Pooled 2-sample t-test (and confidence interval)

important in order to use pooled 2-sample t procedure

Assumptions: ① Both populations are Normal

② Both population distributions have equal variances (that is, $\sigma_1^2 = \sigma_2^2$)

Aim: ① Find two-sample confidence interval for $\mu_1 - \mu_2$

② Perform hypothesis testing involving $\mu_1 - \mu_2$

Using this procedure, you need to calculate an unbiased estimator for the common variance σ^2 based on s_1^2 and s_2^2.

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

pooled

(also remember that degrees of freedom for this procedure is $n_1 + n_2 - 2$)

Example

Find 95% CI for $\mu_1 - \mu_2$ given the following:

from Example 8.4+8.5 in coursenotes

$$\begin{cases} \bar{x}_1 = 22.3 & \bar{x}_2 = 20.7 \\ n_1 = 30 & n_2 = 10 \\ s_1^2 = 2.9^2 & s_2^2 = 2.5^2 \end{cases}$$

95% CI for $\mu_1 - \mu_2$ is

remember that confidence intervals are always 2-sided, so use $\alpha/2$ for CI calculation

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, df=n_1+n_2-2} \cdot se$$

$s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Here's why:

[$var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ assuming independence, and since $\sigma_1^2 = \sigma_2^2$ (by assumption), we use s_p^2 to estimate the common variance

$$\Rightarrow var(\bar{X}_1 - \bar{X}_2) = s_p^2(\frac{1}{n_1} + \frac{1}{n_2})$$

$$\Rightarrow se(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$]

Calculation:

$$\begin{aligned} s_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\ &= \frac{(29)(2.9^2) + (9)(2.5^2)}{30+10-2} \\ &= 7.8984 \end{aligned}$$

note! 95% CI, 2-sided so remember to use 0.025

$$t_{0.025,\, df=30+10-2} = t_{0.025,38}$$

But look at t-table carefully and you'll notice there is no $df = 38$ entry in t-table (you've to choose between 30 or 40). Rule to be conservative: choose the smaller df. [this differs from your coursenotes which rounds it to the nearest df].

Hence, we use $t_{0.025,30}$ instead

=2.042

Ans:

95% CI for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t \cdot se$$

$$\underbrace{(22.3 - 20.7)}_{\approx 1.6} \pm 2.042 \cdot \underbrace{s_p}_{\sqrt{7.8984}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad n_1 = 30 \qquad n_2 = 10$$

$$= (-0.496,\ 3.696)$$

Pooled 2-sample t test example

$H_0 : \mu_1 = \mu_2$

[rewrite as $H_0 : \mu_1 - \mu_2 = 0$
$H_a : \mu_1 - \mu_2 < 0$]

$H_a : \mu_1 < \mu_2$

this is to remind you $\mu_0 = 0$

Under null hypothesis, the test statistic is

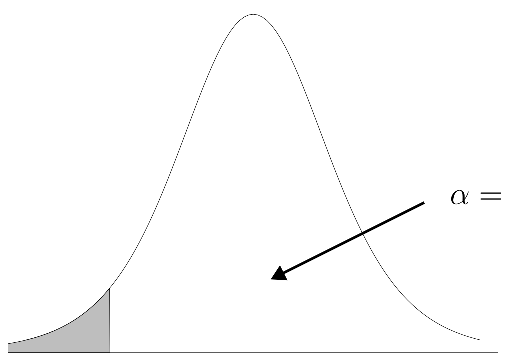
(let's use data from prev eg,
$n_1 = 30, \bar{x}_1 = 22.3, s_1^2 = 2.9^2$,
$n_2 = 10, \bar{x}_2 = 20.7, s_2^2 = 2.5^2$)

$\sqrt{7.8984}$
calculated previously

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(22.3 - 20.7) - 0}{\sqrt{7.8984}\sqrt{\frac{1}{30} + \frac{1}{10}}}$$

$$= 1.56$$



$\alpha = 0.05$, Area=0.95

$-t_{0.05,38} = -1.697$

$df = n_1 + n_2 - 2$

we use df=30 from t table since
38 is not listed in table

Since 1.56 does not fall within the critical region, we do not reject H_0. We conclude there is no statistically significant difference between the two means.

Example 8.6 (Pg 139)

What if you need to find 95% CI for $20\mu_1 - 30\mu_2$?

Not hard at all: just find the correct "se".

Eg.	$\bar{x}_1 = 31.0$	$\bar{x}_2 = 22.7$
	$n_1 = 20$	$n_2 = 20$
	$s_1 = 2.1$	$s_2 = 1.9$

Error on Page 139 (s_2 *should be* 1.9)

95% CI for $20\mu_1 - 30\mu_2$ is

$$20\bar{x}_1 - 30\bar{x}_2 \pm t_{20+20-2} \cdot se$$

calculate as follows:

$$\begin{aligned} Var(20\bar{X}_1 - 30\bar{X}_2) &= 20^2\frac{\sigma_p^2}{n_1} + 30^2\frac{\sigma_p^2}{n_2} \qquad (assume\ \bar{x}_1, \bar{x}_2\ independent) \\ &= \sigma_p^2(\frac{20^2}{20} + \frac{30^2}{20}) \\ \Rightarrow se &= \sigma_p\sqrt{\frac{20^2}{20} + \frac{30^2}{20}} \\ where \quad s_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\ &= \frac{(19)(2.1^2) + (19)(1.9^2)}{20+20-2} \\ &= 4.01 \end{aligned}$$

$$t_{0.025,38} \approx t_{0.025,30} = 2.042$$

Hence, 95% CI for $20\mu_1 - 30\mu_2$ is

$$-61 \pm 2.042 \times \sqrt{4.01}\sqrt{\frac{20^2}{20} + \frac{30^2}{20}} = (-93.97, -28.03)$$

Note:

Your coursenotes do not show an example on this, but we can easily handle hypothesis testing such as:

$H_0 : \mu_1 - \mu_2 = 2$

$H_a : \mu_1 - \mu_2 > 2$

How?

Just note $\mu_0 = 2$, the test statistic is

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - 2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

You can proceed with hypothesis testing as before.

Problem 8.1

$$n_1 = 15 \qquad\qquad n_2 = 12$$

$$\bar{x} = 20 \qquad\qquad \bar{y} = 17$$

$$\sum(x_i - \bar{x})^2 = 28 \qquad\qquad \sum(y_i - \bar{y})^2 = 22$$

$$\Rightarrow \quad s_1^2 = \frac{28}{14} \qquad\qquad \Rightarrow \quad s_2^2 = \frac{22}{11}$$

(a) s_p is needed in calculating 2-sample CI when the variance of the two populations are equal.

$$\begin{aligned} s_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\ &= \frac{(14)\times\frac{28}{14} + 11\times\frac{22}{11}}{15+12-2} = 2 \end{aligned}$$

(b) 95% CI for $\mu_1 - \mu_2$:

$$\bar{x} - \bar{y} \pm t_{0.025,25} \cdot \overbrace{se}^{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$(20-17) \pm 2.06(\sqrt{2}\sqrt{\frac{1}{15}+\frac{1}{12}})$$

$$(1.87,\ 4.13)$$

(c) Check: $\alpha = 0.05$ [matches 95% CI, so can use 95% CI to answer question]

$$H_0 : \mu_1 = \mu_2 \qquad\qquad H_a : \mu_1 \neq \mu_2$$

Since 0 is not in the interval (1.87, 4.13), we reject H_0 at the 5% level.