

## مطالعه‌ای بر نمونه‌گیری‌های احتمالی و غیراحتمالی

حسن گلی ناری<sup>۱</sup>، محمد خراشادی‌زاده<sup>۲\*</sup>، غلامرضا محتشمی برزادران<sup>۳</sup> و فرهاد مهران<sup>۴</sup>

<sup>۱</sup> دانشجوی دکتری آمار، دانشگاه بیرجند

<sup>۲</sup> دانشیار گروه آمار، دانشگاه بیرجند

<sup>۳</sup> استاد بازنشسته گروه آمار، دانشگاه فردوسی مشهد

<sup>۴</sup> دکتری آمار دانشگاه هاروارد و عضو بازنشسته موسسه بین‌المللی نیروی کار

تاریخ دریافت: ۱۴۰۴/۰۲/۱۱

تاریخ پذیرش: ۱۴۰۴/۰۶/۲۷

### چکیده:

تعیین شیوه نمونه‌گیری در روش تحقیق از اهمیت بسیار زیادی برخوردار است. یک تقسیم‌بندی کلی از انواع روش‌های نمونه‌گیری به صورت احتمالی و غیراحتمالی است. روش‌های نمونه‌گیری احتمالی عمدتاً شناخته شده می‌باشند و مطالعات زیادی نیز در این زمینه تاکنون انجام شده است. در این روش‌ها به کمک استنباط‌هایی نتایج به دست آمده از نمونه به کل جامعه تعمیم‌پذیر است، اما در زمینه روش‌های نمونه‌گیری غیراحتمالی علی‌رغم اهمیت کاربردهای آن و همچنین نیاز روزافزون در زمینه تحقیقات روز دنیا و از طرفی پیشینه تاریخی که دارند، مطالعات زیادی صورت نگرفته است، تا آنجا که روش‌های موجود در نمونه‌گیری غیراحتمالی در بین محققان آن‌طور که باید شناخته شده نیست. از این رو در این مقاله مفهوم و فلسفه نمونه‌گیری بیان می‌شود، تاریخچه‌ای از تحولات و استفاده محققان از روش‌های نمونه‌گیری مرور می‌شود و دلایل گرایش به نمونه‌گیری غیرتصادفی تشریح می‌شود. در ادامه دسته‌بندی روش‌های نمونه‌گیری احتمالی و غیراحتمالی ارائه و چگونگی انجام آن‌ها به اختصار توضیح داده می‌شود. در پایان مقاله نیز برخی از ایده‌های نوین که به بهبود کیفیت و افزایش دقت نمونه‌های غیراحتمالی کمک می‌کنند، معرفی می‌شوند.

**واژه‌های کلیدی:** نمونه‌گیری احتمالی، نمونه‌گیری غیراحتمالی، کاهش سوگیری، نمونه‌گیری گلوله‌برفی، هسته، موج، کالیبراسیون، مه‌داده، جوسازی امتیاز تمایل

### ۱ مقدمه

صورت نگرفته است. روش‌های نمونه‌گیری احتمالی ایده‌آل باید برای اطمینان از نماینده بودن نمونه و همچنین برای تعمیم نتایج به جامعه هدف استفاده شود. در صورت عدم استفاده از آن‌ها، باید در تفسیر نتایج حاصل از مطالعه احتیاط کرد. اما محدودیت‌هایی در استفاده از نمونه‌گیری احتمالی وجود دارد از جمله صرف وقت و هزینه زیاد که محققان را به سمت استفاده از روش‌های نمونه‌گیری غیراحتمالی سوق می‌دهند.

برای به دست آوردن نتایج معتبر در هر مطالعه پژوهشی، انتخاب یک روش نمونه‌گیری صحیح و علمی مهم است. لذا یکی از مباحث بسیار مهم در آمار، تعیین روش نمونه‌گیری است که به دو صورت کلی نمونه‌گیری احتمالی (که مبتنی بر روش‌های احتمالی است) و نمونه‌گیری غیراحتمالی (انتخاب نمونه بر پایه روش‌های غیراحتمالی) انجام می‌شود. اگرچه در نمونه‌گیری احتمالی مطالعات بسیار زیادی انجام شده است و به کمک استنباط‌هایی نتایج به دست آمده از نمونه به کل جامعه تعمیم‌پذیر است، اما در موضوع نمونه‌گیری غیراحتمالی علی‌رغم پیشینه تاریخی و اهمیت کاربردهای آن و از طرفی نیاز روزافزون جامعه امروزی، طبق بررسی‌های انجام شده تاکنون مطالعات زیادی

### ۱.۱ انگیزه مطالعه و تحقیق

در اوایل قرن بیستم، تقابل نظری بین «انتخاب تصادفی» و «انتخاب هدفمند» در نمونه‌گیری مطرح بود. مقاله نیمن (۱۹۳۴) با معرفی

بیستم نیمن و پیرسون اصول نمونه‌گیری احتمالی را ضمن اصلاح بیشتر رسمیت بخشیدند و مفاهیمی مانند آزمون فرضیه و بازه‌های اطمینان را معرفی کردند [۱۷]. دهه ۱۹۳۰ مفهوم نمونه‌گیری غیراحتمالی که عمدتاً نمونه‌گیری مبتنی بر قضاوت یا قصدی و نمونه‌گیری سهمیه‌ای بودند به‌ویژه در تحقیقات اجتماعی مورد توجه قرار گرفت [۲۸]. با این حال نگرانی‌هایی در مورد اریبی و تعمیم‌پذیری نتایج حاصل از این روش‌ها وجود داشت. الموریر و جورج گالوپ برای پیش‌بینی نتایج انتخابات آمریکا در دهه ۱۹۳۰ از این روش استفاده کردند [۲۸] و برای به‌دست آوردن نمونه‌ای که نماینده همه رأی‌دهندگان آمریکایی باشد، ابتدا کشور را به تعدادی ناحیه جغرافیایی تقسیم کردند و از هر ناحیه متناسب با تعدادی که حق رأی داشتند تعدادی را با نظر و قضاوت خود برای نظرخواهی انتخاب می‌کردند [۱۱]. در سال ۱۹۳۴ نیمن مقاله‌ای را ارائه داد که روش‌های انتخاب تصادفی و هدفمند را با یکدیگر مقایسه می‌کرد [۱۳] و [۱۸]. نیمن در مقاله خود مطالبی در روش‌شناسی و بحث مفصلی از روش‌هایی برای استنباط از نمونه‌های احتمالی از جامعه‌های محدود از جمله ارائه تعریفی از فاصله اطمینان در این زمینه، ارائه داد [۱۹]. مقاله نیمن نقطه عطفی برای آمارگیری نمونه‌ای بود که منجر به پذیرش گسترده نمونه‌گیری تصادفی، به‌ویژه توسط دفاتر آمار ملی شد [۲۵]. استفاده فراگیر از روش‌های نمونه‌گیری توسط افراد مختلفی در ایالات متحده تا دهه ۱۹۴۰ اتفاق افتاد. بدین صورت که اداره سرشماری آمریکا و مرکز امور کشاورزی آمریکا تکنیک‌های ابداع‌شده برای نمونه‌گیری‌های احتمالی را به کار بردند [۲۶]. در سال ۱۹۴۰ اداره سرشماری آمریکا ضمن سرشماری کلی، نمونه‌ای به نسبت ۵ درصد از جامعه آمریکا انتخاب کرد و نتایج حاصل از برآوردهای این نمونه‌گیری را با نتایج حاصل از سرشماری تطبیق داد و این یک نوآوری مهم در کارهای پژوهشی بود [۲۸] و [۲۷]. نمونه‌گیری احتمالی اصطلاحی است که دمینگ در سال ۱۹۵۰ به کار برد [۲۸]. اواسط قرن بیستم جامعه‌شناسان و محققان پیشگامی مانند پل لازارسفلد شروع به استفاده از روش‌های نمونه‌گیری غیراحتمالی مانند نمونه‌گیری هدفمند و نمونه‌گیری گلوله برفی، در محیط‌های تحقیقات اجتماعی کردند و بر مزایای عملی در زمینه‌های خاص تأکید داشتند. پیدایش حوزه نمونه‌گیری نظرسنجی توسط آندرس کی‌پر (۱۸۹۷) آماردان نروژی و اولین مدیر آمار نروژ، انجام شد [۱۵]. در اواخر قرن بیستم بحث بین نمونه‌گیری احتمالی و غیراحتمالی شدت گرفت [۱۳]. در حالی که نمونه‌گیری احتمالی به اصطلاح نوعی استاندارد برای تعمیم‌پذیری نتایج به جامعه بود، روش‌های غیراحتمالی کاربردهای ویژه‌ای پیدا کردند، به‌ویژه در تحقیقات کیفی، تحقیقات بازار، و مناطقی که اجرای

نمونه‌گیری احتمالی و استنباط طرح‌مبنا، تأثیر تعیین‌کننده‌ای داشت و این روش به استاندارد اصلی دفاتر آمار ملی تبدیل شد [۱۳]. علیرغم این امر، نمونه‌گیری غیراحتمالی همواره در حوزه‌های کاربردی مختلف مورد استفاده گسترده قرار گرفته است. در سال‌های اخیر، شاهد بازگشت به سمت نمونه‌گیری‌های غیراحتمالی بوده‌ایم. این تغییر جهت متأثر از عواملی چند است: محدودیت‌های فزاینده نمونه‌گیری تصادفی از نظر هزینه و کارایی، ظهور فناوری‌های دیجیتال برای جمع‌آوری داده‌های حجیم با هزینه ناچیز و سرعت بالا، و توسعه روش‌های پیشرفته برای اصلاح و تعدیل نمونه‌های غیراحتمالی. این تحول روش‌شناختی، خود نتیجه پیشرفت‌های نظری در تئوری نمونه‌گیری، رشد توان محاسباتی، توسعه چارچوب‌های نمونه‌گیری نوین و پاسخ به چالش‌های عملی فزاینده در کاربردهای نظرسنجی‌های اجتماعی بوده است [۴]. با پذیرش این تغییر رویکرد، باید تأکید کرد که بسیاری از اصول روش‌های نمونه‌گیری کلاسیک که در گذشته بنیان نهاده شده‌اند، کماکان از اعتبار علمی برخوردارند. در واقع، ریشه بسیاری از مباحث معاصر در حوزه نمونه‌گیری غیراحتمالی و داده‌های حجیم را می‌توان در مبانی اولیه تحقیقات نظرسنجی جستجو کرد.

## ۲ تاریخچه تحولات و نقاط عطف روش‌های نمونه‌گیری

در این بخش ما سعی کرده‌ایم به برخی از نقاط عطف و نیز افراد تأثیرگذار در زمینه مباحث نمونه‌گیری مرور خلاصه‌ای داشته باشیم، اگرچه مسلماً افراد دیگری نیز بوده‌اند که در این موضوع تحقیقاتی داشته‌اند [۱۳]. اولین کاربردهای تحقیقات نمونه‌گیری برای مطالعه شرایط اجتماعی جوامع مختلف در اواخر دهه ۱۸۰۰ در انگلیس آغاز شد. آمارگیری‌های نمونه‌ای بزرگ مقیاس چارلز بوث در انگلیس از شرایط اجتماعی جامعه لندن در سال ۱۸۸۶ آغاز شد [۲]. یک دهه بعد سیبوم راونتری شرایط فقر طبقه کارگری در یورک در سال ۱۸۹۶ را مورد بررسی قرارداد [۲۳]. در اواخر قرن نوزدهم فرانسیس گالتون و کارل پیرسون شروع به توسعه نظریه‌های آماری کردند که این مطالعات زمینه را برای روش‌های نمونه‌گیری احتمالی مدرن فراهم کرد [۲۰]. شرایط طبقه کارگری در شهر ریدینگ در سال ۱۹۱۲ توسط باولی انجام شد [۳]. در اوایل قرن بیستم رونالد فیشر مفهوم تصادفی سازی را در نمونه‌گیری معرفی و بر اهمیت آن در استنباط آماری تأکید کرد. مشارکت فیشر در پایه‌گذاری اصول نمونه‌گیری احتمالی بسیار تأثیرگذار بود [۱۰]. اواسط قرن

## ۲.۳ دلایل روی آوردن به نمونه‌گیری غیراحتمالی و برخی از چالش‌های آن

علیرغم اینکه در اغلب مواقع نمونه‌گیری احتمالی تنها راه برای اطمینان از استنتاج بی‌طرفانه در مورد کمیت‌های جامعه است که در آن سرشماری کامل امکان‌پذیر نیست، با این حال، با ورود به عصر «مه‌داده‌ها»، نمونه‌های غیراحتمالی که مکانیسم‌های نمونه‌گیری آن‌ها ناشناخته است، در حال تغییر و تحول هستند. در بسیاری از مطالعات زیست‌محیطی به‌عنوان مثال تنوع و تعداد موجودات زنده یک دریا هیچ سرشماری کاملی وجود ندارد [۴]. از نمونه‌های غیراحتمالی به این دلیل استفاده می‌شود که نمونه‌گیری احتمالی اغلب روشی وقت‌گیر و پرهزینه است و علاوه بر آن ممکن است در بسیاری از وضعیت‌ها امکان‌پذیر نباشد [۲۹]. لذا بیشترین توجه در دفاع از نمونه‌های غیراحتمالی عواملی مانند به‌موقع بودن، سهولت دسترسی و کاهش هزینه طرح‌های آمارگیری است، که برای حصول اهداف بسیار مناسب است. از سویی دیگر چالش‌های آماری عمده‌ای هنگام استفاده از نمونه‌های غیراحتمالی به وجود می‌آید. همچنین مسائل برای نمونه‌های غیراحتمالی پیچیده‌تر است. با اینکه نمونه‌های غیراحتمالی اغلب معرف کل جامعه نیستند، اما روش‌هایی برای کاهش اربیبی در نمونه‌های غیراحتمالی ابداع شده‌اند که آن‌ها را معرفی خواهیم کرد. به دلایلی که ذکر شد نمونه‌گیری غیراحتمالی به‌خصوص در تحقیقات بازار و همچنین نظرسنجی‌ها بسیار فراوان استفاده می‌شوند، همچنین در تحقیقات دانشگاهی و کاربردی و همچنین مؤسسات آماری ملی اهمیت روزافزونی دارد، لذا نیاز به تحقیق و مطالعه عمیق‌تر بر روی این موضوع را بیشتر نمایان می‌کند. ما استدلال می‌کنیم که نمونه‌های غیراحتمالی می‌توانند مفید باشند و در موارد متعددی مورد استفاده قرار بگیرند، سریع و راحت در دسترس قرار بگیرند، راحت تحلیل شوند و ما را به نتایجی مطلوب برسانند، البته مشروط بر اینکه محدودیت‌های آن‌ها ارزیابی شده، در صورت امکان کاهش داده شود و به‌وضوح بیان شود [۴]. تهیه نمونه‌های احتمالی، حتی نمونه‌های نسبتاً کوچک، اغلب بسیار دشوار است. در مقابل، نمونه‌های غیراحتمالی، حتی نمونه‌های بزرگ، نسبتاً آسان هستند. از این رو، اخیراً محققان اغلب استفاده از نمونه‌های غیراحتمالی را بر اساس کمیت داده‌ها توجیه می‌کنند. با وجود مزایای استنباط طرح مینا، چالش‌هایی مانند عدم پوشش و عدم پاسخگویی اغلب منجر به اجرای ناقص طرح‌های نمونه‌گیری می‌شوند. این نواقص، در کنار افزایش هزینه‌ها و محدودیت‌های زمانی، مشکلاتی برای نمونه‌گیری احتمالی

روش‌های احتمالی غیرعملی بودند [۱۳]. از اواخر قرن بیستم تاکنون با پیشرفت گسترده در تکنیک‌های محاسباتی، مه‌داده‌ها و تحقیقات ترکیبی، مرزهای بین نمونه‌گیری احتمالی و غیراحتمالی باریک‌تر شده است. برای مثال محققانی مانند دونالد روبین با ارائه روش‌هایی مانند جورسازی امتیاز تمایل، در پر کردن برخی شکاف‌ها بین رویکردهای احتمالی و غیراحتمالی مشارکت داشته‌اند [۱۳]. در ادامه مروری بر روش‌های نمونه‌گیری احتمالی داشته و سپس برخی از روش‌های اخیر را بر اساس مدل‌هایی که برای نمونه‌های غیراحتمالی استفاده می‌شوند، معرفی می‌کنیم، که می‌تواند به استنتاج‌های مفید از یک نمونه غیراحتمالی به‌تنهایی یا در ترکیب با یک نمونه احتمالی انجام شود.

## ۳ مفهوم نمونه و فلسفه نمونه‌گیری

در هر پژوهشی، بهترین راهبرد، بررسی مسأله در کل جامعه است، اما عملاً همیشه نمی‌توان کل جامعه را مطالعه کرد. با کمک نمونه‌گیری می‌توان هزینه‌های انجام شده، زمان صرف شده برای انجام تحقیق و همچنین نیروی انسانی مورد نیاز برای انجام مطالعه را کاهش داد. نماینده بودن نمونه به سه عامل بستگی دارد: (۱) روش نمونه‌گیری (۲) حجم نمونه و (۳) سطح پاسخ. چارچوب نمونه‌گیری در نمونه‌گیری احتمالی بسیار مهم است، زیرا اگر چارچوب نمونه‌گیری به‌طور مناسب از جامعه ترسیم نشود. نمونه‌گیری تصادفی از آن چارچوب نمی‌تواند مشکل تحقیق را حل کند.

## ۱.۳ برخی چالش‌های به وجود آمده

در طول سال‌ها، اجرای نمونه‌گیری احتمالی در نظرسنجی‌های اجتماعی به‌طور فزاینده‌ای در بسیاری کشورها به دلیل کاهش مستمر تمایل مردم برای شرکت در نظرسنجی‌ها با چالش مواجه شده است. علی‌رغم تلاش‌های بیشتر برای تشویق مشارکت، نرخ پاسخگویی در سال‌های اخیر به‌طور چشمگیری کاهش یافته است. در واکنش، با پیشرفت‌های عمده در تکنیک‌های به‌کاررفته، تلاش‌های بیشتری برای جبران بی‌پاسخی صورت گرفته است. در بسیاری از کشورها، نمونه‌گیری احتمالی با استنباط طرح مینا، دیگر وضعیت خود را به‌عنوان استاندارد بلامنازع حفظ نمی‌کند. برای مثال، کاهش نرخ پاسخگویی در برخی نظرسنجی‌ها به سطح ۴۵ درصد و حتی پایین‌تر تا حد زیادی قالب و شکل نمونه‌گیری را رد می‌کند [۶].

$$R = \text{متغیر نشانگر عضویت (۱ عضو نمونه باشد، ۰ در غیر اینصورت)}$$

$$f = \text{نرخ نمونه‌گیری (} n/N \text{)}$$

$$\sigma_Y = \text{انحراف معیار جامعه}$$

$$n_{\text{eff}} = \text{اندازه مؤثر نمونه}$$

$$\mathbb{E}[\cdot] = \text{امید ریاضی}$$

$$\text{SRS} = \text{نمونه‌گیری تصادفی ساده}$$

$$\text{MSE} = \text{میانگین مربعات خطا}$$

در این فرمول،  $\rho(R, Y)$  نشان‌دهنده همبستگی بین متغیر  $Y$  جامعه و یک متغیر نشانگر  $R$  است که اگر عضو جامعه در نمونه باشد مقدار ۱ و در غیر این صورت ۰ می‌گیرد.  $f$  نرخ نمونه‌گیری  $(n/N)$  است و  $\sigma_Y$  انحراف معیار جامعه برای متغیر  $Y$  است.

$\rho(R, Y)$  جهت و بزرگی سوگیری انتخاب را مشخص می‌کند:

- وقتی  $\rho(R, Y) > 0$  باشد، مقادیر بزرگتر  $Y$  با احتمال بیشتری در نمونه حضور دارند (منجر به برآورد بیش‌ازحد می‌شود)
- وقتی  $\rho(R, Y) < 0$  باشد، مقادیر کوچکتر  $Y$  با احتمال بیشتری در نمونه حضور دارند (منجر به برآورد کمتر از حد می‌شود)
- اگر  $\rho(R, Y) = 0$  باشد، این جزء سایر عبارات را خنثی می‌کند و خطایی وجود ندارد.

معادله دوم، اندازه مؤثر نمونه  $(n_{\text{eff}})$  را ارائه می‌دهد که به‌عنوان حجم نمونه‌ای از نمونه‌گیری تصادفی ساده (SRS) تعریف می‌شود که با استفاده از روش میانگین‌گیری مشابه، همان میانگین مربعات خطا (MSE) را برای برآورد میانگین جامعه تولید می‌کند.  $n_{\text{eff}}$  به صورت زیر است:

$$n_{\text{eff}} = \frac{f}{(1-f)} \frac{1}{\mathbb{E}[\rho(R, Y)^2]} = \frac{n}{1-f} \frac{1}{\mathbb{E}[\rho(R, Y)^2] N}$$

$\mathbb{E}[\rho(R, Y)^2]$  امید ریاضی مجذور ضریب همبستگی  $\rho(R, Y)$  برای یک مکانیزم انتخاب داده مشخص است. این پارامتر به‌صورت امید ریاضی بیان می‌شود زیرا  $\rho(R, Y)$  می‌تواند مقادیر محقق شده متعددی برای یک مکانیزم نمونه‌گیری ثابت داشته باشد. با این حال، در جوامع با  $N$  بزرگ، واریانس  $\rho(R, Y)$  معمولاً ناچیز است، بنابراین می‌توان  $\mathbb{E}[\rho(R, Y)^2]$  را با  $\rho(R, Y)$  جایگزین کرد [۱۶]. پیامد مهم این فرمول آن است که اگر  $\rho(R, Y)^2$  حتی اندکی از صفر انحراف داشته باشد،

ایجاد می‌کند که باعث ظهور روش‌های نمونه‌گیری شبه تصادفی مانند نمونه‌گیری سهمیه‌ای<sup>۱</sup> در برخی از کاربردها می‌شود.

اگرچه جذابیت اصلی نمونه‌گیری احتمالی در عدم وابستگی استنباط طرح مبنا به مفروضات مدل است، باید اذعان کرد که نمونه‌گیری احتمالی در عمل به‌ندرت کاملاً اجرا می‌شود. دو نقص اساسی در این نمونه‌گیری وجود دارد که عبارت‌اند از عدم پوشش و عدم پاسخگویی.

• عدم پوشش، به این دلیل به وجود می‌آید که چارچوب نمونه‌گیری شامل برخی از عناصر جامعه هدف نمی‌شود.

• بی‌پاسخی زمانی اتفاق می‌افتد که یک عنصر نمونه‌گیری شده نتواند پاسخ‌های قابل قبولی برای برخی یا همه سؤالات نظرسنجی ارائه کند

نمونه‌گیری احتمالی دو اشکال اصلی دارد که باید در برابر جذابیت‌های نظری استنباط طرح مبنا در نظر گرفته شود: هزینه و زمان. هزینه‌های اضافی نمونه‌گیری احتمالی شامل هزینه‌های ردیابی نمونه‌ها (از جمله تماس‌های تکراری زمانی که فرد نمونه در ابتدا در دسترس نیست) می‌شود. هنگامی که از نمونه‌گیری منطقه‌ای استفاده می‌شود، هزینه‌های اضافی شامل هزینه‌های فهرست‌برداری واحدها در مناطق نمونه‌گیری نیز می‌شود. به دلایل مشابه، جمع‌آوری داده‌های نظرسنجی از یک نمونه احتمالی بیشتر طول می‌کشد و باعث می‌شود برآوردهای نظرسنجی زمان‌بر شود. به‌موقع بودن برای همه نظرسنجی‌ها مهم است، به‌ویژه برای نظرسنجی‌هایی که نتایج آن به‌شدت وابسته به زمان هستند، مانند نظرسنجی‌های سیاسی یا بررسی شیوع برخی عفونت‌ها و بررسی مناطقی که دچار یک حادثه‌ای شده‌اند. کمیت لزوماً به معنای کیفیت نیست. متأسفانه، کمیت داده‌ها جایگزینی برای کیفیت نمایندگی<sup>۲</sup> نیست [۴]. مینگ [۱۶] فرمولی برای اندازه مؤثر<sup>۳</sup> نمونه ارائه کرد. به این صورت که سوگیری انتخاب<sup>۴</sup> را به دقت میانگین نمونه به‌عنوان برآوردگری برای میانگین جامعه مرتبط می‌سازد. از آنجا که ارائه مشتق کامل فرمول در اینجا ممکن نیست، به ذکر دو معادله مرتبط بسنده می‌کنیم. معادله اول نشان می‌دهد که تفاوت بین میانگین نمونه و میانگین جامعه به صورت زیر است:

$$(\bar{Y}_n) - (\bar{Y}_N) = \rho(R, Y) \sqrt{\frac{1-f}{f}} \sigma_Y$$

$\rho(R, Y)$  = ضریب همبستگی جامعه

$Y$  = متغیر مورد مطالعه

<sup>2</sup>Representativeness

<sup>3</sup>Effective Size

<sup>4</sup>Selection Bias

تحلیلگر این اطمینان را دارد که می‌تواند یافته‌های پیمایش را صرفاً مشمول درجه قابل اندازه‌گیری خطای نمونه‌گیری<sup>۹</sup> بداند، در حالی که در نمونه‌گیری غیراحتمالی همواره این چالش وجود دارد که نمونه هدفمند با توجه به متغیرهای مورد تحلیل، نماینده جامعه نیست. در نمونه‌گیری احتمالی، هر عضو جامعه یک شانس برای انتخاب شدن در مطالعه دارد. از زوایای مختلف استفاده از هر روش نمونه‌گیری مزایا و محدودیت‌های خاص خود را دارد، با این حال، نمونه‌گیری احتمالی ترجیح داده می‌شود، زیرا نتایج آن قابل تعمیم است اما بدلیل محدودیت‌های موجود همیشه نمی‌توان از روش احتمالی استفاده نمود.

## ۱.۴ نمونه‌گیری احتمالی

نمونه‌گیری احتمالی، انتخاب نمونه بر پایه احتمال است. آچارها و همکاران [۱] انواع روش‌های مرسوم در نمونه‌گیری احتمالی را به صورت زیر بیان می‌کنند:

۱. نمونه‌گیری تصادفی ساده<sup>۱۰</sup>

۲. نمونه‌گیری تصادفی سیستماتیک<sup>۱۱</sup>

۳. نمونه‌گیری تصادفی طبقه‌ای<sup>۱۲</sup>

۴. نمونه‌گیری خوشه‌ای<sup>۱۳</sup>

۵. نمونه‌گیری چندفازی<sup>۱۴</sup>

۶. نمونه‌گیری چندمرحله‌ای<sup>۱۵</sup>

کلیه روش‌های نمونه‌گیری بالا از یک شیوه تصادفی برای انتخاب نمونه استفاده می‌کنند.

### ۱.۱.۴ نمونه‌گیری تصادفی ساده

در این روش هر عضو جامعه شانس مساوی برای انتخاب در نمونه از جامعه را دارد. داده‌ها با استفاده از جدول اعداد تصادفی یا لیست اعداد تصادفی تولید شده توسط رایانه انتخاب می‌شوند. این کار به روش

نسبت اندازه مؤثر به اندازه واقعی نمونه ( $n_{eff}/n$ ) با افزایش  $N$  کاهش می‌یابد.

فرمول محاسبه حجم نمونه مؤثر نتایج شگفت‌آوری به همراه دارد. منگ [۱۶] مجموعه داده‌ای با حجم  $n = ۲۳۰۰۰۰۰$  درباره تمایلات رأی‌دهندگان برای انتخابات ریاست‌جمهوری ۲۰۱۶ آمریکا را تحلیل کرد. او دریافت که همبستگی بین عضویت در نمونه و تمایل پاسخ‌دهندگان به رأی دادن به ترامپ بسیار اندک و برابر با  $-۰.۰۰۵$  بود. این همبستگی ظاهراً ناچیز باعث شد تا حجم نمونه مؤثر به حدود ۴۰۰ کاهش یابد (کاهش معادل ۹۸/۹۹ درصد). برادلی و همکارانش از همین فرمول برای تحلیل چندین نظرسنجی درباره مصرف واکسن بیماری کرونا ۲۰۱۹ (COVID-19) در آمریکا استفاده کردند. آن‌ها نشان دادند که بزرگ‌ترین نظرسنجی (اما غیرتصادفی)، با  $۲۵۰۰۰۰$  پاسخ در هفته، می‌تواند برآوردهایی با همان میزان خطا به دست دهد که یک نمونه تصادفی کمتر از ۱۰ نفر تولید می‌کند. لذا کیفیت داده‌ها بسیار مهم‌تر از کمیت داده‌هاست [۴].

## ۴ دسته‌بندی و انواع روش‌ها در نمونه‌گیری

به طور کلی مرسوم‌ترین دسته بندی‌ها در انواع روش‌های رایج در نمونه‌گیری عبارتند از:

- نمونه‌گیری احتمالی<sup>۵</sup> (مبتنی بر روش‌های احتمالی است)
- نمونه‌گیری غیراحتمالی<sup>۶</sup> (انتخاب نمونه بر پایه روش‌های غیراحتمالی)

پس از انتشار مقاله نیمین در سال ۱۹۳۴ نمونه‌های احتمالی و استنباط مبتنی بر طرح<sup>۷</sup> به عنوان استاندارد طلایی [۱۳] برای پیمایش‌های گسترده‌ای که توسط دفاتر آماری ملی انجام می‌شد، تثبیت شد و با اجرای بی‌نقص یک نمونه احتمالی و بدون خطای پاسخ<sup>۸</sup>

<sup>5</sup>Random Sampling

<sup>6</sup>Nonrandom Sampling

<sup>7</sup>Design-based Inference

<sup>8</sup>Rresponse Error

<sup>9</sup>Measurable Sampling Error

<sup>10</sup>Simple Random Sampling - SRS

<sup>11</sup>Systematic

<sup>12</sup>Stratified

<sup>13</sup>Cluster

<sup>14</sup>Multiphase

<sup>15</sup>Multistage

<sup>16</sup>Frame

**نقطه ضعف:** در این روش آن است که از نظر فنی فقط انتخاب نمونه اول یک انتخاب احتمالی است زیرا برای انتخاب‌های بعدی نمونه‌هایی وجود خواهند داشت که شانس انتخاب آنها صفر است.

**مثال ۲۰۴.** اگر فرض کنیم حجم جامعه  $N =$  و حجم نمونه  $n =$ ، بنابراین خواهیم داشت  $k = N/n =$ ، یعنی فاصله نمونه‌گیری برابر ۴ می‌باشد. حال یک عدد تصادفی بین ۱ تا ۴ انتخاب می‌کنیم. فرض کنید «۳» باشد، بنابراین شرکت‌کننده شماره «۳» اولین نمونه ما خواهد بود. سپس با اضافه کردن «۴» به این عدد ادامه می‌دهیم. نمونه‌های بعدی ما ۷، ۱۱، ۱۵، ۱۹، ۲۳، ۲۷، ۳۱ و ... خواهند بود، تا زمانی که ۵۰ نمونه مورد نیاز را تکمیل کنیم.

### ۳.۱.۴ نمونه‌گیری تصادفی طبقه‌ای

در این نوع نمونه‌گیری داده‌ها به زیرگروه‌های مختلف (طبقه یا قشر) با ویژگی‌های مشترک مانند سن، جنس، نژاد، درآمد تقسیم می‌شوند. از هر قشر یک نمونه تصادفی گرفته می‌شود. از مزایای این روش این است که نمایندگی همه گروه‌ها در نمونه مورد نیاز را تضمین می‌کند. مشخصه‌های هر کدام از طبقه‌ها را می‌توان تخمین زد و مقایسه کرد. همچنین تفاوت حاصل از نمونه‌گیری سیستماتیک را کاهش می‌دهد. از جمله محدودیت‌های این روش این است که به اطلاعات دقیق در مورد نسبت‌های هر طبقه نیاز دارد. همچنین تهیه فهرست‌های طبقه بندی شده پرهزینه است.

**مثال ۳۰۴.** در بررسی شیوع دیابت در یک جمعیت بزرگسال، می‌توان جمعیت را بر اساس جنسیت طبقه بندی کرد و سپس تعداد افراد مساوی از زن و مرد را داشت. این امر باعث بررسی شیوع دیابت از نظر جنسیت می‌شود. نمونه را می‌توان بر اساس محل سکونت مانند شهری، روستایی یا حومه شهری طبقه بندی کرد که به ما شیوع دیابت را در منطقه با بازنمایی یکسان از هر گروه نشان می‌دهد.

### ۴.۱.۴ نمونه‌گیری خوشه‌ای

نمونه‌گیری تصادفی خوشه‌ای یک شیوه دو مرحله‌ای است که در آن کل جمعیت به خوشه‌ها یا گروه‌هایی، معمولاً مناطق یا مناطق جغرافیایی مانند روستاها، مدارس، بخش‌ها، بلوک‌ها و غیره تقسیم می‌شوند. این روش در تحقیقات اپیدمیولوژیک بیشتر از تحقیقات بالینی استفاده می‌شود. استفاده از آن در نظرسنجی‌های بزرگ ملی بسیار کاربردی است. خوشه‌ها به صورت تصادفی انتخاب می‌شوند و معمولاً به حجم نمونه بزرگتری نیاز است. نمونه‌گیری خوشه‌ای زمانی بسیار مفید است

قرعه کشی و ... نیز قابل انجام است. در این روش یک چارچوب<sup>۱۶</sup> نمونه‌گیری مورد نیاز است. همه افراد جامعه مورد مطالعه باید به ترتیب صعودی یا نزولی شمارش شوند. از مزایای این روش این است که حداقل دانش جامعه مورد نیاز است، با این حال محدودیت‌هایی نیز داریم. به عنوان مثال یک چارچوب نمونه مورد نیاز است که هزینه آن زیاد است. نمونه‌گیری تصادفی ساده نسبت به نمونه‌گیری طبقه بندی شده با اندازه مشابه دارای خطاهای بزرگتر و دقت کمتری هستند.

**مثال ۱۰۴.** فرض کنید ۲۰۰ خانوار در یک آبادی وجود دارد و ما می‌خواهیم ۱۵ خانوار را با نمونه‌گیری تصادفی ساده انتخاب کنیم. لیست تمام ۲۰۰ خانوار آبادی چارچوب نمونه‌گیری را تشکیل می‌دهد. حال ۱۵ خانوار نمونه می‌توانند با استفاده از جدول اعداد تصادفی یا به روش قرعه کشی انتخاب شوند. هنگامی که یک خانوار انتخاب شد، آن عدد خاص از جدول اعداد تصادفی حذف می‌شود. این روش به نمونه‌گیری تصادفی بدون جایگزینی معروف است. با ادامه روش به این ترتیب ۱۵ خانوار نمونه انتخاب می‌شوند.

### ۲.۱.۴ نمونه‌گیری تصادفی سیستماتیک

نمونه‌گیری سیستماتیک یکی از روش‌های رایج در تحقیقات آماری است. در نمونه‌گیری سیستماتیک انتخاب اولین نمونه به صورت تصادفی انجام می‌شود و سپس نمونه‌های بعدی با فرآیندی دوره‌ای به صورت منظم و با فواصل مشخص از یک لیست یا چارچوب نمونه‌گیری انتخاب می‌شوند [۵]. این روش دارای مزایا و نقاط ضعفی است که به شرح زیر می‌باشد:

**سود متوسط:** در نمونه‌گیری سیستماتیک، با انتخاب واحدها به صورت منظم، احتمال بروز خطاهای تصادفی کاهش می‌یابد. این امر باعث می‌شود که میانگین‌های محاسبه‌شده از نمونه به میانگین‌های جامعه نزدیک‌تر باشند.

**هزینه متوسط:** این روش معمولاً هزینه کمتری نسبت به سایر روش‌های نمونه‌گیری، مانند نمونه‌گیری تصادفی ساده، دارد. زیرا نیاز به انتخاب تصادفی هر واحد به صورت جداگانه نیست و می‌توان به سادگی از یک الگو پیروی کرد.

**اعتبار داخلی و خارجی بالا:** اعتبار داخلی به معنای این است که آیا نتایج به دست آمده از نمونه واقعاً به جامعه هدف تعمیم‌پذیر هستند یا خیر. اعتبار خارجی نیز به تعمیم‌پذیری نتایج به سایر جمعیت‌ها اشاره دارد. با توجه به اینکه نمونه‌گیری سیستماتیک معمولاً نماینده خوبی از جامعه است، اعتبار داخلی و خارجی آن بالا است.

مثال ۶.۴. در یک نظرسنجی ملی، تعداد تصادفی از مناطق در تمام استان‌ها و به دنبال آن تعداد تصادفی از روستاها انتخاب می‌شوند. در مرحله سوم خانه‌ها انتخاب خواهند شد. تمامی خانه‌هایی که واحدهای نهایی نمونه‌گیری هستند بررسی می‌شوند.

این روش به اندازه نمونه‌گیری تصادفی واقعی قوی نیست، اما احتمالاً به رفع محدودیت‌های ذاتی نمونه‌گیری تصادفی کمک می‌کند. این روش بسیار مفید است زیرا شامل مراحل متعدد تصادفی‌سازی می‌شود. نمونه‌گیری چند مرحله‌ای اغلب زمانی استفاده می‌شود که فهرست کاملی از تمام اعضای جامعه وجود نداشته باشد. در نتیجه هزینه‌ها در مقایسه با نمونه‌گیری خوشه‌ای سنتی کاهش می‌یابد. لازم به ذکر است که در نمونه‌گیری چند مرحله‌ای، واحدهای نمونه‌گیری در مراحل مختلف متفاوت است. از سوی دیگر، در نمونه‌گیری چند فازی، از یک واحد نمونه برداری چندین بار نمونه برداری می‌شود.

## ۲.۴ نمونه‌گیری غیراحتمالی

نمونه‌گیری غیراحتمالی یکی از انواع روش‌های نمونه‌گیری می‌باشد که احتمال انتخاب یک نمونه در آنها ناشناخته است. در نمونه‌گیری غیراحتمالی افراد براساس معیارهای غیراحتمالی انتخاب می‌شوند و هر فرد شانس یکسان و معین انتخاب شدن در نمونه را ندارد. نمونه‌ها بر اساس در دسترس بودن و یا معیارهای دیگری است که باعث می‌شود داده‌های اولیه جمع‌آوری گردد. یکی از کاربردهای روش غیراحتمالی بدین صورت است که گاهی مطالعه را در ابتدا برای برآورد تقریبی وضعیت موجود و همچنین رفع نواقص اجرایی تحقیق با نمونه‌های غیراحتمالی انجام می‌دهند که اصطلاحاً به این نوع مطالعه پیش‌آزمون<sup>۱۷</sup> می‌گویند و در گام بعدی پس از بررسی نتایج پیش‌آزمون و رفع نواقص موجود تحقیق را با نمونه‌گیری احتمالی تکرار می‌کنند تا بتوان نتایج آن را به جامعه آماری تعمیم داد. از طرفی این نوع از نمونه‌گیری از نظر هزینه و دسترسی آسان تر است. اما در معرض خطری بالاتری است. لذا از این روش نمونه‌گیری نمی‌توان استنباط آماری کاملاً معتبری را درباره کل جامعه داشته باشیم. استفاده از نمونه‌گیری غیراحتمالی چه در مرحله طراحی و چه در مرحله اجرا ساده تر از نمونه‌گیری احتمالی است و به زمان و سرمایه کمتری نیاز دارد. روش‌های نمونه‌گیری غیراحتمالی اغلب برای تحقیقات کیفی و اکتشافی (نظیر مصاحبه، پرسشنامه رضایت شغلی و...) مفید می‌باشند. در این نوع از تحقیقات هدف تست فرضیات درباره جامعه بزرگتر نیست. اما برای توسعه یک فهم و

که جامعه بسیار پراکنده باشد و انتخاب نماینده از همه عناصر غیرعملی باشد.

مثال ۴.۴. نمونه‌ای را می‌توان از دانش‌آموزان سال اول دبیرستان گرفت که دانش آنها را در مورد نوعی ویروس اندازه‌گیری می‌کند. فرض کنید همه دبیرستان‌های تهران خوشه هستند. ما ۲۰ دبیرستان را با نمونه‌گیری تصادفی ساده یا تصادفی سیستماتیک انتخاب می‌کنیم و سپس هر دبیرستان به یک خوشه تبدیل می‌شود. سپس می‌توانیم با همه دانش‌آموزان مصاحبه کنیم یا به‌طور تصادفی دانش‌آموزان را در هر خوشه انتخابی برای سنجش دانش آنها در مورد ویروس انتخاب کنیم.

## ۵.۱.۴ نمونه‌گیری چندفازی

نمونه‌گیری چندفازی روشی احتمالی است که در آن یک نمونه ثابت در فازهای متوالی بررسی می‌شود. در هر فاز، داده‌های عمیق‌تری پرهزینه‌تر فقط از زیرمجموعه‌ای از نمونه فاز قبل جمع‌آوری می‌شود. هدف اصلی، کاهش هزینه‌ها با حفظ دقت آماری است [۵].

مثال ۵.۴. در مطالعه سلامت گندمزارها در یک ناحیه فازهای زیر را دنبال می‌کنیم:

۱. فاز اول (غربالگری ارزان): بررسی شهودی ۱۰۰ مزرعه گندم برای شناسایی مزارع با علائم آفت (زردی برگ‌ها)، نتیجه ۲۰ مزرعه مشکوک.
۲. فاز دوم (تست تخصصی): آزمایش خاک فقط برای ۲۰ مزرعه مشکوک (داده‌های پرهزینه). در نتیجه رسیدن به ۵ مزرعه آلوده به قارچ.
۳. فاز سوم (تأیید نهایی): تحلیل DNA قارچ فقط در ۵ مزرعه آلوده.

## ۶.۱.۴ نمونه‌گیری چند مرحله‌ای

شکل ترکیبی از نمونه‌گیری خوشه‌ای است که در آن دو یا چند سطح از واحدها یکی در دیگری تعبیه شده‌اند. این روش شامل تکرار دو مرحله اساسی یعنی فهرست بندی و نمونه‌گیری است. به طور معمول، در هر مرحله، اندازه خوشه کوچکتر می‌شود و در پایان نمونه‌گیری از موضوع انجام می‌شود. گاهی برای مراحل مختلف نمونه‌گیری از اصطلاحات خاصی استفاده می‌شود. نمونه‌گیری مرحله اول به‌عنوان واحد نمونه‌گیری اولیه (PSU)، مرحله دوم واحد نمونه‌گیری ثانویه (SSU)، مرحله سوم به‌عنوان واحد نمونه‌گیری سوم (TSU) شناخته می‌شود و به همین ترتیب تا زمانی که به واحدهای نمونه "نهایی" برسیم.

<sup>17</sup>Pretest

مثال ۷.۴. مراجعه بیماران به بخش سرپایی یک بیمارستان و بررسی معیارهای شمول، دانش آموزان یک مدرسه، اعضای یک سازمان اجتماعی و غیره.

#### ۲.۲.۴ نمونه‌گیری سهمیه‌ای

روشی از نمونه‌گیری که تضمین می‌کند که مشخصه خاصی از یک نمونه جمعیت به میزان دقیقی که محقق می‌خواهد اتفاق بیفتد. در نمونه‌گیری سهمیه‌ای، افراد نمونه به صورت غیر احتمالی با توجه به برخی سهمیه بندی‌های تعیین شده انتخاب می‌شوند. دو نوع از نمونه‌گیری سهمیه‌ای وجود دارد که عبارتند از متناسب و نامتناسب. در نمونه‌گیری سهمیه‌ای متناسب هدف محقق ارائه مشخصات و ویژگیهای اصلی جامعه با استفاده از یک نسبتی از جامعه صورت می‌پذیرد. اما نمونه‌گیری سهمیه‌ای نامتناسب تا حدودی محدودیت کمتری دارد. در این روش محقق حداقل تعداد واحدهای نمونه را که در هر گروه می‌خواهد، مشخص می‌کند. در این روش دیگر تناسب مشخصات و ویژگی‌هایی که در بالا ذکر شد وجود ندارد.

مثال ۸.۴. در یک نمونه ۱۰۰ نفری، محقق مایل است ۴۰ درصد مرد و ۶۰ درصد زن در نمونه باشد. زمانی که ۴۰ مرد انتخاب شوند، یعنی «سهمیه» برای مردان تمام شده است و نمونه‌گیری از جمعیت مردان متوقف می‌شود.

از جمله مزایای این روش آن است که هزینه متوسط است، همچنین در سطح گسترده استفاده می‌شود، قابل درک بوده و نیازی به فهرست عناصر جامعه نیست. همچنین برخی از عناصر طبقه بندی را بیان می‌کند. محدودیت‌ها مشابه نمونه‌گیری (راحت) آسان است. نمونه‌گیری طبقه‌ای و نمونه‌گیری سهمیه‌ای از این نظر مشابه هستند که در هر دو، جامعه به دسته‌ها/قشر تقسیم می‌شود و افراد از هر دسته انتخاب می‌شوند. هدف انتخاب یک نمونه نماینده و یا امکان تحلیل زیرگروه است.

ممکن است سوال شود که روش نمونه‌گیری سهمیه‌ای چه تفاوتی

درک اولیه از جمعیت مورد تحقیق و یا جمعیت کوچک مناسب است. انواع روش‌های مرسوم در نمونه‌گیری غیراحتمالی به صورت زیر هستند:

- نمونه‌گیری راحت (در دسترس)<sup>۱۸</sup>
- نمونه‌گیری سهمیه‌ای<sup>۱۹</sup>
- نمونه‌گیری گلوله‌برفی<sup>۲۰</sup>
- نمونه‌گیری متوالی<sup>۲۱</sup>
- نمونه‌گیری قضاوتی<sup>۲۲</sup>
- نمونه‌گیری داوطلبانه<sup>۲۳</sup>
- نمونه‌گیری هدفمند<sup>۲۴</sup>
- نمونه‌گیری مُدی<sup>۲۵</sup>
- نمونه‌گیری اتفاقی<sup>۲۶</sup>
- نمونه‌گیری حداکثر تنوع<sup>۲۷</sup>
- نمونه‌گیری چندمرحله‌ای

#### ۱.۲.۴ نمونه‌گیری راحت (در دسترس)

این رایج‌ترین روش نمونه‌گیری است. همانطور که از نامش پیداست محقق به دنبال راحت‌ترین و در دسترس‌ترین افراد در جامعه هدف است. این روش از نظر هزینه‌ای نیز روش بسیار ارزانی است. نمونه بر اساس راحتی محقق انتخاب می‌شود. در واقع اغلب پاسخ دهندگان به این دلیل انتخاب می‌شوند که در زمان مناسب در مکان مناسب هستند. نمونه‌گیری آسان و معمولاً در تحقیقات بالینی استفاده می‌شود که در آن بیماران که معیارهای شمول را دارند در مطالعه انتخاب می‌شوند. از مزایای این روش این است که بیشتر مورد استفاده قرار می‌گیرند، ارزان تر هستند و نیازی به لیستی از تمام عناصر جامعه نیست. با این حال، بدون محدودیت نیز نمی‌باشد که مهمترین آنها تنوع و اریبی است که نمی‌توان این معیارها را اندازه‌گیری یا کنترل کرد.

<sup>18</sup>Convenience

<sup>19</sup>Quota

<sup>20</sup>Snowball

<sup>21</sup>Sequential

<sup>22</sup>Judgmental (Purposive)

<sup>23</sup>Volunteer

<sup>24</sup>Purposive

<sup>25</sup>Modal

<sup>26</sup>Accidental

<sup>27</sup>Maximum Variation

موج اول : انتخاب ۱۰ نفر اول (به عنوان هسته اولیه).  
 موج دوم : هر فرد از موج اول، ۳ نفر جدید را معرفی می‌کند (۳۰ نفر جدید).  
 موج سوم : هر فرد از موج دوم، ۲ نفر جدید را معرفی می‌کند (۶۰ نفر جدید).  
 بنابراین در نهایت تعداد نمونه‌ها به ۱۰۰ نفر خواهند رسید .

**تعریف ۱۲.۴.** به شیوه گسترش نمونه از طریق معرفی افراد جدید توسط اعضای موجود (ارتباط بین معرفی‌کنندگان و معرفی‌شوندگان) زنجیره ارجاع<sup>۳۰</sup> گویند و طبیعی است هر چه زنجیره طولانی‌تر باشد، خطر آریبی افزایش می‌یابد.

**تعریف ۱۳.۴.** به نمونه‌گیری در چند مرحله متوالی که در هر مرحله تصمیم به ادامه یا توقف گرفته می‌شود نمونه‌گیری موج‌دار<sup>۳۱</sup> می‌گویند. کاربرد آن معمولاً در مطالعات بالینی یا زمانی که جامعه نادر است می‌باشد. برای مثال موج اول شامل نمونه‌گیری از ۱۰۰ بیمار و موج دوم در صورتی که نتایج قطعی نبود، ۵۰ بیمار دیگر اضافه می‌شوند.

**مثال ۱۴.۴.** در مطالعه بر روی نمونه‌ای که درگیر رفتارهای پرخطر یا سوءمصرف مواد هستند، فردی که رفتار پرخطری دارد ممکن است از افراد دیگری که در رفتارهای پرخطر مشابه دارند نام ببرد و این تا زمانی ادامه می‌یابد که تعداد کافی از پاسخ‌دهندگان تکمیل شود. مزایای آن می‌تواند به هزینه پایین آن در شرایط خاص و برای مکان‌های جمعیت‌های نادر اشاره کرد. معایب آن نیز شامل سوگیری است زیرا واحدهای نمونه‌گیری مستقل نیستند.

#### ۴.۲.۴ نمونه‌گیری متوالی

حالتی از نمونه‌گیری در دسترس است. در این روش محقق سعی می‌کند تا آنجا که از لحاظ هزینه و نیروی انسانی در توان اوست از میان نمونه‌های در دسترس به انتخاب بپردازد و به حجم نمونه محاسبه شده توجهی ندارد.

#### ۵.۲.۴ نمونه‌گیری قضاوتی

در این روش، محقق براساس نظر و پیشینه‌ای که در مورد اعضای جامعه آماری دارد، دست به نمونه‌گیری می‌زند. انتخاب یا عدم انتخاب عضوی از جامعه در نمونه بسته به نظر محقق و تجربیات او دارد. معمولاً این

با نمونه‌گیری طبقه‌بندی شده دارد؟ پاسخ این است که نمونه‌گیری طبقه‌ای بدون فراخوان ممکن است در عمل تفاوت زیادی با نمونه‌گیری سهمیه‌ای نداشته باشد. با این حال، تفاوت‌های خاصی بین نمونه‌گیری طبقه‌ای (قشربندی) و سهمیه‌ای وجود دارد. یک چارچوب نمونه برای نمونه‌گیری طبقه‌ای لازم است اما برای نمونه‌گیری سهمیه‌ای لازم نیست. مهم‌تر از آن، نمونه‌گیری طبقه‌ای از نمونه‌گیری احتمالی استفاده می‌کند، بنابراین امکان تخمین خطای نمونه‌گیری را فراهم می‌کند که با نمونه‌های سهمیه‌ای امکان‌پذیر نیست. در نمونه‌گیری طبقه‌بندی شده پس از ایجاد طبقه بندی‌ها انتخاب نمونه‌ها معمولاً با نمونه‌گیری تصادفی ساده انجام می‌شود، اما در نمونه‌گیری سهمیه‌ای انتخاب نمونه‌ها در هر گروه به صورت غیر احتمالی و در دسترس است. بدین صورت که در نمونه‌گیری سهمیه‌ای، مصاحبه‌گر اولین نمونه موجود را انتخاب می‌کند که معیارهای شمول را داشته باشد. از این رو، یک نمونه‌گیری آسان (در دسترس) است.

#### ۳.۲.۴ نمونه‌گیری گلوله برفی

اگر دسترسی به افراد جامعه مشکل باشد از نمونه‌گیری گلوله برفی برای یافتن نمونه از طریق سایر نمونه‌ها استفاده می‌شود. تعداد افرادی که در نمونه‌گیری گلوله برفی قابلیت دسترسی به آنها را پیدا می‌کنید به مراتب بیشتر از تماس مستقیم با مخاطبان است. این نمونه‌گیری در مواقعی که با تمام افراد یک حوزه آشنایی نداشته باشید، بسیار مفید است. در این روش نمونه‌گیری، نمونه‌های اولیه با روش‌های احتمالی یا غیراحتمالی انتخاب می‌شوند و سپس، نمونه‌های بعدی با اطلاعات ارائه شده توسط نمونه‌های اولیه به دست می‌آیند.

**تعریف ۹.۴.** به اعضای اولیه‌ای که نمونه‌گیری از آنها آغاز می‌شود هسته<sup>۲۸</sup> می‌گویند.

در نمونه‌گیری گلوله برفی، هسته معمولاً افراد شناخته‌شده یا در دسترس هستند. بعنوان مثال در مطالعه رفتارهای پرخطر، هسته ممکن است شامل ۵ بیمار مراجعه‌کننده به کلینیک باشد.

**تعریف ۱۰.۴.** به مرحله‌ای از جمع‌آوری داده‌ها که معمولاً در روش‌های طولی یا نمونه‌گیری‌های تکراری استفاده می‌شود موج<sup>۲۹</sup> می‌گویند.

**مثال ۱۱.۴.** به عنوان مثال در اجرای نمونه‌گیری گلوله برفی ممکن است سه موج داشته باشیم:

<sup>28</sup>Seed(Core)

<sup>29</sup>Wave

<sup>30</sup>Referral Chain

<sup>31</sup>Sequential Sampling

#### ۹.۲.۴ نمونه‌گیری اتفاقی

هنگامی که محقق آیت‌های یک مطالعه را به صورت اتفاقی انتخاب می‌کند در تلاش برای شبیه سازی احتمالی است. اگر چه نتایج ممکن است به هیچ وجه احتمالی نباشد و اطلاعات جمع‌آوری شده توسط محقق، به صورت آگاهانه و یا غیر آگاهانه، دچار پراش انتخاب شود.

#### ۱۰.۲.۴ نمونه‌گیری با حداکثر تنوع

استفاده از تنوع نمونه در یک پژوهشی همانند قسمتهای مختلف شهر، کشور، فرهنگ‌ها، کلاسهای اجتماعی و ... تا حد زیادی مورد علاقه محققین است. ابعاد این روش بسیار متنوع است. در واقع این روش نمونه‌گیری نیازمند افرادی است که دامنه تغییرات عنوان پژوهش مد نظر را به خوبی نشان دهد.

#### ۱۱.۲.۴ نمونه‌گیری چند مرحله‌ای

روش نمونه‌گیری چند مرحله‌ای یکی از انواع روش‌های نمونه‌گیری است که در آن از چندین روش نمونه‌گیری توأم استفاده می‌گردد. این روش‌ها می‌توانند همگی روش‌های نمونه‌گیری احتمالی، روش‌های نمونه‌گیری غیراحتمالی و یا ترکیبی از دو روش نمونه‌گیری احتمالی و غیر احتمالی باشند. این نمونه‌گیری در شرایطی که با جوامع بزرگ سرو کار داریم روش مناسبی است و شامل چندین مرحله برای انتخاب نمونه است.

## ۵ چند نکته

اکثر نمونه‌گیری‌هایی که در سطح جمعیت‌های بزرگ انجام می‌شود به صورت چند مرحله‌ای است. رعایت نسبت حجم نمونه به گروه‌ها و طبقات باید همانند نسبت آن در جامعه باشد. نمونه‌های غیراحتمالی لزوماً معرف کل جامعه نیست، ولی باید سعی شود با طراحی مناسب و تا حد امکان آنها را تا حدودی معرف جامعه و دقیقتر انتخاب کرد. یکی از این روش‌ها وارد کردن روش‌های نمونه‌گیری احتمالی در نمونه‌گیری چند مرحله‌ای است. یعنی در سلسله مراتب نمونه‌گیری چند مرحله‌ای، یک یا چند مرحله احتمالی وارد می‌شود، اگر چه بیشتر مراحل غیراحتمالی باشند.

روش در جوامع آماری محدود و با حجم کم به کار می‌رود زیرا محقق باید در مورد تک تک اعضاء اطلاعات قبلی داشته باشد تا بتواند نمونه حاصل را بهتر انتخاب کند.

#### ۶.۲.۴ نمونه‌گیری داوطلبانه

همانند نمونه‌گیری در دسترس یک نمونه‌گیری داوطلبانه بر اساس آسان بودن دسترسی به نمونه‌ها بنا شده است. در واقع به جای اینکه محقق شرکت کنندگان را انتخاب نماید و به صورت مستقیم با آنها در ارتباط باشد افراد به صورت داوطلبانه خودشان وارد مطالعه می‌شوند. در واقع در نمونه‌گیری داوطلبانه نمونه‌های پاسخ دهنده داوطلب همیشه تا حدودی آریبی دارند و انتخاب آنها مغرضانه می‌باشند. دلیل این امر این است که برخی از افراد ذاتاً بیشتر از دیگران داوطلب می‌شوند.

#### ۷.۲.۴ نمونه‌گیری هدفمند

نمونه‌گیری هدفمند شامل طیف بسیار وسیعی از روش‌ها است که در تحقیقات کیفی بسیار مورد استفاده قرار می‌گیرند. در واقع یکی از انواع روش‌های نمونه‌گیری غیراحتمالی است که از تنوع زیادی برخوردار است که هر کدام وابسته به شرایطی که محقق در آن قرار می‌گیرد تعیین می‌گردد. انواع روش‌های نمونه‌گیری هدفمند عبارتند از نمونه‌گیری موارد شایع، ناهمگون، همگون، موارد اضافه یا انتهایی، شدت، موارد بحرانی، تایید کننده و رد کننده، فرصت طلب، سهمیه‌ای، قضاوتی، گلوله برفی و فردماهر. این نوع از نمونه‌گیری از قضاوت خود محقق استفاده می‌کند. نمونه‌گیری هدفمند یعنی خود محقق برای نمونه انتخابی نظر می‌دهد که چه مواردی انتخاب شوند. محقق برای ورود واحدهای نمونه‌گیری به نمونه‌اش، ویژگیهای خاص از آنها را در نظر گرفته و به پیاده سازی ایده‌هایش می‌پردازد. این نوع نمونه‌گیری اغلب در مطالعات کیفی استفاده می‌گردد. یک نمونه هدفمند باید معیارهای مشخص و منطقی برای شمولیت داشته باشد.

#### ۸.۲.۴ نمونه‌گیری مڈی

در این روش نمونه‌گیری افراد و یا مواردی که شایع‌تر و رایج‌تر (بدین معنا که تعداد موارد بیشتری از آنها در جامعه هدف وجود دارد) هستند جمع‌آوری می‌گردند. در واقع اسم این روش نمونه‌گیری از اسم واژه آماری مڈ<sup>۳۲</sup> برگرفته شده است.

## ۱.۵ هدف از جمع آوری داده‌ها

دقت در تعیین هدف نهایی از نمونه‌گیری نکته دیگری است که باید به آن توجه نمود. به منظور دستیابی به اهداف نهایی، سلسله مراتب استنتاجی بصورت زیر را در نظر می‌گیریم که داده‌ها با چه هدفی جمع آوری می‌شوند؟

۱. در جهت استنباط فقط برای فردی که اطلاعات را جمع‌آوری می‌کند؟
۲. در جهت استنباط برای یک سازمان؟
۳. در جهت استنباط‌های منتشر شده در یک نشریه علمی؟
۴. در جهت استنباط عموم مردم (مثلاً یک پایگاه داده عمومی)؟
۵. در جهت شناسایی فرضیه‌های علت و معلولی احتمالی؟
۶. در جهت اثبات روابط علت و معلولی؟

## ۲.۵ توصیه‌ها

لذا با توجه به روش‌های متنوع نمونه‌گیری که در بالا مورد بحث قرار گرفت، هر دو نمونه‌برداری احتمالی و غیراحتمالی باید در جهت توسعه درک ما از وضعیت نقش داشته باشند. برآوردی از هزینه و عملی بودن آن را بررسی کنیم. توصیه‌های زیر را برای توسعه چنین درکی در ذهن پیشنهاد می‌شود.

۱. از نمونه‌گیری غیراحتمالی زمانی استفاده کنید که جامعه هدف و نمونه واقعی یکسان هستند.
۲. زمانی که قرار است تصمیمی سریع گرفته شود، نمونه‌گیری هدفمند را در نظر بگیرید. این روش به طور کلی بسیار ارزان‌تر است و محافظت بیشتری در برابر اندازه نمونه‌های کوچک ارائه می‌دهد.
۳. از نمونه‌گیری احتمالی برای فرمول‌بندی فرضیه‌های مورد آزمایش استفاده کنید. نمونه‌گیری هدفمند معمولاً برای این منظور کارآمدتر است اما می‌تواند گمراه‌کننده‌تر باشد.
۴. از هرگونه ادعا مبنی بر اینکه یک نمونه هدفمند نمونه‌ای احتمالی و در نتیجه نماینده است خودداری کنید. مفروضات انجام شده را به وضوح بیان کنید، اطلاعات نمونه‌برداری شده برای چه مواردی می‌تواند مورد استفاده قرار گیرد و نمی‌تواند مورد استفاده قرار گیرد و بر اساس آن استنتاج کنید.

۵. از نمونه‌گیری احتمالی برای شناسایی علت-معلول و سپس ایجاد علت از طریق آزمایش استفاده کنید یا شاید به طور مؤثرتری از آزمایش برای مستندسازی علت-معلول در شرایط کنترل‌شده استفاده کنید و سپس سعی کنید از طریق نمونه‌گیری احتمالی استنتاج کنید.

## ۶ تحولات اخیر در حوزه نمونه‌گیری غیراحتمالی

در نمونه‌گیری غیراحتمالی با توجه به عدم تصادفی بودن این نوع انتخاب، معمولاً با چالش‌هایی مانند سوگیری و کاهش دقت مواجه هستیم. اما در سال‌های اخیر، مطالعاتی در زمینه بهبود دقت و کاهش سوگیری در انتخاب نمونه‌ها با استفاده از روش‌های نمونه‌گیری غیراحتمالی انجام شده است. در ادامه به بررسی برخی از این تحولات اشاره می‌کنیم.

### ۱.۶ استفاده از روش‌های ترکیبی

یکی از رویکردهای جدید در جهت کاهش سوگیری در نمونه‌گیری غیراحتمالی ترکیب روش نمونه‌گیری غیر احتمالی با روش‌های احتمالی است. به عنوان مثال ترکیب روش‌های غیراحتمالی (مانند نمونه‌گیری سهمیه‌ای یا گلوله‌برفی) با روش‌های احتمالی (مانند تصادفی ساده یا طبقه‌ای) یک راهکار مؤثر برای کاهش سوگیری و افزایش تعمیم‌پذیری نتایج است [۹] و [۷]. این روش از مزایای هر دو رویکرد بهره می‌برد، از یک سو باعث کاهش سوگیری انتخاب و افزایش اعتبار آماری که رویکردی از نمونه‌گیری احتمالی است و از سوی دیگر دسترسی به گروه‌های خاص یا کمیاب که رویکردی از نمونه‌گیری غیراحتمالی است را فراهم می‌کند.

مثال ۱.۶. فرض کنید به دنبال مطالعه رفتار خرید مصرف‌کنندگان در یک فروشگاه زنجیره‌ای با هدف بررسی ترجیحات خریداران با ترکیب روش‌های احتمالی و غیراحتمالی هستیم. در ابتدا به منظور انجام نمونه‌گیری احتمالی تصادفی (طبقه‌بندی شده) مشتریان را به طبقاتی بر اساس سن، جنسیت یا محل سکونت (داده‌های موجود از سیستم فروشگاه) تقسیم می‌کنیم. از هر طبقه (مثلاً ۱۰۰ نفر از هر گروه سنی) به صورت تصادفی انتخاب می‌کنیم. سپس برای دستیابی به مصرف‌کنندگان کم‌مراجعه یا وفادار از برخی مشتریان انتخاب‌شده

۲. محاسبه سهم هر لایه در جامعه: سهم هر لایه در جامعه هدف (مثلاً ۴۰٪ زنان و ۶۰٪ مردان) از منابعی مانند سرشماری یا داده‌های معتبر دیگر استخراج می‌شود.

۳. محاسبه سهم هر لایه در نمونه: نسبت هر لایه در نمونه جمع‌آوری شده محاسبه می‌شود (مثلاً ۳۰٪ زنان و ۷۰٪ مردان).

۴. محاسبه وزن‌ها: برای هر لایه، وزن به صورت نسبت سهم جامعه به سهم نمونه محاسبه می‌شود. مثلاً اگر زنان در جامعه ۴۰٪ باشند اما در نمونه فقط ۳۰٪، وزن زنان ۱٫۳۳ خواهد بود.

۵. اعمال وزن به داده‌ها: پاسخ‌دهندگان هر لایه در تحلیل‌های آماری با وزن محاسبه شده ضرب می‌شوند تا تأثیرشان متناسب با سهم واقعی‌شان در جامعه شود.

مثال ۳.۶. فرض کنید در یک نظرسنجی سیاسی نسبت جنسیت در جامعه به صورت ۵۰٪ مرد، ۵۰٪ زن و نمونه جمع‌آوری شده شامل ۷۰٪ مرد، ۳۰٪ زن باشد. در این حالت وزن مردان ۷۱٪ و وزن زنان ۶۷٫۱ خواهد بود. با اعمال این وزن‌ها، پاسخ‌های زنان (که در نمونه کم‌نمایی شده‌اند) تأثیر بیشتری در نتایج نهایی خواهند داشت و در نتیجه پاسخ‌های وزن‌دار شده نمایندگی بهتری از جامعه واقعی دارند.

### ۳.۶ نمونه‌گیری مبتنی بر مه‌داده‌ها

در سال‌های اخیر، ظهور مه‌داده‌ها تحولی اساسی در روش‌های نمونه‌گیری ایجاد کرده است. یکی از مهم‌ترین کاربردهای این فناوری، کاهش سوگیری در نمونه‌گیری‌های غیراحتمالی است که همواره به دلیل عدم تعادل در توزیع نمونه‌ها با چالش مواجه بوده‌اند. نمونه‌گیری مبتنی بر مه‌داده‌ها<sup>۳۴</sup> با بهره‌گیری از حجم انبوه داده‌های تولیدشده در محیط‌های دیجیتال، امکان دستیابی به نمونه‌هایی با تنوع و نمایندگی بهتر را فراهم می‌سازد.

مثال ۴.۶. در مطالعه‌ای با هدف تحلیل رفتار رأی‌دهندگان در فضای مجازی و پیش‌بینی نتایج انتخابات، پژوهشگران می‌توانند از داده‌های شبکه‌های اجتماعی استفاده کنند. در روش‌های سنتی، به دلیل محدودیت‌های نمونه‌گیری غیراحتمالی، معمولاً گروه‌های خاصی مانند جوانان یا ساکنان مناطق شهری بیش از حد نمایندگی می‌شوند. این در حالی است که با به کارگیری مه‌داده‌ها، می‌توان با مراحل زیر این سوگیری را کاهش داد:

در مرحله قبل خواسته می‌شود که دوستان یا خانواده‌شان را که اخیراً خرید کرده‌اند معرفی کنند (نمونه‌گیری غیراحتمالی گلوله‌برفی). گروه آخر معمولاً در نمونه‌گیری تصادفی کمتر دیده می‌شوند. برای پوشش گروه‌های خاص (مثلاً مشتریان آنلاین یا کسانی که فقط در شب خرید می‌کنند)، نمونه‌هایی به صورت هدفمند اضافه می‌شوند نمونه‌گیری سهمیه‌ای (غیراحتمالی).

ترکیب روش‌ها در مثال فوق با هدف کاهش سوگیری از طریق نمونه‌گیری تصادفی طبقه‌بندی شده اطمینان می‌دهد که همه زیرگروه‌ها نمایندگی می‌شوند و از طرف دیگر با هدف دستیابی به گروه‌های کمیاب و استفاده از روش گلوله‌برفی مشتریان کم‌مراجعه یا وفادار را وارد مطالعه می‌کند. لذا داده‌ها هم تعمیم‌پذیر هستند (بخش احتمالی) و هم کیفی (بخش غیراحتمالی). لذا ترکیب نمونه‌گیری احتمالی (برای پایه علمی) و غیراحتمالی (برای دسترسی به گروه‌های هدف) یک راهکار عملی برای کاهش سوگیری در پژوهش‌های کاربردی است. مثال بالا نشان می‌دهد که چگونه این روش در تحقیقات بازار می‌تواند هم اعتبار و هم کامل بودن داده‌ها را بهبود دهد.

### ۲.۶ کالیبراسیون

کالیبراسیون<sup>۳۳</sup> یا وزن‌دهی پس از لایه‌بندی<sup>۳۴</sup> یک روش آماری برای تعدیل نمونه‌های غیرنماینده است تا با جمعیت هدف (جامعه آماری) همخوانی بیشتری پیدا کند. این روش معمولاً پس از جمع‌آوری داده‌ها و برای اصلاح سوگیری‌های نمونه‌گیری استفاده می‌شود. در این روش با استفاده از اطلاعات کمکی (مانند سرشماری یا داده‌های مرجع)، وزن‌هایی به نمونه‌ها اختصاص داده می‌شود تا توزیع نمونه با توزیع جامعه هدف همخوانی بیشتری پیدا کند [۹]، [۱۴] و [۱۵].

مثال ۲.۶. اگر در یک نظرسنجی (نمونه‌گیری داوطلبانه)، زنان ۲۰٪ تا ۳۰٪ بیش‌نمایی شده‌اند، لذا وزن این گروه کاهش می‌یابد تا با سهم واقعی آنها در جامعه مطابقت داشته باشد.

### ۱.۲.۶ مراحل کالیبراسیون یا وزن‌دهی پس از لایه‌بندی

۱. تعیین لایه‌ها<sup>۳۵</sup>: جامعه آماری بر اساس متغیرهای مهم (مانند سن، جنسیت، تحصیلات، منطقه جغرافیایی و ...) به زیرگروه‌های همگن تقسیم می‌شود.

<sup>33</sup>Calibration

<sup>34</sup>Post-Stratification Weighting

<sup>35</sup>Strata

<sup>36</sup>Big Data Sampling

سهمی‌ای تطبیقی استفاده می‌کند تا اطمینان حاصل شود که ترکیب نمونه از نظر سن، جنسیت، تحصیلات و ... متناسب با جامعه هدف است.

**کاهش بار شناختی<sup>۴۱</sup>:** با حذف سوالات تکراری یا غیرضروری، کیفیت پاسخ‌ها را افزایش می‌دهد و از سوگیری پاسخ‌ها جلوگیری می‌کند.

**مثال ۵.۶.** یک شرکت تحقیقاتی از پرسشنامه‌های هوشمند برای بهبود نمونه‌گیری غیراحتمالی استفاده می‌کند. این سیستم به‌صورت خودکار افرادی را که ویژگی‌های جمعیت‌شناختی خاصی دارند (مثلاً افراد مسن یا اقلیت‌های قومی) هدف قرار می‌دهد تا تعادل نمونه حفظ شود. در نتیجه، داده‌های جمع‌آوری شده به‌رغم استفاده از روش‌های غیراحتمالی، از سوگیری کمتری برخوردار هستند.

## ۵.۶ جوسازی امتیاز تمایل

**تعریف ۶.۶.** متغیر مخدوش‌کننده<sup>۴۲</sup> عاملی است که هم با متغیر مستقل و هم با متغیر وابسته ارتباط دارد و باعث تحریف رابطه واقعی بین آنها می‌شود [۱۲].

**مثال ۷.۶.** در مطالعه بررسی ارتباط بین مصرف قهوه (مستقل) و سرطان ریه (وابسته) سیگار کشیدن را می‌توان به عنوان متغیر مخدوش‌کننده در نظر گرفت بدلیل اینکه سیگار کشیدن با مصرف قهوه مرتبط است (سیگاری‌ها بیشتر قهوه می‌نوشند). سیگار کشیدن مستقل از قهوه، باعث سرطان ریه می‌شود و اگر عامل سیگار کشیدن را در نظر نگیریم، ممکن است اشتباه نتیجه بگیریم قهوه باعث ایجاد سرطان ریه در بیمار شده است.

جوسازی امتیاز تمایل<sup>۴۳</sup> یکی از روش‌های کاهش سوگیری در مطالعات مشاهده‌ای (غیرآزمایشی) است که برای مقایسه گروه‌های درمان شده و درمان نشده با کنترل متغیرهای مخدوش‌کننده به کار می‌رود. این روش با تخمین احتمال دریافت درمان بر اساس متغیرهای پیش‌بین، گروه‌های مشابه را جور می‌کند تا اثر درمان را بدون تأثیر متغیرهای مخدوش‌کننده برآورد کند [۲۲].

**گردآوری داده‌های حجیم:** جمع‌آوری تمامی توییت‌های مرتبط با انتخابات در بازه زمانی مشخص.

**پالایش و طبقه‌بندی داده‌ها:** استفاده از الگوریتم‌های یادگیری ماشین برای شناسایی ویژگی‌های جمعیت‌شناختی کاربران.

**تنظیم نمونه‌ها:** تعدیل ترکیب نمونه نهایی به گونه‌ای که انعکاسی از ساختار جمعیتی جامعه باشد.

**تحلیل نهایی:** اعمال وزن‌های آماری مناسب به هر گروه برای افزایش دقت پیش‌بینی‌ها.

این رویکرد نه تنها امکان دستیابی به نمونه‌های متوازن‌تر را فراهم می‌کند، بلکه با در اختیار گذاشتن حجم بیشتری از داده‌ها، دقت تحلیلی پژوهش را نیز افزایش می‌دهد [۸] و [۲۴].

## ۴.۶ طراحی پرسشنامه‌های هوشمند

توسعه پرسشنامه‌های هوشمند که به صورت پویا و بر اساس پاسخ‌های اولیه شرکت‌کنندگان طراحی می‌شوند، می‌تواند به کاهش سوگیری در جمع‌آوری داده‌ها کمک کند. این روش به محققان اجازه می‌دهد تا بر اساس اطلاعات جمع‌آوری شده، سوالات بعدی را تنظیم کنند و از ایجاد سوگیری‌های احتمالی جلوگیری کنند. پرسشنامه‌های هوشمند با استفاده از الگوریتم‌های تطبیقی و پویایی در طراحی سوالات، سعی در کاهش این سوگیری دارند [۲۱]. ویژگی‌های پرسشنامه‌های هوشمند برای کاهش سوگیری به شرح زیر است:

**تطبیق پویای سوالات<sup>۴۷</sup>:** سوالات بر اساس پاسخ‌های قبلی پاسخ‌دهنده تغییر می‌کنند تا داده‌های متنوع‌تری جمع‌آوری شود. برای مثال اگر پاسخ‌دهنده در سوال اول "زن" باشد، سوال بعدی ممکن است بر اساس جنسیت تنظیم شود تا اطمینان حاصل شود که دیدگاه‌های مردان نیز به اندازه کافی پوشش داده می‌شود.

**تخصیص هوشمند نمونه<sup>۴۸</sup>:** سیستم به‌صورت خودکار گروه‌های کم نمایش شده<sup>۴۹</sup> را شناسایی و هدف قرار می‌دهد. برای مثال در یک نظرسنجی سیاسی، اگر پاسخ‌دهندگان از یک حزب خاص بیش‌ازحد حضور داشته باشند، پرسشنامه به‌صورت خودکار افراد با گرایش‌های دیگر را جذب می‌کند.

**تعادل جمعیت‌شناختی<sup>۴۰</sup>:** از تکنیک‌هایی مانند نمونه‌گیری

<sup>37</sup>Dynamic Questioning

<sup>38</sup>Adaptive Sampling

<sup>39</sup>Underrepresented Groups

<sup>40</sup>Demographic Balancing

<sup>41</sup>Cognitive Load Reduction

<sup>42</sup>Confounding

<sup>43</sup>Propensity Score Matching (PSM)

## ۱.۵.۶ مراحل انجام جورسازی امتیاز تمایل

**جورسازی:** هر فرد درمان‌شده با فردی درمان‌نشده که امتیاز تمایل مشابهی دارد، به کمک الگوریتم‌هایی مانند نزدیکترین همسایگی جور می‌شود<sup>۴۴</sup>.

**ارزیابی تعادل:** پس از جورسازی، توزیع متغیرهای مخدوش‌کننده بین دو گروه باید مشابه باشد.

**تخمین اثر درمان:** مقایسه نتایج مثلاً میانگین تفاوت در نتیجه بین گروه‌های جورشده.

**تخمین امتیاز تمایل:** با استفاده از مدل رگرسیون (معمولاً لجستیک)، احتمال دریافت درمان برای هر فرد بر اساس متغیرهای زمینه‌ای (سن، جنسیت، تحصیلات و...) محاسبه می‌شود.

$$e(X) = P(T = 1 | X)$$

۸۰٪) را با یک دانش آموز مشابه اما غیرشرکت‌کننده (مثلاً حسن با امتیاز ۷۸٪) جور می‌کنیم. دو گروه حالا از نظر ویژگی‌های زمینه‌ای (معدل، وضعیت مالی و...) بسیار شبیه هستند. حال میانگین نمرات دو گروه جورشده را مقایسه می‌کنیم. اگر گروه کلاس‌رفته‌ها نمره بالاتری داشتند، می‌توانیم بگوییم احتمالاً کلاس خصوصی مؤثر بوده، نه به خاطر هوش یا انگیزه بالاتر دانش آموزان. جدول زیر این جورسازی امتیاز تمایل برای دانش آموزان را نمایش می‌دهد.

در مرحله جورسازی نمرات علی با امتیاز تمایل ۸۰ و رضا با امتیاز تمایل ۷۵ و نیز نمرات سارا با امتیاز تمایل ۲۰ و نیما با امتیاز تمایل ۱۸ مقایسه می‌شوند، حال اگر میانگین نمرات علی و سارا (کلاس‌رفته‌ها) از رضا و نیما (کلاس‌نرفته‌ها) بالاتر بود، اثر کلاس تأیید می‌شود. در عمل انجام آزمایش تصادفی (مثل تقسیم تصادفی دانش آموزان به کلاس رفته و کلاس نرفته) همیشه ممکن نیست. *PSM* کمک می‌کند تا از داده‌های مشاهده‌ای (غیرآزمایشی) نتایج معتبرتری بگیریم.

فرض کنید می‌خواهیم بررسی کنیم آیا کلاس خصوصی ریاضی، واقعاً نمرات دانش‌آموزان را بهبود می‌بخشد؟ برای این کار اگر فقط نمرات دانش‌آموزان شرکت‌کننده در کلاس را با بقیه مقایسه کنید، در نتیجه‌گیری اشتباه می‌شود. چون ممکن است دانش‌آموزانی که کلاس می‌روند، از ابتدا با انگیزه‌تر یا باهوش‌تر باشند (متغیرهای مخدوش‌کننده). در حالی که روش *PSM* با مقایسه دانش‌آموزان مشابه (از نظر ویژگی‌هایی مثل معدل قبلی، وضعیت مالی، جنسیت و...) که فقط در کلاس شرکت کرده یا نکرده‌اند، این سوگیری را کاهش می‌دهد. یک مدل آماری (مثل رگرسیون لجستیک) می‌سازیم که پیش‌بینی می‌کند احتمال شرکت یک دانش‌آموز در کلاس بر اساس ویژگی‌هایش چقدر است. برای مثال علی دانش‌آموزی با معدل بالا، از خانواده ثروتمند و احتمال شرکت در کلاس ۸۰٪ رضا دانش‌آموزی با معدل پایین، از خانواده ای کم‌درآمد و احتمال شرکت در کلاس ۱۰٪ باشند. در مرحله جورسازی هر دانش‌آموز شرکت‌کننده (مثلاً علی با امتیاز

جدول ۱: مشخصات دانش‌آموزان و تمایل به شرکت در کلاس

دانش‌آموز	آیا کلاس	معدل	وضعیت مالی	امتیاز تمایل
علی	بلی		بالا	۰.۸۰
رضا	خیر		بالا	۰.۷۵
سارا	بلی		پایین	۰.۲۰
نیما	خیر		پایین	۰.۱۸

## مراجع

[۱] عمیدی، ع. (۱۳۹۲). نظریه نمونه‌گیری و کاربردهای آن. تهران: مرکز نشر دانشگاهی.

[۲] مختاری، گ. و امجدی، ا. (۱۳۸۱). نمونه‌گیری: روش‌ها و کاربردها. تهران: پژوهشکده آمار ایران.

<sup>44</sup>Nearest Neighbor (Caliper Matching)

- [3] Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and how of it. *Indian Journal of Medical Specialties*, **4**(2), 330-333.
- [4] Booth, C. (1889). *Life and Labour of the People in London*. London: Macmillan.
- [5] Bowley, A. L. (1915). *The Measurement of Social Conditions*. *The Economic Journal*, **25**(97), 1-20.
- [6] Boyd, R. J., Powney, G. D., & Pescott, O. L. (2023). We need to talk about nonprobability samples. *Trends in Ecology & Evolution*, **38**(6).
- [7] Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). Wiley.
- [8] Czajka, J. L., & Beyler, A. (2016). Declining Response Rates in Federal Surveys: Trends and Implications. *Mathematica Policy Research*, Final Report - Volume I, June 15, 2016.
- [9] Dever, J. A. (2018). Combining Probability and Nonprobability Samples to Form Efficient Hybrid Estimates: An Evaluation of the Common Support Assumption. *Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference*.
- [10] Einav, L., & Levin, J. (2014). *The Data Revolution and Economic Analysis*. NBER Working Paper No. 19035.
- [11] Elliott, M. R., & Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, **32**(2), 249-264.
- [12] Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- [13] Gallup, G. H., & Rae, R. (1933). *The Gallup Poll: Public Opinion 1935-1971*. New York: Random House.
- [14] Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.
- [15] Kalton, G. (2023). Probability vs. nonprobability sampling: from the birth of survey sampling to the present day. *Statistics in Transition. New Series*, **24**(3), 1-22.
- [16] Little, R. J. (1993). Post-Stratification: A Modeler's Perspective. *Journal of the American Statistical Association*, **88**(423), 1001-1012.
- [17] Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. Wiley.
- [18] Meng, X. L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *Annals of Applied Statistics*, **12**, 685-726.
- [19] Neyman, J., & Pearson, E. S. (1933). *On the Problem of the Most Efficient Tests of Statistical Hypotheses*. *Philosophical Transactions of the Royal Society of London, Series A*, **231**(694-706), 289-337.
- [20] Neyman, J. (1934). On the Two Different Aspects of the Random Sampling Theory: The Method of Stratified Sampling and the Method of Sampling with Replacement. *Journal of the Royal Statistical Society*, **97**(4), 558-606.
- [21] Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society A*, **236**(767), 333-380.

- [22] Pearson, K. (1900). *Mathematical Contributions to the Theory of Evolution*. *Philosophical Transactions of the Royal Society of London*, **195**, 1-47.
- [23] Rivera, S. S., & Rosenbaum, J. E. (2021). Adaptive Survey Designs for Non-Probability Samples: Reducing Bias Through Smart Questionnaires. *Journal of Survey Statistics and Methodology*, **9**(2), 245-267.
- [24] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1), 41-55.
- [25] Rowntree, S. (1901). *Poverty: A Study of Town Life*. London: Macmillan.
- [26] Salganik, M. J. (2018). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.
- [27] Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge University Press.
- [28] U.S. Bureau of the Census. (1940). *Sixteenth Census of the United States: 1940, Volume I, Population*. Washington, D.C.: U.S. Government Printing Office.
- [29] U.S. Bureau of the Census. (1941). *The Sample Census of Population: 1940*. Washington, D.C.: U.S. Government Printing Office.

## **A Study on Probabilistic and Non-Probabilistic Sampling Methods**

**H. Goli Nari<sup>1\*</sup>, M. Khorashadi Zadeh<sup>2</sup>, G. R. Mohtashami Borzadaran<sup>3</sup> and F. Mehran<sup>4</sup>**

<sup>1</sup> PhD Candidate in Statistics, University of Birjand, Iran

<sup>2</sup> Associate Professor, Department of Statistics, University of Birjand, Iran

<sup>3</sup> Professor Emeritus, Department of Statistics, Ferdowsi University of Mashhad, Iran

<sup>4</sup> PhD in Statistics, Harvard University and Retired Member, International Labour Institute

### **Abstract:**

The selection of sampling methods is of paramount importance in research methodology. Sampling techniques are broadly classified into probability and non-probability methods. While probability sampling methods are well-established with extensive literature, non-probability sampling techniques - despite their growing applications in modern research and historical significance - remain understudied and less recognized among researchers. This paper examines the conceptual framework and philosophy of sampling, reviews the historical evolution of sampling methods, and discusses the rationale for adopting non-random sampling approaches. We present a comprehensive classification of both probability and non-probability sampling techniques with brief methodological explanations. Furthermore, we explore recent innovations that enhance the quality and precision of non-probability samples.

**Keywords:** Probability sampling, Non-probability sampling, Bias reduction, Snowball sampling, Core, Wave, Calibration, Big data, Propensity score matching.