

# **LECTURE I**

# **BASIC TERMS AND DEFINITIONS**

**PSY2002**

**Hye Won Suk**

# TOPICS FOR THIS LECTURE

- Basic terms and definitions
  - What is statistics?
  - Population and Sample
  - Parameters and Statistics
  - Descriptive and Inferential Statistics
  - Variables and Constants
  - Scale of Measurement

# WHAT IS STATISTICS?

- “*Statistics is the grammar of science.*” - Karl Pearson
- Statistics is a set of mathematical procedures for organizing, summarizing, and interpreting information.

# WHAT IS STATISTICS?

*“We live in an information age. Computers allow us to collect and store information in quantities that previously would not even have been dreamed of. What is this information? It might be costs, values, sales volumes, measurements, ratings, distances, prices, percentages, counts, times, or market shares. But raw, undigested data stored on computer disks is of no use until we can start to make sense of it. Statistics is the human side of the computer revolution, an information science, the science (and art!) of extracting meaning from seemingly incomprehensible data. In your future life and career, you will need to be able to make good use of such information to make sound decisions.”*

([www.sjsu.edu/statistics/why/](http://www.sjsu.edu/statistics/why/))

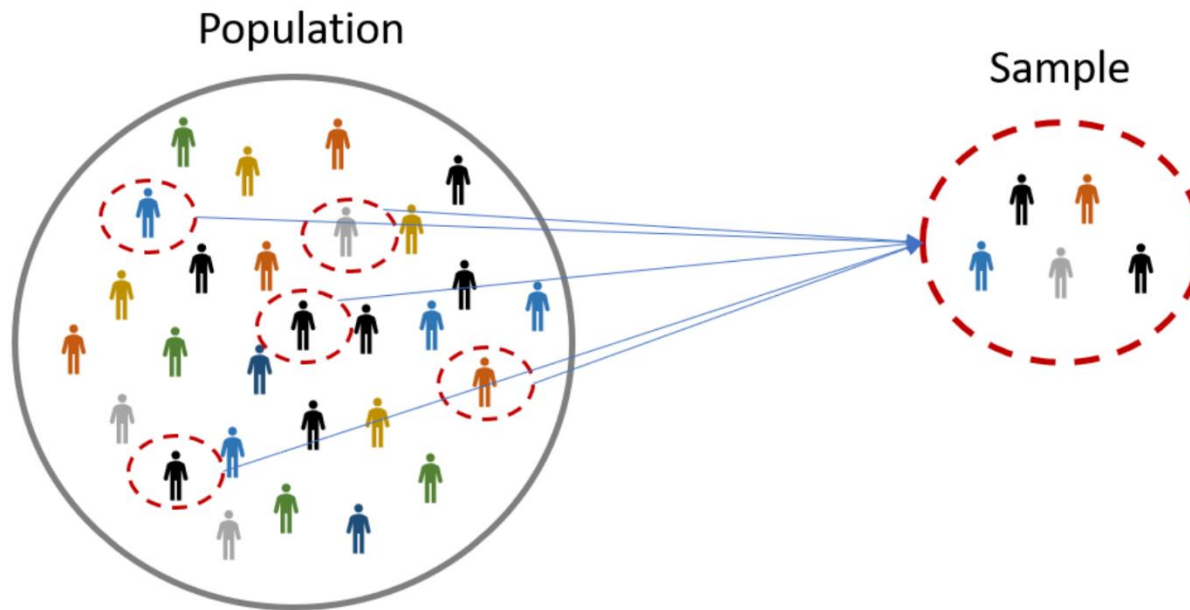
# AN EXAMPLE

- A drug company developed a new drug for depressed patients, and wanted to test the effectiveness of the new drug compared to their standard drug.
- The drug company recruited 100 newly-diagnosed clinically depressed patients. Each patient was randomly assigned to one of the two groups. One group consisting of 50 participants took the new drug, and the other group consisting of another 50 participants took the standard drug. Average depression score for each group after 6 months was measured :
  - New drug = 7.2 (out of 10)
  - Standard drug = 8.3 (out of 10)
- Is the new drug more beneficial than the standard drug?

# AN EXAMPLE

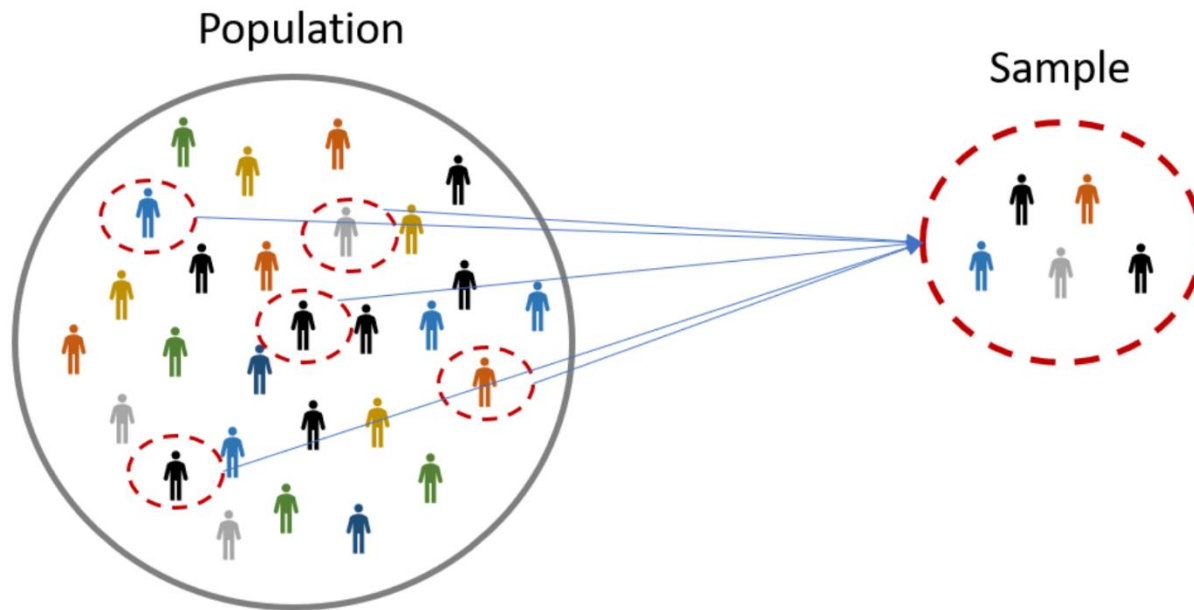
- Before you answer the question, let's think about the followings.
  - What if the drug company recruited different 100 patients and performed the same experiment? Do you expect that the result would be exactly the same?
  - How confident are you in generalizing the result to all other depressed patients?

# POPULATION AND SAMPLE



- A population (모집단) is the complete collection of all individuals, units, or elements of interest in a particular study to which the researcher wishes to generalize the study findings.

# POPULATION AND SAMPLE



- A sample (표본) is a subset of individuals, units, or elements of interest in a particular study; the individuals, units, or elements selected from the population, who are chosen for participation in the study.



# POP QUIZ

- A researcher wants to determine if there is a relationship between the amount of daily caffeine consumption (in milligrams) of young adults living in Korea and their daily sleeping hours (in hours). The researcher examines 200 young adults who live in Seoul.
- What is the population of this study?
- What is the sample of this study?

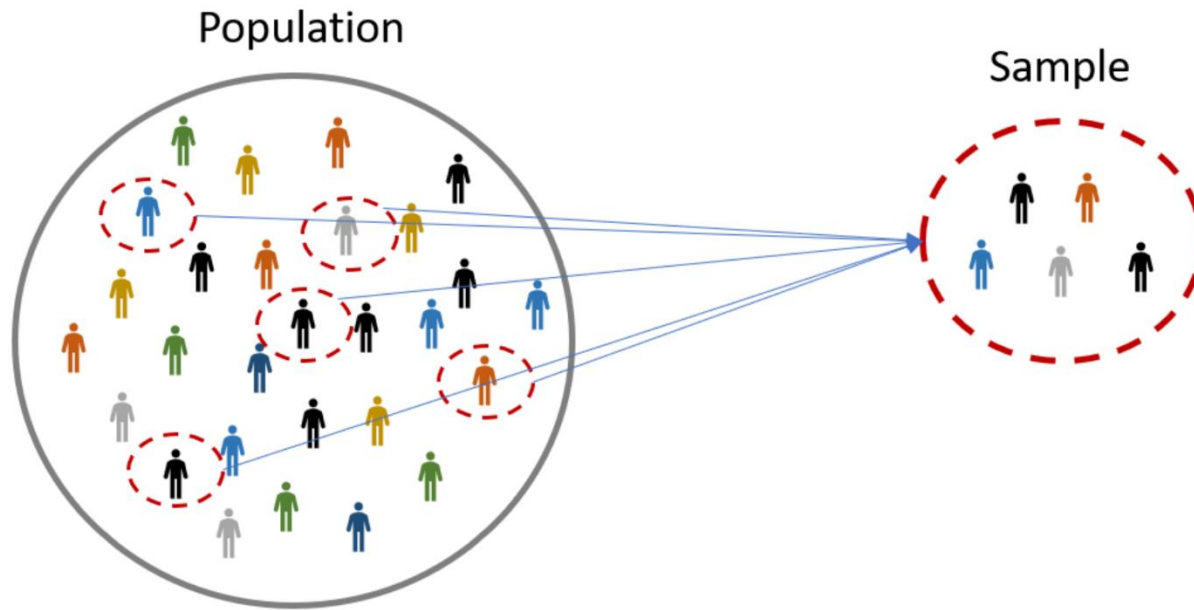
# POP QUIZ

- A hotel manager wants to examine how satisfied their guests were last year. He randomly selects 150 guests who stayed in the hotel last year and sends a guest satisfaction survey questionnaire to them.
- What is the population of this survey?
- What is the sample of this survey?

<b>POPULATION</b>	<b>SAMPLE</b>
<ul style="list-style-type: none"> <li>• The entire set of individuals, units, or elements (e.g., people, animals, objects) of interest in a particular study</li> </ul>	<ul style="list-style-type: none"> <li>• A subset of individuals, units, or elements of interest in a particular study</li> </ul>
<ul style="list-style-type: none"> <li>• The entire set of depressed patients</li> <li>• The entire set of young adults living in Korea</li> <li>• The entire set of guests who stayed in the hotel last year</li> </ul>	<ul style="list-style-type: none"> <li>• 100 newly-diagnosed clinically depressed patients</li> <li>• 200 young adults who live in Seoul</li> <li>• 150 guests who stayed in the hotel last year</li> </ul>
<ul style="list-style-type: none"> <li>• Usually very large in size</li> <li>• Almost impossible to examine every individual in the population due to limitation in time and money</li> </ul>	<ul style="list-style-type: none"> <li>• Researchers decide how many individuals there will be in a sample</li> <li>• Manageable in size</li> </ul>

# PARAMETERS AND STATISTICS

- A POPULATION PARAMETER (모수) is a quantity that describes a population. A parameter is derived from measurements of the entire individuals of the population, which is almost impossible to be obtained.
- A SAMPLE STATISTIC (통계치, 통계량) is a quantity that describes a sample. A sample statistic is derived from measurements of the individuals in the sample, which can be obtained in practice.



The entire set of 11350 guests who stayed in the hotel last year.

Average satisfaction score = 78

This population average is a population parameter.

A group of randomly-selected 150 guests who stayed in the hotel last year.

Average satisfaction score = 72

This sample average is a sample statistic.

# PARAMETERS AND STATISTICS

- In most research, we want to find population parameters, but it is almost impossible to obtain population parameters.
- Therefore, we collect samples from populations and obtain sample statistics instead of population parameters. And then, we *infer* about population parameters based on sample statistics.
- Every population parameter has a corresponding sample statistic.
- There is no such thing like a “*population statistic*” and “*sample parameter*.”

# POP QUIZ

- A drug company developed a new drug for depressed patients, and wanted to test the effectiveness of the new drug compared to their standard drug.
- The drug company recruited 100 newly-diagnosed clinically depressed patients. Each patient was randomly assigned to one of the two groups. One group consisting of 50 participants took the new drug, and the other group consisting of another 50 participants took the standard drug. Average depression score for each group after 6 months was measured :
  - New drug = 7.2 (out of 10)
  - Standard drug = 8.3 (out of 10)
- Are these average scores population parameters or sample statistics?

# POP QUIZ

- A professor wanted to examine the performance of her students last semester. A total of 58 students were enrolled in her PSY2002 class last semester. The average final score of these 58 students was 87 out of 100.
- Is this average score a population parameter or a sample statistic?



# DESCRIPTIVE AND INFERENTIAL STATISTICS

- DESCRIPTIVE STATISTICS (기술 통계) are the statistical procedures used to summarize, organize, and simplify data. They can be used with samples and populations.
  - Frequency Distributions
  - Central Tendency Measures
  - Variability Measures
- INFERENTIAL STATISTICS (추론 통계) are the statistical procedures that allow us to study samples and then make generalizations about the populations from which they were selected. They can only be used with samples.
  - Estimation
  - Hypothesis Testing

# VARIABLES AND CONSTANTS

- A VARIABLE (변수/변인) is a characteristic, property, or condition that takes on different values among the individuals of the sample or the population.
  - Depression, Understanding, Anxiety, Intelligence, Height, Gender, Ethnicity, ...
- A CONSTANT (상수) is a characteristic, property, or condition that has the same value for all the individuals of the sample or the population.

# VARIABLES AND CONSTANTS

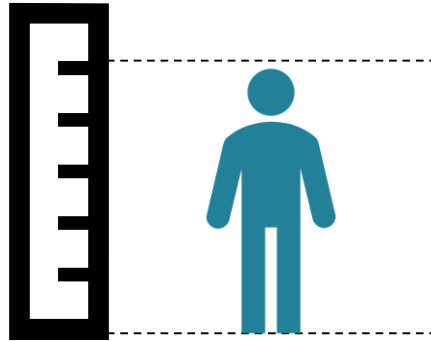
- Suppose a professor wants to examine the relation between motivation and performance among the students who are taking PSY2002.
- For each of the followings, is it a variable or a constant?
  - Motivation
  - Performance
  - Whether a student is taking PSY2002 or not
- In statistics, we are interested in *variables*, not constants.

# SCALE OF MEASUREMENT

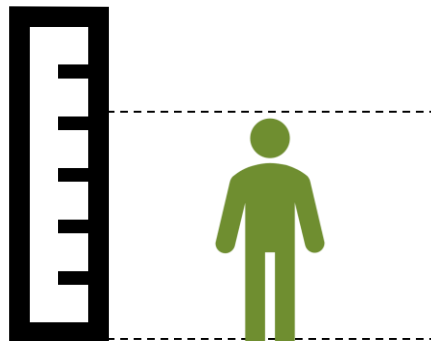
- When we have variables of interest in a study, we should measure them.
- Level of measurement (측정 수준) or scale of measurement (측정 척도) is a classification that describes the nature of information within the values assigned to variables.

# SCALE OF MEASUREMENT

- Let's say that a researcher measures participants' height. It can be measured in different ways.



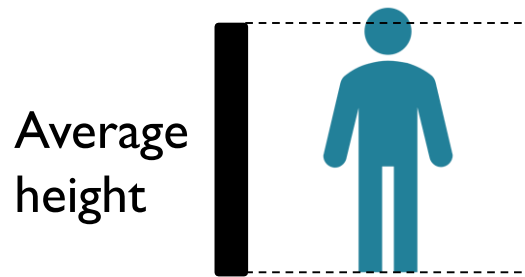
Participant A: 176 cm  
For participant A, the variable (height) value is 176.



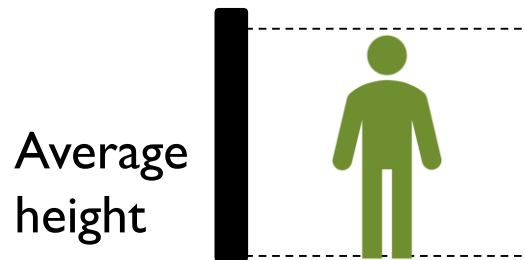
Participant B: 154 cm  
For participant B, the variable (height) value is 154.

# SCALE OF MEASUREMENT

- Let's say that a researcher measures participants' height. It can be measured in different ways.



Participant A: taller than average  
For participant A, the variable value is 1.



Participant B: shorter than average  
For participant B, the variable value is 2.

# SCALE OF MEASUREMENT

- In the previous example, we measured the same characteristic (height) of the participants.
- However, depending on the level of measurement, each participant has different values.
- Also, depending on the level of measurement, the obtained values deliver different information about the measured characteristic.

# SCALE OF MEASUREMENT

- We often classify variables according to the level of measurement.
  - Nominal, Ordinal, Interval, and Ratio
- It is important to take into account the level of measurement.
  - The kind of statistics that we can use varies with the level of measurement.
  - How we measure a variable affects the amount of information that we gather.



# SCALE OF MEASUREMENT

- NOMINAL SCALE (명목 척도)
  - Measurements on a nominal scale label and categorize observations, but do not make any quantitative distinctions between observations (i.e., only permit classification into categories.)
  - Any numbers that are assigned to categories are arbitrary.
  - Examples
    - Sex (male, female)
    - Ethnicity
    - Whether you pass or fail an exam
    - Whether you are taller than the average or not

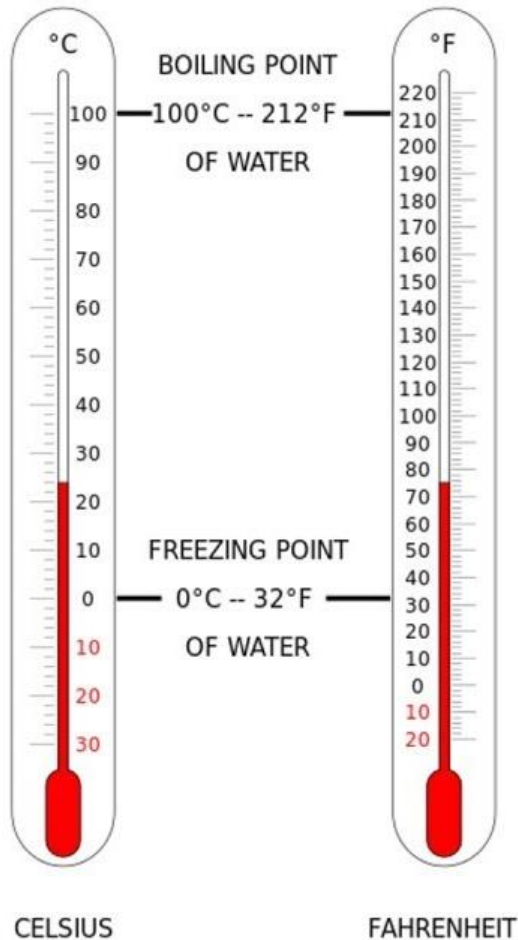
# SCALE OF MEASUREMENT

- ORDINAL SCALE (서열 척도)
  - Measurements on an ordinal scale categorize observations as well as rank observations in terms of size or magnitude.
  - Numbers are assigned to reflect an ordering from high to low, but differences between the numbers is not interpretable. The numbers provides rankings only.
  - Examples
    - Consumer rankings of cars (first, second, third,...)
    - Socioeconomic class (1 = upper, 2 = middle, 3 = lower)
    - Anxiety level (1 = very high, 2 = relatively high, 3 = moderate, 4 = relatively low, 5 = low)
    - Letter grade (A+, A0, A-, B+, B0, B-,...)

# SCALE OF MEASUREMENT

- INTERVAL SCALE (등간 척도)
  - Measurements on an interval scale categorize and rank observations and equal differences between numbers reflect equal differences in magnitude.
  - However, zero point on an interval scale is arbitrary and does not indicate a zero amount of the variable being measured.
  - Does not permit statement about ratios of measurements.
  - Examples
    - Temperature
    - IQ score

# AN EXAMPLE: TEMPERATURE



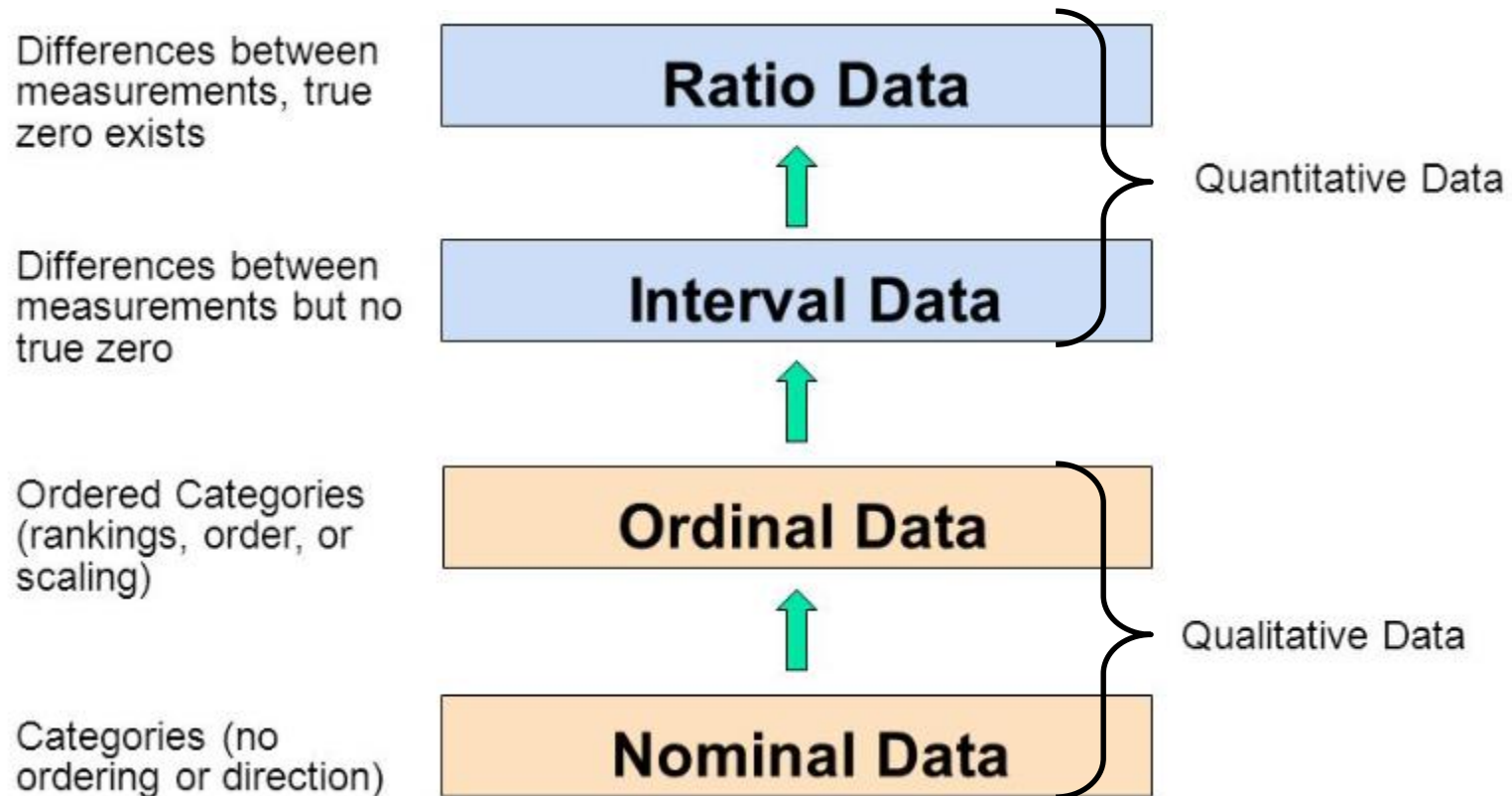
$$^{\circ}\text{F} = \frac{9}{5}^{\circ}\text{C} + 32$$

- Zero point is arbitrary.
  - $0^{\circ}\text{C}$  or  $0^{\circ}\text{F}$  does not indicate that there is no temperature.
- Statement about ratios of measurements is not permitted.
  - $100^{\circ}\text{C}$  is twice as high as  $50^{\circ}\text{C}$ ?
  - $212^{\circ}\text{F}$  is twice as high as  $122^{\circ}\text{F}$ ?

# SCALE OF MEASUREMENT

- RATIO SCALE (비율 척도)
  - Measurements on a ratio scale categorize and rank observations and equal differences between numbers reflect equal differences in magnitude (same as interval scale).
  - A ratio scale has an absolute zero point.
  - Permits statement about ratios of measurements.
  - Examples
    - Height
    - Weight
    - Time to complete a task (measured in minutes)
    - Annual rainfall (measured in millimeters)

# SCALE OF MEASUREMENT



# POP QUIZ

- Nominal, Ordinal, Interval or Ratio?
  - University you belong to
  - Hours worked in a typical week (measured in hours)
  - Whether you work 40 hours or more in a typical week (yes/no)
  - Professor levels (assistant, associate, full)
  - Handedness (left-handed, right-handed)
- For handedness, does it make sense to ask the average?
- For hours worked in a typical week, does it make sense to ask the average?

# SUMMARY

- Basic terms and definitions
  - Population and Sample
  - Parameters and Statistics
  - Descriptive and Inferential Statistics
  - Variables and Constants
  - Scale of Measurement
    - Nominal
    - Ordinal
    - Interval
    - Ratio