Part VII

Unit IV: Statistical Inference (Parametric)

Population and Sample:

Let X be a random variable connected with a random experiment E. Now, if the random experiment is performed once, the result is evidently an event point and corresponding to this event point we will get a value of the random variable X. So that any particular performance of E gives a value of X and hence a sequence of trials of E will give a sequence of values of X. If the random experiment E is repeated infinitely large number of times under uniform conditions, then we will get an infinite sequence of observed values of X, the totality of which will be called the population of the random variable X.

Population and Sample:

Let X be a random variable connected with a random experiment E. Now, if the random experiment is performed once, the result is evidently an event point and corresponding to this event point we will get a value of the random variable X. So that any particular performance of E gives a value of X and hence a sequence of trials of E will give a sequence of values of X. If the random experiment E is repeated infinitely large number of times under uniform conditions, then we will get an infinite sequence of observed values of X, the totality of which will be called the population of the random variable X.

Now if the random experiment is repeated a finite number of times, say, n times which yields a sequence of n observed values $x_1, x_2, x_3, ..., x_n$ of X which is called the realizations of a sample of size n drawn from the population of X. If now we perform another sequence of n trials of E, the earlier sequence will not be generally reproduced, but we shall get a different sequence of values of X: $x'_1, x'_2, ...x'_n$. Thus if different samples of size n are repeatedly drawn under uniform conditions, the sets of observed values of the random variable will fluctuate at random. In this sense, the sample is said to be a random sample and the individual values $x_1, x_2, ..., x_n$ are called the sample values.

This randomness of the sample may be described as follows:

1^{st}	sample value	x_1	may be regarded as the observed value of a r.v.	X_1
2^{nd}	,,	x_2	,,	X_2
- -	- - - -	- -	- -	- -
n^{th}	,,	x_n	,,	X_n

This randomness of the sample may be described as follows:

1^{st}	sample value	x_1	may be regarded as the observed value of a r.v.	X_1
2^{nd}	,,	x_2	,,	X_2
- -	- - - -	- -	- -	- -
n^{th}	,,	x_n	,,	X_n

But all the sample values $x_1, x_2, ..., x_n$ are, in fact, observed values of the parent random variable X, and as such the random variables $X_1, X_2, ..., X_n$ must have the same distribution, viz. that of X, i.e., of the population. Moreover, since the sample values are given by repetitions of the random experiment E under uniform conditions, it follows that the random variables $X_1, X_2, ..., X_n$ should be mutually independent. To sum up, $X_1, X_2, ..., X_n$ are said to be a random sample of size n taken from the population of X. The sample values $x_1, x_2, ..., x_n$ are then regarded as the observed values of the random variables $X_1, X_2, ..., X_n$ which are iid with common distribution of X.

Example:

Let E: random experiment of throwing a die and X: number on the die. If we imagine that the die is repeatedly thrown an infinite large number of times, we will get an infinite sequence of observed values of X which forms the population of the random variable. If now the die is actually thrown 100 times, a sequence of 100 values of X is obtained, which is then a sample of size 100 from the population.

Example:

Let E: random experiment of throwing a die and X: number on the die. If we imagine that the die is repeatedly thrown an infinite large number of times, we will get an infinite sequence of observed values of X which forms the population of the random variable. If now the die is actually thrown 100 times, a sequence of 100 values of X is obtained, which is then a sample of size 100 from the population.

If the random variables $X_1, X_2, ..., X_n$ are iid, then these random variables constitute a random sample of size n from the common distribution of X i.e., F. Their joint CDF is given by

$$F^*(x_1, x_2, ..., x_n) = F(x_1)F(x_2)...F(x_n).$$

Let $T = T(X_1, X_2, ..., X_n)$ be a function of the sample. Then T is called a statistic. For example sample mean $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$, sample variance $S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1} = \frac{\sum X_i^2 - n\overline{X}^2}{n-1}$ – are all statistics.

The probability distribution of the random variable T will be called the sampling distribution of the statistic. We illustrate the reasons why sampling from a population and developing a distribution of these sample statistics would produce a sampling distribution. If we take several samples from a population, the statistics we would compute for each sample need not be the same and most probably would vary from sample to sample.

The probability distribution of the random variable T will be called the sampling distribution of the statistic. We illustrate the reasons why sampling from a population and developing a distribution of these sample statistics would produce a sampling distribution. If we take several samples from a population, the statistics we would compute for each sample need not be the same and most probably would vary from sample to sample.

Suppose our samples each consist of ten 25-years women from a city with a population of one lakh. By computing the mean and s.d. of height for each of these samples, we see that mean and s.d. of each sample would be different. A probability distribution of all these possible means of the samples is a distribution of the sample mean. Statisticians call this a sampling distribution of the mean. We could also have a sampling distribution of proportion. In the remainder, we shall consider the sampling distribution of the mean.

Sampling Distribution of Mean:

The sampling distribution of mean can be partially described by its mean and s.d.

Sampling Distribution of Mean:

The sampling distribution of mean can be partially described by its mean and s.d.

1. The mean of the sampling distribution of mean will equal the population mean regardless of the sample size, even if the population is not normal.

Sampling Distribution of Mean:

The sampling distribution of mean can be partially described by its mean and s.d.

1. The mean of the sampling distribution of mean will equal the population mean regardless of the sample size, even if the population is not normal.
2. As the sample size increases, the sampling distribution of mean will approach- normally, regardless of the shape of the population distribution.

Sampling Distribution of Mean:

The sampling distribution of mean can be partially described by its mean and s.d.

1. The mean of the sampling distribution of mean will equal the population mean regardless of the sample size, even if the population is not normal.
2. As the sample size increases, the sampling distribution of mean will approach- normally, regardless of the shape of the population distribution.

Sampling Distribution of Mean:

The sampling distribution of mean can be partially described by its mean and s.d.

1. The mean of the sampling distribution of mean will equal the population mean regardless of the sample size, even if the population is not normal.
2. As the sample size increases, the sampling distribution of mean will approach- normally, regardless of the shape of the population distribution.

The s.d of the distribution of sample means measures the extent to which we expect the means from the different samples to vary because of this chance error in the sampling process. Rather than say "standard deviation of the distribution of sample means" to describe a distribution of sample means, statisticians refer to the standard error of the mean. Thus, the s.d. of the distribution of a sample statistics is known as the standard error of the statistics.

The standard error of the mean

$$\sigma_{\overline{X}} = \begin{cases} \frac{\sigma}{\sqrt{n}} & \text{(Infinite population)} \\ \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} & \text{(Finite population)} \end{cases}$$

σ: Population s.d.

N: Population size

n: Sample size

The standard error of the mean

$$\sigma_{\overline{X}} = \begin{cases} \frac{\sigma}{\sqrt{n}} & \text{(Infinite population)} \\ \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} & \text{(Finite population)} \end{cases}$$

σ: Population s.d.

N: Population size

n: Sample size

Summary:

If a random sample of size n is taken from a population with mean μ and s.d. σ, then the sample distribution of $\overline{X}$ (sample mean) has a mean equal to population mean i.e., μ and a s.d. $\sigma_{\overline{X}}$ as given above.
Therefore, $Z = \frac{\overline{X}-\mu}{\sigma_{\overline{X}}}$ follows standard normal distribution.

Example:

In a sample of 25 observations from normal distribution with mean 98.6 and s.d. 17.2
(a) What is $P(92 < \overline{X} < 102)$?
(b) Find the corresponding probability given a sample of 36.

Example:

In a sample of 25 observations from normal distribution with mean 98.6 and s.d. 17.2
(a) What is $P(92 < \overline{X} < 102)$?
(b) Find the corresponding probability given a sample of 36.

Example:

Mary Bartel, an auditor for a large credit card company, knows that, on average, the monthly balance of any given customer is $\$112$, and the s.d. is $\$56$. If Marry audits 50 randomly selected accounts, what is the probability that the sample average monthly balance is
(a) Below $\$100$?
(b) Between $\$100$ $\$130$?

Example:

In a sample of 25 observations from normal distribution with mean 98.6 and s.d. 17.2
(a) What is $P(92 < \overline{X} < 102)$?
(b) Find the corresponding probability given a sample of 36.

Example:

Mary Bartel, an auditor for a large credit card company, knows that, on average, the monthly balance of any given customer is $\$112$, and the s.d. is $\$56$. If Marry audits 50 randomly selected accounts, what is the probability that the sample average monthly balance is
(a) Below $\$100$?
(b) Between $\$100$ $\$130$?

Example:

From a population of 125 items with mean of 105 and a s.d. of 17,64 items were chosen.
(a) What is the standard error of the mean?
(b) What is the $P(107.5 < \overline{X} < 109)$?