AMS 7
20 Oct 09

this time: probability models for sums & means

next time: interval estimation

read: DD ch. 11

(1)

take-home mid term handed out next lecture, due 1 week later

lab 2 due by 5pm this Fri 23 Oct

standard error of S = SE =

math fact:

$SE_{IID}(S) = \frac{\sigma\sqrt{n}}{1} = \sigma\sqrt{n}$

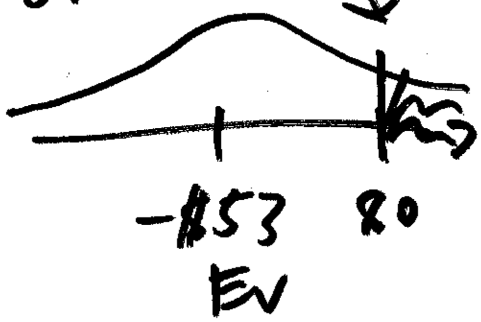
N	X
μ	X
σ ↑	SE ↑
n ↑	SE(S) ↑

$= \begin{pmatrix} \text{top} \\ \text{SD} \end{pmatrix} \sqrt{\text{\# of draws}}$

here $SE(S) = \$5.76\sqrt{1000}$

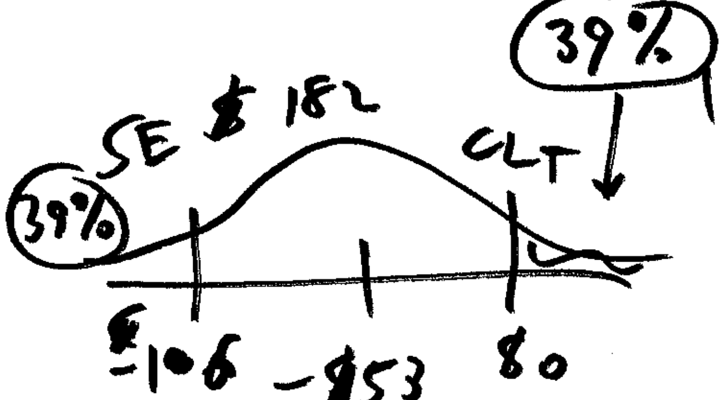
$= \$182$

SE \$182



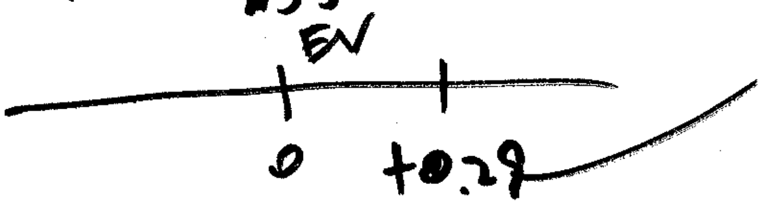
long run hist of S

Central Limit Theorem

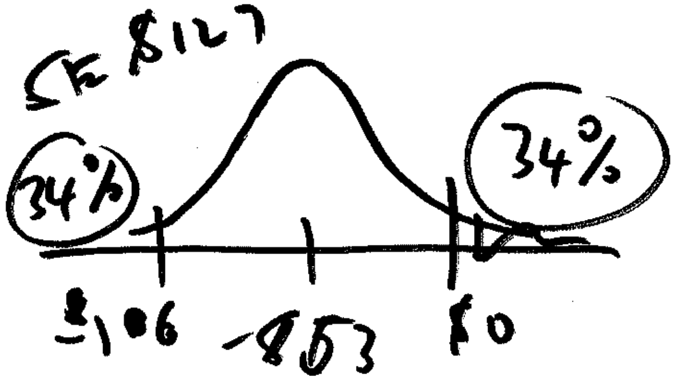


long run hist of $

single #

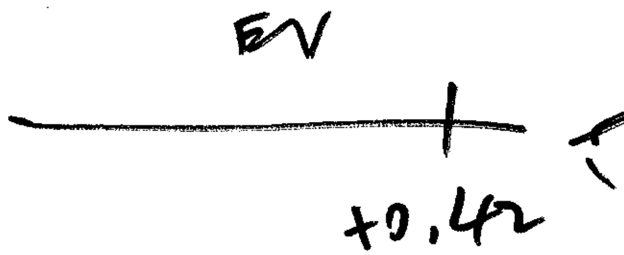


$$\frac{\$0 - (-\$53)}{\$182} = +0.29$$



long run hist of $

split



$$\frac{\$0 - (-\$53)}{\$127} = +0.42$$

single # : risk-seeking

split : more risk-averse

math fact: bold play is optimal (wins end ahead) when game is stacked against you

$$\begin{bmatrix} 16 \text{ oz} \\ 16 \\ \vdots \\ 16 \end{bmatrix} \qquad \begin{bmatrix} 16.0 \\ 16.0 \\ \vdots \\ 16.0 \end{bmatrix} \qquad \begin{bmatrix} 15.97 \\ 16.01 \\ \vdots \\ 15.99 \end{bmatrix}$$

deterministic

stochastic (random)

---

basic measurement error model

---

observable

not observable

IID mean 0

$$\text{observation } 1 = \begin{pmatrix}\text{true}\\ \text{value}\end{pmatrix} + (\text{bias}) + \text{random ``error'' } 1$$

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

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\text{obs } n = \begin{pmatrix}\text{true}\\ \text{value}\end{pmatrix} + (\text{bias}) + \text{random ``error'' } n$$

<u>good</u> data-gathering activity:

unbiased (bias = 0)

spse truth = 3.8 (not hypothetical) & measuring process unbiased

$$3.9 = 3.8 + 0 + (+0.1)$$
$$3.5 = 3.8 + 0 + (-0.3)$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$
$$4.0 = 3.8 + 0 + (+0.2)$$

mean of n obs. $= 3.8 + 0 + \dfrac{(+0.1) + (-0.3) + \ldots + (+0.2)}{n}$

mean of n random errors with mean 0

cancellation of (+), (−) HJ:
mean of n IID random errors (each with mean 0) will likely be a lot closer to 0 than any of the individual errors themselves

5

pop
conceptual: all possible measurements

unbiased sample
true value 3.8
the observed measurements

imag. data
possible $\bar{Y}_s$

potassium

$N = \infty$

like SRS = IID

potassium
Y_1
⋮
Y_4
$n = 4$

mean $\bar{Y} = ?$
(ex. 3.7)

3.7
4.0
⋮
$M = \infty$

long run mean
expected value of $\bar{Y} = \mu = 3.8$

IID

mean $\mu = 3.8$

SD $\sigma = 0.2$

IID

$n = 4$

long run SD
standard error of $\bar{Y} = 0.1$

SD 0.2

pop hist

3.8

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

⋮

long run hist
SE 0.1
CLT

3.8

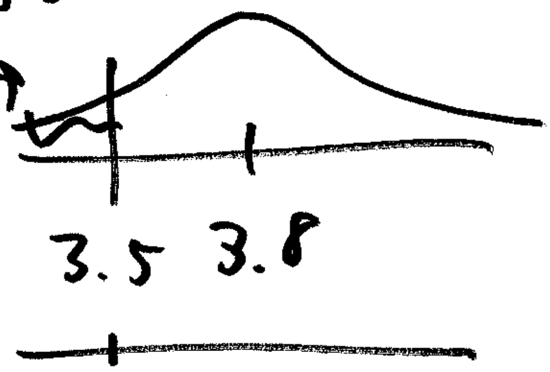
benefit

n	P(misdiagnosis)	cost
1	7%	\$25
4	0.15%	\$100

you need to bring in your values (utility) p(misdiagnosis with $n=1$) =

SD 0.2

pop. hist. = hist of 1 reading at a time



7% $\leftarrow \frac{3.5-3.8}{0.2} = \frac{-0.3}{0.2} = -1.5$

p(misdiagnosis with $n=4$) = ?

~~math fact~~:

expected value of $\bar{y}$ = EV =

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

standard error of $\bar{y}$ = SE = $SE_{IID}(\bar{y}) = \frac{\sigma}{\sqrt{n}}$ math fact

$= \frac{0.2}{\sqrt{4}} = 0.1$

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

note: very different from $SE(s) \uparrow$ as $n \uparrow$