**Project. Designing a stratified sampling plan**

**MSDS 6370**

**Objective:**

* For the student to learn how to design a stratified sampling plan
* For the student to continue learning about SAS procedures for selecting stratified samples and forming estimates using the selected samples.

**Introduction**

The course materials have covered different aspects of implementing stratified sampling using SAS. In this project, students have the opportunity to put all the aspects of implementing stratified sample together from beginning to end in the design of a sampling plan.

**Population data**

In this project, you will use a simulated population of one month’s data for an industry we will call Industry 3. You have Sales and Inventory for the complete population for one month. Your task is to design a sampling plan for a sample of size 500 to be selected for data collection each month for the next 2 years. The purpose of the data collection is to estimate the total Sales and total Inventory for the entire population each month. The quality of the estimate of total Sales has priority over the estimate of total Inventory although both are important.

The data is in the file MSDS\_6370\_Industry3.csv. The variables shown in the table below.

|  |  |
| --- | --- |
| UnitID | Unique identifier |
| Sales | Current month sales for unit |
| Inventory | Current month inventory for unit |

Correlation between Sales and Inventory = 0.82.

In designing the sampling plan, the students have to consider that estimates will be formed for two variables, Sales and Inventory. The methods for determining strata and sample allocation to strata are designed to use one variable. Although the correlation coefficient for sales and inventory is high at 0.82, a design based on one probably will not be perfect for the other. The student will create a sampling design using Sales and form estimates for both Sales and Inventory. Next the student will create sampling design using Inventory and form estimates for both variables. Using the results from both designs, the student will decide which one sampling design is better under the requirement of estimates of total for both.

Part of creating a sampling design is deciding on the number of strata and how to divide the population into the strata. One method is the . The paper at this link describes other methods and compares them with . <http://www.amstat.org/sections/srms/proceedings/y2003/Files/JSM2003-000891.pdf>

**Project Steps:**

1. Provide descriptive statistics for your dataset.

2. Design sampling plan using the Sales variable.

1. Form sampling strata for the population using the Sales variable.
   * A certainty stratum is required. However, the assignment includes deciding on the number of units in the certainty stratum.
   * The total number of strata must be at least 6 but no larger than 11. Since 1 stratum has to be a certainty stratum, the number of non-certainty strata must range between 5 and 10.
   * Choose an approach for dividing the population units into strata and implement it.
2. Allocate the sample of size 500 to the strata using Neyman allocation.
3. Use Proc Surveyselect to select 5 samples using the sample sizes allocated to the strata by the Neyman allocation. Submit your code and seed(s). (Note: A handy option for Proc Surveyselect tht you might want to use is REPS=.)
4. Use SAS Procedure Surveymeans and the samples selected using Neyman allocation to form the following estimates for Sales for each sample: the population mean, the standard error of the mean, the 95% confidence interval for the mean, the population total, the standard deviation, and the 95% confidence interval for the total.
5. Determine how many of the 5 samples produce a 95% confidence interval that includes the population total. Examine the mean of the estimates of total and mean of the estimated variance of the total.
6. Repeat steps 1d - 1e) for Inventory.

3. Design a sampling plan using the Inventory variable by repeating steps 1a) – 1f) with Inventory replacing Sales.

4. Make a recommendation about whether to use your sampling plan based on the variable Sales or the one based on the variable Inventory for monthly collection of both variables for the next 2 years.

5. What have you learned about how to stratify a population?

# Submitting Results

## **Important Dates**

**Presentation during Live Session 13**

**Paper Due Date: Live Session 13**

## **Basic Requirements**

Each 2-person project team is expected to give a 10-minute presentation concerning your results and conclusions in Live Session 13. Each team member will submit written results and conclusions. This is joint work but the submissions do not necessarily have to be identical. Each graphic or table should be clearly labeled and discussed in the text. Your SAS code and log should also be included in the appendix. Please include comments in your SAS code indicating the purpose of each procedure.

## **Evaluation**

In general, expect the written results and conclusion to be weighed about 80% on content and 20% on organization and presentation. Keep in mind that projects are generally much more sensitive to sincere and obvious effort, over and above actual results.

Your Lab 13 grade will be based on your presentation in Live Session 13.

## **Presentation Structure**

Each 2-person team will discuss their results in a 10-minute presentation in Live Session 13. Each team member will give part of the presentation. Prepare at most 10 Powerpoint slides and submit as Lab 13.

Your team presentation will include the following:

* Description of your stratification based on the Sales variable, how you arrived at it, and the performance of estimates of Sales and Inventory.
* Description of your stratification based on the Inventory variable, how you arrived at it, and the performance of estimates of Sales and Inventory.
* The stratification you chose and why.

It will not be possible to show all the details in your presentation. Therefore, emphasize your more interesting results and choices.

Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 4/10/16

**Project Results and Conclusions**

### 1.Descriptive Statistics

Provide a summary of your data set that includes:

* The number of units in the population
* For each variable, the population total, standard deviation, the population mean, and a histogram.

Include your SAS code and log in the appendix.

### 2. Stratification Using Sales

In this section, report the results for using Sales as the stratification variable.

2.1 Describe the process you followed in forming sampling strata for the population using the Sales variable:

* + Describe the number and range of values you chose for the certainty stratum and your criteria for the units in the certainty stratum.
  + Describe your process for deciding on the number of non-certainty strata. Include your calculations such as a table. If you used SAS, include your code and your log put it in an Appendix.
  + Provide you strata and support your choices. Include a table that has a row for each stratum with the stratum number, number of units, and range of values.

2.2 Allocate the sample of size 500 to the strata using Neyman allocation. Show a table with appropriate headings shows your calculations and the allocation of the sample to strata. If you use SAS to calculate table entries, include that SAS code that provides input your SAS code and your log in the appendix.

2.3 Use Proc Surveyselect to select 5 samples using the sample sizes allocated to the strata by the Neyman allocation. Place your SAS code, seeds, and log in an appendix.

2.4 Use SAS Procedure Surveymeans and the 5 samples selected using Neyman allocation to form the following estimates for Sales for each sample: the population mean, the standard error of the mean, the 95% confidence interval for the mean, the population total, the standard deviation, and the 95% confidence interval for the total. Show the estimates for the 5 samples in a table as well as the averages and ranges of 5 population totals and 5 population means and the percentage of the confidence intervals for the estimated totals that includes the true total and the percentage of confidence intervals for the estimated means that includes the true mean. Place your SAS code, seeds, and log in an appendix.

2.5 Repeat Section 3.4 for Inventory.

2.6 Discuss how well the stratification performed for estimates of Sales and Inventory.

Compare the estimated population total and mean with the means the 5 estimates of the total and mean. Discuss how many of the confidence intervals from the 5 samples included the population total and the average of the estimated variance of the population total.

### 3. Stratification Using Inventory

#### Repeat Sections 2.1 to 2.6 for the stratification using Inventory.

#### 4. Conclusion

Decide which stratification to use and support your decision. Discuss any trade-offs and compromises that you had to make in forming your decision. Remember that the quality of the estimate of total Sales has priority over the estimate of total Inventory although both are important

5. Summary

Summarize what you have learned about how to stratify a population and design a sampling plan

Note: references are not required, but if you use material from a book, an article, or a website, you must cite the source, and use quotation marks for any paragraph (or part thereof) that is quoted word-for-word from the source(s).