

# Excel によるアンケート集計と独立性の検定

寺脇 拓

## 1. データ入力前の作業(別紙 2 参照)

以下では、「大阪人の食生活に関するアンケート調査」のデータを用いて、Excel 上でアンケート調査データを分析するための方法について説明する。

まず、データ入力を始める前に、次の準備が必要である。

1. 各アンケート調査票をナンバリングする。
2. 各質問にアルファベットあるいは英数字で変数名をつける(例:年齢 age)。ナンバリングした番号は、通常 Code、Number、Num といった変数名がつけられる。
3. 複数回答は、全ての選択肢をそれぞれ一つの質問とみなし、それら一つ一つに変数名をつける(例:resist1、resist2、...)。順位回答も同様である。

## 2. 入力用ワークシートの作成(pre.xls 参照)

入力準備ができたなら、入力用ワークシートを作る。

1. ワークシートの一行目に、変数名を入力する。一列目はナンバーとする。
2. ワークシートの一列目に、二行目から 1~(サンプルサイズ)の番号を入力する。

### 連続データの作り方例

- 1) A2 のセルに 1 を入力する。
- 2) (サンプルサイズ+1)行目まで選択する。
- 3) [編集] [フィル] [連続データの作成]とクリックし、増分値を 1 として[OK]をクリック。
3. [ウインドウ] [分割]でシートを分割し、一行目と二行目の間、そして一列目と二列目との間でウインドウが分割するように枠線を移動する。
4. その後、[ウインドウ] [ウインドウ枠の固定]で一行目、一列目を固定する。
5. データが入力されるべき箇所全体に、「9999」を入力しておく。「9999」は無回答を表すものとしており、アンケートの回答としてありえない記号であればなんでもよい。

## 3. データ入力方法(kansai.xls 参照)

入力は、次の手順で行う。

1. 入力モードを直接入力にする。
2. 一つのアンケート調査票のデータを(観測値ベクトルという)一つの行に入力する。

### 入力の注意点

- 1) 各データは半角数字で入力する。
- 2) 無回答は「9999」のままにする(矢印キーで入力をスキップすればよい)。
- 3) 単一回答は選択された数字をそのまま入力する。
- 4) 複数回答は、全ての選択肢をそれぞれ一つの質問とみなしているため、がついていれば 1 を、ついていなければ 2 を入力する。
- 5) 順位回答は、順位が記入されているものにはその順位を、それ以外には 0 を入力する。
- 6) 数値回答は、そのまま入力してもよいし、予め適当なカテゴリーに区切っておいて、そのカテゴリーをあらわす数字を入力してもよい。文字による自由回答も同様である。なおここでは、問 5 について、関西在住であれば 1 を、それ以外であれば 2 を入力するものとしている。

#### 4. 単一回答の単純集計

Excel では、ピボットテーブルのコマンドを使って、アンケートを集計する。

1. データが入力されたら、どこでもよいから、そのデータを一つ選択した状態で、[データ] [ピボットテーブルとピボットグラフレポート]を選択する。
2. [次へ] [次へ] [完了]と順にクリックする。
3. 新しいシートが開き、集計用のセルが作られる。
4. 単純集計を行いたい変数(ここではこれをフィールドという)をピボットテーブルウインドウの中から選び、それを「ここに行のフィールドをドラッグします」のところまでドラッグする(例:kote をドラッグ)。
5. さらに同じ変数を、「ここにデータアイテムをドラッグします」のところまでドラッグする(例:kote をドラッグ)。
6. とりあえず単純集計の結果が表示されるが、各行の合計は、それぞれの入力値の合計(2 については、 $2 \times 9 = 18$  となっている)になっているので、これは正しくない。

合計 : kote		
kote	計	
	1	20
	2	18
総計		38

7. ピボットテーブルウインドウから[ピボットテーブル] [フィールドの設定]をクリックし(あるいは左上のセルをダブルクリック)、[データの個数]を選択して、[OK]をクリックする。

データの個数 : kote		
kote	計	
	1	20
	2	9
総計		29

8. 別の変数の単純集計を続けて行いたい場合は、新たに作るピボットテーブルを表示したい場所にカーソルをおき、そのセルをクリックして、ピボットテーブルウインドウから[ピボットテーブル] [ウィザード]をクリックする。
9. [ほかのピボットテーブルまたはピボットグラフ]を選択し、[次へ] [次へ] [完了]と順にクリックする。
10. 新たにピボットテーブルが作られるので、再び、別の変数について 4 からの作業を行う。

#### 5. 単一回答の単純集計結果の整理

1. 新しいシートを開き、B5、B6 のセルを空白として、B7、C5、C6、C7 のセルに次のような数式を入力する。

B5	=B5/B7
B6	=B6/B7
=SUM(B5:B6)	=SUM(C5:C6)

2. 次のように表されるはずである。

	#DIV/0!
	#DIV/0!
0	#DIV/0!

3. ピボットテーブルの合計を除く集計結果のみ (kote の例では、20 と 9) をコピーし、B5 と B6 に貼り付ける。

20	0.689655172
9	0.310344828
29	1

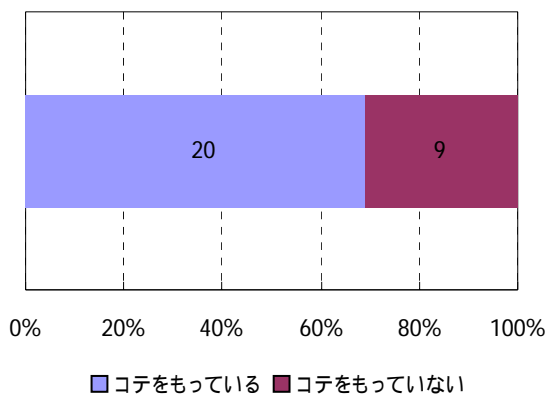
4. C5 ~ C7 を選択し、右クリックで、[セルの書式設定] [表示形式] [パーセンテージ] とクリックし、[小数点以下の桁数] を「1」として[OK]をクリックする。

20	69.0%
9	31.0%
29	100.0%

5. A5、A6、A7、B4、C4 のセルに適当な文字列を入力し、表を整える。

	実数	%
コテをもっている	20	69.0%
コテをもっていない	9	31.0%
合計	29	100.0%

6. 次のように帯グラフなどで図示するのも有効である。



## 6. 複数回答の単純集計

- 複数回答の場合は、選択肢ごとに変数名をつけたので、一つ一つの選択肢について単純集計し、最終的にそれらを統合する必要がある。
- 先に単純集計のところで作ったピボットテーブルのシートで、新しいピボットテーブルをつくる(4の8を参照)。
- 複数回答の選択肢の一つをピボットテーブルウィンドウの中から選び、それを「ここに行のフィールドをドラッグします」のところまでドラッグする(例: resist1 をドラッグ)。
- さらに同じ変数を、「ここにデータアイテムをドラッグします」のところまでドラッグする(例: resist をドラッグ)。
- とりあえず単純集計の結果が表示されるが、先述のように、各行の合計は、それぞれの入力値の合計(2については、 $2 \times 16 = 32$  となっている)になっているので、これは正しくない。

合計 : resist1	
resist1	計
1	13
2	32
総計	45

6. ピボットテーブルウィンドウから[ピボットテーブル] [フィールドの設定]をクリックし(あるいは左上のセルをダブルクリック)、[データの個数]を選択して、[OK]をクリックする。

データの個数 : resist1	
resist1	計
1	13
2	16
総計	29

7. この 13 という値が、この選択肢(ここでは、お好み焼き)を回答した実数となる。  
 8. 以上の作業を全ての選択肢について繰り返し、選択肢の個数だけ、表を作る。

データの個数 : resist1	
resist1	計
1	13
2	16
総計	29

データの個数 : resist2	
resist2	計
1	23
2	6
総計	29

...

#### 7. 複数回答の単純集計結果の整理

1. 選択肢数が 10 の例を挙げる。選択肢数が変わればそれだけ行の数を変えればよいだけで、手続きは同じように進めることができる。
2. 新しいシートを開き、B5～B15 のセルを空白として、C5～C15 のセルに次のような数式、および数値を入力する。

B5	=B5/B15
B6	=B6/B15
B7	=B7/B15
B8	=B8/B15
B9	=B9/B15
B10	=B10/B15
B11	=B11/B15
B12	=B12/B15
B13	=B13/B15
B14	=B14/B15
B15	1

3. 次のように表されるはずである。

	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	#DIV/0!
	1

4. B15 のセルに、有効回答部数(サンプルサイズ)を入力し、B5～B14 のセルに、各選択肢について「1」と回答した実数を入力する。

13	0.448275862
23	0.793103448
9	0.310344828
16	0.551724138
27	0.931034483
14	0.482758621
18	0.620689655
11	0.379310345
27	0.931034483
15	0.517241379
29	1

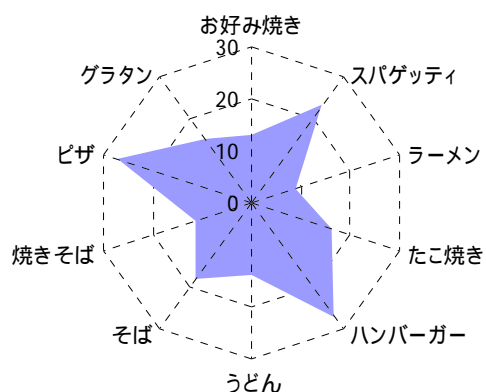
5. C5～C15 を選択し、右クリックで、[セルの書式設定] [表示形式] [パーセンテージ]とクリックし、[小数点以下の桁数]を「1」として[OK]をクリックする。

13	44.8%
23	79.3%
9	31.0%
16	55.2%
27	93.1%
14	48.3%
18	62.1%
11	37.9%
27	93.1%
15	51.7%
29	100.0%

6. A5～A15、B4、C4のセルに適切な文字列を入力し、表を整える。

	実数	%
お好み焼き	13	44.8%
スパゲッティ	23	79.3%
ラーメン	9	31.0%
たこ焼き	16	55.2%
ハンバーガー	27	93.1%
うどん	14	48.3%
そば	18	62.1%
焼きそば	11	37.9%
ピザ	27	93.1%
グラタン	15	51.7%
有効回答部数	29	100.0%

7. 次のようにレーダーチャートなどで図示するのも有効である。



## 8. クロス集計

- 先に単純集計のところを開いたピボットテーブルのシートで、新しいピボットテーブルをつくる(4の8を参照)。
- 表頭にもってきたい変数(「A別にみたB」という表を作るときのA)をピボットテーブルウィンドウの中から選び、それを「ここに列のフィールドをドラッグします」のところまでドラッグする(例: home をドラッグする。ここでは、居住地別にコテの有無が変わってくるかをみる。 )。
- 表側にもってきたい変数(「A別にみたB」という表を作るときのB)をピボットテーブルウィンドウの中から選び、それを「ここに行のフィールドをドラッグします」のところまでドラッグする(例: kote をドラッグ)。
- さらに、表側にドラッグした変数を、「ここにデータアイテムをドラッグします」のところまでドラッグする(例: kote をドラッグ)。
- とりあえずクロス集計の結果が表示されるが、各行の合計は、それぞれの入力値の合計(2の総計については、 $2 \times 9 = 18$ となっている)になっているので、これは正しくない。

合計 : kote	home		
kote	1	2	総計
1	15	5	20
2	12	6	18
総計	27	11	38

6. ピボットテーブルウィンドウから[ピボットテーブル] [フィールドの設定]をクリックし(あるいは左上のセルをダブルクリック)、[データの個数]を選択して、[OK]をクリックする。

データの個数 : kote	home		
kote	1	2	総計
1	15	5	20
2	6	3	9
総計	21	8	29

### 9. クロス集計結果の整理

- 基本的には、5節、7節で述べたことを、各列について行えばよい。ここでは、単一回答同士の2×2クロス集計結果のみについて説明する。
- 新しいシートを開き、B5、B6、D5、D6のセルを空白として、次のような数式を入力する。

B5	=B5/B7	D5	=D5/D7	=B5+D5	=F5/F7
B6	=B6/B7	D6	=D6/D7	=B6+D6	=F6/F7
=SUM(B5:B6)	=SUM(C5:C6)	=SUM(D5:D6)	=SUM(E5:E6)	=SUM(F5:F6)	=SUM(G5:G6)

- 次のように表されるはずである。

	#DIV/0!		#DIV/0!	0	#DIV/0!
	#DIV/0!		#DIV/0!	0	#DIV/0!
0	#DIV/0!	0	#DIV/0!	0	#DIV/0!

- ピボットテーブルの合計を除く集計結果のみ(kote×homeの例では、15、6、5、3)をコピーし、B5、B6、D5、D6に貼り付ける。

15	0.714285714	5	0.625	20	0.689655172
6	0.285714286	3	0.375	9	0.310344828
21	1	8	1	29	1

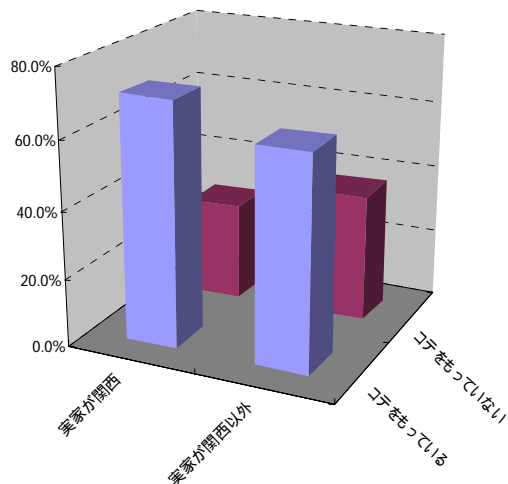
- C、E、G列の5~7行目をそれぞれ選択し、右クリックで、[セルの書式設定] [表示形式] [パーセンテージ]とクリックし、[小数点以下の桁数]を「1」として[OK]をクリックする。

15	71.4%	5	62.5%	20	69.0%
6	28.6%	3	37.5%	9	31.0%
21	100.0%	8	100.0%	29	100.0%

- 表の周辺部のセルに適当な文字列を入力し、表を整える。

	実家が関西		実家が関西以外		全体	
	実数	%	実数	%	実数	%
コテをもっている	15	71.4%	5	62.5%	20	69.0%
コテをもっていない	6	28.6%	3	37.5%	9	31.0%
合計	21	100.0%	8	100.0%	29	100.0%

7. この結果から、実家が関西の人の方が、コテを持っている割合が高いことがわかる。  
 8. 次のようなステレオグラムなどで表示するのも有効である。



9. 複数回答(10)×単一回答(2)の場合の基礎となる表は次のようになる。

B5	=B5/B15	D5	=D5/D15	=B5+D5	=F5/F15
B6	=B6/B15	D6	=D6/D15	=B6+D6	=F6/F15
B7	=B7/B15	D7	=D7/D15	=B7+D7	=F7/F15
B8	=B8/B15	D8	=D8/D15	=B8+D8	=F8/F15
B9	=B9/B15	D9	=D9/D15	=B9+D9	=F9/F15
B10	=B10/B15	D10	=D10/D15	=B10+D10	=F10/F15
B11	=B11/B15	D11	=D11/D15	=B11+D11	=F11/F15
B12	=B12/B15	D12	=D12/D15	=B12+D12	=F12/F15
B13	=B13/B15	D13	=D13/D15	=B13+D13	=F13/F15
B14	=B14/B15	D14	=D14/D15	=B14+D14	=F14/F15
B15	1	D15	1	=B15+D15	1

10. 2×2 クロス集計結果のカイ二乗検定

1. 新しいシートを開き、C6、C7、D6、D7 のセルを空白として、次のような数式を入力する。

	home				
kote	C6	D6	=SUM(C6:D6)		
	C7	D7	=SUM(C7:D7)		
	=SUM(C6:C7)	=SUM(D6:D7)	=SUM(C8:D8)		
	home				
kote	=E11*C13/E13	=E11*D13/E13	=E6		
	=E12*C13/E13	=E12*D13/E13	=E7	p 値	=CHITEST(C6:D7, C11:D12)
	=C8	=D8	=E8	カイ二乗値	=CHIINV(G12,1)



2. 総計を除くピボットテーブルのクロス集計結果のみ (kote×home の例では,15,6,5,3) をコピーし、C6、C7、D6、D7 に貼り付ける。

	home				
kote	15	5	20		
	6	3	9		
	21	8	29		
	home				
kote	14.48	5.52	20		
	6.52	2.48	9	p 値	0.642
	21	8	29	カイ二乗値	0.216

3. この検定では、有意水準 5% で帰無仮説「実家とコテの有無には関係がない」が棄却されず、「関西在住である人はコテをもっている人が多いとはいいいにくい」という結論になる。