

Data Conversion and Data Integrity Testing using XL Audit Commander

XL Audit Commander

data analysis made easier ...

XL Audit Commander

Table of Contents

XL Audit Commander is a data analysis / data mining support tool for performing analytical procedures within the Excel platform. This is done with command language statements or menus, along with a toolbar. The primary purpose of the tool is to support forensic accounting and fraud investigations, although it may find other uses in research and as a general purpose accounting support tool. This is *free software*, there is no license cost, and it may be used without restriction for any purpose, including commercial and educational.

The tool is installed as an Excel add-in, and as such requires Excel 2000 or later. The tool works only on Windows operating systems.

Summary	1
Cautions and Limitations	1
Conversion speeds	2
File Format.....	2
The data conversion process	2
The data conversion reports	3
An example	3
Conversion File.....	3
Processing Parameters	4
Processing Alternatives.....	4
GUI	4
VB Macro.....	5
Status report from the Conversion Process.....	6
Data Integrity testing	6
Control totals.....	6
Date Validation	6
Subtotals.....	7
Benford's Law	7
Numeric Gaps	7
Duplicates	8
Impossible Values.....	8
Round Numbers	8
Conclusion.....	8

XL Audit Commander Usage Guide

Summary

Converting data from its existing format into a format useable for audit testing and data integrity testing is an essential, but often challenging, portion of the overall audit process. This article discusses one approach for data conversion and outlines some of the considerations that may be addressed during data integrity testing.

Cautions and Limitations

Only certain types of data input formats can be converted using this system. If your data does not meet these criteria, then in order to use the system described here, the data must first be converted into a useable format. The input data must meet all of the following criteria:

1. Must be in ASCII code (e.g. EBCDIC not supported)
2. Must be unblocked, i.e. contained in data lines terminated by a carriage return and line feed character
3. Input data columnar data elements can not contain embedded tab characters (ASCII 09)
4. Proprietary data formats for numeric data are not supported, e.g. “packed” data, “zoned decimal”, binary and proprietary floating point amounts.
5. Arrays containing a variable number of elements are not supported
6. Each data element must occur in the same position (offset) in each record and must have a fixed length consistent from record to record

Certain data elements can not be converted or otherwise handled:

1. Dates prior to AD 100
2. Dates after AD 9999
3. Amounts less than negative 920 trillion or greater than positive 920 trillion

Other limitations

1. Record counts in excess of two billion will result in data conversion errors as to counts.
2. Cumulative sub-total amounts for numeric columns in excess of 920 trillion or less than negative 920 trillion will also result in conversion errors
3. Transactions with fractional pennies, i.e. amounts less than .01 will result in rounding errors in reported totals
4. Conversions in excess of 500 million rows can be performed, but the results can not be statistically analyzed with this system, i.e. totals, minimum, maximum, mode, mean, standard deviation, etc.

XL Audit Commander Usage Guide

Conversion speeds

Benchmarks on a 1.53 GHz processor with .5 GB of memory using Windows XP indicate that conversion speeds encountered may be “sluggish”. For example, statistical computations process only about 3,000 transactions per second. A file with 80,000,000 records may require 5 – 6 hours to process.

File Format

Output from the system is a file in “tab separated value” format, whose first row consists of column names separated by tab characters. Each record is terminated with carriage return line feed character. Each column value is separated by a tab character.

The data conversion process

There are six steps in the process:

1. Ensure that the input data conforms with the requirements outlined above
2. For each column of data identify the starting position in the record and the column width (the left most position in the record is considered one).
3. For each data element, determine the data type, which is limited to either alphanumeric (A), numeric (N) or date (D)
 - a. For date types, only four formats are supported:
 - b. VB6 date conversion standard (ISO)
 - c. Six digit date, with month from 01-12, day from 01-31 and year from 00-99, in any sequence, e.g. mdy, ymd, ydm, myd, dym, dmy
 - d. Eighth digit date, with month from 01-12, day from 01-31 and year from 0100-9999, in one of four formats, e.g. mdy, ymd, ydm, dmy
 - e. Variable format with month 1-12, day 1-31, and year 1-9999, with each value separated by either a “/”, “.” Or “-“.
4. Provide a column name for each column
5. Prepare a data conversion specification file in TSV format containing the following information:
 - Row 1 – Column titles – e.g. Name, Start, Length, Column Type
 - Rows 2 and on – Column name, starting position, length, column type

The specification file can be prepared by using a spreadsheet and copying and pasting the data into a text file or else saving the sheet in TSV format. Alternatively, the file can be prepared using notepad or other text processor, separating the column values with the tab character.

6. Perform the data conversion using either the XL Audit Commander GUI, or else perform the conversion in batch using the DLL provided with the system.
7. Review the results.

XL Audit Commander Usage Guide

The data conversion reports

The data conversion process will provide a conversion report which shows, for each column of either numeric or date format, the number of records processed. For numeric columns, the total amount will also be provided. The conversion process stops once the specified maximum number of errors has been encountered.

An example

A basic example is provided to illustrate the process. It consists of six columns of data, each in a fixed position in each record. The first two records of the file are header records, and are not to be converted. The column values, their type and description are provided below:

- Column 1 – Invoice amount numeric – consists of values with commas and decimals, 15 characters wide
- Column 2 – Invoice date – consists of a variable date format, month-day-year with each number separated by a dash (“-“). Variable width from 5 to eight characters, aligned right.
- Column 3 – State code – alphanumeric – two character State code provided by the USPS.
- Column 4 – Invoice due date – consists of fixed width six digit number, in format MMDDYY where each letter is a number. There are no separators.
- Column 5 – Zip code – five digit alphanumeric field.
- Column 6 – Shipping date – consists of fixed width eight digit number, in format MMDDYYYY where each letter is a number. There are no separators.

Conversion File

An example conversion file is as follows. Note that each “{TAB}” below indicates a tab character, and each “(CRLF)” represents a carriage return line feed.

```
Column{TAB}Start{TAB}Length{TAB}Type{CRLF}
Amount{TAB}1{TAB}15{TAB}N{CRLF}
Date1{TAB}28{TAB}8{TAB}mdy{CRLF}
StateCode{TAB}41{TAB}2{TAB}A{CRLF}
Date2{TAB}44{TAB}6{TAB}mdy6{CRLF}
ZipCode{TAB}55{TAB}5{TAB}N{CRLF}
Date3{TAB}62{TAB}8{TAB}mdy8{CRLF}
```

Visually the same information would ordinarily appear as follows (e.g. in a word pad document or other text editor):

XL Audit Commander Usage Guide

Column	Start	Length	Type
Amount	1	15	N
Date1 28	8	mdy	
StateCode	41	2	A
Date2 44	6	mdy6	
ZipCode	55	5	N
Date3 62	8	mdy8	

Processing Parameters

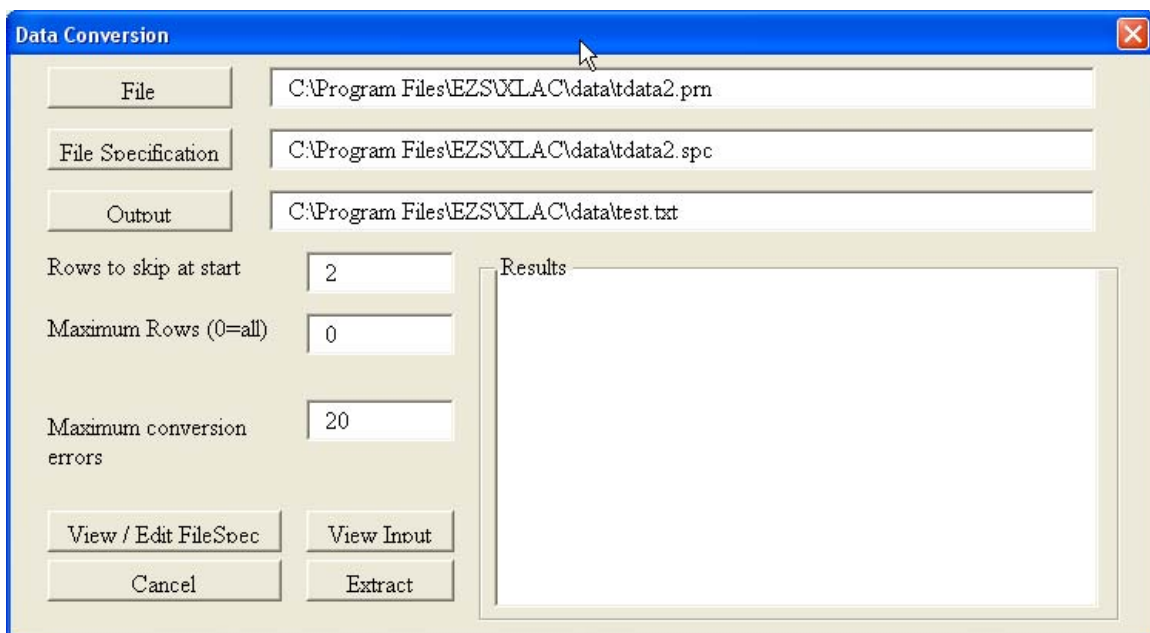
To process the file, six elements of information are required:

1. Name of input file to be converted
2. Name of file containing the conversion specifications
3. Name of file to be output (i.e. the file to be created in tab separated value format)
4. Number of header rows to be skipped, if any
5. Maximum number of rows to be processed (if there is a limit)
6. Maximum number of errors to encountered before the conversion process is halted

Processing Alternatives

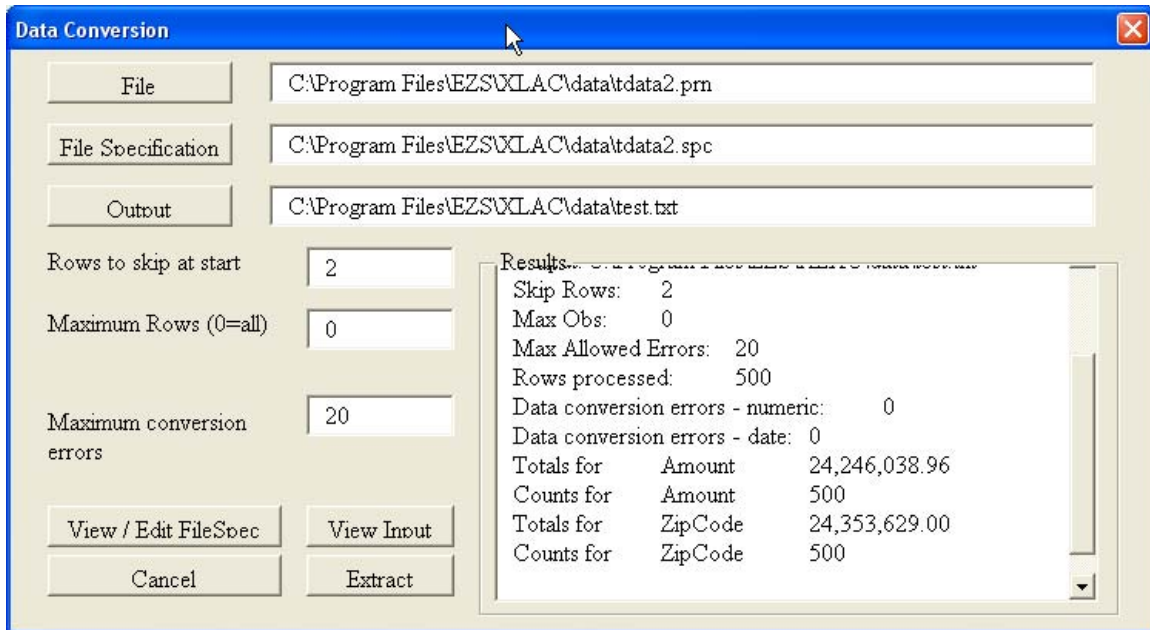
Processing can be done using either a GUI or a Visual basic macro:

GUI



XL Audit Commander Usage Guide

On clicking the “Extract” button, the output file will be created, along with a data conversion report:



VB Macro

```
Function RunTSV(sInfile As String, sSpecFile As String, sOutFile As String, _  
    imaxObs As Long, iSkipRows As Long, imaxErrors As Long) As String  
    '  
    ' perform data conversion  
    '  
    Dim oTSV As New cTSV  
    'Dim sOutFile As String  
    'Dim sInfile As String  
    'Dim sSpecFile As String  
  
    oTSV.FileName = sInfile  
    oTSV.FileSpec = sSpecFile  
    'MsgBox oTSV.ShowSpec  
    oTSV.OutFile = sOutFile  
    oTSV.MaxObs = imaxObs  
    oTSV.SkipRows = iSkipRows  
    oTSV.MaxErrors = imaxErrors  
    Call oTSV.Process  
    RunTSV = oTSV.ErrorReport  
    Set oTSV = Nothing  
  
    Application.StatusBar = "Processed output file " & sOutFile  
End Function
```

XL Audit Commander Usage Guide

Status report from the Conversion Process

Process Report:

Data Source: C:\Program Files\EZS\XLAC\data\tdata2.prn
File Spec: C:\Program Files\EZS\XLAC\data\tdata2.spc
Output: C:\Program Files\EZS\XLAC\data\test.txt
Skip Rows: 2
Max Obs: 0
Max Allowed Errors: 20
Rows processed: 500
Data conversion errors - numeric: 0
Data conversion errors - date: 0
Totals for Amount 24,246,038.96
Counts for Amount 500
Totals for ZipCode 24,353,629.00
Counts for ZipCode 500

Data Integrity testing

Once data has been converted to TSV format, a variety of data integrity tests can be performed. The basic objective of data integrity testing is to obtain assurance that the data being examined is valid data and is therefore eligible for further testing and cross-checking. There are a number of data integrity tests which can be performed, and only of few of the possibilities are described here.

Control totals

Often the very first test is simply that of comparing control totals obtained with external sources of information, such as trial balances or general ledger values. The data conversion process will provide counts and totals for each column which was designated as being numeric. Thus, for certain numeric columns, this information will be the starting point for data integrity testing. Note that data conversion errors for numeric columns are a warning indicator and can indicate that either the data conversion specifications are in error, or else the data being converted contained invalid data elements.

Date Validation

Another very basic test is simply the determination that all columns containing dates do in fact consist of valid date values.

XL Audit Commander Usage Guide

Once these basics as to numeric data and date data have been completed it is then advisable to proceed to a variety of “reasonableness” tests. These tests often consist of a variety of procedures, each of which needs to be custom tailored to meet the requirements of the data population being tested.

However there are some very basic tests which can be performed to at least perform a preliminary analysis of the data. For numeric column values, it may be desirable to obtain univariate statistics for each data element in order to answer the following questions:

What is the minimum value contained in the population and does it seem reasonable/possible?

What is the maximum value contained in the population and does it seem reasonable/possible?

What is the average value contained in the population and does it seem reasonable/possible?

What is the mode (most frequent) value contained in the population and does it seem reasonable/possible?

Prepare a histogram to see if the distribution of the population values seems reasonable and expected.

For date values, it may be desirable to perform an ageing using some amount value, such as account balance. This will help ensure that the distribution of actual date values seems reasonable and the population does not contain any unexpected date values.

Subtotals

Subtotals – It is often desirable to “slice and dice” the population various ways, e.g. obtaining sub-totals of amounts by sort values to ascertain that the amounts seem reasonable and within tolerable ranges. An example includes sales by period, store number, salesman, etc.

Benford’s Law

If certain numeric amounts should be expected to conform with Benford’s law, then it may be desirable to test for this.

Numeric Gaps

Missing information – if the data should contain a complete sequence of numbers, e.g. check numbers, invoices, purchase orders, etc. then it may be desirable to perform a “gap analysis” to identify if there are any missing sequence numbers.

XL Audit Commander Usage Guide

Duplicates

Duplicates – if certain data elements are not expected to contain duplicate values, then it may be desirable to check if this criteria is met.

Impossible Values

“Impossible” situations – generally any collection of data should be subject to certain “rules of thumb” and a check for data elements which are outside of these criteria should be examined more closely. Examples include assets with a negative cost, accumulated depreciation in excess of depreciable amount (i.e. asset cost less salvage), expected life of less than two years, expected life of more than fifty years, etc. Often a brain-storming session will be helpful in identifying possible criteria that should be examined.

Round Numbers

Use of round numbers – Often round numbers are an indicator of an estimate or a value which is not known precisely. A check of round numbers can be helpful, for example, during the review of journal entries, and other accounting estimates.

Conclusion

Converting data from its existing format into a format useable for audit testing and data integrity testing is an essential, but often challenging, portion of the overall audit process. Before any meaningful audit testing can be performed, the data may need to be first converted to a useable format. Once the data has been converted, then it should be subject to rigorous data integrity testing before further extensive audit work is performed in order to ensure that the data being examined does in fact support the financial data being tested.