}^{r_g} w_{gi} y_{gi}}{\sum_{g=1}^G \frac{N_g / N}{1/r}} = \frac{\sum_{g=1}^G \sum_{i=1}^{r_g} w_{gi} y_{gi}}{\sum_{g=1}^G \sum_{i=1}^{r_g} \left(\frac{1}{r_g}\right) \frac{N_g / N}{1/r}} = \frac{\sum_{g=1}^G \sum_{i=1}^{r_g} w_{gi} y_{gi}}{\sum_{g=1}^G \sum_{i=1}^{r_g} w_{gi}}\end{aligned}$$

یعنی وزن، W_{gi} ، حاصل بخش نسبت در جمعیت در g امین پس طبقه و نسبت در نمونه در g امین پس طبقه است: $W_{gi} = (N_g / N) / (r_g / r)$. بنابراین، پس طبقه‌بندی یک نمونه از پاسخگویان خانوارهای تلفنی برای توزیعی براساس همه خانوارها، تعدیلی همزمان برای بی‌پاسخی و بی‌پوششی خانوارهای بدون تلفن فراهم می‌کند.

پس طبقه‌بندی نمونه‌های تلفنی جنبه‌هایی دارد که توجه به آنها حائز اهمیت است. اول، نوعاً، W_g ها داده‌های سرشماری یا سایر داده‌های مربوط به همه خانوارها و نه صرفاً خانوارهای دارای تلفن هستند. دوم، در حالی که تعدیل پس طبقه‌بندی می‌تواند به عنوان تعدیلی هم برای بی‌پاسخی و هم برای بی‌پوششی تلقی شود، در عمل غالباً پس از شکلی از جبران بی‌پاسخی از طریق

وزن دادن، به کار برده می‌شود. سوم، ممکن است به دست آوردن وزنهای جامعه‌ای، W_g ، برای رده‌بندی متقاطع کامل برای جامعه امکان‌پذیر نباشد، ولی توزیعهای حاشیه‌ای ممکن است موجود باشند. شیوه‌های تعدیل نسبتی با دادن وزن (ن. ک. کالتون و کاسپرزیک [۷]) می‌تواند برای تولید یک توزیع کامل از رده‌بندی متقاطع براساس توزیعهای حاشیه‌ای مورد استفاده قرار گیرد. مثلاً وزنهای جامعه‌ای ممکن است برای سن و تحصیلات موجود باشند، ولی برای رده‌بندی متقاطع آنها موجود نباشند. برآورد نسبتی با دادن وزن می‌تواند برای تولید رده‌بندی متقاطع براساس یک مدل «اثرهای اصلی» برای سن و تحصیلات مورد استفاده واقع شود. پس از آن، وزنهای رده‌بندی متقاطع با دادن وزن برای جامعه در مورد توزیع پاسخگو به کار خواهند رفت تا وزنهایی در سطح عنصر به صورتی که در بالا اشاره شد تولید شوند.

۴.۱۵ مقایسه طرحها

۱.۴.۱۵ مبادله هزینه - واریانس

تابع هزینه در بخش ۲.۱.۲.۱۵ و واریانس نمونه‌گیری یک برآورد میانگین در بخش ۱.۳.۱۵ اندازه نمونه‌ای k در درون انبوهه ۱۰۰ تایی را تعیین می‌کنند که هزینه مورد انتظار یک واریانس نمونه‌گیری ثابت را برای طرح میتوفسکی - واکسبرگ به حداقل می‌رساند (یا واریانس را برای یک هزینه ثابت به حداقل می‌رساند). یک عبارت صریح برای مقدار بهینه k را می‌توان در مقاله واکسبرگ [۱۸] یافت. به همین ترتیب، تابع هزینه در بخش ۲.۳.۲.۱۵ و واریانس برآورد میانگین در بخش ۱.۳.۱۵ انتسابی از نمونه را به طبقه‌ها تعیین می‌کنند که هزینه مورد انتظار یک واریانس ثابت را برای طبقه‌بندی مبتنی بر راهنمای تلفن به حداقل می‌رساند (یا واریانس را برای یک هزینه ثابت به حداقل می‌رساند). عبارتهای صریح برای انتساب نمونه را می‌توان در مقاله کاسادی و لپکوفسکی [۳] یافت.

با استفاده از مقادیر مورد قبول عوامل هزینه‌ای و پارامترهای جامعه‌ای برای مدل‌های هزینه‌ای ساده و عبارتهای مربوط به واریانس که در بالا ذکر شدند، کاسادی و لپکوفسکی چنین نتیجه‌گیری کرده‌اند که طرح میتوفسکی - واکسبرگ و طبقه‌بندی مبتنی بر راهنمای تلفن هر دو نسبت به طرح شماره‌گیری رقم تصادفی ساده بهبود قابل ملاحظه‌ای داشته‌اند. آنها همچنین به این نتیجه رسیده‌اند که براساس مدل هزینه‌ای ساده به تنهایی، دو رهیافت مزبور از نظر کارایی تفاوتی اندک با یکدیگر دارند. ولی چنانچه امکان آریبی بیشتر را بتوان تحمل کرد، در آن صورت، طرح برشی از همه کارآمدتر است.

۲.۴.۱۵ ملاحظات اجرایی

ویژگیهای سیستمهای تلفن که بر اجرای طرحهای توصیف شده در بخش اخیر تأثیر می‌گذارند بسیار فراوانند. در این زیربخش، چندین مورد را که مهمترین مورد بحث قرار می‌دهیم.

شناسایی وضعیت خانگی بودن هر شماره تلفن که از شماره‌گیری رقم تصادفی یا نمونه‌هایی به یاری فهرست تولید شده است همیشه فرایندی آسان نیست. شماره‌هایی که جواب می‌دهند باید از نظر کاربری خانگی بررسی شوند و آنهایی که مشترکاً برای مقاصد خانگی و تجاری مورد استفاده قرار می‌گیرند باید به طرز مناسبی رده‌بندی شوند (معمولاً هر کاربری خانگی کافی است که شماره تلفن به عنوان خانگی رده‌بندی شود). بعضی شماره‌ها فوراً به عنوان غیرخانگی شناسایی می‌شوند زیرا دایر نیستند و وضعیت ثبتي آنها نیز به وضوح این وضعیت را نشان می‌دهد. بسیاری از شماره‌هایی که دایر نیستند از نظر ثبتي نیز به جایی ارتباط پیدا نمی‌کنند که وضعیت آنها را نشان دهد، ولی به یک «ماشین زنگ‌زن» متصل‌اند. به این ترتیب، مصاحبه‌گرانی که شماره تلفن‌ها را برای تعیین وضعیت خانگی بودن آنها غربالگری می‌کنند نمی‌توانند بین شماره تلفنهای خانگی که زنگ می‌خورند ولی کسی در منزل نیست و شماره تلفنهایی که فعلاً دایر نیستند تفاوت قایل شوند.

این مشکل اخیر در پیاده‌سازی برخی طرحها دارای اهمیتی خاص است. سامان دادن به شماره‌هایی که بدون جواب زنگ می‌خورند در طرحهای شماره‌گیری رقم تصادفی دو مرحله‌ای که مستلزم جایگذاری شماره تلفنهای غیر خانگی است، به خصوص در ایامی که جمع‌آوری داده‌های آمارگیری با محدودیت زمانی همراه است، بسیار مشکل است. بسیاری از سازمانهای آمارگیری با شماره تلفنهایی که در اوقات گوناگون روز و روزهای مختلف هفته بدون جواب زنگ می‌خورند به عنوان غیر خانگی رفتار می‌کنند. اگر رده‌بندی به عنوان غیر خانگی در طی دوره بررسی دیر انجام شده باشد شماره جایگزین مدت زمان نسبتاً کوتاهی خواهد داشت تا به آن تلفن شود. غالباً شماره‌های جایگزین همان تنوع زمانی را در طی روز و در روزهای هفته که شماره‌های اصلی در اختیار داشته‌اند برای تلفن کردن در اختیار ندارند. سازمانهای آمارگیری شیوه‌های نمونه‌گیری را که نمونه ثابتی از شماره تلفن‌ها ارائه می‌دهند به شیوه‌ای که ممکن است در اواخر دوره آمارگیری شماره تلفنهای جدیدی تولید کند ترجیح می‌دهند.

در پایان دوره بررسی باید شماره تلفنهایی که بدون جواب زنگ خورده‌اند و چندین بار به آنها تلفن شده است به عنوان خانگی یا غیر خانگی رده‌بندی شوند تا بررسی بسته شود. اگر به شماره‌ای در ساعتها و روزهای گوناگون تلفن شده است می‌توان آن را به طور دلخواه به عنوان غیر خانگی رده‌بندی کرد. به این ترتیب در برابر نرخ پاسخ به آمارگیری شمارش نخواهد شد، زیرا به عنوان غیرنمونه‌ای

رده‌بندی شده است. از سوی دیگر، شماره تلفنهایی که بدون جواب زنگ خورده‌اند ولی به دفعات کافی به آنها تلفن نشده است نوعاً به عنوان خانگی و بدون پاسخ رده‌بندی می‌شوند و به محاسبه‌ای محافظه‌کارانه از نرخ پاسخ می‌انجامند.

برای غلبه بر این مشکلات و کاهش هزینه‌های غربالگری شماره تلفن‌ها از لحاظ خانگی بودن، سیستمهای غربالگری خودکار ابداع شده‌اند تا حداقل، شماره تلفنهایی را که به برنامه‌های ضبط شده متصل‌اند شناسایی کرده، معلوم کنند که دایرند یا نه. برنامه ضبط شده برای «خارج از سرویس» نوعاً با یک «فاصله سه گام کامل موسیقی» شروع می‌شود که با صدای زنگ تلفن همراه نیست. نوعی سخت‌افزار و نرم‌افزار انحصاری تهیه شده است که تلفن‌ها را شماره‌گیری و طنین فاصله سه گام را کشف می‌کند. شماره تلفنهایی که دارای برنامه ضبط شده فاصله سه گام هستند از نمونه‌گیری بعدی حذف می‌شوند. شماره‌هایی که برنامه سه گام را ندارند غالباً «صدای قطع زنگ» را تجربه می‌کنند که طی آن در حالی که سخت‌افزار ارتباط را قطع می‌کند صدای زنگ تلفن برای لحظه‌ای شنیده می‌شود.

آمارگیریهایی که در سطح ایالت یا کل کشور اجرا می‌شوند دارای مرزهایی جغرافیایی برای جامعه خود هستند که با مرزهای کد ناحیه‌ای مطابقت دارند. شماره تلفنهای نمونه‌ای که در داخل کدهای ناحیه‌ای نمونه تولید می‌شوند به مکانهای مسکونی داخل ناحیه جغرافیایی هدف اختصاص خواهند داشت. بسیاری از آمارگیریهای جامعه‌های با تعریف جغرافیایی را هدف قرار داده‌اند که مرزهای آنها با کدهای ناحیه‌ای و مرزهای مراکز تلفن جور نمی‌شوند. در این قبیل موارد می‌توان جامعه را دوباره تعریف و آن را به جامعه ساکن در محدوده مراکز تلفن تعیین شده محدود کرد، یا می‌توان نمونه‌ای از یک مجموعه از مراکز تلفن انتخاب کرد که سراسر ناحیه جغرافیایی را پوشش دهد ولی نواحی خارج از هدف را نیز شامل شود. در آن صورت باید شماره تلفن‌ها نه تنها برای تعیین وضعیت خانگی بودن بلکه برای مکان‌یابی اماکن مسکونی براساس خود - گزارش پاسخگو نیز غربالگری شوند. رده‌بندی شماره تلفنهایی که بدون جواب زنگ می‌خورند در این قبیل آمارگیریهای غربالگری، به مراتب مسئله‌سازتر است.

تشخیص دوباره‌کاریها در هر یک از چارچوبها نیز نوعاً مستلزم خود - گزارش پاسخگو است. از خانوارهای پاسخگو سؤال می‌شود که آیا بیش از یک شماره تلفن به آن خانوار اختصاص دارد یا نه و اگر پاسخ مثبت بود شماره این قبیل تلفن‌ها پرسیده می‌شود. این شماره تلفن خود - گزارشی که از طریق آن می‌توان به خانوار دسترسی پیدا کرد بعداً در تولید وزن برای برآورد کردن استفاده می‌شود. بسیاری از سازمانهای آمارگیری، ارتباطهای اشتباه و شماره‌گیریهای غلط تلفنچی را نیز بازبینی می‌کنند.

شماره تلفنهایی که اشتبهاً شماره‌گیری شده‌اند مانند ارتباطهای اشتباه حذف می‌شوند تا از پیچیدگی بیشتر کارها در فرایند وزن دادن برای مدخلهای تکراری خانوار پرهیز شود.

آمارگیریهای علوم اجتماعی و بهداشتی نیز غالباً یک فرد واجد شرایط را در خانوار برای مصاحبه بیشتر انتخاب می‌کنند. مثلاً در یک آمارگیری مربوط به رضایت از وضعیت ازدواج، یک فرد بزرگسال انتخاب خواهد شد تا از تأثیر گفت‌وگوی بزرگسالان درباره محتوای آمارگیری بین مصاحبه‌ها بر پاسخها، جلوگیری شود. انتخاب پاسخگو باید در مراحل اولیه مصاحبه انجام پذیرد. شیوه انتخاب عینی پاسخگو که کیش [۸] توصیف کرده است در آمارگیریهای تلفنی برای این منظور کاربرد وسیعی داشته است ولی این شیوه به نتیجه‌ای نامطلوب - یعنی نرخ افزایش یافته بی‌پاسخی - می‌انجامد. خانوارها تمایل ندارند در آمارگیری شرکت کنند که اولین سؤالهایی که در آن مطرح شده است برای به دست آوردن فهرستی از افراد واجد شرایط ساکن در خانوار است. روشهای دیگر، شامل شیوه توصیف شده توسط ترولدال و کارتر [۱۷]، و روش نزدیکترین تاریخ تولد است (برای توصیف آن نگاه کنید به لاوراکاس [۹]). نشان داده شده است که این شیوه آخر دارای اریبی در انتخاب کردن است ولی همچنان مورد استفاده قرار می‌گیرد زیرا کاربرد آن آسان است و نگرانیهای مربوط به افزایش نرخهای بی‌پاسخی را که با شیوه انتخاب کردن درون خانواری کیش همراه است ندارد.

سرانجام این که دستگاههای پاسخگوی تلفن و تلفنهای همراه مشکلات روزافزونی را برای عملیات نمونه‌گیری تلفنی مطرح ساخته‌اند. ماشینهای پاسخگو تا حدود زیادی تشخیص سریع واحدهای مسکونی را میسر ساخته‌اند. می‌توان پیامهایی گذاشت که از خانوار درخواست شود با یک شماره تلفن مجانی تماس بگیرد و تلفن کردن به خانوارهای دارای ماشین پاسخگو را می‌توان در اوقات گوناگون روز و روزهای مختلف هفته زمانبندی کرد تا برای دستیابی به خانوار در زمانی که کسی به تلفن جواب می‌دهد تلاش شود. تلفنهای همراه مشکلی دیگر را مطرح می‌کنند. آیا این شماره تلفنهای خانگی هستند یا تجاری؟ به علاوه ممکن است مشترک مورد نظر با جواب دادن به تلفن هزینه‌ای متحمل شود. شماره تلفنهای همراه اکثراً در پیش شماره‌های جداگانه‌ای قرار داده نشده‌اند بلکه مانند سایر شماره تلفنهای با همان پیش شماره‌ها مخلوط شده‌اند. این توزیع، شناسایی را مشکل می‌کند. با وجود این به تلفنهای همراه ممکن است سریعتر از تلفن خانگی جواب داده شود. به علاوه، پرسشگری که خوب آموزش دیده باشد می‌تواند ترتیبی اتخاذ کند تا به خانواری با شماره دیگر تلفن بزند و به این ترتیب هزینه مشترک تلفن را کاهش دهد.

۳.۴.۱۵ انتخاب از میان طرحهای گوناگون

همان‌طور که قبلاً اشاره شد، گزینش از میان طرحهای گوناگون تا اندازه‌ی زیادی به در نظر گرفتن ویژگیهای مربوط به هزینه و خطای هر طرح مبتنی است. نوعاً سه عامل عمده‌ی هزینه‌ای در نظر گرفته می‌شود: هزینه‌ی تولید نمونه‌ی شماره تلفنها، هزینه‌ی غربالگری نمونه، و «راحتی» کار با شیوه‌ی نمونه‌گیری موردنظر در پیاده‌سازی (هزینه‌ای که غالباً کمی کردن آن مشکل است). از جهت خطا، دو نگرانی عمده وجود دارند: پوشش جامعه‌ی خانوارهای تلفن‌دار و واریانس نمونه‌گیری.

اگر سه رقیب اصلی را با توجه به این مشخصه‌ها مورد آزمایش قرار دهیم می‌توانیم درک کنیم که چرا سازمانها امروزه از میان طرحهای مختلف، گزینشهای خاصی را به عمل می‌آورند. مثلاً، تولید شماره تلفنها در طرح نمونه‌ای شماره‌گیری رقم تصادفی دو مرحله‌ای میتوفسکی - واکسبرگ کم‌خرج است. غربالگری در مرحله‌ی دوم کارایی قرار دارد زیرا بیش از ۶۰٪ شماره تلفنها خانگی‌اند. طرح میتوفسکی - واکسبرگ در اجرا مشکلات چندی را مطرح می‌کند، از جمله جایگذاری شماره تلفنها غیر خانگی و خوشه‌های خالی شده که می‌توانند برای برخی عملیات آمارگیری، ناراحتیهای قابل توجهی باشند و روشهای دیگری که از این مشکلات دوری کنند جاذبه‌ی زیادی پیدا می‌کنند. از لحاظ خطا، طرح میتوفسکی - واکسبرگ واقعاً پوشش کامل جامعه‌ی خانوارهای تلفن‌دار را تأمین می‌کند. واریانسهای نمونه‌گیری بیشتر از طرحهای نمونه‌ای عنصری است و علت آن، انتخاب نمونه‌ی خوشه‌ای و افزایشهای واریانس ناشی از همگنی درونی انبوهه‌ی ۱۰۰ تایی در میان عناصر نمونه است. یعنی این که اثرهای طرح برای نمونه‌های میتوفسکی - واکسبرگ بیشتر از یک است.

طرح طبقه‌بندی شده دارای مجموعه‌ی مشخصه‌هایی تا حدودی متفاوت است. هزینه‌های تولید نمونه می‌توانند زیاد باشند. نمونه‌ی طبقه‌ای فهرست شده را می‌توان با هزینه‌ی نسبتاً کمی به ازای تعداد نمونه از یک شرکت نمونه‌گیری تجاری خرید ولی نمونه‌ی طبقه‌ای فهرست نشده مستلزم طبقه‌بندی بیشتر شماره تلفنها و نمونه‌های شماره‌گیری رقم تصادفی دو مرحله‌ای است که از هر طبقه‌ی فهرست نشده انتخاب شده باشد. هزینه‌های غربالگری نیز بیشتر از طرح میتوفسکی - واکسبرگ است زیرا تقریباً ۵۰٪ شماره تلفنها در طبقه‌ی فهرست شده خانگی هستند و در طبقه‌ی فهرست نشده درصد بسیار کمی خانگی‌اند. با توجه به این که روشهای نمونه‌گیری مختلف در طبقه‌ها مورد استفاده قرار می‌گیرند، راحتی انتخاب نمونه برای طرح طبقه‌بندی شده کمتر از طرح میتوفسکی - واکسبرگ است. ولی طرح طبقه‌بندی شده نیاز به جایگزین کردن شماره تلفنها و خوشه‌های خالی شده را از میان می‌برد. طرح طبقه‌بندی شده در واقع کل جامعه‌ی خانوارهای تلفن‌دار را پوشش می‌دهد. واریانسهای نمونه‌گیری برای

طرح طبقه‌بندی شده کمتر از طرح میتوفسکی - واکسبرگ خواهد بود زیرا نمونه‌گیری از عناصر، دارای واریانسهای کمتری است و می‌توان بهبودهایی را در دقت ناشی از طبقه‌بندی انتظار داشت.

طرح برشی به نسبت طرحهای میتوفسکی - واکسبرگ و طبقه‌بندی شده دارای عیب عدم پوشش خانوارهای تلفن‌دار در انبوه‌های ۱۰۰ تایی فاقد شماره تلفنهای فهرست شده است. سطح عدم پوشش پایین است و بررسیهای تجربی نشان داده‌اند که تفاوت بین جوامع پوشش داده شده و پوشش داده نشده برای بسیاری از مشخصه‌ها اندک است. نمونه‌های انتخاب شده با استفاده از طرح برشی هنگامی که از شرکتهای نمونه‌گیری تجاری به دست آمده باشند گرانها نیستند. هزینه‌های غربالگری طرح برشی نسبت به طرحهای میتوفسکی - واکسبرگ و طبقه‌بندی شده در حد متوسط است زیرا تقریباً ۵۵٪ شماره تلفنهای تولید شده خانگی خواهند بود. طرح برشی در میان سه طرحی که در اینجا مورد بررسی قرار گرفته‌اند از همه راحت‌تر است زیرا به هیچ نوع جایگذاری شماره تلفن نیاز ندارد، نمونه فقط از طبقه فهرست شده انتخاب می‌شود، و هیچ نمونه‌گیری دو مرحله‌ای از روی طبقه فهرست نشده لازم نیست. واریانسهای نمونه‌گیری برآوردها برای طرح برشی باید از همه کمتر باشد زیرا این یک نمونه عنصری طبقه‌بندی شده بدون نمونه‌گیری خوشه‌ای است.

به این ترتیب، کسی که در حوزه نمونه‌گیری کار می‌کند با موضوع گزینش از میان طرحهایی که پوشش کامل را تأمین می‌کنند ولی با تعدادی نابسامانیها در انتخاب همراه‌اند و یک طرح که پوشش کمتری را تأمین می‌کند ولی با راحتیهایی در انتخاب همراه است مواجه می‌شود. در کاربست فعلی با توجه به شواهد تجربی مربوط به اندازه آریبی ناشی از عدم پوشش خانوارهای تلفن‌دار در انبوه‌های ۱۰۰ تایی فاقد شماره تلفنهای فهرست شده، طرح برشی اخیر را ترجیح می‌دهند. یعنی، دست‌اندرکاران، روشهای نمونه‌گیری برشی را برای آمارگیریهای تلفنی براساس یک روش کلاسیک ولی غیررسمی مبادله هزینه - خطا انتخاب می‌کنند.

۵.۱۵ خلاصه

در این فصل، روشهای مناسبی را برای نمونه‌گیری از خانوارهای تلفن‌دار با استفاده از چارچوب شماره تلفن و چارچوب مبتنی بر راهنمای تلفن شرح و بسط دادیم. شیوه‌های شماره‌گیری رقم تصادفی دو مرحله‌ای به صورتی که میتوفسکی ابداع کرده و واکسبرگ شرح و بسط داده است در سطحی گسترده مورد استفاده قرار می‌گیرند. سایر رهیافتهای جاری به نسبت شیوه مزبور دارای نارساییهایی از نظر پوشش و هزینه هستند. این نارساییها از طریق طرحهای نمونه‌ای تلفنی که از اطلاعات شماره تلفنهای فهرست شده برای بهبود کارایی هزینه شماره‌گیری رقم تصادفی استفاده می‌کنند مورد

رسیدگی قرار گرفته‌اند. چارچوب شماره تلفن به یک طبقه که اطلاعات شماره تلفنهای فهرست شده برای آن در سطح انبوه‌های ۱۰۰ تایی موجود است و یک طبقه که برای آن چنین اطلاعاتی موجود نیست تقسیم می‌شود. کاراییهای طرحهای نمونه‌گیری گوناگون برای این طرح طبقه‌بندی شده با شماره‌گیری رقم تصادفی ساده و فن میتوفسکی - واکسبرگ مقایسه شده است. افزایش نه چندان زیاد کارایی این طرحهای جایگزین را نسبت به فن میتوفسکی - واکسبرگ می‌توان تقریباً برای همه این قبیل طرحها نشان داد. همچنین درباره استفاده از پس طبقه‌بندی برای استنباطهایی از روی جامعه خانوارهای تلفن‌دار برای یک جامعه وسیعتر و درباره مسایل عملیاتی که در اجرای روشهای نمونه‌گیری تلفنی مطرح می‌شوند به طور خلاصه بحث کردیم.

تمرین

- ۱.۱۵ قرار است یک آمارگیری تلفنی با $n=1600$ مصاحبه تکمیل شده مطلوب در یک منطقه کلان‌شهری بزرگ اجرا شود. نرخ مورد انتظار برای تلفنهای خانگی دایر ۳۵٪ است. یک بزرگسال با ۲۱ سال سن یا مستتر قرار است از هر خانوار انتخاب شود (واجد شرایط بودن خانوار مورد انتظار ۹۵٪ است) و نرخ پاسخگویی مورد انتظار ۷۰٪ است. طرح نمونه‌ای شماره‌گیری رقم تصادفی دو مرحله‌ای میتوفسکی - واکسبرگ پیشنهاد می‌شود.
- الف. اگر اندازه خوشه را در $b=8$ شماره تلفن خانگی، ثابت فرض کنیم، چند شماره تلفن اولیه باید تولید کرد و چند خوشه باید به دست آورد؟
- ب. اگر نرخ مورد انتظار دایر بودن تلفن ۶۵٪ باشد چند شماره تلفن باید در مرحله دوم تولید شود؟
- پ. چند مشکل در طی مصاحبه پیش‌بینی می‌شود. با یک جمله تنها توضیح دهید که آیا این شماره تلفنها به عنوان «نمونه یا غیرنمونه» رده‌بندی می‌شوند یا نه و نحوه رفتار با هر یک را چگونه توصیه می‌کنید:
۱. ارتباط غلط
 ۲. تماس با خانوار برقرار شده و قرار مصاحبه برای زمان دیگری گذاشته شده است. وقتی در زمان مقرر دوباره شماره‌گیری می‌شود، شماره تلفن خارج از سرویس است.
 ۳. تلفن شماره‌گیری شده مورد استفاده یک شرکت حسابداری کوچک در یک خانه مسکونی است ولی برای مکالمات شخصی نیز مورد استفاده قرار می‌گیرد.

۴. در پایان بررسی، تعداد کمی از شماره تلفنهای باقی مانده‌اند که پس از یک تا ۲۰ بار شماره‌گیری هنوز بدون جواب زنگ می‌خورند.

۲.۱۵ یک نمونه تلفنی در یک مرکز تلفن انتخاب شده است. یکصد (۱۰۰) شماره تلفن متوالی در یک انبوهه ۱۰۰ تایی که در مرحله اول طرح میتوفسکی - واکسبرگ انتخاب شده‌اند به ترتیب تصادفی قرار گرفته‌اند. فرض کنید که در یکی از انبوهه‌های انتخاب شده به چهل شماره تلفن اول فهرست تصادفی شده تلفن شده و نتایج زیر برای هر یک به دست آمده است:

مصاحبه	۹۳۶-۰۰۲۶ (۲۱)	مصاحبه	۹۳۶-۰۰۷۱ (۱)
کسب و کار	۹۳۶-۰۰۴۹ (۲۲)	کسب و کار	۹۳۶-۰۰۵۳ (۲)
قطع شده	۹۳۶-۰۰۸۴ (۲۳)	مشغول (۶ بار تلفن)	۹۳۶-۰۰۴۶ (۳)
زنگ بدون جواب	۹۳۶-۰۰۰۳ (۲۴)	واگذار نشده	۹۳۶-۰۰۸۴ (۴)
قطع شده	۹۳۶-۰۰۹۴ (۲۵)	واگذار نشده	۹۳۶-۰۰۸۹ (۵)
کسب و کار	۹۳۶-۰۰۱۹ (۲۶)	کسب و کار	۹۳۶-۰۰۷۹ (۶)
مصاحبه	۹۳۶-۰۰۳۱ (۲۷)	امتناع	۹۳۶-۰۰۵۰ (۷)
امتناع	۹۳۶-۰۰۵۶ (۲۸)	مصاحبه	۹۳۶-۰۰۲۴ (۸)
زنگ بدون جواب	۹۳۶-۰۰۱۶ (۲۹)	ارتباط اشتباه	۹۳۶-۰۰۲۱ (۹)
واگذار نشده	۹۳۶-۰۰۹۳ (۳۰)	قطع شده	۹۳۶-۰۰۰۱ (۱۰)
زنگ بدون جواب	۹۳۶-۰۰۸۷ (۳۱)	مصاحبه	۹۳۶-۰۰۳۸ (۱۱)
کسب و کار	۹۳۶-۰۰۴۹ (۳۲)	تلفن پولی	۹۳۶-۰۰۹۰ (۱۲)
نوفه	۹۳۶-۰۰۱۰ (۳۳)	گوشی را می‌گذارند (۵ بار)	۹۳۶-۰۰۴۸ (۱۳)
واگذار نشده	۹۳۶-۰۰۹۳ (۳۴)	زنگ بدون جواب (۱۵)	۹۳۶-۰۰۴۵ (۱۴)
واگذار نشده	۹۳۶-۰۰۹۵ (۳۵)	زنگ بدون جواب (۱۱)	۹۳۶-۰۰۱۴ (۱۵)
کسب و کار	۹۳۶-۰۰۲۲ (۳۶)	پیامده بیمارستان	۹۳۶-۰۰۰۲ (۱۶)
مصاحبه	۹۳۶-۰۰۱۱ (۳۷)	امتناع	۹۳۶-۰۰۱۷ (۱۷)
امتناع	۹۳۶-۰۰۷۵ (۳۸)	ماشین پاسخگو	۹۳۶-۰۰۳۰ (۱۸)
واگذار نشده	۹۳۶-۰۰۳۷ (۳۹)	واگذار نشده	۹۳۶-۰۰۹۶ (۱۹)
واگذار نشده	۹۳۶-۰۰۱۲ (۴۰)	به شماره ۱۱۰۱-۷۴۷	۹۳۶-۰۰۴۴ (۲۰)

تغییر یافته است

برای هر یک از انواع نتایج زیر، بیان کنید که اگر در پایان بررسی این نتایج را به دست می‌آوردید آنها را به عنوان خانگی رده‌بندی می‌کردید یا غیر خانگی:

- الف. قطع شده
 ب. خارج از سرویس
 پ. زنگ بدون پاسخ برای ۱۱ بار تلفن یا بیشتر
 ت. ارتباط اشتباه (با ۲ بار شماره‌گیری)
 ث. ماشین پاسخگو
 ج. شماره تلفن عوض شده
 چ. گوشی را می‌گذارند یا زنگ بدون جواب برای تلفنهای متعدد

۳.۱۵ ویژگیهای طرح زیر را برای یک طرح نمونه‌گیری تلفنی طبقه‌بندی شده در نظر بگیرید

طبقه	شرح	W_h	C_h
۱	انبوهه ۱۰۰ تایی فهرست شده	۰/۳۵	۰/۷۵ دلار
۲	انبوهه ۱۰۰ تایی فهرست نشده پیش‌شماره فهرست شده	۰/۲۴	۷/۳۰
۳	انبوهه ۱۰۰۰ تایی فهرست نشده باقیمانده	۰/۴۱	۱/۴۵

- الف. یک انتساب بهینه برای یک طرح سه طبقه‌ای تهیه کنید که هزینه‌های غربالگری آن از ۱۲۵۰ دلار بیشتر نباشد. فرض کنید واریانسهای مربوط به عنصر در سراسر طبقه‌ها تقریباً یکسان‌اند. به عناصر موجود در هر طبقه، وزنهای مورد نیاز جبران‌کننده احتمالاتی نابرابر برای انتخاب شدن را اختصاص دهید.
- ب. فرض کنید تصمیم بر این است که طبقه ۲ به علت هزینه‌های غربالگری شماره تلفنهای موجود در آن حذف شود (یعنی استفاده از طرح چارچوب برشی). حالا از کدام انتساب بهینه استفاده خواهد شد؟ وزنهای مورد نیاز عناصر برای جبران احتمالاتی نابرابر برای انتخاب شدن کدامها خواهند بود؟

کتابشناسی

1. Brick, J. M., Waksberg, J., Kulp, D., and Starer, A., Bias in list assisted telephone samples. *Public Opinion Quarterly*, 59(2): 218-235, 1995.
2. Brunner, J. A., and Brunner, G. A., Are voluntarily unlisted telephone subscribers really different? *Journal of Marketing Research*, 8: 121-124, 1971.

3. Casady, R. J., and Lepkowski, J. M., Stratified telephone sampling designs. *Survey Methodology*, 19(1): 103-113, 1993.
4. Connor, J., and Heeringa, S., Evaluation of two cost-efficient RDD designs. Paper Presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, FL, 1992.
5. Frankel, M. R., and Frankel, L., Some recent developments in sample survey design. *Journal of Marketing Research*, 14: 280-293, 1977.
6. Groves, R. M., and Lepkowski, J. M., Dual frame mixed mode survey designs. *Journal of Official Statistics*, 1(3): 263-286, 1985.
7. Kalton, G., and Kasprzyk, D., The treatment of missing survey data. *Survey Methodology*, 12: 1-16, 1988.
8. Kish, L., *Survey Sampling*, Wiley, New York, 1965.
9. Lavrakas, P. J., *Telephone Survey Methods: Sampling, Selection, and Supervision*. Sage Publications, Newbury Park, Calif., 1987.
10. Lepkowski, J. M., Telephone sampling methods in the United States. In *Telephone Survey Methodology*, Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L., II, and Waksberg, J., Eds., Wiley, New York, 1988.
11. Mitofsky, W., Sampling of Telephone Households. Unpublished CBS News memorandum, 1970.
12. Potthoff, R. F., Some generalizations of the Mitofsky-Waksberg techniques for random digit dialing. *Journal of the American Statistical Association*, 82(298): 409-418, 1987.
13. Potthoff, R. F., Generalizations of the Mitofsky-Waksberg technique for random digit dialing: Some added topics. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 615-620, 1987.
14. Stock, J. S., How to improve samples based on telephone listings. *Journal of Marketing Research*, 2(3): 50-51, 1962.
15. Sudman, S., The uses of telephone directories for survey sampling. *Journal of Marketing Research*, 10(2): 204-207, 1973.
16. Thornberry, O. T., and Massey, J. T., Trends in U.S. telephone coverage across time and subgroups. In *Telephone Survey Methodology*, Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nichols, W. L., II, and Waksberg, J., Eds., Wiley, New York, 1988.
17. Troidahl, V. C., and Carter, R. E., Jr., Random selection of respondents within households in phone surveys. *Journal of Marketing Research*, 1(2): 71-76, 1964.
18. Waksberg, J., Sampling methods for random digit dialing. *Journal of the American Statistical Association*, 19(1): 103-113, 1978.

The following review was written for the Encyclopedia of Biostatistics and contains, in a different format, much of the material that is in this chapter.

19. Casady, R. J., and Lepkowski, J. M., Telephone sampling. In *The Encyclopedia of Biostatistics*, Armitage, P. A., and Colton, T., Eds., Wiley, Chichester, U.K., 1998.