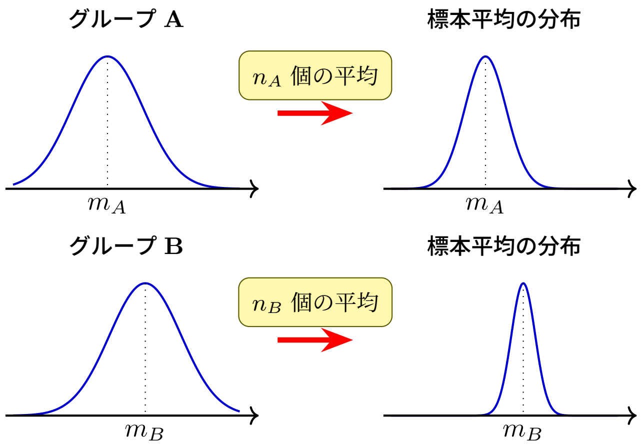
対応のない 2 標本の t 検定

対応のある 2 標本の場合は、差 d_i を 1 標本として t 検定を行ったが、 2 標本の要素に対応の無い場合は差をとることができない。今回は、対応の無い 2 標本で、2 つの母分散が等しい場合の t 検定について考える。すなわち、

$$\sigma_A^2 = \sigma_B^2$$

との仮定のもとで、2 つのグループの情報を統合し、有意な差があるかを検定する。まずは、それぞれの不偏分散から共通する分散 σ^2 を求める方法について考えてみよう。



例 1　数学の小テストの結果について、A 組から 10 人、B 組から 20 人の標本を得たところ以下のようなことが分かった。A 組と B 組の母分散 σ^2 が等しいと仮定して、分散 σ^2 を求めよ。

A 組：$n_A = 10$、不偏分散 $s_A^2 = 7$

B 組：$n_B = 20$、不偏分散 $s_B^2 = 12$

不偏分散の定義は：

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

これを変形すると、

$$\sum(x_i - \bar{x})^2 = (n-1)s^2$$

つまり：

- A 組の偏差平方和 $= (10-1) \times 7 = 63$
- B 組の偏差平方和 $= (20-1) \times 12 = 228$

ここから、母分散 σ^2 の推定値を s_p^2 とすると、

$$\text{分散}：s_p^2 = \frac{63+228}{9+19} = \frac{291}{28} \approx 10.39$$

答　$s_p^2 = 10.39$

観察：B 組の方がサンプルサイズが大きいので、プールした分散 10.39 は B 組の分散 12 に近い値になっている。

このように 2 つの母分散が等しいと仮定し、2 つの異なる分散の値から求めた分散の値を**プールした分散**という。「プールした」(pooled) というのは、2 つのグループの情報を「統合した」という意味である。

2 つの標本 A と B のプールした分散・標準偏差

$$\text{分散}：s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$\text{標準偏差}：s_p = \sqrt{s_p^2}$$

問 1　英語の小テストの結果について、A 組から 5 人、B 組から 8 人の標本を得たところ以下のようなことが分かった。A 組と B 組の母分散 σ^2 が等しいと仮定して、プールした分散 σ^2 を求めよ。

A 組：$n_A = 5$、不偏分散 $s_A^2 = 12$

B 組：$n_B = 8$、不偏分散 $s_B^2 = 10$

答　$s_p^2 =$

母分散 σ^2 の推定値 s_p^2 が求められたので、次は、2 つの標本 A と B の平均の差 $d = \overline{x}_A - \overline{x}_B$ を t 値に変換するための標準誤差 SE を考えよう。

独立な確率変数の和の分散は、

$$V(X-Y) = V(X) + V(Y)$$

となる。そして、母分散 σ^2 と標本の個数 n_A, n_B により、

$$V(\overline{x}_A) = \frac{\sigma^2}{n_A}, \qquad V(\overline{x}_B) = \frac{\sigma^2}{n_B}$$

上記により、

$$\begin{aligned} V(\overline{x}_A - \overline{x}_B) &= V(\overline{x}_A) + V(\overline{x}_B) \\ &= \frac{\sigma^2}{n_A} + \frac{\sigma^2}{n_B} \\ &= \sigma^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right) \end{aligned}$$

したがって、標準誤差 SE は、σ^2 の代わりに s_p^2 を用いて

2 つの標本 A,B の平均の差の標準誤差

$$SE(\overline{x}_A - \overline{x}_B) \approx s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

例 2 数学の小テストの結果について、A 組から 10 人、B 組から 20 人の標本を得た。A 組と B 組のプールした分散の値は 10.39 である。$SE(\overline{x}_A - \overline{x}_B)$ を求めよ。

答 $SE(\overline{x}_A - \overline{x}_B) =$ ________

問 2 英語の小テストの結果について、A 組から 5 人、B 組から 8 人の標本を得た。A 組と B 組のプールした分散の値は 10.73 である。$SE(\overline{x}_A - \overline{x}_B)$ を求めよ。

答 $SE(\overline{x}_A - \overline{x}_B) =$ ________

プールした分散から標準偏差を求め、SE を決定することで、t 検定を行う準備ができた。

対応のない 2 標本の t 検定（6 ステップ）

Step 1：仮説を立てる

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$、$\mu_1 < \mu_2$、または $\mu_1 \neq \mu_2$

Step 2：各グループの平均と標準偏差を求める

グループ 1：$\bar{x}_1, s_1, n_1$

グループ 2：$\bar{x}_2, s_2, n_2$

Step 3：プールした標準偏差を計算

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Step 4：t 値を計算

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Step 5：臨界値を求める 自由度：$\phi = n_1 + n_2 - 2$

Step 6：結論を述べる

例 3 2 つの高校 A と B から 10 人ずつ無作為抽出して数学の得点を調べたところ、次のようになった。2 つの学校の平均点に差があると言えるか。有意水準 5% で検定しなさい。

高校	標本	平均点	標準偏差
A	10	69.0	5.5
B	10	65.5	7.0

問 3 男子と女子の通学時間（分）を調査したところ、次のようになった。男子と女子の平均通学時間に差があると言えるか。有意水準 5% で検定しなさい。

	標本	平均時間	標準偏差
男子	8	33.00	2.00
女子	8	26.25	2.66

補足

Welch の t 検定について

この教材では、等分散を仮定した対応のない 2 標本の t 検定を学んだ。実際には、等分散を仮定しない **Welch の t 検定**という方法もある。Welch の方法は等分散の仮定が不要だが、自由度の計算が複雑になる。統計ソフトでは、Welch の方法が標準的に使われることが多い。

++*+*+*+*+*+*+ **【解答】** *+*+*+*+*+*+*+*+

問 1 英語の小テストの結果について、A 組から 5 人、B 組から 8 人の標本を得たところ以下のようなことが分かった。A 組と B 組の母分散 σ^2 が等しいと仮定して、プールした分散 σ^2 を求めよ。

A 組：$n_A = 5$、不偏分散 $s_A^2 = 12$
B 組：$n_B = 8$、不偏分散 $s_B^2 = 10$

- A 組の偏差平方和 $= (5-1) \times 12 = 48$
- B 組の偏差平方和 $= (8-1) \times 10 = 70$

$$s_p^2 = \frac{48+70}{4+7} \approx 10.73$$

$\boxed{\text{答}}$ $s_p^2 = 10.73$

例 2 数学の小テストの結果について、A 組から 10 人、B 組から 20 人の標本を得た。A 組と B 組のプールした分散の値は 10.39 である。$SE(\overline{x}_A - \overline{x}_B)$ を求めよ。

$$\begin{aligned} SE(\overline{x}_A - \overline{x}_B) &= s_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}} \\ &= \sqrt{10.39} \times \sqrt{\frac{1}{10}+\frac{1}{20}} \\ &= 3.223 \times \sqrt{0.15} \\ &= 1.248 \end{aligned}$$

$\boxed{\text{答}}$ $SE(\overline{x}_A - \overline{x}_B) = 1.248$

問 2 英語の小テストの結果について、A 組から 5 人、B 組から 8 人の標本を得た。A 組と B 組のプールした分散の値は 10.73 である。$SE(\overline{x}_A - \overline{x}_B)$ を求めよ。

$$\begin{aligned} SE(\overline{x}_A - \overline{x}_B) &= s_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}} \\ &= \sqrt{10.73} \times \sqrt{\frac{1}{5}+\frac{1}{8}} \\ &= 3.276 \times 0.570 \\ &= 1.867 \end{aligned}$$

$\boxed{\text{答}}$ $SE(\overline{x}_A - \overline{x}_B) = 1.867$

例 3　2つの高校 A と B から 10 人ずつ無作為抽出して数学の得点を調べたところ、次のようになった。2 つの学校の平均点に差があると言えるか。有意水準 5% で検定しなさい。

高校	標本	平均点	標準偏差
A	10	69.0	5.5
B	10	65.5	7.0

Step 1：仮説を立てる

$H_0 : \mu_1 = \mu_2$ (平均点は等しい)

$H_1 : \mu_1 \neq \mu_2$ (平均点は異なる)

Step 2：各グループの平均と標準偏差

Step 3：プールした標準偏差

$$s_p = \sqrt{\frac{(10-1)\times 5.5^2 + (10-1)\times 7.0^2}{10+10-2}}$$

$$= \sqrt{\frac{713.25}{18}} = \sqrt{39.625} \approx 6.30$$

Step 4：t 値を計算

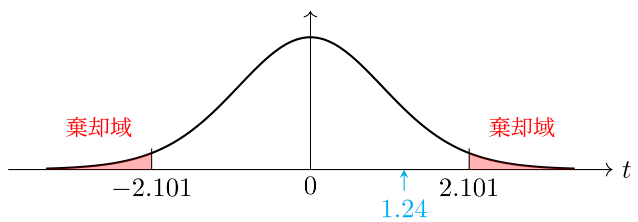
$$t = \frac{69.0 - 65.5}{6.30\sqrt{\frac{1}{10}+\frac{1}{10}}} = \frac{3.5}{6.30\sqrt{0.2}} \approx 1.24$$

Step 5：臨界値を求める

自由度：$\phi = 10 + 10 - 2 = 18$

臨界値（両側検定、自由度 18）：$t_{0.025}(18) = 2.101$

$|t| = 1.24 < 2.101$ なので棄却域に入らない。



Step 6：結論　帰無仮説を棄却できない。

答：有意水準 5% では、2 つの高校の平均点に差があるとは言えない。

問 3　男子と女子の通学時間（分）を調査したところ、次のようになった。男子と女子の平均通学時間に差があると言えるか。有意水準 5% で検定しなさい。

	標本	平均時間	標準偏差
男子	8	33.00	2.00
女子	8	26.25	2.66

Step 1: 仮説を立てる

$H_0 : \mu_1 = \mu_2$ (平均通学時間は等しい)

$H_1 : \mu_1 \neq \mu_2$ (平均通学時間は異なる)

Step 2: 各グループの平均と標準偏差

Step 3: プールした標準偏差

$$s_p = \sqrt{\frac{(8-1)\times 2.00^2 + (8-1)\times 2.66^2}{8+8-2}}$$

$$= \sqrt{\frac{77.56}{14}} = \sqrt{5.54} \approx 2.35$$

Step 4: t 値を計算

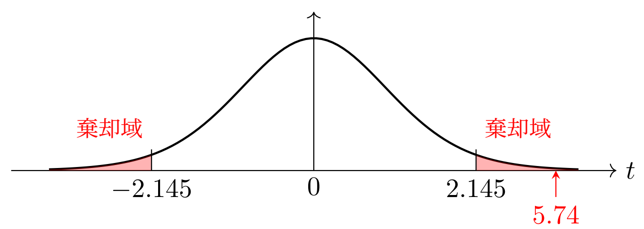
$$t = \frac{33.00 - 26.25}{2.35\sqrt{\frac{1}{8}+\frac{1}{8}}} = \frac{6.75}{2.35\sqrt{0.25}} \approx 5.74$$

Step 5: 臨界値を求める

自由度：$\phi = 8 + 8 - 2 = 14$

臨界値（両側検定、自由度 14）：$t_{0.025}(14) = 2.145$

$|t| = 5.74 > 2.145$ なので棄却域に入る。



Step 6：結論　帰無仮説を棄却する。

答：有意水準 5% で、男子と女子の平均通学時間には差があると言える。