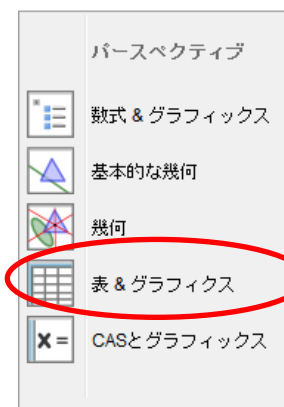
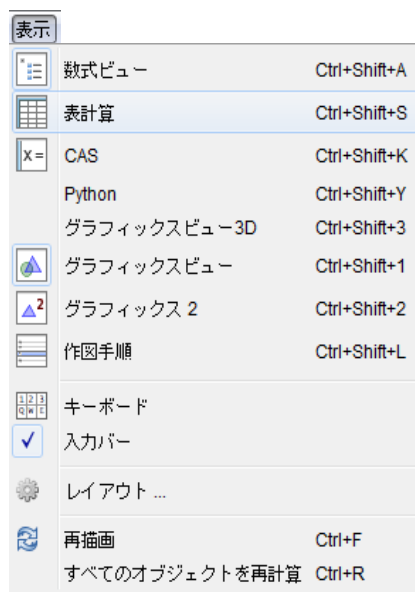


Section2：統計学

GeoGebra は、表計算の為の画面があり、統計関数を取り扱う機能があります。
また、数多くの統計のグラフのオプションも備えています。

表計算の画面を表示するには、方法が2通りあります。

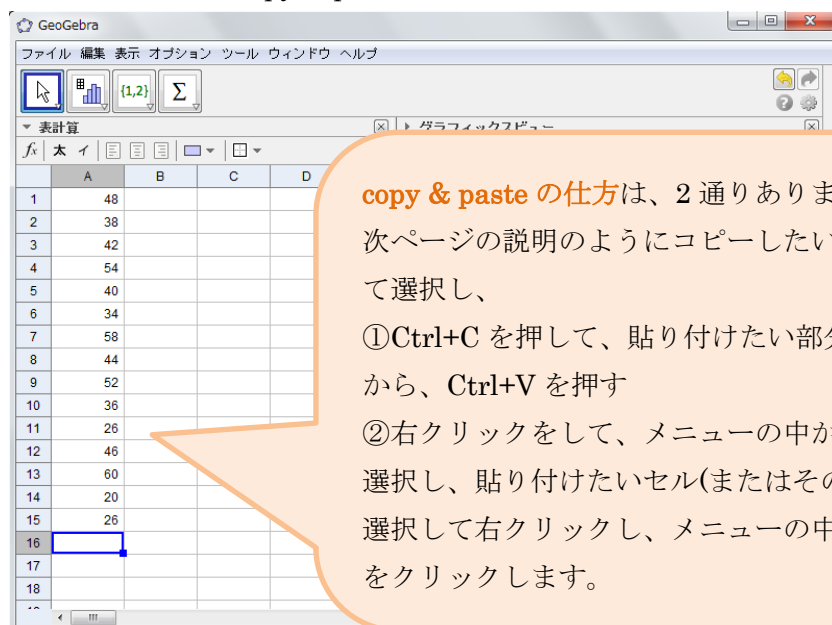
- ① GeoGebra を開いた時に現れるメニューから全体像を選ぶときに、**表&グラフィックス**を選択する方法
- ② 表示メニューの中から、**表計算**をクリックする方法
(**Ctrl+Shift+S** でも同様)



2.1 表と中心傾向を計る

例. 下の数字は、60点満点のある数学のテストの成績です。
GeoGebra を使って、平均値、中央値、最頻値を見つけましょう。
48 38 42 54 40 34 58 44 52 36 26 46 60 20 26

- ① まず最初に、データを入力します。データの数値を表の最初の列に入力します。
Excel からデータを copy & paste することも出来ます。

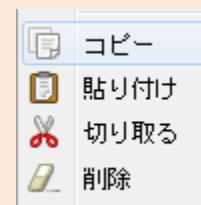


copy & paste の仕方は、2通りあります。

次ページの説明のようにコピーしたいデータをマウスでドラッグして選択し、

① **Ctrl+C** を押して、貼り付けたい部分の最初のセルをクリックしてから、**Ctrl+V** を押す

② 右クリックをして、メニューの中から **コピー** を選択し、貼り付けたいセル(またはその最初のセル)を選択して右クリックし、メニューの中から **貼り付け** をクリックします。



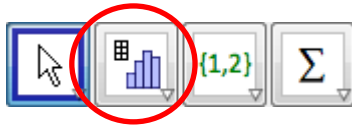
② マウスでドラッグして、データを選択します。

最初のセルで左クリックし、

押したままマウスを下げます。

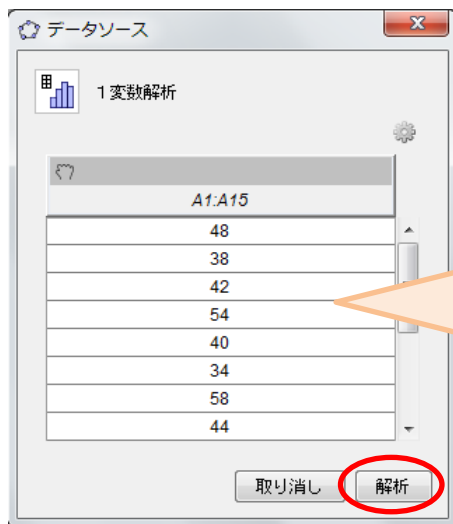
最後のセルで押していたボタンを離します。

③ ツールの中から、1変数解析のアイコンをクリックします。



④ 新しくデータソースの画面が表示されるので、

解析ボタンをクリックします。



表計算		
	A	B
1	48	
2	38	
3	42	
4	54	
5	40	
6	34	
7	58	
8	44	
9	52	
10	36	
11	26	
12	46	
13	60	
14	20	
15	26	
16		

もし、長方形領域でデータを選択した場合には、

	A	B	C	D
1	① 48	② 40	③ 52	④ 60
2	38	34	36	20
3	42	58	26	26
4	54	44	46	

の順番で1変数として読み込まれます。

すると、次のような画面が表示されます。

データソースを表示することが出来ます。

このアイコンをクリックすると、詳細な測定値を知ることが出来、平均値、中央値を見ることが出来ます。

ドロップダウンリストの中から選択すれば、他の表でも表示することが出来ます。

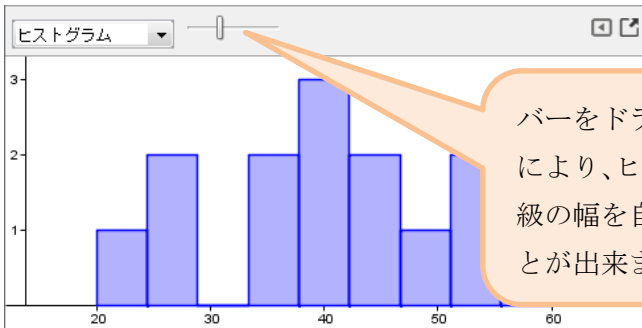
オプションを開いて、詳細を設定することが出来ます。

GeoGebra で表示できる表

ドロップダウンリストの中から選択すれば、様々な表でデータを表示することができます。

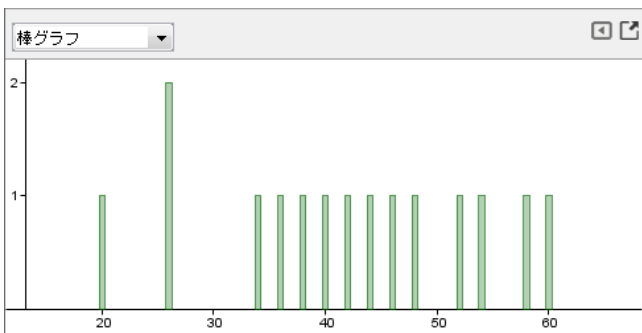
- ヒストグラム
- 棒グラフ
- 箱ひげ図
- ドットプロット
- 幹葉表示
- 正規確率プロット

ヒストグラム

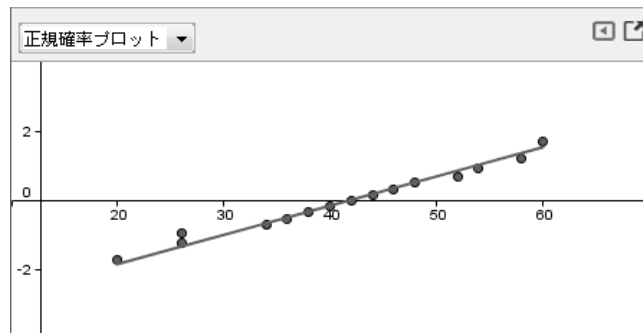


正規確率プロットは、データが正規分布に従っているかどうかを知るために使われます。多くの統計手法ではデータが正規分布に従っていることを仮定しているので、正規確率プロットは、データが正規分布に従っているということを立証したり、誤った仮定に対して警告を与えたりします。

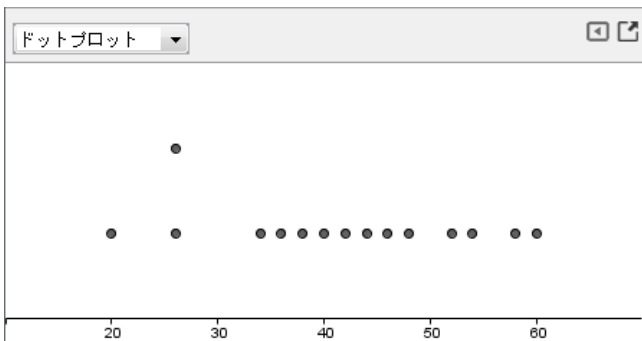
棒グラフ



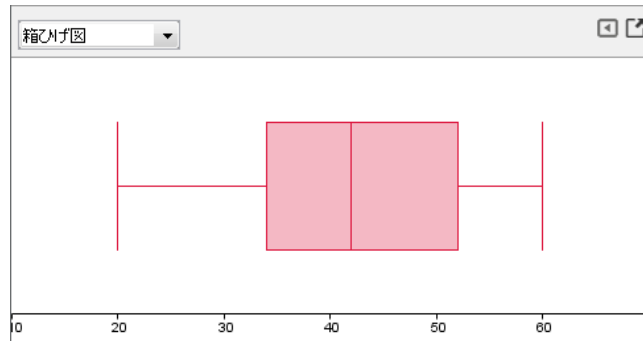
正規確率プロット



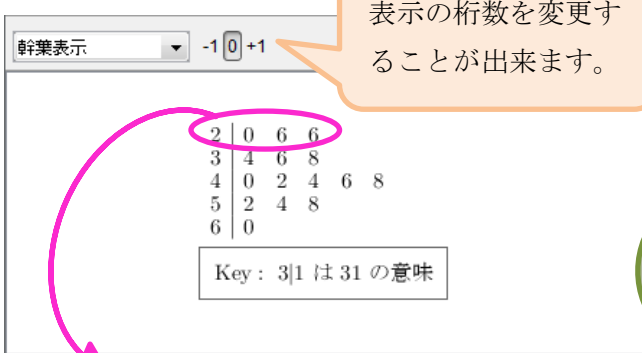
ドットプロット



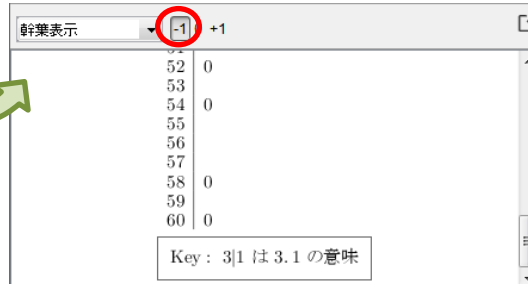
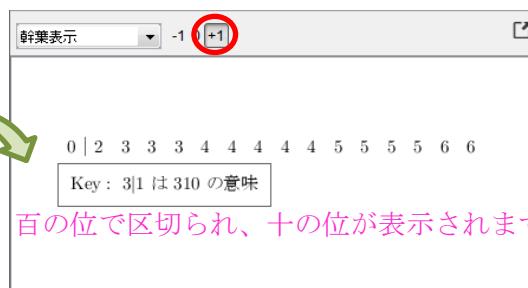
箱ひげ図




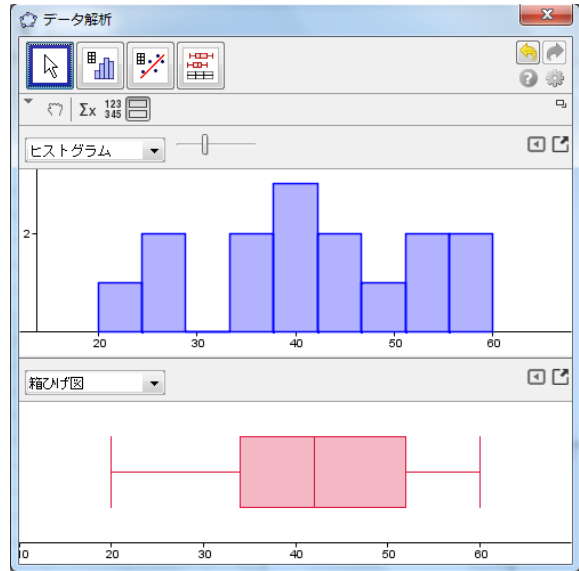
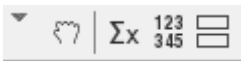
幹葉表示



20代が、20,26,26であることを示しています。



2種類の表を比較する時には、
メニューから  をクリックして下さい。



より詳細な統計量を知るためには、
メニューから Σx をクリックして下さい。

統計	
n	15
平均	41.6
σ	11.5285
s	11.9331
Σx	624
Σx^2	27952
最小	20
Q1	34
中央値	42
Q3	52
最大	60

n : データの個数(度数)

平均 : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

σ : 標準偏差 ($\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$)

s : 標本標準偏差 ($s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$)

ΣX : データの総和

ΣX^2 : データの2乗の和


最小 : データの最小値

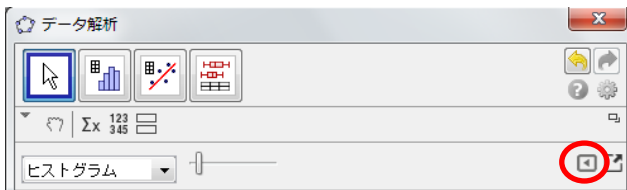
Q1 : 第一四分位数

中央値 : データの中央値

Q3 : 第三四分位数

最大 : データの最大値

より詳細なヒストグラムのオプションを設定するには、 のアイコンをクリックして下さい。



階級をこのメニューから、手入力で変更することが出来ます。

ヒストグラム 開始 0 幅 5

ヒストグラム グラフ

Classes

Set Classes Manually

Class Rule:

$\leq x <$

$< x \leq$

度数のタイプ

累積

Count

相対

正規化された

表示

ヒストグラム

度数分布表

度数多角形

正規分布曲線

標準正規分布の
確率密度関数に
最も近くなるよ
うに縦軸を取り
ます。

折れ線グラフを表
示します。

2.2 散布図と回帰直線

例. 右の表はプレトリアで2002年から2008年にかけて始まった新しい企業の数を表しています。より近い直線を作図しましょう。

年	新しい企業の数
2002	754
2003	881
2004	943
2005	1083
2006	1182
2007	1304
2008	1402

- ① まず最初に、データを入力します。
データを表の最初の列に入力しましょう。
Excel からデータを Copy & Paste することも出来ます。

- ② データをドラッグして選択します。

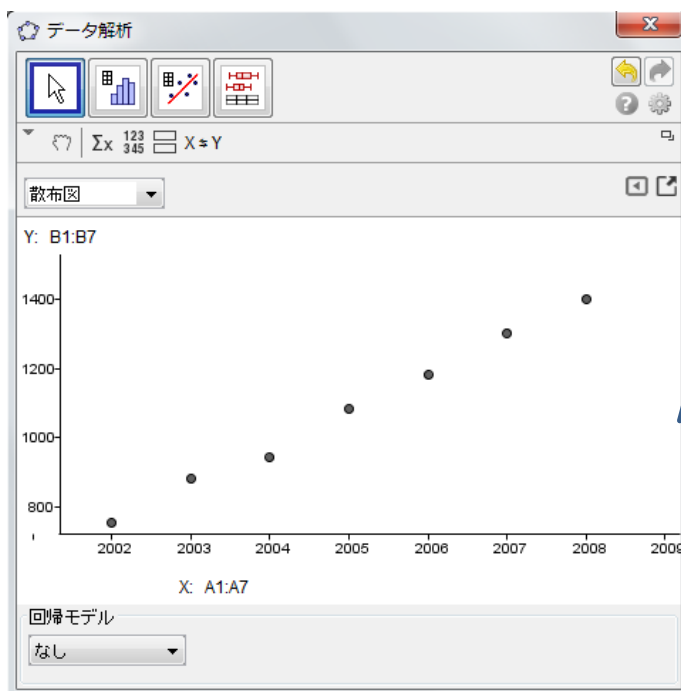
- ③ 表計算のメニューから  の右下の▽をクリックします。

- ④ その中から、2変量回帰分析を選択します。



	A	B
1	2002	754
2	2003	881
3	2004	943
4	2005	1083
5	2006	1182
6	2007	1304
7	2008	1402

- ⑤ 新しく現れる画面から、右下の解析をクリックします。
すると、自動的にこの画面が表示されます。



⑥ 回帰モデルの下のドロップダウンリストから、線型を選択しましょう。

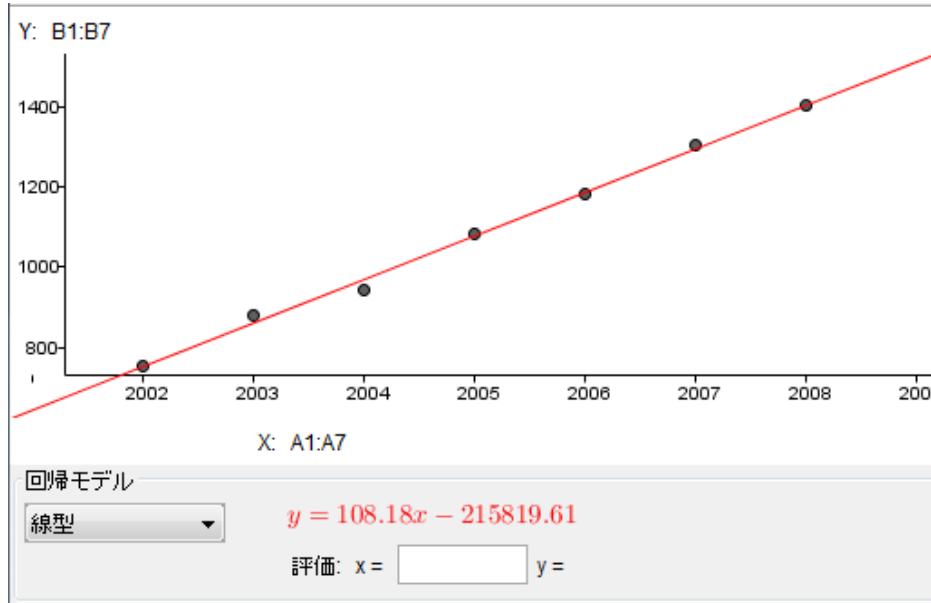
すると、次のような画面になります。

より近い直線の方程式が下に表示されます。

xの値を入力すれば、その直線の方程式に代入したyの値が表示されます。

例)この例の数値を用いるなら、x = 30を入力すれば、

30年後の新しい企業の数を実験することが出来ます。



線型	$y = ax + b$
対数	$y = a + b \log x$
多項式	$y = ax^2 + bx + c$
ベキ	$y = ax^b$
指数	$y = ae^{bx}$
Growth	$y = a \cdot b^x$
Sin	$y = a + b \sin(cx + d)$
ロジスティック	$y = -\frac{a}{1 - be^{-cx}}$

相関係数を求める

相関係数を求めるには、まず上の要領で散布図まで描きます。

次に、 Σx のアイコンをクリックすると、下のように統計が表示されます。

統計	
MeanX	2005
MeanY	1078.4286
Sx	2.1602
Sy	234.1302
r	0.9981
ρ	1
Sxx	28
Syy	328901.7143
Sxy	3029
RSquare	0.9963
SSE	1228.8214

MeanX : X の平均値

MeanY : Y の平均値

Sx : X の標準偏差

Sy : Y の標準偏差

r : 相関係数

ρ : 有意確率(一致している確率)

Sxx : X の分散

Syy : Y の分散

Sxy : 共分散


箱ひげ図の比較をする

例. 次のデータは、2010年度の東京と那覇、札幌における月ごとの平均気温を、気温の低い順にならべたものです。(単位は℃)このデータの箱ひげ図を並べてかき、3つの都市のデータの分布を比較しましょう。

東京 6.5 7.0 9.1 9.9 12.4 13.5 18.9 19.0 23.6 25.1 28.0 29.6

那覇 16.8 18.1 18.3 19.9 21.2 21.4 23.8 25.7 26.7 28.0 28.7 28.9

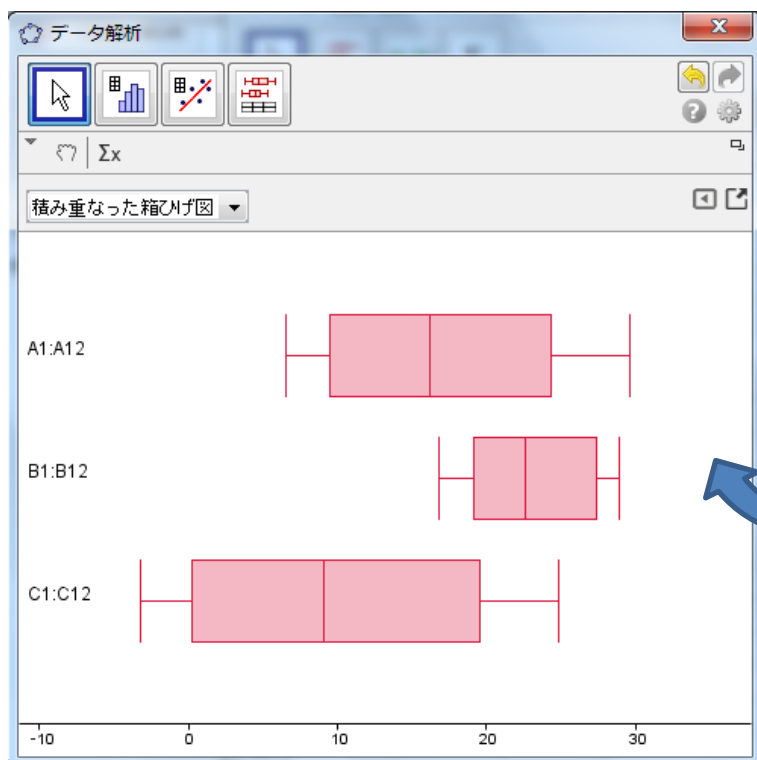
札幌 -3.2 -2.0 -0.1 0.6 5.5 5.9 12.2 12.2 19.2 20.0 22.1 24.8

- ① まず最初に、データを入力します。
データを表の最初の列に入力しましょう。
Excel からデータを Copy & Paste することも出来ます。
- ② データをドラッグして選択します。
- ③ 表計算のメニューから  の右下の▽をクリックします。
- ④ その中から、**多変量解析**を選択します。

	A	B	C
1	6.5	16.8	-3.2
2	7	18.1	-2
3	9.1	18.3	-0.1
4	9.9	19.9	0.6
5	12.4	21.2	5.5
6	13.5	21.4	5.9
7	18.9	23.8	12.2
8	19	25.7	12.2
9	23.6	26.7	19.2
10	25.1	28	20
11	28	28.7	22.1
12	29.6	28.9	24.8
13			



- ⑤ 新しく現れる画面から、右下の解析をクリックします。
すると、自動的にこの画面が表示されます。



◎3つのデータのそれぞれの統計量を、それぞれ1変数解析しなくても、まとめて見ることが出来ます。

上の要領で多変数解析を行い、次に Σx のアイコンをクリックします。
すると、それぞれの変数の統計が表示されます。

統計									
	n	平均	σ	s	最小	Q1	中央値	Q3	最大
A1:A12	12	16.8833	7.9207	8.2729	6.5	9.5	16.2	24.35	29.6
B1:B12	12	23.125	4.2198	4.4075	16.8	19.1	22.6	27.35	28.9
C1:C12	12	9.7667	9.5951	10.0217	-3.2	0.25	9.05	19.6	24.8

n : データの個数(度数)

平均 : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

σ : 標準偏差 ($\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$)

s : 標本標準偏差 ($s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$)

最小 : データの最小値

Q1 : 第一四分位数

中央値 : データの中央値

Q3 : 第三四分位数

最大 : データの最大値