

度数分布表とヒストグラム

1) データ入力

【 課題 5 】 下の表1は全国の体育館の数を都道府県別にまとめたものである。この表を下の作成見本のように入力せよ。

表1 都道府県別体育館数(2002.10.1現在)

北海道	406	埼玉	165	岐阜	184	鳥取	114	佐賀	60
青森	103	千葉	127	静岡	115	島根	89	長崎	108
岩手	178	東京	190	愛知	194	岡山	93	熊本	188
宮城	134	神奈川	221	三重	97	広島	143	大分	69
秋田	159	新潟	250	滋賀	91	山口	83	宮崎	136
山形	99	富山	112	京都	82	徳島	69	鹿児島	154
福島	242	石川	131	大阪	173	香川	77	沖縄	53
茨城	133	福井	86	兵庫	203	愛媛	84		
栃木	91	山梨	90	奈良	124	高知	55		
群馬	126	長野	270	和歌山	53	福岡	187		

< 作成見本 >

	A	B	C	D
1	都道府県別体育館数 (2002.10.1現在)			
2				
3	都道府県	体育館数		
4	北海道	406		
5	青森	103		
6	岩手	178		
7	宮城	134		
8	秋田	159		
9	山形	99		
10	福島	242		
11	茨城	133		
12	栃木	91		
13	群馬	126		
14	埼玉	165		
15	千葉	127		
16	東京	190		
17	神奈川	221		

## 1. 連続データの作成

マウスのドラッグによってセルに入力されているデータを複写することができる。1, 2, 3 ... のような数値データと、月,火,水... のような連続性を持つ特定の文字列データは複写後変化することがある。下表の列(タテ)方向への複写を例として説明する。

	A	B	C	D	E	F	G	H
1	1	1	1	1	第1位	月	Sun	子
2	2	2	1	2	第2位	火	Mon	丑
3	3	1	1	3	第3位	水	Tue	寅
4	4	2	1	4	第4位	木	Wed	卯
5	5	1	1	5	第5位	金	Thu	辰

### (1) 数値データの場合

📖 手順 セル A1 および A2 に 1 と 2 が入力されているものとする。このとき、マウスで A1 から A2 をドラッグして範囲指定をおこなう。左ボタンを離した後、カーソルを反転している長方形の右下隅に移動すると、+ の形状のカーソルに変化する。この状態でボタンを押して下方向にドラッグし、必要なところで左ボタンを開放すれば 3, 4, 5 ... となる。(A 列のようになる)

この手順において、左ボタンを開放する前に **Ctrl** キーを押すとマウスポインタの形状が ++ になる。**Ctrl** キーを押したまま左ボタンを開放すると上の B 列のようになるはずである。

データがひとつしかない場合には機能が逆になる。C 列と D 列を参照せよ。また、前後に文字列があってもよい。E 列がその例である。

### (2) 文字データの場合

メニューから ツール(T) - オプション(O) を選択し、ユーザ設定リストのタグをクリックする。このとき、ユーザ設定リストのウィンドウに表示されている文字列が複写を実行した際に変化するものである。

たとえば、日月火水木金土 という並びがリストに表示されている。この中の 1 文字が入力されているセルを指定して複写すると、この並びにしたがって変化しながらコピーされる。7 文字以上複写先を指定したときには 日月火水木金土日月火... などと反復する。

前ページの表の F、G、H 列にいくつかの例を示しておく。

なお、任意の文字列の並びをユーザ設定リストに加えることができる。都道府県名などはユーザ設定リストに加えておくと便利である。

- 【課題 6】 都道府県名をユーザ設定リストに加えよ。手順は次の通りである。
- メニューから ツール(T) - オプション(O) を選択し、ユーザ設定リストのタグをクリックする。リストの取り込み元範囲の一番右のボタンをクリックし、北海道から沖縄までをドラッグし、右側のボタンを押す。
- リストの取り込み元範囲が **\$A\$4:\$A\$50** となっていることを確認し、インポートボタンを押す。

以上で都道府県名が連続データとして設定された。

## 2) 1 変量データの特性値

統計データは多くの情報を含んでいる。分析者はデータが持つ情報をもとに何らかの判断を行なうのである。しかし、100 個や 200 個のデータを見て、何らかの判断をすることは容易ではない。そこで、データが持つ情報をまとめる必要がある。データ全体の情報を数値でまとめたものが**特性値**であり、視覚的にまとめたものが次節で説明する**度数分布表とヒストグラム**である。

1変量データの分布の傾向を示す特性値には

中心的傾向 ... 算術平均、メディアン（中央値）、モード（最頻値）

バラツキの尺度 ... 分散、標準偏差、レンジ（範囲）、四分位偏差

などが挙げられる。

### 1. 関数SUM

Excel には1変量データの特性値を容易に計算する関数が用意されているので、結果を表示したいセルに関数を用いた式を書き込めば特性値が求まる。ここではまず合計を求める関数SUMを用いてみる。

【課題 7】都道府県別体育館数のデータについて、全国の合計を求めよ。具体的には、B4 から B50 までの 47 個のセルの値の合計を B52 に計算するものとする。

#### 📖 手順

セルB52に `=SUM(` と入力する。SUMは小文字でもよい。なお、日本語入力モードはオフにしておくこと。

マウスポインタをB4におきクリックしたままB50までドラッグする。B4からB50までの連続47個のセルが反転表示される。

マウスを離して、キーボードから `)` を入力する。B52のセルには `=SUM(B4:B50)` と表示される。Enterキーを押すと合計が表示される。

\* なお、マウスでB4:B50の範囲を指定する代わりに、キーボードから入力してもよい。“B4:B50”は「B4からB50までの連続した範囲」という意味である。

### 2. 名前の定義と利用方法

【課題 7】では、関数 SUM を用いて合計を求める場合に、`=SUM(B4:B50)` のように、合計を求める範囲 B4:B50 を関数の引数として指定したが、あらかじめ範囲 B4:B50 に名前（ここでは gym とする）を付けることによって、`=SUM(gym)` とすることができる。

名前には文字、数字、`_`（アンダースコア）などを組合せてつけることができるが、アルファベットの大文字と小文字の区別はない。DATA, Data, data はいずれも同一の名前となる。なお、A5 や X1 などのようなセル番地と混同する命名は避けるのが賢明である。また、数字から始まる名前もつけることができない<sup>1</sup>。

📖 範囲は行や列の1部だけでなく、箱型に指定することもできる。その場合は、左上:右下とする。たとえば、右表で A1 から D3 までの12個のセルは **A1:D3** という形で指定すればよい。

	A	B	C	D
1				
2				

<sup>1</sup> 同一の名前を2度以上定義することはできない。また、名前はブック全体に共通なものであるから、シートがちがっても同じ名前を付けることはできない。

【 課題 8 】 都道府県別体育館数のデータに gym という名前をつけよ。

実際の操作は次の 2 通りのやり方のうち、いずれかの方法を用いれば良い。

- (1) 範囲 B4:B50 を選択する。メニューバーから 挿入-名前-定義 を選択すると、名前を入力するボックスが現われるので、 gym と入力する。
- (2) 範囲 B4:B50 を選択。名前ボックス(通常は列名を表す A B の上方にある)の ↓ をクリックして gym を入力する。なお、すでに 1 個以上の名前が定義されているときには、名前ボックスの ↓ をクリックすると、下方に名前の一覧表が出現する。この一覧表の中の名前をクリックするとその名前が指している範囲を知ることができる。

### 3. 統計関数の利用

関数の一般的な形をあらわすと以下ようになる。

**一般形： =関数名(引数 1, 引数 2, 引数 3, ... 引数 k)**

引数の数は関数によって異なり、0 個のものもあるが、その場合でも( )は必要である。

例 1：関数 AVERAGE はこれまでのように、引数に範囲または名前をとる。あるいは、 のように数値を直接書き込むこともできる。

=AVERAGE(A1:A20)	引数の数は 1 個
=AVERAGE(X1)	引数の数は 1 個
=AVERAGE(5, 3, 6, 8, 9, 5, 8, 9)	引数の数は 8 個

AVERAGE( ) と同様の引数をとる統計関数に、MAX( ), MIN( ), COUNT( ), MODE( ), MEDIAN( ), STDEVP( ), VARP( ) などがある。

【 課題 9 】 表 1 について、データ数(COUNT)、算術平均(AVERAGE)、メディアン(MEDIAN)、モード(MODE)、分散(VARP)、標準偏差(STDEVP)、最大値(MAX)、最小値(MIN) を求めよ。

例 2：関数 QUARTILE(引数 1, 引数 2) は 4 分位点を求める関数である。引数 1 は範囲、引数 2 は 0 から 4 までの数値をとり、以下に示すようなデータを戻り値として与える。

0	データの最小値
1	下位 4 分の 1 (25%) に相当するデータ
2	データの中央値 (50%)
3	上位 4 分の 1 (75%) に相当するデータ
4	データの最大値

第 2 引数 に 0, 2, 4 のいずれかの数値を指定すると、QUARTILE 関数の戻り値は、それぞれ MIN 関数, MEDIAN 関数, MAX 関数の戻り値に等しくなる。

【 課題 10 】 表 1 についてレンジ、四分位偏差を求めよ。なお、レンジは関数 MAX と関数 MIN、四分位偏差は四分位点を求める関数 QUARTILE を利用すればよい。

### 3) 度数分布とヒストグラム

本節では、1変量データの持つ情報を視覚的にまとめる1つの方法である、度数分布表とヒストグラムを作成してみる。度数分布表は適当な大きさの区間（階級という）を作り、その区間に含まれるデータの数を表の形にまとめたものである。ヒストグラムはこの度数分布表をグラフにしたものである。

#### 1. 度数分布表

表 1 から、度数分布表を作成してみよう。度数分布表を作成するときには、関数 **FREQUENCY** を用いる。

まず、表 1 がブック gym.xls の Sheet1 に入力され、体育館数を表す 47 個のデータには範囲名 gym が設定されているものとする。

度数分布表の作成には階級が必要となるが、Excel では、個々の階級を下限と上限に分けて入力しておくことと便利である。ここでは、右図のように階級の下限と上限を - で結んだ。列 F, G, H はそれぞれ列幅を調整してある。階級の上限 (H17:H25) に範囲名 **class** を付けておく。列 I には、(下限 + 上限) / 2 とし、階級値を求めておく。

	F	G	H	I	J
16	階		級	階級値	度数
17	0	-	49	=(f17+h17)/2	=frequency(gym,class)
18	50	-	99		
19	100	-	149		
20	150	-	199		
21	200	-	249		
22	250	-	299		
23	300	-	349		
24	350	-	399		
25	400	-	449		
26	450	-			

関数 **FREQUENCY** の引数は 2 個で第 1 引数がもとのデータの範囲、第 2 引数が階級の範囲である。したがって、**=FREQUENCY(gym,class)** とすれば各度数を結果として得ることができるが、この関数の戻り値は あたえられた階級上限の数 (k) +1 個の要素をもつ配列であり、各要素が対応する階級の度数をあらわす。

関数 **FREQUENCY** の戻り値：

- 1 個目：最初の階級の上限以下の値をもつデータの数
- 2 個目：最初の階級の上限より大で、第 2 の階級の上限を越えない値をもつデータの数
- .....
- k+1 個目 (最後)：k 番目の階級の上限をこえる値をもつデータの数

表 1 の class は 9 個の階級上限を設定したから、戻り値は 10 個の度数からなる配列となる。

また、関数 **FREQUENCY** は配列関数である。Excel で関数は、1 つの数値を返すものであるが、配列関数は複数の配列を返すものである。この場合、配列を記入する範囲を指定し、関数を入力した上で、Enter キーの代わりに、**Ctrl** + **Shift** + **Enter** キーを入力する。具体的な手順は次のとおりである。

#### 📖 手順

- 範囲 J17:J26 を選択する。
- セル J17 に式 **=FREQUENCY(gym,class)** を入力する。
- Ctrl** + **Shift** + **Enter** とする。

このような配列関数は、他に行列の積や逆行列などを求めるものなどがある。

#### 2. ヒストグラムの作成

グラフを作成するにはメニューから 挿入 - グラフ を選ぶ。すると、グラフ・ウィザードが自動的に起動する。

グラフ・ウィザードはどのデータに対してどの種類のグラフを描くかを対話的に問い合わせてくるから、キーボードから入力したり、マウスで範囲を指定したりすれば、自動的にグラフが作成・表示される。

なお、1度描かれたグラフはもとのデータを変更すれば自動的に描きなおされる。

1. グラフ・ウィザードが起動すると、まず、グラフの種類の選択メニューになる。ここでは 縦棒 を選ぶ。
2. すると右側に縦棒グラフのフォーマット（型式）メニューが出るので、左上のものをを選び、**次へ>**をクリックする。
3. この時点でグラフのサンプルが自動的に描かれている。これを修正していく。  
グラフにするデータの範囲：入力ボックスに範囲または範囲名を直接入力してもよいが、シート上にマウスを移動して範囲指定することもできる。ここでは、J17:J25 を指定する。  
系列：列(L)を選ぶ。
4. 階級値を横軸の目盛りに使用する。  
系列のタグをクリックし、項目軸ラベルに使用するボックスの右端のボタンをクリックし、I17:I25 をドラッグ。ウインドウの右端のボタンをクリック  
という操作を行う。
5. 次にグラフのさまざまな書式を設定する。
  - a) 「タイトルとラベル」のタグにおいて  
グラフ タイトル：「都道府県別体育館数」と記入する。  
X/項目軸：「階級値」と記入する。  
Y/数値軸：「度数」と記入する。

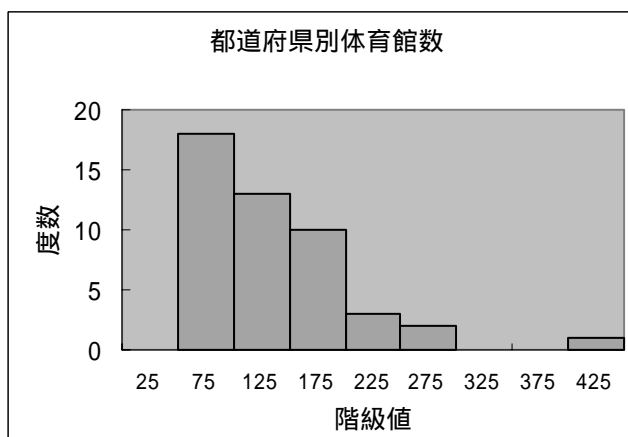
**入力が済んだら次のタグをクリックする。**

  - b) 「目盛線」のタグにおいて  
Y/数値軸の目盛線：チェックをはずす。(クリックすればはずれる)
  - c) 「凡例」のタグにおいて  
凡例を表示する：チェックをはずす。

**a),b),c)のすべての操作が済んだら**次へ>**のボタンをクリックする。  
a),b),c)の順序は入れ替ってもかまわない**

6. オブジェクトかグラフシートの作成かを選ぶ。ここでは、オブジェクトを選択する。  
オブジェクト：データと同一のワークシート上にグラフを描く  
グラフシート：グラフのために新しいシートを作成する
7. 以上で、グラフ・ウィザードは終了する。グラフの移動は $\text{N}$ のカーソルでおこない、幅の変更は $\text{↔}$ などの形のカーソルでおこなう。高さの変更がどのようにおこなえるか各自試みよ。  
さらに細かい修正をするには、グラフの位置にマウスポインタで指してダブルクリックをする。  
グラフ・ウィンドウが現われるので、修正したい要素をダブルクリックすると、ダイアログ・ボックスが現われる。

ヒストグラムでは、縦棒どうしの間隔をあけない。ここでは縦棒どうしの間隔を変更してみよう。まず、適当な縦棒をマウスポインタ（矢印の型）で指してクリックするとすべての縦棒に ■ が現われる。ここで、ダブルクリックをすると、データ系列の書式設定ウィンドウが開く。その「オプション」のタグで、棒の間隔を 0 にする。さらに、項目軸目盛りのフォントなどを変更すると、左図のようなグラフが描ける。



### 3. その他の修正

#### a) 都道府県別に順位を付ける

【課題 11】関数 **RANK** を用いて、体育館数の多い順に都道府県に順位を付けてみよう。

関数 **RANK**(引数 1, 引数 2, 引数 3) は特定のデータ(指定数値)が全体で何番目の大きさになるかを調べる関数である。引数 1 は指定数値、引数 2 は範囲、引数 3 は昇順(小さい順)か降順(大きい順)を決めるためのオプションで、1 のとき昇順、0 のとき降順となる。

、引数 3 を省略すると、0 を指定したことになる。

=RANK(B4, B4:B50, 0)

セル B4 のデータが範囲 B4:B50 の中で何番目に大きいかを調べる。

=RANK(B4, gym) と同じ、0 は省略できる。

=RANK(406, gym, 1)

データ 406 が小さいほうから数えて何番目であるかを調べる。

ここでは、縦にならんだ体育館数データの右隣の列に

= RANK(B4, gym, 0)

として、全都道府県のコピーすればよい。B4 は北海道の体育館数が入力されているセルの番地であるものとする。

#### b) 罫線・フォントの変更

罫線は必ずしも必要なものではないが、引いておくとも表を見やすくすることができる。罫線を引くときは、あらかじめ、罫線を引きたい範囲を選択しておいてから、書式-セル-罫線でおこなう。または、ツールバーの中に罫線設定のアイコンがあるので、これをクリックすることでもおこなうことができるが、細かい設定は意外と難しい。

フォントの変更もまた表を見やすくすることがある。ディスプレイで見ている限りフォントの違いはあまり気にならないが、プリンタで印刷するとフォントによってかなり印象が異なる。フォントの種類は大別すると半角文字専用のもので、半角・全角兼用のものに別れる。前者の代表的なものが、Arial, Courier New, Century, Times New Roman などの英文フォントで、後者の代表的なものは MS <sup>みんちょう</sup>明朝 や MS ゴシックである。通常は MS 明朝にしておけば、特に不都合はないが、半角のアルファベットや数字に英文フォントを指定するとかなり見栄えはよくなる。ただし、全角文字や ¥ に対して英文フォントを指定すると意味不明な表示になるので注意せよ。

また、フォントサイズを変更することによって、文字の大きさを変えたり、太字や斜体字を用いることもできる。各自試してみよ。

### 4. シートの印刷

フォント・レイアウトなどの修正がすんだら印刷出力をする。印刷の際にはレイアウトを確認するために必ず印刷プレビューをしてから印刷すること。

(1) メニューバーの「ファイル」-「印刷プレビュー」を選択すると、印刷時のイメージ画面が出てくる。ここで、出力結果が 1 ページに収まっていることを確認する。1 ページに収まらない場合は、グラフの大きさを変えるなどの修正を加える必要がある。プレビューがすんだら、「閉じる」のボタンを押す。すると、ワークシートに点線で 1 ページの幅が見えるはずである。

(2) 1 ページに収まったことを確認した上で、メニューバーの「ファイル」-「印刷」を選択する。すると「印刷」ウィンドウが開くので、印刷範囲をページ指定 1 ページから 1 ページまでとする。ここで OK ボタンを押すと印刷される。

**パソコン教室のプリンターはコンピュータ 20 台で 1 台のプリンターを共有しているので、出力に少し時間がかかる場合がある。なかなか出力が出ないからといって何度も印刷をせずに、少し待つように！**

☒ 演習問題 1 : 表 1 を A4 用紙縦 1 枚に納まるようにレイアウトして、印刷してみよ。上部の空いているセルに学生番号、名前を記入しておくこと。次ページに仕上がり例を示す。

< 作成見本 >

都道府県別体育館数(2002.10.1現在)			E34-000	徳山 太郎		
<b>都道府県</b>	<b>体育館数</b>	<b>RANK</b>	<b>特性値</b>			
北海道	406	1	データ数	47		
青森	103	29	算術平均	135.98		
岩手	178	12	メディアン	124		
宮城	134	19	モード	91		
秋田	159	15	分散	4509		
山形	99	30	標準偏差	67.149		
福島	242	4	最大値	406		
茨城	133	20	最小値	53		
栃木	91	33	レンジ	353		
群馬	126	23	四分位偏差	43		
埼玉	165	14	<b>度数分布表</b>			
千葉	127	22	階	級	階級値	度数
東京	190	8	0 -	49	24.5	0
神奈川	221	5	50 -	99	74.5	18
新潟	250	3	100 -	149	124.5	13
富山	112	27	150 -	199	174.5	10
石川	131	21	200 -	249	224.5	3
福井	86	37	250 -	299	274.5	2
山梨	90	35	300 -	349	324.5	0
長野	270	2	350 -	399	374.5	0
岐阜	184	11	400 -	449	424.5	1
静岡	115	25	450 -			0
愛知	194	7				
三重	97	31				
滋賀	91	33				
京都	82	40				
大阪	173	13				
兵庫	203	6				
奈良	124	24				
和歌山	53	46				
鳥取	114	26				
島根	89	36				
岡山	93	32				
広島	143	17				
山口	83	39				
徳島	69	42				
香川	77	41				
愛媛	84	38				
高知	55	45				
福岡	187	10				
佐賀	60	44				
長崎	108	28				
熊本	188	9				
大分	69	42				
宮崎	136	18				
鹿児島	154	16				
沖縄	53	46				
全国	6391					

