

Implementation Science and Evaluation #15:

DATA MANAGEMENT (I)

WHAT is good data management?

- ✓ Data is **tidy**
- ✓ Data can be **read by statistical software**
- ✓ Data that is accompanied by a **good data dictionary**

WHY is good data management important?

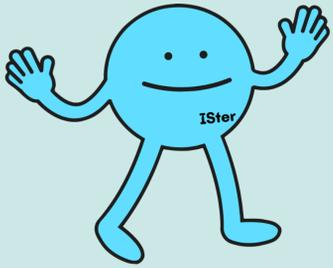
- 1 Reduces **Time and Effort** needed during data cleaning
- 2 Ensures **Usability** of datasets, so that data is readily:



- Linked with other data
- Accessed by various analytical and visualisation softwares
- Shared with others

WHEN is data management done?

Data Management occurs between Data Collection and Data Cleaning



Recap

Data Analysis Process:

1. Define goals
2. Data collection
3. Data cleaning
4. Data analysis
5. Data interpretation
6. Data visualisation

Hey Eva, good news! We have the students' results from the cooking class test! We can start the data analysis!

That's great! But first, let's ensure that we accurately input the data as this can affect how the analytical software classifies it.

Oh, but how do we ensure that we input the data accurately?

Good question ISTER! In order to do this, let's take a look at how analytical softwares classify data.

HOW data is classified by software

1

First, let's understand how data is categorised. Data can be either structured or unstructured. Structured data refers to those that are classified as nominal, ordinal, interval or ratio.

Data Type

- Unstructured
 - Free Text
- Structured
 - Nominal Eg. Gender
 - Ordinal E.g Ranking
 - Interval Eg. Dates
 - Ratio Eg. Weight

2

DATA CLASSIFICATION ON SOFTWARE

- Nominal → Categorical/Factor
- Ordinal → Categorical/Factor
- Interval → Numeric
- Ratio → Numeric
- Free Text → Text/String/Character

When using a data classification software, data in each column must be of the same type.

3

EXAMPLES OF KEYING IN DATA ACCORDING TO THEIR TEST SCORES

Example A		Example B	
Student	Score	Student	Score
Dexton	32	Dexton	32
Charles	57	Charles	57
Dexton	10	Dexton	Ten
Erwin	80	Erwin	80

For example, A shows the correct way to input students' scores. The scores are all in numeric form. In contrast, B is the wrong way to input data because one score is different from the rest (i.e. not in numeric form). This will affect how the software classifies data.

▶ WHAT'S NEXT?

Wow I did not know that! Eva, what else do I need to know to prevent such errors?

Not to worry ISTER. In the next info-poster, we will take a look at the principles and common mistakes of data management. Stay tuned!

References:

Christensen, P. (2011). An Introduction to Statistical Methods and Data Analysis. *Journal of Property Investment & Finance*.

Prybutok, V. R. (1989). An introduction to statistical methods and data analysis.

Proudly brought to you by:

