

Data Preparation

Freddy Rangkuti

Problems

1. Bagaimana mengumpulkan data?
2. Bagaimana membuat tabulasi data?
3. Bagaimana mengolah data?
4. Bagaimana menganalisis data?

Kesalahan yg sering terjadi

1. **Model Analisis** muncul setelah data terkumpul
2. **Data penting yang harus ada**, tidak tercatat di dalam desain penelitian

Persiapan Data

1. **Editing**, bertujuan untuk mengetahui kesalahan-2 yang terdapat dalam sampel sehingga data menjadi akurat, konsisten, lengkap dan siap dilakukan koding dan tabulasi
2. **Jenis Editing:**
 - a. editing lapangan (cek 10%)
 - b. editing menyeluruh (unit analisis)

Persiapan Data

- **Koding**, tujuannya untuk memudahkan memasukkan ke dalam komputer
- Persyaratan dalam koding
 - Kesesuaian (pengelompokan, frekuensi dsb)
 - Klasifikasi (kategorisasi)
 - Jawaban tidak mendua (harus spesifik)
- Cara membuat **buku koding**

Persiapan Data

editing
koding
card (data sheet)
tabulasi

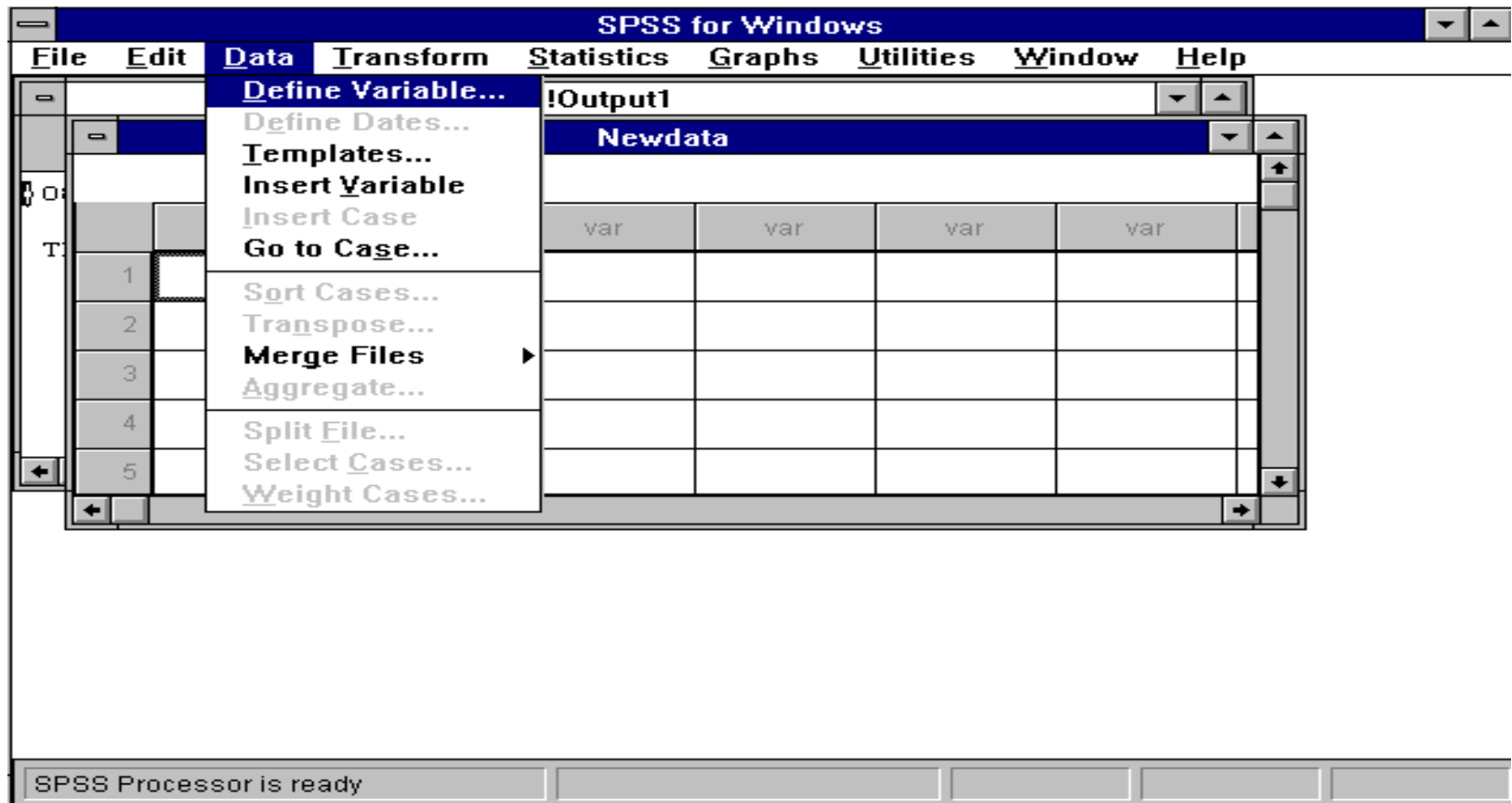
- Proses pemasukan data (data entry)
 - keyboard - instrumen - analisis
- Format pemasukan data
 - Data field
 - Record
 - Data files
 - Data bases

Coding Data for SPSS (1)

Coding is the categorization and numbering of responses obtained from questionnaires. It is used to specify how data is entered for statistical analysis for packages. This is usually necessary for **nominal data**. In SPSS coding can be facilitated using the **Data - Define Variable** menu items.

For example the variable Gender consists of two types. **Male** can be coded **1** and **Female** can be coded **2**

Coding Data Menu



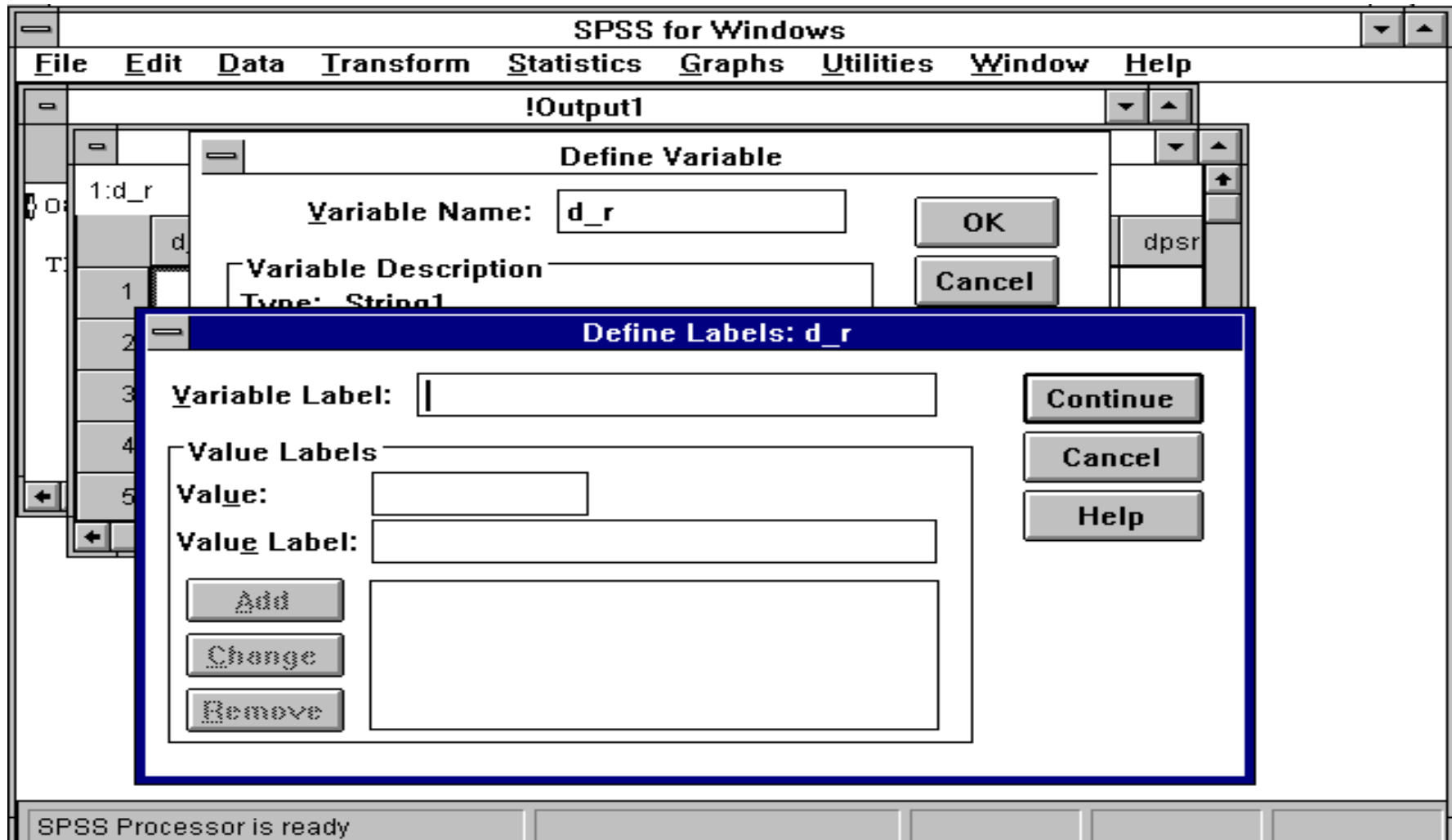
Coding Data for SPSS (2)

Coding for Open-response questions presents difficulties and usually requires **post coding** by the analyst according to a set of category criteria..

For example a question such as “**What is your favorite drink**” may produce many answers, therefore a criteria for categories is required.

Codes for **non-response** and alternative answers such as “**other**” should also be incorporated.

Coding Dialogue Box



Prinsip-Prinsip Analisis Data

- Analisis data awal
 - Tujuan untuk menyederhanakan data ke dalam bentuk yang lebih mudah dibaca dan diinterpretasi
- Tabel Frekuensi
 - Deskripsi ciri-ciri atau karakteristik variabel
 - Distribusi masing-masing variabel
 - Klasifikasi pokok untuk cross tab
- Tabel Silang (DV dan IV)

Pengujian Hipotesis

● Pengujian Hipotesis beberapa Variabel

- ➔ Uji rata-rata sampel
- ➔ Pengujian varian 2 var yang berbeda
- ➔ Pengujian perbedaan sampel berpasangan (paired samples)
- ➔ Pengujian perbedaan sampel tidak berpasangan
- ➔ Pengujian proporsi

● Pengujian Varian

Tabel Silang dgn Var Control

- Jumlah sel jangan terlalu sedikit ($fe < 5$ maks 20%)
- Variabel kontrol dipisahkan
- Variabel antara (intervening variable)

Marketing Research

- VARIABEL ANTARA
(INTERVENING VARIABLE)

- IV → Intervening Variable → DV

TINGKAT KEHADIRAN BERDASARKAN KEANGGOTAAN KLUB

Absentee Status	Anggota Klub	Anggota Klub
	Ya	Tidak
High	40%	20%
Low	60%	80%
Total	100%	100%
n	(60)	(300)

Tingkat Kehadiran berdasarkan Keanggotaan dan Var Kontrol

Absentee Status	Muda		Tua	
	Anggota	Non A	Anggota	Non A
High	48%	46%	14%	15%
Low	52	54	86	85
Total	100%	100%	100%	100%
n	(46)	(50)	(14)	(250)

Tingkat Kehadiran berdasarkan Senioritas dan Var Kontrol (keahlian)

Absentee Status	Tdk Ahli		Ahli	
	Yunior	Senior	Yunior	Senior
High	47%	16%	46%	14%
Low	53	84	54	86
Total	100%	100%	100%	100%
n	(59)	(104)	(37)	(160)

Tingkat Kehadiran berdasarkan Senioritas dan Var Kontrol (etos kerja)


Absentee Status	Menolak Etos kerja		Menerima Etos Kerja	
	Yunior	Senior	Yunior	Senior
High	96%	71%	13%	3%
Low	4	29	87	97
Total	100%	100%	100%	100%
n	(70)	(35)	(15)	(240)

Tingkat Kehadiran berdasarkan Keanggotaan Klub dan Var Kontrol (pengalaman kerja)

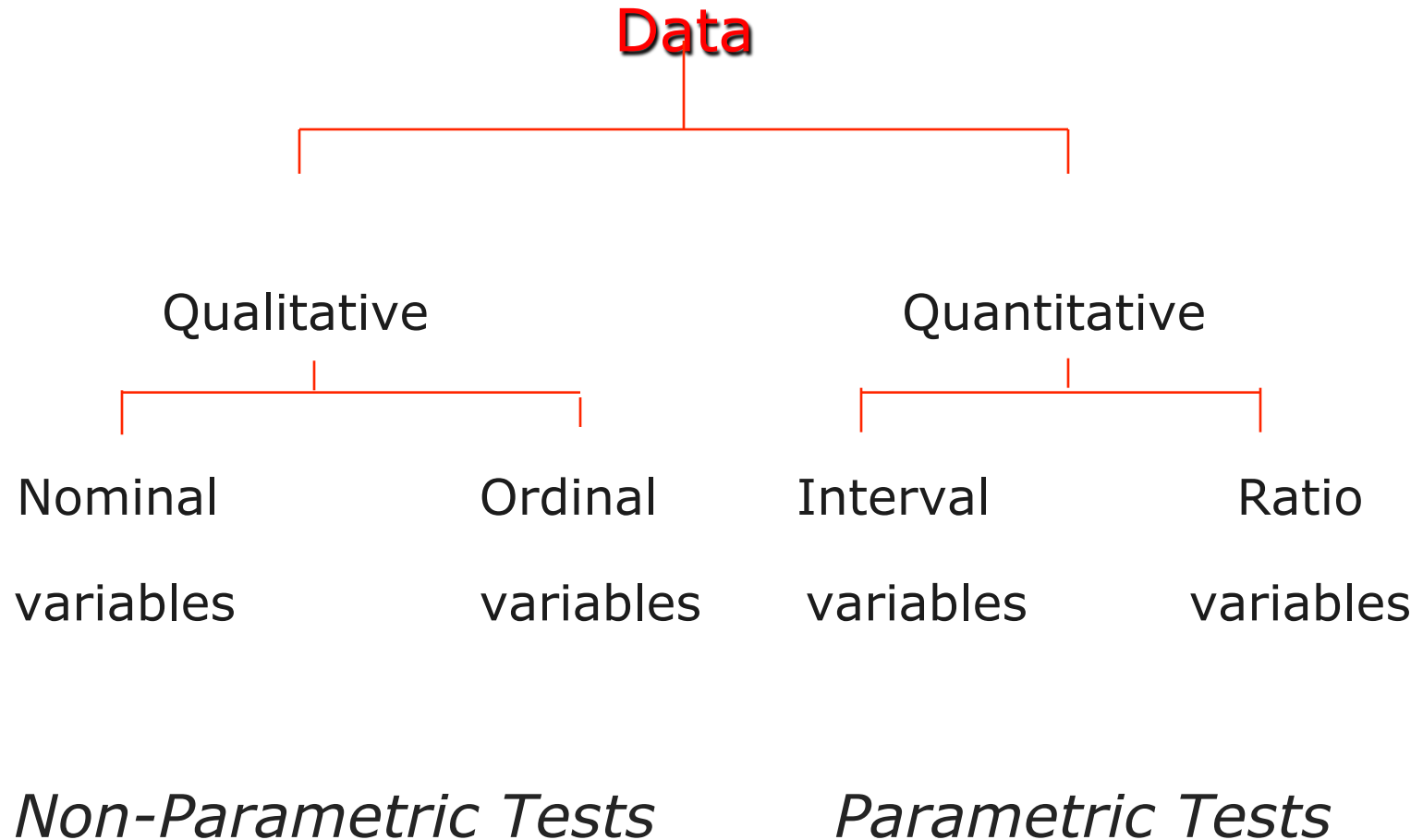
Absentee Status	< 3th		>3th	
	Anggota	Tdk A	Anggota	Tdk A
High	51%	35%	7%	17%
Low	49	65	93	83
Total	100%	100%	100%	100%
n	(45)	(150)	(15)	(249)

REAKSI TERHADAP KENAIKAN HARGA

Reaksi	Bagian A Pendapatan	
	High	Low
High	30%	30%
Low	70	70
Total n	100% (200)	100% (200)



Analysis Divisions for Data



Qualitative and Quantitative Data

Qualitative data lends itself to **non-parametric tests** and Quantitative data lends itself to **parametric tests** in statistics

The non-parametric test you will examine is the **Chi-Square test**

The parametric tests you will examine are **t-tests** and One-way Analysis of Variance (**ANOVA**)

Chi Square

- Tujuan:
 - Untuk melihat hubungan antar variabel
 - Untuk mengukur kekuatan hubungan antar variabel
- Contoh:
 - Untuk mengukur persepsi konsumen
 - Untuk mengukur tingkat kepuasan konsumen

Chi Square

- Soda Consumption - Measured in Days (but recoded to be nominal)
- Gender

Chi-Square from Cross tabs (1)

The **cross tabs** facility within SPSS has a **dialogue box** from which a range of analyses and extra functions can be selected. These include **chi-square** and inclusion of **percentages** in the table.

Interpretation of the chi-square test is quick and can be made by observation of the **significance value** on the top line of the analysis section. It is usual to compare this to 0.05 (5% significance)

Days consume bev * Gender of respondent Crosstabulation

			Gender of respondent		Total
			Male	Female	
Days consume bev	0	Count	19	19	38
		% within Days consume bev	50.0%	50.0%	100.0%
		% within Gender of respondent	31.7%	47.5%	38.0%
		% of Total	19.0%	19.0%	38.0%
	1-3	Count	17	12	29
		% within Days consume bev	58.6%	41.4%	100.0%
		% within Gender of respondent	28.3%	30.0%	29.0%
		% of Total	17.0%	12.0%	29.0%
	4-7	Count	24	9	33
		% within Days consume bev	72.7%	27.3%	100.0%
		% within Gender of respondent	40.0%	22.5%	33.0%
		% of Total	24.0%	9.0%	33.0%
Total	Count	60	40	100	
	% within Days consume bev	60.0%	40.0%	100.0%	
	% within Gender of respondent	100.0%	100.0%	100.0%	
	% of Total	60.0%	40.0%	100.0%	

Chi-square Interpretation

In a chi-square test the null hypothesis should state that there is no association between two variables. The alternative hypothesis should state that the two variables are associated:

H_0 - No association between Gender and day consume beverages

H_1 - Is association between Gender and day consume beverages

as the significance on the printout is below 0.05 (i.e. .00000) you should reject the null hypothesis (**at the 5% level**) and accept the alternative (If it would have been greater than 0.05 the reverse would be the case).