

Data Analytics for Internal Auditors

Benefits

- Improvements to general audit productivity
 - Reduction in audit risk
 - Wrong tests on the right data or drawing erroneous conclusions from the correct analysis
 - Improvement in audit independence
 - The conducting of the analysis itself as well as the reporting of results remains under the control of the individual auditor
 - Improvements in audit assurance
 - Ability to operate at a significantly higher confidence level facilitates the expression of an audit opinion with improved reliability
 - Increased audit opportunities
-
- Overall, data analysis is a judgmental process. One of the primary audit objectives is identification of anomalies in data presented or major changes in trends and relationships of data to facilitate the investigation of the reasons underlying those changes
 - The science of examining raw and unprocessed data with the intention of drawing conclusions from the information thus derived
 - It is tempting to believe that data analysis will involve analyzing 100% of the data in order to ensure that the analysis is accurate. In practice, this is neither desirable nor even possible
 - **Data**
 - The body of facts and figures systematically gathered to achieve specific purposes
 - **Information**
 - Data that has been processed into a form that is or is perceived to be valuable to a recipient and meaningful in the decision-making process

Audit analytical techniques

- These techniques facilitate the following:
 - Computation of statistical factors
 - Validation of transaction parameters

- Identification of duplicate transactions where such duplication should not exist
- Identification of missing transactions where gaps in sequence numbers may be found to be inappropriate
- Identification of calculation or arithmetic errors
- Classification to find patterns and associations
- Analysis of multiple data relationships to identify suspicious transactions

Data modeling

- Process of defining real-world phenomena or geographic features of interest in terms of their characteristics and their relationships with one another

Data input validation

- The process of evaluating collected analytical data against established acceptance criteria to determine data quality and usability in the analysis process prior to conducting the analysis itself
 - Fit for purpose
 - Accuracy
 - Availability or accessibility
 - Completeness
 - Relevance
 - Reliability
 - Timely
 - Valid
- Overall, data analysis has been define as the procedures for analyzing data, techniques for interpreting the results of such procedure, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of statistics which apply to analyzing data

Getting the right data for analysis

- In general, the purpose of an internal audit using data analysis is to seek evidence in order to determine that the control objectives of the area under review have been met, are being met, and will continue to be met
- As with any audit, the first stage is the identification of the business objectives of the area under review. Once these have been agreed upon with the auditee and management, the overall control objectives specific to that business area may be identified in conjunction with management and the auditee so that the controls relied upon by management to achieve the control objectives may also be identified

- It is at this stage that many auditors go wrong in seeking to prove that individual controls are functioning. The critical element is the achievement of the control objectives
- An efficient and effective audit data analytic procedure will follow a predefined program consisting of the following:
 - Defining the audit evidence requirements
 - Identifying the source of the evidence
 - Identifying and acquiring the appropriate skill mix to conduct the analysis
 - Selecting a data analytics strategy
 - Acquiring data access rights
 - Selecting the appropriate analytical architecture

Non-Sampling error

- Typical causes of this type of error include:
 - Lack of understanding of the nature and characteristics of the population
 - Mistakes in application of the techniques for selecting a sample
 - Use of inappropriate audit techniques on the items examined
 - Failure to recognize irregularities or errors in the items examined
 - Other misjudgments or errors by the auditor in carrying out tests or interpreting the results

Population characteristics

- Perhaps the most critical element of data analysis is the determination of the characteristics of the population to be examined
- If the audit seeks to express an opinion based upon a sample drawn from transactions over the previous three months, any opinion expressed can only be valid in terms of the three months

Judgmental sampling

- Nonprobability sampling technique with which the auditor selects the units to be sampled based on his or her professional judgment and experience. The sample to be chosen is not intended to be representative of the whole population, and it is commonly used when the auditor has a specific and limited reason for biasing the sample, and no attempt will be made to extrapolate results to the general population.
- The major drawback to judgmental sampling is that there is no ready mechanism for evaluating the reliability of the auditor's selection and the extent to which bias has influenced the sample examined

The statistical approach

- Intended to produce a sample that can be expected to be representative of the population

Sampling methods

➤ Attribute sampling

- Used to assess how frequently the particular event or attribute occurs within the population as a whole
- Discovery sampling
 - Specific form of attribute sampling, which is used when a single deviation would be classed as a critical finding
- Acceptance sampling
- Stop-or-go sampling

➤ Classic variable sampling

- Technique commonly applied to monetary amounts although other measures can also be handled

Once the sample is chosen:

- Define the objectives of the audit tests
- Decide on the level of confidence required
 - Commonly 95% confidence level
- Establish expected error rate
- Establish a maximum tolerable error rate or value range
- Establish the dispersion of the population
- Calculate the sample size required
- Choose the sample selection method
- More commonly, sampling without replacement is used, and the population number is reduced by one each time a sample item is extracted
- Haphazard selection is a technique in which the auditor samples avoid the introduction of a conscious bias but without following a structured selection methodology

Results analysis and validation

- Any conclusions the auditor draws are only as good as the data used to draw them. Data validation is the process by which the auditor verifies the accuracy and completeness of information obtained or derived prior to analysis:
 - Data type validation

- Format checking
- Length checking
- Existence checks
- Range checks
- Range checks
- Cross-reference verification
- Referential integrity validation
- Check digit validation
- Data cardinality validation
- Hash checking
- Data field uniqueness validation

Data analysis and continuous monitoring

- Continuous monitoring capabilities help the whole data analytic process by:
 - Providing better access to real-time indicators of potentially fraudulent transactions and allowing improved speed and quality of detection and management response
 - Reducing the business impact of errors and omission by reducing the length of time they go undetected
 - Enhancing assurance of corporate compliance with relevant laws and regulations
 - Giving early warning of reduced reliability of computerized internal control systems

IDEA and Data Analysis

- The second most commonly used generalized audit software in use.
 - <http://ideasupport.caseware.com/public/downloadidea/>
 - This is the latest version and is a fully functional version but limited to 1,000 records
- **Excel**
 - Use the file named “Sample” within the Source Files. ILB subdirectory
- **Access**
 - From the Source Files .ILB subdirectory. The database name in this case is ACCESS. The auditor is then given a list of the available tables; in this case, there is only one.
- **Text files**
 - A file called Sales.txt, which is to be found in the Tutorial/Source Files. ILB subdirectory

- Move to the Analysis Windows on the menu bar and select Attribute from the sampling area, then enter the following:
 - Population size: 900
 - % Expected Deviation – 1.00
 - % Tolerable Deviation – 5.00
 - Confidence Level – 95
- Clicking on compute gives the auditor a sample size of 88 and a table showing where errors occur in the sample and percentage error rates. The audit can evaluate the results of the sample testing by clicking on Evaluate and entering the following:
 - The population size: 900
 - The number of deviations in the sample
 - The sample size: 88
 - The desired confidence level: 95%
- Clicking on the representation of a calculator next to the file name, may enter the equation Pay_Type = “CASH”
- The fields may be selected by double-clicking on the field name at the bottom of the editor.
- Your file called Cash_Items will be created containing the 91 records identified as cash payments
- Within the sampling portion of the Analysis Window, the auditor may now select Monetary Unit and click on Plan
- Because this sample technique uses cumulative values, the auditor must tell it which field to accumulate. In this case, use AMOUNT.
- The auditor may decide to use only positive values, only negative values or absolute values

Seeking duplicates and missing items

- Double-clicking the field name to create an index in ascending order or double-clicking again to create another index in descending order.
- Create indexes for Company, Last Name, and Credit Limit. You can now analyze for duplicates by going to the Analysis menu on Duplicate Key, Detection. For this exercise, use the Last Name and build a file called Duplicate Naes.
- Gap Detection – this immediately highlights the missing customer numbers either individually or in groups
- Selecting the Sample-Bank transactions file and under the Analysis tab, select Duplicate Key and choose Detection. Key field to be selected would include Type,

Date and Amount with the output file named as Duplicate. This extraction will produce a file containing six deposits with duplicate dates and amounts.

Use of pivot tables

- Click on the Pivot Table selection within the Categorize box of the Analysis tab. The result name will be left as Pivot Table. Clicking on OK gives the auditor a table in which to drag the appropriate data items.
- Thus, the Salesrep_No can be dragged to the area marked Drop Row Fields Here. The Sales representative numbers will now appear in that column in ascending number order. The Prod_Code field may now be dragged to the area marked Drop Columns Fields Here.
- Sales_Bef_Tax will not be dragged to the area marked Drop Data Items Here.

Correlation and Regression

- Two variables are said to be correlated when they move together in a detectable pattern. The files are called CountryEmissions.XLS and WorldTotalEmissions.xls and may be found in a directory called tutorial
- From within the CountryEmissions file, the auditor will move to the Analysis menu and the Relate box and click to Join. The primary database will show as CountryEmissions database. The secondary database will be selected as the WorldTotalEmissions database matching the two on the key of Year by clicking on Match and selecting primary key and secondary key both as Year.
- Proceed again to the Analysis menu and within the Explore box, may choose Statistics and select Correlation. To answer the audit objectives, the field to be correlated are Carbon and Total. The auditor will select the audit unit field as Country with all the units selected and a database to be created called Correlation.

Trend Analysis

- CountryEmission database and extrapolating it into the succeeding years. Before the future forecasts are created, however, the auditor would create a forecast for 2004 based upon the existing history and compare it to the actual carbon emissions for 2004. This would be carried out by taking the database file and doing a direct extraction from the Analysis menu and creating a file called Pre2004 and using the Equation Editor to create the equation $\text{Year} < > 2004$.
- Using this new file, the auditor can again proceed to the Analysis menu and select Statistics, Trend Analysis.
- The trend that the auditors are interested in is Carbon, the audit units are once again All Countries, and the auditor wishes to Generate Forecasts. The output will involve

the creation of three separate databases: The Trend Analysis database, the Forecast database, and the MAPE database.

- The Mean Absolute Percentage Error is a measure of the fit between the fitted values and the set of actual values, or expressed another way, it indicates the “goodness of fit” or reliability of the forecast.
- To create a 2004 Emissions. These two extracted files may now be joined using the 2004 Emissions file as the primary file and the Trend Analysis Forecast file as the secondary file. The matching key will be Country.
- Fields from the secondary file will only include the Carbon Forecast. The file name will be Joined. Using Joined as the primary file, it may now be joined to the Trend Analysis MAPE file including the fields arbon_Slope and Carbon_MAPE from the secondary database.

Compliance

- The auditor may simply open the file, and by clicking on the Analysis tab, may select Summarization within the Categorize box. The auditor may then choose which fields to summarize by, Salesrep_No and choose the field to total, such as Sales_Plus_Tax, creating a file named Saled by Rep.

Pricing rules not followed

- Under the menu item Data, the auditor can Append a virtual field to the database. Using a field name of Test_Sales, with two decimal places, and using the equation editor to define the field as QTY times Unit_Price.
- It is now simple to do a Direct Extraction where Sales_Bef_Tax does not equal <> Test_Sales.

Identify duplicate employees in the employee master file

- By selecting the Sample_Employees file and conducting a simple Duplicate Key analysis and choosing a Key of Name, the auditor can create a file containing only duplicates.

Benford Analysis

- A form of pattern analysis used in the analysis of high volumes of data looking for evidence of fraudulent transactions. Sample-bank-transactions file as an example, from Analysis tab, select Benford’s Law. I this particular file, there is only one field that could be interested in, and that is the Amount. From field statistics, there are negative entries; therefore, we may wish to exclude the negatives and only include positive values.

- The script can be saved and may be executed in the future by going to the Macros tab, selecting Open, retrieving the macro and clicking on Run.

Clear Writing Techniques

- Clarifying relevant audience perspectives
- Establishing the facts to support an argument
- Prioritizing, editing, and sequencing facts in order to build an argument
- Formulating and stating audit conclusions
- Identify potential emotional reactions both positive and negative from the readers of the analytical report
- Maintaining the focus on logical arguments
- Address potential objections
- Ensure consistency

Basic report structures

- ✚ Executive summary
- ✚ Background, scope, and objectives
- ✚ Summary of major findings
- ✚ Audit opinion
- ✚ Detailed findings and recommendations
- ✚ Technical appendices if required
- Generating the best visuals for displaying data requires the auditor to follow some basic guidelines:
 - Understanding the audience and the type of visual communication medium most acceptable for the message to be delivered
 - Use of a technique that conveys the information in the simplest form for the most effect absorption by the audience
 - Deciding on the kind of information the auditor wishes to convey and what he or she wishes the audience to comprehend
 - The auditor must understand him or herself the data to be visualized, including its cardinality and size

Making visualization effective

- The most critical component of any visualization is the overall quality of data. Data must be sanitized before integration because the very summarizing process involved in visualization can cover up inaccuracies and incompleteness in the data, resulting in misleading visualizations.