WP.17
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Manchester, United Kingdom, 17-19 December 2007)

Topic (ii): Tabular data protection

COMPARATIVE EVALUATION OF FOUR DIFFERENT SENSITIVE TABULAR DATA PROTECTION METHODS USING A REAL LIFE TABLE STRUCTURE OF COMPLEX HIERARCHIES AND LINKS

Invited Paper

Prepared by Ramesh A. Dandekar, U. S. Department of Energy, United States of America

Comparative Evaluation of Four Different Sensitive Tabular Data Protection Methods Using a Real Life Table Structure of Complex Hierarchies and Links

Ramesh A Dandekar
Statistics and Methods Group, EI-70, U. S. Department of Energy, Washington DC 20585
Ramesh.dandekar@eia.doe.gov

Abstract The practitioners of tabular data protection methods in national statistical agencies have some familiarity with commonly used table structures. However, they require some guidance on how to evaluate appropriateness of various sensitive tabular data methods when applied to their own table structure. With that in mind, we use a real life "typical" table structure of moderate hierarchical and linked complexity and populate it with synthetic micro data to evaluate the relative performance of four different tabular data protection methods. The methods selected for the evaluation are: 1) lp-based classical cell suppression 2) lp-based CTA (Dandekar 2001), 3) network flow-based cell suppression as implemented in DiAna, a software product made available to other Federal statistical agencies by the US Census Bureau and 4) a micro data level noise addition method documented in a US Census Bureau research paper. The outcome from the comparative evaluation is available from http://mysite.verizon.net/vze7w8vk/

1 Introduction

To allow comparison of various sensitive tabular data protection methods on a consistent basis, the statistical disclosure control/limitation (SDC/SDL) researchers have long used public domain artificial (synthetic) data sets available from http://webpages.ull.es/users/casc/ website. The format used by these data sets, however, fails to convey visualization aspects of inherent complexities associated with various structural details typical of public use tables. The practitioners of tabular data protection methods are usually familiar with their own table structures. However, they require some assistance to evaluate appropriateness of proposed SDL methods when applied to their own table structure. As a first step to get around this problem, in this paper we use a real life table structure of moderate hierarchical and linked complexity and populate it with artificial (synthetic non-real) data to evaluate the relative performance of four different tabular data protection methods.

2 Table Structure

The templates for the hypothetical linked tables containing hierarchical structure selected for the comparative evaluation are in the Appendix. Appendix A shows column headings of these tables. The rows of the actual tables show geography in a hierarchical structure. However, only US totals are shown in Appendix A. The tables appear as two separate three-dimensional tables: Table 1 "Volumes by Grade, Sales Type, PAD District, and State" and Table 2 "Volumes by Formulation, Sales Type, PAD District, and State". For analytical purposes, these two tables could be considered as two three-dimensional linked portions of a four-dimensional table with

missing two-way interactions between grade and formulation. The tables consist of two independent (separable) components, namely, “Sales to End Users” and “Sales for Resale”. The later component offers a far greater challenge for the sensitive data protection task and therefore is selected for the comparative evaluation.

The four-dimensional table template, without “Sales to End Users” stand alone part of the published table, is populated with artificial micro-data to create this example. The resulting table contains a total of 1556 non-zero cells. The p percent rule with p=10% is used to identify sensitive tabular cells. There are a total of 78 sensitive cells requiring protection from statistical disclosure. Appendix C illustrates the size relationship of sensitive cells and non-sensitive cells in the table. Appendix B is in two parts and contains only a partial listing of Table 1 to illustrate the format used to display the outcome from the four different data protection methods. The entire populated table structure containing artificial data is available in the public domain to SDL researchers from the web site http://mysite.verizon.net/vze7w8vk/ . Part 1 of Appendix B, displays the outcome from classical lp-based cell suppression method in first four columns. Sensitive cells are identified by a symbol ‘w’. Non sensitive cells requiring suppression are identified by a symbol ‘s’. The last four columns of the table display the cell value adjustments from the CTA method. In these columns, the controlled tabular adjustment values to sensitive cells have been shown by a symbol ‘w’. Adjustments to non-sensitive values are displayed by using symbol ‘A’. Similarly, part 2 of Appendix B displays the outcome from DiAna software (network flow model) in the first four columns. Sensitive cells are identified by the symbol ‘p’. Suppressed non-sensitive cells are identified by symbol ‘c’. The last four columns of the table display the adjustments to cell value from the noise method.

3 Tabular Data Protection Methods

The methods selected for the comparative evaluation are: 1) classical lp-based cell suppression 2) lp-based CTA 3) DiAna’s network flow-based cell suppression and 4) micro data level noise addition method described in a US Census Bureau research paper.

The classical lp-based cell suppression method used for the evaluation is similar to that used by CONFID at Statistics Canada since the mid-80. The selection of the complementary cell suppression pattern is done by using a cost proportional to the table cell value as an objective function. This results in higher preference for smaller tabular cells as complementary suppression cells.

The controlled tabular adjustments (CTA) a.k.a. synthetic tabular data method used is the one documented in Dandekar (2001) and Dandekar/Cox (2002) . Large size non-sensitive table cells are targeted for adjustments by using a cost function which is a reciprocal of the table cell value. Such an approach results in relatively small

percentage changes in the cell values and therefore, reduces the overall degradation in the accuracy of the statistical information imbedded in table cell values.

The network flow model in the DiAna software uses a minimal cost flow (mcf) based algorithm from the University of Texas to develop a complementary cell suppression pattern. The PC version of the software used for this evaluation targets smaller sized cells to develop a complementary cell suppression pattern.

The micro data level noise addition method as described in the paper http://www.census.gov/srd/papers/pdf/bte9601.pdf is used for this evaluation. Micro data is perturbed by an average of 10% and standard deviation of 0.005 by using a normal distribution.

4 Comparative Evaluation – Cell Suppression Methods

Complementary cell suppression methods have been used by statistical agencies for many years. Both network flow (DiAna) and classical simplex-based linear programming (Statistics Canada) methods have been used to develop cell suppression patterns. There are pros and cons associated with both methods. Network flow methods are computationally far more efficient than simplex based LP methods and therefore are preferred for large tasks. Auditing of a cell suppression pattern to identify potential problems arising from either insufficient or lack of protection from disclosure is a recommended follow-up procedural step to both cell suppression methods.

Our comparative evaluation of the two suppression methods shows that the DiAna's network flow based procedure results in 479 cells (31% of total non-zero cells) being suppressed. The classical LP based procedure results in 294 cell (19% of total non-zero cells) suppressions. The suppression count includes 78 sensitive cells. A relatively large number of cell suppressions associated with the network flow model is due to the sequential "one two-dimensional section at a time" procedure used by the network flow model. The software also lacks the capability to identify and remove un-necessary secondary cell suppressions. The classical LP-based procedure in the first pass suppresses 321 cells. The second pass through the procedure, which is commonly referred to as a "clean-up" procedure, reduces the suppressions to 294.

5 Comparative Evaluation – Noise vs CTA

The ultimate objective of the noise method and the CTA method is to protect the sensitive tabular data by a sufficient distortion of sensitive tabular cell values without adversely affecting the overall quality of the published non-sensitive tabular cells. The noise method takes an ***indirect approach*** in an ***attempt*** to achieve that objective by a systematic distortion of related micro data records. The CTA method, on the

other hand, takes a ***direct approach*** to achieve that objective by first adjusting the values for sensitive tabular cells by a ***precise*** amount determined by use of the linear cell sensitivity rule. The non-sensitive tabular cells are adjusted "minimally" by using some predetermined criteria. For the noise method, there is no known systematic procedure to determine a direct one-to-one mathematical/statistical relationship between micro data distributional characteristics and the highly aggregated multi-variate public use table structure.[1] As a result of the "ad hoc" nature of the noise method, it does not guarantee enough distortion (therefore, protection from statistical disclosure) of sensitive tabular cells. The noise method also results in unnecessary changes in values for non-sensitive tabular cells. In theory, one advantage of applying methods, like noise addition, directly to micro data is that all tables produced from the micro data will be protected. This would preclude the need for table specific analysis required of the other methods. However, in practice extensive quality control measures are required to ensure adequate protection from statistical disclosure of sensitive tabular cells and to avoid excessive adjustments to non-sensitive tabular cells. We have used the histogram of cell count by percent change in cell value to evaluate relative performance of the noise and the CTA method when applied to two 3-D Linked tables.

CTA vs NOISE - TABULAR DATA QUALITY

CTA frequency Distribution

% From	% To	Non-Sensitive	Sensitive
.00 -	.10	1235	0
.10 -	.50	137	1
.50 -	1.00	60	0
1.00 -	1.50	15	0
1.50 -	2.00	13	1
2.00 -	5.00	15	50
5.00 -	10.00	3	26
10.00 -	15.00	0	0
15.00 -	30.00	0	0
30.00 -	100.00	0	0

1432

Noise Frequency Distribution

% From	% To	Nonsensitive	Sensitive
.00-	.10	96	1
.10-	.50	272	0
.50-	1.00	265	0
1.00-	1.50	215	0
1.50-	2.00	164	0
2.00-	5.00	439	2
5.00-	10.00	27	51
10.00-	15.00	0	24
15.00-	30.00	0	0
30.00-	100.00	0	0

633

Based on 1% or less error as good data quality acceptance criteria, the CTA procedure provides 1432 (92% of total non-zero cells) good quality cells. The noise method, on the other hand, provides 633 (41% of total non-zero cells) good quality cells. Based on these statistics, it is clear that the CTA outperforms noise-based cell perturbation.

---

[1] Highly disaggregated multi-variate table structure in "limiting case" approaches related micro-data and therefore exhibits micro-data characteristics.

6 Comparative Evaluation – Cell Suppression vs Perturbation

Ease of implementation issues aside, in general in addition to protecting sensitive information, the overall objective for the cell suppression method is to "minimize" the information loss. Cell perturbation methods such as CTA or the noise method, on the other hand, are implemented to provide overall high quality information to the end users after adequately protecting imbedded sensitive information. Due to such inherent differences in the strategy, the cell suppression methodology usually targets smaller cells for complementary suppression, while perturbations are "***preferred***" to be targeted on adjusting larger non-sensitive cells. Such a preferential criterion is easy to implement in the CTA method by using an appropriate selection of the objective function. The noise method, unfortunately, does not allow for preferential treatment of tabular cells. This is further confirmed by comparing across four methods selected for the evaluation.

The network flow method performs better than the noise method (69% published cells vs. 41% good quality cells). The CTA method performs better than the classical lp-based cell suppression method (92% good quality cells vs. 81% published cells)

7 Expanding Table Structure to Include Missing Two Way Interactions

If for whatever reason, the agency decides to include for publication a missing two-way interaction between grades and formulation in this example, it would need to create and protect four-dimensional table structures. In Appendix D we provide a summary performance statistics related to four different tabular data protection methods when used on two 3-D linked tables and one 4-D table structure. Based on the summary performance statistics, the relative ranking of the four methods selected for the evaluation remains the same. The detailed output for the 4-D table is available on the website http://mysite.verizon.net/vze7w8vk/.

7 Dream or Reality?

In an ideal situation, a cell suppression method of choice should have a computational speed which is typical of network flow models and should create a cell suppression pattern which is typical of classical lp-based cell suppression methods. A preliminary research performed by this author, during a time frame from 1996 to 1997, shows that lp-based shrinking hypercube method (abstract Dandekar 2002) has a potential to offer such an alternative. In the table below we provide a comparative evaluation of multiple exploratory runs from the lp-based shrinking hypercube method, targeted towards suppressing smaller non-sensitive tabular cells.

LP_BASED HYPERCUBE RELATIVE TO CLASSICAL LP BASED SUPPRESSION

SUPPRESSION METHOD	CELL COUNT	QUANTITY SUPPRESSED
CLASICAL LP	294	886128
HYP(4,2,0.1)	287	954602
HYP(2,5,0.5)	319	1037975
HYP(1,1,0.1)	302	961529
HYP(4,0,0.5)	292	1085394

Conclusion

In this paper we have evaluated the outcome from four different tabular data protection methods by using a common table structure of moderate hierarchical and linked complexity. Our comparative evaluation ranks the CTA highest, followed by classical lp-based cell suppression in second place, network flow based method in third place and the noise based procedure in last place.

The choice of an "***appropriate***" tabular data protection method depends on multiple factors. Factors, such as available technical skills and resources play a critical role in the selection of a method of choice for a statistical agency. We hope that the information presented in this paper will be useful for the statistical agencies in deciding on the appropriateness of their selected tabular data protection method.

References

Dandekar R. A. (2001) "Synthetic Tabular Data: A Better Alternative To Complementary Data Suppression - Original Manuscript Dated December 2001". Energy Information Administration, U. S. Department of Energy. Also available from CENEX-SDC Project International Conference, PSD2006, Rome, Italy, December 13-15, 2006, Companion CD Proceedings ISBN: 84-690-2100-1.

Dandekar R. A. and Cox L. H. (2002), Synthetic Tabular Data: An Alternative to Complementary Cell Suppression, 2002. Manuscript, Energy Information Administration, U. S. Department of Energy.

Dandekar, R.A (2003), Cost Effective Implementation of Synthetic Tabulation (a.k.a. Controlled Tabular Adjustments) in Legacy and New Statistical Data Publication Systems, working paper 40, UNECE Work session on statistical data confidentiality (Luxembourg, 7-9 April 2003)

Dandekar Ramesh A. (2004), Maximum Utility-Minimum Information Loss Table Server Design for Statistical Disclosure Control of Tabular Data, pp 121-135, Lecture Notes in Computer Science, Publisher: Springer-Verlag Heidelberg, ISSN:

0302-9743, Volume 3050 / 2004**,** Title: Privacy in Statistical Databases: CASC Project International Workshop, PSD 2004, Barcelona, Spain, June 9-11, 2004.

Fischetti, M. and J. J. Salazar (2000), "Models and Algorithms for Optimizing Cell Suppression Problem in Tabular Data with Linear Constraints", *Journal of the American Statistical Association* **95**, 916-928.

Evans T., Zayatz L., Slanta j. (1998),"Using Noise for Disclosure Limitation of Establishment Data", USBC paper available from http://www.census.gov/srd/papers/pdf/bte9601.pdf

Appendix A—Templates, Column Headings From Two Linked Tables, Used With Geography

Table 1: Volumes by Grade, Sales Type, PAD District, and State
(Thousand Gallons per Day)

Geographic Area Month	Regular						Midgrade					
	Sales to End Users		Sales for Resale				Sales to End Users		Sales for Resale			
	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total
United States												
November 2006	45,109.5	46,307.2	34,984.4	195,329.6	31,904.2	262,218.1	6,049.4	6,116.4	2,351.5	11,172.9	–	13,524.4
October 2006	46,308.9	47,471.1	33,856.4	194,761.3	35,954.4	264,572.0	6,174.4	6,245.6	2,244.1	11,318.4	–	13,562.4
November 2005	47,001.6	48,145.3	34,836.8	190,758.1	45,057.6	270,652.6	5,602.7	5,678.0	2,779.3	12,955.6	–	15,734.9

Table 1 continue .... Volumes by Grade, Sales Type, PAD District, and State
(Thousand Gallons per Day) — Continued

Geographic Area Month	Premium						All Grades					
	Sales to End Users		Sales for Resale				Sales to End Users		Sales for Resale			
	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total
United States												
November 2006	5,014.9	5,154.0	7,381.6	20,803.2	2,722.2	30,907.0	56,173.8	57,577.5	44,717.5	227,305.7	34,626.3	306,649.5
October 2006	5,118.0	5,256.2	7,236.7	20,719.1	2,945.0	30,900.7	57,601.3	58,972.9	43,337.2	226,798.7	38,899.3	309,035.2
November 2005	5,123.1	5,271.3	7,469.8	20,249.6	2,177.3	29,896.7	57,727.4	59,094.7	45,085.9	223,963.3	47,234.9	316,284.1

Table 2: Gasoline Volumes by Formulation, Sales Type, PAD District, and State
(Thousand Gallons per Day)

Geographic Area Month	Conventional						Oxygenated					
	Sales to End Users		Sales for Resale				Sales to End Users		Sales for Resale			
	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total
United States												
November 2006	31,587.1	32,542.6	7,582.5	164,118.7	32,289.1	203,990.3	2,356.3	2,456.6	1,987.9	9,707.9	–	11,695.8
October 2006	32,839.1	33,779.2	7,805.6	164,138.6	35,224.7	207,168.9	1,849.7	1,947.0	1,330.7	9,054.5	–	10,385.2
November 2005	32,562.5	33,530.7	8,923.5	162,294.3	38,193.8	209,411.6	2,752.0	2,842.1	1,933.7	8,512.0	–	10,445.7

Table 2: continue .... Volumes by Formulation, Sales Type, PAD District, and State
(Thousand Gallons per Day) — Continued

Geographic Area Month	Reformulated						All Formulations					
	Sales to End Users		Sales for Resale				Sales to End Users		Sales for Resale			
	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total	Through Retail Outlets	Total[a]	DTW	Rack	Bulk	Total
United States												
November 2006	22,230.5	22,578.3	35,147.2	53,479.1	2,337.2	90,963.5	56,173.8	57,577.5	44,717.5	227,305.7	34,626.3	306,649.5
October 2006	22,912.6	23,246.7	34,200.9	53,605.6	3,674.6	91,481.1	57,601.3	58,972.9	43,337.2	226,798.7	38,899.3	309,035.2
November 2005	22,412.9	22,721.9	34,228.7	53,157.0	9,041.1	96,426.8	57,727.4	59,094.7	45,085.9	223,963.3	47,234.9	316,284.1

Appendix B

Part 1 – Complete Table available at: http://mysite.verizon.net/vze7w8vk/tableofcontents.pdf

Classical LP-Based Cell Suppresion vs CTA

Classical LP/CTA 01 regular

	DTW	Rack	Bulk	Total	←--- CTA	Solution	--→	
United States	188668.0	218471.0	170021.0	577160.0	-130.A	113.A	-61.A	-78.A
PAD District I	64625.0	72994.0	65620.0	203239.0	-8.A	69.A	-143.A	-82.A
Subdistrict IA	25314.0	28780.0	16952.0	71046.0	-8.A	8.A	0.	0.
Connecticut	6258.0	1494.0	1700.0	9452.0	0.	0.	0.	0.
Maine	3936.0	4719.0	4429.0	13084.0	0.	0.	0.	0.
Massachusetts	172.0 w	3840.0 s	.0	4012.0	-8.w	8.A	0.	0.
New Hampshire	7879.0	.0	3188.0	11067.0	0.	0.	0.	0.
Rhode Island	1748.0 s	6224.0 s	3976.0	11948.0	0.	0.	0.	0.
Vermont	5321.0	12503.0	3659.0	21483.0	0.	0.	0.	0.
Subdistrict IB	19417.0	16493.0	22335.0	58245.0	0.	48.A	-61.A	-13.A
Delaware	6978.0	2400.0	4272.0	13650.0	0.	48.A	0.	48.A
District of Columbia	2253.0	5070.0	11338.0	18661.0	0.	0.	0.	0.
Maryland	3311.0	1836.0 s	1079.0 w	6226.0	0.	0.	-60.w	-60.A
New Jersey	6875.0	.0	144.0	7019.0	0.	0.	0.	0.
New York	.0	648.0 s	784.0 w	1432.0	0.	0.	-39.w	-39.A
Pennsylvania	.0	6539.0	4718.0	11257.0	0.	0.	38.A	38.A
Subdistrict IC	19894.0	27721.0	26333.0	73948.0	0.	13.A	-82.A	-69.A
Florida	.0	10857.0	1847.0	12704.0	0.	0.	-17.A	-17.A
Georgia	9961.0	.0	.0	9961.0	0.	0.	0.	0.
North Carolina	2268.0 s	7226.0 s	8464.0 s	17958.0	0.	13.A	-65.A	-52.A
South Carolina	1195.0	5887.0	7582.0	14664.0	0.	0.	0.	0.
Virginia	3560.0	.0	3625.0	7185.0	0.	0.	0.	0.
West Virginia	2910.0 s	3751.0 s	4815.0 s	11476.0	0.	0.	0.	0.
PAD District II	76174.0	62147.0	54796.0	193117.0	-71.A	0.	126.A	55.A
Illinois	4128.0	.0	.0	4128.0	0.	0.	0.	0.
Indiana	4613.0 s	.0	3846.0 s	8459.0 s	-14.A	0.	14.A	0.
Iowa	1149.0	4196.0	4216.0 s	9561.0 s	0.	0.	0.	0.
Kansas	11996.0	10330.0	1948.0	24274.0	-57.A	0.	112.A	55.A
Kentucky	5826.0 s	2787.0 s	6523.0	15136.0	0.	0.	0.	0.
Michigan	2022.0 s	.0	6668.0 s	8690.0	0.	0.	0.	0.
Minnesota	6400.0	3694.0	1332.0	11426.0	0.	0.	0.	0.
Missouri	5915.0	10385.0	3934.0	20234.0	0.	0.	0.	0.
Nebraska	2652.0	7667.0	942.0	11261.0	0.	0.	0.	0.
North Dakota	4671.0	8286.0	.0	12957.0	0.	0.	0.	0.
Ohio	7197.0	.0	3477.0	10674.0	0.	0.	0.	0.
Oklahoma	4030.0	1864.0	4339.0	10233.0	0.	0.	0.	0.
South Dakota	24.0	11013.0	5526.0	16563.0	0.	0.	0.	0.
Tennessee	2242.0	645.0	8325.0	11212.0	0.	0.	0.	0.
Wisconsin	13309.0	1280.0 s	3720.0 s	18309.0	0.	0.	0.	0.
PAD District III	15248.0	23726.0	26417.0	65391.0	0.	-19.A	0.	-19.A
Alabama	3504.0	259.0 w	2856.0 s	6619.0	0.	25.w	0.	25.A
Arkansas	1598.0	5628.0	6358.0	13584.0	0.	0.	0.	0.
Louisiana	.0	3088.0 s	4667.0 s	7755.0	0.	0.	0.	0.
Mississippi	666.0	8925.0	2980.0	12571.0	0.	0.	0.	0.
New Mexico	8410.0	4928.0	6696.0	20034.0	0.	0.	0.	0.
Texas	1070.0	898.0 w	2860.0 s	4828.0	0.	-44.w	0.	-44.A
PAD District IV	13561.0	23112.0	8479.0	45152.0	-51.A	132.A	-44.A	37.A
Colorado	.0	8772.0	5637.0	14409.0	0.	0.	0.	0.
Idaho	925.0 s	940.0 w	890.0 w	2755.0 s	0.	94.w	-44.w	50.A
Montana	514.0 w	7358.0 s	.0	7872.0 s	-51.w	0.	0.	-51.A
Utah	5676.0 s	382.0 w	.0	6058.0 s	0.	38.w	0.	38.A
Wyoming	6446.0 s	5660.0 s	1952.0 s	14058.0	0.	0.	0.	0.
PAD District V	19060.0	36492.0	14709.0	70261.0	0.	-69.A	0.	-69.A
Alaska	.0	7948.0	4300.0	12248.0	0.	0.	0.	0.
Arizona	2721.0	828.0	2189.0	5738.0	0.	0.	0.	0.
California	3792.0	3728.0	2251.0	9771.0	0.	-69.A	0.	-69.A
Hawaii	1038.0	6141.0	327.0	7506.0	0.	0.	0.	0.
Nevada	2555.0	3522.0	.0	6077.0	0.	0.	0.	0.
Oregon	.0	14325.0	3040.0	17365.0	0.	0.	0.	0.
Washington	8954.0	.0	2602.0	11556.0	0.	0.	0.	0.

Appendix B

Part 2- Complete Table available at: http://mysite.verizon.net/vze7w8vk/tableofcontents.pdf

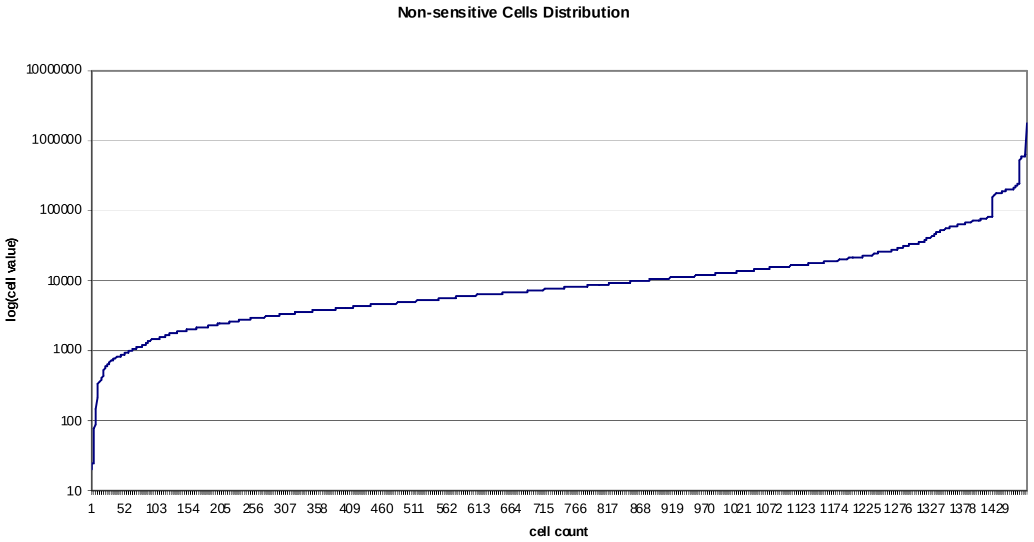
DiAna Cell Suppression Pattern vs Noise Addition

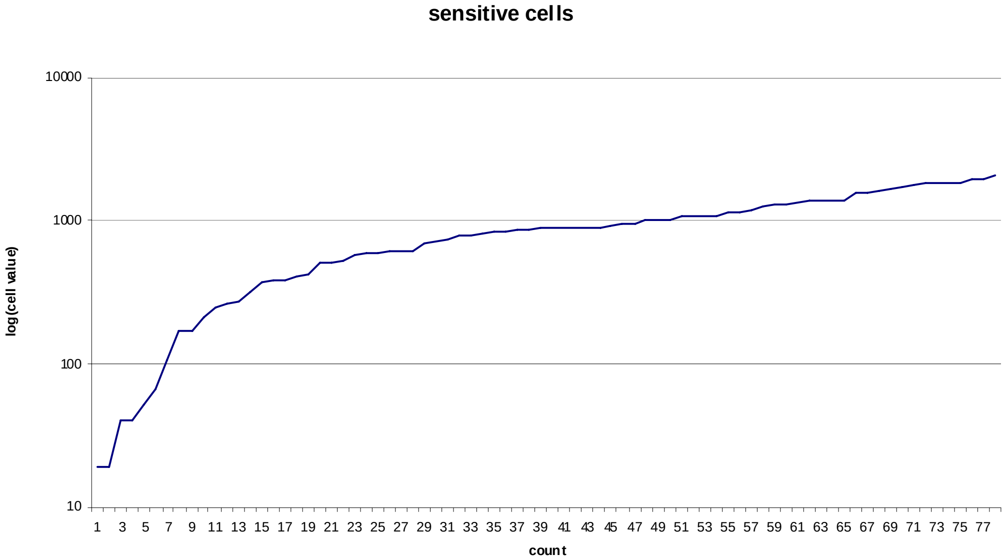
Diana/Noise 01 regular

	DTW		Rack		Bulk		Total		<--- Census' Noise Method-->			
United States	188668.		218471.		170021.		577160.		-16.	357.	81.	422.
PAD District I	64625.		72994.		65620.		203239.		-144.	219.	-53.	22.
Subdistrict IA	25314.		28780.		16952.		71046.		-267.	-115.	83.	-300.
Connecticut	6258.	C	1494.	C	1700.		9452.		-161.	-26.	-35.	-222.
Maine	3936.		4719.		4429.		13084.		150.	-69.	64.	145.
Massachusetts	172.	P	3840.	C	.		4012.		-9.	-65.	0.	-74.
New Hampshire	7879.		.		3188.		11067.		-159.	0.	81.	-78.
Rhode Island	1748.		6224.		3976.		11948.		-53.	143.	79.	168.
Vermont	5321.	C	12503.	C	3659.		21483.		-34.	-98.	-106.	-238.
Subdistrict IB	19417.		16493.		22335.		58245.		123.	51.	-137.	38.
Delaware	6978.	C	2400.	C	4272.		13650.		133.	19.	47.	200.
District of Columbia	2253.		5070.		11338.		18661.		-53.	85.	86.	118.
Maryland	3311.	C	1836.	C	1079.	P	6226.		-120.	-46.	-110.	-275.
New Jersey	6875.	C	.		144.	C	7019.		163.	0.	5.	168.
New York	.		648.	C	784.	P	1432.		0.	22.	-41.	-19.
Pennsylvania	.		6539.		4718.		11257.		0.	-30.	-124.	-154.
Subdistrict IC	19894.		27721.		26333.		73948.		0.	283.	1.	284.
Florida	.		10857.	C	1847.	C	12704.		0.	15.	-15.	1.
Georgia	9961.		.		.		9961.		93.	0.	0.	93.
North Carolina	2268.		7226.	C	8464.	C	17958.		94.	205.	-49.	250.
South Carolina	1195.		5887.		7582.		14664.		25.	-5.	165.	186.
Virginia	3560.		.		3625.		7185.		-187.	0.	-112.	-298.
West Virginia	2910.		3751.	C	4815.	C	11476.		-26.	68.	10.	52.
PAD District II	76174.		62147.		54796.		193117.		12.	186.	22.	220.
Illinois	4128.		.		.		4128.		154.	0.	0.	154.
Indiana	4613.		.		3846.	C	8459.	C	158.	0.	-112.	46.
Iowa	1149.	C	4196.	C	4216.	C	9561.		50.	-49.	60.	62.
Kansas	11996.		10330.		1948.		24274.		-62.	-156.	21.	-197.
Kentucky	5826.	C	2787.	C	6523.	C	15136.		106.	61.	-73.	94.
Michigan	2022.	C	.		6668.	C	8690.	C	73.	0.	-158.	-85.
Minnesota	6400.		3694.		1332.		11426.		-145.	110.	37.	2.
Missouri	5915.		10385.		3934.		20234.		-195.	46.	13.	-136.
Nebraska	2652.		7667.		942.		11261.		-80.	148.	32.	100.
North Dakota	4671.		8286.		.		12957.		30.	-18.	0.	12.
Ohio	7197.		.		3477.		10674.		-101.	0.	-40.	-141.
Oklahoma	4030.		1864.		4339.		10233.		46.	-52.	82.	76.
South Dakota	24.		11013.		5526.		16563.		1.	75.	94.	170.
Tennessee	2242.	C	645.	C	8325.		11212.		-70.	-14.	157.	73.
Wisconsin	13309.		1280.		3720.		18309.		47.	35.	-91.	-9.
PAD District III	15248.		23726.		26417.		65391.		30.	-233.	181.	-22.
Alabama	3504.	C	259.	P	2856.		6619.		-94.	-26.	-51.	-171.
Arkansas	1598.		5628.		6358.		13584.		11.	-43.	80.	48.
Louisiana	.		3088.	C	4667.	C	7755.		0.	-21.	115.	94.
Mississippi	666.	C	8925.	C	2980.		12571.		-23.	-218.	-155.	-396.
New Mexico	8410.		4928.		6696.		20034.		158.	121.	111.	389.
Texas	1070.	C	898.	P	2860.	C	4828.		-22.	-46.	81.	14.
PAD District IV	13561.		23112.		8479.		45152.		-106.	-133.	228.	-11.
Colorado	.		8772.		5637.		14409.		0.	-116.	129.	13.
Idaho	925.	C	940.	P	890.	P	2755.	C	-19.	100.	49.	131.
Montana	514.	P	7358.	C	.		7872.		53.	-23.	0.	30.
Utah	5676.	C	382.	P	.		6058.		-99.	-39.	0.	-138.
Wyoming	6446.		5660.	C	1952.	C	14058.	C	-41.	-54.	49.	-46.
PAD District V	19060.		36492.		14709.		70261.		192.	318.	-297.	213.
Alaska	.		7948.		4300.		12248.		0.	36.	-87.	-51.
Arizona	2721.	C	828.	C	2189.	C	5738.		82.	29.	-67.	44.
California	3792.		3728.	C	2251.	C	9771.		-84.	97.	-61.	-48.
Hawaii	1038.	C	6141.		327.	C	7506.		76.	-196.	11.	-108.
Nevada	2555.	C	3522.	C	.		6077.		-19.	56.	0.	37.
Oregon	.		14325.		3040.		17365.		0.	297.	-64.	233.
Washington	8954.		.		2602.		11556.		136.	0.	-30.	106.

Appendix C

Linked Tables 1 and 2 Cell Distribution





APPENDIX D

SUMMARY PERFORMANCE STATISTICS

TWO 3-D LINKED TABLES — **4-D ENTIRE TABLE**

NETWORK FLOW MODEL

Two 3-D Linked Tables:

- 28 Column Variables
- 15 Column Relations
- 78 Sensitive Cells
- 479 Suppressions
- 31% Suppressions
- 1077 Published

4-D Entire Table:

- 64 Column Variables
- 48 Column Relations
- 267 Sensitive Cells
- 1844 Suppressions
- 62% Suppressed
- 1149 Published

CLASSICAL CELL SUPPRESSION

Two 3-D Linked Tables:

- 1556 Non-Zero Cells
- 2707 Equations
- 78 Sensitive Cells
- 294 Suppressions
- 19% Suppressions
- 1282 Published

4-D Entire Table:

- 2993 Non-Zero Cells
- 6273 Equations
- 267 Sensitive Cells
- 1143 Suppressions
- 38% Suppressions
- 1850 Published

CONTROLLED TABULAR ADJUSTMENT

Two 3-D Linked Tables:

% From	% To	Nonsensitive	Sensitive
.00-	.10	1235	0
.10-	.50	137	1
.50-	1.00	60	0
1.00-	1.50	15	0
1.50-	2.00	14	1
2.00-	5.00	14	50
5.00-	10.00	3	26
10.00-	15.00	0	0
15.00-	30.00	0	0
30.00-	100.00	0	0
1432 Good Quality		124 Poor Quality	

4-D Entire Table:

% From	% To	Non-Sensitive	Sensitive
.00 -	.10	1803	0
.10 -	.50	438	1
.50 -	1.00	214	0
1.00 -	1.50	97	1
1.50 -	2.00	59	0
2.00 -	5.00	103	171
5.00 -	10.00	12	94
10.00 -	15.00	0	0
15.00 -	30.00	0	0
30.00 -	100.00	0	0
2455 Good Quality		538 Poor Quality	

MICRO DATA LEVEL NOISE ADDITION

Two 3-D Linked Tables:

% From	% To	Nonsensitive	Sensitive
.00-	.10	96	1
.10-	.50	272	0
.50-	1.00	265	0
1.00-	1.50	215	0
1.50-	2.00	164	0
2.00-	5.00	439	2
5.00-	10.00	72	51
10.00-	15.00	0	24
15.00-	30.00	0	0
30.00-	100.00	0	0
633 Good Quality		923 Poor Quality	

4-D Entire Table:

% From	% To	Nonsensitive	Sensitive
.00-	.10	137	1
.10-	.50	416	0
.50-	1.00	400	0
1.00-	1.50	334	0
1.50-	2.00	322	0
2.00-	5.00	1069	2
5.00-	10.00	48	172
10.00-	15.00	0	92
15.00-	30.00	0	0
30.00-	100.00	0	0
953 Good Quality		2040 Poor Quality	