

算数・数学科で学ぶ統計分野の内容

数研出版数学編集部

近年、IoT機器の普及によるビッグデータの活用やChatGPTに代表されるAIの進化に伴い、統計的なものの見方・考え方を育成・習得することの重要性がますます高まっています。

初等・中等教育においては算数・数学科がその役割の中心を担っており、学習指導要領で定められる学習内容も近年拡大されてきています。そこで本稿では、小学校～高等学校の算数・数学科で学習する統計分野の内容について、高等学校の公民科・地理歴史科の先生方により具体的にイメージしていただくことを目的として、解説いたします。

1. 統計的手法の大枠

具体的な内容の前に、統計的手法の大枠を簡単に確認します。統計的手法のうち基礎的なものは、次の二つに分けることができます。

記述統計：得られたデータを整理し解釈しやすくする手法

推測統計：得られたデータから、背後にあるさらに大きな集団の状況を推測する手法

この違いを意識することで、これから解説する手法の意味や目的を整理しやすくなります。

2. 算数・数学科での統計分野の内容—総論

現行の学習指導要領下で、算数・数学科で学習する主な内容は以下ようになります(下線が記述統計の内容、それ以外が推測統計の内容です)。

小学校	6年	代表値(平均値, 中央値, 最頻値)
中学校	1年	度数分布表, ヒストグラム, 相対度数, 累積度数
	2年	四分位数・四分位範囲, 箱ひげ図
	3年	標本調査, 標本平均と母平均の関係
高等学校	数学I	分散・標準偏差, 散布図・相関係数, 仮説検定の考え方
	数学B	確率変数, 確率分布(二項分布・正規分布), 標本調査, 区間推定, 仮説検定

小・中学校や数学Iまでは記述統計が中心で、学

年が進むと推測統計の内容が増えます。一般的に推測統計の方が、必要となる確率・統計の知識が幅広く、その分難度も高くなります。まずは記述統計で手元のデータを整理・解釈する力を身につけ、その上で推測統計を用いてデータの背後にある状況を推測する力を養うよう学習する流れになっています。

3. 算数・数学科での統計分野の内容—各論

ここからは、2. で挙げた内容からピックアップし、それらの意味や解釈の注意点を解説します。

3. 0. 使用データ

今回は架空のデータで、あるクラスの生徒30人の数学と英語のテストの点数を使用します。

No.	数	英	No.	数	英	No.	数	英
1	96	75	11	58	63	21	91	79
2	74	70	12	84	70	22	94	88
3	91	70	13	58	64	23	90	78
4	89	73	14	62	73	24	100	91
5	59	90	15	61	72	25	90	78
6	59	84	16	91	80	26	58	68
7	100	90	17	53	44	27	60	70
8	92	76	18	56	49	28	65	74
9	61	83	19	92	85	29	90	78
10	53	40	20	57	55	30	58	67

単位：点(100点満点とする)

3. 1. 代表値(平均値・中央値・最頻値)

代表値は集団の中心的傾向を示す値で、データの特徴を単一の値で要約したものです。

平均値 データの値の合計÷データの個数

中央値 データを値の大きさの順に並べたときの中央の値

最頻値 データに最も多く現れている値

今回のデータではそれぞれ次のようになります。

教科	平均値	中央値	最頻値
数学	74.7	69.5	58
英語	72.6	73.5	70

単位：点

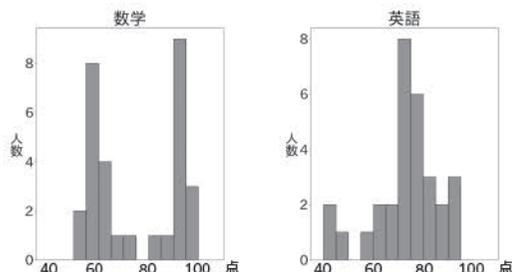
注意 単一の値である代表値だけでデータの様子(分布ともいいます)を捉えるのには限界があります。例えば「クラス全体で数学と英語どちらの点数が高いか」を考察する際、平均値は数学の方が、中央値・最頻値は英語の方が高く、判断が難しいです。

データの分布を正しく捉えるには、代表値だけでなく次のヒストグラムや箱ひげ図も合わせて考察することが大切になります。

3. 2. ヒストグラム

ヒストグラムは、データの値をいくつかの区間に分けて整理し、柱状のグラフで表したものです(この区間を階級といいます)。

今回のデータで、0～100を5点ずつの階級に分けて作ったヒストグラムは次のようになります。



ヒストグラムから、数学は点数の高い生徒と低い生徒に分かれていること、英語は多くの生徒が70～80点前後であることがわかります。このような情報は代表値からだけでは得られないので、ヒストグラムを用いて確認することの重要性がわかります。

補足 数学のように山が複数あるデータを「多峰性」のデータ、英語のように山が一つのデータを「単峰性」のデータといいます。

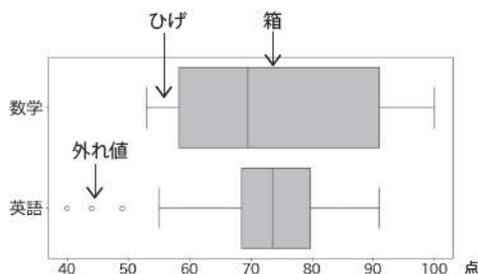
3. 3. 四分位数・箱ひげ図

四分位数はデータを値の大きさの順番に並べたときに4等分する位置にくる三つの値(真ん中の四分位数は中央値に一致します)であり、箱ひげ図は四分位数と最大値・最小値を用いてデータの分布を表現するグラフです。

今回のデータでは、それぞれ次のようになります。

教科	最小値	四分位数		最大値
数学	53	58	69.5	91
英語	40	68	73.5	80

単位：点



箱ひげ図はデータの中央の50%が入る範囲を箱で、その外側のデータの範囲をひげで表します。また、箱ひげ図では分布の中央から大きく外れた値(外れ値)を除いてかくことが多いです。外れ値は四分位数からの離れ具合で決めます。上の図では英語の外れ値のデータ三つを○で表し、それらを除いたデータで箱ひげ図をかいています。このように、箱ひげ図は中央のデータの分布を把握することを主な目的とし、外れ値など中央から離れたデータの影響を必要以上に受けたくない表現方法であるといえます。

3. 4. 分散・標準偏差

分散・標準偏差は、平均値を中心としたデータの散らばりの程度を表す値です。

分散 (データの値-平均値)²の和÷データの個数
標準偏差 $\sqrt{\text{分散}}$ (分散の正の平方根)

分散・標準偏差では、各データと平均値との差を用いて平均値からの散らばりの程度を表現します。2乗することで、正の方向への散らばりと負の方向への散らばりを区別せずに扱うことができます。

今回のデータにおける分散・標準偏差は右のようになります。

教科	分散	標準偏差
数学	285.2	16.9
英語	159.0	12.6

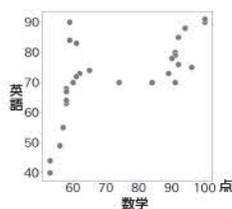
これらの値をもとにすると、数学の方が平均値からの散らばりの程度が大きいと捉えられます。

補足 分散は求める際に2乗するため、全体的な値の大きさは元のデータと大きく異なってしまう場合が多いです(単位が「点²」になってしまうイメージ)。一方で標準偏差は分散の平方根をとるため、全体的な値の大きさは元のデータと同程度に戻ります。

3. 5. 散布図・相関係数

散布図は、二つの変数の関係を見やすく表現したグラフです。

右の散布図は、今回のデータで各生徒の数学と英語の点数を点(ドット)で表したものです。数学の点数が高いと英語の点数も高い生徒が多い傾向が読み取れます。散布図は二つの変数の関係を読み取る際に有効です。



また、一方が増えると他方も増える傾向のことを正の相関関係、逆に他方が減る傾向のことを負の相関関係といいます。相関関係の強さを数値で表すのが相関係数です(具体的な計算式は省略)。

相関係数には、次のような性質があります。

- ・ -1 から 1 の間の値をとる。
- ・ 1 に近いことは正の相関関係が強い目安になる。
- ・ -1 に近いことは負の相関関係が強い目安になる。

今回のデータでは数学の点数と英語の点数の相関係数は約0.62となり、散布図の様子と同様に比較的強い正の相関関係があることを表しています。

3. 6. 標本調査・標本平均と母平均の関係・区間推定

ここからは推測統計の内容を取り上げます。

統計的な調査には、調べたい対象全体のデータを集める全数調査と、一部を抜き出して調べてその結果から全体の状況を推測する標本調査があります。コスト・時間などの制約から、多くの統計的な調査は標本調査で行われています。

標本調査では、調べたい対象全体の集合を母集団といい、抜き出された要素の集合を標本といいます。

例えば興味のあるデータについて、母集団全体の平均値(母平均)を調べたいとします。このとき全数調査を行えない場合には、標本の平均値(標本平均)から母平均を推定する、ということを行います。標本平均はあくまで抜き出された要素をもとにした値であり、母平均とは異なる値です。しかし、要素をランダムに抜き出し、かつ抜き出す要素の数がある程度大きくすれば、標本平均は母平均に近づいていく性質があることが統計的に示されています。

このように、適切な標本であれば精度よく母平均を推定することができます。

なお、標本平均などをもとに、母平均の値ではなくその値の範囲を推定することもあります。これは区間推定とよばれ、標本平均が母平均の推定値としてどの程度の精度をもつかの目安になります。

補足 標本調査の重要な点の一つが要素の選び方です。偏りなく、母集団の特徴をきちんと反映した標本になるように要素の選び方を検討する必要があります(コスト・時間などとのバランスも考慮します)。

3. 7. 仮説検定の考え方

仮説検定は、母集団に関する仮説を立て、標本から得られた結果を用いて仮説が正しいかどうかを判断する方法です。

例えば、数研出版発行教科書『数学 I』では、仮説検定の考え方として次の内容が扱われています。

- ・ ボールペンを製造する会社が、旧製品 A を改良した新製品 B を開発した。
- ・ 30人の消費者を選び、A, B のどちらが書きやすいかを選んで回答してもらったと、21人が新製品 B の方が書きやすいと回答した。
- ・ このデータをもとに、消費者全体において、B が書きやすいと感じる消費者が多いかを判断する。

30人中21人という結果は B を書きやすいと感じる消費者が多そうに感じますが、本当は A, B に差がないのに今回選ばれた30人がたまたま B を好んだだけかもしれません。仮説検定では、結果がたまたまなのか、それとも本当に A, B に差があるのかを統計的に判断します。手法の詳細は省略しますが、仮説検定はマーケティング戦略の策定や医薬品の臨床試験などでも使用される、重要な統計的手法です。興味のある方は、『数学 I』『数学 B』の教科書や一般的な統計の解説書籍をぜひご覧ください。

4. まとめ

本稿では、小学校～高等学校の算数・数学科で学習する統計分野の内容について解説しました。

特に3. でピックアップした内容は数学 I までに学習する内容で、数学 I は必修科目であるため、基本的にはすべての高校生が学習します。本稿が、公民科・地理歴史科での学習内容の検討や、数学との教科横断学習の検討などの一助になれば幸いです。

【参考文献】『数学 I』数研出版

数研出版の数学では、データから色々な統計グラフを簡単に作成できる「統計ツール」をご用意しています。右の二次元コードから、本稿のデータを用いた体験版にアクセスいただけます。

