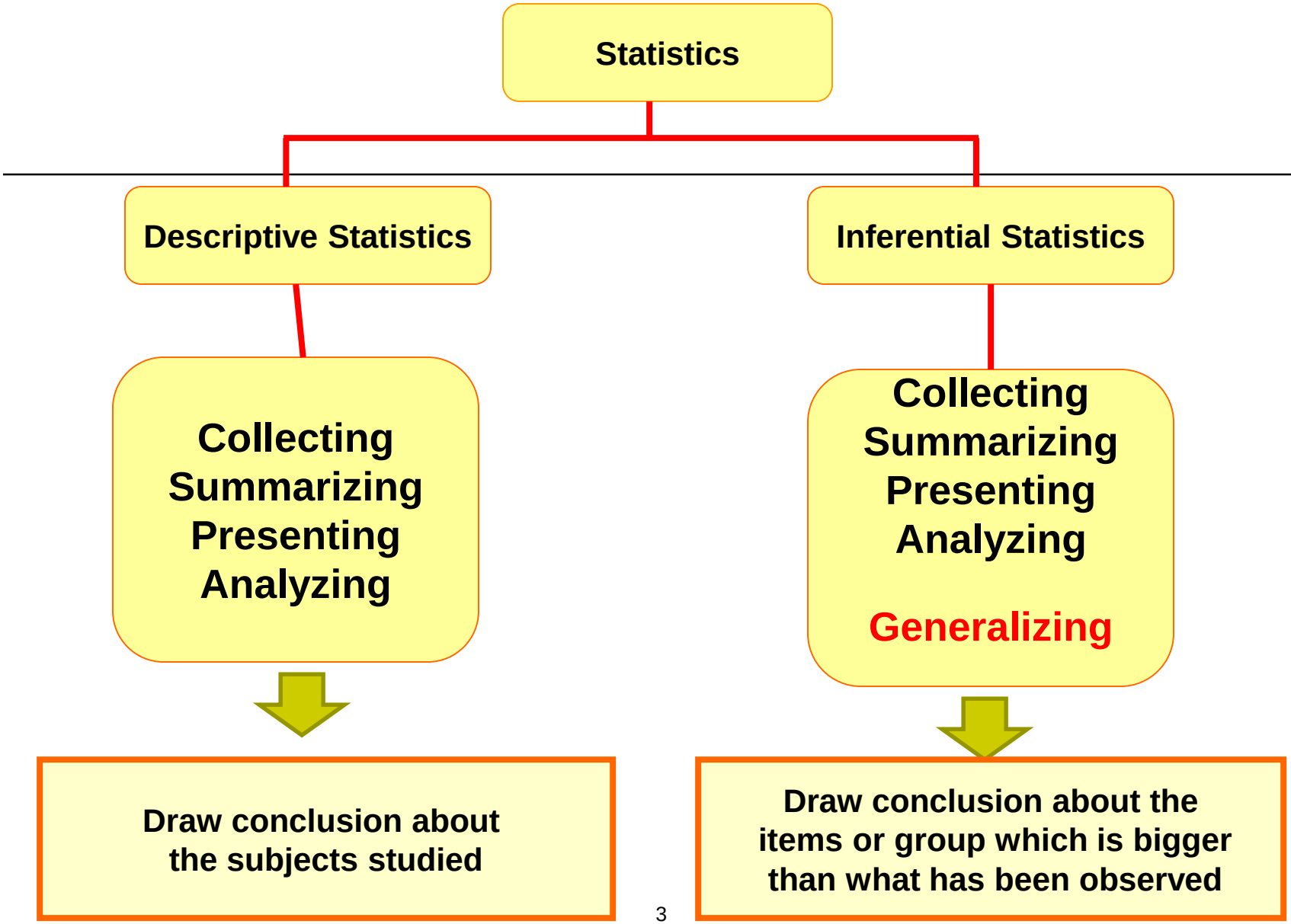
Statistical Issues in Relation to Audit

Statistics are no substitute for judgment.

A Saying: Statistics are used much like a drunk uses a lamppost: for support, not illumination.

What is Statistics ?

STATISTICS: Field of Study Concerned with (a) Collection, Organisation/presentation, Summarisation and analysis of Data - Descriptive Statistics (b) drawing of inference about a set of data (Population) when only a part of data (Sample) is observed - Inferential Statistics.


Statistics
Descriptive Statistics
Inferential Statistics
Collecting
Summarizing
Presenting
Analyzing
Collecting
Summarizing
Presenting
Analyzing
Generalizing
Draw conclusion about the subjects studied
Draw conclusion about the items or group which is bigger than what has been observed

Why Statistics in audit

- To develop an appreciation about averages and variability and how they can be used in audit.
- For making data into information
- Develop understanding of ideas of statistical reliability/precision, probability, Risk/errors etc.
- Use these ideas to develop a proper sampling design including decision about sample size and draw valid inference.

Population and Sample

Population: a complete set of elements (vouchers, bills, audit entities) that possess some common characteristic defined by the study/audit criteria or the entire group of people or objects (vouchers, bills, audit entities) to which the researcher/auditor wishes to generalize the study/audit findings.

Sample: A sample is a part of the population, selected by the investigator/auditor to gather information (measures) on certain characteristics of the original population.

Sampling, Census and Statistical Inference

Sampling: The Process of Selection of a sample from a population to generate precise and valid estimates

Census: The process of collecting relevant information/data in respect of each and every member/unit of the population

Statistical Inference: Drawing Conclusions (Inferences) about a population based on an examination of sample(s) taken from the population

Describing Sample/population

- Descriptive Statistics
  - Measures of Central Tendency
  - Measures of Variability
- Other Descriptive Measures like

  Minimum and Maximum: highly sensitive to extreme observations
- Sample size (n)
- Percentiles like: Median = 50^{th} percentile; Q1, Q3 etc.

Measures of Central Tendency (Averages)

- Measures the “center” of the data
- Measures commonly used for averages
  - Mean
  - Median
  - Mode
- The choice of which measure to use depends on nature of data
- It is okay to report more than one.

Measures of Central Tendency: Mean

- The mean is given by the sum of the observations divided by the number of observations. For example, if observations are 1,3,5,7,9 then the mean is $\frac{1+3+5+7+9}{5}$
- If the data are made up of n observations x_1, x_2,…, x_n. We can calculate the sample mean as:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

where Xi's are observations, Σ is summation, 'n' is sample size and $\overline{X}$ is sample mean.

Measures of Central Tendency: Mean Cont.

- The population mean is usually unknown; so we try to make inference about it.
- According to statistical sampling theory, sample Mean is unbiased estimate of population Mean

Measures of Central Tendency Median & Mode

- Median

  "Middle observation" according to its rank in the data.

  Better than mean if extreme observations are present i.e. for skewed data.

- Mode

  value that occurs most

  Good for Qualitative (ordinal or nominal) data

- If data are symmetric: mean = median = mode

Dispersion: Measures of Variability

- Measure the “spread” in the data
- Measures commonly used
  - Variance
  - Standard Deviation
  - Range
  - Inter-quartile Range

Measures of Variability: Variance

- The sample variance (s^2) may be calculated from the data. It is the average of the square deviations of the observations from the mean.

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)$$

- Where ‘n’ is the sample size and $\overline{X}$ is sample mean
- The population variance is often denoted by S^2. This is usually unknown.
- For sample size n > 30 ‘n’ may be used instead of ‘n-1’

Variance Cont.

- The reason that we divide by *n*-1 instead of *n* has to do with the number of "information units" in the variance. After estimating the sample mean, there are only *n*-1 observations that are *priori* unknown, n -1 is also known as degrees of freedom.
- This makes s^2 (Sample Variance) an *unbiased* estimator of S^2 (population Variance).

Standard Deviation (SD)

- Square root of the variance
  - $s = \sqrt{s^2}$ = sample SD
  - $S = \sqrt{S^2}$ = population SD - Usually unknown
- Merits: Expressed in the same units as the mean (instead of squared units like the variance)
- Demerit: s is not an unbiased estimator of S

  SD is difficult to calculate

Range and Quartile Deviation (QD)

- Range = Maximum - Minimum
- QD = ½*($Q_3 - Q_1$)
- QD is robust than the range to extreme observations
- SD is best and the most useful measure of Variation; however if there are outliers (i.e. if the data are highly skewed) it should not be used.

Coefficient of Variation: A Relative Measure of variation

Coefficient of Variation: The standard deviation of data divided by its mean. It is usually expressed in percent.

$$\text{Coefficient of Variation (CV)} = \frac{\sigma}{\bar{x}} \times 100$$

Where σ is the SD and $\bar{x}$ is Mean.

CV gives an idea of Consistency or variability of the data; a series having smaller CV is more consistent or less variable. The smaller the variability in a series, the smaller would be the sample size required.

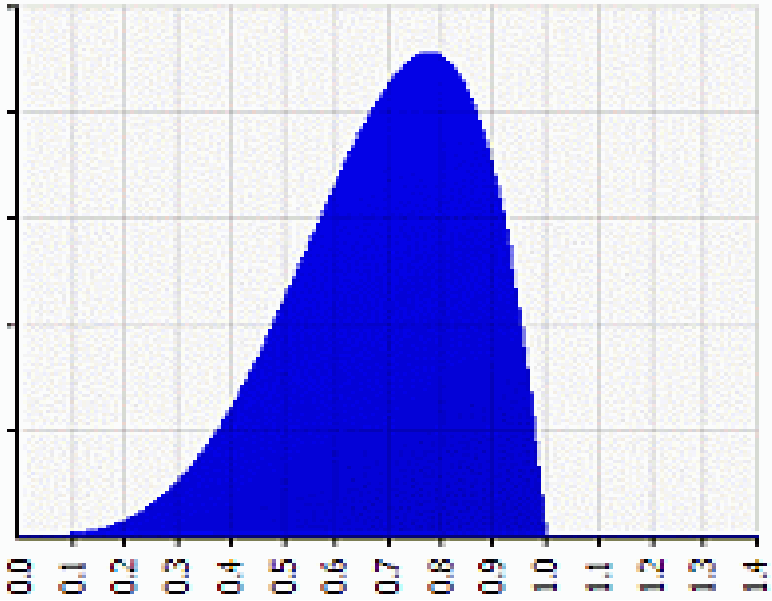
Skewness and Kurtosis: Measures of shape

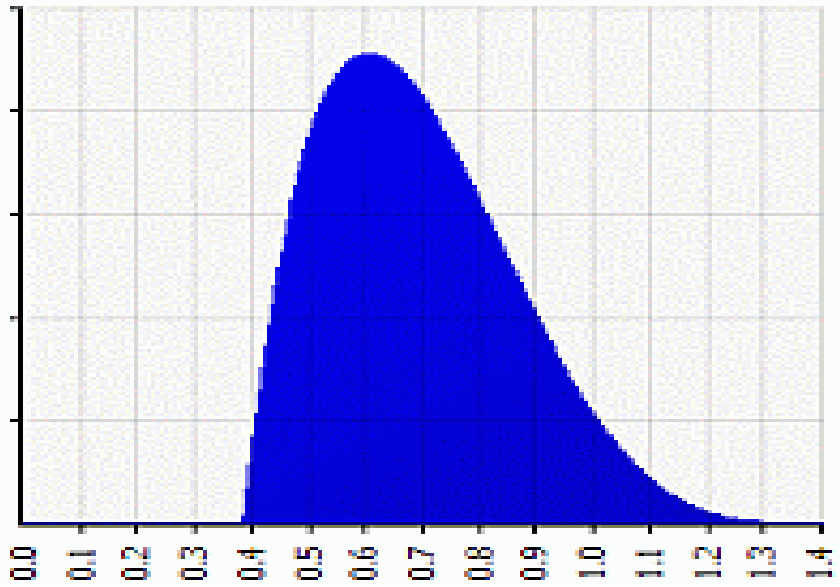
Shape of data is measured by Skewness and Kurtosis

Skewness measures lack of symmetry of data

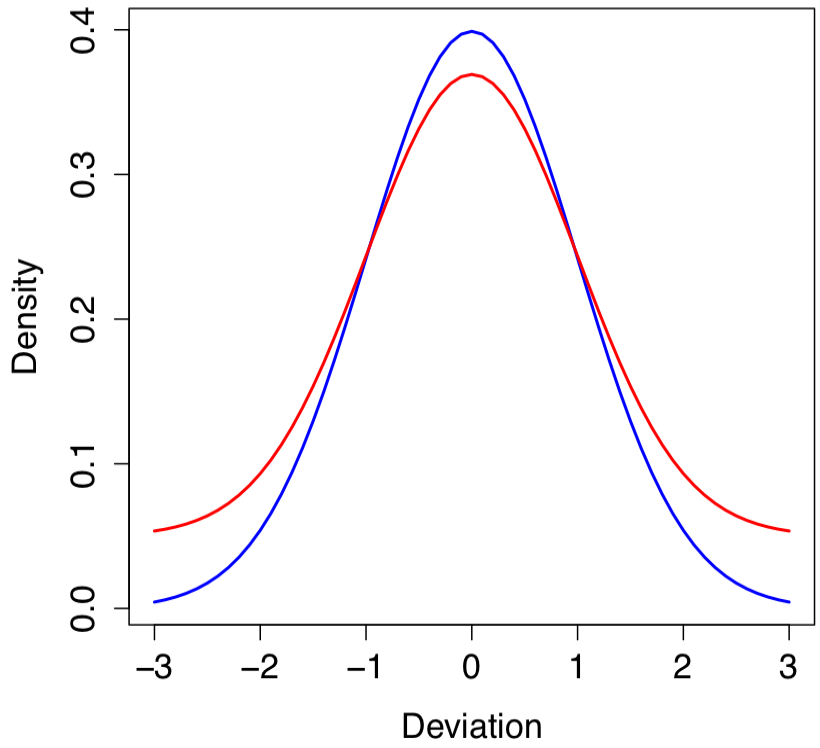
Positive or right skewed: Longer right tail

Negative or left skewed: Longer left tail





Kurtosis: relative flatness or peakedness



Kurtosis relates to the relative flatness or peakedness of a distribution. A standard normal distribution (blue line: $\mu = 0$; $\sigma = 1$) has kurtosis = 0. A distribution like that illustrated with the red curve has kurtosis > 0 with a lower peak relative to its tails.

Sampling: Some Facts

- For very small samples (e.g., <5 observations), summary statistics are not meaningful. Simply list the data.
- Beware that poor samples may provide a distorted view of the population
- In general, larger samples are better representative of the population but they need more resources.

Probability(P)

- Measures the likelihood with which event occurs

$$P = \frac{favourable \;\; number \;\; of \;\; cases}{total \;\; number \;\; of \;\; cases}$$

$$P \; lies \;\; between \;\; 0 \;\; and \;\; 1$$

$$probility \;\; of \;\; an \;\; impossible \;\; event = 0$$

$$probability \; of \; a \;\; certain \;\; event = 1$$

Probability Distributions

- A probability distribution describes the behavior of the character of interest (called variable)
  - It identifies possible values of the variable and provides information about the probability with which these values (or ranges of values) will occur.
- Important probability distributions are Binomial, Poisson and **Normal**

Introduction to Statistical Sampling and Sampling Designs in Audit

Sample and Sampling

A sample is a group of units selected from larger group (the population).

Sampling is the process of selection of *some members of* a population to generate precise and valid estimates of population parameters.

Sampling may be statistical or non-statistical.

What is Statistical Sampling ?

Essential features of statistical sampling are:

1) The sample units (or transactions) should have a known (may not be equal) probability of selection.

2) The sample result should be evaluated mathematically, that is in accordance with probability theory.

Definition of Some Sampling Terms

- **Sampling unit (Basic sampling unit)**

  Example: vouchers, cheques, bills, districts

- **Sampling frame**

  List of all sampling units in the population

- **Sampling scheme:** Method used to select sampling units from the sampling frame

- **Parameter :** (Population characteristics) – Relates to study variable, based on all the units in the population.

- **Statistic:** (Sample characteristics). Corresponding value of the population parameter based on sample values

Advantages of using statistical sampling in audit

- The Statistical Sampling method is scientific and transparent thus credible estimates of population parameters are possible to generate.
- Auditor can take a "*True and fair view*" about the auditable entity as a whole.
- The sample results are **objective and defensible** as these are free from any kind of bias.

Advantages of using statistical sampling in audit Cont.

- The method provides an estimate of the sampling error.
- It helps in estimating the required sample size on an objective basis. Thus auditor can avoid under or over auditing.
- Statistical samples accomplished by different auditors may be combined and evaluated.

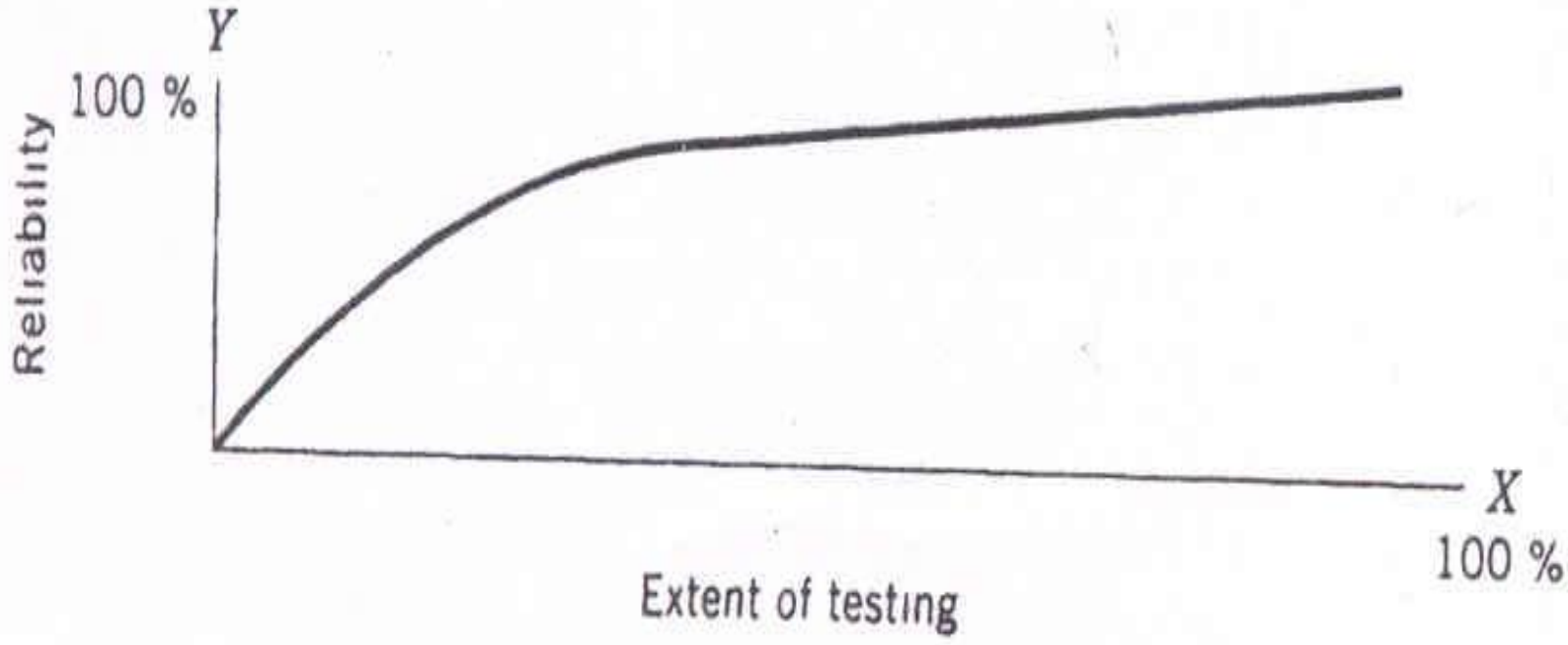
AUDIT VISION

Today the objective of an audit is not served by locating a few objections but an assurance is required for the entire population in order to make the report more effective to the planners and administrators so we need to use sampling.

Audit Sampling

Application of an audit procedure to less than 100% of the auditable population (of vouchers, bills, Districts, audit entities) for the purpose of evaluating some characteristics of the auditable population for audit conclusion. For e.g. to sample Account balance or Class of transactions or selection of audit entities.

Need of Sampling in Audit



Relationship Between Extent of Testing and Reliability

Broad categories of Statistical Sampling: Attribute and Variable: Attribute Sampling

Attribute Sampling: In attribute sampling purpose is to estimate number of times an event occurs in the population. It answers the question: How many? or How often? For e.g. to estimate the number (percentage) of invoices paid twice or no. of vouchers containing error. The auditor may conclude after sampling that (s)he is 95% confident that the % of invoices with double payment does not exceed 2%.

Broad categories of Statistical Sampling: Attribute and Variable: Attribute Sampling cont.

- Used to estimate proportion of a population that possesses a specific characteristic
- Most commonly used for T of C
- Can also be used for dual purpose testing (T of C and Substantive Tests)
- Its three types are: (i) Fixed sample size (ii) Sequential (stop or go) sampling (iii) Discovery sampling.

Two broad categories of Statistical Sampling: Attribute and Variable - Variable Sampling

In variable sampling the purpose is to estimate a quantity like population average or total. It answers the question, How much? The auditor usually uses this sampling in performing substantive tests to estimate the monetary misstatement in an account balance. It has the following types: (i) Un-stratified Mean Per Unit (MPU) (ii) Stratified MPU (iii) Difference Estimation. (iv) PPS - combines attribute and Variable Samplings.

CLASSICAL VARIABLES SAMPLING

- (i) Un-stratified MPU (ii) Stratified MPU and (iii) Difference Estimation are called classical variable sampling as they use normal distribution theory to identify amount of misstatement
- Useful when large no. of differences expected
  - Smaller sample size than MUS
- Effective for both overstatements and understatements
- Can easily incorporate zero balances

MONETARY-UNIT SAMPLING

- AKA probability proportional to size (PPS) sampling, cumulative monetary unit sampling
- Used to estimate Rupee amount of misstatement i.e. each rupee spent has equal representation in the sample

Sampling Risk

- Sampling risk – the risk that the auditors' conclusion, based on a sample, may be different from the conclusion that would be reached if the entire population was subject to the same audit procedure
- Sampling error is generated due to failure of selecting the representative sample. It is measurable under statistical sampling .

SAMPLING RISK/ERRORS

- Possibility that the sample is NOT representative of the population
- As a result, auditor will reach WRONG conclusion
- Decision errors
  - Type I – Risk of incorrect rejection
  - Type II – Risk of incorrect acceptance

Sampling Risk

Actual Conclusion based on 100% check

Sample Results Indicates	**Effective Controls or No material misstatements**	**week Controls or material misstatements**
Effective Controls or No material Misstatements	**Correct Decision**	**Risk of Assessing Control Risk Too Low (β)**
week Controls or material misstatements Indicates	**Risk of Assessing Control Risk Too High (α)**	**Correct Decision**

Risk of incorrect rejection' (Alpha risk) or **Risk of Assessing Control Risk Too High**:

- Risk that sample supports the conclusion that the account balance is materially misstated when it is not.
- arises when the sample indicates a higher level of errors/risk than is actually the case.
- This situation is usually resolved by additional audit work being performed i.e. by large sample
- affects audit efficiency but should not affect the validity of the resulting audit conclusion

'Risk of incorrect acceptance' (Beta risk) Risk of Assessing Control Risk Too Low

- material error is not detected in a population because the sample failed to select sufficient items containing errors.
- This risk, which affects audit effectiveness, can be quantified using statistical sampling techniques
- To control this risk we increase precision

TYPE I – RISK OF INCORRECT REJECTION

- Internal control: Risk that sample supports conclusion that control is NOT operating effectively when it really is
  - AKA – Risk of under reliance, risk of assessing control risk too high
- Substantive testing: Risk that sample supports conclusion that balance is NOT properly stated when it really is

TYPE II – RISK OF INCORRECT ACCEPTANCE

- Internal control: Risk that sample supports conclusion that control is operating effectively when it really isn't.
  - AKA – Risk of overreliance, risk of assessing control risk too low
- Substantive testing: Risk that sample supports conclusion that balance is properly stated when it really isn't.
- Auditor focuses on Type II Risk but also provides coverage for Type I

Non-sampling errors

- Non- Sampling risk – Non Sampling Error includes any misjudgement or mistakes by the auditor that may lead to incorrect conclusions based on the tests carried out by him. These errors may occur even if full population is examined.
- Non-sampling error is generated due to failure of measurement of true characteristic(s) due to: (i) non-response (ii) measurement error (iii)fatigue & (iv) others. These errors can be controlled by better training & management etc. There is no standard procedure to measure these errors.

NONSAMPLING RISK: How to control

- Risk of auditor error
  - Sample wrong population
  - Fail to detect a misstatement when applying audit procedure
  - Misinterpret audit result
- Controlled through
  - Adequate training
  - Proper planning
  - Effective supervision

CONFIDENCE LEVEL and SAMPLING RISK

- Complement of sampling risk
  - 5% sampling risk ↔ 95% confidence level
- The greater the reliance on test results and the more severe the consequences of Type II error, the higher the confidence level needed
- Sample size increases with confidence level (decreases with sampling risk)

TOLERABLE ERROR/MATERIALITY AND EXPECTED ERROR

- **Tolerable error**: The maximum deviation rate established that the auditor would be willing to accept and still conclude that the audit objective have been achieved. - used in compliance test .
- **Materiality:** The value of error that an auditor is willing to accept and still concludes that the audit objective is achieved. The smaller the materiality, the larger is the sample size - used in substantive testing .

Statistical Precision

- ***Statistical Precision:*** *The closeness with which a sample estimate* can be expected to approximate the relevant population value. It is an estimated value.
- This precision is usually estimated using a standard error (Std. Deviation of estimate), that is, the amount of chance fluctuation (or lack of precision) we can expect in sample estimates.
- Sample size increases as the desired precision increases

WHEN DO YOU SAMPLE?

- Inspection of tangible assets, e.g., inventory observation; Inspection of roads and other works
- Inspection of records or documents, e.g., bills, vouchers, cheques
- Selection of Audit entities
- To check effectiveness of internal controls

WHEN IS SAMPLING INAPPROPRIATE?

- Selection of all items with a particular characteristic, e.g., all disbursements > Rs 3 crores
- Testing only one or a few items, e.g. automated IT controls, walkthroughs
- Analytical procedures
- Inquiry

WHEN IS SAMPLING NONSTATISTICAL?

- If sample size determined judgmentally
- If sample selected haphazardly but randomly
- If sample results evaluated judgmentally

Types of Statistical Sampling

- Simple Random Sampling
- Systematic Random Sampling
- Probability Proportional to Size
- Monetary Unit Sampling
- Sequential (Stop or go) sampling
- Discovery Sampling
- Stratified Sampling
- Cluster sampling
- Multistage Sampling

The principal steps in auditing using statistical sampling :

- Audit objective
- Audit Population
- Degree of precision desired for quantitative assessment
- The list of transaction/auditable entities
- Selection of Auditable entities
- Preparation of Questionnaires/Data Format
- Audit of selected entities.
- Recoding of audit observation in the prescribed data format
- Statistical evaluation of audit observations
- Report writing