

社会調査のためのエクセル利用法

小林久高・山本圭三・金政芸・猿渡壮

KOBAYASHI Hisataka, YAMAMOTO Keizo, KIM Jeongwoon, SARUWATARI Takeshi

1 はじめに

社会調査の諸局面では、エクセルはさまざまに利用可能である。それは、作業工程の管理だけでなく、コードブックの作成やデータ入力についても役立つ、また、統計分析においても力を発揮する。このような便利なソフトであるにもかかわらず、エクセルが社会調査において十分に利用されているとは言えないように思う。そこには小さなハードルがあるのかもしれない。そこで本稿では、エクセルについて、(1)ワークシートでの計算の基礎、(2)データ入力とチェックにおける利用法、(3)データ分析における利用法を述べる。本稿は、多少エクセルを使ったことのある読者を対象としている。基本的な使い方については、他の参考書にあたられたい。

2 エクセルでの計算の基礎

2.1 データの「型」

エクセルのセルに入るデータには大きく分けて、数字、式、文字がある。数字と式は半角（直接入力）で入力したほうがいい。数字は通常どおり入力する。式を入力するときには「=」を前につける必要がある（+をつけてもいい）。たとえば「3+3」という式をセルに入れたいときは、「=3+3」と入力する。式をセルに入力するとセルには答えが表示される。そのまま「3+3」と入力すると、それは文字として認識されるので注意が必要である。

アルファベットや日本語は自動的に文字として

扱われる。はじめに「」（shift+7）を入力すると、以下は文字とみなされる。文字は数値的には0（ゼロ）とみなされる。「1000」と入力すると、セルには「1000」と表示されるが、それは数値的には0なのである。

たとえば、A1に「同志社大学」という文字、A2に「123」という数値、A3に「123」という文字、A4に「1+2」という式、A5に「1+2」という文字を入力すると図1のようになる。

図 1

	A
1	同志社大学
2	123
3	123
4	3
5	1+2
6	

2.2 演算

エクセルには演算に用いる次のような演算子がある。加法「+」、減法「-」、乗法「*」、除法「/」。乗法と除法は通常の計算には使わない記号を使うので注意が必要である。これらの演算子を入力する際には、テンキー（キーボードの右にある数字キー）を利用するのが便利である。あるものを2乗、3乗するときには「^」を使う。たとえば3の2乗は「3^2」とすればいい。カッコは「()」を用いるが、大カッコ、中カッコといった区別はない。エクセルは内側のカッコから順番に計算していく。

B1に「2+3」、B2に「2-3」、B3に「2*3」、B4

に「2/3」、B5に「2の3乗」の式を入力すると図2のようになる。文字を入力するのではないことに注意しなければならない。

図 2

	B
1	5
2	-1
3	6
4	0.666666667
5	8

2.3 セル番地を使った計算

エクセルが通常の電卓よりも便利なのは、セルの番地を用いて計算することができるということである。たとえば、B7に「=B1+B2+B3+B4+B5+B6」と入力すると、B7にはB1からB6に入っている値の合計が表示される。さらに便利なのは、セル番地を使った式を入力しておく、セルの中身を変えたとき自動的に再計算してくれることである。上の式の場合、B1の値を変えると、B7の値は自動的に変わる。

D1に100、D2に200、E1に300、E2に400を入力し、セル番地を用いて、D3とE3にそれぞれの縦列の合計が、F1とF2にそれぞれの横行の合計が出るようにし、全体の合計がF3に出るようにしてみよう（図3、4）。図のD1,D2,E1,E2の値を変えたときに各合計が正しく算出される。

図 3

	D	E	F
1	100	300	=D1+E1
2	200	400	=D2+E2
3	=D1+D2	=E1+E2	=F1+F2
4			

図 4

	D	E	F
1	100	300	400
2	200	400	600
3	300	700	1000
4			

2.4 相対番地とコピー・貼り付け

データを連続的に入力する際にコピーと貼り付けは非常に便利である。たとえば、A1からA10に1から10を入力する際、1~10をいちいち入力する必要はない。そんなときには(1)A1に「1」、A2に「=A1+1」を入力し、(2)A2をコピーし、A3からA10に貼り付ければよい。これでA1からA10に1から10までが並ぶ（図5）。

図 5

	A
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10

図 6

	A
1	1
2	=A1+1
3	=A2+1
4	=A3+1
5	=A4+1
6	=A5+1
7	=A6+1
8	=A7+1
9	=A8+1
10	=A9+1

ところで、このときにそれぞれのセルに入力さ

れている式を見てみると、A3では「=A2+1」、A4では「=A3+1」、A5では「=A4+1」という具合になっており、貼り付けられた式がもとの式とは異なっていることがわかる（図6）。

貼り付けられたものはもとの式そのままではなく、位置が調節された式なのである。A2にある「A1+1」という式は、実は「A2にとってのA1」、すなわち「自分の上にあるセル」という意味なのであり、それが貼り付けられた結果、A3はA2+1、A4はA3+1、A5はA4+1という内容になったのである。セル番地を指定するこのような方法（A1、A2、A3、…）は、相対番地による指定法という。それは、当該セルとの相対的な位置関係でセルが指定される方法である。

2.5 絶対番地とコピー・貼り付け

上のような相対番地ではなく、絶対番地を指定することもできる。相対番地はA1というように指定したが、絶対番地では同じ番地を\$A\$1というように指定する。相対番地では、セルの場所は当該セルとの相対的位置でとらえられているが、絶対番地では、セルそのものの場所としてとらえられている。

絶対番地がどのようなものか確認するためには、A1に1を入力し、A2に「=\$A\$1+1」を入力してから、A2をコピーしてA3からA10に貼り付けばいい。A2からA10には2が並ぶことになる（図7）。このときに、それぞれのセルに入力されている式を見てみると、A2からA10までずっと「=\$A\$1+1」が並んでいることがわかる（図8）。\$ \$をつけることで絶対的な番地指定になるのである。

列は絶対で行は相対といった番地指定（\$A1）、あるいは、列は相対で行は絶対といった番地指定（A\$1）もできる。これらの番地指定を駆使することによって、エクセルでの計算表作成の手間は大幅に軽減される。

図 7

	A
1	1
2	2
3	2
4	2
5	2
6	2
7	2
8	2
9	2
10	2

図 8

	A
1	1
2	= \$A\$1+1
3	= \$A\$1+1
4	= \$A\$1+1
5	= \$A\$1+1
6	= \$A\$1+1
7	= \$A\$1+1
8	= \$A\$1+1
9	= \$A\$1+1
10	= \$A\$1+1

2.6 九九の表の作成

相対番地と絶対番地の意味、ならびにそれらの番地指定の使い勝手を理解するためには「九九の表」を作成してみるといい。

A2からA10まで縦に1から9までの数字を入力し、B1からJ1まで横に1から9までの数字を入力する。この後、B2に「=\$A2*B\$1」を入力し、それをコピーして縦横に貼り付けると一挙に九九の表ができあがる（図9、10）。

図 9

	A	B	C	D	E	F	G	H	I	J
1		1	2	3	4	5	6	7	8	9
2	1	=A2*B\$1	=A2*C\$1	=A2*D\$1	=A2*E\$1	=A2*F\$1	=A2*G\$1	=A2*H\$1	=A2*I\$1	=A2*J\$1
3	2	=A3*B\$1	=A3*C\$1	=A3*D\$1	=A3*E\$1	=A3*F\$1	=A3*G\$1	=A3*H\$1	=A3*I\$1	=A3*J\$1
4	3	=A4*B\$1	=A4*C\$1	=A4*D\$1	=A4*E\$1	=A4*F\$1	=A4*G\$1	=A4*H\$1	=A4*I\$1	=A4*J\$1
5	4	=A5*B\$1	=A5*C\$1	=A5*D\$1	=A5*E\$1	=A5*F\$1	=A5*G\$1	=A5*H\$1	=A5*I\$1	=A5*J\$1
6	5	=A6*B\$1	=A6*C\$1	=A6*D\$1	=A6*E\$1	=A6*F\$1	=A6*G\$1	=A6*H\$1	=A6*I\$1	=A6*J\$1
7	6	=A7*B\$1	=A7*C\$1	=A7*D\$1	=A7*E\$1	=A7*F\$1	=A7*G\$1	=A7*H\$1	=A7*I\$1	=A7*J\$1
8	7	=A8*B\$1	=A8*C\$1	=A8*D\$1	=A8*E\$1	=A8*F\$1	=A8*G\$1	=A8*H\$1	=A8*I\$1	=A8*J\$1
9	8	=A9*B\$1	=A9*C\$1	=A9*D\$1	=A9*E\$1	=A9*F\$1	=A9*G\$1	=A9*H\$1	=A9*I\$1	=A9*J\$1
10	9	=A10*B\$1	=A10*C\$1	=A10*D\$1	=A10*E\$1	=A10*F\$1	=A10*G\$1	=A10*H\$1	=A10*I\$1	=A10*J\$1

図 10

	A	B	C	D	E	F	G	H	I	J
1		1	2	3	4	5	6	7	8	9
2	1	1	2	3	4	5	6	7	8	9
3	2	2	4	6	8	10	12	14	16	18
4	3	3	6	9	12	15	18	21	24	27
5	4	4	8	12	16	20	24	28	32	36
6	5	5	10	15	20	25	30	35	40	45
7	6	6	12	18	24	30	36	42	48	54
8	7	7	14	21	28	35	42	49	56	63
9	8	8	16	24	32	40	48	56	64	72
10	9	9	18	27	36	45	54	63	72	81

2.7 関数

エクセルでは合計値や平均値といったさまざまな関数が使えらる。関数を入力する際には、「=関数名(範囲)」という形でセルに入力する。範囲の指定はA1:A3というスタイルで行う(シート上で指定したい範囲をドラッグすると、番地は自動的に入力される)。合計を出す「=SUM()」や平均を出す「=AVERAGE()」などの関数はよく利用される(図 11、12)。

図 11

	A
1	100
2	200
3	300
4	=SUM(A1:A3)

図 12

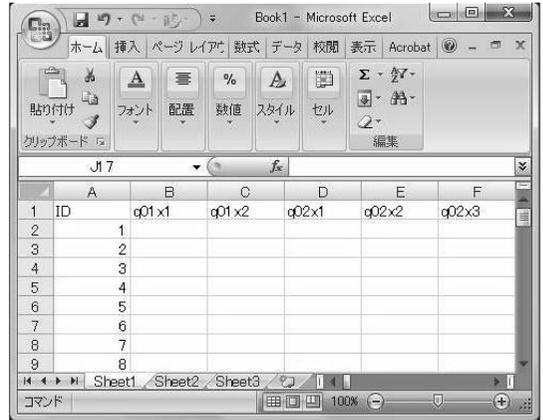
	A
1	100
2	200
3	300
4	600

使いたい関数の関数名がわからないときは「数式」タブにあるfxと書かれたボタンを押すと使える関数のリストがダイアログで示される(図 13)。その中から関数を選び、範囲を指定すると、セルに関数が入力される。

図 13



図 14



3 データ入力とチェックにおける利用法

3.1 データ入力における利用法

(1) セルの移動方向

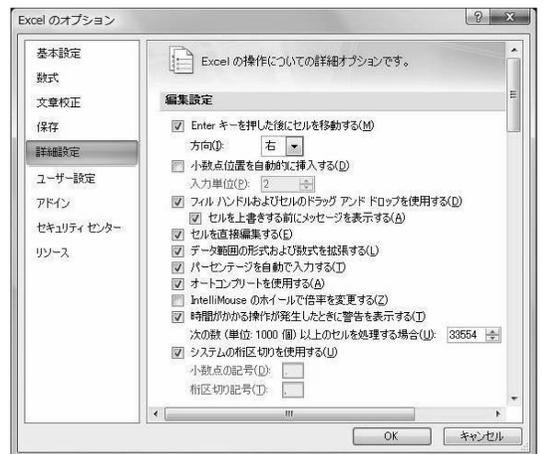
次に、データ入力と入力したデータのチェックにおけるエクセルの利用法について説明していくことにしよう¹⁾。

調査によって得られた情報を分析できるデータにするためには、まずそれをエクセルに入力していく必要がある。実際のデータ入力は、図 14 のようなシートに調査票の情報を打ち込んでいく作業となる。回収された調査票には通し番号が記入され、それをもとに個々の調査票が識別されるのだが、図で A 列に記されているのがその通し番号 (ID) である。B 列以降には調査票の各問につけられた変数名が記されている。

このような入力シートで入力を行う場合、Enter キーを押したときにセルが右に移動する方が都合がいい。しかし設定によっては Enter キーを押したときにセルが下に移動するようになっている場合もあるので、そのような場合は入力作業の前に設定を変更する必要がある。

そのためには、まず「ホーム」タブの左側にある丸いボタンをクリックする。次に「Excel のオプション」→「詳細設定」と進んでいくと図 15 のような画面になる。ここで、編集設定の「Enter キーを押した後にセルを移動する方向」を右に設定すればいい。

図 15



(2) ウィンドウ枠の固定

通常のエクセルの状態では入力作業をしていくと、図 16 のように A 列が表示されなくなり、今どの

ID のデータを入力しているのが画面からはわからなくなってしまう。

図 16

	Y	Z	AA	AB	AC	AD
1	q09s1 x1	q09s1 x2	q09s2x1	q09s2x2	q1 0s1 x1	q1 0s1 x2
2	6	9	F	9	6	9
3	5	9	C	9	6	9
4	7	9	A	9	7	9
5	2	9	D			
6						
7						

また、ある程度のケースまで入力が進むと、1 行目の変数名が画面から表示されなくなってしまう (図 17)。

図 17

	A	B	C	D	E	F
51	50	2	19	社会	メディア	2010
52	51	2	19	社会	社会福祉	2010
53	52	2	19	社会	社会福祉	2010
54	53	2	19	社会	メディア	2010
55	54	1	22			
56						
57						

このようにデータが指定されたセルにきちんと入力されているかが画面上で確認できない状態で作業を進めると、入力ミスにつながる危険性がある。そこで入力作業に入る前に、入力用シートの ID と変数名が常に画面に表示されるように設定しておく必要がある。

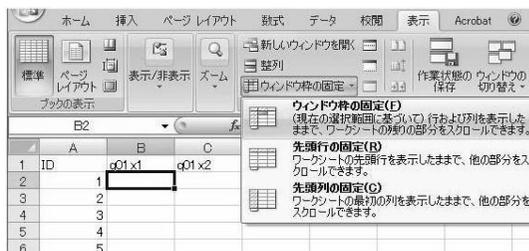
そのためにはまず、常に表示させたい行の 1 行下、列の 1 列右のセルを選択し、「表示」タブをクリックする。

図 18

	A	B	C	D	E	F
1	ID	q01 x1	q01 x2	q02x1	q02x2	q02x3
2	1					
3	2					
4	3					
5	4					
6	5					

次に「ウィンドウ」の欄にある「ウィンドウ枠の固定」をクリックし、もう一度「ウィンドウ枠の固定」を選択する (図 19)。

図 19



そうすると、固定したい行列の内側に黒い線が引かれ、ID と変数名が常に表示されるようになる (図 20)。

図 20

	A	B	C	D	E	F
1	ID	q01 x1	q01 x2	q02x1	q02x2	q02x3
2	1					
3	2					
4	3					
5	4					
6	5					

以上のような準備作業を済ませたら、あとは列と変数の対応関係に注意しながら、2 行目に ID1 のケースのデータ、3 行目に ID2 のケースのデータといった具合にデータを入力していけばいい。

3.2 データチェックにおける利用法

(1) シート間の演算

入力の際、起こりやすいミスには数字の読み違いやキーの打ち間違いなどがある。こうしたミスを防ぐためには、データを 2 度入力し、入力されたデータが同一であるかを照合するといいい。2 度入力されたデータを照合するためにはシート間での演算を利用することになる。

エクセルでは1枚のシートの中だけでなく、異なるシート間でも演算を行うことができる。1枚のシートの中で演算を行う場合は、「=A1+A2」「=B1/B2」といった式を入力する。異なるシートのセルについて演算を行うときには、セル番地の前に「シートの名前!」をつける。たとえば、下図のSheet1のB2(100)とSheet2のB2(500)の足し算の結果をSheet3に出すには、Sheet3のセルを選択した状態で「=Sheet1!B2+Sheet2!B2」と入力すればいい(図21、22、23)。

この式をコピーしてSheet3のB3~B5に貼り付ければ、そこにはB3、B4、B5についてのSheet1とSheet2の足し算の結果が表示される(図24)。

図 21

	A	B	C
1			
2		100	
3		200	
4		300	
5		400	
6			

図 22

	A	B	C
1			
2		500	
3		600	
4		700	
5		800	
6			

図 23

B2		=Sheet1!B2+Sheet2!B2				
	A	B	C	D	E	F
1						
2		600				
3						
4						
5						
6						

図 24

B3		=Sheet1!B3+Sheet2!B3				
	A	B	C	D	E	F
1						
2		600				
3		800				
4		1000				
5		1200				
6						

(2) シート間演算によるデータチェック

シート間の演算によってデータチェックを行う方法は以下の通りである。

まず、全回答者分のデータをエクセルの1枚のシートに入力し、次に別のシートにもう一度全回答者分のデータを入力する。図25および図26は1度目と2度目の入力結果を示したものである(Sheet1:1度目の入力、Sheet2:2度目の入力)。

図 25

	A	B	C	D	E	F
1	ID	q01 x1	q01 x2	q02x1	q02x2	q02x3
2	1	2	18	社会	社会福祉	9999
3	2	2	20	社会	社会	2008
4	3	2	20	社会	メディア	2008
5	4	2	20	社会	社会	2007
6	5	1	21	社会	産業関係	2008
7	6	2	20	社会	社会	2009
8	7	2	20	社会	社会	2008
9	8	1	20	社	社	2010

図 26

	A	B	C	D	E	F
1	ID	q01 x1	q01 x2	q02x1	q02x2	q02x3
2	1	2	18	社会	社会福祉	9999
3	2	2	20	社会	社会	2009
4	3	2	20	社会	メディア	2008
5	4	2	22	社会	社会	2009
6	5	1	21	社会	産業関係	2008
7	6	2	20	社会	社会	2009
8	7	2	20	社会	社会	2008
9	8	1	20	社	社	2010

2度の入力が済んだら、別のシート(Sheet3)において、2つのシートの同じセル番地に同じ値が入力されているかを照合する。たとえばSheet1と

Sheet2 の B2 に同じ値が入力されているかを照合するためには、Sheet3 の B2 に「=IF(Sheet1!B2=Sheet2!B2, "OK", "NG")」という式を入力すればいい。この式は「Sheet1 の B2 と Sheet2 の B2 が同じであれば『OK』を、異なれば『NG』を表示せよ」という意味である。この式をコピーし、シート全体に渡ってそれを貼り付ければ、データ全体でどこに入力ミスが生じているのかわかるようになる。

図 27

	A	B	C	D	E	F
1	ID	q01x1	q01x2	q02x1	q02x2	q02x3
2	1	OK	OK	OK	OK	OK
3	2	OK	OK	OK	OK	OK
4	3	OK	OK	OK	OK	OK
5	4	OK	NG	OK	OK	NG
6	5	OK	OK	OK	OK	OK
7	6	OK	OK	OK	OK	OK
8	7	OK	OK	OK	OK	OK
9	8	OK	OK	OK	OK	OK

図 27 は結果を示したものであるが、C5 と F5 のセルで入力ミスが起きていることがわかる。「NG」のセルがどこにあるかをわかりやすくするには、「ホーム」タブのスタイルの欄にある「条件付き書式」を用いて、Sheet3 全体で「NG」のセルに色をつけるようにするといいだろう。

データの 2 回入力については、1 人で 2 度の入力を行うよりも、2 人 1 組で 1 度ずつ入力するほうがより効果的である。

3.3 ロジカルエラーチェックにおける利用法

(1) オートフィルタ機能

回収された調査票にはロジカルエラーが含まれていることがある。ロジカルエラーというのは、30 歳なのに「自分は 50 年その町に住んでいる」と答えていたり、子どもがいないのに子どもの年齢を答えていたりする場合のように、回答に論理的な矛盾が生じていることをいう。

こうしたロジカルエラーがどこで生じているかを発見するためには、エクセルのオートフィルタ機能が役に立つ。オートフィルタとは、指定した条件に合うデータを抽出するための機能である。

オートフィルタを設定するためには、変数名を入力している行を選択した後、「データ」タブをクリックし、「並べ替えとフィルタ」という欄から「フィルタ」を選択する。そうすると、1 行目に▼印が表示される（図 28）。

図 28

	A	B	C	D	E	F
1	ID	▼ q01x1	▼ q01x2	▼ q02x1	▼ q02x2	▼ q02x3
2	1	2	18	社会	社会福祉	9999
3	2	2	20	社会	社会	2009
4	3	2	20	社会	メディア	2008
5	4	2	22	社会	社会	2009
6	5	1	21	社会	産業関係	2008
7	6	2	20	社会	社会	2009
8	7	2	20	社会	社会	2008
9	8	1	20	社	社	2010

この▼をクリックすると、その列にどのような値が入力されているかが表示される。q01x1 (B1 のセル) の▼をクリックするとエクセルの表示は図 29 のようになり、ここから q01x1 という変数は 1 と 2 の値をとることがわかる。

図 29

	A	B	C	D	E	F
1	ID	▼ q01x1	▼ q01x2	▼ q02x1	▼ q02x2	▼ q02x3
2	昇順(S)			18	社会	社会福祉
3	降順(O)			20	社会	社会
4	色で並べ替え(I)			20	社会	メディア
5	"q01x1" からフィルタをクリア(C)			22	社会	社会
6	色フィルタ(F)			21	社会	産業関係
7	数値フィルタ(E)			20	社会	社会
8				20	社会	社会
9				20	社	社
10						2010

値の左側にはチェックがついているが、このチェックを1だけにつければ、q01x1が1のデータだけが画面に表示されるようになる。

フィルタは1つだけでなくいくつもかけることができる。たとえばq01x1で1にチェックし、さらにq01x2で20にチェックをすると、q01x1が1でq01x2が20のデータだけが表示されるようになる(図30)。フィルタを解除してすべてのデータを表示するには「すべて選択」にチェックをつければいい。

図 30

	A	B	C	D	E	F
1	ID	q01x1	q01x2	q02x1	q02x2	q02x3
9	8	1	20	社	社	2010
38	37	1	20	社会	産業関係	2010
58	57	1	20	社会	産業関係	2010
61	60	1	20	社会	メディア	2010
66	65	1	20	社会	メディア	2010
67						
68						

(2) オートフィルタ機能を用いたロジカルエラーのチェック

オートフィルタ機能を使ってロジカルエラーをチェックする方法について、図31のような簡単なデータを例に説明しよう。

図 31

	A	B	C
1	ID	q01(アルバイト経験)	q02(アルバイトの時給)
2	1	1	920
3	2	2	99998
4	3	2	99998
5	4	1	850
6	5	1	880
7	6	1	900
8	7	1	1000
9	8	1	820

q01はこれまでのアルバイト経験についての変数であり、アルバイトをしたことがある場合は「1」、アルバイトしたことがない場合は「2」がB列に入力されている。q02はこれまでにしたアルバイ

トの中でもっとも高い時給についての変数であり、C列にその金額が入力されている。調査ではq01でアルバイトの経験がないと回答した人はq02に回答しないように指示されており、アルバイト経験がなくq02に回答がなかった場合、C列には「99998」(非該当を示す値)が入力されている。

図31の例では、q01でアルバイト経験なしと回答しているのにもかかわらず、q02でアルバイトの時給を回答している場合、ロジカルエラーが生じていることになる。このような回答者を発見するためには、フィルタをq01にかけ、2(アルバイト経験なし)のものだけを表示すればいい(図32)。

図 32

	A	B	C
1	ID	q01(アルバイト経験)	q02(アルバイトの時給)
3	2	2	99998
4	3	2	99998
10	9	2	99998
14	13	2	99998
16	15	2	99998
17	16	2	99998
20	19	2	99998
21	20	2	950

すると、ID20のケースがアルバイト経験なしと回答しているにもかかわらず時給を950円と回答していることがわかる。以上のようにオートフィルタ機能を利用しながらロジカルエラーを発見し、適切に修正していくことで、データはより質の高いものとなる。

なお、ロジカルエラーの修正は「入力ミス修正済みデータファイル」をそのまま用いるのではなく、これをコピーした新しいファイルで行う方がいい。すべてのロジカルエラーが適切に処理されたら「ロジカルエラー修正済みデータファイル」として保存しておくようにしよう。

4 分析における利用法

4.1 平均と等平均検定

エクセルには、統計分析に役に立つ関数がいくつもあり、SPSS などの汎用統計ソフトを使わなくてもかなりの分析はできる。また、ピボットテーブルを使って簡単に集計表を作ることもできる。ここからは、エクセルの関数を使った分析法と検定方法およびピボットテーブルを使った集計表の作り方について説明する。

まず、エクセルを使って平均の表を作る方法と等平均検定 (t 検定) の方法について説明しよう。

図 33 の左にある表には、アルバイトをしている人 (以下「バイト有り」として) とアルバイトをしていない人 (以下「バイト無し」として) の 1 ヶ月に自由に使えるお金が示されている。この表から平均の表を作っていこう。

平均は数式を使って計算することもできるが、AVERAGE 関数を使うとより簡単に求めることができる。AVERAGE 関数の入力は「=AVERAGE(範囲)」になる。たとえば、図においてバイト有りグループの平均を求めるときは「=AVERAGE(B2:B10)」と入力すればいい。

また、度数を求めるときは COUNT 関数を、標準偏差を求めるときは STDEV 関数を使えばいい。COUNT 関数はセルに「=COUNT(範囲)」と入力し、STDEV 関数の入力は「STDEV(範囲)」とすればいい。したがって、バイト有りグループの度数と標準偏差は、それぞれのセルに「=COUNT(B2:B10)」、「=STDEV(B2:B10)」と入力すれば求められる。

こうして作った図 33 の右にある平均の表では、バイト有りの人がバイト無しの人より 1 ヶ月に自由に使えるお金の平均金額が高くなっている。このことが母集団でもいえるかどうかを確認するためには、等平均検定を行わなければならない。

等平均検定では一般的には t 値を計算する必要があるが、ここでは TTEST 関数を使って t 値を計算せずに直接データから有意確率を求める方法を紹介する。TTEST 関数には 4 つの引数を指定する必要がある。具体的な入力は「=TTEST(グループ 1 の範囲, グループ 2 の範囲, 検定の指定, 検定の種類)」になる。

「グループ 1 の範囲」と「グループ 2 の範囲」は、それぞれのグループの値が入力されているセ

図 33

	A	B	C	D	E	F	G
1	アルバイトの有無(1:有、2無)	自由に使えるお金					
2	1	20000					
3	1	120000					
4	1	50000					
5	1	30000					
6	1	50000					
7	1	70000					
8	1	20000					
9	1	5000					
10	1	20000					
11	2	50000					
12	2	20000					
13	2	30000					
14	2	20000					
15	2	5000					
16	2	20000					

アルバイトの有無			
	有	無	合計
平均	42777.778	24166.667	35333.333
度数			15
標準偏差	35276.684	14972.196	29668.807
有意確率			

関数入力例:

- =AVERAGE(①)
- =AVERAGE(②)
- =COUNT(①)
- =COUNT(②)
- =SDTEV(①)
- =SDTEV(②)

図 34

	A	B	C	D	E	F	G
1	アルバイトの有無(1:有、2無)	自由に使えるお金					
2	1	20000			アルバイトの有無		
3	1	120000			有	無	合計
4	1	50000	①	平均	42777.778	24166.667	35333.333
5	1	30000		度数	9	6	15
6	1	50000		標準偏差	35276.684	14972.196	29668.807
7	1	70000		有意確率	0.248		
8	1	20000			=TTEST(①,②,2,2)		
9	1	5000					
10	1	20000					
11	2	50000					
12	2	20000					
13	2	30000	②				
14	2	20000					
15	2	5000					
16	2	20000					

ルの範囲である。「検定の指定」は、確率の計算に片側だけの分布を使うか、それとも両側の分布を使うかを指定する引数であり、片側ならば 1、両側ならば 2 となる。「検定の種類」については、1 を入力すると対をなすデータ（ケース数が同じ場合）の検定が行われ、2 を入力すると等分散を仮定した t 検定が行われる。3 は等分散を仮定しない t 検定（ウェルチの検定）を行うときに入力する。

t 検定のなかで最もよく使われるのは、等分散を仮定した両側検定である。この方法でバイト有りグループとバイト無しグループの検定を行うには、セルに「TTEST(バイト有りグループの範囲, バイト無しグループの範囲, 2, 2)」と入力すればいい。たとえば、図 34 では「=TTEST(B2:B10, B11: B16, 2, 2)」という式を入力している。

この関数から返された有意確率は 0.248 である。有意水準を 0.05 にすると、求められた有意確率が有意水準より大きいので、「母集団において、バイト有りの人とバイト無しの人のお金の平均には差があるとはいえない」という結果となっている。

4.2 相関分析と無相関検定

次に、相関分析と無相関検定の方法について述べる。まず相関分析の方法について。エクセルを使うと、相関係数も関数を使って簡単に求めることができる。相関係数を求めるときに使う関数は CORREL である。セルには「=CORREL(グループ 1 の範囲, グループ 2 の範囲)」と入力する。

図 35 の左にある表は、地位・収入・競争に勝つこと・他人につくすこと・やりがいのある仕事を見つけることが、重要であるかどうかを 5 段階で問うた項目のデータである。値は 1「重要である」、2「やや重要である」、3「どちらともいえない」、4「あまり重要ではない」、5「重要ではない」となっている。このデータから、CORREL 関数を使って作った相関係数の表が、図 35 の右にある表である。たとえば、「競争」と「やりがい」の相関係数である -0.127 は、セルに「=CORREL(C2:C16, E2: E16)」と入力すると求められる。

では、この表に示された相関係数が母集団でも有意であるかどうかを、無相関検定を行って確認してみよう。無相関検定を行うには、検定統計量を算出し、t 分布の関数を使って有意確率を求めればよい。無相関検定に用いられる検定統計量の

図 35

	A	B	C	D	E	F	G	H	I	J	K	L
1	地位	収入	競争	つくす	やりがい		相関係数	地位	収入	競争	つくす	やりがい
2	2	2	4	2	2		地位					
3	2	1	2	1	2		収入	0.517				
4	1	1	2	①	2	②	競争	0.709	0.637			
5	3	2	4	2	1		つくす	0.121	-0.562	-0.062		
6	2	3	2	1	1		やりがい	-0.062	0.015	-0.127	-0.044	
7	2	2	2	2	1							
8	2	1	2	3	3							
9	4	1	3	4	1							
10	5	2	5	3	1							
11	5	5	5	1	3							
12	4	5	5	1	1							
13	4	2	2	2	2							
14	3	2	3	1	1							
15	2	2	3	2	1							
16	2	2	2	2	2							
17												

図 36

	A	B	C	D	E	F	G	H	I	J	K	L
1	地位	収入	競争	つくす	やりがい		相関係数	地位	収入	競争	つくす	やりがい
2	2	2	4	2	2		地位					
3	2	1	2	1	2		収入	0.517				
4	1	1	2	2	2		競争	0.709	0.637			
5	3	2	4	2	1		つくす	0.121	-0.562	-0.062		
6	2	3	2	1	1		やりがい	-0.062	0.015	-0.127	-0.044	
7	2	2	2	2	1							
8	2	1	2	3	3		度数	15				
9	4	1	3	4	1							
10	5	2	5	3	1		T	地位	収入	競争	つくす	やりがい
11	5	5	5	1	3		地位					
12	4	5	5	1	1		収入	2.176				
13	4	2	2	2	2		競争	3.630	2.982			
14	3	2	3	1	1		つくす	0.440	-2.452	-0.223		
15	2	2	3	2	1		やりがい	-0.225	0.055	-0.461	-0.158	
16	2	2	2	2	2							
17							有意確率	地位	収入	競争	つくす	やりがい
18							地位					
19							収入	0.049				
20							競争	0.009	0.011			
21							つくす	0.667	0.029	0.827		
22							やりがい	0.826	0.957	0.652	0.877	

式は、次のようなものである。

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

(ただし、R は相関係数、n は標本のサイズ)

この検定統計量 T は、母集団において相関が 0 のとき、自由度 n-2 の t 分布に従う。検定統計量

の式は、エクセルでは「=相関係数*SQRT(度数-2)/SQRT(1-相関係数^2)」になる。SQRT は平方根を求めるときに使われる関数である。

この式を使って作成されたものが、図 36 の右にある「T」の表である。ここでは、収入と地位の相関に関する検定統計量 T(2.176)は、「=H3*SQRT(SH\$8-2)/SQRT(1-H3^2)」で求められている。絶対番地を利用しているので、このセルをコピーして

それぞれのセルに貼り付ければ他の相関係数についての検定統計量 T も簡単に算出できる。

いよいよ有意確率の計算である。有意確率の計算には、TDIST 関数を使う。TDIST 関数は、「=TDIST(ABS(t 値),自由度,分布の指定)」というかたちで用いる。TDIST 関数に入力する t 値はマイナスの値が指定できないので、絶対値を返す ABS 関数を使って t 値が常にプラスになるようにする。

上で算出された T に対応する有意確率を求めるときには、「=TDIST(ABS(T),自由度,分布の指定)」とすればよい。また、無相関検定の検定統計量 T は自由度 n-2 の t 分布に従うので、「自由度」には度数から 2 を引いた値を入力する。「分布の指定」には、片側の確率を求めるときは 1、両側の確率を求めるときは 2 を入力する。一般的には両側検定を行うので 2 を入力すればいい。

この式を使って作成されたものが、図 36 の右にある「有意確率」の表である。ここでは、収入と地位の相関に関する有意確率 (0.049) は、「=TDIST(ABS(H12),SH\$8-2,2)」で求められている。絶対番地を利用しているので、このセルをコピーしてそれぞれのセルに貼り付ければ他の相関係数についての有意確率も簡単に算出できる。

収入と地位の相関に関する有意確率は 0.049 であり、有意水準を 0.05 にすると、求められた有意確率は有意水準より小さい。したがって、「母集団において、収入と地位の間に相関がある」という結果となっている。

4.3 ピボットテーブルと集計表

(1) 準備

エクセルにはピボットテーブルという機能がある。この機能を使うと、データから簡単に集計表を作ることができる。ここからは、このピボットテーブルについて説明する。

図 37

	A	B	C	D
1	カウンタ	d	c1	c2
2	1	1	2	22
3	1	2	1	22
4	1	3	1	20
5	1	4	2	21
6	1	5	2	21
7	1	6	2	21
8	1	7	1	21
9	1	8	1	20
10	1	9	2	20
11	1	10	1	21
12	1	11	2	20
13	1	12	2	21
14	1	13	2	21
15	1	14	1	21
16	1	15	1	23

通常のデータシートからピボットテーブルを使って集計表を作るためには、ちょっとした準備をしておくことと便利である。まず図 37 のようにデータシートの一列目の列に新しい列を挿入し、その列の 1 行目に「カウンタ」と入力する。2 行目からサンプルの最後の行には 1 を入力しておく。このカウンタの列は、集計表の度数をカウントするのに使われるものである。

図 38



まず、シート全体、または集計に用いる変数が

含まれる列の範囲を選択し、「挿入」タブの左にある「ピボットテーブル」をクリックする(図 38)。こうすると、図 39 のダイアログが現れる。

ダイアログで、ピボットテーブルを配置する場所を新規ワークシートにしよう。「OK」をクリック

すると新しいシートに図 40 のような画面が現れる。図 40 の右にあるフィールドリストと 4 つのボックスを使って表を作ることになる。

図 39

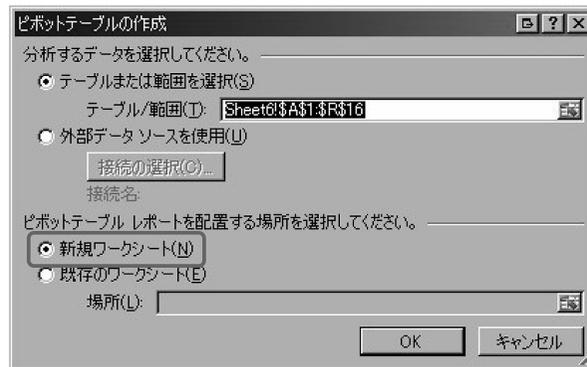


図 40

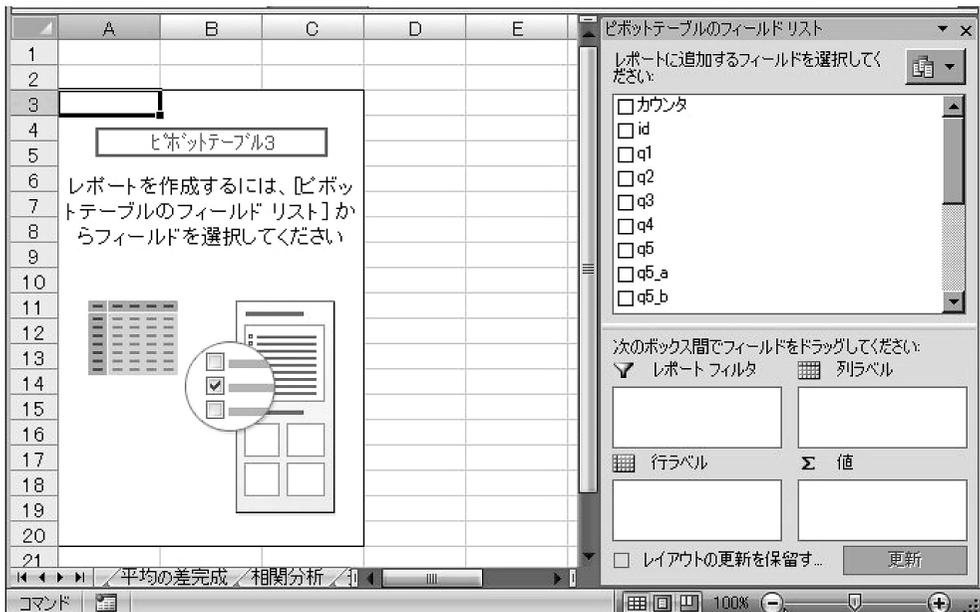
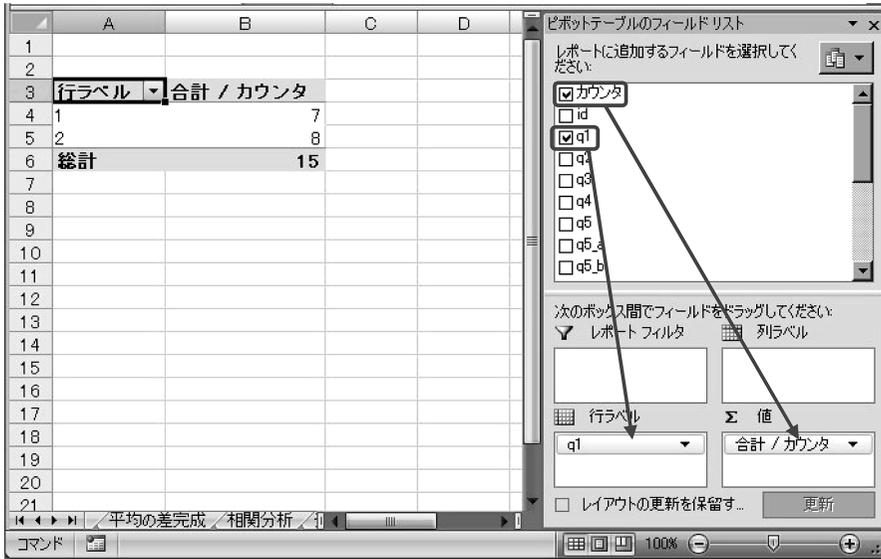


図 41



(2) 度数分布表

まず性別の度数分布表から作ってみよう。ここで使っているデータでは、性別の変数名は「q1」となっており、値は1が「男性」、2が「女性」となっている。

フィールドリストから「カウンタ」を下にある「値」ボックスにドラッグする。それから「q1」（性別）を「行ラベル」ボックスにドラッグする。すると、図 41 の左のような表が作られる。男性の度数が7であり、女性の度数が8となっている。

度数分布表には、一般的に度数だけではなくパーセントも入るので、ここでもパーセントを入れてみよう。まず、フィールドリストから「カウンタ」を「値」ボックスにもう一度ドラッグする。すると「カウンタ」のフィールド名が「カウンタ2」に自動的に変わって「値」ボックスに入る（図 42）。

図 42

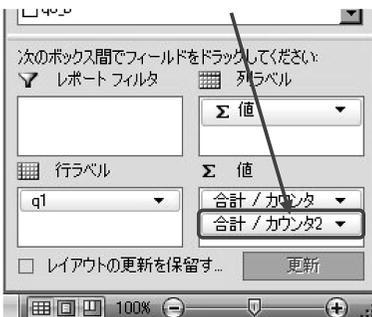
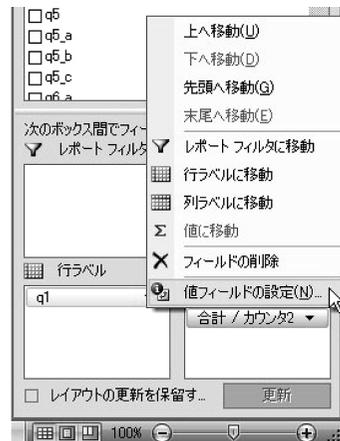


図 43



この「カウンタ 2」をクリックすると図 43 のようにメニューが現れる。このメニューの一番下にある「値フィールドの設定」を選択すると、図 44 のようなダイアログが現れる。

図 44



このダイアログにある「計算の種類」タブを選択し、計算の種類を「全体に対する比率」に変え

る。「OK」をクリックすると、パーセントがセルに入力され、図 45 のような度数分布表が作られる。

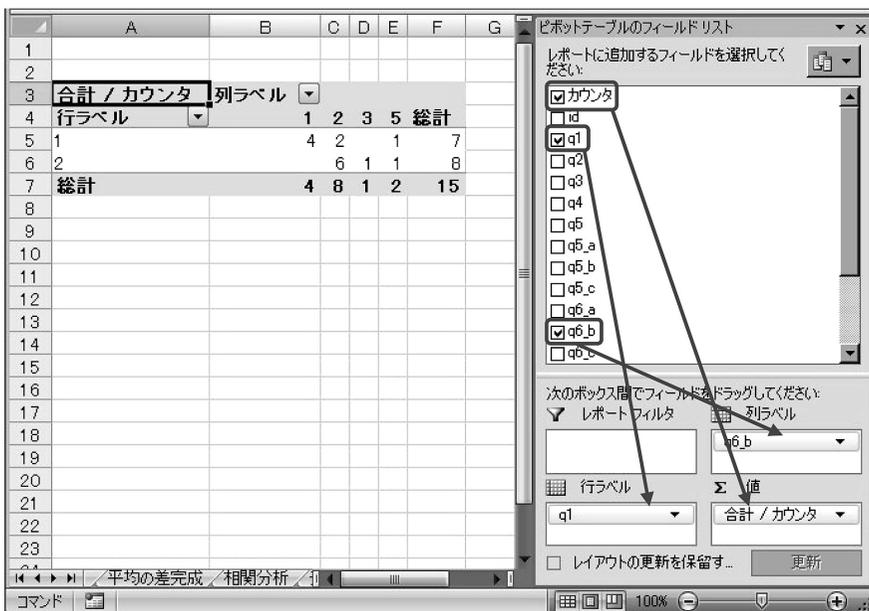
図 45

	A	B	C
1			
2			
3			値
4	行ラベル	合計 / カウンタ	合計 / カウンタ2
5	1	7	46.67%
6	2	8	53.33%
7	総計	15	100.00%

(3) クロス表

次に、クロス表について。性別 (q1) と高い収入を重視すること (以下「高収入志向」と) のクロス表を作ってみよう。「高収入志向」の変数名は「q6_b」であり、回答は 1「重要である」、2「やや重要である」、3「どちらともいえない」、4「あまり重要ではない」、5「重要ではない」となっている。

図 46



まず、フィールドリストから「カウンタ」を「値」のボックスにドラッグする。次に、性別の「q1」を「行ラベル」ボックスに、高収入志向の「q6_b」を「列ラベル」ボックスにドラッグする。こうすると図 46 のように、表側が性別、表頭が高収入志向となるクロス表ができあがる。図 46 のクロス表において値が入っていないセルは、その度数がゼロであることを表す。また表頭に「4」が示されていないのは、高収入志向の項目において 4「あまり重要ではない」と回答したケースがないからである。

4.4 クロス表のカイ 2 乗検定

次に、ピボットテーブルで作成されたクロス表を用いて独立性の検定を行う方法について説明する。クロス表の独立性の検定を行うためには、まずクロス表の期待度数を計算しなければならない。期待度数 は次の式で計算することができる。

$$F_{ij} = \frac{F_{i\bullet} \times F_{\bullet j}}{n}$$

(ただし、 $F_{i\bullet}$ = i 行目の周辺度数、 $F_{\bullet j}$ = j 列目の周辺度数、 n = 標本の大きさ)

図 47 の「観測度数」の表は、先にピボットテーブルを使って作った性別と高収入志向のクロス表である。「期待度数」の表は、「観測度数」の表の周辺度数から作られている。「男性」のなかで「重要である」と回答した人の期待度数は、「=BS\$9*\$F\$7/\$F\$9」と入力すれば求められる。

クロス表の独立性の検定には χ^2 分布が使われるが、観測度数と期待度数がわかれば、 χ^2 値を計算しなくても CHITEST 関数を使って有意確率を求めることができる。CHITEST 関数は「=CHITEST(観測度数範囲,期待度数範囲)」というかたちである。

図 48 の「有意確率」では、「=CHITEST(B2:E3,B7:E8)」と入力している。有意水準を 0.05 にすると、有意確率の値 0.073 は有意水準より大きいので、「母集団においては性別と高収入志向は関係があるといえない」という結果となっている。

図 47

	A	B	C	D	E	F
1	観測度数 f_{ij}	重要 である	やや重要 である	どちらとも いえない	重要 ではない	合計
2	男性	4	2	0	1	7
3	女性	0	6	1	1	8
4	合計	4	8	1	2	15
5						
6	期待度数 F_{ij}	重要 である	やや重要 である	どちらとも いえない	重要 ではない	合計
7	男性	1.866667	3.733333	0.466667	0.933333	7 ②
8	女性	2.133333	4.266667	0.533333	1.066667	8
9	合計	4 ①	8	1	2	15 ③

図 48

	A	B	C	D	E	F
1	観測度数 f_{ij}	重要 である	やや重要 である	どちらとも いえない	重要 ではない	合計
2	男性	4	2	0	1	7
3	女性	0	6	1	1	① 8
4	合計	4	8	1	2	15
5						
6	期待度数 F_{ij}	重要 である	やや重要 である	どちらとも いえない	重要 ではない	合計
7	男性	1.866667	3.733333	0.466667	0.933333	7
8	女性	2.133333	4.266667	0.533333	1.066667	② 8
9	合計	4	8	1	2	15
10						
11	有意確率	0.073045	=CHITEST(①,②)			

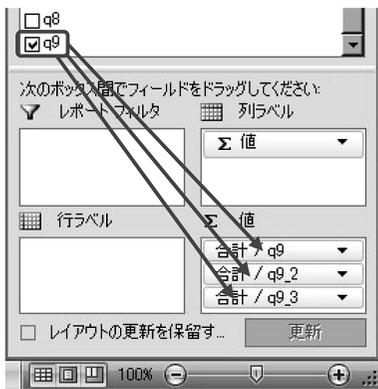
4.5 補足

(1) ピボットテーブルの多様な使用法

ピボットテーブルは、先に紹介した他にも様々な利用可能である。1ヶ月に自由に使えるお金の平均、標準偏差、度数の表を作ってみよう。

1ヶ月に自由に使えるお金の変数名は「q9」である。フィールドリストからこの「q9」を値のボックスに3回ドラッグして入れる(図49)。

図 49



次に、「値」ボックスの「合計 / q9」をクリックし、メニューの「値フィールドの設定」を選択する(図50)。そうすると、図51のダイアログが開く。

図 50

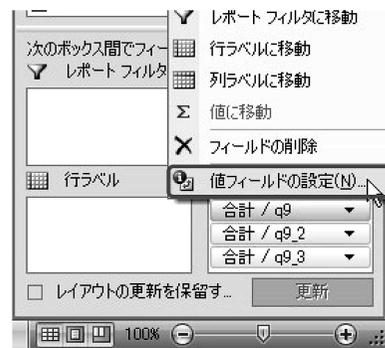
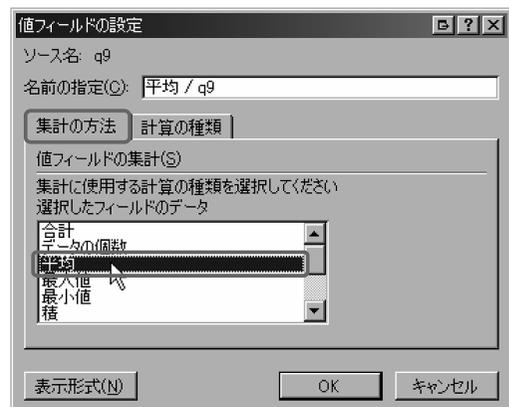


図 51



ダイアログの「集計の方法」タブで、値フィー

ルドの集計から「平均」を選択し「OK」をクリックする。同様に、「合計 / q9_2」については「標準偏差」を選び、「合計 / q9_3」については「データの個数」を選ぶ。こうすると、平均、標準偏差、度数が入った表が作られる (図 52)。

図 52

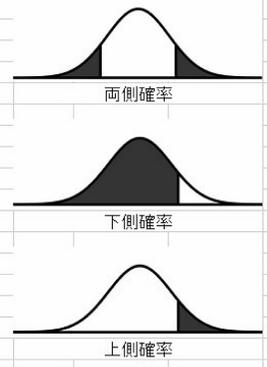
	A	B	C
1			
2			
3	値		
4	平均 / σ	標準偏差 / σ_2	数値の個数 / n_3
5	35.33333333	28.66279044	15
6			

(2) 分析に用いる様々な関数

これまでみてきたように、関数を用いることで作業ははるかに楽になる。図 53 に示しているのは、社会学においてよく用いられる関数とその使用例である。

図 53

	A	B	C	D	E	F	G	H	I
1	関数	引数	出力	式	結果		行番号	データX	データY
2	count	配列	セルの個数	COUNT(H2:H11)	10		2	1	3
3	sum	配列	合計	SUM(H2:H11)	59		3	6	5
4	sumsq	配列	平方和	SUMSQ(H2:H11)	421		4	9	8
5	average	配列	平均	AVERAGE(H2:H11)	5.9		5	4	2
6	median	配列	中央値	MEDIAN(H2:H11)	6.5		6	8	9
7	mode	配列	最頻値	MODE(H2:H11)	9		7	8	9
8	devsq	配列	偏差平方和	DEVSQ(H2:H11)	72.9		8	7	1
9	var	配列	不偏分散	VAR(H2:H11)	8.1		9	2	5
10	varp	配列	分散	VARP(H2:H11)	7.29		10	9	6
11	stdev	配列	母標準偏差推定値	STDEV(H2:H11)	2.84605		11	5	4
12	stdevp	配列	標準偏差	STDEVP(H2:H11)	2.7				
13	percentile	配列, 率	パーセンタイル点	PERCENTILE(H2:H11, 0.25)	4.25				
14	standardize	配列, 平均, 標準偏差	標準得点	STANDARDIZE(H2, E5, E12)	-1.81481				
15	rank	数値, 配列, 昇順のとき1	順位	RANK(H2, H2:H11, 1)	1				
16	correl	配列, 配列	相関係数	CORREL(H2:H11, I2:I11)	0.570265				
17	slope	配列Y, 配列X	単回帰係数	SLOPE(I2:I11, H2:H11)	0.565158				
18	intercept	配列Y, 配列X	切片	INTERCEPT(I2:I11, H2:H11)	1.865569				
19	rsq	配列Y, 配列X	決定係数	RSQ(I2:I11, H2:H11)	0.325203				
20	normsdist	z値	z値の下側確率	NORMSDIST(1.96)	0.975002				
21	normsinv	z値の下側確率	z値	NORMSINV(0.975)	1.959964				
22	zttest	配列, 母平均, 母分散	z値の上側確率	ZTEST(H2:H11, 5, 5)	0.284607				
23	tdist	t値, 自由度, 両側のとき2	t値の両側確率	TDIST(1.96, 100, 2)	0.052779				
24	tinvt	t値の両側確率, 自由度	t値	TINV(0.95, 100)	0.062864				
25	tttest	配列, 配列, 両側のとき2, 等分散のとき2	t値の両側確率	TTEST(H2:H11, I2:I11, 2, 2)	0.587442				
26	chidist	χ^2 値, 自由度	χ^2 値の上側確率	CHIDIST(18.3, 10)	0.050109				
27	chiinv	χ^2 の上側確率, 自由度	χ^2 値	CHIINV(0.05, 10)	18.30704				
28	chitest	実測値範囲, 期待値範囲	χ^2 上側確率						
29	fdist	F値, 自由度1, 自由度2	F値の上側確率	FDIST(2.46, 20, 20)	0.025225				
30	finvt	F値の上側確率, 自由度1, 自由度2	F値	FINV(0.025, 20, 20)	2.464484				
31	fttest	配列, 配列	F値の両側確率	FTEST(H2:H11, I2:I11)	0.979056				



5 おわりに

本稿で解説してきたのは社会調査における一般的な利用法であるが、エクセルの機能を十分に活用すれば、さらに色々な作業を行うことができる。たとえば、データ入力とチェックを行う際には、マクロ機能を使えばより効率的に作業をすすめられる。また、主成分分析などのより複雑な分析も、フリーのアドインなどを利用すればエクセルで行うことができる。本稿で紹介した利用法をひとつおりに身につけた後は、他の文献も参考にしつつ別の利用法を試してみるとよいだろう。

【文献】

- 小林久高, 1998『社会情報演習 コンピュータ入門編』今井書店。
小林久高, 2007「ベクトルと行列の基礎」『同志社社会学研究』11: 45-66。
小林久高・雨森聡・山本圭三, 2008「社会調査データの入力とチェックの方法」『同志社社会学研究』12: 41-50。
小林久高・猿渡壮・山本圭三, 2010「エクセル入門」小林久高編『同志社大学社会調査実習報告書 18 2009年度大学生社会意識調査』257-79。

【付記】

本稿の作成にあたり、雨森聡氏（島根大学）にはさまざまなアイデアを提供していただいた。ここで感謝の意を表したい。

【執筆者紹介】

小林 久高

同志社大学社会学部社会学科 教授

hkobayas@mail.doshisha.ac.jp

山本 圭三

同志社大学大学院社会学研究科社会学専攻 博士後期課程

yamamoke@petite-lapine.com

金 政芸

同志社大学大学院社会学研究科社会学専攻 博士後期課程

freyr80@gmail.com

猿渡 壮

同志社大学大学院社会学研究科社会学専攻 博士前期課程

se_saruwatari_fo@msn.com

【注】

- 1) 本稿のデータの入力とチェックに関しては、小林・雨森・山本（2008）における説明をもとにしている。なお、前稿では旧バージョンのエクセルを使用していたが、本稿ではエクセル 2007 を使用した上で、前稿ではあまり触れていなかったエクセルの諸機能の使用方法についても詳しく説明していくことにする。