**Normal Distribution (contd)**

**Recap**

* Normal distribution

**Today**

* Area under the curve
* Sampling Distribution

**Area under Curve**

Practically the entire area under the normal curve is covered between points which are within 3 standard deviations on each side of the mean.

(Diagram for area within +/- 1 SD of the mean) : 68.26%

(Diagram for area within +/- 2 SD of the mean): 95.44%

(Diagram for area within +/- 3 SD of the mean): 99.74%

+1 SD of the mean covers 34.13% of the area. In other words, probability is 0.3413 that the random variable x will assume values between mean and another point x which is 1 SD away from the mean on either side.

Similarly, when Z +/- 2, then the area is 47.72% on each side of the mean and when Z = +/- 3, the area is 49.87% on each side of the mean.

These values are true for all normal curves.

**Exercise**

1. Find the area to the left of Z = +1
2. Find the area to the right of Z = -1.6.

**Exercise**

The IQ score of students is normally distributed with mean of 120 and standard deviation of 20. What proportion of students have:

1. An IQ between 100 and 130
2. An IQ above 140
3. An IQ above 150
4. An IQ between 140 and 150

**Concept of Sampling Distribution**

**Theory**

The major objective of the field of statistical analysis is to know the true or actual values of different parameters of the population.

Ideally we would like to take the entire population into consideration in determining these values. However, it’s not feasible due to cost, time considerations.

Accordingly random samples of a given size are taken from the population and these samples are properly analysed with the belief that characteristics of these random samples represent the similar characteristics of the population from which these samples are taken.

The results obtained from such analysis lead to generalization that are considered to be valid for the entire population.

**Example**

Suppose that a baby-sitter has 5 children under her supervision. The average age of these five children is 6 years.

|  |  |
| --- | --- |
| Child (X) | Age |
| 1 | 2 |
| 2 | 4 |
| 3 | 6 |
| 4 | 8 |
| 5 | 10 |

Now, calculate Mean and Standard Deviation.

Now, take possible random samples of size 2 without replacement from the population. How many such combinations are possible? (nCx)

|  |  |  |
| --- | --- | --- |
| X1, X2 | 2, 4 | 3 |
| X1, X3 | 2, 6 | 4 |
| X1, X4, | 2, 8 | 5 |
| X1, X5 | 2, 10 | 6 |
| X2, X3 | 4, 6 | 5 |
| X2, X4 | 4, 8 | 6 |
| X2, X5 | 4, 10 | 7 |
| X3, X4 | 6, 8 | 7 |
| X3, X5 | 6, 10 | 8 |
| X4, X5 | 8, 10 | 9 |

Let’s consider the grand mean X(doubled-bar) of these ten samples.

Now, let’s make a probability distribution from this data

|  |  |  |  |
| --- | --- | --- | --- |
| Sample Mean | Frequency | Rel Freq | Prob |
| 3 | 1 | 1/10 | .1 |
| 4 | 1 | 1/10 | .1 |
| 5 | 2 | 2/10 | .2 |
| 6 | 2 | 2/10 | .2 |
| 7 | 2 | 2/10 | .2 |
| 8 | 1 | 1/10 | .1 |
| 9 | 1 | 1/10 | .1 |

Grand mean = Total [ Sample Means x Resp. Prob ]

This probability distribution of the sample means is referred to as sampling distribution of the mean.

It’s a probability distribution of all possible sample means of a given size, selected from a population.

**Central Limit Theorem**

It states that regardless of the shape of the distribution of the population, the distribution of the sample means approaches the normal probability distribution as sample size increases.

Question is how many samples are required to approximate the normal distribution?

In practise, 30 or more samples are considered adequate for this purpose.

As we can see from the sampling distribution of means, the grand mean x of the sample means equals population mean. Realistically speaking, it is not possible to take all the samples of size(n) from the population.

In practise only one random sample is taken and the discussion on the sampling distribution is concerned with the proximity of a sample mean to the population mean.

**Application of the CLT**

If we know the grand mean x(double-bar) and the standard deviation of this distribution (standard error of the mean), then we know from the characteristics of the normal distribution that there is a 68.26% chance that a sample selected at random from a given population will have a mean that lies within one standard error of the population mean.

Similarly, this chance increases to 95.44% that the sample will lie within two standard error of the population mean.