# Lec 8

## Recall that

$\boxed{1}$ Confidence Interval for $\mu$

Let $X_1, X_2, \dots X_n$ R.S. from $N(\mu, \sigma^2)$

$\boxed{A}$ When $\sigma^2$ is Known

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$\boxed{B}$ when $\sigma^2$ is Unknown but $n \geq 30$

replace $\sigma \rightarrow S$

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$\boxed{C}$ When $n < 30$ and $\sigma^2$ is Unknown

$$P\left(\bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

# Recall that

As Confidence level $\gg$

$1-\alpha \gg$

$\alpha \ll$


$Z_{\frac{\alpha}{2}} \gg$

$n \ll$

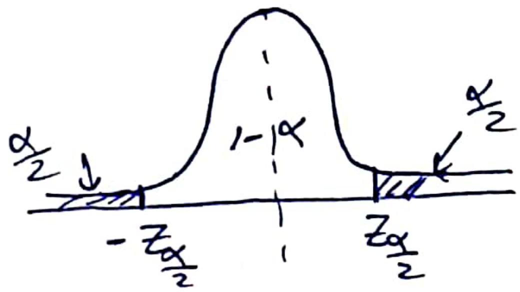
$\sigma \gg$

Range $\gg$

As Conf. level $\ll$

$1-\alpha \ll$

$\alpha \gg$




$Z_{\frac{\alpha}{2}} \ll$

$n \gg$

$\sigma \ll$

Range $\ll$

## 2 Confidence Interval for Difference of means of two populations $\mu_1 - \mu_2$

If $X_1, X_2, \ldots X_n$ , $Y_1, Y_2, \ldots Y_{n_2}$ R.S from $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$

When you want to compare between two methods or Products

**Case 1** [use of the normal distr]

If $\sigma_1^2, \sigma_2^2$ are known or

$n_1 \geq 30$ , $n_2 \geq 30$

sample size

Population #1 : $\mu_1, \sigma_1^2 \Rightarrow n_1$

population 2 : $\mu_2, \sigma_2^2 \Rightarrow n_2$

Sample of Pop 1

$X_1, X_2, \ldots X_{n_1} \Rightarrow \overline{X_1}$

Sample of pop 2

$X_1, X_2, \ldots, X_{n_2} \Rightarrow \overline{X_2}$

Compute the difference $\overline{X_1} - \overline{X_2}$

$\overline{X_1} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$

$\overline{X_2} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$

$\overline{X_1} - \overline{X_2} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$Pr\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$


$$P\left((\overline{X_1} - \overline{X_2}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{X_1} - \overline{X_2}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

* If $\sigma_1^2, \sigma_2^2$ are <u>unknow</u> but $n_1 \geq 30, n_2 \geq 30$ we may replace $\sigma_1^2$ by $s_1^2$ and $\sigma_2^2$ by $S_2^2$ without affecting the confidence interval

$$(\overline{X_1} - \overline{X_2}) - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{X_1} - \overline{X_2}) + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

is $100(1-\alpha)\%$ C.I

EX (3.5) A standard Mathematics test was given to 50 girls and 75 boys at Certain College. The girls made an average grade of 76 with st. dev of 6, while the boys made an average grade of 82 with a st. dev of 8. Find 95% Confidence interval for the difference $(\mu_1 - \mu_2)$ where $\mu_1, \mu_2$ are the mean scores of all boys and all girls respectively

boys $n_1 = 75$, $\bar{x}_1 = 82$, $S_1 = 8$

girls $n_2 = 50$, $\bar{x}_2 = 76$, $S_2 = 6$

Since $n_1, n_2$ are large $(\geq 30)$ then use normal distr to get C.I.

$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$\Phi(Z_{\alpha/2}) = 1 - \frac{\alpha}{2} \Rightarrow \Phi(Z_{\alpha/2}) = 0.975$

Using Minitab $\Rightarrow Z_{\frac{\alpha}{2}} = 1.96$

$$(82-76) - (1.96)\sqrt{\frac{8^2}{75} + \frac{6^2}{50}} \leq \mu_1 - \mu_2 \leq (82-76) + (1.96)\sqrt{\frac{8^2}{75} + \frac{6^2}{50}}$$

$$3.54 \leq \mu_1 - \mu_2 \leq 8.46$$

is the 95% C.I. for $\mu_1 - \mu_2$

Case 2 : (use of the t distr.)

if $\sigma_1, \sigma_2$ are unknown, and $n_1 < 30, n_2 < 30$

Case (1) $\sigma_1^2 = \sigma_2^2 = \sigma^2$ unknown

We estimate $\sigma^2$ by $S_p^2$ Pooled Variance

$S_p^2$ is obtained by combining (pooling) the sample variances

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 \quad \text{of } 1^{st} \text{ sample}$$

$$(n_1 - 1) S_1^2 = \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2$$

Similarly $$(n_2 - 1) S_2^2 = \sum_{i=1}^{n_2} (x_i - \bar{x}_2)^2 \quad 2^{nd} \text{ sample}$$

If we combine of two samples

$$S_p^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_i - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)} \qquad \frac{\text{Sum squares}}{\text{total } \# - 1}$$

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

$$Z_{\frac{\alpha}{2}} \longrightarrow t_{\frac{\alpha}{2}}, \quad \nu = \overbrace{n_1 + n_2 - 2}^{(n_1-1)+(n_2-1)}$$

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \longrightarrow \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$\therefore$ 100 $(1-\alpha)\%$ Conf. Interval for $\mu_1-\mu_2$

$$(\bar{X}_1-\bar{X}_2)-t_{\frac{\alpha}{2}} S_p\sqrt{\tfrac{1}{n_1}+\tfrac{1}{n_2}}<\mu_1-\mu_2 \leq (\bar{X}_1-\bar{X}_2)+t_{\frac{\alpha}{2}} S_p\sqrt{\tfrac{1}{n_1}+\tfrac{1}{n_2}}$$

---

**Ex 3.6** Seven plants of wheat grown in pots and given a standard fertilizer treatment yield respectively 8.2, 4.4, 4.0, 6.3, 4.7, 11.0, 9.7 gram dry weight of seed. A further eight plants from the same source are grown in similar conditions but with different fertilizer and yield respectively 12.5, 7.3, 10.6, 8.2, 13.0, 6.4, 9.6, 13.2 g.

Construct a 95% C.I for the difference of means of two populations weights of seed

$\mu_1$ : mean dry weight of seeds using fertilizer 1

$\mu_2$ : " " " " " " 2

since $n_1 < 30$, $n_2 < 30$ and $\sigma_1, \sigma_2$ are unknown

$n_1 = 7 \Rightarrow \bar{X}_1 = \frac{1}{7}\sum_{i=1}^{7} x_i = 6.9$, $S_1^2 = 7.70$

$n_2 = 8 \Rightarrow \bar{X}_2 = \frac{1}{8}\sum_{i=1}^{8} x_i = 10.1$, $S_2^2 = 7.06$

$\boxed{\because \bar{X}_2 > \bar{X}_1 \quad \bar{X}_2-\bar{X}_1}$

Assume $\sigma_1^2=\sigma_2^2=\sigma^2$

$$S_p^2 = \frac{(6)(7.70)+(7)(7.06)}{7+8-2} = 7.355$$

Pooled variance

$S_p = 2.712$ Pooled St. dev

$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$, $\nu = 7+8-2 = 13$

$$\Phi(t_{\frac{\alpha}{2},13}) = 1-\frac{\alpha}{2} = 1.975$$

by minitab $t_{\alpha/2} = 2.16$

$$(\bar{X}_2-\bar{X}_1) - t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq \mu_2-\mu_1 \leq (\bar{X}_2-\bar{X}_1) + t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(10.1-6.9)-(2.16)(2.71)\sqrt{\frac{1}{7}+\frac{1}{8}} \leq \mu_2-\mu_1 \leq (10.1-6.9)+(2.16)(2.71)\sqrt{\frac{1}{7}+\frac{1}{8}}$$

$$0.17 \leq \mu_2-\mu_1 \leq 6.03$$

**Note**

You can get Confidence interval estimation of $\mu$ using normal, t-distr. by Minitab easily

[watch appendix Video How to use Minitab to solve Ex 3.1 , 3.4]

Also, to get C.I of $\mu_1-\mu_2$ using Minitab easily Ex 3.5, Ex 3.6

⟹ watch appendix Video

Regards

☺