Next time: estimating a population percentage

Read: way to improve class: legibility & logical flow of document camera notes

Solution: 2 or 3 people in class will be paid to create legible & logical notes, which will be posted on web along with my chicken scratches in big syllabus picture.

If pop known, we reason from pop to sample (whole to part) (deductive reasoning);
If sample drawn at random this is probability in action.
(updated copy of p. 6 from lecture notes of 26 Apr)

sample

amount owed

N = 22,984
mean $\mu = 2$
SD $\sigma = 2$

$814.16$ $100.00$

$51$

$\text{historical}$

population

amount owed

2

$\text{n = 2,072}$

$\text{SD}$ $S = 831.40$

mean $\bar{y} = 28.09$

$28.42$

$26.09$

$E(\bar{y}) = \mu$

$\text{long run}$

mean

(estimated)

long run $SE(\bar{y})$

$n = 2,072$

mean $\bar{y} = \frac{1}{n}$ $(\text{ex. } \frac{825.42}{26.09} S = 80.69)$
if instead sample is known & chosen

random

we reason from sample to pop (part - whole) (inductive reasoning)

(b) statistical inference

sample SD $s = 8.3140$ is larger than

$\overline{y}$,

a good single # guess


how much?

Long run mean of $\overline{y}$ is $\bar{y}$

$E(\overline{y}) = \mu$

is $\overline{y} \approx 82.809$, & this variable

$\mu$ is right skewed to tail, (like #)

looks

a (point estimate) for $\mu$ based on sample

is $\overline{y} = 82.809$; but give or take

$1.7$ so boxed value of $\overline{y}$

is $E(\overline{y}) = \mu$.
\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Quantity of interest: & \( \mu = \text{mean amount of money owed in population of utility bills} \) \\
\hline
Estimate: & \( \overline{y} = 828.09 \) \\
\hline
Give or take: & \( SE(\overline{y}) = 0.69 \) \\
\hline
95\% CI for \( \mu \) & \( \overline{y} \pm 2SE(\overline{y}) = (26.71, 29.47) \) \\
\hline
\end{tabular}
\end{table}

The long-run \( S^2 \) of \( \overline{Y} \) in an imaginary data set is the standard error of \( \overline{Y} = SE = SE(\overline{Y}) \)

\[ SE_{\text{SRS}}(\overline{Y}) = \frac{S}{\sqrt{n}} \]

\[ SE_{\text{IID}}(\overline{Y}) = \left[ SE_{\text{IID}}(\overline{Y}) \right] \cdot \left[ \sqrt{\frac{N-n}{N-1}} \right] \]

The correction factor (CF) for \( SE_{\text{SRS}} \)

\[ CF = \begin{cases} 
1 & \text{if } n = 1 \\
0 & \text{if } n = N 
\end{cases} \]

\[ CF = \frac{1}{\sqrt{N-n}} \]

\( \frac{1}{\sqrt{N-n}} \) makes sense when \( n \ll N \).
(cf relevant for #1, hwk 4, p. 54)

& not relevant for any other problems in this class (in other words, in all other problems \( n < N \) & \( CF = 1 \))

\[ SE_{IID}(\bar{X}) = \frac{5}{\sqrt{n}} \]

nice but unusable when (sample known & pop unknown)

fix: \( \bar{X} \) should be a good estimate of \( \sigma \), so use estimated \( SE \) of \( \bar{X} \):

\[ SE_{IID}(\bar{X}) = \frac{\bar{X} \bar{S}}{\sqrt{n}} \]

\[ SE_{IID}(\bar{X}) = \frac{5}{\sqrt{2072}} \approx \frac{5}{45.69} = 0.09 \]

\[ SE_{IID}(\bar{X}) = \frac{0.3140}{\sqrt{2072}} = 0.069 \]

\[ SE_{SRS}(\bar{X}) = \frac{5}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \]

\[ CF = \frac{\sqrt{22984 - 2072}}{\sqrt{22984 - 1}} \]

\[ SE_{SRS}(\bar{X}) = (0.069)(0.95) \approx 0.066 \]

\[ SE_{SRS}(\bar{X}) = \sqrt{0.91} = 0.95 \]

for our purposes here \( CF = 1 \), don't need to calculate it.
meaning: I think that \( \mu \) is around 
\[ \bar{y} \pm 0.69 \] (\( \bar{y} \))

(For the purpose of settling with somebody about where \( \mu \) is on the number line)

\[ SE \approx 0.69 \]

\[ y \]

\[ 95\% \]

\[ -2 \]

\[ 0 \]

\[ 2 \]

\[ \bar{y} \pm 2SE(\bar{y}) \]

\[ 28.09 \pm 2(0.69) = 28.09 \pm 1.38 \]

I am 95\% confident that \( \mu \) is in this interval confidence interval for \( \mu \)

\[ 26.71 \quad 28.09 \quad 29.47 \]
Why 95% confidence? A: it's just conventional. For less accuracy we sometimes just give away with lower degree of confidence (ex. \( \bar{x} \pm 1.5SE \rightarrow 68\% CI \)), at other times need more accuracy (ex. \( \bar{x} \pm 2SE \rightarrow 99.7\% CI \)).

What is meaning of a 95% CI?

Ex. Is it true that \( P(26.71 \leq \mu \leq 29.47) \) = 95%? A: unfortunately, no; \( \mu \) is either between $26.71 and $29.47 or it's not.

Q2: If that's not what CI mean in frequency story for pub, what do they mean?
is the method, not in the method itself.

the time: our confidence is
your method delivers hits 95% of
the time. if you're method delivers
 hits 95%, then

we will never know
another hit is

95% confident
be hit. 95%

95%