

表計算 (Excel) とデータ処理 (6)

1 1つの変数の記述統計

「変数」とは、「あるクラスに属する学生の国語のテストの得点」や、「ポケモンの中でもっとも好きなキャラクタの名前」「さいころの目の数」など、特定の範囲の中で変化しうるものである（数値だけでなく、「ピカチュウ」や「コイル」など値として名前をとる場合もある）。

変数は、尺度水準（統計学的な水準）で分類すると、「名義尺度」「順序尺度」「間隔尺度」「比例尺度」に分類することができる。

この変数は、実際のデータでは多くの具体的な値からなる（例えば1クラス40名分の国語の得点など）。これを一つの値として集約することを「要約」「要約統計量」などという。

要約統計量は大きく分けて「代表値」（変数の代表的な値）と「散布度」（変数の広がり）からなる。

	例	代表値	散布度	計算
名義尺度	好きなポケモン名 車のナンバー	最頻値（カテゴリ中最も度数の大きい値）	なし （カテゴリ数）	=, ≠
順序尺度	競走の順位 嗜好（好き嫌い）順	中央値（並べ替えたときに順番が真中の値） （最頻値）	最大・最小値 第一・第三四分位	=, ≠ >, <
間隔尺度	摂氏温度 カレンダーの日付	平均（算術平均） （中央値, 最頻値）	分散/標準偏差	+, - >, <, =, ≠
比例尺度	絶対温度 年齢, 収入	平均（算術平均） （幾何平均, 中央値, 最頻値）	分散/標準偏差	×, ÷, +, - >, <, =, ≠

※間隔尺度以上では、他に分布の形を数値的に表す歪度（左右の歪み具合）、尖度（尖り具合）がある。

2 変数と代表値

「J1リーグ名簿」ワークシートを使って、変数と各尺度について考えてみよう。

	A	B	C	D	E	F	G	H	I
2		整理NO	クラブ	選手名	位置	身長(cm)	体重(kg)	出身地	
3		1	ベガルタ仙台	桜井 繁	GK	184.0	81.0	茨城	
4		2	ベガルタ仙台	林 卓人	GK	188.0	87.0	大阪	
5		3	ベガルタ仙台	関 憲太郎	GK	177.0	74.0	群馬	
6		4	ベガルタ仙台	鎌田 次郎	DF	179.0	77.0	東京	
7		5	ベガルタ仙台	渡辺 康太	DF	180.0	77.0	千葉	

[練習] このワークシートのデータについて、「変数」と「変数の尺度」について整理してみなさい。

3 代表値（1） 最頻値（mode モード）（名義尺度以上）

「クラブ」や「出身地」はそれぞれ「異なる」ことだけが区別できるので、「名義尺度」である。「登録選手数が最も多いクラブはどこか」「J1リーグの登録選手が最も多い出身地はどこか」のように、尺度の値ごとの頻度分布は「もっとも頻度が高い」ものが代表値となる。これを求めてみよう。

- a) 表を上下に分割し，下半分に表の末尾が見えるようにする。
- b) K6 に以下の式を入力する。

=COUNTIF(\$C\$3:\$C\$545, J6)

COUNTIF(範囲, 条件) は指定範囲の中に値に等しいセルが何個あるかを数える関数である。等しい場合を調べるには，条件にその値を指定すればよい。

- c) 範囲を指定するには，C3 をクリックした後 Shift キーを押しながら B545 をクリックすることも可能である。ただし下方向にコピーして使うので，範囲は絶対指定にする。
- d) G6 を G23 までコピーする。

	A	B	C	D	E	F	G	H	I	J	K	L	M
2		整理NO	クラブ	選手名	位置	身長(cm)	体重(kg)	出身地					
3		1	ベガルタ仙台	桜井 繁	GK	184.0	81.0	茨城					
4		2	ベガルタ仙台	林 卓人	GK	188.0	87.0	大阪					
5		3	ベガルタ仙台	関 憲太郎	GK	177.0	74.0	群馬	クラブ別登録選手数	人数			
6		4	ベガルタ仙台	鎌田 次郎	DF	179.0	77.0	東京	ベガルタ仙台	=COUNTIF(\$C\$3:\$C\$545, J6)			
7		5	ベガルタ仙台	渡辺 広大	DF	180.0	77.0	千葉	鹿島アントラーズ	30			
8		6	ベガルタ仙台	石川 直樹	DF	180.0	73.0	千葉	浦和レッズ	28			
9		7	ベガルタ仙台	角田 誠	DF	180.0	75.0	京都	大宮アルディージャ	31			
10		8	ベガルタ仙台	蜂須賀 孝治	DF	180.0	75.0	栃木	柏レイソル	30			
11		9	ベガルタ仙台	上本 大海	DF	180.0	68.0	鹿児島	FC東京	27			
12		10	ベガルタ仙台	石川 大徳	DF	170.0	65.0	東京	川崎フロンターレ	31			
13		11	ベガルタ仙台	二見 宏志	DF	179.0	75.0	大阪	横浜F・マリノス	32			
14		12	ベガルタ仙台	松下 年宏	MF	174.0	70.0	鹿児島	湘南ベルマーレ	36			
15		13	ベガルタ仙台	梁 勇基	MF	173.0	72.0	大阪	ヴァンフォーレ甲府	31			
16		14	ベガルタ仙台	太田 吉彰	MF	175.0	68.0	静岡	アルビレックス新潟	31			
17		15	ベガルタ仙台	佐々木 勇人	MF	167.0	58.0	宮城	清水エスパルス	32			
18		16	ベガルタ仙台	ヘベルチ	MF	173.0	72.0	ブラジル	ジュビロ磐田	30			
19		17	ベガルタ仙台	富田 晋伍	MF	169.0	64.0	栃木	名古屋グランパス	30			
20		18	ベガルタ仙台	田村 直也	MF	175.0	70.0	東京	セレッソ大阪	26			
21		19	ベガルタ仙台	菅井 直樹	MF	178.0	70.0	山形	サンフレッチェ広島	30			
22		20	ベガルタ仙台	藤村 慶太	MF	180.0	65.0	岩手	サガン鳥栖	30			
23		21	ベガルタ仙台	ジゴゴ	MF	181.0	80.0	ブラジル	大分トリニータ	31			

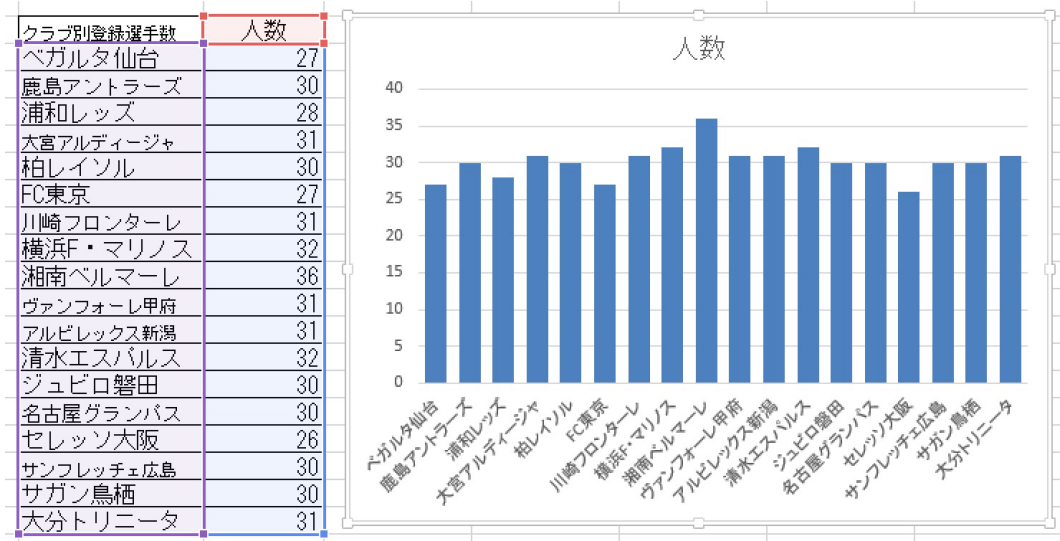


「クラブ登録選手数」の最頻値はいくつで，どのクラブか。

[練習] 出身別選手数について，最頻値とその出身を求めなさい。

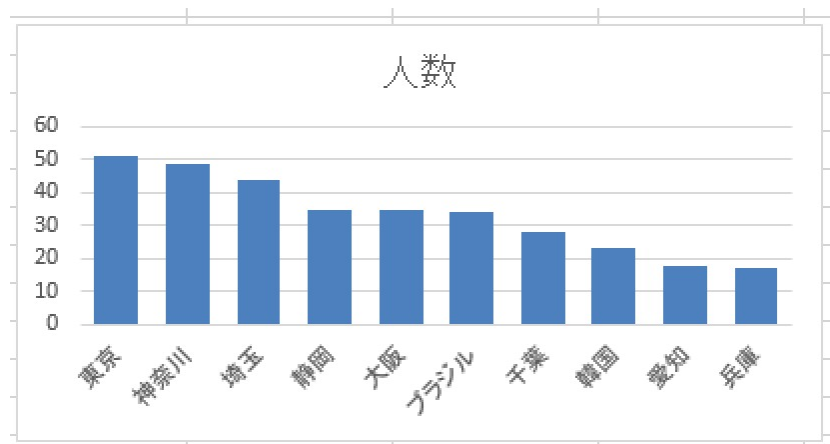
4 ヒストグラム（1） 名義尺度データ

「登録選手数が最も多いクラブはどこか」のヒストグラムを書きなさい。



- a) 表の領域を選択する。
- b) 挿入 - 縦棒グラフ - 2D縦棒 - 集合縦棒を選択する。
- c) グラフ全体の位置と幅を調整する。
- d) 棒の幅は棒を右クリック - データ系列の書式設定-要素の間隔, で調整する。

[練習] 「J1リーグの登録選手が最も多い出身地はどこか」について、上位10都道府県についてヒストグラムを書きなさい。（表全体を選択して - [データ] - [フィルタ] - 降順に並べ替え）



5 代表値（2） 中央値（median メディアン）（順序尺度以上）

「身長」「体重」を並べ替えたとき、ちょうど真ん中の順位の値はいくつで選手は誰か。全税会の練習を思い出して、身長のヒストグラムを作成してみよう。

- a) ヒストグラムの範囲を決めるために、身長 の最大値と最小値を関数 MAX(範囲), MIN(範囲) を使って調べる。
- b) 最大値, 最小値に合わせてヒストグラムの表を作成する (今回はJ87 に作成済み)
- c) K85:K93 を選択し, ヒストグラム関数 =FREQUENCY(F3:F545, J85:J93) を入力して, CTRL+SHIFT+Enter キーを押す。
- d) このヒストグラムから計算する場合は, データ数は 543 あるので, 272 番目の値を含む階級は 175-180。そこでその階級の真ん中の値 177.5 が中央値になる。
- e) 関数 MEDIAN(範囲) で中央値を求めることもできる。

身長	人数	階級値		身長	人数	階級値
160.0	=FREQUENCY (F3:F545, J88:J96)			160.0	1	
165.0	5	162.5	人数	165.0	5	
170.0	68	167.5		170.0	68	
175.0	125	172.5		175.0	125	
180.0	138	177.5		180.0	138	
185.0	142	182.5		185.0	142	
190.0	52	187.5		190.0	52	
195.0	10	192.5		195.0	10	
200.0	2	197.5		200.0	2	

[練習] 体重についても頻度分布と関数をそれぞれ使って中央値を求めなさい。

6 代表値 (3) 算術平均 (average アベレージ) (間隔尺度以上)

「Jリーガーの身長, 体重の平均はいくつか？」

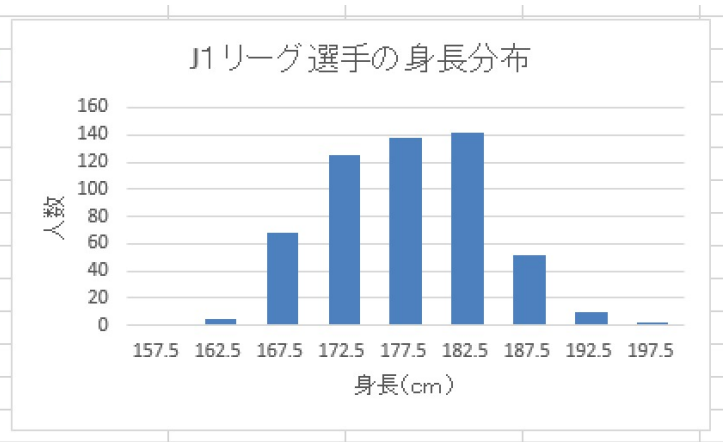
算術平均とは, 総数をカテゴリ数で割った数である。

エクセルでは AVERAGE(範囲) という関数で, 自動的に平均を計算させることができる。「身長」「体重」について平均を求めてみよう。

7 ヒストグラム (2) 間隔/比例尺度データ

1) 身長の高ヒストグラムを作成しなさい。

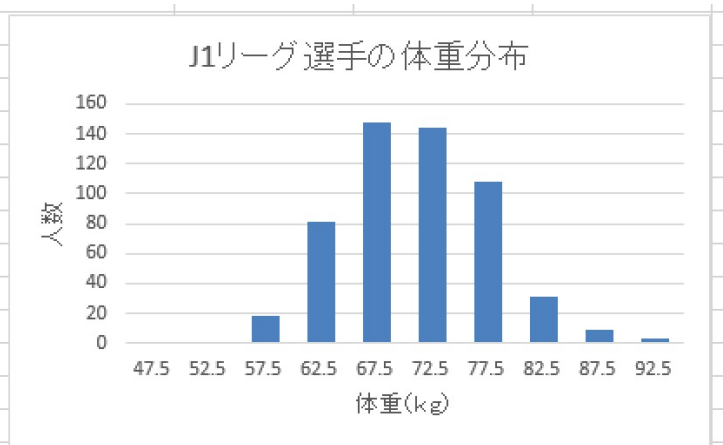
身長	人数	階級値
160.0	1	$=N88-2.5$
165.0	5	162.5
170.0	68	167.5
175.0	125	172.5
180.0	138	177.5
185.0	142	182.5
190.0	52	187.5
195.0	10	192.5
200.0	2	197.5
体重	人数	階級値
50.0	1	



- 作成した表の「身長」は階級の「上限値」なので、グラフに使用する階級値を計算する。(この場合、上限値-2.5)
- 表の「人数」を選択し、2D集合棒グラフを作成する。
- グラフを右クリックし、「データの選択」メニューで「横(項目)軸ラベル」の「編集」ボタンをクリックして、計算した階級値を選択し、項目を入れ替える。
- グラフを選択し、レイアウト - ラベル - グラフタイトル - グラフの上でタイトルを作成する。
- グラフを選択し、グラフツールのデザイン - グラフ要素を追加 - 軸ラベルで、縦横軸のラベルを追加する

2) (練習) 体重の高ヒストグラムを作成しなさい。

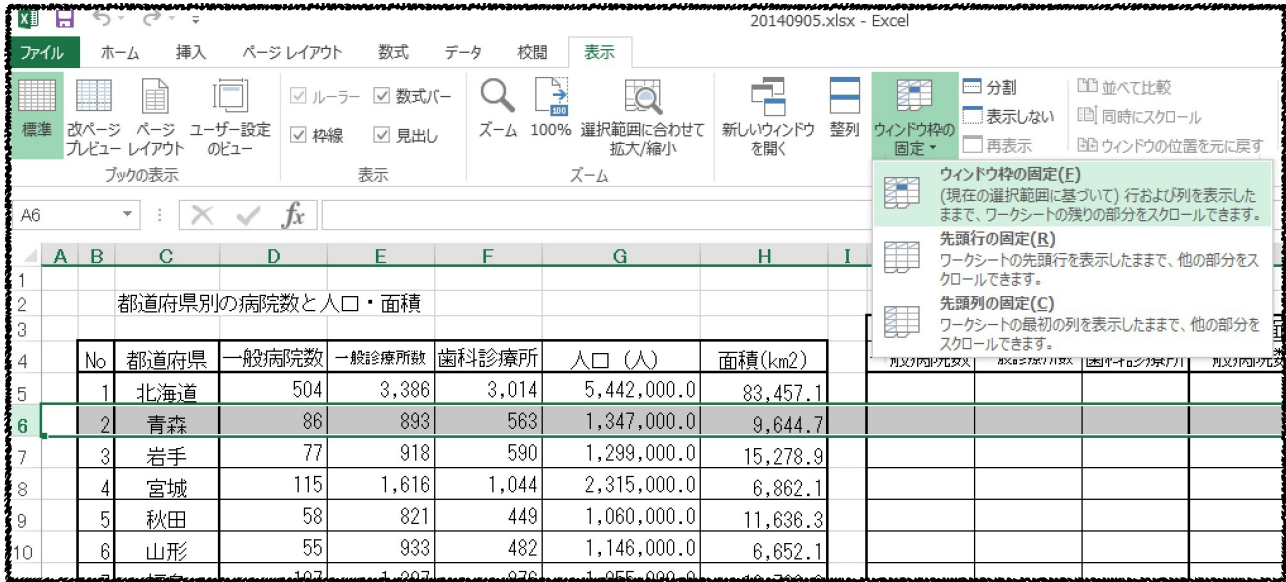
体重	人数	階級値
50.0	1	47.5
55.0	0	52.5
60.0	18	57.5
65.0	81	62.5
70.0	148	67.5
75.0	144	72.5
80.0	108	77.5
85.0	31	82.5
90.0	9	87.5
95.0	3	92.5



表計算 (Excel) とデータ処理 (6) 補足

ワークシートの分割

6行目を選択して、「表示」メニューから「ウィンドウ枠の固定」



フィルタの設定

並べ替えたい表全体を選択して、「データ」メニューから「フィルター」

