This Time: Inference on Proportions (Percentages)

Next Time: Significance Testing

IID vs. SRS

\[ \text{Sample} \]

\[ \text{Population} \]

\[ \begin{bmatrix} \sigma \end{bmatrix} \]

\[ \begin{bmatrix} s \end{bmatrix} \]

\[ \text{Mean} \hat{\mu} \]

\[ \text{SD} \hat{\sigma} \]

\[ E_{\text{IID}}(\hat{\mu}) = \mu \]

\[ SE_{\text{IID}}(\hat{\mu}) = \frac{s}{\sqrt{n}} \]

\[ SE_{\text{IID}}(\hat{\mu}) = \frac{s}{\sqrt{n}} \]

\[ \text{Est.} \]

\[ E_{\text{SRS}}(\hat{\mu}) = \mu \checkmark \]

\[ SE_{\text{SRS}}(\hat{\mu}) = \frac{s}{\sqrt{n}} \]

\[ \leq SE_{\text{IID}}(\hat{\mu}) \]

\[ SE_{\text{SRS}}(\hat{\mu}) = \sqrt{\frac{n-1}{n}} \cdot SE_{\text{IID}}(\hat{\mu}) \]

Correction Factor (CF)
CF MAKES SENSE?

Ex. \( h = 1 \) \( CF = 1 \) \( SE_{rs} = IID \) ✓

\( h = N \) \( CF = 0 \) \( SE_{rs} (\bar{y}) = 0 \) ✓

\( h \ll N \) \( CF = 1 \) \( SE_{rs} = IID \) ✓

Helps with #2 on HNK 4; This is only time all QTr. That CF is relevant.

CS II 1.67

DEVIL'S ADVOCATE (DA): I think REAL 70% is much higher (ex. 90%), but the only reason NAEP got 65% in sample was unlucky sampling. Q: DA's position possible?

A: Yes. A: DA's position plausible?
A: Let's see.
### Inferential Summary

<table>
<thead>
<tr>
<th>Quantity of Interest</th>
<th>$p = (pop.) %$ of all 4th-6th graders who can add $\frac{1}{3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>$\hat{p} = 65% = 0.65$</td>
</tr>
<tr>
<td>Give or Take</td>
<td>$SE_{\text{IID}}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})} = 1.5%$</td>
</tr>
<tr>
<td>95% Confidence Interval for $p$</td>
<td>$\hat{p} \pm 2SE(\hat{p}) = 65% \pm 3%$</td>
</tr>
</tbody>
</table>
\[ \text{E} \bar{p} = E(\bar{p}) = E_{\text{IID}}(\bar{p}) = E_{\text{IID}}(\hat{p}) = \mu = p \]

\[ \begin{align*}
\text{SE of } \hat{p} &= \text{SE}(\hat{p}) = \text{SE}_{\text{IID}}(\hat{p}) = \text{SE}_{\text{IID}}(\hat{p}) \\
&= \frac{\sigma}{\sqrt{n}}; \text{ MATH FACT: with } A \sim B \text{ pop with } p \approx 0.5 \text{ & } (1-p) \approx 0.5, \\
&= \sqrt{p(1-p)} = \sqrt{\frac{\text{frac. of 1's}}{\text{frac. of 0's}}} \\
&= \text{SE}_{\text{IID}}(\hat{p}) = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}} \\
\text{SO NEW FORMULA: } & \text{SE}_{\text{IID}}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
\text{EST. VERSION: } & \text{SE}_{\text{IID}}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
&= \sqrt{\frac{0.65(0.35)}{950}} = 0.015 = 1.5% \]
95% CI for $\mu$: $\bar{x} \pm 2 \hat{\sigma}(\bar{x})$
so 95% CI for $p$: $\hat{p} \pm 2 \sqrt{\hat{p}(1-\hat{p})}$
$= 65\% \pm 2(1.5\%) = 65\% \pm 3\%$
$= (62\%, 68\%)$

$\leftarrow$ 95% CI $\rightarrow$ DA
down
(1)
62%, 65%, 68%
90%

Since DA's 90% is nowhere near our 95% CI, we strongly disagree with DA.

(NOTE)
If they instead wanted $\hat{p}$ to differ from $p$ by no more than 2%, they would have needed more data (this is a sample size calculation):
\[ 2SE = 2\% = 0.02 \]

\[ \sqrt{0.65(0.35)} = 0.01 \rightarrow \frac{0.65(0.35)}{(0.01)^2} \]

\[ \rightarrow (0.65)(0.35) = (0.01)^2 \cdot n \]

\[ \rightarrow n = \frac{0.65(0.35)}{(0.01)^2} = \frac{0.65(0.35)}{(0.01)(0.01)} \]

\[ = \frac{65}{35} = 22.75 \]