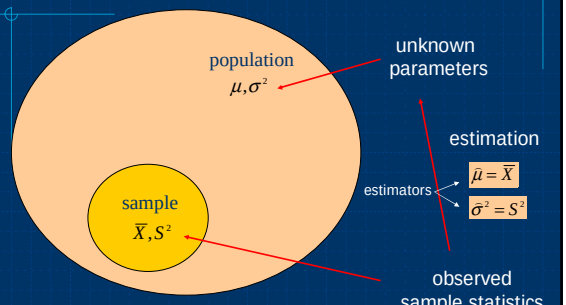
Point Estimation



Desirable Properties of Estimator

parameter : θ estimator : $\hat{\theta}$

$E(\hat{\theta}) = \theta$ for any θ ⟶ $\hat{\theta}$ is unbiased

Example : VSRS

$E(\bar{X}) = \mu$ —— $\bar{X}$ is unbiased for μ

$E(S^2) = \sigma^2$ —— S^2 is unbiased for σ^2

$E(SD^2) = E\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right\} = \frac{n-1}{n}\sigma^2 < \sigma^2$ — SD^2 is biased for σ^2

underestimate

Bias

$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$bias(\hat{\theta}) > 0$ —— overestimate

$bias(\hat{\theta}) < 0$ —— underestimate

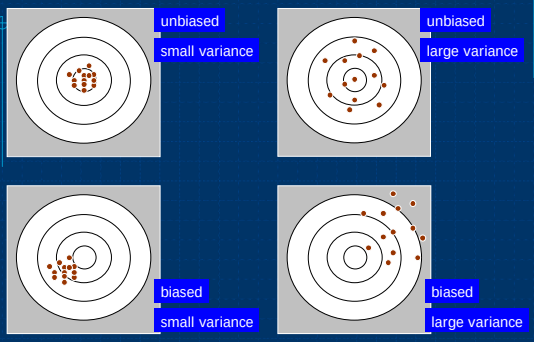
$bias(\hat{\theta}) = 0$ —— unbiased

Example : VSRS

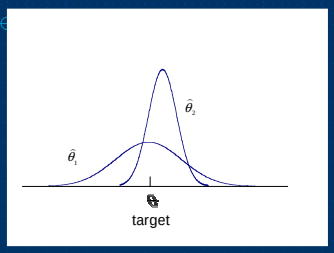
$$SD^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$bias(SD^2) = E(SD^2) - \sigma^2 = -\frac{\sigma^2}{n}$$ underestimate

Mean Squared Error



Mean Squared Error



$\hat{\theta}_1$ is unbiased

$\hat{\theta}_2$ is biased

But $\hat{\theta}_2$ is better than $\hat{\theta}_1$

$MSE(\hat{\theta}_2) < MSE(\hat{\theta}_1)$

$$MSE(\hat{\theta}) = E\left\{(\hat{\theta} - \theta)^2\right\} = Var(\hat{\theta}) + bias(\hat{\theta})^2$$

Efficiency

efficiency of $\hat{\theta}_1$ compared to $\hat{\theta}_2$ —— $eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}$

$\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if $eff(\hat{\theta}_1, \hat{\theta}_2) > 1$

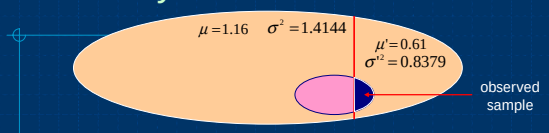
Example : VSRS $\{X_1, X_2\}$ $\tilde{\mu}_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2$ $\tilde{\mu}_2 = \frac{2}{3}X_1 + \frac{1}{3}X_2$

$bias(\tilde{\mu}_1) = bias(\tilde{\mu}_2) = 0$

$MSE(\tilde{\mu}_1) = Var(\tilde{\mu}_1) = \frac{\sigma^2}{2}$ $MSE(\tilde{\mu}_2) = Var(\tilde{\mu}_2) = \frac{5\sigma^2}{9}$

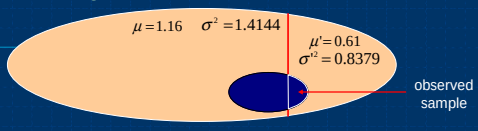
$eff(\tilde{\mu}_1, \tilde{\mu}_2) = \frac{MSE(\tilde{\mu}_2)}{MSE(\tilde{\mu}_1)} = \frac{10}{9} = 111\%$ $\tilde{\mu}_1$ is 11% more efficient than $\tilde{\mu}_2$

Efficiency



	Total population (Target)		Subpopulation who would response	
X = number of children over 5 years old	Frequency	Rel. frequency	Frequency	Rel. frequency
0	20,000	0.40	6,200	0.62
1	12,000	0.24	2,100	0.21
2	10,000	0.20	1,200	0.12
3	6,000	0.12	400	0.04
4	2,000	0.04	100	0.01
Total	50,000	1.00	10,000	1.00

Efficiency



High volume sample : ask 1000 individuals without follow-up

Sample size ≈ 200 $E(\bar{X}') = \mu' = 0.61$ $Var(\bar{X}') = \frac{\sigma'^2}{n} = 0.0041895 = \frac{0.8379}{200}$

$bias(\bar{X}') = 0.61 - 1.16 = -0.55$ $MSE(\bar{X}') = (-0.55)^2 + 0.0041895 = 0.3067$

High quality sample : ask 25 individuals with follow-up

Sample size = 25 $E(\bar{X}) = \mu = 1.16$ $Var(\bar{X}) = \frac{\sigma^2}{n} = 0.05658 = \frac{1.4144}{25}$

$bias(\bar{X}) = 0$ $MSE(\bar{X}) = Var(\bar{X}) = 0.05658$

$$eff(\bar{X}, \bar{X}') = \frac{0.3067}{0.05658} = 5.42$$