

به نام خدا

چند دقیقه با آمار

آمار مقدماتی (جلسه ۶)

موضوع:

نمونه گیری در آمار: انواع و خطاها

تهیه و تدوین:

دکتر اکبر آبروش

محمد سجاد خیاط حسینی

هدیه احمدی

مریم حقی

فاطمه قاسمی

مریم جهانگیری

سینا کوهی

زهرا جوادی



[تلگرام](#)

[اینستاگرام](#)



[تلگرام](#)

[اینستاگرام](#)



[تلگرام](#)

[اینستاگرام](#)



[تلگرام](#)

[اینستاگرام](#)

فروردین ماه ۱۴۰۴

فهرست مطالب

۱ جامعه

۲ انواع متداول نمونه‌گیری

۳ انتخاب روش نمونه‌گیری مناسب

۴ خطای نمونه‌گیری چیست؟

۵ نوع دیگری از خطا

۶ منابع

جامعه

نمونه‌ها بخش‌هایی از یک جامعه هستند. برای مثال، ممکن است فهرستی از اطلاعات ۱۰۰ نفر ("نمونه" شما) از ۱۰۰۰۰ نفر ("جامعه") داشته باشید. می‌توانید از این فهرست برای انجام برخی فرضیات در مورد رفتار کل جامعه استفاده کنید. با این حال، به این سادگی نیست. وقتی آمار انجام می‌دهید، اندازه نمونه شما باید ایده‌آل باشد نه خیلی بزرگ و نه خیلی کوچک. سپس هنگامی که اندازه نمونه را مشخص کردید، باید از یک تکنیک صحیح برای جمع‌آوری نمونه از جامعه استفاده کنید:

نمونه‌گیری احتمالی از تصادفی‌سازی برای انتخاب اعضای نمونه استفاده می‌کند. شما احتمال گنجاندن هر عضو بالقوه در نمونه را می‌دانید. برای مثال، $1/100$. با این حال، لازم نیست که شانس‌ها برابر باشند. برخی از اعضا ممکن است $1/100$ شانس انتخاب شدن داشته باشند، برخی دیگر ممکن است $1/50$ شانس داشته باشند. نمونه‌گیری غیر احتمالی از تکنیک‌های غیر تصادفی (یعنی قضاوت محقق) استفاده می‌کند. شما نمی‌توانید شانس گنجاندن هر مورد، شخص یا چیز خاصی را در نمونه خود محاسبه کنید.

انواع متداول

متداول‌ترین تکنیک‌هایی که احتمالاً در آمار مقدماتی با آن‌ها مواجه می‌شوید عبارتند از گرفتن نمونه با و بدون جایگزینی. تکنیک‌های خاص عبارتند از:

- **نمونه‌های برنولی:** آزمایش‌های برنولی مستقلی روی عناصر جامعه دارند. این آزمایش‌ها تعیین می‌کنند که آیا عنصر به بخشی از نمونه تبدیل می‌شود یا خیر. همه عناصر جامعه از شانس برابری برای گنجاندن در هر انتخاب از یک نمونه برخوردارند. اندازه نمونه‌ها در نمونه‌های برنولی از توزیع دوجمله‌ای پیروی می‌کنند.
 - **نمونه‌های پواسون (کمتر رایج):** یک آزمایش برنولی مستقل تعیین می‌کند که آیا هر عنصر جامعه به نمونه می‌رسد یا خیر.
 - **نمونه‌های خوشه‌ای:** جامعه را به گروه‌هایی (خوشه‌ها) تقسیم می‌کنند. سپس یک نمونه تصادفی از خوشه‌ها انتخاب می‌شود. این روش زمانی استفاده می‌شود که محققان افراد یک جامعه را نمی‌شناسند اما زیرمجموعه‌ها یا گروه‌های جامعه را می‌شناسند.
 - **نمونه‌گیری سیستماتیک:** عناصر نمونه را از یک چارچوب مرتب انتخاب می‌کنید. چارچوب نمونه‌گیری فقط فهرستی از شرکت‌کنندگانی است که می‌خواهید از آنها نمونه بگیرید. برای مثال، در روش احتمال برابر، یک عنصر را از یک لیست انتخاب کنید و سپس با استفاده از معادله $k = N/n$ ، هر k امین عنصر را انتخاب کنید. " n " کوچک نشان‌دهنده اندازه نمونه و " N " بزرگ برابر با اندازه جامعه است.
 - **SRS:** موارد را کاملاً تصادفی انتخاب کنید، به طوری که هر عنصر همان احتمال انتخاب شدن هر عنصر دیگر را داشته باشد. هر زیرمجموعه از عناصر همان احتمال انتخاب شدن هر زیرمجموعه دیگر از k عنصر را دارد.
 - **نمونه‌گیری طبقه‌ای:** از هر زیرجامعه به طور مستقل نمونه بگیرید. ابتدا جامعه را به زیرگروه‌های همگن (بسیار مشابه) قبل از گرفتن نمونه تقسیم کنید. هر عضو جامعه فقط به یک گروه تعلق دارد. سپس روش تصادفی ساده یا سیستماتیک را در هر گروه برای انتخاب نمونه اعمال کنید.
 - **تصادفی‌سازی طبقه‌ای:** یک زیرگروه از طبقه‌بندی است که در آزمایشات بالینی استفاده می‌شود. ابتدا بیماران را به طبقات تقسیم کنید، سپس با تصادفی‌سازی بلوک جایگشت تصادفی کنید.
- به ندرت (اگر نگوییم هرگز) در یک کلاس آمار پایه با این تکنیک‌ها مواجه خواهید شد. با این حال، شما در "دنیای واقعی" با آنها مواجه خواهید شد:
- **نمونه‌گیری پذیرش-رد:** روشی برای نمونه‌گیری از یک توزیع ناشناخته با استفاده از یک توزیع مشابه و راحت‌تر.
 - **نمونه‌گیری تصادفی (همچنین به عنوان نمونه‌گیری فرصت، راحتی یا اتفاقی شناخته می‌شود):** یک نمونه از یک جامعه راحت و در دسترس بکشید. این یک نمونه نماینده برای جامعه ارائه نمی‌دهد اما می‌تواند برای آزمایش مفید باشد.
 - **نمونه‌گیری تطبیقی (همچنین طرح‌های پاسخ تطبیقی نامیده می‌شود):** معیارهای انتخاب خود را با پیشرفت آزمایش، بر اساس نتایج اولیه که به دست می‌آیند، تطبیق دهید.
 - **نمونه بوت استرپ:** یک نمونه کوچکتر از یک نمونه بزرگتر با بوت استرپینگ انتخاب کنید. بوت استرپینگ نوعی نمونه‌گیری مجدد است که در آن تعداد زیادی نمونه کوچکتر با اندازه یکسان، با جایگزینی، از یک نمونه اصلی می‌گیرید.

- **الگوریتم دیو (فیزیک):** اعضای یک مجموعه میکروکانونی (که برای نشان دادن حالت‌های ممکن یک سیستم مکانیکی که انرژی کل دقیقاً مشخصی دارد استفاده می‌شود) را با یک انرژی معین نمونه‌برداری می‌کند. "دیو" نشان‌دهنده درجه‌ای از آزادی در سیستم است که انرژی را ذخیره و تامین می‌کند.
- **نمونه‌های مورد بحرانی:** با این روش، موارد را با دقت انتخاب می‌کنید تا اطلاعاتی را که می‌توانید از تعداد انگشت شماری نمونه دریافت کنید، به حداکثر برسانید.
- **نمونه‌گیری موردی متناقض:** مواردی را انتخاب می‌کنید که به نظر می‌رسد با یافته‌های شما در تضاد هستند.
- **نمونه فاصله:** یک تکنیک پرکاربرد که چگالی یا فراوانی جامعه حیوانات را برآورد می‌زند.
- **روش نمونه‌گیری تجربه:** تجربیات را نمونه‌برداری می‌کند (به جای افراد یا اعضا). در این روش، شرکت‌کنندگان در مطالعه در زمان‌های معینی متوقف می‌شوند و تجربیات خود را همانطور که تجربه می‌کنند یادداشت می‌کنند.
- **نمونه‌گیری تصادفی:** جایی که یک محقق موارد را به طور تصادفی انتخاب می‌کند و سعی می‌کند تصادفی بودن را شبیه‌سازی کند. با این حال، نتیجه ممکن است اصلاً تصادفی نباشد.

احتمالاً در یک کلاس آمار پایه با این موارد مواجه نخواهید شد.

- **نمونه معکوس:** بر اساس نمونه‌گیری دوجمله‌ای منفی. نمونه‌ها را تا زمانی که تعداد مشخصی از موفقیت‌ها اتفاق بیفتد، بگیرید.
- **نمونه‌گیری اهمیت:** روشی برای مدل‌سازی رویدادهای نادر.
- **شبکه Kish:** روشی برای انتخاب اعضای یک خانواده برای مصاحبه و استفاده از جداول اعداد تصادفی برای انتخاب.
- **مکعب بزرگ لاتین:** برای ساخت آزمایش‌های کامپیوتری استفاده می‌شود. نمونه‌هایی از مجموعه‌های قابل قبول مقادیر را برای پارامترها در یک توزیع چند بعدی تولید می‌کند.
- **نمونه‌گیری خط-برخورد:** روشی که در آن یک عنصر را از یک منطقه خاص در یک نمونه قرار می‌دهید اگر یک قطعه خط معین با آن عنصر قطع شود.
- **از نمونه‌های حداکثر تنوع زمانی استفاده کنید** که می‌خواهید موارد شدید (مانند غنی/فقیر یا جوان/پیر) را درج کنید. یک تکنیک مرتبط: نمونه‌گیری موردی شدید.
- **نمونه‌گیری چند مرحله‌ای:** یکی از انواع تکنیک‌های نمونه‌گیری خوشه‌ای که در آن عناصر تصادفی را از یک خوشه انتخاب می‌کنید (به جای هر عضو در خوشه).
- **نمونه‌گیری سهمیه‌ای:** روشی برای انتخاب شرکت‌کنندگان در نظرسنجی. این شبیه به نمونه‌گیری طبقه‌ای است اما محققان اعضای یک گروه را بر اساس قضاوت انتخاب می‌کنند. برای مثال، افرادی که به محقق نزدیک‌تر هستند ممکن است برای سهولت دسترسی انتخاب شوند.

- **نمونه‌گیری مخزن:** در داده‌کاوی برای به دست آوردن نمونه‌ای به اندازه n از یک جریان داده با طول ناشناخته استفاده می‌شود.
- **نمونه‌گیری مبتنی بر پاسخگو:** یک روش نمونه‌گیری زنجیره‌ای-ارجاعی که در آن شرکت‌کنندگان افراد دیگری را که می‌شناسند توصیه می‌کنند.
- یک نمونه متوالی اندازه ثابتی ندارد. موارد را یک (یا چند) مورد در یک زمان بگیرید تا زمانی که به اندازه کافی برای تحقیق خود داشته باشید. این روش معمولاً در بوم‌شناسی استفاده می‌شود.
- **نمونه‌های گلوله برفی:** جایی که شرکت‌کنندگان در مطالعه موجود، شرکت‌کنندگان آینده مطالعه را از افرادی که می‌شناسند جذب می‌کنند.
- **نمونه مغرضانه ریشه دوم:** روشی برای انتخاب افراد برای غربالگری‌های اضافی در فرودگاه‌ها. ترکیبی از SRS و پروفایل‌سازی.

انتخاب روش نمونه‌گیری مناسب:

نمونه

شما در آمار با اصطلاحات زیادی مواجه خواهید شد که روش‌های مختلف نمونه‌گیری را تعریف می‌کنند: نمونه‌گیری تصادفی ساده، نمونه‌گیری سیستماتیک، نمونه‌گیری تصادفی طبقه‌ای و نمونه‌گیری خوشه‌ای. تشخیص تفاوت بین روش‌های مختلف نمونه‌گیری می‌تواند یک چالش باشد. برای کسب اطلاعات بیشتر، ویدیوی زیر را تماشا کنید یا به خواندن ادامه دهید.

روش‌های مختلف نمونه‌گیری: چگونه تفاوت را تشخیص دهیم: مراحل

مرحله ۱: دریابید که آیا مطالعه از افراد نمونه‌برداری شده است (برای مثال، از بین گروهی از افراد انتخاب شده است). نمونه‌گیری تصادفی ساده را در قرعه‌کشی مدرسه خواهید یافت، جایی که نام افراد از یک کلاه انتخاب می‌شود. اما یک روش "سیستماتیک" تر برای انتخاب افراد را می‌توان در "نمونه‌گیری سیستماتیک" یافت، جایی که هر n مین فرد از یک جامعه انتخاب می‌شود. برای مثال، هر صدمین مشتری در یک فروشگاه خاص ممکن است یک هدیه "ویژه" دریافت کند.

مرحله ۲: دریابید که آیا مطالعه گروه‌هایی از شرکت‌کنندگان را انتخاب کرده است. برای تعداد زیادی از افراد (مانند تعداد سربازان وظیفه بالقوه در جنگ ویتنام)، انتخاب افراد به صورت گروهی بسیار ساده‌تر است (نمونه‌گیری تصادفی ساده). در مورد پیش‌نویس، سربازان وظیفه بر اساس تاریخ تولد انتخاب شدند که "رویه را ساده می‌کند".

مرحله ۳: تعیین کنید که آیا مطالعه شما شامل داده‌هایی از بیش از یک گروه با دقت تعریف شده ("طبقه" یا "خوشه") بوده است. برخی از نمونه‌های طبقات می‌توانند این باشند: دموکرات‌ها و جمهوری‌خواهان، مستاجران و صاحبان خانه، مردم روستا در مقابل ساکنان شهر، طرفداران جکسونوایل جگوارز و طرفداران سانفرانسیسکو ۴۹. اگر دو یا چند گروه بسیار مجزا و واضح وجود داشته باشد، شما یک نمونه طبقه‌ای یا یک "نمونه خوشه‌ای" دارید. اگر در مورد افراد گروه‌ها داده دارید، این یک نمونه طبقه‌ای است. برای انجام نمونه‌گیری طبقه‌ای روی این نمونه، می‌توانید نمونه‌گیری تصادفی از هر طبقه را به طور مستقل انجام دهید. اگر فقط در مورد خود گروه‌ها داده دارید (ممکن است فقط مکان افراد را بدانید)، این یک نمونه خوشه‌ای است.

مرحله ۴: دریابید که آیا نمونه به راحتی به دست آمده است. نمونه‌های راحتی مانند فروشگاه‌های راحتی هستند: چرا برای گرفتن نمونه به زحمت بیفتید، در حالی که می‌توانید به فروشگاه سر کوچه بروید؟ یک مثال کلاسیک از نمونه‌گیری راحتی، ایستادن در یک مرکز خرید و پرسیدن نظر از رهگذران است.

خطای نمونه‌گیری چیست؟

هنگامی که به جای استفاده از کل جامعه، یک نمونه از جامعه می‌گیرید، خطا رخ می‌دهد. به عبارت دیگر، این تفاوت بین آماری است که شما اندازه می‌گیرید و پارامتری که در صورت سرشماری کل جامعه پیدا می‌کنید. اگر قرار بود از کل جامعه (مانند سرشماری ایالات متحده) نظرسنجی کنید، هیچ خطایی وجود نخواهد داشت. محاسبه حاشیه خطا تقریباً غیرممکن است. با این حال، هنگامی که به طور تصادفی نمونه می‌گیرید، خطا را برآورد می‌کنید و آن را حاشیه خطا می‌نامید. یک نظرسنجی خوب برنامه‌ریزی شده می‌تواند خطا را کاهش دهد.

برای مثال، اگر می‌خواستید بفهمید که از هر هزار نفر چند نفر زیر ۱۸ سال دارند و به رقم ۱۹,۳۵۷ درصد رسیدید. اگر درصد واقعی برابر با ۱۹,۳۰۰٪ باشد، اختلاف (۱۹,۳۵۷ - ۱۹,۳۰۰) ۵۷,۰ یا ۰.۳٪ = حاشیه خطا. اگر به گرفتن نمونه‌های ۱۰۰۰ نفری ادامه می‌دادید، احتمالاً آمار کمی متفاوت، ۱۹,۱٪، ۱۸,۹٪، ۱۹,۵٪ و غیره به دست می‌آوردید، اما همه آنها در حدود یک رقم بودند. این یکی از دلایلی است که اغلب در نظرسنجی‌ها اندازه نمونه ۱۰۰۰ یا ۱۵۰۰ را مشاهده می‌کنید: آنها حاشیه خطای بسیار قابل قبولی در حدود ۳٪ ایجاد می‌کنند.

فرمول: فرمول حاشیه خطا $1/\sqrt{n}$ است، که در آن n اندازه نمونه است. برای مثال، یک نمونه تصادفی ۱۰۰۰ نفری حدود $1/\sqrt{n} = 3.2\%$ خطا دارد.

خطای نمونه فقط قابل کاهش است، به این دلیل که این یک بده بستان قابل قبول برای جلوگیری از اندازه‌گیری کل جامعه در نظر گرفته می‌شود. به طور کلی، هر چه نمونه بزرگتر باشد، حاشیه خطا کمتر است. یک استثنای قابل توجه وجود دارد: اگر از نمونه‌گیری خوشه‌ای استفاده کنید، این ممکن است به دلیل شباهت‌های بین اعضای خوشه، خطا را افزایش دهد. یک آزمایش یا نظرسنجی با دقت طراحی شده نیز می‌تواند خطا را کاهش دهد.

نوع دیگری از خطا

خطای غیر نمونه‌گیری می‌تواند یکی از دلایل تفاوت بین نمونه و جامعه باشد.^۱ این به دلیل روش‌های جمع‌آوری داده‌های ضعیف (مانند ابزارهای معیوب یا ثبت نادرست داده‌ها)، سوگیری انتخاب، سوگیری عدم پاسخ (جایی که افراد نمی‌خواهند یا نمی‌توانند به یک نظرسنجی پاسخ دهند)، یا سایر اشتباهات در جمع‌آوری داده‌ها باشد. افزایش اندازه نمونه این خطاها را کاهش نمی‌دهد. نکته کلیدی این است که با یک طرح خوب برنامه‌ریزی شده برای نظرسنجی یا آزمایش از ابتدا از این خطاها جلوگیری کنید.

منابع

1. <https://www.statisticshowto.com/>
2. Everitt, B. S.; Skrondal, A. (2010), The Cambridge Dictionary of Statistics, Cambridge University Press.
3. Klein, G. (2013). The Cartoon Introduction to Statistics. Hill & Wang.
4. Kotz, S.; et al., eds. (2006), Encyclopedia of Statistical Sciences, Wiley.
5. Vogt, W.P. (2005). Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences. SAGE.