



2 1 組のデータの解析

2.2 データのグラフ表示と外れ値

テキスト

芳賀敏郎（2011）医薬品開発のための統計解析

第1部 基礎 改訂版、サイエンティスト社、p.275



第1部 基礎

- 1. 統計の基礎
 - 1.1 宝くじの期待値と分散、1.2 サイコロの目の数の期待値と分散
 - 1.3 分散の加法性・中心極限定理・正規分布、1.4 統計的推測、1.5 モデル
- 2. **1組のデータの解析**
 - 2.1 データの特徴の記述、**2.2 データのグラフ表示と外れ値**
 - 2.3 対数変換と対数正規分布、2.4 平均に関する推測（母標準偏差 σ 既知）
 - 2.5 分散に関する推測、2.6 平均に関する推測（母標準偏差 σ 未知）
- 3. **2組のデータの解析**
 - 3.1 データのグラフ化、3.2 平均値の差の t 検定、3.3 分散の違いの検定
 - 3.4 分散が異なる場合の平均値の差の比較
 - 3.5 対応のある場合の平均値の差の t 検定、3.6 検出力と n の決め方
 - 3.7 ノンパラメトリック検定
- 4. 相関・回帰
 - 4.1 散布図、4.2 相関係数、4.3 回帰モデルとモデルの推定
 - 4.4 誤差を考慮した推定、4.5 回帰分析適用上の諸問題



2.2 データのグラフ表示と外れ値

p.70

- (1) Excelによる度数表とヒストグラムの作成
- (2) JMPによるヒストグラム
- (3) 箱ひげ図
- (4) 外れ値の棄却検定と外れ値の表示

テキストの
該当ページ

使用するファイル

Excel ファイル「基本改2.xls」

JMP ファイル「2-演習.jmp」 「基礎改2.xls」

サイエンティスト社ホームページからダウンロード

JMP 10.0.2 を使用

★プレゼンテーションの
スピーカーノートを、
PDF の注釈に変換してあります



●データのグラフ化

基本統計量の意味を理解するためには、データのグラフ化が有効

データの取得後、基本統計量の計算に先立って、グラフ化してデータを丁寧に観察

JMPの「一変量の分布」「二変量の関係」では、グラフ表示が優先

→ これを見ながら解析を進める

●1組のデータ（連続変数）のグラフ化

度数表 → ヒストグラム

四分位値 → 箱ひげ図



(1) Excel による度数表とヒストグラムの作成

ヒストグラムの作成

(1) Excel による度数表とヒストグラムの作成

Excel の FREQUENCY 関数の利用 (学習を目的とした手作業)

Excel のグラフ作成機能による自動作成

(2) JMP によるヒストグラム

JMP による自動作成

●データ

Excel ファイル「基礎改2.xls」を読み込み、シート「§2.1 演習データ」を開く
40人の「血糖」のデータについて、
どのような性質をもっているか調べるためにヒストグラムを作成（課題 2.2）

No.	年齢	身長	体重	圧(上)	圧(下)	TC	HDL	性脂肪	TN	血糖	ln(TC)	ln(HDL)	ln(脂肪)
1	40	169	75	126	78	200	37	141	7.5	95	5.30	3.61	4.95
2	41	164	75	104	70	197	36	155	7.4	88	5.28	3.58	5.04
3	42	161	68	152	92	300	28	580	8.3	117	5.70	3.33	6.36
4	43	169	73	120	92	199	52	96	6.7	83	5.29	3.95	4.56
5	43	177	63	112	64	183	52	95	7.4	89	5.21	3.95	4.55
6	43	166	57	118	82	224	67	113	7.3	94	5.41	4.20	4.73
7	43	170	65	118	72	194	47	183	6.9	83	5.27	3.85	5.21
8	44	163	60	130	88	171	59	109	7.4	80	5.14	4.08	4.69
9	44	157	62	114	76	154	38	161	7.5	93	5.04	3.64	5.08
10	46	166	67	150	98	213	63	169	7.3	103	5.36	4.14	5.13

●ヒストグラムの作成手順

データから度数表を作成し、これを柱状グラフにする

(1) データ

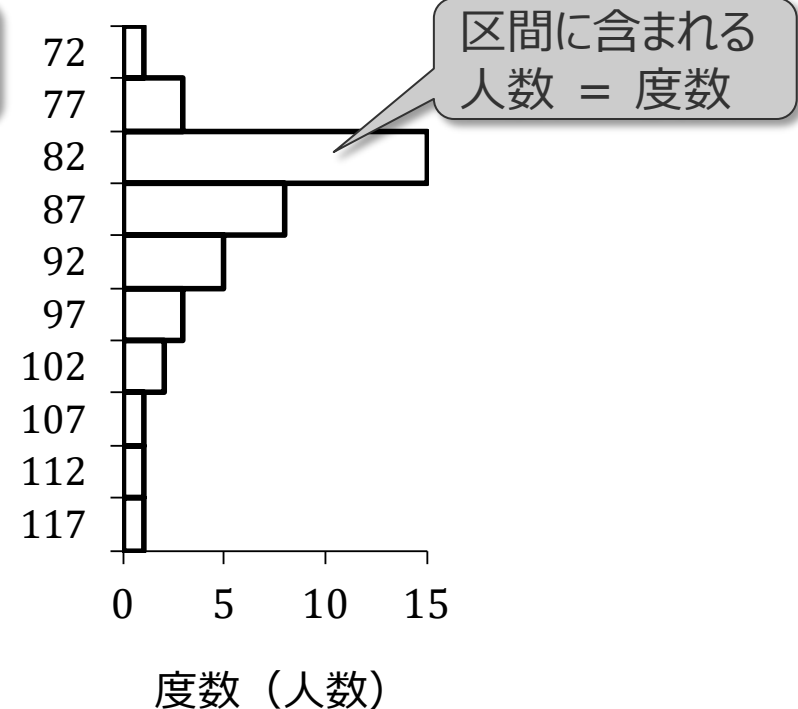
番号	血糖
1	95
2	88
3	117
4	83
・	・
・	・
・	・
・	・
38	79
39	80
40	112

(2) 度数表 (度数分布表)

階級	区間	度数 (人数)
	70 ~ 75	1
	75 ~ 80	7
	80 ~ 85	13
	85 ~ 90	8
	90 ~ 95	5
	95 ~ 100	2
	100 ~ 105	1
	105 ~ 110	1
	110 ~ 115	1
	115 ~ 120	1
	計	40

区間に含まれる
人数 = 度数

(3) ヒストグラム



●ヒストグラムの作成手順：区間の幅と境界値

最小値は 73、最大値は 117、その差は 44

10 前後の区間を設定すると、区間幅は $(117 - 73) / 10 = 4.4 \approx 5$

番号	血糖
1	95
2	88
3	117
4	83
⋮	⋮
⋮	⋮
⋮	⋮
⋮	⋮
⋮	⋮
38	79
39	80
40	112

区間	度数 (人数)
70 ~ 75	1
75 ~ 80	7
80 ~ 85	13
85 ~ 90	8
90 ~ 95	5
95 ~ 100	2
100 ~ 105	1
105 ~ 110	1
110 ~ 115	1
115 ~ 120	1
計	40

初期値 : 70

間隔幅 : 5

90 以上, 95 未満

95 はこちらの区間

境界値

表示 2.1.6 JMPによる
基本統計量の計算 (p.67)

分位点		
100.0%	最大值	117
99.5%		117
97.5%		116.875
90.0%		102.7
75.0%	四分位点	93.75
50.0%	中央値	85
25.0%	四分位点	82
10.0%		79.1
2.5%		73.15
0.5%		73
0.0%	最小値	73

●Excel の FREQUENCY 関数

Excel の FREQUENCY 関数を用いて度数表を作成

番号	血糖	区間	度数 (人数)
1	95	70 ~ 75	1
2	88	75 ~ 80	7
3	117	80 ~ 85	13
4	83	85 ~ 90	8
・	・	90 ~ 95	5
・	・	95 ~ 100	2
・	・	100 ~ 105	1
・	・	105 ~ 110	1
38	79	110 ~ 115	1
39	80	115 ~ 120	1
40	112	計	40

=FREQUENCY(データのセル範囲, 区間の上限値)

Excel の FREQUENCY 関数で
第 1 引数にデータのセル範囲、
第 2 引数に区間の上限値を設定
度数を得る

Excel の FREQUENCY 関数によるヒストグラム

●Excel の FREQUENCY 関数

(1) FREQUENCY(セル範囲, セル)

通常の間数として利用

値

(2) FREQUENCY(セル範囲, セル範囲)

(2-a) 静的配列数式として利用

関数の入力範囲をあらかじめ指定して、
計算式を Ctrl + Shift + Enter キーで確定
(古いタイプ)

(2-b) 動的配列数式として利用

関数の入力範囲のトップのセル (T23) に計算式を入力
スピル機能で範囲全体に自動入力
(新しいタイプ [ブログ](#))
この方法を優先的に使用

(1) FREQUENCY を通常の間数として利用

	R	S	T	U
2	間隔	5		
3	上限値	累積度数	度数	=FREQUENCY(K\$3:K\$42, R4)
4	74	1	1	=FREQUENCY(K\$3:K\$42, R5)
5	79	4	3	=FREQUENCY(K\$3:K\$42, R6)
6	84	19	15	
7	89	27	8	
8	94	32	5	
9	99	35	3	
10	104	37	2	
11	109	38	1	
12	114	39	1	
13	119	40	1	

	R	S	T
22	上限値	代表値	度数
23	74	72	1
24	79	77	3
25	84	82	15
26	89	87	8
27	94	92	5
28	99	97	3
29	104	102	2
30	109	107	1
31	114	112	1
32	119	117	1
33			0

(2) 配列数式として利用

Excel の FREQUENCY 関数によるヒストグラム

●Excel の FREQUENCY 関数

(1) FREQUENCY(セル範囲, セル)

通常の間数として利用

値

(2) FREQUENCY(セル範囲, セル範囲)

(2-a) 静的配列数式として利用

関数の入力範囲をあらかじめ指定して、
計算式を Ctrl + Shift + Enter キーで確定
(古いタイプ)

配列

(2-b) 動的配列数式として利用

関数の入力範囲のトップのセル T23 に計算式を入力

スピル機能で範囲全体に自動入力

(新しいタイプ [ブログ](#))

この方法を優先的に使用

(1) FREQUENCY を通常の間数として利用

	R	S	T	U
2	間隔	5		
3	上限値	累積度数	度数	
4	74	1	1	*
5	79	4	3	***
6	84	19	15	*****
7	89	27	8	*****
8	94	32	5	*****
9	99	35	3	***
10	104	37	2	**
11	109	38	1	*
12	114	39	1	*
13	119	40	1	*

=FREQUENCY(K\$3:K\$42, R4)

=FREQUENCY(K\$3:K\$42, R5)

=FREQUENCY(K\$3:K\$42, R6)

	R	S	T
22	上限値	代表値	度数
23	74	72	1
24	79	77	3
25	84	82	15
26	89	87	8
27	94	92	5
28	99	97	3
29	104	102	2
30	109	107	1
31	114	112	1
32	119	117	1
33			0

関数の入力範囲
=FREQUENCY(K\$3:K\$42, R23:R33)

(2) 配列数式として利用

●通常の間数で利用 (1)

Excel ファイル「基礎改2.xls」
名前ボックスから「表示2.2.1」
(Fig22_01) を選択

シートを複製、そちらで操作

表示 2.2.1 Excel による度数表とヒストグラム (1)

	R	S	T	U	V	W
2	間隔	5				
3	上限値	累積度数	度数	ヒストグラム	*=3	
4	74	1	1	*	1	
5	79	4	3	***	*	
6	84	19	15	*****	*****	
7	89	27	8	*****	**2	
8	94	32	5	*****	*2	
9	99	35	3	***	*	
10	104	37	2	**	2	
11	109	38	1	*	1	
12	114	39	1	*	1	
13	119	40	1	*	1	
14	R5: =R4 + \$S\$2					
15	S5: =FREQUENCY(K\$3:K\$42, R5)					
16	T5: =S5 - S4					
17	U5: = " " & REPT("*", T5)					
18	V5: = " " & REPT("*", T5/3) & IF(MOD(T5,3)=0, "", MOD(T5,3))					

Excel の FREQUENCY 関数によるヒストグラム

●通常の間数で利用 (1)

データ : K列 k3:k42

40人の血糖値, 整数

区間 : R列, 上限値

区間幅 : セルS2

累積度数 : S列

度数 : T列

ヒストグラム : U列

表示 2.2.1 Excel による度数表とヒストグラム (1)

	K	R	S	T	U	V	W	
2	血糖	間隔	5					
3	95	上限値	累積度数	度数	ヒストグラム	*=3		
4	88	74	1	1	*	1		
5	.	79	4	3	***	*		
6	.	84	19	15	*****	*****		
7	.	89	27	8	*****	**2		
8	.	94	32	5	*****	*2		
9	95	99	35	3	***	*		
10	83	104	37	2	**	2		
11	79	109	38	1	*	1		
12	79	114	39	1	*	1		
13	80	119	40	1	*	1		
14	112	R5: =R4 + \$S\$2						
15		S5: =FREQUENCY(K\$3:K\$42, R5)						
16		T5: =S5 - S4						
17		U5: =" "&REPT(""*,T5)						
18		V5: =" "&REPT(""*,T5/3)&IF(MOD(T5,3)=0,"",MOD(T5,3))						

K3:K42 データ
40人の血糖の値、整数

Excel の FREQUENCY 関数によるヒストグラム

●通常の間数で利用 (1)

データ : K列 k3:k42

40人の血糖値, 整数

区間 : R列, 上限値

区間幅 : セルS2

累積度数 : S列

度数 : T列

ヒストグラム : U列

表示 2.2.1 Excel による度数表とヒストグラム (1)

	K	R	S	T	U	V	W
2	血糖	間隔	5				
3	95	上限値	累積度数	度数	ヒストグラム	*=3	
4	88	74	1	1	*	1	
	.	79	4	3	***	*	
	.	84	19	15	*****	*****	
	.	89	27	8	*****	**2	
	.	94	32	5	*****	*2	
37	95	99	35	3	***	*	
38	83	104	37	2	**	2	
39	79	109	38	1	*	1	
40	79	114	39	1	*	1	
41	80	119	40	1	*	1	
42	112						
		R5: =R4 + \$S\$2					
		S5: =FREQUENCY(K\$3:K\$42, R5)					
		T5: =S5 - S4					
		U5: =" "&REPT("*",T5)					
		V5: =" "&REPT("*",T5/3)&IF(MOD(T5,3)=0,"",MOD(T5,3))					

K3:K42 データ
40人の血糖の値、整数

度数の分の「*」を表示

Excel の FREQUENCY 関数によるヒストグラム

● 累積度数と度数の計算 (1)

累積度数の計算

FREQUENCY 関数が、
上限値以下のデータをカウント
\$ マークによる絶対参照

上限値の表示

セル S2 で区間幅を設定
区間幅の変更が容易
1つ上のセル間隔 \$S\$2 を加算
\$ マークによる絶対参照と
オートフィルによるコピー
(Excel を使いこなすための
ポイント [§1.1](#))

表示 2.2.1 Excel による度数表とヒストグラム (1)

	R	S	T	U	V	W
2	間隔	5				
3	上限値	累積度数	度数	ヒストグラム	*=3	セルS2 区間幅
4	74	1	1	*	1	
5	79	4	2	***		
6	84	19	15	*****		
7	89	27	8	*****		
8	94	32	5	*****		
9	99	35	3	***	*	
10	104	37	2	**	2	
11	109	38	1	*	1	
12	114	39	1	*	1	
13	119	40	1	*	1	
14	R5: =R4 + \$S\$2					
15	S5: =FREQUENCY(K\$3:K\$42, R5)					
16	T5: =S5 - S4					
17	U5: =" "&REPT("*",T5)					
18	V5: =" "&REPT("*",T5/3)&IF(MOD(T5,3)=0,"",MOD(T5,3))					

初期値

セル S5
=FREQUENCY (セル範囲, 上限値)
上限値以下のデータ数を返す

\$ マークを使って、
コピーしたときに
変化しないようにする

上限値

Excel の FREQUENCY 関数によるヒストグラム

●累積度数と度数の計算 (1)

FREQUENCY 関数は、
上限値以下の数をカウントする

↓

上限値未満の数をカウントする
ためには工夫が必要

カウントする対象が整数なので

R4 : $x \leq 74 \rightarrow x < 75$ に読替

R5 : $x \leq 79 \rightarrow x < 80$ に読替

T5 : 「74 超、79 以下」
→ 「75以上、80未満」

表示 2.2.1 Excel による度数表とヒストグラム (1)

	R	S	T	U	V	W
2	間隔	5				
3	上限値	累積度数	度数			
4	74	1	1	*	1	
5	79	4	3			
6	84	19	15	*****	*****	
7	89	27	8	***		
8	94	32	5	***		
9	99	35	3	***	*	
10	104	37	2	**	2	
11	109	38	1	*	1	
12	114	39	1	*	1	
13	119	40	1	*	1	
14	R5: =R4 + \$S\$2					
15	S5: =FREQUENCY(K\$3:K\$42, R5)					
16	T5: =S5 - S4					
17	U5: =" "&REPT("*",T5)					
18	V5: =" "&REPT("*",T5/3)&IF(MOD(T5,3)=0,"",MOD(T5,3))					

セル S4 =FREQUENCY(K\$3:K\$42, R4)
 $x \leq 74$ のデータ数は 1 個 → $x < 75$

セル T5 =S5 - S4 = 3 $75 \leq x < 80$

セル S5 =FREQUENCY(K\$3:K\$42, R5)
 $x \leq 79$ のデータ数は 4 個 → $x < 80$

Excel の FREQUENCY 関数によるヒストグラム

●累積度数と度数の計算 (1)

FREQUENCY 関数は、

上限値以下の数をカウントする

↓

上限値未満の数をカウントするためには工夫が必要

カウントする対象が整数なので

R4 : $x \leq 74 \rightarrow x < 75$ に読替

R5 : $x \leq 79 \rightarrow x < 80$ に読替

T5 : 「74 超、79 以下」

→ 「75以上、80未満」

対象が小数第 1 位までである場合

R4 : $x \leq 74.9 \rightarrow x < 75$ に読替

R5 : $x \leq 79.9 \rightarrow x < 80$ に読替

表示 2.2.1 Excel による度数表とヒストグラム (1)

	R	S	T	U	V	W
2	間隔	5				
3	上限値	累積度数	度数			
4	74	1	1	*	1	
5	79	4	3	***		
6	84	19	15	*****		
7	89	27	8	***		
8	94	32	5	***		
9	99	35	3	***	*	
10	104	37	2	**	2	
11	109	38	1	*	1	
12	114	39	1	*	1	
13	119	40	1	*	1	
14	R5:	=R4 + \$S\$2				
15	S5:	=FREQUENCY(K\$3:K\$42, R5)				
16	T5:	=S5 - S4				
17	U5:	= " " & REPT(" ", T5)				
18	V5:	= " " & REPT(" ", T5/3) & IF(MOD(T5,3)=0, "", MOD(T5,3))				

セル S4 =FREQUENCY(K\$3:K\$42, R4)
 $x \leq 74$ のデータ数は 1 個 $\rightarrow x < 75$

セル T5 =S5 - S4 = 3 $75 \leq x < 80$

セル S5 =FREQUENCY(K\$3:K\$42, R5)
 $x \leq 79$ のデータ数は 4 個 $\rightarrow x < 80$

度数の合計、確認

Excel の FREQUENCY 関数によるヒストグラム

●ヒストグラムの作成 (1)

度数の数の「*」を並べて、
その前に「|」を付加

「&」は前後の文字列を連結する
演算子

REPT 関数は、第 1 引数の文字を
第 2 引数で指定する回数だけ
繰り返して表示

表示 2.2.1 Excel による度数表とヒストグラム (1)

	R	S	T	U	V	W
2	間隔	5				
3	上限値	累積度数	度数	ヒストグラム	*=3	
4	74	1	1	*	1	
5	79	4	3	***	*	
6	84	19	15	*****	*****	
7	89	27	8	*****	**2	
8	94	32	5	*****	*2	
9	99	35	3	***	*	
10	104	37	2	**	2	
11	109	38	1	*	1	
12	114	39	1	*	1	
13	119	40	1	*	1	
14	R5: =R4 + \$S\$2					
15	S5: =FREQUENCY(K\$3:K\$42, R5)					
16	T5: =S5 - S4					
17	U5: =" "&REPT("*",T5)					
18	V5: =" "&REPT("*",T5/3)&IF(MOD(T5,3)=0,"",MOD(T5,3))					

Excel の FREQUENCY 関数によるヒストグラム(1)

●ヒストグラムの作成 (1)

短縮したヒストグラム

度数 3 個分を「*」 1 つで表示

余りを数値で表示

条件分岐の関数 IF

IF (条件, 真の場合, 偽の場合)

除算の演算子 MOD

MOD は除算の剰余を返す

MOD(3, 3) → 0 (3/3)

MOD(8, 3) → 2 (8/3)

ヒストグラムの完成 (体験)

表示 2.2.1 Excel による度数表とヒストグラム (1)

	R	S	T	U	V	W
2	間隔	5				
3	上限値	累積度数	度数	ヒストグラム	*=3	
4	74	1	1	*	1	(3 × 1) = 3
5	79	4	3	***	*	
6	84	19	15	*****	*****	
7	89	27	8	*****	**2	(3 × 2) + 2 = 8
8	94	32	5	*****	*2	
9	99	35	3	***	*	
10	104	37	2	**	2	
11	109	38	1	*	1	
12	114	39	1	*	1	
13	119	40	1	*	1	
14	R5: =R4 + \$S\$2					
15	S5: =FREQUENCY(K\$3:K\$42, R5)					
16	T5: =S5 - S4					
17	U5: =" "&REPT("*",T5)					
18	V5: =" "&REPT("*",T5/3)&IF(MOD(T5,3)=0,"",MOD(T5,3))					

T5 / 3
余りは切り捨て

Excel の FREQUENCY 関数によるヒストグラム

●静的配列数式による度数の計算 (2-a)

Excel ファイル「基礎改2.xls」
名前ボックスから「表示2.2.2」
(Fig22_02) を選択

	K
2	血糖
3	95
4	88
・	・
・	・
・	・
37	95
38	83
39	79
40	79
41	80
42	112

表示 2.2.2 Excel による度数表とヒストグラム (2)

	R	S	T	U
22	上限値	代表値	度数	
23	74	72	1	
24	79	77	3	
25	84	82	15	
26	89	87	8	
27	94	92	5	
28	99	97	3	
29	104	102	2	
30	109	107	1	
31	114	112	1	
32	119	117	1	
33			0	
34	R23:	74		
35	S23:	=R23-2		
36	R24:	=R23+5		
37	T23:	{=FREQUENCY(K3:K42,R23:R33)}		
38		上限の最大を超える範囲をあらわす		

Excel の FREQUENCY 関数によるヒストグラム

表示 2.2.2 Excel による度数表とヒストグラム (2)

●静的配列数式による度数の計算 (2-a)

データ：K列、40人の血糖値、整数

区間：R列「上限値」

代表値：S列、表示のみ、計算に無関係

度数：T列、FREQUENCY関数が度数を返す

FREQUENCY関数の第2引数に、

上限値のセル範囲を配列として渡すので、累積度数ではなく、度数が戻り値になる

「超、以下」→「以上、未満」に読替

FREQUENCY関数を静的配列数式で利用

セル範囲 T23:T33 に同じ計算式を入力

特殊な入力方法

	K
2	血糖
3	95
4	88
・	・
・	・
・	・
37	95
38	83
39	79
40	79
41	80
42	112

K3:K42 データ

	R	S	T	U
22	上限値	代表値	度数	
23	74	72	1	
24	79	77	5	初期値
25	84	82	15	
26	89	87	8	
27	94	92	5	
28	99	97	3	84 < x ≤ 89 85 ≤ x < 90
29	104	102	2	
30	109	107	1	
31	114	112	1	
32	119	117	1	
33			0	
34	R23: 74			初期値
35	S23: =R23-2			
36	R24: =R23+5			セル範囲 (配列)
37	T23: {=FREQUENCY(K3:K42,R23:R33)}			
38				上限の最大を超える範囲をあらわす

●静的配列数式による度数の計算 (2-a)

FREQUENCY の静的配列数式の入力手順

- (i) 計算式の入力範囲 T23:T33を範囲指定
- (ii) 範囲指定した状態で関数を入力 ({} 不要)
=FREQUENCY(K3:K42, R23:R33)
- (iii) Ctrl + Shift + Enterキーを押して入力確定
(関数の両脇に {} が自動的に付加される)
- (iv) T23:T33 にFREQUENCY 関数が入力される
全てに同じ計算式が入力される
計算式が中括弧 {} で囲まれる
T23:T33 の一部のセルの変更は不可
- (v) セル T33 は確認用
設定を誤ると 0 にはならない
スピルを使った動的配列数式を優先して利用

表示 2.2.2 Excel による度数表とヒストグラム (2)

	R	S	T
22	上限値	代表値	度数
23	74	72	
24	79	77	
25	84	82	
26	89	87	
27	94	92	
28	99	97	
29	104		
30	109		
31	114		
32	119		
33			

	R	S	T	U
22	上限値	代表値	度数	
23	74	72	=FREQUENCY(K3:K42, R23:R33)	
24	79	77		
25	84	82		
26	89	87		
27	94	92		
28	99	97		
29	104	102		
30	109	107		
31	114	112		
32	119	117		
33				

T23:T33 を範囲指定

=FREQUENCY (K3:K42, R23:R33)

=FREQUENCY(K3:K42, R23:R33)

●静的配列数式による度数の計算 (2-a)

FREQUENCY の静的配列数式の入力手順

(i) 計算式の入力範囲 T23:T33を範囲指定

(ii) 範囲指定した状態で関数を入力 ({} 不要)

=FREQUENCY(K3:K42, R23:R33)

(iii) Ctrl + Shift + Enterキーを押して入力確定
(関数の両脇に {} が自動的に付加される)

(iv) T23:T33 にFREQUENCY 関数が入力される

全てに同じ計算式が入力される

計算式が中括弧 {} で囲まれる

T23:T33 の一部のセルの変更は不可

(v) セル T33 は確認用

設定を誤ると 0 にはならない

スピルを使った動的配列数式を優先して利用

表示 2.2.2 Excel による度数表とヒストグラム (2)

	R	S	T	U
22	上限値	代表値	度数	
23	74	72	1	各区間の度数
24	79	77	3	
25	84	82	15	
26	89	87	8	
27	94	92	5	全てのセルに 同じ計算式が入力
28	99	97	3	
29	104	102	2	
30	109	107	1	
31	114	112	1	
32	119	117	1	確認のため
33			0	
34	R23:	74		
35	S23:	=R23-2		{ } が加わる
36	R24:	=R23+5		
37	T23:	={FREQUENCY(K3:K42,R23:R33)}		
38		上限の最大を超える範囲をあらわす		

●動的配列数式による度数の計算 (2-b)

FREQUENCY の動的配列数式の入力手順

(i) 計算式の入力範囲 T23:T33 の
トップのセルを選択

(ii) 計算式を入力

=FREQUENCY(K3:K42, R23:R33)

(iii) Enter キーを押して入力確定

(iv) T23:T33 に FREQUENCY 関数が入力される

全てのセルに同じ計算式が自動的に入力
(スピル機能)

入力されたセルを選択するとブルー枠が表示

T23:T33 の一部のセルの変更は不可

(v) セル T33 は確認のためにある

設定を誤ると 0 にはならない

表示 2.2.2 Excel による度数表とヒストグラム (2)

	R	S	T
22	上限値	代表値	度数
23	74	72	
24	79	77	
25	84	82	
26	89	87	
27	94	92	
28	99	97	
29	104	102	
30	109	107	
31	114	112	
32	119	117	
33			

	R	S	T	U
22	上限値	代表値	度数	
23	74	72	=FREQUENCY(K3:K42,R23:R33)	
24	79	77		
25	84	82		
26	89	87		
27	94	92		
28	99	97		
29	104	102		
30	109	107		
31	114	112		
32	119	117		
33				

計算式の入力範囲の
トップのセルを選択

=FREQUENCY
(K3:K42, R23:R33)

Excel の FREQUENCY 関数によるヒストグラム

●動的配列数式による度数の計算 (2-b)

FREQUENCY の動的配列数式の入力手順

(i) 計算式の入力範囲 T23:T33 の
トップのセルを選択

(ii) 計算式を入力

=FREQUENCY(K3:K42, R23:R33)

(iii) Enter キーを押して入力確定

(iv) T23:T33 に FREQUENCY 関数が入力される

全てのセルに同じ計算式が自動的に入力
(スピル機能)

入力されたセルを選択するとブルー枠が表示

T23:T33 の一部のセルの変更は不可

(v) セル T33 は確認のため

設定を誤ると 0 にはならない

表示 2.2.2 Excel による度数表とヒストグラム (2)

	R	S	T	U
22	上限値	代表値	度数	
23	74	72	1	
24	79	77	3	ブルー枠
25	84	82	15	
26	89	87	8	
27	94	92	5	
28	99	97	3	全てのセルに 同じ計算式が入力 一部の変更は不可
29	104	102	2	
30	109	107	1	
31	114	112	1	
32	119	117	1	
33			0	確認のため
34	R23:	74		
35	S23:	=R23-2		{ } は付かない
36	R24:	=R23+5		
37	T23:	=FREQUENCY(K3:K42,R23:R33)		
38		上限の最大を超える範囲をあらわす		

Excel の FREQUENCY 関数によるヒストグラム

●グラフ作成機能（集合縦棒グラフ）

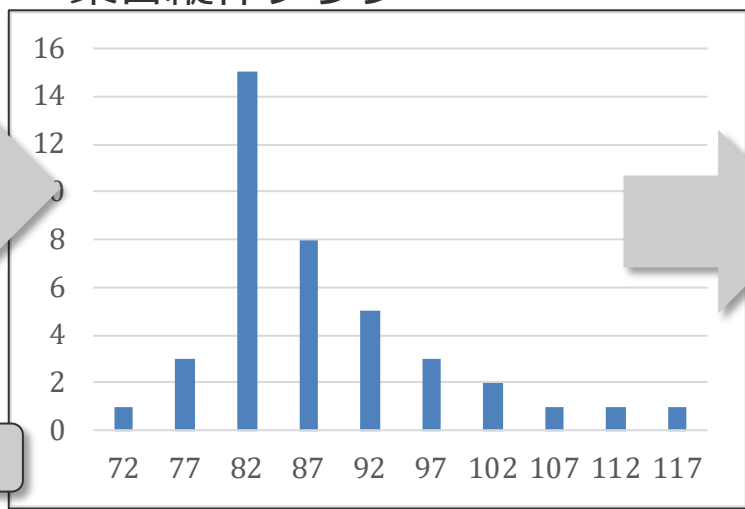
表示 2.2.2

	R	S	T
22	上限値	代表値	度数
23	74	72	1
24	79	77	3
25	84	82	15
26	89	87	8
27	94	92	5
28	99	97	3
29	104	102	2
30	109	107	1
31	114	112	1
32	119	117	1
33			0

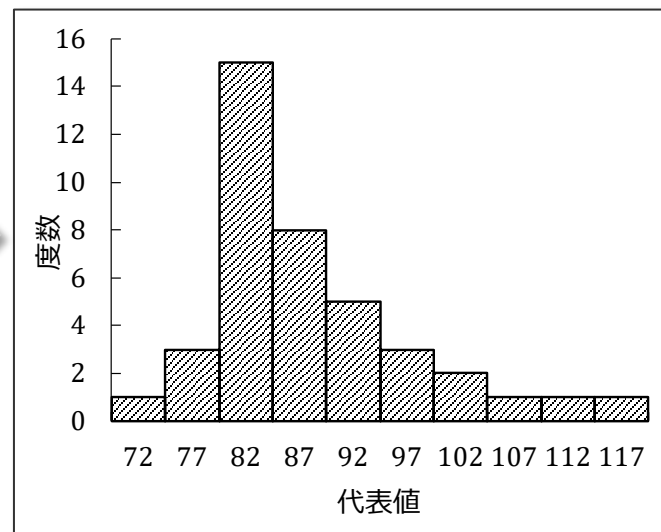
代表値の欄を作成
 $70 \leq x < 75 \rightarrow 72$
度数の欄の左側に作成

代表値と度数の全体を含む S22:T32 を範囲指定
Excelのグラフ機能で「集合縦棒グラフ」を選択
棒グラフからヒストグラムに整形
整形方法は Excel ファイルの記載を参照

集合縦棒グラフ



除外



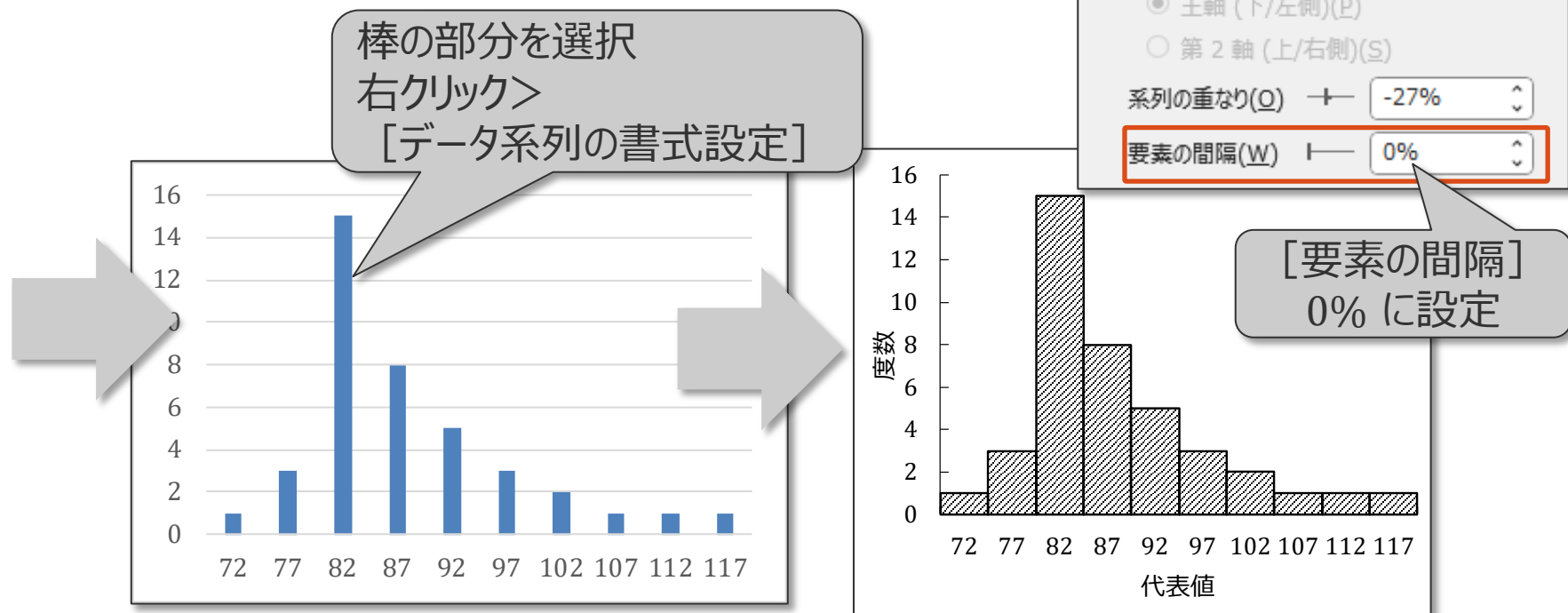
Excel の FREQUENCY 関数によるヒストグラム

●グラフ作成機能（集合縦棒グラフ）

表示 2.2.2

	R	S	T
22	上限値	代表値	度数
23	74	72	1
24	79	77	3
25	84	82	15
26	89	87	8
27	94	92	5
28	99	97	3
29	104	102	2
30	109	107	1
31	114	112	1
32	119	117	1
33			0

ヒストグラムを理解するために、
FREQUENCY 関数を使って手作業で作成
実務では、自動的に作成するツールを利用

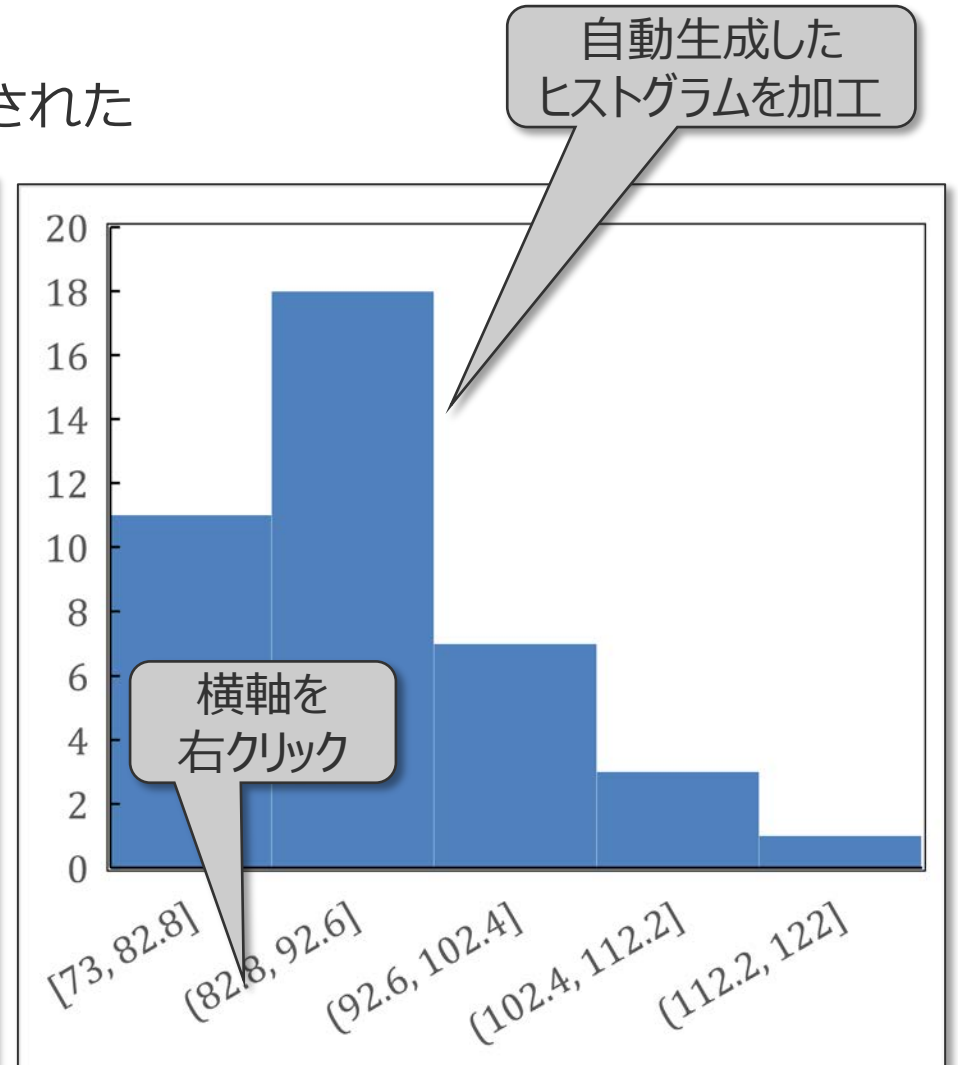
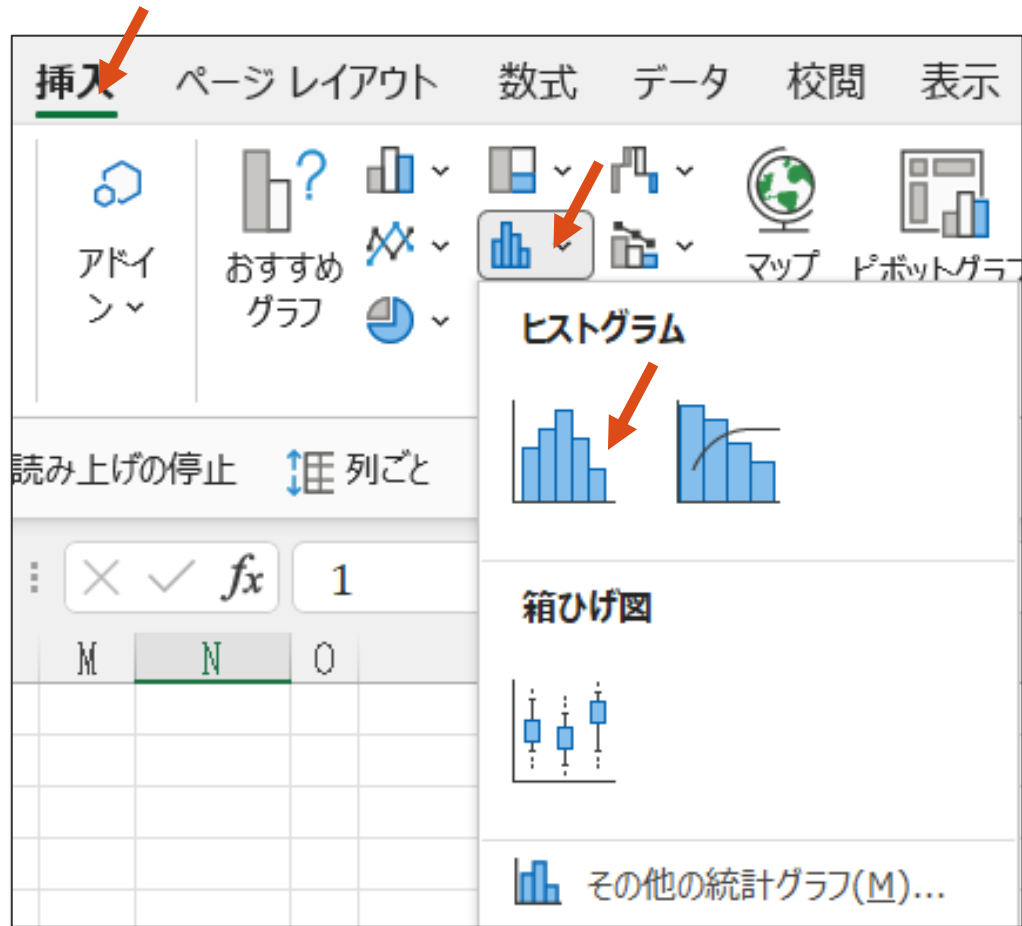


●グラフ作成機能（ヒストグラム）

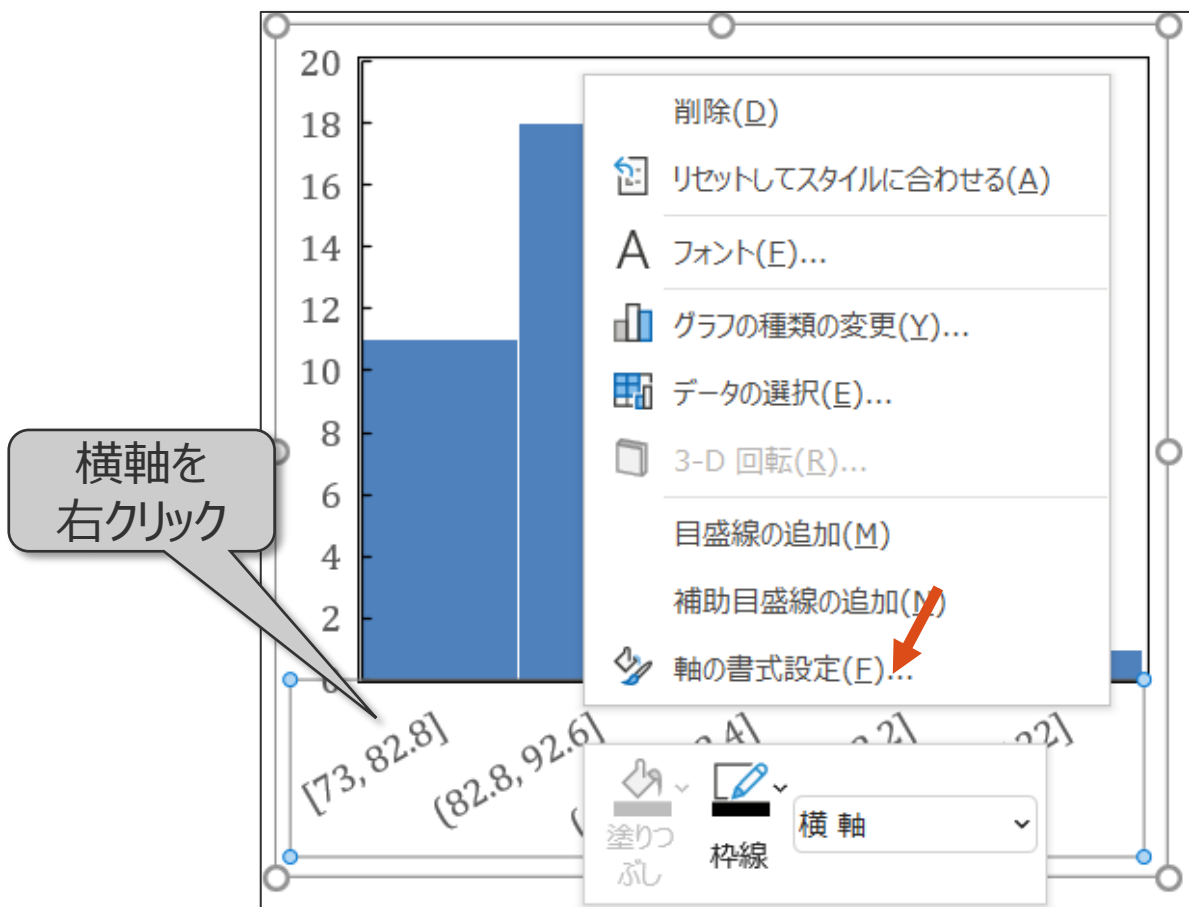
Excel 2016 からヒストグラムを直接作成する機能が追加された

人	血糖
1	95
2	88
3	117
4	83
5	89
6	94
7	83
8	80
9	93
10	103
11	96
12	81
...	...
39	80
40	112

範囲指定



●グラフ作成機能（ヒストグラム）



軸の書式設定

軸のオプション ▼ 文字のオプション

軸のオプション

ピン

- 分類項目別(G)
- 自動(M)
- ピンの幅(B)
- ピンの数(N)
- ピンのオーバーフロー(V) 114.0 リセット
- ピンのアンダーフロー(U) 74.0 リセット

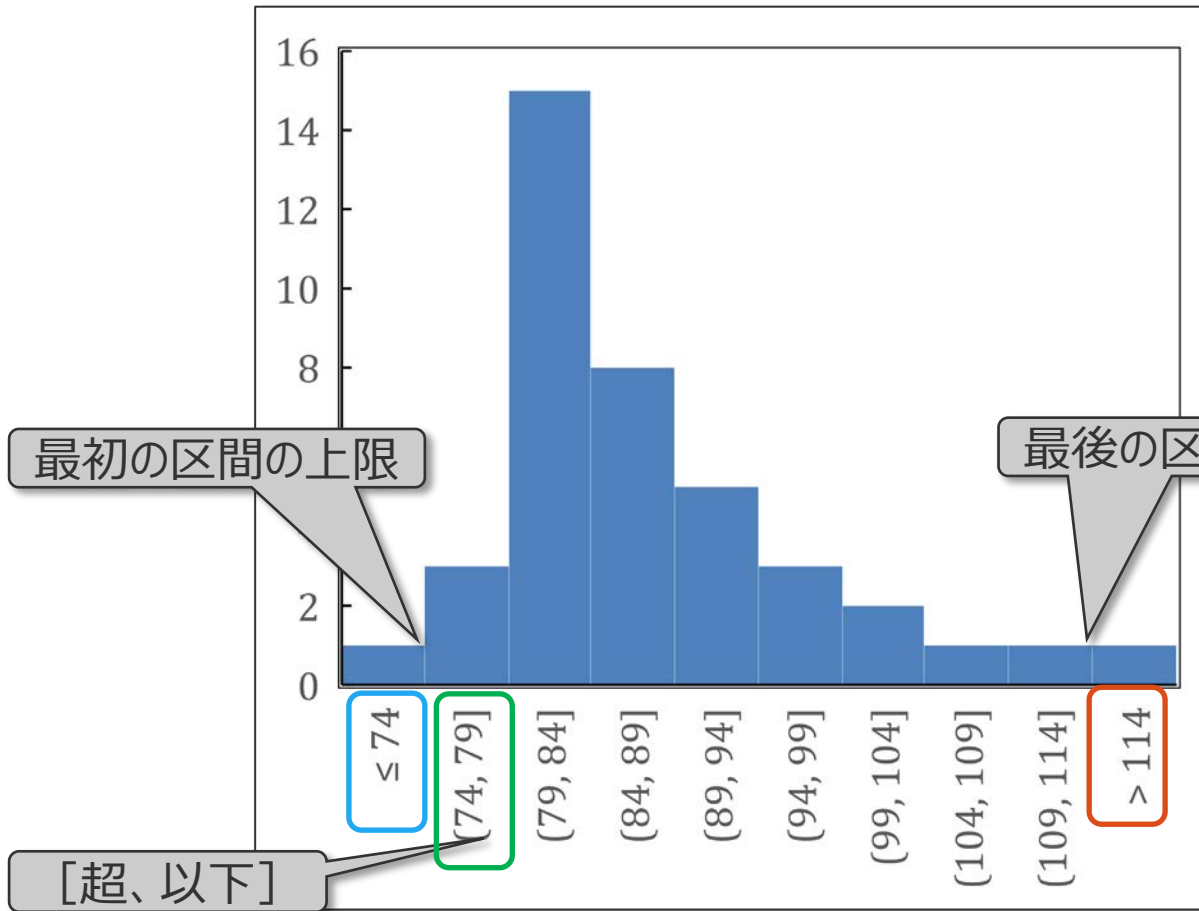
5.0 区間の幅

10 区間の数

Excel のグラフ作成機能によるヒストグラム

●グラフ作成機能（ヒストグラム）

表示2.2.2と同じヒストグラムを得る



軸の書式設定

軸のオプション 文字のオプション

軸のオプション

ピン

分類項目別(G)

- 自動(M)
- ピンの幅(B) 5.0
- ピンの数(N) 10
- ピンのオーバーフロー(V) 114.0 (最後の区間の下限)
- ピンのアンダーフロー(U) 74.0 (最初の区間の上限)

(2) JMPによるヒストグラム



JMP によるヒストグラム

●Excel から JMP へのデータのコピー（Excel での操作）

Excel から JMP にデータをコピー&ペースト、JMP でヒストグラムを作成

Excel と JMP を起動

Excel 「基礎改2.xls」のシート「§1.1 演習データ」を表示（40 人の健康診断のデータ）

データのセル全体を範囲指定（変数名を含む）、指定したセルの上で右クリック> [コピー]

Excel ファイルの作成時の注意

変数名を 1 行に入力（2 行にまたがらない、セルの結合をしない）

セルの中で Alt キー+Enter の改行をしない、分かり易い変数名を命名

全角文字も使用可、長い変数名は使い難い

セル上で
右クリック

列名を含めて
セル全体を
範囲指定

No.	年齢	身長	体重	王(上)	王(下)	TC	HDL	性脂肪	TN	血糖	n(TC)	n(HDL)	n(脂肪)
1	40	169	75	126	78	20					3.01		4.95
2	41	164	75	104	70	19					3.58		5.04
3	42	161	68	152	92	30					3.33		6.36
4	43	169	73	120	92	19					3.95		4.56
5	43	177	63	112	64	18					3.95		4.55
6	43	166	57	118	89	22					4.20		4.79

切り取り(T)

コピー(C)

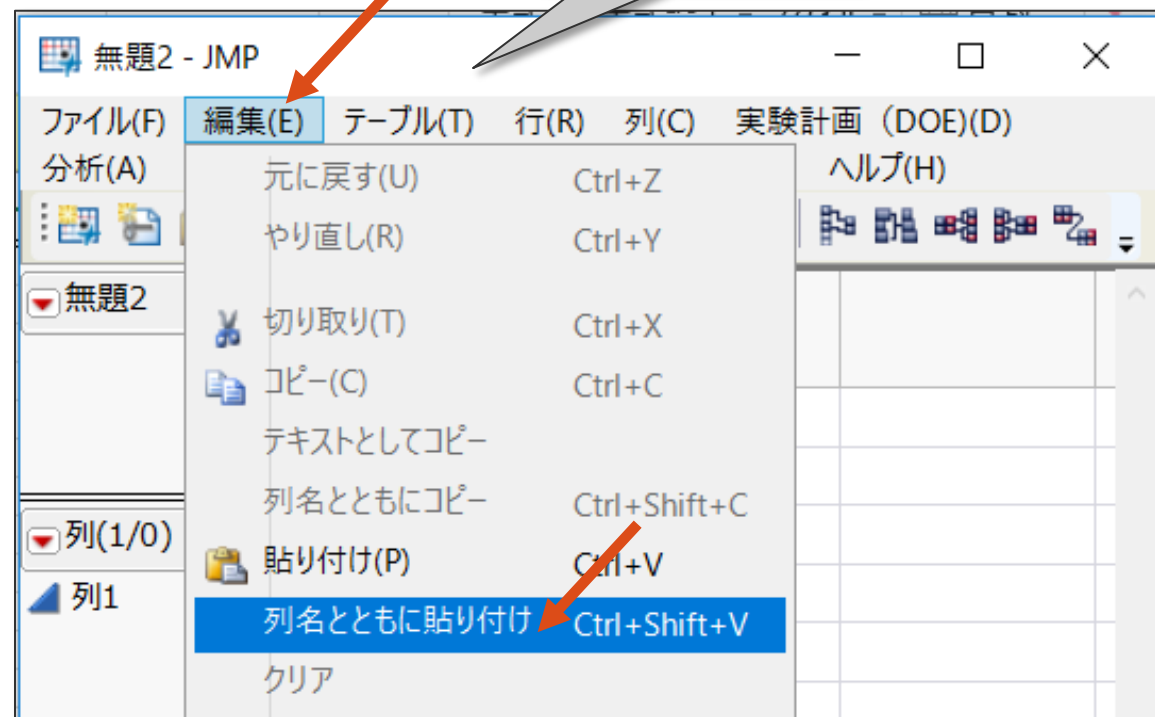
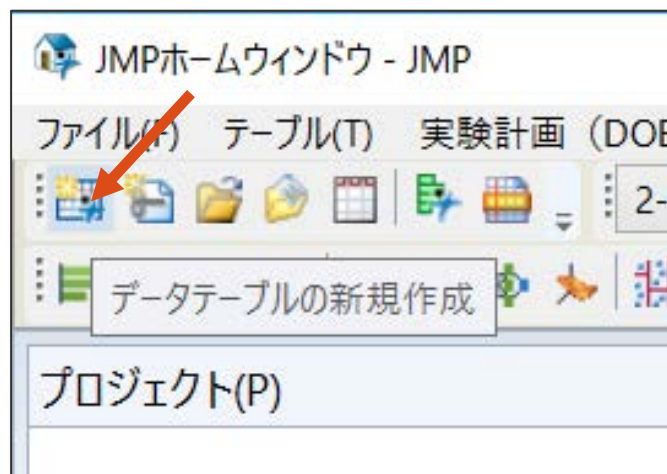
貼り付けのオプション:

貼り付け (P)

貼り付けと形式 (F)

貼り付けと形式と数値 (V)

●Excel から JMP へのデータのコピー (JMPでの操作)



JMP のデータテーブルを新規作成

アイコン [データテーブルの新規作成]

または、トップメニュー [ファイル] > [新規作成] > [データテーブル]

データを JMP に張り付け

新規のデータテーブルで、トップメニュー [編集] > [列名とともに貼り付け]

JMP によるヒストグラム

●Excel から JMP へのデータのコピー (JMPでの操作)

JMP のデータファイルとして保存

トップメニュー [ファイル] > [名前を付けて保存]

The screenshot shows the JMP interface with a data table. The menu bar includes 'ファイル(F)', '編集(E)', 'テーブル(T)', '行(R)', '列(C)', '実験計画 (DOE)', '分析(A)', 'グラフ(G)', 'ツール(O)', '表示(V)', 'ウィンドウ(W)', and 'ヘルプ(H)'. The 'ファイル(F)' menu is open, showing options like '新規作成(N)', '開く(O)...', '閉じる(C)', 'データとして読み込み', '上書き保存(S)', '名前を付けて保存(A)...', and 'データベース(D)'. The '名前を付けて保存(A)...' option is highlighted. The data table contains columns for '血压(下)', 'TC', 'HDL', '中性脂肪', 'TN', and '血糖'. Callouts point to '変数名がコピー' (variable names), 'データがコピー' (data), and '使用するデータ' (data to use). A callout '総窒素量' (Total Nitrogen) points to the 'TN' column. The 'TN' and '血糖' columns are highlighted with a red box.

血压(下)	TC	HDL	中性脂肪	TN	血糖
78	200	37	141	7.5	95
70	197	36	155	7.4	88
92	300	580	580	8.3	117
92	199	52	96	6.7	83
64	183	52	95	7.4	89
82	224	67	113	7.3	94
72	194	47	183	6.9	83

●JMP [一変量の分布] によるヒストグラムの作成

行(R) 列(C) 実験計画 (DOE)(D) 分析(A) グラフ(G) ツール(O)

- 一変量の分布
- 二変量の関係
- 対応のあるペア
- モデルのあてはめ
- モデル化
- 多変量
- 品質と工程
- 信頼性/生存時間分析

	TN	血糖	BI
1	7.5	95	0.002
2	7.4	88	0.002
3	8.3	117	0.002
4	6.7	83	0.002
5	7.4	89	0.002

トップメニュー [分析] > [一変量の分布]

一変量の分布 - JMP

各列の値の分布

列の選択

- No.
- 年齢
- 身長
- 体重
- 血圧(上)
- 血圧(下)
- TC
- HDL
- 中性脂肪
- TN
- 血糖

選択した列に役割を割り当てる

Y, 列

- TN
- 血糖

オプション

重み

オプション(数値)

度数

オプション(数値)

By

オプション

アクション

- OK
- キャンセル
- 削除
- 前回の設定
- ヘルプ

三角のマークは連続尺度に指定されていることを示す
§0.1(2) 参照 (p.2)

Ctrl キーを押しながら
TN と血糖をクリック

JMPによるヒストグラム

●全体の表示の向きを変更

[一変量の分布] の隣の▼> [積み重ねて表示]

表示 2.2.3 JMPによるヒストグラム

一変量の分布

- スケールの統一
- 積み重ねて表示**
- Flash(.SWF)形式で保存...
- スクリプト

血糖

ヒストグラムを回転させ、各変数のレポートを縦に積み重ねる。

箱ひげ図

ヒストグラム

一変量の分布

TN

分位点	値
100.0%	最大値 8.3
99.5%	8.3
97.5%	8.2925
90.0%	7.89
75.0%	四分位点 7.6
50.0%	中央値 7.4
25.0%	四分位点 7.2
10.0%	6.9
2.5%	6.505
0.5%	6.5
0.0%	最小値 6.5

要約統計量	値
平均	7.3725
標準偏差	0.3573316
平均の標準誤差	0.0564991
平均の上側95%	7.4867802
平均の下側95%	7.2582198
N	40

血糖

分位点	値
100.0%	最大値 117
99.5%	117
97.5%	116.875
90.0%	102.7
75.0%	四分位点 93.75
50.0%	中央値 85
25.0%	四分位点 82
10.0%	79.1
2.5%	73.15
0.5%	73
0.0%	最小値 73

要約統計量	値
平均	88.075
標準偏差	9.5283316
平均の標準誤差	1.5065615
平均の上側95%	91.122308
平均の下側95%	85.027692
N	40

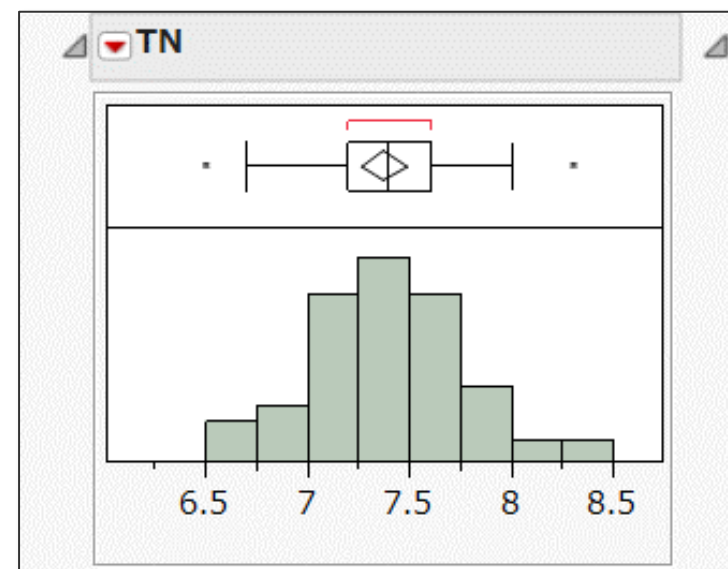
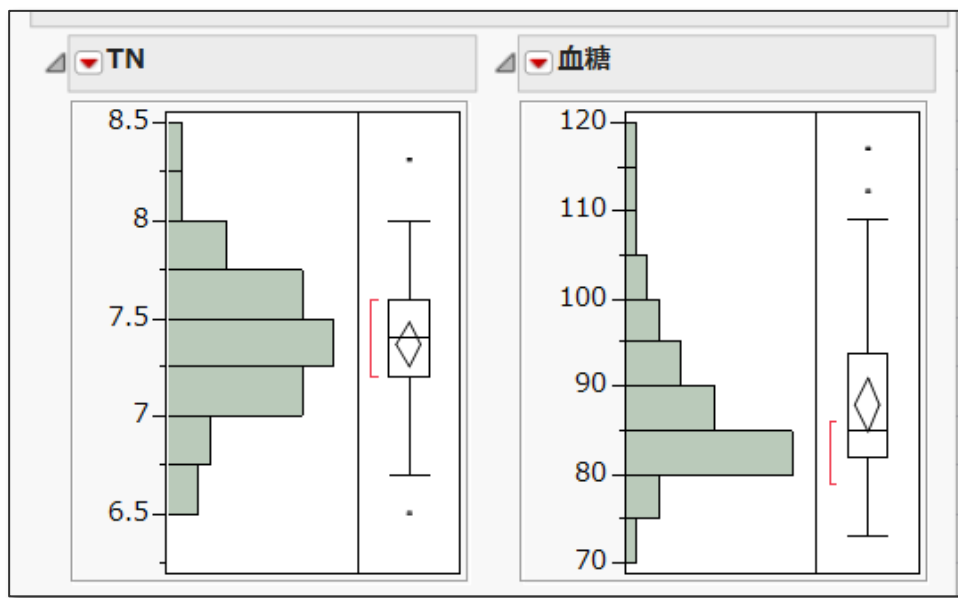
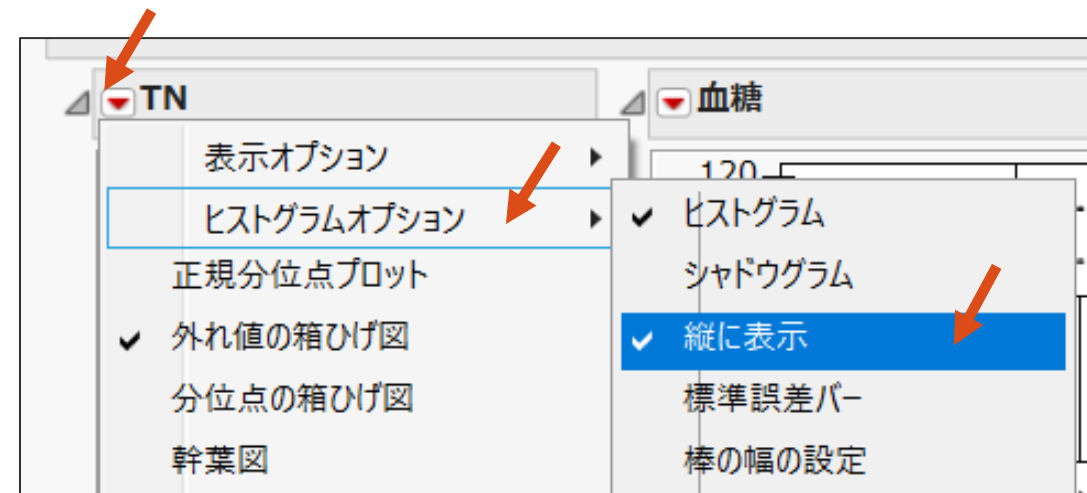
●ヒストグラムの向きを変更

デフォルトでは縦に表示

変数名の隣の▼> [ヒストグラムオプション]

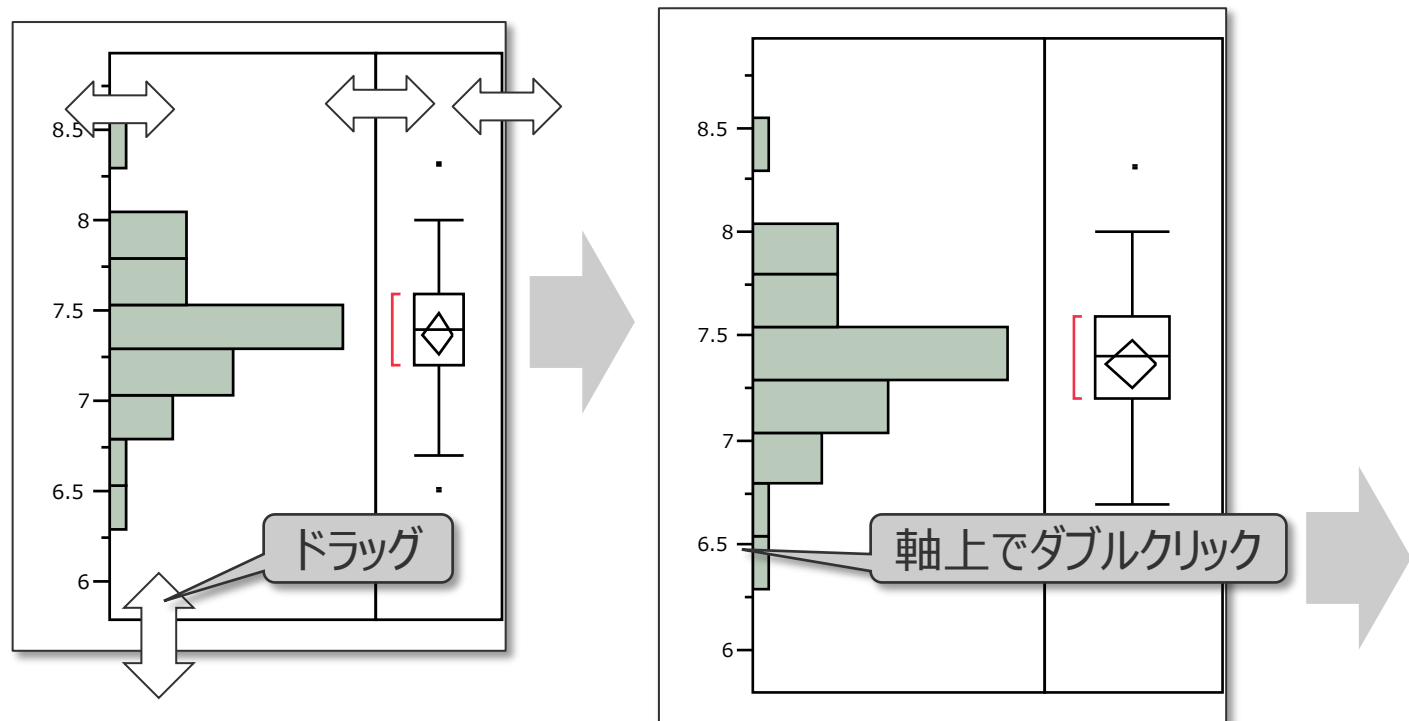
> [縦に表示] (チェックマークが外れる)

▼を押す前からCtrlキーを押しながら作業すると
表示中の全ての変数に同時指定できる



●ヒストグラムの整形

- 軸上にカーソルを置くと両矢印に変化
- ドラッグするとヒストグラムのサイズが変化
- 軸上でダブルクリックするとダイアログが表示
- 最小値、最大値、目盛り間隔などを設定



The screenshot shows the "Y軸の指定" (Specify Y-axis) dialog box. The settings are as follows:

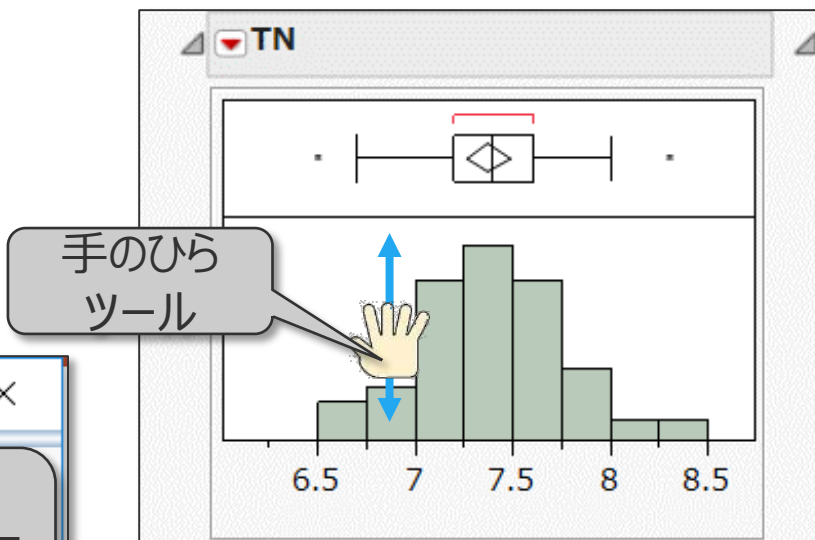
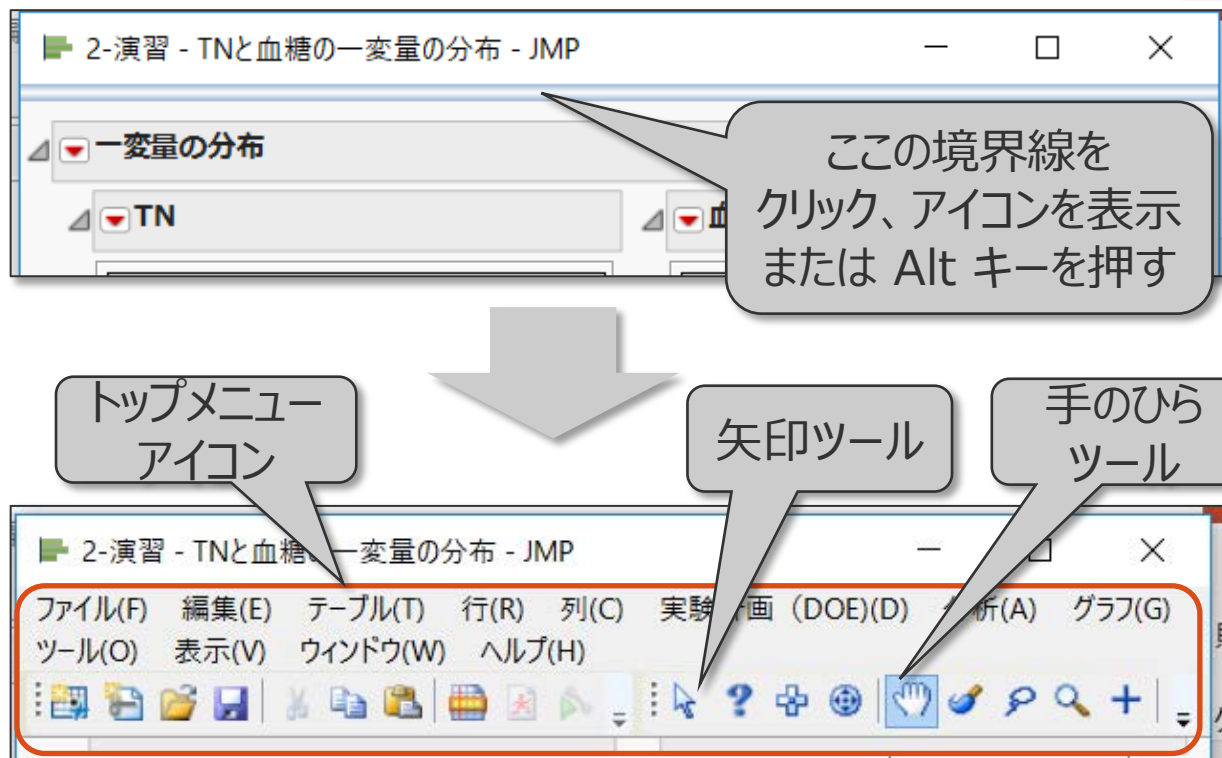
- スケール: 線形 (Linear)
- 形式: 最適 (Optimal), 総桁数: 10
- 最小値: 5.79534363
- 最大値: 8.9258042
- 目盛り間隔: 0.5
- 補助目盛りの数: 1
- 目盛りとグリッド線:
 - 目盛り: 大 (checked), 小 (checked)
 - グリッド線: (unchecked)
 - ラベル: (checked)
 - 目盛りラベルの向き: 横 (Horizontal)
 - 目盛りをグラフフレームの内側に表示: (unchecked)
- 参照線:
 - 値: 7.36057391446075
 - ラベル: (empty)
 - 色: (black)
 - 線種: (solid)

Buttons: OK, キャンセル, ヘルプ

JMP によるヒストグラム

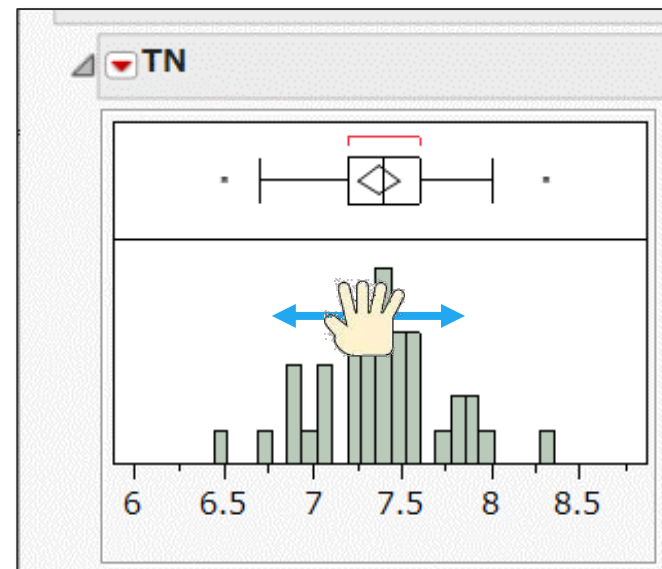
●ヒストグラムの整形

境界値と区間の幅によって形が大きく変化
最初に出力されたヒストグラムだけを見て
結論を下すのは危険



カーソルを
デフォルトの
矢印ツールから
手のひらツールに
変更

柱の上で横軸と
垂直方向に
ドラッグすると
区間幅が変更



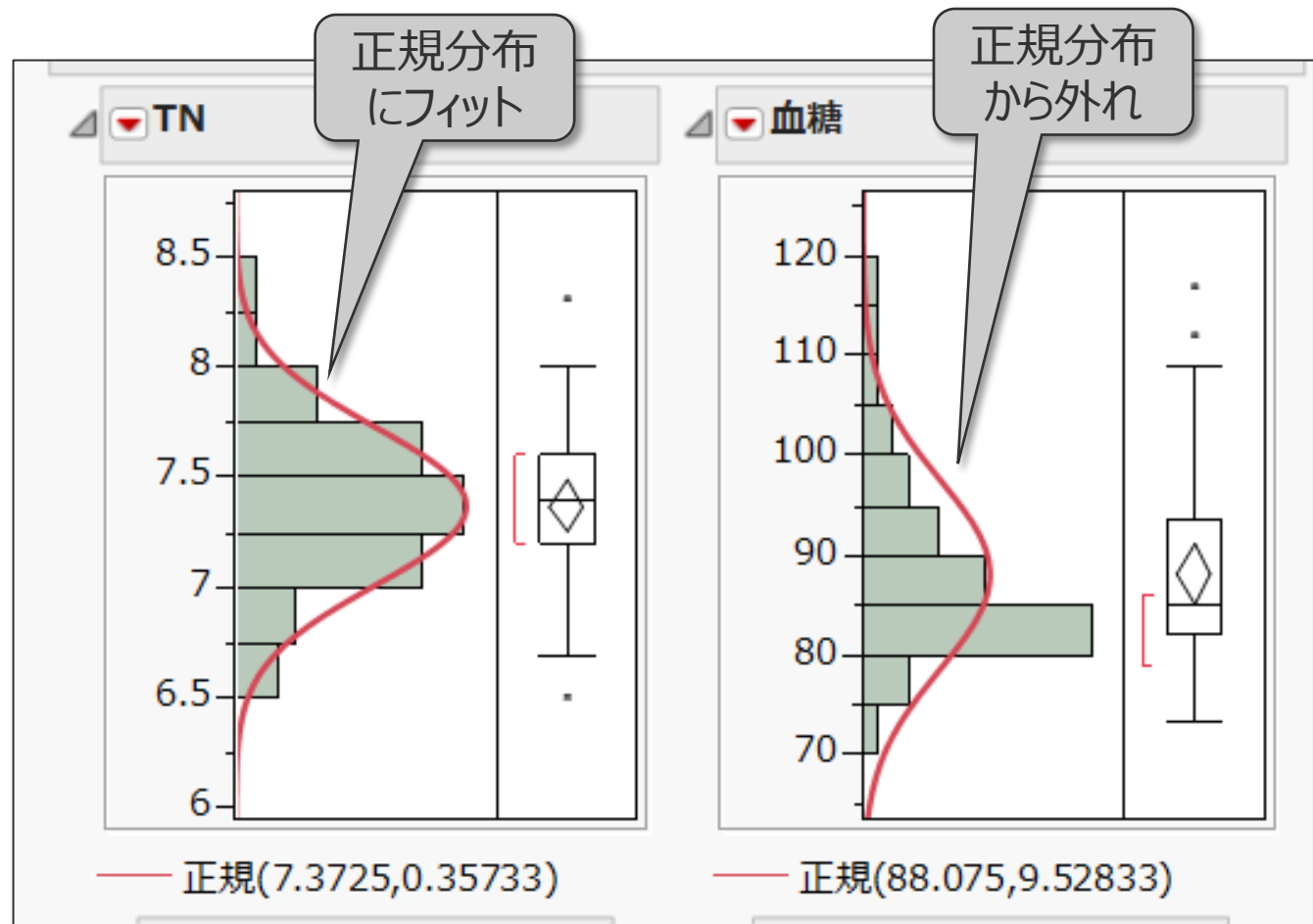
手のひらツールを
横軸と平行に
ドラッグすると
境界が移動

●正規分布のあてはめ



変数名の隣の▼

> [工程能力分析] [連続分布のあてはめ] > [正規]



●ヒストグラムの見方

分布の山

山が複数ある場合、母集団が混合している可能性あり
最初のヒストグラムだけの判断は危険 → 区間を変えて観察

TN(1)、TN(2)、TN(3)
区間の幅を変えると、
分布の見え方が違う

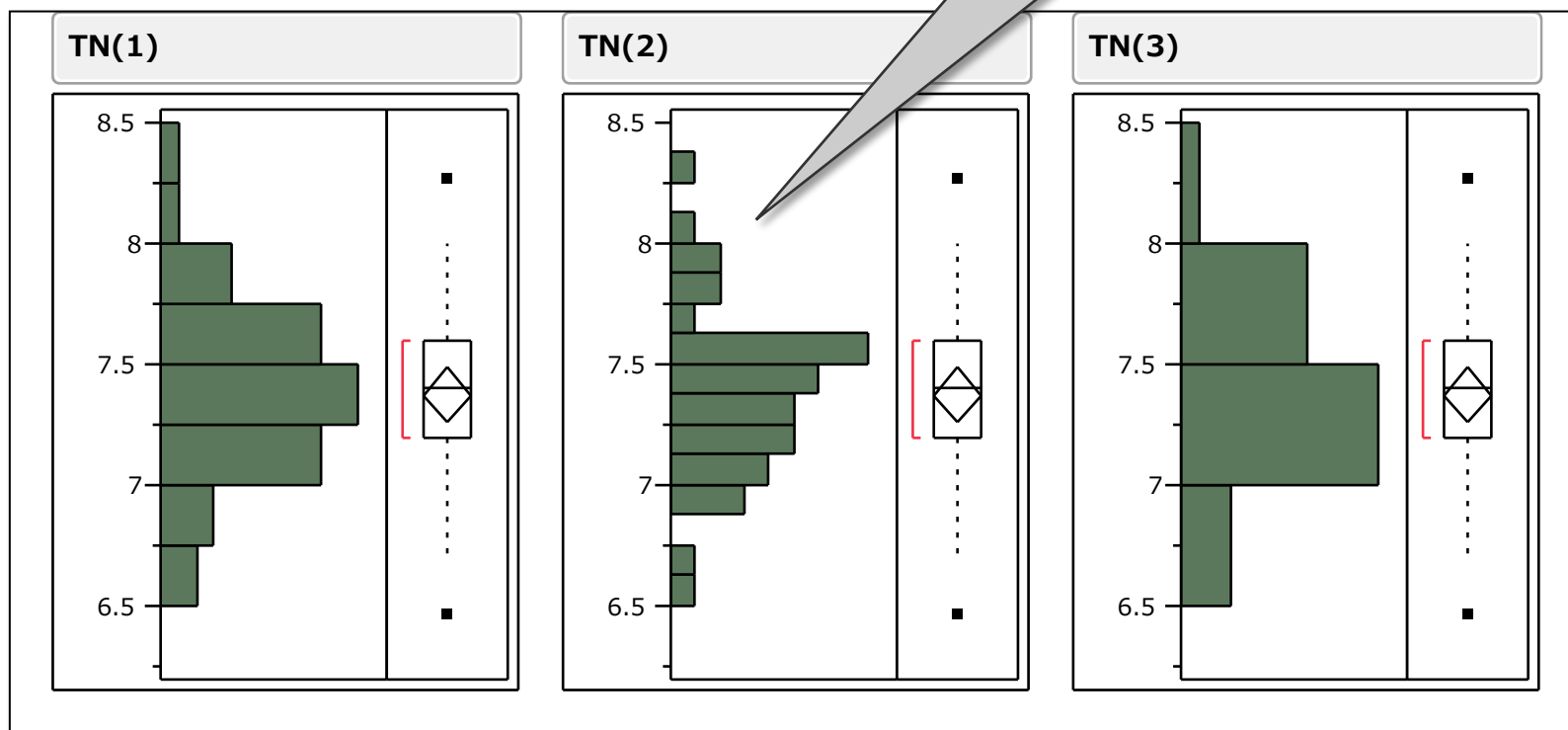
中心の位置

左右対称か

ばらつきの程度

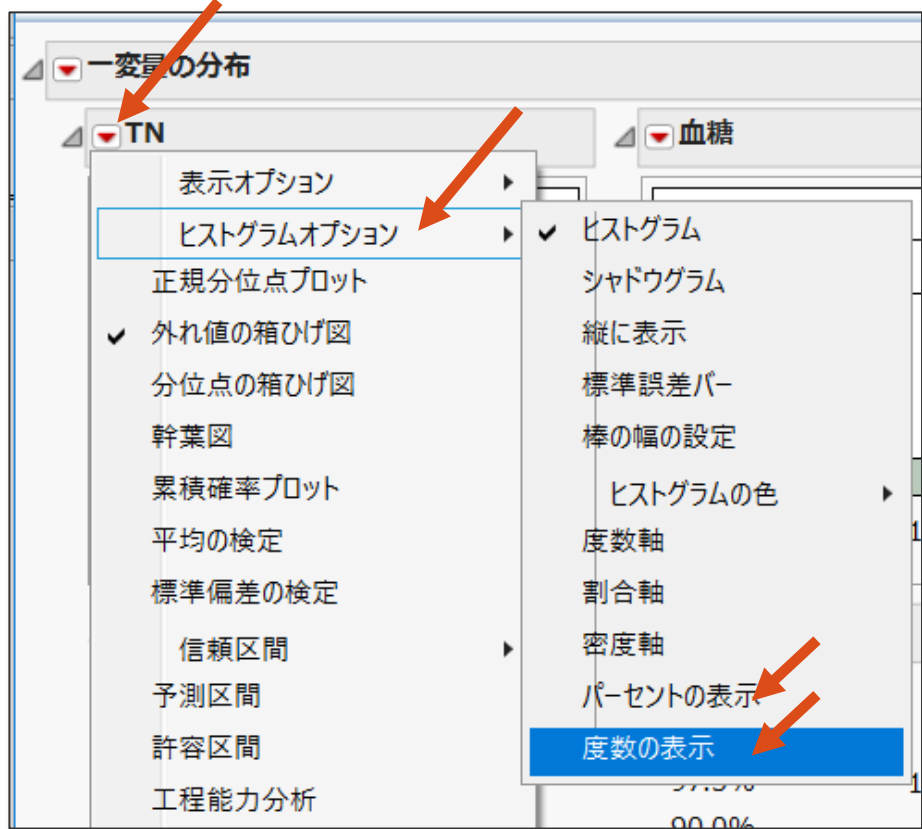
裾の状態

正規分布と見なせるか

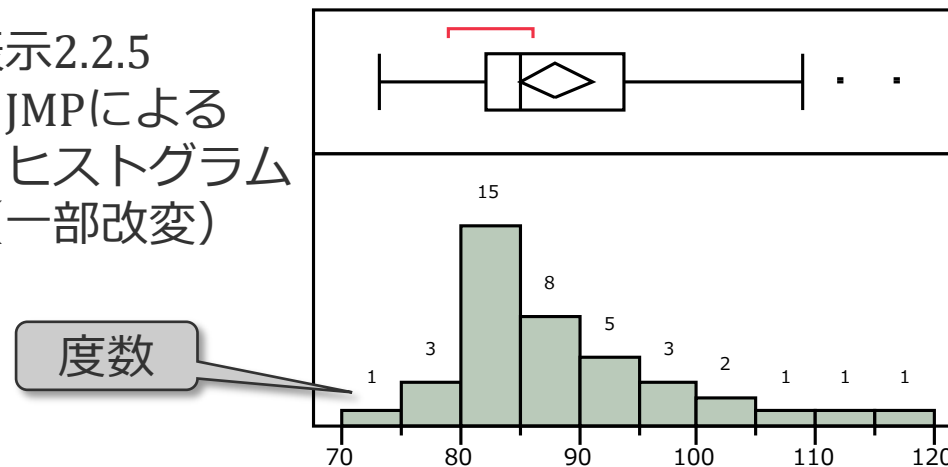


●度数の表示（度数表の作成）

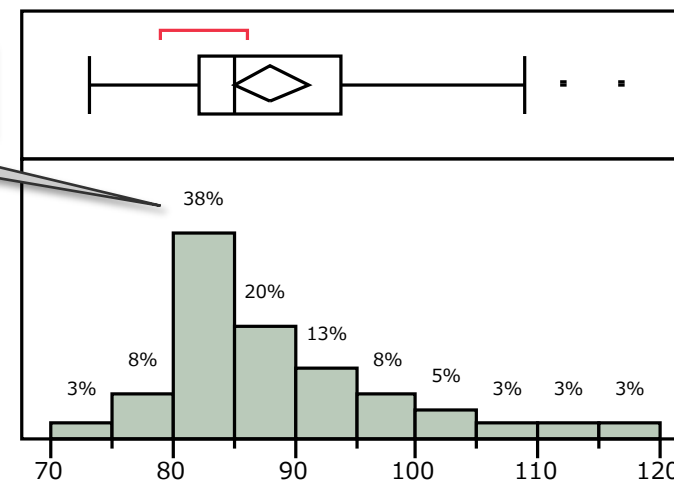
JMPには度数表を表示する機能はない
ヒストグラムに度数と相対度数を付記



表示2.2.5
JMPによる
ヒストグラム
(一部改変)

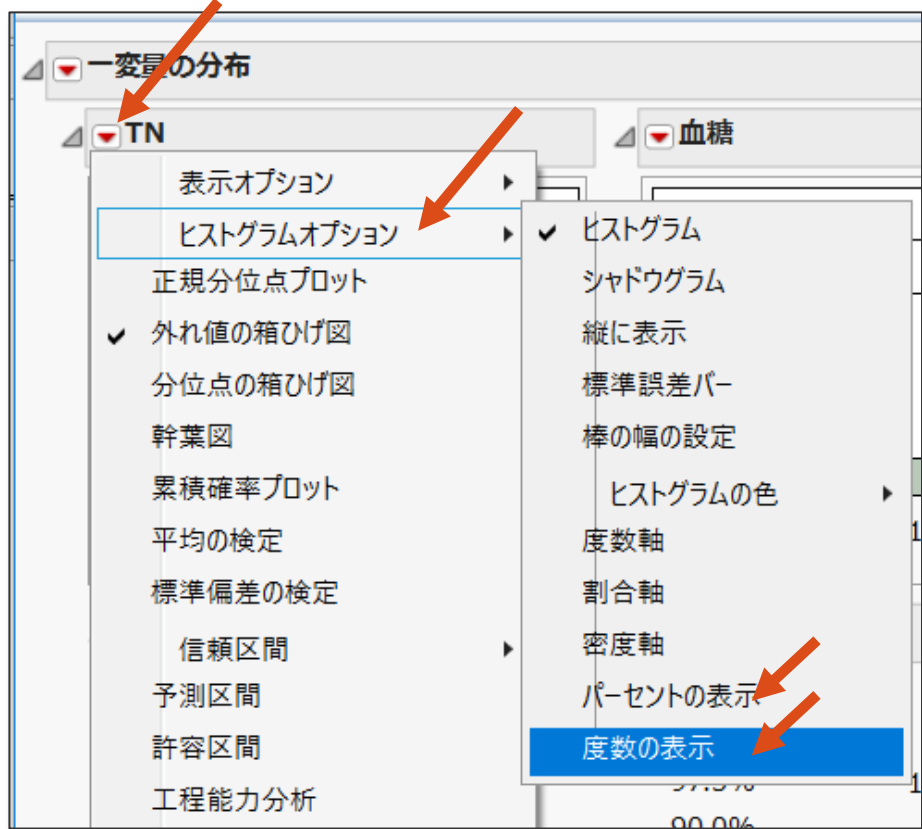


度数の % 表示
相対度数



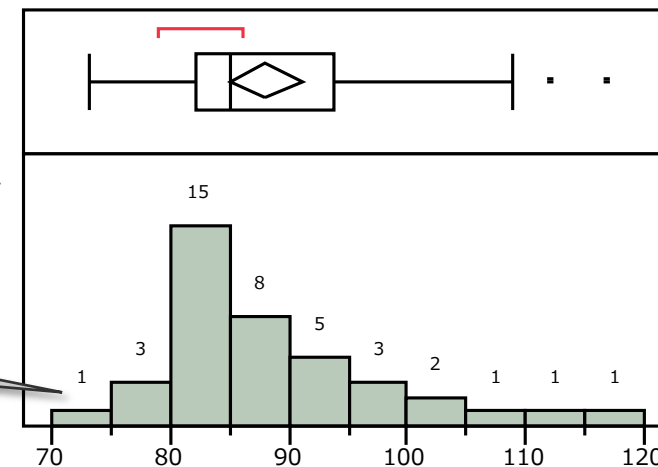
●度数の表示（度数表の作成）

JMPには度数表を表示する機能はない
ヒストグラムに度数と相対度数を付記



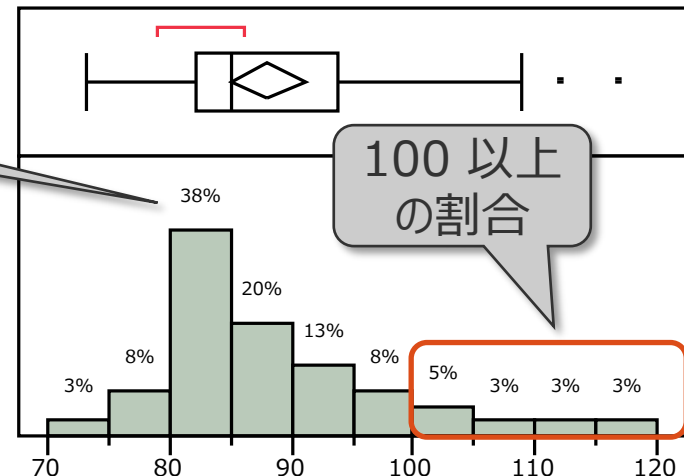
表示2.2.5
JMPによる
ヒストグラム
(一部改変)

度数



度数の % 表示
相対度数

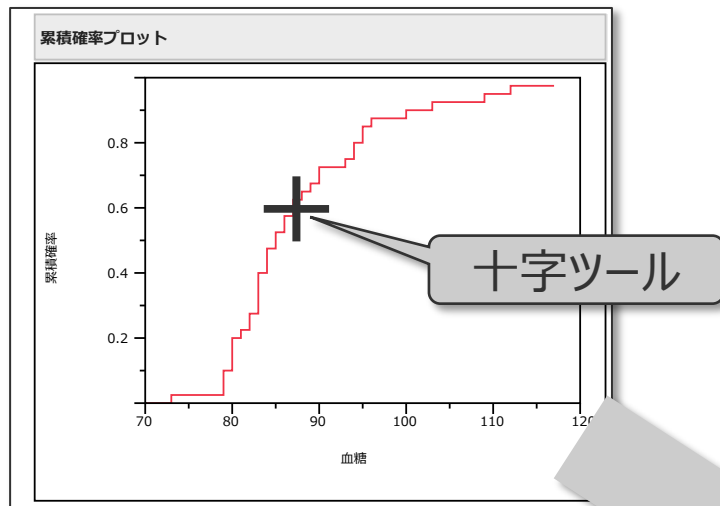
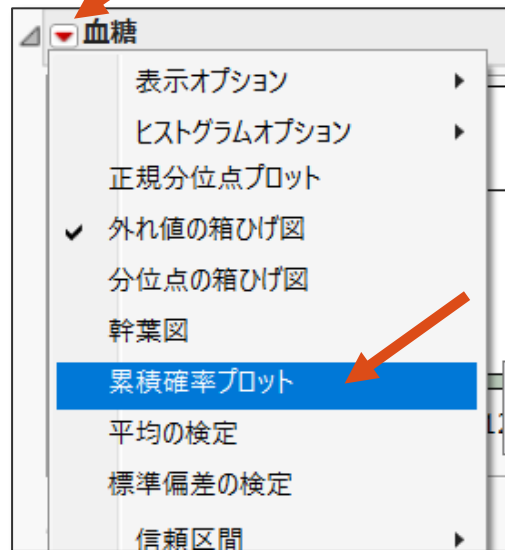
血糖 100 以上の割合
 $5 + 3 + 3 + 3 = 14$
105 以上の割合？
82 以下の割合？



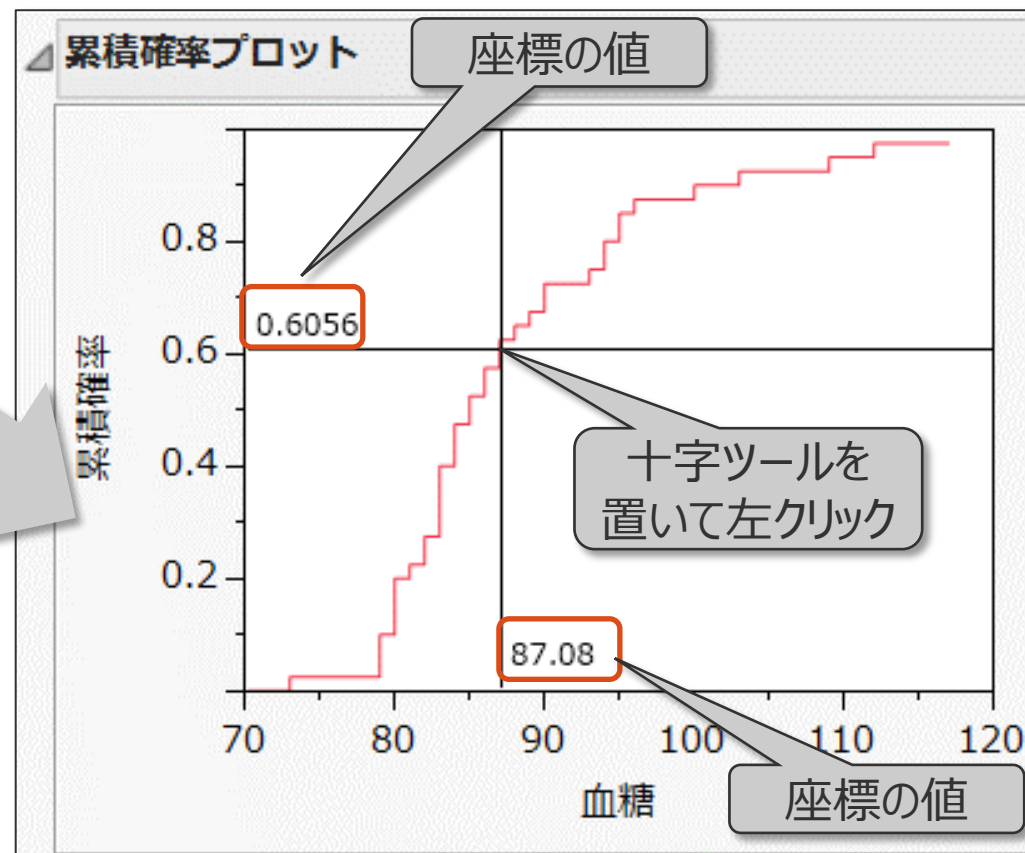
JMPによる累積確率プロット

●ある値以上（以下）の割合の取得

表示2.2.5 右



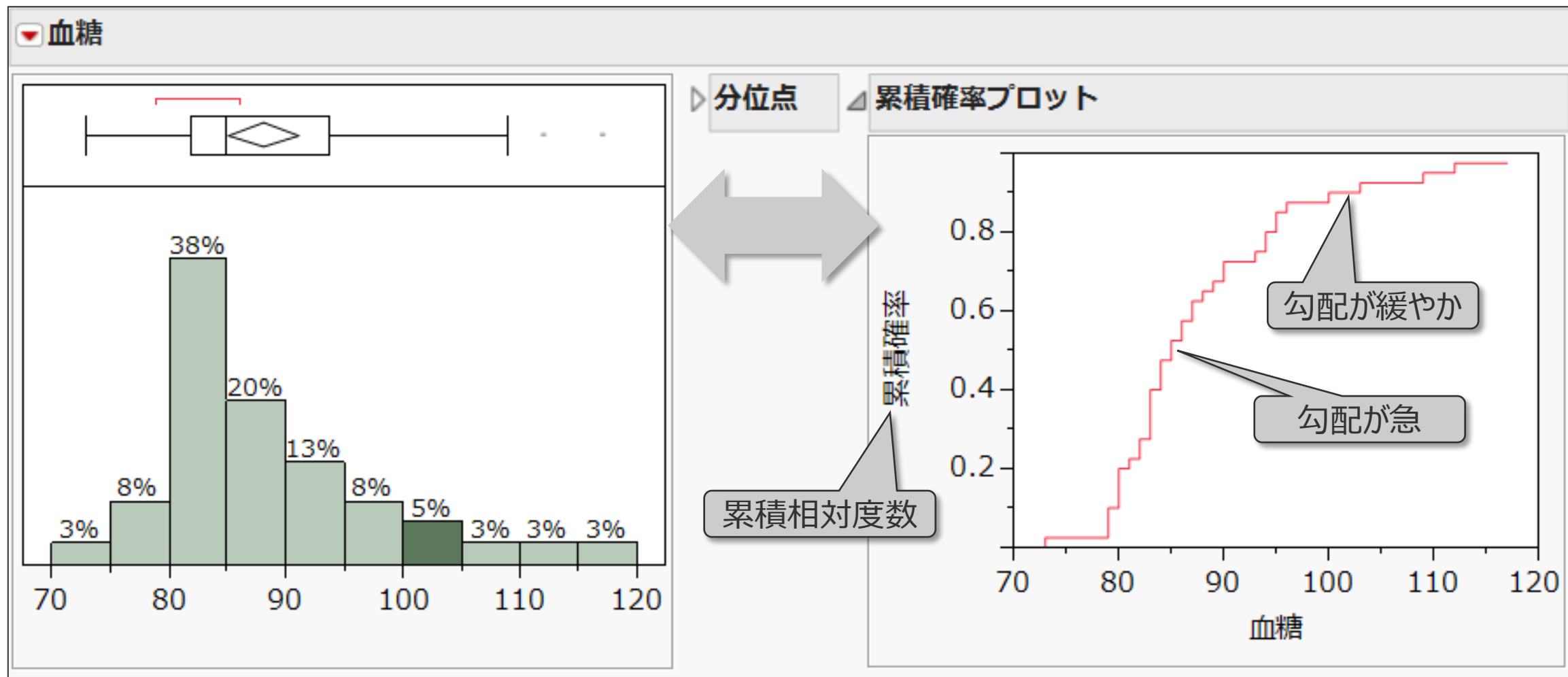
矢印ツールから十字ツールに変更
階段線の上で左クリックすると
任意の x に対する累積確率が求められる



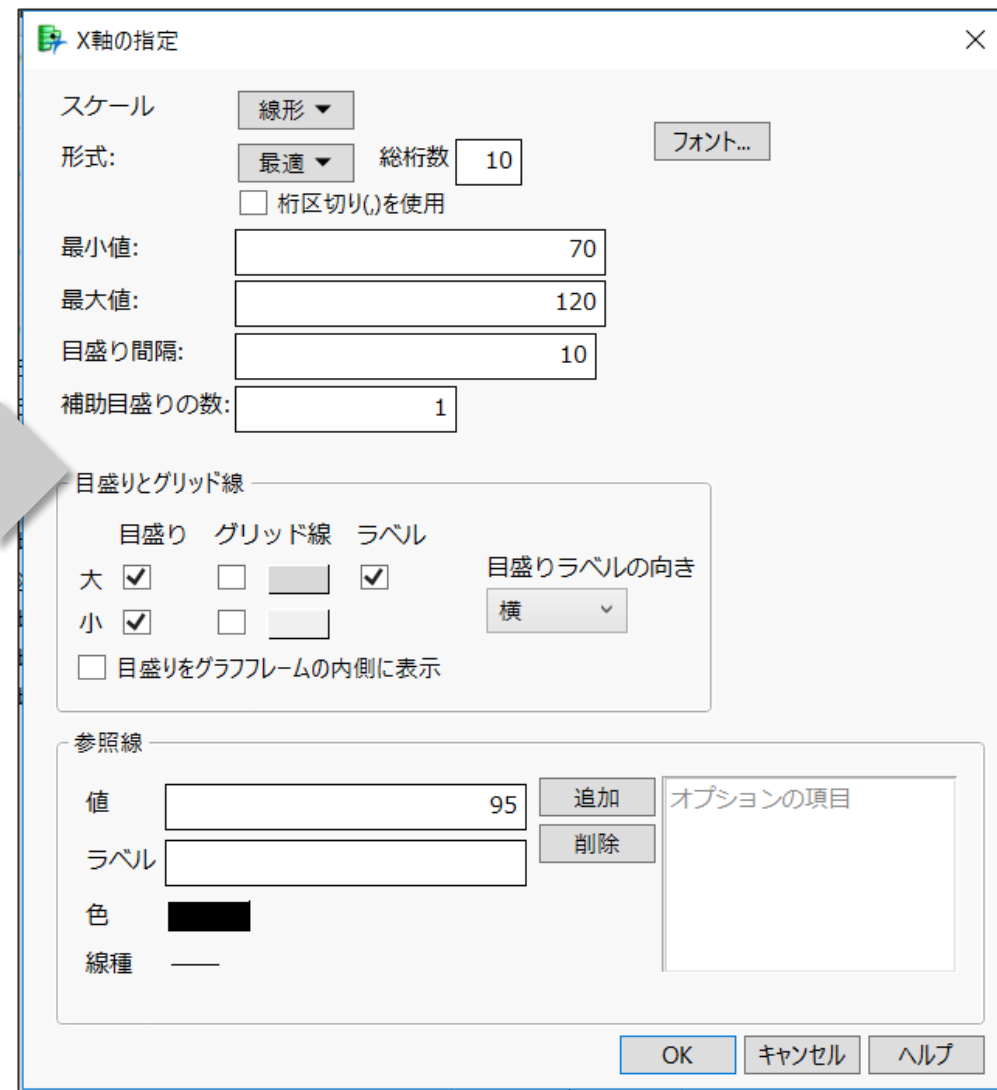
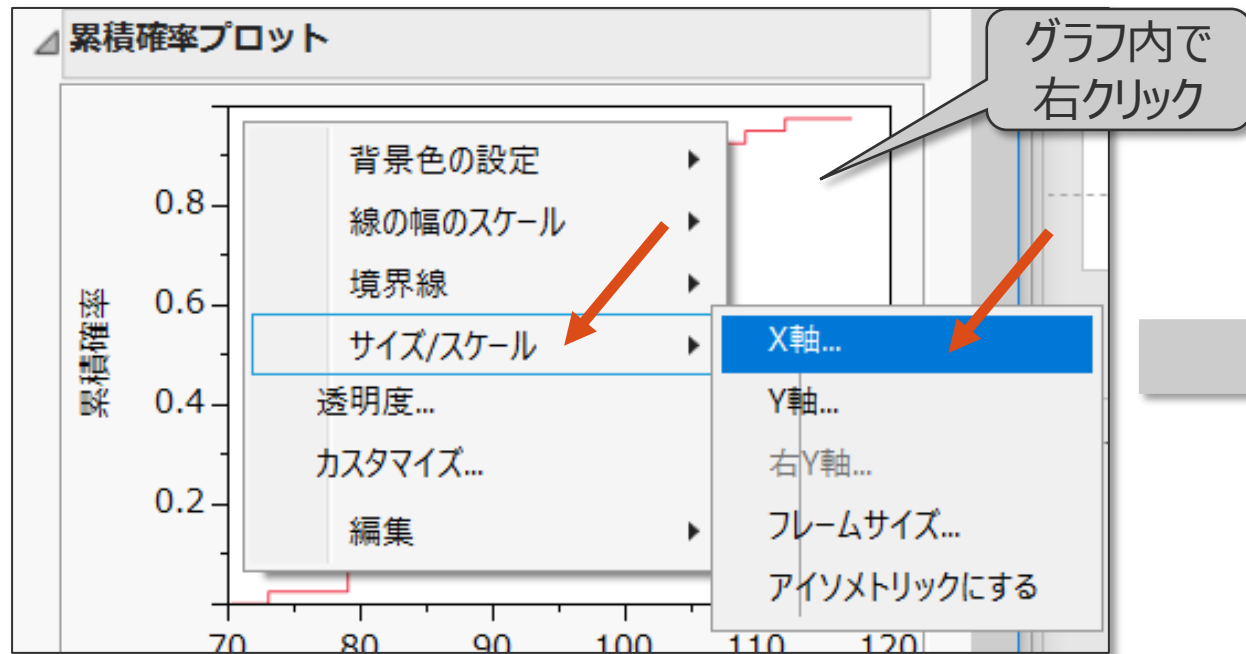
JMP による累積確率プロット

p.75

●ヒストグラムと累積確率プロット



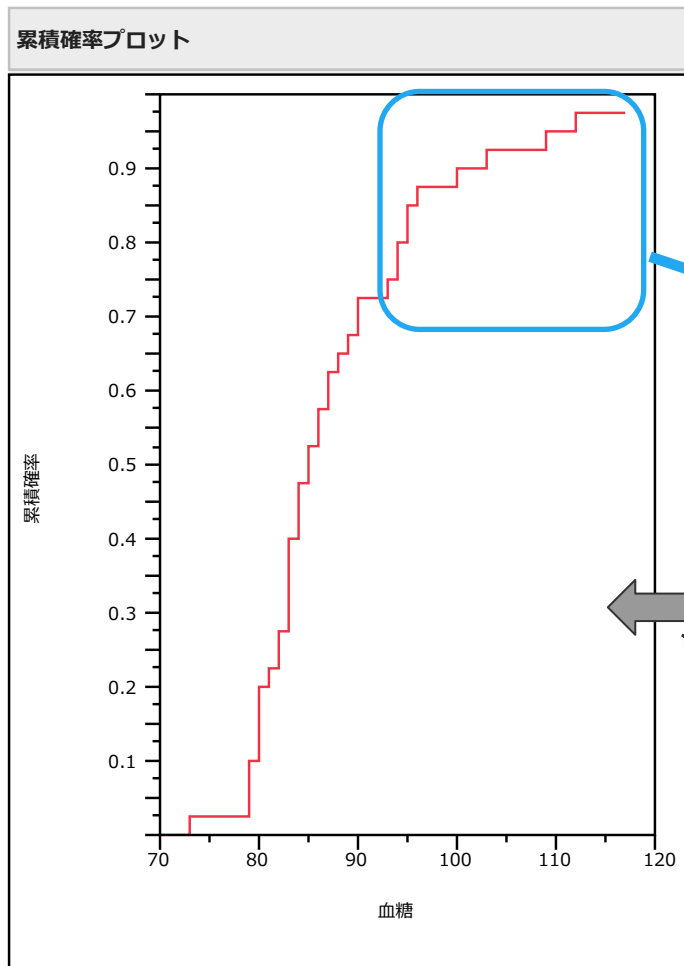
●累積確率プロットの整形



グラフ内で右クリック

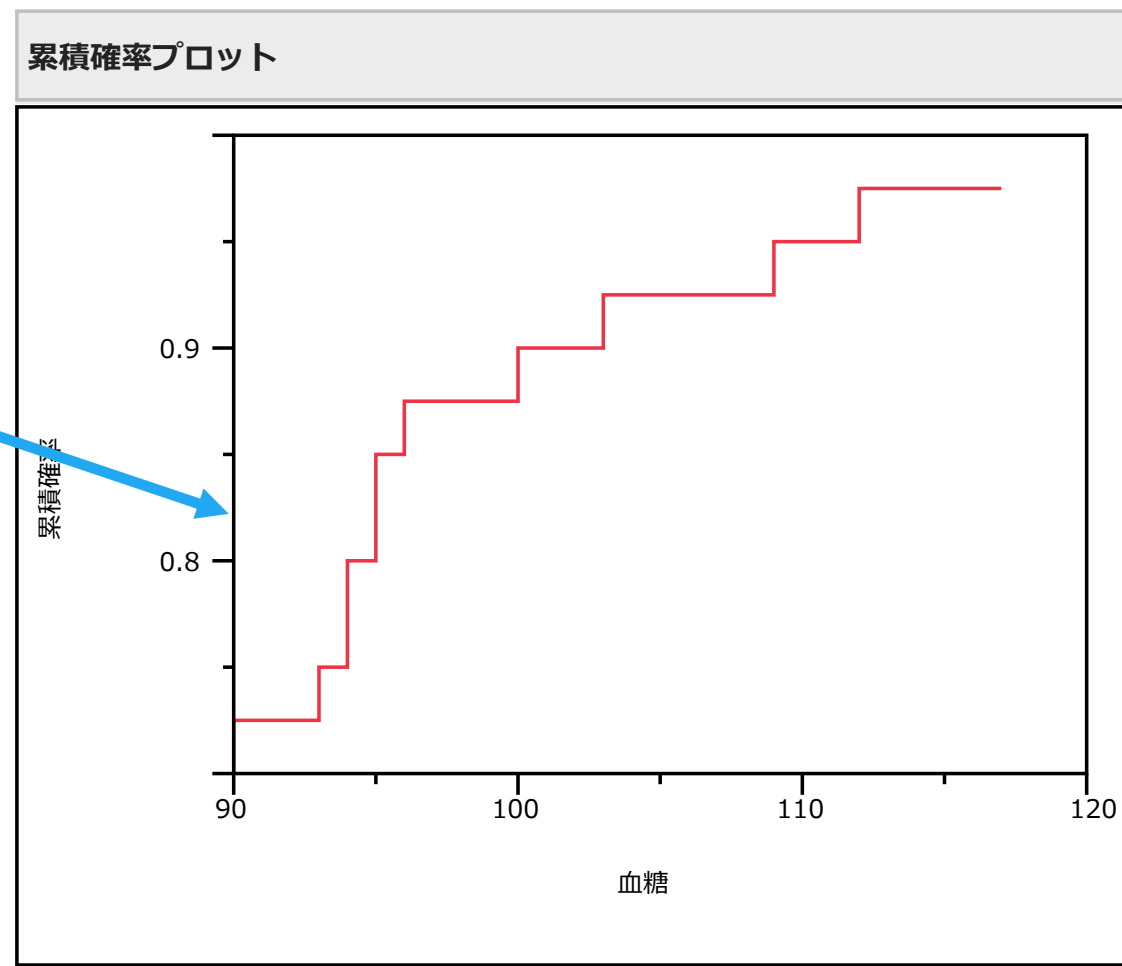
[サイズ/スケール] > [X軸]、[Y軸]
X軸とY軸を選択するとダイアログが表示される
最小値、最大値、目盛り間隔などを設定

● 累積確率プロットの整形



x 軸と y 軸の
設定変更

フレームを
ドラッグして
サイズを変更





(3) 箱ひげ図

Excel のマクロによる箱ひげ図
Excel のグラフ作成機能による箱ひげ図
JMP による箱ひげ図

●前節の事例 1：8都市から1人ずつ1か所に集まる最短の集合場所 (§2.1)

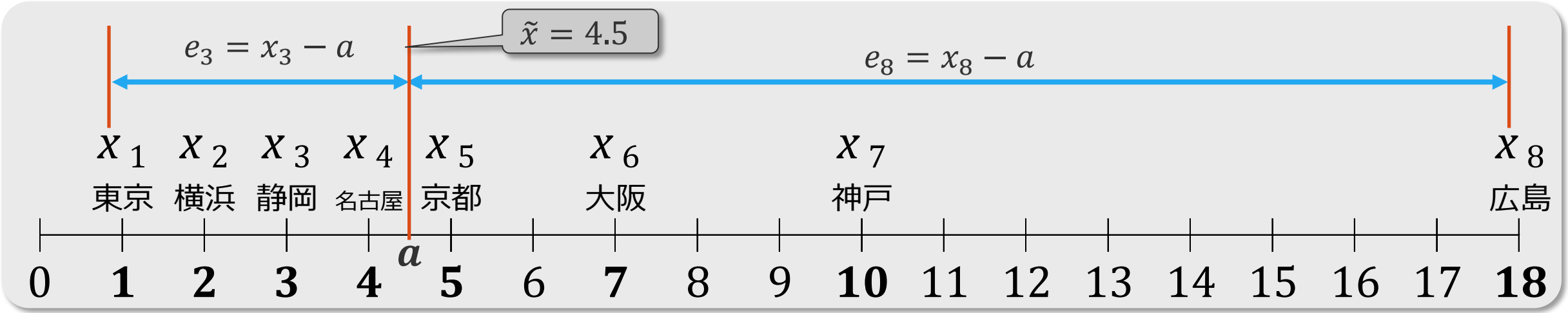
起点から各都市の距離 x_i … 観測値 (一般の場合に拡張するための表現)

集合場所 a … 代表値 (一般の場合に拡張するための表現)

移動距離の絶対値 $|e_i| = |x_i - a|$ 単に e_i を加算しただけでは相殺されてしまう

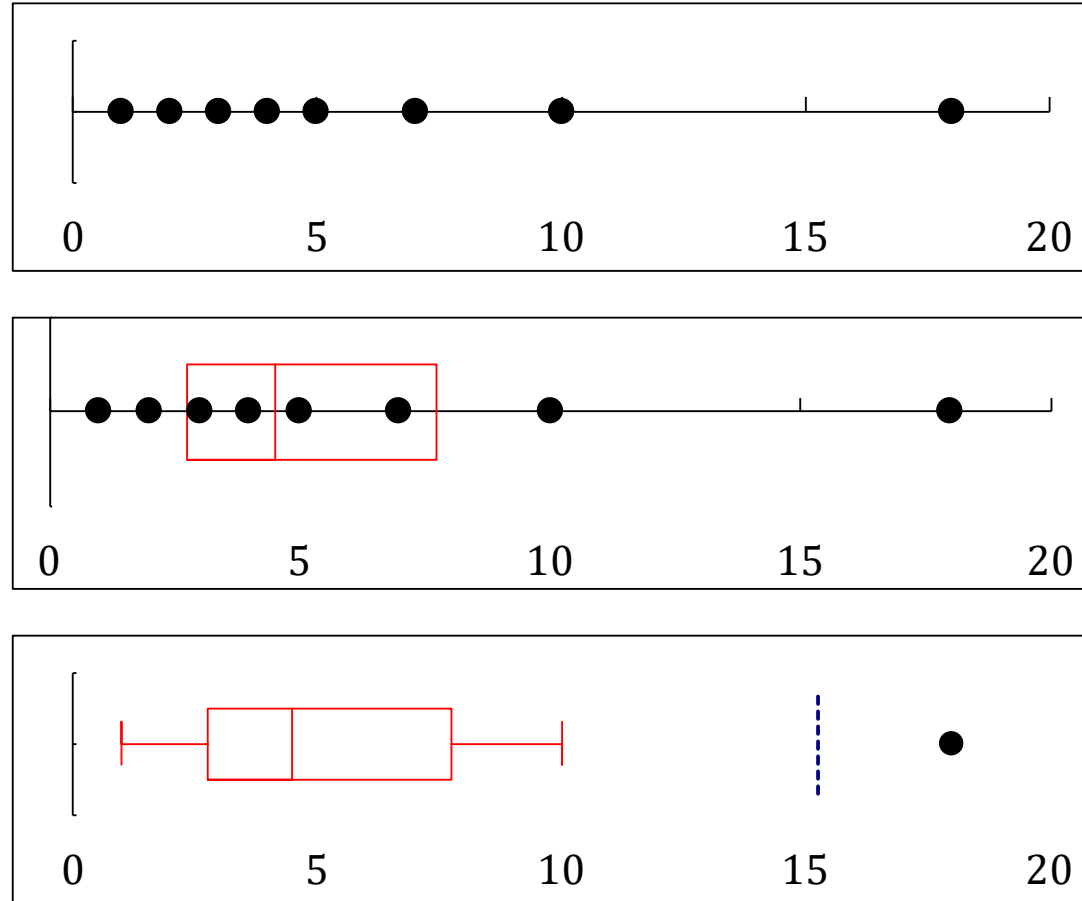
移動距離の絶対値の合計 T
 T が最小になる a $T = \sum_{i=1}^n |e_i| = \sum_{i=1}^n |x_i - a| \Rightarrow \min$ (2.1.1)

T が最小になる a を決める



●箱の作成

表示 2.2.6 箱ひげ図



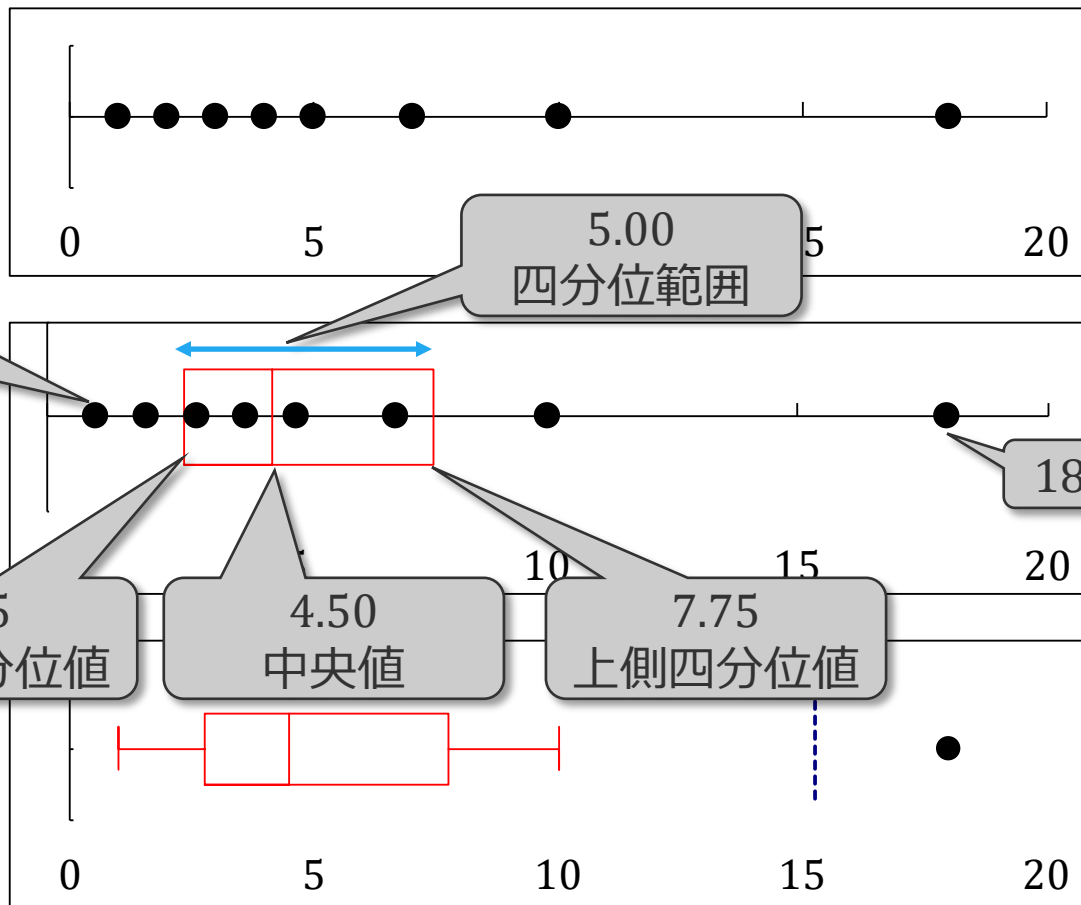
表示 2.1.3 (p.62)

i	x _i
1	1
2	2
3	3
4	4
5	5
6	7
7	10
8	18
n	8
最小値	1.00
下側四分位値	2.75
中央値	4.50
上側四分位値	7.75
最大値	18.00
四分位範囲	5.00
四分位範囲/1.35	3.70
ひずみ	1.57
とがり	2.51
標準偏差	5.55



●箱の作成

表示 2.2.6 箱ひげ図



データは箱の
内に2個ずつ
外に2個ずつ

5.00
四分位範囲

18 外れ値？

2.75
下側四分位値

4.50
中央値

7.75
上側四分位値

ひずみ、とがりが±1.5を超える (§ 2.1)
正規分布から外れた分布？ 外れ値が存在？

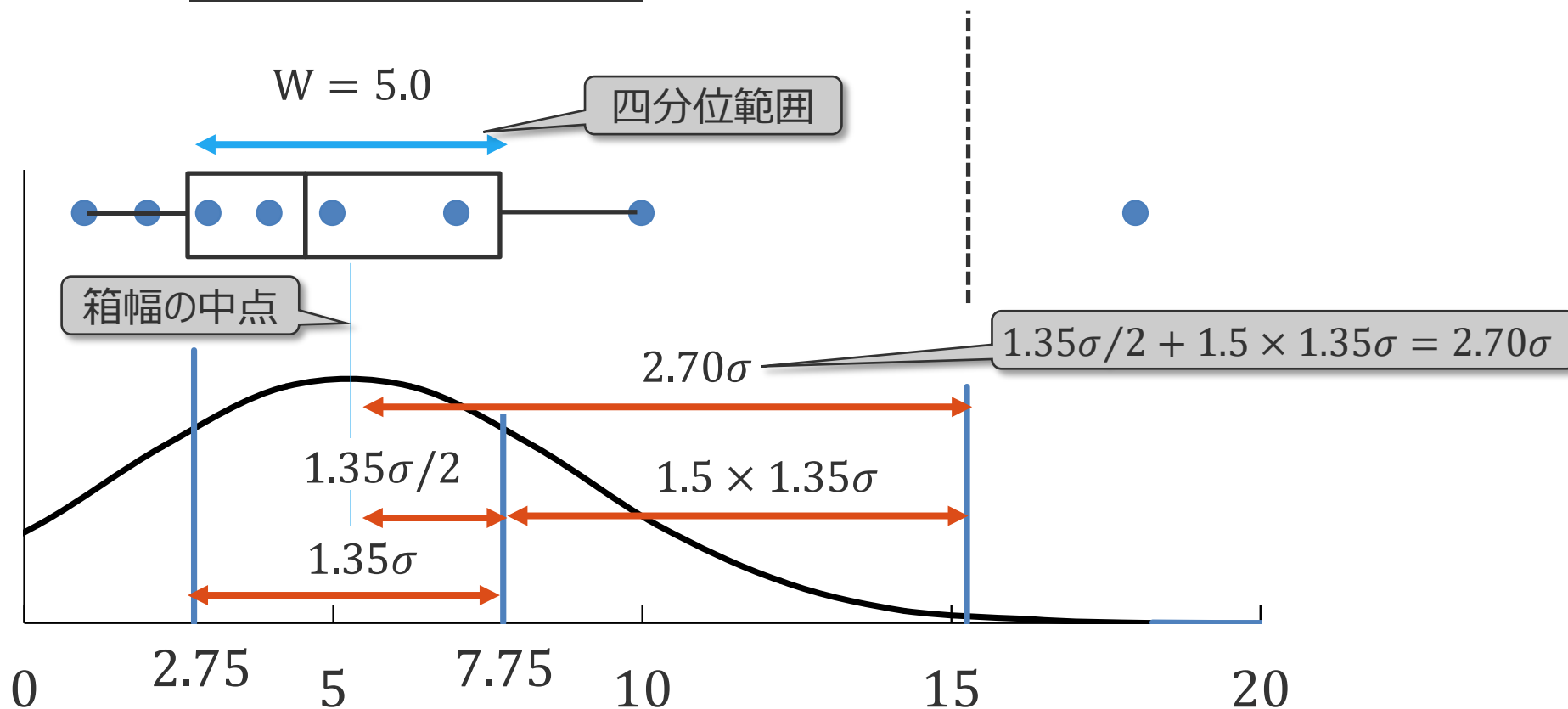
表示 2.1.3 (p.62)

i	x _i
1	1
2	2
3	3
4	4
5	5
6	7
7	10
8	18
n	8
最小値	1.00
下側四分位値	2.75
中央値	4.50
上側四分位値	7.75
最大値	18.00
四分位範囲	5.00
四分位範囲/1.35	3.70
ひずみ	1.57
とがり	2.51
標準偏差	5.55

●箱ひげ図における外れ値の考え方

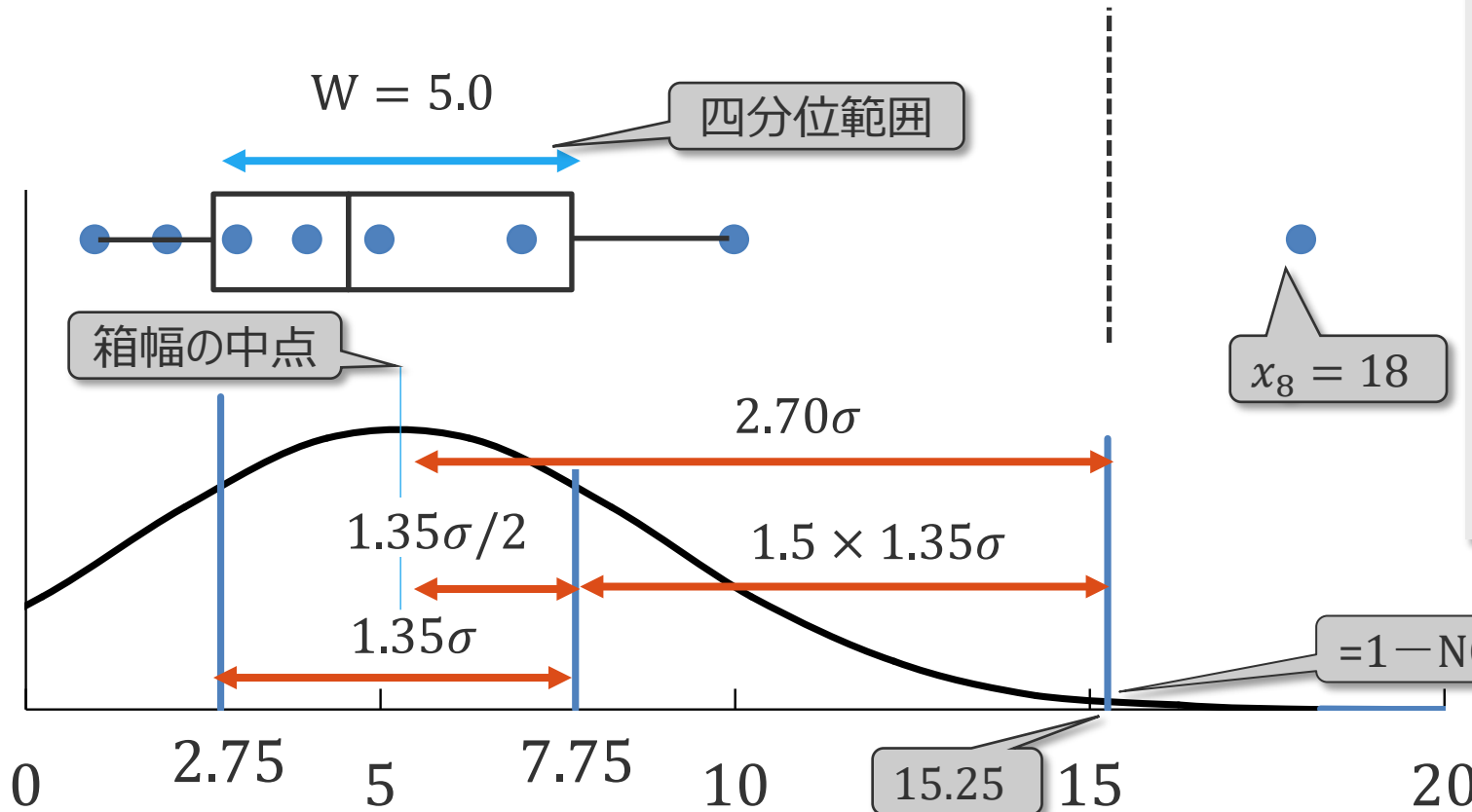
四分位範囲 W は 1.35σ の推定値 ([§2.1\(6\)](#))

(正規分布に従うと仮定して考える)



●箱ひげ図における外れ値の考え方

四分位範囲 W は 1.35σ の推定値 ([§2.1\(6\)](#))
(正規分布に従うと仮定して考える)



$s^* = \text{四分位範囲} / 1.35 = 3.70$
([§2.1](#) p.63)

$\bar{x} = \text{四分位範囲} / 2 + 2.75 = 5.25$

$\mu \pm 2.70\sigma \sim 5.25 \pm 2.70 \times 3.70$
 $= [-4.74, 15.25]$

上側確率 0.003、両側確率 0.006

上側四分位値 + $1.5 \times$ 四分位範囲
 $= 7.75 + 1.5 \times 5.0 = 15.25$

$= 1 - \text{NORM.DIST}(15.25, 5.25, 3.7, \text{TRUE}) = 0.003$

●箱ひげ図における外れ値の考え方

四分位範囲 W は 1.35σ の推定値 (§2.1(6))
(正規分布に従うと仮定して考える)

外れ値に頑健性がある

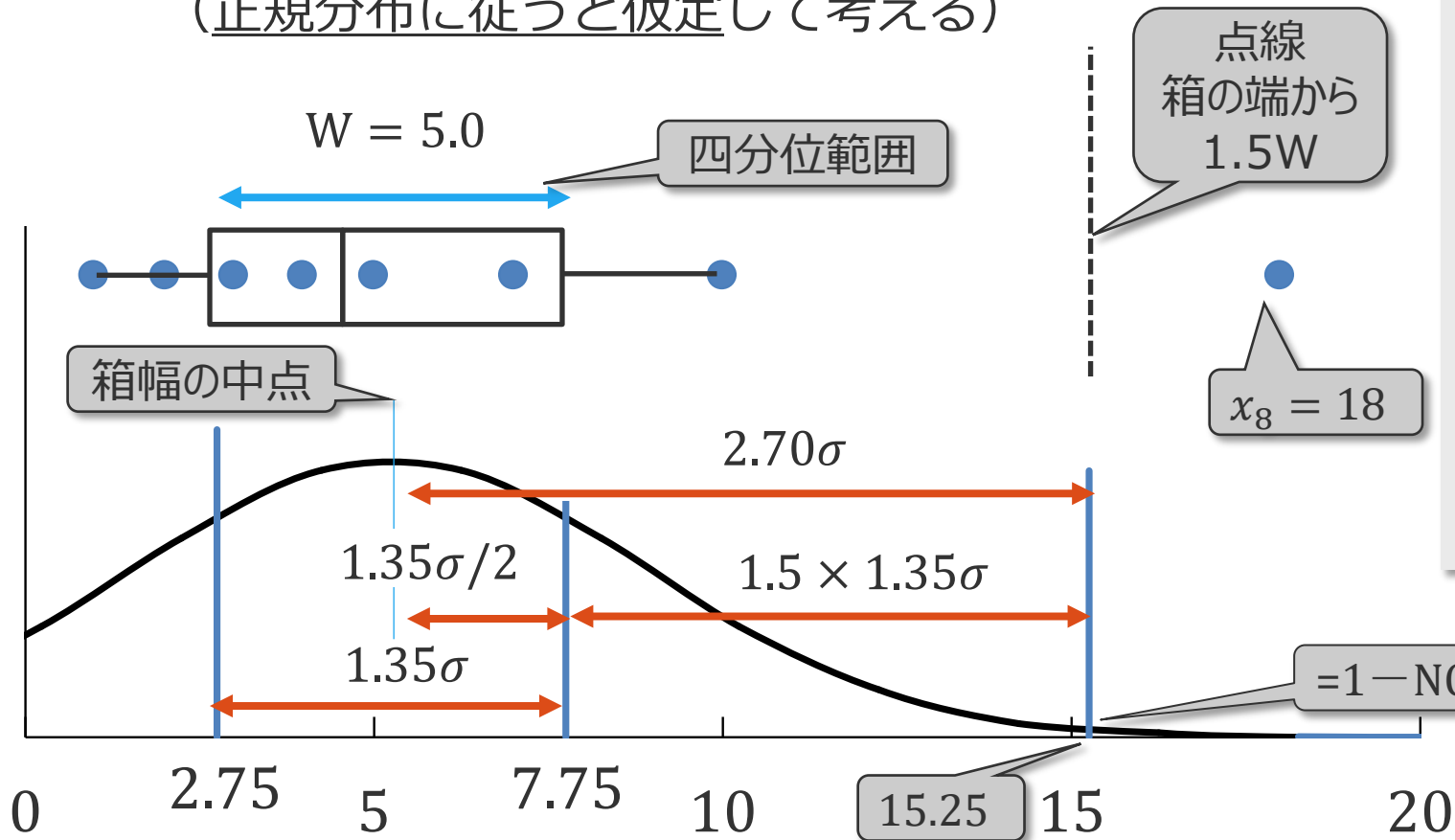
$$s^* = \text{四分位範囲} / 1.35 = 3.70 \quad (\text{§2.1 p.63})$$

$$\bar{x} = \text{四分位範囲} / 2 + 2.75 = 5.25$$

$$\mu \pm 2.70\sigma \sim 5.25 \pm 2.70 \times 3.70 = [-4.74, 15.25]$$

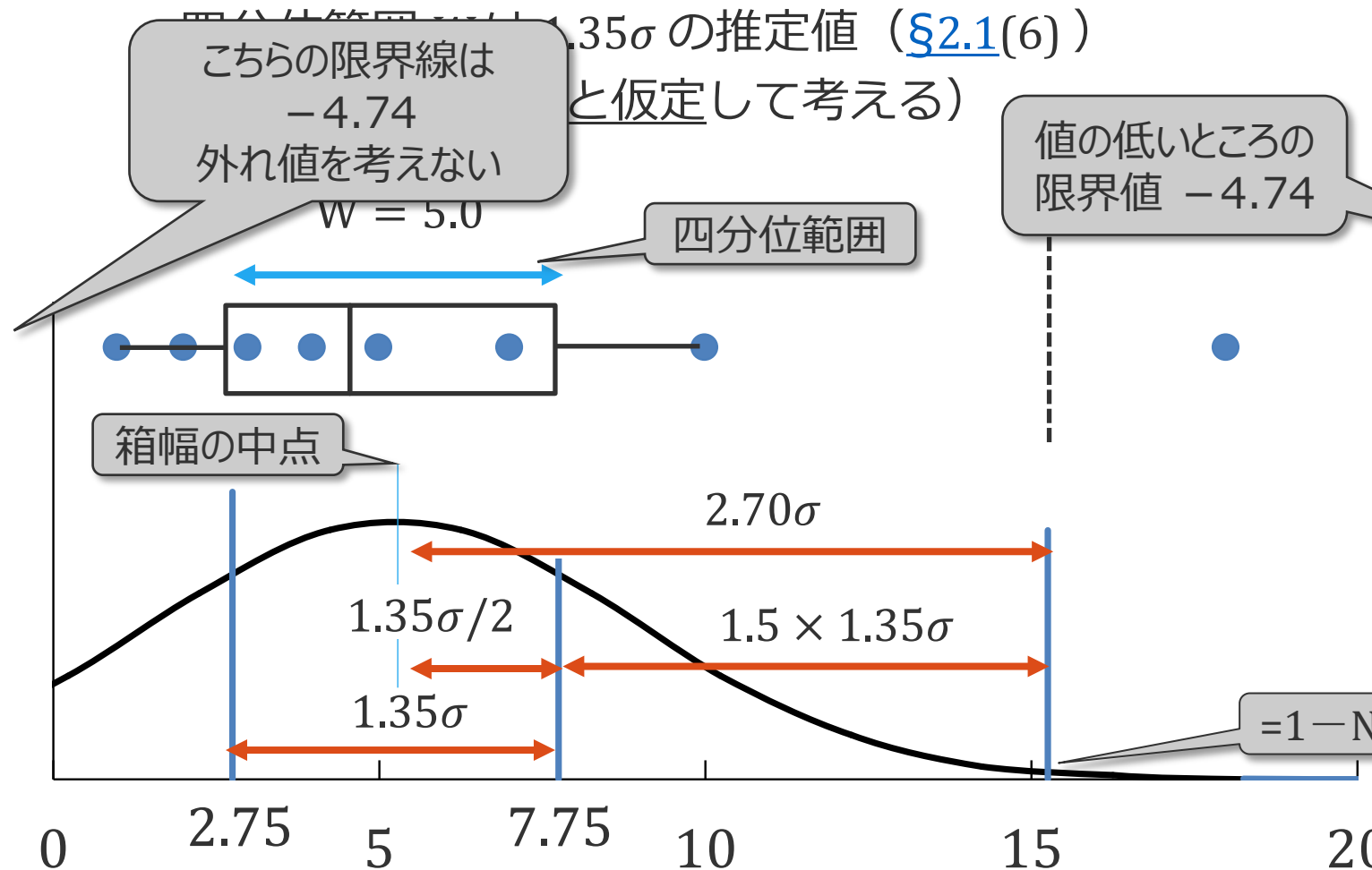
上側確率 0.003、両側確率 0.006

$$\text{上側四分位値} + 1.5 \times \text{四分位範囲} = 7.75 + 1.5 \times 5.0 = 15.25$$



$$= 1 - \text{NORM.DIST}(15.25, 5.25, 3.7, \text{TRUE}) = 0.003$$

●箱ひげ図における外れ値の考え方



$s^* = \text{四分位範囲} / 1.35 = 3.70$
(§2.1 p.63)

$\bar{x} = \text{四分位範囲} / 2 + 2.75 = 5.25$

$\mu \pm 2.70\sigma \sim 5.25 \pm 2.70 \times 3.70$
 $= [-4.74, 15.25]$

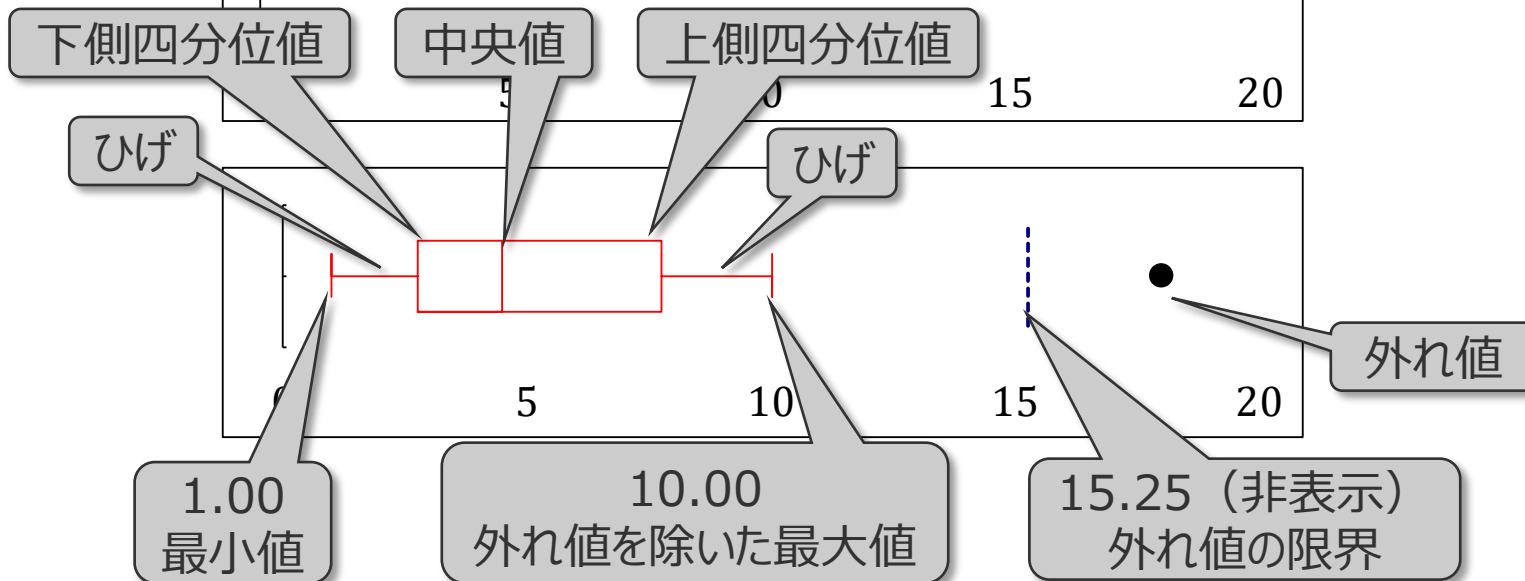
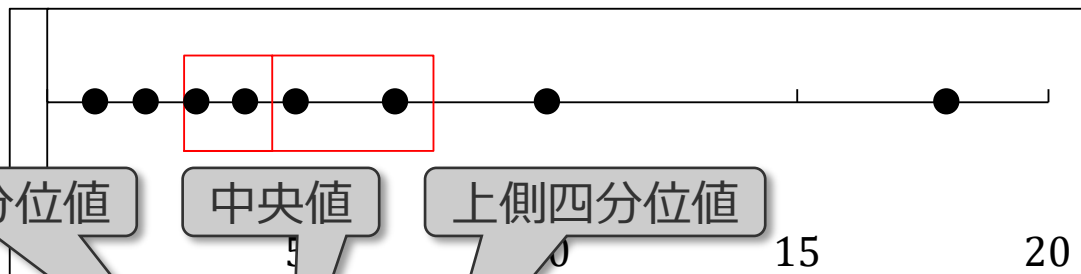
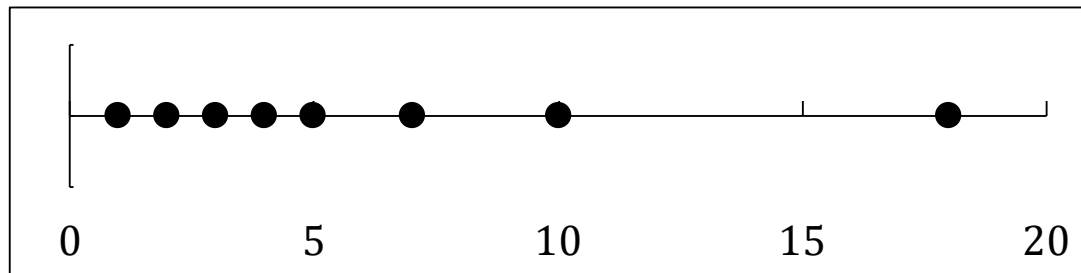
上側確率 0.003、両側確率 0.006

上側四分位値 + $1.5 \times$ 四分位範囲
 $= 7.75 + 1.5 \times 5.0 = 15.25$

$= 1 - \text{NORM.DIST}(15.25, 5.25, 3.7, \text{TRUE}) = 0.003$

●ひげの作成

表示 2.2.6 箱ひげ図



表示 2.1.3
(p.62)

	i	x _i
	1	1
	2	2
	3	3
	4	4
	5	5
	6	7
	7	10
	8	18
n		8
最小値		1.00
下側四分位値		2.75
中央値		4.50
上側四分位値		7.75
最大値		18.00
四分位範囲		5.00
四分位範囲/1.35		3.70
ひずみ		1.57
とがり		2.51

●Excel による箱ひげ図の作成

Excel 2016から箱ひげ図を作成する機能が追加された

範囲指定

	L	M	N
4	i		x
5	1		1
6	2		2
7	3		3
8	4		4
9	5		5
10	6		7
11	7		10
12	8		18

挿入 ページレイアウト 数式 データ 校閲 表示

アドイン おすすめグラフ

読み上げの停止 列ごと

× ✓ fx 1

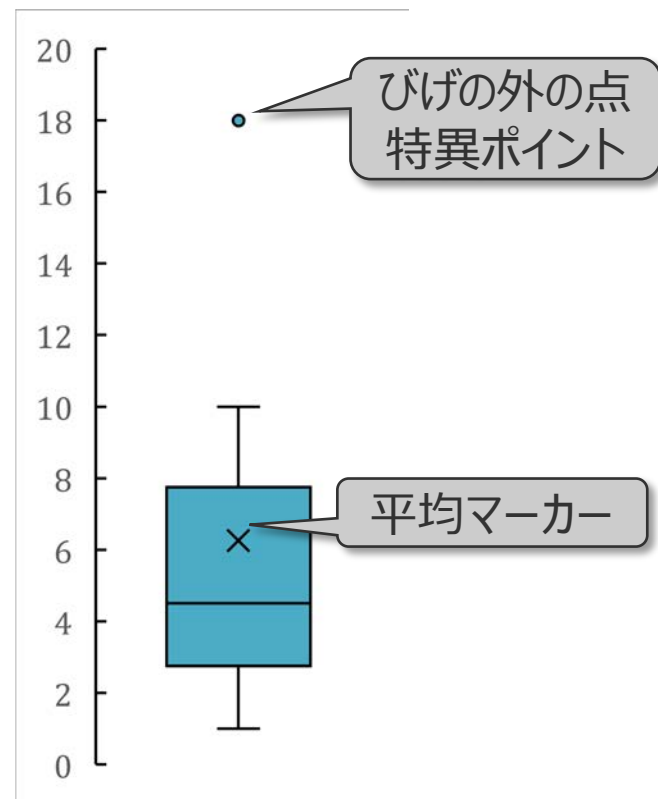
M N 0

ヒストグラム

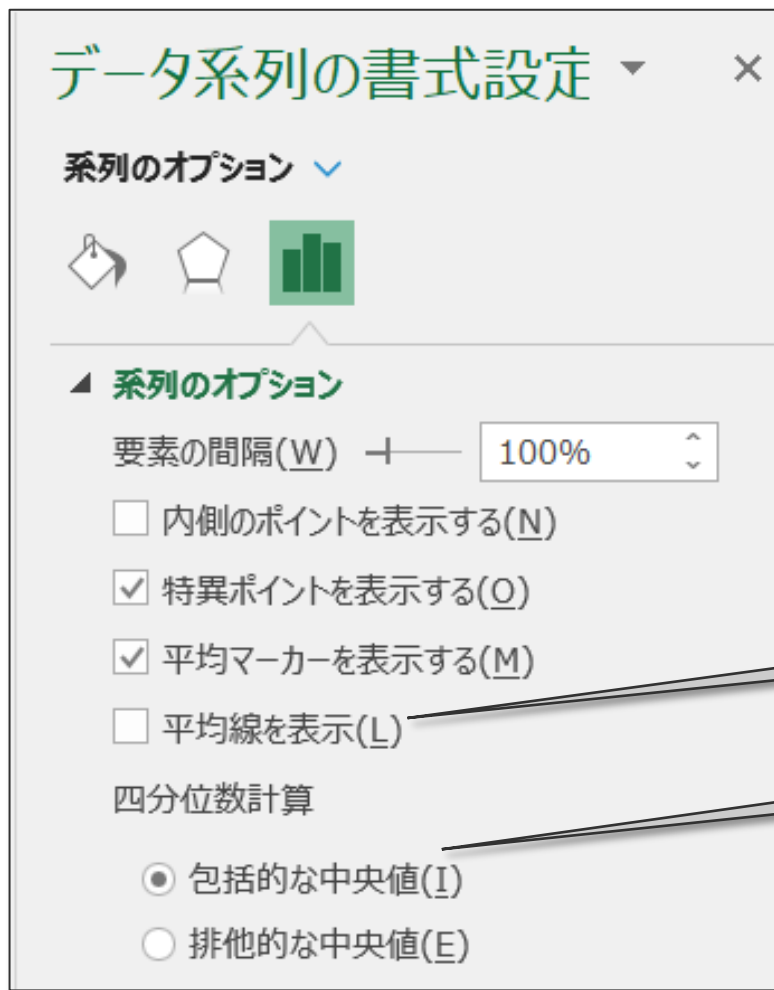
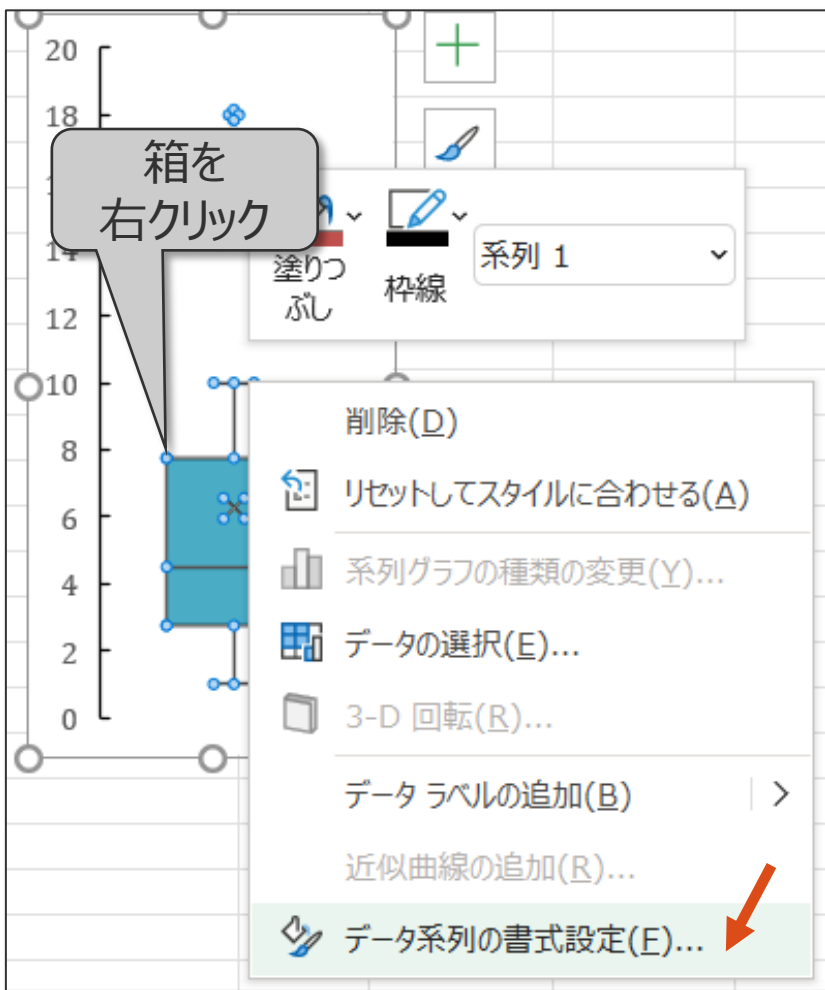
箱ひげ図

その他の統計グラフ(M)...

単一または複数のデータ
(1列または複数列) を範囲指定
[挿入] > [グラフ] > [箱ひげ図]



●Excel による箱ひげ図の作成



箱を右クリック
[データ系列の書式設定]
を選択

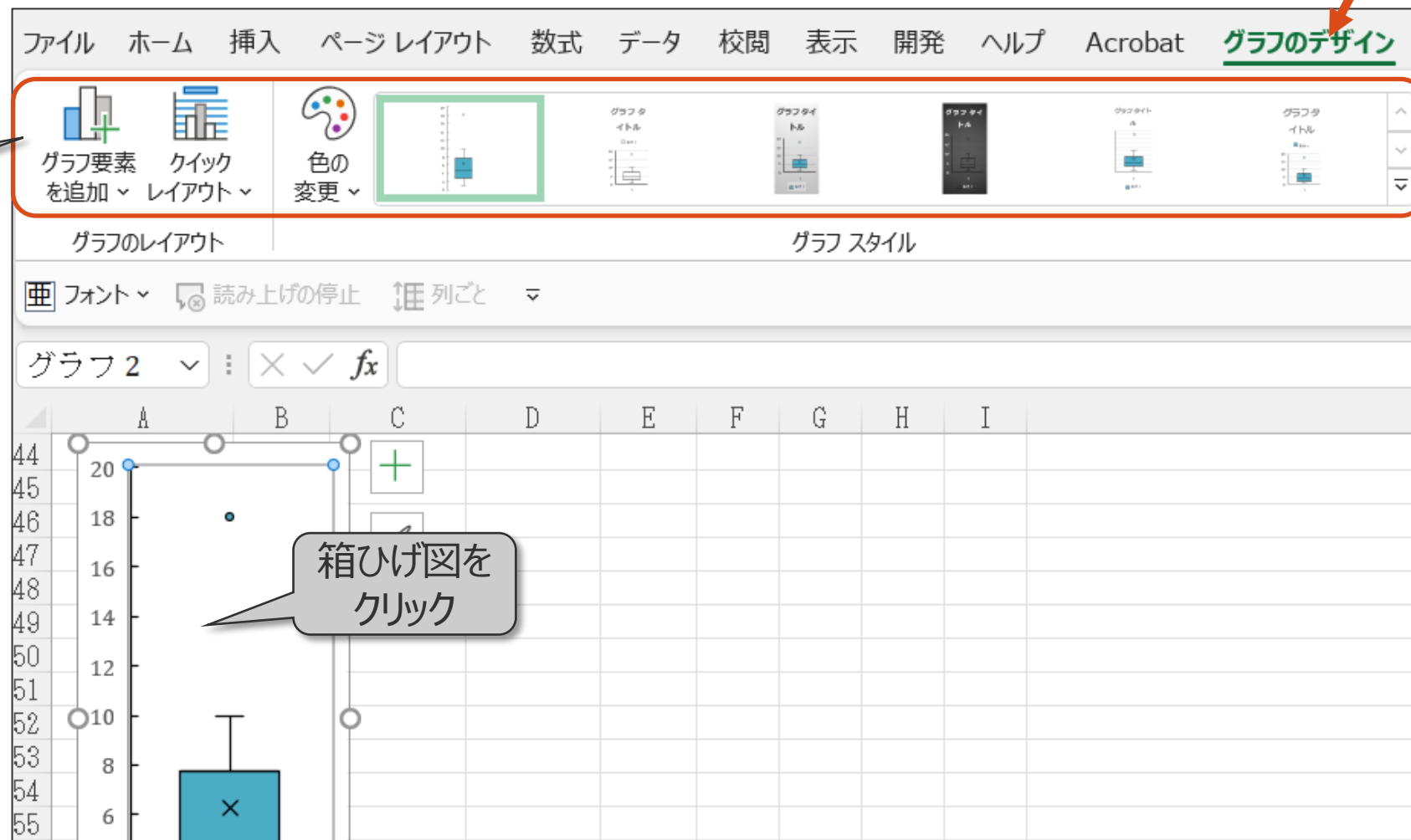
包括的な中央値：中央値を
含めて四分位値を計算
排他的な中央値：中央値を
含めず四分位値を計算

箱が複数ある場合に平均を結ぶ

テキストでの中央値

●Excel による箱ひげ図の作成

箱ひげ図の外観を設定



箱ひげ図の作成

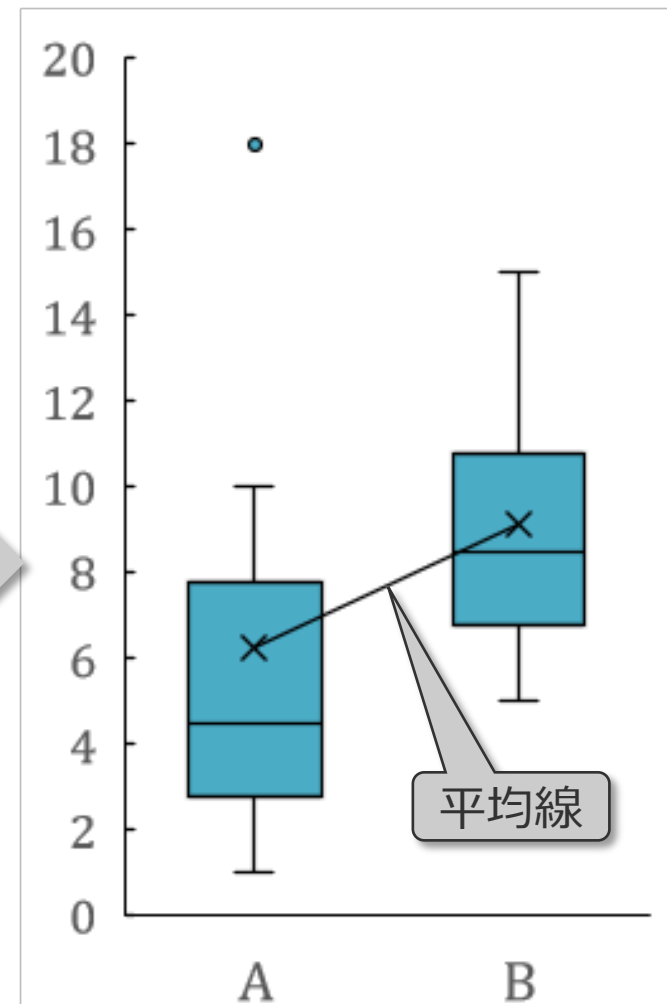
●Excel による箱ひげ図の作成

複数のデータ系列がある場合
データを縦に並べて範囲指定

A	B
1	5
2	6
3	7
4	8
5	9
7	10
10	13
18	15

範囲指定

群	観測値
A	1
A	2
A	3
A	4
A	5
A	7
A	10
A	18
B	5
B	6
B	7
B	8
B	9
B	10
B	13
B	15



●Excel のマクロによる箱ひげ図の作成

テキスト掲載のExcel マクロを利用

[入力] のセルにデータのセル範囲 M5:M12 を入力

[出力] のセルに計算結果の出力位置の

左上の位置 Q4 を入力

このセルの右下は空白にしておく

マクロの実行手順

黄色のセルをクリック

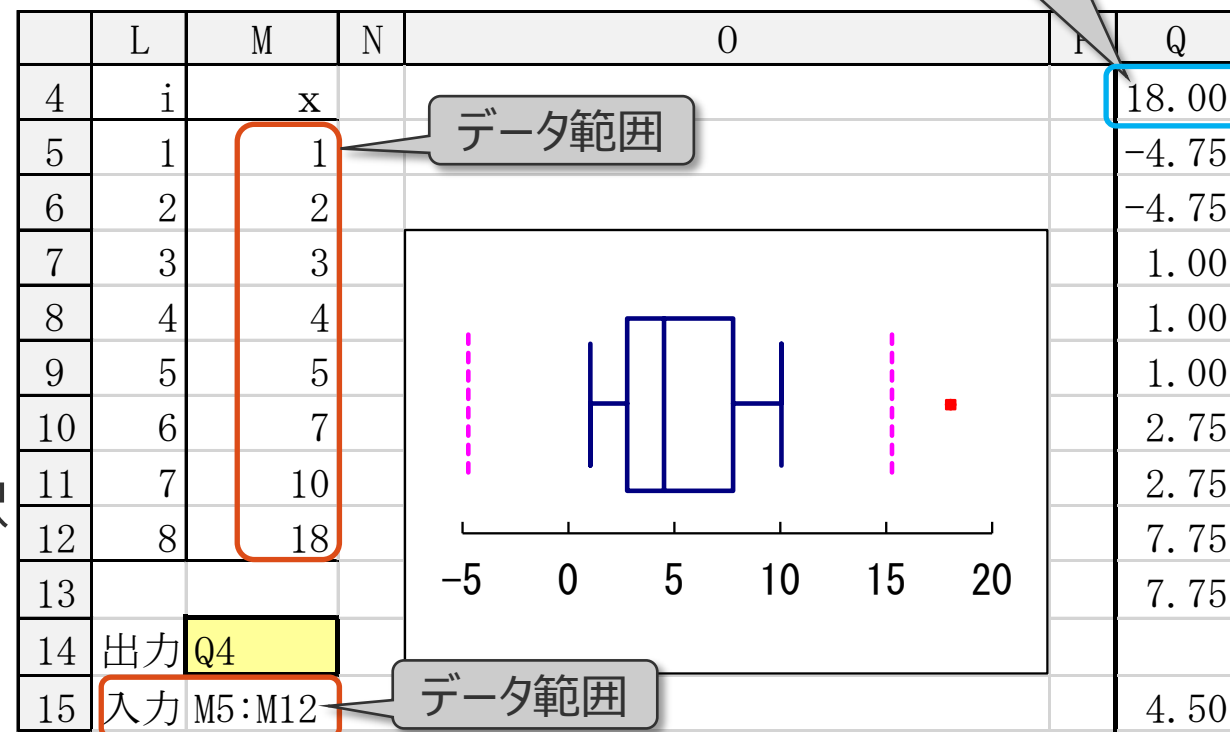
トップメニューの [表示] >

[マクロ] > [マクロの表示] を選択

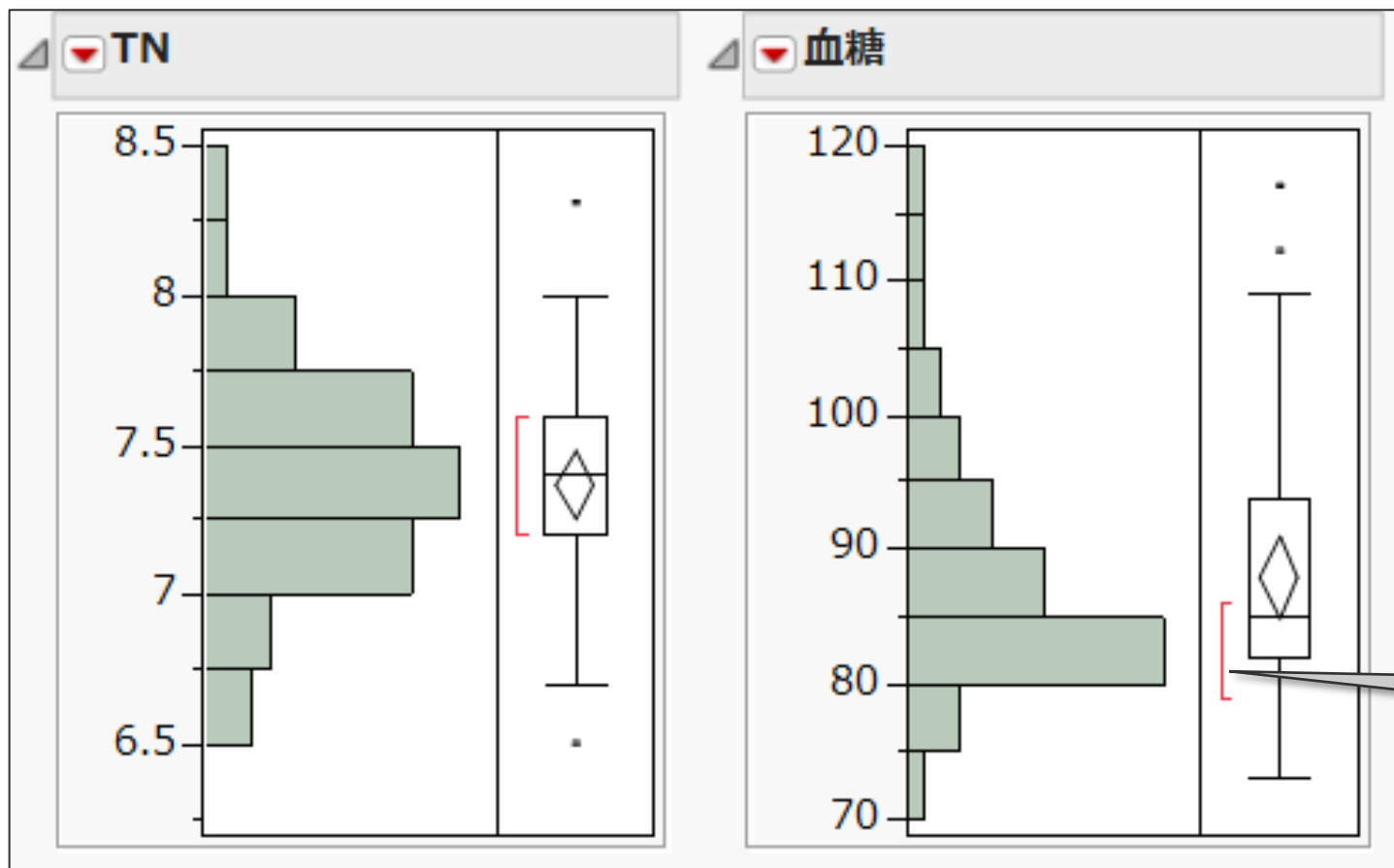
マクロ名リスト「箱ひげ図」を選択

[実行]

表示 2.2.7 マクロによる箱ひげ図



●JMPによる箱ひげ図の構成



箱、ひげ、外れ値の他に、
「平均のひし形（信頼のひし形）」と
「最短の半分」が加わる

平均のひし形

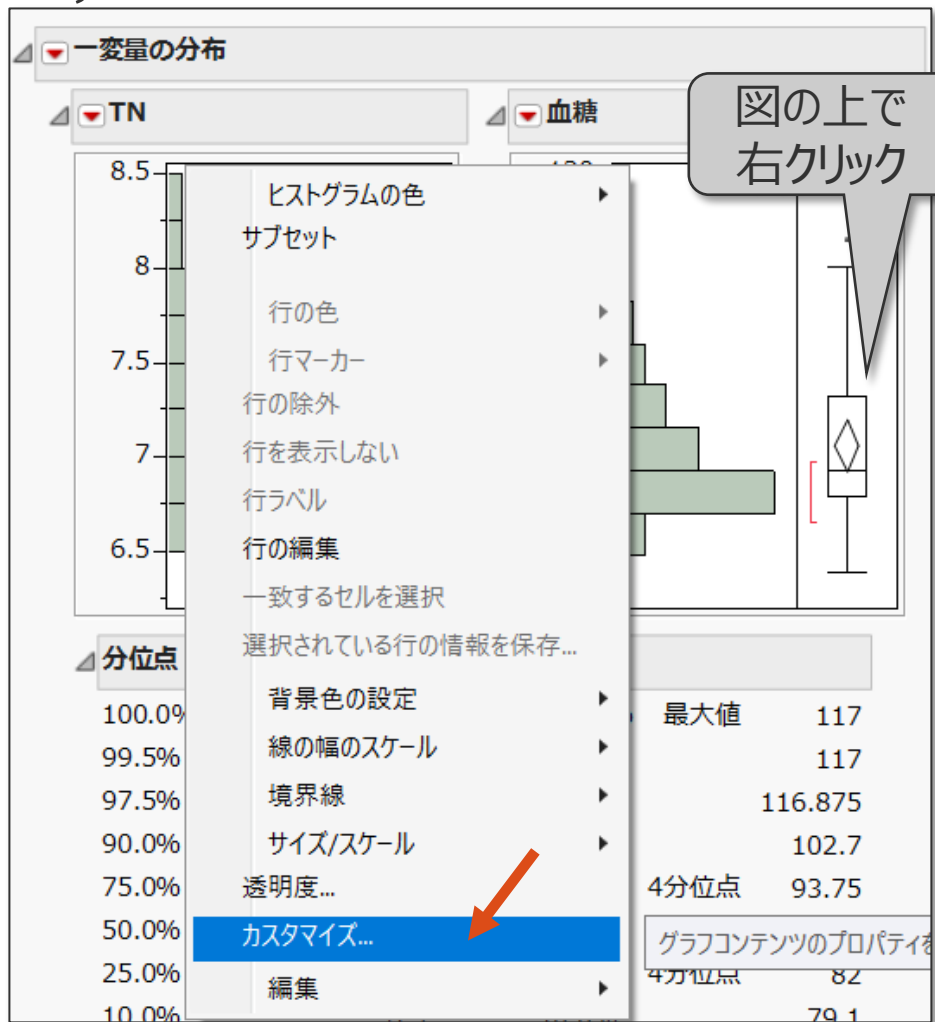
平均 + 95%信頼区間

平均

平均 - 95%信頼区間

最短の半分
観測値の50%を含む
区間の内、最短の範囲

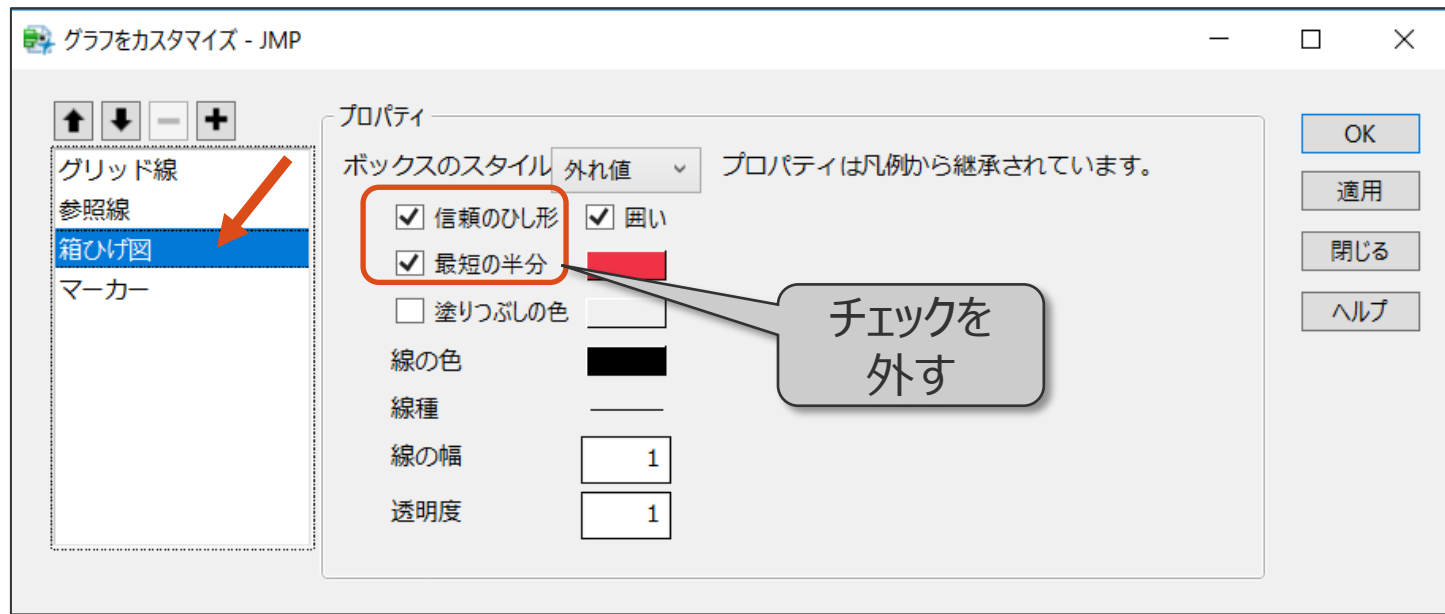
●JMPによる箱ひげ図の作成



平均のひし形、最短の半分を消去する方法
箱ひげ図の上で右クリック>

[カスタマイズ...] > [箱ひげ図]

[信頼のひし形]、[最短の半分] のチェックを外す



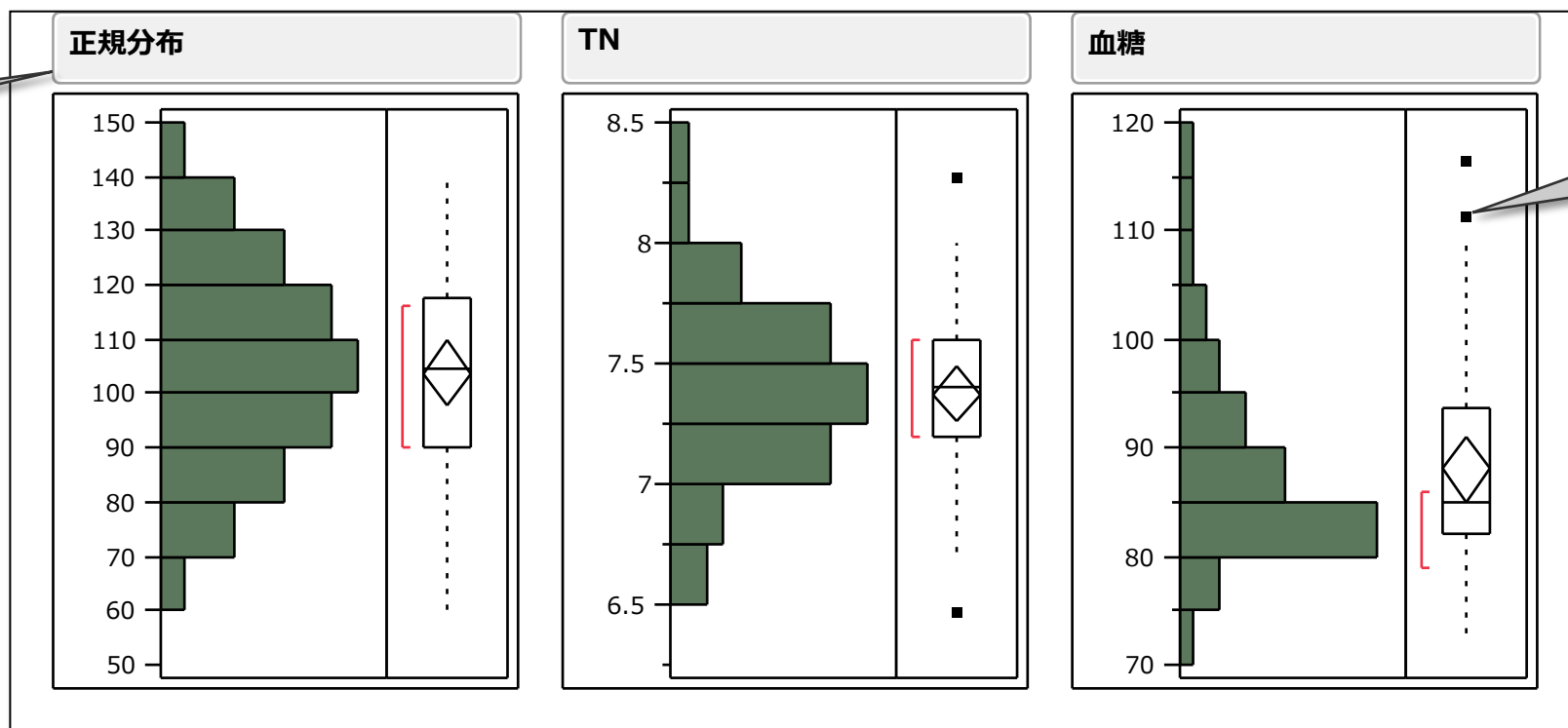
●JMPによる箱ひげ図の作成

中心から距離のあるデータ（外れ値）があるか

分布の形（1つの山か2つの山か等はわからない）

箱とひげの対称性（データの偏り）、データの集中の程度（広がり程度）

正規分布を追加



中心から離れたデータ



補足

グラフ間の連動
グラフとテーブルとの連動
データの抽出
出力結果とグラフのコピー&ペースト

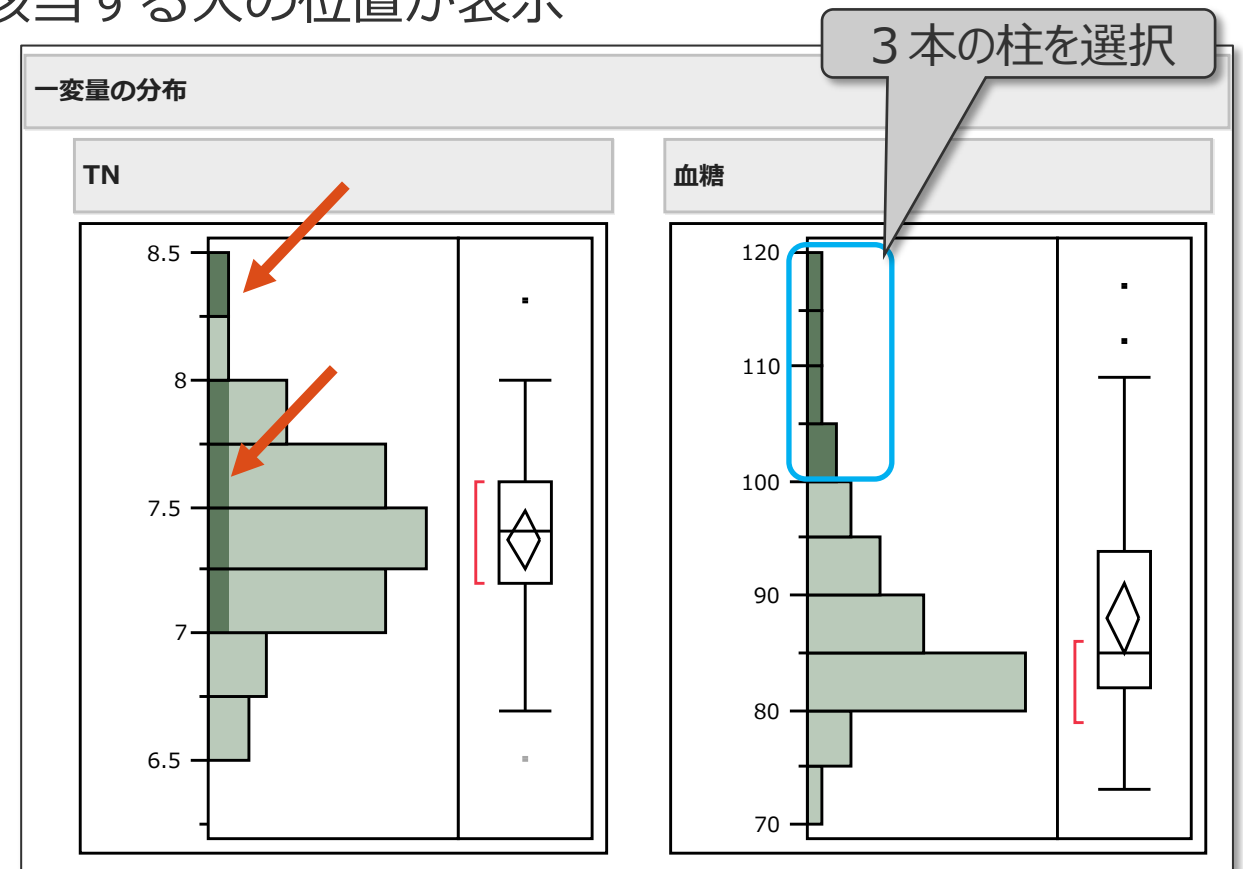
補足：JMP グラフ間の連動、グラフとテーブルとの連動

p.78

●ヒストグラムの相互関係、ヒストグラム→データテーブル

血糖値100以上の人の3本の柱をクリック（Ctrlキーを押しながら）して選択
TNのヒストグラムとデータテーブルに該当する人の位置が表示

	中性脂肪	TN	血糖	BMI
1	141	7.5	95	0.0026259
2	155	7.4	88	0.0027889
3	580	8.3	117	0.0026233
4	96	6.7	83	0.0025559
5	95	7.4	89	0.0020109
6	113	7.3	94	0.0020689
7	183	6.9	83	0.0022499
8	109	7.4	80	0.0022589
9	161	7.5	93	0.0025159
10	169	7.3	103	0.0024314



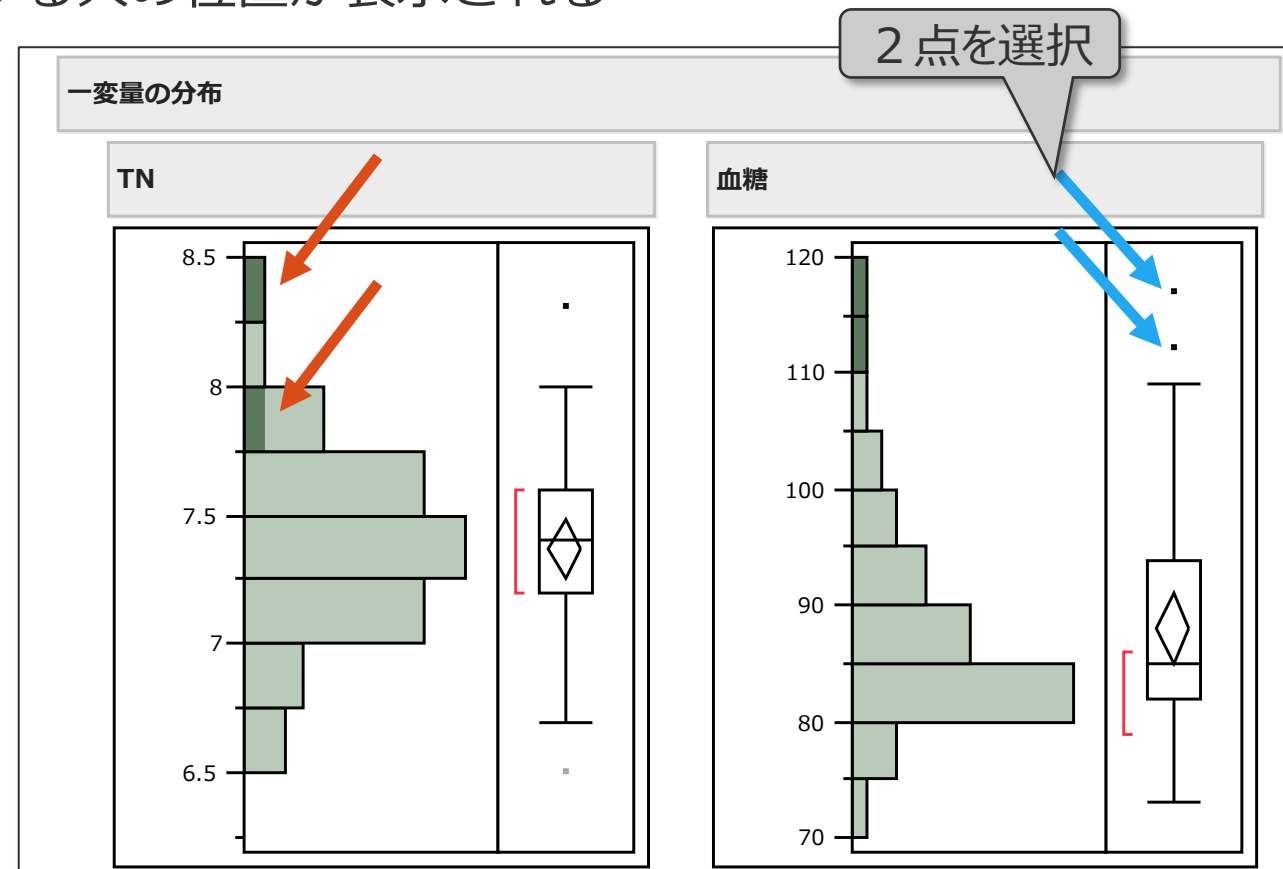
補足：JMP グラフ間の連動、グラフとテーブルとの連動

p.78

●箱ひげ図 → ヒストグラム・データテーブル

箱ひげ図の外れ値をクリックする、または「投げなわツール」で囲むと、ヒストグラムとデータテーブルに該当する人の位置が表示される

	中性脂肪	TN	血糖	BMI
1	141	7.5	95	0.0026259
2	155	7.4	88	0.0027885
3	580	8.3	117	0.0026230
4	96	6.7	83	0.0025559
5	95	7.4	89	0.0020109
6	113	7.3	94	0.0020685
7	183	6.9	83	0.0022491
8	109	7.4	80	0.0022581
9	161	7.5	93	0.0025153
10	169	7.3	103	0.0024314



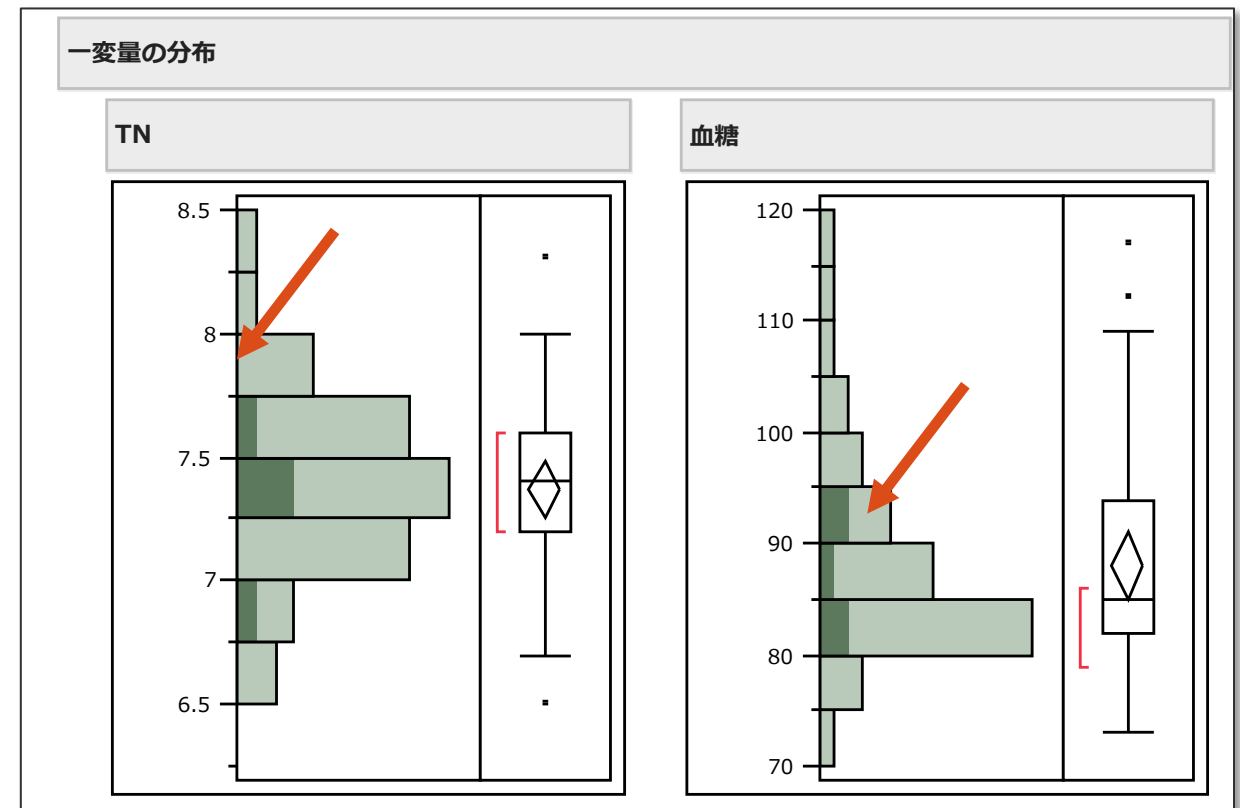
補足：JMP グラフ間の連動、グラフとテーブルとの連動

●データテーブル → グラフ

データテーブルで行をクリック（Ctrlキー、Shiftキーの利用）して選択すると、ヒストグラムに該当する人の位置が表示される

5行を選択

	中性脂肪	TN	血糖	BMI
1	141	7.5	95	0.0026259
2	155	7.4	88	0.0027885
3	580	8.3	117	0.0026230
4	96	6.7	83	0.0025559
5	95	7.4	89	0.0020109
6	113	7.3	94	0.0020685
7	183	6.9	83	0.0022495
8	109	7.4	80	0.0022582
9	161	7.5	93	0.0025150
10	169	7.3	103	0.0024314



補足：JMP グラフ間の連動、グラフとテーブルとの連動

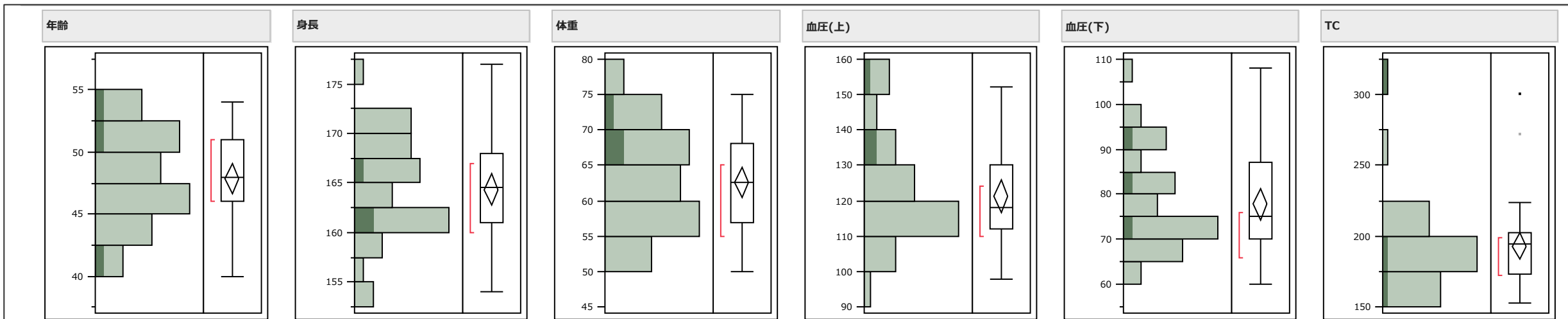
p.78

- [一変量の分布] で、データテーブルにある全てのデータのヒストグラムを表示

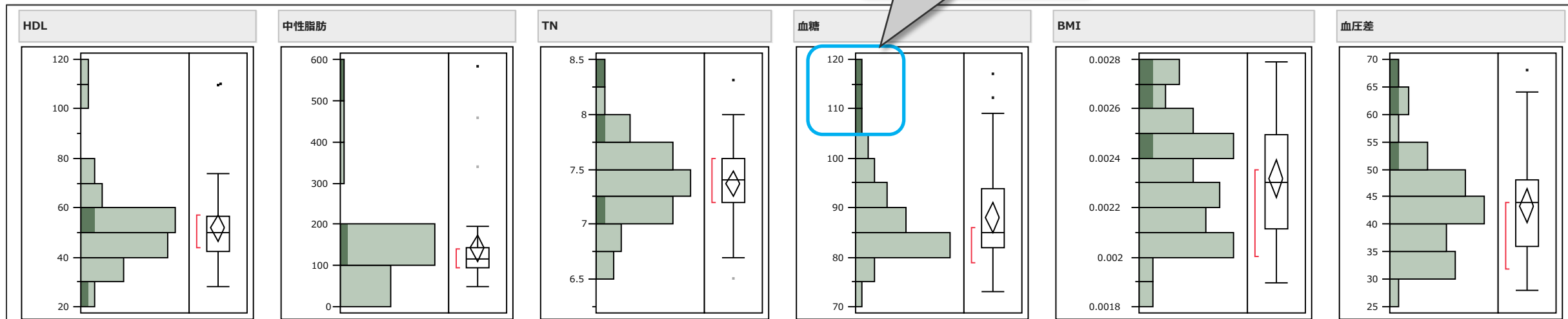
The screenshot shows the JMP interface with a data table and the '一変量の分布' dialog box. The data table has the following columns: No., 年齢, 身長, 体重. The dialog box is titled '一変量の分布 - JMP' and has a 'Y, 列' field containing 'No.', '年齢', '身長', and '体重'. A callout box points to the selection with the text 'No.~ 血圧差まで 全てを選択'.

No.	年齢	身長	体重
1	40	169	
2	41	164	
3	42	161	
4	43	169	
5	43	177	
6	43	166	
7	43	170	
8	44	163	
9	44	157	
10	46	166	
11	46	171	
12	46	162	
13	46	165	
14	46	161	
15	47	166	
16	47	161	

補足：JMP グラフ間の連動、グラフとテーブルとの連動



3本の柱を選択



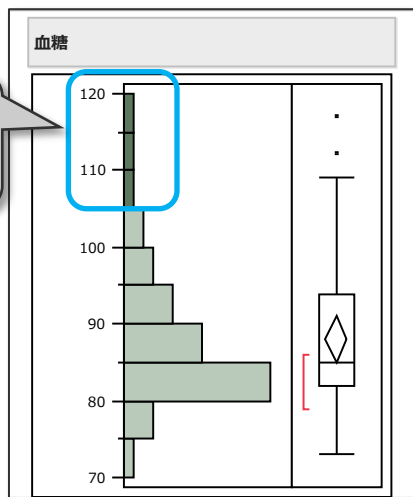
補足：JMP データの抽出

●ヒストグラムを利用した抽出

ヒストグラム、箱ひげ図などで、
特定のデータ（行）を選択し、
データテーブルに反映



行の赤い▼ボタン> [除外する/除外しない]



選択したデータが
データテーブルに
反映

中性脂肪

血糖

除外する/除外しない

表示しない/再表示

ラベルあり/ラベルなし

色

マーカー

次の選択行

前の選択行

行の選択

行の属性をクリア

列の値による色/マーカー分け...

行の編集

行の削除

血糖	値
88	0.0
117	0.0
83	0.0
89	0.0
94	0.0
83	0.0
80	0.0
93	0.0
103	0.0
96	0.0
81	0.0
90	0.0

●データテーブルから直接抽出

[行] >

[行の選択] >

[Where条件で選択]

The screenshot shows the JMP software interface with the 'Rows' menu open. The 'Where condition selection' option is highlighted, and a secondary menu is displayed. Red arrows indicate the navigation path from the 'Rows' menu to the 'Where condition selection' option.

体重	血压(上)	血压(下)	TC
75	126	78	200
75	104	70	197
68	152	92	300
73	120	92	199
63	112	64	183

補足：JMP データの抽出

●データテーブルから直接抽出

[Where条件で選択] のダイアログボックス

該当する列の変数を選択

特定の値を持つセルの選択

次と等しい、 次と等しくない

次より大きい、次を含めそれ以上

次より小さい、次を含めそれ以下

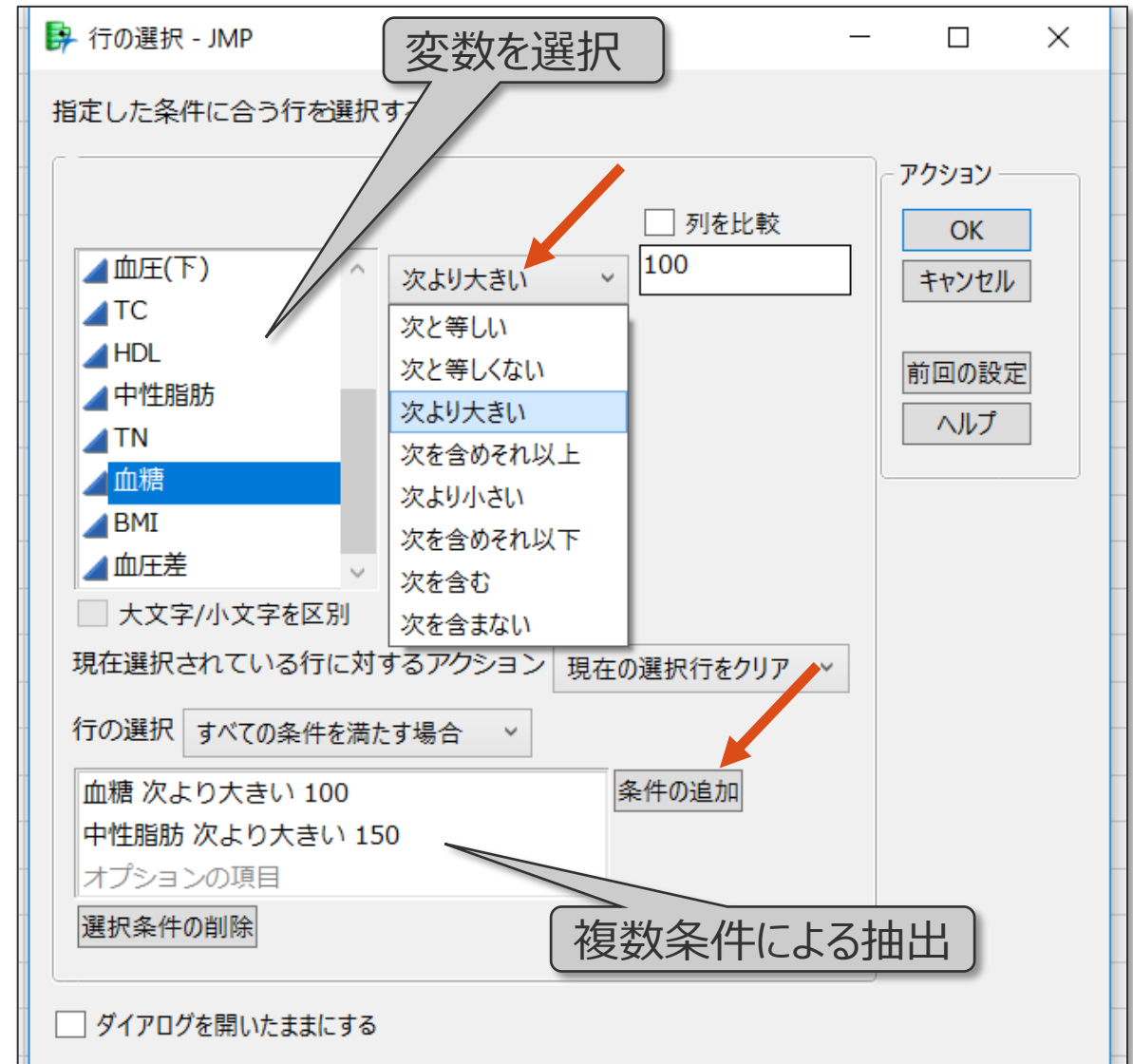
次を含む、 次をふくまない

[条件の追加] をクリック



データテーブルに戻り

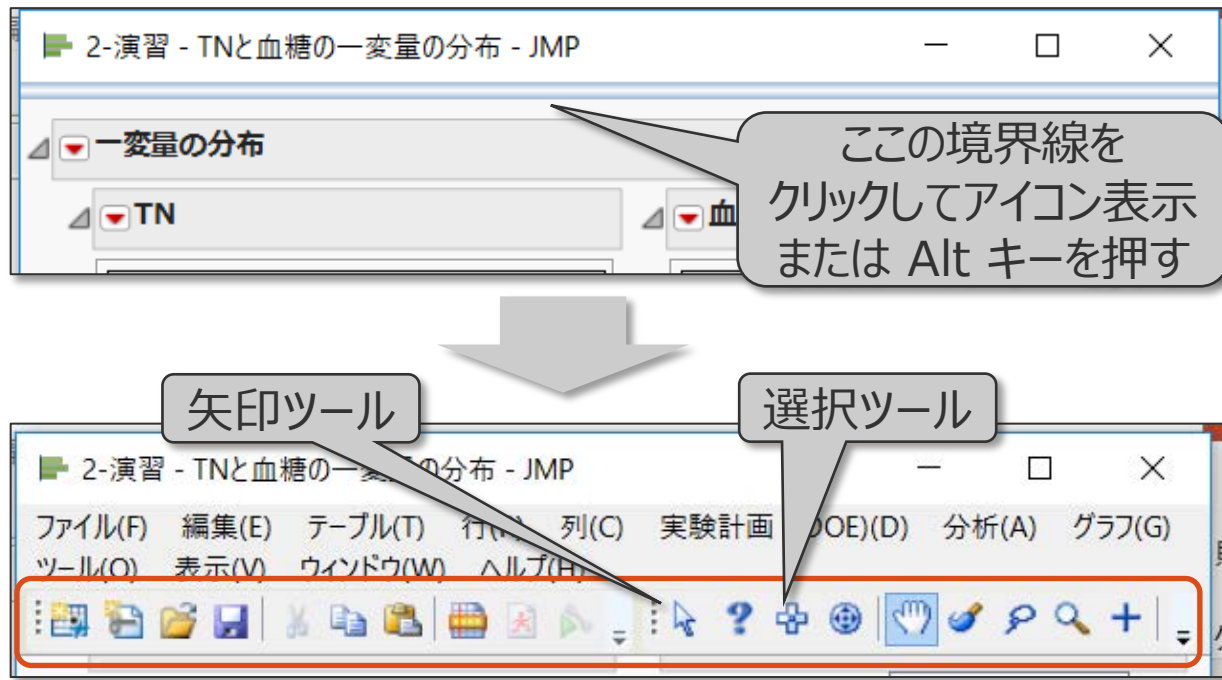
行の赤い▼ボタン> [除外する/除外しない]



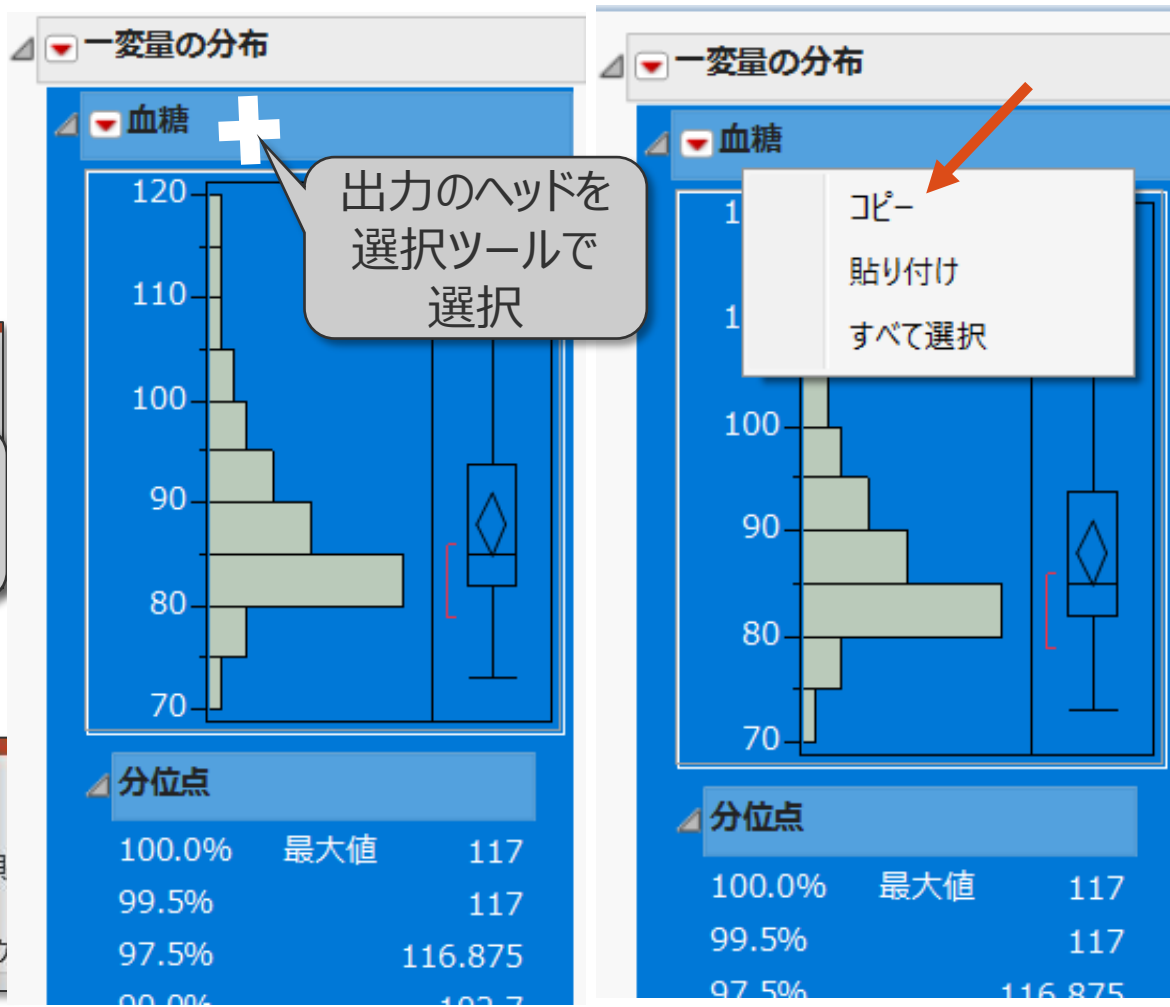
補足：JMP 出力結果とグラフのコピー&ペースト

●計算結果とグラフを JMP 以外にコピー

「選択ツール」を選択（太い十字のマーク）
変数名のあるヘッドの部分をクリック
→ 青く反転した部分を右クリック [コピー]
Excel、ワードにペースト



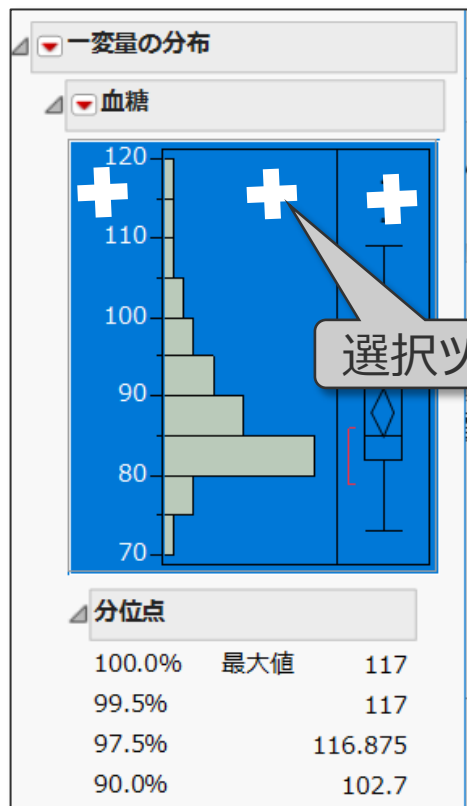
血糖に関する出力全部を選択



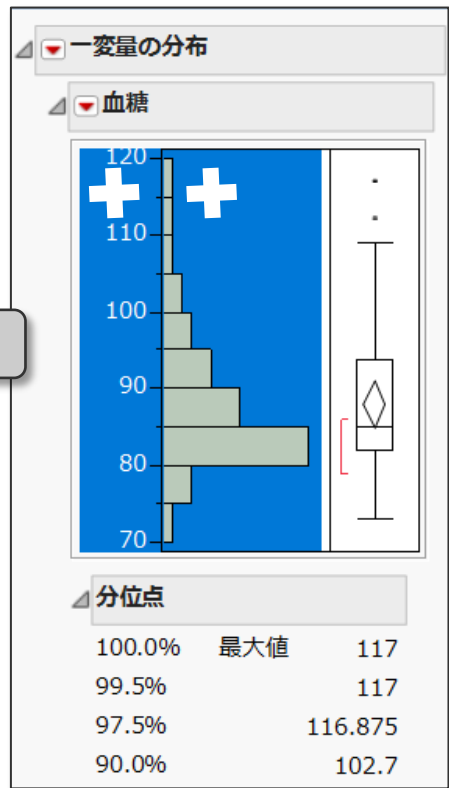
●計算結果とグラフを JMP 以外にコピー

出力の一部を「選択ツール」でクリック、複数の部分選択はCtrlを押しながらクリック

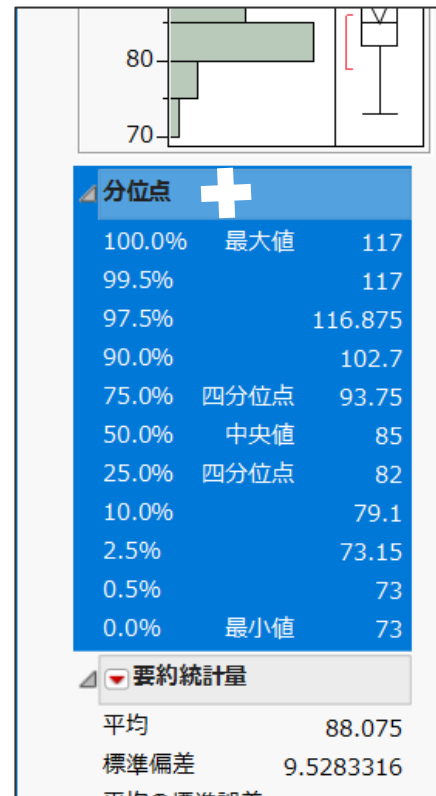
グラフの部分選択



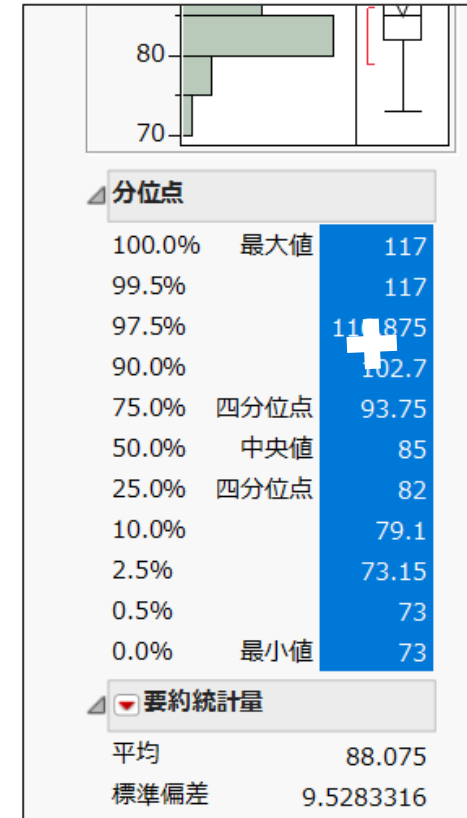
グラフの部分選択



表の部分選択



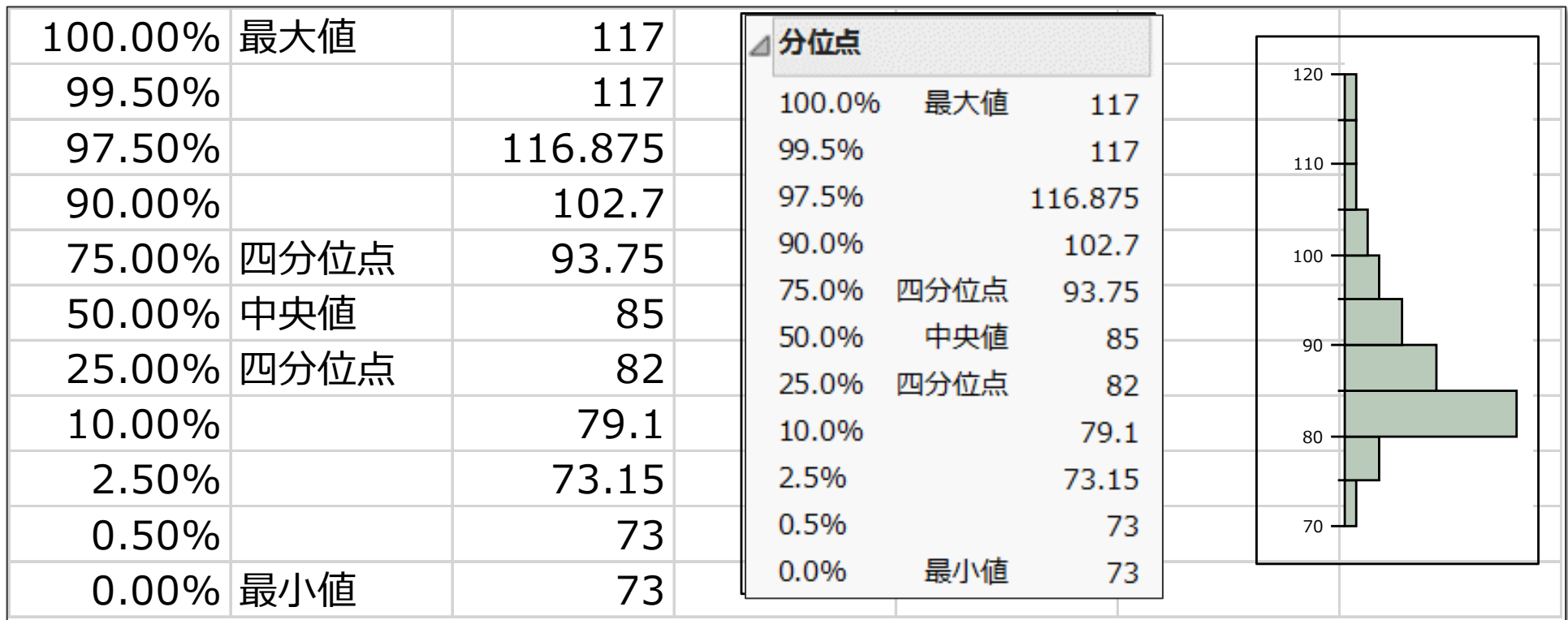
表の一部の部分選択



●計算結果とグラフを JMP 以外にコピー

[貼り付け] → 数値データ、文字データとして貼り付けられる (左)

[貼り付けのオプション] > [形式を選択して貼り付け] > [図 (拡張メタファイル)]
→ 図として貼り付けられる (中、右)





(4) 外れ値の棄却検定と外れ値の表示

●箱ひげ図の外れ値

箱の両端から箱の幅 W (四分位範囲) の 1.5 倍以上離れた値を外れ値とする

正規分布に従うと仮定して、 $\mu \pm 2.70\sigma$ を推定、この範囲を外れる値 ($\alpha = 0.006$)

$\mu \leftarrow$ 四分位範囲の midpoint で推定 $\sigma \leftarrow s^* = \text{四分位範囲} / 1.35$ で推定

外れ値に対して
頑健性がある ([§2.1](#))

●残差の偏差値の絶対値を利用

残差 e_i の偏差値の絶対値が基準値よりも大きい場合に外れ値とする

$$|u_i| = \frac{|e_i|}{\sigma} = \frac{|x_i - \bar{x}|}{\sigma}$$

例えば $|u_i| > 1.96$ ($\bar{x} \pm 1.96s$ から外れる値 $\alpha = 0.05$)

問題点

未知の σ を、サンプル (外れ値を含む) から推定した標準偏差 s で推定

残差 e_i の標準偏差 s は σ より小さくなる ([§2.1](#) p.116) $V[e_i] = ((n-1)/n)\sigma^2$

1 個の e_i を外れ値と誤る確率が $\alpha = 0.05$ でも、 n 個のサンプルだと誤る確率は高くなる

$n=100$ の場合, 少なくとも 1 個を外れ値だと誤る確率は $0.994 (= 1 - 0.95^{100}) \rightarrow$ 第 2 部

●箱ひげ図の外れ値

箱の両端から箱の幅 W (四分位範囲) の 1.5 倍以上離れた値を外れ値とする

正規分布に従うと仮定して、 $\mu \pm 2.70\sigma$ を推定、この範囲を外れる値 ($\alpha = 0.006$)

$\mu \leftarrow$ 四分位範囲の midpoint で推定 $\sigma \leftarrow s^* = \text{四分位範囲} / 1.35$ で推定

外れ値に対して
頑健性がある (§2.1)

●残差の偏差値の絶対値を利用

残差 e_i の偏差値の絶対値が基準値よりも大きい場合に外れ値とする

$$|u_i| = \frac{|e_i|}{\sigma} = \frac{|x_i - \bar{x}|}{\sigma}$$

例えば $|u_i| > 1.96$ ($\bar{x} \pm 1.96s$ から外れる値 $\alpha = 0.05$)

外れ値に影響される

問題点

未知の σ を、サンプル (外れ値を含む) から推定した標準偏差 s で推定

残差 e_i の標準偏差 s は σ より小さくなる (§2.1 p.116) $V[e_i] = ((n-1)/n)\sigma^2$

1 個の e_i を外れ値と誤る確率が $\alpha = 0.05$ でも、 n 個のサンプルだと誤る確率は高くなる

$n=100$ の場合, 少なくとも 1 個を外れ値だと誤る確率は $0.994 (= 1 - 0.95^{100}) \rightarrow$ 第 2 部

●箱ひげ図の外れ値 (方法 1)

箱の両端から箱の幅 W (四分位範囲) の 1.5 倍以上離れた値を外れ値とする

正規分布に従うと仮定して、 $\mu \pm 2.70\sigma$ を推定、この範囲を外れる値 ($\alpha = 0.006$)

$\mu \leftarrow$ 四分位範囲の midpoint で推定 $\sigma \leftarrow s^* = \text{四分位範囲} / 1.35$ で推定

外れ値に対して
頑健性がある (§2.1)

●残差の偏差値の絶対値を利用 (方法 2)

残差 e_i の偏差値の絶対値が基準値よりも大きい場合に外れ値とする

$$|u_i| = \frac{|e_i|}{\sigma} = \frac{|x_i - \bar{x}|}{\sigma}$$

例えば $|u_i| > 1.96$ ($\bar{x} \pm 1.96s$ から外れる値 $\alpha = 0.05$)

外れ値に影響される

問題点

未知の σ を、サンプル (外れ値を含む) から推定した標準偏差 s で推定

残差 e_i の標準偏差 s は σ より小さくなる (§2.1 p.116) $V[e_i] = ((n-1)/n)\sigma^2$

1 個の e_i を外れ値と誤る確率が $\alpha = 0.05$ でも、 n 個のサンプルだと誤る確率は高くなる

$n=100$ の場合, 少なくとも 1 個を外れ値だと誤る確率は $0.994 (= 1 - 0.95^{100}) \rightarrow$ 第 2 部

●外れ値を視覚的に判断する方法

箱ひげ図で外れ値を表示 (方法 1)

Excel、JMP を利用して表示させる
($\mu \pm 2.70\sigma$ の推定値から外れる値)

Excel [条件付き書式] を用いた

外れ値の表示 (方法 2)

外れ限界値 2.7 を入力

$\bar{x} \pm 2.7s$ から外れる値

以下を自動計算

データの平均と標準偏差

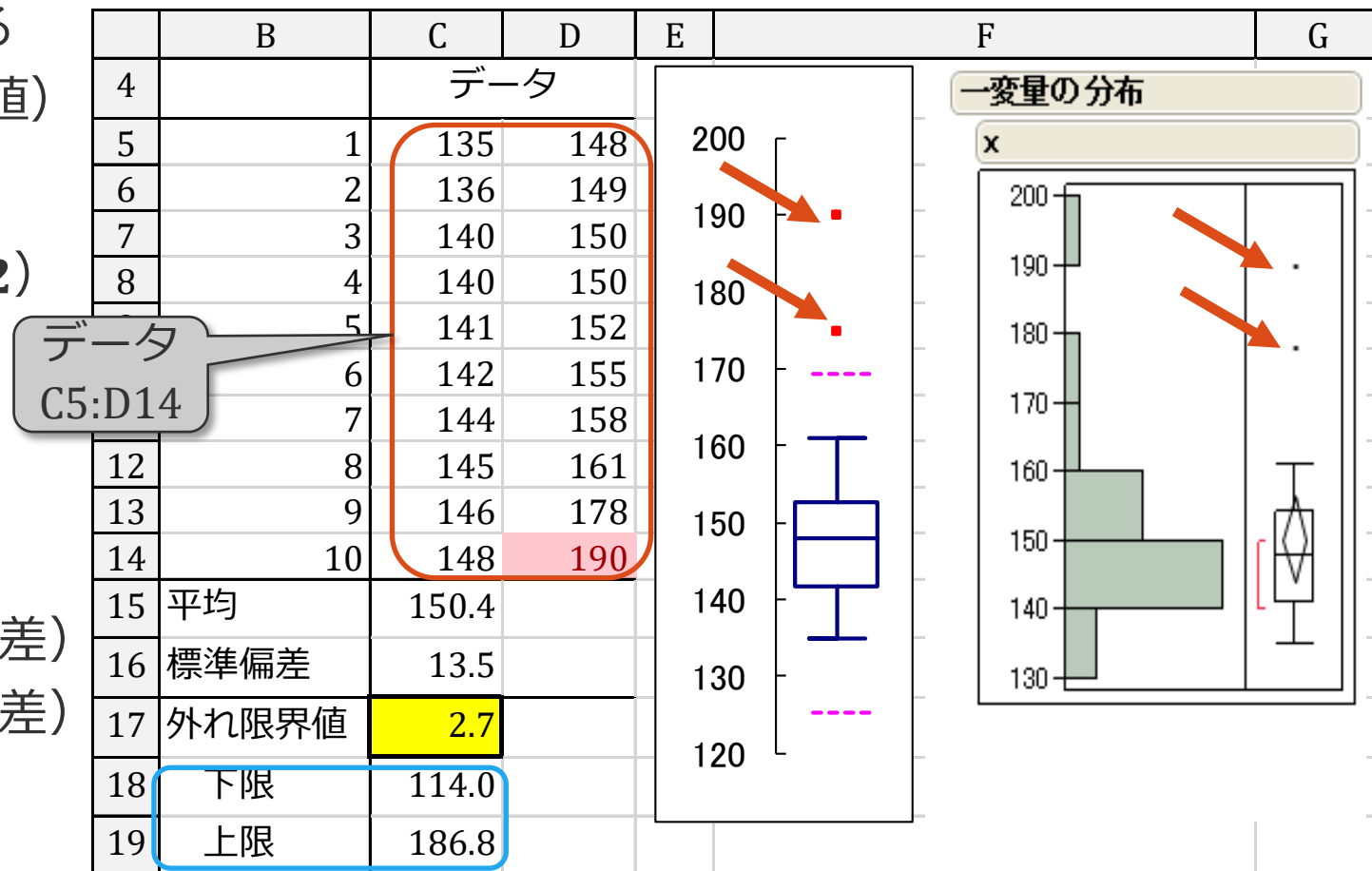
下限 (平均 - 外れ限界値 × 標準偏差)

上限 (平均 + 外れ限界値 × 標準偏差)

↓

[条件付き書式] の設定

表示 2.2.8 [条件付き書式] による外れ値表示



●外れ値を視覚的に判断する方法

箱ひげ図で外れ値を表示 (方法 1)

Excel、JMP を利用して表示させる
($\mu \pm 2.70\sigma$ の推定値から外れる値)

Excel [条件付き書式] を用いた

外れ値の表示 (方法 2)

外れ限界値 2.7 を入力

$\bar{x} \pm 2.7s$ から外れる値

全てのデータから以下を自動計算

データの平均と標準偏差

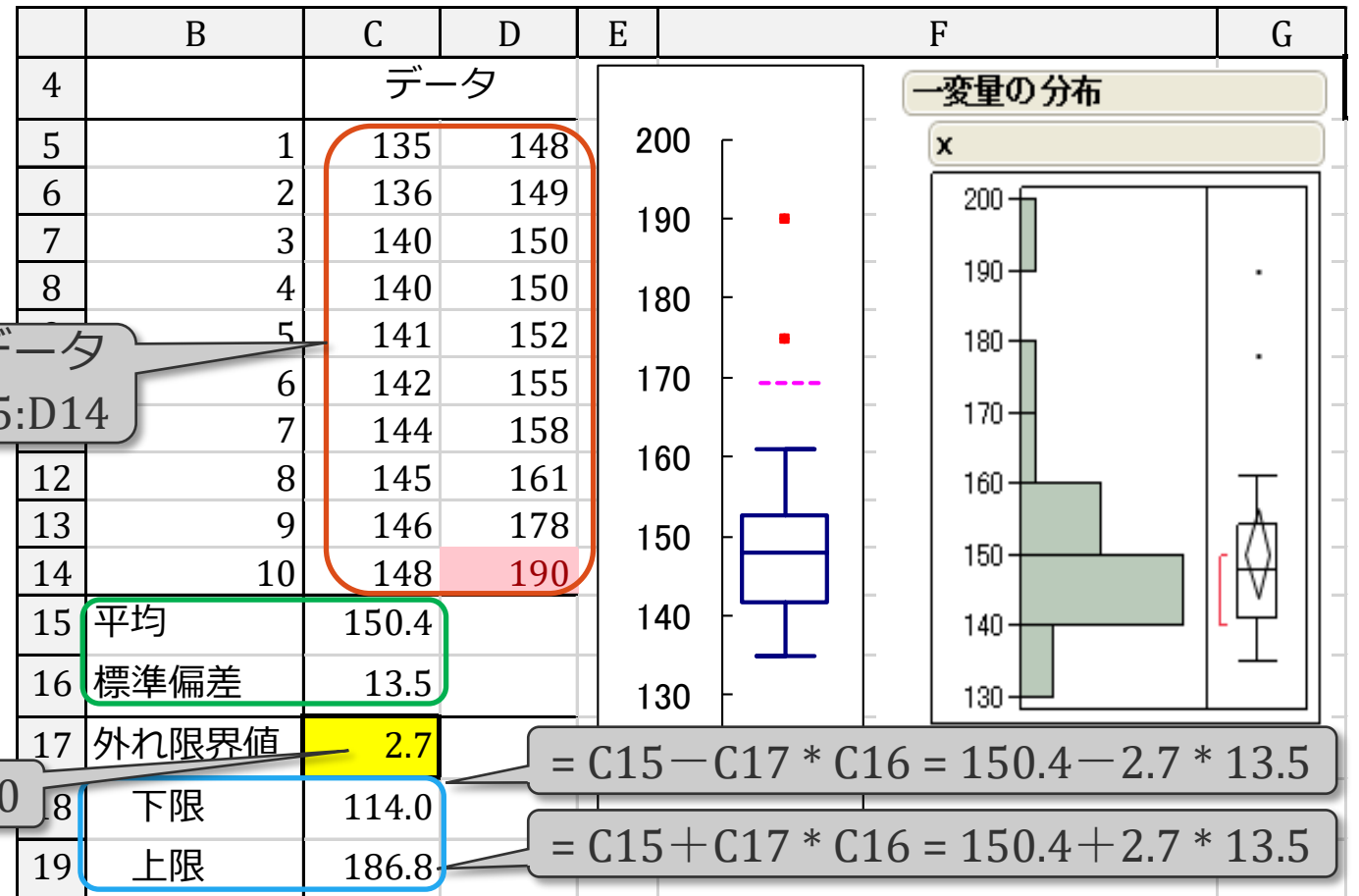
下限 (平均 - 外れ限界値 × 標準偏差)

上限 (平均 + 外れ限界値 × 標準偏差)

↓

[条件付き書式] の設定

表示 2.2.8 [条件付き書式] による外れ値表示



●Excel [条件付き書式] (方法 2)

データを範囲指定 C5:D14

[ホーム] > [スタイル] >

[条件付き書式] >

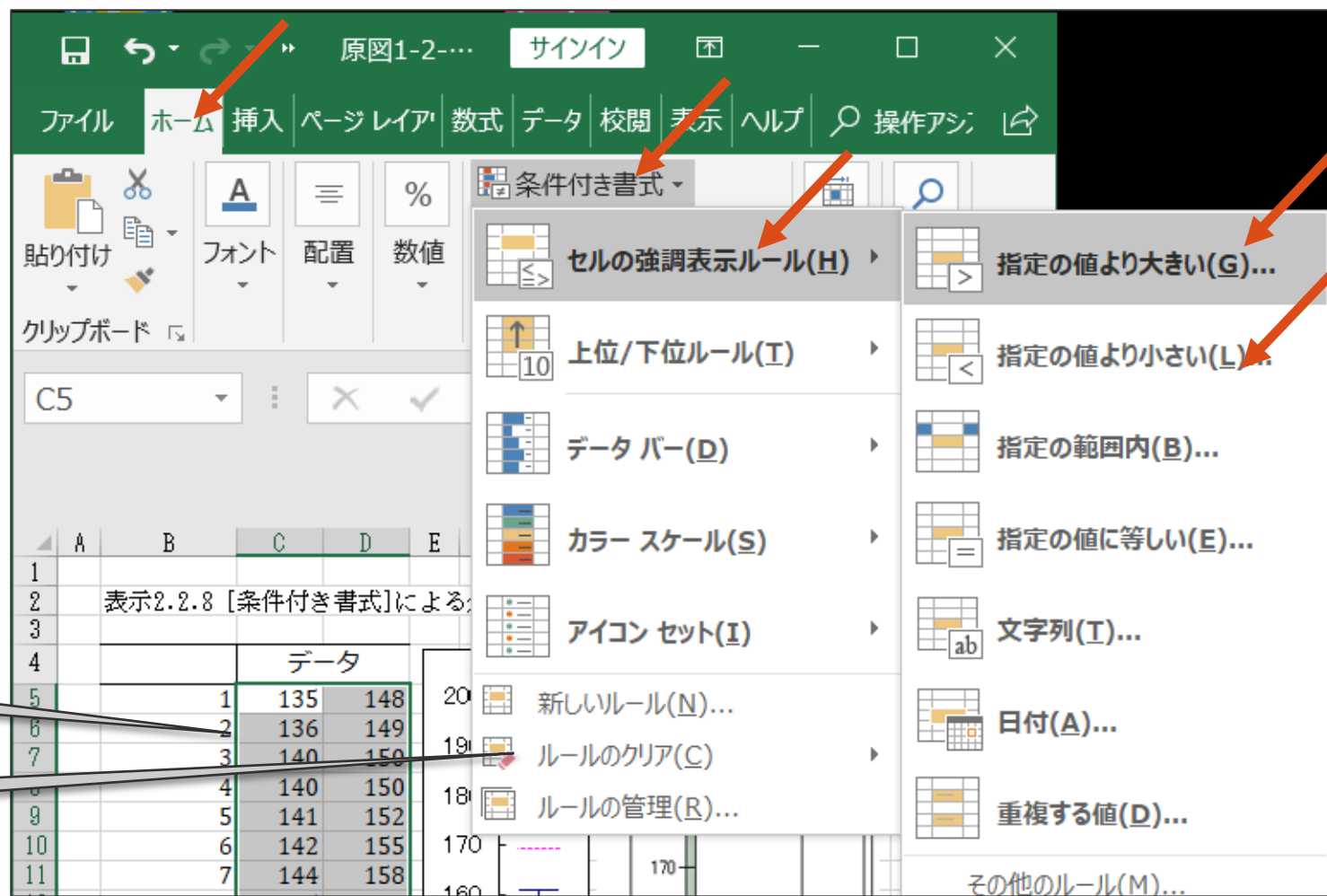
[セルの強調表示ルール] >

[指定の値より大きい]

[指定の値より小さい]

データ
C5:D14

条件付き書式をクリアー



外れ値の表示

●Excel [条件付き書式] (方法 2)

[指定の値より大きい]

上限のセルC19 を選択

書式 (強調の方法) を設定

[指定の値より小さい]

下限のセルC18 を選択

書式 (強調の方法) を設定

書式 (強調の方法)

	データ
	148
	149
	150
	150
	152
	155
	158
	161
	178
	190
平均	150.4
標準偏差	13.5
外れ限界値	2.7
下限	114.0
上限	186.8

● 箱ひげ図（方法 1）と Excel [条件付き書式]（方法 2）

箱ひげ図（方法 1）

外れ値が 2 つ

$$147.3 \pm 2.70 \times 8.15 = [125.3, 169.3]$$

頑健性のある推定値を使用

幅 44

Excel [条件付き書式]（方法 2）

外れ値が 1 つ

$$150.4 \pm 2.70 \times 13.5 = [114.0, 186.8]$$

外れ値の影響を受けた推定値を利用
 平均値と標準偏差の推定値が大
 上限と下限の幅がやや広い

幅 73

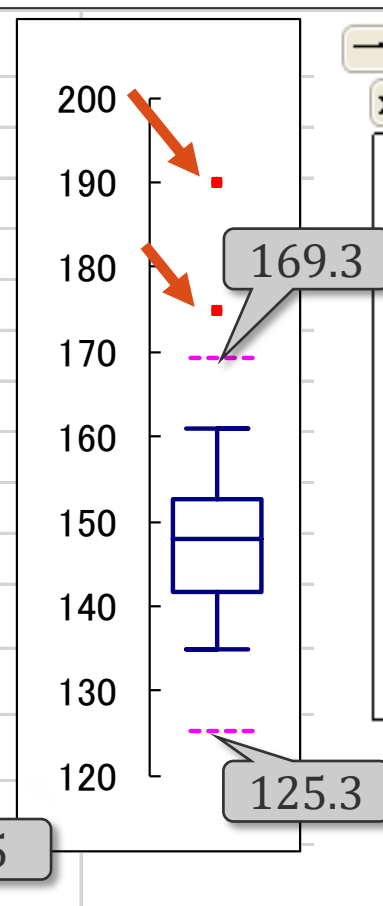
外れ限界値を低くすることも一手段

$$150.4 \pm 2.50 \times 13.5 = [116.7, 184.1]$$

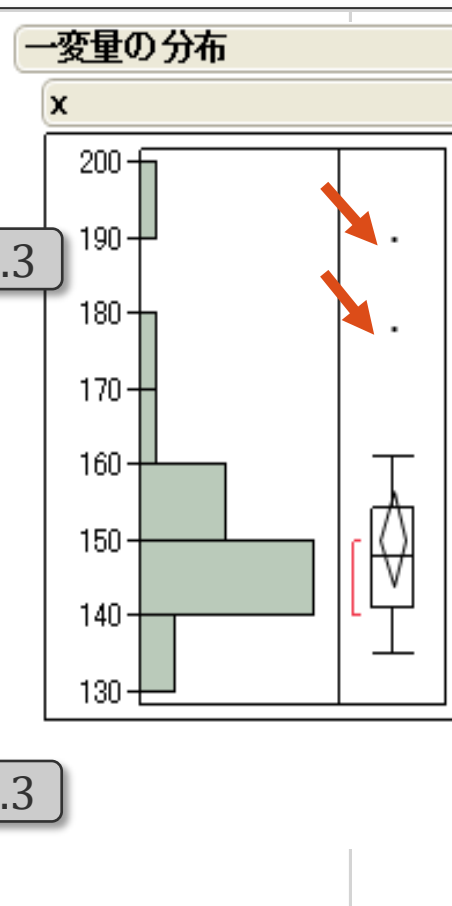
Excel [条件付き書式]

	データ	
1	135	148
2	136	149
3	140	150
4	140	150
5	141	152
6	142	155
7	144	158
8	145	161
9	146	178
10	148	190
平均	150.4	
標準偏差	13.5	
外れ限界値	2.7	
下限	114.0	
上限	186.8	

箱ひげ図



箱ひげ図





●「外れ値」と「異常値」

全体の分布から離れているということだけで、異常値と断定してはならない
統計理論だけから異常とは判断できない
機械的に外れ値を除外してはならない
(Smirnov の棄却検定のみでデータを機械的に削除してはならない)

●薬効、毒性データ

特異体質、併用薬の影響等の原因による場合がある
外れ値を示した個々の実験対象から、新しい知見が得られる可能性もある

●正規分布を前提とした外れ値の判断

分布が正規分布から離れていると外れ値が出やすい (箱ひげ図で多数の外れ値が出現)
正規分布になるように変換してから考える (→次節 [§2.3](#))

- データが得られたら、まずグラフに表して丁寧に観察
- ヒストグラム
 - 区間の幅と区切り値（境界値）を設定 → 度数表
 - この2つパラメータで度数表が大きく変化する
 - 1つのヒストグラムで判断してはならない
 - JMPのグラフ機能は極めて有効
- 箱ひげ図
- これらのグラフを自由に操ることができる→データ解析のプロの必要条件



- 作成 片瀬雅彦
- 監修 松本一彦、長谷文雄
- 作成時期 2018年10月12日
- 改訂 2019年3月10日、2022年5月10日
2024年9月9日