Recall if $X_1, \dots, X_n \overset{iid}{\sim}$ with mean μ and S.E σ and n large

(II) $\dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{d}{\approx} N(0,1)$

(III) $\bar{X} \overset{d}{\approx} N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

(IV) $T \overset{d}{\approx} N\left(n\mu, (\sqrt{n}\sigma)^2\right)$

Shipments are late 2% of the time. In 10,000 shipments, what is the prob more than 3% are late?

$X_1, \dots, X_{10000} \overset{iid}{\sim} \text{Bern}(2\%)$

$P(\bar{X} \geq 3\%) = P(\hat{p} \geq 0.03)$ By (III)

$= P\left(\dfrac{\bar{p} - 0.02}{\sqrt{\frac{0.02(1-0.02)}{10000}}} \geq \dfrac{0.03 - 0.02}{\sqrt{\frac{0.02(1-0.02)}{10000}}}\right)$

$\bar{X} \approx N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

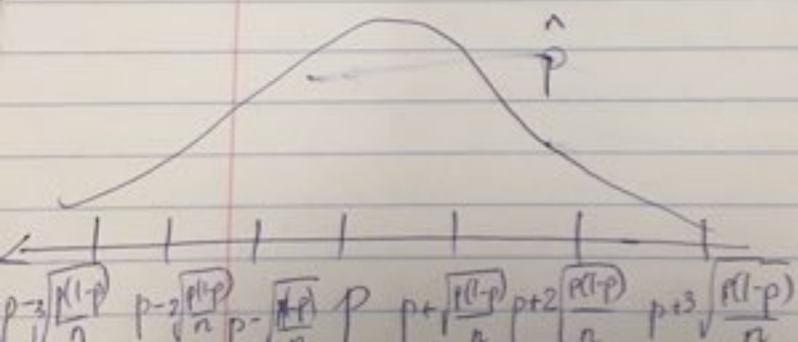
If $X_1, \dots, X_n \overset{iid}{\sim} \text{Bern}(p)$

$\mu = p$

$\sigma = \sqrt{p(1-p)}$

$\hat{p} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$

$\hat{p} := \bar{X}$ in the Bern ca



$\hat{p} := \bar{X} = \dfrac{\sum \#1\text{'s}}{n}$

"Sample proportion"

$\approx P(Z \geq 7.14) \approx 0$

11 like mushroom , $n = 23$ total

$\hat{p} = \frac{11}{23} = 0.48$ ("data")

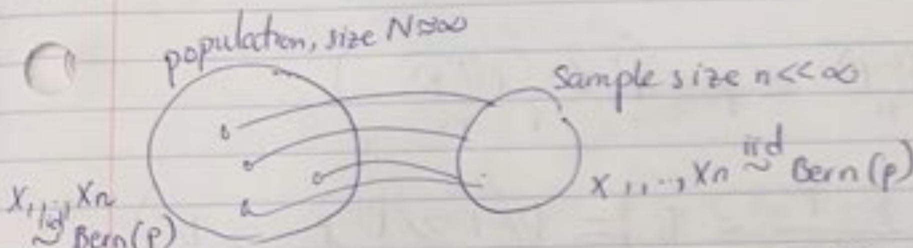
"p" is the "true" proportion paration.

$\emptyset$ know p, ~~because~~ ask each single person

<u>Goal</u>: know something about p

STATISTIC INFERENCE

- Infer the population parameter using a statistic of the data $\hat{p}$



What constitutes a "goal" simple?

A represent sample

Ex completely random sampling

Goals of Inference

1. Give me the best guess of p
2. Give me the reasonable interval of values for p
3. Test theories about p

① Point estimation $p \approx \hat{p}$

② Confidence Intervals

$$\left[\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right] := \left[\hat{p} - \sqrt{\frac{p(1-p)}{n}}, \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right] \sim \infty$$

$$p \in \left[\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right]$$

$$P\Big(p \in \Big[\underset{(big)}{\hat{P}} \pm \sqrt{\frac{p(1-p)}{n}}\Big]\Big) \to p\left(\hat{p} - \sqrt{\frac{p(1-p)}{n}} \le p \le \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right)$$

$$= p\left(-\sqrt{\frac{p(1-p)}{n}} \le p - \hat{P} \le \sqrt{\frac{p(1-p)}{n}}\right)$$

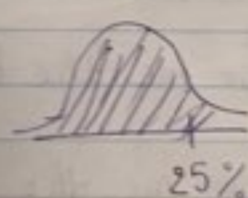
$$= p\left(-1 \le \frac{p - \hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \le 1\right) = \cancel{p\left(-1 \ge \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \ge 1\right)}$$

CLT

$$= p(-1 \le -Z \le 1) = P(Z \in [-1, 1]) = 0.68$$

$$z_{\frac{\alpha}{2}} := F_Z^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$$* \; 1 - \frac{\alpha}{2} = \int_{-\infty}^{z_{\frac{\alpha}{2}}} f(x)\,dx$$



If $\alpha = 10\% \Rightarrow \frac{\alpha}{2} = 5\% \Rightarrow 1 - \frac{\alpha}{2} = 15\%$

$$\ldots = F\left(z_{\frac{\alpha}{2}}\right) - F\left(-z_{\frac{\alpha}{2}}\right) = \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} = \boxed{1 - \cancel{\alpha}\ \alpha}$$

$$F\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$$

$$\int_{z_{\frac{\alpha}{2}}}^{\infty} f_z(x)\,dx = 1 - (1-\alpha) = \frac{\alpha}{2} = \int_{-\infty}^{-z_{\frac{\alpha}{2}}} f_z(x)\,dx$$

$$F\left(-z_{\frac{\alpha}{2}}\right)$$

$$\left[0.48 \pm 2\sqrt{\frac{p(1-p)}{23}}\right]$$

for 95% converge

Interval:

$$\left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \approx \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

If the p is not $p \approx 0$ or $p \approx 1$

$$\left[0.48 \pm 2\sqrt{\frac{0.48(1-0.48)}{23}}\right]$$

Def: A $1-\alpha$ sized confidence intend for population proportion p

$$\left(I_{1-\alpha, p} := \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]\right.$$

$$\left[0.48 \pm 2\sqrt{\frac{0.48(1-0.48)}{23}}\right] = [0.272, 0.688]$$

$$P(p \in [0.272, 0.688]) \neq 95\%$$