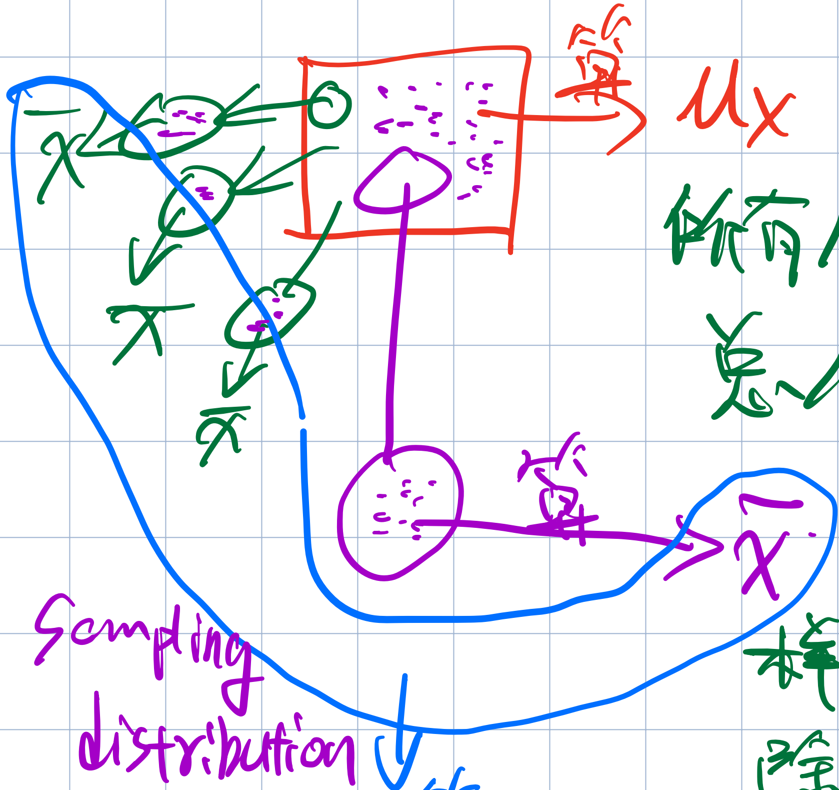
Sampling distribution.

抽样分布

Sampling distribution is distribution of sample statistics like $\bar{x}$, $\hat{p}$

看多个 $\bar{x}$ 的分布.



算 μ_x population mean.

所有人的身高加起来除以总人数

$\bar{x}$ sample mean

样本里的人身高加起来除以样本人数

算

$\mu_{\bar{x}}$ mean of sampling distribution population 的分布

eg. 全校学生身高分布.

sample 的分布

individual 值的分布

eg 某班学生的身高分布.

sample statistics 的分布

eg. 每个班级算个平均身高 $\bar{x}$, 所有 $\bar{x}$ 的分布.

当满足3条件时：

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

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

→ 记

记

使 $\bar{x} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$ 的3条件:

1) SRS. Sample were selected randomly

$\Rightarrow$ unbias $\Rightarrow$ $\mu_{\bar{x}} = E(\bar{x}) = \mu_x$

2) 10% condition: = $n < 0.1N$

$\Rightarrow$ independent sampling $\Rightarrow$ $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$

Large number condition / normal condition

3) $n \geq 30$ or population is normal

$\bar{x} \sim N$

22

使 $\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ 的 3 条件.

1). SRS. $\Rightarrow$ unbias $\Rightarrow$ $\mu_{\hat{p}} = p$.

2) 10% condition $\Rightarrow$ independent $\Rightarrow$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

3). Large count : $np \geq 10$ & $n(1-p) \geq 10$

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

不记

$\hat{p} = \dfrac{X}{n}$

X ← 属于目标类别的个数

n ← sample size.

$X \sim \text{Binomial}(n, p)$.

if unbias $E(X) = np$ $E(\hat{p}) = E\left(\dfrac{X}{n}\right) = \dfrac{np}{n}$

$= p$

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

$std\left(\dfrac{X}{n}\right) = \dfrac{\sqrt{np(1-p)}}{n} = \sqrt{\dfrac{p(1-p)}{n}}$

if $np \geq 10$, $n(1-p) \geq 10$.

$X \sim \text{Binomial} \Rightarrow X \sim N$

$\Rightarrow \hat{p} = \frac{X}{n} \sim N$

不记

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

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

if unbiased $= \frac{n\mu_X}{n} = \mu_X$

$$std(\bar{x}) = std\left(\frac{\sum X_i}{n}\right)$$

if independent $= \frac{1}{n}\sqrt{n}\, std(X)$

$$= \frac{\sigma_X}{\sqrt{n}}$$

忘记 if $n \geq 30$. Central limit Theorem

$$\Rightarrow \sum X_i \sim N$$

or

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

if $X \sim N$. $\sum X_i \sim N$

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