t-value의 여러가지 variation

식(1) 혹은 식(14)로 표현되는 t-value는 실험 상의 가정이나 세팅에 따라 여러가지 변형이 존재하는데 결국은 "pooled standard deviation을 사용하는가?"가 핵심이다.

pooled standard deviation (혹은 pooled variance)라는 것은 두 집단의 표준편차가 같다고 가정하고 하나의 표준편차 값으로 두 집단의 표준 편차를 대체하겠다는 의미를 갖는다.

식 라 식가 차이 없어 보이면, 같다고 가정함.

두 집단의 표준편차(혹은 분산)이 같은 경우에 대해서 (즉, pooled standard deviation을 사용한다고 했을 때), 두 가지의 케이스를 생각해볼 수 있다.

1. 첫번째로 두 표본 집단의 n 수가 $n_1 = n_2 = n$으로 동일하고 두 표본 집단의 분산이 같다고 가정할 수 있는 경우에는 다음과 같이 $s_{\bar{X}_1 - \bar{X}_2}$을 쓸 수 있다.

$$\text{식}(1) \Rightarrow \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{n}}} = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{2}{n}}} \quad \text{where} \quad s_p = \sqrt{\frac{s_1^2 + s_2^2}{2}} \tag{15}$$

즉, 식(15)은 식(14)에서 s_1과 s_2를 s_p로 대체한 것으로 s_p는 pooled standard deviation을 나타내고 있다.

2. 두 번째로 두 표본 집단의 n수는 다르지만 두 표본 집단의 분산은 같다고 가정할 수 있는 경우에는 t-value를 다음과 같이 쓸 수 있다.

$$\text{식}(1) \Rightarrow \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{where} \quad s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{16}$$

식 (15), (16)의 수식은 복잡해 보이지만 사실은 표준 편차값을 pooled 처리 할 것이냐 아니냐에 따라 달린 문제이므로, 실제 연구나 분석에서 적용할 때에는 그 상황에 맞게 적절한 수식을 이용하면 될 것이다.

Two Sample t Test for Independent Samples

Unpooled (assuming distributions have different variances)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Two-Sample Means t Test (unpooled/separate) with df

Pooled (assuming both distributions have same variance)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Two-Sample Means t Test (pooled) with $(n_1 + n_2 - 2)$ df