

Turning Raw Data into Analysis-ready Data

The Basics of Data Entry & Management

Wan Nor Arifin

Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.

email: wnarifin@usm.my

26 June 2018

Objectives in this session:

- Common raw data, proforma and response options.
- How to turn that into an analysis-ready format.

- Continuous: Age, SBP etc.
- Count: number of admission, number of dengue cases etc.

- Nominal:
 - ▶ Common: Gender, Race.
 - ▶ Knowledge question in KAP: Yes/Know/Don't know.
 - ▶ Visual Analog Scale (continuous line): No Pain \longleftrightarrow Worst Pain Ever!
- Ordinal:
 - ▶ Common: Education level, Cancer staging, Disease severity.
 - ▶ Likert-type scale in KAP: Strongly Disagree/Disagree/Neutral/Agree/Strongly Disagree.
 - ▶ Pain Rating Scale: No Pain 1-2-3-4-5-6-7-8-9-10 Worst Pain Ever!

MAKLUMAT SOSIODEMOGRAFIK	
1. Umur:	_____ tahun
2. Jantina:	<input type="checkbox"/> Lelaki <input type="checkbox"/> Perempuan
3. Bangsa:	<input type="checkbox"/> Melayu <input type="checkbox"/> Bumiputera Sabah <input type="checkbox"/> Cina <input type="checkbox"/> Bumiputera Sarawak <input type="checkbox"/> India <input type="checkbox"/> Lain-Lain: _____
4. Taraf perkahwinan:	<input type="checkbox"/> Belum Berkahwin <input type="checkbox"/> Balu <input type="checkbox"/> Berkahwin <input type="checkbox"/> Duda <input type="checkbox"/> Janda
5. Jika berkahwin, nyatakan bilangan anak :	_____
6. Pendapatan kasar bulanan: RM	_____
7. Tahap pendidikan:	<input type="checkbox"/> Tidak bersekolah <input type="checkbox"/> Sek. Rendah (darjah 1-6) <input type="checkbox"/> Sek. Men. Rendah (Tingkatan 1-3) <input type="checkbox"/> Sek. Men. Atas (Tingkatan 4-5) <input type="checkbox"/> Tingkatan 6/Sijil/Diploma <input type="checkbox"/> Lain-Lain, nyatakan _____

Figure 1: Sociodemographic Information

BAHAGIAN C: SIKAP

Arahan: Sila nyatakan tahap persetujuan anda bagi kenyataan berikut dengan menandakan (√) pada skala yang sesuai.

Skala	1	2	3	4	5	
	<i>sangat tidak setuju</i>	<i>tidak setuju</i>	<i>tidak pasti</i>	<i>setuju</i>	<i>sangat setuju</i>	
A1	Saya akan pakai sarung tangan ketika menguruskan sampah	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A2	Saya akan pastikan tong sampah sentiasa ditutup	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A3	Saya perlu bekerjasama dengan pihak kesihatan dalam aktiviti pencegahan dan kawalan penyakit kencing tikus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A4	Saya akan pastikan keluarga saya membersihkan kawasan rumah	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A5	Saya perlu maklumkan kepada pihak kesihatan sekiranya mendapat penyakit kencing tikus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 2: Likert-type Scales

BAHAGIAN B: PENGETAHUAN

Arahan: Tandakan (√) pada pilihan jawapan anda. Jawab semua soalan.

1. Pernahkah anda mendengar tentang penyakit kencing tikus? Ya Tidak
2. Jika ya, anda mendengarnya melalui? Televisyen/Radio Majalah
 Surat Khabar Ceramah Kesihatan
 Doktor Poster/Pamflet
 Lain-Lain, nyatakan _____

Pernyataan	Betul	Salah	Tidak Tahu
K1 Penyakit kencing tikus juga dikenali sebagai penyakit leptospirosis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
K2 Penyakit kencing tikus disebabkan oleh kuman	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
K3 Penyakit kencing tikus adalah penyakit haiwan yang boleh menjangkiti manusia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
K4 Penyakit kencing tikus boleh dikesan melalui ujian darah	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 3: Mixed Scales

- Generally, one question/item, one variable.
- Except, list of presence/absence of disease – multiple options possible. Common question in medical research, but a very common mistake in data entry → unable to analyze data.
- Variable names
 - ▶ **CAPITAL** letters.
 - ▶ “_” in place of space in names, DISEASE_STATUS, FOLLOWUP_1 etc.

- Setup data entry sheet e.g. MS Excel properly for data entry.
- Variable types:
 - ▶ Numerical
 - ★ Continuous – set cell with decimal points.
 - ★ Count – set cell with no decimal point.
 - ▶ Categorical
 - ★ Decide on coding.
 - ★ Set up the codebook, i.e. what the codings/numbers stand for:
 - 1 MALE = 1, FEMALE = 2
 - 2 MILD = 1, MODERATE = 2, SEVERE = 3
 - 3 STRONGLY DISAGREE = 0, DISAGREE = 1, NEUTRAL = 2, AGREE = 3, STRONGLY DISAGREE = 4
 - 4 etc.

Samples

- 1 Registration Number: _____.
- 2 Name: _____.
- 3 Age: _____ years.
- 4 Gender: Male Female
- 5 Race: Malay Chinese Indian Thai
- 6 Number of admission: _____ times.
- 7 Chronic Kidney Disease Stage: 1 2 3A 3B 4 5

Codebook:

- 1 RN: Patient's registration number.
- 2 NAME: Patient's full name.
- 3 AGE: Age on present admission in years.
- 4 GENDER: Gender (MALE = 1, FEMALE = 2).
- 5 RACE: Race (MALAY = 1, CHINESE = 2, INDIAN = 3, THAI = 4).
- 6 ADM: Total number of admission to the hospital since 2000.
- 7 CKD: CKD Stage on present admission (1 = 1, 2 = 2, 3A = 3, 3B = 4, 4 = 5, 5 = 6).

Table 1: Data in columns.

RN	NAME	AGE	GENDER	RACE	ADM	CKD
HRPZ/020391	Ali	20	1	1	5	4
HRPZ/015561	Ganesan	45	1	3	3	5
HRPZ/017754	Cheong	42	1	2	2	3
HRPZ/015432	Botsakorn	49	2	4	7	6

- In MS Excel:
 - ▶ set up the cell to accept only certain data type e.g. number, text, date, specific decimal place, dropdown menus etc.
 - ▶ Built in Google Form when we set up the form.

Next sessions...

- Practical sessions.
- Data entry in:
 - ▶ Spreadsheet program e.g. MS Excel, LibreOffice Calc etc.
 - ▶ Google Form.

Thank you!