Introduction

1 Introduction
- Course Materials
- Software
- References
- Getting Help

2 Statistics
- Defining Statistics
- Two Motivating Scenarios
- Sample Versus Population

Goals of This Course

1. Learn about simple distributions; basic probability; statistical concepts

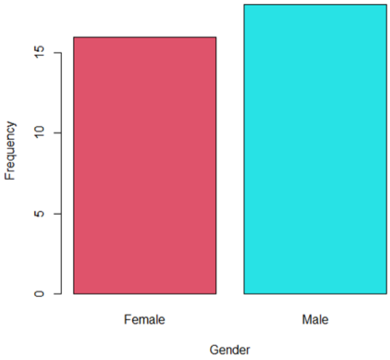
Goals of This Course

1. Learn about simple distributions; basic probability; statistical concepts

2. Know how to produce and to use statistical graphics

Goals of This Course

1. Learn about simple distributions; basic probability; statistical concepts
2. Know how to produce and to use statistical graphics



Goals of This Course

3. Know how to

Goals of This Course

3 Know how to

- determine an appropriate statistical model;

Goals of This Course

3 Know how to

- determine an appropriate statistical model;

- conduct sensible statistical inference;

Goals of This Course

3 Know how to

- determine an appropriate statistical model;

- conduct sensible statistical inference;

- understand statistical portions in a general report or a journal.

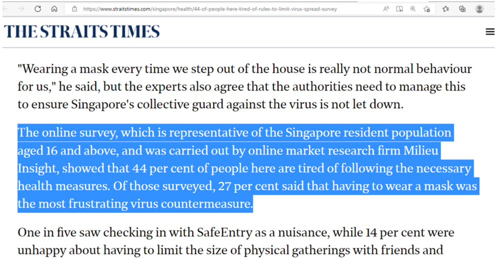
Goals of This Course

3 Know how to

- determine an appropriate statistical model;

- conduct sensible statistical inference;

- understand statistical portions in a general report or a journal.


https://www.straitstimes.com/singapore/health/44-of-people-here-tired-of-rules-to-limit-virus-spread-survey
THE STRAITS TIMES

"Wearing a mask every time we step out of the house is really not normal behaviour for us," he said, but the experts also agree that the authorities need to manage this to ensure Singapore's collective guard against the virus is not let down.

The online survey, which is representative of the Singapore resident population aged 16 and above, and was carried out by online market research firm Milieu Insight, showed that 44 per cent of people here are tired of following the necessary health measures. Of those surveyed, 27 per cent said that having to wear a mask was the most frustrating virus countermeasure.

One in five saw checking in with SafeEntry as a nuisance, while 14 per cent were unhappy about having to limit the size of physical gatherings with friends and

Goals of This Course

4. Know how to
    - analyse appropriately real-world datasets using R
    - solve some statistical questions

Goals of This Course

4. Know how to
    - analyse appropriately real-world datasets using R
    - solve some statistical questions

5. Have practical experience in
    - formulating statistical questions,
    - answering these questions
    - communicating the findings to a non-technical audience.

How We Shall Try To Achieve These Goals

- Formulas and mathematics are kept to a bare minimum.

How We Shall Try To Achieve These Goals

- Formulas and mathematics are kept to a bare minimum.

- Use R to do most of the computations.

How We Shall Try To Achieve These Goals

- Formulas and mathematics are kept to a bare minimum.
- Use R to do most of the computations.
- Focus on

How We Shall Try To Achieve These Goals

- Formulas and mathematics are kept to a bare minimum.

- Use R to do most of the computations.

- Focus on
  - understanding the logic behind statistical decisions,

How We Shall Try To Achieve These Goals

- Formulas and mathematics are kept to a bare minimum.
- Use R to do most of the computations.
- Focus on
  - understanding the logic behind statistical decisions,
  - and interpretation of output from the software.

1. Introduction
    - Course Materials
    - Software
    - References
    - Getting Help
2. Statistics
    - Defining Statistics
    - Two Motivating Scenarios
    - Sample Versus Population

Canvas

- All course materials (lecture notes, tutorial questions and tutorial solutions, datasets, etc.) will be uploaded to the workbin of Canvas.
  `canvas.nus.edu.sg`

Canvas

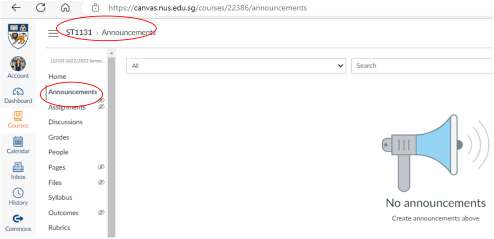
- All course materials (lecture notes, tutorial questions and tutorial solutions, datasets, etc.) will be uploaded to the workbin of Canvas. `canvas.nus.edu.sg`

- All announcements will be made through Canvas.

Canvas

- All course materials (lecture notes, tutorial questions and tutorial solutions, datasets, etc.) will be uploaded to the workbin of Canvas.
  `canvas.nus.edu.sg`

- All announcements will be made through Canvas.



https://canvas.nus.edu.sg/courses/22386/announcements

ST1131 › Announcements

No announcements

Create announcements above

Lectures

- There are pre-recorded lectures and live lectures.

Lectures

- There are pre-recorded lectures and live lectures.

- Before attending any live lecture, please finish all the previous lectures for better understanding.

Tutorials

- Tutorial slot registration is done online.

Tutorials

- Tutorial slot registration is done online.

- Each tutorial slot has 25 - 30 students divided into 5-6 groups.

Tutorials

- Tutorial slot registration is done online.

- Each tutorial slot has 25 - 30 students divided into 5-6 groups.

- Tutorials will be conducted physically by Ms Wong Yean Ling.

1. Introduction
   - Course Materials
   - Software
   - References
   - Getting Help
2. Statistics
   - Defining Statistics
   - Two Motivating Scenarios
   - Sample Versus Population

R

- R is free. Either RGui or RStudio is accepted.

R

- 4.1.0 is the oldest version that is accepted.

R

- 4.1.0 is the oldest version that is accepted.
- The test/exam and the quizzes will require the use of R and will contain R output.

R

- 4.1.0 is the oldest version that is accepted.

- The test/exam and the quizzes will require the use of R and will contain R output.

- You **will be tested** on how to use R to produce the output (numerical/graphical).

1. Introduction
   - Course Materials
   - Software
   - References
   - Getting Help
2. Statistics
   - Defining Statistics
   - Two Motivating Scenarios
   - Sample Versus Population

Recommended Book

- Recommended book to read:
  Statistics: The Art and Science of Learning from Data
  4th edition
  Authors: Agresti, Franklin and Klingenberg.

Recommended Book

- Recommended book to read:
  Statistics: The Art and Science of Learning from Data
  4th edition
  Authors: Agresti, Franklin and Klingenberg.

- Two thirds of the topics follow closely to the content of this book.

1. Introduction
    - Course Materials
    - Software
    - References
    - Getting Help
2. Statistics
    - Defining Statistics
    - Two Motivating Scenarios
    - Sample Versus Population

Asking Questions

- Check the FAQs folder on Canvas.

Asking Questions

- Check the FAQs folder on Canvas.
- Post it on Forum of Canvas.

Asking Questions

- Check the FAQs folder on Canvas.
- Post it on Forum of Canvas.
- Send email to the lecturer or the tutor or your student tutor. Put “ST1131” somewhere in the title of your email.

Asking Questions

- Check the FAQs folder on Canvas.
- Post it on Forum of Canvas.
- Send email to the lecturer or the tutor or your student tutor. Put "ST1131" somewhere in the title of your email.
- Book a time slot for consultation.

Facilitators and Emails

- Lecturer: Ms Daisy Pham, staptkc@nus,edu.sg

Facilitators and Emails

- Lecturer: Ms Daisy Pham, staptkc@nus,edu.sg

- Tutor: Ms Wong Yean Ling, stawyl@nus.edu.sg

Facilitators and Emails

- Lecturer: Ms Daisy Pham, staptkc@nus,edu.sg

- Tutor: Ms Wong Yean Ling, stawyl@nus.edu.sg

- Student tutor: each is to support two groups of 5-6 students each.

Topics

1. Introduction to R programming
2. Exploratory Data Analysis (EDA)
3. Collecting Data
4. Probability
5. Random Variables (Discrete and Continuous Random Variables)
6. Sampling Distribution
7. Statistical Estimation
8. Hypothesis Testing
9. Linear Regression
10. Some Limitations of Linear Regression (extra topic if time permits)

1. Introduction
    - Course Materials
    - Software
    - References
    - Getting Help

2. Statistics
    - Defining Statistics
    - Two Motivating Scenarios
    - Sample Versus Population

1. Introduction
    - Course Materials
    - Software
    - References
    - Getting Help
2. Statistics
    - Defining Statistics
    - Two Motivating Scenarios
    - Sample Versus Population

How Statistics is Used

Statistics is used in almost all domains today.

- In business managers analyse the results of marketing studies about new products, to help predict the sales.

How Statistics is Used

Statistics is used in almost all domains today.

- In business managers analyse the results of marketing studies about new products, to help predict the sales.
- In finance, people study stock returns and investment opportunities.

How Statistics is Used

Statistics is used in almost all domains today.

- In business managers analyse the results of marketing studies about new products, to help predict the sales.
- In finance, people study stock returns and investment opportunities.
- In medicine, people evaluate if a new drug is better than an existing one.

How Statistics is Used

Statistics is used in almost all domains today.

- In business managers analyse the results of marketing studies about new products, to help predict the sales.
- In finance, people study stock returns and investment opportunities.
- In medicine, people evaluate if a new drug is better than an existing one.
- Even if you never use statistics in your job, it's important to understand statistics. Why?

The Heart of Statistics

- In examples above, statistics is being used to answer specific questions, or to make a decision.

The Heart of Statistics

- In examples above, statistics is being used to answer specific questions, or to make a decision.

    - Will customers buy this new products?

The Heart of Statistics

- In examples above, statistics is being used to answer specific questions, or to make a decision.
  - Will customers buy this new products?
  - Does smoking lead to lung cancer?

The Heart of Statistics

- Statistics needs information, typically gathered from experiments or surveys.

The Heart of Statistics

- Statistics needs information, typically gathered from experiments or surveys.

- The information that we gather is collectively called **data**.

The Heart of Statistics

- Statistics needs information, typically gathered from experiments or surveys.
- The information that we gather is collectively called **data**.
  - Smoking habits of patients, whether they developed lung cancer or not, age, gender, etc.

The Heart of Statistics

- Statistics needs information, typically gathered from experiments or surveys.
- The information that we gather is collectively called **data**.
    - Smoking habits of patients, whether they developed lung cancer or not, age, gender, etc.
    - A poll of randomly selected customers from Starbucks for example, on whether they like the new flavor of coffee or not.

What is Statistics?

Definition 1 (Defining Statistics)

Statistics is the art and science of designing studies and analyzing data that those studies produce. Its ultimate goal is to translate data into knowledge, that allows us to make informed and objective decisions.

Using Statistics

Overall, there are 4 steps to investigating questions using statistics:

- Formulate a *Statistical Question*.

Using Statistics

Overall, there are 4 steps to investigating questions using statistics:

- Formulate a *Statistical Question*.
- Collect data. *Design*

Using Statistics

Overall, there are 4 steps to investigating questions using statistics:

- Formulate a *Statistical Question*.
- Collect data. *Design*
- Analyze data. *Description*

Using Statistics

Overall, there are 4 steps to investigating questions using statistics:

- Formulate a *Statistical Question*.
- Collect data. *Design*
- Analyze data. *Description*
- Interpret results. *Inference*

1. Introduction
   - Course Materials
   - Software
   - References
   - Getting Help
2. Statistics
   - Defining Statistics
   - Two Motivating Scenarios
   - Sample Versus Population

Predicting Price

Example 1 (Predicting Resale-flat Price)

- Suppose we are interested in what factors that affect the price of a HDB resale-flat in Singapore.

Predicting Price

Example 1 (Predicting Resale-flat Price)

- Suppose we are interested in what factors that affect the price of a HDB resale-flat in Singapore.
- Data could be retrieved from HDB's website. `https://data.gov.sg/dataset/resale-flat-prices`

Predicting Price

Example 1 (Predicting Resale-flat Price)

- Suppose we are interested in what factors that affect the price of a HDB resale-flat in Singapore.
- Data could be retrieved from HDB's website.
  `https://data.gov.sg/dataset/resale-flat-prices`
- Then, based on this data, we not only could use statistics to predict the price but we also could make some inferences about the price.

Design, Describe and Infer in Example 1

- Full data of all the resale-flats that were sold from Jan 1990 until the current month.

Design, Describe and Infer in Example 1

- Full data of all the resale-flats that were sold from Jan 1990 until the current month.

- Data collection (*Design* step) is the skipped.

Design, Describe and Infer in Example 1

- Full data of all the resale-flats that were sold from Jan 1990 until the current month.

- Data collection (*Design* step) is the skipped.

- Focus on how to analyze the data (*Describe*)

Design, Describe and Infer in Example 1

- Full data of all the resale-flats that were sold from Jan 1990 until the current month.

- Data collection (*Design* step) is the skipped.

- Focus on how to analyze the data (*Describe*)

- Produce the estimation and/or statistical statements (*Inference*).

Aspirin and Heart Disease

Example 2 (Does Aspirin Reduce Heart Disease?)

- Heart disease is the most common cause of death in industrialized nations. Does regular aspirin intake reduce death from heart attacks?

[a] *Statistics: The Art and Science of Learning from Data, 4th edition*, Agresti, Franklin and Klingenberg

Aspirin and Heart Disease

Example 2 (Does Aspirin Reduce Heart Disease?)

- Heart disease is the most common cause of death in industrialized nations. Does regular aspirin intake reduce death from heart attacks?
- Harvard Medical School conducted a study and found that, of those who took aspirin, 0.9% had heart attacks during the study. Of those who did not take aspirin, the percentage was 1.7%.[a]

[a] *Statistics: The Art and Science of Learning from Data, 4th edition*, Agresti, Franklin and Klingenberg

Aspirin and Heart Disease

Example 2 (Does Aspirin Reduce Heart Disease?)

- Heart disease is the most common cause of death in industrialized nations. Does regular aspirin intake reduce death from heart attacks?
- Harvard Medical School conducted a study and found that, of those who took aspirin, 0.9% had heart attacks during the study. Of those who did not take aspirin, the percentage was 1.7%.
  [a]
- Is this sufficient evidence for the benefit of aspirin in preventing heart attacks?

[a] *Statistics: The Art and Science of Learning from Data, 4th edition*, Agresti, Franklin and Klingenberg

Design, Describe and Infer in Example 2

- Design step: must consider who to include in the study, how to assign aspirin or a placebo, how long to follow them, etc.

Design, Describe and Infer in Example 2

- Design step: must consider who to include in the study, how to assign aspirin or a placebo, how long to follow them, etc.

- Description step: use the percentage of individuals who contracted heart disease in two groups.

Design, Describe and Infer in Example 2

- Design step: must consider who to include in the study, how to assign aspirin or a placebo, how long to follow them, etc.

- Description step: use the percentage of individuals who contracted heart disease in two groups.

- Inference step: can we infer that in *general*, aspirin can reduce the chances of heart disease?

1. Introduction
   - Course Materials
   - Software
   - References
   - Getting Help
2. Statistics
   - Defining Statistics
   - Two Motivating Scenarios
   - Sample Versus Population

Our Interest is in Population

Definition 2

The **population** is the total set of subjects in which we are interested. A **sample** is the subset of the population for whom we have, or plan to have, data for. This subset is often randomly selected.

Our Interest is in Population

Definition 2

The **population** is the total set of subjects in which we are interested. A **sample** is the subset of the population for whom we have, or plan to have, data for. This subset is often randomly selected.

- Resale-flat example: the population is all the HDB resale-flats in Singapore. Sample is all flats in the provided data.
- Aspirin example: If the sample has US people with age from 35–65 then the results are for population of US people aged 35 – 65.

Our Interest is in Population

Definition 2

The **population** is the total set of subjects in which we are interested. A **sample** is the subset of the population for whom we have, or plan to have, data for. This subset is often randomly selected.

- Resale-flat example: the population is all the HDB resale-flats in Singapore. Sample is all flats in the provided data.
- Aspirin example: If the sample has US people with age from 35–65 then the results are for population of US people aged 35 – 65.
- How we select our sample affects what population we can generalize the results to.

Parameter and Statistic

- A **parameter** is a numerical summary of the population. It is unknown.

Parameter and Statistic

- A **parameter** is a numerical summary of the population. It is unknown.

- A **statistic** is a summary of a sample taken from the population. We compute it based on the data in our sample.

Parameter and Statistic

- A **parameter** is a numerical summary of the population. It is unknown.
- A **statistic** is a summary of a sample taken from the population. We compute it based on the data in our sample.
- Our goal is to use statistics to make a conclusion about a population parameter.

Parameter and Statistic

- A **parameter** is a numerical summary of the population. It is unknown.
- A **statistic** is a summary of a sample taken from the population. We compute it based on the data in our sample.
- Our goal is to use statistics to make a conclusion about a population parameter.
- When doing so, we have to take into account the randomness in our sample.

Randomness and Variability

- It is crucial that the sample is representative of the population.

Randomness and Variability

- It is crucial that the sample is representative of the population.
- One way we can do this is to ensure that every subject has an equal chance of being selected.

Randomness and Variability

- It is crucial that the sample is representative of the population.
- One way we can do this is to ensure that every subject has an equal chance of being selected.
- This is known as *random sampling*.

Randomness and Variability

- It is crucial that the sample is representative of the population.
- One way we can do this is to ensure that every subject has an equal chance of being selected.
- This is known as *random sampling*.
  - It allows for powerful inferences about the population.

Randomness and Variability

- It is crucial that the sample is representative of the population.
- One way we can do this is to ensure that every subject has an equal chance of being selected.
- This is known as *random sampling*.
    - It allows for powerful inferences about the population.
    - It is crucial for performing experiments as well.

Randomness and Variability

- It is crucial that the sample is representative of the population.
- One way we can do this is to ensure that every subject has an equal chance of being selected.
- This is known as *random sampling*.
  - It allows for powerful inferences about the population.
  - It is crucial for performing experiments as well.
- Every time we take a random sample of subjects, it will vary. Hence the statistics will change as well.