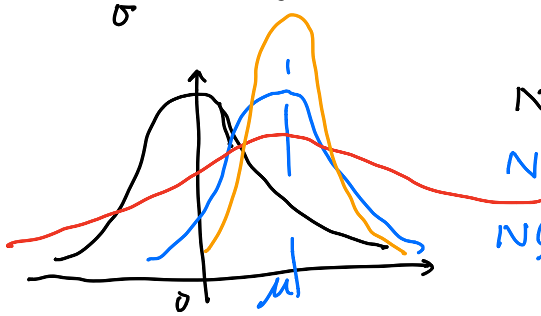
Common Distribution

I. χ^2, t, F

1. Normal. $X \sim N(\mu, \sigma^2)$ $\quad Z \sim N(0,1)$

$$\frac{X-\mu}{\sigma} = Z \sim N(0,1)$$



$N(0,1) \to N(\mu, 1)$

$N(\mu,1) \to N(\mu, \sigma_1^2)$ $\quad \sigma_1^2 < 1$

$N(\mu,1) \to N(\mu, \sigma_2^2)$ $\quad \sigma_2^2 > 1$

2. $Z \to \chi^2$

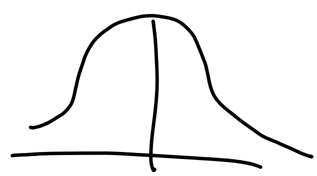
① $Z \sim N(0,1)$, Define $U := Z^2 \sim \chi_1^2$

② $U_i \overset{iid}{\sim} \chi_1^2$, Define $V := \sum_{i=1}^{n} U_i \sim \chi_n^2 \longrightarrow S^2$

3. $Z \to t$

$Z \sim N(0,1)$, $U \sim \chi_n^2$, Z and U independent

Define $t := \frac{Z}{\sqrt{U/n}} \sim t(df = n) \to t_n$



4. $Z \to F$ (ANOVA)

$U \sim \chi_m^2$, $V \sim \chi_n^2$ $\quad$ Define $W = \frac{U/m}{V/n} \sim F_{m,n}$

II. Sample Mean and Sample Variance.

Δ X $\quad x$

S^2 $\quad s^2$

$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$

$\bar{X} = \frac{1}{n} \sum X_i$ $\qquad \underline{S}^2 = \frac{1}{\underline{n-1}} \sum (X_i - \bar{X})^2$

Thm 1.

① $(X_1-\bar{X}, X_2-\bar{X}, \cdots, X_n-\bar{X})$ are independent

② $\bar{X}$ and S^2 are independent

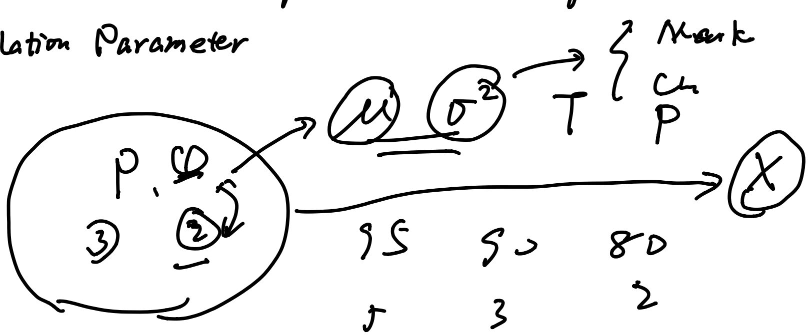
Thm 2. ★

① $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ $\Longrightarrow$ Confidence Int for σ^2

② $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ $\Longrightarrow$ Confidence Interval for μ.

Simple Random Sample

I. Population Parameter



(mu) $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

(tao) $\tau = \sum_{i=1}^{N} x_i = N\mu$

(sigma) $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$

$= \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right) - \mu^2$

N: Population Size

Numerical

△ Recall RV: $\mu = EX = \sum_{x\in\Omega_X} p(x)\cdot x$ $\quad p(x) = \frac{1}{N}$

$\sigma^2 = E(X-\mu)^2 = \sum (x-\mu)^2\cdot p(x)$

$= EX^2 - \mu^2 =$

Dichotomous Case: $X_i = 0, 1$ → Population Prop.

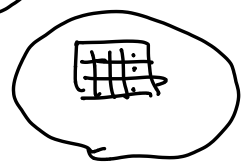
$\mu = p = \frac{n}{N}, \quad \mu = \frac{1}{N}\sum x_i \qquad n = \sum x_i \qquad x_i^2 = x_i$

$\sigma^2 = \frac{1}{N}\sum x_i^2 - \mu^2 = p - p^2 \qquad \sum x_i^2 = \sum x_i$

$= p(1-p)$

II. Point Estimate of Population Parameter

1. Simple Random Sample (SRS)



① Def:

Sample that makes each unit has equal chance to be chosen.

$\begin{cases} \text{with Replacement} \\ \text{w/o Replacement} \end{cases}$

$\boxed{1 \quad 2 \quad 3}$ pool $\qquad \frac{n}{N}$

w $\quad p(1) = \frac{2}{3}$

w/o $\quad p(1) = \frac{1}{3} + \frac{2}{3}\cdot\frac{1}{2} = \frac{2}{3}$

△. Question? $\quad N \to n$

1 2 3 4

a). How many possible samples? $\quad$ Error: $\underline{\underline{N}}$

12 13 14 23 24 34

b). What is the probability of a certain sample?

a).

"w" $\quad (N, N, \cdots, N) \to N^n$

"w/o" $\quad \binom{N}{n} = \frac{N!}{n!(N-n)!}$

b) Inverse of # of possible Sample

② Sampling Distribution $\qquad \underline{100} \to \bar{x} \qquad \bar{x} \qquad \bar{x} = 1.5$

$P(\bar{x} = 1.0)$

a). Common Sense based on Analogy

Sample Mean: $\hat{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad$ Vs. $\mu = \frac{1}{N}\sum x_i$

Sample Total: $\hat{T} = N\bar{X}$

Sample Variance: $\begin{cases} \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \quad \text{类比} \\ S^2 = \frac{1}{n-1}\sum(X_i - \mu)^2 \end{cases}$

$n-1$

△ Dichotomous Case.

Sample Proportion: $\hat{p} = \bar{X} = \frac{\sum X_i}{n} = \frac{m}{n}$

P 11:01 X_1
11:02 X_2

b). Definition of Statistic

△ Mathematical Definition of a Sample.

A sequence of Random Variables (finite sequence) $n \in \mathbb{N}$

$\{X_1, X_2, \dots, X_n\} \to \{X_k\}_{k \in \mathbb{N},\ k \le n}$

Statistic: ~~A summary~~ function of a sample

$T = T(X_1, X_2, \cdots, X_n)$ T μ

$y = T(x_1, x_2, \cdots, x_n)$ $\bar{x}$

μ σ^2 τ $N\mu$ Estimate / Estimation

2. Estimation of $\mu \to \bar{X}$

① Thm: $\bar{X}$ is an unbiased esitimate of μ.

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

$E(T) = \tau$ $\bar{x}_1, \bar{x}_2, \bar{x}_3 \cdots \bar{x}_{\binom{N}{n}}$

② Thm.:

$Var(\bar{X}) = \begin{cases} \sigma^2/n & \to 0 \quad \text{SRSWR} \to N \\ \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) & \text{SRSWOR} \quad N \to \infty \end{cases}$

$\hookrightarrow$ finite population correction

$\bar{x}$

$Var(X)$

3. Estimation of $\sigma^2 \to \begin{cases} \hat{\sigma}^2 \to \\ s^2 \to \text{unbiased} \end{cases}$

① Thm:

$$E(\hat{\sigma}^2) = \begin{cases} \sigma^2 \cdot \frac{n-1}{n} & \text{SRWR} \\ \sigma^2 \frac{N}{N-1} \frac{n-1}{n} & \text{SRWOR} \end{cases}$$

$\sigma^2 = \frac{1}{n}\Sigma \Rightarrow \Sigma = n\sigma^2$

$s^2 = \frac{1}{n-1}\Sigma = \frac{1}{n-1} n \cdot \sigma^2$

$= \frac{n}{n-1}\sigma^2$

S^2

$$E(S^2) = \begin{cases} \sigma^2 \cdot & \text{SRWR} \\ \sigma^2 \cdot \frac{N}{N-1} & \text{SRWOR} \end{cases}$$

② Unbiased estimate of $Var(\bar{X})$

$$\left.\begin{matrix} \text{SRSWOR} & \frac{N-n}{Nn} s^2 \\ \text{SRSWR} & \frac{s^2}{n} \end{matrix}\right\} = S_{\bar{X}}^2 \longleftrightarrow S_X^2$$

Week 2: Probability Distributions

Summary of Key Results

Summary of Key Results

Population Parameter (μ)	Estimate, ($\bar{X}$)	Variance of Estimate, ($\sigma^2_{\bar{X}}$)	Estimated Variance ($S^2_{\bar{X}}$)
population want to know μ	$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$	$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$S^2_{\bar{X}} = \frac{s^2}{n}\left(1 - \frac{n}{N}\right)$
p	$\hat{p}$ – sample proportion	$\sigma^2_{\hat{p}} = \frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)$	$S^2_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n-1}\left(1 - \frac{n}{N}\right)$
τ	$T = N\bar{X}$	$\sigma^2_T = N^2\sigma^2_{\bar{X}}$	$S^2_T = N^2 S^2_{\bar{X}}$
σ^2	$\left(1 - \frac{1}{N}\right)s^2$		

where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$

$$\begin{cases} \text{What:} & \bar{X} \Rightarrow \mu \\ \text{How:} & S^2_{\bar{X}} \Rightarrow Var(\bar{X}) = \sigma^2_{\bar{X}} \end{cases}$$

III. Interval Estimation

1. Two facts:

If $X_1 \sim N(\mu, \sigma^2) \to$ Population Dist, then

$$\begin{cases} ① \ \dfrac{\bar{X}-\mu}{\sqrt{Var(\bar{X})}} \sim N(0,1) \\[2ex] ② \ \dfrac{\bar{X}-\mu}{S_{\bar{X}}} \sim t(n-1) \end{cases}$$

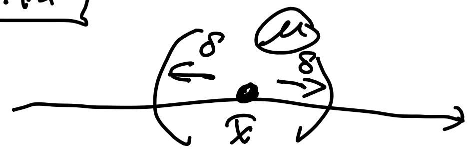
$$\boxed{\frac{\bar{X}-\mu}{S/\sqrt{n}}}$$

WR

$$S_{\bar{X}}^2 = \frac{S^2}{n}$$

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

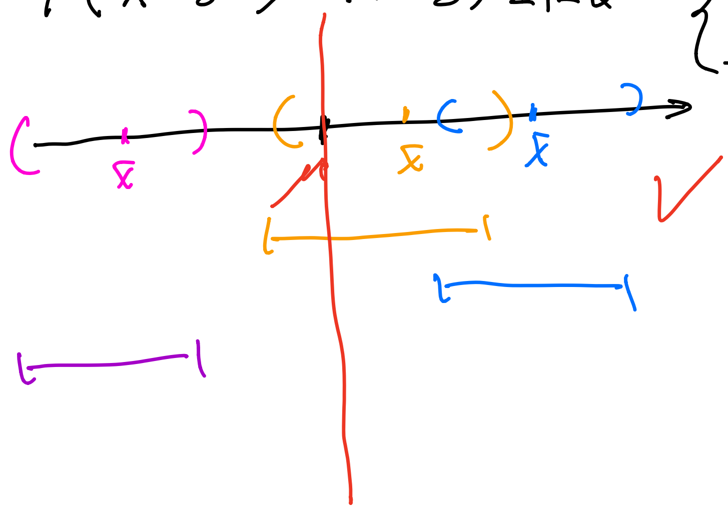
2. Confidence Interval $\bar{X}$

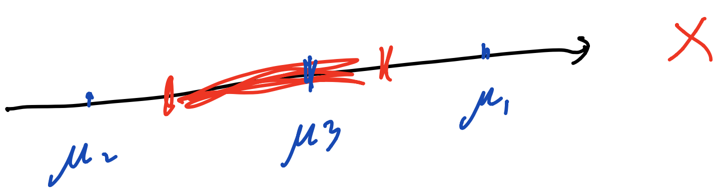


$$P(\bar{X} - \delta \le \mu \le \bar{X} + \delta) = 1-\alpha$$

$\begin{cases} \text{the chance that } \mu \text{ in } \bar{x} \pm \delta. \ ✗ \\ \text{the } \bar{x} \pm \delta \text{ contain } \mu \ ✓ \end{cases}$

prob that





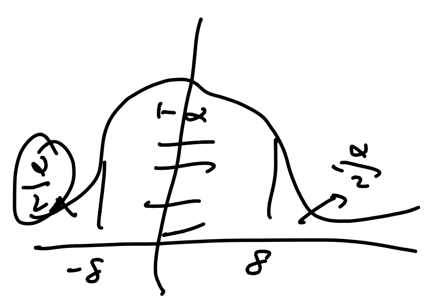
Confidence Level: $1-\alpha$

$$P(\bar{x} - \delta \le \mu \le \bar{x} + \delta) = 1-\alpha$$

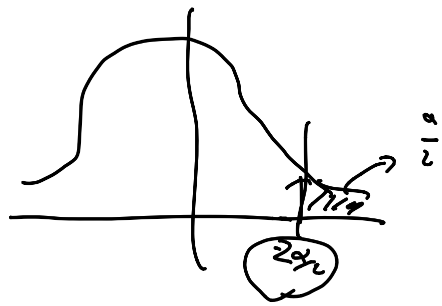
$$\Leftrightarrow P(-\delta \le \bar{x} - \mu \le \delta) = 1-\alpha$$

$$\Leftrightarrow P\left(-\delta \le \frac{\bar{X}-\mu}{\sqrt{Var(\bar{X})}} \le \delta\right) = 1-\alpha$$

$$\Leftrightarrow P\left(\frac{\bar{X}-\mu}{\sqrt{Var(\bar{X})}} > \delta\right) = \frac{\alpha}{2}$$



$\delta = z_{\alpha/2}$ ✓



$\bar{x} \pm z_{\alpha/2} \cdot \sigma_{\bar{x}}$

$\bar{x} \pm t_{n-1, \frac{\alpha}{2}} \cdot S_{\bar{x}}$

Sample $\Rightarrow$ Population Parameter

$\downarrow$

Decision Making

n, N Large $\boxed{\text{Normal}}$

IV. Estimation of Ratio.

Variable of interest: Y

Known Population : X $\Leftarrow$ Auxiliary Information

$T_y = \sum^{N} Y_i \qquad T_x = \sum_{i=1}^{N} X_i$

$T_r = \frac{T_y}{T_x} = \frac{\mu_y}{\mu_x} \Rightarrow T_y = \frac{\mu_y}{\mu_x} \cdot T_x \qquad \mu_y \Leftarrow \bar{y}$

$\mu_x \Leftarrow \bar{x}$

$\hat{T}_r = \frac{\bar{y}}{\bar{x}} \cdot T_x \qquad T_y$

A: $Var(\hat{T}_r) < Var(N\bar{Y})$

B: N is unknown. $N\bar{Y}$ cannot used