

クロス集計とピボットテーブル

クロス集計とは

数値データを積算して合計を求めることが「集計」ですが、何のために集計を行うかという、集計で得たデータ等をもとに数量的な傾向や特性をつかみ(これを「統計」と言います)、さらには、数値を比較検討して差異の原因や理由を推測したり、仮説を立てて証明したり(これを「分析」と言います)、するためです。良い統計や分析のためには、統計や分析を行う人がデータの意味するところや傾向などが分りやすい形で集計がされていることが重要と言えます。

集計は、単純集計とクロス集計の大きく2つに分けられます。

単純集計はGT(グランドトータル)とも呼ばれ、調査結果などのデータを整理する際に、異なる属性や基準などに分類せず、すべてを単純に合計する集計方法のことです。アンケート結果のデータであれば、設問ごとの回答選択肢別にデータを合計して、比率などを算出するというもので、集計の基本的な方法となります。ちなみに集計する前のデータのことをローデータと呼びます。ロー(Raw)は生のという意味です。単純集計では分析軸が1つしかありません。

一方、クロス集計とは、調査結果などのデータを2つ以上の分析軸から集計していく方法のことです。アンケート結果のデータであれば、分析軸として主に用いられるのは、回答者の属性(例:性別、年齢別、居住地域別など)や設問の回答内容です。これらはそれぞれ「属性間クロス集計」「設問間クロス集計」と呼ばれています。

クロス集計における分析軸は2軸である必要はありません。必要であれば、3軸以上にすることもできます。これを多重クロス集計と言い、それぞれ「n重クロス集計(n=3,4,5...)」と呼びます。例えば、アンケート結果のデータで、ある設問の回答データを「性別+年齢別」にクロス集計をすることは、3重クロス集計となります。

クロス集計は、調査の種類・目的によってさまざまな分析軸(切り口)で単純集計結果を細分化し、データの詳細を掴むために活用されるのです。

クロス集計の方法と分析

クロス集計には、おおまかに以下の3種類があります。

(1) カテゴリ x カテゴリ

例: 設問の選択肢 x 性別

(2) カテゴリ x 数量

例: 設問の選択肢 x [年収]

(3) 数量 x 数量

例: [年収] x [預金額]

※[年収]や[預金額]は、選択肢としてではなく、数量データに基づいて得たもの

一般的には、回答や集計のし易さから、(1)が最も利用されます。例えばアンケート調査においては、年収や預金額も、その範囲を選択肢として回答させる設問にすれば(2)(3)ではなく、(1)になります。

組合せを自在に選択することで、多種多様な分析が可能になります。視覚化するために棒グラフや円グラフも作れますし、必要とあれば性別・年齢別と数量といった3重クロス集計も可能で、3次元立体グラフさえも描画できます。

クロス集計は、非常に有効な分析手法なのですが、サンプルサイズ(標本の大きさ=抽出したデータの個数、一般に n で表示される数)が少ないと、分析結果で得られる差が統計学的に意味のあるもの(有意差)とは言えなくなってしまいます。一方でクロス集計での分析精度を高めようとするほど、サンプルサイズが大きくなって、調査等にかかる費用が高くなり、日程も長くなりがちです。サンプルサイズ、コスト、日程において妥協点を見極めておくべきでしょう。統計学に基づけば、正規分布と仮定した場合、サンプルサイズは30以上であることが推奨されています。

クロス集計表

クロス集計は例えば実験結果のデータ分析などでも使われますが、一番よく使われるのは、アンケート調査結果の分析の場面でしょう。

クロス集計結果を表にしたものをクロス集計表と呼びます。

ここではアンケート調査の例を使って、クロス集計表について解説します。

右の表は、列別に設問の選択肢、行別に年齢層とするクロス集計表の例です。

表頭と表側

最も上の行にある分析軸(この場合は設問の選択肢)を「表頭(ひょうとう)」、最も左の列にある分析軸(この場合は年齢層)を「表側(ひょうそく)」と言います。

上記の表はクロス集計のみの表でしたが、実際には、右の表のように、単純集計とクロス集計を組み合わせ、それぞれの合計の欄も付けた形で示されることが多いでしょう。

度数(実数)と割合(比率)

上記の表ではそれぞれの分析軸で集計した実数(度数)が示されています。分析結果を考察するには、実数でなく、%表記などの割合(比率)で示した方が分かりやすい場合がよくあります。度数と割合の両方を載せている表も少なくありません。

横%表と縦%表

割合を求めるのに、横方向の合計値で割った割合を示した表を横%表と言います。一方、縦方向の合計値で割った割合を示した表を縦%表と言います。

単一回答の分析軸の場合、横%表では横方向の%の合計が 100%に、縦%表では縦方向の%の合計が 100%になります。

【設問】 地域活動への参加の必要性 **表頭**

	必要だと思う	ある程度必要だと思う	あまり必要だと思わない	必要だと思わない
29歳以下	49	142	14	6
30代	38	126	16	5
40代	75	232	29	3
50代	101	219	25	3
60代	63	139	13	2
70歳以上	105	157	18	2

表側
年齢層別

クロス集計表の例

【設問】 地域活動への参加の必要性 **表頭**

単純集計	必要だと思う	ある程度必要だと思う	あまり必要だと思わない	必要だと思わない	合計
全体	431	1,015	115	21	1,582
29歳以下	49	142	14	6	211
30代	38	126	16	5	185
40代	75	232	29	3	339
50代	101	219	25	3	348
60代	63	139	13	2	217
70歳以上	105	157	18	2	282

表側
年齢層別
クロス集計

単純集計とクロス集計表を組み合わせた例

【設問】 地域活動への参加の必要性

	回答数	必要だと思う	ある程度必要だと思う	あまり必要だと思わない	必要だと思わない
全体	1,582	27.2%	64.2%	7.3%	1.3%
29歳以下	211	23.2%	67.3%	6.6%	2.8%
30代	185	20.5%	68.1%	8.6%	2.7%
40代	339	22.1%	68.4%	8.6%	0.9%
50代	348	29.0%	62.9%	7.2%	0.9%
60代	217	29.0%	64.1%	6.0%	0.9%
70歳以上	282	37.2%	55.7%	6.4%	0.7%

年齢層別

横%表の例

【設問】 地域活動への参加の必要性

	回答数	必要だと思う	ある程度必要だと思う	あまり必要だと思わない	必要だと思わない
全体	1,582	431	1,015	115	21
29歳以下	13.3%	11.4%	14.0%	12.2%	28.6%
30代	11.7%	8.8%	12.4%	13.8%	23.8%
40代	21.4%	17.4%	22.9%	25.2%	14.3%
50代	22.0%	23.4%	21.6%	21.7%	14.3%
60代	13.7%	14.6%	13.7%	11.3%	9.5%
70歳以上	17.8%	24.4%	15.5%	15.7%	9.5%

年齢層別

縦%表の例

多重クロス集計表

右図は、性別 x 年齢層別 x 設問の選択肢という3つの分析軸の3重クロス集計表の例です。

例えば、性別 x 年齢層別 x 居住地別 x 設問の選択肢といった風に、4重クロス集計やそれ以上の多重クロス集計表の作成も可能です。ただし多重度が大きくなるほど表も大きくなりますので、3重ないしは4重クロス集計程度に留めることをお勧めします。

【設問】 地域活動への参加の必要性

	回答数	必要だと思う	ある程度必要だと思う	あまり必要だと思わない	必要だと思わない
全件	1,582	431	1,015	115	21
男性 計	827	233	533	55	6
29歳以下	112	22	82	6	2
30代	101	21	70	8	2
40代	172	39	120	13	0
50代	180	55	110	14	1
60代	117	38	72	6	1
70歳以上	145	58	79	8	0
女性 計	755	198	482	60	15
29歳以下	99	27	60	8	4
30代	84	17	56	8	3
40代	167	36	112	16	3
50代	168	46	109	11	2
60代	100	25	67	7	1
70歳以上	137	47	78	10	2

3重クロス集計表の例

クロス集計のメリット

アンケート調査の結果をわかりやすく可視化できる

クロス集計は、単純集計を細分化する形で集計されるので、数値の意味がはっきりして比較がしやすく、統計に詳しくない人でもわかりやすいのが大きなメリットです。またグラフへの加工も簡単で、クロス集計表と併せてグラフを活用することで、企画書やプレゼンテーションの説得力が増します。

Excel にクロス集計のためのツールが用意されている

データ量が極端に大きくなければ、Excel で簡単にクロス集計を実施できます。フィルタと SUBTOTAL 関数を用いて分析軸単位にローデータを拾って、個別にクロス集計表を作るという方法もありますが、「ピボットテーブル」機能を使えば、さらに簡単にクロス集計表を作成することができます。グラフへに置き換えることも可能ですので、様々な視点からアンケート結果を分析・可視化することができます。

手軽に視点を変えて分析できる

クロス集計の掛け合わせる分析軸を変えるだけで視点を変えた分析ができるという点もメリットです。アンケート調査の設問を2つ以上組み合わせるとクロス集計を実施すると、色々面白い事実が判明するという点もあるでしょう。また、組み合わせる情報を1つ変えるだけでも分析の視点を変えられるので、1回の調査で様々な知見が得られるというのもクロス集計分析がよく利用される理由です。

ピボットテーブルとは

ピボットテーブル(pivot table)とは、大量のデータの分析を可能にする Excel の機能です。リスト形式で用意されたローデータを任意の形式で分類し、元のデータを「顧客ごとの売上金額」「支社ごとの売上金額」などの複数の項目別に集計したり、それらの集計項目を入れ替えたりするなど、できることは多岐にわたります。

ピボットテーブルを用いれば、複雑な数式や特別な関数を使わずにクロス集計ができます。元のデータを変更せずに集計方法を切り替えられるため、コピー&ペーストを繰り返す必要もありません。

数値や数式が入力されただけの一般的な表とは異なり、柔軟性に富んだ結果の表示に強く、さまざまな調整が可能です。例えば、全取引先への支払履歴を表として Excel に打ち込んだあと、それに基づいてピボットテーブルを起動すれば「月別の支出額」や「〇〇社への支払い総額」などを即座に掴むこともできるのです。

ピボットテーブルの使い方等は別ページで詳細に解説します。