Confidence Interval 2

Mainak Thakur

15.09.2020

Example

Example

A study is conducted to estimate the mean BP (blood pressure) for a single target pop^n. A sample is collected and mean, s.d. and sample size are tabulated. Find 95% CI for mean BP for the target population.

Given: The pop^n BP follows Normal dist^n.

Ref: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Confidence_Intervals/BS704_Confidence_Intervals_print.html

Example

Case 1.

$n = 3534, \quad \bar{x} = 127.3, \quad s.d. = 19.$

$n \geq 30$, Normal popⁿ, $1-\alpha = .95, \quad \alpha = 0.05,$

$\alpha/2 = 0.025$

$$95\% \text{ CI} \rightarrow \left(\bar{x} - z_{0.025} \times \frac{s}{\sqrt{n}},\ \bar{x} + z_{0.025} \times \frac{s}{\sqrt{n}}\right)$$

$$\left(127.3 - 1.96 \times \frac{19}{\sqrt{3524}},\ 127.3 + 1.96 \times \frac{19}{\sqrt{3524}}\right)$$

$(126.7, 127.9)$

Ref: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Confidence_Intervals/BS704_Confidence_Intervals_print.html

Example

<u>Case 2.</u> $n = 10$, $\bar{x} = 121.2$, $s = 11.1$.

$n < 30$, Normal popⁿ, $\frac{\alpha}{2} = 0.025$

$$95\%\ CI: \left(\bar{x} - t_{0.025,9} \times \frac{s}{\sqrt{n}},\ \bar{x} + t_{0.025,9} \times \frac{s}{\sqrt{n}}\right)$$

$$t_{0.025,9} = 2.262 \quad \text{(from t-table)}.$$

$$95\%\ CI: \left(121.2 - 2.262 \times \frac{11.1}{\sqrt{10}},\ 121.2 + 2.262 \times \frac{11.1}{\sqrt{10}}\right)$$

$$(113.3,\ 129.1)$$

<u>Interpretation</u>:

Ref: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Confidence_Intervals/BS704_Confidence_Intervals_print.html

CI for Variance

CI for Variance

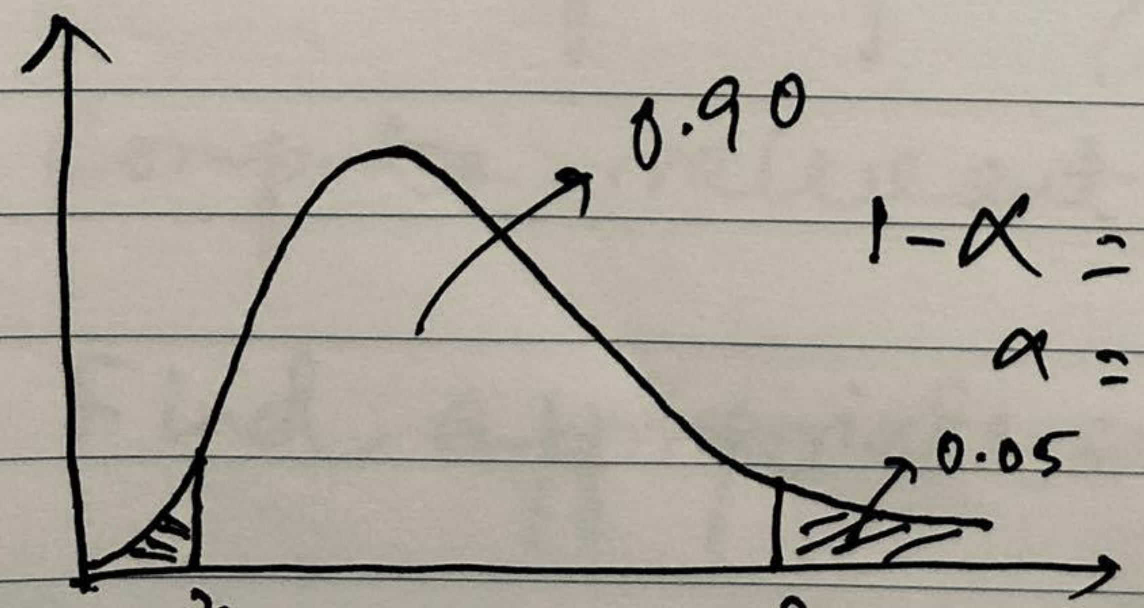
Assumption: Population is normal.

$X_1, X_2, \ldots, X_n$ are iid random sample from $N(\mu, \sigma^2)$. Let $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Then $X^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$

Sampling distn.

$$P\left(\chi^2_{1-\alpha/2, n-1} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{\alpha/2, n-1}\right) = 1-\alpha$$



$\chi^2_{0.95,10}$ $\quad$ $\chi^2_{0.05,10}$

3.94 $\quad$ 18.31

$1-\alpha = 0.90$

$\alpha = 0.1, \ \frac{\alpha}{2} = 0.05$

Ref: Applied Statistics and Probability for Engineers (5th ed.) by Montgomery, D. C., & George, C. R..

CI for Variance

unknown parameter : σ^2

Population : Normal

Sample variance : s^2

random sample of n obs. from Normal popn

100(1−α)% CI : $$\frac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}}$$

Ref: Applied Statistics and Probability for Engineers (5th ed.) by Montgomery, D. C., & George, C. R..

Algorithm for developing CI

General Algorithm for obtaining C.I.

1. Collect sample of size n
2. Compute relevant statistic.
3. Find appropriate sampling distn
4. Choose α
5. Calculate lower and upper limits from the statistical table.

Ref: Applied Statistics and Probability for Engineers (5th ed.) by Montgomery, D. C., & George, C. R..

CI for Difference in Means

CI for Difference in Means

Case 1: Variances known (σ_1^2, σ_2^2).

pop^n 1, pop^n 2

$x_{11}, x_{12}, \ldots, x_{1n_1}$ i.i.d random sample pop^n 1.

$x_{21}, x_{22}, \ldots, x_{2n_2}$ . . . . from pop^n 2.

pop^n 1 (X_1), pop^n 2 (X_2) are independent

Both populations are normal.

Estimator of $\mu_1 - \mu_2 \rightarrow \bar{X}_1 - \bar{X}_2$

$E(\bar{X}_1 - \bar{X}_2) = ?$ $\quad$ $Var(\bar{X}_1 - \bar{X}_2) = ?$

Sampling Dist

Both populations are normal.

Estimator of $\mu_1 - \mu_2 \rightarrow \bar{x}_1 - \bar{x}_2$

$E(\bar{x}_1 - \bar{x}_2) = ?$ $\quad Var(\bar{x}_1 - \bar{x}_2) = ?$

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$

$$Var(\bar{x}_1 - \bar{x}_2) = Var(\bar{x}_1) + Var(\bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Then, $$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Sampling Distribution

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\alpha/2} \le \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \le z_{\alpha/2}\right) = 1-\alpha.$$

$$\Rightarrow P\left(\bar{X}_1-\bar{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1-\mu_2 \le \bar{X}_1-\bar{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right) = 1-\alpha.$$

Note: When σ_1^2, σ_2^2 are unknown, the sample variances s_1^2, s_2^2 can substitute σ_1^2, σ_2^2 for large samples. If $n_1, n_2 \geq 30$, this holds even if the populations are not normally distributed.

Ref: Applied Statistics and Probability for Engineers (5th ed.) by Montgomery, D. C., & George, C. R..

CI for Case 1

sample sizes : n_1 & n_2 , sample means: $\bar{x}_1$ & $\bar{x}_2$

Independent random samples from independent normal populations with known vars σ_1^2 & σ_2^2 resp.

$100(1-\alpha)\%$ CI for $(\mu_1 - \mu_2)$:

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} .$$

S-pooled

Case 2: Variances unknown

2A: $\sigma_1^2 = \sigma_2^2 = \sigma^2$, Normal pop^n, small samples

$$\hat{\sigma}_1^2 = S_1^2, \quad \hat{\sigma}_2^2 = S_2^2$$

But $S_1^2 \neq S_2^2$

Pooled estimator of σ^2: S_p^2

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

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

CI for Case 2A

$\bar{x}_1, \bar{x}_2, s_1^2, s_2^2$: sample means and vars

i.i.d random samples of sizes n_1 and n_2

two independent Normal populations

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

$100(1-\alpha)\%$ CI on the difference in means $(\mu_1 - \mu_2)$:

$$\left[(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

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

CI for Case 2B

2B: $\sigma_1^2 \neq \sigma_2^2$, small samples, Normal pop^{n}s

$\bar{x}_1, \bar{x}_2, s_1^2, s_2^2$: means and vars of two i.i.d random samples

sample size: n_1 & n_2

two independent Normal pop^{n}s with

$\sigma_1^2 \neq \sigma_2^2$ (unknown)

$100(1-\alpha)\%$ CI for $(\mu_1 - \mu_2)$ (Approx):

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, \nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, \nu} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$, take closest integer