INTERVAL ESTIMATION

Interval estimation refers to the use of sample data to estimate the range of values within which a population parameter is expected to lie. Confidence intervals are used.

A Confidence Interval is a specific interval/range of values within which the parameter is expected to fall, with a certain degree of confidence. Confidence intervals are preferred to point estimates, because confidence intervals indicate (a) the precision/accuracy of the estimate and (b) the uncertainty of the estimate.

A confidence interval is defined by the ***sample statistic*** $\pm$ ***margin of error***.

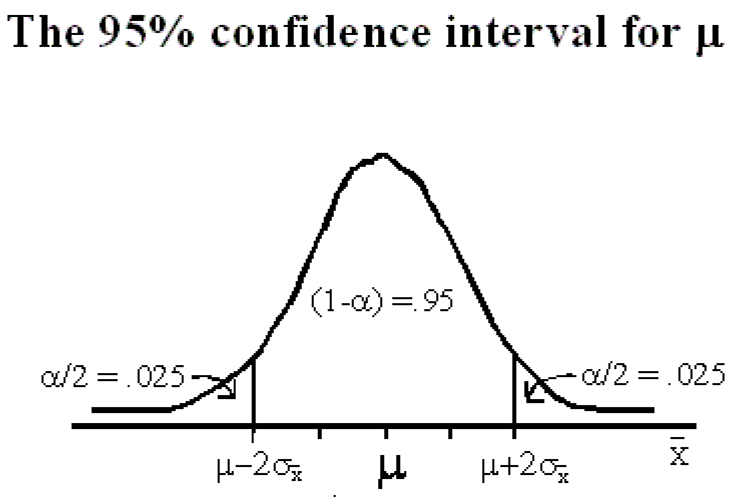
A confidence interval consists of three parts.

1)A confidence level.

2)A statistic/point estimate.

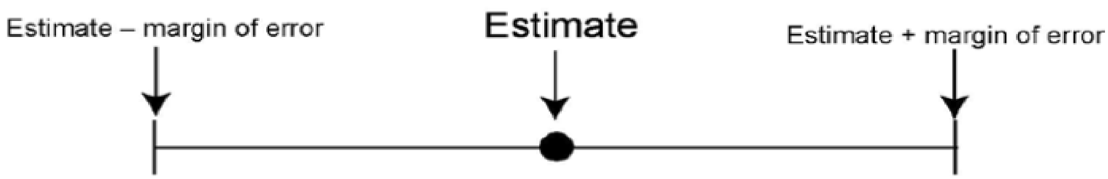
3)A margin of error/maximum error of estimate.

Confidence Level: This is the probability that the interval estimate will contain the parameter. The confidence level describes the likelihood that a particular sampling method will produce a confidence interval that includes the true population parameter. **{Confidence level = (1−α)100%}.** A 95% confidence level means that 95% of the intervals contain the true population parameter or that one is 95% confident that the true population parameter lies in the interval; a 90% confidence level means that 90% of the intervals contain the population parameter or that one is 90% confident that the true population parameter lies in the interval; and so on.



Statistic: A statistic that is used here is the point estimate obtained from the sample data.

Margin of Error: In a confidence interval, the range of values above and below the sample statistic is called the margin of error.



Margin of error = Critical value x Standard deviation of the statistic.

or

Margin of error = Critical value x Standard error of the statistic.

The lower end of the confidence interval is the **lower confidence limit (LCL)**. The upper end is the **upper confidence limit (UCL)**.

N.B. Confidence width is the difference between the upper and lower limits of a confidence interval.

Qtn. State and explain 3 factors that determine the margin of error.

Qtn. Explain what happens to the confidence width when;

i) Sample size increases
ii) Confidence level increases
iii) Standard deviation increases

Qtn. What's better between the small and large confidence intervals?

The wider the confidence interval, the more confident one can be that the given interval contains the unknown population parameter. However, narrower confidence intervals with high degree of confidence are preferred.

CONFIDENCE INTERVALS FOR A POPULATION MEAN

Possibly, the most common use of confidence intervals is to provide an estimate of the population mean. The following cases will be examined

- A known population standard deviation
- An unknown population standard deviation

But we take consideration of the sample size.

a) Population Standard Deviation Known

Irrespective of n, CI is given as;

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is the z-value with probability $\alpha/2$ to the right of it.

b) Population Standard Deviation Unknown

i) For n > 30, replace σ with s, the sample standard deviation in the formula.

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

ii) For n < 30,

CI is given as;

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where t has n-1 degrees of freedom.

Example 1:

Suppose that we check for clarity in 50 locations in Lake Tahoe and discover that the average depth of clarity of the lake is 14 feet. Suppose that we know that the standard deviation for the entire lake's depth is 2 feet. What can we conclude about the average clarity of the lake with a 95% confidence level?

Example 2:

Ten randomly selected people were asked how long they slept at night. The mean time was 7.1 hours, and the standard deviation was 0.78 hour. Find the 95% confidence interval of the mean time. Assume the variable is normally distributed.

Exercise:

Suppose a student measuring the boiling temperature of a certain liquid observes the following readings (in degrees Celsius) 102.5,101.7,103.1,100.9,100.5, and 102.2. What is the confidence interval for the population mean at a 95% confidence level?

Sample Size Requirements

One of the questions a statistician often faces is "How much data should be collected?" Collecting too much data may be a waste of resources, and collecting too little data renders an estimate too imprecise to be useful.

To address the question of sample size requirements, let E represent the **margin of error** of a confidence interval and $z\alpha/_2$ be a critical value obtained from a given confidence level. Assuming that the population standard deviation σ of the variable is known or has been estimated from a previous study,

Sample size, $n = \left(\frac{z\alpha/_2 * \sigma}{E}\right)^2$

Example:

A scientist wishes to estimate the average depth of a river. He wants to be 99% confident that the estimate is accurate within 2 feet. From a previous study, the standard deviation of the depths measured was 4.33 feet. What should be the sample size that a scientist should use?

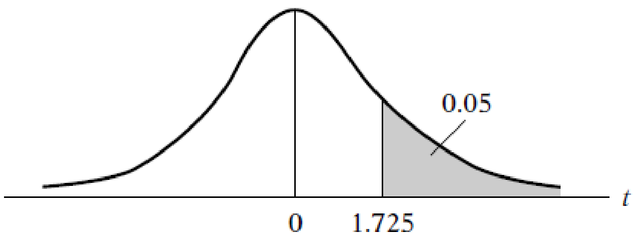
PERCENTAGE POINTS OF THE t DISTRIBUTION

Example

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ for df = 20

$\Pr(|t| > 1.725) = 0.10$



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Note: The smaller probability shown at the head of each column is the area in one tail; the larger probability is the area in both tails.

Source: From E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians,* vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966. Reproduced by permission of the editors and trustees of *Biometrika.*