MATH1. Part II

Probability and Statistics

Credits for MATH1 and your UoE/ZJE degrees

- You require 25 ZJU credits to progress to Y2
- You require 120 UoE credits to progress to Y2

This course is worth	4 ZJU credits 20 UoE credits

- This course contributes **2.6%** (= 4/152.5 %) of the final ZJU GPA, or
- **2.9%** (= 4/138 %) to the final GPA for international students

- Your final UoE mark and classification is calculated from course marks in years 3 and 4
- Marks for this course will still be included in your UoE degree transcript

Timetable

Week #	Lecture	Lecture	Tutorial / Exam
5	10/11 Description of samples and populations - C1,2	10/12 Probability - C3	10/13 Midterm 1 (Thursday)
6	10/18 Distribution - C4	10/19 Distribution – C5	10/21 Tutorial 2-1
7	10/25 Confidence intervals - C6	10/26 Comparison of two samples - C7,8	10/28 Tutorial 2-2
8	11/1 Categorical data - C9	11/2 Categorical data - C10	11/4 Tutorial 2-3
9	11/8 Comparing the means of many independent samples - C11	11/9 Summary and review - C13	11/11 Tutorial 2-4
10	11/15 Limits	11/16 Limits	**11/18 Midterm 2**

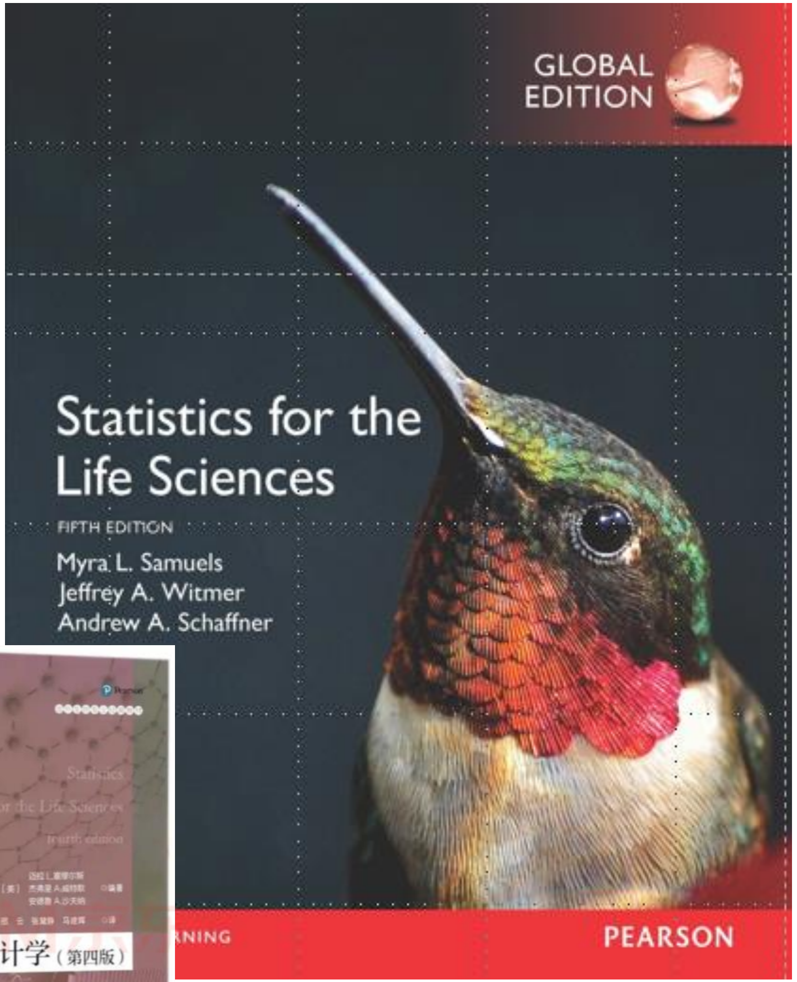
Textbook:

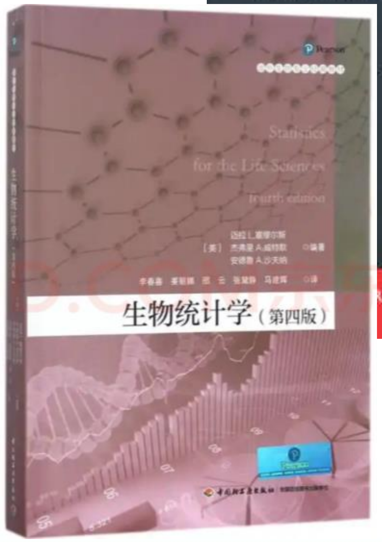
"Statistics for the Life Sciences", 5th ed., Pearson,

By M.L. Samuels, J.A. Witmer and A.A. Schaffner

Lecture notes & solutions:

Will be posted on Blackboard for review.

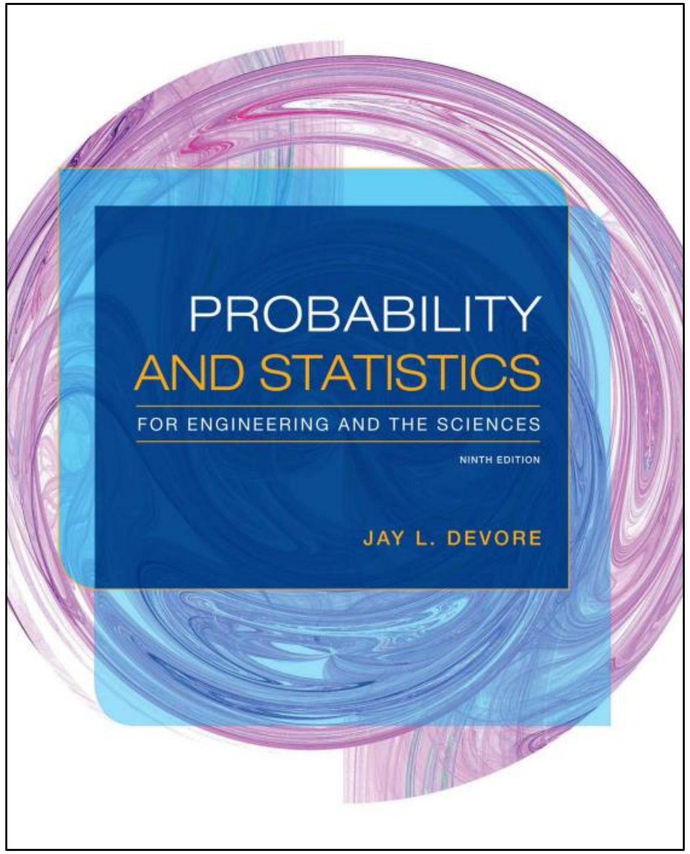




Pearson

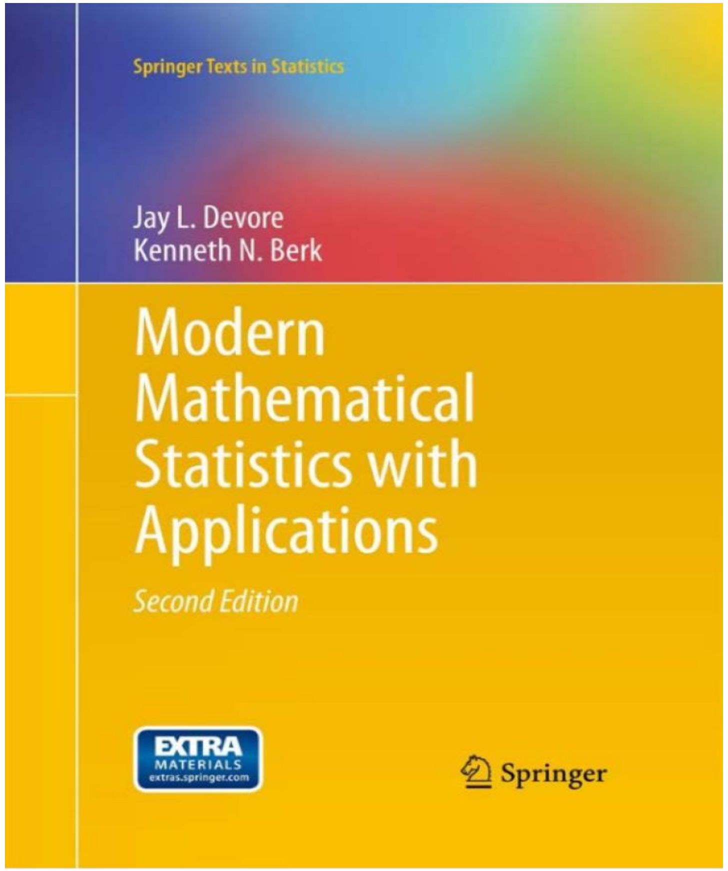
Reference books:

“Probability and Statistics for Engineering and the Sciences”, Ninth Edition, Cengage, By Jay L. Devore



Cengage

“Modern Mathematical Statistics with Applications”, 2th ed., Springer,
By Jay L. Devore, Kenneth N. Berk



Spinger

Chapter 1

Introduction

1.1 Statistics and the Life Sciences

History of Probability and Statistics

- Probability has its origin in the study of gambling and insurance in the 17th century.
- Statistics was developed in the early 19th century as the study of populations and economies.
- Probability theory and statistical methods now are indispensable tool of both social and natural sciences.

Blaise Pascal, an early pioneer on the mathematics of probability.

1.1 Statistics and the Life Sciences

Scope of Modern Statistics

These days statistical methodology is employed by investigators in virtually all disciplines, including such areas as

- **public health** (identifying sources of diseases and ways to treat them)
- **materials engineering** (studying properties of various treatments to retard corrosion)
- **marketing** (developing market surveys and strategies for marketing new products)
- **civil engineering** (assessing the effects of stress on structural elements and the impacts of traffic flows on communities)
- **molecular biology** (analysis of microarray data)
- **ecology** (describing quantitatively how individuals in various animal and plant populations are spatially distributed)

1.1 Statistics and the Life Sciences

- **Statistics** is the science of understanding data and of making decisions in the face of **variability** and uncertainty.
- The conclusions of a statistical data analysis are often stated in terms of **probability**.

Example 1.1.1 Vaccine for Anthrax

- Anthrax is a serious disease of sheep and cattle.
- A group of 24 sheep were vaccinated;
- another group of 24 unvaccinated sheep served as controls.
- Then, all 48 animals were inoculated with a virulent culture of anthrax bacillus.

Table 1.1.1 Response of sheep to anthrax

	Treatment	
Response	Vaccinated	Not vaccinated
Died of anthrax	0	24
Survived	24	0
Total	24	24
Percent survival	100%	0%

1.2 Types of Evidence

- **Anecdotal evidence:** a short story or an example of an interesting event.
- **Observational study:** In an observational study the researcher systematically collects data from subjects, but only as an observer and not as someone who is manipulating conditions.
- **Experiment:** In an experiment, researchers imposed the conditions on the subjects.
  - Often human subjects in experiments are given a **placebo**—an inert substance, such as a sugar pill. It is well known that people often exhibit a ***placebo response***; that is, they tend to respond favorably to any treatment, even if it is only inert.
  - In experiments on humans, particularly those that involve the use of placebos, **blinding** is often used. This means that the treatment assignment is kept secret from the experimental subject.
  - An experiment in which both the subjects and the persons making the evaluations of the response are blinded is called a **double-blind experiment**.

1.2 Types of Evidence

Types of Evidence

Example. Heart attack risk vs. taking aspirin

- An article in the New York Times (Jan. 27, 1987) reported that heart attack risk could be reduced by taking aspirin.
- If you want to check if the above conclusion is true or not, what type of evidence that you are going to collect?
- If you are going to use blinding when collecting evidence?

1.3 Random Sampling

- **Population:** The population consists of <u>all</u> subjects/animals/specimens/plants, and so on, of interest.
- **Sample**: Typically, we are unable to observe the entire population; therefore, we must be content with gathering data from a <u>subset</u> of the population, a **sample** of size n. From this sample we make inferences about the population as a whole.

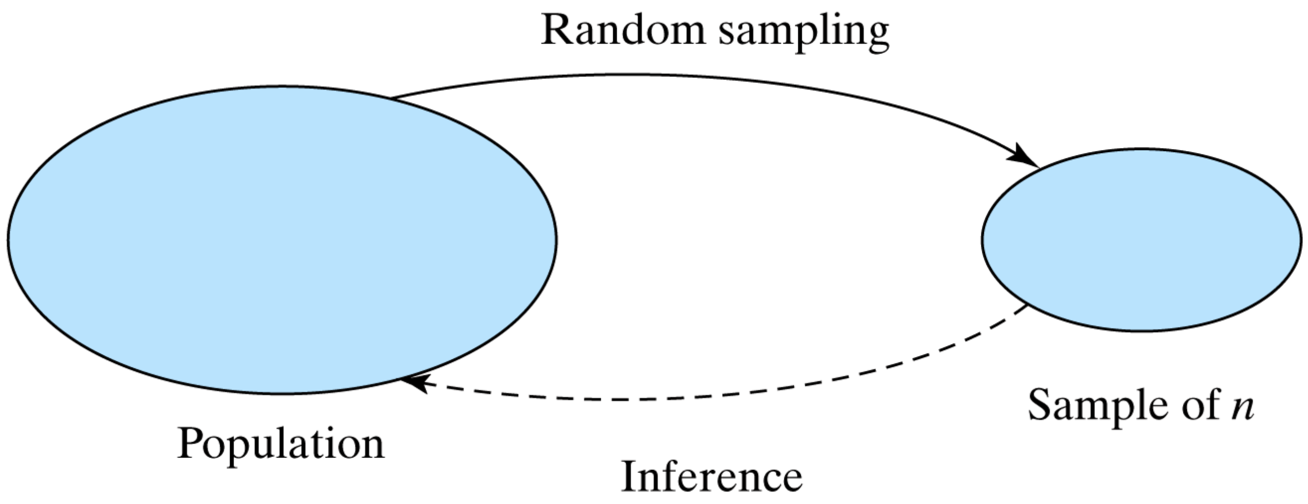


Figure 1.3.1 Sampling from a population

1.3 Random Sampling

- **Population:** The population consists of all subjects/animals/specimens/plants, and so on, of interest.
- **Sample:** Typically, we are unable to observe the entire population; therefore, we must be content with gathering data from a subset of the population, a **sample** of size n. From this sample we make inferences about the population as a whole.

Example 1.3.4 Lengths of Fish

- A biologist plans to study the distribution of body length of rock cod (*Epinephelus puscus*) in the Chesapeake Bay.
- What is the population?
- Give an example of the sample?

1.3 Random Sampling

We may define random sampling as follows:

- **A Simple Random Sample:** a simple random sample of n items is a sample in which
  - a) every member of the population has the same chance of being included in the sample,

  and
  - b) the members of the sample are chosen independently of each other. [Requirement (b) means that the chance of a given member of the population being chosen does not depend on which other members are chosen.]

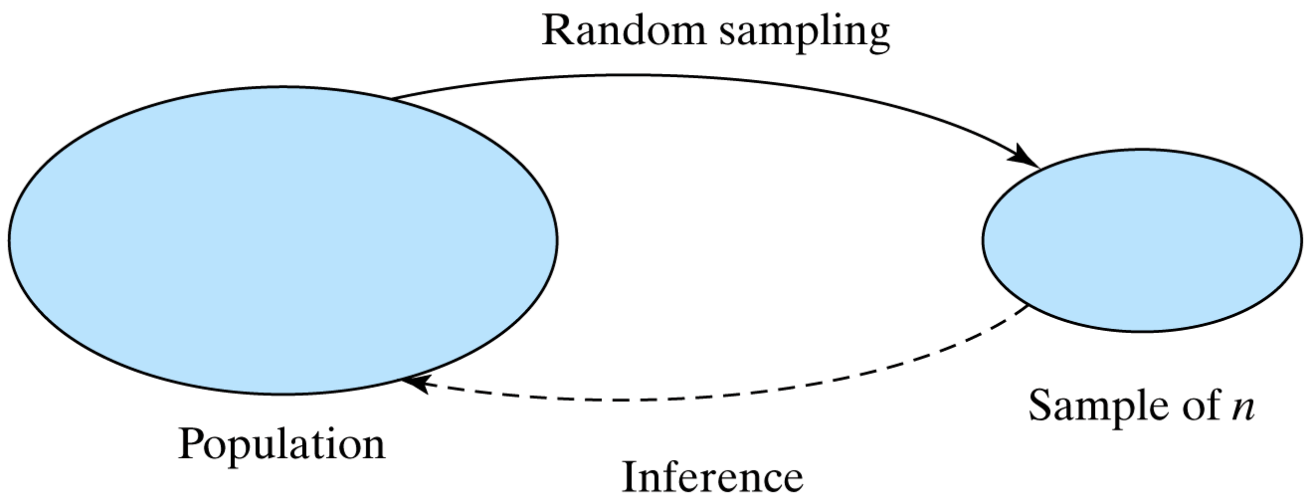


Figure 1.3.1 Sampling from a population

1.3 Random Sampling

Collecting Data

- **Sampling error:** The discrepancy between the sample and the population is called chance error due to sampling or sampling error.
  - When the sampling procedure is biased, the sample may not accurately represent the population, because it is systematically distorted.
- **Non-sampling error:** a non-sampling error is an error that is NOT caused by the sampling method.
  - Nonresponse bias, which is bias caused by persons not responding to some of the questions in a survey or not returning a written survey.
  - Missing data—that is, observations that were planned but could not be made.

1.3 Random Sampling

Collecting Data

Example 1.3.4 Lengths of Fish (continue)

- A biologist plans to study the distribution of body length of rock cod (*Epinephelus puscus*) in the Chesapeake Bay.
  - **Population:** all rock cod (*Epinephelus puscus*) in the Chesapeake Bay.
  - **Sample:** thirteen (n = 13) rock cod captured in the Chesapeake Bay.
  - **Variable:** body length.
- The sample will be collected using a fishing net.
- Will the above method induce sampling error?

Summary

Chapter 1

- 1.1 Statistics and the Life Sciences
- 1.2 Types of Evidence
- 1.3 Random Sampling