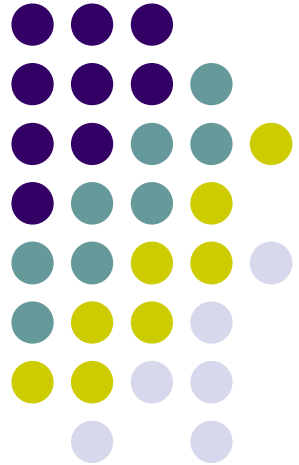
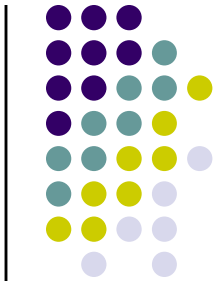
# Sampling and Investigating Hard Data

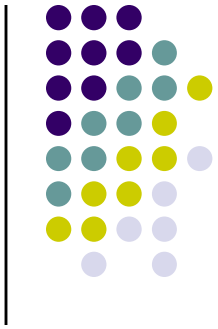


# Major Topics



- Sampling
- Hard data
- Qualitative document analysis
- Workflow analysis
- Business process reengineering
- Archival documents

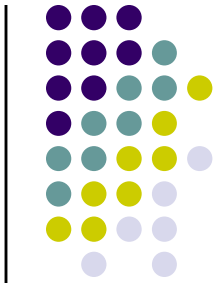
# What is Sampling?



- Sampling is a process of systematically selecting representative elements of a population

- Key issues?
  - Which of the key documents and Web sites should be sampled?
  - Which people should be interviewed or sent questionnaires?

# Why do we need Sampling?



- The reasons systems analysts do sampling are
  - Reduction of costs

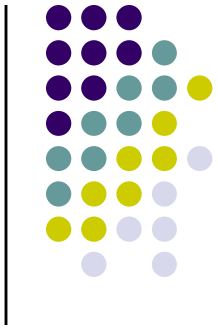
    (copying documents and talking to everyone)
  - Speeding up the data-gathering process

    (Due to selective data processing)
  - Improving effectiveness

    (Talking to few people but asking detailed questions)
  - Reduction of data-gathering bias

    (views of the executives, who have already handled the existing IS, may provide biased evaluation)

# Sampling Design Steps



- To design a good sample, a systems analyst needs to follow four steps:
  - Determining the data to be collected or described (only relevant data and the methods of analysis))
  - Determining the population to be sampled ( Who to interview? Not only the employee but alos the customer sometimes)
  - Choosing the type of sample (1&2 Unrestricted; 3&4 Selective)

    1-**Cnvenience** or unrestricted, e.g. System Analysts may call a meeting on any issue (**Reliability ?**);
    2-**Simple Random** - equal chance to the each people to be questioned (**non practical**)
    3- **Purposive**- based on judgment after meeting with few knowledgeable people (Reliability?)
    4-**Complex random**- (Systematic, Stratified (making subgroups and taking samples) and clustered (group of documents or people to study))

  - Deciding on the sample size ( by the way of knowing the population thinking in a single way and having certain characteristics)

# Sample Size

- Sometimes by knowing the population thinking in a single way and having certain characteristics and sometimes by the mistakes in the forms- referred as attribute data

- The sample size decision should be made according to the specific conditions under which a systems analysts works with such as

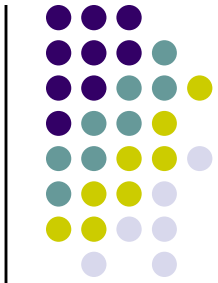
  - Sampling data on attributes

    (Subjective decisions e.g. Confidence level)
  - Sampling data on variables

    (on actual numbers e.g. gross sales, items returned, number of mistakes)
  - Sampling qualitative data

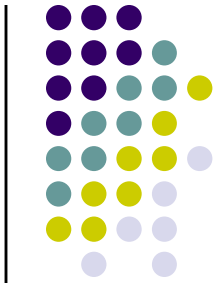
    ( by interviewing and not through searching in files, reports or documents)

# Types of Sampling



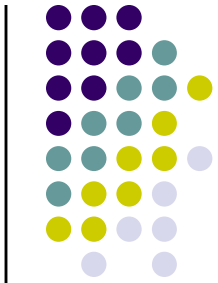
- There are four types of sampling
  - Convenience
  - Purposive
  - Simple random
  - Complex random

# Convenience Sampling



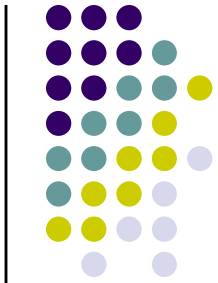
- Convenience samples are unrestricted, nonprobability samples
- Easy to arrange
- Most unreliable

# Purposive Sampling



- Based on judgment
- Analyst chooses group of individuals to sample
- Based on criteria
- Nonprobability sample
- Moderately reliable

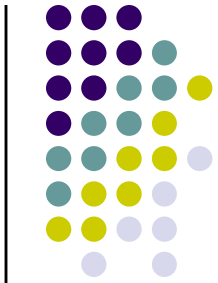
# Simple Random Sampling



- Based on a numbered list of the population
- Each person or document has an equal chance of being selected

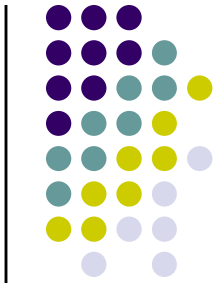
Hence Non Practical

# Complex Random Sampling



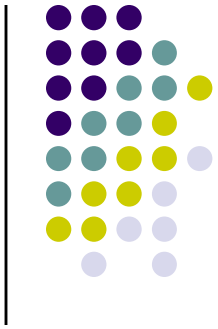
- Has three forms
  - Systematic sampling
  - Stratified sampling
  - Cluster sampling

# Systematic Sampling



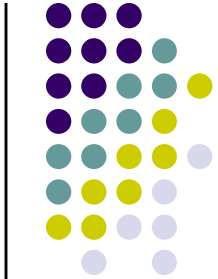
- Simplest method of probability sampling
- Choose every *k*th person on a list
- Not good if the list is ordered

# Stratified Sampling



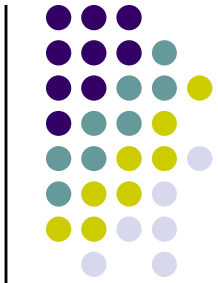
- Identifying subpopulations or strata
- Selecting objects or people for sampling from the subpopulation
- Compensates for a disproportionate number of employees from a certain group
- Most important to the systems analyst

# Cluster Sampling



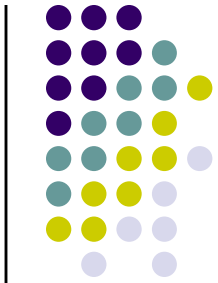
- Select group of documents or people to study
- Select typical groups that represent the remaining ones

# Deciding Sample Size for Attribute Data



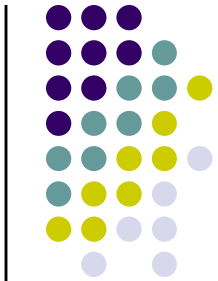
- Seven steps to determine sample size
  - Determine the attribute to sample
  - Locate the database or reports where the attribute is found
  - Examine the attribute and estimate *p,* the proportion of the population having the attribute

# Deciding Sample Size for Attribute Data



- Steps to determine sample size (continued)
  - Make the subjective decision regarding the acceptable interval estimate, *i*
  - Choose the confidence level and look up the confidence coefficient (*z* value) in a table

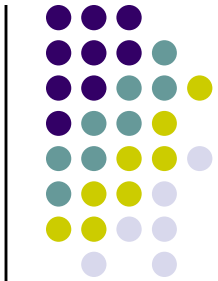
# Deciding Sample Size for Attribute Data



- Steps to determine sample size (continued)
  - Calculate $\sigma_p$, the standard error of the proportion as follows:

$$\sigma_p = \frac{i}{z}$$

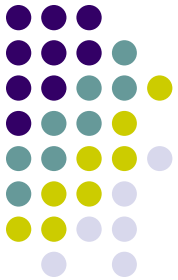
# Deciding Sample Size for Attribute Data



- Steps to determine sample size (continued)
  - Determine the necessary sample size, *n*, using the following formula:

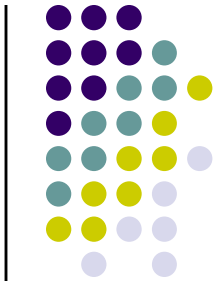
$$n = \frac{p(1-p)}{\sigma_p^2} + 1$$

# Confidence Level Table



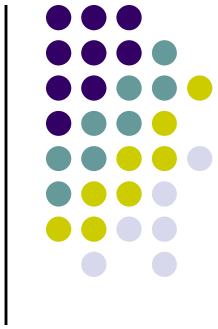
| | |
|---|---|
| 99% | 2.58 |
| 98% | 2.33 |
| 97% | 2.17 |
| 96% | 2.05 |
| 95% | 1.96 |
| 90% | 1.65 |
| 80% | 1.28 |
| 50% | .67 |

# Sample Size for Data on Variables



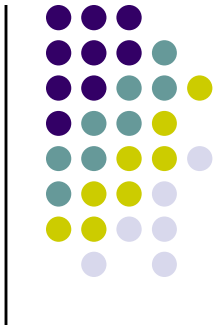
- The steps to determine the sample size when sampling data on variables are
  - Determine the variable you will be sampling
  - Locate the database or reports where the variable can be found

# Sample Size for Data on Variables



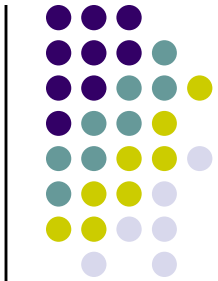
- The steps to determine variable sample size (continued)
  - Examine the variable to gain some idea about its magnitude and dispersion
    - It would be useful to know the mean to determine a more appropriate interval estimate i and the standard deviation, *s* to determine sample size (in the last step)

# Sample Size for Data on Variables



- The steps to determine variable sample size (continued)
  - Make a subjective decision regarding the acceptable interval estimate, *i*
  - Choose a confidence level and look up the confidence coefficient (*z* value)

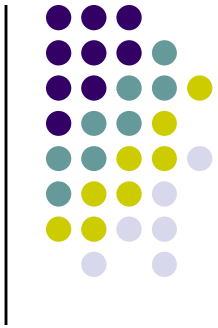
# Sample Size for Data on Variables



- The steps to determine variable sample size (continued)
  - Calculate $\sigma_{x,}$ the standard error of the mean as follows:

$$\sigma_x = \frac{i}{z}$$

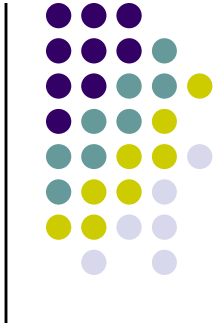
# Sample Size for Data on Variables



- The steps to determine variable sample size (continued)
  - Determine the necessary sample size, *n*, using the following formula:

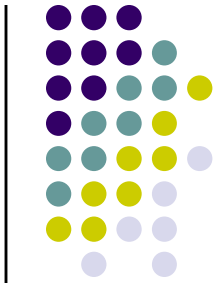
$$n = \left( \frac{s}{\sigma_x} \right)^2 + 1$$

# Hard Data



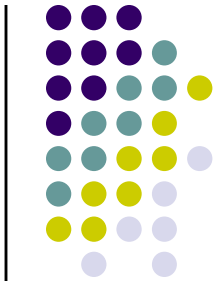
- In addition to sampling, investigation of hard data is another effective method for systems analysts to gather information

# Obtaining Hard Data



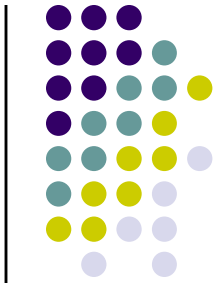
- Hard data can be obtained by
  - Analyzing quantitative documents such as records used for decision making
  - Performance reports
  - Records
  - Data capture forms
  - Ecommerce and other transactions

# Qualitative Documents



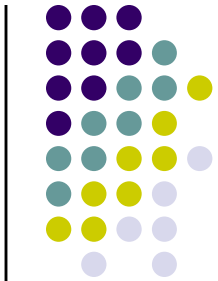
- Examine qualitative documents (includes memos, signs, bulletin boards, manual, policy handbook etc.) for the following:
  - Key or guiding metaphors
  - Insiders vs. outsiders mentality
  - What is considered good vs. evil
  - Graphics, logos, and icons in common areas or Web pages
  - A sense of humor

# Analyzing Qualitative Documents



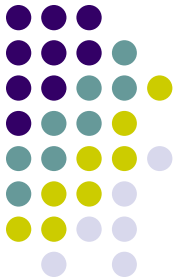
- Qualitative documents include
  - Memos
  - Signs on bulletin boards
  - Corporate Web sites
  - Manuals
  - Policy handbooks

# Workflow Analysis



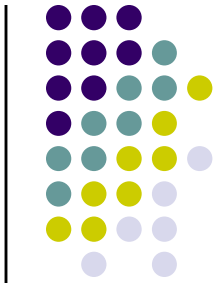
- Workflow analysis may reveal signs of larger problems, such as
  - Data or information doesn't flow as intended
  - Bottlenecks in the processing of forms
  - Access to online forms is cumbersome
  - Unnecessary duplication of work occurs because employees are unaware that information is already in existence
  - Employees lack understanding about the interrelatedness of information flow

# Business Process Reengineering



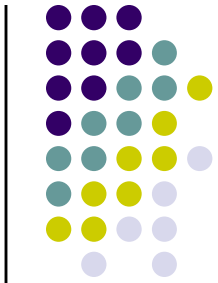
- Business process reengineering software includes the following features:
  - Modeling of the existing system
  - Analysis of possible outcomes
  - Simulation of proposed work flow

# Archival Documents



- A systems analyst may obtain some valuable information by abstracting data from archival documents
- Generally, archival documents are historical data, and they are prepared and kept by someone else for specific purposes

# Guidelines for Abstracting Archival Data



- Fragment data into subclasses and make cross-checks to reduce errors
- Compare reports on the same phenomenon by different analysts
- Realize the inherent bias associated with original decisions to file, keep, or destroy reports
- Use other methods to obtain data