

情報検索：ロシアの美しいスパイの話 ～アンケートとデータ集計～

2012年6月8日(金)
ワンコイン経営セミナー

富山国際大学子ども育成学部
上坂 博亨 (うえさかひろゆき)
uesaka@tuins.ac.jp

1

アンケート調査の流れ

アンケートの立案

- 目的の明確化
- 質問内容・実施方法・実施場所・実施時期など

質問票の作成

- 単一選択・複数選択・数値回答・語句回答・自由記述
- 選択肢の決定のために予備調査実施

調査の実施

- 想定した回答者にアンケートが届く *郵送は2ヶ月(回答)がわかる*
- 目的や趣旨が理解される

回答結果の回収

- 回収率に影響を与える
- 時間・コストを考慮

回答結果の入力

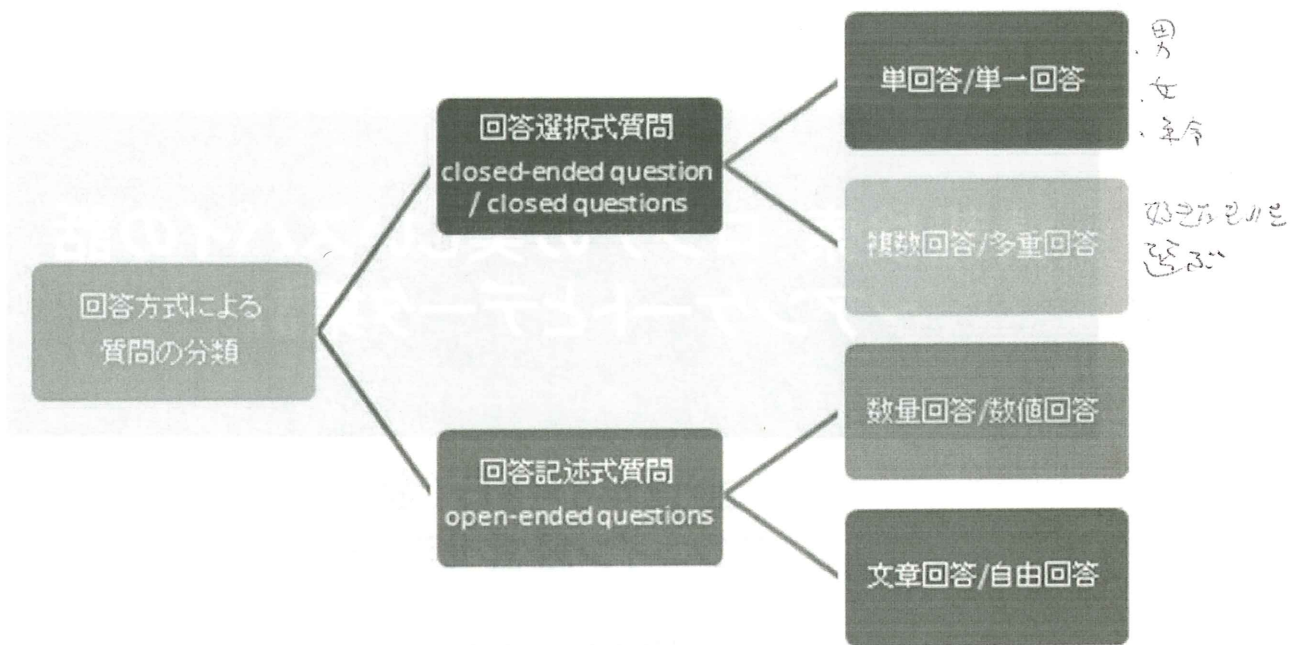
- EXCELへのデータ入力：1回答1レコード
- データの配置と符号化

集計と解析

- 基本的統計計算(平均・分散・標準偏差・中央値・最頻値)
- 関数の利用・グラフ化・クロス集計・相関係数

2

回答選択肢の種類



統計Web ホームページより

<http://software.ssri.co.jp/statweb2/column/column1102.html#c099>

3

基本的なデータ配置

通し番号	質問01	質問02	質問03	質問04
001	回答	回答	回答	回答
002	回答	回答	回答	回答
003	回答	回答	回答	回答
004	回答	回答	回答	回答

- 改修したアンケートに通し番号を付加
 - 後で照合の必要が生じた場合への準備
- 質問票の回答選択肢に番号を付けて符号化
 - コンピュータでのデータ集計は自然言語では困難 符号化
- EXCELに入力する際は、1回答を1レコードで
 - 「レコード」とはデータベース用語で、表の「行」に当たる
 - EXCEL表の限界は65536行 2¹⁶ 2⁸ bit
- 先頭行に項目行、2行目以降から回答データ

性別
(男、女)



性別
1. 男
2. 女

4

Excelへのデータ入力

フィールド名には、データとは異なる書式を付ける

色付け

列見出し行は1行のみとし、複数行からなる事は禁止



けっこうたいへんだぜ!

リストの周辺は、1行あるいは1列の完全な空白領域を作る(これによってEXCELがデータ領域を認識)

リスト内部には完全な空白行や空白列があってはならない

	A	B	C	E	F	G	H	I	J	K
1										
2		No.	Q1	Q2	Q3	Q3-SQ-1	Q3-SQ-2	Q3-SQ-3	Q3-SQ-4	Q3-SQ-5
3		1	2	50	2	1	0	0	0	0
4		2	3	50	1	-	-	-	-	-
5		3	4	30	1	-	-	-	-	-
6		4	3	30	1	-	-	-	-	-
7		5	4	30	1	-	-	-	-	-
8		6	1	30	1	-	-	-	-	-
9		7	4	30	1	-	-	-	-	-
10		8	2	50	1	-	-	-	-	-
11		9	2	40	1	-	-	-	-	-
12		10	1	12	1	-	-	-	-	-
13		11	3	30	1	-	-	-	-	-
14		12	3	30	2	0	1	1	0	0
15			3	40	1	-	-	-	-	-
16		:	:	:	:	:	:	:	:	:
17		:	:	:	:	:	:	:	:	:
18										
19										
20										

5

単一回答のデータ集計方法

	A	B	C	D	E	F
1						
2		Q1		件数	割合	
3		1 徒歩				
4		2 自転車				
5		3 バス				
6		4 鉄道				
7		合計				
8						

集計するためのワークシートを作ります。

集計はデータリストとは別のシートに作成することをお勧めします。

	A	B	C	D	E	F	G
1							
2		Q1		件数	割合		
3		1 徒歩		=COUNTIF(データ!\$B\$15:\$B\$214,B3)			
4		2 自転車					
5		3 バス					
6		4 鉄道					
7		合計					
8							

COUNTIF関数を用いて件数を求めます。COUNTIF関数は、[範囲]内で[検索条件]を満たす数値の個数を求める関数です。

COUNTIF(範囲, 検索条件)

集計表内のセル範囲「D3:D6」を選択後、下図のように数式を入力します。[範囲]に「データ」シートのQ1のデータ範囲、[検索条件]に集計表のカテゴリコードのセルを指定します。また、選択範囲内で同じ式を利用できるように、[範囲]を絶対参照で参照します。絶対参照とするには、数式のテキストにカーソルを置いて[F4]キーを押し、\$マークを付けます。

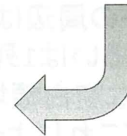
複数回答のデータ集計方法

集計するためのワークシートを作ります。

COUNTIF関数を用いて全体の件数を求めます。全体の件数は、Q3で「いいえ」と回答した人の数です。

	A	B	C	D	E	F	G
1							
2		Q3-SQ		全体	件数	割合	
3		1	食べている時間がないから	=COUNTIF(データ!\$D\$15:\$D\$214,2)			
4		2	朝は食欲がないから				
5		3	ダイエットのため				
6		4	朝食を作ってもらえないから				
7		5	この中にあてはまる理由はない				
8							

	A	B	C	D	E	F	G	H	I
1									
2		Q3-SQ		全体	件数	割合			
3		1	食べている時間がないから	57	=SUM(INDEX(データ!\$E\$15:\$I\$214,0,B3))				
4		2	朝は食欲がないから	57					
5		3	ダイエットのため	57					
6		4	朝食を作ってもらえないから	57					
7		5	この中にあてはまる理由はない	57					
8									



INDEX関数は、[参照]範囲の[行番号]と[列番号]の交差するセル範囲を返す関数です。ここでは、[行番号]に0を指定することで[列番号]で指定した列全体のセル範囲を返り値とします。

SUM関数で各選択肢の件数を求めます。各選択肢のデータ範囲が異なるので、SUM関数の参照範囲をINDEX関数で切り替えて指定します。

white + Alt + ニ

http://software.ssri.co.jp/statweb2/enquete/eq_2.html

7

数量回答のデータ集計方法

	A	B	C	D	E	F	G	H	I
1									
2		Q2		件数	割合		検索条件		
3		1	0~9分				>=0		
4		2	10~19分				>=10		
5		3	20~29分				>=20		
6		4	30~39分				>=30		
7		5	40~49分				>=40		
8		6	50~59分				>=50		
9		7	60~69分				>=60		
10		8	70~79分				>=70		
11		9	80~89分				>=80		
12		10	90分以上				>=90		
13		合計							

COUNTIF関数を用いて件数を求めるために、検索条件を作成します。

階級「0~9分」の場合、「>=0」を満たすセルの個数から「>=10」を満たすセルの個数を引きます。

数式の指定は以下の様になります

	A	B	C	D	E	F	G	H	I	J	K
1											
2		Q2		件数	割合		検索条件				
3		1	0~9分	=COUNTIF(データ!\$D\$16:\$D\$215,G3)-COUNTIF(データ!\$D\$16:\$D\$215,G4)							
4		2	10~19分				>=10				
5		3	20~29分				>=20				
6		4	30~39分				>=30				
7		5	40~49分				>=40				
8		6	50~59分				>=50				
9		7	60~69分				>=60				
10		8	70~79分				>=70				
11		9	80~89分				>=80				
12		10	90分以上				>=90				
13											

この式をD4からD12までコピーします。

http://software.ssri.co.jp/statweb2/enquete/eq_3.html

8

アンケート例題

高校生の生活調査 (架空)

Q1.あなたは、どのようにして学校に通っていますか？

1. 徒歩 2. 自転車 3. バス 4. 鉄道

Q2.通学にかかる時間はどれくらいですか？

1. 0～9分 3. 20～29分 5. 40～49分 7. 60～69分 9. 80～89分
2. 10～19分 4. 30～39分 6. 50～59分 8. 70～79分 10. 90分以上

Q3.平日は、毎日、朝食を摂っていますか？

1. はい 2. いいえ

Q3-SQ.朝食を摂らない理由は何ですか？

1. 食べている時間がないから
2. 朝は食欲がないから
3. ダイエットのため
4. 朝食を作ってもらえないから
5. この中にあてはまる理由はない

Q4.あなたは授業中眠たくなることがありますか？

1. ほぼ毎日のように眠たくなる
2. 毎日ではないが、時々眠たくなることがある
3. ほとんど眠たくなることはない

Q5.あなたは部活動を行っていますか？

1. はい
2. いいえ

Q6.あなたは進学塾や予備校に通っていますか？

1. 週に3回以上通っている
2. 週に1,2回通っている
3. 進学塾や予備校には通っていない

Q7.あなたの平日の就寝時刻は何時くらいですか？

1. 午後10時頃までに就寝
2. 午後11時前後

3. 深夜 12 時前後
4. 深夜 1 時前後
5. 深夜 2 時以降

Q8.あなたは卒業後の進路をどのように考えていますか？

1. 短期大学に進学する
2. 4年制大学に進学する
3. 専門学校に通って技術を身に着ける
4. 就職する
5. まだ決めていない

Q9-1.自分にしかできない事をしてみたい

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-2.他人に指図されるのは嫌いである

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-3.自分には他の人よりも優れたところがある

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-4.他の人と同じ服を着るのは嫌だ

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-5.こつこつと努力することが好きだ

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-6.いつも何か目標を持っている

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-7.今日は何をしようかと考えることは楽しい

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-8.みんなに喜んでもらえることをしたい

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-9.変わり者と見られないよう皆と同じにしている

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-10.古い習慣や伝統は守っていかなければいけない

1. あてはまる
2. ややあてはまる

3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-11.あまり自分の意見を言わない方である

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q9-12.誰からも嫌われたくない

1. あてはまる
2. ややあてはまる
3. どちらともいえない
4. あまりあてはまらない
5. あてはまらない

Q10.あなたは将来どのような仕事につきたいですか?

F1.性別

1. 男性
2. 女性

F2.学年

1. 1年生
2. 2年生
3. 3年生

F3.同居の家族

1. 父親
2. 母親
3. 兄や姉
4. 弟や妹
5. 祖父
6. 祖母
7. 家族とは同居していない

F4.父親の学歴

1. 中卒
2. 高卒
3. 高専卒
4. 短大卒
5. 大卒
6. 大学院卒
7. わからない

F5.母親の学歴

1. 中卒
2. 高卒
3. 高専卒
4. 短大卒
5. 大卒
6. 大学院卒
7. わからない

お答えします

Excelでグラフの順番を直すには

Q

Excelで横棒グラフを作ると、表の順番とは逆に項目が並んでしまっています。これを正しい順番に変更したいのですが、どのようにすればよいでしょうか。
(京都府、53歳、男性)

A

Excelを使って右図のような横棒グラフを作ると、各項目の並び順が、表の並び順と逆転してしまいます。これはExcelの仕様。Excelでは表のそれぞれの要素が項目名に近いほど、横棒グラフの数値軸に近く配置されるようになっています。

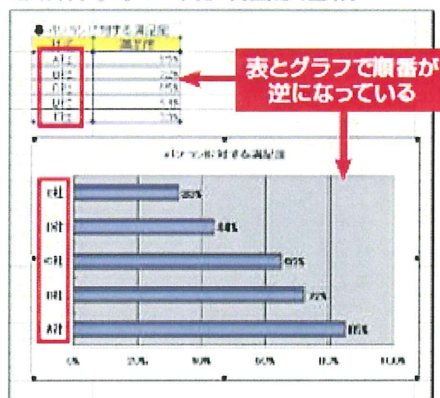
ここでいう項目名とは表の先頭にある、項目を説明している語句のこと。要素は表中のデータの実際の数値を指します。右図でいえば、項目名が「社名」で、その下にあるA社、

B社、C社、D社、E社が要素になります。図の順番で表を作成した場合、グラフの並び順は、表とは逆転し、上からE社→A社となるわけです。

書式で順番を入れ替える

これを変更する方法として、表自体の並び方を逆転してしまう方法があります。つまり、表中のデータの並び順をE社からA社の順番に並べ替えるわけです。ただ、確かにグラフの並び方は正しくなりますが、表を作り直すのは面倒です。

●横棒グラフでは項目が逆転



そのまま横棒グラフを作成すると、図のように表の項目に対してグラフの項目順が逆転し、E社→A社の順に並んでしまう

ここではもっと簡単な方法を紹介しましょう。グラフの設定を変更すれば、項目の並び順を入れ替えられます(左下図)。

まず、通常通りグラフを作成します。次に、作成した表の項目軸の部分を右クリック。表示されたメニューで「軸の書式設定」を選びます。書式設定の画面が開いたら、「目盛」タブを選んで「軸を反転する」「最大項目でY/数値軸と交差する」の2つにチェックを入れます。後は「OK」ボタンを押すだけです。

「軸を反転する」を選ぶと、上からA社→E社の順に並びます。ただ、数値軸、つまり0%から100%の数字が書かれている部分も反転し、上部に表示されます。数値軸を下に表示するのが「最大項目でY/数値軸と交差する」設定です。最大項目とは、表の項目名から一番離れた項目のこと。ここでいえばE社のデータです。これで数値軸がE社のデータに最も近い位置、すなわちグラフの下部に配置されます。

●正しいグラフにするには

1 並びを逆転するには、表の項目軸にポインターを合わせてマウスの右ボタンをクリック。表示されたメニューで「軸の書式設定」を選ぶ

2 軸の書式設定の画面で「目盛」タブを開き、「軸を反転する」「最大項目でY/数値軸と交差する」にチェックを付ける

3 これで表の並びと同じ順番になった横棒グラフが完成する

