

はじめに

～統計学を学ぶということ～

統計的手法やそれを学問的に体系化した統計学をなぜ学ぶのでしょうか。みなさんは、統計学を学ぶことのメリットや目的はなんだと思いますか。統計学を学習することによって、次の3つのことができるようになります。

1. 多くのデータを整理して法則を見つけ出すことができる

データをたくさん集め、それを統計的に処理していくと、法則(モデル)を見つけ出すことができます。ドイツの著名な社会学者マックス・ウェーバー(Max Weber 1864-1920)は、欧米の経済的に成功している人たちのデータを整理していくと、キリスト教の一派であるプロテスタントの居住する地域が裕福であることに気がつきました。彼は、質素、勤勉、儉約というプロテスタントの倫理が資本主義を興隆させる原動力になったという説を論文に著しました。有名な「プロテスタンティズムの倫理と資本主義の精神」です。彼は、帰納法的な視点からの仮説を科学的に証明(検証)して、多くの人々が納得のいく根拠を示したのです。

2. 部分を調べて全体(未知)を知る(推測)ことができる

選挙の投票が終わって開票が始まると、マスコミ各社は、いち早く当選者の予想を報道します。開票後、1時間もたっていないのになぜ当選確実といえるのでしょうか。しかも、ほとんどが報道の通りか、ほぼ予想に近い結果になっています。有権者全員に確認したわけでもないし、また、そのようなことは不可能です。統計的手法を使うと少しのデータで全体を知ることができるようになります。

3. 未来を予測し、対策をたてることができる

日本の人口を定点観測していくと、年齢の低い人は少なく、年齢が高い層ほど多くなってきていることがわかります。少子高齢社会の到来です。これらの過去のデータをもとに、統計的な予測にもとづいて、未来の人口ピラミッドを描くことができます。今後ますます少子高齢化が進展することが予測できます。予測ができると年金や税金などの政策の立案に役に立ちます。また、労働

人口の減少からは、定年制度の年齢の引き上げ、女性の活用、外国人労働者導入などの対策を考えることができます。

さて、本書の特色は主に2点あげられます。

1点目は、統計的手法の基本である「記述統計」を中心に解説していることです。記述統計は、特にデータの法則を見つけ出すのに優れた手法です。データを分析するには、じっくりと丹念にデータを読み取ることが大切です。そのために、各章にはワークシートを設定しています。ワークシートの内容はExcelの操作や計算問題ではなく、データを読み取って考察を行う問題を中心に設定しています。みなさん自身の理解度を確認するために活用してください。

2点目は、シナリオ・ラーニングの発想で身近な視点から統計的手法に触れられ、しかも応用ができそうと感じられるように工夫したことです。シナリオは、学園祭で模擬店を出店して、その利益を旅行サークルの資金にしようという設定です。実際にアンケートを行って、そのデータから模擬店の出店プランを立てていくという一貫したストーリーで構成しており、興味をもって学習できるようにしています。統計的手法を理解するとともに統計的なモノの見方を身につけることをねらいにしています。

その他、本書はなるべく平易で、理解しやすいように記述することに配慮しています。数式を少なくしたほか、パソコンを使って実践的に学習できる構成になっています。ぜひ「統計的思考」を鍛えて実践に活用して下さい。

なお、本書の執筆にあたって、忍耐強く支え、励ましてくださった、斉藤章明氏に感謝の意を捧げます。

竹田 茂生・藤木 清

目次

はじめに.....	i
本書の効果的な使い方.....	Vii

第1章 データをまとめる 1

シナリオ 2次データの収集……過去のデータを集める.....	2
1. 表とグラフ.....	4
2. 表のマナー.....	4
3. グラフの種類.....	6
4. グラフのマナー.....	9
5. Excelを使ったグラフの描き方.....	9

ワークシート

1-1. 表のマナーを身につけよう.....	13
1-2. マナーに注意して表を改善しよう.....	14

第2章 データの視覚化による発見 15

シナリオ 学年別に学園祭来場者数を考察.....	16
1. Excelによる表の作成.....	18
2. Excelによるグラフの作成.....	19
コラム 視覚化のウソ—グラフの表現によって2倍が4倍に！—.....	24

ワークシート

2. グラフを読み取ろう.....	25
-------------------	----

第3章 仮説を検証するための道具 27

シナリオ アンケート票の作成……質問項目を考える.....	28
1. 変数と尺度.....	30
2. 仮説.....	32
コラム パレートの法則からマイレージへ.....	33

ワークシート

3-1. アンケート調査項目を4つの尺度に分類しよう	34
3-2. 4つの尺度を使って質問文と選択肢をつくろう	35

第4章 全体の傾向をつかむ

シナリオ アンケート票の集計……データを入力して質問項目ごとに集計する	38
1. 変数とデータの入力	40
2. Excelによる単純集計	41

ワークシート

4-1. ピボットテーブルを用いてF2(住まいの形態)を集計しよう	47
4-2. Q3(出かけたことのある祭りやイベント)を集計しよう	48

第5章 細部の検討

シナリオ 質的データの分析……属性別に集計して比較する	50
1. クロス集計	52
2. 比率による集計	55
3. Excelによるクロス集計	57
コラム 行動生態学から探る殺人行動	63

ワークシート

5. 2つの変数を選んでクロス集計をしよう	64
-----------------------------	----

第6章 カテゴリーでまとめる

シナリオ 量的データの整理……データの分布を把握する	66
1. 量的データとカテゴリー化	68
2. Excelによる階級別度数分布表の作成	69
3. 分布の歪み(ゆがみ・ひずみ)	78
コラム 一連の行動をはかる	79

ワークシート

6. 大学創立年度のヒストグラムを作成しよう	80
------------------------------	----

第7章 データの中心を要約する	81
[シナリオ] 量的データの分析 1……データの水準を比較する	82
1. 代表値	84
2. Excel による代表値の分析	86
[コラム] オリジナルな物差しをつくる——高感度人間——	92
[ワークシート]	
7. 代表値を比較しよう	93

第8章 データの散らばり具合をみる	95
[シナリオ] 量的データの分析 2……データのばらつきを比較する	96
1. 散布度	98
2. 分散と標準偏差の計算例	100
3. 分析ツールの結果から	102
[ワークシート]	
8. ばらつきを比較しよう	104

第9章 データの関係を調べる	105
[シナリオ] 量的データの関連性分析 1……2変数の関係をみる	106
1. 相関	108
2. 散布図	108
3. Excel による相関係数の算出	111
4. 相関係数の計算例	115
[コラム] "女の子の登校率が上がると、子どもの死亡率が減る"	117
[ワークシート]	
9. 相関行列から2変数の関係性を評価しよう	118

第10章 データの一方向的関係を調べる	119
[シナリオ] 量的データの関連性分析 2……2変数の関係から予測する	120

1. 回帰モデルの考え方	122
2. Excel による回帰直線の描画.....	122
3. 最小二乗法.....	126
4. 決定係数	128
5. 回帰係数と決定係数の計算例.....	128

ワークシート

10. 回帰直線を求め、あてはまりを評価しよう	131
-------------------------------	-----

第 11 章 データの違いを判断する 133

シナリオ アンケート結果を検証……一部から全体を推し測る	134
1. 統計的仮説検定の考え方	136
2. Excel によるカイ二乗検定.....	137
3. Excel による t 検定	139
4. Excel による回帰係数の t 検定.....	143
5. 検定における 2 種類の誤り.....	146
コラム 損益分岐点の話	147

ワークシート

11. 回帰分析の結果を整理しよう	148
-------------------------	-----

総合課題 1 模擬店のプランを作成しよう	149
生活やイベントに関するアンケート.....	153
総合課題 2 企業・ブランドイメージ調査をしよう.....	159
参考文献	165
索引.....	167
巻末附録	169

本書では Excel 2007 を使用しています。お使いの OS やソフトのバージョンにより画面デザインや操作(グラフ作成、ピボットテーブルなど)が異なります。しかし、関数計算や分析ツールの内容は、Excel 2003～2000 とほぼ同じです。

■本書の構成と使い方

本書は Excel を利用した記述統計分析を中心に説明しています。各章の学習内容は以下のとおりです。第 1 章, 第 2 章は表とグラフの表現やマナーをおさらいします。第 3 章は統計分析をする上で必要となる変数と尺度の概念について説明しています。第 4 章, 第 5 章は質的なデータの集計方法と結果の読み取り方を学びます。第 6 章から第 10 章は量的なデータの整理の仕方や分析手法について説明します。その中で, 第 6 章, 第 7 章, 第 8 章は量的データの分布を捉えデータ間の比較検討を行います。第 9 章と第 10 章は 2 つの量的データの関係性を分析する手法を学習します。第 11 章では分析結果に意味があるかどうかを判断する方法を学びます。

各章にはワークシートを付けています。ワークシートの目的は単に Excel による作業を行うことではありません。結果からどのようなことが読み取れるのかを考察することが大切です。考察は論理的に結論を導き出しているかがポイントです。本書の目的は, 得られたデータを自ら加工し, 自分なりの解答を論理的に導き出す力を伸ばすことです。是非ワークシートを活用してステップアップしてください。

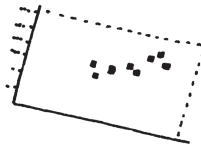
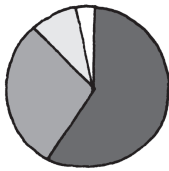
巻末には, 2 つの総合課題を用意しました。

総合課題 1 は, アンケート結果から模擬店を立案する課題です。データは, 本書専用ウェブサイト (<http://www.9640.jp/toukei>) から 103 ケースのデータをダウンロードして活用して下さい。みなさんなら, どのようなプランをたてるでしょうか。本書のシナリオでは, 4 人の大学生が模擬店の立案に挑戦します。みなさんも彼らに負けないような, もっとすてきな模擬店プランを作成してみてください。

総合課題 2 は, シナリオから離れて, 企業やブランドのイメージに関する意識調査を実施する課題です。調査票の作成, データ入力, 集計・分析および報告書作成といった調査の一連の流れを体験して, 総合実践力を養ってください。

第1章

データをまとめる



KEYWORD

- 表・グラフのマナー
- グラフの種類

2 次データの収集……過去のデータを集める



武谷

：この前の打ち合わせでは、どんな模擬店がいいのかニーズをつかむためにアンケートを実施することにしたわね。その前にターゲットをしぼるためのデータを集めることにしたのよね。



遠藤

：学園祭に関する一般的な既存資料は、ほとんど見つからなかったね。



橋田

：簡単なアンケートは毎年行っているみたいだけど、整理していないんだって。でも、学生会には、学園祭の来場者数のデータがあった。



佐々木

：じゃあ、その来場者数のデータを分析してみましょう。



武谷

：でも、どういった分析ができるかしらね。



遠藤

：数字をみてもわかりにくい。まずは、表をグラフにするところから始めよう。

過去(5年間)の学園祭の来場者数

	2004年	2005年	2006年	2007年	2008年
合計	1,916人	1,782人	1,479人	1,680人	1,777人

蓮浦大学学園祭実行委員会の調べ

☆ニーズ=お客様の望んでいることは何？

魅力を感じる商品・サービスは……。

→ アンケート調査を実施

☆ターゲット=メインとなるお客さんは誰？

1年生か2年生か，男性か女性か，タイプは……。

→過去のデータを分析

↓↓

表とグラフの活用

アンケート調査をする前に，既存のデータ
(2次データ)を分析することが重要です。
分析の第一歩は数値を視覚化することです。
データを表やグラフに整理して，
読み取ることからはじめましょう。



1-1 表(table)とグラフ(graph)

調査、観察、実験などによって多くの統計データが得られますが、値そのものから、増減の傾向や集団の構成、あるいは相関関係などを直接理解することは困難です。そこで、それらの数値を目的や用途にそって「表」の形に整理すると、数値のもっている意味が理解しやすくなります。

たとえば、毎年の売り上げのデータを店舗ごとやエリアごとにまとめるといったような工夫で、店舗間やエリア間の比較ができます。あるいは、毎年の売り上げを3年ごとにまとめることによって、細かい変動にとらわれず全体的な傾向を読み取ることが可能になります。また、レポートなどに記述する場合には、文章で数値を羅列するよりも、表に整理することによって読み手にわかりやすく伝達することができます。

グラフは「目で見える統計」ともいわれるように、グラフの役割は、表の役割に加え、視覚的にわかりやすくすることです。グラフに表すことによって、値のもつ意味をより単純化することができます。その結果、複雑な構造や原因の探求、また規則性をとらえることが容易になります。視覚化によって意外な事実気がついたりすることも少なくありません。

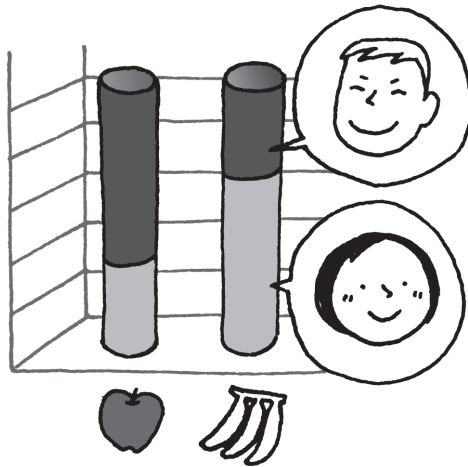
また、統計データをグラフ化することは、見る人に情報をわかりやすく提供するという目的をもっています。そのためには大勢の人々に、どのようなことを知ってほしいのか、また、いかに正確に情報を伝達できるかという視点が重要になります。グラフは、統計データを通して共通の理解を得るためのコミュニケーション・ツールといえるでしょう。

1-2 表のマナー

表を作成するにあたって、いくつか注意点があります。まず第1に、データの内容が的確にわかる表題(タイトル)をつけることです。表題は内容がよくわかるよう簡潔にし、長すぎないように気をつけましょう。特に、表をレポートに挿入するときは、「表番号」(表1, 表2, ……)をつけるようにしましょう。分量の多いレポートや報告書などでは、後で表を追加することも

第5章

細部の検討



KEYWORD

- クロス集計
- 多次元クロス集計
- 行の比率，列の比率，全体の比率

質的データの分析……属性別に集計して比較する



橋田

：「参加したことがあるイベントや祭り」の結果では、初詣、花火大会と同じくらい高校の学園祭も高いスコアが出ているね。



遠藤

：イベントや祭りに出かけるのはどんな人かな。友達の中でも、いろいろな人がいるね。男子は、あまり出たがらないゲームなどのインドア派と、スポーツや登山などのアウトドア派に分かれる。



武谷

：女の子は、イベントやお祭りが好きな人が比較的多いわ。基本的にはお買物は好き。それにオシャレできるチャンスでもあるし。女の子同士でもよく行くわ。どちらかというと、わたしは誘われる方だけど。



橋田

：他の質問項目では、どんな結果が出てる？



佐々木

：「お祭りで食品を購入するときに重視すること」として、“持ち運び易さ”を重視する人がたくさんいるわ。商品企画に生かせそうね。



遠藤

：“清潔さ”も重要なポイントになっている。そういえば、衛生的とはいえないような模擬店も多いな。データを、もう少し詳しくみてみよう。



武谷

：まずは、男女別に集計してみましよう。ほかに、こづかい別とかいろいろ仮説をたてて検証してみましようよ。

参加したことがあるイベントや祭り

N = 103	初詣	商店街のイベント	盆踊り	クリスマスイベント	花火大会	大学の学園祭	高校の学園祭	地域主催のイベント	友人とのパーティ	その他
実数	80	21	44	40	76	21	80	49	44	7
%	77.7	20.4	42.7	38.8	73.8	20.4	77.7	47.6	42.7	6.8

☆祭りによく出かける人は

性別	よく出かける	あまり……
男性	〇〇%	
女性		

☆こづかい

性別	～3,000円未満	3,000円～5,000円未満
男性	〇〇%	
女性		

単純集計の次は、クロス集計を使って詳細な分析を行います。クロス集計の知識を身につけてじっくり分析してみましょう。単純集計では得られない発見がある場合もあります。



5-1 クロス集計(cross tabulation)

データの分析では、単純集計だけでなく、収集したデータを多面的に分析することが重要です。たとえば、性別(男性、女性)、年齢(10歳代、20歳代とか若年層、高齢層)、未婚(未婚、既婚)、住居形態(持ち家一戸建て、持ち家マンション、賃貸マンション、賃貸アパートなど)といった、属性データやカテゴリーごと(使用頻度：ヘビーユーザー、ミドルユーザー、ライトユーザー)に集計し分析することで新しい知見を得ることができます。このように、属性、カテゴリーなどと他の項目を、表側(ひょうそく)と表頭(ひょうとう)に分けて掛け合わせてデータを集計することを、クロス集計といいます。

一般的には、独立変数を表側に、従属変数を表頭に配置します。つまり、表側の項目を基準として、表頭の項目間における度数の比較をするということになります。表1のクロス集計の事例をみてみましょう。「全体」の行のデータが、単純集計のデータにあたります。「ヘアスタイルや服装などの流行に関心ある」人は、「すごく興味がある」23.7%ですが、性別でみると、女性は男性の3倍近くあることがわかります。

表1 クロス集計の例

Q1. あなたは、ヘアスタイルや服装(ファッション)などの流行に関心がありますか

N=704	すごく 興味がある	ちょっと 興味がある	あまり 興味はない	興味はない	無回答
男性	11.4	44.5	34.4	9.7	0.0
女性	32.8	58.0	8.6	0.2	0.4
全体	23.7	52.3	19.6	4.3	0.1

「団塊ジュニアの消費と行動意識」日経産業消費研究所 2002

しかし、Excelを使えば、どのような変数でもクロス集計を実施することはできます。だからこそ、仮説を検証するための論理的な思考が大切なのです。因果関係を説明するにあたって、どのような独立変数と従属変数を用いるか、また、集計結果をみて新たな仮説を設定し、さらに深く考えることを繰り返すといった思考の過程を重視するようにしてください。ここが、良い