

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は全国の野球場・ソフトボール場(以下「球場」とあらわす)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 都道府県別球場数(2005.10.1現在)

北海道	582	埼玉	413	岐阜	159	鳥取	47	佐賀	45
青森	82	千葉	292	静岡	193	島根	54	長崎	65
岩手	106	東京	560	愛知	341	岡山	97	熊本	104
宮城	134	神奈川	195	三重	75	広島	100	大分	103
秋田	113	新潟	187	滋賀	46	山口	29	宮崎	83
山形	58	富山	56	京都	89	徳島	32	鹿児島	62
福島	129	石川	92	大阪	126	香川	66	沖縄	48
茨城	239	福井	71	兵庫	123	愛媛	46		
栃木	236	山梨	32	奈良	35	高知	23		
群馬	219	長野	198	和歌山	76	福岡	188		

データ出典： 文部科学省 『平成17年社会教育調査』

<作成見本>

	A	B	C	D	E
1	都道府県別球場数(2005.10.1現在)				
2					
3	都道府県	球場数			
4	北海道	582			
5	青森	82			
6	岩手	106			
7	宮城	134			
8	秋田	113			
9	山形	58			
10	福島	129			
11	茨城	239			
12	栃木	236			
13	群馬	219			
14	埼玉	413			
15	千葉	292			
16	東京	560			
17	神奈川	195			

1. 連続データリストの登録

連続データの作成については前回説明したが、登録されていない連続データは自分でユーザー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ー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ー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、ユーザー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 変量データの特徴値

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**特徴値**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徴値には

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変量データの特徴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徴値が求まる。ここではまず合計を求める関数SUMを用いてみる。

【課題 7】都道府県別球場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 ballpark とする）を付けることによって、=SUM(ballpark) 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				
3				

<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 】 都道府県別球場数のデータに ballpark 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、 ballpark 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 ballpark 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つの方法なのである、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 **ballpark.xls** の Sheet1 に入力され、球場数を表す 47 個のデータには範囲名 **ballpark** が設定されているものとする。

	F	G	H	I	J
16	階		級	階級値	度数
17	0	-	49	$=(f17+h17)/2$	$=FREQUENCY(ballpark,class)$
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	-	499		
27	500	-	549		
28	550	-	599		
29	600	-			

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:H28）に範囲名 **class** を付けておく。列 I には、 $(\text{下限} + \text{上限}) / 2$  として、階級値を求めておく。

関数 **FREQUENCY** の引数は 2 個で第 1 引数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、 $=FREQUENCY(ballpark,class)$  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 12 個の階級上限を設定したから、戻り値は 13 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:J29 を選択する。
- セル J17 に式  $=FREQUENCY(ballpark,class)$  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:J28 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のボックスの右端のボタンをクリックし、I17:I28 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：「都道府県別球場数」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。

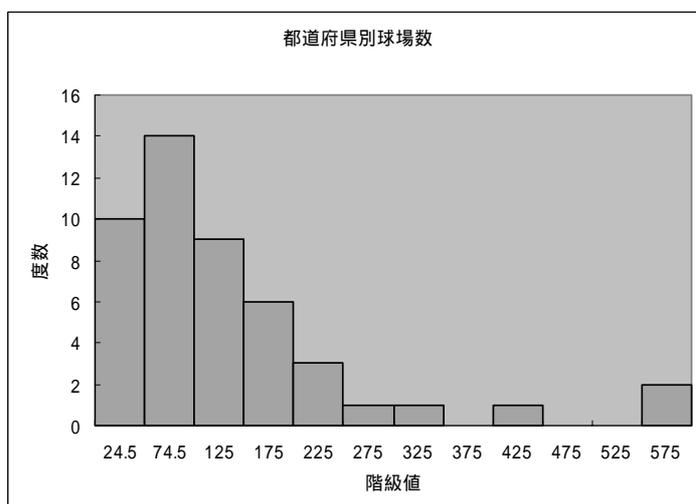
- b) 「目盛線」のタグにおいて  
Y/数値軸の目盛線：チェックをはずす。(クリックすればはずれる)
- c) 「凡例」のタグにおいて  
凡例を表示する：チェックをはずす。

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インドウが開く。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、球場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, ballpark) と同じ。0 は省略できる。

=RANK(582, ballpark, 1)

データ 582 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、縦にならんだ球場数データの右隣の列に

=RANK(B4, ballpark, 0)

として、全都道府県のコピーすればよい。B4 は北海道の球場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、細かい設定は案外と難しい。

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」ウインドウが開くので、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 >

都道府県別球場数(2005.10.1現在)			E37-000	徳山 太郎		
<b>都道府県</b>	<b>球場数</b>	<b>RANK</b>	<b>特性値</b>			
北海道	582	1	データ数	47		
青森	82	28	算術平均	137.213		
岩手	106	20	メディアン	97		
宮城	134	15	モード	32		
秋田	113	19	分散	15415.7		
山形	58	35	標準偏差	124.16		
福島	129	16	最大値	582		
茨城	239	6	最小値	23		
栃木	236	7	レンジ	559		
群馬	219	8	四分位偏差	65.25		
埼玉	413	3	<b>度数分布表</b>			
千葉	292	5	階	級	階級値	度数
東京	560	2	0 -	49	24.5	10
神奈川	195	10	50 -	99	74.5	14
新潟	187	13	100 -	149	124.5	9
富山	56	36	150 -	199	174.5	6
石川	92	25	200 -	249	224.5	3
福井	71	31	250 -	299	274.5	1
山梨	32	44	300 -	349	324.5	1
長野	198	9	350 -	399	374.5	0
岐阜	159	14	400 -	449	424.5	1
静岡	193	11	450 -	499	474.5	0
愛知	341	4	500 -	549	524.5	0
三重	75	30	550 -	599	574.5	2
滋賀	46	40	600 -			0
京都	89	26				
大阪	126	17				
兵庫	123	18				
奈良	35	43				
和歌山	76	29				
鳥取	47	39				
島根	54	37				
岡山	97	24				
広島	100	23				
山口	29	46				
徳島	32	44				
香川	66	32				
愛媛	46	40				
高知	23	47				
福岡	188	12				
佐賀	45	42				
長崎	65	33				
熊本	104	21				
大分	103	22				
宮崎	83	27				
鹿児島	62	34				
沖縄	48	38				
全国	6449					
	6449					

