資料の散らばりと代表値

2015 年 2 月 12 日

目次

このテキストの使いかた

日頃の学習では・・・

- テキストをていねいに読んでいきましょう。

  このテキストは、きちんと言葉を使っててていねいな説明が書かれています。記号や数式が並んでいるだけの、意味不明のものではありません。ひとことひとこと言葉を大切にして、解き方ではなく考え方を学び取るようにしてください。そして、書いてあることに対して、「あーそういうことか」とか「えーよくわからない」とか「これ、ちがうんじゃないの？」といった反応をしてください。数学は自分の頭を使って考えていく科目ですから立ち止まって考えることがとても大切なのです。

- テスト直前に勉強を始めるのではなく、テストで力を発揮できるように前もって準備をしておきましょう。

  数学のように、自分の頭を使って「あーでもない、こーでもない」と考えながら学んでいく科目では、学習を始めてからしばらくの間はなかなか成果が出ない事があります。しかし少し我慢をして学習を続けていくうちに、あるとき、驚くような力が付いていることに気づくことがあります。 つまり、実力は初めのうちはゆっくり伸びていき、あとからぐんと伸びることが多いのです。

- 例題を学ぶときには、解答を読む前に、できればまず自分の力で解くことができるかどうか試してみましょう。

  数学の学習では、誰かから教わっただけのことよりも自分で悩んで考えたことの

ほうがよく身につきます。紙と筆記用具を使って実際に答案を書いてみてください。それができたらテキストの解答をていねいに読んで自分の考えと比べてみましょう。

- 例題の学習ができたら、この問題がテストに出ても自分の力だけで解けるかどうか想像してみましょう。そして心配なものには印を付けておきましょう。

  日頃から自分の実力をつかんでおくとテスト対策がしやすくなります。

- 問はもちろん、まず自分の力で解くことができるかどうか試してみてください。

  紙と筆記用具を使って実際に答案を書いてみてください。それができたらテキストの解答をていねいに読んで自分の考えと比べてみましょう。そうすれば理解が深まるでしょう。

- 問を解き終わって答え合わせをしたら、間違ったものには印を付けておきましょう。

  日頃から自分の実力をつかんでおくとテスト対策がしやすくなります。

- ひとつのひとつの節を読み終わったら、どんなことをその節で学んだのか思い出して「あらすじ」を言えるようにしておきましょう。紙と筆記用具を用意して、誰かにあらすじを伝えるにはどんなふうに説明すればよいか考え、文章を書いてみるととても効果があります。

  中学生に「今日は学校の数学の授業でどんなことを勉強したの？」と聞いてみると、「えーと、何だっけ、そうだ、傾きとか習った。」と断片的なことを言えたりすることはあるのですが、改めて、「へぇ、ところで傾きってなんなの？」と聞いてみると「えー、何だっけ、そうだ、なんか計算したり直線を描いてた。」ぐらいの答えしか返ってこないことが多いのです。専門用語を正しく言えるようになることも必要なことかもしれませんが、そんなことより大切なのは「どんなお話を学んだのか」ということです。数学は意味の無い記号操作を学ぶ科目ではなく、ちゃんとしたストーリーがあるものを学んでいるのです。ですから「お話のあらすじ」を理

解しておくことが大切なのです。

定期テスト対策では・・・

「日頃の学習」のところにも書いてありますが、数学のような科目は力がつくまでに時間のかかる科目です。テストに備えて十分な日数を確保しておきましょう。そして、「日頃の学習」で心配な例題や間違った問にちゃんと印を付けているとテスト対策が楽になります。

- テキストから試験範囲の例題や問を探して、印のついていないものがちゃんと解けるかどうか試してみましょう。
- 印を付けた例題や問を繰り返し復習して、テストに出ても大丈夫な問題を少しでも増やしておきましょう。

第1章

データを集めると大切なことがわかるかもしれない

1.1　現代社会はデータ分析で動いています

今の私たちの社会では、様々なデータが集められ分析されています。例えば、コンビニエンスストアで売られているお弁当のことを考えてみてください。コンビニエンスストアを経営している人は、毎日、お弁当を作っている工場からお弁当を仕入れています。そして仕入れたお弁当をお客さんに売るわけです。しかし、悩まなくてはならないことがあります。毎日、「明日はどのお弁当がいくつ売れのだろううか？」ということを考えないといけないのです。もし、仕入れたお弁当が売れ残ってしまうと大変です。お弁当には消費期限というものがあり、消費期限を過ぎてしまったお弁当を売ることはできないのです。普通、そのようなお弁当は廃棄されることになります。お弁当を仕入れるためにお店の人はもちろんお金を払っているわけですが、売れ残ってしまうとその分お金を損してしまうのです。つまり、お店の人にとってはお金を捨てたのと同じです。（お金を損してしまうことも深刻な問題ですが、古くなったお弁当を捨ててしまうということは食べ物を無駄にすることにもなりますし、環境にもよくありませんね。）では、そんなことにならないよ

うにするために、いったいどうやって仕入れるお弁当の種類や数を決めているのでしょうか。

昔はかなり多くの会社経営者が「勘」と「経験」と「度胸」でいろいろなことを決めていたようです。しかし今では、「データ集め」をし、「集められたデータを分析」して、「一番良いと思われる答えを出す」ということが行われています。では、コンビニエンスストアではどのようなデータを集めているいるのでしょうか。コンビニエンスストアには「レジ」がありますよね。お客さんがお店の人にお金を支払うところに置いてある機械ですね。今ではこの機械には「POS システム」と呼ばれている仕掛けが入っています。お客さんが商品を買うときに、POS システムでは「売れた時間帯」、「お客さんの年齢層」、「天候」、「気温」、「曜日」などのデータを集めています。また「お店のある地域ではどんな行事がいつ行われるのか」ということや「お店の近くでいつどんな工事が行われるのか」といったデータも入力されています。そのようなデータを毎日毎日たくさん集めておき、データを数学の力を使って分析すると、明日どのお弁当がどれぐらいそのお店で売れるのか予想ができるのです。もちろんその予想が完璧に当たるというわけでもありません。しかし、「勘」と「経験」と「度胸」に頼るよりはるかに良い予想がたてられるのです。

これまで説明してきたコンビニエンスストアの POS システムの例でもわかるように、現代社会ではデータを集めて分析することがとても重要になっています。そしてコンビニエンスストアの POS システムに限らず、例えば国や自治体、民間企業などは様々な場面でデータを集め、分析をし意思決定をしています。「どのようなデータを集めれば良いのか」ということや「どのような方法で分析すれば良いのか」ということは目的によって違うので、分析をする人が一番良い方法を決めなくてはなりません。ですからそこがデータ分析をする人の腕の見せ所というわけです。一方今では、コンピュータの発達により大量のデータを速く処理することができるようになってきました。そして現代では、「データ集めの方法」や「分析の方法」はどんどん複雑で高度になってきています。そのような複雑で高度な話をここでするわけにはいきませんが、最も基本となる事柄をこれから私たち

は学ぼうと思います。

1.2　散らばりがあるデータの扱い方

1.2.1　度数分布表：散らばりのあるデータを表にまとめて散らばり具合を調べよう

私たちがこれから学ぶのは「散らばりのあるデータ」です。これからいくつかの例をお見せすることにしましょう。

例 1　睡眠時間の調査を例にして、調査の方法や度数分布表、階級、階級の幅、度数という言葉について学ぼう

ある小学校の 6 年生のあるクラスの生徒達の睡眠時間を調査をしようと思います。ではまずどんなことをすればよいでしょうか？

もちろん一人ひとりの生徒に睡眠時間を教えてもらわないといけないわけです。そんな時は「睡眠時間アンケート」をすれば良いですね。ですから例えば、右のようなアンケート用紙をみんなに配って睡眠時間を記入してもらい、回収すれば良いわけです。

> 睡眠時間アンケート
>
> あなたの睡眠時間を教えてください。
>
> 私の睡眠時間は毎日
>
> だいたい　　時間　　分ぐらいです。
>
> 記入した人は教室の後ろロッカーの上の
> 回収箱に提出してください。

では次の図を見てください。

データ番号	睡眠時間	データ番号	睡眠時間
No.1	8 時間 10 分	No.19	9 時間 40 分
No.2	9 時間 12 分	No.20	9 時間 00 分
No.3	10 時間 20 分	No.21	8 時間 15 分
No.4	9 時間 35 分	No.21	8 時間 50 分
No.5	7 時間 00 分	No.23	8 時間 40 分
No.6	7 時間 55 分	No.24	8 時間 15 分
No.7	6 時間 40 分	No.25	7 時間 45 分
No.8	9 時間 40 分	No.26	7 時間 25 分
No.9	8 時間 35 分	No.27	7 時間 30 分
No.10	9 時間 25 分	No.28	9 時間 20 分
No.11	8 時間 10 分	No.29	9 時間 17 分
No.12	10 時間 10 分	No.30	8 時間 10 分
No.13	6 時間 45 分	No.31	8 時間 20 分
No.14	8 時間 20 分	No.32	8 時間 20 分
No.15	9 時間 43 分	No.33	8 時間 50 分
No.16	8 時間 00 分	No.34	7 時間 50 分
No.17	9 時間 15 分		
No.18	8 時間 10 分		

これは回収したアンケートから睡眠時間を転記して作った表です。それぞれの生徒の睡眠時間が全部記録されているわけです。

この表を見るとアンケートに答えてくれた生徒の数は 34 人であることがわかりますね。

また、答えてもらった睡眠時間の数値を眺めていくと、結構人によって長かったり短かったりと幅があることがわかります。つまり、睡眠時間の数値は「散らばっている」のです。このような調査をした時に大事なのは「どのように散らばっているのか」ということを調べることです。「散らばり方を調べること」はデータ分析の第一歩なのです。

それではどのようにして散らばり方を調べるのかこれから説明していきます。

まず、集められたデータをいくつかの階級にわけます。

いま、「階級」という専門用語が出てきました。「階級」とは何のことなのかゆっくり説

明することにしましょう。

　調査結果からもわかるように睡眠時間は人それぞれです。そこで、細かい時間の違いを無視してデータをいくつかの組に分けるのです。例えば、睡眠時間が 6 時間以上 7 時間未満のデータを 1 つの組にまとめたり、7 時間以上 8 時間未満のデータを 1 つの組にまとめたり、8 時間以上 9 時間未満のデータを 1 つの組にまとめたり・・・としていくのです。このようにすると、例えば、データ番号 No.1 の 8 時間 10 分とデータ番号 No.9 の 8 時間 35 分は同じ組のデータということになりますね。また例えば、データ番号 No.2 の 9 時間 12 分とデータ番号 No.4 の 9 時間 35 分は同じ組のデータということになります。

　ではここであなたに質問です。

質問

(1)　睡眠時間が 6 時間以上 7 時間未満のデータはいくつありますか。

(2)　睡眠時間が 7 時間以上 8 時間未満のデータはいくつありますか。

(3)　睡眠時間が 8 時間以上 9 時間未満のデータはいくつありますか。

(4)　睡眠時間が 9 時間以上 10 時間未満のデータはいくつありますか。

(5)　睡眠時間が 10 時間以上 11 時間未満のデータはいくつありますか。

　では 5 分待ちます。さっきの表をよく見てしっかり数えてこの質問に答えてください。

・・・・・・・・・・・・・・・・・・・

・・・・・・・・

・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・

・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・

・・・・・・・・・・・・

はい 5 分たちました。表を見てしっかり数えるだけなので大丈夫だと思いますが、念のため質問の答えを教えることにしましょう。

質問の答え

(1) 睡眠時間が6時間以上7時間未満のデータは2個です。

(2) 睡眠時間が7時間以上8時間未満のデータは6個です。

(3) 睡眠時間が8時間以上9時間未満のデータは14個です。

(4) 睡眠時間が9時間以上10時間未満のデータは10個です。

(5) 睡眠時間が10時間以上11時間未満のデータは2個です。

どうでしたか？数え間違いはありませんでしたか？

睡眠時間 (時間)	人数 (人)
以上　未満	
6 〜 7	2
7 〜 8	6
8 〜 9	14
9 〜 10	10
10 〜 11	2
合計	34

それでは、今あなたに調べてもらったことをわかりやすく表にまとめてみることにしましょう。右の表を見てください。これは、「いくつかに分けられた組」と、「その組の中に入っているデータがいくつあるのか」ということをまとめた表です。この表のように、「いくつかに分けられた組」と、「その組の中に入っているデータがいくつあるのか」ということをまとめた表を数学では**度数分布表**と呼んでいます。

この表を見ても、もう細かいことはわかりません。例えば、睡眠時間が6時間以上7時間未満の生徒が2人いるということはわかりますが、その2人の生徒の正確な睡眠時間はもうわからないわけです。しかしこの表を見れば睡眠時間の散らばり具合がうっすらとわかります。例えば睡眠時間が8時間以上9時間未満の生徒が一番多いということがわかります。また、睡眠時間が「8時間以上9時間未満」を中心にして、睡眠時間の少ない方や多い方へ離れていくと生徒の数が減っていくこともわかりますね。

ここまでの説明でわかってもらえたと思いますが、「データの散らばり具合」を知るために「集めたデータをいくつかの組に分ける」ということをするわけです。数学では、い

くつかに分けられている組のそれぞれを**階級**と呼んでいます。ですからこの例では、「6時間以上7時間未満の階級」とか「7時間以上8時間未満の階級」などがあるわけです。また、階級には**幅**があります。例えば「6時間以上7時間未満の階級」の幅は1時間ですし、「6時間以上7時間未満の階級」の幅も1時間です。（この例ではどの階級の幅も1時間ですね。）そして、それぞれの階級に入っているデータの個数をその階級の**度数**と呼びます。例えば「6時間以上7時間未満の階級」にはデータが2個あるわけですから「6時間以上7時間未満の階級」の度数は2というわけです。

問 1.　次の表は、ある中学校の1年生50人について身長を調査した結果です。

データ番号	身長 (cm)	データ番号	身長 (cm)	データ番号	身長 (cm)
No.1	142.7	No.21	151.5	No.41	152.4
No.2	164.7	No.22	163.8	No.42	158.4
No.3	158.8	No.23	156.9	No.43	143.5
No.4	146.2	No.24	159.9	No.44	156.2
No.5	162.9	No.25	170.8	No.45	169.6
No.6	155.1	No.26	145.1	No.46	166.3
No.7	157.3	No.27	170.3	No.47	154.7
No.8	171.8	No.28	159.7	No.48	168.4
No.9	160.6	No.29	167.0	No.49	157.5
No.10	167.8	No.30	147.3	No.50	161.8
No.11	136.4	No.31	153.8		
No.12	161.3	No.32	163.1		
No.13	148.3	No.33	150.9		
No.14	169.1	No.34	138.5		
No.15	141.2	No.35	164.2		
No.16	157.8	No.36	159.3		
No.17	151.3	No.37	152.0		
No.18	167.5	No.38	171.5		
No.19	142.6	No.39	162.2		
No.20	154.0	No.40	146.9		

この調査結果について以下の問に答えなさい。

(1) 調査結果を整理して右の度数分布表を完成しなさい。

(2) (1) で完成した度数分布表では階級の幅はいくつですか。

(3) No.36 の生徒はどの階級に入りますか。

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	
145 ～ 150	
150 ～ 155	
155 ～ 160	
160 ～ 165	
165 ～ 170	
170 ～ 175	
合計	

答えを見る

問 2.　右の表はある中学校の 3 年生のあるクラスで睡眠時間の調査をして作った度数分布表です。以下の問に答えなさい。

(1) 階級の幅をいいなさい。

(2) このクラスの生徒である A さんの睡眠時間は 7 時間 45 分です。A さんはどの階級に入りますか。

(3) 度数が一番大きい階級をいいなさい。

(4) 6 時間以上 7 時間未満の階級の度数をいいなさい。

(5) 度数が一番大きい階級の度数をいいなさい。

睡眠時間 (時間)	人数 (人)
以上　未満	
5 ～ 6	3
6 ～ 7	10
7 ～ 8	12
8 ～ 9	5
9 ～ 10	1
合計	31

答えを見る

問 3.　あるバス会社では A 町から B 町へ行く路線バスを運行しています。この会社では次のダイヤ改正に向け、バスの運行状況を調査しています。調査担当の P さんは、これまで大体どのくらいの時間がかかっていたのか知りたいと思い、平日の午前 9 : 10 に A 町を出発する B 町行きのバスについて、とりあえず過去 40 回の運行状況の記録を調べてみました。そして、右のような度数分布表ができました。この度数分布表について以下の問に答えなさい。

かかった時間 (分)	回数 (回)
以上　未満	
30 ～ 35	1
35 ～ 40	8
40 ～ 45	15
45 ～ 50	13
50 ～ 55	3
合計	40

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表

(1) 階級の幅をいいなさい。
(2) 度数が一番大きい階級をいいなさい。
(3) 45 分以上 50 分未満の階級の度数をいいなさい。
(4) 度数が一番大きい階級の度数をいいなさい。

答えを見る

例題 1　右の表を見てください。これは、ある中学校の 2 年女子 36 名の 50 m 走の記録を度数分布表に整理したものです。以下の問に答えなさい。

タイム (秒)	人数 (人)
以上　未満	
7.5 ～ 8.0	3
8.0 ～ 8.5	8
8.5 ～ 9.0	8
9.0 ～ 9.5	5
9.5 ～ 10.0	1
10.0 ～ 10.5	4
10.5 ～ 11.0	5
11.0 ～ 11.5	2
合計	36

(1) 記録（つまりタイム）が 8.9 秒の生徒はどの階級に入りますか。
(2) 記録 (つまりタイム) が 9.0 秒の生徒はどの階級に入りますか。
(3) タイムが短い方 (つまり記録の良い方) から数えたとき、12 番目の生徒はどの階級に入っていますか。
(4) タイムが短い方（つまり記録の良い方）から数えたとき、19 番目の生徒はどの階級に入っていますか。

解答

「以上」とか「未満」という言葉の意味は大丈夫ですか？念のため少しだけおさらいしておきましましょう。

例えば「5 より大きい数」というのと「5 以上の数」というのは少し違います。5.0 は「5 より大きい数」の仲間ではありません。だって 5.0 は 5 より大きくないですよね。5.0 は 5 と同じなんですよね。しかし、5.0 は「5 以上の数」の仲間です。つまり、「5 以上の数」にはきっかり 5.0 も含まれているのです。また例えば、「5 より小さい数」というのと「5 以下の数」というのは少し違います。5.0 は「5 より小さい数」の仲間ではありません。だって 5.0 は 5 より小さくないですよね。5.0 は 5 と同じなんですよね。しかし、5.0 は「5 以下の数」の仲間です。つまり、「5 以下の数」にはきっかり 5.0 も含まれているのです。

ここまでの説明は大丈夫ですか？では今度は「未満」という言葉の使い方をおさらいします。

例えば「5 未満」と書いてあったらこれは「5 より小さい」と書いてあるのと同じことです。ですから 4.9 とか 4.99 とか 4.999 とか 4.9999 とか $\cdots$ は「5 未満」ですが 5.0 ちょうどは「5 未満」ではないのです。

では本題に入ることにしましょう。

(1) 記録が 8.9 秒の生徒は「8.5 秒以上 9.0 秒未満」の階級に入りますね。

(2) 記録が 9.0 秒の生徒は「9.0 秒以上 9.5 秒未満」の階級に入りますね。

いいですか、この解答の初めに説明したように、9.0 は「9.0 以上」ですが「9.0 未満」ではないですよね。ですから 記録が 9.0 秒の生徒は「8.5 秒以上 9.0 秒未満」の階級には入らないんですよ。

(3) 右の表を見てください。あなたのためにもう一度度数分布表を書いておきました。

タイム (秒)	人数 (人)
以上　未満	
7.5 ～ 8.0	3
8.0 ～ 8.5	8
8.5 ～ 9.0	8
9.0 ～ 9.5	5
9.5 ～ 10.0	1
10.0 ～ 10.5	4
10.5 ～ 11.0	5
11.0 ～ 11.5	2
合計	36

度数分布表ではそれぞれの生徒の記録 (つまりタイム) がどれだけだったのかはわかりませんよね。ですがこの度数分布表を見ると、例えば、「タイムが 7.5 秒以上 8.0 秒未満だった生徒が 3 人いた」ということはわかるわけです。そしてこの 3 人は、最も良い記録の階級に入っているわけです。さらに度数分布表を見てみると、次に良い階級（つまり「タイムが 8.0 秒以上 8.5 秒未満の階級」）の中には 8 人の生徒がいますよね。ということはここまでで合計、$3+8=11$ 人の生徒がいるわけです。ということは 12 番目に良い記録をだした生徒（つまり第 12 位の記録を出した生徒）はその次の階級に入っていることになりますね。つまり、12 番目の生徒は「8.5 秒以上 8.9 秒未満の階級」に入っていることになるわけです。

(4) 最も良い階級である「タイムが 7.5 秒以上 8.0 秒未満」の生徒は 3 人で、次に良い階級である「タイムが 8.0 秒以上 8.5 秒未満」の生徒は 8 人ですからここまでで合計、$3+8=11$ 人の生徒がいるのでしたね。ということは 19 番目に良い記録をだした生徒（つまり第 19 位の記録を出した生徒）はまだここまでの階級には出てきません。その次の階級、つまり「タイムが 8.5 秒以上 9.0 秒未満」には 8 人いますよね。ということはここまでで、合計、$11+8=19$ 人生徒が出てくるわけです。ということは、19 番目の生徒はこの階級、つまり「8.5 秒以上 9.0 秒未満の階級」に入っていることになりますね。

問 4.　右の表は、ある中学校の 1 年生 50 人について身長を調査して作った度数分布表です。以下の問に答えなさい。

(1) 身長の低い方から数えていくとき、身長が 21 番目の生徒はどの階級に入っていますか。

(2) 身長の高い方から数えていくとき、身長が 20 番目の生徒はどの階級に入っていますか。

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	6
145 ～ 150	5
150 ～ 155	8
155 ～ 160	11
160 ～ 165	9
165 ～ 170	7
170 ～ 175	4
合計	50

答えを見る

1.2.2 ヒストグラム：度数分布表からグラフを作って散らばりの様子をわかりやすくしてみよう

数学では、「数値で表されていることをさらにグラフにしてわかりやすくする」ということをよく行います。グラフを作ると「ぱっと見ただけでも全体の様子が見て取れるようになる」からです。ここではデータの散らばり具合をわかりやすくするためによく使わている「ヒストグラム（別名、柱状グラフ）」と呼ばれるグラフを作ります。

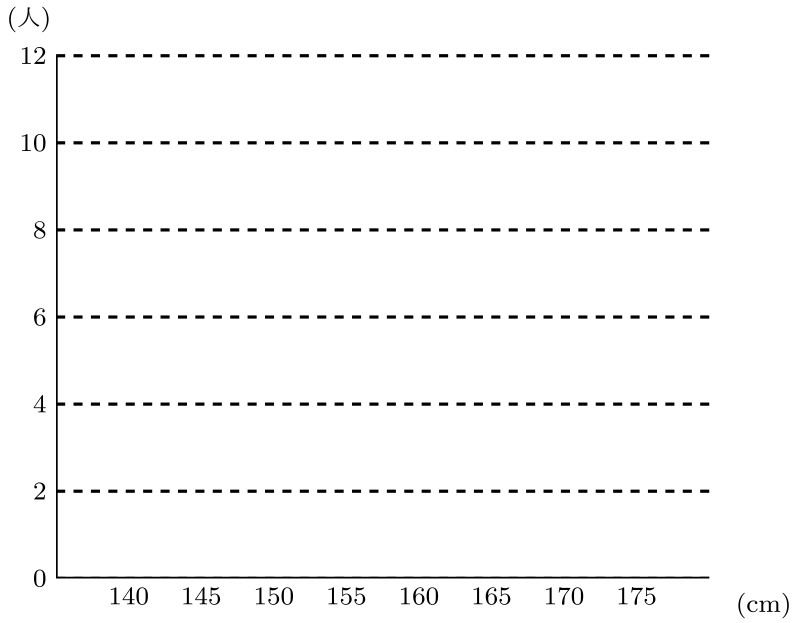
例 2 身長を調査して作られた度数分布表の例を使って**ヒストグラム**の作り方を学ぼう

右の表は、ある中学校の 1 年生 50 人について身長を調査して作った度数分布表です。これからこの度数分布表をもとに「ヒストグラム」と呼ばれているグラフを作ることにします。それでは作り方の手順を説明することにしましょう。

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	6
145 ～ 150	5
150 ～ 155	8
155 ～ 160	11
160 ～ 165	9
165 ～ 170	7
170 ～ 175	4
合計	50

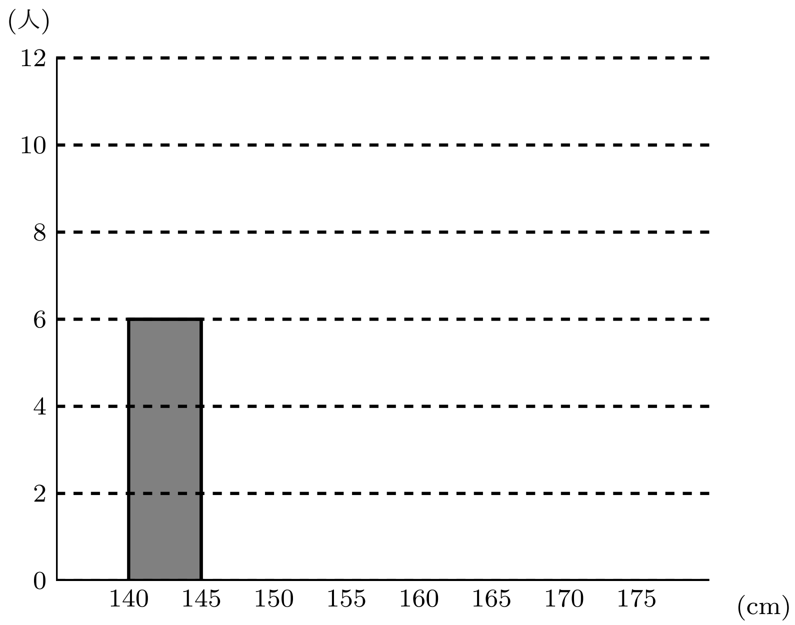
手順 1 まず、横軸と縦軸を作ります。横軸は調査している数値（この例の場合は身長）をあらわし、縦軸は度数（この例の場合は人数）をあらわすために使います。

では度数分布表を見てください。すると、調査している数値（つまり身長）をあらわすための横軸は 140 cm から 175 cm まで用意すればよいことがわかります。また度数（つまり人数）は最大でも 11 人ですから、度数をあらわす縦軸は 0 から 12 まで用意することにします。ですから例えば、次の図のように横軸と縦軸を描いておけばよいわけです。



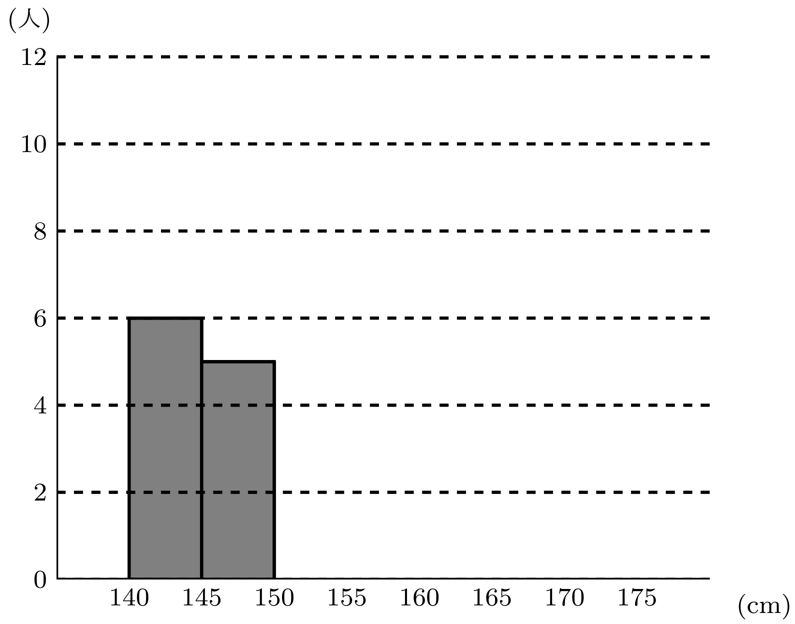
手順 2　度数分布表を見ながらそれぞれの階級の度数をあらわす「柱」を立てていきます。

ではまず、度数分布表で「140 cm 以上 145 cm 未満の階級」を見てください。この階級の度数は 6 ですよね。ですからまず次の図のように、高さが 6 の柱を立てればよいのです。

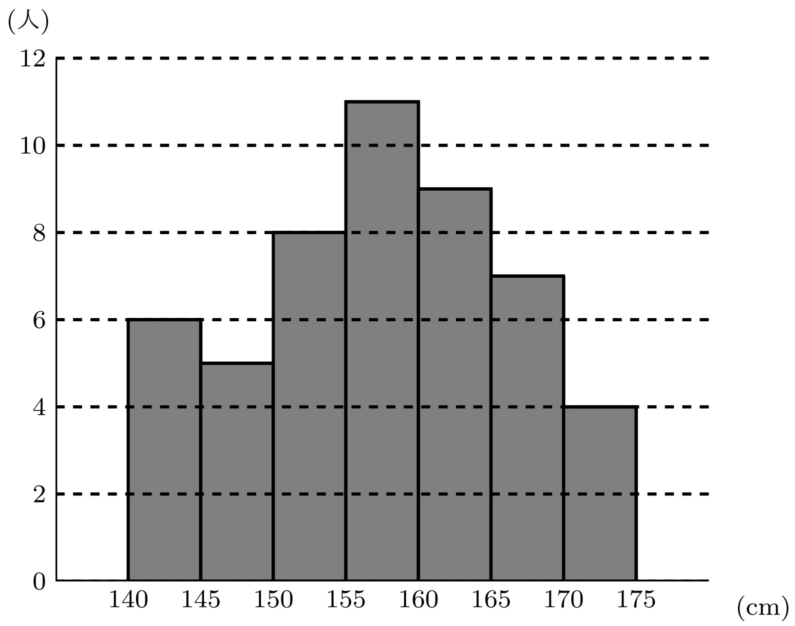


次は度数分布表で「145 cm 以上 150 cm 未満の階級」を見てください。この階級の度数は 5 ですよね。ですから今度は次の図のように高さが 5 の柱を立てればよい

のです。



以下同じようにして度数分布表を見ながら次々に柱を立てていきます。すると最後には次の図のようなものが完成します。

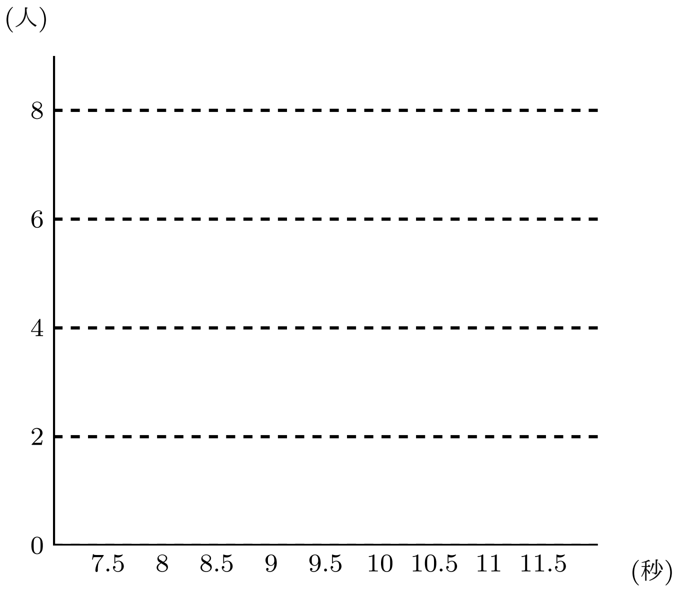


このようなグラフを作ると、「調査していたデータを分けてできたそれぞれの階級」にどれだけの「度数」あるのかということがぱっと見てとれるようになるわけです。このように「調査していたデータを分けてできたぞれぞれの階級」にどれだけの「度数」あるのか

ということを柱で表したグラフ（つまり度数分布表を柱でグラフにしたもの）を**ヒストグラム**（別名、柱状グラフ）と呼んでいます。

問 5.　次を見てください。

タイム (秒)	人数 (人)
以上　未満	
7.5 ～ 8.0	3
8.0 ～ 8.5	8
8.5 ～ 9.0	8
9.0 ～ 9.5	5
9.5 ～ 10.0	1
10.0 ～ 10.5	4
10.5 ～ 11.0	5
11.0 ～ 11.5	2
合計	36



左側の表は、ある中学校の 2 年女子 36 名の 50 m 走の記録を度数分布表に整理したものです。この度数分布表をよく見て、右にヒストグラムを完成しなさい。　答えを見る

1.2.3　度数折れ線：ヒストグラム以外にも、散らばりの様子をわかりやすくしたグラフを作ることができる

ヒストグラムに少し手を加えると「度数折れ線」と呼ばれるグラフを作ることができます。ヒストグラムはデータの散らばり具合を読み取るために役立ちましたが、これから学ぶ「度数折れ線」も散らばりの様子を読み取るために役立ちます。これから例を使って「度数折れ線」の作り方を学びます。

例 3　50 m 走のタイムを調査して作られた度数分布表の例を使って**度数折れ線**の作り方を学ぼう

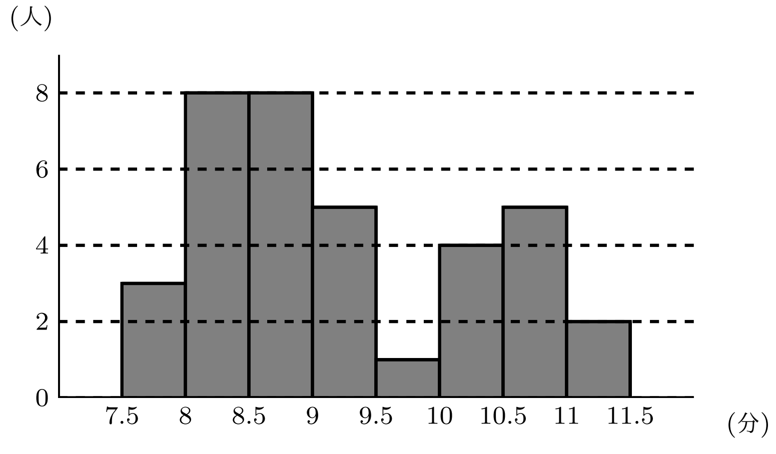
右の表は、ある中学校の 2 年女子 36 人について 50 m 走のタイムを調査して作った度数分布表です。これからこの度数分布表をもとに「 度数折れ線」と呼ばれているグラフを作ることにします。それでは作り方の手順を説明することにしましょう。

タイム (秒)	人数 (人)
以上　未満	
7.5 〜 8.0	3
8.0 〜 8.5	8
8.5 〜 9.0	8
9.0 〜 9.5	5
9.5 〜 10.0	1
10.0 〜 10.5	4
10.5 〜 11.0	5
11.0 〜 11.5	2
合計	36

手順 1　まず、度数分布表を使ってヒストグラムを作ります。

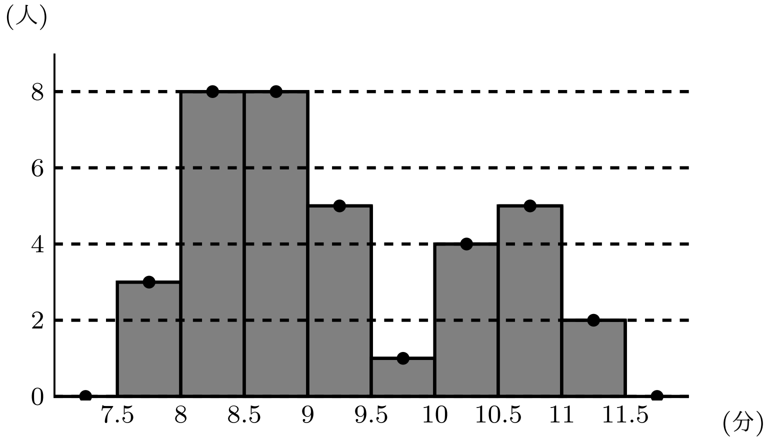
ヒストグラムの作り方は前に詳しく学んでいるのでここではもう教えません。忘れてしまった人は今すぐ 19 ページを開いて例 2 を復習してください。

この 50 m 走の度数分布表からヒストグラムを作ると次のようになりますよね。



手順 2　次は「それぞれの柱の一番上のところの真ん中」に点を打っておきます。ただし、「左端のなんにも柱のないところ」と「右端のなんにも柱のないところ」にも「高さが 0 の柱」があると思うことにして点を打ちます。

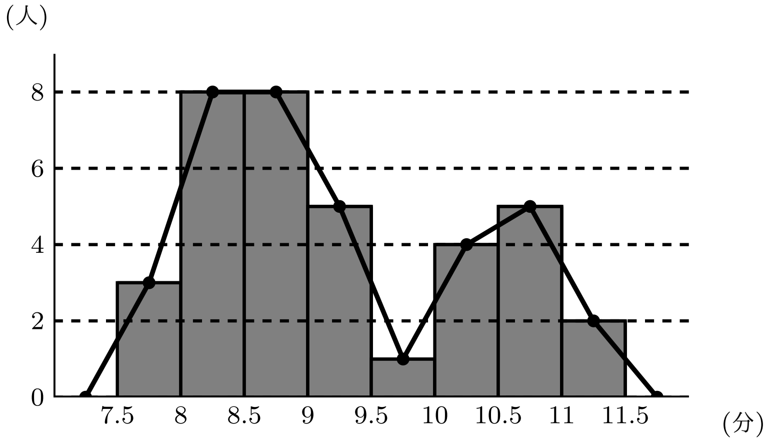
つまり、次の図のようになるわけです。



もう一度言っておきますが、「左端のなんにも柱のないところ」と「右端のなんにも柱のないところ」にも高さ 0 の点が打ってあることに注意してください。（どうしてこんなことをするのか不思議に思う人がいるかもしれませんが、気にすることはありません。深い理由は特にないのです。）

手順 3　最後に手順 2 で打った点を順番にまっすぐ結びます。

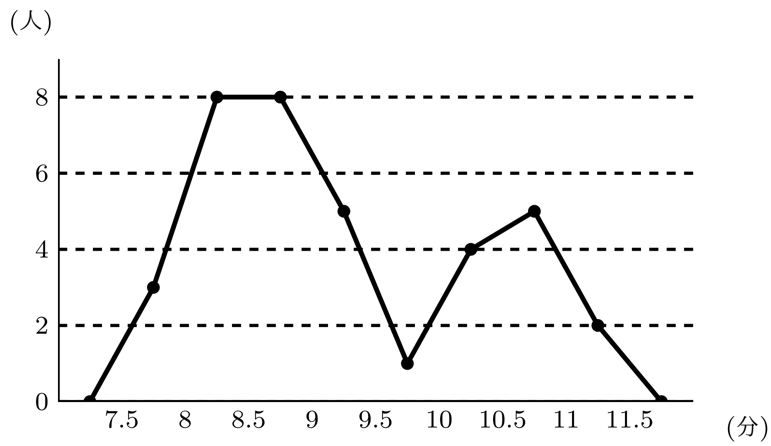
つまり、次の図のようになるわけです。



これで完成です。ヒストグラムから折れ線を描くことができましたね。

ヒストグラムの他にこのようなグラフを作っても、「調査していたデータを分けてできたぞれぞれの階級」にどれだけの「度数」あるのかということがぱっと見てとれるようになるわけです。このように「調査していたデータを分けてできたぞれぞれの階級」にどれだけの「度数」あるのかということを折れ線で表したグラフ（つまり度数分布表を折れ線でグラフにしたもの）を**度数折れ線**と呼んでいます。

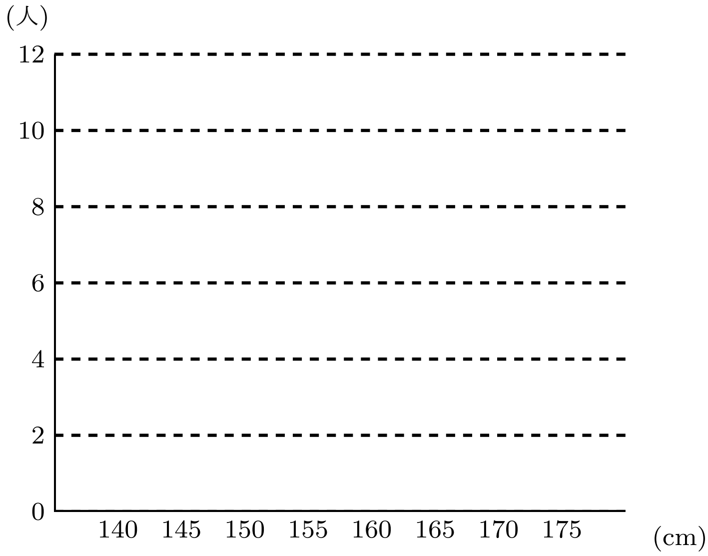
補足：念のための補足です。右のグラフを見てください。



これはさっき手順 3 で完成した図から、「柱」をすべて取り去って折れ線だけを残したものです。度数折れ線とはこの折れ線の事なのです。ですから度数折れ線を作るとき、柱を描かなくても大丈夫な人は、度数分布表からいきなり折れ線を描いても構いません。

問 6.　次を見てください。

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	6
145 ～ 150	5
150 ～ 155	8
155 ～ 160	11
160 ～ 165	9
165 ～ 170	7
170 ～ 175	4
合計	50



左側の表は、ある中学校の 1 年生 50 人について身長を調査して作った度数分布表です。この度数分布表をもとにして右側に度数折れ線を作りなさい。

答えを見る

1.2.4　相対度数：調べたデータの数が違う 2 つの調査を比べる時に重要になる量

前置きの話

あなたに質問です。

質問　A 中学校の生徒全員と B 中学校の生徒全員に一番好きなスポーツを答えてもらうアンケートをしてみたところ、サッカーが一番好きと答えた人が A 中学校では 87 人、B 中学校では 52 人でした。ですから、サッカーの好きな人の人数は A 中学校のほうが B 中学校よりかなり多かったわけです。ということは、A 中学校と B 中学校を比べたとき、学校の中でサッカーの人気が高いのは A 中学校であると判断しても良いのでしょうか。

では 10 分待ちます。考えてください。

・・・・・・・・・・・・・・・・・・・・・・・・・

. . . .

. . . .

. . . .

. . . .

. . . .

. .

. . . .

. . . .

はい 10 分たちました。自分の頭を使ってしっかり考えてくれた人は次を読んでください。

質問の答え　サッカーが一番好きと答えた人は A 中学校では 87 人、B 中学校では 52 人だったのですよね。つまり、サッカーの好きな人の人数は A 中学校のほうが多いわけです。でも、気楽に、「A 中学校のほうが B 中学校よりサッカーの人気が高い」なんて言えないですよね。どうして気楽にそんなことは言えないのかというと、それぞれの中学校で何人生徒がいるのかわからないからです。だって例えば、A 中学校の生徒数が 870 人だったら、A 中学校ではサッカーが一番好きな人は 10 人に 1 人の割合ということになりますね。そして例えば B 中学校の生徒数が 104 人だったら、B 中学校ではサッカーが一番好きな人は 2 人に 1 人の割合ということになります。（ちゃんと計算して確認してみてくださいね。）つまりこの場合、サッカーが一番好きな人は A 中学校では 10 人中たった 1 人だけですが、B 中学校では 2 人に 1 人もいるわけです。ですからもし、A 中学校の生徒数が 870 人で B 中学校の生徒数が 114 人だったら、サッカーの人気が高いのは B 中学校ということになりますね。

このように考えてみると、この質問では、「それぞれの中学校の生徒数がわからないので、サッカーが一番好きな人の割合もわからないということになり、どちらの中学校のほうがサッカーの人気が高いのかもわからない」ということになりますね。

この話からわかるのは、調べたデータの数が違う 2 つの調査を比べるときは「割合」で比べたほうが良いということです。

例題 2　A 中学校の生徒全員と B 中学校の生徒全員に一番好きなスポーツを答えてもらうアンケートをしてみたところ、サッカーが一番好きと答えた人が A 中学校では 87 人、B 中学校では 52 人でした。また A 中学校の生徒数は 580 人、B 中学校の生徒数は 270 人です。それでは A 中学校と B 中学校を比べたとき、サッカーの人気が高いのはどちらの中学校であると言えますか。

解答

こういうことを比べるときは「割合」を比べると良いのでしたね。つまり、この問題では、「サッカーが一番好きという生徒の割合」を A 中学校と B 中学校で比べてみるわけです。

A 中学校では 580 人のうち 87 人が「サッカーが一番好き」なのですから、

$$\text{A 中学校でサッカーが一番好きという生徒の割合} = \frac{87}{580} = 0.15$$

となります。（百分率で言うと、A 中学校ででサッカーが一番好きという生徒は 15% ということですね。）

B 中学校では 270 人のうち 52 人が「サッカーが一番好き」なのですから、

$$\text{B 中学校でサッカーが一番好きという生徒の割合} = \frac{52}{270} = 0.192\cdots$$

となります。（百分率で言うと、B 中学校でサッカーが一番好きという生徒は約 19% ということですね。）

ということは、「B 中学校のほうが A 中学校より少しサッカーの人気が高い」と言えますね。

問 7. A 中学校の生徒全員と B 中学校の生徒全員に一番好きなスポーツを答えてもらうアンケートをしてみたところ、サッカーが一番好きと答えた人が A 中学校では 87 人、B 中学校では 52 人でした。また A 中学校の生徒数は 510 人、B 中学校の生徒数は 320 人です。それでは A 中学校と B 中学校を比べたとき、サッカーの人気が高いのはどちらの中学校であると言えますか。 答えを見る

例題 3 ある町に住んでいる 281 人の人に一番好きなスポーツを答えてもらうアンケートをしてみたところ、次の表のような結果になりました。それぞれのスポーツについてあなたが「割合」を計算し、この表の「割合」の欄に数値を記入してください。ただし、この表では「割合」は少数第三位を四捨五入して少数第二位までの数値を書いてください。

一番好きなスポーツ	票数	割合
野球	52	
サッカー	42	
テニス	20	
自転車	14	
水泳・競泳	14	
バレーボール	14	
陸上	13	
ゴルフ	11	
バスケットボール	11	
フィギュアスケート	11	
ラグビー	11	
モータースポーツ（四輪）	7	
バドミントン	6	
モータースポーツ（二輪）	5	
その他	50	
合計	281	

この表の「その他」には、スキー、卓球、関心がない、アメリカンフットボール、柔道、スノーボード、水球、相撲、体操競技、フットサル、ボクシング、モータースポーツ（水上)、カヌー、剣道、セーリング、ソフトボール、ハンドボール、ボウリングなどが含まれています。

解答

例えば「野球が一番好きな人」は 281 人中の 52 人ですから、

$$野球が一番好きな人の割合 = \frac{52}{281} \cdots\cdots ①$$
$$= 0.1850\cdots\cdots ②$$

となりますね。ですから表の「野球」の欄の「割合」のところには少数第三位を四捨五入して 0.19 と書けばよいですね。というわけでまず右のようになるわけです。

他のスポーツについても同じようにして割合を計算し、表に記入していけば良いわけです。この先はあなたに任せることにします。次の問を解いてください。

一番好きなスポーツ	票数	割合
野球	52	**0.19**
サッカー	42	
テニス	20	
自転車	14	
水泳・競泳	14	
バレーボール	14	
陸上	13	
ゴルフ	11	
バスケットボール	11	
フィギュアスケート	11	
ラグビー	11	
モータースポーツ（四輪）	7	
バドミントン	6	
モータースポーツ（二輪）	5	
その他	50	
合計	281	

問 8.　例題 3 の解答の続きをやってもらう問題です。

ある町に住んでいる 281 人の人に一番好きなスポーツを答えてもらうアンケートをしてみたところ、右の表のような結果になりました。以下の問に答えなさい。

(1) 野球についてはもう「割合」が記入されていますが、他のそれぞれのスポーツについてはあなたが「割合」を計算し、この表の「割合」の欄に数値を記入してください。ただし、「割合」は少数第三位を四捨五入して少数第二位までの数値を書いてください。

(2) すべてのスポーツについて割合を合計するといくつになるはずですか。

一番好きなスポーツ	票数	割合
野球	52	0.19
サッカー	42	
テニス	20	
自転車	14	
水泳・競泳	14	
バレーボール	14	
陸上	13	
ゴルフ	11	
バスケットボール	11	
フィギュアスケート	11	
ラグビー	11	
モータースポーツ（四輪）	7	
バドミントン	6	
モータースポーツ（二輪）	5	
その他	50	
合計	281	

答えを見る

前置きの話おわり

調べたデータの数が違う 2 つの調査の度数分布表を比べる時も割合で比べよう

度数分布表がでてくる話では、「割合」という言葉の代わりに「相対度数」という言葉を使うことがあります。難しく感じるかもしれない言葉ですが、**相対度数**というのは「割合」のことだと思っていれば良いのです。

例題 4　あるバス会社では A 町から B 町へ行く路線バスを運行しています。この会社では次のダイヤ改正に向け、バスの運行状況を調査しています。以下の問に答えなさい。

(1) 調査担当の P さんは、これまで大体どのくらいの時間がかかっていたのか知りたいと思い、平日の午前 9 : 10 に A 町を出発する B 町行きのバスについて、とりあえず過去 40 回の運行状況の記録を調べてみました。そして、右のような度数分布表ができました。それぞれの階級についてあなたが相対度数を計算し、この度数分布表の相対度数の欄に記入しなさい。ただし、「相対度数」は少数第四位を四捨五入して少数第三位までの数値を書いてください。

かかった時間 (分)	回数 (回)	相対度数
以上　未満		
30 〜 35	2	
35 〜 40	9	
40 〜 45	13	
45 〜 50	12	
50 〜 55	4	
合計	40	

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表 (40 回分の調査)

(2) 調査担当の P さんは、たった 40 回分のデータでは真実に迫れないと考え、平日の午前 9 : 10 に A 町を出発する B 町行きのバスについて、過去 160 回の運行状況の記録を調べてみました。そして、右のような度数分布表ができました。それぞれの階級についてあなたが相対度数を計算し、この度数分布表の相対度数の欄に記入しなさい。ただし、「相対度数」は少数第四位を四捨五入して少数第三位までの数値を書いてください。

かかった時間 (分)	回数 (回)	相対度数
以上　未満		
30 〜 35	3	
35 〜 40	37	
40 〜 45	64	
45 〜 50	45	
50 〜 55	11	
合計	160	

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表 (160 回分の調査)

(3)　(1) で求めた相対度数の数値と (2) で求めた相対度数の数値を比べてみるとどんなことがわかりますか。

解答

まずはじめに確認しておきますが、相対度数というのは割合のことでしたね。

(1)　では (1) の度数分布表を見てください。

まず、「30 分以上 35 分未満の階級」相対度数を計算しましょう。「30 分以上 35 分未満」は 40 回中の 2 回ですから、

$$30\text{ 分以上 }35\text{ 分未満の階級の割合} = \frac{2}{40} = 0.05$$

となりますね。ですから表の「30 分以上 35 分未満の階級」の欄の「割合」のところには少数第四位を四捨五入して 0.050 と書けばよいことになります。

他の階級についてもも同じようにして割合を計算し、表に記入していけば良いわけです。そうすると右のようになりますよね。（電卓を使っても良いからちゃんと自分でも計算してくださいね。）

かかった時間 (分)	回数 (回)	相対度数
以上　未満		
30 ～ 35	2	0.050
35 ～ 40	9	0.225
40 ～ 45	13	0.325
45 ～ 50	12	0.300
50 ～ 55	4	0.100
合計	40	1

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表 (40 回分の調査)

(2) (1) と同じようにして計算していけばよいですよね。例えば「30 分以上 35 分未満」は 160 回中の 3 回ですから、

$$30\text{ 分以上 }35\text{ 分未満の階級の割合} = \frac{3}{160} = 0.01875$$

となりますね。ですから表の「30 分以上 35 分未満の階級」の欄の「割合」のところには少数第四位を四捨五入して 0.019 と書けばよいことになります。

他の階級についてもも同じようにして割合を計算し、表に記入していけば良いわけです。そうすると右のようになりますよね。（電卓を使っても良いからちゃんと自分でも計算してくださいね。）

かかった時間 (分)	回数 (回)	相対度数
以上　未満		
30 ～ 35	3	0.019
35 ～ 40	37	0.231
40 ～ 45	64	0.400
45 ～ 50	45	0.281
50 ～ 55	11	0.069
合計	160	1

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表 (160 回分の調査)

(3) (1) で求めた相対度数の数値と (2) で求めた相対度数の数値を比べてみるとどんなことがわかるのか考えるのでしたね。では次の表を見てください。

かかった時間 (分)	相対度数	
	40 回調査の場合	160 回調査の場合
以上　未満		
30 ～ 35	0.050	0.019
35 ～ 40	0.225	0.231
40 ～ 45	0.325	0.400
45 ～ 50	0.300	0.281
50 ～ 55	0.100	0.069
合計	1	1

あなたのために「40 回調査」と「160 回調査」の相対度数を 1 つの表にまとめてみました。

ひとつひとつの階級について、「調査の回数を 40 回から 160 回へ増やすと相対度数がどのように変わるか」この表で見ていくことにしましょう。

- 30 分以上 35 分未満では・・・

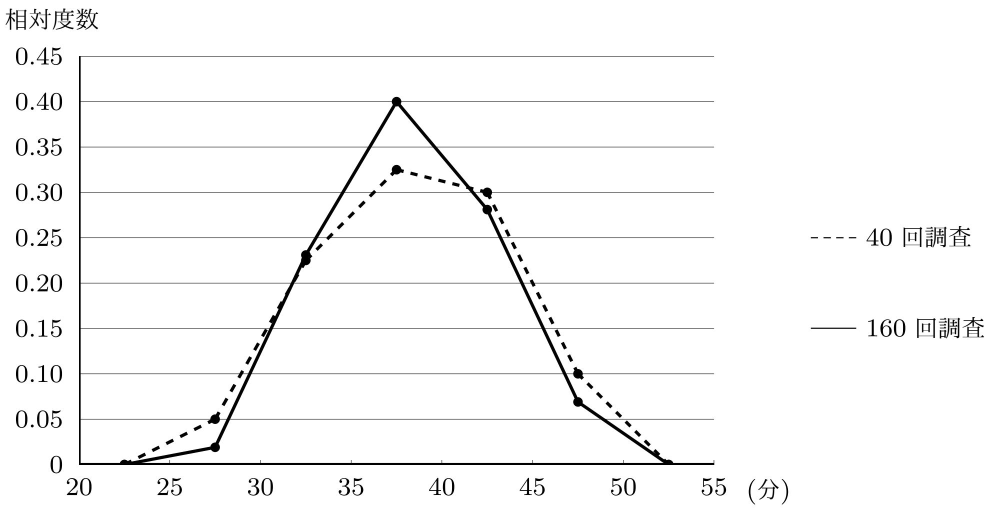
  「割合」がどのように変わったか見てみましょう。「40 回調査」では 0.050、「160 回調査」では 0.019 ですね。ですから調査回数を増やすと「割合」は減っています。

- 35 分以上 40 分未満では 0.225 から 0.231 へと「割合」は増えています。
- 40 分以上 45 分未満では 0.325 から 0.400 へと「割合」は増えています。
- 45 分以上 50 分未満では 0.300 から 0.281 へと「割合」は減っています。
- 50 分以上 55 分未満では 0.100 から 0.069 へと「割合」は減っています。

どうも、真ん中あたりの階級では相対度数は増えているようですが、端っこの階級では相対度数は減っているようです。もうすこし詳しく言うと、調べたバスの数を増やしたら、「かかった時間が 35 分から 40 分のバスの割合」や「かかった時間が 40 分から 45 分のバスの割合」は増えましたが、「早く到着するバスの割合」や「遅く到着するバスの割合」は減ったということがわかるわけです。

補足：この例題を通して、データの数の違う 2 つの調査を比べるときは「割合」で比べると良いということがわかったと思います。この例題では「表」に「相対度数（つまり割合)」を記入して 2 つのデータを比べましたが、グラフを作って比べるという方法もあります。そこでこれから、この例題のデータを使って作ったグラフをお見せすることにしましょう。

以前「度数折れ線」と呼ばれるグラフのことを学びましたね。（忘れてしまった人は 25 ページを開いて復習してください。）度数折れ線では縦軸はそれぞれの階級の度数を表していましたね。でも度数をそのまま使うとデータの数の違う 2 つの調査を比べるときは不便なわけです。この例題をしっかり学んだ人はもうおわかりだと思いますが、このようなときは「度数」の代わりに「相対度数（つまり割合）」を使えばよいですよね。ですからグラフの縦軸を「度数」ではなく「相対度数（つまり割合）」に変えてグラフを作ればよいわけです。次のグラフを見てください。



これは、縦軸を相対度数にして 40 回調査の結果と 160 回調査の結果を同時に折れ線グラフにしたものです。このように、2 つの調査を 1 つのグラフとしてまとめて描いておけば、2 つの調査でどのような違いが出ているのかわかりやすくなります。このグラフを見ると、40 回調査のグラフと 160 回調査のグラフはどちらも山のような形をしていますが、160 回調査のグラフは 40 回調査のグラフと比べて山の形がやや細くなり、山の頂上が高くなっていることがわかりますね。つまり、調査の回数を 40 回から 160 回に増やしてみたら、データの散らばりは少なくなり、データの数値は真ん中に集まってきているということがわかるのです。

問 9. ある中学校の 1 年 A 組と 1 年 B 組で男子の 50 m走のタイムがどのように違うのか比べるため、データを分析してみることにしました。以下の問に答えなさい。

(1) 右の表は 1 年 A 組の男子 20 人の 50 m 走のタイムをまとめたものです。それぞれの階級に対して相対度数を計算し、この表の相対度数の欄に記入しなさい。ただし、「相対度数」は少数第三位を四捨五入して少数第二位までの数値を書いてください。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 ～ 7.0	2	
7.0 ～ 7.5	3	
7.5 ～ 8.0	4	
8.0 ～ 8.5	2	
8.5 ～ 9.0	2	
9.0 ～ 9.5	1	
9.5 ～ 10.0	2	
10.0 ～ 10.5	3	
10.5 ～ 11.0	1	
11.0 ～ 11.5	0	
合計	20	

1 年 A 組の男子 20 人の 50 m 走のタイムの度数分布と相対度数分布

(2) 右の表は 1 年 B 組の男子 22 人の 50 m 走のタイムをまとめたものです。それぞれの階級に対して相対度数を計算し、この表の相対度数の欄に記入しなさい。ただし、「相対度数」は少数第三位を四捨五入して少数第二位までの数値を書いてください。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 ～ 7.0	1	
7.0 ～ 7.5	2	
7.5 ～ 8.0	4	
8.0 ～ 8.5	5	
8.5 ～ 9.0	3	
9.0 ～ 9.5	2	
9.5 ～ 10.0	2	
10.0 ～ 10.5	2	
10.5 ～ 11.0	1	
11.0 ～ 11.5	0	
合計	22	

1 年 B 組の男子 22 人の 50 m 走のタイムの度数分布と相対度数分布

(3) 右の表に、(1) で計算した A 組の相対度数と (2) で計算した B 組の相対度数を記入しなさい。

かかった時間 (分)	相対度数	
	A 組男子	B 組男子
以上　未満		
6.5 〜 7.0		
7.0 〜 7.5		
7.5 〜 8.0		
8.0 〜 8.5		
8.5 〜 9.0		
9.0 〜 9.5		
9.5 〜 10.0		
10.0 〜 10.5		
10.5 〜 11.0		
11.0 〜 11.5		
合計		

1 年 A 組の男子と 1 年 B 組の男子の 50 m 走のタイムの相対度数分布

(4) (3) で作った表を見てください。A 組男子と B 組男子では 50 m 走のタイムにはどのような違いがあると言えますか。

答えを見る

問 10. P さんはある中学校の 1 年 A 組の生徒です。P さんは「自分のクラスの男子の 50 m走のタイム」と「自分の中学校の 1 年男子全体の 50 m走のタイム」がどのように違うのか比べるため、データを分析してみることにしました。以下の問に答えなさい。

(1) 右の表は 1 年 A 組の男子 20 人の 50 m 走のタイムをまとめたものです。それぞれの階級に対して相対度数を計算し、この表の相対度数の欄に記入しなさい。ただし、この表では「相対度数」は少数第三位を四捨五入して少数第二位までの数値を書いてください。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 ～ 7.0	2	
7.0 ～ 7.5	3	
7.5 ～ 8.0	4	
8.0 ～ 8.5	2	
8.5 ～ 9.0	2	
9.0 ～ 9.5	1	
9.5 ～ 10.0	2	
10.0 ～ 10.5	3	
10.5 ～ 11.0	1	
11.0 ～ 11.5	0	
合計	20	

1 年 A 組の男子 20 人の 50 m 走のタイムの度数分布と相対度数分布

(2) 右の表はこの中学校の 1 年男子全体 90 人の 50 m 走のタイムをまとめたものです。それぞれの階級に対して相対度数を計算し、この表の相対度数の欄に記入しなさい。ただし、この表では「相対度数」は少数第三位を四捨五入して少数第二位までの数値を書いてください。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 ～ 7.0	1	
7.0 ～ 7.5	2	
7.5 ～ 8.0	12	
8.0 ～ 8.5	28	
8.5 ～ 9.0	19	
9.0 ～ 9.5	13	
9.5 ～ 10.0	10	
10.0 ～ 10.5	3	
10.5 ～ 11.0	1	
11.0 ～ 11.5	1	
合計	90	

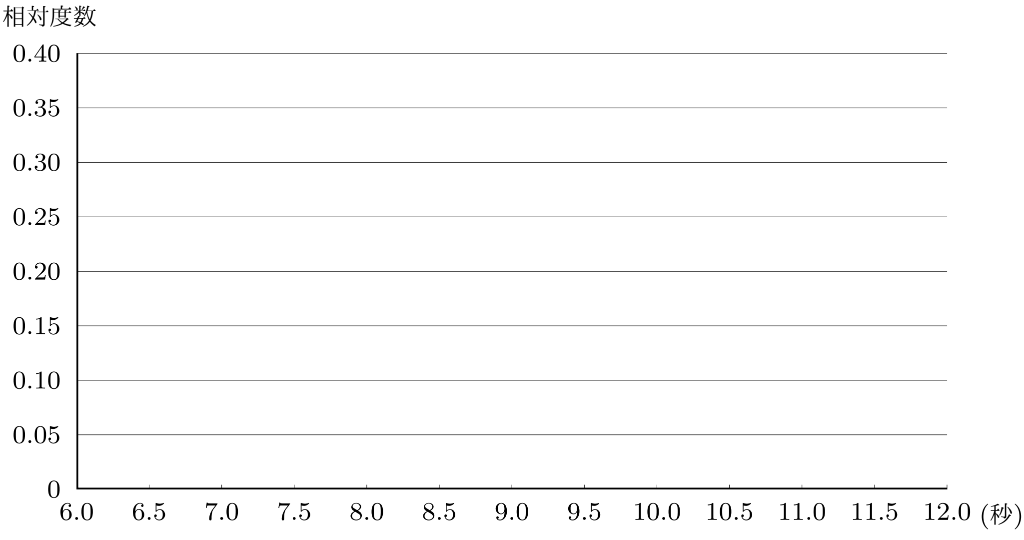
1 年男子全体 90 人の 50 m 走のタイムの度数分布と相対度数分布

(3) 右の表に、(1) で計算した A 組の相対度数と (2) で計算した全体の相対度数を記入しなさい。

かかった時間 (分)	相対度数	
	A 組男子	男子全体
以上　未満		
6.5 〜 7.0		
7.0 〜 7.5		
7.5 〜 8.0		
8.0 〜 8.5		
8.5 〜 9.0		
9.0 〜 9.5		
9.5 〜 10.0		
10.0 〜 10.5		
10.5 〜 11.0		
11.0 〜 11.5		
合計		

1 年 A 組の男子と 1 年男子全体の 50 m 走のタイムの相対度数分布

(4) (3) で作った表を見て、「1 年 A 組の男子の 50 m 走のタイムの相対度数分布折れ線」と「1 年男子全体の 50 m 走のタイムの相対度数分布折れ線」を次に完成しなさい。



1 年 A 組の男子の 50 m 走のタイムの相対度数分布折れ線」と「1 年男子全体の 50 m 走のタイムの相対度数分布折れ線」

(5) (4) で作った相対度数分布折れ線を見てください。A 組男子と男子全体では 50 m 走のタイムにはどのような違いがあると言えますか。

答えを見る

1.3　範囲と代表値：散らばりのあるデータの特徴をいくつかの数値であらわすには

1.3.1　データの特徴は分布の形にあらわれる

散らばりのあるデータを度数分布表にまとめることは、調査結果が持っている特徴をとらえるための第一歩となります。そしてさらに、度数分布がどのようになっているのかわかりやすくするためにはヒストグラムや度数分布折れ線を描くいて度数分布の「形」を良く見てみることが大切です。そうすると調べた集団が持っている特徴が見えてくるからです。度数分布の「形」は大抵の場合、ある値を中心にしてまとまっています。そして大抵の場合、

(1) データはどこからどこまでの範囲に散らばっているのか

(2) データの分布の中心はだいたいどこらへんにあるのか

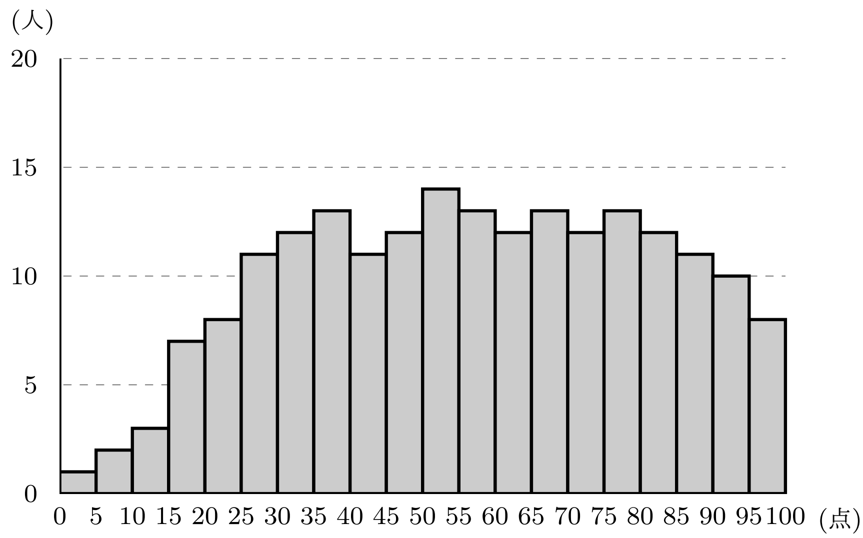
(3) データの散らばり方は中心に対して左右対称かそれともどちらかにかたよっているのか

などに注目すると、度数分布の「形」の特徴をわりと良くつかむことができます。例を使って説明することにしましょう。

例 4　次の図は、グラフはある中学校の中学 3 年生全員の英語のテストの結果をまとめた度数分布表とヒストグラムです。

点数	人数
以上 未満	
0 ～ 5	1
5 ～ 10	2
10 ～ 15	3
15 ～ 20	7
20 ～ 25	8
25 ～ 30	11
30 ～ 35	12
35 ～ 40	13
40 ～ 45	11
45 ～ 50	12
50 ～ 55	14
55 ～ 60	13
60 ～ 65	12
65 ～ 70	13
70 ～ 75	12
75 ～ 80	13
80 ～ 85	12
85 ～ 90	11
90 ～ 95	10
95 ～ 100	8
合計	198

注：この度数分布表では、点数が 100 点の人は最後の階級に含めてある



注：このヒストグラムでは、点数が 100 点の人は最後の階級に含めてある

この度数分布表やヒストグラムからどんなことが読み取れるか考えてみます。すると、

- テストの点数は 0 点ぐらいから 100 点ぐらいまで広く分布していることがわかります。
- データは 0 点ぐらいから 100 点ぐらいまで広く分布しているのでデータの分布の中心はだいたい 50 点ぐらいと考えることができます。
- テストの点がとても低かった人（つまり 0 点ぐらいから 14 点ぐらいの人）は少ないことがわかります。一方 25 点から 94 点ぐらいになると、どの階級にも似たような人数がいることがわかります。つまり、25 点ぐらいから 94 点ぐらいまでほぼ同

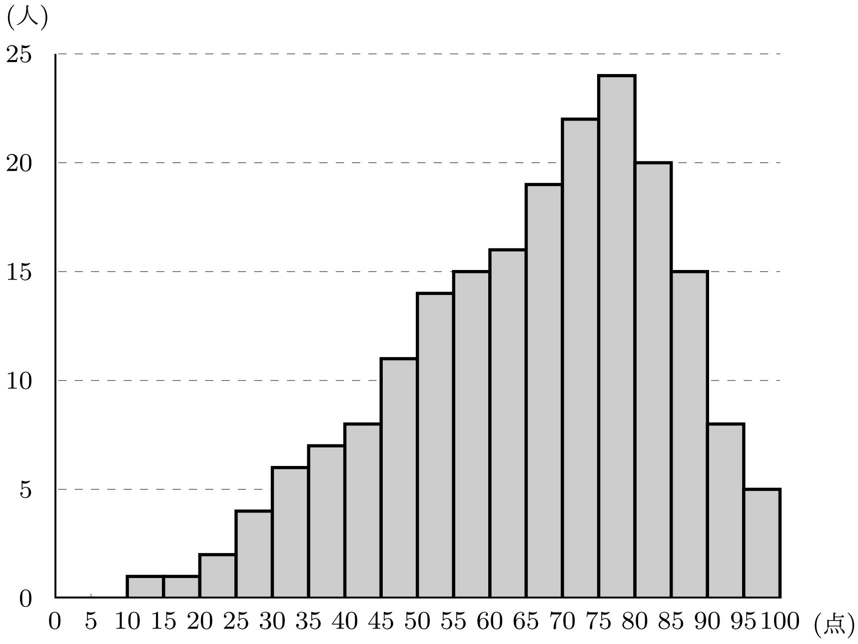
じように広い範囲に点数がひろがっていて、結構点数の低かった人、わりと真ん中辺りの点数だった人、結構良い点だった人の人数はだいたい似たようなものになっています。また、95 点から 100 点だった人も結構いることがわかります。

データの散らばり方はまあまあ左右対称と言って良いかもしれません。

例 5　次の図は、グラフはある中学校の中学 3 年生全員の国語のテストの結果をまとめた度数分布表とヒストグラムです。

点数	人数
以上　未満	
0 ～ 5	0
5 ～ 10	0
10 ～ 15	1
15 ～ 20	1
20 ～ 25	2
25 ～ 30	4
30 ～ 35	6
35 ～ 40	7
40 ～ 45	8
45 ～ 50	11
50 ～ 55	14
55 ～ 60	15
60 ～ 65	16
65 ～ 70	19
70 ～ 75	22
75 ～ 80	24
80 ～ 85	20
85 ～ 90	15
90 ～ 95	8
95 ～ 100	5
合計	198

注：この度数分布表では、点数が 100 点の人は最後の階級に含めてある



注：このヒストグラムでは、点数が 100 点の人は最後の階級に含めてある

この度数分布表やヒストグラムからどんなことが読み取れるか考えてみます。すると、

- テストの点数は 10 点ぐらいから 100 点ぐらいまで広く分布していることがわかります。

- データは 10 点ぐらいから 100 点ぐらいまで広く分布しているのでデータの分布の中心はだいたい 55 点ぐらいと考えることができます。

- 点数ががとても低かった人や結構低かった人（つまり 0 点ぐらいから 29 点ぐらいの人）は少ないことがわかります。また 75 点から 84 点ぐらいの人がとても多ことがわかります。さらに、点数がかなり高かった人（つまり 90 点ぐらいから 100 点の人）は少ないことがわかります。

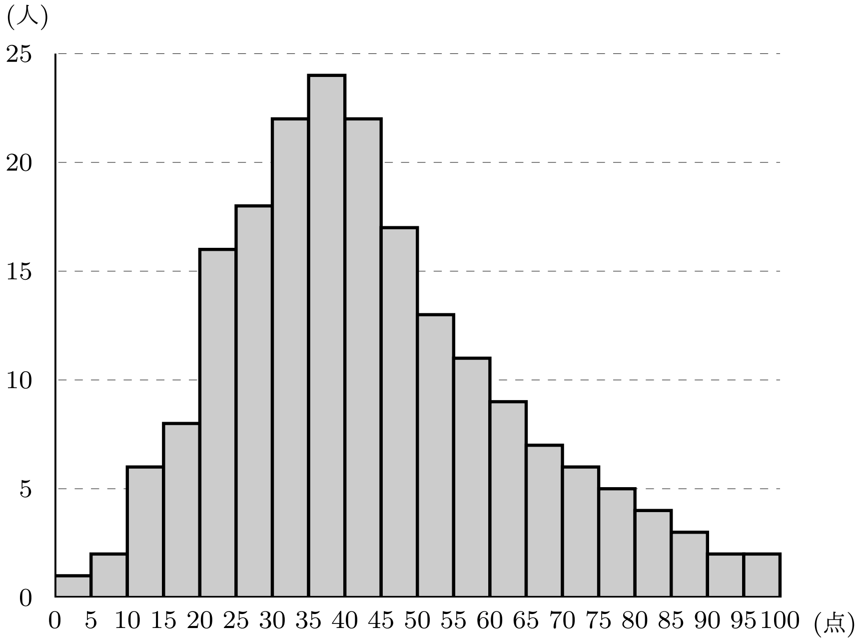
  データの散らばり方は、中心である 55 点ぐらいからかなり右にかたよっていると言えます。

ではもう少し、度数分布表やヒストグラムからどんなことを読み取ることができるのか例を見てみましょう。

例 6 次の図は、グラフはある中学校の中学 3 年生全員の数学のテストの結果をまとめた度数分布表とヒストグラムです。

点数	人数
以上　末満	
0 〜 5	1
5 〜 10	2
10 〜 15	6
15 〜 20	8
20 〜 25	16
25 〜 30	18
30 〜 35	22
35 〜 40	24
40 〜 45	22
45 〜 50	17
50 〜 55	13
55 〜 60	11
60 〜 65	9
65 〜 70	7
70 〜 75	6
75 〜 80	5
80 〜 85	4
85 〜 90	3
90 〜 95	2
95 〜 100	2
合計	198

注：この度数分布表では、点数が 100 点の人は最後の階級に含めてある



注：このヒストグラムでは、点数が 100 点の人は最後の階級に含めてある

この度数分布表やヒストグラムからどんなことが読み取れるか考えてみます。すると、

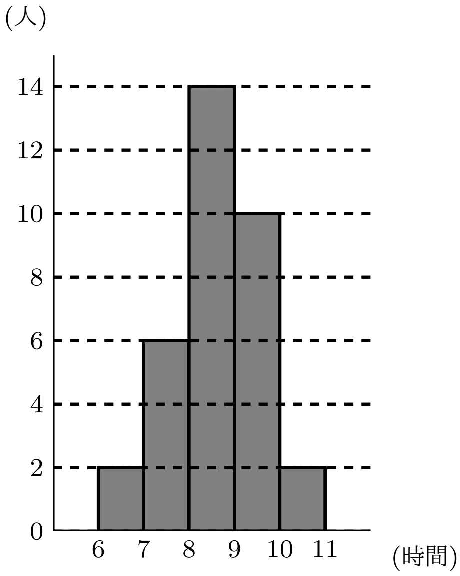
- テストの点数は 0 点ぐらいから 100 点ぐらいまで広く分布していることがわかります。
- データは 0 点ぐらいから 100 点ぐらいまで広く分布しているのでデータの分布の中心はだいたい 50 点ぐらいと考えることができます。
- 点数ががとても低かった人や結構低かった人（つまり 0 点ぐらいから 19 点ぐらいの人）は少ないことがわかります。また 20 点から 49 点ぐらいの人がとても多ことがわかります。さらに、点数が結構り高かった人（つまり 80 点ぐらいから 100 点

の人）は少ないいことがわかります。

データの散らばり方は、中心である 50 点ぐらいからかなり左にかたよっていると言えます。

問 11. 次の図は、ある小学校の 6 年生のあるクラスの生徒達の睡眠時間を調査した結果をまとめた度数分布表とヒストグラムです。

睡眠時間 (時間)	人数 (人)
以上　未満	
6 ～ 7	2
7 ～ 8	6
8 ～ 9	14
9 ～ 10	10
10 ～ 11	2
合計	34



この度数分布表やヒストグラムからどんなことが読み取れるか考えてみます。以下の問に答えなさい。

(1) データは何時間から何時間まで分布していますか。

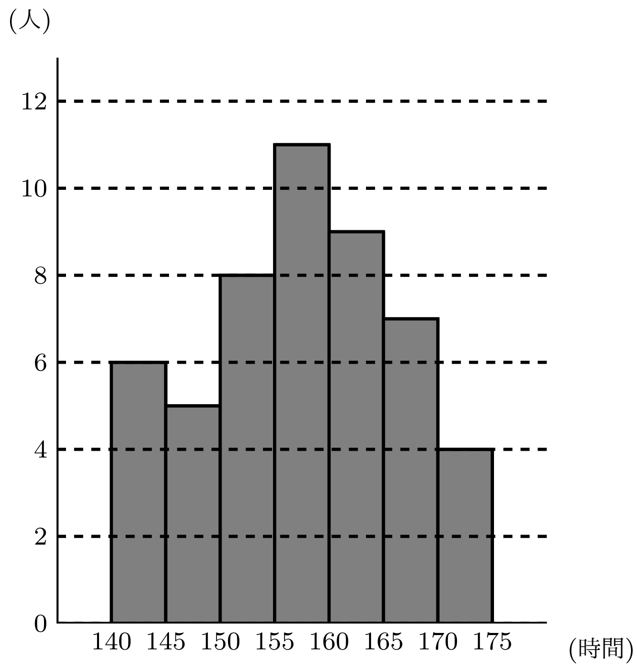
(2) データの分布の中心はだいたい何時間何分ぐらいですか。

(3) データの散らばり方の特徴を言いなさい。

答えを見る

問 12. 次の表は、ある中学校の 1 年生 50 人について身長を調査した結果をまとめた度数分布表とヒストグラムです。

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	6
145 ～ 150	5
150 ～ 155	8
155 ～ 160	11
160 ～ 165	9
165 ～ 170	7
170 ～ 175	4
合計	50



この度数分布表やヒストグラムからどんなことが読み取れるか考えてみます。以下の問に答えなさい。

(1) データは何 cm から何 cm まで分布していますか。

(2) データの分布の中心はだいたい何 cm ぐらいですか。

(3) データの散らばり方の特徴を言いなさい。

答えを見る

1.3.2　範囲：データの広がりをあらわす数値

これまで学んできたように、例えばいくつかの集団にたいして同じ調査をしたとしても、得られるデータには違いがあるのが普通です。また 1 つの集団に何種類かの似ている調査をしても、得られるデータには違いがあるのが普通です。このようなとき、特に、データの散らばり具合に違いが出てくるわけです。ですから例えば、調査の結果得られたデータの値がいくつからいくつまで広がっているのかということに注目すると集団の違いが見えてきたり、調査の種類による違いが見えてくることになるわけです。そこで、「得

られたデータの最大の値」から「得られたデータの最小の値」をひいて 1 つの数を作ることにします。そのようにして作られる数を分布の**範囲**と呼びます。

具体的な話がないとわかりにくいかもしれませんね。そこでこれから、例を使って説明することにしましょう。

例 7　ある中学校の 1 年生と 3 年生の身長を調査してみたら・・・

次の表は、ある中学校の 1 年生 50 人について身長を調査した結果です。

データ番号	身長 (cm)	データ番号	身長 (cm)	データ番号	身長 (cm)
No.1	142.7	No.21	151.5	No.41	152.4
No.2	164.7	No.22	163.8	No.42	158.4
No.3	158.8	No.23	156.9	No.43	143.5
No.4	146.2	No.24	159.9	No.44	156.2
No.5	162.9	No.25	170.8	No.45	169.6
No.6	155.1	No.26	145.1	No.46	166.3
No.7	157.3	No.27	170.3	No.47	154.7
No.8	171.8	No.28	159.7	No.48	168.4
No.9	160.6	No.29	167.0	No.49	157.5
No.10	167.8	No.30	147.3	No.50	161.8
No.11	136.4	No.31	153.8		
No.12	161.3	No.32	163.1		
No.13	148.3	No.33	150.9		
No.14	169.1	No.34	138.5		
No.15	141.2	No.35	164.2		
No.16	157.8	No.36	159.3		
No.17	151.3	No.37	152.0		
No.18	167.5	No.38	171.5		
No.19	142.6	No.39	162.2		
No.20	154.0	No.40	146.9		

この表を一生懸命見て見てください。身長の最も低い人は No.11 の人で 136.4 cm ですね。また、身長の最も高い人は No.8 の人で 171.8 cm ですね。(「得られたデータの最大の値は 171.8」で「得られたデータの最小の値は 136.4」ということです。) つまり、この

中学校の 1 年生の身長は 136.4 cm から 171.8 cm まで広がっているのです。ですから、広がりの幅、つまり範囲は、

$$\text{中学 1 年 50 人の身長の分布の範囲} = 171.8 - 136.4 = 35.4$$

ということになるわけです。

次の表は、同じ中学校の 3 年生 42 人について身長を調査した結果です。

データ番号	身長 (cm)	データ番号	身長 (cm)	データ番号	身長 (cm)
No.1	155.5	No.21	161.5	No.41	162.4
No.2	163.8	No.22	168.5	No.42	167.2
No.3	165.2	No.23	170.1		
No.4	169.5	No.24	157.6		
No.5	169.5	No.25	154.0		
No.6	164.0	No.26	160.5		
No.7	157.8	No.27	162.5		
No.8	162.5	No.28	172.9		
No.9	161.0	No.29	165.5		
No.10	169.5	No.30	174.3		
No.11	152.5	No.31	164.3		
No.12	166.0	No.32	166.8		
No.13	163.5	No.33	158.4		
No.14	163.8	No.34	165.2		
No.15	164.0	No.35	161.2		
No.16	158.2	No.36	167.5		
No.17	170.1	No.37	173.0		
No.18	162.4	No.38	163.5		
No.19	174.8	No.39	172.5		
No.20	159.5	No.40	171.2		

この表を一生懸命見てください。身長の最も低い人は No.11 の人で 152.5 cm ですね。また、身長の最も高い人は No.19 の人で 174.8 cm ですね。(「得られたデータの最大の値は 174.8」で「得られたデータの最小の値は 152.5.4」ということです。）つまり、この中学校の 3 年生の身長は 152.5 cm から 174.8 cm まで広がっているのです。ですから、広が

りの幅、つまり範囲は、

$$\text{中学 3 年 42 人の身長の分布の範囲} = 174.8 - 152.5 = 22.3$$

ということになるわけです。

もうおわかりだと思いますが、このようにして計算される「範囲」という数値は、データの広がり具合をあらわす量となっているわけです。この例では、中学 1 年生の身長の範囲（広がり）は 35.4 cm で中学 3 年生身長の範囲（広がり）は 22.3 cm ですから、データは中学 3 年生より中学 1 年生の方が広い範囲に広がっているということですね。

問 13. 2 つのチーム A と B があります。A チームと B チームの人数はどちらも 11 人です。2 つのチームでハンドボール投げの測定をしました。その結果、記録は右のようになりました。以下の問に答えなさい。

(1) A チームの記録の範囲を求めなさい。

(2) B チームの記録の範囲を求めなさい。

(3) 分布の範囲が広いのはどちらのチームですか。

A チーム

選手番号	記録 (m)
No.1	29
No.2	28
No.3	28
No.4	29
No.5	27
No.6	25
No.7	31
No.8	29
No.9	27
No.10	27
No.11	28

B チーム

選手番号	記録 (m)
No.1	26
No.2	29
No.3	28
No.4	24
No.5	30
No.6	26
No.7	25
No.8	30
No.9	28
No.10	28
No.11	34

答えを見る

1.3.3　代表値：散らばりのあるデータを代表するいくつかの数値

先に学んだ「範囲」と呼ばれている数値も散らばりのあるデータの特徴をあらわす数値の 1 つですが、これだけでは分布の様子を詳しくあらわすことはできません。そこでこの他にも、分布の特徴をあらわすための方法が色々と考えられています。そのようなもののうち、よく使われるものが平均値、中央値（別名、メジアン）、最頻値（別名、モード）と呼ばれているものです。これらの数値は、「調査をしているデータの特徴を代表する数値」と考えることができるのです。そこで、平均値、中央値（別名、メジアン）、最頻値（別名、モード）のように、調査をしているデータの特徴を代表する数値のことを**代表値**と呼んでいます。（つまり、平均値、中央値（別名、メジアン）、最頻値（別名、モード）はどれも代表値の仲間です。）これから、平均値、中央値（別名、メジアン）、最頻値（別名、モード）についてゆっくり学んでいくことにします。

平均値ってなに

「得られたデータの数値の合計」を「データの個数」で割って得られる数を**平均値**と呼びます。

例 8　ある中学校の 1 年のあるクラスで 10 点満点の小テストをしました。その結果は右のようになりました。このクラスの小テストの平均点を求めてみます。

生徒番号	点数	生徒番号	点数
No.1	7	No.11	6
No.2	4	No.12	4
No.3	5	No.13	5
No.4	5	No.14	5
No.5	3	No.15	6
No.6	3	No.16	4
No.7	6	No.17	2
No.8	6	No.18	5
No.9	7	No.19	2
No.10	3	No.20	4

まず右の表をよく見て、「得られたデータの数値の合計」を求めます。つまり点数の合計を求めるわけです。すると、

$$
\begin{aligned}
\text{得られたデータの数値の合計} &= 7+4+5+5+3+3+6+6+7+3 \\
&\quad +6+4+5+5+6+4+2+5+2+4 \\
&= 92
\end{aligned}
$$

となりますね。

次に、「得られたデータの数値の合計」を「データの個数」で割れば平均値が求められるわけです。生徒の人数は 20 人ですから「データの個数」は 20 です。ですから、

$$\text{平均値} = 92 \div 20 = 4.6$$

となりますね。

以上でこのクラスの小テストの平均点は 4.6 点であることがわかりました。

問 14. ある中学校の 1 年のあるクラスで 10 点満点の小テストをしました。その結果は右のようになりました。以下の問に答えなさい。

(1) 小テストの点数の平均値を求めなさい。

(2) 小テストの点数の分布の範囲を求めなさい。

生徒番号	点数	生徒番号	点数
No.1	3	No.11	4
No.2	6	No.12	5
No.3	6	No.13	6
No.4	3	No.14	4
No.5	3	No.15	5
No.6	4	No.16	6
No.7	3	No.17	3
No.8	5	No.18	6
No.9	6		
No.10	3		

答えを見る

問 15. 右の表はある中学校の 1 年生である A 君と B 君の過去 10 回の 50 m 走の記録です。以下の問に答えなさい。

A 君

データ番号	記録 (秒)
No.1	7.4
No.2	7.2
No.3	7.2
No.4	7.1
No.5	7.1
No.6	7.3
No.7	7.1
No.8	7.2
No.9	7.3
No.10	7.1

B 君

データ番号	記録 (秒)
No.1	7.2
No.2	7.3
No.3	7.5
No.4	7.4
No.5	7.2
No.6	6.9
No.7	7.0
No.8	7.2
No.9	7.1
No.10	7.2

(1) A 君のタイムの平均値を求めなさい。

(2) B 君のタイムの平均値を求めなさい。

(3) 分布の範囲が広いのは A 君ですかそれとも B 君ですか。

(4) この中学の 1 年に C 君という生徒がいます。C 君の 50 m 走のタイムの平均値は 7.0 秒です。A 君と B 君のうちのどちらかを選んで C 君と 50 m 走で 1 回だけ対決することとします。C 君に勝てる可能性の高いのは A 君でしょうか、それとも B 君でしょうか。

答えを見る

中央値（別名メジアン）ってなに

ある調査を行ってデータが得られたとします。このとき、次の手順で得られる数値をそのデータの**中央値（別名、メジアン）** と呼びます。

手順 まず、調査の結果得られたデータを大きさの順に並べます。小さい方から順に並べても構いませんし、大きい方から順に並べても構いません。

手順 2 小さい方から数えてもよいですし、大きい方から数えても良いのですが、ちょうど真ん中にあるデータの値を読み取ります。このようにして読み取った数値をそのデータの**中央値 (別名メジアン)** と呼ぶのです。

ただし、ここで注意しておくことがあります。データの数が偶数個の時は、ちょうど真ん中には何もデータはないですよね。こういう時はちょうど真ん中の前後に 2 つデータがあるわけです。データの数が偶数のときはこの 2 つのデータをたして 2

でわった値を中央値と考えることにします。

例 9　右の表は、ある中学校の 1 年のあるクラスで女子のソフトボール投げの記録をとったものです。このデータの中央値を求めてみることにしましょう。

データ番号	記録 (m)	データ番号	記録 (m)
No.1	8	No.11	13
No.2	14	No.12	15
No.3	16	No.13	11
No.4	13		
No.5	10		
No.6	7		
No.7	12		
No.8	14		
No.9	14		
No.10	10		

まず、得られたデータを大きさの順に並べるのですよね。ではこの表を見ながら間違わないように慎重に並べることにしましょう。小さい値から並べても良いですし、大きい値から並べても良いわけですが、ここでは小さい値から並べてみます。すると、

7、8、10、10、11、12、13、13、14、14、14、15、16

となりますね。

次は大きさの順になったデータを見て、ちょうど真ん中に並んでいるるデータを読み取るのでしたね。今、データの数は 13 個なので、7 番目に並んでいるデータを探せばよいわけです。すると、13 が見つかりますね。ですから、女子のソフトボール投げの中央値（別名、メジアン）は 13 m となるわけです。

例 10　右の表は、ある中学校の 1 年のあるクラスで男子のソフトボール投げの記録をとったものです。このデータの中央値を求めてみることにしましょう。

データ番号	記録 (m)	データ番号	記録 (m)
No.1	20	No.11	17
No.2	11	No.12	21
No.3	13		
No.4	14		
No.5	19		
No.6	22		
No.7	28		
No.8	11		
No.9	18		
No.10	14		

まず、得られたデータを大きさの順に並べるのですよね。ではこの表を見ながら間違わないように慎重に並べることにしましょう。小さい値から並べても良い

ですし、大きい値から並べても良いわけ
ですが、ここでは小さい値から並べてみます。すると、

11、11、13、14、14、17、18、19、20、21、22、28

となりますね。

次は大きさの順になったデータを見て、ちょうど真ん中に並んでいるるデータを読み取るのでしたね。でも今、データの数が 12 個なので、ちょうど真ん中にはデータがありません。このようなときは、「ちょうど真ん中」の前後にある 2 つのデータの値をたして 2 でわるのでしたね。「ちょうど真ん中」の前後にある 2 つのデータは 17 と 18 ですよね。というわけで、ソフトボール投げの中央値 (別名メジアン) は、

$$(17+18)\div 2=35\div 2=17.5\,(\mathrm{m})$$

となるわけです。

問 16. 右の表は、ある中学校の 1 年のあるクラス女子の立ち幅とびの記録をとったものです。このデータの中央値を求めなさい。

データ番号	記録 (cm)	データ番号	記録 (cm)
No.1	180	No.11	195
No.2	184	No.12	210
No.3	195	No.13	180
No.4	181	No.14	165
No.5	167	No.15	176
No.6	190		
No.7	132		
No.8	170		
No.9	195		
No.10	169		

答えを見る

問 17. 右の表は、ある中学校の 1 年のあるクラス男子の立ち幅とびの記録をとったものです。このデータの中央値を求めなさい。

データ番号	記録 (cm)	データ番号	記録 (cm)
No.1	166	No.11	204
No.2	177	No.12	187
No.3	200	No.13	188
No.4	144	No.14	163
No.5	171		
No.6	193		
No.7	178		
No.8	153		
No.9	175		
No.10	202		

答えを見る

最頻値（別名、モード）ってなに

　ある調査を行ってデータを集め、そのデータを度数分布表にまとめたとします。そのとき、度数の最も大きい階級の真ん中の数値のことを、そのデータの**最頻値 (別名モード)** と呼びます。（度数が最も大きい階級というのは、調査の中で最も頻繁に出てきた階級ですよね。ですからその階級の真ん中の値を最頻値と呼ぶのです。）

例 11　右の表は、ある中学校の 1 年生 50 人について身長を調査して作った度数分布表です。

　この度数分布表を見ると、度数の最も大きい階級は、

155 cm〜160 cm の階級

　であることがわかります。（11 人より多いところはないですよね。）ということは、最頻値は、「155 cm〜160 cm」の階級の真ん中の値なので 155.5 cm ということになりますね。

身長 (cm)	人数 (人)
以上　未満	
140 〜 145	6
145 〜 150	5
150 〜 155	8
155 〜 160	11
160 〜 165	9
165 〜 170	7
170 〜 175	4
合計	50

問 18. あるバス会社では、平日の午前 9 : 10 に A 町を出発する B 町行きのバスについて、過去 160 回の運行状況の記録を調べてみました。そして、右のような度数分布表ができました。かかる時間の最頻値を求めなさい。

かかった時間 (分)	回数 (回)
以上　未満	
30 〜 35	3
35 〜 40	37
40 〜 45	64
45 〜 50	45
50 〜 55	11
合計	160

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表 (160 回分の調査)

答えを見る

1.3.4　平均値、中央値（別名、メジアン）、最頻値（別名、モード）のうちどれを使うといいの？

散らばりのあるデータの分析では、そのデータを代表する数値として、平均値、中央値（別名、メジアン）、最頻値（別名、モード）というものがあるということを学びました。しかしこれらの数値はどれも、散らばりあるデータが持っている特徴をたった 1 つ数値だけで表そうとしたものです。ですからかなり無理なことをしているわけです。データの散らばり具合によってはそのデータの集まりが持っている特徴をうまくとらえられないことがあるのです。どういうことなのか、いくつか例を使って説明しましょう。

例 12　たった 1 つでも極端なデータがあると平均値は結構変わってしまうこともある

P さんは野球のチームに入っています。P さんは自分のチームの得点力を分析しようと思い、過去 5 試合の P さんのチームの得点を調べてみました。右のデータはその結果です。

試合日	得点
7/8	3
7/15	1
7/22	2
7/29	2
8/5	19

P さんはとりあえず、得点の平均値を求めてみることにしました。

得点の合計は、

$$\text{合計} = 3 + 1 + 2 + 2 + 19 = 27\,(\text{点})$$

で、試合数は 5 ですから、得点の平均値は、

$$\text{得点の平均値} = 27 \div 5 = 5.4\,(\text{点})$$

となりますね。

どうでしょう。この計算結果を見て、「平均得点は 5.4 か。ということは、自分のチームは試合をすると、だいたい 5 点から 6 点ぐらい得点できるということだな。」と考えても良いのでしょうか。あんまりそんな気はしないですよね。だって、気になる試合結果が 1 つありますよね。まあ、この 5 試合の相手がどんなチームっだったのかということも気になりますが、1 試合だけ極端なデータがあります。5 試合のうちの 4 試合の得点は 1 点から 3 点 ですが、1 試合だけ 19 点もとっているのです。（8/5 の試合のことですよ。）この試合だけは例外と考えたほうが良いのかもしれませんよね。そこで得点が 19 のデータを除外して、4 試合の得点の平均値を求めることにします。

その 4 試合の得点の合計は、

$$\text{合計} = 3 + 1 + 2 + 2 = 8\,(\text{点})$$

で、試合数は 4 ですから、得点の平均値は、

$$\text{その 4 試合の得点の平均値} = 8 \div 4 = 2\,(\text{点})$$

となりますね。

得点が 19 の試合も含めた場合の得点の平均値を計算すると 5.4 点で、得点が 19 の試合を含めない場合の得点の平均値を計算すると 2 点となりました。というわけで、得点が 19 の試合のせいで、平均得点が 3.4 も上がってしまうのです。

例 13　たった 1 つ極端なデータがあっても平均値は少ししか変わらないこともある

Q さんは野球のチームに入っています。Q さんは自分のチームの得点力を分析しようと思い、過去 20 試合の Q さんのチームの得点を調べてみました。右のデータはその結果です。

試合日	得点	試合日	得点
7/3	5	9/11	3
7/10	4	9/18	5
7/17	4	9/25	19
7/24	6	10/2	4
7/31	7	10/98	6
8/7	3	10/16	8
8/14	6	10/23	2
8/21	5	10/30	6
8/28	4	11/6	4
9/4	0	11/13	7

Q さんはとりあえず、得点の平均値を求めてみることにしました。

得点の合計は、

$$\begin{aligned}\text{合計} &= 5+4+4+6+7+3+6+5+4+0 \\ &\quad +3+5+19+4+6+8+2+6+4+7 \\ &= 108\,(\text{点})\end{aligned}$$

で、試合数は 20 ですから、得点の平均値は、

$$\text{得点の平均値} = 108 \div 20 = 5.4\,(\text{点})$$

となりますね。

どうでしょう。この計算結果を見て、「平均得点は 5.4 か。ということは、自分のチームは試合をすると、だいたい 5 点から 6 点ぐらい得点できるということだな。」と考えても良いでしょうか。だって、気になる試合結果が 1 つありますよね。まあ、この 20 試合の相手がどんなチームっだったのかということも気になりますが、1 試合だけ極端なデータがありますね。20 試合のうちの 19 試合は得点は 0 点から 8 点 ですが、1 試合だけ 19

点もとっているのです。(9/25 の試合のことですよ。) この試合だけは例外と考えたほうが良いのかもしれませんよね。そこで、得点が 19 のデータを除外して、19 試合の得点の平均値を求めることにします。

その 19 試合の得点の合計は、20 試合分の合計点から 19 点を引けば求めることができますね。というわけで、

$$\text{合計} = 108 - 89\,(\text{点})$$

となります。試合数は 19 ですから、得点の平均値は、

$$\text{その 19 試合の得点の平均値} = 89 \div 19 = 4.68 \cdots\ (\text{点})$$

となりますね。

得点が 19 の試合も含めて得点の平均値を計算すると 5.4 点で、得点が 19 の試合を含めないで得点の平均値を計算すると 4.68… 点となりました。得点が 19 の試合のせいで、平均得点が 0.7 点程度上がっています。ですから、平均得点は、「まあ、そんなに極端に上がるわけでもない」と考える人もいるかもしれませんし、「0.7 点ででも結構大きい違いだよ」と考える人もいるかもしれませんね。

問 19. 例 12 と例 13 が理解できた人への問題です。

例 12 では得点が 19 の試合が 1 試合あるせいで平均得点が 3.4 も上がりましたが、例 13 では得点が 19 の試合が 1 試合あっても平均得点は 0.7 ぐらいしか上がりませんでした。どうしてこんな違いが出たのでしょう。原因として最もふさわしいものを次から選びなさい。

a. 得点が 19 の試合を除くと、例 12 の P さんのチームの得点は 1 点から3 点ばかりで得点はあまり散らばっていないが、13 の Q さんのチームの得点は 0 点から8 点まで結構散らばりがある。この「散らばりの広さの違い」が原因である。

b. 得点が 19 の試合を除くと、例 12 の P さんのチームの平均得点は 2 点と低いが、13 の Q さんのチームの平均得点は 4.68… 点とわりと高い。この 「得点が 19 の試合を除

いた場合の平均得点の違い」が原因である。

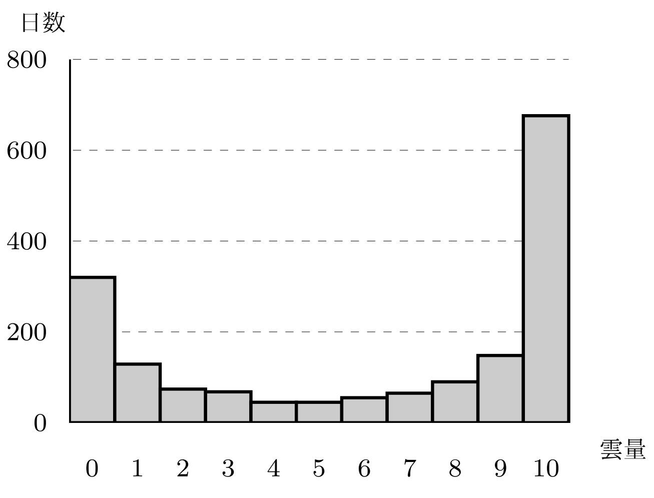
c. 得点が 19 の試合を除くと、例 12 の P さんのチームの試合はたった 4 試合であるが、13 の Q さんのチームは 19 試合もある。この、「 得点が 19 の試合を除いた試合数の違い」が原因である。

答えを見る

例 14　集められたデータの特徴を中央値を使ってあらわすのは適当ではないこともある。

次の図は、ある町で 1715 日にわたって毎日雲の量を観察した結果をまとめた度数分布表とヒストグラムです。

雲量	日数
0	320
1	129
2	74
3	68
4	45
5	45
6	55
7	65
8	90
9	148
10	676
合計	1715



まず、雲量について簡単に説明しておきましょう。

雲の量は 10 段階で表されます。空全体を見渡しても雲がないときは雲量は 0、空全体の広さのうち 1 割程度が雲に覆われているときは雲量は 1、空全体の広さのうち 2 割程度が雲に覆われているときは雲量は 2、・・・、空全体が雲に覆われているとき雲量は 10 とするのです。

それではとりあえず、中央値を調べてみることにします。

中央値を求めるには、観察された 1715 日分のデータを大きさの順に並べなくてはいけ

ませんよね。そうすると、「えっ、じゃあ無理じゃん。今ここには、度数分布表ならあるけど、おおもとの 1715 日分の観察データはないんだから。」って思う人もいることでしょう。でも、この場合、「おおもとの 1715 日分の観察データ」はなくても「おおもとの 1715 日分の観察データを大きさの順に並べたもの」を度数分布表から復元できるんです。どうすればよいかわかりますか？

例えばおおもとのデータを小さい順に並べるとどうなるのか考えてみましょう。

度数分布表で雲量が 0 の階級を見ると、雲量 0 のデータは 320 個あったということがわかりますよね。ですから、おおもとのデータを小さい順に並べるとまず、次のように 0 が 320 個並ぶのです。

$\underbrace{0, 0 \cdots, 0}_{320\text{個}}$

次にに雲量が 1 の階級をみると、雲量 1 のデータは 129 個あったということがわかりますよね。ですから、おおもとのデータを小さい順に並べるとま次のように 0 が 320 個並んだあとに、1 が 129 個並ぶのです。

$\underbrace{0, 0 \cdots, 0}_{320\text{個}}, \underbrace{1, 1 \cdots, 1}_{129\text{個}}$

ここまで並べたデータの数は、

$$320 + 129 = 449\text{個}$$

ですよね。

次にに雲量が 2 の階級をみると、雲量 2 のデータは 74 個あったということがわかりますよね。ですから、おおもとのデータを小さい順に並べると、次のように、さっきのあとに、2 が 74 個並ぶのです。

$\underbrace{0, 0 \cdots, 0}_{320\text{個}}, \underbrace{1, 1 \cdots, 1}_{129\text{個}}, \underbrace{2, 2 \cdots, 2}_{74\text{個}}$

ここまで並べたデータの数は、

$$320 + 129 + 74 = 523\,個$$

ですよね。

この先も同じように考えていきましょう。すると例えば、雲量が 7 のところまで調べてみれば、おおもとのデータを小さい順に並べると、次のように並んでいることがわかりますね。

$$\underbrace{0,0\cdots,0}_{320\,個},\underbrace{1,1\cdots,1}_{129\,個},\underbrace{2,2\cdots,2}_{74\,個},\underbrace{3,3\cdots,3}_{68\,個},\underbrace{4,4\cdots,4}_{45\,個},\underbrace{5,5\cdots,5}_{45\,個},\underbrace{6,6\cdots,6}_{55\,個},\underbrace{7,7\cdots,7}_{65\,個}$$

ここまで並べたデータの数は、

$$320 + 129 + 74 + 68 + 45 + 45 + 55 + 65 = 801\,個$$

ですよね。

ところで、データは全部で 1715 個あるのでしたね。では、中央値は何個目に出てくるデータでしょうか。$1715 \div 2 = 857.5$ となりますから、中央値は 858 個目に出てくるデータですよね。今、雲量が 7 のところまでデータを 801 個並べたのですから、中央値はもう少しあとに出てくるわけですね。

というわけで、次にに雲量が 8 の階級を見てみましょう。すると、雲量 8 のデータは 90 個あったということがわかりますよね。ですから、おおもとのデータを小さい順に並べると、次のように、さっきのあとに、8 が 90 個並ぶのです。

$$\underbrace{0,0\cdots,0}_{320\,個},\underbrace{1,1\cdots,1}_{129\,個},\underbrace{2,2\cdots,2}_{74\,個},\underbrace{3,3\cdots,3}_{68\,個},\underbrace{4,4\cdots,4}_{45\,個},\underbrace{5,5\cdots,5}_{45\,個},\underbrace{6,6\cdots,6}_{55\,個},\underbrace{7,7\cdots,7}_{65\,個},\underbrace{8,8\cdots,8}_{90\,個}$$

ここまで並べたデータの数は、

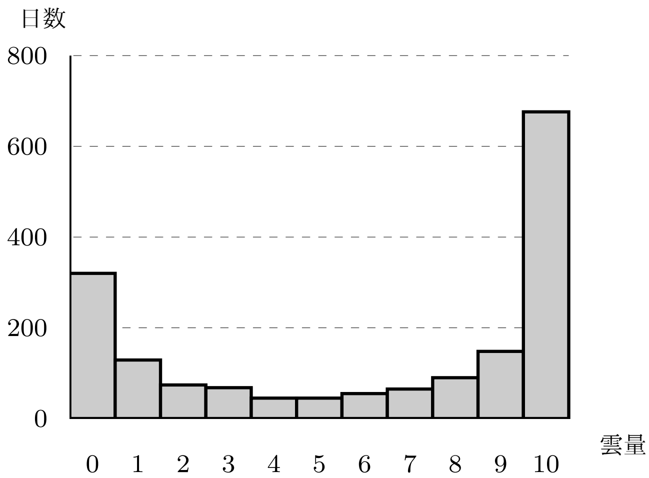
$$320 + 129 + 74 + 68 + 45 + 45 + 55 + 65 + 90 = 891\,個$$

ですよね。851 を超えました。

中央値は858個目に出てくるデータですから、中央値は8ということになりますね。これで中央値がわかりました。

ところでいま求めた中央値である雲量8ですが、この町の雲量を代表する数値としてふさわしいでしょうか。度数分布表やヒストグラムをもう一度見てみましょう。あなたのためにもう一度、次に度数分布表やヒストグラムを載せておきます。

雲量	日数
0	320
1	129
2	74
3	68
4	45
5	45
6	55
7	65
8	90
9	148
10	676
合計	1715



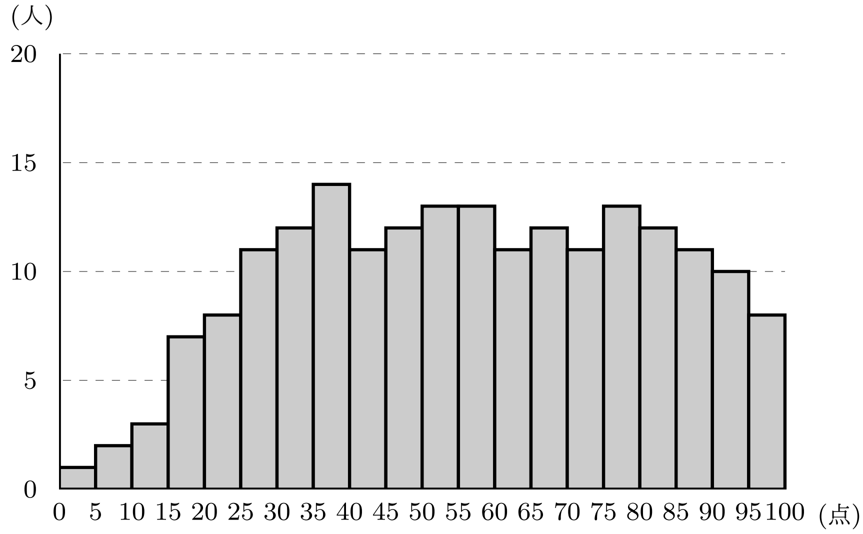
これを見ると、この街では雲量8の日ってそんなに多くはないですよね。雲量10の日が圧倒的に多いんですよね。雲量8というのは空がほとんど雲に覆われているということで、雲量10というのは空が完全に雲に覆われているということなのですからそんなに違いはないと感じる人もいるかもしれません。ですが、一方、雲量が0の日は雲量8の日に比べるとかなり多いわけです。つまり、ほとんど雲に覆われている雲量8の日に比べると、すっかり晴れている雲量0の日のほうがかなり多いわけです。このようなことを考えに入れると、この町の雲量を代表する数値として雲量8を使うのは気が引けますよね。つまりこの町の雲量を中央値で代表させるのはちょっと無理があると言えるかもしれません。この街では完全に曇っている日が圧倒的に多いのですが、すっかり晴れている日も結構多いのですから。

例 15　集められたデータの特徴を最頻値を使ってあらわすのは適当ではないこともある。

次の図は、ある中学校の 3 年生 195 人の英語のテストの結果をまとめた度数分布表とヒストグラムです。

点数	人数
以上　未満	
0 ～ 5	1
5 ～ 10	2
10 ～ 15	3
15 ～ 20	7
20 ～ 25	8
25 ～ 30	11
30 ～ 35	12
35 ～ 40	14
40 ～ 45	11
45 ～ 50	12
50 ～ 55	13
55 ～ 60	13
60 ～ 65	11
65 ～ 70	12
70 ～ 75	11
75 ～ 80	13
80 ～ 85	12
85 ～ 90	11
90 ～ 95	10
95 ～ 100	8
合計	195

注：この度数分布表では、点数が 100 点の人は最後の階級に含めてある



注：このヒストグラムでは、点数が 100 点の人は最後の階級に含めてある

これを見ると、度数が最も大きい階級は「35 点位上 40 点未満」の階級であることがわかりますね。（つまり、35 点位上 40 点未満だった生徒が 14 人で一番多かったわけですよね。）ですから、最頻値は、

$$最頻値 = (35 + 40) \div 2 = 35.5\,(点)$$

ということになりますよね。（最頻値とは、最頻の階級の真ん中の値のことでしたね。）ところで、今求めた 35.5 点という最頻値ですが、この 195 人の生徒たちの英語のテストの結果を代表する数値としてふさわしいでしょうか。度数分布表やヒストグラムをもう一度よく見てください。25 点以上 30 点未満だった生徒の人数、30 点以上 35 点未満だった生徒の人数、・・・、85 点以上 90 点未満だった生徒の人数は 11 人から 14 人なのであまり変わらないといえますよね。つまり、25 点から 90 点ぐらいまで、だいたい同じような人数がいるわけです。ですから、いくら 35 点位上 40 点未満だった生徒が一番多かったからといって、35.5 点という最頻値をつかってこの生徒たちの英語の実力を代表させるのは気が引けますよね。

問 20. 次の表は、ある中学校の 1 年生 10 人の 1 ヶ月のおこづかいの金額を表したものです。

生徒名 (仮名)	A	B	C	D	E	F	G	H	I	J
金額 (円)	800	1200	900	8200	800	1000	1100	1100	1500	800

(1) この 10 人の生徒のおこづかいの額の平均値を求めなさい。

(2) この 10 人の生徒のおこづかいの額の中央値を求めなさい。

(3) 階級の幅を 100 円にして度数分布表を作りなさい。そしておこづかいの額の最頻値を求めなさい。

(4) (1)、(2)、(3) の答えが正しく求められた人は、平均値は 1740 円、中央値は 1050 円、最頻値は 850 円になっているはずです。

実は、プライバシーを守るため、おこづかいを調査したこの表はは秘密になっていたのですが、平均値、中央値、最頻値は発表されましたその発表を聞いた A 君、B さん、I さん、E さんの考えについて、以下の質問に答えなさい。

（a）A 君は、「平均値は 1740 円かぁ。ということは、おこづかいが 1750 円以下の人とおこづかいが 1750 円以上の人が 5 人づついるってことだな。」と考えました。あなたは A 君の考えについてどう思いますか。

（b）B さんは、「中央値は 1050 円かぁ。ということは、おこづかいが 1050 円以下

の人とおこづかいが 1050 円以上の人が 5 人づついるってことだね。」と考えました。あなたは B さんの考えについてどう思いますか。

(c) I さんは、「平均値は 1740 円かぁ。私のおこづかいは 1500 円だから、みんなに比べて私のおこづかいって少ないんだね。」と考えました。あなたは I さんの考えについてどう思いますか。

(d) E さんは、「最頻値は 850 円かぁ。私のおこづかいは 800 円だから、みんなと比べても私のおこづかいってほんの少し少ないだけだね。」と考えました。あなたは E さんの考えについてどう思いますか。

答えを見る

問の解答

問 1.　次の表の、ある中学校の 1 年生 50 人について身長を調査した結果についての問題でしたね。

データ番号	身長 (cm)	データ番号	身長 (cm)	データ番号	身長 (cm)
No.1	142.7	No.21	151.5	No.41	152.4
No.2	164.7	No.22	163.8	No.42	158.4
No.3	158.8	No.23	156.9	No.43	143.5
No.4	146.2	No.24	159.9	No.44	156.2
No.5	162.9	No.25	170.8	No.45	169.6
No.6	155.1	No.26	145.1	No.46	166.3
No.7	157.3	No.27	170.3	No.47	154.7
No.8	171.8	No.28	159.7	No.48	168.4
No.9	160.6	No.29	167.0	No.49	157.5
No.10	167.8	No.30	147.3	No.50	161.8
No.11	136.4	No.31	153.8		
No.12	161.3	No.32	163.1		
No.13	148.3	No.33	150.9		
No.14	169.1	No.34	138.5		
No.15	141.2	No.35	164.2		
No.16	157.8	No.36	159.3		
No.17	151.3	No.37	152.0		
No.18	167.5	No.38	171.5		
No.19	142.6	No.39	162.2		
No.20	154.0	No.40	146.9		

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	6
145 ～ 150	5
150 ～ 155	8
155 ～ 160	11
160 ～ 165	9
165 ～ 170	7
170 ～ 175	4
合計	50

(1) 調査結果を整理して度数分布表を完成すると右のようになります。

(2) (1) で完成した度数分布表では階級の幅は 5 cm です。

(3) No.36 の生徒の身長は 159.3 cm です。ですから No.36 の生徒は「155 cm 以上 160 cm 未満の階級」に入ります。

本文へ戻る

問 2.　ある中学校の 3 年生のあるクラスで睡眠時間の調査をして作った度数分布表についての問題でしたね。

睡眠時間 (時間)	人数 (人)
以上　未満	
5 ～ 6	3
6 ～ 7	10
7 ～ 8	12
8 ～ 9	5
9 ～ 10	1
合計	31

(1) 階級の幅は 1 時間ですね。

(2) A さんの睡眠時間は 7 時間 45 分ですから、A さんは「7 時間以上 8 時間未満の階級」に入ります。

(3) この場合、度数とは「人数」です。そして人数が一番多いのは 12 人です。ですから、度数が一番大きい階級は「7 時間以上 8 時間未満の階級」です。

(4) 6 時間以上 7 時間未満の階級の度数は 10 です。

(5) 度数が一番大きい階級は「7 時間以上 8 時間未満の階級」です。そしてその階級の度数は 12 です。

本文へ戻る

問 3.　平日の午前 9 : 10 に A 町を出発する B 町行きのバスについて、過去 40 回の運行状況の記録をまとめた右の度数分布表についての問題でしたね。

(1) 階級の幅は 5 分です。

(2) この場合、度数とは「回数」です。度数が一番大きい階級は「40 分以上 45 分未満の階級」ですね。

(3) 45 分以上 50 分未満の階級の度数は 13 です。

(4) 度数が一番大きい階級は「40 分以上 45 分未満の階級」です。そしてその階級の度数は 15 です。

かかった時間 (分)	回数 (回)
以上　未満	
30 〜 35	1
35 〜 40	8
40 〜 45	15
45 〜 50	13
50 〜 55	3
合計	40

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表

本文へ戻る

問 4.　ある中学校の 1 年生 50 人について身長を調査して作った右の度数分布表についての問題でしたね。

(1) 身長の低い方から数えていくと、

「140 cm 以上 145 cm 未満の階級」までに 6 人
「145 cm 以上 150 cm 未満の階級」までに

$$6 + 5 = 11 \text{ 人}$$

「150 cm 以上 155 cm 未満の階級」までに

$$6 + 5 + 8 = 19 \text{ 人}$$

いるわけです。まだ 21 人にはなりませんね。

身長 (cm)	人数 (人)
以上　未満	
140 〜 145	6
145 〜 150	5
150 〜 155	8
155 〜 160	11
160 〜 165	9
165 〜 170	7
170 〜 175	4
合計	50

ですが次の階級を考えると、「155 cm 以上 160 cm 未満の階級」までに

$$6 + 5 + 8 + 11 = 30\text{ 人}$$

となり、21 人より多くなりました。ですから、身長の低い方から数えていくとき、身長が 21 番目の生徒は

「155 cm 以上 160 cm 未満の階級」

に入っているわけです。

(2) 身長の高い方から数えていくと、

「170 cm 以上 175 cm 未満の階級」までに 4 人

「165 cm 以上 170 cm 未満の階級」までに

$$4 + 7 = 11\text{ 人}$$

「160 cm 以上 165 cm 未満の階級」までに

$$4 + 7 + 9 = 20\text{ 人}$$

いるわけです。ちょうど 20 人になりままし た。ですから身長の高い方から数えていくとき、身長が 20 番目の生徒は

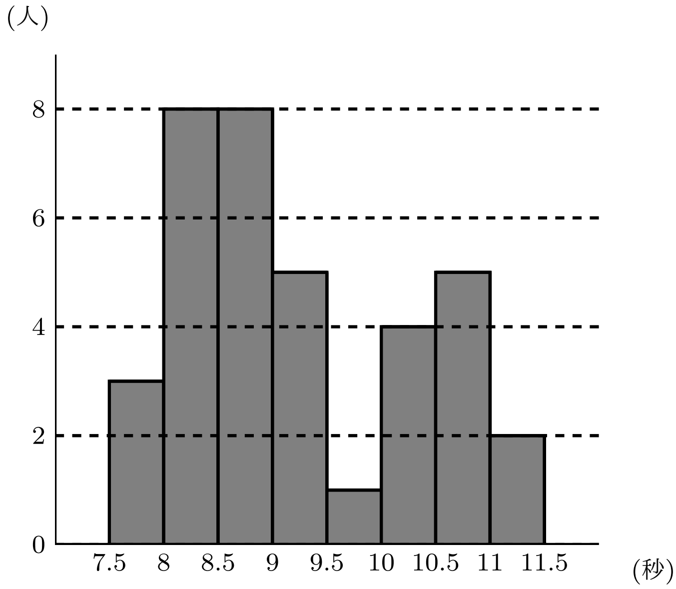
「160 cm 以上 165 cm 未満の階級」

に入っているわけです。

本文へ戻る

問 5.

タイム (秒)	人数 (人)
以上　未満	
7.5 ～ 8.0	3
8.0 ～ 8.5	8
8.5 ～ 9.0	8
9.0 ～ 9.5	5
9.5 ～ 10.0	1
10.0 ～ 10.5	4
10.5 ～ 11.0	5
11.0 ～ 11.5	2
合計	36

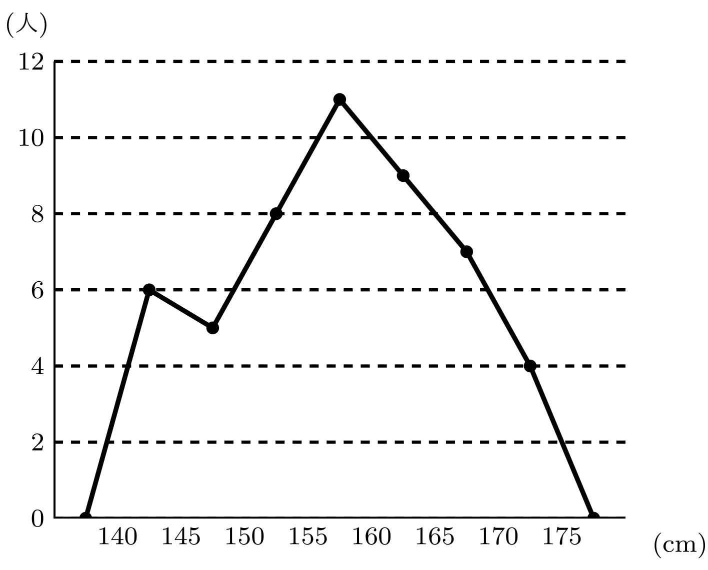


左側の度数分布表をよく見て、右にヒストグラムを作るとこのようになりますね。

本文へ戻る

問 6.　次を見てください。

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	6
145 ～ 150	5
150 ～ 155	8
155 ～ 160	11
160 ～ 165	9
165 ～ 170	7
170 ～ 175	4
合計	50



左側の度数分布表をもとにして度数折れ線を作ると右のようになりますね。

本文へ戻る

問 7. A 中学校では 510 人のうち 87 人が「サッカーが一番好き」なのですから、

$$\text{A 中学校でサッカーが一番好きという生徒の割合} = \frac{87}{510} = 0.1705\cdots$$

となります。（百分率で言うと、A 中学校ででサッカーが一番好きという生徒は約 17% ということですね。）

B 中学校では 320 人のうち 52 人が「サッカーが一番好き」なのですから、

$$\text{B 中学校でサッカーが一番好きという生徒の割合} = \frac{52}{320} = 0.1625$$

となります。（百分率で言うと、B 中学校でサッカーが一番好きという生徒は 16.25% ということですね。）

ということは、「A 中学校のほうが B 中学校より少しサッカーの人気が高い」と言えますね。

本文へ戻る

問 8. 『ある町に住んでいる 281 人の人に一番好きなスポーツを答えてもらうアンケートをしてみたところ、右の表のような結果になりました。以下の問に答えなさい。』ということでした。

一番好きなスポーツ	票数	割合
野球	52	0.19
サッカー	42	0.15
テニス	20	0.07
自転車	14	0.05
水泳・競泳	14	0.05
バレーボール	14	0.05
陸上	13	0.05
ゴルフ	11	0.04
バスケットボール	11	0.04
フィギュアスケート	11	0.04
ラグビー	11	0.04
モータースポーツ（四輪）	7	0.02
バドミントン	6	0.02
モータースポーツ（二輪）	5	0.02
その他	50	0.18
合計	281	

(1) 『野球についてはもう「割合」が記入されていますが、他のそれぞれのスポーツについてはあなたが「割合」を計算し、この表の「割合」の欄に数値を記入してください。ただし、「割合」は少数第三位を四捨五入して少数第二位までの数値を書いてください。』という問題でした。

ちゃんと計算すると右のような数値になります。

(2) 『すべてのスポーツについて割合を合計するといくつになるはずですか。』という問題でした。

割合を全てたすと 1 になるはずですよね。

補足 : (1) で記入した割合の数値を全部たすと 1.01 になってしまい 1 ちょうどにはなりません。これは四捨五入をしたせいですね。

本文へ戻る

問 9. ある中学校の 1 年 A 組と 1 年 B 組で男子の 50 m走のタイムがどのように違うのか比べるため、データを分析する問題でしたね。

(1) 1 年 A 組の男子 20 人の 50 m 走のタイムについて、それぞれの階級に対して相対度数を計算すると右の表のようになります。ただし、「相対度数」は少数第三位を四捨五入して少数第二位までの数値が書いてあります。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 〜 7.0	2	0.10
7.0 〜 7.5	3	0.15
7.5 〜 8.0	4	0.20
8.0 〜 8.5	2	0.10
8.5 〜 9.0	2	0.10
9.0 〜 9.5	1	0.05
9.5 〜 10.0	2	0.10
10.0 〜 10.5	3	0.15
10.5 〜 11.0	1	0.05
11.0 〜 11.5	0	0.00
合計	20	1.00

1 年 A 組の男子 20 人の 50 m 走のタイムの度数分布と相対度数分布

(2) 1 年 B 組の男子 22 人の 50 m 走のタイムについて、それぞれの階級に対して相対度数を計算すると右の表のようになります。ただし、「相対度数」は少数第三位を四捨五入して少数第二位までの数値が書いてあります。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 〜 7.0	1	0.05
7.0 〜 7.5	2	0.09
7.5 〜 8.0	4	0.18
8.0 〜 8.5	5	0.23
8.5 〜 9.0	3	0.14
9.0 〜 9.5	2	0.09
9.5 〜 10.0	2	0.09
10.0 〜 10.5	2	0.09
10.5 〜 11.0	1	0.05
11.0 〜 11.5	0	0.00
合計	22	1.00

1 年 B 組の男子 22 人の 50 m 走のタイムの度数分布と相対度数分布

(3) (1) で計算した A 組の相対度数と (2) で計算した B 組の相対度数を記入すると右のようになります。

かかった時間 (分)	相対度数	
	A 組男子	B 組男子
以上　未満		
6.5 〜 7.0	0.10	0.05
7.0 〜 7.5	0.15	0.09
7.5 〜 8.0	0.20	0.18
8.0 〜 8.5	0.10	0.23
8.5 〜 9.0	0.10	0.14
9.0 〜 9.5	0.05	0.09
9.5 〜 10.0	0.10	0.09
10.0 〜 10.5	0.15	0.09
10.5 〜 11.0	0.05	0.05
11.0 〜 11.5	0.00	0.00
合計	1.00	1.00

1 年 A 組の男子と 1 年 B 組の男子の 50 m 走のタイムの相対度数分布

(4) (3) で作った表を見ると以下のようなことがわかります。

- A 組男子では「7.5 秒以上 8.0 秒未満」の生徒の割合が最も高く、B 組男子では「8.0 秒以上 8.5 秒未満」の生徒の割合が最も高い。
- 「6.5 秒以上 7.0 秒未満」、「7.0 秒以上 7.5 秒未満」、「7.5 秒以上 8.0 秒未満」の階級（つまり、結構速く走る生徒の入っている階級）の割合は、A 組男子のほうが B 組男子より大きい。
- 「9.5 秒以上 10.0 秒未満」、「11.0 秒以上 11.5 秒未満」階級（つまり、結構遅く走る生徒の入っている階級）の割合は、A 男子と B 組男子であまり変わらない。
- A 組男子も B 組男子も、最も相対度数が高い階級を中心にしてその階級からはなれていくにつれ、相対度数はだいたい減っていく。

本文へ戻る

問 10. Pさんが「自分のクラス（A 組）の男子の 50 m走のタイム」と「自分の中学校の 1 年男子全体の 50 m走のタイム」がどのように違うのか比べるため、データを分析してみる話でしたね。

(1) 1 年 A 組の男子 20 人の 50 m 走のタイムについて、それぞれの階級に対して相対度数を計算すると右のようになります。ただし、「相対度数」は少数第三位を四捨五入して少数第二位までの数値が書いてあります。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 〜 7.0	2	0.10
7.0 〜 7.5	3	0.15
7.5 〜 8.0	4	0.20
8.0 〜 8.5	2	0.10
8.5 〜 9.0	2	0.10
9.0 〜 9.5	1	0.05
9.5 〜 10.0	2	0.10
10.0 〜 10.5	3	0.15
10.5 〜 11.0	1	0.05
11.0 〜 11.5	0	0.00
合計	20	1.00

1 年 A 組の男子 20 人の 50 m 走のタイムの度数分布と相対度数分布

(2) この中学校の 1 年男子全体 90 人の 50 m 走のタイムについて、それぞれの階級に対して相対度数を計算すると右のようになります。ただし、「相対度数」は少数第三位を四捨五入して少数第二位までの数値が書いてあります。

タイム (秒)	人数 (人)	相対度数
以上　未満		
6.5 〜 7.0	1	0.01
7.0 〜 7.5	2	0.02
7.5 〜 8.0	12	0.13
8.0 〜 8.5	28	0.31
8.5 〜 9.0	19	0.21
9.0 〜 9.5	13	0.14
9.5 〜 10.0	10	0.11
10.0 〜 10.5	3	0.03
10.5 〜 11.0	1	0.01
11.0 〜 11.5	1	0.01
合計	90	1.00

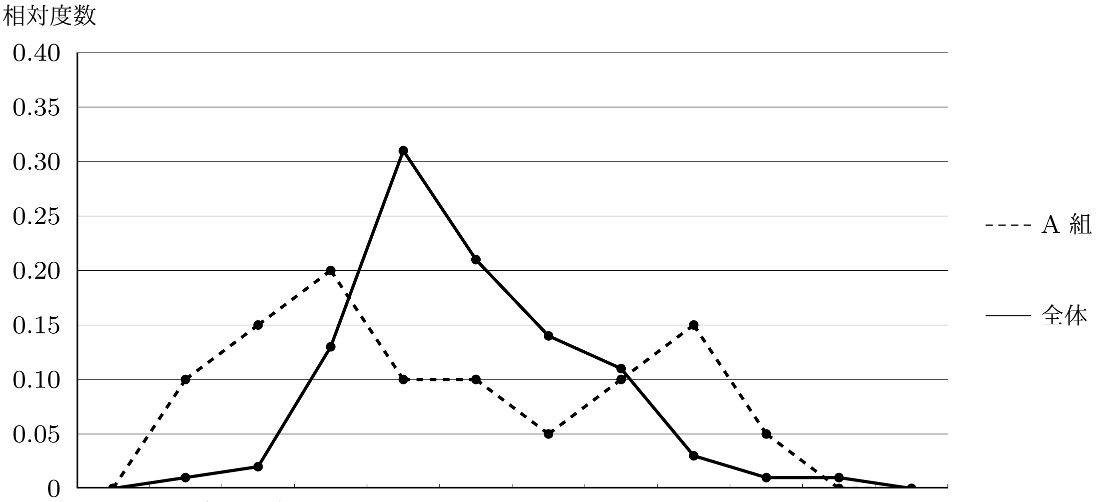
1 年男子全体 90 人の 50 m 走のタイムの度数分布と相対度数分布

(3) (1) で計算した A 組の相対度数と (2) で計算した全体の相対度数を記入すると右のようになります。

かかった時間 (分)	相対度数	
	A 組男子	男子全体
以上　未満		
6.5 〜 7.0	0.10	0.01
7.0 〜 7.5	0.15	0.02
7.5 〜 8.0	0.20	0.13
8.0 〜 8.5	0.10	0.31
8.5 〜 9.0	0.10	0.21
9.0 〜 9.5	0.05	0.14
9.5 〜 10.0	0.10	0.11
10.0 〜 10.5	0.15	0.03
10.5 〜 11.0	0.05	0.01
11.0 〜 11.5	0.00	0.01
合計		

1 年 A 組の男子と 1 年男子全体の 50 m 走のタイムの相対度数分布

(4) (3) で作った表を見て、「1 年 A 組の男子の 50 m 走のタイムの相対度数分布折れ線」と「1 年男子全体の 50 m 走のタイムの相対度数分布折れ線」を作ると次のようになります。



1 年 A 組の男子の 50 m 走のタイムの相対度数分布折れ線」と「1 年男子全体の 50 m 走のタイムの相対度数分布折れ線」

(5) (4) で作った相対度数分布折れ線を見ると以下のようなことがわかります。

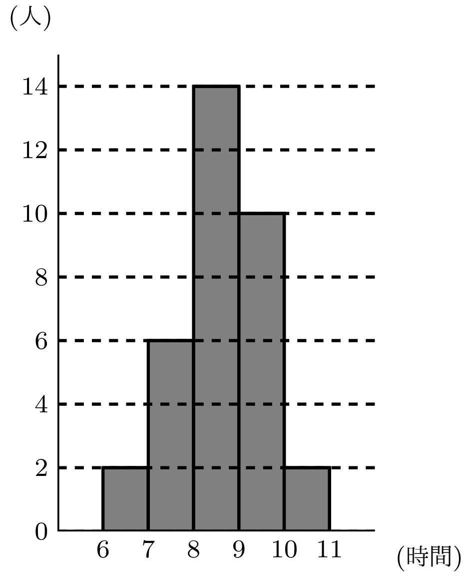
- A 組男子では「7.5 秒以上 8.0 秒未満」の生徒の割合が最も高く、男子全体では「8.0 秒以上 8.5 秒未満」の生徒の割合が最も高い。
- 「6.5 秒以上 7.0 秒未満」、「7.0 秒以上 7.5 秒未満」、「7.5 秒以上 8.0 秒未満」の階級（つまり、結構速く走る生徒の入っている階級）の割合は、A 組男子のほうが男子全体より大きい。
- 「8.0 秒以上 8.5 秒未満」、「8.5 秒以上 9.0 秒未満」、「9.0 秒以上 9.5 秒未満」の階級（つまり、まあまあ普通の速さで走る生徒の入っている階級）の割合は、A 組男子のほうが男子全体より小さい。
- 「10.0 秒以上 10.5 秒未満」、「10.5 秒以上 11.0 秒未満」階級（つまり、結構遅く走る生徒の入っている階級）の割合は、A 男子のほうが男子全体より大きい。
- A 組男子の相対度数分布は「7.5 秒以上 8.0 秒未満」の階級と「10.0 秒以上 10.5 秒未満」の階級のところに山ができている。それに対して、男子全体の相対度数分布は「8.0 秒以上 8.5 秒未満」のところに山が 1 つあるだけで、最も相対度数が高い階級を中心にしてその階級からはなれていくにつれ、相対度数は減っていく。

以上のことから、A 組男子を男子全体と比べると、「結構速く走る生徒の割合」や「結構遅く走る生徒の割合」が大きいが、「まあまあ普通に走る生徒の割合」が小さいということになりますね。

本文へ戻る

問 11. 次の図は、ある小学校の 6 年生のあるクラスの生徒達の睡眠時間を調査した結果をまとめた度数分布表とヒストグラムです。

睡眠時間 (時間)	人数 (人)
以上　未満	
6 ～ 7	2
7 ～ 8	6
8 ～ 9	14
9 ～ 10	10
10 ～ 11	2
合計	34



この度数分布表やヒストグラムからどんなことが読み取れるか考える問題でしたね。

(1) データは 6 時間から 11 時間まで分布しています。

(2) データの分布の中心はだいたい 8 時間 30 分ぐらいです。

(3) データの散らばり方の特徴

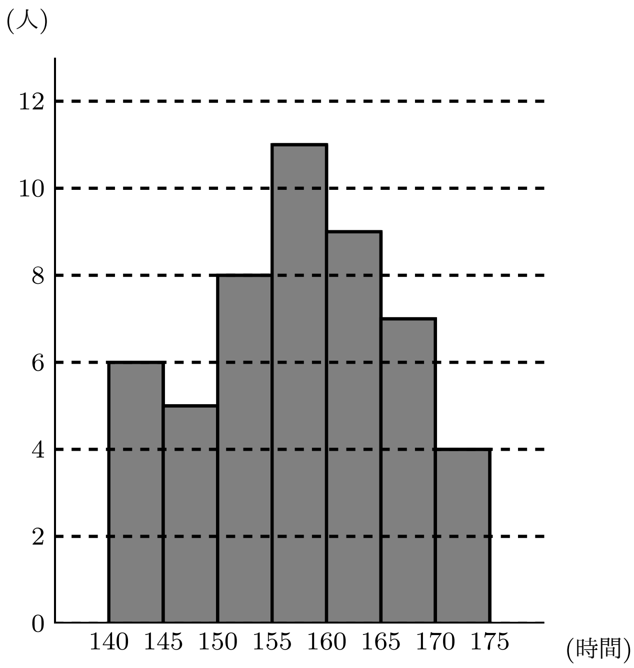
睡眠時間が短い人や長い人（つまり 7 時間未満の人や 10 時間以上の人）は少ないことがわかります。また 8 時間から 10 時間の人がとても多ことがわかります。

データの散らばり方はまあまあ左右対称と言って良いかもしれません。

本文へ戻る

問 12. 次の表は、ある中学校の 1 年生 50 人について身長を調査した結果をまとめた度数分布表とヒストグラムです。

身長 (cm)	人数 (人)
以上　未満	
140 ～ 145	6
145 ～ 150	5
150 ～ 155	8
155 ～ 160	11
160 ～ 165	9
165 ～ 170	7
170 ～ 175	4
合計	50



この度数分布表やヒストグラムからどんなことが読み取れるか考えてみる問題でしたね。

(1) データは 140 cm ぐらいから 175 cm ぐらいまで分布しています。

(2) データの分布の中心はだいたい 157.5 cm ぐらいです。

(3) データの散らばり方の特徴

「150 cm～160 cm の階級」の人数が最も多くなっています。

「150 cm～160 cm の階級」から身長が高くなればなるほどその階級に入っている人の人数は減っていきます。

「150 cm～160 cm の階級」から身長が低くなればなるほどその階級に入っている人の人数は一応減っていきます。ただし、身長が最も低い人の入っている階級（つまり、「140 cm～145 cm の階級」）に入っている人も意外に多くいます。

「140 cm～145 cm の階級」のところをあまり気にしないことにすると、データの散らばり方はまあまあ左右対称と言って良いかもしれません。

本文へ戻る

問 13. 2 つのチーム A と B のハンドボール投げの記録についての問題でしたね。

A チーム

選手番号	記録 (m)
No.1	29
No.2	28
No.3	28
No.4	29
No.5	27
No.6	25
No.7	31
No.8	29
No.9	27
No.10	27
No.11	28

B チーム

選手番号	記録 (m)
No.1	26
No.2	29
No.3	28
No.4	24
No.5	30
No.6	26
No.7	25
No.8	30
No.9	28
No.10	28
No.11	34

(1) A チームの「最高の記録」は 31 m で「最低の記録」は 25 m ですね。ということは、

$$\begin{aligned}\text{A チームの記録の範囲} &= 31 - 25 \\ &= 6\end{aligned}$$

ですね。

(2) B チームの「最高の記録」は 34 m で「最低の記録」は 24 m ですね。ということは、

$$\text{B チームの記録の範囲} = 34 - 24 = 10$$

ですね。

(3) 分布の範囲が広いのは B チームですね。

本文へ戻る

問 14. ある中学校の 1 年のあるクラスで 10 点満点の小テストをした結果についての問題でしたね。

生徒番号	点数	生徒番号	点数
No.1	3	No.11	4
No.2	6	No.12	5
No.3	6	No.13	6
No.4	3	No.14	4
No.5	3	No.15	5
No.6	4	No.16	6
No.7	3	No.17	3
No.8	5	No.18	6
No.9	6		
No.10	3		

(1) 小テストの点数の平均値を求るのでしたね。

まず右の表をよく見て、「得られたデータの数値の合計」を求めます。つまり点数の合計を求めるわけです。すると、

$$\begin{aligned}\text{得られたデータの数値の合計} &= 3+6+6+3+3+4+3+5+6+3\\ &\quad +4+5+6+4+5+6+3+6\\ &= 81\end{aligned}$$

となりますね。

次に、「得られたデータの数値の合計」を「データの個数」で割れば平均値が求められるわけです。生徒の人数は 18 人ですから「データの個数」は 18 です。ですから、

$$\text{平均値} = 81 \div 18 = 4.5$$

となりますね。

以上でこのクラスの小テストの平均点は 4.5 点であることがわかりました。

(2) 最高点は 6 点で、最低点は 3 点です。ですから、

$$\text{小テストの点数の分布の範囲} = 6 - 3 = 3$$

ですね。

本文へ戻る

問 15. ある中学校の 1 年生である A 君と B 君の過去 10 回の 50 m 走の記録についての問題でしたね。

A 君

データ番号	記録 (秒)
No.1	7.4
No.2	7.2
No.3	7.2
No.4	7.1
No.5	7.1
No.6	7.3
No.7	7.1
No.8	7.2
No.9	7.3
No.10	7.1

B 君

データ番号	記録 (秒)
No.1	7.2
No.2	7.3
No.3	7.5
No.4	7.4
No.5	7.2
No.6	6.9
No.7	7.0
No.8	7.2
No.9	7.1
No.10	7.2

(1) A 君のタイムの平均値を求めるのですね。

まず右の表をよく見て、「得られたデータの数値の合計」を求めます。つまり記録（秒）の合計を求めるわけです。すると、

$$\begin{aligned}\text{得られたデータの数値の合計} &= 7.4 + 7.2 + 7.2 + 7.1 + 7.1 \\ &\quad +7.3 + 7.1 + 7.2 + 7.3 + 7.1 \\ &= 72.0\end{aligned}$$

となりますね。

次に、「得られたデータの数値の合計」を「データの個数」で割れば平均値が求められるわけです。「データの個数」は 10 です。ですから、

$$\text{平均値} = 72 \div 10 = 7.2$$

となりますね。

以上で A 君のタイムの平均値は 7.2 秒であることがわかりました。

(2) B 君のタイムの平均値を求めるのですね。

まず右の表をよく見て、「得られたデータの数値の合計」を求めます。つまり記録（秒）の合計を求めるわけです。すると、

$$
\begin{aligned}
\text{得られたデータの数値の合計} &= 7.2+7.3+7.5+7.4+7.2\\
&\quad +6.9+7.0+7.2+7.1+7.2\\
&= 72.0
\end{aligned}
$$

となりますね。

次に、「得られたデータの数値の合計」を「データの個数」で割れば平均値が求められるわけです。「データの個数」は 10 です。ですから、

$$\text{平均値} = 72 \div 10 = 7.2$$

となりますね。

以上で B 君のタイムの平均値は 7.2 秒であることがわかりました。

(3) A 君の最も良い記録は 7.1 秒で最も悪い記録は 7.4 秒です。ですから、

$$\text{A 君の記録の分布の範囲} = 7.4 - 7.1 = 0.3$$

となります。

B 君の最も良い記録は 6.9（秒）で最も悪い記録は 7.5 秒です。ですから、

$$\text{A 君の記録の分布の範囲} = 7.5 - 6.9 = 0.6$$

となります。

分布の範囲が広いのは B 君ですね。

(4) この中学の 1 年生である C 君の 50 m 走のタイムの平均値は 7.0 秒なのでしたね。

また、A 君のタイムの平均値は 7.2 秒で B 君のタイムの平均値も 7.2 秒ということがわかったのでしたね。

平均値で見る限り、A 君も B 君も C 君に勝てる可能性はあまり無いようです。しかし、A 君と B 君のどちらかを選んで C 君と対決するんですよね。そこで、平均

値だけではなく、分布の範囲や最も良い記録をもう一度詳しく見てみましょう。

A 君の最も良い記録は 7.1 秒で最も悪い記録は 7.4 秒なので、分布の範囲が狭くなっています。ですから A 君を選ぶときっと、よくても 7.1 秒ぐらいで走ってしまい、B 君に勝てそうな感じがしません。

一方、B 君の最も良い記録は 6.9 秒で最も悪い記録は 7.5 秒なので、分布の範囲が広くなっています。つまり、良いときと悪いときの差が大きいわけです。ですから、調子が良い時は C 君の平均タイムより速く走れるわけです。

というわけで、C 君に勝てる可能性の高いのは B 君であると判断するのが良さそうです。

本文へ戻る

問 16. ある中学校の 1 年のあるクラス女子の立ち幅とびの記録をとった、右のデータの中央値を求めるのでしたね。

データ番号	記録 (cm)	データ番号	記録 (cm)
No.1	180	No.11	195
No.2	184	No.12	210
No.3	195	No.13	180
No.4	181	No.14	165
No.5	167	No.15	176
No.6	190		
No.7	132		
No.8	170		
No.9	195		
No.10	169		

まず、得られたデータを大きさの順に並べるのですよね。ではこの表を見ながら間違わないように慎重に並べることにしましょう。小さい値から並べても良いですし、大きい値から並べても良いわけですが、ここでは小さい値から並べてみます。すると、

132、165、167、169、170、176、180、180、181、184、190、195、195、195、210

となりますね。

次は大きさの順になったデータを見て、ちょうど真ん中に並んでいるるデータを読み取るのでしたね。今、データの数は 15 個なので、8 番目に並んでいるデータを探せばよいわけです。すると、180 が見つかりますね。ですから、女子の立ち幅とびの中央値（別名、

メジアン）は 180 m となるわけです。

本文へ戻る

問 17. ある中学校の 1 年のあるクラス男子の立ち幅とびの記録をとった、右のデータの中央値を求める問題でしたね。

データ番号	記録 (cm)	データ番号	記録 (cm)
No.1	166	No.11	204
No.2	177	No.12	187
No.3	200	No.13	188
No.4	144	No.14	163
No.5	171		
No.6	193		
No.7	178		
No.8	153		
No.9	175		
No.10	202		

まず、得られたデータを大きさの順に並べるのですよね。ではこの表を見ながら間違わないように慎重に並べることにしましょう。小さい値から並べても良いですし、大きい値から並べても良いわけですが、ここでは小さい値から並べてみます。すると、

144、153、163、166、171、175、177、178、187、188、193、200、202、204

となりますね。

次は大きさの順になったデータを見て、ちょうど真ん中に並んでいるるデータを読み取るのでしたね。でも今、データの数が 14 個なので、ちょうど真ん中にはデータがありません。このようなときは、「ちょうど真ん中」の前後にある 2 つのデータの値をたして 2 でわるのでしたね。「ちょうど真ん中」の前後にある 2 つのデータは 177 と 178 ですよね。というわけで、ソフトボール投げの中央値（別名、メジアン）は、

$$(177 + 178) \div 2 = 178.5\,(\mathrm{m})$$

となるわけです。

本文へ戻る

問 18. あるバス会社で、平日の午前 9 : 10 に A 町を出発する B 町行きのバスについて、過去 160 回の運行状況の記録を調べてみた右のような度数分布表について、かかる時間の最頻値を求める問題でしたね。

この度数分布表を見ると、度数の最も大きい階級は、

40 分以上〜45 分未満の階級

であることがわかります。(64 回より多いところはないですよね。）ということは、最頻値は、「40 分以上〜45 分未満」の階級の真ん中の値なので 42.5 分ということになりますね。

かかった時間 (分)	回数 (回)
以上　未満	
30 〜 35	3
35 〜 40	37
40 〜 45	64
45 〜 50	45
50 〜 55	11
合計	160

平日の午前 9 時 10 分に A 町を出発する B 町行きのバスが A 町から B 町へ行くのにかかった時間の度数分布表 (160 回分の調査)

本文へ戻る

問 19. 答えは c. です。

本文へ戻る

問 20. ある中学校の 1 年生 10 人の 1 ヶ月のおこづかいの金額について調査した次の結果についての問題でしたね。

生徒名 (仮名)	A	B	C	D	E	F	G	H	I	J
金額 (円)	800	1200	900	8200	800	1000	1100	1100	1500	800

(1) この 10 人の生徒のおこづかいの額の平均値を求めるのでしたね。

合計を求めると

$$\begin{aligned}\text{合計} &= 800 + 1200 + 900 + 8200 + 800 + 100 + 1100 + 1100 + 1500 + 800 \\ &= 17400\,\text{円}\end{aligned}$$

となります。また、全部で 10 人の人がいます。ですから、

$$\begin{aligned}\text{平均値} &= 17400 \div 10 \\ &= 1740\,\text{円}\end{aligned}$$

となりますね。

(2) この 10 人の生徒のおこづかいの額の中央値を求めるのでしたね。

データを小さい順に並べると、

800、800、800、900、1000、1100、1100、1200、1500、8200

となります。

データの個数は 10 個ですから、「ちょうど真ん中」にはデータがありません。こういう時は「ちょうど真ん中」の前後にある 2 つのデータの値をたして 2 でわるのでしたね。「ちょうど真ん中」の前後にある 2 つのデータは 1000 と 1100 ですよね。というわけで、おこづかいの金額の中央値（別名、メジアン）は、

$$(1000 + 1100) \div 2 = 1050\,\text{円}$$

となるわけです。

(3) 階級の幅を 100 円にして度数分布表を作り、おこづかいの額の最頻値を求めるのでしたね。

もともとの調査結果をよく見て度数分布表を作ると右のようになります。

この度数分布表を見ると、最も人数の多いのは「800 円以上 900 円未満の階級」ですね。ということは、「おこづかいの額の最頻値」は 850 円ですね。

（念のための注意ですが、最頻値は最も度数の大きい階級の真ん中の値ですね。）

金額（円）	人数
以上　　未満	
800 ～ 900	3
900 ～ 1000	1
1000 ～ 1100	1
1100 ～ 1200	2
1200 ～ 1300	1
1300 ～ 1400	0
1400 ～ 1500	0
1500 ～ 1600	1
1600 ～ 1700	0
···	0
···	0
···	0
8200 ～ 8300	1
合計	10

(4) 平均値は 1740 円、中央値は 1050 円、最頻値は 850 円になりましたね。そして A 君、B さん、I さん、E さんの考えについて、以下の質問に答えるのでしたね。

（a）　A 君は、「平均値は 1740 円かぁ。ということは、おこづかいが 1750 円以下の人とおこづかいが 1750 円以上の人が 5 人づついるってことだな。」と考えました。あなたは A 君の考えについてどう思いますか。

という質問でした。

どうも、A 君は、「平均より上にいる人の数」と「平均より下にいる人の数」は同じになると信じているようです。でも、この考えは正しいとは限りません。実際この問題でも、A 君の考えのようにはなっていません。たしか、データを小さい順に並べると、

800、800、800、900、1000、1100、1100、1200、1500、8200

となりましたよね。おこづかいが「平均値である 1750 円」以上の人は 1 人しかいませんね。

(b) B さんは、「中央値は 1050 円かぁ。ということは、おこづかいが 1050 円以下の人とおこづかいが 1050 円以上の人が 5 人づついるってことだね。」と考えました。あなたは B さんの考えについてどう思いますか。

という質問でした。

B 君は、「中央より上にいる人の数」と「中央より下にいる人の数」は同じになると考えています。この考えはもちろん正しいですよね。

(c) I さんは、「平均値は 1740 円かぁ。私のおこづかいは 1500 円だから、みんなに比べて私のおこづかいって少ないんだね。」と考えました。あなたは I さんの考えについてどう思いますか。

という質問でした。

どうも、I さんは、「平均より下である」ということは「みんなより少ない」という考えをしているようです。でも、この考えは正しいとは限りません。

実際この問題でも、I さんの考えは正しくありません。たしか、データを小さい順に並べると、

800、800、800、900、1000、1100、1100、1200、1500、8200

となりましたよね。「I さんのおこづかいである 1500 円」より多くおこづかいをもらっている人はたった 1 人しかいませんね。

(d) E さんは、「最頻値は 850 円かぁ。私のおこづかいは 800 円だから、みんなと比べても私のおこづかいってほんの少し少ないだけだね。」と考えました。あなたは E さんの考えについてどう思いますか。

という質問でした。

どうも、E さんは、「最頻値と同じぐらいである」ということは「だいたいみんなと同じ」という考えをしているようです。でも、この考えは正しいとは限

りません。しかし、この問題に限れば、まあまあ正しいと言っても良いかもしれません。「みんな」というのをどのように考えるのかによって違ってくるのです。たしか、データを小さい順に並べると、

800、800、800、900、1000、1100、1100、1200、1500、8200

となっていました。

800 円の人が最も多くいるので、この 3 人を多数派と考えて「みんな」だとすれば、E さんのおこづかいである 800 円は「みんなより少し少ない」どころか「みんなと同じ」ということになります。

しかし、8200 円という一人だけやたら多くおこづかいをもらっている人以外を「みんな」だとすれば、1100 円とか 1200 円とか 1500 円もらっている人も合わせて 4 人いるので、800 円というおこづかいは結構少ないと感じるかもしれません。

本文へ戻る