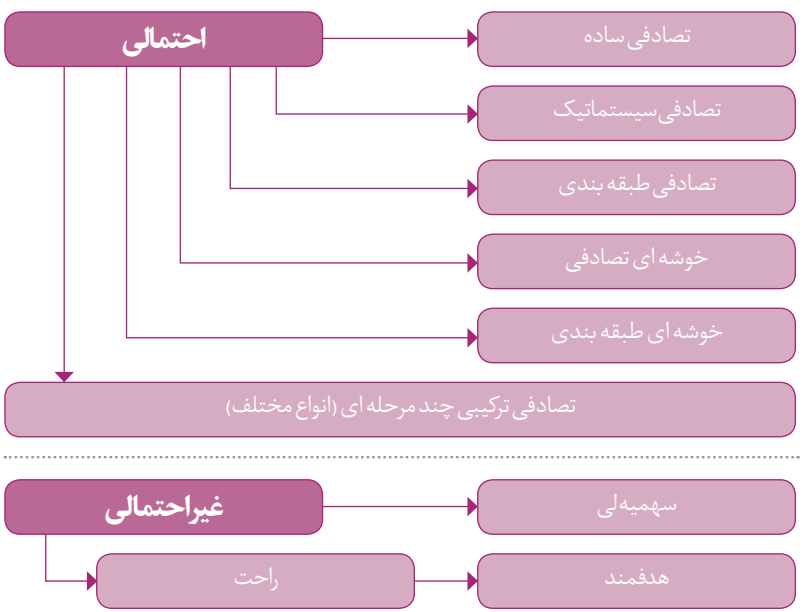
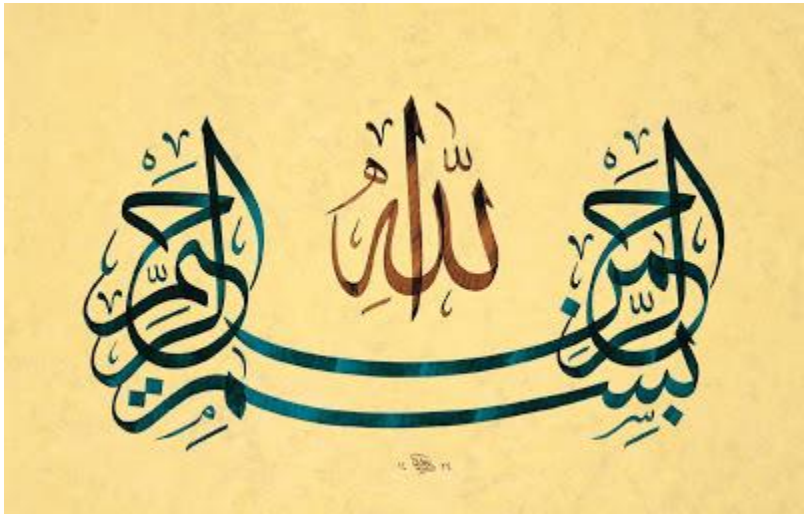
روش‌های نمونه‌گیری

گردآوری، ترجمه و تدوین: دکتر فرشید علاءالدینی




بسم الله الرحمن الرحيم

سرشناسه : علاءالدینی، فرشید، ۱۳۴۸ -
عنوان و نام پدیدآور : روشهای نمونه‌گیری/ گردآوری، ترجمه و تدوین فرشید علاالدینی.
مشخصات نشر : تهران: فرشید علاءالدینی، ۱۳۹۶.
مشخصات ظاهری : ۷۹ ص.: مصور، جدول.
شابک : ۲۰۰۰۰۰ ریال978-600-04-8348-7:
وضعیت فهرست نویسی : فیپا
موضوع : نمونه سنجی
موضوع : Sampling
موضوع : آمارگیری نمونه‌ای
موضوع : Sampling (Statistics)
رده بندی کنگره : ۱۳۹۶ ۸ع۱آ۱۸۰Q/
رده بندی دیویی : ۰۰۱/۴۳۳
شماره کتابشناسی ملی : ۴۸۳۵۹۳۸

روشهای نمونه گیری

گردآوری، ترجمه و تدوین : دکتر فرشید علاءالدینی

ناشر : مولف

تیراژ : 1000 جلد

نوبت چاپ : اول

قیمت : 20000 تومان (جهت خرید کتاب به آدرس اینترنتی www.inimkat.com/eshop مراجعه نمایید)

سال 1396

روش‌های نمونه‌گیری

گردآوری، ترجمه و تدوین: دکتر فرشید علاءالدینی

فهرست مطالب

مقدمه ۱

اصطلاحات نمونه‌گیری ۳

روشهای نمونه گیری ۹

مثال در رابطه با استفاده از روش‌های نمونه‌گیری ۵۱

نمونه‌گیری غیر احتمالی ۶۱

رفع مشکلات نمونه‌گیری ۶۷

خطاهای نمونه‌گیری ۷۷

مقدمه

بسیاری از حرفه‌ها (تجاری، دولتی، مهندسی، علمی، تحقیقات اجتماعی، کشاورزی، و غیره) به دنبال گسترده‌ترین مبنای واقعی برای تصمیم‌گیری هستند. در غیاب داده‌های مرتبط با یک موضوع، تصمیم‌گیری دقیقاً شبیه به پریدن درون تاریکی است. نمونه‌گیری فرایندی است که در آن بخشی از داده‌ها از مجموعه بزرگی از داده‌ها استخراج شده و سپس نتیجه بدست آمده از نمونه به کل مجموعه تعمیم داده می‌شود. کار اصلی بررسی کننده (فرد یا موسسه‌ای که مسئول جمع‌آوری و ثبت داده‌ها است) یا محققان این است که توجیه منطقی برای استفاده از نمونه‌گیری در تحقیقاتشان فرموله کنند. اگر نمونه‌گیری برای تحقیق مناسب تشخیص داده شد، محقق باید:

۱. تا آن جا که ممکن است و به روشی متناسب با هدف تحقیق، جمعیت هدف را با دقت تعیین کند.
۲. لیستی از جمعیت هدف که قرار است نمونه از درون آن انتخاب شود را تهیه کند. این لیست توسط بسیاری از متخصصان علم آمار بازه (یا به طور دقیق‌تر بازه لیست) نامیده شده است.
۳. نمونه را انتخاب کرده و در مورد تکنیک نمونه‌گیری تصمیم‌گیری کند، و؛
۴. در مورد جمعیت نتیجه‌گیری کند.

تمامی این چهار مرحله وابسته به هم هستند و نمی‌توانند از یکدیگر تفکیک شوند. نمونه‌گیری تصادفی ساده، نمونه‌گیری سیستماتیک، نمونه‌گیری طبقه بندی جزو تکنیک‌های نمونه‌گیری ساده هستند. تنها زمانی که مجموعه بزرگی از داده‌های تجربی در دسترس باشند؛ وقتی کارآیی مورد نیاز باشد و همزمان قرار است برآورد دقیقی در مورد گروه‌های نسبتاً کوچکی درون جمعیت‌های بزرگ انجام شود از تکنیک‌های پیچیده نمونه‌گیری استفاده می شود.

اصطلاحات نمونه‌گیری

جمعیت

جمعیت، که جامعه نیز نامیده می‌شود، مجموعه‌ای از افراد یا موسساتی است که قرار است یافته‌های مرتبط با آنها تعمیم داده شوند. جمعیت باید قبل از نمونه‌گیری به طور صریح تعریف شده باشد. باید مراقب بود که تعمیم فراتر از جامعه نباشد. انجام این کار یکی از متداول‌ترین خطاها در نوشته‌های آماری است.

نمونه‌گیری تصادفی

نمونه‌گیری تصادفی گردآوری داده است که بر اساس آن هر فرد درون جمعیت شانس این که انتخاب شود از قبل مشخص است. معمولاً شانس انتخاب شدن برابر است. اگر داده‌ها به صورت یک نمونه تصادفی باشد، محقق باید علاوه بر شدت روابط کشف نشده، سطح معناداری آنها (شانس این که یافته‌ها ناشی از شانس نمونه‌گیری است) نیز گزارش کند. انواع مختلف نمونه‌گیری وجود دارد.

بازه نمونه‌گیری

بازه نمونه‌گیری لیستی از هویت‌های نهایی نمونه‌گیری است که ممکن است افراد، صاحبخانه‌ها، سازمان‌ها یا سایر واحدهای تحلیلی باشند. به عنوان مثال، لیست دانشجویان ثبت نام شده ممکن است یک بازه نمونه‌گیری برای ارزیابی تعداد دانشجویان یک دانشگاه باشد. در نمونه‌گیری چند مرحله‌ای در هر مرحله یک بازه نمونه‌گیری وجود دارد (مثلاً در تحقیقی که افراد از خانه‌ها، خانه‌ها از بلوک‌های سرشماری، بلوک‌های سرشماری از نشریات سرشماری، نشریات سرشماری از استان، و استان‌ها از کشور نمونه‌گیری شده‌اند، پنج بازه وجود دارد: لیست ساکنین، خانه‌ها، بلوک‌های سرشماری، نشریات سرشماری، و

استان‌ها).

جمعیت هدف که در حالت ایده‌ال همان بازه نمونه‌گیری است. در واقعیت، جمعیت هدف ممکن است بزرگتر یا کوچکتر از بازه نمونه‌گیری در دسترس باشد که در این صورت محقق باید به پرسش‌هایی در مورد نمایندگی بازه نمونه‌گیری پاسخ دهد. اگر بازه نمونه‌گیری کوچکتر از جمعیت هدف باشد، پوشش کم وجود دارد. اگر بازه نمونه‌گیری بزرگتر از جمعیت هدف باشد، پوشش بیش از حد وجود دارد. در هر دو صورت انحراف بازه نمونه‌گیری وجود دارد. به عنوان مثال، از دفاتر تلفن اغلب به عنوان بازه‌های نمونه‌گیری استفاده شده ولی این دفاتر ضعف (تلفن‌های کمی دارند یا ندارند) و غنی بودن (نداشتن شماره‌های ثبت نشده) را کمتر از حد نشان می‌دهند. شماره‌گیری تصادفی اعداد (RDD) باعث دستیابی به شماره‌های لیست نشده می‌شود ولی آنهایی که تلفن ندارند را مشخص نمی‌کند، و صاحبخانه‌هایی که مالک چند شماره تلفن هستند بیش از حد نشان داده می‌شوند.

طبقات

طبقات گروه‌بندی متغیرهای استفاده شده در نمونه‌گیری هستند. در «نمونه‌گیری منطقه» گروه‌ها به صورت جغرافیایی هستند مثلاً بلوک‌های سرشماری، نشریات سرشماری، مناطق شهری و استان‌ها. واحد اولیه نمونه‌گیری (PSU) سطح کلی است که بلافاصله بعد از سطح فردی قرار دارد، مثلاً بلوک‌های سرشماری حاصل از نمونه‌گیری از ساکنین. البته، طبقات ممکن است جغرافیایی نباشد، مثل گروه‌های جنسی یا افراطی. نمونه‌گیری طبقه بندی از اطلاعات راجع به طبقات برای افزایش دقت نمونه استفاده می‌کند.

معناداری

معناداری درصد شانسی است که یک رابطه را می‌توان در داده‌های یک نمونه بد پیدا کرد، و اگر محقق نمونه دیگری را انتخاب کند، هیچ رابطه‌ای یافت نمی‌شود. یعنی معناداری به معنی شانس خطای تیپ یک است: شانس این که رابطه‌ای که وجود ندارد را پیدا کنیم. دانشمندان علوم اجتماعی اغلب از سطح به عنوان برش استفاده می‌کنند: اگر ۵ درصد

یا کمتر شانس این وجود داشته باشد که رابطه فقط ناشی از شانس باشد، نتیجه گرفته می‌شود که رابطه واقعی است. تست معناداری برای نمونه‌های غیر تصادفی یا شمارش شده/سرشماری مناسب نیست. محققان دوست دارند که نتایج مشابهی برای نمونه‌های غیرتصادفی بگیرند ولی چنین چیزی غیرممکن است.

فواصل اطمینان

فواصل اطمینان «حاشیه خطای» است که معمولاً برای نظرسنجی‌ها گزارش شده است. فواصل اطمینان ارتباط مستقیمی با ضریب معناداری دارند. برای یک متغیر مشخص در یک نمونه مشخص، فرد می‌تواند خطای استاندارد را محاسبه کند، که (با فرض نمونه‌گیری تصادفی، توزیع نرمال، و ۰،۰۵ سطح برش معناداری برای رد فرضیه صفر) دارای ۹۵ درصد فاصله اطمینان حدوداً به علاوه و منهای ۱/۹۶ برابر خطای استاندارد است. اگر تعداد بسیار زیادی نمونه گرفته شود، و میانگین برآورد (احتمالاً متفاوت) و فاصله اطمینان ۹۵ درصد برای هر نمونه ایجاد شده باشد، در این صورت ۹۵ درصد این فواصل اطمینان شامل مقدار واقعی جمعیت خواهد بود. فرمول محاسبه فاصله اطمینان، سطوح معناداری، و خطاهای استاندارد، به نوع نمونه‌گیری تصادفی، چه ساده و چه پیچیده، بستگی دارد.

شماره‌گذاری

که سرشماری نیز نامیده می‌شود، مجموعه‌ای از داده‌های مربوط به هر فرد یا هویت درون جمعیت است. اگر داده‌ها به صورت یک شمارش باشد، محقق نباید نگران معناداری که شکلی از برآورد خطای نمونه‌گیری است، باشد. یعنی این که، اگر سطح معناداری ۰/۰۵ باشد، این بدان معنی است که اگر نمونه تصادفی دیگری انتخاب شود ۵ درصد شانس این وجود دارد که نتیجه‌ای قوی یا قوی‌تر از آن چه که مشاهده شده به دست آوریم. با استفاده از شمارش، هیچ احتمالی برای «یک نمونه تصادفی دیگر» وجود ندارد و بنابراین تست معناداری انجام نمی‌شود. هر رابطه‌ای، بدون توجه به این که چقدر کم باشد، یک رابطه واقعی برای شمارش است (به جز خطای اندازه‌گیری). هرچند برخی از محققان سطوح

معناداری را برای داده‌های شمارشی گزارش می‌کنند ولی ضریب معناداری با اندازه نمونه و شدت رابطه ترکیب شده و بنابراین وقتی به عنوان معیاری برای اندازه تاثیر استفاده شود، عملکرد ضعیفی دارد.

تاثیر طرح

تاثیر طرح، D، ضریبی است که نشان می‌دهد طرح نمونه‌گیری چگونه بر محاسبه سطوح معناداری در مقایسه با نمونه‌گیری تصادفی ساده تاثیرگذار است. ضریب تاثیر طرح ۱ به این معنی است که طرح نمونه‌گیری معادل با نمونه‌گیری تصادفی ساده است. تاثیر طرح بالاتر از ۱ به این معنی است که طرح نمونه‌گیری باعث کاهش دقت برآورد در مقایسه با نمونه‌گیری تصادفی ساده می‌شود (مثلاً نمونه‌گیری خوشه‌ای دقت را کاهش می‌دهد). تاثیر طرح کمتر از ۱ به این معنی است که طرح نمونه‌گیری باعث افزایش دقت در مقایسه با نمونه‌گیری تصادفی ساده می‌شود (به عنوان مثال، نمونه‌گیری طبقه بندی باعث افزایش دقت می‌شود). متاسفانه، بیشتر برنامه‌های رایانه‌ای ضرایب معناداری و فواصل اطمینان را براساس فرمول‌های در نظر گرفته شده برای نمونه‌گیری تصادفی ساده تولید می‌کنند.

جایگزینی

نمونه‌ای که در آن هر واحد دارای احتمال برابر برای انتخاب شدن است را نمونه تصادفی می‌نامند. اگر تکرار مجاز نباشد، به آن نمونه تصادفی ساده می‌گویند که بدون جایگزینی انتخاب شده است. اگر تکرار مجاز باشد، نمونه با استفاده از جایگزینی انتخاب شده است.

ضریب درجه

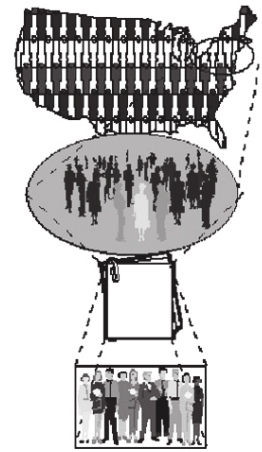
این ضریب برابر با خارج قسمت اندازه جمعیت و اندازه نمونه است، N/n. این ضریب نشان می‌دهد که هر عنصر درون نمونه نشاندهنده چه تعداد از عناصر موجود در جمعیت است. فرض کنید که ما از قبل یک نمونه انتخاب شده را داریم. از یک دبیرستان با ۵۶۰ دانش‌آموز، نمونه‌ای متشکل از ۲۸ دانش‌آموز انتخاب کرده‌ایم تا بدانیم آیا آنها در خانه

اینترنت دارند یا خیر. تعداد عناصر جمعیت را بر تعداد عناصر نمونه تقسیم و خارج قسمت را به دست می‌آوریم: ۲۰ = ۵۶۰/۲۸ که به این معنی است هر کدام از عناصر نمونه نشاندهنده ۲۰ دانش‌آموز دبیرستانی است.

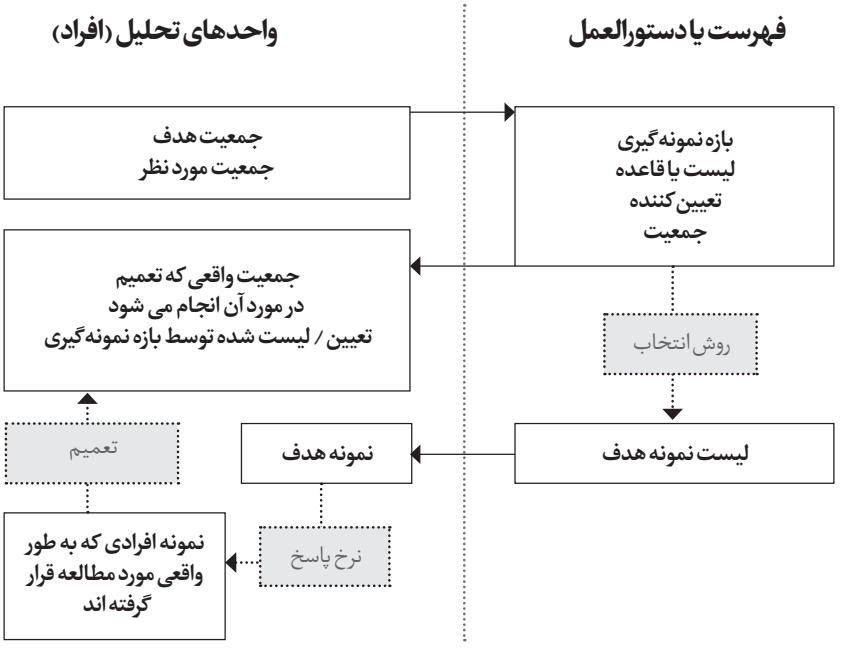
ضریب نمونه‌گیری

این ضریب برابر با خارج قسمت اندازه نمونه و اندازه جمعیت است، n/N. اگر این خارج قسمت در ۱۰۰ ضرب شود، درصد جمعیت نشان داده شده در نمونه را به دست می‌آوریم. برای محاسبه نسبتی از دانش‌آموزان که با آنها مصاحبه می‌کنیم، اندازه نمونه را به اندازه جمعیت تقسیم می‌کنیم، یعنی: 0.05 = 28/560 ، و این بدان معنی است که از ۵٪ جمعیت نظرسنجی می‌کنیم.

مفاهیم کلیدی نمونه‌گیری



فرآیند نمونه‌گیری



روشهای نمونه گیری

نمونه‌گیری احتمالی و غیر احتمالی

نمونه‌گیری احتمالی یک فرایند نمونه‌گیری است که در آن از انواع انتخاب تصادفی استفاده شده است. در نمونه‌گیری احتمالی، هر واحد با احتمال مشخص بیرون کشیده شده، یا این که شانس غیر صفر برای انتخاب شدن در نمونه را دارد. چنین نمونه‌هایی معمولاً با استفاده از اعداد تصادفی انتخاب شده‌اند.

نمونه‌گیری غیراحتمالی یا نمونه‌گیری قضاوتی به قضاوت ذهنی بستگی دارد. روش غیراحتمالی نمونه‌گیری فرایندی است که در آن احتمالات را نمی‌توان به صورت هدفمند به واحدها تخصیص داد و بنابراین تعیین قابلیت اطمینان نتایج نمونه بر اساس احتمال دشوار می‌شود. در نمونه‌گیری غیراحتمالی، در اغلب موارد، ارزیاب بر اساس راحتی خود ماهیت یک نمونه را انتخاب می‌کند. نمونه‌گیری غیراحتمالی برای تحقیق اکتشافی که در آن هدف ایجاد ایده‌های جدیدی است که بعداً به صورت سیستماتیک بررسی خواهند شد مناسب است. البته، اگر هدف این باشد که در مورد یک جمعیت بزرگ یاد بگیریم، اجتناب از قضاوت در مورد نمونه غیراحتمالی در تحقیقات ارزیابی الزامی است. بر خلاف تکنیک‌های نمونه‌گیری احتمالی، هیچ راهی برای شناخت دقت برآورد در نمونه غیراحتمالی وجود ندارد.

با توجه به مواردی که تاکنون مشاهده کردیم می‌توانیم بگوییم که وقتی نمونه نماینده جمعیت نیست، یک نمونه انحرافی داریم.

تا پیش از این روی اهمیت انتخاب صحیح عناصر نمونه به نحوی که نماینده جمعیت ما باشد تاکید کرده‌ایم ولی چگونه می‌توانیم روش‌های مختلف انتخاب نمونه را طبقه‌بندی کنیم؟ می‌توانیم بگوییم سه نوع نمونه‌گیری وجود دارد:

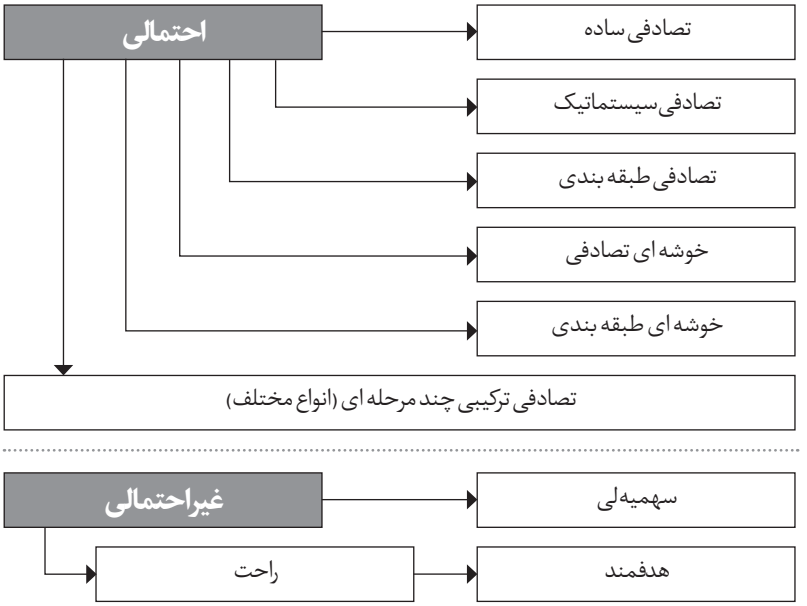
۱. نمونه‌گیری احتمالی: نمونه‌گیری است که در آن هر نمونه احتمال برابر برای انتخاب شدن دارد.
۲. نمونه‌گیری هدفمند: نمونه‌گیری است که در آن فردی که نمونه را انتخاب می‌کند

کسی است که سعی دارد با توجه به نظر یا هدف خودش نمونه نماینده را ایجاد کند و بنابراین ذهنی است.

۳. نمونه‌گیری بدون قاعده: نمونه را بدون هیچ قاعده‌ای انتخاب می‌کنیم و در صورتی که جمعیت همگن باشد و هیچ انحرافی در انتخاب نداشته باشیم، نمونه نماینده است.

ما همیشه نمونه‌گیری را به صورت احتمالی انجام می‌دهیم زیرا اگر روش مناسبی انتخاب کنیم به ما تضمین می‌دهد که نمونه نماینده است و می‌توانیم خطاهای نمونه‌گیری را برآورد کنیم. انواع مختلف نمونه‌گیری احتمالی وجود دارد:

- نمونه‌گیری تصادفی با و بدون جایگزینی.
- نمونه‌گیری طبقه بندی.
- نمونه‌گیری خوشه‌ای.
- نمونه‌گیری سیستماتیک.
- انواع دیگر روش‌های نمونه‌گیری.



پس از این که استفاده از طرح نمونه احتمالی انتخاب شد، لازم است که نوع نمونه‌گیری احتمالی انتخاب شود. چهار نوع اصلی طرح نمونه احتمالی وجود دارد: نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه بندی، نمونه‌گیری سیستماتیک، و نمونه‌گیری خوشه‌ای. نمونه‌گیری تصادفی ساده شناخته شده‌ترین روش نمونه‌گیری احتمالی است. نمونه‌گیری طبقه بندی بهبود قابل توجهی در نمونه‌گیری تصادفی ساده ایجاد می‌کند. نمونه‌گیری سیستماتیک احتمالاً ساده‌ترین روش نمونه‌گیری و نمونه‌گیری خوشه‌ای عملی‌ترین روش برای ارزیابی‌های بزرگ ملی است.

نمونه‌گیری تصادفی ساده

نمونه‌گیری تصادفی ساده چیست؟

نمونه‌گیری تصادفی ساده یک روش نمونه‌گیری احتمالی است که به هر عنصر درون جمعیت هدف، و هر نمونه احتمالی با اندازه مشخص، شانس برابر برای انتخاب شدن را میدهد. بنابراین، نمونه‌گیری تصادفی ساده یک روش انتخاب با احتمال برابر است.

مراحل انتخاب نمونه تصادفی ساده چیست؟

برای انتخاب نمونه تصادفی ساده شش مرحله اصلی وجود دارد.

۱. تعریف جمعیت هدف.
۲. شناسایی بازه نمونه‌گیری موجود در جمعیت هدف یا ایجاد یک بازه جدید.
۳. بررسی بازه نمونه‌گیری از نظر پوشش ناقص، پوشش بیش از حد، پوشش چندگانه، و خوشه‌بندی، و در صورت لزوم انجام اصلاحات.
۴. تخصیص یک عدد منحصر به فرد به هر عنصر درون بازه.
۵. تعیین اندازه نمونه
۶. انتخاب تصادفی تعداد هدفگذاری شده از عناصر جمعیت.

برای انجام مرحله ۶ معمولاً از سه روش استفاده شده است: روش قرعه‌کشی، جدول اعداد تصادفی، و اعدادی که با استفاده از یک برنامه رایانه‌ای به صورت تصادفی ایجاد شده‌اند

(یعنی مولد عدد تصادفی). در استفاده از روش قرعه‌کشی (که به آن «روش بیرون کشیدن کور» و «مدل کلاه» نیز گفته می‌شود)؛ اعداد نشاندهنده هر عنصر درون جمعیت هدف روی مهره‌ها قرار داده می‌شوند (یعنی کارتها، کاغذ، یا اشیا دیگر). سپس مهره‌ها داخل یک ظرف قرار داده و آنها را مخلوط می‌کنیم. سپس، بدون نگاه کردن به ظرف، مهره‌ها را از درون ظرف بیرون می‌آوریم تا این که اندازه نمونه مورد نظر به دست آید. معایب این روش برای انتخاب نمونه این است که زمان‌بر بوده و محدود به جمعیت‌های کوچک است.

از جدول اعداد تصادفی نیز می‌توان استفاده کرد. اعداد درون جدول اعداد تصادفی بدون الگوی مشخصی چیده شده‌اند. آنها را می‌توان به هر روشی قرائت کرد، یعنی افقی، عمودی، قطری، رو به جلو یا رو به عقب. برای استفاده از جدول اعداد تصادفی، محقق باید بدون نگاه کردن به جدول یک نقطه شروع را انتخاب و سپس به صورت سیستماتیکی در ستون اعداد درون جدول رو به جلو (عقب) حرکت می‌کند. تعداد اعداد استفاده شده باید متناسب با اندازه کلی جمعیت هدف باشد. هر عنصری که شماره تخصیص داده شده با آن منطبق با عددی است که محقق به آن می‌رسد به عنوان نمونه انتخاب شده است. اعدادی که محقق به دست می‌آورد ولی با اعداد تخصیص داده شده به عناصر درون جمعیت هدف تطابق ندارند، نادیده گرفته می‌شوند. همانند استفاده از روش قرعه‌کشی، استفاده از جدول اعداد تصادفی نیز یک کار طاقت‌فرسا و زمان‌بر است و برای جمعیت‌های بزرگ توصیه نشده است. در واقع، برای جمعیت‌های بزرگ باید از نرم‌افزار آمار استفاده کرد. بیشتر نرم‌افزارهای آماری و نرم‌افزار صفحه گسترده دارای روش‌هایی برای تولید اعداد تصادفی هستند. عناصری از جمعیت که اعداد تخصیص داده شده به آنها با اعداد ایجاد شده توسط نرم‌افزار مطابقت داشته باشند در نمونه قرار می‌گیرند. فرد می‌تواند یک عدد از جدول اعداد تصادفی را به عنوان عدد شروع فرایند انتخاب کند.

انواع فرعی نمونه‌گیری تصادفی ساده چیست؟

دو نوع نمونه‌گیری تصادفی ساده وجود دارد: نمونه‌گیری با جایگزینی و نمونه‌گیری بدون جایگزینی. در نمونه‌گیری با جایگزینی، بعد از این که یک عنصر از بازه نمونه‌گیری انتخاب شد، به بازه برگردانده شده و می‌تواند دوباره انتخاب شود. در نمونه‌گیری بدون جایگزینی،

بعد از این که عنصر از بازه نمونه‌گیری انتخاب شد، از جمعیت جدا شده و به بازه نمونه‌گیری برگردانده نمیشود. برای تولید نمونه‌های نماینده، نمونه‌گیری بدون جایگزینی کارآمدتر از نمونه‌گیری با جایگزینی است. نمونه بدون جایگزینی امکان ورود چند باره یک عنصر جمعیت را به نمونه نمی‌دهد. نمونه‌گیری بدون جایگزینی متداول‌تر از نمونه‌گیری با جایگزینی است. موضوع این متن نمونه‌گیری بدون جایگزینی است.

نقاط قوت و ضعف نمونه‌گیری تصادفی ساده چیست؟

نمونه‌گیری تصادفی ساده وقتی که با روش‌های نمونه‌گیری غیر احتمالی مقایسه می‌شود، نقاط قوت و ضعف روش‌های نمونه‌گیری احتمالی را دارد. نکته جالب توجه این است که در بین نقاط قوت آن، می‌توان به ایجاد نمونه‌های نماینده، و امکان استفاده از آمارهای استنتاجی برای تحلیل داده‌های جمع‌آوری شده اشاره کرد. در مقایسه با سایر روش‌های نمونه‌گیری احتمالی، نمونه‌گیری تصادفی ساده نقاط قوتی دارد که در زمان انتخاب نوع طرح نمونه احتمالی مورد استفاده بایستی مد نظر قرار گیرند (جدول ۱ را ببینید). برخی از این نقاط قوت عبارتند از:

- نیازی به اطلاعات کمکی پیشرفته در مورد عناصر درون جمعیت نیست. چنین اطلاعاتی برای روش‌های دیگر نمونه‌گیری احتمالی، مثل نمونه‌گیری طبقه بندی، مورد نیاز است.
- هر انتخاب مستقل از انتخاب‌های دیگر است، و هر ترکیب احتمالی از واحدهای نمونه‌گیری شانس برابر و مستقل برای انتخاب شدن دارد. در نمونه‌گیری سیستماتیک، شانس‌های انتخاب شدن مستقل از یکدیگر نیستند.
- برای شناخت و ارتباط با سایرین این روش ساده‌تر از سایر روشهای نمونه‌گیری احتمالی (همانند نمونه‌گیری خوشه‌ای چند مرحله‌ای) است.
- روش‌های آماری لازم برای تحلیل داده‌ها و محاسبه خطاها ساده‌تر از روش‌های آماری لازم برای روش‌های دیگر نمونه‌گیری احتمالی هستند.
- روش‌های آماری برای محاسبه آمارهای استنتاجی در بیشتر نرم‌افزارهای آماری قرار

داده شده و در بیشتر کتاب‌های مقدماتی آمار توضیح داده شده‌اند.

از طرف دیگر، نمونه‌گیری تصادفی ساده نقاط ضعفی هم دارد. در مقایسه با سایر روش‌های نمونه‌گیری احتمالی، نمونه‌گیری تصادفی ساده نقاط ضعف زیر را دارد:

- به بازه نمونه‌گیری عناصر در جمعیت هدف نیاز است. ممکن است بازه نمونه‌گیری مناسب برای جمعیت مورد هدف وجود نداشته باشد یا این که ایجاد این بازه ممکن یا عملی نباشد. روش‌های دیگر نمونه‌گیری، همانند نمونه‌گیری خوشه‌ای، نیازی به بازه نمونه‌گیری عناصر جمعیت هدف ندارد.
- نمونه‌گیری تصادفی ساده خطاهای نمونه‌گیری بزرگتر و دقت کمتری در مقایسه با نمونه‌های طبقه بندی در اندازه نمونه یکسان دارد.
- پاسخ‌دهندگان ممکن است گستردگی زیادی داشته باشد؛ بنابراین هزینه جمع‌آوری داده‌ها ممکن است بیشتر از طرح‌های دیگر نمونه احتمالی، همانند نمونه‌گیری خوشه‌ای، باشد.
- نمونه‌گیری تصادفی ساده ممکن است تعداد کافی از عناصر را در زیرگروه‌های کوچک به دست ندهد. این باعث می‌شود که نمونه‌گیری تصادفی ساده انتخاب خوبی برای مطالعاتی که نیاز به تحلیل مقایسه‌ای گروه‌های کوچکی از جمعیت با گروه‌های بسیار بزرگتر جمعیت دارند خوب نباشد.

جدول ۱) نقاط قوت و ضعف نمونه‌گیری تصادفی ساده در مقایسه با سایر روش‌های نمونه‌گیری احتمالی

نقاط قوت	نقاط ضعف
نیازی به اطلاعات کمکی پیشرفته در مورد عناصر درون جمعیت نیست.	به بازه نمونه‌گیری عناصر در جمعیت هدف نیاز است.
هر ترکیب احتمالی از واحدهای نمونه‌گیری شانس برابر و مستقل برای انتخاب شدن دارد.	نیازی به این ندارد که محقق در مورد جمعیت شناخت داشته باشد.
شناخت و ارتباط ساده‌تر با دیگران.	ممکن است خطاهای نمونه‌گیری بزرگتر و دقت کمتری در مقایسه با سایر طرح‌های نمونه‌گیری احتمالی با اندازه نمونه برابر داشته باشد.
تمایل به ایجاد نمونه‌های نماینده.	اگر زیر گروه‌های جمعیت مورد توجه خاص باشند، ممکن است این زیر گروه‌ها تعداد کافی در نمونه نداشته باشند.
روش‌های آماری لازم برای تحلیل داده‌ها و محاسبه خطاها ساده‌تر هستند.	اگر جمعیت گستردگی زیادی داشته باشد، ممکن است هزینه جمع‌اوری داده‌ها نسبت به سایر روش‌های نمونه احتمالی بیشتر باشد.
روش‌های آماری برای محاسبه استنتاجی در بیشتر نرم‌افزارهای آماری وجود دارند.	ممکن است خیلی پرهزینه باشد، به خصوص وقتی که جمعیت‌ها از نظر جغرافیایی گستردگی داشته باشند و یا به دلیل تغییر نام خانوادگی به دلیل ازدواج یا مهاجرت تعیین مکان افراد دشوار باشد.

نمونه‌گیری طبقه بندی

نمونه‌گیری طبقه بندی چیست؟

نمونه‌گیری طبقه بندی یک روش نمونه‌گیری احتمالی است که در آن جمعیت هدف ابتدا به بخش‌های مانع‌الجمع، همگن (طبقات) تقسیم شده و سپس یک نمونه ساده تصادفی از هر بخش (طبقه) انتخاب شده است. سپس نمونه‌های انتخاب شده از طبقات مختلف در یک نمونه واحد ترکیب می‌شوند. برخی اوقات به این روش نمونه‌گیری، «نمونه‌گیری سهمیه‌ای تصادفی» نیز گفته می‌شود.

مراحل انتخاب یک نمونه طبقه بندی چیست؟

برای انتخاب نمونه تصادفی طبقه بندی هشت مرحله اصلی وجود دارد:

۱. تعریف جمعیت هدف.

۲. شناسایی متغیر(های) طبقه بندی و تعیین تعداد طبقاتی که قرار است مورد استفاده قرار گیرند. متغیرهای طبقه بندی باید با اهداف مطالعه ارتباط داشته باشند. اگر هدف تحقیق این است که برآوردهای زیرگروه را انجام دهد، متغیرهای طبقه بندی باید با این زیر گروه‌ها مرتبط باشد. غالباً در دسترس بودن اطلاعات کمکی تعیین کننده متغیرهای طبقه بندی هستند که استفاده می‌شوند. ممکن است به بیش از یک متغیر طبقه بندی نیاز باشد. البته، برای تامین منافع مورد نظر، این متغیرها باید با متغیرهای مورد نظر در مطالعه مرتبط بوده و مستقل از هم باشند. توجه داشته باشید که با افزایش تعداد متغیر طبقه بندی، احتمال این که برخی از متغیرها تاثیر روی سایر متغیرها را خنثی کنند نیز افزایش خواهد یافت، بنابراین نباید بیش از چهار تا شش متغیر طبقه بندی و بیش از شش طبقه برای یک متغیر خاص استفاده شود.

۳. شناسایی یک بازه نمونه‌گیری موجود یا ایجاد یک بازه نمونه‌گیری که شامل اطلاعاتی در مورد متغیر(های) طبقه بندی برای هر عنصر درون جمعیت هدف باشد. در صورتی که بازه نمونه‌گیری شامل اطلاعات راجع به متغیرهای طبقه بندی نباشد، طبقه بندی نمی‌تواند انجام شود.

۴. بررسی بازه نمونه‌گیری از نظر پوشش ناقص، پوشش بیش از حد، پوشش چندگانه، و خوشه‌بندی، و در صورت لزوم انجام اصلاحات.

۵. تقسیم بازه نمونه‌گیری با طبقات، طبقه‌بندی متغیرهای طبقه بندی، ایجاد بازه نمونه‌گیری برای هر طبقه. اختلاف درون طبقه باید به حداقل رسانده شود، و اختلاف بین طبقات نیز باید حداکثر شود. طبقات نباید همپوشانی داشته باشند و همراه با هم باید کل جمعیت را تشکیل دهند. طبقات باید مستقل و زیر مجموعه‌های مانع‌الجمعی از جمعیت باشند. هر عنصر جمعیت باید در یک و فقط یک طبقه قرار داشته باشد.
۶. تخصیص یک عدد منحصر به فرد به هر عنصر.
۷. تعیین اندازه نمونه برای هر طبقه. توزیع عددی عناصر نمونه‌گیری شده بین طبقات مختلف تعیین کننده نوع نمونه طبقه بندی اجرا شده است. این نمونه‌گیری می‌تواند به صورت نمونه‌گیری طبقه بندی تناسبی یا یکی از انواع نمونه‌گیری طبقه بندی غیر تناسبی باشد.
۸. انتخاب تصادفی تعداد هدفگذاری شده عناصر از هر طبقه. از هر طبقه حداقل باید یک عنصر به عنوان نماینده درون نمونه انتخاب شود؛ و برای محاسبه حاشیه خطای برآوردهای محاسبه شده با استفاده از داده‌های جمع‌اوری شده حداقل باید دو عنصر از هر طبقه انتخاب شود.

انواع نمونه‌گیری طبقه بندی چیست؟

دو نوع اصلی نمونه‌گیری طبقه بندی وجود دارد: نمونه‌گیری طبقه بندی تناسبی و نمونه‌گیری طبقه بندی غیرتناسبی (شکل ۱ را ببینید). نمونه‌گیری طبقه بندی غیرتناسبی دارای زیر گروه‌های مختلفی است.

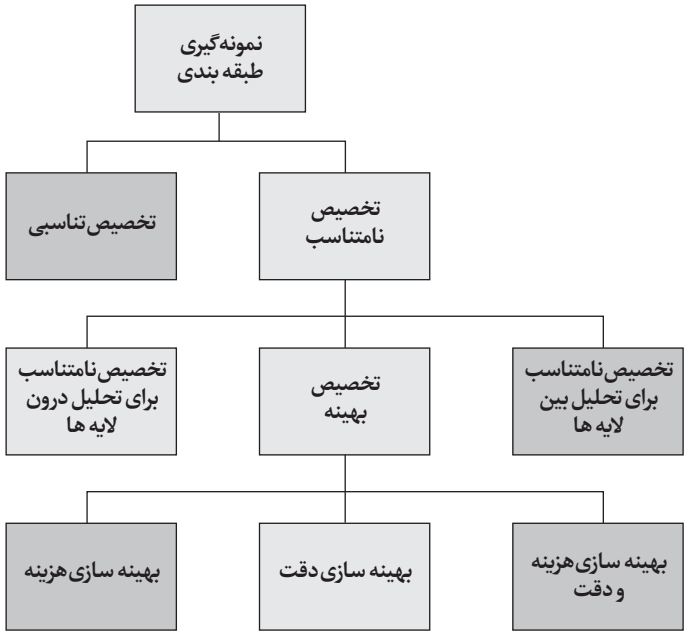
■ نمونه‌گیری طبقه بندی تناسبی

در نمونه‌گیری طبقه بندی تناسبی، تعداد عناصر تخصیص داده شده به هر طبقه متناسب با نمایندگی طبقه در جمعیت هدف است. یعنی این که، اندازه نمونه استخراج شده از هر طبقه متناسب با اندازه نسبی آن طبقه در جمعیت هدف است. بنابراین، این روش یک روش نمونه‌گیری خود وزن‌گذار و EPSEM است. با استفاده از یک روش نمونه‌گیری برای هر طبقه، عناصر درون جمعیت شانس برابر برای انتخاب شدن خواهند داشت. نمونه به دست آمده یک نمونه خود وزن‌گذار است. از این روش نمونه‌گیری زمانی استفاده می‌شود که هدف

از تحقیق برآورد پارامترهای جمعیت باشد.

یک مثال فرضی برای تخصیص تناسبی در جدول ۲ نشان داده شده است. در این مثال، عناصر نمونه‌گیری شده در چهار ناحیه متمایز بازاریابی قرار داده شده‌اند و بنابراین نسبت عناصر نمونه‌گیری شده برای هر ناحیه برابر با نسبت عناصر هر ناحیه به کل جمعیت است. نسبت نمونه‌گیری در هر طبقه برابر یک بر ۲۲ نمونه است. هر ناحیه به صورت برابر در نمونه نشان داده شده است.

شکل ۱) زیر نمونه‌های نمونه‌گیری طبقه بندی براساس تخصیص طبقات



برخی مواقع، محقق ممکن است قصد داشته باشد که علاوه بر برآورد پارامترهای جمعیت، تحلیل جامعی در درون یک طبقه نسبتاً کوچک داشته باشد و یا این که طبقات را با هم مقایسه کند. در برخی از طبقات نمونه‌گیری طبقه بندی تناسبی نمی‌تواند تعداد کافی

حالت برای چـنین تحـلیل‌هایی به دسـت بدهد. به عنوان نمونه، مثال توضـیح داده شـده در جدول ۲ را در نظر بگیرید، این امکان وجود ندارد که تحلیل جامعی در مورد عناصر ناحیه ۲ انجام داد زیرا در نمونه فقط ۱۲ عتصر وجود دارد. به علاوه، مقایسه عناصر ناحیه ۲ با عناصر نواحی دیگر نیز جای تردید دارد. نمونه‌گیری طبقه بندی تناسبی یک نمونه‌گیری ضعیف برای انجام چنین تحلیل‌هایی است. نمونه‌گیری طبقه بندی غیر تناسبی ممکن است انتخاب بهتری باشد.

جدول ۲) مثال در مورد نمونه‌گیری لایه‌ای تناسبی

ناحیه بازاریابی	جمعیت		نمونه طبقه بندی تناسبی	
	تعداد	درصد	تعداد	درصد
ناحیه۱	۱۸۰۰۰	٪۳۳	۳۹۶	٪۳۳
ناحیه۲	۶۰۰	٪۱	۱۲	٪۱
ناحیه۳	۱۲۰۰۰	٪۲۲	۲۶۴	٪۲۲
ناحیه۴	۲۴۰۰۰	٪۴۴	۵۲۸	٪۴۴
جمع	۵۴۶۰۰	٪۱۰۰	۱۲۰۰	٪۱۰۰

■ **نمونه‌گیری طبقه بندی غیرتناسبی**

نمونه‌گیری طبقه بندی غیرتناسبی یک روش نمونه‌گیری طبقه بندی است که در آن تعداد عناصر نمونه‌گیری شده از هر طبقه، تناسبی با حضور آن در جمعیت کلی ندارد. عناصر جمعیت شانس برابر برای قرار گرفتن در نمونه را ندارند. برای هر طبقه از یک نسبت نمونه‌گیری استفاده نشده است. از طرف دیگر، طبقات دارای نسبت‌های نمونه‌گیری متفاوت بوده و بنابراین این روش نمونه‌گیری یک روش نمونه‌گیری EPSEM نیست. برای برآورد پارامترهای جمعیت، از ترکیب جمعیت بایستی به عنوان وزنها استفاده شود تا عدم تناسب در نمونه جبران شود. البته، در برخی از پروژه‌ها، نمونه‌گیری طبقه بندی غیرتناسبی

ممکن است بهتر از نمونه‌گیری طبقه بندی تناسبی باشد.

نمونه‌گیری طبقه بندی غیرتناسبی می‌تواند بر اساس هدف تخصیص انجام شده به سه زیر گروه تقسیم شود. هدف از تخصیص می‌تواند تسهیل تحلیل درون طبقه، تحلیل بین طبقات، یا تخصیص بهینه باشد. در تخصیص بهینه ممکن است روی بهینه‌سازی هزینه‌ها، بهینه‌سازی دقت، یا بهینه‌سازی دقت و هزینه‌ها تمرکز شده باشد.

تخصیص غیر تناسبی برای تحلیل درون طبقات : ممکن است با توجه به هدف تحقیق محقق ملزم باشد تحلیل‌های جامعی درون طبقات نمونه انجام دهد. در صورت استفاده از طبقه بندی تناسبی، اندازه نمونه یک طبقه بسیار کوچک است و بنابراین تامین اهداف تحقیق می‌تواند دشوار باشد. تخصیص تناسبی ممکن است تعداد حالت‌های کافی برای چنین تحلیل‌های جامعی در اختیار قرار ندهد. یک گزینه این است که از طبقه کوچک یا نادر نمونه بیش از اندازه گرفته شود. این نمونه‌گیری بیش از حد باعث توزیع غیرتناسبی طبقه در نمونه در مقایسه با جمعیت می‌شود.

با وجود این، ممکن است تعداد حالت‌های کافی برای تحلیل‌های درون طبقه بندی لازم که با توجه به اهداف تحقیق مورد نیاز هستند ایجاد شوند. مثال‌های تحقیقاتی که در آنها این روش نمونه‌گیری می‌تواند مناسب باشد عبارتند از مطالعه روی افرادی با یک بیماری نادر، یا مطالعه روی افرادی که بیشتر دوران جوانی خود را صرف مراقبت از کودکان بی‌سرپرست کرده‌اند. با استفاده از مثال فرضی توضیح داده شده در جدول ۲، اگر هدف این باشد که تحلیل جامعی در مورد ناحیه ۲ انجام دهیم، بایستی عناصر این ناحیه را بیش از حد نمونه‌گیری کنیم؛ به عنوان مثال به جای نمونه‌گیری فقط ۱۲ عنصر، ۱۳۰ عنصر نمونه‌گیری کنیم. برای انجام یک تحلیل معنادار و جامع در ناحیه ۲، اندازه نمونه برای این ناحیه باید بزرگتر از ۱۲ عنصر باشد. توزیع حاصل از عناصر نمونه بر اساس ناحیه ممکن است شبیه به توزیع نشان داده شده در جدول ۳ باشد.

تخصیص غیر تناسبی برای تحلیل بین طبقات : با توجه به هدف تحقیق ممکن است لازم باشد محقق طبقات را با هم مقایسه کند. در این صورت، برای هر گروه باید تعداد مناسبی از

عناصر انتخاب شوند. ممکن است محقق قصد داشته باشد اندازه نمونه هر طبقه را حداکثر کند. برای چنین تحقیقی، تخصیص برابر (که به آن «تخصیص متعادل» و «نمونه‌گیری فاکتوریل» نیز گفته می‌شود) می‌تواند مناسب باشد. محقق ممکن است به دنبال این باشد که تعداد برابری از عناصر را از هر طبقه انتخاب کند.

یک نمونه از تخصیص غیرتناسبی با تخصیص برابر در جدول ۴ نشان داده شده است. در این مثال، عناصر نمونه‌گیری شده بین چهار ناحیه از مثال فرضی ما تخصیص داده شده‌اند بنابراین تعداد عناصر نمونه‌گیری شده برای هر ناحیه برابر است. در مقایسه با توزیع تناسبی نمونه جدول ۲، تخصیص‌های نمونه‌گیری نشان داده شده در جدول ۴ حداقل تعداد عناصر را برای هر ناحیه در اختیار قرار می‌دهد، و تحلیل مقایسه‌ای متعادل‌تری بین نواحی ایجاد می‌کند. به علاوه، تعداد نسبتاً زیادی از عناصر از ناحیه ۲ نمونه‌گیری شده که امکان تحلیل جامع درون آن طبقه را فراهم می‌سازد.

جدول ۳) مثال در مورد تخصیص نا متناسب نمونه‌گیری طبقه بندی

ناحیه بازاریابی	جمعیت		نمونه طبقه بندی تناسبی	
	تعداد	درصد	تعداد	درصد
ناحیه۱	۱۸۰۰۰	٪۳۳	۳۵۷	٪۳۰
ناحیه۲	۶۰۰	٪۱	۱۳۰	٪۱۱
ناحیه۳	۱۲۰۰۰	٪۲۲	۲۳۸	٪۲۰
ناحیه۴	۲۴۰۰۰	٪۴۴	۴۷۵	٪۳۹
جمع	۵۴۶۰۰	٪۱۰۰	۱۲۰۰	٪۱۰۰

جدول ۴) مثال در مورد تخصیص نامتناسب نمونه‌گیری طبقه بندی برای تسهیل تحلیل بین طبقات

ناحیه بازاریابی	جمعیت		نمونه طبقه بندی تناسبی	
	تعداد	درصد	تعداد	درصد
ناحیه۱	۱۸۰۰۰	٪۳۳	۳۰۰	٪۲۵
ناحیه۲	۶۰۰	٪۱	۳۰۰	٪۲۵
ناحیه۳	۱۲۰۰۰	٪۲۲	۳۰۰	٪۲۵
ناحیه۴	۲۴۰۰۰	٪۴۴	۳۰۰	٪۲۵
جمع	۵۴۶۰۰	٪۱۰۰	۱۲۰۰	٪۱۰۰

تخصیص بهینه : اگرچه نمونه‌گیری طبقه بندی تناسبی می‌تواند حاشیه خطای کمتری نسبت به نمونه‌گیری تناسبی ساده برای برآورد پارامترهای جمعیت به دست بدهد، ولی این وضعیت می تواند بهتر نیز شود. تخصیص بهینه برای دستیابی به دقت کلی بیشتر نسبت به آن چه با استفاده از طراحی نمونه‌گیری طبقه بندی تناسبی بدست می آید، در نظر گرفته شده است. تخصیص بهینه اندازه نمونه طبقاتی مختلف را تعیین کرده و دو جنبه مهم برای انجام تحقیق را در نظر می‌گیرد: هزینه‌ها و دقت. نسبت نمونه‌گیری بر اساس هزینه‌ها و تغییرپذیری درون طبقاتی مختلف تغییر می‌کند. نمونه‌گیری طبقه بندی غیرتناسبی، به خصوص، تخصیص بهینه، وقتی که طبقات بر اساس هزینه‌های جمع‌آوری داده و تغییرپذیری متغیرهای مورد نظر با هم تفاوت داشته باشند ممکن است برای تحقیق بهتر از نمونه‌گیری طبقه بندی تناسبی باشد. تخصیص بهینه می‌تواند روی فقط هزینه، فقط دقت، یا هم هزینه و هم دقت تمرکز کند.

طبقاتی همگن با اندازه نمونه کوچکتر می‌توانند دقتی برابر با طبقاتی ناهمگن با اندازه نمونه بزرگتر داشته باشند. بر اساس این اصل، بهتر این است که تعداد عناصر انتخاب شده از هر طبقه ارتباط مستقیمی با انحراف معیار متغیر مورد نظر در طبقه داشته باشد. هر چه

تغییرپذیری متغیر درون طبقه بیشتر باشد، اندازه نمونه نیز باید بزرگتر باشد. به علاوه، با در نظر گرفتن هزینه‌های جمع‌آوری داده، هر چه هزینه‌های جمع‌آوری داده یک طبقه بالاتر باشد، اندازه نمونه هدفگذاری شده کوچکتر است.

داده‌های فرضی نشان دهنده تخصیص بهینه در جدول ۵ نشان داده شده‌اند.

داده‌های فرضی نشان داده شده در جدول ۵ نشان می‌دهند که هزینه‌های جمع‌آوری داده در چهار ناحیه بین ۱۰ تا ۳۹ دلار به ازای هر واحد است (ستون ۴ را ببینید). تفاوت‌ها در توزیع جغرافیایی عناصر طبقاتی مختلف می‌توانند عامل این اختلاف‌ها باشند. توزیع اندازه‌های نمونه برای طبقات مختلف در ستون ۷ این تغییر در هزینه‌های جمع‌آوری داده‌ها را نشان می‌دهد.

انحراف معیار فرضی متغیر مورد نظر برای چهار ناحیه در ستون ۵ جدول ۵ نشان داده شده است. انحراف‌های استاندارد بین ۴,۳ تا ۹,۴ هستند. اگر هزینه‌های جمع‌آوری داده برای نواحی مختلف در دسترس نباشند یا در واقع برابر باشند، فرد می‌تواند با تخصیص اندازه نمونه هر طبقه و با در نظر گرفتن تغییرپذیری طبقات، اندازه‌های نمونه را بهینه کند. این نوع تخصیص اولین بار توسط جرزی نیمان (۱۹۳۴) پیشنهاد و اغلب به آن تخصیص نیمان گفته می‌شود. توزیع اندازه نمونه طبقاتی مختلف در مثال فرضی، با در نظر گرفتن تغییرپذیری طبقات، در ستون ۸ نشان داده شده است. استفاده از این روش بهینه‌سازی به داده‌های مربوط به تغییرپذیری متغیر مورد نظر برای طبقاتی مختلف دارد. غالباً این داده‌ها در دسترس نیست. به علاوه، اگر مطالعه اهداف مختلف و بیش از یک متغیر مورد نظر داشته باشید، بهینه‌سازی آنها ممکن است با یکدیگر تعارض داشته باشند.

ناحیه بازاریابی	جمعیت تعداد	درصد جمعیت توزیع	هزینه جمع آوری داده ها واحد (j)	تغییرپذیری	$\frac{S}{\sqrt{C_j}}$	نمونه بهینه‌سازی اندازه هزینه ها	نمونه بهینه‌سازی اندازه تغییرپذیری	نمونه بهینه‌سازی هزینه‌ها و تغییرپذیری
ناحیه۱	۱۸۰۰۰	٪۳۳	۱۸ دلار	۴/۳	۱/۰۱۴	۳۰۰	۱۹۰	۲۰۳
ناحیه۲	۶۰۰	٪۱	۱۰ دلار	۶/۴	۲/۰۲۴	۵۳۸	۲۸۲	۴۰۵
ناحیه۳	۱۲۰۰۰	٪۲۲	۳۹ دلار	۹/۴	۱/۵۰۵	۱۳۸	۴۱۵	۳۰۲

ناحیه۴	۲۴۰۰۰	٪۴۴	۲۴ دلار	۷/۱	۱/۴۴۹	۲۲۴	۳۱۳	۲۹۰
جمع	۵۴۶۰۰	٪۱۰۰				۱۲۰۰	۱۲۰۰	۱۲۰۰

اگر داده‌های مربوط به هزینه‌های جمع‌آوری داده و تغییرپذیری متغیر مورد نظر در دسترس باشد، در این صورت باید هزینه‌ها و دقت را بهینه کرد. ضریب وزن‌گذاریی که هزینه‌های جمع‌آوری داده و انحراف معیار را مد نظر دارد را می‌توان به صورت $\frac{s}{\sqrt{j}}$ محاسبه کرد که در آن "s" نشاندهنده انحراف معیار درون طبقات، و "j" هزینه‌های جمع‌آوری داده به ازای هر واحد درون طبقه است. توزیع این ضریب برای مثال فرضی ما در ستون ۶ جدول قبل نشان داده شده است. با در نظر گرفتن این ضریب، اندازه نمونه طبقاتی مختلف در نظر گرفتن هزینه‌های جمع‌آوری داده و تغییرپذیری با طبقه بهینه شده و در ستون ۹ نشان داده شده است.

نقاط قوت و ضعف نمونه‌گیری طبقه بندی چیست؟

نمونه‌گیری طبقه بندی همانند بسیاری از روش‌های نمونه‌گیری احتمالی در مقایسه با روش‌های نمونه‌گیری غیراحتمالی، دارای نقاط قوت و ضعف است. برای تعیین این که آیا از نمونه‌گیری طبقه بندی استفاده شود یا خیر، بایستی نقاط قوت و ضعف آن را با نمونه‌گیری تصادفی ساده مقایسه کرد (جدول ۶ را ببینید). در مقایسه با نمونه‌گیری تصادفی ساده، نقاط قوت نمونه‌گیری طبقه بندی عبارتند از:

- توانایی برآورد پارامترهای جمعیت و نتایج درون طبقات و مقایسه بین طبقات. داده‌های کافی در مورد زیر گروه‌ها را نمی‌توان با استفاده از روش نمونه‌گیری تصادفی ساده به دست آورد. نمونه‌های طبقه بندی نسبت به نمونه‌های به دست آمده از یک نمونه تصادفی ساده با همان اندازه، به خصوص در صورتی که از تخصیص بهینه استفاده شده باشد، خطاهای نمونه تصادفی کمتری ایجاد خواهد کرد. طبقه بندی باعث افزایش دقت و حذف تغییر متغیرهای مورد استفاده برای طبقه بندی می‌شود. مقدار افزایش دقت با استفاده از این که مقدار تغییرات درون طبقه بندی متغیرهای مورد مطالعه تا چه اندازه به حداقل و تغییر متغیرهای مورد مطالعه بین طبقات تا چه اندازه حداکثر شده تعیین شده است. طبقه

بندی نمونه‌ای را به دست خواهد که در اندازه نمونه برابر حداقل دقتی برابر با نمونه تصادفی ساده به دست خواهد داد. اگر این کار باعث افزایش سطح دقت نشود، در صورت استفاده از نمونه‌گیری تصادفی ساده نتایج را بدتر نخواهد کرد.

■ نمونه‌های طبقه بندی شده نمایندگی بیشتر جمعیت را دارند زیرا آنها تضمین می‌کنند که عناصر هر طبقه جمعیت در نمونه نشان داده شده‌اند. برای اطمینان از این که نمونه روی زیر نواحی جغرافیایی و زیر گروه‌های جمعیت گسترده شده‌اند، نمونه‌گیری می‌تواند طبقه بندی شود.

■ در استفاده از نمونه‌گیری طبقه بندی، مزیت این است که محقق در مورد جمعیت شناخت داشته باشد.

■ اگر متغیرهای طبقه بندی جمعیت را به نواحی جغرافیایی تقسیم کنیم، هزینه‌های جمع‌آوری داده‌ها ممکن است کمتر از نمونه‌گیری تصادفی ساده باشد.

■ استفاده از نمونه‌گیری طبقه بندی به محقق این امکان را می‌دهد تا از روش‌های مختلف نمونه‌گیری در طبقاتی مختلف استفاده کند.

■ برای استفاده از نمونه‌گیری طبقه بندی، محقق می‌تواند برای انجام مطالعه راحتی کارهای مدیریتی را مد نظر داشته باشد. محقق می‌تواند خوشه‌بندی جمعیت بر اساس مناطق شهری، بخش‌های سازماندهی شده جمعیت، و توزیع کارکنان جمع‌آوری کننده داده‌ها را مد نظر قرار دهد.

در مقایسه با نمونه‌گیری تصادفی ساده، نقاط ضعف نمونه‌گیری طبقه بندی عبارتند از:

■ نمونه‌گیری طبقه بندی در مقایسه با نمونه‌گیری تصادفی ساده الزام بیشتری به اطلاعات کمکی دارد. برای هر عنصر درون جمعیت به اطلاعاتی در مورد متغیرهای طبقه بندی نیاز است. برخی از این اطلاعات عبارتند از نسبتی از جمعیت هدف که به هر طبقه تعلق دارد؛ در صورتی که از تخصیص بهینه استفاده شده باشد اطلاعات مربوط به تغییرپذیری و هزینه‌های جمع‌آوری داده برای هر طبقه مورد نیاز هستند. کسب چنین اطلاعاتی ممکن

است زمان‌بر و پر هزینه باشد.

- در صورتی که تحقیق شامل تعداد زیادی متغیر باشد، انتخاب متغیرهای طبقه بندی می‌تواند دشوار باشد. این متغیرها باید با متغیرهای مورد نظر در تحقیق همبستگی داشته باشد.

- در نمونه‌گیری طبقه بندی برای آماده شدن برای نمونه‌گیری، اجرای طرح نمونه، و تحلیل داده‌های جمع‌آوری شده نیاز به کار بیشتری دارد.

- برای محاسبه برآوردهای نمونه‌گیری، حداقل دو عنصر از هر طبقه باید انتخاب شوند.

- تحلیل داده‌های جمع‌آوری شده در مقایسه با داده‌های جمع‌آوری شده با استفاده از روش نمونه‌گیری تصادفی ساده پیچیده‌تر است.

- طبقه‌بندی اشتباه عناصر به طبقات باعث افزایش تغییرپذیری می‌شود.

- در صورتی که از تخصیص غیر تناسبی استفاده شده باشد، برای برآورد پارامترهای جمعیت، داده‌های جمع‌آوری شده باید اصلاح (وزن‌گذاری) شوند. وزن‌گذاری باعث می‌شود که برخی از برآوردهای جمعیت دقت کمتری داشته باشند.

جدول ۶) نقاط قوت و ضعف نمونه‌گیری طبقه بندی در مقایسه با نمونه‌گیری تصادفی ساده

نقاط قوت	نقاط ضعف
برخلاف نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه بندی:	برخلاف نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه بندی:
توانایی بیشتری برای استنتاج درون یک طبقه و مقایسه بین طبقات دارد.	نیاز به اطلاعات در مورد نسبت مجموع جمعیتی که به هر طبقه تعلق دارد، دارد.

نقاط قوت	نقاط ضعف
دارای خطاهای نمونه‌گیری تصادفی کمتری برای نمونه‌های با اندازه برابر است، بنابراین برای حاشیه خطای برابر نیاز به نمونه‌هایی با اندازه کوچکتر دارد.	برای هر عنصر درون جمعیت نیاز به اطلاعاتی در مورد متغیرهای طبقه بندی است. در صورتی که چنین اطلاعاتی به راحتی در دسترس نباشد، جمع‌آوری این اطلاعات می‌توانند هزینه‌بر باشد.
نمونه نماینده‌تری به دست می‌دهد زیرا تضمین می‌کند که هر عنصر از هر طبقه در نمونه نشان داده شده است.	نسبت به نمونه‌گیری تصادفی ساده هزینه بیشتر، زمان بیشتر و پیچیده‌تر است.
شناخت محقق در مورد جمعیت مزیت بیشتری دارد.	در صورتی که تحقیق شامل تعداد زیادی متغیر باشد، انتخاب متغیرهای طبقه بندی ممکن است دشوار باشد.
اگر متغیرهای طبقه بندی جمعیت را به نواحی جغرافیایی تقسیم کنیم، هزینه‌های جمع‌آوری داده‌ها ممکن است کمتر از نمونه‌گیری تصادفی ساده باشد.	برای محاسبه برآوردهای نمونه‌گیری، حداقل دو عنصر از هر طبقه باید انتخاب شوند.
این امکان را فراهم می‌کند که در طبقاتی مختلف از روش‌ها و دستورالعمل‌های مختلف استفاده شود.	تحلیل داده‌های جمع‌آوری شده در مقایسه با داده‌های جمع‌آوری شده با استفاده از روش نمونه‌گیری تصادفی ساده پیچیده‌تر است.
امکان تحلیل الگوهای درون طبقه و گزارشگری مجزای نتایج هر طبقه را فراهم می‌کند.	در صورتی که از تخصیص غیرتناسبی استفاده شده باشد، برای برآورد دقیق پارامترهای جمعیت، داده‌های جمع‌آوری شده باید وزن‌گذاری شوند.

تفاوت بین نمونه‌گیری طبقه بندی و نمونه‌گیری سهمیه‌ای چیست؟

نمونه‌گیری طبقه بندی و نمونه‌گیری سهمیه‌ای تاحدودی مشابه هم هستند. هر دو جمعیت هدف را به گروه‌ها تقسیم کرده و سپس تعداد مشخصی از عنصر را از هر گروه انتخاب می‌کنند (جدول ۷ را ببینید).

در هر دو روش هدف اولیه این است که نمونه نماینده انتخاب و یا تحلیل زیر گروه‌ها

تسهیل شود. البته تفاوت‌های مهمی نیز وجود دارد. نمونه‌گیری طبقه بندی پس از این که گروه‌ها ایجاد شدند از نمونه‌گیری تصادفی ساده استفاده می‌کند؛ نمونه‌گیری سهمیه‌ای از نمونه‌گیری در دسترس استفاده می‌کند. برای نمونه‌گیری طبقه بندی به بازه نمونه‌گیری نیاز است ولی برای نمونه‌گیری سهمیه‌ای نیازی به بازه نمونه‌گیری نیست. از همه مهمتر این که، نمونه‌گیری طبقه بندی یک روش نمونه‌گیری احتمالی است که امکان برآورد خطای نمونه‌گیری را فراهم می‌کند. این برآورد با استفاده از نمونه سهمیه‌ای میسر نیست.

جدول ۷) مقایسه نمونه‌گیری طبقه بندی و نمونه‌گیری سهمیه‌ای

نمونه‌گیری طبقه بندی	نمونه‌گیری سهمیه‌ای
نمونه‌گیری طبقه بندی و نمونه‌گیری سهمیه‌ای در رابطه با موارد زیر مشابهت دارند:	
جمعیت به گروه‌ها تقسیم شده است؛ سپس عناصر از هر گروه انتخاب شده‌اند.	جمعیت به گروه‌ها تقسیم شده است؛ سپس عناصر از هر گروه انتخاب شده‌اند.
هدف این است که یک نمونه نماینده انتخاب و یا تحلیل زیر گروه تسهیل شود.	هدف این است که یک نمونه نماینده انتخاب و یا تحلیل زیر گروه تسهیل شود.
نمونه‌گیری طبقه بندی و نمونه‌گیری سهمیه‌ای در رابطه با موارد زیر با هم تفاوت دارند:	
عناصر درون هر گروه با استفاده از نمونه‌گیری تصادفی ساده انتخاب شده‌اند و در نتیجه:	عناصر درون هر گروه با استفاده از در دسترس بودن نمونه‌گیری انتخاب شده‌اند و بنابراین:
به بازه نمونه‌گیری نیاز است.	به بازه نمونه‌گیری نیازی نیست.
خطای نمونه‌گیری تصادفی را می‌توان برآورد کرد.	خطای نمونه‌گیری تصادفی را نمی‌توان برآورد کرد.
یکسونگری انتخاب حداقل شده است.	یکسونگری انتخاب حداقل نشده است.
هدف این است که خطای نمونه‌گیری کاهش یابد.	

نمونه‌گیری سیستماتیک

نمونه‌گیری سیستماتیک چیست؟

نمونه‌گیری سیستماتیک (یا نمونه‌گیری تصادفی فاصله‌ای) یک روش نمونه‌گیری احتمالی است که در آن انتخاب تصادفی از اولین عنصر شروع می‌شود و عناصر بعدی با استفاده از یک فاصله ثابت یا سیستماتیک انتخاب می‌شوند تا این که به اندازه نمونه مورد نظر دست پیدا کرد. شروع تصادفی وجه تمایز این روش نمونه‌گیری با روش نمونه‌گیری غیراحتمالی، نمونه‌گیری سیستماتیک غیراحتمالی است. برخی مواقع، از بازه نمونه‌گیری استفاده نشده است. لازم نیست که جمعیت هدف شماره‌گذاری شود و در صورتی که نمایش فیزیکی همانند جریان مستمر عناصر جمعیت در محل‌های خاصی وجود داشته باشد، بازه نمونه‌گیری ایجاد می‌شود. به عنوان مثال، بعد از شروع تصادفی، فرد باید به صورت سیستماتیکی هر iامین بیمار در بخش اورژانس یک بیمارستان، مشتریان فروشگاهی که به صف ایستاده‌اند، یا پرونده‌های درون کشوهای فایل را انتخاب کند.

مراحل انتخاب نمونه سیستماتیک چیست؟

به طور کلی، برای انتخاب یک نمونه به صورت سیستماتیک هشت مرحله اصلی وجود دارد:

۱. تعریف جمعیت هدف.
۲. تعیین اندازه نمونه مورد نظر (n).
۳. شناسایی بازه نمونه‌گیری فعلی یا ایجاد یک بازه نمونه‌گیری برای جمعیت هدف.
۴. ارزیابی بازه نمونه‌گیری از نظر پوشش ناقص، پوشش بیش از حد، پوشش چندگانه، خوشه‌بندی و تناوب و در صورت لزوم انجام اصلاحات. در حالت ایده‌ال، لیست به صورت تصادفی و بر اساس متغیر مورد مطالعه مرتب شده است یا این که بهتر این است که بر حسب متغیر مورد نظر یا متغیر وابسته به آن و در نتیجه ایجاد طبقه بندی صریح مرتب شود. اگر بازه نمونه‌گیری تصادفی باشد، نمونه‌گیری سیستماتیک تقریب خوبی از نمونه‌گیری تصادفی سیستماتیک خواهد بود.
۵. تعیین تعداد عناصر درون بازه نمونه‌گیری (N).
۶. محاسبه فاصله نمونه‌گیری (i) با تقسیم کردن تعداد عناصر در بازه نمونه‌گیری (N)

بر اندازه نمونه هدف (n). فرد باید باقیمانده را نادیده گرفته و عدد را باید به سمت پایین و نزدیکترین عدد روند یا گرد کند. روند کردن به سمت پایین یا گرد کردن باعث می‌شود که اندازه نمونه بزرگتر از اندازه مورد نظر باشد. در این صورت فرد باید به صورت تصادفی انتخاب‌های اضافی را حذف کند. در صورتی اندازه دقیق جمعیت مشخص نباشد و تعیین آن عملی نباشد، فرد باید نسبت نمونه‌گیری را ثابت کند.

۷. انتخاب تصادفی یک عدد، r، از «۱» تا i.

۸. انتخاب برای نمونه r ، r + i ، r + 2i ، r + 3i و الی آخر، تا این که بازه تکمیل شود.

در سطح فنی، نمونه‌گیری سیستماتیک یک نمونه کاملاً تصادفی را ایجاد نمی‌کند. به این نوع نمونه‌گیری، «نمونه‌گیری تصادفی کاذب»، «نمونه‌گیری تصادفی ساده کاذب» یا «نمونه‌گیری نیمه تصادفی» نیز می گویند. فقط انتخاب اولین عنصر در نمونه‌گیری سیستماتیک یک انتخاب احتمالی است. پس از این که اولین عنصر انتخاب شد، بقیه عناصر احتمال صفر برای انتخاب شدن دارند. به علاوه، ترکیبات خاصی از عناصر، همانند عناصری که در بازه نمونه‌گیری کنار هم قرار دارند، احتمالاً انتخاب نمی‌شوند. برای رفع این مشکل می‌توان از نمونه‌گیری سیستماتیکی تکراری، که در ادامه توضیح داده شده، استفاده کرد.

زیرنمونه‌های نمونه‌گیری سیستماتیک چیست؟

نمونه‌گیری سیستماتیک را می‌توان به سه دسته اصلی تقسیم کرد: نمونه‌گیری سیستماتیک خطی، نمونه‌گیری سیستماتیک دایره‌ای، و نمونه‌گیری سیستماتیک تکراری (تکرار شده). نمونه‌گیری سیستماتیک خطی متداولترین شکل نمونه‌گیری سیستماتیک است. مراحل انتخاب یک نمونه سیستماتیک خطی همان هایی هستند که در بالا گفته شد. نمونه‌گیری سیستماتیک دایره‌ای را می‌توان به عنوان زیر نمونه‌ای از نمونه‌گیری سیستماتیک خطی در نظر گرفت. برای استفاده از این روش، در مرحله ۷، به جای انتخاب یک عدد تصادفی بین ۱ و i، اندازه فاصله، یک عدد تصادفی بین ۱ و N انتخاب می‌شود. وقتی در انتخاب نمونه به انتهای لیست می‌رسیم، باید از ابتدای لیست شروع کنیم. این روش باعث ایجاد یک الگوی دایره‌ای برای انتخاب نمونه می‌شود.

نمونه‌گیری سیستماتیک خطی و نمونه‌گیری سیستماتیک دایره‌ای یک نمونه واحد را

به دست می‌دهند. نمونه‌گیری سیستماتیک تکراری شامل انتخاب نمونه‌های متعدد از جمعیت هدف و سپس ترکیب این نمونه‌ها به شکل یک نمونه واحد است. به جای شروع از یک نقطه شروع تصادفی، با استفاده از شروع‌های تصادفی چند نمونه سیستماتیک کوچکتر انتخاب شده است. این کار باعث می‌شود که زمان انجام کار نسبت به نمونه‌گیری سیستماتیک خطی طولانی‌تر باشد. البته، نمونه‌گیری تکراری تاثیر یکسونگری ناشی از متناوب بودن و الگوی منظمی که در بازه نمونه‌گیری ایجاد می‌شود را کاهش می‌دهد. به علاوه، به دلیل این که نمونه‌گیری سیستماتیک خطی فقط یک خوشه از عناصر ایجاد می‌کند (البته ممکن است خوشه شامل چند عنصر باشد)، از نظر فنی، برآورد بدون انحراف خطای نمونه‌گیری را نمی‌توان بدون در نظر گرفتن برخی فرضیات به دست آورد. حداقل بایستی دو خوشه به صورت مستقل انتخاب شوند. نمونه‌گیری تکراری بیش از یک خوشه از عناصر را ایجاد کرده و محاسبه واریانس‌ها و خطای استاندارد برآوردها با استفاده از نمونه را تسهیل می‌کند.

نقاط قوت و ضعف نمونه‌گیری سیستماتیک چیست؟

نمونه‌گیری سیستماتیک در مقایسه با روش‌های نمونه‌گیری غیراحتمالی دارای نقاط قوت و ضعفی همانند بیشتر روش‌های نمونه‌گیری احتمالی است. برای مشخص کردن نقاط قوت و ضعف نمونه‌گیری سیستماتیک آن را با نمونه‌گیری تصادفی ساده مقایسه می‌کنیم. از نمونه‌گیری سیستماتیک زمانی استفاده می‌شود که استفاده از نمونه‌گیری تصادفی ساده دشوار یا غیرممکن باشد. این روش در مقایسه با نمونه‌گیری تصادفی ساده، در برخی از موارد قوی‌تر و در برخی از موار ضعیف‌تر عمل می‌کند. در مقایسه با نمونه‌گیری تصادفی ساده، نقاط قوت نمونه‌گیری سیستماتیک عبارتند از:

- اگر روند انتخاب به صورت دستی باشد، نمونه‌گیری سیستماتیک راحت‌تر، سریعتر، و اقتصادی‌تر از روش نمونه‌گیری تصادفی ساده است. فرد فقط برای انتخاب اولین عنصر باید از فرایند تصادفی استفاده کند. از طرف دیگر، اگر فرایند انتخاب به صورت رایانه‌ای باشد، سهولت در فرایند انتخاب نمونه‌گیری سیستماتیک و نمونه‌گیری تصادفی ساده می‌تواند

مشابه با هم باشد.

■ در صورتی که بازه نمونه‌گیری ترتیب یکنواختی از متغیر مورد مطالعه داشته باشد (مثلاً مرتب کردن فروشگاه‌ها بر حسب ارزش ریالی، لیست کردن کارمندان بر اساس سال‌های خدمت، و لیست کردن دانشگاه‌ها بر اساس نرخ فارغ‌التحصیلی)، طبقه بندی ضمنی می‌تواند منجر به بازدهی آماری برابر با نمونه‌گیری طبقه بندی تناسبی شده و بنابراین کارآمدتر از نمونه‌گیری تصادفی ساده است. در صورتی که ترتیب به صورت تصادفی باشد، نمونه‌گیری سیستماتیک ممکن است منجر به نتایجی مشابه با نمونه‌گیری تصادفی ساده شود.

■ نمونه‌گیری سیستماتیک تضمین می‌کند که نمونه بین جمعیت پراکندگی بیشتری دارد.

نقاط قوت	نقاط ضعف
برخلاف نمونه‌گیری تصادفی ساده	برخلاف نمونه‌گیری تصادفی ساده
اگر روند انتخاب به صورت دستی باشد، نمونه‌گیری سیستماتیک راحت‌تر، سریعتر، و اقتصادی‌تر است.	اگر فاصله نمونه‌گیری با ترتیب دوره‌ای عناصر در بازه نمونه‌گیری مرتبط باشد، ممکن است تغییرپذیری افزایش یابد.
در صورتی که نمایش فیزیکی وجود داشته باشد، نیازی نیست که جمعیت هدف شماره‌گذاری و بازه نمونه‌گیری تهیه شود.	ترکیبی از عناصر که احتمالات متفاوتی دارند باید انتخاب شود.
در صورتی که ترتیب عناصر در بازه نمونه‌گیری تصادفی باشد، نمونه‌گیری سیستماتیک ممکن است منجر به نتایجی مشابه با نمونه‌گیری تصادفی ساده شود.	از نظر فنی، فقط انتخاب اولین عنصر به صورت احتمالی است زیرا برای انتخاب‌های بعدی، عناصر درون جمعیت هدف شانس صفر برای انتخاب شدن دارند.

نقاط قوت	نقاط ضعف
در صورتی که بازه نمونه‌گیری ترتیب یکنواختی از متغیر مورد مطالعه داشته باشد، طبقه بندی ضمنی می‌تواند منجر به بازدهی آماری برابر با نمونه‌گیری طبقه بندی تناسبی شده و بنابراین کارآمدتر از نمونه‌گیری تصادفی ساده است.	برای انتخاب اولین عنصر که تعیین کننده انتخاب عناصر دیگر است، اصل استقلال نادیده گرفته شده است.
نمونه‌گیری سیستماتیک احتمال همبستگی خودکار را از بین می‌برد.	واریانس برآورد نسبت به نمونه‌گیری تصادفی ساده پیچیده‌تر است.
نمونه‌گیری سیستماتیک تضمین می‌کند که نمونه درون جمعیت گسترده شده است.	

شباهت عناصر مجاور درون یک لیست باعث همبستگی خودکار و همبستگی عناصر درون جمعیت می‌شود. ممکن است به ندرت در نمونه‌گیری تصادفی ساده نیز این وضعیت بروز کند. در لیست آدرس‌ها ممکن است همبستگی خودکار بخشی وجود داشته باشد. افرادی که در آدرس‌هایی که به یکدیگر هستند زندگی می‌کنند نسبت به افرادی که در آدرس‌های نزدیک به هم زندگی نمی‌کنند، احتمالاً مشابهت بیشتری با یکدیگر دارند، مثلاً بر اساس وضعیت اقتصادی اجتماعی. در مقایسه با نمونه‌گیری تصادفی ساده، همبستگی خودکار مثبت دقت را کاهش و همبستگی منفی دقت را افزایش می‌دهد. البته، نمونه‌گیری سیستماتیک باعث از بین رفتن احتمال همبستگی خودکار می‌شود. به عنوان مثال، در استفاده از لیست رای دهندگان برای انتخاب نمونه برای بررسی ترجیحات رای دهندگان، این امکان وجود دارد که با استفاده از نمونه تصادفی ساده اعضای یک خانواده انتخاب شوند ولی با استفاده از نمونه‌گیری سیستماتیک این امکان وجود ندارد.

در مقایسه با نمونه‌گیری تصادفی ساده، نمونه‌گیری سیستماتیک نقاط ضعف کمتری دارد. برخی از این نقاط ضعف عبارتند از:

- هر چند بروز این وضعیت نسبتاً نادر است ولی تناوب در بازه نمونه‌گیری یک نگرانی دایمی در نمونه‌گیری سیستماتیک است. در صورتی که الگوی دوره‌ای یا سیکلی متناسب

با نسبت نمونه‌گیری در بازه نمونه‌گیری وجود داشته باشد، ممکن است نمونه دچار انحراف شود. در صورتی که نسبت نمونه‌گیری برابر یا ضریبی از فاصله تناوب لیست باشد، این مشکل وجود خواهد داشت. به عنوان مثال، در صورتی که دانشجویان بر اساس کلاس لیست شده باشند و درون هر کلاس بر اساس عملکرد در آزمون رتبه‌بندی شده باشند، ممکن است نمونه سیستماتیک دانشجویان دچار انحراف شود. در صورتی که کلاس‌ها تعداد تقریباً برابری از دانشجویان داشته باشد، انحراف دوره‌ای رخ خواهد داد.

- به علاوه، بر خلاف نمونه‌گیری تصادفی ساده که هر ترکیب از n عنصر شانس برابری برای انتخاب شدن دارد، این وضعیت برای نمونه‌گیری سیستماتیک وجود ندارد.
- از نظر فنی، فقط انتخاب اولین عنصر به صورت احتمالی است زیرا برای انتخاب‌های بعدی، عناصر درون جمعیت هدف شانس صفر برای انتخاب شدن دارند.
- برای انتخاب اولین عنصر که تعیین کننده انتخاب عناصر دیگر است، اصل استقلال نادیده گرفته شده است.
- واریانس برآورد نسبت به نمونه‌گیری تصادفی ساده پیچیده‌تر است.

نمونه‌گیری خوشه‌ای

نمونه‌گیری خوشه‌ای چیست؟

در اغلب موارد ایجاد یک بازه نمونه‌گیری از جمعیت هدف غیرممکن یا غیرعملی است، و یا جمعیت هدف از نظر جغرافیایی آنقدر گسترده است که جمع‌آوری داده هزینه نسبتاً بالایی دارد. در چنین وضعیتی بهترین کار استفاده از نمونه‌گیری خوشه‌ای است. نمونه‌گیری خوشه‌ای یک روش نمونه‌گیری احتمالی است که در آن عناصر جمعیت به صورت تصادفی از بین گروه‌ها (خوشه‌ها) انتخاب می‌شوند. در رابطه با نمونه‌گیری خوشه‌ای، «خوشه» مجموعه یا گروهی از عناصر جمعیت است. نمونه‌گیری عناصر به صورت انتخاب مجزای عناصر جمعیت و به یکباره است. از طرف دیگر، نمونه‌گیری خوشه‌ای شامل انتخاب عناصر جمعیت نه به صورت فردی که به صورت گروهی است. واحدهای نمونه‌گیری یا خوشه‌ها ممکن است مبتنی بر فضا، مثلاً واحدهایی که به طور طبیعی به صورت جغرافیایی یا فیزیکی ایجاد می‌شوند (مثلاً ایالت‌ها، شهرستان‌ها، نشریات سرشماری، یا ساختمان‌ها)؛ مبتنی بر سازمان، مثلاً واحدهایی به صورت مناطق دانشگاهی، دانشگاه‌ها، سطوح نمره، یا کلاس‌ها؛ یا مبتنی بر تلفن، همانند کد مناطق یا تغییر شماره تلفن‌ها باشند.

برای داشتن یک طرح نمونه خوشه‌ای خوب، ناهمگنی خوشه‌ها بسیار مهم است. در حالت ایده‌ال، اختلاف درون خوشه‌ها باید زیاد و اختلاف بین آنها کم باشد. خوشه‌ها باید شبیه به هم باشند. از طرف دیگر، عناصر درون هر خوشه باید به اندازه جمعیت هدف ناهمگن باشد. در حالت ایده‌ال، خوشه‌ها کوچک هستند ولی نه آنقدر که همگن باشند.

مراحل انتخاب در نمونه خوشه‌ای چیست؟

برای انتخاب نمونه خوشه‌ای، شش مرحله اصلی وجود دارد:

۱. تعریف جمعیت هدف.
۲. تعیین اندازه نمونه مورد نظر.
۳. شناسایی بازه نمونه‌گیری موجود در جمعیت هدف یا ایجاد یک بازه جدید.
۴. بررسی بازه نمونه‌گیری از نظر پوشش ناقص، پوشش بیش از حد، پوشش چندگانه، و خوشه‌بندی، و در صورت لزوم انجام اصلاحات. در حالت ایده‌ال، خوشه‌ها باید

همانند جمعیت ناهمگن، دو به دو مانع‌الجمع، و در کل جامع باشند. در صورتی که عناصر جمعیت به بیش از یک خوشه تعلق داشته باشند، ممکن است تکرار عناصر در یک نمونه بروز کند. از قلم افتادن‌ها منجر به یکسونگری در پوشش خواهد شد.

۵. تعیین تعداد خوشه‌هایی که باید انتخاب شوند. این کار می‌تواند با تقسیم کردن اندازه نمونه بر میانگین برآورد شده تعداد عناصر جمعیت در هر خوشه انجام شود. میزان همگنی و ناهمگنی خوشه‌ها با جمعیت تفاوت دارد، زیرا هر چه تعداد خوشه‌ها افزایش یابد، دقت نیز افزایش می‌یابد. از طرف دیگر، هر چه اختلاف بین خوشه‌ها افزایش یابد، دقت کاهش می‌یابد.

۶. انتخاب تصادفی تعداد هدفگذاری شده خوشه‌ها.

انواع نمونه‌گیری خوشه‌ای چیست؟

برای طبقه‌بندی انواع مختلف نمونه‌گیری خوشه‌ای از دو جنبه مهم استفاده شده است. یک جنبه مبتنی بر تعداد مراحل در طرح نمونه، و دیگری مبتنی بر نمایش تناسبی خوشه‌ها به کل نمونه است.

انواع مبتنی بر تعداد مراحل

در اغلب موارد نمونه‌گیری خوشه‌ای در بیش از یک مرحله انجام شده است. مرحله گامی در روند نمونه‌گیری است که در آن نمونه انتخاب شده است. با در نظر گرفتن تعداد مراحل طرح، سه نمونه اصلی برای نمونه‌گیری خوشه‌ای وجود دارد: نمونه‌گیری خوشه‌ای تک مرحله‌ای، نمونه‌گیری خوشه‌ای دو مرحله‌ای، و نمونه‌گیری خوشه‌ای چند مرحله‌ای.

نمونه‌گیری خوشه‌ای تک مرحله‌ای. در نمونه خوشه‌ای تک مرحله‌ای، نمونه‌گیری در یک مرحله انجام شده است. به عنوان یک مثال از نمونه‌گیری خوشه‌ای تک مرحله‌ای، حالتی را در نظر بگیرید که در آن فردی علاقه به بررسی افراد بی‌خانمانی که در پناهگاه‌ها زندگی می‌کنند دارد. در صورتی که در یک شهر پنج پناهگاه وجود داشته باشد، محقق به صورت تصادفی یکی از آنها را انتخاب و سپس تمامی افراد بی‌خانمانی که در این پناهگاه ساکن هستند را مورد بررسی قرار می‌دهد. یک محقق بازار نیز می‌تواند از روش نمونه خوشه‌ای تک مرحله‌ای استفاده کند. مثلاً محققی که به تست بازاریابی یک محصول

علاقمند است. محقق می‌تواند به صورت تصادفی کدهای پستی را انتخاب کند؛ نمونه‌های محصول را همراه با یک پرسشنامه که با پست برگردانده می‌شود را به هر آدرسی که درون خوشه‌های انتخاب شده قرار دارند ارسال کند.

نمونه‌گیری خوشه‌ای دو مرحله‌ای. روش نمونه خوشه‌ای دو مرحله‌ای شامل تمامی مراحل روش نمونه خوشه‌ای یک مرحله‌ای است، ولی در آخرین مرحله یک تفاوت دارد. به جای در نظر گرفتن تمامی عناصر در خوشه‌های انتخابی در نمونه، از عناصر درون هر کدام از خوشه‌های انتخابی یک نمونه تصادفی (به صورت نمونه تصادفی ساده، نمونه طبقه بندی، یا نمونه سیستماتیک) انتخاب می‌شود. نمونه‌گیری که بعد از اولین مرحله گرفته می‌شود را برخی اوقات زیر نمونه‌گیری می‌نامند. به طور کلی، در صورتی که خوشه‌ها همگن باشند، روش نمونه خوشه‌ای دو مرحله‌ای بهتر از روش نمونه خوشه‌ای یک مرحله‌ای است. در صورتی که نمونه‌گیری تک مرحله‌ای همراه با احتمالی متناسب با اندازه باشد، یک نمونه خود وزن‌گذار ایجاد خواهد شد. در مثال ذکر شده در مورد بررسی روی افراد بی‌خانمان، به جای انتخاب تمامی افرادی که در پناهگاه انتخابی برای قرار گرفتن در تحقیق، محقق به صورت تصادفی یک زیر مجموعه از ساکنین پناهگاه انتخاب خواهد کرد.

نمونه‌گیری خوشه‌ای چند مرحله‌ای. انجام ارزیابی در مناطق وسیع جغرافیایی تا حدودی نیاز به روش نمونه‌گیری پیچیده‌تری نسبت به آن چه که تاکنون گفته شد دارد. معمولاً، باید از روش نمونه‌گیری خوشه‌ای چند مرحله‌ای استفاده کرد. نمونه‌گیری خوشه‌ای چند مرحله‌ای شامل تکرار دو مرحله اصلی است: لیست برداری و نمونه‌گیری. معمولاً، در هر مرحله، خوشه‌ها از نظر اندازه کوچکتر می‌شوند و در آخرین مرحله از نمونه‌گیری عنصر استفاده شده است. روش‌های نمونه‌گیری (نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه بندی، یا نمونه‌گیری سیستماتیک) در هر مرحله متفاوت هستند. لزومی ندارد که روش‌های نمونه‌گیری در هر مرحله یکسان باشد. تعداد مراحلی که استفاده می‌شوند اغلب توسط در دسترس بودن بازه‌های نمونه‌گیری در مراحل مختلف تعیین شده است.

برای اشاره به واحدهای مختلف نمونه‌گیری از اصطلاحات خاصی استفاده شده است. واحد نمونه‌گیری که در اولین مرحله استفاده شده را واحد نمونه‌گیری اولیه (PSU) می‌نامند. واحدهای نمونه‌گیری بعدی را واحد نمونه‌گیری ثانویه (SSU)، واحدهای نمونه‌گیری سومی

(TSU)، والی آخر می‌نامند تا این که در نهایت به واحد نمونه‌گیری «نهایی» یا «آخر» می‌رسیم.

معمولاً، هر چه روند نمونه‌گیری از سمت انتخاب واحدهای نمونه‌گیری اولیه به سمت مراحل بعدی نمونه‌گیری پیش میرود، واحدهای نمونه‌گیری نیز همگن‌تر می‌شوند. خوشه‌های بزرگ نسبت به خوشه‌های کوچک تمایل بیشتری به ناهمگنی دارند. به دلیل ناهمگنی بیشتر واحدهای نمونه‌گیری اولیه، در صورتی که فرد نمونه‌ای داشته باشد که بیشتر عناصر آن PSU باشد تا SSU، بیشتر SSU باشد تا TSU و همین طور الی آخر، خطای نمونه‌گیری به حداقل خواهد رسید.

زیرنمونه‌های مبتنی بر نمایش تناسبی خوشه‌ها در نمونه

خوشه‌ها ممکن است به روش نمونه‌گیری EPSEM انتخاب شده باشند؛ یعنی این که هر عنصر درون جمعیت شانس برابری برای قرار گرفتن در نمونه داشته باشد. در صورتی که خوشه‌های نمونه‌گیری شده اندازه تقریباً برابری داشته باشند، روش نمونه را می‌توان به صورت یک روش نمونه EPSEM در نظر گرفت. در صورتی که خوشه‌ها اندازه برابر نداشته باشند، با استفاده از روش انتخاب با احتمال متناسب با اندازه (PPS)، می‌توان به روش نمونه EPSEM دست پیدا کرد. احتمال انتخاب یک خوشه به توزیع تناسبی عناصر آن در جمعیت هدف بستگی دارد. با استفاده از PPS، به نمونه خود وزن‌گذار می‌توان دست یافت. روش نمونه‌گیری با احتمال نامتناسب با اندازه (PDS) شامل انتخاب خوشه‌ها بدون در نظر گرفتن توزیع تناسبی عناصر در جمعیت هدف است.

روش انتخاب پاسخ‌دهندگان

معمولاً، در ارزیابی خانوار که از روش نمونه خوشه‌ای دو مرحله‌ای یا چند مرحله‌ای استفاده می‌شود، هر عنصر در آخرین مرحله روش نمونه انتخاب شده است. در صورتی که خانوار شامل بیش از یک عضو در جمعیت هدف باشد، در این صورت باید فقط یک عنصر انتخاب شود. برای انتخاب عنصر به عنوان جمع‌آوری داده از هر دو روش تناسبی و غیرتناسبی می توان استفاده نمود.

از دو روش کلی برای انتخاب غیراحتمالی پاسخ دهنده در خانوار استفاده شده است:

انتخاب سرپرست خانوار و انتخاب اولین فرد بزرگسال. در استفاده از انتخاب سرپرست خانوار، محقق به راحتی می‌خواهد که با سرپرست خانوار صحبت کند. همچنین ممکن است فرد بخواهد با سرپرستان مرد و زن خانوار صحبت کند. روش اولین فرد بزرگسال شامل انتخاب اولین بزرگسالی است که با وی تماس برقرار شده و در این صورت او عضوی از جمعیت هدف خواهد بود. کنترل این روش‌ها ساده است، نیاز به وقت زیادی ندارد و سرزده نیستند. البته، آنها منجر به انحراف در انتخاب می‌شوند و احتمالاً از زنان نمونه بیش از اندازه گرفته می‌شود زیرا زنان در زمان مصاحبه بیشتر از مردان در دسترس هستند. در روش سرپرست خانوار گرایش به این سمت است که از زنان بیش از اندازه نمونه‌گیری شود، به خصوص در مناطق شهری، زیرا تعداد سرپرست خانوارهای تک سرپرست اکثراً زن هستند. روش انتخاب اولین فرد بزرگسال تمایل به نمونه‌گیری بیش از حد از زنان دارد زیرا احتمال این که زنان در خانه حضور داشته باشند بیشتر است.

این روش‌های انتخاب پاسخ‌دهنده به هر عضو درون جمعیت هدف شانسی برای قرار گرفتن در نمونه نمی‌دهد. ترکیب انتخاب احتمالی خوشه‌ها با انتخاب غیراحتمالی اعضای خانوار باعث می‌شود که روش نمونه‌گیری تبدیل به یک روش ترکیبی شود.

چندین روش برای انتخاب احتمالی پاسخ دهنده خانوار وجود دارد. متداولترین روش‌های احتمالی استفاده شده عبارتند از جداول کیش، جداول تروداهال-کارتر-بریانت، روش انتخاب هاگان و کارتر، و روش روز تولد آخری/بعدی. این روش‌ها نشاندهنده تلاش بین محققان برای به حداقل رساندن خطای سیستماتیک است. معمولاً، انجام مصاحبه طولانی است زیرا نیاز به دو مجوز دارد: مجوز اولیه برای تماس اول با خانوار و مجوز دوم برای مصاحبه با فرد انتخاب شده. این وضعیت باعث شده که انحراف ناشی از پوشش ناقص کاهش ولی نرخ انصراف افزایش یابد. به علاوه، اگر فرد انتخاب شده در خانه نباشد، مصاحبه کننده برای انتخاب فرد دیگری از خانوار دارای محدودیت است. در این صورت باید دوباره تماس گرفته شود. موفقیت تماس‌های مجدد بر انحراف عدم پاسخ واحد مورد مطالعه تاثیرگذار است.

نقاط قوت و ضعف روش نمونه‌گیری خوشه‌ای چیست؟

نمونه‌گیری خوشه‌ای همانند بیشتر روش‌های نمونه‌گیری احتمالی و در مقایسه با روش‌های نمونه‌گیری غیر احتمالی دارای نقاط قوت و ضعف است. البته، این روش در مقایسه با سایر روش‌های نمونه‌گیری احتمالی، همانند نمونه‌گیری تصادفی ساده، دارای نقاط قوت و ضعف خاصی است (جدول ۸ را ببینید). برخی از نقاط قوت نمونه‌گیری خوشه‌ای در مقایسه با نمونه‌گیری تصادفی ساده عبارتند از:

- در صورتی که خوشه‌ها از نظر جغرافیایی مشخص شده باشند، نمونه‌گیری خوشه‌ای نیاز به زمان، پول و نیروی کار کمتری در مقایسه با نمونه‌گیری تصادفی ساده دارد. این روش به صرفه‌ترین روش نمونه‌گیری احتمالی است.
- در سطح ثابت هزینه، نمونه‌گیری خوشه‌ای با اندازه نمونه بزرگتر خطای نمونه‌گیری کمتری در مقایسه با نمونه‌گیری تصادفی ساده با اندازه نمونه کوچکتر دارد.

جدول ۸) نقاط قوت و ضعف نمونه‌گیری خوشه‌ای در مقایسه با نمونه‌گیری تصادفی ساده

نقاط قوت	نقاط ضعف
در مقایسه با نمونه‌گیری تصادفی ساده:	در مقایسه با نمونه‌گیری تصادفی ساده:
در صورتی که خوشه‌ها از نظر جغرافیایی مشخص شده باشند، نمونه‌گیری خوشه‌ای نیاز به زمان، پول و نیروی کار کمتری در مقایسه دارد.	نمونه خوشه‌ای، همانند نمونه تصادفی ساده با اندازه برابر، ممکن است نماینده جمعیت نباشد.
نمونه‌گیری خوشه‌ای امکان نمونه‌گیری بعدی را فراهم می‌کند زیرا خوشه‌های نمونه‌گیری شده مجموعی از عناصر هستند.	واریانس‌های نمونه‌های خوشه‌ای بالاتر از واریانس‌های نمونه‌های تصادفی ساده هستند.

نقاط قوت	نقاط ضعف
فرد می‌تواند مشخصات خوشه‌ها و هم چنین جمعیت را تعیین کند.	نمونه‌گیری خوشه‌ای باعث افزایش پیچیدگی تحلیل داده‌ها و تفسیر نتایج حاصل از تحلیل‌ها می‌شود.
نمونه‌گیری خوشه‌ای نیازی به بازه نمونه‌گیری شامل تمامی عناصر جمعیت هدف ندارد.	در نمونه‌هایی با اندازه برابر، نمونه‌گیری خوشه‌ای منجر به خطاهای نمونه‌گیری بیشتری در مقایسه با نمونه‌گیری‌های احتمالی می‌شود.

- نمونه‌گیری خوشه‌ای امکان نمونه‌گیری بعدی را فراهم می‌کند زیرا خوشه‌های نمونه‌گیری شده مجموعی از عناصر هستند.
- بر خلاف نمونه‌گیری تصادفی ساده، نمونه‌گیری خوشه‌ای امکان تعیین خصوصیات زیرمجموعه‌ها (خوشه‌ها) و همچنین جمعیت هدف را فراهم می‌کند.
- نمونه‌گیری خوشه‌ای یک مرحله‌ای نیاز به فقط بازه نمونه‌گیری خوشه‌ها دارد، و نمونه‌گیری خوشه‌ای دو مرحله‌ای و نمونه‌گیری خوشه‌ای چند مرحله‌ای نیاز به بازه نمونه‌گیری عناصر جمعیت برای فقط خوشه‌های نمونه‌گیری شده در آخرین مرحله فرایند دارد.
- اجرای نمونه‌گیری خوشه‌ای بسیار ساده‌تر از نمونه‌گیری تصادفی ساده است.

برخی از نقاط ضعف نمونه‌گیری خوشه‌ای در مقایسه با نمونه‌گیری تصادفی ساده عبارتند از:

- در اندازه نمونه برابر، خوشه‌های نمونه‌گیری شده ممکن است همانند نمونه تصادفی ساده نماینده جمعیت نباشد.
- ترکیب واریانس دو خوشه همگن مجزا ممکن است واریانس کل نمونه بالاتر از نمونه‌گیری

تصادفی ساده باشد.

■ نمونه‌گیری خوشه‌ای باعث پیچیدگی بیشتر تحلیل داده‌ها می‌شود. تعیین و تفسیر تحلیل استنتاجی آماری داده‌های جمع‌آوری شده با استفاده از نمونه‌گیری خوشه‌ای بسیار دشوارتر از تحلیل آمار استنتاجی داده‌های جمع‌آوری شده با استفاده از نمونه‌گیری تصادفی ساده است. نرم‌افزار آماری استفاده شده برای تحلیل داده‌های جمع‌آوری شده بایستی از فرمول‌هایی استفاده کند که استفاده از روش نمونه خوشه‌ای را در نظر داشته باشد. بسیاری از برنامه‌های نرم‌افزاری آماری از فرمول مربوط به نمونه‌گیری تصادفی ساده استفاده می‌کنند و در نتیجه سطوح معناداری را بیش از اندازه برآورد می‌کنند.

■ هر چه تعداد مراحل در روش نمونه خوشه‌ای بیشتر باشد، خطای کلی نمونه‌گیری بیشتر خواهد بود.

■ در صورتی که خوشه‌ها با یکدیگر مشابه نباشند، هر چه تعداد خوشه‌ها کمتر باشند، خطای نمونه‌گیری بیشتر خواهد بود.

■ در نمونه‌هایی با اندازه برابر، نمونه‌گیری خوشه‌ای منجر به خطاهای نمونه‌گیری بیشتری در مقایسه با سایر نمونه‌گیری‌های احتمالی می‌شود. در صورتی که خوشه‌ها با یکدیگر مشابه باشند، این خطا به حداقل رسیده می‌شود. به علاوه با افزایش تعداد خوشه‌ها، می‌توان این خطاها را کاهش داد. توجه داشته باشید که این کار روی افزایش هزینه جمع‌آوری داده‌ها نیز تاثیرگذار است.

■ هر چه تعداد خوشه‌هایی که یک فرد انتخاب می‌کند بیشتر باشد، اختلاف هزینه‌های جمع‌آوری داده بین نمونه‌گیری خوشه‌ای و نمونه‌گیری تصادفی ساده کمتر می‌شود.

■ چون عناصر درون یک خوشه تمایل به مشابهت دارند، وقتی عنصر دیگری از آن خوشه به جای خوشه دیگر انتخاب می‌کنیم، اطلاعات کمتری در مورد جمعیت دریافت می‌کنیم. این عدم وجود اطلاعات جدید باعث می‌شود که نمونه خوشه‌ای دقت کمتری نسبت به نمونه تصادفی ساده داشته باشد.

اختلاف بین نمونه‌گیری خوشه‌ای و نمونه‌گیری طبقه بندی چیست؟

نمونه‌گیری خوشه‌ای مشابه با نمونه‌گیری طبقه بندی است زیرا هردو روش نیاز به تقسیم جمعیت به گروه‌ها و سپس نمونه‌گیری از درون گروه‌ها دارند (جدول ۹ را ببینید). هردو روش امکان تحلیل مجزای گروه‌ها (طبقات یا خوشه‌ها) و همچنین تحلیل کل نمونه را دارند. البته تفاوت‌های مهمی نیز وجود دارد. برخی از این تفاوت‌ها عبارتند از:

- در نمونه‌گیری طبقه بندی پس از این که گروه‌ها (طبقات) ایجاد شدند، یک نمونه تصادفی از هر گروه (طبقه) استخراج می‌شود. از طرف دیگر، در نمونه‌گیری خوشه‌ای، عناصر از تمامی خوشه‌ها انتخاب نمی‌شوند. در نمونه‌گیری خوشه‌ای یک مرحله‌ای، پس از این که گروه‌ها (خوشه‌ها) ایجاد شدند، یک نمونه تصادفی از خوشه استخراج می‌شود. تمامی عناصر در خوشه انتخاب شده در نمونه وارد شده‌اند. در نمونه‌گیری خوشه‌ای دو مرحله‌ای و چند مرحله‌ای، یک نمونه تصادفی از خوشه استخراج و سپس عناصر به صورت تصادفی از خوشه‌های انتخاب شده انتخاب می‌شوند.

جدول ۹) مقایسه نمونه‌گیری طبقه بندی و نمونه‌گیری خوشه‌ای

نمونه‌گیری طبقه بندی	نمونه‌گیری خوشه‌ای
جمعیت به گروه‌هایی تفکیک و سپس نمونه‌گیری از درون هر طبقه انجام شده است.	جمعیت به خوشه‌هایی تفکیک و سپس نمونه‌گیری از درون هر خوشه انجام شده است.
علاوه بر تحلیل نمونه کلی، تحلیل هر طبقه نیز امکان‌پذیر است.	علاوه بر تحلیل نمونه کلی، تحلیل هر گروه (خوشه) نیز امکان‌پذیر است.
برای به حداقل رساندن خطای نمونه‌گیری، اختلاف بین طبقات نیز باید به حداقل رسانده شود و اختلاف بین گروه‌ها و بین طبقات نیز باید حداکثر شوند.	برای به حداقل رساندن خطای نمونه‌گیری، اختلاف درون گروه باید مطابق با اختلاف آنها در جمعیت باشد و اختلاف بین گروه بین خوشه‌ها نیز باید حداقل شوند.

نمونه‌گیری طبقه‌بندی	نمونه‌گیری خوشه‌ای
برای کل جمعیت هدف به یک بازه نمونه‌گیری نیاز است.	در نمونه‌گیری خوشه‌ای تک مرحله‌ای، فقط برای خوشه‌ها به یک بازه نمونه‌گیری نیاز است. در نمونه‌گیری خوشه‌ای دو و چند مرحله‌ای، فقط برای عناصری که در خوشه‌های انتخاب شده در مرحله نهایی انتخاب شده‌اند به بازه نمونه‌گیری برای هر عنصر نیاز است.
هدف اصلی: افزایش دقت و نمایندگی.	هدف اصلی: کاهش هزینه‌ها و افزایش بازده عملیاتی.
طبقه‌بندی‌ها توسط محقق تعیین شده‌اند.	طبقه‌بندی‌ها به طور طبیعی بین گروه‌هایی که از قبل وجود دارند ایجاد شده است.
دقت بیشتر در مقایسه با نمونه‌گیری تصادفی ساده.	دقت کمتر در مقایسه با نمونه‌گیری تصادفی ساده.
متغیرهای استفاده شده برای طبقه بندی باید با مساله مورد تحقیق ارتباط داشته باشد.	متغیرهای استفاده شده برای خوشه‌بندی نباید با مساله مورد تحقیق ارتباط داشته باشد.
متغیرهای متداول برای طبقه بندی: سن، جنسیت، درآمد، نژاد.	متغیرهای متداول برای طبقه‌بندی: ناحیه جغرافیایی، دانشگاه، سطح نمره.
نیاز به اطلاعات اولیه بیشتر نسبت به نمونه‌گیری خوشه‌ای.	نیاز به اطلاعات اولیه کمتر در مقایسه با نمونه‌گیری طبقه بندی.

■ در نمونه‌گیری طبقه بندی برای به حداقل رساندن خطای نمونه‌گیری، اختلاف‌های بین گروه و بین طبقات باید حداقل و طبقات تا حد ممکن باید همگن شوند. در نمونه‌گیری خوشه‌ای، برای به حداقل رساندن خطای نمونه‌گیری، اختلاف‌های بین گروه باید مطابق با اختلاف آنها در جمعیت باشد و خوشه‌ها باید تا حد امکان همانند جمعیت ناهمگن باشند. وضعیت ایده‌ال برای نمونه‌گیری طبقه بندی این است که درون هر طبقه همگنی داشته و طبقات با هم متفاوت باشند. وضعیت ایده‌ال برای نمونه‌گیری خوشه‌ای این است که درون

خوشه‌ها ناهمگنی داشته و خوشه‌ها با هم تفاوتی نداشته باشند.

■ در نمونه‌گیری طبقه بندی برای به حداقل رساندن خطای نمونه‌گیری، اختلاف بین طبقات باید به حداکثر رسانده شود. در نمونه‌گیری خوشه‌ای، برای به حداقل رساندن خطای نمونه‌گیری، اختلافات بین گروه و بین خوشه‌ها باید به حداقل رسانده شود.

■ در نمونه‌گیری طبقه بندی، گروه‌ها توسط محقق ایجاد شده است. در نمونه‌گیری خوشه‌ای، طبقه‌بندی‌ها به طور طبیعی گروه‌ها را شکل می‌دهند.

■ در نمونه‌گیری طبقه بندی، برای کل جمعیت هدف نیاز به بازه نمونه‌گیری است. در نمونه‌گیری خوشه‌ای یک مرحله‌ای، فقط برای خوشه‌ها به بازه نمونه‌گیری نیاز است. در نمونه‌گیری خوشه‌ای دو و چند مرحله‌ای، علاوه بر بازه نمونه‌گیری خوشه‌ها در اولین مرحله فرایند، فقط برای عناصر هر کدام از خوشه‌های انتخاب شده به بازه نمونه‌گیری نیاز است.

■ هدف اصلی نمونه‌گیری طبقه بندی این است که دقت و نمایندگی را افزایش دهیم. هدف اصلی نمونه‌گیری خوشه‌ای این است که هزینه‌ها کاهش و بازده عملیاتی افزایش دهیم.

■ در مقایسه با نمونه‌گیری تصادفی ساده، نمونه‌گیری طبقه بندی دقت بالاتر و نمونه‌گیری خوشه‌ای دقت کمتری دارد. افزایش دقت با استفاده از طبقه بندی خیلی زیاد نیست. اما، خوشه‌بندی باعث کاهش قابل توجه دقت می‌شود.

■ متغیرهای مورد استفاده برای طبقه بندی باید با متغیرهای مورد مطالعه ارتباط داشته باشد. متغیر مورد استفاده برای خوشه‌بندی نباید ارتباطی با متغیرهای مورد استفاده داشته باشد.

■ متغیرهای متداول استفاده شده برای طبقه بندی عبارتند از سن، جنس، و درآمد. متغیرهای متداول استفاده شده برای طبقه‌بندی در نمونه‌گیری خوشه‌ای عبارتند از منطقه جغرافیایی، دانشگاه و سطح نمره.

■ نمونه‌گیری طبقه بندی نیاز به اطلاعات اولیه بیشتری نسبت به نمونه‌گیری خوشه‌ای دارد؛ نمونه‌گیری خوشه‌ای نیاز به اطلاعات اولیه کمتری در مقایسه با نمونه‌گیری طبقه بندی دارد.

■ در نمونه‌گیری طبقه بندی، محقق جمعیت هدف را به چند زیر گروه، هر کدام با تعداد زیادی عنصر، تقسیم کرده است. در نمونه‌گیری خوشه‌ای، محقق جمعیت هدف را به زیر گروه‌های متعدد، هر کدام با چند عنصر، تقسیم کرده است.

اختلاف بین نمونه‌گیری چند مرحله‌ای و نمونه‌گیری چند فازی چیست؟

نمونه‌گیری چند مرحله‌ای (نمونه‌گیری دو و چند مرحله‌ای) اغلب با نمونه‌گیری چند فازی (که به آن نمونه‌گیری دو فازی و نمونه‌گیری بعد از طبقه بندی نیز اطلاق می‌شود) اشتباه گرفته می‌شود. هر دو روش نمونه‌گیری شامل نمونه‌گیری چندگانه در مراحل یا فازهای مختلف بوده و برخی از مواقع می‌توان آنها را به عنوان نمونه‌گیری با استفاده از روش‌های ترکیبی در نظر بگیریم. در نمونه‌گیری چند مرحله‌ای، واحدهای نمونه‌گیری برای مراحل مختلف متفاوت هستند. از طرف دیگر، در نمونه‌گیری چند فازی از یک واحد نمونه‌گیری چند بار نمونه‌گیری شده است.

معمولاً از نمونه‌گیری چند فازی زمانی استفاده می‌شود که یک بازه نمونه‌گیری با اطلاعات کمکی مناسب که امکان طبقه بندی را فراهم می‌کنند، نداشته باشیم. از اولین فاز برای غربالگری استفاده شده است. با استفاده از بازه نمونه‌گیری در دسترس فرد باید به روش زیر عمل کند:

۱. انتخاب یک نمونه اولیه از عناصر با استفاده از بازه نمونه در دسترس.
۲. انجام مصاحبه کوتاه غربالگری برای جمع‌آوری اطلاعات کمکی لازم برای نمونه‌گیری بیشتر و طبقه بندی.
۳. طبقه بندی بعدی نمونه اولیه به طبقات با استفاده از اطلاعات کمکی جمع‌آوری شده.
۴. با استفاده از طبقه برای جمع‌آوری اطلاعات اضافی برای آن، یا باید تمامی عناصر

درون طبقه انتخاب شوند یا این که برای جمع‌آوری اطلاعات بیشتر نمونه احتمالی از عناصر درون طبقه تهیه شود.

نمونه‌گیری چند فازی معمولاً برای افزایش دقت، کاهش هزینه، و کاهش عدم پاسخ انجام می شود. همان طور که قبلاً گفته شد، نمونه‌های طبقه بندی دقت بالاتری نسبت به نمونه‌های تصادفی ساده با اندازه نمونه برابر دارند. البته، بازه نمونه‌گیری باید شامل اطلاعاتی راجع به متغیر(های) طبقه بندی برای تمامی عناصر جمعیت باشد تا طبقه بندی انجام شود. از نمونه‌گیری چند فازی زمانی استفاده می‌شود که بازه نمونه‌گیری شامل این اطلاعات نباشد.

از نمونه‌گیری چند فازی می‌توان برای کاهش هزینه‌های جمع‌آوری داده، در صورتی که برای جمع‌آوری داده برای برخی متغیرها نسبت به متغیرهای دیگر به زمان و کار زیادی نیاز باشد، استفاده کرد. در فاز۱، داده‌هایی که به راحتی در دسترس هستند را می‌توان از نمونه کلی جمع‌آوری کرد. در فاز ۲ و فازهای بعدی، در صورتی که مد نظر باشد یا ضرورت داشته باشد، داده‌هایی که نیاز به کار یا هزینه بیشتر برای جمع‌آوری شدن داشته باشند از یک زیر نمونه کوچکتر جمع‌آوری شده‌اند. هزینه‌های جمع‌آوری داده به حداقل رسیده است.

از نمونه‌گیری چند فازی می‌توان برای به دست آوردن اطلاعات مربوط به نمونه های عدم پاسخ استفاده کرد.

معمولاً، جمع‌آوری داده در مورد افرادی که در ابتدا از مشارکت در تحقیق سر باز زده‌اند و افراد دیگری که به جمع‌آوری داده پاسخ نداده‌اند، هزینه بیشتری دارند. چنین هزینه‌هایی با انجام نمونه‌گیری چند فازی در بین نمونه های عدم پاسخ به حداقل رسیده است.

رهنمودهایی برای انتخاب نوع طرح نمونه احتمالی

با در نظر گرفتن ویژگی‌هایی ذکر شده در بالا برای طرح‌های نمونه، و نقاط قوت و ضعف آنها، رهنمودهای زیر را می‌توان برای انواع طرح‌های نمونه زیر ارائه کرد:

- نمونه‌گیری تصادفی ساده
- نمونه‌گیری طبقه بندی
- نمونه‌گیری سیستماتیک
- نمونه‌گیری خوشه‌ای
- طرح‌های مختلط نمونه

رهنمود۱) نمونه‌گیری تصادفی ساده. در صورتی که فرد به بازه نمونه‌گیری کامل و دقیقی از جمعیت هدف کامل و دقیق دسترسی باشد ولی اطلاعات کمکی که می‌توانند برای اهداف طبقه بندی استفاده شوند را نداشته باشد در این صورت انتخاب نمونه‌گیری تصادفی ساده را مد نظر قرار می‌دهد.

رهنمود ۲) نمونه‌گیری طبقه بندی. در حالت زیر نمونه‌گیری طبقه بندی را مد نظر قرار دهید:

- این امکان وجود داشته باشد که جمعیت را به دو یا چند طبقه همگن تقسیم و برای هر طبقه یک بازه نمونه‌گیری ایجاد شود.
- فرد به بازه نمونه‌گیری کامل و دقیقی از جمعیت هدف که حاوی اطلاعات کمکی که می‌توانند برای اهداف طبقه بندی استفاده شوند، دسترسی داشته باشد.
- برخی از زیر گروه‌های جمعیت تفاوت زیادی با زیر گروه‌های دیگر داشته باشد.
- به حداقل رساندن خطای نمونه‌گیری بسیار مهم باشد.
- در رابطه با نمایندگی کمتر زیرگروه‌های کوچکتر نگرانی وجود داشته باشد.
- جمعیت ناهمگن است.
- قصد این باشد که از روش‌های انتخاب مختلف برای طبقات مختلف استفاده شود.
- این احتمال وجود دارد که پاسخ‌ها به پرسش‌های تحقیقی یک تحقیق برای زیرگروه‌های مختلف متفاوت باشد.

- گزارش کردن مجزای نیازهای مختلف هر طبقه مفید باشد.
- تحلیل جامع طبقات مد نظر باشد.

رهنمود ۳) نمونه‌گیری طبقه بندی تناسبی. در صورتی که قرار است زیرگروه‌هایی که تقریباً اندازه برابر دارند بررسی یا مقایسه شوند، انتخاب نمونه‌گیری طبقه بندی تناسبی را مد نظر قرار دهید.

رهنمود ۴) نمونه‌گیری طبقه بندی غیر تناسبی. در موارد زیر انتخاب نمونه‌گیری طبقه بندی غیر تناسبی را مد نظر قرار دهید:

- قرار باشد زیرگروه‌هایی با اندازه‌های بسیار متفاوت بررسی یا مقایسه شوند.
- مهم باشد که تعداد زیادی از عناصر یک بخش کوچک جمعیت در نظر گرفته شوند.
- فرد عمدتاً علاقمند به تشابهات و اختلافات کلیدی بین طبقات باشد.
- برخی از مشاهدات محدود یا به دست آوردن آنها دشوار است.
- مهم باشد که گفته‌ها در مورد زیر گروه‌ها از نظر آماری معتبر باشند.
- زیر گروه‌های جمعیت، واریانس‌های مختلفی برای متغیرهای مورد نظر داشته باشد.
- هزینه‌های جمع‌آوری داده بین زیر گروه‌های جمعیت متفاوت باشد.

رهنمود ۵) نمونه‌گیری سیستماتیک. در موارد زیر انتخاب نمونه‌گیری سیستماتیک را مد نظر قرار دهید:

- اگر شناسایی موارد با استفاده از روش نمونه‌گیری تصادفی ساده دشوار باشد.
- در صورتی که مهم باشد از روش نمونه‌گیری احتمالی که اجرای ساده دارد، استفاده کنیم.
- در صورتی که بازه نمونه‌گیری در دسترس نباشد یا تهیه آن غیر عملی باشد ولی جریان عناصر نماینده جمعیت در دسترس باشد.
- لیست‌برداری از جمعیت اصولاً تصادفی بوده یا این که بتواند تصادفی شود.

مثال در رابطه با استفاده از روش‌های نمونه‌گیری

تصمیم گرفته‌ایم که یک دبیرستان را بررسی کنیم. می‌خواهیم داده‌هایی در مورد تعداد دانش‌آموزان چپ دست، تعداد دانش‌آموزانی که در خانه اینترنت دارند، قد دانش‌آموزان و پول توجیبی که هر هفته دریافت می‌کنند داشته باشیم.

درک فایده دانستن تعداد دانش‌آموزان چپ دست یک دبیرستان راحت است، زیرا دبیرستان باید تجهیزات مناسبی، مثلاً صندلی‌های مناسب، را برای آنها داشته باشد.

داشتن اینترنت در خانه نیز یکی از اطلاعات مهم است. از این اطلاعات می‌توان برای کنترل این که آیا این احتمال وجود دارد برخی از تکالیف را از طریق اینترنت بدهیم و همچنین آیا آنها به اطلاعات مشخص شده دیگری در وب دسترسی دارند یا خیر استفاده کنیم.

تحقیق در مورد قد نیز یک موضوع کلاسیک است. همچنین جالب است که بدانیم آیا قد دانش‌آموزان سالانه تغییر کرده است.

پول توجیبی یک داده اجتماعی است. همچنین جالب است بدانیم که دانش‌آموزان چقدر پول دارند و چگونه آن را خرج کرده و وقت خودشان را چگونه صرف می‌کنند.

پس از آن که آن چه که می‌خواهیم را به دست آوردیم، تصمیم گرفتیم نمونه‌گیری کرده و بدون پرسش از آنها نتایجی در مورد دانش‌آموزان دبیرستان به دست آوریم. اطلاعات در دسترس ما اطلاعات راجع به توزیع دانش‌آموزان بر اساس سال و کلاس است.

	A	B	C	D	E	جمع
سال اول	۳۳	۲۰				۵۳
سال دوم	۲۰	۱۵	۳۰			۶۵
سال سوم	۲۰	۱۵	۲۶	۱۴		۷۵
سال چهارم	۲۷	۲۷	۲۵			۷۹
سال پنجم	۳۳	۲۸	۳۰	۳۱	۲۳	۱۴۵
سال ششم	۳۰	۳۴	۳۲	۳۱		۱۲۷

بنابراین ما با جمعیتی متشکل از ۵۴۴ دانش‌آموز دبیرستانی سروکار داریم.

ما با طرح این موضوع شروع می‌کنیم که می‌خواهیم از نمونه‌ای با اندازه حدود ۶۰ دانش‌آموز که حداکثر مجاز است استفاده کنیم و فکر می‌کنیم این مقدار برای تحقیقی که قرار است انجام دهیم کافی باشد.

بنابراین می‌توانیم اولین اطلاعات را به دست آوریم، نسبت نمونه‌گیری ما برابر است با:

$$f = \frac{n}{N} = \frac{60}{544} = 0.1102$$

یعنی این که، می‌خواهیم حدود ۱۱ درصد جمعیت را نمونه‌گیری کنیم. همچنین می‌توانیم ضریب نما را به صورت زیر محاسبه کنیم:

$$E = \frac{N}{n} = \frac{544}{60} = 9.1$$

یعنی این که هر دانش‌آموز نماینده ۹ همکلاسی است.

حالا باید مشخص کنیم که از چه روشی برای نمونه‌گیری از خصوصیات مختلفی که قرار است روی آنها بررسی کنیم استفاده نماییم. فرض کنید که برای نشان دادن این خصوصیات از علایم زیر استفاده شده است:

- X نشاندهنده قد.
- Y نشاندهنده پول توجیبی.
- Z نشاندهنده متغیر چپ دست بودن است که اگر دانش‌آموز چپ دست باشد یک و راست دست باشد مقدار صفر را دارد.
- I نشاندهنده متغیر «داشتن اینترنت در خانه» است که اگر جواب مثبت باشد مقدار آن یک و در غیر این صورت صفر است.

ما ۴ متغیر را به دو حالت متفاوت تقسیم می‌کنیم. اول این که، ما از خودمان سوال می‌کنیم که اگر ما جمعیت را به گروه‌ها و سطوح تقسیم کنیم، می‌توانیم در نظر بگیریم که این تقسیم‌بندی بر این متغیرها تاثیرگذار است؟ یعنی این که، به عنوان مثال در هر سطحی می‌توانیم فکر کنیم که میانگین قد می‌تواند تغییر کند؟ پاسخ به این پرسش این است که منطقی این است فکر کنیم که این متغیر تغییر خواهد کرد. از روی قیاس، می‌توانیم فرض کنیم که سن تاثیر زیادی روی قد دارد. و برای پول توجیبی چه؟ پس سن نیز مهم است زیرا

هر چه سن ما بالاتر می‌رود می‌توانیم پول توجیبی بیشتری از پدر و مادرمان بگیریم. آیا این وضعیت برای چپ دست بودن نیز اتفاق می‌افتد؟ پاسخ منفی است زیرا اگر ما چپ دست باشیم، این وضعیت از همان روزی که به دنیا می‌آییم رخ می‌دهد بنابراین سن روی این متغیر تاثیری ندارد. همین وضعیت برای داشتن اینترنت در خانه نیز مطرح است. بنابراین برای این دو حالت از روش‌های نمونه‌گیری متفاوتی استفاده می‌کنیم.

حالت I: متغیرهای پول توجیبی و قد

قبلاً گفتیم که ما جمعیت را به سطوح و گروه‌هایی تقسیم کرده‌ایم. برای ما، تقسیم‌بندی به سطوح همان تقسیم بندی به طبقه است زیرا سطوح در درون خودشان نسبت به سن همگن هستند (و همچنین می‌توانیم فکر کنیم که همین وضعیت برای پول توجیبی و قد نیز اتفاق می‌افتد)، و همان طور که قبلاً گفتیم، سن تاثیر زیادی روی این متغیرها دارد و این باعث می‌شود که ما علاقمند به داشتن این طبقات در نمونه باشیم. بنابراین برای این حالت از نمونه‌گیری تصادفی طبقه بندی استفاده می‌کنیم.

کار بعدی که باید انجام دهیم این است که در مورد اندازه نمونه درون هر لایه تصمیم‌گیری کنیم. ما شش لایه با اندازه‌های زیر داریم:

طبقات	اندازه
سطح اول (طبقه اول)	N۱ = ۵۳
سطح دوم (طبقه دوم)	N۲ = ۶۵
سطح سوم (طبقه سوم)	N۳ = ۷۵
سطح چهارم (طبقه چهارم)	N۴ = ۷۹
سطح پنجم (طبقه پنجم)	N۵ = ۱۴۵
سطح ششم (طبقه ششم)	N۶ = ۱۲۷

کار منطقی در این وضعیت آن است که از اندازه نمونه هر طبقه متناسب با اندازه آن

استفاده کنیم، به نحوی که اندازه نمونه‌ها همان نسبت اندازه‌های طبقه را داشته باشند. سپس اندازه نمونه هر طبقه را با استفاده از عبارت زیر محاسبه می‌کنیم:

$$n_i = n.\frac{N_i}{N}$$

و اندازه‌های نمونه را به صورت زیر به دست می‌آوریم:

$$n_1 = 6, \quad \text{بنابراین داریم} \quad n_1 = 60.\frac{53}{544} = 5.84$$

$$n_2 = 8, \quad \text{بنابراین داریم} \quad n_2 = 60.\frac{65}{544} = 7.16$$

$$n_3 = 8, \quad \text{بنابراین داریم} \quad n_3 = 60.\frac{75}{544} = 8.27$$

$$n_4 = 8, \quad \text{بنابراین داریم} \quad n_4 = 60.\frac{79}{544} = 8.71$$

$$n_5 = 16, \quad \text{بنابراین داریم} \quad n_5 = 60.\frac{145}{544} = 15.99$$

$$n_6 = 14, \quad \text{بنابراین داریم} \quad n_6 = 60.\frac{127}{544} = 14.00$$

که در آن اندازه‌ها به گونه‌ای حفظ شده که اندازه نمونه ۶۰ که گفته بودیم حفظ شود. بنابراین اندازه‌های نمونه مورد نیاز را داشته و می‌توانیم درون هر طبقه از نمونه تصادفی استفاده کنیم تا تعداد دانش‌آموزانی که از قبل تصمیم گرفته بودیم انتخاب شود.

داده‌های ما به شرح زیر هستند: برای قد داریم:

طبقه۱	۱۶۵	۱۶۱	۱۵۳	۱۵۰	۱۵۱	۱۵۳										
طبقه۲	۱۵۷	۱۶۱	۱۶۸	۱۶۲	۱۶۵	۱۷۱	۱۶۹	۱۶۴								
طبقه۳	۱۶۸	۱۶۵	۱۷۵	۱۶۵	۱۶۵	۱۶۳	۱۶۵	۱۶۵								
طبقه۴	۱۶۴	۱۷۱	۱۷۷	۱۶۳	۱۷۰	۱۶۵	۱۶۰	۱۷۵								
طبقه۵	۱۷۵	۱۷۳	۱۶۱	۱۵۸	۱۷۵	۱۶۴	۱۵۸	۱۶۱	۱۵۸	۱۷۱	۱۷۵	۱۷۰	۱۸۷	۱۶۸	۱۷۰	۱۸۵
طبقه۶	۱۹۰	۱۷۸	۱۹۴	۱۸۳	۱۶۵	۱۷۰	۱۷۶	۱۷۳	۱۶۸	۱۸۳	۱۷۳	۱۸۳	۱۷۴	۱۷۷		

and for the pocket money:

طبقه۱	۱۰	۰	۳/۵	۰	۰	۳										
طبقه۲	۰	۵	۰	۱۵	۰	۳	۲	۰								
طبقه۳	۵	۸	۸	۰	۲۰	۵	۱۰	۱۰								
طبقه۴	۱۲	۶	۵	۱۲	۱۲	۶	۰	۰								
طبقه۵	۵	۱۰	۱۲	۱۵	۱۰	۱۲	۳۰	۱۲	۳۰	۱۰	۶	۵	۱۰	۲۱	۴۰	۱۵
طبقه۶	۱۲	۱۰	۹	۶	۸	۹/۴	۱۵	۰	۲۰	۱۰	۱۵	۱۰	۰	۰		

حالا شروع به برآورد می‌کنیم. اولین کاری که باید انجام دهیم این است که مقدار میانگین هر طبقه را حساب کنیم که به ما اطلاعاتی در مورد رفتار متغیرهای درون طبقه می‌دهد. در ادامه، میانگین قد و پول توجیبی دانش‌آموزان دبیرستان را محاسبه کرده و همراه با برآوردی از خطا مشخص می‌کنیم.

ما روند هر کار را برای هر متغیر به صورت تصادفی انجام می‌دهیم:

برای قد داریم

Stratus	Average	Std deviation
1	$\overline{x}_1 = 155.5$	$S^2_{x1} = 36.7$
2	$\overline{x}_2 = 164.625$	$S^2_{x2} = 21.4107$
3	$\overline{x}_3 = 167.625$	$S^2_{x3} = 22.5535$
4	$\overline{x}_4 = 168.125$	$S^2_{x4} = 36.6964$
5	$\overline{x}_5 = 169.3125$	$S^2_{x5} = 81.6958$
6	$\overline{x}_6 = 177.642857$	$S^2_{x6} = 67.478$

مستقیماً می‌توانیم ببینیم که بعضی ارقام عجیب هستند. با افزایش سطح، میانگین نیز افزایش می‌یابد.

این باعث می‌شود که فکر کنیم انتخاب نمونه‌گیری لایه‌ای برای این وضعیت درست بوده است.

حالا همین محاسبات را برای پول توجیبی انجام می‌دهیم:

Stratus	Average	Std deviation
1	$\overline{y}_1 = 2.75$	$S^2_{y1} = 4.026$
2	$\overline{y}_2 = 3.125$	$S^2_{y2} = 26.4107$
3	$\overline{y}_3 = 8.25$	$S^2_{y3} = 33.3571$
4	$\overline{y}_4 = 6.625$	$S^2_{y4} = 25.4107$
5	$\overline{y}_5 = 15.1875$	$S^2_{y5} = 101.2291$
6	$\overline{y}_6 = 8.8857$	$S^2_{y6} = 35.229$

حالا میانگین برآورد شده حاصل از نمونه کلی و برآورد خطای برآورد واریانس ۲ متغیر مورد مطالعه را محاسبه می‌کنیم. برای قد:

$$\widehat{\overline{X}} = \sum_{h=1}^{6} w_h \overline{x}_h = \sum_{h=1}^{6} \frac{N_h}{N} \overline{x}_h = \frac{53}{544} \cdot 155.5 + \frac{65}{544} \cdot 164.625 + \frac{75}{544} \cdot 167.625 + \frac{79}{544} \cdot 168.125$$

$$+ \frac{145}{544} \cdot 169.3125 + \frac{127}{544} \cdot 177.642857 = 168.9463.$$

حالا همین محاسبات را برای پول توجیبی انجام می‌دهیم. ما با محاسبه میانگین برآوردی شروع می‌کنیم:

$$\widehat{\overline{Y}} = \sum_{h=1}^{6} w_h \overline{y}_h = \sum_{h=1}^{6} \frac{N_h}{N} \overline{y}_h = \frac{53}{544} \cdot 2.75 + \frac{65}{544} \cdot 3.125 + \frac{75}{544} \cdot 8.25 + \frac{79}{544} \cdot 6.625$$

$$+ \frac{145}{544} \cdot 15.1875 + \frac{127}{544} \cdot 8.8857 = 8.8633.$$

حالت II: متغیرهای «چپ دست بودن» و «داشتن اینترنت در خانه»

حالا می‌خواهیم متغیرهای «چپ دست بودن» و «داشتن اینترنت در خانه» را بررسی کنیم. به راحتی می‌توان مشاهده کرد که تقسیم کردن به لایه برای این حالت مناسب نیست بنابراین روی استفاده از روش‌های دیگر نمونه‌گیری فکر می‌کنیم. ما بازهم می‌خواهیم نمونه‌ای با ۶۰ دانش‌آموز داشته باشیم. در رابطه با این متغیرها، می‌توانیم فکر کنیم که گروه‌هایی که جمعیت به آنها تقسیم شده شبیه به جمعیت‌های کوچک است یعنی می‌توانیم فکر کنیم که گروه‌ها همانند کل دبیرستان عمل می‌کنند. به علاوه برای ما احتمال نمونه‌گیری از برخی گروه‌ها جذاب است زیرا انتخاب یک نمونه تصادفی از دانش‌آموزان، یافتن آنها و مصاحبه با آنها کار ساده‌ای نیست.

ولی حالا در مورد گروه‌ها چه؟ ما قبلاً گفتیم که گروه‌ها در درون خودشان برای متغیرهای ما همانند جمعیت‌های کوچک عمل می‌کنند در حالی که بین خودشان مشابه هستند. این بدان معنی است که ما باید جمعیت را به خوشه‌ها تقسیم کنیم و بنابراین برای این وضعیت از نمونه‌گیری خوشه‌ای استفاده می‌کنیم.

کار بعدی که باید انجام دهیم شماره‌گذاری گروه‌هایی که قرار است نمونه‌گیری شوند است. می‌دانیم که اندازه گروه‌ها برابر نیست ولی ۲ یا ۳ گروه به ما تضمین می‌دهند که نمونه شامل حدود ۶۰ دانش‌آموز خواهد بود. برای اجتناب از احتمال داشتن نمونه‌ای از دو گروه کوچک و به دست آوردن نمونه بسیار کوچکی برای اهدافمان، تصمیم گرفتیم که از دبیرستان سه گروه را انتخاب کنیم.

بنابراین داده‌هایی که به دست آوردیم به شرح زیر است. برای متغیر «چپ دست بودن»:

خوشه ۱: 00000100000000010001,

خوشه ۲: 00000000000000000000,

خوشه ۳: 00000000100010000000000,

که در آن ۱ به معنی چپ دست بودن و صفر به معنی راست دست بودن است. حالا برای متغیر «داشتن اینترنت در خانه» داریم:

خوشه ۱: 00100101111010101001,

خوشه ۲: 01111111110010110110111,

خوشه ۳: 11110111011111110111101011,

که در آن ۱ به معنی داشتن اینترنت در خانه و صفر به معنی نداشتن اینترنت در خانه است.

حالا شروع به تعیین تعداد کلی دانش‌آموزان چپ دست، نسبت دانش‌آموزان چپ دست و هم چنین تعداد کلی دانش‌آموزانی که دارای اینترنت در خانه هستند و نسبت این دانش‌آموزان به کل دبیرستان می‌کنیم.

تعداد کل و نسبت هر گروه و متغیر را محاسبه می‌کنیم:

	چپ دست بودن		اینترنت	
خوشه	جمع	نسبت	جمع	نسبت
۱	۳	۰/۱۵	۱۰	۰/۵
۲	۰	۰	۱۷	۰/۷۳۹۱
۳	۲	۰/۰۸	۲۰	۰/۸

حالا می‌توانیم برآوردهایی برای نسبت و کل متغیرهای Z و I محاسبه کنیم. ما با متغیر Z شروع می‌کنیم:

$$\widehat{Z} = M \cdot \frac{\sum_{i=1}^{n} \widehat{Z}_i}{\sum_{i=1}^{n} M_i} = 544 \cdot \frac{\sum_{i=1}^{3} \widehat{Z}_i}{\sum_{i=1}^{3} M_i} = 544 \cdot \frac{3+0+2}{20+23+25} = 544 \cdot \frac{5}{68} = 40,$$

$$\widehat{P_Z} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} M_i} = \frac{3+0+2}{20+23+25} = \frac{5}{68} = 0.0735,$$

و همین کار را برای متغیر I انجام می‌دهیم :

$$\widehat{I} = M \cdot \frac{\sum_{i=1}^{n} \widehat{I}_i}{\sum_{i=1}^{n} M_i} = 544 \cdot \frac{\sum_{i=1}^{3} \widehat{I}_i}{\sum_{i=1}^{3} M_i} = 544 \cdot \frac{10+17+20}{20+23+25} = 544 \cdot \frac{47}{68} = 376,$$

$$\widehat{P_I} = \frac{\sum_{i=1}^{n} A_i}{\sum_{i=1}^{n} M_i} = \frac{10+17+20}{20+23+25} = \frac{47}{68} = 0.6911.$$

مثالهای تکنیک‌های اساسی نمونه‌گیری احتمالی

نمونه‌گیری تصادفی ساده

فرض کنید که N=۸۵۰ دانش‌آموز در یک مدرسه وجود دارند و قرار است از بین آنها یک نمونه متشکل از n=۱۰ دانش‌آموز انتخاب شوند. دانش‌آموزها از ۱ تا ۸۵۰ شماره‌گذاری شدند. چون جمعیت ما سه رقمی است از اعداد تصادفی که شامل سه عدد هستند استفاده می‌کنیم. تمامی اعدادی که بالاتر از ۸۵۰ هستند نادیده گرفته شده‌اند زیرا آنها تناسبی با اعداد درون جمعیت ندارد. در صورتی که عدد تکراری آمد، عدد تکراری نادیده گرفته می‌شود. با استفاده از این قواعد نمونه تصادفی ساده متشکل از ۱۰ دانش‌آموز به دست آمده که برای آن از جدول اعداد تصادفی استفاده شده است.

۰۷۴	۴۹۵	۲۱۴	۵۴۶	۲۵۱
۶۲۸	۵۱۳	۵۰۲	۴۰۷	۸۰۰

نمونه‌گیری سیستماتیک

فرض کنید جمعیتی داریم که شامل فقط N=۱۰۰ نفر بوده و می‌خواهیم یک نمونه n=۲۰ نفر را از بین آن انتخاب کنیم. برای استفاده از نمونه‌گیری سیستماتیک، جمعیت را باید به صورت تصادفی لیست کنیم. نسبت نمونه‌گیری برابر با n/N = ۲۰/۱۰۰ = ۲۰٪ خواهد بود. در این صورت، اندازه تناوب، k، برابر است با N/n = ۱۰۰/۲۰ = ۵. حالا، یک عدد صحیح بین ۱ تا ۵ انتخاب می‌کنیم. در مثال ما، فرض کنید که شما ۴ را انتخاب کرده‌اید. حالا، برای انتخاب نمونه، از واحد ۴ام درون لیست شروع کرده و هر واحد kام را انتخاب می‌کنیم (هر ۵ام زیرا k=۵). شما باید واحدهای نمونه‌گیری ۴، ۹، ۱۴، ۱۹ و الی آخر را انتخاب و ۲۰ واحد را در نمونه‌تان داشته باشید.

نمونه‌گیری غیراحتمالی

اختلاف بین نمونه‌گیری احتمالی و غیر احتمالی این است که نمونه‌گیری غیر احتمالی شامل انتخاب تصادفی نیست ولی نمونه‌گیری احتمالی شامل انتخاب تصادفی می‌باشد. نمونه‌های غیر احتمالی به مبنای تئوری احتمال بستگی ندارند. در مورد نمونه احتمالی احتمال این که جمعیت را به درستی نشان دهیم را می‌دانیم. ما قادر به برآورد فاصله اعتماد آماری هستیم. در نمونه‌های غیر احتمالی ممکن است جمعیت را نیز نشان دهیم ولی در اغلب موارد دانستن این که تا چه اندازه این کار را به خوبی انجام داده‌ایم دشوار است. به طور کلی، محققان روش‌های نمونه‌گیری احتمالی یا تصادفی را بر روش‌های غیر احتمالی ترجیح داده و آنها را دقیق‌تر و منسجم‌تر می‌دانند. البته، در تحقیقات اجتماعی ممکن است شرایطی وجود داشته باشد که انجام نمونه‌گیری تصادفی ممکن یا عملی نباشد یا این که از نظر تئوری انجام آن عاقلانه نباشد. در این صورت انواع روش‌های غیر احتمالی را مد نظر قرار می‌دهیم.

می‌توانیم روش‌های نمونه‌گیری احتمالی و غیر احتمالی را به دو نوع کلی تقسیم کنیم: تصادفی یا هدفمند. بیشتر روش‌های نمونه‌گیری ماهیت هدفمند دارند زیرا ما معمولاً با استفاده از یک برنامه مشخص در ذهن سراغ مساله نمونه‌گیری می‌رویم. مهمترین تفاوت بین این روش‌های نمونه‌گیری، تفاوت بین انواع مختلف روش‌های نمونه‌گیری هدفمند است.

نمونه‌گیری تصادفی، اتفاقی، یا راحت

یکی از متداول‌ترین روش‌های نمونه‌گیری که عناوین مختلف دارد در زیر معرفی شده است. در این گروه بطور مثال مصاحبه های معمولی با افراد داخل خیابان قرار می گیرد که که همواره توسط برنامه‌های خبری تلویزیون برای به دست آوردن (البته نه نماینده) نظرات عموم انجام می‌شوند. همچنین دلیل عمده استفاده از دانشجویان در بیشتر تحقیقات روان‌شناسی بحث راحتی کار است. (شما که واقعا باور ندارید که روانشناسان به این دلیل که دانشجویان نماینده جمعیتی بزرگ هستند، از آنها استفاده می‌کنند؟) در روش بالینی، می‌توانیم از مشتریانی که به عنوان نمونه در دسترس ما هستند استفاده کنیم. در بسیاری

از حوزه‌های تحقیقی، ما خیلی راحت با جستجوی افراد داوطلب نمونه‌گیری را انجام می‌دهیم. به طور مشخص، مشکلی که در رابطه با این نوع نمونه‌ها وجود دارد این است که ما شواهدی دال بر این که نشان دهد آنها نماینده جمعیت‌های که ما مایل به تعمیم به آنها هستیم هستند وجود ندارد- و در بسیاری از موارد مشخصاً انتظار هم نداریم که آنها نماینده جمعیت مورد نظر باشند.

نمونه‌گیری هدفمند

در نمونه‌گیری هدفمند، نمونه‌گیری با توجه به هدفی که در ذهن وجود دارد انجام شده است. ما معمولاً یک یا چند گروه مشخص از پیش تعیین شده داریم که در موردشان تحقیق می‌کنیم. به عنوان مثال، آیا شما تا به حال در بازار یا خیابان با افرادی برخورد کرده‌اید که پلاکاردی را با خود حمل می‌کنند و افراد مختلف را متوقف کرده و از آنها می‌پرسند که می‌توانند با آنها مصاحبه کنند؟ به احتمال زیاد آنها نمونه‌گیری هدفمند را انجام می‌دهند (و به احتمال زیاد در مورد بازار تحقیق می‌کنند). ممکن است آنها به دنبال خانم‌های سفید پوست ۴۰-۳۰ ساله بگردند. آنها افراد عبوری را تحت نظر گرفته و هر کسی که به نظر برسد در این گروه قرار دارد را متوقف کرده و از آنها می‌پرسند که می‌توانند در کارشان مشارکت داشته باشند. یکی از اولین کارهایی که احتمالاً آنها انجام می‌دهند این است که مشخص کنند آیا پاسخ دهندگان واقعاً معیار لازم برای قرار گرفتن در نمونه را دارند یا خیر. نمونه‌گیری هدفمند می‌تواند برای مواردی مفید باشد که در آنها لازم است به سرعت به نمونه هدف دست پیدا کنید و نگرانی اصلی نمونه‌گیری تناسبی نیست. شما با استفاده از نمونه هدفمند احتمالاً به دنبال این هستید که نظرات جمعیت هدف‌تان را به دست آورید، ولی احتمالاً به دنبال این نیز خواهید بود که وزن بیشتری به گروه‌های فرعی درون جمعیت‌تان که راحت‌تر در دسترس هستند بدهید.

تمامی روش‌هایی که دنبال می‌شوند را می‌توان به عنوان گروه‌های فرعی از روش‌های نمونه‌گیری هدفمند در نظر گرفت. ممکن است از گروه‌های خاص یا انواع افراد به عنوان نمونه‌گیری نمایی، حرفه‌ای یا سهمیه‌ای، نمونه‌گیری کنیم. ممکن است از تنوع به عنوان نمونه‌گیری ناهمگن نمونه‌گیری کنیم. یا این که ممکن است روی شبکه‌های اجتماعی غیر

رسمی سرمایه‌گذاری کنیم تا پاسخ‌دهندگان خاصی که پیدا کردن آنها به غیر از این روش دشوار هستند را شناسایی کنیم، مثلاً نمونه‌گیری گلوله برفی. در تمامی این روش‌ها می‌دانیم که چه می‌خواهیم- نمونه‌گیری را با هدف انجام می‌دهیم.

نمونه‌گیری از نمونه نمایی

در آمار، نما تکرارترین مقدار در یک توزیع است. در نمونه‌گیری، وقتی که از یک نمونه نمایی نمونه‌گیری می‌کنیم، متداول‌ترین حالت یا حالت بارز را نمونه‌گیری می‌کنیم. به عنوان مثال، در برخی از نظرسنجی‌های عمومی غیررسمی با یک رای دهنده بارز مصاحبه می‌شود. در رابطه با این نوع روش نمونه‌گیری مشکلاتی وجود دارد. اول، چگونه بدانیم حالت بارز یا نما چیست؟ می‌توانیم بگوییم که رای‌دهنده نمایی فردی است که در جمعیت دارای سن، سطح تحصیلات و درآمد متوسط است. ولی، مشخص نیست که استفاده از متوسط این موارد بی‌طرفانه باشد (به عنوان مثال توزیع چوله در درآمد را در نظر بگیرید). و چگونه می‌دانید که این سه متغیر--سن، تحصیلات، درآمد-- فقط یا حتی مرتبط‌ترین عوامل برای طبقه‌بندی رای‌دهنده نماینده است. آیا مذهب یا قومیت یک متمایزکننده مهم نیست؟ مشخص است که نمونه‌گیری از نمونه نمایی فقط در حوزه نمونه‌گیری غیررسمی منطقی است.

نمونه‌گیری خبره‌ای

نمونه‌گیری خبره‌ای شامل سرهم کردن نمونه‌ای از افراد با تجربه مشخص یا قابل اثبات و خبره در برخی مناطق است. در اغلب موارد، چنین نمونه‌ای را تحت نظر گروهی از خبرگان گردآوری می‌کنند. در واقع دو دلیل برای انجام نمونه‌گیری خبره‌ای وجود دارد. اول، این روش بهترین راه برای استخراج دیدگاه افرادی که دارای تجربه خاصی هستند می‌باشد. در این صورت نمونه‌گیری خبره‌ای یک حالت فرعی خاص از نمونه‌گیری هدفمند است. دلیل دیگر برای استفاده شما از نمونه‌گیری خبره‌ای فراهم کردن شواهدی برای تایید روش دیگر نمونه‌گیری است که انتخاب کرده‌اید. به عنوان مثال، شما می‌خواهید نمونه‌گیری از یک نمونه‌گیری نمایی را انجام داده و نگران این هستید که معیار استفاده شده برای تعریف نمونه

نمایی مورد انتقاد قرار گیرد. ممکن است شما یک گروه خبره متشکل از افرادی با تجربه تایید شده و صاحب‌نظر در آن حوزه یا موضوع تشکیل داده و از آنها بخواهید که تعاریف نمای شما بررسی کرده و در مورد مناسب بودن و تایید آن اظهار نظر کنند. مزیت انجام این کار آن است که شما به تنهایی از تصمیمات‌تان دفاع نمی‌کنید-- شما افراد متخصصی دارید که از شما پشتیبانی می‌کنند. عیب این روش آن است که حتی متخصصان نیز، و غالباً، اشتباه می‌کنند.

نمونه‌گیری سهمیه‌ای

در نمونه‌گیری سهمیه‌ای، شما افراد را به صورت غیر تصادفی و بر اساس سهمیه ثابت انتخاب می‌کنید. دو نوع نمونه‌گیری سهمیه‌ای وجود دارد: تناسبی و غیر تناسبی در نمونه‌گیری سهمیه‌ای تناسبی شما می‌خواهید با نمونه‌گیری از مقدار تناسبی جمعیت مشخصه‌های اصلی جمعیت را نشان دهید. به عنوان مثال، اگر شما بدانید که جمعیت شامل ۴۰ درصد زن و ۶۰ درصد مرد است و بخواهید اندازه کلی نمونه ۱۰۰ نفر باشد، نمونه‌گیری را آن قدر ادامه می‌دهید تا به این درصدها دست پیدا کرده و پس از دستیابی به این درصدها کار را متوقف می‌کنید. بنابراین، اگر شما از قبل ۴۰ زن را در نمونه داشته باشید ولی تعداد مردان ۶۰ نفر نباشند، به نمونه‌گیری از مردان ادامه می‌دهید حتی اگر زنانی که مشخصات مورد نظر را دارند وجود داشته باشند شما از آنها نمونه‌گیری نخواهید کرد زیرا قبلاً سهمیه‌تان را کسب کرده‌اید. مشکلی که در این جا وجود دارد (همانند بسیاری از نمونه‌گیری‌های هدفمند) این است که شما باید در مورد مشخصه‌های خاصی که مبنای سهمیه هستند تصمیم‌گیری کنید. آیا سهمیه بر اساس جنس، سن، تحصیلات، نژاد، مذهب و غیره است؟

نمونه‌گیری سهمیه‌ای غیر تناسبی سخت‌گیری کمتری دارد. در این روش، شما حداقل تعداد واحدهای نمونه‌گیری شده‌ای که می‌خواهید در هر گروه وجود داشته باشند را تعیین می‌کنید. در این جا، شما نگران این نیستید که تعدادی که مطابق با تناسب درون جمعیت باشد داشته باشید. در عوض شما خیلی راحت به دنبال این هستید که تعداد کافی داشته باشید تا تضمین شود شما قادر به صحبت در مورد گروه‌های کوچکتر درون جمعیت هستید. این روش نوع غیر احتمالی نمونه‌گیری تصادفی طبقه بندی است که معمولاً برای اطمینان از این که گروه‌های کوچکتر به طور مناسب در نمونه شما نشان داده شده، استفاده شده است.

نمونه‌گیری ناهمگن

ما زمانی از ناهمگنی نمونه‌گیری می‌کنیم که بخواهیم تمام نظرات یا دیدگاه‌ها را در نظر بگیریم، و در مورد نشان دادن این دیدگاه‌ها به صورت تناسبی نگرانی نداریم. نام دیگر این نمونه‌گیری، نمونه‌گیری برای تنوع است. در بسیاری از فرایندهای طوفان ذهنی یا گروه اسمی (شامل تعیین مفهوم)، ما از اشکالی از نمونه‌گیری ناهمگن استفاده می‌کنیم زیرا هدف اصلی ما این است که طیف وسیعی از ایده‌ها را به دست اوریم نه این که ایده‌های «میانگین» یا «نمونه نمایی» را شناسایی کنیم. در واقع آن چه که می‌خواهیم نمونه‌گیری کنیم ایده‌ها هستند نه افراد. ما تصور می‌کنیم که جامعه‌ای از تمامی ایده‌های احتمالی مرتبط با چند موضوع وجود دارند و ما می‌خواهیم از این جمعیت نمونه‌گیری کنیم نه از جمعیتی از افرادی که دارای ایده هستند. مشخص است که برای به دست آوردن تمامی ایده‌ها و به خصوص ایده‌های برجسته یا غیرعادی، باید شرکت‌کنندگان گسترده و متنوعی را در نظر بگیریم. در این صوت نمونه‌گیری ناهمگن تقریباً در نقطه مقابل ناهمگنی از نمونه نمایی قرار دارد.

نمونه‌گیری گلوله برفی

در نمونه‌گیری گلوله برفی، شما از شناسایی افرادی که معیار لازم برای قرار گرفتن در تحقیق شما را دارند شروع می‌کنید. سپس از آنها می‌خواهید که به دیگرانی که می‌دانند آنها نیز معیار مورد نظر را دارند سفارش کنند. هر چند این روش به سختی منجر به نمونه‌های نماینده می‌شود، ولی مواقعی است که این روش بهترین روش در دسترس است. نمونه‌گیری گلوله برفی به خصوص زمانی مفید است که می‌خواهید به جمعیت‌های دست پیدا کنید که در دسترس نیستند یا این که یافتن آنها دشوار است. به عنوان مثال، اگر شما در مورد افراد بی‌خانمان تحقیق می‌کنید، احتمالاً قادر به یافتن لیست خوبی از افرادی بی‌خانمان در یک منطقه جغرافیایی خاص نخواهید بود. البته، اگر شما به منطقه بروید و یک یا دو نفر را شناسایی کنید، ممکن است متوجه شوید که آنها اطلاعات بسیار خوبی در مورد سایر افراد بی‌خانمان در آن اطراف و این که چگونه می‌توانید آنها را پیدا کنید دارند.

رفع مشکلات نمونه‌گیری

رفع مشکلات مرتبط با بازه‌های غیر منطبق نمونه‌گیری

در اغلب موارد، بازه نمونه‌گیری با واحد اولیه نمونه‌گیری انطباق ندارد. به عنوان مثال، بازه نمونه‌گیری ممکن است لیستی از ساکنین باشد ولی واحدهای نهایی نمونه‌گیری افراد باشند. مصاحبه کننده به دستورالعملی نیاز دارد تا بر اساس آن فرد درون یک خانه را برای به عنوان یک موضوع انتخاب کند. این کار با استفاده از شبکه انتخاب انجام می‌شود. شبکه انتخاب ممکن است اشکال متفاوتی داشته باشد. به عنوان مثال، در ارزیابی مالیات‌دهندگان، برای سادگی کار فرض می‌شود که یک خانه می‌تواند حداکثر سه مالیات دهنده داشته باشد. فرض کنید که محقق مالیات دهندگان را بر اساس سن از ۱ تا ۳ شماره‌گذاری کرده و از شبکه انتخاب زیر استفاده می‌کند:

		If no. of taxpayers in residence is		
		1	2	3
Survey form	% of all forms	Interview taxpayer #		
A	1/3	1	1	1
B	1/6	1	1	2
C	1/6	1	2	2
D	1/3	1	2	3

در این مثال، چهار شکل مختلف ارزیابی وجود دارد (A,B,C,D)، که این چهار ارزیابی چاپ و به صورت تصادفی و بر اساس تناسب جدول بالا توزیع شده‌اند. هر فرم ارزیابی مربوط به یکی از سطرهای جدول است. به عنوان مثال، در فرم B، اگر دو مالیات دهنده در خانه وجود داشته باشد در این صورت مصاحبه کننده با مالیات دهنده شماره ۱ مصاحبه خواهد کرد. برای فرم‌های C یا D، با مالیات دهنده شماره ۲ مصاحبه می‌شود. اگر در خانه یک مالیات دهنده وجود داشته باشد، در این صورت بدون توجه به فرم با آن مالیات دهنده (شماره ۱)

مصاحبه شده است. اگر دو یا سه مالیات دهنده وجود داشته باشد، هر دو یا هر سه شانس برابر برای این که مورد مصاحبه قرار گیرند دارند.

کالتون (۱۹۸۳:۶۱) جدول مشابهی برای فرض وجود حداکثر چهار موضوع قابل انتخاب در هر خانه ارائه کرده است. برای هر فرضی می‌توان جداول مشابهی تهیه کرد. تمامی این جداول شبکه احتمال این که هر کدام از افراد در نمونه قرار بگیرند را برابر می‌کند.

به طور مشابه، روش ساده‌تری وجود دارد به نام روش «آخرین روز تولد»، که در آن محقق تعداد افراد بالغ درون خانه را سوال کرده، سپس با فرد مجرد درون خانه‌های یک نفره مصاحبه می‌کند؛ در دفعات بعدی با جوانترین فرد درون یک خانه دو نفره و دفعه دیگر با نفر دیگر مصاحبه می‌کند؛ دفعه سوم با افراد درون خانه‌های سه نفره و همین طور الی آخر مصاحبه می‌کند. به جای این که سن اهمیت داشته باشد از تاریخ‌های تولد سوال می‌شود. برخی شواهد نشان می‌دهند که در ارزیابی‌های تلفنی، این روش نیاز به زمان بیشتری برای غربال‌گیری و مصاحبه دارد و ممکن است تماس‌های تلفنی مجدد یا امتناع از پاسخگویی زیادی وجود داشته باشد. در مصاحبه‌های رو در رو تغییرات موضوعی بیشتری برای سوالات غربالگری وجود دارد.

افزایش نرخ پاسخ

نرخ پاسخ در مصاحبه‌های رو در رو معمولاً حدود ۷۵ درصد است، و برای ارزیابی‌های پستی در مورد برخی از ارزیابی‌های بازاریابی نرخ ۱۰ درصد نرخ خوبی در نظر گرفته شده است. البته، «خوب» به هدف و منابع سازمانی که نظرسنجی می‌شود بستگی دارد. نرخ پاسخ را می‌توان با استفاده از موارد زیر افزایش داد:

- داشتن حامی مالی قانونی برای ارزیابی، حمایت حامیان مالی که در جامعه مورد ارزیابی اعتبار بالایی دارند.
- توضیح خوب و کوتاه برای تایید ارزیابی
- اعلام قبلی به افراد در مورد برگزاری ارزیابی

- کوتاه نگه داشتن سند ارزیابی و اطلاع‌رسانی به موضوعات تا این موضوع را بدانند.
- تضمین قابلیت اطمینان و گمنامی
- ارائه برنامه زمان‌بندی براساس راحتی موضوع
- در صورت لزوم برقراری تماس مجدد (معمولاً چهار بار). در ارزیابی‌های پستی، هر بار یک نسخه جدید از سند ارسال شود.
- سند با پرسش‌های غیرتهدیدآمیزی که باعث جلب توجه می‌شوند، شروع شود.
- بعداً نتایج نهایی ارائه شود.
- در ارزیابی‌های پستی، از پست سفارشی دو قبضه و پیشتاز استفاده شود. در ارزیابی‌ها با پست الکترونیک، از پست الکترونیکی اولویت‌دار استفاده شود. کارت پستال و یادآورهای پست الکترونیک نیز مفید است. در ارزیابی‌های پستی، برای دریافت تمامی پاسخ‌ها حدود شش هفته صبر کنید.
- ارائه پاداشی برای یادگاری (مثلاً ضمیمه کردن ۱ دلار در ارزیابی‌های پستی؛ ۵ دلار در مصاحبه‌های رو در رو؛ یا استفاده از خودکارها، جاکلیدی، تقویم، یا سایر یادگاری‌ها برای ایجاد این احساس در افراد که در قبال همکاری چیزی دریافت می‌کنند). مرتبط ساختن نتیجه ارزیابی با ورود به قرعه‌کشی مبلغی قابل توجه (مثلاً ۱۰۰۰ دلار) می‌تواند نرخ پاسخ را افزایش دهد.

تحلیل عدم پاسخ

محقق باید فرض کند که افرادی که پاسخ نمی‌دهند از نظر متغیرهای مورد نظر تفاوت قابل توجهی با افراد پاسخ دهنده دارند. برای ارزیابی تاثیر عدم پاسخ سه استراتژی کلی وجود دارد. این استراتژی‌ها مانع‌الجمع هستند.

مقایسه جمعیت) میانگین ارزیابی می‌تواند با میانگین یک جمعیت شناخته شده مقایسه شود، به خصوص عملی بودن متغیرهای مربوط به آمارگیری نفوس همانند جنسیت، توزیع سنی، درآمد، اشتغال، و سایر متغیرهایی که توسط سرشماری تحت پوشش قرار گرفته‌اند. در تحقیقات مربوط به آموزش، ممکن است داده‌های جمعیتی در مورد توزیع جنسی، سن،

کلاس، و عملکرد آزمون جمعیت در دسترس باشد. محقق به دنبال این است که متغیرهایی را شناسایی کند که میانگین نمونه آنها با میانگین جمعیت متفاوت بوده و تاثیر چنین تورشی بر متغیرهای وابسته مورد نظر را پیش‌بینی کند (ترجیحاً بر اساس مطالب قبلی).

نمونه‌گیری گسترده بعدی) برای مصاحبه با نمونه‌ای از افرادی که پاسخ نداده‌اند، لازم است تلاش گسترده‌ای با استفاده از نمونه‌گیری بعدی برای ارزیابی افرادی که پاسخ نداده‌اند انجام شود. اضافه کردن نمونه بعدی به نمونه باعث کاهش تورش نمونه از مقدار برآورد شده با این روش که هزینه‌بر و در نتیجه به ندرت انجام می‌شود، می‌گردد. اگر از بین افرادی که پاسخ نداده‌اند نمونه‌گیری شود، محقق ممکن است بخواهد از پرسش‌های ارزیابی برای تعیین این که چرا در نمونه اولیه عدم پاسخ وجود دارد استفاده کرده و در نتیجه ماهیت عدم پاسخ را تعیین کند. اگر نمونه بعدی از افراد پاسخ نداده به روشی متفاوت تهیه شود (مثلاً تلفن به جای ارزیابی اولیه مبتنی بر وب)، محقق ممکن است پرسش‌هایی را در مورد کیفیت اولیه مطرح کند (مثلاً، آیا افرادی که پاسخ ندادند به اینترنت دسترسی دارند؟ آدرس پست الکترونیکی دارند؟ آیا پست الکترونیک خودش را به طور منظم کنترل می‌کند؟ برای بارگذری صفحه وب ارزیابی مشکل دارد؟ و غیره)

برون‌یابی موج) محقق مجموعه پاسخ‌های اولیه و هر کدام از چهار پاسخ به تماس مجدد را کدگذاری کرده، میانگین متغیرهای کلیدی هر کدام از پنج مجموعه را محاسبه کرده و سپس کنترل می‌کند که آیا برای هر کدام از متغیرهای امکان برون‌یابی وجود دارد یا خیر. به عنوان مثال، هر کدام از مجموعه‌های متوالی ممکن است نشاندهنده این باشند که زنان کمتری پاسخ داده‌اند که نشان می‌دهد انتظار این است که مجموعه بدون پاسخ نسبت کمتری از زنان داشته باشد. این روش هزینه حاشیه‌ای کمی دارد و بنابراین به شدت و به خصوص همراه با مقایسه جمعیت توصیه شده است.

پاسخ‌های اسنادی

همه پاسخ‌دهندگان به تمامی موارد مطرح شده در یک پرسشنامه پاسخ نداده‌اند. آندرو گلمن، گری کینگ، و چوتنهایی لیو (۱۹۹۸) روشی برای تجزیه و تحلیل یک سری از ارزیابی‌های متقابل مستقل معرفی کرده‌اند که در آن برخی از سوالات در برخی از ارزیابی‌های پاسخ داده نشده‌اند و تعدادی از پاسخ دهندگان به برخی از پرسش‌های مطرح شده پاسخ نداده‌اند. این روش می‌تواند برای یک ارزیابی واحد نیز استفاده شود که در آن سوالات مختلفی مطرح شده یا این که در طبقات یا خوشه‌های مختلف از روش‌های مختلف نمونه‌گیری استفاده شده است. این روش شامل استفاده از استناد چند باره موارد از قلم افتاده و پرسش‌ها با اضافه کردن به روش‌های موجود استناد طراحی شده برای ارزیابی‌های واحد در یک مدل رگرسیون سلسله مراتبی است که امکان کوواریانس در سطوح فردی و ارزیابی را فراهم می‌کند.

اطلاعات حاصل از وزن‌های ارزیابی درتحلیل متغیرهایی که وزن‌ها بر اساس آنها بوده مورد استفاده قرار گرفته و سپس از وزن‌گذاری مجدد پاسخ‌های فردی (مشاهده شده و انتسابی) برای برآورد مقادیر جمعیت استفاده شده است. آنها همچنین بر اساس مقایسه داده‌های انتسابی و غیرانتسابی تشخیص‌هایی برای کنترل تطبیق مدل انتسابی ایجاد کرده‌اند. آنها با استفاده از مثالی که عامل انجام این پروژه بوده این موضوع را نشان دادند- بررسی نظرسنجی افکار عمومی قبل از انتخاب، که در آن تمامی سوالات مورد نظر در تمامی ارزیابی‌های پاسخ داده شده‌اند و بنابراین این امکان وجود دارد که هر کدام از ارزیابی‌های به طور مستقل مستند شود.

وزن‌گذاری

دلایل مختلفی برای وزن‌گذاری وجود دارد، در واقع، وزن‌گذاری می‌تواند یک فرایند چند مرحله‌ای باشد زیرا محقق به صورت ترتیبی احتمال انتخاب، عدم پاسخ، طبقه بندی، و اهداف دیگر را تنظیم می‌کند. برخی از محققان با وزن‌گذاری موضوعات مخالف هستند زیرا مستلزم تکرار مصنوعی حالت‌ها است که معمولاً تغییری در نتایج ایجاد نکرده یا این که تغییر اندکی در نتایج ایجاد می‌کنند، و ممکن است تایید نادرستی بر داده‌ها منتقل کرده و توضیح انحرافات نمونه را غیر ممکن سازند.

طرفداران وزن‌گذاری می‌گویند که وزن‌گذاری شکل منطقی تقریب برای اصلاح نمونه بر اساس انحراف‌های مشخص است و چنین اصلاحی بهتر از عدم اصلاح است.

وزن‌گذاری برای عدم پاسخ) برخی اوقات این امکان وجود دارد که پاسخ‌دهندگان و و آنهایی که پاسخ نداده‌اند را بر اساس جنسیت و ویژگی‌های دیگر مقایسه کنیم. به عنوان مثال، در نمونه تصادفی از یک شهر، داده‌های مربوط به سرشماری برآورد خوبی از نسبت واقعی مردان و زنان به دست می‌دهد. در ارزیابی پستی، نسبت واقعی مردان و زنان ممکن است با توجه به نام تعیین شود. اگر در چنین مواقعی فردی کشف کند که توزیع مشاهده شده مطابق با جمعیت واقعی نیست، ممکن است بخواهد برای اصلاح توزیع، پاسخ‌ها را وزن‌گذاری کند. به عنوان مثال، اگر تعداد زنان خیلی کمی بین پاسخ‌دهندگان وجود داشته باشد ممکن است فرد بخواهد وزن بیشتری به پاسخ‌دهندگان زن نسبت به مرد بدهد. به عنوان مثال، اگر نسبت واقعی بر اساس جنسیت ۵۰-۵۰ باشد، و اگر ۴۰ زن و ۶۰ مرد داشته باشیم در این صورت فرد می‌تواند پاسخ هر زن را ۱/۵ وزن‌گذاری کند.

این صورت ۶۰ مرد و ۶۰ زن داریم. البته، برای جلوگیری از افزایش مصنوعی اندازه نمونه از ۱۰۰ به ۱۲۰، لازم است که وزن‌گذاری بعدی به نحوی باشد که مقیاس به ۱۰۰ برگردد. این کار می‌تواند با وزن‌گذاری مردان و زنان با وزن ۵/۶ انجام شود. این منطق را می‌تواند به حالتی که افراد پاسخ نداده برای موضوعات در مجموعه پاسخ‌دهندگان قرار گرفته‌اند نیز تعمیم داده شود. اصلاح قابل اطمینان عدم پاسخ نیاز به نرخ پاسخ بالا (مثلاً >۷۰ درصد) دارد. بورکیو و کلارک (۱۹۹۲:۶۰) می‌گویند: « این که استفاده از وزن‌ها منجر به تغییر قابل توجه برآوردهای میانگین نمونه نشود به تجربه ما بستگی دارد مگر این که افراد پاسخ نداده به اندازه کافی از افراد پاسخ داده تفکیک شده و نسبت قابل توجهی از افراد پاسخ نداده وجود داشته باشد.»

وزن‌گذاری برای اصلاح بعد از طبقه بندی همین منطق و همین استراتژی وزن‌گذاری را می‌توان برای یک طبقه مشخص (مثلاً مردان سیاه پوستی که بین ۱۸-۲۵ سال سن دارند) که به دلیل عدم پاسخ یا به دلیل نمونه‌گیری طبقه بندی نامتناسب کمتر از حد نمایش داده شده استفاده کرد.

در هر صورت هدف این است که حالت‌های موجود را به نحوی وزن‌گذاری کنیم که باعث

افزایش نمایش در نمونه اصلاح شده لایه‌ای که در داده‌های خام کمتر از واقع نشان داده شده شود.

وزن‌گذاری برای در نظر گرفتن احتمالی انتخاب) در نمونه تصادفی از افراد، هر فرد باید شانس برابر برای انتخاب شدن داشته باشد. البته، واقعیت روند نمونه‌گیری باید به گونه‌ای باشد که هر خانوار، نه فرد، شانس برابر داشته باشد. افراد درون خانوارهای با جمعیت بیشتر شانس کمتری برای انتخاب شدن دارند: آنها کمتر از حد نشان داده شده‌اند باید وزن بیشتری داشته باشند. در صورتی که محقق فکر کند که برای داشتن یک مورد باید به دنبال تعداد افراد واجد شرایط در خانوار باشد در این صورت می‌توان وزن‌گذاری را اصلاح کند. با جمع کردن پاسخ‌ها به این مورد و تقسیم آن بر تعداد خانوارهای ارزیابی شده، با فرض یک ارزیابی برای هر خانوار، میانگین تعداد افراد هر خانوار، مثلاً ۲/۵ ، به دست می‌آید. در این صورت وزن هر کدام از افراد درون نمونه برابر با تعداد افراد درون خانوار تقسیم بر این متوسط است. به عنوان مثال، اگر یک خانوار مشخص دارای ۵ فرد واجد شرایط باشد در این صورت وزن این حالت برابر با ۲ = ۵/۲,۵ است.

برخورد با داده‌های از قلم افتاده

عدم پاسخ به موارد ذکر شده در ارزیابی فردی این پرسش را مطرح می‌کند که آیا مواردی که دارای داده‌های از قلم افتاده هستند را حذف یا این که پاسخ‌ها را درون‌یابی کنیم. برای کمک به رفع این مشکل نرم‌افزار تخصصی وجود دارد. یک نمونه از این نرم‌افزارها مدول «تحلیل مقدار از قلم افتاده» است که به SPSS پایه اضافه شده است.

برآورد اندازه نمونه قبل از ارزیابی

در اغلب موارد لازم است که اندازه نمونه در مرحله طراحی تحقیق تعیین شود. در صورتی که یک یا چند حالت زیر وجود داشته باشد لازم است که اندازه نمونه بزرگتر باشد:

- روابط ضعیف‌تری شناسایی شده باشد
- از سطح معناداری بالاتری استفاده شده باشد

- از متغیرهای کنترلی بیشتری استفاده شود
- تعداد حالت‌های کمتری در کوچکترین گروه از هر متغیر وجود داشته باشد
- واریانس متغیرها بیشتر باشد

ترکیب این عوامل باعث ایجاد پیچیدگی‌هایی میشود که همراه با عدم شناخت از جمعیتی که قرار است نمونه‌گیری شود معمولاً منجر به تعیین اندازه نمونه به صورت بی‌طرفانه می‌شود - برآورد اختیاری. توجه داشته باشید که اندازه نمونه به طور کلی به اندازه جمعیتی که قرار است نمونه‌گیری شود بستگی ندارد. حتی در پیچیده‌ترین تحلیل‌ها، به ندرت به نمونه‌های بالاتر از ۱۵۰۰ نیاز است. نرم‌افزار تخصصی، همانند مدول "SamplePower" که می‌تواند به SPSS پایه اضافه شود، وجود دارد که به محقق در محاسبه اندازه مورد نیاز نمونه کمک می‌کند.

تست معناداری فقط برای نمونه‌های تصادفی مناسب است

فرض شده که نمونه‌گیری تصادفی برای آمارهای استنتاجی است (تست معناداری). «استنتاجی» اشاره به این واقعیت دارد که نتایج راجع به روابط درون داده‌ها بر اساس نتیجه حاصل از شناخت توزیع نمونه‌گیری استخراج شده است. تست‌های معناداری مبتنی بر تئوری نمونه‌گیری است که بر اساس آن هر حالت از قبل شانس مشخصی، معمولاً برابر، برای انتخاب شدن دارد. نتیجه آماری، معناداری برآورد را با استفاده از نمونه‌های تصادفی ارزیابی می‌کند. در مورد شماره‌گذاری و سرشماری، به چنین نتیجه‌ای نیاز نیست زیرا برآوردها دقیق هستند. خطای نمونه‌گیری مرتبط نبوده و بنابراین آمارهای استنتاجی مربوط به خطای نمونه‌گیری نیز موضوع نامرتبطی هستند. برخی اوقات به صورت اختیاری از تست‌های معناداری برای نمونه‌های غیر تصادفی استفاده می‌شود ولی هیچ روشی برای ارزیابی اعتبار این برآوردها وجود ندارد، هر چند تحلیل عدم پاسخ ممکن است تا حدودی مرتبط باشد.

چگونه فرد می‌تواند اندازه نمونه را از قبل تعیین کند؟

محاسبه اندازه نمونه به تعدادی عوامل پیچیده بستگی دارد. در عمل، محققان از نرم‌افزار تخصصی طراحی شده برای این محاسبات استفاده می‌کنند. در تست‌های معمولی

منحنی از قاعده سرانگشتی مبتنی بر خطاهای استاندارد استفاده شده است: اندازه نمونه $n = (s * z/T)^2$، که در آن s خطای استاندارد متغیر با بالاترین واریانس (شاید در نمونه قبل از تست تعیین شده باشد)، z تعداد واحدهای استاندارد مربوط به بخش مورد نظر حالت‌ها (z=۱/۹۶ برای تست‌های دو طرفه با سطح معناداری ۰/۰۵) و T تغییرات قابل قبول در نمونه است.

تاثیرات خوشه‌بندی و چند سطحی در نمونه‌های چند سطحی

نمونه‌گیری خوشه‌ای یک مرحله‌ای مترادف نمونه‌گیری طبقه‌بندی است. از این نمونه‌گیری زمانی استفاده می‌شود که محقق نمونه‌های تصادفی ساده‌ای از افراد را از مجموعه‌ای از واحدهای مورد نظر خاص (مثلاً از گروهی از بلوک‌های سرشماری خاص) استخراج می‌کند. نمونه‌گیری خوشه‌ای چند مرحله‌ای مترادف نمونه‌گیری طبقه بندی چند مرحله‌ای است. وقتی طبقات به صورت واحدهای جغرافیایی باشد، این روش را برخی اوقات نمونه‌گیری منطقه‌ای می‌نامند. توجه داشته باشید که برخی از محققان از عبارت «نمونه‌گیری خوشه‌ای» برای اشاره به یک فرایند دو مرحله‌ای استفاده می‌کنند که در آن از خوشه‌ها نمونه‌گیری شده و سپس از بین خوشه‌های انتخاب شده، افراد انتخاب می‌شوند، و از واژه «نمونه‌گیری خوشه‌ای چند مرحله‌ای» برای فرایندهای سه مرحله‌ای و بالاتر استفاده می‌کنند.

خوشه‌بندی، که بخش ذاتی نمونه‌گیری چند مرحله‌ای است، اغلب منجر به مشاهدات همبسته می‌شود که نوعی تخطی از فرض موارد نمونه‌گیری شده مستقل است- فرضی که در بسیاری از روش‌های آماری مطرح است. هر چند معمول این است که وقتی داده‌ها به صورت تصادفی نمونه‌گیری شده‌اند برای این داده‌ها از نمونه‌گیری چند مرحله‌ای استفاده شود، این کار فقط بایستی زمانی انجام شود که همبستگی درون لایه‌ای (inter cluster correlation ICC-) معنادار نباشد، یعنی این که متغیر گروه‌بندی یا خوشه‌بندی مستقل از متغیر وابسته باشد. اگر ICC معنادار باشند، از اشکالی از مدل‌سازی چند سطحی (که مدل‌سازی ترکیبی خطی یا مدل‌سازی خطی سلسله مراتبی نیز نامیده می‌شوند) استفاده می‌شود.

خطاهای نمونه‌گیری

انتخاب نادرست عناصر نمونه می‌تواند منجر به خطاهای بیشتری در برآورد پارامترهای مورد نظر در جمعیت می‌شود. البته می‌توانیم خطاهای متفاوت‌تری را نیز پیدا کنیم: مصاحبه کننده می‌تواند بی‌طرف نباشد یعنی این که می‌تواند برخی از پاسخ‌های را تقویت کند. همچنین ممکن است فردی که قرار است با او مصاحبه شود مایل به پاسخ دادن به برخی از پرسش‌ها نباشد (یا این که نتواند پاسخ دهد). ما این خطاهای احتمالی را به صورت زیردسته‌بندی می‌کنیم:

۱. خطای انتخاب: اگر هر کدام از عناصر جمعیت احتمال بیشتری برای انتخاب شدن نسبت به سایر عناصر داشته باشد. فرض کنید می‌خواهیم میزان رضایت مشتریان یک ورزشگاه را اندازه‌گیری کنیم و به همین دلیل بین ساعت ۱۰ تا ۱۲ صبح با برخی از مشتریان مصاحبه می‌کنیم. این بدان معنی است که افرادی که بعد از ظهر به ورزشگاه می‌روند نشان داده نخواهند شد و بنابراین نمونه نمی‌تواند نشاندهنده تمامی مشتریان باشد. یک راه برای اجتناب از این نوع خطاها انتخاب نمونه به نحوی که تمامی مشتریان احتمال برابر برای انتخاب شدن داشته باشند، است.

۲. خطای عدم پاسخ: همچنین ممکن است برخی از عناصر جمعیت نخواهند یا نتوانند به برخی از سوالات پاسخ دهند. این وضعیت می‌تواند زمانی رخ بدهد که ما پرسشنامه‌ای داشته باشیم که برخی از اعضای جمعیت از روی صداقت به آنها پاسخ ندهند. دوری از این خطاها بسیار سخت است ولی در مواردی که می‌خواهیم درستی پاسخ‌ها را کنترل کنیم، می‌توانیم پرسش‌هایی (پرسش‌های فیلتر) را مطرح کنیم.

۳. خطاهای نمونه‌گیری: در نتیجه محاسبه برآورد (میانگین برآوردی، مجموع، نسبت و غیره) بر اساس یک نمونه و نه کل جمعیت رخ می‌دهند. خطاهای نمونه‌گیری ناشی از این واقعیت هستند که وضعیت تعیین شده بر اساس یک نمونه نمی‌تواند به طور دقیق برابر با مقدار واقعی جمعیت باشد.

وقتی نتایج حاصل از یک ارزیابی نمونه‌ای گزارش شده باشد، غالباً این نتایج به شکل «به علاوه یا منفی» واحدهای مرتبط استفاده شده بیان شده‌اند. این «به علاوه یا منفی» نشاندهنده خطاهای نمونه‌گیری هستند. آمارهای مبتنی بر نمونه‌های استخراج شده از یک جمعیت همیشه با هم (و از مقدار واقعی جمعیت) متفاوت هستند و دلیل آن هم شانس است. این تغییر خطای نمونه‌گیری است و معیاری که برای تعیین خطای نمونه‌گیری استفاده شده، خطای استاندارد است.

از خطاهای استاندارد معمولاً برای تعیین دقت برآوردها استفاده شده است. تئوری توزیع نمونه، می‌گوید که حدود ۶۸ درصد برآوردهای مبتنی بر یک خطای استاندارد یا انحراف از میانگین استاندارد، ۹۵ درصد مبتنی بر دو انحراف استاندارد و ۹۹ درصد تمامی برآوردها مبتنی بر سه انحراف استاندارد هستند. خطاهای نمونه‌گیری را می‌توان با انتخاب مناسب نمونه‌ها به حداقل رساند، و در در رابطه با طرح نمونه‌ها سه عامل بر خطاهای نمونه‌گیری تاثیرگذار هستند - روش نمونه‌گیری، تغییرات درون نمونه نسبت به تغییرات مورد نظر، و اندازه نمونه. نمونه بزرگتر منجر به خطای کمتر نمونه‌گیری می‌شود.

خطاهای غیر نمونه‌گیری

دقت برآورد نیز تحت تاثیر خطاهای ناشی از عواملی همانند پوشش ناقص و دستورالعمل‌های اشتباه برای برآورد قرار داشته و همراه با خطاهای مشاهداتی باعث خطاهایی می‌شود که خطاهای غیرنمونه‌گیری نامیده می‌شود. هدف از ارزیابی معمولاً این است که اطلاعاتی در مورد ارزش واقعی جمعیت به دست آید. هدف این است که تا آنجا که ممکن است منابع در دسترس برای ارزیابی را به دست آوریم. تفاوت بین مقدار ارزیابی و مقدار واقعی خطای مشاهداتی یا خطای پاسخ نامیده می‌شود. خطاهای پاسخی غیر نمونه‌گیری در نتیجه ثبت نادرست تغییرات مورد نظر، گزارشگری غیر دقیق داده‌ها، یا اصلاح عمدی داده‌ها توسط جمع‌آوری کننده‌ها و ثبت‌کننده‌های داده‌ها مطابق با نظر خودشان بروز می‌کنند. خطای غیر پاسخی زمانی رخ می‌دهد که تعداد قابل توجهی از افراد در نمونه ارزیابی وجود ندارند، یا به پرسشنامه پاسخ نمی‌دهند، یا به نحوی با افرادی که برای مطالعه اهمیت دارند متفاوت هستند.

یکسونگری

هر چند نمونه‌گیری قضاوتی سریعتر از نمونه‌گیری احتمالی است ولی در معرض خطاهای سیستماتیک قرار دارد. به عنوان مثال، اگر برای بررسی میانگین تعداد صفحات یک کتاب از بین ۲۰۰ کتاب ۲۰ کتاب انتخاب شوند، ارزیاب می‌تواند پیشنهاد کند که این کتاب‌ها را از بین کتاب‌هایی برداشته شوند که اندازه متوسط دارند. دشواری چنین روشی آن است که عمداً یا سهواً، نمونه‌بردار تمایل به ایجاد خطاهای قضاوتی در راستای انتخاب کتاب‌هایی که بزرگتر از متوسط مورد نظر هستند می‌باشد. چنین خطاهای سیستماتیکی منجر به یکسونگری می‌شوند.