

Confidence Intervals: Estimators

Point Estimate:

- a specific value that estimates a parameter
- e.g., “best estimator” of the population mean () is a sample mean (\bar{X})
- problem is that there is no way to determine how close a point estimate is to the parameter

Properties of a Good Estimator:

1. must be an **unbiased estimator** -expected value of estimator or mean obtained from samples of a given size must be equal to the parameter
2. must be **consistent** -as sample size increases estimator approaches value of the parameter
3. must be **relatively efficient** -estimator must have smallest variance of all other estimators

Interval Estimate:

- precise probabilities can be assigned the validity of the interval
- range of values that estimate a parameter
- e.g., mean \pm standard deviation (), 25 ± 10 kg, 5th to 95th-percentile

Confidence Intervals when σ is Known and n is Large

Confidence Interval:

- interval estimate based on sample data and a given confidence level

Confidence Level:

- probability that a parameter will fall within an interval estimate
- related to alpha (α) level, that is,

$$\text{Confidence Level (CL)} = 1 - \alpha$$

E.g., CL = 95% means $\alpha = 0.05$

CL = 99% means $\alpha = 0.01$

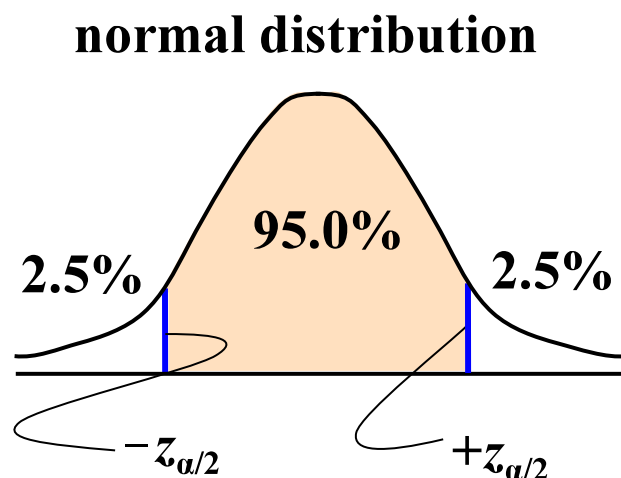
Formula for Computing Confidence Intervals

$$\bar{X} - z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right) < \mu < \bar{X} + z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

where $z_{\alpha/2}$ is the z-score that places the area $\alpha/2$ in each tail of the normal distribution.

For example, if α is 5% then $z_{\alpha/2}$ is the z-score that places 2.5% in the right tail and 2.5% in the left tail.

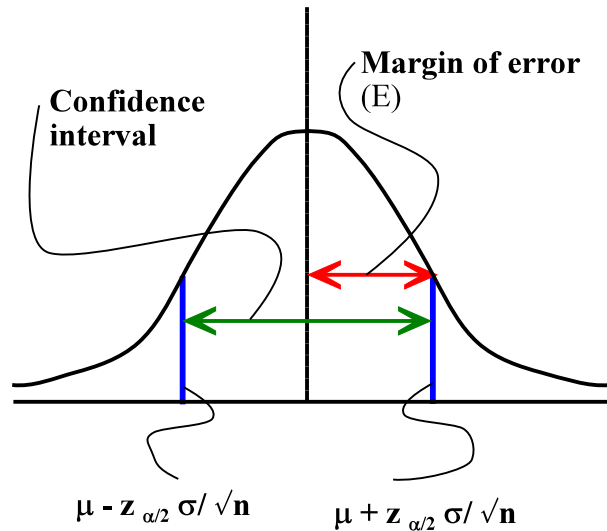
That is $z_{\alpha/2} = \pm 1.960$.



Margin of Error (Maximum Error of Estimate)

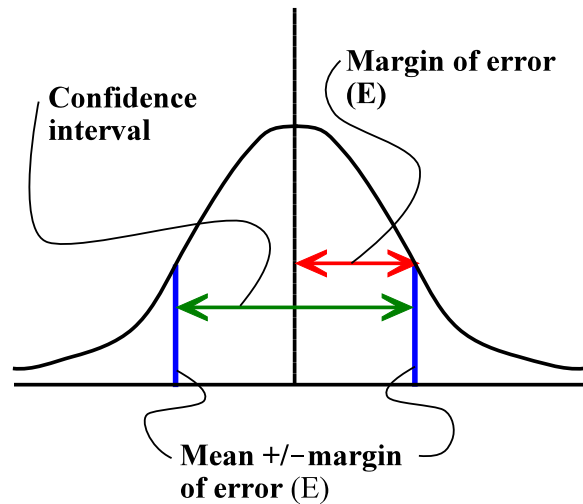
Margin of Error also called Maximum Error of Estimate (E):

$$E = z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$



Example:

Compute the 95th percentile confidence interval from a sample of size 30 which has a standard deviation of 5.00 and a mean of 25.0. Since σ is unknown use s .

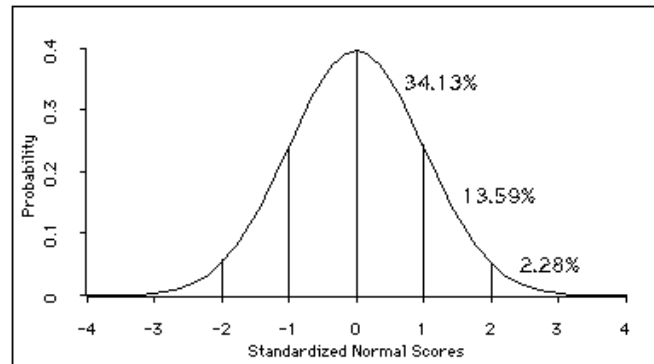


From z-table:

$$z_{\alpha/2} = \pm 1.960$$

Therefore,

$$E = 1.960 \times \frac{5.00}{\sqrt{30}} = 1.789$$



Confidence interval = CI = $25.0 \pm 1.789 = 23.2$ to 26.8

Confidence Intervals when σ is Known and n is Large

- Data must be randomly sampled
- Each datum must be independent of other data
- Data must be “approximately” normally distributed
- If not normally distributed, sample size should be greater than 30
- When population standard deviation is known, calculate margin of error (E) from: $E = z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$
- Add and subtract E from the sample mean

Example

Heights of a random sample of 100 people were collected and after determining they were normally distributed, the mean was computed. Given that the population’s standard deviation was 8.25 cm and the sample mean was 150.5 cm, compute the 95th-percentile confidence interval.

First, determine the z -values that define the 95th-percentile confidence interval. I.e., $z_{\alpha/2} = \pm 1.960$

Next, compute margin of error: $E = (z_{\alpha/2} \times \sigma) / \sqrt{n}$

$$= (1.960 \times 8.25) / \sqrt{100} = 16.17 / 10 = 1.617 \text{ cm}$$

Finally, add and subtract E from sample mean to define the confidence interval:

$$\text{CI: } 148.9 \text{ cm} < \bar{x} < 152.1 \text{ cm}$$

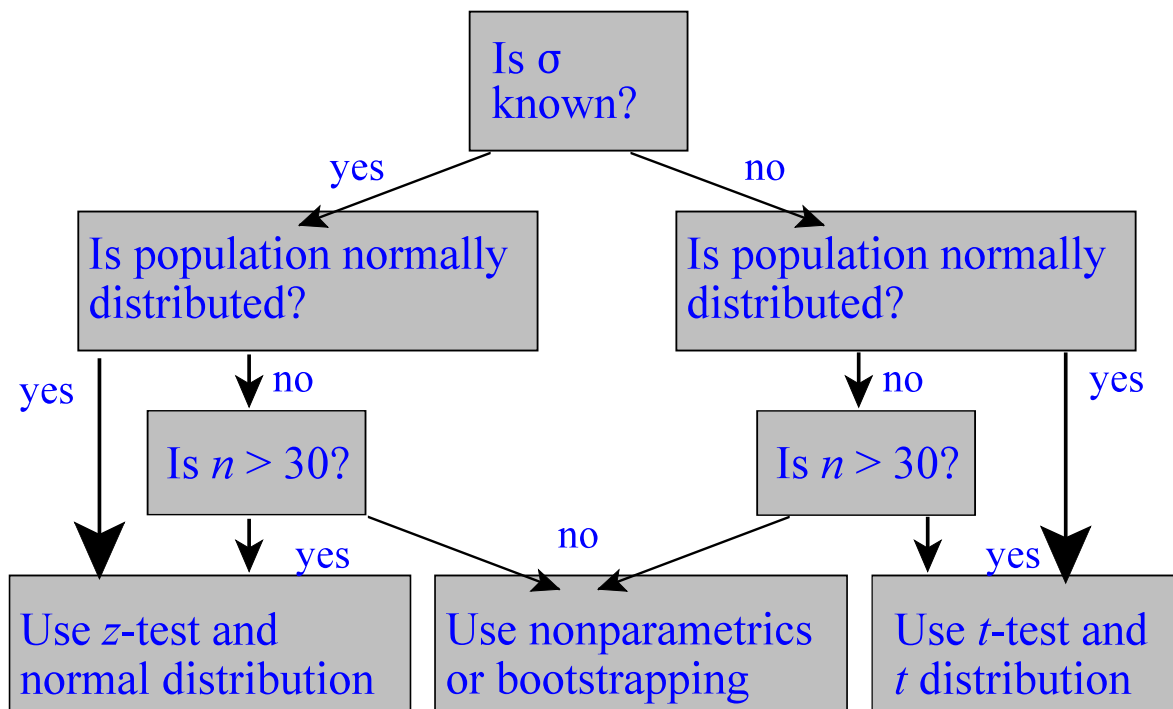
Confidence Intervals when σ is Unknown and n is Small

- Data must be randomly sampled and independent.
- If data are normally distributed and σ is **not known**, which is often the case, use sample standard deviation, s , and use the t -distribution with $n-1$ degrees of freedom.
- If **not** normally distributed use t -distribution as long as $n > 30$.

Formula for Computing Confidence Intervals for Small Sample Sizes

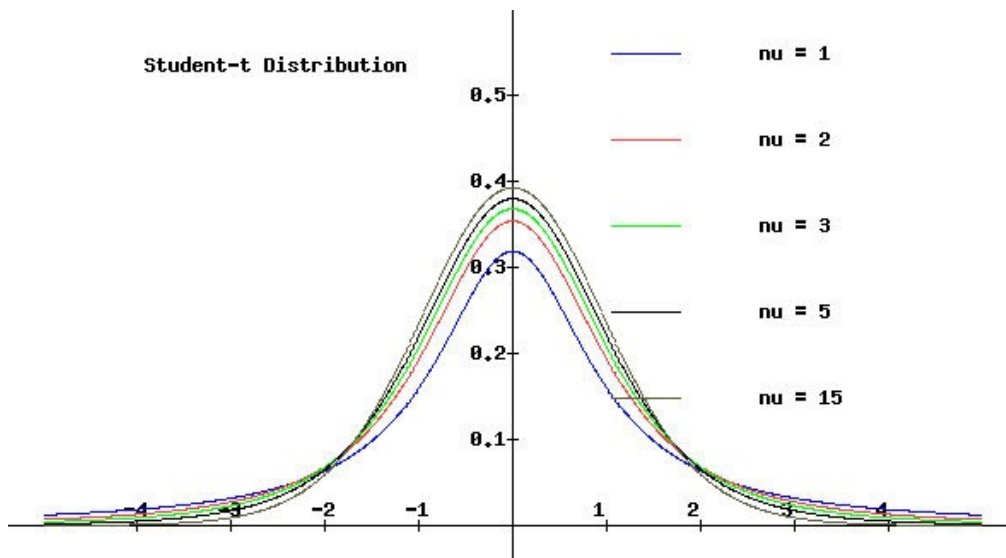
$$\bar{X} - t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right) < \mu < \bar{X} + t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right)$$

Decision Tree for Selecting Statistical Method



t-distribution

- family of curves similar to z-distribution which become more platykurtic (flatter) as sample size decreases
- select distribution using degrees of freedom (df) that is usually $n-1$
- mean is 0, area under curve is 1, is asymptotic to x-axis, symmetrical about mean, SD is greater than 1, approaches shape of z-distribution as sample size increases (very similar when $df=29$ or greater).



Example:

Compute the 95th-percentile confidence interval from a sample of size of 10 which has a standard deviation of 5.00 and a mean of 25.0. (Similar to previous example.)

From t -table with degrees of freedom (df) = $n-1 = 9$:

$$t_{/2} = +/-2.262$$

(Use columns labelled "area in two tails" or "two tails".)

Therefore,
$$E = 2.262 \times \frac{5.00}{\sqrt{10}} = 3.58$$

Confidence interval = CI = $25.0 \pm 3.58 = 21.4$ to 28.6

Sample Size Estimation

Minimum Sample Size for Interval Estimate of Population Mean:

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

where n is sample size, σ is the population standard deviation and E is the maximum error of estimate. Note, always “round up” to the next highest integer when there is a fraction.

When σ is unknown it may be estimated from the sample standard deviation, s .

Example:

Calculate the sample size needed to estimate muscle strength from a population that has a standard deviation of 140.0 newtons if you want to be 95% confident and within 50.0 newtons.

$$n = (1.960 \times 140.0 / 50.0)^2 = 30.1$$

You will need a sample size of 31.