

# Introduction to Statistics



Berlin Chen  
Department of Computer Science & Information Engineering  
National Taiwan Normal University

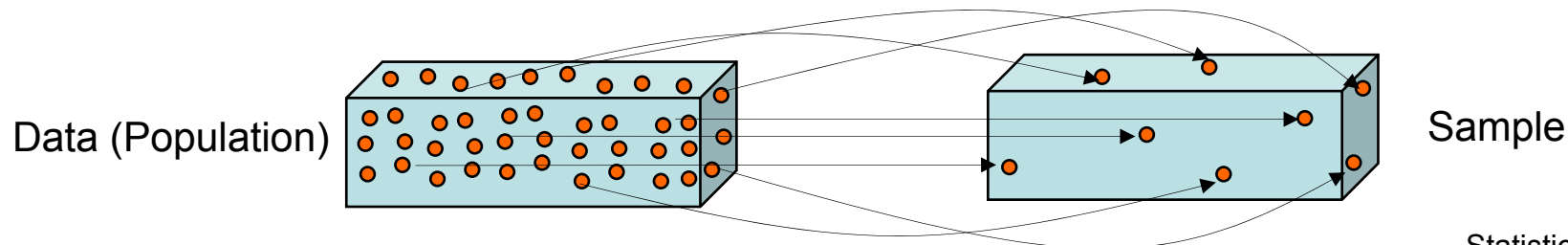


# What is Statistics?

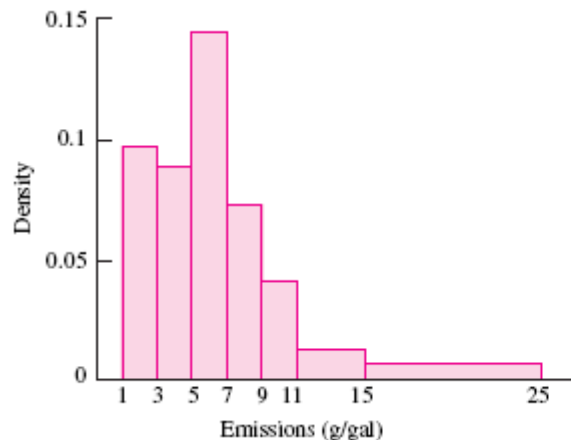
- Statistics is the field of study concerned with the collection, analysis, and interpretation (making decisions on) of uncertain data
  - E.g., the explanation of social or economic trends through the analysis of data
- Or, in more common usage, statistics refers to numerical facts of the data
  - E.g., the age of a student, the allowance of a student, the height of a student, etc.

# Types of Statistics (1/3)

- Broadly speaking, statistics can be divided into two areas
  - Descriptive statistics
  - Inferential statistics
- Descriptive Statistics
  - Concerned with the methods of collecting data and of summarizing clearly the basic information they contain
  - Collecting data refers to sampling, i.e., choosing a subset of data (a sample)
  - Summarizing data refers to organizing, displaying, and describing data by tables, graphs, and summary measures



# Types of Statistics (2/3)



Class Interval (g/gal)	Frequency	Relative Frequency	Density
1 - < 3	12	0.193	0.0965
3 - < 5	11	0.178	0.0890
5 - < 7	18	0.290	0.1450
7 - < 9	9	0.146	0.0730
9 - < 11	5	0.082	0.0410
11 - < 15	3	0.048	0.0120
15 - < 25	4	0.063	0.0063

- Frequency table for PM emissions of 62 vehicles driven at high altitude
- Inferential statistics
  - Concerned with the methods that use sample results to help make decisions or predictions about the data (population)
  - Or, the methods that draw conclusions from the data

# Types of Statistics (3/3)

The screenshot shows the Amazon.com search results for 'machine learning'. The page is divided into several sections:

- So You'd Like to... Offer your advice:** A recommendation for 'Learn more about Artificial Intelligence (AI) and Games' by John David Funge.
- You may also like:** A red dashed circle highlights the book 'Pattern Classification (2nd)' by Richard O. Duda, et al. A blue arrow points from this section to the 'Customers who searched for machine learning ultimately chose:' section.
- Customers who searched for machine learning ultimately chose:** A list of three books:
  - Machine Learning** -- by Tom M. Mitchell; Hardcover (Rate it) Buy new: \$133.75 -- Used & new from: \$59.90
  - Introduction to Machine Learning (Adaptive Computation and Machine Learning)** -- by Ethem Alpaydin; Hardcover (Rate it) Buy new: \$39.00 -- Used & new from: \$36.00
  - The Elements of Statistical Learning** -- by T. Hastie, et al; Hardcover (Rate it) Buy new: \$63.71 -- Used & new from: \$61.45
- 1 - 10 of 866 results for machine learning:** A search filter set to 'Relevance' and a 'GO!' button. The results list:
  - Introduction to Machine Learning (Adaptive Computation and Machine Learning)** by Ethem Alpaydin (Hardcover - October 1, 2004) Avg. Customer Rating: ★★★★★ (Rate this item) Usually ships in 24 hours List Price: ~~\$60.00~~ Buy new: \$39.00 Used & new from \$36.00
  - Machine Learning** by Tom M. Mitchell (Hardcover - March 1, 1997) Avg. Customer Rating: ★★★★★ (Rate this item) Other Editions: Hardcover | Paperback
- Listmania! Add your list:** Recommendations for 'Essential data mining books', 'AI\*: machine learn, mind, brain, NN, SVM, kernel, PatRecog', and '"Artificial Intelligence (AI) for Games" related books (1)'.

- E.g., Association Rule:  
 $P(\text{buying "Pattern Classification"} | \text{buying "Machine Learning"}) = ?$

# Textbook and Reference

- Textbook
  - William C. Navidi, “Statistics for Engineers and Scientists,” McGraw-Hill (2 edition, 2007)
- Reference
  - Prem S. Mann, "Introductory Statistics," Wesley, (6 edition, 2007)
  - D. P. Bertsekas, J. N. Tsitsiklis, “Introduction to Probability,” Athena Scientific (2002)

# Topics to be Covered

Date	Content
03/02, 03/09	Descriptive Statistics (Chapter 1)
03/16, 03/23	Probability and Common Used Distributions (Chapters 2 & 4, quick review)
03/23, 03/30	Propagation of Error (Chapter 3)
04/20, 04/27	Confidence Intervals (Chapter 5)
05/04, 05/11	Hypothesis Testing (Chapter 6)
05/18, 05/25	Correlation and Simple Linear Regression (Chapter 7)
06/01~	More Topics ...

# Grading (Tentatively)

- Midterm and Final: 45%
- Homework: 30%
- Attendance/Other: 25%
  
- TA: 李鴻欣同學 (碩一)
  - E-mail: winleetw@gmail.com
  - Tel: 29322411ext 208 (資工系208室)