Today: Comparing 2 samples

Friday: Comparing 2 sample

HW 5 p. 107-109 Due Fri June 2

Case Study 13½

Null Hypothesis \( \mu = 70 \) (Recalibration not needed)

Alt. Hypothesis \( \mu \neq 70 \) (Recalibration needed)

T-test

Histogram of \( \bar{y} - \mu \) if null is true

Here

\[
p = \frac{S}{\sqrt{n}}
\]

\[
t = \frac{y - \mu}{S/\sqrt{n}} = 2.16
\]

Evidence: Running against null but not very strongly - Borderline Recalibrate

\[
\begin{bmatrix}
18 \\
83 \\
68 \\
84 \\
42
\end{bmatrix}
\]

\( n = 5 \)

Mean \( \bar{y} = 77.8 \)

SD \( s = 8.1 \)

Why are there \( (n-1) \) degrees of freedom w/ a sample size of \( n \)?

Whole Premise: Use \( \bar{y} \) to estimate \( \mu \); This is like holding \( \bar{y} \) constant.

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5
\end{bmatrix}
\]

\( n = 5 \)

\( n \) \( S \) samples out of only \( n-1 \): \( 4 \)

Free to vary if the mean is fixed at 77.8

A Pitfall of Significance Testing:

(Statistical) \( \neq \) (Practical) i.e. New blood to combat high blood pressure

(After-Before)

Null Hyp: Drug Doesn't Work \( \mu = 0 \)

Alt Hyp: \( \mu \neq 0 \)

\( s = 20 \text{ mm}/\sqrt{8000} \)

\[
z = \frac{\bar{y} - \mu}{s} = 4.5
\]

Long Run Hist of \( \bar{y} \)

Here \( p \) is far less than 0.05

5% (1%) So the difference is highly statistically significant if null is true.
But a decline of only 1 mm on average is too small to be practically significant. So here (StatSig: \text{Yes}, \text{No})

How did this happen? Too much data

- StatSig b/c p is small \( \rightarrow \) p is small because z is big \( \rightarrow \)
  \[
  z = \frac{(-10 \text{ mm}) - (0 \text{ mm})}{1 \text{ mm}} = -10
  \]
  Fraction is big when numerator is big (not true here), denominator is small (true here) \( \rightarrow \) Denominator = SE small = \( \frac{s}{\sqrt{n}} \) Fraction small if numerator is small (not true)

  \[
  \frac{s}{\sqrt{n}} = \frac{20 \text{ mm}}{\sqrt{5000}} \Rightarrow \text{denominator is big (which it is)}
  \]

BAD EXPERIMENTAL DESIGN: Too much data

<table>
<thead>
<tr>
<th>After</th>
<th>Before</th>
</tr>
</thead>
<tbody>
<tr>
<td>-25</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>( n = 8 )</td>
<td>( n = 8 )</td>
</tr>
<tr>
<td>Mean ( \bar{y} ) = 0 \text{ mm}</td>
<td>Mean ( \bar{y} ) = 0 \text{ mm}</td>
</tr>
<tr>
<td>( s = 20 )</td>
<td>( s = 20 )</td>
</tr>
</tbody>
</table>

\[
SE(\bar{y}) = \frac{s}{\sqrt{n}} = \frac{20 \text{ mm}}{\sqrt{8}} \approx 7 \text{ mm}
\]

\[
2 = \frac{(-10 \text{ mm}) - (0 \text{ mm})}{1 \text{ mm}} = -1.4
\]

Long Run Hist of \( \bar{y} \) if null is true

95% CI: -10.14

The difference is not Stat-Sig (\( p = 10\% > 5\% \)) but it is Practically Significant \( (\text{StatSig: No, PracSig: Yes}) \)

Problem: Too Little Data

Fix: Design experiment with right n so that StatSig \( \approx \) PracSig
VII. COMPARING 2 SAMPLES

2 Topics:
1) Paired Comparisons
2) Analysis of 2 independent samples.

Paired Comparisons: Outcome = Sales volume (in cases) of product.

Treatment: Discount (T) vs. Standard (C) Marketing Plans.

One possible design: Get a bunch of stores (i.e., n = 120),
Randomize design: Go to T, go to C.

Design 1

<table>
<thead>
<tr>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>n₁ = 60</td>
<td>n₂ = 60</td>
</tr>
</tbody>
</table>

Mean y₁, Mean y₂

Estimate of treatment effect = (\bar{y}_1 - \bar{y}_2)

Better design possible?

PCFs? Don't have to worry about PCFs in Design 1 as far as validity of experiment is concerned, because randomization will make T & C groups similar on PCFs.

Can we increase the efficiency (accuracy) of experiment?
Yes, think about PCFs at design time.

Main PCF: Overall sales volume in all products

How can we improve design by using this PCF?

PCF | T | C |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>($87, 916)</td>
<td>[857, 991]</td>
</tr>
<tr>
<td>Low</td>
<td>($787, 649)</td>
<td>[787, 699]</td>
</tr>
<tr>
<td>Med</td>
<td>($814, 812)</td>
<td>[814, 812]</td>
</tr>
</tbody>
</table>

Basic Idea: Hold PCF constant in creating matched pairs of stores.

Assign T & C in pairs at Random.

Design 2 is also valid, but will be more efficient because PCF was controlled in design.

Paired Comparisons: \[Design 2\]

Analysis of 2 Independent Samples: \[Design 1\]