PSYC 2317 - Lecture 8

Recall: we can estimate a 95% confidence interval for an unknown population mean μ by using the sample mean $\bar{X}$ and the population standard deviation σ as

$$\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{N}}$$

or equivalently

$$\bar{X} \pm 1.96 \cdot SE$$

What if we are not given σ?

Can we use our estimate $\hat{\sigma} = \sqrt{\frac{SS}{N-1}}$?

Well, yes - sort of - but we have to adjust the 1.96

Why?

* 1.96 is used because for a normal distribution, 95% of sample means fall between $-1.96 \cdot SE$ and $1.96 \cdot SE$. This assumes σ is known.

* if estimating σ with $\hat{\sigma}$, we get a t-distribution for the sample means. The exact shape of this distribution depends on the size of the sample.

In light of this, let's define a generalized confidence interval

$$\bar{X} \pm t^*_{df} \cdot SE$$

where

* the value of t^*_{df} depends on sample size
    → defined as the value of t which leaves 5% of the distribution in the two tails (combined).
        ↳ sometimes called the critical value of the t-distribution
        ↳ easy to find from distribution calculator app

* the formula for SE depends on design:
    * for single sample (or repeated measures), we have

$$SE = \frac{\hat{\sigma}}{\sqrt{N}}$$

    * for independent samples, we have

$$SE = \hat{\sigma}_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

Example 1 - single sample design

A sample of 25 people is given a treatment. After treatment, we find $\overline{X} = 22.2$ with $SS = 384$. Construct a 95% confidence interval for μ, the population mean of the treatment group.

$$95\%\ CI = \overline{X} \pm t^*_{df} \cdot SE$$

(1) $df = 24$, so from app we find $t^*_{24} = 2.06$

(2) single sample $\longrightarrow$ $SE = \dfrac{\hat{\sigma}}{\sqrt{N}}$

$$\hat{\sigma} = \sqrt{\frac{SS}{N-1}} = \sqrt{\frac{384}{25-1}} = \sqrt{\frac{384}{24}} = \sqrt{16} = 4$$

so $SE = \dfrac{\hat{\sigma}}{\sqrt{N}} = \dfrac{4}{\sqrt{25}} = \dfrac{4}{5} = 0.8$

Putting it all together:

$$95\%\ CI = \overline{X} \pm t^*_{df} \cdot SE$$

$$= 22.2 \pm 2.06 \cdot 0.8$$

$$= 22.2 \pm 1.648 = (20.55,\ 23.85)$$

For independent groups designs, the goal is to estimate the "mean difference" $\mu_1 - \mu_2$. The resulting formula for 95% CI is

$$(\bar{X}_1 - \bar{X}_2) \pm t^*_{df} \cdot SE$$

$$\text{where } SE = \hat{\sigma}_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

Example 2: Does watching educational TV as a kid predict better high school grades?

Educ. TV	No educ. TV
$N_1 = 10$	$N_2 = 10$
$\bar{X}_1 = 93$	$\bar{X}_2 = 85$
$SS_1 = 200$	$SS_2 = 160$

Compute a 95% CI for the mean difference $\mu_1 - \mu_2$.

solution:

$$95\%\ CI = (\bar{X}_1 - \bar{X}_2) \pm t^*_{df} \cdot SE$$

(1) $df = df_1 + df_2 = 9 + 9 = 18$, so $t^*_{df} = 2.10$

(2) two independent samples, so $SE = \hat{\sigma}_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$

$$\hat{\sigma}_p = \sqrt{\frac{SS_1 + SS_2}{df_1 + df_2}} = \sqrt{\frac{200 + 160}{9 + 9}} = \sqrt{\frac{360}{18}} = \sqrt{20}$$
$$= 4.47$$

$$SE = \hat{\sigma}_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} = 4.47 \sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$= 2.00$$

Putting it all together:

$$95\%\ CI = (\bar{X}_1 - \bar{X}_2) \pm t^*_{df} \cdot SE$$
$$= (93 - 85) \pm 2.10(2.00)$$
$$= 8 \pm 4.2$$
$$= (3.8,\ 12.2)$$