

2. データサイエンス・AI の事例



アウトライン

1. データサイエンス
2. 表計算ソフトウェア Excel
3. 散布図 (Excel を使用)
4. 合計、平均 (Excel を使用)
5. 分布、密度 (Excel を使用)
6. 人工知能による分類、特徴抽出

2-1 データサイエンスでできること

データサイエンスでできること

- データから，正しく知見や結論を導くこと

平均，分散などの算出.

要因や連関の度合いの算出.

- データサイエンスと人工知能は大きく関係する
 - 人工知能で得られた結果の検証にデータサイエンスを使う
 - データの分析のために，人工知能の技術を使う

2-2 表計算ソフトウェア Excel

パソコンの威力

- **ワープロ**

文書の編集、清書、目次、表の作成など

- **表計算**

データの管理、計算、グラフ作成など

- **プレゼン**

ビジュアル資料作成

- **インターネット**

情報収集、コミュニケーション

データはすべて**デジタル**（ファイル）。

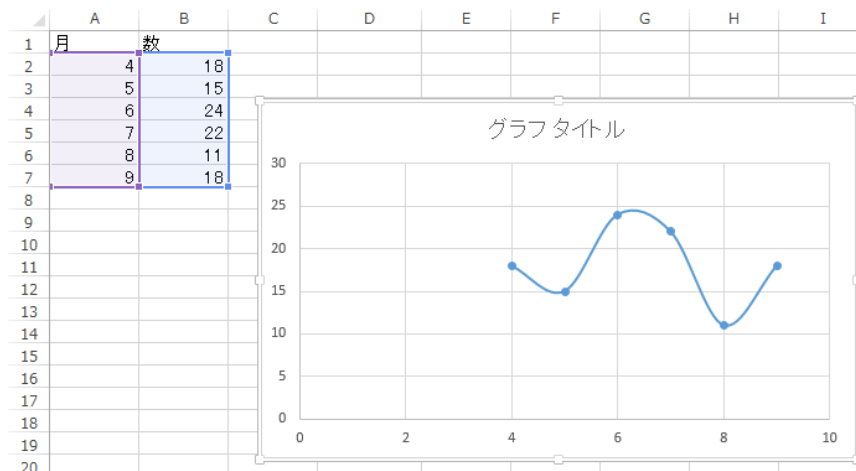
管理、共有、交換が簡単

表計算ソフトウェアは何の役に立つのか

- データの記録、保管、共有
- 表計算の機能 = 集計・集約、グラフ作成など

	A	B	C	D
1	品名	単価	数量	合計
2	りんご	100	10	1000
3	みかん	50	5	250
4				1250

	A	B	C
1	AA	算数	90
2	AA	国語	85
3	BB	算数	92
4	CC	国語	75
5	CC	理科	95
6			



例えば、こんなことが簡単にできます

表の作成

	A	B	C	D
1	品名	単価	数量	合計
2	りんご	100	10	1000
3	みかん	50	5	250
4				1250

単価を書き変えると

Excel の画面

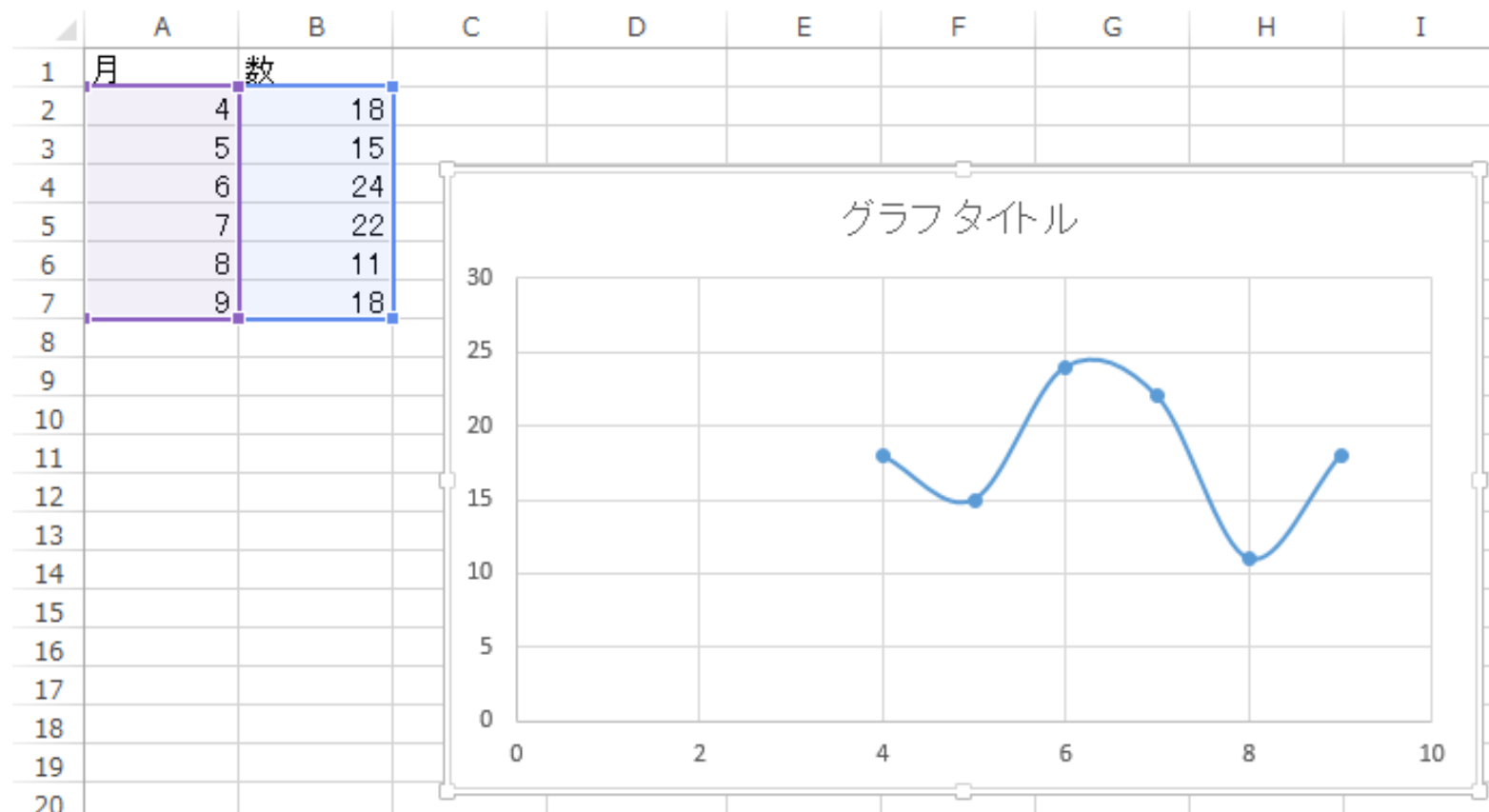
	A	B	C	D
1	品名	単価	数量	合計
2	りんご	80	10	800
3	みかん	50	5	250
4				1050

合計が**自動**で再計算される

Excel の画面

例えば、こんなことが簡単にできます

グラフ



例えば、こんなことが簡単にできます

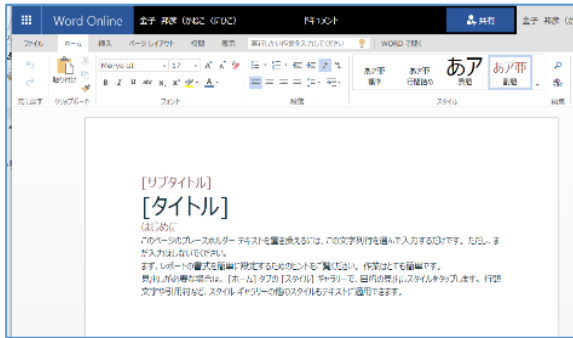
条件に合致するデータの
強調表示

	A	B	C
1	AA	算数	90
2	AA	国語	85
3	BB	算数	92
4	CC	国語	75
5	CC	理科	95
6			

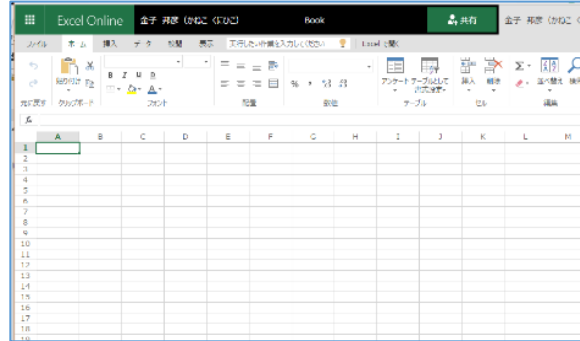
並べ替え

	A	B	C
1	CC	国語	75
2	AA	国語	85
3	AA	算数	90
4	BB	算数	92
5	CC	理科	95
6			

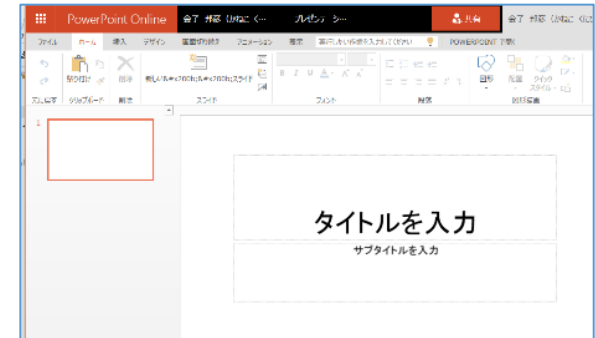
Office 365 の主な機能



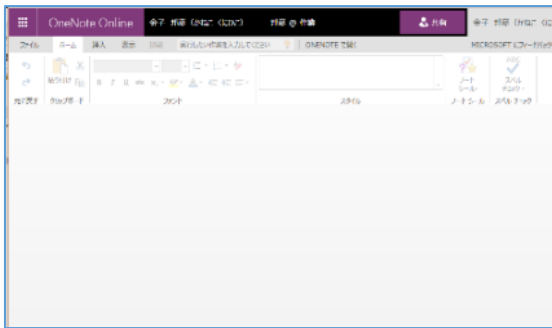
ワード (文書作成)



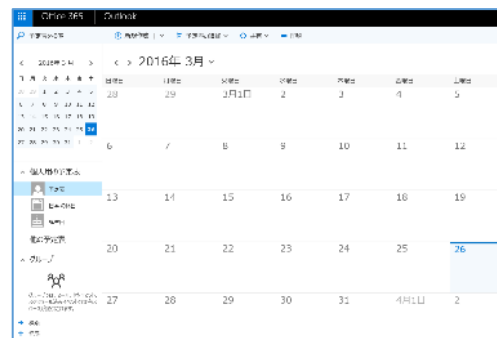
エクセル (表計算)



パワーポイント
(プレゼン)



ワンノート (電子ノート)



アウトルック (電子メール)

- ・パソコンでレポートを作成したり，発表したり，データをまとめたりで便利

Office 365 の種類

- Office 365 の**オンライン版**

WEBブラウザで使う。

<https://portal.office.com>

各自の **ID とパスワード**でサインインが必要。

- Office 365 の**アプリ版**

前もってインストールが必要。

インストールでは、大量の通信が行われる。

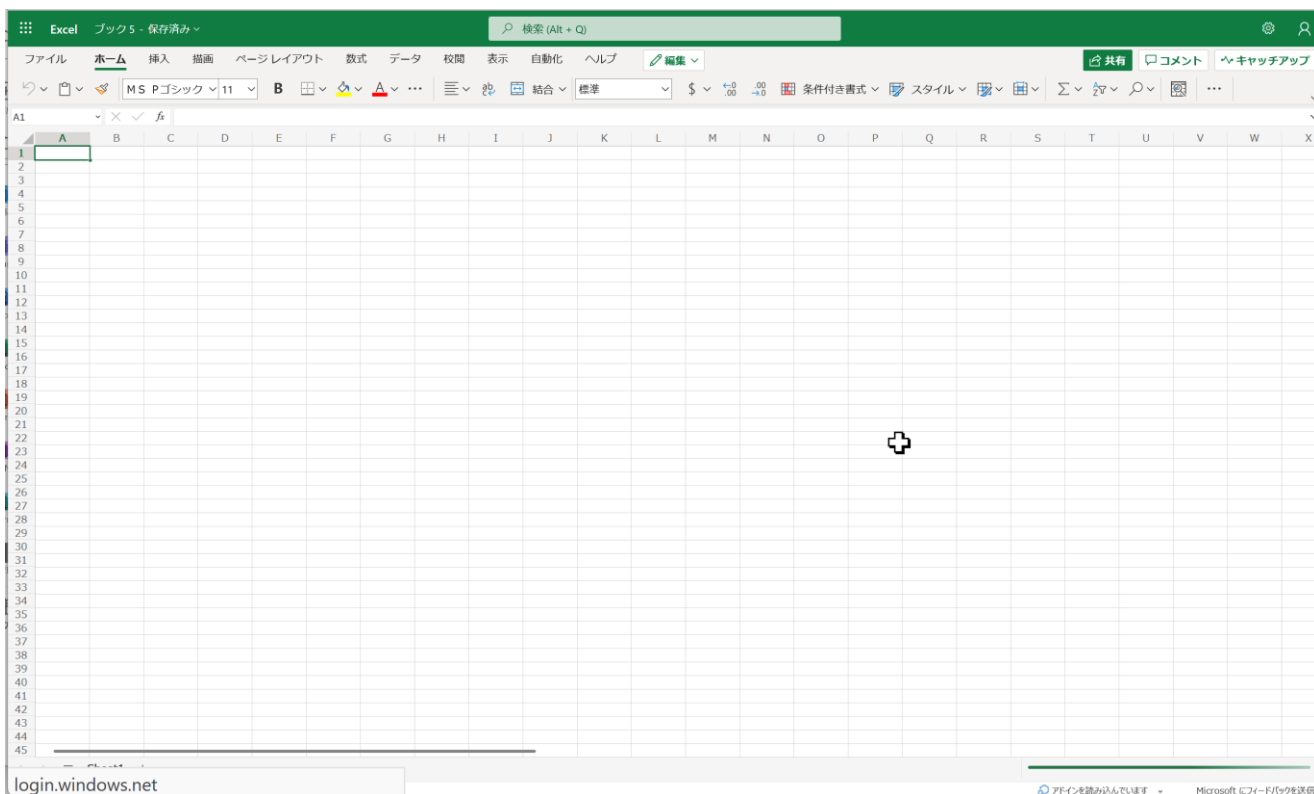
(時間がかかる。通信費用にも注意)

2種類ある。この授業では、**どちらを使用しても問題ない**

Office 365 オンライン版で Excel を起動

【要点】 Web ブラウザで、次のページを開き、各自の ID とパスワードでサインイン

<https://portal.office.com>



Office 365 オンライン版で Excel を起動

① Web ブラウザで、次のページを開く

<https://portal.office.com>

② 電子メールアドレスを入れる。「次へ」をクリック。

(例) p1234567@fukuyama-u.ac.jp



サインイン

メール、電話、Skype

アカウントをお持ちではない場合、[作成](#)できます。

[アカウントにアクセスできない場合](#)

戻る

次へ

Office 365 オンライン版で Excel を起動

③ パスワードを入れ, 「サインイン」をクリック

パスワードは, 各自が設定したもの

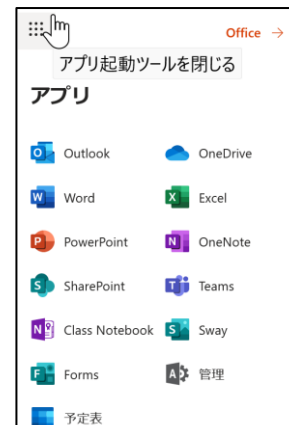
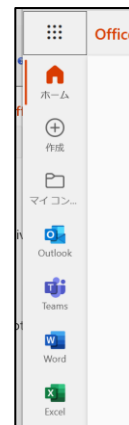
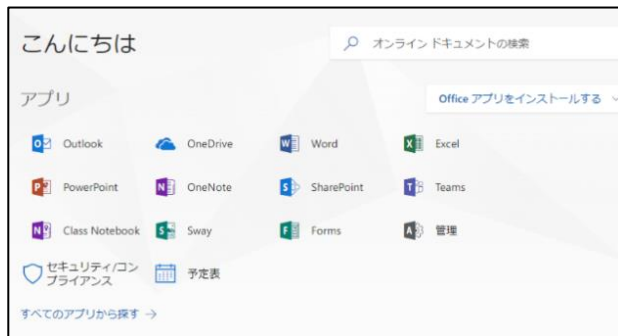
パスワードの入力

パスワード

[パスワードを忘れた場合](#)

サインイン

④ Excel を使いたいときは, **メニュー**で Excel を選ぶ



さまざまなメニュー

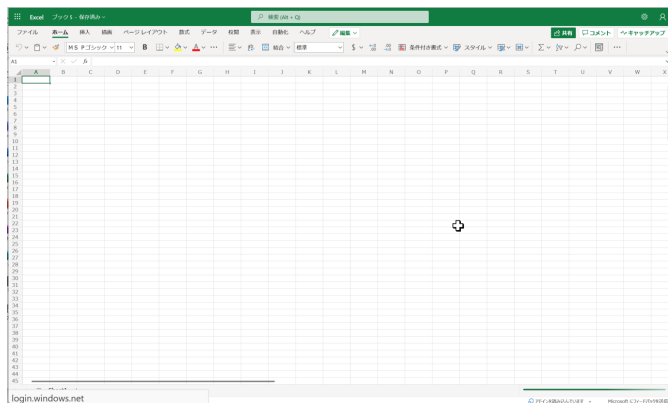
Office 365 オンライン版で Excel を起動

⑤ Excel のブックの種類を選ぶ



この授業では「新しい空白のブック」を使う

⑥ Excel の画面が開く



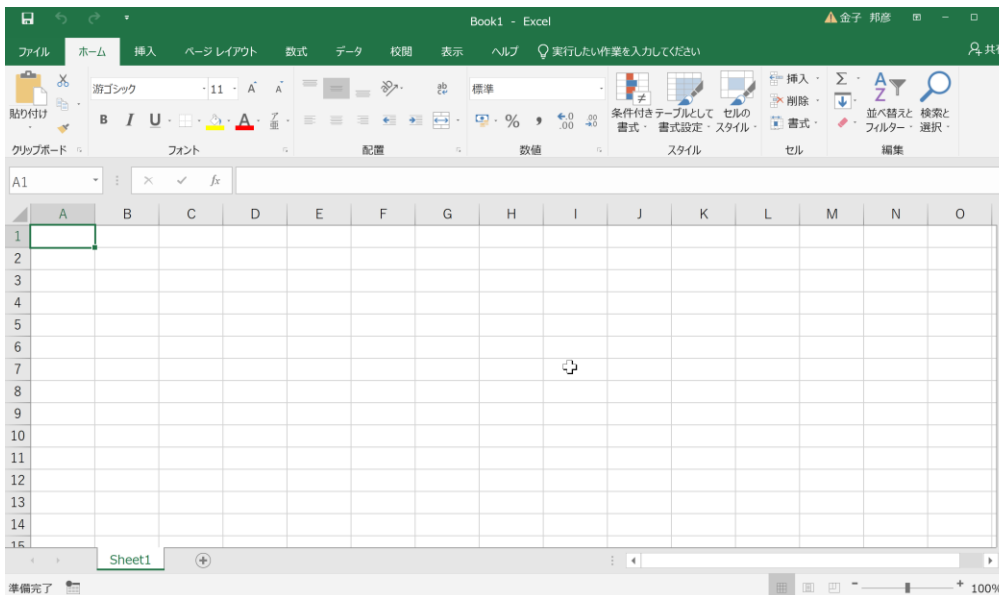
Office 365 アプリ版のインストールと Excel の起動

【要点】 **インストール**は、Office 365 アプリ版を使えるようにするための作業（最初に行う）。

そのとき、次のページを開き、各自の **ID とパスワード** でサインイン

<https://portal.office.com>

インストールが終わったら、**スタートメニュー**等で Excel を起動



Office 365 アプリ版のインストールと Excel の起動

① Web ブラウザで、次のページを開く
https://portal.office.com

② 電子メールアドレスを入れる。「次へ」をクリック。
(例) **p1234567@fukuyama-u.ac.jp**



サインイン

メール、電話、Skype

アカウントをお持ちではない場合、[作成](#)できます。

[アカウントにアクセスできない場合](#)

戻る

次へ

Office 365 アプリ版のインストールと Excel の起動

- ③ パスワードを入れ, 「サインイン」をクリック
パスワードは, 各自が設定したもの

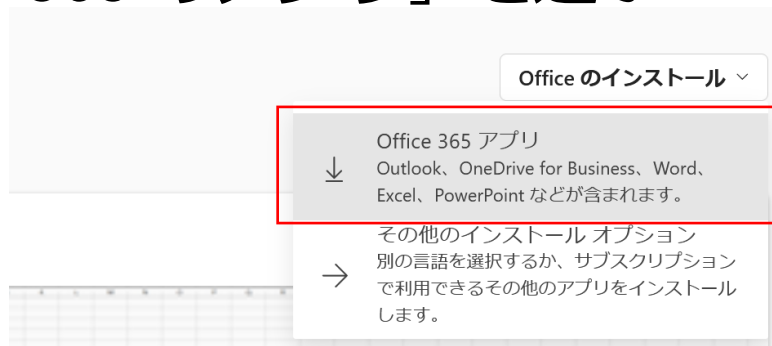
パスワードの入力

パスワード

[パスワードを忘れた場合](#)

サインイン

- ④ 画面で「Office のインストール」をクリック. メニューで
「Office 365 のアプリ」を選ぶ

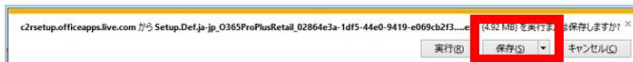


Office 365 アプリ版のインストールと Excel の起動

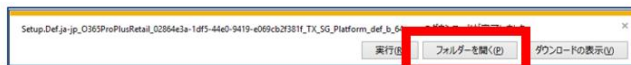
⑤ 画面の指示に従い、インストールを行う

インストールでは、大量の通信が行われる。
(時間がかかる。通信費用にも注意)

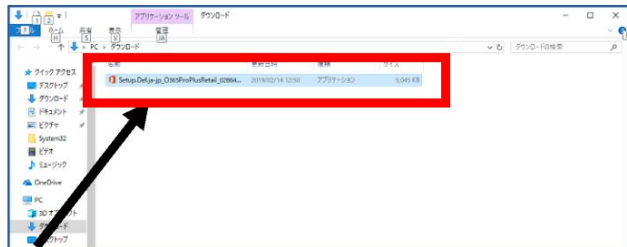
次のような指示がでる



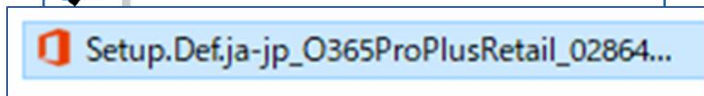
1. 保存する



2. フォルダを開く



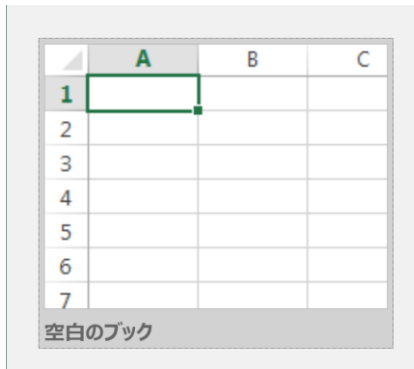
3. 実行し、その後も、画面の指示に従う



Office 365 アプリ版のインストールと Excel の起動

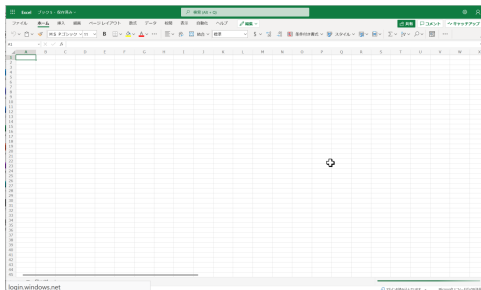
⑥ Excel を使うときは、スタートメニューなどで Excel を選ぶ

⑦ **Excel のブックの種類**を選ぶ



この授業では「新しい空白のブック」を使う

⑧ Excel の画面が開く

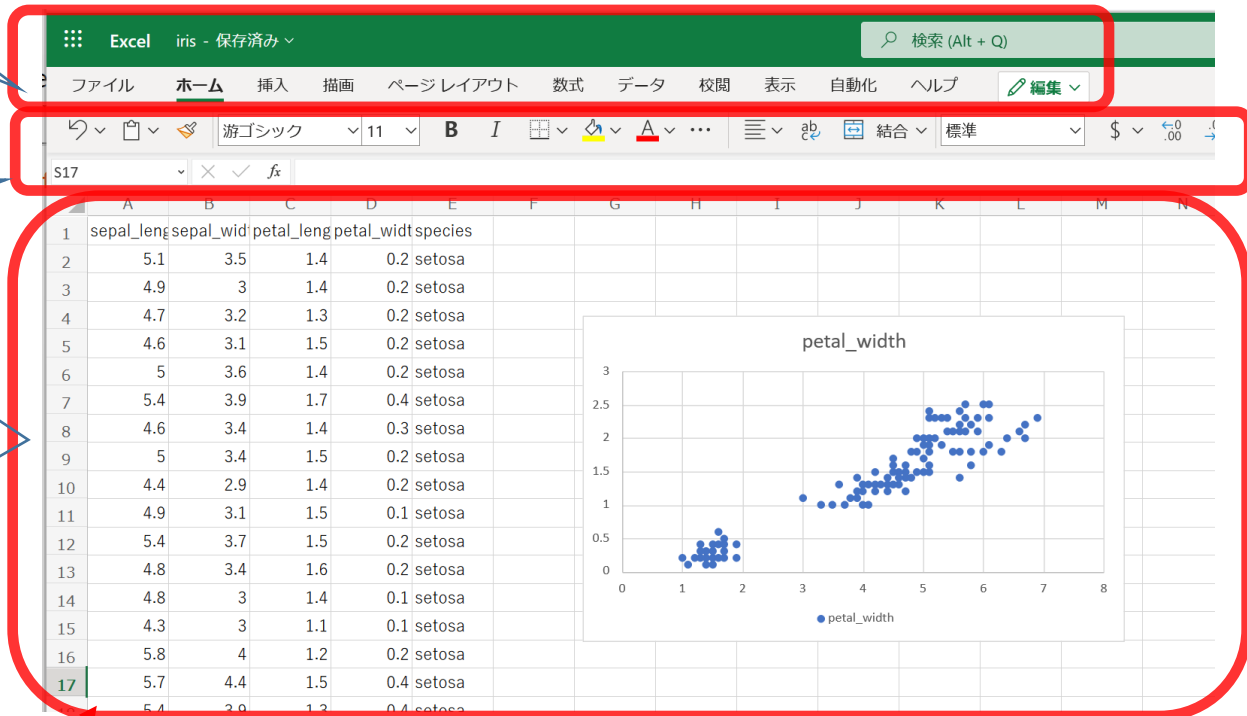


オンライン版の Excel の画面（メニュー、リボン、ワークシートなど）

メニュー

リボン

ワークシート
表形式で値などが
入る。
グラフの挿入な
ども可能



表形式で、値や数式を並べる

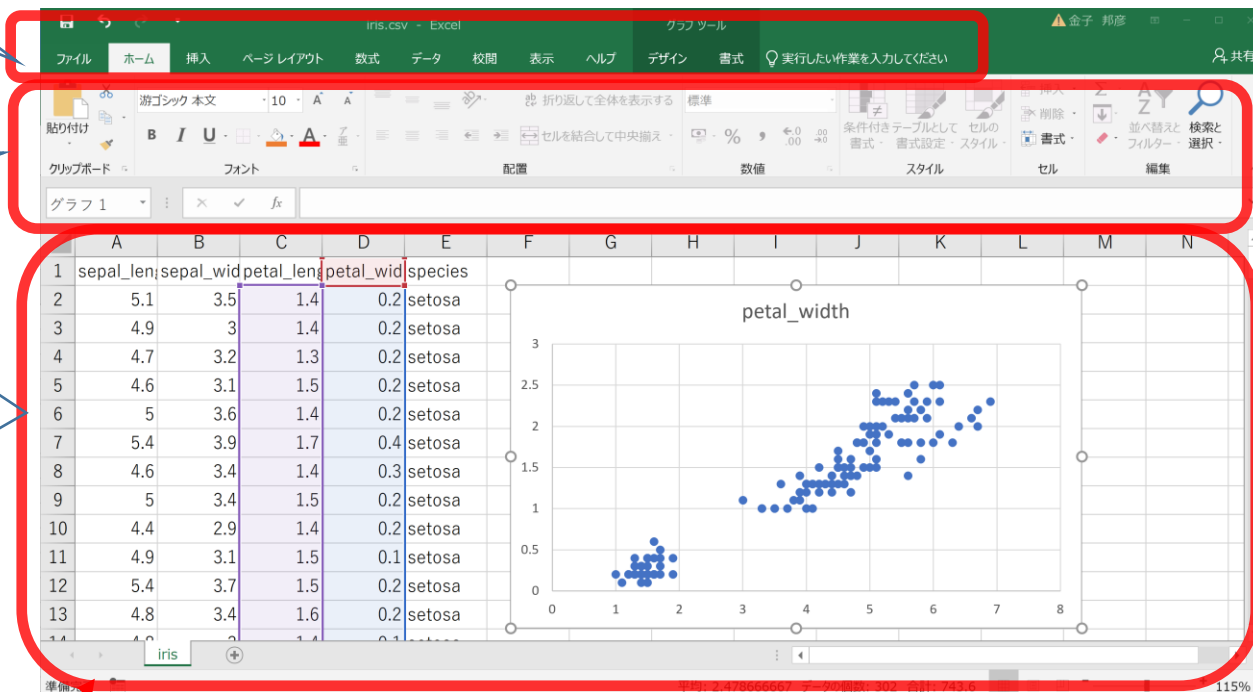
オンライン版の Excel の画面

アプリ版の Excel の画面（メニュー、リボン、ワークシートなど）

メニュー

リボン

ワークシート
表形式で値などが
入る。
グラフの挿入な
ども可能



表形式で、値や数式を並べる

アプリ版の Excel の画面
(Excel 2019 の画面を示している) 23

Excel のワークシート

Excel の**ワークシート**（シートともいう）には、表形式で、値や数式を並べる。

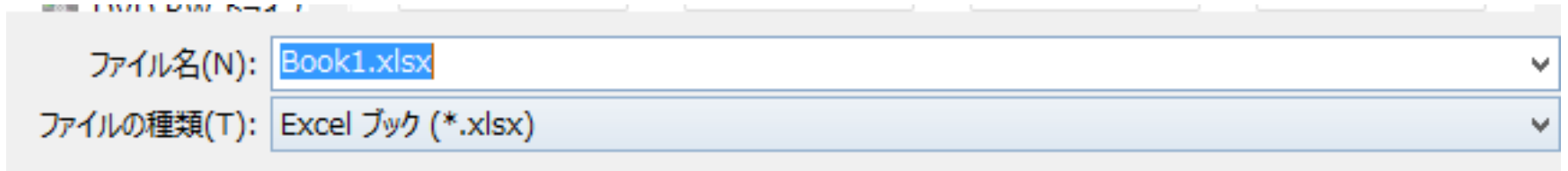
グラフの挿入なども可能

	A	B	C	D
1	品名	単価	数量	合計
2	りんご	100	10	1000
3	みかん	50	5	250
4				1250
5				

Excel のワークシートの例

Excel のブック

- Excel の**ブック**は、Excel の**ファイル**のこと
- 1つあるいは複数の**ワークシート**を、1つの**ブック**に保存することができる



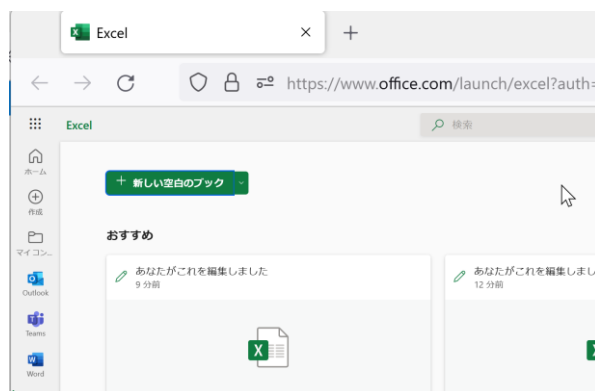
Excel で保存するときに、ファイル名などを設定できる。

Excel のスタート画面

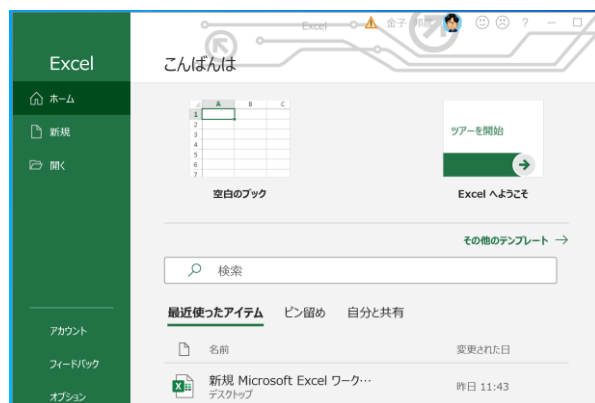
Excel を**起動**すると、**最初**にスタート画面が表示される。
作成したいブックの種類を選ぶことができる。
過去の履歴の確認もできる

この授業では「新しい空白のブック」を使う

オンライン版

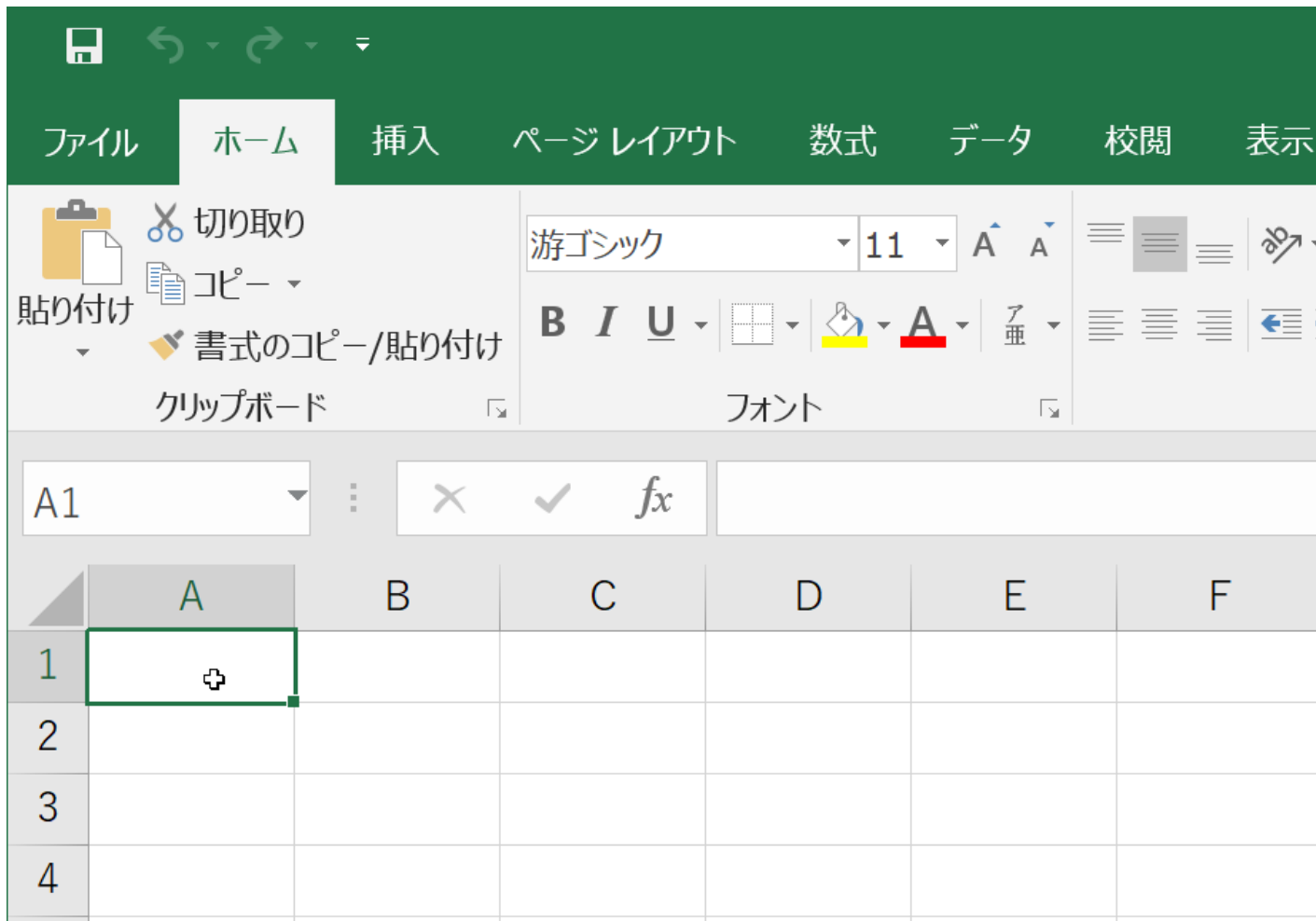


アプリ版



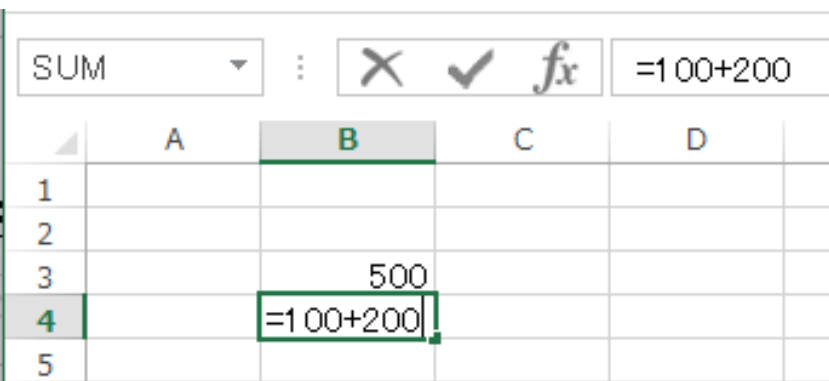
アクティブセル

- Excel での編集集中のセル

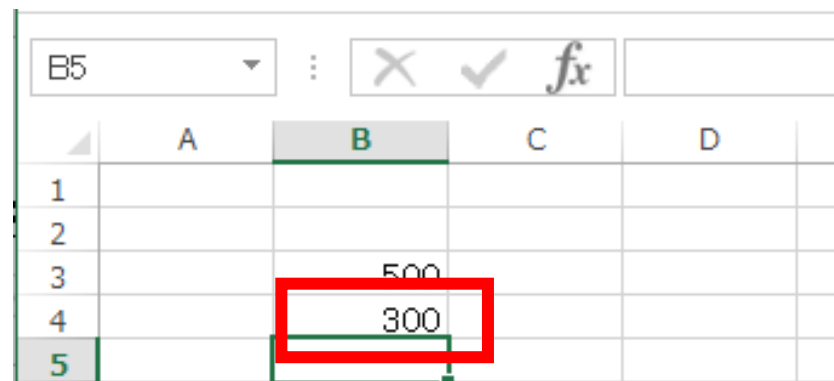


アクティブセルでの数式の入力

数式を入力したいときは、頭に半角の「=」を付ける



A screenshot of an Excel spreadsheet. The formula bar at the top shows the formula $=100+200$. The active cell is B4, which also contains the formula $=100+200$. The spreadsheet grid shows columns A, B, C, D and rows 1, 2, 3, 4, 5. Cell B3 contains the value 500.



A screenshot of an Excel spreadsheet. The formula bar at the top is empty. The active cell is B4, which now contains the calculated result 300. The spreadsheet grid shows columns A, B, C, D and rows 1, 2, 3, 4, 5. Cell B3 contains the value 500. A red rectangle highlights the cell B4.

キーボードで「**=100+200**」と打つと、**アクティブセル**に数式が入る

入力を終わりたいので **Enter キー**を押す。
すると、数式が**自動計算**される

数式バーで数式の確認①

数式バーに
数式が表示される
(ここで修正もできる)

	A	B	C	D
1				
2				
3		500		
4		300		
5				



	A	B	C	D
1				
2				
3		500		
4		300		
5				

「300」のところを
クリック

アクティブセルが動く

数式バーで数式の確認②

	A	B	C	D
1				
2				
3		500		
4		300		
5				



	A	B	C	D
1				
2				
3		500		
4		=100+200		
5				

「300」のところを
ダブルクリック

数式バーに
数式が表示される
(ここで修正もできる)

アクティブセルのところに
数式が表示される
(ここでも修正できる)

アクティブセルでの数式の入力

「=B3+B4」のような数式もある

	A	B	C	D
1				
2				
3		500		
4		300		
5		=B3+B4		



	A	B	C	D
1				
2				
3		500		
4		300		
5		800		

キーボードで「=B3+B4」と
打つと、**アクティブセル**に数式が入る

入力を終わりたいので
Enter キーを押す。
すると、数式が**自動計算**される

Excel の数式

- Excel では、数式の頭に、**半角の「=」** を付ける
- 数式は、**半角文字**である
- 数式の中には、**番地**（「B3」や「B4」など）を書くことができる

セルの数式と値のクリア

	A	B	C	D	E	F
1						
2		Sepal Length (cm)	Sepal Width (cm)			
3		7	3.2	4.7	1.4	Iris-versicolor
4		6.4	3.2			
5		6.9	3.1			
6		5.5	2.3			
7		6.5	2.8			
8		5.7	2.8			
9		6.3	3.3			
10		4.9	2.4			
11		6.6	2.9			
12		5.2	2.7			
13		5	2			
14		5.9	3			
15		6	2.2			
16		6.1	2.9			
17		5.6	2.9			

右クリックメニュー:

- 切り取り(I)
- コピー(C)
- 貼り付けのオプション:
- 形式を選択して貼り付け(S)...
- 挿入(I)...
- 削除(D)...
- 数式と値のクリア(N)**
- クイック分析(Q)

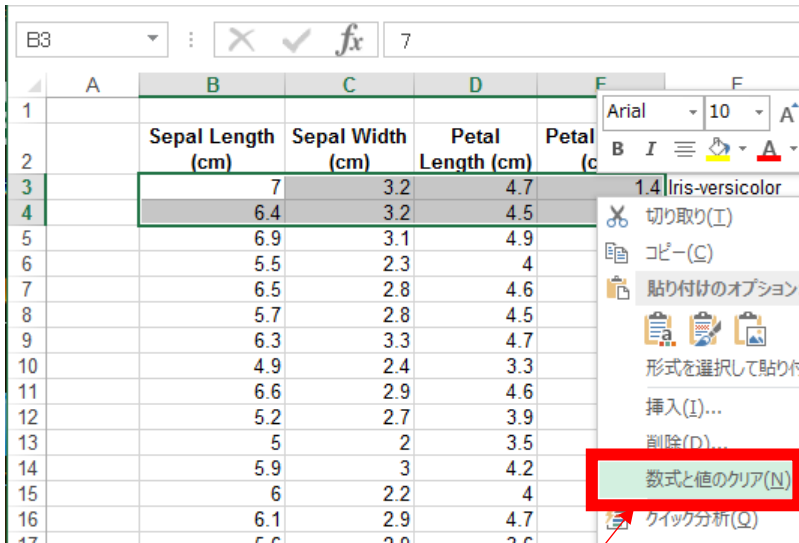


	A	B	C	D	E	F
1						
2		Sepal Length (cm)	Sepal Width (cm)	Petal Length (cm)	Petal Width (cm)	Class
3		7	3.2	4.7	1.4	Iris-versicolor
4		6.4		4.5	1.5	Iris-versicolor
5		6.9	3.1	4.9	1.5	Iris-versicolor
6		5.5	2.3	4	1.3	Iris-versicolor
7		6.5	2.8	4.6	1.5	Iris-versicolor
8		5.7	2.8	4.5	1.3	Iris-versicolor
9		6.3	3.3	4.7	1.6	Iris-versicolor
10		4.9	2.4	3.3	1	Iris-versicolor
11		6.6	2.9	4.6	1.3	Iris-versicolor
12		5.2	2.7	3.9	1.4	Iris-versicolor
13		5	2	3.5	1	Iris-versicolor
14		5.9	3	4.2	1.5	Iris-versicolor
15		6	2.2	4	1	Iris-versicolor
16		6.1	2.9	4.7	1.4	Iris-versicolor
17		5.6	2.9	3.6	1.3	Iris-versicolor

消えた!

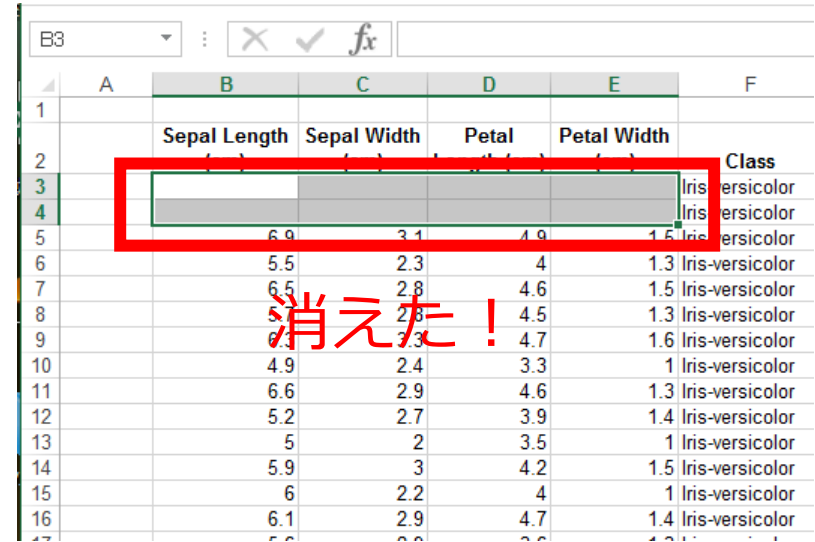
クリアしたいセルを右クリックして、
「数式と値のクリア」を選ぶ

セルの数値と値のクリア



	A	B	C	D	E	F
1						
2		Sepal Length (cm)	Sepal Width (cm)	Petal Length (cm)	Petal Width (cm)	
3		7	3.2	4.7	1.4	Iris-versicolor
4		6.4	3.2	4.5		
5		6.9	3.1	4.9		
6		5.5	2.3	4		
7		6.5	2.8	4.6		
8		5.7	2.8	4.5		
9		6.3	3.3	4.7		
10		4.9	2.4	3.3		
11		6.6	2.9	4.6		
12		5.2	2.7	3.9		
13		5	2	3.5		
14		5.9	3	4.2		
15		6	2.2	4		
16		6.1	2.9	4.7		
17		5.8	2.6	3.9		

Context menu options: 切り取り(I), コピー(C), 貼り付けのオプション:, 形式を選択して貼り付, 挿入(I)..., 削除(D)..., **数式と値のクリア(N)**, クイック分析(Q)



	A	B	C	D	E	F
1						
2		Sepal Length	Sepal Width	Petal	Petal Width	Class
3						Iris-versicolor
4						Iris-versicolor
5		6.9	3.1	4.9	1.5	Iris-versicolor
6		5.5	2.3	4	1.3	Iris-versicolor
7		6.5	2.8	4.6	1.5	Iris-versicolor
8		5.7	2.8	4.5	1.3	Iris-versicolor
9		6.3	3.3	4.7	1.6	Iris-versicolor
10		4.9	2.4	3.3	1	Iris-versicolor
11		6.6	2.9	4.6	1.3	Iris-versicolor
12		5.2	2.7	3.9	1.4	Iris-versicolor
13		5	2	3.5	1	Iris-versicolor
14		5.9	3	4.2	1.5	Iris-versicolor
15		6	2.2	4	1	Iris-versicolor
16		6.1	2.9	4.7	1.4	Iris-versicolor
17		5.8	2.6	3.9	1.3	Iris-versicolor

消えた!

クリアしたいセルを範囲選択（マウスでドラッグ）したあと、右クリックして、「数式と値のクリア」を選ぶ

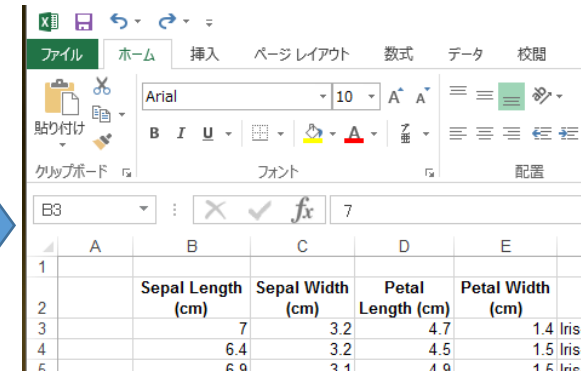
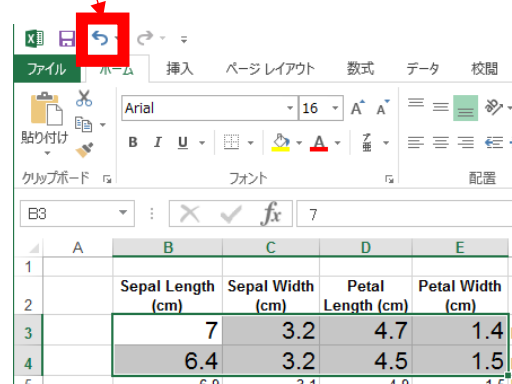
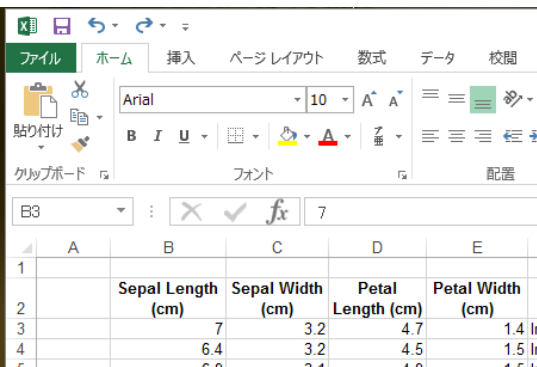
元に戻す操作



「元に戻す」
ボタン



オンライン版の Excel でも
「元に戻す」ボタンはある



何かの操作を
したとする

「元に戻す」ボタン
を押すと元に戻る

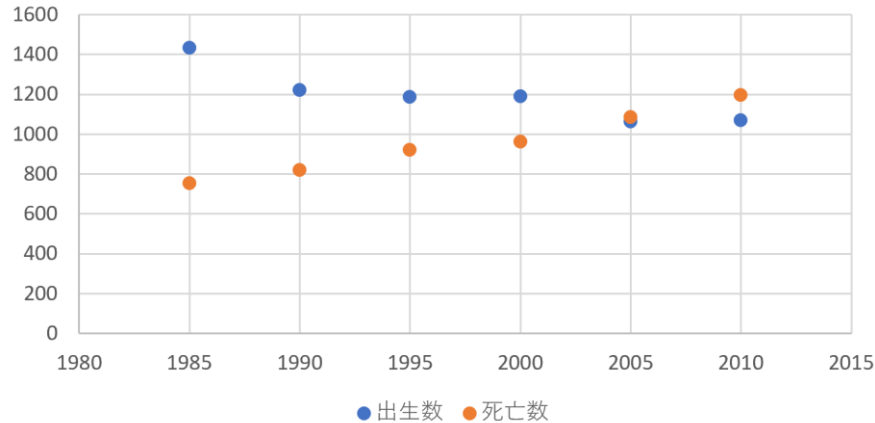
もとに戻す操作は CTRL + Z

(コントロールキーと「Z」を同時押し)でも、できる

2-3 散布図 (Excel を使用)

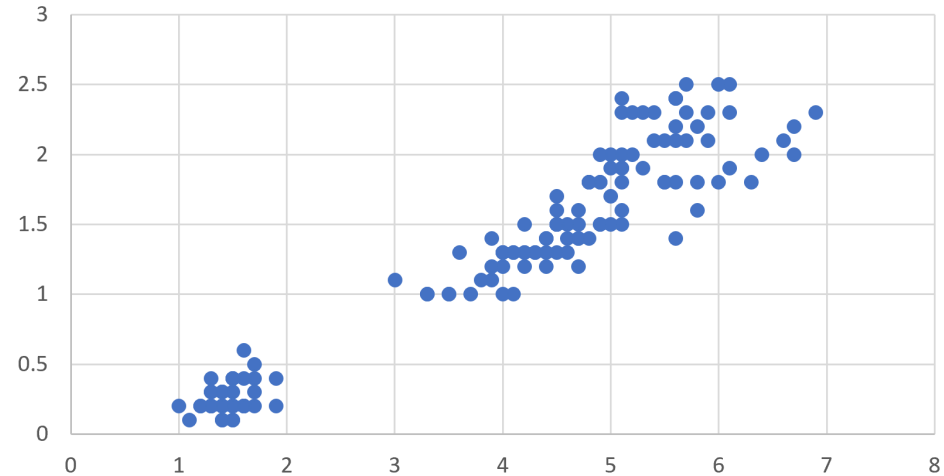
散布図の用途

- 時間変化



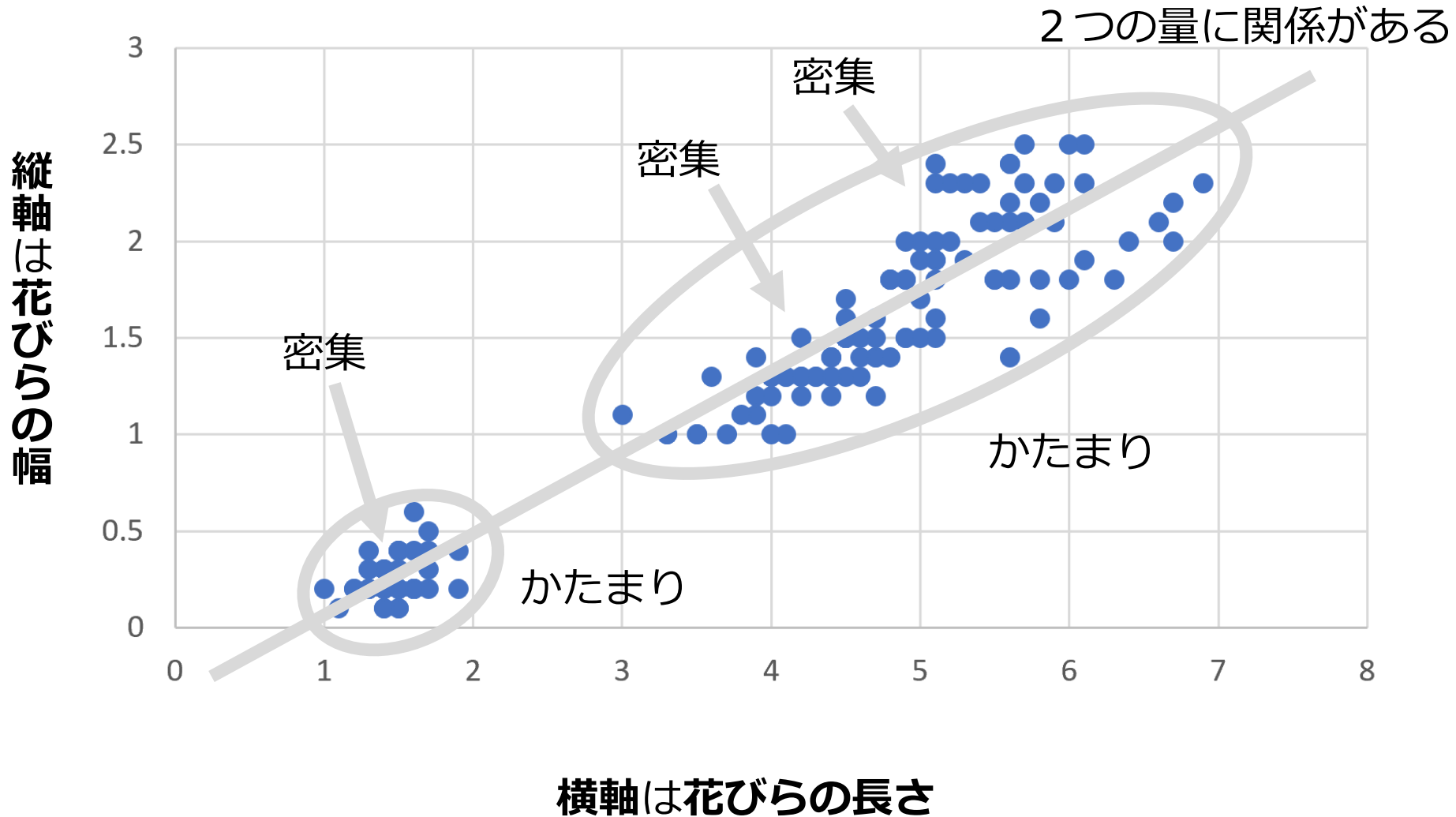
横軸は時間。
散布図から、
時間変化を読み取る

- 分布



横軸と縦軸は、2つの量。
散布図から、
2つの量の間の**関係**を見る

分布から読み取れること



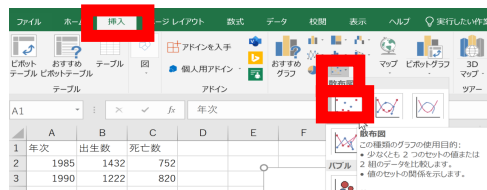
Excel での散布図の作成手順

	A	B	C
1	年次	出生数	死亡数
2	1985	1432	752
3	1990	1222	820
4	1995	1187	922
5	2000	1191	962
6	2005	1063	1084
7	2010	1071	1197

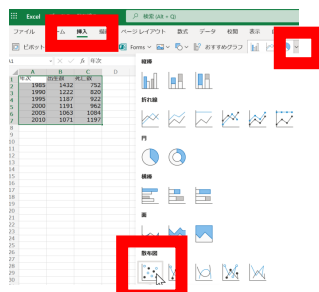
元データ

	A	B	C	D
1	年次	出生数	死亡数	
2	1985	1432	752	
3	1990	1222	820	
4	1995	1187	922	
5	2000	1191	962	
6	2005	1063	1084	
7	2010	1071	1197	
8				

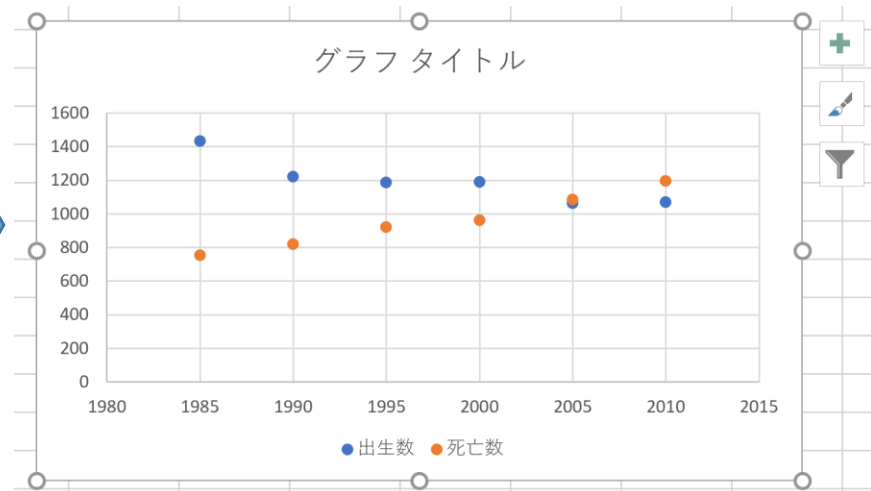
① グラフ化したい部分を範囲選択



アプリ版の Excel



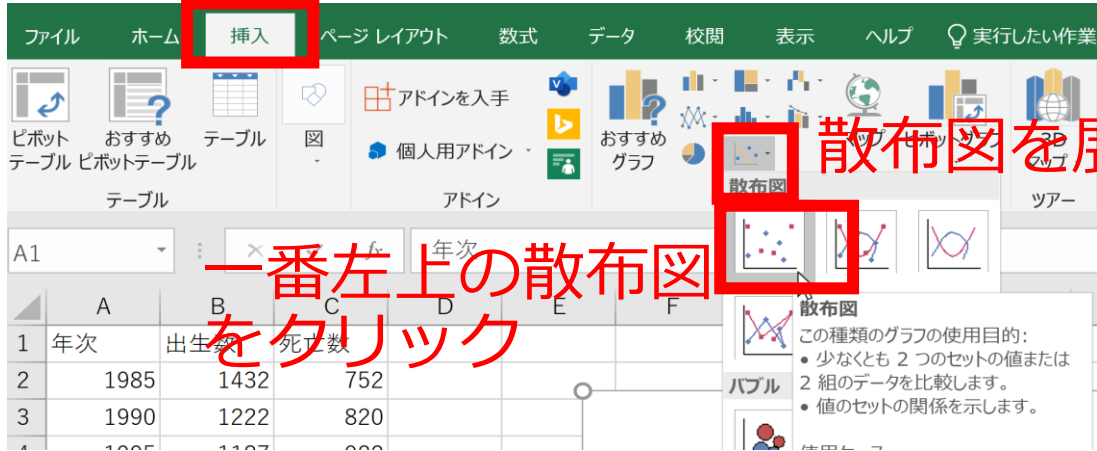
オンライン版の Excel



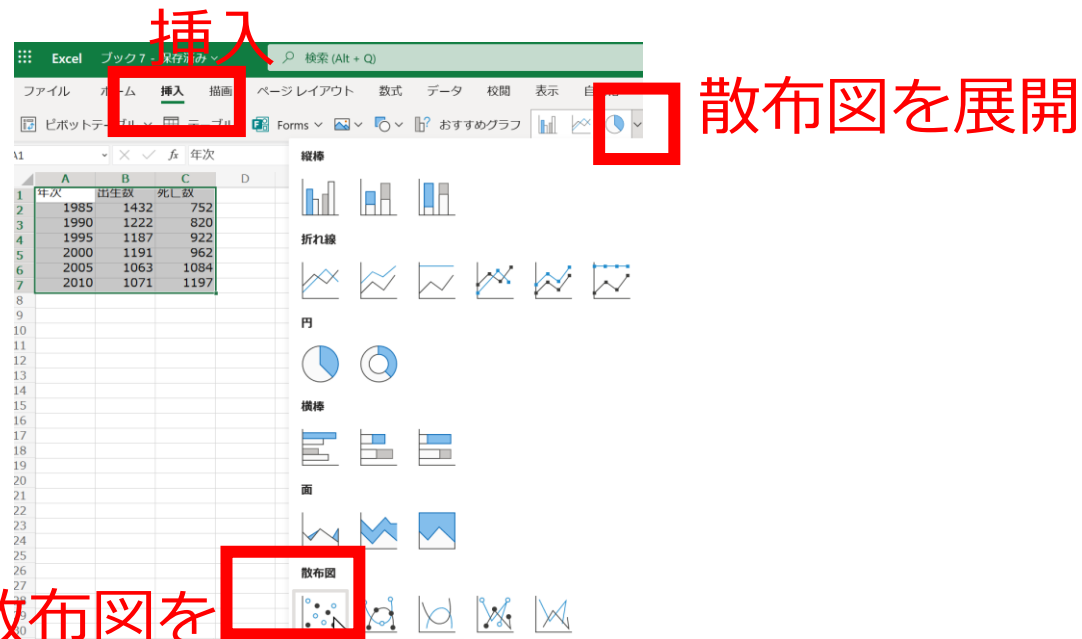
散布図が得られる

② リボンで「挿入」→散布図

Excel での散布図の種類を選択 挿入



アプリ版の Excel

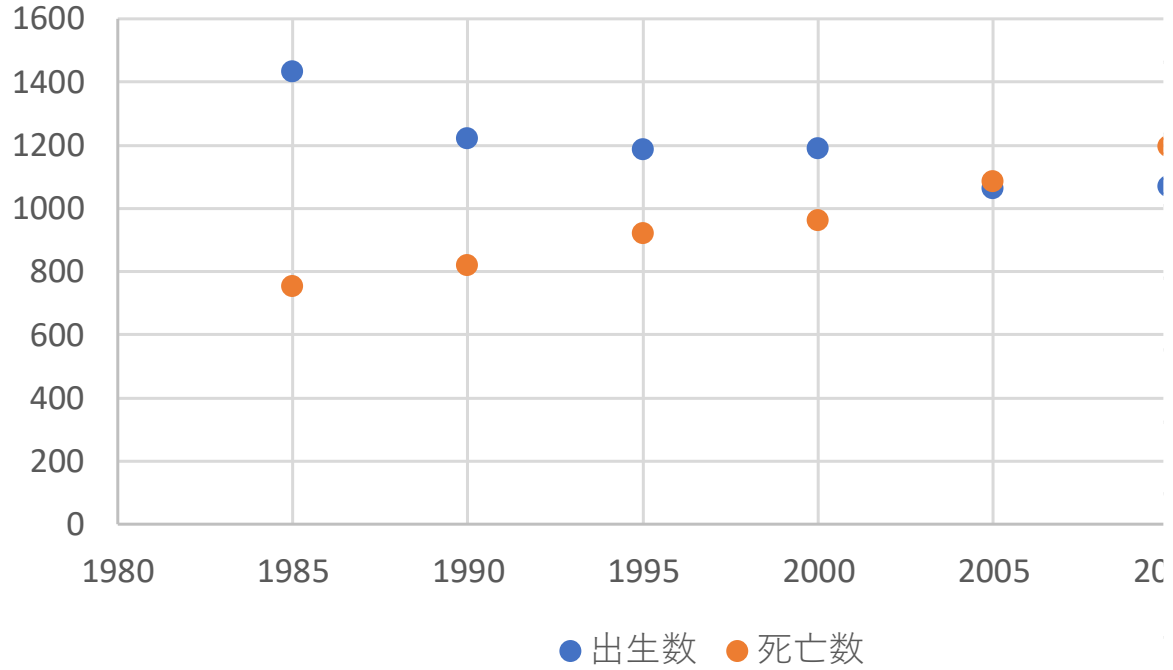


オンライン版の Excel

元データ

年次	出生数	死亡数							
1985	1432	752							
1990	1222	820							
1995	1187	922							
2000	1191	962							
2005	1063	1084							
2010	1071	1197							

グラフタイトル



散布図

2-4 データの合計、平均 (Excel を使用)

Excel で合計を求める SUM

	A	B	C	D	E
1	年次	出生数	死亡数		
2	1985	1432	752		
3	1990	1222	820		
4	1995	1187	922		
5	2000	1191	962		
6	2005	1063	1084		
7	2010	1071	1197		
8			5737		

=SUM(C2:C7) は、
範囲 **C2** から **C7** の
合計を求める

Excel で平均を求める AVERAGE

	A	B	C	D	E
1	年次	出生数	死亡数		
2	1985	1432	752		
3	1990	1222	820		
4	1995	1187	922		
5	2000	1191	962		
6	2005	1063	1084		
7	2010	1071	1197		
8		1194.333	5737		

=AVERAGE(B2:B7) は、
範囲 **B2 から B7** の
平均を求める

平均

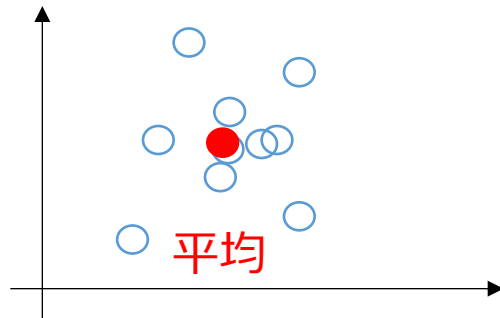
- **平均**の基本, **合計**して, **データの個数で割る**

10, 40, 30, 40 の**平均**: $120 \div 4$ で **30**

- **複数の値の組の平均**を考えることもある

(10, 5), (40, 10), (30, 5), (40, 20) の平均:

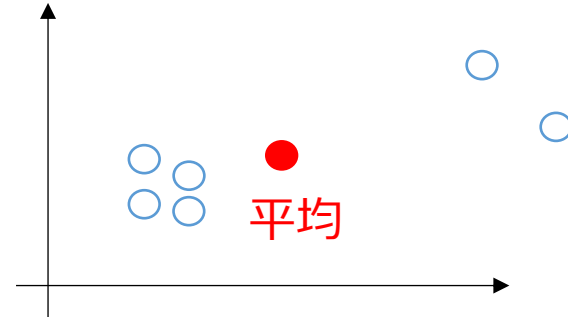
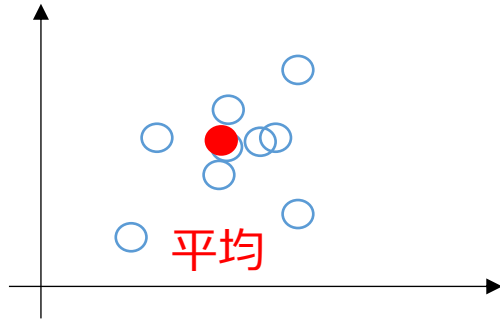
合計は 120 と 40. 4で割って (30, 10)



平均は, **データ集合の代表**とみることができるところがある

計測に**誤差**があるとき,
複数の計測を繰り返して, **平均**をとることで, **誤差を軽減**できることも

平均を使うときの注意点



このような平均に、
意味があるでしょうか？

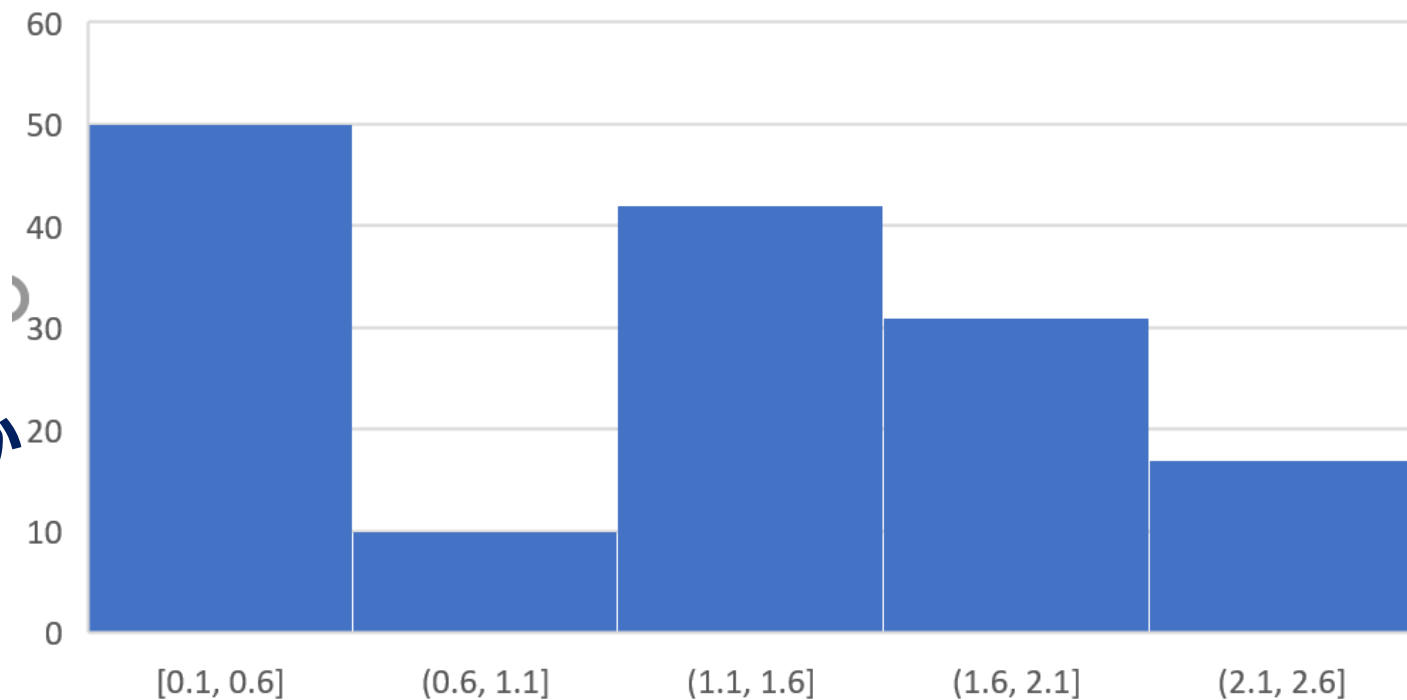
データの分布によっては、平均では役に立たないこともある。
(平均は万能ではない)

2-5 データの分布、密度 (Excel を使用)

ヒストグラム

ヒストグラムは、区間ごとに、データを数え上げたもの

↑
データが
何個あるのか



↑
区間 0.6 ~ 1.1 の
データは 10個

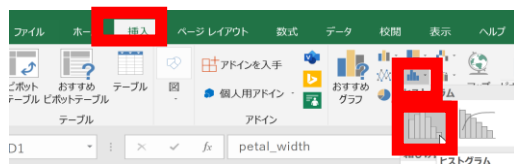
Excel でのヒストグラムの作成手順

	A	B	C	D	E
1	sepal_len	sepal_wid	petal_len	petal_wid	species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	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	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	3.4	1.5	0.2	setosa
8	4.6	3.4	1.4	0.3	setosa
9	5	3.4	1.5	0.2	setosa

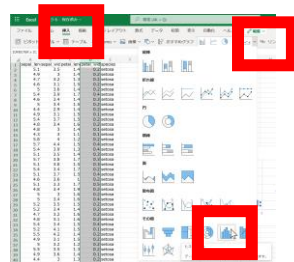
元データ

	A	B	C	D	E
1	sepal_len	sepal_wid	petal_len	petal_wid	species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	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	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	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
12	5.4	3.7	1.5	0.2	setosa

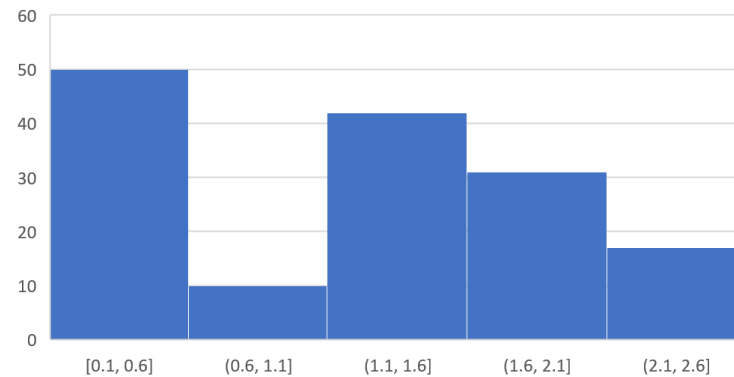
① ヒストグラム化したい列を選択



アプリ版の Excel



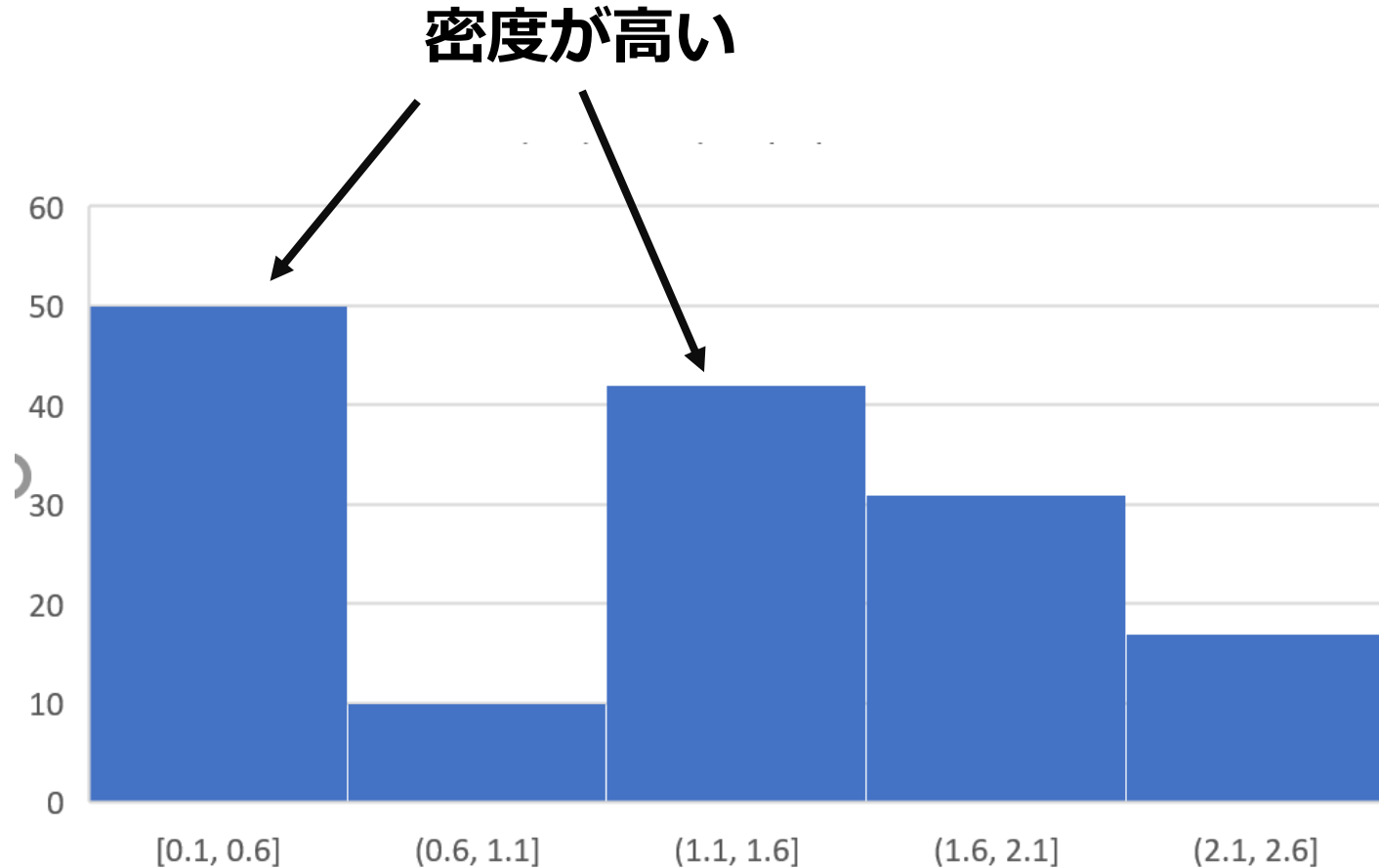
オンライン版の Excel



ヒストグラムが得られる

② リボンで「挿入」→ヒストグラムの選択

ヒストグラムから読み取れること

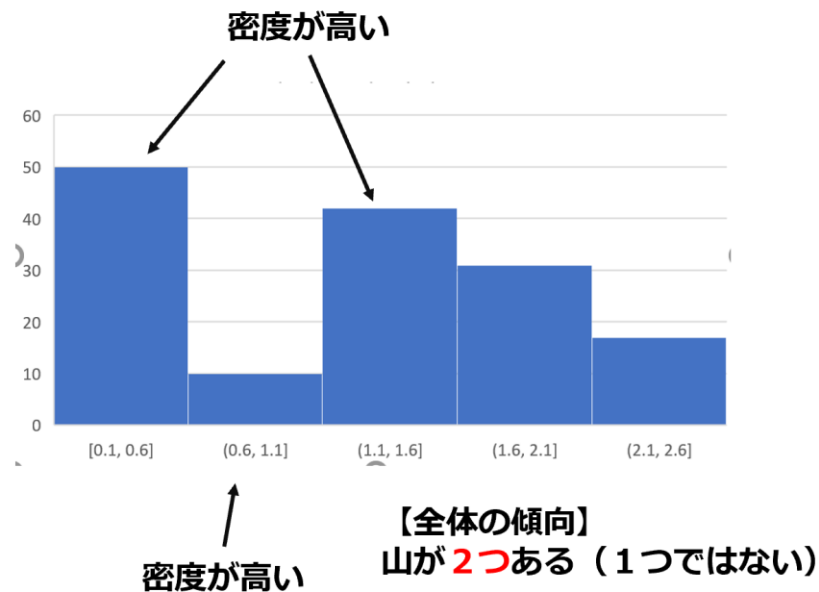
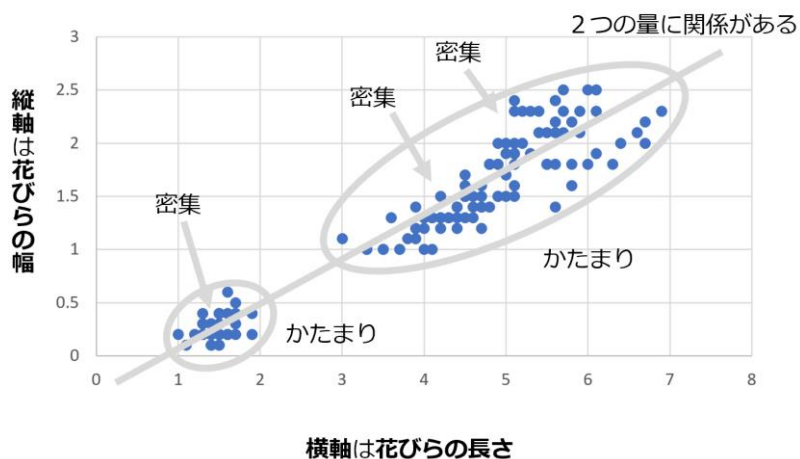


密度が低い

【全体の傾向】
山が2つある（1つではない）

データサイエンスの要点

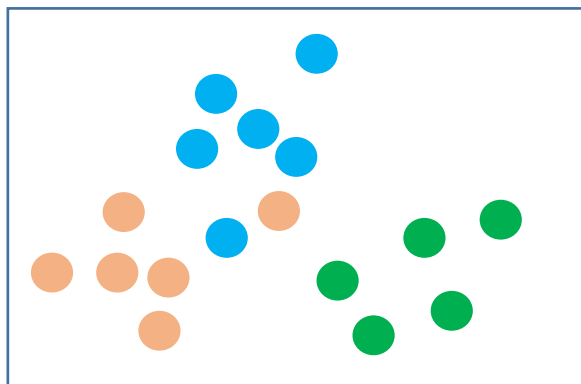
- データから、正しく知見や結論を導くこと



- 決して「難解な数式が出てきて難しい」ものではない
- 正しい手順を踏んで、データから知見や結論を導くことにつながる

2-6 データの分類, 特徴抽出 での人工知能の応用

大量の**分類済み**
データ



3 種類に分類済み

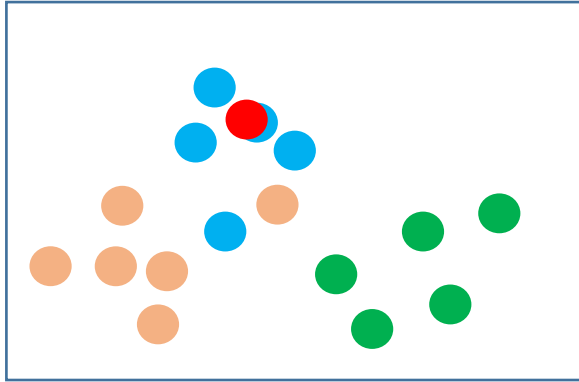
学習



学習者

**データの傾向, どのような
基準で分類されているか
などを学習 (特徴抽出)**

上にあれば：水色
左下にあれば：オレンジ
右下にあれば：緑色



新しいデータ



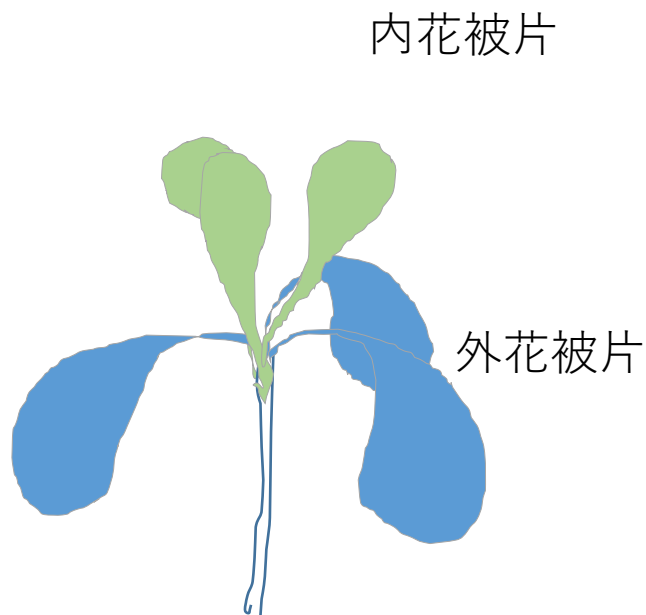
学習者

新しいデータを自動で
分類できる能力を獲得

新しいデータは：赤色

- 間違えることもある
- 学習に使うデータは多いほど良い結果になる

アヤメ属 (Iris)



- 多年草
- 世界に 150種. 日本に 9種.
- 花被片は 6個
- 外花被片 (がいかひへん) Sepal 3個 (大型で下に垂れる)
- 内花被片 (ないかひへん) Petal 3個 (直立する)

Iris データセット

Iris データセットのうち、先頭 10 行

	A	B	C	D	E
1	sepal_len	sepal_wid	petal_len	petal_wid	species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	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	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	3.4	1.5	0.2	setosa
10	4.4	2.9	1.4	0.2	setosa

外花被片 (Sepal) の長さ
内花被片 (Petal) の長さ
種類

◆ 3種のアヤメの**外花被**、**内花被片**を計測

◆ **種類**のデータも

setosa

versicolor

virginica

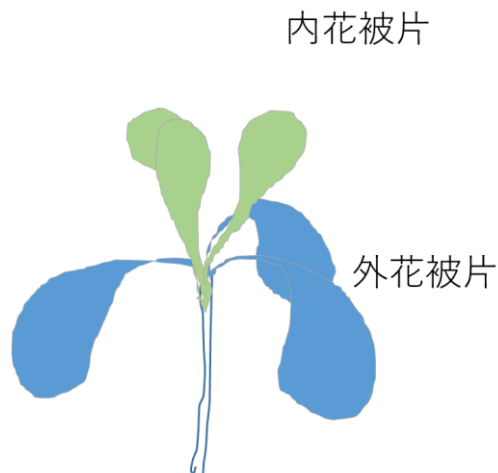
◆ データ数は **50 × 3**

作成者 : Ronald Fisher

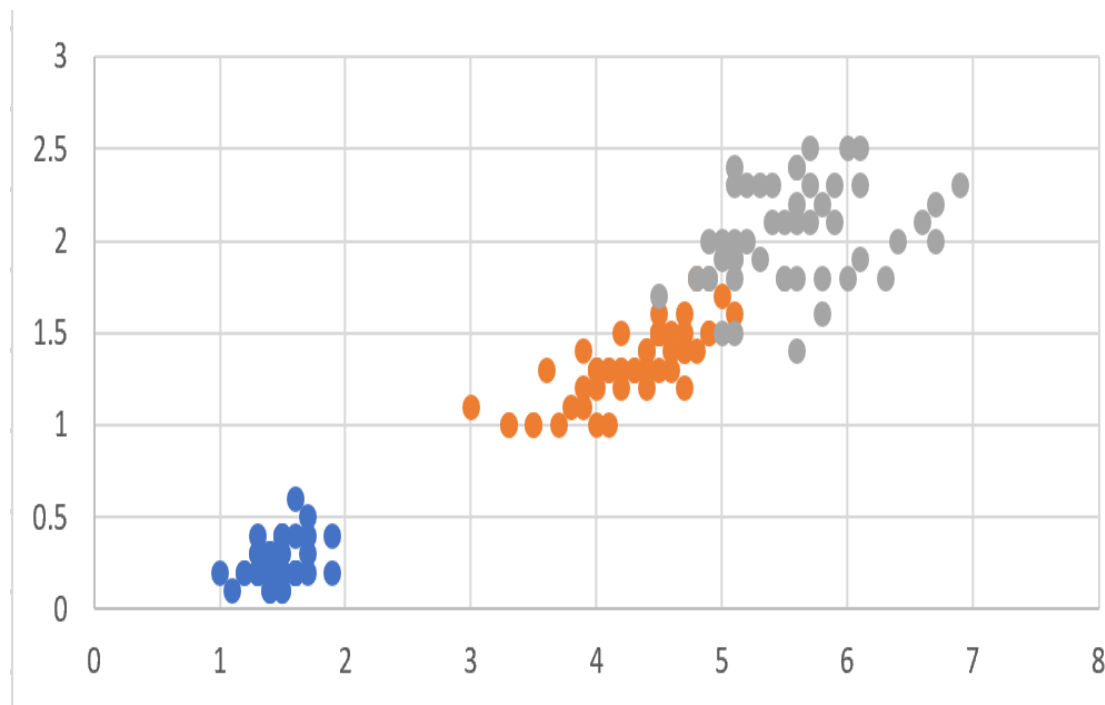
作成年 : 1936

Iris データセットの散布図

アヤメ属 (Iris)



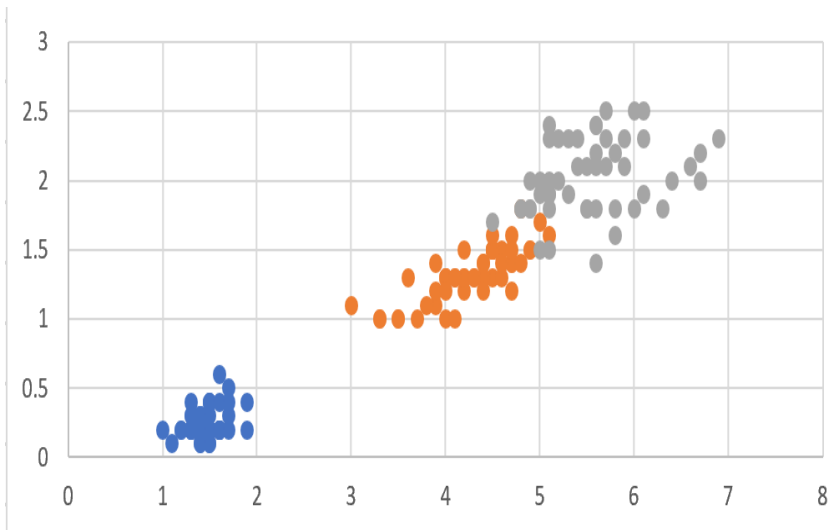
縦軸
..
内花被片の幅



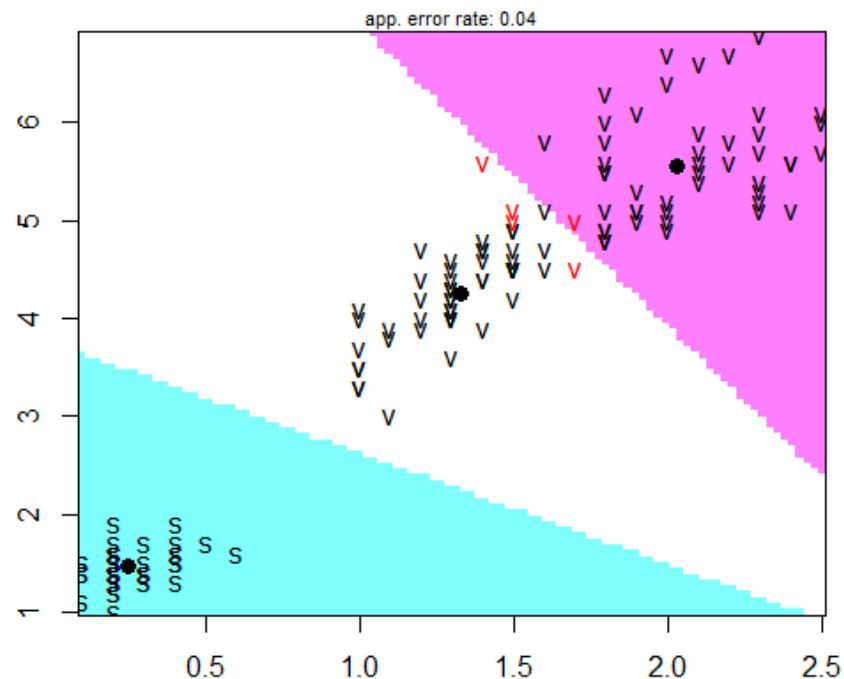
横軸：内花被片の長さ

次の **3種類** の分類済みのデータ
setosa
versicolor
virginica

人工知能分野の手法（LDA法）による分析例



元データ



**LDA 法の結果
空間が分けされた。**

**新しいデータを自動で
分類できる能力を獲得**

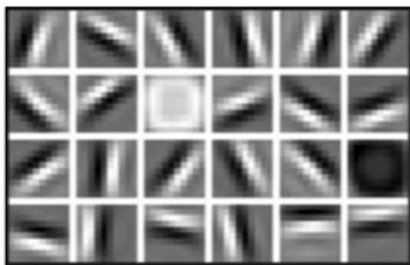
人工知能の一種であるニューラルネットワークは、「データからのパターンの抽出」を行っているという考え方も

Why Deep Learning?

Hand engineered features are time consuming, brittle, and not scalable in practice

Can we learn the **underlying features** directly from data?

Low Level Features



Lines & Edges

Mid Level Features



Eyes & Nose & Ears

High Level Features



Facial Structure



ニューラルネットワークが扱うさまざまなレベルのパターン

画像分類を行うオンラインサービス



元画像

Faces Objects **Labels** Properties Safe Search

The screenshot shows the 'Labels' tab of the Google Cloud Vision API interface. It displays a list of detected labels with their corresponding confidence percentages and progress bars. The labels are: Forehead (98%), Glasses (98%), Chin (97%), Eyebrow (95%), Vision Care (94%), Dress Shirt (90%), Jaw (88%), and Smile (88%).

Label	Confidence
Forehead	98%
Glasses	98%
Chin	97%
Eyebrow	95%
Vision Care	94%
Dress Shirt	90%
Jaw	88%
Smile	88%

126.png

画像分類の結果

URL : <https://cloud.google.com/vision/docs/drag-and-drop>

データの分類, 特徴抽出での人工知能の応用

- データによる**学習**を行う
- **学習を重ねることで上達する**
- **データは大量にあればあるほど良い**

- 「**学習**によって, **未知のデータに対しても当てはまるパターンや規則を抽出 (特徴抽出)** している」という考え方もある

人工知能の利用での注意点

思い込みを疑い，根拠を確認することが大切。

《思い込みの例》

- 「人工知能だから，100%正解」
- 「人工知能の能力は，必ず，人間を上回る」
- 「人工知能に，すべてを任せても大丈夫」

人工知能は，データによる学習，知的なプログラムで動く．
間違った結果を出すこともある．万能ではない．