

# AI Analytics for Small Business

## Techniques for Reshaping Competitiveness

Auditmetrics® AI V6.2 included\*

 Microsoft Certified

\*For Universal Windows Platform (UWP)



**Empowering Small Business**

*The analytical procedures discussed in this review are automatically handled by the Auditmetrics AI assisted software. Also generated by the software are Excel spread sheets which provides summary documentation that can be shared with clients.*

*A major saving in time and effort while maintaining the standards of the AICPA & IRS. All you have to do is decide on a margin of error as in political polling and AI does the rest.*

*For more background: [Auditmetrics.com](https://www.auditmetrics.com)*

## Small Business AI Overview

- The following slides provide a quick look at the book “**AI Analytics for Small Business**”
- Some discussion may be unfamiliar territory. This presentation is to get a look at the big picture.
- The book starts with monitoring the accuracy of cash flow contained in business accounts using statistical standards set by the AICPA and IRS.
- The book then covers how to use that audited data to project future revenues and expenses and conduct market research based on customer feedback.
- If you are new to statistical analytics and desire some background, the book has an Appendix covering a summary of statistical analytics and model building at a basic level.

# Auditmetrics® Book and Software

## auditmetrics.com

### **1. New Hardcover Book on Amazon:**

[AI Analytics for Small Business](#)

### **2. Small Business Power Series:**

[3 book Kindle & Paperback Series](#)

### **3. AI Assisted Software at Microsoft Store:**

[Auditmetrics® AI Professional V6.2](#)

Designed for small business in auditing and managing accounts, forecasting cash flow and conducting market research. It is “Auditmetrics AI Assistance” that makes it possible to conduct timely and efficient sophisticated analytics. Any book purchase includes free learning version of Auditmetrics AI V6.2 system.

# Auditmetrics

## “AI Analytics for Small Business”

The following slides provide a summary covering:

1. Statistical Auditing of Cash Flow, Using AI Assisted Techniques\*
2. Forecasting Revenues and Expenses for Small Business
3. Market Research for Small Business

*Included with book from Amazon is:*

*learning version Auditmetrics AI V6.2*

*\* Originally Developed for the Massachusetts Department of Revenue Audit Division*

# **Advantages of the Statistical Audit of Cash Flow**

- Statistical random sampling of accounts results in efficient defensible audit outcomes by eliminating bias.
- It provides for advance estimation of sample size that provides reasonableness of the sample and improves audit efficiency.
- Such efficiencies are based on universally accepted standards.
- Statistical methods in sampling an account also provides an estimate of the sampling error.
- Sampling error indicates how far the sample projection might deviate from the value that could be obtained by a 100 percent examination of the account.

- Statistical Sampling especially with AI assistance saves time and money.
- The advance estimate of the required sample size usually results in a smaller sample size than might be arrived at by using a judgmental approach.
- Multiple samples may be combined and evaluated.
- When the entire audit has an accepted objective and scientific basis, it is possible for different auditors to participate independently on the same audit.

# *Is it Worth the Effort?*

- 1. Some small business managers may be concerned that the process to obtain sophisticated analysis is complicated and time consuming. The time may be better spent on selling.**
- 2. It is the purpose of the book/software to show how to quickly get up to speed in an evolving market place where survival is more and more dependent on mastering technology and analytics.**

## ***Make inroads on the time issue:***

- Regular timely random samples allows the business manager to deal with small workable subsets of account data representative of the total book.
- The Auditmetrics Software generates a sample that meets AICPA and IRS statistical standards.
- It was originally developed for the Massachusetts Department of Revenue, Audit Division to speed up statistical audits of sales and use tax.
- The time saver is the Auditmetrics AI assistance, software feature that makes it possible to rapidly generate random samples by simply deciding on the desired ***margin of error***.

- The Validated Sample now paves the way to move onto forecasting and market research
- Regular sampling is like putting a business on a monitor
- A small business version is offered for free with purchase of the book. There are limitations on files size and forensic accounting. V6.2 SB cannot handle the sampling of millions of records as with V6.2 Pro available on the Microsoft store.

## **Steps in the Statistical Monitoring of Cash Flow:**

Step 1 - Define audit population

Step 2 - Stratify the audit population

Step 3 - Calculate the sample size

Step 4 - Allocate sample to the different strata

Step 5 - Conduct the audit to determine & rectify errors

## **Post Audit Account Management & Forecasting:**

Step 6 - Post audit financial projections & market research

# **Steps in the Analysis and Projection of Cash Flow:**

## **Step 1 -Define Audit Population**

Audit population in statistical terms is the account total book from which the sample is to be drawn. The most important decision is to determine what should be included in the audit population. A distinction should be made between cash flow items and those that are a function of tax law such as allowable depreciation deductions. The mixing of non-cash flow and cash flow items can lead to confusion.

Issues such as:

- A depreciation book expense can lead to an increase in cash inflow by means of non-cash adjustments that do not represent cash outflow.
- Loan premium payments are cash outflows, but to rely on income statement expenses ignores them and can lead to underestimate of cash outflow.
- Only a clear-cut cash flow sample can set the stage for the statistical analysis of future potential operational business health.

- With statistical tools, only cash flow is the clearest way to project future economic viability.
- Cash flow is difficult metric to manipulate.
  - It is often seen as a reliable way of determining the health of a business
  - Profit, on the other hand, can be calculated and interpreted in a number of different ways to paint a particular picture.
  - Profit is notoriously difficult to predict, and the profit that is reported can be tied up in assets
  - Cash flow is something a business can project with much more clarity
  - Planning for cash inflows and outflows, will highlight any potential problems that may arise in the future, giving ample time to avoid them.

## Step 2- Stratify the Audit Population

- Stratification is the process of dividing the population of transactions into segments (strata) based on a certain characteristic
- In the cash flow audit one would stratify the population based on the dollar amount of the transactions or invoices of interest

A stratified random sample will yield more precise results than an unrestricted random sample. One of the advantages of stratification is to offset the effect of extreme values. To minimize the effect of large invoices on sample based estimates, it is better for the auditor look at 100% of the invoices in the largest dollar stratum. This stratum is called the “**Detail Stratum**”

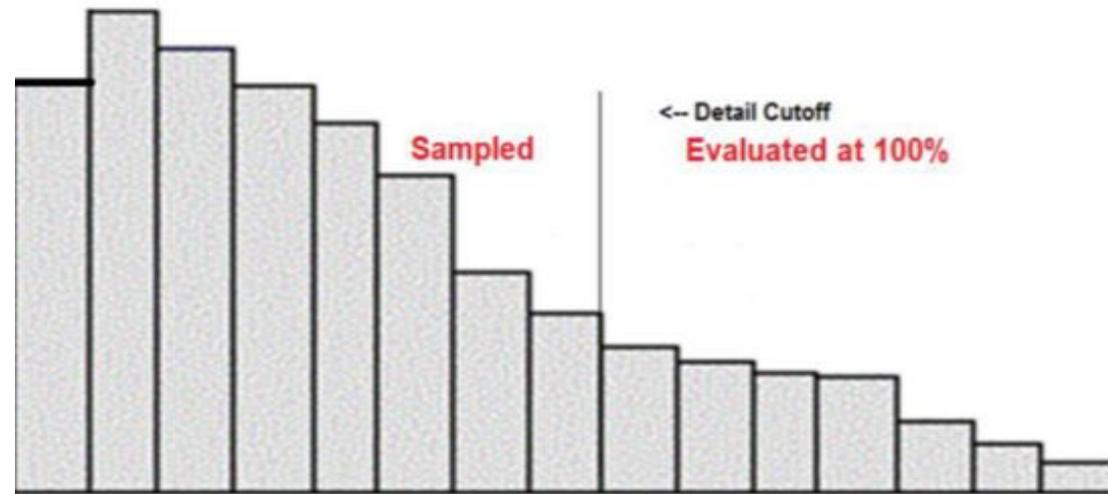
The number of strata can vary from 3 to 10. The proper number is where the Auditmetrics AI software can be of great help. The specific method to set strata boundaries is “cumulative square root of the frequency method”.

Most statistical packages do not generally focus on this specific method but its algorithm is pretty straight forward but somewhat tedious. This method is documented in Chapter 2 and is the most efficient method for determining strata boundaries.

### Example Stratification by Dollar Value

Acme Inc.	Sample	Audit Population				
Strata:	n	Mean	Std. Dev.	Total Value	Pop. Total	% Dollars
0 -99.99	54	\$27.43	\$28	\$437,090	15,934	2%
100-374.99	56	\$192.47	\$93	\$941,549	4,892	5%
375-1124.99	77	\$604.54	\$263	\$1,369,291	2,265	7%
1125-2724.99	103	\$1,676.30	\$741	\$1,731,614	1,033	8%
2725-6549.99	153	\$3,907.85	\$2,157	\$2,082,884	533	10%
6550-24500	105	\$10,340.92	\$8,275	\$3,629,662	351	18%
> 24500 Detail (100%)*	144	\$72,694.04		\$10,467,941	144	51%
<b>Totals Including Detail</b>	<b>692</b>			<b>\$20,660,031</b>	<b>25,152</b>	
Sampling Efficiency**	0.90					

The general rule of thumb is the detail stratum should reflect 30% of the total dollar volume of the total book. Auditmetrics AI uses percentile rankings in determining detail cutoff and as indicated in the previous exhibit found a detail cutoff that represents 51% of the total dollar volume.



Typical shape of financial data is skewed to the right, most dollar values bunch together with a few extreme values. To the right of the detail cutoff, 100% of the transactions are reviewed. This increases sampling efficiency because the large dollar amounts do not rely on a statistical estimate. It is what it is. The dollars to the left are then what will be subject to random sampling.

### Step 3 - Calculate the sample size

The next step is to actually calculate the sample size for each stratum omitting the detail stratum. This technique involves the Neyman allocation which will be discussed in the next section. Optimal size determination takes into account a standard set for confidence (95%) and for this example 3% precision. That is, the calculated sample size of each stratum is weighted in part by the standard deviation of that stratum. The formula is:

$$n_i = (N_i SD_i) (\text{sum}(N_i SD_i) / (A/U_R)^2 + \text{sum}(N_i SD_i^2))$$

**Where:**  $n_i$  = sample size per stratum  $i$

$N_i$  = population size of stratum  $i$

$SD_i$  = standard deviation of stratum  $i$

$A$  = acceptable precision (Margin of Error)

$U_R$  = reliability factor (Confidence Interval)

Luckily this calculation is handled by Auditmetrics AI.

## **Contained in the formula is confidence vs precision (margin of error)**

Confidence - Generally expressed as a percentage, confidence refers to the probability range, based on the normal curve, around a sample statistic estimate which includes the true population parameter. The built in confidence in this calculation is set at 95%.

*Of 100 random samples, 95 estimates would be within our confidence interval range based on the Normal Curve.*

*Precision (margin of error) - is how close a sample estimate would approximate the true audit population value*

The formula is used to seek the sample size that would give us 95% confidence that our sample estimate is within a 3% precision range of the true audit population value.

## Step 4 - Allocate sample among the different strata

- Each stratum sample size allocation is proportional to the standard deviation (variability) of each stratum in the population. This method is called the Neyman Allocation.
- Standard deviation or data variability is a measure of statistical uncertainty.

In this example total sample of size  $n$  is 698. Strata sample size  $n_i$  in each strata with the greater variability has the greater number allocated. As one moves up the strata ladder, the standard deviation get larger, that means increased variability and in turn statistical uncertainty. Increasing sample size reduces sampling error or uncertainty so the net effect of Neyman allocation is to assure the strata with greater uncertainty has a larger sample size.

# Neyman Allocation:

Sample Size = 698

	N	Std. Dev.	N x Std. Dev.	Neyman Allo.	n
0 -149.99	5,996 (39.4%)	\$31.88	191,152	6.7%	47
150-399.99	3,636 (23.9%)	\$105.33	382,980	13.4%	93
400-799.99	2,250 (14.8%)	\$189.56	426,510	14.9%	104
800-1524.99	1,404 (9.2%)	\$378.80	531,835	18.6%	130
1525-2624.99	1,029 (6.8%)	\$596.87	614,179	21.5%	150
2625-3999.99	902 (5.9%)	\$789.95	712,535	24.9%	174
			2,859,192	100%	698
<b>Total N</b>	<b>15,217</b>				
<b>Detail</b>	<b>175</b>				
<b>Total Pop.</b>	<b>15,392</b>				

## **Step 5 – Conduct the audit to determine Error Rate**

At this stage it is now up to the Auditor to review the sample for errors. But before the audit commences there are statistical validity steps that are conducted by Auditmetrics AI. These are critical steps that assure the validity of the sample and assure the ability to defend audit results. Part II of Book I covers this issue in detail.

Once the sample specification is completed, Auditmetrics AI generates a spreadsheet with the random sample so the auditor can conduct the audit. In the results section of the spreadsheet is where the auditor enters dollar amount in “Amount Error”. At this stage the auditor must correct any errors encountered before proceeding on to the next phase of projecting revenue and expenses.

## Computer Generated Spread Sheet: Medical Claims Audit Example:

HMO	Audit Results								Validity Test
	Population			Sample			Audit Results		
Strata	Pop. Mean	Pop.Total Value	N	Sample Mean	Sample Total	n	Amount Error	Error Rate	
0 -124.99	\$51.78	\$7,612	147	\$51.78	\$7,612	41	\$152.23	2.0%	pass
125-249.99	\$181.01	\$28,600	158	\$181.01	\$28,600	41	\$429.00	1.5%	pass
250-399.99	\$315.96	\$32,228	102	\$315.96	\$32,228	43	\$580.11	1.8%	pass
400-524.99	\$390.81	\$60,575	155	\$390.81	\$60,575	45	\$666.33	1.1%	pass
525-749.99	\$509.98	\$43,348	85	\$509.98	\$43,348	47	\$953.66	2.2%	pass
750-1092	\$887.46	\$79,871	90	\$887.46	\$79,871	49	\$718.84	0.9%	pass
> 1092 (detail)	\$2,723.26	\$476,570	175	\$2,723.26	\$476,570	175	\$4,765.70	1.0%	
<b>Total (Excluding Detail)</b>		<b>\$252,234</b>			<b>\$252,234</b>	<b>266</b>	<b>\$8,071.48</b>		
<b>Total (Including Detail)</b>		<b>\$728,804</b>			<b>\$728,804</b>	<b>441</b>			
<b>Error (Excluding Detail)</b>	<b>Std.Error (\$)=</b>	<b>\$2976</b>					<b>Sample Adj. error rate=</b>	<b>1.4%</b>	
<b>Error (Including Detail)</b>							<b>Overall error Rate=</b>	<b>1.1%</b>	
		Lower Bound	\$524						
		Mid-Point	\$3,500		Detail Error	\$4,766			
		Upper Bound	\$6,476						

 Auditor enters amount in error

**Audit results are automatically generated**

## **Step 6 – Post Statistical Audit Financial Projections**

Quantitative methods will help expand control of day to day operations. There is an old adage in business, “if you can measure it you can control it”. Also when seeking funding for current operations and new business plans, universally accepted quantitative methods are easy to defend. Included in the book is a level of statistical sophistication that usually is available only to large corporations that have the funding to hire statisticians.

Is the effort worth it?

- Financial institutions will put money in a business if it’s able to provide a set of thoughtful forecasts
- More important, proper financial forecasts will help develop operational and staffing plans that will help improve business efficiency.
- Auditmetrics AI is set up such that an manager can set confidence and precision levels with the software doing the heavy duty calculations in the background.

## Projections Using Regression\*

- This example projection will use 3 years (36 months) of sales data.
- If the data was from a fairly small business we may not have to use a random sample.
- That is fine up to a point but account data with large a volume of transactions it is more efficient to select a random sample

Regression Model:  **$Y = a + bX + \text{error}$**  where

**Y is total sales**

**X is month count**

**$Y = (\text{Model}) + \text{error as measured by } S^2_{y.x};$**

**In terms of the account variables: Projected Sales =  $a + b(\text{month count})$**

*\*More fully covered in Book's Appendix*

## Excerpt of Sales Data (36 months)

<b>Month</b>	<b>year</b>	<b>month</b>	<b>sales</b>
1	1	Jan	\$129,646
2	1	Feb	\$126,227
3	1	March	\$151,966
4	1	April	\$165,168
5	1	May	\$182,977
6	1	June	\$145,694
7	1	July	\$156,814
8	1	August	\$150,373
9	1	September	\$146,155
10	1	October	\$202,621
11	1	November	\$217,081
12	1	December	\$222,458
13	2	January	\$198,338
"	"	"	"

**Excel was used to build the mathematical model:**

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
R	0.95			
R Square	0.91			
<b>Adjusted R Square</b>	<b>0.91</b>			
Standard Error	47809			
Observations	36			
<i>Coefficient</i>				
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	
<b>Intercept</b>	<b>55617</b>	16274.25419	3.417518	<b>0.0000</b>
<b>Month</b>	<b>14598</b>	767.0336696	19.0321	<b>0.0000</b>

**This indicates our model is:**

$$\text{Total projected Sales} = \$55,617 + \$14,95 \times \text{Month count}$$



- Correlation coefficients are used to measure how strong a relationship is between two variables, in this case the relationship between time in months and cash flow is strong.
- The data presented is of a business that is in a very active growth phase as denoted by 91%  $R^2$ , the higher the better.
- A low  $R^2$  denotes a weak relationship between cash flow over time.
- A low  $R^2$  is still important feedback:
  1. *It may represent a very stable even cash flow over time with a healthy sustainable cash flow with no need for any action other than a desire to expand the business.*
  2. *Low  $R^2$  with cash flow problems is a warning shot, especially if a plot of cash flow over time is erratic with a high standard deviation.*
  3. *The need is to explore administrative and/or marketing issues.*

**Though we have a very good fit, there is still a problem with our model so far:**

- Data is that of a wholesaler that supplies retail outlets
- If we make a prediction for the next month or quarter it will always be higher than the previous month or quarter based on the simple linear function
- Business activity does have seasonal fluctuations, so a straight line needs to be modified
- The fourth quarter of the year with its holiday activity will always be higher than the following first quarter of the following year
- **The model is not complete without allowing for seasonal fluctuations**

The full regression equation taking into account quarterly adjustments:

$$Y = a + b x + Q1 + Q2 + Q3$$

$$a = \$83,272$$

$$b = \$14,317$$

$$X = \text{Month}$$

Where Q1 is 1 if in Quarter 1 else 0

Where Q2 is 1 if in Quarter 2 else 0

Where Q3 is 1 if in Quarter 3 else 0

Q inputs are dichotomous *dummy variables* which take a value of 0 or 1.

$$\begin{aligned}\text{Total Sales} &= \$83,049 + (\$14,317 \times \text{Month}) \\ &+ (-\$38,272 \times \text{Q1}) \\ &+ (-\$8,334 \times \text{Q2}) \\ &+ (-\$42,340 \times \text{Q3})\end{aligned}$$

You may notice the data matrix has inputs for three of the four quarters. Why not add a fourth dummy variable Q4?

The problem is that the fourth variable (Q4) is redundant. If the first three dummy variables are 0 then it is a given that Q4 has to be 1.

**So the rule is:**

1. The number of dummy variables necessary to represent a single attribute variable is equal to the number of levels (categories) in that variable minus one.
2. For a given attribute variable, none of the dummy variables constructed can be redundant. That is, one dummy variable cannot be a constant, multiple, simple linear relation or defined by another. This would violate the required randomness of the input matrix required by statistical theory.

Let's now project the estimated total sales for Quarter 4 and Month

$$\begin{aligned}\text{Quarter 4: Total Sales Month 36} &= \$83,049 + (\$14,317 \times 36) \\ &= \$598,461*\end{aligned}$$

$$\begin{aligned}\text{Quarter 1: Total Sales Month 37} &= \$83,049 + (\$14,317 \times 37) \\ &+ (1 \times -\$38,272) = \$574,506\end{aligned}$$

**Due to seasonal fluctuations as expected quarter 1 projection is lower than previous 4th quarter value.**

\* This is a quarter when all three dummy variables are zero

Regression is very valuable in adjusting predictions using categorical adjustments for various demographics:

- Geographic region and other characteristics such as gender
- Geographic region can also be a surrogate for income distribution which is readily available from government published data
- For example, there is available through IRS sources income tax collections by zip code
- This leads to a more detailed discussion in projections regarding Market research which involves market research methods
- All major market research assessments rely heavily on regression techniques and is fully covered in Chapter 4.

**The key is matching account data and customer inputs:**

For customer input, Likert customer ratings can be linked with customer sales data. Book 1 and 3 cover how data can be derived from business software such as QuickBooks<sup>®</sup> reports such as total sales by which one could add customer inputs using Likert Scales:

The marketing relevant input variables are:

**Zip Area-** that is the customers zip code aggregated by areas based on census data that delineate areas of different socioeconomic characteristics based on median family income.

**L-price-** Is average Likert customer rating evaluating price and value of the product.

**L-Satisfied-** Average rating for satisfaction with interaction with business staff

**L-Recommend** – Average rating for repeat business and recommend to others