Mathematical Statistics

Zhang, Lixin and Dai, Jialing

Course Website:
www.math.zju.edu.cn/zlx/teaching.htm

Miscellaneous

- SAS, SPSS, R, S-plus, EViews, Excel, Matlab...

Miscellaneous

- SAS, SPSS, R, S-plus, EViews, Excel, Matlab...
- Prerequisite: Probability (3 credit units)

Miscellaneous

- SAS, SPSS, R, S-plus, EViews, Excel, Matlab...
- Prerequisite: Probability (3 credit units)
- Course syllabus and lecture slides:
  www.math.zju.edu.cn/zlx/teaching.htm

Miscellaneous

- SAS, SPSS, R, S-plus, EViews, Excel, Matlab...
- Prerequisite: Probability (3 credit units)
- Course syllabus and lecture slides: www.math.zju.edu.cn/zlx/teaching.htm
- Course components:

Miscellaneous

- SAS, SPSS, R, S-plus, EViews, Excel, Matlab...
- Prerequisite: Probability (3 credit units)
- Course syllabus and lecture slides: www.math.zju.edu.cn/zlx/teaching.htm
- Course components:
    1. weekly homework/project(s)/quizzes

Miscellaneous

- SAS, SPSS, R, S-plus, EViews, Excel, Matlab...
- Prerequisite: Probability (3 credit units)
- Course syllabus and lecture slides: www.math.zju.edu.cn/zlx/teaching.htm
- Course components:
  1. weekly homework/project(s)/quizzes
  2. Attendance

Miscellaneous

- SAS, SPSS, R, S-plus, EViews, Excel, Matlab...
- Prerequisite: Probability (3 credit units)
- Course syllabus and lecture slides: www.math.zju.edu.cn/zlx/teaching.htm
- Course components:
    1. weekly homework/project(s)/quizzes
    2. Attendance
    3. Final exam

- **Textbook**: "**数理统计**", 韦来生编著，科学出版社
- **References**:

❖ "**Statistical Inference**", by George Casella and Roger L. Berger and Duxbury Thomson Learning

❖ "**Theory of Point Estimation**", by E.L Lehmann and George Casella, Springer

❖``**数理统计** ", 峁诗松、王静龙著 ,华东师范大学出版社.

Topics

1. Statistics and Their Distributions
2. Parameter Estimation
   - Point Estimation
   - Interval Estimation
3. Hypothesis Testing
   - Parametric Hypothesis Testing
   - Non-Parametric Hypothesis Testing
   - Distribution Testing

Topics-Spring Quarter

1. Statistics and Their Distribution
2. Parameter Estimation
   - Point Estimation
   - Interval Estimation

First Project

Topics-Summer Quarter

Hypothesis Testing

1. Parametric Hypothesis Testing
2. Non-Parametric Hypothesis Testing
3. Distribution Testing

Mathematical Statistics

Chapter One. Introduction

Your company has created a new drug that may cure arthritis. How would you conduct a test to confirm the drug's effectiveness?

Your company has created a new drug that may cure arthritis. How would you conduct a test to confirm the drug's effectiveness?

The latest sales data have just come in, and your boss wants you to prepare a report for management on places where the company could improve its business. What should you look for? What should you not look for?

Your company has created a new drug that may cure arthritis. How would you conduct a test to confirm the drug's effectiveness?

The latest sales data have just come in, and your boss wants you to prepare a report for management on places where the company could improve its business. What should you look for? What should you not look for?

You want to conduct a poll on whether your school should use its funding to build a new athletic complex or a new library. How many people do you have to poll? How do you ensure that your poll is free of bias? How do you interpret your results?

Your company has created a new drug that may cure arthritis. How would you conduct a test to confirm the drug's effectiveness?

The latest sales data have just come in, and your boss wants you to prepare a report for management on places where the company could improve its business. What should you look for? What should you not look for?

You want to conduct a poll on whether your school should use its funding to build a new athletic complex or a new library. How many people do you have to poll? How do you ensure that your poll is free of bias? How do you interpret your results?

A widget maker in your factory that normally breaks 4 widgets for every 100 it produces has recently started breaking 5 widgets for every 100. When is it time to buy a new widget maker? (And just what is a widget, anyway?)

Your company has created a new drug that may cure arthritis. How would you conduct a test to confirm the drug's effectiveness?

The latest sales data have just come in, and your boss wants you to prepare a report for management on places where the company could improve its business. What should you look for? What should you not look for?

You want to conduct a poll on whether your school should use its funding to build a new athletic complex or a new library. How many people do you have to poll? How do you ensure that your poll is free of bias? How do you interpret your results?

A widget maker in your factory that normally breaks 4 widgets for every 100 it produces has recently started breaking 5 widgets for every 100. When is it time to buy a new widget maker? (And just what is a widget, anyway?)

These are some of the many real-world examples that require the use of statistics.

How would you approach the problem statements? There are some stepwise human algorithms, but is there a general problem statement?

How would you approach the problem statements? There are some stepwise human algorithms, but is there a general problem statement?

"Find possible solutions, decide on a solution, plan the solution, implement the solution, learn from the results for future solutions (or re-solution)."

How would you approach the problem statements? There are some stepwise human algorithms, but is there a general problem statement?

“Find possible solutions, decide on a solution, plan the solution, implement the solution, learn from the results for future solutions (or re-solution).”

“**SOAP** - **s**ubjective - the problem as given, **o**bjective - the problem after examination, **a**ssessment - the better defined problem, **p**lan - decide if guidelines to management already exist, and blueprint the solution for this case, or generate a risk-minimizing, new solution path”.

How would you approach the problem statements? There are some stepwise human algorithms, but is there a general problem statement?

"Find possible solutions, decide on a solution, plan the solution, implement the solution, learn from the results for future solutions (or re-solution)."

"**SOAP** - **s**ubjective - the problem as given, **o**bjective - the problem after examination, **a**ssessment - the better defined problem, **p**lan - decide if guidelines to management already exist, and blueprint the solution for this case, or generate a risk-minimizing, new solution path". "**HAMRC** - **h**ypothesis, **a**im,

methodology, **r**esults, **c**onclusion" - the concept that there is no real difference is the null hypothesis.

Then there is the joke that compares the different ways of thinking:

"A physicist, a chemist and a statistician were working collaboratively on a problem, when the wastepaper basket spontaneously combusted (they all sweared they had stopped smoking).

The chemist said, 'quick, we must reduce the concentration of the reactant which is oxygen, by increasing the relative concentration of non-reactive gases, such as carbon dioxide and carbon monoxide. Place a fire blanket over the flames. '

The physicist, interjected, 'no, no, we must reduce the heat energy available for activating combustion; get some water to douse the flame'.

Meanwhile, the statistician was running around lighting more fires. The others asked with alarm, 'what are you doing?'. 'Trying to get an adequate sample size' ."

Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty; and it thereby provides the navigation essential for controlling the course of scientific and societal advances.

http://www.amstat.org

In short,

Statistics is the study of the **collection**, **organization**, **analysis**, **interpretation**, and **presentation of data**.

In short,

Statistics is the study of the **collection**, **organization**, **analysis**, **interpretation**, and **presentation of data**.

It deals with all aspects of data including the planning of data collection in terms of the design of surveys and experiments.

Collecting $\Rightarrow$Organize$\Rightarrow$Analyzing $\Rightarrow$Making inference/predicting

§1.1 What is Mathematical Statistics?

Wikipedia:

Mathematical statistics is the study of statistics from a mathematical standpoint, using probability theory as well as other branches of mathematics such as linear algebra and analysis. The term "mathematical statistics" is closely related to the term "statistical theory" but also embraces modeling for actuarial science and non-statistical probability theory.

§1.1 What is Mathematical Statistics?

Wikipedia:

Mathematical statistics is the study of statistics from a mathematical standpoint, using probability theory as well as other branches of mathematics such as linear algebra and analysis. The term "mathematical statistics" is closely related to the term "statistical theory" but also embraces modeling for actuarial science and non-statistical probability theory.

Statistics deals with gaining information from data. In practice, data often contain some randomness or uncertainty. Statistics handles such data using methods of probability theory.

Statisticians apply statistical thinking and methods to a wide variety of scientific, social, and business endeavors in such areas as astronomy, biology, education, economics, engineering, genetics, marketing, medicine, psychology, public health, sports, among many. “The best thing about being a statistician is that you get to play in everyone else’s backyard.” (John Tukey, Bell Labs, Princeton University)

Statisticians apply statistical thinking and methods to a wide variety of scientific, social, and business endeavors in such areas as astronomy, biology, education, economics, engineering, genetics, marketing, medicine, psychology, public health, sports, among many. "The best thing about being a statistician is that you get to play in everyone else's backyard." (John Tukey, Bell Labs, Princeton University)

Many economic, social, political, and military decisions cannot be made without statistical techniques, such as the design of experiments to gain federal approval of a newly manufactured drug.

http://www.amstat.org

$$\text{Statistics} \begin{cases} \text{Mathematical Statistics} \\ \text{Biostatistics} \\ \text{Medical Statistics} \\ \text{Economic Statistics} \\ \text{Sociological Statistics} \\ \ldots \end{cases}$$

Mathematical Statistics studies the fundamental theory of statistics.

Random Data

Recall:
"Statistics is the study of the collection, organization, analysis, interpretation and presentation of **data**..."

Data is the key word.

Data in Statistics are random data.

Example

Example 1 To check whether all light bulbs produced meet the standard, a quality control engineer randomly select 10 light bulbs and test the lifetime (in hours) of each item. The measurements are presented as follows:

$$1980, \quad 2800, \quad 3060, \quad 4500, \quad 2760, \quad 3270, \quad 1560, \quad 0, \quad 3200, \quad 1940.$$

Notice that

1. The sample data are random, because ten light bulbs are randomly selected from a large number of light bulbs.
2. What we mean by "randomly selected " will become clear in a moment.
3. As small number as "10", the lifetimes of those ten light bulbs tell us, to certain degree, the lifetime distribution of the whole batch of light bulbs produced.

Example

Example 2 Quality control requires light bulbs last least 3000 hours. Any light bulb with lifetime less than 3000 hours is considered to be defective. In terms of defective or non-defective of each item, the sample data observed from the sample in Example 1 would be as follows:

Defective	Defective	Non-defective	Non-defective	Defective
Non-defective	Defective	Defective	Non-defective	Defective.

Remark: Notice that even though both set of data are random and observed from the sample, they contain different information of the sample. The data collected in Example 1 are different type of data collected in Example 2.

The data in Example 1 is called **numerical (quantitative)**, and the data in Example 2 is called **categorical (or qualitative)**.

Example

Example 3 A retail business plans to expand its business to a new city. Its first market research project is to investigate a few demographics of the city. A sample of 500 residents in the city are randomly selected, and their age, profession, education level, and annual income are recorded. The observations for the first five people are shown below.

OBS*	Age	Profession	Education	Annual Income (Thousand Yuan)
1	42	Government Official	High School	102.8
2	35	Worker	Middle School	62.6
3	50	Physician	College	99.2
4	47	Worker	Elementary School	65.6
5	36	Teacher	College	84.0
⋮	⋮	⋮	⋮	⋮

* "OBS" =Observations

There are four variables in this study.

There are four variables in this study.

Age and **Annual Income** are numerical variables;

There are four variables in this study.

Age and **Annual Income** are numerical variables;

Profession and **Level of Education** are categorical variables.

There are four variables in this study.

Age and **Annual Income** are numerical variables;

Profession and **Level of Education** are categorical variables.

Again the sample is a random sample, and the data are random data. The goal is to gain some insight of demographics of the city via those 500 people in the sample.

In general, for a study of a sample of size n and m variables, the data may be recorded in a $n \times m$ table:

Data Table

OBS	Variable 1	Variable 2	$\cdots$	Variable m
1	x_{11}	x_{12}	$\cdots$	x_{1m}
2	x_{21}	x_{22}	$\cdots$	x_{2m}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	x_{n1}	x_{n2}	$\cdots$	x_{nm}

Example

Example 4 A research project is interested in gathering information of heights of 4-year-old children in China. 500 four-year-old children are randomly **with replacement**, and their heights(in meters) are measured:

$$x_1, x_2, \cdots, x_n.$$

Example

Example 4 A research project is interested in gathering information of heights of 4-year-old children in China. 500 four-year-old children are randomly **with replacement**, and their heights(in meters) are measured:

$$x_1, x_2, \cdots, x_n.$$

The arithmetic mean of those 500 measurements yields

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = 1.05m.$$

What can you say about height of ALL 4-year-old children in China?

Example

Example 4 A research project is interested in gathering information of heights of 4-year-old children in China. 500 four-year-old children are randomly **with replacement**, and their heights(in meters) are measured:

$$x_1, x_2, \cdots, x_n.$$

The arithmetic mean of those 500 measurements yields

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = 1.05m.$$

What can you say about height of ALL 4-year-old children in China?

We claim that: *the average height of ALL 4-year-old children in China is about 1.05 meters.*

On what basis may we make such a claim?

On what basis may we make such a claim?

What questions do you have regarding about this claim?

Suppose the total number of four-year-old children in China is N. Let

$$H = \{h_1, h_2, \cdots, h_N\}$$

be their heights in meters.

To pick a random sample of 500 children from this population, we put N paper slips with names of four-year-old children in this population in a big box and mix them well. Then $n = 500$ paper slips are selected one by one **with replacement** and the heights of children in the sample are denoted by

$$X_1, X_2, \cdots, X_n.$$

Notice that $X_1, X_2, \cdots, X_n$ are **random variables** taking values in H.

The **observed values** of the sample are denoted by $x_1, x_2, \cdots, x_n$.

Notice that $X_1, X_2, \cdots, X_n$ are **random variables** taking values in H.

The **observed values** of the sample are denoted by $x_1, x_2, \cdots, x_n$.

Since the sample is selected **with replacement**, $X_1, X_2, \cdots, X_n$ are independent and identically distributed (**i.i.d** with the probability mass function (pmf):

$$P(X_k = h_j) = \frac{1}{N}, \quad j = 1, 2, \ldots, N.$$

The expected value of each X_i is

$$\mathbb{E}X_i = \frac{h_1 + h_2 + \cdots + h_N}{N} = \overline{h}$$

which is exactly the average height of ALL four-year-old children in China.

And it is easy to see that the expected value of

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is $\mathbb{E}\overline{X} = \overline{h}$.

And it is easy to see that the expected value of

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is $\mathbb{E}\overline{X} = \overline{h}$.

From the (strong) law of large number,

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \to \mathbb{E}\overline{X} = \overline{h} \quad a.s.$$

This implies that as long as the sample size n is large enough, $\overline{X}$ is approximately equal to $\overline{h}$.

The sample size $n = 500$ seems large enough, and one observed value of $\overline{X}$ is $\overline{x}$. Therefore

$$\overline{x} \approx \overline{h}.$$

And it is easy to see that the expected value of

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is $\mathbb{E}\overline{X} = \overline{h}$.

From the (strong) law of large number,

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \to \mathbb{E}\overline{X} = \overline{h} \quad a.s.$$

This implies that as long as the sample size n is large enough, $\overline{X}$ is approximately equal to $\overline{h}$.

The sample size $n = 500$ seems large enough, and one observed value of $\overline{X}$ is $\overline{x}$. Therefore

$$\overline{x} \approx \overline{h}.$$

Any questions?

Now we should ask:

- What is the error of estimation? $|\overline{x} - \overline{h}| \leq \epsilon$?
- How much confidence do we have about the estimation? $P\left(|\overline{X} - \overline{h}| \leq \epsilon\right) \geq ?$
- $\cdots$

Now we should ask:

- What is the error of estimation? $|\overline{x} - \overline{h}| \leq \epsilon$?
- How much confidence do we have about the estimation? $P\left(|\overline{X} - \overline{h}| \leq \epsilon\right) \geq ?$
- $\cdots$

Those are typical questions Statistics addresses.

Basic Terms in Statistics

Population

In Statistics, the collection of all objects under study H is call the **population**.

Basic Terms in Statistics

Population

In Statistics, the collection of all objects under study H is call the **population**.

Sample

A subset $X_1, X_2, \cdots, X_n$ of the population is called a sample of size on n.

Probability versus Statistics

Probability versus Statistics

- In Probability, given the distribution of the population H, we deduce the properties of the sample $X_1, X_2, \cdots, X_n$.
  Population$\Rightarrow$ Sample. **Deductive reasoning**.

Probability versus Statistics

- In Probability, given the distribution of the population H, we deduce the properties of the sample $X_1, X_2, \cdots, X_n$.
  Population$\Rightarrow$ Sample. **Deductive reasoning**.

Probability versus Statistics

- In Probability, given the distribution of the population H, we deduce the properties of the sample $X_1, X_2, \cdots, X_n$.
  Population$\Rightarrow$ Sample. **Deductive reasoning**.
- In Statistics, the population H is not completely known, we want to gain understanding of the population H via studying the sample $X_1, X_2, \cdots, X_n$.
  Sample $\Rightarrow$ Population. **Inductive reasoning**.

Probability versus Statistics

- In Probability, given the distribution of the population H, we deduce the properties of the sample $X_1, X_2, \cdots, X_n$.
  Population$\Rightarrow$ Sample. **Deductive reasoning**.
- In Statistics, the population H is not completely known, we want to gain understanding of the population H via studying the sample $X_1, X_2, \cdots, X_n$.
  Sample $\Rightarrow$ Population. **Inductive reasoning**.

Probability versus Statistics

- In Probability, given the distribution of the population H, we deduce the properties of the sample $X_1, X_2, \cdots, X_n$.
  Population$\Rightarrow$ Sample. **Deductive reasoning**.
- In Statistics, the population H is not completely known, we want to gain understanding of the population H via studying the sample $X_1, X_2, \cdots, X_n$.
  Sample $\Rightarrow$ Population. **Inductive reasoning**.
- Since a sample is just a subset of the population, a sample is not equal to the population. Results obtained from a sample cannot be generalized to population without any mistakes.

  How to control the mistakes in statistical inference is one of the most important questions Statistics investigates.

Difference between Mathematics and Statistics

- Mathematics exploits "deductive reasoning". In Mathematics, results are logically deducted from axioms, definitions, and known facts.
- Statistics uses more "inductive reasoning". In Statistics, conclusions are drawn inductively based upon what are observed from many individuals.

Therefore, Statistical reasoning is inductive reasoning, and indicative results are not 100% reliable. However, its reliability (or its confidence level on conclusions) can be measured by probability.

Relationship between Statistics and Other Disciplines

Statistics is a science that makes inferences about the possible rules of development of an object, based upon analyzing external data.

Statistics cannot explain the possible rules. The statistical results need to be carefully interpreted by one who understands the methods used as well as the subject matter.

Miscellaneous

- History, development, and Application of Statistics

Miscellaneous

- History, development, and Application of Statistics
- Probability vs. Statistics

History, Development, and Application of Statistics

According to Wikipedia,

- Statistical methods date back at least to the 5th century BC.

History, Development, and Application of Statistics

According to Wikipedia,

- Statistical methods date back at least to the 5th century BC.
- Some scholars pinpoint the origin of statistics to 1663, with the publication of "*Natural and Political Observations upon the Bills of Mortality*" by John Graunt.

History, Development, and Application of Statistics

According to Wikipedia,

- Statistical methods date back at least to the 5th century BC.
- Some scholars pinpoint the origin of statistics to 1663, with the publication of "*Natural and Political Observations upon the Bills of Mortality*" by John Graunt.
- Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data, hence it's stat-etymology.

History, Development, and Application of Statistics

According to Wikipedia,

- Statistical methods date back at least to the 5th century BC.
- Some scholars pinpoint the origin of statistics to 1663, with the publication of "*Natural and Political Observations upon the Bills of Mortality*" by John Graunt.
- Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data, hence it's stat-etymology.
- The scope of the discipline of statistics broadened in the early 19th century to include the collection and analysis of data in general. Today, statistics is widely employed in government, business, and natural and social sciences.

The modern field of statistics emerged in the late 19th and early 20th century in three stages:

- At the turn of the century, was led by the work of Sir Francis Galton and Karl Pearson, who transformed statistics into a rigorous mathematical discipline used for analysis, not just in science, but in industry and politics as well.

The modern field of statistics emerged in the late 19th and early 20th century in three stages:

- At the turn of the century, was led by the work of Sir Francis Galton and Karl Pearson, who transformed statistics into a rigorous mathematical discipline used for analysis, not just in science, but in industry and politics as well.
- The second wave of the 1910s and 20s was initiated by William Gosset, and reached its culmination in the insights of Sir Ronald Fisher, who wrote the textbooks that were to define the academic discipline in universities around the world.

- The final wave, which mainly saw the refinement and expansion of earlier developments, emerged from the collaborative work between Egon Pearson and Jerzy Neyman in the 1930s. They introduced the concepts of "Type II" error, power of a test and confidence intervals. Jerzy Neyman in 1934 showed that stratified random sampling was in general a better method of estimation than purposive (quota) sampling.

Statistical methods are applied in all fields that involve decision making, for making accurate inferences from a collated body of data and for making decisions in the face of uncertainty based on statistical methodology.

Statistical methods are applied in all fields that involve decision making, for making accurate inferences from a collated body of data and for making decisions in the face of uncertainty based on statistical methodology.

The use of modern computers has expedited large-scale statistical computational, and has also made possible new methods that are impractical to perform manually.

Figure: Karl Pearson, the founder of mathematical statistics.

Probability vs. Statistics

Statistics is closely related to probability theory.

Probability vs. Statistics

Statistics is closely related to probability theory.

The difference is, roughly, that probability theory starts from the given parameters of a total population to deduce probabilities that pertain to samples.

Probability vs. Statistics

Statistics is closely related to probability theory.
The difference is, roughly, that probability theory starts from the given parameters of a total population to deduce probabilities that pertain to samples.
Statistical inference, however, moves in the opposite directioninductively inferring from samples to the parameters of a larger or total population.

Probability vs. Statistics

Statistics is closely related to probability theory.

The difference is, roughly, that probability theory starts from the given parameters of a total population to deduce probabilities that pertain to samples.

Statistical inference, however, moves in the opposite directioninductively inferring from samples to the parameters of a larger or total population.

Statistics has many ties to machine learning and data mining.

§1.2 Fundamental Concepts in Statistics

Part I. Population and Its Distribution A population in a statistical study is the complete set of people or things being studied. Each element in the population is called individual.

§1.2 Fundamental Concepts in Statistics

Part I. Population and Its Distribution A population in a statistical study is the complete set of people or things being studied. Each element in the population is called individual.

The sample is the subset of the population from which the raw data are actually obtained.

$$\begin{array}{c}\text{Population}\\ \updownarrow\\ \text{All possible values of individuals in the population}\\ \updownarrow\\ \text{Random variable (or vector) and its distribution}\end{array}$$

The distribution of the variable X under study in population is called distribution of the population.

For example: In Example 4, H is the population, and the variable under consideration is the height (X) of four-year-old childern, which takes value in H.

The distribution of X is

$$F_N(x) = \frac{\#\{h_i : h_i < x\}}{N},$$

which contains all information of this population. Furthermore, any

randomly selected child from this population, his/her height X_i also follows this distribution.

Usually, the distribution of the population X is described by the relative frequency of each value in the population.

Usually, the distribution of the population X is described by the relative frequency of each value in the population.

The distribution of a randomly selected individual from the population has the same distribution as that of the population.

The distribution of the random variable X of the population is called the **population distribution** of the population.

Hence, we usually denote a population by a random variable X (vector $\mathbf{X}$), or by the distribution function $F(x)$ ($F(\mathbf{X})$) of the random variable X (vector $\mathbf{X}$).

Remark:

- When the population is a finite set (or the random variable takes finitely many values), the population is called a **finite population**.

Remark:

- When the population is a finite set (or the random variable takes finitely many values), the population is called a **finite population**.
- For the same individual in the population, different studies can be conducted.

Remark:

- When the population is a finite set (or the random variable takes finitely many values), the population is called a **finite population**.
- For the same individual in the population, different studies can be conducted.
- The primary goal of statistical analysis is not focus on the individuals, but on the population, that is, the population distribution.

Remark:

- When the population is a finite set (or the random variable takes finitely many values), the population is called a **finite population**.
- For the same individual in the population, different studies can be conducted.
- The primary goal of statistical analysis is not focus on the individuals, but on the population, that is, the population distribution.
- In fact, it is impossible to know the exact distribution of the population, unless we study all individuals in the postulation.

Part II. Distribution Family

Even the distribution of the population X is not completely known, we usually make some assumptions about the population distribution.

Part II. Distribution Family

Even the distribution of the population X is not completely known, we usually make some assumptions about the population distribution.

In Statistics, we usually assume that:

The distribution of the population X is from a certain distribution family, or X follows a certain type of the distribution.

For instance, in Example 4, we want to investigate the height of four-year-old children in China. The population

$$H = \{h_1, h_2, \cdots, h_N\}$$

is a finite population. The distribution of the population variable X is given by

$$F_N(x) = \frac{\#\{h_i : h_i < x\}}{N}.$$

For instance, in Example 4, we want to investigate the height of four-year-old children in China. The population

$$H = \{h_1, h_2, \cdots, h_N\}$$

is a finite population. The distribution of the population variable X is given by

$$F_N(x) = \frac{\#\{h_i : h_i < x\}}{N}.$$

Notice that this distribution function is a step function, which is hard to analyze mathematically.

For instance, in Example 4, we want to investigate the height of four-year-old children in China. The population

$$H = \{h_1, h_2, \cdots, h_N\}$$

is a finite population. The distribution of the population variable X is given by

$$F_N(x) = \frac{\#\{h_i : h_i < x\}}{N}.$$

Notice that this distribution function is a step function, which is hard to analyze mathematically.
When the number of individuals in the population is very large, the finite population may be approximately treated as a infinite population.

Practically, the population variable (X) for the height only takes non-negative values, but we can assume that X takes any real number $(-\infty, \infty)$. Past experience suggests that the height follows a normal (bell-shaped) distribution

$$N(\mu, \sigma^2).$$

That is, the distribution function of X is

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp\Big\{ -\frac{(y-\mu)^2}{2\sigma^2} \Big\} dy,$$

Here μ is the population mean, σ the population standard deviation, and both are unknown parameters in this example.

On what basis do we make those assumptions?

On what basis do we make those assumptions?

The assumptions are made based upon the following reasons:

- From the past experience, when N is large enough, $F_N(x) \approx F(x)$.
- The height is affected by many factors. We may assume that the population H under study is part of even larger population. Understanding the larger population helps understanding H.
- Mathematically, a random variable taking infinitely many values is more treatable.

In Example 4, we assume that the population variable X is a normal random variable and it follows a normal distribution $N(\mu, \sigma^2)$.

Once we know the true values of μ and σ we know the distribution of the height of all four-year-old children in China.

By assuming the type of distribution of X, statistical inference on parameters of the distribution of four-year-old children becomes making inference on parameters μ and σ.

When μ and σ are unknown, we do not know the actual distribution of X, but we know that its distribution belongs to the family:

$$\mathcal{F} = \{N(\mu, \sigma^2) : \mu \geq 0, \sigma > 0\}.$$

$\mathcal{F}$ is called the distribution family of the population.

In Example 1, the population consists of ALL light bulbs in that batch. Again the population is finite, but we treat it as a infinite population. The population variable under investigation is the lifetime X (in hours). It is reasonable to assume that X takes values in the interval $[0, \infty)$. The historical records indicate that the lifetime of light bulbs follows an exponential distribution

$$F(x) = 1 - e^{-\lambda x}, \qquad 0 \le x < \infty,$$

where $\lambda > 0$ is the unknown parameter, and $1/\lambda$ is the population mean.

The distribution of this population belongs to the family of exponential distribution:

$$\mathcal{F} = \{E(\lambda) : \lambda > 0\}.$$

Similarly, for the same batch of light bulbs in Example 2, instead of studying the lifetime of the light bulbs, we are interested whether the light bulbs are defective or not. The population variable X is no longer numerical, but categorical (or qualitative). However, we can code the categories of X by numerical values as follows

$$X = \begin{cases} 1, & \text{if the light bulb is defective} \\ 0, & \text{if the light bulb is non-defective.} \end{cases}$$

Suppose there are total of N light bulbs, among which there are M defective (M is unknown), then the defective rate for this batch is $p = M/N$. The distribution of X (the population) is

$$P(X = 0) = 1 - p\,; \qquad P(X = 1) = p.$$

where the defective rate p is a unknown parameter.

The family of distribution is

$$\mathcal{F} = \{b(1, p) : 0 < p < 1\}.$$

Example

An experimenter wants to measure a physical item μ. The measurement, a random variable X, may be any number in $(-\infty, +\infty)$. The all possible measurements $(-\infty, +\infty)$ constitutes the population. The population can be denoted by the random X.

As you might expect, the measurements are affected by many random factors. Experience indicates that the measurements

$$X = \mu + \epsilon$$

where ϵ denotes the random error of the measurement.

1. Usually, random measurement error $\epsilon \sim N(0, \sigma^2)$. Hence we may assume that X follows a normal distribution:

$$\mathcal{F}_1 = \{N(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma > 0\}.$$

1. Usually, random measurement error $\epsilon \sim N(0, \sigma^2)$. Hence we may assume that X follows a normal distribution:

$$\mathcal{F}_1 = \{N(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma > 0\}.$$

2. Suppose we also know σ^2 (i.e. σ_0^2) (say, we know how precise the measuring instrument is). Then the family of the distribution becomes even smaller

$$\mathcal{F}_2 = \{N(\mu, \sigma_0^2) : -\infty < \mu < \infty\}.$$

3. On the other hand, if we do not have much information regarding the distribution of random error ϵ, but we do know that it is continuous distribution or the second moment exists, then we would have a larger distribution family:

$$\mathcal{F}_3 = \{F(x) : F\text{continous distribution}\}$$

or

$$\mathcal{F}_4 = \{F(x) : F\text{the second moment exists}\}.$$

Parametric and Non-parametric Distribution Families:

Parametric and Non-parametric Distribution Families:

Distribution Family with Parameters: A distribution family with finitely many unknown parameters is denoted by

$$\mathcal{F} = \{F(x; \theta) : \theta \in \Theta\},$$

where θ stands for the unknown parameters (or vectors), and Θ the set of possible values of θ-it is called the **parameter space**.

Parametric and Non-parametric Distribution Families:

Distribution Family with Parameters: A distribution family with finitely many unknown parameters is denoted by

$$\mathcal{F} = \{F(x;\theta) : \theta \in \Theta\},$$

where θ stands for the unknown parameters (or vectors), and Θ the set of possible values of θ-it is called the **parameter space**.

Statistical inference based upon a distributions with parameters is called parametric statistical method.

Statistical inference based upon a distributions without parameters is called non-parametric statistical method.

Parametric and Non-parametric Distribution Families:

Distribution Family with Parameters: A distribution family with finitely many unknown parameters is denoted by

$$\mathcal{F} = \{F(x;\theta) : \theta \in \Theta\},$$

where θ stands for the unknown parameters (or vectors), and Θ the set of possible values of θ-it is called the **parameter space**.

Statistical inference based upon a distributions with parameters is called parametric statistical method.

Statistical inference based upon a distributions without parameters is called non-parametric statistical method.
Most commonly used distribution families are **normal distributions, binomial distributions, Poisson distributions, and exponential distributions, Γ distributions**, and etc.

Methods for studying a population:

- Census

Methods for studying a population:

- Census
- Sampling

Methods for studying a population:

- Census
- Sampling
- Random experiment

Part III. Sampling Methods

Recall: A sample is a subset of the population from which the raw data are actually obtained.

A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.

Part III. Sampling Methods

Recall: A sample is a subset of the population from which the raw data are actually obtained.

A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.

- Sampling Method: A sampling method is a process of choosing a group of individuals from the given population.

Part III. Sampling Methods

Recall: A sample is a subset of the population from which the raw data are actually obtained.

A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.

- Sampling Method: A sampling method is a process of choosing a group of individuals from the given population.
- Sample size: The number of individuals in the sample.

Part III. Sampling Methods

Recall: A sample is a subset of the population from which the raw data are actually obtained.

A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.

- Sampling Method: A sampling method is a process of choosing a group of individuals from the given population.
- Sample size: The number of individuals in the sample.
- For a random sample, before observations are made, the sample of size n is a random vector of size n.

Remark: Dual Properties of Samples

Remark: Dual Properties of Samples

Suppose a sample of size n are randomly selected from a population X. The sample is denoted by

$$X_1, X_2, ..., X_n.$$

Remark: Dual Properties of Samples

Suppose a sample of size n are randomly selected from a population X. The sample is denoted by

$$X_1, X_2, ..., X_n.$$

Before measurements are taken, $(X_1, X_2, ..., X_n)$ forms a n-dimensional random vector.

Remark: Dual Properties of Samples

Suppose a sample of size n are randomly selected from a population X. The sample is denoted by

$$X_1, X_2, ..., X_n.$$

Before measurements are taken, $(X_1, X_2, ..., X_n)$ forms a n-dimensional random vector.

Once the observations are made, the observed values are $x_1, x_2, ..., x_n$, and $(x_1, x_2, ..., x_n)$ is called a realized value of the n-dimensional random vector $(X_1, X_2, ..., X_n)$.

Remark: Dual Properties of Samples

Suppose a sample of size n are randomly selected from a population X. The sample is denoted by

$$X_1, X_2, ..., X_n.$$

Before measurements are taken, $(X_1, X_2, ..., X_n)$ forms a n-dimensional random vector.

Once the observations are made, the observed values are $x_1, x_2, ..., x_n$, and $(x_1, x_2, ..., x_n)$ is called a realized value of the n-dimensional random vector $(X_1, X_2, ..., X_n)$.

$(X_1, X_2, \cdots, X_n)$ is called a sample of size n, and $(x_1, x_2, ..., x_n)$ is called an observed sample value.

All possible values that the n-dimensional random vector $(X_1, X_2, ..., X_n)$ forms the "sample space"

$$\mathcal{H} = \{(x_1, x_2, ..., x_n) : x_i \in \mathbb{R},\ i = 1, 2, \cdots, n\}$$

Part IV Simple Random Samples

A sample can be drawn in many different ways.

Part IV Simple Random Samples

A sample can be drawn in many different ways. A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.

Simple Random Sample

The most commonly used sampling method is "**simple random sampling method**", which must satisfy the following two conditions:

- **Representability(randomness)**. Each individual in the population is equally likely being selected. This condition implies that each selected individual X_k and the population X share the same distribution.
- **Independence**. All individuals in the sample are independent. Thats is, $X_1, X_2 \cdots, X_n$ are independent.

In short, **Simple random sampling method** is a method that we choose a sample of n items in such a way that every subset of size n has an **equal chance** of being selected.

Definition of Simple Random Samples

Definition of Simple Random Samples

Definition

A collection of random variables $X_1, X_2 \cdots, X_n$ is called a **simple random sample (SRS) of size** n from a population X if

1. $X_1, X_2 \cdots, X_n$ are independent; and
2. $X_1, X_2 \cdots, X_n$ have the same distribution as the population X.

Definition of Simple Random Samples

Definition

A collection of random variables $X_1, X_2 \cdots, X_n$ is called a **simple random sample (SRS) of size** n from a population X if

1. $X_1, X_2 \cdots, X_n$ are independent; and
2. $X_1, X_2 \cdots, X_n$ have the same distribution as the population X.

A group of **independent identically distributed** random variables $X_1, X_2 \cdots, X_n$ are abbreviated to **iid random variables**, denoted by

$$X_1, X_2 \cdots, X_n \text{ i.i.d.} \sim F(x),$$

or

$$X_1, X_2 \cdots, X_n \text{ i.i.d.} \sim f(x),$$

or

$$X_1, X_2 \cdots, X_n \text{ i.i.d.} \sim X,$$

where $F(x)$, $f(x)$, X are the common distribution function

Suppose $F(x)$ is the distribution function of the population X. The joint distribution of a SRS of size n from this population is

$$F_n(x_1, x_2, \cdots, x_n) = F(x_1)F(x_2)\cdots F(x_n).$$

Remark: In this course, the distribution of a random variable X is defined as

$$F(x) = \mathrm{P}\{X < x\}, \quad x \in \mathbb{R},$$

which is a left-continuous function in $\mathbb{R}$.

For example, "**sampling with replacement**" produces simple random samples.

For example, “**sampling with replacement**” produces simple random samples.

When the **population is infinite**, selecting finite many items from the population will not change the distribution of the rest. Therefore, “**sampling without replacement**” also produces simple random samples.

For example, "**sampling with replacement**" produces simple random samples.

When the **population is infinite**, selecting finite many items from the population will not change the distribution of the rest. Therefore, "**sampling without replacement**" also produces simple random samples.

As for **finite population**, "**sampling without replacement**" will **NOT** produce simple random samples.

For example, "**sampling with replacement**" produces simple random samples.

When the **population is infinite**, selecting finite many items from the population will not change the distribution of the rest. Therefore, "**sampling without replacement**" also produces simple random samples.

As for **finite population**, "**sampling without replacement**" will **NOT** produce simple random samples.

However, when the population size N is significantly greater than the sample size n, the impact of "**sampling without replacement**" on the population distribution is negligible. In this case, samples obtained by "**sampling without replacement**" can be viewed **approximately simple random samples**.

Other Common Sampling Methods

1. **Systematic sampling**: We use a simple system to choose the sample, such as selecting every 10th or every 50th member of the population.
2. **Convenience sampling**: We use a sample that is convenient to select, such as people who happen to be in the same classroom.
3. **Stratified sampling**: We use this method when we are concerned about differences among subgroups, or *strata*, within a population. We first identify the subgroups and then draw a simple random sample within each subgroup. The total sample consists of all the samples from the individual subgroups.

Regardless of what type of sampling method is used, we should always keep the following two key ideas in mind:

- No matter how a sample is chosen, the study can be successful only if the sample is **representative** of the population.
- Even if a sample is chosen in the best possible way, it is still just a sample (as opposed to the entire population). Thus, we can never be sure that a sample is representative of the population. In general, a larger sample is more likely to be representative of the population, as long as it is chosen well.

In this course, almost all samples we work with are simple random samples. From now on, samples are simple random samples, unless otherwise stated.

Sampling Distribution Family and Statistical Model

If the distribution of a population belongs to a distribution family $\mathcal{F}$, then the distribution of the simple random sample

$$\mathbf{X} = (X_1, \ldots, X_n)$$

belong to the following family of distributions

$$\mathscr{F} = \{F(x_1) \cdots F(x_n) : F \in \mathcal{F}\}.$$

This family is called the **sampling distribution family**; it may also called the **statistical model**.

Since the distribution of a simple random sample is completely determined by the distribution of the population, sometimes the distribution family of the population is also called the statistical model.

While $(X_1, X_2, \cdots, X_n)$ is NOT a simple random sample, the distribution of the sample cannot be determined by the distribution of the population.

But the distribution of the sample still contains the information of the sample.

Therefore, the distribution of the sample is generally called the statistical model.

Part V Understand Population from Samples

In Statistics, the primary objective is to infer some unknown characteristics of the population from the sample characteristics.

Part V Understand Population from Samples

In Statistics, the primary objective is to infer some unknown characteristics of the population from the sample characteristics.

How can we make inference based upon samples? What is the basis for such practice? What theory validates such statistical inference?

Empirical Distribution Function

Definition

Definition 1.3.2 The empirical distribution function of a sample $(X_1, X_2, \cdots, X_n)$ is defined by

$$F_n(x) = \frac{1}{n}\sharp\{X_i : X_i < x, i = 1, ..., n\} \qquad \forall x \in \mathbb{R},$$

where $\sharp\{\cdot\}$ denotes the number of elements in set $\{\cdot\}$.

Alternatively, the empirical distribution function of a sample can be defined as follows:

Arrange the sample $X_1, X_2, \cdots, X_n$ in ascending order:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

Then

$$F_n(x) = \begin{cases} 0, & x \leq X_{(1)}, \\ k/n, & X_{(k)} < x \leq X_{(k+1)} \ \ (k = 1, 2, \cdots, n-1) \\ 1, & X_{(n)} < x. \end{cases}$$

For the empirical distribution function $F_n(x)$, we observe that:

A. For a given sample, $F_n(x)$ is a function of x, and it satisfies all properties of a distribution function. That is,

1. $F_n(x)$ is a nondecreasing, left-continuous function of x;
2. $\lim_{x\to-\infty} F_n(x) = 0, \quad \lim_{x\to\infty} F_n(x) = 1.$

For the empirical distribution function $F_n(x)$, we observe that:

A. For a given sample, $F_n(x)$ is a function of x, and it satisfies all properties of a distribution function. That is,

1. $F_n(x)$ is a nondecreasing, left-continuous function of x;
2. $\lim\limits_{x \to -\infty} F_n(x) = 0, \quad \lim\limits_{x \to \infty} F_n(x) = 1.$

B. For any given x, $F_n(x)$ is a function of the sample, and its value is uniquely determined by the sample values of $X_1, X_2, \cdots, X_n$.

The distribution of the population may be characterized by $F_n(x)$. To see how this works, let's express $F_n(x)$ in a different form

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i < x\}}.$$

Define $Y_i = I_{\{X_i < x\}}$, $i = 1, ..., n$, then $Y_i, i = 1, ..., n$ are **i.i.d.** random variables and

$$E(Y_i) = F(x), \quad Var(Y_i) = F(x)(1 - F(x)).$$

For **any given** x, by the strong law of large number, we have

$$P[\lim_{n\to\infty} F_n(x) = F(x)] = 1.$$

That is, for any given x, $F_n(x)$ converges to $F(x)$ in probability 1.

As a matter of fact, something stronger is true:

Theorem

Theorem 1.3.1 *(Glivenko) Let $X_1, ..., X_n$ be a simple random sample and i.i.d.$\sim F(x)$, and $F_n(x)$ the empirical distribution function of the sample. Then*

$$P(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0) = 1.$$

Remark:

Theorem 1.3.1 shows that: $F_n(x)$ converges uniformly to $F(x)$ for all $x \in \mathbb{R}$ in probability 1. This result is stronger than the result deduced from the strong law of large number.

Therefore, when the sample size n is large enough, $F_n(x)$ is a good-fit of the distribution function $F(x)$ of the population. Consequently, the population can be better understood by $F_n(x)$, since all information of the population is contained in $F(x)$.

Review: Modes of Convergence

Definition

Let $\{X_n\}_{n=1}^{\infty} = X_1, X_2, ...X_n, ...$ be a sequence of random variables, and X be a random variable. We define

- if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \varepsilon) = 0,$$

then $\{X_n\}_{n=1}^{\infty}$ converges to X **in probability** as $n \to \infty$.

Review: Modes of Convergence

Definition

Let $\{X_n\}_{n=1}^{\infty} = X_1, X_2, \ldots X_n, \ldots$ be a sequence of random variables, and X be a random variable. We define

- if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \varepsilon) = 0,$$

  then $\{X_n\}_{n=1}^{\infty}$ converges to X **in probability** as $n \to \infty$.

- Let $F_n(x), n = 1, 2, \ldots$ and $F(x)$ be distribution functions of $X_n, n = 1, 2, \ldots$ and X, respectively. If

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \text{for all } x \text{ at which } F(x) \text{ is continuous},$$

  then $\{X_n\}_{n=1}^{\infty}$ converges to X **in distribution** as $n \to \infty$.

- if,

$$P(\lim_{n\to\infty} X_n = X) = 1,$$

then $\{X_n\}_{n=1}^{\infty}$ converges to X **almost surely (a.s.)** as $n \to \infty$, or $\{X_n\}_{n=1}^{\infty}$ converges to X **in probability 1**.

- if,

$$P(\lim_{n\to\infty} X_n = X) = 1,$$

then $\{X_n\}_{n=1}^{\infty}$ converges to X **almost surely (a.s.)** as $n \to \infty$, or $\{X_n\}_{n=1}^{\infty}$ converges to X **in probability 1**.

- if $\mathbb{E}X_n^r < \infty$, $\mathbb{E}X^r < \infty$, and

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0,$$

then $\{X_n\}_{n=1}^{\infty}$ converges to X **in the rth moment** as $n \to \infty$.

Relation of Four Modes of Convergence

Almost surely $\Rightarrow$ In probability $\Rightarrow$ In distribution

1st moment convergence $\Rightarrow$ 2nd moment convergence $\Rightarrow$ In probability $\Rightarrow$ In distribution

Kolmogorov-Smirnov Distance

Definition

The Kolmogorov-Smirnov distance between two functions is defined as

$$D_n := ||F_n - F||_\infty = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|,$$

which is to say, we take the largest gap between the two functions at any point.

Theorem

Theorem 1.3.1 *(Glivenko) Let $X_1, ..., X_n$ be a simple random sample and i.i.d.$\sim F(x)$, and $F_n(x)$ the empirical distribution function of the sample. Then*

$$P(\lim_{n\to\infty} D_n = 0) = 1.$$

Proof of Theorem 1.3.1: First of all, we discretize x: for any given positive integer r, let $x_{r,k}$ be the smallest x satisfying the inequality:

$$F(x-0) = F(x) \leq \frac{k}{r} \leq F(x+0),$$

$k = 1, 2, \cdots, r$.

The Borel strong law of large number implies

$$P(\lim_{n\to\infty} F_n(x_{r,k}) = F(x_{r,k})) = 1.$$

Similarly, we have

$$P(\lim_{n\to\infty} F_n(x_{r,k}+0) = F(x_{r,k}+0)) = 1.$$

Define events

$$\begin{aligned}
A_k^r =& \{\lim_{n\to\infty} F_n(x_{r,k}) = F(x_{r,k})\},\\
B_k^r =& \{\lim_{n\to\infty} F_n(x_{r,k}+0) = F(x_{r,k}+0)\},\\
A^r =& \bigcap_{k=1}^{r}(A_k^r \bigcap B_k^r), \quad A = \bigcap_{r=1}^{\infty} A^r.
\end{aligned}$$

Then $P(A_k^r) = P(B_k^r) = 1$

and

$$A^r = \{\lim_{n\to\infty} \max_{1\le k\le r} \{\max(|F_n(x_{r,k}) - F(x_{r,k})|,$$
$$|F_n(x_{r,k}+0) - F(x_{r,k}+0|))\} = 0\},$$

Notice that

$$P(\overline{A^r}) = P(\bigcup_{k=1}^{r}(\overline{A_k^r} \cup \overline{B_k^r}))$$
$$\le \sum_{k=1}^{r}(P(\overline{A_k^r}) + P(\overline{B_k^r})) = 0$$

and

$$A^r = \{\lim_{n\to\infty} \max_{1\le k\le r} \{\max(|F_n(x_{r,k}) - F(x_{r,k})|, \\ |F_n(x_{r,k}+0) - F(x_{r,k}+0|))\} = 0\},$$

Notice that

$$P(\overline{A^r}) = P(\bigcup_{k=1}^{r}(\overline{A_k^r} \cup \overline{B_k^r})) \\ \le \sum_{k=1}^{r}(P(\overline{A_k^r}) + P(\overline{B_k^r})) = 0$$

Therefore

$$P(\overline{A}) = P(\bigcup_{r=1}^{\infty} \overline{A^r}) = \lim_{n\to\infty} P(\bigcup_{r=1}^{n} \overline{A^r}) \le \lim_{n\to\infty} \sum_{r=1}^{n} P(\overline{A^r}) = 0.$$

Hence
$$P(A) = 1.$$
Let
$$E = \{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\}.$$
It suffices to prove
$$A \subset E.$$

For any x satisfying $x_{r,k} < x \leq x_{r,k+1} (k = 1, 2, \cdots, r-1)$, we have

$$F_n(x_{r,k} + 0) \leq F_n(x) \leq F_n(x_{r,k+1}),$$

$$F(x_{r,k} + 0) \leq F(x) \leq F(x_{r,k+1}).$$

When $k = 1, 2, \cdots, r-1$

$$
\begin{aligned}
F_n(x) - F(x) \leq & F_n(x_{r,k+1}) - F(x_{r,k} + 0) \\
= & F_n(x_{r,k+1}) - F(x_{r,k+1}) + F(x_{r,k+1}) - F(x_{r,k} + 0) \\
\leq & \max_k |F_n(x_{r,k}) - F(x_{r,k})| + \frac{1}{r}.
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
F(x) - F_n(x) \leq & F(x_{r,k+1}) - F_n(x_{r,k} + 0) \\
\leq & \max_k |F(x_{r,k} + 0) - F_n(x_{r,k} + 0)| + \frac{1}{r}.
\end{aligned}
$$

In addition, we can similarly prove that when $x \le x_{r,1}$, we have

$$|F_n(x) - F(x)| \le |F_n(x_{r,1}) - F(x_{r,1})| + \frac{1}{r};$$

In addition, we can similarly prove that when $x \leq x_{r,1}$, we have

$$|F_n(x) - F(x)| \leq |F_n(x_{r,1}) - F(x_{r,1})| + \frac{1}{r};$$

when $x > x_{r,r}$, we have

$$|F_n(x) - F(x)| \leq |F_n(x_{r,r} + 0) - F(x_{r,r} + 0)| + \frac{1}{r}.$$

In addition, we can similarly prove that when $x \le x_{r,1}$, we have

$$|F_n(x) - F(x)| \le |F_n(x_{r,1}) - F(x_{r,1})| + \frac{1}{r};$$

when $x > x_{r,r}$, we have

$$|F_n(x) - F(x)| \le |F_n(x_{r,r}+0) - F(x_{r,r}+0)| + \frac{1}{r}.$$

Both results together yield:

$$\begin{aligned} &\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \\ \le &\max_{1\le k\le r} \{\max\{|F_n(x_{r,k}) - F(x_{r,k})|, |F_n(x_{r,k}+0) - F(x_{r,k}+0)|\}\} \\ &+ \frac{1}{r} \end{aligned}$$

Notice that if event A occurs, then event A^r occurs, hence

$$
\begin{aligned}
&\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \\
\leq &\lim_{n\to\infty} \max_{1\leq k\leq r} \{\max\{|F_n(x_{r,k}) - F(x_{r,k})|, \\
&\qquad |F_n(x_{r,k}+0) - F(x_{r,k}+0)|\}\} + \frac{1}{r} \\
= &0 + \frac{1}{r}.
\end{aligned}
$$

Since r is an arbitrary positive integer,

$$\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0.$$

$A \subset E$. This completes the proof. $\square$

§1.3 Data Reduction–Statistics

Roughly speaking, statistical analysis aims to understand the statistical characteristics of the population. In particular, in a statistical model with parameters, the task is to find the values of parameters in the model.

§1.3 Data Reduction–Statistics

Roughly speaking, statistical analysis aims to understand the statistical characteristics of the population. In particular, in a statistical model with parameters, the task is to find the values of parameters in the model.

Statistical analysis is based upon samples, since samples contain information of the population.

We use the information in a sample $X_1, X_2, \cdots, X_n$ to make inference about unknown parameters of the population.

If the sample size n is large, the observed sample values $x_1, x_2, \cdots, x_n$ is a long list of numbers that may be hard to analyze and interpret.

It is highly desirable to summarize the information in a sample by determining a few key features of the sample. This is usually done by computing statistics, function of the sample.

Statistics are the functions of the sample.
You may think of samples as the *raw material*, and statistics as the *products*.

Definition

Definition 1.3.1 Given a distribution family $\mathcal{F} = \{F\}$, and a simple random sample $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ from the family, a real-valued (Borel measurable) function $T(\mathbf{X})$ of the sample

$$T = T(\mathbf{X}) = T(X_1, X_2, \cdots, X_n),$$

independent of $\mathcal{F}$, is called a **statistic** of the distribution family $\mathcal{F} = \{F\}$.

Definition

Definition 1.3.1 Given a distribution family $\mathcal{F} = \{F\}$, and a simple random sample $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ from the family, a real-valued (Borel measurable) function $T(\mathbf{X})$ of the sample

$$T = T(\mathbf{X}) = T(X_1, X_2, \cdots, X_n),$$

independent of $\mathcal{F}$, is called a **statistic** of the distribution family $\mathcal{F} = \{F\}$.

In the case of that $\mathcal{F}$ contains parameter: $\{F(x;\theta) : \theta \in \Theta\}$, $T(\mathbf{X})$ needs to be independent of the unknown parameter θ as well.

Definition

Similarly, a real vector-valued (Borel measurable) function $\mathbf{T}(\mathbf{X})$

$$\mathbf{T} = \mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}), \cdots, T_k(\mathbf{X})),$$

independent of $\mathcal{F}$, is called a **a vector statistic** of the distribution family $\mathcal{F} = \{F\}$.

Sampling Distribution

Let $F_T(t)$ denote the distribution function of a statistic $T(\mathbf{X})$. Then

$$
\begin{aligned}
F_T(t) \quad &= \mathrm{P}\{T(\mathbf{X}) < t\} \\
&= \mathrm{P}\{(X_1, X_2, \cdots, X_n) \in B\} \\
&\qquad (\text{where} B = \{(x_1, x_2, \cdots, x_n) : T(x_1, x_2, \cdots, x_n) < t\}) \\
&= \int \cdots \int_B dF(x_1) dF(x_2) \cdots dF(x_n),
\end{aligned}
$$

where $F(x)$ is the distribution of the population.

Common Statistics

- **Sample mean**:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{——Population mean } \mu = \mathbb{E}X.$$

Common Statistics

- **Sample mean**:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{---Population mean } \mu = \mathbb{E}X.$$

- **Sample variance**:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{---Population variance } \sigma^2 = \mathrm{Var}(X).$$

Common Statistics

- **Sample mean**:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{——Population mean } \mu = \mathbb{E}X.$$

- **Sample variance**:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{——Population variance } \sigma^2 = \mathrm{Var}(X).$$

- **Sample k-th moment**:

$$a_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \text{———} \mathbb{E}X^k.$$

Common Statistics

- **Sample mean**:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{——Population mean } \mu = \mathbb{E}X.$$

- **Sample variance**:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{——Population variance } \sigma^2 = \mathrm{Var}(X).$$

- **Sample k-th moment**:

$$a_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \text{———} \mathbb{E}X^k.$$

- **Sample k-th central moment**:

$$m_{n,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \text{———} \mathbb{E}(X - \mathbb{E}X)^k.$$

- **Sample Coefficient of Variation**:

$$\frac{S_n}{\overline{X}}.$$

It contains the information of the population coefficient of variation

$$\frac{\sqrt{Var(X)}}{\mathbb{E}(X)}.$$

The coefficient of variation of X shows the extent of variability in relation to mean of the population.

- **Sample Coefficient of Variation**:

$$\frac{S_n}{\overline{X}}.$$

It contains the information of the population coefficient of variation

$$\frac{\sqrt{Var(X)}}{\mathbb{E}(X)}.$$

The coefficient of variation of X shows the extent of variability in relation to mean of the population.

- **Sample skewness**:

$$\frac{\sqrt{n}\sum_{i=1}^{n}(X_i-\overline{X})^3}{\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]^{3/2}}.$$

which estimates the population skewness

$$\frac{\mathbb{E}(X-\mathbb{E}X)^3}{\sigma^3}.$$

- **Sample kurtosis**:

$$\frac{n\sum_{i=1}^{n}(X_i-\overline{X})^4}{[\sum_{i=1}^{n}(X_i-\overline{X})^2]^2}-3.$$

It reflects the population kurtosis

$$\frac{\mathbb{E}(X-\mathbb{E}X)^4}{\sigma^4}-3.$$

- Empirical distribution function

- Empirical distribution function
- **Sample covariance of bivariate random variables**

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

- Empirical distribution function
- **Sample covariance of bivariate random variables**

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

- **Order statistics**–will be covered later.

Independence and Integrality of Statistics

Definition

Definition Let T_1 and T_2 be two statistics of a distribution family $\mathcal{F} = \{F\}$. If for each distribution F from the family, T_1 and T_2 are mutually independent,then it is said that the **two statistics are independent**.

Independence and Integrality of Statistics

Definition

Definition Let T_1 and T_2 be two statistics of a distribution family $\mathcal{F} = \{F\}$. If for each distribution F from the family, T_1 and T_2 are mutually independent,then it is said that the **two statistics are independent**.

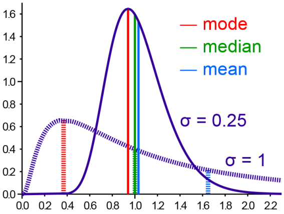
Definition

Definition Let T be a statistic of a distribution family $\mathcal{F} = \{F\}$. If for each distribution F from the family, T is integrable (i.e. the expected value of T exists), then **the statistic is called integrable**.

Population parameters

A **parameter** is a characteristic of a population. A **statistic** is a characteristic of a sample. Inferential statistics enables you to make an educated guess about a population parameter based on a statistic computed from a sample randomly drawn from that population

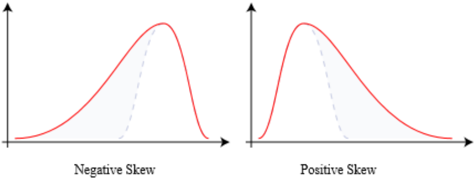
- Population mean: μ
- Population standard deviation: σ



Comparison of the arithmetic mean, median and mode of two skewed (log-normal) distributions.

- Population skewness:

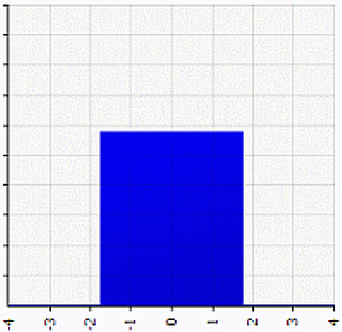
$$\gamma_1 = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{\sigma^3}$$



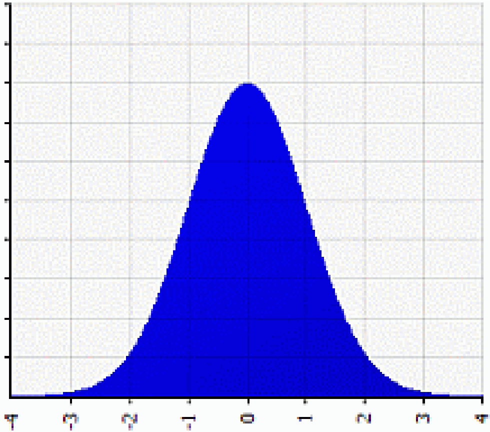
- Population kurtosis:

$$\text{Kurtosis:} \quad \beta_2 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{\sigma^4}$$

$$\text{Excess Kurtosis:} \quad \gamma_2 = \beta_2 - 3.$$

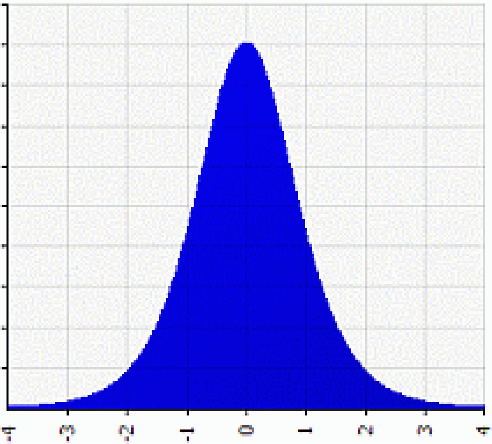


Uniform($\sqrt{3}, \sqrt{3}$)
$\beta_2 = 1.8, \gamma_2 = -1.2$



Normal$(\mu = 0, \sigma = 1)$

$\beta_2 = 3, \gamma_2 = 0$



Logistic($\alpha = 0, \beta = 0.55153$)

$\beta_2 = 4.2, \gamma_2 = 1.2$