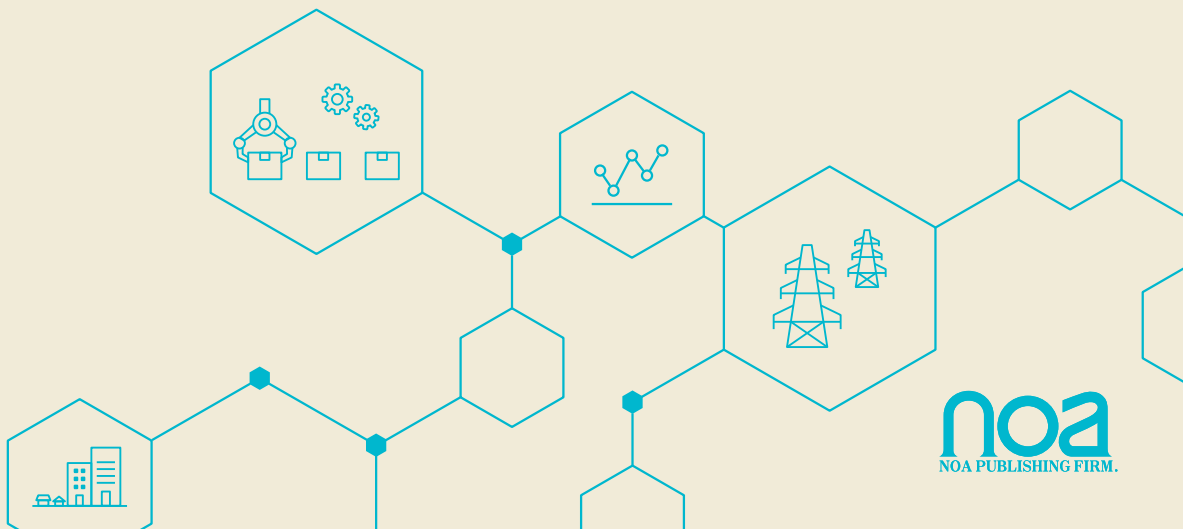


はじめの第一歩

基礎から  
はじめる データ  
サイエンス

編：保本 正芳

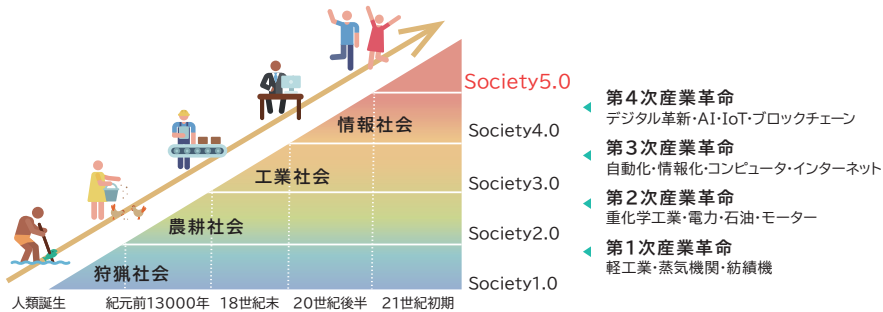


# STEP 2 これからの社会

## Lesson 1 Society5.0

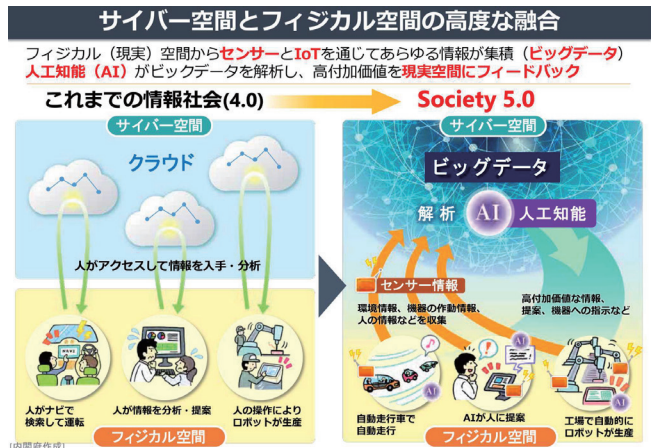
### 1 Society5.0

人類は狩猟社会から農耕社会、工業社会を経て、現代の情報社会へと変化・発展してきました。情報社会の次に日本が目指す新たな社会として、内閣府は2016年の第5期科学技術基本計画において「**Society5.0**」を提唱しました。最先端技術を取り入れ、誰もが快適で活力に満ちた質の高い生活を送ることのできる人間中心の社会が、これからの私たちの社会とされています。



社会の変化 (日本経済団体連合会 Web ページ) より作成

情報社会では、人が仮想空間に存在するクラウドサービスにインターネット経由でアクセスして情報やデータを入手し、分析を行なってきました。Society5.0ではIoT、AI、ビッグデータ、ロボットの4つの技術が私たちの生活に浸透し、サイバー空間とフィジカル空間が融合した「超スマート社会」とも言われています。Society5.0では、よりよい社会を築くために、経済発展と社会的課題の解決を両立する人間中心の社会を目指します。誰もが、いつでもどこでも安心して、自然と共生しながら、価値を生み出す社会を私たち1人1人が実現していくことで新しい未来社会が作り出されていくのです。最新の技術を用いた技術革新のことは「第4次産業革命」と呼ばれ、国際的にも関心が高まっています。Society5.0と第4次産業革命の実現には、IoTやAI、ビッグデータ、ロボットに精通する人材育成が必要です。



内閣府が示している Society5.0 の仕組み (https://www8.cao.go.jp/cstp/society5\_0/society5\_0.pdf) より引用

### 3 「ばらつき」をグラフから確認しよう

価格帯ごとの購入者数のばらつきを調べてみましょう。

#### 価格帯を決める

まずは価格帯を決めます。購入金額のデータの個数と最大値、最小値、データの幅を確認しましょう。

最小値と最大値を含み、データの幅の17,905を適度に区分する価格帯を設定します。今回は**0～19,999円の範囲**で、**2,000円区切りの10段階**に区分しましょう。この価格帯のようにデータを定数で区分したものを階級と言います。

データの個数	100
最大値	18,802
最小値	897
データの幅	17,905


データの個数 p94 ▶

最大値 p95 ▶

最小値 p95 ▶

#### 価格帯ごとの購入者数を調べる

次に、各価格帯の購入者関数を使って求めます。各級数の個数を度数と言います。

- ① J11からJ20を範囲指定して、 (関数の挿入)をクリックする
- ② 【関数の挿入】ダイアログボックスが表示されるので、[関数の分類(C)] から“統計”、[関数名(N)] から“FREQUENCY”を選び、[OK]をクリックする
- ③ 【関数の引数】ダイアログボックスが表示されるので、[データ配列] に「D3:D102」、[区間配列] に「H11:H20」を指定する
- ④ [Ctrl] キー + [Shift] キーを押しながら、[OK]をクリックする
- ⑤ 購入人数の列に値が挿入される

J11      fx      =FREQUENCY(D3:D102,H11:H20)

	A	B	C	D	E	F	G	H	I	J	K
8	C01006	28	男性	2,463							
9	C01007	45	女性	7,489							
10	C01008	39	女性	9,404							
11	C01009	38	男性	10,750							
12	C01010	26	女性	3,688							
13	C01011	25	女性	2,963							
14	C01012	22	女性	9,550							
15	C01013	32	女性	8,595							
16	C01014	37	女性	12,738							
17	C01015	34	女性	9,701							
18	C01016	35	女性	8,525							
19	C01017	36	女性	7,341							
20	C01018	40	女性	8,051							
21	C01019	29	女性	3,181							

価格帯ごとの購入人数	以上	以下	価格帯 (円)	購入人数	
	0	～	1,999	0～1,999	9
	2,000	～	3,999	2,000～3,999	19
	4,000	～	5,999	4,000～5,999	9
	6,000	～	7,999	6,000～7,999	9
	8,000	～	9,999	8,000～9,999	31
	10,000	～	11,999	10,000～11,999	11
	12,000	～	13,999	12,000～13,999	4
	14,000	～	15,999	14,000～15,999	3
	16,000	～	17,999	16,000～17,999	3
	18,000	～	19,999	18,000～19,999	2

## One Point

ピボットテーブルで分析すると

- ① [購入金額(円)] フィールドを [行] に追加
- ② [伝票ID] フィールドを [値] に追加

※購入金額のグループ値を変更するには、行フィールドの値を右クリックして [グループ化(G)] を選択して、先頭の値、末尾の値、単位を変更する

	A	B
1		
2		
3	行ラベル	値 / 伝票ID
4	0-1999	9
5	2000-3999	19
6	4000-5999	9
7	6000-7999	9
8	8000-9999	31
9	10000-11999	11
10	12000-13999	4
11	14000-15999	3
12	16000-17999	3
13	18000-19999	2
14	総計	100

グループ化 ? ×

自動

先頭の値(S): 0

末尾の値(E): 19999

単位(B): 2000

OK      キャンセル

## 度数分布表

ヒストグラムを作成する場合には、いくつかの値の範囲を設定し、それぞれの範囲に収まるデータの個数を数えます。このとき設定した値の範囲を**階級**と呼び、数えたデータの個数のことを**度数**と呼びます。

この階級ごとの度数をまとめた表のことを**度数分布表**と言い、この表からヒストグラムを作成すると、データがどのように分布しているかを一目でわかるようになります。

度数分布表の例

階級	度数
～ 19	1
20 ～ 39	3
40 ～ 59	5
60 ～ 79	7
80 ～ 99	12
100 ～ 119	10
120 ～ 139	9
140 ～ 159	6
160 ～ 179	3
180 ～	0

データを区切る  
値の範囲

階級の範囲に含ま  
れるデータの数

Excel

分布

=FREQUENCY( データ配列 , 区間配列 )

※データ配列：調査の対象データを指定

※区間配列：度数分布の基準となる数値を指定

※引数を指定し、[Ctrl] キーと [Shift] キーを押しながら [OK] ボタンをクリックする

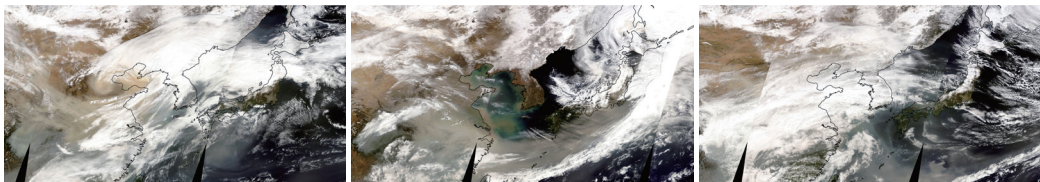
### 練習問題 44

- ①ブック『ばらつき\_01\_度数分布表』を開く
- ②H4からH9に得点帯ごとの人数を求める  
※平均点以下がクラスの半数を占めることを確認しましょう
- ③ファイル名『ばらつき\_01\_度数分布表完成』で保存し、閉じる

## Column 外れ値は貴重なデータ

黄砂は東アジア内陸部のタクラマカン砂漠やゴビ砂漠または乾燥地域の砂塵が、風によって上空に巻き上げられ、偏西風に乗って日本に飛来し、大気中に浮遊あるいは地上に降り注ぐ現象です。2010年3月20日に大規模黄砂が発生し、日本には2010年3月20日～21日にかけて飛来し、日本各地で高濃度の浮遊粒子状物質 (SPM) が観測されました。

下図は、2010年3月20日～21日までの米国の衛星 Terra/MODIS から見た東アジア域の様子です。21日には大規模黄砂が日本に飛来している様子が茶色の帯で確認できます。



2010年3月20日

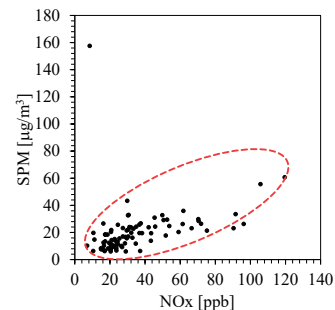
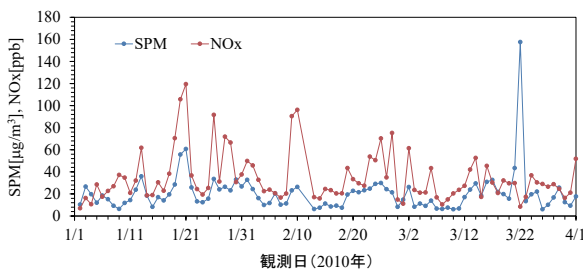
2010年3月21日

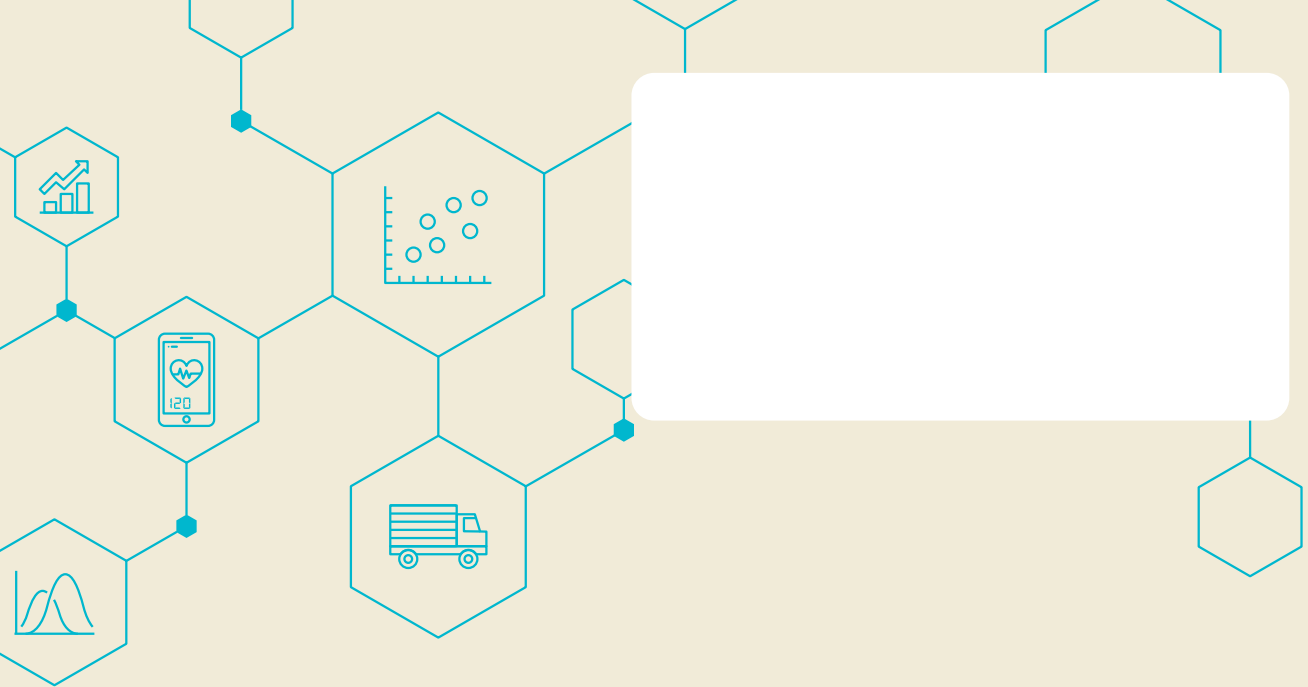
2010年3月22日

また、2010年1月1日～4月1日に大阪市（国設大阪）で取得された SPM(単位:  $\mu\text{g}/\text{m}^3$ )、窒素酸化物 (NO<sub>x</sub>, 単位: ppb) の日平均値を下の図で示します。左の日変化では、20、21日にかけて、SPM 濃度値が急上昇し、NO<sub>x</sub> は変化していないのが確認できます。

NO<sub>x</sub> は、工場や自動車、家庭等からガス状物質として放出されたもので一酸化窒素、二酸化窒素が主となっています。SPM は粒径が  $10\mu\text{m}$  ( $0.01\text{mm}$ ) 以下の大気中に浮遊する粒子状物質と定義されており、黄砂や海塩粒子等の自然起源の粒子や工場や自動車から排出される人為起源の粒子、NO<sub>x</sub> 等のガス状物質が化学的に変化して2次的に生成される粒子があります。3月20日、21日の SPM 濃度値の上昇は、国内で発生した汚染物質でなく、中国から飛来した黄砂による影響とわかります。

右の図は、NO<sub>x</sub> と SPM の散布図を示しています。赤の破線の部分は、NO<sub>x</sub> と SPM に相関が見られますが、1つのデータ (21日のデータ) が大きく外れていることがわかります。この外れ値があると相関係数に強い影響が及ぶため、外れ値は分析対象から除く場合がありますが、ここで示した外れ値は中国から飛来した黄砂粒子を分析するには非常に価値があるデータとも言えます。





**noa**  
NOA PUBLISHING FIRM.

