Case Study 15 continued

- 2 independent samples
  - crucial question: Is there a linkage between the two samples?
  - No!
  - Run models in parallel

Tribe 1

**Population**
- all tribe 1 adult females at relevant time
- \( N_1 = \text{Big} \)
- \( \mu_1 = ? \)
- \( \sigma_1 = ? \)

**Sample**
- size \( n_1 = 25 \)
- \( \bar{y}_1 = 59.4'' \)
- \( s_1 = 1.8'' \)

**Inferential Summary**
- \( M_1 \) = mean adult height of all adult females in tribe 1
- \( \bar{y}_1 = 59.4'' \)
- \( SE_{\bar{y}_1} = 0.36'' \)

Tribe 2

- \( M_2 = ? \)
- \( \sigma_2 = ? \)

**Population**
- all tribe 2 adult females of relevant time
- \( N_2 = \text{Big} \)

**Sample**
- size \( n_2 = 27 \)
- \( \bar{y}_2 = 61.3'' \)
- \( s_2 = 2.4'' \)

**Inferential Summary**
- \( M_2 \) = mean adult height of all adult females in tribe 2
- \( \bar{y}_2 = 61.3'' \)

**Possible \( y_2 \) with long run histogram norm CLT**
Inferential Summary

**Known quantities:**
- $M_2$ = mean height of all adult females in Tribe 2
- $\bar{y}_2 = 61.3''$
- $SE(\bar{y}_2) = 0.46''$
- $\sigma$ = unknown

**Best estimate:**
- $\mu_2 - M_1 = \text{mean difference between adult females in Tribe 2 vs. Tribe 1}$
- $\bar{y}_2 - \bar{y}_1 = (61.3'' - 59.4'') = 1.9''$

**Is this diff. pratical?**
- Yes! Most of the Tribe 2 women would be taller than most Tribe 1 women.

**Give or Take:**
*Need new formula for this

- $95\%$ CI.

**Calculate:**
- $SE(\bar{y}_2 - \bar{y}_1) = \sqrt{SE(\bar{y}_1)^2 + SE(\bar{y}_2)^2} = \sqrt{0.36''^2 + 0.46''^2} = 0.58''$

- See next page for calculations

- $(\bar{y}_2 - \bar{y}_1) \pm 2 \times SE(\bar{y}_2 - \bar{y}_1)$

- $= 1.9'' \pm 2 \times 0.58'' = 1.9'' \pm 1.16''$

- $0.7'' \quad 1.4'' \quad 3.1''$
Math Fact

Uncertainty combines with 2 independent samples like the legs of a right triangle

\[ a^2 + b^2 = c^2 \]

\[ \sqrt{a^2 + b^2} = c \]

\[ SE(\bar{y}_2 - \bar{y}_1) = \sqrt{(SE(\bar{y}_1))^2 + (SE(\bar{y}_2))^2} \]

For 2 independent samples:

\[ SE(\bar{y}_1 - \bar{y}_2) = \sqrt{(SE(\bar{y}_1))^2 + (SE(\bar{y}_2))^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]

Long run histogram of \((\bar{y}_2 - \bar{y}_1)\)

\[ SE = 0.58 \]

95% CI:

\( (0.7, 3.1) \)

Null Hypothesis: \( M_2 - M_1 = 0 \) no real difference between tribes

But: 0 is not in the C.I. for \((\bar{y}_2 - \bar{y}_1)\) so the null looks wrong.

Size biased (length biased) sampling: big things are easier to find than small things.

her estimate of \( \bar{y}_1 \) of \( M_1 \) is likely to have been biased on the high side and ditto for \( \bar{y}_2 \) as an estimate of \( M_2 \).
However this bias should (largely) cancel itself when looking at the difference \((y_2 - y_1)\) as an estimate of \((\mu_2 - \mu_1)\), so her method seems ok in this case.

* Biased measurement methods are bad for single samples, but can be ok for comparing 2 samples, given that the same bias was present for each sample.

### Case Study 16

2 independent samples with 1's & 0's (yes or no)

- "black" \(n_1 = 100\)
- "white" \(n_2 = 900\)

#### Black

<table>
<thead>
<tr>
<th></th>
<th>(P = \frac{y}{n})</th>
</tr>
</thead>
<tbody>
<tr>
<td>pop</td>
<td>all black people in US in 1977</td>
</tr>
<tr>
<td></td>
<td>observed black people</td>
</tr>
<tr>
<td></td>
<td>favor?</td>
</tr>
<tr>
<td>(N)</td>
<td>big [i]</td>
</tr>
<tr>
<td>(Y)</td>
<td>[100]</td>
</tr>
<tr>
<td>(P_1)</td>
<td>(\hat{P}_1 = 27%)</td>
</tr>
<tr>
<td>(SE(P))</td>
<td>(\sqrt{\frac{P(1-P)}{n}})</td>
</tr>
</tbody>
</table>

#### White

<table>
<thead>
<tr>
<th></th>
<th>(P = \frac{y}{n})</th>
</tr>
</thead>
<tbody>
<tr>
<td>pop</td>
<td>all white people in US in 1977</td>
</tr>
<tr>
<td></td>
<td>observed white people</td>
</tr>
<tr>
<td></td>
<td>favor?</td>
</tr>
<tr>
<td>(N)</td>
<td>big [i]</td>
</tr>
<tr>
<td>(Y)</td>
<td>[900]</td>
</tr>
<tr>
<td>(P_2)</td>
<td>(\hat{P}_2 = 38%)</td>
</tr>
<tr>
<td>(SE(P))</td>
<td>(\sqrt{\frac{P(1-P)}{n}})</td>
</tr>
</tbody>
</table>

#### Summary

The quantity of interest is \((P_1 - P_2)\), the difference in proportion between "B" and "W" on opinion on preferential treatment.
Practically significant? 
Give or take:

\[
(\hat{p}_1 - \hat{p}_2) = 27\% - 8\% = 19\%
\]

Yes: \[
\frac{27}{8} = 3.4\text{ times greater }\% \text{ for black than white.}
\]

\[
SE_{\text{diff}}(\hat{p}_1 - \hat{p}_2) = 4.5^\circ.
\]

Remember:

\[
SE_{\text{diff}}(\hat{p}_1 - \hat{p}_2) = SE(\hat{p}_2 - \hat{p}_1)
\]

\[
= \sqrt{\left(\frac{SE(\hat{p}_1)^2}{n_1}\right) + \left(\frac{SE(\hat{p}_2)^2}{n_2}\right)}
\]

\[
SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]

So:

\[
SE(\hat{p}_1 - \hat{p}_2) = \sqrt{(4.4\%)^2 + (0.9\%)^2} = 4.5^\circ.
\]

95\% CI:

\[
(\hat{p}_1 + \hat{p}_2) \pm 2SE(\hat{p}_1 - \hat{p}_2)
\]

Null: \( \hat{p}_1 - \hat{p}_2 = 0 \)

19°6 ± 7°

(10°, 19°, 28°6)

\( \theta \) is nowhere near C.I. so thus result is **Stat. Sig.**