this time: estimating a population percentage

5 ① May

AMS5

next time: significance testing

read: FPP ch. 26

If you want to be considered for one of the note-taker jobs, please give me (soon) a copy of your notes for the last 2 classes, & put your name & email address on your notes

finishing CS 10 p. ⑧①

2 survival skills concerning diagrams (stat. models):

① which dataset (pop, sample, imag.) is being referred to?

② is the question about individual elements of the dataset or a summary?

big picture: p. ㉒

V B

point & interval estimates

ex. (#1 in hwk 4) ① (T or F) About 68% of the waybills in the sample were in the range $25.09 ± $0.69.

This is (F) because it's a statement about individual elements of sample dataset, & correct give or take

for elements of sample dataset is ② sample sd $s = \$31.40$ To make it ① should replace $0.69 by $31.40

[2] About 68% of the mugbills in pop were in the range $25.99 ± $0.69 (OK) X.

(F) Because elements in pop dataset differ from mean μ by an amount roughly σ in size, & our best est. of σ is $s = \$31.40$ Devil's Advocate: "I think real % of all 17-yr olds who can add fractions is 90%, & you got a really different answer in sample by unlucky sampling." (Plausible & in need of rebutting.)

pop: all 17-yr-olds in U.S. in 1975

1 = correct ans.
0 = wrong ans.

sample: the observed 17-yr-olds

estimate: imag. lots of possible $\hat{p}$'s

can add fractions?

N = big $\Rightarrow$ $\begin{bmatrix} 1\text{s} \\ 1 \\ 0\text{s} \end{bmatrix}$ $\doteq$ IID

like SRS

actual $\begin{bmatrix} 1\text{s} \\ 2 \\ 0\text{s} \end{bmatrix}$ $n = 950$

can add fractions?

mean $\bar{y} = \hat{p} = \frac{617}{950} \doteq 65\%$

$\begin{bmatrix} 65\% \\ 62\% \\ \vdots \end{bmatrix}$ a $\hat{p}$ $\uparrow$ ∞ $\downarrow$

long run mean (est.) $E(\hat{p}) = p$

long run SD $\hat{SE}(\hat{p}) = 1.5\%$

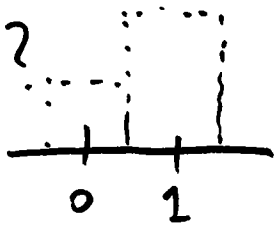
long run hist (**)

mean $\mu = p = ?$

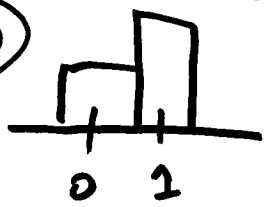
SD $\sigma = \sqrt{p(1-p)}$ hypothetical

IID

pop hist.



sample hist.



$\begin{bmatrix} 1\text{s} \\ 2\text{s} \\ 0\text{s} \end{bmatrix}$ $n = 950$

mean $\hat{p} = ?$ (ex. 62%)

This is a statistics problem because sample is known & pop. is unknown

---

$SE_{IID}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$ not usable because p is unknown

---

Fix: $\hat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.65)(0.35)}{950}} \cdot 100\%$

$= (0.015)(100\%)$

$= 1.5\%$

inferential summary

(unknown) quantity of interest	p = pop. % of all U.S. 17 yr olds who can add fractions in 1975
estimate	$\hat{p} = 65\%$
give or take ⊛	$\hat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.5\%$
95% interval ⊛ estimate (CI)	$\hat{p} \pm 2\hat{SE} = 65\% \pm 3\%$

⊛ requires imag. dataset

$SE_{IID}(\hat{p}) =$

$E_{IID}(\hat{p}) = E_{IID}(\bar{y}) = \mu = p$

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

$SE(\hat{p}) = \frac{\sigma}{\sqrt{n}}$

math fact:

SD σ of a pop with 0s & 1s is $\sqrt{p(1-p)}$

$= \sqrt{(\text{fract. of 1s})(\text{fract. of 0s})}$

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

(new useful formula)