

Research Methodology : Data Collection



Data Needs



Data Types



Data Collection Methods

Anil Minhans

B.Tech. (Civil Engineering), M.Pl.(Transport Planning)

Dr.-Ing.(Traffic Engineering and Transport Planning)

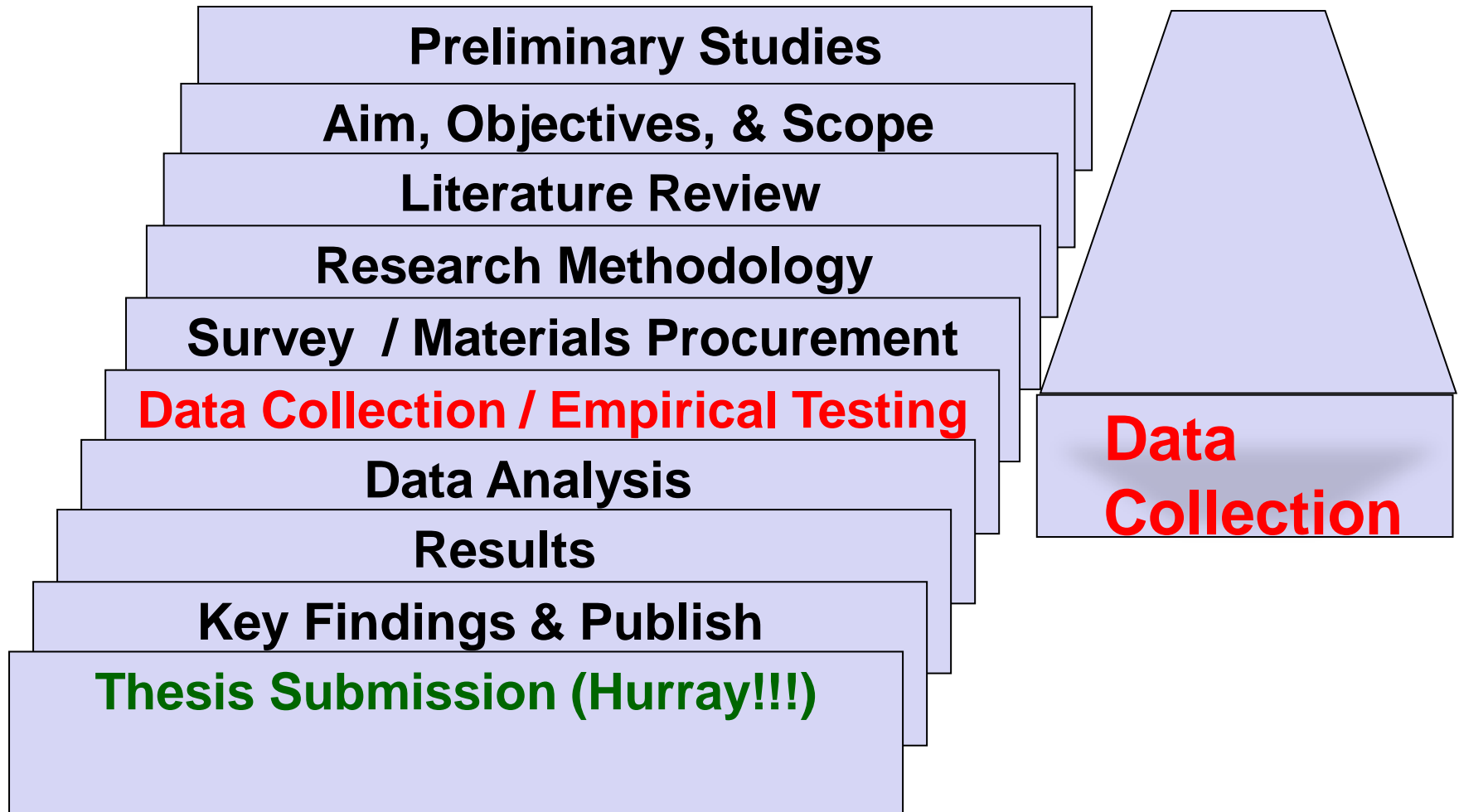
Faculty of Geotechnics and Transportation

Venue: Universiti Teknologi Malaysia

18. Mar, 2012

Skudai, Johor Bahru

Research: Life Span



Research Framework



Research statement
"(need of research), Title of research"

e.g. Climate change will affect structural reliability

e.g. Climate change will affect soil salinity

e.g. Climate change will affect housing

Goal 1

Goal 2

Goal 3

Goal 4

Research Goals

Objective 1

Objective 2

Objective 3

Objective 4

Objective 1

Objective 2

Objective 3

Objective 1

Objective 2

Objective 3

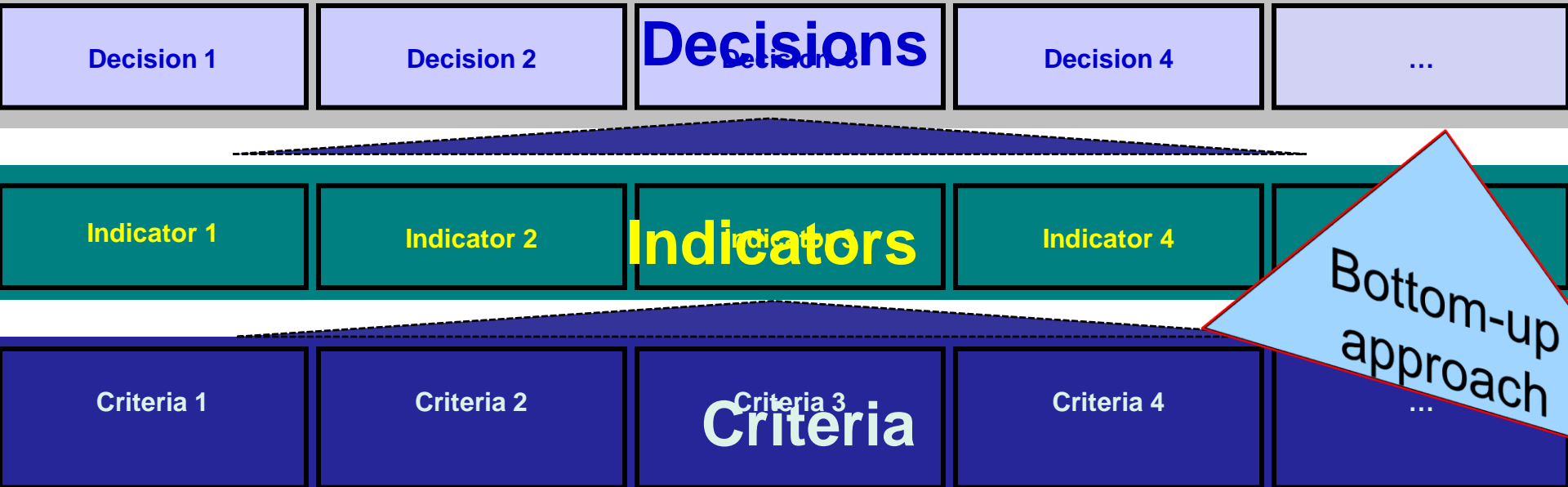
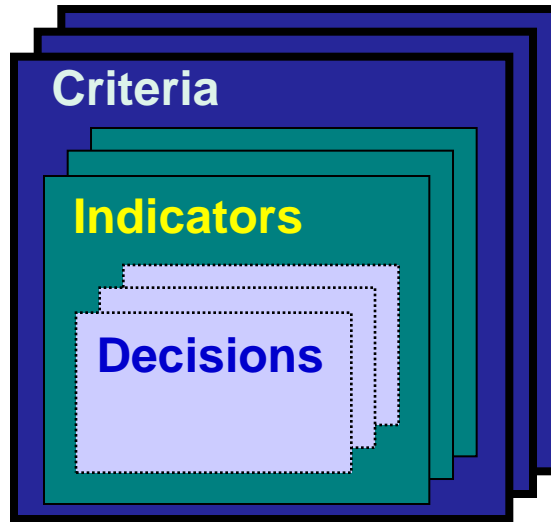
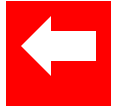
Objective 1

Objective 2

Objective 3

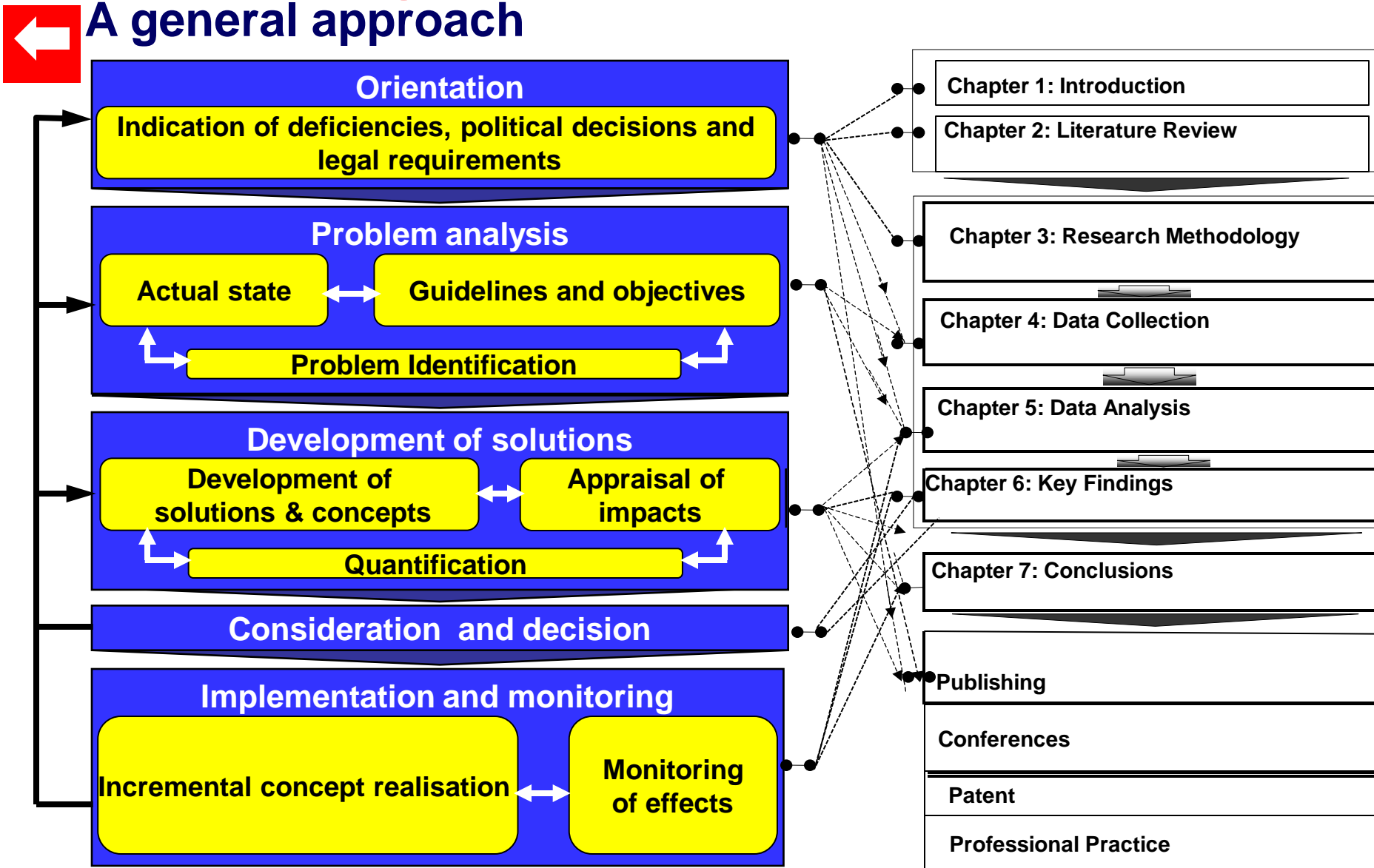
Research Objectives

Research Framework



Research Design

A general approach



Adapted from Source :FGSV, 2001

Dissertation Chapters/ Post Research

Research Design

A general approach

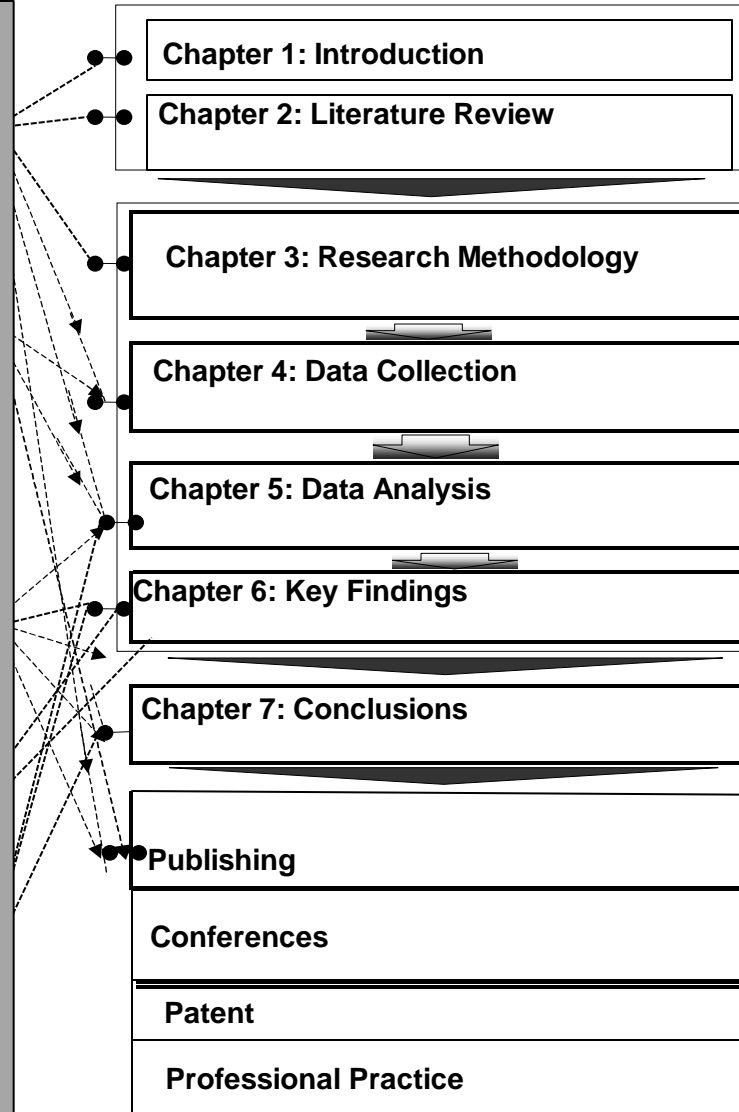
Objectives/Scope of enquiry,
Statistical Units to be used,
Sources of Information,
Type of Enquiry
ARE FIXED



DATA COLLECTION IS CARRIED OUT



**DATA ANALYSIS ON DATA COLLECTION
IS CONDUCTED**



Dissertation Chapters/ Post Research

Choice of Data

Proposing a study or experiment to collect meaningful data to answer the problem.

- What variables have to be measured?
- How many data/samples have to be collected?
- How the samples should be selected?
- Etc..



The most crucial element in data collection is the manner in which the sample is selected.

Choice of Data

Proposing a study or experiment to collect meaningful data to answer the problem.

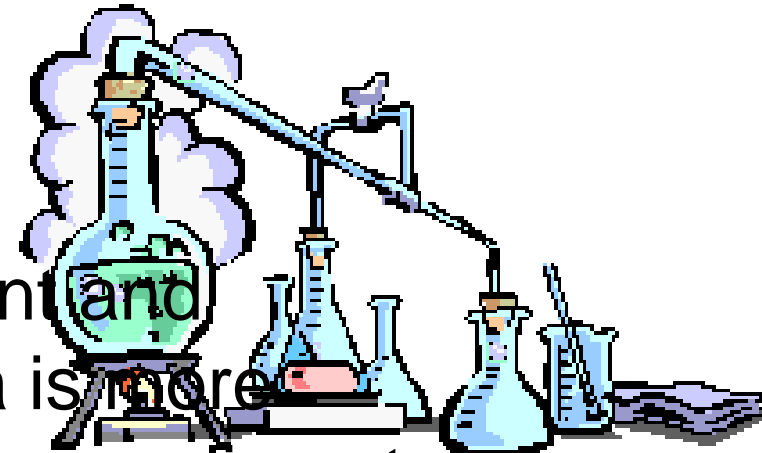
- **Primary Data?**

(direct observation, experiments, study of natural or manmade phenomena)

- **Secondary Data?**

Detached from the original

- Both data sources compliment and supplements, secondary data is more supplementary in nature than complimentary

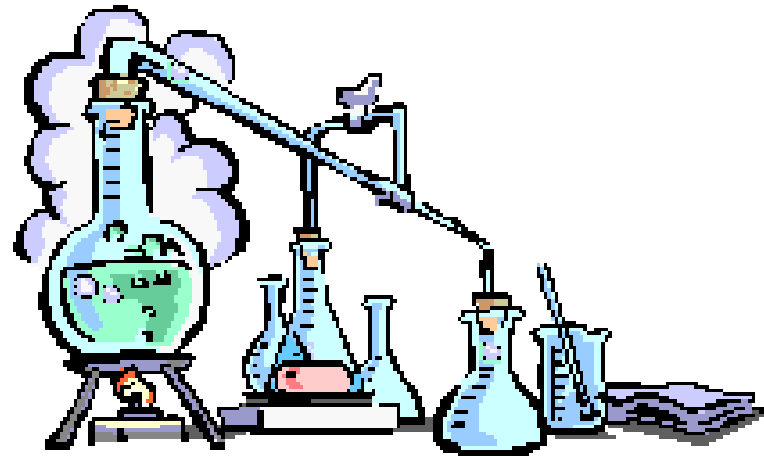


Choice of Data

- Degree of accuracy
- Objectives and Scope of study (open-confidential, original-repetitive, official-non, census or sample)
- Time of collection
- Statistical Units to be used (simple or composite units)
- Status of the investigator
- Degree of detail

Where secondary data is used,

- Availability of funds
- Trained investigators



Population and Samples

Population

The population is the total set of measurements which could hypothetically be taken from the entity being studied. It is the set of all measurements of interest. For example, trace element composition of all stream sediments in a area, groundwater depth as any point of a catchment, etc.

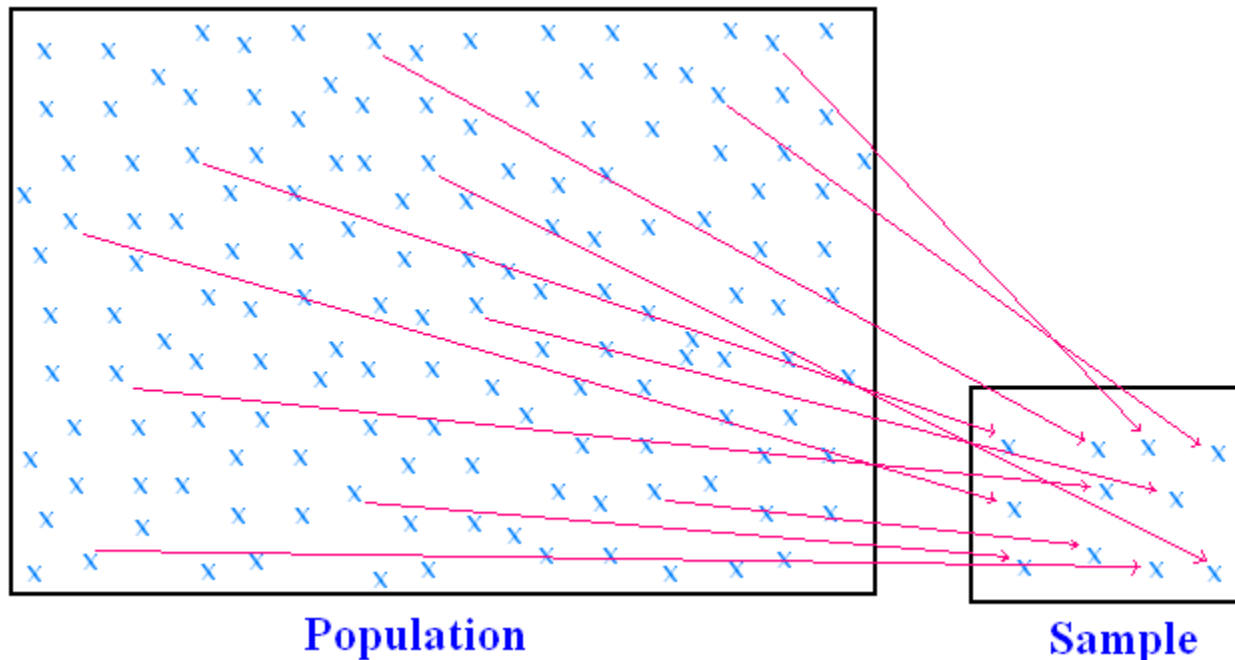
The Statistical Sample (make sure it is fraction!!!)

A sample is any subset of measurements selected from the population. There may be confusion about sample in water resources and sample in statistical. Water sample usually means one bottle of water. In statistics, sample is a data that is actually available for analysis.

**Sampling is the process of learning about popln.
On the basis of sample drawn from it!!!**

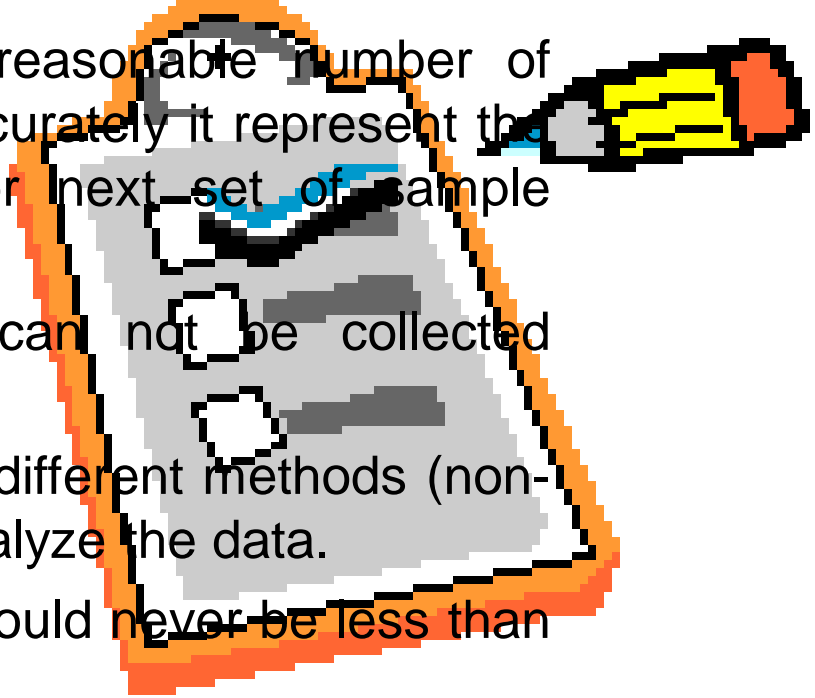
Population and Samples

- ✓ Sample is a subset of population.
- ✓ Sample should be collected in such a way that it represents the whole population
- ✓ All the items «(Sample) » Too few items



Sampling Size

- Generalization of sample size is not possible to made.
- It depends on the degree of intricacy of the problem being addressed.
- Many times, sample size is possible to know in advance.
- Many cases, hydrologists collect reasonable number of sample and analyze to see how accurately it represent the population. According, they go for next set of sample collection.
- However, many cases samples can not be collected according to need.
- If the sample size is not adequate, different methods (non-parametric) methods are used to analyze the data.
- Whatever it may be, sample size should never be less than 6.

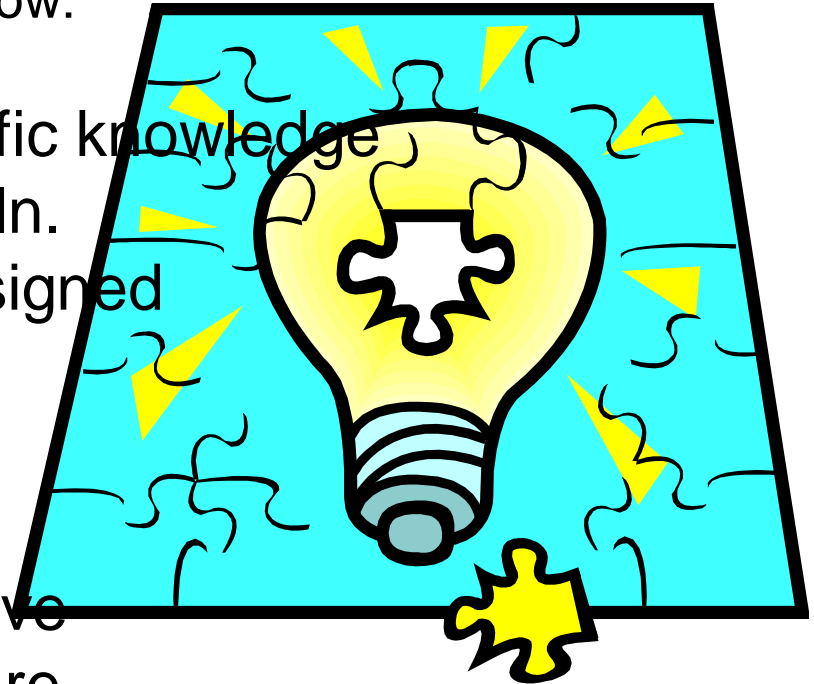


Sampling Techniques

In civil engineering, one of the main decision about sampling population is where, what and how to sample. Sampling techniques can be classified as below:

-Probability Sampling, scientific knowledge of drawing samples from popln. each unit has definite pre-assigned probability of being chosen

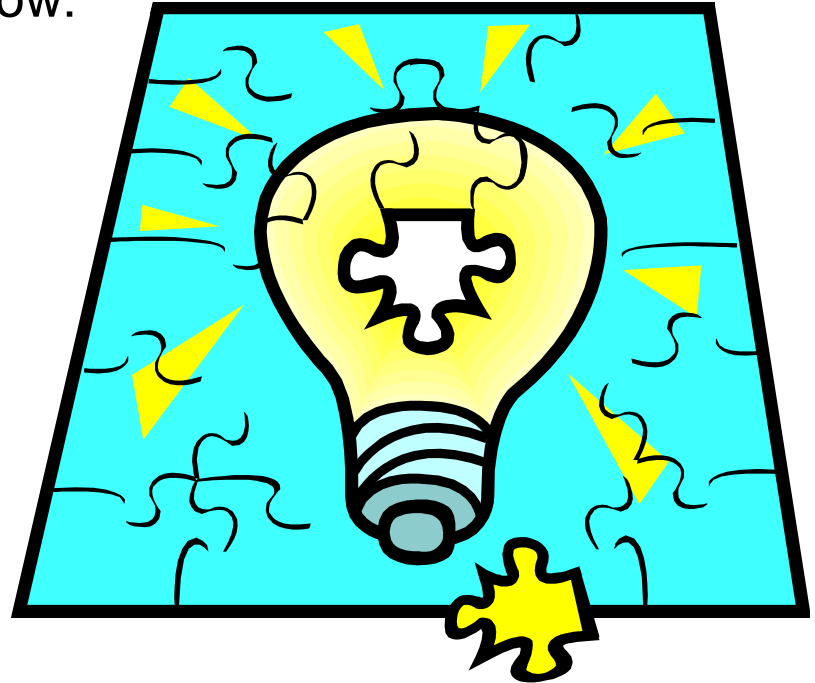
-Non-probability Sampling, Based on judgments, purposive sampling, desired # of units are selected deliberately



Sampling Techniques

In civil engineering, one of the main decision about sampling population is where, what and how to sample. Sampling techniques can be classified as below:

- i. Random Sampling
- ii. Stratified Sampling
- iii. Uniform Sampling
- iv. Regular Sampling
- v. Clustered Sampling
- vi. Traverse Sampling

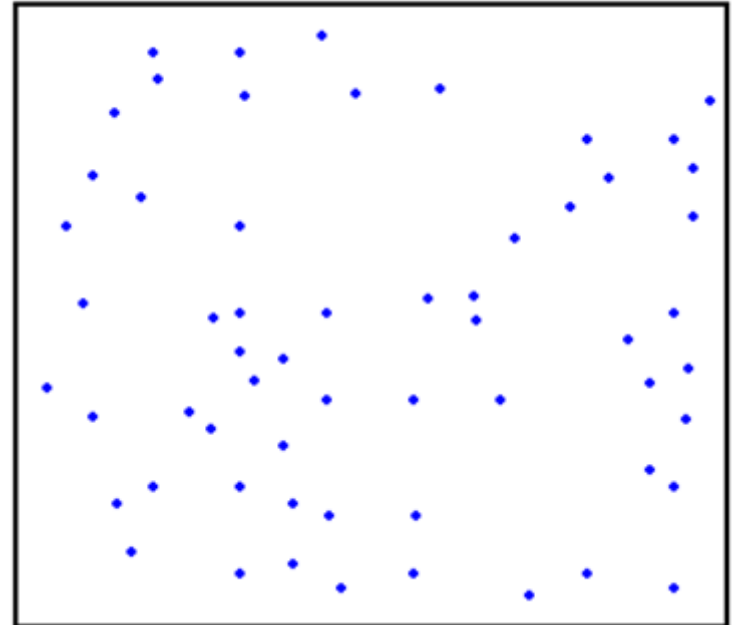


Random Sampling

A random sample of n measurements selected from a population containing N measurements ($N > n$). A sample of n measurements selected from a population is said to be a random sample if every different sample of size n from the population has an equal probability of being selected.

Sample data selected in a nonrandom fashion are frequently distorted by a selection bias. A selection bias exists whenever there is a systematic tendency to over-represent or under-represent some part of the population.

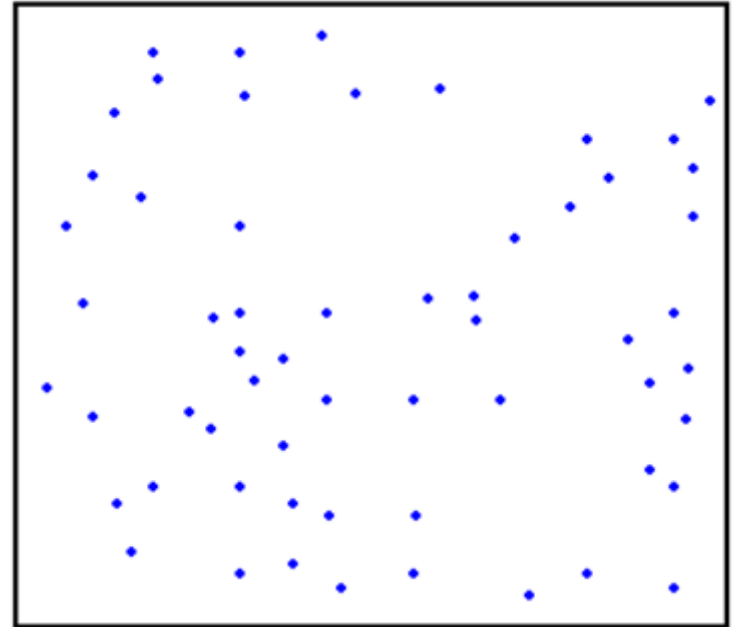
Random



Selection of a Random Sample

- i. **Lottery Method**, blindfold selection is made to constitute the desired size of sample
- ii. **Tables of random #s**, #s occurs with approximately the same frequency and independently. Tippet, Fisher & Yates, Kendall & Bobington Smith. Sample regards accuracy and representativeness.
- iii. **Sequential list**, e.g. every tenth student from class, based on arrival and departure pattern etc.
- iv. **Grid system**, for selecting a sample area, grid is placed on map to select areas falling in selected squares.

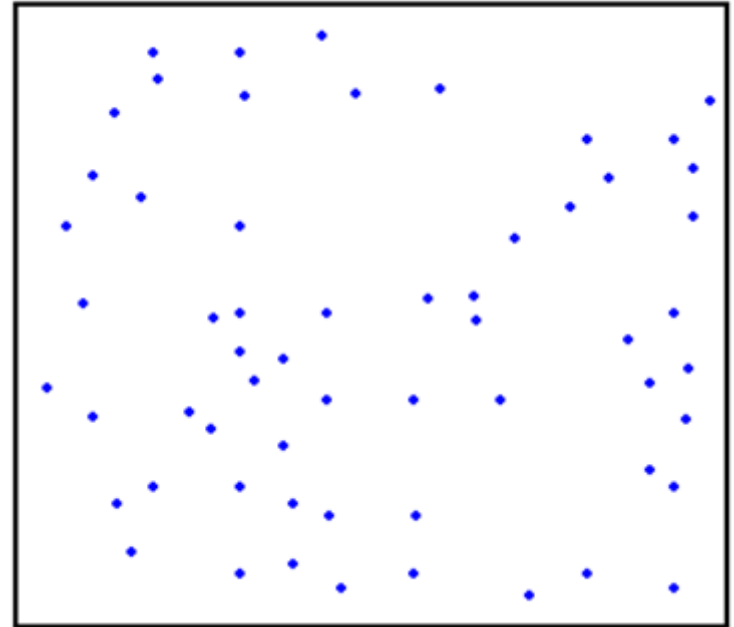
Random



Selection of a Random Sample

- i. **Precaution**, units must clearly defined, approx equal size, independent, method of selection: independent, fully accessible
- ii. **Merits**, most scientific method, no scope of bias, no knowledge required, accurate assessment possible by sample error estimation, simple and easy, most reliable and max.info. @ least cost, time, money and labor. True representative of universe
- iii. **Demerits**, full popln. must, may be time consuming- due to geographical scatter, in small samples: not true representative of popln, leads to result with low probability, larges samples required for more accuracy

Random

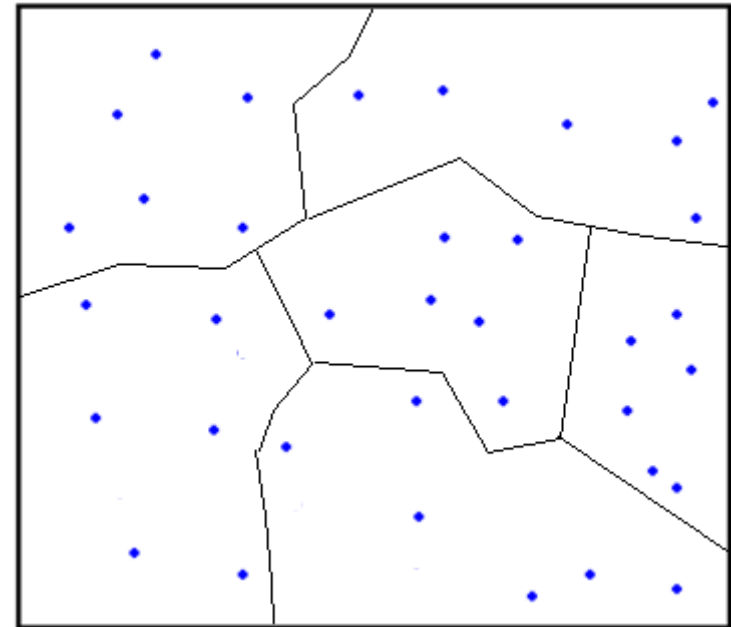


Stratified Sampling

The whole population is divided into a number of groups or strata according to a property and each group is sampled differently. This type of sampling is called Stratified Sampling.

Let, we want to study transmissivity of groundwater in a catchment. Transmissivity of groundwater heavily depends on local geology. Therefore, if we divided the catchment according to geological units to study the impact of geology on transmissivity.

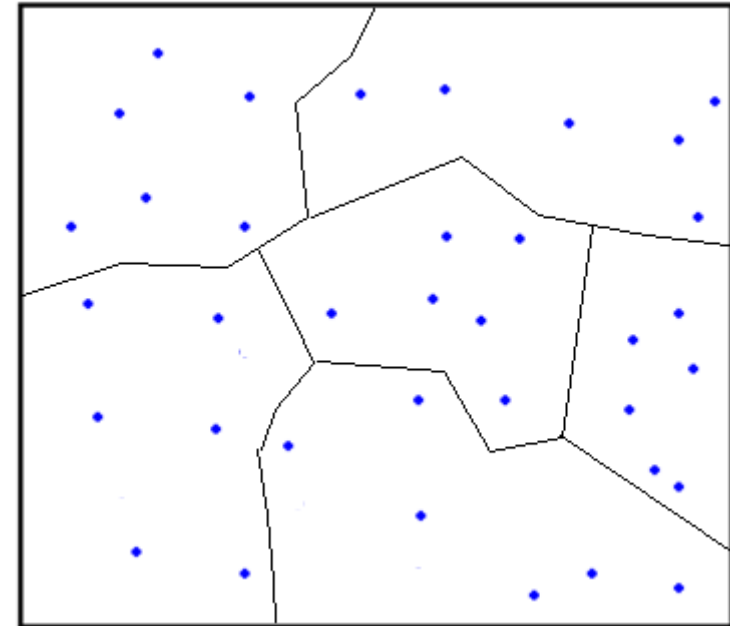
Stratified



Process of Stratifying

- i. **Stratification**, formation into groups, samples selected at random.
- ii. **Uniformity in selection of units**, items in one stratum must be similar, and different in units when other strata are considered.
- iii. **Uniqueness**, must belong to one stratum only, no overlapping allowed.
- iv. **Largeness**, must be large to allow drawing samples.
- v. **Proportionality**, can be proportional or disproportional to the size of a stratum

Stratified



Selection of a Sample

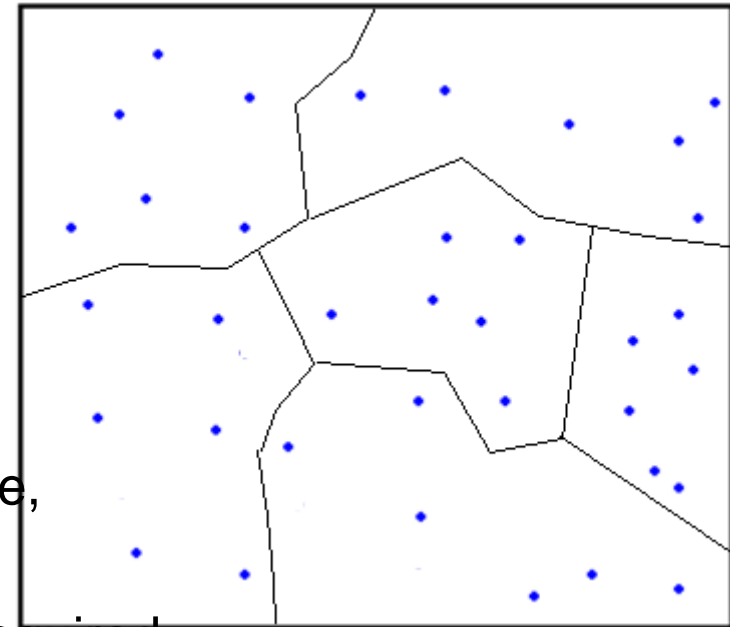
Merits

- Small units form representative sample
- No group is under or over-represented
- Avoid bias @ random selection
- More precise, save time and cost
- Different 's of accuracy \leftrightarrow different strata

Demerits

- Homogeneous strata are difficult to form
- Overlapping, unsuitable or disproportionate,
- Faulty stratification: biased results
- Disproportionate stratification: weighting required
- Introduces factor in sample, leading to further chaos

Stratified

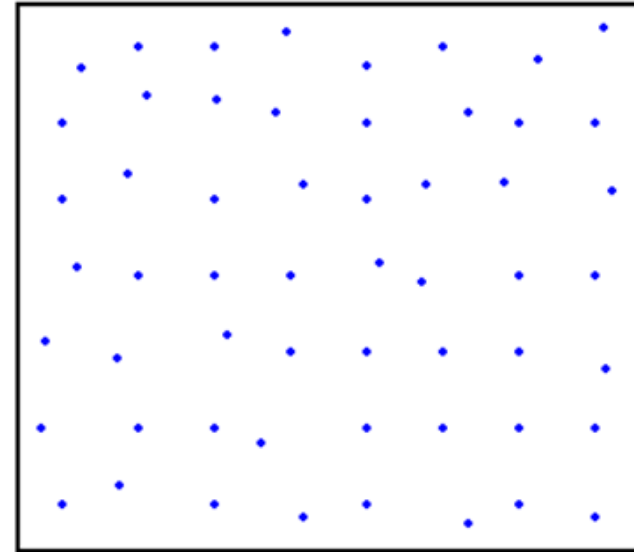


Systematic Sampling

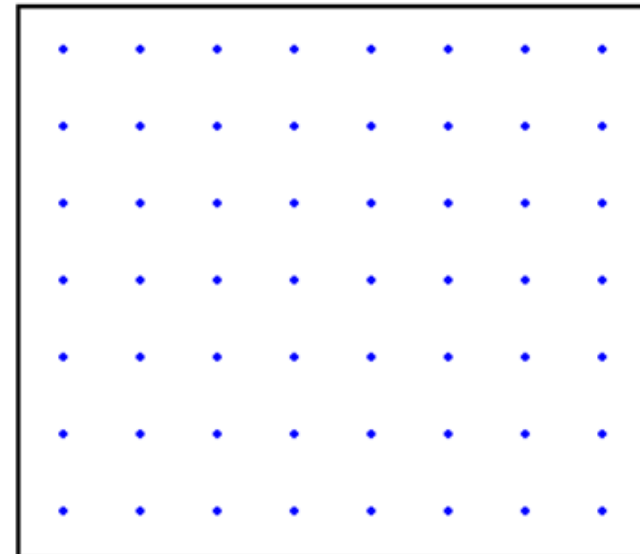
Uniform Sampling: Planned by randomization within grid squares.

Regular or gridded: Planned on rectangular or triangular grid.

Uniform



Regular



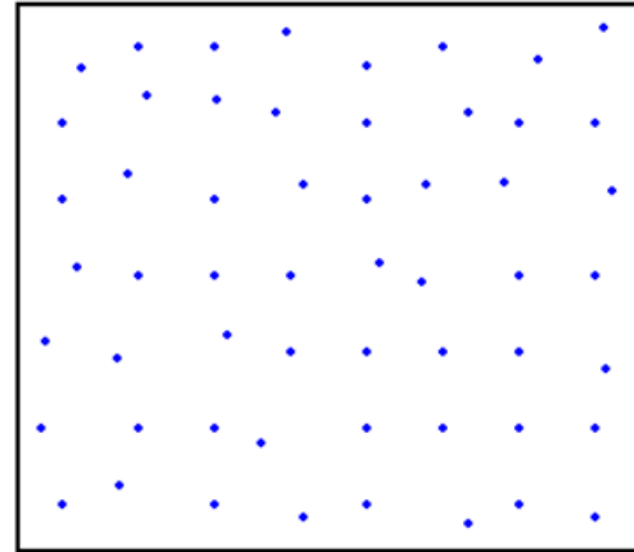
Process of Systematic Sampling

- The whole population is arranged in serial numbers from 1 to N
- Size of the sample n is the determined,
- The sampling interval is determined (K)

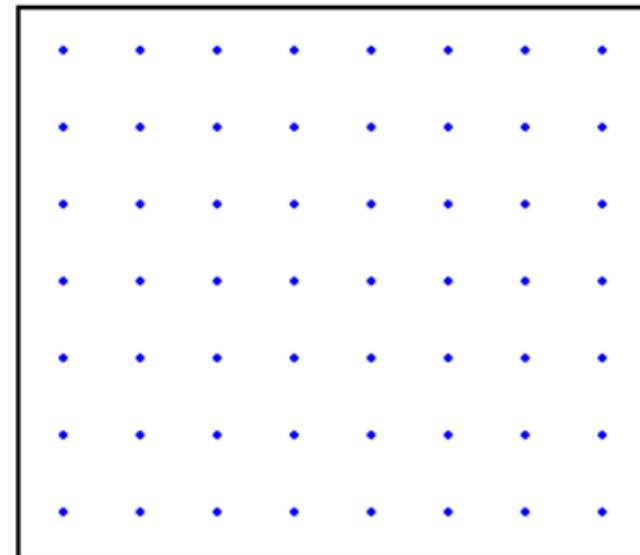
$$N / n = K$$

- Any number is selected from the first sampling interval and subsequent samples are selected at equal or regular intervals.

Uniform



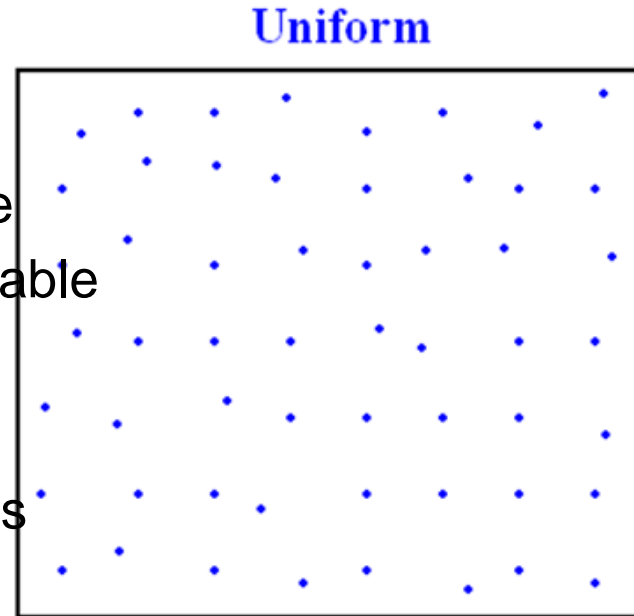
Regular



Process of Systematic Sampling

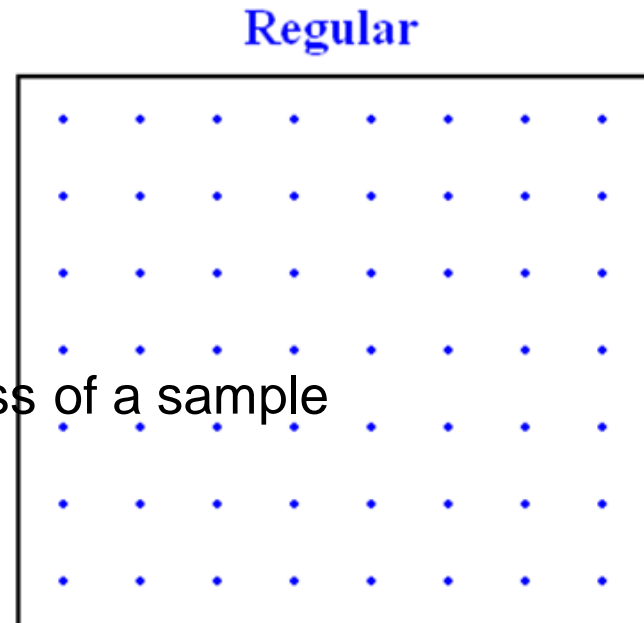
Merits

- Easy to operate and checking can quickly done
- Randomness and probability features are available
- Sample interval can be adjusted
- Basis is the alphabetic order, house no. etc.
- Selection of first random sample is a good basis



Demerits

- Comprehensive or full data is difficult to obtain
- Units are difficult to arrange randomly
- Hidden seasonal effects: biased results
- Hidden periodicity: biased results
- Above factors affect the true representativeness of a sample

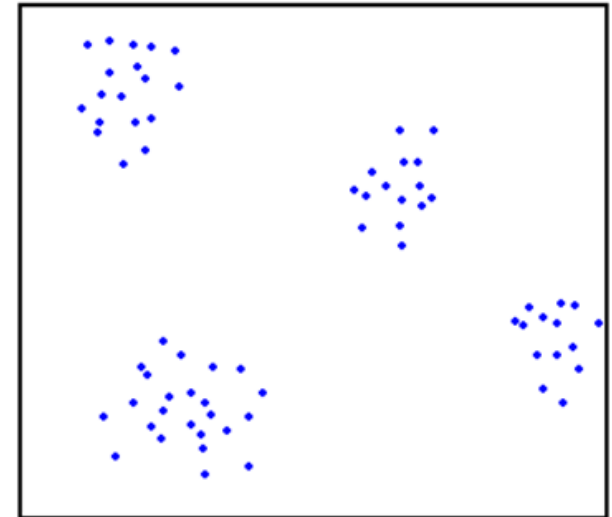


Clustered Sampling

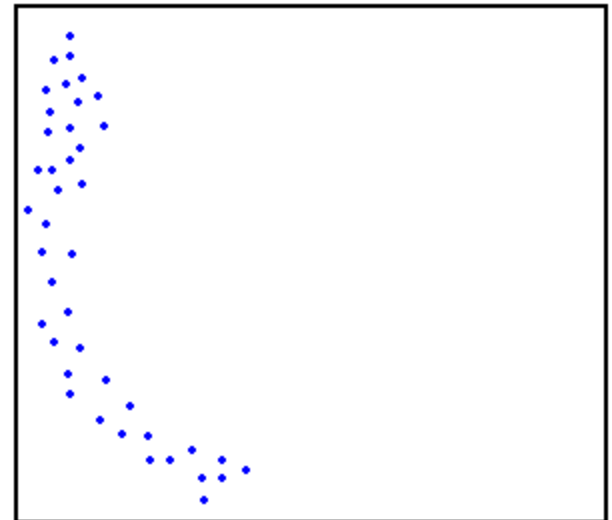
Clustered: It is focused on patchy distribution.

Traverse: Often forced by access and exposure constraints or logistics.

Clustered

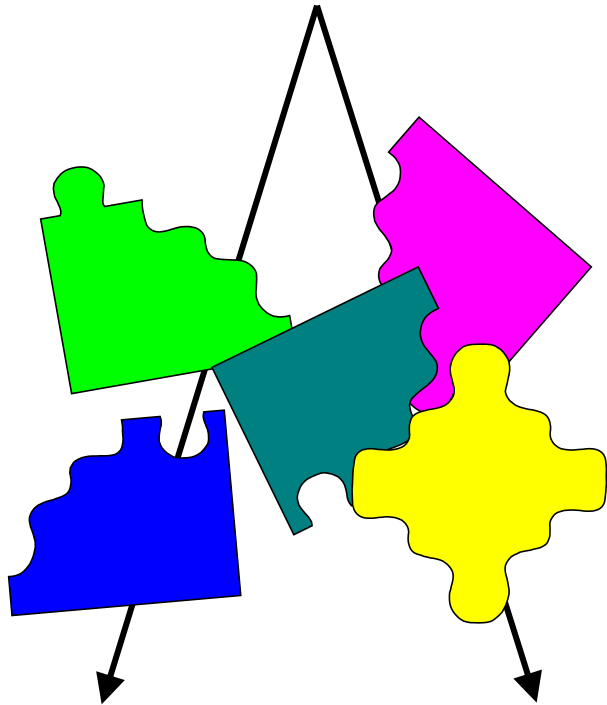


Traverse

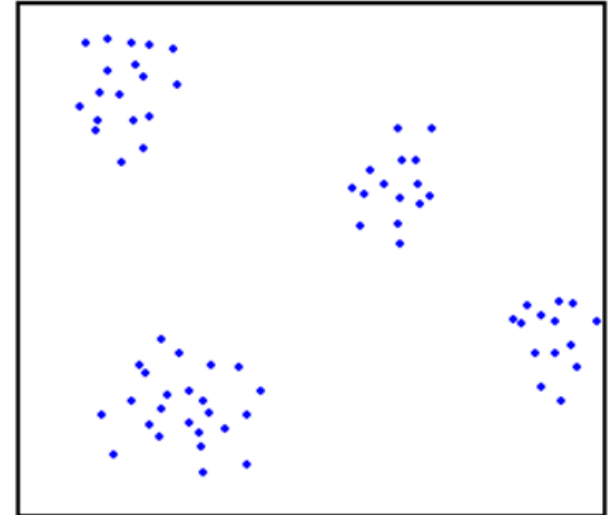


Clustered Sampling

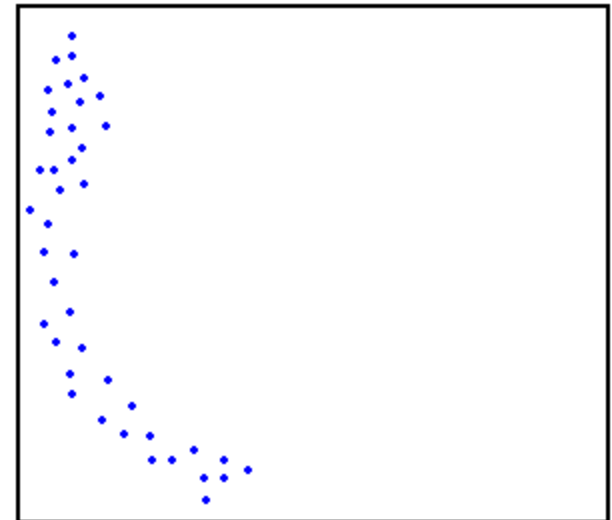
-The whole population is divided into recognizable sub-divisions which are called clusters and a simple random sample of these clusters is drawn and then the survey of each and every unit of in this cluster is made.



Clustered



Traverse



Clustered Sampling

Principles

- Must be manageable cluster, so make it small
- Cluster must represent constraints of survey
- #s of units in each cluster must be approximately same

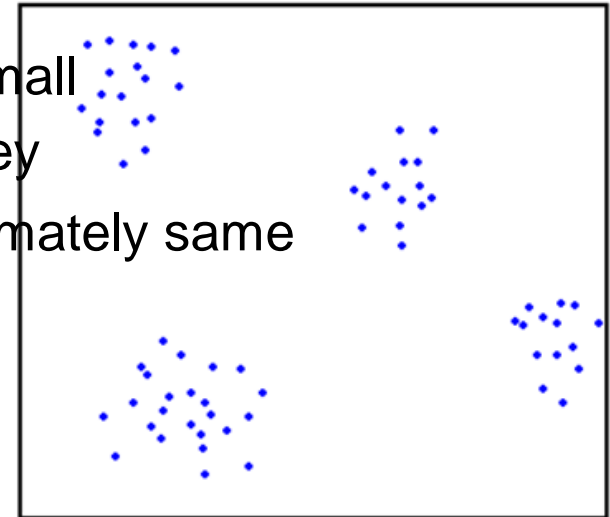
Merits

- Significant cost gain
- Most practical, facilitates field work

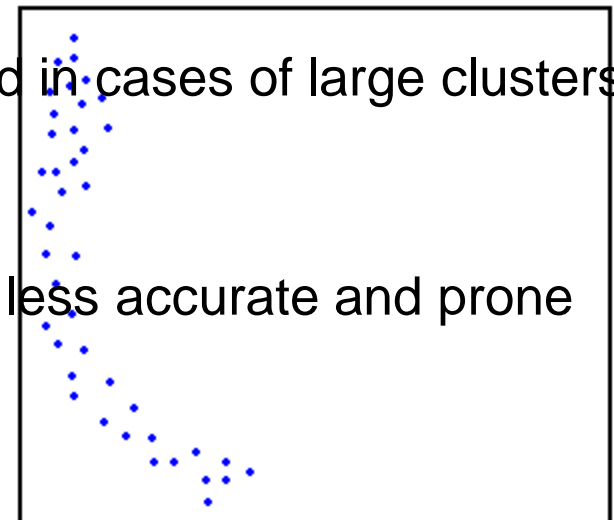
Demerits

- Probability and representativeness affected in cases of large clusters
- Hidden seasonal effects: biased results
- Hidden periodicity: biased results
- If sampling units are not same, results are less accurate and prone to biases.

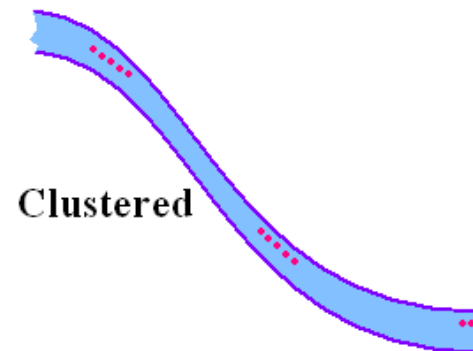
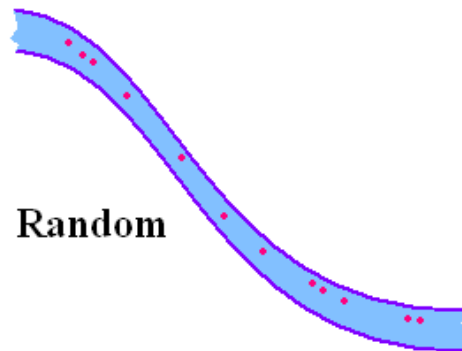
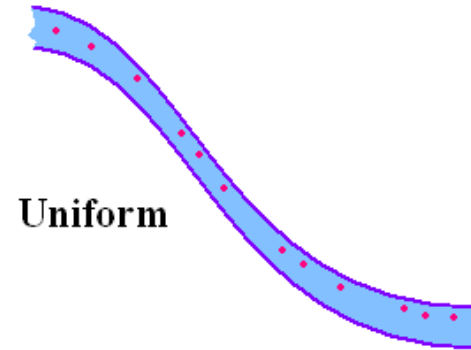
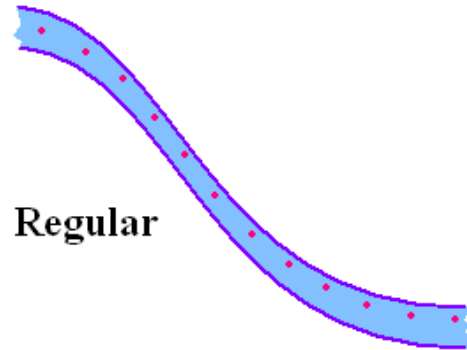
Clustered



Traverse

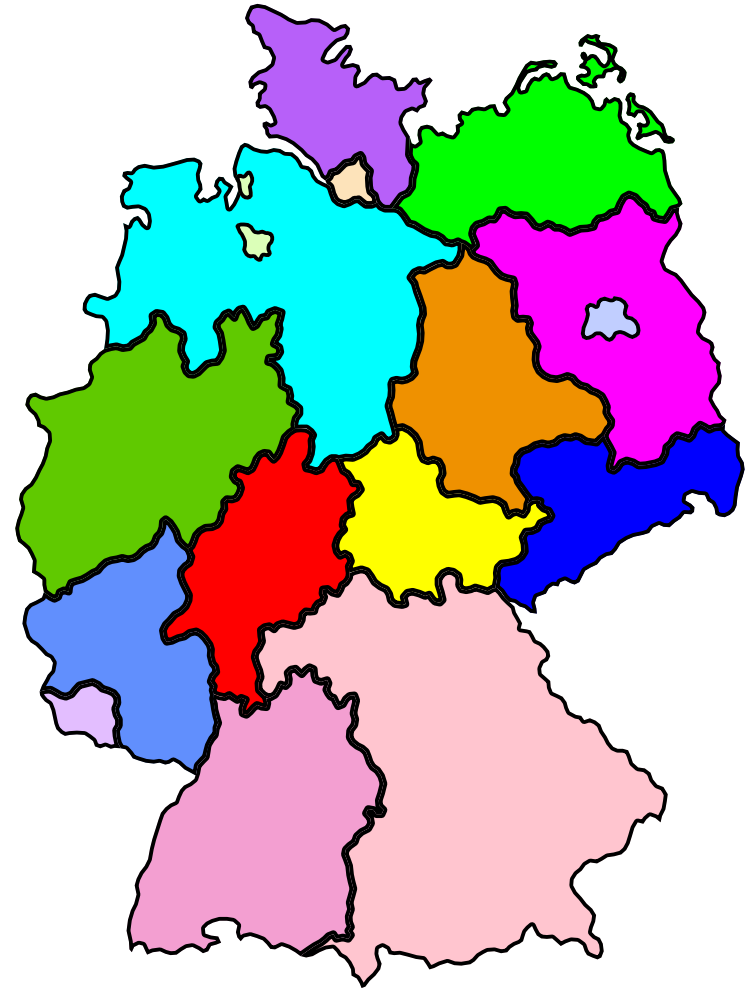
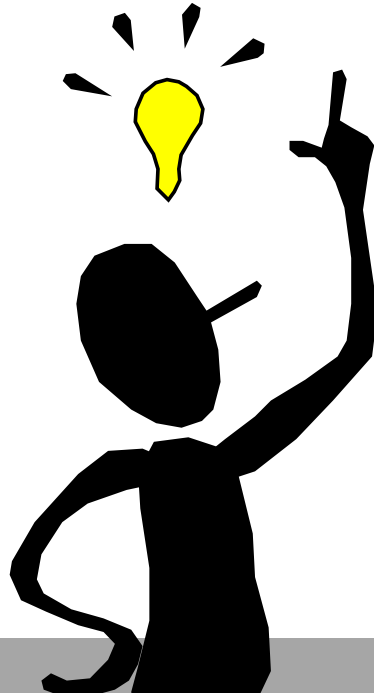


Sample Techniques- e.g. Hydraulics & Traffic Engineering



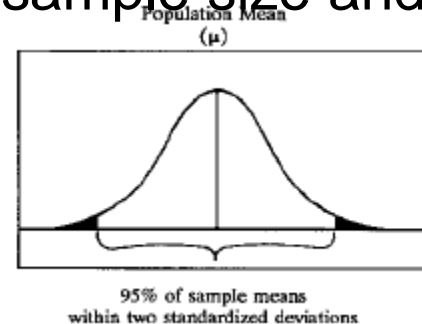
Sample Size Determination

- Purpose of the study
- Population size
- Risk of selecting a „bad“ sample
- Allowable sampling error



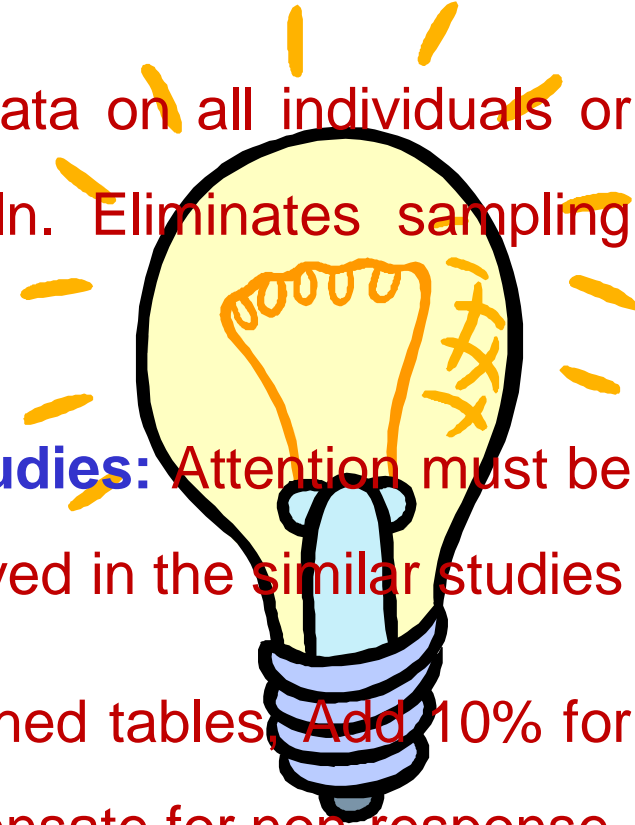
Sample Size Criteria

- **Level of Precision** (sometimes called sampling error): Range in which true value is estimated to be. e.g. Precision rate (5%)
- **Level of Confidence or risk:** When repeatedly sampled, the average exhibit true population value, Normally distributed, e.g. 95 % confidence level, risk is reduced @ 99%.
- **Degree of variability** refers to distribution of attributes in population. More heterogeneous popln.= large sample size and vice versa, 0.5 represent the largest degree. Sample size will be larger if true variability is considered....



Strategies for determining Sample Size

- **Census for small populations:** All data on all individuals or processes in the experiment or popln. Eliminates sampling error, fixed cost (cost optimum),
- **Imitating a sample size of similar studies:** Attention must be paid to determine the procedures involved in the similar studies
- **Using published tables:** Rely on published tables, Add 10% for unable to contact cases, 30% to compensate for non-response.
- **Apply formulae to calculate sample size:** Different combinations of levels of precision, confidence and variability can be used to obtain require sample sizes.



Formulae: Sample Size

Where

$$n = \frac{N}{1 + N(e)^2}$$

$$n^0 = \frac{Z^2 pq}{(e)^2}$$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

- n = sample size
- N = Population size
- N^0 = Adjusted sample size
- Z^2 = abscissa of normal curve, found in stats. table
- e = desired precision level
- p = estimated proportional of attribute in the population
- $q = 1 - p$

Formulae: Sample Size Examples

Where

$$n^0 = \frac{Z^2 pq}{(e)^2}$$

- n = sample size
- N = Population size
- N^0 = Adjusted sample size
- Z^2 = abscissa of normal curve, found in stats. table
- $Z=1.96$ at 95% confidence level
- e = desired precision level (here 5% precision)
- p = estimated proportional of attribute in the population (assume $p=.5$ maximum variability)
- $q = 1-p$

$$n_0 = \frac{Z^2 pq}{e^2} = \frac{(1.96)^2 (.5)(.5)}{(.05)^2} = 385 \text{ farmers}$$

Formulae: Sample Size Examples

Where

If the population is small the sample size can be reduced slightly !!!

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} = \frac{385}{1 + \frac{(385 - 1)}{2000}} = 323 \text{ farmers}$$

- n = sample size
- N = Population size
- N^0 = Adjusted sample size
- Z^2 = abscissa of normal curve, found in stats. table
- e = desired precision level
- p = estimated proportional of attribute in the population
- $q = 1 - p$
- Works only when the population is too small in this case, 2000

Formulae: Sample Size Examples

Where

$$n = \frac{N}{1 + N(e)^2}$$

- n = sample size
- N = Population size
(Assume 2000 people)
- e = desired precision level, it can be assumed to be at 3 %, 5%, 7% and 10%

$$n = \frac{N}{1 + N(e)^2} = \frac{2000}{1 + 2000(.05)^2} = 333 \text{ farmers}$$

Table 1. Sample size for $\pm 3\%$, $\pm 5\%$, $\pm 7\%$ and $\pm 10\%$ Precision Levels Where Confidence Level is 95% and $P=.5$.

Sample Size Tables, examples

| Size of Population | Sample Size (n) for Precision (e) of: | | | |
|--------------------|---------------------------------------|-----------|-----------|------------|
| | $\pm 3\%$ | $\pm 5\%$ | $\pm 7\%$ | $\pm 10\%$ |
| 500 | a | 222 | 145 | 83 |
| 600 | a | 240 | 152 | 86 |
| 700 | a | 255 | 158 | 88 |
| 800 | a | 267 | 163 | 89 |
| 900 | a | 277 | 166 | 90 |
| 1,000 | a | 286 | 169 | 91 |
| 2,000 | 714 | 333 | 185 | 95 |
| 3,000 | 811 | 353 | 191 | 97 |
| 4,000 | 870 | 364 | 194 | 98 |
| 5,000 | 909 | 370 | 196 | 98 |
| 6,000 | 938 | 375 | 197 | 98 |
| 7,000 | 959 | 378 | 198 | 99 |
| 8,000 | 976 | 381 | 199 | 99 |
| 9,000 | 989 | 383 | 200 | 99 |
| 10,000 | 1,000 | 385 | 200 | 99 |
| 15,000 | 1,034 | 390 | 201 | 99 |
| 20,000 | 1,053 | 392 | 204 | 100 |
| 25,000 | 1,064 | 394 | 204 | 100 |
| 50,000 | 1,087 | 397 | 204 | 100 |
| 100,000 | 1,099 | 398 | 204 | 100 |

| | | | | |
|----------|-------|-----|-----|-----|
| >100,000 | 1,111 | 400 | 204 | 100 |
|----------|-------|-----|-----|-----|

a = Assumption of normal population is poor (Yamane, 1967).
The entire population should be sampled.

Table 2. Sample size for $\pm 5\%$, $\pm 7\%$ and $\pm 10\%$ Precision Levels Where Confidence Level is 95% and $P=.5$.

| Size of Population | Sample Size (n) for Precision (e) of: | | |
|--------------------|---------------------------------------|-----------|------------|
| | $\pm 5\%$ | $\pm 7\%$ | $\pm 10\%$ |
| 100 | 81 | 67 | 51 |
| 125 | 98 | 78 | 56 |
| 150 | 110 | 86 | 61 |
| 175 | 122 | 94 | 64 |
| 200 | 134 | 101 | 67 |
| 225 | 144 | 107 | 70 |
| 250 | 154 | 112 | 72 |
| 275 | 163 | 117 | 74 |
| 300 | 172 | 121 | 76 |
| 325 | 180 | 125 | 77 |
| 350 | 187 | 129 | 78 |
| 375 | 194 | 132 | 80 |
| 400 | 201 | 135 | 81 |
| 425 | 207 | 138 | 82 |
| 450 | 212 | 140 | 82 |

Minimum Sample Size

$$\sqrt{n} = \frac{(1.96)\sigma}{\textit{precision}}$$

Where

- n = sample size
- 1.96 = 95% confidence level
- σ = population standard deviation

Data Quality

Most important !!!

Garbage in «» Garbage out.

Civil Engineers in all streams must be confident about data quality before processing of data.

Selection of data analysis technique depends on quality of data.

Use of any measurement device must be accompanied by awareness of precision and accuracy.



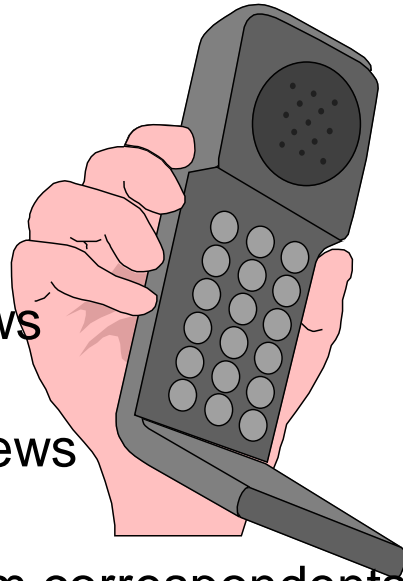
Data Quality

Precision - A measurement is precise if repeated measurements of the same entity are similar.

Accuracy – A measurement is accurate if it is close to the true value. In water resources, the true value is usually unknown, although there are standard that can be used for calibrating analytical equipment.

| | | Precision | |
|----------|------|----------------------------|----------------------------|
| | | High | Low |
| Accuracy | High | 49, 50, 50, 52, 50, 49, 51 | 55, 47, 50, 52, 44, 53, 57 |
| | Low | 54, 55, 55, 57, 55, 54, 56 | 60, 52, 55, 57, 49, 58, 62 |

Methods of collecting primary data



■ Direct Personal Interviews

■ Indirect Personal Interviews

■ Information received from correspondents

■ Mailed Questionnaire

■ Questionnaire filled by enumerators



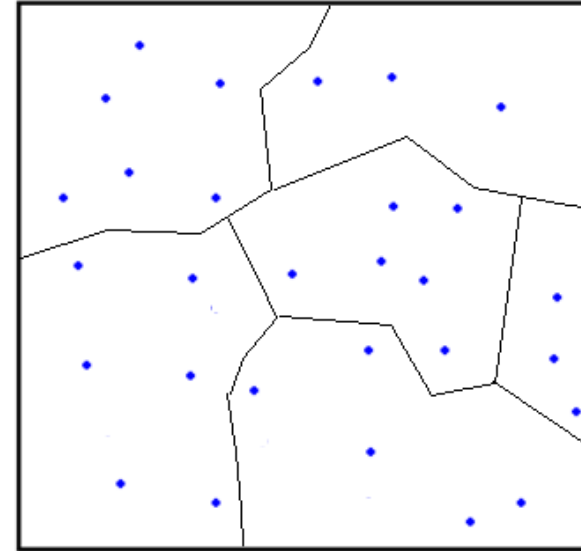
Direct Personal Interviews

Merits

- More clarified information is obtained
- Prone to changes in scheme, remove doubts
- More miscellaneous info by twisting the question
- Language and communication- good tools
- Additional info. Adds to better interpretation

Limitations

- Can't be implemented on large scale
- Time and resource intensive
- Police, authorities intervention needed, recommendations from UTM etc.
- Subjectivity is largely involved
- Needs thorough training and supervision.



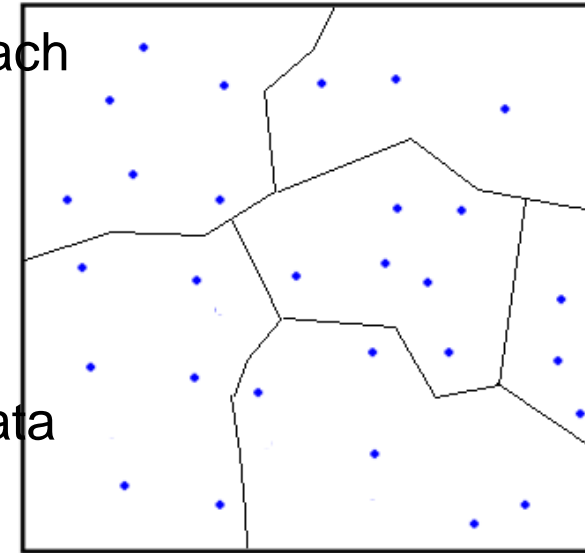
Indirect Personal Interviews/ Expert Interviews/Delphi Techniques

Precautions

- Selection of whom really possess knowledge on subject matter
- Prejudices in favor or against
- Interviewers must not be subjective in their approach
- Allowance for optimism and pessimism
- Additional info. Adds to better interpretation

Suitability

- Where direct sources do not exist, e.g accident data
- The enquiry is extensive
- Impossible to get quantitative data
- Rating scales can be applied



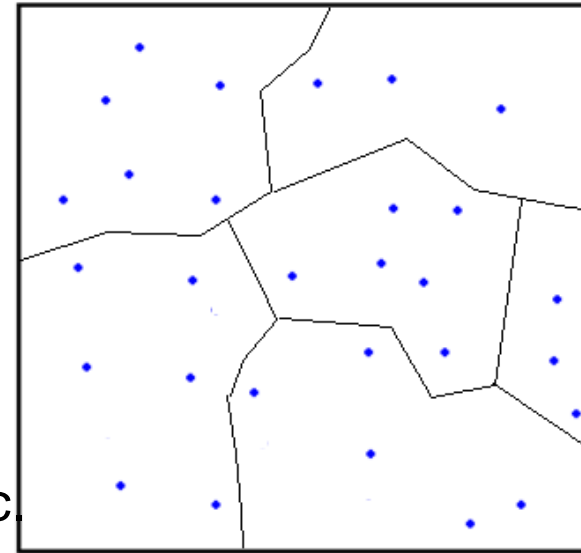
Information from correspondents

How? Investigators appoint correspondents

- Method is cheap for the data collected
- Results are approximate and rough
- Promptly and easily collected
- Personal bias may enter the data source

Suitability

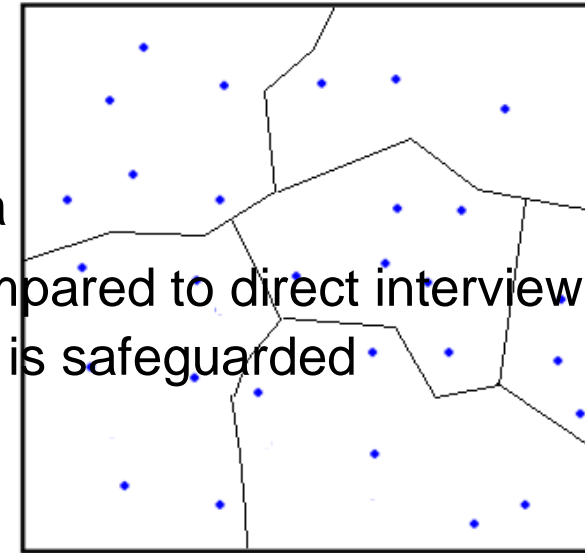
- Regular information requirements
- Wide area information is required, climate data etc.
- Approximate estimates are good enough



Mailed questionnaire- Respondent

Merits

- Extensive field survey may be surveyed
- Information available for a wide geographical area
- Cost is economical, printouts, return posts etc. compared to direct interview
- Rights are safeguarded, right to deny and answer is safeguarded
- More qualified info. Can be available



Limitations

- 10-15% people don't respond
- Results may sometimes may not be reliable enough
- Convincing required, data to be anonymised
- Don't forget to send the letter of appreciation or token of thanks!!!

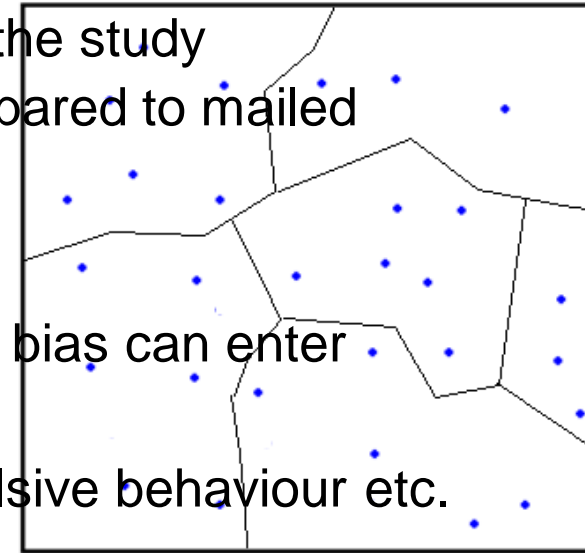
Questionnaire- Enumerators

Merits

- Cases where informants are not qualified
- Eliminates the problem of non-response cases
- Enumerators can explain the use and purpose of the study
- Data collection is more accurate and reliable compared to mailed
- More qualified info. Can be available

Limitations

- Enumerators may not fill the information correctly, bias can enter
- Proper training necessary
- State of mind of the information, moody or compulsive behaviour etc.



Suitability

- Where finance and trained enumerators are available
- Significance is attached to the accuracy of results

Questionnaire Design

Non Disguise

Non-structure,
Non disguise

Structure,
Non disguise

Non-
Structure

Structure

Non-Structure,
Disguise

Structure,
Disguise

Disguise

Developing a Good Questionnaire

- **Determining questionnaire contents:** format depends on the data source
Goals, objectives, hypotheses must be clear and well defined, use literature to discover, generate possible content areas, create a list
- **Developing items to be used,** formulate Questions or statements, Structure or non-S questions, rank orders, rating and attitude scales
- **Preparing the first draft,** trial and error methods, review and revising
Until appropriate, logical groups must be created, write instruction to fill
Format, title, margins, spacing, emphasis, key points etc.
- **Pre-testing, pilot survey,** check how well it serves the purpose, is it Representative, recognise any shortcoming and eliminate them through Corrections and further updation.
- **Revising and writing the Final Questionnaire**

Good to go!!!

Secondary Data

Once a primary data is used, it loses its significance & becomes Secondary data

■ **Published sources:**

- Official Publications of Central Governments
 - Publications of Semi-govt. Statistical Organisation
 - Publications of Research Institutions
 - Publications of commercial and financial institutions
 - Reports of committees and commissions by the govt.
-
- News papers and periodicals
 - International publications
 - Websites and other sources
 - Social Media Sites

Data Type

There are a variety of categories of data which may be seen in lets say- in Hydrology. It is important to know the data type before selecting the appropriate data analysis technique.

Ratio scale data

Ordinary measurements such as amount of rainfall, depth of groundwater level, etc. This is the best quality and most versatile data type.

Interval scale data

Interval scale data differ from ratio scale data in that the zero point is not a fundamental termination of the scale. The classical example of interval scale data is temperature measured in Centigrade.

Data Type

Ordinal Scale Data

This category is of considerable lower quality than the Ratio or Interval Scale Data. Only purpose of the scale is to place observations in relative order. Consequently, it is not valid to apply addition or subtraction, as well as division, to ordinal scale data. Non-parametric methods are used to analysis the Ordinal Scale Data

Nominal or Categorical Data

Information is sometimes prescribed in the form of names. Such as flood sometimes recorded as normal flood, severe flood, very severe flood, etc. Sometimes, drought is recoded according to it's occurrence, such as drought occur in 1974, 1981, 1993, 1990, 2008, etc.

Data Type

Directional Data

Data that is expressed in angle. Example of directional data: direction of cyclone, direction of surface runoff, etc. Directional data require special methods of analysis as the numerical values cycle around through 360 degree.

Closed Data

There are lower and upper limits of this type of data. These are data in the form of percentage, parts per million (ppm), etc. Such data require cautious treatment, especially in bivariate and multivariate methods, because variables are fundamentally interdependent.

Discrete and Continuous- Data Types

Discrete Data

When observations on a quantitative random variable can assume only a countable number of values, the variable is called a discrete random variable. It can have on only a countable number of values.

Example: Number of days with rainfall > 20 mm, number of flood in a year, etc.

Continuous Data

When observations on a quantitative random variable can assume any one of the uncountable number of values in a line interval, the variable is called a continuous random variable. A *continuous variable* can take on any value over some range.

Example: Daily rainfall, River Discharge, Groundwater Depth, etc.

Psychological Scales

Likert Scale

„I can catch a bus easily any time of the day“

Strongly
agree

Agree

Neither
Dis/agree

Disagree

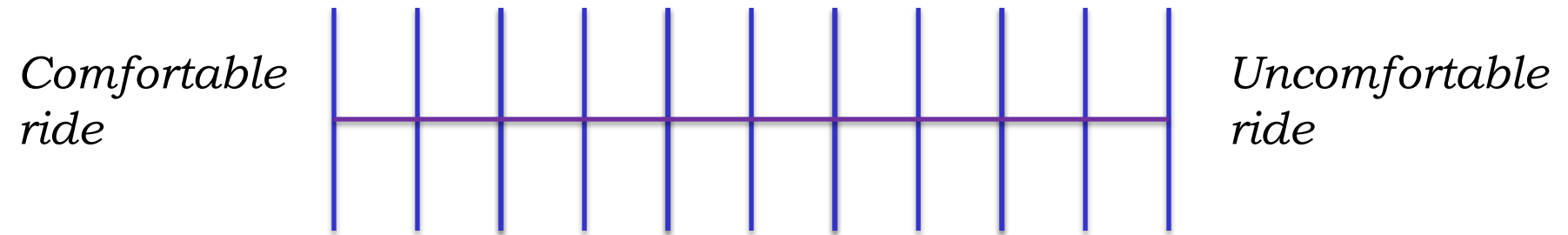
Strongly
Disagree



Psychological Scales

Semantic Differential

„Comfortability“

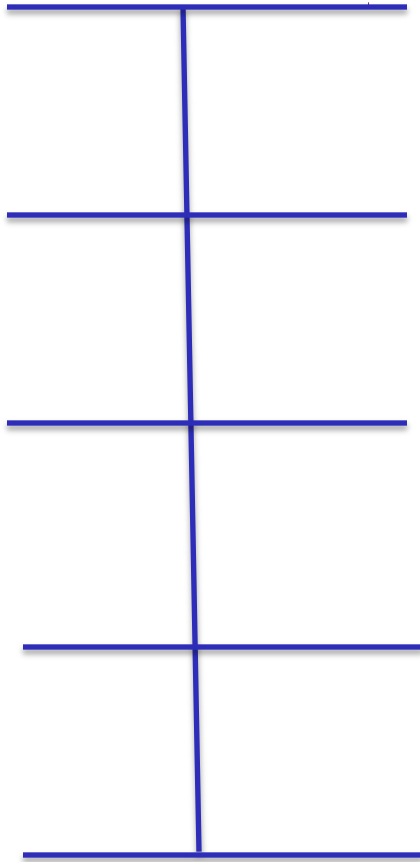


Psychological Scales

Graphical (marked)

„Atmosphere in buses and terminals“

Good



Bad



Psychological Scales

Graphical (unmarked)

„service reliability“

Low

High



Psychological Scales

Itemised

„driver friendliness-warm, friendly, personal approach, eager to help“

Extremely
poor

Very poor

Poor

Satisfactory

Good

Very good



Psychological Scales

Pairs

„allocate 100 points among the four objectives to reflect your preference“

Minimise
cost

Minimise
in-vehicle
travel time

Minimise
out-of-
vehicle
travel time

Maximise
comfort



Scaling

„process of turning scale measures into numerical values is called scaling “ e.g.

Minimise
cost

Minimise
in-vehicle
travel time

Minimise
out-of-
vehicle
travel time

Maximise
comfort



Scale of Measurement

Nominal Scale

„Labels represent various levels of a categorical variable, lowest measurement level, without any order or structure – Male/ Female“. Determination of equality, any one-to-one substitution.

In Statistics- non-parametric group, use
Mode or crosstabulation with chi-square



e.g. Gender

Male-1

Female- 2

e.g. Ethnic Group

Malay-1

Chinese-2

Indian-3

Others-4

Scale of Measurement

Ordinal Scale

„Labels represent an order that indicates either preference or ranking – Dean, HOD, Professors etc.“ Determination of greater or less, Any increasing monotonic function.

In Statistics- non-parametric group, use

Median or Mode, rank order correlation, analysis of variance (ANOVA)



e.g. Age

< 18 years old -1

18-25 years old-2

> 25 years old-3

Scale of Measurement

Interval Scale

„Numerical labels that indicate order and distance between elements is equal (equidistant)– Examination marks“.
Determination of equality of intervals, Any linear transformation.

e.g. Likert Scale

In Statistics- parametric group, use
Mean, Standard Deviation, Correlation,

Strongly Disagree-1

Disagree-2

Somewhat Agree-3

Agree-4

Strongly Agree-5

ANOVA, Factor Analysis



Scale of Measurement

Ratio Scale

„Numerical labels that indicate order and distance between elements is equal, there is an absolute zero– Age“.
Determination of equality of ratios. Any linear transformation retaining natural origin. Length, density, temperature etc.
Mostly engineering scale and not often in social research

In Statistics- parametric group, use

Mean, Standard Deviation, Correlation,
ANOVA, Factor Analysis



e.g. Your age

-Is 25 years old

Present week

-18 day

w.r.t ?.

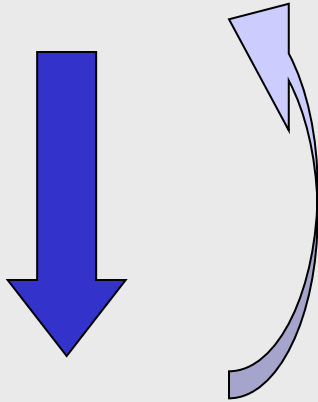
Data collection examples: Traffic Engineering

Purpose of traffic data collection

quantitative data collection – description of type, intensity and time-space distribution of traffic demand.

qualitative data collection- description of the reasons and behaviour of road user and internal relationships and dependencies.

Data collection aims



Data collection
methods

Data collection methods

Traffic counts

Determination of time and space distribution of traffic streams and traffic flow

Observations

Determination of visible behaviour of people and other aspects.

Surveys

Determination of past, actual or induced road user/traffic behaviour and the reasons thereof.

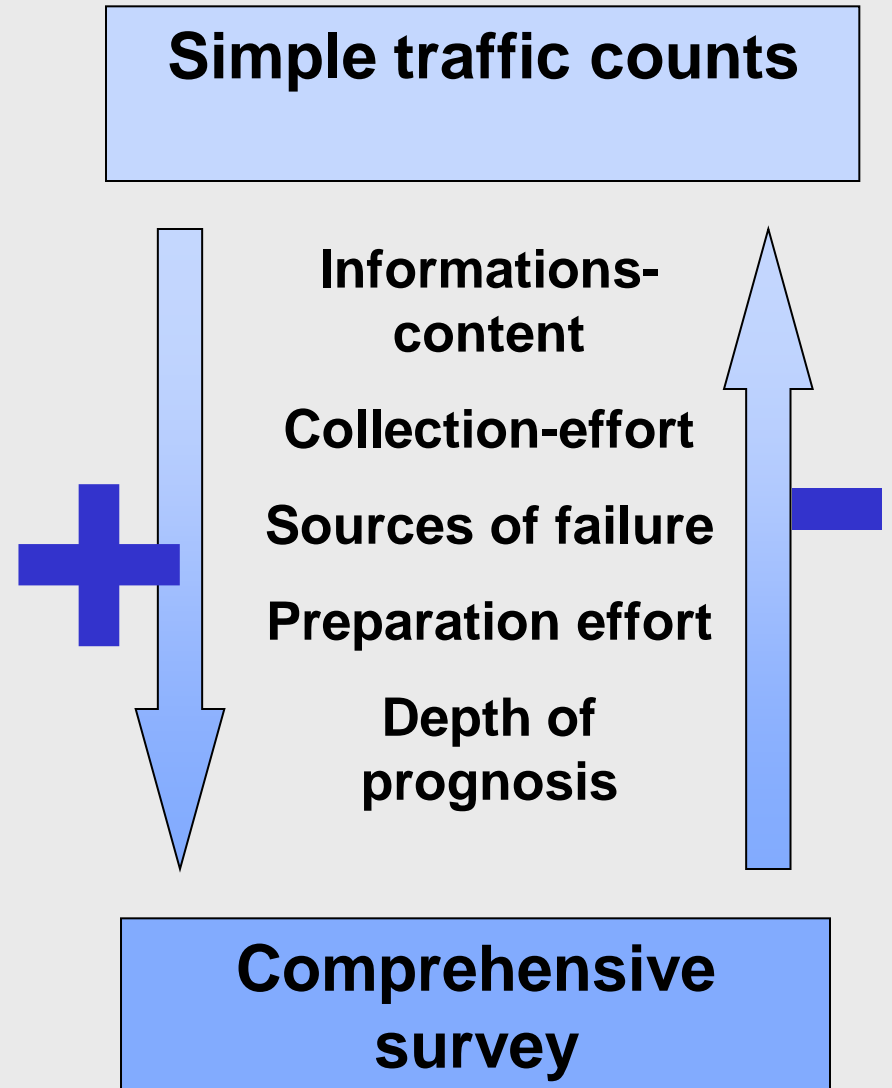
Choice of data collection

Rational criteria for the choice of data collection methods

- required data
- necessity to obtain differences in data
- possible collection efforts
- statistical precision
- preparation methods
- Demarcation of investigation area in cells and regions

- Duration and time of collection

Quelle: nach FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESSEN (FGSV): Empfehlungen für Verkehrserhebungen (EVE) Köln 1991



Work steps of a data collection (I)

Basic conception

- Need of the study
- prognosis quality
- secondary statistics
- other considerations
- Study aim
- Boundary conditions
- Hypotheses

Data collection preparatio

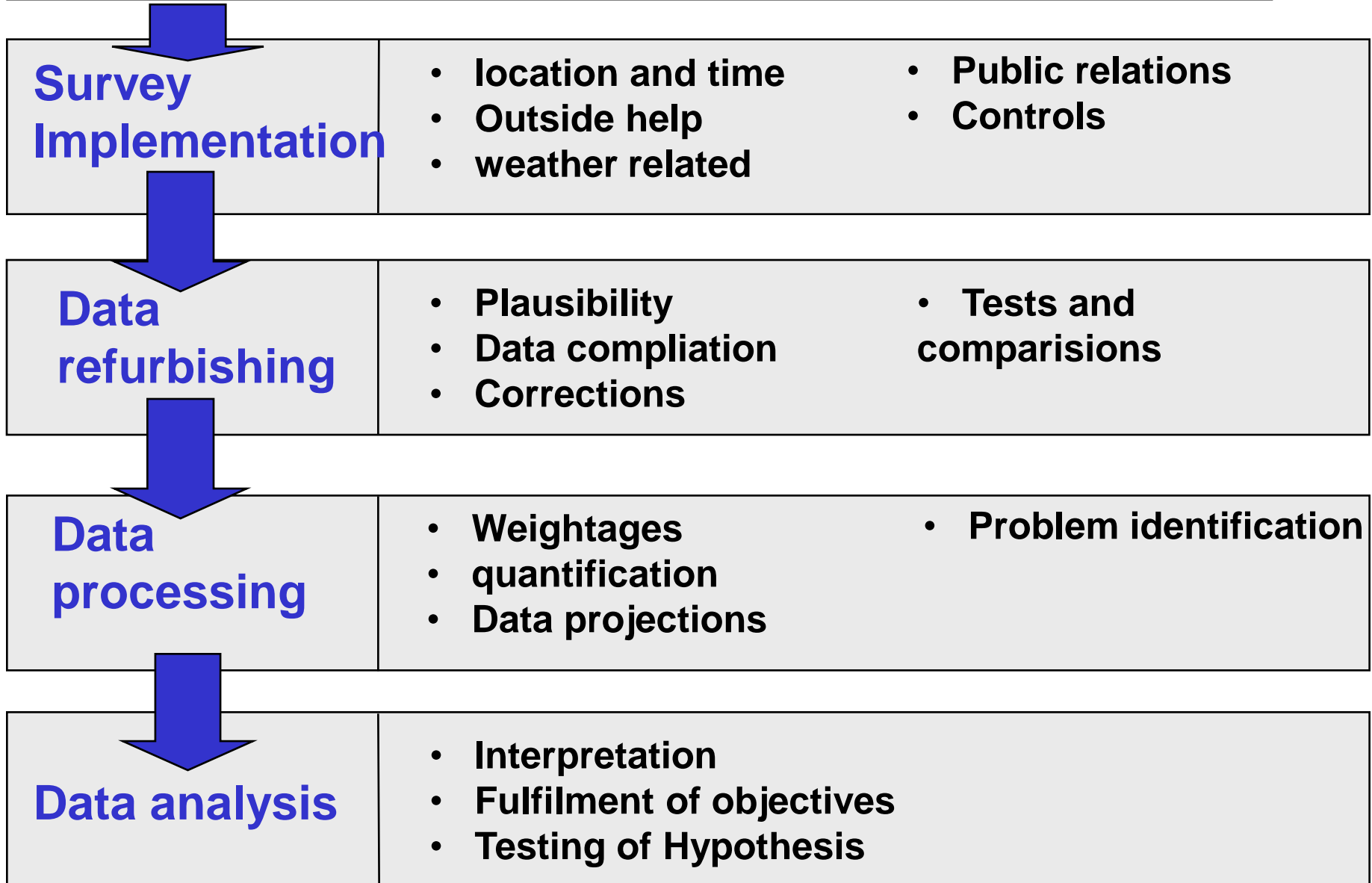
- collection methods
- Survey/
Questionnaire
- Appropriateness of data/data precision
- Data sample
- Time and duration
- Space limitation

Data collection organisaton

- Survey manual
- Public relations
- Personal training
- Materials

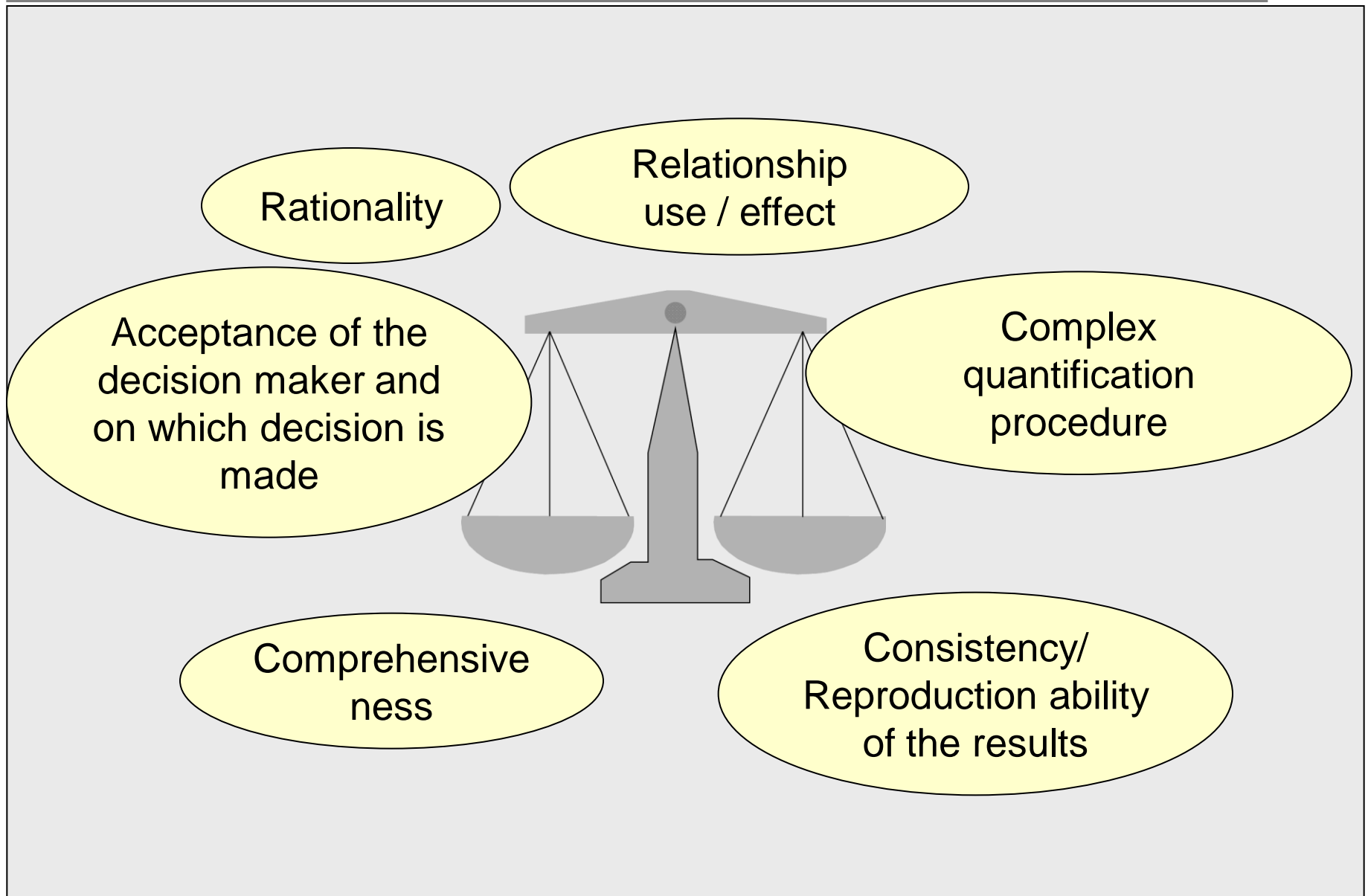
Quelle: nach FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (FGSV): Empfehlungen für Verkehrserhebungen (EVE) Köln 1991

Work steps of a data collection (II)



Quelle: nach FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (FGSV): Empfehlungen für Verkehrserhebungen (EVE) Köln 1991

Problems of Quantification



Subjectivity in data collection procedure

Work-procedure steps

Intentional or unintentional impact on results...

Planning assignment

... Abgrenzung des Planungsgegenstandes
(⇒ Ausschluß von Lösungsansätzen)
... Auswahl derer, die planen dürfen

Formulation and quantification of
concept of aims and objectives

... Ignore particular aims

Analysis of present state

... Incomprehensive and wrong representative data

Establishment of problems

... Type and form of expression:
„the road is not adequately efficient“
or
„traffic flow is very high“?

Development of
measures

... Ignored measures, highly effective
measures

Determination of
impacts

... Incomprehensive and wrong effect-determination

Quantification

... Not proper weightages

What's next ? Data Analysis



What's next ? Data Analysis in next Lecture



Thank you for your consideration

Any Questions!!!