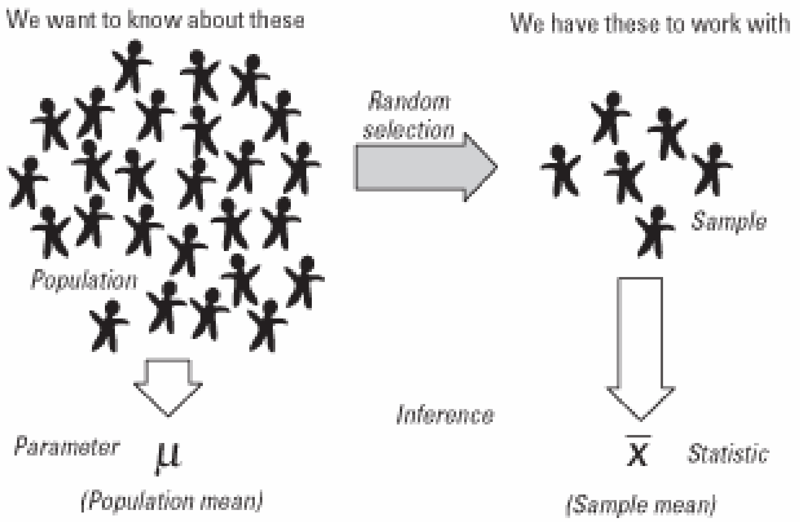
Sampling Distribution

Inferential Statistics

A Sample

- When we draw a sample and study it, we find its characteristics by calculating its measures.
- We may calculate its mean and standard deviation.
- By calculating the measures, we intend to find the estimates of corresponding population parameters.



Courtesy: www.clifffnotes.com

Example

- Consider that we want to estimate the average salaries of Oracle Functionals working in IT industry with 5 to 6 years of experience in India.
- So we draw a random sample of size 10 and calculate the measures. (Figures in Lakhs)

8.9	9.3	6.7	8.5	5.66	7.66	10.2	11.3	12.4	9.21
		Mean =	8.983						

- Hence we have the sample mean as 8.983 lakhs.

Sane's STATS Academy of Statistics

Our sample
Mean = 8.983

Sample from
person B:
Mean = 9.683

Hence we see that if the experiments is performed by different people we get different values for the sample mean.

Sample from
person C:
Mean = 10.09

Sample from
person D:
Mean = 8.77

Sample from
person A:
Mean = 8.64

So we implies that...

- The values from different samples namely by persons A, B, C, D etc. also follow a random pattern.
- This pattern of randomness is the sampling distribution.

Sane's STATS Academy of Statistics

Population and Sample Notations

	Population	Sample
Mean	μ	$\bar{x}$
Variance	σ^2	s^2
Standard Deviation	σ	s
Correlation	ρ	r

Population Parameters

Statistics

- A Statistic is said to be an estimator of a population parameter
- An estimator is a formula and estimate is its particular value

Estimation Terminology

- **Point estimate** is calculated as being a “best guess” of the population parameter. e.g. Sample mean is point estimate of population mean.
- A **confidence interval** is an interval around the point estimate calculated from the sample data, where it is strongly believed that the true value of the population parameter lies.
- The difference between the point estimate and the true value of the population parameter is called **estimation error** or **sampling error**.
- The standard deviation of the sampling distribution of the estimate is called **standard error** of the estimate.

Point Estimate

Interval Estimate

Confidence Interval

- Interval $(x1,x2)$ is said to be 95% confidence interval of population parameter μ,
  - If $P(x1 \leq \mu \leq x2) = 0.95$
- Similarly,
- Interval $(x1,x2)$ is said to be 99% confidence interval of population parameter μ,
  - If $P(x1 \leq \mu \leq x2) = 0.99$

Sane's STATS Academy of Statistics

C.I. of μ

- Assuming Normal Distribution, C.I. of mean μ with known standard deviation is given by the formula(Sample size = n):

$$(\bar{x} - z.value\frac{\sigma}{\sqrt{n}}, \bar{x} + z.value\frac{\sigma}{\sqrt{n}})$$

- Assuming Normal Distribution, C.I. of mean μ with unknown standard deviation is given by the formula(Sample size = n):

$$(\bar{x} - t.value\frac{s}{\sqrt{n}}, \bar{x} + t.value\frac{s}{\sqrt{n}})$$

Margin of Error

- The quantity t.value* s / √n is margin of error.
- More is the margin of error wider is the C.I.
- If its 95% C.I. then its confidence coefficient is 0.95.
- If its 99% C.I. then its confidence coefficient is 0.99.
- Confidence coefficient is denoted by (1- α)
- More the confidence coefficient wider is the C.I.
- Also greater is the sample size n, lesser would be the margin of error.

Thank You