



آمار استنباطی



مجتبی جهانی فر
دانشکده روانشناسی و علوم تربیتی دانشگاه شهید چمران اهواز
نیم سال اول سال تحصیلی ۹۹-۹۸

آشنایی با مفاهیم بنیادی آمار استنباطی (نمونه گیری و برآورد)



موضوعات اساسی در آمار استنباطی

فرض آزمایی

آزمونهای آماری

فرضیه

برآورد

آماره

پارامتر

نمونه

جامعه

خطای نمونه گیری

روشهای نمونه گیری

جامعه

واحد نمونه گیری



عنصر

نمونه

- به مشخصه مربوط به جامعه پارامتر گفته می شود. پارامترها مقادیر ثابت و مجهول هستند.

- به مشخصه مربوط به نمونه آماره گفته می شود، آماره متغیر تصادفی است که می تواند مقادیر زیادی را اختیار کند.

- مقدار پارامتر هر جامعه از طریق آماره نمونه برآورد می شود. این کار با مفاهیم و قوانین آمار استنباطی ممکن شده است.

- برای محاسبه آماره باید از جامعه نمونه گیری کرد.

روش های نمونه گیری احتمالاتی

نمونه گیری
احتمالاتی

نمونه گیری
سیستماتیک

نمونه گیری
خوشه ای

نمونه گیری
طبقه ای

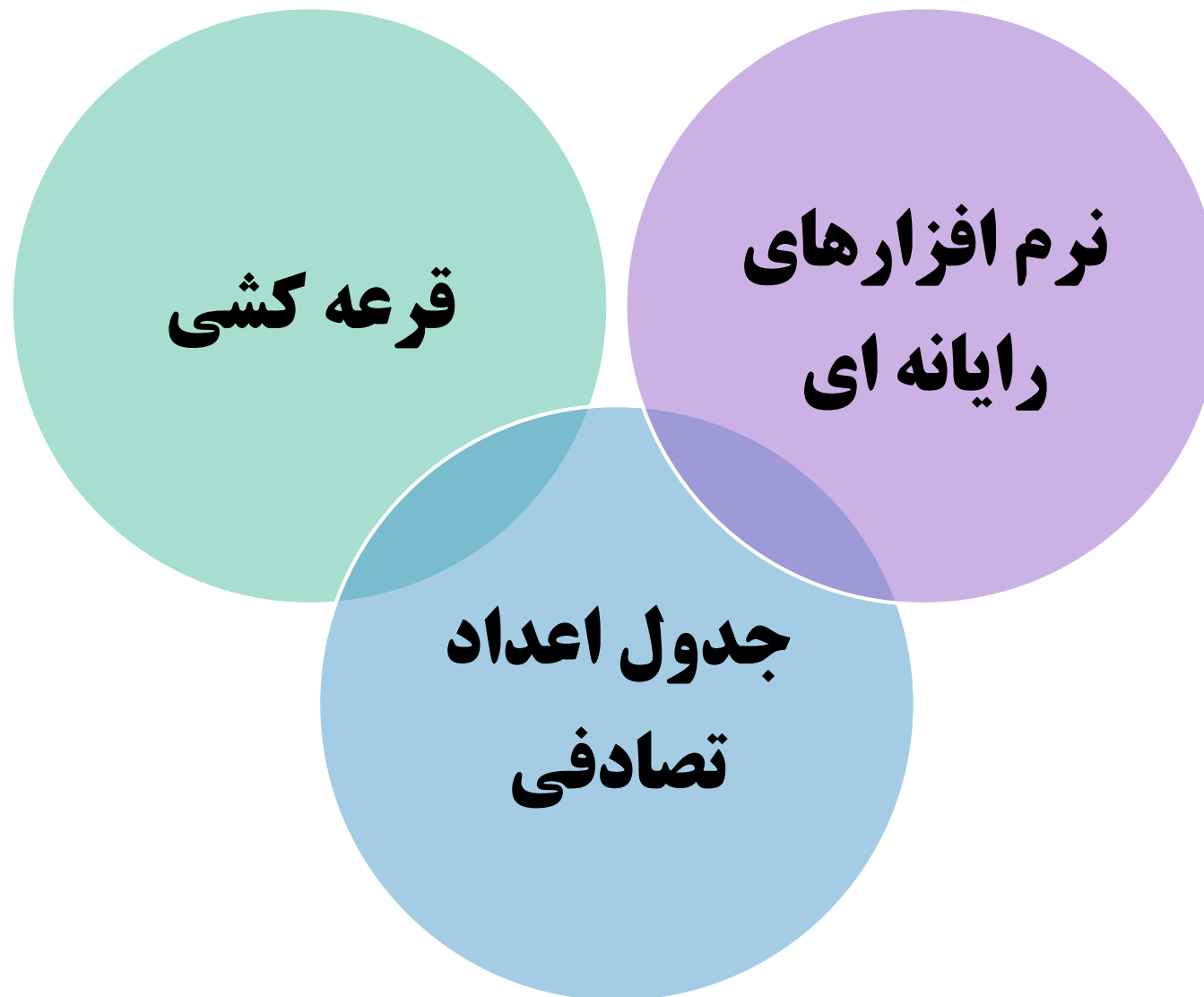
نمونه گیری
تصادفی ساده

نمونه گیری تصادفی ساده



- ساده ترین روش نمونه گیری احتمالاتی **نمونه گیری تصادفی ساده** است.
- درک این روش، کمک زیادی به فهم سایر نمونه گیری ها خواهد کرد، به همین خاطر به آن **نمونه گیری مادر** می گویند.
- روش نمونه گیری ساده به دو روش **با جایگذاری** و **بدون جایگذاری** انجام می شود.
- در این روش، n واحد از جامعه ای به حجم N واحد انتخاب می شود. ($n < N$)
- از میان N واحد جامعه، می توان k نمونه به حجم n انتخاب کرد به طوری که $k = \binom{N}{n}$
- شانس انتخاب همه k نمونه ممکن **یکسان** و برابر با $\frac{1}{k}$ است. (بدون جایگذاری)
- شانس انتخاب همه واحدهای جامعه **یکسان** و برابر با $\frac{1}{N}$ است. (**استخراج در نوبت اول، نوبت دوم و... با هم برابر است**) (بدون جایگذاری)
- بیشتر برای جوامع کوچک، و یکنواخت به کار می رود که در دسترس هم باشد.

واحدهای نمونه گیری را چگونه انتخاب کنیم؟



جدول اعداد تصادفی

- هر کدام از رقم های جدول از اعداد تصادفی از ۰ تا ۹ تشکیل شده اند.
- همه رقمها به طور تصادفی و مستقل از توزیع تولید شده اند.
- رقم های ۰ تا ۹ شانس یکسانی برای انتخاب شدن در هر خانه را دارند.
- ساختار جدول به گونه ای است که اگر بخواهیم اعداد چند رقمی بسازیم احتمال انتخاب ثابت باقی می ماند.
- در این روش باید همه عناصر جامعه شماره گذاری شوند.

68 4 2 5 7 954 1 2 5 6 32140

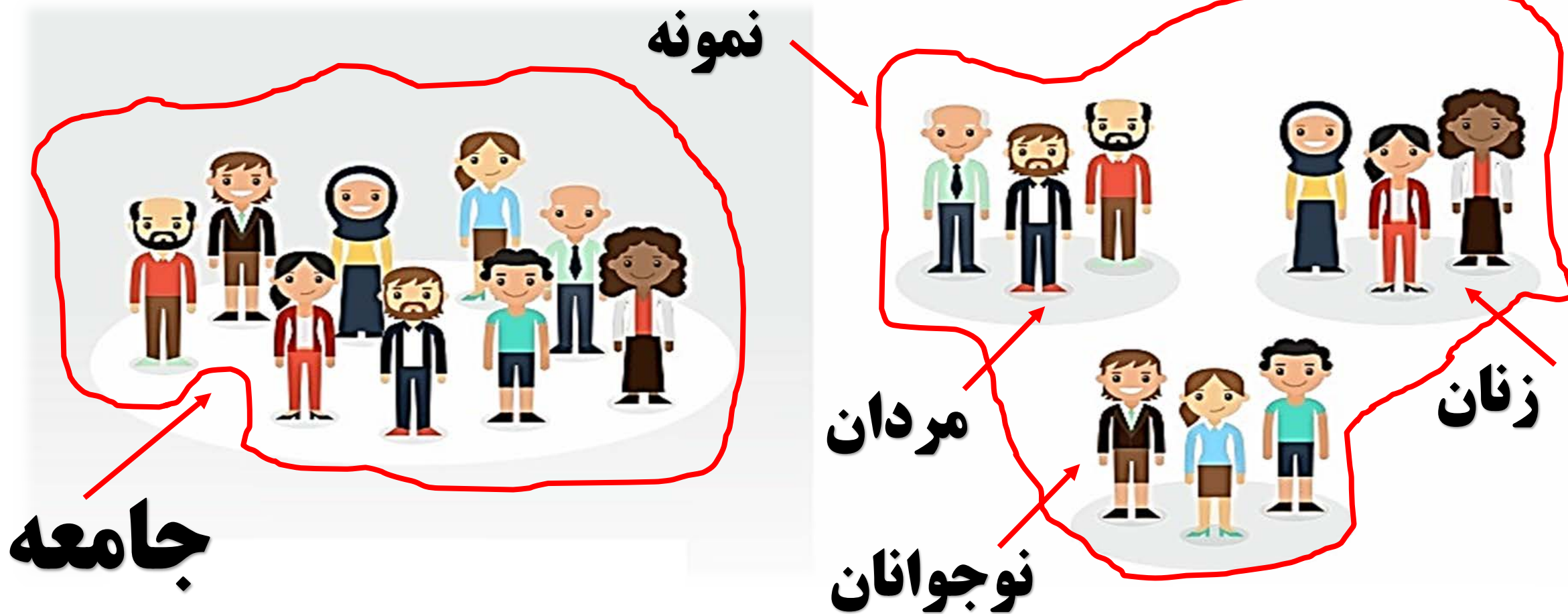
58 2 0 3 2 154 7 8 5 9 62024

36 2 3 3 3 254 7 8 9 1 20325

98 5 2 6 3 0 1 7 4 2 4 5 0 3 6 8 6

نمونه گیری تصادفی با طبقه بندی

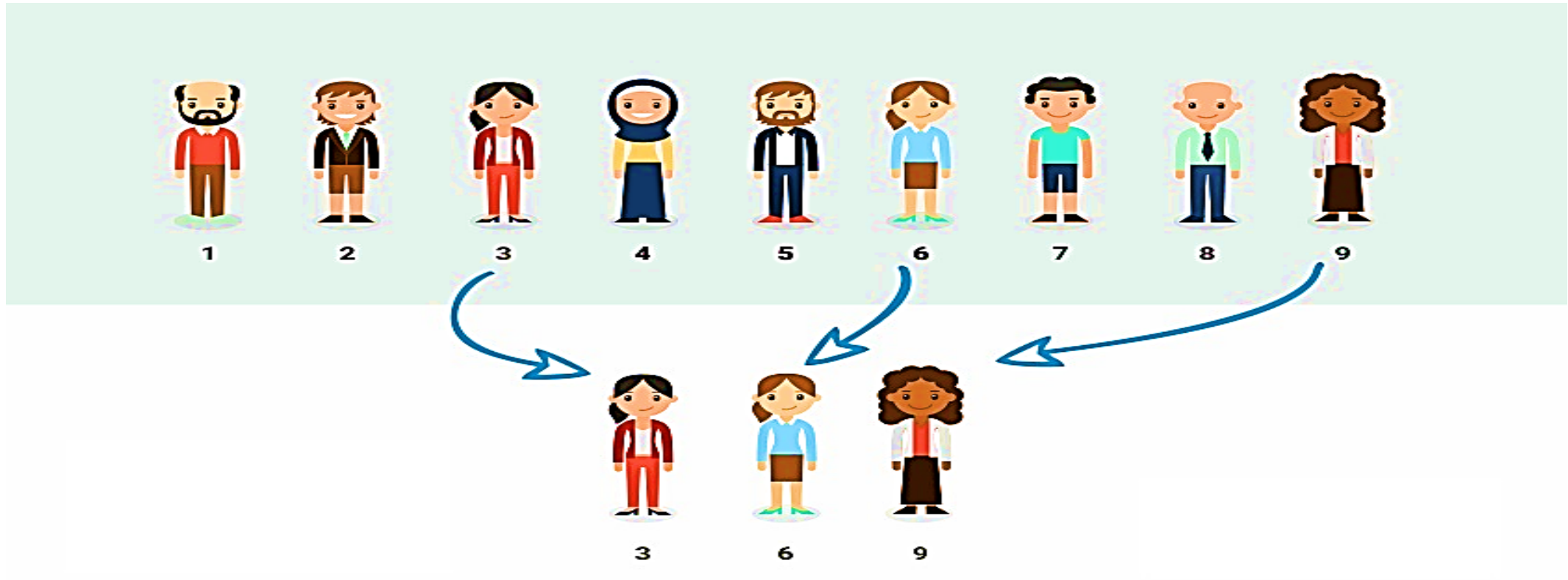
- هرگاه جامعه خیلی نا همگن باشد، دیگر با افزایش حجم نمونه و با وجود محدودیت های اقتصادی نمی توانیم با استفاده از نمونه گیری تصادفی ساده به برآوردهای دقیقی برسیم.
- نوسانات شدید یک صفت در جامعه موجب می شود تا به طبقه ها و گروههای مختلف جامعه نیز توجه شود.
- برای حجم نمونه های برابر خطای برآورد در طبقه بندی از تصادفی ساده کمتر است.
- هزینه یک پژوهش با طبقه بندی به مناطق جغرافیایی و... کاهش پیدا می کند.
- نمونه گیری طبقه ای باعث شناسایی بهتر زیر جامعه ها می شود.
- طبقه بندی در جامعه بر حسب ویژگی های متفاوت آن جامعه است.
- در برخی از موارد هم طبقه ها به صورت طبیعی در جامعه تشکیل شده اند.



- واریانس درون طبقه ها باید کوچک باشد. (طبقه ها از درون باید همگن باشند).
- طبقه ها در بیرون باید نا همگن باشند و واریانس بین طبقه ها زیاد باشد.
- واریانس کل از مجموع واریانس مستقل طبقه ها محاسبه می شود پس بهتر است واریانس درون طبقه ها کم باشد.
- جنسیت، نژاد، شغل، میزان درآمد، وضعیت تاهل و..... مواردی هستند که جامعه بر حسب آنها طبقه بندی می شود.

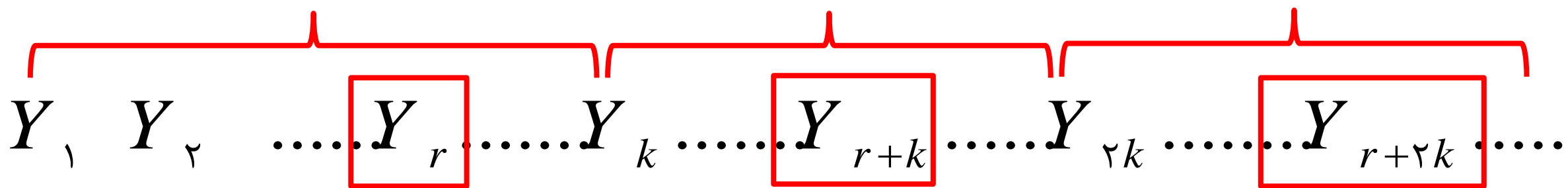


- در برخی از طرح های نمونه گیری یا **فهرست واحد ها** در دسترس نیست، و یا تهیه آن مشکل و مقرون به صرفه نیست.
- نمونه گیری سیستماتیک در بسیاری از موارد **دارای خطای کمتری نسبت به نمونه گیری تصادفی** است
- در این روش تنها اولین واحد تصادفی انتخاب می شود و سایر واحدهای نمونه طبق الگویی از پیش تعیین شده انتخاب می شوند.



فرض کنیم حجم جامعه تحت بررسی N باشد، و واحدهای جامعه را از 1 تا N شماره گذاری کنیم. نمونه سیستماتیک یک در k (یعنی از هر k واحد یک واحد انتخاب می شود) به حجم n به صورت زیر تعریف و انتخاب می شود:

■ از واحدهای شماره 1 تا k واحدی را به تصادف انتخاب می کنیم، مثلاً واحد شماره r ، سپس به آن به ترتیب مضرب های صحیح k را اضافه می کنیم، تا واحدهای بعدی نیز مشخص شوند، این کار را تا قبل از اینکه شمارش بزرگتر از N شود ادامه می دهیم. k را فاصله نمونه گیری می نامیم.

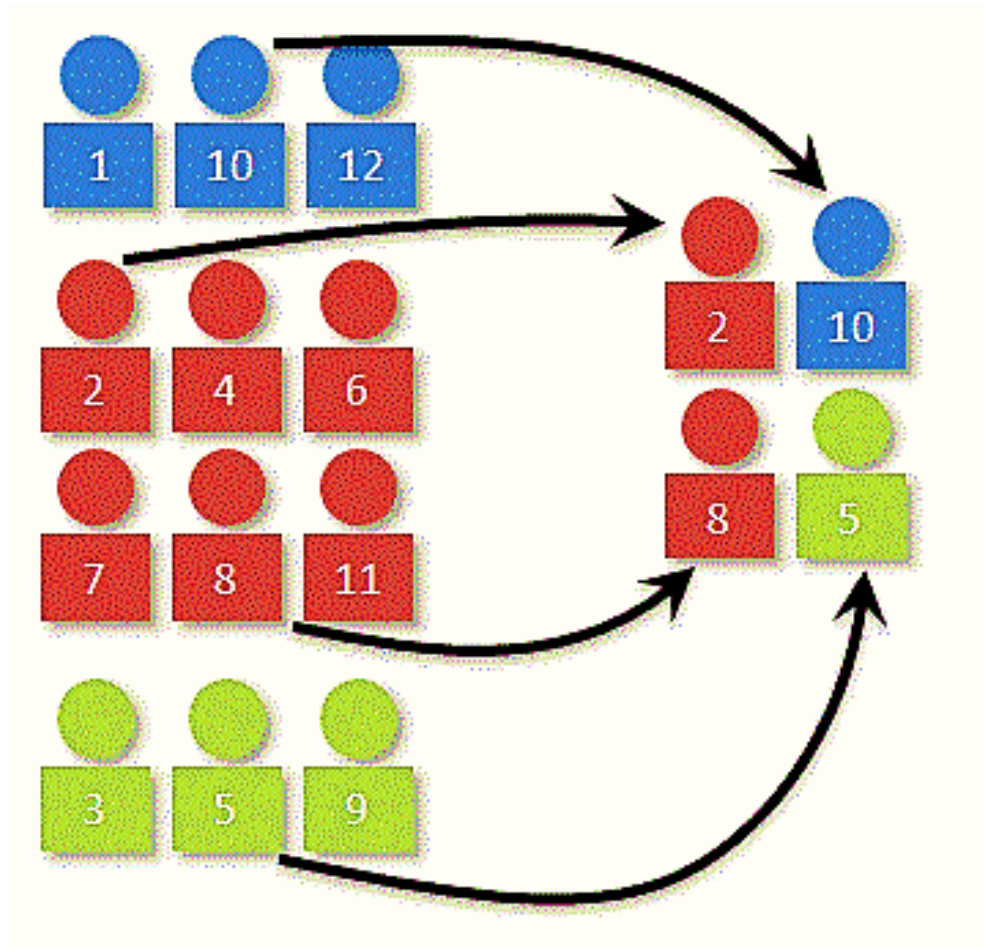


نمونه $\longrightarrow Y_r \quad Y_{r+k} \quad Y_{r+2k} \quad \dots Y_{r+jk} \quad \dots Y_{r+(n-1)k}$

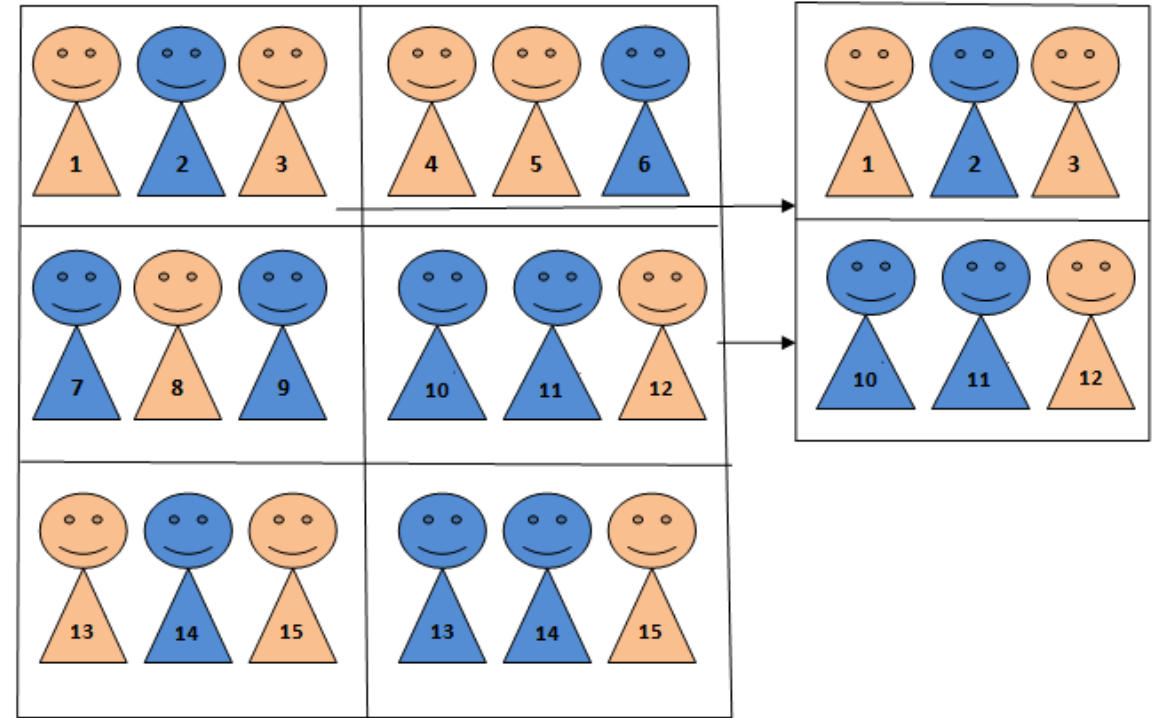
نمونه گیری خوشه ای



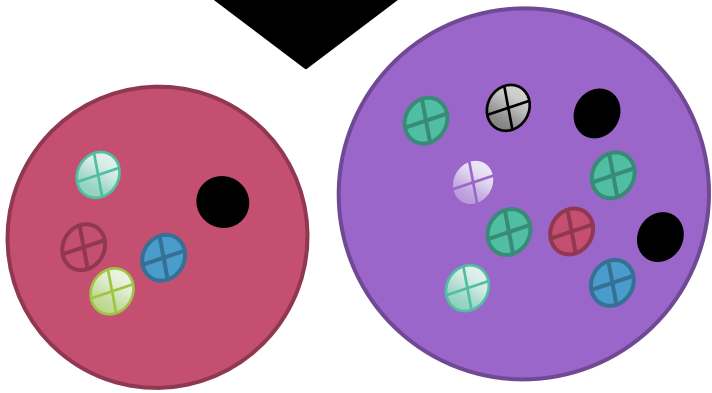
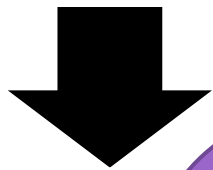
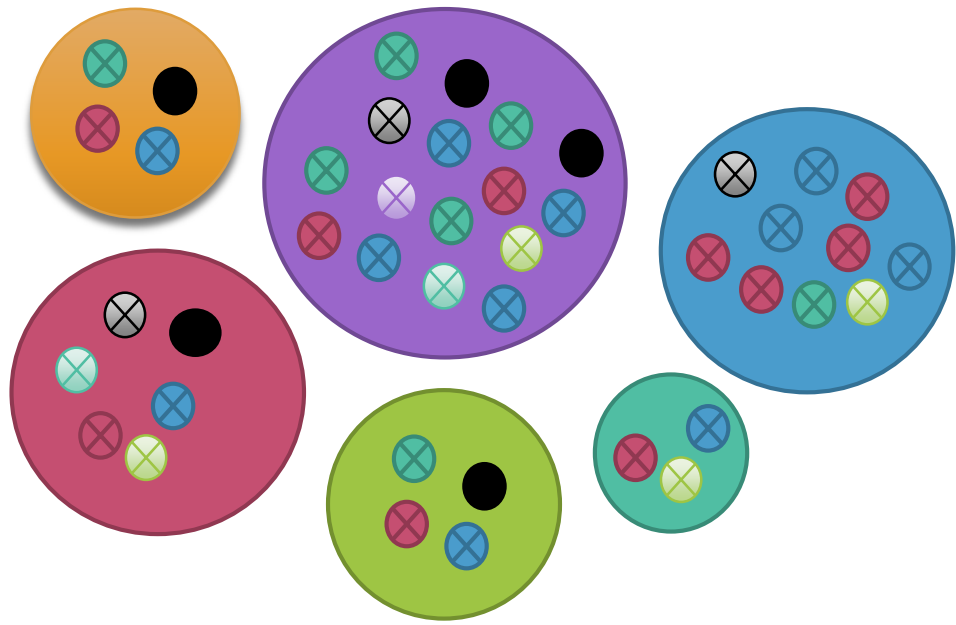
- هرگاه واحدهای نمونه به گونه ای انتخاب شوند که هر یک شامل تعدادی عنصر باشد، این واحد نمونه گیری خوشه نام دارد.
- چرا از نمونه گیری خوشه ای استفاده می کنیم؟
- در برخی طرح های نمونه گیری هزینه دستیابی به چارچوب نمونه گیری و شمارش عناصر بالاست.
- در برخی موارد چارچوب نمونه گیری مشخصی در دسترس نیست.
- در برخی موارد هزینه انجام مشاهدات و اندازه گیری ها با افزایش پراکندگی ها افزایش می یابد.



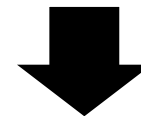
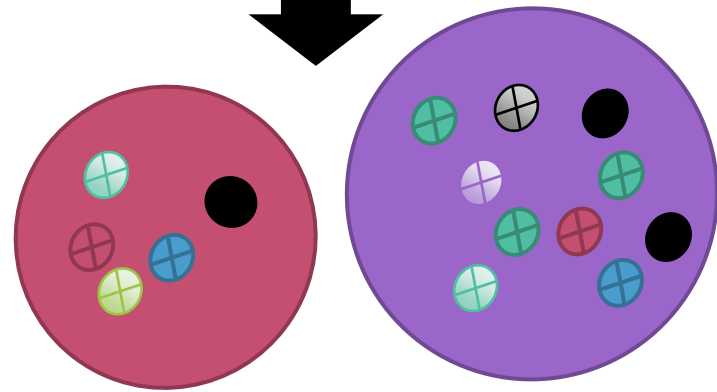
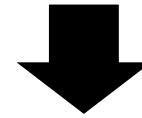
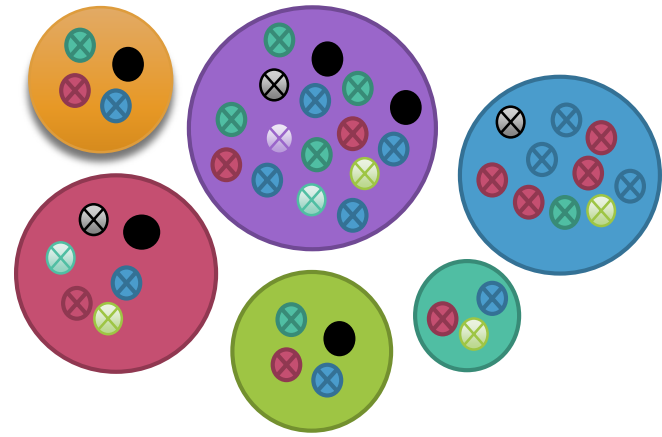
نمونه گیری تصادفی با طبقه بندی



نمونه گیری خوشه ای



نمونه گیری خوشه ای یک مرحله ای



نمونه گیری خوشه ای دو مرحله ای

خطای غیر نمونه گیری

Non-Sampling Errors

- عدم پاسخ به سؤالات پرسشنامه‌ها،
- پاسخ اشتباه،
- اثر مصاحبه کننده
- اثر پرسش بر پاسخگو،
- انتخاب نمونه بد که ناشی از چارچوب بد به دلیل انتخاب واحدهایی که جزء جمعیت نیستند ولی در چارچوب انتخاب شده‌اند،
- خطاهای پر کردن پرسشنامه مانند اشتباه انتخاب کردن یک گزینه (تیک زدن) در پرسشنامه و...

خطای نمونه گیری

Sampling Errors

- **تفاوت** بین مقدار بدست آمده از نمونه و مقدار واقعی یک پارامتر در جمعیت است.
- خطاهای نمونه گیری **حتی وقتی نمونه به روش صحیحی** انتخاب شده باشد ممکن است بوجود آیند .
- میزان خطای نمونه گیری بستگی به **حجم نمونه** دارد. نمونه بزرگتر خطای نمونه گیری کمتری دارد و این **بستگی به حجم جمعیت** ندارد.
- نکته مهم در مورد خطای نمونه گیری این است که به شرط این که روش نمونه گیری بر اساس انتخاب تصادفی انجام گرفته باشد امکان محاسبه احتمال خطا، برای حجم نمونه داده شده وجود دارد.

■ اگر جامعه ای به حجم ۵۰ نفر داشته باشیم و بخواهیم نمونه ای به حجم ۱۰ نفر از آن به طور تصادفی اختیار کنیم تعداد نمونه های ممکن **۱۰ ۲۷۲۲۷۸۱۷۰** است.

$$\binom{50}{10} = \frac{50!}{40! \times 10!} = 10,272,278,170$$

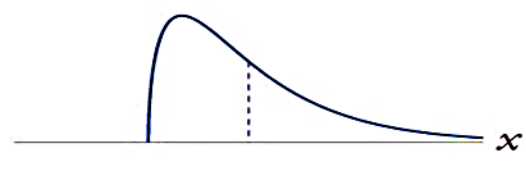
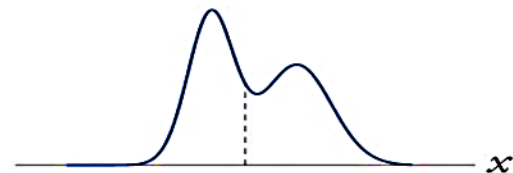
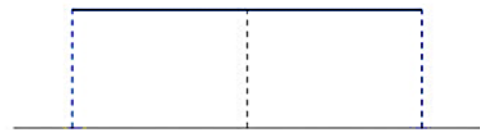
■ خطای نمونه گیری الزاماً نتیجه اشتباه در نمونه گیری نیست بلکه می تواند به دلیل اختلاف فاحش بین نمونه ها باشد.

■ **قضیه حد مرکزی** می توان برای توضیح این موضوع مناسب است.

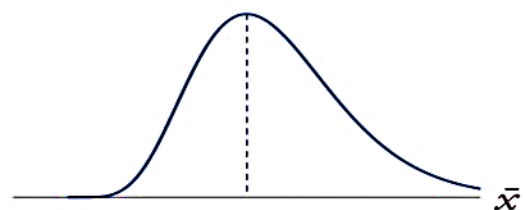
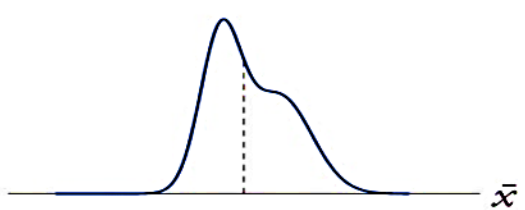
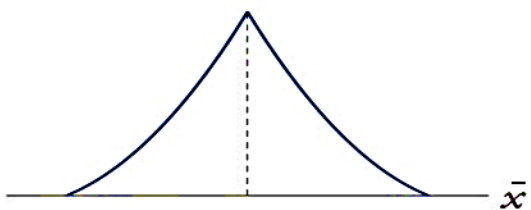
قضیه حد مرکزی central limit theorem

- اگر از یک جامعه به صورت تصادفی نمونه‌های زیاد، با اندازه‌های مساوی انتخاب کنیم، و میانگین این نمونه‌ها را محاسبه کنیم، توزیع فراوانی این میانگین‌ها یک توزیع طبیعی خواهد بود و میانگین میانگین‌ها انتخاب شده تقریباً برابر با میانگین جامعه خواهد بود.
- نرمال شدن توزیع میانگین‌ها از توزیع جامعه مستقل است.

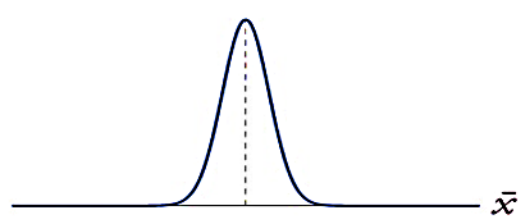
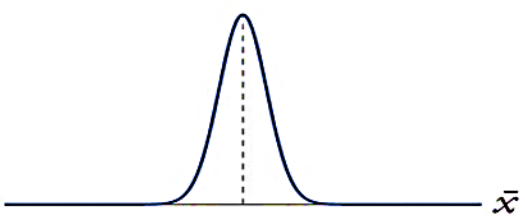
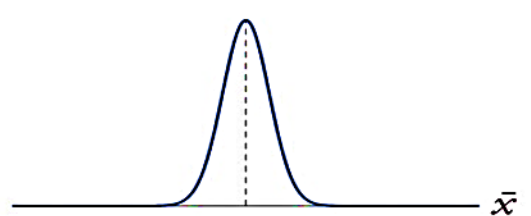
توزیع در جامعه

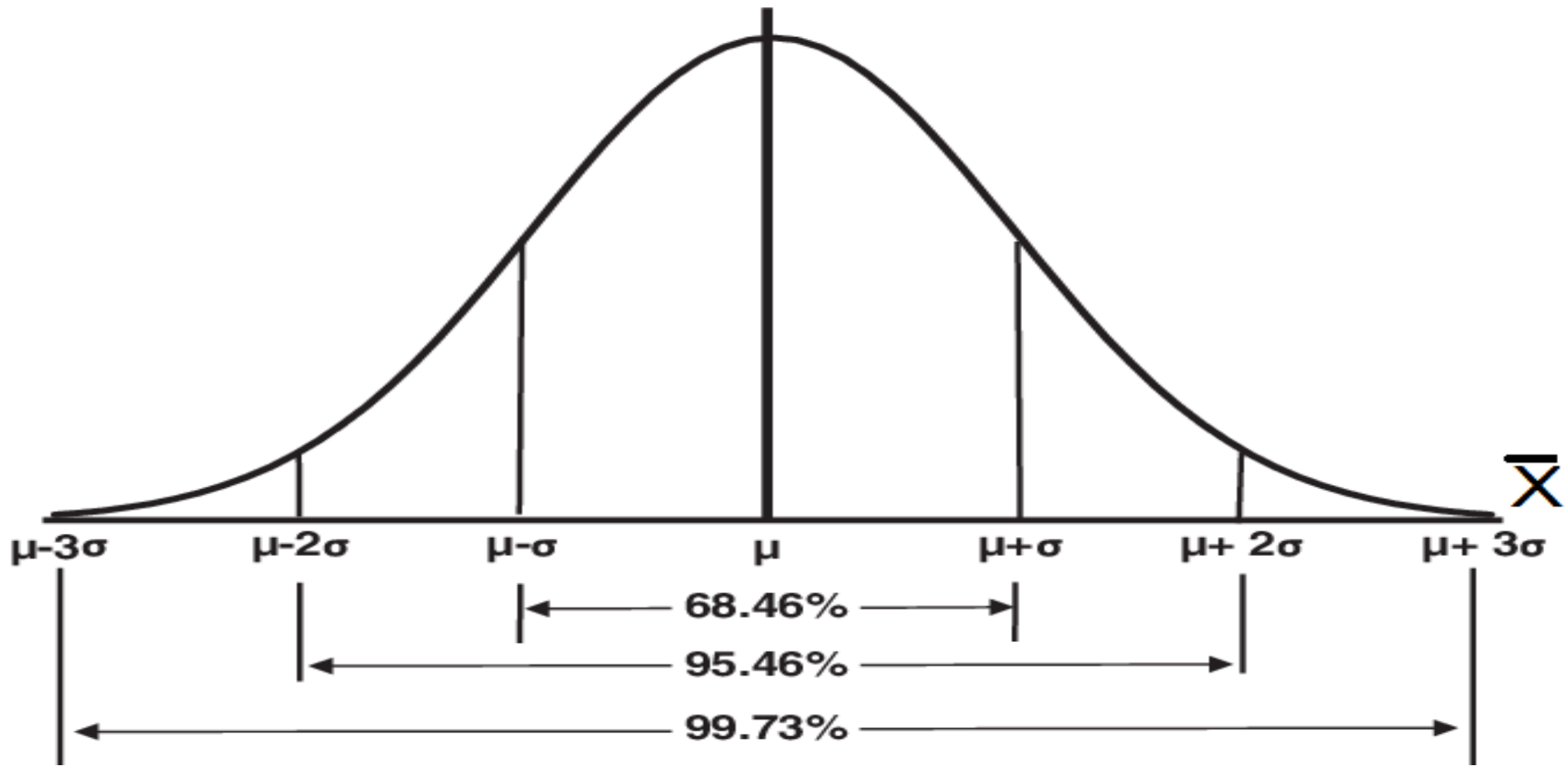


توزیع میانگین برای ۵ نمونه



توزیع میانگین برای ۳۰ نمونه





$$\bar{x} \sim N(\mu, SE)$$

میانگین جامعه یا همان میانگین میانگین ها

خطای نمونه گیری یا واریانس میانگین ها

اگر میانگین نمونه انتخابی را با \bar{X} و میانگین جامعه را با μ نمایش دهیم، گزاره های زیر در مورد میانگین جامعه برقرار است:

■ با ۶۸/۲۶ اطمینان می توان گفت که میانگین جامعه در بازه روبرو قرار دارد: $(\bar{x} - S_{\bar{x}}, \bar{x} + S_{\bar{x}})$

■ با ۹۹ درصد اطمینان میانگین جامعه در بازه روبرو قرار دارد: $(\bar{x} - ۲/۵۸ \times S_{\bar{x}}, \bar{x} + ۲/۵۸ \times S_{\bar{x}})$

■ با ۹۵ درصد اطمینان میانگین جامعه در بازه روبرو قرار دارد: $(\bar{x} - ۱/۹۶ \times S_{\bar{x}}, \bar{x} + ۱/۹۶ \times S_{\bar{x}})$

$$S_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}}$$

$$\hat{S}_{\bar{x}} = \frac{s_X}{\sqrt{n}}$$

$$\hat{S}_{md} = \frac{1/\sqrt{203} s_X}{\sqrt{n}}$$

$$\hat{S}_p = \sqrt{\frac{pq}{n}} \qquad \hat{S}_p = 1 \cdot \sqrt{\frac{pq}{n}}$$

$$\hat{S}_s = \frac{s}{\sqrt{2n}}$$

$$\hat{S}_f = \sqrt{npq}$$

برآورد و برآوردگر Estimation & Estimator

- به مشخصه مربوط به جامعه **پارامتر** گفته می شود. پارامترها مقادیر **ثابت** و **مجهول** هستند.
- به مشخصه مربوط به نمونه **آماره** گفته می شود، آماره متغیر تصادفی است که می تواند مقادیر **زیادی** را اختیار کند.
- مقدار پارامتر هر جامعه از طریق آماره نمونه **برآورد** می شود.
- برای محاسبه آماره باید از جامعه **نمونه گیری** کرد.

به آماره برآوردگر نیز گفته می شود.

ویژگی های برآوردگر

■ برآوردگر باید **نا اریب** باشد و **بدون سوگیری** پارامتر جامعه را برآورد کند. unbiased

اگر پارامتر مجهول جامعه θ باشد و از طریق مشاهده و نمونه گیری برای آن برآورد کننده $\hat{\theta}$ را بدست آورده باشیم. اگر میانگین برآورد کننده (یعنی میانگین آماره همه نمونه ها) با پارامتر جامعه برابر باشد، آنگاه $\hat{\theta}$ برآورد کننده نااریب برای θ است.

■ برآوردگر باید **سازگار** باشد. Contingency

اگر پارامتر مجهول جامعه θ باشد و از طریق مشاهده و نمونه گیری برای آن برآورد کننده $\hat{\theta}$ را بدست آورده باشیم. **انتظار داریم که برآورد به مقدار واقعی نزدیک باشد.** هر گاه در جامعه های نامتناهی حجم نمونه به سمت بی نهایت و یا در جامعه های متناهی حجم نمونه به سمت حجم جامعه میل کند و مقدار برآورد به سمت پارامتر جامعه میل کند، آنگاه برآورد کننده **سازگار** خوانده می شود.

▪ برآورد کننده باید **کارآمد** باشد. Efficiency

هرگاه برای پارامتر مجهول جامعه از روی یک نمونه به ترتیب دو برآورد کننده θ_1 و θ_2 را برآورد کرده باشیم. و $V(\hat{\theta}_1) \leq V(\hat{\theta}_2)$ آنگاه ضریب کارایی e را چنین تعریف می کنیم:

$$e(\theta_2, \theta_1) = \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}$$

$$0 \leq e \leq 1$$

برآورد کننده ای که واریانس کوچکتری داشته باشد کارآتر خواهد بود. در برخی مواقع کارایی را به صورت $e(\theta_2, \theta_1) = V(\hat{\theta}_1) - V(\hat{\theta}_2)$ تعریف می شود، در صورت مثبت بودن عبارت برآوردگر دوم کارآتر خواهد بود.

مثال ص ۲۹۰ و خود آزمایی ۲-۱۲