Applications in Economic Sector Statistics Service (ESSS)

Desiree R. Robles
Senior Statistical Specialist
SSO/ESSS/FSD

Regional Workshop on the Use of R for Statistical Business Processes
17 April 2024
Quezon City
Outline:

1. Brief Background on ESSS

2. Preparatory Activities for Survey Operations:
   a. Computation of Sample Size based on Target CV
   b. Sample Selection

3. Post-survey operations (package: survey)
   a. Generation of Estimate, Standard Error, Coefficient of Variation (%) and Confidence Interval
Outline:

1. Brief Background on ESSS

2. Preparatory Activities for Survey Operations:
   a. Sample Selection

3. Post-survey operations (package: survey)
   a. Generation of Estimate, Standard Error, Coefficient of Variation (%) and Confidence Interval
   b. Computation of Sample Size based on Target CV (for the next survey round)
Sectoral Statistics Office – Economic Sector Statistics Service
Organizational Goal

Deliver relevant and reliable statistical information and services on:
agriculture, price, trade, industry, and services
 Sectoral Statistics Office – Economic Sector Statistics Service

**Functional Objectives**

1. Strengthen operational planning of statistical programs
2. Develop comprehensive and relevant operations manuals and other standard operating procedures
3. Develop clear survey instruments adhering to statistical standards
4. Ensure complete and timely collection of data
5. Ensure sound and efficient processing of data
6. Generate timely and reliable statistical tables, reports, and derived indicators
7. Intensify dissemination of statistical information and products
8. Establish and maintain an updated sector-specific database of statistical information, indicators, and resources
Preparatory Activities for Survey Operations:

a. **Sample Selection**

**General Process:**

1. Prepare the input data (frame)
2. Know the survey design
3. Write the program in R
Preparatory Activities for Survey Operations:

a. **Sample Selection**

Example: establishment-based survey

**Survey:** Annual Survey on Philippine Business and Industry (ASPBI)

1. **Sampling Frame:** List of Establishments

2. **Sampling Design:**
   - **Certainty Stratum** – Take all
   - **Non-Certainty Stratum** – Stratified Systematic Sampling
Preparatory Activities for Survey Operations:

a. **Sample Selection**

Example: establishment-based survey

**Sampling Design:**
Stratified-Systematic Sampling Design

- For each region, industry group and employment stratum, the establishment is sorted by total employment from largest to smallest.
- Samples are then selected through systematic sampling.

Requirements for systematic sampling:
- Random start
- Sampling interval: $k = \frac{N}{n}$
Preparatory Activities for Survey Operations:

a. **Sample Selection**

Example: establishment-based survey

**Stratum Composition:**

```
01    A014    Micro
02    A015    Small
...   ...   ...
19    S963    Large
```

**REGION-INDUSTRY-EMPLOYMENT**

e.g.
- 01-A014-Micro
- 01-A014-Small
- 01-A014-Medium
- 01-A014-Large
Preparatory Activities for Survey Operations:

a. **Sample Selection**

Example: establishment-based survey

For each stratum, samples are selected _systematically_.

**Example**: 01-A032-Small

Total Establishments (N) = 18

Samples (n) = 5

Samples need to be sorted in descending order based on Total Employment.
Preparatory Activities for Survey Operations:

a. **Sample Selection**

**Example: establishment-based survey**

For each stratum, samples are selected systematically

e.g. 01-A032-Small

Total Establishments (N) = 18

Samples (n) = 5

Sampling Interval (k) 
= N/n = 3.6 = 4

Random Start (i) = 2

Samples: 
i, i+k, i+2k, ..., i+(n-1)k

Sorted units based on Total Employment:
Preparatory Activities for Survey Operations:

a. Sample Selection

Example: establishment-based survey

For each stratum, samples are selected systematically

e.g. 01-A032-Small

Total Establishments (N) = 18
Samples (n) = 5
Sampling Interval (k) = N/n = 3.6 = 4
Random Start (i) = 2

Samples:
i, i+k, i+2k, … , i+(n-1)k

Samples for the given stratum code:
Preparatory Activities for Survey Operations:

a. **Sample Selection**

Hands-on example:

Download the files in Folder: “1a_Sample Selection”

- Data
- Output

1a_Sample Selection (One Sector).R
1a_Sample Selection (All Sectors).R

**Try the following:**

1. Select Samples for Sector S using the program: 1a_Sample Selection (One Sector).R
2. Select Samples for Sectors I, J, and K using the program: 1a_Sample Selection (All Sectors).R
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

**General Process:**

1. Prepare the input data (survey returns)
2. Know the survey design (estimation procedure)
3. Write the program in R
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

   **package “survey”** Analysis of Complex Survey Samples

   i. Creating a Survey Design
   ii. Generation of Estimates
      a. svytotal
      b. svymean
      c. svyby
   iii. Options for lonely PSU
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

Hands-on example (for discussion):

Download the files in Folder: "2a_Generation of Estimate, SE, CV, CI"

Data
Output
2a_Generation of Est,SE,CV,CI (One Sector).R
2a_Generation of Est,SE,CV,CI (All Sectors).R
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

```R
package "survey"
```

Analysis of Complex Survey Samples

Creating a Survey Design

Syntax:

```
> svydesign(ids, strata, fpc, weights, data)
```

where:

- `ids` = unique id of each observation
- `strata` = stratum variable (if applicable)
- `weights` = weight of each observation
- `data` = dataset containing the variables
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

```
package "survey"
```

Analysis of Complex Survey Samples

**Creating a Survey Design**

**Syntax:**

```
> svydesign(ids, strata, fpc, weights, data)
```

where:
- `ids` = unique id of each observation
- `strata` = stratum variable (if applicable)
- `weights` = weight of each observation
- `data` = dataset containing the variables

**Hands-on example:**

0. Open `2a_Generation of Est,SE,CV,CI (One Sector).R`
1. Set the working directory, load the needed libraries, and read the datafile "Section R.xlsx" into object `DATA`.
2. Create a survey design with the following parameters:
   - `ids` = `~id`
   - `strata` = `~strcode`
   - `weights` = `~weight`

Save the design on object `DESIGN`. 
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

package “survey” Analysis of Complex Survey Samples

svytotal

**Syntax:**

```
> svytotal(variable, design)
```

**where:**

- `variable` = object of the estimate to be computed
- `design` = survey design defined using svydesign
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

package “survey”

Analysis of Complex Survey Samples

svytot

Syntax:

> svytot(variable, design)

where:

variable = object of the estimate to be computed

design = survey design defined using svydesign

Hands-on example:

3. Compute for the total revenue and total expense for DATA using the following variables:

   total revenue: revenue
   total expense: expense

Save the estimates on the following objects:

totrev
totexp
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

package “survey” Analysis of Complex Survey Samples

svymean

Syntax:

> svymean(variable, design)

where:

variable = object of the estimate to be computed

design = survey design defined using svydesign
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

Syntax:

\[ \text{svymean}(\text{variable}, \text{design}) \]

where:
- **variable** = object of the estimate to be computed
- **design** = survey design defined using svydesign

Hands-on example:

4. Compute for the mean revenue and mean expense for DATA using the following variables:
   - revenue: \texttt{revenue}
   - expense: \texttt{expense}

Save the estimates on the following objects:
- meanrev
- meanexp
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

**package “survey”**

Analysis of Complex Survey Samples

**svyby**

**Syntax:**

```
> svyby(variable, by variable, function, design)
```

where:

- `variable` = object of the estimate to be computed
- `by variable` = level of estimation
- `function` = svymean, svytotal
- `design` = survey design defined using svydesign
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

package “survey”

Analysis of Complex Survey Samples

Syntax:

> svyby(variable, by variable, function, design)

where:
variable = object of the estimate to be computed
by variable = level of estimation
function = svymean, svytotal
design = survey design defined using svydesign

Hands-on example:

5. Compute for the total revenue and total expense for DATA by Industry Group using the following variables:
   revenue: revenue
   expense: expense
   industry group: ind_3d
   Save the estimates on the following objects:
   totrev_ind, totexp_ind

6. Compute for the total revenue and total expense for DATA by Region using the following variables:
   region:
   Save the estimates on the following object:
   totrev_reg, totexp_reg
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

**package “survey”** Analysis of Complex Survey Samples

**Options for Lonely PSUs**

**certainty**

- a single-PSU stratum makes no contribution to the variance (for multistage sampling it makes no contribution at that level of sampling). This is an alternative to specifying fpc and might be useful for compatibility with other software.

**Syntax:**
```
options(survey.lonely.psu="certainty")
```
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

package “survey” Analysis of Complex Survey Samples

Options for Lonely PSUs

adj

• the stratum contribution to the variance is taken to be the average of all the strata with more than one PSU. This might be appropriate if the lonely PSUs were due to data missing at random rather than to design deficiencies.

Syntax:
```r
options(survey.lonely.psu="adjust")
```
Post-Survey Operations

a. Generation of Estimate, SE, CV, and Confidence Interval

Coefficient of Variation (%)

Description

- A measure of relative variability. It is the ratio of the standard deviation to the mean (average).

\[ CV(\%) = \frac{\text{standard error}}{\text{estimate}} \times 100\% \]

- Useful when comparing results from two different surveys or tests that have different measures or values.

- Example: Comparing the results from two tests that have different scoring mechanisms. If sample A has a CV of 15% and sample B has a CV of 25%, you would say that sample B has more variation, relative to its mean.
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

Confidence Interval

**Description**

- The range of values where the true population value falls
- Constructing a Confidence Interval:

  \[ [\text{Lower Limit}, \text{Upper Limit}] \]

  where:

  \[ \text{Lower Limit} = \hat{x} - z \times \text{standard error} \]
  \[ \text{Upper Limit} = \hat{x} + z \times \text{standard error} \]
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

**Hands-on example:**

7-8. Compute for the Coefficient of Variation (%) and 95% Confidence interval for the **total revenue** and **total expense** for DATA. Save the CV (%) and CI as new columns named CV, LCI, UCI on the previously created variables: totrev, totexp

9. Change the working directory to the Output folder.

10. **Export** the objects into **one excel file** with sheets: totrev, totexp, meanrev, meanexp, totrev_ind, totexp_ind, totrev_reg, totexp_reg

Excel file name: “2a_Sector_R_Output.xlsx”
Post-Survey Operations

a. *Generation of Estimate, SE, CV, and Confidence Interval*

**Try the following:**

1-6. Do Items #1 to 6 for Sector I, filename: “Section I.xlsx”

7-8. Compute for the Coefficient of Variation (%) and 95% Confidence interval for the total revenue by industry group and total expense by industry group for I_data.

Save the CV (%) and CI as new columns named CV, LL, UL on the previously created variables: totrev_ind, totpexp_ind

9. Change the working directory to the Output folder.

10. **Export** the objects into one excel file with sheets: totrev, totpexp, meanrev, meanexp, totrev_ind, totpexp_ind, totrev_reg, totpexp_reg

Excel file name: “2a_Sector_I_Output.xlsx”
Post-Survey Operations

a. **Generation of Estimate, SE, CV, and Confidence Interval**

**Hands-on example (using user-defined functions):**

0. Open "2a_Generation of Est,SE,CV,CI (All Sectors).R"
1. Compute for the Estimate, SE, CV, and 95% CI for the **total revenue** by Region and Industry of all sectors. Export results onto file named: “All_Sectors_TotRev_RegInd.xlsx”
2. Compute for the Estimate, SE, CV, and 95% CI for the **mean revenue** by Region and Industry of all sectors. Export results onto file named: “All_Sectors_MeanRev_RegInd.xlsx”

**Try the following:**

1. Compute for the Estimate, SE, CV, and 95% CI for the **total expense** by Region and Industry of all sectors. Export results onto file named: “All_Sectors_TotExp_RegInd.xlsx”
2. Compute for the Estimate, SE, CV, and 95% CI for the **mean expense** by Region and Industry of all sectors. Export results onto file named: “All_Sectors_MeanExp_RegInd.xlsx”
Post-Survey Operations

b. Computation of Sample Size based on Target CV

Initial Sample Size

\[ n_i = \frac{CV'}{CV_{target}} \times n_{old} \]

where \( n_i \) = initial sample size
\( CV' \) = actual CV
\( CV_{target} \) = target CV
\( n_{old} \) = Total no. of samples based on the previous survey round
Post-Survey Operations

b. Computation of Sample Size based on Target CV

\[ n_{adj} = \frac{n_i}{1 + \frac{n_i}{N}} \]

where \( n_{adj} \) = adjusted sample size
\( n_i \) = initial sample size
\( N \) = total number of population units
Post-Survey Operations

b. Computation of Sample Size based on Target CV

In cases where the computed sample size is greater than the population size, $n$ would be adjusted accordingly.

If $n > N$, $n = N$
else $n = n$
b. Computation of Sample Size based on Target CV

\[ n_{nr} = n_{adj} \times (1 + x) \]

where

- \( n_{nr} \) = adjusted sample size based on non-response
- \( n_{adj} \) = adjusted sample size
- \( x \) = adjustment factor for non-response

Adjusted by non-response
b. Computation of Sample Size based on Target CV

Adjusted by minimum

Adjustment of sample size based on a minimum value. (e.g. 5).

Let the final sample size be equal to $n_f$.

**Conditions:**
1) If the $n_{nr}$ is less than the minimum sample size and its $N > 5$ then $n_f$ would be equal to 5
2) If the $n_{nr}$ is less than the minimum sample size and its $N <= 5$, then $n_f$ would be equal to $N$

Otherwise retain $n_f = n_{nr}$
Post-Survey Operations

b. Computation of Sample Size based on Target CV

Hands-on example:

Using the output for the generated output for total revenue with SE, CV, CI by region and industry group in 2a. Compute for the sample size with a target CV of 5% and estimated non-response rate of 5%.
Thank you!

http://www.psa.gov.ph

http://openstat.psa.gov.ph

https://twitter.com/PSAgovph

https://www.facebook.com/PSAgovph