本書の特徴

算数の言葉で説明

できる限り多くのみなさんに統計学の真髄を伝えたいと考えたので、すべて算数の言葉で説明することにしました。文字式を変形したり、方程式を解いたりといった文字式の運算はしません。

算数には負の数が出てきませんから、この本でも負の数を用いずに統計学の数理を解説していきます。逆に、中学校で負の数を習った人にとっては、まだるこしく感じる箇所もあるかもしれません。

数学を使わないで解説するということは、数学が分かる人にとってはかえって読みづらいものです。これは、漢字が読める人にとって、平仮名だけで書かれた文章の方が漢字交じりで書かれた文章よりも読みづらいのと似ています。易しく書きすぎるとかえって読みにくいのです。易しすぎて読みづらいと思った人は、自分が数学ができるからそう感じるのだと納得し自信を持っていただければと考えます。

小学校卒業レベルの実力をお持ちであれば、小数・分数の四則演算はおできになるでしょう。しかし、割合の3用法、速さの3用法などの分野は意外とてこずるのではないでしょうか。そこで、算数で既習の分野であっても復習しながら説明を進めています。

応用できる検定の例を紹介

統計の数理を説明しようとすると、どうしても教科書っぽくなってしまいます。そこで、なるべくみなさんに馴染みのあるような話題で検定の例を紹介することにしました。

一つ目は、独立性の検定です。2018年、私立医学部の入試で女子受験者に不利な得点操作をしていたことが話題になりました。医学部の男女の合格率の比を計算して一覧表にした記事もありましたが、どれくらいの差があれば不正が行なわれているかまでは言及していませんでした。独立性の検定という手法を用いると、男女別の合格・不合格の人数から、不正が行なわれている可能性のある大学を統計学的にあぶりだすことができます（3章§6）。

二つ目は、適合度検定です。みなさんが所属しているクラスやサークル、コミュニティで、「このコミュニティは血液型B型の人が多いなあ」などと特定の血液型の割合が多いと感じたことはありませんか。実は、調べてみるとこういう場合でも日本人の血液型組成から見ると誤差の範囲内であることが分かったりします。これはちょうど、コインを100回投げて表と裏の出る回数がぴったり50回ずつになることは逆に珍しく、50回から少しずれた回数になることと似ています。コミュニティの血液型組成のずれも、そのような確率的なずれの範囲に収まっている場合が多いのです。

AKB48、乃木坂46、欅坂46のメンバー（以下、AKBメンバーと称する）168人の血液型を調べてみると、O型の人数が一番多くなっています（2019年1月現在）。ご存知のように日本人の血液型で一番多いのはA型です。このAKBメンバーの血液型組成は、日本人の血液型組成から比べて確率的なずれの範囲に収まっているのでしょうか。収まっていなければ、AKBメンバーは日本人の中でも血液型と相関のある基準によって選抜された特別な集団、真のスーパーアイドル（？）ということになるかもしれません。

適合度検定を用いると、AKBメンバーの血液型組成が、日本人の血

液型組成と比べてO型が多い特別な集団であるか否かを統計学的に判定することができます（3章§7）。

これらの例が学習の駆動力となれば幸いです。

復習問題

本文中では「問題」、「復習問題」が数多く掲載されています。

「問題」の方は解く必要はありません。「問題」とあるので問題を解いてからでないと読み進めてはいけないと思う人がいるかもしれませんが、その必要はありません。「問題」の意図をくみ取ったら、すぐにそのあとに続く解説・答えを読み進めていきましょう。「問題」は説明のポイントを明確にするためのもので、みなさんの実力を試そうとして出しているわけではありません。

ただし、「復習問題」の方は既出の説明事項の確認ですから、ぜひ解いてみてください。すべての章に「復習問題」が付いているわけではありませんが、統計学の本筋でここは押さえておいて欲しいというところでは、「復習問題」を用意してあります。

「問題」の解説を読んで理解したと思っていても、理解が浅い場合がほとんどです。手を動かして類題を解いてみないと、理解は深まらないものです。「復習問題」を解いて頭と体で統計学を理解して欲しいと思います。「復習問題」は解き易いように穴埋め式になっています。理解を深めるためにぜひチャレンジして欲しいと思います。

高校までに習う統計学分野の事項を網羅

本文中では「小学校〇年生で習ったように」という文言が何か所か出てきます。これはみなさんに既習事項を思い出して欲しいと思って書いているわけではありません。ましてや「小学校〇年生で習っただろ」とみなさんを強迫しているわけでもありません。むしろ、多くのみなさんにとって、小学校〇年生で習ったことがない事項である場合も多いと思われます。

あえて触れたのは、「この部分は、現在の文科省のカリキュラムでは小学校〇年生で習う事項である」ということを確認して欲しいからです。統計学への社会的要請が高まる中、文科省の指導要領では、統計学に関連する単元を低学年に移行するように改変しつつあります。いまの人たちは、小学校〇年生からこんなことまで勉強するんだと、社会の現状を実感して欲しいのです。

この本では、現在の小・中・高で習う統計学に関する分野の事項を網羅しました。これは社会人が身に付ける最低限の統計学的素養であると思うからです。

Excelでの計算法

「データの平均・標準偏差の値を計算する」「2つの量に関するデータの相関係数を計算する」などは表計算ソフトExcel（エクセル）を用いて処理することが可能です。小学校の算数の授業でも表計算ソフトを用いています。一度手計算をして値の求め方を知った後は、表計算ソフトを用いて値を計算できるようになれば現実の問題にも対応できます。そこで、この本では表計算ソフトの使い方まで説明することにしました。

本書の構成と読み方

本書の構成を説明するとともに、読み方についても触れておきましょう。

本書は、本編、Column（コラム）、補習の3つのパートから成り立っています。

本編にはこの本のテーマとなる統計学の真髄が、Columnには本編の補足や学校で習う統計学が、補習では読者がつまずきそうな単元の詳しい説明が書かれています。

次ページの表にあるように、本編は1章の「記述統計」で4節、2章の「正規分布」で4節、3章の「推測統計」で8節からなります。これら本編をしっかりと読めば、統計学の真髄を理解することができます。

ですから、初めてこの本を読む方は、ここだけを選んで読むことをおすすめします。本を全部読むのに越したことはありませんが、本編で統計学の幹となるところを固めてから、枝葉に手を伸ばしていくのがよいでしょう。

さらには、本編の中でも3章の§6 独立性の検定、§7 適合度検定は、身近で馴染みのある話題を提供し、読者のモチベーションを高めるために用意した節ですから、読まなくても統計学の真髄を理解することはできます。また、少し納得ができないところが出てくるかもしれませんが、1章の§4も読まないで済ませることもできるでしょう。

ですから、この本を読むと決めた方には、最低でも次を読んでいただきたいです。

第1章　§1、§2、§3　　第2章　§1、§2、§3、§4
第3章　§1、§2、§3、§4、§5、§8

	本編	Column（コラム）	補習
第1章 記述統計	1：データを整理しよう	1：階級の取り方の目安を知ろう	1：比と割合を復習しよう
	2：平均を計算しよう	2：代表値は３つある 3：仮平均で楽々計算	2：√（ルート）って何だ
	3：「散らばり」を捉える	4：箱ひげ図で散らばりを知ろう 5：Excel で計算しよう	
	4：度数分布表から平均・分散を求める	6：もう１つの「分散の求め方」	
第2章 正規分布	1：一方向に図形を伸ばす 2：モデルにあてはめる 3：正規分布の形を知ろう	7：無限和でも有限値になる 8：エクセルで正規分布を知る	
	4：正規分布をモデルとして使おう	9：偏差値なんて怖くない 10：正規分布で近似できるとき、できないとき	
第3章 推測統計	1：確率って何？ 2：「平均データ」を使いこなそう 3：推測統計の枠組みを知ろう 4：これが検定だ	11：検定の結果が間違うとき	
	5：標本が大きい場合に検定しよう	12：標本が小さい場合に検定しよう	
	6：独立性を検定しよう 7：適合度を検定しよう 8：区間推定しよう（標本のサイズが大きいとき）	13：区間推定しよう（標本のサイズが小さいとき） 14：視聴率調査には誤差がある	
		Appendix　２変量のデータの相関を知ろう	

これだけ読めば本書を読み切ったと言っても過言ではありません。この本を手に取っていただいたからには、必ずや統計学の真髄を掴んでほしいと願っています。

なお、目次を見ると分かるように、これら本編は連続して読めるわけではありません。間にColumnが挟まっています。ですから、初めてこの本を読む人は、本編だけを選んで読んでいってください。

では、Columnには何が書いてあるのか。Columnには、真髄の理解にはさほど重要でない周辺の話題や本編で理由なしで用いている事実の説明が書かれています。

Column 2、3、4、Appendixは、真髄を理解するためには必要ありませんが、高校までで習う統計学の事項なので、この本でも解説しました。

Column 5、8は、Excelの使い方についての基本的な説明です。Excelの詳しい使い方は、バージョンにより微妙な差がありますから、ネットで最新情報を得るのがよいと考えます。

Column 1、9、10、11は、さらっと読める統計学の知識的内容です。本編に比べれば難しい内容ではないので、本編を読んでいて疲れたときに気晴らしで読んでみるとよいでしょう。

Column 6、7は、算数の言葉で書かれていますが、内容的には数学を扱っていて難しいです。本編で説明されている事実が気になった人だけ読めばよいでしょう。

Column 12、13は、小さい標本についての理論です。本編にある3章の§5、§8での大きい標本についての理論が分かっていれば、小さい標本の場合もパラレルに理解できるであろうと考えColumnにしました。

Column 14は視聴率の誤差について調べるという興味ある話題ですが、難易度が高いのでColumn扱いにしました。

補習は2編あります。小学校で習う「割合」と中学校で習う「ルート」です。「割合」はご存じだと思いますが、間違いやすい箇所があるので言及しておきました。「ルート」も日常生活の中では用いませんから、一度習ったことがあっても忘れている人が多いでしょう。今一度確認してみてください。

次に、本書に限らず解説本・参考書を読むときのコツについて話しておきます。

この本は参考書ですから、小説を読むように最初から読んでそのまますらすらと理解できる本ではありません。初めて学ぶ統計学の用語が多く出てきます。参考書を読むときにネックとなることの一つは未知なる用語なのです。

用語を一度聞いただけで納得し、次に聞いたときにありありと意味が分かるという人は稀です。また、公式や計算手順についても、一度なぞっただけですぐに再現できる人はそうはいません。解説を読み進めていく中で、あれっ、この用語ってどういう意味だっけ、計算の仕方が分からない、と立ち止まってしまうことの方が普通です。そういうときは、面倒くさがらずにその用語や計算手順が初めて出てくるところに戻って、用語の意味、使い方、計算手順を確認してほしいと思います。 公式・計算手順の確認のためには、「復習問題」を解いてみるとよいでしょう。

用語の意味を確かめるには索引からたどるのが通常ですが、それが面倒だという意見もあったので、リクエストのあった用語に関してWeb上に簡単なリストを作りました。意味が想起できるような表現にしてあ

「の」が付いている数（12）が「もとにする数」です。30％＝0.3ですから、

$$\underset{(\text{もとにする数})}{12} \times \underset{(\text{割合})}{0.3} = \underset{(\text{比べられる数})}{0.36}$$

となります。

復習問題1

(1)　12 をもとにすると、9 の割合はいくつですか。

(2)　2.4 を 1 とすると、0.6 はいくつに当たりますか。

(3)　5 をもとにすると、2 の割合はいくつですか。

(4)　2 を 1 とすると、0.7 はいくつに当たりますか。

(5)　9 の 40％はいくつですか。

(1)　ア ÷ イ ＝ ウ　　9の割合はウです。

(2)　エ ÷ オ ＝ カ　　0.6の割合はカです。

(3)　キ ÷ ク ＝ ケ　　2の割合はケです。

(4)　コ ÷ サ ＝ シ　　0.7の割合はシです。

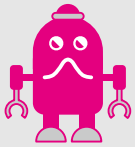
(5)　ス × セ ＝ ソ　　9の40％はソです。

解答

ア 9	イ 12	ウ 0.75	エ 0.6	オ 2.4	カ 0.25
キ 2	ク 5	ケ 0.4	コ 0.7	サ 2	シ 0.35
ス 9	セ 0.4	ソ 3.6			

記述統計！

SECTION 1



データを整理しよう

さて、ここから統計学の本題に入っていきます。

「統計を取る」、「データを取る」という言葉を聞いたことはあるでしょう。「データを取る」と一口にいいますが、解析に使えるようなデータを取ることは、実は難しいことなのです。東京都に住んでいるすべての12歳男子の身長の値を知ることは現実的にはできません。また、ある製品の耐用年数を調べるには、壊れるまで待たなければなりませんから時間がかかります。データを得るための方法論で1冊の本になってしまうくらい語るべきことはあります。しかし、そこから解説していくのは統計学の真髄を掌握するには迂遠（うえん）ですから、データはすでに手元にあるものとして、データの整理の仕方から解説していくことにします。

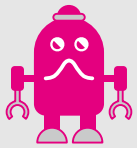
「あるクラスで、男子の身長のデータを取る」といえば、このクラスの男子の身長を表す数値を集めることです。「統計を取る」、「データを取る」といえば、特定の集団についての数値を集めることを指しています。

数値を集めただけでは、その集団が持っている数値に関する特性を理解することはできません。集めた数値から表やグラフを作ることで、集団が持つ数値の特性を理解することが可能になります。

すでに、みなさんは小学校3年生で、表の作り方と棒グラフの描き方を勉強しました。この節では、**ヒストグラム**と呼ばれるグラフの描き方

推測統計！

SECTION 8



区間推定しよう

（標本のサイズが大きいとき）

前の節までは検定を紹介しました。この節からは推定を紹介します。

推定の中でも区間推定を紹介しましょう。

例えば、東京都に住んでいる20代の平均年収は「250万円から350万円までの間に入っているのではないかな」と予測したとします。これは平均年収をある幅を持って範囲で予測しています。これが“区間”を予測したということです。しかし、統計学の答えとしては、これだけでは足りません。予測区間だけでなく、その予測がどれくらい信頼のおける予測であるかという信頼度にまで言及すると、統計学の区間推定になります。

区間推定とは、母平均や母分散の値をある範囲を持って予想することです。次の問題は、標本から母平均を区間推定する問題です。この本では母平均の推定しか取り上げません。問題文中の95％信頼区間という言葉は解説の中で説明します。

問題 34 **K製菓のあるラインで作られたせんべい100枚を取り出して重さについて調べたところ、平均は16.0g、標準偏差は0.7gでした。このとき、このラインで作られるせんべいの重さの平均について95％信頼区間を求めてください。**

この問題では、標本は手元にある100枚のせんべいです。母集団は何でしょうか。母集団はK製菓のラインで作られるせんべい全体です。

しかし、せんべい全体の枚数は何枚かと聞かれると困ります。時間をかければ何枚でも作ることができるからです。無限枚のせんべいを作ることができると考えて、このような母集団を無限母集団といいます。実際には、無限枚のせんべいを作ることはできませんが……。

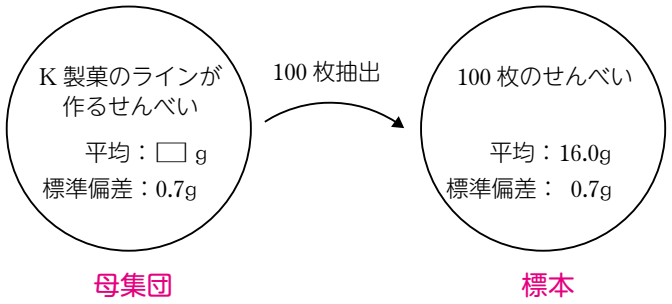
母集団の平均のことを母平均といいました。区間推定とは、ある範囲で母平均や母分散を推定することです。この問題は標本から母平均を区間推定しなさいという問題です。

区間推定でなく、「1つの値で母平均を推定せよ」と言われた場合には何と答えますか。この場合は素直に標本の平均16.0gを答えればよいですね。これを**点推定**といいます。

区間推定の場合には、分からない母平均をとりあえず□gとします。

母標準偏差は、標本の標準偏差に等しいとして、0.7gであると考えます。標本が大きいときの検定と同じように0.7gと決めてしまいます。

母集団の分布は、平均□g、標準偏差0.7gです。



この母集団から100枚を取り出してその平均を考えます。この平均を