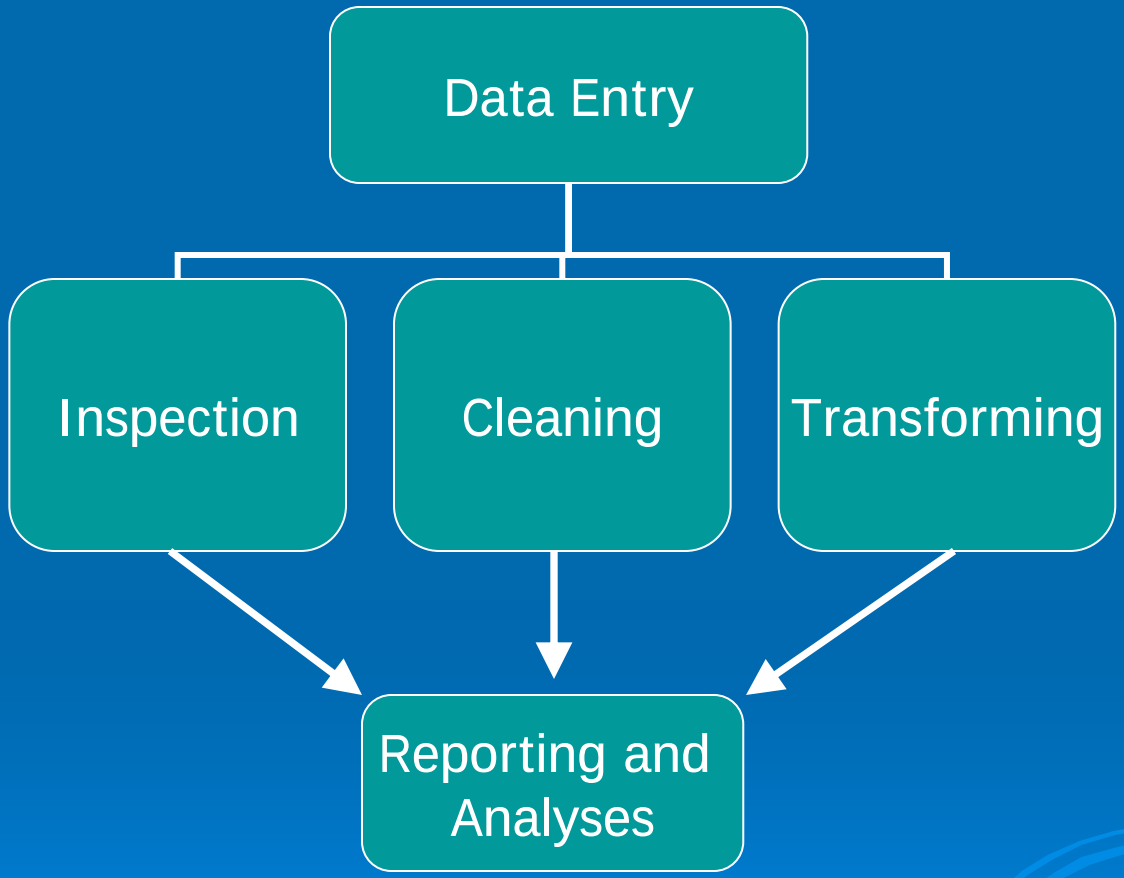
SAS Fundamentals For Survey Data Processing

Renato G. Villacorte

Fairbank, Maslin, Maullin & Associates

Opinion Research & Public Policy Analysis

Survey Data Processing



Sample Survey Questions

- Demographic Questions
  - Gender
  - Age
- Attitudinal Questions
  - Ideology or Political Identification
  - Outlook on the future
- Opinion Questions
  - Favorability or Job Approval
  - Importance or Satisfaction Ratings

Opinion Question Example

- Please tell me if you have a favorable, neutral, or unfavorable opinion of:
  - A. Anytown City Government
    1 – Very favorable
    2 – Somewhat favorable
    3 – Neutral
    4 – Somewhat unfavorable
    5 – Very unfavorable

Original Sample Data

Obs	SEX	AGE	Q4A	Q4B	Q4C	Q4D	Q4E	Q4F	Q4G	Q4H
1	1	8	3	4	3		4	2	3	3
2	1	11	3	4			4	1		
3	1	10	1	3	2	2		2	3	1
4	1	5	4	5	4		2	1	4	2
5	2	7	4	3	3	5	5	1	3	3
6	2	4	4	3				1		
7	2	7	3	1	3		3	1		
8	2	10	3	3	4	3	5	2	3	1
9	1	4	2	4	1	2	3	1	3	1
10	2	5	3			3		1		1
11	2	2	4	2				2		
12	2	10	3	3				1	2	2
13	2	1	2		2	2		2	4	2
14	2	5	5	5	2	2	3	5	1	5
15	1	5	3	4	3	3	3	2	3	3

Transform Missing Values

- Method 1: Listed IF/THEN statements

```
Data NewData1; set SampleData;
        if Q4a=. then Q4a=6;
        if Q4b=. then Q4b=6;
        if Q4c=. then Q4c=6;
        if Q4d=. then Q4d=6;
        if Q4e=. then Q4e=6;
        if Q4f=. then Q4f=6;
        if Q4g=. then Q4g=6;
        if Q4h=. then Q4h=6;
run;
```

Transform Missing Values

- Method 2: Simple Array

```
Data NewData2; set SampleData;
array Q4Process (8) Q4a Q4b Q4c Q4d Q4e Q4f Q4g Q4h;
do X = 1 to 8;
        if Q4Process (X)=. then Q4Process (X)=7;
end;
run;
```

Transform Missing Values

- Method 3: Flexible Array

```
Data NewData3; set SampleData;
array Q4Process (*) Q4:;
do X = 1 to Dim(Q4Process);
        if Q4Process(X)=. then Q4Process(X)=8;
end;
run;
```

Reversing Scales

- Method 1: Listed IF/THEN statements

```
Data NewData4; set SampleData;
array Q4Process (*) Q4:;
do X = 1 to Dim(Q4Process);
            if Q4Process(X)=. then Q4Process(X)=6;
end;
OLD_Q4A=Q4A;
if Q4A=1 then Q4A=5;
if Q4A=2 then Q4A=4;
if Q4A=3 then Q4A=3;
if Q4A=4 then Q4A=2;
if Q4A=5 then Q4A=1;
run;
```

Reversing Scales

- Method 2: SELECT Statement

```
Data NewData5; set SampleData;
array Q4Process (*) Q4:;
do X = 1 to Dim(Q4Process);
        if Q4Process(X)=. then Q4Process(X)=6;
end;
OLD_Q4A=Q4A;
SELECT (Q4A);
        WHEN (1) Q4A=5;
        WHEN (2) Q4A=4;
        WHEN (3) Q4A=3;
        WHEN (4) Q4A=2;
        WHEN (5) Q4A=1;
        otherwise;
END;
run;
```

Reversing Scales

- Method 3: ARRAY+SELECT Statements

```
Data NewData6; set SampleData;
array Q4Process (*) Q4:;
OLD_Q4A=Q4A;
do X = 1 to Dim(Q4Process);
        if Q4Process(X)=. then Q4Process(X)=6;
        IF X=1 THEN DO;
                SELECT (Q4Process(X));
                        WHEN (1) Q4Process(X)=5;
                        WHEN (2) Q4Process(X)=4;
                        WHEN (3) Q4Process(X)=3;
                        WHEN (4) Q4Process(X)=2;
                        WHEN (5) Q4Process(X)=1;
                        otherwise;
                END;
        END;
end;
run;
```

Reversing Scales

- Method 4: Temporary Value Array

```
Data NewData7; set SampleData;
array Q4Process (*) Q4:;
array reverser (6) _temporary_ (5,4,3,2,1,6);
OLD_Q4A=Q4A;
OLD_Q4D=Q4D;
do X = 1 to Dim(Q4Process);
        if Q4Process(X)=. then Q4Process(X)=6;
        IF X in (1,4) THEN Q4Process(X)=Reverser(Q4Process(X));
end;
run;
```

Simple Tabulations

- Method 1: Proc FREQ

```
Title 'Simple Tabulation';
proc freq data=SampleData;
table _ALL_;
run;
```

Simple Tabulations

- Method 2: Proc FREQ with Formats

```
proc format;
value AGE
        1-2="18-29"
        3-4="30-39"
        5-6="40-49"
        7-9="50-64"
        10="65+"
        11="DK/NA";
value SEX
        1='Men'
        2='Women';
value RATE
        1='Very favorable'
        2='Somewhat favorable'
        3='Neutral'
        4='Somewhat unfavorable'
        5='Very unfavorable'
        6='DK';
  picture pctfmt low-high='009%';
run;
```

Simple Tabulations

- Method 2: Proc FREQ with Formats

```
Title 'Simple Tabulation with Formats';
proc freq data=SampleData;
table _ALL_;
        format
                sex sex.
                age age.
                Q4: rate.;
run;
```

Simple Tabulations

- Method 3: 2 Dimensions

```
Title 'Simple Tabulation: 2 Dimension';
proc freq data=SampleData;
        table Q4A*SEX;
        format
                sex sex.
                age age.
                Q4: rate.;
run;
```

Customized Tables

- Method 1: Proc Tabulate-Single Dimension

```
Title 'PROC TABULATE: Single Dimension';
proc tabulate data=SampleData;
        class Q4A;
        table Q4A,
                ALL='Total'*pctn<Q4A>=' '*f=pctfmt9.;
        format Q4A rate.;
run;
```

Customized Tables

- Method 2: Proc Tabulate-Two Dimension

```
Title 'PROC TABULATE: Two Dimension';
proc tabulate data=SampleData;
        class Q4A SEX;
        table Q4A,
                ALL='Total'*pctn<Q4A>=' '*f=pctfmt9.
                SEX='Gender'*pctn<Q4A>=' '*f=pctfmt9.;
        format
                Q4A rate.
                SEX SEX.;
run;
```

Customized Tables

- Method 3: Two Dimension with n-size

```
Title 'PROC TABULATE: Two Dimension With N-size';
proc tabulate data=SampleData;
        class Q4A SEX AGE;
        table Q4A,
                ALL='Total'*
                        (pctn<Q4A>='%'*f=pctfmt6.
                        N='N'*f=4.0)
                SEX='Gender'*pctn<Q4A>=' '*f=pctfmt6.
                AGE='Age'*pctn<Q4A>=' '*f=pctfmt6.;
        format
                Q4A rate.
                SEX SEX.
                AGE AGE.;
run;
```

Advanced Tables

- Method 1: Formats with multiple labels

```
proc format;
value RATEB (Multilabel)
            1='Very favorable'
            2='Somewhat favorable'
            3='Neutral'
            4='Somewhat unfavorable'
            5='Very unfavorable'
            6='DK'
            1-2='Total favorable'
            4-5='Total unfavorable'
            ;
run;
```

Advanced Tables

- Method 1: Formats with multiple labels

```
Title 'PROC TABULATE: Multilabel Method 1';
proc tabulate data=SampleData;
        class Q4A SEX AGE
                / mlf ;
        table Q4A,
                ALL='Total'*pctn<Q4A>='%'*f=pctfmt6.
                SEX='Gender'*pctn<Q4A>=' '*f=pctfmt6.
                AGE='Age'*pctn<Q4A>=' '*f=pctfmt6.;
        format
                Q4A rateb.
                SEX SEX.
                AGE AGE.;
run;
```

Advanced Tables

- Method 2: Ordering multiple labels

```
proc format;
value RATEB (Multilabel)
            1='02-Very favorable'
            2='03-Somewhat favorable'
            3='04-Neutral'
            4='06-Somewhat unfavorable'
            5='07-Very unfavorable'
            6='08-DK'

            1-2='01-Total favorable'
            4-5='05-Total unfavorable'
            ;
run;
```