Confidence ~~Interval~~ Interval 28th April 2023

Interpretation of the CI

A $100(1-\alpha)\%$ CI of θ based on the random sample $X_1, \ldots, X_n$ is $[T_1, T_2]$

$$\Updownarrow$$

If we repeatedly draw samples from the underlying distributions, then the random interval $[T_1, T_2]$ contains the true value of θ approximately $\underbrace{100(1-\alpha)\%}_{CI}$ cases.

$\alpha = 0.05$ Then $100(1-\alpha)\% = 95\%$

The random interval $[T_1, T_2]$ contains the true value of θ approximately 95% of the times when random sample are repeatedly from the underlying distribution.

1) Testing for μ when σ^2 is known

~~comparison~~ Here the concerned test statistic is

$$T = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim Z = N(0,1)$$

To get 95% C.I, we have $\left[\alpha = 0.05\right]$ significance level

$$Pr\left(-z_{\alpha/2} < T < z_{\alpha/2}\right) = 0.95 = (1-\alpha)$$

$$Pr\left(-z_{\alpha/2} < \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} < z_{\alpha/2}\right) = 0.95$$

$\Rightarrow$ Interval is $\quad -z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < (\bar{x} - \mu) < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$\Rightarrow \underbrace{\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{T_1} < \mu < \underbrace{\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{T_2}$$

$\therefore$ $[T_1, T_2]$ serves as the $(1-\alpha)100\%$ confidence interval for μ.

2) Testing for μ when σ^2 is unknown

Comparison value be μ'

Here the concerned test statistic is

$$T = \frac{\sqrt{n}(\bar{x} - \mu')}{s} \quad \text{where}$$

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

$$\text{and } \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

So $T \sim t_{n-1}$

To get 95% CI, we have

$$Pr\left(-t_{n-1, \alpha/2} < \frac{\sqrt{n}(\bar{x} - \mu)}{s} < t_{n-1, \alpha/2} \right) = 1 - \alpha$$

$$\therefore \quad \bar{x} - t_{n-1; \alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1; \alpha/2} \frac{s}{\sqrt{n}}$$

$$\Downarrow$$

$$T_1 < \mu < T_2$$

Then $[T_1, T_2]$ serves as the $100(1-\alpha)\%$ confidence interval for μ.

- Population with low variation leads to similar samples with lower variation leads to narrow CI.

3) Testing for σ^2 when μ is known $= (\mu_0)$

$X \sim N(\mu_0, \sigma^2)$. Comparison value $(\sigma^0)^2$

Define $s^2 = \frac{1}{n} \Sigma (x_i - \mu_0)$

we know $\frac{ns^2}{(\sigma^0)^2} \sim \chi^2_n$

Suppose our CI is $1-\alpha$, $0 < \alpha < 1$
then we have,

$$P_r \left\{ \chi^2_{n;1-\alpha/2} < \frac{ns^2}{(\sigma^0)^2} < \chi^2_{n;\alpha/2} \right\} = 1-\alpha$$

$\Rightarrow \frac{ns^2}{\chi^2_{n;\alpha/2}} < (\sigma^0)^2 < \frac{ns^2}{\chi^2_{n;1-\alpha/2}}$ is the CI interval for σ^2

let $T_1 = \frac{ns^2}{\chi^2_{n;\alpha/2}}$ $T_2 = \frac{ns^2}{\chi^2_{n;1-\alpha/2}}$

(T_1, T_2) is CI for σ^2 when μ is known.

4) Testing for σ^2 when μ is unknown.

~~let ... be the~~

$X \sim N(\mu, \sigma^2)$ where μ is unknown. Let $X_1, \dots, X_n$ be n iid samples of X

$\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$

Define $s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$

Consider the test statistic $T = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

Consider the confidence coefficient to be $(1-\alpha)$. $0 < \alpha < 1$.

then $P_r \{ \chi^2_{n-1; 1-\alpha/2} < T < \chi^2_{n-1; \alpha/2} \} = (1-\alpha)$

$\Rightarrow$ The interval is $\frac{(n-1)s^2}{\chi^2_{n-1; \alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1; 1-\alpha/2}}$

is the CI for σ^2

5) Testing for difference of $(\mu_1 - \mu_2)$

$X_1, \dots, X_{n_1} \overset{IID}{\sim} X$

$X \sim N(\mu_1, \sigma_1^2)$ $\quad$ $Y_1, \dots, Y_{n_2} \overset{IID}{\sim} Y$

$Y \sim N(\mu_2, \sigma_2^2)$.

(a) when σ_1^2 and σ_2^2 are known $\sigma_{10}^2, \sigma_{20}^2$

$$\bar{X} = \frac{1}{n_1}\left(\sum_{i=1}^{n_1} X_i\right) \quad \text{and} \quad \bar{Y} = \frac{1}{n_2}\left(\sum_{i=1}^{n_2} Y_i\right)$$

test statistic $T = \dfrac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_{10}^2}{n_1} + \frac{\sigma_{20}^2}{n_2}} = u} \sim Z \equiv N(0,1)$

T_{obs} uses (μ_1', μ_2') the comparison values.

Suppose the confidence coefficient be $1-\alpha$, $(0<\alpha<1)$

Then,

$$P_r\left\{-\tau_{\alpha/2} < T < \tau_{\alpha/2}\right\} = 1-\alpha$$

$\Rightarrow$ The interval is

$$-\tau_{\alpha/2} u < (\bar{X}-\bar{Y}) - (\mu_1 - \mu_2) < u\tau_{\alpha/2}$$

$$\Rightarrow (\bar{X}-\bar{Y}) - u\tau_{\alpha/2} < \mu_1 - \mu_2 < (\bar{X}-\bar{Y}) + u\tau_{\alpha/2}$$

Hence $T_1 = (\bar{X}-\bar{Y}) - u\tau_{\alpha/2}$

$T_2 = (\bar{X}-\bar{Y}) + u\tau_{\alpha/2}$

Thus (T_1, T_2) is the interval serves $100(1-\alpha)\%$ CI for $(\mu_1 - \mu_2)$

(b) when σ_1^2, σ^2 are unknown

By assumption of homoscedasticity, assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$ though unknown

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i) \quad \text{and} \quad \bar{Y} = \frac{1}{n_2}\left(\sum_{i=1}^{n_2} Y_i\right).$$

$$S_X^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \quad \text{and} \quad S_Y^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$$

Let $S^2 = \dfrac{(n_1-1)S_X^2 + (n_2-1)S_Y^2}{n_1+n_2-2}$. Let $S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = S'$

Consider the test statistic $T = \dfrac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

Consider the confidence coefficient to be $(1-\alpha)$, $\alpha < 1$

Then

5) Testing for difference of $(\mu_1 - \mu_2)$ $X_1, \ldots, X_{n_1} \overset{IID}{\sim} X$

$X \sim N(\mu_1, \sigma_1^2)$ $\quad Y_1, \ldots, Y_{n_2} \overset{IID}{\sim} Y$

$Y \sim N(\mu_2, \sigma_2^2)$.

(a) when σ_1^2 and σ_2^2 are known. $\sigma_{10}^2, \sigma_{20}^2$.

$$\bar{X} = \frac{1}{n_1}\left(\sum_{i=1}^{n_1} X_i\right) \quad \text{and} \quad \bar{Y} = \frac{1}{n_2}\left(\sum_{i=1}^{n_2} Y_i\right)$$

test statistic $T = \dfrac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_{10}^2}{n_1} + \frac{\sigma_{20}^2}{n_2}}} \sim Z \equiv N(0,1)$

$\sqrt{\frac{\sigma_{10}^2}{n_1} + \frac{\sigma_{20}^2}{n_2}} = u$

T_{obs} uses (μ_1', μ_2') the comparison values.

Suppose the confidence coefficient be $1-\alpha$, $(0<\alpha<1)$

Then,

$$Pr\left\{ -\tau_{\alpha/2} < T < \tau_{\alpha/2} \right\} = 1-\alpha$$

$\Rightarrow$ The interval is

$$-\tau_{\alpha/2}\, u < (\bar{X}-\bar{Y}) - (\mu_1 - \mu_2) < u\,\tau_{\alpha/2}$$

$$\Rightarrow (\bar{X}-\bar{Y}) - u\,\tau_{\alpha/2} < \mu_1 - \mu_2 < (\bar{X}-\bar{Y}) + u\,\tau_{\alpha/2}$$

Hence $T_1 = (\bar{X}-\bar{Y}) - u\,\tau_{\alpha/2}$

$T_2 = (\bar{X}-\bar{Y}) + u\,\tau_{\alpha/2}$

Thus (T_1, T_2) is the Interval serves $100(1-\alpha)\%$ CI for $(\mu_1 - \mu_2)$

(b) when σ_1^2, σ^2 are unknown

By assumption of homoscedasticity, assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$ though unknown

$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i)$ and $\bar{Y} = \frac{1}{n_2}\left(\sum_{i=1}^{n_2} Y_i\right)$.

$$S_X^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \quad \text{and} \quad S_Y^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$$

Let $S^2 = \dfrac{(n_1-1)S_X^2 + (n_2-1)S_Y^2}{n_1+n_2-2}$. Let $S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = s'$

Consider the test statistic $T = \dfrac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

Consider the confidence coefficient to be $(1-\alpha)$, $\alpha < 1$

Then

$$Pr\left\{ -t_{n_1+n_2-2;\,\alpha/2} < T < t_{n_1+n_2-2;\,\alpha/2} \right\} = 1-\alpha$$

$\Rightarrow$ Interval is

$$-t_{n_1+n_2-2;\,\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S'} < t_{n_1+n_2-2;\,\alpha/2}$$

$$\Rightarrow \quad (\bar{X}-\bar{Y}) - s'\,t_{n_1+n_2-2;\,\alpha/2} < (\mu_1-\mu_2) < (\bar{X}-\bar{Y}) + s'\,t_{n_1+n_2-2;\,\alpha/2}$$

let $T_1 = \bar{X}-\bar{Y} - s'\,t_{n_1+n_2-2;\,\alpha/2}$ and

$T_2 = \bar{X}-\bar{Y} + s'\,t_{n_1+n_2-2;\,\alpha/2}$. Then the Intervral (T_1, T_2) that serves $100(1-\alpha)\%$ CI for $(\mu_1-\mu_2)$.

6) Testing for σ_1/σ_2

$$X_1, \dots, X_{n_1} \overset{iid}{\sim} X \quad \text{where } X \sim N(\mu_1, \sigma_1^2)$$
$$Y_1, \dots, Y_{n_2} \overset{iid}{\sim} Y \quad \text{where } Y \sim N(\mu_2, \sigma_2^2)$$

(a) When μ_1, μ_2 is known.

let $\sigma_1 = \sigma_X$ and $\sigma_2 = \sigma_Y$

let $\bar{X} = \frac{1}{n_1}\left(\sum_{i=1}^{n_1} X_i\right)$ and $\bar{Y} = \frac{1}{n_2}\left(\sum_{i=1}^{n_2} Y_i\right)$

and $S_X^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_1)^2$ and $S_Y^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - \mu_2)^2$

$\therefore$ Consider the test statistic

$$T = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} = \frac{(n_1 S_X^2/\sigma_X^2)/(n_1)}{(n_2 S_Y^2/\sigma_Y^2)/(n_2)} \sim F_{n_1, n_2}$$

$\therefore$ Consider the confidence coefficient as $(1-\alpha)$, $0<\alpha<1$

then $\Pr\left(F_{n_1, n_2, 1-\alpha/2} < T < F_{n_1, n_2, \alpha/2}\right) = 1-\alpha$

$\Rightarrow$ the interval is $F_{n_1, n_2, 1-\alpha/2} < \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} < F_{n_1, n_2, \alpha/2}$

$$\Rightarrow \underbrace{\frac{S_X^2/S_Y^2}{F_{n_1, n_2, \alpha/2}}}_{=T_1} < \frac{\sigma_X^2}{\sigma_Y^2} < \underbrace{\frac{S_X^2/S_Y^2}{F_{n_1, n_2, 1-\alpha/2}}}_{=T_2}$$

Then the interval $[T_1, T_2]$ serves $100(1-\alpha)\%$ confidence for $\frac{\sigma_X^2}{\sigma_Y^2}$

(b) when μ_1, μ_2 is unknown

let $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ $\quad \bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$

and $s_X^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})$ and $s_Y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})$

Now consider the test statistic, $T = \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2}$

where $T \sim F_{n_1 - 1, n_2 - 1}$

Consider the confidence coefficient $(1-\alpha)$ $0 < \alpha < 1$, then

$$\Pr\left\{ F_{n_1 - 1; n_2 - 1; 1-\alpha/2} < T < F_{n_1 - 1; n_2 - 1; \alpha/2} \right\} = 1 - \alpha$$

So the interval is,

$$\underbrace{\frac{(S_X^2 / S_Y^2)}{F_{n_1 - 1; n_2 - 1; \alpha/2}}}_{T_1} < \frac{\sigma_X^2}{\sigma_Y^2} < \underbrace{\frac{(S_X^2 / S_Y^2)}{F_{n_1 - 1; n_2 - 1; 1-\alpha/2}}}_{T_2}$$

Then the interval $[T_1, T_2]$ serves $100(1-\alpha)\%$ confidence for $\frac{\sigma_X^2}{\sigma_Y^2}$