

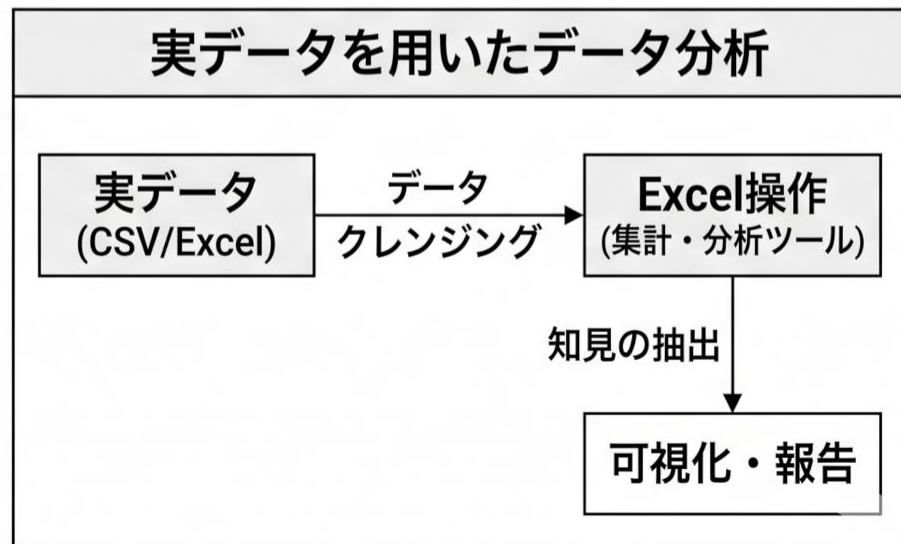
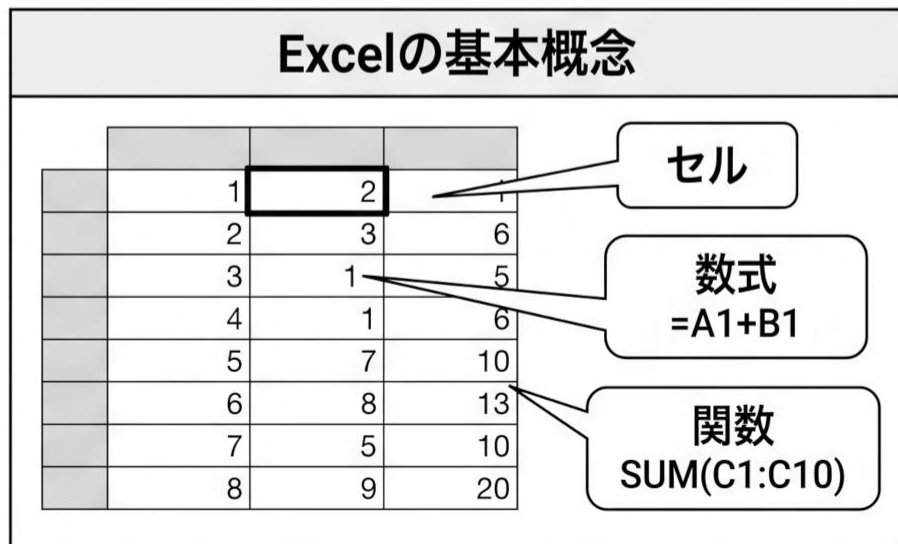
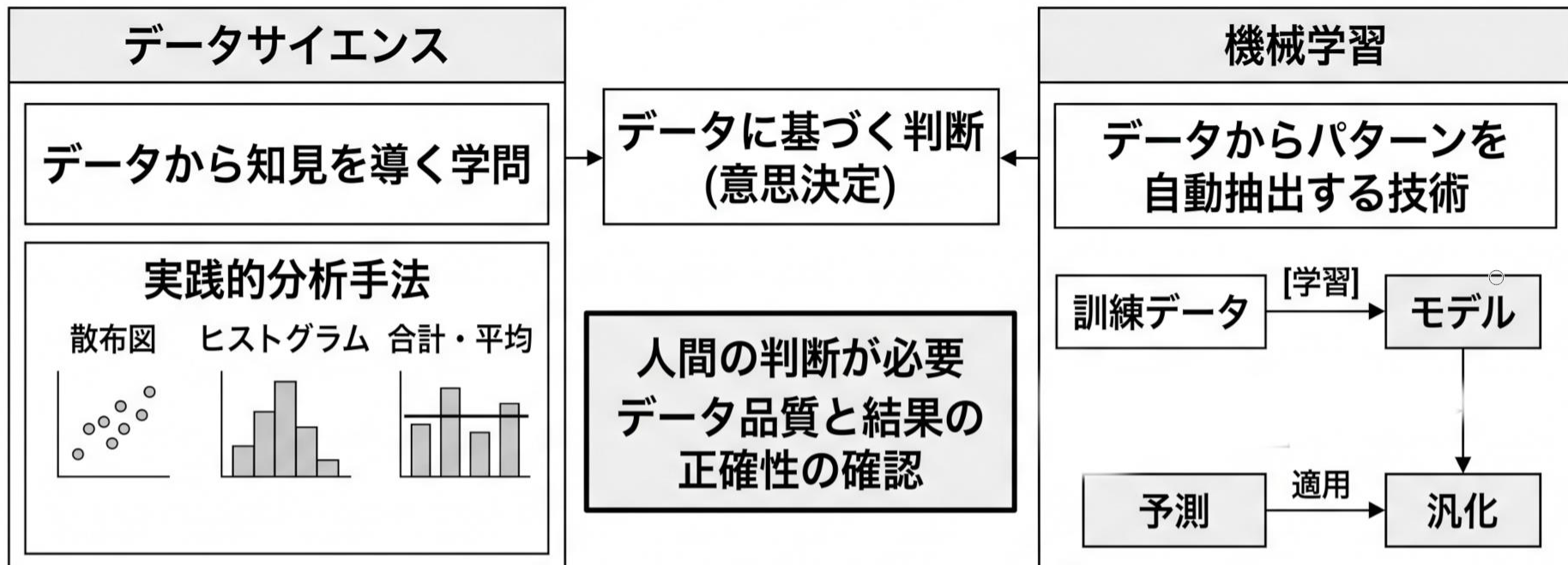
# aa-2. データサイエンス・AI の事例、 技術

(人工知能)

金子邦彦



# 内容



- **データサイエンス**は**データから知見を導く**学問である。**散布図・ヒストグラム・合計・平均**などの実践的分析手法がある。
- **機械学習**は**データからパターンを自動抽出する技術**であり、**訓練データ**による**学習・予測・汎化**の仕組みを理解する。
- **データサイエンス、機械学習**は**データに基づく判断を行うための**ものであり、広く活用されている。**結果の正確性やデータ品質には人間の判断が必要**。
- Excel の基本概念（セル・数式・関数）、実データをを用いたデータ分析

# 2-1 データサイエンス

- 情報化社会において**多くのデータが**生み出されており、**さまざまな分野で活用**されている。
- **データサイエンス**も、後述の**機械学習**も、「**データに基づく**」という共通の原理を持つ。

- **データから正しい知見や結論を導き出すための学問.**
- **大量のデータを扱う**
- 人工知能や情報処理とも大きく関連する.
- **データの分析, グラフなどのデータの可視化を用いて, データから有益な情報を引き出す**
- **さまざまな分野でデータを活用する実力につながるため, 大学生にとって身につけるべきスキルである.**

# データサイエンスの主要な活用分野



- **医療分野**：病気の早期発見や効果的な治療法の開発
- **工学分野**：製品品質の改善や予測保全など，生産の最適化
- **ビジネス分野**：顧客の嗜好やニーズの分析，マーケティング戦略の立案

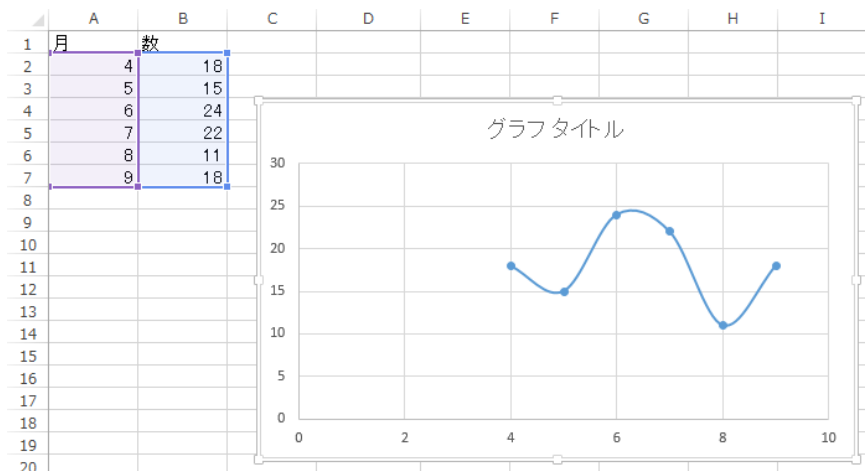
## 2-2 表計算ソフトウェア Excel

# 表計算ソフトウェア Excel

- 機能：データの記録、保管、共有、集計・集約、グラフ作成など
- 表の作成，合計の自動再計算，条件に合致するデータの強調表示，並べ替えなどが簡単にできる。

|   | 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 |    |    |    |



# Excelの便利機能① 自動計算

## 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

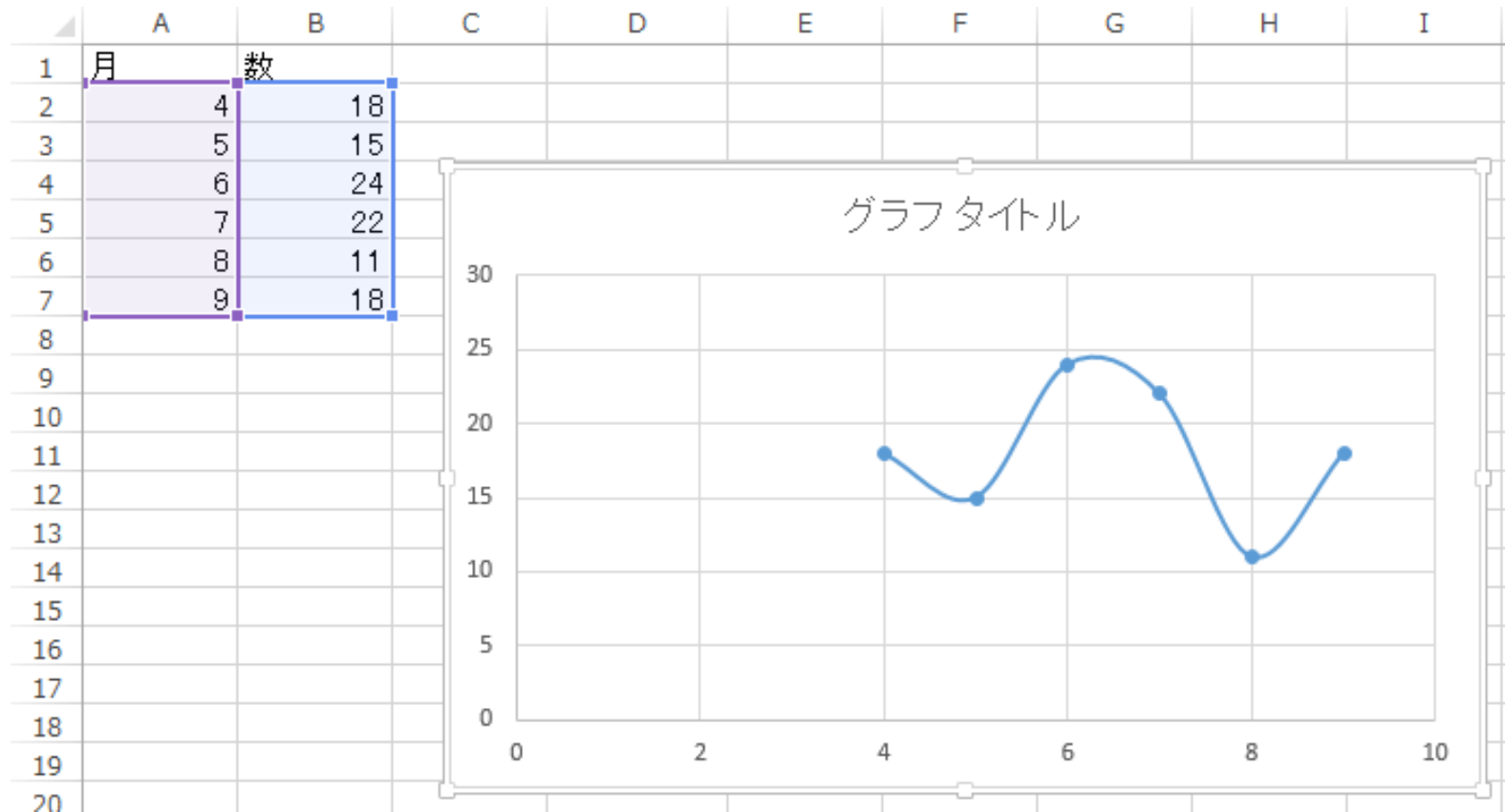
|   | A   | B  | C  | D    |
|---|-----|----|----|------|
| 1 | 品名  | 単価 | 数量 | 合計   |
| 2 | りんご | 80 | 10 | 800  |
| 3 | みかん | 50 | 5  | 250  |
| 4 |     |    |    | 1050 |

合計が**自動**で再計算される

Excel の画面

# Excelの便利機能② グラフ作成

## グラフ



# Excelの便利機能③ データ整理

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 |    |    |    |

# パソコンの基本機能



- **文書作成**

文書の編集、清書、目次や表の作成

- **表計算**

データの管理、計算、グラフ作成

- **プレゼン**

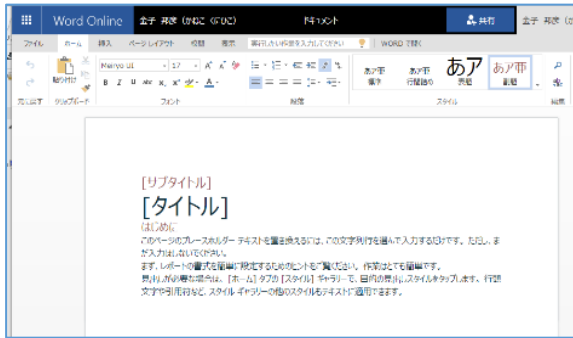
ビジュアル資料作成

- **インターネット**

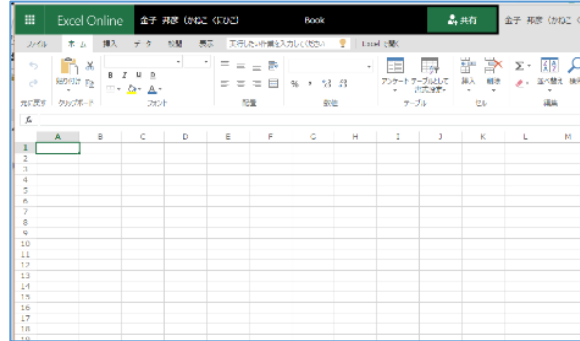
情報収集、コミュニケーション

データはすべて**デジタル**

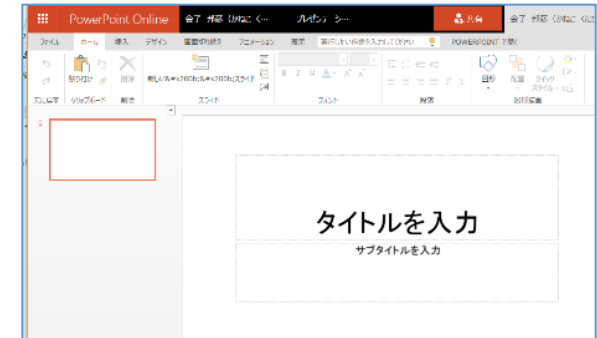
# Microsoft 365 の主な機能



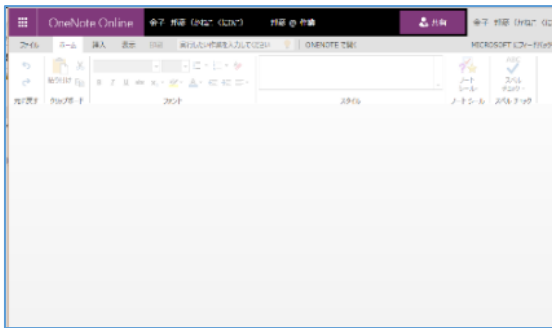
ワード (文書作成)



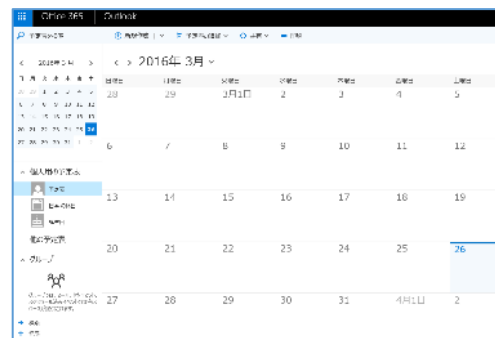
エクセル (表計算)



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

# Microsoft 365 の種類



**2種類**ある。この授業では、**どちらを使用しても問題ない**

- **Microsoft 365 のオンライン版**

**WEBブラウザ**で使う。

.

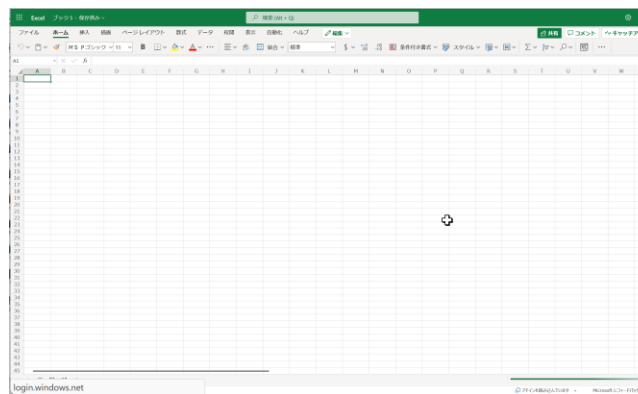
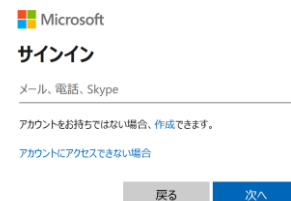
- **Microsoft 365 のアプリ版**

**前もってインストールが必要。**

WEBブラウザで使用。 <https://portal.office.com> に各自サインイン。

## 【起動手順】

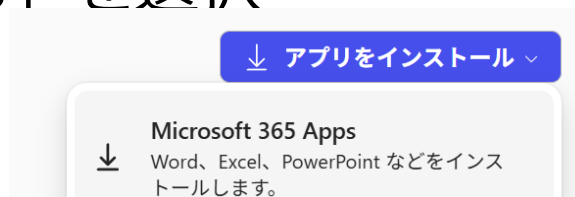
- ① Webブラウザで <https://portal.office.com> を開く
- ② **電子メールアドレス**を入力し「次へ」をクリック  
(例) p1234567@fukuyama-u.ac.jp
- ③ **パスワード**を入力し「サインイン」をクリック
- ④ **メニューでExcelを選択**
- ⑤ **ブックの種類を選択**
- ⑥ **Excelの画面が開く**



事前にインストールが必要。インストール時に大量の通信が発生する（時間・通信費用に注意）。

## 【インストール手順】

- ① Webブラウザで <https://portal.office.com> を開く
- ② **電子メールアドレス**を入力し「次へ」をクリック  
（例） p1234567@fukuyama-u.ac.jp
- ③ **パスワード**を入力し「サインイン」をクリック
- ④ 「**アプリ**」をクリック。「**アプリをインストール**」 → 「**Microsoft 365 Apps**」を選択

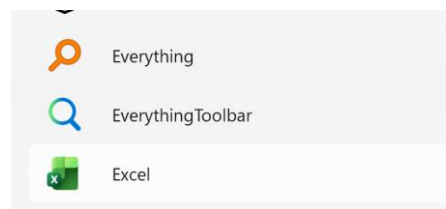


- ⑤ 画面の指示に従いインストール

【新しく Excel を使う場合】

※既存の Excel ファイルを開くと Excel が起動する

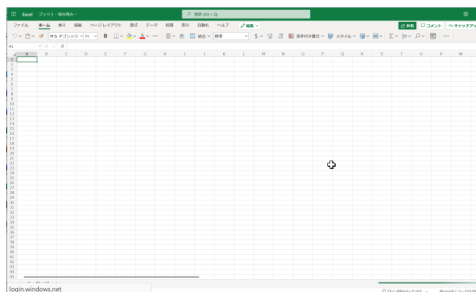
① スタートメニュー等でExcelを選択



② ブックの種類を選択



③ Excelの画面が開く



## 2-3 Excel の基本概念

# 画面構成

メニュー, リボン, 数式バー, ワークシートで構成される。  
オンライン版・アプリ版ともに同様の構成。

## メニュー

画面上部に並ぶ項目（「挿入」等）

## リボン

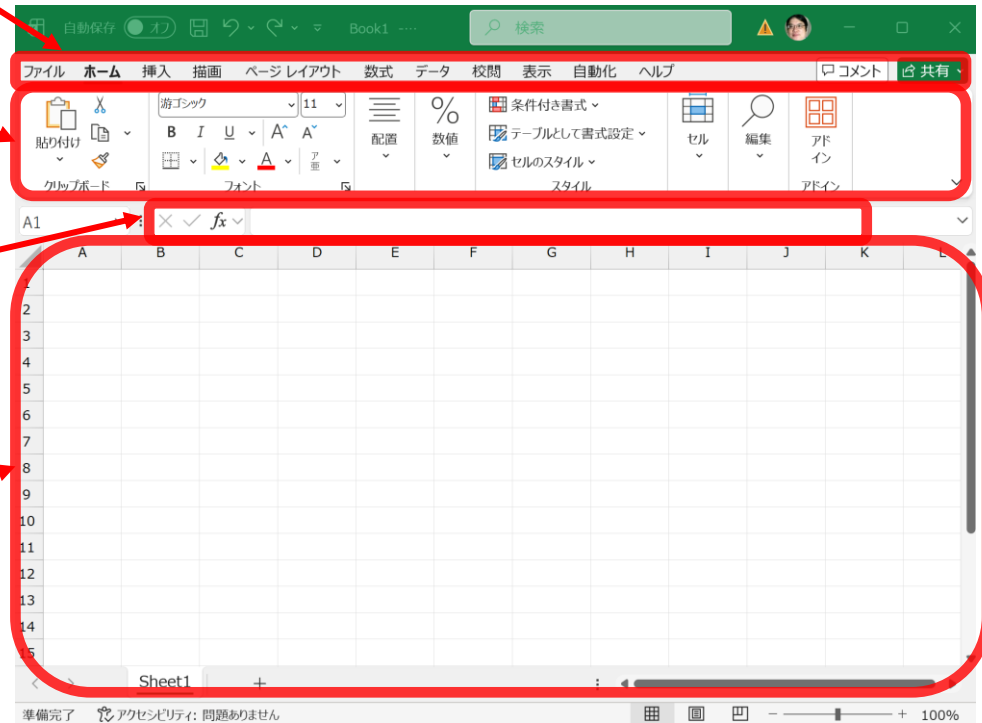
メニュー選択時に表示される操作ボタン群の領域。グラフの挿入等はリボンから操作

## 数式バー

アクティブセルの内容（値や数式）を表示・編集する領域。

## ワークシート

シートともいう。表形式で値や数式を並べる領域。グラフの挿入も可能。

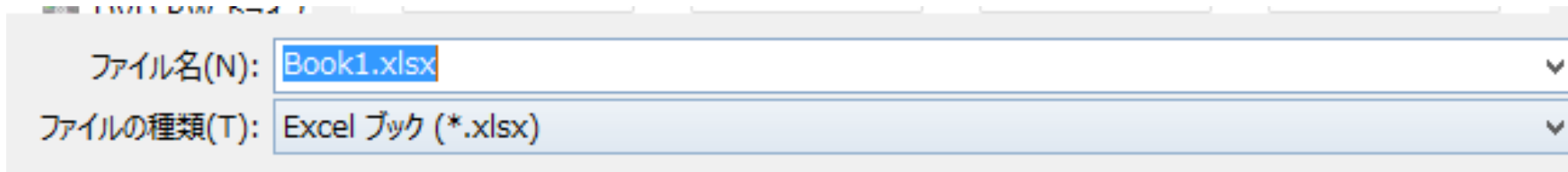


アプリ版の Excel の画面

# Excel のブック

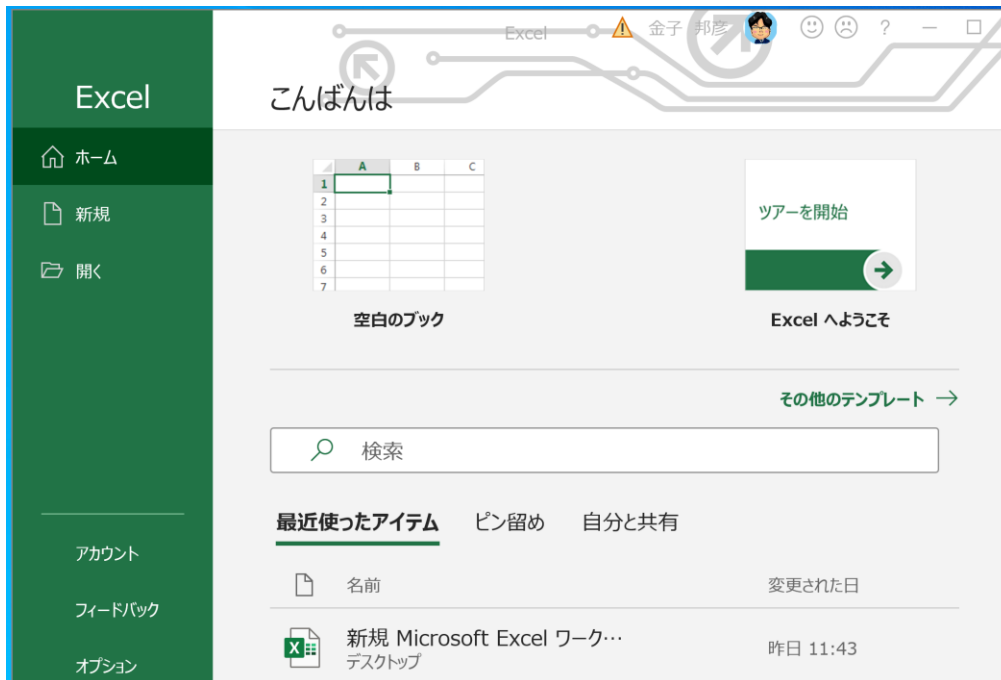


- Excel の**ブック**は、Excel の**ファイル**のこと
- 1つまたは複数の**ワークシート**を1つのブックに保存できる。
- 保存時にファイル名を設定可能。



# スタート画面

Excel起動時に最初に表示される画面。ブックの種類  
の選択や、過去の履歴の確認ができる

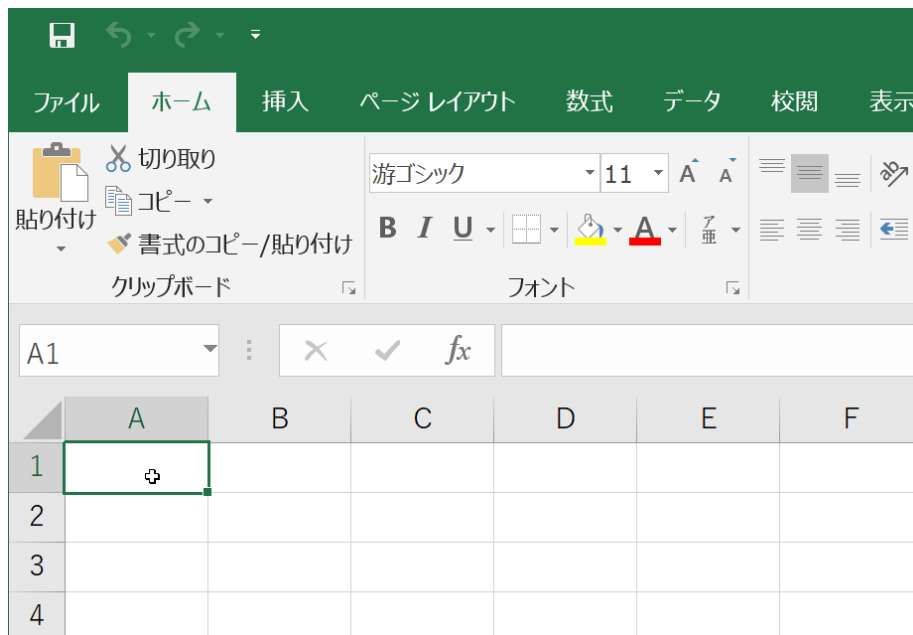


アプリ版の Excel の画面

# セル、番地（セル参照）、アクティブセル



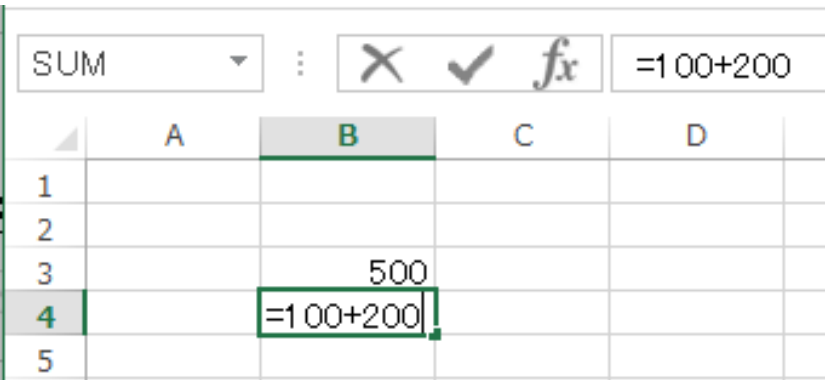
- **セル**：ワークシートを構成する1つ1つのマス目。値や数式を入力する基本単位。
- **番地（セル参照）**：列のアルファベットと行の数字の組み合わせでセルの位置を示す。例：**B3**は列B・行3のセル。
- **アクティブセル**：現在選択されている編集集中のセル。



## 2-4 Excel の基本操作

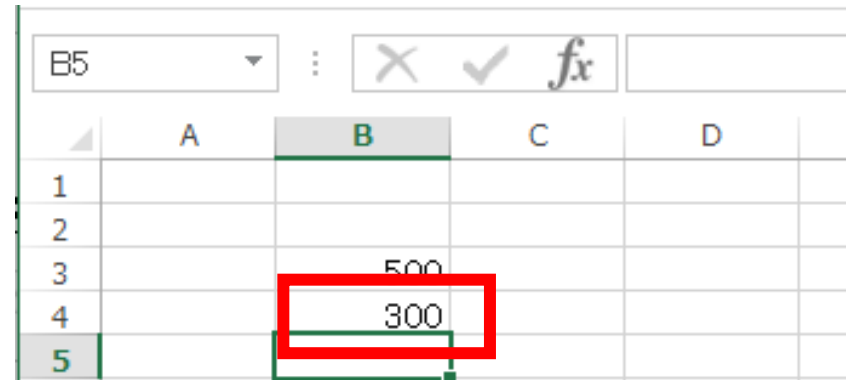
# 数式の入力

数式を入力したいときは、頭に半角の「=」を付ける



SUM : X ✓ fx =100+200

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



B5 : X ✓ fx

|   | A | B   | C | D |
|---|---|-----|---|---|
| 1 |   |     |   |   |
| 2 |   |     |   |   |
| 3 |   | 500 |   |   |
| 4 |   | 300 |   |   |
| 5 |   |     |   |   |

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

# 数式の入力

- 先頭に半角の「=」を付ける。（例）「=100+200」
- 数式は半角文字で記述する。
- 番地（「B3」「B4」等）を数式内に記述できる。  
（例）「=B3+B4」
- Enterキーを押すと自動計算される。

|   | 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

## 数式バーでの確認・修正

- セルをクリックすると数式バーに数式が表示される（修正も可能）
- セルをダブルクリックすると数式バーとセル内の両方に数式が表示される（どちらでも修正可能）。

|   | A | B   | C | D |
|---|---|-----|---|---|
| 1 |   |     |   |   |
| 2 |   |     |   |   |
| 3 |   | 500 |   |   |
| 4 |   | 300 |   |   |
| 5 |   |     |   |   |



数式バー

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

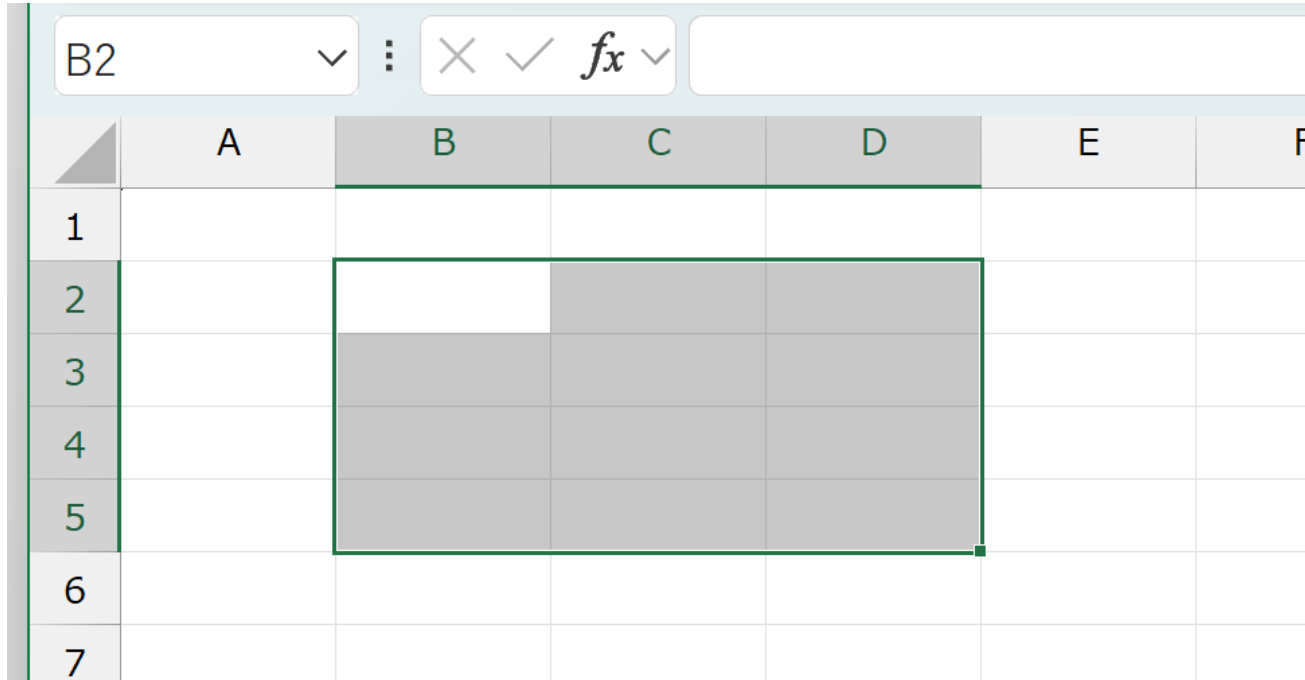
「300」のところを  
ダブルクリック

アクティブセル

# 範囲選択



- マウスのドラッグで複数セルを選択する操作.



# セルの数式と値のクリア

C4 : X ✓ fx 3.2

|    | 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              |     |     |                 |

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け(S)..., 挿入(I)..., 削除(D)..., **数式と値のクリア(N)**, クリック分析(Q)



C4 : X ✓ fx

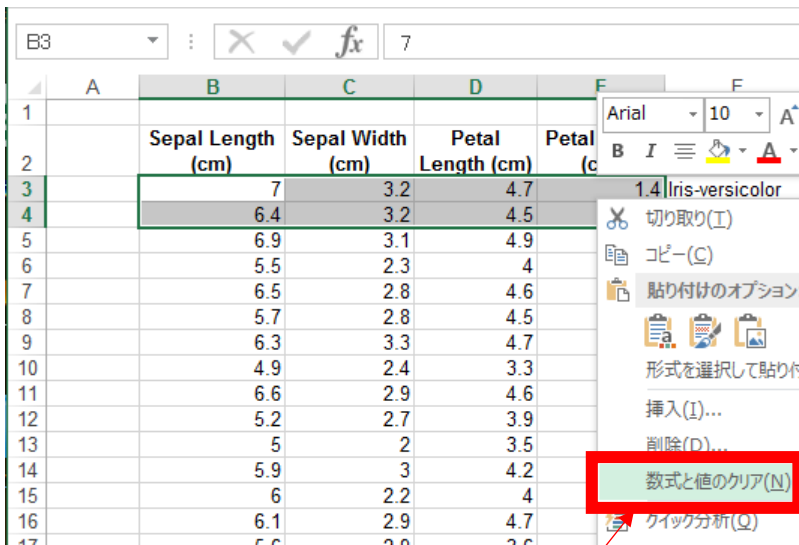
|    | A | B                 | C                | D                 | E                | F               |
|----|---|-------------------|------------------|-------------------|------------------|-----------------|
| 1  |   |                   |                  |                   |                  |                 |
| 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 |

Red box highlights the empty cell C4. Red text "消えた!" (Disappeared!) is written over the table.

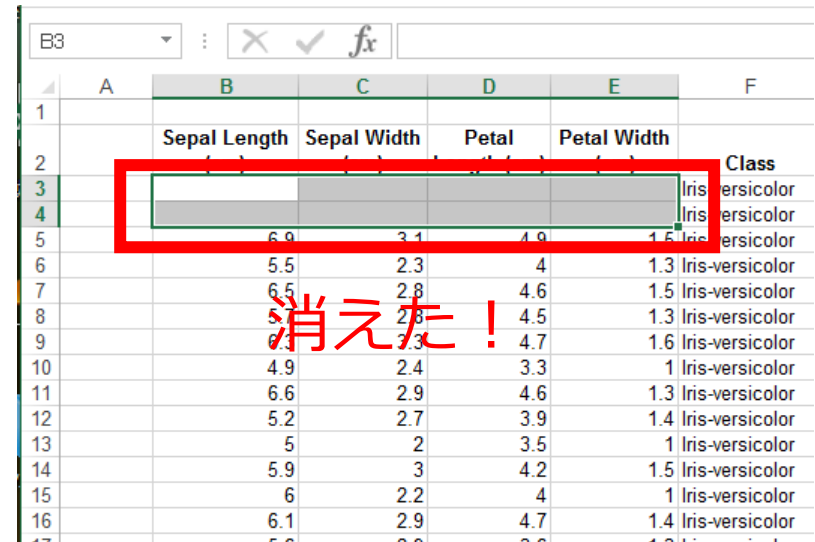
クリアしたいセルを右クリックして、  
「数式と値のクリア」を選ぶ

# セルの数値と値のクリア

セルを右クリックし「数式と値のクリア」を選択する。  
範囲選択後に右クリックしても同様の操作が可能。



|    | 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               | 3.2              | 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 |   | 5.8               | 2.8              | 3.9               | 1.4              | Iris-versicolor |



|    | 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          | 2.8         | 3.9   | 1.4         | 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

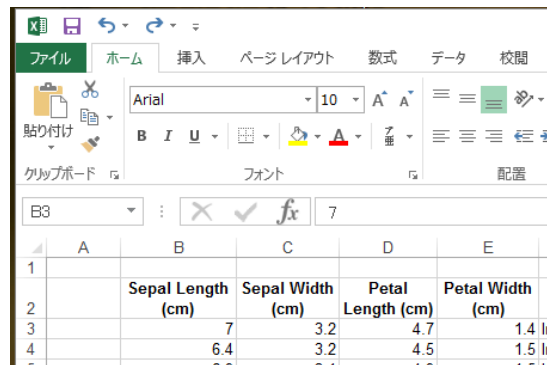
「元に戻す」ボタン，またはCtrl+Z（同時押し）で直前の操作を取り消せる。オンライン版でも同様。



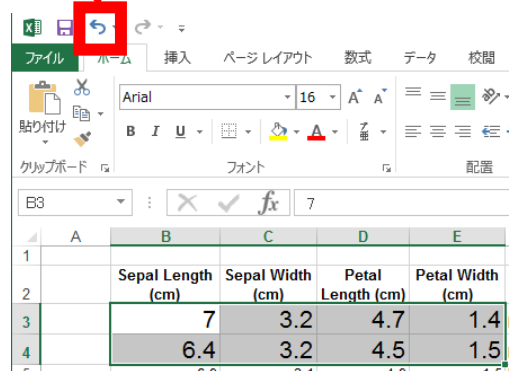
「元に戻す」  
ボタン



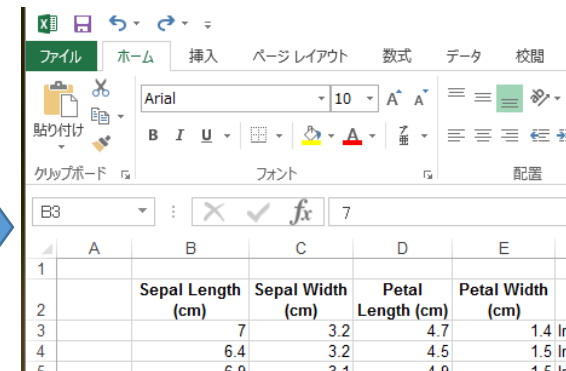
オンライン版の Excel でも  
「元に戻す」ボタンはある



|   | A | B                 | C                | D                 | E                |
|---|---|-------------------|------------------|-------------------|------------------|
| 1 |   |                   |                  |                   |                  |
| 2 |   | Sepal Length (cm) | Sepal Width (cm) | Petal Length (cm) | Petal Width (cm) |
| 3 |   | 7                 | 3.2              | 4.7               | 1.4              |
| 4 |   | 6.4               | 3.2              | 4.5               | 1.5              |



|   | A | B                 | C                | D                 | E                |
|---|---|-------------------|------------------|-------------------|------------------|
| 1 |   |                   |                  |                   |                  |
| 2 |   | Sepal Length (cm) | Sepal Width (cm) | Petal Length (cm) | Petal Width (cm) |
| 3 |   | 7                 | 3.2              | 4.7               | 1.4              |
| 4 |   | 6.4               | 3.2              | 4.5               | 1.5              |



|   | A | B                 | C                | D                 | E                |
|---|---|-------------------|------------------|-------------------|------------------|
| 1 |   |                   |                  |                   |                  |
| 2 |   | Sepal Length (cm) | Sepal Width (cm) | Petal Length (cm) | Petal Width (cm) |
| 3 |   | 7                 | 3.2              | 4.7               | 1.4              |
| 4 |   | 6.4               | 3.2              | 4.5               | 1.5              |

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

# Excel の関数と範囲指定



- **関数の引数で範囲を指定**するとき、**コロン「:」**を使用する。

(例) 「**C2:C7**」は**セルC2からC7までの範囲**を意味する。

- **SUM (合計)**

**=SUM(C2:C7)** で**範囲C2～C7の合計**を求める。

- **AVERAGE (平均)**

**=AVERAGE(B2:B7)** で**範囲B2～B7の平均**を求める。

# 合計 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

# 平均 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**は、データの**合計**を、**データの個数**で割ったもの

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

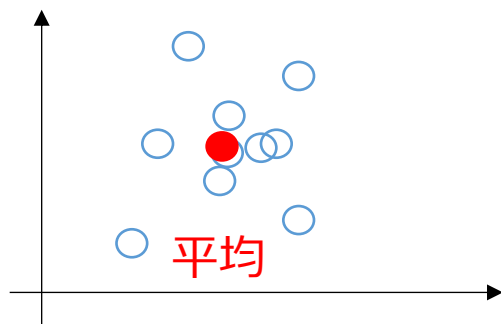
- **複数の値の組の平均**を考えることもある (例)  
(例) (10,5), (40,10), (30,5), (40,20) → 合計(120, 40)  $\div 4 = (30, 10)$

- **代表値としての平均**

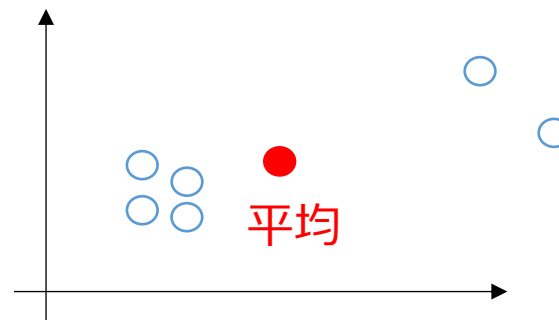
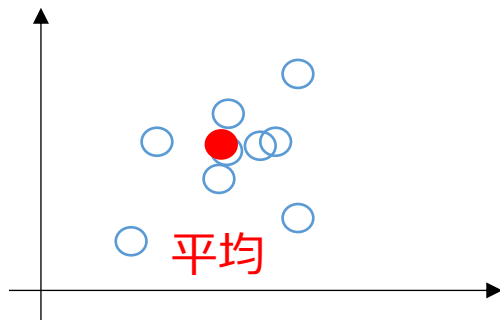
平均はデータ集合の代表とみなせる場合がある。

- **誤差軽減のための平均**

計測に誤差があるとき、複数回の計測の平均をとることで誤差を軽減できる場合がある。



# 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？

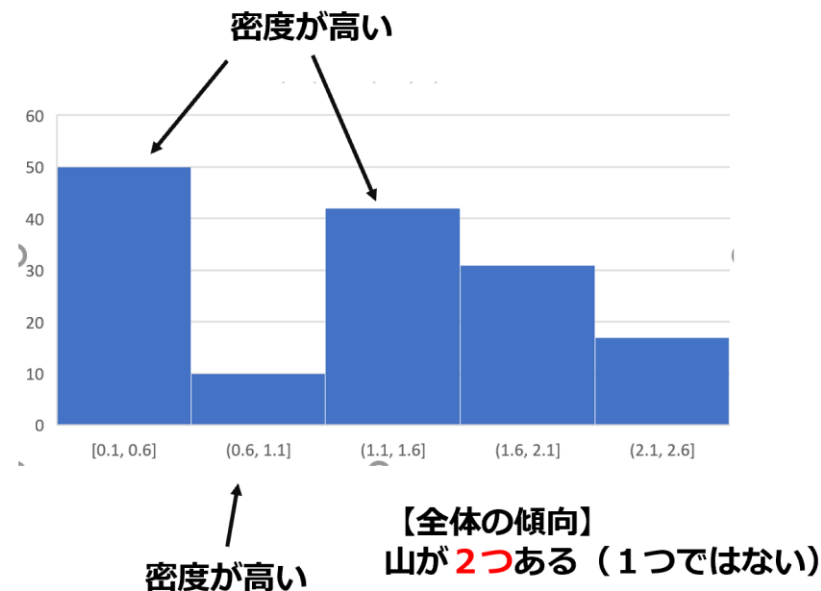
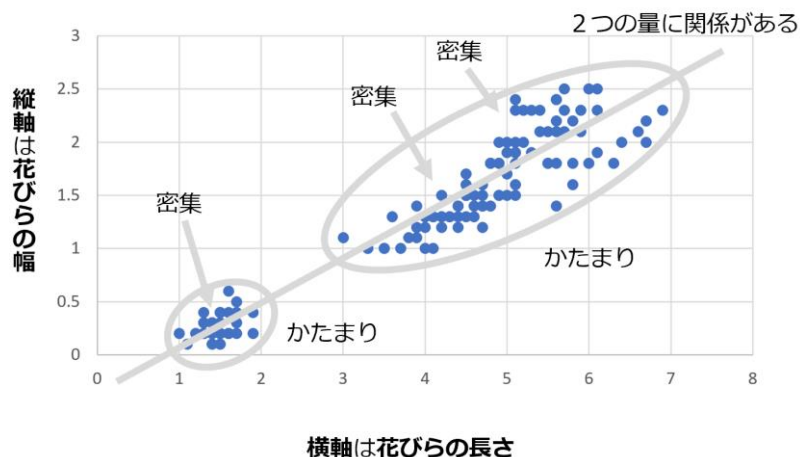
**データの分布によっては、平均では役に  
立たないこともある。**  
(平均は万能ではない)

# 2-5 Excelを用いたデータ分析 の実践

# Excelを用いたデータ分析の実践



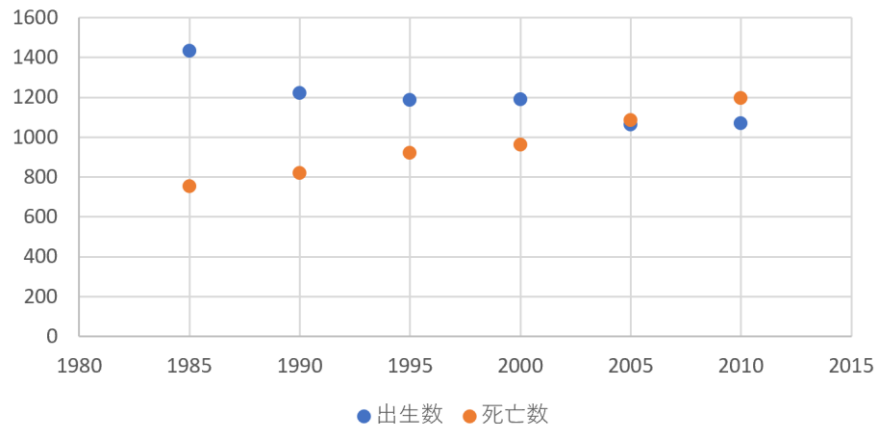
- ここでは、**データ**に対して**散布図**、**ヒストグラム**を用いた分析を説明する
- これらのスキルは、後述の機械学習でも扱うデータの可視化・理解の基盤となる。



# 散布図の用途

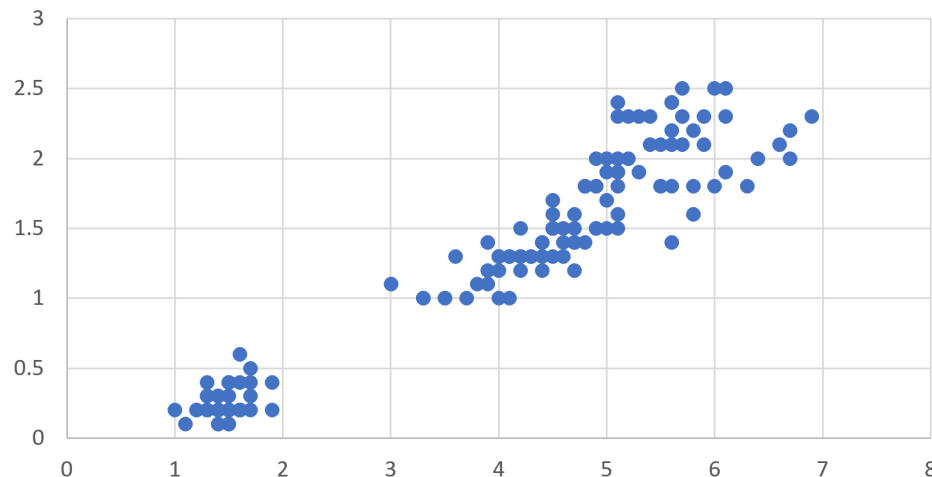


## ・ 時間変化の読み取り



横軸を時間とし、散布図から時間変化を読み取る。  
例：年次ごとの出生数・死亡数の変化。

## ・ 2つの量の関係の把握



横軸と縦軸に2つの量を取り、両者の関係を見る

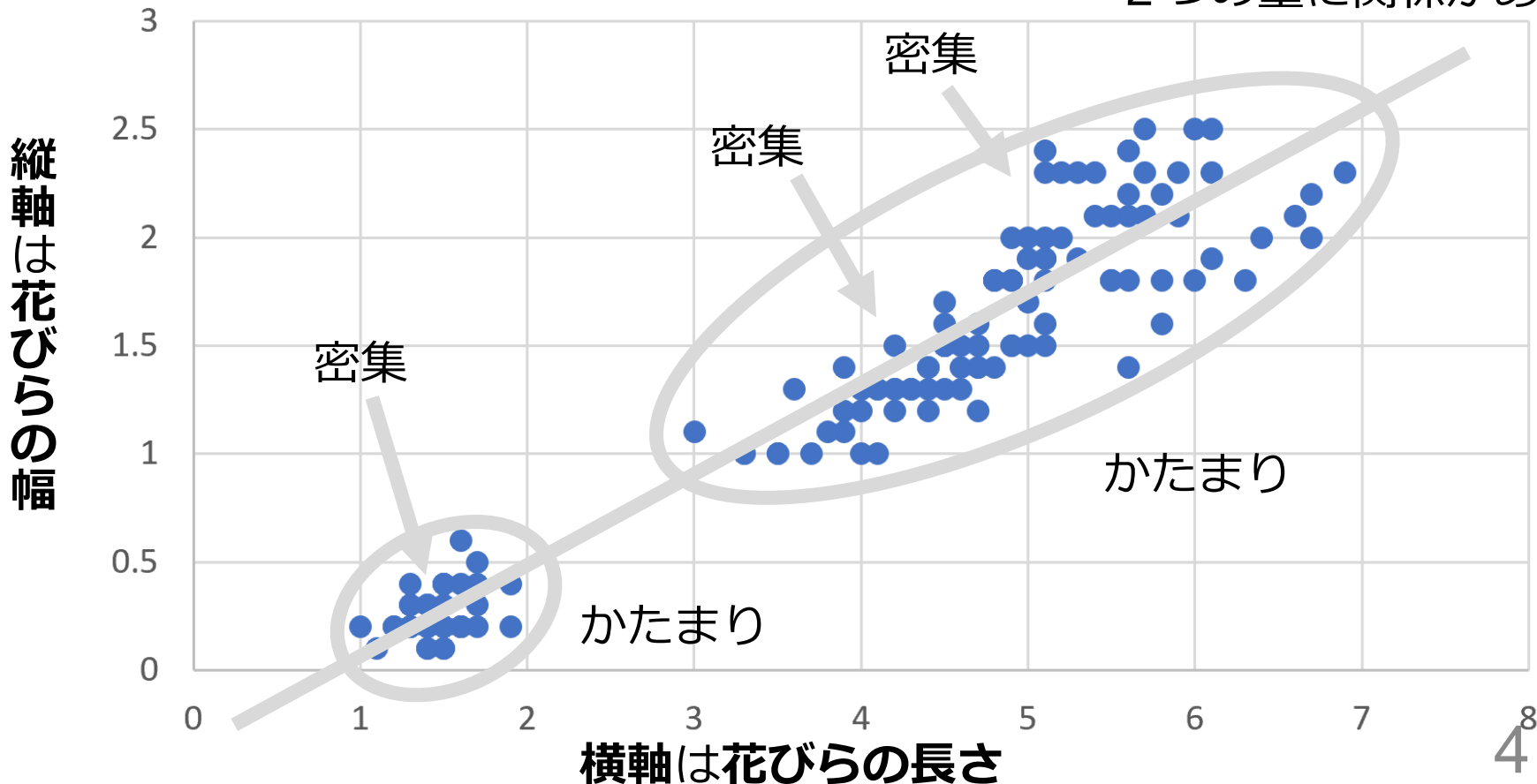
# 分布から読み取れること



密集やかたまりの存在, 2つの量の間の関係 (相関) を読み取れる.

(例) 横軸と縦軸を花びらの長さ と 幅 とすると, **かたまり**が見られ, **種ごとにデータが分かれています** ことが分かる.

2つの量に関係がある



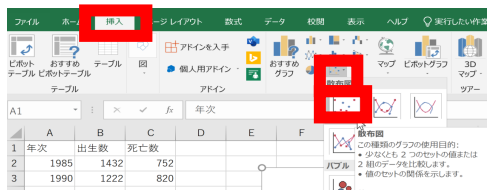
# Excel での散布図の作成手順

|   | 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 |

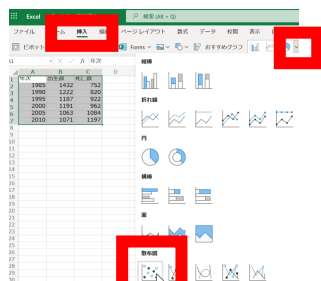
元データ

|   | 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 |      |      |      |   |

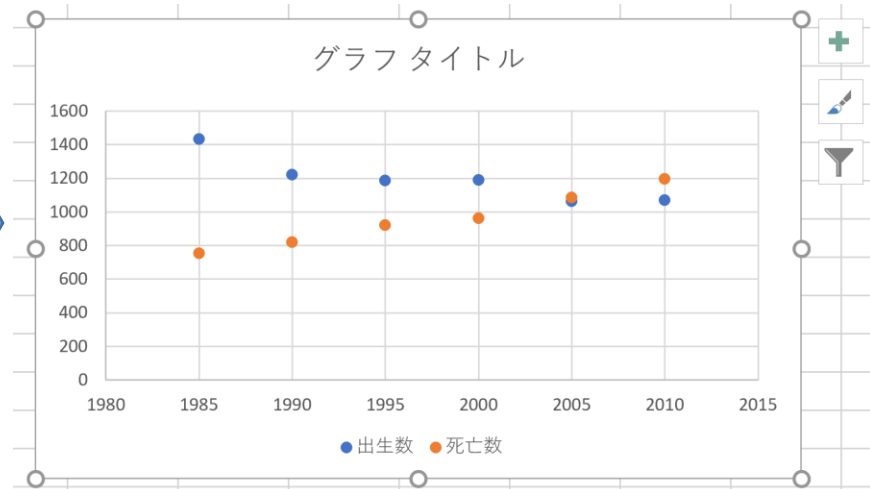
① グラフ化したい部分を範囲選択



アプリ版の Excel



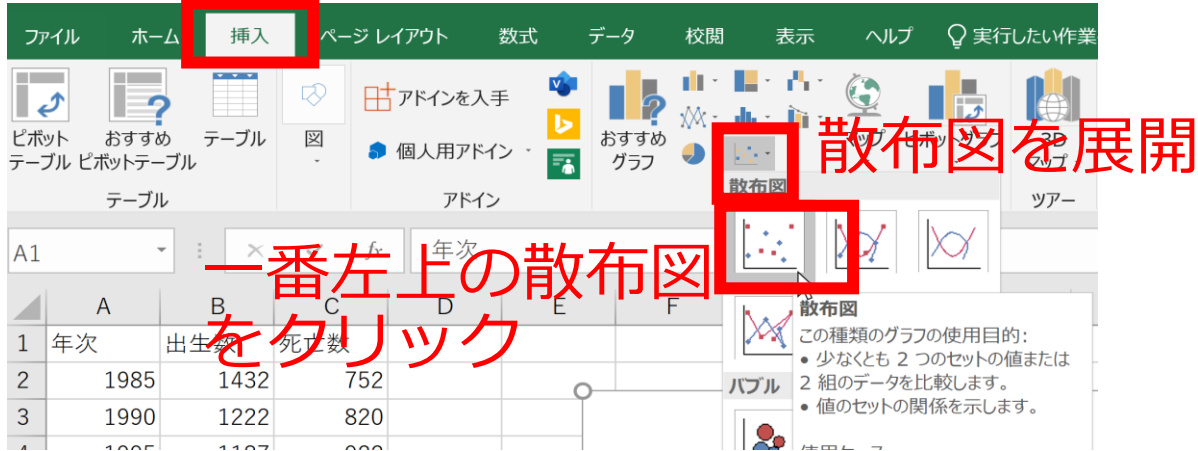
オンライン版の Excel



散布図

② リボンで「挿入」→散布図

# Excel での散布図の種類を選択 挿入

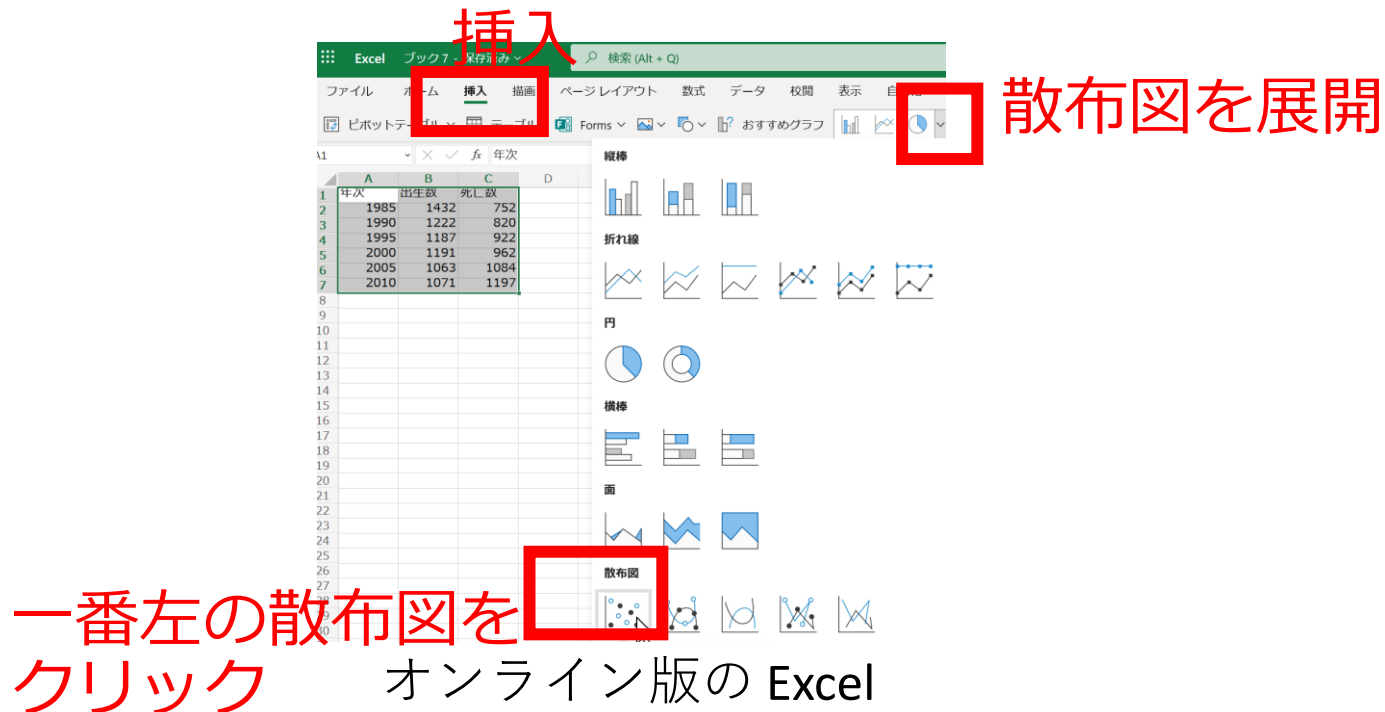


一番左上の散布図をクリック

散布図を展開

| 年次   | 出生数  | 死亡数  |
|------|------|------|
| 1985 | 1432 | 752  |
| 1990 | 1222 | 820  |
| 2005 | 1063 | 1084 |
| 2010 | 1071 | 1197 |

アプリ版の Excel



一番左の散布図をクリック

オンライン版の Excel

散布図を展開

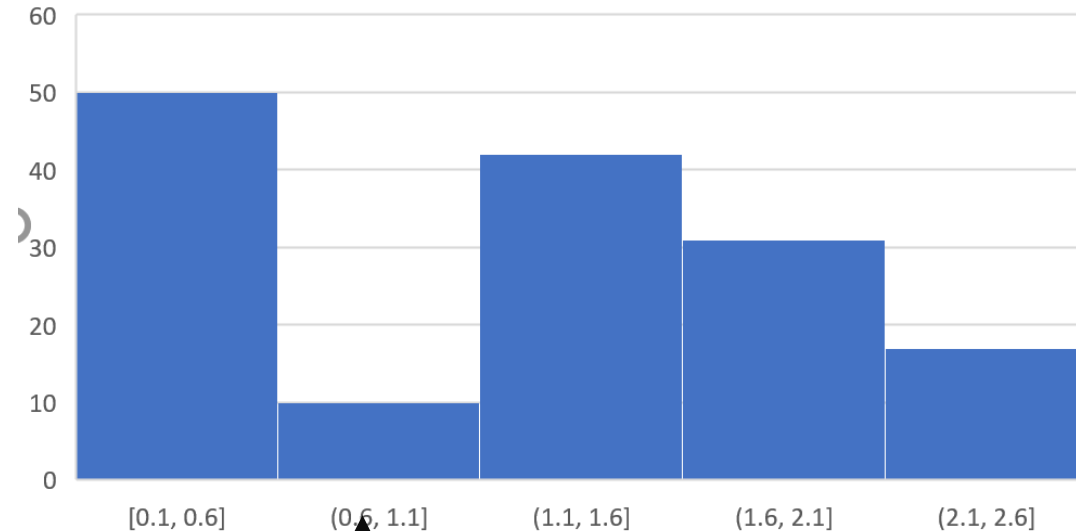
# ヒストグラムの概念



- **ヒストグラム**は、**区間ごとにデータを数え上げたもの**。

- データの分布（散らばり方）を可視化する手法。

（例）区間0.1~0.6, 0.6~1.1, ...のように区切り, 各区間に該当する花びらの数を数え上げる。

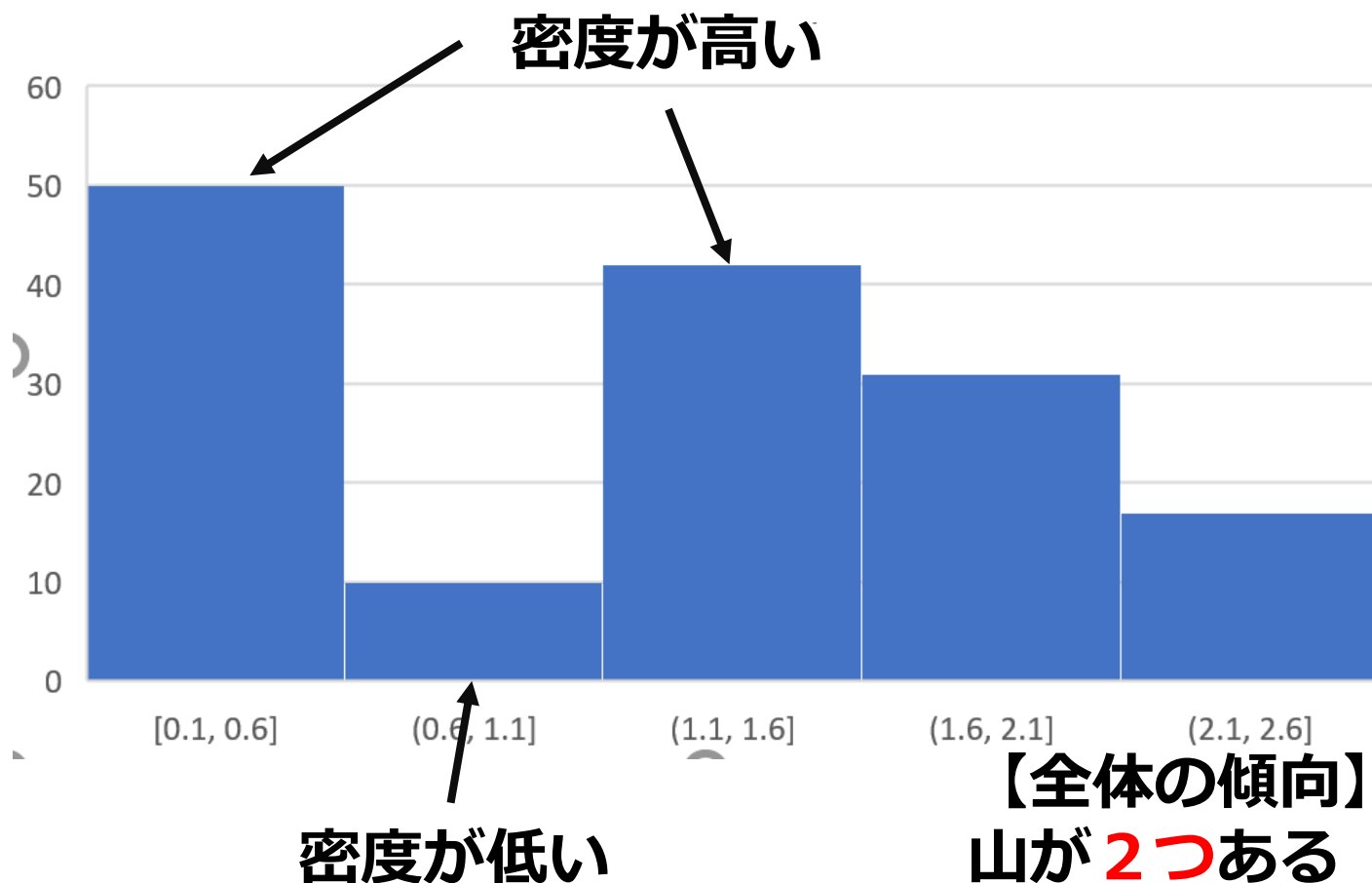


↑  
**区間 0.6 ~ 1.1 の  
データは 10個**

# ヒストグラムから読み取れること



密度の高低を読み取れる。全体の傾向として山が2つある等の特徴が分かる。平均だけでは把握できないデータの偏りや形状を確認できる。



# Excel でのヒストグラムの作成手順

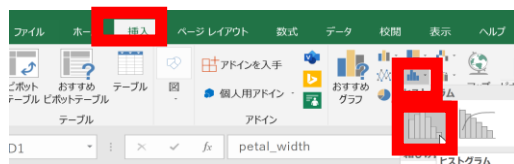


|   | A         | B         | C         | D         | E       |
|---|-----------|-----------|-----------|-----------|---------|
| 1 | sepal_len | sepal_wid | petal_len | petal_wid | species |
| 2 | 5.1       | 3.5       | 1.4       | 0.2       | setosa  |
| 3 | 4.9       | 3         | 1.4       | 0.2       | 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  |

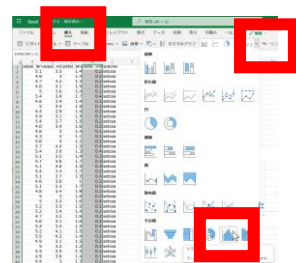
元データ

|    | 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  |

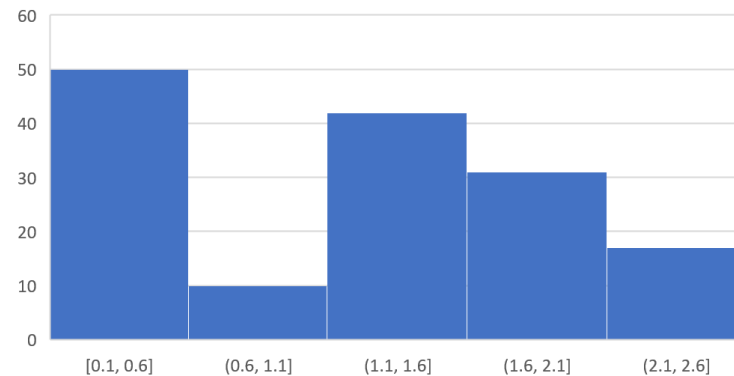
① ヒストグラム化したい列を選択



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

# 研究レポートに書くべき6つの要素



データ分析の結果をレポートにまとめる際にも、データサイエンスのスキルは有用である。

- ① **問題**：背景と目的を明確に説明し，対処する問題を示す
- ② **仮説**：問題を解決するための自分のアイデアや仮説
- ③ **実験手順など**：手順を明確に示す
- ④ **結果**：結果を正確かつ明確に示す．**グラフや表による視覚的な表現や，数値・統計的な情報の提供も有効**
- ⑤ **考察**：研究結果に基づいた分析
- ⑥ **引用文献**：関連する先行研究や参考文献の明示

## 2-6 人工知能と機械学習

- **機械学習もデータサイエンスと同様に「データに基づく」技術である。**
- 今までは、**散布図でデータの分布やかたまり、相関を人間が目で読み取った。**
- **機械学習では、こうしたパターンの発見や分類をコンピュータが自動で行う。**

(身近な例) スマートフォンの顔認証, 動画サイトのおすすめ表示, 音声アシスタント, 自動翻訳, 生成AI (ChatGPT等) ⇒ 機械学習の技術が使われている。

# 機械学習の定義と特徴



**機械学習**は、**コンピュータ**が**データ**を使用して**学習**し**知的能力を向上**させる技術

- **情報の抽出**：データからパターンや関係性を自動で発見する能力
- **簡潔さ**：人間が設定していたルールを自動生成できる
- **限界の超越**：他の方法では困難だった課題にも解決策を得られる可能性がある

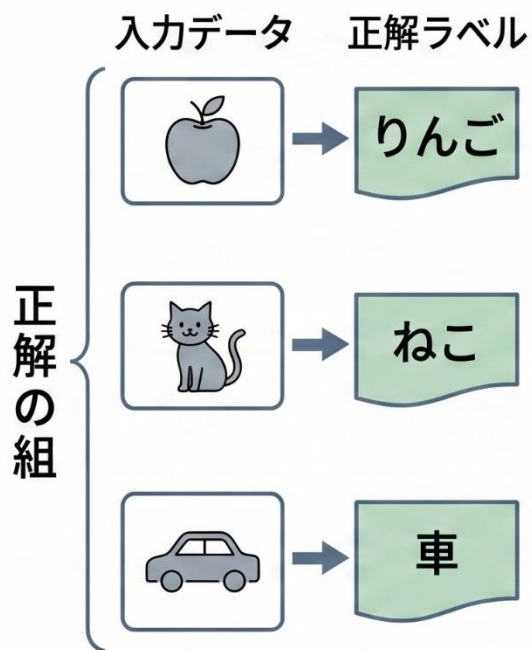


# 教師あり学習の仕組み



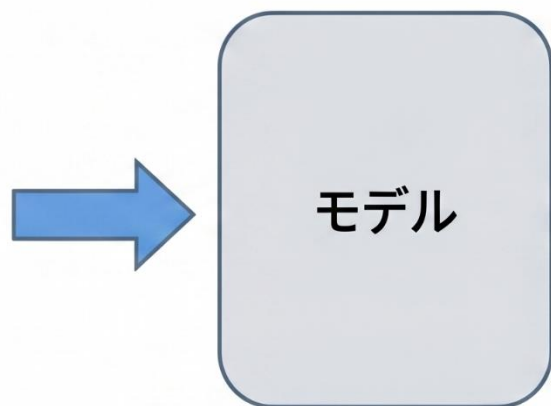
正解の組を用いる方式を 教師あり学習 と呼ぶ。  
(このほかに教師なし学習や強化学習といった方式もある。)

## 1. 大量の訓練データ



## 2. 学習プロセス

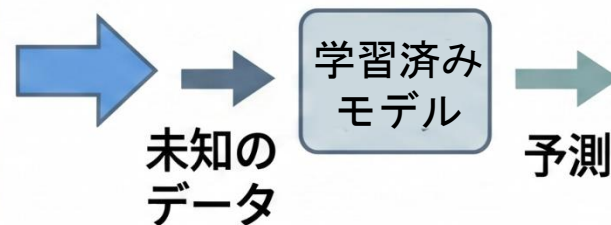
大量の訓練データを用いて学習を行う。



正解の組を用いる方式を教師あり学習と呼ぶ。

## 3. 学習済みモデル

未知のデータから予測が可能になる。



# 教師あり学習の仕組み

入力と正解がペアになったデータを使用

(例) 入力9→正解500, 入力11→正解500, 入力12→正解1000, 入力14→正解1000.

⇒ 学習済みのプログラムが新しいデータに対して自動で結果を出力.

(例) 入力10→予測結果500, 入力13→予測結果1000.



# Palmer Penguins (パーマーペンギン) データセット

## ペンギン・データセットの概要

- 収集: Kristen Gorman博士 (アラスカ大学フェアバンクス校)
- 期間: 2007年-2009年
- 場所: 南極パーマー諸島
- 公開: 2020年 Allison Horst氏ら
- 目的: 教育用パッケージ

## 3種のペンギンの特徴

- Adelie (アデリーペンギン)  
目の周りに白い縁取り、3島すべてに分布
- Chinstrap (ヒゲペンギン)  
あごの下に黒い線、Dream島のみ分布
- Gentoo (ジェンツーペンギン)  
頭頂部に白い帯模様、体が最も大きい

## データの構造

- サンプル数: 344個体
- 観測地点: Torgersen島, Biscoe島, Dream島
- 特徴量: くちばしの長さ, くちばしの奥行き, フリッパーの長さ, 体重, 種, 島, 性別
- 欠損値: 11個体分に存在

## このデータで実施可能な分析

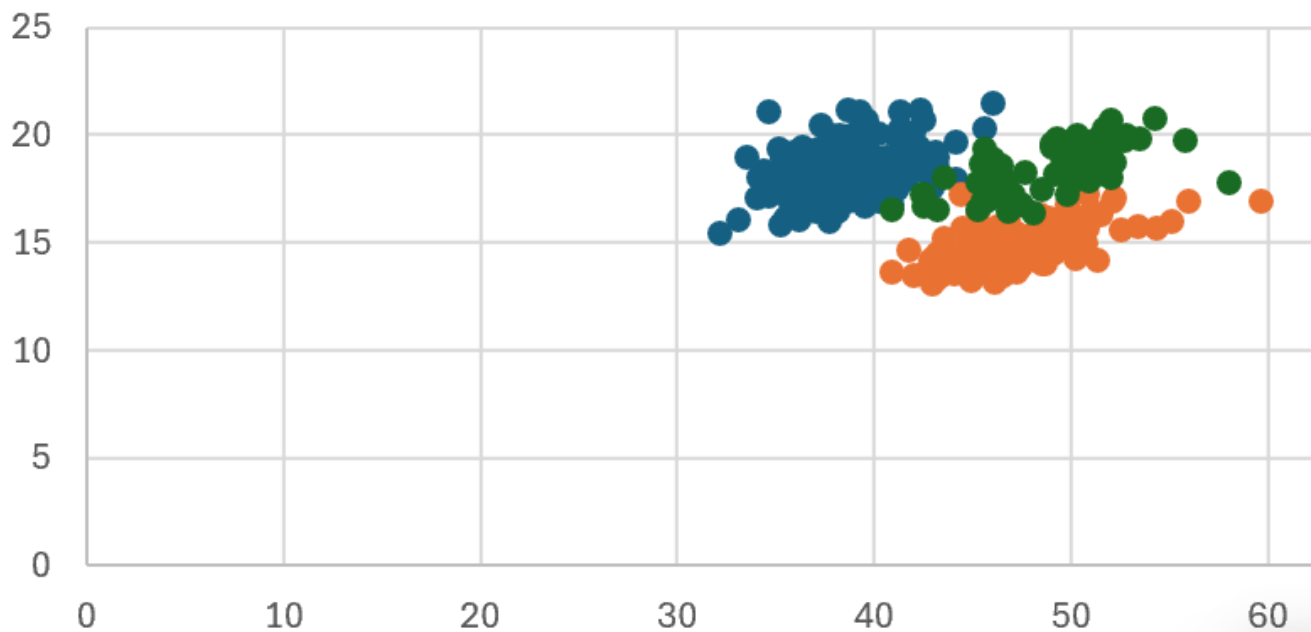
- 探索的データ分析: 分布の確認, 変数間の関係の可視化
- 統計的検定: 種間の体格差の検証
- 分類: 特徴量からペンギンの種を予測するモデルの構築
- クラスタリング: 種のラベルなしでグループを発見する実験

# Palmer Penguins (パーマーペンギン) データセット

## 分布の確認例

横軸：くちばしの長さ  
縦軸：くちばしの深さ

3種類のペンギンを色分け



色を付けた散布図

青 Adelie

オレンジ Gentoo

緑 Chinstrap

# Palmer Penguins (パーマーペンギン) データセット

- テーブル形式のデータ
- Excelやプログラミング (Pythonなど) で取り扱い可能

## (例) 種間の体格差の検証

くちばしの長さ、くちばしの深さ、フリッパーの長さ、体重について、3種のペンギンで差があるか、**平均**を用いて検証

|   | A      | B           | C           | D            | E     | F |
|---|--------|-------------|-------------|--------------|-------|---|
| 1 | 種類     | くちばしの長さ(mm) | くちばしの深さ(mm) | フリッパーの長さ(mm) | 体重(g) |   |
| 2 | Adelie | 39.1        | 18.7        | 181          | 3750  |   |
| 3 | Adelie | 39.5        | 17.4        | 186          | 3800  |   |
| 4 | Adelie | 40.3        | 18          | 195          | 3250  |   |
| 5 | Adelie | 36.7        | 19.3        | 193          | 3450  |   |
| 6 | Adelie | 39.3        | 20.6        | 190          | 3650  |   |

データの一部

|           |             |             |             |             |
|-----------|-------------|-------------|-------------|-------------|
| Adelie    | 38.79139073 | 18.34635762 | 189.9536424 | 3700.662252 |
| Gentoo    | 47.50487805 | 14.98211382 | 217.1869919 | 217.1869919 |
| Chinstrap | 48.83382353 | 18.42058824 | 195.8235294 | 3733.088235 |

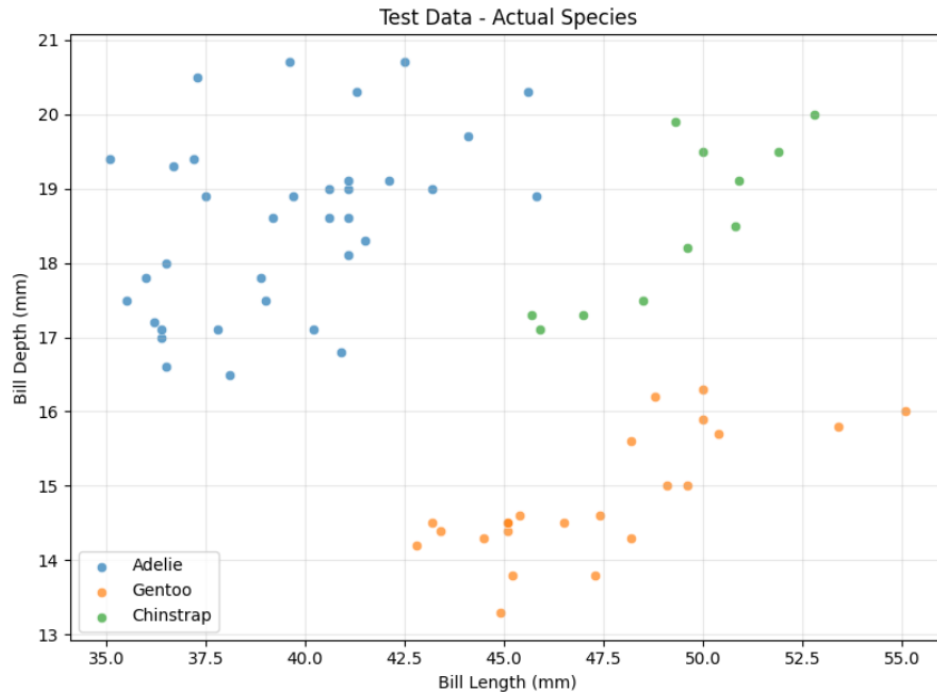
## 3種のペンギンの対格差

左から順に、「くちばしの長さ(mm)」, 「くちばしの深さ(mm)」, 「フリッパーの長さ(mm)」, 「体重(g)」の平均

# Palmer Penguins (パーマーペンギン) データセット

**特徴量** (くちばしの長さ、くちばしの深さ、フリッパーの長さ、体重) から、**ペンギンの種類を予測する機械学習モデル = データに基づく予測**

6. 散布図での比較 (正解 vs 予測)



実データ

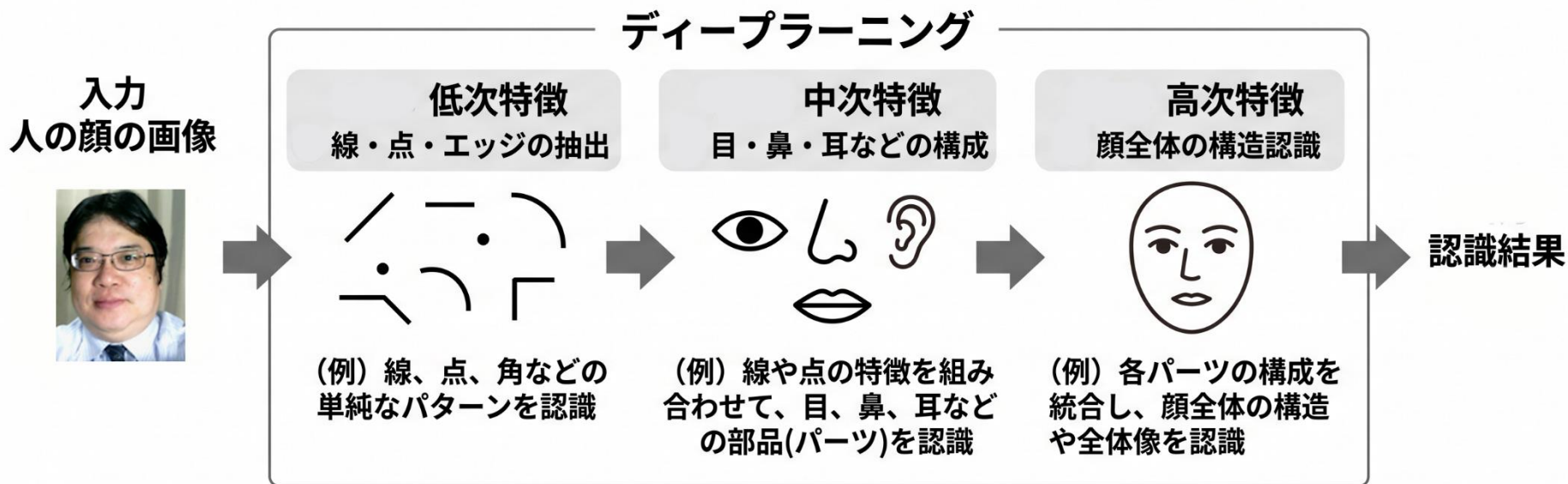


機械学習による予測 (3種を予測)

# ディープラーニングへの期待



- ディープラーニングは、さまざまなレベルのパターンを段階的に抽出できる技術。
- この技術により、複雑なパターンの認識が可能となり、画像認識や自然言語処理の精度が大きく向上した。



(例) 線や点 → 目・鼻・耳 → 顔の構造 と、段階的にパターンを認識

- **データサイエンス**は、**データから有益な情報を導き出すための学問**.
- **機械学習**は、**大量のデータからパターンを抽出し、精度の高いデータ分類**などを可能にする手法。大量の学習用データにより、**パターン抽出やデータ分類の精度が向上する**.
- **いずれもデータに基づく技術**
- **さまざまな分野でデータを活用する実力**につながるため、**これらの技術を理解し実践的な経験を積むことは将来の活躍に役立つ**.

データサイエンスも機械学習も，データを活用して知見や結論を導くものである。

- **結果に100%の正確性は保証されない。** 結果の根拠の確認が大切
- **データの選択・前処理**（分析や学習の前にデータを整形・修正する作業）に**注意が必要**。誤りのあるデータが混入すると結果も誤ったものになる
- データの漏洩や改竄の危険にも注意が必要