

کار در کلاس



الف) کدام روش برای گردآوری هر یک از داده‌ها مناسب است؟

۱) تعداد قلم‌های هر دانش‌آموز در یک کلاس. مشاهده

۲) ساعات خواب دانش‌آموزان کلاس درس شما در شب گذشته. مصاحبه

۳) طول قد دانش‌آموزان یک کلاس. پرسش‌نامه

ب) می‌خواهیم طول قد دانش‌آموزان یک کلاس یا مدرسه را به یکی از سه روش زیر آمارگیری کنیم. هر یک از این روش‌ها محدودیت‌هایی دارند. چگونه می‌توان این محدودیت‌ها را از بین برد؟

پرسش‌نامه: اگر تعداد واحدهای نمونه زیاد باشد، این روش زمان‌بر است.

مشاهده: اگر به دقت زیادی نیاز داشته باشیم، مناسب نیست.

دادگان‌ها: همیشه اطلاعات ثبتی در اختیار نیست.

مصاحبه می‌تواند سرعت بیشتری داشته باشد. دانش‌آموزان قد خود را اعلام می‌کنند. مقدار پیامک بدهند.

خواندنی

آمارگیری را می‌توان به روش‌هایی بسیار سریع‌تر یا کم‌هزینه‌تر مانند آمارگیری پستی، تلفنی، اینترنتی یا پیامکی انجام داد. همچنین می‌توان با ابزاری نظیر «گوگل فرم» یک پرسش‌نامه طراحی کرد و آن را به نشانی نمونه انتخابی ارسال کرد و نتایج را از «گوگل فرم» بازیابی کرد.

فعالیت

قرار است درباره افرادی که از کوه دنا بالا رفته‌اند، پژوهشی آماری انجام دهیم. واحدهای آماری این پژوهش، همه افرادی هستند که توانسته‌اند به قله برسند. هدف از این پژوهش می‌تواند فرهنگی، یا علمی باشد. بسته به نوع پژوهش، یک یا چند ویژگی این افراد (مانند طول قد یا جنسیت) مورد نیاز است. به هر یک از این ویژگی‌ها که مورد پژوهش قرار می‌گیرد، متغیر می‌گویند. سایر متغیرها می‌توانند مواردی مانند: سن، وزن، ملیت، میزان تحصیلات و درآمد باشند. متغیرهای مورد بررسی در یک پژوهش ممکن است کمی یا کیفی باشند.

یادآوری

متغیر: هر ویژگی از اشخاص یا اشیا که قرار است بررسی شود.

متغیر کمی: متغیری است که مقادیر عددی می‌گیرد و برای آن عملیات ریاضی از قبیل

جمع، تفریق و معدل‌گیری قابل انجام است.

متغیر کیفی: متغیری است که صرفاً برای دسته‌بندی افراد یا اشیا در گروه‌ها به کار

می‌رود و لزوماً مقدار عددی نمی‌گیرد.

در مثال کوهنوردان دنا، سن، وزن، قد و درآمد یک کوهنورد، متغیرهای کمی اند. متغیرهای کیفی معمولاً از نوع مشاهدات غیر عددی اند و در مثال کوهنوردان دنا، جنسیت و ملیت را در بر می گیرند. به عنوان مثال، جنسیت برای دسته بندی افراد به مرد و زن استفاده می شود.

پارامتر یا پارامتر جامعه: یک مشخصه عددی است که توصیف کننده جنبه ای خاص از جامعه است و در صورتی که داده های کل جامعه در اختیار باشند، قابل محاسبه است.

مثلاً اگر داده های مربوط به تک تک کوهنوردان را داشته باشیم؛ یعنی به داده های جامعه دسترسی داریم. نسبت مردان در کل جامعه کوهنوردان، معرّف یک پارامتر است. اگر داده های بعضی از کوهنوردان را داشته باشیم؛ یعنی داده های نمونه را در اختیار داریم. نسبت مردان کوهنورد به این داده های نمونه ای را، آماره (مقدار آماره) گویند. آماره ها از نمونه ای به نمونه دیگر تغییر می کنند؛ این در حالی است که پارامترهای جامعه همیشه ثابت اند. چرا؟ در بسیاری از موارد، آمارگیری از کل جامعه امکان پذیر نیست. بنابراین، به رغم اینکه پارامتر مقدار ثابتی دارد، این مقدار مجهول است و به همین دلیل از آماره ها برای تخمین پارامترها استفاده می کنند.

آماره^۱ یا آماره نمونه: مشخصه ای عددی که توصیف کننده جنبه ای خاص از نمونه است و از داده های نمونه به دست می آید.

مثال: اداره کشاورزی استان خوزستان در حال ارزیابی هندوانه های آماده برداشت است. در این بررسی، هندوانه ها همان واحدهای آماری اند. اگر پژوهشگران وزن هندوانه ها را مورد بررسی قرار دهند، متغیر، «وزن» آنهاست. وزن یک متغیر کمی است؛ زیرا با مقادیر عددی ارائه می شود. اگر وزن تک تک هندوانه های این زمین بررسی شود، سرشماری از جامعه انجام داده ایم (که امکان پذیر نیست). متوسط وزن همه هندوانه های قابل برداشت در این زمین، «پارامتر» است. حال فرض کنیم پژوهشگران تصمیم دارند بر اساس معیار «مزه» هندوانه ها را مورد بررسی قرار دهند. در این حالت، مزه هندوانه ها را می توان به سه دسته تقسیم کرد: بد، قابل قبول و خوب. حال که می خواهیم مزه هندوانه ها را امتحان کنیم، مطالعه به بخشی از کل هندوانه ها محدود می شود. در اینجا «مزه» متغیری کیفی است؛ زیرا نمی توانیم همه هندوانه ها را مزه مزه کنیم، فقط بخشی از هندوانه ها مورد مطالعه قرار می گیرند؛ پس باید «نمونه» بگیریم. نسبت تعداد هندوانه های دارای مزه «خوب» در نمونه، یک «آماره» است.

فرایند نتیجه گیری درباره پارامترهای جامعه بر اساس نمونه، آمار استنباطی^۲ است.

۱- Statistic

۲- Inferential Statistics

روش‌های تولید ارقام (اعداد) تصادفی

سال‌ها پیش، اعداد تصادفی به صورت دستی و پس از آن به کمک الگوریتم‌های ریاضی ارقام شبه تصادفی تولید شده است و در کتاب‌هایی با همین عنوان، در اختیار محققان قرار می‌گرفت. امروزه ماشین‌حساب‌های علمی و نرم‌افزارهای رایانه‌ای این اعداد را در کسری از ثانیه تولید می‌کنند. از رقم‌های عدد پی نیز می‌توان به عنوان ارقام و اعداد تصادفی طبیعی استفاده کرد. خروجی تمامی روش‌های یادشده به صورت زیر است:

۴۶۹۱۴۳۹۳۲۸۲۶۵۸۰۲۵۴۶۴۸۱۶۲۹۹۷۶۱۱۵۷۳۱۳۹۹۴۱۷۳۶۷۴۳۲۳۶۲۰۳۶۴۷۳۴۳۴۲۶۹۴۹

تمرین

تمرین

- ۱ در نمونه‌گیری تصادفی ساده، احتمال اینکه فرد به خصوصی در اولین انتخاب عضو نمونه باشد، چقدر است؟ اگر مسئله با جای‌گذاری باشد، احتمال اینکه او در دومین انتخاب عضو نمونه باشد، چقدر است؟ اگر مسئله بدون جای‌گذاری باشد، و از نتیجه انتخاب اول اطلاع نداشته باشیم، احتمال اینکه او در دومین انتخاب عضو نمونه باشد، چقدر است؟
- ۲ آیا در نمونه‌گیری خوشه‌ای احتمال انتخاب واحدهای آماری برابر است؟ چرا؟ احتمال انتخاب خوشه‌ها چطور؟ آیا این روش نمونه‌گیری احتمالی است؟
- ۳ روش‌های نمونه‌گیری احتمالی چه مزیتی بر نمونه‌گیری‌های غیر احتمالی دارند؟
- ۴ برای هریک از روش‌های نمونه‌گیری احتمالی دو مثال واقعی بیاورید.
- ۵ اگر اندازه جامعه بزرگ باشد، نمونه‌گیری با جای‌گذاری و بدون جای‌گذاری تقریباً مثل هم هستند. در این صورت، آیا می‌توانید راه حل کلی برای انتخاب تصادفی n نمونه از یک فهرست N تایی ارائه کنید؟
- ۶ آیا احتمال انتخاب واحدهای آماری در نمونه‌گیری طبقه‌ای برابر است؟ در هر طبقه چطور؟
- ۷ فرق بین داده و متغیر چیست؟
- ۸ فرق بین آماره با پارامتر چیست؟
- ۹ در یک جامعه آماری، آیا ممکن است که یک پارامتر تغییر کند؟ اگر سه نمونه با اندازه یکسان از یک جامعه داشته باشیم، می‌توان سه مقدار متفاوت از یک آماره به دست آورد؟
- ۱۰ در یک مطالعه از ۱۲۶۱ مشتری غذاهای گیاهی، سؤال شده است که برای کدام وعده غذایی (ناهار یا شام) سفارش داده‌اند؟
- الف) متغیر را مشخص کنید. این متغیر کمی است یا کیفی؟
- ب) کدام روش گردآوری داده‌ها برای مطالعه مناسب است؟
- پ) جامعه آماری در اینجا چیست؟ در این مطالعه پارامتر و آماره چه چیزی می‌توانند باشند؟

حل تمرین صفحه ی ۱۱۶ (آمار و احتمال)

۱: اگر تعداد اعضای جامعه برابر N در نظر بگیریم، احتمال انتخاب اولین فرد $\frac{1}{N}$ خواهد بود. اگر مسئله با جایگذاری انجام شود، احتمال انتخاب دومین فرد باز $\frac{1}{N}$ خواهد شد. اما اگر بدون جایگذاری باشد، $\frac{1}{N-1}$ می شود.

۲: خیر، احتمال انتخاب خوشه ها برابر است. زیرا در مرحله ی اول واحد های آماری همان خوشه ها هستند. چون احتمال انتخاب هر واحد آماری درون هر یک از خوشه ها معلوم و برابر است. لذا می تواند نمونه گیری احتمالی باشد.

۳: در نمونه گیری احتمالی، شانس انتخاب واحدهای آماری معلوم و گاهی یکسان است، در حالی که در نمونه گیری غیراحتمالی چنین ویژگی وجود ندارد و ممکن است عضوی از جامعه برای انتخاب شدن در نمونه شانس نداشته باشد یا اینکه شانس اعضا یکسان نباشد.

۴:

روش تصادفی ساده	
مثال ۱	انتخاب چند نفر کارمند از بین تمامی کارمندان یک اداره ی کوچک
مثال ۲	انتخاب ۳ دبیرستان از بین دبیرستان های پسرانه ی یک شهر
روش خوشه ای	
مثال ۱	انتخاب چند نفر از بین محلات یک شهر به جهت بررسی فشار خون آنها
مثال ۲	اگر بخواهیم از بین دانش آموزان دبستانهای دخترانه تهران نمونه ای برای بررسی بهداشت دهان و دندان انتخاب کنیم.
روش طبقه ای	
مثال ۱	انتخاب چند دانش آموزان از بین دانش آموزان پایه ی دهم منطقه (طبقات : دختر و پسر)
مثال ۲	از یک جامعه آماری ۱۰۰۰۰ نفری که ۱۵ درصد آن دانشجو ، ۲۰ درصد کارمند اداری ، ۳۰ درصد کارگر و ۳۵ درصد کشاورز هستند می خواهیم ۴۰۰ نفر نمونه انتخاب کنیم.
روش سیستماتیک	
مثال ۱	تحويل پرسش نامه به مراجعه کنندگان یک رستوران در و نظرخواهی در مورد اینکه چگونه نام و آدرس رستوران را فهمیده اند.
مثال ۲	نمونه گیری به جهت تعیین میزان رضایت بیماران مراجعه کننده به بیمارستان ، هنگام ترخیص آنها

۵: استفاده از اعداد تصادفی که به کمک ماشین حساب، یا جداول اعداد تصادفی صورت می گیرد.

۶: احتمال انتخاب واحد های آماری در هر طبقه یکسان است.

۷: داده واقعی است در باره ی یک موضوع می باشد. این موضوع ممکن است از هر فرد جامعه یا نمونه

تغییر کند، که به آن متغیر می نماند. مثلاً نمره ی درس ریاضی هر یک از دانش آموزان (۱۵ و ۱۷ و ۱۸ و ...) داده است. ولی چون این نمره از فردی به فرد دیگر تغییر می کند. یک متغیر می باشد.

۸:

پارامتر، یک مشخصه ی عددی است که توصیف کننده ی جنبه ای خاص از جامعه است. مانند نسبت معلّمان

فوق لیسانس به کل معلمان یک منطقه ی آموزشی

نمونه، یک مشخصه ی عددی است که توصیف کننده ی جنبه ای خاص از نمونه است. نسبت معلّمان فوق

لیسانس به کل معلمان مدرسه ای که به تصادف انتخاب شده است.

۹: خیر، پارامتر جامعه ثابت است. اما آماره از یک نمونه به نمونه ای دیگر متفاوت است.

۱۰:

الف: وعده ی غذایی (ناهار یا شام)، کیفی

ب: پرسش شفاهی، البته می توان پیام دریافت کرد.

پ: تمامی مشتریان غذاهای گیاهی، مُد جامعه پارامتر و مُد نمونه آماره است.

نَیبه کُنده:

گروه ریاضی دوره ی دوم متوسطه و انجمن معلمان ریاضی، استان خوزستان

khuzmath1394@chmail.ir

۱۱۶,۲۵۵

۱۱ کدام روش گردآوری داده‌ها برای موارد زیر مناسب است؟ یک دلیل برای انتخاب خود ذکر کنید.

- میزان رضایت مشتریان بانک از نحوه برخورد و رسیدگی به درخواست‌های آنها.
- سن همه دانش‌آموزان مدرسه بر حسب ماه در پایه دهم.
- تعداد سرنشینان خودروهای سواری در یکی از محورهای خروجی شهر.

۱۲ فرض کنید جامعه‌ای از $N = 100$ عضو تشکیل شده و می‌خواهیم نمونه‌ای به اندازه $n = 20$ از آن انتخاب کنیم. در هر یک از حالت‌های زیر احتمال انتخاب هر عضو جامعه به‌عنوان نمونه چقدر است؟ نام هر روش نمونه‌گیری را بگویید.

(الف) اگر جامعه به دو قسمت ۵۰ تایی تقسیم شود و بخواهیم از هر قسمت نمونه تصادفی ۱۰ تایی انتخاب کنیم.

(ب) اگر جامعه به تصادف به ۱۰ قسمت مساوی تقسیم شود و دو قسمت را به‌عنوان نمونه انتخاب کنیم.

(پ) اگر جامعه به تصادف به ۲۰ قسمت مساوی تقسیم شود، و از قسمت اول یک عضو به تصادف انتخاب شود. فرض کنید عضو انتخابی دومین عضو باشد و از قسمت‌های بعدی نیز دومین عضو انتخاب شود.

۱۳ دلایل ارببی در نمونه‌گیری‌های زیر را ذکر کنید. کدام روش گردآوری داده‌ها برای آنها مناسب‌تر است؟

- (الف) نمونه‌گیری راحت: افراد در دسترس را به‌عنوان نمونه انتخاب می‌کنیم.
- (ب) نمونه غیر تصادفی: عامل شانس در انتخاب نمونه نقشی ندارد.
- (پ) نمونه‌گیری ایمیلی (رایانامه‌ای): پرسش‌نامه‌ای به ایمیل‌های انتخاب شده ارسال می‌شود.
- (ت) نمونه‌گیری تلفنی: از دفترچه راهنمای تلفن تعدادی شماره به تصادف انتخاب می‌شود.
- (ث) برخی از اعضای انتخاب شده در نمونه حاضر به پاسخگویی نمی‌باشند.

۱۴ نوع روش نمونه‌گیری مناسب‌تر را انتخاب کنید.

- (الف) شرکت واردکننده خودروهای سنگین برای بررسی عملکرد سامانه ترمز آنها می‌خواهد ده درصد از خودروهایی را که به‌مرور زمان وارد کشور می‌شوند بازرسی کند.
- (ب) مدیر مدرسه ۶۰ نفری می‌خواهد نظر دانش‌آموزان را برای تغییر ساعت تعطیلی مدرسه براساس یک نمونه ۱۲ تایی بداند.
- (پ) در قسمت قبل اگر مدرسه، شش پایه داشته باشد و ما حدس بزنیم که نظر ۶ پایه باهم تفاوت دارد (با فرض برابر بودن تعداد دانش‌آموزان در پایه‌های مختلف).

۱۵ چگونه از ارقام ۰ تا ۹ عدد تصادفی انتخاب می‌کنید؟ آیا با روش پیشنهادی شما می‌توان عدد تصادفی بین اعداد ۰ تا ۹۹ انتخاب کرد؟ آیا امکان توسعه روش پیشنهادی شما به انتخاب تصادفی از فهرستی ۱۰۰۰ تایی امکان‌پذیر است؟

۱۱:

الف : پرسش نامه. چون مشتریان در دسترس هستند. راحت بودن در پاسخ آنها، راحت بودن و سریع بودن

ب : دادگان : مراجعه به پرونده های دانش آموزان ، تنها روش راحت و سریع

ج : مشاهده ، با گذر خودروها از محوار های خروجی به آنها توجه می شود. عدم لزوم توقف خودروها

۱۲:

نام روش	احتمال
الف	طبقه ای $\frac{1}{50} = \frac{1}{5}$
ب	خوشه ای $\frac{1}{10}$
پ	سیستماتیک احتمال اولی $\frac{1}{5}$ و بقیه ۱ است.

۱۳:

نویه کننده:

گروه ریاضی دوره ی دوم متوسطه و انجمن معلمان ریاضی ، استان خوزستان

khuzmath1394@chmail.ir

الف : افراد شانس یکسان برای انتخاب ندارند

ب : تمام صفات جامعه معلوم نمی شود.

پ : تمان افراد ممکن است ایمیل نداشته باشند. لذا شانس انتخاب ندارند.

ت : افرادی که تلفن آنها را در دفترچه نیست، شانس انتخاب ندارند.

ث : تمام صفات جامعه معلوم نمی شود.

۱۴:

الف : سیستماتیک ، هر چند اتومبیل با فواصل معین انتخاب می شوند.

ب : تصادفی ساده: چون لیست دانش آموزان موجود است. می توان قرئه کشی کرد. (البته سیستماتیک هم

قابل انجام است.)

پ : طبقه ای، چون می خواهیم از تمام طبقات در نمونه انتخاب شود.

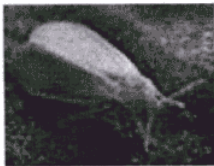
۱۵: با انتخاب یک عدد تصادفی بین صفر و یک و ضرب آن در تعداد اعضای نمونه. بله ، کاملاً این روش

قابل تعمیم است.

درس ۱

گردآوری داده‌ها

فعالیت



می‌خواهیم برخی از ویژگی‌های مگس‌های سفید مزاحم در شهر تهران را بررسی کنیم. آیا برای انجام این کار می‌توانیم ویژگی‌های همه مگس‌های سفید را اندازه‌گیری کنیم؟ آیا همه آنها در دسترس‌اند؟ آیا زمان و هزینه لازم برای این کار در اختیار داریم؟ **خیر**

واحد آماری^۱ به هر یک از افراد یا اشیا می‌گویند که داده‌های مربوط به آنها در یک بررسی آماری گردآوری می‌شود.

مجموعه کل واحدهای آماری را جامعه آماری^۲ می‌نامند.

هر زیرمجموعه از جامعه آماری را که با روش مشخصی انتخاب شده باشد، یک نمونه می‌نامند.

نمونه‌گیری^۳، فرایند انتخاب نمونه‌ای از یک جامعه، به منظور تعمیم اطلاعات آن به جامعه است.



بیشتر مطالعات آماری بر روی بخشی از جامعه است. رابطه بین جامعه و بخشی از آن که نمونه نامیده می‌شود، در شکل نشان داده شده است.

کار در کلاس

در فعالیت قبل هر مگس سفید **یک واحد آماری** است. همه مگس‌های سفید، که کل واحدهای آماری هستند، **جامعه آماری** را تشکیل می‌دهند. اگر سن همه مگس‌های سفید را در اختیار داشته باشیم، داده‌های **آماری جامعه** را داریم. ۱۰۰ مگس سفید معرف یک **نمونه** است.

۱- Statistical Unit

۲- Statistical Population

۳- Sampling

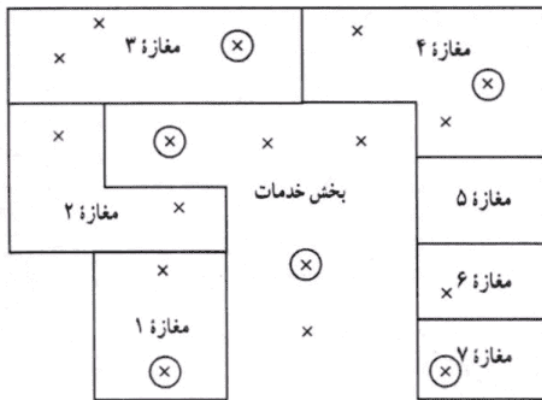
نپیه کننده:

گروه ریاضی دوره ی دوم متوسطه و انجمن مطلقان ریاضی، استان خوزستان

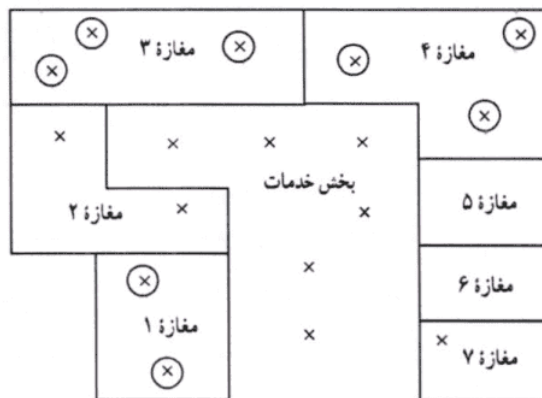
نمونه‌گیری تصادفی ساده^۱ نوعی روش نمونه‌گیری که در آن همه واحدهای آماری برای انتخاب شدن در نمونه، احتمال یکسان دارند.

فعالیت

۱ می‌خواهیم متوسط درآمد کارکنان یک مجتمع تجاری را محاسبه کنیم. اگر این مجتمع از ۷ مغازه و یک بخش خدمات تشکیل شده باشد، که روی هم ۱۷ کارکن دارند، چگونه از بین ۱۷ نفر، ۶ نفر را به تصادف انتخاب می‌کنید؟ یک راه ساده برای انجام این کار نوشتن اسامی کارکنان یا شماره کارمندی آنها روی ۱۷ برگه کوچک و انتخاب تصادفی ۶ تا از آنهاست. آیا این روش نمونه‌گیری، نمونه‌گیری تصادفی ساده است؟ آیا همه واحدهای جامعه احتمال برابری برای انتخاب دارند؟ **بله - بله**



در شکل روبه‌رو نقشه‌ای از مجتمع تجاری ترسیم شده که کارکنان با X و دور انتخاب‌شدگان یک دایره رسم شده است. انجام نمونه‌گیری تصادفی ساده در عمل با دشواری‌هایی همراه است. اگر اندازه جامعه بزرگ باشد، یعنی تعداد واحدهای آماری زیاد باشند، دسترسی به فهرستی از اعضای جامعه و دسترسی به اعضای انتخابی، دشوار و ممکن است هزینه‌بر باشد.



۲ هر یک از ۷ مغازه و بخش خدمات را به‌صورت یک گروه فرض می‌کنیم. حال از بین ۸ گروه در نظر گرفته شده، سه تا از آنها را به تصادف انتخاب می‌کنیم و در هر یک سرشماری انجام می‌دهیم. آیا این روش نمونه‌گیری سریع‌تر است؟ **بلی**

نویسنده:

گروه ریاضی دوره ی دوم متوسطه و انجمن طلمان ریاضی، استان خوزستان

khuzmath1394@chmail.ir

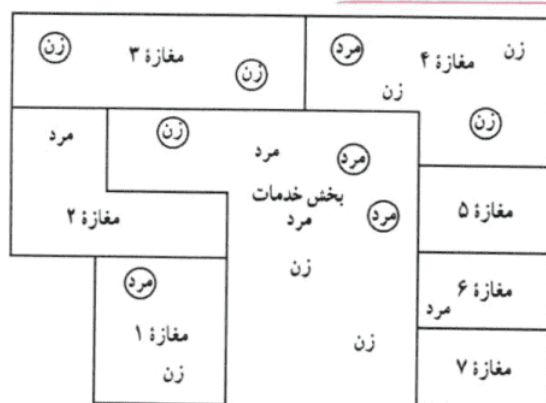
نمونه‌گیری خوشه‌ای^۲: نمونه‌گیری که در آن، واحدهای نمونه‌گیری اولیه در جامعه، گروه‌ها یا خوشه‌ها باشند. سپس همه واحدهای آماری خوشه‌های انتخاب شده را به عنوان نمونه در نظر می‌گیریم.

۱- Simple Random Sampling

۲- Cluster Sampling

سؤال : می‌خواهیم میانگین نمرات ریاضی دانش آموزان شهر تهران را محاسبه کنیم. اگر فهرست همه دانش‌آموزان را نداشته باشیم، اما فهرست مدارس موجود باشد، نمونه‌گیری خوشه‌ای، راه مناسبی برای گردآوری داده‌هاست. اگر بودجه کافی یا زمان لازم برای نمونه‌گیری تصادفی ساده نداشته باشیم آیا این روش مقرون به صرفه است؟ **بلی**

۲ اگر بخواهیم یک نمونه ۸ تایی شامل دقیقاً ۴ مرد و ۴ زن از مجتمع تجاری بگیریم، چگونه این کار را انجام می‌دهیم؟ زمانی که جامعه به دو یا چند بخش تقسیم می‌شود که عضو مشترکی ندارند، می‌توان از هر بخش جداگانه نمونه‌گیری کرد. این کار با افزایش هزینه یا زمان همراه است، ولی انتظار داریم که **دقت** را نیز افزایش دهد.



در واقع داده‌ها به دو طبقه مجزا تقسیم شده‌اند که طبقات از نظر ویژگی مورد بررسی همگن هستند. در صورتی که در نمونه‌گیری خوشه‌ای، درون خوشه‌ها هرچه ویژگی مورد بررسی تفاوت بیشتری داشته باشند بهتر است.

تپیه کننده:

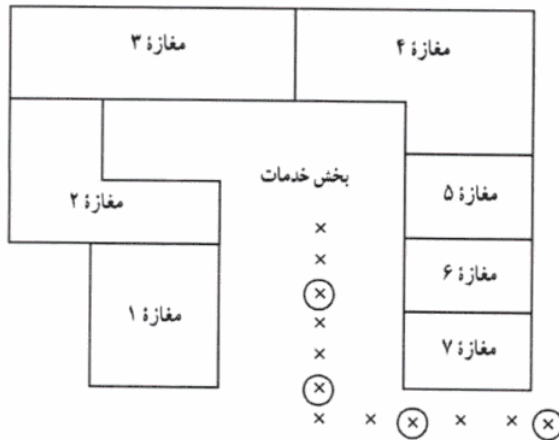
گروه ریاضی دوره دوم متوسطه و انجمن معلمان ریاضی، استان خوزستان

khuzmath1394@chmail.ir

نمونه‌گیری طبقه‌ای ۱: روش نمونه‌گیری که در آن با طبقه‌بندی جامعه به زیرجامعه‌های مجزا یک نمونه تصادفی ساده از هر طبقه انتخاب می‌شود.

علاقه‌مند به نمونه‌گیری از نمرات درس ریاضی دانش‌آموزان استان تهران هستیم. اگر فهرست همه دانش‌آموزان را در اختیار داشته باشیم، می‌توانیم از نمونه‌گیری تصادفی ساده استفاده کنیم؛ ولی این روش نمونه‌گیری هیچ تضمینی ندارد که دانش‌آموزان از تمامی شهرهای استان در نمونه حضور داشته باشند. در صورتی که اگر از هر شهر متناسب با تعداد دانش‌آموزان آن شهر نمونه‌گیری تصادفی ساده انجام دهیم، مشکل قبلی رفع می‌شود. به عبارت دیگر از نمونه‌گیری طبقه‌ای استفاده می‌کنیم. حال فرض کنید فقط فهرست مدارس را داشته باشیم. چه روش نمونه‌گیری را پیشنهاد می‌کنید؟ (راهنمایی: شما می‌توانید از دو روش نمونه‌گیری پشت سر هم استفاده کنید.) **نمونه‌گیری چند مرحله‌ای (انجام شود در فصل بعد)**

۴ فرض کنید در مجتمع، ۱۲ نفر حضور دارند. صبر می‌کنیم که مجتمع تجاری تعطیل شود و هنگام خروج کارکنان می‌خواهیم نمونه ۴ نفری انتخاب کنیم. برای این منظور، همانند شکل صفحه بعد عمل کرده‌ایم. ابتدا از ۳ نفر یکی را به تصادف انتخاب می‌کنیم. در این شکل، نفر اول انتخاب شده است. حال با همین رویه برای سه نفر بعد هم، نفر اول را انتخاب می‌کنیم و ادامه می‌دهیم. این روش نمونه‌گیری شباهت بیشتری به کدام یک از روش‌های نمونه‌گیری قبلی دارد؟ خوشه‌ای یا طبقه‌ای؟ این کار باعث چه نوع صرفه‌جویی می‌شود؟ **زمان و هزینه**



به نظر شما این نوع نمونه‌گیری در کدام یک از مثال‌های زیر امکان دارد:

- گردآوری اطلاعات از مبدأ و مقصد مسافران در خروجی - ورودی یک شهر
- کنترل کیفیت یک خط تولید ✓
- انتخاب نمونه از ماهی‌های یک حوضچه
- زمانی که فهرستی از واحدهای جامعه وجود نداشته باشد.
- فهرست واحدهای آماری ترتیب تصادفی داشته باشند. ✓

نمونه‌گیری سیستماتیک^۱ یا سامانمند، نوعی نمونه‌گیری طبقه‌ای است که در آن اندازه طبقات باهم برابر است. فقط از طبقه اول، واحد آماری به تصادف انتخاب می‌شود و با همان رویه از طبقات دیگر، این کار انجام می‌گیرد.

آیا اعضای جامعه برای انتخاب شدن در نمونه‌گیری سامانمند شانس برابر دارند؟ چرا؟

کار در کلاس

جدول زیر را کامل کنید.

روش نمونه‌گیری	مزیت	محدودیت
تصادفی ساده	همه اعضا شانس مساوی برای انتخاب دارند.	هزینه زیاد (بویژه در جامعه‌ها بزرگ)
خوشه‌ای	هزینه کم، سریع	نداشتن شانس مساوی برای انتخاب
طبقه‌ای	اطمینان از انتخاب از همه طبقات	زمان بزرگ و هزینه بزرگ / عدم مساوی بودن شانس در انتخاب اعضا
سامانمند	شانس انتخاب مساوی / هم‌جوشی در هر مرحله	اگر فهرست اعضا موجود نباشد، قابل انجام نیست.

فعالیت

از مگس‌های سفید با چه روشی می‌توان نمونه‌گیری کرد؟ فهرستی از آنها نداریم، تعداد آنها را هم نمی‌دانیم. می‌توان چند منطقه از تهران را به تصادف انتخاب کرد و در هر منطقه نمونه در دسترس را انتخاب و بررسی کنیم. آیا این روش نمونه‌گیری به تمامی واحدهای جامعه شانس انتخاب می‌دهد؟

نمونه‌گیری احتمالی: نمونه‌گیری است که همه واحدهای آماری احتمالی معلوم برای انتخاب در نمونه داشته باشند. و از روشی تصادفی برای انتخاب واحدهای نمونه استفاده شود.

^۱ Systematic

نمونه‌گیری‌های چهار فعالیت قبل، همگی احتمالی‌اند. در کدام یک همه واحدهای آماری احتمال برابری برای انتخاب دارند.

نمونه‌گیری ساده / سیستماتیک

کار در کلاس

راه حلی ارائه کنید که نمونه‌گیری‌های غیر احتمالی زیر را احتمالی می‌کند، هر چند که به صورت غیر واقعی باشد.

مثال	نمونه‌گیری غیر احتمالی	نمونه‌گیری احتمالی
نمونه‌گیری از یک قفس بزرگ خرگوش‌های یک آزمایشگاه	بدون برنامه‌ریزی خرگوش‌هایی را برمی‌دارد که دستش به آنها می‌خورد.	از چند طرف قفس نمونه‌برداری می‌شود.
در مطالعاتی که در آنها فرایند سنجش برای شخصی که سنجیده می‌شود ناخوشایند یا دردسرافرین است.	داوطلبانی که حاضر به پاسخ به سؤالات شما در یک نظرسنجی می‌شوند.	فهرست تمام اشخاصی دارد می‌شود و سپس نمونه‌گیری ساده انجام می‌گیرد.
نمونه‌گیری از زغال‌سنگ‌های یک واگن	نمونه در دسترس انتخاب می‌شود.	ابتدا چند واگن آبی را می‌شمارد و از واگن‌های آبی به عنوان نمونه‌گیری می‌شود.

فعالیت

شاید در نگاه اول این‌طور به نظر برسد که انجام نمونه‌گیری تصادفی ساده کاری آسان است، در حالی که در دنیای واقعی، گاهی چنین نیست. روش‌های نمونه‌گیری که به ذهن می‌رسند، گاهی مشکلات و ایرادهایی دارند که در نگاه اول دیده نمی‌شوند و استفاده از آنها ما را به نتایجی بسیار دور از واقعیت می‌رساند. این موضوع را با چند مثال بهتر متوجه خواهید شد. فرض کنید آمارگیری می‌خواهد بداند در یک شهر خانواده‌ها چند نفره‌اند. او برای این کار صد نفر را به تصادف انتخاب می‌کند و از آنها می‌پرسد: «خانواده شما چند نفر است؟»

آیا این روش برای نمونه‌گیری درست است؟ جواب منفی است! دلیل آن هم این است که واحدهای آماری مورد نظر در این مسئله خانواده‌ها هستند نه افراد. آیا خانواده‌های مختلف احتمال حضور برابر در این نمونه‌گیری را دارند؟ واضح است که احتمال حضور هر خانواده متناسب با تعداد اعضای آن است و مثلاً احتمال حضور یک خانواده شش نفره دو برابر احتمال حضور یک خانواده سه نفره است و این، یعنی شرایط نمونه‌گیری ساده برقرار نیست.

نتیجه چنین ایرادی در نمونه‌گیری این است که هر چه تعداد نمونه‌ها را افزایش دهیم، نتایج به مقداری اشتباه نزدیک‌تر می‌شود. مثلاً فرض کنید آمار واقعی تعداد افراد خانواده‌ها چنین باشد:

تعداد افراد	۱	۲	۳	۴	۵ و بیشتر
درصد	۸/۵	۲۰/۷	۲۸/۵	۲۷/۶	۱۴/۷

در این جامعه تعداد خانواده‌های دو نفره تقریباً $\frac{20.7}{100} = 1/4$ برابر تعداد خانواده‌های پنج نفره (و بیشتر) است، ولی با آمارگیری نادرستی که توضیح داده شد به نتیجه دیگری خواهیم رسید: احتمال حضور یک خانواده پنج نفره (و بیشتر) در

نمونه‌ها بیشتر از ۲۵۰۰... برابر احتمال حضور یک خانواده دو نفره است و لذا عددی که در روش نادرست آماری به دست می‌آید کمتر از ۵۶٪ = $\frac{1/4}{3/5}$ است. نتیجه اینکه هر چند واقعیت این است که نسبت خانواده‌های دو نفره بسیار بیشتر از خانواده‌های ۵ نفره (و بیشتر) است، ولی ما با نمونه‌گیری اشتباه به نتیجه‌ای بسیار متفاوت می‌رسیم. برای برطرف کردن این مشکل راه‌های مختلفی دارد. مثلاً اینکه فقط از سرپرست خانواده‌ها در مورد تعداد اعضای خانواده‌ها پرسیم. (جدول صفحه قبل، برگرفته از آمار واقعی کشور در سرشماری سال ۱۳۹۵ است.)

کار در کلاس

فرض کنید در شهری جمعیت کلاس‌های پایه ششم دبستان به شکل زیر باشد:

تعداد دانش‌آموز	۲۵	۲۶	۲۷	۲۸	۲۹	۳۰	۳۱	۳۲	۳۳	۳۴	۳۵
تعداد کلاس	۷	۹	۱۰	۱۱	۷	۱۱	۸	۸	۷	۵	۴

الف) چه تعداد کلاس پایه ششم در این شهر وجود دارد؟ تعداد دانش‌آموزان پایه ششم چند تاست؟
 ب) چه درصدی از کلاس‌های پایه ششم بیشتر از ۳۰ دانش‌آموز دارند؟
 پ) اگر به تصادف یک دانش‌آموز ششم دبستانی را انتخاب کنیم، احتمال اینکه کلاسی که در آن درس می‌خواند بیشتر از ۳۰ دانش‌آموز داشته باشد، چقدر است؟

ت) فرض کنید فردی برای اینکه بفهمد کلاس‌های ششم دبستان چند نفری‌اند، تعداد زیادی دانش‌آموز ششم دبستانی را به تصادف انتخاب کند و از آنها بپرسد «کلاس شما چند نفره است؟» این کار چه ایرادی دارد؟
 ث) اگر با روش قبل درصد کلاس‌های با بیش از ۳۰ دانش‌آموز را محاسبه کنیم، نتیجه از جواب واقعی چند درصد فاصله خواهد داشت؟

مثال: فرض کنید می‌خواهیم میزان آلودگی خودروهای در حال تردد در شهری را بررسی کنیم. برای این کار چگونه باید نمونه‌گیری کنیم؟

اگر نمونه‌گیری را در تعمیرگاه‌ها انجام دهیم، هرچند هر خودرویی ممکن است گاهی سر از تعمیرگاه درآورد، ولی این نمونه‌گیری، تصادفی ساده نیست؛ زیرا احتمال اینکه یک خودرو در نمونه ما باشد، متناسب با ساعتی است که در تعمیرگاه بوده است و لذا درصد خودروی‌های آلاینده بسیار بیشتر از واقعیت نشان داده خواهد شد.
 اگر نمونه‌گیری را در خیابان انجام دهیم، مشکل آن کمتر است، ولی باز هم نمونه‌گیری ما مشکل دارد؛ زیرا خودروهایی که از آنها بیشتر استفاده می‌شود، احتمال بیشتری دارد که به عنوان نمونه انتخاب شوند.

کار در کلاس

فرض کنید بخواهیم میزان مطالعه غیردرسی دانش‌آموزان یک مدرسه را بررسی کنیم. برای این کار این سؤال را طراحی کرده‌ایم و می‌خواهیم از یک نمونه ۳۰ تایی آن را بررسییم:

۱- سرشماری (Census) فرایند انتخاب همه واحدهای آماری جامعه است.

«در یک سال گذشته چند کتاب غیردرسی خوانده‌اید؟»

روش‌های زیر را نقد کنید :

- الف) پرسیدن سؤال از تعدادی از دانش‌آموزانی که در کتابخانه هستند. *نمونه‌گیری ساده تصادفی*
- ب) گذاشتن تعدادی پرسش‌نامه در محل رفت و آمد دانش‌آموزان. *نمونه‌گیری تصادفی*
- پ) پرسیدن از دانش‌آموزانی که صبح وارد مدرسه می‌شوند و مایل اند به سؤال مذکور جواب دهند. *نمونه‌گیری تصادفی*
- شما چه روشی را پیشنهاد می‌کنید که به نمونه‌گیری تصادفی ساده نزدیک‌تر باشد؟
- نمونه‌گیری طبقه‌ای «طبقه‌بندی»*

کار در کلاس

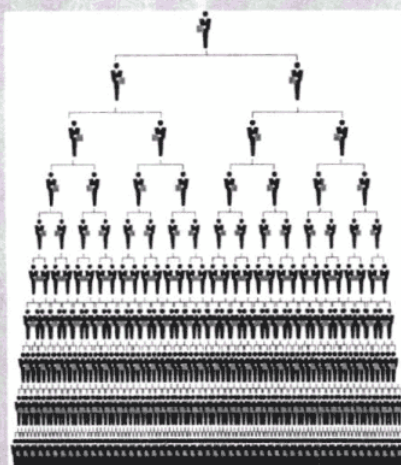
از جمله مسائلی که مردم در مورد آنها به نظرسنجی‌ها علاقه زیادی نشان می‌دهند، انتخابات است. با این وجود، گاهی مردم گاهی گروه‌های سیاسی از روش‌هایی برای کشف نظر مردم استفاده می‌کنند که آنها را گمراه می‌کند.

در این مورد روش‌های زیر را نقد کنید :

- الف) پرسیدن نظر دوستان و اطرافیان. *نمونه‌گیری اطرافیان «طبقه‌بندی»*
- ب) طراحی یک نظرسنجی در وبگاهی برطرف‌دار و لحاظ کردن ساز و کاری که از یک آدرس بیش از یک بار رأی گرفته نشود. *افزودن به جهت طرف‌دار بر روی سایت و اداری «پس‌گویی» شده اند.*
- در تمام مثال‌هایی که ذکر شد، روش نمونه‌گیری به شکلی بود که داده‌هایی به سمتی انحراف پیدا می‌کردند و لذا افزایش تعداد نمونه‌ها نیز به کاهش این انحراف کمک نمی‌کرد. در علم آمار اصطلاحی خاص برای این مشکل وجود دارد :

اگر یک روش نمونه‌گیری از نمونه‌گیری ایده‌آل فاصله بگیرد و به سمتی خاص انحراف پیدا کند می‌گویند آن روش نمونه‌گیری اریب است. لذا آمارشناسان تلاش می‌کنند تا با شناسایی منابع تولید اریبی، نمونه‌گیری‌ها را تا جایی که می‌توانند تا اریب کنند.

خواندنی



شرکت‌های هرمی که سال‌ها در دنیا کلاهبرداری کرده‌اند از اواخر دهه ۷۰ شمسی (حدود سال ۱۳۷۸) وارد ایران شدند. یک شرکت هرمی از مشتریان خود می‌خواهد که افراد جدیدی را به عضویت شرکت درآورند و به آنها وعده می‌دهد که اگر زیرشاخه تولیدشده به اندازه کافی بزرگ شود، جوایزی به او تعلق خواهد گرفت. معمولاً اعضای این شرکت‌ها برای اینکه بفهمند شانس رسیدن به جایزه‌ها چقدر است به اطرافیان خود و به خصوص بالاسری‌های خود نگاه می‌کنند. مثال ساده صفحه بعد نشان می‌دهد که این نوع نمونه‌گیری تا چه حد گمراه کننده است :

شرکتی هر می از مشتریان خود می خواهد که بعد از خرید یک میلیون تومان کالا، دو نفر را به عنوان زیرشاخه خود معرفی کنند تا آنها هم همین کار را انجام دهند. شرکت به هر کس که زیر مجموعه هایش تا سه ردیف رشد کند جایزه می دهد. عموم مشتریان تصور می کنند تا قبل از یک ماه به جایزه خواهند رسید.

درخت دودویی صفحه قبل نشان می دهد که مجموعه مشتریان این شرکت چگونه رشد می کند؛ اگر این درخت n ردیف داشته باشد، تعداد اعضای آن برابر است با:

$$1+2+4+\dots+2^{n-1}=2^n-1$$

همه این اعضا، مگر آنهایی که در سه ردیف آخر قرار دارند، از شرکت جایزه گرفته اند. تعداد آنها برابر است با:

$$1+2+4+\dots+2^{n-2}=2^{n-1}-1$$

پس نسبت کسانی که جایزه گرفته اند، به کل اعضای برابر کسر زیر است:

$$\frac{2^{n-1}-1}{2^n-1} \approx \frac{2^{n-1}}{2^n} = \frac{1}{2}$$

یعنی از هر ۸ نفر فقط یکی موفق به گرفتن جایزه شده است. اکنون از زاویه دید کسی به ماجرا نگاه کنید که در ردیف آخر این مجموعه است و به بالاسری های خود نگاه می کند. خودش و افراد بالای سرش n نفر هستند و غیر از سه نفر پایینی همه جایزه گرفته اند. لذا این عضو، نسبت جایزه گرفته ها را به کل اعضا به این شکل برآورد می کند:

$$\frac{n-3}{n} = 1 - \frac{3}{n}$$

و این عبارت وقتی n بزرگ باشد، تقریباً برابر یک است. مثلاً به ازای $n=12$ که تعداد اعضای شرکت

$$2^{12}-1=4095$$

نفر است. یک عضو تازه وارد تصور می کند که نسبت کسانی که جایزه گرفته اند برابر با:

$$1 - \frac{3}{12} = 0.75$$

است؛ یعنی ۷۵ درصد. در حالی که مقدار واقعی کمتر از ۱۲/۵ درصد است!

اگر تعداد ردیف ها به $n=18$ برسد، تعداد اعضای شرکت $2^{18}-1=262143$ نفر می شود. در این صورت، عضو تازه واردی که فقط به خودش و بالاسری های خودش نگاه می کند، تصور می کند که نسبت کسانی که جایزه گرفته اند برابر $1 - \frac{3}{18} = 0.83$ یعنی ۸۳ درصد است؛ در حالی که در واقع این نسبت تقریباً همان ۱۲/۵ درصد است!

این شرکت ها، اگر از طریق قانون، متوقف نشوند، آن قدر رشد می کنند که دیگر ادامه کار برای اعضا به صرفه نیست و در آن مقطع مشخص می شود که مثلاً از حدود یک میلیون عضو، ۸۷۵ هزار نفر ضرر کرده اند؛ از آن بدتر وضعیت کسانی است که جایزه ای از شرکت گرفته اند، ولی زیرشاخه های آنها که عموماً دوست و فامیلشان هستند ضرر کرده اند؛ این افراد آبرو و اعتبار خود را از دست داده اند.

فعالیت

می‌خواهیم طول قد دانش‌آموزان یک مدرسه را گردآوری کنیم. برای این منظور چه راهی پیشنهاد می‌کنید؟

آمارگیری: گردآوری داده‌ها به یکی از روش‌های ممکن
آمارگیر: کسی که آمارگیری را انجام می‌دهد.

اگر قرار شد آمارگیر باشیم، می‌توانیم جدولی به صورت زیر تکمیل کنیم.
 مثالی از جدول طراحی شده برای ثبت داده‌ها

اندازه طول قد	چوب خط برای شمارش	تعداد دانش‌آموزان
کوتاه‌تر از ۱۴۰ سانتی‌متر		
۱۴۰-۱۴۹ سانتی‌متر		
۱۵۰-۱۵۹ سانتی‌متر		
۱۶۰-۱۶۹ سانتی‌متر		
۱۷۰ سانتی‌متر یا بلندتر		

آمارگیری زحمت زیادی برای آمارگیر دارد. آیا راه‌حل ساده‌تری برای انجام آن دارید؟ یکی از مرسوم‌ترین روش‌های آمارگیری، استفاده از پرسش‌نامه است. پرسش‌نامه شبیه همان جدولی است که هنگام ثبت‌نام در مدرسه، آن را تکمیل کرده‌اید. واحدهای جامعه یا نمونه می‌توانند پرسش‌نامه تکمیل کنند.

مثالی از پرسش‌نامه طراحی شده

سلام.

می‌خواهیم طول قد دانش‌آموزان مدرسه را آمارگیری کنیم.
 لطفاً یکی از گزینه‌ها را انتخاب کنید.
 طول قد شما چقدر است؟

☐ کوتاه‌تر از ۱۴۰ سانتی‌متر

☐ ۱۴۰-۱۴۹ سانتی‌متر

☐ ۱۵۰-۱۵۹ سانتی‌متر

☐ ۱۶۰-۱۶۹ سانتی‌متر

☐ ۱۷۰ سانتی‌متر یا بلندتر

- ۱ چه راه دیگری برای آمارگیری طول قد دانش آموزان یک مدرسه پیشنهاد می کنید؟ *نمونه پرسش و پاسخ به علوم*
- ۲ فرض کنید زمان لازم را برای گردآوری همه داده های دانش آموزان در اختیار نداشته باشید. اگر بخواهیم نمونه ای را انتخاب و آمارگیری کنیم، چه راهی پیشنهاد می کنید که نمونه به صورت تصادفی انتخاب شود؟ *این پرسش آموزنی که هم تکمیلی و هم آسان باشد.*



- ۱ مشاهده: گردآوری داده ها بدون نیاز به فرد پاسخگو، مانند شمارش تعداد وسایل نقلیه عبوری از یک تقاطع در هر ساعت یا اندازه گیری وزن محصولات یک باغ میوه.
- ۲ پرسش نامه: مجموعه سؤالات از پیش تعیین شده که توسط تعدادی پاسخگو تکمیل می شود. این روش مرسوم ترین ابزار گرفتن اطلاعات از مردم است. مرکز آمار ایران هر ۱۰ سال یک بار با استفاده از پرسش نامه اطلاعات همه خانوارهای ساکن در ایران را گردآوری می کند. به این فرایند، سرشماری نفوس و مسکن می گوئیم.
- ۳ مصاحبه: معمولاً بین دو نفر صورت می گیرد: یکی مصاحبه گر (همان آمارگیر) و دیگری مصاحبه شونده، یا پاسخگو است. مثلاً اگر بخواهیم درباره مسائل فرهنگی کاهش شدآمد (ترافیک) پژوهش کنیم، مصاحبه از صاحب نظران راه حل مناسبی برای گردآوری داده ها است. از این روش، بیشتر زمانی استفاده می شود که آمارگیر از همه پاسخ های ممکن اطلاع کافی ندارد.
- ۴ دادگان: شامل مجموعه ای از اطلاعات ذخیره شده است. در بسیاری از موارد، داده ها را می توان از اطلاعاتی که قبلاً ذخیره شده است، به دست آورد. اگر قرار است تحقیقی در مورد نمره های دروس ریاضی استان ها انجام شود، اطلاعات ثبتی اداره کل آموزش و پرورش راه گشا خواهد بود. از سوی دیگر به دلیل تولید داده ها به صورت خودکار، در بسیاری از مؤسسات و سامانه ها، استفاده از این روش برای گردآوری داده ها به سرعت رواج یافته است.

- ۱- Observation
۲- Questionnaire
۳- Interview
۴- Database