Data Collection and Preparation

Data Quality

Data analyst must have confidence in the quality of data before commencing any analysis.

Is the data set good enough?

- Natural randomness effect ➔ may need more samples!
  (Ex: Irregularly sampled data)
- Failure in measurement procedures ➔ follow proper guidelines

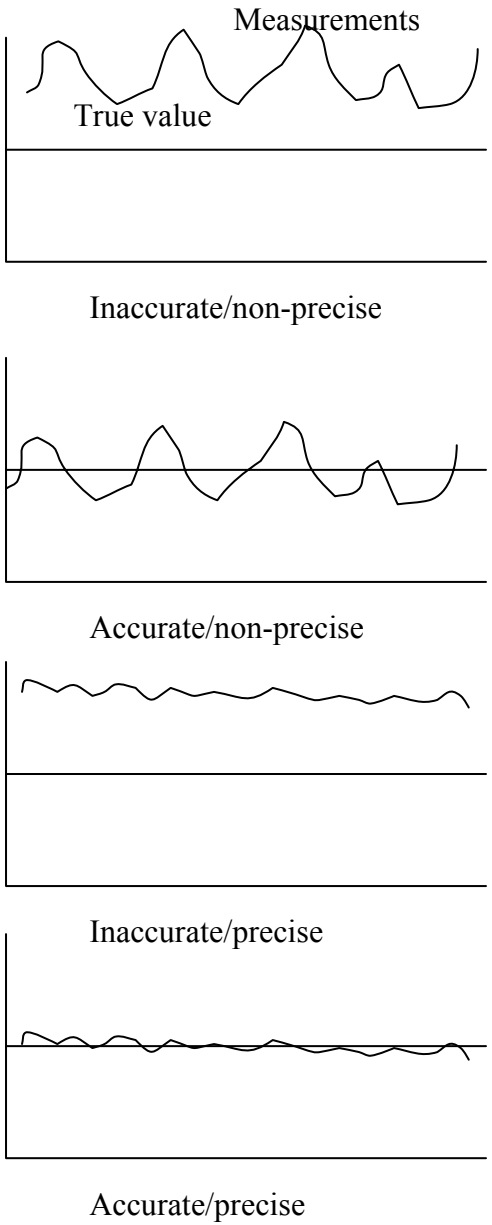
Put into consideration:

- a) Number of samples

  ➔ higher # of samples = good results (cost?)

- b) Global picture of the problem (i.e. understand geology!)

- c) Calibration of measuring devices against lab. analysis (e.g. porosity measurements vs. logging techniques)
  - ➔ precision and accuracy



Data and Variables

A **data set** is a group of variables that are related to different properties.

- Geological (Strike & dip values, … etc.)
- Physical (Porosity, permeability, … etc.)
- Chemical (mineral %, … etc.)

A **variable** is a numerical measurement or qualitative description of one particular property of a geologic sample.

Numerical measurement ➔ Porosity, grain size … etc.

Qualitative description ➔ lithology

Types of Geological Data

1- Ratio scale data

Ordinary measurements that have zero as a termination point. ➔Length, weight, porosity, permeability… etc.

Ratio here has a meaning like a ratio of 2:1 in length means doubling the length and so on.

2- Interval scale data

Ordinary measurements in which zero point is not a fundamental termination. Intervals along the data scale are constant ➔ temperature, logarithm of hydraulic conductivity.

3- Closed data

Data expressed in form of a proportion of a fixed total like percentage and part per million (ppm) ➔ mineral percentages, geochemical data, … etc.

4- Directional data

Data expressed in angles ➔ bearing from north, strike, dip, … etc.

5- Ordinal scale data

Data category that has irregular scale that depends on qualitative information ➔ Richter earthquake scale, Moh's hardness scale, … etc.

6- Categorical or nominal data

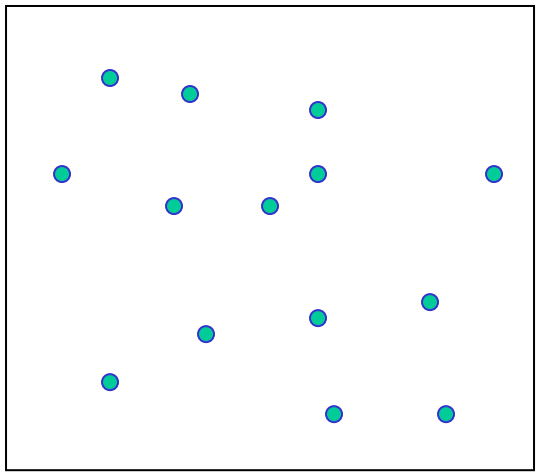
Data presented in form of names or description. It is used to classify objects. It can be expressed in BINARY format (0 and 1) ➔ in a stratigraphic unit presence of LS = 1 and absence of LS = 0

Sampling Strategies

Sample size: Could be decided on the basis of a pilot study.

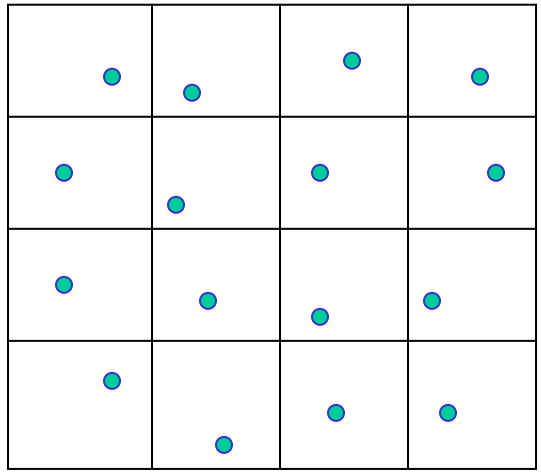
Spatial sampling schemes:

- **Random** ➔ use random number generator to decide about sampling plan

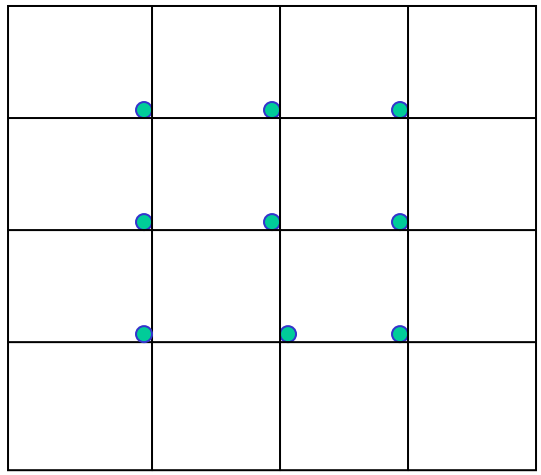


- **Uniform** ➔ use randomization sampling process within pre defined squares

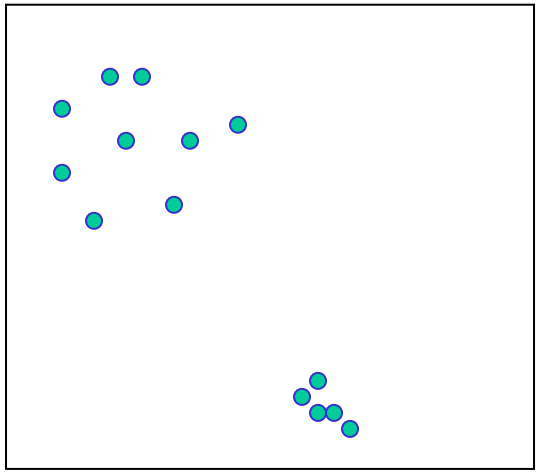
  ➔ Best to reduce bias



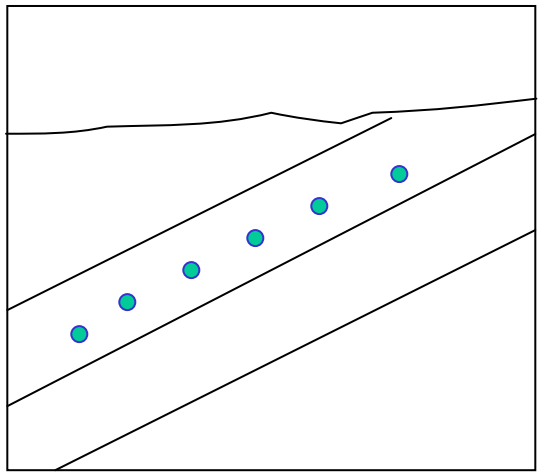
- **Regular** ➔ use regular squares or rectangles to sample



- **Clustered** ➔ not usually used in sampling unless forced by accessibility restrictions.



- **Traverse** ➔ sampling along a road-side, cliff, … etc.



Types of Analysis

Analytical questions:

1. Is the grade of ore high enough to be economical?
2. Is the sandstone sufficiently permeable enough to be a good reservoir?
3. What is the relationship between porosity & neutron log?
4. What is the shape and volume of a carbonate aquifer?
5. Is the data set good enough?

To answer these questions the data analyst has to have:

1. An estimate of the true value (e.g. mean value)
2. A reliability measure of the estimate (e.g. probabilistic models ➔ 95% sure that the ore is economical to mine)

Computational Techniques

Statistics and data analysis

Science of collecting, analyzing, and interpreting data in order to draw a valid conclusion.

1. Parametric ➔ deals with numerical data (e.g. porosity, permeability, … etc.)

Univariate Statistics: Each variable analyzed in isolation (e.g. mean, variance, … etc.)

Bivariate Statistics: Two variables analyzed together to investigate the relationship between them (e.g. correlation coefficient, covariance, … etc.)

Multivariate Statistics: More than two variables analyzed together to understand the relationship between them (e.g. multiple regression analysis, … etc.)

2. Non-parametric ➔ deals with nominal or categorical data (e.g. presence or absence of SS in a rock unit, … etc.)

Spatial analysis

The analysis of 3, 4 or more variables together where 2 or 3 of which are spatial coordinates

Examples: contour maps, GIS maps, … etc.

Flow simulators

The **imitation** of a real geological system characteristics or response by solving space / time constrained mathematical equations.

Examples: oil simulators, groundwater simulators, contaminant transport simulator … etc.

Analytical methods

The description of a geological feature / characteristic by mathematical or graphical tools.

Examples: rose diagram, ternary diagrams … etc.

Data Quality Objective (DQO) Process

Specifications needed to develop a sampling and analysis plan

- Qualitative procedures
- Quantitative procedures

What are they?

1) State the problem ➔ e.g. **Reservoir Characterization**

Objectives (why study?)

- Knowledge (literature review)
- Conceptual models

2) Identify the decision (Is it feasible to develop an oil field associated with the reservoir?)

3) Identify the required data set

- Variables: Porosity, permeability, … etc
- Geological information: lithology, structural, … etc.
- Sampling techniques: regular, irregular, … etc.

4) Define the boundaries of the decision ➔ dimensions of the field

5) Develop a decision rule

- How data will be summarized ➔ statistical results
- What action to take? (proceed or quit?)

6) Specify limits on the decision errors ➔ Cost-effect studies of different scenarios.

7) Optimize the design

- Understand the statistical/spatial variability of the available variables
- Develop and evaluate design alternatives

  Select the most cost-efficient design