Survey Sampling

△ Sample $\downarrow n$ Statistics $\hat{\theta}(X_1, \cdots, X_n)$ $\xrightarrow{\text{estimate}}$ Population $\downarrow N$ parameter θ

△ Simple random sampling with / without replacement

with $\to$ i.i.d. X_i; without $\to$ dependent X_i

equal chance for each individual to be selected.

↳ E.g. Sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\downarrow$ estimate

population mean μ.

$E(\bar{X}) = \mu$, unbiased.

$$\mathrm{Var}(\bar{X}) = \begin{cases} \frac{\sigma^2}{n}, & \text{with replacement} \\ \frac{\sigma^2}{n}\underbrace{\left(1-\frac{n-1}{N-1}\right)}_{\text{finite population correction}}, & \text{without} \end{cases} \qquad (*)$$

The sampling distribution of $\bar{X}$:

In most cases, μ, σ^2 are unknown. We need to further estimate σ^2, so that to estimate $\mathrm{Var}(\bar{X})$.

△ Estimate the population variance σ^2

Consider

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \leftarrow \text{naive estimator of } \sigma^2$$

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n}E\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \\ &= \frac{1}{n}\sum_i E(X_i^2) - E(\bar{X}^2) \\ &= \frac{1}{n}\cdot n(\mu^2+\sigma^2) - [\mathrm{Var}(\bar{X}) + E(\bar{X})^2] \\ &= \mu^2 + \sigma^2 - [\mathrm{Var}(\bar{X}) + \mu^2] \\ &= \sigma^2 - \mathrm{Var}(\bar{X}) \\ &= \begin{cases} \sigma^2\left(1-\frac{1}{n}\right), & \text{with} \\ \sigma^2\left[1-\frac{1}{n}\cdot\left(1-\frac{n-1}{N-1}\right)\right], & \text{without.} \end{cases} \quad > \text{both biased} \end{aligned}$$

An unbiased estimator for $\boxed{\sigma^2}$:

$$(\triangle) \quad \begin{cases} \hat{\sigma}^2 \Big/ \left(1-\frac{1}{n}\right), & \text{with} \\ \hat{\sigma}^2 \Big/ \left[1-\frac{1}{n}\left(1-\frac{n-1}{N-1}\right)\right], & \text{without} \end{cases}$$

↳ in slides / book

An unbiased estimator for $\mathrm{Var}(\bar{X})$:

Subsitute $(\triangle)$ back into $(*)$ to obtain

$$S_{\bar{X}}^2 = \begin{cases} \left[\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}_{\text{estimate } \sigma^2}\right]\cdot\frac{1}{n}, & \text{with} \\ \underbrace{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]\cdot\frac{n(N-1)}{N(n-1)}}_{\text{unbiased estimator of } \sigma^2}\cdot\frac{1}{n}\left(1-\frac{n-1}{N-1}\right), & \text{without} \end{cases}$$

△ How "good" it is by using $\bar{X}$ to estimate μ?

Want to know $P(|\bar{X}-\mu| \leq \varepsilon)$ and have it stay small.

but unable to compute since the sampling distribution of $\bar{X}$ is unknown!

↳ Solution : CLT !

$$\triangle \quad \bar{X} \overset{\text{approx.}}{\sim} N(\mu, \sigma_{\bar{X}}^2)$$

$\sigma_{\bar{X}}^2$: $\frac{\sigma^2}{n}$; $\frac{\sigma^2}{n}\left(1-\frac{n-1}{N-1}\right)$

To construct a $100(1-\alpha)\%$ confidence interval for μ :

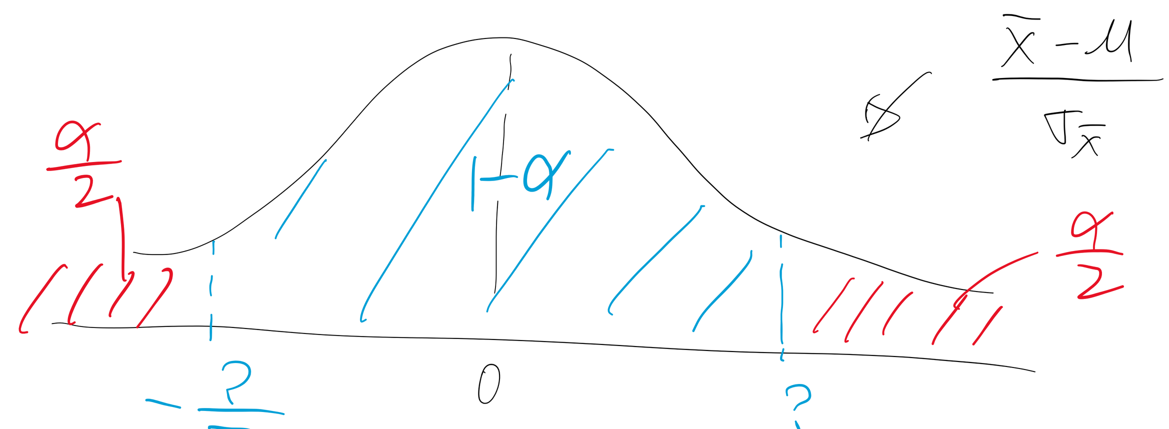
($\alpha = 0.1, 0.05, 0.01 \ldots$)

$$P(|\bar{X}-\mu| \leq ?) = 1-\alpha$$

↳ standadize

$$P\left(-\frac{?}{\sigma_{\bar{X}}} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq \frac{?}{\sigma_{\bar{X}}}\right) = 1-\alpha$$

$$\cancel{\phi} \quad \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \sim N(0,1)$$



$-\frac{?}{\sigma_{\bar{X}}} = -Z_{\frac{\alpha}{2}}$; $\frac{?}{\sigma_{\bar{X}}} = Z_{\frac{\alpha}{2}}$ ← notation

$Z_{\frac{\alpha}{2}}$: the number so that $\phi(Z_{\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$

$$? = Z_{\frac{\alpha}{2}}\cdot\sigma_{\bar{X}}$$

Thus, an approximate $100(1-\alpha)\%$ C.I. for μ

$$-Z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq Z_{\frac{\alpha}{2}} \iff \underbrace{\left(\bar{X} - Z_{\frac{\alpha}{2}}\sigma_{\bar{X}},\ \bar{X} + Z_{\frac{\alpha}{2}}\sigma_{\bar{X}}\right)}_{\text{random interval}}$$

random variable

i.e. $\bar{X} - Z_{\frac{\alpha}{2}}\sigma_{\bar{X}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}}\sigma_{\bar{X}}$