Inferential Statistics: Sample Statistics to estimate population parameters. Sample $\subset$ Population $\{X_i\}_{i=1}^{n}$. Estimation of parameters / Hypothesis Testing. Good estimators: (i) Unbiased (ii) Precise. Let $\{X_i\}^n$ random sample with mean: μ, s.d: σ $\therefore$ $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{X}$ unbiased: $\mathbb{E}[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{n\mu}{n} = \mu$. Standard Error $(\bar{X})$: $\sigma_{\bar{X}} = \sqrt{Var(\frac{1}{n}\sum^{n} X_i)} = \frac{\sigma}{\sqrt{n}}$

If population $\sim N(\mu, \sigma^2) \Longrightarrow \bar{X} \sim N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$, Z-Value for $\bar{X}$: $Z = \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ $\therefore Z \sim N(0,1)$, $E[\bar{X}] = E[X] = \mu$, $Var[\bar{X}] < Var[X]$

Central Limit Theorem: $\{X_i\}^n$, mean: μ, var $= \sigma^2$, $\bar{X}_n$, $n \longrightarrow \infty \Longrightarrow Z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \sim N$. Sampling Distributions (proportions): $\{X_i\}^n, X \in \{0,1\}$ $\mathbb{P}(X=1) = p$, $\mathbb{P}(X=0) = 1-p$ $\therefore$ $\mathbb{E}[X_i] = p$, $Var[X_i] = p(1-p)$, $E(\bar{X}) = \frac{1}{n}\sum^{n} E(X_i) = p$ $\therefore$ proportion unbiased e. of p., $\sigma_{\bar{X}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$

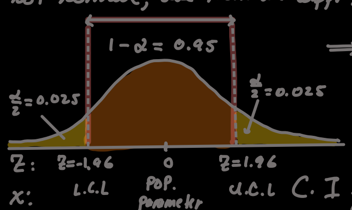
CLT (Proportions): $\{X_i\}^n \overset{iid}{\sim} Bernoulli(p)$, $n \to \infty \Longrightarrow Z = \frac{\bar{X}-p}{\sqrt{p(1-p)/n}} \sim N(0,1)$. A point estimator $\hat{\theta}$ unbiased of parameter θ if

Pop. Parameter	Sample statistic
Mean μ	$\bar{x}$
Propor. P	$\hat{p}$

$\mathbb{E}(\hat{\theta}) = \theta$. Ex: $\bar{X}$ U.E. μ, $\hat{p}$ U.E P. $\therefore$ Bias$(\hat{\theta}) = E(\hat{\theta}) - \theta$. Most efficient estimator of θ is unbiased with smallest variance. Point Estimate: Single number, Confidence Interval: Additional Info variability

L.C.L   point E.   U.C.L.

α% Confidence Interval: True population parameter is contained in α% intervals calculated this way. Confidence Interval Estimation for the Mean (σ^2 known): $\{X_i\} \sim N$. if population not normal, use normal approximation. C.I.: $\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, $Z_{\alpha/2}$ s.t $\mathbb{P}(Z \geq z_{\alpha/2}) = \alpha/2$. Finding $Z_{\alpha/2}$: Consider 95% C.I.



$\Longrightarrow Z_{0.25} = 1.96$, $LCL = \bar{x} - Z\frac{\sigma}{\sqrt{n}}$, $UCL: \bar{X} + Z\frac{\sigma}{\sqrt{n}}$. Ex: $n = 11$, $\bar{X}_n = 20$, $\sigma = 5$ $\therefore$ 90% C.I for μ

C.I: $20 \pm Z_{0.05}\frac{\sigma}{\sqrt{n}} = 20 \pm 1.645\left(\frac{5}{\sqrt{11}}\right) = 20 \pm 2.48 \Longrightarrow 17.52 < \mu < 22.48$ $\therefore$ We are 90% confident that $\mu \in$

C.I for mean (σ^2 unknown): $\{X_i\}^n$, $\bar{x}$, S $\therefore$ $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ (T-distribution): $\bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$

C.I for Pop. Proportion: $\sqrt{Var(\hat{p})} = \sqrt{\frac{P(1-P)}{n}}$, since P unknown, $\hat{p} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\therefore$

C.I for P: $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ Ex: $n = 1526$, $\hat{p} = 0.64$, 95% C.I $\therefore$ $0.64 \pm 1.96\sqrt{\frac{0.64(0.36)}{1526}}$, $0.616 < p < 0.664$

Dependent Samples: Population Parameter of interest is $E(y) - E(x)$ $\therefore$ $d_i := y_i - x_i$, point estimate for population mean difference is $\bar{d} := (\sum^{n} d_i)/n$, Confidence Interval for $\bar{d}$: $\bar{d} \pm Z_{\alpha/2}\frac{\sigma_d}{\sqrt{n}}$. Difference Between two means

Independent Samples: $\bar{x} - \bar{y}$ (σ_x^2, σ_y^2 known) $\therefore$ $\sqrt{Var(\bar{x}-\bar{y})} = \sqrt{Var(\bar{x}) + Var(\bar{y})} = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$, $\{X_i\}^n, \{Y_i\}^n \sim N$

C.I. $(\bar{x} - \bar{y}) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ Ex: $n_x = 4592$, $n_y = 1423$, $\bar{x} = 106.8$, $\bar{y} = 37.8$, $\sigma_x = 439.2$, $\sigma_y = 70.2$

$\therefore$ 95% C.I: $(106.8 - 37.8) \pm 1.96\sqrt{\frac{439.2^2}{4562} + \frac{70.2^2}{1423}} = 69 \pm 13.2$. Two population proportions: point estimate $P_x - P_y$ $\hat{p}_x - \hat{p}_y$ $\therefore$ $SE = \sqrt{Var(\hat{p}_x - \hat{p}_y)} = \sqrt{Var(\hat{p}_x) + Var(\hat{p}_y)} = \sqrt{\frac{P_x(1-P_x)}{n_x} + \frac{P_y(1-P_y)}{n_y}}$ $\therefore$ C.I: $(\hat{P}_x - \hat{P}_y) \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}_x(1-\hat{P}_x)}{n_x} + \frac{\hat{P}_y(1-\hat{P}_y)}{n_y}}$

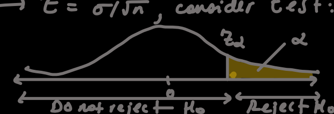
Ex: $\hat{P}_x = \frac{420}{751} = 0.56$, $\hat{P}_y = \frac{560}{775} = 0.72$, $\therefore$ 90% C.I: $(0.56 - 0.72) \pm 1.645\sqrt{\frac{0.56(0.44)}{751} + \frac{0.72(0.28)}{775}} = -0.199 < P_x - P_y < -0.121$

Hypothesis: Claim about a population parameter. Null Hypothesis (H_0): states (numerical) assumption to be tested about pop. parameter. Alternative Hypothesis (H_1): opposite of null hypothesis: Ex (H_0: $P \leq 0.5$) $\Longrightarrow H_1: P > 0.5$ is $\hat{p} = 0.8$ likely if $P \leq 0.5$? If not, reject H_0. Type (I) Error: Reject a true (H_0). Probability of Type (I) error is α. α := level of significance. Type (II) error: Fail to reject false H_0. Probability Type (II) Error is β determined by α, n, etc.

Decision	H_0 TRUE	H_0 False
Fail to Reject (H_0)	Correct Decision $(1-\alpha)$	TYPE II (β)
Reject (H_0)	TYPE I α	Correct d. $(1-\beta)$

Type I can only occur if H_0 is True, Type II can only occur if H_0 false $\therefore$ $(\alpha)\downarrow \longrightarrow (\beta)\uparrow$. Test: Convert $\bar{X} \longrightarrow Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$, consider test:

$\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ Desicion Rule if $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > Z_\alpha$



Ex: ($\sigma = 10$ known) $H_0: \mu \leq 52$, $H_1: \mu > 52$. Suppose $\alpha = 0.10$. Find Rejection region $Z = 1.28$ $\therefore$ Reject H_0 if $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > 1.28$. Suppose $\{X_i\}^{64}$, $\bar{x} = 53.1$ $\therefore Z = \frac{53.1-52}{10/\sqrt{64}} = 0.88 < 1.28$ $\therefore$ Do not reject H_0

P-Value approach to testing: Convert $\bar{X} \longrightarrow Z$, Calculate p-value: Probability of obtaining a test statistic more extreme than the observed sample value given that h_0 is true i.e $p = \mathbb{P}(\bar{X} > 53.1 \mid \mu = 52)$, Decision rule: if p-value $< \alpha$, reject H_0, if p-value $\geq \alpha$, do not reject. $\therefore$ $\mathbb{P}(\bar{X} \geq 53.1 \mid \mu = 52.0) = \mathbb{P}(Z \geq \frac{53.1-52}{10/\sqrt{64}}) = \mathbb{P}(Z \geq 0.88) = 1 - 0.81$

$\therefore p = 0.1894 > \alpha = 0.10$ $\therefore$ fail to reject H_0