

統計學 (一)

Anderson's Statistics for Business & Economics (14/E)

Chapter 1: Data and Statistics

上課時間地點: 二 D56, 學思 040103

授課教師: 吳漢銘 (國立政治大學統計學系副教授)

教學網站: <http://www.hmwu.idv.tw>

Overview: What is Statistics?

1. The term statistics can refer to _____ such as averages, medians, percentages, and maximums that help us _____ a variety of business and economic situations.
2. Wikipedia: Statistics is the discipline (art and science) that concerns the _____, _____, _____, _____, and _____ of data.

1.1 Applications in Business and Economics

Accounting

Public accounting firms use statistical _____ procedures when conducting audits for their clients.

Economics

Economists use statistical information in making _____ about the future of the economy or some aspect of it.

Finance

Financial advisors use price-earnings _____ and dividend yields to guide their investment advice.

Marketing

Electronic point-of-sale scanners at retail checkout counters are used to _____ data for a variety of marketing research applications.

Production

A variety of statistical _____ are used to monitor the output of a production process.

Information Systems

A variety of statistical information helps administrators assess the _____ of computer networks.

1.2 Data

1. Data are the _____ collected, analyzed, and summarized for presentation and interpretation.
2. All the data collected in a particular study are referred to as the _____ for the study.

Elements, Variables, and Observations

1. Elements are the _____ on which data are collected.
2. A variable is a _____ of interest for the elements.
3. The set of measurements obtained for a particular element is called an _____.
4. A data set with _____ elements contains _____ observations.

5. The total number of _____ in a complete data set is the number of elements multiplied by the number of variables.

6. **Example** WTO, 惠譽國際 (Fitch Group)

TABLE 1.1 Data Set for 60 Nations in the World Trade Organization

Nation	WTO Status	Per Capita GDP (\$)	Fitch Rating	Fitch Outlook
Armenia	Member	3,615	BB-	Stable
Australia	Member	49,755	AAA	Stable
Austria	Member	44,758	AAA	Stable
Azerbaijan	Observer	3,879	BBB-	Stable
Bahrain	Member	22,579	BBB	Stable
Belgium	Member	41,271	AA	Stable
Brazil	Member	8,650	BBB	Stable
Bulgaria	Member	7,469	BBB-	Stable
Canada	Member	42,349	AAA	Stable
Cape Verde	Member	2,998	B+	Stable
Chile	Member	13,793	A+	Stable
China	Member	8,123	A+	Stable
Colombia	Member	5,806	BBB-	Stable
Costa Rica	Member	11,825	BB+	Stable
Croatia	Member	12,149	BBB-	Negative

Scales of Measurement

1. Scales of measurement include: _____.
2. The scale determines the _____ contained in the data.
3. The scale indicates the _____ and statistical analyses that are most appropriate.
4. **Nominal scale**
 - (a) Data are _____ used to identify an attribute of the element.
 - (b) A _____ or _____ may be used.
 - (c) **Example:** Students of a university are classified by the school in which they are enrolled using a nonnumeric label such as Business, Humanities, Education, and so on. Alternatively, a numeric code could be used for the school variable (e.g., 1 denotes Business, 2 denotes Humanities, 3 denotes Education, and so on).

5. Ordinal scale

- (a) The data have the properties of nominal data and the _____ of the data is meaningful.
- (b) A _____ or _____ may be used.
- (c) Example: Students of a university are classified by their class standing using a non-numeric label such as Freshman, Sophomore, Junior, or Senior. Alternatively, a numeric code could be used for the class standing variable (e.g., 1 denotes Freshman, 2 denotes Sophomore, and so on).

6. Interval scale

- (a) The data have the properties of ordinal data, and the _____ between observations is expressed in terms of a _____.
- (b) Interval data are always _____.
- (c) Example: Melissa has an SAT score of 1985, while Kevin has an SAT score of 1880. Melissa scored 105 points more than Kevin.

7. Ratio scale

- (a) Data have all the properties of interval data and the _____ of two values is meaningful.
- (b) Ratio data are always _____.
- (c) _____ is included in the scale.
- (d) Example: Price of a book at a retail store is \$200, while the price of the same book sold online is \$100. The ratio property shows that retail stores charge twice the online price.

Categorical and Quantitative Data

1. Data can be further classified as being _____ or _____.
2. The statistical analysis that is appropriate depends on whether the data for the variable are categorical or quantitative.
3. **Categorical Data**

- (a) Labels or names are used to identify an _____ of each element.
- (b) Often referred to as _____.
- (c) Use either the _____ or _____ scale of measurement.
- (d) Can be either _____ or nonnumeric.
- (e) Appropriate statistical analyses are rather _____.

4. Quantitative Data

- (a) Quantitative data indicate _____ or _____.
- (b) Quantitative data are always _____.
- (c) _____ operations are meaningful for quantitative data.

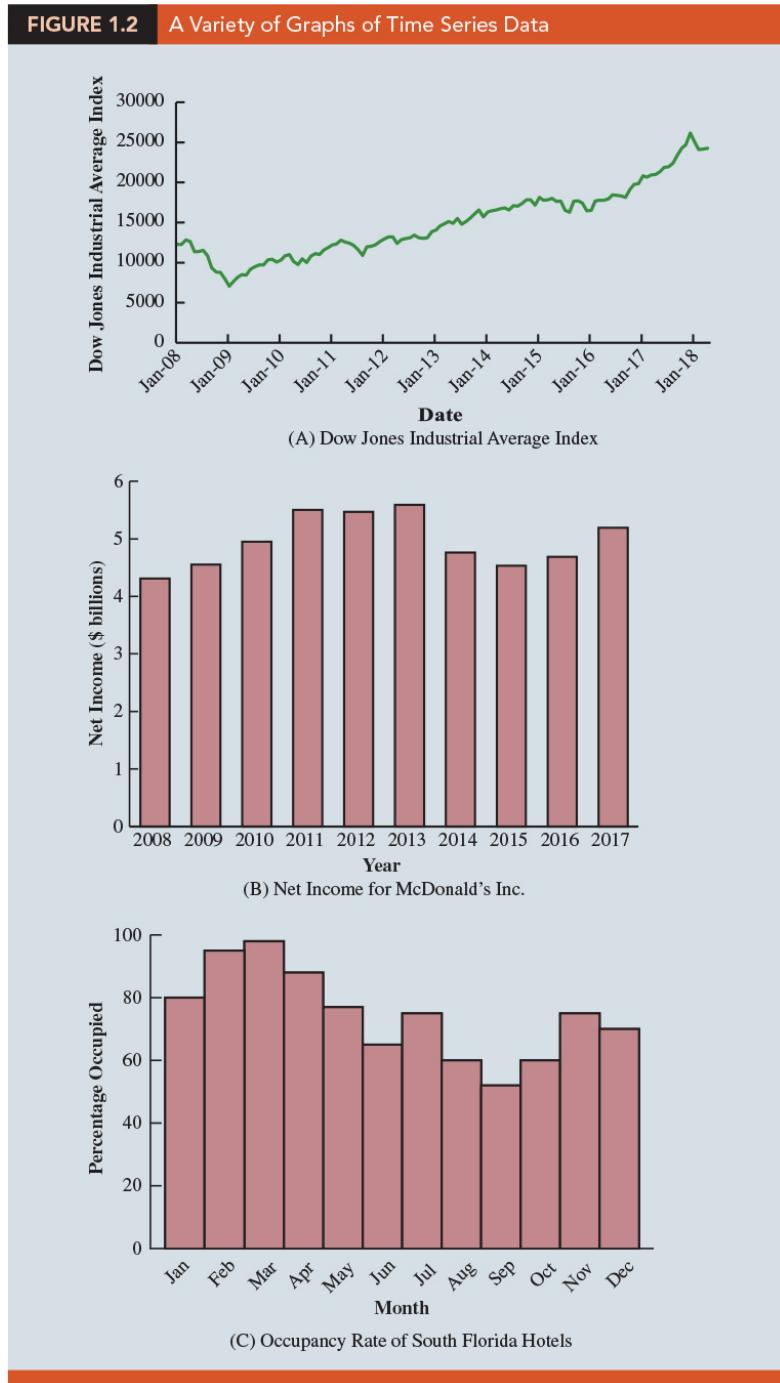
Cross-Sectional and Time Series Data

1. Cross-sectional data

- (a) Cross-sectional data are collected _____ or _____ point in time.
- (b) Example: The data in Table 1.1 are cross-sectional because they describe the five variables for the 60 World Trade Organization nations at the same point in time.

2. Time Series Data

- (a) Time series data are collected _____.
- (b) Example: Data detailing the number of building permits issued in Lucas County, Ohio in each of the last 36 months.
- (c) Graphs of time series data help analysts understand: (i) what happened in the _____, (ii) identify any _____ over time, and (iii) _____ future levels for the time series.



1.3 Data Sources

Existing Sources

1. Internal company records – almost any department
2. Business database services – Dow Jones & Co.
3. Government agencies – U.S. Department of Labor
4. Industry associations – Travel Industry Association of America
5. Special-interest organizations – Graduate Management Admission Council (GMAT)
6. Internet –e.g., 政府資料開放平臺: <https://data.gov.tw>

Observational Study

1. In observational (nonexperimental) studies _____ is made to _____ the variables of interest.
2. Example: Survey and public opinion, e.g., studies of smokers and nonsmokers are observational studies because researchers do not determine or control who will smoke and who will not smoke.

Experiment

1. In experimental studies the variable of interest is first _____. Then one or more other variables are _____ so that data can be obtained about how they influence the variable of interest.
2. Example: The largest experimental study ever conducted is believed to be the 1954 Public Health Service experiment for the Salk polio vaccine. Nearly two million U.S. children (grades 1- 3) were selected.

Time and Cost Issues

1. Searching for information can be _____.
2. Information may _____ by the time it is available.

Data Acquisition Errors

1. Organizations often charge for information even when it is not their primary business activity.
2. Using any data that happen to be available or were acquired with little care can lead to _____ information.

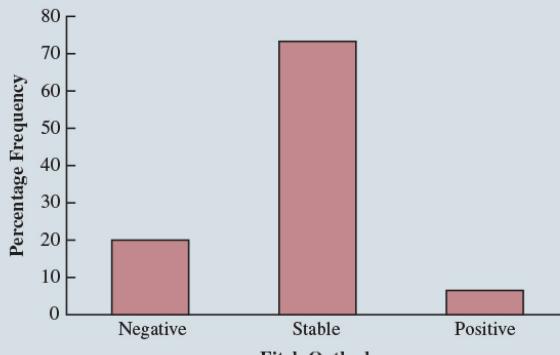
1.4 Descriptive Statistics

1. Most of the statistical information in the media, company reports, and other publications consists of data that are summarized and presented in a form that is easy for the reader to understand. Such _____, which may be _____, _____, or _____, are referred to as _____.
2. **Example** A tabular summary of the data showing the number of nations with each of the Fitch Outlook ratings (Table 1.4). A graphical summary of the same data, called a _____ (Figure 1.4). We can see that the majority of Fitch Outlook credit ratings are stable, with 73.3% of the nations having this rating. More nations have a negative outlook (20%) than a positive outlook (6.7%).

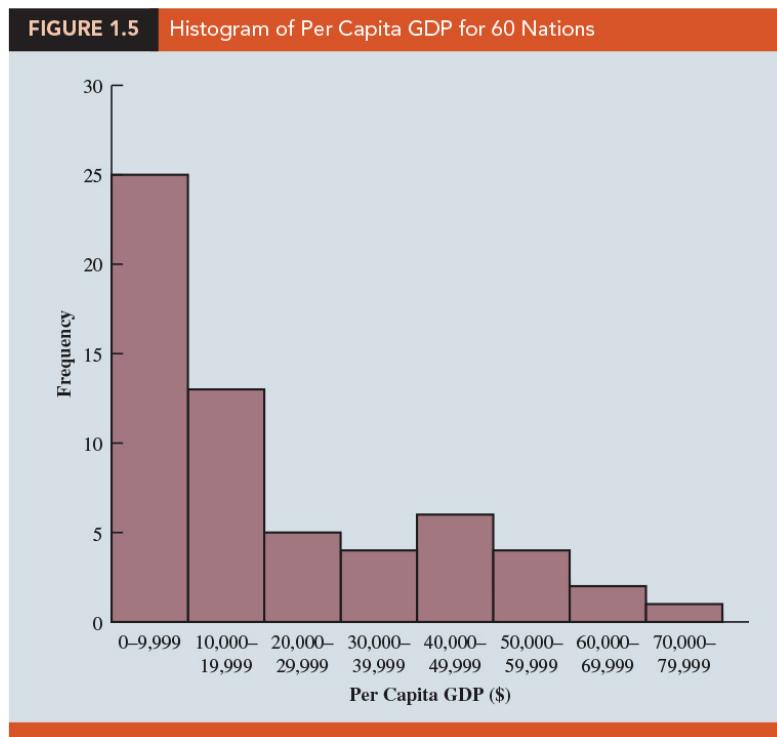
TABLE 1.4 Frequencies and Percent Frequencies for the Fitch Credit Rating Outlook of 60 Nations

Fitch Outlook	Frequency	Percent Frequency (%)
Positive	4	6.7
Stable	44	73.2
Negative	12	20.0

FIGURE 1.4 Bar Chart for the Fitch Credit Rating Outlook for 60 Nations



3. **Example** A graphical summary of the data for the quantitative variable Per Capita GDP in Table 1.1, called a _____ (Figure 1.5). Using the histogram, it is easy to see that Per Capita GDP for the 60 nations ranges from \$0 to \$80,000, with the highest concentration between \$0 and \$10,000. Only one nation had a Per Capita GDP exceeding \$70,000.



4. The most common numerical descriptive statistic is the _____.
5. The mean demonstrates a measure of the _____, or _____ of the data for a variable.

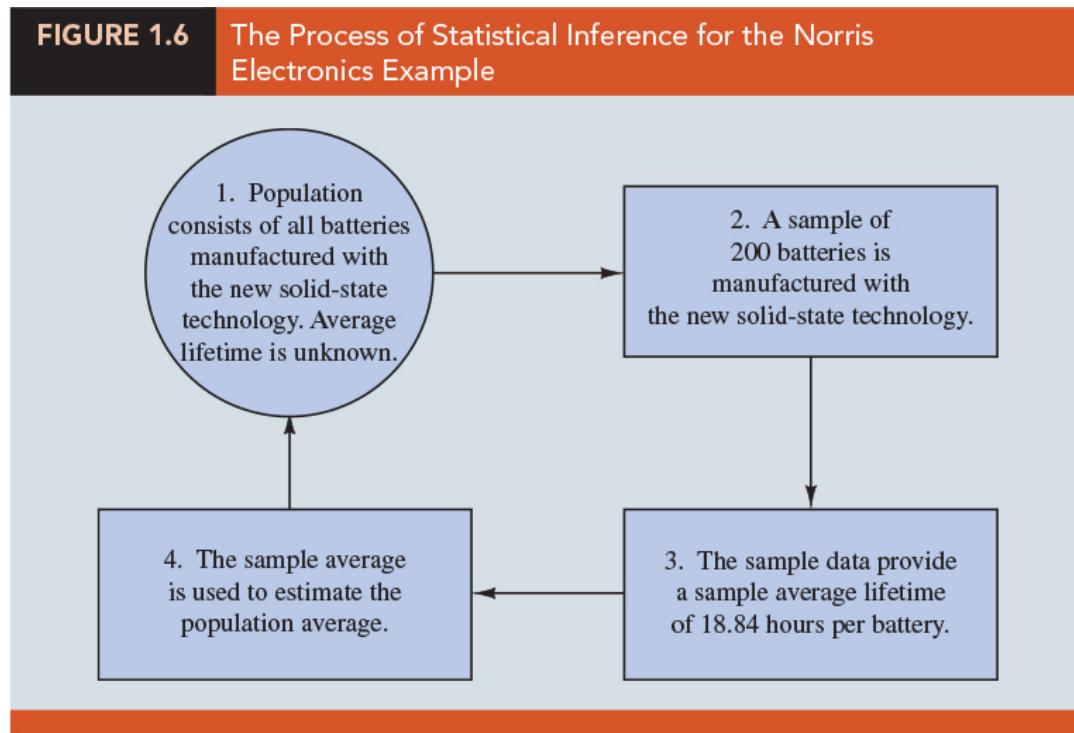
1.5 Statistical Inference

1. **Population:** The set of _____ in a particular study.
2. **Sample:** A _____ of the population.
3. **Statistical inference:** The process of using data obtained from a sample to _____ about the characteristics of a population.

4. **Census:** Collecting data for the _____ population.
5. **Sample survey:** Collecting data for a _____.
6. **Example** Rogers Industries manufactures lithium batteries used in rechargeable electronics such as laptop computers and tablets. In an attempt to increase battery life for its products, Rogers has developed a new solid-state lithium battery that should last longer and be safer to use.
- (a) population: all lithium batteries that could be produced using the new solid-state technology.
 - (b) sample (data): 200 batteries manufactured with the new solid-state technology were tested. (Table 1.5) the number of hours each battery lasted before needing to be recharged under controlled conditions.

TABLE 1.5 Hours Until Recharge for a Sample of 200 Batteries for the Rogers Industries Example									
Battery Life (hours)									
19.49	18.18	18.65	19.45	19.89	18.94	17.72	18.35	18.66	18.23
19.08	19.92	19.01	18.84	17.73	19.70	18.37	18.69	19.98	18.80
19.11	18.26	19.05	17.89	19.61	18.52	18.10	19.08	18.27	18.29
19.55	18.81	18.68	17.43	20.34	17.73	17.66	18.52	19.90	19.33
18.81	19.12	18.39	19.27	19.43	19.29	19.11	18.96	19.65	18.20
19.18	20.07	18.54	18.37	18.13	18.29	19.11	20.22	18.07	18.01

- (c) Suppose Rogers wants to use the sample data to make an inference about the average hours of battery life for the population of all batteries that could be produced with the new solid-state technology.
- (d) The sample average battery life: _____. We can use this sample result to estimate that the average lifetime for the batteries in the population is _____ hours.
- (e) (Figure 1.6) a graphical summary of the statistical inference process for Rogers Industries.



1.6 Analytics

1. Analytics is the scientific process of transforming data into _____ for making better _____.
2. Three categories of techniques:
 - (a) **Descriptive analytics:** This describes what has happened in the _____.
 - (b) **Predictive analytics:** Use models constructed from past data to _____ or to assess the _____ of one variable on another.
 - (c) **Prescriptive analytics:** The set of analytical techniques that yield a _____

1.7 Big Data and Data Mining

1. **Big data:** Large and complex data set. Three V's of Big data:
 - (a) _____: Amount of available data.
 - (b) _____: Speed at which data is collected and processed.

(c) _____ : Different data types.

2. Data Mining

- (a) Methods for developing useful _____ information from _____.
- (b) Using a combination of procedures from statistics, mathematics, and computer science, analysts _____ to convert it into useful information.
- (c) The most effective data mining systems use automated procedures to discover relationships in the data and predict future outcomes prompted by general and even vague queries by the user.
- (d) _____ such as multiple regression, logistic regression, and correlation are heavily used.
- (e) Also needed are computer science technologies involving _____ and _____.
- (f) With the enormous amount of data available, the data set can be partitioned into a _____ (for _____ development) and a _____ (for _____ the model).
- (g) Careful interpretation of results and extensive testing is important.

1.8 Computers and Statistical Analysis

1. Computer: laptop, PC, Server, ...
2. OS: Windows, Mac OS, Linux, ...
3. Software for statistical analysis: Microsoft Excel, JMP, SPSS, SAS, R, Python

1.9 Ethical Guidelines for Statistical Practice

1. In a statistical study, unethical behavior can take a variety of forms including: improper _____, inappropriate _____ of the data, development of _____, use of inappropriate _____, _____ of the statistical results.

2. One should strive to be _____, _____, _____, and _____ as you collect, analyze, and present data.
3. As a consumer of statistics, one should also be aware of the possibility of _____ by others.
4. The American Statistical Association (ASA) developed the report "Ethical Guidelines for Statistical Practice". It contains 67 guidelines organized into 8 topic areas:
 - (a) Professionalism.
 - (b) Responsibilities to Funders, Clients, and Employers.
 - (c) Responsibilities in Publications and Testimony.
 - (d) Responsibilities to Research Subjects.
 - (e) Responsibilities to Research Team Colleagues.
 - (f) Responsibilities to Other Statisticians/Practitioners.
 - (g) Responsibilities Regarding Allegations of Misconduct.
 - (h) Responsibilities of Employers Including Organizations, Individuals, Attorneys, or Other Clients.

⌚ SUPPLEMENTARY EXERCISES

3, 4, 5, 9, 13, 14, 19, 25