

# Lecture 01: Introduction— Statistics with Applications

## Study Objectives:

1. A brief introduction to the genesis of statistics.
2. On statistical sampling schemes.
3. Examples of application in public health, epidemiology, and biomedical studies.



國立中興大學  
National Chung Hsing University

Applied Math./NCHU: Statistics

## What is “Statistics”

(Textbook Sec. 1.5, pp. 3 ~7)

- State
- Granut's mortality table
- Halley's knowledge of 'insurance'
- Numerical science of society and descriptive science of states
- Drawing inferences about individuals from the data
- Important figures: Bernoulli, Gauss, Laplace, Galton, Pearson, Gosset, Fisher, Neyman,....

---

## Sampling scheme, data collection process, and study design

- Goal: to reflect the 'situation' of the true situation (population).
  - Possible approaches:
    - Simple random sampling
    - Stratified sampling
    - Cluster sampling
    - Systematic sampling
- 

---

## Statistics with applications: examples in epidemiology and clinical studies

流行病學 (Epidemiology; **Webster's** dictionary: )

- The branch of medicine dealing with the **incidence** and **prevalence** of disease in large populations and with detection of the source and cause of epidemics.
  - The **factors** contributing to the presence or absence of a disease.
-

---

## Study Design

- 橫斷型研究(cross sectional study)
  - 病例對照研究法(case-control study)
  - 世代(追蹤)研究法(cohort study, follow-up study, longitudinal study)
  - 重疊病例對照研究法 (nested case-control study)
  - 病例世代研究法 (case-cohort study)
- 

---

## Data types

- **baseline** : sex, age, height, weight, race, socioeconomic status...(often fixed and can be measured or stated without error.)
  - **vital data** : blood type, genotype or gene polymorphism,...
  - **biological data** : RBC count, WBC count, serum urine, serum cholesterol level, GOT, GPT, blood pressure....often varies in time and measured with small error. )
  - **qualitative data** (obtained from questionnaire) : food intake, exercise, habit, risk taking behavior, attitude,...(often obtained with large error, relies on the interviewer, needs standardization.)
-

# Example 1: Case-control Study

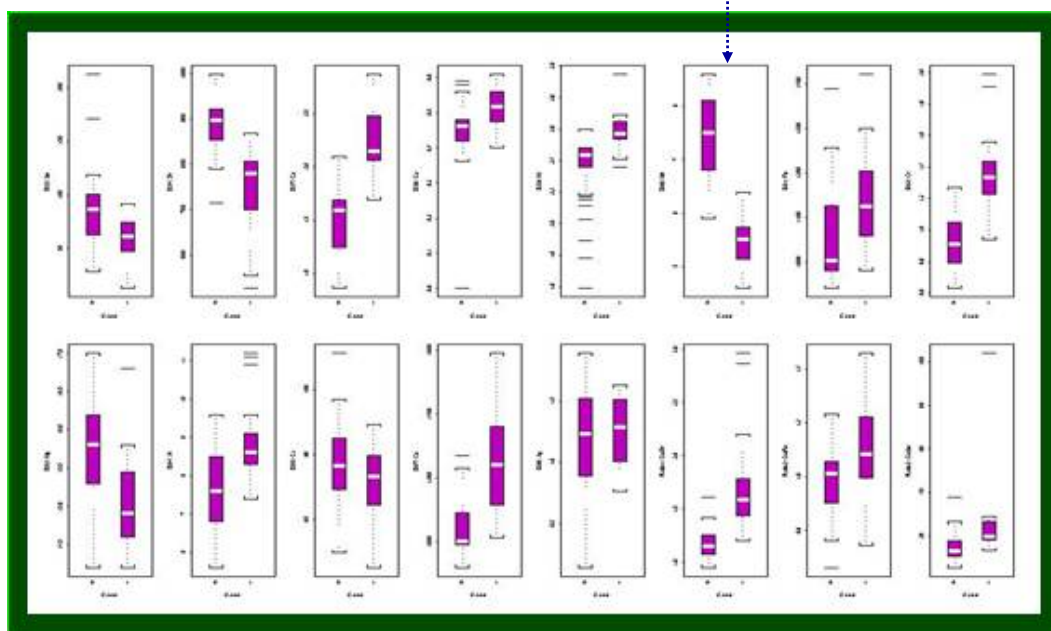
女性乳癌的診斷 (with hospital controls)

Element	Normal	Benign	Malignant	P <sub>NB</sub>	P <sub>NM</sub>	P <sub>BM</sub>
Se	99.50±25.83	83.87±33.48	60.94±17.78	<0.05	<0.01	<0.01
Zn	976.7±92.2	884.9±55.5	753.8±86.4	<0.01	<0.01	<0.01
Cd	1.13±0.42	1.50±0.34	2.25±0.28	<0.01	<0.01	<0.01
Co	0.49±0.08	0.46±0.09	0.51±0.05	NS	NS	<0.01
Ni	2.44±0.18	2.37±0.21	2.58±0.11	NS	<0.01	<0.01
<b>Mn</b>	<b>9.09±1.25</b>	<b>7.47±0.74</b>	<b>5.50±0.47</b>	<0.01	<0.01	<0.01
Fe	1040.4±99.9	1055.8±107.0	1142.5±109.4	NS	<0.01	<0.01
Cr	0.64±0.20	0.93±0.17	1.36±0.26	<0.01	<0.01	<0.01
Mg	16.14±0.82	15.70±0.62	15.04±0.62	<0.05	<0.01	<0.01
Al	7.94±1.40	7.66±1.03	8.84±1.00	NS	<0.05	<0.01
Ca	89.16±7.76	88.53±6.49	86.32±5.49	NS	NS	NS
Cu	964.95±70.3	1038.93±82.75	1252.20±150.9	<0.01	<0.01	<0.01
Ag	1.29±0.39	1.15±0.39	1.22±0.31	NS	NS	NS

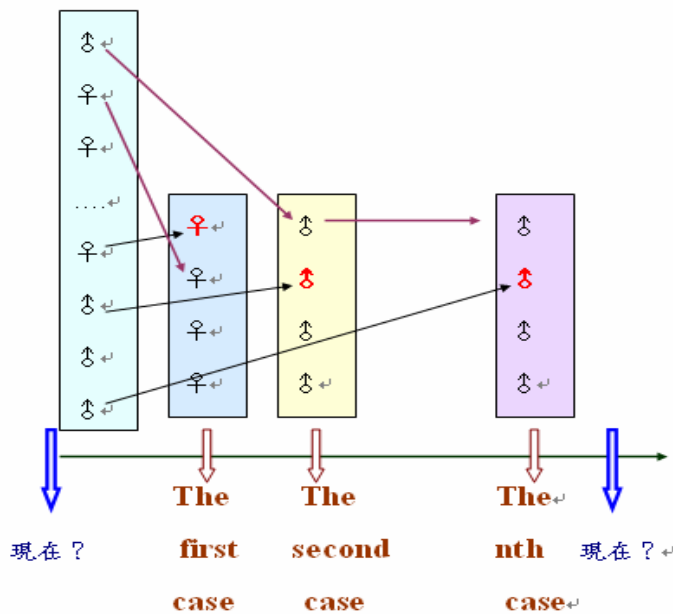
# Example 1: Case-control Study (cont'd)

Box plot: benign vs. malignant

Mn

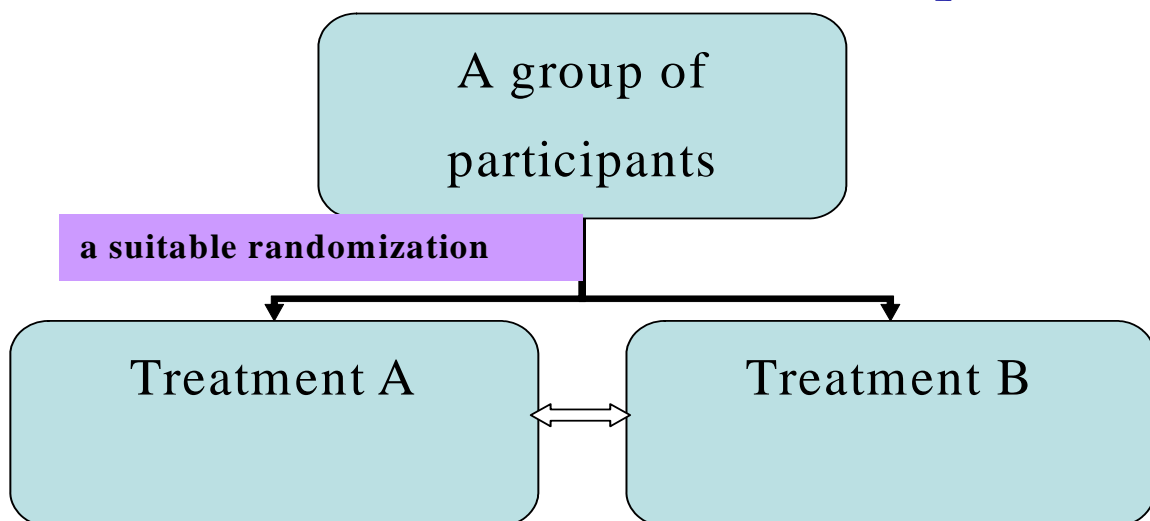


## Example 2: A retrospective or prospective cohort study



- Nested case-control study
- Time matching !!
- Matching on other variables? (sex, age, ...)

## Example 3. Clinical trials: a comparison between treatments for some 'endpoint'



# Example 4: Creutzfeldt-Jakob Disease

## (庫賈氏症, CJD) and S100 protein (as a 'marker' of diagnosis/early detection?)

BMJ 1999; 319: 557-562

### Diagnosis of Creutzfeldt-Jakob disease by measurement of S100 protein in serum: prospective case-control study

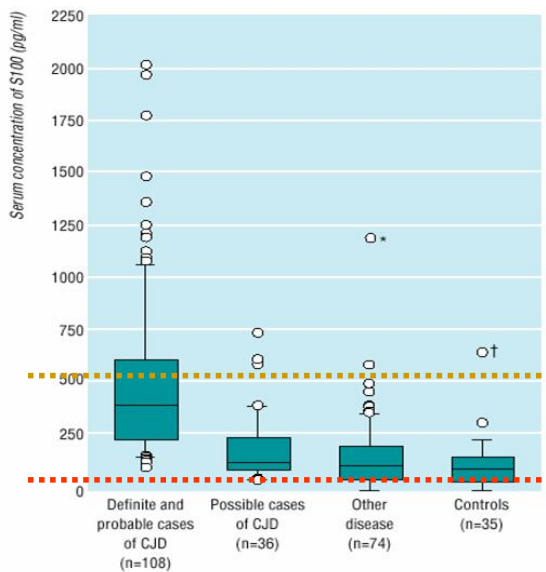
Markus Otto, Jens Wiltfang, Ekkehard Schütz, Inga Zerr, Anke Otto, Annette Pfahlberg, Olaf Gefeller, Manfred Uhr, Armin Giese, Thomas Weber, Hans A Kretzschmar, Sigrid Poser

## Example 4 (cont.): on CJD and S100 protein

**Table 3** Clinical and electroencephalographic signs and symptoms of 224 patients with suspected Creutzfeldt-Jakob disease at time of entry into the study. Values are numbers of patients (percentages)

Symptoms	Definite or probable CJD (n=108)	Possible CJD (n=36)	Other disease (n=74)
Rapidly progressive dementia of <2 years' duration	101 (94)	36 (100)	32 (43)
Periodic sharp wave complexes on electroencephalography	89 (82)	0	2 (3)
Myoclonus	93 (86)	26 (72)	26 (35)
Visual or cerebellar symptoms, or both	90 (83)	34 (86)	31 (42)
Pyramidal or extrapyramidal signs, or both	72 (67)	32 (89)	38 (51)
Akinetic mutism	62 (57)	12 (33)	7 (9)

## Example 4 (cont.): Box-plots

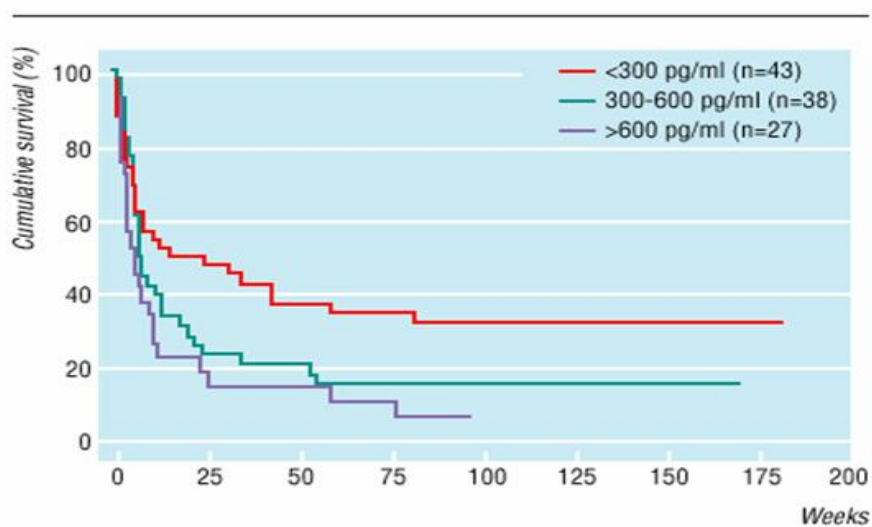


**Fig 1** Serum concentrations of S100 protein by final diagnosis. 10th, 25th, 75th, and 90th percentiles are shown. CJD=Creutzfeldt-Jakob disease; \*Patient with meningoencephalitis; †Patient with amyotrophic lateral sclerosis

- Cut-off point?
- To maximize (sensitivity+specificity) (**Youden index**= sensitivity+specificity-1)
- Cut-off point S100=213pg/ml
- Sens.=77.8%
- Spec.=81.1% (**Youden index=0.59**)

## Example 4 (cont.):

### S100 as a differentiation between survival?



**Fig 3** Kaplan-Meier survival curves at different serum concentrations of S100 protein (P=0.023)

---

## Exercises and Homework

- 1. Is 'random sampling' an important mechanism for statistical inferences? Why?
  - 2. Please sketch how to conduct a simple random sampling and stratified random sampling from the population of NCHU-students for answering the question:  
“What are the students’ opinion towards the Innovation (Reform) on Teaching of NCHU”.
- 

---

## Exercises and Homework (cont.)

- 3. Try to give several examples on the applications or possible applications of STATISTICS.
  - 4. Textbook exercises: (pp. 7~8)  
1, 2, 4, 5
-