www.Paadars.com

درس ۱ گردآوری داده‌ها

واحد آماری

به هر یک از افراد یا اشیایی که داده‌های مربوط به آن‌ها در یک بررسی آماری گردآوری می‌شود، واحد آماری می‌گویند.

جامعهٔ آماری

مجموعهٔ کل واحدهای آماری را جامعهٔ آماری می‌نامیم. در واقع جامعهٔ آماری مجموعهٔ کل افراد یا اشیایی هستند که می‌خواهیم موضوعی را دربارهٔ آن‌ها مطالعه کنیم.

سرشماری اگر تمام اعضای جامعه را مورد بررسی و مطالعه قرار دهیم، می‌گوییم «سرشماری» انجام داده‌ایم.

مشکلات

۱. در دسترس‌نبودن همهٔ افراد
۲. وقت‌گیربودن و هزینهٔ بالای دسترسی و بررسی همهٔ افراد
۳. امکان از بین رفتن قسمتی از جامعه در برخی مطالعات در زمان طولانی سرشماری

نمونه هر زیرمجموعه از جامعهٔ آماری که با روش مشخصی انتخاب‌شده باشد را یک «نمونه» می‌گوییم.

نمونه‌گیری فرایند انتخاب نمونه از یک جامعه به منظور «تعمیم اطلاعات آن به جامعه» را نمونه‌گیری می‌گوییم.

چون بررسی تمام اعضای جامعهٔ آماری هزینه‌بر و وقت‌گیر است، تنها بخشی از اعضای جامعه را به عنوان نمونه به نمایندگی از کل جامعه مورد بررسی قرار می‌دهیم. برای این‌که نمونه‌گیری دقیق‌تر باشد نمونه باید تصادفی انتخاب شود و همهٔ افراد جامعه شانس برابر برای انتخاب‌شدن در نمونه داشته باشند، هم‌چنین نمونه باید با حجم جامعه متناسب باشد و به اندازهٔ کافی بزرگ انتخاب شود تا خصوصیات اصلی جامعه را داشته باشد.

نکته شرط اصلی یک مطالعهٔ موفق، انتخاب یک روش درست برای نمونه‌گیری است تا نمونهٔ انتخاب‌شده بتواند نمایندهٔ خصوصیات جامعهٔ اولیه باشد. روش‌های نمونه‌گیری به صورت زیر دسته‌بندی می‌شوند:

- انواع نمونه‌گیری
  - ۱) نمونه‌گیری احتمالی (اتفاقی یا تصادفی)
    - آ) نمونه‌گیری تصادفی ساده
    - ب) نمونه‌گیری خوشه‌ای
    - پ) نمونه‌گیری طبقه‌ای
    - ت) نمونه‌گیری سیستماتیک یا سامانمند
  - ۲) نمونه‌گیری غیراحتمالی

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

آ) نمونه‌گیری تصادفی ساده

نوعی روش نمونه‌گیری که در آن همهٔ واحدهای آماری برای انتخاب شدن در نمونه، احتمال یکسان دارند.

در این روش، انتخاب افراد مستقل از یکدیگر هستند، یعنی انتخاب هر شخص در نمونه، هیچ تأثیری در انتخاب‌شدن یا انتخاب‌نشدن فرد دیگری در نمونه ندارد.

مراحل انجام نمونه‌گیری تصادفی:

۱ مشخص‌کردن جامعهٔ هدف که می‌خواهیم نمونه‌گیری را از اعضای آن انجام دهیم.

۲ تهیهٔ لیست به روز شده برای انجام نمونه‌گیری از تمام اعضای جامعه

۳ اختصاص شماره به اعضای جامعهٔ هدف (این قسمت اجباری نیست.)

۴ انتخاب نمونه با یکی از روش‌های احتمالی مثل قرعه‌کشی، روش‌های کامپیوتری و ...

یادداشت در نمونه‌گیری تصادفی ساده اگر بخواهیم از میان جامعهٔ آماری n نفره، k نفر را به عنوان نمونه انتخاب کنیم، شانس انتخاب همهٔ افراد در نمونه با هم برابر و مساوی $\frac{k}{n}$ خواهد بود.

مزایای روش نمونه گیری تصادفی ساده:

۱ ساده‌بودن روش نمونه‌گیری

۲ برای نمونه‌گیری از جوامع کوچک، روشی مناسب، عملی و سریع می‌باشد.

۳ انتخاب کاملاً تصادفی بوده و شانس انتخاب همهٔ افراد با هم یکسان و جهت‌گیری انتخاب در این روش وجود ندارد.

معایب روش نمونه گیری تصادفی ساده:

۱ نیاز به لیست کامل و جدید از اعضای جامعه داریم.

۲ در جوامع بزرگ و پراکنده دسترسی به همهٔ افراد و تهیهٔ لیست، کاری مشکل، هزینه‌بر و زمان‌بر است.

۳ با این‌که افراد تصادفی انتخاب شده‌اند اما ممکن است این افراد معرف همهٔ افراد و با خصوصیات مختلف در جامعه نباشند.

ب) نمونه گیری خوشه‌ای:

نمونه‌گیری که در آن، واحدهای نمونه‌گیری اولیه در جامعه، گروه‌ها یا خوشه‌ها هستند؛ سپس همهٔ واحدهای آماری خوشه‌های انتخاب شده را به عنوان نمونه درنظر می‌گیریم.

تذکر در واقع درون همهٔ خوشه‌های انتخاب‌شده سرشماری انجام می‌دهیم.

هر چه‌قدر گروه‌ها و خوشه‌های مختلف از نظر ویژگی مورد بررسی به هم نزدیک‌تر باشند (یعنی واریانس بین خوشه‌ها در ویژگی مورد بررسی کم باشد.) ولی درون هر خوشه اعضای خوشه تفاوت بیشتری از نظر ویژگی مورد بررسی داشته باشند (یعنی واریانس ویژگی مورد نظر درون خوشه زیادتر باشد)، نتایج حاصل از نمونه به واقعیت کل جامعه نزدیک‌تر خواهد بود.

مراحل انجام نمونه گیری خوشه ای:

۱ مشخص‌کردن جامعهٔ هدف

۲ مشخص‌کردن تقسیم‌بندی‌های جامعهٔ هدف به واحدهای کوچک‌تر و فرعی به عنوان خوشه

۳ انتخاب چند خوشه به طور تصادفی از میان کل خوشه‌ها

۴ سرشماری از همهٔ واحدهای آماری خوشه‌های انتخاب‌شده

۵ جمع‌بندی نتایج و تعمیم نتایج به کل داده‌ها

مزایای این روش:

۱ صرفه‌جویی در وقت و هزینه

۲ اگر تفاوت خوشه‌ها در ویژگی مورد بررسی زیاد نباشد نتایج به خوبی به کل جامعه قابل تعمیم خواهد بود.

www.Paadars.com

معایب این روش: اگر خوشه‌ها در ویژگی مورد بررسی تفاوت زیادی با هم داشته باشند ممکن است نتایج نمونه‌گیری با نتایج سرشماری کل افراد جامعه منطبق نباشد و فاصلهٔ زیادی داشته باشد.

(پ) نمونه‌گیری طبقه‌ای

روش نمونه‌گیری که در آن با طبقه‌بندی جامعه به زیرجامعه‌های مجزا یک نمونه تصادفی ساده از هر طبقه، انتخاب می‌شود. معمولاً تعداد افراد انتخابی از هر طبقه با نسبت تعداد اعضای آن طبقه نسبت به کل افراد جامعه، متناسب است.

مراحل انجام نمونه گیری طبقه ای:

۱) جامعهٔ هدف را برای نمونه‌گیری مشخص می‌کنیم.

۲) جامعه را به طبقات و بخش‌های مختلف طوری افراز (طبقه‌بندی) می‌کنیم که افراد هر طبقه از نظر ویژگی‌های مورد بررسی همگن باشند و دو یا چند طبقه با هم اشتراکی نداشته باشند.

۳) از میان افراد هر طبقه متناسب با تعداد اعضای آن طبقه (نسبت به کل افراد جامعه) نمونه‌گیری تصادفی ساده انجام می‌دهیم.

۴) افراد انتخاب‌شده از هر طبقه را در یک‌جا جمع کرده و بررسی‌ها را روی آن‌ها انجام می‌دهیم.

نکته اگر بتوانیم جامعه را به دو یا چند طبقهٔ مختلف تقسیم‌بندی کنیم که تفاوت‌ها و واریانس بین طبقات مختلف زیاد اما تفاوت‌ها و واریانس داخل طبقات کم باشد، بهتر است از روش «نمونه‌گیری طبقه‌ای» استفاده کنیم.

ولی اگر تفاوت‌ها و واریانس بین طبقات کم اما تفاوت‌ها و واریانس در داخل طبقات زیاد باشد، بهتر است از روش نمونه‌گیری خوشه‌ای استفاده شود.

نکته اگر N نفر را در گروه‌های $n_1, n_2, \ldots, n_k$ نفری طبقه‌بندی کنیم برای انتخاب یک نمونهٔ n نفره متناسب با تعداد افراد هر گروه باید از گروه iاُم تعداد $\frac{n_i}{N} \times n$ نفر انتخاب کنیم.

مثال اگر یک جامعهٔ ۵۰۰ نفری را به طبقات $n_1 = ۱۰۰$، $n_2 = ۲۰۰$، $n_3 = ۵۰$ و $n_4 = ۱۵۰$ نفری تقسیم کرده باشیم، برای انتخاب یک نمونهٔ ۱۰ نفری متناسب با تعداد اعضای هر گروه باید از این گروه‌ها به ترتیب ۲، ۴، ۱ و ۳ نفر انتخاب شوند

زیرا:

$$\frac{n_1}{N} \times n = \frac{۱۰۰}{۵۰۰} \times ۱۰ = ۲ \qquad \frac{n_2}{N} \times n = \frac{۲۰۰}{۵۰۰} \times ۱۰ = ۴$$

$$\frac{n_3}{N} \times n = \frac{۵۰}{۵۰۰} \times ۱۰ = ۱ \qquad \frac{n_4}{N} \times n = \frac{۱۵۰}{۵۰۰} \times ۱۰ = ۳$$

مزایای نمونه‌گیری طبقه‌ای:

۱) چون نمونه از همهٔ اقشار و گروه‌ها انتخاب می‌شود نتایج آن به نتایج بررسی کل جامعه نزدیک‌تر است.

۲) اگر واریانس و تفاوت در داخل طبقات کم باشد، دقت آمارگیری بالا می‌رود.

معایب این روش:

۱) ممکن است طبقه‌بندی هزینه داشته باشد و زمان‌بر باشد.

۲) ممکن است در بعضی موارد اطلاعات کافی برای طبقه‌بندی مناسب اعضای جامعه را در اختیار نداشته باشیم.

www.Paadars.com

تذکر (نمونه گیری چند مرحله ای)

گاهی می‌توان از دو روش نمونه‌گیری یا از یک روش، در چند مرحله و پشت سرهم استفاده کرد. مثلاً می‌توانیم در بعضی مراحل انتخاب نمونه، اعضای یک خوشه را طبقه‌بندی کرده و از روش نمونه‌گیری طبقه‌ای برای انتخاب نمونهٔ بهتر استفاده کنیم. این روش نمونه‌گیری را روش نمونه‌گیری **«چندمرحله‌ای»** می‌نامیم.

معمولاً وقتی در ویژگی مورد بررسی هم واریانس و تفاوت بین طبقات و گروه‌ها زیاد باشد و هم واریانس و تفاوت در داخل گروه‌ها زیاد باشد، استفاده از روش نمونه‌گیری چندمرحله‌ای مفید می‌باشد.

ت نمونه‌گیری سیستماتیک یا سامانمند

نوعی نمونه‌گیری طبقه‌ای است که در آن اندازهٔ طبقات با هم برابر است. فقط از طبقهٔ اول واحد آماری به تصادف انتخاب می‌شود و با همان رویه از طبقات دیگر، این‌کار انجام می‌گیرد.

یعنی از همهٔ طبقات دیگر عضو متناظر با واحد آماری انتخاب‌شده از طبقهٔ اول برای حضور در نمونه انتخاب می‌شود.

- در این روش تمام اعضای جامعه دارای شانس برابر برای قرارگرفتن در نمونه خواهند بود.
- اگر افراد جامعه به صورت تصادفی فهرست شده باشند، می‌توان روش نمونه‌گیری سیستماتیک یا منظم را به جای روش نمونه‌گیری تصادفی ساده استفاده کرد، اما اگر افراد جامعه با توجه به یک نظم معین براساس ویژگی‌های خاصی مرتب شده باشند، باید از نمونه‌گیری تصادفی ساده استفاده کرد.

توضیح این روش با یک مثال:

فرض کنید بخواهیم از بین ۱۰۰ نفر یک نمونهٔ ۲۰ نفره انتخاب کنیم. برای نمونه‌گیری به روش سیستماتیک باید مراحل زیر را طی کنیم:

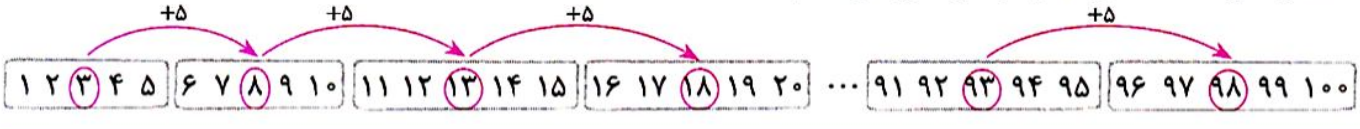
۱) تمام اعضای جامعه را به صورت تصادفی فهرست می‌کنیم و به آن‌ها شماره‌های ۱ تا ۱۰۰ را نسبت می‌دهیم.

۲) تعداد کل افراد جامعه را به تعداد اعضای نمونه‌ای که می‌خواهیم انتخاب کنیم، تقسیم می‌کنیم تا عددی به نام نسبت یا فاصلهٔ نمونه‌گیری به دست آید. در مسئلهٔ بالا فاصلهٔ نمونه‌گیری $\frac{۱۰۰}{۲۰} = ۵$ می‌باشد. با این کار جامعهٔ ۱۰۰ نفره را به ۲۰ گروه ۵ نفره به شکل زیر افراز (تقسیم‌بندی) کرده‌ایم:

۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	...	۹۶	۹۷	۹۸	۹۹	۱۰۰

۳) در این مرحله با استفاده از روش‌های نمونه‌گیری تصادفی عددی کم‌تر یا مساوی «فاصلهٔ نمونه‌گیری» انتخاب می‌کنیم. مثلاً در مثال بالا فرض کنید از عدد ۱ تا ۵ عدد ۳ انتخاب شود.

۴) فردی‌که شماره‌اش را انتخاب کرده‌ایم را در نمونه قرار داده و به شمارهٔ انتخاب‌شده «فاصلهٔ نمونه‌گیری» را اضافه می‌کنیم و همهٔ کسانی که شماره‌شان به دست می‌آید را در نمونه قرار می‌دهیم تا تعداد اعضای نمونه کامل شود.



مزایا این روش:

۱) ساده‌بودن روش ۲) کم‌هزینه‌تر و عملی‌تربودن نسبت به روش نمونه‌گیری تصادفی ساده

معایب این روش:

۱) ممکن است افراد انتخاب‌شده از همهٔ سطوح مختلف و متغیر در جامعه انتخاب نشوند.

۲) اگر در فواصل نمونه‌گیری منظم، افراد دارای خصوصیات یکسان قرار داشته باشند، نمونهٔ انتخاب‌شده، نمایندهٔ جامعه نخواهدبود.

www.Paadars.com

روش نمونه گیری غیراحتمالی یا غیر تصادفی

در این نوع نمونه‌گیری شانس حضور همهٔ اعضای جامعه در نمونهٔ انتخابی یکسان نیست و تصمیم‌گیری و قضاوت فرد انتخاب کننده و عوامل انسانی در انتخاب نمونه‌ها نقش دارد.

مواردی از نمونه گیری غیر احتمالی

۱) مواردی که در آن‌ها بخشی از جامعه بر مبنای قضاوت آمارگیر به عنوان نمونه انتخاب شوند. (آمارگیری قضاوتی)

۲) مواردی که قسمتی از جامعه به دلیل در دسترس‌بودن و سهولت و سادگی انتخاب به عنوان نمونه انتخاب شوند. (نمونه‌گیری در دسترس یا داوطلبانه)

۳) برای انتخاب نمونه، جامعه را به چند بخش مختلف تقسیم کرده و از هر بخش به تناسب تعدادشان موارد در دسترس را برای نمونه‌گیری انتخاب کنیم (نمونه‌گیری سهمیه‌ای)

۴) اگر یک فرد در دسترس را برای نمونه انتخاب کنیم و بقیهٔ افراد را با معرفی این شخص یا انتخاب او انتخاب کنیم، از روش غیراحتمالی استفاده کرده‌ایم. (نمونه‌گیری زنجیره‌ای)

نمونه گیری اریب:

اگر روش نمونه‌گیری درست انتخاب نشود ما را به نتایجی می‌رساند که دور از واقعیت است و حتی افزایش حجم نمونه هم باعث افزایش انحراف از واقعیت شود و **به سمتی خاص انحراف پیدا کند، در این حالت می‌گویند آن روش نمونه‌گیری اریب است.**

این انحراف از واقعیت در نتایج نمونه‌گیری می‌تواند به دلایل زیر باشد:

۱) روش نادرست انتخاب نمونه‌ها

۲) انتخاب نادرست روش‌گرفتن اطلاعات از هر واحد آماری

۳) تحلیل آماری نامناسب

هر کدام از عوامل زیر هم می‌توانند در خطاهای نمونه‌گیری مؤثر باشند:

- نداشتن فهرست کامل از افراد جامعه و شرایط آن‌ها (مثلاً در بررسی بیماران مبتلا به ایدز)
- انتخاب افرادی در نمونه که این افراد نمایندهٔ کل افراد موجود در جامعه نباشند و یا حتی مثلاً در بررسی یک پدیده افرادی که به آن پدیده مربوط نمی‌شوند در داخل نمونه اشتباهاً انتخاب شوند.
- پاسخ‌ندادن و یا حتی پاسخ اشتباه عمدی یا غیرعمدی به سؤالات پرسش‌نامهٔ طراحی‌شده برای جمع‌آوری اطلاعات
- تأثیر شخص مصاحبه‌کننده و یا سؤال او در نوع جواب شخصی که به عنوان نمونهٔ انتخاب شده و عدم پاسخ دقیق و درست
- اشتباهات ناشی از مشاهدهٔ غیردقیق برای جمع‌آوری اطلاعات، دقیق‌نبودن ابزار اندازه‌گیری، دقیق ثبت‌نشدن اطلاعات، عدم استخراج درست اطلاعات از منابع اطلاعاتی و آماری موجود و... .
- مقدار حجم نمونه که با بزرگ‌تر انتخاب‌کردن حجم نمونه می‌توانیم این خطا را کم کنیم.

در موارد زیر، علت خطای نمونه در چیست؟

۱) در آمارگیری برای تعیین این‌که خانواده‌های یک روستا چند نفره هستند، یک نمونهٔ ۱۰۰ نفره از مردم روستا انتخاب می‌کنیم و از هر کدام می‌پرسیم «خانوادهٔ شما چند نفره است؟»

۲ در آمارگیری برای این که «چند درصد از مردم تهران برای رفتن به محل کارشان از خودرو شخصی استفاده می‌کنند؟» یک نمونهٔ ۶۰ نفره از مردم تهران را انتخاب می‌کنیم و از هرکدام می‌پرسیم «برای رفتن به محل کارشان از خودرو شخصی استفاده می‌کنند؟»

۳ در یک مدرسه تعداد کلاس‌ها و دانش‌آموزان آن‌ها به صورت زیر است:

پایه	اول متوسطه	دوم متوسطه	سوم متوسطه
تعداد کلاس‌ها	۱۰	۸	۲
تعداد دانش‌آموزان هر کلاس	۱۰	۱۵	۴۰

تعداد کل کلاس‌ها در مدرسه $۲۰=۱۰+۸+۲$ کلاس

تعداد کل دانش‌آموزان $۳۰۰=۴۰\times۲+۱۵\times۸+۱۰\times۱۰$ نفر

نسبت تعداد کلاس‌های اول، دوم و سوم به کل کلاس‌ها به ترتیب $\frac{۱۰}{۲۰}\times۱۰۰=٪۵۰$، $\frac{۸}{۲۰}\times۱۰۰=٪۴۰$ و $\frac{۲}{۲۰}\times۱۰۰=٪۱۰$

حالا فرض کنید یک شخص آمارگیر نداند کلاس‌های اول، دوم و سوم از نظر تعداد چه نسبتی با هم دارند و تنها بداند که مدرسهٔ مورد بررسی مجموعاً ۲۰ کلاس اول، دوم و سوم دارد و برای فهمیدن تعداد کلاس‌های اول، دوم و سوم، ۶۰ نفر از دانش‌آموزان مدرسه را به عنوان نمونه در نظر بگیرد و از آن‌ها بپرسد: «کلاس شما چندنفره است؟»

با توجه به این‌که تعداد دانش‌آموزان سال اول، دوم و سوم به ترتیب $۱۰\times۱۰=۱۰۰$، $۱۵\times۸=۱۲۰$ و $۴۰\times۲=۸۰$ نفر است، پس به احتمال زیاد در نمونهٔ تصادفی ۶۰ نفره، افراد با همین نسبت انتخاب می‌شوند،

پس در نمونهٔ ۶۰ نفره تعداد دانش‌آموزان از هر پایه به ترتیب $\frac{۱۰۰}{۳۰۰}\times۶۰=۲۰$، $\frac{۱۲۰}{۳۰۰}\times۶۰=۲۴$ و $\frac{۸۰}{۳۰۰}\times۶۰=۱۶$ نفر می‌باشد.

که می‌گویند از پایه‌های اول، دوم و سوم هستند. پس از نظر آمارگیر

تعداد کلاس‌های اول، دوم و سوم با نسبت $\frac{۲۰}{۶۰}\times۱۰۰\simeq٪۳۳/۳$، $\frac{۲۴}{۶۰}\times۱۰۰=٪۴۰$ و $\frac{۱۶}{۶۰}\times۱۰۰\simeq٪۲۶/۶$ می‌باشند

که این با نتایج اولیه و واقعی اختلاف دارد.

این اختلاف به خاطر این است که در نمونهٔ انتخاب‌شده احتمال حضور دانش‌آموزان هر کلاس در نمونه برابر نیست و کلاس‌های با تعداد بالاتر احتمال حضور بیشتری در نمونه دارند، پس تعداد کلاس‌های پایهٔ آن‌ها به طور غیرواقعی افزایش می‌یابد.

برای این‌که نتایج این نمونه‌گیری درست به‌دست آید، باید تعداد افرادی که از هر کلاس در نمونه انتخاب می‌شوند با هم برابر باشند.

گردآوری داده ها

روش‌های گردآوری داده‌ها:
۱ روش‌های میدانی
۲ روش استفاده از دادگان‌ها: استفاده از مجموعه اطلاعات ذخیره‌شدهٔ قبلی

روش‌های میدانی:
آ مشاهده: گردآوری داده‌ها بدون نیاز به فرد پاسخگو مثلاً شمارش تعداد ماشین‌های عبوری از یک خیابان در یک ساعت خاص
ب پرسش‌نامه: مجموعهٔ سؤالات از پیش تعیین‌شده که باید توسط تعدادی پاسخگو تکمیل شود.
پ مصاحبه: به صورت حضوری یا تلفنی یا ... بین آمارگیر و پاسخگو انجام می‌گیرد.

www.Paadars.com

مشاهده : در این روش اطلاعات و داده‌های مربوط به موجودات زنده، اشیا یا پدیده‌های مختلف از طریق مشاهده یا دیگر حواس پنج‌گانه ملاحظه و ثبت می‌شود. ثبت وقایع می‌تواند با کمک «ابزارهای» مخصوص این کار صورت گیرد.

مزایا:
۱) در جوامع کوچک، اگر اندازه‌گیری‌ها با دقت انجام شوند، روش مشاهده می‌تواند دقیق و قابل اطمینان باشد.
۲) ساده، سریع و در دسترس است.
۳) اگر از ابزار دقیق و مخصوص استفاده شود، اطلاعات زیادی در زمان کم قابل دستیابی است.

معایب:
۱) اگر به حواس پنج‌گانه تکیه کنیم ممکن است دقت بالا نداشته و در مشاهدات خطا داشته باشیم.
۲) اگر تعداد موارد مورد بررسی زیاد باشد، وقت‌گیر است و باید از ابزارهای دقیق و سریع استفاده کنیم که پرهزینه است.
۳) تمایلات شخصی مشاهده‌گر و توانایی او، در مشاهده و ثبت دقیق فعالیت مورد مشاهده می‌تواند اثر داشته باشد.
۴) در ابزار مورد استفاده هم امکان خطا وجود دارد.

پرسشنامه: استفاده از پرسش‌نامه مرسوم‌ترین ابزار گرفتن اطلاعات از مردم است.
پرسش‌نامه مجموعه‌ای از سؤالات از پیش تعیین شده است که باید توسط تعدادی پاسخگو تکمیل شوند.

مزایا:
۱) اگر اعضای جامعه در دسترس باشند، آن‌ها را در زمان کم پوشش می‌دهد.
۲) تجزیه و تحلیل نتایج آسان است و معمولاً به روش‌های کامپیوتری امکان‌پذیر است.
۳) پاسخگو بدون این‌که نگران باشد جواب صادقانه می‌دهد و می‌تواند در موارد لازم هویتش مخفی بماند.

معایب:
۱) «حذف افراد بی‌سواد»
۲) «عدم مشارکت بعضی افراد در جواب‌دادن به پرسش‌ها»
۳) «زمان‌بر بودن آن اگر تعداد افراد خیلی زیاد باشد»

مصاحبه: معمولاً بین ۲ نفر انجام می‌گیرد یکی مصاحبه‌گر یا آمارگیر و دیگری مصاحبه‌شونده یا پاسخگو.
بیشتر در زمانی که آمارگیر اطلاع کافی از همهٔ پاسخ‌های ممکن ندارد استفاده می‌شود تا مسائل مورد بررسی روشن شود.

مزایا:
۱) عدم حذف بی‌سوادان
۲) رفع ابهام سؤالات مبهم
۳) رفع عدم تمایل مصاحبه‌شونده بر پاسخگویی
۴) تحریک شخص به بیان حقیقت

معایب:
۱) وقت‌گیر و پرهزینه‌بودن
۲) تحلیل مشکل‌تر داده‌ها و نتایج
۳) نیاز داشتن به مصاحبه‌گر علاقه‌مند، توانا، بی‌طرف و آشنا با تحقیق
۴) محدودیت در جمعیت‌های زیاد اشاره کرد.

دادگان: در بعضی از تحقیق‌ها اطلاعات موردنیاز آمارگیر از قبل توسط افراد، ارگان‌ها، مؤسسه‌های آمارگیری یا ... جمع‌آوری شده‌اند و یا حتی به صورت کتاب یا مجله یا بولتن و... چاپ شده‌اند و نیازی به جمع‌آوری دوبارهٔ آن‌ها نیست و می‌توان از اطلاعات قبلی استفاده کرد.

مزایا:
۱) ارزان‌بودن اطلاعات
۲) صرفه‌جویی در وقت و جلوگیری از دوباره‌کاری
۳) ارزیابی اطلاعات و روند تحقیقات گذشته و انتخاب مسیر درست برای جمع‌آوری اطلاعات تکمیلی

معایب:
۱) موجودنبودن اطلاعات ثبت‌شدهٔ دقیق یا ناقص‌بودن آن‌ها در بعضی موارد
۲) وجود مانع برای دسترسی به اطلاعات ثبت‌شده مثلاً در مسائل امنیتی یا در موارد آبروی اشخاص
۳) قدیمی و کهنه‌بودن و یا عدم ثبت دقیق و یا جهت‌داربودن اطلاعات ثبتی در بعضی موارد

مفاهیم جمع آوری آمار

آمارگیری: گردآوری داده‌ها به یکی از روش‌های ممکن.

آمارگیر: کسی که آمارگیری را انجام می‌دهد.

متغیر: هر ویژگی از اشخاص یا اشیا که قرار است بررسی شود مثل قد، وزن، IQ، گروه خونی، جنسیت، نمرهٔ درس ریاضی و ... را متغیر می‌گویند. این متغیرها معمولاً می‌توانند از یک عضو به عضو دیگر تغییر کنند.

متغیرها
۱) **کمی:** متغیرهایی که مقادیر عددی می‌گیرند و برای آن عملیات ریاضی مثل جمع، تفریق و معدل‌گیری قابل انجام است. مثل: سن، قد، وزن، درآمد یک شخص و ...
۲) **کیفی:** متغیرهایی که صرفاً برای دسته‌بندی افراد یا اشیا در گروه‌ها به کار می‌روند، لزوماً مقدار عددی نمی‌گیرند و معمولاً انواع مختلف یک صفت هستند. مثل: جنسیت افراد (زن، مرد)، گروه خونی (A، B، AB و O) و ...

داده ها: مقادیر اندازه‌گیری‌شدهٔ (مشاهده‌شدهٔ) یک متغیر هستند که با اندازه‌گیری، آزمایش، مشاهده و ... به دست می‌آیند و در محاسبات، برنامه‌ریزی‌ها و پیش‌بینی‌ها به کار می‌روند.

آماره یا آماره نمونه: مشخصه‌ای عددی که توصیف‌کنندهٔ جنبه‌ای خاص از نمونه است و از داده‌های نمونه به‌دست می‌آید.
در واقع مشخصه‌های عددی هر نمونه مثل میانگین، واریانس و ... را «آماره» می‌گویند.
آماره‌ها متغیرهای تصادفی هستند یعنی ثابت نیستند و از نمونه‌ای به نمونهٔ دیگر می‌توانند تغییر کنند.

پارامتر یا پارامتر جامعه: یک مشخصه عددی است که توصیف‌کنندهٔ جنبه‌ای خاص از جامعه است و در صورتی که داده‌های کل جامعه در اختیار باشند، قابل محاسبه است.

فرض کنید بخواهیم میانگین وزن دانشجویان یک دانشگاه که ۱۰۰۰ دانشجو دارد را بررسی کنیم. چند نمونهٔ ۲۰ نفری از دانشجویان انتخاب کرده‌ایم و پس از اندازه‌گیری‌های لازم میانگین وزن این نمونه‌های انتخاب‌شده به شکل زیر بوده است:

شمارهٔ نمونهٔ ۲۰ نفره	۱	۲	۳	۴	۵	...
میانگین وزن نمونهٔ ۲۰ نفره	۸۰	۸۳	۸۶	۸۵	۸۲	...

همان‌طور که می‌بینید میانگین‌ها در نمونه‌های انتخاب شده متفاوت هستند. **که به آنها آماره می گویند**

در این مثال میانگین وزن کل دانشجویان عددی ثابت است و با سرشماری و داشتن اطلاعات همهٔ عضوهای جامعه قابل محاسبه است. **این مقدار را پارامتر جامعه می گویند (و مقداری ثابت است)**

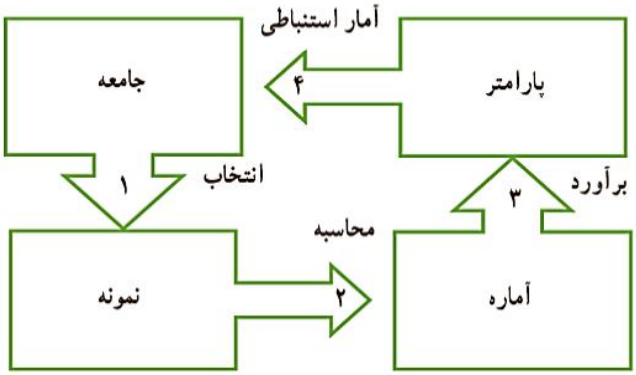
توجه: گاهی نمی توان پارامتر جامعه را از روی سرشماری پیدا کرد، در این حالت مجبوریم از آماره های نمونه ها استفاده نماییم!

آمار استنباطی: فرایند نتیجه‌گیری دربارهٔ پارامترهای جامعه بر اساس نمونه، **آمار استنباطی** نام دارد.

www.Paadars.com

درس ۲ برآورد

فرض کنید بخواهیم «میانگین تعداد اعضای خانوارهای کشور» را که آن را «بعد خانوار» می‌گویند، تعیین کنیم.
این مقدار یک «پارامتر» از جامعه به حساب می‌آید و برای به دست آوردن آن باید از جامعهٔ موردنظر سرشماری کنیم، چون انجام سرشماری برای به دست آوردن مقدار یک پارامتر، معمولاً کاری وقت‌گیر و هزینه‌بر است و مشکلاتی نظیر در دسترس‌نبودن همهٔ افراد و ... را دارد، می‌توانیم: «یک نمونهٔ تصادفی از جامعه انتخاب و آمارهٔ نظیر پارامتر مجهول را در این نمونه به‌دست آوریم. حالا با استفاده از مقدار آمارهٔ به دست آمده، مقدار پارامتر را حدس بزنیم. این کار را **برآوردکردن پارامتر جامعه** می‌نامیم.»



در آمار ۲ نوع برآورد وجود دارد:

۱ برآورد نقطه‌ای پارامتر جامعه

برابر است با مقدار عددی حاصل از جای‌گذاری اعداد نمونه تصادفی در آمارهٔ نظیر آن پارامتر.
به بیان دیگر مقدار عددی آماره را برآورد یا برآورد نقطه‌ای می‌نامند.

فرض کنید جامعه‌ای n عضوی دارای یک پارامتر ثابت و مجهول باشد. با در نظر گرفتن نمونه‌های ۱، ۲، ۳، ... و n عضوی می‌توان آمارهٔ نظیر پارامتر جامعه را به دست آورد و پارامتر مورد بررسی را براساس عدد آمارهٔ به دست آمده، برآورد کرد.
هر چه‌قدر تعداد اعضای نمونه بیشتر شود، امکان نزدیک‌شدن برآورد به پارامتر بیشتر است تا جایی که اگر اندازهٔ نمونه با اندازهٔ جامعه برابر شود، برآورد دقیقاً برابر با پارامتر می‌شود.

۲ برآورد بازه‌ای یا بازه اطمینان پارامتر جامعه

عبارت است از بازه‌ای عددی برای پارامتر به همراه یک درصد اطمینان که به **ضریب اطمینان** شهرت دارد.

اگر برآورد شود که پارامتر در بازهٔ $[a,b]$ قرار دارد، مقدار $b-a$ را فاصلهٔ اطمینان می‌گوییم. بدیهی است هر چه‌قدر فاصلهٔ اطمینان کوتاه‌تر و ضریب اطمینان بالاتر باشد، برآورد فاصله‌ای بهتر انجام شده است.

مثال: نمرات دانش‌آموزان شرکت‌کننده در یک آزمون به صورت ۸، ۱۱، ۱۴، ۱۷ و ۲۰ می‌باشد، **اگر** پارامتر جامعه را برابر با میانگین نمرات دانش‌آموزان در نظر بگیریم و آن را با μ نمایش دهیم

الف) مقدار پارامتر را بدست آورید؛

حالا فرض کنید بخواهیم به روش برآورد نقطه‌ای با استفاده از نمونه‌ها، میانگین کلی جامعه یعنی μ را برآورد کنیم.

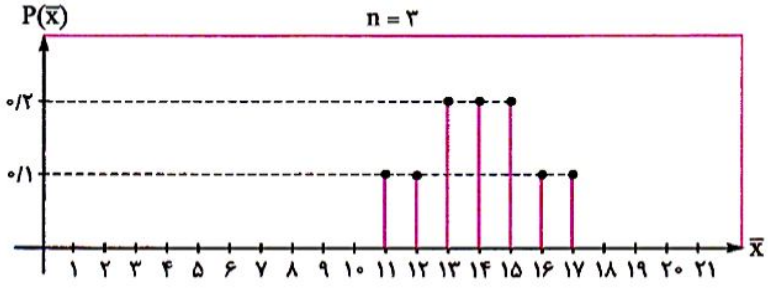
ب) چند نمونه ی ۳ عضوی از این جامعه می توان داشت؟

در جدول زیر همه ی نمونه های سه عضوی را به همراه میانگین شان (آماره متناظر با پارامتر) و احتمال هر کدام از میانگین ها آمده است:

نمونه‌ها	{۸,۱۱,۱۴}	{۸,۱۱,۱۷}	{۸,۱۱,۲۰} یا {۸,۱۴,۱۷}	{۸,۱۴,۲۰} یا {۱۱,۱۴,۱۷}	{۸,۱۷,۲۰} یا {۱۱,۱۴,۲۰}	{۱۱,۱۷,۲۰}	{۱۴,۱۷,۲۰}
$\bar{x}$	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷
احتمال $p(\bar{x})$	۰/۱	۰/۱	۰/۲	۰/۲	۰/۲	۰/۱	۰/۱

ج) میانگین آماره های جدول بالا (که شامل همه ی نمونه های ۳ عضوی جامعه است) را نوشته و نتیجه ی خود را بنویسید.

رسم توزیع آماره‌ای **جدول بالا:**



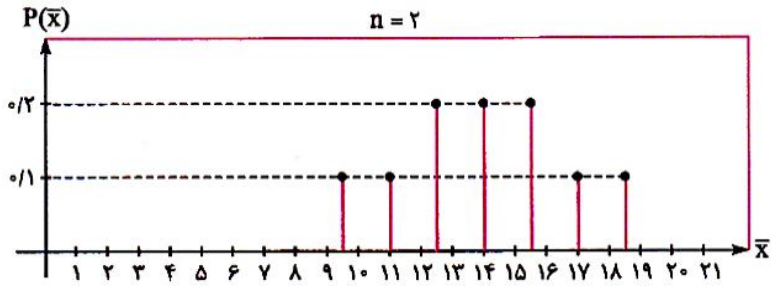
اگر همین کار را برای **همه ی نمونه های دو عضوی و چهارعضوی** هم انجام بدهیم، نمودارها به شکل زیر خواهند بود:

برآورد پارامتر میانگین به وسیلهٔ نمونه‌های ۲ عضوی:

نمونه‌ها	{۸,۱۱}	{۸,۱۴}	{۸,۱۷} یا {۱۱,۱۴}	{۸,۲۰} یا {۱۱,۱۷}	{۱۱,۲۰} یا {۱۴,۱۷}	{۱۴,۲۰}	{۱۷,۲۰}
میانگین $\bar{x}$	۹/۵	۱۱	۱۲/۵	۱۴	۱۵/۵	۱۷	۱۸/۵
احتمال	۰/۱	۰/۱	۰/۲	۰/۲	۰/۲	۰/۱	۰/۱

www.Paadars.com

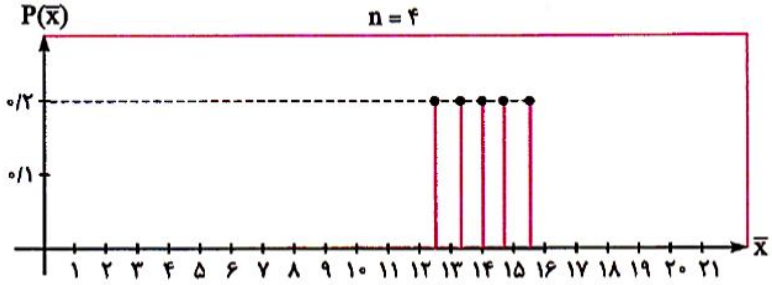
نمودار توزیع آمارهٔ $\bar{x}$ در نمونه‌های دو عضوی:



برآورد پارامتر میانگین به وسیلهٔ نمونه‌های ۴ عضوی:

نمونه‌های ۴ عضوی	{۸,۱۱,۱۴,۱۷}	{۸,۱۱,۱۴,۲۰}	{۸,۱۱,۱۷,۲۰}	{۸,۱۴,۱۷,۲۰}	{۱۱,۱۴,۱۷,۲۰}
$\bar{x}$	۱۲/۵	۱۳/۲۵	۱۴	۱۴/۷۵	۱۵/۵
$P(\bar{x})$	۰/۲	۰/۲	۰/۲	۰/۲	۰/۲

نمودار توزیع آمارهٔ $\bar{x}$ در نمونه‌های ۴ عضوی:



سوال: با افزایش تعداد اعضای نمونه، وضعیت برآوردها و پارامتر جامعه چگونه است؟

با افزایش اندازهٔ نمونه‌ها برآوردها ($\bar{x}$ ها) به میانگین جامعه که پارامتر است (μ)، می‌شوند.

نکته: در هر نمودار با زیادشدن اندازهٔ نمونه انحراف‌معیارِ «برآوردهای پارامتر» کم‌تر می‌شود و بدیهی است هر چه‌قدر انحراف معیار برآوردها حول میانگین جامعه کم‌تر باشد، احتمال این‌که با نمونه‌گیری به عددی نزدیک به پارامتر برسیم بیشتر بوده و آن برآورد بهتر خواهد بود.

یادآوری: پارامتر، یک ویژگی جامعه بوده و آماره، یک ویژگی نمونه است. **در این** مثال میانگین جامعه یعنی μ یک پارامتر است و میانگین **هر کدام از نمونه ها**، $\bar{x}$، یک آماره است و همان‌طور که دیدیم مقدار آماره، ممکن است از نمونه‌ای به نمونهٔ دیگر تفاوت داشته باشد.

انحراف معیارِ برآورد میانگین جامعه

اگر در جامعه‌ای که میانگین داده‌های آن μ و انحراف معیار داده‌های آن σ باشد، برای برآورد نقطه‌ای پارامترِ μ، همهٔ نمونه‌های تصادفی nعضوی ممکن را از جامعه انتخاب کرده و میانگین $\bar{x}$ را در آن‌ها حساب کنیم انحراف معیار آماره‌های به دست آمده ($\bar{x}$ ها) از رابطهٔ زیر به دست می‌آید:

$$\text{انحراف معیار برآورد میانگین جامعه} = \frac{\text{انحراف معیار جامعه}}{\text{جذر اندازهٔ نمونه}}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

سوال: بر اساس این رابطه، توضیح دهید که چگونه افزایش نمونه در انحراف معیار آماره ها تاثیر گذار است؟

www.Paadars.com

مساله: اگر انحراف‌معیار «بعد خانوار» در کشوری ۳ نفر باشد:
۱) انحراف‌معیار برآورد میانگین تعداد اعضای خانوار را برای اندازهٔ نمونه ۹۰۰ نفری بیابید.
۲) اندازهٔ نمونه را چند برابر کنیم تا انحراف‌معیار برآورد میانگین ۰/۱ برابر شود؟

سوال: اگر اندازهٔ نمونه ۱۰ برابر شود، انحراف معیار برآورد میانگین چند برابر می‌شود؟

مساله: در یک مطالعه، انحراف‌معیار برآورد میانگین در نمونه‌ها از $\frac{۱}{۲۰}$ انحراف‌معیار جامعه کم‌تر است. حداقل حجم نمونه‌ها کدام است؟

برآورد بازه‌ای برای میانگین جامعه: اگر نمونه‌ای تصادفی به اندازهٔ n در اختیار داشته باشیم، با اطمینان بیش از ۹۵ درصد می‌توانیم بگوییم:

$$\bar{x}-\frac{۲\sigma}{\sqrt{n}} \le \mu \le \bar{x}+\frac{۲\sigma}{\sqrt{n}}$$

μ میانگین جامعه
σ انحراف معیار جامعه

تذکر: ضریب اطمینان ۹۵٪ در برآورد بازه‌ای به این معنا است که اگر ۱۰۰ نمونهٔ مختلف با حجم برابر انتخاب کنیم و برای هر کدام $\bar{x}$ را حساب کنیم و با استفاده از میانگین به دست آمده از نمونه بازهٔ $(\bar{x}-\frac{۲\sigma}{\sqrt{n}}, \bar{x}+\frac{۲\sigma}{\sqrt{n}})$ را برای برآورد پارامتر میانگین بسازیم، فقط ۵تا از این ۱۰۰ بازهٔ برآوردشده شامل عدد μ نیستند و ۹۵٪ آن‌ها شامل μ می‌باشند.

مساله: ۵ نفر از کشاورزان یک منطقه را به عنوان نمونه انتخاب کرده و میزان درآمد سالانهٔ آن‌ها را ۶، ۸، ۱۰، ۱۴ و ۱۲ میلیون تومان محاسبه کرده‌ایم. اگر واریانس درآمد کشاورزان منطقه، طبق آمارهای قبلی $\sigma^۲=\frac{۵}{۴}$ باشد، با اطمینان ۹۵٪، بازه‌ای را برای میانگین درآمدهای کشاورزان منطقه برآورد کنید.

نکته: وقتی میانگین در فاصلهٔ (a,b) برآورد می‌شود مقدار $\bar{x}$ نمونه برابر با $\frac{a+b}{2}$ و مقدار $\frac{2\sigma}{\sqrt{n}}$ برابر $\frac{b-a}{2}$ خواهد بود. هم‌چنین طول بازهٔ اطمینان $b-a$ است.

مساله: با اندازه‌گیری وزن ۱۰ کشتی‌گیر بازهٔ (۸۰,۱۰۰) را با اطمینان ۹۵٪ برای وزن کل کشتی‌گیران شرکت‌کننده در مسابقات المپیک برآورد کرده‌ایم. میانگین وزن کشتی‌گیران و ضریب تغییرات را برای وزن کشتی‌گیران به دست آورید.

برآورد بازه‌ای نسبت با اطمینان ۹۵ درصد:

اگر از جامعه‌ای نسبتاً بزرگ، n نمونه‌ای تصادفی ساده انتخاب کنیم و m تا از آن‌ها ویژگی مورد مطالعه ما را داشته باشند، آن‌گاه نسبت واقعی افرادی از جامعه که آن ویژگی را دارند با اطمینان ۹۵ درصد در بازهٔ زیر است:

$$(p-2\sqrt{\frac{p(1-p)}{n}}\ ,\ p+2\sqrt{\frac{p(1-p)}{n}})$$

که در آن $p=\frac{m}{n}$ برآورد پارامتر نسبت ویژگی در جامعه است.

بیان چند نکته:

- طول بازهٔ اطمینان برابر با $4\sqrt{\frac{P(1-P)}{n}}$ می‌باشد.
- میزان خطا در محاسبهٔ $P=\frac{m}{n}$ طبق رابطهٔ بالا حداکثر $2\sqrt{\frac{P(1-P)}{n}}$ می‌باشد.
- اگر در بازهٔ اطمینان تعداد نمونه‌ها k برابر شود، طول بازهٔ اطمینان تقسیم بر $\sqrt{k}$ خواهد شد،

نتیجه: اگر تعداد نمونه‌ها ۱۰۰ برابر شود، طول بازهٔ اطمینان بر $\sqrt{100}=10$ تقسیم خواهد شد. بنابراین دقت محاسبهٔ نسبت P، یک رقم اعشار بهتر می‌شود.

- اگر بخواهیم بدون تغییر تعداد نمونه با اطمینان بیش از ۹۵٪ نسبت را برآورد کنیم، لازم است که به جای عدد ۲ ضریبی بزرگ‌تر گذاشته شود.

مساله: برای این‌که بدانیم چند درصد دانش‌آموزان شهر تهران از تدریس معلم ریاضی خود راضی هستند ۱۲۰۰ نفر از آن‌ها را به عنوان نمونه انتخاب کرده و متوجه شدیم ۹۰۰ نفر از آن‌ها از تدریس معلم خود راضی هستند. نسبت واقعی افرادی که از تدریس معلم ریاضی‌شان راضی هستند با اطمینان ۹۵٪ در چه بازه‌ای قرار دارد؟

نکته: (بخصوص در هنگامی که برآوردی از نسبت مورد مطالعه نداشته باشیم)

در برآورد بازه‌ای نسبت، بازهٔ روبه‌رو شامل بازهٔ اطمینان ۹۵ درصدی است: $(p=\frac{m}{n})$ $(p-\frac{1}{\sqrt{n}},p+\frac{1}{\sqrt{n}})$

مثال: اگر بخواهیم نسبت دانش‌آموزان شهر تهران را که روزانه بیش از ۱ ساعت مطالعهٔ غیردرسی دارند، محاسبه کنیم و بخواهیم طول بازهٔ اطمینان ۹۵ درصدی کم‌تر از چهار درصد باشد، حداقل چند نفر از دانش‌آموزان را باید به عنوان نمونه انتخاب کنیم؟