Aug 22, 2019

Clinical epidemiology study
~ Analysis of clinical data~

Hirohide Yokokawa, M.D., Ph.D.
Department of General Medicine,
Juntendo University School of Medicine

Contents

- Research question
- Steps of data analysis
- Presentation of analyzed data
- Implications of obtained results

What is a research question?

- ✓ The research question sets out what you hope to learn about the topic.
- ✓ This question, together with your approach, will guide and structure the choice of data to be collected and analyzed.

http://www.socscidiss.bham.ac.uk/research-question.html

http://twp.duke.edu/uploads/media_items/research-questions.original.pdf

Research Question

Hypothesis

Specific Aim

Guidelines highlight some of the features of good questions

- Relevant
- Manageable in terms of research and in terms of your own academic abilities
- Substantial and with original dimensions
- Consistent with the requirements of the assessment
- Clear and simple
- Interesting

- **Relevant:** Arising from issues raised in literature and/or practice, the question will be of academic and intellectual interest.

- **Manageable:** You must be able to access your sources of data (be they documents or people), and to give a full and nuanced answer to your question.

- **Substantial and original**: The question should showcase your imaginative abilities, however far it may be couched in existing literature.

- **Fit for assessment**: Remember, you must satisfy the learning outcomes of your course. Your question must be open to assessment, as well as interesting.

- **Clear and simple**: A clear and simple research question will become more complex as your research progresses.

- **Interesting**: Make your question interesting, but try to avoid questions which are convenient or flashy.

Steps of clinical epidemiology

Descriptive study
（To know distribution and characteristics）

Analytic study
（To know associations）

Intervention study
（To know effectiveness）

Descriptive study

- **Checking distribution and characteristics of the participants**: Knowing the distribution and characteristics leads to adequate further analytic analysis.
- **Checking errors:** Data errors must be corrected before main analysis.

Analytic study

- **Estimating associations**: To know associations between outcome and factors.
- **Exploring associated factors:** To explore/identify factors associated with the outcome.

An example of a research question

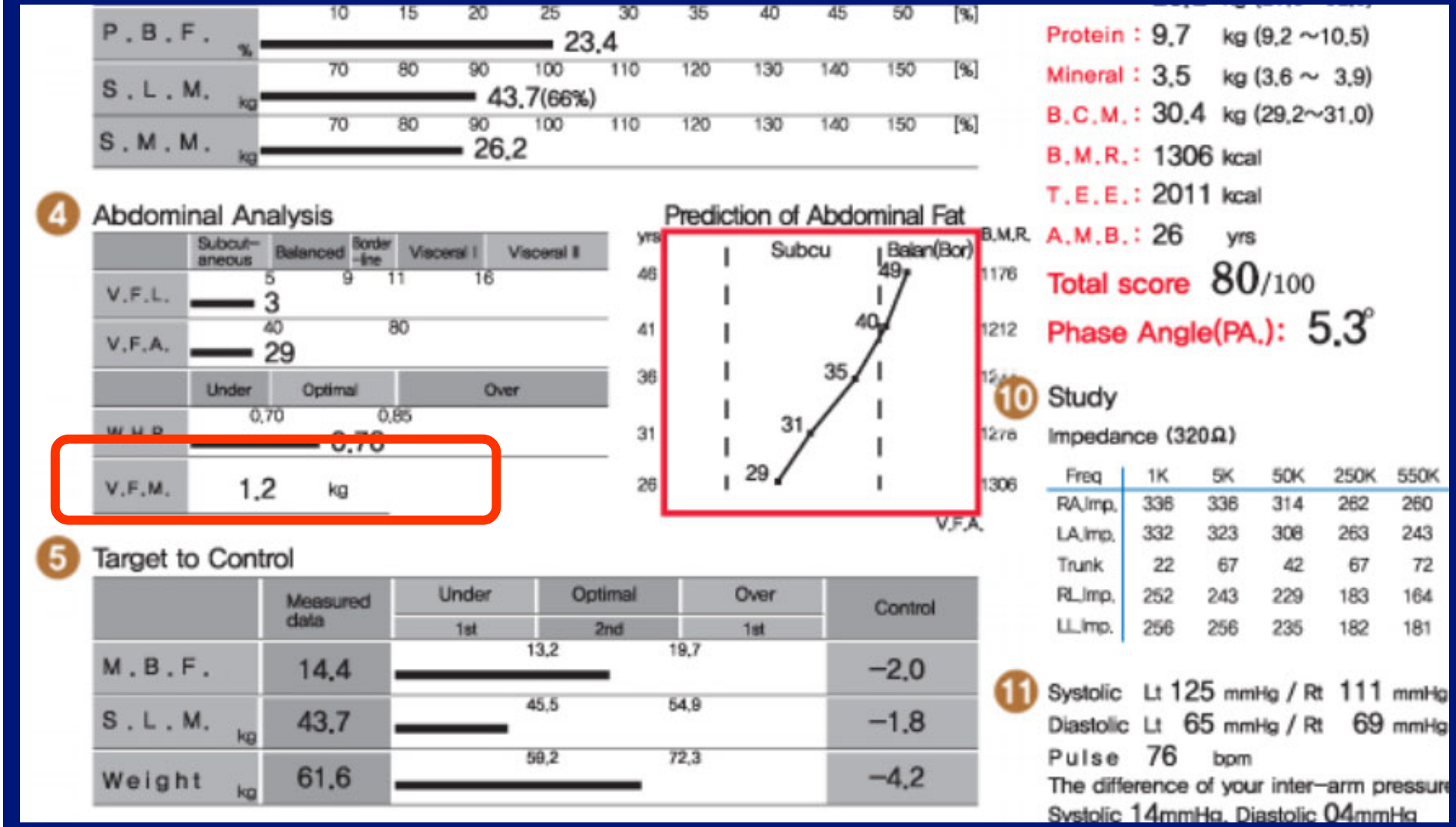
The association between waist circumference and lifestyle related disorders has been reported already.

However, there are few studies examining the association between actual visceral fat mass and lifestyle-related disorders.

Automated bioelectrical impedance analysis (BIA)

http://www.sowamedical.com/

Summary report



http://jawon.com/japanese/bca/x-scanplus950kr.html

Let’s make a research question and discuss how to analyze the data.

A research question

To clarify the association between visceral fat mass and lifestyle-related disorders, and estimate an appropriate cutoff value for visceral fat mass that associates with an increased risk of developing lifestyle-related disorders.

What is the first step to analyze the clinical data?

www.spanease.com

- **Establish the inclusion and exclusion criteria**
- **Checking distribution and characteristics of the participants**
- **Checking errors**

Inclusion and exclusion criteria

Inclusion criteria; Completed voluntary medical check-ups, and had their amounts of visceral fat measured using a BIA automated body composition analyzer.

Exclusion criteria; Use of medications for hypertension, dyslipidemia, or diabetes mellitus, and/or past history of cardiovascular disease or cerebrovascular disease.

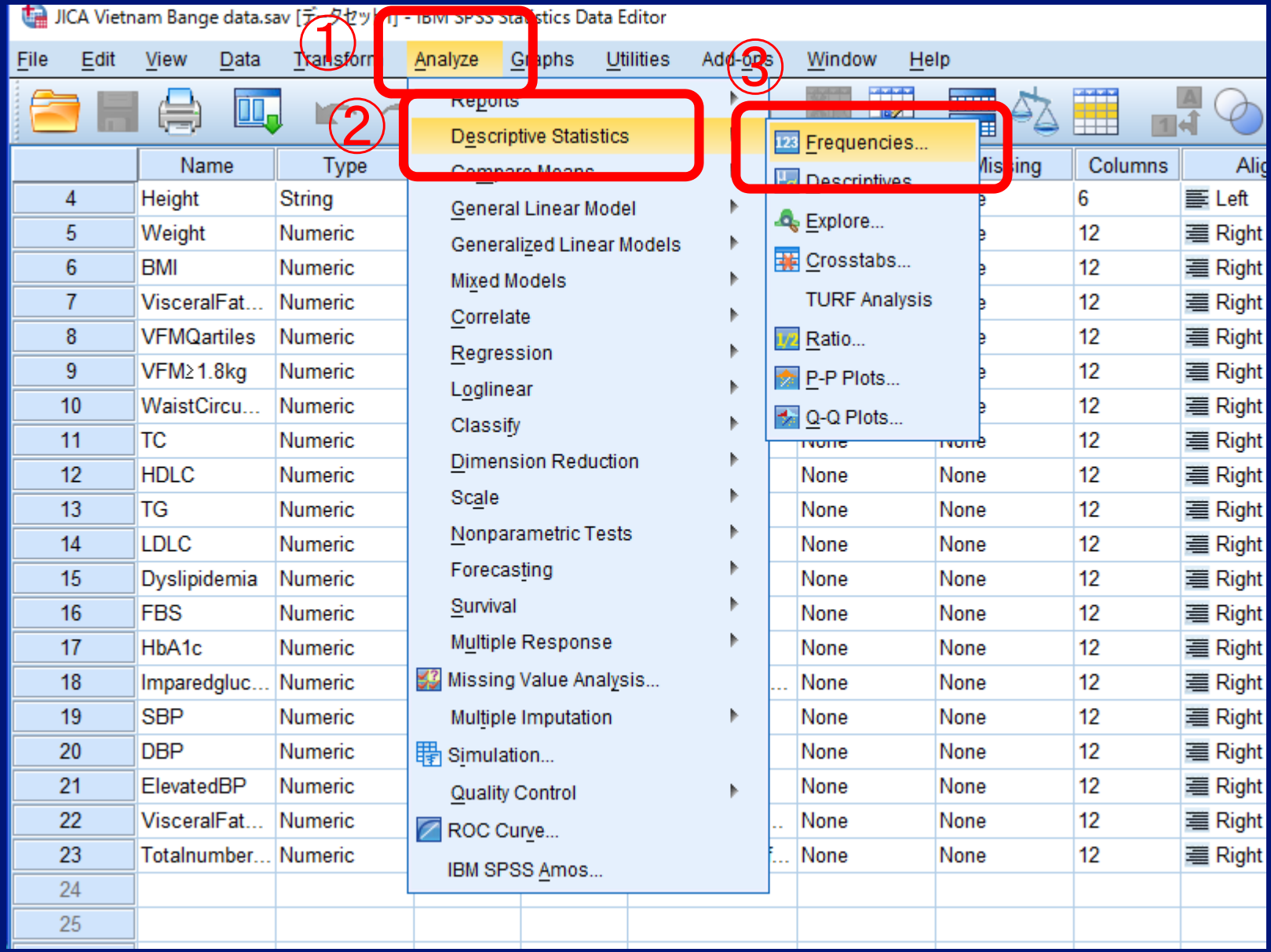
How to present distribution and characteristics?

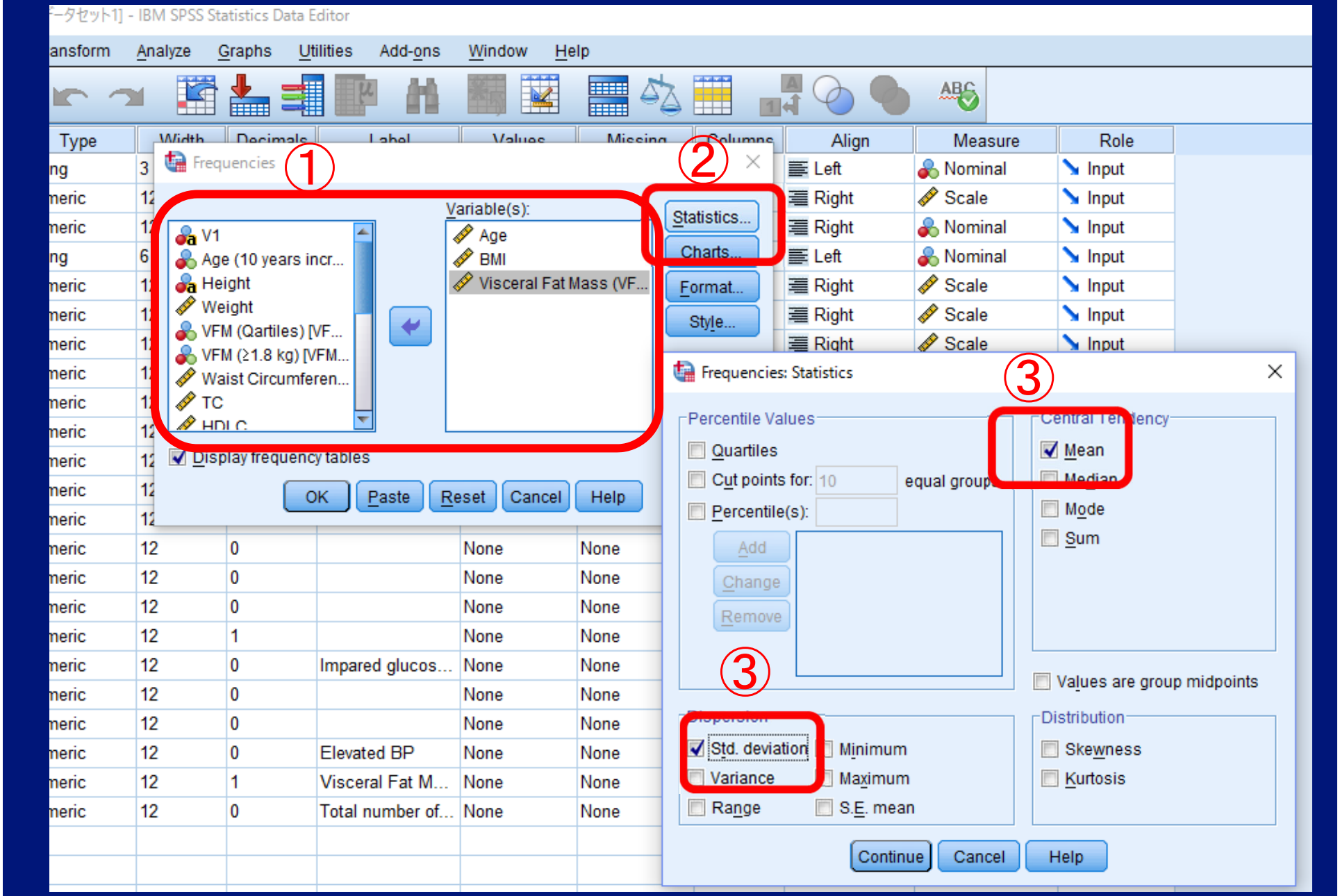
- **The first Table (Table 1) usually shows distribution and characteristics.**

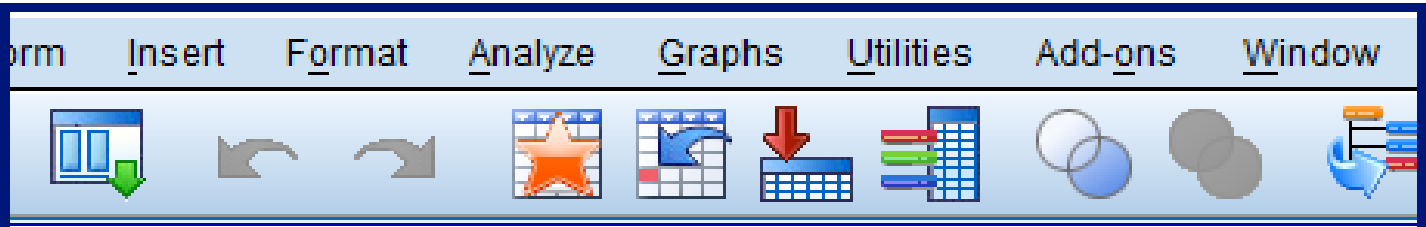
- **As representative values, mean (standard deviation) for parametric data, median (minimum, maximum) for non-parametric data, percentage for categorical data are usually used.**

Let's calculate mean (standard deviation) for the items below.

Age, BMI, Wait circumference (WC), Visceral fat mass (VFM), Lipid profiles (LDL-C, HDL-C, TG), Blood pressures, HbA1c.


JICA Vietnam Bange data.sav [データセット] - IBM SPSS Statistics Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Add-ons
Window
Help
①
②
③
Reports
Descriptive Statistics
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Simulation...
Quality Control
ROC Curve...
IBM SPSS Amos...
Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...
Q-Q Plots...
Name
Type
Missing
Columns
Align
4 Height String 6 Left
5 Weight Numeric 12 Right
6 BMI Numeric 12 Right
7 VisceralFat... Numeric 12 Right
8 VFMQartiles Numeric 12 Right
9 VFM≥1.8kg Numeric 12 Right
10 WaistCircu... Numeric 12 Right
11 TC Numeric 12 Right
12 HDLC Numeric None None 12 Right
13 TG Numeric None None 12 Right
14 LDLC Numeric None None 12 Right
15 Dyslipidemia Numeric None None 12 Right
16 FBS Numeric None None 12 Right
17 HbA1c Numeric None None 12 Right
18 Imparedgluc... Numeric None None 12 Right
19 SBP Numeric None None 12 Right
20 DBP Numeric None None 12 Right
21 ElevatedBP Numeric None None 12 Right
22 VisceralFat... Numeric None None 12 Right
23 Totalnumber... Numeric None None 12 Right
24
25


ータセット1] - IBM SPSS Statistics Data Editor
ansform Analyze Graphs Utilities Add-ons Window Help
Type Width Decimals Label Values Missing Columns Align Measure Role
Frequencies
①
②
Variable(s):
V1
Age (10 years incr...
Height
Weight
VFM (Qartiles) [VF...
VFM (≥1.8 kg) [VFM...
Waist Circumferen...
TC
HDL C
Age
BMI
Visceral Fat Mass (VF...
Statistics...
Charts...
Format...
Style...
Display frequency tables
OK Paste Reset Cancel Help
Frequencies: Statistics
③
Percentile Values
Quartiles
Cut points for: 10 equal group
Percentile(s):
Add
Change
Remove
Central Tendency
Mean
Median
Mode
Sum
Values are group midpoints
Dispersion
Std. deviation
Minimum
Variance
Maximum
Range
S.E. mean
Distribution
Skewness
Kurtosis
Continue Cancel Help




```
FREQUENCIES VARIABLES=Age BMI VisceralFatMassVFMkg
  /FORMAT=NOTABLE
  /STATISTICS=STDDEV MEAN
  /ORDER=ANALYSIS.
```

Frequencies

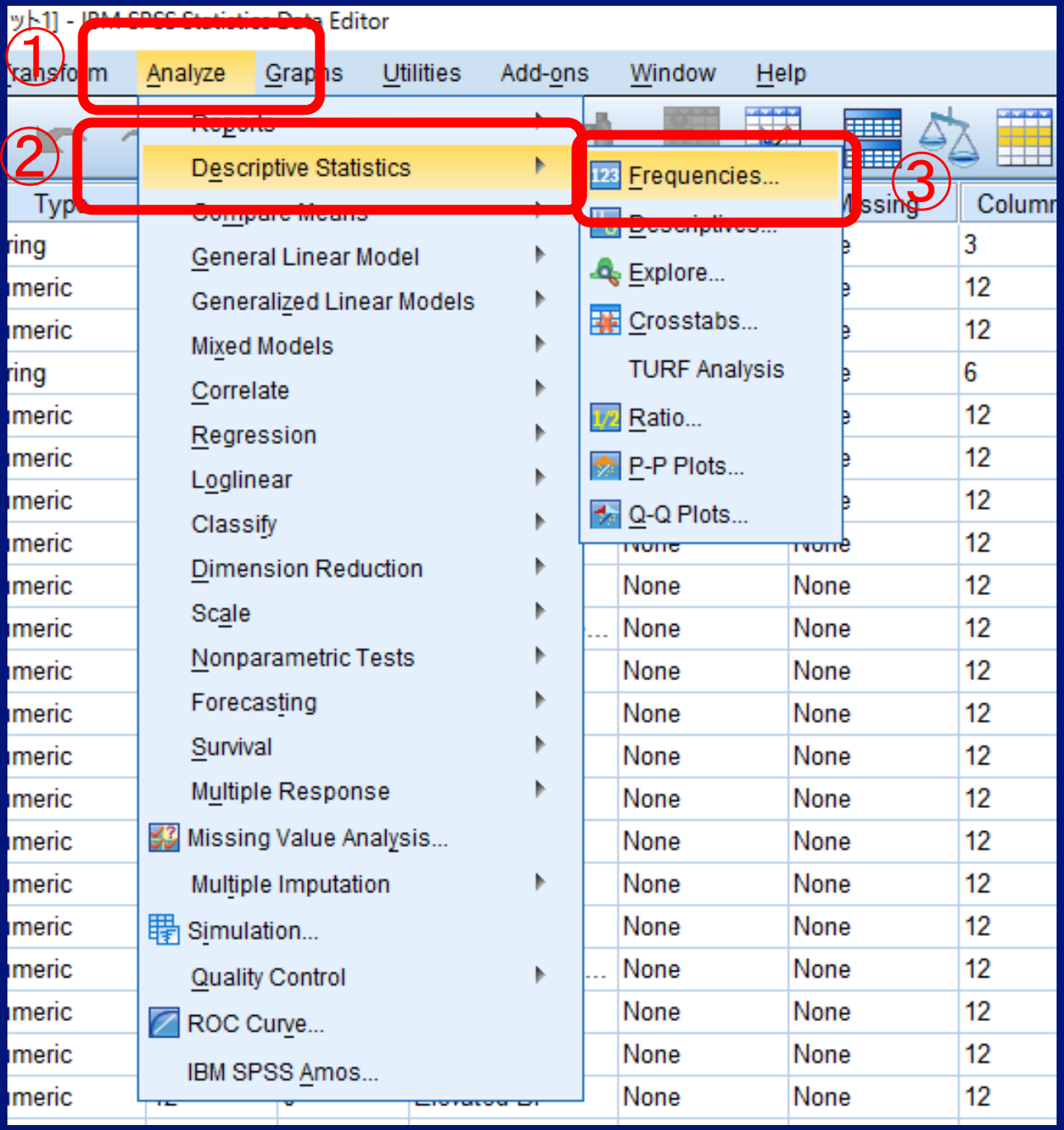
[データセット1] C:¥Users¥Hirohide Yokokawa¥Documents¥福島医大公

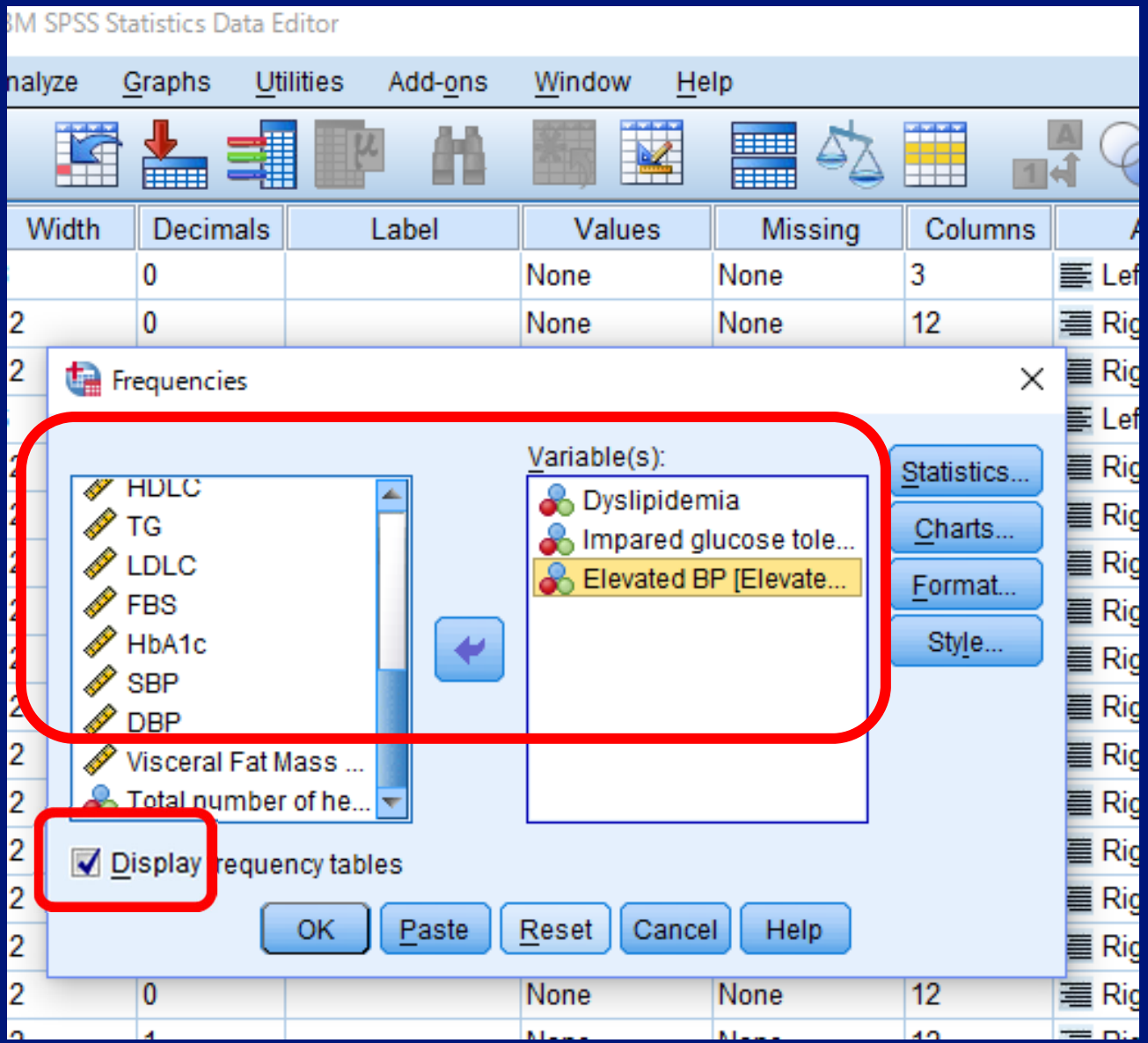
Statistics

		Age	BMI	Visceral Fat Mass (VFM) (kg)
N	Valid	442	442	442
	Missing	0	0	0
Mean		56.02	24.085	2.054
Std. Deviation		8.759	3.0612	.9550

Let's calculate percentage for the items below.

High blood pressure
Dyslipidemia
Impaired glucose tolerance


① Analyze
② Descriptive Statistics
③ Frequencies...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...
Q-Q Plots...
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Simulation...
Quality Control
ROC Curve...
IBM SPSS Amos...
Transform
Graphs
Utilities
Add-ons
Window
Help


M SPSS Statistics Data Editor
nalyze
Graphs
Utilities
Add-ons
Window
Help
Width
Decimals
Label
Values
Missing
Columns
0
None
None
3
2
0
None
None
12
Frequencies
HDLC
TG
LDLC
FBS
HbA1c
SBP
DBP
Visceral Fat Mass ...
Total number of he...
Variable(s):
Dyslipidemia
Impared glucose tole...
Elevated BP [Elevate...
Statistics...
Charts...
Format...
Style...
Display
requency tables
OK
Paste
Reset
Cancel
Help

Frequency Table

Dyslipidemia

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	201	45.5	45.5	45.5
	1	241	54.5	54.5	100.0
	Total	442	100.0	100.0	

Impared glucose tolerance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	402	91.0	91.0	91.0
	1	40	9.0	9.0	100.0
	Total	442	100.0	100.0	

Elevated BP

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	375	84.8	84.8	84.8
	1	67	15.2	15.2	100.0
	Total	442	100.0	100.0	

Let’s make a table for basic characteristics.

Table . Basic characteristics (N=422)

	Number (%) or Mean (Standard deviation)
Age (years)	56.0 (8.8)
Anthropometric measurements	
Body mass index (BMI) (kg/m^2)	
Waist circumference (WC) (cm)	
Visceral fat mass (kg)	
Blood pressure-related factors	
Systolic blood pressure (mmHg)	
Diastolic blood pressure (mmHg)	
High blood pressure	
Lipid-related items	
High-density lipoprotein cholesterol (HDL-C) (mg/dL)	
Low-density lipoprotein cholesterol (LDL-C) (mg/dL)	
Triglycerides (TG) (mg/dL)	
Dyslipidemia	
Glucose-related items	
Hemoglobin A1c (HbA1c) (%)	
Impaired glucose tolerance	

Let’s discuss following questions.

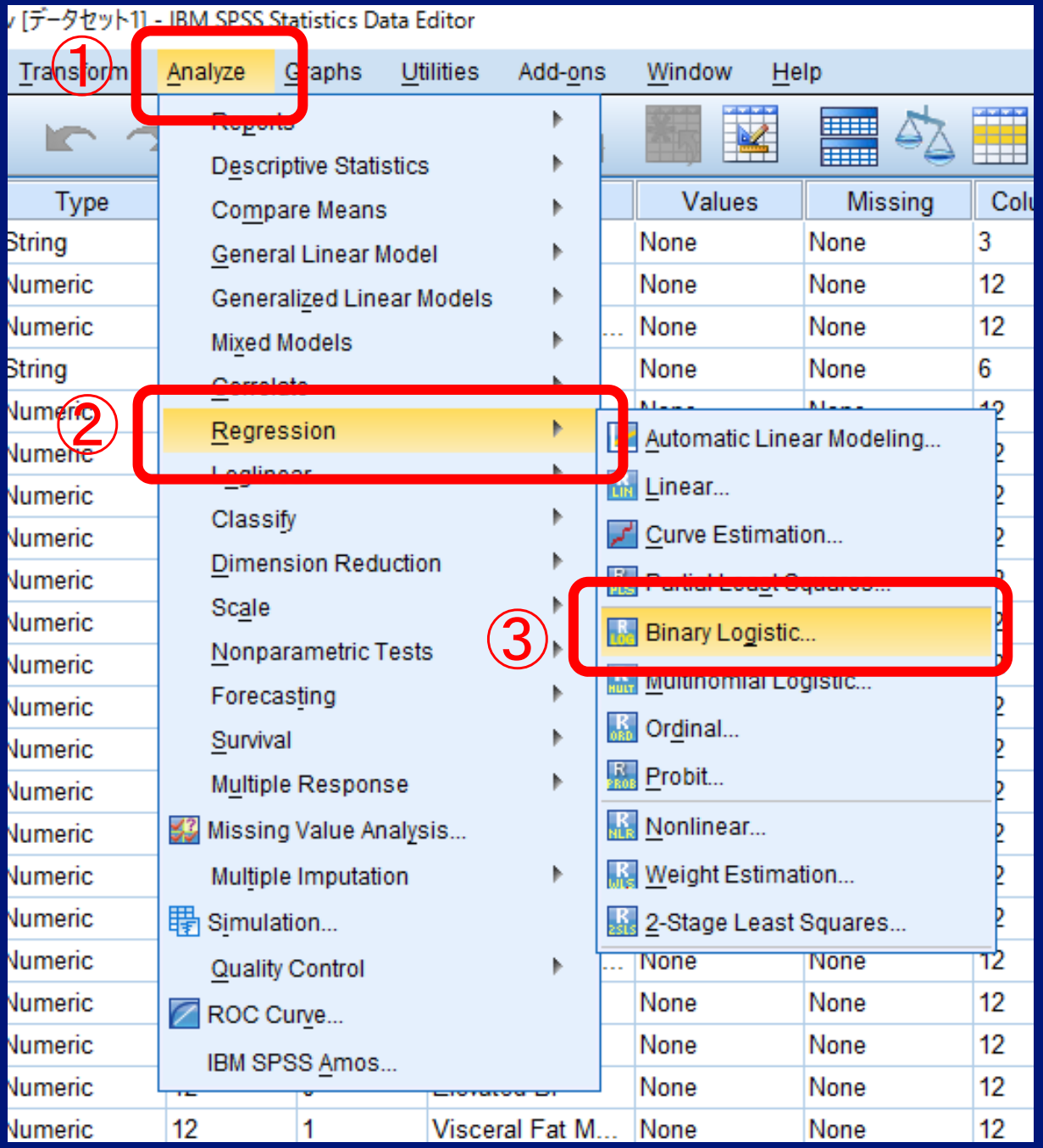
1. Which lifestyle related disorders is the most frequent?
2. What kind of intervention will be effective?

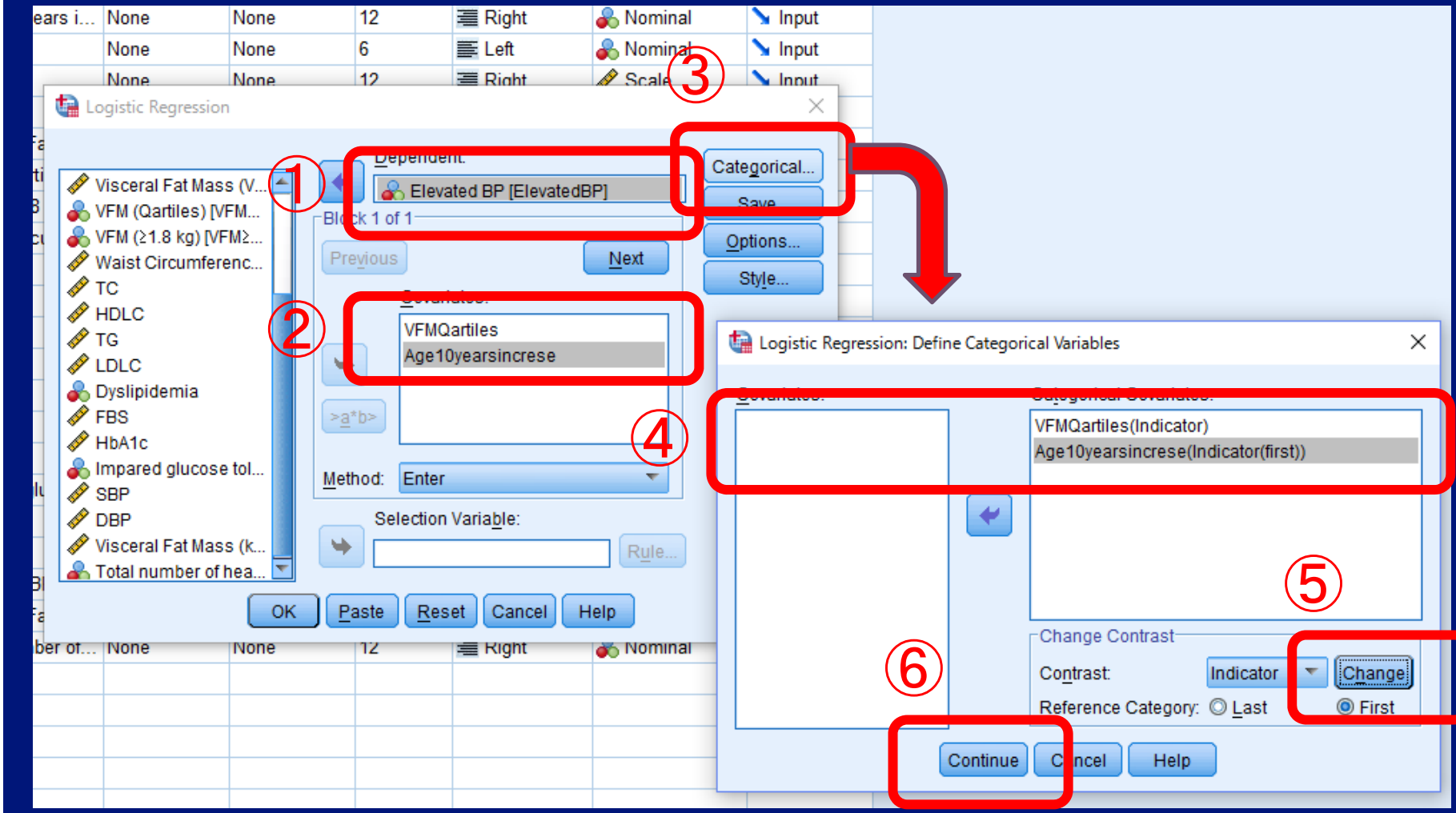
How to present analytic study data?

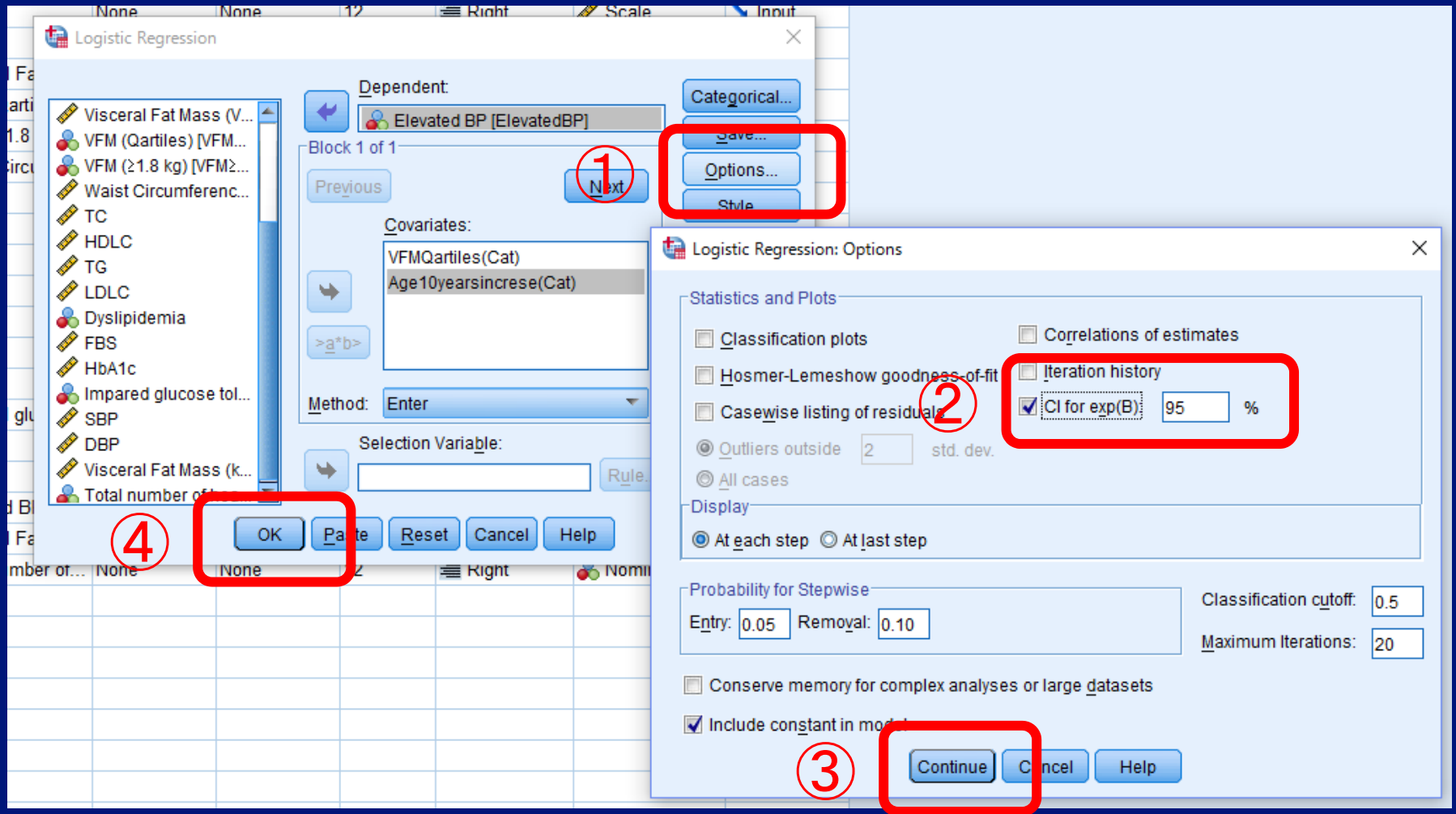
www.brandwithin.com

Let’s conduct Logistic regression analysis

- Bivariate regression analysis; adjusting for age (10-year increase)


[データセット1] - IBM SPSS Statistics Data Editor
Transform
Analyze
Graphs
Utilities
Add-ons
Window
Help
①
②
③
Descriptive Statistics
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Regression
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Simulation...
Quality Control
ROC Curve...
IBM SPSS Amos...
Automatic Linear Modeling...
Linear...
Curve Estimation...
Binary Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Type
Values
Missing
String
Numeric
None
12
6
Visceral Fat M...


Logistic Regression
Visceral Fat Mass (V...
VFM (Qartiles) [VFM...
VFM (≥1.8 kg) [VFM≥...
Waist Circumferenc...
TC
HDLC
TG
LDLC
Dyslipidemia
FBS
HbA1c
Impared glucose tol...
SBP
DBP
Visceral Fat Mass (k...
Total number of hea...
Dependent:
Elevated BP [ElevatedBP]
Block 1 of 1
Previous
Next
Covariates:
VFMQartiles
Age10yearsincrese
>a*b>
Method: Enter
Selection Variable:
Rule...
Categorical...
Save...
Options...
Style...
OK
Paste
Reset
Cancel
Help
Logistic Regression: Define Categorical Variables
Covariates:
Categorical Covariates:
VFMQartiles(Indicator)
Age10yearsincrese(Indicator(first))
Change Contrast
Contrast: Indicator
Change
Reference Category: Last First
Continue
Cancel
Help
1
2
3
4
5
6


Logistic Regression
Visceral Fat Mass (V...
VFM (Qartiles) [VFM...
VFM (≥1.8 kg) [VFM≥...
Waist Circumferenc...
TC
HDLC
TG
LDLC
Dyslipidemia
FBS
HbA1c
Impared glucose tol...
SBP
DBP
Visceral Fat Mass (k...
Dependent:
Elevated BP [ElevatedBP]
Block 1 of 1
Previous
Next
Covariates:
VFMQartiles(Cat)
Age10yearsincrese(Cat)
>a*b>
Method:
Enter
Selection Variable:
Rule...
Categorical...
Save...
Options...
Style...
Bootstrap...
OK
Paste
Reset
Cancel
Help
1
2
3
4
Logistic Regression: Options
Statistics and Plots
Classification plots
Hosmer-Lemeshow goodness-of-fit
Casewise listing of residuals
Outliers outside
2
std. dev.
All cases
Correlations of estimates
Iteration history
CI for exp(B)
95
%
Display
At each step
At last step
Probability for Stepwise
Entry:
0.05
Removal:
0.10
Classification cutoff:
0.5
Maximum Iterations:
20
Conserve memory for complex analyses or large datasets
Include constant in model
Continue
Cancel
Help

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Age10yearsincrese			3.992	3	.262			
	Age10yearsincrese(1)	.160	.576	.078	1	.781	1.174	.380	3.631
	Age10yearsincrese(2)	.310	.522	.352	1	.553	1.363	.490	3.789
	Age10yearsincrese(3)	.648	.529	1.501	1	.221	1.911	.678	5.386
	VFMQartiles			11.363	3	.010			
	VFMQartiles(1)	.066	.311	.045	1	.833	1.068	.581	1.964
	VFMQartiles(2)	.612	.301	4.139	1	.042	1.844	1.023	3.326
	VFMQartiles(3)	.804	.307	6.862	1	.009	2.233	1.224	4.075
	Constant	-1.260	.543	5.398	1	.020	.284		

a. Variable(s) entered on step 1: Age10yearsincrese, VFMQartiles.

Odds ratio

Table . Odds Ratios for Lifestyle-Related Disorders According to Visceral Fat Mass (Logistic Regression Analysis)

	Bivariate[a]			Multivariate					
				Model 1[d]			Model 2[e]		
Visceral Fat Mass (kg)	OR[b]	95% CI[c]	*P*	OR[b]	95% CI[c]	*P*	OR[b]	95% CI[c]	*P*
Elevated blood pressure									
Q1≤1.2	Reference			Reference			Reference		
1.2<Q2≤1.8	1.07	0.58-1.96	0.83						
1.8<Q3≤2.5	1.84	1.02-3.33	0.04						
2.5<Q4	2.23	1.22-4.08	<0.01						
Dyslipidemia									
Q1≤1.2	Reference			Reference			Reference		
1.2<Q2≤1.8									
1.8<Q3≤2.5									
2.5<Q4									
Glucose tolerance impairment									
Q1≤1.2	Reference			Reference			Reference		
1.2<Q2≤1.8									
1.8<Q3≤2.5									
2.5<Q4									

Let’s conduct multiple Logistic regression analysis.

1. Which are the adjusting factors?

Let’s interpret the results.

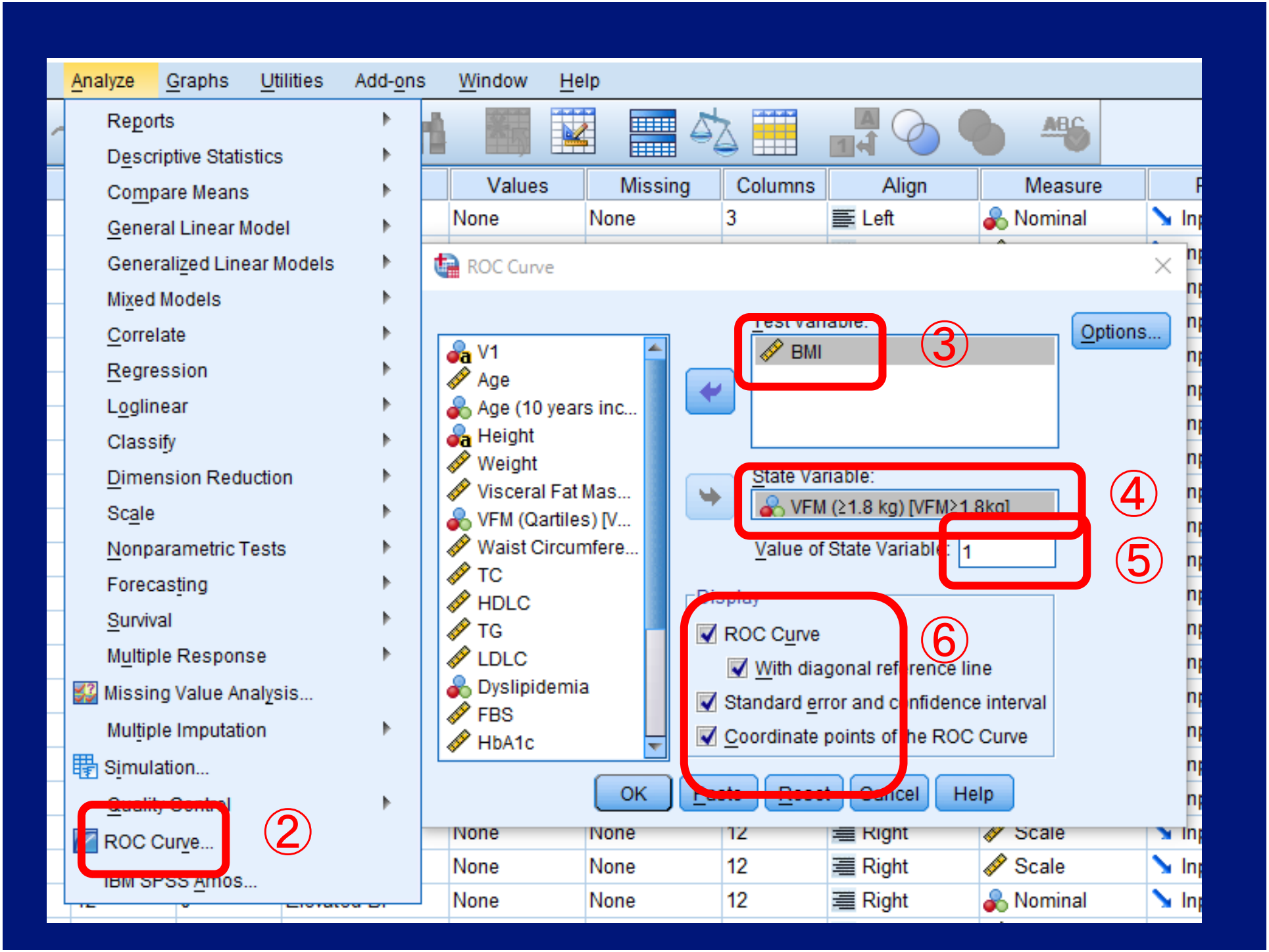
Which is the most appropriate cutoff of VFM for lifestyle related disorders?
(1.2, 1.8, or 2.5 kg)

Implication of bivariate multivariate analysis

	Bivariate[a]			Multivariate					
				Model 1[d]			Model 2[e]		
Visceral Fat Mass (kg)	OR[b]	95% CI[c]	*P*	OR[b])	95% CI[c]	*P*	OR[b]	95% CI[c]	*P*
Elevated blood pressure									
Q1≤1.2	Reference			Reference			Reference		
1.2<Q2≤1.8	1.07	0.58-1.96	0.83						
1.8<Q3≤2.5	1.84	1.02-3.33	0.04						
2.5<Q4	2.23	1.22-4.08	<0.01						
Dyslipidemia									
1.2<Q2≤1.8									
1.8<Q3≤2.5									
2.5<Q4									

Q4 was 2.23 times likely to have elevated blood pressure compared to reference (Q1) after adjusting by10 years increase.

Let’s make a Receiver operating characteristics curve (ROC curve) and estimate the appropriate cut off values of BMI and WC for visceral fat mass (≥1.8 kg)


Analyze
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Simulation...
ROC Curve...
IBM SPSS Amos...
②
Values
Missing
Columns
Align
Measure
None
None
3
Left
Nominal
ROC Curve
V1
Age
Age (10 years inc...
Height
Weight
Visceral Fat Mas...
VFM (Qartiles) [V...
Waist Circumfere...
TC
HDLC
TG
LDLC
Dyslipidemia
FBS
HbA1c
BMI
③
Options...
State Variable:
VFM (≥1.8 kg) [VFM≥1.8kg]
④
Value of State Variable:
1
⑤
ROC Curve
⑥
With diagonal reference line
Standard error and confidence interval
Coordinate points of the ROC Curve
OK
Paste
Reset
Cancel
Help
None
None
12
Right
Scale
None
None
12
Right
Scale
None
None
12
Right
Nominal

BMI

22.250	.996	.348
22.350	.996	.328
22.450	.992	.304
22.550	.992	.279
22.650	.983	.250
22.750	.979	.230
22.850	.950	.221
22.950	.933	.206
23.050	.924	.181
23.150	.920	.162
23.250	.916	.142
23.450	.870	.123
23.550	.866	.118

Waist circumference

83.200	.941	.235
83.450	.937	.230
83.600	.924	.225
83.800	.920	.225
83.950	.916	.225
84.100	.903	.181
84.300	.899	.181
84.450	.895	.181
84.550	.891	.181
84.700	.882	.181
84.900	.878	.176
85.300	.849	.167
85.550	.836	.152
85.800	.836	.147

Environ Health Prev Med
DOI 10.1007/s12199-014-0411-9

REGULAR ARTICLE

Association between lifestyle-related disorders and visceral fat mass in Japanese males: a hospital based cross-sectional study

Hironobu Sanada · Hirohide Yokokawa · Junichi Yatabe · Scott M. Williams · Robin A. Felder · Pedro A. Jose · Seiichi Takenosita

Received: 2 July 2014 / Accepted: 3 September 2014

Environ Health Prev Med. 2014; 19:429-435.