this time: measurement error; prob. models for means

next time: statistical models for means; interval estimation

read: DD (B) ch 11 (AMS 7 220 ct 15)

LN: pp. L-(137)-(156) (1)

this time: LN pp. L-(127) →

3.9

#is > 3.5

[16 oz, 16, 16, ⋮] → [16.0, 16.0, 16.0, ⋮]

deterministic (same value every time)

[16.01, 15.98, }, 15.99]

probabilistic (stochastic) (different value every time)

basic model

$$\begin{pmatrix}\text{obs.}\\1\end{pmatrix} = \begin{pmatrix}\text{true}\\\text{value}\end{pmatrix} + (\text{bias}) + \begin{pmatrix}\text{random}\\\text{error}\\1\end{pmatrix}$$

y_1

IID mean 0

$$(\text{obs. } 2) = \begin{pmatrix}\text{true}\\\text{value}\end{pmatrix} + (\text{bias}) + \begin{pmatrix}\text{random}\\\text{error}\\2\end{pmatrix}$$

y_2

⋮

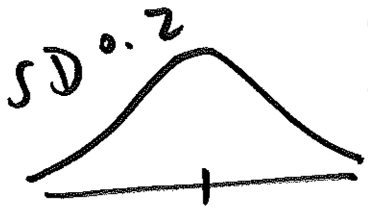
conceptual pop. data [2]

all possible ~~potassium~~ readings

$N = \infty$

mean $\mu = 3.8$

SD $\sigma = 0.2$



sample [1]

the observed ~~blood samples~~

potassium

y_1 3.9

3.5

y_n

(IID) →

$n = 4$

mean $\bar{y} = ?$

(ex. 3.6)

imagine

$n = 4$

mean $\bar{y} = ?$

(ex. 3.9)

imag. data (2)

possible $\bar{y}$s [3]

3.6

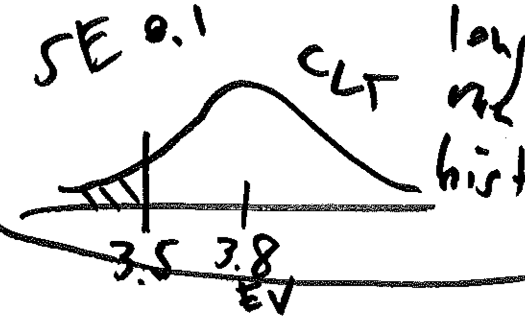
3.9

a lot

$M = \infty$

long run mean (*) EV of $\bar{y}$ = 3.8

long run SD (**) SE of $\bar{y}$ = 0.1



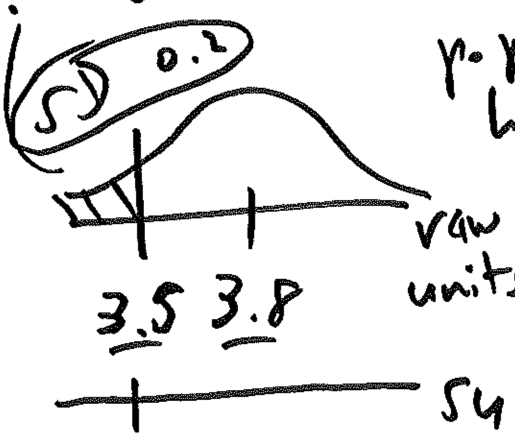
(*) expected value of $\bar{y}$

= EV of $\bar{y}$ = $E(\bar{y})$ = $E_{IID}(\bar{y})$ = ?

math fact:

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

$.0668 \doteq 7\%$



pop. hist. (1 reading at a time)

saying you're / hypo. when not true ↓

$P(\text{misclassification})$

$= P(Y_1 < 3.5)$

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

$= -1.50$

9.03

(**) long run SD of $\bar{y}$ = standard error of $\bar{y}$

= SE of $\bar{y}$ = $SE(\bar{y})$ =

$SE_{IID}(\bar{y}) = ?$

N	no
μ	no
$\sigma\uparrow$	$SE(\bar{y})\uparrow$
$n\uparrow$	$SE(\bar{y})\downarrow$

math fact:

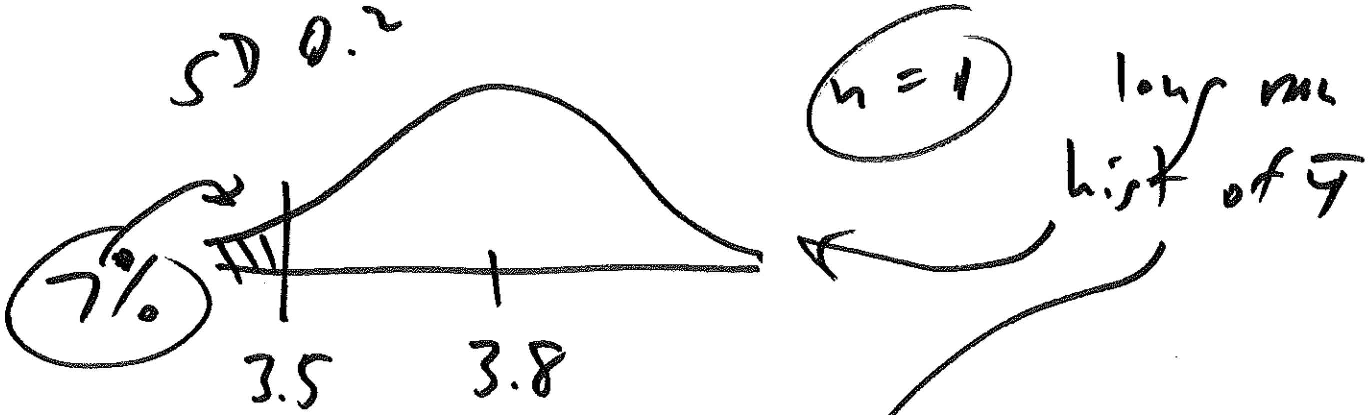
$$SE_{IID}(\bar{y}) = \frac{\sigma}{\sqrt{n}}$$

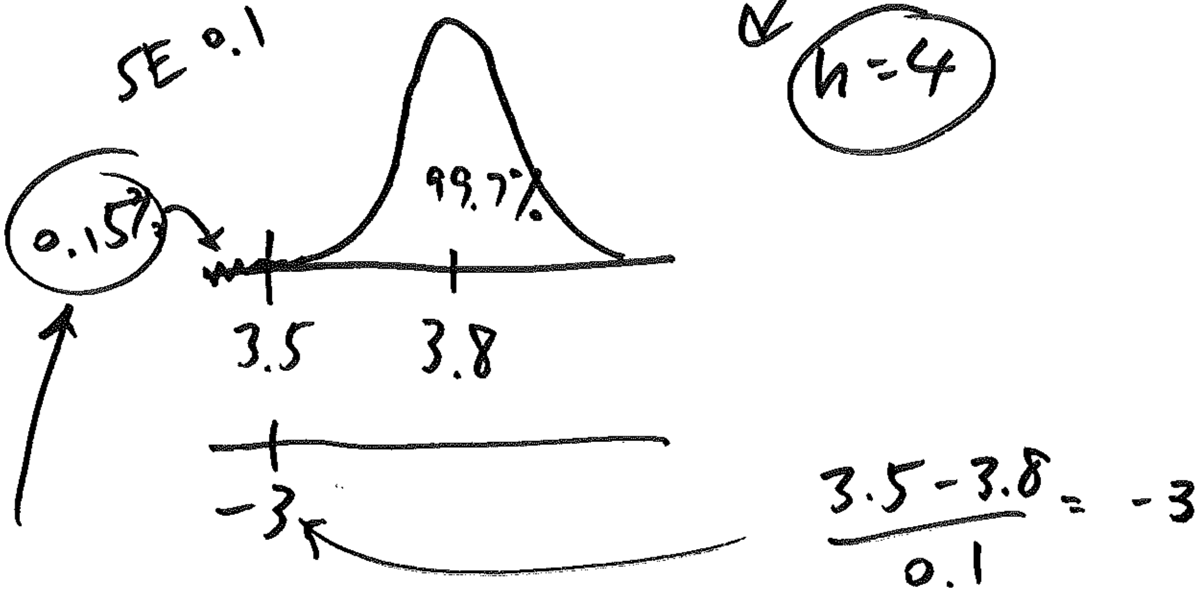
square root law

to cut the SE in half, you have to quadruple the sample size

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

$= 0.1$





P(misdiagnosis w/ n = 4) = 0.15%

n	error rate	cost
1	7%	$25
4	0.15%	$100

(big n + longer cost)

benefit (small error rate with big n)