AMS 7
Lecture 8
7/6/16

R-58 Case Study: Hypokalemia

lets talk about butter + measurements

I.J.
1lb pack
butter

$\begin{bmatrix} 16.02 \\ \vdots \\ 16 \end{bmatrix}$ $\begin{bmatrix} 16.0 \\ 16.0 \\ \vdots \end{bmatrix}$ $\begin{bmatrix} 15.98 \\ 16.01 \\ 15.97 \end{bmatrix}$

__Deterministic__: always get same thing

__Probabilistic/Stochastic__: the measurement varies haphazardly "at random" from repititi to repetition

__Basic measurement error model__

raw value

$(\text{observation } 1)_{15.98} = (\text{true value})_{16.00} + (\text{bias})_{0.00} + (\text{"random error 1"})$ $\uparrow = -.02$

...

$(\text{observation } n)_{16.00} = (\text{true value})_{16.00} + (\text{bias})_{0.00} + (\text{"random error n"})_{+.00}$

y_i $\qquad \theta \qquad B \qquad e_i$

$y_1 = \theta + b + c$

$y_n = \underbrace{\theta + b + e_n}_{\text{not observable}}$

$\bar{y} = \theta + b + \frac{e_1 + \ldots + e_n}{n}$

Assumption: errors have mean 0
errors are IID

* Cancellation of (+) and (−) errors means $\bar{e}$ is likely closer to 0 than any of the ind. errors going into $\bar{e}$. Averaging replications = good.

As n increases, $\bar{e} \to 0$
So $\bar{y}$ approaches $(\theta + b)$

truth + bias

With unbiased measuring process, replication followed by averaging is guaranteed to get closer to truth.

unbiased → replication + mean
biased → replication doesn't help bc bias is not removed by taking mean

pop	sample	imag data set
conceptual population	observed measure.	all possible $\bar{y}$s

$N = \infty$

pop: K+

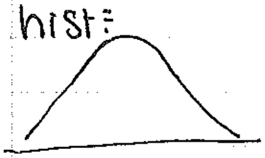
sample: K+, y_1 … y_n, n=

$\bar{y}$ mean

imag data set: $\bar{y}_1$, $\bar{y}_2$, …, $\bar{y}_n$

EV of $\bar{y} = \mu = 3.8$
SE of $\bar{y} = 0.1$

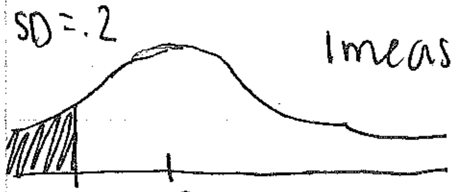
$\mu =$
$\sigma = 0.2$
hist:



SE = 0.1

3.8

SD = .2



3.6 3.8

1measurement
n=1

P(misdiagnosis) =

7%

$\frac{3.6 - 3.8}{.2} = -1.5$ .0668 = 7%

Now, P(misdiagnosis) when $n=4$

so $\underbrace{P(\bar{y} < 3.5)}_{\text{misdiagnosis}}$ with $n=4$

Probability $\bar{y} < 3.5$, get a lot of $\bar{y}$ and find proportion < 3.6

Now we are finding "a lot" of $\bar{y}$s

all possible $\bar{y}$s

imag data set

3.45
3.9

~~n~~ now called $M \Rightarrow \infty$

μ must be 3.8
$\sigma = .2$

Long Run Mean $\bar{y}$ = Expected Value $\bar{y}$
$= E_{IID}(\bar{y}) = \mu = 3.8$

Long Run SD = Standard Error $\bar{y}$
SE of $\bar{y}$ = $SE_{IID}(\bar{y})$ =

$\bar{y}$ is a good estimate of μ

SE of $\bar{y}$ = uncertainty of $\bar{y}$ as estimate of μ

$SE_{IID}(\bar{y})$ =
$\sigma \uparrow SE(\bar{y}) \uparrow$
$n \uparrow SE(\bar{y}) \downarrow$

Square Root Law: to cut your uncertainty about μ on the basis of $\bar{y}$ in half, you have to quadruple the sample size

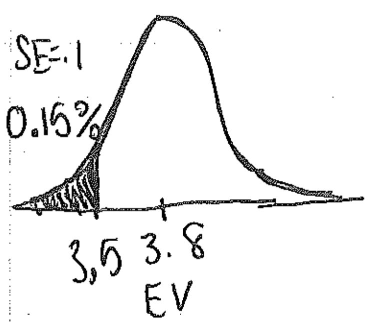
At the casino, we were looking for netgains

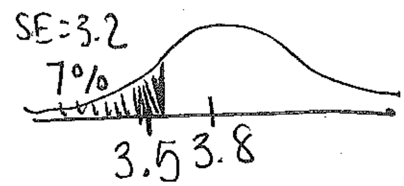
Here, we are focused on means, $\bar{y}$.

SE of $\bar{y}$ $\frac{\sigma}{\sqrt{n}} = \frac{.2}{\sqrt{4}} = .1$

Long Run SD = SE $\bar{y}$ = 0.1

Long Run Hist.
$\bar{y}$ w/ $n=4$ | w/ $n=1$ (POP)

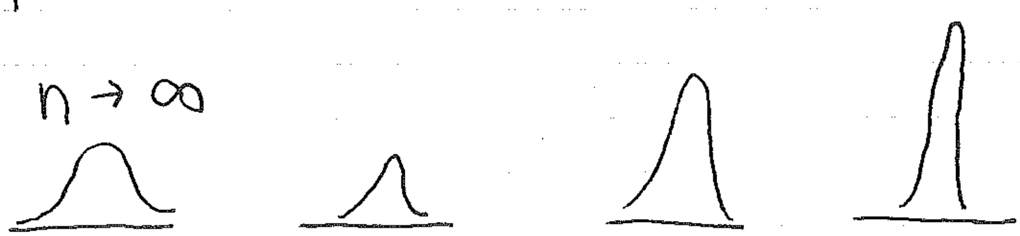




$\frac{3.5-3.8}{.1} = .0013$

7% too high
0.13% okay

As $n \to \infty$



$n \to \infty$

small = good

n	benefit misdiagnosis probability	cost
1	7%	\$25
4	.15%	\$100

AMS 1
Class|Lecture 8 cont.
7/6/16

Section 4: Statistical Inferences

L-13a

Pop
all intertidal crabs (similar)
T°C

N = ?
(big)

μ = ?
σ = ?
Pop Hist. ?

?

like at random

like IID

Sample
observed crabs
int. T (°C)

y_1 26.8
24.6
⋮
y_n 25.4

n = 25

$\bar{y} = 25.0$
$s = 1.34$

sample hist.

imag data