Plan for Data Processing and Analysis

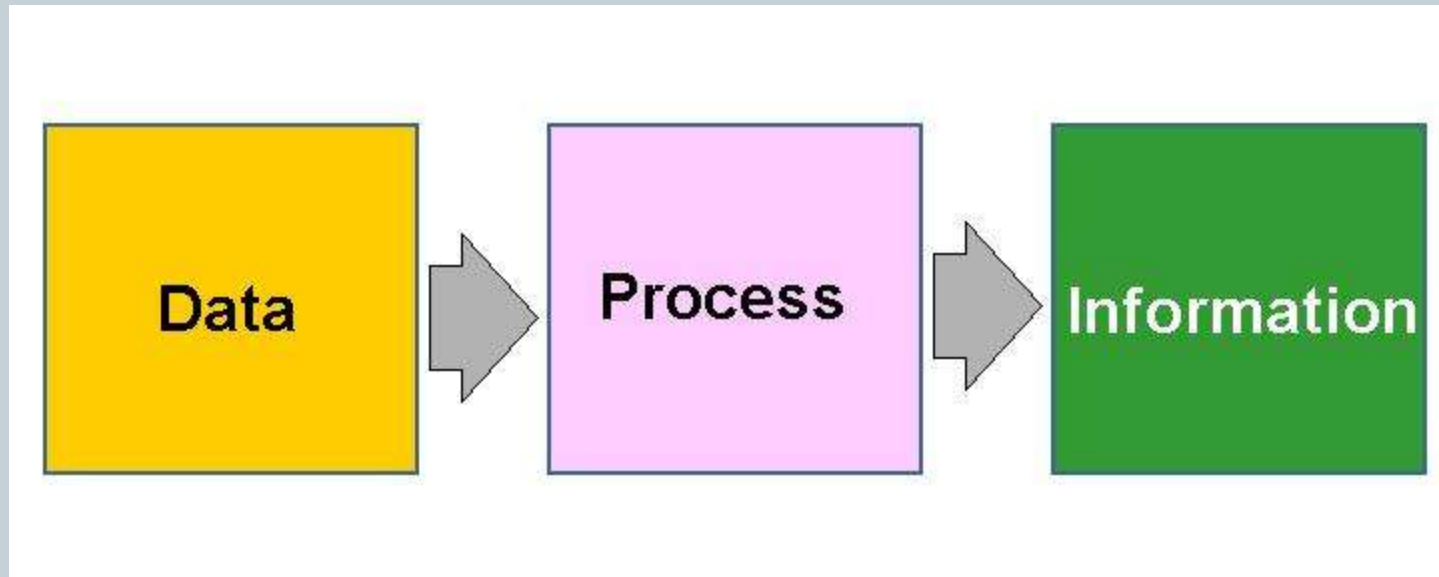## KNGMANY CHALEUNVONG

Why plan for data processing & analysis?

To ensure that:

- All information needed has really been collected, and in a standardized way.

- We have not collected unnecessary data which will never be analyzed.

What Is Data Processing?

Process of data processing

Data from your study

- Data from questionnaires, case record from (CRF), patients' records, patients' charts, etc.
- To be keyed in database software or Program data entry: Excel Sheet, FileMaker Pro, Microsoft Access and Epidata.

Data processing & analysis

- Data quality Audit and control

- Coding

- Data order

- Data processing

- Data analysis

- Training data collectors
- Validity and data record
- Consistency check

Coding

- We can run statistical models.
- Our computer programs will understand the variables.
- Accountability – we can run models "blind," or without knowing what variables stand for, in order to reduce programming / author bias.
- Be Consistent in your coding.
- Know what you are coding!
- When in doubt, have someone coding a sample of your data, and see the level of consistency.
- Keep track of what you do! Use a codebook!

Data processing

- Data entry
- Double data entry
- Validation
- Exploratory data analysis
- Cross tabulation
- Transformation
- Transfer data

EpiData Entry

Use EpiData when you have collected data on paper and you want to do statistical analyses or tabulation of data. Your data could be collected by questionnaires or any other kind of paper based information. EpiData Entry is **not made for analysis.**

http://www.epidata.dk/downloads/epitour.pdf

EpiData Entry...

- **Controlled data entry:**

  EpiData will only allow the user to enter data which meets certain criteria.

- **Double Entry of Data**

  Enter data separately in two different files and compare them afterwards.

- **Data validation**

  Compare the two files and then check the discordances against the original paper copy and correct the errors.

- **Cleaning data and verification data**

EpiData Entry...

- Have three files:
  - **.qes**
  - **.rec**
  - **.chk**

- From Epidata: transfer to statistic programs for analysis (SPSS, STATA, Epi Info, etc.)

- Make your own codes for each variable and categorical data and have separate sheet for explanations of each code (very important!!!)
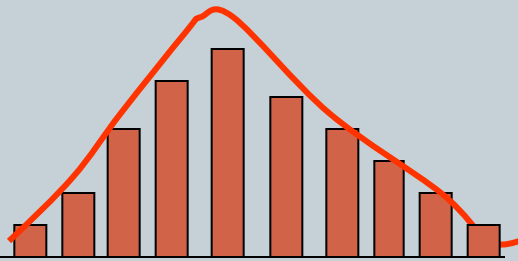
Data analysis

- Please consult a statistician or statistics textbook if you are not confident of your own data analysis capacity

- Any statistics programs can be used: SPSS, STATA, Epi Info, R, etc.

- Digest your study objectives before doing analysis
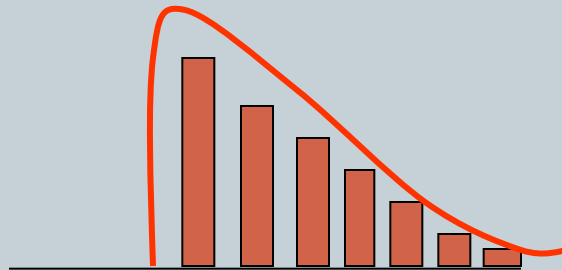
- Use simple statistics first!
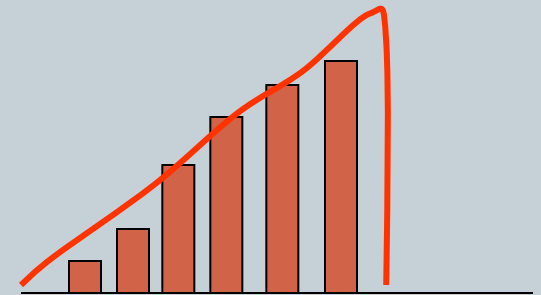
Data analysis...

- Before applying any tests: check if the data are normally distributed or not (using histogram chart)

**Normal distribution
(Bell shape)**

**Not Normal distribution
(Skewed shape)**

$$CV = \frac{SD}{\overline{X}} \times 100 < 25\%$$ Normal distribution

Data analysis...

- Descriptive statistics:

  + Proportion/frequency/percentage

  + Mean (SD) or Mean (95%CI)

  + Median (range)

Data analysis...

- Comparisons or Inferential statistics:

  + Z-test (compare proportion one or two group)

  + Chi-square (association of categorical outcome variable)

  + Student t-test (compare mean one or two group)

  + Mann Whitney-U test (non parametric test)

  + ANOVA (compare mean more than two group)

  + ANCOVA (repeated measurement)

Data analysis...

+ Correlation (two continuous variables)

+ Regression (continuous variable)

+ Logistic regression (categorical variable)

+ Log Linear Model (continuous variable)

+ Poisson Regression (Count variable)

+ GEE (Generalized Estimating Equation)

+ GLM (Generalized Linear Models)

+ Survival (COX hazard model)

Data analysis…

- Normal distributed data:

  + Mean (SD) or Mean (95%CI)

  + Comparison:

    - Paired or unpaired t-test

    - Chi-square

    - ANOVA

    - MANAVA

    - ANCOVA

    -Correlation

    - Regression

Data analysis…

- Not normal distributed data:

  + Median (range)

  + Comparison:

    - Wilcoxon signed rank test

    - Wilcoxon Matched Pairs Signed
      Rank Test

    - Wilcoxon Rank Sum test or Mann Whitney

    - Fisher's exact test

      - Kruskal Wallis test

    - Spearman's and Kendall's correlation

Technique

- Outline of research writing and plan for data analysis
- Coding manual
- Data collection manual

Research Outline

1. Title page
2. Summary)
3. INTRODUCTION
3.1 Background
 3.2 Research problem
 3.3 Objectives
 3.3 Literature review
4. METHODOLOGY)
 4.1 Study design
 4.2 Population and Sample
 4.3 Sample size
 4.4 Variable

4.5 Data processing and analysis
4.6 Ethic
4.7 Pre-test
5. PROJECT MANAGEMBNT
6. BUDGET
7. REFERENCES
8. APPENDICES
 8.1 Questionnaire form
 8.2 Plan for data analysis

Dummy table for research question

Factor	n	% disease	Crude Odds Ratios	Adjusted Odds Ratios	p-value

Smoke
    Yes
    No
    Total					

Alcohol use
    Yes
    No
    Total					

Age
    < 20 yrs
    20- 29 yrs
    >= 30 yrs
    Total					

Over weight
    Yes
    No
    Total					