

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(2005.10.1現在)

北海道	413	埼玉	168	岐阜	157	鳥取	124	佐賀	72
青森	100	千葉	130	静岡	127	島根	105	長崎	111
岩手	185	東京	212	愛知	201	岡山	94	熊本	200
宮城	136	神奈川	268	三重	92	広島	142	大分	74
秋田	180	新潟	249	滋賀	90	山口	102	宮崎	137
山形	97	富山	123	京都	80	徳島	74	鹿児島	156
福島	262	石川	136	大阪	170	香川	78	沖縄	56
茨城	132	福井	84	兵庫	223	愛媛	106		
栃木	95	山梨	102	奈良	124	高知	61		
群馬	127	長野	264	和歌山	68	福岡	187		

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
1	都道府県別体育館数(2005.10.1現在)			
2				
3	都道府県	体育館数		
4	北海道	413		
5	青森	100		
6	岩手	185		
7	宮城	136		
8	秋田	180		
9	山形	97		
10	福島	262		
11	茨城	132		
12	栃木	95		
13	群馬	127		
14	埼玉	168		
15	千葉	130		
16	東京	212		
17	神奈川	268		

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。
- 左上の **Office ボタン** を押し、下にある「Excel のオプション」のボタンをクリックする。
 - 「基本設定」の中にある「ユーザー設定リスト」のボタンをクリックする。
 - 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 tennis とする）を付けることによって、=SUM(tennis) 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¹。

📖 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				

¹ 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。数式のタブをクリックし、リボンの中から「名前の定義」をクリックし、「名前」として、 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)の右側の 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、名前ボックスの右側の 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

Excel2007 では、描きたいグラフに近いグラフをとりあえず描き、その上でレイアウトを変更したり、タイトルや軸の名前を描き入れたりしていく。(Excel2003 まであったグラフ・ウィザードは無くなった。)

都道府県別体育館数の度数分布表をヒストグラムに表すには、次のような手順でおこなう。

1. 最初にグラフに描く範囲を範囲指定する。ここでは、I17:I25(最後の0は除く)を範囲指定する。
2. グラフを作成するには、**挿入タブ**をクリックすることで、リボン内にグラフのグループが表示される。ヒストグラムは縦棒グラフの1種なので、**縦棒**のボタンをクリックする。
3. すると縦棒グラフのフォーマット(型式)メニューが出るので、**集合縦棒(2-D縦棒の中の左端)**をクリックする。
4. 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。
まず、横軸ラベルに階級値を用いる。リボンの中の「データの選択」ボタン(「データ」のグループにある)をクリックし、横(項目)軸ラベルの「編集」ボタンをクリックし、H17:H25を範囲指定し、OKボタンを押す。

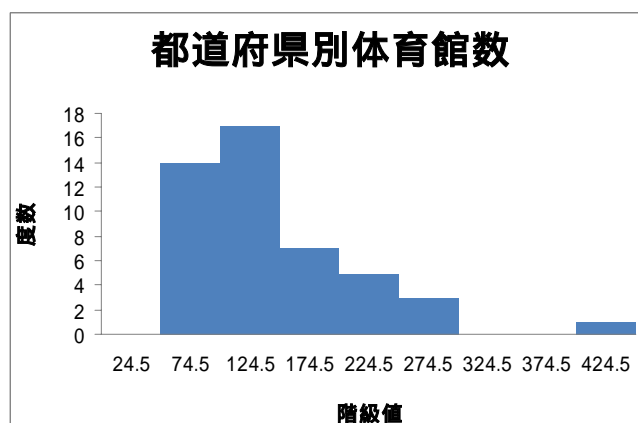
5. 次に、レイアウトを変更する。リボンの中の「グラフのレイアウト」を展開し、レイアウト8(3段目の真ん中)をクリックし、タイトルや軸ラベルなどが書き込めるようにする。そして、

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。

6. 以上で、左図のようなヒストグラムが描けたはずである。グラフの移動はグラフの白い部分をドラッグすることで、大きさの変更はグラフの4隅または4辺の真ん中にマウスを合わせ、ドラッグすることでできる。各自試みよ。



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(413, gym, 1)

データ 413 が小さいほうから数えて何番目であることを調べる。

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、全都道府県の分コピーすればよい。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、リボン内のフォントのグループ(ホームタブをクリックすることで表示される。)にある、罫線設定のボタンを展開し、罫線を引く場所を選択する。ここで、最下部にあるか「その他の罫線」を選択すれば、Excel2003まででおこなわれていた方法(セルの書式設定)で罫線を引くことができる。

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 ^{みんちょう}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、¥ など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ことがあるので注意せよ。

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) 左上のOfficeボタンを押し、「印刷」-「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、「印刷プレビューを閉じる」のボタンを押し。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、Officeボタンを押し、「印刷」-「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ボタンを押し印刷される。

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(2005.10.1現在)			E37-000	徳山 太郎		
都道府県	体育館数	RANK	特性値			
北海道	413	1	データ数	47		
青森	100	33	算術平均	142		
岩手	185	11	メディアン	127		
宮城	136	19	モード	136		
秋田	180	12	分散	4735.57		
山形	97	34	標準偏差	68.8155		
福島	262	4	最大値	413		
茨城	132	21	最小値	56		
栃木	95	35	レンジ	357		
群馬	127	23	四分位偏差	40.25		
埼玉	168	14	度数分布表			
千葉	130	22	階	級	階級値	度数
東京	212	7	0 -	49	24.5	0
神奈川	268	2	50 -	99	74.5	14
新潟	249	5	100 -	149	124.5	17
富山	123	27	150 -	199	174.5	7
石川	136	19	200 -	249	224.5	5
福井	84	39	250 -	299	274.5	3
山梨	102	31	300 -	349	324.5	0
長野	264	3	350 -	399	374.5	0
岐阜	157	15	400 -	449	424.5	1
静岡	127	23	450 -			0
愛知	201	8				
三重	92	37				
滋賀	90	38				
京都	80	40				
大阪	170	13				
兵庫	223	6				
奈良	124	25				
和歌山	68	45				
鳥取	124	25				
鳥根	105	30				
岡山	94	36				
広島	142	17				
山口	102	31				
徳島	74	42				
香川	78	41				
愛媛	106	29				
高知	61	46				
福岡	187	10				
佐賀	72	44				
長崎	111	28				
熊本	200	9				
大分	74	42				
宮崎	137	18				
鹿児島	156	16				
沖縄	56	47				
全国	6674					
	6674					

