STA 248 - March 15/05

Assignment 3, Additional Problem #5

6 = 10' in both groups

Use normal distribution.

Jeff office hours Mon 21 2-4 SS 2114

Rule-of-thumb for whether or not 2 variances could be equal.

Most common

large sd. \( < 2 \)

small sd.

Can be worse if sample sizes are small.

\( t \)-test with pooled variance is more robust against departures from equal variance if sample sizes in 2 groups are close to equal.
Example:
Is brain size associated with schizophrenia?
To control for other factors, study looked at 15 pairs of identical twins. One diagnosed with schizophrenia, other was not.

Measured brain size using MRI.
Data: left hippocampus in cm³

Paired comparisons

Want to compare \( \mu_{Y_1} \) to \( \mu_{Y_2} \)
where each value of \( Y_1 \) and \( Y_2 \)
is measured on same "individual".

Let \( d = Y_1 - Y_2 \)

Estimate of \( \mu_{Y_1} - \mu_{Y_2} \) is \( \bar{d} \)

Carry out tests (CI on differences using 1-sample methods)
A test: $H_0: \mu_{\text{schiz}} = \mu_{\text{unaffected}}$

Equiv. to $\mu_1 = 0$

$H_a: \mu_1 > 0$

Test statistic:

$$t_{obs} = \frac{\bar{d} - \mu_0}{sd_{\bar{d}}}$$

$t_{obs}$ is the s.d. of the differences

If differences are normally distributed and $H_0$ is true, $t_{obs}$ is an observation from a t-distribution with $n - 1$ df.

Schizophrenia data:

$t = 1.991$

$p$-value = .056

Strong evidence of a difference in mean brain size.
What if data don't follow a Normal distribution?

Example: Stereograms

Data: time to see clear:
2 groups - one got verbal information
- other got verbal + visual information
43 people  $\rightarrow$ 35 people.

Is the mean time different?

Log transformation
- useful for making right-skewed distributions close to normal.
Also square root (not as strong as log)
Compare means of the log of data using t-tests

CI for mean of logged data:
lower \leq \log \mu_x - \log \mu_y \leq upper

Back-transform:
\exp(\text{lower}) \leq \frac{x}{y} \leq \exp(\text{upper})

Bootstrap tests - 2-sample
- No assumptions necessary about the distribution of the data
- Can test any statistic (mean, median, etc)

Two samples of size n₁ and n₂
X₁, ..., Xₙ₁
y₁, ..., yₙ₂
Suppose we want to test:

\[ H_0: \mu_Y - \mu_X = 0 \]

vs. \[ H_a: \mu_Y - \mu_X > 0 \]

Test statistic: \[ V = \overline{Y} - \overline{X} \]

Observed: \[ V = \overline{y} - \overline{x} \]

\[ p\text{-value:} \]

\[ p = P(V \geq \overline{y} - \overline{x} \mid H_0 \text{ is true}) \]

Want bootstrap estimate of \( p \)-value

Must generate bootstrap samples under assumption that \( H_0 \) is true.

So combine 2 samples into one of size \( n_1 + n_2 \)

Then re-sample with replacement into 2 new groups: (do B times)

- \( j \)th bootstrap sample: \( 1 \) of size \( n_1 \): \( x^*_1, \ldots, x^*_n \); \( \overline{x}^*_j \)
- \( j \)th bootstrap sample: \( 1 \) of size \( n_2 \): \( y^*_1, \ldots, y^*_n \); \( \overline{y}^*_j \)

(\( j = 1, \ldots, B \))

and calculate

\[ \overline{v}^*_j = \overline{y}^*_j - \overline{x}^*_j \]
The bootstrap estimated p-value is
\[ p^* = \frac{\#_{j=1}^{b} \{ v_j^* \geq y-x \}}{B} \]