



第2学年 課題研究
小石川フィロソフィーⅡ

共 | 通 | テ | キ | ス | ト



都立小石川中等教育学校

目 次

1	ガイダンス	3
2	PPDAC サイクルの考え方	7
3	記述統計学：資料の整理，データの分析	9
4	科学レポートを書いてみよう	23
5	科学レポートを読んでみよう	27
6	PPDAC サイクルの実践編	28
7	統計グラフコンクールの作品を作成しよう	35
8	統計グラフコンクールの作品を発表しよう	36
9	確率論の基礎	39
10	推測統計学Ⅰ：推定	45
11	推測統計学Ⅱ：確率分布	51
12	推測統計学Ⅲ：検定（統計的仮説検定）	57
13	参 考	61
	EXCEL（表計算ソフト）を使ってみよう	
	いろいろなグラフを使い分けよう	
	統計グラフコンクール出品作品をつくろう！（平成30年度版）	
14	振り返り&自己評価（学年末）	86



ガイダンス～この1年で身につける力～

🐾 はじめに

昨年度の小石川フィロソフィーⅠでは、図書館の利用法や意見文の書き方、スピーチやディベートの仕方など、探究活動の土台となる読解力、表現力、思考力を身につけました。

今年度の小石川フィロソフィーⅡでは、それらの力を踏まえつつ、**自然科学**（物理学、化学、生物学、地学、数学など）の分野に目を向け、小石川フィロソフィーⅢから始まる**課題研究**の基礎を築きます。

◎ 「研究」にもさまざまな種類がある！

同じ映画を見ても、人によってその感想はさまざまです。それは、映像から切り取られる情報が人によって異なり、映像をそれぞれの立場・視点で部分的に見ているからです。

「研究」も、これとまったく同じことが言えます。日常で生じた疑問を解決するために研究を進めようとしたとき、どんな立場で研究するかによって、同じものであっても得られる結論や捉え方が異なってきます。

① 的研究

研究対象を数値で表し、その数値を処理することで研究を進めていく方法です。基本的には大量のデータを扱うことになり、**統計学**の知識が必要となります。自然科学では、何らかの法則を見つけることによって自然をコントロールすることが図られます。

医療の分野では、医薬品の効果に関して、観察や治療結果の統計学による比較を根拠とする「evidence-based medicine；EBM」という手法がとられることもあります。

② 的研究

研究対象を数値化せずに捉えて研究を進めていく方法です。インタビューによる調査などは、この研究の最たる例です。

また、犯罪捜査もこの研究の一種と捉えることができます。捜査の過程で得られたデータをもとに最も妥当なストーリーが組み立てられ、そのストーリーが信ずるに値するかどうかを裁判で争われるわけです。

小石川フィロソフィーⅡでは、上記の①、②の研究方法の分類のうち、①の方法について学んでいきます。特に①の研究の基本となる「統計学」について、その基本的な部分を中心に扱うことにします。

小学校で習ったいろいろなグラフの知識も踏まえつつ、それらを利用して大量のデータを処理したり、統計の意味や考え方を正確に学んだりすることで、さまざまな視点で疑問を解決できるようにしましょう！！

◎ どんな力をつけるか？

🐾 データやグラフに騙されるな！

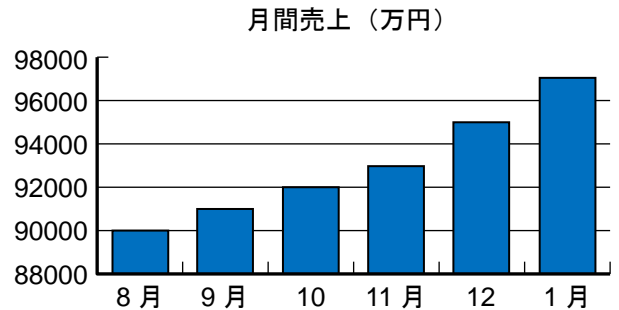
統計学は、大量のデータを扱うのにとっても便利ですが、その扱い方を誤るととんでもない誤解を生んだり、受ける印象が異なったりしてしまいます。

テーマ1 グラフを正しく読み取ろう！～棒グラフ編～

右の棒グラフは、あるスーパーマーケットにおけるあるお菓子の売り上げを、ここ6ヵ月分調査したものです。

このグラフを見て、このお菓子メーカーの担当者は、「このお菓子の人気はこの半年で急上昇している！」と言っています。

この主張に対する反論を考えてください。



テーマ2 グラフを正しく読み取ろう！～円グラフ編～

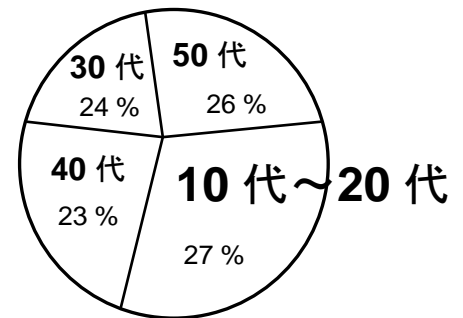
ある化粧品メーカーが、新発売の化粧品について「購買者の年齢層」を示した右のような円グラフをつくって、

「この商品は10代～20代の若者に大人気です」と言っています。

この主張に対する反論を考えてください。

この主張に対する反論を考えてください。

(このグラフは、とある報道番組で実際に使われたものをアレンジしたものです)



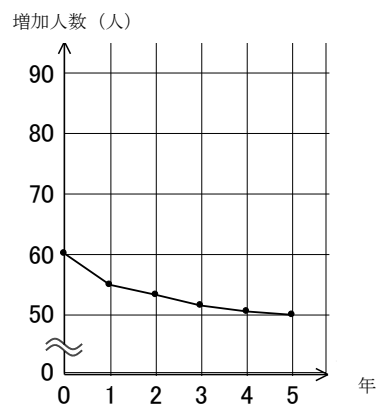
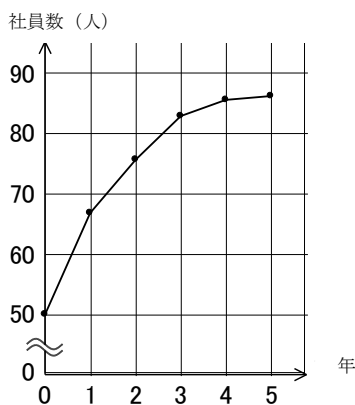
購買者の年齢層

テーマ3 グラフを正しく読み取ろう！～折れ線グラフ編～

下の2つのグラフは、A社、B社のここ5年間の社員数の変化を示したグラフです。

<A社>

<B社>



このグラフを見て、A社の社長は

「わが社の社員数はB社と比べて順調に伸びている！」

と言っています。

この主張に対する反論を考えてください。

テーマ 4 データの扱いに気をつけよう！①

生徒数 830 人の学校で、今年度、インフルエンザにかかった生徒は 43 人もいました。その 43 人を調べたところ、約半数の 22 人が予防注射を受けていて、残りの 21 人は受けていないことがわかりました。

予防注射を受けた人の方が、ほんの少しですが、受けなかった人より多いという結果が出てしまいました。この結果を見て、ある人が

「今年度は予防注射が効かなかったんだねえ」

と言っています。

この主張に対する反論を考えてください。

テーマ 5 データの扱いに気をつけよう！②

ある中学校の 2 年生 20 人が 2 つのグループに分かれて、数学の試験に挑戦しました。それぞれのグループの点数は次の通りです。

グループ① 73, 85, 86, 72, 84, 96, 75, 66, 73, 85

グループ② 82, 85, 81, 72, 84, 96, 90, 11, 85, 92

この結果を見て、ある人が

「グループ①の方が、平均点が高いから、グループ①の方が数学の力がある」

と言っています。

この主張に対する反論を考えてください。

テーマ 6 データの扱いに気をつけよう！③

昔、ある国で貨物自動車を、年間台数にして 1000 台増産する計画を立て、5 年後に、生産台数にしてわずかに 200 台しか増産できませんでした。ところが、その国の発表を見ると、

「はじめの目標の 80 % を達成することができました」

とあります。

1000 台のところを 200 台しか増産できなかったのならば、達成率は

$$200 \div 1000 = 0.2 = 20 \%$$

のはずです。

国の発表は正しいとして、達成率 80 % の謎を説明してみてください。

選んだテーマ：

話し合いメモ

◎ 各班の発表

班	テーマ	発表内容
1班		
2班		
3班		
4班		
5班		
6班		
7班		
8班		

次回以降の持ち物

- 筆記用具（色ペンなどが何種類かあるとよい）
- A4 ファイル（フィロソフィー用）
- 「生徒のための統計活用～基礎編～」
- 電卓（あった方が便利）



PPDAC サイクルの考え方

2年 組 番氏名

PPDAC サイクルとは

「PPDAC サイクル」という言葉を聞いたことはあるでしょうか？

統計的探究を行うための枠組みの一つで、問題解決における各段階を Problem (問題), Plan (調査の計画), Data (データ), Analysis (分析), Conclusion (結論) に分割した考え方です。

この「PPDAC サイクル」を回すことで、段階を追って問題解決に近づくことができます。各段階を詳しく見ていきましょう。

STEP 1

Problem 問題

テーマを設定し、具体的に問題をとらえます。

STEP 2

Plan 計画

明らかにしたい仮説を決めます。
必要なデータの収集計画を立てます。

STEP 3

Data 収集

必要なデータ・統計資料を集めます。

STEP 4

Analysis 分析

表やグラフを作成し、代表値（平均値など）を計算して、統計資料をまとめます。全体の傾向や関連性、因果関係を分析します。

STEP 5

Conclusion 結論

分析の結果をまとめます。

※得られた結論が不十分だった場合、もう一度 STEP 1 に戻って PPDAC サイクルを回します。

では、PPDAC サイクルを回して問題解決にあたる例を見てみましょう。

【例】

太郎くんの学校では、毎年体育祭でクラス対抗短距離リレーが行われます。この種目は配点も高く、1位になれば他のクラスに圧倒的な差をつけられます。

太郎くんが所属しているB組は、優勝候補と言われているA組に勝つため、作戦を考えようとしています。体育祭実行委員である太郎くんは、PPDACサイクルを使って作戦を立てることにしました。

◎ クラス対抗短距離リレーのルール

各クラスとも、決められた人数の代表を決め、その代表者のみが走る。

STEP 1

Problem 問題

知りたい問題を決めよう

太郎くんは、体育祭実行委員として、リレーの走順を提案することができます。太郎くんは自分のクラスが有利になるように、リレーの走順を提案しようと考えています。

STEP 2

Plan 計画

どのようなデータを集めるか考えよう

体育の授業で、短距離走（50 m 走）のタイムを測ったので、A組、B組の分を集めることができればいいのではないかと考えました。

STEP 3

Data 収集

必要なデータ・統計資料を集めよう

体育の先生にきいたところ、A組、B組の全員の50 m 走の記録をもらうことができました。

STEP 4

Analysis 分析

グラフや表、統計量で傾向をとらえよう

ここからがメインパートです。これから学ぶ平均値や中央値・最頻値などの代表値や標準偏差などを算出し、グラフや表を作成して勝つための戦略を分析します。

STEP 5

Conclusion 結論

わかったことをまとめ・読み取ろう

これまで調べたことをもとに、太郎くんが所属しているB組の立場から、リレーの走順を決めます。

◎ この続きは……

PPDACサイクルを1回まわしても、目的に合った結論が得られない場合、さらに視点を変えて別の問題（Problem）を考えることにより、PPDACサイクルをもう1回（あるいはもう2回）まわします。

3

記述統計学：資料の整理，データの分析

右の表は，A組とB組の50m走のタイムです。A組とB組のタイムの傾向を分析するとき，この表では分析しにくいと感じるのではないのでしょうか。わかりやすく整理することを考えましょう。

(1) 表で整理してみよう

まず，「何秒以上何秒未満が何人いるか」が一目でわかる下のような表をつくってみました。

タイム(秒)		人数(人)		階級値(秒)
		A組	B組	
以上	未満			
6.5	～ 7.5	9		
7.5	～ 8.5	15		
8.5	～ 9.5	6		
合 計		30	30	

表 A組, B組の50m走タイム

番号	A組	B組
1	6.5	6.7
2	6.6	6.8
3	6.9	6.8
4	7.0	6.8
5	7.0	7.1
6	7.2	7.2
7	7.3	7.2
8	7.4	7.3
9	7.4	7.3
10	7.5	7.5
11	7.5	7.6
12	7.5	7.7
13	7.6	7.8
14	7.7	7.9
15	7.7	7.9
16	7.8	8.3
17	7.8	8.4
18	7.9	8.4
19	8.0	8.4
20	8.1	8.5
21	8.1	8.6
22	8.2	8.9
23	8.2	8.9
24	8.3	8.9
25	8.5	9.0
26	8.5	9.1
27	8.6	9.1
28	8.6	9.4
29	9.0	9.4
30	9.3	9.4

◎ 用語をおさえておきましょう！

①

上のように，ある区間を設定して資料（データ）の個数を数えた表

②

区切られた各区間のこと。区間の幅のことを，と

いい，区間の中央の値のことを，という。

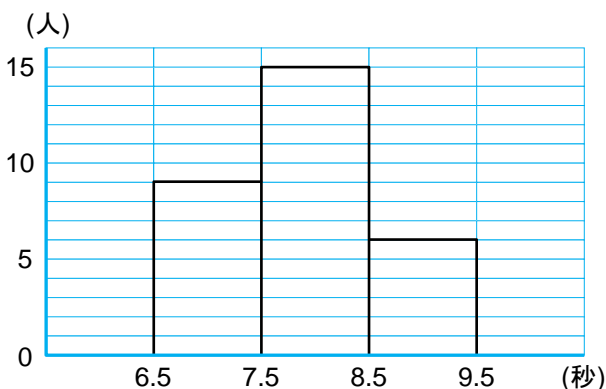
③

各階級に含まれる資料（データ）の個数。上の表の場合は人数。

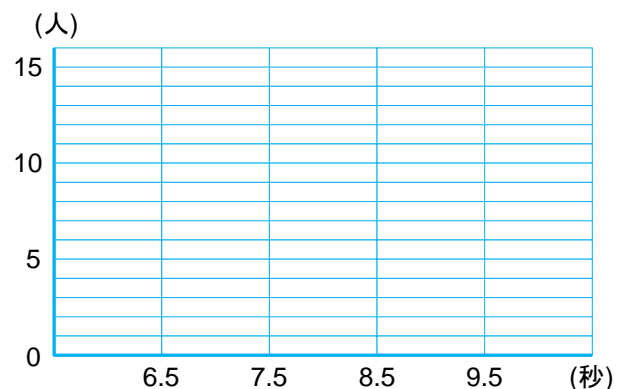
(2) グラフで整理してみよう

上の度数分布表を，グラフ化してみます。下のようなグラフを，**ヒストグラム（柱状グラフ）**といいます。

A組の短距離走のヒストグラム



B組の短距離走のヒストグラム

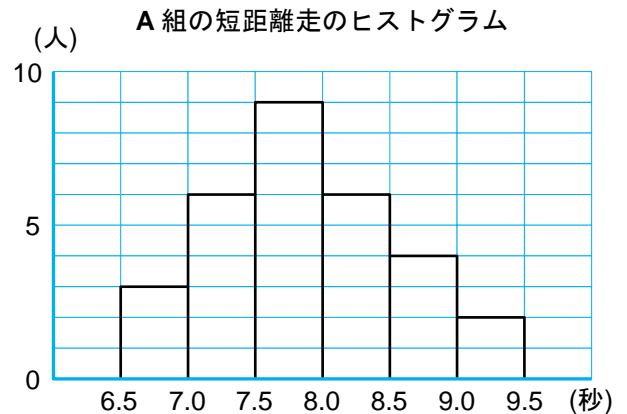


ヒストグラムは，資料全体のような山を山の形で視覚的に捉えることができるので便利です。また，山の形がはっきりするように，ヒストグラムの各長方形の上の辺の中点を結んでできる折れ線グラフをつくることもあります。この折れ線グラフを といいます。

(3) 階級の幅を変えてみよう

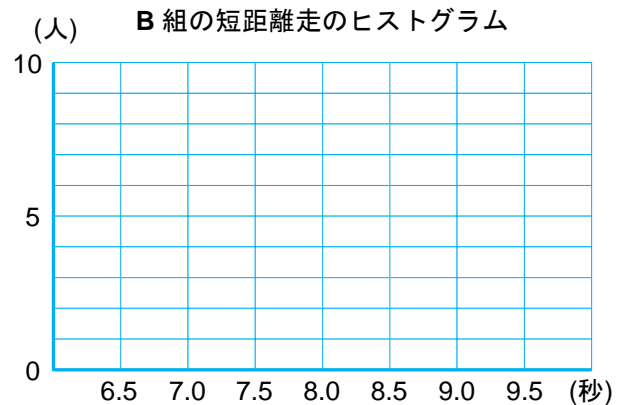
もう少し詳しく分析してみます。さきほどは階級の幅を 1.0 秒としましたが、今度は 0.5 秒として度数分布表とヒストグラムをつくってみましょう。

タイム(秒)	人数(人)		階級値(秒)
	A組	B組	
以上 未満			
6.5 ~ 7.0	3		
7.0 ~ 7.5	6		
7.5 ~ 8.0	9		
8.0 ~ 8.5	6		
8.5 ~ 9.0	4		
9.0 ~ 9.5	2		
合計	30	30	



◎ どんなことがわかるでしょうか？

山の形などを見て、A組、B組の走力にどのような特徴があるか話し合ってみましょう。また、階級の幅を変えることによって、何が変わったのでしょうか。



(4) 他の分析方法はないか考えよう

資料全体の特徴を 1 つの数値で表すことを考えてみましょう。そのような値を、資料の**代表値**といいます。

①

資料の値の合計を、資料の個数で割った値。

資料が度数分布表にまとめられている場合は、各階級に含まれる資料の個々の値がわからないので、ある階級に含まれる資料は、すべてその階級の階級値をとるものと考えて平均値を求める。

$$\frac{\{(\text{階級値}) \times (\text{度数})\} \text{の合計}}{\text{度数の合計}}$$

② (メジアン)

資料を大きさの順に並べたとき、その中央の順位にくる値。資料の個数が偶数のときは、中央に 2 つの値が並ぶから、その 2 つの値の平均値を中央値とする。

③ (モード)

資料において、最も個数の多い値。

資料が度数分布表に整理されているときは、度数が最も大きい階級の階級値を最頻値とする。複数ある場合はそのすべてを最頻値とする。

代表値としてよく用いられるのは①～③の 3 種類ですが、この他にも資料全体の傾向をつかむ量としていくつかの値があります。

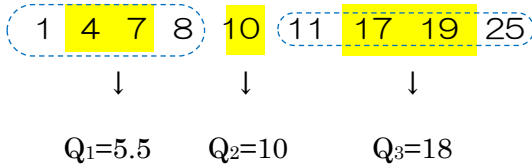
④

資料のとり値のうち、最大値から最小値をひいた差。

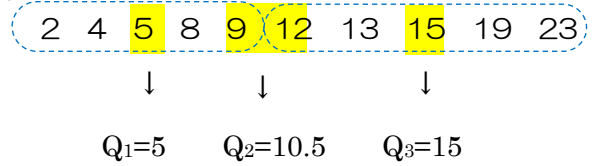
⑤

資料を値の大きさの順に並べたとき、4等分する位置にくる値。四分位数は、小さい方から順に第1四分位数、第2四分位数、第3四分位数といい、順に Q_1 、 Q_2 、 Q_3 で表す。

<例1>



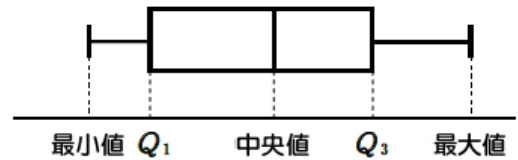
<例2>



これらの例を見てわかるように、 Q_2 は資料全体の中央値、 Q_1 は資料の下半分の中央値、 Q_3 は資料の上半分の中央値です。

(5) 箱ひげ図のつくり方を覚えよう

右の図のように、資料の最小値、第1四分位数、中央値、第3四分位数、最大値の5つの値を箱と線（ひげ）で表したのが箱ひげ図です。



箱ひげ図は、縦に表示することもあります。

(6) A組とB組のタイムの記録について、代表値などを使ってもう少し深く分析してみましょう。

タイム(秒)	人数(人)		階級値(秒)	(階級値)×(度数)	
	A組	B組		A組	B組
以上 未満					
6.5 ~ 7.0	3				
7.0 ~ 7.5	6				
7.5 ~ 8.0	9				
8.0 ~ 8.5	6				
8.5 ~ 9.0	4				
9.0 ~ 9.5	2				
合計	30	30			

① 平均値

○ 資料から求めた平均値……………A組：, B組：

○ 上の度数分布表から求めた平均値……A組：, B組：

② 中央値

A組：, B組：

③ 最頻値

上の度数分布表より、

A組：, B組：

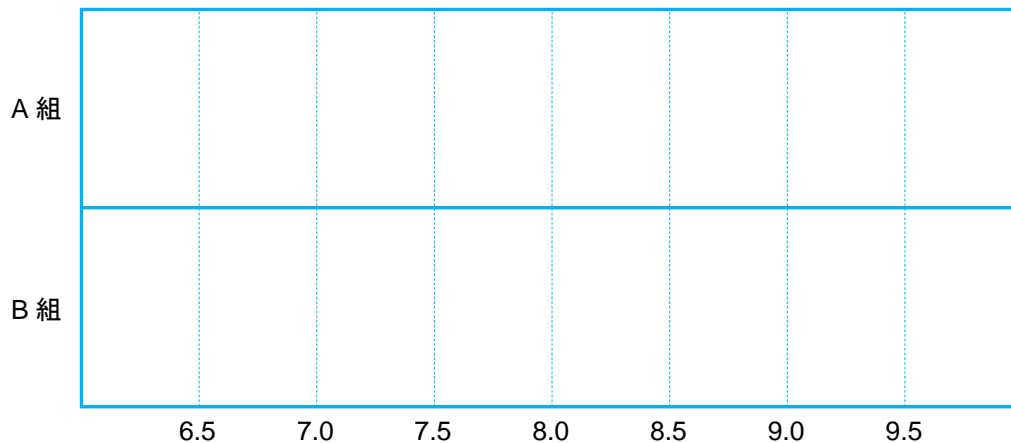
④ 第1四分位数

A組: , B組:

⑤ 第3四分位数

A組: , B組:

⑥ 箱ひげ図



このように、箱ひげ図は、ヒストグラムほどにはデータの分布が詳しく表現されませんが、大まかな様子はわかります。

また、複数の資料の分布を比較したいとき、箱ひげ図は大変便利です。

(7) 戦略を練るために分析の仕方をいろいろと変えてみよう

太郎くんは、どうすればA組に勝てるかという戦略を決めるために、もう少しいろいろな視点で考えることにしました。

<考えるポイント>

- 階級の幅をもっと小さくしてみたらどうだろう?
- 男女別に整理してみたらどうなるだろう?
- 男女別に代表値を求めてみたらどうなるだろう?

などなど……

Conclusion 結論

わかったことをまとめ・読み取ろう

これまで調べたことをもとに、B組の立場から、リレーに勝つ戦略を決めよう。

- ◎ リレーに勝つためには、代表を 人にする!
- ◎ 戦略は……

🐾 度数の違いや累積を比較するには…

次のような例を比較する方法を考えてみましょう。

例1 どちらの病院にする？

右の表は、A病院とB病院の患者の待ち時間について、1週間調べた結果をまとめた度数分布表です。

待ち時間が少ない病院に行きたいと考えているとき、どちらの病院に行くかによいでしょうか。理由も含めて考えてみましょう。

待ち時間(分)		人数(人)	
		A病院	B病院
以上	未満		
0	～ 10	16	8
10	～ 20	24	38
20	～ 30	40	55
30	～ 40	91	9
40	～ 50	53	26
50	～ 60	16	12
60	～ 70	9	9
70	～ 80	1	3
合 計		250	200

◎ 用語をおさえておきましょう！

①

各階級の度数を度数の合計で割った値

②

最も小さい階級から各階級までの度数の合計、最も小さい階級から各階級までの相対度数の合計を

という。

(1) 相対度数を求めて表にしてみましょう。

待ち時間(分)	A病院		B病院	
	度数(人)	相対度数	度数(人)	相対度数
以上	未満			
0	～ 10	16	0.064	8
10	～ 20	24		38
20	～ 30	40		55
30	～ 40	91		49
40	～ 50	53		26
50	～ 60	16		12
60	～ 70	9		9
70	～ 80	1		3
合 計		250		200

(2) 累積度数、累積相対度数を求めて表にしてみましょう。

待ち時間(分)	A 病院		B 病院	
	累積度数	累積相対度数	累積度数	累積相対度数
以上 未満				
0 ~ 10	16	0.064		
10 ~ 20	40			
20 ~ 30	80			
30 ~ 40	171			
40 ~ 50	224			
50 ~ 60	240			
60 ~ 70	249			
70 ~ 80	250			

(3) ここからわかることをまとめてみましょう。

累積度数分布表は、全体に対する上側および下側の部分の割合を調べるときに便利です。

累積相対度数を表す折れ線グラフ

右の図は、A 病院の待ち時間の累積相対度数をグラフに表したものです。

グラフを見ると、例えば、全体の 40% の患者の待ち時間は、累積相対度数の 0.4 を見るとおよそ 32 分未満であることが読み取れます。

(累積相対度数)

1.0

0.4

0

0 10 20 30 40 50 60 70 80 (分)

🐾 分散と標準偏差があらわすものは？

資料（データ）を整理する方法として、度数分布表やヒストグラム、代表値などを学びました。代表値とは、そのデータの傾向や性質を示す数値のことで、平均値、中央値、最頻値を取り扱いましたね。

今回は、データをより精密に分析するために、これまで学んだもの以外にどんな統計量があるかを探っていきます。

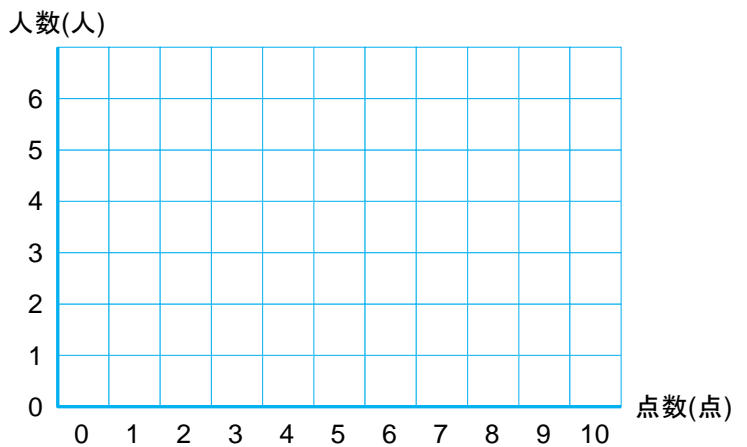
その前に、復習がてら、次の問題を考えてみましょう。

問題 1

10 人の生徒が 10 点満点のテストを受けて、次の表のような結果となりました。

点数(点)	0	1	2	3	4	5	6	7	8	9	10
人数(人)	3	3	1	0	0	0	0	0	0	0	3

(1) この結果を、右の図にヒストグラムとして表してください。ただし、ここでは階級の幅を 0 点と考えるので見た目は棒グラフのようになります。



(2) このデータの代表値としてふさわしいものは、平均値、中央値、最頻値のいずれでしょうか？

それぞれ求めた上で考えましょう。

① 平均値…… 点

② 中央値…… 点

③ 最頻値…… 点

このデータの代表値としてふさわしいものは、 であると考えます。

◎その理由は……？

さて、本題です。問題 1 を踏まえて、次の問題 2 を考えてください。

問題 2

A 組, B 組で, それぞれ 10 人ずつの生徒が, 10 点満点のテストを受けて, 下の表のような結果となりました。

<A 組>

点数(点)	0	1	2	3	4	5	6	7	8	9	10
人数(人)	1	0	0	0	2	4	2	0	0	0	1

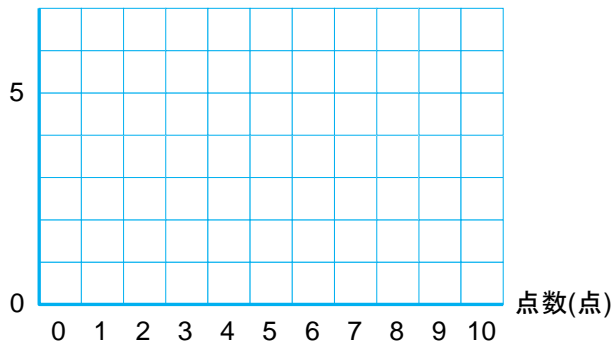
<B 組>

点数(点)	0	1	2	3	4	5	6	7	8	9	10
人数(人)	2	1	0	0	0	4	0	0	0	1	2

(1) それぞれのクラスのヒストグラムをつくってください。

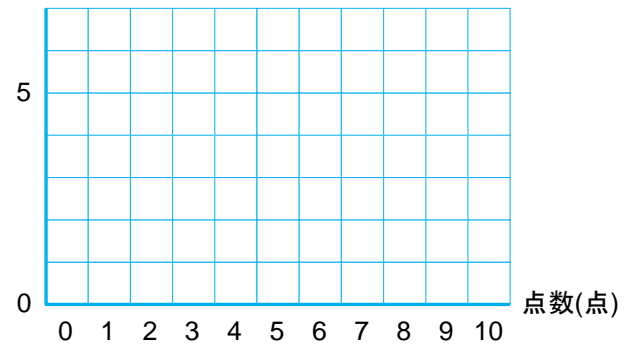
<A 組>

人数(人)



<B 組>

人数(人)



(2) それぞれのクラスにおける平均値, 中央値, 最頻値を求めてください。ついでに, 範囲も求めてみましょう。

<A 組>

○ 平均値…… 点

○ 中央値…… 点

○ 最頻値…… 点

○ 範囲…… 点

<B 組>

○ 平均値…… 点

○ 中央値…… 点

○ 最頻値…… 点

○ 範囲…… 点

(3) この 2 つのデータが区別できるような統計量を考えたいと思います。そこで, 2 つのデータからつくったヒストグラムを見て, 特徴の違いを挙げてみてください。

(4) (3)で調べたことを数値化するために、「個々の点数が平均値（平均点）からどのくらい離れているか」を調べてみます。

データに含まれている1つ1つの数値から平均値を引いた値を、それぞれの数値の といいます。

A組、B組のそれぞれについて、次の表の空欄を埋めてください。

<A組>

点数	0	1	2	3	4	5	6	7	8	9	10
偏差											
人数	1	0	0	0	2	4	2	0	0	0	1
偏差×人数											

<B組>

点数	0	1	2	3	4	5	6	7	8	9	10
偏差											
人数	2	1	0	0	0	4	0	0	0	1	2
偏差×人数											

したがって、「偏差×人数」の合計は、

A組…… , B組……

となります。

◎以上からわかること。

(5) (4)の問題点を克服するために、偏差の絶対値を考えてみます。次の表の空欄を埋めてください。

<A組>

点数	0	1	2	3	4	5	6	7	8	9	10
偏差の絶対値											
人数	1	0	0	0	2	4	2	0	0	0	1
偏差の絶対値×人数											

<B組>

点数	0	1	2	3	4	5	6	7	8	9	10
偏差の絶対値											
人数	2	1	0	0	0	4	0	0	0	1	2
偏差の絶対値×人数											

したがって、「偏差の絶対値×人数」の合計は、

A組…… , B組……

となって、これならA組とB組の区別ができそうです。

(6) 「偏差の絶対値」を考える代わりに、「偏差の2乗」を考えることもできます。次の表の空欄を埋めてください。

<A組>

点数	0	1	2	3	4	5	6	7	8	9	10
偏差の2乗											
人数	1	0	0	0	2	4	2	0	0	0	1
偏差の2乗×人数											

<B組>

点数	0	1	2	3	4	5	6	7	8	9	10
偏差の2乗											
人数	2	1	0	0	0	4	0	0	0	1	2
偏差の2乗×人数											

したがって、「偏差の2乗×人数」の合計は、

A組…… , B組……

となります。

ここで、上で求めた「偏差の2乗×人数」の合計を、人数でわってみます。つまり、「偏差の2乗の平均」を求めるといことです。

A組…… , B組……

いま求めた、「偏差の2乗の平均」を といいます。

◎どんな統計量？

(7) A組, B組のそれぞれについて, (6)で計算した統計量のルートをとってみましょう。

A組…… , B組……

このように, 分散の平方根のうち, 負でない方を といいます。

◎どんな統計量？

計算の練習をしておこう

これまで学習したことを、文字を使ってまとめておきます。

◎まとめ

n 個の数値があるとします。この n 個の数値のまとまりを**データ**と呼びます。データに含まれている n 個の数値を、

$$x_1, x_2, \dots, x_n$$

と表すことにします。

平均 \bar{x}

このとき、

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

を、このデータの**平均**と呼びます。

分散 s^2

データ x_1, x_2, \dots, x_n の平均を \bar{x} で表すことにします。このとき、数値 x_1 と平均 \bar{x} の差、つまり、

$$x_1 - \bar{x}$$

を、 x_1 の**偏差**と呼びます。 x_1 と同様に、 x_1, x_2, \dots, x_n の偏差をそれぞれ考えることができます。

x_1, x_2, \dots, x_n の偏差をそれぞれ2乗した値の平均を、**分散**と呼び、記号 s^2 で表します。すなわち、

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

です。

標準偏差 s

分散の平方根のうち、負でない方を**標準偏差**といい、記号 s で表します。すなわち、

$$s = \sqrt{s^2}$$

です。

では、これらを踏まえて、計算の練習をしてみましょう。

問題 3

5人の小テストの得点のデータが、次のように与えられています。

$$5 \quad 7 \quad 5 \quad 10 \quad 8$$

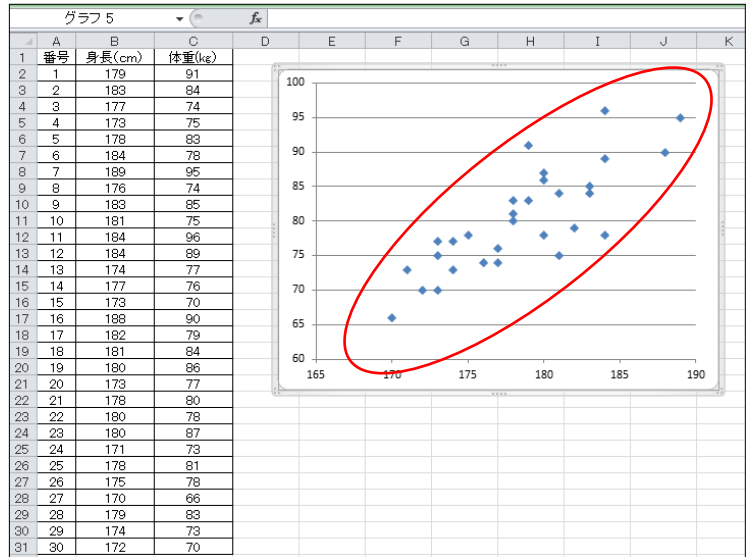
このデータの分散 s^2 と標準偏差 s を求めてください。

🐾 散布図と相関係数

身長と体重の関係や、国語と英語の成績の関係など、2つの量の間に関係があるかないかを調べたいときがあります。

(1) 散布図

右の図の表は、30人の身長と体重の一覧です。身長と体重の間に何らかの関係があるかを調べたいとき、右のように、横軸に身長、縦軸に体重を取り、身長、体重の値の組を座標とする点を平面上にとった図を作ることができます。



この図を、**散布図**といいます。散布図の上にとどのように点が配置されるかによって、2つの量の間を把握することができます。

右の図では、赤丸で囲ったところを見てわかるように、点の集まりが左下から右上に向かって伸びている様子がうかがえます。

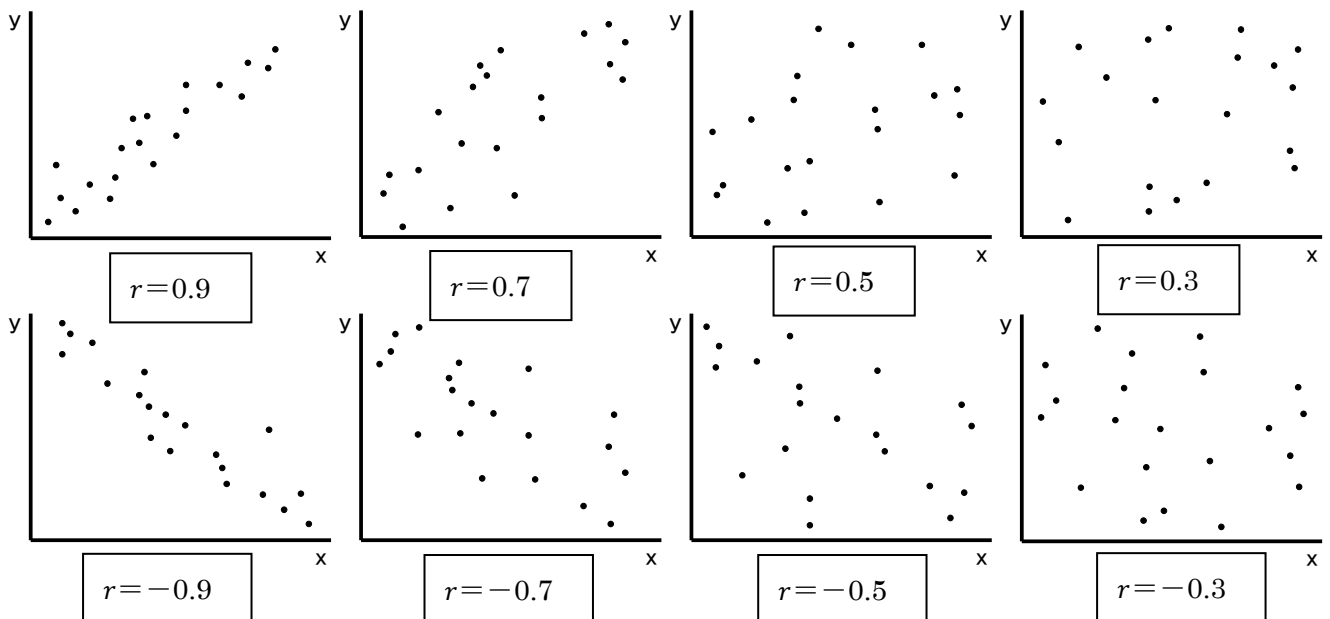
このように、一方が増えると他方も増える傾向が認められるとき、2つの量の間**正の相関**があるとといいます。逆に、一方が増える

と他方が減る傾向が認められるとき、2つの量の間**負の相関**があるとといいます。どちらの傾向も認められない場合（つまり、点の集まりがバラバラの場合）は、**相関がない**とといいます。

(2) 相関係数

相関の強さを測る指標として、**相関係数**というものがあります。相関係数 r には、次のような性質があります。

- ① r は-1から1までの値をとる。つまり、 $-1 \leq r \leq 1$
- ② r の値が**1**に近いとき、**強い正の相関関係**がある。このとき、散布図の点は右上がりの直線に沿って分布する傾向が強くなる。
- ③ r の値が**-1**に近いとき、**強い負の相関関係**がある。このとき、散布図の点は右下がりの直線に沿って分布する傾向が強くなる。
- ④ r の値が0に近いとき、直線的な相関はない。



(3) Excelを使って相関係数を求めてみよう

相関係数は、手で計算するとかかなり大変です。高校で学ぶ記号を使って書くと、

$$r = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^n (x_k - \bar{x})^2} \sqrt{\sum_{k=1}^n (y_k - \bar{y})^2}}$$

となります (x_k やら \bar{x} の記号の意味はともかくとして、これを計算するのが大変なのはわかりますね)。Excelを使うと、この計算が命令1つでできてしまうわけです。

さきほどの身長と体重について、その相関係数を計算してみます。

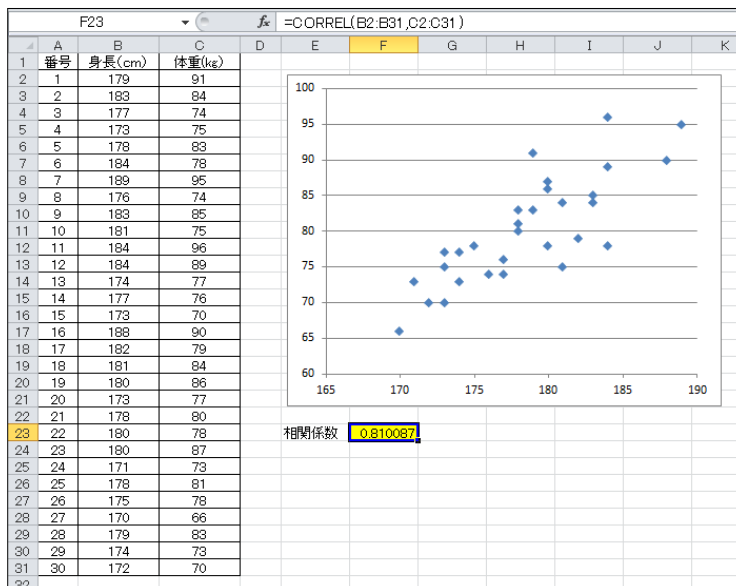
右の図は、F23セルに身長と体重の相関係数を出力したものです。

F23セルには

「=CORREL(B2:B31,C2:C31)」

と数式が入っていますが、これは、B2からB31までの値と、C2からC31までの値の相関係数を求めなさい、と指示する関数になっています。

この例では、相関係数が約0.8なので、「強い相関がある」と言えます。



(4) 相関を扱うときの注意～相関関係と因果関係の違いを見極めよう！～

相関関係と似た言葉に**因果関係**というものがあります。この違いをきちんとおさえておかないと、思わぬところでミスをします。

「相関関係」とは、**2つのことからのうち、一方が変わると、もう一方も変化する**という関係です。一方が増えたとき、もう一方も増えるなら「正の相関」、一方が増えたときもう一方が減るなら「負の相関」、どちらでもなければ「相関がない」と呼ばれるのでした。

これに対して、「因果関係」とは、2つのことからの一方が**原因**で、他方が**結果**となる関係をいいます。つまり、「AだからBである」という関係です。

相関係数が1に近いとき、散布図の点は右上がりの直線の上に多くのっていることのみを表し、因果関係が強いことを示しているわけではありません。

例1 「イケメン度」と「バレンタインデーにもらったチョコの数」との関係

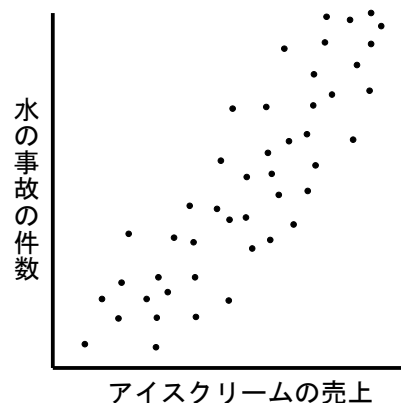
イケメン度が高い人は、より多くのチョコをもらう傾向にあり、相関関係があるといえます。また、イケメンだから、もらったチョコが多いという因果関係も成り立つでしょう。

例2 「アイスクリームの売り上げ」と「水の事故の件数」との関係

ある調査の結果、これらの間に右の図のような相関関係が見られたとします。ここで勘違いするのが、『アイスクリームがよく売れる日は水の事故が起きやすいから気をつけよう』と考えてしまうことです。

アイスクリームの売上と水の事故の件数の間には、何の因果関係もありません。

このように、因果関係がない相関関係のことを**疑似相関**（または**偽相関**）といいます。



◎ 疑似相関では、何が見落とされている！？

疑似相関があったとき、その要因として考えられるのは2つパターンがあります。

① まったくの偶然

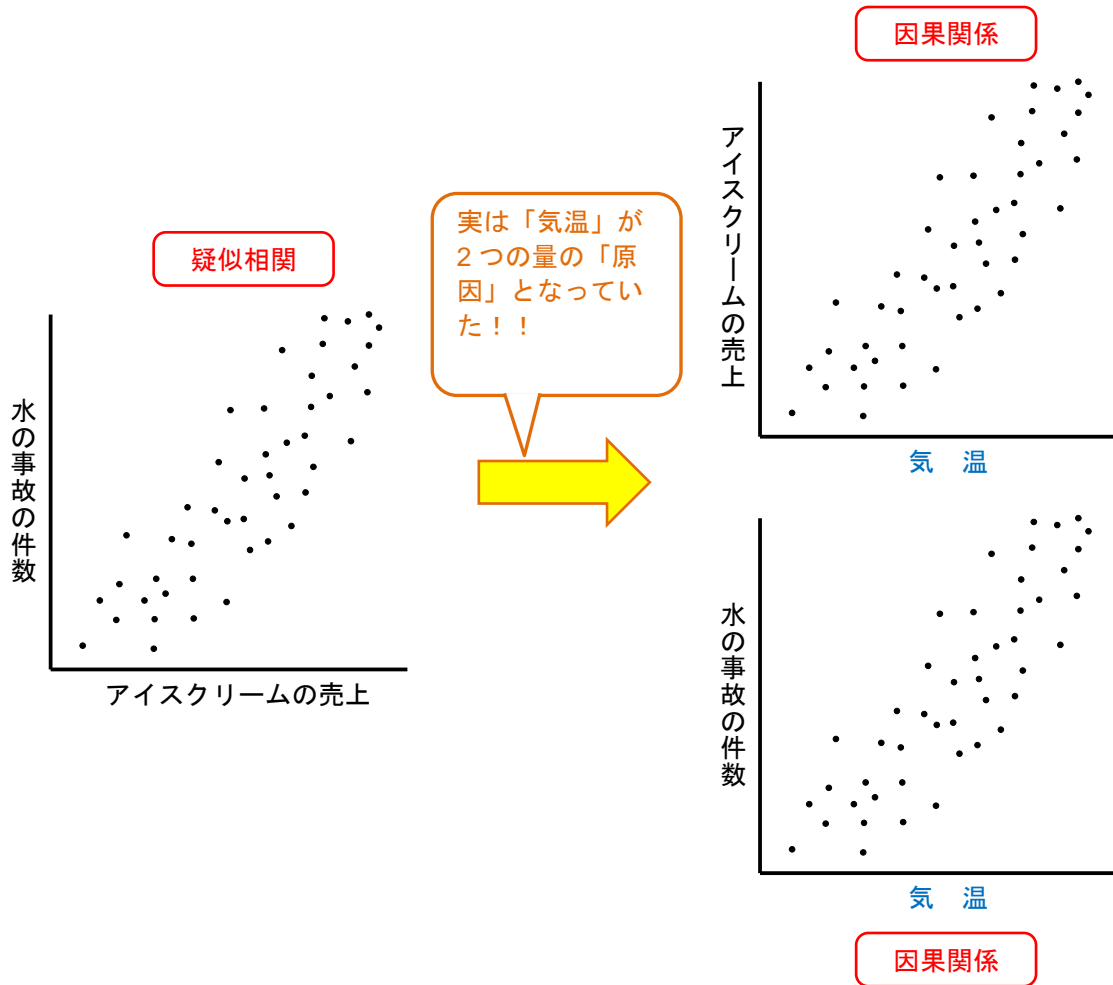
2つの量の間の相関関係がまったくの偶然によって生まれた場合です。

② 別の量との間に因果関係がある場合

さきほどのアイスクリームの売上と水の事故の件数の例で考えてみます。

アイスクリームの売上が増える原因として、例えば気温の上昇が考えられます。また、水の事故の原因も気温の上昇が考えられます（気温が高くなれば水遊びをすることも多くなるので、自然と事故も増えます）。

つまり、下の図のようなイメージです。



ということで、相関関係を扱う際には、いろいろと注意が必要です。

問題

次の例はいずれも「疑似相関」です。第3の変数は何かを考えてみましょう。

- (1) 靴のサイズが大きい子供は文章の読解能力が高い。だから、足を見ればその子の読解力が分かります。
- (2) カラオケが多い街ほど犯罪の件数が多い。街にもう一つカラオケを作ったら、犯罪件数が増えるだろう。

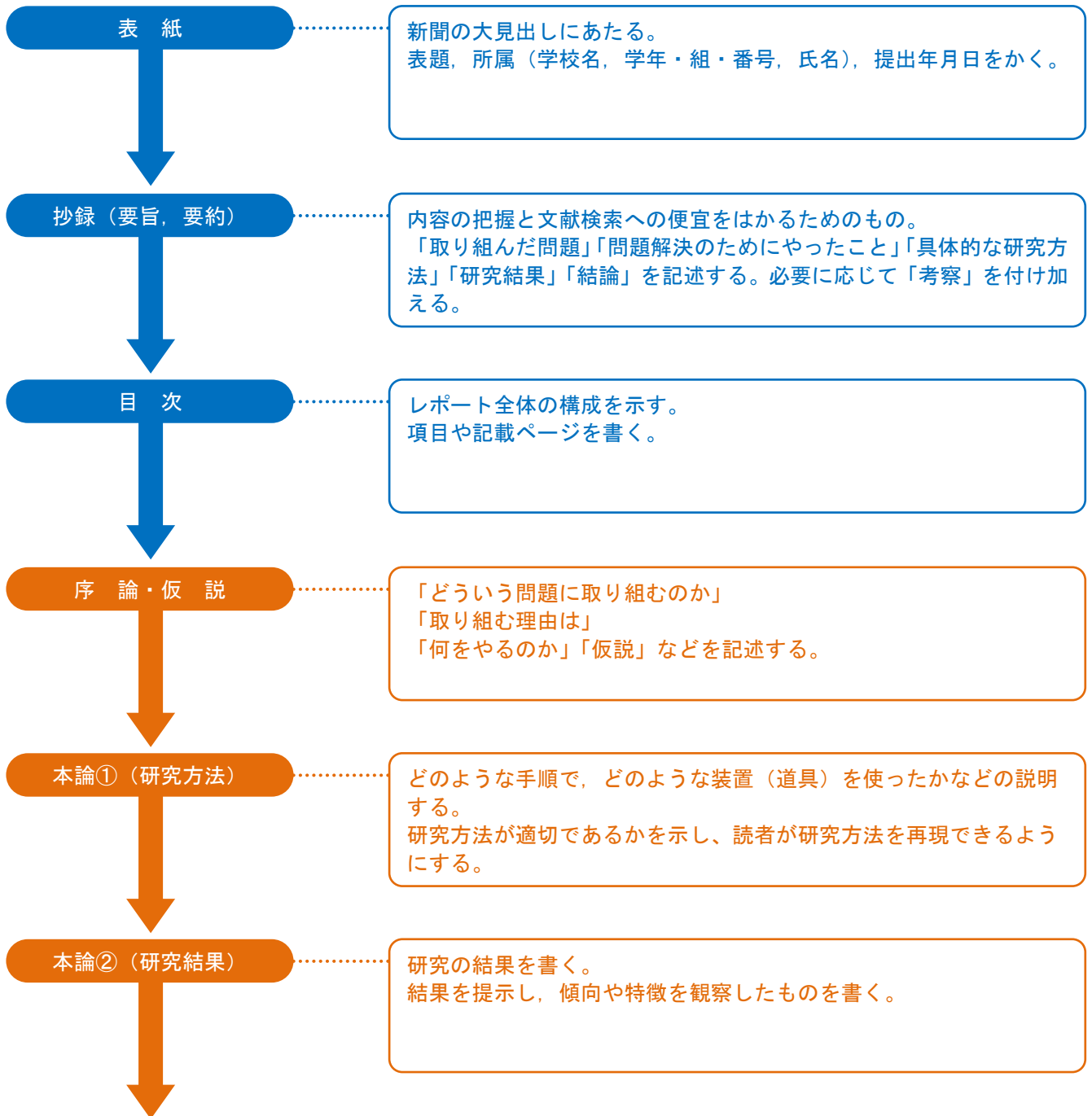
4

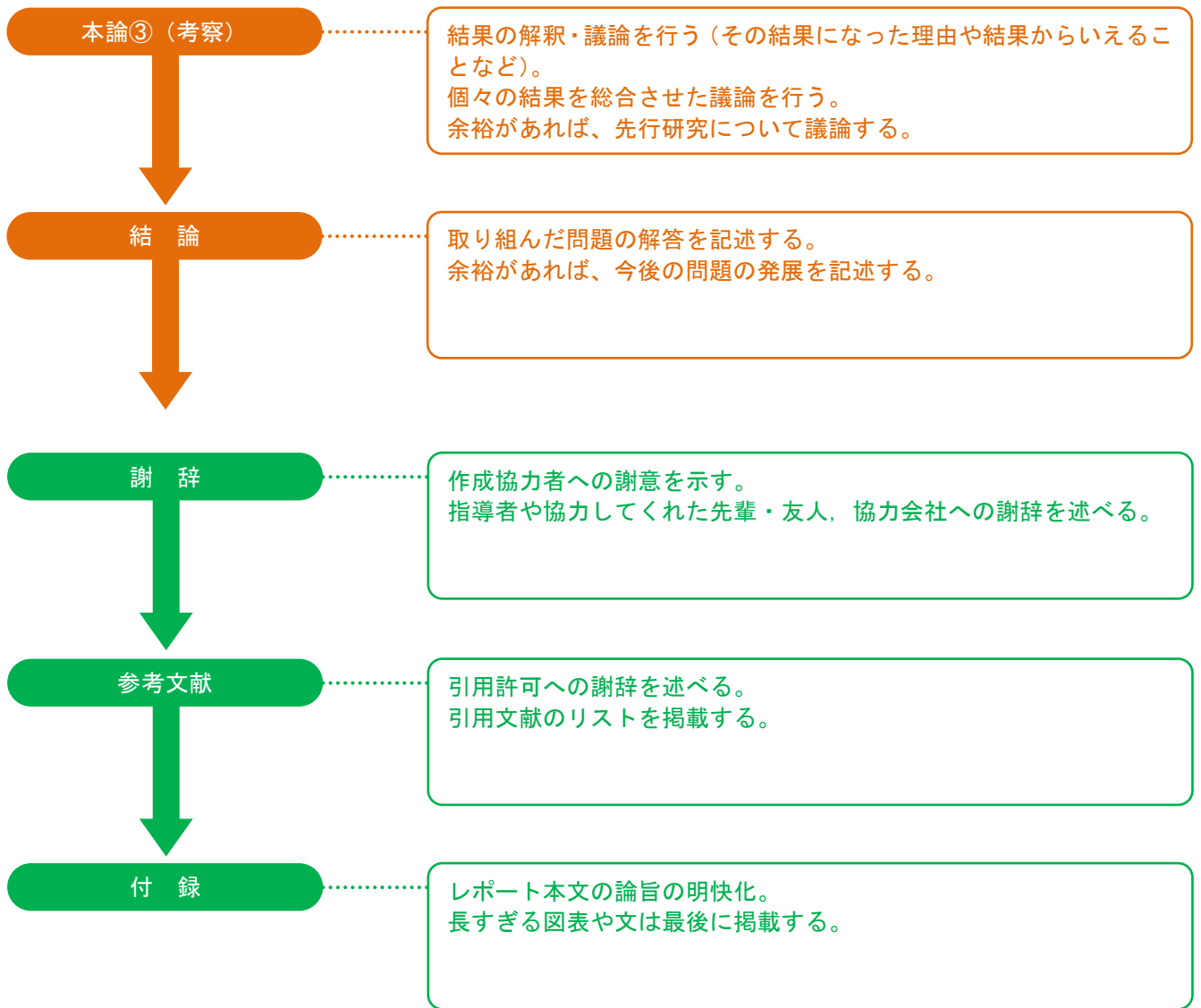
科学レポートを書いてみよう

「小石川フィロソフィー」では、実際に課題研究を行い、その成果について発表したり、論文にまとめたりします。その練習として、この章では「科学レポート」の書き方について学習しましょう。

🐾 レポート・論文の書き方

(1) レポートや論文の構成





<注意>

ここに掲載したものは、長いレポートや論文を書くときに必要となるものです。短めのレポートや論文については、上記のうち、

「抄録」、 「目次」、 「謝辞」、 「付録」
を省略することができます。

(2) 具体的に見てみよう

【序論で書くこと】

「どういう問題に取り組むのか」、「取り組む理由は」、「何をやるのか」「仮説」などを書きます。ここでの問いに対応した解答が結論になります。

【研究方法で書くこと】

「研究対象」、「実験・観察・調査の狙い」、「実験・観察・調査の方法」、「統計処理の方法」をしっかり説明することが大切です。

- 研究対象…実験・観察・調査を行った対象の特徴や由来を示します。例えば、ある花を実験に使ったとしたら、その花の正式な種名や、実験に使用した花をどこで採取したのかまで記述します。
- 実験・観察・調査の狙い…調べる理由を書きます。例えば、実験方法を書くだけでなく、何を狙いとしてその実験を行うのかをしっかり書きます。
- 実験・観察・調査の方法…読者が研究を再現できるように、丁寧に書きましょう。
- 統計処理の方法…これから統計で学習していきます。本格的な論文を書く5年生までに理解しておきましょう。

【研究結果で書くこと】

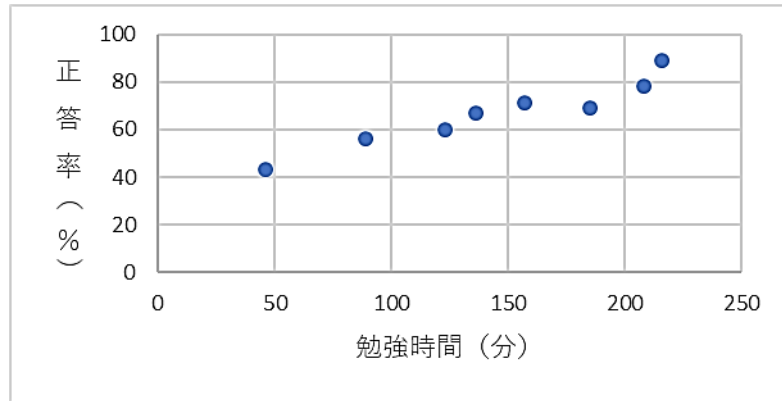
得られたデータを表や図などでわかりやすく提示するとともに、その研究結果の説明を書きます。ここでは、図や表を用いるときの注意点を見ておきます。

・図と表の選択

図は、データ全体から言えることを伝えたいとき、表は、個々のデータを伝えたいとき、に用いるとよいでしょう。

(例) 勉強時間と点数の関係

勉強時間 (分)	正答率 (%)
46	43
89	56
123	60
136	67
157	71
185	69
208	78
216	89



この場合、勉強時間と点数の関係の傾向を読み取ってほしいので、図を用いるほうが適切です。

・図表には説明文をつける

図表には説明文をつけましょう。本文を読まなくても、図表が理解できるような文章をつけます。

(例)

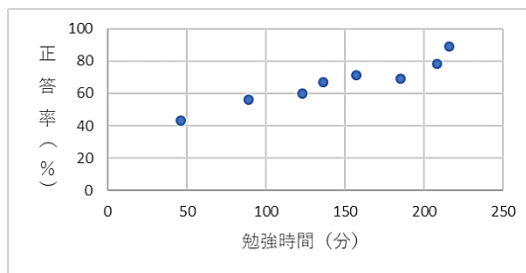


図1 家庭学習の平均時間のとテストの正答率の関係

なお、図の説明文は図の下に、表の説明文は表の上につけることが一般的です。

・図を作成するときの注意点

目盛りの単位を必ずつけます。これを忘れては、図の価値はほとんどありません。

また、横軸には原因(図1では勉強時間)を、縦軸には結果(図1では正答率)をとります。

・表を作成するときの注意点

表1 を見てください。新聞などでみられる勝敗表です。「チーム、勝ち数、負け数」が組になっており、組が横に並んでいることで、見やすくなっています。このように、データの組を横方向にならべ、各組を縦に並べるのが一般的です。

表1 3チームの勝敗表

チーム	勝ち数	負け数
A	4	2
B	3	3
C	2	4

他にも、図や表を見やすくする工夫はたくさんあります。ぜひ、多くの科学レポートを読んで、研究してください。

【考察で書くこと】

実験結果について、そのような結果になった理由や、結果から言えることを書きます。また、この項で先行研究に言及してもよいでしょう。

【結論で書くこと】

研究した内容の解答を書きます。例えば、序論で「なぜ寿司は外国人に人気があるのか」とテーマを決めたとすると、その解答「寿司は外国では健康によいと認識されているから」（この解答は例です）を書きます。結論を書かない論文は、調べ学習と変わりません。

また、今後さらに研究していきたいこと、今後の展望も書けるとなるとよいでしょう。

5

科学レポートを読んでみよう

科学レポートを書くにあたり，はじめに優秀な科学レポートを読んでみましょう。
インターネットを検索すると，入賞した中学生の論文が多く出てきます。興味をもった論文を読んで，その内容をわかりやすく伝えましょう。以下の枠に沿って発表すると発表内容がわかりやすくなります。

【テーマ】

【序論】…「どういう問題に取り組むのか」「取り組む理由は」「何をやるのか」

【仮説】…どのような結果になるかを予想して仮説をたてる

【研究方法】…どのような手順で，どのような装置（道具）を使ったかなど

【研究結果】…データの傾向や観察した特徴など

【考察】…結果の解釈，議論など

【結論】…取り組んだ問題の解答にあたるもの



PPDAC サイクルの実践編

🐾 統計学で生活リズムの改善を！

2章で学んだ PPDAC サイクルを利用して、自分の生活リズムを見直し、改善するためにどんなことが必要か分析してみましょう。 ☞「生徒のための統計活用～基礎編～」 p.70～p.79

STEP 1

知りたい問題を決めよう

Problem 問題

- ◎ 課題（テーマ）設定
自分の生活リズムのどんなところに課題を感じているか？
直したい生活習慣はないか？
改善したい生活リズムは、平日のものか土日のものか？



話し合いメモ

STEP 2

どのようなデータを集めるか考えよう

Plan 計画

- ◎ 調査方法はどうするか？
自分の生活リズム（生活時間）はどうなっているか？
クラス（学年）の生活リズムはどうなっているか？ ☞アンケート調査
全国の中学生の生活リズムはどうなっているか？ ☞e-Stat で統計データ収集



話し合いメモ

🐾 データを集めるには……？

いよいよ調査にとりかかります。データを集めるには、どのような方法があるでしょうか？

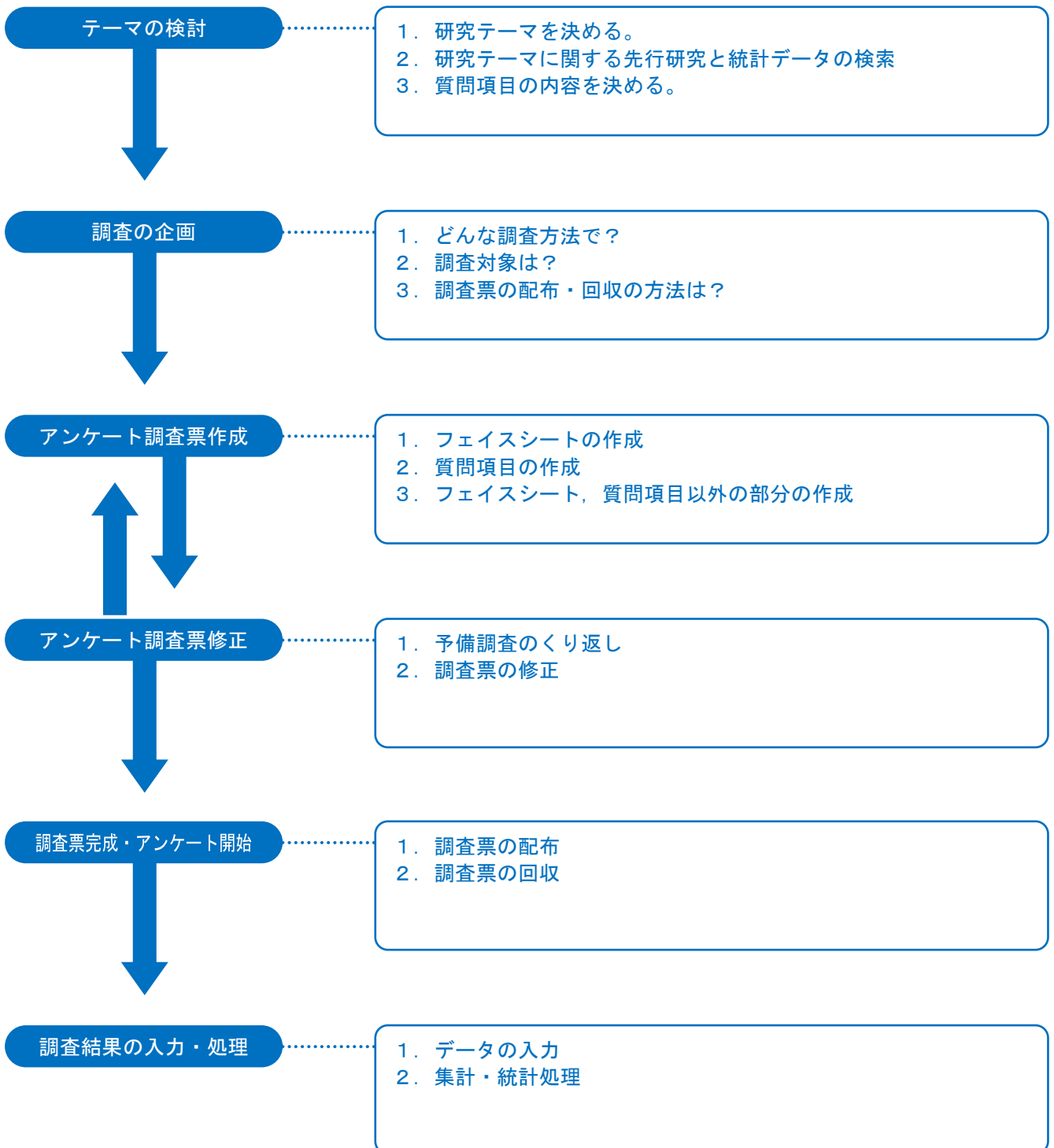
STEP 3

Data 収集

必要なデータ・統計資料を集めよう

(1) アンケート調査によるデータ集め

身近な仲間（クラス，学年，部活動など）のデータを集めるには，アンケート調査が有効です。アンケートをつくるにあたっては，次のプロセスを踏むのが一般的です。



では，このプロセスを1つずつ見ていくことにしましょう。

テーマの検討

研究テーマが決まったら、アンケートの質問項目を考えます。

◎ 図書館の文献やインターネットでキーワードを集める

例えば、「勉強と部活動の両立」をテーマにするなら、平日の時間をどのように使うかがポイントとなります。そこで、放課後の時間をどのようなことに使っているか調査したいと考えるわけです。

「放課後の時間」と一口に言っても、その使い方は人によってさまざまなので、ある程度おおまかな分類を考える必要があります。

「睡眠時間」、「テレビ・ラジオ・新聞・雑誌を読む時間」、「休養・くつろぎの時間」、「遊び（趣味・娯楽）の時間」などに大別しておくといよいでしょう。

そして、それらをまとめた統計資料がないか、図書館やインターネットで調べておくと、アンケートによって集めたデータと、調べた統計資料を比較することができて分析しやすくなります。

◎ 質問項目を考える

質問項目は、あとで集計することを考え、「はい・いいえ」や「1. ○○ 2. △△ 3. □□」のような選択式で解答できるものがよいでしょう。

例えば、「平日の睡眠時間」についてアンケートをとる場合、

（良くない例）『普段の睡眠時間はどれくらいですか？』

とするよりは、

（良い例）『普段の睡眠時間で、自分に近いものに○をつけてください。』

1. 6時間, 2. 7時間, 3. 8時間, 4. 9時間, 5. それ以外』

とした方が、あとで集計しやすくなります。

調査の企画

調査方法には、アンケート調査の他にインタビュー調査もあります。研究テーマをもとに、適している方を選びましょう。

① アンケート調査

アンケート調査票を作成し、調査対象者に回答してもらい、データを集める方法。

<メリット>

- すべての調査対象者に対し、同じ質問項目に回答してもらうので、比較検討に適している。
- 一度に多くの調査ができるので、時間も費用も比較的少なくて済む。

<デメリット>

- 調査対象者が深く考えずに回答したり、質問項目を読み違えたりしてしまうことがある。

② インタビュー調査

調査をする人が調査対象者と直接会い、調査の目的に沿って質問をし、その回答によってデータを集める方法。

<メリット>

- 質問に対する調査対象者の反応を観察できる。
- 調査対象者の回答に応じて質問内容の変更や追加ができる。

<デメリット>

- 一度に多くの調査を実施することができないので、調査に多くの時間や費用がかかる。

③ 表紙

例えば、次のようなものを用意するとよいでしょう。

〇〇〇〇年〇月〇日

生活習慣に関するアンケート調査

私は、〇年〇組の〇〇〇〇です。
生活習慣を改善し、勉強と部活動の両立を図るため、普段の生活の実態を調査しています。

回答の内容は、研究以外の目的に使用することは決してありません。
調査は無記名です。回答の内容が外部の方にもれることはありません。
ありのままをお答えくださいますようお願いいたします。

＜記入についてのお願い＞

○ 他の方と相談することなく、お一人でお答えください。
○ 回答後は、回答欄に記入漏れがないか、ご確認ください。

＜調査結果および質問について＞

調査結果をお知りになりたい方やご質問がある方は、下記の連絡先にお知らせください。

連絡先：〇年〇組

アンケート調査を行っている人や機関を書き、どんな目的で調査を行っているかを説明しましょう。

調査結果をどのように利用するか明記し、プライバシーが侵害されることがないことを伝えましょう。

記入にあたって注意して欲しいことがあれば書いておきましょう。

調査票修正・完成へ

作った調査票は、第三者に一度見てもらいましょう。その上で、

- わかりづらい質問項目はないか
- 誤解を招くような質問項目はないか

などをチェックしてもらいましょう。

最後に、誤字・脱字がないか、レイアウトは適切かなどを確認し、完成です！

(2) インターネットによるデータ集め 「生徒のための統計活用～基礎編～」 p.124～

政府統計の総合窓口である「e-Stat」を利用してみましょう。今回例として挙げている生活時間の統計データは、「社会生活基本調査（総務省）」があります。

- ① インターネットブラウザで「e-Stat」にアクセス
検索するか、URL「<http://www.e-stat.go.jp/>」を直接入力。
- ② 「主要な統計から探す」→「社会生活基本調査」とクリック



- ③ 目的のものを探して、**Excel** ボタンをクリック
そうすると、表計算ソフト Excel が立ち上がり、データを閲覧することができます。Excel を利用すれば、データを加工したり、すでにあるデータから別のデータをつくったりすることもできます。
いずれ授業でも Excel の使い方を学習します。

練習してみましょう！

「生徒のための統計活用～基礎編～」の p.72 に掲載されている、「全国（10～14 歳）の睡眠時間」の度数分布表は、「社会生活基本調査」のデータをもとにして作っています。
この度数分布表の元になっているデータを、e-Stat から探してみましょう。

元データが掲載されている表のタイトルを記入しておこう

いよいよデータの分析です！

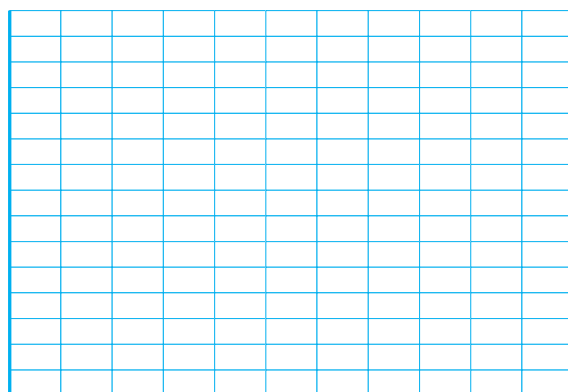
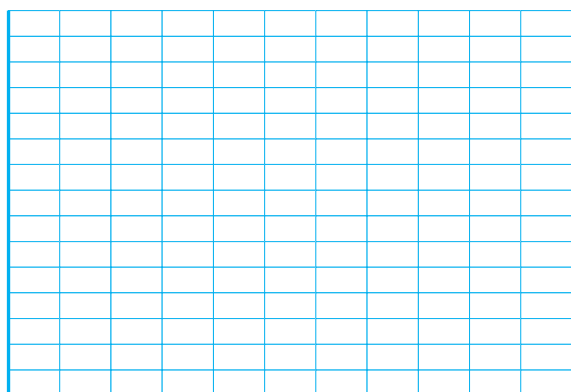
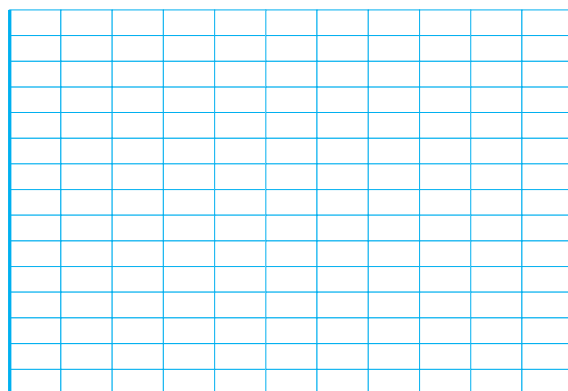
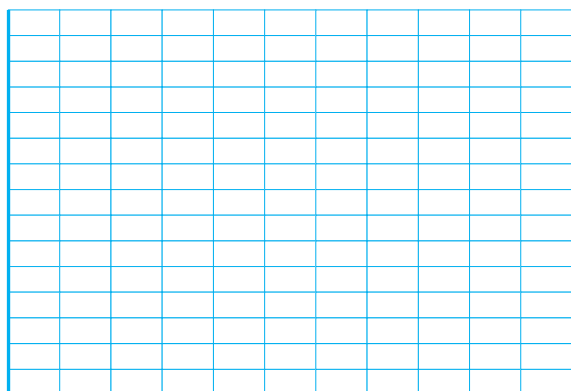
アンケート調査によって得られた身近な人たちのデータと、全国のデータを度数分布表やヒストグラムにまとめたり、代表値を求めたりして、傾向をつかみましょう。

STEP 4

Analysis 分析

グラフや表、統計量で傾向をとらえよう

以下の方眼は、下書きとして自由に利用してください。



☆ 生徒のための統計活用～基礎編～」 p.72～p.77 を参考にして、グループで話し合い、まとめてみましょう。

STEP 5

Conclusion 結論

わかったことをまとめ・読み取ろう

「生徒のための統計活用～基礎編～」p.79 を参考にして、以下の欄に自分なりに結論をまとめてみましょう。

結論



統計グラフコンクールの作品を発表しよう

発表に向けて

みんなに作ってもらったグラフコンクールの作品を、校内の発表会で発表してもらおう予定です。そのための発表原稿づくりやクラス内発表会を行います。よりよい発表ができるように、チームのメンバーで話し合って、内容を深めましょう。

(1) クラス内発表について

① 発表の方法

発表の提示資料として、グラフコンクール用に制作した作品を、黒板に貼り付けて、その前に並んで発表します。(机を後ろに移動して、ブース形式で発表を行う。)

② 発表の条件

- 何も見ずに発表すること。**

原稿を書きますが、当日はその原稿を見ずに発表してもらいます。そのためには、①内容を理解すること、②簡潔にまとめること、③難しい言葉はわかりやすいものに直すことが必要です。

- チームの全員が発表すること。**

誰か一人が発表して終わりとするのではなく、チームの全員が発表できるような構成としてください。

- 持ち時間は 1 チーム 5 分。**

発表時間は 3 分以内とします。必要な原稿の量を考え、「短すぎず長すぎず」を心がけるようにしましょう。

- 作品以外の資料の提示は可。**

発表で、特に強調したい部分などは、画用紙等書き写して、聴衆に提示しても構いません。資料を提示すると、聴衆に具体的なイメージを持たせることができてわかりやすく、また、集中させる効果もあります。

- どのような質問がくるか想定して答えを準備してください。**

③ 発表する項目

発表の際は、以下の項目を必ず盛り込んでください。

- テーマは何か
- そのテーマを選んだ理由
- 調査の方法
- 結論（自分たちの考えや主張）とその根拠

※ 根拠は、統計資料を使ってわかりやすく説明できるように。

どのようにすれば聴衆にわかりやすいか考えながら原稿をつくりましょう。

<参考>

- 先行研究・調査はあるか？

同じテーマについて調べた資料がすでにあれば、その資料と自分たちの調査結果を比較できます。

- 今後の課題と展望

今回調べたことが、今後どのように発展していくかを発表すると、聴衆の興味をひきます。

- 先に結論を述べて、その根拠を説明していくとわかりやすいです。

原稿づくり

(1) 発表の構成を考えよう

① オープニング

発表テーマ：



テーマ設定の理由

② ボディ

発表を通じてもっとも主張したいことや結論：



調査の方法

主張の根拠・理由となる統計資料等

③ クロージング

今後の課題・展望，調査をしてみたの感想等

(2) 発表の要点をまとめよう

① オープニング

スピーチの「つかみ」の部分です。本題に入る前の数十秒から数分の間で、関心や好意的な姿勢を、聴衆から引き出す必要があります。

オープニングの案

② ボディ

聴衆にわかりやすくシンプルな構成にします。特に、自分たちの主張したいこととその根拠を明確にする必要があります。

ボディの案

③ クロージング

締めくくり方は発表の印象に大きな影響を与えます。自分たちの主張したいことを簡潔に要約し、記憶に残るようにする必要があります。

クロージングの案

発表の練習をしよう

実際に時間を測って、発表の練習をしてみましょう。発表のたびに改善点を話し合い、よりよい発表になるように磨いていきましょう。

実際の発表は、1つのチームにつき、

- ① 発表（3分）
- ② 質疑応答、評価シートの記入（2分）

という流れになります。



🐾 確率は2種類あります

正しく作られたさいころを1回投げるとき、どの目が出るかは誰にもわかりません。ところが、さいころの目は1から6までなので、さいころを1回投げたときに起きる結果の可能性は、1の目が出るか、2の目が出るか、3の目が出るか、4の目が出るか、5の目が出るか、6の目が出るか、の6つの結果のうちのどれか1つです。

よって、さいころを1回投げたときに1の目が出る確率は、6つの（同程度に期待できる）可能性のうちの1つだから、 $\frac{1}{6}$ になります。このような考え方から求めた確率のことを、 といいます。

一方、さいころを100回投げたときに、そのうち20回1の目が出たとすると、1の目が出る確率は、 $\frac{20}{100} = \frac{1}{5}$ と計算できます。このような考え方から求めた確率のことを、 といいます。

ここでは、**数学的確率 $\frac{1}{6}$ と統計的確率 $\frac{1}{5}$ は異なる値となりました。**

このように、試行する回数が少ないと、統計的確率と数学的確率の間にずれが生じます。

ところが、さいころを投げる回数を何百回、何千回、何万回とどんどん増やしていくにしたがい、1の目が出る割合は数学的確率 $\frac{1}{6}$ に限りなく近づいていきます。

このように、ある試行の回数を増やせば、統計的確率と数学的確率の誤差をいくらでも小さくできるということが、確率の という定理によって保証されています。

🐾 さいころの1の目が出る確率

(統計的確率を求めてみよう)

さいころを投げる回数を増やすほど、1の目の出る統計的確率が、数学的確率に近づいていくのか、実際にさいころを投げて実験してみましょう。

手順は、次の通りです。

(1) グループをつくり、役割分担を確認する

投げる人、確認する人、記録する人、連絡する人を決めます。



(2) さいころをひたすら投げて出た目を記録する

できるだけ多くの試行（さいころを投げる）を行います。このとき、投げる人の作為が入らないよう、きちんとさいころが転がるように投げます。

終了の合図があったら、連絡担当の人はさいころを教卓にもどし、先生に結果を伝えます。

記録

(3) 集計表から作成されたグラフを観察しよう

◎グラフからわかること

数学的確率の求め方

正しく作られたさいころを投げるとき、どの目が出るかの結果は全部で 6 通りあり、その 6 通りのどれが起こることも同じ程度に期待できます。

このようなとき、それぞれの結果が起こることは といいます。

確率論では「起こりうる事柄」を**事象**というが、どの事象の起こる結果も同様に確からしいときに限り、次の計算によって**数学的確率**を求めることができます。

起こりうる結果が全部で n 通りあり、どの場合が起こることも同様に確からしいとする。

このとき、ある事象 A の起こる場合が a 通りあるとすると、事象 A が起こる確率は、 $\frac{a}{n}$ で求められる。

例 1 1 枚のコインを投げるとき、表が出る確率は

例 2 1 個のさいころを 1 回投げたとき、出る目が 3 の倍数である確率は

例 3 赤玉 3 個、青玉 2 個、黄玉 5 個が入った袋から玉を 1 個取り出すとき、赤玉が出る確率は

例 4 1 組のトランプのカード 52 枚からカードを 1 枚引いたとき、エースが出る確率は

例 5 白玉 5 個が入った袋から玉を 1 個取り出すとき、白玉が出る確率は

例 6 白玉 5 個が入った袋から玉を 1 個取り出すとき、黒玉が出る確率は

例 5、例 6 からわかるように

事柄 A が必ず起こる確率は、

事柄 A が決して起こらない確率は、

です。一般に、事柄 A の起こる確率を p とすると、

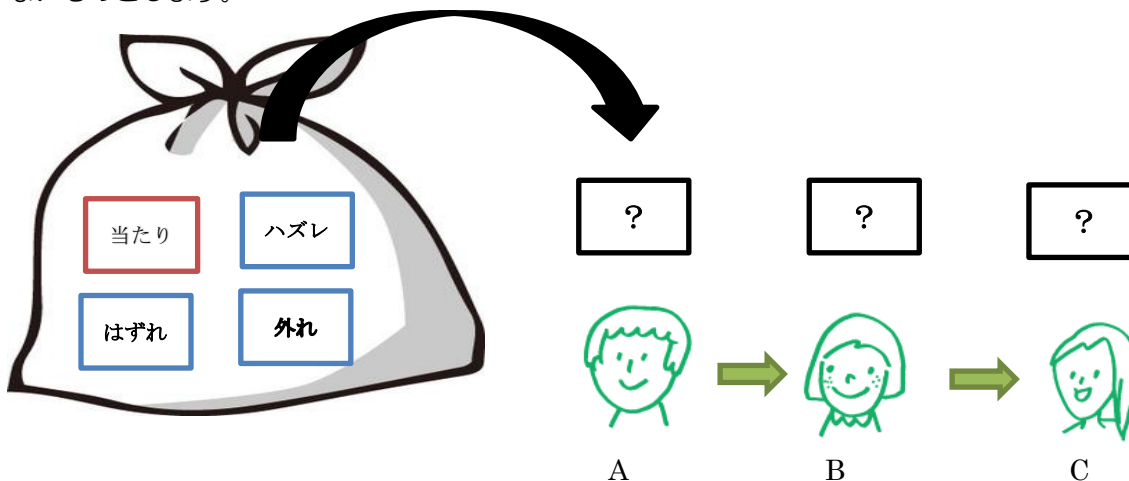
$$\text{} \leq p \leq \text{}$$

が成り立ちます。

🐾 勘違いしやすい確率の話

直感的に正しいと感じても、それが論理的には正しくない、といったことが確率論にはよくあります。ここでは、典型的な3つの例を紹介しましょう。

1本の当たりくじと、3本のはずれくじの計4本のくじが入った袋があります。この袋から、A、B、Cの3人がこの順で1本ずつくじを引くとき、誰が最も当たりやすいでしょうか？ ただし、引いたくじは袋の中に戻さないものとします。



(1) 直観で予想してみよう

①Aが当たりやすい ②Bが当たりやすい ③Cが当たりやすい ④変わらない ⑤わからない

(2) 考えてみよう

メモ

◎ 結論

2. モンティ・ホール問題

A, B, C の3つのドアがあります。

そのうちの1つのドアの向こうに景品があり、残りの2つのドアの向こうには何もありません。

ゲームの参加者がこの3つのドアのうち1つのドアを開けて、そこに景品があれば、景品をもらえます。ただし、参加者がドアを開けるまでに、司会者と参加者の間で次のようなかけひきがあります。

- ① 参加者がドアを1つ選ぶ（ドアはまだ開けない）。
- ② 景品が入っているドアを知っている司会者が、残りの2つのドアから景品が入っていないドアを1つ開ける。
- ③ 司会者が参加者に「ドアの選択を変更してもよい」と告げる。
- ④ 参加者が2つのドアから最終的に選んだドアを開ける。



では、問題。このゲームで、③の後に景品を当てたい参加者はドアを変更すべきでしょうか。

(1) 予想してみよう

- ①変更した方がよい ②変更しない方がよい ③変わらない ④わからない

(2) トランプで実験してみよう

赤のトランプ1枚をあたり、黒のトランプ2枚をはずれとします。

司会者役の人が相手に見せないようにトランプ3枚を持ちます。

参加者役の人が1枚のトランプをひき、札を見ずに机の上に置きます。

司会者が残りの2枚のうち、はずれ（黒）のトランプを見せます。

参加者は 取り替える または 取り替えない で、自分のカードの札を確認します。

記録者は、当たったか、はずれたかを記録します。

【1】グループをつくり、役割分担を確認する。

【2】集計を記録しよう

記録

【3】集計結果を報告し、それぞれの仮説にもとづく結果から確率を求めます。

仮説1：選択肢を変える（選びなおす）				仮説2：選択肢を変えない（そのまま選ぶ）			
当たり	回	はずれ	回	当たり	回	はずれ	回
カードを変えたときに当たる確率→				カードを変えないときに当たる確率→			

(3) 結論 参加者は、選択肢を 変更した方がよい ・ 変更しない方がよい ・ 変わらない

◎ その理由は？

3. 精度99%の検査

症状がひどいとされるたちの悪い新型のウイルスが発生したとします。ある病院の、ウイルスの感染検査では、99%の精度で「陽性」か「陰性」かを判別できるそうです。あなたはその検査で「陽性」という判定を受け、悲観してしまいました。

それを聞いた数学の T 先生は、「そのウイルスは 1 万人に 1 人の割合で感染しているものだろ。2 人に 1 人ならともかく。心配しすぎるな」と、謎の励ましをしてきました。珍しいウイルスだからこそ悲観しているのに T 先生は何を言っているのでしょうか。それともこの励ましに何か根拠があるのでしょうか。

(1) 予想してみよう

- ①根拠のない励ました ②T 先生を信用する ③わからない

(2) 100 万人がこの検査を受けたとして、感染する割合が 1 万人に 1 人か、2 人に 1 人かで何がかわるのか考えてみよう。

メモ

◎ 結論

1, 2 学期に学んだ内容は、統計学の中でも特に「**記述統計学**」と呼ばれる分野です。これは、すべてのデータを工夫して整理することによって、全体がどのような傾向にあるか、どんな性質をもっているかを探るものでした。このように、集団のすべてを対象に調査することを、統計学では**全数調査**と呼びます。

ところが、統計量が大きいときには対象のすべてを調査することができるとは限りません。そこで、偏りなく集団の一部のみを調査することによって、全体の性質を推測することを考えます。この考え方を「**推測統計学**」といいます。

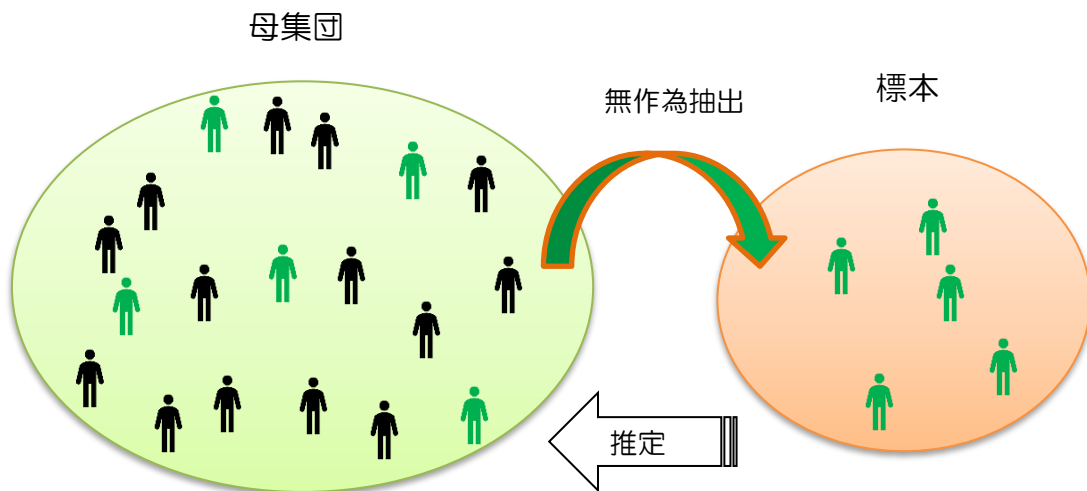
🐾 一部だけ取り出して調べる方法

対象とする集団の一部を調べ、その結果から、集団全体の状況を推定する調査を といいます。

この調査において、調査の対象全体を といい、調査のために母集団から抜き出されたものの集まりを 、母集団から標本を抜き出すことを標本の といいます。

また、母集団に含まれるものの数を 、標本に含まれるものの個数を といいます。

このように、集団の一部のみに注目して全体の性質を推論する統計学の方法を といいます。



標本調査は

- ① 適切な標本の大きさで
- ② まったくバラバラの（性質に偏りのない）集団を選んで

行うことが大切です。

くじ引きなどの方法で、母集団から偏りなく標本を選ぶことを、標本を といいます。街頭アンケートやインターネットによるモニター調査は無作為に抽出したとはいえません。

無作為抽出では、人間の意志がまったく入らないように、さいころを振ったり、コンピュータでランダムにつくった数を並べた**乱数表**などを使ったりするのが一般的です。



乱数さい

例 表計算ソフト（EXCEL）では、「RAND」「RANDBETWEEN(*,*)」で発生できます。

🐾 標本調査の実際

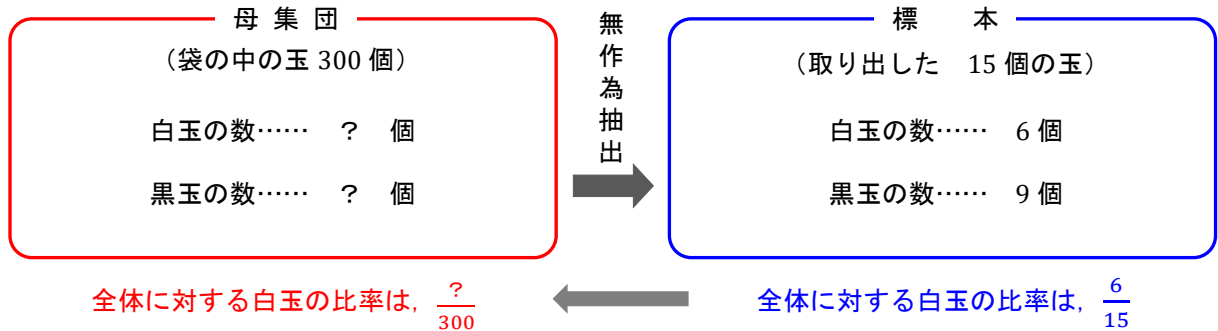
標本調査の例を2つ紹介します。

例1 袋の中に白玉は何個？（母比率の点推定）

袋の中に大きさが等しい白玉と黒玉が合計 300 個入っています。この袋の中の玉をよく混ぜてから 15 個の玉を取り出したところ、白玉が 6 個、黒玉が 9 個でした。

このとき、最初に袋の中に入っていた白玉の個数は、およそ何個と考えられるでしょうか？

(1) 話を整理してみよう



(2) 計算してみよう

標本における白玉の全体に対する比率は、

これが、母集団における白玉の全体に対する比率と同じだと推定することができます。

よって、最初に袋の中に入っていた白玉の個数は、およそ

$$300 \times \text{} = \text{} \text{ (個) と推定できます。}$$

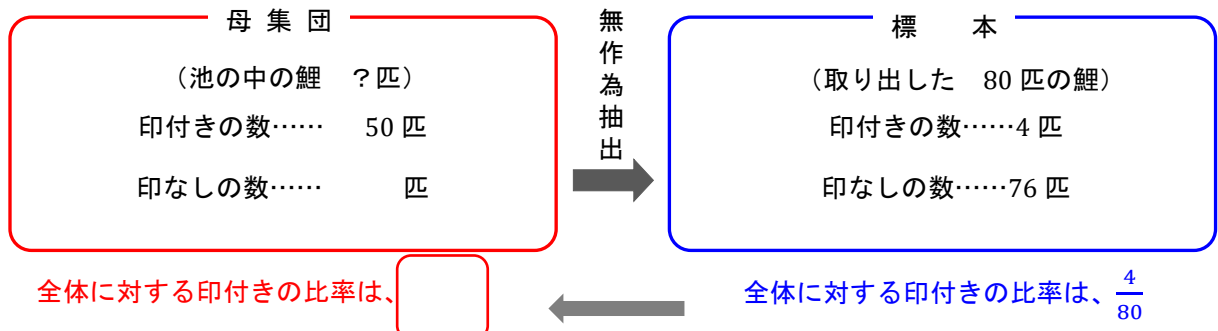
例2 湖に鯉は何匹いるだろうか？（母集団の大きさの点推定）

広く深い池にいる鯉の数をすべて捕らえて数えるのは不可能です。このようなときには、一部の鯉を捕獲し、母集団の個体数を推定する「捕獲再捕獲法」という方法があります。

- ① まず、池のあちこちから全部で 50 匹の鯉を捕獲し、それらに印をつけて、池に放す。
- ② その魚がちらばるのを待ち、数週間後に同じようにして池から全部で 80 匹の鯉を捕獲したところ、そのうちの 4 匹に印がついていたとする。

この結果から、池にいる鯉の数を推定してみましよう。

(1) 話を整理してみよう



(2) 計算してみよう

標本における印付きの鯉の全体に対する比率は、

これが、母集団における印付きの鯉の全体に対する比率と等しいと推定することができます。

よって、池にいる鯉の数は、 と推定できます。

🐾 標本調査の実験（視聴率調査）

ここでは、標本調査の例として、視聴率調査の模擬実験を行います。

(1) 視聴率とは？

視聴率は、テレビの番組やCMがどのくらいの世帯や人々に見られているかを示す一つの指標です。

関東地方の世帯数はおおむね1987万世帯（2016年1月1日現在、総務省公表）です。これらのうち、ある番組を見ている世帯の割合を表したものが視聴率ですが、視聴率は全数調査ではなく、無作為に900世帯を抽出して標本調査を行っています。

(2) 実験の目的

袋の中に入っている80個の玉を国民の人数、取り出す玉の数が協力世帯数、取り出した赤玉の数が番組を見た世帯数を表しているとし、視聴率調査の実験を行うことで、標本の大きさが、推測の結果にどのように影響するのか調べる。

(3) 準備

グループをつくり、視聴率調査キットを持っていく。

- ① 白玉と赤玉がたくさん入っている袋
- ② 標本を取り出すための容器

(4) 方法

- ① 標本の大きさがそれぞれ4個、8個、12個、15個、25個、30個で推定するグループとする。
- ② 袋の中身をよくかき混ぜ取り出す。決められた個数に足りなかった場合は、袋の中身を見ずに手で何個か取り出して調整する。
- ③ 赤玉が何個入っていたかを確認し、＜実験の記録＞に記録する。
- ④ 容器の中身を袋に戻した後、よくかき混ぜ、③をくり返す（できるだけ多くの試行を繰り返す）。
- ⑤ 他の班の結果も写して、すべてのグラフを仕上げる。
- ⑥ 赤玉の出る割合（視聴率）がどのくらいになるか予想する。

[比率の換算表]

標本の大きさが4個の場合

個	1	2	3	4
%	25	50	75	100

標本の大きさが8個の場合

個	1	2	3	4	5	6	7	8
%	13	25	38	50	63	75	88	100

標本の大きさが12個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12
%	8	17	25	33	42	50	58	67	75	83	92	100

標本の大きさが15個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
%	7	13	20	27	33	40	47	53	60	67	73	80	87	93	100

標本の大きさが25個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
%	4	8	12	16	20	24	28	32	36	40	44	48	53	56	60	64	68	72	76	80	84	88	92	96	100

標本の大きさが30個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	...
%	3	7	10	13	17	20	23	27	30	33	37	40	43	47	50	53	57	60	63	67	70	73	77	80	...

(5) 記録

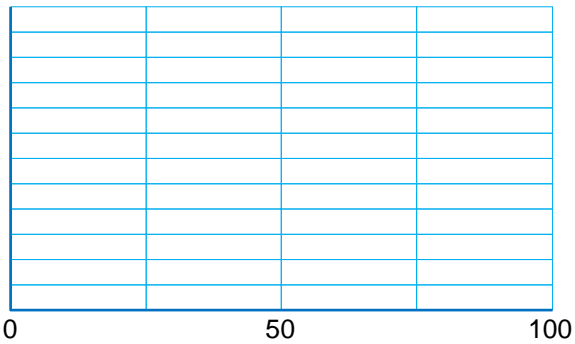
<実験の記録>

回	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
個																									
%																									

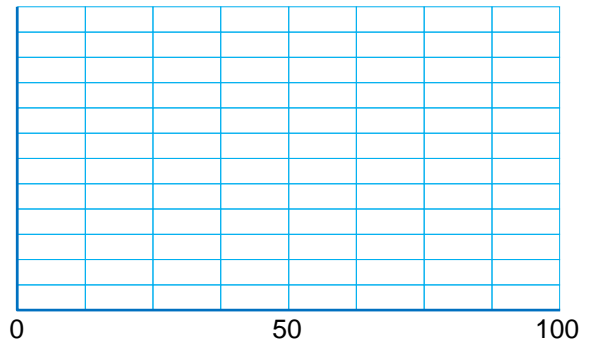
回	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
個																									
%																									

<ヒストグラム> (横軸は%)

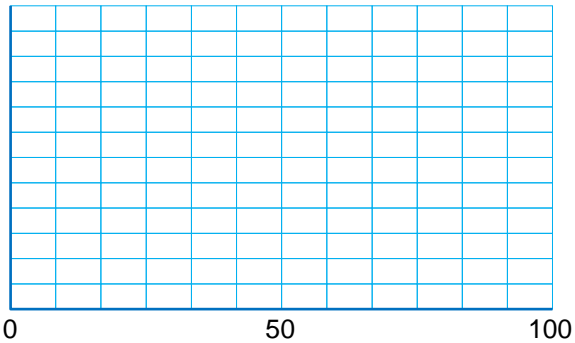
標本の大きさが 4 個



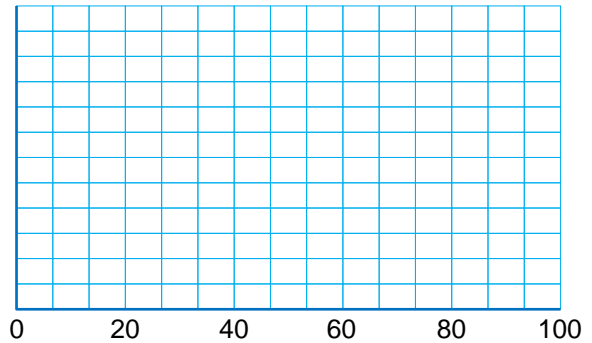
標本の大きさが 8 個



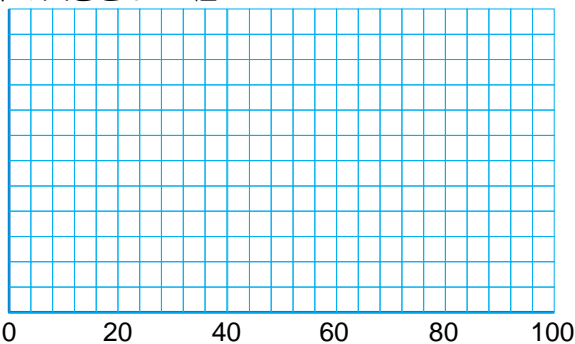
標本の大きさが 12 個



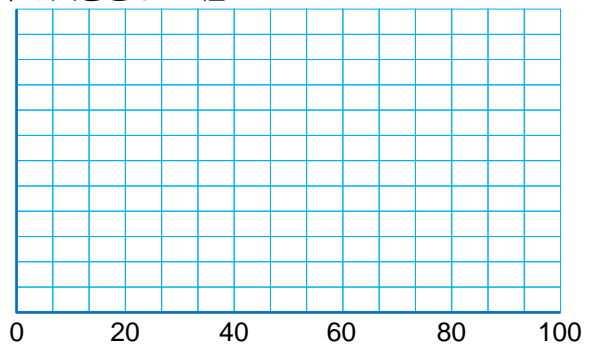
標本の大きさが 15 個



標本の大きさが 25 個



標本の大きさが 30 個



<結論>

これらの結果から、何がわかるか考えてみよう。

🐾 幅を持たせて予想しよう（視聴率）

(1) 信頼度とは？

視聴率調査から、視聴率について、一つの値で推定の結果とすること（点推定）は危険であることがわかります。

そこで、誤差を考慮し、幅を持たせた区間を推定の結果とすること（区間推定）が一般的です。つまり、『**%から**%の間』といったように推定を行います。

標本の大きさが十分大きいときには、この区間を求める下のような計算式が存在します。ただしその際、推定結果がどれだけ信用できるかの指標として信頼度という値を定める必要があります。

例えば、信頼度 95% (0.95) というのは、「100 回調査したら約 95 回はこの範囲に視聴率がある」ということです。

信頼度を上げた方が文字通り確からしさは上がりますが、その分区間の幅も広がるため、実用性は下がります。

分野にもよりますが、一般的に信頼度は 95%か 99%を使うことが多いです。

※ややこしいですが、標本や母集団の比率についての%と、信頼度（推定自体の確からしさ）の%を混同しないようにしてください。

母比率（母集団の比率）の区間推定

母比率を \hat{p} 、標本の大きさを n 、標本比率を p 、信頼度によって定まる定数を z とすると

$$p - z \sqrt{\frac{p(1-p)}{n}} \leq \hat{p} \leq p + z \sqrt{\frac{p(1-p)}{n}}$$

※信頼度 95%のとき $z = 1.96$,

99%のとき $z = 2.58$

$z \sqrt{\frac{p(1-p)}{n}}$ を 標本誤差 といいます

(2) 視聴率の信頼区間を計算で求めてみよう。

いま、「900 世帯の標本のうち 225 世帯が視聴していた」とします。このとき、信頼度 95%で区間推定を行ってみましょう。（結果のことを 95%信頼区間 と呼びます）

標本比率 p は となります。

次に標本誤差 $z \sqrt{\frac{p(1-p)}{n}}$ を小数第五位まで求めると $1.96 \times \sqrt{\frac{0.25(1-0.25)}{900}} =$

よってこの視聴率の 95%信頼区間は %から %の間にあると考えることができます。

(3) (2)の例で視聴率の 99%信頼区間を推定してみよう。

次ページの 2 つの問題について、母比率の推定の考え方をういて解いてみよう。

問題 1 製造された不良品の割合（不良率）は？

小石川社では、毎日頑丈な瓦『小石瓦』を大量生産している。ある日製造した『小石瓦』の中から無作為に 400 個抽出したところ、8 個の割れた瓦（不良品）を発見した。この日製造した製品全体の不良率（不良品の比率）について 95%信頼区間を求めてみよう。

問題 2 信頼区間を縮めるために・・・

ある都市で有権者の内閣支持率を調べるために標本調査を行うことになった。信頼度 95%で推定したときに信頼区間の幅を 0.2 以内（標本誤差が ± 0.1 以内）にするためには、何名以上の有権者を無作為抽出すればよいか考えてみよう。

ヒント：標本比率を p とすると

$$p(1-p) = -p^2 + p = -\left(p - \frac{1}{2}\right)^2 + \frac{1}{4} \leq \frac{1}{4}$$

より $p(1-p)$ は（0 以上） $\frac{1}{4}$ 以下の値をとる。

母比率以外にも、母平均や母分散など推測したい値（母数）や母集団について分かっている情報によって、使われる区間推定の計算式が異なります。（深入りはしません。）



推測統計学Ⅱ：確率分布

前の章では、推測統計学の入門として、区間推定を謎の式を用いて行いました。その背景にあるのは確率論です。ある統計量について、母集団あるいはそこからとった標本が、どの事象を、どの程度の割合でとると期待できるのかを知ることによって、その値を推測することができます。

ある試行について、その結果（**確率変数**）1つ1つとそれぞれの確率との対応を**確率分布**といいます。この章では、いくつかの確率分布についてみていきましょう。

二項分布

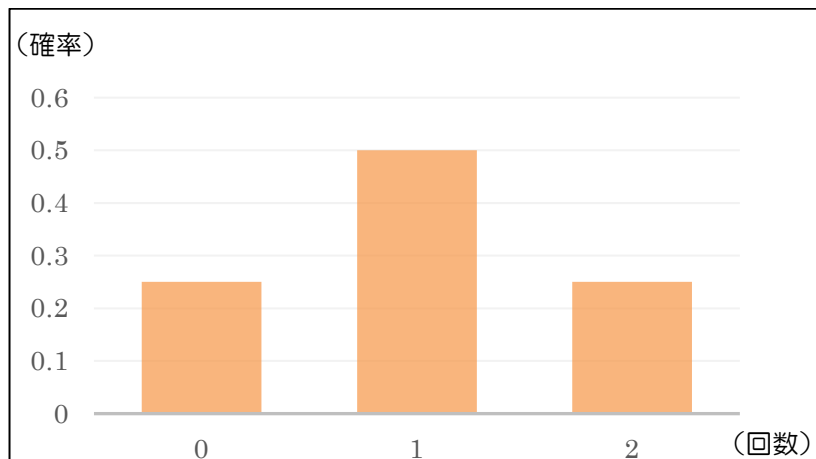
例 1 コインを何回か投げて、表が出る回数の確率を求めよう。

コインを何回か投げて表が出る回数（横軸）とその確率（縦軸）をグラフにしてみます。

(1) コインを2回投げたとき

コインのすべての出方は、（表，表）（裏，裏）（表，裏）（裏，表）のいずれかです。よって、

表が0枚出る確率は $\frac{1}{4} = 0.25$ 、表が1枚出る確率は $\frac{2}{4} = 0.5$ 、表が2枚出る確率は $\frac{1}{4} = 0.25$ であり、グラフは次のようになります。



(2) コインを8回投げたとき、表が出る確率についてまとめてみよう。

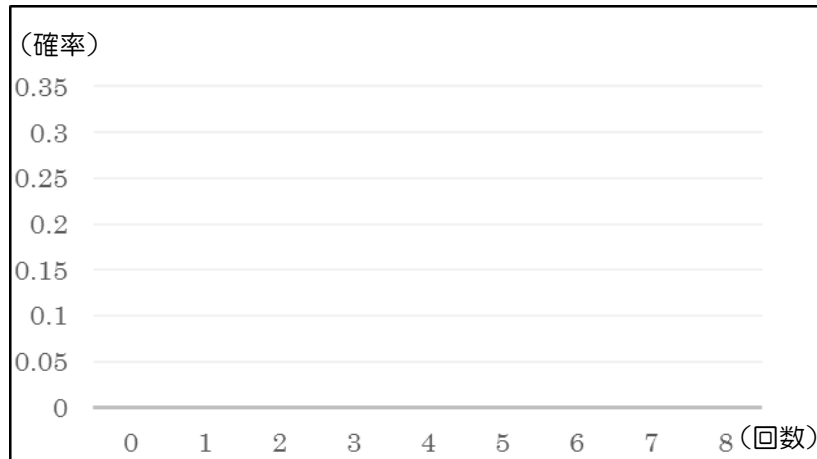
メモ

ヒント：例えば8回中2回表が出る場合の数は、8個のうちの2個を選ぶ組合せの総数と等しい。

表にすると

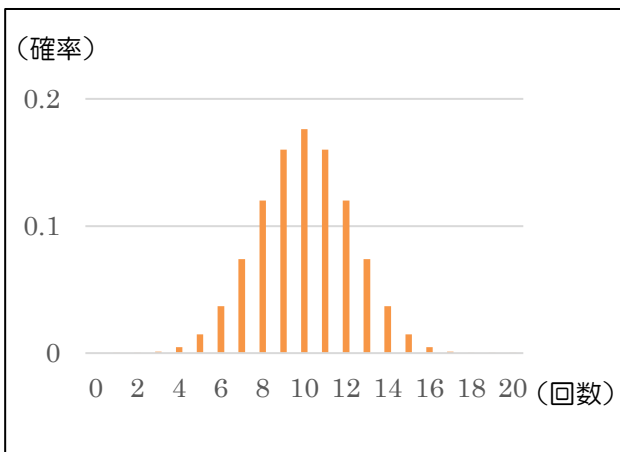
表の回数	0	1	2	3	4	5	6	7	8	計
確率 (分数)										
(小数)										

グラフにすると

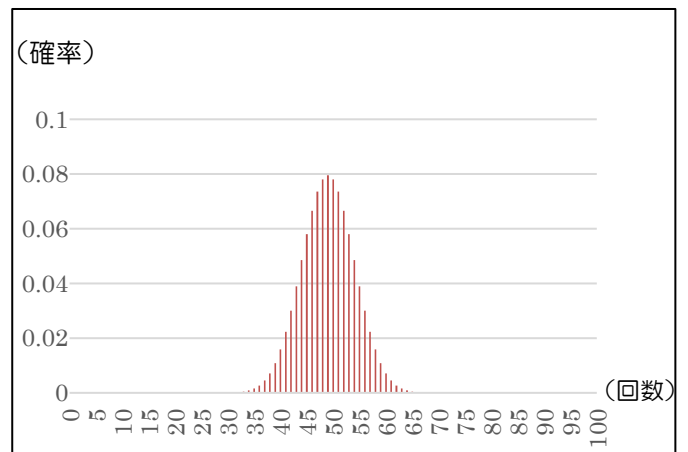


(3) さらにコインを投げる回数を増やす

20 回



100 回



このように、1 回の試行につき、2 つの可能性がある試行を繰り返したときに、その片方が起こる回数を確率変数とすると、その変数は、二項分布とよばれる確率分布に従います。

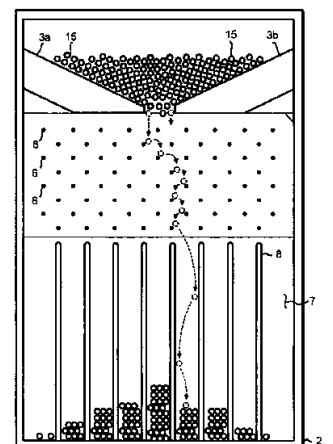
※試行を繰り返しても、1 回あたりの確率が変わらない試行であることが前提です。

例 2 二項分布の統計的確率

右のような器具によってできるピンボールの玉の分布を考えてみましょう。

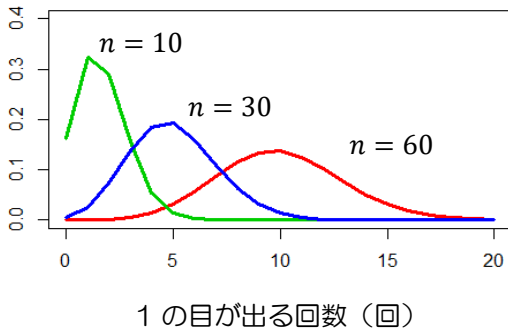
上のコインの場合で「表が出る」、「裏が出る」という可能性を、釘に当たったら「右に進む」、「左に進む」とこと同様だと考えると、コインの実験の統計的確率の分布とみることができます。

右の器具で 512 個の玉が上から落ちたとしたら、それぞれの枠に入る玉は何個ぐらいだと考えられるでしょう。



🐾 推測統計の肝（正規分布）

下の図は、さいころを n 回振って「1の目が出る回数」を確率変数とする、二項分布のグラフを比較したものです。縦軸は確率ではなく、確率密度（次ページ参照）という値です。



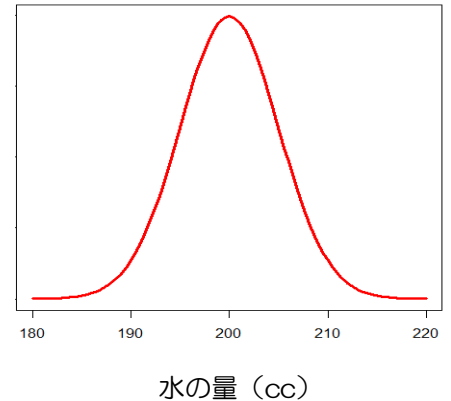
前ページやこのグラフから、二項分布は、1回あたりの確率に関係なく、試行回数が多くなると、**左右対称な山型の分布**になることが予想できます。

二項分布は試行回数が多いとき、**正規分布**と呼ばれる確率分布に近似できることが知られています。

正規分布は、いろいろな自然現象や、社会現象を説明する際に利用されています。

例えば、ある工場で水 200 cc をボトルに詰めるとします。ところが、どんな優秀な装置を使ってもピッタリ 200 cc の水をボトルに詰めることは困難です。実際には 200 cc より多かったり少なかったりまちまちです。

このとき、ボトルに詰めた水の量と 200 cc との差が、誤差になります。水が詰められたたくさんのボトルについて誤差を求め、その分布のグラフをつくると右の図のようになります。



例3 パン屋の不正を見抜け！

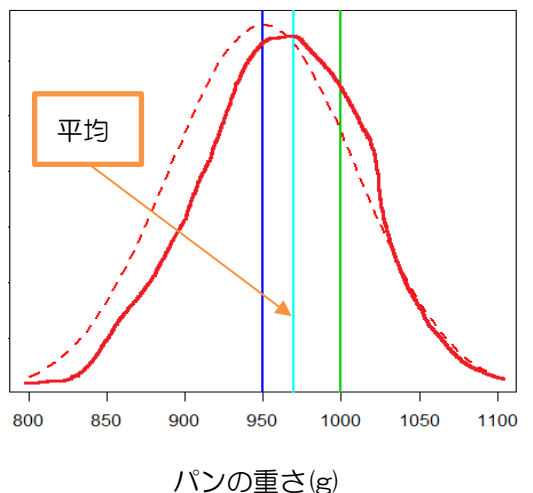
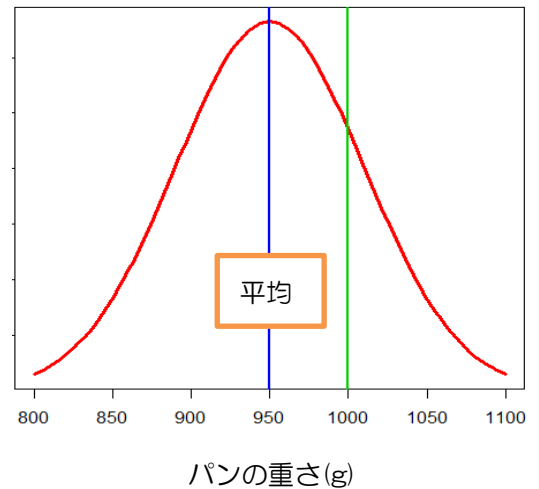
フランスの数学者アンリ・ポアンカレ（1854~1912）についての逸話です。

ポアンカレは「1kgの重さのパン」を売るパン屋に通っていました。1kgといっても誤差はあるので、920gの日もあれば、1020gの日もあります。実は、このパン屋は1kgより軽いパンを故意に作って、誤差だとかまかしていたのですが、ポアンカレに指摘されてしまいます。

ポアンカレは1年間毎日買ったパンの重さを計りました。すると、その重さのグラフは右の図のような正規分布を描きました。このことから、パン屋は50g少なく、950gを基準にパンを作っていたことが判明したのです。

話はまだ続きます。ポアンカレの指摘以降、ポアンカレが「1kgの重さのパン」を買うと、平均して950gより重いパンがもらえるようになりました。

一件落着と思いきや、ポアンカレはパンの重さを計り続けました。すると今度は右下の図の、実線のようなグラフになったのです。ここから何がわかるでしょう。



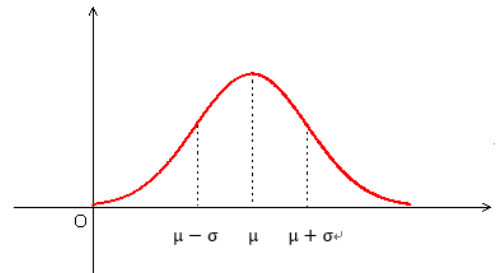
製造業などの現場においても、部品の重さや大きさの平均が正規分布を描くかどうかで品質を確認しています。

次に挙げる4つは推測統計学の大切な部分ですが、雰囲気が分かれば十分です。

正規分布の特徴

正規分布は左右対称な山型の分布です。山の頂上に対する x の値が正規分布の「平均値」であり、山の中腹（変曲点）に対する x の値が、「平均値±標準偏差」です。

図では平均値を μ （ミュー）、標準偏差を σ （シグマ）で表しています。



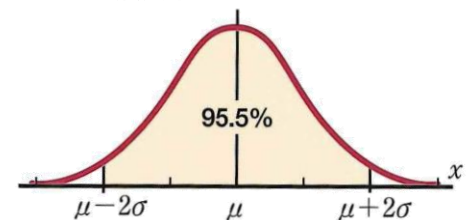
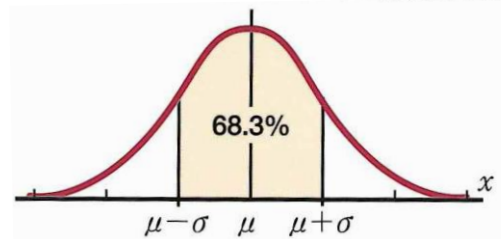
確率密度

確率密度 y を簡単に説明すると、ある区間におけるその関数のグラフと横軸で囲まれる面積と、その区間に確率変数が含まれる確率とを同一視するために、縦軸を調整した値です。

例えば正規分布において、右図の色付き部分 $\mu - \sigma \leq x \leq \mu + \sigma$ の面積は 0.683 であり、その区間に確率変数がある確率も 68.3% (0.683) です。

同様に、 $\mu - 2\sigma \leq x \leq \mu + 2\sigma$ の面積は 0.955 で確率も 95.5% です。

μ から3 σ まで離れると、確率は 99.7%になります。

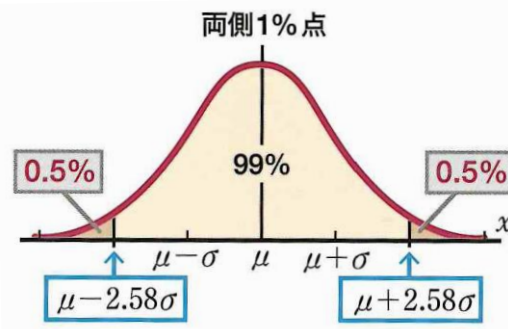
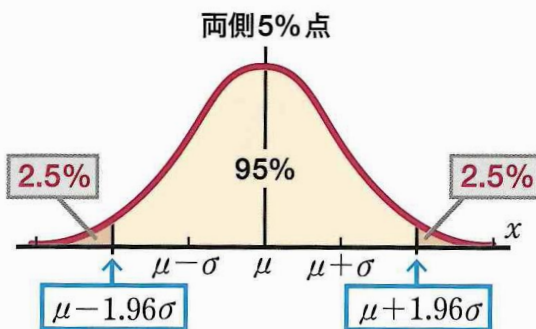


パーセント点

μ を対称の軸とし「95%と99%を覆う範囲」は統計学でよく利用されます。

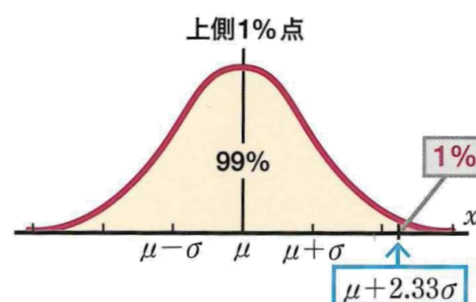
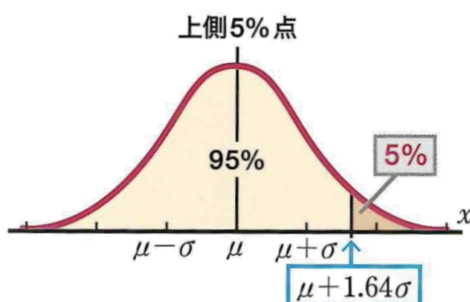
これらの右側の境界点をそれぞれ、正規分布の両側5%点、両側1%点といいます。

前者は $x = \mu + 1.96\sigma$ 、後者は $x = \mu + 2.58\sigma$ です。区間推定で登場した1.96や2.58はここから来ています。



また、左側から右側にわたる95%、99%の範囲もよく利用されます。これらの境界点をそれぞれ、正規分布の上側5%点、上側1%点といいます。

前者は $x = \mu + 1.64\sigma$ 、後者は $x = \mu + 2.33\sigma$ です。



中心極限定理

母集団の平均を μ 、分散を σ^2 とします。このとき、標本の大きさを十分大きくとると、母集団の分布がどんな形であっても、その標本平均 \bar{X} は、近似的に正規分布に従います。

また、この正規分布の平均は母平均と等しく μ 、分散は母分散を標本の大きさを割った $\frac{\sigma^2}{n}$ です。

中心極限定理は重要な定理なのですが、誤解しやすいので例を示します。

例えば、全国の主婦の「へそくり額の平均値」を標本調査で調べるとします。

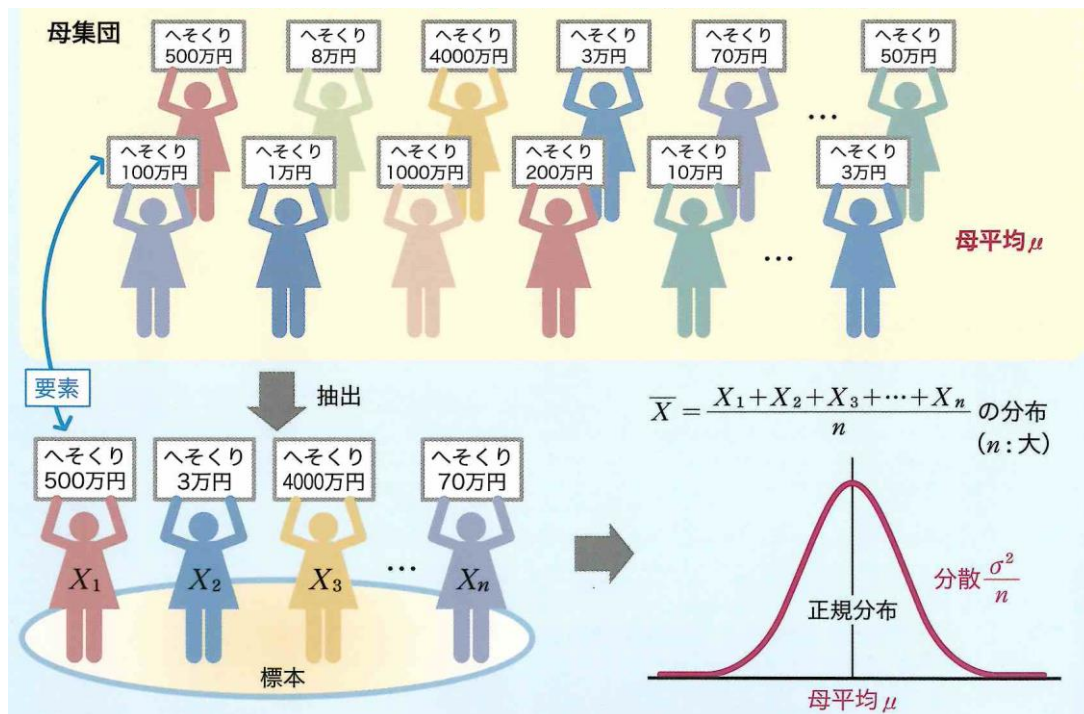
このとき、母集団である「へそくり額」自体が正規分布に従うとは限りません。へそくり額3万円の主婦もいれば、3000万円持っている主婦もあり、偏りがあるはずで

さて、この母集団から n 人の無作為抽出をし「へそくり額の平均値」を求めることを考えます。 n 人の標本の取り方はいろいろあるので、その1つ1つに対して「へそくり額の平均値」という値が定まります。

中心極限定理は、 n が大きいきこの「へそくり額の平均値」が正規分布に従う、と主張しています。

また、この正規分布の平均は、母集団の平均と等しいことも分かっています。

これによって母集団の確率分布が不確かでも、1つの標本平均の値から、実際のへそくり額の平均値について信頼区間を推定できるわけです。



技術評論社『統計学の図鑑』より抜粋

🐾 偏差値と正規分布

模擬試験を受けると偏差値がわかります。
 点数を X 、平均点を μ 、標準偏差を σ とすると、**偏差値**は、

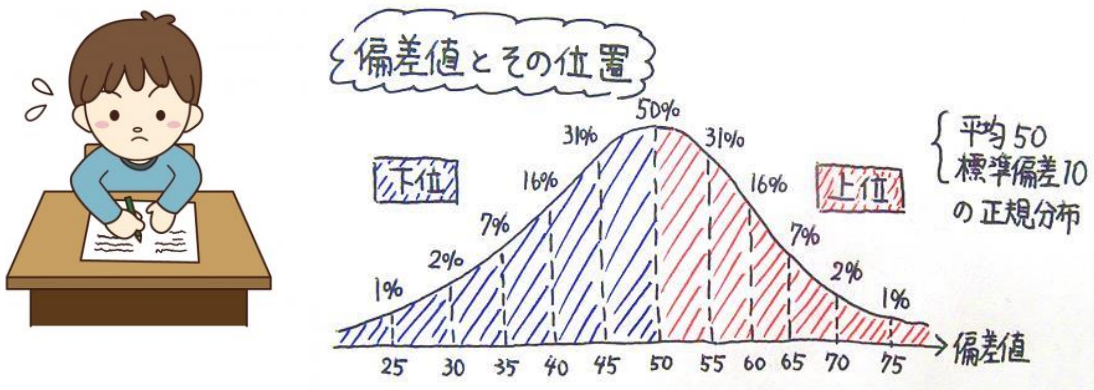
$$\text{(偏差値)} = \frac{X - \mu}{\sigma} \times 10 + 50$$

と定義されています。

$\frac{X - \mu}{\sigma}$ を X の標準化といいます。点数について偏差をとり、標準偏差で割ることで、確率分布の平均が0、標準偏差が1となるよう調整されています。

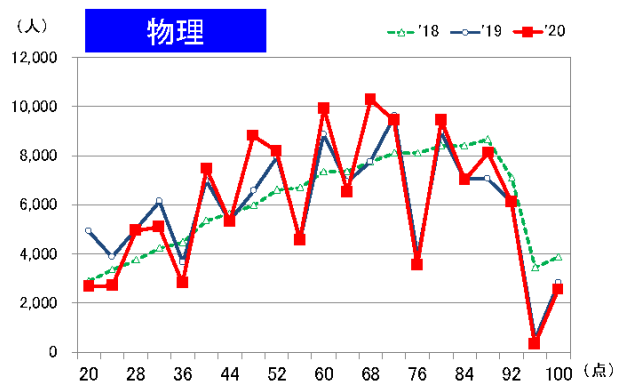
さらに標準化した値に10をかけて50を足しているのが、**母集団の平均を50、標準偏差を10になるよう調整した値が偏差値**です。

仮に母集団が正規分布に従うとすると、偏差値40から60に全体の68.3%の人がいるということになり、偏差値70を超えたと受験者全体の上位2.3%といえます。つまり、もしも受験者が100人だったら上位3位以内の素晴らしい成績であると考えられます。



ただし、注意してほしいのは**母集団が正規分布に従わない場合、偏差値だけで全体における正確な位置はわからない**ということです。偏差値は標本平均の分布ではないので、正規分布に従うとは限りません。母集団の分布がどうなっているかが大切です。

右はセンター試験物理の得点分布ですが…偏差値をもとに自分の位置を考えるのは難しそうです。



データネット科目別度数分布

ベネッセコーポレーション/駿台予備学校

問題

Sさんは数学のテストで前回80点取れたのに、今回は70点しか取れませんでした。平均点はどちらも60点であり、標準偏差は前回が20点、今回が8点です。

生徒全員の数学の得点分布が前回も今回も正規分布に従うとき、Sさんは喜ぶべきか悲しむべきか考えてみましょう。

帰無仮説と対立仮説

例 1 ラーメン代値上げ交渉

Yさんは大のラーメン好きです。ラーメン屋に行くとき、親に毎回ラーメン代600円をもらうのですが、Yさんの行くラーメン屋さんには、600円では足りないことが多いのです。Yさんは、ラーメン代の値段交渉をしようとデータを集めることにしました。

自分が住んでいる地域のラーメン屋さんは5000軒以上あり、すべてを調べるのは難しいので、無作為にラーメン屋100軒を抽出し、ラーメン一杯の平均額を計算したところ690円でした。これをもとに、値上げ交渉をしようとしたYさんは、親にこう言われました。

「たまたま調べた100件の平均が690円だけで、誤差の範囲じゃない？」
これに対する反論の糸口はあるでしょうか。



統計的仮説検定（検定）とは、主張したい仮説を確率論に基づいて訴える方法です。つまり、親の主張は現実的にほとんどあり得ない、とYさんが伝えられれば勝ちです。

検定は次の流れで行います。難しい言葉が並びます。例1より簡単な例2を後に示しますので、このページと照らし合わせながら確認してください。

- ① 母集団に関する**帰無仮説**、**対立仮説**の2つをたてる
- ② **検定統計量**を定め、帰無仮説を前提としたときの確率分布を考える
- ③ **有意水準**を定め、**棄却域**を確認
- ④ **実現値**から、帰無仮説を**棄却**するか帰無仮説を**受容**するか結論を出す

① 母集団に関する帰無仮説、対立仮説の2つをたてる

検定の大筋は、「ある仮説を前提として考えていった結果、その仮説は確率的にありえない」というもので、断定でないことを除けば背理法によく似ています。

前提として話を進め、最終的に否定したい仮説を**帰無仮説**、帰無仮説を否定することによって最終的に主張したい仮説を**対立仮説**といいます。

確率によって仮説を否定することを**棄却**、認めることを**採択**といいます。

② 検定統計量を定め、帰無仮説を前提としたときの確率分布を考える

平均値や分散など、検定に使う値（**検定統計量**）と、その値が従う確率分布を考えます。

③ 有意水準を定め、棄却域を確認

どの程度低い確率の結果が示されたら帰無仮説を棄却するか、という基準が**有意水準**です。推定における信頼度と似たようなものです。5%または1%に設定されることが多いです。

確率分布から、起こる可能性が有意水準以下となる範囲（**棄却域**）が特定できます。

④ 実現値から、帰無仮説を棄却するか帰無仮説を受容するか結論を出す

検定統計量について、抽出した標本から計算した値（**実現値**）が、棄却域に入れば帰無仮説を**棄却**、棄却域に入らなければ帰無仮説を**受容**します。

注意 帰無仮説の受容とは、「帰無仮説がありえないとはいえない」という消極的なものです。それだけを根拠として帰無仮説を採択することはできません。（検出力という話になってきますが、割愛します。）

例2 コインに細工はあるか

コインを8回投げたところ、7回表が出ました。このコインに特別な細工があるかどうか有意水準5%で検定してみましょう。

① 仮説を立てます。

帰無仮説：コインに細工はない（表が出る確率は1回あたり0.5）

対立仮説：コインに細工がある（表が出る確率は1回あたり0.5より大きい）

帰無仮説を前提とするので、表が出る確率が一つに定まる「細工はない」を帰無仮説とするのがよいです。

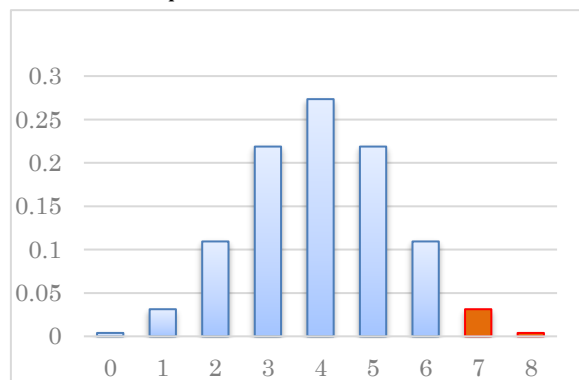
② コインを8回投げた時の表の出る比率を検定統計量とします。これは、p.52 でかいた二項分布に従います。

③ 有意水準は5%とします。つまり8回中7回以上表が出る確率が5%以内であれば、帰無仮説を棄却できます。

④ 8回中7回以上表が出る確率は である

ので、帰無仮説は されます。

よって



といえます。

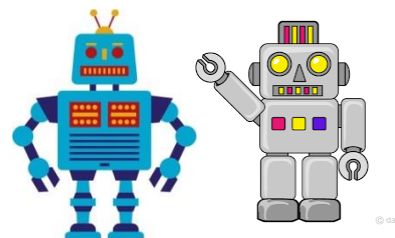
問題

2体のロボットA、Bがある対抗型ゲームを何回か行う。ただし、各回のゲームでは引き分けはなく、A、Bの勝つ確率は変わらないものとする。

これまでの実績からAの方が強いと見られていたが、Bもいくつかの改良を加えてAと互角の強さになったといわれている。

今回、10回のゲームを行ったところ、Aが8回、Bが2回勝った。

このことから「やはりロボットAの方がロボットBより強い」といってよいかどうかを有意水準5%で検定しよう。

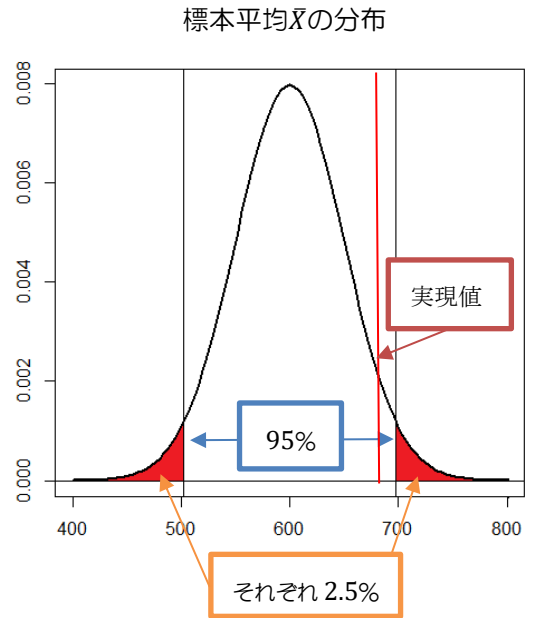


🐾 両側検定と片側検定

例 1 ラーメン代値上げ交渉再び

それでは、冒頭の例 1 について有意水準 5% で検定してみましょう。実は情報として、もう 1 つ母集団の標準偏差が必要です。ですが一般的に、推測したい母集団についての標準偏差は、分かる方が珍しいです。代わりに標本の標準偏差を使う検定もあるのですが、少し複雑です。今回は母集団の標準偏差が 500 円と分かっていることにして話を進めます。

- ① 自分の住んでいる地域について
 帰無仮説：ラーメン一杯の平均額は 600 円
 対立仮説：ラーメン一杯の平均額は 600 円ではない
- ② 平均値を検定統計量とします。
 中心極限定理より、標本平均 \bar{X} は正規分布に従います。
 母平均が 600 という前提で話を進めるので、 \bar{X} の正規分布の平均は 600、分散は $\frac{500^2}{100} = 2500$ です。
 すなわち標準偏差は $\sqrt{2500} = 50$ です。
- ③ 有意水準は 5% です。右図の赤く塗られた部分に実現値が入る確率は 5% 以内です。境界の値（**臨界値**）は、
 $600 \pm 1.96 \times 50 = 502, 698$ です。
 （式の意味は p.54、パーセント点の項参照）
- ④ 棄却域は 502 円以下か 698 円以上です。実現値は 690 円ですから棄却域に入っていない。
 よって、帰無仮説は棄却できず、「ラーメン一杯の値段の平均額は 600 円でない」とはいえません。



このままでは Y さんは泣き寝入りですが、対立仮説の立て方に問題があります。

帰無仮説 : ラーメン一杯の平均額は 600 円に等しい。

に対して、3 通りの対立仮説の立て方があります。

対立仮説① : ラーメン一杯の平均額は 600 円に等しくない。

対立仮説② : ラーメン一杯の平均額は 600 円よりも高い。

対立仮説③ : ラーメン一杯の平均額は 600 円よりも安い。

対立仮説①は、600 円よりも高いか安いかわからないが 600 円ではないということです。これを**両側検定**といいます。

対立仮説②のように、「600 円よりも高い（大きい）」とした検定を**右片側検定**といいます。

対立仮説③のように、「600 円よりも安い（小さい）」とした検定を**左片側検定**といいます。

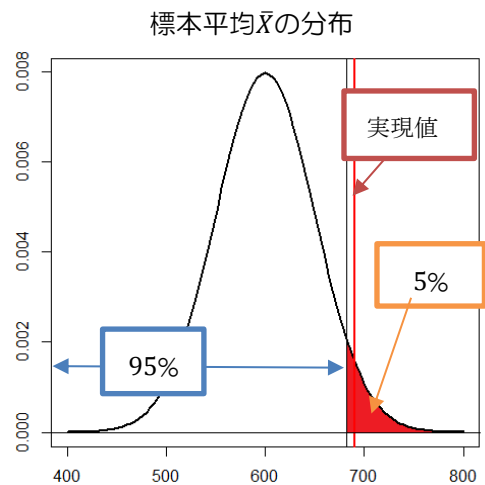
先ほど行ったのは両側検定でしたが、Y さんが主張したいことは、「600 円よりも高い！」ということなので、右片側検定を用いるのが正しいです。つまり、棄却域を左右に 2.5% ではなく、右片側に 5% とし、実現値が平均よりも高いことを反映しやすくします。

このとき、臨界値を求める係数は 1.96 ではなく、1.64 を使います。（上側 5% 点を求めればよいということです）

$$600 + 1.64 \times 50 = 682$$

となり、めでたく実現値が棄却域に入ったため、帰無仮説は棄却され、対立仮説が採択されます。すなわち、600 円よりも平均額は高いといえます。

これをもとに、ラーメン代交渉は有利に進められそうです。



係数早見表

有意水準	5%	1%
両側検定	1.96	2.58
片側検定	1.64	2.33

問題

2000年の小学5年生の全国平均身長は148.5cm、分散は 7.8^2 でした。食生活の変化などのため、子供の成長に変化が起きたと思われます。そこで、現代の小学5年生の児童を無作為に100人抽出したところ、その平均身長は149.9cmでした。小学5年生の子供の身長が変わったといえるかどうかを有意水準5%で検定してみましょう。また、身長が伸びたといえるかどうかも有意水準5%で検定してみましょう。ただし、分散は変わっていないものとします。

2つの過ち（第一種の過誤、第二種の過誤）

確率的な決断を下す検定では、当然「過ち」を犯す危険があります。次の2つに分類されます。

第一種の過誤：帰無仮説が本当は正しいにもかかわらず、棄却してしまうという過ち

第二種の過誤：帰無仮説が本当は誤りなのににもかかわらず、受容してしまうという過ち



裁判の例で考えてみます。ある被告人について

帰無仮説：「有罪である」

対立仮説：「有罪ではない（無罪である）」

とすると、

第一種の過誤は

第二種の過誤は

第一種の過誤を犯す確率は有意水準と同じです。それでは、有意水準を小さくすれば検定の精度が上がるのでしょうか。上の例で考えてみましょう。第一種の過誤を犯す確率を下げるために、有罪であると判断する基準を上げることになると、そのぶん、本来無罪であるはずの人を有罪としてしまう、つまり第二種の過誤を犯す確率を上げることになってしまいます。

逆に有罪であると判断する基準を下げて、第二種の過誤を犯す確率を下げると、そのぶん、本来有罪となるべき人まで無罪としてしまう、つまり第一種の過誤を犯す確率を上げることになってしまいます。

2つの過誤を犯す確率について、バランスをとった標本の大きさを考える必要があるということです。

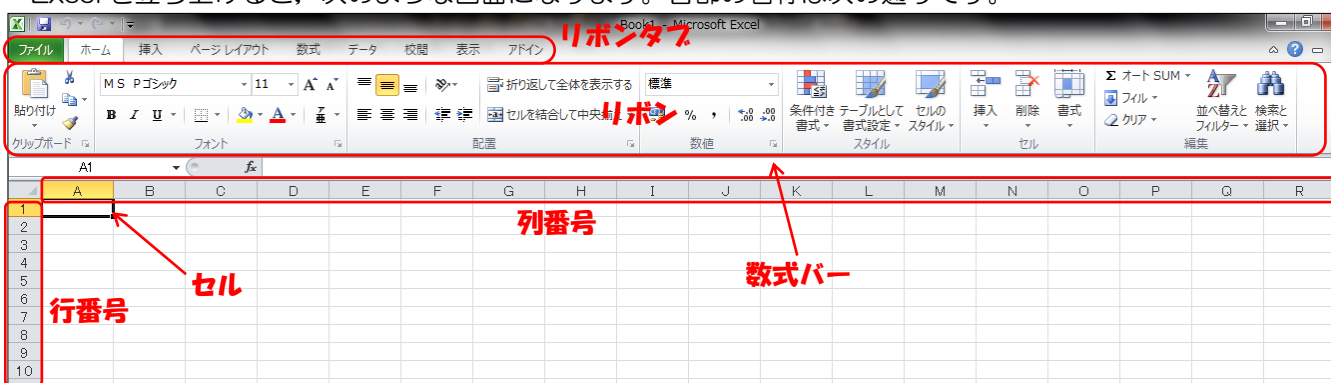
【EXCEL（表計算ソフト）を使ってみよう】

🐾 Excel で大量のデータを処理しよう

Excel（エクセル）は、Windows でおなじみの米国 Microsoft 社が開発したアプリケーションソフトです。表計算ソフトと呼ばれるものの 1 つに分類されます。表計算ソフトとは、データを入力して様々な方法で集計させたり、数式を入力して計算させたりすることができるソフトです。

1. 画面の説明

Excel を立ち上げると、次のような画面になります。各部の名称は次の通りです。



① リボンタブ

「ファイル」、「ホーム」、「挿入」、「ページレイアウト」、「数式」、「データ」、「校閲」、「表示」、「アドイン」などを切り替えられます。

② リボン

いろいろな操作をするためのボタンや、ドロップダウンリストが機能ごとに各タブに格納されています。

③ セル

Excel のデータを入力できる場所で、文字や数字、数式などを入力できます。

④ 数式バー

選択（クリック）したセルの内容が表示されます。数式バーをクリックすると内容を編集できます。

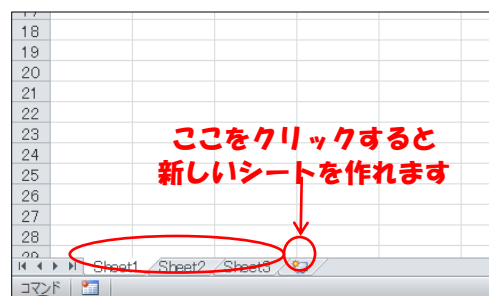
⑤ 行番号・列番号

行番号は数字で行の位置を、列番号はアルファベットで列の位置を示します。列番号は、A から始まり Z の次は AA, AB,のように続きます。

2. ブックとワークシート

Excel の表のファイルのことを**ブック**といいます。Excel を立ち上げると、画面の左下に右図のように「Sheet1」、「Sheet2」、「Sheet3」と表示されていることがわかります。

これらを**ワークシート**と呼びます。それぞれのワークシートで別々の作業をすることができるし、ワークシート間で関連するデータを扱うこともできます。

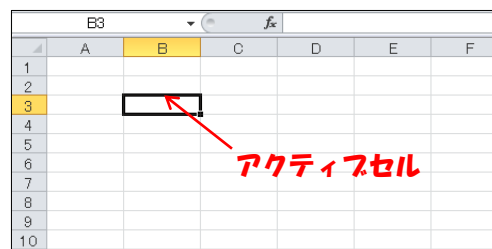


3. データの入力

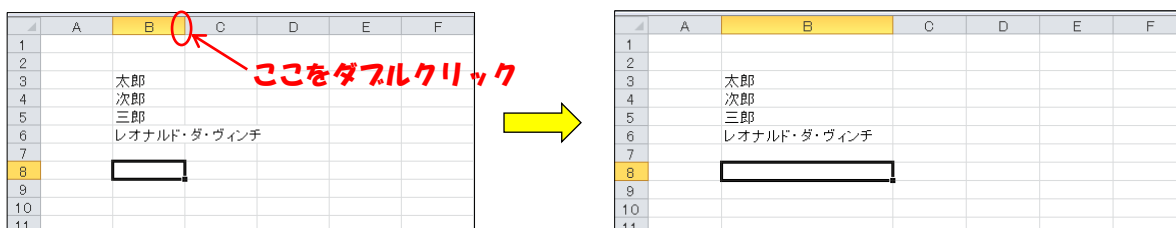
(1) 入力の基本

データを入力したいセルをクリックすると、そのセルが太い枠で囲まれます。このセルを**アクティブセル**といいます。

キーボードをおして文字を入力したとき、このアクティブセルに入力されます。



B列に文字を入力したのが下の図です。「太郎」、「次郎」、「三郎」についてはセルの枠内におさまっていますが、「レオナルド・ダ・ヴィンチ」は枠からはみ出してしまっています。これを枠内におさめたいのならば、列番号の右端をダブルクリックします。



(2) 入力の修正

入力した文字を修正するには、次の3通りの方法があります。

- ① 修正したいセルをダブルクリックする。
- ② 修正したいセルをクリックしたあと、キーボードのF2を押す。
- ③ 修正したいセルをクリックしたあと、数式バーをクリックする。

(3) 入力内容の削除

入力した文字などをそっくりそのまま消したいときは、そのセルをクリックして選択したあと、キーボードの「Delete」キーを押します。

(4) 連続したデータの入力

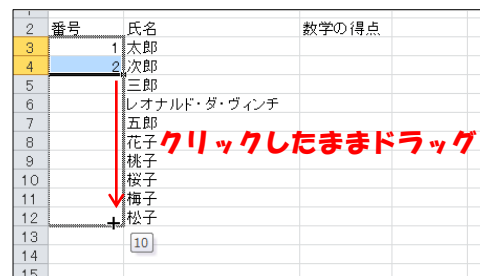
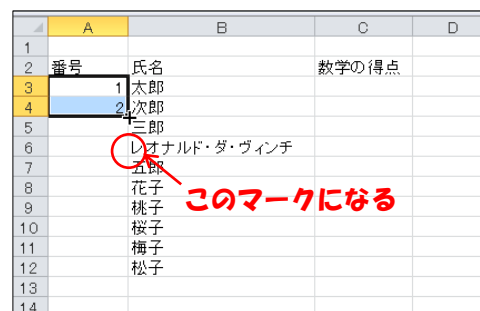
Excelには入力を簡略化するため、**オートフィル**という機能があります。

例えば、右の図でA列の「番号」の下に「太郎」から「松子」まで通し番号を振りたい場合、1から10までの数を直接入力するのは大変なので、オートフィルの機能を使って10までの数を簡単に入力することができます。

アクティブセルの太枠の右下の角にマウスカースールをおくと、図のような十字のマークに変わります。

このマークが変わったら、角をクリックしたまま値を入力したい方向にマウスを動かすと、右の図のようになり、1から10までの数字があっという間に入力できます。

Excelでデータ処理をするときに、高頻度で使う機能なので、覚えておきましょう。



(5) フォントなどの変更

フォントの種類や文字の大きさを変更する場合、リボンの「ホーム」タブ内で操作できます。セル内で文字を中央に配置したり、セルの色を変えたりすることも「ホーム」タブ内でできるのでいろいろ試してみましょう。

4. データの処理～数式の使用～

(1) セルの参照

それぞれのセルには、番地が割り振られています。行番号・列番号を組み合わせて、「A1」、「B6」などと表します（例えば、右の図のアクティブセルは G3 セルです）。

これがセルの番地

	A	B	C	D	E	F	G	H
1								
2		番号	氏名	数学の得点	英語の得点	社会の得点	合計	
3		1	太郎	55	23	86		
4		2	次郎	98	46	21		
5		3	三郎	12	68	89		
6								
7								

(2) 計算と数式

Excel が「表計算ソフト」と呼ばれるように、その中心となるのは「計算機能」です。Excel で計算・数式を扱うには、以下のような決まりがあります。

① 数式（計算式）の入力は、**数式バー**で行う。

② 数式は、必ず「=」（半角のイコール）から始め、「=100+200+300」のように入力します（ちなみに、このように入力したあとエンターキーを押すと、そのセルには「600」と表示されます）。

③ 通常は、数式バーに入力した数式が表示され、その答えがセルに表示されます。

※ この例では、加法ですが、減法の場合は「=200-100」、乗法の場合は「=200*100」、除法の場合は「=200/100」と記述します。

(3) 得点の合計を計算してみよう

上の図の G3 セルから G5 セルにかけて、太郎、次郎、三郎の 3 教科の得点の合計を入力したい場合、数式を利用することができます。

① 直接入力の場合

G3 セルをクリックして、数式バーに「=55+23+86」と入力すると、計算結果 164 が G3 セルに表示されます。

	A	B	C	D	E	F	G	H
1								
2		番号	氏名	数学の得点	英語の得点	社会の得点	合計	
3		1	太郎	55	23	86	164	
4		2	次郎	98	46	21		
5		3	三郎	12	68	89		
6								
7								

もっとも簡単な方法ですが、例えば、あとで得点の入力ミスが発覚して一部のデータを修正しなくてはいけない事態になったとき、合計までは修正されないの、もう一度同じように入力しなくてはならないので面倒です。

② セルの参照機能を利用する場合

右の図のように、G3 セルをクリックして、数式バーに「=D3+E3+F3」と入力しても①と同じ結果が得られます。これは、『D3, E3, F3 に入力されている値をすべて足しなさい』という命令になるので、あとで得点を修正するようなことがあっても、修正後に合計の値も追従して変更されます。

	A	B	C	D	E	F	G	H
1								
2		番号	氏名	数学の得点	英語の得点	社会の得点	合計	
3		1	太郎	55	23	86	164	
4		2	次郎	98	46	21		
5		3	三郎	12	68	89		
6								
7								

③ 関数を使う場合

Excel には「**関数**」と呼ばれる強力な機能が用意されています。関数とは、目的の計算をより簡単にできるようあらかじめ Excel に用意されている数式のことです。これを使えば、面倒な計算も簡単に行うことができます。

例えば、複数の値の合計を計算する「**SUM 関数**」という関数があり、上の図のように G3 セルに

「=SUM(D3:F3)」

と入力すると、①、②と同じ結果が得られます。「D3:F3」は『D3 セルから F3 セルまでの範囲』を意味します。これは、

「=SUM(D3,E3,F3)」

と入力しても同じです（ただし、この方法はセルが多くなると大変です）。

	A	B	C	D	E	F	G	H
1								
2		番号	氏名	数学の得点	英語の得点	社会の得点	合計	
3		1	太郎	55	23	86	164	
4		2	次郎	98	46	21		
5		3	三郎	12	68	89		
6								
7								

5. 関数を使ってみよう

(1) 関数の種類

リボンの「数式」タブの「関数の挿入」をクリックすると、関数の一覧を見ることができます。関数は、扱うデータの種類に応じて、いくつかに分類されています（例えば、数値を扱う関数は「数学/三角」という分類に、文字列を扱う関数は「文字列操作」という分類に入っています）。

(2) 関数の基本構造

さきほどの SUM 関数を例に説明します。

SUM(* * * *)

引数

SUM 関数の場合、かっこ内にはセル番地（セル範囲）を入力します。

関数は、一般的にその関数名のあとにかっこをつけて、かっこ内にその関数が必要とするものを入力します。かっこ内に入れるべき値は**引数**（ひきすう）と呼ばれ、その種類は数値、文字列、セル参照などさまざまです。

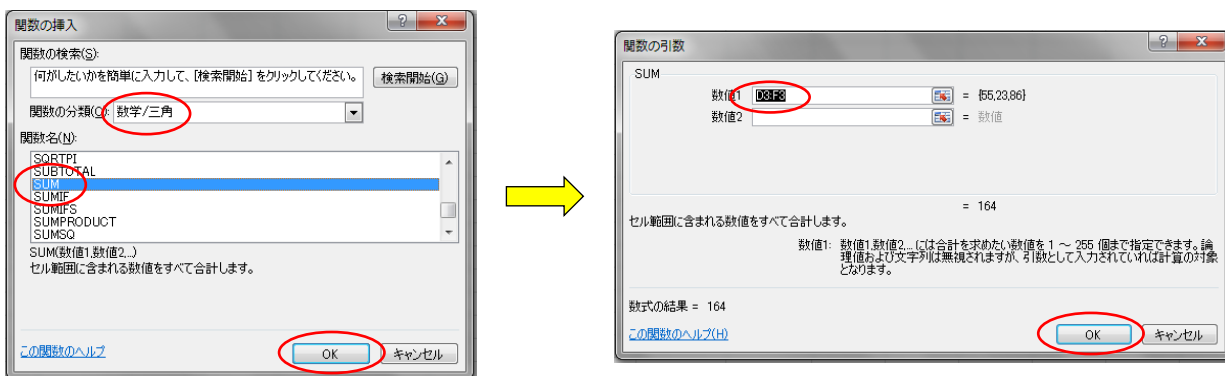
(3) 関数を使ってみよう

セルに関数を入力する場合、大きく分けて 2 通りの方法があります。

① 関数のリストから選ぶ方法

リボンの「数式」タブから「関数の挿入」を選択すると、関数の一覧が表示されるので、その中から必要な関数を選んでセルに挿入することができます。

例えば、さきほどの SUM 関数は、「数学/三角」の分類にあるので、その中から「SUM」を選択し、右下の「OK」をクリックすると、下の図の右側のようなウィンドウが表示されます。



その関数が必要としている引数（この場合はセル範囲）を入力し、右下の「OK」をクリックすると、シート上の選択したセルに関数が挿入されます。

② 直接入力する方法

数式バーに関数を直接入力することもできます。慣れてくるとこちらの方が圧倒的に早いです。

(4) よく使う関数の例

SUM 関数も非常によく使いますが、それ以外によく使う関数を挙げておきます。

① AVERAGE 関数

指定した範囲の数値の平均値を求める関数です。

右の図の D6 セルに、3 人の数学の得点の平均点を表示したい場合、

「=AVERAGE(D3:D5)」

と入力すれば OK です。

	A	B	C	D	E	F	G	H
1								
2		番号	氏名	数学の得点	英語の得点	社会の得点	合計	
3		1	太郎	55	23	86	164	
4		2	次郎	98	46	21	165	
5		3	三郎	12	68	89	169	
6			平均点	55				
7								

② RANK 関数

指定した数値が、指定した範囲の中で第何位になるのかを求める関数です。

右の図の H3 セルに、太郎の合計点が 3 人のうち何位になるのかを表示したい場合、

「=RANK(G3,G3:G5,0)」

と入力すれば OK です。引数がいくつかある場合、上のように「,」で区切ります。この場合、

『G3 セルの値が、G3 から G5 までの範囲で第何位かを表示しなさい』

という意味になります。3 番目の引数の「0」は、順位を得点の高い方を 1 位として第何位になるかを表示するよう指定するためのものです（これを「1」にすると、得点の低い方を 1 位とした場合が表示されます）。

	A	B	C	D	E	F	G	H	I
1									
2		番号	氏名	数学の得点	英語の得点	社会の得点	合計	順位	
3		1	太郎	55	23	86	164	3	
4		2	次郎	98	46	21	165		
5		3	三郎	12	68	89	169		
6			平均点	55					
7									

③ MIN 関数, MAX 関数

指定した範囲の最小値、最大値を求める関数です。

右の図の D7 セルには、

「=MIN(D3:D5)」

と入力されています。

	A	B	C	D	E	F	G	H	I
1									
2		番号	氏名	数学の得点	英語の得点	社会の得点	合計	順位	
3		1	太郎	55	23	86	164		
4		2	次郎	98	46	21	165		
5		3	三郎	12	68	89	169		
6			最高点	98					
7			最低点	12					
8									

④ COUNTIF 関数

指定した範囲に、指定した値（文字列）がいくつあるかを求める関数です。

例えば、右の図の E12 セルに、男子が何人いるか表示したい場合、

「=COUNTIF(E4:E11,"男")」

と入力します（上のように文字列の個数を数える場合、ダブルクォーテーションで囲む必要があります）。

	A	B	C	D	E	F
1						
2						
3					性別	
4				太郎	男	
5				花子	女	
6				松子	女	
7				次郎	男	
8				三郎	男	
9				五郎	男	
10				桃子	女	
11				梅子	女	
12				男性	4	人
13				女性		人
14						

⑤ MEDIAN 関数, MODE 関数

中央値および最頻値を求める関数です。いずれも引数として、MEDIAN(セル範囲), MODE(セル範囲)として使います。

ここからは、少しレベルが高くなりますが、よく使うものなので今のうちに覚えておきましょう。

⑥ IF 関数

指定した条件（論理式）によって、対象が真（TRUE）か偽（FALSE）かを判定し、それぞれに指定した値を求める関数です。

書式は、IF(論理式, 真の場合, 偽の場合)となります。

言葉だけではわからないと思うので、実際の使い方を見てみましょう。

右の図で、E 列に、数学の得点が 60 点以上なら「合格」、60 点未満なら「不合格」と表示させたいとします。このとき、

「=IF(D3>=60,"合格","不合格")」

のように入力すれば OK です。つまり、「D3>=60」（D3 は 60 以上）という式が真ならば「合格」と表示させ、偽ならば「不合格」と表示させる指示になっています。太郎の数学の得点は 55 点で、「D3>=60」を満たさない（つまり偽である）ので、第 3 引数の「不合格」が適用され、「不合格」が表示されています。

	A	B	C	D	E	F	G
1							
2		番号	氏名	数学の得点	合格		
3		1	太郎	55	不合格		
4		2	次郎	98	合格		
5		3	三郎	12	不合格		
6							

ちなみに、IF 関数は、関数の構文の中にさらに IF 関数を入れることができます（こういう構造を、**入れ子構造**といいます）。

例えば、右の図の数式バーを見ると、IF 関数の中に IF 関数を入れていますが、これは、『数学が 60 点以上の場合は（国語は何点であっても）「合格」、数学は 60 点未満だが国語は 60 点以上なら「数学がんばれ」、両方 60 点未満なら「不合格」と表示せよ』という命令になります。

	A	B	C	D	E	F	G	H
1								
2		番号	氏名	数学の得点	国語の得点	合格		
3		1	太郎	55	65	数学がんばれ		
4		2	次郎	98	24	合格		
5		3	三郎	12	30	不合格		
6								

⑦ VLOOKUP 関数

リストや表の指定した範囲の中で、左端の列内から指定した文字や数を検索し、指定した列からそれに該当するデータを求める関数です。

これも例を見てみます。

右の図で、H2 セルに名前を入力すると、I2 セルにその人の数学の得点が表示されるようにしたいとき、

「=VLOOKUP(H2,C3:E5,2,FALSE)」

と入力します。この意味は、

『H2 に入力された名前を、範囲 C3:E5 の左端の列 C から探し、その範囲の 2 列目の文字（数値）を求める』

となります（「FALSE」は、H2 に入力したものと完全一致したものを検索することを意味しています。ここを「TRUE」にした場合は完全一致ではなく部分一致となります）。

	A	B	C	D	E	F	G	H	I	J
1									得点	
2		番号	氏名	数学の得点	国語の得点	氏名		太郎	55	
3		1	太郎	55	65					
4		2	次郎	98	24					
5		3	三郎	12	30					
6										

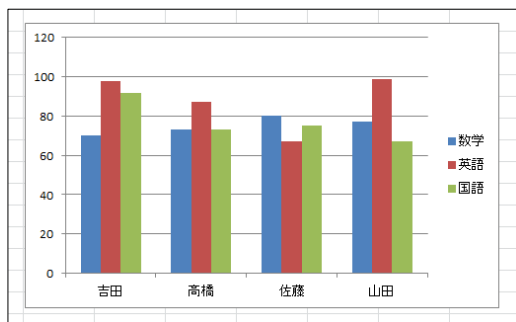
6. グラフをつくってみよう

Excelには、つくったデータを元にして、棒グラフや円グラフなど、いろいろなグラフをつくる機能が備わっています。

ここでは、

	A	B	C	D	E	F	G
1							
2			数学	英語	国語	合計	
3		吉田	70	98	92	260	
4		高橋	73	87	73	233	
5		佐藤	80	67	75	222	
6		山田	77	99	67	243	
7							

という表をもとにして、



というグラフをつくる手順を説明します。

① 表のグラフ化したい部分を選択する

ここでは、各生徒の数学・英語・国語の得点を棒グラフにしたいので、項目も含まれるように B2 から E6 までの範囲を長方形になるように選択します。

ここを選択

	A	B	C	D	E	F	G
1							
2		数学	英語	国語	合計		
3		吉田	70	98	92	260	
4		高橋	73	87	73	233	
5		佐藤	80	67	75	222	
6		山田	77	99	67	243	
7							

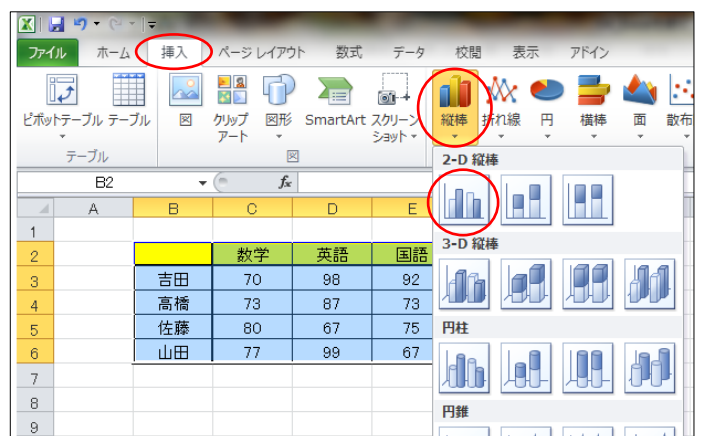
② リボンの「挿入」タブから表示したいグラフを選択する

グラフに必要な範囲を選択したら、リボンの「挿入」タブから、「縦棒」を選びます。

そうすると、棒グラフの種類が選択できるようになるので、ここでは「2-D 縦棒」のいちばん左のものを選びましょう。

.....実は手順はこれだけです。たった 2 ステップで、さきほどのグラフをつくることができます。

あとは、このグラフのレイアウトなどに手を加えたい場合に次のステップに進みます。



③ レイアウトを変える



グラフの軸を変更したり、棒の色を変えたり、パーツをつけ足したりしたい場合は、できあがったグラフを選択した状態で、リボンタブとして加わった「グラフツール」を使って修正していきます。

あとは、いろいろと試してみるとよいでしょう。

いよいよ実戦編！

ここまで学んだことを使うと、ある程度の操作は Excel を使って行うことができます。Excel を使って、以下の問題にチャレンジしてみましょう。

問題 1 : 数式を利用してみる

- Excel を起動し、右の図のように入力しましょう。
- D 列にある各日付の「残席」の欄に、定員から各日付の申込者数を引いた数が表示されるようにしましょう。
- D9 セルに表示されている数式を、下の欄に記入しましょう。

	A	B	C	D	E
1	日別申込者数一覧				
2			定員	40	
3					
4	日付	曜日	申込者数	残席	
5	8月20日	月		16	
6	8月21日	火		37	
7	8月22日	水		9	
8	8月23日	木		40	
9	8月24日	金		25	
10					
11					

問題 2 : 合計や平均を求める

- Excel を起動し、右の図のように入力しましょう。
- D, E 列と 5, 6 行目の各合計と平均を、関数を使って表示してみましょう。
- E5 セルに表示されている数式を、下の欄に記入しましょう。

	A	B	C	D	E
1	商品	上半期	下半期	合計	平均
2	りんご	5800	2790		
3	みかん	4580	4680		
4	バナナ	7540	7720		
5	合計				
6	平均				

問題 3 : 代表値の計算

- Excel を起動して、右の図のように入力しよう
[ヒント] B10 セルの「8.0」はそのまま入力してもこのように表示されません。小数第 1 位が表示されるようにするためには、B10 セルを右クリックして表示されるメニューから「セルの書式設定」を選択し、「表示形式」タブから「数値」を選んで、その右にある「小数点以下の桁数」を 1 に設定して「OK」をクリックします。
- E2, E3 および E5, E6 セルに、男子（番号 1~15）、女子（番号 21~35）の平均値、中央値が表示されるように関数を入力しましょう。
- E6 セルに表示されている数式を、下の欄に記入しましょう。

	A	B	C	D	E
1	番号	A組			
2	1	8.2		男子平均	
3	2	8.5		男子中央値	
4	3	8.5			
5	4	9.3		女子平均	
6	5	7.7		女子中央値	
7	6	8.3			
8	7	7.9			
9	8	8.6			
10	9	8.0			
11	10	7.2			
12	11	7.6			
13	12	6.6			
14	13	7.7			
15	14	7.8			
16	15	8.1			
17	21	7.3			
18	22	7.4			
19	23	7.5			
20	24	8.2			
21	25	7.0			
22	26	7.0			
23	27	8.1			
24	28	6.5			
25	29	6.9			
26	30	7.4			
27	31	9.0			
28	32	7.5			
29	33	7.8			
30	34	7.5			
31	35	8.6			

問題4：ちょっと発展その1

- (1) 問題3の表を使って、E8セルに、A組の中で50m走の記録が8.0秒以下の人数を表示させましょう。

[ヒント] COUNTIF関数を使います。

「8.0秒以下」は、Excelでは「<=8」と入力することで表現できます。これらをうまく組み合わせてみましょう。

- (2) E8に表示されている数式を、下の欄に記入しましょう。

	A	B	C	D	E
1	番号	A組			
2	1	8.2	男子平均		
3	2	8.5	男子中央値		
4	3	8.5			
5	4	9.3	女子平均		
6	5	7.7	女子中央値		
7	6	8.3			
8	7	7.9	8.0秒以下の人数		
9	8	8.6			
10	9	8.0			
11	10	7.2			
12	11	7.6			
13	12	6.6			
14	13	7.7			
15	14	7.8			
16	15	8.1			
17	21	7.3			
18	22	7.4			
19	23	7.5			
20	24	8.2			
21	25	7.0			
22	26	7.0			
23	27	8.1			
24	28	6.5			
25	29	6.9			
26	30	7.4			
27	31	9.0			
28	32	7.5			
29	33	7.8			
30	34	7.5			
31	35	8.6			

問題5：ちょっと発展その2

- (1) Excelを起動して、右の図のように入力しましょう。

- (2) G2セルに「男子」か「女子」を入力し、G3セルに名前を入力すると、H3セルにその人の数学の得点が表示されるようにしましょう。

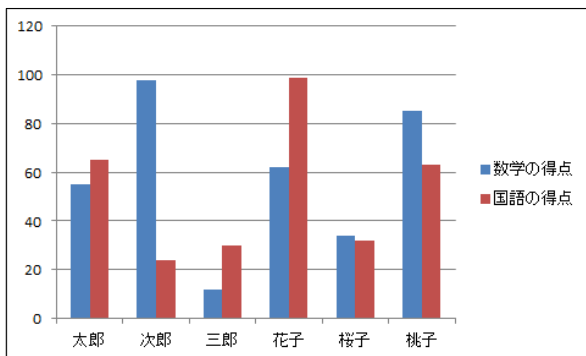
[ヒント] IF関数とVLOOKUP関数を組み合わせてみましょう。

	A	B	C	D	E	F	G	H
1								
2		番号	氏名	数学の得点	国語の得点		女子	
3		1	太郎	55	65		桜子	34
4		2	次郎	98	24			
5		3	三郎	12	30			
6		1	花子	62	99			
7		2	桜子	34	32			
8		3	桃子	85	63			
9								

- (3) H3に表示されている数式を、下の欄に記入しましょう。

問題6：グラフの練習

- (1) 問題5の表を利用して、下のようなグラフを作ってみましょう。



さらに学習したい人へ……

1. 分析ツールを使ってみよう

せっかく Excel を使うので、これまで学習した度数分布表やヒストグラムも作れたらいいですね。そこでここでは発展編として、「分析ツール」を紹介します。

ただ、この機能は Excel の初期状態で装備されているものではないので、少し操作が必要です。まず、リボンタブの「データ」をクリックしてください。



上の図のように、「データ」タブの右端に「ソルバー」、「データ分析」と表示されていれば準備完了です。もしこの表示がなければ、以下の作業をしてください。

- (1) リボンの「ファイル」タブをクリックし、「オプション」を選択。
- (2) 左端のメニューから「アドイン」を選択。
- (3) 下方の「設定」をクリック。
- (4) 「有効なアドイン」の一覧から、「ソルバーアドイン」と「分析ツール」にチェックを入れ、「OK」をクリック。

これで、分析ツールが有効になります。

2. ヒストグラムをつくってみよう

ここでは、p.64 の問題3の表について、1 で導入した分析ツールを使って、階級の幅 0.5 秒のヒストグラムをつくってみます。

(1) データの準備

右の図のように、生徒それぞれの 50 m 走の記録を示した表を用意します。

① 階級

ここでは、階級の幅を 0.5 秒にし、

「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」

「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 6 つの階級をつくることを考えます。

この後の操作の都合上、D 列に(階級の最小値-0.1)から(階級の最大値-0.1)まで数値を入力しておきます。

Excel の仕様上、例えば 6.5 秒のような階級の境界にある値は、「6.5 秒以上 7.0 秒未満」の階級ではなく、「6.0 秒以上 6.5 秒未満」の階級に入ってしまいます。上で 0.1 を引いているのは、6.5 秒が「6.5 秒以上 7.0 秒未満」に入るようにするための措置です。

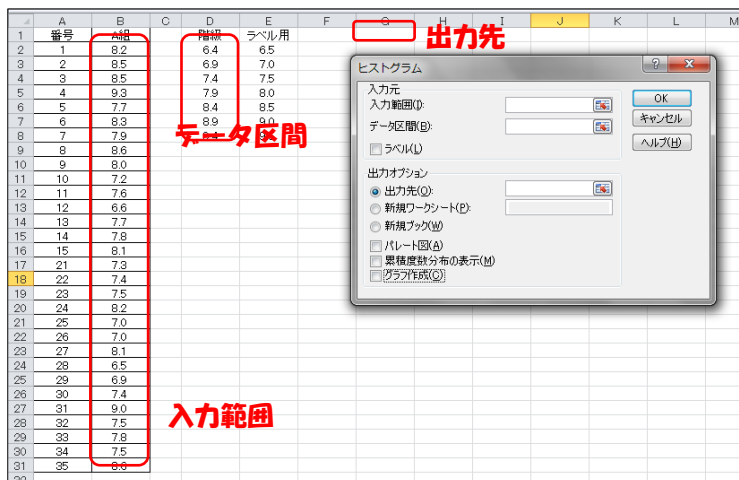
	A	B	C	D	E	F
1	番号	A組		階級	ラベル用	
2	1	8.2		6.4	6.5	
3	2	8.5		6.9	7.0	
4	3	8.5		7.4	7.5	
5	4	9.3		7.9	8.0	
6	5	7.7		8.4	8.5	
7	6	8.3		8.9	9.0	
8	7	7.9		9.4	9.5	
9	8	8.6				
10	9	8.0				
11	10	7.2				
12	11	7.6				
13	12	6.6				
14	13	7.7				
15	14	7.8				
16	15	8.1				
17	21	7.3				
18	22	7.4				
19	23	7.5				
20	24	8.2				
21	25	7.0				
22	26	7.0				
23	27	8.1				
24	28	6.5				
25	29	6.9				
26	30	7.4				
27	31	9.0				
28	32	7.5				
29	33	7.8				
30	34	7.5				
31	35	8.6				
32						

② ラベル用

このままだとヒストグラムをつくったときに横軸に表示される値が D 列に入力したものになってしまうので、横軸のラベルを変更するために、E 列に表示させたい実際の数値を入力しておきます。

(2) ヒストグラムをつくってみよう

ここまで準備ができたなら、リボンの「データ」タブの右端に表示されている「データ分析」をクリックします。表示されるメニューから「ヒストグラム」を選択し、「OK」をクリックすると、右の図のようなウィンドウが出ます。



① 入力範囲

実際にグラフ化したいデータの数値を選択します。📁をクリックして、B2 から B31 までの範囲を選択してください。

② データ区間

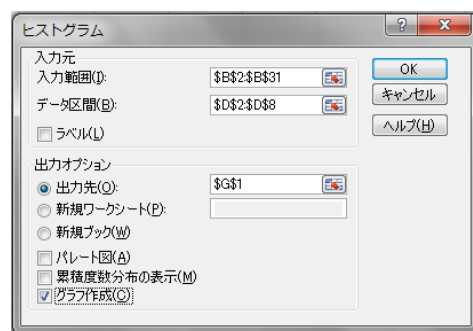
階級を決めるためのデータの数値を選択します。📁をクリックして、D2 から D8 までの範囲を選択してください。

③ 出力先

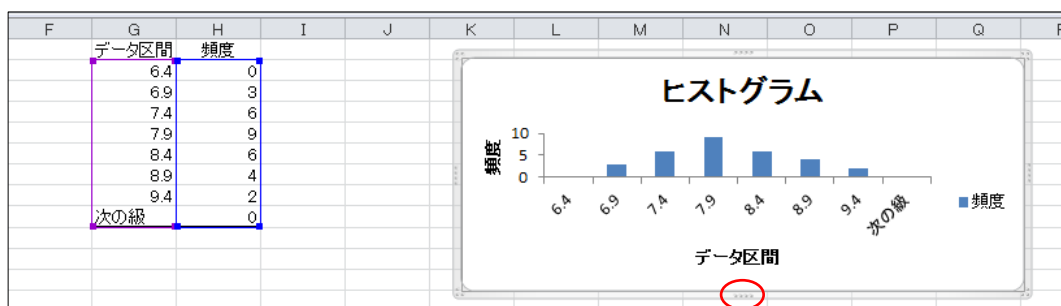
度数分布表を出力するセルを選択します。ここでは、G1 セルを指定します。

④ グラフ作成

チェックボックスにチェックを入れます。



ここまでの作業が完了すると、右の図のようになります。「OK」をクリックすれば、度数分布表とヒストグラムが表示されます。それが下の図です。



このままだとヒストグラムが異様に小さくなっているので、上の赤丸の部分をクリックしながら動かし、グラフの大きさを変えます。

(3) ヒストグラムのレイアウトを修正しよう

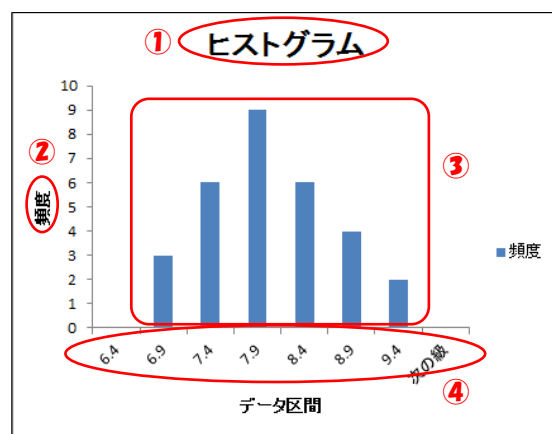
変更したいところは、右の図の①～④の部分です。

① タイトル

タイトル部分をクリックすると編集できます。例えば、「50 m 走の記録」などとするとよいでしょう。

② 縦軸の名称

クリックすると編集できます。ここでは、「人数」としておきます。



③ グラフ

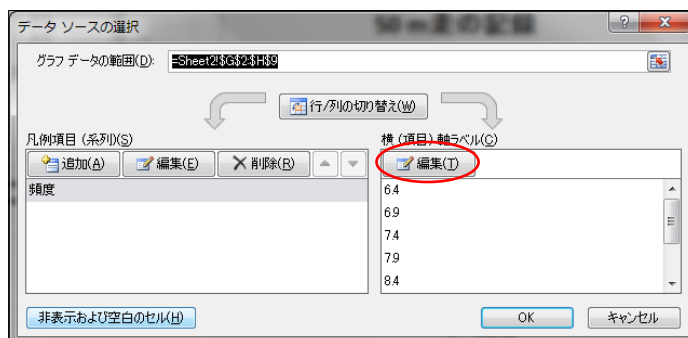
本来ヒストグラムは棒と棒の間隔がないので、棒の間隔を 0 にします。グラフ部分をクリックして、リボンに表示される「デザイン」タブの「グラフのレイアウト」の中から「レイアウト 8」を選択してください。



④ 横軸のラベル

このままだと D 列に入力した階級が表示されているので、ここを E 列に入力した数値に変えます。横軸のラベル上で右クリックし、「データの選択」を選ぶと、右のようなウィンドウが出ます。

「編集」をクリックし、E2 から E8 までの範囲を選択すると、データラベルが変更されます。



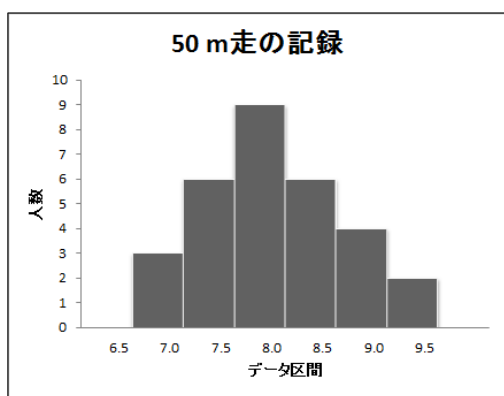
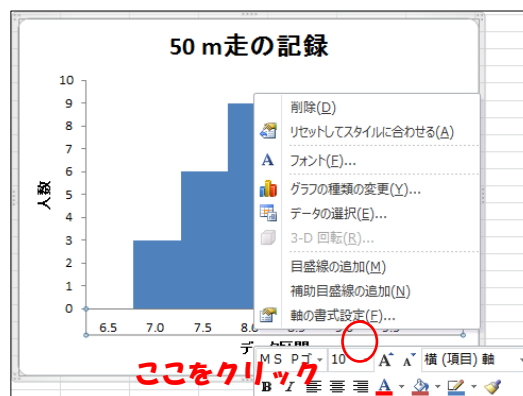
このままだとラベルの位置が棒の真ん中になっていて、以前つくったヒストグラムのようにないので、ラベルの表示位置を変更します。

ラベル上で右クリックすると右のようなメニューが出るので、赤丸をつけた部分をクリックしてください。

⑤ あとはデザインをいろいろ変えてみる

右クリックで表示されるメニューから「○○の書式設定」を選択すると、デザインをいろいろ変更することができます。

で、ひとまず完成したのが下の図です。



【いろいろなグラフを使い分けよう】

総務省統計局ホームページ なるほど統計学園(<http://www.stat.go.jp/naruhodo/index.html>)から抜粋

統計をグラフにあらわそう(種類と特徴)



棒グラフ
折れ線グラフ
円グラフ
帯グラフ
ヒストグラム
レーダーチャート
箱ひげ図
散布図

統計は、データを集めて集計しただけでは、単なる数字の集まりであり、そこから何が読み取れるか必ずしも明らかではありません。

統計を作成するときは、必ず、「〇〇について知りたい!」という目的があるはずですから、**得られた結果を、その目的に合わせて上手に使うことが重要**です。グラフは、結果を視覚的に表す便利な道具であり、グラフをうまく使うことによって、**自分の考えていることを相手に的確に伝える**ことができます。

グラフにはいくつかの種類があり、それぞれ、**得手・不得手があります**。自分が伝えたい目的に応じて、適切なグラフを使うことにより、説明力もぐっと高まります。ここでは、そういったグラフの種類やそれぞれの用途、注意点について説明します。

量の大小をあらわすときには、どのグラフを使えばいいですか？

→**棒グラフ**を使えば、量の大小を、**棒の高さ**であらわすことができます。

増えている、減っているといったことをあらわすには、どのグラフを使えばいいですか？

→**折れ線グラフ**を使えば、**線の傾き**で、増減をあらわすことができます。

割合をあらわすときには、どのグラフを使えばいいですか？

→**円グラフ**や**帯グラフ**を使います。三角グラフを使うこともあります。

データの散らばり具合をあらわすときには、どのグラフを使えばいいですか？

→**ヒストグラム**や**箱ひげ図**を使います。

複数の指標をまとめてあらわすときには、どのグラフを使えばいいですか？

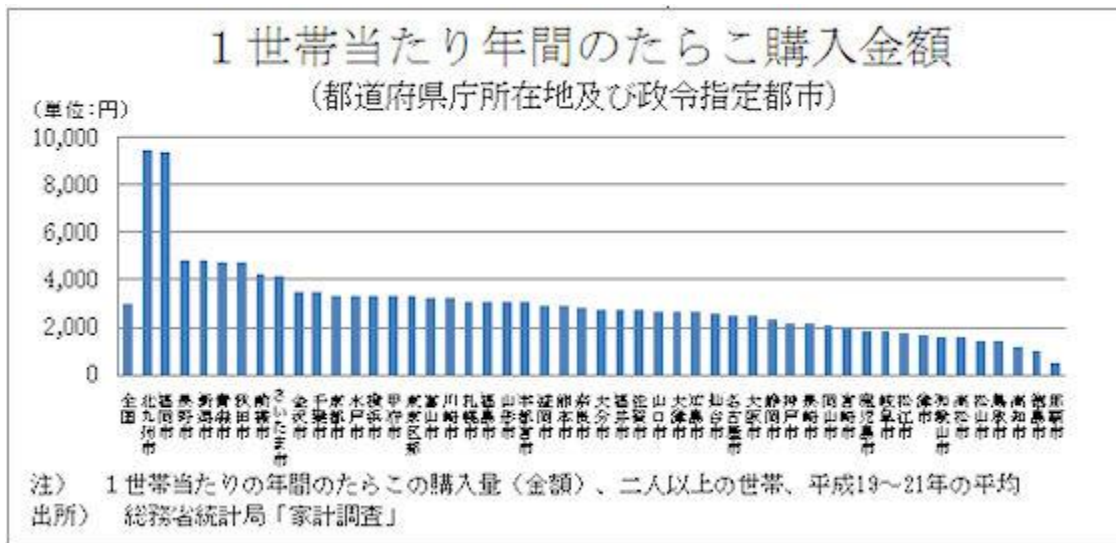
→棒グラフや折れ線グラフでもあらわすことができますが、**レーダーチャート**を使えば、**平均と比較したときにどの項目が大きい**か等をわかりやすくあらわすことができます。

2つの量に関係があるかどうかをあらわすときには、どのグラフを使えばいいですか？

→**散布図**を使います。

棒グラフ

日本列島は東西・南北に長いので、どんなものを食べているか、地域によってかなり特色があります。ここでは、総務省統計局「家計調査」を使って、たらこの消費量にどのような地域特色があるか、みてみましょう。



これをみると、福岡県の北九州市、福岡市でのたらこ購入量が飛び抜けて大きいことが一目でわかりますね。福岡は「辛子明太子」で有名ですから、ご家庭でたらこを買う量も多い、ということでしょう。(「世帯」という言葉は耳慣れないかも知れませんが、「いっしょに住んでいる家族のこと」と思って下さい。)

どんなときに使うの？

棒グラフは、縦軸にデータ量を取り、棒の高さでデータの大小を表したグラフです。(稀に縦横が逆の場合もあります。)

データの大小が、棒の高低で表されるので、データの大小を比較するのに適しています。

上記の例だと、北九州市、福岡市でのたらこ購入量が他と比べてどんなに多いか、一目でわかりますよね。

気をつけることは？

棒グラフを描くときに、データをどの順に並べるかは、とくに決まりはありません。ただし、あまり意味もなく並べてもグラフが見にくいので、

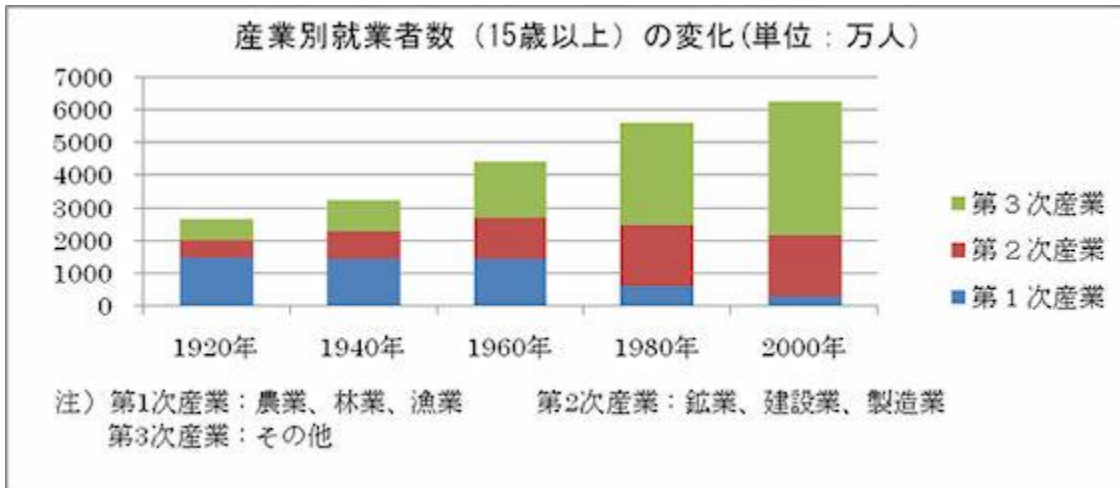
- ① データの多い順(または少ない順)に並べる。・・・上記の例
 - ② 都道府県データの場合、北から順に並べることもある。
 - ③ 五十音順に並べる。
 - ④ 横軸が年や月といった時間軸である場合、時間の順に並べる。
 - ⑤ 質問に対する回答を並べる場合、質問票の順と同じ順に並べる。
- などの方法をとります。

その他には？

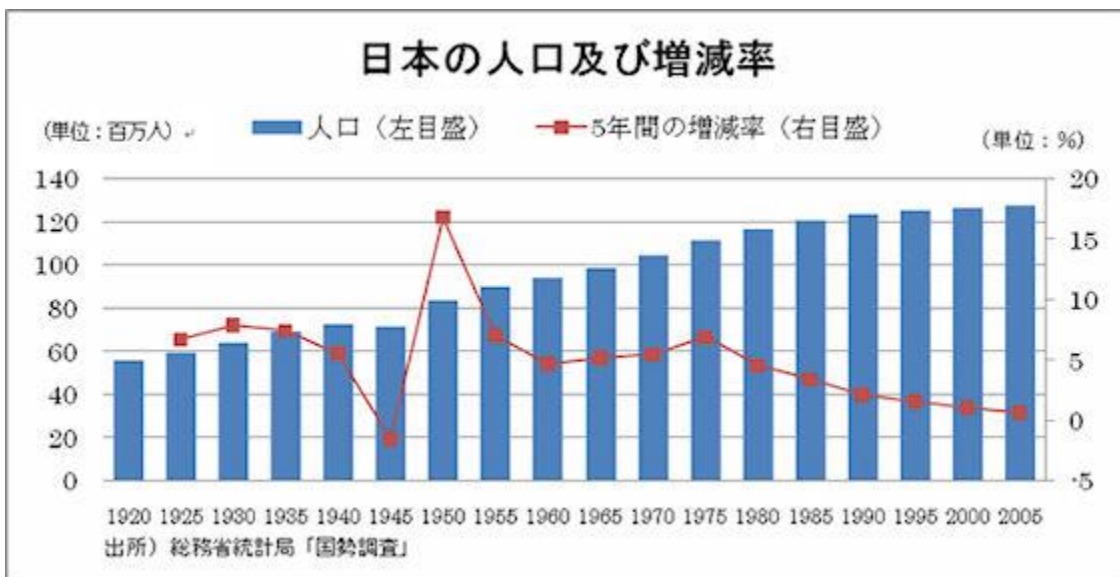
先ほど紹介したのは、棒グラフのうちでも、もっとも単純なものですが、もっと複雑な棒グラフもあります。そういった複雑なグラフのいくつかを紹介しましょう。

①積み上げ棒グラフ

1本の棒に、複数のデータを積み上げて表示したものの。



②棒グラフと折れ線グラフの複合グラフ



やってみよう

[総務省統計局「家計調査」](#)では、たらこ以外の品目についてもデータを掲載しているので、それらについてもグラフを描いてみましょう。

折れ線グラフ

たまごは、昔は「物価の優等生」と呼ばれ、様々なものの値段が変動する中で価格が安定していたのですが、最近では、「たまごの値段が上がった」というニュースも見かけるようになりました。たまごの値段の変化を、総務省「消費者物価指数」で見してみましょう。



ここでは、たまご（鶏卵）の価格の年平均をグラフにしました。ここで使った「消費者物価指数（平成 17 年基準）」は、平成 17 年の平均価格がちょうど 100 になるように計算されています。

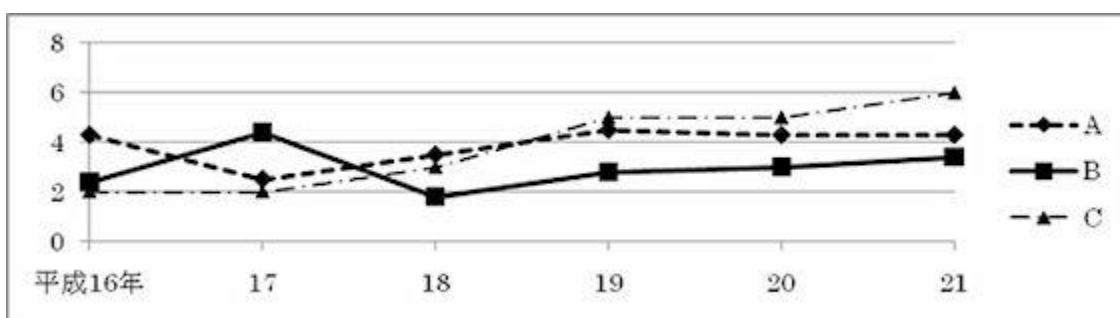
これをみると、たまごの値段もかなり上下動していることがわかります。ニワトリの飼料は輸入穀物に頼っていること、養鶏場の暖房等の費用もかかることから、たまごの価格も、国際的な穀物価格や原油価格の動向に左右されやすい、という側面もあります。

どんなときに使うの？

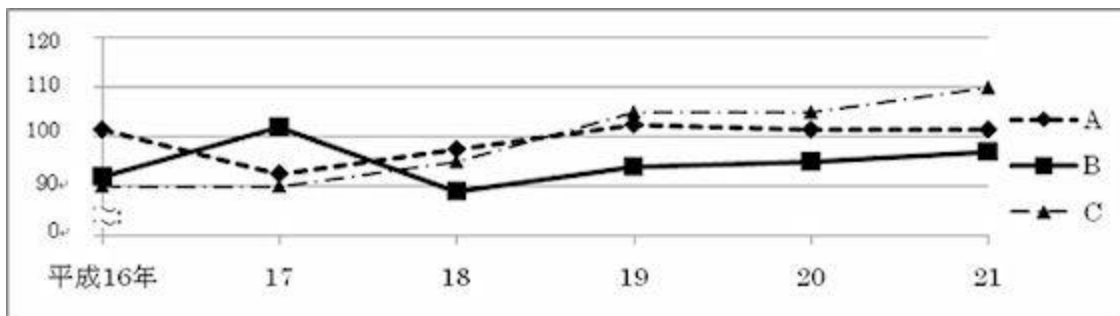
折れ線グラフは、横軸に年や月といった時間を、縦軸にデータ量を取り、それぞれのデータを折れ線で結んだグラフです。線が右上がりならその期間はデータが増加（上昇）、右下がりならデータが減少（下降）していることになるので、データの増減を見るのに適しています。

気をつけることは？

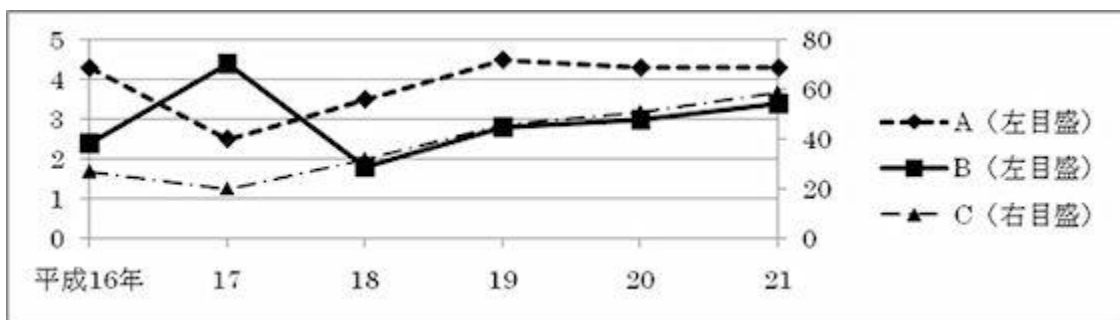
折れ線グラフでは、複数のデータを一つのグラフに重ねて描き、「こちらのデータは増加しているが、こちらは減少」という比較をすることがありますが、その場合、線の区別が付きやすいよう、線を色分けしたり、実線と破線を使い分けたりします。



また、折れ線グラフの縦軸の目盛りは、ゼロから始めるのが普通ですが、あまり数字に変化がなくて見にくいときは、縦軸に波線を入れて途中を省略することもあります。



また、Cだけ数値が大きく違って、一つのグラフに描きにくい、という場合もあります。その場合は、Cだけ右目盛りで表示する、ということにして、一つのグラフに収めることもできます。



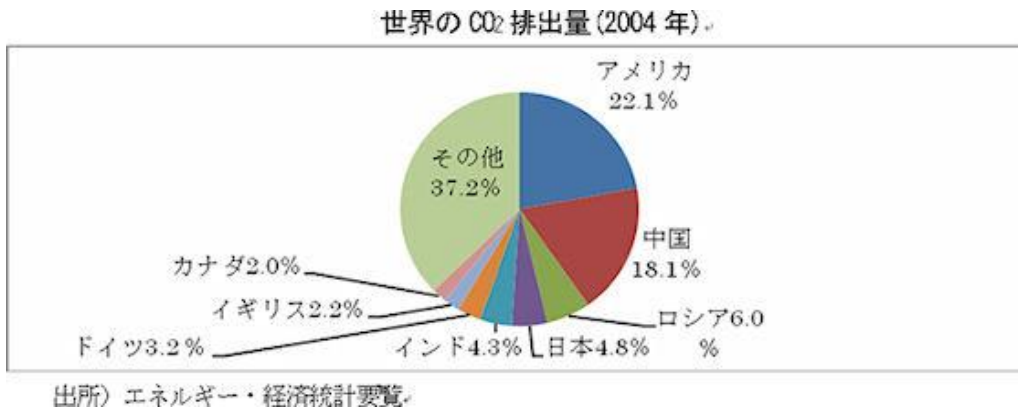
やってみよう

[総務省統計局「消費者物価指数」](#)では、その他の様々な品物の価格を調べています。

たまご以外のいろいろな品物やサービスの価格について調べ、グラフにしてみましょう。

円グラフ

最近、「地球温暖化防止のため、二酸化炭素（CO₂）排出量を減らそう」「日本が率先してCO₂削減に努力すべきだ」とのニュースをよく見かけます。いったい、どの国がどのくらい、二酸化炭素を排出しているのでしょうか。



アメリカと中国のCO₂排出量が飛び抜けて多いこと、上位4か国（アメリカ、中国、ロシア、日本）で全体の半分以上を占めている、等がわかります。

どんなときに使うの？

円グラフは、円を全体として、その中に占める構成比を扇形で表したグラフです。扇形の面積により構成比の大きさがわかるので、構成比を示すのに使われます。

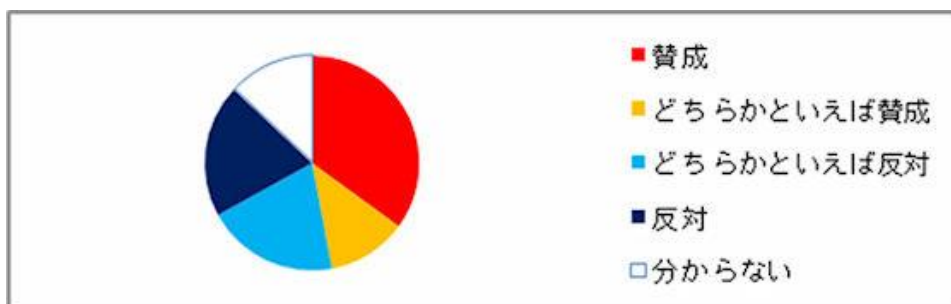
気をつけることは？

データは、時計の針の12時の位置から時計回りに、大きい順に並べます。

「その他」はいくら大きくても、一番最後に表示します。

あまり扇形が小さくなるとグラフが見にくくなるので、構成比が小さいものは、まとめて「その他」にしてもいいでしょう。

なお、「〇〇に賛成ですか」とような質問に対する回答を円グラフに表示する場合など、必ずしもデータの大きい順に並べないこともあります。「賛成」と「どちらかといえば賛成」を合わせて過半数かどうか、ということが一目でわかるようにする方が便利だからです。



円グラフを描くときは、合計が100%になるようにします。左の例で「わからない」を除いてしまうと合計が100%にならなくなりますが、それで円グラフを描いてはいけません。

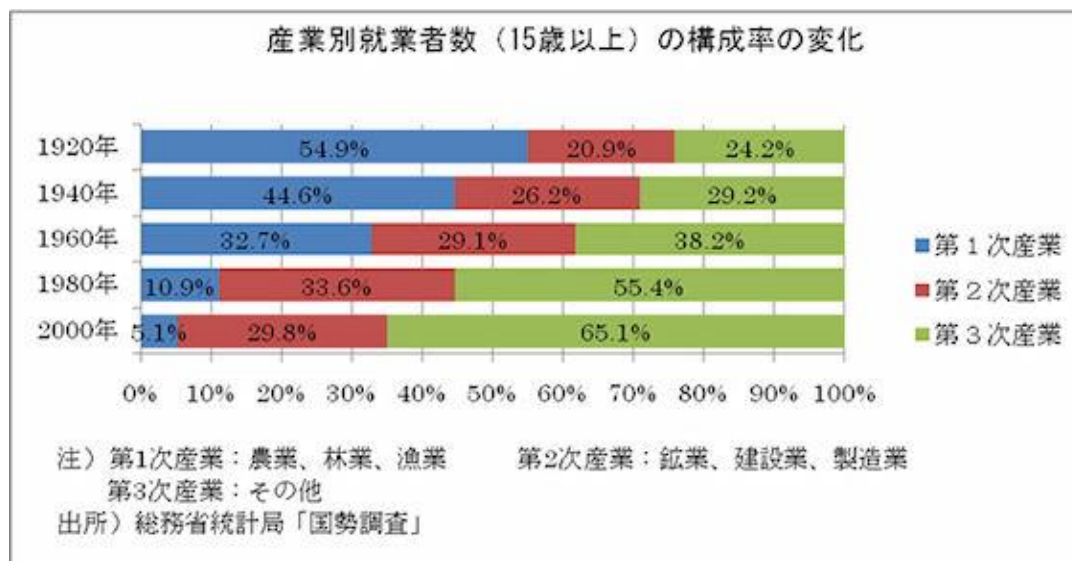
また、複数回答を認める質問の場合、合計が100%を超えますが、このような場合も円グラフにはしてはいけません。

やってみよう

[総務省統計局「家計調査」](#)では、ご家庭で、どのようなものを買っているか、調べています。円グラフを描いて、どのようなものにお金を使っているか、見てみましょう。

帯グラフ

「最近、農業をする人の割合が減ってきた」との話を聞きますが、総務省「国勢調査」で、第1次産業・第2次産業・第3次産業別の就業者数の割合の変化をみてみましょう。



ここでは、仕事をしている人の数ではなく割合を知りたいので、棒の長さをそろえ、それぞれでの構成比をみてみました。1920年では第1次産業で働いている人が全体の半分以上であったのに、この割合は低下を続け、現在では5%程度であることがわかります。

どんなときに使うの？

帯グラフは、長さをそろえた棒を並べ、それぞれの棒の中に構成比を示すことによって、構成比の比較をするためのグラフです。

気をつけることは？

構成比をみるのが目的なので、棒の長さは全て同じにします。

また、項目を並べる順番を途中で変えると、割合の変化がグラフを見てわからなくなってしまうので、一つのグラフの中では順番は変えないでおきます。

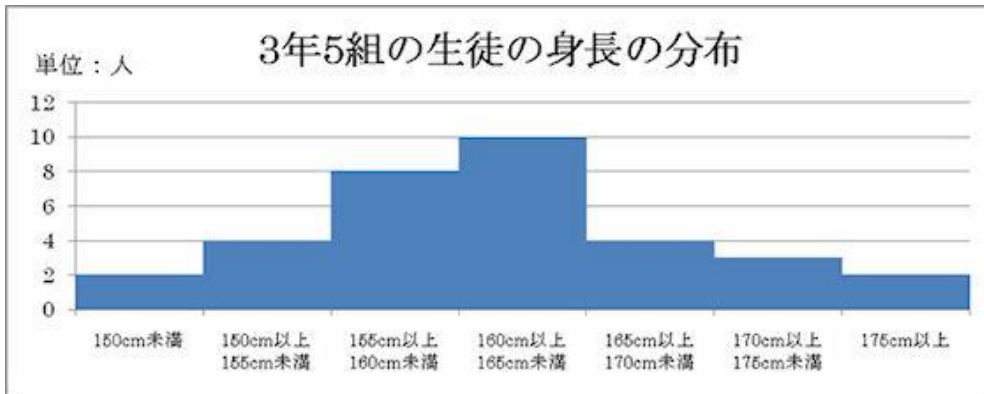
（上の例では、途中から第3次産業の就業者数が一番多くなっていますが、多い順に並べることはせずに、すべての年次にわたって第1次産業・第2次産業・第3次産業、の順に並べています）

やってみよう

[総務省統計局「社会生活基本調査」\(PDF:429KB\)](#)では、1日の時間の使い方（睡眠、食事、学業、仕事、家事、趣味、等）を調べています。年代別に、時間の使い方どのような違いがあるか、帯グラフを描いて比べてみましょう。

ヒストグラム

クラスの中にも、背の高い人、低い人いろいろいます。どれくらいの身長の人が何人いるか、グラフで示してみましよう。



160cm以上 165cm未満の人が一番多いことがわかりますね。

どんなときに使うの？

ヒストグラムは、データの散らばり具合をみるのに使われます。

横軸にデータの階級を、縦軸にその階級に含まれるデータの数（人数、個数など）をとって棒グラフで表します。

気をつけることは？

階級数でヒストグラムを描いてみて、あまりデータの散らばりが見られなければ階級数を変えてみる、という方法をとります。データの階級幅が異なるときには、棒グラフとは異なった扱いが必要です。

例えば、上記の例で、「165cm以上 170cm未満と 170cm以上 175cm未満のところは人数も少ないから、まとめてしまおう」とした場合、「165cm以上 175cm未満 7人」となりますが、それをそのままグラフに描くと、階級をまとめて人数が増えただけなのに、グラフの高さが高くなってしまって、誤解してしまうおそれがあります。そのため、そのような場合には、階級幅が2倍なので横幅を2倍にし、その代わりに高さを半分にして表示します（つまり、面積がデータの個数に比例するように描きます）。

また、グラフの両端（上の例では「150cm未満」と「175cm以上」）も、高さを隣よりも、（その階級の幅は5cmより長いので）低く表示します。

ヒストグラムを描く際には、「階級数をいくつに分ければよいか」というのが問題になります。通常は、5～10くらいの階級数でヒストグラムを描いてみて、あまりデータの散らばりが見られなければ階級数を変えてみる、という方法をとります。

やってみよう

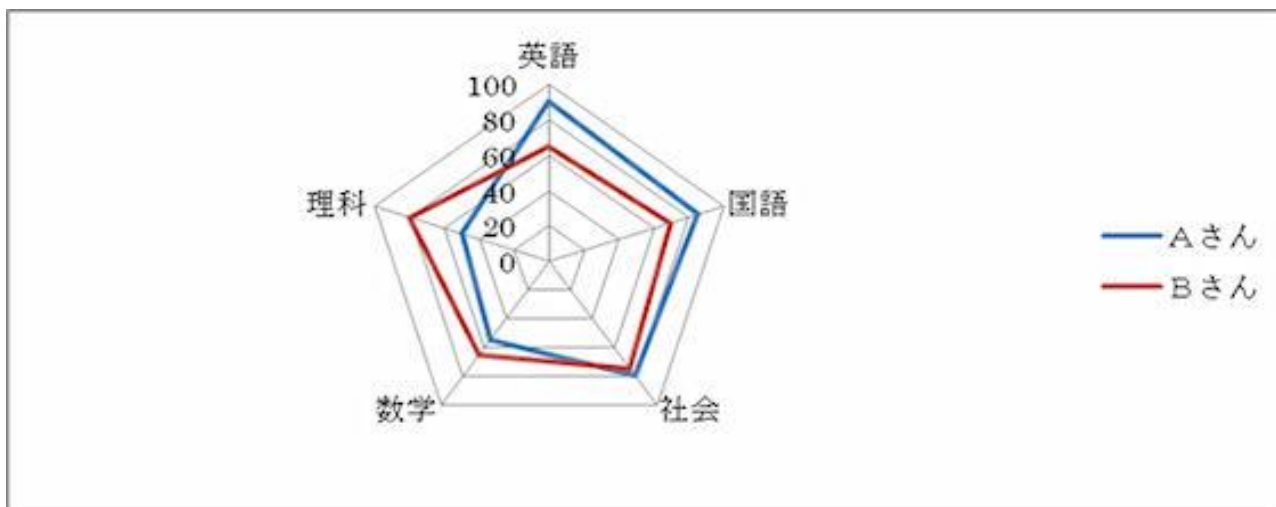
総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。これを使って、

「市町村別の中学校数の分布」などのヒストグラムを描いてみましょう。

[「統計でみる都道府県・市区町村（社会・人口統計体系）」](#)

レーダーチャート

みなさんの中には、「理科系には強いが、文科系がちょっとねえ」とか、逆に「私は文科系の方が得意だ」という人もいるでしょう。そういったことを一目で示せないでしょうか。



テストの成績をレーダーチャートで示すことにより、Aさんは英語や国語といった文科系は得意だが数学や理科といった理科系は苦手、Bさんは逆に理科系が強い、といったことがみてとれます。

どんなときに使うの？

レーダーチャートは、複数のデータ（指標）を一つのグラフに表示することにより、全体の傾向をつかむのに用いられるグラフです。

円を、データの項目数に応じて中心から放射状に線を引き、それぞれの線上に、データを表示します。そしてそれらを線で結んで、その形状を見ます。

気をつけることは？

レーダーチャートを描く際には、通常、「外に行くほど（データ値が大きいほど）良い」となるようにデータを選びます。

例えば、体力測定の結果をレーダーチャートに描く場合、幅跳びや反復横飛びは数字が大きいほど良いのですが、50m走などはタイムが短いほど良いので、単純に50m走のタイムでレーダーチャートを描くとおかしいグラフになってしまいます。そのような場合には、タイムではなくその逆数（つまり、「秒速〇〇m」とか）に直して、「データ値が大きいほど良い」データに変換する必要があります。

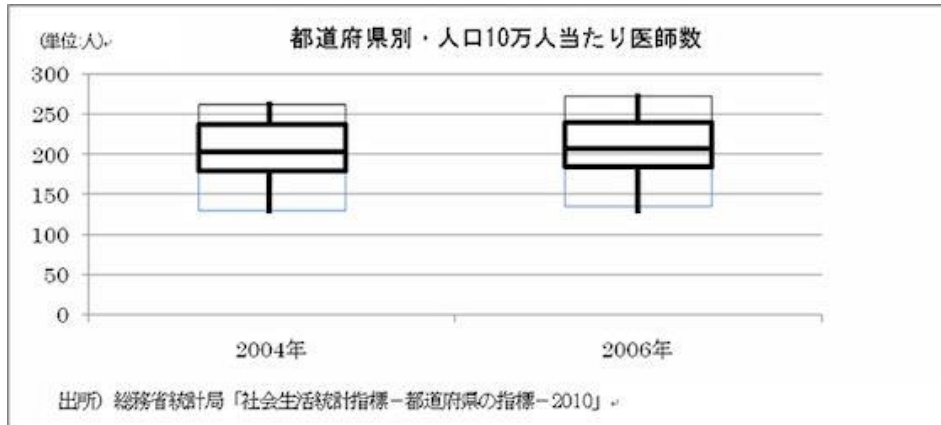
やってみよう

総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。これを使って、皆さんの住んでいる都道府県・市町村の指標をレーダーチャートに描き、全国平均と比べてどうか、見てみましょう。

箱ひげ図

お医者さんの数は、地域的に偏りがあると言われています。

都道府県別に見て、人口10万人当たり医師数（医療施設に従事する医師数）にどの程度ばらつきがあるか、グラフに表してみましょう。



出所) 総務省統計局「社会生活統計指標—都道府県の指標—2010」

四角い箱の上下に、ひげが生えている形をしており、このようなグラフを「箱ひげ図」と呼びます。

この図では、ひげの一番下が最小値、箱の下部の辺が第1四分位数、真ん中の線が第2四分位数（中央値）、上部の辺が第3四分位数、そしてひげの一番上が最大値を表しています。

2004年と2006年とを比べると、全体として増加しているものの、最大値の増加が大きいことが見てとれます。

どんなときに使うの？

データのばらつき具合を比較するのに用います。1つのデータのばらつきを見るだけでしたらヒストグラムでも見られますが、ヒストグラムでは異なる複数のデータのばらつきを比較することは困難です。そのような場合、箱ひげ図により比較することができます。

気をつけることは？

ここで紹介したのは、ひげの両端がそれぞれ最小値、最大値になっている箱ひげ図ですが、中には、上下から10%点をひげの両端として描いている箱ひげ図もあります。

新聞やインターネットで箱ひげ図を見かけたときは、ひげの両端が何を表しているか、きちんと確認するとよいでしょう。

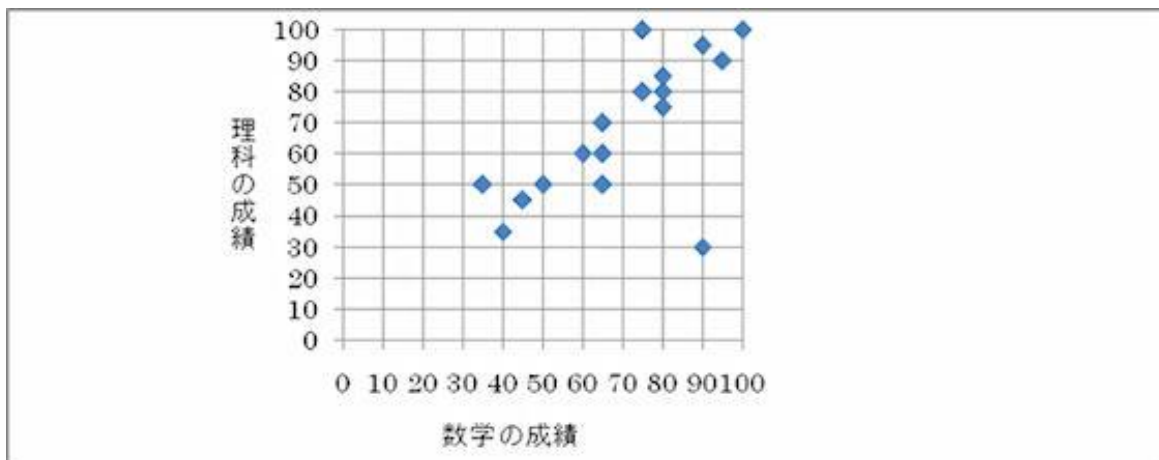
やってみよう

総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。いろいろなデータを使って、箱ひげ図を描いてみましょう。

[「統計でみる都道府県・市区町村（社会・人口統計体系）」](#)

散布図

「数学が得意な人は理科も強い」とよく言われますが、本当にそうなのか、クラスの人の点数をグラフにしてみました。



これをみると、確かに、数学の成績と理科の成績には関係がありそうに見えますね。

どんなときに使うの？

散布図は、縦軸と横軸にそれぞれ別の量を取り、データがあてはまる場所に点を打って示す（「プロットする」といいます。）グラフです。2つの量に関係があるかどうかをみるのに使います。

気をつけることは？

散布図でわかることは、2つの量の間に関係があるかということだけであり、因果関係（どちらかが原因となって、もう一方が起こる）を示すものではありません。

例えば、働いている人について、血圧と給料の関係を散布図に描いたとします。普通は、年齢が高い人ほど血圧が高く、また年齢が高いほど給料も高くなるので、「血圧と給料には関係があります」という散布図になります。

でも、これから、「血圧を高くすると、給料が上がります」ということにはなりませんよね。

やってみよう

総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。これを使って、県民所得や水道普及率、ごみの量などで、散布図を描いてみましょう。

[「統計でみる都道府県・市区町村（社会・人口統計体系）」](#)

グラフをつくるツールとして、下記のサイトも参考するとよい。

○ <http://rika-net.com/>

○ 科学の道具箱

【 統計グラフコンクール出品作品をつくろう！（平成 30 年度版）】

調べ学習ではなく、問題解決学習。根拠となることを数値やグラフを使って表現します。

●出品上の注意

- ア 紙の大きさはB2版（縦横どちらでもよい）。白の模造紙を避ける。
- イ 裏面の板張りや表面のセロハンカバーを使用不可。
- イ 1グループ5人以内。
- ウ グラフ部分は手書き。
- エ 下書きの消し忘れはないか。色紙や絵を貼った場合、しっかり接着してあるか。
- オ 第三者（応募者以外）が作成したイラストや写真等を使用していないか。
- カ 商品・企業・キャラクター・スポーツ選手・芸能人の名前や写真を使用していないか。
- キ 国旗を用いる場合には国旗の縦・横の比率やデザインが正しいか。
- ク 自己の観察又は調査によった場合は、その観察又は調査の記録を別紙（A4版かB5版）として添付する。
- ケ 外から資料を得た場合は、その資料出所を作品表面に明記（URL表記不可）するとともに、統計表（取材資料）を別紙（A4版かB5版）として添付する。

- 1 テーマを決める
- 2 仮説をたてる
- 3 調べる方法と進め方の計画をたてる
- 4 データを集める

自分でデータをとる（いつ・何を、対象は？・どうやって調べたか。写真や絵を加えても。）

公的データを集める（資料の出典を必ずメモしておく→書き添える）

※東京都の統計 <http://www.toukei.metro.tokyo.jp/kidsc/kc-1link.htm>

※統計情報研究開発センター“提供データ” <http://www.sinfonica.or.jp/datalist/index.html>

※政府統計の総合窓口「e-Stat」データを地図上に表示する「jSTAT MAP」など

など

<1～4の構成>

現状の把握 そのテーマを選んだ理由やきっかけ

<解決したい問題が読み取れるグラフ>

<何に着目すればよいか、現象の本質をとらえたグラフ>

因果 どうすればその問題を解決できるか仮説をたてる

<要因をもれなく考える> 議論する

<この要因（横軸）をこうしたら、結果がこうなる（縦軸）がわかるグラフ>

調べた結果（データ・グラフ）とそこからわかったことを書く 議論する

対策 感想・提言・新たな課題

<（できれば）解決案がどの程度の効果をもたらすかを示すグラフ>

- 5 グラフや文章の下書きをして、レイアウトや配色、書くのか塗るのか貼るのかなどの方針を決めていく。

既存のグラフから探す。

データを処理して新たなグラフをつくる。分担可。

議論したことをまとめていく。

●東京都の審査基準

応募作品は、次の基準によって審査されます。

ア 共通基準

- ・誤りはないか（目盛りを書いたか・〇から始まっているか、単位の取り方、凡例を正しく書いてあるか、誤字、脱字、数字の誤り、記入漏れ）
- ・書き落としはないか（資料の出所（URL 表記不可）、観察やアンケート調査の方法、）
- ・的確か（見出し（主題）の表現、配色）

イ 各部別基準（中学生の部に出品）

- （1）統計データを正しく理解し、グラフ化することによってデータのもつ事象が理解されやすくなるよう、工夫されているか。
- （2）訴えたい主題が的確にグラフに表れているか。また、主題は斬新で興味を喚起するものであるか。

●審査方法及び審査員

ア 第一次審査

東京都総務局統計部職員による審査

イ 第二次審査

学識経験者等による審査

ウ 本審査（最終審査）

学識経験者及び関係行政機関の職員による本審査会で入賞作品を決定します。

6 タイトルの見直し（どんなことを取り上げたのかが伝わり、読んだみたい！思えるようなタイトル） サブタイトルをつけるときは、注目してほしいことを短くまとめる

例 1：「日本の人口の将来について ～子どもたちがドンドン減っていく～」

例 2：「地球温暖化・海面上昇 ～沈んじゃう国がある！？～」

7 清書をする

- ・タイトルは本文より大きめの太い文字（レタリング）
- ・本文の説明の文字は全体を通じて同じ色・大きさの方がよい。
- ・図・表・写真・グラフにはタイトルや解説をつける。
- ・色をうまく使って目立たせたい部分を強調する。

8 出品の準備をする

友達や保護者、先生などに誤字脱字がないか、最後に確認してもらおう。

観察・アンケート調査の記録や利用したデータの統計表を別紙に書いたかな。

活動別自己評価

I. 統計学の正しい知識を身につけることの意義を知る

***** 1. ガイダンス～この1年で身につける力～ *****

統計データにおける数値の扱い方や、グラフの見せ方を変えることで、相手に誤解を与えたり、受け取る人の印象が大きく変わってしまったりすることがある。

6つのテーマについてグループで話し合い、統計データやグラフを扱う上でどんな点に注意しなければならないかを発表した。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

II. 記述統計学の基礎を学ぶ

***** 3. 記述統計学 I：資料の整理 *****

大量のデータを、どのように整理したらその集団の特徴をつかみやすいか、集団を特徴づける数値にはどのようなものがあるかを学んだ。

また、平均値・中央値・最頻値といった代表値だけでは見えてこないデータの特徴を、分散や標準偏差を用いることで、他の視点から捉えた。具体的なデータを用いて、分散や標準偏差を計算し、それらの統計量を用いて分析することの意義を学んだ。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

III. 科学レポートの書き方を学ぶ

***** 4. 科学レポートを書いてみよう *****

***** 5. 科学レポートを読んでみよう。 *****

科学レポートや論文を読んで、内容が明確に伝わるようにするためには、どんな工夫をしたらよいか、どのように表現に気をつけるべきか考えた。

また、興味や疑問をもったことを1つとりあげ、簡単な実験をすることで、仮説から研究方法、結果、考察といったレポートの流れを書いた。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

IV. PPDAC サイクル

***** 2. PPDAC サイクルの考え方 *****

***** 6. PPDAC サイクルの実践編 *****

統計的な分析を進めるときの基本的な手順となる PPDAC サイクルについて具体例を通して学び、それを実践した。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

V. グラフコンクールの作品を完成させよう

***** 7. 統計グラフコンクールの作品を作成しよう *****

これまでに学んだ統計的な分析の手法を用いて、東京都グラフコンクールに出品する作品を、制作した。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

VI. 統計グラフコンクールの作品を発表しよう

***** 8. 統計グラフコンクールの作品を発表しよう *****

1年生のフィロIで学んだスピーチを構成するための手法を用いて、統計グラフコンクールに応募した作品について、発表を行った。

1つのグループにつき3分のスピーチにまとめ、自分たちが調べたことや分析したこと、それによってどんな結論が得られ、研究がどのように発展していくかを発表した。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

Ⅶ. 不確かな現象を確率で表現する

***** 9. 確率論の基礎 *****

統計的な分析をする上で、その土台となる確率論の基礎を学んだ。

また、さいころを使って統計的確率と数学的確率の違いを学んだ。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

Ⅷ. 推測統計学の基礎を学ぶ

***** 10. 推測統計学Ⅰ：推定 *****

***** 11. 推測統計学Ⅱ：確率分布 *****

***** 12. 推測統計学Ⅲ：検定 *****

集団の特徴を調査するとき、その一部だけを取り出して全体の傾向を調べる手法を学んだ。

具体的な問題を通じて、標本調査の原理を理解するとともに、視聴率調査の模擬実験を行い、標本の大きさによって、結果がどのように異なるか考えた。統計的仮説検定の手法の基礎を学んだ。

◇自己評価 5 … 4 … 3 … 2 … 1
◆評価の理由

◎ どんな力を身につけたか

◎ コメント・感想

1年間のまとめ

◎ 今後、小石川フィロソフィーⅡで身につけた「どんな力」を「どのような場面」で生かしていきますか。

◎ この1年間、多くの具体例を通じて、数量的な課題探究の基礎的スキル（データの分析、確率論、推測統計学など）を身につけてきました。今後、小石川フィロソフィーⅢ以降での課題研究を行う際、これらの力をどのように発展させていきたいと思いませんか。

◎ 1年間の授業の感想

第2学年 課題研究
小石川フィロソフィーⅡ 共通テキスト

平成30年 第1版 第1刷発行

平成31年 第2版 第1刷発行

令和2年 第3版 第1刷発行

発行所……都立小石川中等教育学校数学科

〒113-0021

東京都文京区本駒込2丁目29番29号

電話 03-3946-7171

FAX 03-3946-7397

<http://www.koishikawachuto-e.metro.tokyo.jp/>



都立小石川中等教育学校 2年 組 番

氏
名