Data Analytics and Use of CAATTs

Seminar On Investigation and Forensic Accounting & Audit

CA Mahesh Bhatki

Mumbai, 28/December/2013

Agenda

- Overview of CAATTs (Tools and Techniques)
- Some Useful Techniques for Investigation
- MS Excel as Audit Tool
- Mathematical Methods (Benford's Law, RSF, etc)

CAATTs

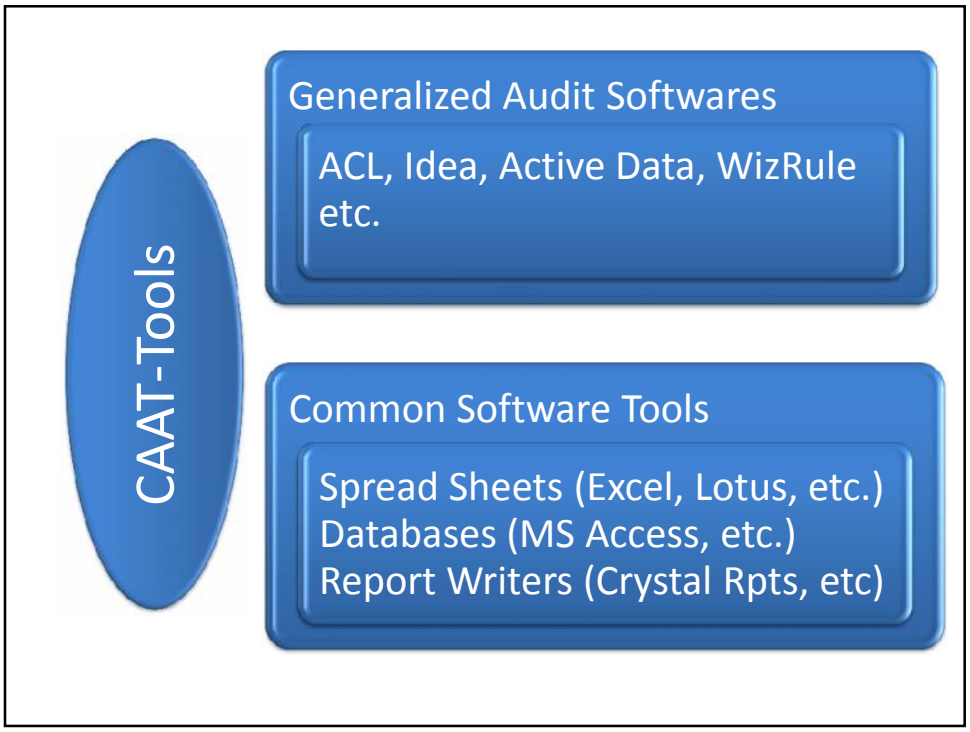
- Audit (Regulations, Opinion, Transactions, IS-Controls, Checks, Sampling ….)
- Scarce Resources (time, audit staff, cost,...)
- CAATT is indispensable (Data Volume, online transactions, paperless records, complex data, timelines,….)

CAAT-Tool : Use of Software Tool

- Learning the syntax and how to use various functions of the software
- Knowledge of Software Application - Understanding of Database

CAAT-Technique :

- What tests, methods, procedures, practices to be applied. What questions to ask.
- Knowledge of audit objective, processes, internal control framework, how database built-up ….Auditors Judgment



Why Excel ?

- It is the most simplest tool that you have on any PC.
- It is extremely versatile

Its good to know many things, but it is always better to make best use of whatever little you know

8Checks

1. Check Missings
2. Check Duplicates
3. Round Numbers
4. Repetitive Odd-Numbers
5. Classification
6. Stratification
7. Single Transactions
8. Isolated Outliers

RSF BLaw

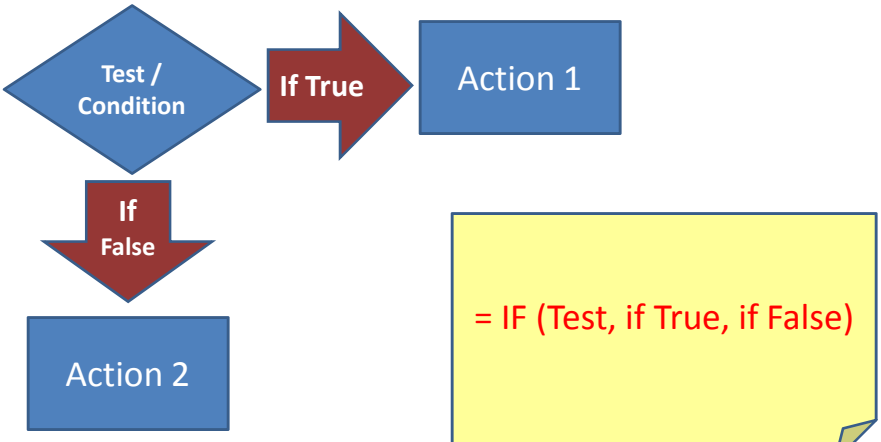
Database Functions

- 'IF'
- 'IF' in Combination with 'AND' & 'OR'
- 'CountIF' and 'SUMIF'
- 'SUMIFS'
- 'VLOOKUP'
- Pivot Table Function
- Setting Filters,
- Conditional Formatting
- Formula Auditing

Missing

'IF' function

- Very powerful function in decision making



Duplicates

Duplicates

1. Duplicated items in serial control mechanism (discussed earlier)

← test of compliance….

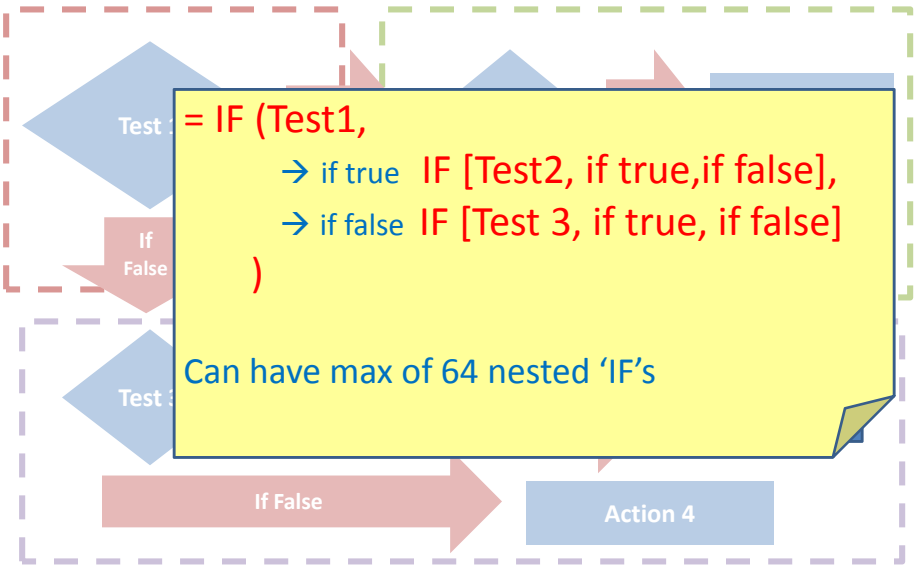
If these are generated by a good application system, chances of duplication are very low

E.g. One vendor_bill paid in installments

2. Duplicated Transactions

Combination Fields

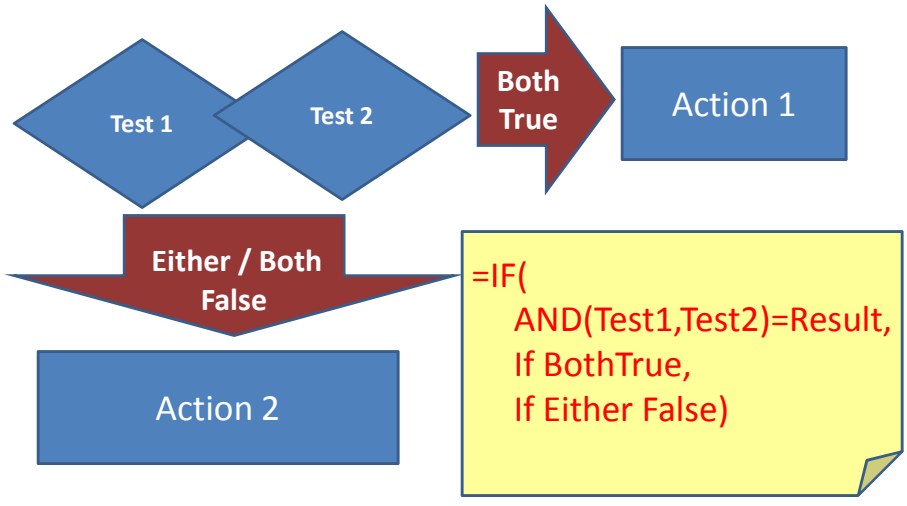
Nested'IF' function



= IF (Test1,
→ if true IF [Test2, if true,if false],
→ if false IF [Test 3, if true, if false]
)

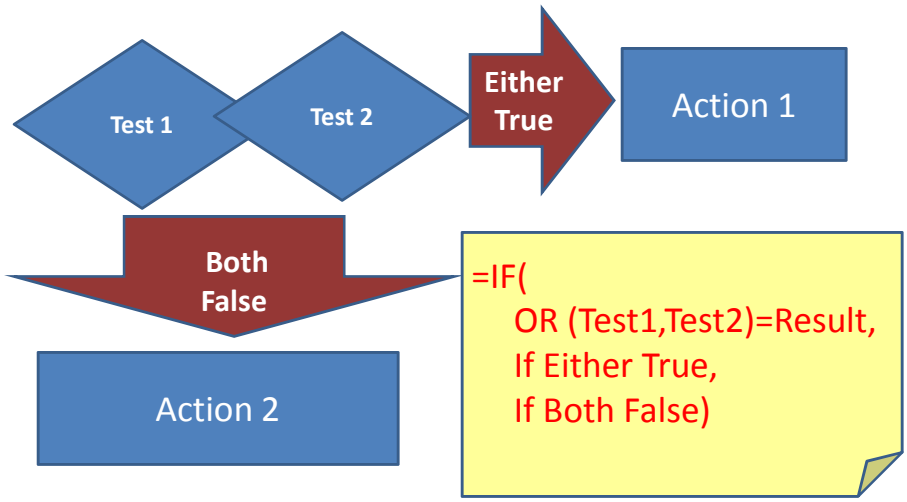
Can have max of 64 nested 'IF's

'IF' with 'AND' function



=IF(
AND(Test1,Test2)=Result,
If BothTrue,
If Either False)

'IF' with 'OR' function



Applications of 'IF'

- Detecting Gaps
- Finding Duplicate
- Locating Multiple Records
- Flagging Records
- Ageing Analysis or Advance Analysis
- Extracting Records meeting certain criteria (Combination with filter commands or with Pivot Table commands)

Round Numbers

Round Numbers

- Are there transactions of round nos.?
- Examples - monthly rentals, professional fees, audit remuneration, repairs & maintenance, hire charges, etc.
- Round Nos. most often used for estimations
- Sometimes Round-No. transactions are symptomatic of casualness / mysterious deals or settlements or frauds.

Repetitive Odd-Nos.

Repetitive Odd Nos.

- This is converse of Round-numbers.
- Unlike the round numbers, repetition of odd numbers (particularly repetitions at decimals levels) are very rare coincident.
- Unless of course there is apparent reasons say, like for telco having promotional offer of Rs 199/- pre-paid packs – but in that case, the repetitions will be by volumes and not a few stray incidences here and there.
- Repeated odd-number transactions can be filtered for detailed verification and most often these will throw up some irregularities.

Repetitive Odd-Nos.

1. Use 'Count IF' function to take counts of the numeric value field
2. Use the 'Right' Function to get the right most digit
3. Apply Filter On Step 1 : Count > 1
4. Apply Filter On Step 3 : Non-Zero's

'CountIF' Function

- Counts records satisfying given criteria
- Extremely useful in analysis of tel nos / fax nos/ email_ids etc.

= COUNTIF(criteria_ range, criteria,)

Single Transaction

Single Transaction

- What is the problem with Single Transaction?
- Generally anyone would be interested to regularly do business with an organisation. Therefore stray transactions could be symptomatic of some errors, disputes, advantages of opportunity or frauds

'VLOOKUP' FUNCTION

- Useful in fetching value from another database
- Useful for linking two databases

=VLOOKUP (lookup_value , table_array , Col_ Index, range_lookup)

'VLOOKUP' FUNCTION

- Useful in fetching value from another database
- Useful for linking two databases

=VLOOKUP (lookup_value , table_array , Col_ Index, range_lookup)

=VLOOKUP (lookup_value , table_array , Col_ Index, range_lookup)

- **Lookup_value** : Criteria value to search in the first column of the table array
- **Table_array :** The Database - the values in the first column of table_array are the values searched by lookup_value. These values can be text, numbers, or logical values. Uppercase and lowercase text are equivalent.
- **Col_index_num :** The column number in table_array from which the matching value must be returned. Generally 2, 3,4,5 . . . . .
- **Range_lookup :** A logical value true or false → to find an exact match or an approximate match.

=VLOOKUP (lookup_value , table_array , Col_ Index, range_lookup)

Range_lookup : A logical value true or false → exact match or an approx match.

If TRUE or omitted :

exact or approximate match is returned.

If an exact match is not found, the next largest value that is less than lookup_value is returned.

- **The values in the first column of table_array must be placed in ascending sort order; if not incorrect value.**

If FALSE, VLOOKUP will only find an exact match.

- **The first col. need not be sorted. If there are multi-values in the first col., the first value found is used.**
- **If exact match is not found → #N/A**

'SUMIF' Function

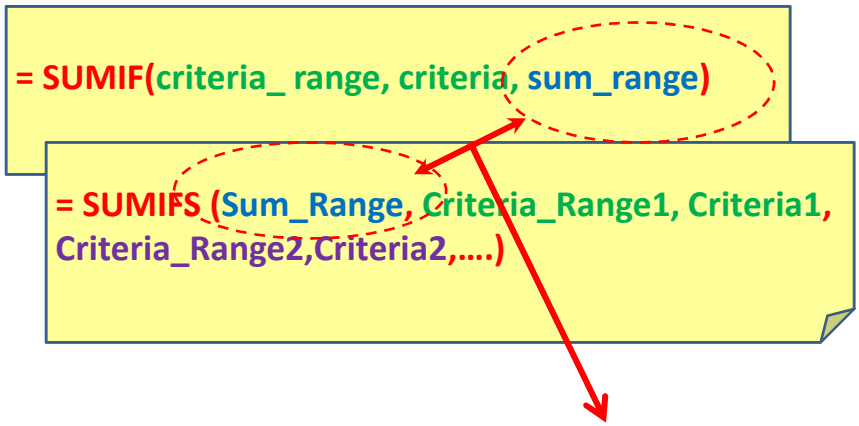
- Aggregates records satisfying given criteria.
- Useful for Summarizing, Grouping, etc.

= SUMIF(criteria_ range, criteria, sum_range)

'SUMIFS' Function

- Aggregates records satisfying given multiple-criterias. (Upto 127 criterias)
- Useful for Summarizing, Grouping, etc. (particularly cross tab summaries)

= SUMIFS (Sum_Range, Criteria_Range1, Criteria1, Criteria_Range2,Criteria2,….)



Note sum_range appears differently

Pivot Tables

A Kaleidoscopic peek at Data

Classification

The process of arranging data into homogenous group or classes according to some common characteristics present in the data

Classification :

- → **Party-wise, account-wise, stock-wise, period-wise…….**
- → **Attributes : Sum of Values, Counts, Max /Min, Average…….**

Stratification

Strata : levels, bands, groups

The process dividing large data and rearranging it into Strata

Isolated Outliers

An Isolated outlier is an observation in a data set which is far extreme in value from the others in the data set. It is an unusually large or an unusually small value compared to the others.

Spotting Techniques :

- ❑ Very simply by using Charts
- ❑ Limits Overrun (e.g Financial powers, authority levels, etc.)
- ❑Mathematical Tools (RSF, Bendford's Law, etc.)

Relative Size Factor
(RSF)

What is RSF ?

- RSF is the ratio of Largest Number to the Second Largest Number of a relevant set.

$$\text{RSF} = \frac{\text{Largest Number}}{\text{Second Largest Number}}$$

Relevance Of RSF

- Scrutiny of individual parties account is humanly ineffective/ impracticable.
- Therefore, RSF could be conveniently applied on full database
- It highlights all unusual fluctuations which may be stemming from frauds or errors

How RSF Works ?

bank payments voucher no.	Rs.
SB-211	50,000
SB-642	5,00,000
SB-547	5,00,000
SB-1864	20,000
SB-4755	23,000
SB-8347	8,500

- Any set of transactions take place in certain range. E.g. A vendor XYZ may have normal pattern of bill value range of Rs. 13k to 50k.
- If there is any stray instance of single transaction which is way beyond the normal range than that ought to be looked into. E.g. in above case, if there is bill of Rs. 5 lacs.
- RSF is above case will give a ratio of 10 (I.e. ratio of Rs. 5lac to Rs. 0.50 lacs)
- These single instances could be cases where there is some foul play.

Benford's Law

Benford's law – a tool for every auditor

A tool for Sampling → Risk based Sampling

Benford's Law is a mathematical tool that can be applied on the full database to overview millions of transactions

Benford's Law is also popularly known as the first digit law

Benford's law : Simple, simple, simple

- This law is very simple to use and gives you a overall view instantaneously.
- This law is being used by auditors across countries
- Empirical studies and research is on in many countries

Benford's Law

expected frequencies for first digit

Digit	
1	30.10 %
2	17.60 %
3	12.50 %
4	9.70 %
5	7.90 %
6	6.70 %
7	5.80 %
8	5.10 %
9	4.60 %

Z Test

$$Z = \frac{|Po - Pe| - (1/2\ n)}{\sqrt{(Pe\ X\ (1 - Pe)\ /n)}}$$

Pe = Expected Proportion

Po = Observed Proportion

n = Total Number of observations

Limitations of Benford's Law

- 32 % chance of failure
- Will not work on small database (<500-600)
- Will not work on data where limits are imposed
- Will not work on serial control numbers
- Will not work on random numbers
- Works only on numbers

Benford's set - implication

It means that there is 68% chance that there is no error / frauds

Larger the Data, better the results

Many receive advice;…

……only the wise profit by it

mkbhatki@hotmail.com