การแปลผลและ
การเขียนผลทางสถิติ

ผศ.ดร. ณัฐนารี เอมยงค์

Study designs

Observational study

- Cohort study
- Case-control study
- Cross-sectional study
- Longitudinal study

Possible bias or confounding

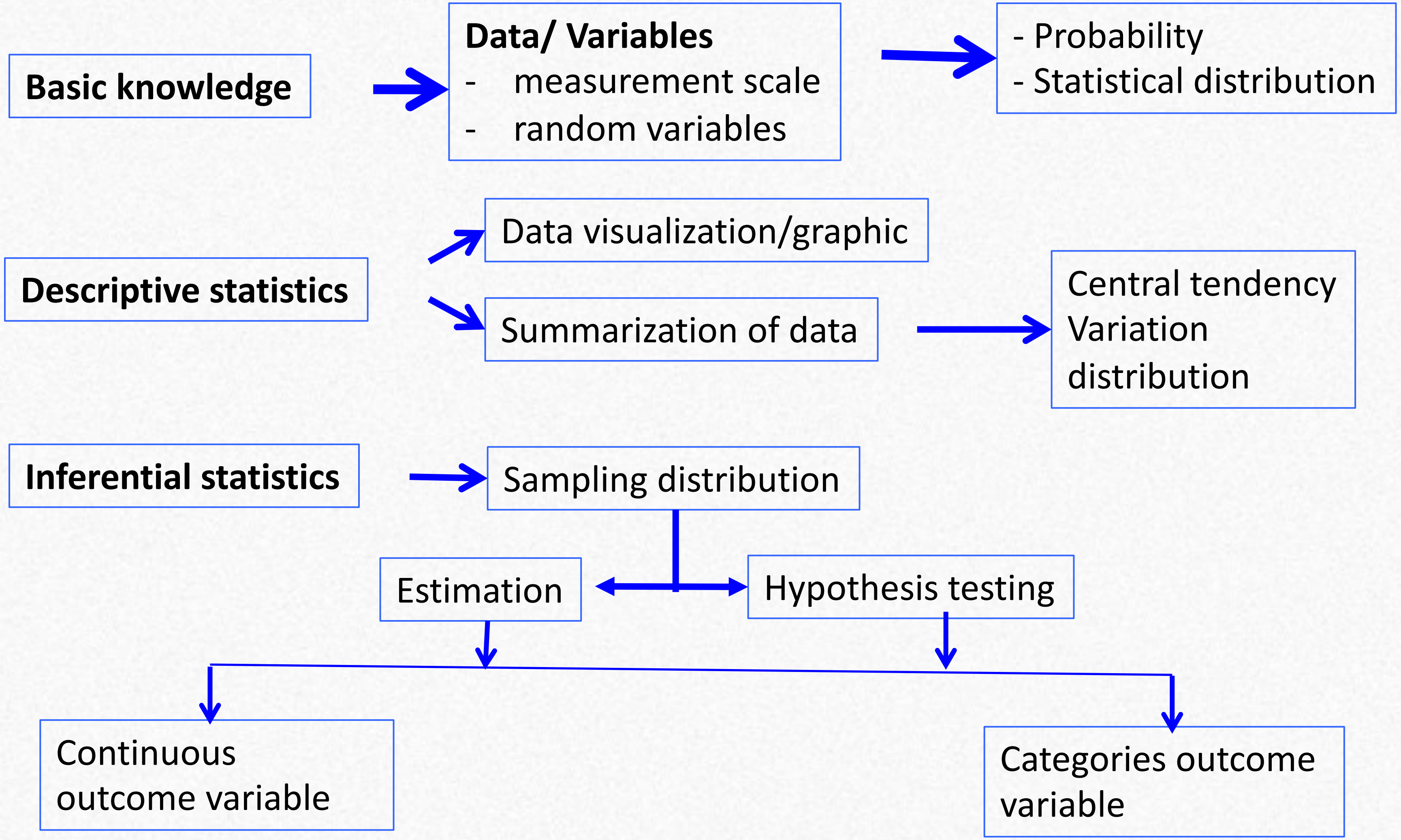
Experimental study

- True experiment
- Clinical experiment
- Quasi experimental
- Longitudinal study

Control bias by study design, some bias still.

What do we learn from statistics class?

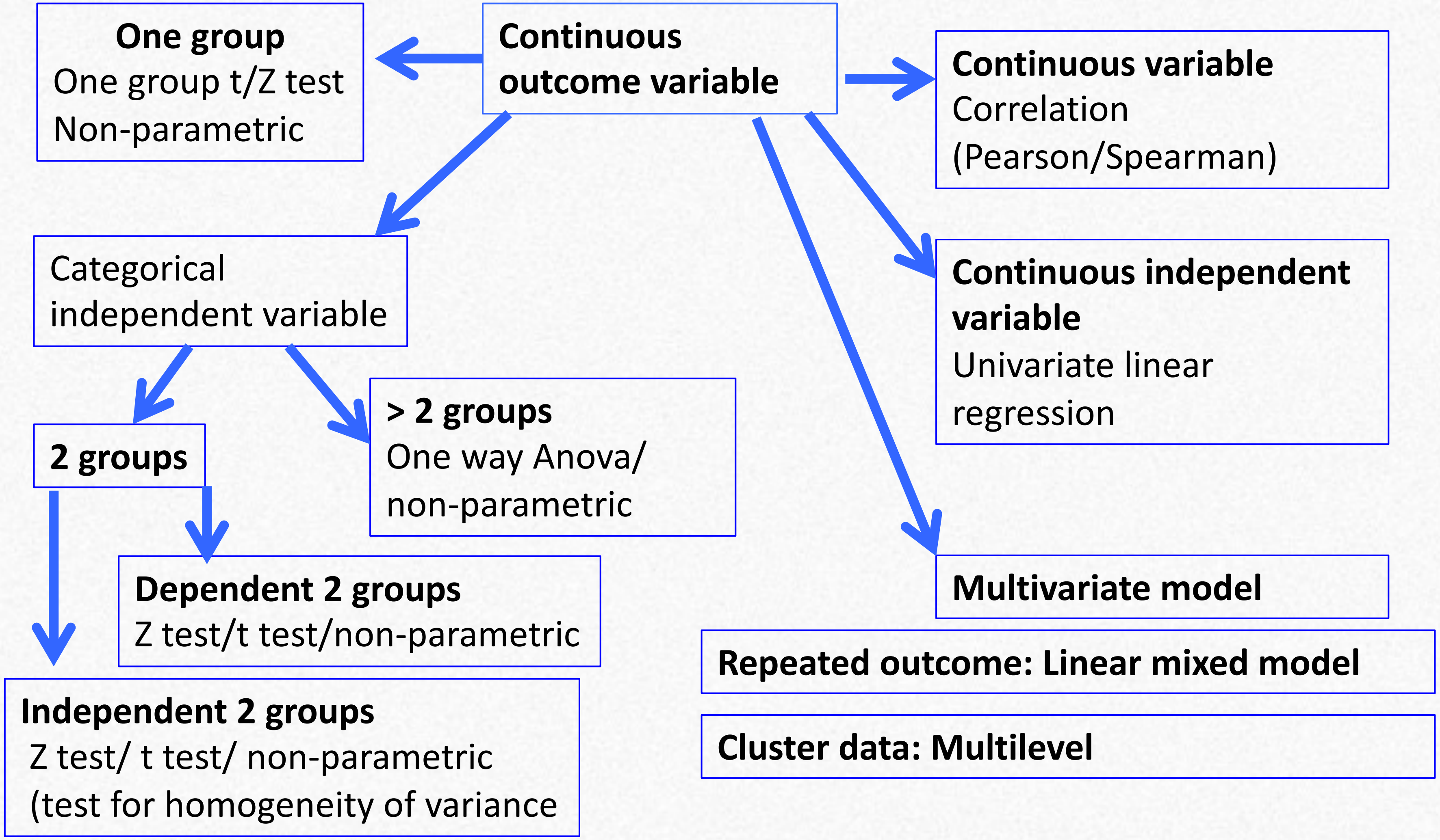
Topics in Biostatistics



Continuous outcome/primary outcome

- Identify association/correlation
  - Pearson/Spearman correlation
- Categorical independent variable (compare means),
  - 2 groups: t-test/Z-test (how to control bias?)
  - Categorical independent, >2 groups: ANOVA/ Linear Regression
- Continuous independent variable(s)
  - Linear regression (put confounding factors into the model)

Topics in Biostatistics (cont.)



Continuous outcome/primary outcome

- Identify association/correlation
  - Chi-Square test
- Categorical independent variable,
  - 2 groups or more: Chi-square test(how to control bias?)
- Continuous independent variable(s)
  - Logistics regression (put confounding factors into the model)
  - Multinomial logistic regression (ordinal/nominal logistic regression)

Topics in Biostatistics (cont.)

Categories outcome variable

One group
one sample binomial test/ McNemar's test

Categorical independent variable

>2 groups
Chi-square test for homogeneity

2 groups
two sample binomial test
Chi-square test for homogeneity

Continuous variable
Chi-square test for association/ Kappa

Continuous independent variable
Generalize linear model

Multivariate model
Generalize linear model

Repeated outcome: GLMM

Cluster data: Multi-level

Topics in Biostatistics (cont.)

- Counting outcome variable → Poisson Model/Negative binomial model
- Time to event outcome variable → Survival analysis/Cox Model
- Time series

Add up topics

- Power/ Sample size estimation
- Method to control confounding: such as Propensity score/ Matching/Cluster/stratification etc

Descriptive and inferential statistics

- Descriptive statistics are concerned about presentation, organization and summarization of data.
  - Indices to summarize the data
  - Organizing and graphing the data

- Inferential statistics allow us to generalized from sample to population.
  - What was the design?
  - Guard against bias: comparability, representative of target population,…
  - Source/controlling/quantification of uncertainty/variation

Measurement scale

- Nominal scale
- Ordinal scale
- Interval scale
- Ratio scale

Descriptive Statistics

- Measures of location
- measures of spread
- Describe shape

Measurement of location

- ✧Mean
- ✧Median
- ✧Percentile
- ✧Mode
- ✧Minimum
- ✧Maximum
- ✧Percent/proportion

Measurement of spread

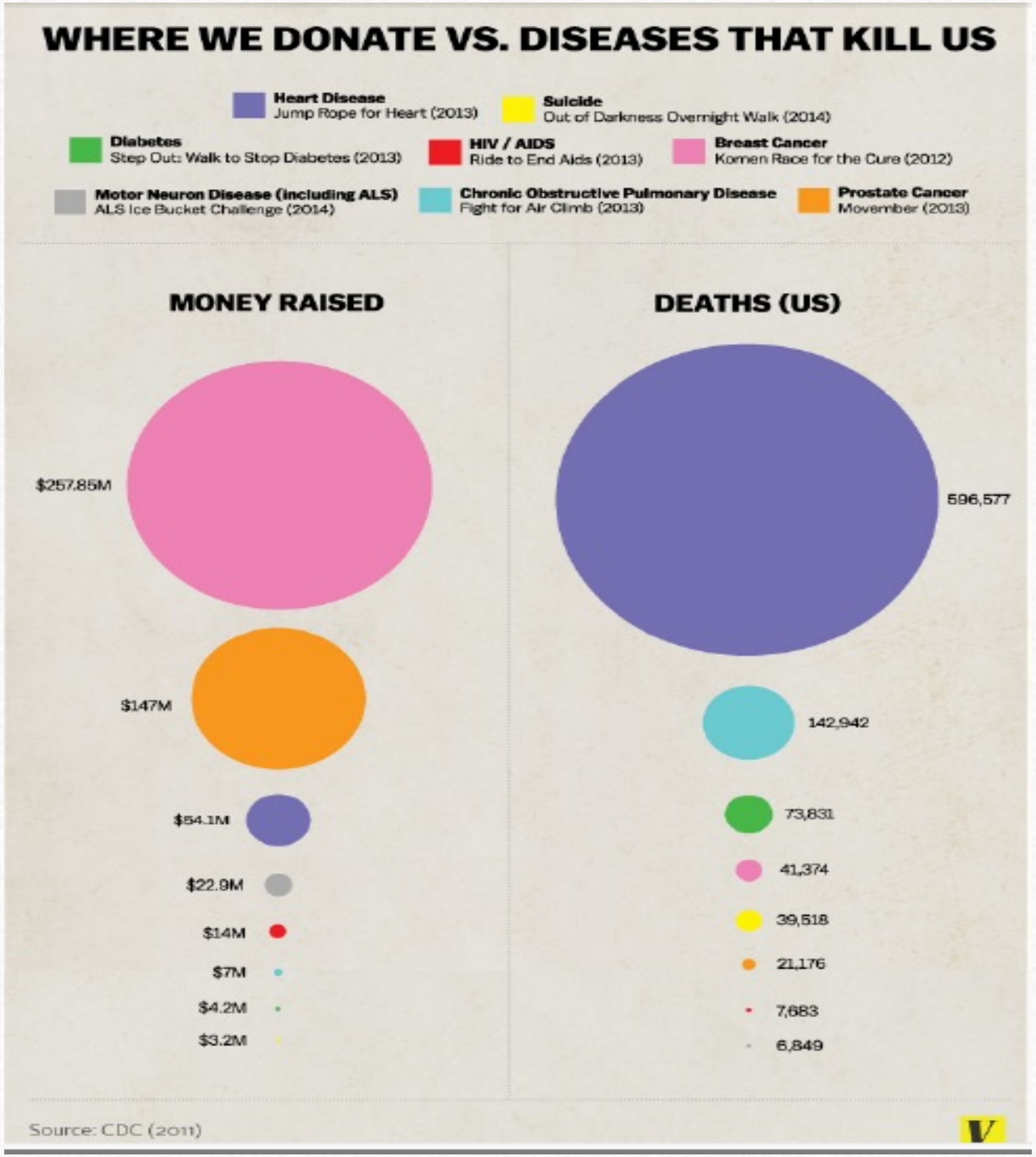
- ✧Range
- ✧Interquartile range
- ✧Variance

Measurement of location

Rate/proportion/ratio

Rate is a measure of the frequency with which an event occurs in a defined population in a defined time, e.g., number of deaths per hundred thousand Thai in one year. It has a time dimension, whereas a **proportion** does not, e.g., number of Thai with cancer divided by the total population.

Ratios. The value obtained by dividing one quantity by another: the male to female ratio. A ratio often compares two rates (the 'rate ratio'), for example comparing death rates for women and men at a given age.



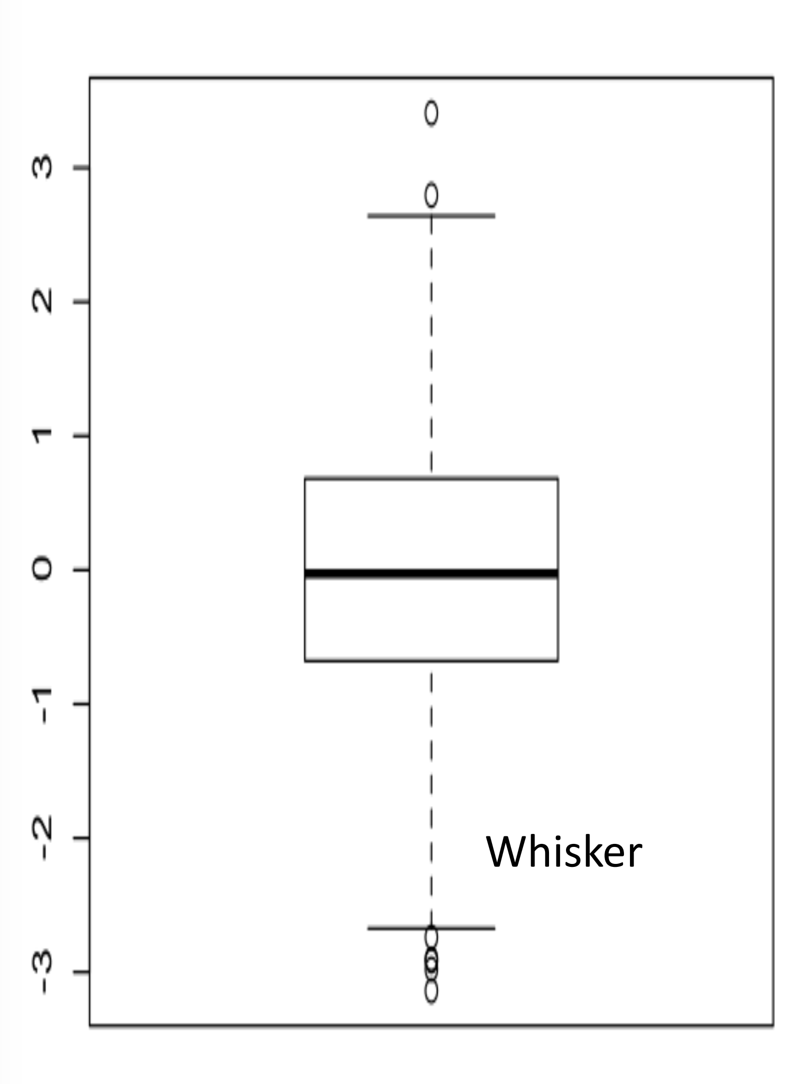
It is supposed to compare the relative amounts of funds raised for “disease research” vs. the relative number of deaths.

Boxplots

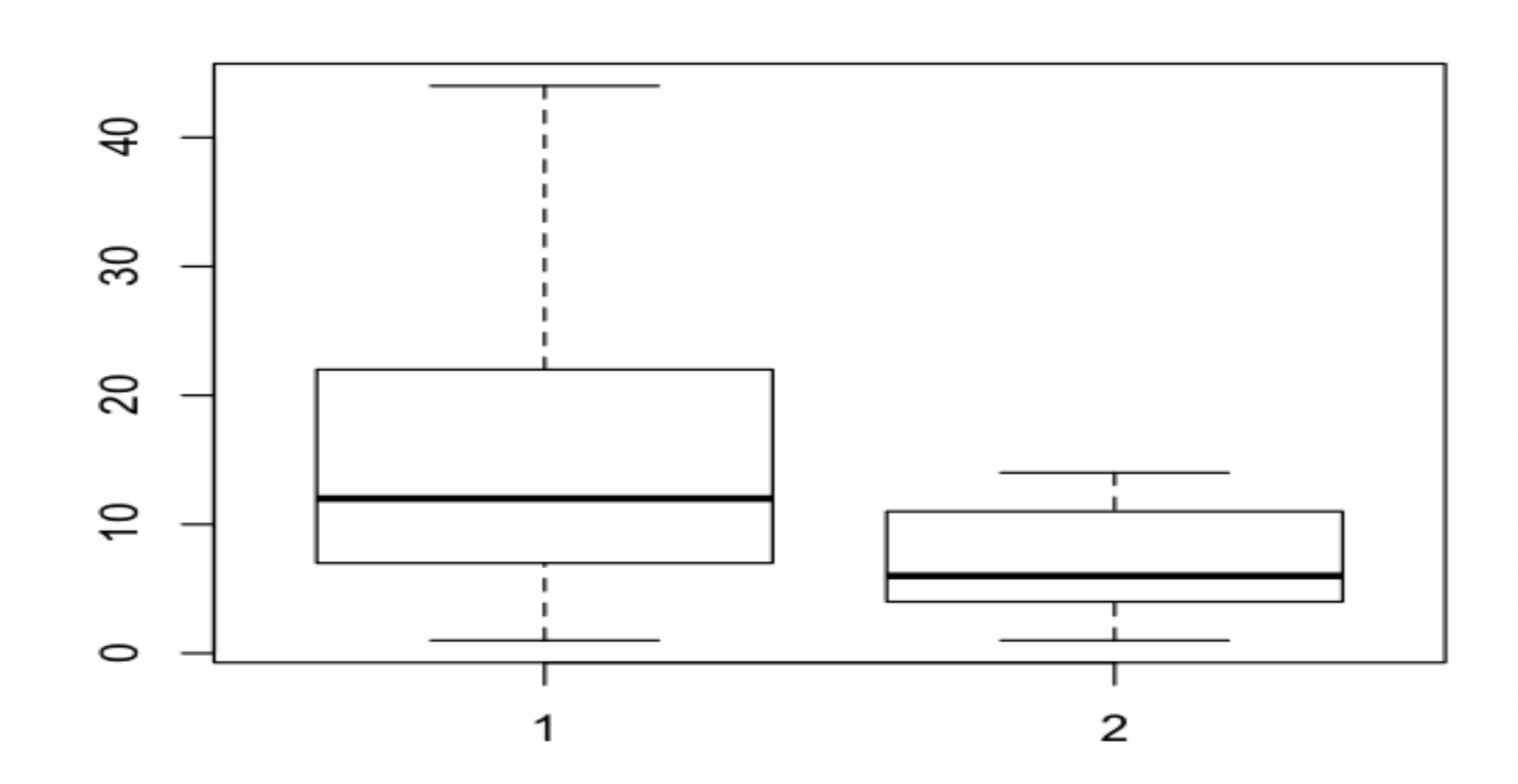
Whiskers: range of values that are not outliners

Outliner = Q1- 1.5 IQR and Q3 +1.5 IQR

Extreme = Q1 -3 IQR   and Q3+3 IQR



Boxplot



Displaying data well การนำเสนอที่ดี

- Be accurate and clear ชัดเจน
- Let the data speak ให้ข้อมูลอธิบายตัวเอง
  - Show as much information as possible, taking care not to obscure the message.
- Science no sales.เป็นวิทยาศาสตร์
  - Avoid unnescessary frills- esp. gratuitous 3D
- In tables, every digit should be meaningful. Don’t drop ending 0’s. ทุกทศนิยมมีความหมาย

- Source: H Wainer, How to display data badly, American Statistician,1984, 38(2):137-147

Guidelines for tabular presentation
การนำเสนอตาราง

- ชื่อตารางชัดเจน
- หัวตารางทั้งแนว row/column ชัดเจน
- การเรียบลำดับตัวแปร ตามความมากน้อย/พื้นที่/ตัวอักษร
- ตารางไม่ซับซ้อนเกินไป
- ไม่ทับซ้อนกันของตัวแปร
- จำนวนกลุ่มตัวแปรไม่มากหรือน้อยไป

Guideline for diagram presentation
การนำเสนอด้วยแผนภาพ

- Simple เรียบง่าย
- Easy to understand เข้าใจง่าย
- Save a lot of words ประหยัดคำพูด
- Self explanatory อธิบายตัวเองได้
- Clear title indicating its content มีชื่อแผนภาพที่ชัดเจน
- Fully labeled อธิบายทุกตัวแปร
- Use a good scale การใช้แกนของกราฟที่เหมาะสม

Table

Table 4. Population variation in hatch success (mean percent) of unfertilized eggs for females from populations sampled in1997. N = number of females tested. <--Table legend

<--Column titles

Population	mean (%)	Standard deviation	Range	N
Beaver Creek [T]	7.31	13.95	0-53.16	15
Honey Creek [T]	4.33	7.83	0-25.47	11
Rock Bridge Gans Creek [T]	5.66	13.93	0-77.86	38
Cedar Creek [P]	6.56	9.64	0-46.52	64
Grindstone Creek [P]	8.56	14.77	0-57.32	19
Jacks Fork River [P]	5.28	8.28	0-30.96	28
Meramec River [P]	5.49	10.25	0-45.76	45
Little Dixie Lake [L]	7.96	14.54	0-67.66	71
Little Prairie Lake [L]	6.86	7.84	0-32.40	36
Rocky Forks Lake [L]	3.31	4.12	0-16.14	43
Winegar Lake [L]	10.73	17.58	0-41.64	5
Whetstone Lake [L]	7.36	12.93	0-63.38	57

<--Table body (data)

<--Lines demarcating the different parts of the table

[T] = temporary stream, [P] = permanent streams, [L] = lakes. <--footnotes

Table 1 Number and Percentage of General Data among Female Academic University Employees (N= 2,000).

Data	Number	Percentage
Number of female academic university : Groups of new university	1,000	50.00
: Groups of old university	1,000	50.00
Age (years) : 24-33	396	19.80
: 34-43	938	46.90
: 44-53	746	37.30
Sex : Female	2,000	100.00
Education : Master degree	1,800	90.00
: Doctoral degree	200	10.00
Periods of duty (year) : 0-1	80	4.00
: 2-3	120	6.00
: 4-5	400	20.00
: 6-7	800	40.00
: 8-9	300	15.00
: 10-11	200	10.00
: 12-13	100	5.00
: 14-15	0	0
Wages (baht) : 15,001-20,000	200	10.00
: 20,001-25,000	1,600	80.00
: 25,001-30,000	100	5.00
: 30,001-35,000	100	5.00

Table 2 Statistic Analysis among Female Academic University Employees (N= 2,000).

Variable	Mean	Minimum	Maximum	SD	Skewedness	Kurtosis
1. Age (year)	40	24	53	7.040	-0.086	-0.740
2. Periods of duty (year)	6.10	1	14	1.340	0.090	0.139
3. Wages (Baht)	20,001-25,000	15,001	35,000	0.589	1.456	4.358
4. Job and environmental conditions	2.55 (more)	1 (least)	4 (most)	0.447	0	2.016
5. Family support	1 (Some of the time)	0 (Hardly ever)	2 (Almost always)	0.472	-0.891	-0.256
6. Stress level	39 (Moderate)	22 (Low)	65 (Severe)	0.622	1.012	1.389

table

Table 2a: Mean intakes of milk, supplement and water and mean growth rates for four diets (artificial data). The table is poorly presented.

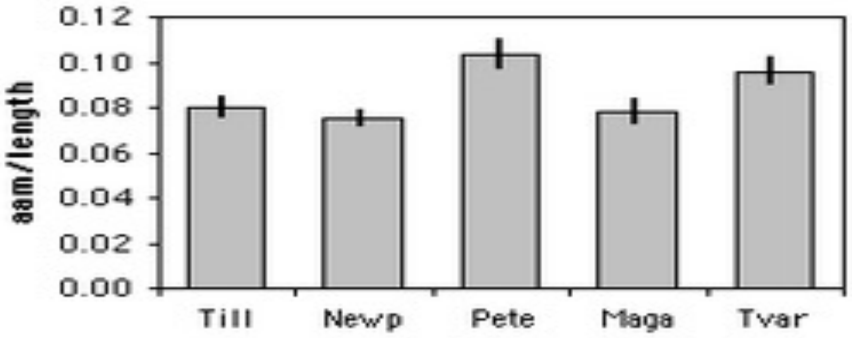
	Diet[1]			
Variable	I	II	III	IV
Milk Intake	9.82	10.48	8.9	9.15
Supplement intake	0	449.5	363.6	475.6
Growth rate	89	145.32	127.8	131.5
Water intake	108.4	143.6	121.29	127.8

[1] Diet I = Control; Diet II = Lucerne supplement;
Diet III = Leucaena; Diet IV Sesbania.

Table 2b: Mean growth rate and intakes of supplement, milk and water for four diets.

Supplement	Growth rate (g/day)	Supplement intake (g/day)	Milk intake (ml/kg$^{0.75}$)	Water intake (ml/kg$^{0.75}$)
Lucerne	145	450	10.5	144
Sesbania	132	476	9.2	128
Leucauna	128	364	8.9	121
None	89	0	9.8	108

Graph or table



Length of the anterior adductor muscle scar divided by total length in *Mytilus trossulus*. Means ±one standard error are shown for five locations.

Length of the anterior adductor muscle scar divided by total length in *Mytilus trossulus*. SE: standard error. N: sample size.

Location	Mean AAM/ length	SE	N
Tillamook	0.080	0.0038	10
Newport	0.075	0.0030	8
Petersburg	0.103	0.0061	7
Magadan	0.078	0.0046	8
Tvarminne	0.096	0.0053	6

chart

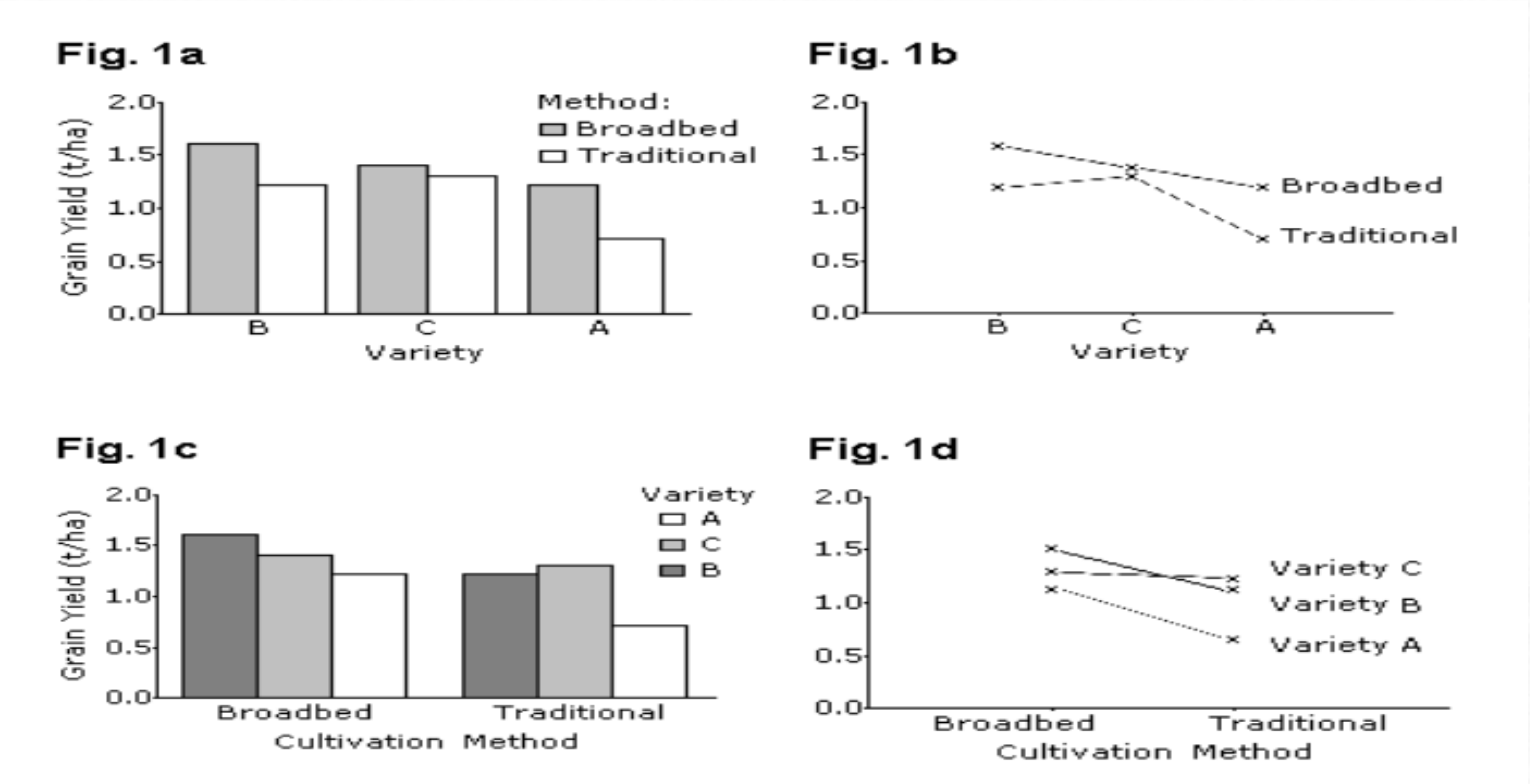


Fig. 1a
Grain Yield (t/ha)
2.0
1.5
1.0
0.5
0.0
B
C
A
Variety
Method:
Broadbed
Traditional
Fig. 1b
2.0
1.5
1.0
0.5
0.0
B
C
A
Variety
Broadbed
Traditional
Fig. 1c
Grain Yield (t/ha)
2.0
1.5
1.0
0.5
0.0
Broadbed
Traditional
Cultivation Method
Variety
A
C
B
Fig. 1d
2.0
1.5
1.0
0.5
0.0
Broadbed
Traditional
Cultivation Method
Variety C
Variety B
Variety A

Qualitative data (categories data)

- Percent/rate/proportion

A	Group A		Group B	
	Number	Percent	Number	Percent
total	30	100	80	100
Knowledge				
Good	25	32.5	52	67.5
Modera	3	13.0	20	87.0
Poor	2	20.0	8	80.0

B	Group A		Group B	
	Number	Percent	Number	Percent
total	30	100	80	100
Knowledge				
Good	25	83.3	52	65.0
Modera	3	10.0	20	25.0
Poor	2	6.7	8	10.0

A cross-sectional study of exposure to mercury in schoolchildren living near the eastern seaboard industrial estate of Thailand

Punthip Teeyapant, Siriwan Leudang, Sittiporn Parnmen

The study was carried out in four senior schools at four sites (S1 to S4) near the MTPIEs in the Rayong province of Thailand. Sites S1 and S2 are located closer to the petrochemical and chemical industrial plants than sites S3 and S4.

The 873 schoolchildren (405 boys and 468 girls) aged 6–13 years (grades 1 to 6). The participants were divided into two different age groups, with 327 children in the younger age group (6–9 years) and 546 children in the older age group (10–13 years).

Table 1: Whole blood total mercury levels (Hg-B) from schoolchildren living around Map Ta Phut Industrial Estate, Thailand					
Site, sex and age	*n*	Mean ± SD (µg/L)	95% CI	Median[a] (IQR)	<LOQ[b,c] (n)
S1					
Boys					
6–9 years	39	1.5 ± 0.9	1.17 to 1.75	0.75 (0.75 to 2.17)	22
10–13 years	61	2.3 ± 2.0	1.76 to 2.92	1.80 (0.75 to 2.80)	21
Girls					
6–9 years	36	2.7 ± 2.6	1.79 to 3.57	2.05 (1.55 to 2.65)	8
10–13 years	82	2.5 ± 1.2	2.17 to 2.73	2.20 (1.80 to 3.00)	11
S2					
Boys					
6–9 years	30	1.9 ± 1.3	1.47 to 2.41	1.80 (0.75 to 2.50)	11
10–13 years	59	2.7 ± 2.9	1.96 to 3.46	2.30 (1.60 to 3.00)	8
Girls					
6–9 years	32	2.0 ± 1.0	1.66 to 2.35	1.95 (1.50 to 2.55)	7
10–13 years	72	2.3 ± 1.0	2.04 to 2.53	2.20 (1.65 to 2.80)	8

1. อธิบายและเปรียบเทียบค่า median and IQR ของเด็กจากโรงเรียน S1
2. เด็กชาย อายุ 10-13 ปี ของโรงเรียน S1 มีปริมาณสารปรอทต่างจากเด็กชาย อายุ 10-13 ปี ของโรงเรียน S2 หรือไม่
3. เด็กหญิง อายุ 6-9 ปี กับเด็กชายอายุ 6-9 ปี ของโรงเรีย S1 มีปริมาณสารปรอทต่างกันหรือไม่

Prevalence of household drinking-water contamination and of acute diarrhoeal illness in a periurban community in Myanmar

Su Latt Tun Myint, Thuzar Myint, Wah Wah Aung, Khin Thet Wai

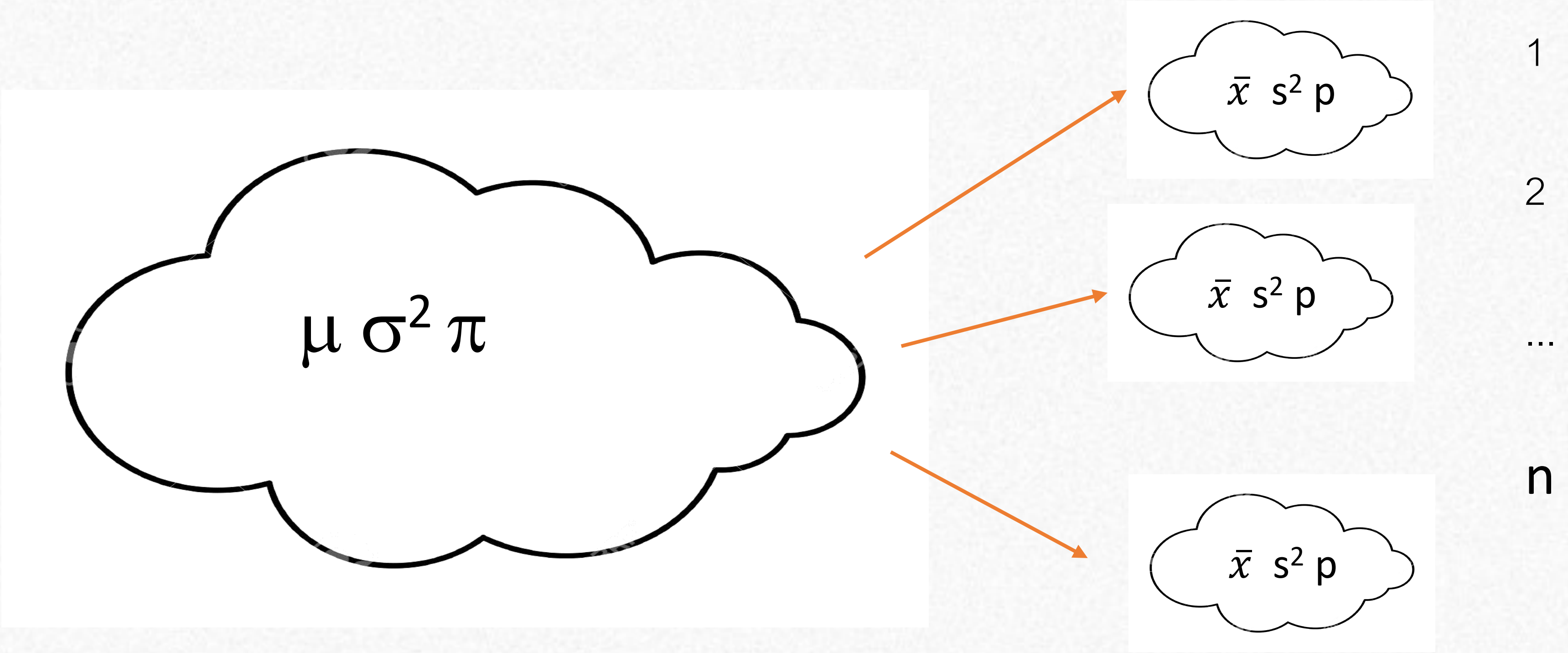
Table 2: Drinking-water management in study households		
Characteristic	Number	Percentage
Source of drinking water	(*n* = 211)	
Reservoir	74	35
Tube well only	58	27.5
Purified bottled water only	78	37.0
Tube well and bottled water	1	0.5
Storage of drinking water[a]	(*n* = 211)	
Clay pots	59	28.0
Large plastic bottles (20 litre)	93	44.1
Ceramic jars	50	23.7
Small plastic bottles (1 litre)	17	8.1
Steel pots	2	0.9
Drinking-water purification	(*n* = 211)	
No	52	24.6
Yes	159	75.4
Households with drinking-water purification	(*n* = 159)	
Cloth filter	131	82.4
Boiling	53	33.3
Ceramic filter	24	15.0
Sedimentation	26	16.4
Chlorine products	4	2.5

[a] Some households used more than one type of storage vessel.

จากตาราง

1. จากการสำรวจพบว่าครัวเรือนที่มีน้ำดื่มที่ผ่านการกรองเท่ากับร้อยละ 75.4 หากเกณฑ์ของประเทศกำหนดให้ครัวเรือนมากกว่าร้อยละ 80 มีน้ำสะอาดดื่ม จากข้อมูลสามารถสรุปได้หรือไม่ว่าชุมชนที่สำรวจผ่านเกณฑ์
2. คำนวณค่า 95% CI ของสัดส่วนการมีน้ำสะอาดดื่มในครอบครัว

Sampling



Confidence Interval for μ

- $100(1-\alpha)\%$ CI for μ

$$\bar{Y} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

or

$$(\bar{Y} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{Y} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$

- Values of $z_{1-\alpha/2}$

α	$z_{1-\alpha/2}$
0.10	1.645
0.05	1.960
0.01	2.576

Confidence Interval for μ

- If we draw 100 different random samples, on average 100(1-α)% of them will contain μ.
- To decrease the width of CI – decrease α

  - increase sample size

Why we need to show sample size in ethical document?

Confidence Interval for mean

Normal	σ^2 known	n large	Confidence Interval
√	√	–	$\bar{Y} \pm z_{1-\alpha/2}(\sigma/\sqrt{n})$
–	√	√	$\bar{Y} \pm z_{1-\alpha/2}(\sigma/\sqrt{n})$
√	–	–	$\bar{Y} \pm t_{n-1,1-\alpha/2}(s/\sqrt{n})$
–	–	√	$\bar{Y} \pm z_{1-\alpha/2}(s/\sqrt{n})$
–	–	–	Transform; nonparametrics

Hypothesis testing

- Estimation is a sample statistics and make a statement about the population parameter. The confidence interval make probabilistic about an uncertain of parameter.
- Hypothesis testing, starting with assuming a value for a parameter and probability statement is made about the value of the corresponding statistics

Introduction to hypothesis testing

- A hypothesis may be defined as a statement about one or more population.
- The hypothesis is frequently concerned with the parameters of the populations about which statement is made.
- Ex.
  - The average length of stay of patients admitted to the hospital is 5 days.
  - The education program associated with healthy lifestyle profile.

Null hypothesis

- Null hypothesis, H_0, the hypothesis to be tested.

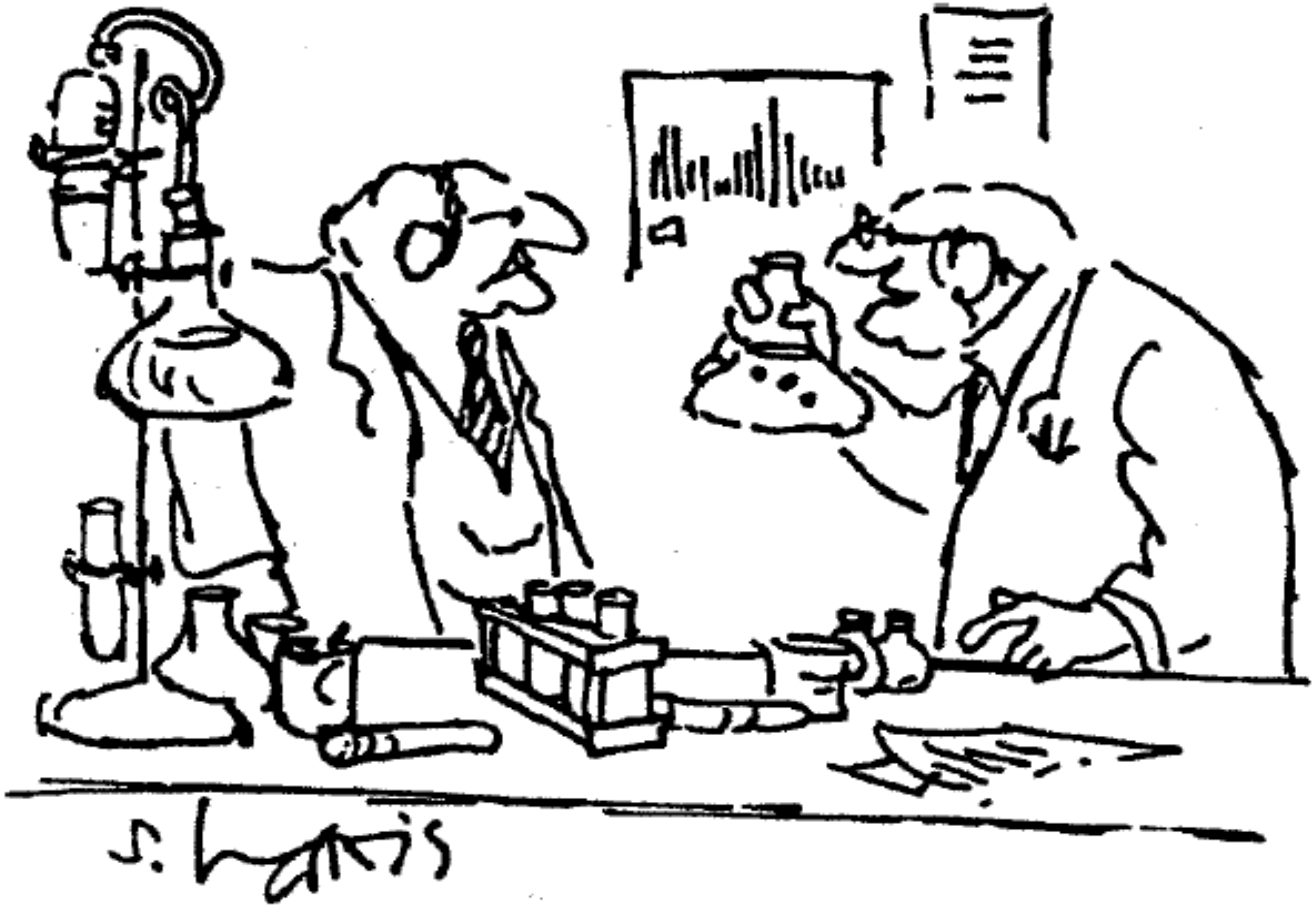
How do you decide which of two hypotheses is the null?

- The null hypothesis the hypothesis of “indifference”—which is a poor scientific attitude

Interpretation

- Usually if the probability is small, we conclude H_0 is not true– we 'reject' H_0
- If the probability is large, we have not proved H_0– we 'failed to reject H_0'

We can never prove H_0 is true and we never accept the alternative hypothesis.



"It may very well bring about immortality, but it will take forever to test it."

Cartoon 4.1 Testing some hypotheses can be tricky. (From *American Scientist*, March–April 1976.)

How to report results?

- Descriptive statistics make reader to understand your sample or population of study.
- Inferential statistics: apply estimation or hypothesis testing to infer statistics to parameter.

How to select type of statistics

- fit to measurement scale of outcome variable
- achieved goal/objective of study or can answer research questions

- garbage in garbage out: statistics cannot improve quality of data.
- The strong research methodology is important!

Example

วัตถุประสงค์การวิจัย

- เพื่อประมาณระยะเวลาการเข้ารับบริการในแผนกผู้ป่วยนอกของรพ.แห่งหนึ่ง
- เพื่อหาความสัมพันธ์ระหว่างระยะเวลาการเข้ารับบริการ กับ ปัจจัยส่วนบุคคล ปัจจัยและด้านการแพทย์
- เพื่อทำนายระหว่างระยะเวลาการเข้ารับบริการ

สถิติที่ใช้

Mean/ SD/ median…etc

ตัวแปรต้นเป็น continuous

Correlation/univariate linear regression

ตัวแปรต้นเป็น categories

T-test/z-test/one-way anova

multivariate linear regression (confounding control)

+show percent prediction or model fit/ sensitivity analysis…etc

Example

วัตถุประสงค์การวิจัย

- เพื่อประมาณร้อยละของเจ้าหน้าที่หน้าด่านโควิด19ที่เกิดภาวะ Burnout

- เพื่อหาความสัมพันธ์ระหว่างการเกิดภาวะ burnout กับปัจจัยส่วนบุคคล ปัจจัยการทำงานและปัจจัยสิ่งแวดล้อม

- เพื่อทำนายการเกิดburnout ของเจ้าหน้าที่หน้าด่าน

สถิติที่ใช้

ร้อยละ (#burnout/#total)

Chi-square/univariate logistics

multivariate logistic (confounding control)

+show percent prediction or model fit/ sensitivity analysis…etc

Confounding control methods

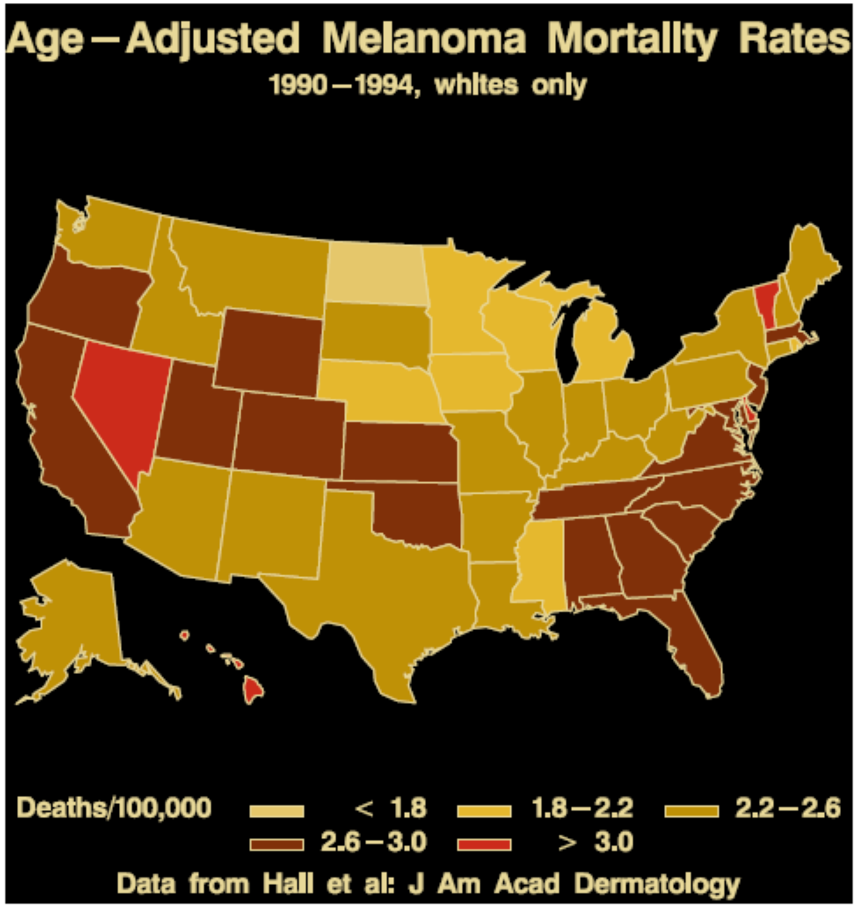
- Put variable into the model
- Stratification
- matching

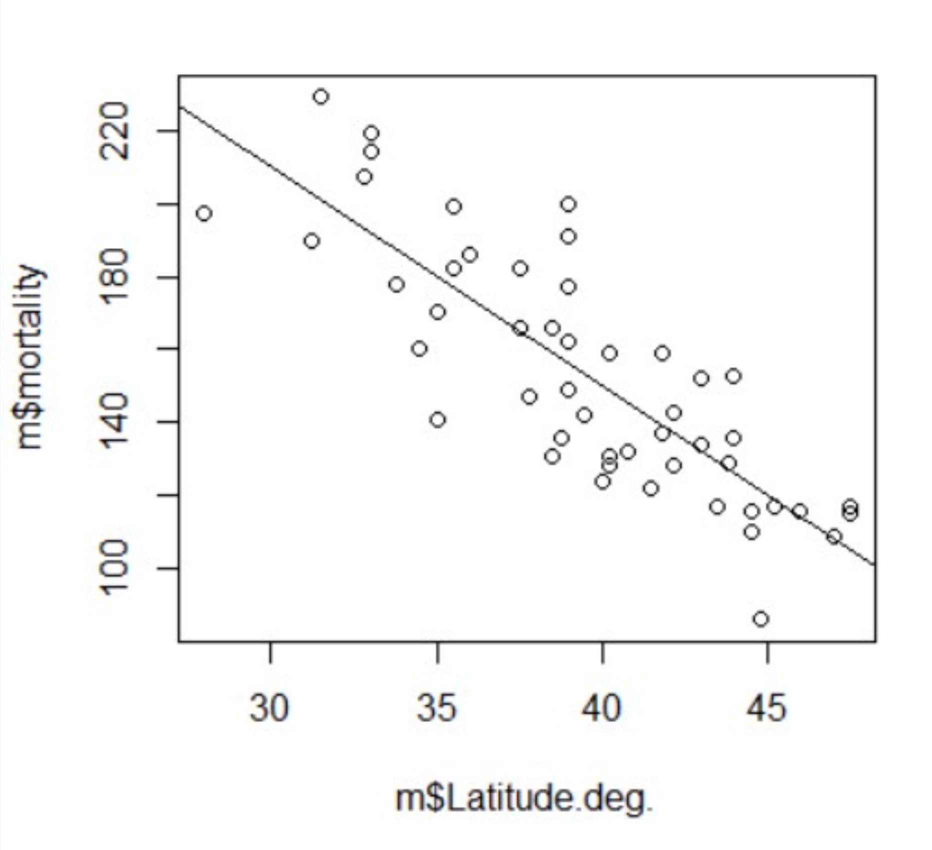
Assumption Checking

- Normality
- Equality of variance
- Etc…

P-value

- Probability of obtaining test statistic as unlikely or more unlikely than the observed test statistics if the null hypothesis is true.
- The probability of an observed result assuming that the null hypothesis is true





Guess!
How to interprete?

Mortality = 389.189 -5.978 Latitude

Mortality - 360.690 -5.489latitude+ 20.430 Ocean

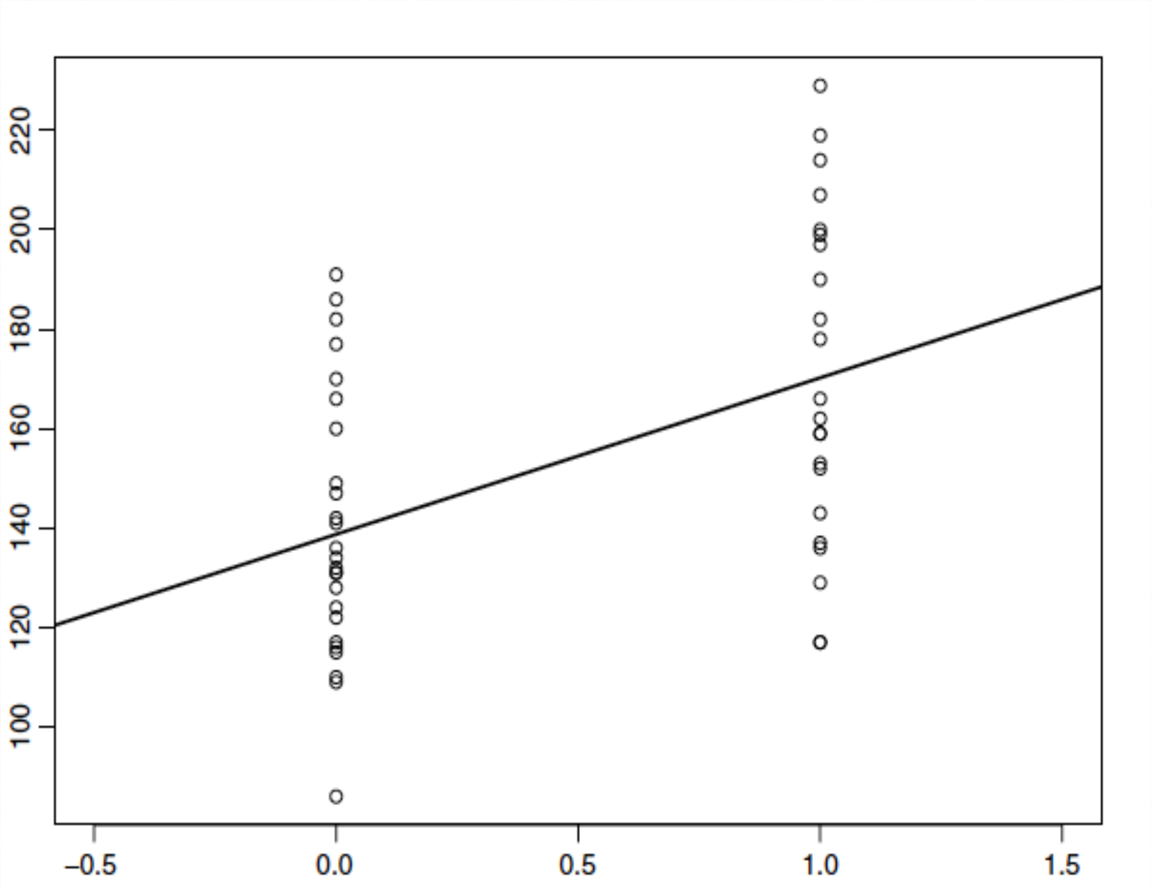
Parameters	parameter	SE	p-value
Intercept	-23.51	10.63	0.029
Sex (Ref : Male)			
Female	3.95	1.39	0.005*
Marital status (Ref : Widow Divorced Separated)			
Single	-5.84	2.63	0.028*
Marital	-0.19	1.76	0.910
Income sufficiency (Ref : Not enough)			
Sometimes enough	6.72	1.55	0.001*
Enough to save money	9.81	1.77	0.001*
Family relationship	0.38	0.15	0.015*

Outcome = score of knowledge about liver cancer

Female showed the significant higher average score of knowledge about liver cancer around 4 score compare to male .

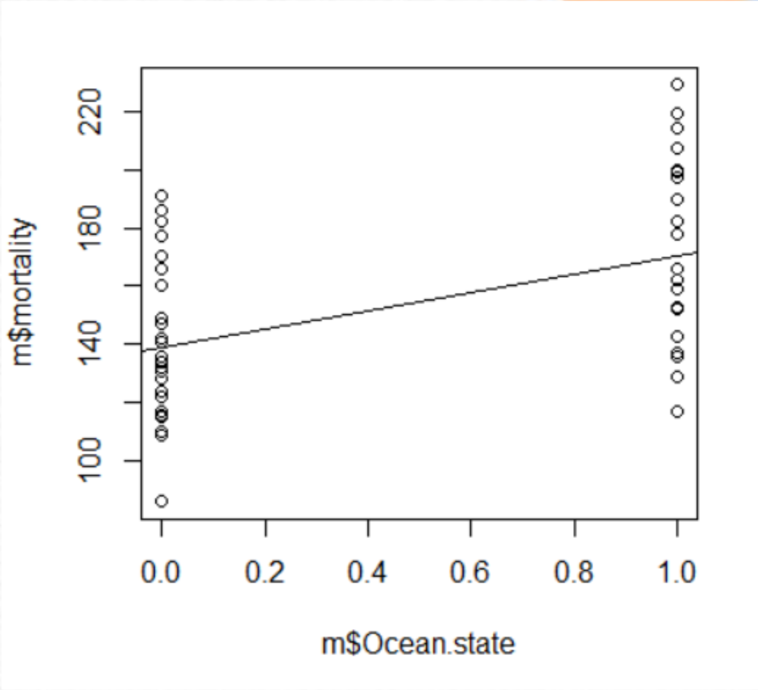
Increasing one score of family relationship, the average knowledge about liver cancer increased 0.38 score.

Linear regression and two sample t-test



By linear regression

Mortality = 138.74 + 31.49 Ocean

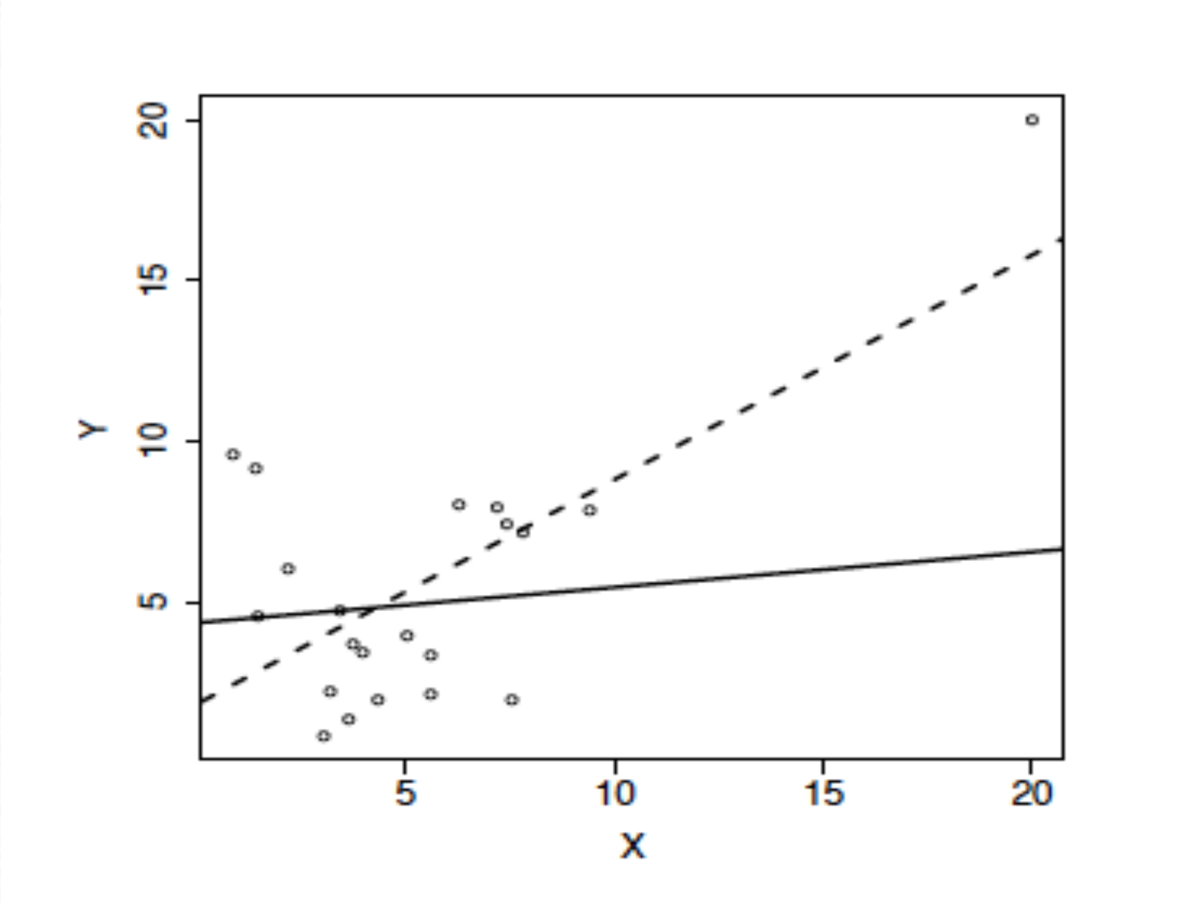


Welch Two Taster t-test

data: m$Ocean.state and m$mortality
t = -31.916, df = 48.022, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval:
-162.0312 -142.8259
sample estimates:
mean of x mean of y
0.4489796 152.8775510

Diagnostics

Example of data with possible extreme data.



Multiple selection strategy overview

Goal: to use purely exploratory analysis to find the best fitting model.

1) Specify the maximum model under consideration.

2) Specify criteria for model selection.

R^2, $adjustedR^2$, F-test, AIC, BIC, Mallow C_p

3) Specify a strategy for applying the criterion.

All possible model, Backward elimination, Forward selection, Stepwise selection.

4) Conduct the analysis.

ตารางที่ 4.23 เปรียบเทียบความสัมพันธ์ระหว่างปฏิสัมพันธ์ระหว่างบุคคล กับพฤติกรรมการกินอาหารเสี่ยง

ปฏิสัมพันธ์ระหว่างบุคคล	n	พฤติกรรมเสี่ยงสูง		Sig.
		จำนวน	ร้อยละ	
อิทธิพลจากครอบครัว				
สูง	156	124	64.2	
ปานกลาง	69	60	31.1	0.003
ต่ำ	18	9	4.7	
อิทธิพลจากเพื่อน				
สูง	57	50	26.0	
ปานกลาง	94	72	37.5	0.202
ต่ำ	91	70	36.5	

อิทธิพลจากครอบครัวมีความสัมพันธ์กับพฤติกรรมเสี่ยงในการกินอาหารสุกๆ ดิบๆ อย่างมีนัยสำคัญทางสถิติ โดยนักเรียนที่ได้รับอิทธิพลจากครอบครัวสูง ร้อยละ 64.2 มีพฤติกรรมเสี่ยงในการกินอาหาร ในขณะที่นักเรียนที่ได้รับอิทธิพลจากครอบครัวต่ำเพียงร้อยละ 4.7 ที่มีพฤติกรรมเสี่ยงในการกินอาหาร

ตารางที่ 4.6 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการรับรู้ความสามารถในการดูแลตนเอง การสนับสนุนทางสังคมของผู้ป่วยโรคมะเร็งตับ กับพฤติกรรมการดูแลตนเองระหว่างได้รับการรักษา

ตัวแปรที่ศึกษา	ค่าสัมประสิทธิ์สหสัมพันธ์(r)	P-value
การรับรู้ความสามารถในการดูแลตนเองกับพฤติกรรมการดูแลตนเองระหว่างได้รับการรักษา		
ด้านการรับประทาน	0.378	<0.01
ด้านการบริโภคเครื่องดื่ม	0.259	<0.01
ด้านการตรวจติดตามอาการตามนัด	0.119	0.114
ด้านการสังเกตอาการผิดปกติต่อตนเอง	0.094	0.212
การรับรู้ความสามารถในการดูแลตนเองโดยรวม	0.276	<0.01

การรับรู้ด้านการรับประทานมีความสัมพันธ์กับความสามารถในการดูแลตนเอง ร้อยละ 37.8

Table 4.4.1 Statistics of the Logistic Regression model for the level of low compliance

	OR	LB	UB	p-value
Gender (Ref.=Female)	1.585	0.200	1.986	0.431
Age	3.716	0.764	18.081	0.104
Marital status(Ref.= Married)	1.147	0.435	1.748	0.699
Education level (Ref.= high school and lower)	1.166	0.406	1.816	0.690

Male showed the higher risk of low compliance 1.6 times higher than female but the higher risks did not present the statistically significant (p-value=0.431).

Increasing one year of age, the risk of low compliance increased 3.716 times.

Logistic model

- the study population consists of people who visited a clinic on a walk-in basis and required a catheterization. The response, presence of coronary artery disease (CA), is dichotomous, as are the explanatory variables, sex and ECG (<0.1 ST segment depression/ other).

These data were analyzed were analyzed by Mantel-Haenzel method, ECG was clearly associated with disease status (CA) adjusted for sex.

CA data

Sex	ECG	Disease	No Disease	Total
Female	< 0.1 ST segment depression	4	11	15
Female	≥ 0.1 ST segment depression	8	10	18
Male	< 0.1 ST segment depression	9	9	18
Male	≥ 0.1 ST segment depression	21	6	27

h=0 for females, h=1 for males; i=0 for ECG < 0.1, i=1 for ECG 0.1;
j=1 for disease, j=0 for no disease

Model fitting

- For CA data, main effect is sex and ECG

$$\begin{bmatrix} \text{logit}(\theta_{11}) \\ \text{logit}(\theta_{12}) \\ \text{logit}(\theta_{21}) \\ \text{logit}(\theta_{22}) \end{bmatrix} = \begin{bmatrix} \alpha & & & & \\ \alpha & & & + & \beta_2 \\ \alpha & + & \beta_1 & & \\ \alpha & + & \beta_1 & + & \beta_2 \end{bmatrix}$$

- α is log odds of CA for females with ECG <0.1 (reference)
- β1 is the increment in log odds for male
- β2 is in increment in log odds for ECT<=0.1

Model predicted probability and Odds

Sex	ECG	Pr{CA Disease}=θ_{hi}	Odds of CA Disease
Females	< 0.1	$e^{\alpha}/(1+e^{\alpha})$	e^{α}
Females	≥ 0.1	$e^{\alpha+\beta_2}/(1+e^{\alpha+\beta_2})$	$e^{\alpha+\beta_2}$
Males	< 0.1	$e^{\alpha+\beta_1}/(1+e^{\alpha+\beta_1})$	$e^{\alpha+\beta_1}$
Males	≥ 0.1	$e^{\alpha+\beta_1+\beta_2}/(1+e^{\alpha+\beta_1+\beta_2})$	$e^{\alpha+\beta_1+\beta_2}$

Odds ratio

- Odds ratio for males versus female for either low or high ECG is

$$\frac{e^{\alpha+\beta_1}}{e^{\alpha}} = e^{\beta_1} \quad \text{or} \quad \frac{e^{\alpha+\beta_1+\beta_2}}{e^{\alpha+\beta_2}} = e^{\beta_1}$$

- Odds ratio for high ECG versus low ECG for either sex is

$$\frac{e^{\alpha+\beta_1+\beta_2}}{e^{\alpha+\beta_1}} = e^{\beta_2} \quad \text{or} \quad \frac{e^{\alpha+\beta_2}}{e^{\alpha}} = e^{\beta_2}$$

Interpretation of main effects model

- Model can be written a $\text{logit}(\theta_{hi}) = -1.1747 + 1.2770\ \text{SEX} + 1.0545\ \text{ECG}$

Parameter	Estimate	Standard Error	Interpretation
α	-1.1747	0.485	log odds of coronary disease for females with ECG < 0.1
β_1	1.2770	0.498	increment to log odds for males
β_2	1.0545	0.498	increment to log odds for high ECG

Sex	ECG	Logit	Odds of Coronary Artery Disease
Female	< 0.1	$\hat{\alpha} = -1.1747$	$e^{\hat{\alpha}} = e^{-1.1747} = 0.3089$
Female	≥ 0.1	$\hat{\alpha} + \hat{\beta}_2 = -0.1202$	$e^{\hat{\alpha}+\hat{\beta}_2} = e^{-0.1202} = 0.8867$
Male	< 0.1	$\hat{\alpha} + \hat{\beta}_1 = 0.1023$	$e^{\hat{\alpha}+\hat{\beta}_1} = e^{0.1023} = 1.1077$
Male	≥ 0.1	$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 = 1.1568$	$e^{\hat{\alpha}+\hat{\beta}_1+\hat{\beta}_2} = e^{1.1568} = 3.1797$

Interpretation of main effects model

- Odds ratio of males compared to females is the ratio of predicted odds of CA disease for males versus females, which is $\exp(\beta 1)=\exp(1.277)=3.586$

Men in the study have 3.6 times higher odds for coronary artery disease than women in the study.

- Odds ratio for high and low ECG is the ratio of predicted odds of CA disease for high versus low ECG, which is $\exp(\beta 2)=\exp(1.054) = 2.87$
- Those persons with ECG>=1 have nearly three times the odds of coronary artery disease as those with ECG <1.