Biostatistics

NCD Epidemiology Research Center (NERC)
Department of Medical statistics

Akiko HARADA

0. Course series orientation

How to do this course series

- Aim of the course
  - The main aims of this course are to understand and use statistical methods in medical studies.
  - We will focus on applied epidemiological study examples rather than clinical study.
- Course series
  - We will review basic statistical methods using statistical packages.

Lecture schedule

The sessions marked with * will be conducted onsite; otherwise, students will register for the web-based class, watch the video, do the exercises on their own, and submit their assignments.

1 *Introduction 1-A JMP(May 8) 、1-B Python (May 15)
<on site: MMC browsing room>

2 Group Comparison <On-Demand >

3 General linear model (1) linear regression <On-Demand >

4 Generalized linear model Logistic, Poisson <On-Demand >

5 Survival analysis <On-Demand >

6 Cluster analysis, Latent class analysis and ML <On-Demand >

7 General linear model (2) ANOVA MIXED <On-Demand >

8 *Wrap up (July 31)
<on site: NCD Epi center meeting room>

How to proceed with the course series

◆Lecture type: On-demand

1st(1-A,1-B) , 8th Face-to-face lecture

2nd~7th Lecture (conducted on demand)

◆Login address: Posted on Web Class

◆Earn credits:

All students are required to submit a report for credit ・・ Submitted via Webclass

Credits will be earned in the Basic Course, which will be completed by the end of June

◆Lecture materials: via Webclass

◆Preparation for the course: Installation of some statistical package (software).

Please send any questions about the lectures and exercises at any time. Email or webclass

E-mail: aharada@belle.shiga-med.ac.jp

Webclass:(course URL)

https://latte.shiga-med.ac.jp/webclass/login.php?group_id=38da601dea59e8e42b5724e6fd83f91a&auth_mode=SHIB

This lecture will be given using JMP, Python

The first two lectures will provide exercises on basic operations.

- JMP
  - JMP is an academic license, which means that students and staff can use it freely and free of charge. It can also be installed on your own PC.
  - GUI-based software that is intuitive and easy to use for beginners
- Python
  - Python is freeware. It has many users in many fields such as program development, AI, statistical analysis, etc., and there are plenty of teaching materials available.
  - It is also possible to create code and use support such as chatbots that utilize generated AI.
  - Google's Collaboratory is a convenient environment for beginners, with low hurdles such as PC environment settings.

Webclass

- Access to lecture materials and submission of reports will be done using Webclass.
  - 「NCD Biostatistics」
    - https://latte.shiga-med.ac.jp/webclass/login.php?group_id=38da601dea59e8e42b5724e6fd83f91a&auth_mode=SHIB

- Please submit your reports to the folder for each lecture session.

Report

- **A. Reports that must be submitted (Minimum)**
  - Set within the range that does not necessarily require the use of statistical packages.

- **B. Assignments to be practiced using the package** (this is recommended)

Please choose either A or B and submit your report via Webclass.

Webclass: [Report Submission]

Webclass

Report Submission •••

New Report A レポート	更新 7分前 実行者数 0	•••
New Report:B #01 Q1-Q2 *To be implemented in either JMP or Python. レポート	更新 1分前 実行者数 0	•••
New Report:B #2-Q1,Q2 レポート	更新 1分前 実行者数 0	•••
New Report B: #3-Q1-Q2 レポート	更新 1分前 実行者数 0	•••
New Report:B #4 Q1,Q2 レポート	更新 0分前 実行者数 0	•••
New Report B: #5-Q1-Q2 Report レポート	更新 0分前 実行者数 0	•••

Report A

NCD course："Biostatistics" Assignments (Report)

Watch the on-demand lecture and answer the following 1 to 7 questions.*
**You may skip the exercises parts using the statistical analysis package if not necessary.*

1. t-test, Wilcoxon test

Describe the characteristics of the two tests (t-test and Wilcoxon test) with respect to outliers.
You may use the exercises from the lecture and explain them by showing the results of your analysis.

2. Describe the analysis method that should be used in the following case study and the reasons for its use.

We asked junior high school students, "With whom do you feel more comfortable discussing your problems, your father or your mother? The same number of people（n=100）answered each question, and the results are shown in A and B.
For each of the results of A and B, consider whether there is a difference in whether they would choose their fathers or mothers as their consulting partners.
Indicate the method of analysis that should be selected and the results of the test.

<A>

		Father		
		Talk about (+)	Talk about (−)	Total
Mother	Talk about (+)	50	0	50
	Talk about (−)	0	50	50
Total		50	50	100

<B>

		Father		
		Talk about (+)	Talk about (−)	Total
Mother	Talk about (+)	32	48	80
	Talk about (−)	8	12	20
Total		40	60	100

<A>

- Consultation rate was 50% for both fathers and mothers.
- In the fathers' consultation group, the consultation rate for mothers is 100%.
- The consultation rate for mothers in the non-father consultation group was 0%.

Watch the on-demand lecture and answer the following 1 to 7 questions.*

- **You may skip the exercises parts using the statistical analysis package if not necessary.*

Please access the following

Webclass

[#0 Preparation for attending lectures]

- [Report(for credit)]

Report B
(this is recommended)

- Report to be practiced using the stat.packages
- Please submit each Report for each session.

<u>The report section is set up as follows</u>

#02

Q1 t-test

Q2 chi-square test, McNemar test

#03

Q1 General regression

Q2 Variable selection

#04

Q1 Logistic regression

Q2 Poisson regression

#05

Q1 Q2 Survival analysis

#06, #07

No report

This is the end of the lecture orientation.

Questions about the lectures and exercises are always welcome.

Biostatistics

Course #01- part A

NCD Epidemiology Research Center (NERC)
Department of Medical statistics

Akiko HARADA

Course Content

This course will be divided in 7 chapters as follow

1 (1-A,1-B) Introduction

2 Group Comparison

3 General linear model (1) linear regression

4 Generalized linear model Logistic, Poisson

5 Survival analysis

6 Principal component analysis, Factor analysis, Cluster analysis, Latent class analysis

7 General linear model (2) ANOVA MIXED

Ondemand

8 Wrap up

1 Introduction

Using the statistical package

Goals of Todays Lectures:
Become familiar with the use of statistical packages.
1) Be able to load data to be used for analysis
2) Understand the characteristics of data by using various methods

After the lecture, please work on the assignment and submit the report.

Today's overview
Use JMP package

1. Usage of Statistical Software (package)
   1) Brief Introduction to JMP
   2) Reading Data

In the next course (#02),
we will try the same content in Python.

2. Description of Data
   1) Location: mean, median, percentile
   2) Dispersion: range, inter-quartile range, CV
   3) Distribution: skewness, kurtosis, Normality
   4) Association between two variables: Correlation

Graphics

Software packages for statistical analysis of data

- SAS
- R
- Python
- STATA

Programming statement (code) type

- SPSS
- JMP

Point-and-click interface

It is desirable to be able to use one (or more if possible) of the packages !

This lecture will be given using JMP and Python.

- JMP

- Python
  - Google Colaboratory

JMP

How to install JMP Pro

JMP or JMP PRO

- Analysis package JMP is now available for use at our universities (Univ. license)
  - This package can be downloaded from MMC and installed on individual PCs for use on both Win and Mac computers (requires annual renewal).
  - JMP is another excellent software package from SAS.
    It has a lot of features that are easier to use than SAS,
    - such as machine learning tools, experimental design (DOE), Structured Equation Modeling (SEM), etc.
    - The user interface has been wonderfully improved with each new version.

JMP PRO

- JMP Pro covers basic and advanced analysis.
- JMP Pro is more intuitive and better at visualizing data than coding-type software.
- JMP Pro is useful for creating data sets and cleaning data for use with SAS, SPSS, and other statistical packages.

JMP can be downloaded to your PC by SUMS students, faculty, and staff.
Download instructions are given below.

Go to the SUMS Multimedia Center website.
URL:https://www.shiga-med.ac.jp/mmc/

Download JMP Pro" from the various services

Application for use of JMP Pro

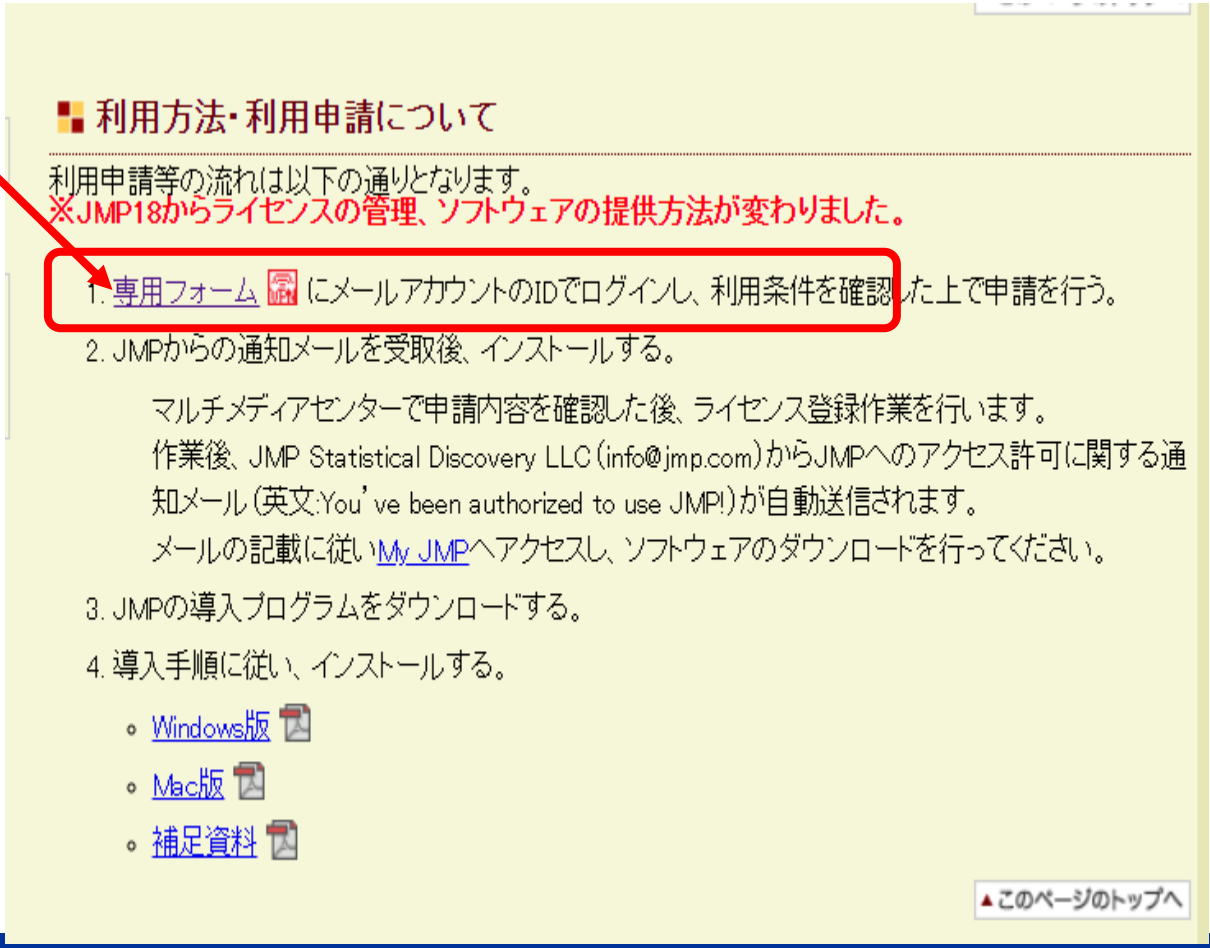
1.Log in to the dedicated form with your e-mail account ID, check the terms and conditions of use, and submit the application.

2.Install the software after receiving a notification e-mail from JMP.JMP Statistical Discovery LLC (info@jmp.com) will automatically send you an email notification that you've been authorized to use JMP!

Follow the instructions in the email to access

3.My JMP and download the software. Download the JMP installation program.

4.Follow the installation instructions.



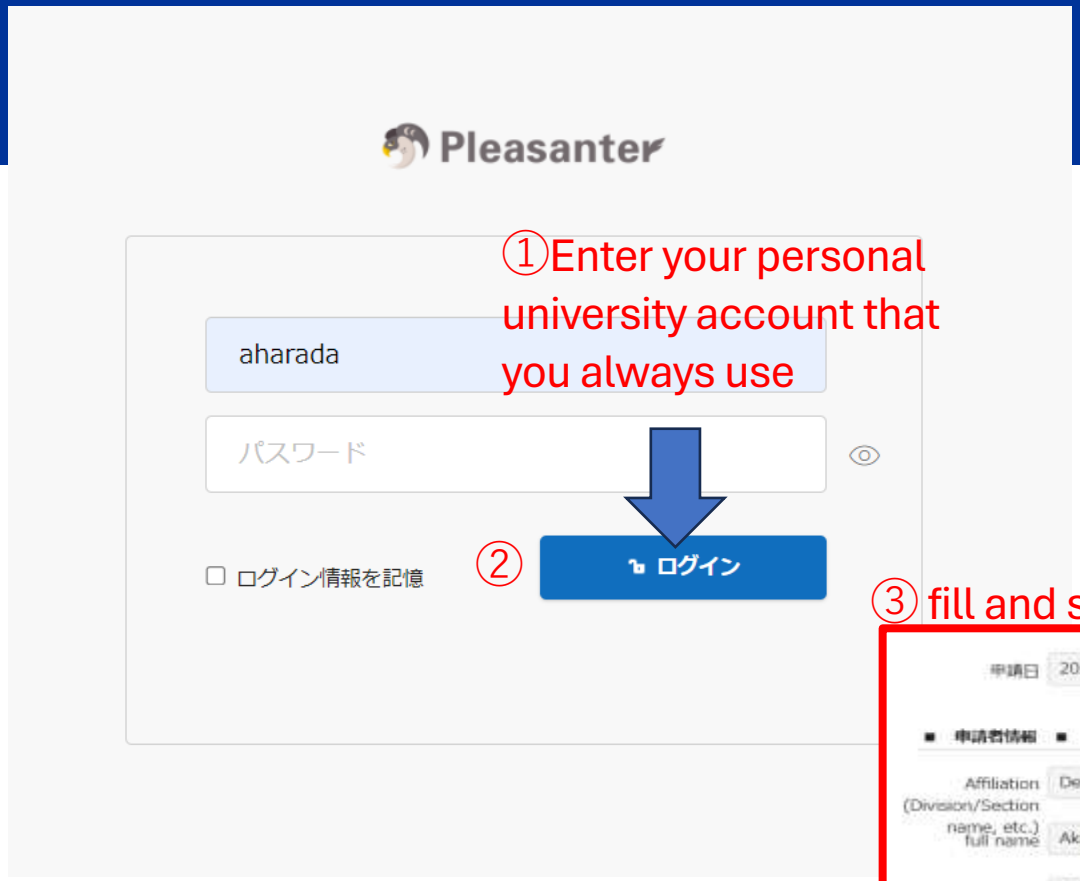
利用方法・利用申請について

利用申請等の流れは以下の通りとなります。
※JMP18からライセンスの管理、ソフトウェアの提供方法が変わりました。

1. 専用フォーム にメールアカウントのIDでログインし、利用条件を確認した上で申請を行う。
2. JMPからの通知メールを受取後、インストールする。

   マルチメディアセンターで申請内容を確認した後、ライセンス登録作業を行います。
   作業後、JMP Statistical Discovery LLC(info@jmp.com)からJMPへのアクセス許可に関する通知メール(英文:You've been authorized to use JMP!)が自動送信されます。
   メールの記載に従いMy JMPへアクセスし、ソフトウェアのダウンロードを行ってください。
3. JMPの導入プログラムをダウンロードする。
4. 導入手順に従い、インストールする。
   - Windows版
   - Mac版
   - 補足資料

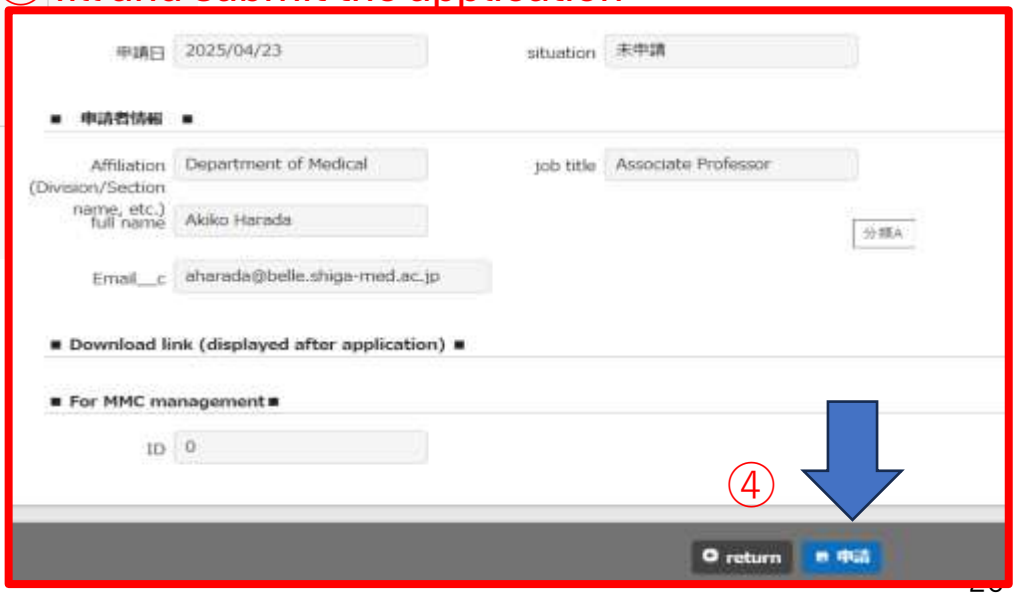
▲このページのトップへ



①Enter your personal university account that you always use

②

③ fill and submit the application



④

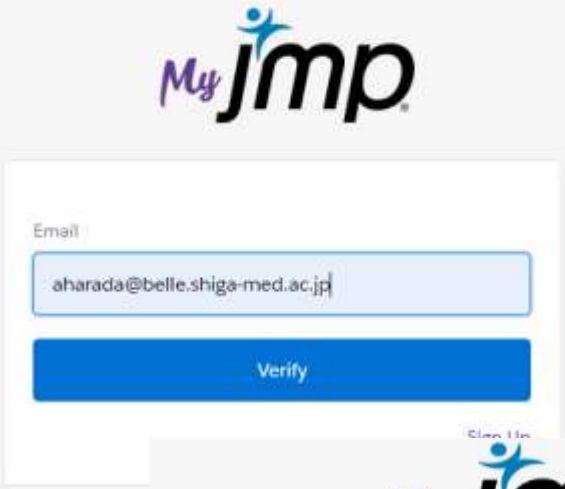
Next step

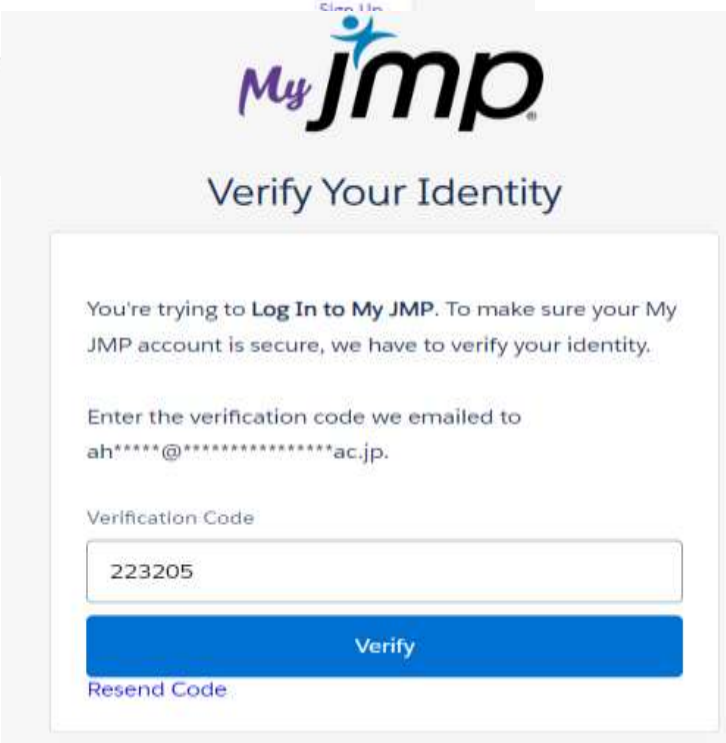
It may take some time to work and you may not get an immediate response

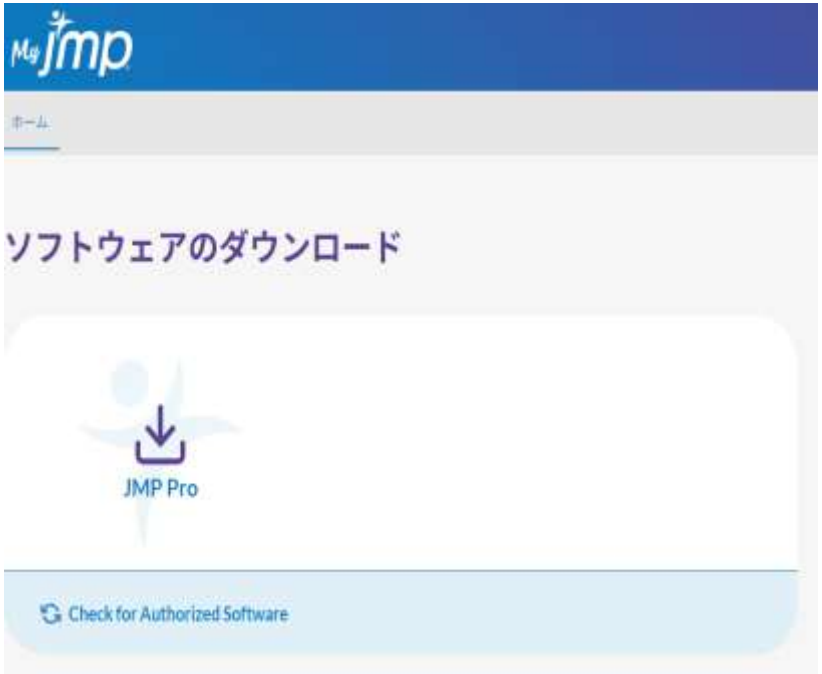
- After completing the steps in the JMP18 Multimedia Center, you will automatically receive a notification email (in English) from JMP Statistical Discovery LLC (info@jmp.com) regarding your access permission to JMP.

- Follow the instructions in the email to access "My JMP*" and download the software.

Next step


My JMP
Email
aharada@belle.shiga-med.ac.jp
Verify


My JMP
Verify Your Identity
You're trying to Log In to My JMP. To make sure your My JMP account is secure, we have to verify your identity.
Enter the verification code we emailed to ah*****@**********ac.jp.
Verification Code
223205
Verify
Resend Code


My JMP
ホーム
ソフトウェアのダウンロード
JMP Pro
Check for Authorized Software

Flow of use JMPPro18

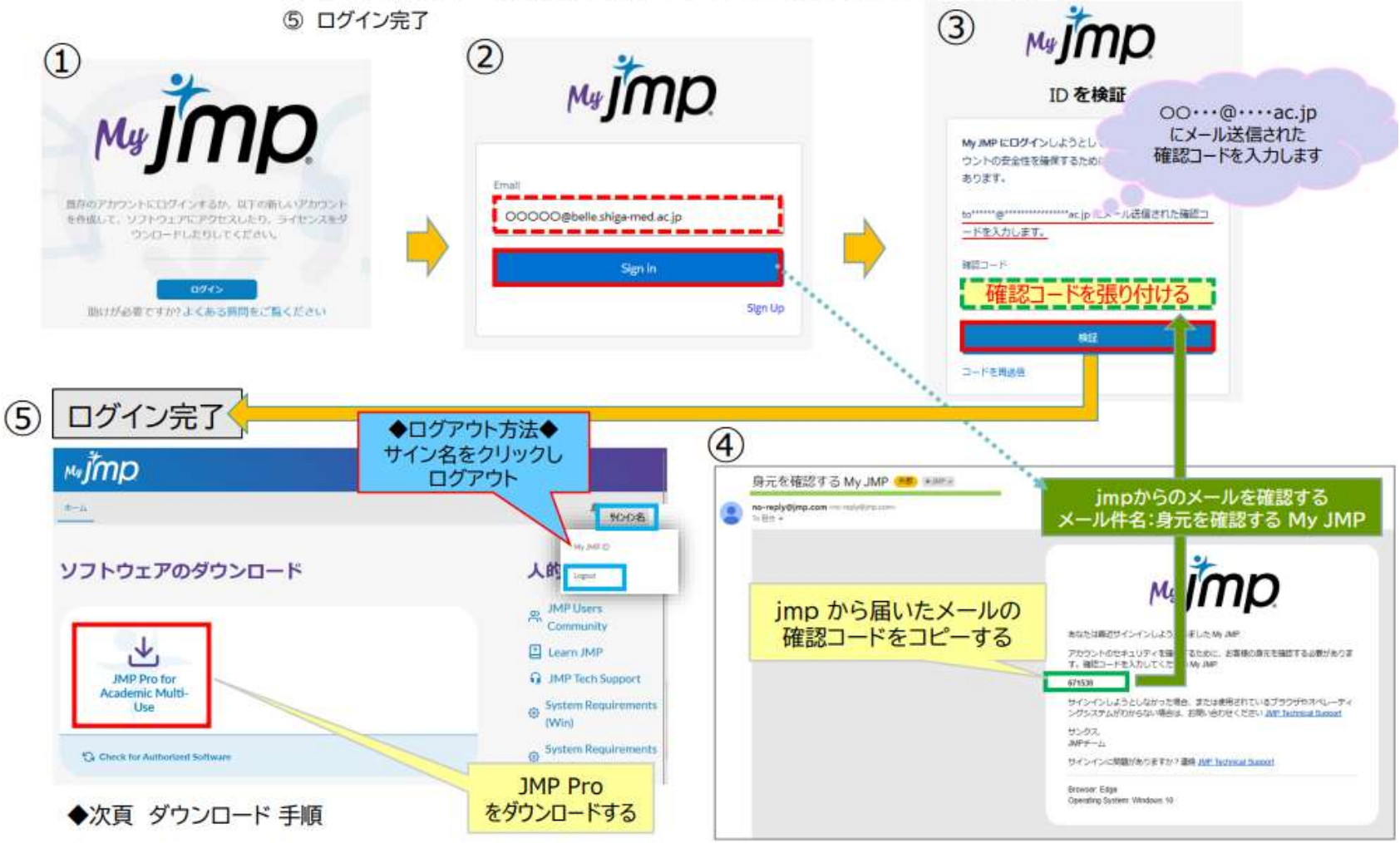
- 1.Log in to the dedicated form with your e-mail account ID, check the terms and conditions of use, and submit the application.

- 2.Install the software after receiving a notification e-mail from JMP.JMP Statistical Discovery LLC (info@jmp.com) will automatically send you an email notification that you've been authorized to use JMP!
  - Follow the instructions in the email to access

(Win)

1. jmpログイン

① my.jmp.com:へログイン する
② Emailアドレスを入力し、サインインする
③ 「IDを検証」画面が立ち上がる → jmp からのメールが届く
④ jmpから届いたメールを確認し、確認コード を「IDを検証」画面にコピー＆ペーストする
⑤ ログイン完了



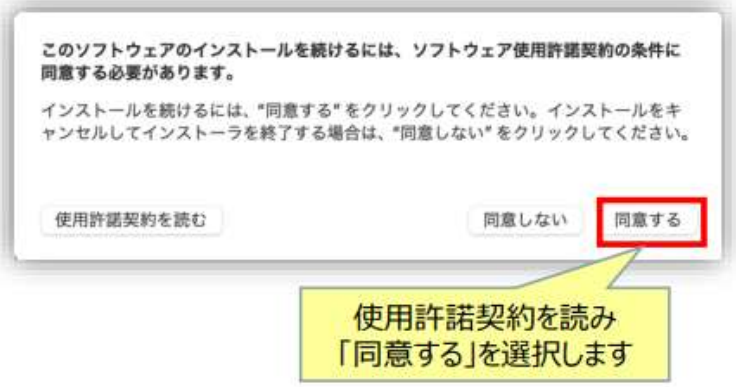
◆次頁　ダウンロード 手順

- 3.My JMP and download the software. Download the JMP installation program.

(Mac)

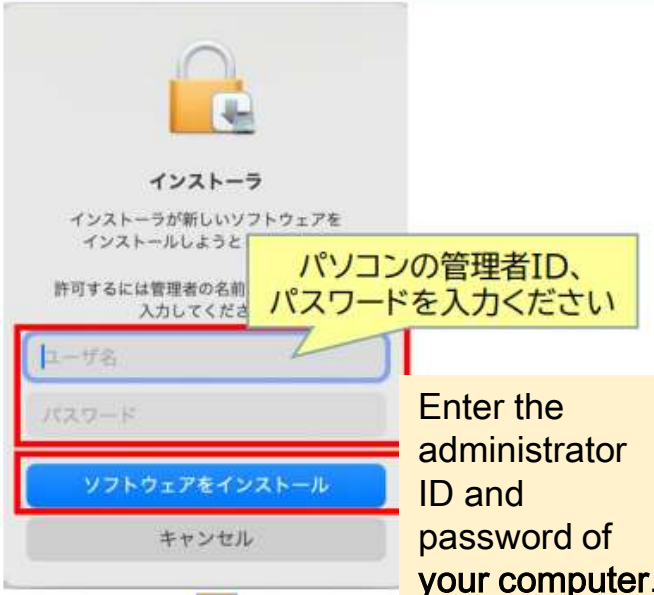
2. インストーラの選択画面にてMac Installerを選択し、ダウンロードする。
3. ダウンロードしたdmgファイル(jmppro_18**_mac.dmg)を解凍します。
4. 解凍したファイルを実行し、導入を進めてください。

1. Extract the downloaded file



使用許諾契約を読み
「同意する」を選択します

5. 導入(続き)

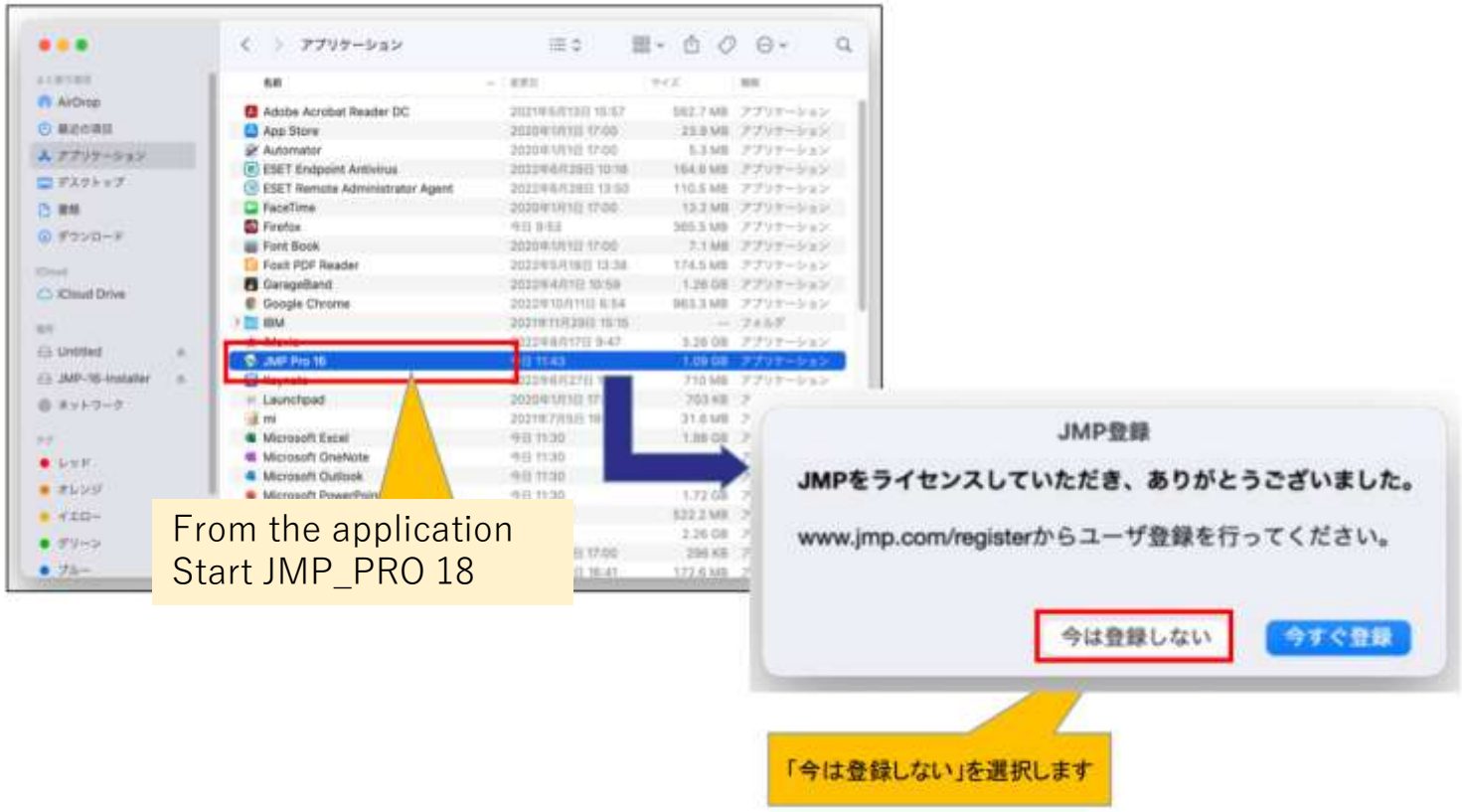


Enter the administrator ID and password of <u>your computer</u>.

Installation Complete

導入完了

3. 導入後のJMP登録は必要ありません



From the application
Start JMP_PRO 18

Do not register now

- “JMPPro” is already installed on MMC's PC

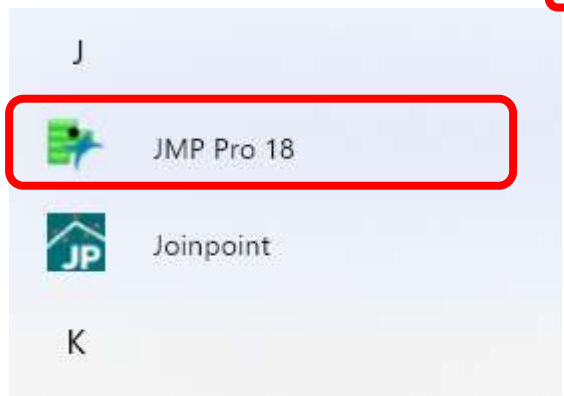
- After this, as long as the installation have been complete, you may use your own PC, or you may use the JMP in MMCs PC.

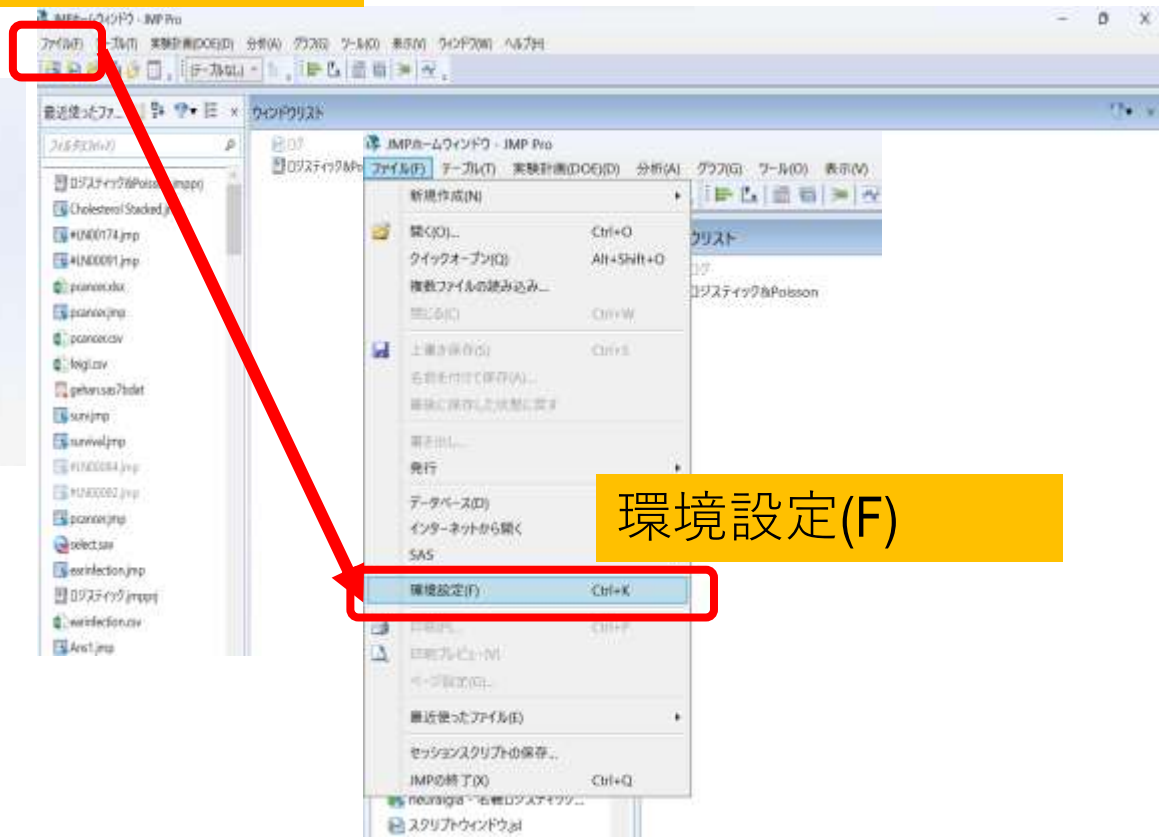
Software Language Support

- Windows needs to be changed
- Mac supports PC languages. No need to change language settings

Japanese screen ： Windows のみ

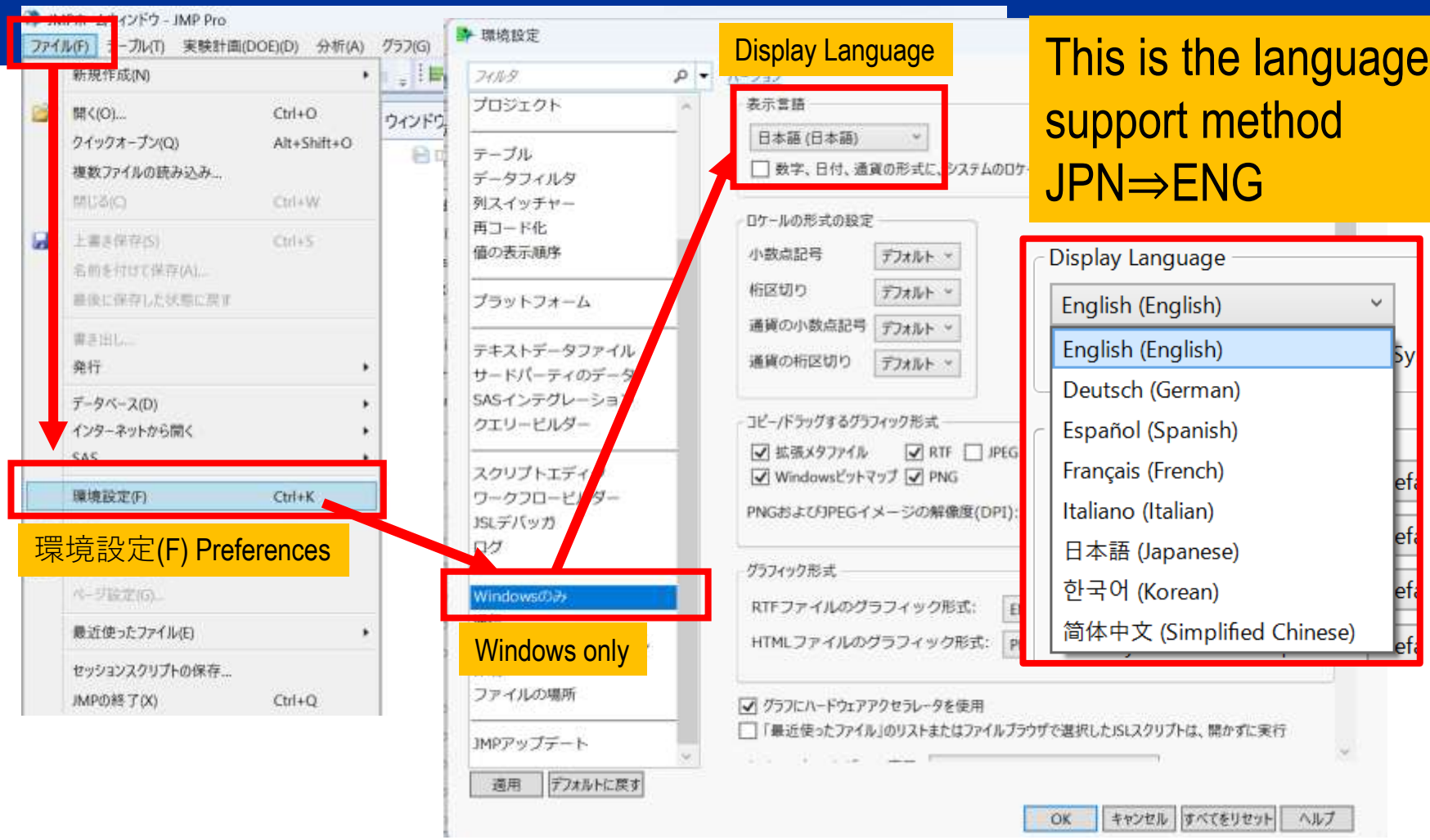
ファイル　File





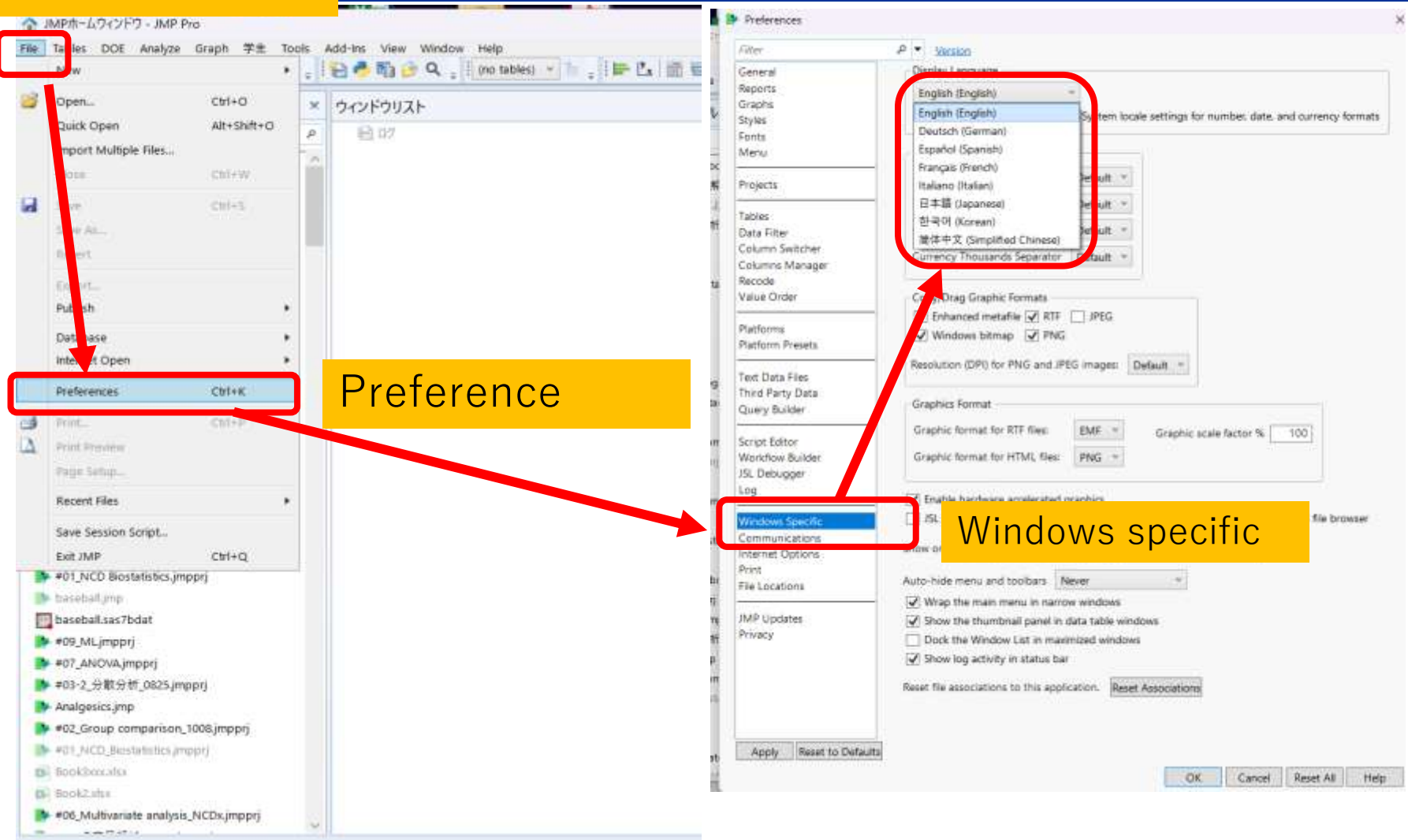
環境設定(F)

Japanese screen



English screen: Windows specific

ファイル　File



Mac OS

- As for Macs, they are basically displayed in the language of the operating system.

- However, JMP supports
  - English
  - German
  - French,
  - Spanish
  - Italian
  - Korean
  - Chinese (Simplified Chinese)

SAS has prepared the JMP Learning Library for learning purposes, please refer to it if necessary

- JMP documentation
  - https://www.jmp.com/en/support/jmp-documentation

  Add-in
  - https://community.jmp.com/t5/JMP-Add-Ins/JMP-Documentation-Add-In/ta-p/722342

- JMP Learning Library
  - https://www.jmp.com/en_us/learning-library.html

- On-Demand Webinars
  - https://www.jmp.com/en_us/events/ondemand/webcasts/by-application-area/all-application-areas.html

Supplementary material for JMP exercises

How to get the data files for this lecture

①From Webclass

②From Github

①Webclass

#0 Preparation for attending lectures

Usage of the statistics package (software) アンケート	更新 1ヵ月前 実行者数 2	···
SAS academic edition 資料	更新 1ヵ月前 実行者数 0	···
New Install_JMP 資料	更新 0分前 実行者数 1	···
Preparation for the lecture_Python 資料	更新 1ヵ月前 実行者数 1	···

#1_Introduction (1-A:JMP part, 1-B: Python part)

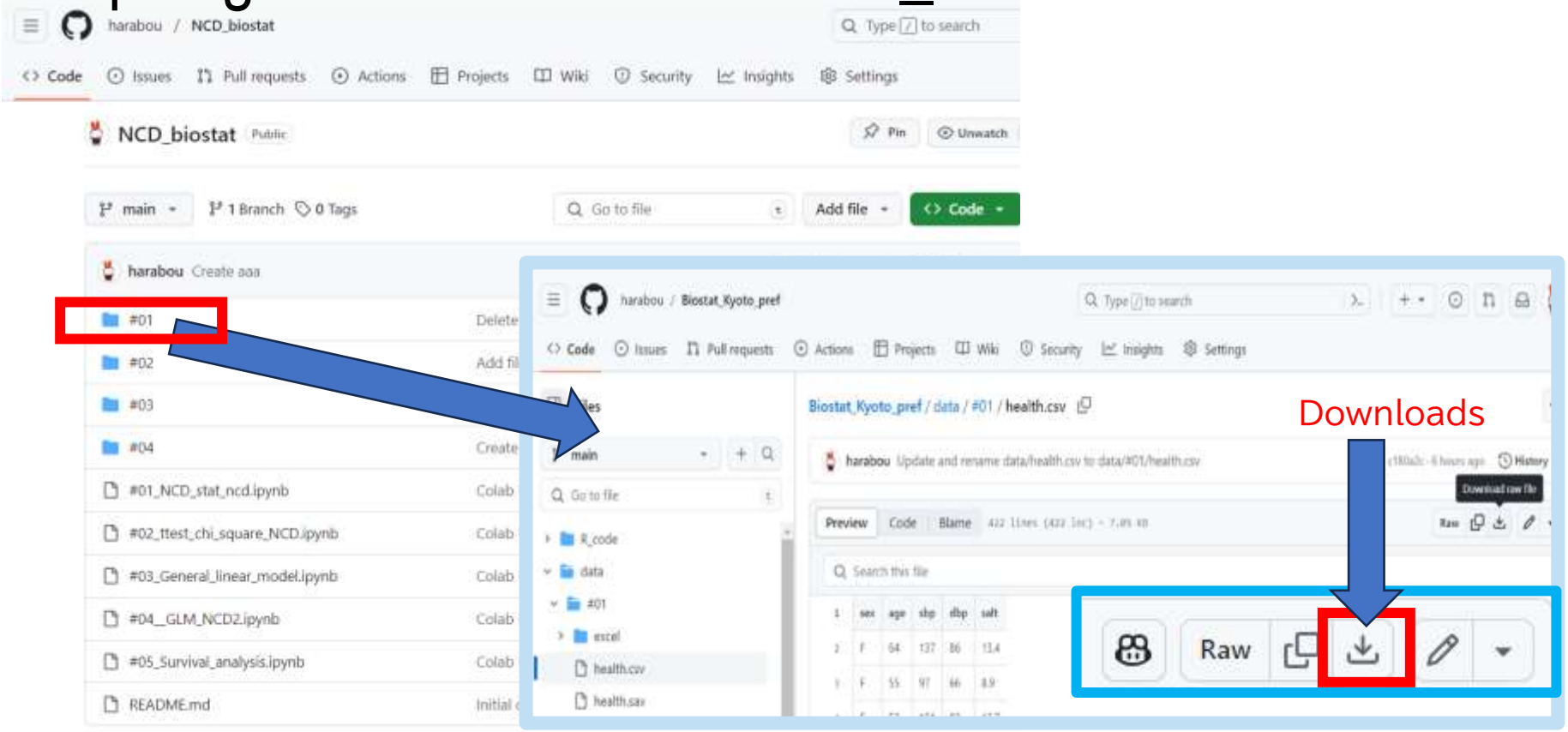
1_Introduction 資料	更新 14日前 実行者数 0	···
#01_JMP file 資料	更新 14日前 実行者数 0	···
#01_Python access 資料	更新 14日前 実行者数 0	···
#01_supporting material R 資料	更新 1ヵ月前 実行者数 0	···
New Report (for credit) 資料	更新 54分前 実行者数 0	···

資料を閉じる
NCD Biostatistics course
#01_JMP file
原田　亜紀子 さんがログイン中
前のページ
次のページ
①
目次を隠す
資料を閉じる
1節 #01_JMP file
1
添付資料
2節 csv
2
添付資料
Forbidden
ファイルのダウンロード - Google...
latte.shiga-med.ac.jp/webclass/file_d...
ファイルのダウンロード
②
» 01_NCD_Biostatistics.jmpprj
ファイルをダウンロードできない場合は
ファイル名をマウスの右ボタンでクリックし
[名前をつけて保存]を選択してください。
CSVファイルの文字コードはShift JISです。
このウィンドウを閉じる

②Github

■Github

- https://github.com/harabou/NCD_biostat



harabou / NCD_biostat

Type / to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

NCD_biostat

Public

Pin Unwatch

main 1 Branch 0 Tags Go to file t Add file Code

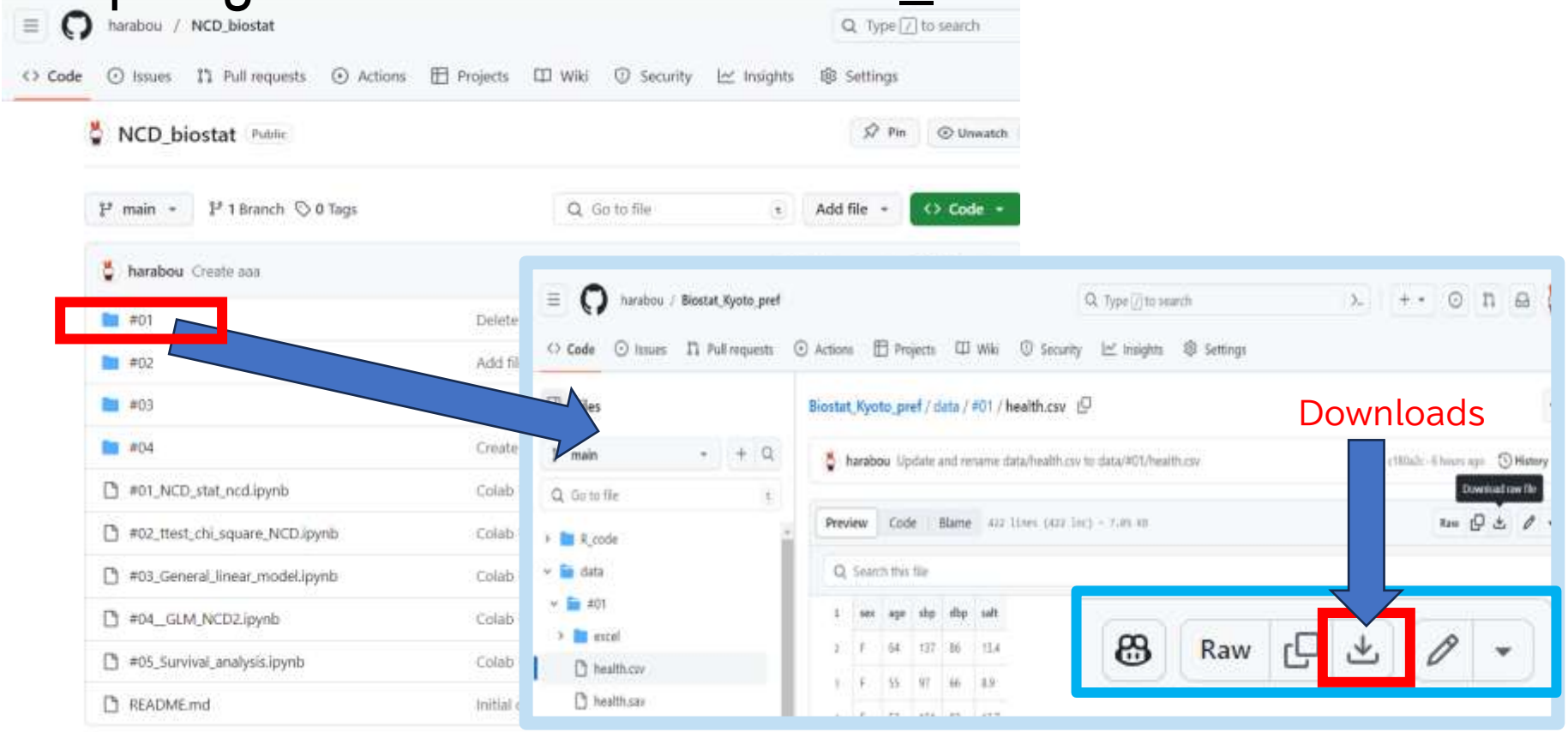
harabou Add files via upload 4550170 · 2 weeks ago 22 Commits

#01	Add files via upload	2 weeks ago
#02	Add files via upload	2 months ago
#03	Create aaa	2 months ago
#04	Create aaa	2 months ago
#01_NCD_stat_ncd.ipynb	Colab を使用して作成されました	2 months ago
#02_ttest_chi_square_NCD.ipynb	Colab を使用して作成されました	2 months ago
#03_General_linear_model.ipynb	Colab を使用して作成されました	2 months ago
#04__GLM_NCD2.ipynb	Colab を使用して作成されました	2 months ago
#05_Survival_analysis.ipynb	Colab を使用して作成されました	2 months ago

②Github

■Github

- https://github.com/harabou/NCD_biostat



harabou / NCD_biostat

Type / to search

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

Files

main

Go to file t

- #01
  - #01_NCD B[illegible]tics.jmpprj
  - Solubility.csv
- #02
- #03
- #04

NCD_biostat / #01 /

harabou Add files via upload

Name	Last commit message
..	
#01_NCD Biostatistics.jmpprj	Add files via upload
Solubility.csv	Add files via upload

NCD_biostat / #01 / Solubility.csv

harabou Add files via upload 4550170 · now

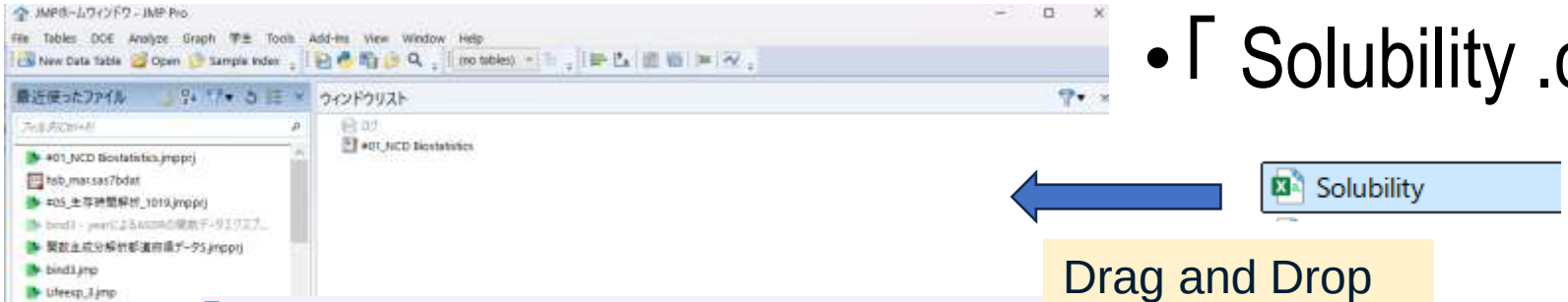
Preview | Code | Blame | 73 lines (73 loc) · 3.77 KB

Search this file

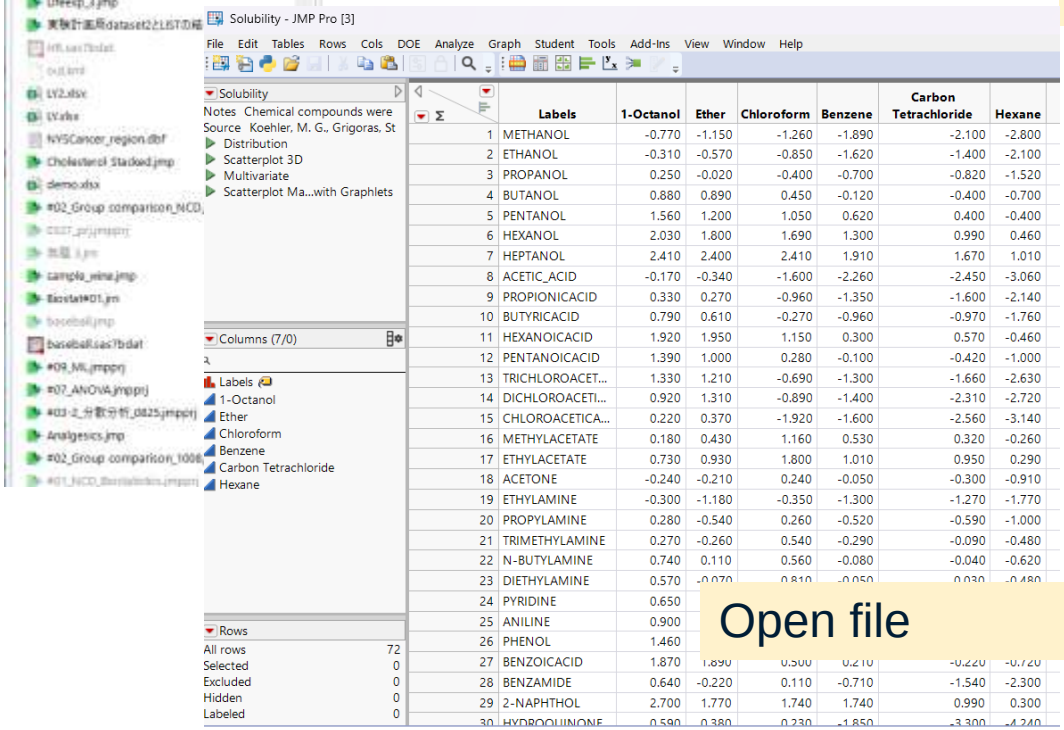
1	Labels	1-Octanol	Ether	Chloroform	Benzene	Carbon Tetrachloride	Hexane
2	METHANOL	-0.770	-1.150	-1.260	-1.890	-2.100	-2.800
3	ETHANOL	-0.310	-0.570	-0.850	-1.620	-1.400	-2.100
4	PROPANOL	0.250	-0.020	-0.400	-0.700	-0.820	-1.520
5	BUTANOL	0.880	0.890	0.450	-0.120	-0.400	-0.700
6	PENTANOL	1.560	1.200	1.050	0.620	0.400	-0.400
7	HEXANOL	2.030	1.800	1.690	1.300	0.990	0.460
8	HEPTANOL	2.410	2.400	2.410	1.910	1.670	1.010
9	ACETIC_ACID	-0.170	-0.340	-1.600	-2.260	-2.450	-3.060
10	PROPIONICACID	0.330	0.270	-0.960	-1.350	-1.600	-2.140

Import ①

- 「Solubility .csv」



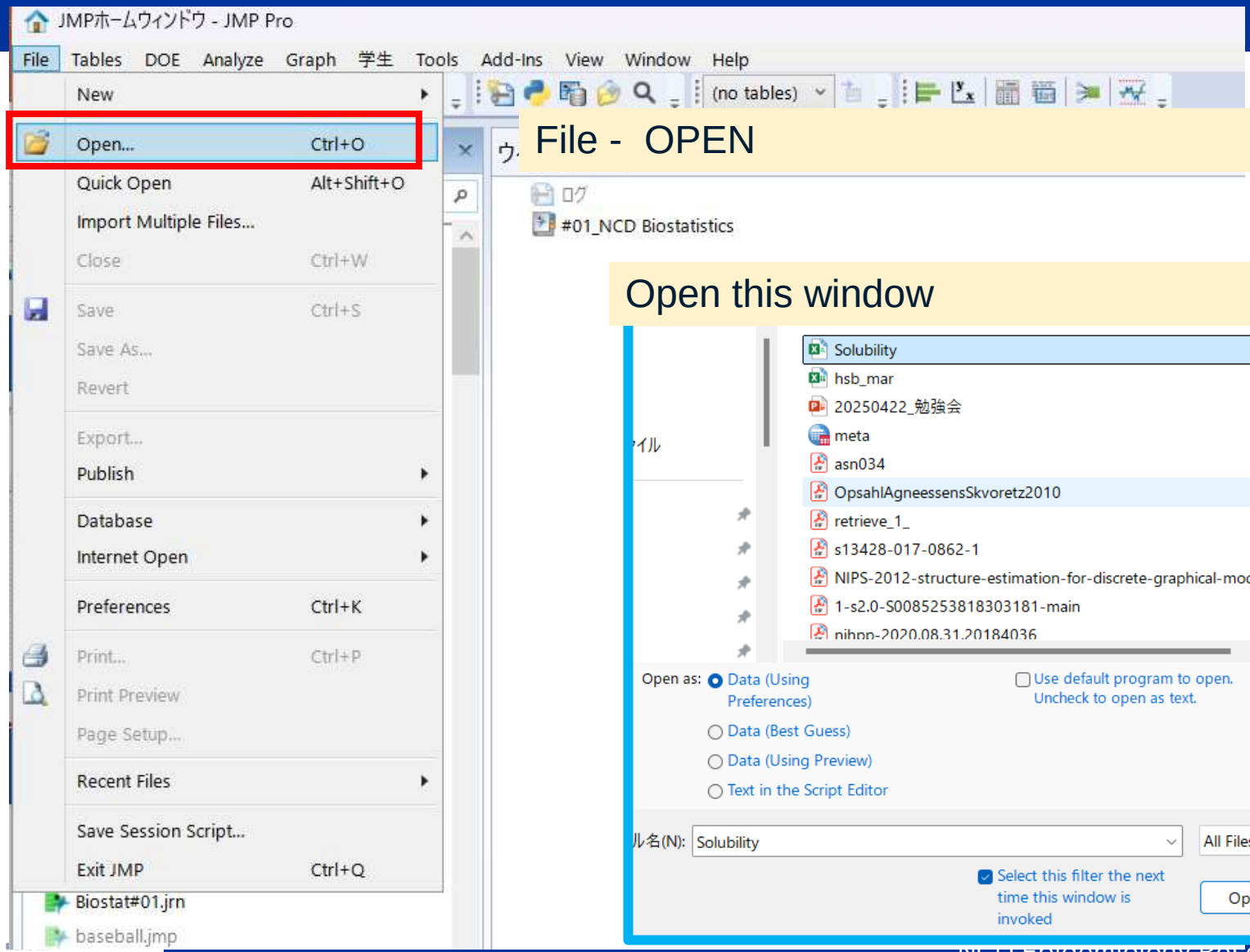
Drag and Drop



	Labels	1-Octanol	Ether	Chloroform	Benzene	Carbon Tetrachloride	Hexane
1	METHANOL	-0.770	-1.150	-1.260	-1.890	-2.100	-2.800
2	ETHANOL	-0.310	-0.570	-0.850	-1.620	-1.400	-2.100
3	PROPANOL	0.250	-0.020	-0.400	-0.700	-0.820	-1.520
4	BUTANOL	0.880	0.890	0.450	-0.120	-0.400	-0.700
5	PENTANOL	1.560	1.200	1.050	0.620	0.400	-0.400
6	HEXANOL	2.030	1.800	1.690	1.300	0.990	0.460
7	HEPTANOL	2.410	2.400	2.410	1.910	1.670	1.010
8	ACETIC_ACID	-0.170	-0.340	-1.600	-2.260	-2.450	-3.060
9	PROPIONICACID	0.330	0.270	-0.960	-1.350	-1.600	-2.140
10	BUTYRICACID	0.790	0.610	-0.270	-0.960	-0.970	-1.760
11	HEXANOICACID	1.920	1.950	1.150	0.300	0.570	-0.460
12	PENTANOICACID	1.390	1.000	0.280	-0.100	-0.420	-1.000
13	TRICHLOROACET...	1.330	1.210	-0.690	-1.300	-1.660	-2.630
14	DICHLOROACETI...	0.920	1.310	-0.890	-1.400	-2.310	-2.720
15	CHLOROACETICA...	0.220	0.370	-1.920	-1.600	-2.560	-3.140
16	METHYLACETATE	0.180	0.430	1.160	0.530	0.320	-0.260
17	ETHYLACETATE	0.730	0.930	1.800	1.010	0.950	0.290
18	ACETONE	-0.240	-0.210	0.240	-0.050	-0.300	-0.910
19	ETHYLAMINE	-0.300	-1.180	-0.350	-1.300	-1.270	-1.770
20	PROPYLAMINE	0.280	-0.540	0.260	-0.520	-0.590	-1.000
21	TRIMETHYLAMINE	0.270	-0.260	0.540	-0.290	-0.090	-0.480
22	N-BUTYLAMINE	0.740	0.110	0.560	-0.080	-0.040	-0.620
23	DIETHYLAMINE	0.570	-0.070	0.810	-0.050	0.030	-0.480
24	PYRIDINE	0.650					
25	ANILINE	0.900					
26	PHENOL	1.460					
27	BENZOICACID	1.870	1.890	0.500	0.210	-0.220	-0.720
28	BENZAMIDE	0.640	-0.220	0.110	-0.710	-1.540	-2.300
29	2-NAPHTHOL	2.700	1.770	1.740	1.740	0.990	0.300
30	HYDROQUINONE	0.590	0.380	0.230	-1.850	-3.300	-4.240

Open file

Import ②



File - OPEN

Open this window

File Import

#01_NCD Biostatistics - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph 学生

New	
Open...	Ctrl+O
Quick Open	Alt+Shift+O
Import Multiple Files...	
Close	Ctrl+W
Close Project	
Save	Ctrl+S
Save As...	
Revert	
Save Project	Ctrl+Shift+S
Save Project As...	
Export...	
Publish	
Database	
Internet Open	
Preferences	Ctrl+K
Print...	Ctrl+P
Print Preview	
Page Setup...	
Recent Files	
Save Session Script...	
Exit JMP	Ctrl+Q

All JMP Files
Data Files (*.jmp;*.xls;*.xlsx;*.xlsm;*.txt;*.csv;*.dat;*.tsv;*.xml;*.json;*.pdf;*.fcs;*.sav;*.sa
JMP Data Tables (*.jmp)
Excel Files (*.xlsx;*.xlsm;*.xls)
Text Files (*.txt;*.csv;*.dat;*.tsv)
JMP Scripts (*.jsl)
JMP Journals (*.jrn)
JMP Reports (*.jrp)
JMP Projects (*.jmpprj;*.jmpprjarchive)
JMP SQL Queries (*.jmpquery)
JMP Customizations (Add-Ins, Menus) (*.jmpaddin;*.jmpaddindef;*.def;*.jmpcust)
JMP Application Files (*.jmpappsource;*.jmpapp)
SAS Data Sets (*.sas7bdat;*.xpt;*.stx;*.v8xpt;*.xpt8)
Source Code (SAS, R, Python, MATLAB, SQL, C/C++, JavaScript) (*.sas,*.r;*.py;*.m;*.s
PDF Files (*.pdf)
HTML Files (*.htm;*.html)
XML Data Files (*.xml)
JSON Data Files (*.json)
FACS Files (*.fcs)
Genomics Text Files (*.txt;*.tsv;*.csv;*.gff;*.gff3;*.gtf;*.mtx;*.vcf)
IDAT Files (*.idat)
Binary Bed Files (*.bed)
SPSS Data Files (*.sav)
Triple-S Survey Files (*.sss;*.xml)
xBase Data Files (*.dbf)
Shapefiles (*.shp)
Teradata Databases (*.trd)
HDF5 Data Files (*.h5;*.hdf5;*.hdf)
SQLite database (*.sqlite;*.db;*.sqlite3;*.db3)
Minitab Portable Worksheet Files (*.mtp)

Various data sets can be read

After Import

JMP Pro [3]

Rows Cols DOE Analyze Graph Student Tools Add-Ins View Window Help

Solubility

Notes Chemical compounds were

Source Koehler, M. G., Grigoras, St

- Distribution
- Scatterplot 3D
- Multivariate
- Scatterplot Ma...with Graphlets

Columns (7/0)

- Labels
- 1-Octanol
- Ether
- Chloroform
- Benzene
- Carbon Tetrachloride
- Hexane

Rows

All rows	72
Selected	0
Excluded	0
Hidden	0
Labeled	0

	Labels	1-Octanol	Ether	Chloroform	Benzene	Carbon Tetrachloride	Hexane
1	METHANOL	-0.770	-1.150	-1.260	-1.890	-2.100	-2.800
2	ETHANOL	-0.310	-0.570	-0.850	-1.620	-1.400	-2.100
3	PROPANOL	0.250	-0.020	-0.400	-0.700	-0.820	-1.520
4	BUTANOL	0.880	0.890	0.450	-0.120	-0.400	-0.700
5	PENTANOL	1.560	1.200	1.050	0.620	0.400	-0.400
6	HEXANOL	2.030	1.800	1.690	1.300	0.990	0.460
7	HEPTANOL	2.410	2.400	2.410	1.910	1.670	1.010
8	ACETIC_ACID	-0.170	-0.340	-1.600	-2.260	-2.450	-3.060
9	PROPIONICACID	0.330	0.270	-0.960	-1.350	-1.600	-2.140
10	BUTYRICACID	0.790	0.610	-0.270	-0.960	-0.970	-1.760
11	HEXANOICACID	1.920	1.950	1.150	0.300	0.570	-0.460
12	PENTANOICACID	1.390	1.000	0.280	-0.100	-0.420	-1.000
13	TRICHLOROACET...	1.330	1.210	-0.690	-1.300	-1.660	-2.630
14	DICHLOROACETI...	0.920	1.310	-0.890	-1.400	-2.310	-2.720
15	CHLOROACETICA...	0.220	0.370	-1.920	-1.600	-2.560	-3.140
16	METHYLACETATE	0.180	0.430	1.160	0.530	0.320	-0.260
17	ETHYLACETATE	0.730	0.930	1.800	1.010	0.950	0.290
18	ACETONE	-0.240	-0.210	0.240	-0.050	-0.300	-0.910
19	ETHYLAMINE	-0.300	-1.180	-0.350	-1.300	-1.270	-1.770
20	PROPYLAMINE	0.280	-0.540	0.260	-0.520	-0.590	-1.000
21	TRIMETHYLAMINE	0.270	-0.260	0.540	-0.290	-0.090	-0.480
22	N-BUTYLAMINE	0.740	0.110	0.560	-0.080	-0.040	-0.620
23	DIETHYLAMINE	0.570	-0.070	0.810	-0.050	0.030	-0.480
24	PYRIDINE	0.650	0.080	1.430	0.410	0.230	-0.210
25	ANILINE	0.900	0.850	1.420	1.000	0.600	-0.300
26	PHENOL	1.460	1.640	0.370	0.360	-0.360	-0.700
27	BENZOICACID	1.870	1.890	0.500	0.210	-0.220	-0.720
28	BENZAMIDE	0.640	-0.220	0.110	-0.710	-1.540	-2.300
29	2-NAPHTHOL	2.700	1.770	1.740	1.740	0.990	0.300
30	HYDROQUINONE	0.590	0.380	0.230	-1.850	-3.300	-4.240

List of variable

Continuous

Ordinal

Nominal

Number of data read

Change variable types

Solubility - JMP Pro [3]

File Edit Tables Rows Cols DOE Analyze Graph Student

Solubility

Notes Chemical compounds were

Source Koehler, M. G., Grigoras, St

- Distribution
- Scatterplot 3D
- Multivariate
- Scatterplot Ma...with Graphlets

Columns (7/1)

- Labels
- 1-Octanol
- Ether
- Chlorof
- Benzen
- Carbon
- Hexane

Filter

Column Info...

Columns Manager

Delete Columns

Label/Unlabel

View or cha and propert

Σ	Labels
1	METHANOL
2	ETHANOL
3	PROPANOL
4	BUTANOL
5	PENTANOL
6	HEXANOL
7	HEPTANOL
8	ACETIC_ACID
9	PROPIONICAC
10	BUTYRICACI
11	HEXANOICACI
12	PENTANOICAC
13	TRICHLOROAC
14	DICHLOROACI
	HLOROACETI
	CETONE
	THYLAMINE
	ROPYLAMINE
	RIMETHYLAM

1-Octanol - JMP Pro [3]

'1-Octanol' in table 'Solubility'

Column Name 1-Octanol

Data Type Numeric

Modeling Type Continuous

Format Fixed Dec Width 8 Dec 3

Use thousands separator (,)

Column Properties

Notes

Remove

Notes

Click to edit and key in comments about column

Solvent - Use as Y variable.

OK

Cancel

Apply

Help

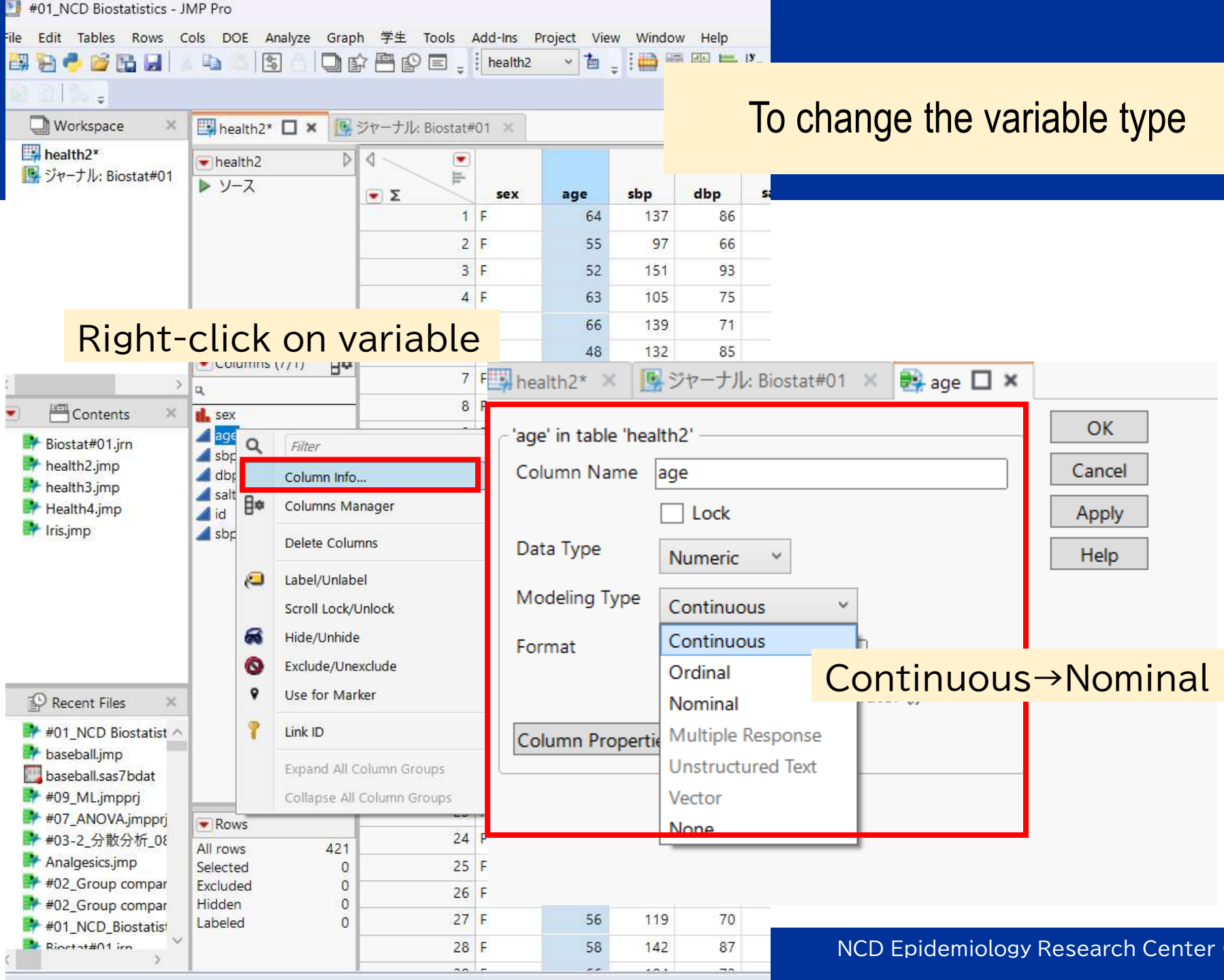
List of variable

Continuous

Ordinal

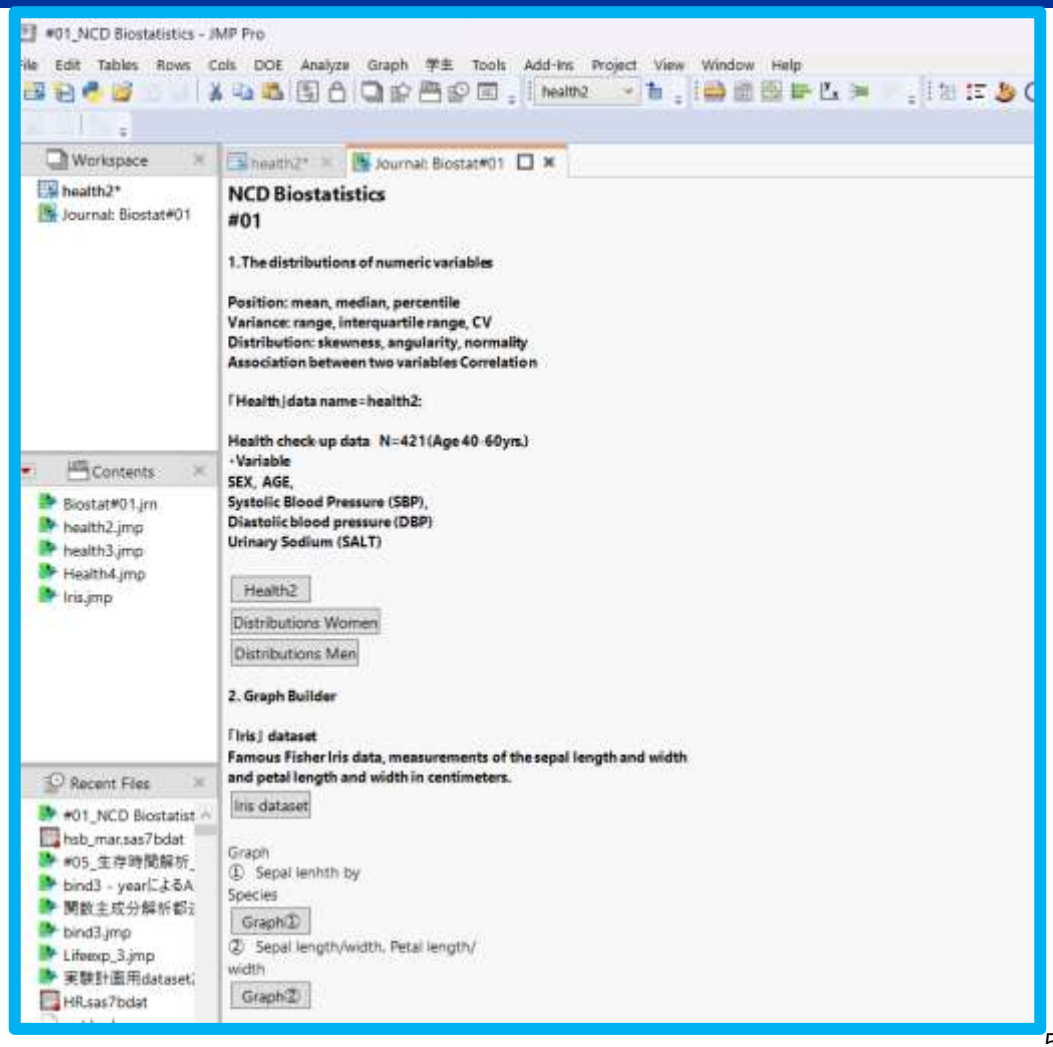
Nominal

To change the variable type



#01_NCD Biostatistics .jmppr

- 「#01_NCD Biostatistics .jmppr 」



From Webclass
「 NCD Biostatistics course 」

#0 Preparation for attending lectures

New
Usage of the statistics package (software)
アンケート

New
SAS academic edition
資料

New
Install JMP
資料

New
Preparation for the lecture Python
資料

#1_Introduction

New
1 Introduction
資料
更新 2日前
実行者数 0

New
#01 JMP file
資料
更新 2日前
実行者数 0

New
#01 Python access
資料
更新 2日前
実行者数 0

New
Course Report (B course)
資料
更新 1時間前
実行者数 0

New
Report Q2-Q4
更新 2日前
実行者数 0

教材を作成する

教材 資料を閉じる

> NCD Biostatistics course
#01_JMP file

原田 亜紀子 さんがログイン中

前のページ 次のページ

目次を隠す 資料を閉じる

click

第1節 #01_JMP file 1 添付資料

Forbidden

You don't have permission to access this resource.

#4 Generalized linear models (Logistic, Poisson)
Survival analysis
#6 Primcipal component analysis, Factor analysis, Cluster analysis,

ファイルのダウンロード

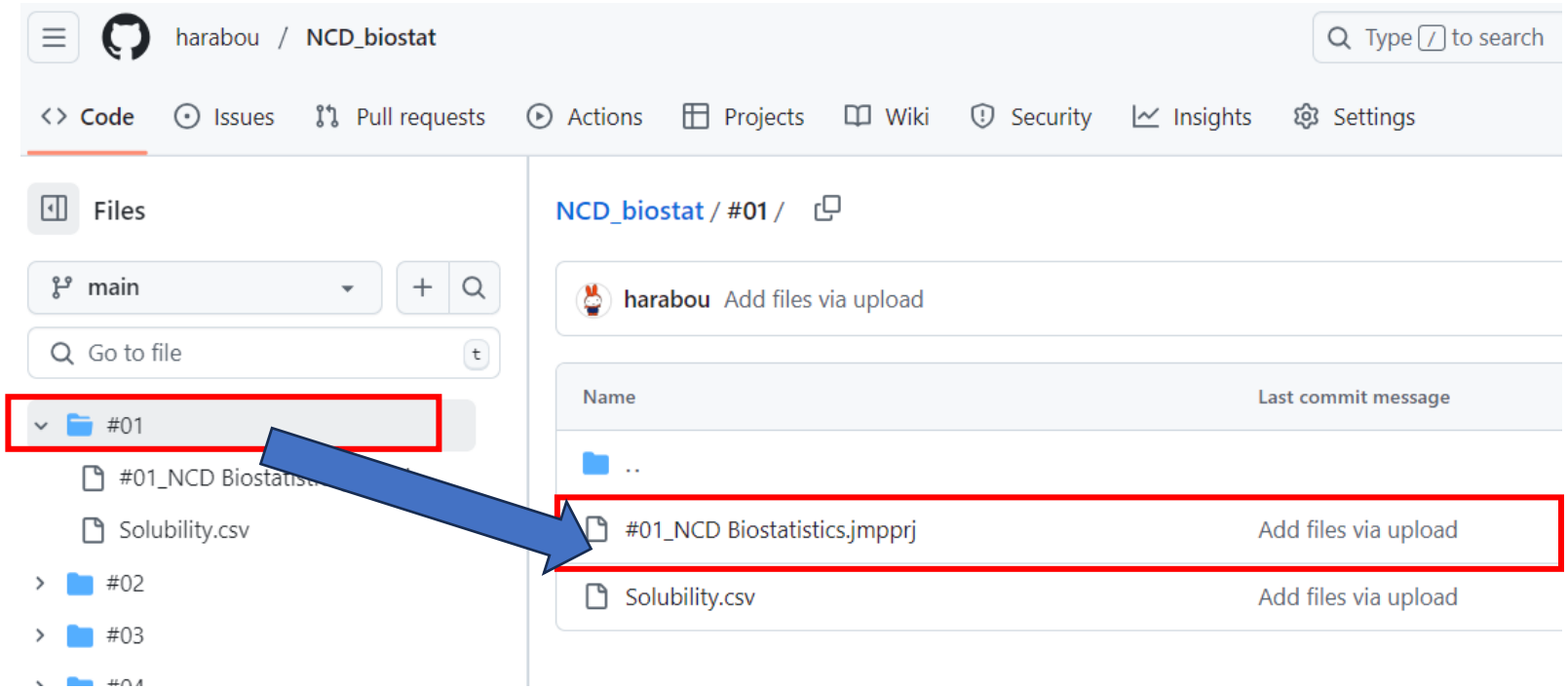
» 01_NCD_Biostatistics.jmpprj

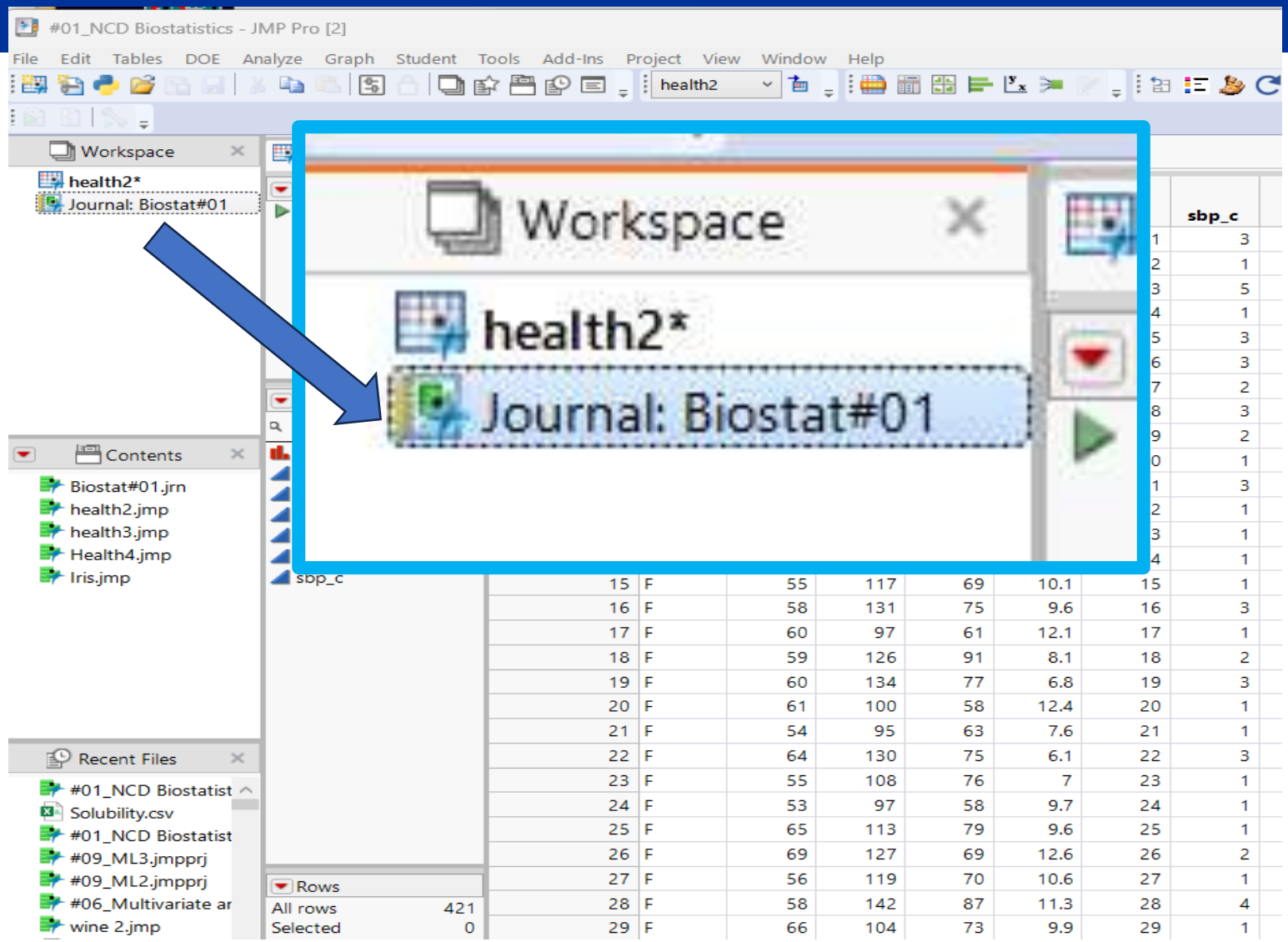
ファイルをダウンロードできない場合は
ファイル名をマウスの右ボタンでクリックし
[名前をつけて保存]を選択してください。

CSVファイルの文字コードはShift JISです。

このウィンドウを閉じる

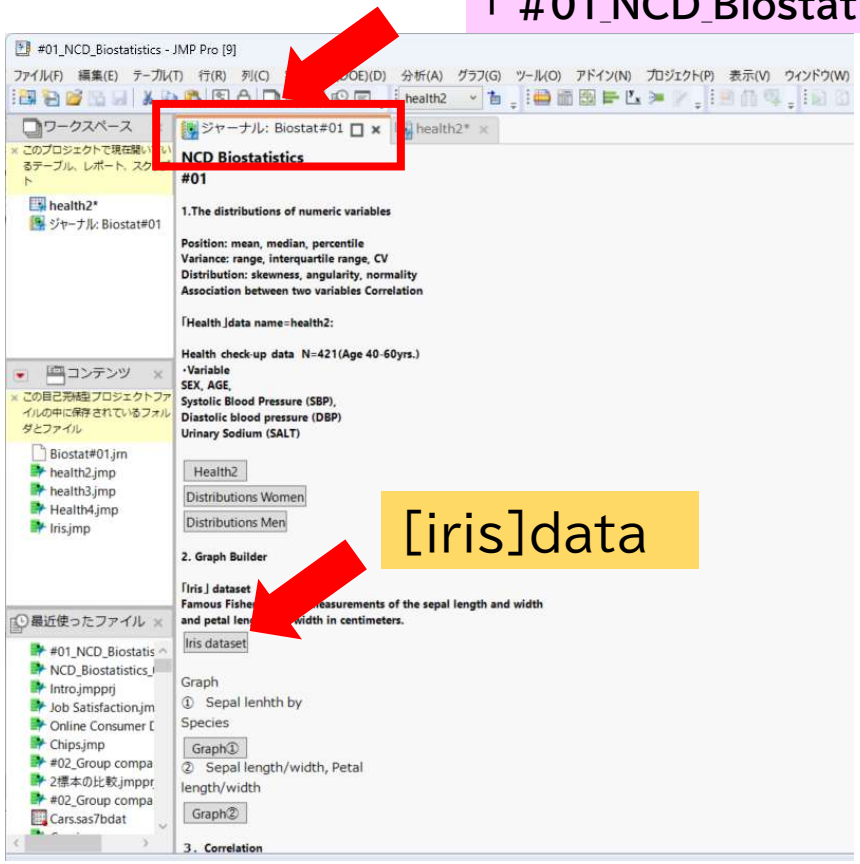
From Github


harabou / NCD_biostat
Type / to search
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
Files
main
Go to file
t
#01
#01_NCD Biostatist
Solubility.csv
#02
#03
NCD_biostat / #01 /
harabou Add files via upload
Name
Last commit message
..
#01_NCD Biostatistics.jmpprj
Add files via upload
Solubility.csv
Add files via upload


#01_NCD Biostatistics - JMP Pro [2]
File Edit Tables DOE Analyze Graph Student Tools Add-Ins Project View Window Help
health2
Workspace
health2*
Journal: Biostat#01
Contents
Biostat#01.jrn
health2.jmp
health3.jmp
Health4.jmp
Iris.jmp
Recent Files
#01_NCD Biostatist
Solubility.csv
#01_NCD Biostatist
#09_ML3.jmpprj
#09_ML2.jmpprj
#06_Multivariate ar
wine 2.jmp
sbp_c
Rows
All rows 421
Selected 0

JMP file 『#01_NCD_Biostatistics』

JMP project file
「#01_NCD_Biostatistics」



[iris]data

JMP project files

- Creation of portal
- Screens Button
- Data sets
- Analysis results

This data file can be distributed with the data set, analysis operation buttons and results in one package.

#01_NCD_Biostatistics - JMP Pro [9]

ファイル(F) 編集(E) テーブル(T) 行(R) 列(C) 実験計画(DOE)(D) 分析(A) グラフ(G) ツール(O) アドイン(N) プロジェクト(P) 表示(V) ウィンドウ(W)

health2

ワークスペース

このプロジェクトで現在開いているテーブル、レポート、スクリプト

health2*

ジャーナル: Biostat#01

コンテンツ

この自己完結型プロジェクトファイルの中に保存されているフォルダとファイル

Biostat#01.jrn
health2.jmp
health3.jmp
Health4.jmp
Iris.jmp

最近使ったファイル

#01_NCD_Biostatis
NCD_Biostatistics_
Intro.jmpprj
Job Satisfaction.jm
Online Consumer D
Chips.jmp
#02_Group compa
2標本の比較.jmppr
#02_Group compa
Cars.sas7bdat

ジャーナル: Biostat#01 | health2*

NCD Biostatistics
#01

1.The distributions of numeric variables

Position: mean, median, percentile
Variance: range, interquartile range, CV
Distribution: skewness, angularity, normality
Association between two variables Correlation

「Health」data name=health2:

Health check-up data N=421(Age 40-60yrs.)
・Variable
SEX, AGE,
Systolic Blood Pressure (SBP),
Diastolic blood pressure (DBP)
Urinary Sodium (SALT)

Health2

Distributions Women

Distributions Men

2. Graph Builder

「Iris」dataset
Famous Fisher Iris data, measurements of the sepal length and width and petal length and width in centimeters.

Iris dataset

Graph
① Sepal lenhth by Species

Graph①

② Sepal length/width, Petal length/width

Graph②

3. Correlation

[iris]data

『iris』dataset
Fisher's iris data

For JMP sample data, analysis examples (scripts) are included.

ジャーナル: Biostat#01 | health2* | Iris*

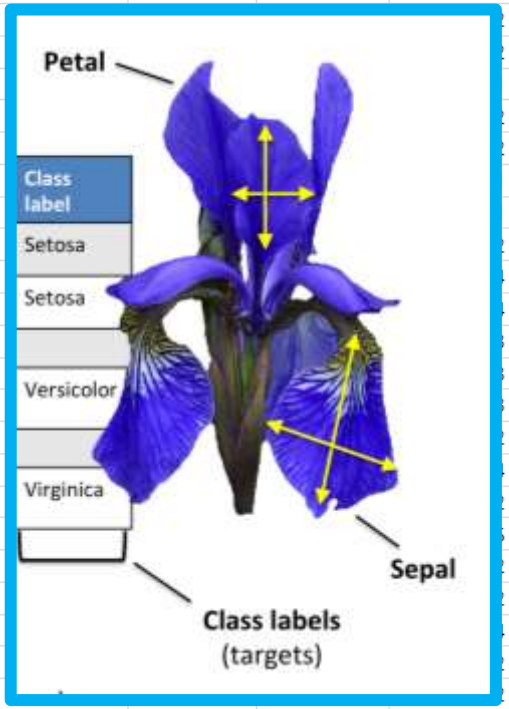
Iris
Reference D.F. Andrews, D.F. a
Notes Famous Fisher Iris data,
- Nominal Logistic
- Discriminant
- Scatterplot 3D
- Graph Builder Contour Plot
- Graph Builder Violin Plot
- Graph Builder Box Plot

Columns (5/0)
- Sepal length
- Sepal width
- Petal length
- Petal width
- Species

Rows

All rows	150
Selected	0
Excluded	0

	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8					setosa
9					setosa
10					setosa
11					setosa
12					setosa
13					setosa
14					setosa
15					setosa
16					setosa
17					setosa
18					setosa
19					setosa
20					setosa
21					setosa
22					setosa
23					setosa
24					setosa
25					setosa
26					setosa
27					setosa
28					setosa
29					setosa
30	4.7	3.2	1.6	0.2	setosa

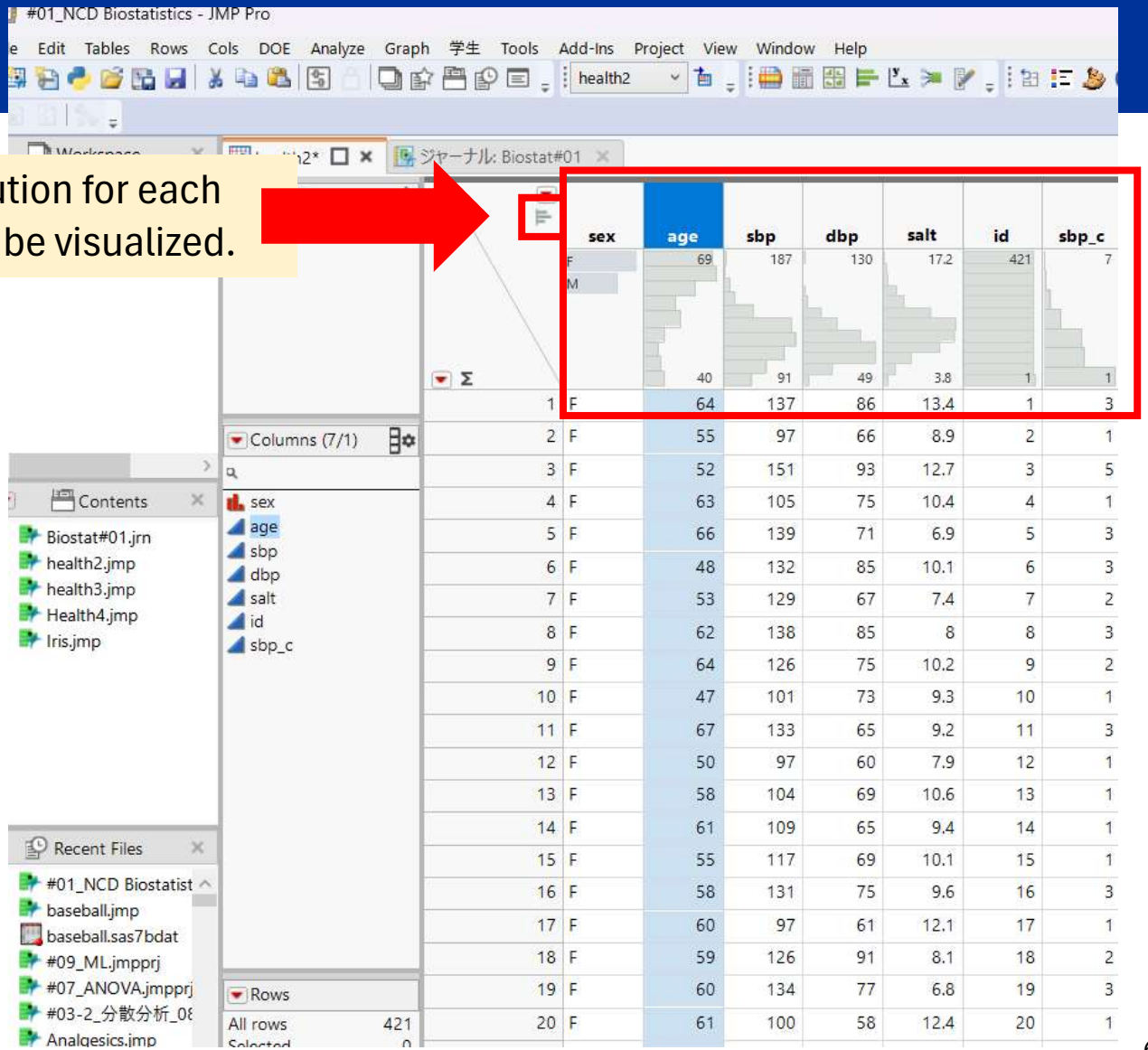


The distributions of numeric variables

Using JMP, try to output the following indicators

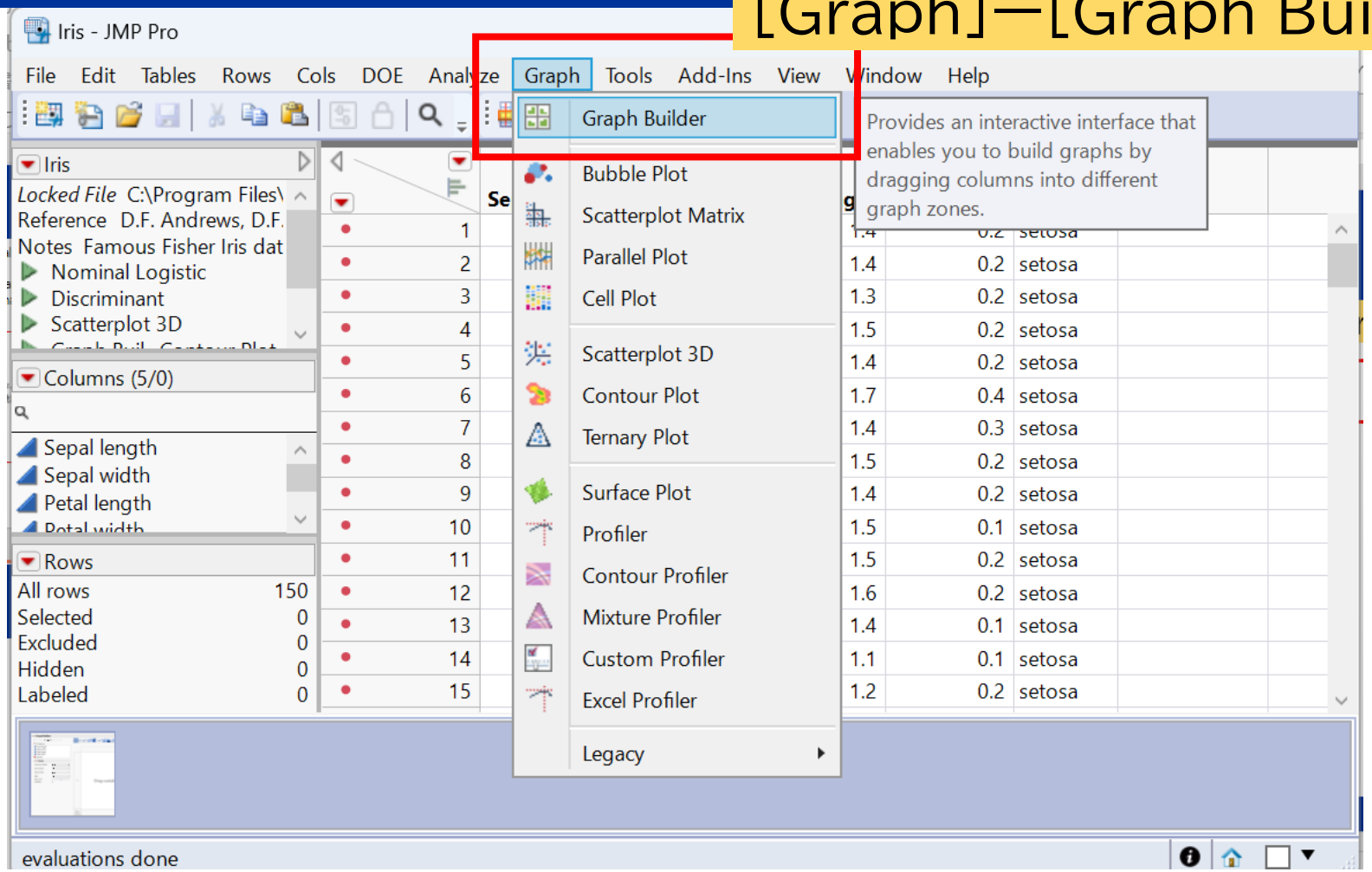
1）Creating graphs
2）Position: **mean, median, percentile**
3）Variance: **range, interquartile range, CV**
4）Distribution: **skewness, angularity, normality**
5）Association between two variables **Correlation**

Data distribution for each variable can be visualized.

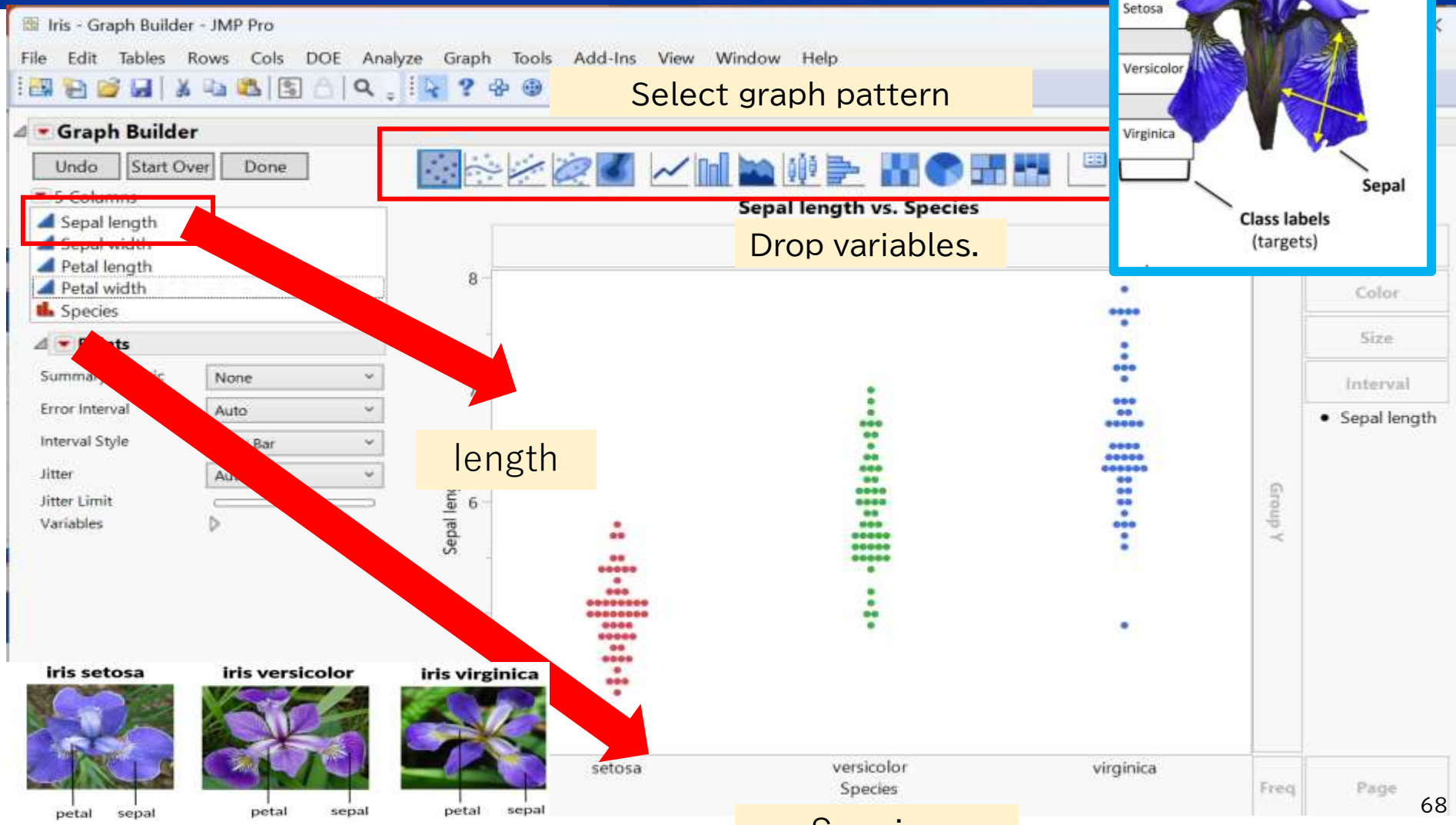


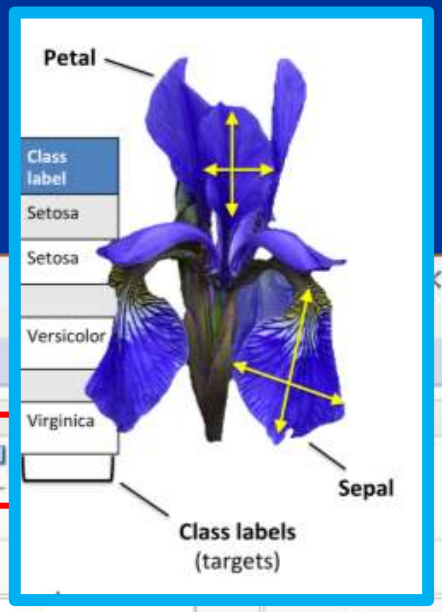
Graph Builder

[Graph]—[Graph Builder]

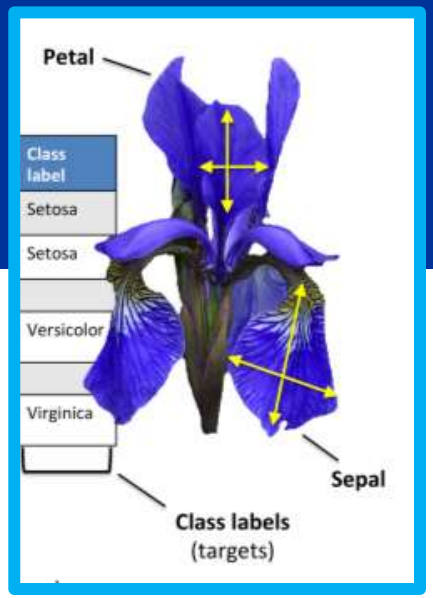


Graph Builder

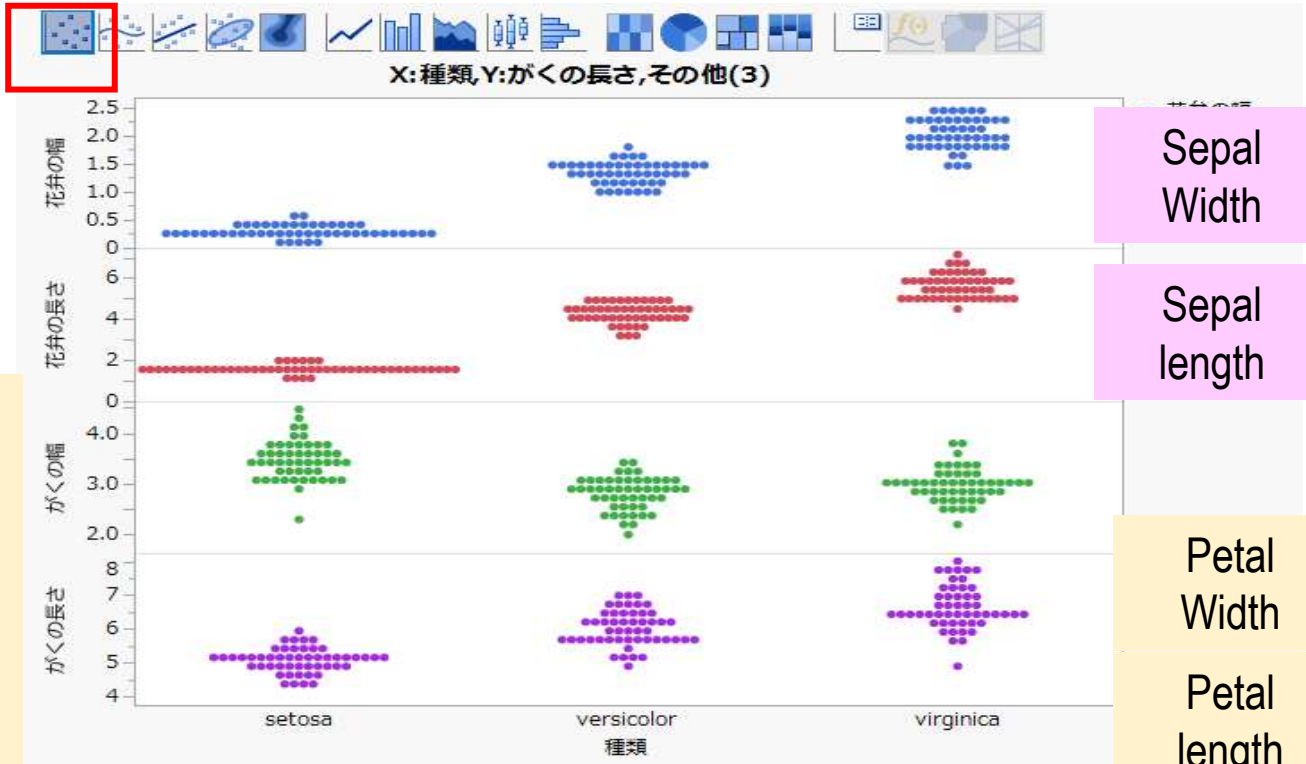




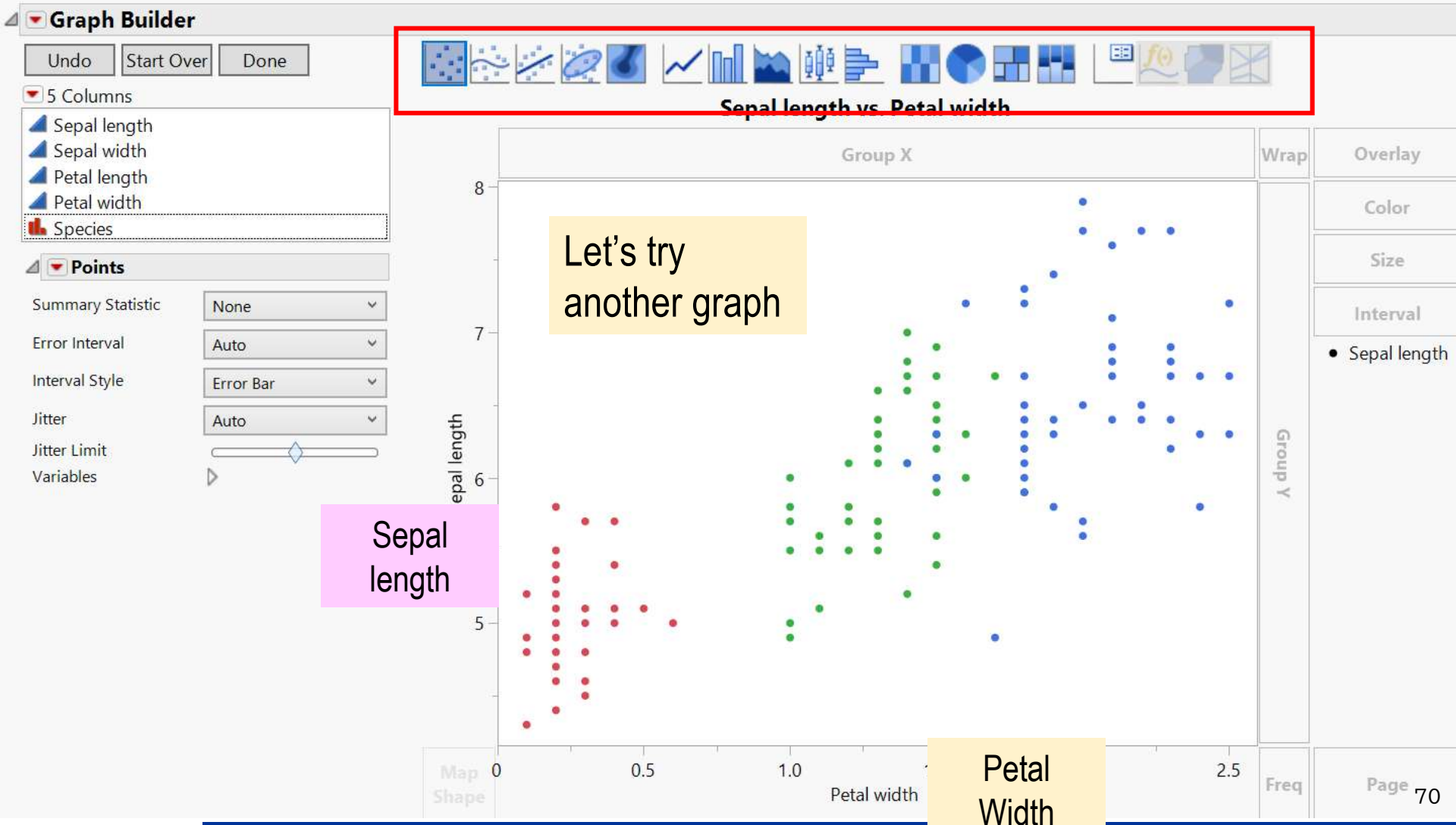
Graph Builder



Graphs can also be created by dropping other variables and arranging them vertically.



Graph Builder



#01_NCD_Biostatistics - JMP Pro [9]

ファイル(F) 編集(E) テーブル(T) 行(R) 列(C) 実験計画(DOE)(D) 分析(A) グラフ(G) ツール(O) アドイン(N) プロジェクト(P) 表示(V) ウィンドウ(W)

health2

ワークスペース

このプロジェクトで現在開いているテーブル、レポート、スクリプト

health2*

ジャーナル: Biostat#01

コンテンツ

この自己完結型プロジェクトファイルの中に保存されているフォルダとファイル

Biostat#01.jrn
health2.jmp
health3.jmp
Health4.jmp
Iris.jmp

最近使ったファイル

#01_NCD_Biostatis
NCD_Biostatistics_
Intro.jmpprj
Job Satisfaction.jm
Online Consumer [
Chips.jmp
#02_Group compa
2標本の比較.jmppr
#02_Group compa
Cars.sas7bdat

ジャーナル: Biostat#01 health2*

NCD Biostatistics
#01

1.The distributions of numeric variables

Position: mean, median, percentile
Variance: range, interquartile range, CV
Distribution: skewness, angularity, normality
Association between two variables Correlation

「Health」data name=health2:

Health check-up data N=421(Age 40-60yrs.)
・Variable
SEX, AGE,
Systolic Blood Pressure (SBP),
Diastolic blood pressure (DBP)
Urinary Sodium (SALT)

Health2

Distributions Women

Distributions Men

2. Graph Builder

「Iris」dataset
Famous Fisher Iris data, measurements of the sepal length and width and petal length and width in centimeters.

Iris dataset

Graph
① Sepal lenhth by Species

Graph①

② Sepal length/width, Petal length/width

Graph②

3. Correlation

[iris]data graph

Example

- Health check-up data N=421
  - Age 40-60yrs.
  - Variable
    - SEX, AGE, Systolic Blood Pressure (SBP), Diastolic blood pressure (DBP)
    - Urinary Sodium (SALT)
  - Dataset name=health

Health check-up data

ジャーナル: Biostat#01 health2* Iris*

NCD Biostatistics
#01

1.The distributions of numeric variables

Position: mean, median, percentile
Variance: range, interquartile range, CV
Distribution: skewness, angularity, normality
Association between two variables Correlation

「Health」data name=health2:

Health check-up data N=421(Age 40-60yrs.)
・Variable
SEX, AGE,
Systolic Blood Pressure (SBP),
Diastolic blood pressure (DBP)
Urinary Sodium (SALT)

Health2

Distributions Women

Distributions Men

2. Graph Builder

「Iris」dataset
Famous Fisher Iris data, measurements of the sepal length and width and petal length and width in centimeters.

Iris dataset

Graph
① Sepal lenhth by Species

Graph①

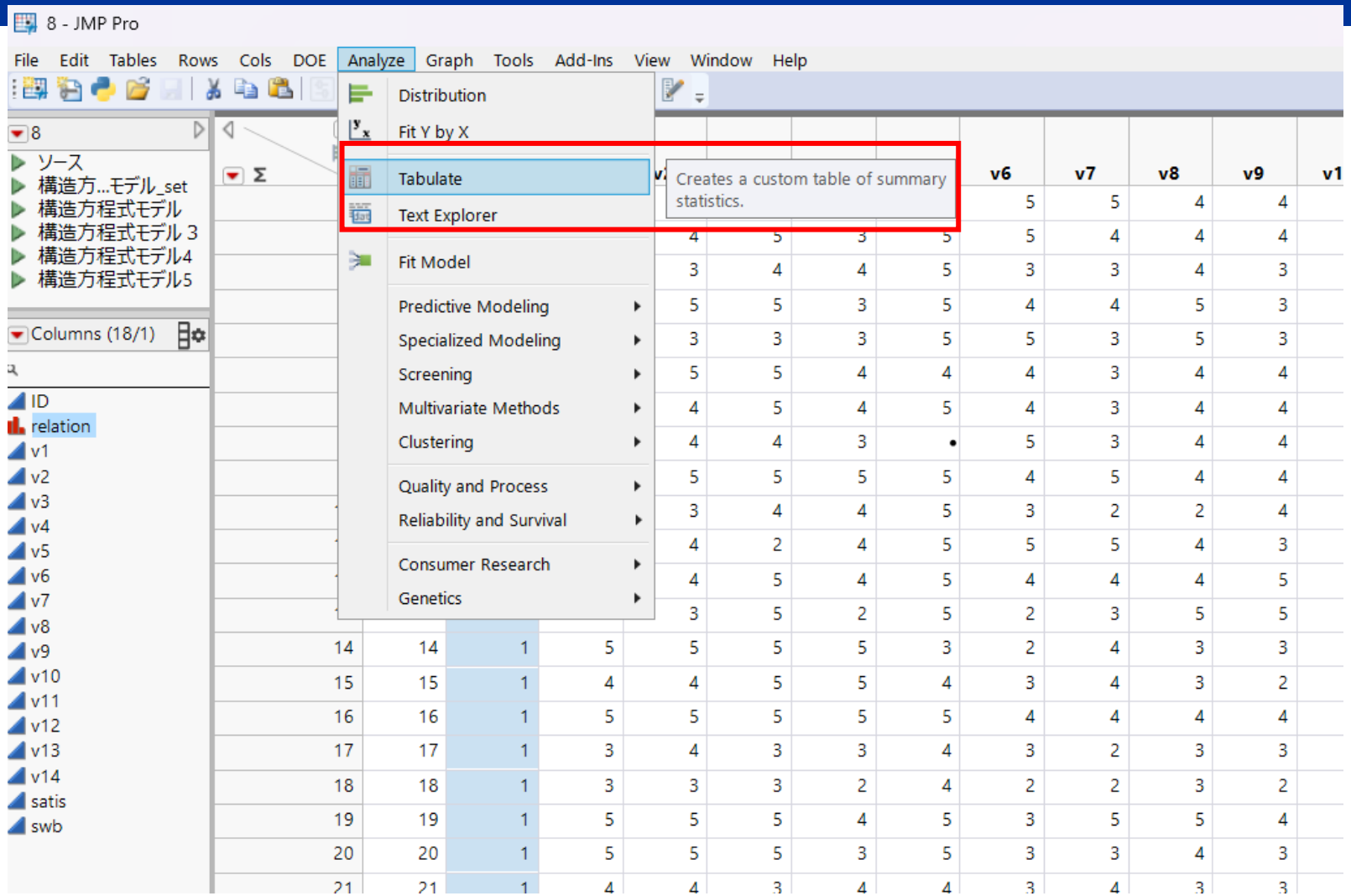
② Sepal length/width, Petal length/width

Graph②

3. Correlation

Health2

Tabulate
[Analyze]-[Tabulate]



Tabulate
[Analyze]-[Tabulate]

Drop and add variables and statistics

Tabulate

To add to the table, drag and drop columns or statistics into the column header or row label area of the table.

Undo | Start Over | Done

7 Columns

- sex
- age
- sbp
- dbp
- salt
- id
- sbp_c

Freq

Weight

ID

Page Column

Mean
Std Dev
Min
Range
% of Total
N Missing
N Categories
Sum
Sum Wgt
Variance
Std Err
CV
Median
Median Absolute Deviation
Geometric Mean

		sex	
		F	M
age	Mean	58.45	59.53
	Std Dev	8.06	7.45
sbp	Mean	122.73	125.49
	Std Dev	16.27	15.80
dbp	Mean	72.23	76.62
	Std Dev	9.50	10.70
salt	Mean	9.33	9.41
	Std Dev	2.10	1.93

Default Statistics

Format

☑ Use the same decimal format

○ Best
◉ Fixed Dec
○ Percent

Field Width: 10 Number of decimals 2

Set Format | Cancel | OK

Change Decimal point

Check the shape, location, and distribution indices of the distribution

1. Location
   1)Mean
   2)Median
   3)Mode
   4)Percentile

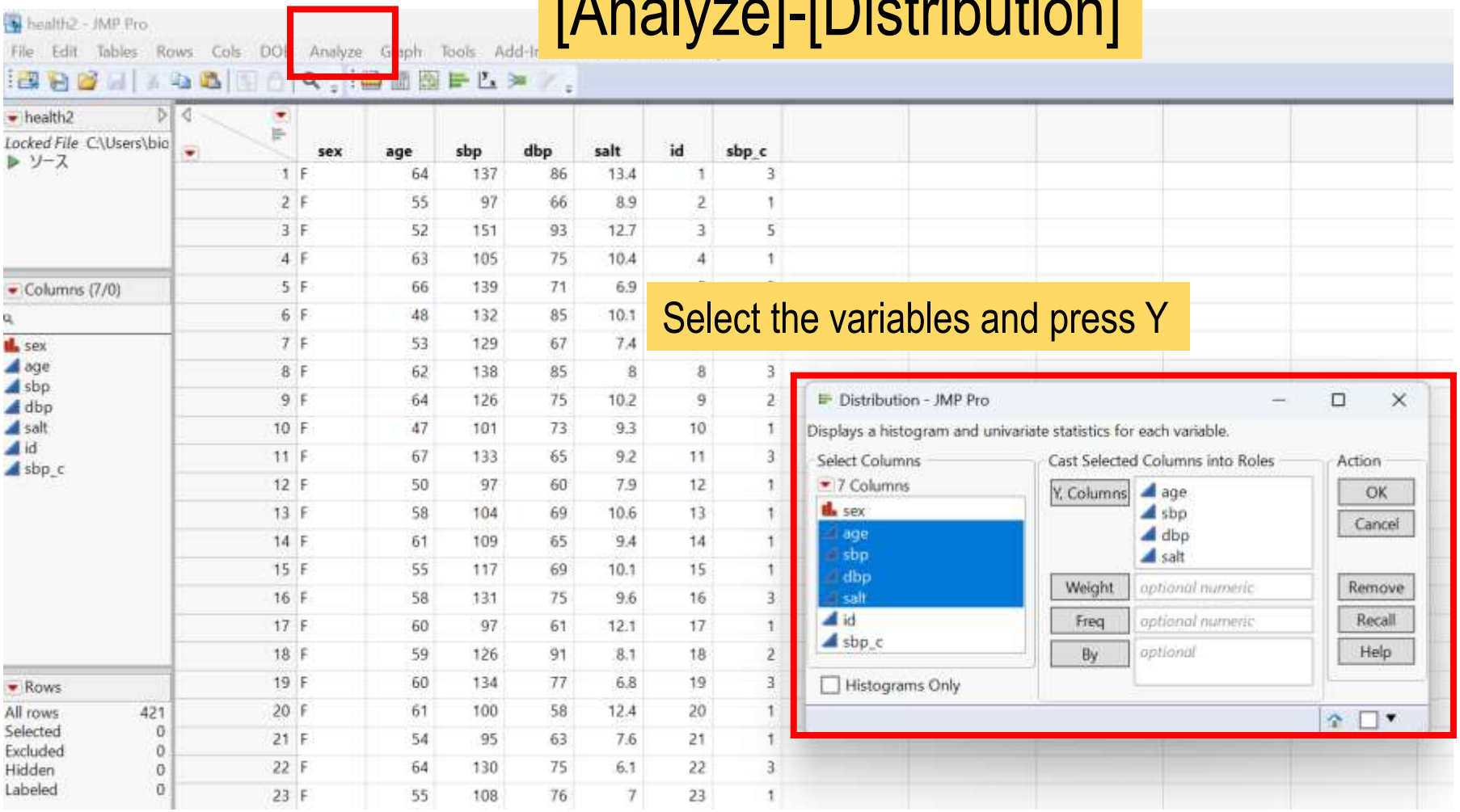
2. Variability dispersion
   1) Variance
   2) SD
   3) Range
   4) Inter-quartile range

3. Shape
   1) Skewness 歪度
   2) Kurtosis 尖度

Distribution

[Analyze]-[Distribution]

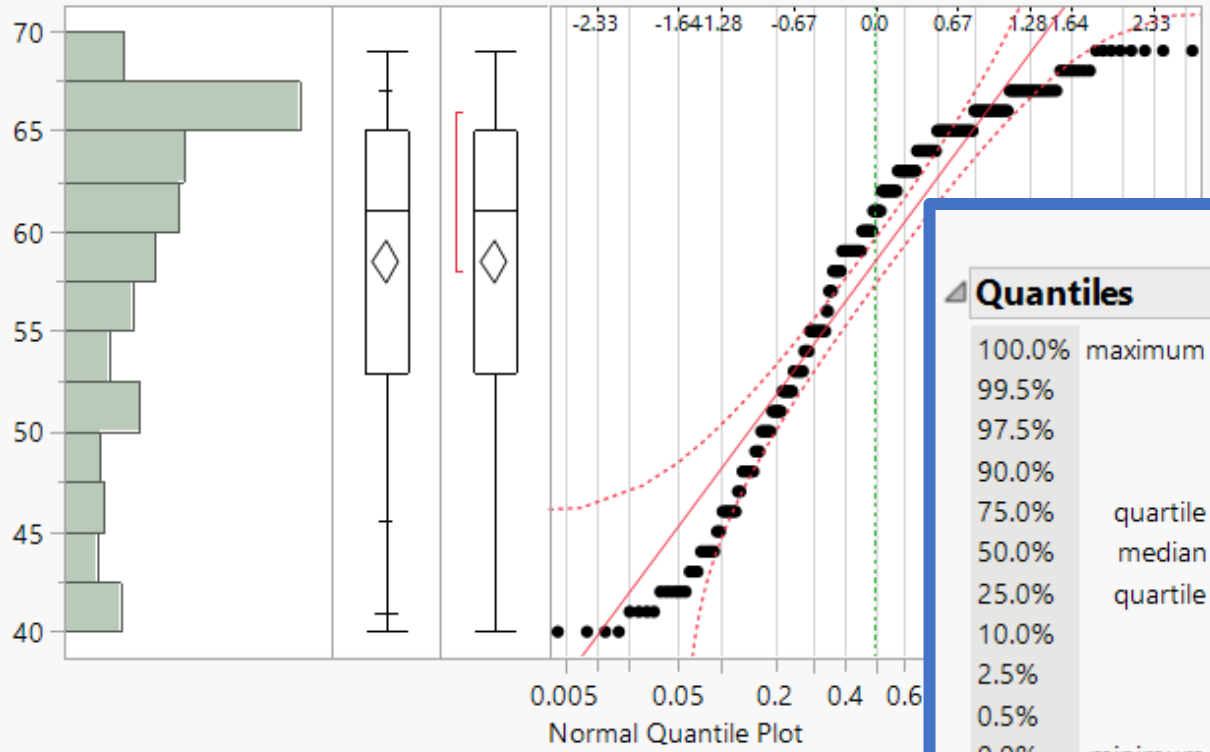
Select the variables and press Y



[Distribution]

Distributions

age



Quantiles

100.0%	maximum	69
99.5%		69
97.5%		69
90.0%		67
75.0%	quartile	65
50.0%	median	61
25.0%	quartile	53
10.0%		45.5
2.5%		41
0.5%		40
0.0%	minimum	40

Quantiles

100.0%	maximum	69
99.5%		69
97.5%		69
90.0%		67
75.0%	quartile	65
50.0%	median	61
25.0%	quartile	53
10.0%		45.5
2.5%		41
0.5%		40
0.0%	minimum	40

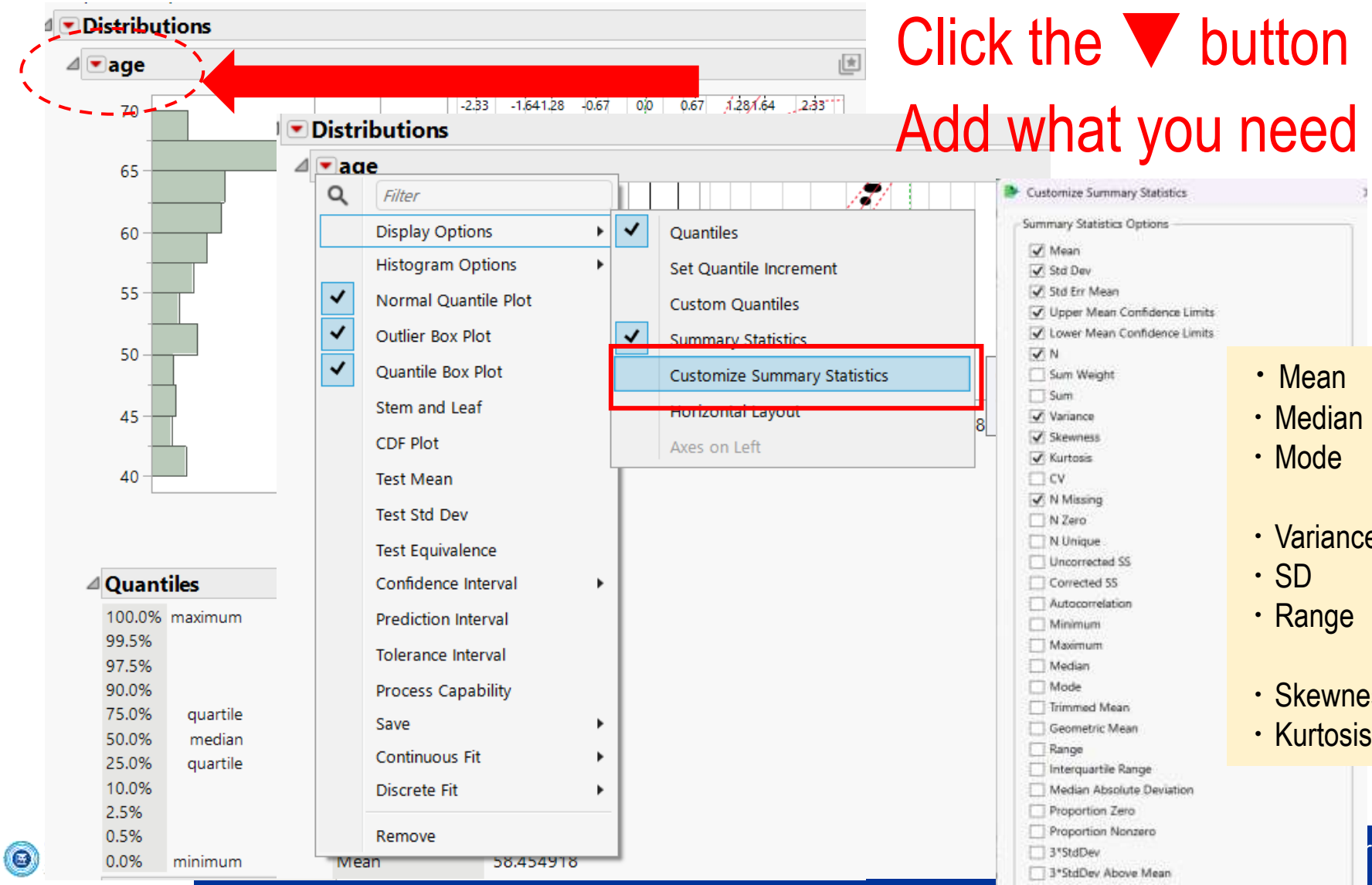
Summary Statistics

Mean	58.454918
Std Dev	8.0580466
Std Err Mean	0.5158636
Upper 95% Mean	59.471053
Lower 95% Mean	57.438783
N	244
Variance	64.932116
Skewness	-0.737276
Kurtosis	-0.587694
N Missing	0

When you want to add a statistic display

Click the ▼ button

Add what you need



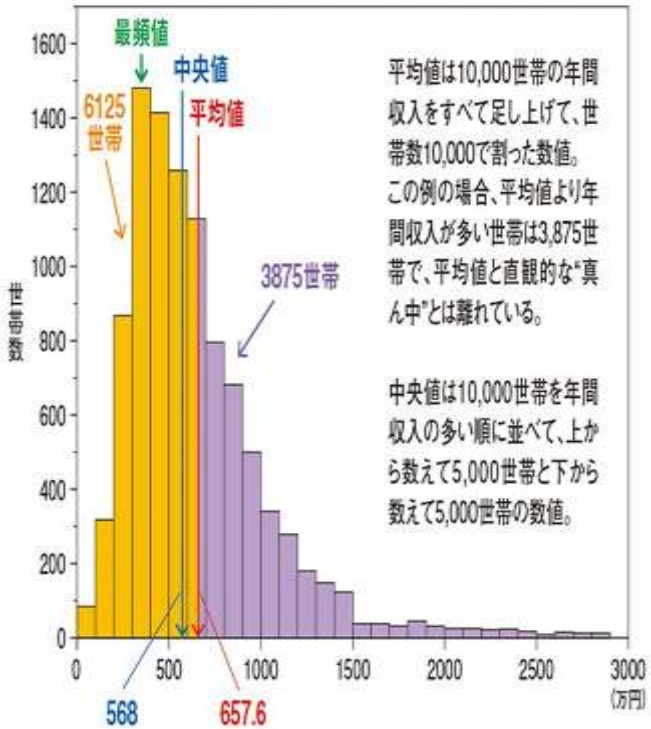
- Mean
- Median
- Mode

- Variance
- SD
- Range

- Skewness
- Kurtosis

Mean・Median・Mode

Example of
household income



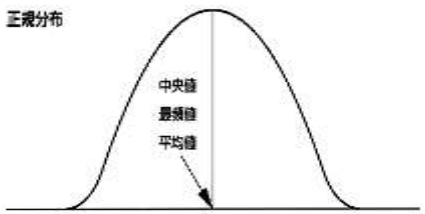
（左）

（右）

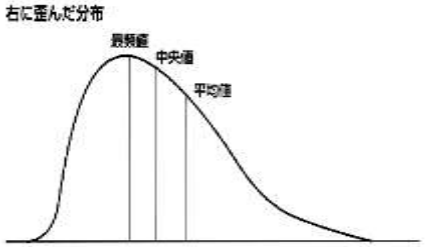
最頻値(Mode)
The class value with the largest frequency in the frequency distribution

中央値(Median)
Value that comes in the middle when sorted in ascending or descending order

平均値(Mean)
summing the sample data and dividing by the sample size



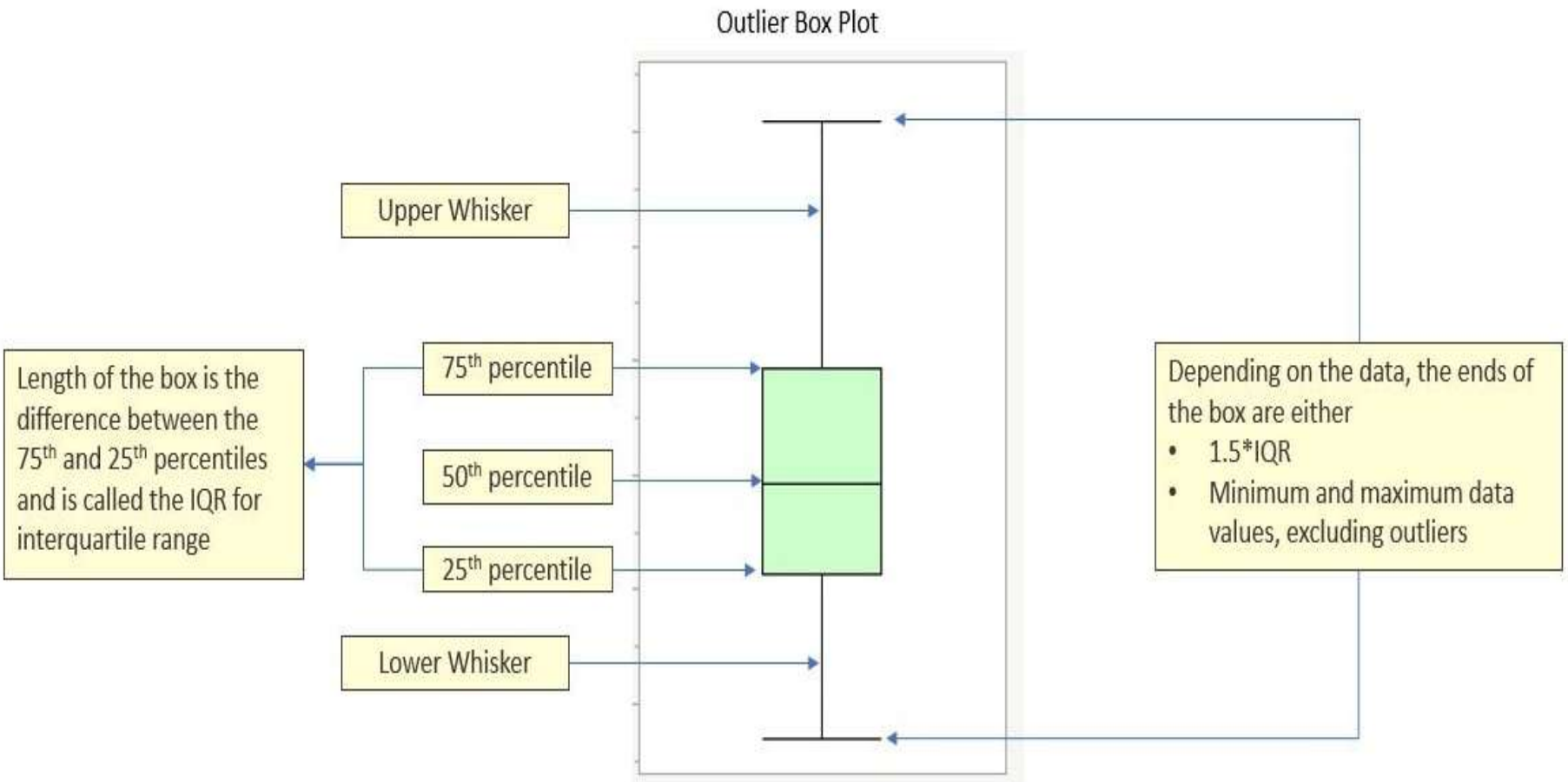
If the distribution is normal, the three indices are consistent



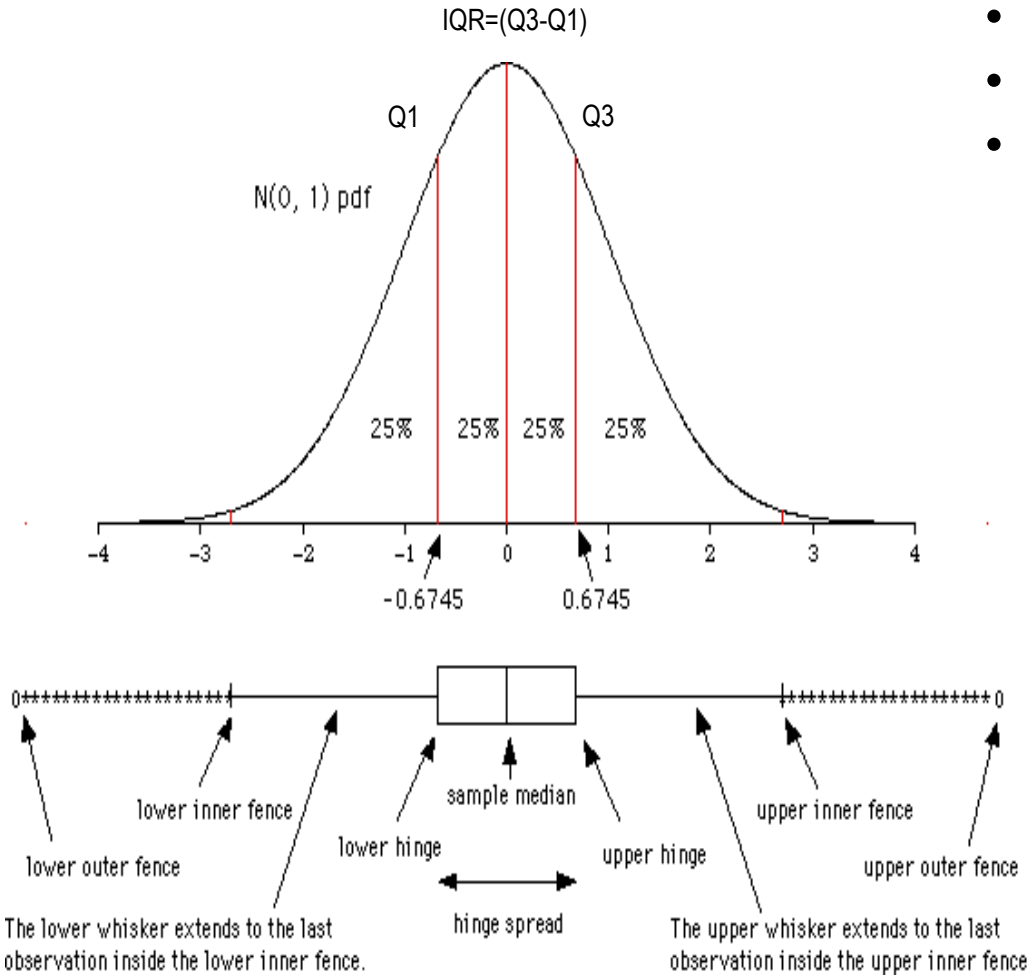
If the distribution is skewed to the right, From left to right, mode, median, and mean

図3　正規分布と右に歪んだ分布における代表値の関係

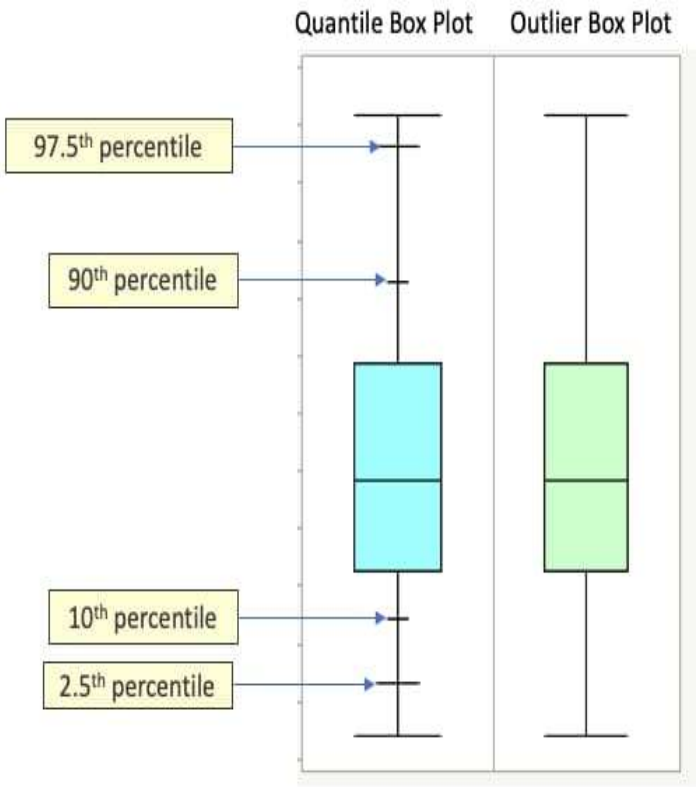
Boxplot


Outlier Box Plot
Upper Whisker
75th percentile
50th percentile
25th percentile
Lower Whisker
Length of the box is the difference between the 75th and 25th percentiles and is called the IQR for interquartile range
Depending on the data, the ends of the box are either
• 1.5*IQR
• Minimum and maximum data values, excluding outliers

Boxplot



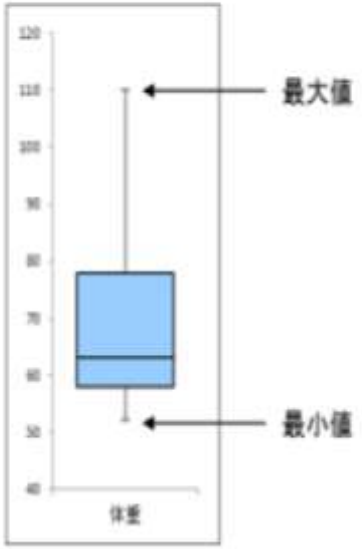
- Median is center of box
- 25th percentile is bottom of box
- 75th percentile is top of box



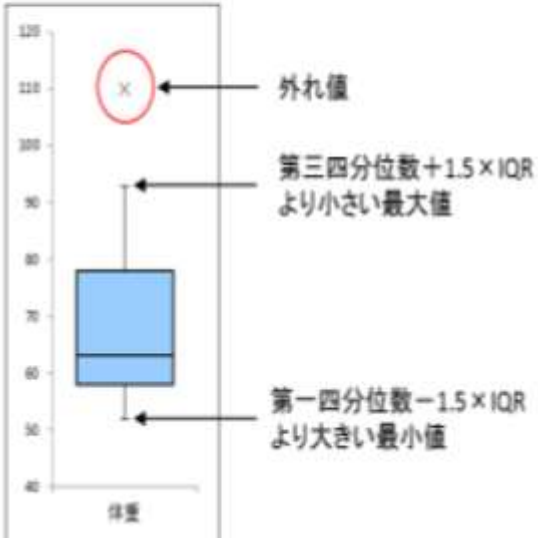
Boxplot

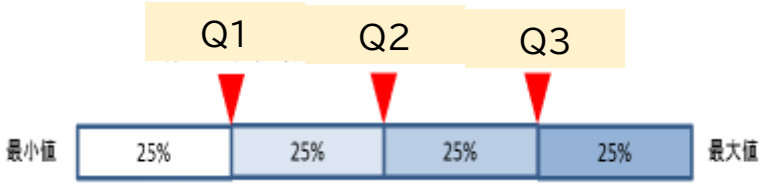
There are many different definitions of a boxplot

Upper and lower edges of the whiskers are maximum and minimum



Maximum length of the whiskers is 1.5 times the quartile range





Box plot in JMP

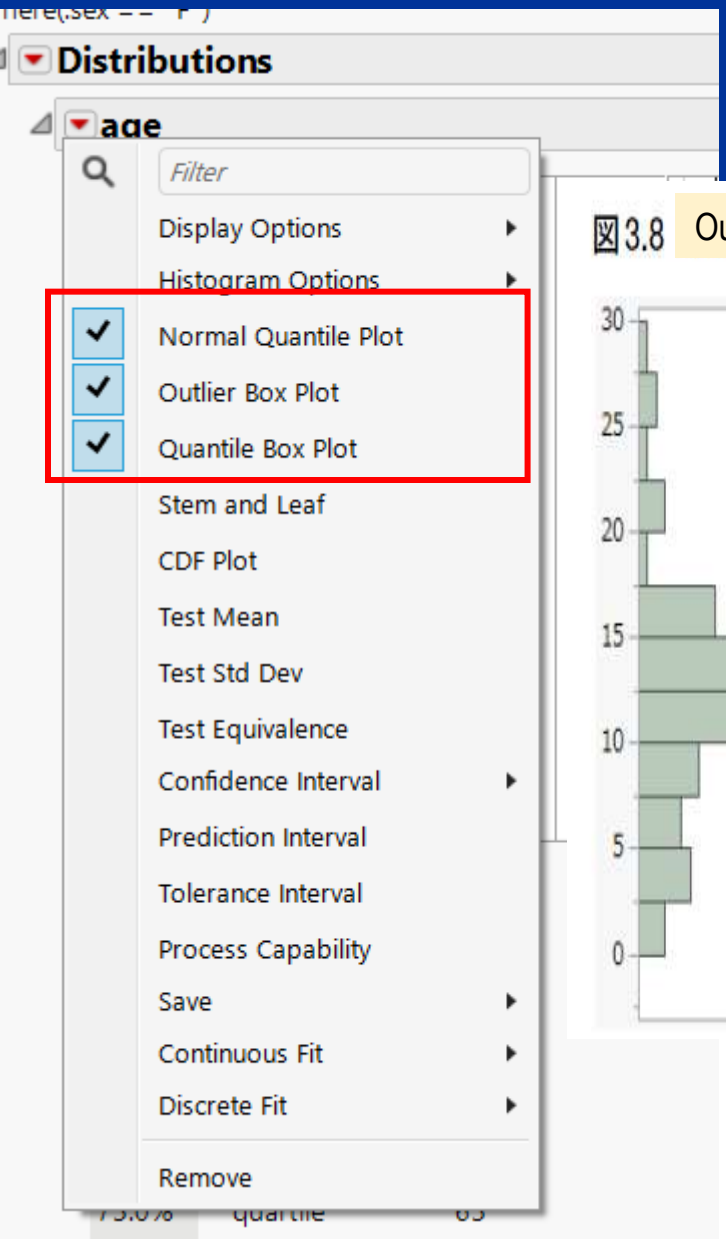


図3.8 Outlier Box plot

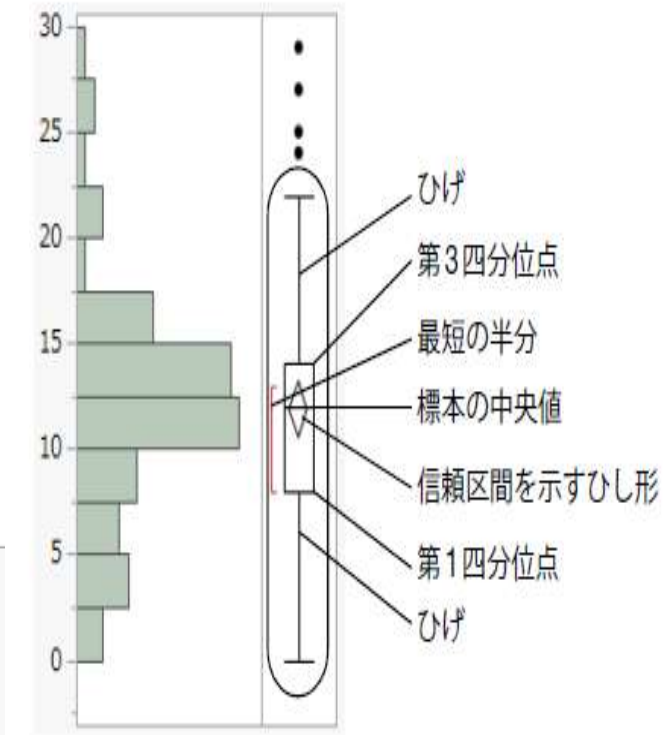
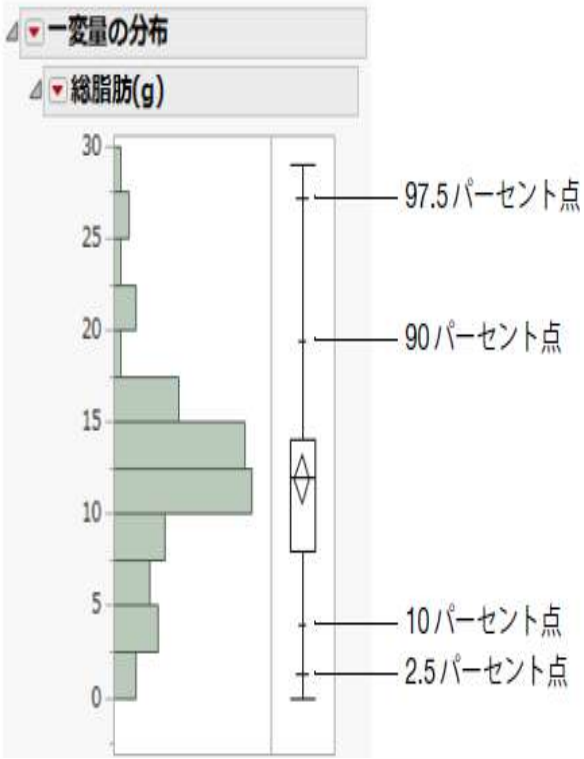
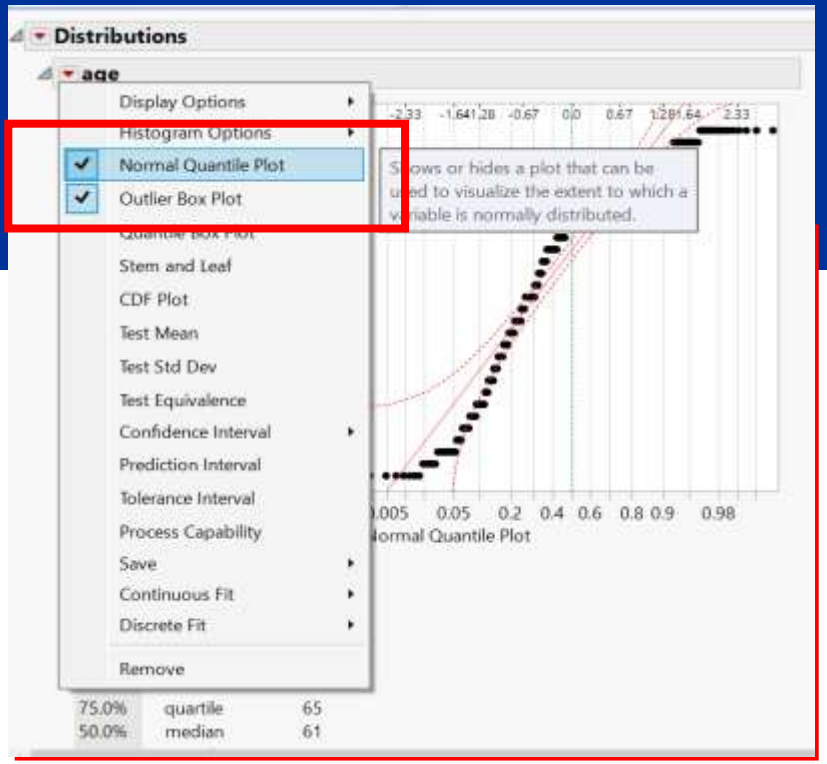


図3.9 Quantile Box plot

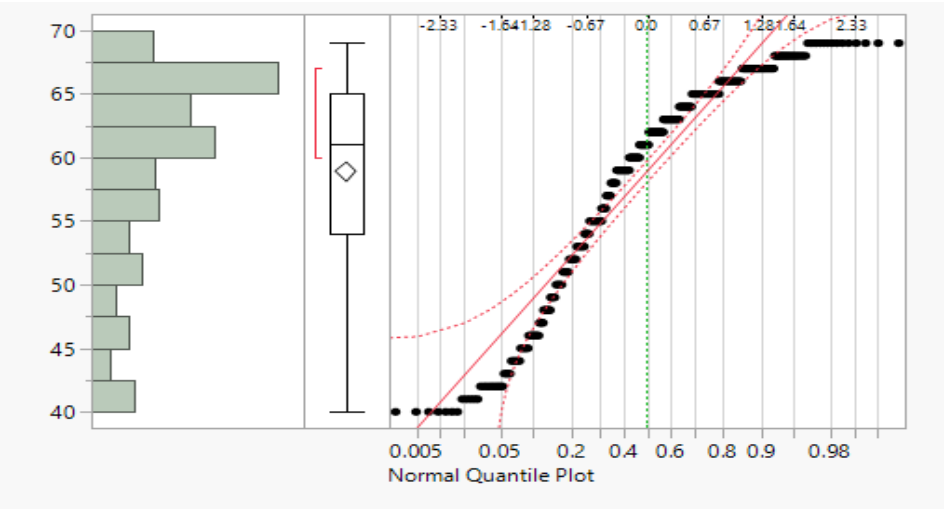


Skewness(b_1), Kurtosis(b_2)
Normal probability plot (Q-Qplot)



分位点		
100.0%	最大値	187
99.5%		179.23
97.5%		157.9
90.0%		144
75.0%	四分位点	135
50.0%	中央値	123
25.0%	四分位点	111.5
10.0%		101
2.5%		95.55
0.5%		93
0.0%	最小値	91

要約統計量	
平均	123.88836
標準偏差	16.115783
平均の標準誤差	0.7854352
平均の上側95%	125.43223
平均の下側95%	122.34449
N	421
歪度	0.3612619
尖度	0.3200493

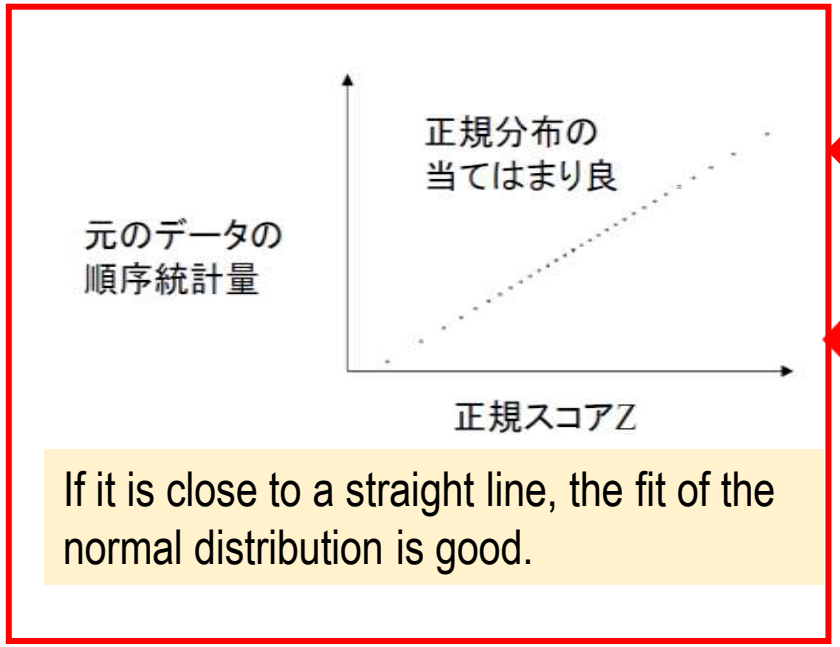


- A graphical method for evaluating whether a data set is approximately normally distributed
- The normal probability plot is constructed as follows
  - Vertical axis: ordered response values
  - Horizontal axis: median value of the normal ordered statistic

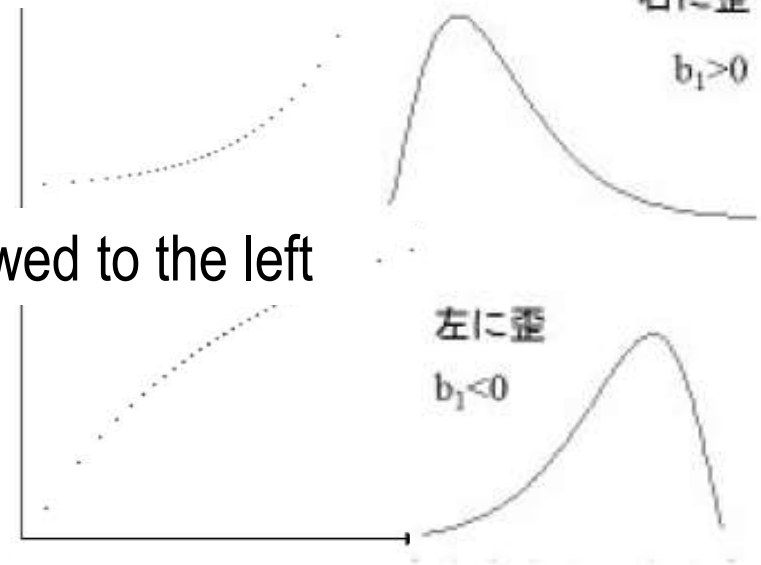
Normal probability plot

1.Normal probability plot to determine if the data are normally distributed

2.What are the deviations from normality (data skewed, tails shorter than expected, tails longer than expected)?

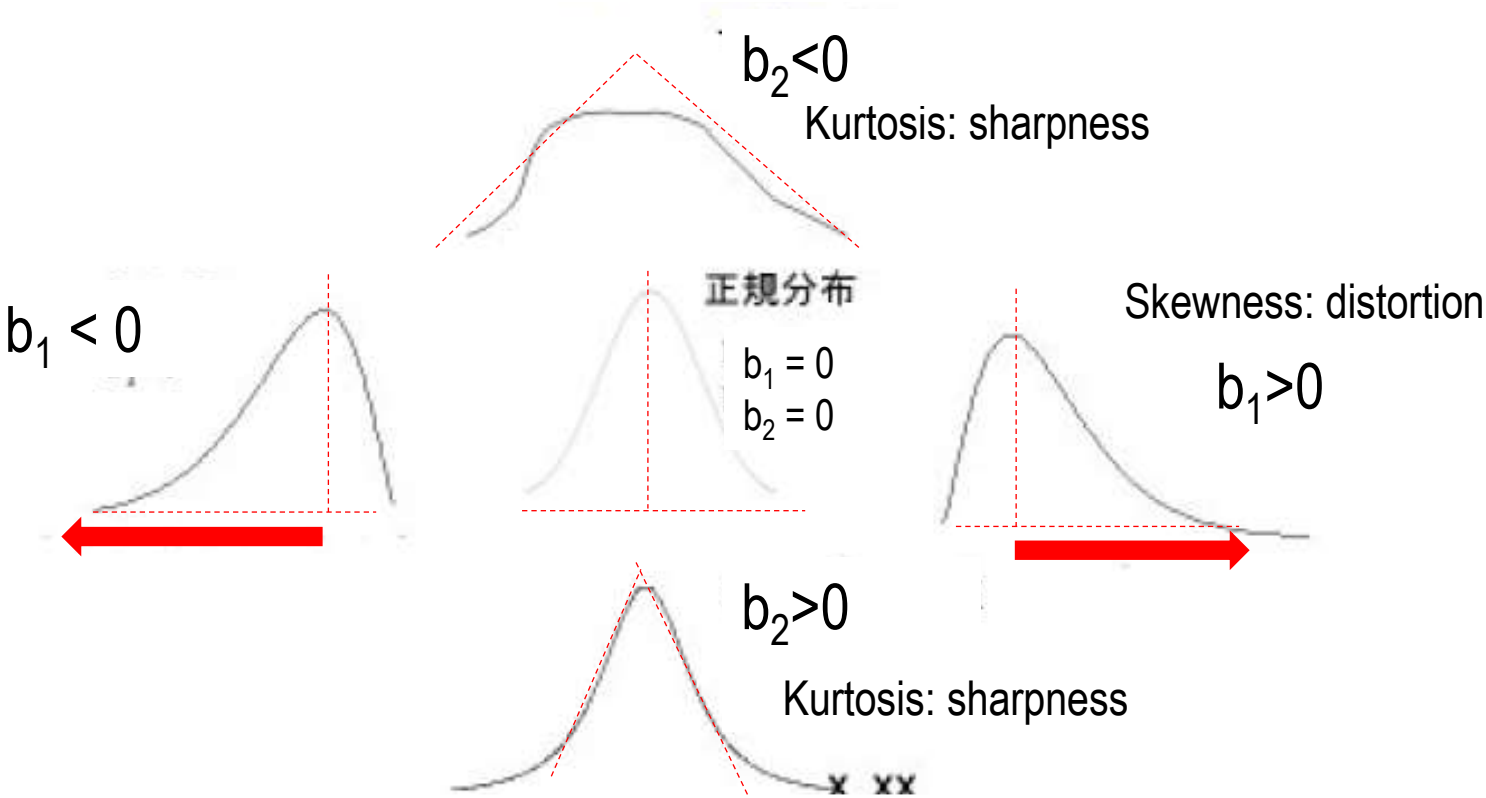


skewed to the right



skewed to the left

Shape
Skewness(b_1), Kurtosis(b_2)



Skewness: distortion

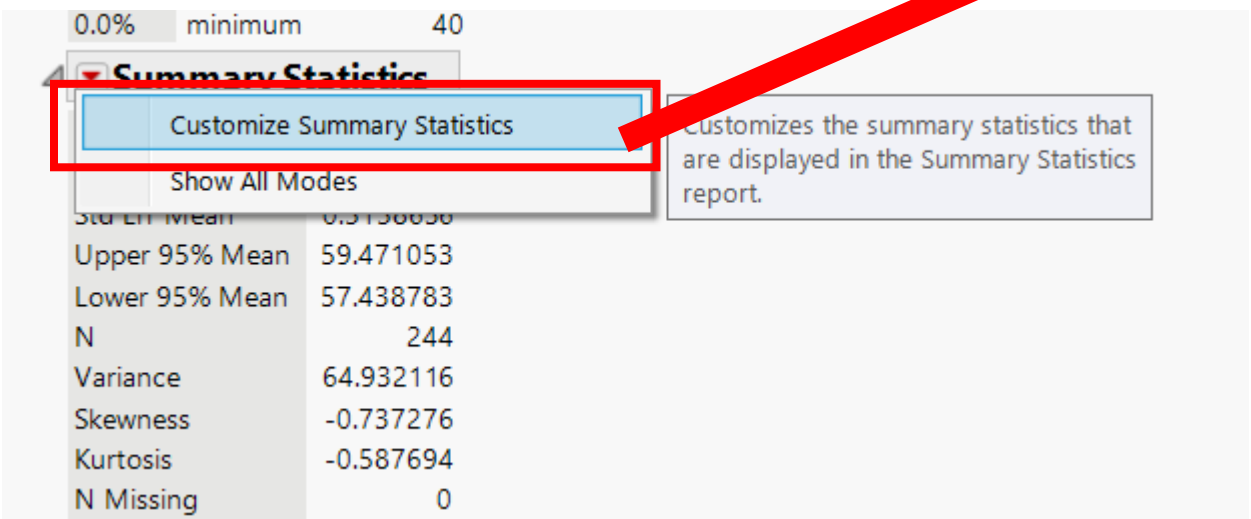
Kurtosis: sharpness

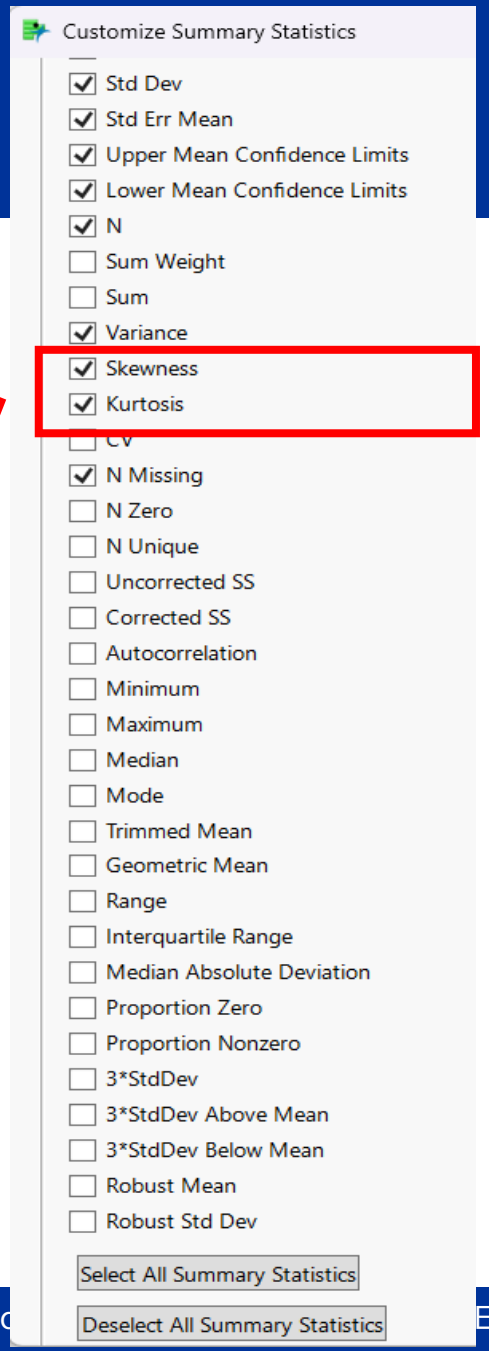
$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^3/(n-1)}{SD^3}, \quad b_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^4/(n-1)}{SD^4} - 3$$

To add Skewness and Kurtoisis displays

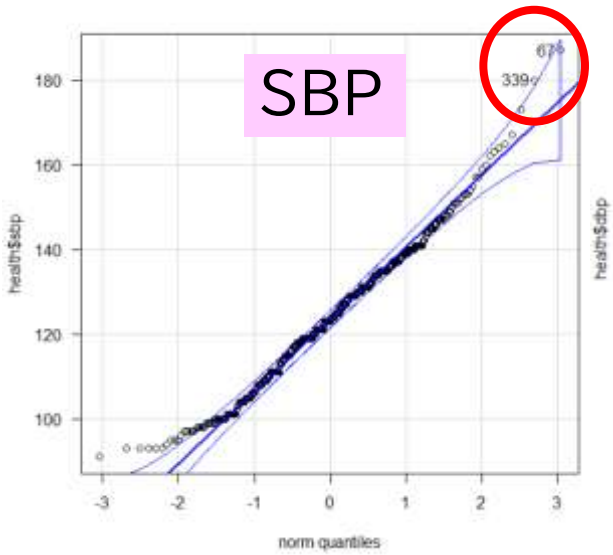
Click the ▼ button

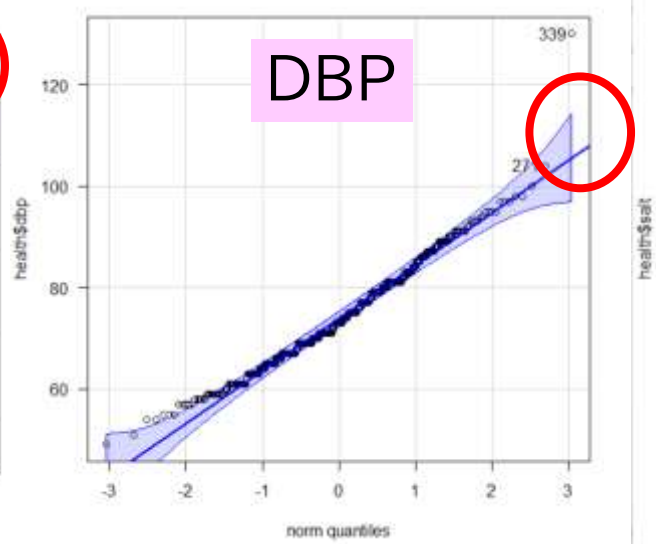
Add what you need

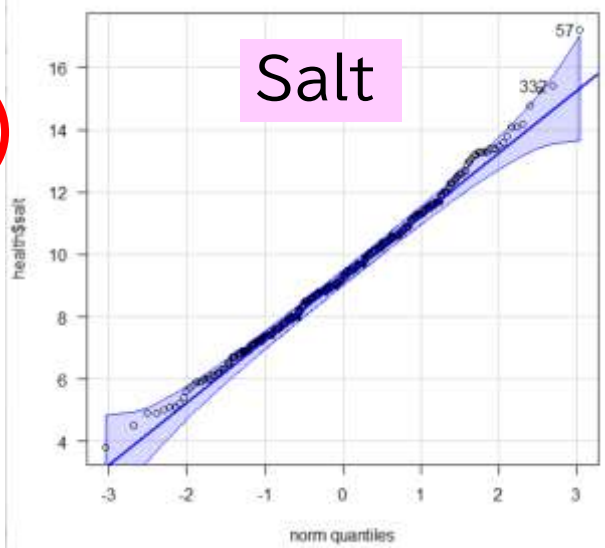




Shape
Skewness(b_1), Kurtosis(b_2)







Reference for kurtosis and skewness:
If the absolute value does not exceed 1, it is close to a normal distribution

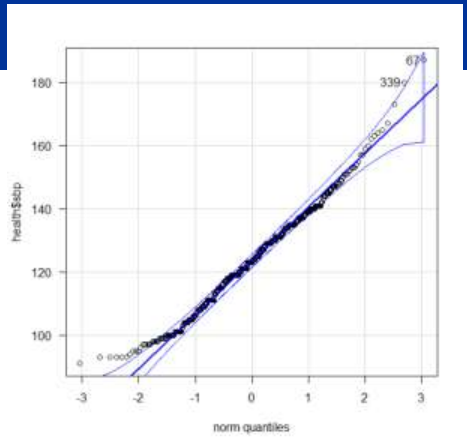
> skewness.kurtosis(health$sbp)

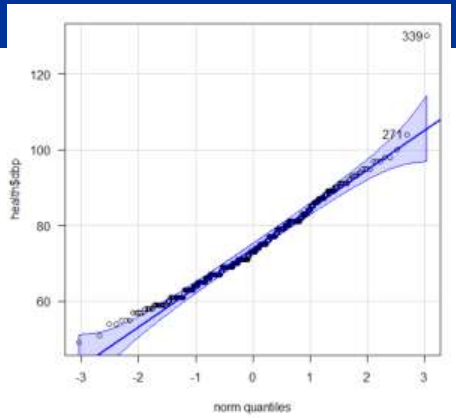
Skewness 0.361
Kurtosis 0.320

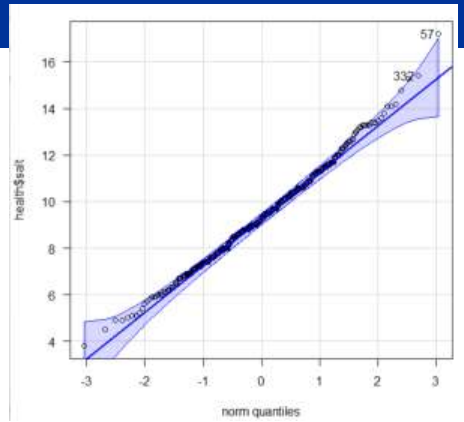
> skewness.kurtosis(health$dbp)

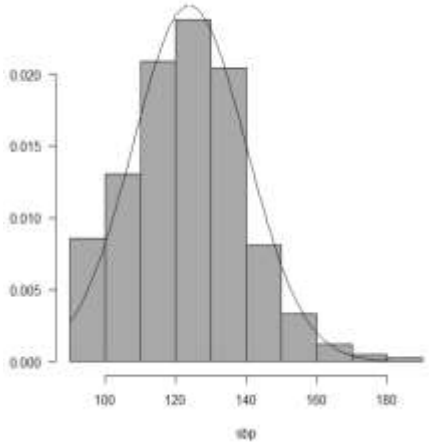
Skewness 0.625
Kurtosis 1.450

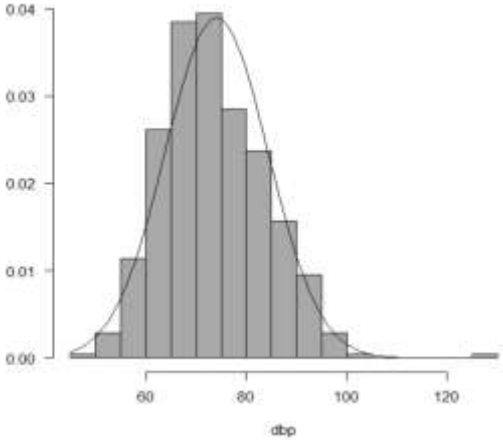
Skewness 0.30
Kurtosis 0.32

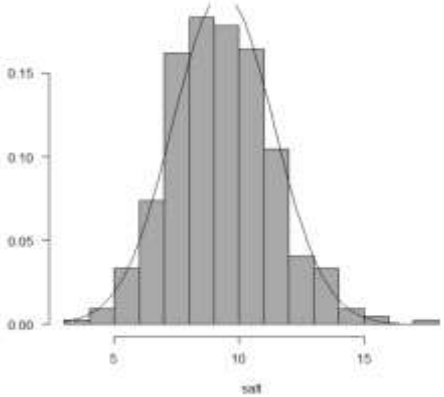












Skewness 0.361
Kurtosis 0.320

Skewness 0.625
Kurtosis 1.450

Skewness 0.30
Kurtosis 0.32

Today's overview

1. Usage of Statistical Software (package)
   1) Brief Introduction to SAS
   2) Reading Data

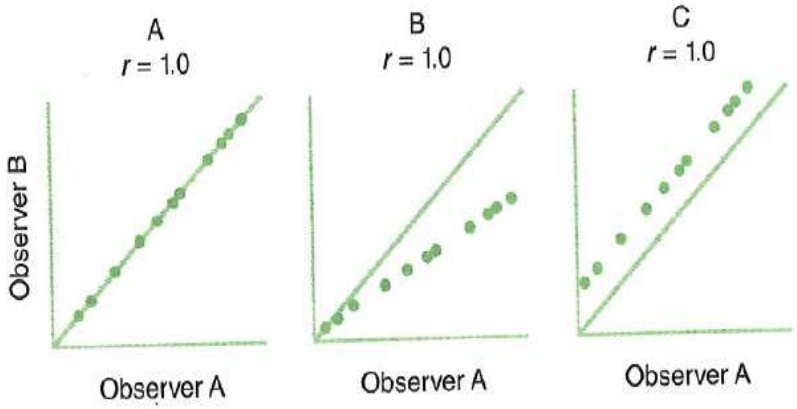
2. Description of Data
   1) Location: mean, median, percentile
   2) Dispersion: range, inter-quartile range, CV
   3) Distribution: skewness, kurtosis, Normality
   4) Association between two variables: Correlation

Correlation

- Correlation is a bivariate analysis that measures the strength of association between two variables and the linear association

→ The correlation coefficients of A, B and C are all 1.0

→ It is not necessarily a good measure of agreement.



- The value of the correlation coefficient varies between +1 and -1.
- As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker.

Pearson product-moment correlation coefficient (usually denoted by "r")

$$\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}} = Cov(x,y)/\sqrt{V(x)V(y)}$$

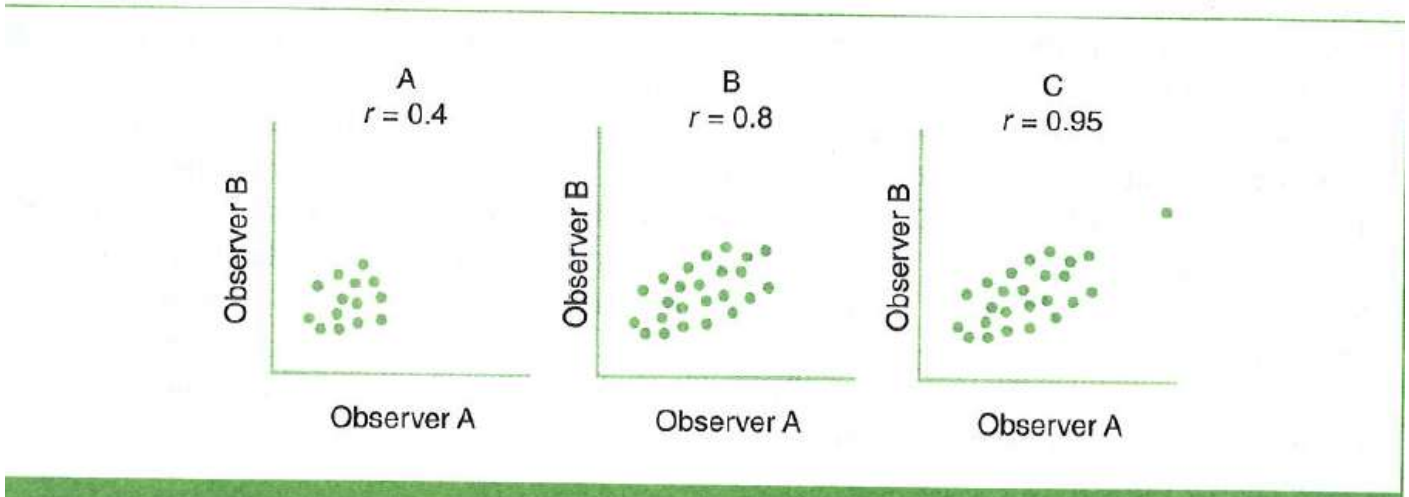


FIGURE 8-11 Hypothetical examples illustrating the high sensitivity of Pearson's correlation coefficient to the range of values and to the presence of outliers (extreme values).

Spearman correlation coefficient or rank correlation coefficient (denoted by "ρ")

- Non-parametric test that measures the strength of dependence between two variables.

$$\rho = 1 - \frac{6}{n^3 - n}\sum_{i=1}^{n}(R_i - R_i')^2$$

ρ= Spearman rank correlation
Ri,Ri'= rank
Ri-Ri'= the difference between the ranks of corresponding variables
n= number of observations

- (x,y) translate as rank order data (1 to n) and calculate Pearson correlation coefficient. If there are tie data, these are given as mean orders.

A note on correlation coefficients

- A large correlation coefficient does not indicate the magnitude of the causal relationship
  Be careful about making judgments based on tests and significance probabilities
  It only indicates that "correlation is not zero".
  (It is meaningless to compare the correlation coefficients by significance probability when n is different.)

- Does it follow a normal distribution?
  For relationships such as floor and ceiling effects, use the "rank correlation coefficient".
- Beware of outliers
- Beware of mixed distributions

『#01_NCD_Biostatistics』

ジ…1 h…2* Iris* Ir…— Ir… 2

Distributions Women

Distributions Men

2. Graph Builder

「Iris」dataset
Famous Fisher Iris data, measurements of the sepal length and width and petal length and width in centimeters.

Iris dataset

Graph
① Sepal lenhth by Species

Graph①

② Sepal length/width, Petal length/width

Graph②

3. Correlation

Health2

Correlation [Analyze]−[Fit Y by X]

Correlation①

Correlation [Analyze]−[Multivariate]−[Multivariate]

Correlation②

4. Dataset "join"
Health2 & Health3 →Health4

A: "Health2"

B: "Health3"

A&B→ NEW dataset

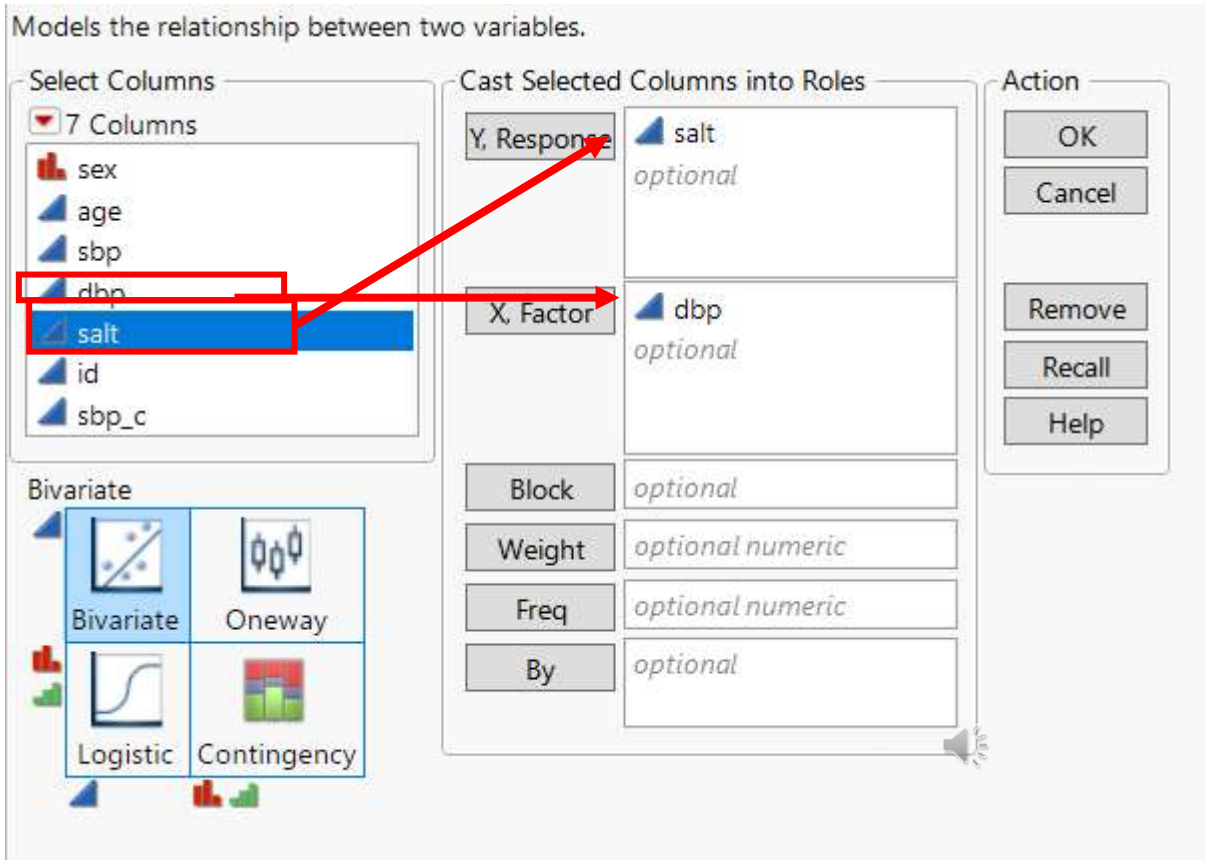
Health4

Correlation

Check the strength of linearity

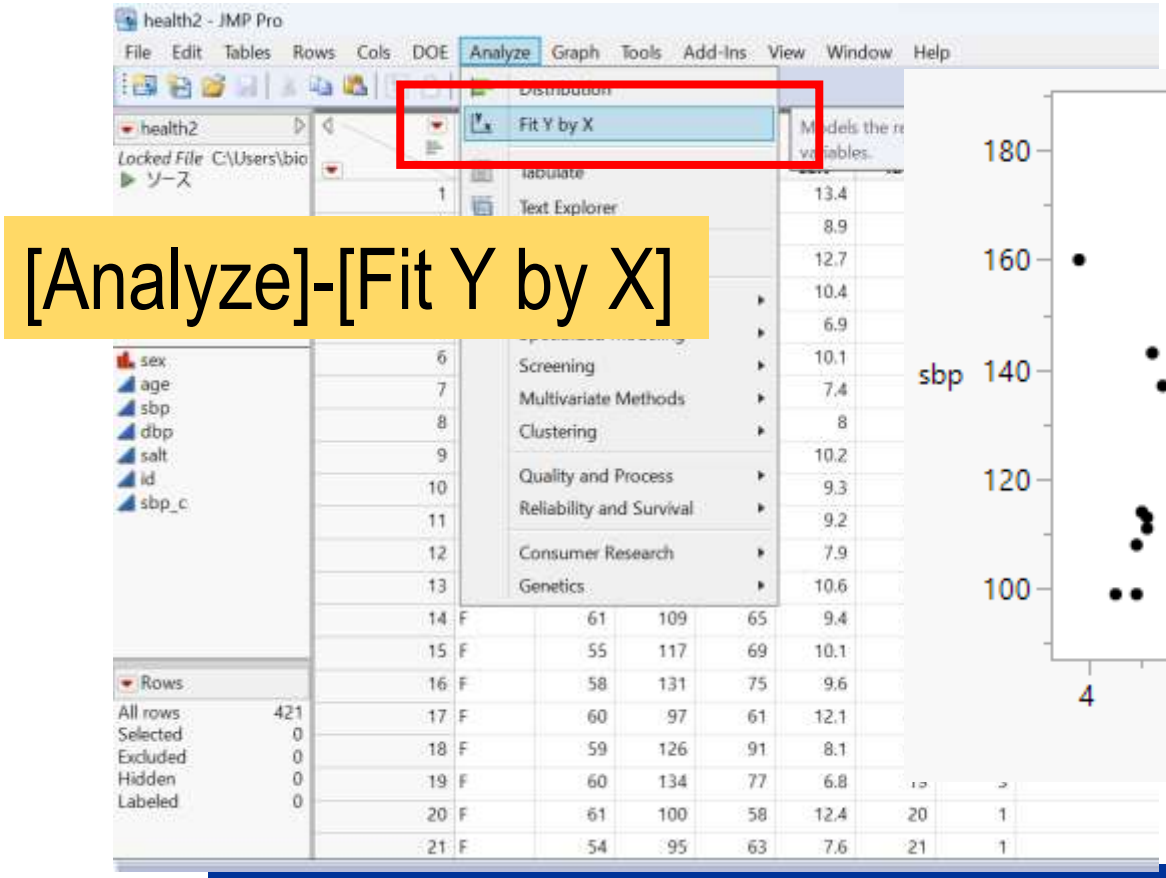
[Analyze]-[Fit Y by X]

Calculate the correlation coefficient between blood pressure and salt intake using the「health2」dataset

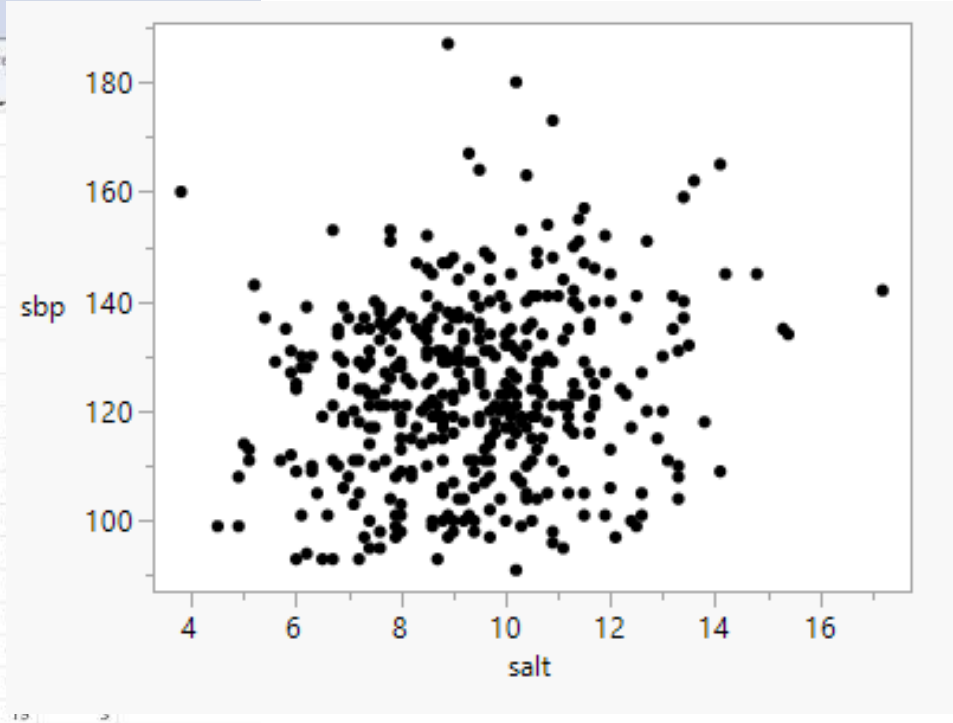


Correlation

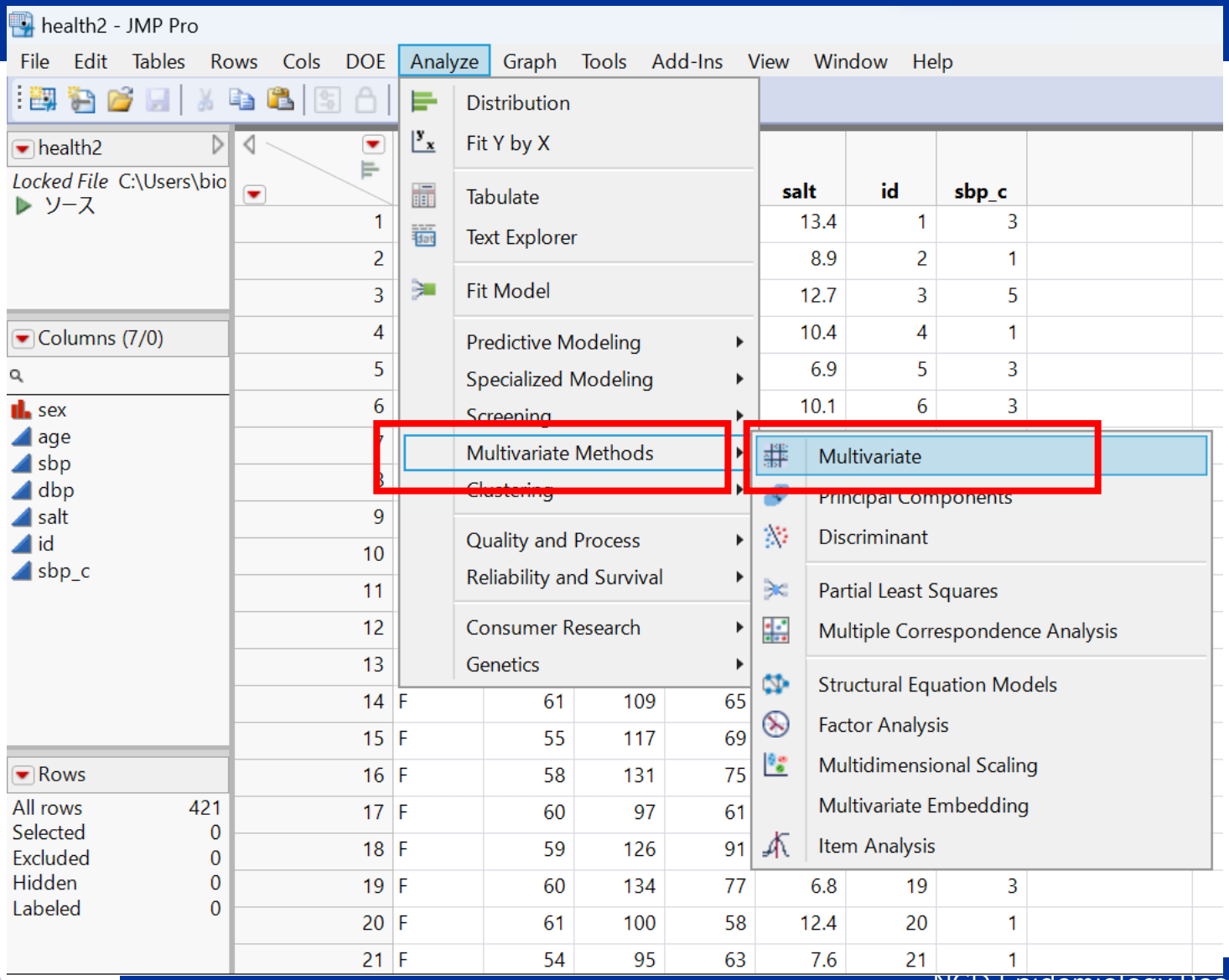
Correlation coefficient between blood pressure and salt intake using the health2 dataset



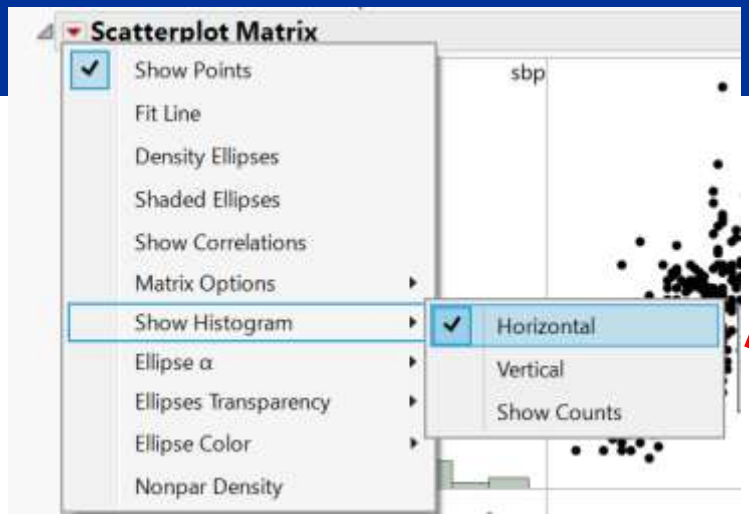
[Analyze]-[Fit Y by X]

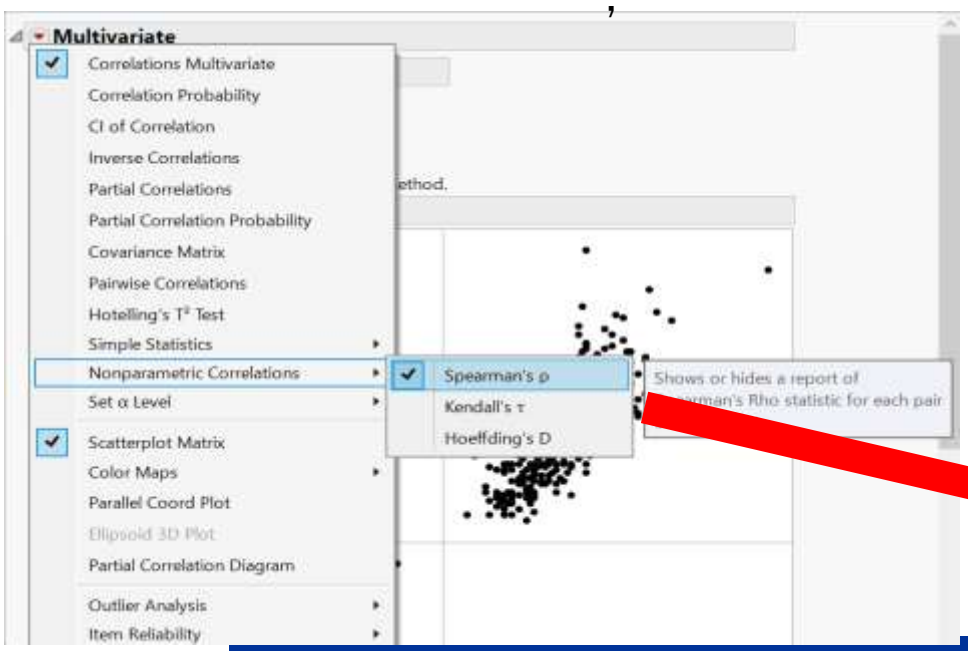


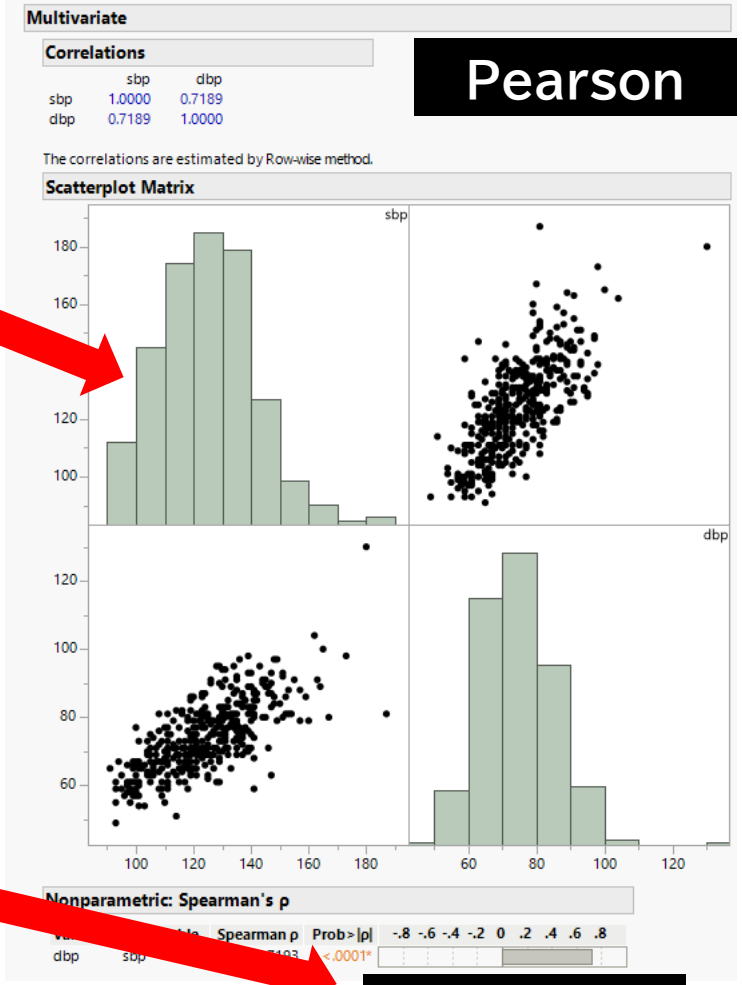
[Analyze]—[Multivariate]—[Multivariate]



Correlation Matrix（JMP）



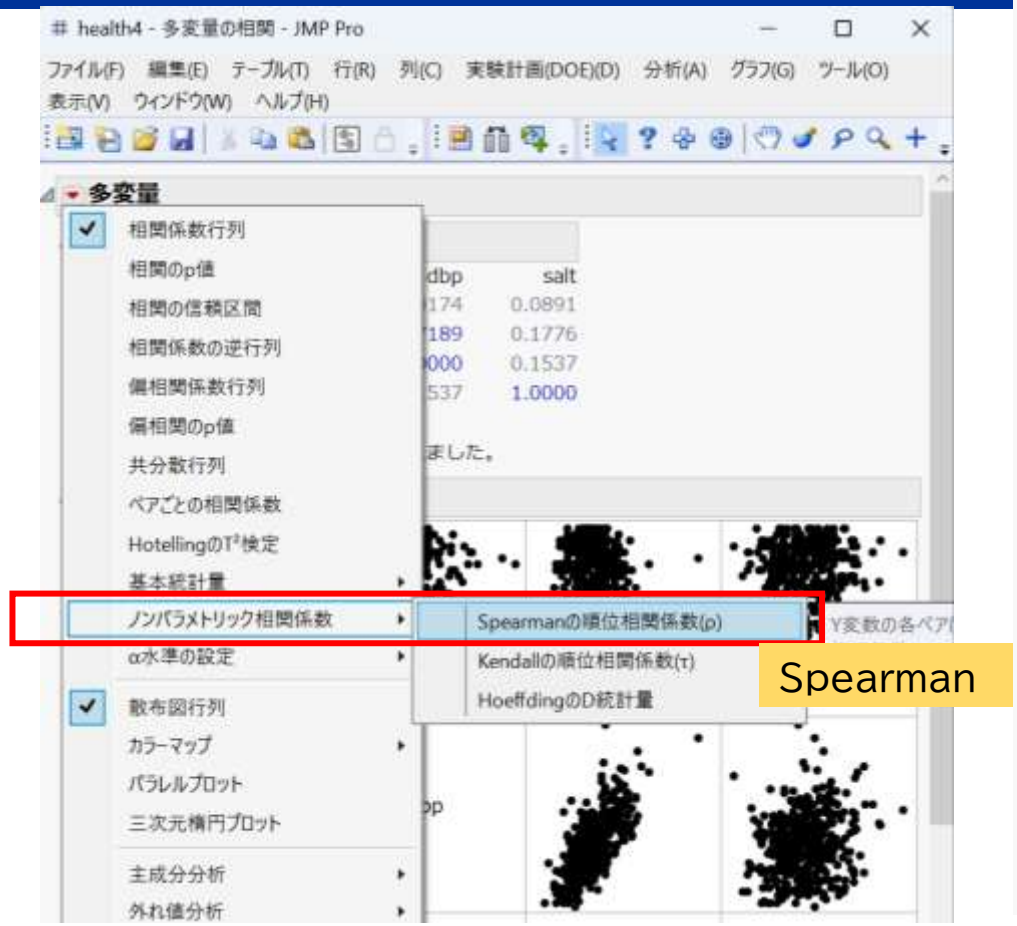


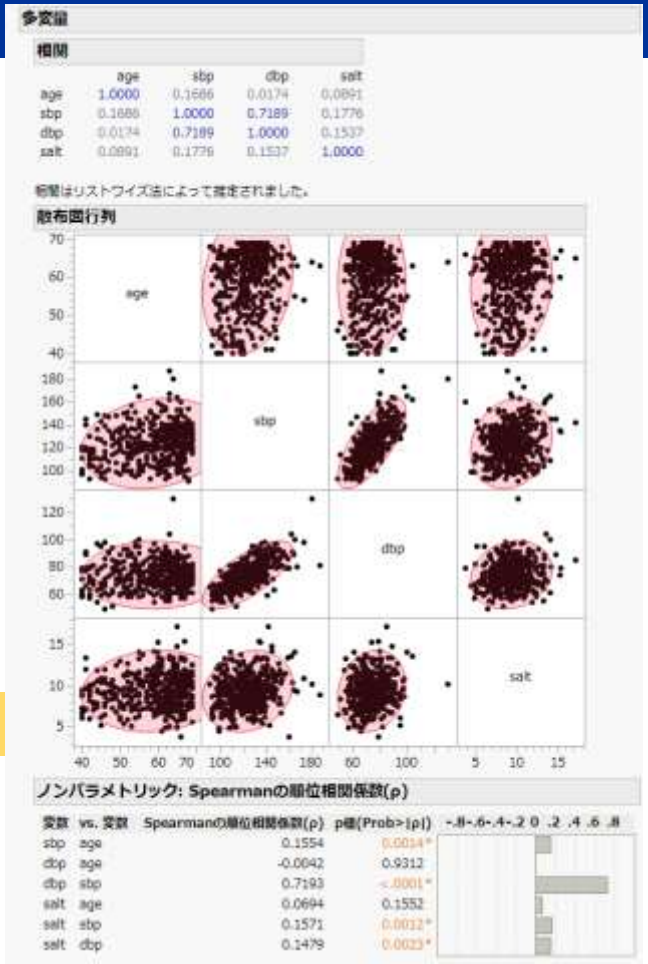


Pearson

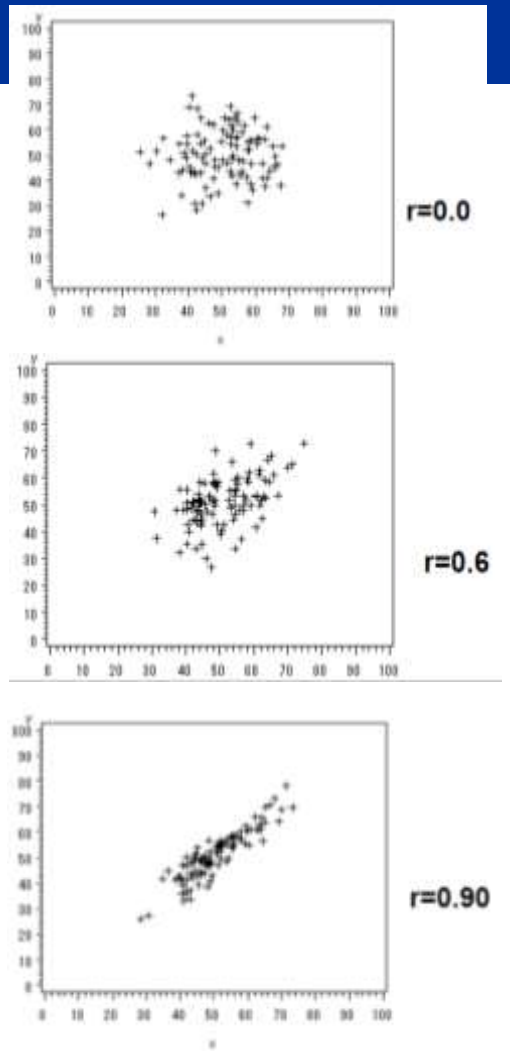
Spearman

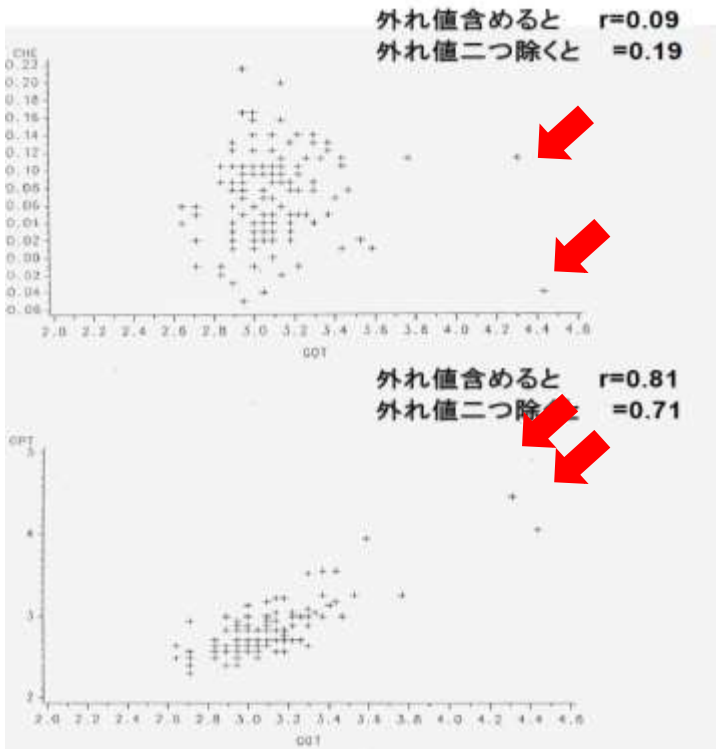
Correlation（JMP）





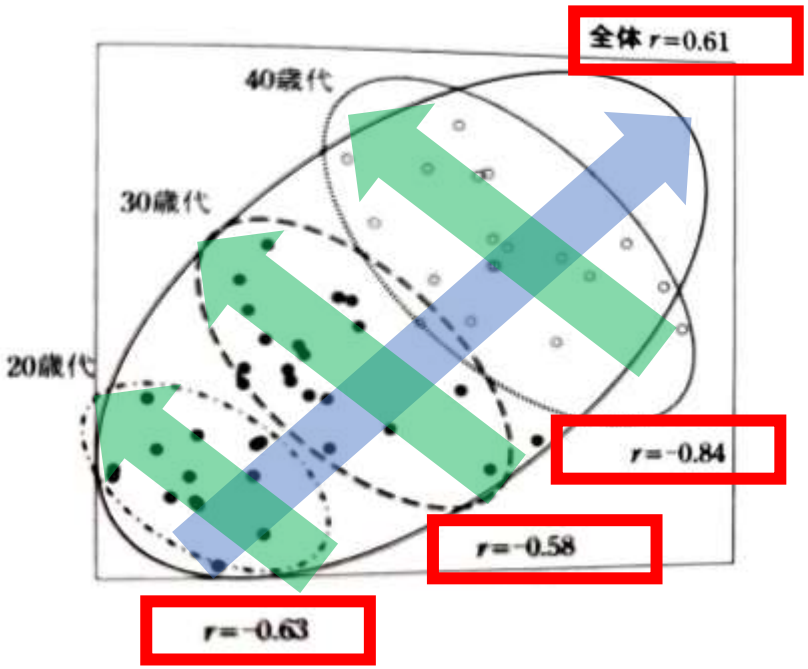
Impact of outliers





Beware of outliers. Check the distribution.

Example of mixed distribution

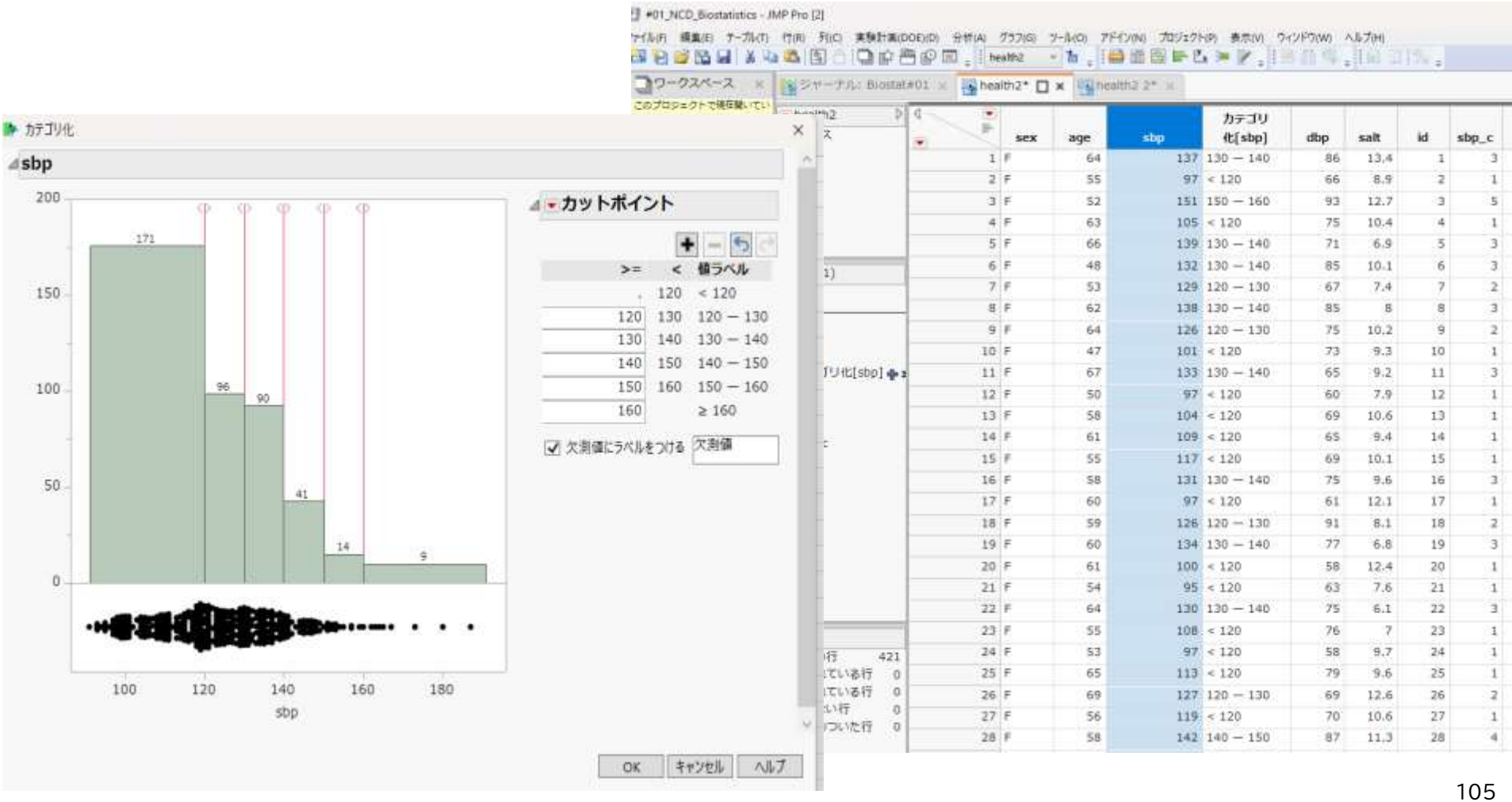


Negative correlation for each age group (Green arrow), but overall (Blue arrow) shows a positive correlation

True association is the negative association indicated by the Green arrow

JMP Tips

Categorization
Systolic blood pressure

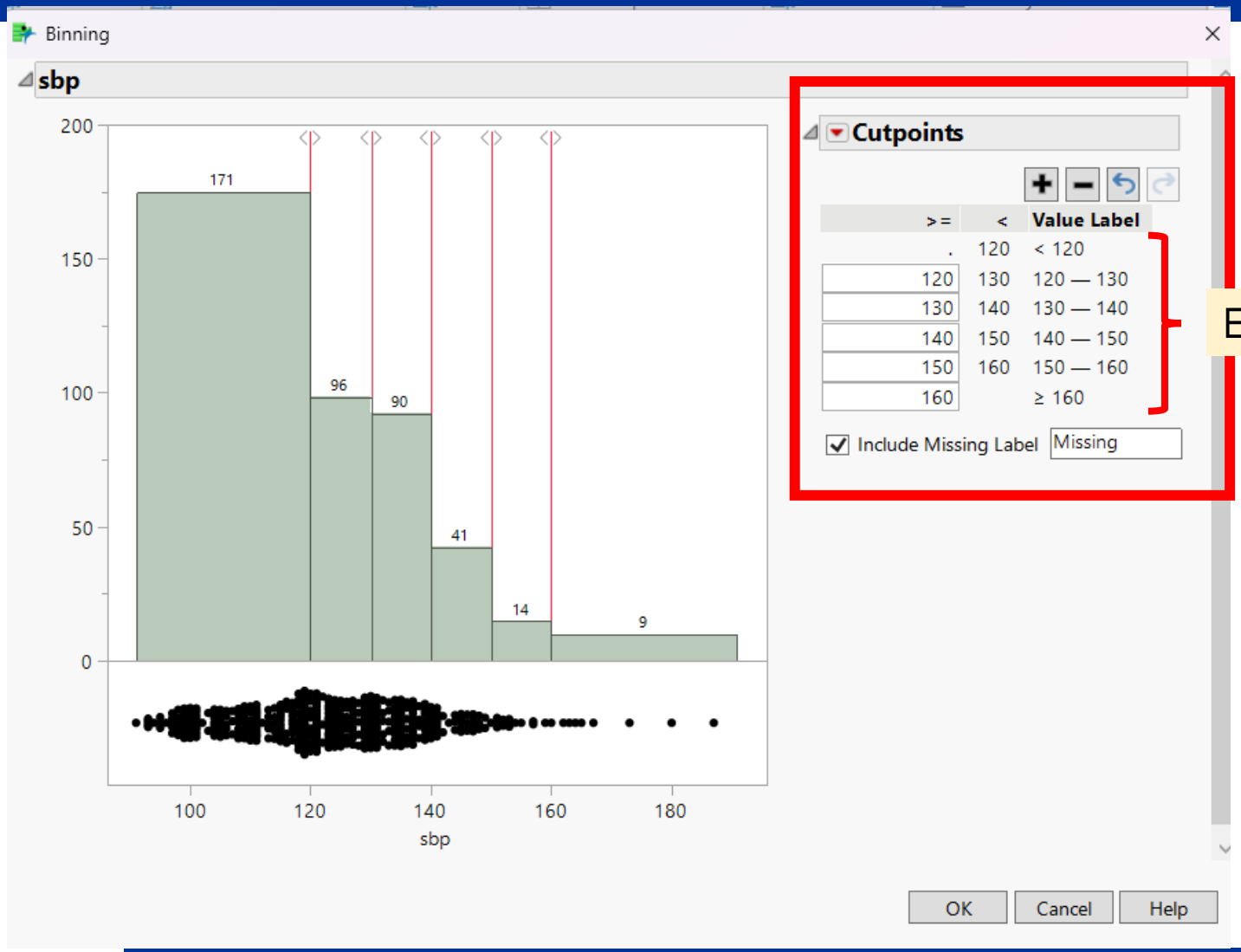


ャーナル: Biostat#01 | Iris* | Iris - Graph Builder | health2 2* | Fit Y by X - Contextual | health2 3*

Σ	sex	age	sb			
1	F	64				
2	F	55				
3	F	52				
4	F	63				
5	F	66				
6	F	48				
7	F	53				
8	F	62				
9	F	64				
10	F	47				
11	F	67				
12	F	50				
13	F	58				
14	F	61				
15	F	55				
16	F	58				
17	F	60				
18	F	59				
19	F	60				
20	F	61				
21	F	54				
22	F	64	130	75	6.1	22
23	F	55	108	76	7	23
24	F	53	97	58	9.7	24
25	F	65	113	79	9.6	25
26	F	69	127	69	12.6	26
27	F	56	119	70	10.6	27

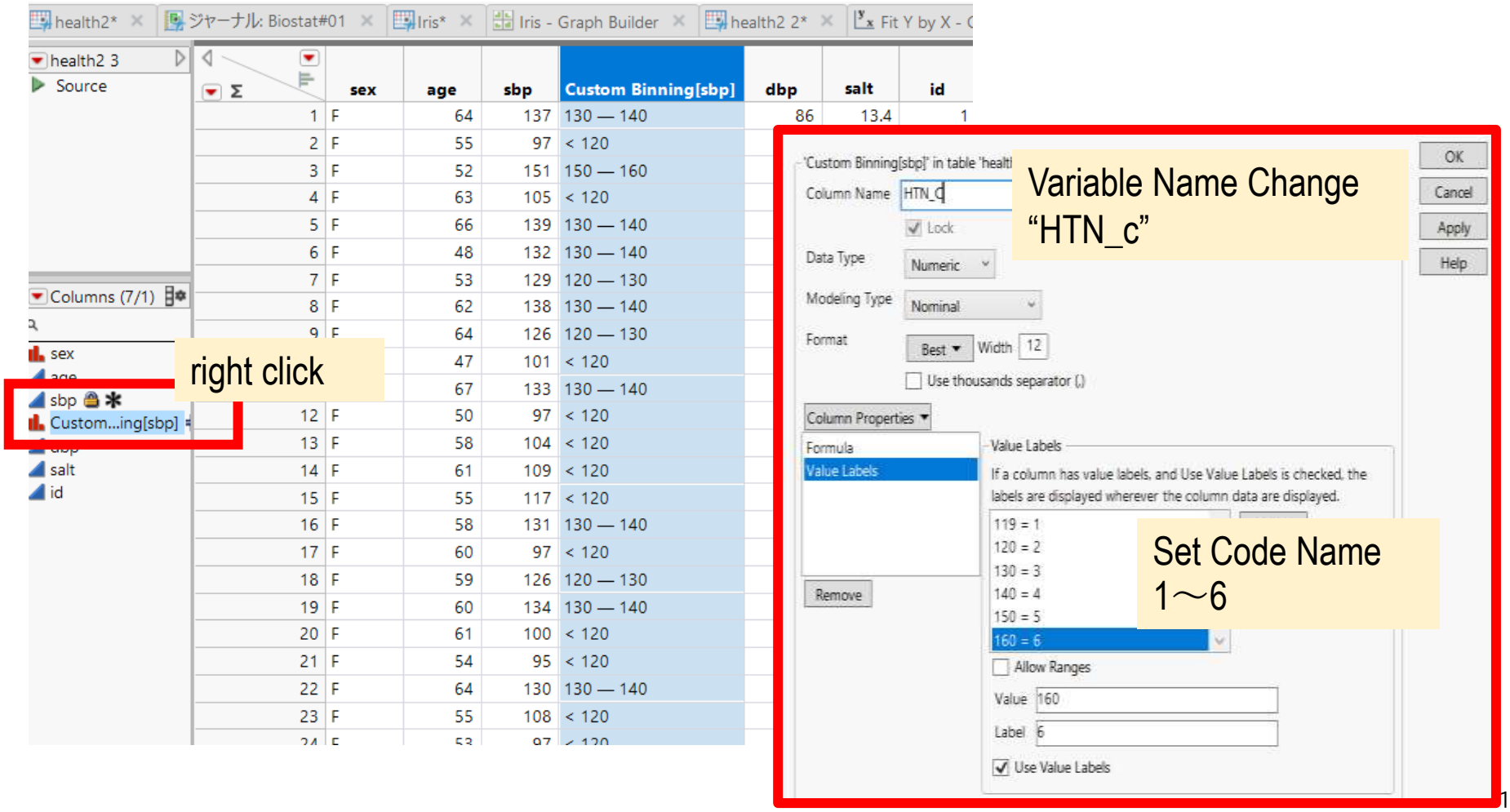
- Filter
- Column Info...
- Standardize Attributes...
- Column Properties
- Formula...
- Recode...
- New Formula Column
  - Transform
  - Log
  - Specialty
  - Aggregate
  - Distributional
    - Center
    - Custom Binning...
    - Standardize
    - Range 0 to 1
    - Box Cox...
    - Johnson Normalizing
    - Informative Missing
    - Rank
    - Rank (reverse order)
    - Cumulative Probability
    - Normal Quantile
    - SHASH...
  - Random
  - Row
  - Formula...
  - Group By
- Insert Columns
- Delete Columns
- Label/Unlabel
- Link ID
- Sort
- Header Background Color
- Header Text Color
- Copy Column Properties
- Copy Columns

Custom binning

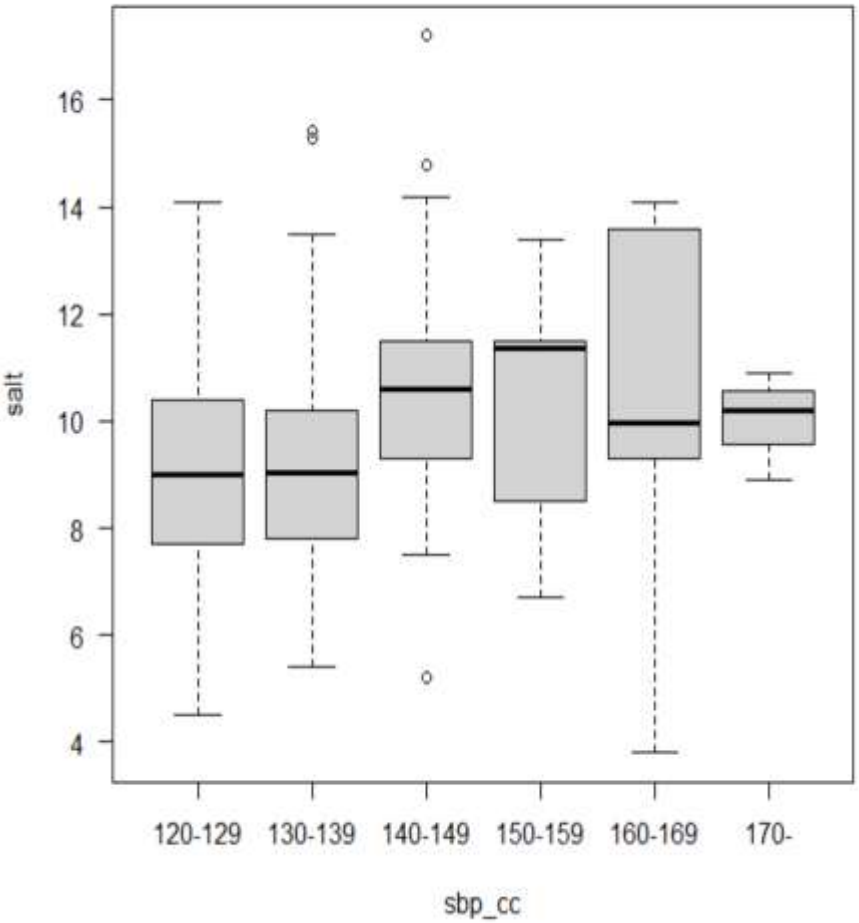


Enter cutoff value

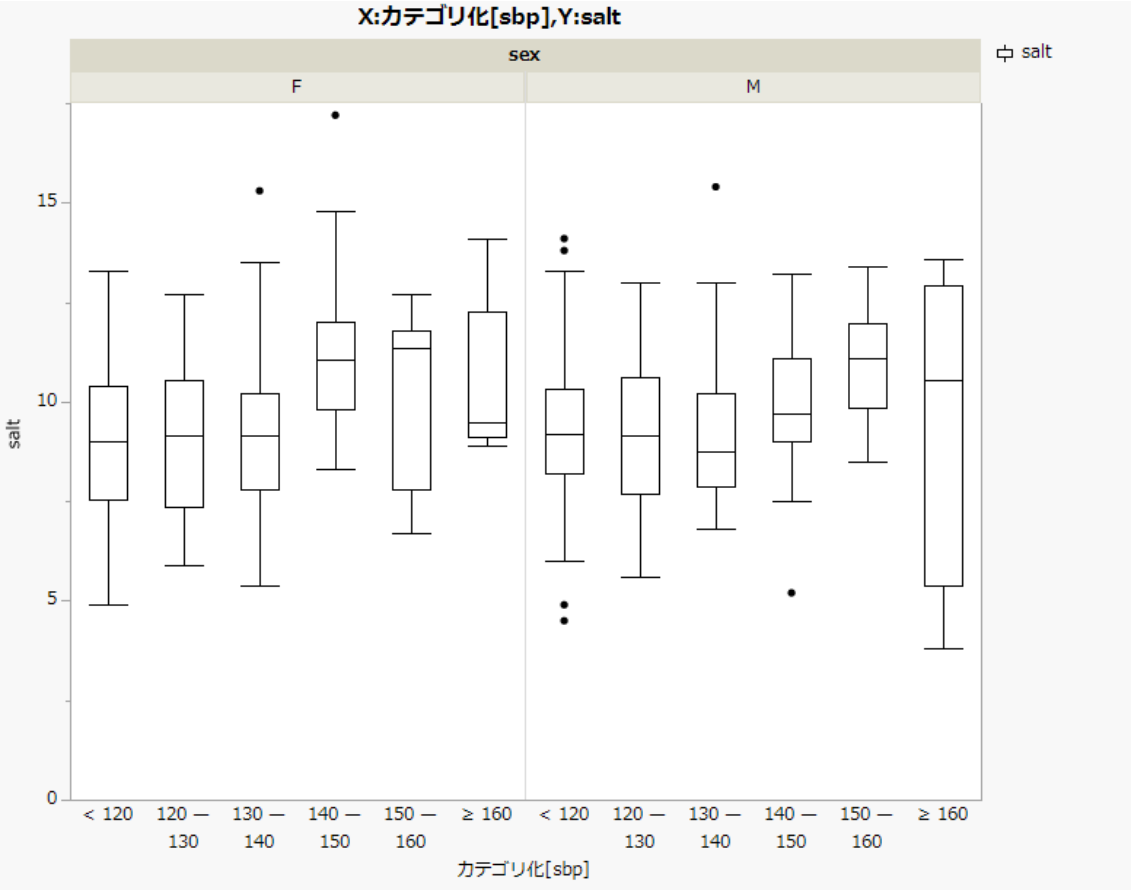
Name, Code change

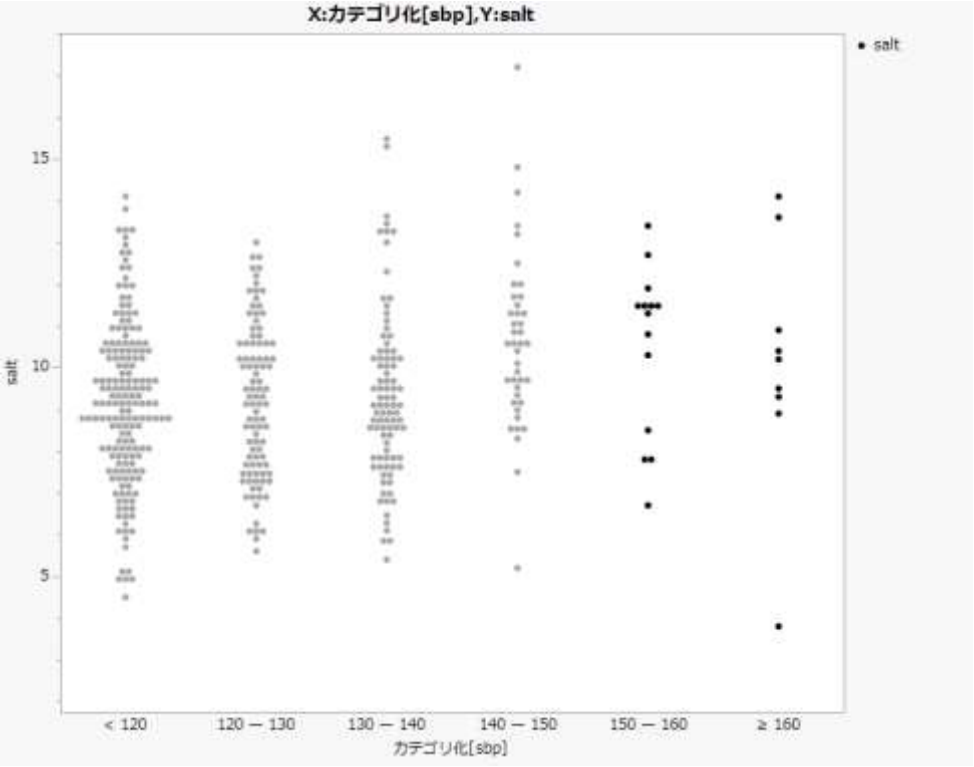


Box plot: SBP-cc × Salt intake using graph builder



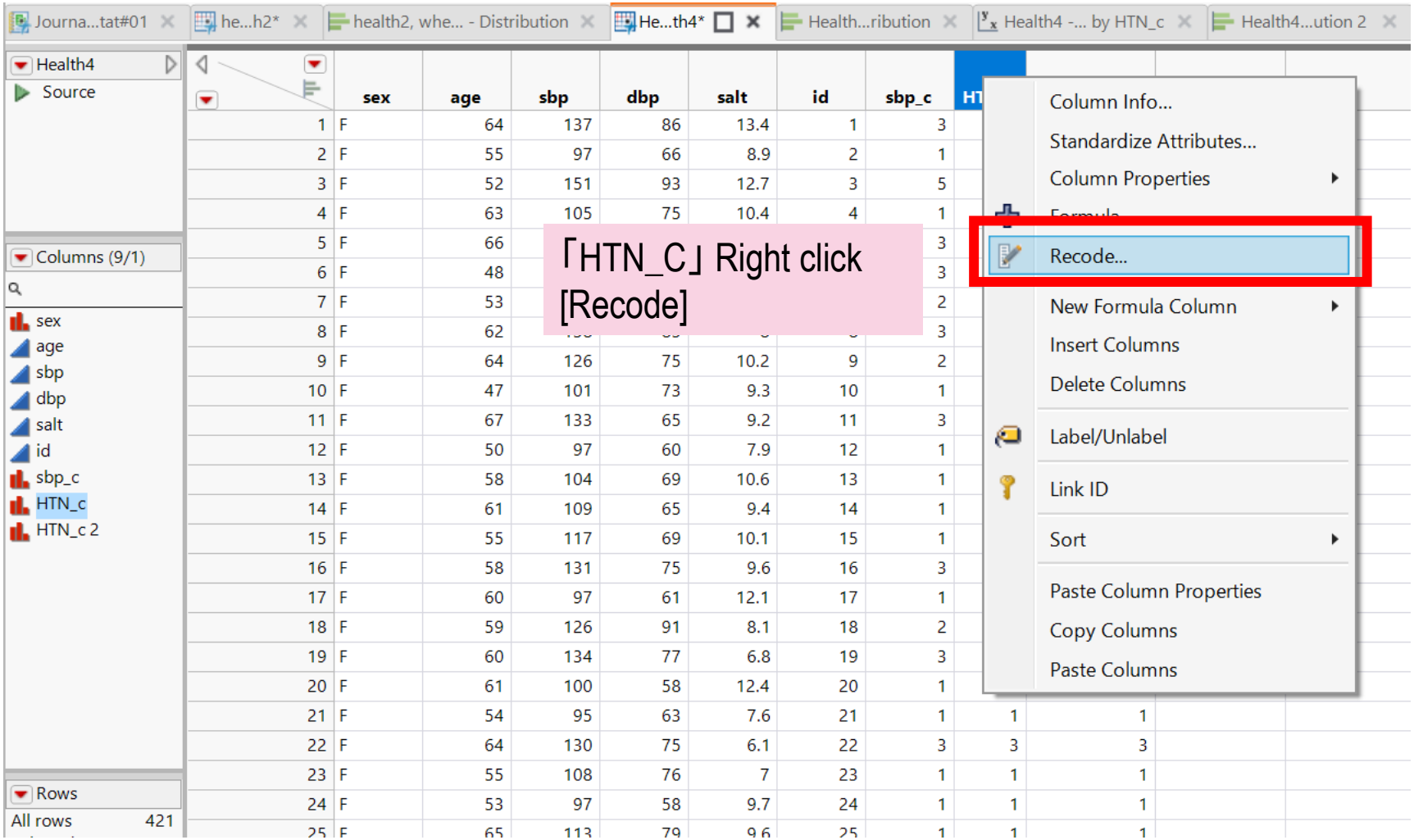
Box plot: SBP-cc × Salt intake （by sex）





- HTN_c's 5 and 6 want to be merged into 4 because of the small number of people

“Re-code”
HTN_c's 5 and 6 want to be merged into 4 because of the small number of people



Re-coding

Journ...at#01 | he...h2* | health2, w...istribution | He...h4* | Healt...

New Column Name: HTN_c_new

Count	Old Values (6)	New Values (4)
161	1	1
74	2	2
151	3	3
31	4	4
3	5 *	
1	6 *	

Bind codes 4-6 and set as 4

Code integration (grouping) and label change → “Recode”.

New Column | Name: HTN_C 2

Count	Old Values (6)	Old Labels	New Values (6)	New Labels
171	119	1	119	1
96	120	2	120	2
90	130	3	130	3
41	140	4	140	4
14	150	5	150	4
9	160	6	160	4

Filter

Group controls
- [x] View Groups
- [ ] Show Only Grouped
- [ ] Show Only Ungrouped

Group

- (•) All
- ( ) Only Modified
- ( ) Only Unmodified

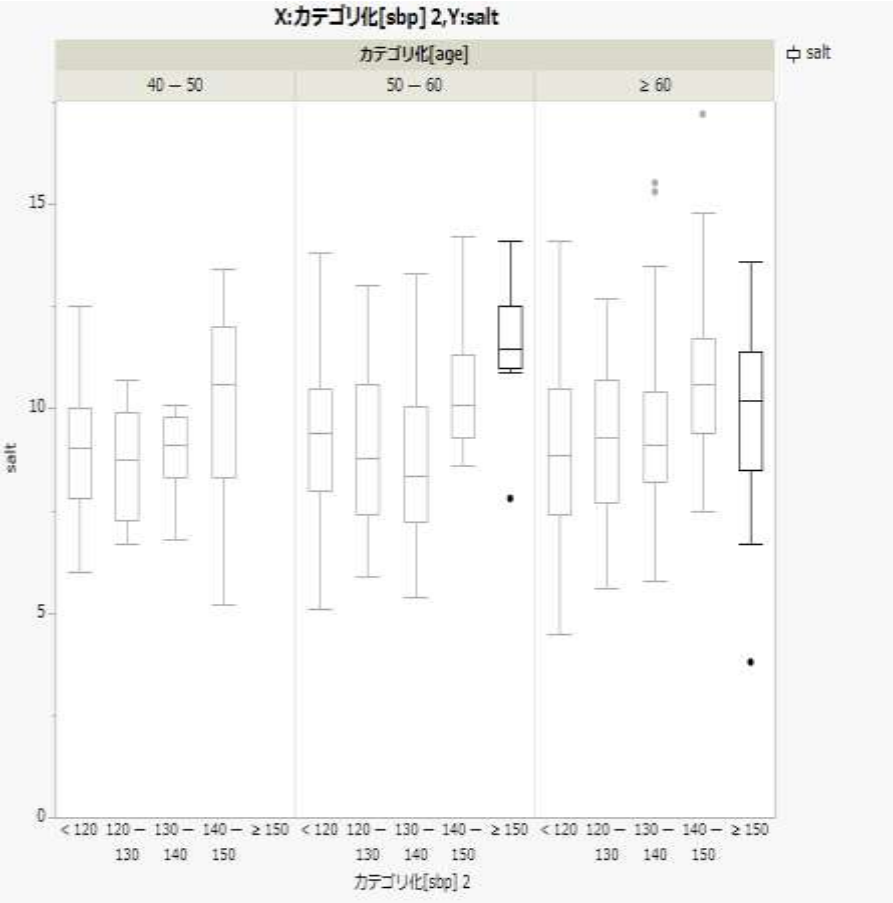
Changes

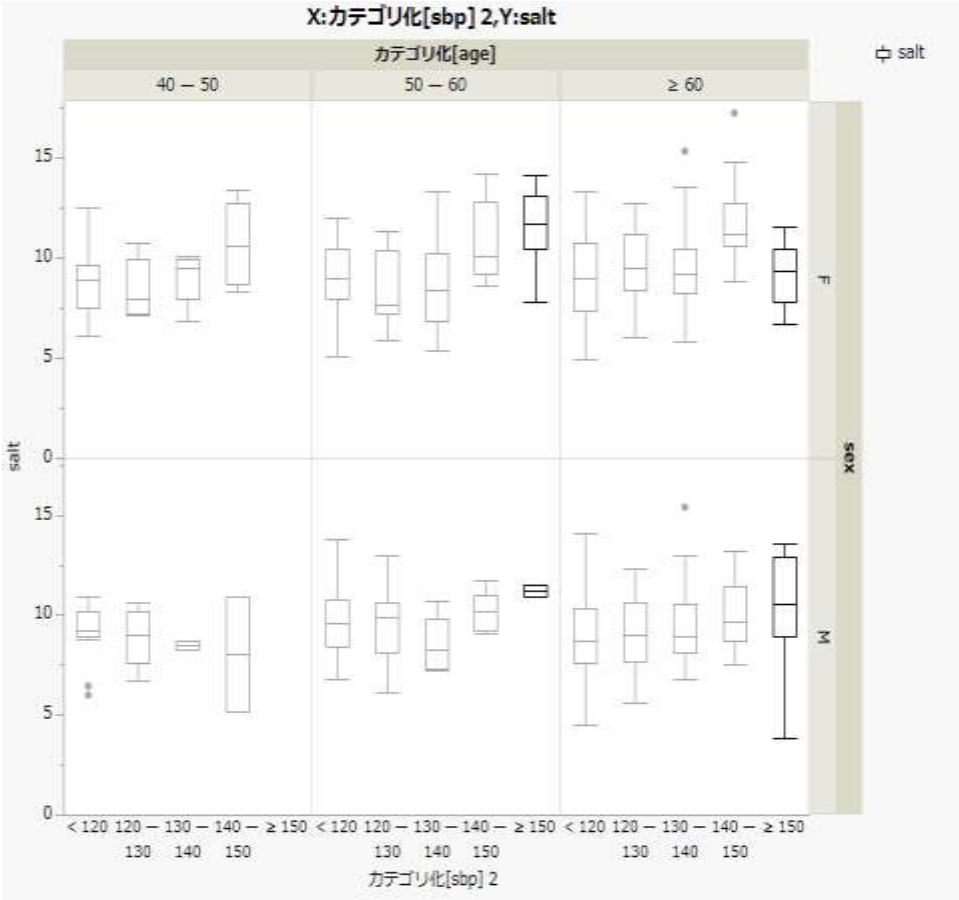
Change Label: "6" -> "4"
Change Label: "5" -> "4"

65	F	68	≥ 60	144	140 — 150	140 — 150
66	F	40	40 — 50	118	< 120	< 120
67	F	63	≥ 60	187	≥ 160	≥ 150
68	F	66	≥ 60	141	140 — 150	140 — 150
69	F	67	≥ 60	119	< 120	< 120
70	F	65	≥ 60	109	< 120	< 120
71	F	62	≥ 60	134	130 — 140	130 — 140

Scripting
- [x] Script sequence of actions
  - [x] Compress sequence

Box plot: SBP-cc × Salt intake (by age) using graph builder





Wide dataset → Long data set
Dataset Direction Conversion(Transpose)

- [Table]-[Transpose]

Data per measurement point expands horizontally

Time point

Blood Pressure - JMP Pro [6]

ファイル(F) 編集(E) テーブル(T) 行(R) 列(C) 実験計画(I) ル(O) アドイン(N) 表示(V) ウィンドウ(W) ヘルプ(H)

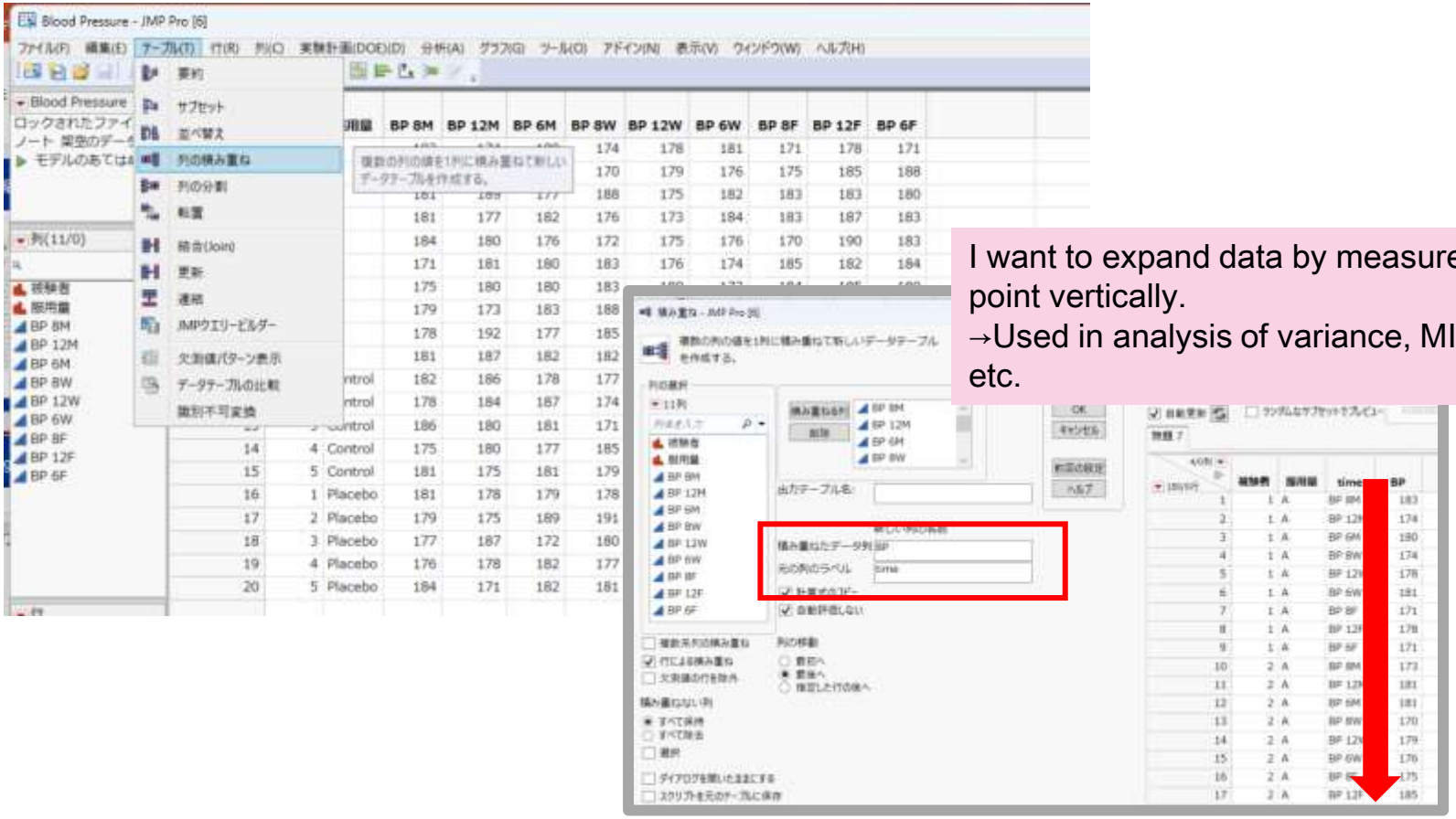
Blood Pressure
ロックされたファイル C
ノート 架空のデータ。複
モデルのあてはめ

列(11/0)
被験者
服用量
BP 8M
BP 12M
BP 6M
BP 8W
BP 12W
BP 6W
BP 8F
BP 12F
BP 6F

	被験者	服用量	BP 8M	BP 12M	BP 6M	BP 8W	BP 12W	BP 6W	BP 8F	BP 12F	BP 6F
1	1	A	183	174	180	174	178	181	171	178	171
2	2	A	173	181	181	170	179	176	175	185	188
3	3	A	181	189	177	188	175	182	183	183	180
4	4	A	181	177	182	176	173	184	183	187	183
5	5	A	184	180	176	172	175	176	170	190	183
6	1	B	171	181	180	183	176	174	185	182	184
7	2	B	175	180	180	183	180	173	184	185	180
8	3	B	179	173	183	188	193	170	179	185	170
9	4	B	178	192	177	185	172	179	187	184	177
10	5	B	181	187	182	182	177	177	184	183	184
11	1	Control	182	186	178	177	177	177	183	191	175
12	2	Control	178	184	187	174	175	176	180	181	184
13	3	Control	186	180	181	171	173	181	179	181	189
14	4	Control	175	180	177	185	171	182	176	185	170
15	5	Control	181	175	181	179	185	184	184	173	185
16	1	Placebo	181	178	179	178	182	179	188	183	173
17	2	Placebo	179	175	189	191	178	177	182	188	171
18	3	Placebo	177	187	172	180	183	178	183	190	178
19	4	Placebo	176	178	182	177	178	177	176	184	184
20	5	Placebo	184	171	182	181	186	183	184	172	179

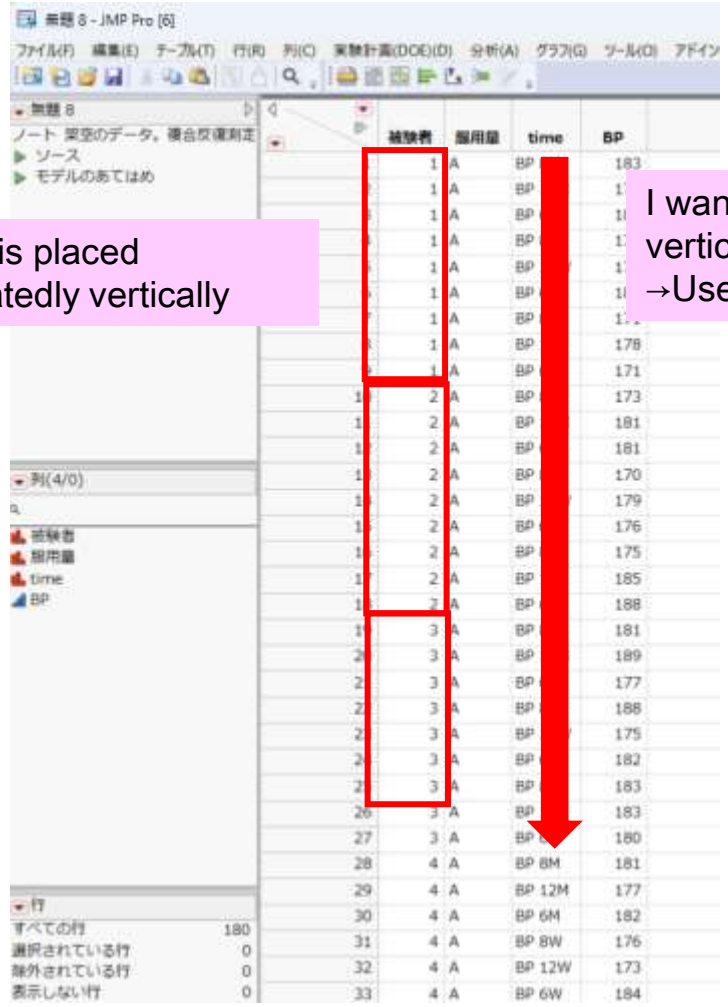
Wide dataset → Long data set
Dataset Direction Conversion(Transpose)

[Table]-[Transpose]



I want to expand data by measurement point vertically.
→Used in analysis of variance, MIXED, etc.

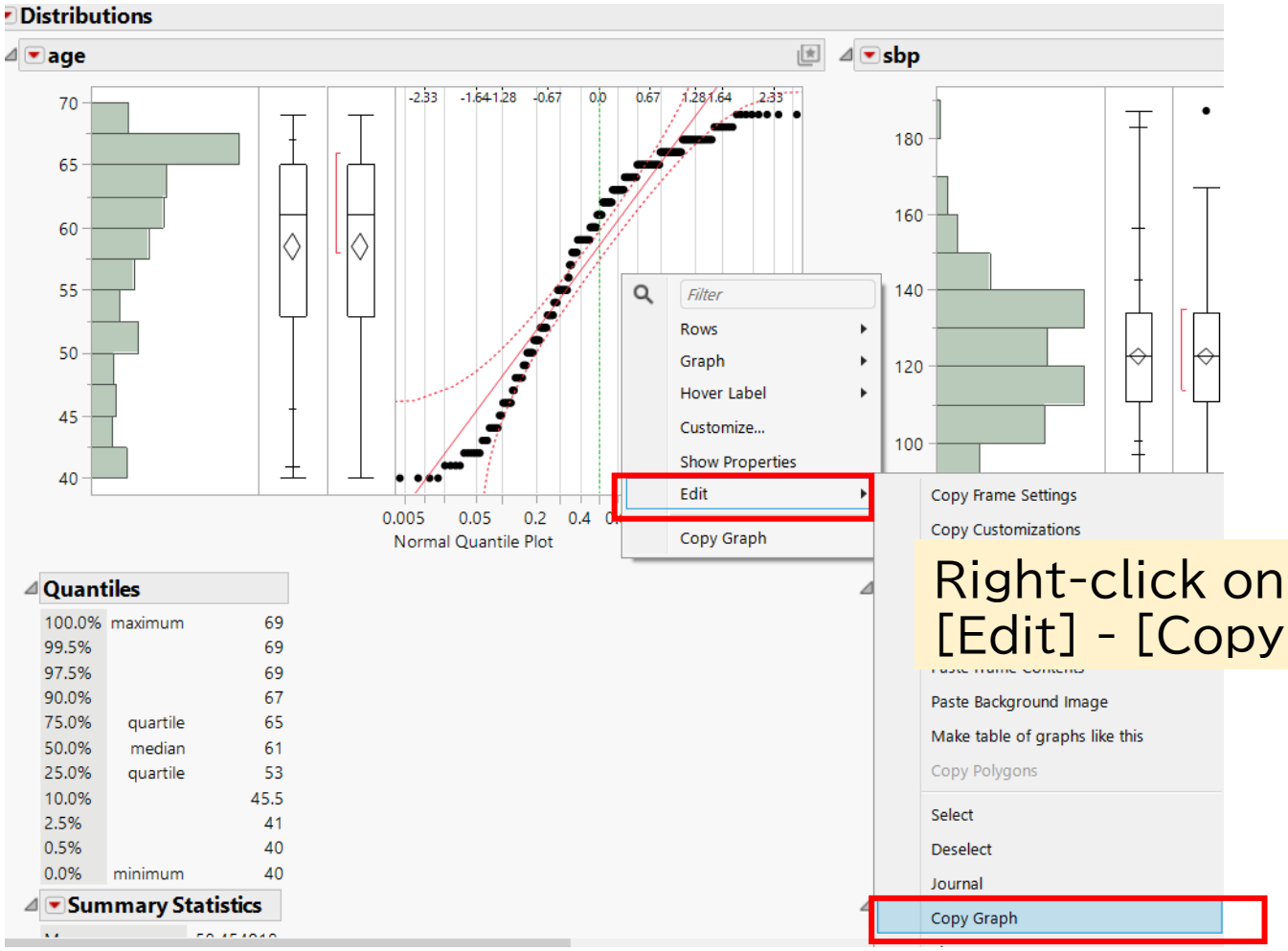
Wide dataset → Long data set
Dataset Direction Conversion(Transpose)



Obs. is placed repeatedly vertically

I want to expand data by measurement point vertically.
→Used in analysis of variance, MIXED, etc.

Copy Figure and tables



Right-click on a graph, [Edit] - [Copy Graph].

Subset the data

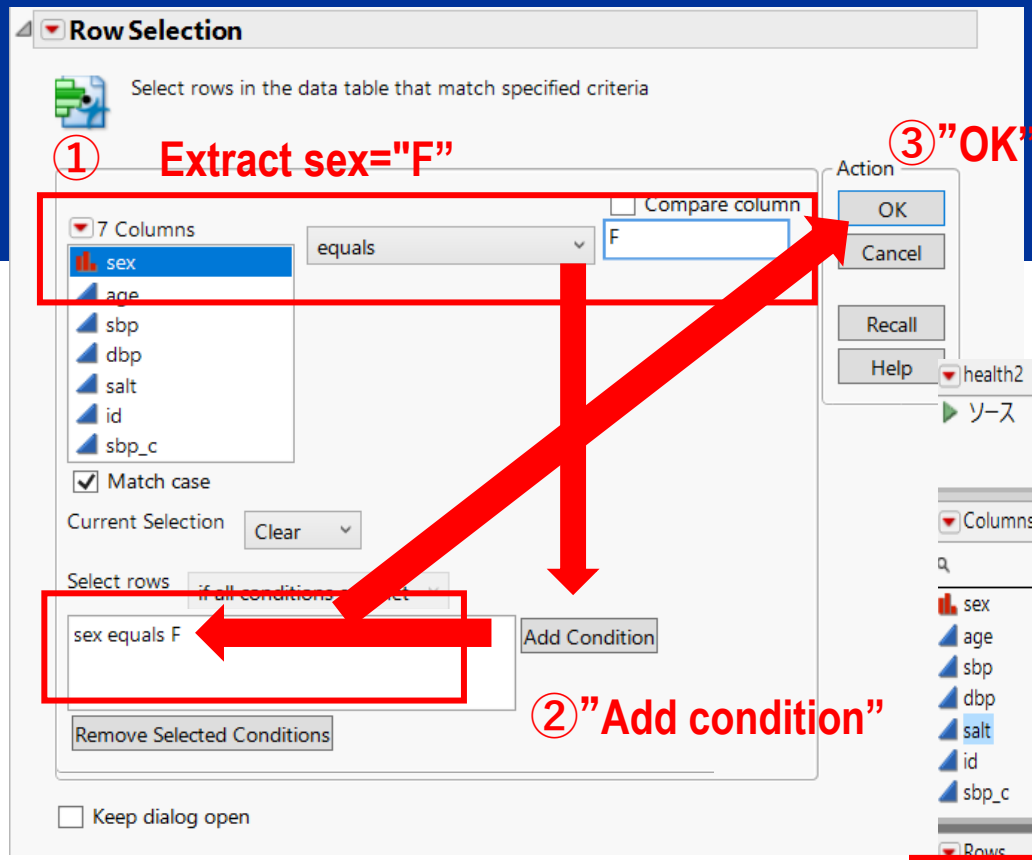
#01_NCD Biostatistics - JMP Pro [3]

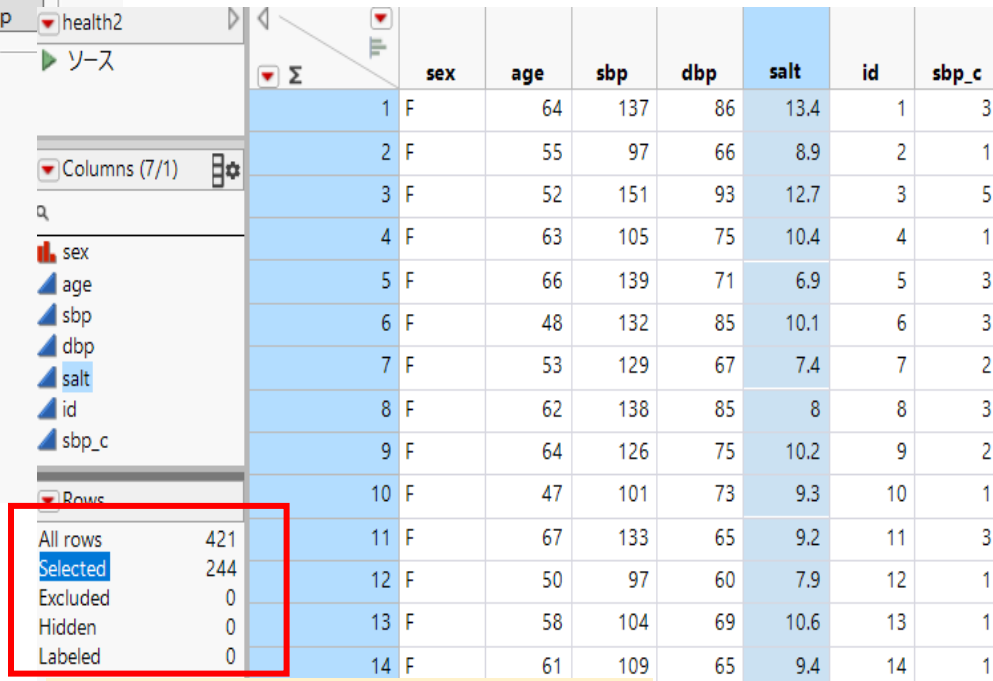
File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins Project View Window Help

Hide and Exclude Ctrl+Shift+E
Exclude/Unexclude Ctrl+E
Hide/Unhide
Label/Unlabel
Colors
Markers
Next Selected F7
Previous Selected F6
Row Selection
Clear Row States
Clear Selected Row States
Color or Mark by Column...
Row Editor
Delete Rows
Add Rows...
Move Rows...
Data Filter

Go to Row...
Invert Row Selection
Select All Rows
Select Excluded
Select Hidden
Select Labeled
Select Where... Ctrl+Shift+W
Select Matching Cells
Select All Matching Cells
Select Randomly...
Select Dominant...
Select Duplicate Rows
Name Selection in Column...

sex	age	sbp	dbp	salt	id	sbp_c
F	64	137	86	13.4	1	3
F	55	97	66	8.9	2	1
F	52	151	93	12.7	3	5





	sex	age	sbp	dbp	salt	id	sbp_c
1	F	64	137	86	13.4	1	3
2	F	55	97	66	8.9	2	1
3	F	52	151	93	12.7	3	5
4	F	63	105	75	10.4	4	1
5	F	66	139	71	6.9	5	3
6	F	48	132	85	10.1	6	3
7	F	53	129	67	7.4	7	2
8	F	62	138	85	8	8	3
9	F	64	126	75	10.2	9	2
10	F	47	101	73	9.3	10	1
11	F	67	133	65	9.2	11	3
12	F	50	97	60	7.9	12	1
13	F	58	104	69	10.6	13	1
14	F	61	109	65	9.4	14	1

Female (F) n=244 are extracted

Output of results, etc.

health2 - Multivariate of sbp, dbp - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

New
Open... Ctrl+O
Quick Open Alt+Shift+O
Import Multiple Files...
Close Ctrl+W
Save Ctrl+S
Save As...
Revert
Export...
Publish
Database
Internet Open
SAS
Preferences Ctrl+K
Print... Ctrl+P
Print Preview
Page Setup...
Recent Files
Save Session Script...
Exit JMP Ctrl+Q

dbp

100 120 140 160 180

60 80 100 120

Nonparametric: Spearman's ρ

Variable	by Variable	Spearman ρ	Prob>|ρ|	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
dbp	sbp	0.7193	<.0001*	

Export

Select the format to use for exporting this report:

- Text
  Creates a text document that can be viewed and edited by many applications.
- Image — Format: PNG
  Creates a single image document that can be viewed and edited with an image editing application. Several raster and vector image formats are available.
- PDF
  Creates a PDF document, which can be viewed with a web browser or PDF reader.
- HTML
  Creates an HTML document and associated image documents that can be viewed with a web browser.
- Interactive HTML with Data
  Creates an HTML document containing data and dynamic graphics that can be explored in a web browser.
- RTF
  Creates a Rich Text Format document that can be viewed and edited with most word processing applications.
- Microsoft Word
  Creates a Microsoft Word document.
- Microsoft PowerPoint
  Export the result from a JMP Report as a Microsoft PowerPoint presentation.

Next... Cancel

Save Analysis Results as “Script”

health2 - Multivariate of sbp, dbp - JMP Pro

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

Multivariate

- Correlations Multivariate
- Correlation Probability
- CI of Correlation
- Inverse Correlations
- Partial Correlations
- Partial Correlation Probability
- Covariance Matrix
- Pairwise Correlations
- Hotelling's T^2 Test
- Simple Statistics
- Nonparametric Correlations
- Set α Level
- Scatterplot Matrix
- Color Maps
- Parallel Coord Plot
- Ellipsoid 3D Plot
- Partial Correlation Diagram
- Outlier Analysis
- Item Reliability
- Local Data Filter
- Redo
- Platform Preferences
- Save Script

To Data ...
To Journal
To Script Window
To Log
To Workflow
To Report
To Clipboard

analysis, and append it to the current Script text window.

Script Window - JMP Pro

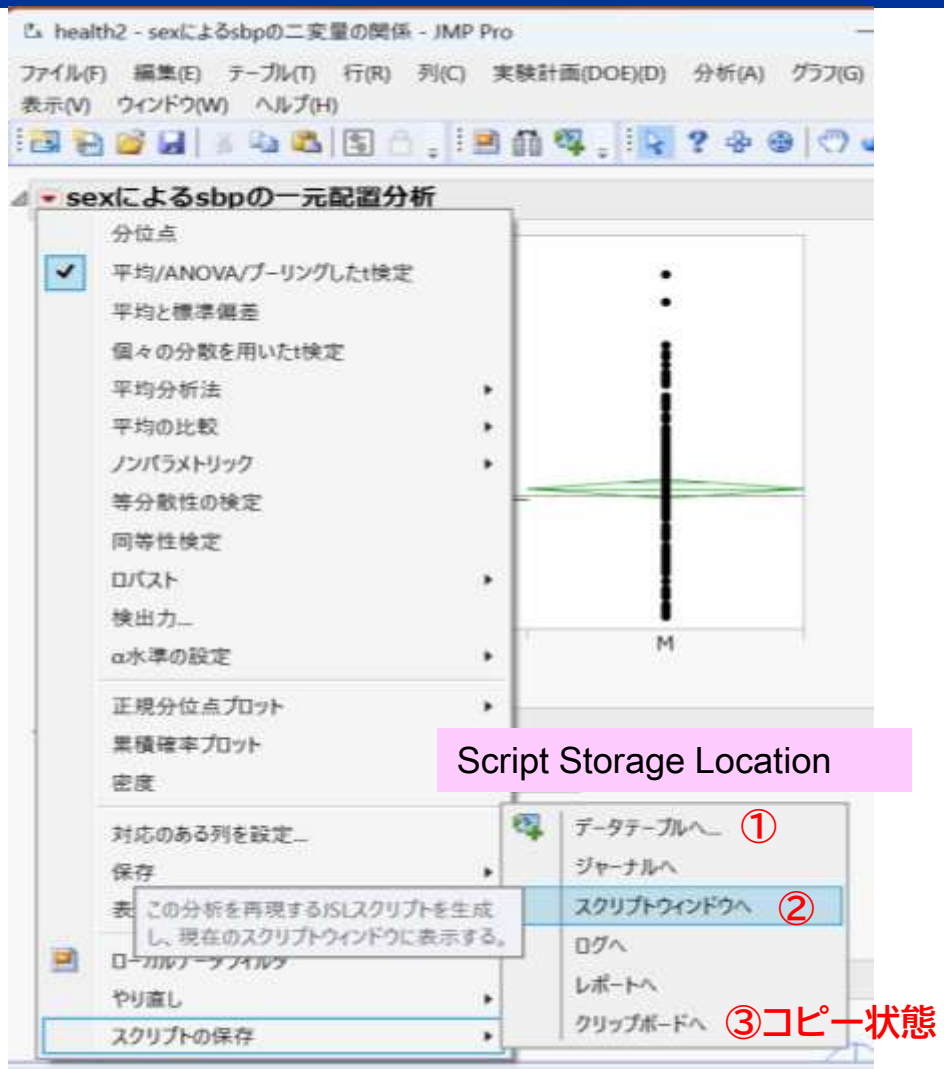
File Edit Tables DOE Analyze Graph Tools Add-Ins View Window Help

health2

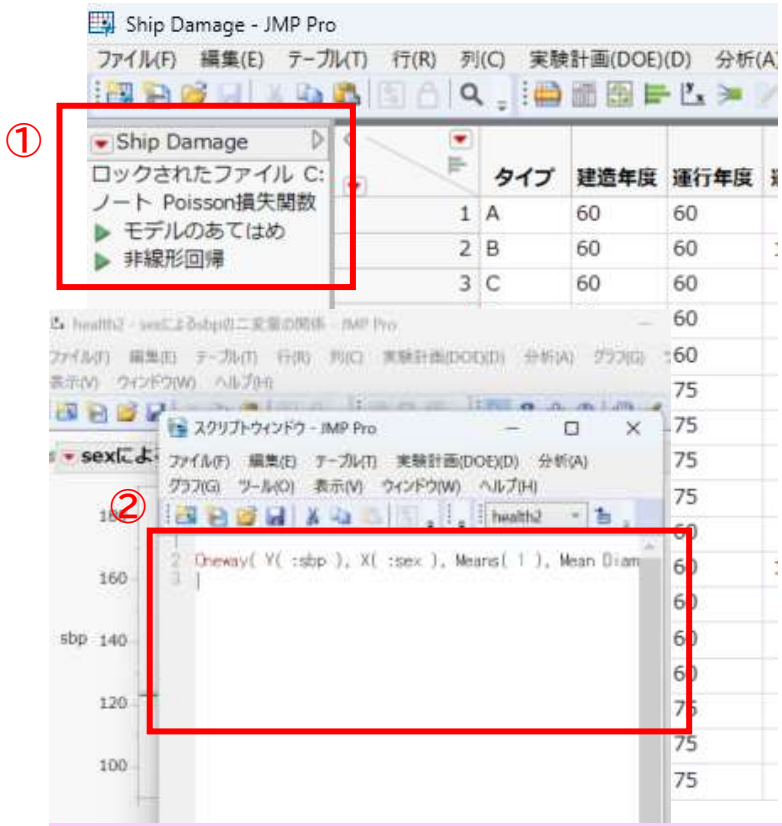
```
Multivariate(
    Y( :sbp, :dbp ),
    Variance Estimation( "Row-wise" ),
    Scatterplot Matrix( Horizontal( 1 ) ),
    Spearman's ρ( 1 ),
    Parallel Coord Plot
);
```

Save the script to analyze it in the same way in the future.

How to save scripts



Script Storage Location



For Click-type packages such as JMP and SPSS, it is desirable to save scripts (syntax) to guarantee reproducibility of analysis methods.

Report

- For the following assignments shown in the lecture slides, check that the results can be reproduced using a statistical package JMP.

  - Check the shape, location, and distribution (var= SBP, DBP, SALT).
  - Using “health2” dataset, calculate correlation coefficients between blood pressure and salt intake.
  - The ”Health4” dataset created in the course of the exercise has a blood pressure category variable (HTN_c) added to it (HTN_c should Recode 4-6 as 4).

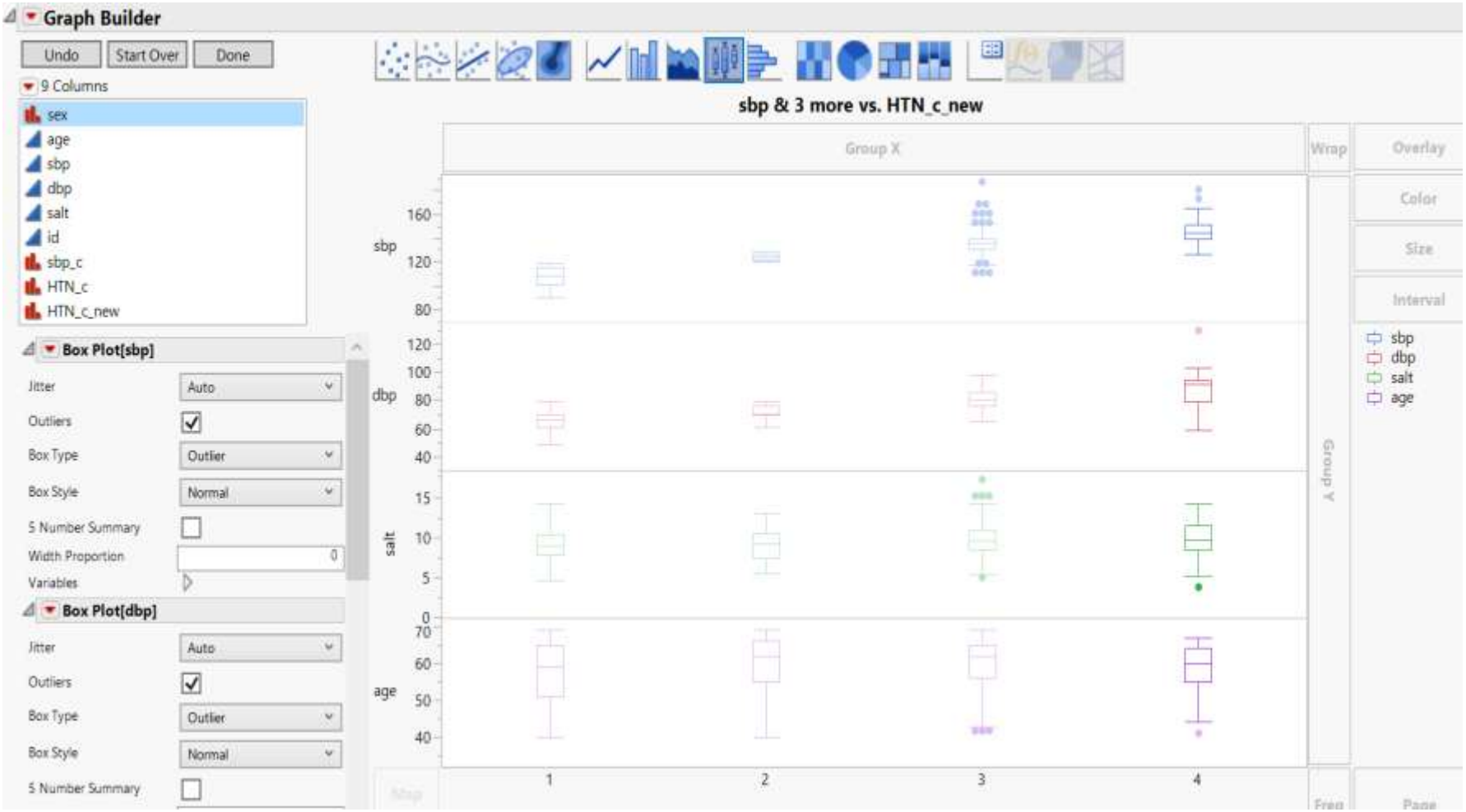
<u>Report</u>

Q1：Create a box-and-whisker diagram of SBP, DBP, and SALT for each blood pressure category.

→Use Graph builder.

Q2: Examine the relationship between blood pressure (SBP, DBP), age, and salt intake by examining scatter plots and correlations, and describe what you notice about the relationship between the four variables.Use and explain necessary graphs as appropriate.

Hint


Graph Builder
Undo
Start Over
Done
9 Columns
sex
age
sbp
dbp
salt
id
sbp_c
HTN_c
HTN_c_new
Box Plot[sbp]
Jitter
Auto
Outliers
Box Type
Outlier
Box Style
Normal
5 Number Summary
Width Proportion
0
Variables
Box Plot[dbp]
Jitter
Auto
Outliers
Box Type
Outlier
Box Style
Normal
5 Number Summary
sbp & 3 more vs. HTN_c_new
Group X
Wrap
Overlay
Color
Size
Interval
sbp
dbp
salt
age
Group Y
sbp
dbp
salt
age
1
2
3
4
Freq
Page

Biostatistics

Course #01- part B

NCD Epidemiology Research Center (NERC)
Department of Medical statistics

Akiko HARADA

JMP → Python

Today's overview

1. Usage of Statistical Software (package)

1) Brief Introduction to JMP

2) Reading Data

In the next course (#02),
we will try the same content in Python.

2. Description of Data

1) Location: mean, median, percentile
2) Dispersion: range, inter-quartile range, CV
3) Distribution: skewness, kurtosis, Normality
4) Association between two variables: Correlation

Graphics

Supplementary material for Python exercises

Google Colaboratory
https://colab.research.google.com/?hl=en-GB#

Welcome to Colaboratory

File Edit View Insert Runtime Tools Help

Share

Table of contents

Getting started

Data science

Machine learning

More resources

Featured examples

+ Section

+ Code + Text Copy to Drive

Connect

Google login

Welcome to Colab!

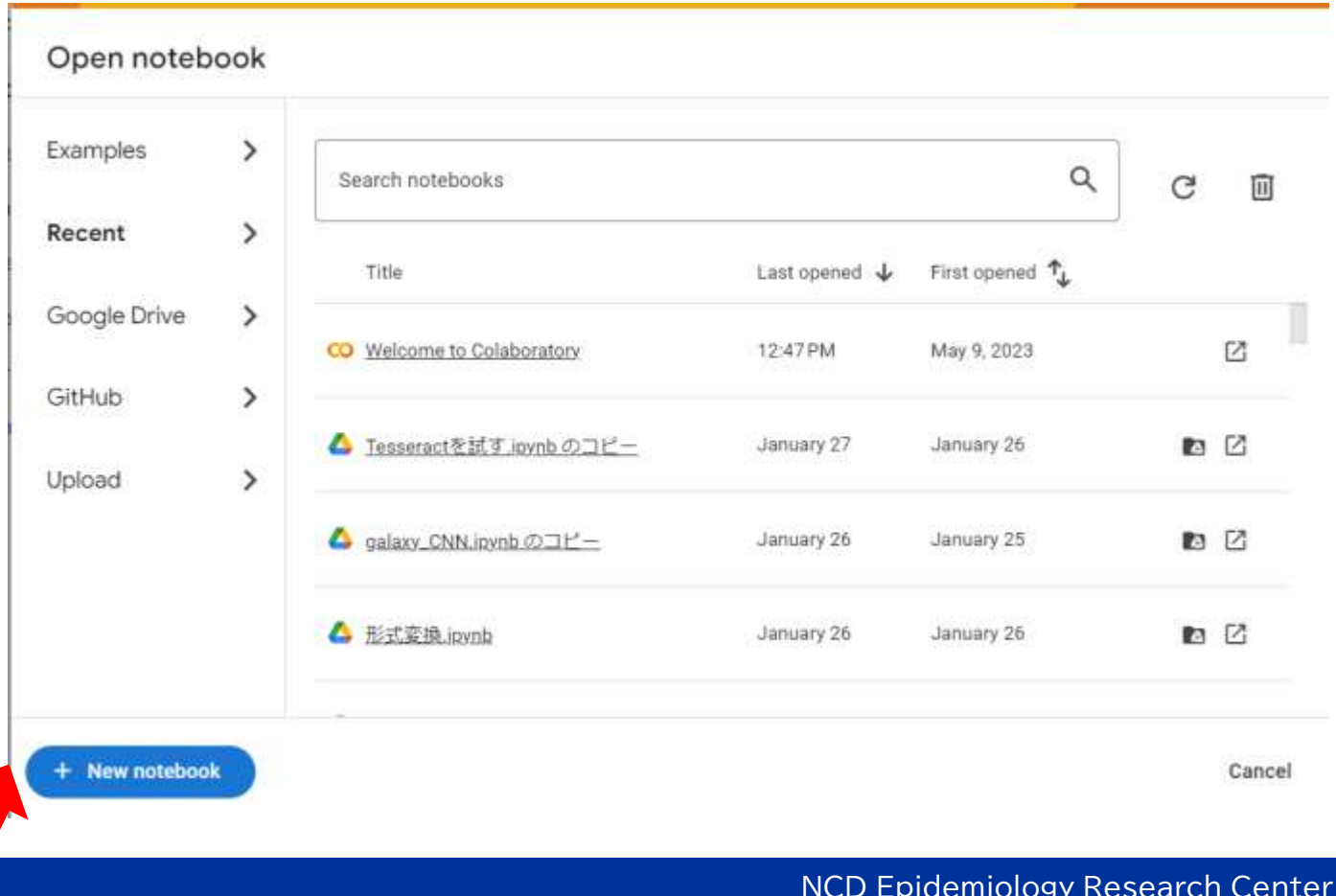
(New) Try the Gemini API

- Generate a Gemini API key
- Talk to Gemini with the Speech-to-Text API
- Gemini API: Quickstart with Python
- Gemini API code sample
- Compare Gemini with ChatGPT
- More notebooks

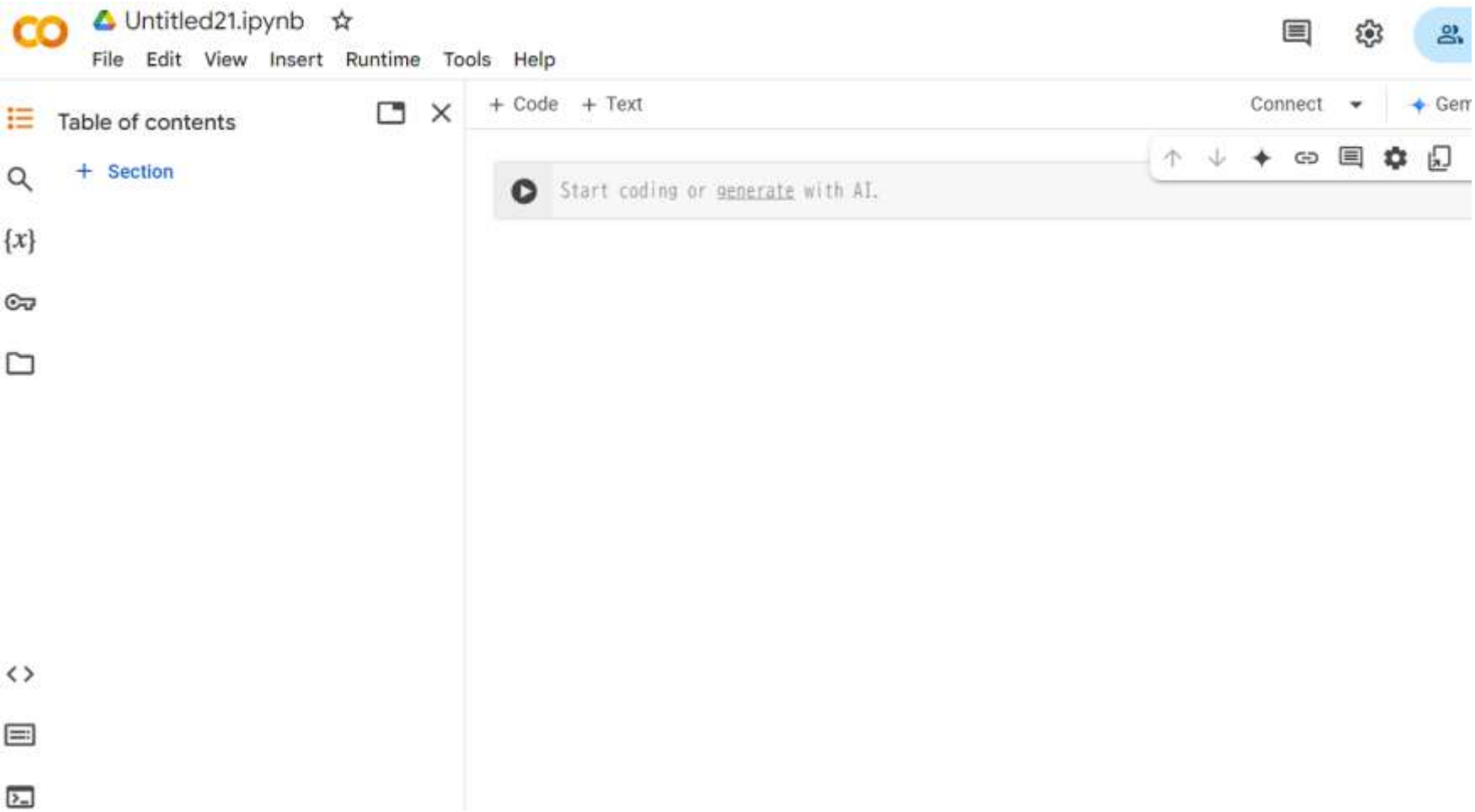
If you're already familiar with Colab, check out this video to learn about interactive tables, the executed code history view and the command palette.

Google Colaboratory

https://colab.research.google.com/?hl=en



Google Colaboratory



notebook

Welcome To Colab

File Edit View Insert Runtime Tools Help

Make Code cell

Share

Table of contents

Getting started

Data science

Machine learning

More Resources

Featured examples

+ Section

+ Code + Text Copy to Drive Connect Gemini

Getting started

The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

```
[ ] seconds_in_a_day = 24 * 60 * 60
    seconds_in_a_day
```

```
86400
```

Cell: Cells contain either explanatory text or executable code and its output. Click a cell to select it.

To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut "Command/Ctrl+Enter". To edit the code, just click the cell and start editing.

Variables that you define in one cell can later be used in other cells:

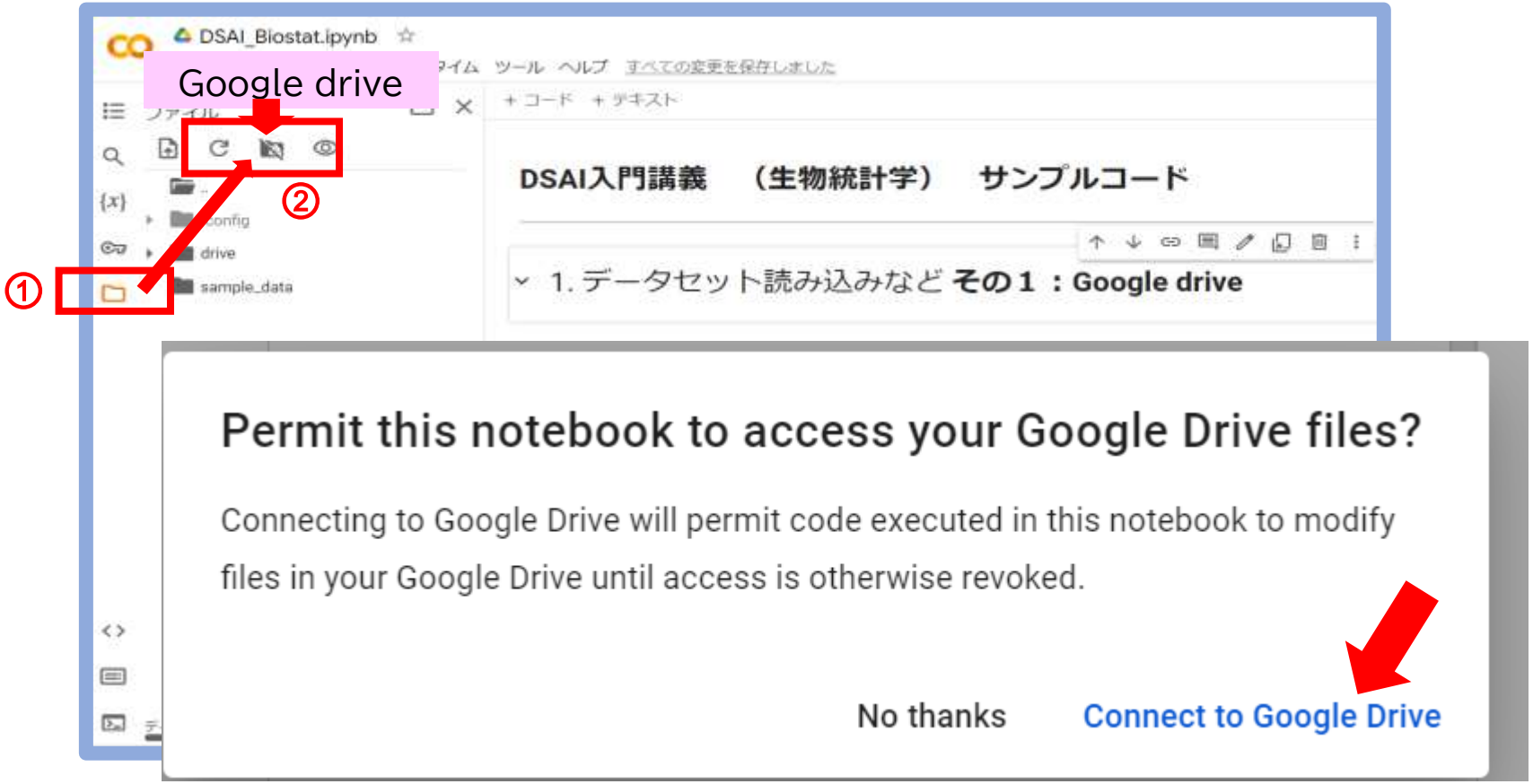
```
[ ] seconds_in_a_week = 7 * seconds_in_a_day
    seconds_in_a_week
```

Setup for analysis with Google Colaboratory

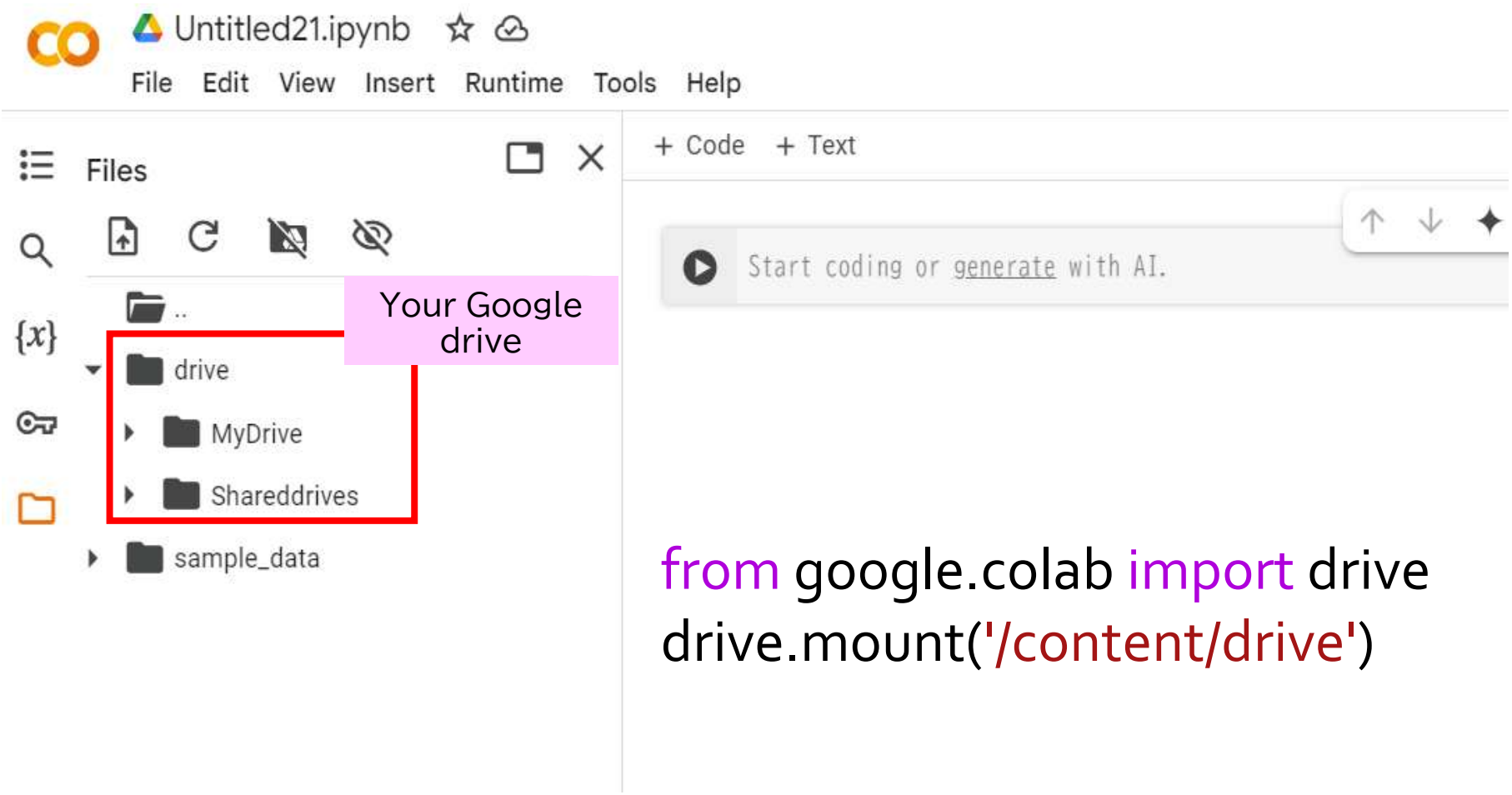
Integration with Drive

- Colaboratory is integrated with Google Drive. It allows you to share, comment, and collaborate on the same document with multiple people:
  - The **SHARE** button (top-right of the toolbar) allows you to share the notebook and control permissions set on it.
  - **File->Make a Copy** creates a copy of the notebook in Drive.
  - **File->Save** saves the File to Drive.

Pattern1. Drive mount



My Drive, a Google Drive, is added under "/content/drive".



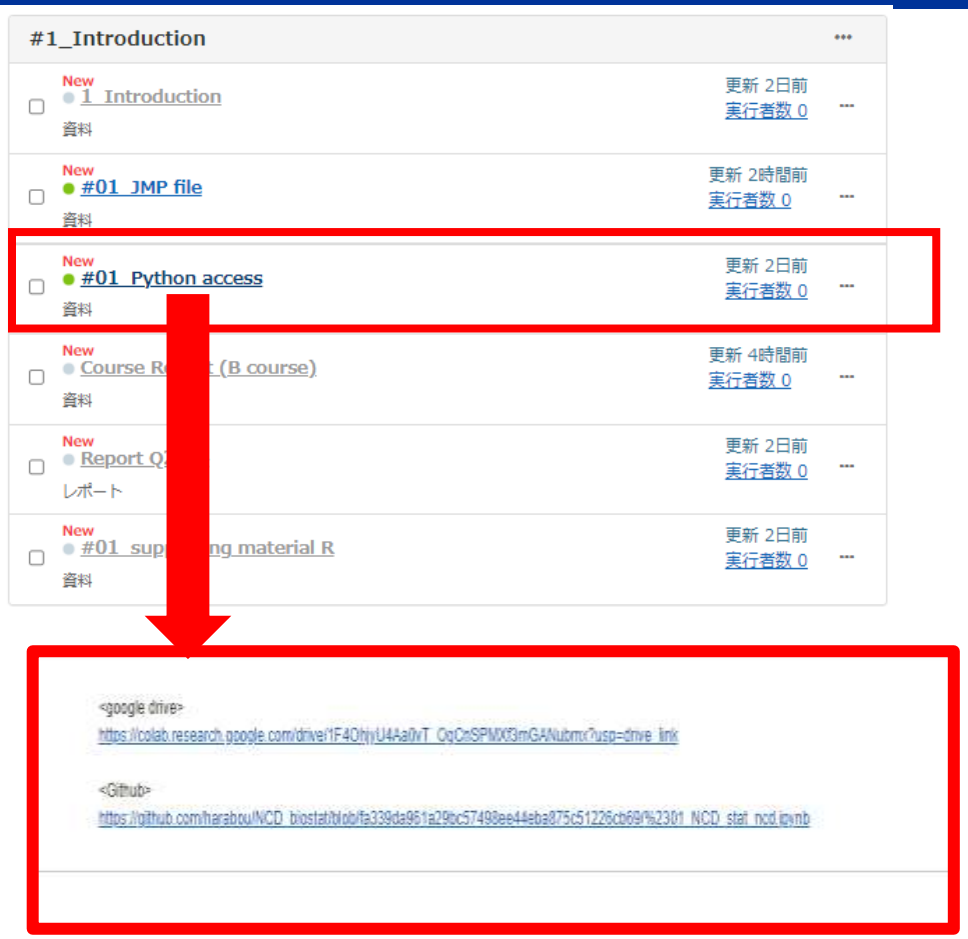

```
from google.colab import drive
drive.mount('/content/drive')
```

Files for this exercise course

- Webclass
  - Google drive LINKs are stored.

- Google drive
  - https://colab.research.google.com/drive/1F4OhjyU4Aa0vT_OgCnSPMXf3mGANubmx?usp=drive_link

- Github
  - https://github.com/harabou/NCD_biostat/blob/fa339da961a29bc57498ee44eba875c51226cb69/%2301_NCD_stat_ncd.ipynb



#1_NCD_stat _ncd course

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ

目次

01 NCD course biostat

Pattern 1 : Drive mount

Pattern2: Drive mount

2 Graph

Fisher's Iris data

1)Histgram

2)Box plot

Health2 data

3)QQ plot

4)Correlation

3.Via. Web data

コマンド + コード + テキスト

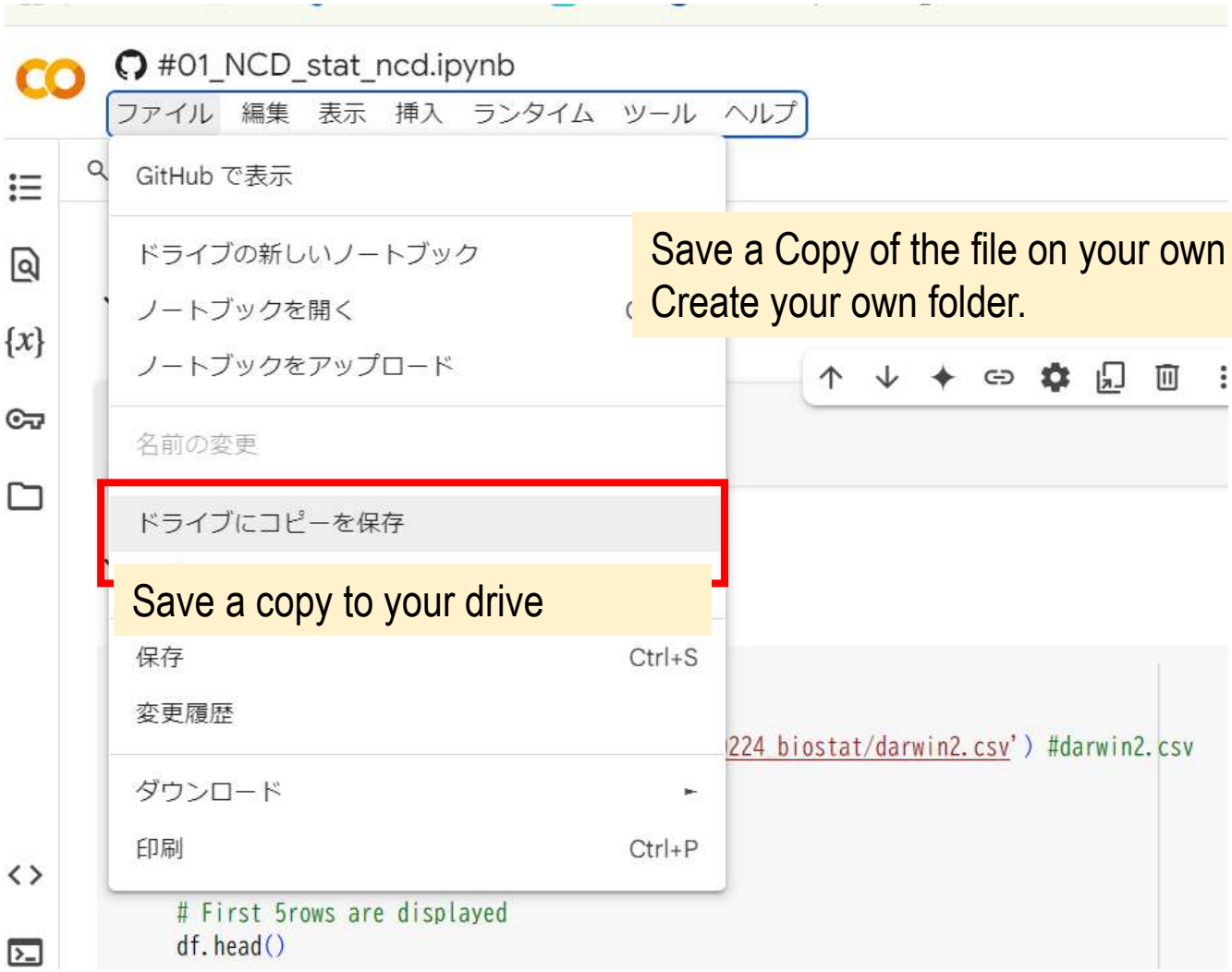
#01 NCD course biostat

```
from google.colab import drive
drive.mount('/content/drive')
```

+ コード + テキスト

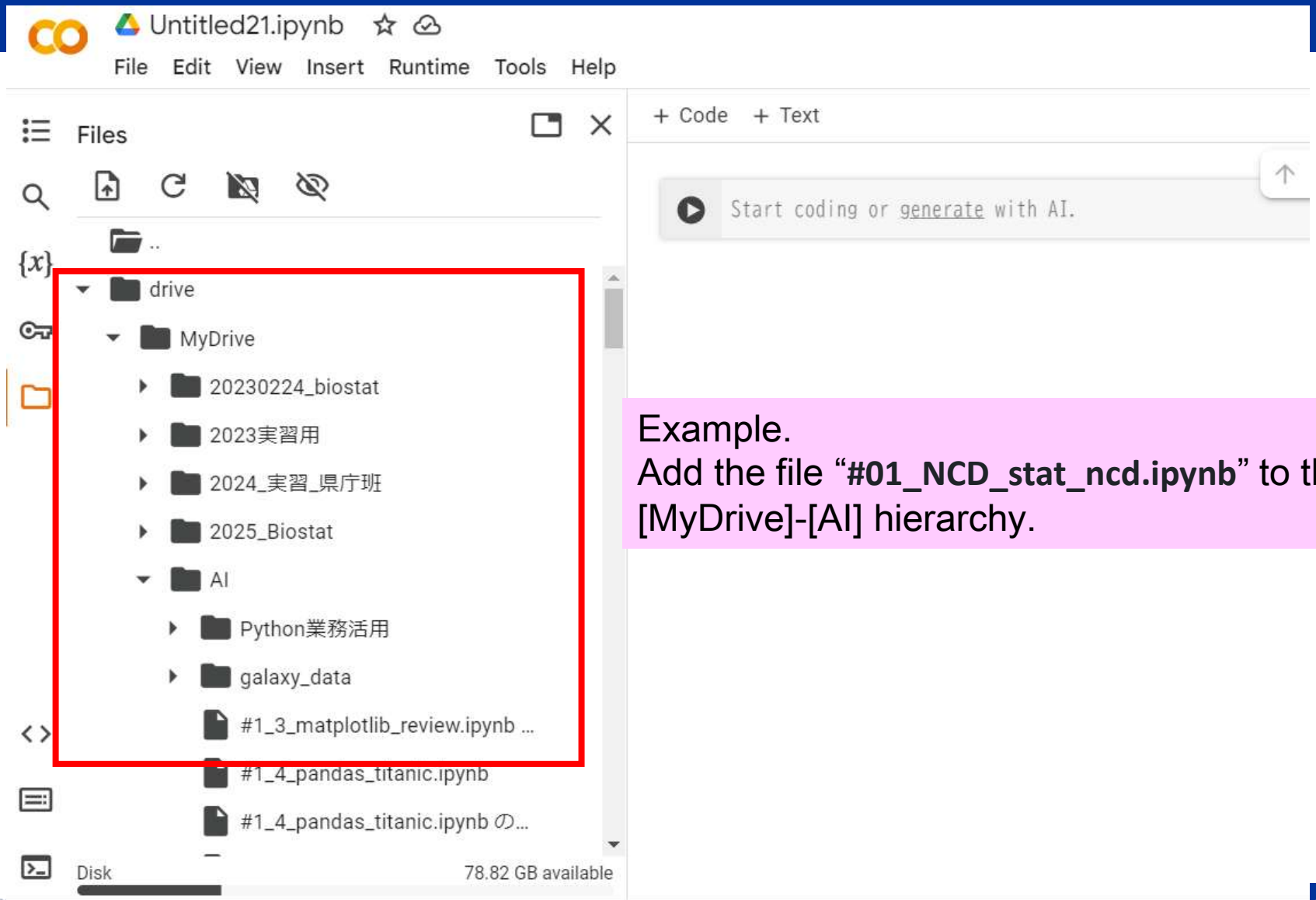
1. Pattern 1 : Drive mount

```
[ ] #Importing from Google drive
    import pandas as pd
    df= pd.read_csv('/content/drive/MyDrive/20230224_biostat/darwin2.csv') #darwin2.c

    # Check (row, column)
    df.shape

    # First 5rows are displayed
    df.head()
```



Save a Copy of the file on your own Google drive.
Create your own folder.

Save a copy to your drive



Example.
Add the file "**#01_NCD_stat_ncd.ipynb**" to the [MyDrive]-[AI] hierarchy.

From “Welcome to Colab” https://colab.research.google.com/?hl=en

Open notebook

Examples

Recent

Google Drive

GitHub

Upload

Search notebooks

Title	Last opened	First opened
#01_NCD_stat_ncd.ipynb	7:49 PM	7:35 PM
		9, 2023
#1_NCD_stat_ncd course	7:42 PM	February 12

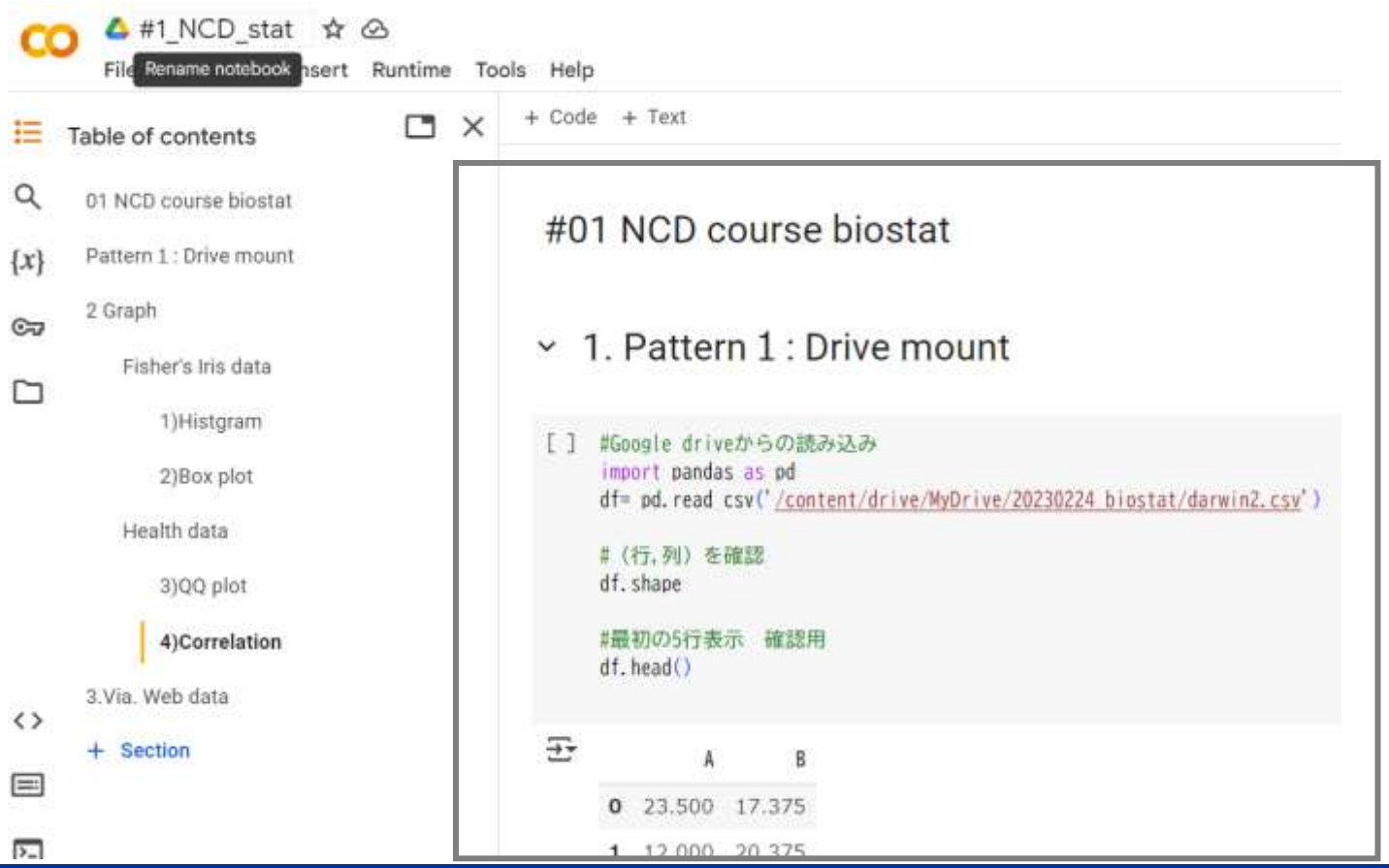
Open “**#01_NCD_stat_ncd.ipynb**”

+ New notebook

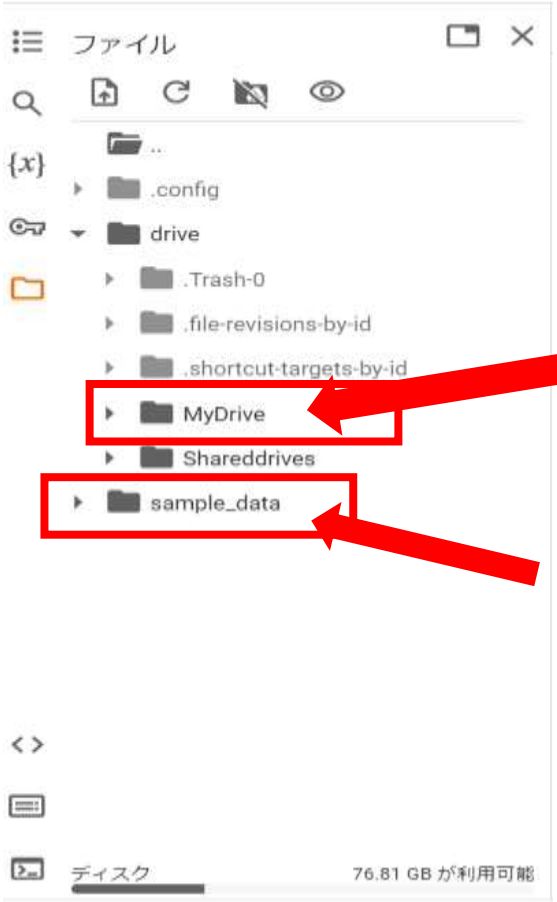
Cancel

Code「.ipynb」

- #01_NCD_stat.ipynb

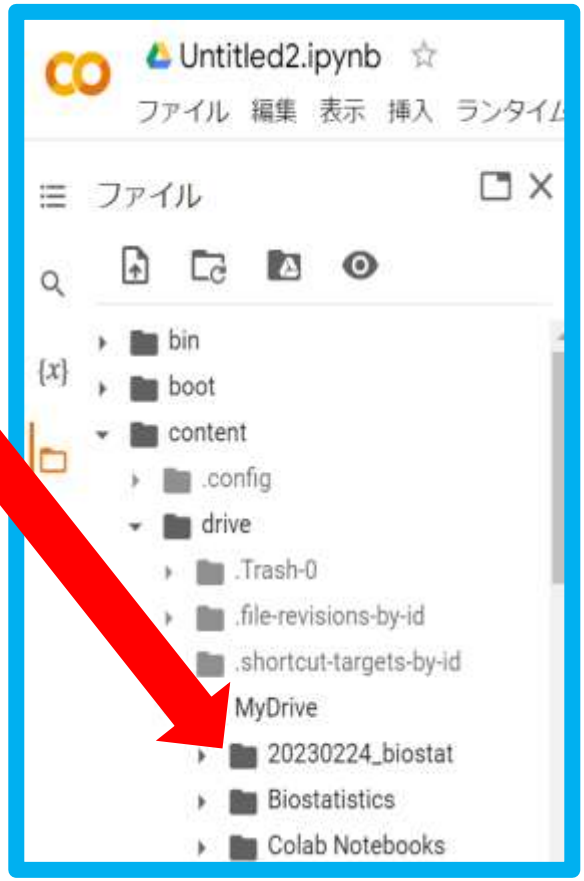


Attention.　Contents in「Sample folder」disappear after the session.



Saved after the session ends.

Disappears after the session.



Pattern2. Via. Github

- Data is saved on DSAI's Github, so data can be used from there.

→ The data can be used as it is without moving it.

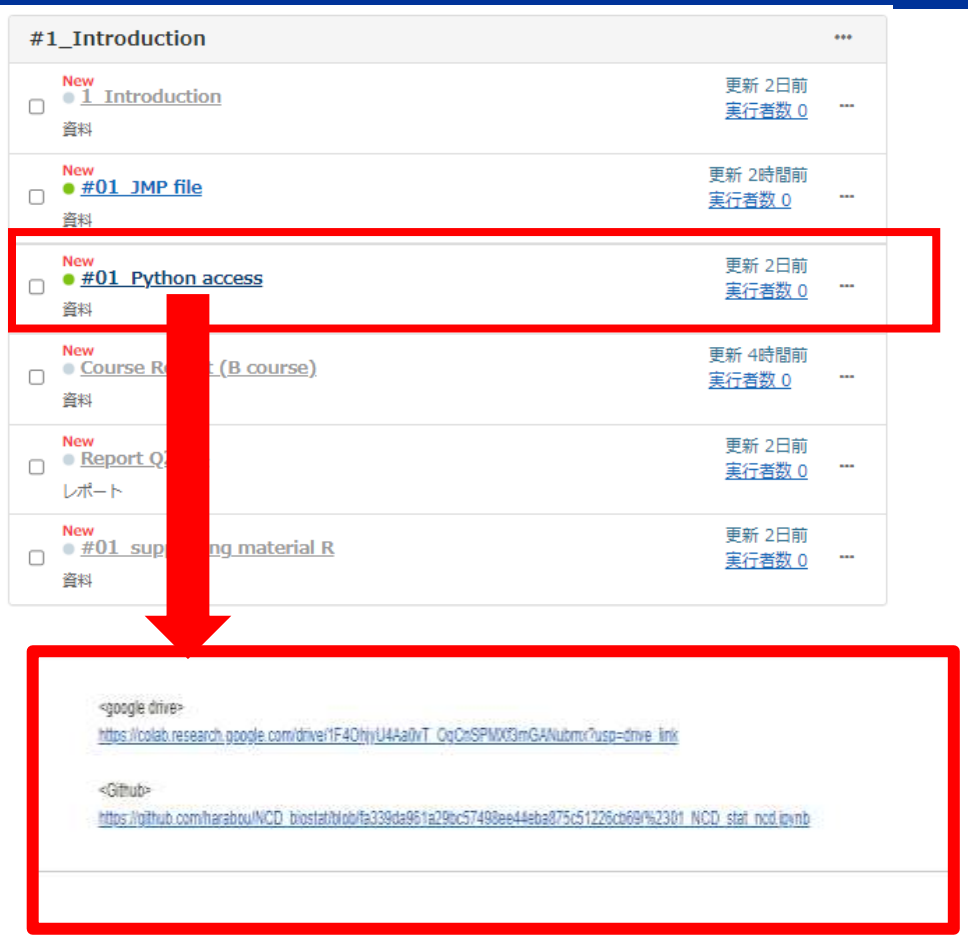
- Most of the provided code is as direct links from Github.

Files for this exercise course

- Webclass
  - Google drive LINKs are stored.

- Google drive
  - https://colab.research.google.com/drive/1F4OhjyU4Aa0vT_OgCnSPMXf3mGANubmx?usp=drive_link

- Github
  - https://github.com/harabou/NCD_biostat/blob/fa339da961a29bc57498ee44eba875c51226cb69/%2301_NCD_stat_ncd.ipynb



Github
https://github.com/harabou/NCD_biostat

harabou / NCD_biostat

Code Issues Pull requests

NCD_biostat Public

main 1 Branch 0 Tags

harabou Colab を使用して作成さ

- #01
- #02
- #03
- #01_NCD_stat_ncd.ipynb
- #02_ttest_chi_square_NCD.ipynb
- #03_General_linear_model.ipynb
- #04__GLM_NCD2.ipynb
- #05_Survival_analysis.ipynb

Python Code

Open in Colab

#01 NCD course biostat

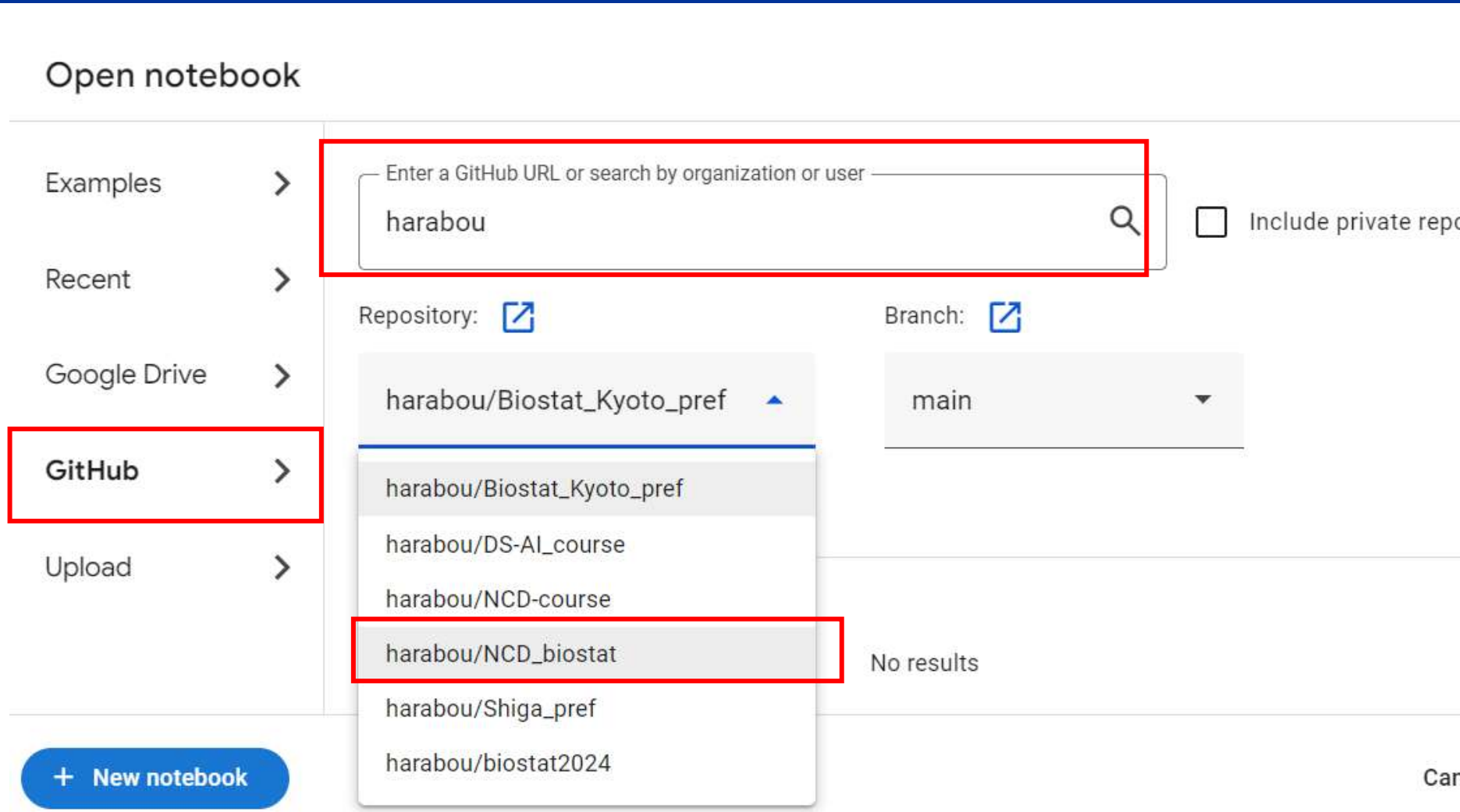
In [ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

1. Pattern 1 : Drive mount

In [ ]:
```
#Importing from Google drive
import pandas as pd
df= pd.read_csv('/content/drive/MyDrive/20230224_biostat/darwin2.csv') #darwin2.csv

# Check (row, column)
df.shape

# First 5rows are displayed
df.head()
```

Out[ ]:

	A	B
0	23.500	17.375
1	12.000	20.375


Open notebook
Examples
Recent
Google Drive
GitHub
Upload
Enter a GitHub URL or search by organization or user
harabou
Include private repo
Repository:
Branch:
harabou/Biostat_Kyoto_pref
main
harabou/Biostat_Kyoto_pref
harabou/DS-AI_course
harabou/NCD-course
harabou/NCD_biostat
harabou/Shiga_pref
harabou/biostat2024
No results
New notebook

Open notebook

Examples

Recent

Google Drive

GitHub

Upload

harabou

Include private repos

Repository:

harabou/NCD_biostat

Branch:

main

Path

#01_NCD_stat_ncd.ipynb

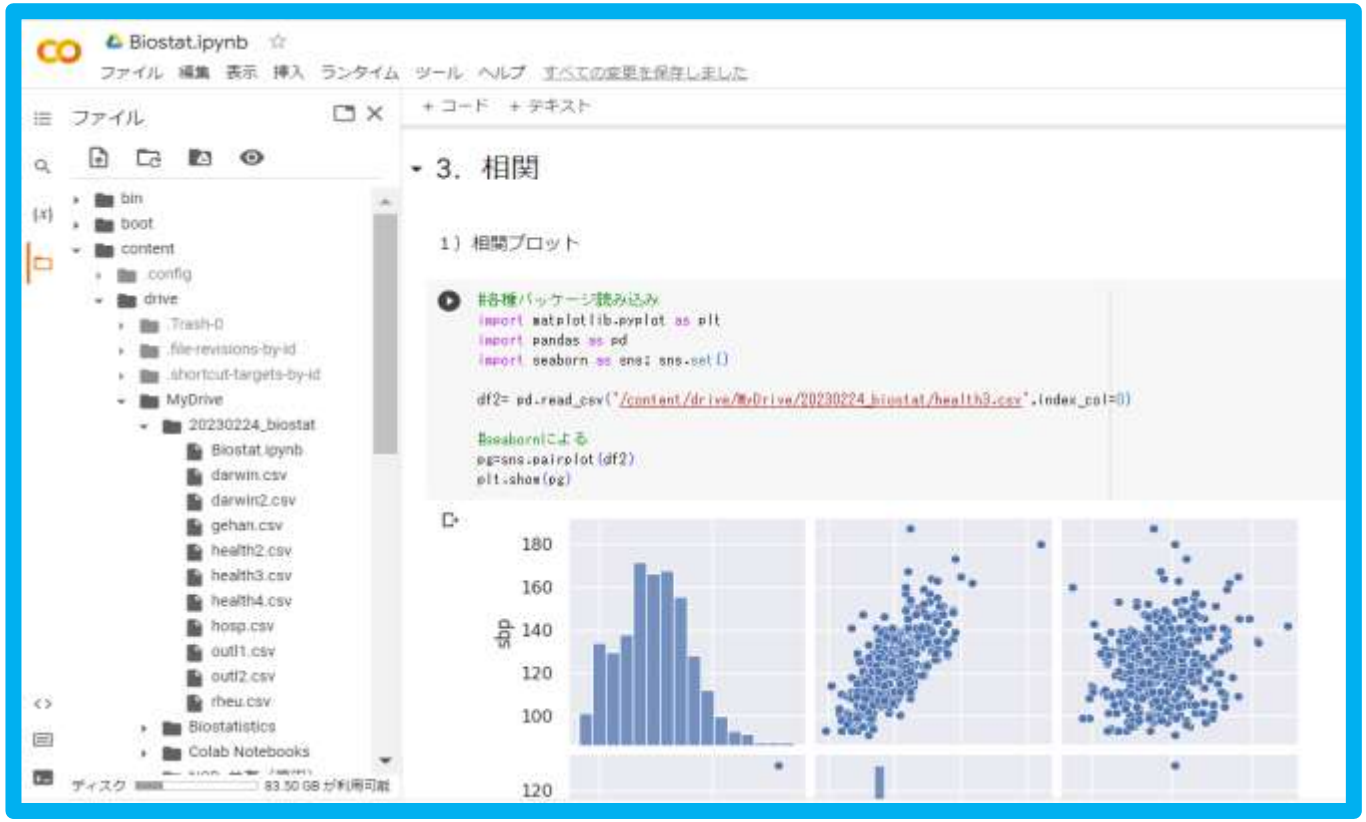
#02_ttest_chi_square_NCD.ipynb

+ New notebook

Cancel

Python code

- **#01_NCD_stat_ncd.ipynb**



This is a comment in Python, indicated by the # symbol. It translates to "Loading various packages" and is simply a note for readers of the code.

Loading packages

```
#Loading various packages
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import seaborn as sns
```

To bring external libraries (packages of pre-written code) into your current script, allowing you to use their functionalities.

matplotlib.pyplot is a library widely used for creating static, interactive, and animated visualizations in Python. as plt creates a shorter alias plt for convenience, so you can refer to it with plt instead of the full name.

pandas is a powerful library for data manipulation and analysis. It provides data structures like DataFrames for efficiently working with tabular data. as pd creates a shorter alias pd.

numpy is the fundamental package for scientific computing in Python. It offers tools for working with arrays, mathematical functions, and more. as np creates a shorter alias np.

seaborn is a library built on top of matplotlib that provides a high-level interface for creating informative and attractive statistical graphics. as sns creates a shorter alias sns.

Loading Dataset

```
fpath='https://raw.githubusercontent.com/harabou/DS-AI_course/main/%2301/Iris.csv'

df = pd.read_csv(fpath,header=0)

df.head()
```

This variable is then used by the pd.read_csv() function to read the data from the CSV file located at that URL into a pandas DataFrame named df.

This allows the user to work with the data in their Python code.

Pattern1: “fpath” Github

harabou / DS-AI_course

Type / to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

DS-AI_course Public

Pin Unwatch

main 1 Branch 0 Tags Go to file t Add file Code

harabou Add files via upload

#01
#02
#03
#04
#05
#06
#07
#SSDSE Add files via upload 4 months ago

harabou Add files via upload

Name	Last commit message
..	
#01_NCD_Biostatistics.jmpprj	Add files via upload
#08_Compare.jmpprj	Add files via upload
Iris 3.jmp	Add files via upload
Iris.csv	Update Iris.csv
Iris.jmp	Add files via upload

https://raw.githubusercontent.com/harabou/DS-AI_course/refs/heads/main/%2301/Iris.csv

151 lines (151 loc) · 3.89 KB

Preview Code Blame Raw

Search this file

1	gaku_l	gaku_w	kaben_l	kaben_w	group
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3.0	1.4	0.2	setosa
4	4.7	3.2	1.3	0.2	setosa
5	4.6	3.1	1.5	0.2	setosa
6	5.0	3.6	1.4	0.2	setosa
7	5.4	3.9	1.7	0.4	setosa
8	4.6	3.4	1.4	0.3	setosa
9	5.0	3.4	1.5	0.2	setosa
10	4.4	2.9	1.4	0.2	setosa
11	4.9	3.1	1.5	0.1	setosa

raw.githubusercontent.com/harabou/DS-AI_cour

クリックすると前に戻ります。押したまま待つと履歴が表示されます。

ログイン

```
gaku_l,gaku_w,kaben_l,kaben_w,group
5.1,3.5,1.4,0.2,setosa
4.9,3.0,1.4,0.2,setosa
4.7,3.2,1.3,0.2,setosa
4.6,3.1,1.5,0.2,setosa
5.0,3.6,1.4,0.2,setosa
5.4,3.9,1.7,0.4,setosa
4.6,3.4,1.4,0.3,setosa
5.0,3.4,1.5,0.2,setosa
4.4,2.9,1.4,0.2,setosa
4.9,3.1,1.5,0.1,setosa
5.4,3.7,1.5,0.2,setosa
4.8,3.4,1.6,0.2,setosa
4.8,3.0,1.4,0.1,setosa
4.3,3.0,1.1,0.1,setosa
5.8,4.0,1.2,0.2,setosa
5.7,4.4,1.5,0.4,setosa
5.4,3.9,1.3,0.4,setosa
5.1,3.5,1.4,0.3,setosa
5.7,3.8,1.7,0.3,setosa
5.1,3.8,1.5,0.3,setosa
5.4,3.4,1.7,0.2,setosa
5.1,3.7,1.5,0.4,setosa
4.6,3.6,1.0,0.2,setosa
5.1,3.3,1.7,0.5,setosa
4.8,3.4,1.9,0.2,setosa
5.0,3.0,1.6,0.2,setosa
5.0,3.4,1.6,0.4,setosa
5.2,3.5,1.5,0.2,setosa
5.2,3.4,1.4,0.2,setosa
4.7,3.2,1.6,0.2,setosa
4.8,3.1,1.6,0.2,setosa
5.4,3.4,1.5,0.4,setosa
5.2,4.1,1.5,0.1,setosa
5.5,4.2,1.4,0.2,setosa
```

Pattern2: “fpath”　Google drive

/content/drive/MyDrive/がマイドライブに相当

CSVを保存しているフォルダを指定
（最後の/を忘れないように注意）

```
f_path = '/content/drive/MyDrive/AI/'

df = pd.read_csv(f_path + ‘ilis.csv')
```

CSVファイルをデータフレームとして読み込む

This updates the default settings (rcParams) for matplotlib plots. Specifically, it sets the font size to 15, making the text in your plots larger and easier to read.

Graph –matplotlib-

```
x = np.random.normal(size = 1000)
fpath = 'https://raw.githubusercontent.com/harabou/DS-AI_course/main/%2301/Iris.csv'

df = pd.read_csv(fpath,header=0, encoding='shift-jis')

#matplotlibを使用したプロット
plt.style.use('ggplot')

plt.rcParams.update({'font.size':15})

%matplotlib inline
```

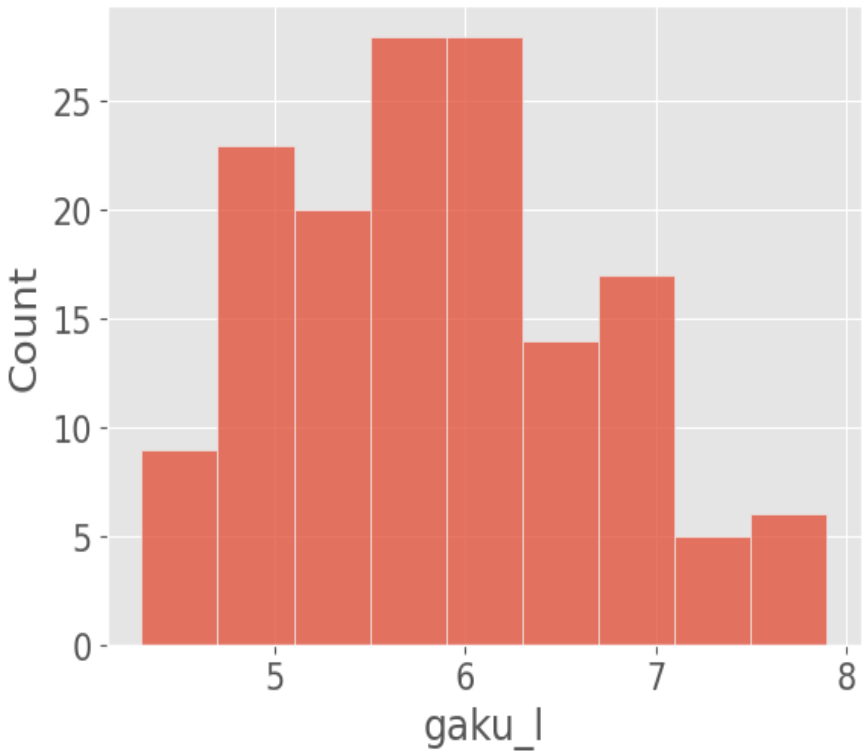
This line sets the plotting style to 'ggplot'. Styles control the visual appearance of your plots, and 'ggplot' is known for its clean and modern aesthetic, often used in data visualization.

This updates the default settings (rcParams) for matplotlib plots. Specifically, it sets the font size to 15, making the text in your plots larger and easier to read.

This is a "magic command" specific to Jupyter Notebooks. It ensures that your plots are displayed directly within the notebook output cells, rather than in separate windows.

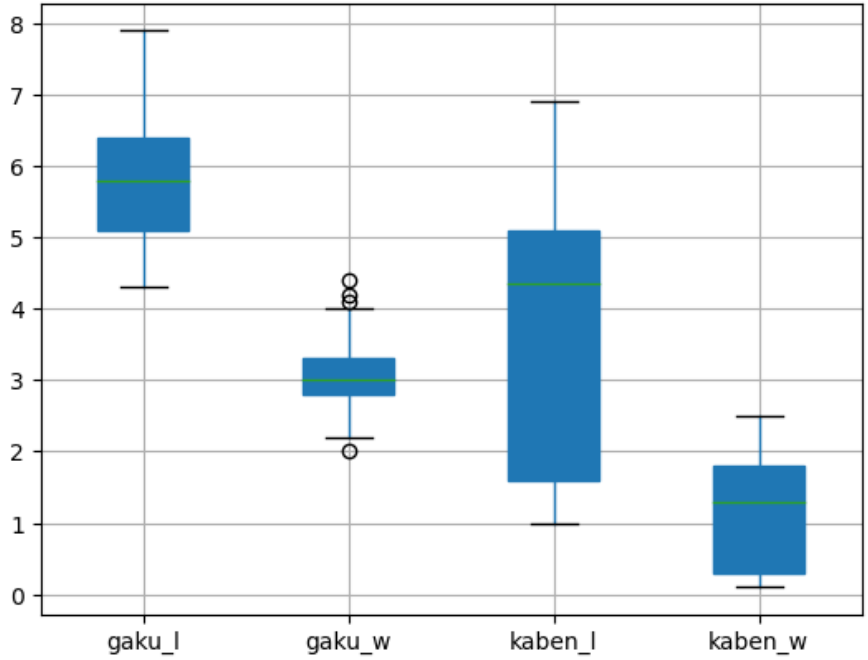
Graph –seaborn-

```
x = np.random.normal(size = 1000)
df = pd.read_csv(fpath,header=0,
encoding='shift-jis’)

# use seaborn
sns.histplot(df.gaku_l)
```

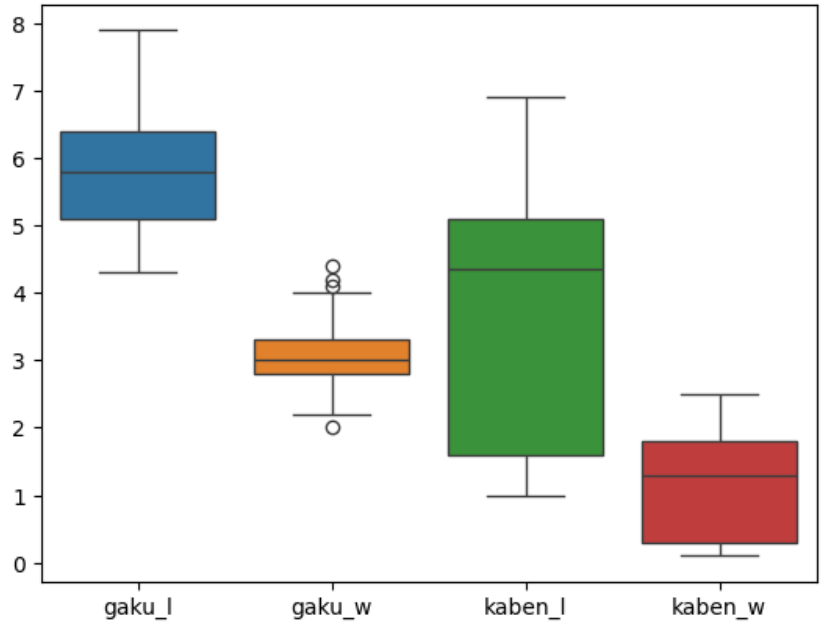


Box plot

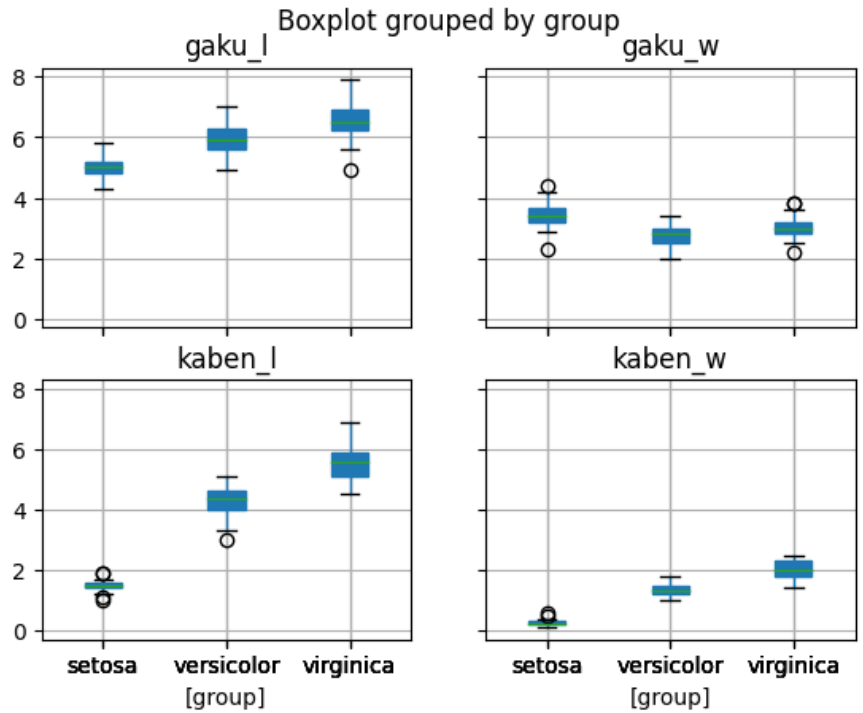
```
df.boxplot(patch_artist = True)
```



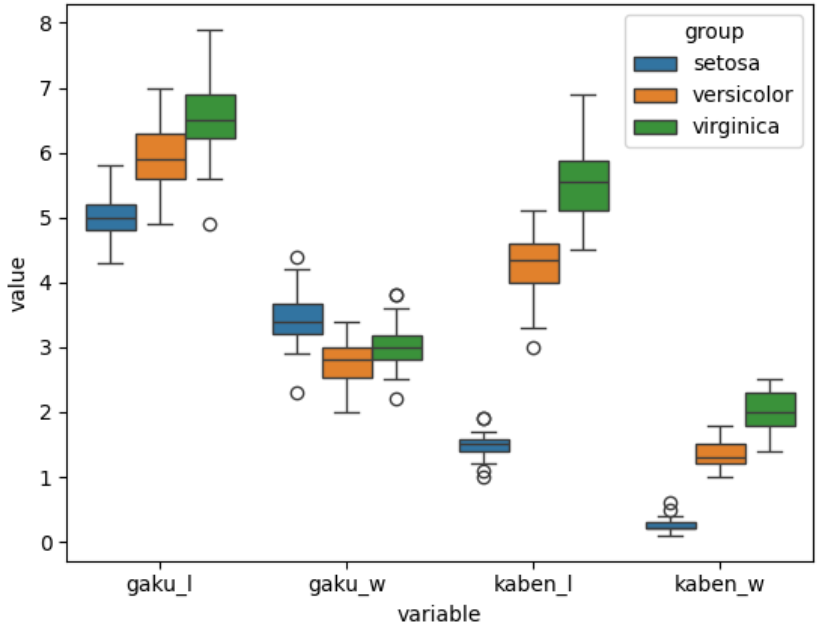

```
sns.boxplot(data=df)
```



```
df.boxplot(by='group', patch_artist = True)
```




```
w = pd.melt(df, id_vars = ["group"])
display(w.head())

sns.boxplot(x = "variable", y = "value", data = w, hue = "group")
```



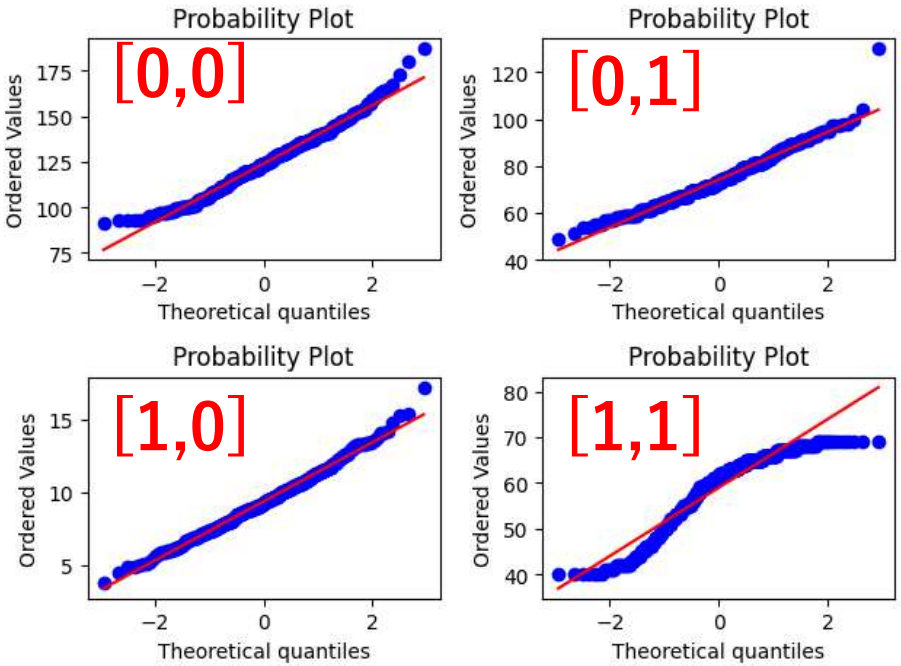
Health data

```
fpath =
'https://raw.githubusercontent.com/harabo
u/DS-AI_course/main/%2301/health2.csv’

df = pd.read_csv(fpath,header=0)

df.head()
```

Q-Q plot

```
import scipy.stats as stats
from scipy.stats import norm
import matplotlib.pyplot as plt

sbp = df['sbp']
dbp = df['dbp']
salt = df['salt']
age = df['age']

fig, ax = plt.subplots(2, 2)
probplot(sbp, plot=ax[0, 0])
probplot(dbp, plot=ax[0, 1])
probplot(salt, plot=ax[1, 0])
probplot(age, plot=ax[1, 1])
plt.tight_layout()
plt.show()
```

import scipy.stats as stats:
This line imports the scipy.stats module, which contains a wide range of statistical functions, including those for creating Q-Q plots. It's aliased as stats for easier use.



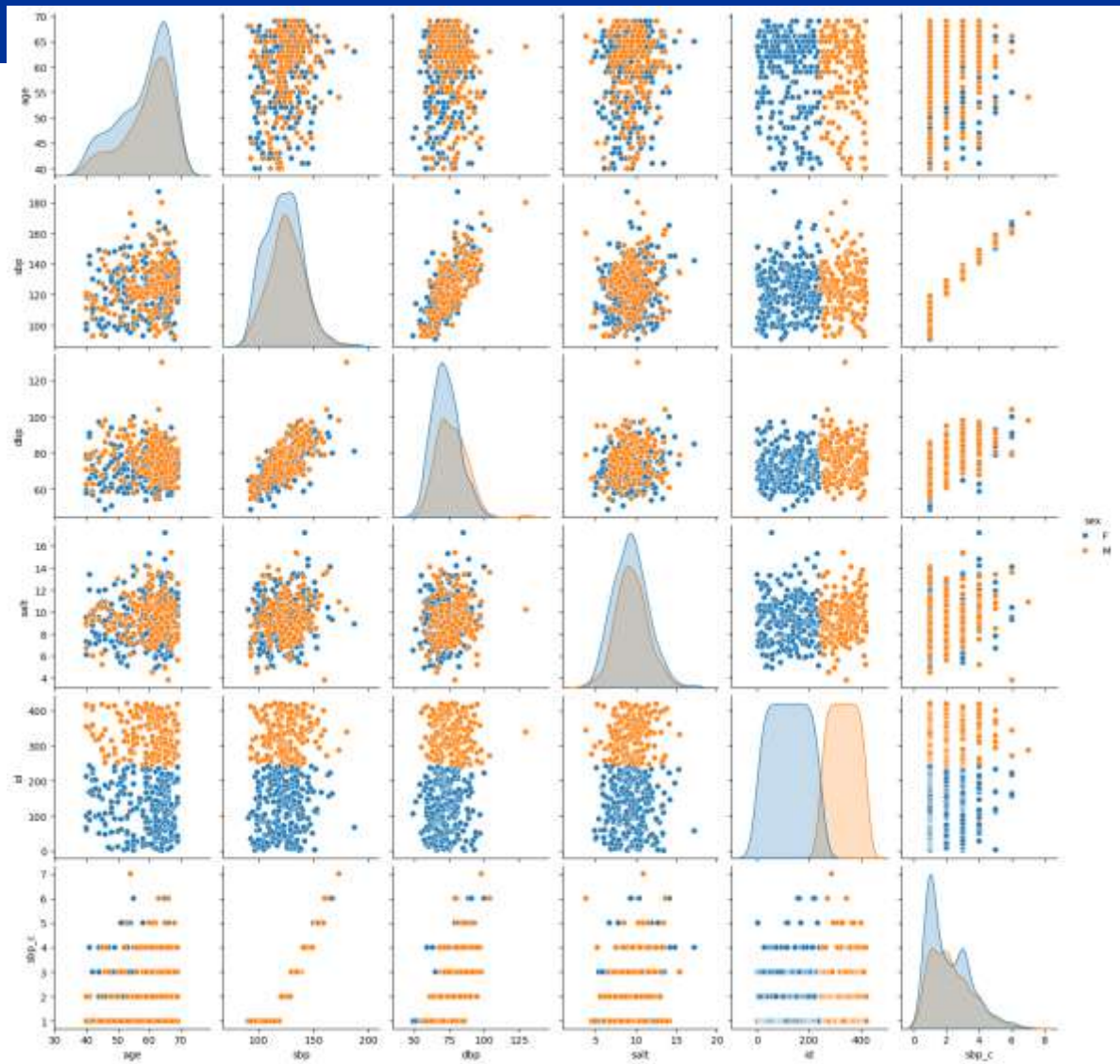
Correlation

```
import pandas as pd
df_subset = pd.concat([df['sbp'],df['salt'],df['dbp']], axis=1)

corr = df_subset.corr()
corr
```

	sbp	salt	dbp
sbp	1.000000	0.177644	0.718942
salt	0.177644	1.000000	0.153706
dbp	0.718942	0.153706	1.000000

```
from scipy.stats import spearmanr
correlation, pvalue = spearmanr(df.sbp, df.dbp)
print(correlation)
```

```
0.719316718120455
```

Seaborn’s pair plot

```
pg2=sns.pairplot(df,hue='sex')
plt.show(pg2)


SBP_F = df[df["sex"]=="F"]["sbp"]
SBP_M = df[df["sex"]=="M"]["sbp"]
DBP_F = df[df["sex"]=="F"]["dbp"]
DBP_M = df[df["sex"]=="M"]["dbp"]
```