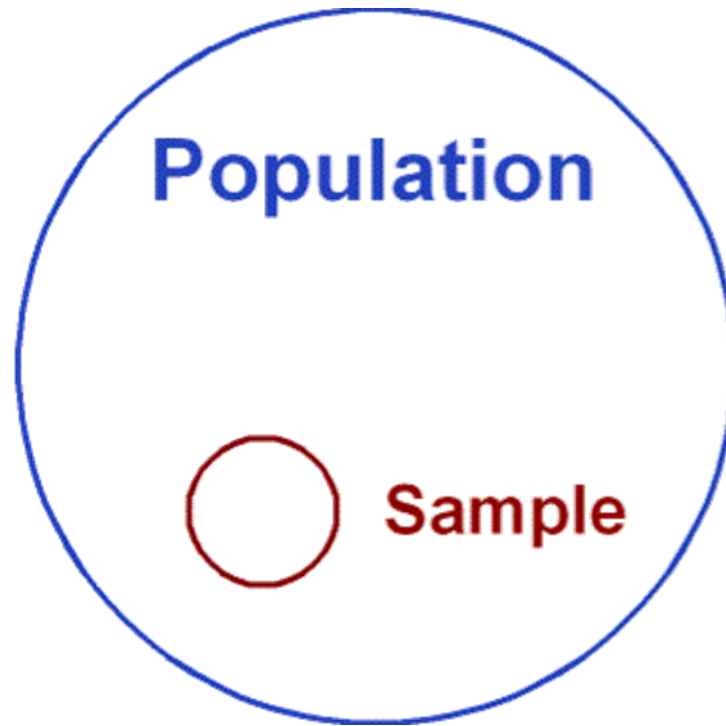


Stratified Sampling



Adnan Alkhouri

Bank of Montreal – TASS Sept 2016

Objectives:

- What and Why
- Advantage
- SAS Example (with code and results)

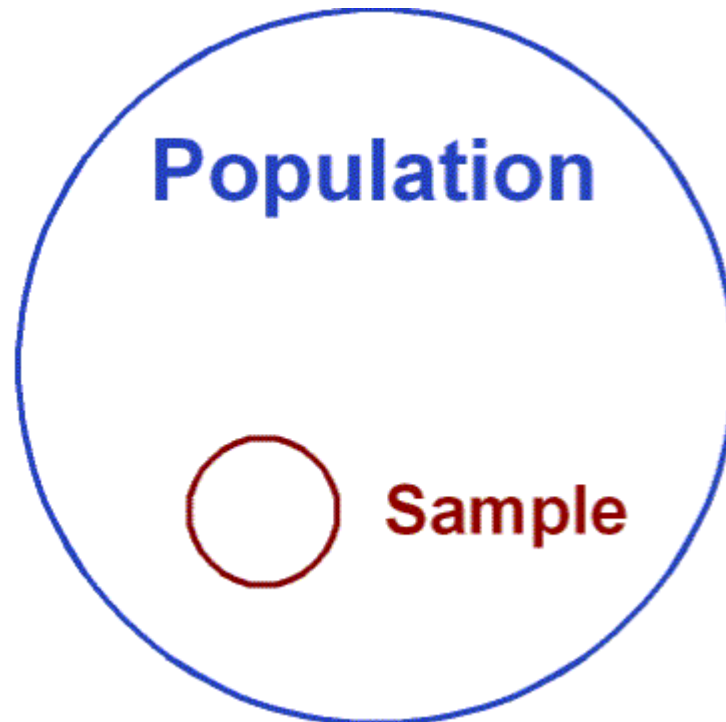
Quotes

“Next time someone tells you they don’t believe a small sample poll can possibly tell you anything, just say to them ‘OK, then. Next time you have to have a blood test, why don’t you ask them to take the whole lot?’”

– British opinion pollster Nick Moon

CONCEPTS

- **Census:** an accounting of the complete population
- **Sample:** a subset of the population that should represent the entire group



Stratified Sampling

- **What is Stratified Sampling**
- **How to Sample**
- **Advantage**

NEIGHBOURHOOD INCOME AND DEMOGRAPHICS STATISTICS, 2007

| Forward Sortation Area, BC | Married Taxfilers (#) | | | Males | Males | Females | Females | Total | Total |
|----------------------------------|-----------------------|---------|--------|--------|--------------------------------|---------|--------------------------------|--------|--------------------------------|
| | Males | Females | Total | # With | Total | # With | Total | # With | Total |
| | | | | Income | Amt of Total Inc. (\$000's) | Income | Amt of Total Inc. (\$000's) | Income | Amt of Total Inc. (\$000's) |
| V0A | 2,980 | 2,900 | 5,890 | 5,330 | 252,932 | 5,020 | 150,159 | 10,360 | 403,267 |
| V0B | 7,600 | 7,390 | 15,040 | 12,100 | 580,287 | 11,760 | 311,144 | 23,910 | 892,802 |
| V0C | 5,840 | 5,630 | 11,470 | 9,990 | 623,977 | 8,650 | 245,343 | 18,640 | 869,338 |
| V0E | 15,710 | 15,260 | 30,990 | 25,220 | 1,063,589 | 24,750 | 623,622 | 49,980 | 1,687,740 |
| V0G | 6,030 | 5,810 | 11,850 | 9,880 | 427,871 | 9,630 | 257,455 | 19,520 | 685,408 |
| V0H | 13,410 | 13,170 | 26,590 | 20,470 | 856,850 | 21,400 | 626,476 | 41,880 | 1,483,497 |
| V0J | 11,370 | 11,010 | 22,400 | 19,260 | 887,503 | 17,800 | 437,200 | 37,070 | 1,324,944 |
| V0K | 7,930 | 7,620 | 15,590 | 12,860 | 503,739 | 12,220 | 279,415 | 25,110 | 784,286 |

Source: Canada Revenue Agency

Stratified Sampling **SAS Code 1**

```
1. PROC SORT  
2.     DATA=SASHELP.PRDSAL2  
3.     OUT=WORK.STATE;  
4. BY  
5. STATE;  
6. RUN;
```

Stratified Sampling **SAS Code 2**

```
1. PROC SURVEYSELECT  
   DATA=WORK.STATE  
2.   N=2304  
3.   OUT=WORK.STR_SAMPLE ;  
4.   STRATA STATE / ALLOC=PROP;  
5. RUN;
```


Stratified Sampling **SAS Results 1**

The SURVEYSELECT Procedure

| | |
|---------------------------|-------------------------------|
| Selection Method | Simple Random Sampling |
| Strata Variable | STATE |
| Allocation | Proportional |
| Input Data Set | STATE |
| Random Number Seed | 323918840 |
| Number of Strata | 16 |
| Total Sample Size | 2304 |
| Output Data Set | STR_SAMPLE |

State/Province (Entire Population) SAS Results 2

| STATE | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|----------------------------------|------------------|----------------|---------------------------------|-------------------------------|
| Baja California Norte | 1152 | 5.00 | 1152 | 5.00 |
| British Columbia | 1152 | 5.00 | 2304 | 10.00 |
| California | 1152 | 5.00 | 3456 | 15.00 |
| Campeche | 1152 | 5.00 | 4608 | 20.00 |
| Colorado | 1152 | 5.00 | 5760 | 25.00 |
| Florida | 1152 | 5.00 | 6912 | 30.00 |
| Illinois | 5760 | 25.00 | 12672 | 55.00 |
| Michoacan | 1152 | 5.00 | 13824 | 60.00 |
| New York | 1152 | 5.00 | 14976 | 65.00 |
| North Carolina | 1152 | 5.00 | 16128 | 70.00 |
| Nuevo Leon | 1152 | 5.00 | 17280 | 75.00 |
| Ontario | 1152 | 5.00 | 18432 | 80.00 |
| Quebec | 1152 | 5.00 | 19584 | 85.00 |
| Saskatchewan | 1152 | 5.00 | 20736 | 90.00 |
| Texas | 1152 | 5.00 | 21888 | 95.00 |
| Washington | 1152 | 5.00 | 23040 | 100.00 |

State/Province (SAMPLE) SAS Results 3

| STATE | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|----------------------------------|------------------|----------------|---------------------------------|-------------------------------|
| Baja California Norte | 115 | 4.99 | 115 | 4.99 |
| British Columbia | 115 | 4.99 | 230 | 9.98 |
| California | 115 | 4.99 | 345 | 14.97 |
| Campeche | 115 | 4.99 | 460 | 19.97 |
| Colorado | 115 | 4.99 | 575 | 24.96 |
| Florida | 115 | 4.99 | 690 | 29.95 |
| Illinois | 576 | 25.00 | 1266 | 54.95 |
| Michoacan | 116 | 5.03 | 1382 | 59.98 |
| New York | 116 | 5.03 | 1498 | 65.02 |
| North Carolina | 116 | 5.03 | 1614 | 70.05 |
| Nuevo Leon | 115 | 4.99 | 1729 | 75.04 |
| Ontario | 115 | 4.99 | 1844 | 80.03 |
| Quebec | 115 | 4.99 | 1959 | 85.03 |
| Saskatchewan | 115 | 4.99 | 2074 | 90.02 |
| Texas | 115 | 4.99 | 2189 | 95.01 |

Questions

