



はじめに



「データの分析」は2012年施行の学習指導要領から高校数学の数学Ⅰに登場した単元で、ここで学習する内容は大きく分けて「データの読み取り」と「代表値の計算」になります。

「データの分析」については高校の教科書で扱われている内容は決して難しいものではなく、一つ一つを丁寧に理解して行けばよいのですが、実際に大学入試センター試験（以下、センター試験）、大学入学共通テスト（以下、共通テスト）で出題されてきた内容を見ると、教科書で扱っている内容から逸脱していることはないが、データの数が非常に多くて着眼点をしっかりもたないと判断が難しかったり、図を選ぶにも訓練が必要になるものが散見されています。共通テストでは数学Ⅰで扱われる単元は必修であるため対策が必要であり、また2次・私大試験でも出題が増えています。

そこでこの問題集は自学自習することを想定して、内容をできる限り絞り、短時間で「データの分析」の学習が一通りできるように作成しました。

- 図の読み取りについては、センター試験、共通テストなどで実際に出題された問題を「ヒストグラム」「箱ひげ図」「散布図」に分けてパターン分析し、解法をまとめました。

具体的には、「ヒストグラム」から「散布図」を選ぶとか、「散布図」から「箱ひげ図」を選ぶなど出題パターンを分析し、パターンごとに図を選別する際の着眼点が確認できるよう配慮しました。また、図の微妙な違いがわかることが必要なものについては多くの図を掲載し、違いが判断できるようにしました。

- 代表値の計算については実際に問題を解くとき効率よく計算を行うことを最優先して、その計算方法を徹底的にマスターできるよう配慮しました。また、2次・私大試験での出題を想定して他分野との融合問題も扱っています。

この問題集を仕上げることで、「データの分析」の理解を深め、自信をもって試験に臨めることを期待しています。

本書の特徴と使い方

本書の構成は次の通りです。

第1章 用語の確認 … 「データの分析」で必要な用語の定義を中心にまとめてあります。理解の助けに具体例を掲載しています。一つ一つ手を動かして自分で確認してください。

第2章 重要テーマの演習 … 重要テーマごとに定義、重要事項の確認と計算方法の確認をします。

例題と問題で一つのテーマが完成します。

例題を解き、内容を理解したら**問題**を自分の力だけで解いてみてください。

最初はゆっくりでよいので確実に！

問題の解答は別冊の解答編に掲載してあります。答え合わせをして次のテーマに進みましょう。

第3章 共通テストの対策 … 共通テストでの出題を想定した総合問題です。

第4章 2次・私大試験の対策 … 2次・私大試験での出題を想定して、教科書に掲載されていない内容、入試問題として出題可能な他分野との融合問題を掲載しています。

すべての読者は第1章から順番に学習を進めてください (p.5「本書の進め方」参照)。共通テスト対策としては、第3章まで終了したら完成です。さらに深い内容まで学習したい読者(多くの場合、2次・私大試験の対策として必要性を感じる読者)が第4章の問題を学習してください。

もくじ

第1章 用語の確認	9
1. 度数分布表	10
2. ヒストグラム	11
3. 代表値	12
4. 四分位数	14
5. 5数要約と箱ひげ図	16
6. 外れ値	18
7. 分散	20
8. 散布図・相関表	22
9. 共分散	24
10. 相関係数	25
11. 仮平均	26
12. 変量の変換	28
13. 仮説検定	30
第2章 重要テーマの演習	33
1. 度数分布表・ヒストグラム	34
2. 四分位数	38
3. 箱ひげ図(1)	40
4. 箱ひげ図(2)	44
5. 平均値	48
6. 分散・標準偏差(1)	52
7. 分散・標準偏差(2)	56
8. 散布図・相関表	60
9. 共分散, 相関係数(1)	64
10. 相関係数(2)	68
11. 変量の変換	70
第3章 共通テストの対策	75
第4章 2次・私大試験の対策	99
別冊 演習問題の解答	

1 度数分布表

調査や実験などで得られた測定値の集まりを**データ**という。

また、データを構成する値の個数を**データの大きさ**という。

データを構成する値をいくつかの区間に分け、その区間に入る値の個数を数えてまとめたものを**度数分布表**という。

度数分布表で設定される区間を**階級**、区間の幅を**階級幅**、区間の中央の値を**階級値**、各階級に属する値の個数を**度数**という。

また、各階級の度数をその階級まで合計したものを**累積度数**といい、全体に占める各階級の度数の割合を**相対度数**、各階級の相対度数をその階級まで合計したものを**累積相対度数**という。

次のデータ A は、10 人の生徒に関する英語のテストの得点の記録である。

データ A : 61, 84, 75, 71, 47, 77, 69, 96, 61, 79

階級幅 10 で度数分布表を作成すると、次のようになる。

階級(点)	階級値(点)	度数(人)	累積度数(人)	相対度数	累積相対度数
40 以上 50 未満	45	1	1	0.1	0.1
50 以上 60 未満	55	0	1	0.0	0.1
60 以上 70 未満	65	3	4	0.3	0.4
70 以上 80 未満	75	4	8	0.4	0.8
80 以上 90 未満	85	1	9	0.1	0.9
90 以上 100 未満	95	1	10	0.1	1.0
合計		10		1.0	

(注) 度数分布表は、その作成意図により、階級値、累積度数、相対度数、累積相対度数の欄のいくつか(あるいは全部)を省略することもある。

例題 2・1

度数分布表・ヒストグラム

次の20個の値からなるデータについて考える。

7, 8, 16, 10, 9, 9, 17, 19, 7, 4,
19, 16, 14, 10, 9, 15, 5, 18, 16, 3

- (1) 与えられたデータをもとに次の度数分布表を作成せよ。

階級	度数	累積度数	相対度数
0以上 4未満			
4以上 8未満			
8以上 12未満			
12以上 16未満			
16以上 20未満			
合計			

- (2) (1)の度数分布表をもとにヒストグラムを作成せよ。
 (3) (1)の度数分布表をもとにしたときのデータの最頻値を求めよ。

解答

- (1) 20個の値を、値の小さい方から順に並べる。

3, 4, 5, 7, 7, 8, 9, 9, 9, 10, 10,
14, 15, 16, 16, 16, 17, 18, 19, 19

度数分布表は次のようになる。

※階級ごとに区切りを入れた。

問題 2・3

- (1) 最小値, 第1四分位数, 第2四分位数(中央値), 第3四分位数, 最大値をそれぞれ m , Q_1 , Q_2 , Q_3 , M とすると, 箱ひげ図は次のようになる.



A, B, C それぞれに対応する箱ひげ図は,

A クラス…④, B クラス…③, C クラス…⑧.

- (2) A, B, C それぞれについて, 四分位範囲を求めると,

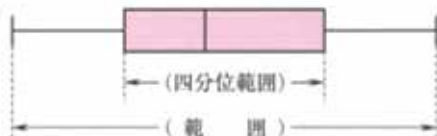
$$\text{A クラス} : 70 - 53 = 17,$$

$$\text{B クラス} : 82 - 53 = 29,$$

$$\text{C クラス} : 82 - 50 = 32.$$

よって, 四分位範囲で得点の分布の散らばり度合いを比較したとき, 散らばり度合いが最も小さいのは A クラスである.

(参考)



(1) の ③, ④, ⑧ の箱ひげ図を比較すると, 四分位範囲(箱の部分の長さ)が最も短いのは ④ である. これより, ④ に対応するクラス, すなわち, A クラスが散らばり度合いが最も小さいとわかる.

- (3) A, B, C それぞれについて, $Q_1 - 1.5 \times L$, $Q_3 + 1.5 \times L$ の値を求めると

$$\ast \ast (\text{四分位範囲}) = Q_3 - Q_1$$

$$\ast \ast 17 < 29 < 32$$

$\ast \ast L$ は(2)で求めた四分位範囲の値.