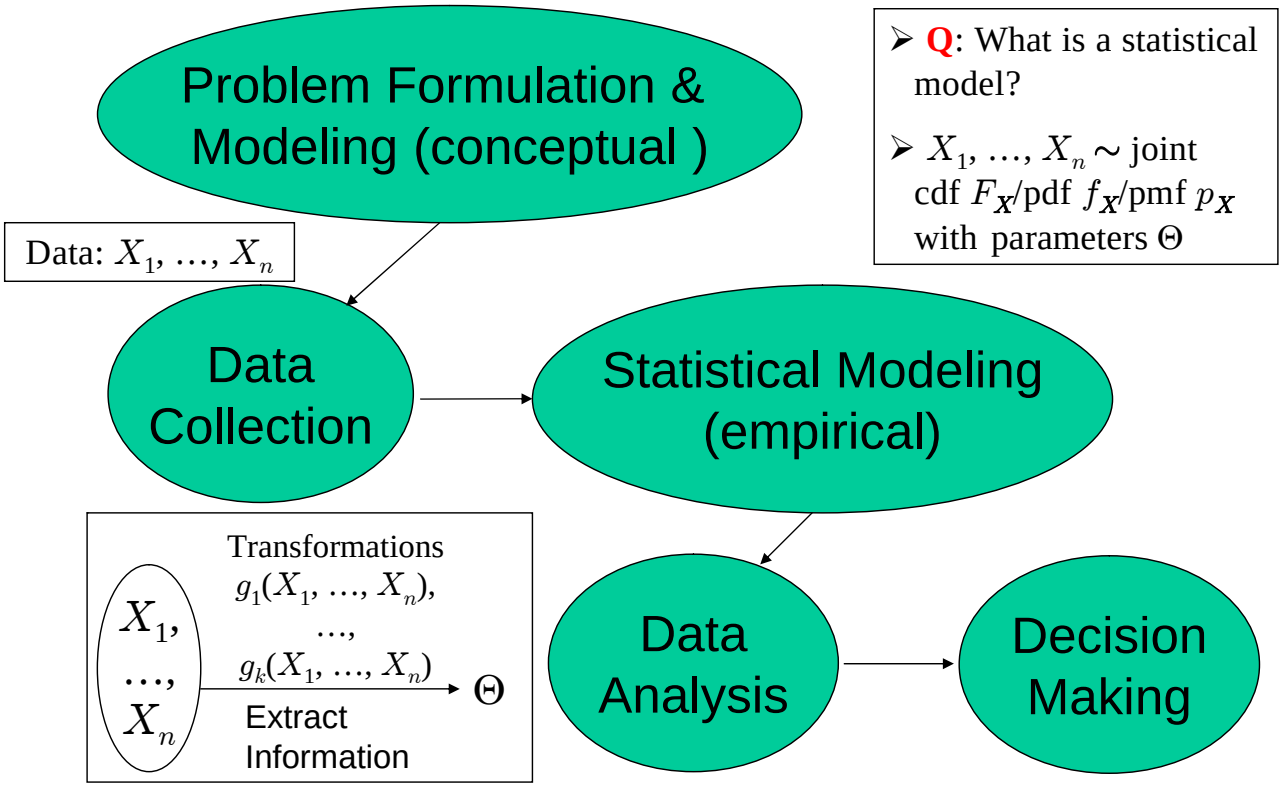
Basic Procedures of Statistics

• Statistics divides the study of data into *five* steps:



1. Problem formulation & modeling (conceptual approach)

- Communication with client
  - Communicate in a language client can understand
  - Understand client's requirement
  - Understand the physical/social/political/biological/ medical/... background to avoid the missing of important conditions that should be included in model
- Problem formulation: use statistical/probabilistic/ mathematical language to "*clearly*" define the problem and the objective of study
- modeling (conceptual approach): use the information that we possessed *prior to obtaining data* to develop a *representation of the underlying system*, also account for uncertainty in data

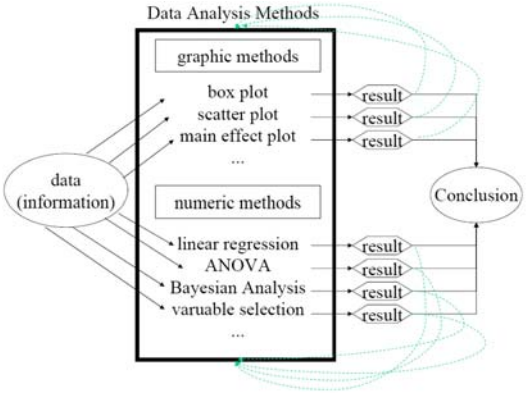
2. **Data collection:** producing *representative* data for drawing correct information

   - Some data-collection techniques
     - survey sampling
     - design of experiment
     - observational data
   - Evaluate the quality of data - is it a representative data?
   - Some preliminary investigation of data
     - observational or experimental?
     - is there non-response?
     - are there missing values?
     - qualitative or quantitative variables?
     - how are the data coded?
     - what are the units of measurement?
     - beware of data entry errors
     - initial data analysis

3. **Statistical modeling (empirical approach):** use *empirical* information contained in the data to build a model or to justify/adjust the (conceptual) model developed in **1.**, also account for uncertainty in data

   - A statistical model is a description of the joint distribution of data
   - A statistical model may contain the following components:
     - parametric component: (fixed, random) effects
     - nonparametric component
     - distribution component
   - There might exist many ways to model a data

4. **data analysis:** squeeze information out of data as much as possible
   - ➢ graphical methods
   - ➢ numerical methods
     - (point, interval) estimation
     - hypothesis testing, confidence interval
   - ➢ Data analysis is inherently interactive



5. **Inference/decision making:** drawing conclusions & answering questions based on results obtained in the step **4**.

Some Notes in Data analysis and Decision making

- if possible, most available analysis methods should be performed.
- assumptions and analysis results between different methods/models could be (slightly) different
- important information usually consistently appear in the results of every methods
- report all your findings - conclusions should be summarized based on consistent results one-by-one
- quantitative (量性) and qualitative (質性) conclusions
- evaluate the level of evidence
- the final presentation and report to the client should use the language that the client can understand to answer client's questions