Case Study 15:

tribe 1

tribe 2

\[
\begin{align*}
&\text{height} \quad \text{linked?} \\
&n_1 = 25 \\
&n_2 = 27 \\
&\text{mean } \bar{y}_1 = 59.4 \text{ in} \\
&\text{SD } s_1 = 1.8 \text{ in} \\
&\text{mean } \bar{y}_2 = 61.3 \text{ in} \\
&\text{SD } s_2 = 2.4 \text{ in}
\end{align*}
\]

2 independent samples - no connection between the two

Tribes:

population

all tribe 1 adult females at relevant time

sample

observed skeletons

1. O.S.

\[ \text{possible } \bar{y}_i \text{'s} \]

\[
\begin{align*}
&\text{height} \quad \text{like SRS} \quad \text{like IID} \\
&n_1 = 25 \\
&\text{mean } \mu_i = ? \\
&\text{SD } \sigma_i = ? \\
&\text{pop. hist.} \\
&\text{hypothetical mean } \bar{y}_i = 59.4 \text{ in} \\
&\text{SD } s_i = 1.8 \text{ in} \\
&\text{long run mean?} \\
&\text{long run SD?} \\
&\text{long run histogram?}
\end{align*}
\]

\[
\begin{align*}
&\text{mean } \bar{y}_i = ? \text{(ex: 59.9 in)} \\
&\text{SE}_{\bar{y}_i} = \frac{\sigma_i}{\sqrt{n_i}} = \frac{1.8}{\sqrt{25}} = 0.36 \text{ in} \\
&\text{SE} = 0.36 \\
&\text{normal curve (CLT)}
\end{align*}
\]
Inferential Summary for Tribe 1

unknown quantity of interest
\[ \mu_1 = \text{mean height of all adult females in tribe 1} \]

estimate
\[ \bar{y}_1 = 59.4 \text{ in} \]

give or take
\[ \hat{SE}_{11D}(\bar{y}_1) = 0.36 \text{ in} \]

Tribe 2

population
all tribe 2 adult females at relative time

sample
observed skeletons

possible \( y_2 \)'s

Inferential Summary for Tribe 2

unknown quantity of interest
\[ \mu_2 = \text{(same as tribe 1)} \]

estimate
\[ \bar{y}_2 = 61.3 \text{ in} \]

give or take
\[ \hat{SE}_{11D}(\bar{y}_2) = \frac{\sigma_2}{\sqrt{n_2}} = \frac{5}{\sqrt{27}} = 2.4 \text{ in} = 0.46 \text{ in} \]
Real Inferential Summary

<table>
<thead>
<tr>
<th>Quantity of Interest</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_2 - \mu_1 )</td>
<td>( \bar{y}_2 - \bar{y}_1 ) = (61.3 in - 59.4 in) = 1.9 in</td>
</tr>
</tbody>
</table>

Is this difference pract. sig?

- Yes
- Most Tribe 2 women would be taller than most Tribe 1 women

Give or take for estimate

SE (\( \bar{y}_2 - \bar{y}_1 \)) = 0.6 in

95% CL for \( (\mu_2 - \mu_1) \)

\( (\bar{y}_2 - \bar{y}_1) \pm 2SE(\bar{y}_2 - \bar{y}_1) = 1.9 \pm 2(0.6) = 1.9 \pm 1.2 = (0.7, 3.1) \)

Math fact: Uncertainty combines w/ 2 independent samples like the legs of a right triangle

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

\[
(\text{SE}(\bar{y}_1))^2 = \frac{S_1^2}{n_1}, \quad (\text{SE}(\bar{y}_2))^2 = \frac{S_2^2}{n_2}
\]

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

null: no real difference = \( (\mu_2 - \mu_1) = 0 \)

\( \text{SE} = 0.6 \)

95% CI for \( (\mu_2 - \mu_1) \)

- Bottom: 0.7 in
- Top: 3.1 in

\( (\bar{y}_2 - \bar{y}_1) - 2\text{SE} \)

\( (\mu_2 - \mu_1) + 2\text{SE} \)

0 is not in 95% CI, so null looks wrong \( \rightarrow \) stat sig
Is this data gathering method like SRS or might it instead be biased?

- **Size-bias/length-bias sampling**: big things are easier to find than small things.
- Her estimate of $\bar{y}_1$, of $\mu_1$, is likely to have been biased on high side, and ditto for $\bar{y}_2$ as estimate of $\mu_2$.
- But this bias should largely cancel in looking at the difference ($\bar{y}_2 - \bar{y}_1$) as an estimate of $(\mu_2 - \mu_1)$.
- So this is okay.

(Bad measurements methods are bad for one sample at a time, but are okay for comparing 2 samples)

(End of case study 15)

---

**Case Study 16 p. 102**

<table>
<thead>
<tr>
<th>&quot;black&quot;</th>
<th>&quot;White&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1's+]</td>
<td>[1's+]</td>
</tr>
<tr>
<td>[0's]</td>
<td>[0's]</td>
</tr>
<tr>
<td>$n_1 = 100$</td>
<td>$n_2 = 900$</td>
</tr>
</tbody>
</table>

2 independent samples w/ 1's + 0's

---

**Black population**

- All black people in US in 1977
- 1=yes, 0=no

**Sample**

- Observed black people
- Favor?

**IDS possible $\hat{p}$'s**

---

**Pop Hist**

- N: big
- Mean $\hat{p}_1 = ?$

---

**Long run mean:**

- $E_{1,n_1}(\hat{p}) = \hat{p}_1$
- Long run SD:

$$s_{\hat{p}_1} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}$$

- Estimated SD:

$$s_{\hat{p}_1} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}$$

- $s_{\hat{p}_1}(\hat{p}_1) = \hat{p}_1(1-\hat{p}_1)$
- $\sqrt{(0.27)(0.73)} = 4.4^{\%}$
long run histogram:

\[ \hat{SE} = 4.4\% \]

normal curve (CLT)

---

White population

all whites in 1977

favor?

1 ± 5

N_2 = [1's + 0's]

like SRS

N_2 = 900

mean \( \hat{p}_2 = ? \)

hypothesized

\[ \frac{15^*}{0's} \]

mean \( \hat{p}_2 = 8\% \)

IDS possible \( \hat{p}_2 \) 's

---

sample

observed white people

favor?

1 ± 5

n_2 = 900

[8\%]

[7\%]

long run mean:

\[ E_{100} (\hat{p}_2) = \mu_2 \]

mean \( \hat{p}_2 = ? \) (ex: 7\%)

long run SD:

\[ SE_{100} (\hat{p}_2) = \sqrt{\hat{p}_2 (1-\hat{p}_2) / n_2} \]

\[ \sqrt{(0.05)(0.95)} = 0.9\% \]

\[ \sqrt{900} \]

long run histogram:

\[ \hat{SE} = 0.9\% \]

normal curve (CLT)

---

Real Inferential Summary

<table>
<thead>
<tr>
<th>Quantity of Interest</th>
<th>( (p_1 - p_2) ) = difference in proportions between blacks + whites on preferential treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>( \hat{p}_1 - \hat{p}_2 ) = 27% - 8% = 19%</td>
</tr>
<tr>
<td>Pract Sig?</td>
<td>yes: ( \frac{27}{5} = 3.4 ) times greater % for blacks than whites</td>
</tr>
<tr>
<td>Give or Take</td>
<td>( SE(\hat{p}_1 - \hat{p}_2) = 4.5% )</td>
</tr>
<tr>
<td>95% CI for</td>
<td>( (\hat{p}_1 - \hat{p}_2) \approx 2SE(\hat{p}_1 - \hat{p}_2) = 19% \pm 2(4.5%) = 19% \pm 9% = (10%, 28%) )</td>
</tr>
</tbody>
</table>
\[
\text{SE}_{1, P} (\hat{p}_1 - \hat{p}_2) = \sqrt{\left(\frac{\text{SE}(\hat{p}_1)}{n_1}\right)^2 + \left(\frac{\text{SE}(\hat{p}_2)}{n_2}\right)^2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \approx 4.5\% \\
\frac{0.9\%}{\text{Estimated}} \quad \frac{4.1\%}{\text{Actual}}
\]