

Practical Considerations in Selecting Statistical Disclosure Methodology for Tabular Data

BTS Confidentiality Seminar Series, March 2003



National Agricultural Statistics Service
Washington, D.C.

Examples in the Presentation are Based on Actual Occurrences



National Agricultural Statistics Service
Washington, D.C.

Data have been
changed to protect the
innocent



National Agricultural Statistics Service
Washington, D.C.

Goals:

- Discuss Decision Factors
- Discuss Alternatives
- Present Illustrations
- Provide a Perspective for Future Decisions



DECISION FACTORS - 1

What is your data structure?

Household, establishment, opinion, census data, sample results, etc...



National Agricultural Statistics Service
Washington, D.C.

DECISION FACTORS - 2

What are your population characteristics?

Limited variation, extremely clustered, highly skewed, etc...



DECISION FACTORS - 3

Who are your customers?

- Casual readers
- In-depth analysts
- People (companies) from the data population
- Etc...



DECISION FACTORS - 4

What decisions will be made by your customers?

- Social Research Studies
- Economic Investment Decisions
- Implementation of Government Programs
- Etc...



BASIC SUPPRESSION APPROACHES

- (n, p) Rules
- (p, k) Rules
- (n, p, t) Rules



Should You Publish Your Suppression Rule?

- Conventional wisdom says NO!
- It might be essential for certain situations



BASIC NASS APPROACHES

- (p, k) Rule for Ag Census
- (n, p) Rule for Periodic Surveys
- Special Techniques



SPECIAL TECHNIQUES

- Signed Release Agreements (Waivers)
- Geographic Combinations
- Size Group Combinations
- Reclassification of Categories



WAIVERS

- Large Operations
- Permission to Publish is Requested
- Must Agree in Writing
- Updated Annually



Waiver Example

Hypothetical Pecan Production Example 1/

<u>State</u>	<u>Production - 1,000lbs</u>
AL	15,000
GA	80,000
LA	18,000
NM	35,000
TX	30,000
All Other 2/	31,850
US	209,850

1/ Assumes One person controls all Arizona Production

2/ Includes AZ, AR, CA, FL, KS, MS, NC, OK, SC

Waiver Example

Pecan Production Example, Actual Results

<u>State</u>	<u>2000 Production/1,000lbs</u>
AL	15,000
AZ	14,500
AR	900
CA	3,400
FL	3,300
GA	80,000
KS	550
LA	18,000
MS	3,500
NM	35,000
NC	1,600
OK	2,500
SC	1,600
TX	30,000
US	209,850

Geographic Combinations

Hypothetical Olive Acreage

<u>States</u>	<u>Growers</u>	<u>Acres</u>
CA	1,363	37,714
FL	2	2
U.S. Total	1,365	37,716

Traditional Approach

CA	1,363	(D)
FL	2	(D)
U.S. Total	1,365	(D)

(D) Withheld to avoid disclosing data for individual farms.



Geographic Combinations

Handling hypothetical Olive data
Situation by moving Florida Olives to "Other"
Olive Acreage

<u>States</u>	<u>Growers</u>	<u>Acres</u>
CA	1363	37,714
U.S. Total	1363	37,714



Size Group Example

Percent of Farms, by Economic Sales Class, U.S. 2001

<u>Economic Sales Class</u>	<u>Percent of Total Farms</u>
\$1,000 - \$2,499	25.3
\$2,500 - \$4,999	15.3
\$5,000 - \$9,999	13.4
\$10,000 - \$24,999	12.5
\$25,000 - \$49,999	8.7
\$50,000 - \$99,999	8.6
\$100,000 - \$249,999	8.9
\$250,000 - \$499,999	4.1
\$500,000 - \$999,999	2.0
\$1,000,000+	1.2
Total	100.0

Size Group Example

Number of Farms: Economic Sales Class by State and Region, 2001

State and Region	Economic Sales Class		
	<u>\$1,000 - \$9,999</u>	<u>\$10,000 - \$99,999</u>	<u>\$100,000 & Over</u>
	<i>Number</i>	<i>Number</i>	<i>Number</i>
Northeast			
CT	2,450		
ME	4,200		
MA	3,200		
NH	2,100		
NJ	6,400		
NY	17,800	11,000	8,700
PA	32,800	16,500	9,700
RI	350		
VT	3,500		
Other States		10,600	3,800
Total	72,800	38,100	22,200

Reclassification Example

Hypothetical Potato Variety Data, 1 State

Fall Potatoes: Percent of Major Varieties Planted,

State and Varieties

Percent of Planted Acres

R Burbank	71.0
Ranger R	12.0
R Norkotah	7.5
Shepody	3.4
Silverton R	2.4
De Wolf	1.0
Other	2.7
TOTAL	100.0

However, assume that deWolf is a new variety planted only on one large farms

Reclassification Example

Actual Potato Variety Data, 1 State

Fall Potatoes: Percent of Major Varieties Planted,

State and Varieties

Percent of Planted Acres

R Burbank	71.0
Ranger R	12.0
R Norkotah	7.5
Shepody	3.4
Silverton R	2.4
Other	3.7
TOTAL	100.0

CASE STUDY

Agricultural Marketing Service
implementation of mandatory price
reporting law illustrates basic
decision factor principles.



OUTLINE

- Define the problem
- Explain the approach
- Describe the solution
- Present some results



**Large meat packers must
report details of all
purchase transactions.**



National Agricultural Statistics Service
Washington, D.C.

**Data must be reported by
fixed times each day.
AND
Summaries published
an hour later.**



**AMS must protect
confidentiality in issuing
summaries.**



National Agricultural Statistics Service
Washington, D.C.

**90% of slaughter
is in 114 plants, owned
by 65 companies.**



National Agricultural Statistics Service
Washington, D.C.

4 Companies:

- 80% of fed cattle
- 80% of fed lambs
- 55% of all hogs



**AMS adopted a 3/60
standard for each data
cell.**



National Agricultural Statistics Service
Washington, D.C.

**At least three companies
and no company
exceeding 60%.**



National Agricultural Statistics Service
Washington, D.C.

AMS never identified how many companies were in an aggregate.



National Agricultural Statistics Service
Washington, D.C.

**AMS did not distinguish
blank cells from
suppressed cells.**



National Agricultural Statistics Service
Washington, D.C.

**Many (24 %) daily reports
were suppressed plus
many cells in released
reports.**



**Some observers calculated
90% suppression.**



National Agricultural Statistics Service
Washington, D.C.

When in Doubt - Examine the Data!



National Agricultural Statistics Service
Washington, D.C.

SEQUENCE DAY

Plant	1	2	3	4	5	6	7	8	9	10	11	12
A	X			X	X	X			X	X		X
B								X				
C	X	X	X	X	X	X	X		X	X	X	X
D	X	X	X	X					X	X	X	
E				X		X			X			
F	X	X	X	X	X	X	X		X	X	X	X
G	X		X	X	X	X	X		X			
H	X		X	X	X	X	X	X	X	X	X	X
I			X									



When are proprietary data not unique?



National Agricultural Statistics Service
Washington, D.C.

**A new 3/70/20
confidentiality standard
based on 60-days of
data.**



National Agricultural Statistics Service
Washington, D.C.

**At least 3
companies operated
50% of the time.**



National Agricultural Statistics Service
Washington, D.C.

**No company had
more than 70
percent of the
volume.**



National Agricultural Statistics Service
Washington, D.C.

**No company would be
exposed more than
20% of the time.**



National Agricultural Statistics Service
Washington, D.C.

**Nearly all cattle daily
reports and all cells
now released.**



National Agricultural Statistics Service
Washington, D.C.

EXAMPLE TALLY

Cattle Open Market, One Region, Daily

3 or more

68%

Concen.

47%

Exposure

2%



EXAMPLE TALLY

Cattle Forward Contracts, National, Daily

3 or more	41%
Concen.	51%
Exposure	6%



Modification 1

Cattle forward
contracts are now
published weekly



National Agricultural Statistics Service
Washington, D.C.

Modification 2

All swine packer sales
alternatives weighted
together daily.



National Agricultural Statistics Service
Washington, D.C.

Modification 3

Lamb purchases
accumulated until the
3/70/20 is reached.



National Agricultural Statistics Service
Washington, D.C.

Author Contact Information

- Rich Allen
- Deputy Administrator, NASS
- Room 5905 South Building
- 1400 Independence Ave., S.W.
- Washington, D.C. 20250-2001
- 202-690-8141 Fax 202-690-1311
- rich_allen@nass.usda.gov



National Agricultural Statistics Service
Washington, D.C.