مقدمه ای بر تحلیل آماری با نرم افزار R

مدرس:
پیمان نیک چی
فارغ التحصیل کارشناسی ارشد آمار از دانشگاه تهران

سرفصل مباحث جلسه ششم

- قضیه حد مرکزی
- آزمون فرض
- آزمون Z
- آزمون t
- لزوم نرمال سازی داده ها

مقدمه

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2) \quad \mu = \mu_X \quad \sigma^2 = \sigma_X^2$$

$$\Rightarrow \quad \overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

قضیه حد مرکزی

$$X_1, X_2, \ldots, X_n \qquad \mu = \mu_X \qquad \sigma^2 = \sigma_X^2$$

$$\xrightarrow{\text{CLT}} \qquad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

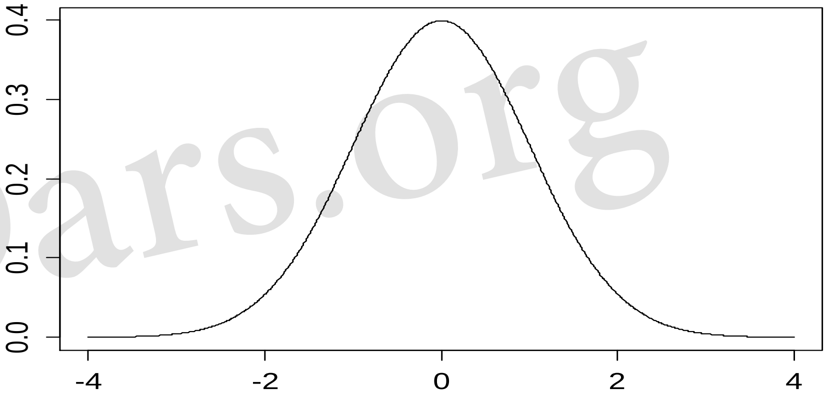
قضیه حد مرکزی

$$X_1, X_2, \ldots, X_n \xrightarrow{\text{CLT}} \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$P(\overline{X} > a)$$

$$= P\left(\frac{\overline{X} - \mu}{\sigma} > \frac{a - \mu}{\sigma}\right)$$

$$= P\left(Z > \frac{a - \mu}{\sigma}\right)$$



آزمون فرض

- در علم آمار، آزمون فرض یک روش برای بررسی ادعا یا فرض های انجام شده درباره پارامترهای نامعلوم جامعه است.
- در این روش یک فرض در مقابل فرض دیگر قرار می گیرد و صحت آنها نسبت به همدیگر مشخص می شود.
- بررسی این که کدام فرضیه صحیح است بر اساس نمونه گیری از جامعه صورت می گیرد.

یک مثال از آزمون فرض

- فرض کنید در یک تحقیق علاقه مند به بررسی قد مردان و زنان در جامعه هستیم. به طور مشخص سوال فرضیه به این صورت بیان می شود:

آیا تفاوت معناداری بین میانگین قد در بین مردان و زنان یک جامعه از قبل تعیین شده وجود دارد؟

آزمون Z

$$X_1, X_2, \ldots, X_n \sim N(\mu_1, \sigma_1^2) \Rightarrow \overline{X} \sim N(\mu_1, \frac{\sigma_1^2}{n})$$

$$Y_1, Y_2, \ldots, Y_m \sim N(\mu_2, \sigma_2^2) \Rightarrow \overline{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m})$$

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1)$$

آزمون t

$$X_1, X_2, \ldots, X_n \sim N(\mu_1, \sigma_1^2) \Rightarrow \overline{X} \sim N(\mu_1, \frac{\sigma_1^2}{n}) \qquad \sigma_1^2 = ?$$

$$Y_1, Y_2, \ldots, Y_m \sim N(\mu_2, \sigma_2^2) \Rightarrow \overline{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m}) \qquad \sigma_2^2 = ?$$

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2$$

$$\Rightarrow t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \underset{m \to \infty}{\overset{n \to \infty}{\sim}} N(0,1)$$

آزمون t

$$X_1, X_2, \ldots, X_n \sim N(\mu_1, \sigma_1^2) \Rightarrow \overline{X} \sim N(\mu_1, \frac{\sigma_1^2}{n}) \quad \sigma_1^2 = ?$$

$$Y_1, Y_2, \ldots, Y_m \sim N(\mu_2, \sigma_2^2) \Rightarrow \overline{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m}) \quad \sigma_2^2 = ?$$

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2$$

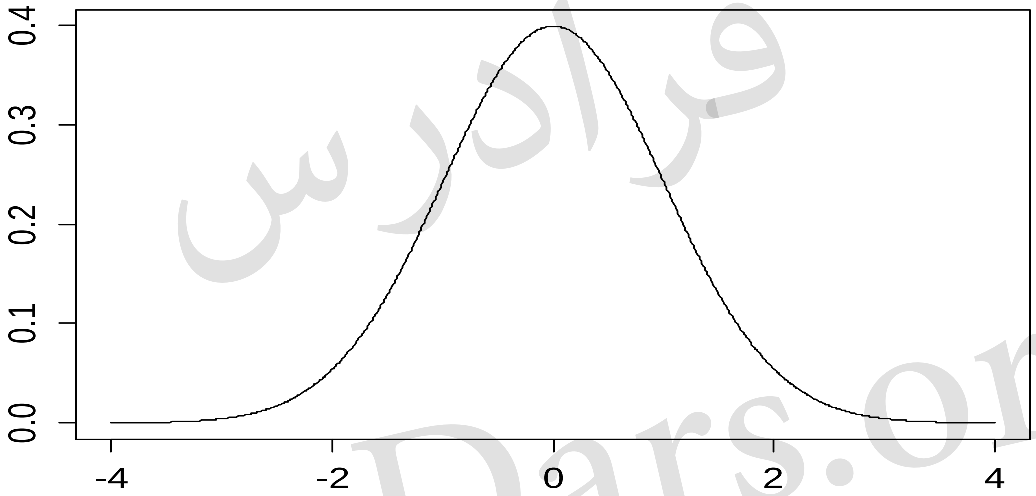
$$\Rightarrow S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

$$\sigma_1^2 = \sigma_2^2$$

آزمون t

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right) S_p^2}} \sim t(n+m-2)$$

پی مقدار



$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1)$$

لزوم نرمال سازی داده ها

- در تمامی موارد بررسی شده توزیع اصلی جامعه نرمال بود.

- در مواردی که حجم نمونه زیاد باشد، می توان به طور تقریبی نرمال بودن توزیع میانگین را فرض کرد، اما در بسیاری از تحقیق ها به علت هزینه بر بودن، امکان اتخاذ حجم نمونه بالا امکان پذیر نیست.

- در اغلب تحلیل های زیستی حجم نمونه کم است و لازم است قبل از انجام تحلیل های مرتبط با آزمون t ، داده ها نرمال شوند.

این اسلاید ها بر مبنای نکات مطرح شده در فرادرس

«آموزش برنامه نویسی R و نرم افزار R Studio»

تهیه شده است.

برای کسب اطلاعات بیشتر در مورد این آموزش به لینک زیر مراجعه نمایید

faradars.org/fvr9311