

無断使用をお断りします。日科技連出版社

高校数学から
はじめる
統計学

竹士 伊知郎 著

JUJIE

日科技連

はじめに

近年、高校数学に「統計学」が本格的に「入ってくる」ということが大きな話題になっている。

2022年度に入学する新高校生に対する文部科学省の高等学校学習指導要領によれば、具体的には、「数学Ⅰ」の「データの分析」で「仮説検定の考え方」も扱うことになり、さらに、「数学Ⅱ」が「統計学的な推測」，「数列」，「数学と社会生活」の3項目となった。

高校の数学の内容がそれほど話題になるのは、一つは「大学の入学試験」に大きな影響を及ぼすからであろう。従来の学習指導要領にあった「確率分布と統計学的な推測」は、多くの大学で入試の出題範囲から外されていたため、高校では実際には教えられないことも多かったと聞く。しかし、今後は、数学Ⅰはもちろん数学Ⅱも理系文系問わず必須科目であるので、各大学の数学の入試科目や大学入学共通テストの範囲に「統計学」が一定の割合を占めることが予想される。

本書は、原則として2023年度以降の高校2年生が学習する数学Ⅱの「統計学」に関する内容を出発点にして、多くの読者に、本来実学であるはずの「統計学」への興味をもっていただき、さらに「統計学」を利用した各種の問題解決や課題達成に役立ててほしいという思いから執筆した。

もとより、高校数学の「統計学」の内容を批判したり、改善を提言するものではないことを最初に断っておきたい。

高校数学の「中」に入ってしまった「統計学」としては、2023年度以降の教科書に記述される内容は、妥当なものであり、十分納得できるものである。そのうえで、いくつかの検討すべき点も垣間見える。

本文中で何度も言及するように、筆者は、統計学は実学であって、「数学」とは異なる分野の学問であると考え、「統計学」が実学として、政治・経済、医療、製造、サービスといった多くの分野で、なぜ重宝され、広く使われているのか、その理由は何なのかといった疑問に対して、残念ながら高校数学の統計学では、十分記述されているとはいえない。

教科書では、多くの例題が示されてはいるが、その例は、さいころであったり、硬貨の表裏であったり、高校生の身長やテストの成績であったりするものが大半である。これらは高校生にとって身近なもので受け入れやすい例示ではあるが、他方「こんなことにしか統計学は使えないのか」、「高校で勉強する数学は社会に出たらもう使わない」との誤解を招きかねないと危惧している。

そこで、本書は「高校数学での統計学」の内容に触れたあと、この内容が「実学としての統計学」にどうつながるかの道程を示したい。すなわち、高校数学の内容が「実学としての統計学」へどのように昇華していくのかを述べている。

また、各章には実践的な例題を設けている、理解の一助としてほしい。

[高校生、高校教員の方へ]

高校数学での統計学を学んだ後、「統計学」がどのように成り立っているのか、実際にどのような場面で使われているのか、どのような注意が必要なのかなどについて、「実学としての統計学」を読んで学んでいただきたい。また、教科書や参考書の記述だけで十分理解できないことがあれば、該当の箇所をご覧になることで「そんな意味なのか、そんなことなのか、そう考えればいいのか」といった新たな理解の促進や発見が必ずあるはずである。

すべての高校生とはいわないが、文系理系問わず、多くの皆さんは、今後大学や企業などの職場において、「品質管理」、「品質保証」、「品質(クオリティ)マネジメント」といったものに必ず出会うはずである。ここで使われるのが「統計学」を背景とした「統計的方法」、「統計手法」、「QC手法」などと呼ばれる問題解決のためのツールである。つまり、高校数学の統計学は、こういっ

た一連のツールにつながっていくのであるから、将来必ず役に立つ。10年後、20年後のあなたが、高校生であったころの自分に言い聞かせるのは、「今しっかり統計を勉強しておけば、将来必ず役に立つよ」といった言葉かもしれない。

また、大学進学や就職といった進路に関連して、話題のデータサイエンスに興味がある方もおられるかもしれない。しかし、データサイエンスの勉強に際して、まず前提になるのは「統計学」の知識である。キャッチボールのやり方や基本のルールも知らずに野球選手にはなれない。

[大学生、社会人・一般の方へ]

「統計学」を初めて学ぶ、あるいは過去に少しかじっただけといった方には、「高校数学での統計学」で、今の高校生はこんなことを学習している、ということをもっと知っていただきたい。そのうえで、「実学としての統計学」に読み進んでいただき、統計学のおもしろさ、重要性、そしてご自身の問題を解決する道具としての使い方を、ぜひ学んでいただきたい。

数年後には、高校時代にこのような形で統計学を学んだ方たちが、確実に皆さんの仲間に加わる。変に構える必要はないが、彼らに対して、実社会ではこんな風に統計学を活用して、日々の品質管理などの活動を行っているということを実例とともに示してあげてほしい。自信のない方は、本書でもう一度学習をしてもらえれば幸いである。

なお、現在多くの統計学に関する書籍が出版されているが、そこで使われている統計用語や統計学に関する記号と、高校数学での表記が異なるものも多い。読者の混乱を避けるために、これらの対照表を掲載している。本来、ひとつの用語や記号で統一すべきであるが、現実として、多くのものが混在していることは事実であるので、はじめて学ぶ高校生や大学生・社会人の方のために「同義語、同義の記号である場合と厳格に区別すべき場合」を具体的に示している。また同様に、統計学で極めて重要な役割を果たす「正規分布表」などの数値表についても、高校数学で用いられているものと、一般的に用いられているもの

は体裁が異なる。これも、双方の数値表の見方・使い方を併記することで、より理解が深まるよう工夫をしている。

筆者は数学の専門家ではない。しかしながら、かれこれ40年以上も、統計学を使って品質管理を行うための一連の学問体系としての「統計的品質管理」を生業の一つとしてきた。企業人であったころは、自ら各種の統計的方法を駆使して、所属した会社の品質問題に取り組むとともに部下や後輩の指導も行った。あわせて、主として(一財)日本科学技術連盟が主催する社会人向け各種セミナーの講師、大学での「統計学」に関する講義、「統計的方法」、「TQM」、「品質管理検定のテキスト・模擬問題集」などに関する書籍の執筆を継続して行っている。

現在は、「使える統計的方法をわかりやすく説く」ことを目的にして、各種の講義・講演、企業指導、書籍の執筆などを通して、大学生や企業の方のみならず、高校生、一般の方を含めてできるだけ多くの方に、知識や経験をお伝えしたいと念じている。

この小さな本が、今拙宅の窓から眺める咲き始めた花水木のように、読者の手の中で一輪一輪開花してくれれば望外の喜びである。

2023年5月

竹士 伊知郎

目 次

はじめに *iii*

第1章 データの整理	1
1.1 高校数学でのデータの整理	2
1.1.1 データの整理	2
1.1.2 データの代表値	4
1.1.3 データの散らばり	5
1.2 実学としてのデータの整理	9
1.2.1 データの種類	10
1.2.2 ヒストグラム・度数分布表	12
1.2.3 基本統計量	16
1.3 本章の例題	19
第2章 標本調査と母集団の分布	25
2.1 高校数学での標本調査と母集団の分布	26
2.1.1 母集団と標本	26
2.2 実学としての標本調査と母集団の分布	28
2.2.1 母集団とサンプル	28
2.2.2 母集団の推測	32
第3章 確率変数と確率分布	35
3.1 高校数学での確率変数と確率分布	36
3.1.1 確率変数と確率分布	36

3.1.2	確率変数の平均と分散	37
3.1.3	確率変数の和と積	40
3.1.4	連続型確率変数	44
3.2	実学としての確率変数と確率分布	47
3.2.1	確率変数と確率分布	49
3.2.2	期待値と分散	51
3.3	本章の例題	55
第4章	正規分布と二項分布	59
4.1	高校数学での正規分布と二項分布	60
4.1.1	正規分布	60
4.1.2	二項分布	62
4.2	実学としての正規分布と二項分布	66
4.2.1	正規分布	68
4.2.2	標準正規分布・正規分布表の見方	69
4.2.3	二項分布	74
4.3	本章の例題	76
第5章	統計量の分布	85
5.1	高校数学での統計量の分布	86
5.1.1	標本平均の確率分布	86
5.2	実学としての統計量の分布	87
5.2.1	統計量の分布	88
5.2.2	大数の法則と中心極限定理	92
5.3	本章の例題	93
第6章	検定・推定	95
6.1	高校数学での検定・推定	96

6.1.1	母平均の推定	96
6.1.2	母比率の推定	99
6.1.3	仮説検定	100
6.2	実学としての検定・推定	102
6.2.1	検定	104
6.2.2	推定	112
6.2.3	計量値の検定・推定	114
6.2.4	計数値の検定・推定	116
6.3	本章の例題	118
	おわりに	129
	付 表	131
	引用・参考文献	137
	索 引	139

る。このようなばらつきについて、サンプル間のばらつきをサンプリング誤差、測定のばらつきを測定誤差とよぶ。

2.2.2 母集団の推測

高校数学の統計では、平均や分散、標準偏差といった用語が頻出するが、これらが母集団についてのものなのか、サンプル(標本)についてのものなのかがあいまいとなっている場合も多い。

教科書では、あるクラスの男子生徒の身長やテストの点数などがデータの例としてよく扱われている。この場合、あるクラス全体を母集団と考えると、クラスすべての生徒のデータから求めた平均や分散は、母集団の平均、分散、すなわち母平均、母分散ということになるだろう。

しかし、日本全体の高校2年生男子の身長を母集団と考えると、母集団を構成するすべての生徒の身長を測定してまとめることは容易ではない。この場合、母集団の一部の生徒を指定して身長を測定して、それらのデータから平均値や分散を求めることを行う。指定された生徒はサンプル(標本)であり、サンプルは母集団全体からランダムにサンプリングされるものとする。

このようにして得られたサンプルから求めた平均値や分散は、母平均や母分散を推測する目的に使われるのだが、母平均、母分散と必ずしも一致するものではない。

統計学では、一般に未知である**母平均**や**母分散**(これらを**母数**と呼ぶ)を推測するために、サンプルから求めた平均値や分散(これらを**統計量**と呼ぶ)を扱うのである。したがって、母集団に関することを述べているのか、サンプルに関することを述べているのかを常に意識することは極めて重要である。

慣れてくれば文脈からでも、母集団についての母数のことか、サンプルについての統計量のことかは、判断できると思われるが、万一にも誤解を招かないためにも、きちんと区別をして表現したり理解することが求められる。

母集団・母数とサンプル・統計量の関係を図 2.3 に表す。

統計学では、図のように、一般に未知である母平均や母分散をデータから求

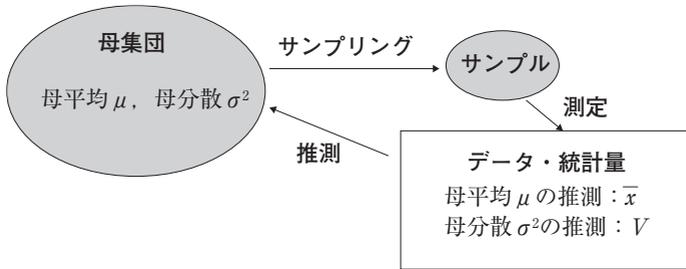


図 2.3 母集団とサンプル

表 2.3 母数と統計量

区分	用語	記号	式
母数	母平均 (平均, 期待値ともいう)	μ (m : 高校数学)	$E(X) = \mu$ $E(X)$ は確率変数 X の期待値を表す(第3章参照).
	母分散 (分散ともいう)	σ^2	$V(X) = E[(X - E(X))^2] = \sigma^2$ $V(X)$ は確率変数 X の分散を表す(第3章参照).
	母標準偏差 (標準偏差ともいう)	σ	$\sqrt{V(X)} = D(X) = \sigma(X) = \sigma$
統計量	平均値 (標本平均ともいう)	\bar{x}	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
	分散 (不偏分散, 標本分散ともいう)	V	$V = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
	標準偏差 (標本標準偏差ともいう)	s	$s = \sqrt{V} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$

めた統計量を使って推測を行うのである。表 2.3 に母数と統計量について整理する。

表 3.13 に第 3 章における用語と記号の対照表を示す。

表 3.13 用語と記号の対照表

高校数学	実学	意味と注
平均 $E(X)$ 期待値ともいう。	期待値 $E(X)$ 平均 $E(X)$, 母平均 μ ともいう。	確率変数 X の平均 $E(X)=\mu$ サンプルについては, 平均値 \bar{x} を求める. $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ これを標本平均ともいう。
分散 $V(X)$	分散 $V(X)$ 母分散 σ^2	確率変数 X からその平均を引いた変数の 2 乗の期待値, $V(X)=E\{(X-\mu)^2\}=\sigma^2$ であり, 母分散ということもある. ただし, この定義は母集団についてのものである. サンプルについては, 平均値 \bar{x} からの偏差の 2 乗の和を $(n-1)$ (自由度) で割ったものであって, $V = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ と求める. これを不偏分散 V , 標本分散 V ともいう。
標準偏差 $\sigma(X)$	標準偏差 $D(X)$, 母標準偏差 σ	分散の正の平方根 $\sqrt{V(X)}=D(X)=\sigma(X)=\sigma$ であり, 母標準偏差ということもある. 分散の場合と同様に, こ

表 3.13 つづき

高校数学	実学	意味と注
		<p>の定義は母集団のものである。</p> <p>サンプルについては、不偏分散の正の平方根として求める。</p> $s = \sqrt{V} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ <p>これを標本標準偏差 s ともいう。</p>

3.2.1 確率変数と確率分布

(1) 確率変数と確率分布

母集団からサンプルをとるたびに、そのサンプルは異なり、値はばらつく。では、同じ母集団からとられたデータの一つひとつやその全体の様子には、何か性質や規則のようなものはないのだろうか。

統計では、これらを確率変数とその分布である確率分布とよぶ。確率変数とは、「とってみないとわからない、とるたびに異なる値のこと」、分布は、「ばらつきをもった集団の姿形」のことである。したがって確率分布は、「確率変数の集団としての性質や規則性を示すもの」ということになる。

(2) 連続型確率変数

質量、長さ、強度、時間などのように「はかる」量を計量値という。計量値はどこまでも細かく測定できるので、とり得る値が連続的であると考えられる。このような場合に用いられる確率変数を連続型確率変数といい、その分布を連続分布という。計量値の分布の代表的なものが正規分布である。

連続分布は確率密度関数 $f(x)$ を用いて表現され、以下のような性質がある。

① $f(x) \geq 0$

索引

【英数字】

1つの母不適合品率の検定・推定 126
 1つの母分散の検定・推定 123
 1つの母平均の検定・推定(母分散既知) 118
 1つの母平均の検定・推定(母分散未知) 121
 p 値 109
 s 10
 t 表 90
 t 分布 89,121
 V 10
 σ 10
 σ^2 10
 χ^2 表 91
 χ^2 分布 91

【か行】

階級 2
 ——値 3
 確率分布 36,49
 確率変数 29,36,49
 ——の標準偏差 37
 ——の分散 37
 ——の平均値 37,50
 確率密度関数 49
 仮説検定 102,103
 片側仮説 107
 片側検定 107
 規格 12

棄却 102
 ——域 107
 危険率 103,107
 期待値 37,48,51
 ——の性質 52
 基本統計量 16
 帰無仮説 102,107
 共分散 53
 区間 113
 ——推定 112
 計数值 11,29
 ——の検定・推定 116
 計量値 11,29
 ——の検定・推定 114
 検出力 107
 検定 104
 検定統計量 108
 工程 30
 誤差 31

【さ行】

最頻値 5
 算術平均 10
 サンプルング 29,30
 サンプル 29
 ——サイズ 29
 ——の大きさ 29
 ——の分散 10
 ——の平均値 10
 四分位数 5
 四分位範囲 6
 四分位偏差 6

自由度 18
 順位データ 11
 試料 29
 信頼区間 98,113
 信頼度 98,103
 信頼率 103,113
 推定 98
 正規分布 67,68
 — 曲線 60,61
 — 表 63,67

【た 行】

第1四分位数 6
 第1種の誤り 103,107
 第1種の過誤 103,107
 第2四分位数 6
 第2種の誤り 107
 第2種の過誤 107
 第3四分位数 6
 大数の法則 92
 代表値 4
 対立仮説 102,107
 中央値 5,16
 抽出 29
 中心極限定理 92
 散らばり 5,9
 点推定 112,113
 統計的仮説検定 103
 統計量 16,114
 — の分布 88
 同時確率分布 41
 独立 42,53
 度数 2
 — 分布表 2,4,13

【な 行】

二項分布 64,67,74

【は 行】

箱ひげ図 6,7
 外れ値 7
 ばらつき 9,16
 範囲 5
 ヒストグラム 3,12
 非復元抽出 27
 標準化 67,70
 標準正規分布 61,67,70
 標準偏差 10,37,48
 標本 29
 — の大きさ 29
 — 標準偏差 99
 — 比率 99,103
 — 分散 33,48,117
 — 平均 33,86,88
 復元抽出 27
 不適合品率 11,67
 不偏分散 33,48,114
 分散 13,17,37
 — の加法性 54
 分散の性質 54
 分布 12
 分類データ 11
 平均値 4,10,16
 平方和 17,91
 偏差 8
 — 平方和 7
 変動係数 18
 変量 26,29
 母集団 27,29

母数 32,33,114

母標準偏差 10,48

母比率 99,103

母不適合品率 67,103

母分散 10,33

母平均 10,33

——の推定 96,98

【ま 行】

無限母集団 29

無作為抽出 27,29

無作為標本 27,29

メジアン 5

メディアン 16

【や 行】

有意水準 102,103,109

有限母集団 29

【ら 行】

ランダムサンプリング 29,31

ランダムサンプル 29

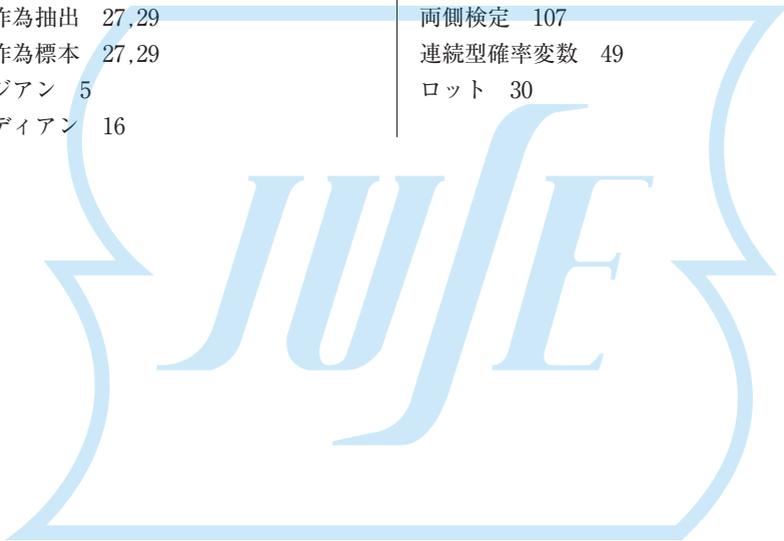
離散型確率変数 50

両側仮説 107

両側検定 107

連続型確率変数 49

ロット 30



●著者紹介

竹士 伊知郎(ちくし いちろう)

1979年 京都大学工学部卒業、(株)中山製鋼所入社。

金沢大学大学院自然科学研究科博士後期課程修了、博士(工学)。

現在 QMビューローちくし代表、関西大学化学生命工学部非常勤講師、

(一財)日本科学技術連盟嘱託。

日本科学技術連盟などの団体、大学、企業において、品質管理・統計分野の講義、指導、コンサルティングを行っている。

主な品質管理・統計分野の著書に、『学びたい 知っておきたい 統計的方法』、『ことばの式でわかる統計的方法の極意』(単著、日科技連出版社)、『QC検定受検テキストシリーズ』、『QC検定対応問題・解説集シリーズ』、『QC検定模擬問題集シリーズ』、『速効! QC検定シリーズ』、『TQMの基本と進め方』(いずれも共著、日科技連出版社)がある。

高校数学からはじめる統計学

2023年6月30日 第1刷発行

著者 竹士 伊知郎

発行人 戸羽 節文

発行所 株式会社 日科技連出版社

〒151-0051 東京都渋谷区千駄ヶ谷5-15-5

DSビル

電話 出版 03-5379-1244

営業 03-5379-1238

検印

省略

Printed in Japan

印刷・製本 港北メディアサービス(株)

© Ichiro Chikushi 2023

ISBN 978-4-8171-9775-7

URL <https://www.juse-p.co.jp/>

本書の全部または一部を無断でコピー、スキャン、デジタル化などの複製をすることは著作権法上での例外を除き禁じられています。本書を代行業者等の第三者に依頼してスキャンやデジタル化することは、たとえ個人や家庭内での利用でも著作権法違反です。