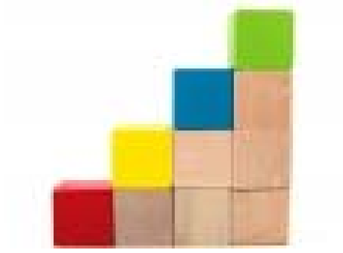
Course VI-2, 2015

Data Management & Descriptive Analysis

Aya Goto

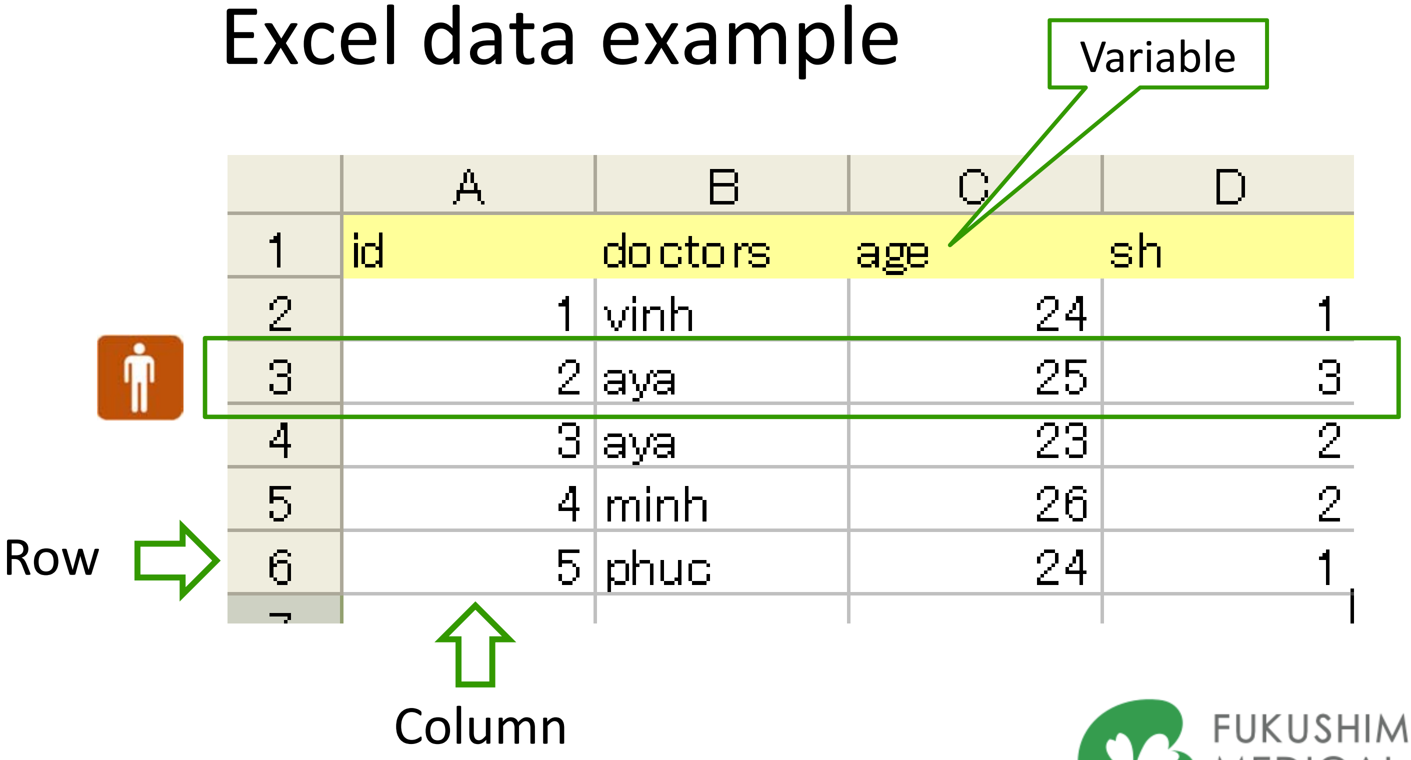
Department of Public Health

Fukushima Medical University



Data structure

Excel data example



	A	B	C	D
1	id	doctors	age	sh
2	1	vinh	24	1
3	2	aya	25	3
4	3	aya	23	2
5	4	minh	26	2
6	5	phuc	24	1
7				

Steps to develop a dataset

1. Check collected questionnaires for missing answers or mistakes
2. Prepare a list of codes
3. Enter data into computer
   DOUBLE CHECK!
4. Check frequencies of all variables, perform logic check and correct mistakes

Checking collected questionnaires

Examples

Subject A:

Q1. How many times did you take Pap smear test during the last 5 years?
(2-3) times

Subject B:

Q2. What do you think about your health?
(Circle one)
① Excellent ② Good
3. Fair 4. Poor

Notes

Q1. If ( )-( ), take the middle.
E.g. Correct 2-3 into 2.5.
(Other options: correct into 2, 3, or missing.)

Q2. If more than one answer, code as missing.

Cont.

Subject D:
Q3. Your sex
1. Female 2. Male
Q4. If female, how many times did you take Pap smear test during the last 5 years?
(2) times

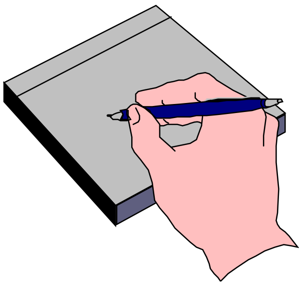
Subject E:
Q5. How frequently do you drink?
1. Less than once / week
[✓ 2. 1-2 times / week]
3. 3-4 times / week
4. Almost everyday

Q3. If sex is missing, but the respondent gives the number of Pap smear, code the person as female.
Q5. If more than one answer, take the smaller number.
(Example of the logic:
Drinking habit and infertility: When you have a result that drinking is a significant risk factor, you can be confident with your result that the result was significant even though you selected the smaller number in such cases.)

FUKUSHIMA
MEDICAL
UNIVERSITY

There will be many unexpected answers.
The way you clean the answers should always be recorded!

Do not change the way you clean during data processing.



FUKUSHIMA
MEDICAL
UNIVERSITY

List of codes

Q1. What do you think about your health? (Circle one)
1. Excellent 2. Good
3. Fair 4. Poor

Q2. How many times did you take Pap smear test during the last 5 years? ( ) times

Q3-1. Do you exercise regularly?
1. Yes 2. No

Q3-2. If yes, what kinds of exercise? (Circle all exercises you do)
1. Walking 2. Jogging
3. Swimming 4. Ball games
5. Dancing 6. Others

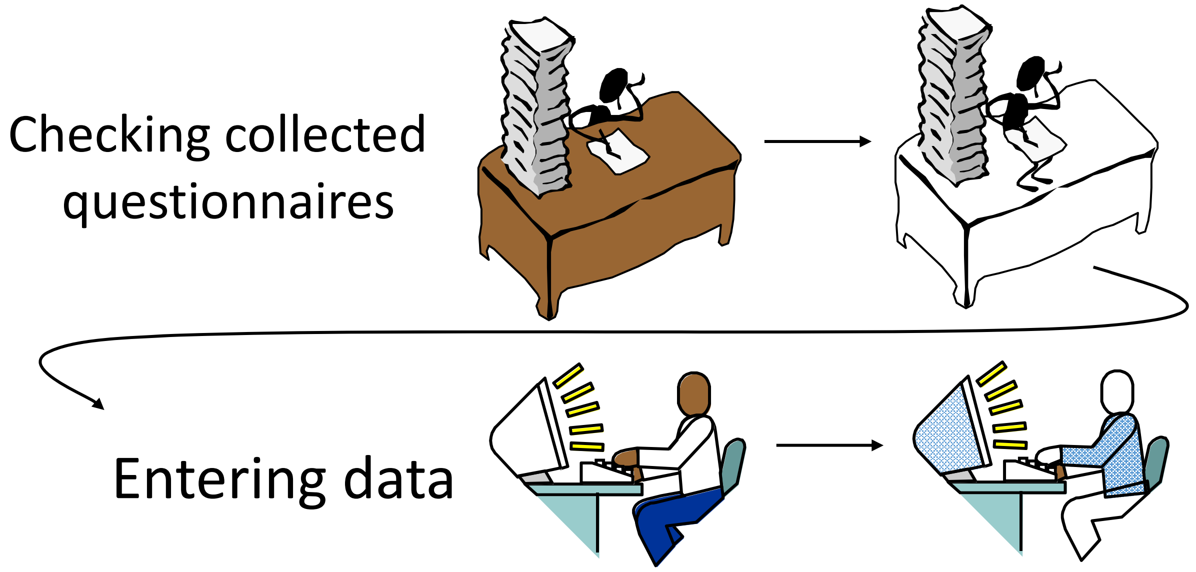
No.	Variables	Codes
Q1	sh **Single choice**	1=excellent 2=good 3=fair 4=poor
Q2	pap	years
Q3-1	exc	1=yes 0=no

Q3-2	exc1=walking
	exc2=jogging
	exc3=swimming
	exc4=ball games
	exc5=dancing
	exc6=others
		1=yes
		0=no
		Multiple choice

FUKUSHIMA MEDICAL UNIVERSITY

Data entry

Never forget to double check!



Checking collected questionnaires

Entering data

FUKUSHIMA MEDICAL UNIVERSITY

Tabulation (One-way)

Do not jump into analysis right after data entry.

Tabulate all variables first!!!

Sex

```
. tabulate sex
1(female) |
2( male)   | Freq. Percent  Cum.
------------+-------------------
        1 |   6   60.00   60.00
        2 |   3   30.00   90.00
      0.1 |   1   10.00  100.00
------------+-------------------
    Total |  10  100.00
```

Total number of pregnancies

```
. tabulate tp

total number of |
pregnancies | Freq.  Percent    Cum.
------------+-----------------------------------
          0 |       5      16.67       16.67
         .1 |       1       3.33       20.00
          1 |       6      20.00       40.00
          2 |      10      33.33       73.33
          3 |       6      20.00       93.33
          5 |       1       3.33       96.67
         60 |       1       3.33      100.00
------------+-----------------------------------
      Total |       30      100.00
```

Tabulation (Two-way)

Logic check is required for conditional questions.

Q12-1. Have you ever been pregnant before? (Circle one)

1. Yes   2. No

Q12-2. If yes, how many time?

Total (   ) times

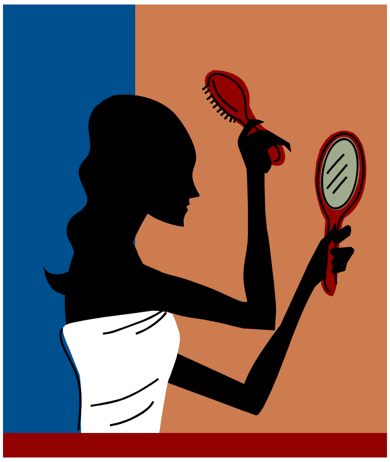
- Live birth (   ) times
- Still birth/miscarriage (   ) times
- Abortion (   ) times

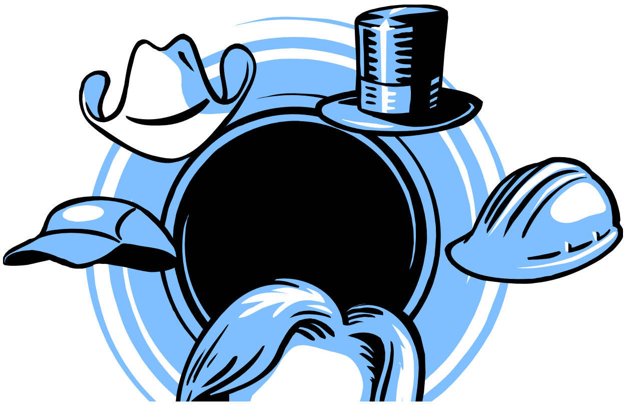
```
. tabulate  preg tn

1(yes)  |         total No. of pregnancies
  2(no) |       0        1        2        3 |     Total
-----------+------------------------------------------
      1 |       1        1        1        3 |         6
      2 |       0        1        0        0 |         1
-----------+-----------------------------------------+-
  Total |       1        2        1        3 |         7
```

FUKUSHIMA MEDICAL UNIVERSITY

Making Tables and Graphs

Aya Goto





IMPORTANT: Tables

- Let’s make this table with EXCEL.

Goto A, et al. Association of pregnancy intention with parenting difficulty in Fukushima, Japan. J Epidemiol. 2005;15(6):244-6.

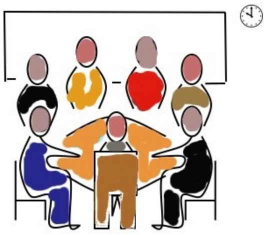
The median age of the 197 children in question was 10.5 months (min=3, max=19); 51% (N=51) were male; and 10 were born with low birth weight. The median age of their mothers was 29 (min=18, max=40); and 58% (N=114) were housewives.

Tips

- Categorize information.
- Write title
- Write headings in the top row.
- Format only with horizontal lines.
- Utilize indent to clarify hierarchy.
- Align numbers to the right.
- Distinguish categorical and continuous variables.

FUKUSHIMA MEDICAL UNIVERSITY

Graph



- Let's make this graph with EXCEL.

Goto A, Fujiyama-Koriyama C, et al. Abortion trends in Japan, 1975-95. Stud Fam Plann. 2000;31(4):301-8.

Figure 1 shows the incidence of abortion for all women (the abortion rate) between 1975-1995. The only age group in which the abortion rate increased was for women under 20 years old, increasing by 109.1% from 1975 to 1995. Women aged 20-24 years showed a lower reduction in abortion rate (32.9% decrease) than the reduction in women aged 25-39 years and 40-44 years (50.0% and 43.5% decrease, respectively) in the study period.

	under 20	20-24	25-39	40-44
1975	3.3	25.2	33.6	13.8
1980	4.8	23.7	30.0	12.4
1985	6.3	21.5	27.2	11.0
1990	6.4	19.8	22.6	9.9
1995	6.9	16.9	16.8	7.8

FUKUSHIMA MEDICAL UNIVERSITY

Table and figure formats

Table

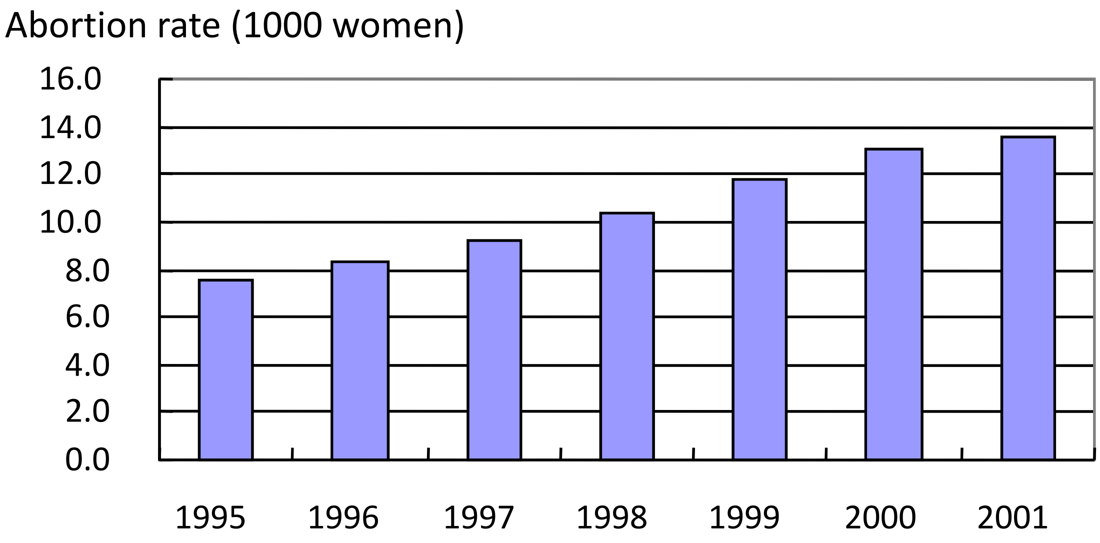
Table 1. Age-specific abortion rates and ratios in Japan and Fukushima, 2000

		Age group						
	Total	<20	20-24	25-29	30-34	35-39	40-44	45-49
Abortion rate (per 1000 women)								
Japan	11.6	12.2	20.1	15.1	14.2	13.2	6.2	0.5
Fukushima	20.7	18.4	30.8	23.8	22.9	19.9	10.2	0.7
Abortion ratio (per 1000 live births)								
Japan	285	2249	512	154	156	420	1624	5775
Fukushima	390	2627	462	203	237	607	2324	13500

FUKUSHIMA MEDICAL UNIVERSITY

Bar chart

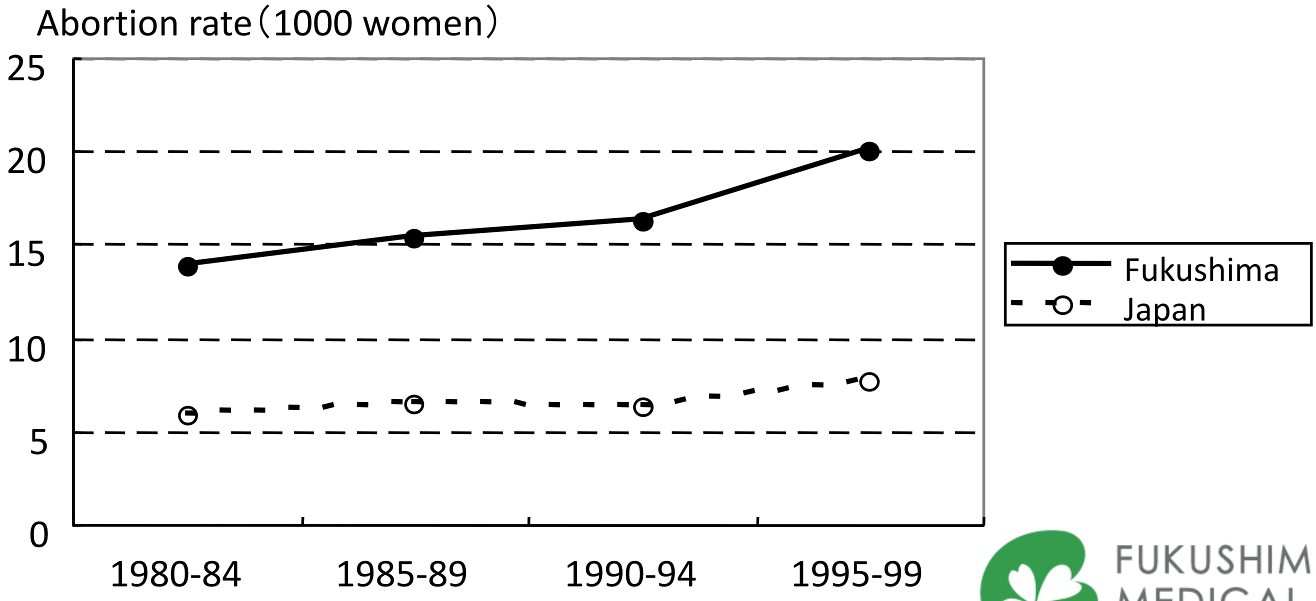
Figure 1. Trends in abortion rate, Japan, 1995-2001



FUKUSHIMA MEDICAL UNIVERSITY

Line graph

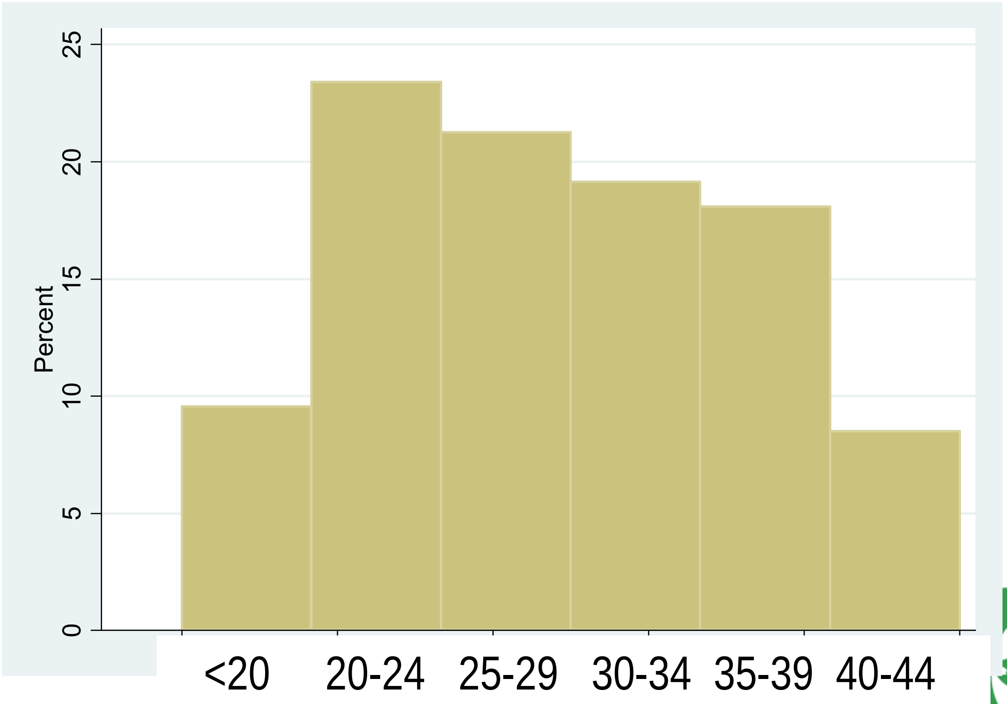
Figure 2. Trends in abortion rate,
Japan and Fukushima, 1980-1999



FUKUSHIMA MEDICAL UNIVERSITY

Histogram

Figure 3. Age-specific proportion of abortion cases, Japan, 1999



FUKUSHIMA MEDICAL UNIVERSITY

USAGE

❖Table	Precise data
❖Bar chart	Trend
❖Line graph	Trend (Comparing several groups)
❖Histogram	Distribution

FUKUSHIMA MEDICAL UNIVERSITY

Table 1. Characteristics of enrolled families

Characteristics	N(%) or *Median (min, max)* Total N=197
Mothers	
Age (years)	*29 (28, 40)*
Occupation	
Housewife	114 (58)
Employeed	83 (42)
Children	
Age (years)	*10.5 (3, 19)*
Sex	
Male	100 (51)
Female	97 (49)
Birth weight	
Less than 2500g	10 (5)
2500g or higher	187 (95)

FUKUSHIMA
MEDICAL
UNIVERSITY

Arithmetic rules can NOT be applied.

→ N (%)

- Categorical data

  Blood type: 1= O, 2=A, 3=B, and 4=AB

  Injury: 1=fatal, 2=severe, 3=moderate, 4=minor

Arithmetic rules can be applied.

→ Summary measures

- Continuous data

  Hb level

  Number of births that woman has given.

FUKUSHIMA
MEDICAL
UNIVERSITY

Summary measures of continuous data

- Mean=average

  Standard Deviation(SD)

  Mean±2SDs = a range in which "most" of your subjects fit. (About 95% lie within 2 SDs)

- Median=50th percentile

  Range=Min/Max

- Mode=most frequent value(s)

FUKUSHIMA MEDICAL UNIVERSITY

Age: 6, 6, 7, 8, and 25

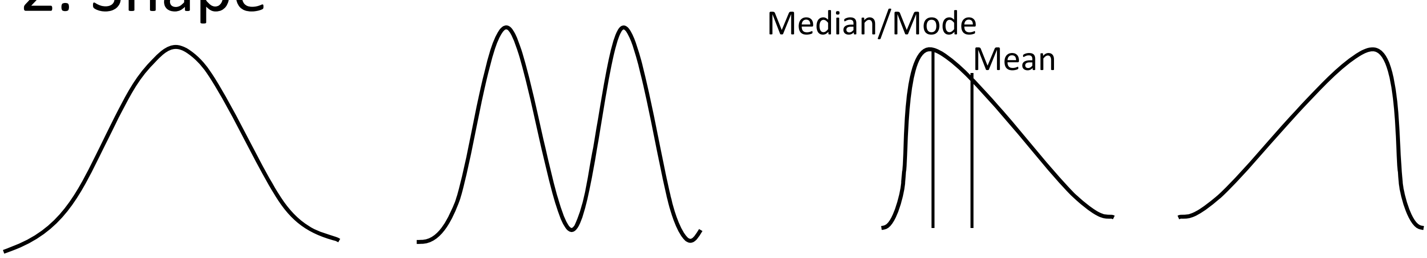
- Mean =
- Median =
- Mode =

FUKUSHIMA MEDICAL UNIVERSITY

Selection of summary measures

1. Sample size: Large (>30) --> Mean (SD)
Small --> Median (Min, Max)

2. Shape



Bell-shape
→ Mean
(= Median = Mode)

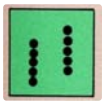
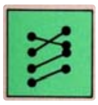
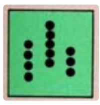
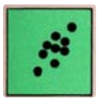
More than one peaks
→ Mode

Skewed to the right (left)
→ Median

Basic statistical tests

Aya Goto

Frequently used statistical tests

Data type		Parametric or large N	Non-parametric or small N
Contingency table	A B D + D -	Chi-square test	Fisher's exact test
Comparison of means			
(2 groups, independent)		T-test	Mann-Whitney U test
(2 groups, paired)		Paired t-test	Wilcoxon signed rank test
(≥3 groups, independent)		ANOVA	Kruskal-Wallis test
Correlation		Pearson's correlation	Spearman's correlation

Relationship of residence and prevalence of hypertension

	City A	City B
Disease positive	20	80
Disease negative	40	60

⬇

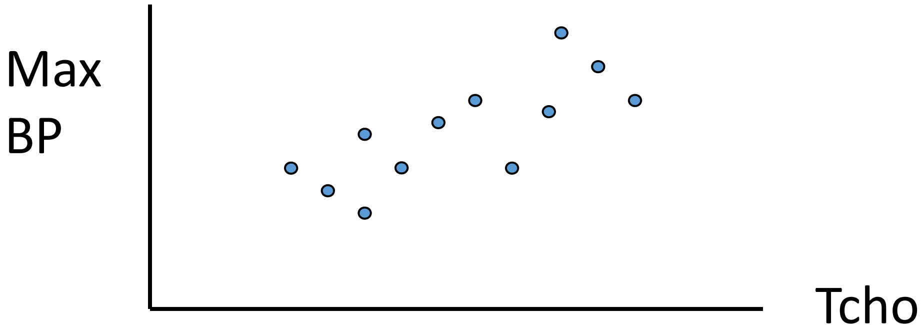
Analysis of contingency table

Relationship of residence and blood pressure level

	City A	City B
max BP(mean)	160	140

Comparison of means

Relationship of total cholesterol and blood pressure



Correlation

Paired or unpaired (independent) ?

Before-after study or
matched case-control study → **Paired**

Others → **Unpaired (independent)**

FUKUSHIMA MEDICAL UNIVERSITY

Para or Nonpara ?

Data type: Categorical

Sample size: small

Distribution (graph): not bell shape

Non-PARA

Data type: Not categorical

Sample size: large (>30)

Distribution (graph): bell shape

PARA

FUKUSHIMA MEDICAL UNIVERSITY

Data presentation

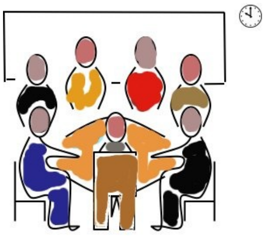
	Mean (SD)		
	City A N=200	City B N=1000	p-value*
Systolic blood pressure			
Total cholesterol			

* T-test was used.

	Median (min, max)		
	Village A N=10	Village B N=30	p-value*
Systolic blood pressure			
Total cholesterol			

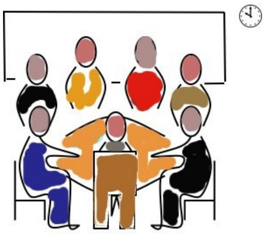
* Mann-Whitney U test was used.

- Let’s analyze the sample EXCEL data.



Sample data: Smoking survey among medical students in two countries.

Items: Country, age, sex, smoking status (1=smoker, 2=past smoker, 3=non-smoker), 10-item knowledge test (1=correct)



Assignments

1. Check distribution of age, sex and smoking status of students in each country.
2. Calculate a summary measure of the knowledge test score of each country, and perform a statistical test to examine the difference.
3. Develop tables and graphs to tell what you found.

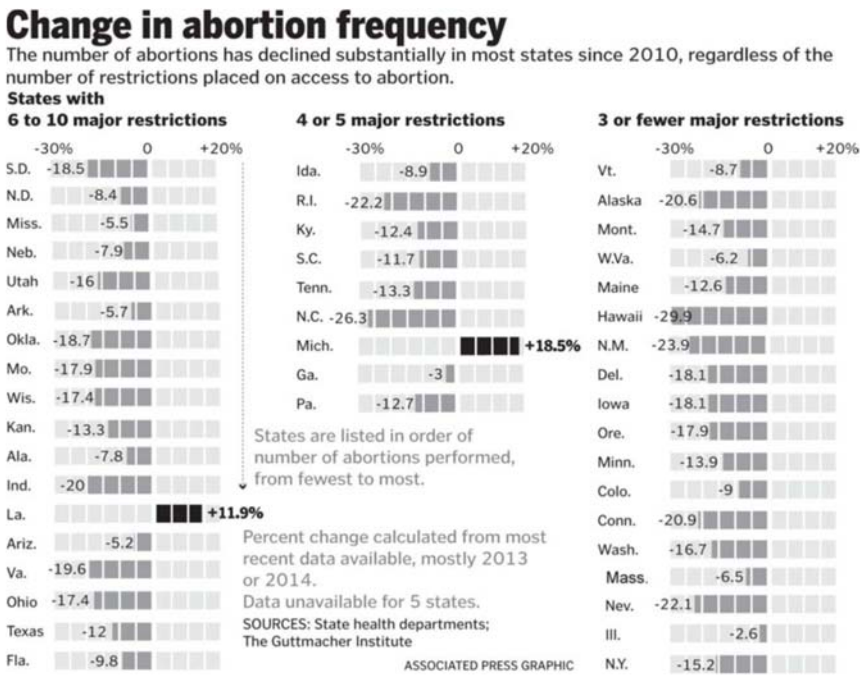
Additional assignment

Abortions declining greatly across most of US

Changes in laws do not appear to affect trend

ASSOCIATED PRESS JUNE 08 , 2015

NEW YORK — Abortions have declined in states where new laws make it harder to have them — but they've also waned in states where abortion rights are protected, an Associated Press survey finds.



Change in abortion frequency

The number of abortions has declined substantially in most states since 2010, regardless of the number of restrictions placed on access to abortion.

States are listed in order of number of abortions performed, from fewest to most.

Percent change calculated from most recent data available, mostly 2013 or 2014.

Data unavailable for 5 states.

SOURCES: State health departments; The Guttmacher Institute

ASSOCIATED PRESS GRAPHIC

Really?