Data Management

CT051-3-M

Topic 3 – Data Types

Topic & Structure of Lesson

- What is Dataset
- Types of Variables
- Six Basic Ways to Identify Variables

What is Dataset?

- Collection of data objects and their attributes

- An attribute is a property or characteristic of an object
  - Examples: eye color of a person, temperature, etc.
  - Attribute is also known as **variable, field, characteristic, parameter, factor, or feature**
- A collection of attributes describe an object
  - Object is also known as **record, point, case, sample, entity, observation, or instance**

Attributes

Objects

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Types of Attributes

- There are different types of attributes
  - Nominal
    - Examples: ID numbers, eye color, zip codes
  - Ordinal
    - Examples: rankings (e.g., taste of potato chips on a scale from 1-10), grades, height in {tall, medium, short}
  - Interval
    - Examples: calendar dates, temperatures in Celsius or Fahrenheit.
  - Ratio
    - Examples: temperature in Kelvin, length, time, counts

Exercise

Primary Scales of Measurement

Scale				
Numbers Assigned to Runners				Finish
Rank Order of Winners	 Third place	 Second place	 First place	Finish
Performance Rating on a 0 to 10 Scale	8.2	9.1	9.6	
Time to Finish, in Seconds	15.2	14.1	13.4	

© 2007 Prentice Hall

Exercise

Scale	Basic Characteristics	Common Examples	Marketing Examples
	Numbers identify & classify objects	Social Security nos., numbering of football players	Brand nos., store types
	Nos. indicate the relative positions of objects but not the magnitude of differences between them	Quality rankings, rankings of teams in a tournament	Preference rankings, market position, social class
	Differences between objects can be compared, zero point is arbitrary	Temperature (Fahrenheit) Celsius)	Attitudes, opinions, index nos.
	Zero point is fixed, ratios of scale values can be compared	Length, weight	Age, sales, income, costs

R Dh@ker, Lecturer, PCNMS

Question 1 out of 5.

1. Identify the scale of measurement for the following: military title -- Lieutenant, Captain, Major.

- ❑nominal
- ❑ordinal
- ❑interval
- ❑ratio

Correct Answer: The scale is **ordinal**. There is an inherent ordering in that a Major is higher than a Captain, which is higher than a Lieutenant.

Question 2 out of 5.

2. Identify the scale of measurement for the following categorization of clothing: hat, shirt, shoes, pants

- ❑nominal
- ❑ordinal
- ❑interval
- ❑ratio

Correct Answer: Since clothes are categorized and have no inherent order, the scale is **nominal**

Question 3 out of 5.

3. Identify the scale of measurement for the following: heat measured in degrees centigrade.

- ❑nominal
- ❑ordinal
- ❑interval
- ❑ratio

Correct Answer: The scale is **interval** because there are equal intervals between temperatures but no true zero point.

Question 4 out of 5.

4. A score on a 5-point quiz measuring knowledge of algebra is an example of a(n)

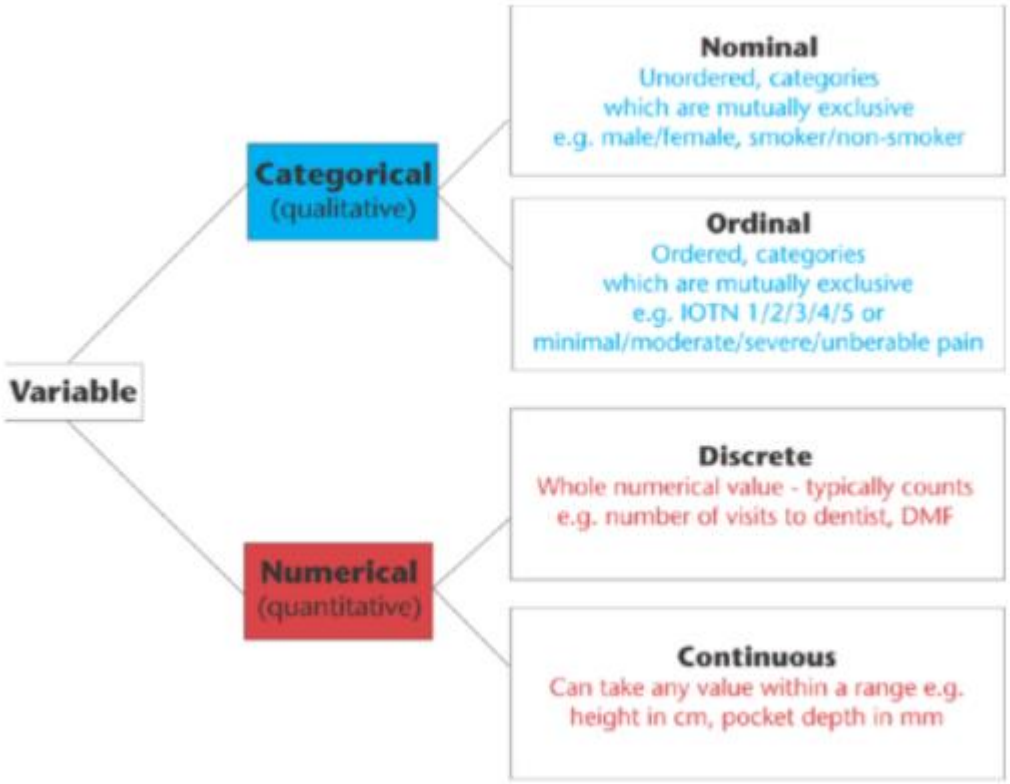
- ❑nominal
- ❑ordinal
- ❑interval
- ❑ratio

Correct Answer: It is **ordinal** because higher scores are better than lower scores. However, there is no guarantee that the difference between, say, a 2 and a 3 represents the same difference in knowledge as the difference between a 4 and a 5.

Question 5 out of 5.

5. City of birth is an example of a(n)
   - ❑nominal
   - ❑ordinal
   - ❑interval
   - ❑ratio

Correct Answer: The city that someone was born in has no inherent order, thus can only be a **nominal** scale.


Variable
Categorical
(qualitative)
Nominal
Unordered, categories
which are mutually exclusive
e.g. male/female, smoker/non-smoker
Ordinal
Ordered, categories
which are mutually exclusive
e.g. IOTN 1/2/3/4/5 or
minimal/moderate/severe/unberable pain
Numerical
(quantitative)
Discrete
Whole numerical value - typically counts
e.g. number of visits to dentist, DMF
Continuous
Can take any value within a range e.g.
height in cm, pocket depth in mm

6 Basic Ways to Identify Variables

- Independent Variables (Predictor Variables)
- Dependent Variables (Criterion Variables)
- Variables of Interest
- Confounding (Control) Variables
- Moderating Variables
- Mediating Variables

Independent & Dependent Variables

ASIA PACIFIC UNIVERSITY OF TECHNOLOGY & INNOVATION

INDEPENDENT VARIABLE

What I CHANGE

DEPENDENT VARIABLE

What I OBSERVE

Independent & Dependent Variables

Independent Variable [OR] Input Variable

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Dependent Variable [OR] Output Variable

6 Basic Ways to Identify Variables

- Variables of Interest - To find the correlation analysis between two variables (dependent, independent variables)
- Confounding (Control) variables - are variables that influence the dependent variable
- Moderating variables - influence the strength of a relationship between two other variables
- Mediating variables - explains the relationship between the two other variables

Mediator

- A mediating variable explains the relation between the independent (predictor) and the dependent (criterion) variable. It explains how or why there is a relation between two variables.
- A mediator can be a potential mechanism by which an independent variable can produce changes on a dependent variable.

Mediator

- When you fully account for the effect of the mediator, the relation between independent and dependent variables may go away. For instance, imagine that you find a positive association between note-taking and performance on an exam. This association may be explained by number of hours studying, which would be the mediating variable.

Moderator

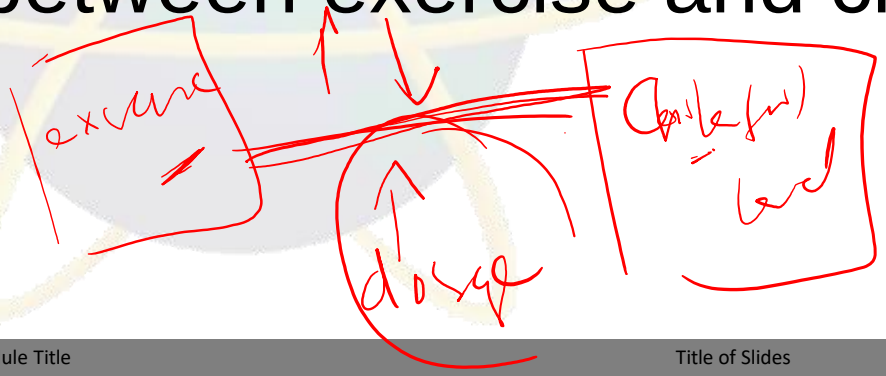
- A moderator is a variable that affects the strength of the relation between the predictor and criterion variable. Moderators specify when a relation will hold. It can be qualitative (e.g., sex, race, class…) or quantitative (e.g., drug dosage or level of reward). Moderating variable are typically an interaction term in statistical models

Moderator

- For instance, imagine researchers are evaluating the effects of a new cholesterol drug. The researchers vary the participants in minutes of daily exercise (predictor/independent variable) and measure their cholesterol levels after 30 days (criterion/dependent variable).

Moderator

- They find that at low drug doses, there is a small association between exercise and cholesterol levels, but at high drug doses, there is a huge association between exercise and cholesterol levels. Drug dosage moderates the association between exercise and cholesterol levels.



Question & Answer Session

A · P · U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

Q & A