教養科目／B 自然の摂理の探求

2026 年度 統計学（集中１期）

第１回

- 記述統計

  統計データを記述する

- 推測統計

  統計データから推測する

1. 1. データ

例 1.1

A 君はある花の種を蒔いてから発芽するまでの日数がどの程度かを知りたくて実際にこの花の種 10 粒を蒔いて発芽するまでの日数を計ってみました.
その結果は

$$8,\quad 5,\quad 5,\quad 9,\quad 5,\quad 8,\quad 9,\quad 9,\quad 7,\quad 9$$

でした.

1．1．データ

例 1.2

あるラーメン店は客がどう感じているかを知るためにアンケートをとってみました．たとえば，ある項目は質問内容が「味はどうですか？」であり，それに対する回答は「うまい」,「普通」,「まずい」から選ぶ方式でした．この項目についての 20 人のアンケート結果は

普通，普通，うまい，まずい，普通，普通，まずい，普通，まずい，まずい，うまい，普通，普通，普通，うまい，普通，うまい，うまい，普通，まずい

でした．

1．1．データ

例 1.3

あるクラスの学生の体重 (kg) を測ってみました．その結果は

$$62.6,\quad 73.1,\quad 56.3,\quad 47.8,\quad 69.9,\quad 50.0,\quad 65.0,\quad 74.3,\quad 77.5,\quad 62.1,$$
$$46.6,\quad 70.6,\quad 67.8,\quad 63.1,\quad 52.4,\quad 55.2,\quad 57.5,\quad 64.1,\quad 63.3,\quad 70.3,$$
$$71.3,\quad 66.2,\quad 60.8,\quad 63.1,\quad 60.3,\quad 57.7,\quad 71.9,\quad 79.6,\quad 55.5$$

でした．

1．2．データの種類

「1」,「2」,「3」のような何らかの量として得られるデータ

⇩

量的データといいます.

「うまい」,「普通」,「まずい」のように量を表さないデータ

⇩

質的データといいます.

1. 2. データの種類

例 1.1〜例 1.3 では …

量的データは, 例 1.1 （「日数」）と例 1.3 (「重さ」)
質的データは, 例 1.2 (「ラーメンの味」)

例 1.1 での「日数」，例 1.3 での「重さ」は量的データですが, さらに，これらにも違いがあることがわかります.

・「日数」は「とびとびの値」だけをとる.

⇩

離散型データ

・「重さ」は「とびとびの値」以外の値もとる.

⇩

連続型データ

１．３ 度数分布表とヒストグラム

例 1.1

A 君はある花の種を蒔いてから発芽するまでの日数がどの程度かを知りたくて実際にこの花の種 10 粒を蒔いて発芽するまでの日数を計ってみました.
その結果は

$$8,\quad 5,\quad 5,\quad 9,\quad 5,\quad 8,\quad 9,\quad 9,\quad 7,\quad 9$$

でした.

・目的:

「種を蒔いてから発芽するまでの日数がどの程度か」を知りたい.

離散型データと連続型データに分けて考える.

１．３．１ 離散型データの場合

・元のデータ：8,　5,　5,　9,　5,　8,　9,　9,　7,　9

・度数分布表

表 1.1: 例 1.1 の度数分布表

階級値 (日数)	度数 (種の個数)
5	3
6	0
7	1
8	2
9	4
計	10

階級値

度数

度数の合計はデータの個数

１．３．１ 離散型データの場合

・度数分布表

表 1.1: 例 1.1 の度数分布表

階級値 (日数)	度数 (種の個数)
5	3
6	0
7	1
8	2
9	4
計	10

・ヒストグラム

右の図を, 離散型データの **ヒストグラム** という.

視覚化！

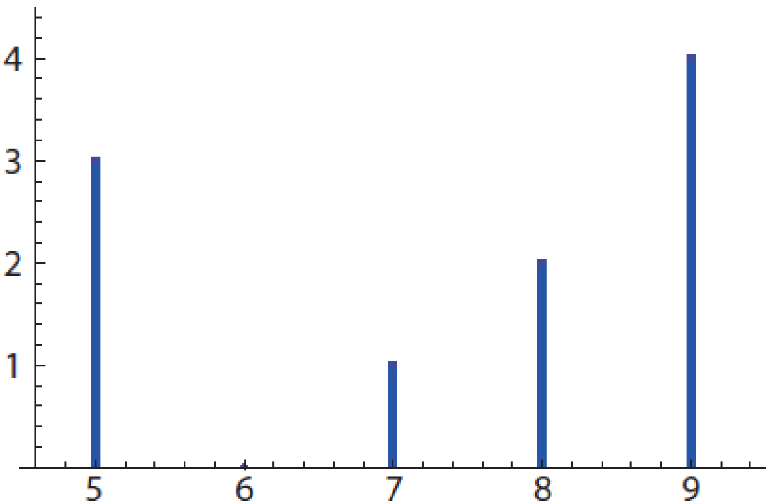


図 1.1: 例 1.1 のヒストグラム

1．3．1 離散型データの場合

・ヒストグラムを作成することで，データがもつ特徴をみつけやすくなる!

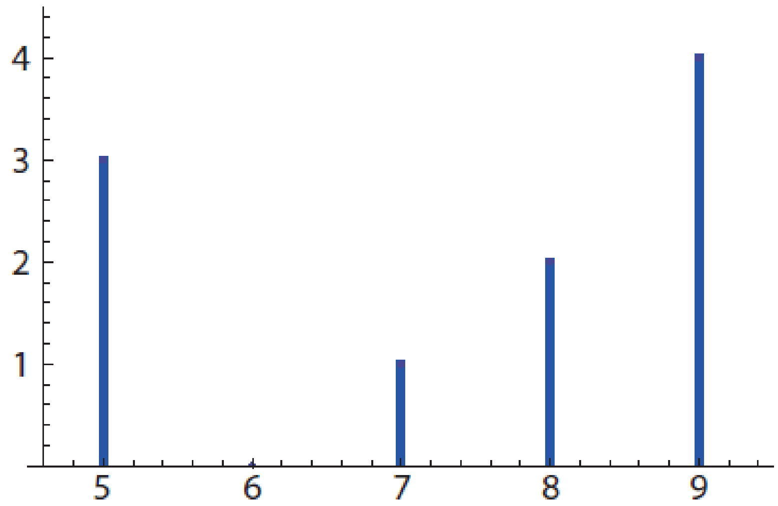


図 1.1: 例 1.1 のヒストグラム

・ヒストグラムより得られる知見

・最短で 5 日間，最長で 9 日間で発芽した．
・比較的両端 (5 日間や 9 日間) で発芽した種が多い．
・6 日間，7 日間で発芽した種は少ない．

１．３．２ 連続型データの場合

例 1.3

あるクラスの学生の体重 (kg) を測ってみました．その結果は

62.6,	73.1,	56.3,	47.8,	69.9,	50.0,	65.0,	74.3,	77.5,	62.1,
46.6,	70.6,	67.8,	63.1,	52.4,	55.2,	57.5,	64.1,	63.3,	70.3,
71.3,	66.2,	60.8,	63.1,	60.3,	57.7,	71.9,	79.6,	55.5	

でした．

とりあえず, 離散型データと同じように扱ってみると ···

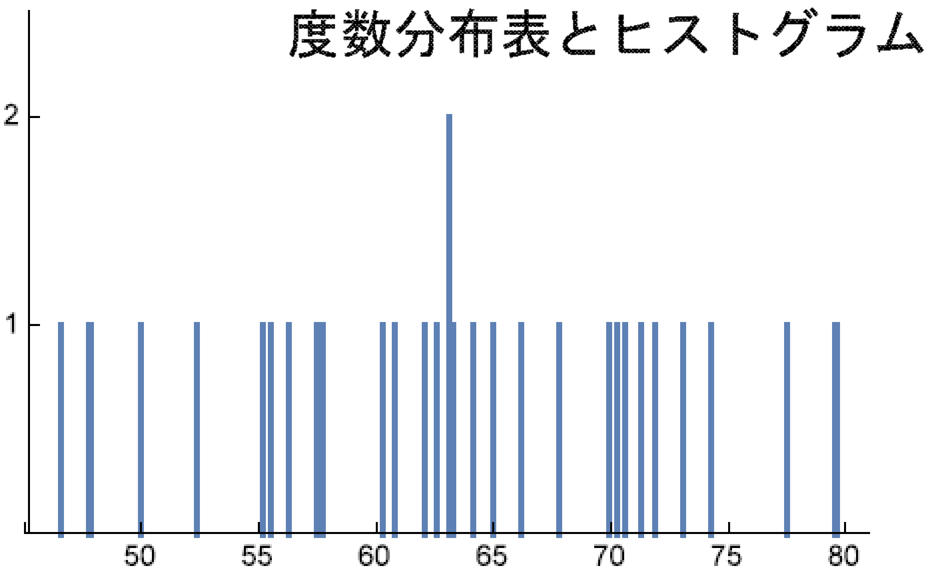
１．３．２ 連続型データの場合

・例 1.3 のデータを離散型とみなして作成した度数分布表とヒストグラム

表 1.2

階級値	度数
46.6	1
47.8	1
50.0	1
⋮	⋮
63.1	2
⋮	⋮
79.6	1
計	29

図 1.2



・29 個のデータはほとんど違う値をとっている.
・データがもっている特徴を表しているとはいえない.

これらは, 連続型データがとびとびの値以外の値もとることに起因する.

1．3．2 連続型データの場合

・連続型の場合の度数分布表は，次の手順で作成する．

1. 適度な幅をもついくつかの区間，たとえば，

$$45 \sim 51, \quad 51 \sim 57, \quad 57 \sim 63, \quad 63 \sim 69, \quad 69 \sim 75, \quad 75 \sim 81$$

のような区間を考える．

・このような区間を**階級**という．

2. 各階級に属するデータの個数を数える．これを**度数**という．

表 1.3

階級	度数
45 ∼ 51	3
51 ∼ 57	4
57 ∼ 63	6
63 ∼ 69	7
69 ∼ 75	7
75 ∼ 81	2
計	29

・**注意** 51 は 45 ∼ 51 と 51 ∼ 57 のどちらに含めるべきか?

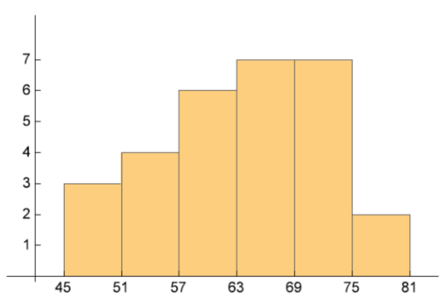


図 1.3: 例 1.3 のヒストグラム

・境界の値を大きい方の階級に含める:

45 ∼ 51 を 45 以上 51 未満，

51 ∼ 57 を 51 以上 57 未満，

⋮

75 ∼ 81 を 75 以上 81 未満．

１．３．２ 連続型データの場合

・度数分布表に**級中央値**と**相対度数**といわれるものを追加した方が便利なことがある.

・**級中央値**:

$$\text{級中央値} = \frac{\text{階級の下側の値} + \text{階級の上側の値}}{2}.$$

(例) $45 \sim 51$ の級中央値は, $\frac{45+51}{2} = 48.$

・**相対度数**:

$$\text{相対度数} = \frac{\text{階級の度数}}{\text{データの個数}}.$$

(例) $45 \sim 51$ の度数は 3 であり，データの個数が 29 なので,

この階級の相対度数は $\frac{3}{29} \fallingdotseq 0.10.$

１．３．２ 連続型データの場合

表 **1.4**　例 **1.3** のより詳しい度数分布表

階級	級中央値	度数	相対度数
45 ～ 51	48	3	0.10
51 ～ 57	54	4	0.14
57 ～ 63	60	6	0.21
63 ～ 69	66	7	0.24
69 ～ 75	72	7	0.24
75 ～ 81	78	2	0.07
計	—	29	1

１．３．２ 連続型データの場合

・階級の個数の決め方について考える．

図 1.4 (i) 階級の個数が 1 :

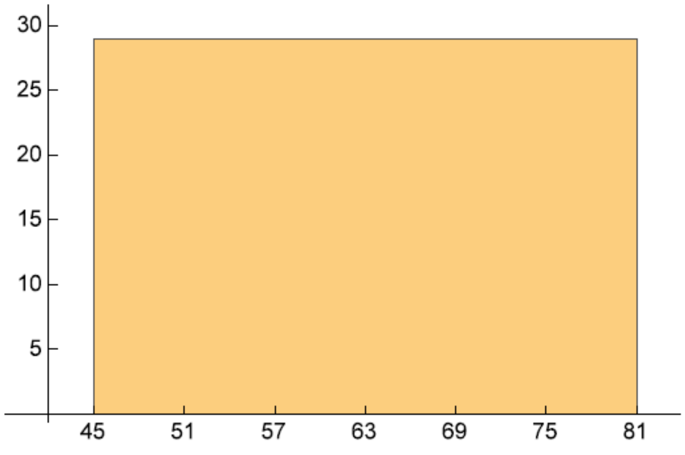
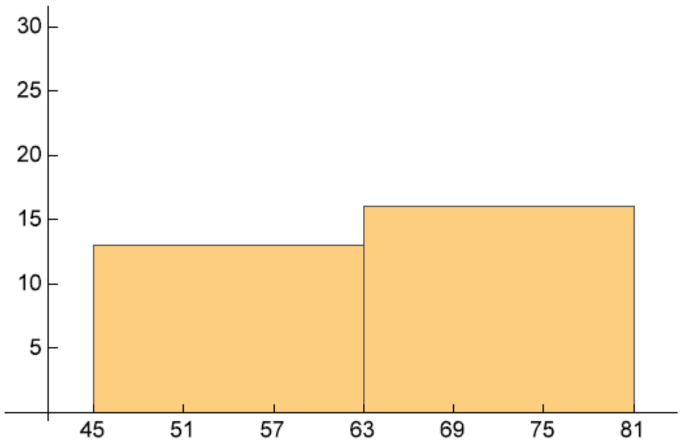
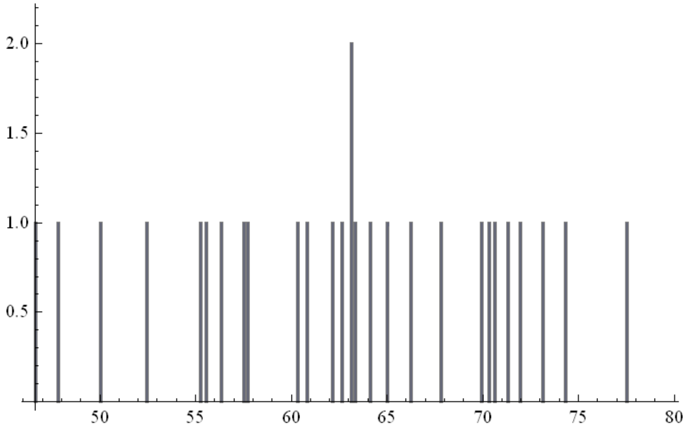


図 1.4 (ii) 階級の個数が 2:



階級の個数が 29:



- 階級の個数を多くするとデータの分類が細か過ぎる．
- 階級の個数を少なくするとデータの分類が粗過ぎる．
- いずれの場合もデータがもつ特徴を引き出すのは難しい．

1．3．2 連続型データの場合

・階級の個数の決め方について紹介する.

・表 1.5 [階級の個数の決め方 (スタージェスの方法)]

データの個数	12～22	23～45	46～90	91～181	182～362
階級の個数	5	6	7	8	9

・注意：この方法はあくまで１つの経験則であって，必ずこれを使わなければいけないというわけではない.

・例 (例 1.3 の場合)

データの個数は 29 なので，スタージェスの方法による階級の個数は 6 となる.

１．３．２ 連続型データの場合

連続型データの度数分布表の作成手順

ステップ 1 データの個数を求め，表 1.5 を参考に階級の個数を決定.

ステップ 2 データの中で 1 番小さい値と 1 番大きい値を求め，この 1 番大きい値と 1 番小さい値の差より大きい値で適度な値を決定．その際，次のステップ 3 で決定する階級の幅が自然数または区切りのよい小数となるようにする.

ステップ 3 ステップ 2 で決定した値を階級の個数で割ることにより階級の幅を決定．ここで 1 番小さい値が 1 番小さい階級に含まれ，1 番大きい値が 1 番大きい階級に含まれるように階級を決定.

ステップ 4 各階級の度数，つまり各階級に含まれるデータの個数を数える．さらに，級中央値，相対度数を求め，階級，度数と共に表にまとめる.

例1.5

表 1.6　10 万人あたりの結核罹患者数（2010 年，厚生労働省）

北海道	12	栃木	13	石川	16	滋賀	15	岡山	15	佐賀	21
青森	14	群馬	11	福井	14	京都	19	広島	16	長崎	23
岩手	12	埼玉	16	山梨	15	大阪	30	山口	16	熊本	17
宮城	11	千葉	17	長野	9	兵庫	21	徳島	18	大分	19
秋田	14	東京	23	岐阜	20	奈良	17	香川	15	宮崎	13
山形	11	神奈川	17	静岡	17	和歌山	21	愛媛	19	鹿児島	21
福島	12	新潟	12	愛知	23	鳥取	14	高知	18	沖縄	19
茨城	14	富山	13	三重	16	島根	18	福岡	19		

このデータは連続型データとみなして考える.

１．３．２ 連続型データの場合

ステップ 1 データの個数は 47 $\Rightarrow$ 表 1.5 より階級の個数は 7.

ステップ 2 データの中で 1 番小さい値は長野の 9 であり， 1 番大きな値は大阪の 30.

$30 - 9 = 21$ であり，階級の個数は 7 なので， $\frac{21}{7} = 3$ となりぴったりのように感じられるが, 21 より大きい値を考える必要がある. ここでは 24.5 を採用する.

ステップ 3 階級の幅は $\frac{24.5}{7} = 3.5$ となる．9 が 1 番小さい階級に含まれ, 30 が 1 番大きい階級に含まれるように階級を決める方法は無数にあるが, ここでは階級を

$$6.0 \sim 9.5, 9.5 \sim 13.0, 13.0 \sim 16.5, 16.5 \sim 20.0,$$

$$20.0 \sim 23.5, 23.5 \sim 27.0, 27.0 \sim 30.5$$

とする.

ステップ 4 階級 $6.0 \sim 9.5$ に入っているのは長野だけである．

階級 $9.5 \sim 13.0$ に入っているのは北海道，岩手，宮城，山形，福島，群馬，新潟の 7 個である．

$$\vdots$$

階級 $27.0 \sim 30.5$ に入っているのは大阪だけである．

級中央値は順番に

$$\frac{6.0+9.5}{2} = 7.75, \quad \frac{9.5+13.0}{2} = 11.25, \quad \cdots, \quad \frac{27.0+30.5}{2} = 28.75$$

となる．

データの個数は 47 なので相対度数は順番に

$$\frac{1}{47} \fallingdotseq 0.02, \quad \frac{7}{47} \fallingdotseq 0.15, \quad \cdots, \quad \frac{1}{47} \fallingdotseq 0.02$$

となる．

1．3．2　連続型データの場合

以上をまとめることによって，表 1.7 の度数分布表と図 1.5 のヒストグラムが得られる．

表 **1.7**　例 **1.5** の度数分布表

階級	級中央値	度数	相対度数
6.0 ∼ 9.5	7.75	1	0.02
9.5 ∼ 13.0	11.25	7	0.15
13.0 ∼ 16.5	14.75	17	0.36
16.5 ∼ 20.0	18.25	13	0.28
20.0 ∼ 23.5	21.75	8	0.17
23.5 ∼ 27.0	25.25	0	0.00
27.0 ∼ 30.5	28.75	1	0.02
計	—	47	1

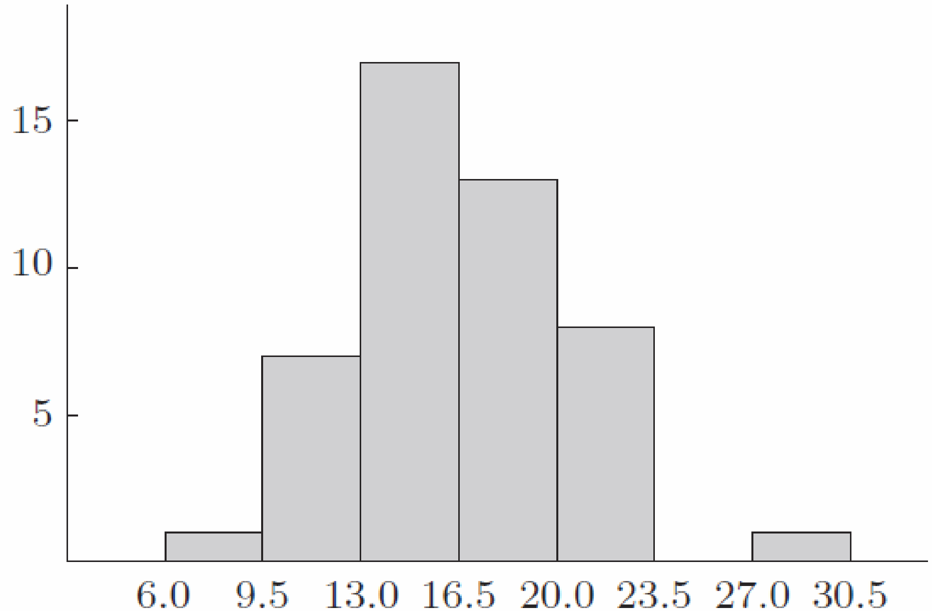


図 **1.5**　例 **1.5** のヒストグラム

まとめ

- 離散型

・データ: 8,　5,　5,　9,　5,　8,　9,　9,　7,　9

階級値 (日数)	度数 (種の個数)
5	3
6	0
7	1
8	2
9	4
計	10

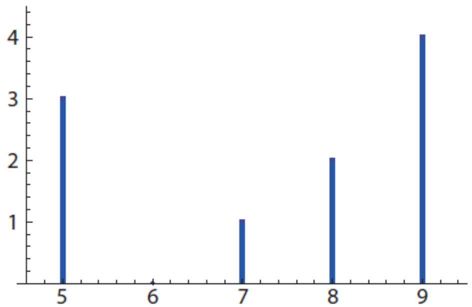


図 1.1: 例 1.1 のヒストグラム

- 連続型

・データ: 12,　14,　12,　11,　14,　11, $\cdots$, 19

表 1.7　例 1.5 の度数分布表

階級	級中央値	度数	相対度数
6.0 ～ 9.5	7.75	1	0.02
9.5 ～ 13.0	11.25	7	0.15
13.0 ～ 16.5	14.75	17	0.36
16.5 ～ 20.0	18.25	13	0.28
20.0 ～ 23.5	21.75	8	0.17
23.5 ～ 27.0	25.25	0	0.00
27.0 ～ 30.5	28.75	1	0.02
計	—	47	1

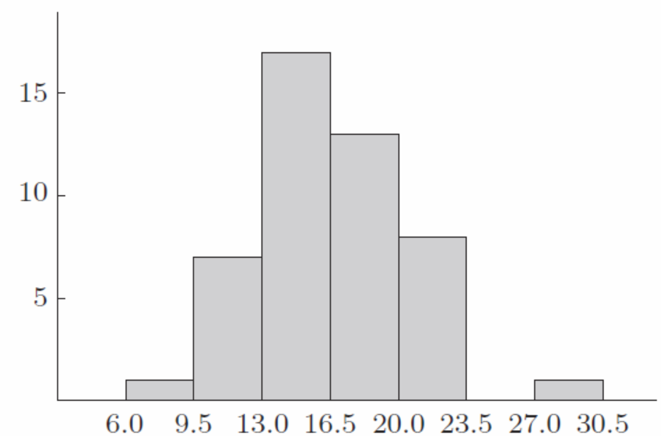


図 1.5　例 1.5 のヒストグラム

スタージェスの方法により，階級の個数，階級の幅を決める．