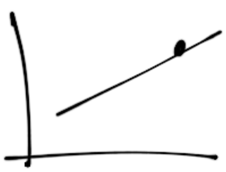
Review - 1

Data

↓

Stats → Probability

(μ, –, –, m_0)



x

Arithmetic

~~Variance~~

$\left.\boxed{\begin{matrix}\sigma \\ (\sigma^2)\end{matrix}}\right\}$ dispersion

✓

BINOMIAL DISTRIBUTIONS

Expt / Trials $\boxed{(n)}$

Success → (p)

Failure → $(1-p) = q$

$\boxed{\text{Binom}(n, p)}$

$\mu = np \qquad \sigma = \sqrt{n(p)(q)}$ → $(1-p)$

$$\binom{n}{C_p} = \frac{n!}{p!\,(n-p)!}$$

$\boxed{\begin{matrix} 5 \\ 20 \\ 100 \end{matrix}}$

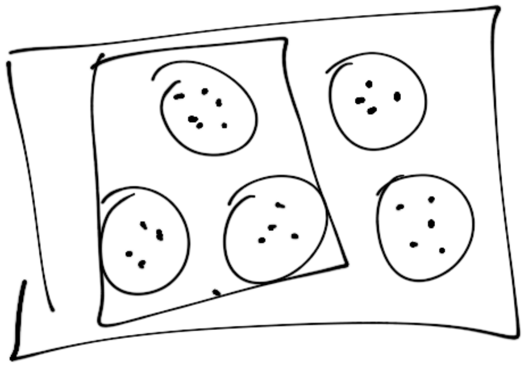
$n \to \infty$

$p \neq 1 \Rightarrow p$ - small

POISSON

(Avg rate of succes)

$\boxed{\lambda}$

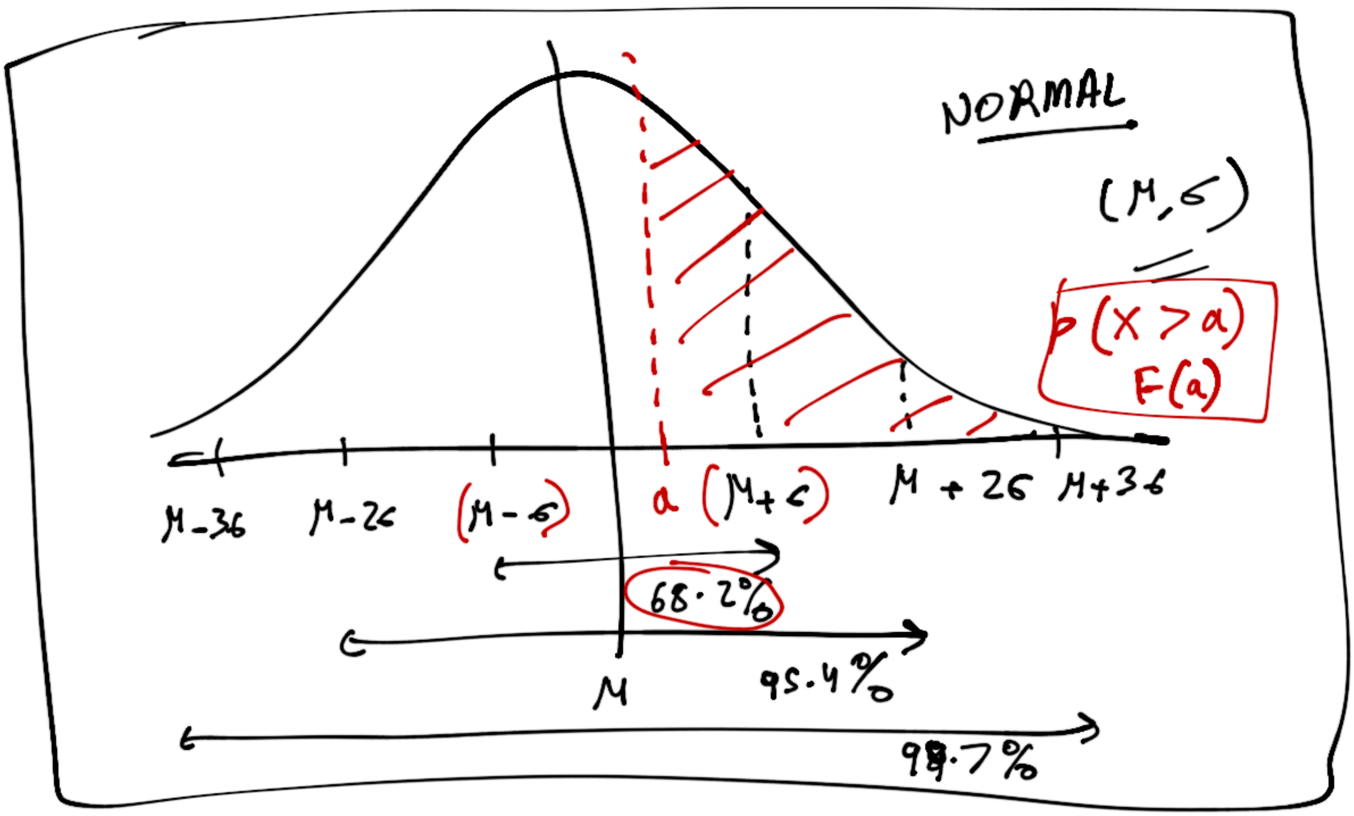


λ = threshold

$(3/5)$

0.6

$\boxed{\lambda \to \infty}$



$\boxed{\mu = 0}$ STANDARD

$\boxed{\sigma = 1}$ NORMAL

0.6

Random Variables

$\{T, T, T, T, H, H\}$

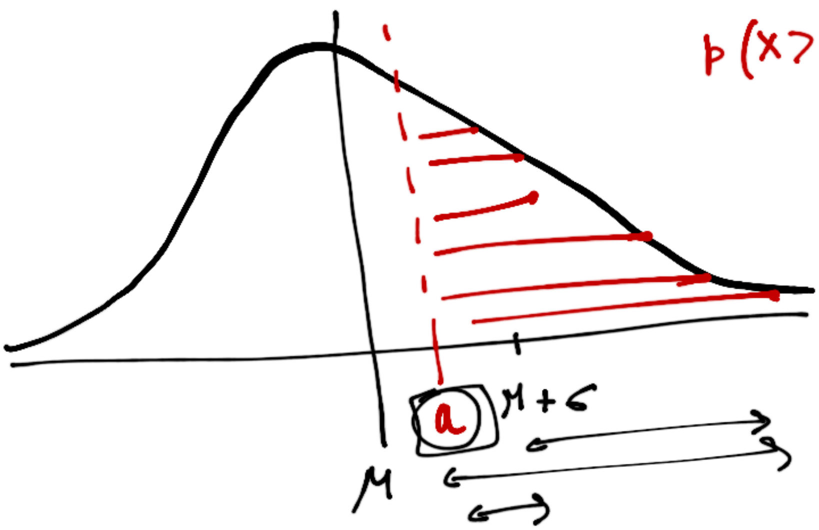
$\boxed{X} = \{0, 0, 0, 0, 1, 1\}$ YES, the coin is biased?

NO

Sample

$T = 0$

$H = 1$



$p(x > a)$

Z values

$Z = \left(\frac{x - \mu}{\sigma}\right)$

$\boxed{F(a)} = p(x > a)$

Random Variables $\underline{[X]} = \{1, 0, 1, 0, 1, 1\}$

$E[x] = \mu_x$, $Var[x] = \sigma_x^2$

Expected value of a single RV = Mean

$W = bX + a \rightarrow$ error

$\boxed{E[W] = b\,E[X] + a}$, $\mu_w = b\,\mu_x + a$

$$\sigma_w^2 = \boxed{b^2}\sigma_x^2 + \cancel{a^2} \to 0$$

$b \equiv$ constant

$b^2 = ?$

{dependance parameter}

$$\boxed{\sigma_w^2 = b^2 \sigma_x^2}$$

CO$\boxed{\text{VARIANCE}}$

$$\mathrm{Cov}(X,Y) = \boxed{\sigma_{XY}}$$

$\boxed{X, Y}$ → independent

$$\mathrm{Cov}(X,Y) = \underline{0}$$

$$\mathrm{Cov}(X,Y) = \underline{E[XY]} - \underbrace{E[X]E[Y]}_{\mu_x \ \mu_y}$$

E

$$X = a x^3$$

$$Y = \beta y^4 + \gamma y^2$$

$$E[X] \neq a\,E[x^3]$$

$$E[Y] = \boxed{\beta E[y^4] + \gamma E[y^2]}$$

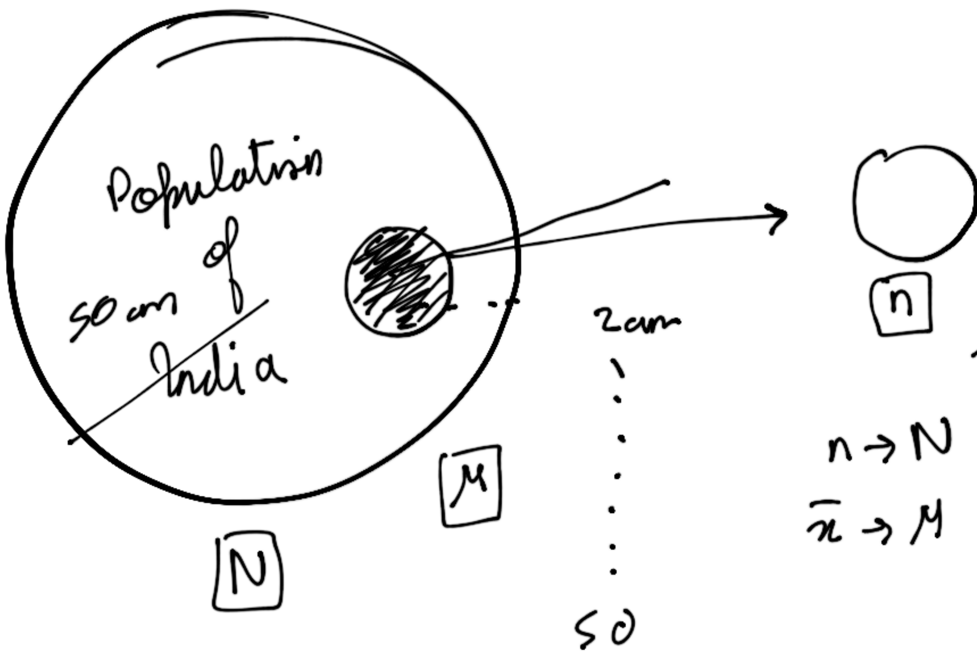
$$E\left[a\beta x^3 y^4 + a\gamma x^3 y^2\right]$$

$$\boxed{E[XY] \neq E[X]\,E[Y]}$$

( )

$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

SAMPLING

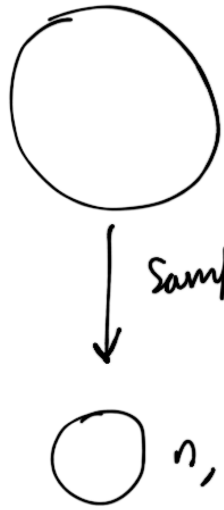


Population of City X

$\boxed{n}$, $\bar{x}$

$n \to N$
$\bar{x} \to \mu$ $\Rightarrow$ Achieved Population Generalization.

$N \to \infty$

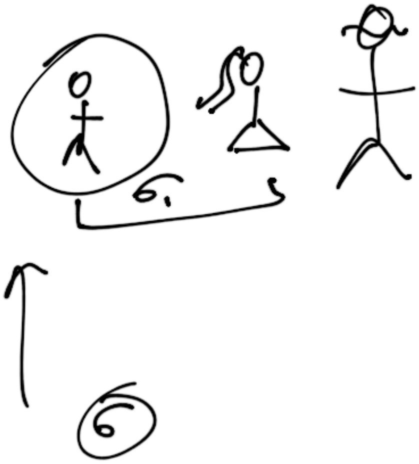


N, μ

Sample

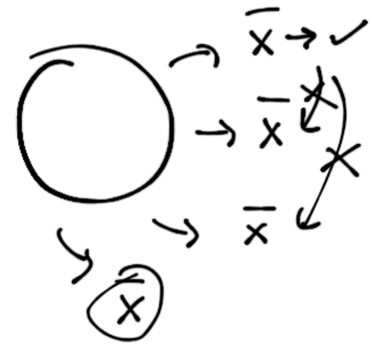
$n, \bar{x}$

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



σ_1

σ



independent

(n)

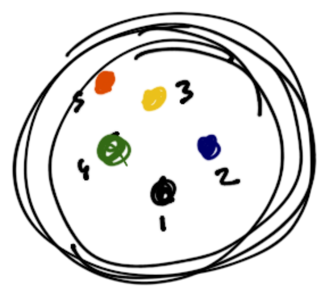
$\bar{x} = \boxed{\mu}$

UNBIASED ESTIMATOR

$\sigma_{\bar{x}} \neq \sigma$

$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

STANDARD ERROR

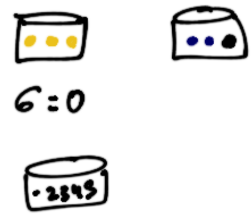


$\sigma_{12} = 0.2$

$\sigma_{34} = 0.4$

Camille

Florence



$\sigma = 0$

σ_{12345} (n)

$n \geq 15$ mil

σ_{12345} (N)

Distributted Normally ✓

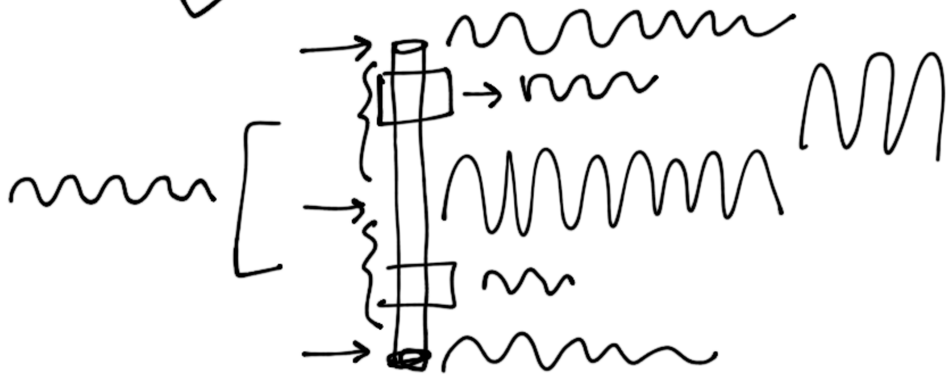
or the werent ✓

CLT

(Central Limit Theorem)

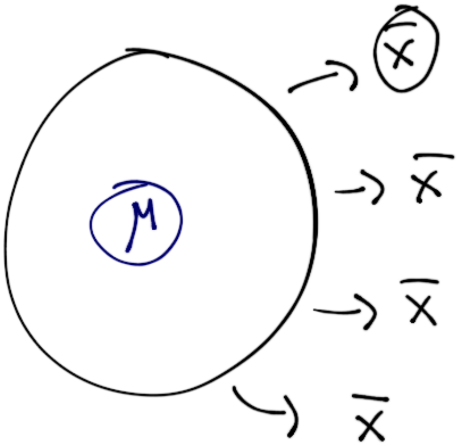
Large $\boxed{n \geq 30}$

Sample will be distributed normally.

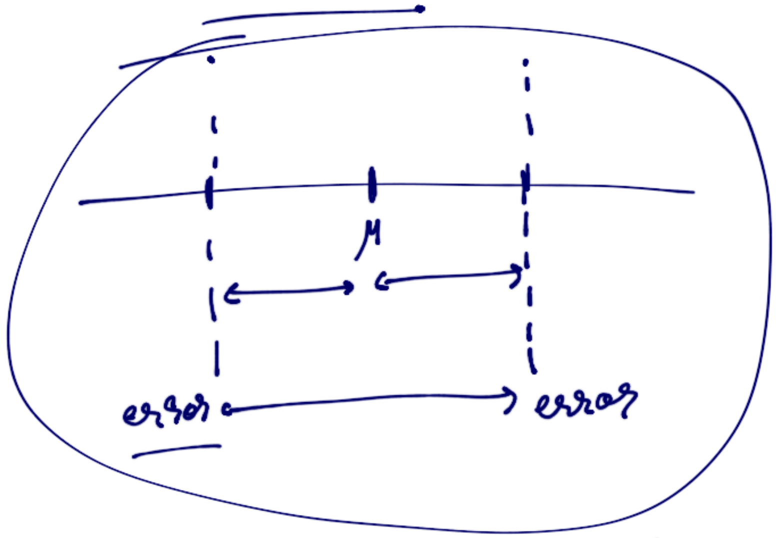


CONFIDENCE INTERVALS

How confident are you that your samples are representative.



95% sure



$\boxed{100(1-\alpha)}\%$ of the time, $\bar{x}$ falls between $\left(Z_{\alpha/2}\,(SE)\right)\mu$

$100(1-\alpha)\%$ of the time $\bar{x} \in \left(\mu \pm Z_{\alpha/2}\,\cancel{SE}\right)$ $\frac{\sigma}{\sqrt{n}}$

$100(1-\alpha)\%$ of the time $\mu \in \left(\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

75% confidence interval

$\frac{.75}{10\cancel{0}}\cancel{\%} = 1\cancel{00}(1-\alpha)\cancel{\%}$

$\alpha = 0.25$

$Z_{\alpha/2} = \underbrace{Z_{0.125}\frac{\sigma}{\sqrt{n}}}_{\text{Standard error}}$

$$\boxed{\mu \in \bar{x} \pm Z_{0.125}\frac{\sigma}{\sqrt{n}}}$$

95%

$\alpha/2 = 0.125$

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

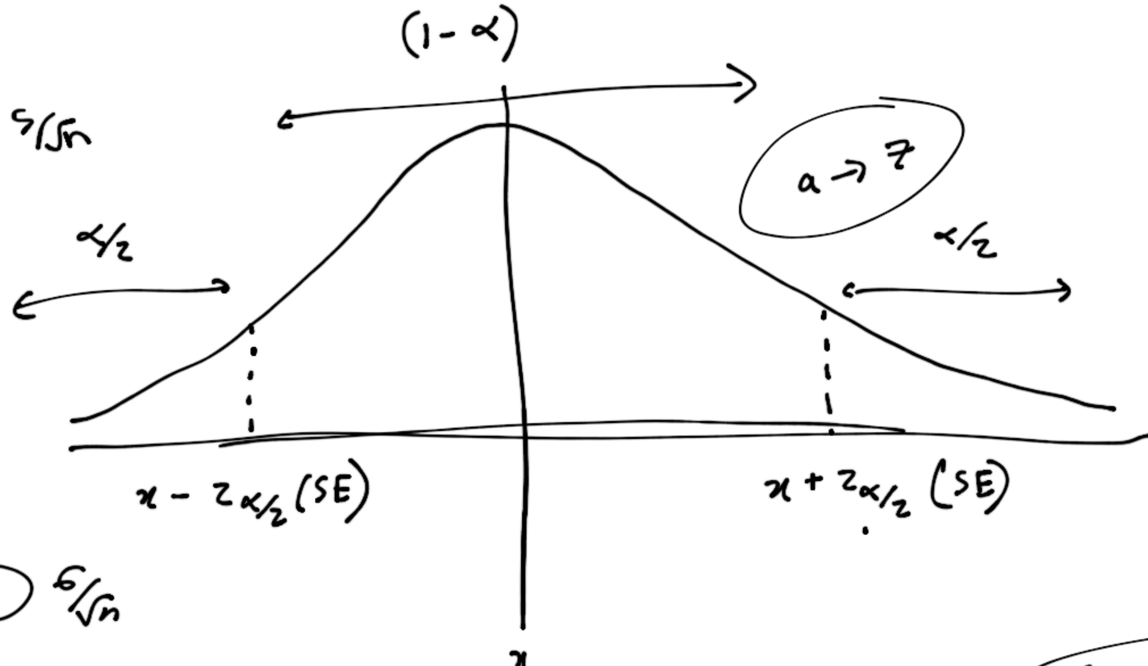
$$0.05 = z_{0.125} \frac{\sigma}{\sqrt{n}}$$

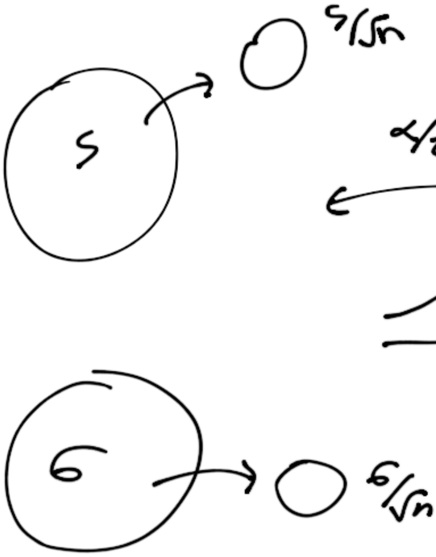
How many trials you need to do, to get a certain threshold (or less) of error given your variance & CI.

$$E^2 = z_{\alpha/2}^2 \frac{\sigma^2}{n}$$

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2$$

0.5





$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$z_{\alpha/2} \frac{s}{\sqrt{n}}$$

σ, s

$$\sigma_{\hat{p}} = p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

$$\mu_{\hat{p}} = p$$

$$SE_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n}$$

$$E(SE_{\hat{p}}) = \sigma_{\hat{p}}$$

$$E = z_{\alpha/2}\underbrace{\sqrt{\frac{p(1-p)}{n}}}$$

$$E^2 = z_{\alpha/2}^2\left(\frac{p(1-p)}{n}\right)$$

$$\boxed{n} = \frac{z_{\alpha/2}^2 \,\boxed{p(1-p)}}{E^2}$$

$p=0.5$

Worst case
biggest n

$$n = \frac{z_{\alpha/2}^2}{E^2}\,\underline{(0.5\times0.5)} = \frac{0.25\, z_{\alpha/2}^2}{E^2}$$

$$n = \frac{1}{4}\times\frac{z_{\alpha/2}^2}{E^2} = \frac{z_{\alpha/2}^2}{4E^2}$$

$$\boxed{n = \frac{z_{\alpha/2}^2}{4E^2}}$$

4

$\boxed{2H, 2T}$

$1H, 3T$

$\boxed{4H, 0T}$

$p=1, q=0$

$\boxed{0}$

$0.5\times0.5 = \boxed{.25}$

$$\frac{d}{dp}(p-p^2)$$

$$1-2p=0 \qquad \boxed{p=0.5}$$

$$\boxed{p=\tfrac{1}{2}}$$

↑ Useless

↓ Amazing

Q1) 95% CI

2% Margin of error

How big should my sample size be?

$100(1-\alpha)\% = 95\%$

$1-\alpha = 0.95$

$\boxed{\alpha = 0.05}$

$\boxed{E = 0.02}$

$\alpha/2 = 0.025$

$$n = \frac{z_{0.025}^2}{4\times(0.02)^2}$$

$$n = \frac{1.96\times1.96}{4\times0.0004} = \underset{\text{CLT}}{2401}$$

Samples

Small sample $(n < 30)$

$(n \geq 30)$ Large Sample.
(Normal or Not)

Small sample:

- Not Normal — Youre fucked
- Normal
  - σ known: $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
  - **σ unknown** → **t-distribution**

Large sample:

- σ known: $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
- σ unknown: $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

$$W = \frac{x - \mu}{SE} = \frac{x - \mu}{s/\sqrt{n}}$$

$t_{(n-1)}$

$n-1 \equiv$ degrees of freedom

$z_{\alpha/2} \rightarrow t_{n-1,\alpha}$

$100(1-\alpha)\%$ CI for x: $x \pm t_{n-1,\alpha}\ SE$!

$$\boxed{x \pm t_{n-1,\alpha} \frac{s}{\sqrt{n}}}$$

$\frac{s}{\sqrt{n}} \rightarrow \sigma$ unknown.

z value replaced

$\alpha/2 \rightarrow$

19

Role of Size

WITH REPLACEMENT } Nothing changes.

something changes.

WITHOUT REPLACEMENT

Population correlation factor

$$\boxed{\sqrt{\frac{N-n}{N-1}}}$$

σ known → $SE_x = \sqrt{\frac{N-n}{N-1}} \frac{\sigma}{\sqrt{n}}$

σ unknown → $\sqrt{\frac{N-n}{N-1}} \frac{s}{\sqrt{n}}$

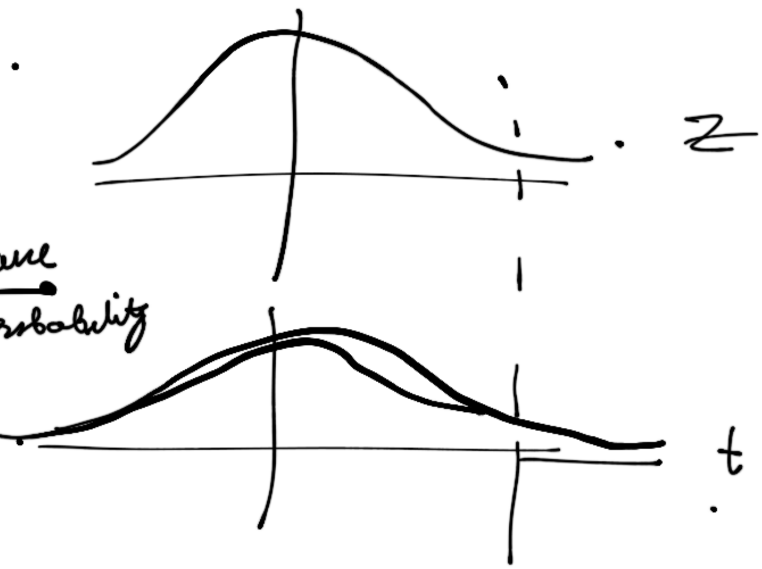
probability sample was succ/exciting/worked

p

$$SE_p = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}}$$

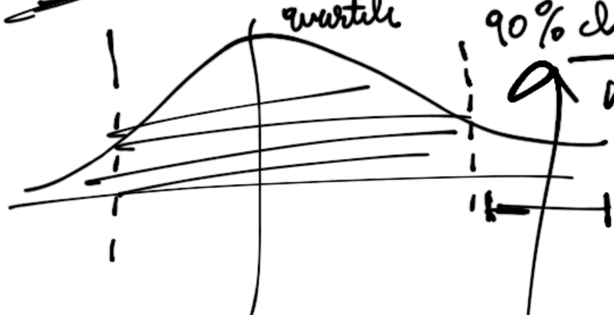
$t_{n-1,\ \alpha/2}$

95th percentile $\boxed{t_{0.05}} =$



90% chance

probability



5%
↓
10%

$t_{0.05} = 1.860$

10

90% my value is b/w -1.860 & 1.860

---

90 percentile $t_{0.1} \rightarrow$ 10% $\downarrow \times 2$ 20% $\Rightarrow$ 80% chance

8.3

300 clients

256 interested | 44 not interested

⇓

BINOMIAL

success | failure

$\sigma \propto$

$p(1-p)$ ✓

n ✓

(a) We want to say $p > 0.7$

$H_0: p = 0.7$ (1) $p \neq 0.7$

$H_A: p > 0.7 \rightarrow$ (2)

(b) $Z_{observed} = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$

$$\frac{\bar{X} - \mu_0}{\sqrt{\frac{p(1-p)}{n}}}$$

46/

$$\boxed{\frac{\frac{256}{300} - \frac{210}{300}}{\sqrt{\frac{0.7(0.3)}{300}}}}$$

~~$P(Z > \ldots)$~~

$P(Z > 5.8)$

$\boxed{3.3 \times 10^{-9}}$

less 0.01

HYPOTHESIS

$\mu = 32$ → His hypothesis : H_0 (null hypothesis)

$\mu < 32$ → My hypothesis : H_A (alternate hypothesis)

Null hypothesis rejected

↓

Alternate hypothesis accepted.

TYPE I error — α — reject H_0 but it happens to be true

TYPE II error — β — Accept H_0 but its false

$\alpha = 0$, $\beta = 0$: UTOPIA

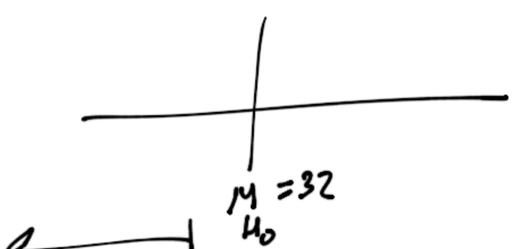
$H_0 : \mu = 32$

NO

$H_A : \mu < 32$

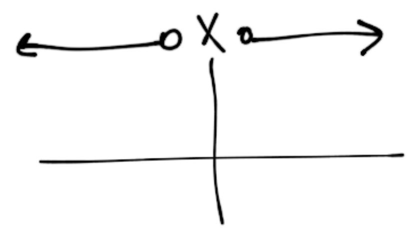
↓

one sided alternate hypothes



NO

$\boxed{\mu \neq 32}$



$H_0 : \mu = 32$

$H_A : \mu < 32$

Adding proof

$\bar{X} < 31$

Sample size is large enough

$$\boxed{Z} = \frac{\bar{X} - \mu}{6/\sqrt{n}}$$

$$z \leq \frac{31 - 32}{6/\sqrt{2}} \sim \alpha$$

$$P(Z \leq \alpha)$$

(a) $\boxed{H_0} : \mu = 120$

$\boxed{H_A} : \mu < 120$ This is what they want to prove. (2)

(b) $n = 70$ (≥ 30)

$\bar{X} = 105$ — we dont care.

$s = 65$

$$Z_{observed} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{105 - 120}{65/\sqrt{70}}$$

$$\boxed{p\text{ value} = P(Z < Z_{observed})}$$

p value basically tests the null hypothesis

$pval = P(Z < -1.93)$

a low p-value $p < 0.05$ } reject null hypothesis
95%

$pval = 0.02680$

$\alpha = 0.05$

$pval < \alpha$

$p > 0.05$ } accept null hyp
95%

$H_0 \rightarrow$ wrong

$H_A \rightarrow$ correct

$\boxed{\text{YES}}$

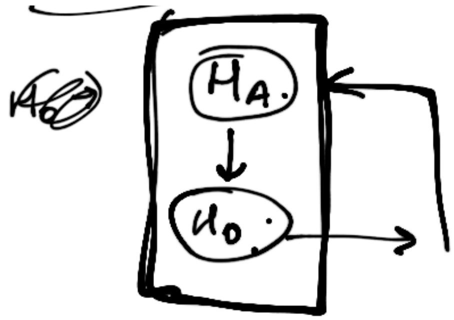
No

120

$pval < 0.05$

$\mu \neq 120$

$\mu < 120$ YES



$H_0 \quad \mu = 120 \rightarrow$ disproved

$H_A: \ \mu < 120$

8.2

$n = 6$ (small sample)
$(6 < 30)$

$\bar{x} = 150$ days

$s = 60$ days

before $\bar{x}_1 = 300$ days

σ

t-test

$t_{n-1, \alpha}$

$t_{5, 0.01} = 4.032$

$t_{observed} =$

(?) $t_{n-1, \alpha} \rightarrow$ 1 tail t table

$t_{n-1, \alpha/2} \rightarrow$ 2 tail t table

$t_{observed} \dfrac{150 - 300}{60 / \sqrt{6}} = -6.1237$

$t_{observed} < t_{value}$: Reject H_0