Unpaired T test

Miaoyan Wang

Department of Statistics
UW Madison

Example: Dog bark distance

Table: Bark distance (m)

urban	rural
29	40
10	47
15	38
41	59
18	45
18	52
12	57
45	50
34	50
30	49
22	50
26	43
18	

Notation

- Y_{1i}: Random variable of the ith response in the first sample for $i = 1, \ldots, n_1$.
- Y_{2i}: Random variable of the ith response in the second sample for $i = 1, \ldots, n_2$.
- $\mu_1 = E(Y_{1i})$: Population mean response of the first group.
- $\mu_2 = E(Y_{2i})$: Population mean response of the second group.
- Our goal is to test

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_A : \mu_1 \neq \mu_2.$$

Assumptions

#1 The first sample $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ is an i.i.d. sample of size n_1 from $N(\mu_1, \sigma_1^2)$.

#2 The second sample $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ is an i.i.d. sample of size n_2 from $N(\mu_2, \sigma_2^2)$.

#3 The two samples $\{Y_{1i}\}$ and $\{Y_{2i}\}$ are independent.

#4 The (unknown) variances are the same $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Test Statistic

- To test
$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_A : \mu_1 \neq \mu_2.$$
- The main idea is to consider the difference in mean
$$\bar{Y}_1 - \bar{Y}_2,$$
and construct a T-type test statistic
$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - \mathbb{E}_0(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{Var(\bar{Y}_1 - \bar{Y}_2)}}.$$

Test Statistic

Under the null hypothesis:

- What is the distribution of $\bar{Y}_1$?

$$\bar{Y}_1 \sim N(\mu_1, \frac{\sigma^2}{n_1})$$

- What is the distribution of $\bar{Y}_2$?

$$\bar{Y}_2 \sim N(\mu_2, \frac{\sigma^2}{n_2})$$

- What is the expectation of $\bar{Y}_1 - \bar{Y}_2$?

$$\mu_{\bar{Y}_1 - \bar{Y}_2} = E(\bar{Y}_1 - \bar{Y}_2) = E(\bar{Y}_1) - E(\bar{Y}_2) = \mu_1 - \mu_2$$

- What is the variance of $\bar{Y}_1 - \bar{Y}_2$?

$$\sigma^2_{\bar{Y}_1 - \bar{Y}_2} = Var(\bar{Y}_1) + Var(\bar{Y}_2) = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$$

Estimated Variance of $\bar{Y}_1 - \bar{Y}_2$

- If S_p^2 is an estimator of $\sigma^2 = \text{Var}(Y_1) = \text{Var}(Y_2)$, then we can estimate $\sigma^2_{\bar{Y}_1 - \bar{Y}_2}$ by

$$\text{Var}(\bar{Y}_1 - \bar{Y}_2) = S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

- We propose a **pooled variance estimate** of σ^2:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \quad \text{where}$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (Y_{1i} - \bar{Y}_1)^2, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_{2i} - \bar{Y}_2)^2$$

- Interpretation of S_p^2: A weighted average of the two sample variances, weighted by the d.f.'s.

T Test Statistic

- Our goal is to test

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_A : \mu_1 \neq \mu_2$$

- Under the H_0, the test statistic follows t-distribution with df$=n_1 + n_2 - 2$.

$$t = \frac{\bar{Y}_1 - \bar{Y}_2 - 0}{\sqrt{S_P^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim T_{n_1+n_2-2}.$$

Example: Bark Distance

- From the bark distance data, we have $\bar{y}_1 = 48.33, s_1^2 = 38.97, n_1 = 12$ and $\bar{y}_2 = 24.46, s_2^2 = 118.77, n_2 = 13.$
- The pooled sample variance is:

$$s_p^2 = \frac{(12-1) \times 38.97 + (13-1) \times 118.77}{12+13-2} = 80.60$$

- The observed test statistic is:

$$t = \frac{48.33 - 24.46 - 0}{\sqrt{80.60\left(\frac{1}{12} + \frac{1}{13}\right)}} = 6.642$$

- The degrees of freedom are: $\text{df} = 12 + 13 - 2 = 23.$
- The p-value is: The p-value is $2 \times P(T_{23} \geq 6.642)$, which is less than 0.002.
- The conclusion is: Reject H_0 at the 5% level. There is very strong evidence that the mean bark distances for rural and urban prairie dogs are different.