AMS 206
22 Jan 19

(1)

this time: samples & populations; next time: frequentist inference

read: D&S ch. 7, 8

J ch. 2, 3

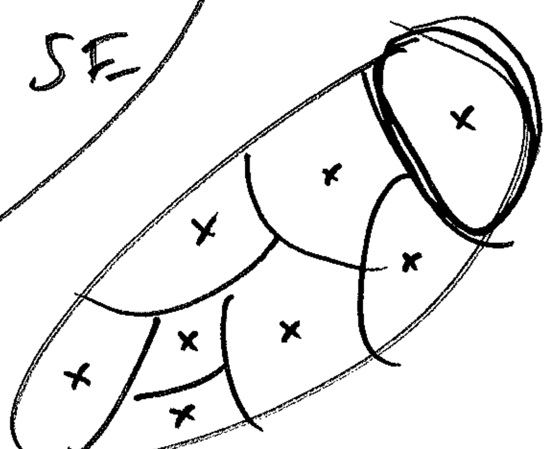
SRS
$\downarrow$

$$P(Y_1 = 7, Y_2 = 7 \mid B_2)$$

$$= \underbrace{P(Y_1 = 7 \mid B_2)}_{\frac{1}{3}} \cdot P(Y_2 = 7 \mid Y_1 = 7, B_2)$$

$$\frac{1}{3} \cdot 0 = 0 \checkmark$$

SE



window method

Q: What is the broadest scope of valid generalizability outward from my data?

$$SE_{IID}(\bar{y}) = \frac{\sigma}{\sqrt{n}}$$

finite population correction (FPC)

$$SE_{SRS}(\bar{y}) = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

$$\boxed{(Y_i \mid \theta\, \mathcal{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta) \quad i = 1, \dots, n}$$

$$\boxed{\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i}$$

$$E(\bar{Y}_n) = \theta$$

$$V(\bar{Y}_n) = \frac{\theta(1-\theta)}{n}$$

$$\bar{Y}_n \xrightarrow{P} \theta$$

WLLN

$n \to \infty$

$$\left(P(|\bar{Y}_n - \theta| > \epsilon) \to 0 \right)$$

$$\text{CLT: } \left(\frac{\bar{Y}_n - \cancel{\mu}}{\sqrt{\frac{\theta(1-\theta)}{n}}} \right) \xrightarrow{D} N(0, 1)$$

$$P_F(\text{Bob is HIV+} \mid B) = \text{undefined}$$

1 = HIV+
0 else

pop P^*_{Bob}

all people similar to Bob in all $\boxed{\text{relevant}}$ ways

Fisher: Bob's relevant subpopulation

$\xrightarrow{\text{at random}}$ $Y_1 [\, y_1 \,]$

IID

WLLN

(Bob is HIV+) $\longleftrightarrow$ P_F(an individual chosen at random from P^* is HIV+) ✓

P_F(Bob is HIV+ | Bob's blood test says HIV+)

WLLN $\quad Y_i \overset{IID}{\sim} \quad E(Y_i) = \mu$

$V(Y_i) = \sigma^2 < \infty$

$\bar{Y}_n \xrightarrow{P} \mu$

2 approaches to frequentist inference

① Neyman CS2

pop

p^*

sample

the obs.

AMI patients

1 = died
0 = didn't die?

die?

$N = ?$

1_s

0_s

like IID

1_s

0_s

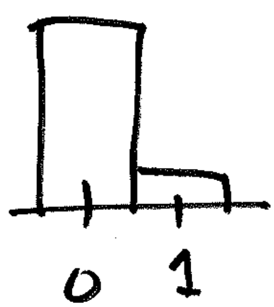
$n = 403$

mean $\theta = ?$

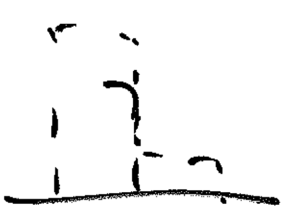
SD $\sigma = \sqrt{\theta(1-\theta)} = ?$

mean $\hat{\theta} = \frac{s}{n}$

$= \frac{72}{403} \doteq 0.18$



sample hist



pop. distn

inferential summary

unknown pop. quantity of main interest	θ = pop. 30-day mortality rate
estimate of θ (in absence of external info.)	$\hat{\theta} = \frac{s}{n} \doteq 0.18$
give-or-take for $\hat{\theta}$ as est. of θ	← $\hat{SE}(\hat{\theta}) \doteq 0.019$
95% interval for θ	

$$SE_{IID}(\hat{\theta}) = \sqrt{\frac{\theta(1-\theta)}{n}} = \sqrt{V_{RS}(\hat{\theta})} = \sqrt{V_F(\hat{\theta})}$$

$$\hat{SE}_{IID}(\hat{\theta}) \triangleq \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = \sqrt{\frac{\left(\frac{72}{403}\right)\left(\frac{403-72}{403}\right)}{403}}$$

is defined to be

$$\doteq 0.019$$