بسم الله الرحمن الرحیم

دانشگاه شهید چمران اهواز

روشهای نمونه گیری

مجتبی جهانی فر
دانشکده روانشناسی و علوم تربیتی
نیم سال اول سال تحصیلی ۹۹–۹۸

درس دوم

مرور برخی مفاهیم آماری
(مورد استفاده در نظریه نمونه گیری)

مقدار مورد انتظار(امید ریاضی) یا مقدار چشم داشتی
Expected value

- مقادیر عددی را که برای خلاصه کردن مشخصات یک جامعه به کار می روند مقادیر مورد انتظار می گویند.
- هر گاه **x** متغیری تصادفی با مقادیر $x_1, x_2, \ldots, x_k$ باشد. و احتمال متناظر با هر یک از مقادیر آن به صورت $p_1, p_2, \ldots, p_k$ تعریف شده باشد ، آنگاه امید ریاضی متغیر تصادفی **x** به صورت زیر تعریف می شود:

$$E(X) = \sum_{1}^{k} p_i x_i = \overline{X}$$

$$p_i = \frac{1}{k} \Rightarrow E(X) = \frac{1}{k} \sum_{1}^{k} x_i$$

برخی ویژگی های مهم امید ریاضی

- اگر مقدارهای **a** و **b** ثابت باشند :

$$E(aX+b)=aE(X)+b$$

- اگر **x** متغیر تصادفی باشد:

$$E(X_۱+X_۲+\dots+X_n)=E(X_۱)+E(X_۲)+\dots+E(X_n)$$

واریانس variance

- تغییر پذیری مقادیر در یک جامعه آماری را می توان با واریانس اندازه گرفت.
- اگر **x** متغیر تصادفی و **E(X)** امید ریاضی آن باشد :

$$V(X) = \sum_{۱}^{k} p_i (x_i - \overline{x})^۲ = E[(x - \overline{x})^۲]$$

$$p_i = \frac{۱}{k} \Rightarrow V(X) = \frac{۱}{k} \sum_{۱}^{k} (x_i - \overline{x})^۲$$

برخی ویژگی های مهم واریانس

- اگر **a** و **b** مقادیر ثابت باشند:

$$V(X+b)=V(X)$$

$$V(aX+b)=a^2V(X)$$

- انحراف استاندارد ریشه دوم واریانس است :

$$\sigma(X)=\sqrt{V(X)}$$

- اگر متغیرهای تصادفی **X** دو به از هم مستقل باشند:

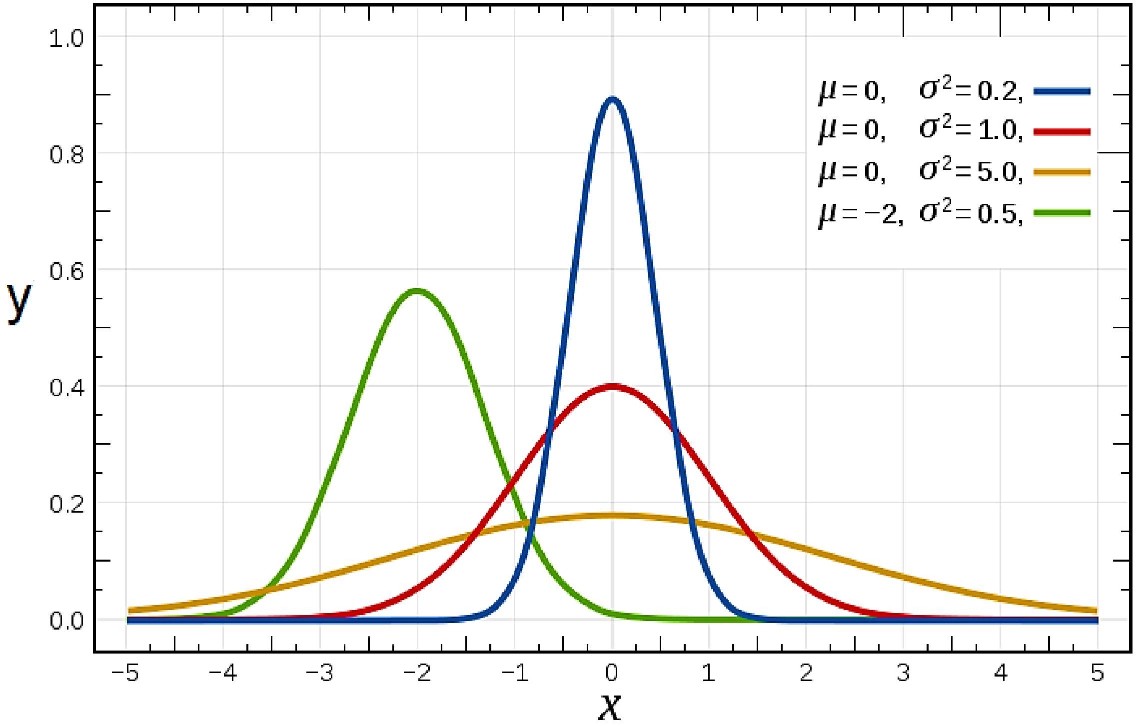
$$V(X_1+X_2+\dots+X_n)=V(X_1)+V(X_2)+\dots+V(X_n)$$

منحنی طبیعی و منحنی طبیعی استاندارد

- یکی از جالب ترین اکتشافات بشر، تعیین رابطه بین اندازه گیری پدیده های طبیعی و قوانین احتمال است.
- توزیع فراوانی بسیاری از پدیده های طبیعی شکلی شبیه زنگوله دارند.
- توزیع فراوانی نمره های بسیاری از آزمون های روان شناختی هم تقریبا زنگوله مانند است.
- این شکل زنگوله ای همان منحنی طبیعی یا خم نرمال است که رابطه و شکل آن به صورت زیر است:

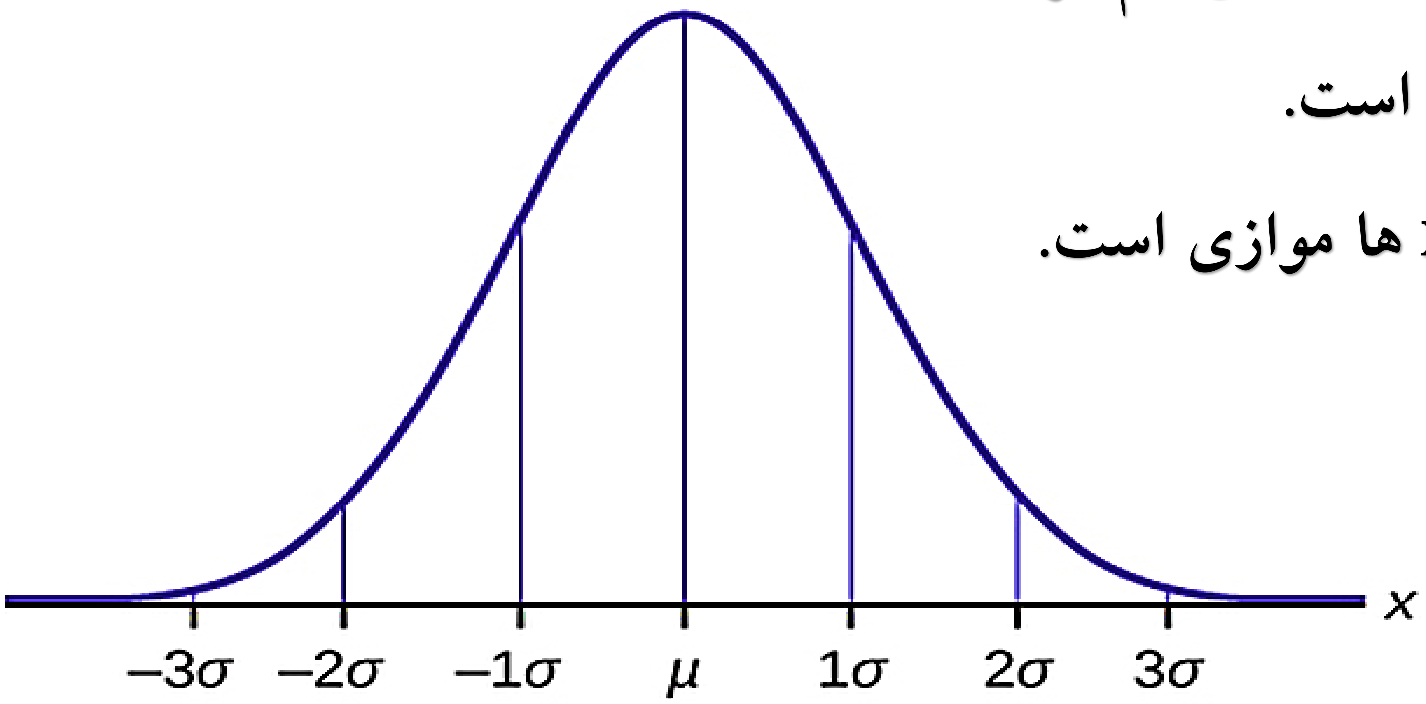
$$y = \frac{۱}{\sqrt{۲\pi}} e^{[-\frac{(x-\mu)^{۲}}{۲\sigma^{۲}}]}$$

$$y \sim N(\mu, \sigma)$$



ویژگی های منحنی طبیعی

- منحنی طبیعی متقارن است و حداکثر ارتفاع آن در میانگین قرار دارد.
- در منحنی طبیعی میانگین، میانه و نما روی هم قرار دارند.
- این منحنی دارای دو نقطه عطف است.
- دنباله های این منحنی با محور **x** ها موازی است.



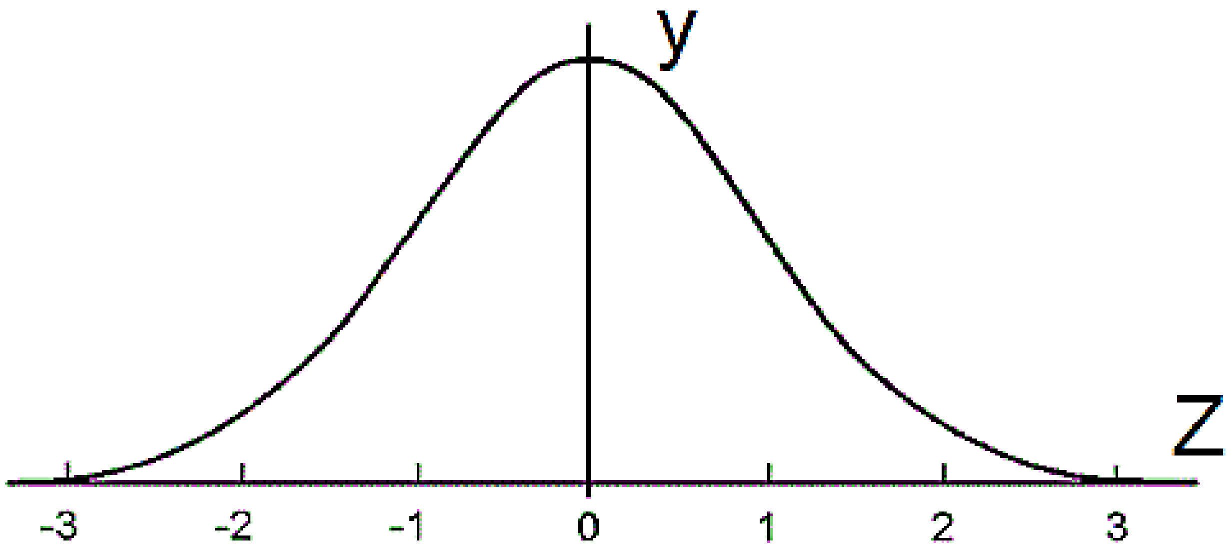
منحنی طبیعی استاندارد

- هرگاه میانگین و انحراف استاندارد منحنی طبیعی به ترتیب به ۰ و ۱ برده شوند، یا به عبارتی نمره ها استاندارد شوند، آنگاه منحنی طبیعی نیز به منحنی طبیعی استاندارد تبدیل می شود.

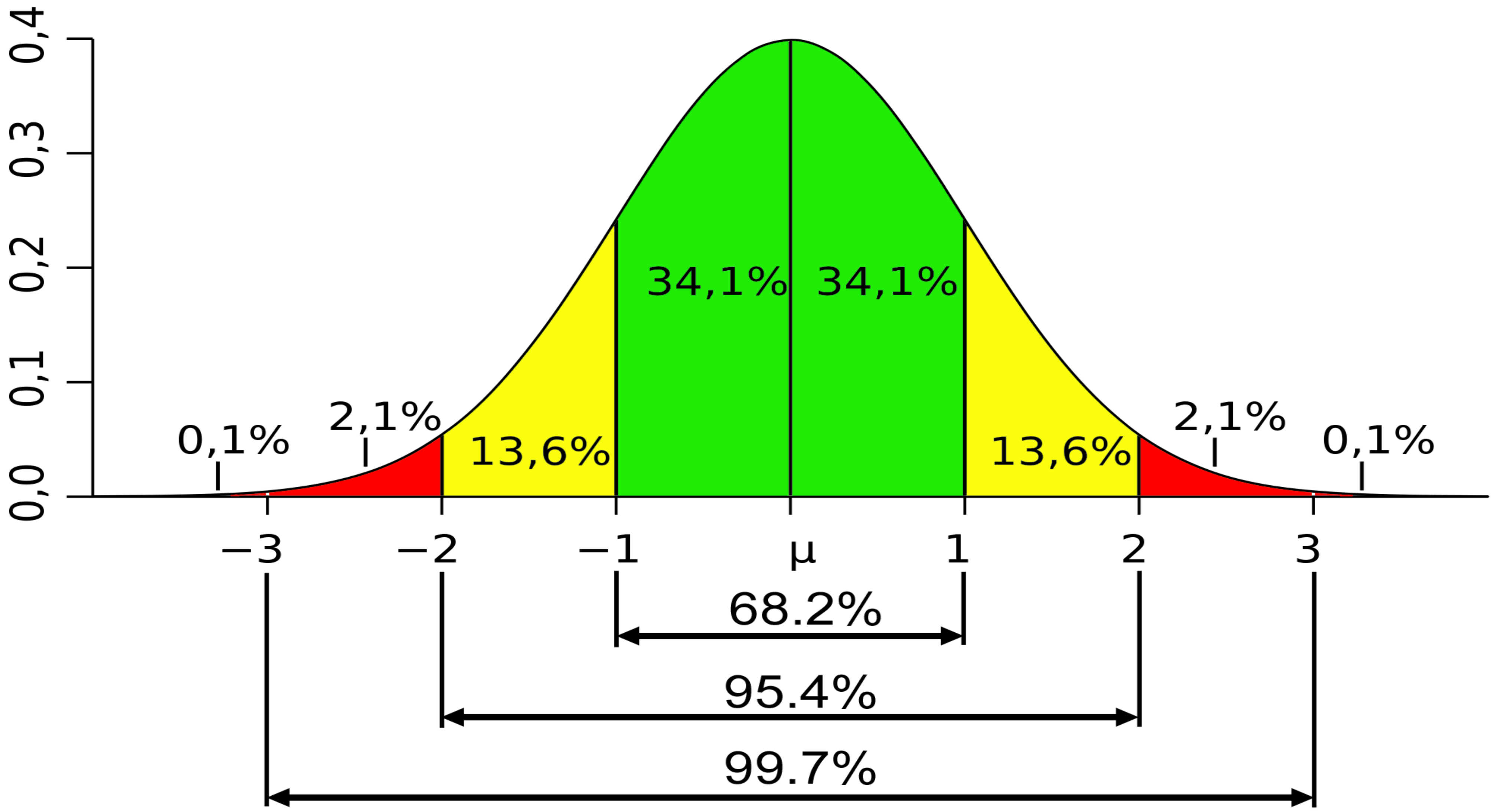
$$y = \frac{۱}{\sqrt{۲\pi}} e^{[-\frac{Z^{۲}}{۲}]}$$

$$z = \frac{x - \mu}{\sigma}$$

$$y \sim N(۰,۱)$$



استفاده از منحنی طبیعی استاندارد



خطا ی 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 پرشسنامه و...

خطا ی نمونه گیری
Sampling Errors

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

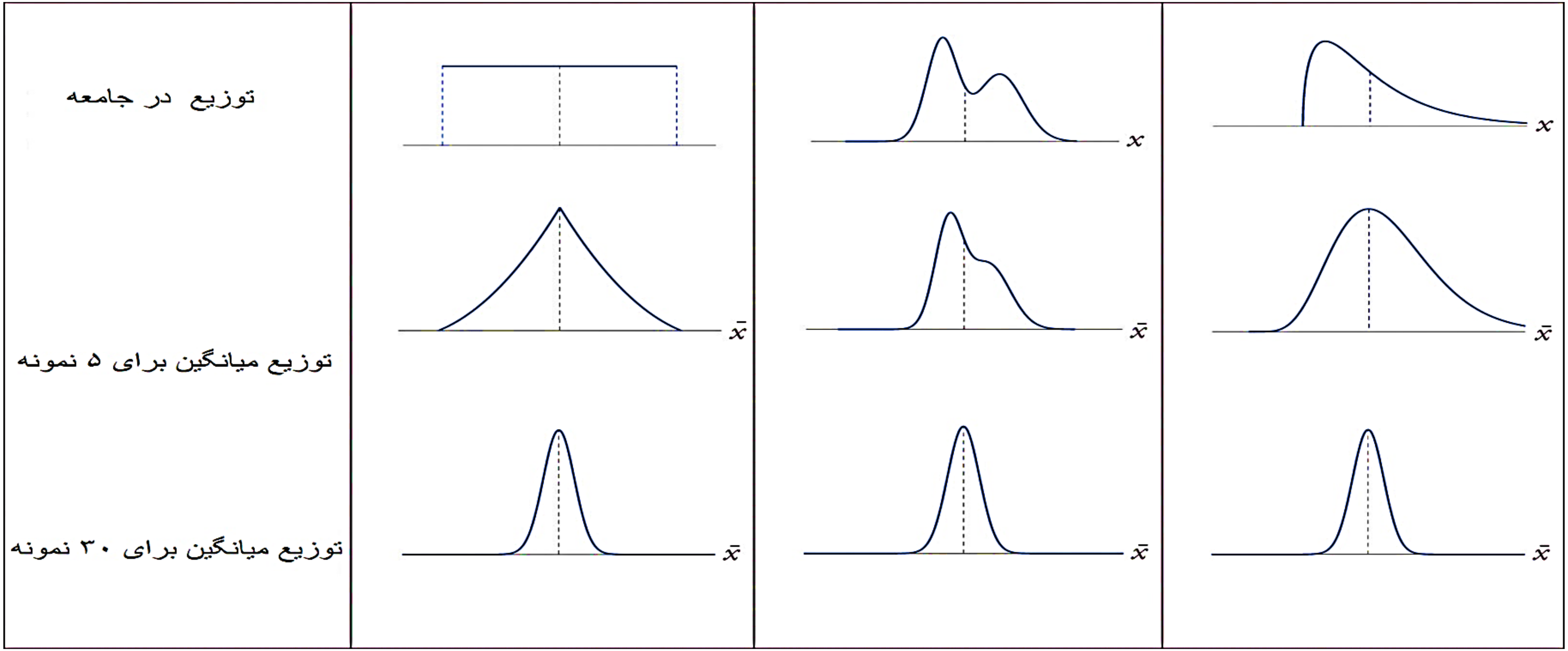
- اگر جامعه ای به حجم ۵۰ نفر داشته باشیم و بخواهیم نمونه ای به حجم ۱۰ نفر از آن به طور تصادفی اختیار کنیم تعداد نمونه های ممکن ۱۰۲۷۲۲۷۸۱۷۰ است.

$$\binom{۵۰}{۱۰} = \frac{۵۰!}{۴۰! \times ۱۰!} = ۱۰۲۷۲۲۷۸۱۷۰$$

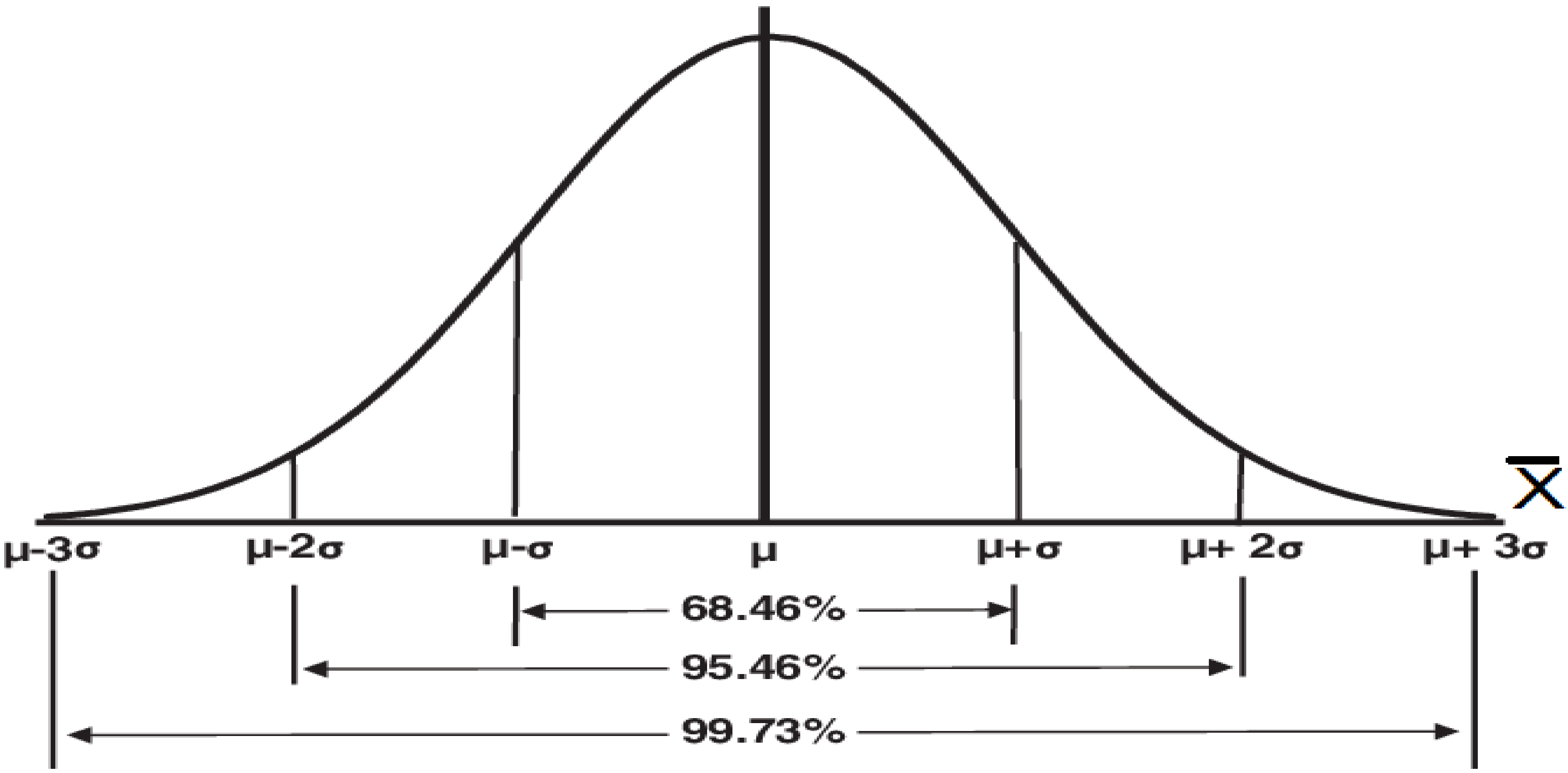
- خطای نمونه گیری الزاماً نتیجه اشتباه در نمونه گیری نیست بلکه می تواند به دلیل اختلاف فاحش بین نمونه ها باشد.
- **قضیه حد مرکزی** می توان برای توضیح این موضوع مناسب است.

قضیه حد مرکزی
central limit theorem

- اگر از یک جامعه به صورت تصادفی نمونه های زیاد، با اندازه های مساوی انتخاب کنیم، و میانگین این نمونه ها را محاسبه کنیم، توزیع فراوانی این میانگین ها یک توزیع طبیعی خواهد بود و میانگین میانگین های انتخاب شده تقریبا برابر با میانگین جامعه خواهد بود.
- نرمال شدن توزیع میانگین ها از توزیع جامعه مستقل است.



توزیع در جامعه
توزیع میانگین برای ۵ نمونه
توزیع میانگین برای ۳۰ نمونه
x
$\bar{x}$



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

میانگین جامعه یا همان میانگین میانگین ها

خطای نمونه گیری یا واریانس میانگین ها

اگر میانگین نمونه انتخابی را با $\overline{X}$ 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: $(\bar{x}-SE, \bar{x}+SE)$
- با ۹۹ درصد اطمینان میانگین جامعه در بازه روبرو قرار دارد: $(\bar{x}-۲/۵۸\times SE, \bar{x}+۲/۵۸\times SE)$
- با ۹۵ درصد اطمینان میانگین جامعه در بازه روبرو قرار دارد: $(\bar{x}-۱/۹۶\times SE, \bar{x}+۱/۹۶\times SE)$

روش های نمونه گیری متفاوت وجود دارد که هر کدام دارای SE مربوط به خودش است و هنگام معرفی هر روش نمونه گیری SE مربوط به آن هم معرفی خواهد شد.

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.

$$E(\hat{\theta}) = \theta$$

- 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 نا متناهی 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 $V(\hat{\theta}_2) \leq V(\hat{\theta}_1)$ 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.

تکالیف

تمرین های منتخب **فصل دوم کتاب آمارگیری نمونه ای**