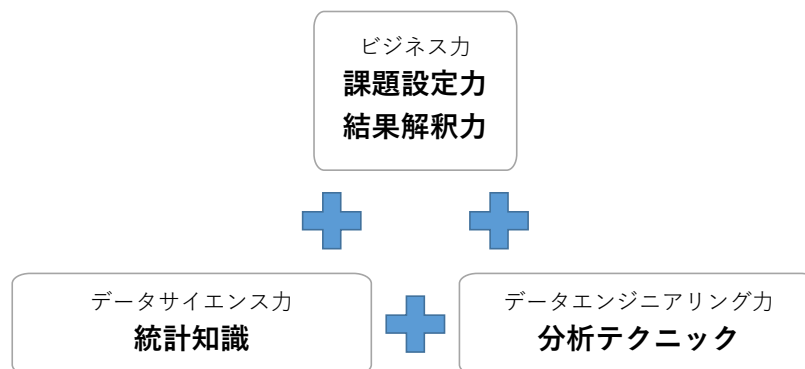


# 統計基礎

## データを扱う際に必要となるスキル

### □ （一社）データサイエンティスト協会による定義

- 「ビジネス力」「データサイエンス力」「データエンジニアリング力」  
ーデータ分析の成否の決め手は「ビジネス力」



# データサイエンス力として持っておきたい知識

## □ 以下のような項目の理解は必須

- ーデータの種類（量的変数、質的変数）
- ー基本的なグラフ[棒グラフ・折れ線グラフ・円グラフなど]の見方・読み方
- ーデータの集計（1変数データ、2変数データ、度数分布表、ヒストグラム[柱状グラフ]）
- ーデータの代表値（平均値・中央値）
- ーデータの散らばりの指標（分散、標準偏差、はずれ値）
- ークロス集計表（2次元の度数分布表、行比率、列比率）
- ー相関と回帰（散布図、擬相関、相関係数、相関と因果、回帰直線）

## □ 統計検定4～3級の内容におおむね合致

# 代表値

## □ 代表値とは、データの概要を把握するために有効な指標

配布先	配り方①	配り方②	配り方③
A	5	7	10
B	5	7	8
C	5	7	8
D	5	7	6
E	5	7	6
F	5	3	3
G	5	3	3
H	5	3	3
I	5	3	3
J	5	3	0
合計	50	50	50
平均値	5	5	5
中央値	5	5	4.5
最頻値	5	7	3

# 平均値

平均値は最も基本的でかつ重要な統計指標

	配布先	配り方		
		配り方①	配り方②	配り方③
上級生	A	5	7	10
	B	5	7	8
	C	5	7	8
	D	5	7	6
	E	5	7	6
下級生	F	5	3	3
	G	5	3	3
	H	5	3	3
	I	5	3	3
	J	5	3	0
合計		50	50	50
平均値		5	5	5

Excel関数    =average(範囲)  
=sum(範囲) / count(範囲)

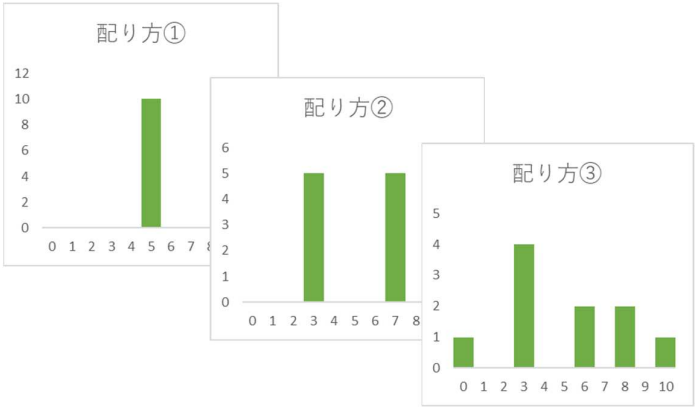
50個入りのお土産  
10人への配り方

# 中央値

昇順（または降順）に並べたときに真ん中にくる値

	配布先	配り方		
		配り方①	配り方②	配り方③
上級生	A	5	7	10
	B	5	7	8
	C	5	7	8
	D	5	7	6
	E	5	7	6
下級生	F	5	3	3
	G	5	3	3
	H	5	3	3
	I	5	3	3
	J	5	3	0
合計		50	50	50
平均値		5	5	5
中央値		5	5	4.5

Excel関数    =median(範囲)  
※Excelで計算する場合は並べ替える必要はない。



# 最頻値

最頻値は多数派の集団を抽出するために利用

	配布先	配り方① 配り方② 配り方③		
		5	7	10
上級生	A	5	7	10
	B	5	7	8
	C	5	7	8
	D	5	7	6
	E	5	7	6
下級生	F	5	3	3
	G	5	3	3
	H	5	3	3
	I	5	3	3
	J	5	3	0
合計		50	50	50
平均値		5	5	5
中央値		5	5	4.5
最頻値		5	7	3

Excel関数

=mode(範囲)  
※同数のデータがある場合ははじめて出てきた値を返します。

# 代表値から分かること

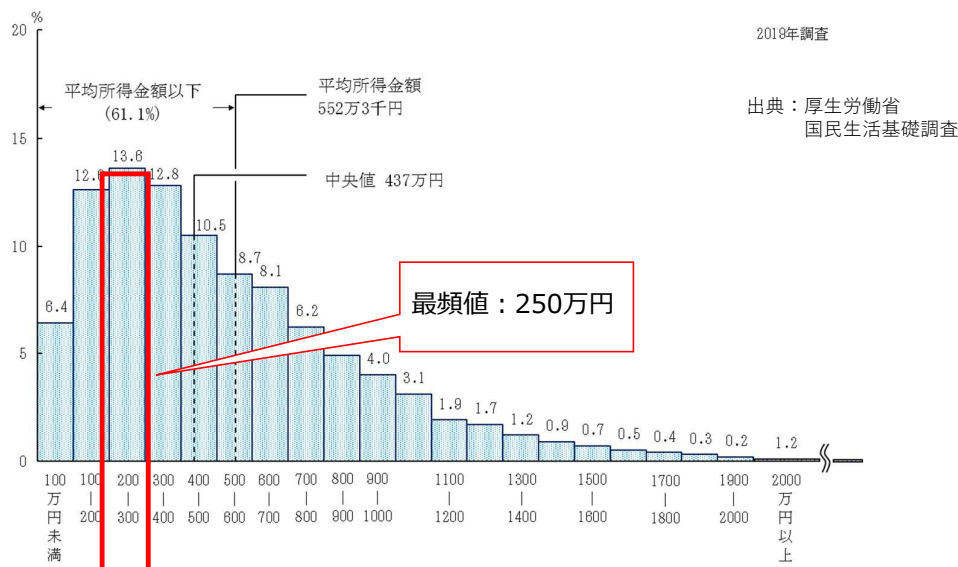
平均値・中央値・最頻値が持つ特徴を捉える

- データの分布に偏りがない場合（自然現象など）  
平均値 ≒ 中央値 ≒ 最頻値
- データの分布に偏りがある場合（社会現象など）  
平均値 ≠ 中央値 ≠ 最頻値

	平均値	中央値	最頻値
メリット	認知度が高く わかりやすい	外れ値の影響を 受けない	ボリュームゾーンが 分かる
デメリット	外れ値の影響を 強く受ける	データ数が多いと 特定に手間がかかる	全体傾向がつかめない

# データの分布が偏っている例

格差が発生しやすいデータは分布に偏りが生じる



Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

# 偏差

偏差は「平均値からのズレ」を表す

- 標準偏差を計算するためにはまず先に偏差を算出します

配布先	配り方①	偏差1	配り方②	偏差2	配り方③	偏差3
A	5	0	7	2	10	5
B	5	0	7	2	8	3
C	5	0	7	2	8	3
D	5	0	7	2	6	1
E	5	0	7	2	6	-1
F	5	0	3	-2	3	-2
G	5	0	3	-2	3	-2
H	5	0	3	-2	3	-2
I	5	0	3	-2	3	-2
J	5	0	3	-2	0	-5
合計	50		50		50	
平均値	5		5		5	
標準偏差		0		2		2.93

ズレ 2

0 1 2 3 4 5 6 7 8 9 10

平均値

Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

# 標準偏差

標準偏差は「“平均値からのズレ”の平均」を表す

配布先	配り方①	偏差1	配り方②	偏差2	配り方③	偏差3
A	5	0	7	2	10	5
B	5	0	7	2	8	3
C	5	0	7	2	8	3
D	5	0	7	2	6	1
E	5	0	7	2	6	1
F	5	0	3	-2	3	-2
G	5	0	3	-2	3	-2
H	5	0	3	-2	3	-2
I	5	0	3	-2	3	-2
J	5	0	3	-2	0	-5
合計	50		50		50	
平均値	5		5		5	
標準偏差		0		2		2.93

Excel関数

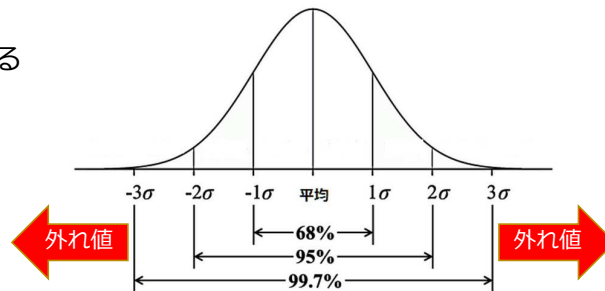
=stdev.p (範囲)  
※stdev.pとstdev.sがある。

# 外れ値

# 外れ値とは

## □ 正規分布に従うデータは、平均値の周りに以下の確率で分布

- $\pm 1\sigma$ で約70%
- $\pm 2\sigma$ で95%
- $\pm 3\sigma$ で99%以上が含まれる



## □ 平均値から $\pm 3\sigma$ 以上離れたデータは相当珍しいデータ ( $0.3\%=3/1000$ ) なので、「外れ値」として取り扱う

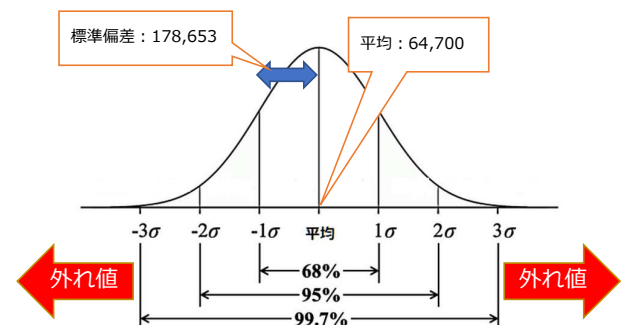
# 外れ値の見つけ方

## □ 偏差と標準偏差を計算

利用者	月間利用額
1	0
2	0
3	0
4	0
5	0
6	2,000
7	5,000
8	10,000
9	30,000
10	600,000
合計	647,000
平均値	64,700
中央値	1,000
最頻値	0



利用者	月間利用額	偏差
1	0	-64,700
2	0	-64,700
3	0	-64,700
4	0	-64,700
5	0	-64,700
6	2,000	-62,700
7	5,000	-59,700
8	10,000	-54,700
9	30,000	-34,700
10	600,000	535,300
合計	647,000	
平均値	64,700	
中央値	1,000	
最頻値	0	
標準偏差		178,653



# 外れ値の見つけ方

- 偏差が標準偏差の3倍以上になっているものを見つける

利用者	月間利用額	偏差
1	0	-64,700
2	0	-64,700
3	0	-64,700
4	0	-64,700
5	0	-64,700
6	2,000	-62,700
7	5,000	-59,700
8	10,000	-54,700
9	30,000	-34,700
10	600,000	535,300
合計	647,000	
平均値	64,700	
中央値	1,000	
最頻値	0	
標準偏差		178,653



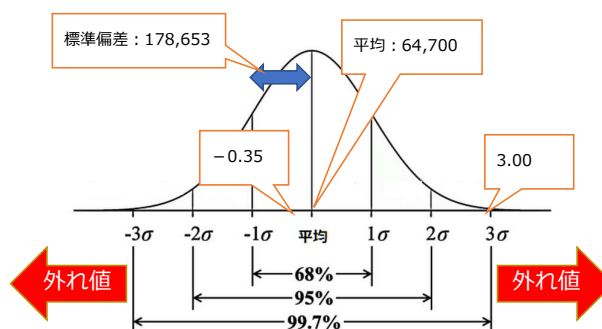
利用者	月間利用額	偏差	偏差 ÷ 標準偏差
1	0	-64,700	-0.36
2	0	-64,700	-0.36
3	0	-64,700	-0.36
4	0	-64,700	-0.36
5	0	-64,700	-0.36
6	2,000	-62,700	-0.35
7	5,000	-59,700	-0.33
8	10,000	-54,700	-0.31
9	30,000	-34,700	-0.19
10	600,000	535,300	3.00
合計	647,000		
平均値	64,700		
中央値	1,000		
最頻値	0		
標準偏差		178,653	

Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

15

## 位置関係の確認

利用者	月間利用額	偏差	偏差 ÷ 標準偏差
6	2,000	-62,700	-0.35
10	600,000	535,300	3.00
平均値	64,700		
中央値	1,000		
最頻値	0		
標準偏差		178,653	



Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

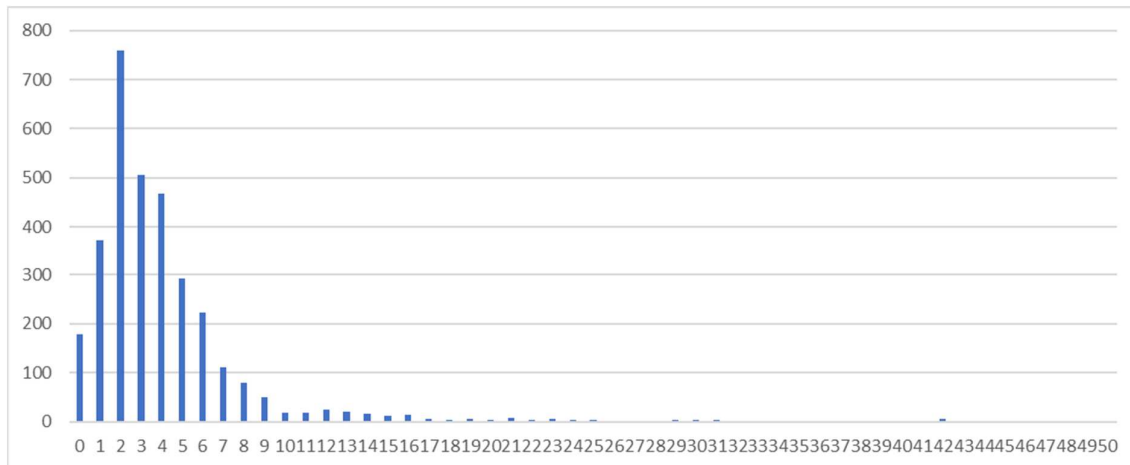
16



# 対数変換

□ 偏りのあるデータを扱う方法として対数変換という手法があります

- 東京都の住宅価格の例

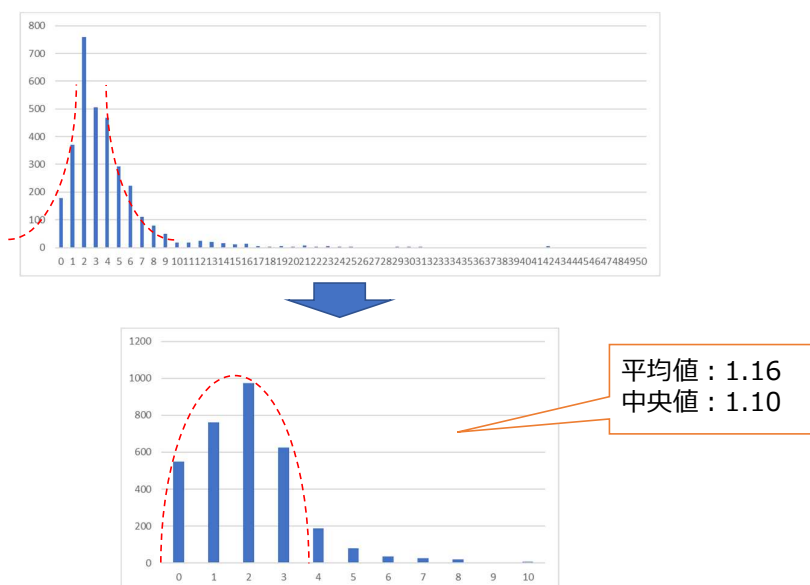


Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

17

# 対数変換

□ 対数変換を行うことで正規分布に近い形を作ることができる

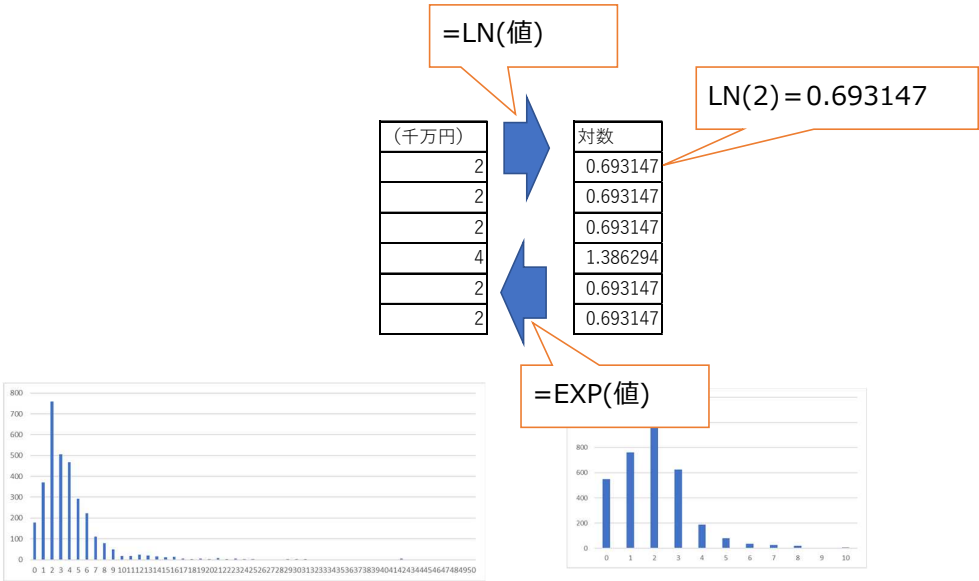


Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

18

# 対数変換

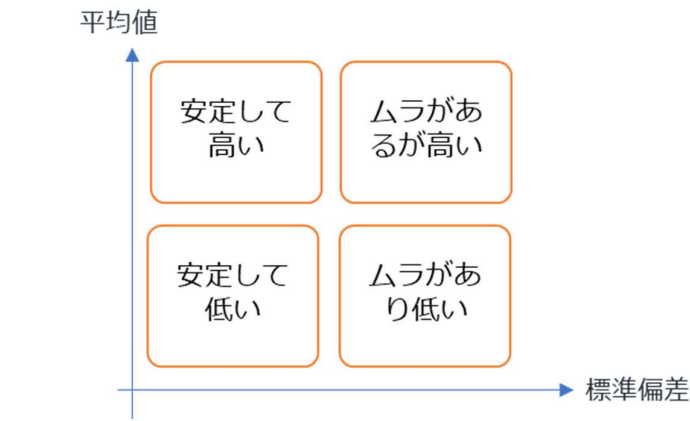
対数変換の処理はExcel関数を使って簡単に行うことができる



# 分析の手法

# 統計量のみでできる分析手法

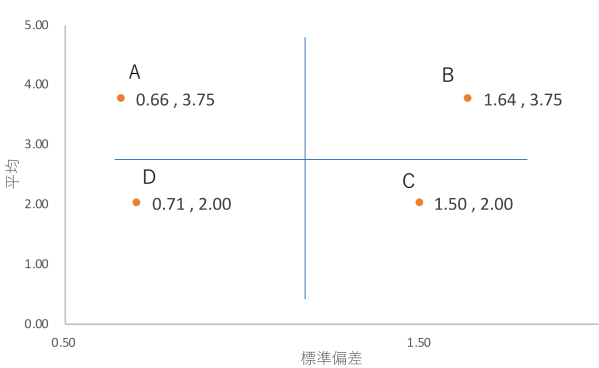
- 平均値を縦軸、標準偏差を横軸にして散布図を描く。
- 散布図の領域を4つのエリアに分けて捉えることで簡単にカテゴライズすることが可能。



Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

# 統計量のみでできる分析手法

	HP	MP	攻撃力	防御力	すばやさ	ちから	運	魔力	統計量		
									平均値	標準偏差	
A	5	3	4	3	3	4	4	4	3.75	0.66	熟練タイプ
B	5	5	1	4	2	2	5	5	3.75	1.64	特化タイプ
C	2	1	5	1	4	1	1	1	2.00	1.50	成長段階
D	1	2	3	2	2	1	3	2	2.00	0.71	初期段階



解釈の例	
A	どんな環境でも活躍できる。
B	特定のシチュエーションにおいてトップレベルの能力があるが、どこでも活躍できるかは相性による。
C	強みが現れ始めている可能性が高く、環境によっては一軍として使える。
D	まずはベースの向上が必要。経験を積ませて、強みとなる要素を作る。

Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

# 散布図分析

□ 配布データを使用して、平均値と標準偏差の散布図を作成しましょう

□ データの準備

- e-Stat (<https://www.e-stat.go.jp/>)



23

## データの準備

□ e-Stat (<https://www.e-stat.go.jp/>)

### 政府統計一覧



政府統計コード	政府統計名	概要
00000002	一般職国家公務員在職状況統計表（人事統計報告）	<a href="#">詳細</a>
00000003	国家公務員退職手当実態調査	<a href="#">詳細</a>
00020112	国家公務員死因調査	<a href="#">詳細</a>
00020131	国家公務員災害補償統計	<a href="#">詳細</a>
00020151	退職公務員生活状況調査	<a href="#">詳細</a>
00020211	一般職の国家公務員の任用状況調査	<a href="#">詳細</a>
00020312	国家公務員給与等実態調査	<a href="#">詳細</a>
00100502	SDGsに関する全国アンケート調査	<a href="#">詳細</a>
00160003	地方消費者行政の現況調査	<a href="#">詳細</a>

# データの準備

## 令和5年度のデータを使用します

SDG sに関する全国アンケート調査

一覧形式で表示

本調査は、自治体のSDG s 達成に向けた取組の実施状況を調査することを目的に、貴自治体におけるSDG s の認知度や取組度合いに関してお伺いさせて頂くものです。また、令和4年12月に閣議決定された「デジタル田園都市国家構想総合戦略」に「地方公共団体によるSDG s 達成に向けた取組割合の把握を行う」ことが明記されています。

SDG sに関する全国アンケート調査	データベース	件数 更新日	ファイル	件数 更新日	概要
令和2年度 SDGsに関する全国アンケート調査				3件 2021-04-01	
令和3年度SDG sに関する全国アンケート調査				3件 2021-12-22	
令和4年度SDG sに関する全国アンケート調査				3件 2023-12-06	
令和5年度SDG sに関する全国アンケート調査				3件 2024-01-19	

# データの準備

SDG sに関する全国アンケート調査

一覧形式で表示

本調査は、自治体のSDG s 達成に向けた取組の実施状況を調査することを目的に、貴自治体におけるSDG s の認知度や取組度合いに関してお伺いさせて頂くものです。また、令和4年12月に閣議決定された「デジタル田園都市国家構想総合戦略」に「地方公共団体によるSDG s 達成に向けた取組割合の把握を行う」ことが明記されています。

令和5年度SDG sに関する全国アンケート調査	公開（更新）日
- [3件]	2024-01-19

データセット一覧

URLをコピー

一覧形式で表示

政府統計名

SDG sに関する全国アンケート調査

提供統計名

令和5年度SDG sに関する全国アンケート調査

表番号	統計表	調査年月	公開（更新）日	データダウンロード
自治体別	地方創生SDGs達成に向けた取組状況	2023年度	2024-01-19	EXCEL
都道府県別	地方創生SDG s 達成の取組を推進している自治体割合	2023年度	2024-01-19	EXCEL
	地方創生SDG s を推進している自治体一覧	2023年度	2024-01-19	EXCEL

# グラフ作成の手順

※今回使用するデータシートに保護がかけられていて編集不可になっているため

- 元表から一部分だけコピーして別シートに貼り付けます

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	自治体別個別状況													
2														
3														
4														
5														
6														
7														
8														
9	自治体別 地方創生SDGs達成に向けた取組状況													
10														
11	自治体名	SDGsについて推進しているか	自治体内部における普及啓発活動	2030年のあるべき姿を設定した施策のビジョンづくり	推進組織の設置等による体制づくり	SDGs推進のために、自治体で作成している地方版総合戦略との紐づけ反映	SDGs推進のために、自治体で作成している地方版総合戦略との紐づけ反映	SDGsに関する取組の推進状況、成果の共有	国内外で定めたステークホルダー（関係者）との連携	ローカル指標（自治体独自の評価指標）の設置	その他SDGsと推進に向けた取組	「地方創生SDGs推進」に関する取組の推進	担当部署	
12														
13	北海道	4	3	3	3	3	3	3	3	3	3	2	北海道総合政策部計画課計画推進	
14	札幌市	4	3	3	2	3	2	3	2	3	2	2	まちづくり政策局政策企画課企画	
15	函館市	4	1	1	1	3	3	1	3	1	2	1	企画部計画推進課計画課課長	
16	小樽市	4	2	2	1	2	3	2	2	2	2	1	経済部企画政策課	
17	旭川市	4	2	3	2	3	3	2	2	2	2	1	総合政策部政策課課長	
18	室蘭市	2	2	1	1	2	2	2	2	1	1	1	企画財政部企画課	
19	網走市	4	3	3	3	3	3	2	2	1	1	1	総合政策部 都市経営課 企画	

Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

27

# グラフ作成の手順

- 値を「数値」に変換します

	A	B	C	D	E	F	G	H	I	J	K	L
1	札幌市	4	3	3	2	3	2	3	2	3	2	2
2	函館市				1	3	3	1	3	1	2	1
3	小樽市				1	2	3	2	2	2	2	1
4												
5												
6												
7												

Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

28

# グラフ作成の手順

## 標準偏差を計算します

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	札幌市	4	3	3	2	3	2	3	2	3	2	2	
2	函館市	4	1	1	1	3	3	1	3	1	2	1	
3	小樽市	4	2	2	1	2	3	2	2	2	2	1	
4													

関数メニュー

関数の挿入

関数の検索(S):

関数の分類(C): 統計

関数名(N):

STDEV.P

STDEV.S

STDEVA

STDEVPA

STEYX

T.DIST

STDEV.P(数値1,数値2,...)

引数を母集団全体であるとして、母集団の標準偏差を返します。論理値、および文字列は無視されます。

この関数のヘルプ

OK

キャンセル

# グラフ作成の手順

## 標準偏差を計算します

	A	B	C	D	E	F	G	H	I	J	K	L
1	札幌市	4	3	3	2	3	2	3	2	3	2	2
2	函館市	4	1	1	1	3	3	1	3	1	2	1
3	小樽市	4	2	2	1	2	3	2	2	2	2	1
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												

関数の引数

STDEV.P

数値1 B1:L1

数値2

2は空白のままでOK

= {4,3,3,2,3,2,3,2,3,2,2}

= 数値

= 0.642824347

引数を母集団全体であるとして、母集団の標準偏差を返します。論理値、および文字列は無視されます。

数値1: 数値1,数値2,... には母集団に対応する数値、または数値を含む参照を、1 ~ 255 個まで指定できます。

数式的結果 = 0.642824347

この関数のヘルプ(H)

OK

キャンセル

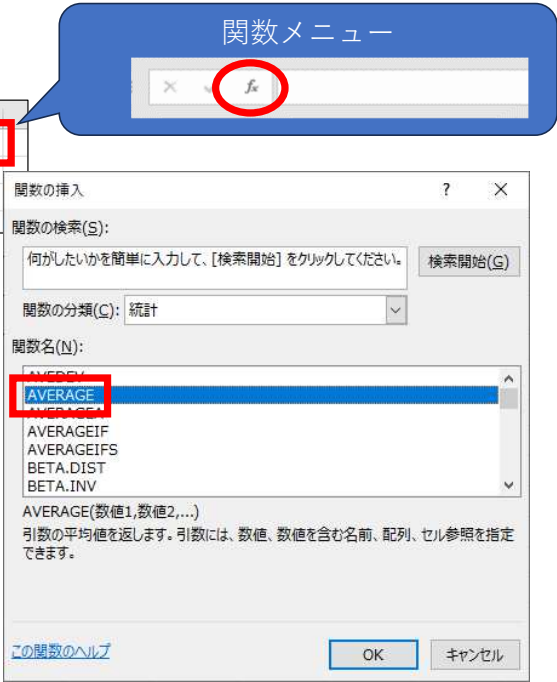


# グラフ作成の手順

## 平均値を計算します

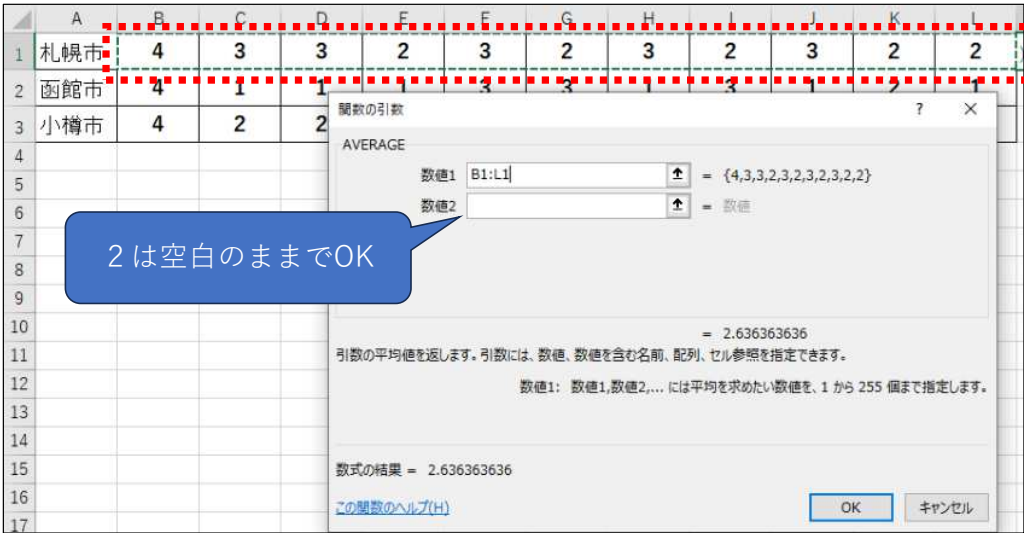
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	札幌市	4	3	3	2	3	2	3	2	3	2	2		
2	函館市	4	1	1	1	3	3	1	3	1	2	1		
3	小樽市	4	2	2	1	2	3	2	2	2	2	1		
4														

関数メニュー



# グラフ作成の手順

## 平均値を計算します

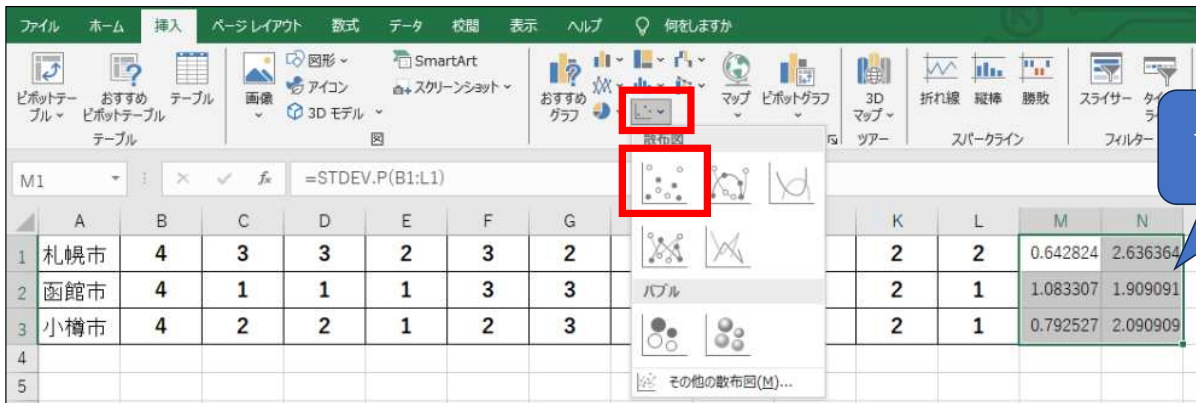




# グラフ作成の手順

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	札幌市	4	3	3	2	3	2	3	2	3	2	2	0.642824	2.636364
2	函館市	4	1	1	1	3	3	1	3	1	2	1		
3	小樽市	4	2	2	1	2	3	2	2	2	2	1		
4														

下へコピー



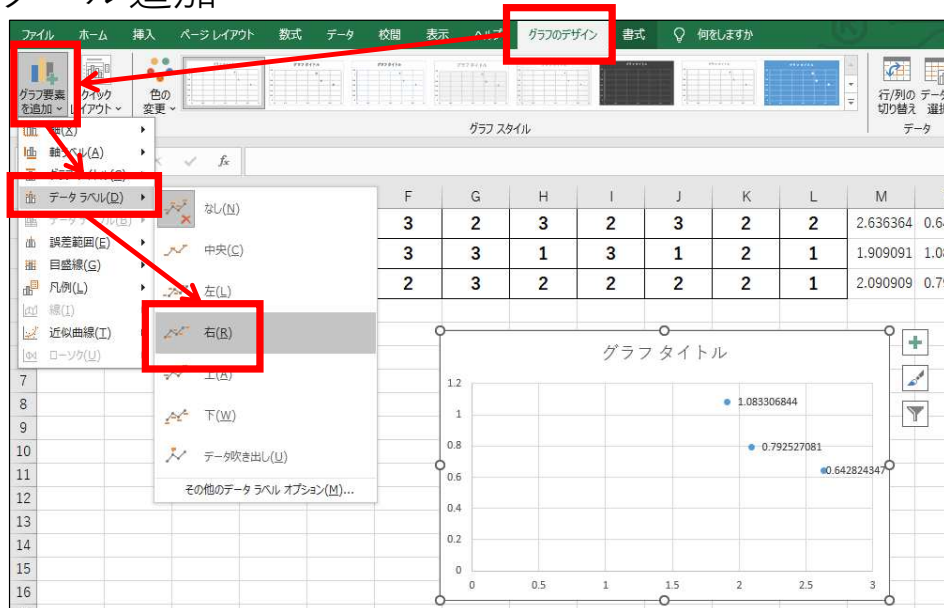
エリア選択して  
散布図を作成

Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

33

## グラフの体裁を整える

### □ データラベル追加



Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

34

# グラフの体裁を整える

## □ 書式設定する

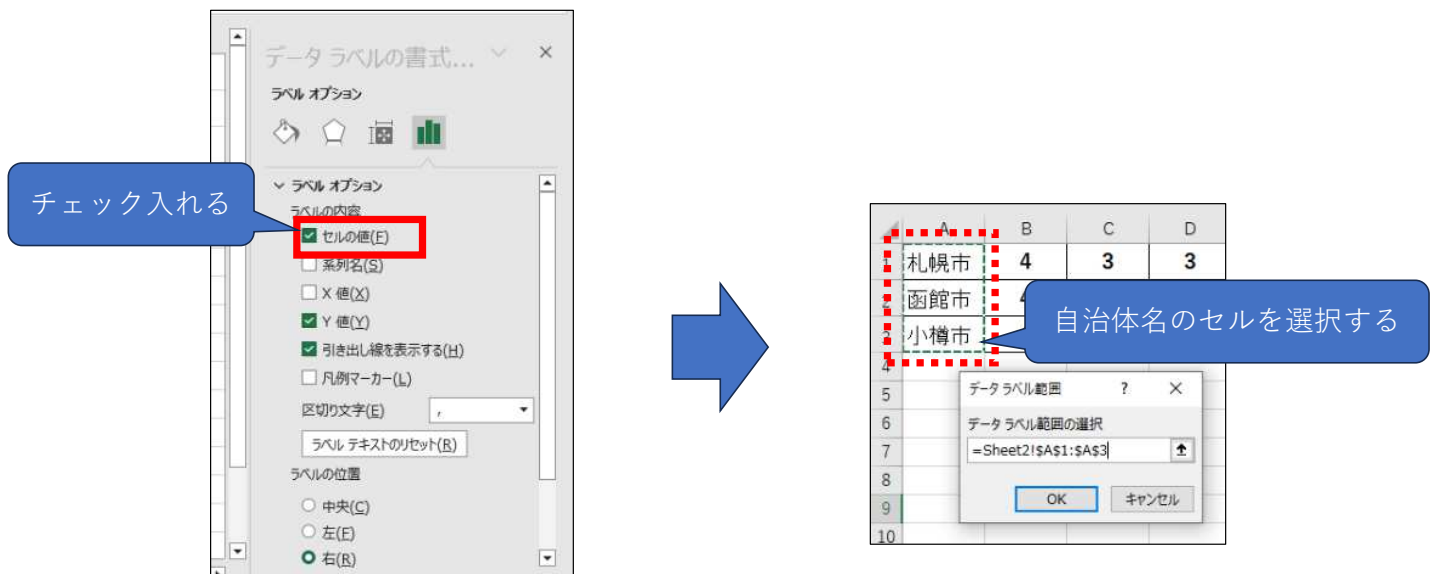


Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

35

# グラフの体裁を整える

## □ A列をラベルの内容に設定



Copyright UCHIDA HUMAN DEVELOPMENT all rights reserved.

36

# グラフの体裁を整える

## □ 完成

