Chapter 9: Comparing two paired populations

Duzhe Wang

Part 2: the sign test
https://dzwang91.github.io/stat371/

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Assumptions of the paired T-test

- All of the *pairs* are indepdendent.
- The differences follow a normal distribution.

Assumptions of the paired T-test

- All of the *pairs* are indepdendent.
- The differences follow a normal distribution.

- What if the differences don't follow a normal distribution?

Assumptions of the paired T-test

- All of the *pairs* are indepdendent.
- The differences follow a normal distribution.

- What if the differences don't follow a normal distribution?
- Let's consider a specific case that the distribution of differences is symmetric.

Setup

The data consists of paired observations. Let:

- $X_{1,i}$ = i-th data point from population 1
- μ_1 = true mean of population 1
- $X_{2,i}$ = i-th data point from population 2
- μ_2 = true mean of population 2
- $D_i = X_{1,i} - X_{2,i}$ = the difference for pair i
- n = number of pairs
- σ_D^2 = true variance of the differences

We wish to test: $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_A : \mu_1 - \mu_2 \neq 0$.

Setup

The data consists of paired observations. Let:

- $X_{1,i}$ = i-th data point from population 1
- μ_1 = true mean of population 1
- $X_{2,i}$ = i-th data point from population 2
- μ_2 = true mean of population 2
- $D_i = X_{1,i} - X_{2,i}$ = the difference for pair i
- n = number of pairs
- σ_D^2 = true variance of the differences

We wish to test: $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_A : \mu_1 - \mu_2 \neq 0$.
$\iff$ $H_0 : \mu_D = 0$ vs. $H_A : \mu_D \neq 0$ where $\mu_D = \mu_1 - \mu_2$.

Setup

The data consists of paired observations. Let:

- $X_{1,i}$ = i-th data point from population 1
- μ_1 = true mean of population 1
- $X_{2,i}$ = i-th data point from population 2
- μ_2 = true mean of population 2
- $D_i = X_{1,i} - X_{2,i}$ = the difference for pair i
- n = number of pairs
- σ_D^2 = true variance of the differences

We wish to test: $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_A : \mu_1 - \mu_2 \neq 0$.
$\iff$ $H_0 : \mu_D = 0$ vs. $H_A : \mu_D \neq 0$ where $\mu_D = \mu_1 - \mu_2$.
$\iff$ $H_0 : \mathrm{m}_D = 0$ vs. $H_A : m_D \neq 0$ where m_D is the population median of the differences.

Sign test in two paired samples

- **Assumption.** $D_1, \ldots, D_n$ are an i.i.d. sample.
- **Hypothesis.** Let m be the population *median*. We wish to test $H_0 : m = 0$.

Sign test in two paired samples

- **Assumption.** $D_1, \ldots, D_n$ are an i.i.d. sample.
- **Hypothesis.** Let m be the population *median*. We wish to test $H_0 : m = 0$.
- **Test statistic.** $B = \#$ of data values greater than 0 (Ignore values tied with 0.) Note that if H_0 is true, $B \sim \text{Binomial}(n^*, 0.5)$, where n^* is the number of data points not equal to 0.

Sign test in two paired samples

- **Assumption.** $D_1, \ldots, D_n$ are an i.i.d. sample.
- **Hypothesis.** Let m be the population *median*. We wish to test $H_0 : m = 0$.
- **Test statistic.** $B = \#$ of data values greater than 0 (Ignore values tied with 0.) Note that if H_0 is true, $B \sim \text{Binomial}(n^*, 0.5)$, where n^* is the number of data points not equal to 0.
- **P-value**. Let b be the observed number of data points greater than 0. If:
  - $H_A : m > 0$:
    $P(B \geq b) = P(B = b) + P(B = b + 1) + \ldots + P(B = n^*)$.
  - $H_A : m < 0$:
    $P(B \leq b) = P(B = b) + P(B = b - 1) + \ldots + P(B = 1) + P(B = 0)$.
  - $H_A : m \neq 0$: $2 \min\{P(B \geq b), P(B \leq b)\}$.