Pharmaceutical statistics

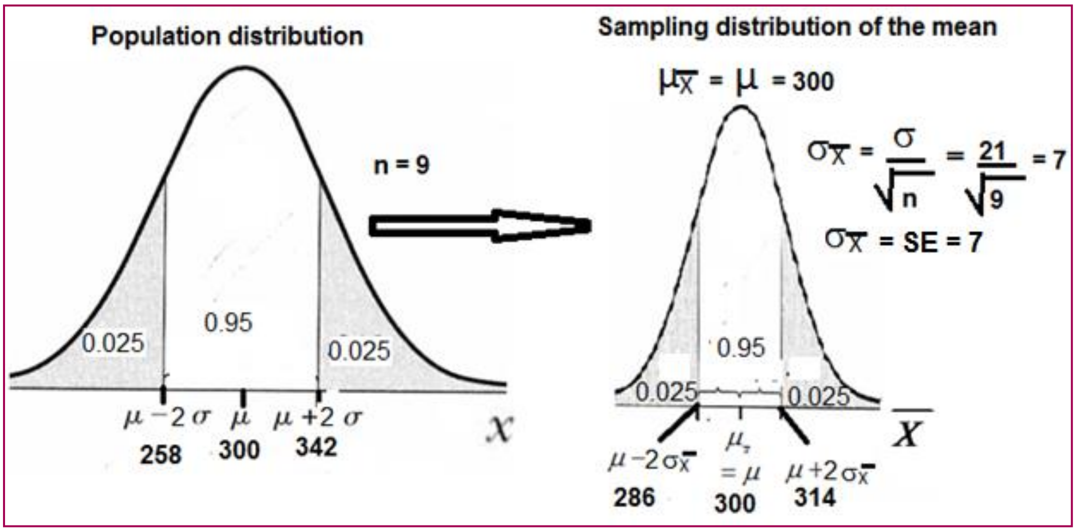
2025-2024

Dr. Rola Aqeel

ARKAN
ACADEMY

Confidence Interval for Population Mean

- A **Confidence Interval (CI) for a Population Mean** is a range of values that is likely to include the true average (mean) of a population. It's used when you want to estimate the population mean based on data from a sample.
  - Population mean (μ) is unknown.
  - Population variance or standard deviation is known or can be fairly estimated.
  - Suppose that the mean (μ) a population of normal distribution is known or estimated as **300** and its standard deviation **(σ)** is **21.** All possible samples **(n=9)** were taken from a population, and the distribution of the sample mean (X) was obtained as given below.



- ✓ **95%** of all possible samples would have mean values between 286 and 314 or the probability a single sample to given mean value between 286 and 314 is 0.95.
- ✓ **5%** of the samples would have mean values higher than 314 or lower than 286.

- Let us construct intervals for the following sample means by adding $2\sigma_{\bar{X}}$ or 2SE to and subtracting $2\sigma_{\bar{x}}$ or 2SE from the sample means as: $\bar{x} \pm 2\sigma_{\bar{x}}$ **OR** $\bar{x} \pm 2SE$

  - ✓ Any sample within 2 standard deviations around the mean of the distribution of the sample mean would give an interval that would capture the population mean.
  - ✓ The probability that a random sample drawn from the population would give an interval that overlaps with population mean is **0.95**.

Sample mean	Interval	Is μ (300) within interval
286	**272-300**	YES
294	280-308	YES
300	286-314	YES
307	293-321	YES
314	300-328	YES

  - ✓ Any sample with mean outside 2 standard deviations around the mean of the distribution of the sample mean (in the upper or lower tail) gives an interval that won't capture the population mean.
  - ✓ The probability that a random sample drawn from the population gives an interval that does not overlap with population mean is **0.05**.

Sample mean	Interval	Is μ (300) within interval
284	**270-298**	No
280	266-294	No
278	264-296	No
316	302-330	No
318	304-332	No
320	306-334	No

- **Explanation points:**
  - The probability that a random sample would have a mean within $2\sigma_{\bar{x}}$ or 2SE of the distribution of the sample mean is 0.95. This the same as what is the percent of samples of all possible samples that give means within $2\sigma_{\bar{x}}$ (2SE) of the distribution of sample mean.
  - For these samples if we construct intervals for each sample mean as $\bar{x} \pm 2\sigma_{\bar{x}}$ **or** $\bar{x} \pm 2SE$, 95% of the intervals would capture or overlap with the population mean (μ).
  - For any random sample we are 95% confident that $\bar{x} \pm 2\sigma_{\bar{x}}$ ($\bar{x} \pm 2SE$) as interval captures the population mean (μ). Thus, the ***confidence level*** is 0.95 and this constructed interval is called ***confidence interval.***
  - The probability that a random sample has a mean outside $\mu \pm 2\sigma_{\bar{x}}$ ($\mu \pm 2SE$) of the distribution of the sample mean is 0.05 **(α)**, because area of the upper tail = area of the lower tail = (α/2) = 0.025.
  - This the same as what is the percent of all possible samples which give means higher than $\mu + 2\sigma_{\bar{x}}$ or lower than $\mu - 2\sigma_{\bar{x}}$ for such samples if you constructed interval for each sample mean as $\bar{x} \pm 2\sigma_{\bar{x}}$ **($\bar{x} \pm 2SE$)**, the confidence intervals will not capture or overlap with the population mean (μ).

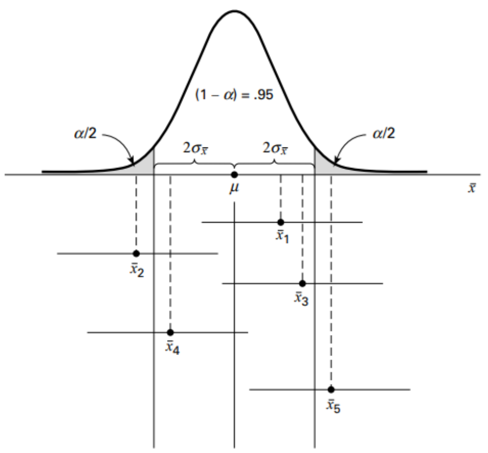


FIGURE 6.2.1 The 95 percent confidence interval for μ.

  - Accordingly, there is 5% (α) chance that a random sample will not contain the population mean within its confidence interval at confidence level of 0.95. Such samples are unusual samples and poorly estimate the population means.
  - **α is called Type 1 error.**
  - α =1- (CL)confidence level or (CL)confidence level =**1-α**

- **Summary:** when a sample is taken from a population and its mean ($\bar{x}$) is to be used to estimate the population means. There is two chances or probabilities as follows:
  - The sample mean is within certain standard deviations around the population mean in the distribution of sample mean at certain level of confidence. E.g. within $2\sigma_{\bar{x}}$ or confidence level of 0.95. At this level the probability the sample mean to be within $2\sigma_{\bar{x}}$ around the population mean is 0.95. If this was the real case, then fair estimation of the population mean is obtained.
  - The sample happened to be one of those rare samples, of which its confidence interval does not capture the population mean, and thus it poorly estimates the mean of the population.

- **How confidence interval is employed and interpreted**
  - For a population that has **unknown mean (μ)**, but we know or have good estimate of its variance or standard deviation **(σ is known)** we do the following.
  - A sample of certain size (n) is drawn.
  - The mean value of the sample ($\bar{x}$) is calculated and used to estimate the population mean (μ) by the construction of a confidence interval.
  - Using confidence level of 0.95, the confidence interval is constructed by adding and subtracting 2 standard deviations of the distribution of the sample mean ($2\sigma_{\bar{x}}$ or 2SE) to and from the sample mean as: $\bar{x} \pm 2\sigma_{\bar{x}}$ **or** $\bar{x} \pm 2SE$
  - With 95% confidence, we can state that the true population mean is estimated to be between the lower limit and upper limit of the confidence interval i.e. is any single value in this range.

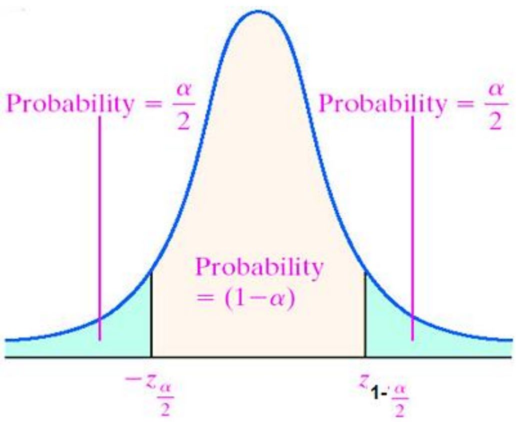
- There is 5% chance (α or type I error) that confidence interval did not accurately estimate the population mean as the probability that the population mean is outside the confidence interval. If the sample came from the upper tail area of the distribution of the sample mean, the confidence interval would overestimate the population mean (P = 0.025) and if it was from the lower tail area, the population mean would be underestimated (P = 0.025). In either case we would commit Type I error with total probability of 0.05.

★ Example:

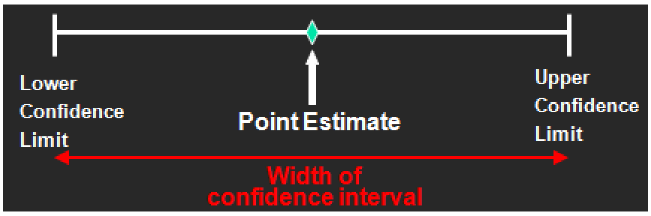
- ✓ Suppose a researcher, interested in obtaining an estimate of the average level of some enzyme in a certain human population, takes a sample of **10** individuals, determines the level of the enzyme in each, and computes a sample mean of ($\bar{x}$) = **22**. Suppose further it is known that the variable of interest is approximately normally distributed with a variance of **45**. We wish to estimate μ at 0.95 confidence level.
- ✓ We are 95% confident that the interval from 17.76 to 26.24 does contain the true value of the population mean of enzyme level. OR the population mean falls between 17.76 to 26.24 or is any single value in this interval range (≥17.76 and ≤26.24)

$$\bar{x} \pm 2\sigma_{\bar{x}}$$
$$22 \pm 2\sqrt{\frac{45}{10}}$$
$$22 \pm 2(2.1213)$$
$$17.76 - 26.24$$

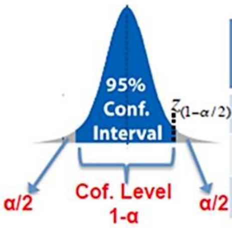
- **General formula for Confidence Interval for μ (σ Known)**
  - Assumptions:
    - ✓ Population standard deviation σ is known
    - ✓ Population is normally distributed
    - ✓ If population is not normal, use large sample (n > 30)
  - $\bar{X} \pm z_{(1-\alpha/2)} * \sigma_{\bar{x}}$
  - $\bar{X} \pm z_{(1-\alpha/2)} * \frac{\sigma}{\sqrt{n}}$
  - $\bar{X} \pm z_{(1-\alpha/2)} *$ SE
    (Estimator ± (reliability coefficient) * (standard error))
  - $Z_{\alpha/2}$ is the normal distribution critical value for a probability of α/2 in each tail.
  - Where $Z_{(1-\alpha/2)}$ is the value of z to the left of which lies 1-α/2 and to the right of which lies α/2 of the area under its curve.



  - Width of CI is = 2 * $z_{(1-\alpha/2)}$ * SE



Common Levels of Confidence!!



α	Conf. Level 1-α	$z_{(1-\alpha/2)}$
10%(0.1)	90% (0.9)	$z_{0.95} = 1.65$
5%(0.05)	95% (0.95)	$z_{0.975} = 1.96$
1%(0.01)	99% (0.99)	$z_{0.995} = 2.58$

$$90\%\,CI\,of\ \mu = \overline{X} \pm 1.645\left(\sigma_x/\sqrt{n}\right)$$

$$95\%\,CI\,of\ \mu = \overline{X} \pm 1.96\left(\sigma_x/\sqrt{n}\right)$$

$$99\%\,CI\,of\ \mu = \overline{X} \pm 2.576\left(\sigma_x/\sqrt{n}\right)$$

Any z can be used (from table)

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.9	.97128	.97193	.97257	.97320	.97381	.97441	$\boxed{.97500}$	.97558	.97615	.97670

★ Example 2:

- A physical therapist wished to estimate, with 99% confidence, the mean maximal strength of a particular muscle in a certain group of individuals. He is willing to assume that strength scores are approximately normally distributed with a variance of 144. A sample of 15 subjects who participated in the experiment yielded a mean of 84.3.
- Solution:

✓ The z value corresponding to a confidence coefficient of 0.99 is found in Appendix Table to be 2.58. This is our reliability coefficient.

✓ The standard error is $12/\sqrt{15} = 3.0984$

✓ Our 99% confidence interval for μ is: $84.3 \pm 2.58(3.0984)$

$$84.3 \pm 8$$

$$76.3 - 92.3$$

✓ Width of confidence interval = 8×2=16

✓ Point estimate is 84.3

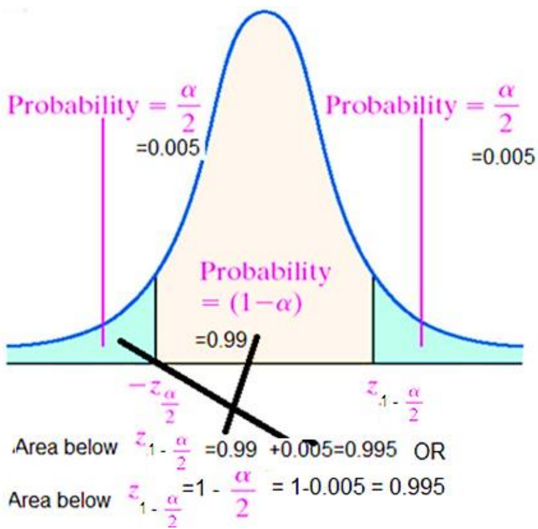
✓ Lower confidence limit is 76.3

✓ Upper confidence limit is 92.3

✓ α =1-confidence level (0.99) = 0.01

✓ α/2 = 0.01/2 = 0.005

✓ $Z_{1-\alpha/2}$ of 0.99 confidence is $z_{0.995}$ with areas above of 0.005 and below of 0.995, which is 2.58 (average of 2.57 and 2.58) as shown in the table below



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60

ARKAN
ACADEMY

علم في كل مكان

Arkan academy

Arkanacademy

www.arkan-academy.com

+962 790408805