innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

Chapter 7: Statistical Intervals Based on a Single Sample

Sumanta Pasari

sumanta.pasari@pilani.bits-pilani.ac.in

BITS Pilani
Pilani Campus

Sample Proportion

The statistic that estimates the parameter p, a proportion of a population that has some property, is the sample proportion

$$\hat{p} = \frac{\text{number in sample with the trait (success)}}{\text{sample size}} = \frac{X}{n}$$

Properties:

(i) As the sample size increases (n large), the sampling distribution of $\hat{p}$ becomes approximately normal (WHY?)

(ii) The mean of $\hat{p}$ is p, and variance of $\hat{p}$ is $\frac{p(1-p)}{n}$ (WHY?)

(iii) Can we get a point estimators of p? (See Ex. 6.15, page no. 258)

Interval Estimation

- A point estimate cannot be expected to provide the exact value (close value) of the population parameter.
- **Usually**, an interval estimate can be obtained by adding and subtracting a margin of error to the point estimate. Then,

Interval Estimate = Point Estimate + / – Margin of Error

- Interval estimation provides us information about **how close** the point estimate is to the value of the parameter.
- Why we use the term ***confidence interval?***

Interval (CI) Estimation

– Instead of considering a statistic as a point estimator, we may use *random intervals* to trap the parameter.

– In this case, the end points of the interval are RVs and we can talk about the probability that it traps the parameter value.

Confidence Interval : A $100(1-\alpha)\%$ confidence interval for a parameter is a **random interval** $[L_1, L_2]$ such that

$$P[L_1 \leq \theta \leq L_2] = 1-\alpha$$, regardless the value of θ.

Theorem 7.1: Interval estimation for μ: σ known

Let $X_1,\ X_2,\ \cdots, X_n$ be a random sample from a normal population with mean μ (unknown) and the variance σ^2 (known). Then, we know,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)$$

Taking two points $\pm z_{\alpha/2}$ symmetrically about the origin, we get

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) = 1-\alpha$$

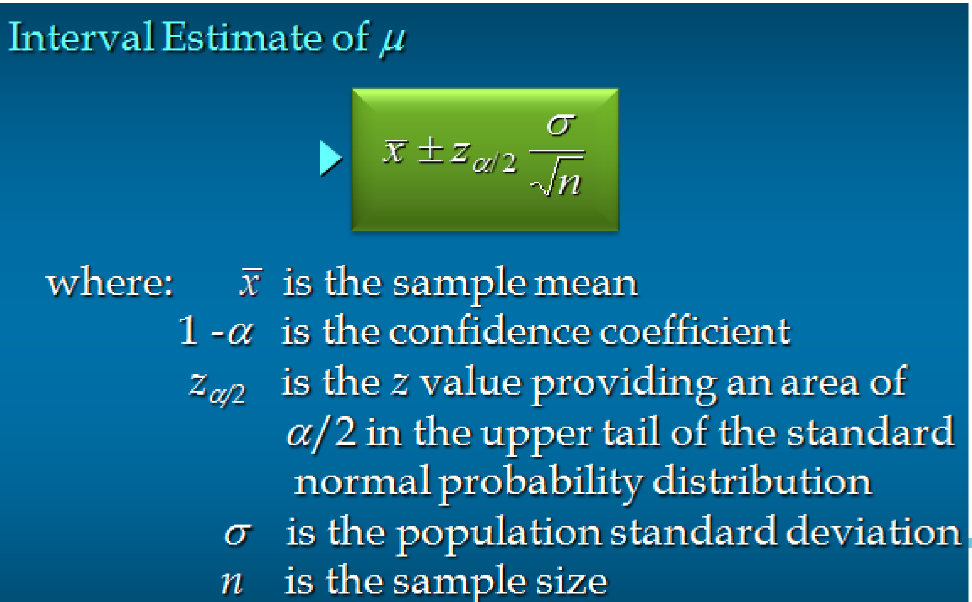
Here $(1-\alpha)$ is known as confidence level, and α is the level of significance.

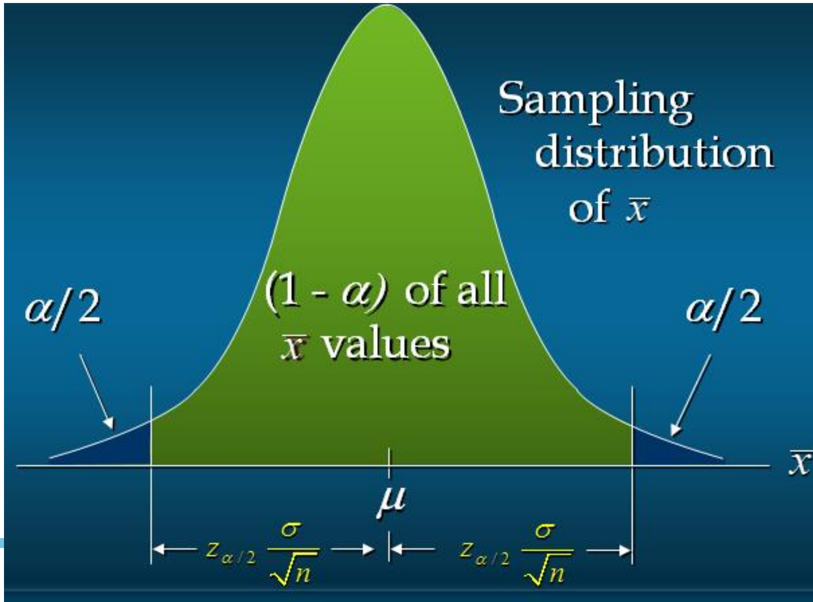
Interval estimation for μ: σ known

$$P\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)=1-\alpha$$

Hence, the confidence interval for population mean μ having confidence level $100\times(1-\alpha)\%$ is given as $\left(\bar{x}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{x}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$.

The endpoints of the confidence interval is called confidence limits.





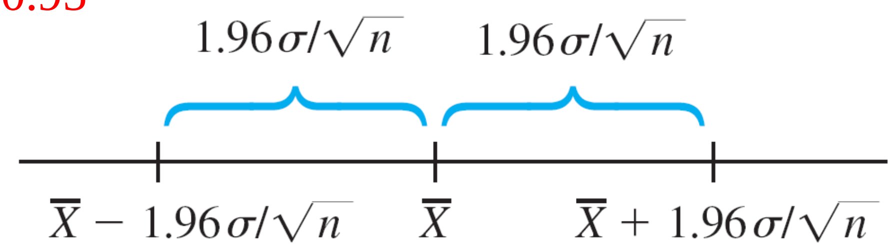
Interval estimation for μ: σ known

Most commonly used confidence levels:

Confidence Level	α	$\alpha/2$	Table Look-up Area	$z_{\alpha/2}$
90%	.10	.05	.9500	1.645
95%	.05	.025	.9750	1.960
99%	.01	.005	.9950	2.576

Hence, 95% CI for μ is given as $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$.

That is, $P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$



Practice Problems

Ex.1. The mean of a sample size 50 from a normal population is observed to be 15.68. If the s.d. of the population is 3.27, find (a) 80% (b) 95%, (c) 99% confidence interval for the population mean. Can you find out the respective margin of errors? What is the length of CI for each case?

Sol. (b) First check the two assumptions : (i) normality (ii) σ known

Step 1: Here $n = 50$, $\bar{x} = 15.68$, $\sigma = 3.27$, and $\alpha = 0.05$. We need CI for μ.

Step 2: As $\alpha = 0.05$, we need to find $z_{\alpha/2}$ such that $P\left(Z < z_{\alpha/2}\right) = 0.975$.

From cumulative normal distribution table, we see $z_{\alpha/2} = 1.96$.

Step 3: The CI for $\mu(\sigma \text{ known})$ is $\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{0.025}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{0.025}\right) = (14.77, 16.59)$

Interval estimation for μ: σ known

Confusions and confusions?? ☹ ☹

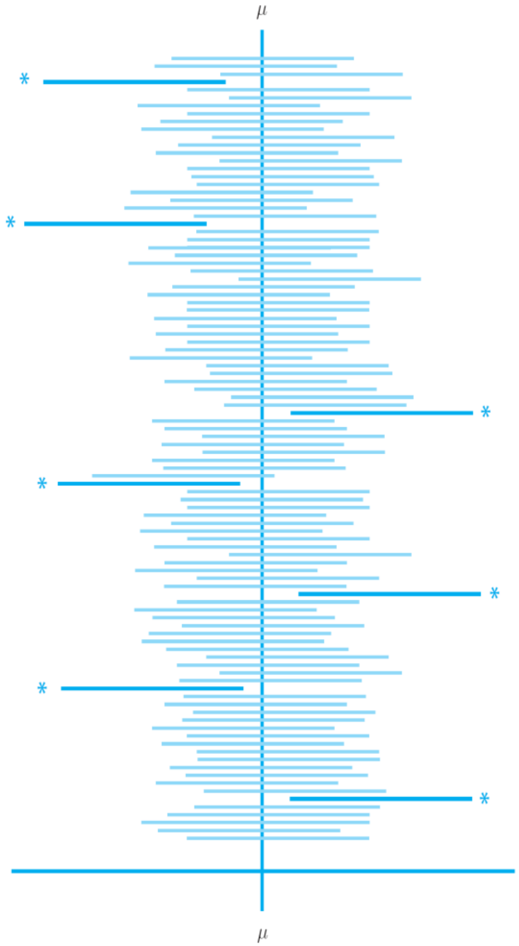
In the above example, we found 95% CI for μ is (14.77, 16.59). This means, the unknown μ lies within the fixed interval with probability 0.95. That is,

P [μ lies in (14.77, 16.59)]= 0.95 **– right?**

If not, then what is the interpretation of "95% confidence"?

- Long run relative frequency?
- A single replication/realization of random interval is not enough! Not satisfactory, at least.

95% CIs for population mean



One hundred 95% CIs (asterisks identify intervals that do not include μ).

Practice Problems

HW 1. Studies have shown that the random variable X, the processing time required to do a multiplication on a new 3-D computer, is normally distributed with mean μ and standard deviation 2 microseconds. A random sample of 16 observations is to be taken

(a) These data are obtained

42.65	45.15	39.32	44.44
41.63	41.54	41.59	45.68
46.50	41.35	44.37	40.27
43.87	43.79	43.28	40.70

Based on these data, find an unbiased estimate for μ.

(b) Find a 95% confidence interval for μ.

Confidence Level, Precision, and Sample Size

Ex.2. Extensive monitoring of a computer time-sharing system has suggested that response time to a particular editing command is normally distributed with standard deviation 25 millisec.

A new operating system has been installed, and we wish to estimate the true average response time μ for the new environment.

Assuming that response times are still normally distributed with $\sigma = 25$, what sample size is necessary to ensure that the resulting 95% CI has a width of (at most) 10?

Confidence Level, Precision, and Sample Size

The sample size n must satisfy

$$10 = 2 \cdot (1.96)(25/\sqrt{n})$$

Rearranging this equation gives

$$\sqrt{n} = 2 \cdot (1.96)(25)/10 = 9.80$$

So

$$n = (9.80)^2 = 96.04$$

Since n must be an integer, a sample size of 97 is required.

Impracticality of Assumptions in CI

In practice, we usually face mainly two problems in application of previous C.I. formula.

- What if the population is not normal? **(large sample size is needed) - *Can we take help from CLT?***

- What if the population variance is unknown?

7.2: Large Sample CI for μ

Let $X_1,\ X_2,\ \cdots, X_n$ be a random sample from a large sample $(n \geq 40)$ with mean μ (unknown) and sample variance S^2. Then, we know,

$$\frac{\bar{X}-\mu}{\left(\frac{S}{\sqrt{n}}\right)} \sim N(0,1) \quad \Rightarrow \quad P\left(\bar{X}-\frac{S}{\sqrt{n}} z_{\alpha/2} < \mu < \bar{X}+\frac{S}{\sqrt{n}} z_{\alpha/2}\right) = 1-\alpha$$

Hence, the large-sample confidence interval for population mean μ having confidence level $100 \times (1-\alpha)\%$ is approximately given as $\left(\bar{x}-\frac{s}{\sqrt{n}} z_{\alpha/2},\ \bar{x}+\frac{s}{\sqrt{n}} z_{\alpha/2}\right)$, $n \geq 40$ is needed.