IF 2220 Probabilitas & Statistika

Semester 2
2020/2021

Prodi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Silabus IF2220

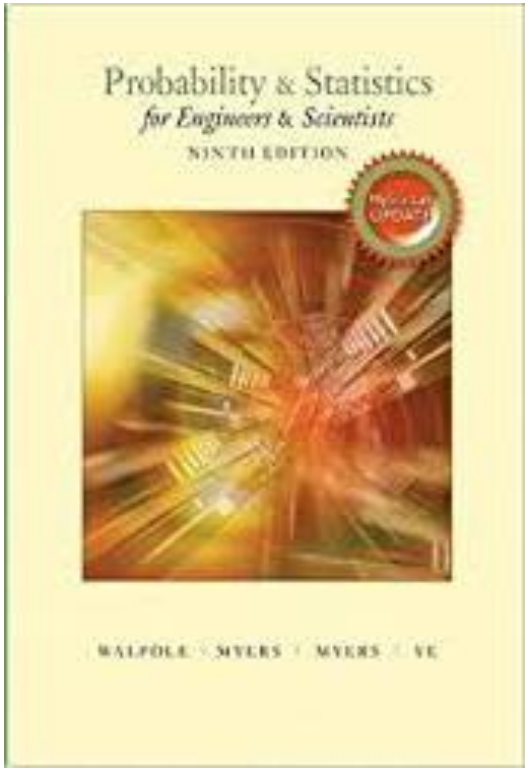
- Kredit : 3 SKS
- Semester : Genap
- KK/Lab: Informatika/IRK
- Sifat: Wajib
- Kelompok Kuliah: MK Dasar Engineering
- Silabus ringkas: Peluang, ruang sample, kejadian, operasi dengan kejadian, peluang sebuah kejadian, beberapa hukum peluang, peluang bersyarat dan aturan Bayes. Peubah acak, peubah acak diskrit, peubah acak kontinu. Matematika ekspektasi, Beberapa distribusi diskrit dan distribusi kontinu, Distribusi sampling, Teori penaksiran dan Pengujian hipotesis, Regresi
- Luaran: Mahasiswa memahami dan dapat menyelesaikan persoalan distribusi peluang variabel random diskrit dan kontinu, dan mahasiswa mampu menyelesaikan persoalan untuk menarik kesimpulan mengenai parameter populasi yang diperoleh dari data hasil eksperimen.
- Prerequisit: Kalkulus I, Kalkulus II

Atmosfer Perkuliahan

- 1 sks kuliah: 1 jam tatap muka di kelas + 1 jam mengerjakan berbagai tugas + 1 jam belajar mandiri
- Jenis tugas:
  - PR (setiap selesai satu topik)
  - Tugas dengan bantuan komputer (skala: kecil, besar)
  - Belajar mandiri (KUIS, UTS, UAS)
- Kegiatan mahasiswa selain kuliah terkait akademik
  - HMIF:
    - Tutorial
  - Kompetisi (internal & eksternal)
  - Asisten perkuliahan
  - Asisten lab

Silabus - Buku Acuan

- Probability- Aturan Bayes
- Variabel random
- Ekspektasi matematika
- Distribusi peluang variabel random (diskrit, kontinu)
- Distribusi sampel
- Estimasi parameter distribusi peluang
- Tes hipotesa
- Regresi



Buku Acuan: Probability & Statistics for Engineers & Scientists, Ninth Edition, Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, Pearson International Edition, 2014

Probability dalam Kehidupan Sehari-hari

- Menurut Bidang Sektor dalam Kehidupan:

1. Pemerintahan (asal kata Statista = negarawan (italia))
2. Kesehatan
3. Keuangan
4. Pertambangan
5. Pertanian
6. Konsumsi dan Retail
7. Pendidikan
8. Teknologi
9. Hiburan dan Seni
10. Jasa
11. Transportasi
12. Manufaktur dan lainnya

10 Everyday Reasons Why Statistics Are Important

1. Weather Forecasts
2. Emergency Preparedness
3. Predicting Disease
4. Medical Studies
5. Genetics
6. Political Campaigns
7. Insurance
8. Consumer Goods
9. Quality Testing
10. Stock Market

Source: http://www.mathworksheetscenter.com/mathtips/statsareimportant.html

Manfaat Probability & Statistics

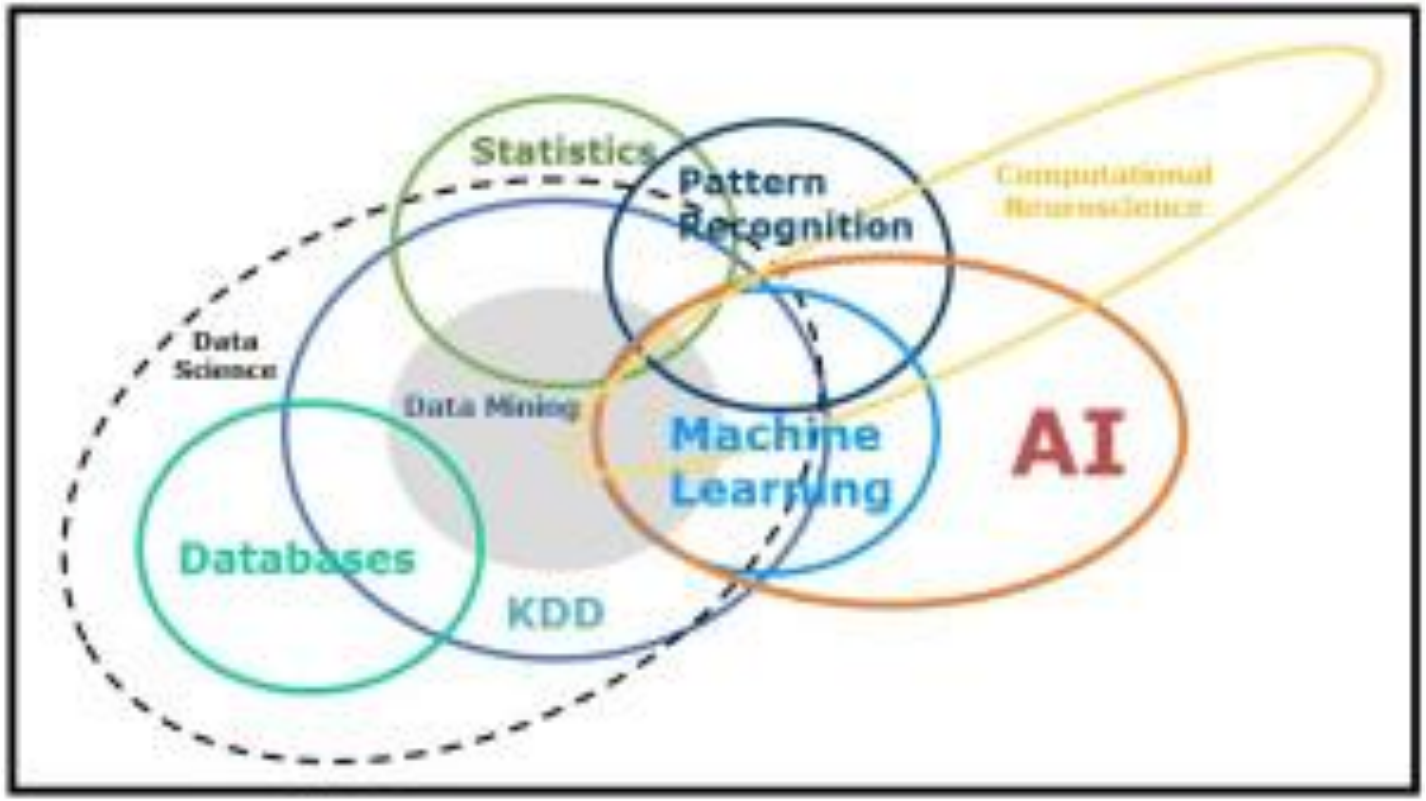
- Memperbaiki Proses/Produk suatu sistem.
- Kinerja (performansi) sistem: suatu ukuran/besaran efektifitas sistem.
- Prediksi: suatu ukuran / besaran hasil histori untuk rencana yang akan datang.
- Keandalan: suatu ukuran/ besaran untuk mengurangi kegagalan.
- Garansi(warranty), Asuransi: suatu ukuran/besaran harapan hidup.
- Deteksi keaslian gambar, video, dll
- Hubungan variable-variabel dalam suatu sistem.

Di ITB semua fakultas / sekolah, prodi – prodi mengajarkan Probability & Statistika walaupun nama MK berbeda-beda: Pengolahan data, Deskripsi data, dll

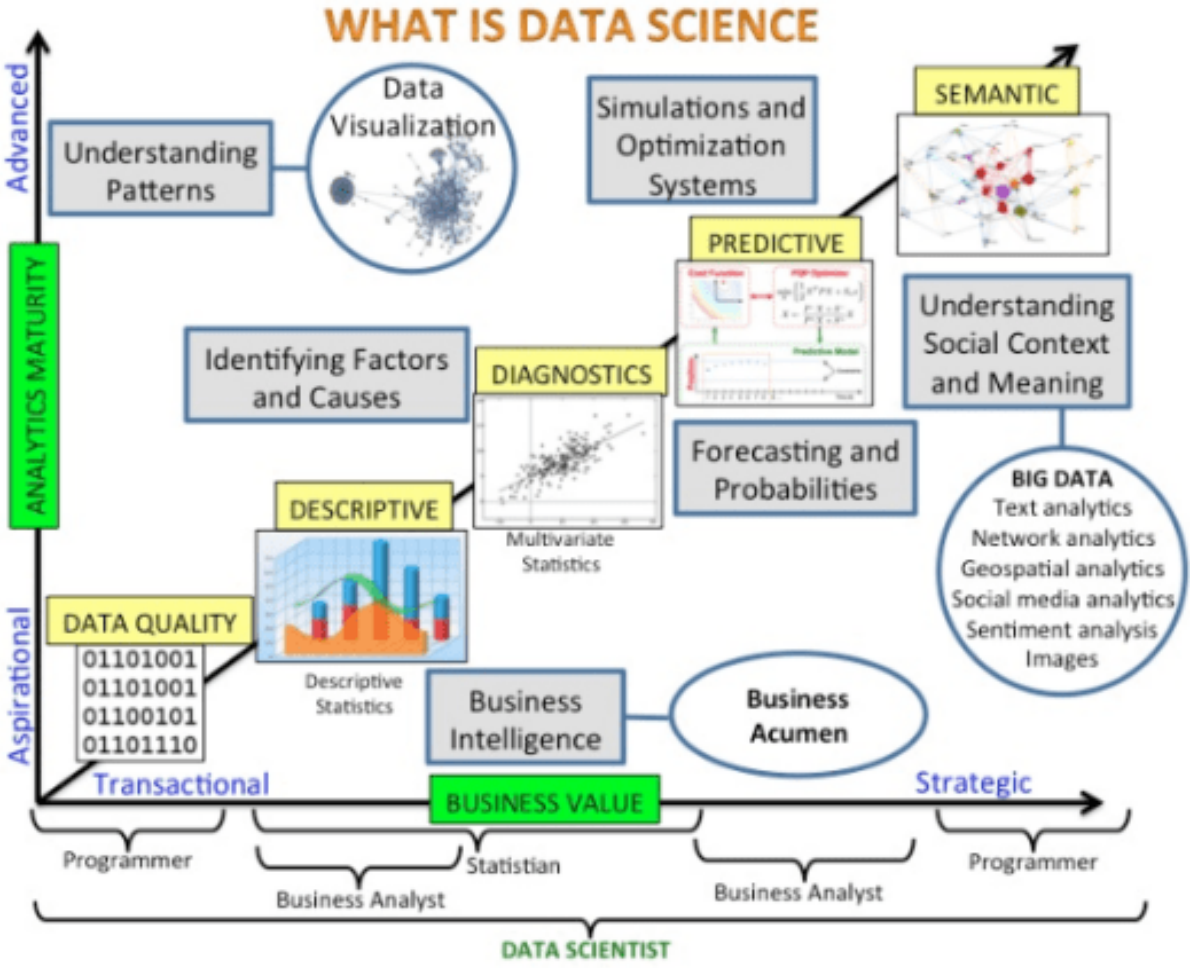
Kaitan Probability-MK Prodi IF

- Kecerdasan Buatan
- Pembelajaran Mesin
- Pemodelan dan Simulasi
- Jaringan Komputer Lanjut
- Pemodelan Data Lanjut
- Sistem Temu Balik Informasi
- Pemrosesan Bahasa Alami

Statistics and Computer Science



Statistics in Data Science



Machine Learning	Statistics
Problem Framing	exploratory data analysis and data mining
Data Understanding	summary statistics and data visualization
Data Cleaning	data sampling and feature selection methods
Data Selection	data sampling and feature selection methods
Data Preparation	data transforms, scaling, encoding and much more
Model Evaluation	experimental design and resampling methods
Model Configuration	statistical hypothesis tests and estimation statistics
Model Selection	statistical hypothesis tests and estimation statistics
Model Presentation	estimation statistics such as confidence intervals
Model Predictions	estimation statistics such as prediction intervals

Statistics in Machine Learning

Chapter 1: Introduction to Statistics

Statistics & Probability

Source: UCI Department of Statistics:

- Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting and presenting empirical data.
- Two fundamental ideas in the field of statistics are uncertainty and variation.
- Probability is a mathematical language used to discuss uncertain events and probability plays a key role in statistics. Any measurement or data collection effort is subject to a number of sources of variation.

Variables

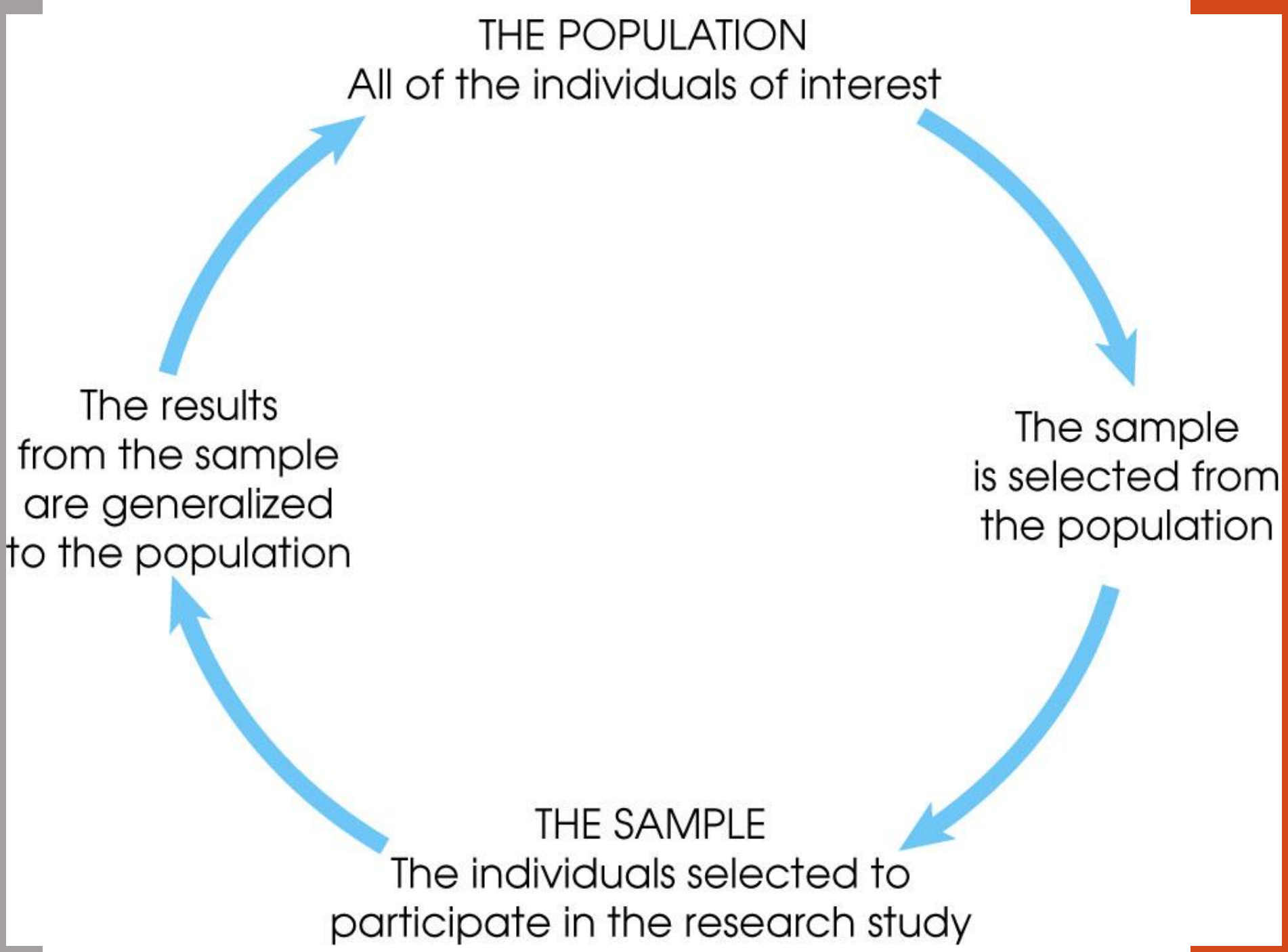
- A **variable** is a characteristic or condition that can change or take on different values.
- Most research begins with a general question about the relationship between two variables for a specific group of individuals.

Population

- The entire group of individuals is called the **population**.
- For example, a researcher may be interested in the relation between class size (variable 1) and academic performance (variable 2) for the population of third-grade children.

Sample

- Usually populations are so large that a researcher cannot examine the entire group. Therefore, a **sample** is selected to represent the population in a research study. The goal is to use the results obtained from the sample to help answer questions about the population.


THE POPULATION
All of the individuals of interest
The sample is selected from the population
THE SAMPLE
The individuals selected to participate in the research study
The results from the sample are generalized to the population

Types of Variables

- Variables can be classified as discrete or continuous.
- **Discrete variables** (such as class size) consist of indivisible categories, and
- **Continuous variables** (such as time or weight) are infinitely divisible into whatever units a researcher may choose. For example, time can be measured to the nearest minute, second, half-second, etc.

Real Limits

- To define the units for a continuous variable, a researcher must use **real limits** which are boundaries located exactly half-way between adjacent categories.

Measuring Variables

- To establish relationships between variables, researchers must observe the variables and record their observations. This requires that the variables be **measured**.
- The process of measuring a variable requires a set of categories called a **scale of measurement** and a process that classifies each individual into one category.

4 Types of Measurement Scales

1. A **nominal scale** is an unordered set of categories identified only by name. Nominal measurements only permit you to determine whether two individuals are the same or different.

2. An **ordinal scale** is an ordered set of categories. Ordinal measurements tell you the direction of difference between two individuals.

4 Types of Measurement Scales

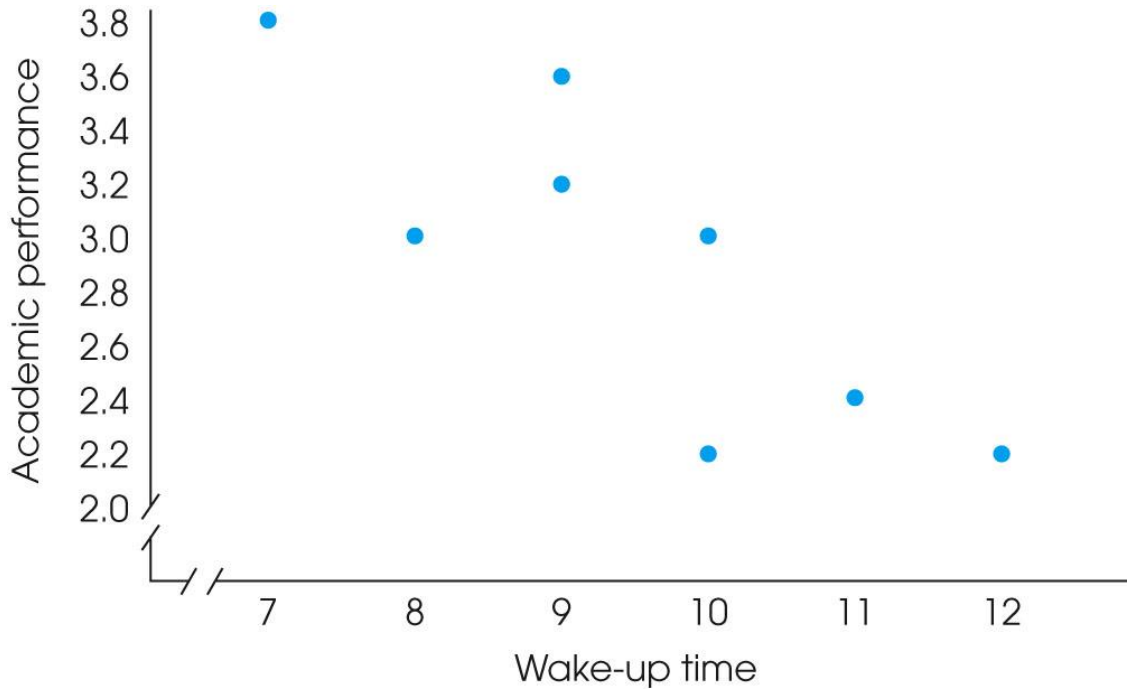
3. An **interval scale** is an ordered series of equal-sized categories. Interval measurements identify the direction and magnitude of a difference. The zero point is located arbitrarily on an interval scale.

4. A **ratio scale** is an interval scale where a value of zero indicates none of the variable. Ratio measurements identify the direction and magnitude of differences and allow ratio comparisons of measurements.

Correlational Studies

- The goal of a **correlational** study is to determine whether there is a relationship between two variables and to describe the relationship.
- A **correlational** study simply observes the two variables as they exist naturally.

Child	Wake-up Time	Academic Performance
A	11	2.4
B	9	3.6
C	9	3.2
D	12	2.2
E	7	3.8
F	10	2.2
G	10	3.0
H	8	3.0



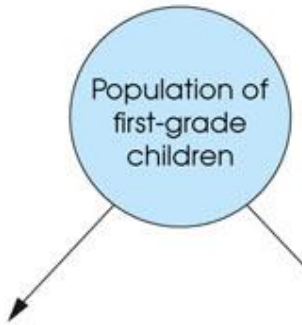
Experiments

- The goal of an **experiment** is to demonstrate a cause-and-effect relationship between two variables; that is, to show that changing the value of one variable causes changes to occur in a second variable.

Experiments (cont.)

- In an **experiment**, one variable is manipulated to create treatment conditions. A second variable is observed and measured to obtain scores for a group of individuals in each of the treatment conditions. The measurements are then compared to see if there are differences between treatment conditions. All other variables are controlled to prevent them from influencing the results.
- In an experiment, the manipulated variable is called the **independent variable** and the observed variable is the **dependent variable**.

Step 1
Experiment:
Compare two teaching methods



Data
Test scores for the students in each sample

Sample A
Taught by Method A

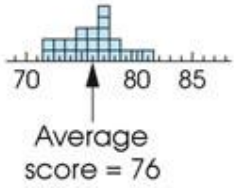
73	75	72	79
76	77	75	77
72	75	76	78
80	74	76	78
73	77	74	81
77	77		

Sample B
Taught by Method B

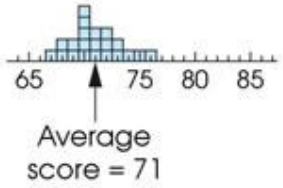
68	70	73	71
67	72	70	71
75	68	70	71
72	74	69	72
76	73	70	70
69			

Step 2
Descriptive statistics:
Organize and simplify

Sample A



Sample B



Step 3
Inferential statistics:
Interpret results

The sample data show a 5-point difference between the two teaching methods. However, there are two ways to interpret the results:

1. There actually is no difference between the two teaching methods, and the sample difference is due to chance (sampling error).
2. There really is a difference between the two methods, and the sample data accurately reflect this difference.

The goal of inferential statistics is to help researchers decide between the two interpretations.

Descriptive Statistics

- **Descriptive statistics** are methods for organizing and summarizing data.
- For example, tables or graphs are used to organize data, and descriptive values such as the average score are used to summarize data.
- A descriptive value for a population is called a **parameter** and a descriptive value for a sample is called a **statistic**.

Inferential Statistics

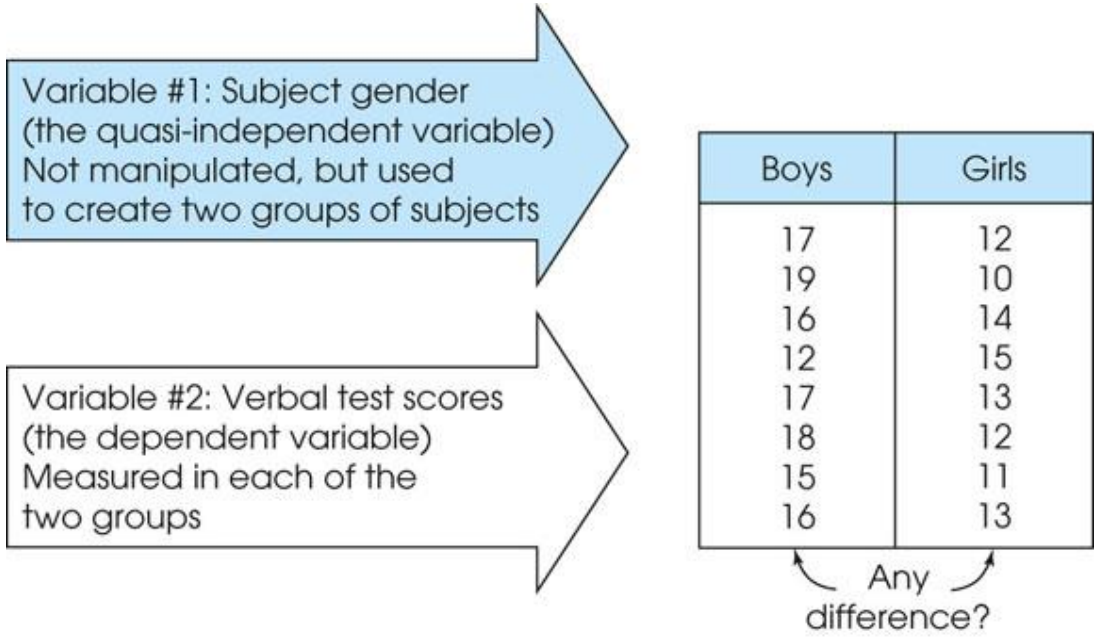
- **Inferential statistics** are methods for using sample data to make general conclusions (inferences) about populations.
- Because a sample is typically only a part of the whole population, sample data provide only limited information about the population. As a result, sample statistics are generally imperfect representatives of the corresponding population parameters.

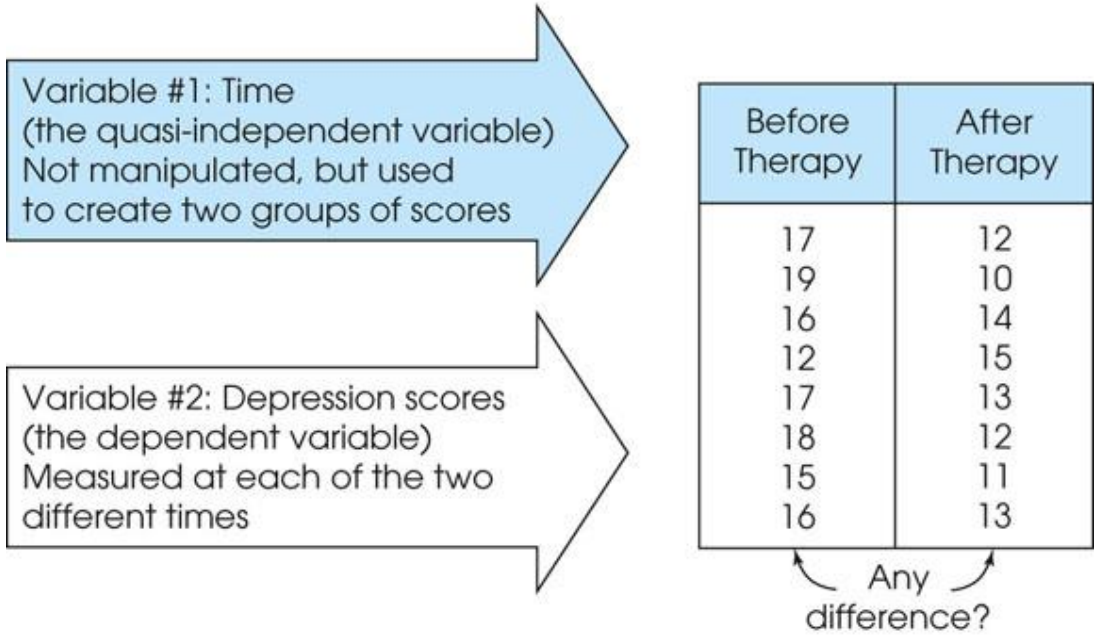
Other Types of Studies

- Other types of research studies, know as **non-experimental** or **quasi-experimental**, are similar to experiments because they also compare groups of scores.
- These studies do not use a manipulated variable to differentiate the groups. Instead, the variable that differentiates the groups is usually a pre-existing participant variable (such as male/female) or a time variable (such as before/after).

Other Types of Studies (cont.)

- Because these studies do not use the manipulation and control of true experiments, they cannot demonstrate cause and effect relationships. As a result, they are similar to correlational research because they simply demonstrate and describe relationships.


Variable #1: Subject gender
(the quasi-independent variable)
Not manipulated, but used
to create two groups of subjects
Variable #2: Verbal test scores
(the dependent variable)
Measured in each of the
two groups
Boys
Girls
17 12
19 10
16 14
12 15
17 13
18 12
15 11
16 13
Any
difference?

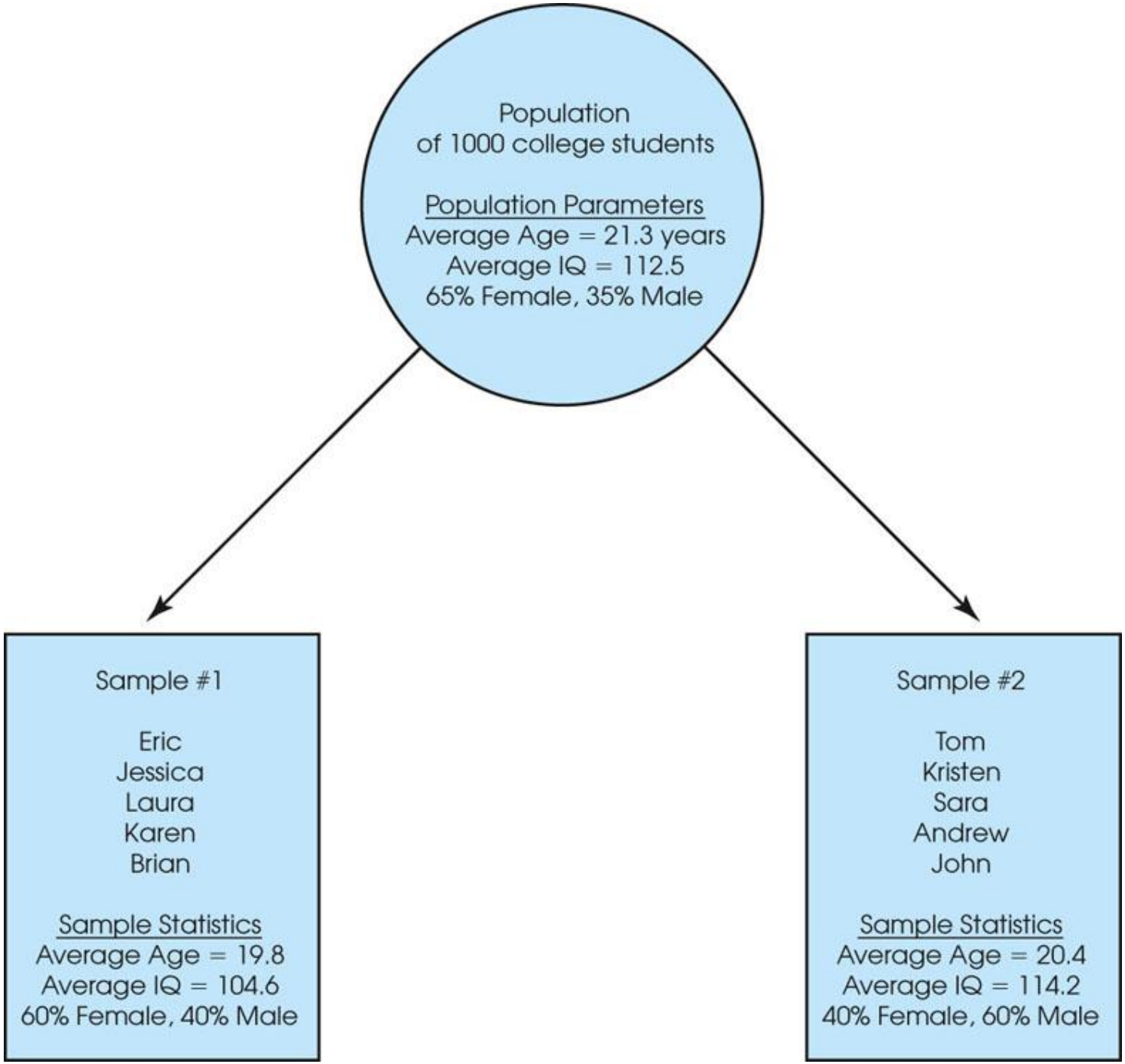

Variable #1: Time
(the quasi-independent variable)
Not manipulated, but used
to create two groups of scores
Variable #2: Depression scores
(the dependent variable)
Measured at each of the two
different times
Before Therapy
After Therapy
17 12
19 10
16 14
12 15
17 13
18 12
15 11
16 13
Any
difference?

Data

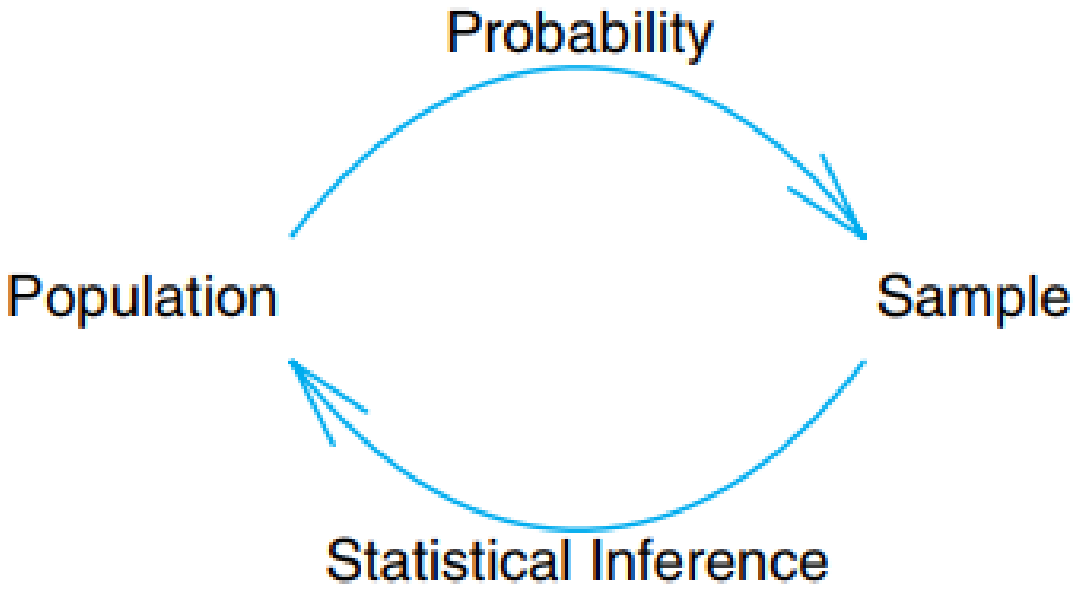
- The measurements obtained in a research study are called the **data**.
- The goal of statistics is to help researchers organize and interpret the data.

Sampling Error

- The discrepancy between a sample statistic and its population parameter is called **sampling error**.
- Defining and measuring sampling error is a large part of inferential statistics.


Population
of 1000 college students
Population Parameters
Average Age = 21.3 years
Average IQ = 112.5
65% Female, 35% Male
Sample #1
Eric
Jessica
Laura
Karen
Brian
Sample Statistics
Average Age = 19.8
Average IQ = 104.6
60% Female, 40% Male
Sample #2
Tom
Kristen
Sara
Andrew
John
Sample Statistics
Average Age = 20.4
Average IQ = 114.2
40% Female, 60% Male

Probability and Statistics (work together)



Any questions?