People
Innovation
Excellence

GREATER JAKARTA • BANDUNG • MALANG

STAT6171001
Basic Statistics

What is Statistics? & Descriptive and Inferential Statistics
Session 1

Maria Seraphina Astriani
seraphina@binus.ac.id

Session Learning Outcomes

Upon completion of this session, students are expected to be able to

- LO 1. Explain basic statistics concept

Topics

- What is statistics?
- Descriptive and inferential statistics

What is Statistics?

Statistics

- Statistics deals with variability
- Statistics include numerical facts and figures
- The study of statistics involves math and relies upon calculations of numbers
- It also relies heavily on how the numbers are chosen and how the statistics are interpreted.

Descriptive and Inferential Statistics

Descriptive Statistics

- Descriptive statistics are numbers that are used to summarize and describe data.

- The word “data” refers to the information that has been collected from an experiment, a survey, an historical record, etc.

- Descriptive statistics are just descriptive.

- They do not involve generalizing beyond the data at hand.

Descriptive Statistics

- Statistics exists because of the prevalence of variability in the real world.

- In its simplest form, known as descriptive statistics, statistics provides us with tools—tables, graphs, averages, ranges, correlations—for organizing and summarizing the inevitable variability in collections of actual observations or scores.

Descriptive Statistics

Examples:

1. A tabular listing, ranked from most to least, of the total number of movies
2. A graph showing the annual change in global temperature during the last 30 years
3. A report that describes the average difference in grade point average (GPA) between different programs

Descriptive Statistics

Example: Average salaries for various occupations in 1999.

$112,760	pediatricians
$106,130	dentists
$100,090	podiatrists
$76,140	physicists
$53,410	architects,
$49,720	school, clinical, and counseling psychologists
$47,910	flight attendants
$39,560	elementary school teachers
$38,710	police officers
$18,980	floral designers

Inferential Statistics

- Statistics also provides tools—a variety of tests and estimates—for generalizing beyond collections of actual observations.
- This more advanced area is known as inferential statistics.

- Tools from inferential statistics permit us to use a relatively small collection of actual observations to evaluate

- Inferential statistics can be defined as a field of statistics that uses analytical tools for drawing conclusions about a population by examining random samples

Inferential Statistics

Examples:

1. A researcher’s hypothesis that, on average, meditators report fewer headaches than do nonmeditators
2. An assertion about the relationship between job satisfaction and overall happiness

Inferential Statistics

Case:

A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score. He concludes from their report that the class did extremely well.

What is the sample?

What is the population?

Can you identify any problems with choosing the sample in the way that the teacher did?

Inferential Statistics

Case – Analyse:

The population consists of all students in the class.

The sample is made up of just the 10 students sitting in the front row. The sample is not likely to be representative of the population.

Those who sit in the front row tend to be more interested in the class and tend to perform higher on tests.

Hence, the sample may perform at a higher level than the population.

Three Types of Data

- Any statistical analysis is performed on data, a collection of actual observations or scores in a survey or an experiment
- The precise form of a statistical analysis often depends on whether data are:
  1. Quantitative
  2. Ranked
  3. Qualitative
- To determine the type of data, focus on a single observation in any collection of observations

Table 1.1

Table 1.1
QUANTITATIVE DATA: WEIGHTS (IN POUNDS) OF MALE STATISTICS STUDENTS

160	168	133	170	150	165	158	165
193	169	245	160	152	190	179	157
226	160	170	180	150	156	190	156
157	163	152	158	225	135	165	135
180	172	160	170	145	185	152	
205	151	220	166	152	159	156	
165	157	190	206	172	175	154	

Quantitative Data

- Quantitative data consist of numbers (weights of 238, 170, . . . 185 lbs) that represent an amount or a count.
- For example, the weights reported by 53 male students are quantitative data (Table 1.1), since any single observation, such as 160 lbs, represents an amount of weight.

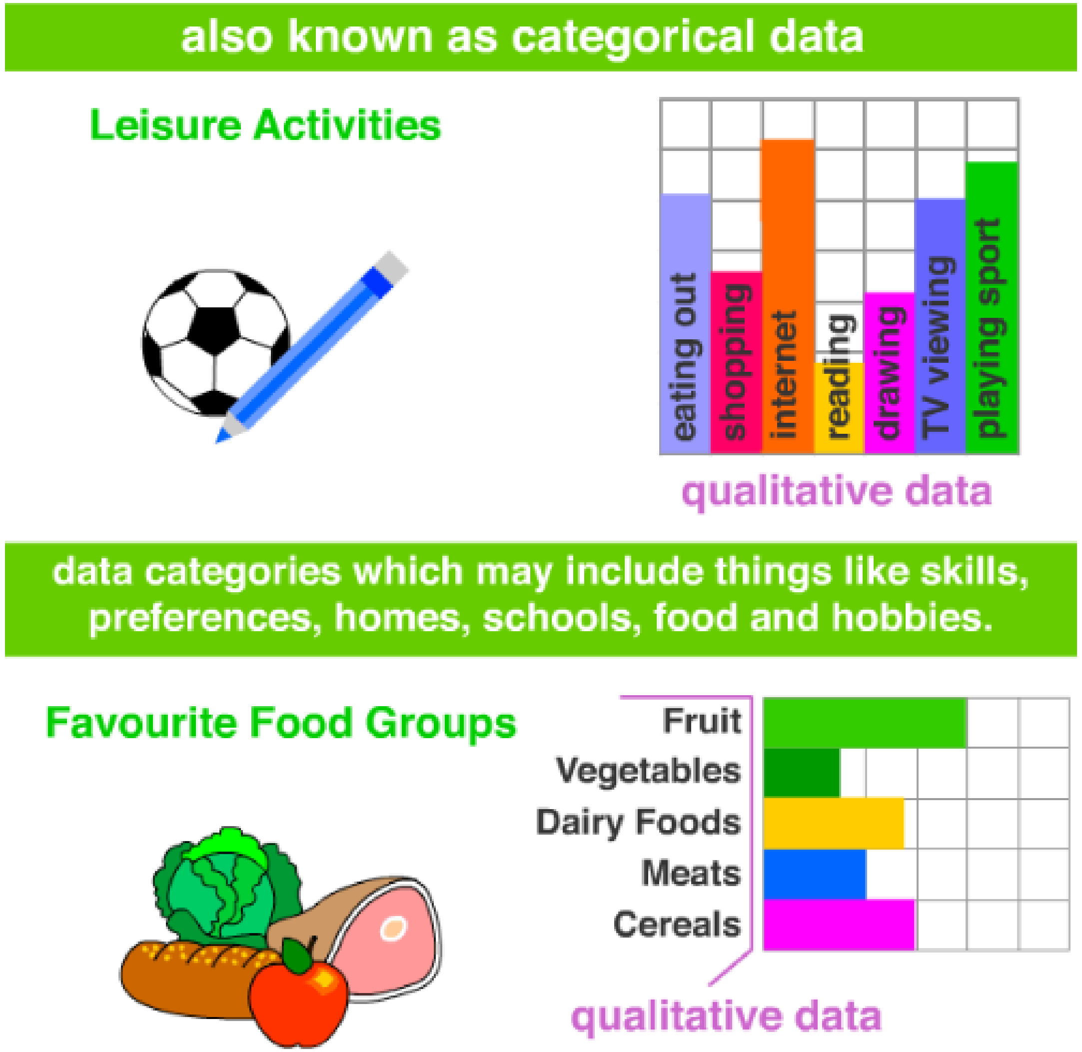
Ranked Data

- Ranked data consist of numbers (1st, 2nd, . . . 40th place) that represent relative standing within a group.
- If the weights in Table 1.1 had been replaced with ranks, beginning with a rank of 1 for the lightest weight of 133 lbs and ending with a rank of 53 for the heaviest weight of 245 lbs, these numbers would have been ranked data, since any single observation represents not an amount, but only relative standing within the group of 53 students.

Qualitative Data

- If there were a group of people in a room, qualitative data could describe how they feel, what they look like, what clothes they are wearing, or the motivations of why they're here.
- Whereas quantitative data about the same group may include the number of people in the group, their age, or the temperature in the room.

Qualitative Data



source: http://www.amathsdictionaryforkids.com/qr/q/qualitativeData.html

Level of Measurement

- The level of measurement specifies the extent to which a number (or word or letter) actually represents some attribute and, therefore, has implications for the appropriateness of various arithmetic operations and statistical procedures.
- There are **three levels** of measurement—**nominal**, **ordinal**, and **interval/ratio**—and these levels are paired with qualitative, ranked, and quantitative data, respectively.

Qualitative Data and Nominal Measurement

- If people are classified as either male or female (or coded as 1 or 2), the data are qualitative, and measurement is nominal.
- The single property of nominal measurement is classification—that is, sorting observations into different classes or categories.
- Examples of nominal measurement include classifying mood disorders as manic, bipolar, or depressive.

Ranked Data and Ordinal Measurement

- When any single number indicates only relative standing, such as first, second, or tenth place in a horse race or in a class of graduating seniors, the data are ranked, and the level of measurement is ordinal.
- The distinctive property of ordinal measurement is order. Comparatively speaking, a first-place finish reflects the fastest finish in a horse race or the highest GPA among graduating seniors.
- Although first place in a horse race indicates a faster finish than second place, we don’t know how much faster

Ranked Data and Ordinal Measurement

- Since ordinal measurement fails to reflect the actual distance between adjacent ranks, simple arithmetic operations with ranks are inappropriate.
- For example, it’s inappropriate to conclude that the arithmetic mean of ranks 1 and 3 equals rank 2, since this assumes that the actual distance between ranks 1 and 2 equals the distance between ranks 2 and 3. Instead, these distances might be very different.
- For example, rank 2 might be virtually tied with either rank 1 or rank 3.

Quantitative Data and Interval/Ratio Measurement

- Often the products of familiar measuring devices, such as rulers, clocks, or meters, the distinctive properties of interval/ratio measurement are equal intervals and a true zero.
- Equal intervals imply that the difference between 60 and 70 kg represents an amount of weight equal to the difference between 70 and 80 kg, and it’s appropriate to describe one person’s weight as a certain amount greater than another’s
- A true zero signifies that the scale registers 0 when not in use—that is, when weight is completely absent.

Measurement of Nonphysical Characteristics

- When numbers represent nonphysical characteristics, such as intellectual aptitude, psychopathic tendency, or emotional maturity, the attainment of interval/ratio measurement often is questionable.
- For example, there is no external standard (such as 10 kg) to demonstrate that the addition of a fixed amount of intellectual aptitude always produces an equal increase in IQ scores.
- There also is no instrument (such as the unoccupied scale) that registers an IQ score of 0 when intellectual aptitude is completely absent (true zero).

Measurement of Nonphysical Characteristics

- In the absence of equal intervals, inappropriate to claim that the difference between IQ scores of 120 & 130 represents the same amount of intellectual aptitude as the difference between 130 & 140.
- Likewise, in the absence of a true zero, it would be inappropriate to claim that an IQ score of 140 represents twice as much intellectual aptitude as an IQ score of 70.
- One possibility is to treat IQ scores as attaining only ordinal measurement—that is, for example, a score of 140 represents more intellectual aptitude than a score of 130—without specifying the actual size of this difference.

Types of Data and Levels of Measurement

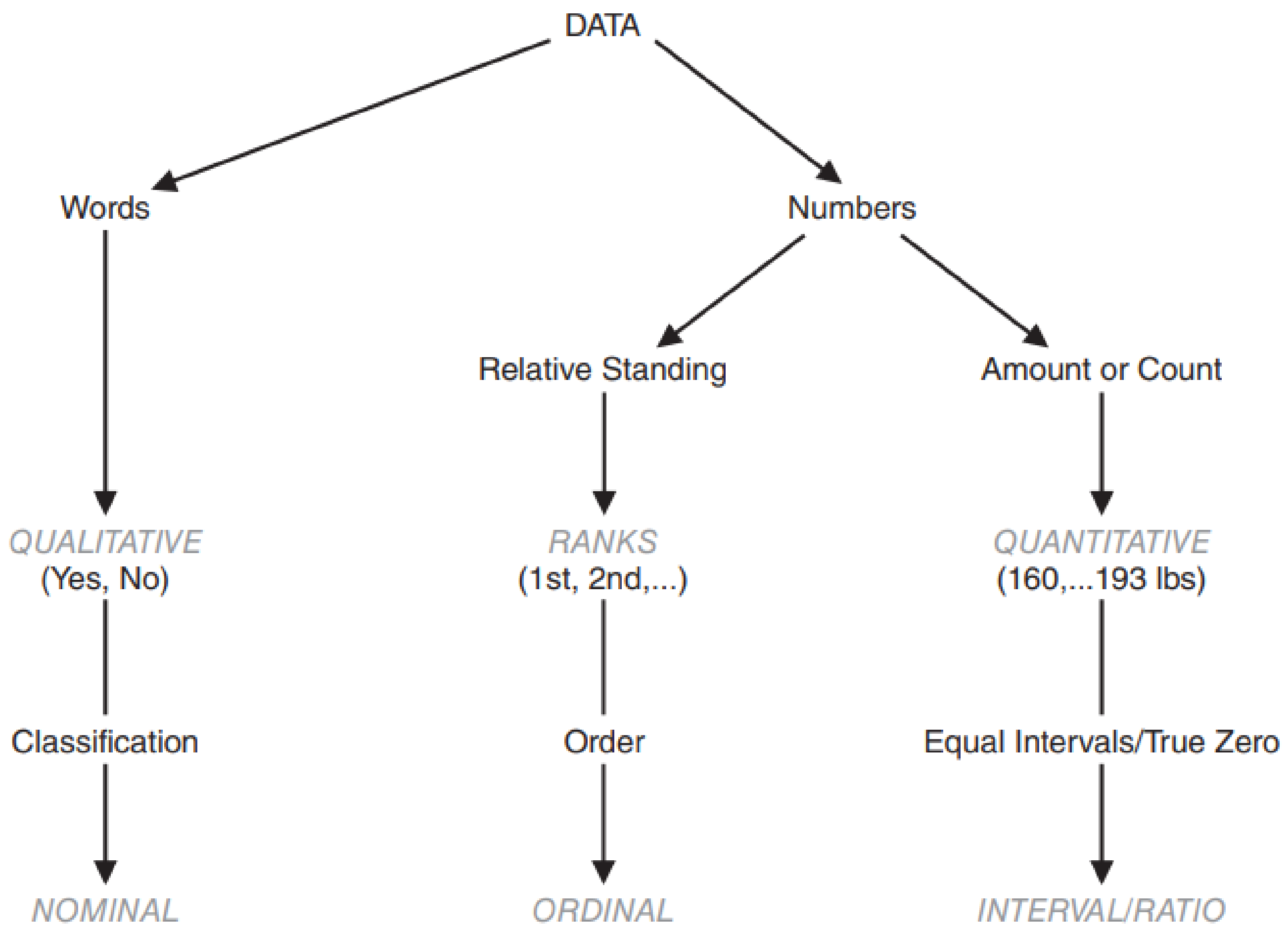


FIGURE 1.2
Overview: types of data and levels of measurement.

Types of Variables

- A variable is a characteristic or property that can take on different values.
- Any single observation in Table 1.1 can be described as a constant, since it takes on only one value.

Discrete and Continuous Variables

- Quantitative variables can be further distinguished in terms of whether they are discrete or continuous.
- A discrete variable consists of isolated numbers separated by gaps.
- Examples: the number of children in a family; the number of foreign countries you have visited.
- A continuous variable consists of numbers whose values, at least in theory, have no restrictions.
- Examples: include amounts, such as weights of male statistics students; durations, such as the reaction times of grade school children to a fire alarm; and standardized test scores, such as those on SAT.

Approximate Numbers

- In theory, values for continuous variables can be carried out infinitely far. Someone's weight, in pounds, might be 140.01438, and so on, to infinity! Practical considerations require that values for continuous variables be rounded off.
- Whenever values are rounded off, as is always the case with actual values for continuous variables, the resulting numbers are approximate, never exact.

Approximate Numbers

- For example, the weights of the students in Table 1.1 are approximate because they have been rounded to the nearest pound.
- A student whose weight is listed as 150 lbs could actually weigh between 149.5 and 150.5 lbs.

Independent and Dependent Variables

- Variables are properties or characteristics of some event, object, or person that can take on different values or amounts.
- When conducting research, experimenters often manipulate variables.
- For example, an experimenter might compare the effectiveness of four types of antidepressants. In this case, the variable is “type of antidepressant.”
- When a variable is manipulated by an experimenter; it is called an independent variable.

Independent and Dependent Variables

- The experiment seeks to determine the effect of the independent variable on relief from depression.
- In this example, relief from depression is called a dependent variable.
- In general, the independent variable is manipulated by the experimenter and its effects on the dependent variable are measured.

Example #1
Can blueberries slow down aging?

- A study indicates that antioxidants in blueberries may slow down the process of aging. In this study, 19-month-old rats (60-year-old humans) were fed by either blueberry, strawberry, or spinach powder. After 8 weeks, the rats were given memory and motor skills tests. Although all supplemented rats showed improvement, supplemented with blueberry powder showed the most notable improvement.

1. What is the **independent** variable? (dietary supplement: none, blueberry, strawberry, and spinach)
2. What are the **dependent** variables? (memory test and motor skills test)

Example #2
Does beta-carotene protect against cancer?

- Beta-carotene supplements have been thought to protect against cancer. However, a study published in the Journal of the National Cancer Institute suggests this is false. The study was conducted with 39,000 women aged 45 and up. These women were randomly assigned to receive a beta-carotene supplement or a placebo, and their health was studied over their lifetime. Cancer rates for women taking the beta-carotene supplement did not differ systematically from the cancer rates of those women taking the placebo.

1. What is the **independent** variable? (supplements: beta-carotene or placebo)
2. What is the **dependent** variable? (occurrence of cancer)

Example #3
How bright is right?

- An automobile manufacturer wants to know how bright brake lights should be in order to minimize the time required for the driver of a following car to realize that the car in front is stopping and to hit the brakes.

1. What is the **independent** variable? (brightness of brake lights)
2. What is the **dependent** variable? (time to hit brakes)

References

- Witte, R.S.&Witte, J.S. (2017). Statistics (11th ed.). Wiley. ISBN: 978-1119386056.
- Lane, D.M., Scott, D., Hebl, M., Guerra, R., Osherson, D.& Zimmer, H. (2003). Introduction to Statistics. Online edition at https://open.umn.edu/opentextbooks/textbooks/459
- Levine, D.M., Stephan, D.F. & Szabat, K.A. (2017). Statistics for Managers Using Microsoft Excel (8th ed.). Pearson. ISBN: 978-0134566672

Thank you