

Module 2: Types of Data

This module describes the types of data typically encountered in public health applications.

Recognizing and understanding the different data types is an important component of proper data use and interpretation.

Data and Variables

Data are often discussed in terms of variables, where a **variable** is:

Any characteristic that *varies* from one member of a population to another.

A simple example is height in centimeters, which varies from person to person.

Types of Variables

There are two basic types of variables: *numerical* and *categorical* variables.

Numerical Variables: variables to which a number is assigned as a quantitative value.

Categorical Variables: variables defined by the classes or categories into which an individual member falls.

Types of Numerical variables

- **Discrete:** Reflects a number obtained by counting—no decimal.
- **Continuous:** Reflects a measurement; the number of decimal places depends on the precision of the measuring device.
 - **Ratio scale:** Order and distance implied. Differences *can* be compared; has a true zero. Ratios *can* be compared. Examples: Height, weight, blood pressure
 - **Interval scale:** Order and distance implied. Differences *can* be compared; no true zero. Ratios *cannot* be compared. Example: Temperature in Celsius.

Ratio Scale

SYSTOLIC

DIASTOLIC

5th PHASE

11. 1 2 0 1 0 0 First reading (R1)
12. 2 0 1 5 RZ1 (ranges from 0 - 40)
13. 1 0 0 8 5 First corrected (R1 - RZ1)
14. 2 0 0 1 2 0 Second reading (R2)

Categorical Variables

Defined by the classes or categories into which an individual member falls.

- **Nominal Scale:** Name only--Gender, hair color, ethnicity
- **Ordinal Scale:** Nominal categories with an implied order--Low, medium, high.

NOMINAL SCALE

b. Appearance of plasma:	b.
1. Clear.....	1.
2. Turbid.....	2.
9. Not done.....	9.

ORDINAL SCALE

81. Urine protein (dipstick reading):	81.
1. Negative.....	1.
2. Trace.....	2.
3. 30 mg% or +.....	3.
4. 100 mg% or ++.....	4.
5. 300 mg% or +++.....	5.
6. 1000 mg% or ++++.....	6.
<i>If urine protein is 3+ or above, be sure subject gets a 24 hour urine collection container and instruction</i>	

Likert Scale

Question: Compared to others, what is your satisfaction rating of the National Practitioner Data Bank?

1	2	3	4	5
Very Satisfied	Somewhat Satisfied	Neutral	Somewhat Dissatisfied	Very Dissatisfied

Datasets and Data Tables

Dataset: Data for a group of variables for a collection of persons.

Data Table: A dataset organized into a table, with one column for each variable and one row for each person.

Typical Data Table

OBS	AGE	BMI	FFNUM	TEMP(°F)	GENDER	EXERCISE LEVEL	QUESTION
1	26	23.2	0	61.0	0	1	1
2	30	30.2	9	65.5	1	3	2
3	32	28.9	17	59.6	1	3	4
4	37	22.4	1	68.4	1	2	3
5	33	25.5	7	64.5	0	3	5
6	29	22.3	1	70.2	0	2	2
7	32	23.0	0	67.3	0	1	1
8	33	26.3	1	72.8	0	3	1
9	32	22.2	3	71.5	0	1	4
10	33	29.1	5	63.2	1	1	4
11	26	20.8	2	69.1	0	1	3
12	34	20.9	4	73.6	0	2	3
13	31	36.3	1	66.3	0	2	5
14	31	36.4	0	66.9	1	1	5
15	27	28.6	2	70.2	1	2	2
16	36	27.5	2	68.5	1	3	3
17	35	25.6	143	67.8	1	3	4
18	31	21.2	11	70.7	1	1	2
19	36	22.7	8	69.8	0	2	1
20	33	28.1	3	67.8	0	2	1

Definitions for Variables

- AGE: Age in years
- BMI: Body mass index, $\text{weight}/\text{height}^2$ in kg/m^2
- FFNUM: The average number of times eating “fast food” in a week
- TEMP: High temperature for the day
- GENDER: 1- Female 0- Male
- EXERCISE LEVEL: 1- Low 2- Medium 3- High
- QUESTION: Compared to others, what is your satisfaction rating of the National Practitioner Data Bank?
1- Very Satisfied 2- Somewhat Satisfied 3- Neutral
4- Somewhat dissatisfied 5- Dissatisfied