

レポート

1 アンケートの設問間の関係を分析する

(一財)和歌山社会経済研究所 主任研究員

長谷川 強

1. はじめに

通常、アンケートを行うときには複数の設問の項目があり、項目に対する回答は相互に関連があるという可能性を想定する。特に、注目している項目の回答結果には、その要因や要素として他の回答に求めようとする、ということがよくある。

そのような場合は、クロス集計表を作って2つの項目の関連を評価する。ところがそれは意外に難しい。2つ項目 A,B の回答が共に「はい」「いいえ」の二者択一であり、A が「はい」なら B も「はい」、A が「いいえ」なら B も「いいえ」、というような傾向が明らかである場合(表 1.1) はよいが、それが微妙な配分の場合(表 1.2) に関連がある、もしくはないと言い切ることに躊躇することが少なくない。また、回答が3つ以上の場合はますます評価が困難になる。(こういったクロス集計表を評価するには「カイ二乗検定」という手法があるが、ここでは割愛する。)

さらに、クロス集計表はあくまで2つの項目間の関係だけの評価に留まり、3つ以上の項目の関係性を評価しようとする、2つのときと比べて一層複雑で項目数の多い表(表 1.3)となり、評価が難しくなる。

これらの課題を解決する手法として、多変量解析の一つである数量化理論を用いる。

表 1.1 評価が容易なクロス集計表の例

項目 A \ 項目 B	はい	いいえ
はい	80	10
いいえ	5	40

表 1.2 評価が容易でないクロス集計表の例

項目 A \ 項目 B	はい	いいえ
はい	40	25
いいえ	30	35

表 1.3 3 設問のクロス集計表の例

項目 B	はい			いいえ		
項目 A \ 項目 C	良い	ふつう	悪い	良い	ふつう	悪い
はい	10	25	5	30	20	10
いいえ	30	35	10	10	10	10

(数値はサンプル数、以下同じ)

2. 数量化理論

2.1. 数量化理論とは

「数量化理論」とは、非常に大雑把に言うと、回答に数値（点数）を割り付け、その値の大小や合計値で事象を説明しようとする理論である。

本稿では、実際のデータに数量化理論を適用して、その特徴や留意点を述べる。そのデータはある地域で行った観光動態調査であり、その調査項目の一つである観光客の全体的な満足度（以下「全体満足度」という）と他の設問項目との関連を分析することをテーマとする。

このように「満足（した・しなかった）」のような「質的データ」を判断基準（数量化理論では「外的基準」という）として、「個別項目の回答」のような「質的データ」を数量化する手法は、「数量化Ⅱ類」と呼んで区別している。

2.2. 数量化の流れ

ここでは、各項目の回答に点数付け、言うなれば配点を行い、その回答の有無で合計値を求めて、全体満足度の満足「した」「しなかった」を判定できるようにすることを目的とする。例えば、「景観・雰囲気良かったか」という項目に対して「はい」に 1 点、「いいえ」に -0.5 点、というように配点し、各回答者における各項目の合計値で全体満足度の「した」「しなかった」の判定が出せるようにする。

3. 実データへの適用

3.1. 個別の満足度との関連

3.1.1. 分析対象の概要

計算の元となる理論及び数式化の過程は専門書に譲ることにして、本稿では数量化Ⅱ類が実用的に使えるか、分析対象に実データを用いて検証する。

アンケートは、全体満足度を含めた 9 項目について、満足度を 6 段階で回答する形式である。これらの回答それぞれに配点を行い、合計値を算出する。ある回答者 A の配点例を表 3.1 に示す。表中の○印が A の回答箇所であり、その合計値で判定することになる。

表 3.1 ある回答者 A の配点例

項目	非常に満足	やや満足	どちらでもない	やや不満	不満	わからない	(無回答)
満足度-景観・雰囲気※	2.04				△10.97		
満足度-宿泊施設	4.29	13.95	△64.49	△98.25	0.00	△2.40	△2.55
満足度-おもてなし	5.78	5.80	△10.84	△39.39	4.48	4.48	0.89
満足度-特色	12.61	8.81	△9.81	△31.59	1.22	1.22	△8.71
満足度-旅行前情報	7.11	1.63	△12.59	7.32	△2.58	12.23	0.74
満足度-旅行中情報	△8.74	△6.52	8.45	23.07	△7.48	△12.30	2.57
満足度-飲食	5.12	13.95	△53.10	△101.15	△150.22	13.40	21.20
満足度-買い物	4.87	5.19	△3.28	△31.00	0.00	△4.26	△11.60
配点(○印)の合計値	△124.87						
全体満足度(実際の回答)	満足しなかった						

※分析の都合上、2 値とした。

なお、判定の対象とする全体満足度は「大変満足」「やや満足」と回答したサンプルを「満足した」とし、それ以外（「わからない」、無回

答を含む）を「満足しなかった」というように 2 値に置き換える。A の場合は「満足しなかった」である。

ひとまずこの配点をふまえた上で、全回答者の配点の合計値を表すグラフを図 3.1 に示す。グラフの折れ線のデータは以下の流れで作成した。

- 全体満足度「満足した」「満足しなかった」で回答者のグループを作る。
- 「満足しなかった」回答者を満足度の合計値の数値の小さい順に並べる。

- 小さい順に並んだ回答者の順位（全数を 1 としたときの割合）を縦軸に、回答の合計値を横軸に折れ線を作成する。
- 「満足した」回答者を数値の大きい順に並べる。
- 大きい順に並んだ回答者の順位（全数を 1 としたときの割合）を縦軸に、回答の合計値を横軸に折れ線を作成する。

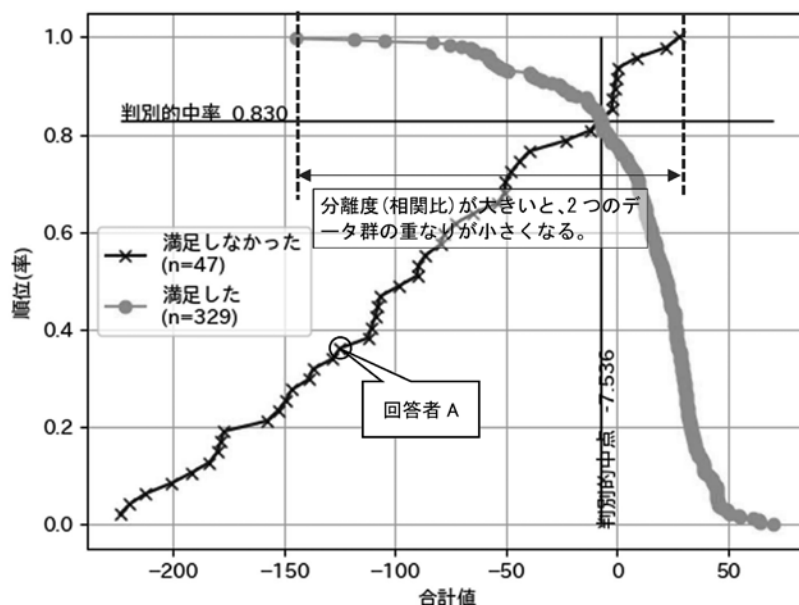


図 3.1 回答者の配点合計値と回答者順位（率）の関係（プロット図）

3.1.2 判定の精度

このグラフを作成することで、2つの折れ線の交点が、回答者の満足度の合計値で「満足した」「しなかった」を判断する境目、と考えることができる（「判別的中点」という）。この境目から右にある、すなわち合計値が大きい回答者は「満足した」グループの判定が成功となる。一方、境目から左にある（合計値が小さい）回答者は「しなかった」グループの判定が成功となる。

同時に、交点より下の回答者は「満足した」「しなかった」の判定がいずれも成功していることから、交点における順位（率）の値を「成功率」ということができる（「判別の中率」という）。

判別の中率は高いに越したことはないが、そ

もそも2つの折れ線の重なりが小さければ自ずと判別の中率は高くなる。この重なり（の小ささ、言うなればグループの「分離度」と連動する数値として「相関比」というものがある。相関比は0と1の間であり、大きいほどグループ分けがうまくできている、と言える。

本稿の例では判別の中率は0.830と0.75を超えておりそこそこの精度があると言えるが、相関比は0.486と0.5を下回っており、この点ではあまり精度が良くない結果となった。

3.1.3 満足度の判定に対する項目及び配点の影響

配点の合計値と配点の関係に着目すると、配点（合計値）が大きいほど「全体的に満足した」と判定さ

れるようになっている。したがって、項目における配点の最大値と最小値の差（「配点差」）は、

回答者の合計値を左右することから、判定との関連度とみなすことができる（表 3.2）。

表 3.2 個別の満足度の配点差

項目\回答	非常に満足	やや満足	どちらでもない	やや不満	不満	わからない	(無回答)	配点差
満足度-景観・雰囲気※	2.04				△ 10.97			13.01
満足度-宿泊施設	4.29	13.95	△ 64.49	△ 98.25	0.00	△ 2.40	△ 2.55	112.21
満足度-おもてなし	5.78	5.80	△ 10.84	△ 39.39	4.48	4.48	0.89	45.20
満足度-特色	12.61	8.81	△ 9.81	△ 31.59	1.22	1.22	△ 8.71	44.20
満足度-旅行前情報	7.11	1.63	△ 12.59	7.32	△ 2.58	12.23	0.74	24.82
満足度-旅行中情報	△ 8.74	△ 6.52	8.45	23.07	△ 7.48	△ 12.30	2.57	35.38
満足度-飲食	5.12	13.95	△ 53.10	△ 101.15	△ 150.22	13.40	21.20	171.41
満足度-買い物	4.87	5.19	△ 3.28	△ 31.00	0.00	△ 4.26	△ 11.60	36.19

※分析の都合上、2 値とした。

この表から、「満足度 - 宿泊施設」「満足度 - 飲食」は配点差が大きく、他の項目と比較して全体満足度の判定との関連度が大きいと言える。なお、「満足度 - 旅行中情報」については、全体的な満足との相関が非常に微妙なため、配点の計算結果は直観と逆になっている。一方で配点差は小さく、全体満足度の判定との関連度の小ささが反映されている。

結果への影響は小さい、という事象が発生している。これは、「多重共線性（マルチコリニアリティ、multicollinearity）」と言われる。分析ではどちらか一方を除外して多重共線性を切り離す。

表 3.3 相関関係にある項目での配点

項目\回答	触れた	触れなかった	配点差
史跡名所1	△ 43.26	15.04	58.29
風景2	57.29	△ 20.19	77.48

項目の組合せ	配点の合算	サンプル数 (全体の割合)
両方触れた	14.03	25.8%
両方触れなかった	△ 5.16	73.9%
風景2だけ触れた	72.33	0.3%
史跡名所1だけ触れた	△ 63.45	0.0%

3.2. 他の設問との関連

3.2.1. 観光資源

全体満足度との関連があるものとして、観光資源に触れる機会を挙げることができる。ここではそれらとの関連を分析する。

アンケートでは、史跡・名所、飲食店、物販店、みやげ物、文化・自然体験などの観光資源を選択肢として列挙し（「その他」を含む全 73 種類）、触れたものがあれば選択する（複数可）という形式で回答していただいた。回答は各観光資源 1 つにつき「触れた（行った、買った、体験したなど）」「触れなかった」の二者択一となる。

3.2.2. 多重共線性

「史跡名所 1」と「風景 2」については、配点がそれぞれ逆転した値となっている（表 3.3）。データを確認したところ、「風景 2」を訪れた人の 99%（=25.8/（25.8+0.3）*100）が「史跡名所 1」を訪れていた。その結果、配点差は大きいですが、回答者ごとに配点を合算した

3.3. 全体的な関連

観光資源、個別の満足度に回答者属性（年齢・性別、出発地など）を加え、改めて分析した結果として、設問の配点差における上位 10 項目を表 3.4 に示す。ここから、各回答との相対的な関連度の差を知ることができる。たとえば、「満足度 - 宿泊施設」と「満足度 - 飲食」の配点差は満足度以外の他の回答に比して大きく、表 3.2 における分析と同じ傾向である。ここから、「全体満足度は飲食、宿泊施設と関連が最も高く、他の要素との関連は相対的に小さい」ということができる。なお、判別の中率は 0.899 とまずまず、相関比は 0.532 で一応の精度を確保した。

表 3.4 全体的な項目の配点差 上位 10 項目

項目	非常に満足	やや満足	どちらでもない	やや不満	不満	わからない	(無回答)	配点差
満足度-飲食	7.17	13.41	△ 48.37	△ 87.72	△ 132.33	1.96	17.35	149.68
満足度-宿泊施設	9.25	25.87	△ 56.32	△ 79.72	0.00	△ 12.15	△ 11.87	105.59
満足度-特色	11.34	9.52	△ 5.84	△ 39.47	△ 13.14	△ 13.14	△ 8.08	50.81
満足度-おもてなし	3.37	4.23	△ 8.00	△ 36.34	1.87	1.87	7.23	43.57
満足度-買い物	5.47	9.03	△ 7.14	△ 31.63	0.00	△ 15.27	△ 11.67	40.66
満足度-旅行中情報	△ 0.03	△ 7.90	9.38	18.83	10.16	△ 12.63	△ 0.96	31.46
温泉5	行った △ 19.02	行かなかった 5.64						24.66
出発地	南大阪 12.53	北大阪 2.99	県内 △ 8.87	近畿 △ 10.48	その他 △ 3.25			23.01
年齢	30代以下 14.57	40代 △ 2.19	50代 0.63	60代 △ 7.07	70代以上 2.90			21.64
旅行の主目的	温泉 4.99	宿泊 △ 13.95	飲食・買い物 △ 6.56	その他 △ 4.43				18.93

一般的に、説明変数を増やせば増やすほど判別の中率及び相関比は増加する。ただし、計算の所要時間は加速度的に長くなること、変数間に相関があると多重共線性が発生して、各回答と全体満足度の関係を判断することが難しくなることなどのデメリットも生じる。

4. まとめと今後の課題

以上のことから、アンケートを全体的に分析、評価する手法として数量化理論（Ⅱ類）が実用面から適用が可能であることを明らかにすることができた。一方で、多重共線性とそれに伴う説明変数の選び方など注意すべき点があることも明らかになった。

説明変数の選び方に関して実務においては、説明変数の追加・削除、再計算、他の説明変数への影響をチェック、という作業の繰り返しになる。このプロセスに、改善の余地があると考ええる。

それらをふまえながら、クライアントにより良い情報を提供できるよう、今後も研鑽に努めたい。

参考文献：

涌井良幸、涌井貞美「多変量がわかる」技術評論社,2016

菅民郎「多変量解析の実践（下）」現代数学社,1993

酒井隆「図解 アンケート調査と統計解析がわかる本 [新版]」日本能率協会マネジメントセンター,2015

岡本安晴「いまさら聞けない Python でデータ分析」丸善出版,2019

堀田敬介「判別分析・数量化Ⅱ類・数量化Ⅲ類」,文教大学 Hotta Lab. 講義資料 最新版資料 倉庫 (https://www.bunkyo.ac.jp/~hotta/lab/courses/2003seminar/ch4-4_5_hotta.pdf)

福井正康・細川光浩「社会システム分析のための統合化プログラム7ー多変量解析ー」福山平成大学経営学部紀要 (<https://www.heisei-u.ac.jp/ba/fukui/pdf/analysis7.pdf>)