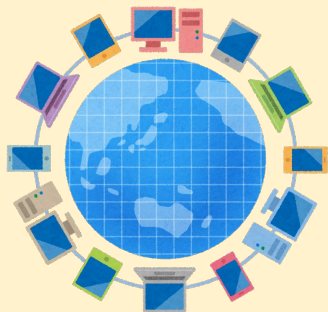


# 【第6回】 データの収集と整理

*Information and Communication Technology*



# 本日の学習目標 & 授業内容

## ★学習目標★

統計データの分析に必要なスプレッドシートの操作を理解する。

1. **【済】** 着席 → ログイン → タイピング練習
2. データの収集と整理
3. スプレッドシート総合演習1

# データを活用する意義

↳ IPv6 (128ビットのIPアドレス) の技術

[1: **IoT** ] (Internet of Things)

膨大なデータのやりとりが可能に



[2: **AI** ] (Artificial Intelligence)

収集された膨大なデータを分析・活用することが容易に

数学や統計学、機械学習などの理論を活用して、莫大なデータの分析や解析を行い、様々なデータから知見や洞察を引き出そうとする研究分野を[3: **データサイエンス** ]という。



# オープンデータの利用

---

国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用できるよう、次のいずれの項目にも該当する形で公開されたデータを

[4: **オープンデータ**]という。

- 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
- 機械判読（プログラムが自動的にデータを再利用できること）に適したもの
- 無償で利用できるもの

# e-Stat と RESAS



## e-Stat (イー・スタット)

各府省等が公表する統計データを一つにまとめ、  
統計データを検索したり、地図上に表示できるなど、  
統計を利用する上で便利な機能を備えた  
政府統計のポータルサイト

## RESAS (リーサス)

Regional Economy Society Analyzing System  
産業構造や人口動態、  
人の流れなどの官民ビッグデータを  
集約し、可視化するシステム



# 数値データの整理

【5: 誤差】 測定値と真の値の間に生じる「ずれ」

## ◆ 精密なデータ

測定値のばらつきが小さい

バリエーション

## ◆ 正確なデータ

測定値の平均と真の値の差が小さい

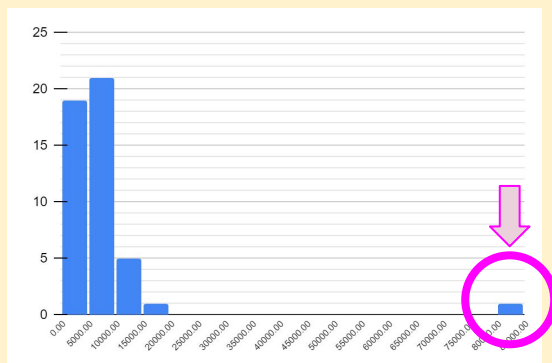
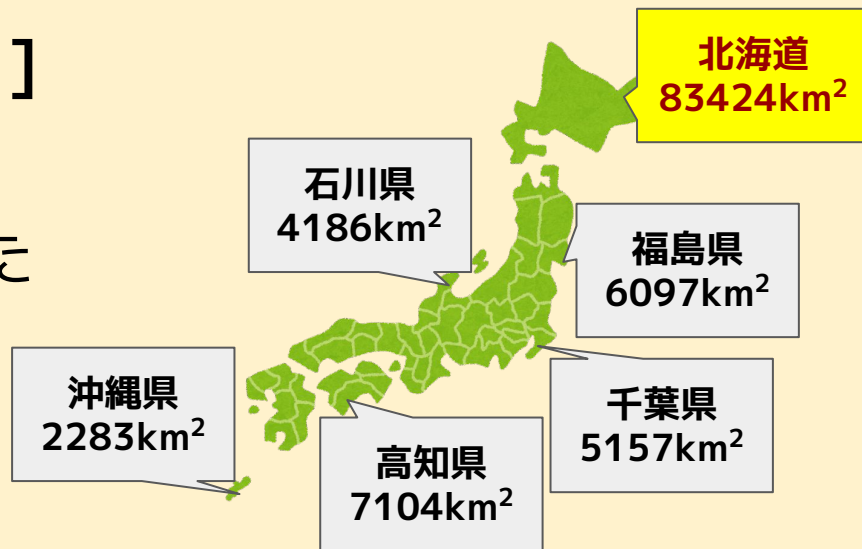
バイアス

		正確さ	
		高い	低い
精密さ	高い		
	低い		

# 数値データの整理

## [6: 外れ値]

他の測定値と  
大きく異なった  
特徴を示す値



## [7: 欠損値]

何らかの原因で値が取得できず、  
変数に値が格納されていないデータ

	性別	年齢	身長	体重
0	男	24.0	171	72.0
1	女	53.0	152	50.0
2	女	48.0	163	64.0
3	男	19.0	184	NaN
4	男	NaN	177	80.0
5	女	61.0	149	46.0
6	女	52.0	158	56.0
7	男	41.0	160	NaN
8	男	29.0	170	74.0
9	女	NaN	151	53.0
10	男	48.0	177	77.0

# おもな統計量

データ  $x_1, x_2, \dots, x_n$  に対して、

平均

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

分散

偏差平方(平均との差の2乗)の平均

$$s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

標準偏差

分散の正の平方根

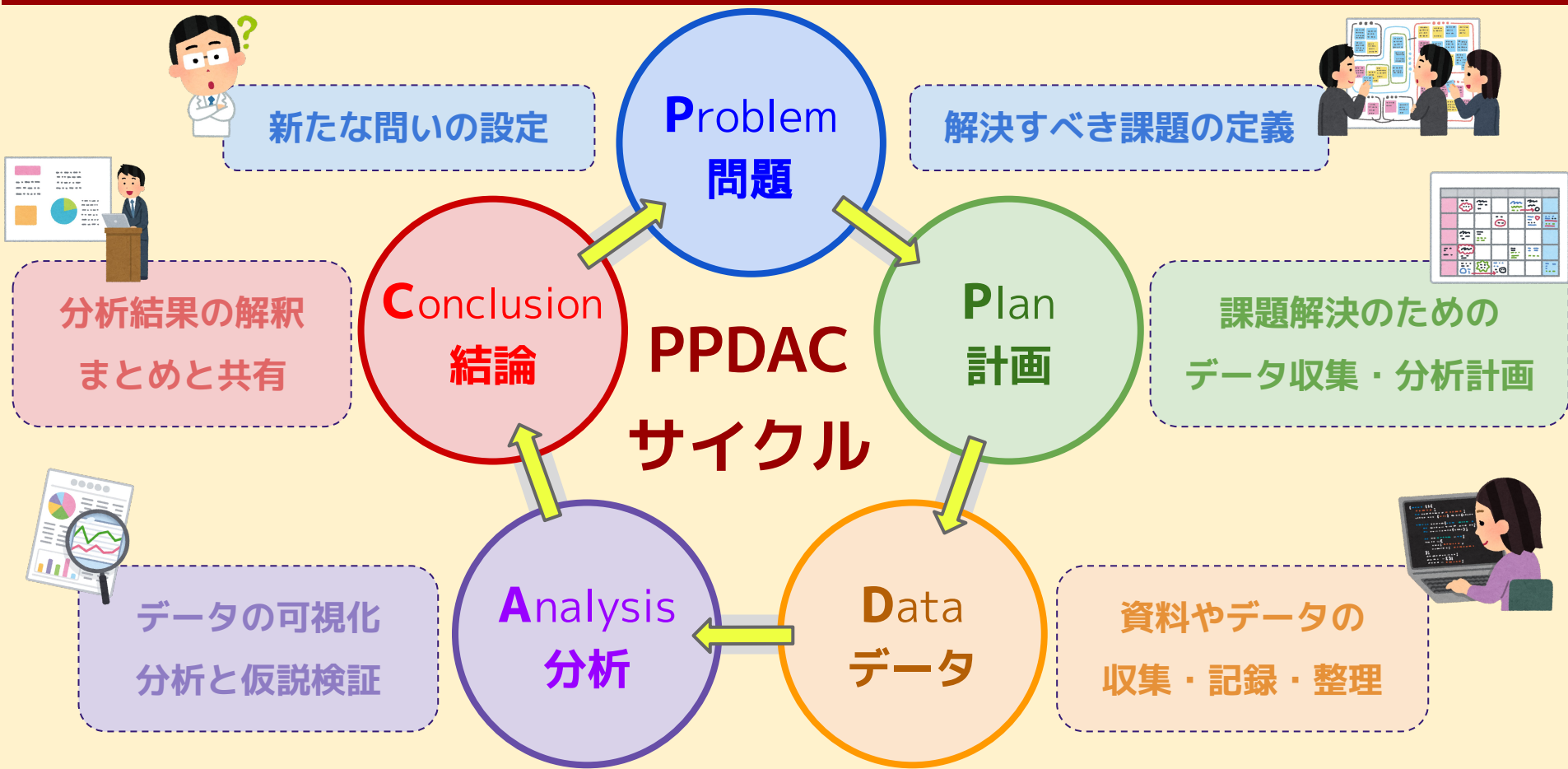
$$s_x = \sqrt{s_x^2}$$



# 尺度水準

	尺度水準	例	順序性	等間隔	正比例
質的 データ	[8: 名義 ]尺度	血液型 氏名 郵便番号	×	×	×
	[9: 順序 ]尺度	10段階評定 服のサイズ 書道の段級	○	×	×
量的 データ	[10: 間隔 ]尺度	西暦 温度(摂氏) 知能指数	○	○	×
	[11: 比率 ]尺度	距離 年収 質量	○	○	○

# データを活用した問題解決のフレームワーク



2024年度 SSI（情報） 全クラス共通

# 【第6回・課題】 スプレッドシート演習 入門編

*Information and Communication Technology*



# スプレッドシート演習 入門編①

Classroom【第6回・課題】スプレッドシート演習 入門編

[基本の演算]シート



列番号

表計算ソフト

行番号

ワークシート

アクティブセル

	A	B	C
1			
2			
3			
4			
5			

	A	B	C	D	E	F
1	【基本の演算】					
2	和	8	+	3	=	<div>11 × =B2+D2</div>

セル

セル番地：B2

★大原則★

数式は半角英数で入力し、  
先頭に「=」をつける。

=B2+D2

# スプレッドシート演習 入門編②

## Classroom【第6回・課題】スプレッドシート演習 入門編

### [関数入門]シート

	A	B	C	D	E	F	G	H	I
1	組	番号	氏名	国語	数学	英語	合計	順位	合否判定
2	9	1	アオキ ヤスヒコ	93	31	29			
3	9	2	イナバ サダジ	83	58	100			
4	9	3	ウチダ ヨシフミ	88	79	75			
5	9	4	オカダ ユキオ	65	13	33			
6	9	5	オグラ ヤスゾウ	84	25	94			
7	9	6	カワイ カツヒロ	95	42	37			

演習内容	
1	セルG2に9組1番の生徒の国語・数学・英語の合計点を出力する関数を入れましょう。
2	セルG2に入れた数式をセルG3～G21にコピーしましょう。
3	セルD22にこのクラスの国語の平均点を算出する関数を入れましょう。
4	セルD22に入れた数式をセルE22～G22にコピーしましょう。
5	セルD23にこのクラスの国語の標準偏差を算出する関数を入れましょう。
6	セルD23に入れた数式をセルE23～G23にコピーしましょう。
7	セルD24にこのクラスの国語の最高点を算出する関数を入れましょう。
8	セルD24に入れた数式をセルE24～G24にコピーしましょう。
9	セルD25にこのクラスの国語の最低点を算出する関数を入れましょう。
10	セルD25に入れた数式をセルE25～G25にコピーしましょう。
11	セルH2に9組1番の生徒の合計点昇順の順位を出力する関数を入れましょう。
12	セルH2に入れた数式をセルH3～H21にコピーしましょう。
13	セルH2の数式の範囲を「相対参照」から「絶対参照」に変えましょう。
14	セルH2に入れた数式をセルH3～H21にコピーしましょう。

※以降「中級者向け」です。ここまで問題なくできた人は挑戦してみましょう。

15	セルI2に出席番号1番の生徒の合否判定を出力する関数を入れます。 180点以上ならば「合格」、179点以下ならば「不合格」が出力されるようにします。
16	セルI2に入れた数式をセルI3～I21にコピーしましょう。
17	セルI23に合格した人の人数を出力する関数を入れましょう。 セルI2～I21の範囲の合格の人数をカウントする関数を入れます。