this time: measurement error; prob. models for means (averages) | 28 Apr AMS 5 | 1

next time: estimating averages | read: (D) ch. 11) FPP ch. 23, 24

measurement error (FPP ch. 6): No matter how carefully made, any measurement if repeated, could come out differently. why?

basic measurement error equation:

(each indiv. meas.) = (exact ("true") value) + (bias) + (a random error) (*)

systematic error → bias

bias → 0 in (this case study)

3.7 = potassium reading #1 = 3.8 + 0 + (−0.1)

4.0 = #2 = 3.8 + 0 + (+0.2)

⋮

observable: 3.7, 4.0, … — vary around truth with an SD of 0.2

unobservable: 3.8, 0, (−0.1)

(*) these fluctuate around 0) (the mean of random errors is 0 & SD is 0.2)

pop — conceptual: all possible potassium readings for you

sample — the observed potassium readings

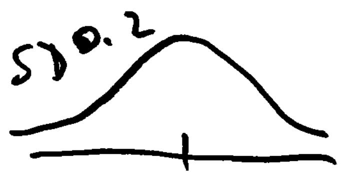
imag. data — possible $\bar{Y}$s

potassium

$N = \infty$

mean $\mu = 3.8$

SD $\sigma = 0.2$



pop. hist.

IID →

potassium

$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ $n = 4$

mean $\bar{Y} = ?$ (ex. 3.6)

IID →

$n = 4$

mean $\bar{Y} = ?$ (ex. 4.0)

$\begin{bmatrix} 3.6 \\ 4.0 \\ \vdots \end{bmatrix}$ a $\bar{Y}$ ↕ ∞

long run mean: $E(\bar{Y}) = 3.8$

long run SD: $SE(\bar{Y}) = 0.1$

long run hist — 3.8

hist of Y_1

SU: -1.5, $+1.5$

$-1.5 \leftarrow \dfrac{3.5 - 3.8}{0.2}$

$P(\text{misdiagnosis with } n = 1)$

$= P(Y_1 < 3.5)$

$= 6\frac{1}{2}\%$

($n = 4$) P(misdiagnosis with $n = 4$)

$= P(\bar{y} < 3.5$ with $\bar{y}$ based on $n = 4$ readings)

Expected value of $\bar{y}$ = $E(\bar{y})$ = EV

$= E_{IID}(\bar{y}) = \mu = 3.8$

---

standard error of $\bar{y}$ = $SE(\bar{y})$ = SE = $SE_{IID}(\bar{y})$

N ✗
μ ✗
$\sigma \uparrow$ $SE(\bar{y}) \uparrow$
$n \uparrow$ $SE(\bar{y}) \downarrow$

(noise of uncertainty of $\bar{y}$ as an estimate of μ)

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

most important formula in statistics

uncertainty about μ goes down as $n \uparrow$ but only at a $\sqrt{n}$ rate: to cut SE in <u>half</u> I need to <u>quadruple</u> (n).

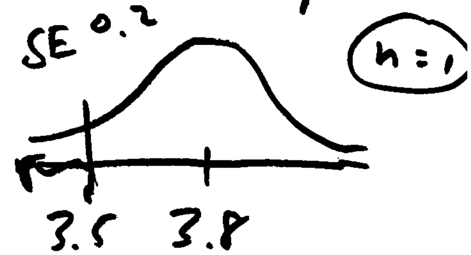
$$SE_{IID}(S) = \sigma\sqrt{n}$$

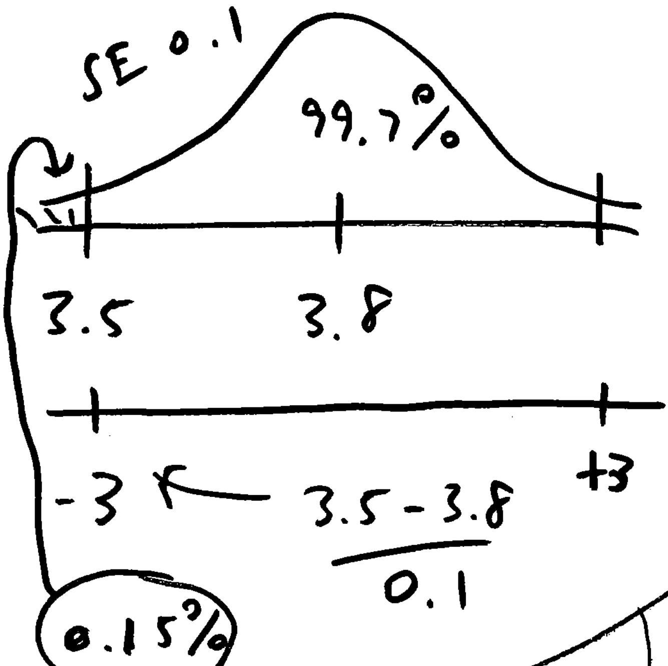
(S = sum)

please don't confuse these 2 formulas

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

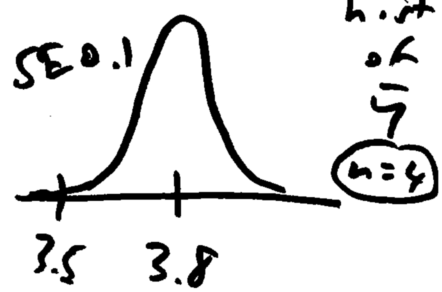
hist of $\bar{y}$





long run hist of $\bar{y}$ with $n = 4$

hist of $\bar{y}$



Cost	n	P(incorrect diagnosis)
\$25	1	6.5%
\$100	4	0.15% (benefit)

here

consequences of misdiagnosis are relatively mild (eat some bananas when you didn't need to)

(decision theory?) (utility) ← how much worth you attach to correct diagnosis)