Data
collection

L2

Data Collection

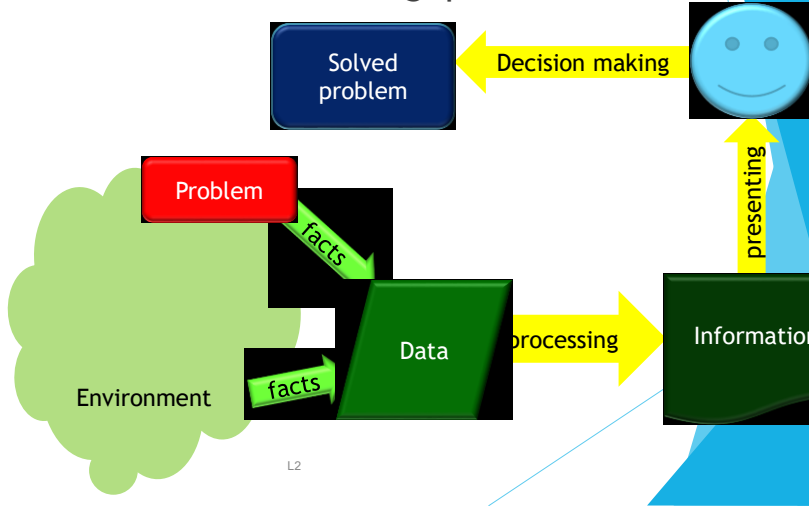
- The **management of data** is an important skill to develop.
- In some situations, the data requirement is clear, and in other is less clear. Mostly you will find some data and need to add some.
- Data have to be:
  - Appropriate
  - Adequate
  - Without bias

Data Collection

- It is always possible to collect more and more data. So where do we stop?
- You will:
  - need to be clear about problem boundaries;
  - need to know what the problem owner or client expect from you;
  - need to know if any important data is missing
  - be expected to work within time and resource constraints
  - need to decide whether the current data is sufficient for the purpose or whether additional data should be acquired

Data Collection

- Data in decision-making process.



Timing and quantity of data collection

- Data are collected for a specific purpose and the way they are used should have an effect on the way they are collect.
- We should design data collection to meet its specific purpose, and not the other way around.

Data collection process should be designed **after** deciding the use of the data.

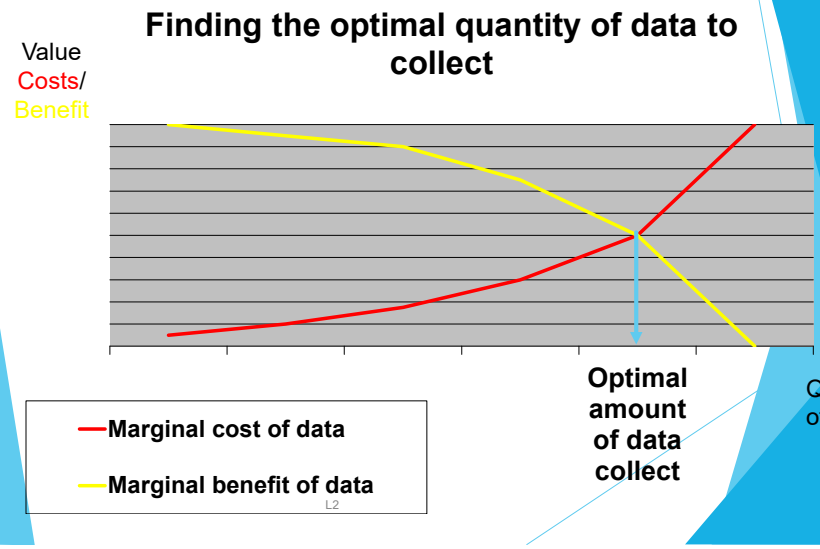
Timing and quantity of data collection

- How much data to collect?

- In many cases, there is an almost limitless amount of data which could be collected and might be useful.
- Data collection and processing costs money and collecting unnecessary data is wasteful.
- You should find the optimal amount of data to collect.

The marginal benefit of data is the benefit of the last 'unit' of data collected.

Timing and quantity of data collection



L2

7

Timing and quantity of data collection

- Collecting more data than optimal amount will be wasteful, but collecting fewer data would lose some potential benefit.
- Problem – difficulty of defining the cost and benefit of the data collected.
- Suggestion – not calculate but take in mind previous experience.

L2

8

Timing and quantity of data collection

- Another factor – available time for data collection.
- The time available can limit both the type of data that can be collected and the amount of data.
- It is a common view that some data, even if they are slightly inaccurate, are better than no data at all.
- In many circumstances, however, wrong data can be worse than no data at all.

- Why is data collection important for an organization?

Types of data

- Data of different types are collected in different ways.
- Classification of data:

  Qualitative and quantitative;

  Depending on how well data can be estimated:
  - Nominal (categorical);
  - Ordinal;
  - Cardinal (metric);

Types of data

- Nominal (categorical data). This is the kind of data which really cannot be quantified with any meaningful units.
  - The fact that a company is a manufacturer or a country operates a centrally planned economy, or a cake has cream in it, are examples of nominal data.
- A common analysis of nominal data defines a number of different categories and says how many observations fall into each.
  - A survey of companies in a particular area might show that there are 7 manufacturers, 16 service companies and 5 in primary industries.

Types of data

- Ordinal data – one step more quantitative, in that the categories into which observations are divided can be ranked in some order. The order of the categories is important.
  - Sweaters may be described as extra large, large, medium, small or extra small.
- Sometimes, when there are few observations, they can all be ranked individually rather than put into ranked categories.

Types of data

- Cardinal data have some attribute which can be directly measured. The measures give a precise description of a particular characteristic.
  - The weight of a product, time to perform a task, temperature in an office.
- Cardinal data is generally the easiest to analyze and are most relevant to quantitative methods.
- Cardinal data can be divided into two types depending on whether they are discrete or continuous.

Types of data

Measurement data

- Discrete data can only take integer values.
  - A number of children in family, cars owners, machines operated, people employed.
- Continuous data can take any value and are not restricted to integers.
  - The weight of a bag of biscuits, time period, the length of metal bars.
- Sometimes there is a mismatch in data types.
  - The circumferences of men's necks are continuous data, but shirt collars use a discrete measure.

Types of data

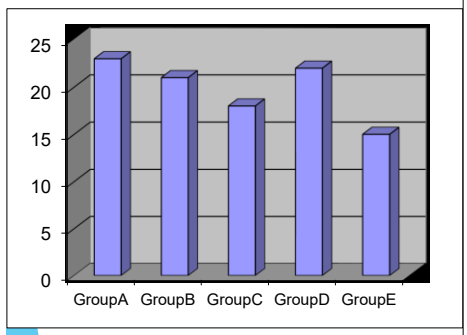
- Discrete or Continuous data?

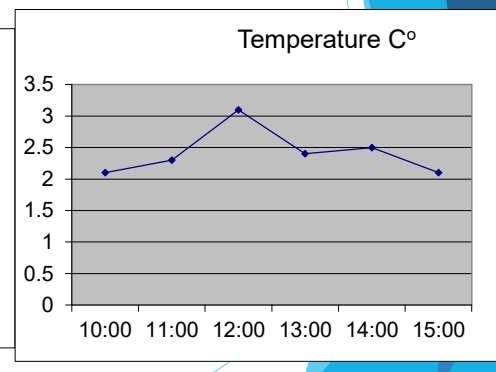
	Number of children
GroupA	23
GroupB	21
GroupC	18
GroupD	22
GroupE	15

Time	Temperature C°
10:00	2.1
11:00	2.3
12:00	3.1
13:00	2.4
14:00	2.5
15:00	2.1

Types of data

- Graphical presentation of Discrete and Continuous data.





Types of data

Depending on the method of data collection it may be primary or secondary.

- Primary data are collected by the organization itself for the particular purpose.
- Secondary data are collected by other organizations for other purposes.
- What are the benefits of the primary and secondary data?

Sampling methods

- Sometimes, the entire population will be sufficiently small, and the researcher can include the entire population in the study. This type of research is called a census study because data is gathered on every member of the population.
- Population in its statistical sense is the set of all items or people which could supply data.

All letters which are posted first class, all pote
customers of a product, all people in a regio

Sampling methods

- Census – data are collected from every member of the population. The sample is the same as the population.
- Usually, the population is too large for the researcher to attempt to survey all of its members. A small, but carefully chosen sample can be used to represent the population. The sample reflects the characteristics of the population from which it is drawn.

Sampling methods

- The purpose of sampling – obtaining primary data to get over missing secondary data, and to get reliable results using only a sample of the whole population.
- Data are collected from a representative sample of items or people, and these are used to infer characteristics of all items or people.

Type of sample

- Sampling methods are classified as either *probability* or *nonprobability*.
- In probability samples, each member of the population has a known non-zero probability of being selected.
  - Random sample
  - Systematic sample
  - Stratified sample

Type of sample

- Random sample – every member of the population has exactly the same chance of being selected for data collection. When there are very large populations, it is often difficult or impossible to identify every member of the population, so the pool of available subjects becomes biased.
- Excel

Type of sample

- Systematic sample – collect data at regular intervals. It is often used instead of random sampling. It is also called a N^{th} name selection technique. After the required sample size has been calculated, every N^{th} record is selected from a list of population members. As long as the list does not contain any hidden order, this sampling method is as good as the random sampling method. Its only advantage over the random sampling technique is simplicity.

Type of sample

- Stratified sample –is commonly used probability method that is superior to random sampling because it reduces sampling error. A stratum is a subset of the population that shares at least one common characteristic.
- Examples of stratum might be males and females, or managers and non-managers.

Type of sample

- The researcher first identifies the relevant stratum and their actual representation in the population.
- Random sampling is then used to select a sufficient number of subjects from each stratum. "Sufficient" refers to a sample size large enough for us to be reasonably confident that the stratum represents the population. Stratified sampling is often used when one or more of the stratum in the population have a low incidence relative to the other stratums.

Type of sample

- The advantage of probability sampling is that sampling error can be calculated.
- Sampling error is the degree to which a sample might differ from the population.
- When inferring to the population, results are reported plus or minus the sampling error.

Type of sample

- In nonprobability sampling, members are selected from the population in some nonrandom manner.
  - Convenience sampling,
  - Judgment sampling,
  - Quota sampling,
  - Snowball sampling.

Type of sample

- Convenience sampling is used in exploratory research where the researcher is interested in getting an inexpensive approximation of the truth. As the name implies, the sample is selected because they are convenient. This nonprobability method is often used during preliminary research efforts to get a gross estimate of the results, without incurring the cost or time required to select a random sample.

Type of sample

- **Judgment sampling** is a common nonprobability method. The researcher selects the sample based on judgment. This is usually and extension of convenience sampling.
- For example, a researcher may decide to draw the entire sample from one "representative" city, even though the population includes all cities.
- When using this method, the researcher must be confident that the chosen sample is truly representative of the entire population.

Type of sample

- **Quota sampling** is the nonprobability equivalent of stratified sampling. Like stratified sampling, the researcher first identifies the stratum and their proportions as they are represented in the population.
- Then convenience or judgment sampling is used to select the required number of subjects from each stratum. This differs from stratified sampling, where the stratum are filled by random sampling.

Type of sample

- **Snowball sampling** is a special nonprobability method used when the desired sample characteristic is rare. It may be extremely difficult or cost prohibitive to locate respondents in these situations.

Type of sample

- **Snowball sampling** relies on referrals from initial subjects to generate additional subjects. While this technique can dramatically lower search costs, it comes at the expense of introducing bias because the technique itself reduces the likelihood that the sample will represent a good cross section of the population.

Type of sample

- In nonprobability sampling, the degree to which the sample differs from the population remains unknown.

Type of sample

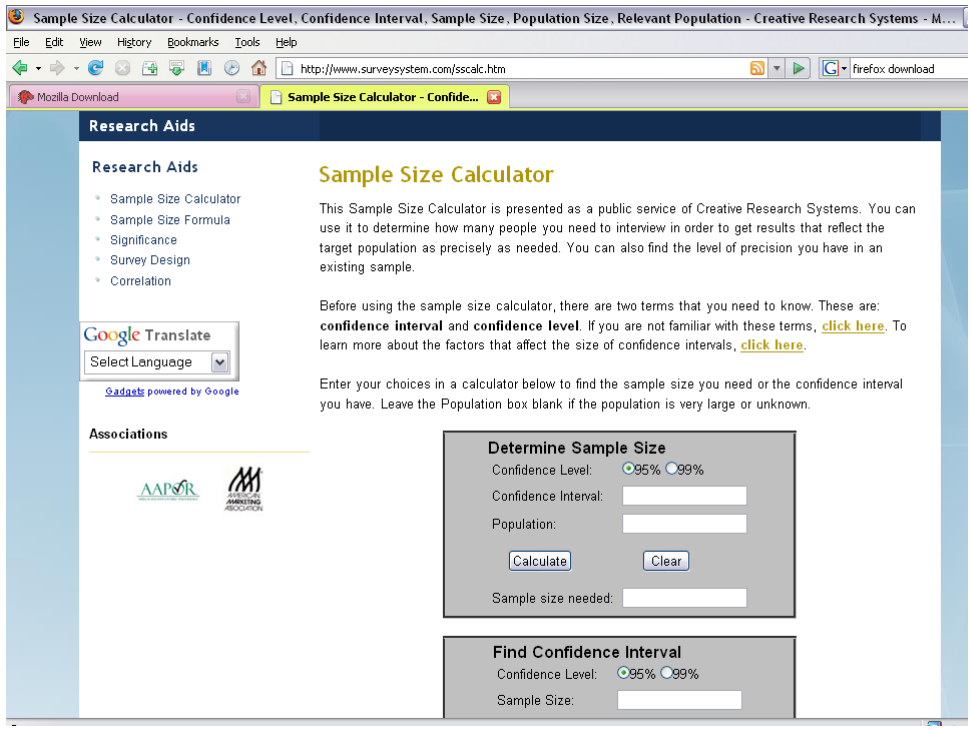
- Two additional sampling methods are used, when the population is too big and nonhomogeny.
  - Multi-stage sample
  - Cluster sample

Type of sample

- Multistage sample – using other sampling methods in two or more stages to find reliable samples.
  - An organization could simply take a random sample of a population, the take a sample of it, for example wi quota method.

Type of sample

- Cluster sampling – chooses the items in a sample not individually, but in clusters.
  - From people living in a town, we visit a sample in a single area than to visit a sample spread over the whole town


Sample Size Calculator - Confidence Level, Confidence Interval, Sample Size, Population Size, Relevant Population - Creative Research Systems - M...
File Edit View History Bookmarks Tools Help
http://www.surveysystem.com/sscalc.htm
firefox download
Mozilla Download
Sample Size Calculator - Confide...
Research Aids
Research Aids
Sample Size Calculator
Sample Size Formula
Significance
Survey Design
Correlation
Google Translate
Select Language
Gadgets powered by Google
Associations
AAPOR
Sample Size Calculator
This Sample Size Calculator is presented as a public service of Creative Research Systems. You can use it to determine how many people you need to interview in order to get results that reflect the target population as precisely as needed. You can also find the level of precision you have in an existing sample.
Before using the sample size calculator, there are two terms that you need to know. These are: confidence interval and confidence level. If you are not familiar with these terms, click here. To learn more about the factors that affect the size of confidence intervals, click here.
Enter your choices in a calculator below to find the sample size you need or the confidence interval you have. Leave the Population box blank if the population is very large or unknown.
Determine Sample Size
Confidence Level: 95% 99%
Confidence Interval:
Population:
Calculate
Clear
Sample size needed:
Find Confidence Interval
Confidence Level: 95% 99%
Sample Size:

20

Sample Size

- The **confidence interval** (also called margin of error) is the plus-or-minus figure usually reported in newspaper or television opinion poll results. For example, if you use a confidence interval of 4 and 47% percent of your sample picks an answer you can be "sure" that if you had asked the question of the entire relevant population between 43% (47-4) and 51% (47+4) would have picked that answer.

Sample Size

- The **confidence level** tells you how sure you can be. It is expressed as a percentage and represents how often the true percentage of the population who would pick an answer lies within the confidence interval. The 95% confidence level means you can be 95% certain; the 99% confidence level means you can be 99% certain. Most researchers use the 95% confidence level.

Sample Size

- When you put the confidence level and the confidence interval together, you can say that you are 95% sure that the true percentage of the population is between 43% and 51%. The wider the confidence interval you are willing to accept, the more certain you can be that the whole population answers would be within that range.

Sample Size