Jumping into Statistics: Introduction to Study Design and Statistical Analysis for Medical Research Using JMP Pro Statistical Software

FALL 2022/SPRING 2023

DR. TERRIE VASILOPOULOS, DR. CYNDI GARVAN,
& DR. PENNY REYNOLDS

Meet the Instructors

TERRIE VASILOPOULOS, PHD

Research Associate Professor in Anesthesiology and Orthopaedics and Rehabilitation

CYNTHIA GARVAN, PHD

Research Professor in Anesthesiology

PENNY REYNOLDS, PHD

Research Assistant Professor in Anesthesiology and Veterinary Medicine

Course Objectives

- Review fundamentals of study design and research methodology
- Understand how to choose best statistical test for your research question
- Practice basic statistical analysis use JMP Pro Software

Course Topics

- Life Cycle of Research and Asking a Good Research Question
- Choosing the Right Study Design for Your Research
- Clinical Trial Design
- Populations, Samples, and Hypothesis Testing in Medical Research
- Introduction to Data Types
- Best Practices in Data Collection and Database Management: Getting Started with SAS JMP Pro
- Summarizing and Visualizing Data
- Statistical Methods and How to Choose Them
- Risk Assessment Methods
- Introduction to Regression and Correlation
- Time-to-Event (Survival) Analysis
- Methods for Clinical Diagnostic Testing

Introduction to Data Types

11/02/2022

Learning Objectives

Participants will be able to:

1) Explain importance of understanding data types

2) Summarize different ways to classify variables

3) Identify data type of a variable.

Why is this topic important?

The fundamental unit of statistics and statistical analyses is the **variable**.

The **variable type** (or data type) determines:

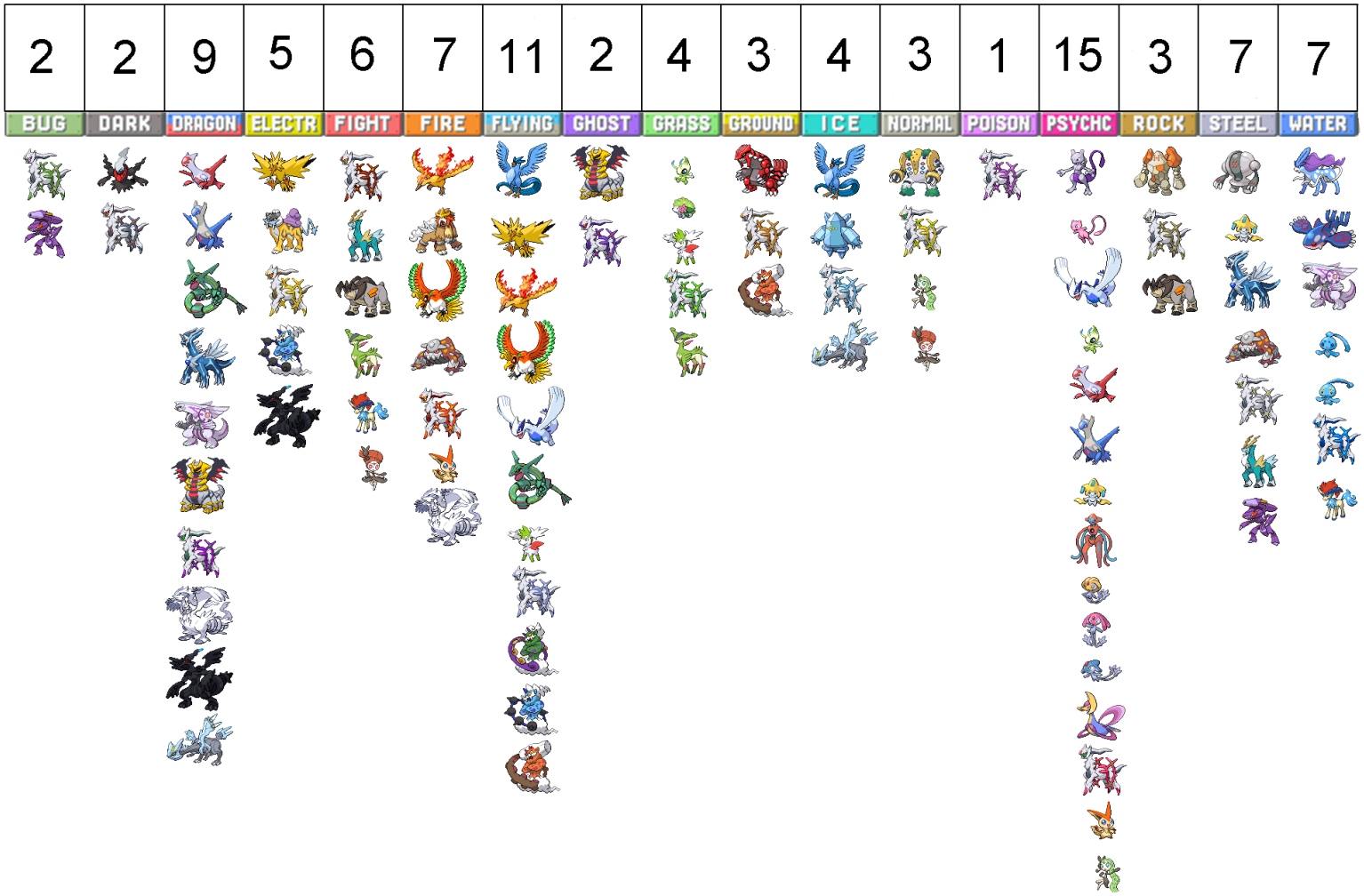
- how the variable can be *described* (summary statistics)
- how the variable can be *analyzed* (variable-appropriate analytical methods)

Why is this topic important?

You need to know data type in order to choose the correct statistical method for analysis.

LIAM AGE 5 – Really into Pokemon

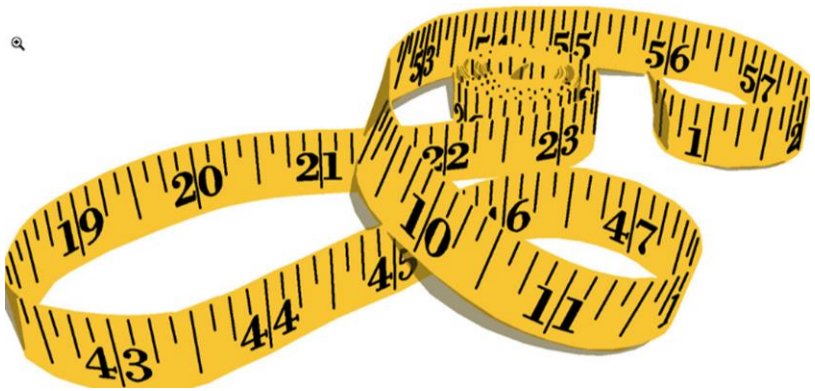

BUG
DARK
DRAGON
ELECTRIC
FAIRY
FIGHTING
FIRE
FLYING
GHOST
GRASS
GROUND
ICE
NORMAL
POISON
PSYCHIC
ROCK
STEEL
WATER


2
2
9
5
6
7
11
2
4
3
4
3
1
15
3
7
7
BUG
DARK
DRAGON
ELECTR
FIGHT
FIRE
FLYING
GHOST
GRASS
GROUND
ICE
NORMAL
POISON
PSYCHC
ROCK
STEEL
WATER

1. Introduction and Warning

Introduction to Data Types: Measurement

Measurement is the assigning of numbers to characteristics of an individual entity or an event in a systematic way as a quantitative method of representing properties of that entity or event.



Introduction to Data Types

The fundamental unit of statistics and statistical analyses is the **variable**.

The **variable *type*** determines:

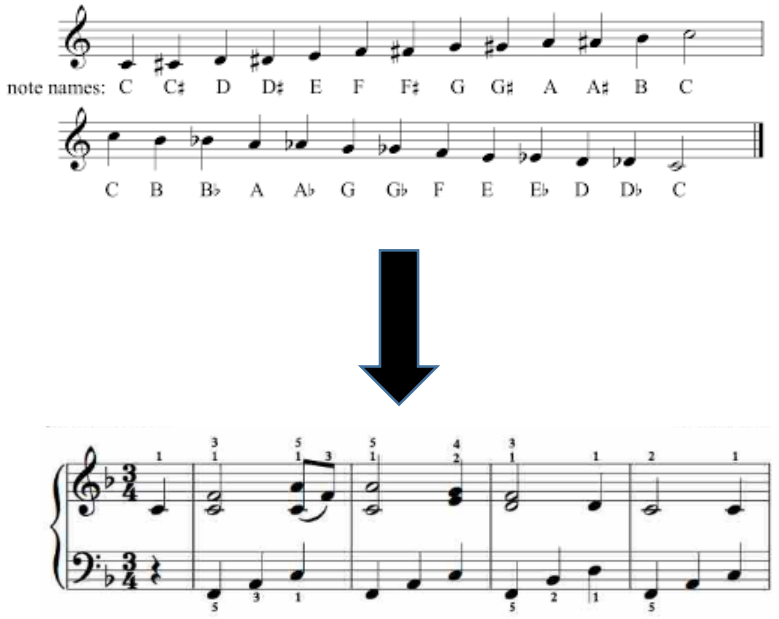
- how the variable can be *described* (summary statistics)
- how the variable can be *analyzed* (variable-appropriate analytical methods)

Introduction to Data Types

Developing a thorough understanding of variable types will improve skills in:

- troubleshooting data
- statistical analysis
- identification of analytic pitfalls
- presentation and interpretation of results
- critical assessment of published results

Understanding the data type of a variable is analogous to understanding the musical *scale* as the basis for developing musical *proficiency*.



Warning

Many investigators usually skip over information about variable data type, believing that the topic is obviously ‘too elementary.’ Nevertheless, **most of the major errors in subsequent analyses and graphics** are a direct result of applying the wrong statistical method to analyze a given variable.

Misunderstanding data type is a threat to validity.

KEY POINT

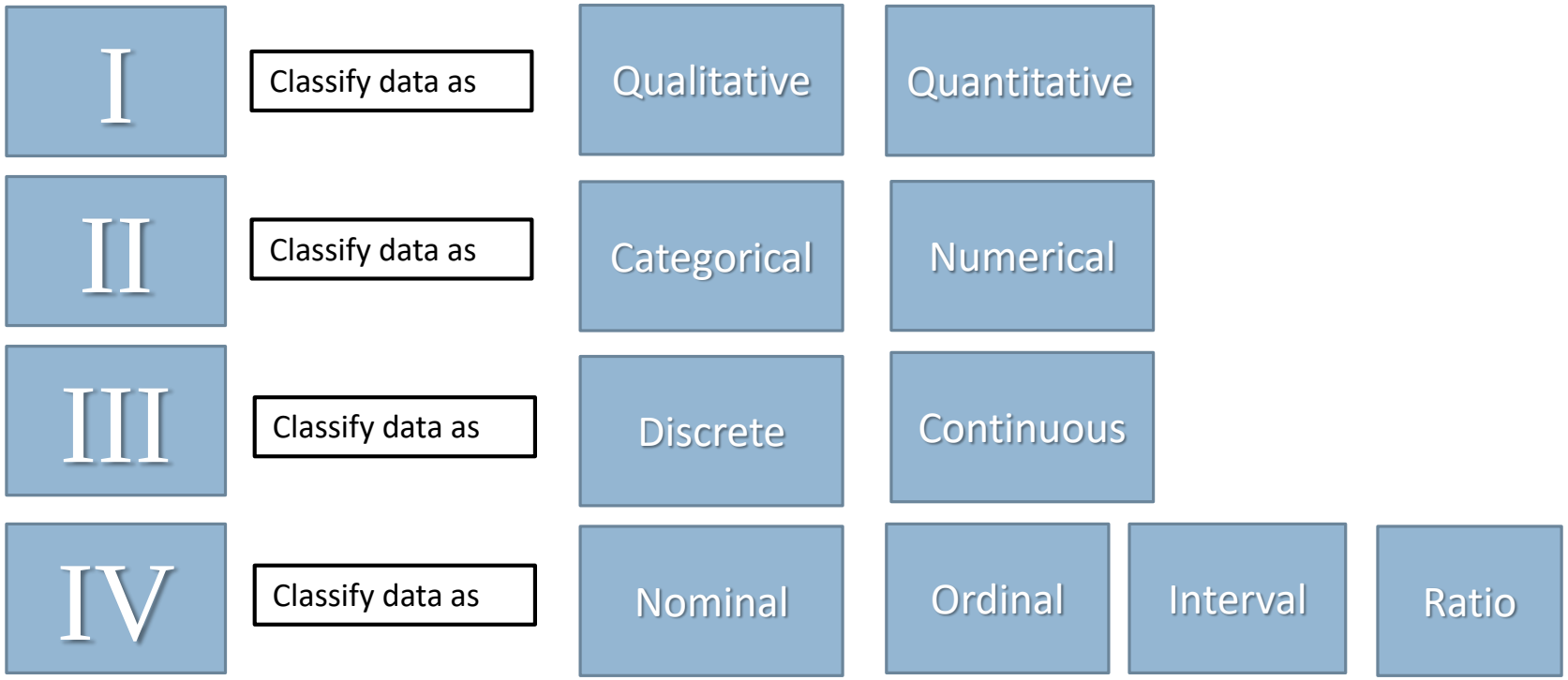
Dr. Garvan wants you to review the slides on Data Type before the “Choosing the Correct Stats Method” lecture on December 14, 2022.

2. Ways to Classify Data Types

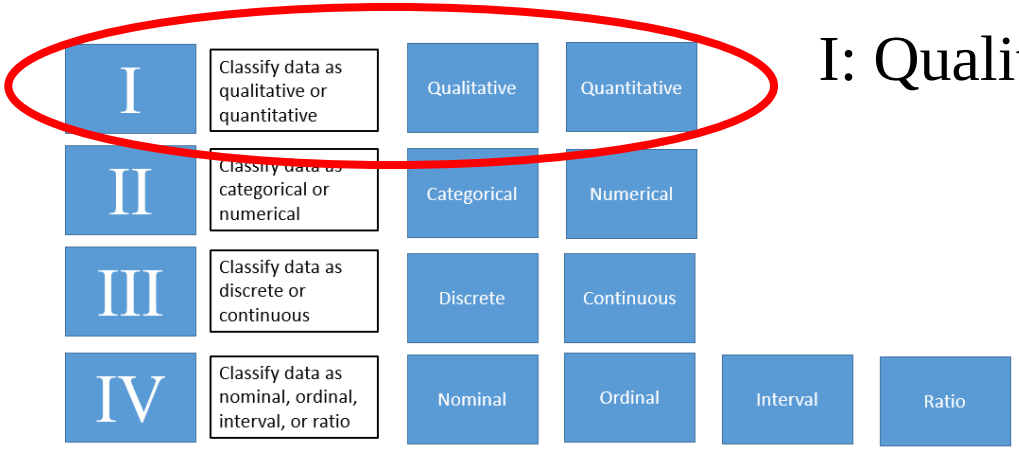
Data Types

FOUR CLASSIFICATION SYSTEMS

There are four ways to classify data type:



Analogous to different ways of classifying data, there are a different ways of classifying risk for postoperative outcomes of heart surgery patients. These include the ASA, the Cardiac Anesthesia Risk Evaluation score, EuroScore II, and the Clinical Frailty Scale.



I: Qualitative or Quantitative

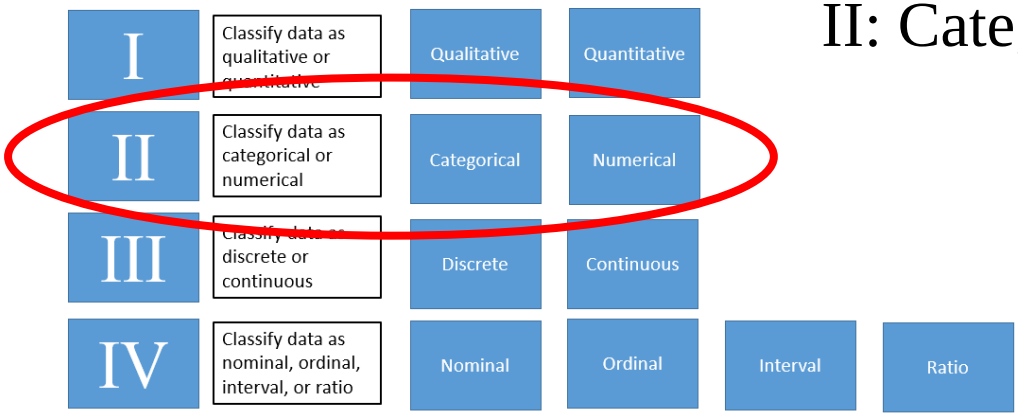
- ***Qualitative*:** A scale of measurement is a set of categories that vary in some quality but not in magnitude.
- ***Quantitative*:** A scale of measurement is a set of categories that vary in magnitude.

Examples of Qualitative and Quantitative data

Variable	Qualitative or Quantitative
Heart rate bpm	Quantitative
History of MI	Qualitative
ASA classification	Quantitative
Number of pRBCs given in surgery	Quantitative
Modified Fatigue Impact Scale (MFIS)	Quantitative
Cancer stage	Quantitative
Surgery type	Qualitative
Body temperature ^{o}C	Quantitative

Think of an example of a Qualitative variable.
Think of an example of a Quantitative variable.

Your Turn



II: Categorical or Numerical

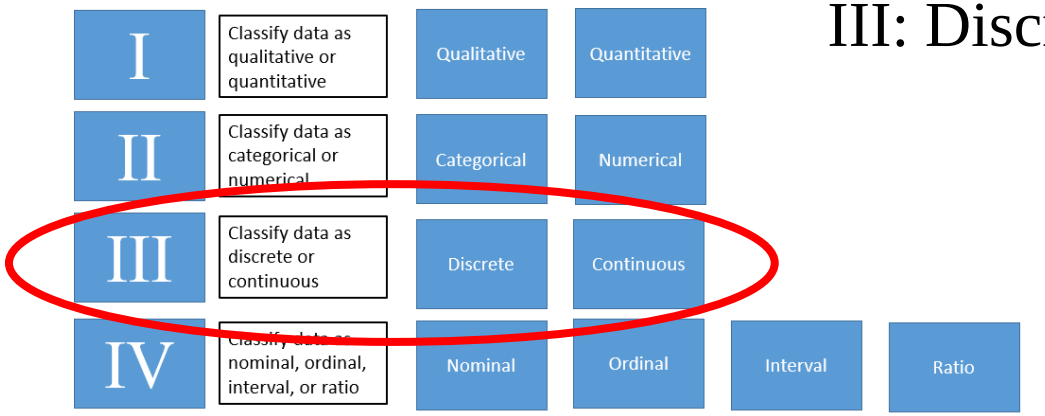
- ***Categorical***: A scale of measurement where levels are a set of categories.
- ***Numerical***: A scale of measurement where levels are a set of meaningful numbers such as integers or decimals.

Examples of Categorical and Numerical data

Variable	Categorical or Numerical
Heart rate bpm	Numerical
History of MI (YES/NO)	Categorical
ASA classification	Categorical
Number of pRBCs given in surgery	Numerical
Modifed Fatigue Impact Scale (MFIS)	Numerical
Cancer stage	Categorical
Surgery type	Categorical
Body temperature ^{o}C	Numerical

Think of an example of a Categorical variable.
Think of an example of a Numerical variable.

Your Turn



III: Discrete or Continuous

- ***Discrete:*** A variable that can take only selected values.
- ***Continuous:*** A numerical variable whose levels include (conceptually) all values between any two levels.

Discrete data are counted

Continuous data are measured.

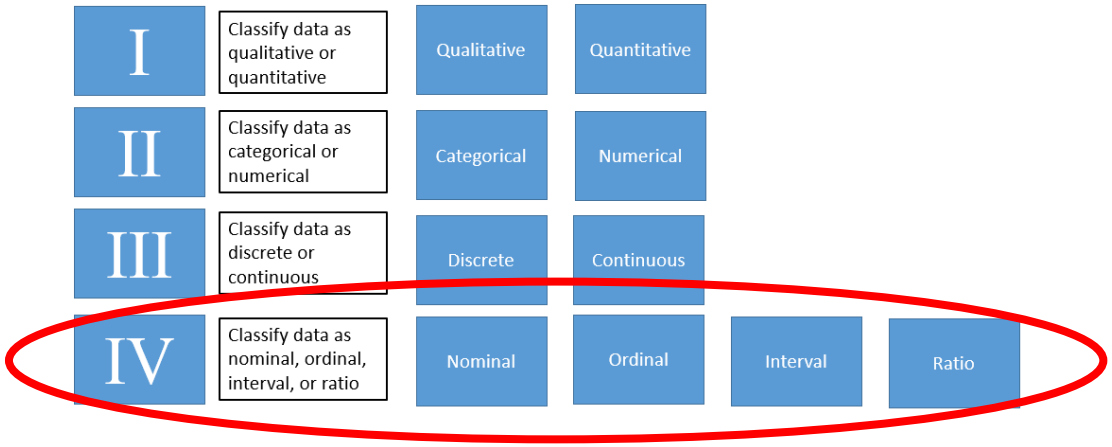
Examples of Discrete and Continuous data

Variable	Discrete or Continuous
Heart rate bpm	Continuous
History of MI (0/1; YES/NO)	Discrete
ASA classification	Discrete
Number of pRBCs given in surgery (1,2,..)	Discrete
Modified Fatigue Impact Scale (MFIS)	Continuous
Cancer stage	Discrete
Surgery type	Discrete
Body temperature ^{o}C	Continuous

Think of an example of a Discrete variable.
Think of an example of a Continuous variable.

Your Turn

IV: Nominal, Ordinal, Interval, Ratio



- ***Nominal:*** A scale of measurement where levels are distinct but do not vary in magnitude.
- ***Ordinal:*** A scale of measurement where levels vary in order of magnitude but equal intervals between levels cannot be assumed.
- ***Interval:*** The interval level of measurement has the characteristics of distinct levels, ordering in magnitude, and equal intervals.
- ***Ratio:*** The ratio level of measurement has characteristics of distinct levels, ordering in magnitude, equal intervals, and an absolute zero. A measurement has an absolute zero when a measurement of zero represents the absence of the property being measured.

Understanding Nominal, Ordinal, Interval, and Ratio Data

Measurement can take place at four different levels - Nominal, Ordinal, Interval, and Ratio.

The level of measurement is determined by noting the presence or absence of four characteristics:

1) distinctiveness,
2) ordering in magnitude,
3) equal intervals, and
4) an absolute zero.

Understanding Nominal, Ordinal, Interval, and Ratio Data

Characteristic	Meaning
Distinctiveness	Values assigned to an individual represent different attributes.
Ordering in magnitude	Values assigned to an individual can be ordered in a meaningful way.
Equal intervals	Equal intervals are obtained if equivalent differences between measurements represent the same amount of difference in the property being measured.
Absolute zero	A measurement has an absolute zero if a score of 0 represents an absence of the property being measured.

Distinctiveness

Values assigned to an individual represent different attributes.

For example, if we assign the number 1 to males and the number 2 to females, the measurement variable has the characteristic of ***distinctiveness***.

Ordering in magnitude

Values assigned to an individual can be ordered in a meaningful way.

For example, if a test score of 7 represents more knowledge of history than a test score of 5, then the scores have ordering in magnitude.

Equal intervals

Equal intervals are obtained if equivalent differences between measurements represent the same amount of difference in the property being measured.

For example, if a two-point difference between the scores of 108 and 110 represents the same amount of difference in vocabulary level as a two-point difference between the scores of 92 and 94, the measurement variable has equal intervals.

Example: The Modified Fatigue Impact Scale (MFIS)

Fatigue is a feeling of physical tiredness and lack of energy that many people experience from time to time. But people who have medical conditions like Parkinson's disease experience stronger feelings of fatigue more often and with greater impact than others.

The MFIS is a 21 item scale where each item can take a value from 0 (never) to 4 (almost always). Scores can range from 0 to 84.

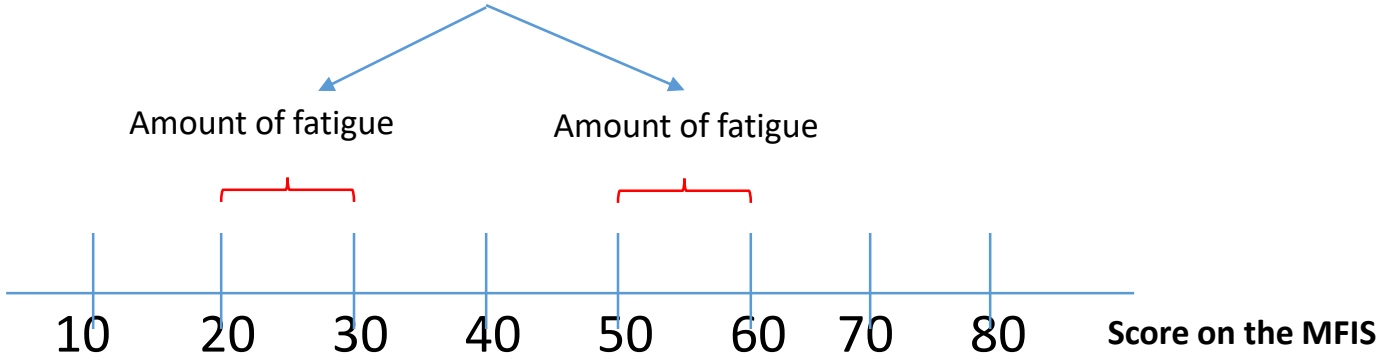
The first 12 items of the Modifed Fatigue Impact Scale

Because of my fatigue during the past 4 weeks

		Never	Rarely	Sometimes	Often	Almost Always
1.	I have been less alert.	0	1	2	3	4
2.	I have had difficulty paying attention for long periods of time.	0	1	2	3	4
3.	I have been unable to think clearly.	0	1	2	3	4
4.	I have been clumsy and uncoordinated.	0	1	2	3	4
5.	I have been forgetful.	0	1	2	3	4
6.	I have had to pace myself in my physical activities.	0	1	2	3	4
7.	I have been less motivated to do anything that requires physical effort.	0	1	2	3	4
8.	I have been less motivated to participate in social activities.	0	1	2	3	4
9.	I have been limited in my ability to do things away from home.	0	1	2	3	4
10.	I have trouble maintaining physical effort for long periods.	0	1	2	3	4
11.	I have had difficulty making decisions.	0	1	2	3	4
12.	I have been less motivated to do anything that requires thinking	0	1	2	3	4

Interval data = equal spacing

The MFIS is considered interval data. This means that the amount of fatigue between scores of 20 and 30 is the same amount of fatigue between the scores of 50 and 60.



Equal intervals are obtained if equivalent differences between measurements represent the same amount of difference in the property being measured.

Hoehn and Yahr Scale for staging Parkinson's disease does not have equally spaced intervals.

H & Y	Modified Hoehn and Yahr Scale for Staging Parkinson's Disease
1	Unilateral involvement only
1.5	Unilateral and axial involvement
2	Bilateral involvement without impairment of balance
2.5	Mild bilateral disease with recovery on pull test
3	Mild to moderate bilateral disease; some postural instability; physically independent
4	Severe disability; still able to walk or stand unassisted
5	Wheelchair bound or bedridden unless aided

The difference between Stage 1 and Stage 2 and the difference between Stage 3 and Stage 4 **DO NOT** represent the same amount of difference in the property being measured, i.e., advancement of disease.

Absolute zero

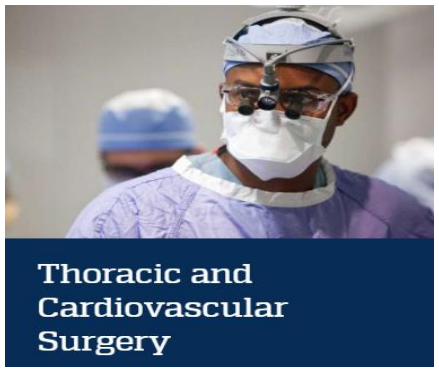
A measurement has an absolute zero if a score of 0 represents an absence of the property being measured.

For example, if a score of 0 represents an absence of errors made while running a maze, the measurement has an absolute zero.

The four characteristics of measurement determine the four major levels of measurement: nominal, ordinal, interval, and ratio.

Scale have values that are:	**Nominal**	**Ordinal**	**Interval**	**Ratio**
Distinctive	X	X	X	X
Ordered		X	X	X
Equally spaced			X	X
Has an absolute zero				X

Nominal data: A scale of measurement where levels are distinct but do not vary in magnitude. **Example: Surgery Type**



Examples of Nominal Data

Variable	Possible Values
Blood Group	A, B, AB, O
Sex	Male, Female
Disease Status	Present, Absent
Myers-Briggs Personality Type Extroversion (E)/Introversion (I) Sensing (S)/Intuition (N) Thinking (T)/Feeling (F) Judgment (J)/Perception (P)	ENFJ,ENFP,ENTJ,ENTP, ESFJ, ESFP, ESTJ,ESTP, INFJ, INFP, INTJ, INTP, ISFJ, ISFP, ISTJ, ISTP
Zip code	32605, 18901, 30066, etc.
Binary Outcome	Success or Failure (i.e., 0 or 1)

Think of an example of a Nominal data type variable.

Your Turn

Ordinal data: Data which can be ordered. Even though these data may be coded as numbers, they are not meaningful numbers. Example: Hoehn and Yahr Scale.

H & Y	Modified Hoehn and Yahr Scale for Staging Parkinson's Disease
1	Unilateral involvement only
1.5	Unilateral and axial involvement
2	Bilateral involvement without impairment of balance
2.5	Mild bilateral disease with recovery on pull test
3	Mild to moderate bilateral disease; some postural instability; physically independent
4	Severe disability; still able to walk or stand unassisted
5	Wheelchair bound or bedridden unless aided

Examples of Ordinal Data

Variable	Possible Values
Likert Scale data	Strongly Agree, Agree, Disagree, Strongly Disagree
Cancer Stage	Stage 0, Stage I, Stage II, Stage III, Stage IV
Glasgow Coma Score	3 = no response to 15= mild to no head trauma
Modified Hoehn and Yahr Scale	1, 1.5, 2, 2.5, 3, 4, 5
Chest Roentgenogram Score	No alveolar consolidation = 0; Alveolar consolidation confined to 1 quadrant = 1; Alveolar consolidation confined to 2 quadrants = 2; Alveolar consolidation confined to 3 quadrants = 3; Alveolar consolidation in all 4 quadrants = 4’

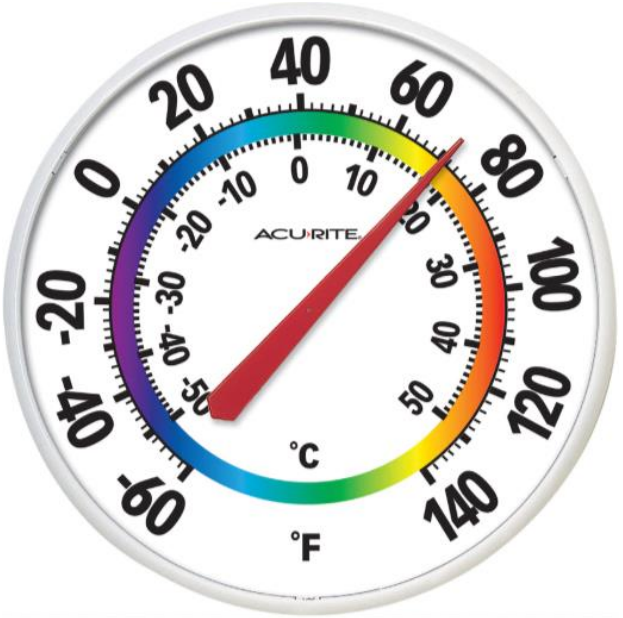
Think of an example of an Ordinal variable.

Your Turn

Interval Data: The interval level of measurement has the characteristics of distinct levels, ordering in magnitude, and equal intervals, but no true zero.

Temperatures in ^{o}F or ^{o}C are not ratio data

Both the Fahrenheit and Celsius temperature measurement scales are interval data. Both scales have the characteristics of distinct levels, ordering in magnitude, and equal intervals. However zero is an arbitrary point on both scales. Temperature of 10^{O}C is not twice as "hot" as 5^{O}C - the equivalent temperatures in Fahrenheit are 50^{O}F and 41^{O}F.



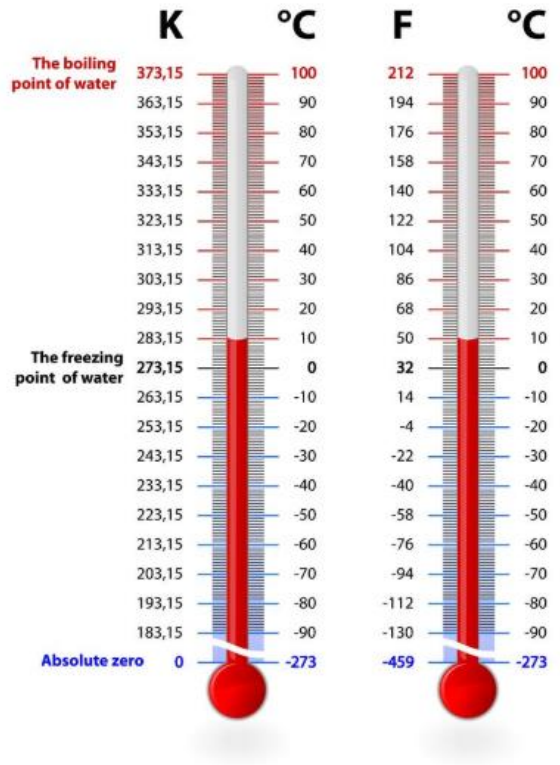
Examples of Interval Data

Variable	Possible Values
Modified Fatigue Impact Scale	0 - 84
36-Item Short Form Health Survey (SF-36)	The SF-36 consists of eight scaled scores. The eight sections are: vitality, physical functioning, bodily pain, general health perceptions, physical role functioning, emotional role functioning, social role functioning, and mental health. Each scale ranges from 0 – 100.
Geriatric Depression Scale	Scale score range from 0 - 30
Body temperature ^{o}C	Normal body temperature ranges from 36.1°C to 37.2°C

Ratio: The ratio level of measurement has characteristics of distinct levels, ordering in magnitude, equal intervals, and an absolute zero. A measurement has an absolute zero when a measurement of zero represents the absence of the property being

Example: What is the outdoor temperature in Kelvin?

In the Kelvin temperature scale, $^{o}0$ K is the point where all motion stops. So the Kelvin scale does have an absolute zero since the value of 0 represents the absence of the property being measured (i.e., molecular motion).



Unique Property of Ratio Data

Meaningful ratios of measurements or scores can be calculated only when measurement is made on a ratio level. If height were measured on an ordinal level using ranks (say 1 = less than 5’ in height, 2 = between 5’ and 5.5’ in height, 3 = between 5.5’ and 6’ in height, and 4 = greater than 6’), then ratios of scores (ranks) would not make any sense. A person with a rank of 2 would not necessarily be twice as tall as a person with a rank of 1. But when we measure height at the ratio level, we can calculate meaningful ratios. For example, a 6’ tall person is twice as tall as a 3’ person.

Examples of Ratio Data

Variable	Ranges
Age	0 – 100+ years
Blood pressure	Systolic (mm Hg) less than 120, Diastolic (mm Hg) less than 80
Grip strength Grip strength is a measure of muscular strength or the maximum force/tension generated by one's forearm muscles.	Less than 12 (kg) – 60+ (kg)
Anesthesia gas flow	0.3 – 6.0 L/min

Think of an example of a Ratio variable.

Your Turn

A variable can be classified using each classification schema.

KEY POINT

Examples of Data Classified in Different Ways

	I	II	III	IV
Variable	**Qualitative or Quantitative**	**Categorical or Numerical**	**Discrete or Continuous**	**Nominal Ordinal Interval Ratio**
Race/ethnicity	Qualitative	Categorical	Discrete	Nominal
Age	Quantitative	Numerical	Continuous	Ratio
Number of medications taken	Quantitative	Numerical	Discrete	Ratio
SF-36 - short-form health survey	Quantitative	Numerical	Continuous	Interval
Chest Roentgenogram Score	Quantitative	Numerical	Discrete	Ordinal

Fill in the blanks in the table on the next slide.

Knowledge Check

Complete the following table:

	I	II	III	IV
Variable	**Qualitative or Quantitative**	**Categorical or Numerical**	**Discrete or Continuous**	**Nominal Ordinal Interval Ratio**
Heart rate bpm				
History of MI				
ASA classification				
Number of pRBCs given in surgery				
Modified Fatigue Impact Scale (MFIS)				
Cancer stage				
Surgery type				

Answers are on next slide!

Answers to Knowledge Check

	I	II	III	IV
Variable	**Qualitative or Quantitative**	**Categorical or Numerical**	**Discrete or Continuous**	**Nominal Ordinal Interval Ratio**
Heart rate bpm	Quantitative	Numerical	Continuous	Ratio
History of MI	Qualitative	Categorical	Discrete	Nominal
ASA classification	Quantitative	Categorical	Discrete	Ordinal
Number of pRBCs given in surgery	Quantitative	Numerical	Discrete	Ratio
Modified Fatigue Impact Scale (MFIS)	Quantitative	Numerical	Continuous	Interval
Cancer stage	Quantitative	Categorical	Discrete	Ordinal
Surgery type	Qualitative	Categorical	Discrete	Nominal


QUESTIONS?
GREAT SMOKY MOUNTAINS
NATIONAL PARK

References

Allen, M.J. and Yen, W.M., 2001. *Introduction to measurement theory*. Waveland Press.