This time: probability + statistical models for means

Next time: " " " " " May 10, 2017

Read: DD (B) Ch. 11

LN pp. L-137 + L-144

Today: LN p. L-128 →

Extra office hours today @ 1:15-2:15 Jack's Lounge

Hypokalemia Case Study p. R-55

basic measurement error model: $\boxed{L-127}$

$$\begin{pmatrix}\text{obs.}\\ 1\end{pmatrix}_{Y_1} = \begin{pmatrix}\text{true}\\ \text{value}\end{pmatrix}_{\theta} + \begin{pmatrix}\text{bias}\end{pmatrix}_{b} + \begin{pmatrix}\text{random}\\ \text{error } 1\end{pmatrix}_{e_1}$$

e_1 → mean 0, → SD σ, → IID normal

$\vdots$

$$\begin{pmatrix}\text{obs.}\\ N\end{pmatrix}_{y_n} = \begin{pmatrix}\text{true}\\ \text{value}\end{pmatrix} + \begin{pmatrix}\text{bias}\end{pmatrix} + \begin{pmatrix}\text{random}\\ \text{error } n\end{pmatrix}_{e_n}$$

$$\begin{pmatrix}\text{mean of}\\ n \text{ obs } \bar{y}\end{pmatrix} = \begin{pmatrix}\text{true}\\ \text{value}\end{pmatrix}_{\theta} + \begin{pmatrix}\text{bias}\end{pmatrix}_{b} + \begin{pmatrix}\text{mean of } n\\ \text{random errors}\\ \frac{e_1+e_2+\dots+e_n}{n}\end{pmatrix}_{\bar{e}}$$

$\bar{y} = \theta + b + \bar{e}$

$\bar{y}$ → (mean of n obs.); θ (truth); b → (bias); $\bar{e}$ → (mean of random n errors)

assume that your measuring process is biases

ex. assume b = 0 →

ex. hypokalemia

pretend we know $\theta = 3.8$

obs.

$$3.9 = 3.8 + 0 + (+0.1)$$
$$3.5 = 3.8 + 0 + (-0.3)$$
$$\vdots$$
$$3.8 = 3.8 + 0 + (+0.0)$$
$$\bar{y} = 3.8 + 0 + \frac{(+0.1) + (-0.3) + \ldots + (0.0)}{n}$$

we will get to "enjoy" ~~canca~~ cancellation of ⊕ or ⊖ errors, with the result that the typical size $|\bar{e}|$ of e will be smaller than any of the e

as $n \uparrow$, the expected value of $|\bar{e}|$ goes to 0, so $\bar{y} \to \theta + b + 0$ (as $n \to \infty$)

Ie. if $b = 0$, $\bar{y}$ get close + closer to the truth θ as ~~the~~ n increases

<u>but</u> if $b \neq 0$, $\bar{y} \to \theta + b$ not θ: w/ a bias measuring process you can't make the bias go away just by taking more obs. + averaging them

(to decrease uncertainty about θ), get more + more good data

good → unbiased measuring process

conceptual pop. of all possible blood samples from you

pop.

pretend θ/truth $= 3.8$

sample (the observed)

blood sample from you (potassium)

repeated sampling (imaginary data set)

$N = \infty$

$$\begin{bmatrix} \; \\ \; \end{bmatrix}$$

at random w/ or w/out replacement IID

$$\begin{bmatrix} 3.9 \\ 3.6 \\ \vdots \end{bmatrix} n = 4$$

$$\begin{bmatrix} 3.85 \\ 3.67 \\ \vdots \end{bmatrix} \to \bar{y} \quad n = \infty$$

mean $\mu = \theta = 3.8$

SD $\sigma = 0.2$

mean $= \bar{y} = ?$

ex. (3.85)

IID

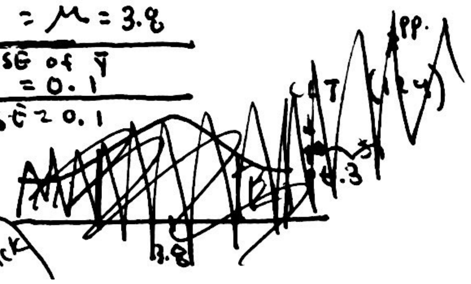
$$\begin{bmatrix} \; \end{bmatrix} n = 4$$

$\bar{y} = ?$ (ex. 3.67)

pop. hist.

normal curve

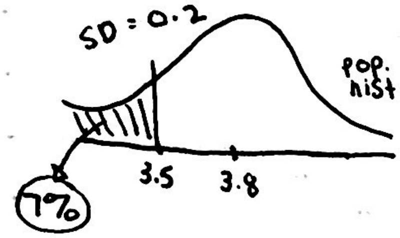
long run mean	EV of $\bar{y}$ $= \mu = 3.8$
long run SD	SE of $\bar{y}$ $= 0.1$
	$\hat{SE} \approx 0.1$
long run hist	

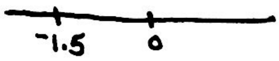


on back

P(~~dia~~ misdiagnosis w/ $n=1$)

$= P(Y_1 < 3.5) =$





$\frac{3.5 - 3.8}{0.2} = \frac{-0.3}{0.2} = -1.5$

P(misdiagnosis w/ $n=4$)

$= P(\bar{y}$ based on $< 3.5) = ?$

$n=4$ obs.

EV of $\bar{y}$: $\boxed{E_{IID}(\bar{y}) = \mu}$ math fact

$E_{IID}(S) = n\mu$ but $\bar{y} = \frac{S}{n}$

$\boxed{\text{SE of } \bar{y} = \frac{\sigma}{\sqrt{n}}}$

N	X
μ	X
σ	$\sigma\uparrow$, SE($\bar{y}$)$\uparrow$
n	$n\uparrow$, SE($\bar{y}$)$\downarrow$
M	X

$\bar{y}$ is our best estimate of μ; SE of $\bar{y}$ represents how much uncertainty (noise) there is in the process of using $\bar{y}$ to get estimate θ

uncertainty about μ on the bias of $\bar{y}$ as an estimate of ~~μ~~

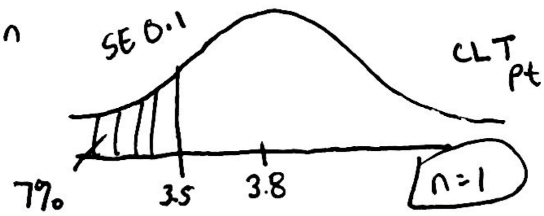
μ goes down w/ n, but only at a $\sqrt{n}$ rate

ie. unfortunately, to get SE ($\bar{y}$) in half ~~y has to~~ you have to quadruple the sample size

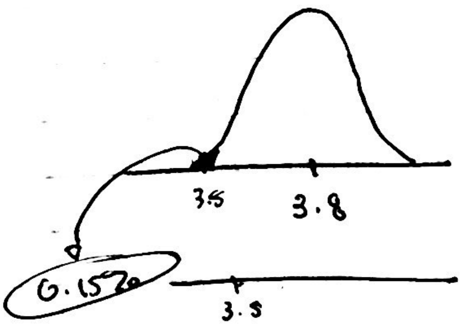
[square root ~~cute~~ law]

here SE $(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{0.2}{\sqrt{4}} = 0.1$

long run hist.



CLT pt III



$$\frac{3.5 - 3.8}{0.1} = -3$$

n	cost	P(mis diagnosis)
1	\$25	7%
4	\$100	0.15%
	cost	benefit