

Topics in Mathematics

- Financial Mathematics
- Operation Research
- Game Theory
- Probability
- Statistics
- Logic
- ...

Chapter 1 - 1

Statistics

Biostatistics
Engineering Statistics
Social Statistics

...

Chapter 1 - 2

What is Statistics?

- *Lies and nothing but lies!?*
- *Scientific method for understanding and solving problem, and making decision!*

Chapter 1 - 3

Decision

- *National Medical Statistics*: 1 out of 200 patients had miscarriage after a medical procedure.
Failure rate: $1/200 = .005 = .5\%$,
Success rate: $199/200 = .995 = 99.5\%$
- Should you have this medical procedure done by Dr. ###?
- How about if 1 out of 200 randomly selected cases from the doctor had miscarriage?

Chapter 1 - 4

Poll Study Report 1

The poll showed that 25% voted for Policy 2000.

What do we know from the statistic 25%?

Chapter 1 - 5

Poll Study Report 2

The poll showed that 25% voted for Policy 2000, with a margin of error $\pm 3\%$ at 95% confidence level.

Measure of Reliability

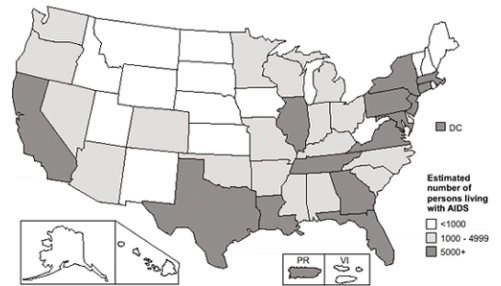
Chapter 1 - 6

Number of Hospital Admissions for Viral Pneumonia, 1983-1998, and 1990 Census Population: Females Aged 0-17, in Sacramento, San Francisco, and Los Angeles, California

	Cases	Population
Sacramento	951	134,492 (.71%)
San Francisco	837	126,162 (.66%)
Los Angeles	3743	140,913 (2.66%)

Chapter 1 - 7

Estimated number of persons living with AIDS at the end of 1999, by state of residence



Chapter 1 - 8

Estimation of Average Blood Pressure for Hypertensive Adults

	Men (n = 1100)	Women (n=1300)
Age	62 ± 10	64 ± 9
Systolic	153 ± 15	160 ± 19
Diastolic	92 ± 10	89 ± 10

Chapter 1 - 9

AIDS Epidemic

Findings:

Out of 840 subjects randomly selected in the study, 32 tested HIV positive, (95% CI, [2.7%, 4.9%] or $3.8\% \pm 1.1\%$)

The difference between the percentages of subjects tested HIV positive from regions A and B is statistically significant at 5% level.

Chapter 1 - 10

What is Statistics?

Statistics is the science of *understanding data* and of *making decisions* in the face of “variability” and “uncertainty”.

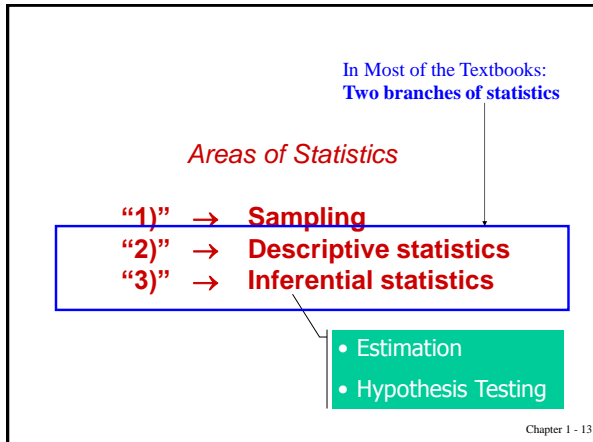
Chapter 1 - 11

Statistics is a field of study concerned with the

- 1) data collection, [**Producing data**]
- 2) *organization, summarization, examination* and *providing an overview* of the general features of **data**, [**Exploring Data**]
- 3) and the *drawing of inferences* about a body of data (**population**) based on the properties of a part of the data (**sample**) observed. [**Statistical Inference**]



Chapter 1 - 12



Some Basic Terms In Statistics

Population (n.): the collection of all individuals of interest in a statistical study. [**Census (v.):** obtain information from the whole population.]

Sample (n.): any subset of individuals from the population. [**Sample (v.):** obtain a sample from the population.]

Data: the facts and figures that are collected, and analyzed in a statistical study.

Parameter: Numerical summary of a population.
Statistic: Numerical summary of a sample.

Chapter 1 - 14

Individuals (subjects, experimental unit): the entities on which data are collected.

Variable: a characteristic of interest for the individual which takes on different values in different individual.

Chapter 1 - 15

Variable Types

- **Quantitative Variables** (can be measured) [height, number of subscriptions, ...]
 - **Discrete:** a variable that has a countable number of possible values.(counts)
 - **Continuous:** a variable that has an uncountable number of possible values.(measurements)
- **Qualitative (Categorical) Variables** [hair color, gender, ...]
 - **Variable contains only two categories of outcomes is called a binary variable.**

Chapter 1 - 16

Quantitative or Qualitative?

- Number of patients per hour in a clinic
- Marital Status
- Time is take to get to school
- Rating of wine (excellent, good, fair, poor)
- Body temperature
- Gender

Which of the variables above is **binary** variable?

Chapter 1 - 17

Measurement Scales

- **Nominal:** consists of labels, names or categories.
- **Ordinal:** data that the order or rank is meaningful.
- **Interval:** numerical data that arithmetic operations are meaningful.
- **Ratio:** data that the ratio of two data is meaningful.

Chapter 1 - 18

What is the Measurement Scale?

- ACT score
- Marital Status
- Time is take to get to school
- Rating of wine (excellent, good, fair, poor)
- Body temperature
- Gender

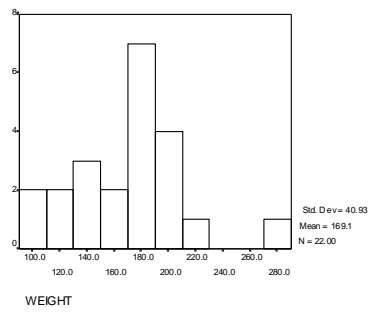
Two Basic Types of Analyses in Introductory Statistics

Univariate analysis: is an analysis that involves only a single variable.

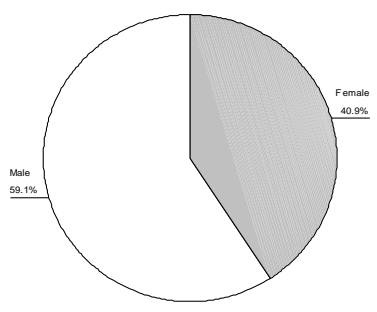
Bivariate analysis: is an analysis that involves two variables.

Weights of a random sample of 22 subjects in a study.

Univariate

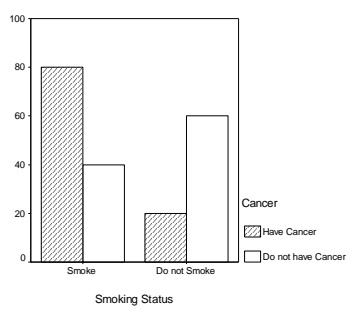


Examine Distribution of Gender



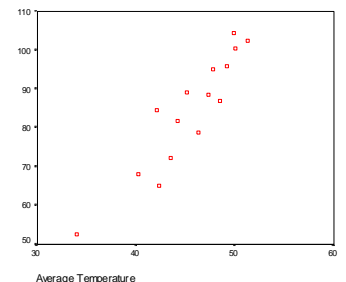
Examine correlation between smoking and lung cancer

Bivariate



Average annual temperature and the mortality index for a type of breast cancer in women in certain region of Europe

Bivariate



Response/Explanatory Variables

- Response, (Dependent, Outcome) Variable:
Have Lung Cancer or not
- Explanatory (Independent, Predictor) Variables:
Smoking or not, Age, ...

Producing Data

- Study Design
- Data Collection

Sources of Data

- Routinely kept records
- Surveys
- Designed Experiments
- External sources (internet, public or private organizations)
- ...

Preparing Data

- Retain the raw data source.
- Study log: data received, investigation, statisticians, discussions and results, data source, changes made.
- Create machine readable data base
 - Prepare a "work" file for data analysis.
- **Data cleaning** (logic check, correcting, clarifying).

Sampling Techniques (For obtaining representative sample)

- **Simple random sampling (SRS):** A sample obtained from a population in such a manner that all samples of the same size have equal likelihood of being selected.

Random sampling using numbers:

- Label each individual with number.
- Use random digits table, random numbers generator or other random selection methods to select the numbers.

Table of Randomly Generated Digits From 0 To 9.

6	7	5	7	3	7	4	6	1	0	4	6	4	4	1	5	4	0	2	2	6	0	2	0	4	1	2	6	7	6
2	7	1	4	6	2	7	9	6	3	1	3	4	0	8	6	8	3	2	9	6	6	3	1	1	4	1	3	0	3
2	2	1	6	6	4	6	8	6	1	2	6	9	3	2	0	3	4	8	2	8	3	6	4	6	0	0	1		
7	1	2	7	3	6	8	4	8	2	7	1	8	6	0	8	6	6	6	4	7	5	3	7	0	3	7	3	5	4
5	7	8	8	2	8	9	0	5	7	7	1	1	9	6	6	7	8	9	6	7	9	4	2	3	7	5	7		
7	8	4	4	3	6	2	1	2	4	8	2	5	3	7	1	6	4	4	1	7	5	4	8	3	7	7	1	6	7
5	5	6	1	7	8	9	6	3	7	3	0	4	1	1	2	5	5	9	2	8	1	8	0	6	8	1	7	2	3
4	7	5	0	4	2	1	7	1	7	7	4	8	3	2	6	1	8	8	4	3	8	6	4	0	8	2	1	6	
4	2	7	6	1	8	5	1	3	4	7	2	2	8	6	4	6	8	9	6	4	8	2	8	6	6	3	2	2	8
8	7	3	4	1	9	6	8	7	9	8	1	6	1	5	8	2	2	6	5	6	7	4	4	3	4	7	6	1	1
2	3	4	4	6	8	9	8	5	2	9	6	5	7	2	6	5	4	6	5	4	1	7	5	4	7	3	1	8	0

2 7 9 2 6 6 3 8 8 1 1 5 1 7 5 7 9 2 5 7 1 3 9 6 4 8 7 6 3 2

3 8 4 4 8 4 8 7 8 9 1 8 3 3 8 2 4 6 9 7 3 9 3 6 4 4 2 0 0 6

9 3 0 8 8 1 6 5 3 8 3 3 6 9 4 1 7 1 4 5 7 2 8 7 8 2 2 8 1 5

2	4	1	6	3	4	0	8	9	1	7	1	7	7	5	3	8	7	5	4	0	3	9	6	2	6	9	4	8	
3	7	8	5	8	1	4	8	1	4	8	4	5	2	7	6	1	6	2	5	7	8	1	7	5	2	4	0	0	
6	8	9	4	7	7	5	2	7	4	9	1	8	2	5	4	3	8	8	4	4	5	1	6	3	7	9	2	4	
8	6	3	0	3	4	5	5	3	5	1	7	6	6	2	3	6	7	1	7	4	4	1	5	7	8	7	4	8	
4	8	9	5	6	3	3	2	8	1	6	8	4	9	3	4	3	1	2	8	2	7	6	8	9	4	6	4	8	
7	2	2	5	8	5	4	8	8	5	3	6	4	1	5	5	2	9	4	7	2	1	9	6	3	8	5	1	8	1
3	2	8	5	1	6	7	2	9	0	1	7	6	6	4	6	3	6	7	4	0	2	0	7	1	3	4	3	6	
8	1	2	8	3	9	2	4	4	5	9	1	4	3	5	3	1	8	4	5	6	5	3	8	8	3	5	8		
8	5	2	8	8	9	6	0	8	6	3	2	2	1	2	1	6	7	8	5	1	8	1	7	1	5	3	8	3	
0	3	8	4	1	7	7	1	5	7	8	3	6	1	3	1	5	8	2	8	4	3	8	5	3	4	2	5	4	
5	4	9	4	7	6	9	4	0	7	1	7	6	1	2	1	7	9	4	9	3	9	2	8	1	0	5	3	4	
3	8	4	4	7	3	1	2	4	6	2	7	4	9	8	9	3	4	7	7	5	4	5	7	6	6	8	2	5	4
0	3	3	8	7	1	9	7	1	4	7	3	4	7	6	2	4	4	1	5	9	1	1	3	1	8	1	7	8	7
5	5	3	8	9	2	2	6	1	0	1	8	6	2	3	6	8	2	8	6	4	7	7	2	7	7	3	7	1	2

How to take a random sample of 5 cars from 50 cars?



Line 110: (Select two digits a time)

3844 4878 1833 2469 39364 42006

2 7 9 2 6 6 3 8 8 1 1 5 1 7 5 7 9 2 5 7 1 3 9 6 4 8 7 6 3 2
 3 8 4 4 8 4 8 7 8 9 1 8 3 3 8 2 4 6 9 7 3 9 3 6 4 4 2 0 0 6
 9 3 0 8 8 1 6 5 3 8 3 3 6 9 4 1 7 1 4 5 7 2 8 7 8 2 2 8 1 5

You are going to take a random sample of five individuals from a company that has 300 employees. Their ID's are from 001 to 300.

Use the random digits above row by row from top left to the bottom to generate the five 3-digit ID's.

Ex: Create a random digit table with 50 random digits.

Other Sampling Techniques (For obtaining representative sample)

- **Stratified random sampling:** Classify the population into at least two strata, then draw a sample from each.
- **Cluster sampling:** Divide the population into sections, randomly select few of those sections, and then choose all members in them.
- **Systematic sample:** Select every i-th member in the population according to their ID.

** **Convenient sample:** Use results that are already available or using data that is convenient to obtain. This approach often can not obtain representative sample.

Bias

- **Response bias:** influenced response, measurement bias, lies, ...
- **Non-response bias:** shy, conservative, ...

Voluntary response sample are generally biased because people with strong opinions are most likely to respond.

Statistical Studies

- **Observational Study:** conditions to which subjects are exposed are not controlled by the investigator. (No attempt is made to control or influence the variables of interest.)
- **Experimental Study:** conditions to which subjects are exposed to are controlled by the investigator. (Manipulated treatments are used to see their effect on other variables.)

Results from observing behavior and outcomes from the use of medicine for 200 randomly selected patients.
(Patients chose their medicine)

Treatment	Hypertension		Total
	Yes	No	
Drug A	44	56	100
Drug B	29	71	100
Total	73	127	200

Drug A: $44/100 = 44\%$
 Drug B: $29/100 = 29\%$

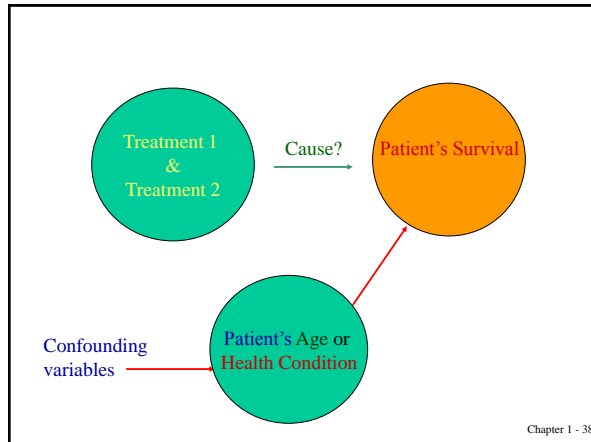
Treatment	Hypertension					
	Below 65			65+		
	Yes	No	Total	Yes	No	Total
Drug A	5	18	23	39	38	77
Drug B	17	60	77	12	11	23

* Older patients prefer Drug A

OR <65: Drug A: $5/23 = 22\%$
Drug B: $17/77 = 22\%$

OR 65+: Drug A: $39/77 = 51\%$
Drug B: $12/23 = 52\%$

Chapter 1 - 37



Confounding Effect

- Variables are said to be confounded when their effects on the outcome cannot be distinguished from each other, whether they are part of a study or not,

Chapter 1 - 39

Experimental Design

- Completely Randomized Design:** a design of experiment in which all the experimental units are allocated at random among all the treatments.
 - Control Group, Placebo Group
 - Treatment Group

Possible problem?

- **Placebo effect.**
- **Hawthorne effect:** bias from the fact that subjects know that they are participated in the experiment.

Chapter 1 - 40