

できる
やさしく学ぶ

Excel統計入門

難しいことはパソコンにまかせて
仕事で役立つデータ分析ができる本

Excel 2013/2010/2007 対応

羽山 博 & できるシリーズ編集部

練習用ファイルを
無料でダウンロードできる!



こんな比較や検証も
6日でマスターできます!

- ✓ 競合商品との評価の違いを数値化して解析
- ✓ 人気ランキングと売り上げの関係を分析
- ✓ 年齢や職業による購買数の違いを比較

さらなるデータ分析は……本文で!

感覚や思い込みではなく、
数字で違いを説明できます!

シリーズ
6000万部
突破
売上
No.1
ベストセラー

インプレス

ご購入・ご利用の前に必ずお読みください

本書は、2015年1月現在の情報をもとに「Microsoft Excel 2013」「Microsoft® Office Excel 2010」「Microsoft® Office Excel 2007」の操作について解説しています。下段に記載の「本書の前提」と異なる環境の場合、または本書発行後に「Microsoft Excel 2013」「Microsoft® Office Excel 2010」「Microsoft® Office Excel 2007」の機能や操作方法、画面などが変更された場合、本書の掲載内容通りに操作できない可能性があります。

本書発行後の情報については、弊社のWebサイト (<http://book.impress.co.jp/>) など可能な限りお知らせいたしますが、すべての情報の即時掲載ならびに、確実な解決をお約束することはできません。なお本書の運用により生じる、直接的、または間接的な損害について、著者ならびに弊社では一切の責任を負いかねます。あらかじめご理解、ご了承ください。

本書で紹介している内容のご質問につきましては、巻末をご参照の上、封書にてお問い合わせください。電話やFAX、電子メール、アンケートはがきでのご質問には対応しておりません。また、本書の発行後に発生した利用手順やサービスの変更に関しては、お答えしかねる場合があることをご了承ください。

練習用ファイルについて

本書で使用する練習用ファイルは、弊社Webサイトからダウンロードできます。練習用ファイルと書籍を併用することで、より理解が深まります。

▼練習用ファイルのダウンロードページ

<http://book.impress.co.jp/books/1114101032>

●用語の使い方

本文中では、「Microsoft Excel 2013」のことを「Excel 2013」、「Microsoft Office Excel 2010」のことを「Excel 2010」「Microsoft® Office Excel 2007」のことを「Excel 2007」または「Excel」と記述しています。また、本文中で使用している用語は、基本的に実際の画面に表示される名称に則っています。

●本書の前提

本書では、「Windows 8.1」と「Office Professional Plus 2013」がインストールされているパソコンで、インターネットに常時接続されている環境を前提に画面を再現しています。

「できる」「できるシリーズ」は、株式会社インプレスの登録商標です。

Office、Excel 2013は、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。

そのほか、本書に記載されている会社名、製品名、サービス名は、一般に各開発メーカーおよびサービス提供元の登録商標または商標です。

なお、本文中には™および®マークは明記していません。

Copyright © 2015 Rogue International and Impress Corporation. All rights reserved.

本書の内容はすべて、著作権法によって保護されています。著者および発行者の許可を得ず、転載、複写、複製等の利用はできません。

まえがき

多少大げさかもしれませんが、私たちの文化がここまで進化したのは、人類が過去の経験を分析し、未来を予測する能力に恵まれていたからと言っても間違いではないでしょう。統計学はその能力を大きく拡大させるのに役立つ知識です。コンピューターが一般的でない時代には、多くの計算を必要とする統計学の手法は一部の人にしか使いこなせませんでした。今の私たちにはExcelなどの便利な道具があります。考え方さえ理解すれば、簡単に分析や予測ができるようになったのです。

「統計」と一口に言っても、人によってイメージが違っているかもしれません。誤解を恐れずざっくりと分けると、少量のデータから全体を推し量る方法と、大量のデータの中から隠れた特徴を見つけ出す方法があります。この本では、どちらかというと前者の方法を説明します。一般に基礎的な統計学として位置付けられる領域で、一部のデータから平均や分散などを求め、全体の性質を推測する方法です。後者にあたるものとしては、多変量解析と呼ばれる方法があります。最近注目を集めているデータマイニングなどもその応用例と考えていいでしょう。この本ではそういった内容を直接には取り扱いませんが、それらの手法を理解するための基礎知識はしっかりと身に付けられます。

したがって、本書の対象となる読者像は、統計学に初めて接する人です。「ビジネスに活用したいんだけど、数学が苦手だから統計学はちょっと……」と敬遠していた人がいるかもしれませんが、心配は無用です。数式は登場しますが、基本的に小中学校で学んだ四則演算だけでほとんどの計算ができます。しかも、実際の計算はExcelが代わりにやってくれます。そういう面倒な部分はコンピューターに任せることができるのです。

重要なのは、統計学の考え方を理解することです。

そのため、この本では、これまでの「できるシリーズ」の体裁にはなかった、会話やレシピなどの要素を紙面に盛り込み、着目すべき点や考え方、手法の適用場面、意外な落とし穴などを強調しています。この本を読んで、統計学をより身近に感じるとともに、ビジネスなどでの利用に手応えを感じていただけるようになれば幸いです。

最後になりましたが、この本を世に出す機会をくださった株式会社インプレスできる編集部の藤井編集長、大塚副編集長、企画・編集のすべてにわたって大変お世話になった編集担当の井上薫さん、すてきなイラストを描いてくださった野津あきさん、そのほかご尽力くださった皆さまに感謝の意を表します。

2015年1月 羽山 博

本書の読み方

この本は、Excel を操作しながら統計学の基礎を学べるように構成されています。先輩と後輩の会話を糸口として、テーマに沿った問題解決の手法や操作方法、関連知識が身に付きます。

4日目 平均値の差の検定

A 1 商品の評価に差があるかどうかを検証する

B 1-1 モニター調査で得られた評価には本当に差があるのか?

C 練習用ファイル
4_1_1.xlsx

D 統計レシピ
2群の平均値に差があるかどうかを検定するには (対応のあるデータの場合)
方法 1検定を行う
利用する関数 T.TEST 関数 (検定の種類) に 1 を指定する
前提 母集団の分布が正規分布に従っている
帰無仮説 母集団の平均は等しい
参照 平均値の差の検定 (対応のないデータで母集団の分散が等しい場合) → 134 ページを参照
平均値の差の検定 (対応のないデータで母集団の分散が等しくない場合) → 137 ページを参照

E キーワード
1検定 - P-215
正規分布 - P-219
検定 - P-221
母集団 - P-222
有意差 - P-223
T.TEST 関数 - P-233

F Point!
いずれが大きいかを調べるときは片側検定。
差があるかどうかを調べるときは両側検定。

Step Up 有意差の判定を記号で表してみよう
一般に、検定の結果が 1% 有意の場合は「**」、5% 有意の場合は「*」、有意でない場合は「ns」と表記します。求められた確率によって、それらの文字列を表示できるようにしてみましょう。

G キーワード
関数の形式 IF (論理式, [真の場合], [偽の場合])
関数の意味 TRUE (真) か FALSE (偽) を返す式を指定する
入力例 =IF(E4<=0.01,"**",IF(E4<=0.05,"*",ns))

練習用ファイル
4_1_1.xlsx
検定 - P-217

4 ccs 5

A 大項目
各章では実際の仕事で直面しそうな課題や問題をテーマごとに解説します。大項目を見ればテーマの課題やレッスンの目的などがひと目で分かります。

B 中項目
大項目をさらに分類して、中項目で区切っています。おおまかな流れがタイトルから理解できます。

C 練習用ファイル
読み進めるだけでも知識が得られますが、実際に操作すればより確実に手法が身に付きます。練習用ファイルがある項目にはファイル名を明記しています。

D 統計レシピ
操作を始める前に「方法」や「利用する関数」など確認できます。統計レシピを参照すれば、問題解決のために必要なことや関連する項目がすぐに分かります。

E キーワード
そのページで覚えておきたい用語の一覧です。用語集や関数 INDEX の該当ページを掲載しているので、用語の意味をすぐに調べられます。

F Point
項目を理解するために必要な要点を簡潔に解説しています。



うーむ、別にウチの商品が勝つてというわけではないんですね。

そうね、まあ、勝っていてもそうでなくても、結果の上にあくらをかいたり、逆に心配しすぎたりするの良くないけど、少しでも良くする努力は必要ね。

● 引数の指定方法

引数	指定数	意味
尾部	1	片側検定 (しずれか大きいかを調べる)
	2	両側検定 (差があるかどうかを調べる)
検定の種類	1	対応のあるデータの検定
	2	母集団の分散が等しいと仮定される場合
	3	母集団の分散が等しいと仮定されない場合

なお、t検定を行うためには母集団が正規分布であることが仮定できる必要があります。実は、サンプルデータは正規分布に従っていないのですが、計算方法を補正するために、そのまま使っています。実際のところ、正規分布から多少はずれていても、サンプル数が十分多ければ (一般には30以上と言われています)、t検定の適用は可能です。その根拠については146ページで紹介する中心極限定理をご覧ください。

H **Tip**
t検定のように補正を多少必要にしてもいいけど、正しい結果が得られる性質のことを「信頼性」と呼びます。「信頼」とは簡単に言えば、不利な状況があっても大丈夫という意味です。

Column

第一種の過誤と第二種の過誤

仮説検定には、第一種の過誤と第二種の過誤と呼ばれる2つの誤りの可能性があります。第一種の過誤とは、仮説が正しいにもかかわらず棄却してしまう誤りです。これはよく警報器の動作たとえられます。火事が起こっているにもかかわらず、警報器が鳴らないのが第一種の過誤というわけです。以下の表ではその確率を α と表しています。

一方、第二種の過誤は仮説が誤っているにもかかわらず採用してしまう誤りです。火事が起こっていないのに、警報器が鳴ると

いった事例です。以下の表ではその確率を β と表しています。T.TEST関数での検定では、第二種の過誤については考慮されていません。つまり、平均値が実際には異なるのに平均値が等しいと言ってしまう誤りもあります。平均値が小さい場合には第二種の過誤を犯す確率が高くなります (そのような場合には、サンプル数を多く取る必要があります)。なお、第二種の過誤を犯さない確率、つまり $1-\beta$ のことを検出力と呼びます。

G Step Up

項目の内容を応用した、ワンランク上の使いこなしワザを解説しています。ワザによっては、練習用ファイルも用意しています。

H Tips

項目に関連したさまざまな機能や一歩進んだ使いこなしのヒントなどを解説しています。

I Column

読むとためになるプラス α の知識や情報を紹介しています。後でじっくり読めば統計やデータ分析に関する理解が深まります。

※ここに掲載している紙面はイメージです。実際のページとは異なります。

手順 必要な手順を、画面と操作を掲載して解説しています。

操作説明

「○○と入力」「○○をクリック」など、それぞれの手順での実際の操作です。番号順に操作してください。

解説

操作の前提や意味、操作結果に関して解説しています。

	A	B	C	D	E	F	G	H	I
1	お菓子の食べ比べの結果				T検定				
2									
3	サンプル	性別	できるサンプル	飲社サンプル		片側確率			
4	1 F		2	7		=T.TEST(C4:C23,D4:D23,1,1)			
5	2 M		5	6					
6	3 F		8	6					
7	4 M		7	5					
8	5 M		7	6					
9	6 F		3	5					
10	7 M		4	5					
11	8 F		9	5					
12	9 F		6	6					
13	10 F		8	7					

① セルF4に「=T.TEST(C4:C23,D4:D23,1,1)」と入力

② [Enter]キーを押す

片側確率が求められる

目次

まえがき	3
本書の読み方	4
登場人物の紹介	11
フロログ	12

第1章

調査結果から顧客の特徴を把握しよう	15
1日目 収集したデータをどう扱う？	16
1 データ入力	
1 データの入力方法を知って集計や分析のしやすい表を作る	18
1-1 アンケートをデータ化する「鉄板のルール」とは	18
Column 作業の流れと頭の中で起こっていること	20
1-2 売上伝票は「明細」と「頭書き」に注目！	21
Column アンケート調査の落とし穴	23
2 度数分布表	
2 全体の傾向や特徴を表にまとめて整理する	24
2-1 意外に簡単！ データ集計と整理の極意	24
2-2 データを区切れれば傾向が見える	25
Step Up データの変更に対応する表を作ろう	28
Step Up 一瞬にして度数分布表を作成するワザ！	30
3 ヒストグラム	
3 団体の全体像と特徴を表すグラフを作成する	31
3-1 データの個数や分布がひと目で分かるグラフを作る	31
3-2 まずは集合縦棒グラフを作る	32
3-3 ドラッグ操作でグラフから不要なデータを除外！	34
Step Up ボタン1つで系列が指定できる	36
3-4 棒の間隔や色を変更してグラフを仕上げる	37
4 ピボットグラフ	
4 もっと簡単にダウンロード数をグラフ化する	40
4-1 元のデータから直接ヒストグラムを作成できる！	40
4-2 ダウンロード数を5個ずつの区切りでまとめよう	43
4-3 性別によってデータに違いがあるかを比較してみよう	45
Step Up 男性と女性のデータを1つのグラフに表示しよう	47
この章のまとめ	48

	商品に対する評価を掘り下げて調べてみよう	49
	2日目 集団を代表する値って何のこと？.....	50
	1 平均値・中央値・最頻値	
	1 ライバル商品との評価の違いを調べる.....	52
1-1	自社製品と他社製品の評価を比較しよう.....	52
	Column 印象操作にだまされないために.....	54
	Step Up 広いセル範囲をサクッと指定.....	55
1-2	男性と女性で評価が違うかどうかを確認しよう.....	56
1-3	平均値という信仰は捨てよう！.....	59
	Column 平均値の落とし穴.....	61
1-4	平均値だけが代表値じゃない！真ん中にある値を調べる「中央値」.....	62
1-5	最もよく現れる値を調べる「最頻値」.....	63
	Column 配列から特定の要素を取り出すには.....	65
	2 歪度・尖度	
	2 商品モニター調査の分布の形を見る.....	66
2-1	分布の形で商品の評価を見てみよう.....	66
2-2	できるサプレの好みには偏りがある？.....	67
	Step Up セルの左上に三角形のマークが表示されたら.....	69
2-3	できるサプレは評価の分かれる商品なのか？.....	71
	Step Up 名前を使って引数を分かりやすくしよう.....	73
	Step Up 正規分布って何？.....	74
	3 分散・標準偏差	
	3 商品モニター調査の分布のばらつきを見る.....	76
3-1	分布のばらつきを数値で求めてみよう！.....	76
3-2	分散や標準偏差はどう使い分けるの？.....	78
	Column Excel 2007 以前と互換性のある関数を使うには.....	81
	Step Up データベース関数を使って性別ごとに分散を求める.....	81
3-3	そもそも分布のばらつきって何？.....	83
3-4	関数を使わずに不偏分散を求めてみよう.....	86
3-5	どうして不偏分散を求めるときは $n-1$ で割るの？.....	88
	Step Up ピボットテーブルでも分散や標準偏差が求められる.....	89
	Step Up 平均値や分散の推定にある程度の幅を持たせるには.....	90

4

偏差値

試験結果から集団内での位置を知る..... 92

4-1 偏差値で本当の実力が分かる..... 92

この章のまとめ..... 96

第3章

売り上げに何が関係しているかを見極めよう..... 97

3日目 売り上げに影響する要因とは..... 98

1

相関係数

店舗への訪問回数と売り上げの関係を調べる..... 100

1-1 訪問回数と売上金額の関係を見える化してみよう..... 100

1-2 相関関係＝因果関係ではない..... 102

1-3 訪問回数と売り上げてどれくらい関係があるの?..... 104

1-4 相関係数は自分で計算できる!..... 107

Step Up 見かけの数値にだまされないように..... 109

1-5 Column 結果は何けたまで表示すればいい?..... 110

2

順位相関

お菓子の人気ランキングと売り上げの関係を調べる..... 111

2-1 売上金額は人気ランキングを反映しているか..... 111

Column 尺度のいろいろ..... 113

3

回帰分析・重回帰分析

商品の売り上げを予測する..... 114

3-1 営業担当の訪問回数から売上金額を予測できるか..... 114

3-2 営業担当の訪問回数と経験から売り上げを予測してみよう..... 118

3-3 似たような変数を使って予測しても意味がない!..... 122

この章のまとめ..... 126

第4章

他社商品との評価の差やばらつきの差を検証しよう..... 127

4日目 その差は本当に意味のある差なのか?..... 128

1

平均値の差の検定

商品の評価に差があるかどうかを検証する..... 130

1-1 モニター調査で得られた評価には本当に差があるのか?..... 130

1-2 研修のテキストによって資格試験の成績は異なるか..... 134

Step Up 有意差の判定を記号で表してみよう..... 136

1-3	一般の顧客と専門家によってデザインの評価は異なるか	137
1-4	帰無仮説と対立仮説を理解して検定を使いこなそう	139
1-5	差があるかどうかを判定するための基準とは？	142
1-6	差があるかどうかを判定するための値をどうやって求めるの？	145
1-7	中心極限定理はすべての基礎！	147
	Column 第一種の過誤と第二種の過誤	147
1-8	平均値の差の検定を手作業でやってみよう	149
1-9	母集団が正規分布していない場合の平均値の差の検定は？	154
	Step Up 同順位がある場合のマン・ホイットニー検定	158

2

分散の差の検定

商品の評価のばらつきに違いがあるかどうかを検証する	159
---------------------------	-----

2-1	評価のばらつきの差も検定できる！	159
2-2	分散が大きい小さいかを検定したい	161
	この章のまとめ	164

第5章

性別によって好みに違いがあるかどうかを調べてみよう	165
---------------------------	-----

5日目 マーケティングやターゲティングに役立つ検定	166
---------------------------	-----

1

χ^2 検定

性別によって好きなペットは異なるか	168
-------------------	-----

1-1	性別とペットの好みをクロス集計表で確認しよう	168
1-2	女性は猫好き、男性は犬好きって本当？	170
	Step Up 一瞬で縦横の合計を求める！	175
1-3	返品の手数は特定のパターンを持つのか？	175
	Column 離散分布と連続分布	179

2

相関係数の検定

店舗への訪問回数と売り上げの関係を検定する	180
-----------------------	-----

2-1	相関係数が大きいとホントに相関があると言っていいのか？	180
-----	-----------------------------	-----

3

回帰分析・重回帰分析の検定

訪問回数と経験で売り上げが本当に説明できるのか	183
-------------------------	-----

3-1	その回帰直線はホントに役に立つ？	183
3-2	予測に役立つ係数は訪問回数それとも年齢？	186
	この章のまとめ	188

第6章

性別や職業による購入数の差を調べよう…………… 189

6日目 「会社員」「学生」「無職」の平均値の差を検定する…………… 190

1

一元配置分散分析・多重比較

職業によってお菓子の購入数が異なるかどうかを検証する…………… 192

1-1 3群以上の場合は平均値の差の検定が使えない?…………… 192

1-2 職業によって購入数に差があるか…………… 193

1-3 水準間と水準内の変動ってどういうこと?…………… 198

1-4 差があるのは会社員と学生? 学生と無職?…………… 200

Step Up シェッフエの方法による多重比較とは…………… 205

2

二元配置分散分析

性別と職業によってお菓子の購入数が異なるかどうかを検証する…………… 206

2-1 要因が2つある場合は二元配置の分散分析!…………… 206

2-2 分析する前に Excel のアドインを有効にしよう…………… 207

2-3 性別による平均値の差と職業による平均値の差を検定しよう…………… 208

2-4 交互作用っていったい何?…………… 210

この章のまとめ…………… 212

エピソード…………… 213

あとがきにかえて…………… 214

用語集…………… 215

関数 INDEX…………… 224

索引…………… 234

本書を読み終えた方へ…………… 237

読者アンケートのお願い…………… 238

登場人物の紹介



市川 学 (いちかわまなぶ)

できる製菓株式会社の営業企画部勤務（入社2年目の24歳）。一浪して地元のS大学法学部になんとか潜り込み、平凡な成績で卒業。仕事に関してはまだまだ新人の域を出ないが、意外に努力家。趣味のギターはなかなか上達しない。

網島 久美 (つなしまくみ)

株式会社カイ・プランニングのチーフプランナー（27歳）。学の高校時代の先輩。生徒会長で、成績優秀、スポーツ万能。名門T大学経済学部にて現役合格し、主席で卒業。数々の大きなプロジェクトを成功に導いている。趣味はバイク。

藤井 経堂 (ふじいきやうどう)

できる製菓株式会社の営業企画部部長（40歳）。高校卒業後の入社以来、常に営業の最前線に立っていたが、アイデアあふれる働きぶりが評価され、営業企画部の部長に抜擢された。常に前向きで部下思い。趣味は釣りと囲碁。

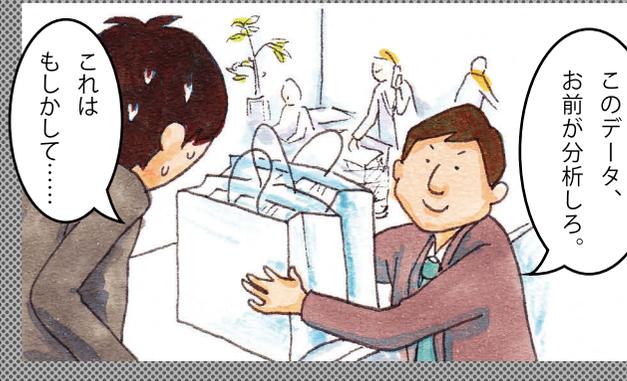
練習用ファイルについて

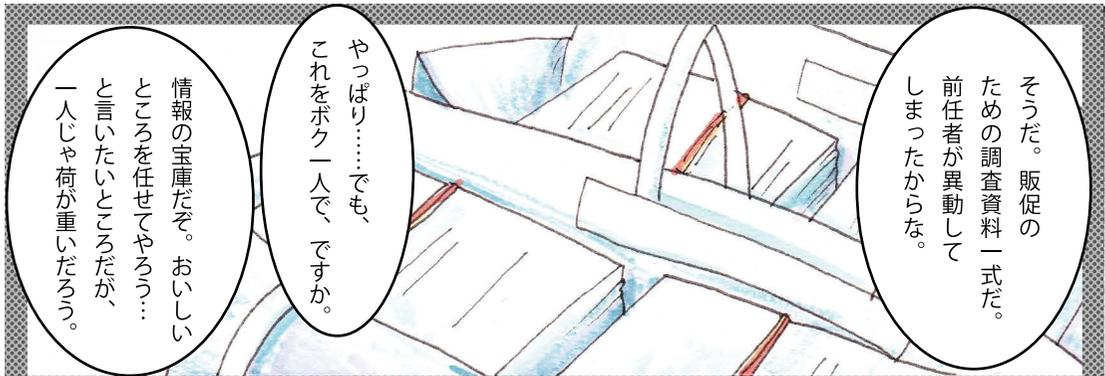
本書は、事例を中心に分かりやすく説明しているので、読み進めるだけでもひと通りの知識が得られますが、練習用ファイルと書籍を併用すると手法や考え方の理解が深まります。Excelの操作が必要な項目には練習用ファイルを用意しています。以下のホームページからダウンロードして、操作してみてください。

▼練習用ファイルのダウンロードサイト

<http://book.impress.co.jp/books/1114101032>

プロローグ





そうだ。販促の
ための調査資料一式だ。
前任者が異動して
しまったからな。

やっぱり……でも、
これをボク一人で、ですか。

情報の宝庫だぞ。おいしい
ところを任せてやろう……
と言いたいところだが、
一人じゃ荷が重いだらう。



カイ・プランニングさんは
知ってるな。1週間勉強
させてもらってこい。

あの、
調査とか企画を
手伝ってくれている
頭脳集団ですね……

ふふふふふふふふふふ

ガチャッ

—そして
今日—



市川さんですね、
どうぞこちらへ

失礼しま…

あれ、
久美先輩！



高校のときの…

meeting room

ごめんなさい、私、覚えてなくて

いえいえ、ボクは地味でしたけど、先輩は超有名人でしたから。

そっだったかなあ。

それに…

それに？

データ分析のお話なんですがあ、いえ！

あこがれの先輩でした。

なんて、恥ずかしくて言えないよな

藤井部長から伺ってるわ。みっちり鍛えてやれって。

高校の後輩だから、マナブくんがいいわね。

！

じゃ、じゃあ、先輩は「先輩」で。

あれ、そのまんまだ。

ところで、その紙袋は？

あ。

マナブくんには分析以前の問題があるわね。

資料一式です。

これをどっやって分析すればいいの？

なるほど、分かったわ。

ええっ、まだ資料を1つも見ていないのに？



第1章

調査結果から顧客の 特徴を把握しよう

久美先輩との再会に驚いたマナブ君。いよいよ久美先輩からデータ分析の方法をレクチャーしてもらうことになりました。プロローグで久美先輩が指摘したマナブ君の問題とはいったい何なのでしょう。その問題を明らかにした後、集団がどのような特徴を持つのかを大まかにつかむ方法として、度数分布表の作り方やヒストグラムの作り方を学びます。

1	データ入力 データの入力方法を知って集計や分析のしやすい表を作る	18
2	度数分布表 全体の傾向や特徴を表にまとめて整理する	24
3	ヒストグラム 集団の全体像と特徴を表すグラフを作成する	31
4	ピボットグラフ もっと簡単にダウンロード数をグラフ化する	40

第1章を始める前に 収集したデータをどう扱う？



久美先輩、今日からよろしくお願いします！



よろしくね、マナブ君。藤井部長に頼まれたからには、頑張って教えるわよ！
どうやら、マナブ君はデータ分析以前の問題があるようね！



えっ！「データ分析以前の問題」って、いったい何なんですか？



分析以前にデータの整理方法を知らないってことよ！ 収集したアンケートをどう
いう形式で入力すればいいかわかっていないんじゃない？



確かにそうかも。Excelを使えばいいんだろうな、とは分かるんですけど……。
でも、何でいきなりそこまで分かるんですか。



だって、アンケートの束を持ち歩いているじゃない。ってことは、入力の段階で
つまずいているってことでしょ。



データの入力に決まった方法があるんですか。



あるわよ。みんな何となくやっているけど、原則をちゃんと理解しているのとし
ていないのでは大違い。後でデータを加工したり、集計したりするときに効率
が全然違ってくるわよ。



そうなんですな！ 頑張りますっ！

こんなことが できるようになります

スマートフォンアプリのダウンロード数に関するアンケートの束を持っているマナブ君。Excel を起動したものの、そこで行き詰まっているようです。この章では、調査票から集計表やグラフを作成して、「どの年代がスマートフォンのアプリを最もインストールしているか?」「男女でダウンロード数に違いがあるのか?」といった特徴を大まかにつかめるようにします。

統計レシピ

- 調査票のデータを Excel のワークシートに入力するには
- 伝票形式のデータを Excel のワークシートに入力するには
- 度数分布表を作成するには
- ヒストグラムを作成するには
- グラフから余計な系列を除外するには
- 棒グラフの間隔を詰めるには
- 元のデータからさまざまなグラフを簡単に作成するには
- ピボットグラフから度数分布表とヒストグラムを作成するには
- ピボットグラフで複数の系列を比較するには



1

データの入力方法を知って 集計や分析のしやすい表を作る

1-1 アンケートをデータ化する「鉄板のルール」とは

データを分析するためには、表にデータを入力する必要があります。しかし、どのようにデータを入力すればいいのか、きちんと説明されることはあまりなかったと思います。データ分析の方法がよく分からないという人は、Excelの関数や統計の分析手法以前に「データの入力方法」という出発点があやふやだったのかもしれません。

出発点をおろそかにせず、きちんと確認しておくことは大切なことです。そこから始めましょう。

原則は極めて簡単です。1件のデータを1行に入力する、これだけです。

📎 練習用ファイル

1_1_1.xlsx

📌 Tips

本書で使用する練習用ファイルは、弊社 Web サイトからダウンロードできます。練習用ファイルと書籍を併用すれば、より理解が深められます。

▼練習用ファイルのダウンロードページ

<http://book.impress.co.jp/books/1114101032>

統計レシピ

調査票のデータを Excel のワークシートに入力するには

方法 1件のデータを1行に入力する

留意点 通常は、1枚の調査票が1件分のデータにあたる。項目数が多い場合は複数枚の調査票が1件分のデータになることもある

ただし、「1件のデータ」が何を指すのかがきちんと理解できていないと表の作成ができません。例えば、図 1-1 のような調査票を使ってスマートフォンアプリの利用に関するアンケートを取った例を見てみましょう。

📌 Tips

本書で取り扱っているデータは架空のデータです。また、分析方法を理解しやすくするため、一部を単純化して示しています。

図1-1 マナブ君が100人に聞いたアンケートの内容

この調査では、1人分のデータが1件のデータになる



スマートフォンアプリに関するアンケート
No. 036
性別：男性・女性
年齢：45歳
アプリのダウンロード数 15個

この場合、調査票1枚が1件のデータにあたります。つまり1人分の調査結果が1件のデータです。1件のデータに「No.」「性別」「年齢」「アプリのダウンロード数」という項目があることも分かります。それらを1行ずつ入力すればいい、というわけです。「No.」は「サンプル」、「アプリのダウンロード数」は「DL数」という見出しに変えています、図1-2に示したものが実際に入力したデータです。

図1-2 1件のデータは1行に入力する

	A	B	C	D	E
1	スマホアプリ利用調査				
2					
3	サンプル	性別	年齢	DL数	
4	1	F	19	37	
5	2	M	44	27	
6	3	F	52	8	
7	4	M	49	21	
8	5	M	20	19	
9	6	F	77	0	
10	7	M	42	21	
	8	F		46	
96			45		
97	94	F	44	13	
98	95	F	76	4	
99	96	F	21	19	
100	97	M	34	53	
101	98	M	25	16	
102	99	F	45	21	
103	100	M	19	8	
104					

19歳の女性がスマートフォンのアプリを37個ダウンロードしたという結果は、このように入力する

調査票のデータを1行ずつ入力していく



サンプルとは全体から取り出した個々の人や物のことで、標本とも呼ばれます。また、全体のことを母集団と呼びます。この例なら、「スマートフォンを利用している人すべて」が母集団にあたります。ユーザー全体の数はあまりにも多いので、全員からアンケートを取ることは不可能です。そこで100人のサンプルを抽出して調査したというわけです。

性別の「F」は女性(Female)を表し、「M」は男性(Male)を表します。「女」とか「男」のように日本語で入力するのは面倒なので、半角英数字を使ったというわけです。女性を0、男性を1というようにコード化して表すこともあります。

年齢については説明するまでもありませんね。「DL数」はアプリを何個ダウンロードしたかということです。

♂ キーワード

サンプル…P.218

標本…P.221

母集団…P.222



なるほど！ 調査票の束を前にして途方に暮れていましたが、データの入力方法がよく分かりました。



こういう1件1件のデータのことをレコードと呼ぶこともあるわ。Excelではレコードって言葉はあまり使わないけど。

なお、国勢調査のように調査項目が多いアンケートの場合には、何枚かの用紙の内容が1件のデータになることもあります。つまり、用紙が何枚であっても、1件のデータとは「1つのサンプルから得られたデータ」と考えられます。

Point!

1つのサンプルから得られたデータが1件のデータ。1件のデータを1行に入力する。

Column

作業の流れと頭の中で起こっていること

私たちは、データを入力するという単純な作業の中でも「モデル化」という操作を行い、それを適用しています。この例なら「1枚の調査票(具体的なもの)」→「1件のデータ(抽象的なもの)」→「ワークシートの1行(具体的なもの)」という流れになります。

簡単な作業ならこの操作が無意識のうちにできますが、複雑な作業になってくるとどう手を付けていいのかわからなくなることがあります。そういうときに、この流れを意識して紙に書いてみると、どのデータをどのように取り扱えばいいのかが見えてきます。

1-2

売上传票は「明細」と「^{かしらが}頭書き」に注目！

前項のような簡単な調査票の場合、1枚の用紙の内容が1件分のデータにあたります。しかし、売上传票のような複雑な帳票の場合、何が1件分のデータにあたるか分かりにくいことがあります。そのような例を見てみましょう。

統計レシピ

伝票形式のデータを Excel のワークシートに入力するには

方法 商品の売り上げをまとめた伝票では、明細の行数だけデータを入力する。伝票の共通部分は各行の先頭に入力する

図 1-3 の例は、簡略化するために商品コードや消費税などは省いてありますが、売上传票はだいたいこのような形式になっています。^{かしらが}頭書きと明細に分かれていることに注目してください。

図 1-3 売上传票のイメージ

売上传票 1 枚には複数件のデータがある



売上传票			
No. 101	日付		2015/3/5
得意先名: ^{かしらが} できる商事(株)			
商品名	単価	数量	金額
マーカー黒	150	5	750
名前ペン	120	4	480
製図鉛筆	100	2	200
		合計	1,430

^{かしらが}頭書き

明細

売上傳票のように、明細が何行かある帳票の場合、それぞれの明細が1件のデータになります。したがって、1枚の用紙に複数件のデータが含まれます。伝票番号や日付、得意先名のような頭書きは1回しか書かれませんが、すべての明細に共通する内容です。そこで、このような共通部分は明細の最初に入力しておきます。

なお、「金額」は項目に含めないこともあります。金額の値を保存しておかなくても「単価×数量」という計算で求められるからです。金額の合計は複数のデータを元に計算して求められるので、1件1件の明細データとは取り扱いが異なります（合計の部分は^{あしが}脚書きとも呼ばれます）。

図1-4 売上傳票は頭書きを左に、明細を右に入力する

売上傳票

No.101 日付 2015/3/5

得意先名: できる商事(株)

商品名	単価	数量	金額
マーカー黒	150	5	750
名前ペン	120	4	480
製図鉛筆	100	2	200
		合計	1,430

頭書き + 1件の明細を
1行に入力する

伝票番号	日付	得意先名	商品名	単価	数量	金額
101	2015/3/5	できる商事(株)	マーカー黒	150	5	750
101	2015/3/5	できる商事(株)	名前ペン	120	4	480
101	2015/3/5	できる商事(株)	製図鉛筆	100	2	200



売上傳票の場合は、データの入力方法が違いますね。



そうですね、共通のデータがあるかどうかでデータの入力方法が異なることを覚えておくといいわね。

さまざまな帳票に「1件のデータ」が通常どのように記録されているかをまとめると、以下のようになります。

- ・簡単な調査票：1枚の用紙に1件のデータ
- ・項目数の多い調査票：複数枚の用紙で1件のデータ
- ・伝票：1枚の用紙に複数件のデータ

Point!

複数のデータが1ページに記録されているときは、共通部分を各行の先頭に入力する！

Column

アンケート調査の落とし穴

アンケート調査の結果は根拠のはっきりしない主観的な主張と異なり、実際に得られたデータなので信頼できるものと思われています。しかし、調査の方法によっては実態を反映していない結果が得られることも多いので注意が必要です。本書のデータは架空のもので、取りあえずそういった判断は保留にしていますが、例えば、丸の内などのビジネス街で実施したアンケートと、吉祥寺などのファッション街で実施したアンケートでは、結果が違ってくる可能性が大です。住宅街だとさらに異なる結果になるかもしれません。そもそも、アンケートに答えてくれる人と答

えてくれない人で違いがあるかもしれません。多数のデータを集めたいからといって、インターネットでアンケートを取ると、パソコンやインターネットの使い方に慣れた人の回答しか得られなかったり、興味のない人からは回答がもらえなかったりする可能性もあるのです。

こういった「サンプルの偏り」をバイアスと呼びます。調査する場合にも、分析結果を読み解く場合にもバイアスには十分に注意する必要があります。目的があって特定のサンプルを集める場合もありますが、普通はランダムにサンプルを選ぶのが理想的です。

2

全体の傾向や特徴を
表にまとめて整理する

2-1 意外に簡単！ データ集計と整理の極意



データが入力できたら、細かい分析をする前に全体像を見ておくといいわね。



数字ばかりで全体像ってあんまり見えないんですけど。



適当に区切ってみるといいわよ。アプリのダウンロード数が0～4個の人は何人、5～9個の人は何人って感じの、度数分布表を作ってみましょう。



度数？……って何ですか。



^{ひんど}頻度のことね。要するに「何人いるか」とか「何回登場したか」といった値のこと。度数分布表を作れば全体的な傾向や特徴が分かるわよ。

各行に入力された1件1件のデータは生のデータなので、分析のためにはデータを並べ替えたり集計したりして整理しておく必要があります。ここでは、まず、度数分布表を作成してデータを集計します。さらに、その表のデータをグラフ化して全体像や特徴をひと目で分かるようにします。具体的な操作は後ほど見ることにして、ここではどのような流れでグラフを作成するかを確認しておきましょう。

なお、**分布とは、どの値がどれぐらいの頻度で（あるいはどれぐらいの確率で）現れるか**ということです。

♂ キーワード

度数…P.220
度数分布表…P.221