Today: Inference on Proportions (Percentages)  Read FPP ch 26 & 29

Wednesday: Significance Testing

HW 4 p. 95-97 Due Mon May 22 or [Wed May 24] (Will decide soon)

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\bar{y}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$s$</td>
</tr>
</tbody>
</table>

**I.I.D** $\Rightarrow$ $\text{SE}_{\text{I.I.D}}(\bar{y}) = \frac{s}{\sqrt{n}}$

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\bar{y}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$s$</td>
</tr>
</tbody>
</table>

**SRS**

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\bar{y}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$s$</td>
</tr>
</tbody>
</table>

**SRS** $\Rightarrow$ Correction Factor $\left(\frac{N-n}{N-1}\right)^{\frac{1}{2}}$ $\text{SE}_{\text{SRS}}(\bar{y}) = \frac{s}{\sqrt{n}} \cdot \left(\frac{N-n}{N-1}\right)^{\frac{1}{2}}$

**CF makes sense?**

- If $n \ll N$, $\text{CF} \approx 1$, $\text{SRS} \approx \text{I.I.D}$ [OK!]
- If $n < N$, $\text{CF} \approx 1$, $\text{SRS} \approx \text{I.I.D}$ [OK!]

Helps w/ #1 on HW4; This is the only time all quarter that CF is relevant.

---

**Case Study 11 p. 98**

Devil's Advocate (DA): I think real population 90% is much higher (i.e. 90%+) & the only reason NAED got 68% in sample was unlucky sampling.

Q) DA's position possible?
A) Yes.

Q) DA's position plausible?
A) Let's see
Inferential Summary

Quantity of Interest: \( \hat{p} = \frac{\text{pop}}{\text{total}} \) of all US 17 yr olds who can add

Estimate: \( \hat{p} = 0.55 \) or 0.65

Give or Take: \( \text{SE}_{\text{IID}}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \) = 1.5%

95% Confidence Interval for \( \hat{p} \):

\[ \hat{p} \pm 2 \times \text{SE}(\hat{p}) \]

For long run SD w/ a 0/1 pop, w/ \( p \% \) 1's a (1-p)% 0's,

\[ \sigma = \sqrt{p(1-p)} = \frac{1}{\sqrt{n}} \text{ (Fraction of 1's)} \] (Fraction of 0's)

So the new formula is:

\[ \text{SE}_{\text{IID}}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} \]

So here \( \text{SE}_{\text{IID}}(\hat{p}) = \frac{0.55(0.45)}{\sqrt{950}} = \frac{0.445}{950} = 0.046 < 1.5\% \)

95% CI for \( \mu \) = \( \bar{y} \pm 2 \times \text{SE}(\bar{y}) \). So 95% CI for \( \hat{p} \) = \( \hat{p} \pm 2 \times \text{SE}(\hat{p}) \)

= 0.65 \pm 2(0.03) = 0.65 \pm 0.06 = (0.59, 0.71)

Since DA's 90% is nowhere near our 95% CI, we strongly disagree w/ DA
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):

$\pm 2\% = 0.02 = 2\sqrt{\frac{(0.65)(0.35)}{n}}$

$0.02 = \frac{2}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{0.02}{0.01} \Rightarrow n = \left(\frac{0.02}{0.01}\right)^2 = \left(0.02\right)^2$

$\Rightarrow n = \frac{(0.65)(0.35)}{(0.01)^2} = \frac{(0.65)(0.35)}{(0.01)(0.01)} = 2275$