X is a random variable.
with any distribution.

Central limit theorem

The sum of X (many many independent and identically distributed X) will become normal as the number of X goes larger.

---

rule of thumb: when $n > 30$ the sum of X is already close to normal.

→ 记.

Sampling distribution : distribution of $\bar{x}$, $\hat{p}$

sample mean

sample proportion.

For $\bar{x}$ sample mean.

1) assume sample size $n \geq 30$

2) assume sample size $n < 0.1N$

Because $n < 0.1N$. we have independent sampling.

Because $n \geq 30$ we can use C.L.T.

$Y = X_1 + \cdots + X_n \sim N$ Because of C.L.T.

$\bar{X} = \frac{Y}{n} \sim N$

$$E(\bar{X}) = E\left(\frac{Y}{n}\right)$$

$$= \frac{1}{n} E(Y)$$

$$= \frac{1}{n} \cdot n E(X)$$

Sample mean is normal because of CLT.

individual's distribution is unknow in shape.

$$E(\bar{X}) = E(X)$$

Sample mean's center of distribution

population center of distribution

$$\mathrm{Std}(\bar{X}) = \mathrm{Std}\left(\frac{Y}{n}\right) = \frac{1}{n}\mathrm{Std}(Y)$$

$$= \frac{1}{n}\cdot\sqrt{n}\,\mathrm{Std}(X)$$

$$= \frac{\mathrm{Std}(X)}{\sqrt{n}}$$

当 population 为正态时，$X_1, \cdots X_n$ 天然为正态，

$Y = X_1 + \cdots X_n$ 也为正态，无须 $n \geq 30$

$\bar{X} = \frac{Y}{n} \sim N$.

如何可知 population 为正态

① Sample 无 skewness、无 outlier | 常考.

⇓

② use Q-Q plot. | 很少考.

⇓

不考. | ③. Shapiro-Wilk & Kolmogorov-Smirnov.

PBL用.

For sample proportion. $\hat{p}$

$\hat{p} = \frac{Y}{n}$. Y is → Binomial distribution

→ Sum of Bernouli distribution

已知当 $np \geq 10$ 且 $n(1-p) \geq 10$ 时.

$Y \sim Binom(n, p) \Rightarrow Y \sim Normal(np, \sqrt{np(1-p)})$

$\hat{p} \sim Normal$.

$$E(\hat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = p.$$

Mean of Sample proportion

population ~~proportion~~ prob.

$$Std(\hat{p}) = Std\left(\frac{Y}{n}\right) = \frac{1}{n}Std(Y)$$

$$= \frac{1}{n}\sqrt{np(1-p)}$$

$$= \sqrt{\frac{p(1-p)}{n}}$$

$$n\uparrow \Rightarrow Std(\hat{p})\downarrow.$$