

# STATISTICAL SAMPLING METHODS USED IN OAS AUDITS

---

**NEW ENGLAND INTERGOVERNMENTAL AUDIT FORUM  
MEETING NOVEMBER 8-9, 2018**

WESTIN PORTLAND HARBORVIEW



# INTRODUCTIONS

---

➤ Presenter:

Stacie Last, National Statistical  
Specialist, HHS-OIG-OAS

➤ Moderator:

Stephen Conway, Director, Advanced  
Audit Techniques, HHS-OIG-OAS

# PRESENTER FROM OAS AT THE 2016 SPRING NEIAF:

---

OAS Statistician: Jared Smith, PhD, CIA



In the interest of disclosure,  
Stacie Last is NOT a PhD!

# SUMMARY OF THE OAS SAMPLING METHODOLOGY

---

During audit planning, the audit team makes the decision to use sampling to accomplish the objective(s) of the audit.

The Regional Statistical Specialist and/or the OAS Statistician then begin to assist the audit team during the following phases of the sampling methodology:

- I. Designing the sample
- II. Conducting the sample
- III. Interpreting the sample results

# I. DESIGNING THE SAMPLE (THE FOLLOWING STEPS ARE OFTEN DONE CONCURRENTLY.)

---

- **Creating and Refining the Sampling Frame**
- **Choosing a Sample Design**
- **Documenting the Sampling Methodology**

# CREATING AND REFINING THE SAMPLING FRAME

---

- **Sampling frame** – A listing of the items from which the sample will be drawn. Ideally, this list would only include items in the *target population* and no other items.
- **Target population** – Items relevant to the audit; group of items about which the audit team wants to make an estimate.
- **Ideally**, the sampling frame would be a listing all items in the target population and no other items.



# CREATING AND REFINING THE SAMPLING FRAME

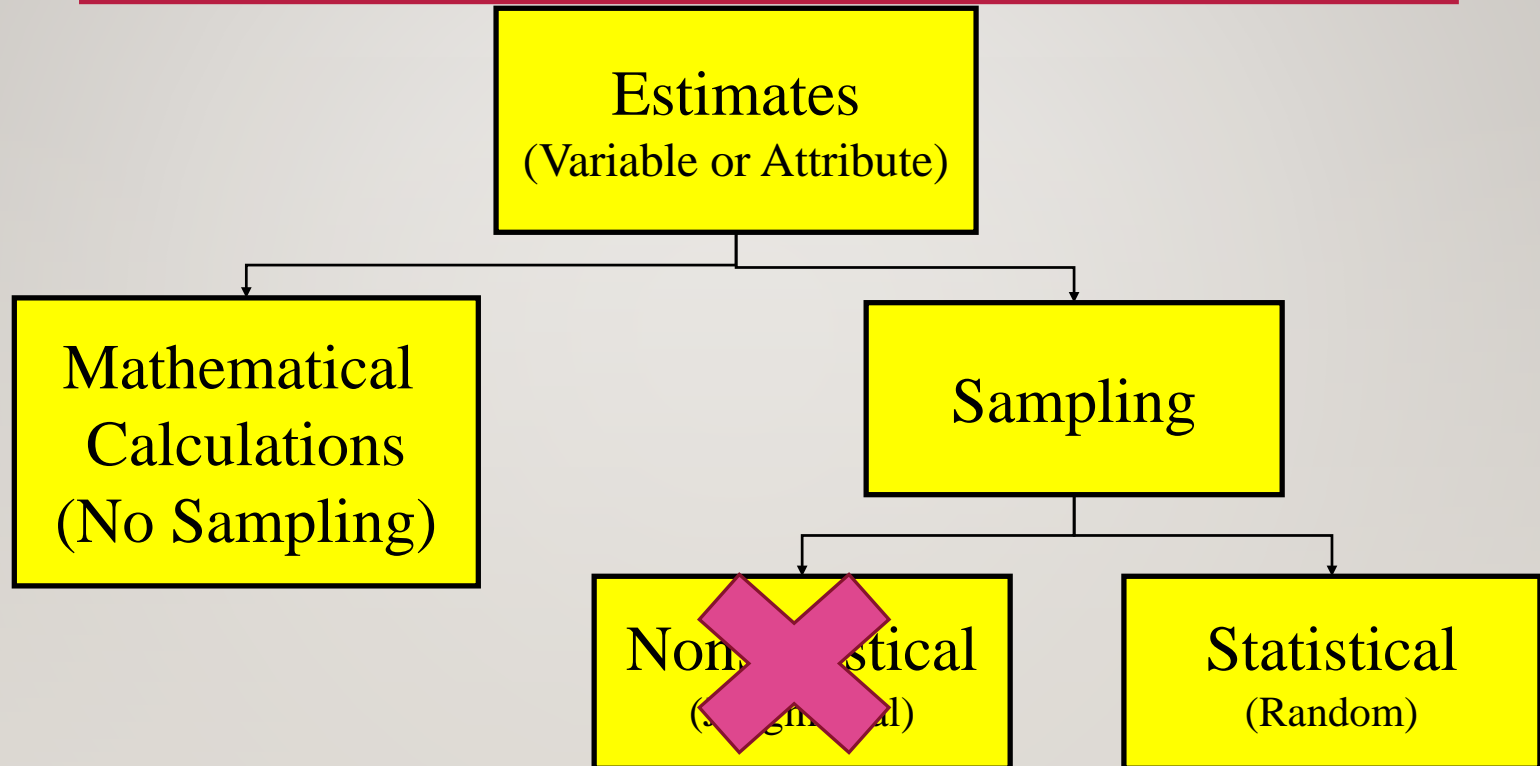
---

## Resolving Deficiencies in the Sampling Frame:

- Make sure the data used to create the sampling frame comes from a reliable source. (Is the data regularly tested and validated?)
- Make sure there are no duplicated sample items.
- If the sampling frame includes duplicated items or items that are not part of the target population, decide on a statistically valid way to handle such items if they are selected in the sample.
- Resolve deficiencies **BEFORE** pulling the sample.

# DESIGNING THE SAMPLE - DETERMINE THE TYPE OF ESTIMATE(S)

---



No estimates from nonstatistical samples.



# DESIGNING THE SAMPLE - AUDIT ESTIMATES

---

- Variable Estimate

Quantitative characteristic

(e.g., Federal dollars paid)

- Attribute Estimate

Rate or Proportion of a Characteristic

(e.g., Number of claims paid incorrectly)

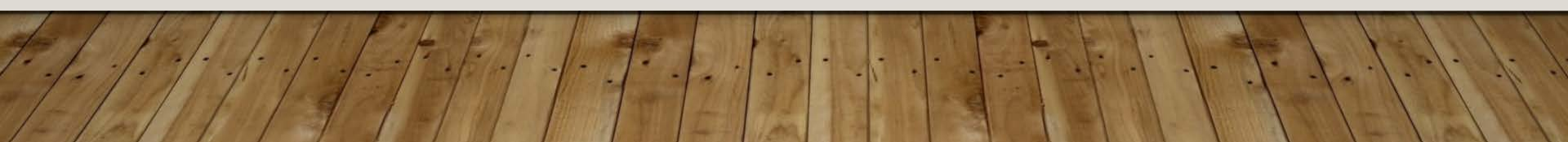
# DESIGNING THE SAMPLE

---

## Most Common Sample Designs:



- Simple (Unrestricted) Random Sampling
- Stratified Sampling

## Other Sample Designs OAS Uses:

- Multistage Sampling
  - Stratified Multistage
  - RHC Sampling
- 

# DESIGNING THE SAMPLE – WHAT TO CONSIDER

---

- Precision 
- Expected error rate
- Logistics not related to statistical sampling 

# DESIGNING THE SAMPLE – CHOOSING A SAMPLE SIZE

---

30 200 50  
120  
150 75 100

How Many is Enough?

# DESIGNING THE SAMPLE – THE STRATACHECK WORKSHEET

---

- Programmed in Excel by the OAS Statistician in December of 2014.
- Used by OAS audit teams since 2015.
- The purpose of the worksheet is to assess the efficiency of a stratified design compared to an unrestricted design.

# DESIGNING THE SAMPLE – THE STRATACHECK WORKSHEET

---

The assessment of efficiency is based on actual amounts that have not yet been audited, so there are some **assumptions** being made by the worksheet.

1. Error rates are the same **within a stratum**. For example, high-dollar items are not associated with lower error rates.
2. Sample items are either all correct or all in error; there are no partially correct sample items.
3. Underpayments, if any, make up an insignificant portion of the sampling frame.



# DESIGNING THE SAMPLE – THE STRATACHECK WORKSHEET

---

If some, but not all, of these assumptions are true, the worksheet can still be a useful guide for designing a sample.

However, if none of the assumptions are true, proportional allocation of sample items among stratum may be a better design.



# STRATACHECK WORKSHEET DEMONSTRATION

	A	B	C	D	E	F	G
1	<b>Unique ID (Optional)</b>	<b>Amount</b> <small>Clear Columns</small>	<b>Strata (Optional)</b>		<b>Number of Strata</b>		3
2		0.00	1		<b>Error Type</b>	All or Nothing	
3		0.00	1		<b>Possible Error Rate ( 0 to 1)</b>	0.5	
4		0.00	1		<b>Sample Size</b>	100	
5		0.00	1		<b>Sample Allocation Method</b>	Equal	
6		0.00	1		<b>Amount Source (Optional)</b>		
7		0.00	1				
8		0.00	1		Stratify for Attribute Estimate		
9		0.00	1		Stratify for Variable Estimate		
10		0.00	1				
11		0.00	1		Stratify for Att and Var Estimates		
12		0.00	1				
13		0.00	1				
14		0.01	1		Stratify Based on Labels in Column C		
15		0.01	1				

# QUESTIONS ON THE STRATACHECK WORKSHEET?

---



# DOCUMENTING THE SAMPLING METHODOLOGY: SAMPLING PLAN

---

- 1. Audit Objective
- 2. Target Population
- 3. Sampling Frame
- 4. Sample Unit
- 5. Survey Information
- 6. Sample Design
- 7. Sample Size
- 8. Source of Random Numbers
- 9. Method of Selecting Sample Units
- 10. Characteristics to Be Measured
- 11. Treatment of Missing Sample Units
- 12. Estimation Methodology
- 13. Other Evidence
- 14. Description of How Results Will Be Reported
- 15. Sources of Data
- 16. Assessment Made on Data Reliability

# DOCUMENTING THE SAMPLING METHODOLOGY: SAMPLING PLAN

---

- A statistically valid random sample is one that can be replicated.
- Replication of the sample is necessary to allow others (third parties) to verify the results.

## II. CONDUCTING THE SAMPLE

---

### Providing the Audit Team with the Random Numbers

- OAS uses “Random Number Sampling” to select samples:
  - Select random numbers
  - Match random numbers to corresponding items in sampling frame
  
- Sources of random numbers:
  - RAT-STATS random number generator
  - R software function



# RAT-STATS

---

- Initially developed in DOS (in the '80s) by Doug Rennie, an OAS AATS Audit Manager (retired).
  - RAT-STATS Windows version programmed by Dr. Al Kvanli, OAS Statistical Consultant.
  - RAT-STATS “508-compliant” version programmed by Contractor.
- Modules -
- “R” Software
    - Ratio and Regression Estimation
    - Discovery and Acceptance Sampling

## RAT-STATS Main Menu

RANDOM NUMBERS  
ATTRIBUTE APPRAISALS  
VARIABLE APPRAISALS  
SAMPLE SIZE DETERMINATION

### Random Numbers

SINGLE STAGE NUMBERS  
SETS OF TWO NUMBERS  
SETS OF THREE NUMBERS  
SETS OF FOUR NUMBERS  
FRAMES - SINGLE STAGE  
FRAMES - SETS OF TWO  
RHC SAMPLE SELECTION

### Attribute Appraisals

UNRESTRICTED  
STRATIFIED  
TWO-STAGE UNRESTRICTED  
THREE-STAGE UNRESTRICTED  
TWO-STAGE RHC  
THREE-STAGE RHC  
STRATIFIED CLUSTER  
STRATIFIED MULTISTAGE

### Variable Appraisals

UNRESTRICTED  
STRATIFIED  
TWO-STAGE UNRESTRICTED  
THREE-STAGE UNRESTRICTED  
TWO-STAGE RHC  
THREE-STAGE RHC  
STRATIFIED CLUSTER  
STRATIFIED MULTISTAGE  
POST STRATIFICATION  
UNKNOWN UNIVERSE SIZE

### Sample Size Determination

VARIABLE SAMPLE SIZE DETERMINATION  
• UNRESTRICTED  
• STRATIFIED  
ATTRIBUTE SAMPLE SIZE DETERMINATION

# RAT-STATS DEMONSTRATION

---

## **First Module:** Random Numbers

- The RAT-STATS Random Number Generator (RNG) has passed 13 certification program tests for various aspects of randomness.
- These tests were developed by the National Institute of Standards and Technology (NIST).

# RAT-STATS

## RANDOM NUMBER GENERATOR DEMO

---

RAT-STATS will prompt the user for the following information:

- A seed number (or RAT-STATS can provide one)
- Name of the audit/review
- Quantity of numbers to be generated (sample size)

*(continued on next slide)*



# RAT-STATS RANDOM NUMBER GENERATOR DEMO

---

## Information for the RAT-STATS Random Number Generator (continued):

- Numbering of the sampling frame (low number and high number)
- Where to save the random numbers – printer, text file, Access file, Excel file, or flat file

# RAT-STATS

## RANDOM NUMBER GENERATOR DEMO

---

### Other Options for Random Numbers:

- Sets of Two Numbers
- Sets of Three Numbers
- Sets of Four Numbers
- Frames – Single Stage
- Frames – Sets of Two
- RHC Sample Selection



# CONDUCTING THE SAMPLE

---

Once the random numbers have been drawn, the statistical specialist (or a member of the audit team approved by the statistical specialist) matches the random numbers to the sampling frame to select the sample.

The audit team can then continue conducting the sample by evaluating each sample unit and preparing the sample results.

# III. INTERPRETING SAMPLE RESULTS

---

## Includes:

- Using RAT-STATS to Calculate Estimates Based on Sample Results
- Interpreting the RAT-STATS Estimates
  - For the auditors
  - For the audit report

# RAT-STATS

## ATTRIBUTE APPRAISAL DEMO

---

RAT-STATS will prompt the user for the following information:

- Name of the audit/review
- Universe (sampling frame) size
- Sample size
- Number of items with the characteristic of interest (attribute)
- Where to save the attribute estimates

# RAT-STATS

## ATTRIBUTE APPRAISAL DEMO

### Example Output

```
Windows RAT-STATS
Statistical Software
Date: 4/28/2009      Single Stage Attribute Appraisal      Time: 12:45
                    AUDIT/REVIEW: AA Scenario1
OUTPUT FILE: C:\My Files\Files after C crash\RAT-STATS section of Day1 of
Basic\AA Scenario1 Output.txt

UNIVERSE SIZE                      1,000
SAMPLE SIZE                        30
CHARACTERISTIC(S) OF INTEREST
  QUANTITY IDENTIFIED IN SAMPLE      8
  PROJECTED QUANTITY IN UNIVERSE    267
  PERCENT                          26.667%
STANDARD ERROR
  PROJECTED QUANTITY                 81
  PERCENT                          8.088%

                    CONFIDENCE LIMITS

                    90% CONFIDENCE LEVEL
LOWER LIMIT - QUANTITY              142
              PERCENT              14.200%
UPPER LIMIT - QUANTITY              427
              PERCENT              42.700%
```

# RAT-STATS

## ATTRIBUTE APPRAISAL DEMO

---

### **Simple Example:** Audit of payments to an HMO for beneficiaries enrolled

- 1-year audit period
- Total monthly beneficiary payments = 100,000
- Sample unit = monthly beneficiary payment
- Sample size = 100 monthly beneficiary payments
- Characteristic of Interest = “unallowable payment”



# RAT-STATS

## ATTRIBUTE APPRAISAL DEMO

---

**Simple Example (continued):** The 100 randomly selected monthly beneficiary payments have been evaluated.

Twenty-five of the 100 payments were unallowable.



# RAT-STATS

## ATTRIBUTE APPRAISAL DEMO

---

### **Other Attribute Appraisal Options:**

- Stratified
- Two-stage Unrestricted
- Three-stage Unrestricted
- Two-stage RHC
- Three-stage RHC
- Stratified Cluster
- Stratified Multistage

# RAT-STATS VARIABLE APPRAISAL DEMO

---

**RAT-STATS will prompt the user for the following information:**

- Name of the audit/review
- Universe (sampling frame) size
- Data input file

# RAT-STATS

## VARIABLE APPRAISAL DEMO

---

### Format for Sample Results

<b>Sample Item</b>	<b>Paid (Examined)</b>	<b>Correct (Audited)</b>	<b>Unallowed (Difference)</b>
<b>1</b>	<b>\$276.23</b>	<b>\$276.23</b>	<b>\$0.00</b>
<b>2</b>	<b>365.39</b>	<b>0.00</b>	<b>365.39</b>
<b>3</b>	<b>430.37</b>	<b>0.00</b>	<b>430.37</b>

Variable Appraisal Input File

Variable Appraisal Output File



# RAT-STATS

## VARIABLE APPRAISAL OUTPUT

Windows RAT-STATS  
Statistical Software  
Date: 4/28/2009 VARIABLE UNRESTRICTED APPRAISAL Time: 13:59  
AUDIT/REVIEW: VA Scenario1

DATA FILE USED: C:\My Files\Files after C crash\RAT-STATS section of Day1 of  
Basic\VA Scenario1 Input.xls

SAMPLE SIZE	VALUE OF SAMPLE	NONZERO ITEMS
100	8,683.15	25

----- D I F F E R E N C E -----	
MEAN / UNIVERSE	86.83
STANDARD DEVIATION	159.96
SKWENESS	1.45
KURTOSIS	3.37
STANDARD ERROR (MEAN)	15.99
STANDARD ERROR (TOTAL)	1,598.754
POINT ESTIMATE	8,683,150

100,000

### CONFIDENCE LIMITS

80% CONFIDENCE LEVEL	
LOWER LIMIT	6,620,499
UPPER LIMIT	10,745,801
PRECISION AMOUNT	2,062,651
PRECISION PERCENT	23.75%
T-VALUE USED	1.290161442034

90% CONFIDENCE LEVEL	
LOWER LIMIT	6,028,593
UPPER LIMIT	11,337,707
PRECISION AMOUNT	2,654,557
PRECISION PERCENT	30.57%
T-VALUE USED	1.660391156017

95% CONFIDENCE LEVEL	
LOWER LIMIT	5,510,875
UPPER LIMIT	11,855,425
PRECISION AMOUNT	3,172,275
PRECISION PERCENT	36.53%
T-VALUE USED	1.984216951586

# RAT-STATS

## VARIABLE APPRAISAL DEMO

---

### **Stratified Variable Appraisal**

- Basis for stratification
- Number of strata

# RAT-STATS

## VARIABLE APPRAISAL DEMO

---

### **Stratified Variable Appraisal Input File**

- Number of items in each stratum
- Sample size for each stratum



# RAT-STATS

## VARIABLE APPRAISAL DEMO

---

### **Stratified Variable Appraisal Input File (cont.)**

- Worksheet showing number of items in each stratum frame and sample sizes per stratum (sample summary data)
- Worksheet showing sample results by stratum

# RAT-STATS

## VARIABLE APPRAISAL DEMO

---

### Sample Summary

Stratum	Sampling Frame Size	Sample Size
1	1000	30
2	100000	100

Variable Appraisal Input File

Variable Appraisal Output File

# RAT-STATS

## VARIABLE APPRAISAL DEMO

---

### **Other Variable Appraisal Options:**

- Two-stage Unrestricted
- Three-stage Unrestricted
- Two-stage RHC
- Three-stage RHC
- Stratified Cluster
- Stratified Multistage
- Poststratification
- Unknown Universe Size

# RAT-STATS

## SAMPLE SIZE DETERMINATION

---

- Variable Sample Size
  - Unrestricted
    - Using a Probe Sample
    - Using Estimated Error Rate
  - Stratified
- Attribute Sample Size

# INTERPRETING SAMPLE RESULTS

---

- The statistical specialist provides the audit team with the output files from the RAT-STATS appraisal(s).
- The statistical specialist also reviews the audit report before they are issued to ensure the sample results and appraisals are not misrepresented in the audit report.

---

QUESTIONS ?