# Lecture 28

## One-Sample Estimation Problems

In the last lecture we have studies about the sampling distributions of sample mean and sample varience. In this lecture, we give a brief introduction about the purpose of **Statistical Inference**. We follow this by discussing the Problem of **Interval Estimation or Confidence Intervals**.

## 1 Statistical Inference

It is a branch of Statistics which consists of those methods by which one can make inferences about a populations. Many of the real world situations are stochastic or probabilistic in nature. So, a population may contain some unknown parameters whose values we may not be knowing. In order to infer about the population parameters, we take a random sample from the population and then try to infer it. For example,
1. Amount of rain fall during a monsoon season is a random variable. It's not sure that how much rain will fall? How much rain will fall tomorrow?
2. Time taken by patients to get cured by a disease while going under a particular treatment? or say the effect of medicine on different patients is a study of our concern.
3. Number of persons in a service queue or a ticket counter. Some day we may have large people standing on a queue and some other day there may be less.
Some typical scientific problems may be adressed in this connection namely, **Quality of production, Avarage monthly salary, How much increase in temparature will be there globally, effectiveness of biochemical new drug** etc. To all these we can apply certain methods to infer about these populations. Broadly, Statistcal Inference is divided into the following categories depending upon the nature of the problem.
1. **Confidence Intervals.**
2. **Point Estimation.**
3. **Testing of Statistical Hypothesis.** We will study in detail about this in our future classes.

## 2 Confidence Intervals

An interval estimate of a population parameter $\theta$, is an interval of the form $\hat{\theta_L} < \theta < \hat{\theta_U}$, where $\hat{\theta_L}$ and $\hat{\theta_U}$ depend upon the random sample

$X_1, X_2, \ldots, X_n$. Moreover, we would be interested to estimate the confidence interval of $\theta$ with a confidence level $(1-\alpha)$ i.e

$$P(\hat{\theta_L} < \theta < \hat{\theta_U}) = 1 - \alpha, \quad 0 < \alpha < 1. \tag{1}$$

The above interval is called the $100(1-\alpha)\%$ C.I for $\theta$. $(1-\alpha)$ is called the confidence level or degree of confidence. Normally, we take $1-\alpha = 90\%, 95\%, 99\%$ etc. For example, the average life of a TV lies within an interval.

# 3 Estimating The Mean $\mu$ Of A Normal Population When $\sigma^2$ Is Known

Here, we try to find out $\hat{\mu_L}$ and $\hat{\mu_U}$ such that

$$P_{\sigma^2=known}(\hat{\mu_L} < \theta < \hat{\mu_U}) = 1 - \alpha$$

Let $X_1, X_2, \ldots, X_n$ be a random sample taken from $\mathcal{N}(\mu, \sigma^2)$ distribution where $\sigma^2$ is known. From the Normal graph, we see that

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha \tag{2}$$

where $Z = \sqrt{n}\frac{(\overline{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$. So, from eq.(2), we have

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

$$\Rightarrow \quad P\left( \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

So, we have

$$\begin{aligned} \hat{\theta_L} &= \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \\ \hat{\theta_U} &= \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \end{aligned}$$

where $Z_{\alpha/2}$ is such that $P(Z > Z_{\alpha/2}) = \frac{\alpha}{2}$.

So, $\left( \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right)$ is a $100(1-\alpha)\%$ C.I for $\mu$.

In the next class we would like to derive an interval estimation for $'\mu'$ when $\sigma^2$ is not known, i.e, we wish to find

$$P(\hat{\theta_L} < \theta < \hat{\theta_U}) = 1 - \alpha.$$

Next, we will solve some problems in this connection.
**Question:-** The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 gm/ml. Find a 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume normality and $\sigma = 0.3$.

**Answer:-** The point estimate of $\mu = \overline{x} = 2.6$, $(1 - \alpha = 0.95, 0.99)$
$\implies \alpha = 0.05, \alpha/2 = 0.025$ or $\alpha = 0.01, \alpha/2 = 0.005$.
So, $Z_{0.025} = 1.96$ and $Z_{0.005} = 2.575$.
So, the 95% and 99% C.I for $\mu$ is thus given by

$$\begin{aligned} & 2.6 - (1.96)\Big(\frac{0.3}{\sqrt{36}}\Big) < \mu < 2.6 + (1.96)\Big(\frac{0.3}{\sqrt{36}}\Big) \\ \text{and} \qquad & 2.6 - 2.575\Big(\frac{0.3}{\sqrt{36}}\Big) < \mu < 2.6 + 2.575\Big(\frac{0.3}{\sqrt{36}}\Big) \end{aligned}$$

**Theorem 3.1** *If $\overline{X}$ is used as an estimate of $\mu$, we can be $100(1-\alpha)\%$ confident that the error will not exceed $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.*

**Proof:-** Check

$$\begin{aligned} P\Big(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) &= 1 - \alpha \qquad \text{or} \\ P\Big(0 < \mid \overline{X} - \mu \mid < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) &= 1 - \alpha \end{aligned}$$

The term $\Big[e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big]$ is called error in estimating $'\mu' \Rightarrow n = \Big(\frac{Z_{\alpha/2}\sigma}{e}\Big)^2$.

**Example 1** *How large a sample is required if we want to be 95% confident that our estimate of $\mu$ in the above example is off by less than 0.05.*

**Answer:-** $\sigma = 0.3$ and $e = 0.05$, hence $n = \Big[\frac{(1.96)(0.3)}{0.05}\Big]^2 = [138.3] = 139$.
Therefore, we can be 95% confident that a random sample of size 139 will provide an estimate $\overline{X}$ differs from $\mu$ by an amount less than 0.05.

**Example 2** *Q-2 An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 40 hours. If a sample of 30 bulbs has an average life of 780 hours, find a* 96% *confidence interval for the population mean of all bulbs produced by this firm.*

**Answer:-**Here $n = 30$, $\overline{x} = 780$ and $\sigma = 40$. Also $Z_{0.02} = 2.054$, so a 96% confidence interval for the mean of the population is given by $765 < \mu < 795$.

**Example 3** *Q-3 Many cardiac patients wear an implanted pacemaker to control their heartbeat. A plastic connector module mounts on the top of the pacemaker. Assuming a standard deviation of 0.0015 inch and an approximately normal distribution, find a* 95% *confidence interval for the mean of the depths of all connector modules made by a certain manufacturing company. A random sample of 75 modules has an average depth of 0.310 inch.*

**Answer:-** Here $n = 75$, $\overline{x} = 0.310$ and $\sigma = 0.0015$. Also $Z_{0.025} = 1.96$. A 95% confidence interval for the mean of the population is given by $0.3097 < \mu < 0.3103$.