



第VI部

調査結果のまとめ方

第14章 調査が終わったら

14.1 報告書作成の流れ

 図 14.1 は、調査票を回収してから報告書を作成するまでの流れを示したものである。

 まず、回収した「調査票の点検・編集」によって、不適当な調査票や回答を無効とするなどする (14.2 節参照)。

次に、「集計のために回答を数値化」し (14.3 節参照)、数値化した「回答の入力」

を行って (14.4 節参照)、集計用データを作成する。

「回答の集計と集計結果のグラフ化」によって得られた「集計結果の検討」を行い、場合によっては、別の観点から再度「回答の集計と集計結果のグラフ化」を行う。

そして最終的に、検討結果を基に「報告書の作成」を行う (第 15 章参照)。

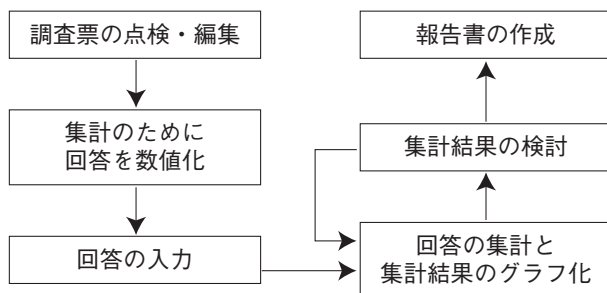





図 14.1: 報告書作成の流れ

14.2 回答を編集する—エディティング—

 回収した調査票やその回答を点検し、不適当な調査票は取り除いたり、指示に反した回答を無効としたりすることを **エディティング** あるいは **データクリーニング** と呼ぶ。

 用意した標本あるいは調査票は、調査を実施した結果に応じて、**回収票** と **不能票**（未回収票）とに分類することができる。さらに回収票は、集計に用いる **有効票** と **無効票** とに分類することができる。

- 回収票：他記式調査法において調査に応じた調査対象の調査票、あるいは自記式調査法において回収した調査票
 - － 有効票：無効票以外の回収票
 - － 無効票：調査対象以外が回答した調査票、一部あるいは全ての回答が記入されていない調査票
- 不能票：回収票以外の調査票

 次に、特に自記式調査法（5.1 節参照）では、有効票の回答内容を点検し、場合によっては以下のようなエディティングを行う。

- 回答がない調査項目は、「無回答」あるいは「D.K.（わからない）」などとする。


- 単数回答法（9.2.2 節参照）の調査項目に対して、複数の選択肢を選んでいる場合、制限回答法の調査項目に対して、制限を超えた数の選択肢を選んでいる場合には、「その他」あるいは「D.K.」などとする。
- 選択肢「その他」に具体的に記入された内容が、明らかに選択肢として既に用意されていれば、その選択肢を選んだものとする。
- 数値を記入する自由回答法では、回答者が数値の単位を間違えていないか点検・確認する。
- 濾過質問（11.7 節参照）に関する項目の回答状況を点検し、濾過質問に対する回答を修正したり、濾過質問後の項目を「無回答」としたりする。




ワンポイント アドバイス


論理的にあり得ない回答の点検（これを **論理チェック** と呼ぶ）は、人が行うよりも、コンピュータが行う方が確実である。そのため、とりあえず回答通りにデータ入力（14.4 節参照）を行い、コンピュータプログラムによってエディティングを行うのがよい。

14.3 回答を数値化する—コーディング—


 個々の回答に対して、集計のため数値を与えることを **コーディング** と呼ぶ。また、コーディングの方法を整理したものを **コードブック** と呼ぶ。

 選択肢回答法 (9.2.1 節参照) の項目のうち、単数回答法 (9.2.2 節参照) では、調査票を作る段階で、あらかじめ各選択肢に数値を割り振っておけばよい。例えば、選択肢「当てはまる」は「1」とし、選択肢「当てはまらない」は「2」とする、等である。なお、「無回答」や「D.K.」などに対しても、例えば「-99」など何らかの数値を割り当てる。

複数回答法の場合には、一つの実験肢に一つの数値を与えるとよい。例えば、その選択肢が選ばれていれば「1」、そうでなければ「0」とすると、その合計値が、選択肢を選んだ調査対象の数となって便利である。

 自由回答法のうち、統計的な集計を行う調査項目については、回答のコーディングを行う。自由意見欄 (11.12 節参照) の内容などは、コーディング・集計するよりも、そのままの文章の方が有意義な情報となることが多い。


コーディングをする場合には、まず全ての回答内容に目を通すことが必要である。

 自由回答法によって人数や金額、日

数などの数値の記入を求める項目では、集計方法に応じて二通りのコーディング方法がある。


数値の平均値や総計値を求めることが目的であれば、有効な回答はそのまゝの数値とする。「無回答」や無効な回答に対しては、有効な回答と混同しない数値を与える。


範囲を区切っていくつかのカテゴリ (9.3 節参照) に分類する場合には、あるカテゴリだけに回答が集中しないような区切りを設定する。設定するカテゴリの数は、調査目的と標本サイズによって決める。また「無回答」あるいは「無効」などといったカテゴリも設定する。そして、各カテゴリに例えば「1」から順に数値を割り当てる。

 自由回答法によって文字や文章の記入を求める項目では、回答をコーディングするために、まず **KJ法** (4.1 節参照) などを用いて、回答内容をいくつかのカテゴリに分類する。カテゴリの数は、数値の場合と同様に、調査目的と標本サイズによって決める。「その他」や「無回答」といったカテゴリも設定するとよい。

それぞれの回答をカテゴリに分類する作業は、できれば複数の人が別々に行い、その結果をつきあわせるとよい。不一致があまりに多い場合には、カテゴリの作り方を再検討すべきである。

14.4 回答を入力する—データ入力—

 多くの場合、結果の集計や分析はコンピュータ上で行うことになる。そのためには、コンピュータにデータを入力しなければならない。


 データ入力には、例えば図 14.2 のような表計算ソフトを用いると便利である。

通常、一つの調査対象の回答を一行とする。図 14.2 は五番目の調査対象のデータを入力しているところである。

また、一つの調査項目を一行とする。ただし、複数回答法 (9.2.2 節参照) の場合には、一つの選択肢を一行とする。図 14.2 では、「性別」や「年齢層」といった単数回答法の調査項目はそれぞれ一行とし、複数回答法である「学習活動の内容」は、「一般教養」や「趣味」といった選択肢を一行としている。

そして各セルには、コーディング (14.3

節参照) された数値をそのまま入力すればよい。

 入力を業者に委託する場合には、納品のためのデータフォーマットをあらかじめ指定するとよい。なお、入力に不要な情報は業者には一切渡さないようにし、入力データは納品後直ちに破棄させる契約を交わすべきである。さらに、調査票や入力データが外部に漏れないよう、その受け渡し方法も含め、秘密保持には細心の注意が必要である (13.6 節参照)。




ワンポイント アドバイス

データ入力は、簡単そうに見えて意外に間違いやすいものである。二人以上の人が同じ調査票を別々に入力し、それを付き合わせるなど、ダブルチェックを行うとよい。

	A	B	C	D	E	F	G
1							
2	番号	性別	年齢層	学習活動の有無	一般教養	趣味	技術
3	001	1	2	2	0	0	0
4	002		2	3	1	1	0
5	003		2	5	1	0	1
6	004	1	4		1	0	0
7	005						
8							
9							
10							
11							
12							
13							


図 14.2: データ入力


14.5 無回答はどうするか

 調査においては、二種類の無回答がある。


一つは、調査対象単位での無回答である（これを **unit-nonresponse** と呼ぶ）。標本として選び出した調査対象が実際には見つからなかったり、調査対象が調査への協力を拒否したりした場合などに、この無回答が生じる。不能票や無効票（14.2 節参照）が、この無回答に相当する。

もう一つは、調査項目に対する「無回答」や「D.K. (わからない)」などである（これを **item-nonresponse** と呼ぶ）。この無回答は、質問の内容を理解できない、回答を選択肢の中から選ぶことができない、質問に答えたくない、誤った回答方法をしたり回答しなかった、といった場合に生じる。

 いずれの無回答に対しても、その対処法には大きく二つの方法がある。一つは、無回答をそのまま無回答として扱い、集計から取り除いたり、無回答として集計したりする方法である。もう一つは、無回答部分をそれ以外の回答部分を基に推測し、無回答を推測した回答で置き換えて集計する方法（これを代入法と呼ぶ）である。


 調査対象単位での無回答（不能票や無効票）の場合には、無回答をそのまま無回答として扱うことが多い。つまり、不能票や無効票は集計から完全に除外し、有効票だけを集計することが多い。


また、調査項目に対する無回答のうち、選択肢回答法の場合や、自由回答法の回答をコーディング（14.3 節参照）した場合にも、無回答をそのまま無回答として扱うことが多い。つまり、無回答がある調査票も集計に含め、無回答は「無回答」あるいは「D.K.」という選択肢を選んだものとして扱うことが多い。

 一方、調査項目に対する無回答のうち、人数や金額などの数量を自由回答法によって回答してもらい、その母集団平均値や総計値を推定しようとする場合には、次のいずれかの方法をとることが多い。つまり、有効票の中からさらに有効回答がある調査票だけを取り出して集計する方法と、無回答を例えば「0」などの数値で置き換え、全ての有効票を用いて集計する方法である。どちらの方法がよいかは、調査内容による。


第15章 報告書作成の基本原則

15.1 結果だけでなく方法も報告する

 社会教育調査の目的は、統計的方法によって母集団に関する情報を得ようとすることである (1.2 節参照)。

 そのため調査報告書には、母集団推定値 (8.1 節参照) だけでなく、その数値を得た方法についてもできる限り開示するのがよい。結果数値が母集団に関する情報としてどの程度信頼できるものであるかは、その数値を得た方法によって決まるからである。

別の言い方をすれば、調査結果の数値には誤差が含まれている (1.5 節参照)。誤差のうち、標本誤差は標本サイズを基にその大きさを理論的に数値で示すことができるが、非標本誤差の大きさは算出することができない。非標本誤差の大きさを報告書の読者に判断してもらうためには、どのように調査を実施し結果数値を得たのかといった方法を示さなければならないのである。

 調査報告書には、母集団推定値とし

ての調査結果の数値以外にも、以下の項目が必要である。


1. 調査の目的
2. 調査主体と調査実施者
3. 母集団
4. 標本抽出方法
5. 標本サイズ
6. 調査実施時期
7. 調査実施方法
8. 回収率
9. 調査票




ワンポイント アドバイス


報告書に調査結果の数値を得た方法の記述があると、比較調査を企画する場合にも役立つ。調査結果は、調査方法をわずかに変えるだけでも、大きく変わってしまうことがあるからである。

15.2 個別の回答は公表しない

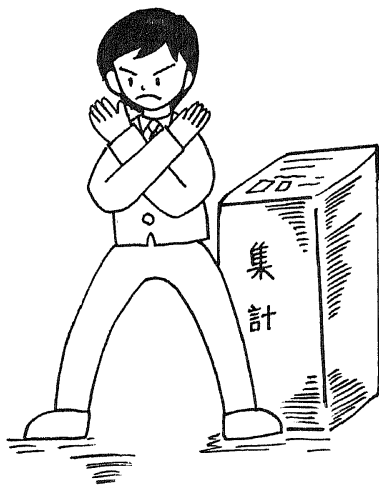
 統計的方法 (1.2 節参照) による調査では、調査結果としては、母集団推定値 (8.1 節参照) としての集計結果のみを公表し、個別の回答は公表しないのが原則である (13.6 節参照)。

自由回答法 (9.2.1 節参照) による回答や、自由意見 (11.12 節参照) の内容などをそのまま公表する場合には、回答者を特定できない形とするのが原則である。


 標本調査 (1.4 節参照) では、どの調査対象が標本として選ばれたのか、という情報も公表しないのが原則である。集落抽出法 (7.3 節参照) や多段抽出法 (7.4 節参照) では、それぞれの段において選ばれた抽出単位の情報についても、基本的には公表しないのがよい。選ばれた抽出単位を公表することで、最終的に標本となった調査対象がある程度の範囲に絞られてしまうことがあるからである。


 クロス集計 (16.4 節参照) を行う場合には、各セルに入る標本サイズが小さく

なり過ぎないように注意する必要がある。極端な場合を挙げれば、あるセルに入る調査対象が一つとなってしまうと、もしその調査対象が特定されてしまえば、個別の回答も知られてしまうからである。例えば、社会教育施設を調査対象とし、施設規模とのクロス集計を行ったとする。一般に、大規模な施設は数が少なく、そのため大規模施設の個別の回答は、クロス集計の結果からある程度推測できてしまう。



15.3 数値の表現に注意する

 調査結果を数値で表示する場合には、その桁数に注意する必要がある。例えば、パーセントであれば、普通、小数第一位を四捨五入し整数に丸めるのがよい。


 その理由は、調査結果の数値には標本誤差や非標本誤差 (1.5 節参照) が含まれるからである。表 8.1 から分かる通り、標本誤差だけでも、その大きさは数ポイントとなることが多い。したがって、小数点以下の小さな値は誤差の範囲内と考えるべきである。細かな数値を出しても結果がより正確になるわけではない。

また、可能であれば、理論的に求めた標準誤差 (8.2 節参照) の大きさも提示するとよい。




ワンポイント アドバイス

パーセントを整数表示すると、それがパーセントなのか実際の数なのか分かりにくくなることもある。そのような場合には、あえて小数点以下第一位まで表示し、パーセントであることを分かりやすくすることもある。

 また、調査結果を示す際には、その結果を得るために用いた標本サイズを明示する必要がある。例えば、パーセントの場


合には、それが何に対する割合なのかを明確にしなければならない。濾過質問 (11.7 節参照) などによって一部の調査対象のみが回答した場合、パーセントが、全体に対する割合なのか、回答すべき調査対象に対する割合なのか、分からなくなるおそれがあるからである。また、標本サイズを明示することは、無回答をどのように扱ったのかを示すことにもなる (14.5 節参照)。

結果を得るために用いた標本サイズは、($n=〇〇〇$) などと表示することが多い。

 パーセントの差は、**ポイント** という単位で呼ぶ。例えば、40%から50%に増えたことを言う場合、「10ポイントの増加」と言う。「10%の増加」とは言わない。その理由は、40%から10%増加した、と言ってしまうと、50% ($=40+10$) となったのか、44% ($=40 \times 1.1$) となったのかが分からなくなってしまうからである。



15.4 適切なグラフを選ぶ

 報告書で主張したい調査結果は、数値として示すだけでなく、グラフを描いて示すとよい。また、調査結果を検討する際にも、数値を見るよりもグラフを眺める方がよい。細かな差違にとらわれず、大局的に結果を見ることができからである。

なお、グラフに数値を付ける場合には、値の小さな数値や小数点以下などの小さな


位は表示しないのがよい。グラフは大まかな傾向を見るためのものだからである。




ワンポイント アドバイス

グラフのタイトルは、グラフの下に付けるのが原則である。なお、表のタイトルは、表の上に付ける。

15.4.1 円グラフ

 **円グラフ** は、単数回答法 (9.2.2 節参照) の結果を一つだけ示すのに適している。

 段階評定法 (9.2.3 節参照) など、カテゴリ (9.3 節参照) に自然な順序があれば、その順序に従って右回りにカテゴリを並べる。

カテゴリに順序がなければ、割合の大きなカテゴリから順に並べるとよい (図 15.1

左)。ただし、例えば図 15.1 右の「ひとりで」のように、割合の大きなカテゴリを円グラフの下部に配置すると、そのカテゴリの割合が大きい、ということ的印象づけるグラフとなる。また、強調したいカテゴリの部分だけ切り離して描く方法もある。

「その他」や「無回答」などは、右回りの順で最後に配置する。

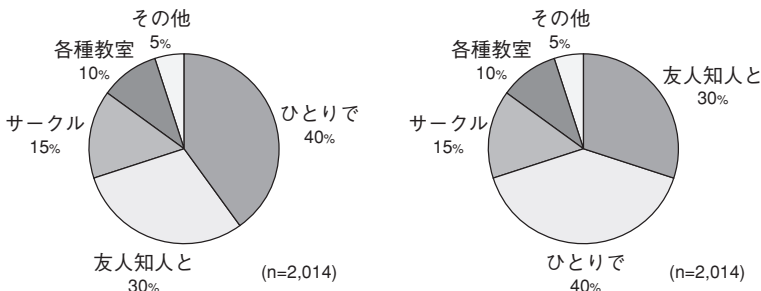




図 15.1: 円グラフ

15.4.2 帯グラフ

 **帯グラフ** は、単数回答法 (9.2.2 節参照) の結果が複数あるとき、その間の比較を示したり、クロス集計結果 (16.4 節参照) を示すのに適している。帯グラフが一本となる場合には、円グラフ (15.4.1 節参照) を用いる方が見やすいことが多い。

 一本の帯の中でのカテゴリ (9.3 節参照) の配列は、カテゴリに自然な順序があれば、その順序に従ってカテゴリを並べる。カテゴリに順序がなければ、割合が大きい順に左から右へ並べる、最も強調したいカテゴリを左端に置く、帯の間での割合の違いが大きいカテゴリ (図 15.2 では、「ひとりで」や「サークル」など) を両端に置く、といった配列にするとよい。

複数の帯の配列は、帯の間に順序があれ

ば、その順序に従って並べる。順序がなければ、帯の間での、最も強調したい違いが明確になるよう並べるのがよい。例えば、結果が似ている帯どうしは隣りに置く、帯の左端のカテゴリの割合が大きい帯から小さい帯へ並べる、といった配列が有効である。



ワンポイント アドバイス

一本の帯に含まれるカテゴリの数が多くなりすぎると、見にくいグラフになってしまう。その場合は、割合が小さなカテゴリをいくつかまとめてカテゴリの数を減らす、あえて棒グラフ (15.4.3 節参照) を用いる、といった方法をとるとよい。

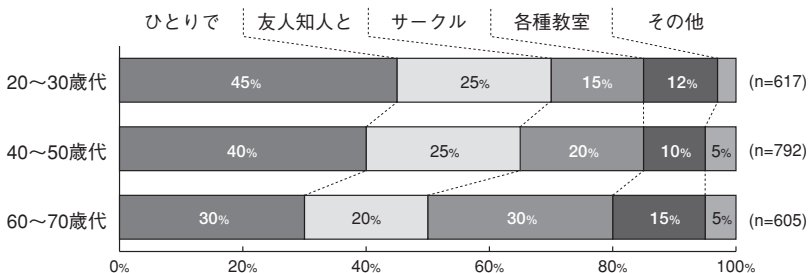





図 15.2: 帯グラフ

15.4.3 棒グラフ

 **棒グラフ** は、複数回答法 (9.2.2 節参照) の結果やクロス集計の結果 (16.4 節参照) を示すのに適している。なお、人数や金額などの量的データ (9.3 節参照) をコーディング (14.3 節参照) した結果は、ヒストグラム (15.4.4 節参照) を用いるべきである。

 複数回答法では、カテゴリ (9.3 節参照) の間に順序がないことが多い。そこで、グラフでは、選ばれた割合が大きい順にカテゴリを並べるとよい。ただし、「その他」や「どれにも当てはまらない」などのカテゴリは最後に置く。

 複数の棒グラフを同時に表示するには、主に三つの方法がある。

一つ目は、「スポーツ」や「趣味」といったカテゴリごとに棒を並べる方法である (図 15.3 左)。この場合、「スポーツ」や「教養」は男性が多く、「趣味」や「健康」は女性

が多い、といった各カテゴリの中での比較に重点を置くことになる。

二つ目は、カテゴリの配列順序を同じにした棒グラフを並べる方法である (図 15.3 右)。この場合、「男性」の中ではスポーツが最も多く、「女性」の中では趣味が最も多い、といった複数の棒グラフの間での傾向の違いに重点を置くことになる。

三つ目は、棒を積み重ねて描く方法である。



ワンポイント アドバイス

棒グラフに限らず、グラフの中でのカテゴリの配置順序は、自然な順序がある場合を除いて、必ずしも調査票通りとする必要はない。グラフは、調査の結果明らかになったことを主張するために用いるものであるから、主張内容が一目で分かるようなグラフを作成するのがよい。

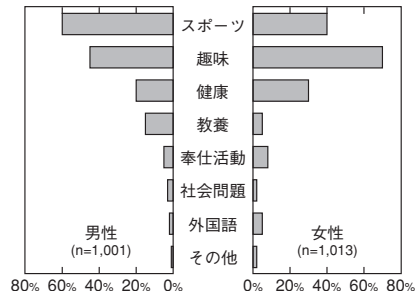
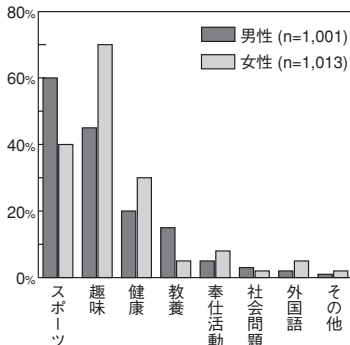



図 15.3: 棒グラフ

15.4.4 その他のグラフ

 **折線グラフ** は、変化を示すのに適している。同じグラフであっても、縦軸の長さによってグラフから受ける印象が変わってしまう場合があることに注意する必要がある。

散布図 は、数量で表される二つの項目の間の相関関係 (15.5 節参照) を示すのに適している。散布図上の点が、対角線上に集まっているほど直線的な相関関係が強いことを示す。直線的な相関関係の強さは、相関係数 (16.5 節参照) という指標によ

て表される。

レーダーチャート は、同じ単位で表される五から八つ程度の項目の間の比較を示すのに適している。複数個のレーダーチャートを並べたり、重ねて描くこともある。

ヒストグラム は、数量の頻度を示すのに適している。例えば、図 15.4 右下は、0 点以上 10 点未満が 15 人いるということを示している。数量を区切る階級幅の取り方によって、グラフの形状が変わってしまうことに注意する必要がある。

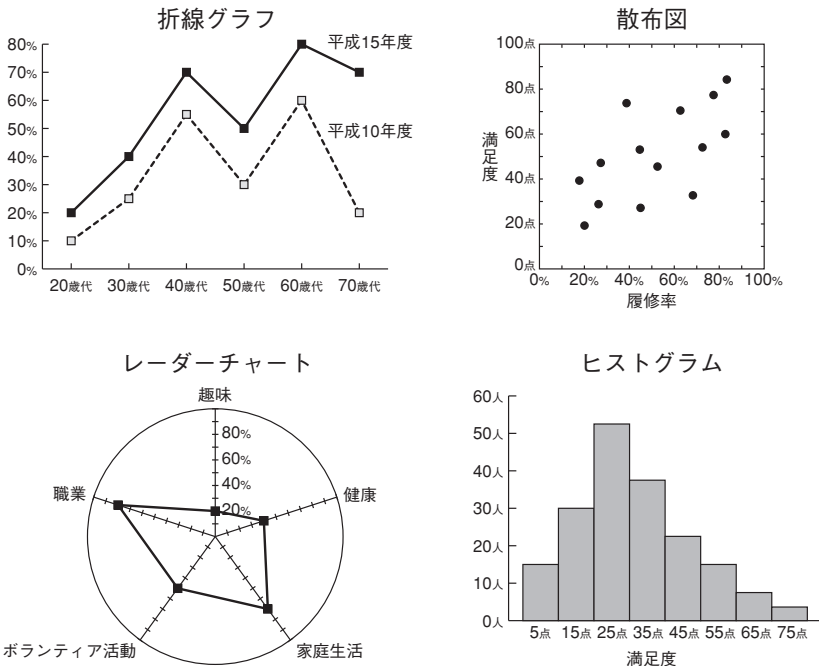




図 15.4: その他のグラフ

15.5 相関関係を因果関係としない


 **相関関係**とは、二つの事柄の間での、一方が変化すれば、もう一方も変化するような関係のことを言う。これに対し、**因果関係**とは、一方がもう一方の原因となっている関係のことを言う。

調査では、普通、相関関係しか調べられない。調査の結果、相関関係が見られたからといって、ただちに因果関係があると結論づけることはできない。相関関係が因果関係の直接の証拠になるとは限らないからである。

 例えば、学習活動の有無とボランティア活動の有無とのクロス集計(16.4節参照)を行ったところ、ボランティア活動を行っている人ほど学習活動を行っている割合が高い、という相関関係が見られたとする。しかし、この結果から直ちに、ボランティア活動を行うと学習活動も行うようになる、あるいは、学習活動を行うとボランティア活動も行うようになる、といった因果関係を結論づけることはできない。


なぜなら、例えば、好奇心が強く何事にも積極的な人ほど、その積極性のために学習


活動も盛んであり、また同時にボランティア活動も行っている割合が高いであろう。この場合、学習活動の有無やボランティア活動の有無は、積極性の有無が原因となっているのであって、学習活動とボランティア活動のいずれかがもう一方の原因となっているのではない。

 横断的方法(5.8節参照)である一回限りの調査では、因果関係を調べることは難しい。因果関係をより慎重に調べるためには、調査目的としている事柄以外の要因を統制した実験的方法や、パネル調査法などの縦断的方法を用いる必要がある。




15.6 比較を心がける

 統計的方法の特徴は、結果が数値で表されることである (1.2 節参照)。しかし単一の数値は、その大きさを評価しにくい。複数の数値を比較提示し、それらの大きさを相対的に評価できるよう心がけるとよい。

 例えば、「ボランティア活動への参加経験率は 10% である」という結果が得られたとする。10% という数値を、ある人は高いと感じ、別の人は低いと感じるかもしれない。

高いと感じた人は、「多分数% 程度であろう」といった事前の予想を心の中に持っており、その予想との比較によって「10% は高い」と感じたのである。逆に、低いと感じた人は、「3 割はいるのではないか」といった心の中の期待との比較によって、「10% は低い」と感じたのである。

このように、10% という数値の大きさを評価するためには、他の数値との比較が必要と言える。

 当然、人が心の中に持っている数値は人によって異なるため、他の客観的な数値が必要である。

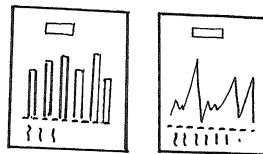
例えば、

- 過去の調査や他の調査主体が実施した調査との比較
- 性・年齢・地域等の異なる属性の間での比較
- 調査項目の間での比較 (項目 A の経験率と項目 B の経験率など)

などといった比較が考えられる。


報告書で結果数値を示すときには、一つの数値だけを単独で示すだけではなく、このような比較ができるような形でも示すとよい。数値の持つ意味がよりはっきりとし、情報として価値が高まるからである。

また、あらかじめ調査の企画段階において、このような比較ができるような調査計画を立案しておくとうよい (4.5 節参照)。




第16章 調査結果の集計・分析方法

16.1 集計・分析用語の基礎知識 (その1)

 調査によって知ろうとしている特性のことを **変数** と呼ぶ。例えば、学習経験の有無、生涯学習に関する情報源、性別、年齢、公民館で実施している事業プログラムの数、などである。

大きさ N の母集団において、各調査対象は変数の値をそれぞれ持っており、それらを X_1, \dots, X_N と表す。例えば、母集団の1番目の人が46歳であれば、年齢という変数では、 $X_1 = 46$ である。


 **平均値** (算術平均値) は、変数の値の状態を示す代表値の一つである。例えば、母集団平均値は、

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N} = \frac{1}{N} \sum_{i=1}^N X_i$$

によって求められる。

パーセントなどの割合も、0または1をとる変数の平均値として表すことができる。例えば、通学合宿の実施率 (1.2 節参


照) は、通学合宿を実施していれば1、実施していなければ0という変数の平均値に一致する。

 変数の値の状態を示す代表値としては、他に、**中央値** や **最頻値** などがある。

中央値とは、変数の N 個の値を小さい方から順に並べていったとき、ちょうど真ん中の順番に位置する値のことである。例えば、 $N = 15$ であれば、小さい方から8番目の値、 $N = 14$ であれば、小さい方から7番目と8番目の値の平均値となる。変数の値の中に、大きく外れた値が含まれていると、平均値はその影響を大きく受けるが、中央値は影響が小さい。

最頻値とは、 N 個の値のうち、度数 (16.3 節参照) が最も多い値のことである。つまり、棒グラフ (15.4.3 節参照) を描いたとき、最も棒が長くなる値 (数値とは限らない) のことである。

16.2 集計・分析用語の基礎知識 (その 2)


 **分散** は、変数の値のばらつきを示す指標の一つである。例えば、図 16.1 に示す二つのヒストグラム (15.4.4 節参照) を比べると、どちらも平均値 (16.1 節参照) は同じであるが、分布 A に比べて分布 B の方がばらつき・分散は大きい。

母集団分散は、

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

によって求められる。

なお、分散の平方根を **標準偏差** と呼ぶ。


 社会教育調査の目的の一つは、調査結果を集計し、母集団平均値を知ることである (1.4 節参照)。そのための推定量 (8.1 節参照) として最もよく用いられているのは、**HT 推定量** (Horvitz-Thompson 推定量) である。

母集団平均値に対する HT 推定量は、大きさ n の標本の変数の値を x_1, \dots, x_n とし、その包含確率 (6.2 節参照) を

π_1, \dots, π_n とすると、

$$\text{HT 推定量} = \frac{1}{N} \sum_{i=1}^n \frac{x_i}{\pi_i}$$

によって与えられる。

 包含確率が全て等しい自動加重標本 (6.1 節参照) では、母集団平均値に対する HT 推定量は、標本平均値に一致する。つまり、調査の結果得られた平均値が、そのまま母集団平均値の推定値となる (8.1 節参照)。例えば、単純無作為抽出 (7.1 節参照) では、 $\pi_i = n/N$ なので、

$$\text{HT 推定量} = \frac{1}{n} \sum_{i=1}^n x_i$$

である。

また、標本における分散が、そのまま母集団分散の推定量となる。そこで、推定した母集団分散を用いて、推定量の標準誤差 (8.2 節参照) を求めたり、信頼区間 (8.3 節参照) を求めることができる。

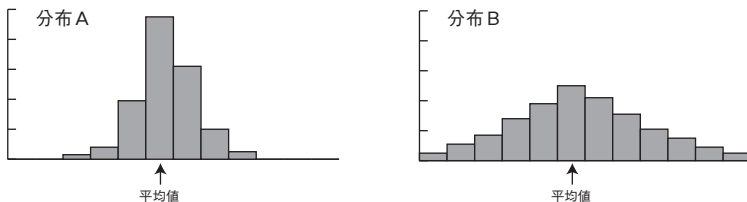




図 16.1: 分散

16.3 集計・分析の基本—単純集計—

 **単純集計**は、選択肢回答法 (9.2.1 節参照) などによって得られた質的データ (9.3 節参照) の集計方法の一つである。カテゴリごとに、そのカテゴリを選んだ調査対象の数 (これを **度数**と呼ぶ) やその割合を計算する。結果をまとめた表 (表 16.1*参照) は、単純集計表、GT 表、度数分布表などと呼ぶ。

 単純集計表の合計欄は、度数の場合には、有効票 (14.2 節参照) の数、あるいは濾過質問後の項目などでは、その項目に回答すべき調査対象の数を示す。パーセントの場合には、表中のパーセントの合計を示すのがよい。

単数回答法 (9.2.2 節参照) では、パーセントの合計は、理論的には 100 となるが、数値を丸めたために合計が 100 とならない場合もある。そのときは、無理に 100 に

合わせる必要はない。

複数回答法の場合、表中のパーセントとしては、度数の合計欄に対するそれぞれのカテゴリの度数の割合を示す。そのパーセントの合計を 100 で割った値は、一つの調査対象が選んだカテゴリ数の平均値に一致する。例えば、パーセントの合計が 238 であれば、調査対象当たり 2.38 個のカテゴリを選んでいる、ということが分かる。そのため、複数回答であるということを明記した上でパーセントの合計を示すとよい。


 パーセントは、そのカテゴリに当てはまれば 1、当てはまらなければ 0 という値をとる変数の平均値に一致する (16.1 節参照)。そのため、自動加重標本 (6.1 節参照) であれば、標本におけるパーセントは、母集団におけるパーセントの推定値に一致する。


表 16.1: 単純集計表


日常生活の中で、あなたは、若い世代の人たちとの交流がありますか。

	よくある	時々ある	あまりない	全くない	無回答	合計
度数	414	764	371	69	26	1,644 名
%	25	46	23	4	2	100 %

*世代間交流活動研究会「青少年及び高齢者の異世代に対する意識調査報告書」平成 12 年 1 月 より


16.4 二つの変数の間の関連を知りたい (その 1) — クロス集計 —

 **クロス集計** は、二つの質的な変数 (9.3 節参照) の間の関連を知るための集計方法の一つである。一方の変数のあるカテゴリと、もう一方の変数のあるカテゴリとを同時に選んだ調査対象の数やその割合を計算する。結果をまとめた表 (表 16.2[†] 参照) は、クロス集計表あるいは **分割表** などと呼ぶ。

 クロス集計表のパーセントは、度数を横に合計したものに対する割合 (これを **行パーセント** と呼ぶ)、あるいは縦に合計したものに対する割合 (これを **列パーセント** と呼ぶ) を示すことが多い。

行パーセントあるいは列パーセントの間の比較をすることで、二つの変数間にどのような関連があるのかを知ることができる。例えば、表 16.2 では、「よくある」あ

るいは「時々ある」に対する男性の行パーセントが、女性のそれよりも小さいことから、男性の方が女性に比べ、若い世代との交流が少ないことが分かる。

 調査結果の数値には誤差が含まれている (1.5 節参照)。特にクロス集計では、パーセントを求めるときの標本サイズが、単純集計を行うときよりも小さくなる。そのため、標本誤差がより大きくなっている (8.2 節参照) ことに注意する必要がある。

パーセントの間の差が、誤差の範囲内なのかどうかを調べるためには、統計的仮説検定 (16.9 節参照) を行えばよい。二つの変数の間に関連があるかどうか、つまり行パーセントや列パーセントの分布が異なるかどうか、は検定方法の一つである χ^2 検定によって調べることができる。


表 16.2: クロス集計表

日常生活の中で、あなたは、若い世代の人たちとの交流がありますか。


		よくある	時々ある	あまりない	全くない	無回答	合計
男	度数	182	343	202	44	13	784 名
	行%	23	44	26	6	2	101 %
女	度数	222	408	155	24	12	821 名
	行%	27	50	19	3	1	100 %

[†] 世代間交流活動研究会「青少年及び高齢者の異世代に対する意識調査報告書」平成 12 年 1 月 より

16.5 二つの変数の間の関連を知りたい (その2)—相関係数—

 **相関係数** は、二つの量的な変数 (9.3 節参照) の間の直線的な相関関係 (15.5 節参照) を示す指標の一つである。


二つの変数を用いて散布図 (15.4.4 節参照) を描いたとき、例えば図 16.2 の左端のような図や、右から二番目のような図が得られれば、二つの変数の間には直線的な関連がある、あるいは直線的な相関関係がある、と言える。一方、例えば図 16.2 の左から二番目のような図が得られれば、二つの変数の間には相関関係は見られない、と言える。

 相関係数は、二つの変数の値を X_1, \dots, X_N と Y_1, \dots, Y_N としたとき、

$$R = \frac{1}{N-1} \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sigma_X \sigma_Y}$$

によって求められる。ただし、 σ_X と σ_Y はそれぞれ変数 X と変数 Y の標準偏差 (16.2 節参照) である。

相関係数の最大値は 1、最小値は -1 であり、相関係数の絶対値が大きいほど強い相関関係を意味する。ただし、図 16.2 の右端のように、相関関係はあってもそれが直線的ではない場合には、相関係数が大きくなるとは限らない。

 なお、質的な変数と量的な変数との間の関連を示す指標としては、**相関比** などがある。



ワンポイント アドバイス

二つの変数間の関連を調べるためには、相関係数を求めるだけでなく、散布図を描くことも重要である。図 16.2 右端のように曲線的な関連は、相関係数ではとらえられない。また、一つだけ大きく外れた値がある場合には、実質的な関連がないにもかかわらず、相関係数が大きくなることもある。

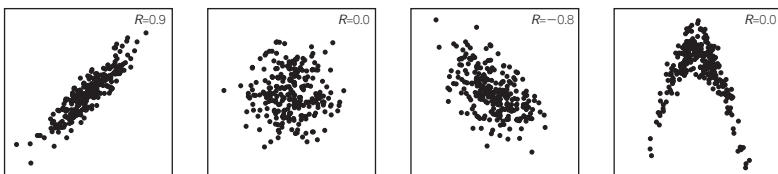




図 16.2: 相関係数

16.6 散布図に直線を引きたい—回帰分析—

 **回帰分析** は、ある一つの量的変数（これを **基準変数** と呼ぶ）と、その他いくつかの変数（これを **説明変数** と呼ぶ）との関係を、総合的かつ簡潔にとらえようとする方法の一つである。

特に、説明変数が、一つの量的変数の場合を **単回帰分析**、二つ以上の量的変数の場合を **重回帰分析**、質的変数の場合を **数量化 I 類** と呼ぶ。

 二つの量的変数があるとき、その間の直線的な相関関係 (15.5 節参照) の強さは、相関係数 (16.5 節参照) によって表すことができる。もし、相関係数の絶対値が大きければ、普通、散布図 (15.4.4 節参照) 上にプロットされる点は、ほぼ直線的に並んでいる。単回帰分析は、この散布図上の点の集まりの中心を通るように直線を引くことで、二つの変数の間の関係を簡潔にとらえようとする方法である。

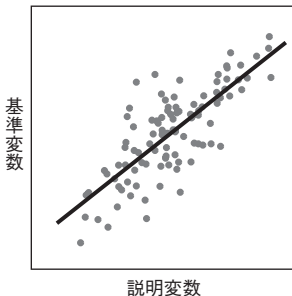



図 16.3: 単回帰分析

また、二つの量的変数の間に強い相関関係があれば、一方の変数（説明変数）の値を基に、もう一方の変数（基準変数）の値をほぼ言い当てることのできるであろう。つまり、散布図の横軸の値を基に、縦軸の値を **予測** することになる。単回帰分析は、実際の基準変数値からなるべく外れないよう、予測値（散布図上の直線）を求める方法であるとも言える。

単回帰分析によって引かれる直線の式は、基準変数の予測値を \hat{Y} (Y ハットと読む)、説明変数の値を X_1 とすると、

$$\hat{Y} = a + b_1 X_1$$


と書くことができる。 \hat{Y} や a 、 b_1 は、回帰分析の結果求められる値であり、特に b_1 を **回帰係数** と呼ぶ。


 一般に、説明変数が m 個の場合には、

$$\hat{Y} = a + b_1 X_1 + \dots + b_m X_m$$


とし、基準変数の予測値 \hat{Y} と実際の基準変数値との差がなるべく小さくなるよう、 a 、 b_1, \dots, b_m の値を求める。求めた値を見ることで、基準変数と説明変数との関係を総合的にとらえようとするのである。

16.7 グループの間でどこが違うのか—判別分析—

 **判別分析**は、ある一つの質的変数（これを **基準変数** と呼ぶ）と、その他いくつかの変数（これを **説明変数** と呼ぶ）との関係を、総合的かつ簡潔にとらえようとする方法の一つである。判別分析は、基準変数が質的変数である重回帰分析（16.6 節参照）と見なすこともできる。さらに、特に説明変数が質的変数の場合の判別分析を **数量化 II 類** と呼ぶ。


 例えば、日常生活における世代間交流が多い人と少ない人とは、どこが違うのかを調べたいものとする。世代間交流の有無や程度には、その人の性別や年齢、健康状態などの属性、同居している異世代の有無といった生活環境、さらには積極性といったその人の性格など、様々な要因が影響していると考えられる。

世代間交流の程度と一つ一つの要因との関係は、それぞれクロス集計（16.4 節参照）を行って知ることができる。例えば、表 16.2 によれば、世代間交流が多いのは女性であることが分かる。

 判別分析は、クロス集計表を一つ一つ検討する代わりに、世代間交流が多い群と少ない群とを最もよく区別（**判別**）できるように要因の組み合わせ方を見つけ出

す方法である。

判別分析を行うことによって、基準変数（世代間交流の有無）にはどの説明変数（要因）がどのように影響しているのか、さらに基準変数に対する説明変数の影響は全体としてどの程度なのか、といったことが分かる。

 例えば、図 16.4[‡]は、基準変数を若い世代との交流の程度、説明変数を性別と年齢層としたときの判別分析の結果、得られた係数を棒グラフとしたものである。係数が相対的に大きい（棒が長い）ほどその要因の影響が大きいことを表す。

図から、世代間交流が多いのは、女性や 70 歳未満の高齢者であること、世代間交流の程度には、年齢層よりも性別の方が大きく影響していること、が分かる。

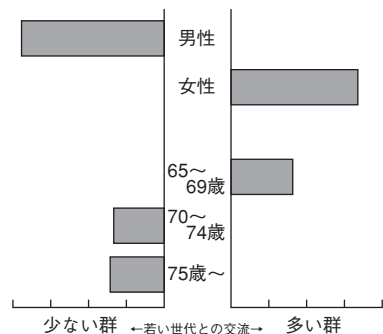



図 16.4: 判別分析

[‡] 世代間交流活動研究会「青少年及び高齢者の異世代に対する意識調査報告書」平成 12 年 1 月より


16.8 いくつかの項目をまとめて総合得点を出したい—尺度構成法—

 **尺度構成法**は、複数の変数をまとめ、より少数の（あるいは一つの）新たな変数を作り出す方法である。例えば学力テストでは、いくつかのテスト項目の結果を基に、総合的なテスト得点を求める。普通、得られた新たな変数を **尺度** と呼び、その値を **尺度得点** と呼ぶ。



ワンポイント アドバイス


厳密に言えば、尺度とは、測定対象に数値を付与する規則のことであり、尺度構成法とは、その規則を作り出す方法のことである。そのため、一つの変数だけから構成される尺度もあり得る。

 **リカート法**は、段階評定法（9.2.3 節参照）による項目を用いた尺度構成法の一つである。例えば、五件法であれば各選択肢を 1 から 5 とコーディングし、選ばれ

た選択肢のコードを足していくことで各調査対象の尺度得点を求める。

主成分分析法 や **因子分析法** は、量的変数（9.3 節参照）を基にする方法である。尺度得点を求めるために、基の変数の値を単に足し合わせるのではなく、最適な重み付けを行う。

数量化 III 類 は、質的変数を基にする方法である。**対応分析**、**尺度解析法**、**双対尺度法** などとも呼ばれる。

 図 16.5[§]は、まずリカート法によって、子どもの生活体験と生活習慣に関する尺度をそれぞれ構成し、次に得られた尺度得点を五段階に分類して、その結果どうしをクロス集計（16.4 節参照）したものである。生活体験の豊富な子どもほど生活習慣が身についている、といった相関関係があることが分かる。

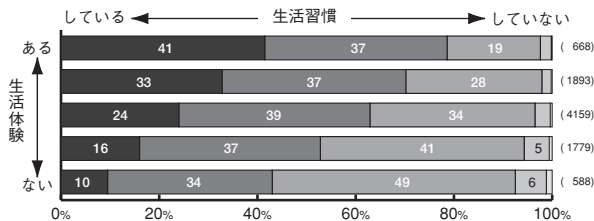




図 16.5: 尺度得点間のクロス集計

[§] 青少年教育活動研究会「子どもの体験活動等に関するアンケート調査報告書」平成 11 年 3 月 より

16.9 数値間に差はあると言えるのか—統計的仮説検定—

 調査結果の数値には誤差が含まれている (1.5 節参照)。そのため、調査結果において数値間に差が見られたからといって、ただちに母集団においても差があるとは言えない。例えば、表 16.2 (136 ページ) の「よくある」は男性が 23%、女性が 27%であり、4 ポイントの差がある。この 4 ポイントの差は誤差かもしれない、母集団においては男女差はないのかもしれない。

統計的仮説検定 は、例えば調査結果の数値間の差が、誤差の範囲内であり母集団においては差がないのか、あるいは誤差の範囲外であり母集団においても差があるのか、を調べる方法である。


 統計的仮説検定の基本的な考え方は次の通りである。


まず、母集団においては差がないと仮定する (これを **帰無仮説** と呼ぶ)。次に、もし帰無仮説が正しければ、調査結果の数値間の差はどの程度の確率 (これを **p 値** と呼ぶ) で生じるものなのか、を計算する。

p 値が大きい場合には、帰無仮説の下でそのような調査結果はしばしば得られるということを意味し、帰無仮説が間違っているとは言えない。つまり、積極的に差があるとは言えない。一方、p 値が非常に小さい場合には、二通りの解釈ができる。一つ

は、母集団において差がないという帰無仮説は正しく、得られた調査結果は非常に珍しいものであるという解釈である。もう一つは、そもそも帰無仮説が間違っており、差があるという仮説 (これを **対立仮説** と呼ぶ) の方が正しいという解釈である。

統計的仮説検定では、p 値が、あらかじめ定めておいた値 (これを **有意水準** と呼ぶ) よりも小さければ、帰無仮説を棄却し、対立仮説を採択する。

 有意水準としては、5%や 1%を用いることが多い。検定の結果は、帰無仮説を棄却しなかった場合には n.s.、有意水準 5%で棄却した場合には $p < .05$ 、1%で棄却した場合には $p < .01$ などと示す。なお、実際の p 値の計算は、統計ソフトを利用して行うとよい。

 調査結果について解釈を行う際には、統計的仮説検定を行い、数値間に有意差があるのかどうか調べておくとよい。

ただし、標本サイズが大きくなるほど、標本誤差は小さくなる (8.2 節参照)。そのため、標本サイズが大きくなれば、例えば 0.1 ポイントのような小さな差であっても有意となる。有意差が見られたからといって、それが実質的に意味のある差、とはならないことに注意しなければならない。

