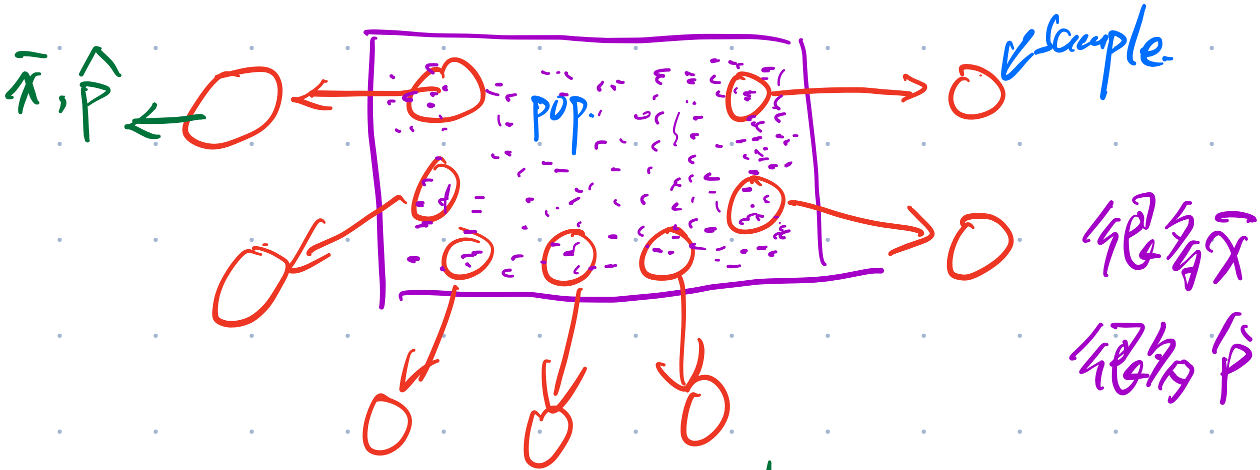

quantitative data $\xrightarrow{\text{calc}}$ sample mean $\bar{x}$

height — mean height $\bar{x}$

categorical data $\xrightarrow{\text{calc}}$ sample proportion

gender — distribution of female proportion $\hat{p}$

样本统计量 Sample statistics

Sampling distribution

抽样分布

distribution of individual data point

Include all individual ⇒ population distribution

Include a small number of individual ⇒ sample distribution

distribution of sample statistics

Include s.s. from all possible sample

在满足特殊条件的情况下,

$\bar{x}$ 和 $\hat{p}$ 呈现 特殊分布

利用这种特殊分布,做推断.

statististical inference

对于 $\hat{p}$

当：1) SRS. 样本被随机抽出.

$\Downarrow$

unbiasness 无偏性.

$\Downarrow$

$E(\hat{p}) = p$. ← 全体百分比

样本百分比的均值

当 $n < 0.1N$: 10% condition.

$$\Downarrow$$

Independent sampling

$$\Downarrow$$

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

当 $np \geq 10$ 且 $n(1-p) \geq 10$ ｜ Large count condition

$\hat{p} \sim Binom \Rightarrow \hat{p} \sim N$

当 1) SRS. 2) 10% condition 3) Large count

三条件满足时.

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

当 1) SRS $\Rightarrow \mu_{\bar{x}} = \mu_x$.

sample size

pop. size

2) $n < 0.1N \Rightarrow \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$.

3) $n \geq 30$

或

population 是 normal

$\rightarrow \bar{x} \sim$ 正.

成立时.

$$\bar{x} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$$

当 $n < 0.1N$. ⇒ independent sample.

$$\bar{X} = \frac{\sum X_i}{n}.$$

$$Var(\bar{X}) = Var\left(\frac{\sum X_i}{n}\right)$$

$$= \frac{1}{n^2} Var(\sum X_i)$$

$$= \frac{1}{n^2}\left(Var(X_1) + Var(X_2) + \cdots + Var(X_n)\right)$$

$$= \frac{1}{n^2} \cdot n\,Var(X)$$

$$= \frac{Var(X)}{n}$$

$$std(\bar{X}) = \sqrt{\frac{std^2(X)}{n}} = \frac{std(X)}{n}$$

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}.$$

$$\bar{X} = \frac{\sum X_i}{n} \sim N$$

$\sum X_i \sim N$

X_i independent identically distribution

当 $n \geq 30$.

or

population is normally distributed

$$X_i \sim N \Rightarrow \sum X_i \sim N \Rightarrow \frac{\sum X_i}{n} \sim N$$

$$\bar{x} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$$

对于 $\hat{p}$.　① SRS $\Rightarrow E(\hat{p}) = p$.

$$\hat{p} = \frac{X}{n}$$　X 成功的个数.

$X \sim B(n, p)$

$\downarrow$ $np \geq 10$　$n(1-p) \geq 10$

$E(X) = np$

$std(X) = \sqrt{np(1-p)}$

山$\hat{p}$八

$$E(\hat{p}) = \frac{E(X)}{n} = p$$

样本百分比

$$\mathrm{std}(\hat{p}) = \mathrm{std}\left(\frac{X}{n}\right) = \frac{1}{n}\mathrm{std}(X)$$

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

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

当 $np \geq 10$, $n(1-p) \geq 10$

$$X \sim B \Rightarrow X \sim N$$

$$\hat{p} = \frac{X}{n} \sim N \quad \hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$p = 0.80$ $N = 800$. $n = 60$. 0.52.

(1) 定假设.

Hypothesis Testing

假设检验

(2). 验条件.

(3) 算 P-值.

(4). 做判断.

1. 定假设

他觉得 $p>0.8$

你觉得 $p<0.8$

先定H_A $\Rightarrow$ $H_a: p<0.8$

↓

再定H_0 $\Rightarrow$ $H_0: p=0.8$ 将不等号改等号

通过假定 H_0成立时出现了不该出现的现象，推翻 H_0，证明 H_a

H_0 永远是"$=$"　H_a 是"$>$""$<$"或"$\neq$"

不能是"$=$"

2. 验条件

1) SRS. ✓

2) 10%. $n = 60 < 0.1N = 80$ ✓

3). $np = 60 \times 0.52 \geq 10$. ✓

$n(1-p) = 60 \times 0.48 \geq 10$.

3) 算 p-value. 假定 H_0 成立.

$$\hat{p} \sim N\left(\underline{\underline{0.8}},\ \sqrt{\frac{0.8\times 0.2}{60}}\right)$$

p-value是当H_0成立时, 取到该样本与比样本更符合H_a方向样本的概率.

$$P(\hat{p} < 0.52)$$

$$= \text{normalcdf}\left(\begin{array}{ll} \text{lower} & 0 \\ \text{upper} & 0.52 \\ \mu: & 0.8 \\ \sigma: & \sqrt{\frac{0.8\times 0.2}{60}} \end{array}\right)$$

$\approx 0.$