

Module 6: Sampling



Objectives

This session will address

- why we use sampling
- how sampling can create efficiencies for data collection
- sampling techniques, processes and choices



Sampling: Links to Data Collection

Can we collect data from the entire population?

- *All files, all streets, all students, all people?*

- When we can, we can talk about what is true for the entire population.
- But often we cannot collect data from the entire population due to time, cost and other constraints.

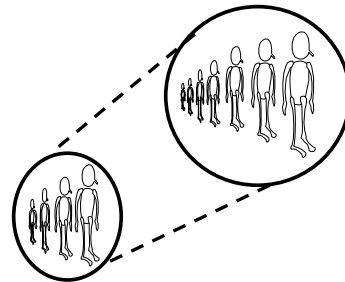
3

Europe and Central Asia Region and World Bank Institute Evaluation Group



Sampling

- It is a data collection strategy
- It is widely used in all sectors
- Offers distinctive efficiencies for gathering data about populations



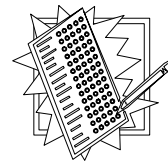
4

Europe and Central Asia Region and World Bank Institute Evaluation Group



Why Sample?

1. A **sample** allows to draw conclusions about the larger population based on what we learn from a subset
2. There are two general types of sampling:
 - Random sampling
 - Non-random sampling



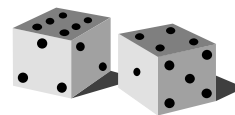
5

Europe and Central Asia Region and World Bank Institute Evaluation Group



Random Sampling

- In statistical terms, a random sample is a set of items that are drawn from a population in such a way that each item of the population has equal (or positive) opportunity to appear in the sample.
 - **Advantages**
 - Eliminates selection bias
 - Able to generalize to the population
 - Cost-effective
 - **Challenge**
 - To locate a complete listing of the entire population from which to select a sample.



6

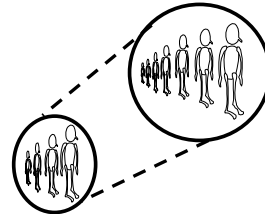
Europe and Central Asia Region and World Bank Institute Evaluation Group



Random Sampling

Sampling Concepts

- **Population** (also referred to as target population or universe)
 - the total set of units
 - e.g., first-year teachers across Freelandia's 13 provinces, in all types of schools.
- **Sampling Frame** (or survey population)
 - list from which to select your sample
 - e.g., a list of all first-year teachers
- **Sample**
 - a subset of the population
 - e.g., a selection from the list of first-year teachers



7

Europe and Central Asia Region and World Bank Institute Evaluation Group



Random Sampling

Sampling Concepts (continued)

- **Sample Design**
 - methods of sampling
- **Parameter**
 - characteristic of the population
- **Statistic**
 - characteristic of a sample



8

Europe and Central Asia Region and World Bank Institute Evaluation Group



Types of Random Samples

Simple Random Sample

- Simplest
- Subset of the entire population

Stratified Random Sample

- Population is separated into strata (or groups).
- Each stratum is randomly sampled

Cluster Sample

- Useful when you don't have a complete listing of the entire population.
- If you want to survey all pregnant women, you probably don't have a list.

9

Europe and Central Asia Region and World Bank Institute Evaluation Group

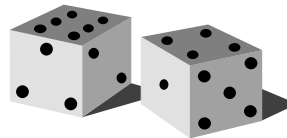


Types of Random Samples

Simple Random Sample

- simplest
- subset of the entire population

Example: A sample drawn from a list of all graduates of the teachers college



10

Europe and Central Asia Region and World Bank Institute Evaluation Group



Types of Random Samples

Stratified Random Sample

- population is separated into strata (or groups)
- each strata is randomly sampled
- ensures that we have enough in each group for statistical analysis
- May need a larger sample than for simple random sample

**Example: Population of graduates stratified by sex.
A random sample of men and a random sample of women are selected**

11

Europe and Central Asia Region and World Bank Institute Evaluation Group

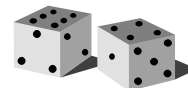


Types of Random Samples

Cluster Samples

Useful when you don't have a complete listing of the entire population.

- If you want to survey parents of primary school children in your country, you probably don't have a list.
 - Randomly select schools
 - Obtain list of parents by schools
 - Randomly select parents



12

Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

1. Samples have a probability of error
2. Statisticians have figured out how to estimate that probability
3. **Statistics:** estimates for the probability that the sample results are representative of the population parameter as a whole.

13

Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

Need to decide:

1. How confident do you want to be that your results are accurate? **The confidence interval**
2. How precise do you want to be in your estimates? **The margin of error**

Answers to these questions influence the sample size you need and, ultimately, the costs of the survey!

14

Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

Confidence Interval

- ✓ How **confident** do you want to be that your sample is reasonably accurate?
- ✓ Standard is a 95% confidence level:
 - ✓ means that 19 out of 20 samples would have found similar results
 - ✓ means that we are 95% certain that the sample results are an accurate estimate of the population

15

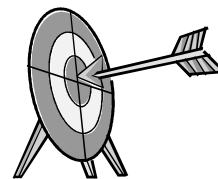
Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

Margin of Error

- How **Precise** do you want to be in your estimates?
- Survey results:**
 - 45% oppose building a dam and 55% favor building a dam.
 - The margin of error is $\pm 3\%$.



16

Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

± 3% points

No

42% - 45% - 48%

Yes

52% - 55% - 58%

17

Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

Margin of error

- This means that if we surveyed everyone, between 42-48% oppose building a dam and between 52-58% of the population favor building a dam.
- We are 95% certain that the majority of the citizens favor building the dam.



18

Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

Margin of error

- The social science standard for margin of error is $\pm 5\%$.
- Survey results:**
 - 45% oppose building a dam and 55% favor building a dam.
 - The margin of error is $\pm 5\%$.

19

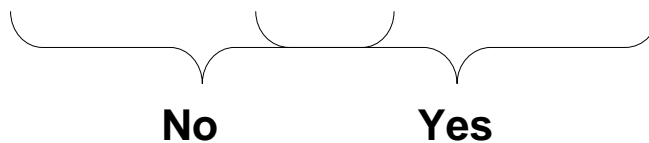
Europe and Central Asia Region and World Bank Institute Evaluation Group



Samples are Imperfect

$\pm 5\%$ points

40% - 45% - 50% - 55% - 60%



20

Europe and Central Asia Region and World Bank Institute Evaluation Group



Sample Sizes for Large (Infinite) Populations

Discussion

| Margin of Error | Confidence Level | | |
|-----------------|------------------|-----|-----|
| | 90% | 95% | 99% |
| $\pm 5\%$ | | | |
| $\pm 3\%$ | | | |
| $\pm 2\%$ | | | |
| $\pm 1\%$ | | | |

21

Europe and Central Asia Region and World Bank Institute Evaluation Group



Sample Sizes for Large (Infinite) Populations

Discussion

| Margin of Error | Confidence Level | | |
|-----------------|------------------|-------|--------|
| | 90% | 95% | 99% |
| $\pm 5\%$ | 271 | 384 | 883 |
| $\pm 3\%$ | 752 | 1,067 | 1,848 |
| $\pm 2\%$ | 1,691 | 2,301 | 4,144 |
| $\pm 1\%$ | 6,765 | 9,604 | 16,576 |

22

Europe and Central Asia Region and World Bank Institute Evaluation Group



Guide to Sample Size

| Population Size | Sample Size | Population Size | Sample Size |
|-----------------|-------------|-----------------|-------------|
| 10 | 10 | 550 | 226 |
| 20 | 19 | 600 | 234 |
| 40 | 36 | 700 | 248 |
| 50 | 44 | 800 | 260 |
| 75 | 63 | 900 | 269 |
| 100 | 80 | 1,000 | 278 |
| 150 | 108 | 1,200 | 291 |
| 200 | 132 | 1,300 | 297 |
| 250 | 152 | 1,500 | 306 |
| 300 | 169 | 3,000 | 341 |
| 350 | 184 | 6,000 | 361 |
| 400 | 196 | 9,000 | 368 |
| 450 | 207 | 50,000 | 381 |
| 500 | 217 | 100,000+ | 385 |

*95% Confidence Level and +/- 5% sample error

23

Europe and Central Asia Region and World Bank Institute Evaluation Group



Sample Size

In general, accuracy and precision is improved by increasing the sample size.

Using the previous example, the sample size would be:

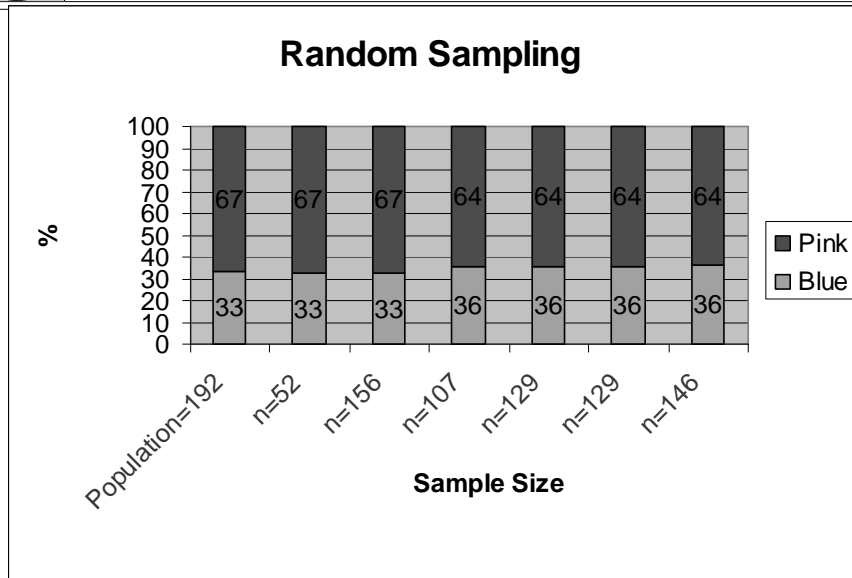
- 384, if we are to be 95% certain, $\pm 5\%$.
- 1,067, if we are to be 95% certain, $\pm 3\%$.

24

Europe and Central Asia Region and World Bank Institute Evaluation Group



Random Sampling



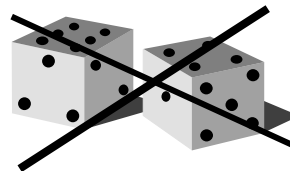
25

Europe and Central Asia Region and World Bank Institute Evaluation Group



Non-Random Sampling

- Quota
- Accidental
- Snowball
- Judgmental
- Convenience



26

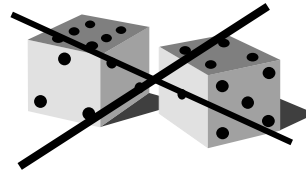
Europe and Central Asia Region and World Bank Institute Evaluation Group



Non-Random Sampling

Potential Bias

- Were these people selected in a biased way?
- Are they substantially different from the rest of the population?
- It helps to collect some data to show that the people selected are fairly similar to the larger population (e.g. demographics)



27

Europe and Central Asia Region and World Bank Institute Evaluation Group



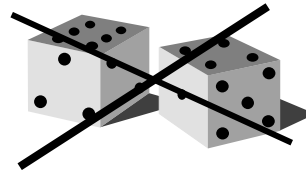
Non-Random Sampling

The results of non-probability samples cannot be generalized

- Data are reported in terms “Of the respondents....”

Sample size not that important

- Enough so it seems reasonable
- Enough to ensure variation



28

Europe and Central Asia Region and World Bank Institute Evaluation Group



Discussion: Sampling

What sampling strategy would you use to determine the quality of the roads after implementing a road improvement project?

Measure: number of holes per street per mile.

29

Europe and Central Asia Region and World Bank Institute Evaluation Group

WORLD BANK INSTITUTE

Promoting knowledge and learning for a better world

The Power of Random Sampling

Practical Example 25.6.07



© 2007. The World Bank Group. All rights reserved.



The Power of Random Sampling

The French Presidential Election 2007



31

Europe and Central Asia Region and World Bank Institute Evaluation Group



The Power of Random Sampling

French Presidential Election 2007

- First Round of Vote - 22 April 2007 (12 candidates)
- Second Round of Vote - 06 May 2007 (2 candidates)
- The Random Sampling Technique was employed by several Polling firms to predict the winners of each Round.
- This example will focus on some recent predictions of opinion polls conducted by the firm IPSOS prior to the 1st and 2nd Rounds and the proximity to the actual results
- General Methodology used by IPSOS: Varying samples of persons were selected at random from among the French Registered Voting population, over different periods of time, and telephone interviews were conducted during two day intervals.

32

Europe and Central Asia Region and World Bank Institute Evaluation Group



The Power of Random Sampling

Prior to the 1st Round of Voting, 4 of the 12 Candidates consistently registered over 10% in the Opinion Polls and were regarded as having a reasonable chance of reaching the 2nd Round :



Nicolas Sarkozy



Segolene Royal



Francois Bayrou



Jean-Marie Le Pen

33

Europe and Central Asia Region and World Bank Institute Evaluation Group



The Power of Random Sampling

IPSOS Opinion Poll- French Presidential Election 2007 – (1st Round)
IPSOS: "Who would you vote for in the Presidential Election?"

| Candidate | Polling Date April 13 2007 | Polling Date April 15 2007 | Polling Date April 19 2007 | Polling Date April 20 2007 | Actual Result |
|-------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|------------------|
| Nicolas Sarkozy | 30 % | 29.5 % | 28 % | 28 % | 31.18 % |
| Segolene Royal | 24 % | 25 % | 22.5 % | 22.5 % | 25.87 % |
| Francois Bayrou | 18.5 % | 17.5 % | 20 % | 20 % | 18.57 % |
| Jean-Marie Le Pen | 13.5 % | 13.5 % | 13 % | 13 % | 10.44 % |
| 8 Others | 14 % | 14.5 % | 16 % | 16.5 % | 13.94 % |

•**Source:** IPSOS/SFR/Dell/Le Point

•**Margin of Error :** + / - 3% for the leading candidates and + / - 2% for the others

•**Methodology :** Telephone interviews conducted with 1598 persons between 19 and 20 April 2007 selected from among 44M Registered Voters

34

Europe and Central Asia Region and World Bank Institute Evaluation Group

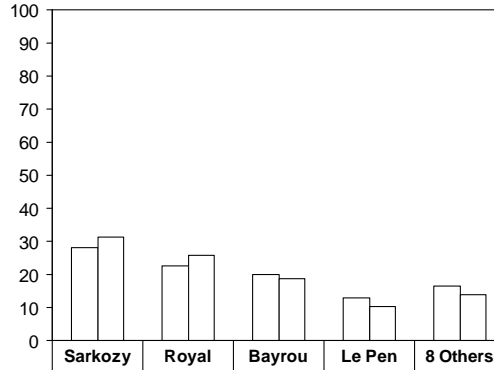


The Power of Random Sampling

French Presidential Election 2007 – (1st Round)

- Date of 1st Round of Voting: 22 April 2007
- Date of Poll: 20 April 2007
- Sampling Technique: Random Sampling
- Sampling Frame: List of 44 M registered voters
- Sample Size: 1598 Registered Voters
- Methodology: Telephone interviews conducted between April 19 and 20, 2007
- Margin of Error: +/- 3% for the leading candidates and +/- 2% for the Others

IPSOS OPINION POLL vs ACTUAL RESULT



| | Sarkozy | Royal | Bayrou | Le Pen | 8 Others |
|--------|---------|-------|--------|--------|----------|
| IPSOS | 28 | 22.5 | 20 | 13 | 16.5 |
| ACTUAL | 31.18 | 25.87 | 18.57 | 10.44 | 13.94 |

35

Europe and Central Asia Region and World Bank Institute Evaluation Group



The Power of Random Sampling

French Presidential Election 2007 – (2nd Round)

Nicolas Sarkozy

Segolene Royal



36

Europe and Central Asia Region and World Bank Institute Evaluation Group



The Power of Random Sampling

IPSOS Opinion Poll – French Presidential Election 2007 - (2ndRound) IPSOS: “Who would you vote for in the Presidential Election?”

| Candidate | Polling Date May 1 2007 | Polling Date May 2 2007 | Polling Date May 3 2007 | Actual Result May 6 2007 |
|-----------------|-------------------------------|-------------------------------|-------------------------------|--------------------------------|
| Nicolas Sarkozy | 53.5 % | 53.5 % | 54 % | 53.04 % |
| Segolene Royal | 46.5 % | 46.5 % | 46 % | 46.94 % |

- **Source:** Ipsos/SFR/Le Point Newspaper
- **Methodology:** Telephone Interviews conducted with 1,414 French Registered Voters on May 2 and May 3, 2007
- **Total Registered Voters:** 44 million
- **Margin of Error:** +/-3%

37

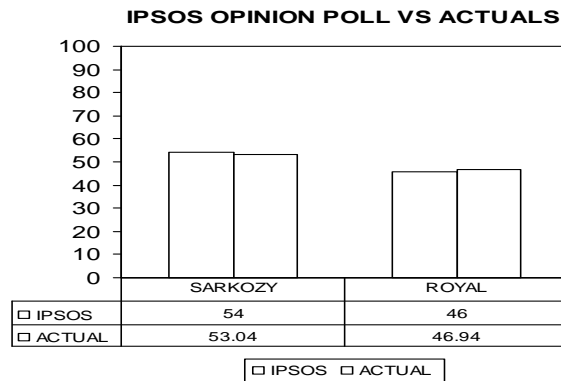
Europe and Central Asia Region and World Bank Institute Evaluation Group



The Power of Random Sampling

French Presidential Election 2007 – 2nd Round IPSOS: “Who would you vote for in the Presidential Election?”

- Date of 2nd Round of Voting: 6 May 2007
- Dates of Poll: 3rd May 2007
- Sampling Technique: Random Sampling
- Sampling Frame: List of 44 million Registered Voters
- Sample Size: 1,414 Registered Voters
- Methodology: Telephone Interviews conducted on May 2nd - 3rd 2007
- Margin of Error: +/-3%
- Source: IPSOS/SFR/Le Point Newspaper



38

Europe and Central Asia Region and World Bank Institute Evaluation Group



The Power of Random Sampling

Summary:

- The polls were conducted over a period of time
- The sample of persons representing the registered voting population varied over the different periods
- The estimates of the polls were relatively close to the actual results in both instances despite the varying numbers of persons chosen and the different times that the polls were conducted
- It was not necessary to interview the entire Registered Voting population in order to arrive at a reasonable estimate of the Election results
- Significance of the Margin of Error

39

Europe and Central Asia Region and World Bank Institute Evaluation Group



Group Project Exercise

Your project – for evaluation

- For each question or sub-question, decide if sampling is needed
- Determine the best sampling method for collecting the data
- Insert sampling information into the sampling column

Your project – for monitoring

- For each objective, decide if sampling is needed
- Determine the best sampling method for collecting the data

40

Europe and Central Asia Region and World Bank Institute Evaluation Group