STAT 131
6 Jun 20

THT3 #1(b)

DD extra office 1.5-hour session

$$Y_i = \theta + b + e_i$$

truth — bias — random "error"

$$(i = 1, \dots, n)$$

$$(e_i \mid \sigma^2) \overset{IID}{\sim} N(0, \sigma^2)$$

---

constants — r.v.

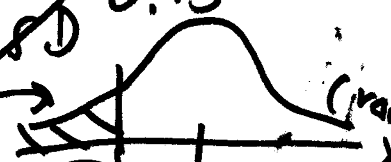
$$Y_1 = \theta + b + e_1 \sim N(0, \sigma^2)$$

$E(e_1) = 0$

true pH of lake (5.1) — bias of pH kit 0

$\sigma = 0.15$

SD 0.15 — PDF of Y_1



(raw units) y

$\doteq 25\%$ — want small

5.0 — 5.1 ← not threatened

standard units (z)

$$\frac{5.0 - 5.1}{0.15} = -0.67$$

$$E(Y_1) = E(\theta + b + e_1)$$
$$= \theta + b + E(e_1)$$
$$= \theta + b = 5.1 + 0 + 0$$

$$SD(Y_1) = \sqrt{V(Y_1)} = \sigma$$

$$V(Y_1) = V(\theta + b + e_1)$$
$$= V(e_1) = \sigma^2$$

population

all British adult hypertensive patients in mid 1970s

sample

the observed hypertensive patients

repeated sampling ~~dataset~~

possible $\bar{D}$s

$D \mid A \mid B$

representatively

~~like~~ actual

IID

$D = B - A \mid A \mid B$

$d_1 = +9$

$\vdots$

$d_n = +33$

$n = 12$

18.6

17.1

$\vdots$

$M \to \infty$

$N = ?$ (big) $\gg n$

$\bar{D} \to$ mean $\bar{d} = 18.6$

SD $s_D = 10.1$

sample hist.

pop mean $\Delta = ? = E(D_i)$

SD $\sigma_D = ? < \infty$

hyp. IID

pop PDF of D_i

$\bar{D}$

$n = 12$

mean $\bar{D} = ?$

(ex. 17.1)

long run mean

WLLN

$E_{IID}(\bar{D}) = \Delta$ ✓

estimated long run

$SD_{IID}(\bar{D}) = \sqrt{\hat{V}_{IID}(\bar{D})}$

$\hat{SD} = \hat{SE}(\bar{D}) = \frac{\cancel{\sigma_D}\, s_D}{\sqrt{n}}$

long run hist.

SD $\frac{\sigma_D}{\sqrt{n}}$

part 3 of CLT

PDF of $\bar{D}$

Δ

$$\hat{SE}_{IID}(\bar{D}) = \frac{\cancel{\sigma_D}\, s_D}{\sqrt{n}} = \frac{10.1 \text{ mmHg}}{\sqrt{12}} \doteq 2.92 \text{ mmHg}$$

t distribution

"the population represents the broadest scope of valid generalizability outward from your sample" ③

---

inferential summary

unknown pop. quantity of main interest	Δ = pop. mean improvement in s.d.p. under Captopril
estimate of Δ	$\bar{d}$ = 18.6 mm Hg
give or take for $\bar{d}$ as est. of Δ	$\widehat{SE}_{IID}(\bar{D}) \doteq 2.92$ mm Hg
99.9% CI for Δ	

repeated-sampling SD of rep. samples

confidence interval

---

math: If A then C

assumptions

conclusions / consequences

best math can offer in practical problem solving

If A_1 then C

A_2 C

$\vdots$

A_k C

$\{A_1, \ldots, A_k\}$ are the only assumption sets reasonably motivated by problem context

$\boxed{\vdots \quad C}$

how accurate is $\bar{D}$ as est of Δ?

r.v.

want

$$P_F\left(|\bar{D} - \Delta| \le \text{small (close to 0)} \right) = \text{big (close to 1)}$$

want small

frequentist