Discussion section, week of 4-8 Nov 19

R - 56, 57 #1, 2

STAT 7
4 Nov 19

G = (card really good) 1

B = ( ——— bad)

truth

$\oplus$ = (system says card is bad)

$\ominus$ = ( ——— good)

system says

$1\% = P(B)$

$97\% = P(\ominus \mid G)$

$98\% = P(\oplus \mid B)$

show:

$P(B \mid \oplus) = 25\%$

P(raining | clouds) = low

P(clouds | raining) = high

almost never true that $P(A \mid B) = P(B \mid A)$

a lot like problem 6 on midterm

truth

| what system says | B | G | |
|---|---|---|---|
| + | 98 | 297 | 395 |
| − | 2 | 9,603 | 9,605 |
| | 100 | 9,900 | 10,000 |

2 x 2 contingency table (gender & MLP)

$P(B) = .01$

(!)

$$\boxed{P(B \mid +) = \frac{98}{395} \doteq 25\%}$$

$$\boxed{P(- \mid G) = 0.97}$$

$(0.97) \cdot (9900) = 9,603$

$P(+ \mid \underline{\underline{B}}) = 0.98$

$(0.98)(100) = 98$

real world

Pimlico

the total weight of a fully loaded escalator at rush hour

is like

the sum $S$ of $n = 192$ IID draws from pop. (*)

math world

pop. *

London adults riding Pimlico esc.

weight (lb.)

$N = ?$ (big)

like IID

sample

the observed people

weight (lb.)

$n = 192$

imag. data

possible values of $S$

30502

31820

⋮

$M \to \infty$

sum $S = ?$ (ex. 30,502 lb.)

mean $\mu = 158$ lb.

SD $\sigma = 33$ lb.

like IID

$n = 192$

pop. hist.

⋮

long run mean: expected value of $S = 30336$

long run SD: standard error of $S = 457$ lb.

sum $S = ?$ (ex. 31,820)

**

long run hist.

30336

$\frac{1}{1000^2}$ .0001

SE 457

30336

3.719

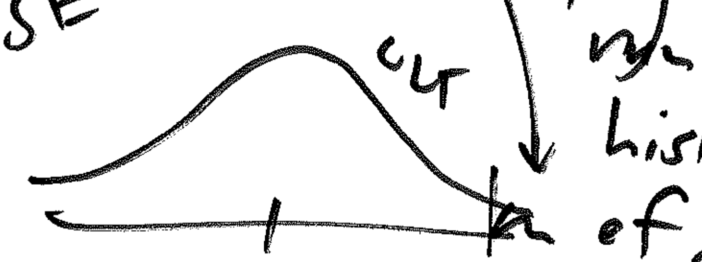
P(breaking) = P( $S > 31{,}400$ lb.) = ?

$$\begin{pmatrix}\text{expected}\\\text{value}\\\text{of } S\end{pmatrix} = \begin{pmatrix}\text{EV}\\\text{of}\\S\end{pmatrix} = E_{IID}(S) = n\mu$$

$$= \begin{pmatrix}\#\text{ of}\\\text{draws}\end{pmatrix}\begin{pmatrix}\text{pop.}\\\text{mean}\end{pmatrix} = 192\,(158\text{ lb.})$$

$$= 30{,}336\text{ lb.}$$

SE 457 lb.




$$\begin{pmatrix}\text{standard error}\\\text{of } S\end{pmatrix} = \begin{pmatrix}\text{SE of } S\end{pmatrix} =$$

$$2.33 \doteq \frac{31400\,\cancel{\text{lb.}} - 30336\,\cancel{\text{lb}}}{457\,\cancel{\text{lb}}}$$

$$SE_{IID}(S) = \sigma\sqrt{n} = (33\text{ lb.})\sqrt{192} \doteq 457\text{ lb.}$$

1% way too big or error tolerance: fails every 100 trips, but 90 full trips / day: $\left(\doteq 1 \frac{}{\text{fails}}/\text{day}\right)$

R-56 5

#1

G = (card really is good)

B = ( bad)

truth

$\oplus$ = (system says card is bad)

$\ominus$ = ( good)

system says

$1\% = P(B)$

$97\% = P(\ominus | G)$

$98\% = P(\oplus | B)$

problem 6 on midterm

show: $P(B | \oplus) \doteq 25\%$

rare to have $P(A | B) = P(B | A)$

P( raining | clouds) = low

P( clouds | raining) = high

gender, MCP case study

2×2 contingency table

cross-tabulation

truth

| | | B | G | |
|---|---|---|---|---|
| what system says | ⊕ | 98 | 297 | 395 |
| | ⊖ | 2 | 9,603 | 9,605 |
| | | 100 | 9,900 | 10,000 |

$P(B \mid \oplus) = \frac{98}{395} \doteq 25\%$ (!)

$P(B) = 1\%$

$(0.01) \cdot 10{,}000$

$P(\ominus \mid G) = 0.97$

$(0.97) \cdot (9900) = 9{,}603$

$P(\oplus \mid B) = 0.98$ : $(0.98) \cdot (100) = 98$

R-56, 57 #2

real-world

the total weight fully loaded at rush hour of the Pimlico esc. $\boxed{\text{is like}}$ the sum of $n$ IID draws from population ⊛

math world

(*) pop. (possible people) British adults who ride Limtico and weight of r.o.

sample: the observed people on esc. weight

imag. data: possible values of $S$

$N = ?$, big

like IID

$n = 192$

sum $S = ?$ (ex. 30,500)

30,500 lb.

31,150

⋮

$M \to \infty$

mean $\mu = 158$ lb.

SD $\sigma = 33$ lb.

pop. hist.

like IID

$n = 192$

sum $S = ?$ (ex. 31,150)

long run mean: expected value of $S = 30336$ lb

long run SD: standard error of $S$

$$\left(\begin{array}{c}\text{expected value of}\\ S\end{array}\right) = \left(\begin{array}{c}\text{EV}\\ \text{of } S\end{array}\right) = 30{,}336 \text{ lb.}$$

$$= E_{IID}(S) = n \cdot \mu$$

$$= \left(\begin{array}{c}\#\text{ of}\\ \text{draws}\end{array}\right) \cdot \left(\begin{array}{c}\text{pop}\\ \text{mean}\end{array}\right) = 192(158 \text{ lb.})$$

(just like roulette)

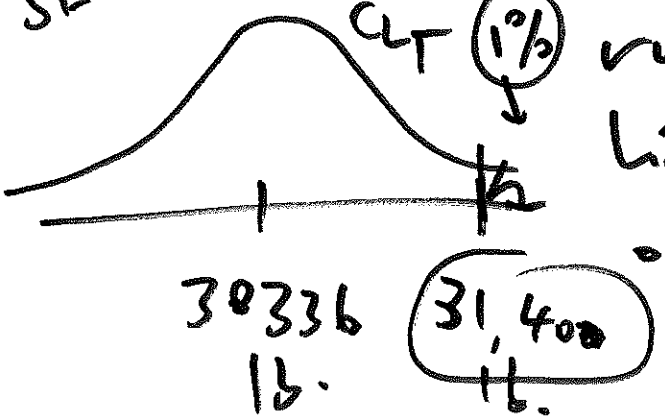
(**) long run hist.

30336

$$P(\text{overload}) = P(S > 31{,}400 \text{ lb.}) = ?$$

SE 457 lb.

long run hist. of $S$

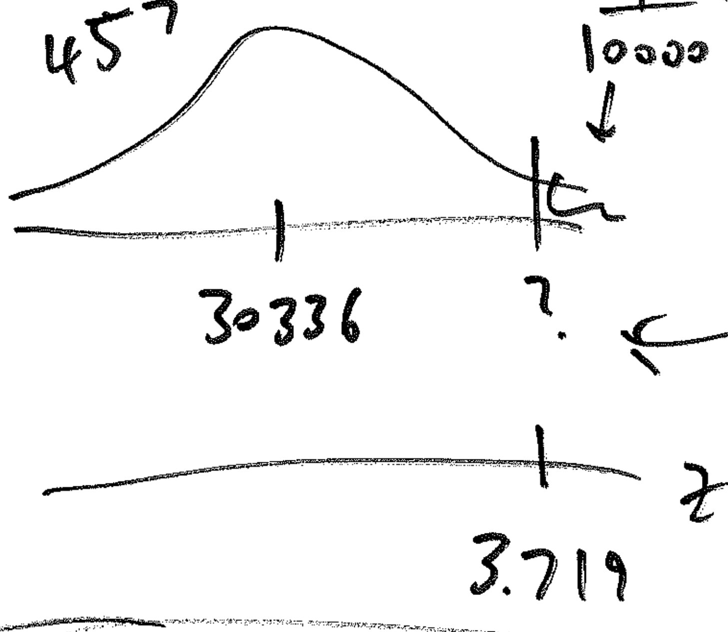



$$\left(\text{standard error of } S\right) = \left(\text{SE of } S\right) = SE_{IID}(S) = \sigma\sqrt{n} = (33 \text{ lb.})\sqrt{192} \doteq 457 \text{ lb.}$$

SU(z)

$$\frac{31{,}400 \cancel{\text{lb.}} - 30{,}336 \cancel{\text{lb.}}}{457 \cancel{\text{lb.}}} = \underline{2.33}$$

$\frac{1}{10000} = .0001$




$(3.719)(457 \text{ lb.}) + 30336 \text{ lb.}$

like #4 on midterm