

理工系のための
数学入門 Introduction to Mathematics
for Science and Engineering
learners

確率・統計

菱田 博俊・著

Probability and Statistics

B

A

C

はじめに

ビッグデータ時代、AI 第三世代などと言われていますが、人間が何をすべきかという、人間の役目（評価基準？）が根底から変わりつつあります。

その中で、データの取捨選択から分析評価までを適切に行う能力は、文部科学省もうたっている通り、とても重要と言えます。他方、高校の数学 A には場合の数と確率の話が、数学 I 及び B には統計学の基礎部分の話が載っていますが、高校ではいまだ幾何学や解析学のように力を入れていない（入れられない）とも聞きます。

そこで、高校生（場合によっては中学生）から社会人までを広く対象とした、「基本的な統計学的手法を即道具として使えるようにする」入門書を執筆することとしました。データや AI の話に触れつつ、確率、集合の話から推定、検定の話までの基本と本質的な考え方を網羅しました。また、読者の皆さんがその手で確認しながら体得していく演習形式になっています。統計学ユーザーにこれからだろうとしている老弱男女の皆さんが、取りあえずこれだけでければ、身近なちょっとしたデータ処理は 9 割方できるようになると期待しています。

大学受験や企業内研修などにも、広く役立てて頂ければ幸いです。

本書は 12 章構成です。順番に学習してください。大学で、連続した 14 回程度の講義で使用する場合などには、原則として毎回 1 章分を学習した後に、試験を実施し、その結果を振り返って頂ければよいでしょう。

各章のページ数は極力合わせましたが、それでもなお、内容の違いにつき、やむを得ずバラつきが発生しました。特に、第 12 章は初学者にはいささか高度な内容なので、状況次第では省略して、代わりにページ数の多い章を 2 回分に分割して頂ければよいと思います。

各章とも、必要に応じて「例」を挙げて原理や概要を解説し、「類題」でそれを確認しながら補強しています。そして、「Excel の問題」で実際に読者の皆様に PC を使って数値を取り扱って頂く構成になっています。例と類題は有機的に配列していますので、飛ばさずに学習してください。

はじめに

また、学習した内容の定着に役立つよう、各章とも、章末にできるだけたくさん「練習問題」を用意しました。の数は難度を示します。簡単な問題から頑張って挑戦してみてください。

統計学 (science of statistics) は、見えている一部のデータから、見えていない真実を予測するための道具だと考えます。

予測するには、単なる希望や憶測ではいけません。そこに論理展開が必要です。集合や確率の考え方は、基盤理論としてそれを支えます。

また、学問と言わずに道具と言ったのは、使いこなせて初めて意味があるという思いがあります。つまり、知っているのではなく、使えることが重要なのです。道具は、良いに越したことはありませんが、たとえ、その道具がそこそこでも、それを使いこなせたときは、おそらく良い道具を使いこなせなかったときより良い効果を得られるものと確信しています。つまり、統計学を学ぶ目的は、統計学を道具として使いこなせるようになることと言えます。

ところで、見えない真実は、とても貴重、……宝物のようです。データの中に宝物が隠れていることが、たくさんあるはずです。統計学は宝探しと言えますね。さあ、この本を読んで、隠れた宝を探し出しましょう。

最後に、本書の編集・組版を頂いた Green Cherry、それと出版の協力を頂いたオーム社に謝意を表します。

令和2年3月

菱田 博俊

工学は、

人間社会の為に有意義な物や仕組みを、創出する
学問です。

確率統計学は、

それを実現する為の助けとして、
見えている現象から見えない真理に少しでも近づく為の
道具です。

目 次

第 1 章 データ取扱いの心得

1.1	何のための確率・統計か.....	1
1.2	データの流れ.....	4
1.3	データの取得.....	5
1.4	データ倫理.....	9
1.5	データの分析と考察.....	11
	章末問題.....	12

第 2 章 確率集合論の基礎

2.1	集合論の考え方.....	17
2.2	順 列.....	20
2.3	組合せ.....	22
2.4	確 率.....	22
	章末問題.....	22

第 3 章 第 3 章标题

3.1	节 3.1 标题.....	23
3.2	节 3.2 标题.....	23
3.3	节 3.3 标题.....	24
	章末問題.....	24

第 4 章 第 4 章标题

4.1	节 4.1 标题.....	25
4.2	节 4.2 标题.....	25
4.3	节 4.3 标题.....	26
	章末問題.....	26

第 5 章 第 5 章标题

5.1	节 5.1 标题	27
5.2	节 5.2 标题	27
5.3	节 5.3 标题	28
	章末問題	28

第 6 章 第 6 章标题

6.1	节 6.1 标题	29
6.2	节 6.2 标题	29
6.3	节 6.3 标题	30
	章末問題	30

第 7 章 第 7 章标题

7.1	节 7.1 标题	31
7.2	节 7.2 标题	31
7.3	节 7.3 标题	32
	章末問題	32

第 8 章 第 8 章标题

8.1	节 8.1 标题	33
8.2	节 8.2 标题	33
8.3	节 8.3 标题	34
	章末問題	34

第 9 章 第 9 章标题

9.1	节 9.1 标题	35
9.2	节 9.2 标题	35
9.3	节 9.3 标题	36
	章末問題	36

第 10 章 第 10 章标题

10.1 节 10.1 标题	37
10.2 节 10.2 标题	37
10.3 节 10.3 标题	38
章末問題	38

第 11 章 第 11 章标题

11.1 节 11.1 标题	39
11.2 节 11.2 标题	39
11.3 节 11.3 标题	40
章末問題	40

第 12 章 第 12 章标题

12.1 节 12.1 标题	41
12.2 节 12.2 标题	41
12.3 节 12.3 标题	42
章末問題	42

章末問題の解答例	143
----------------	-----

参考文献	143
------------	-----

索 引	143
-----------	-----

本書中の「Excel の問題」について

本書中にある「Excel の問題」の内容を PC で実行したことによる直接あるいは間接的な損害に対して、著作者およびオーム社は一切の責任を負いかねます。

この「Excel の問題」で解説している実行方法は、2020 年 3 月時点のものです。将来にわたって保証されるものではありません。Excel は頻繁にバージョンアップがなされています。このため、本書で解説している実行方法で実行できなくなることもありますので、あらかじめご了承ください。本書の発行にあたって、読者の皆様に問題なく実践していただけるよう、できる限りの検証をしておりますが、以下の環境以外では構築・動作を確認しておりませんので、あらかじめご了承ください。

- ・ PC 本体：Windows 10 Pro 64 bit (CPU：Intel Core i5、メモリ：8 GB)
- ・ Excel 環境：Microsoft Excel 2016

また、上記環境を整えたいかなる状況においても動作が保証されるものではありません。ネットワークやメモリの使用状況、および同一 PC 上にある他のソフトウェアの動作によって、本書のプログラムが動作できなくなることがあります。併せてご了承ください。

本書の購入者に対する限定サービスとして、本書に掲載しているソースコードは、以下の手順でオーム社の Web ページからダウンロードできます。

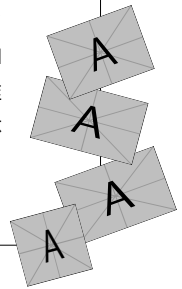
- ① オーム社の Web ページ「<https://www.ohmsha.co.jp/>」を開きます。
- ② 書籍検索で『理工系のための数学入門確率・統計』を検索します。
- ③ 本書のページの「ダウンロード」タブを開き、ダウンロードリンクをクリックします。
- ④ ダウンロードしたファイルを解凍します。

なお、本書に掲載しているソースコードについては、オープンソースソフトウェアの BSD ライセンス下で再利用も再配布も自由です。

第 1 章

データ取扱いの心得

まず本章で、データの取り扱いについて学びましょう。例えば、健康診断では、身長と体重を測り、これらをもとに健康の指標として BMI 指数を計算し、肥満度を考察します。ここで、身長、体重、BMI 指数がデータであり、これらのデータを分析することで肥満度が推論できます。この章で、データを取得してから推論するまでの全体的なイメージを押さえてください。ビッグデータ？時代にしっかり対応できるよう、データ取り扱いの基本を学んでいきましょう。



1.1 何のための確率・統計か

学歴社会の是非はともかく、多くの日本の高校生が大学受験に際して偏差値 (117 ページ) に振り回されています。最後の模擬試験でとった偏差値が 58 の人は、偏差値 65 が必要な希望の国立大学を受けるか、ランクを下げるか迷うこともあるでしょう。偏差値の高い大学を卒業すれば、幸せになれるとは限りませんが、入れる大学に入るという考え方も本末転倒かも知れません。

第一志望の大学に何浪してでも入りたいという信念があれば迷わないのですが、実際にはいろいろな問題がのしかかってきます。妥協したからといって、一概に責められるべきではないでしょう。

ここで、妥協の根拠は、覚悟（納得）するうえで重要です。例えば、家が貧しく給料のよい会社に早く就職したいという事情があれば、合格可能な範囲で

就職有利な大学（得てして偏差値の高い大学）を受験する決断ができます。もちろん必ず入試に受かるとは限りませんので、家計状況や社会経済状況などを勘案し、どの大学を受けるか、最悪一浪は我慢するか、などと検討をた重ねることになります。

このように、人生ではとにかく情報に基づいて判断することを求められます。さらに、AI が台頭した昨今、人間が頭に蓄えている知識自体の価値は下がり、専門家ですら AI に適わないと言われ始めています。このような時代に適応して、人間でないとできないこと、すなわち、課題を設定し、それに対して情報を入手し覚悟をもって判断することこそ、できるようになりたいものです（図 1.1）。

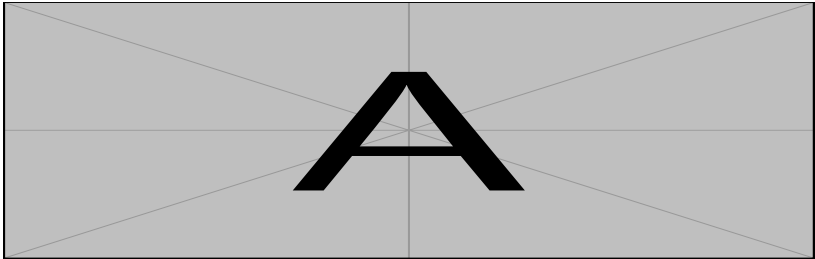


図 1.1 情報を用いた問題解決への道（模式図）

類題 1.1

過酷な環境で使う機械を、1 種類の材料だけで作るとします。調査の結果、最終候補材は材料 A と B になりました。これらについて、主要な四つの強さの指標を測定し、表 1.1 の結果を得ました。そして、互いに比較できるよう偏差値に換算しました。

表 1.1 強さ試験の結果

	材料 A	材料 B
韧性	66	62
硬度	71	75
耐食性	69	64
破断強度	68	73
平均	68.5	68.5

- (1) 最も必要な強さが、確率 24% で軸性、29% で硬度、18% で耐食性、22% で破断強度、7% でその他であるらしいとの情報を入手しました。どちらの材料を選びますか。
- (2) 全くそういった情報が入手できない場合、どちらの材料を選びますか。

答え

- (1) まずは簡単に考えてみましょう。捉性と耐食性は材料 A がよいので、足して確率 $24 + 18 = 42\%$ で材料 A が有利です。他方、硬度と破断強度は材料 B がよいので、足して $29 + 22 = 51\%$ で材料 B が有利です。 $51\% > 42\%$ ですから、材料 B を選ぶほうが確率論的には正しいと言えます。
- (2) 情報は、表 1.1 に限られます。平均がわかりやすい指標になりますが、あいにく同じですね 1 上。しからば、例えば「強さの最大値」と「安定した強さ」のどちらかを優先させることにしましょう。

前者ならば強さ 75 を硬度でマークした材料 B を、後者ならば強さ間のバラつきが小さい材料 A を選ぶべきでしょう。了臨機応変に判断しながら、納得することが重要です。

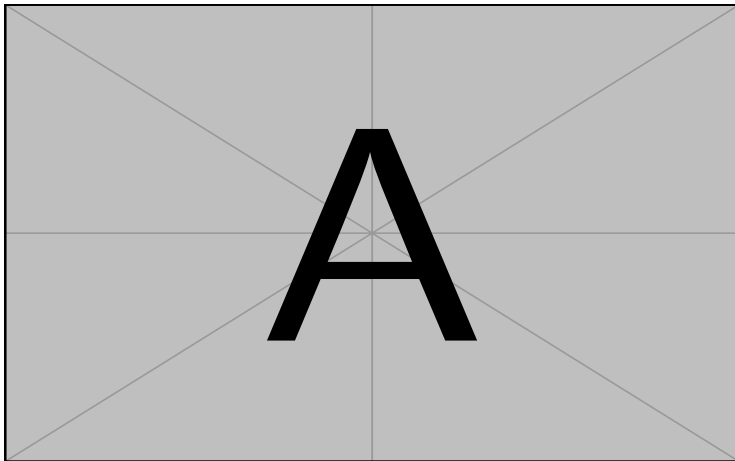


図 1.2 強さ試験の結果

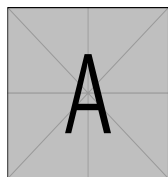


図 1.2 に、強さ試験の結果を示します。数値をグラフにすると見た目でよくわかりますよ。

1.2 データの流れ

上記の通り決断するためには納得感、言い換えれば理屈を必要とします。確率統計は、不確かなことを決断するための理屈を提供してくれる「道具」です。

確率統計により、情報を数学的に処理して推論を導き出します。推論するための定量的な情報を、特に**データ**と呼ぶことにしましょう。

この節では、データを取得してから何らかの結論を出すまでの過程を大まかに把握しましょう。図 1.3 (次ページ) に、データの流れを示します。

- ① データ取得: 必要なデータとその高品質な取得方法を熟考します。
- ② データ管理: 保管や処分のやり方やルールなどを議論し、いつでも安全に使えるようにします。
- ③ データ分析: 取得したデータの特徴を求め、そこから隠れた真実を推定します。
- ④ 結論導出: 仮説、条件、前提等を踏まえて、推定内容を正確に理解したうえで、覚悟を決めて判断をします。

それぞれの過程で実施する詳細な内容については、後続の節で述べます。

細かい作業を実際に行っていると、その意義や目的を見失ってしまうことも出てきます。具体的なデータ処理を行っているときにも、自分がいま行っている処理が、全体の流れのどの部分で、どんな役割を持っているのかを常に意識するようにしましょう。そうすれば、データの取り扱い能力も向上します。

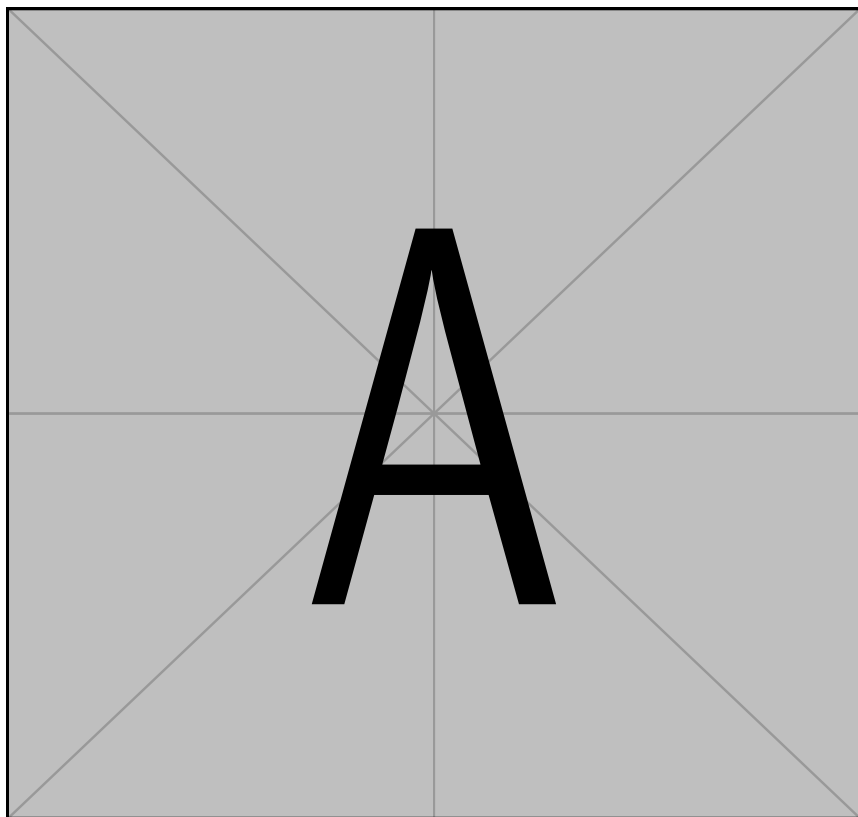


図 1.3 データの流れ

1.3 データの取得

取得したデータから、まずは現状を把握します。続いて、未来を推定 (予測、予見) します。これらの作業には、何より高品質なデータが不可欠です。たとえば、深層学習ができるニューラルネットワークのような優れた道具を用いたとしても、高品質なデータがなければ高精度の推定は不可能です。

ビッグデータ時代の今日、高品質なデータを持つ意義は大きいです。無駄なく効率よく高品質のデータを得ることに、多くの人が知恵を傾けています。

(1) 目的データ

時間と費用を掛けて調査をする際には、調査目的を検討、理解し、最終的にどのようなデータを取得したいのかがブレないようにすることが肝心です。最終的に取得したいデータを**目的データ**と言います。目的データとは結論を導くために必要なデータです。

例えば、あなたのグループで、メンバーの健康度に基づき、今後の活動を検討することになります。この場合、健康度が目的データとなります。ここで、データを提供するメンバーは**被験者**と呼ばれます。

なお、データを物や事象から得ることもあります。

(2) 主観調査と客観調査

健康度には心理的な側面と生理的な側面があり、それらが一致しないこともあります。そこで、両方を調査する必要があります。

心理的健康度は、被験者に尋ねることでわかります。具体的には、図 1.4 のような目盛りを用意し、チェックしてもらうような調査になります。このような被験者に直接聞く調査を**主観調査**と言います。主観調査で得たデータには、真偽を確かめる有効な手段が乏しく、信じるしかないという特徴があります。

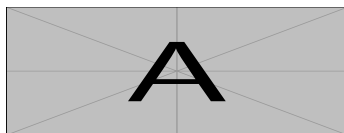


図 1.4 データの流れ

生理的健康度は、血圧や肥満度などの測定値や、BMI 指数などの指標値で表されます。BMI 指数は、身長と体重から簡便に計算できるので、精度は落ちますが多用されています。これらの測定値や指標値は、誰が調査してもほぼ同じになるという特徴があり、このような調査を**客観調査**と言います。

(3) 直接データと間接データ

次に、データを別の視点からみてみましょう。

身長と体重などの測定データは、それ自体を直接取得できるので**直接データ**と言います。直接データは、調査しさえすれば確実に取得できます。

対して、BMI 指数のように別の情報から計算されるデータもあります。このデータは直接取得できないので、**間接データ**と言います。間接データの精度は、元データの精度を上回ることあり得ません。したがって、統計調査においては、できるだけ直接データを取得すべきです。

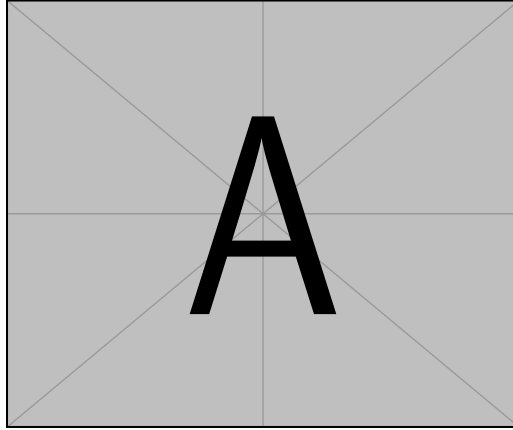


図 1.5 データの分類

(4) 予定データと予定外データ

また、別の分類をします。予定して得たデータを**予定データ**、そうでないデータを**予定外データ**と言います。

先の例では、BMI 指数を求めるために身長と体重を調査したのだから、これらは全て予定データです。統計調査を成功させるには、しっかり計画し、予定データを過不足なく取得することが肝心です。被験者への負担や分析の手間を考えて、取得すべきデータを厳選しましょう。

しかし、最初から目的データが明確とは限りません。この場合には、次の予定外データをとります。

まず、目的データに近いと思われるデータを目指して、いろいろとデータの収集や処理を試みます。このようにして収集や処理されたデータを、かゆい部位に手が届かずかけないみたいなので、**隔靴データ**と呼ぶことにしましょう。かくかまた、データ処理をしているうちに、有益なデータの存在に気づくこともあります。これを遭遇データと呼びましょう。

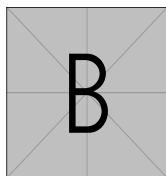
本当の発見は予定外データにあることが多いです。取得したデータを、一生懸命分析して、隠れて見えなかった予定外データを見つけ出すことが肝心です。

類題 1.2

I 君は、高精度と定評のある深層学習ソフトウェアを入手しました。さっそく、Web 上から簡単に入手できるいくつかのデータを使って学習させました。高品質な推定結果を得られるでしょうか。

答え

Web 上から簡単に入手したデータが高品質とは思えません。ソフトウェアは良いに越したことはありませんが、データから導かれる推定精度を上げるためには、データ自体の品質こそが重要です。



よい教育を受けることが大事なのは、人も機械も同じですね。

類題 1.3

近年、イヤフォンが原因の難聴が、若者を中心に世界的に増加しています。そこで、周囲のイヤフォン使用者の耳年齢について調べました。本来、年齢相応に聞こえる周波数がありますが、耳年齢とはどの高さの周波数まで音が聞こえるかを測定し、相応の年齢を逆算したものです。調査に当たっては、イヤフォンをどの程度使っているかについても併せて尋ねることにします。

- (1) 耳年齢は直接データと間接データのいずれでしょうか。
- (2) 耳年齢は目的データでしょうか。
- (3) 「イヤフォンをどの程度使っていましたか」という尋ね方は適切でしょうか。

答え

- (1) 間接データ。聞こえた音の周波数が直接データです。
- (2) 目的データです。イヤフォンをどの程度使っているかも目的データです。
- (3) 不適切です。「毎晩寝る前に少し」「う～ん、あまり」「塾の行き帰りにたまたま」「仕事の必需品だ」などのさまざまな回答が出てきて、調査後の処理に困ります。

1.4 データ倫理

(1) データの管理

ータを取得したら、それを使える状態に保持しつつ、特に個人情報が漏洩しないように管理しなければなりません。さらに、万一漏洩しても内容がわからないように暗号化する、個人名を切り離すなどの工夫を施します。

一方、取得したデータが将来また必要になることもあります。そのときに備えて、データをいつ、どこで、どうやって、なぜ取得したのか、どのようなデータなのかをできるだけ詳細に記録しておくことも大切です。

(2) 被験者への配慮

個人情報を取得する際には、被験者に説明書を渡して説明し、同意を得て同意書に署名をもらう必要があります。

情報提供の際には、アンケートへの記入や採血などのさまざまな負担が、被験者に掛かります。狭義には被験者を実際に傷つけるデータ取得方式を、広義には被験者に許容以上の苦痛を与えるデータ取得方式を、**侵襲式**と言います。対義語は非侵襲式です。基本はもちろん、「極力、非侵襲式で」データを収集するのが望ましいと言えます。

(3) データに対する心構え

身のまわりには、膨大な情報がれかえっています。情報化社会においては、有益あるいは有害な情報を選別し、利用や棄却する能力が問われます。情報に埋もれず、悪意のある第三者に個人情報を盗られないように身を守らねばなり

ません。

取得したデータと、そこから導き出された結論には、データを取り扱った人の価値観や人間性が表れます。誰もがデータを扱う時代ですが、データの取り扱い方は人によって千差万別なのです。人から安心してデータを任してもらえる人になることが大切です。

類題 1.4

18 人に対して、身長、体重、自覚健康度を無記名アンケートします。アンケート用紙の管理上の注意事項を挙げましょう。

答え

少人数なので、筆跡などで個人を特定されない配慮が必須です。そのためにも、適切に保管します。アンケート用紙は施錠できる閉空間に入れ、伴は責任者が持ちます。また、電子化したデータがあれば、それをパスワードでアクセス制限し、念のため暗号化します。

一方、適切な人がデータを正しく、スムーズに使えるような配慮も必要です。そのために、いつ、どこでなどの調査内容に関する記録も併せて保管します。

類題 1.5

次の各調査が侵襲式かどうかを考えましょう。

- (1) 脳波計を使って、睡眠の質を調べました。
- (2) 友人のせきがひどいので、本人に断って血中酸素濃度を測定しました。
- (3) ストレス度の測定をするために、指先から少しだけ採血しました。
- (4) 人体にかかわる調査をするために、自分の骨格を CT スキャンしました。
- (5) イヤフォンをこれまで何時間使ってきたかをアンケート調査しました。
- (6) 病歴についてアンケート調査しました。

答え

どちらか迷うときは、侵襲式として考えて注意するのが無難です。

- (1) 侵襲式（心の作用を測定されてしまいます）。
- (2) 非侵襲（指先にクリップ形の測定子を挟むだけで、危険はありません）。
- (3) 侵襲式（採血は明らかに侵襲式です）。
- (4) 侵襲式（放射線を浴びてしまいます）。
- (5) 非侵襲式（被験者にかかる負担はアンケートに答える手間ぐらいでしょう）。
- (6) 回答が非強制であれば非侵襲式（治療に必要であれば、強制でも仕方ありません……）。

1.5 データの分析と考察

(1) データの数学的处理

いくら時間をかけて集めた高品質なデータを基にしても、その後の処理に信頼性がなければ説得力はありません。つまり、誰もが立場によらず信頼することができる、客観性が担保された結果を得るために、数学を使って処理をするのです。統計学は、この数学的处理の方法論を掘り下げます。そしてそれを、集合論や確率論などの論理学に属する理論が援護射撃しています。

統計学を用いたデータ処理は、さまざまです。例えば、取得したデータ群の特徴を明らかにしたり（第3、6、7、8、12章）、複数のデータ群の特徴を比較したり（第4章）、それらの関係性を議論したり（第4、5章）、さらには複数のデータ群から新しいデータを作り出したり（前述のBMI指数）します。

実際には、調べたい対象全数（**母集団**と言います）からデータを取得できないことが大半でしょう。したがって、取得した一部のデータ（**標本集団**と言います）から対象全体を推定（第9～2章）する方法も必要です。

(2) 結論の導出

そして、データからわかることを総合して、場合によってはデータからは読み取れない情勢や仮説も加えながら、目的としている課題に対してどんなこと

が言えるかを論理的に考察し、結論を導きます。すなわち、データを取り扱うまでは「見えなかった」知見を、見つけ出すのです。

データの取得から処理までが同じでも、人によって結論が異なることも少なくありません。それもまた、人間の成す人間のための行為ならではでしょう。データ処理をする人は、可能性の探索（想像）、今回達成できなかったこと（今後の課題）なども、結論とともに列挙し、それをメッセージとして残すことが大切です。

章 末 問 題

◆1.1 「予」が付く語句について、次の各問に答えなさい。

予測・予見（＝推定）：推論により未来を論じる行為。

予想・予期：ある基準に基づき期待して未来を論じる行為。

予報：科学モデルに則り合理的に未来を論じる行為。

予言：ある物事を実現前に言明する行為。主として神秘的現象を対象とし、呪術や宗教に用いられる。

予知・予感：前もって認識する行為。

予断：不完全な証拠に基づく未来への見解（判断）。

予定：意志を伴う未来の行為。

- (1) 上記の語句のうち、内容に対する客観的根拠がないのはどれですか。
- (2) 上記の語句のうち、確率論的根拠に基づいてなされるのはどれですか。
- (3) 以下の五つの行為を、左から理論的といえる順番に並べなさい。

予断 予報 予測 予言 予想

◆1.2 次のいくつかの語句どうしの関係について簡単に説明しなさい。

(1) 情報 データ

(2) 予測 予報 予言 予定

(3) 主観調査 客観調査

(4) 直接データ 間接データ 予定データ 隔靴データ
遭遇データ

◆1.3 被験者からデータを得る際について、次の各問に答えなさい。

- (1) 同意書には必ず署名してもらうべきか述べなさい。
- (2) 自由回答形式の質問を 30 個用意しました。質問数が多いかどうか述べなさい。
- (3) 謝礼は用意すべきか述べなさい。
- (4) スマホやインターネットで回答を得ることに問題はるか述べなさい。
- (5) 被験者へのデータ分析結果の報告義務はあるか述べなさい。

◆1.4 統計学とは本質的にどのような学問と本書では解説しているか、簡潔にまとめなさい。

◆1.5 データと真実の関係について、次の各問に答えなさい。

- (1) 全数調査は常にできるか述べなさい。
- (2) 身のまわりのデータや情報を列挙しなさい。
- (3) 上記で挙げたデータや情報を、日頃それと認識して暮らしているか述べなさい。
- (4) 情報は常に真実を表していると言えるか述べなさい。
- (5) 情報は常に正確に得られると言えるか述べなさい。

◆1.6 データ（情報）取得の際の注意事項について、次の各問に答えなさい。

- (1) 被験者に負担が掛かるデータ（情報）の例を挙げ、なぜ負担が掛かるのかを簡単に説明しなさい。
- (2) データ（情報）取得の際に重要なことを、簡単に述べなさい。

◆1.7 類題 1.4 (10 ページ) の続きで、「累積使用時間」を目的データとして、イヤフォンをどの程度使っていたかを尋ねることにしました。

- (1) 累積使用時間を目的データとしてよいか述べなさい。
- (2) 累積使用時間だけで耳がどの程度傷んでいるかを論じられるか述べなさい。
- (3) 累積使用時間は、直接データとして取得できるか述べなさい。
- (4) 累積使用時間のデータを得るにはどのような尋ね方をすればよいか答えなさい。

◆1.8 統計調査の報告書作成に関して、次の文章の空欄を埋めなさい。

〔①〕データを得た後には、結果を〔②〕やグラフ等を用いてわかりやすく整理します。〔①〕データが得られなかった場合には、できるだけそれに〔③〕データが得られるように統計学的な工夫を行います。

データから〔④〕を考察する際には、論理的に導かれる〔⑤〕、〔⑥〕に基づいた可能性、単なる希望的な憶測などを明確に区別して、もともとの統計調査目的から外れないように、言いたいこと、言えることをまとめます。また、「データが足りなかった」等の〔⑦〕も記述しましょう。


◆1.9 統計学でデータをどのように取り扱っていくかについて、次の各問に簡潔に答えなさい。

- (1) 収集したデータの管理の際に厳守すべきことを挙げなさい。
- (2) 収集したデータの分析に関する基本理念を述べなさい。


◆1.10 ビッグデータを説明した次の文章を読み、各問に答えなさい。

ビッグデータの定義はさまざまですが、「従来のデータベース管理システムでは記録、保管、解析が難しいような巨大なデータ群 17)」と一般的に解釈されています。ビッグデータは概して〔①〕がさまざまで、時々刻々と〔②〕が増え続けます。従来はデータとして扱い切れなかったビッグデータを③記録、保管し、④素早く解析することで、斬新な事実や仕組みを見出すことができると期待されています。

- (1) 文中の①と②に適切な単語を入れなさい。
- (2) Web 上のビッグデータに該当するデータを挙げなさい。
- (3) 一般の生活環境中のビッグデータに該当するデータを挙げなさい。
- (4) 下線③に「記録、保管」とあるが、Web 上のビッグデータはどこに記録、保管されているか、述べなさい。
- (5) 下線④に「素早く解析する」とありますが、そのために必要なことを挙げなさい。
- (6) ビッグデータにあてはまるのは、次のうちどれか、述べなさい。
構造化データ 非構造化データ 定型データ 非定型データ
自己免疫性 環境対応性 時系列性 リアルタイム性
- (7) ビッグデータを運用するうえで注意すべきことを考え、述べなさい。

 **1.11** それぞれの場合に、データを新たに取得する具体的手法を述べなさい。

- (1) 個人に直接尋ねる場合
- (2) 結果的に個人から取得する場合
- (3) 現象や社会を観察する場合

 **1.12** 次の文献を、信頼性がある順番に並べかえなさい。

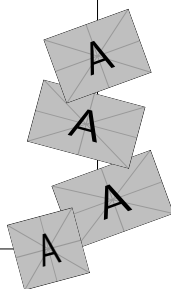
査読論文（真偽の審査を受けた論文） 学会報告 新聞記事や TV 情報
特許 広告 行政広報 週刊誌 Web 上の情報（発信者無記名）

第 2 章

確率集合論の基礎

この章では、高校で学ぶべき集合と確率の基礎を復習しながら、集合の考え方が確率の考え方につながっていることを学びます。順列や組合せについて具体的な例で考える際に、集合の考え方を利用すると全体を明るく見通せます。

それぞれの問題を解くときに、順列を使うのか、組合せを使うのかを判定できるようになりましょう。



2.1 集合論の考え方

この節では、基本的な用語や式を**ヴェン図**^{*1}を介してまとめます。式を駆使して問題を解くのもよいですが、ヴェン図から集合同士の関係性を考えることこそが**集合の本質**です。式とヴェン図が頭の中で連動できることが理想的です。

例 2.1

1 ～ 20 の自然数を分類してみましょう。ここで、分類すべき対象である 1 ～ 20 を、**要素＝元**と言います。要素数が少ない場合には、イメージを得るためにも、全ての要素をヴェン図で分類するのがお勧めです（図 2.1）。

ヴェン図では、全ての要素を外枠で囲い**全体集合** U とします。ここでは、 U の要素は 1 ～ 20 で、 U の要素数 $n(U)$ は 20 です。数学的記述は以下の通りで

*1 考案したイギリスの数学者 John Venn の名にちなんでいます。

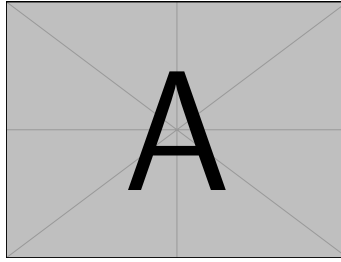


図 2.1 例 2.1 のヴェン図

す。

$$U = \{1, 2, \dots, 19, 20\}, \quad n(U) = 20.$$

部分集合として、2 の倍数の集合 A 、3 の倍数の集合 B 、4 の倍数の集合 C 、23 の倍数の集合 D を考えます。これらを、同様に次のように記述します。部分集合は、全体集合に含まれます。

$$A = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}, \quad n(A) = 10$$

$$B = \{3, 6, 9, 12, 15, 18\}, \quad n(B) = 6$$

$$C = \{4, 8, 12, 16, 20\}, \quad n(C) = 5$$

$$D = \{ \} = \emptyset, \quad n(D) = 0$$

ここで、 \emptyset は要素がないことを表し、要素のない D を**空集合**と言います。

U において、例えば A の要素を全て除いた集合を、 A の**補集合** \bar{A} と言います。以下が成立します。

$$A + \bar{A} = U, \quad n(A) + n(\bar{A}) = n(U) \quad (2.1)$$

全ての要素を、必ずいずれかの集合に属するように分類することを、「ミシィ*²に分類する」と言います。補集合はこの考え方の一種で、例えばケース A 、ケース B 、ケース C と考えた後、「その他のケース」を考えることに対応

* 2 ミシィは、mutually exclusive and collectively exhaustive の頭文字 MECE の日本語読みです。「相互に重複せず、全体として漏れがない」と言う意味です。

します。数学の答案論述をする際にも、研究をするうえでも、考え落としがないように常に補集合を意識しましょう。

A と B の両方に属する要素、すなわち 6 の倍数の集合を、**積集合（共通集合）**と言ひ、 $A \cap B$ と表記します。また、 A と B のどちらかに属する要素の集合を、**和集合（融合集合）**と言ひ、 $A \cup B$ と表記します。 $A \cap B$ と $A \cup B$ について、以下の式 (2.2) が成立します。

$$\begin{aligned} n(A \cup B) &= n(A) + n(B) - n(A \cap B) \\ (A \cap B) &\subset A, \quad (A \cap B) \subset B, \quad (A \cup B) \supset A, \quad (A \cup B) \supset B \end{aligned} \quad (2.2)$$

この部分集合 C は A に完全に含まれるので、以下の関係が成立します。

$$A \supset C, \quad A \cap C = C, \quad A \cup C = A \quad (2.3)$$

さらに、この場合には派生的に以下の関係も成立します。

$$(A \cup B) \supset (C \cup B), \quad (A \cap B) \supset (C \cap B) \quad (2.4)$$

類題 2.1

大小のサイコロを振って、出た目で 2 桁の数を作ります。具体的には、大きいサイコロの目を十の位、小さいサイコロの目を一の位にします。集合 A を 3 の倍数、集合 B を 30 以上 50 未満としたとき、次の各問に答えましょう。

- (1) $n(A)$ と $n(B)$ を求めましょう。
- (2) A と B の積集合の要素を求めましょう。
- (3) A と B の和集合の補集合を求めましょう。
- (4) 空集合を一つ考えてみましょう。

答え

ヴェン図は図 2.2 です。描ける限り描く習慣をつけましょう。

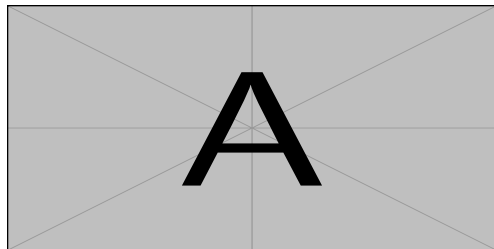


図 2.2 類題 2.1 のヴェン図

$$U = \{11, \dots, 16, 21, \dots, 26, 31, \dots, 36, 41, \dots, \\ 46, 51, \dots, 56, 61, \dots, 66\}$$

です。

- (1) $A = \{12, 15, 21, 24, 33, 36, 42, 45, 51, 54, 63, 66\}, n(A) = 12$
 $B = \{31, 32, 33, 34, 35, 36, 41, 42, 43, 44, 45, 46\}, n(B) = 12$
- (2) $A \cap B = \{33, 36, 42, 45\}, n(A \cap B) = 4$
- (3) $\overline{A \cap B} = \{11, 13, 14, 16, 22, 23, 25, 26, 52, 53, 54, 56, 61, 62, 64, 65\}, n(\overline{A \cap B}) = 16$
- (4) 「70 以上」「19 の倍数」「一の位が 8 の数」等。

2.2 順 列

順列とは、ある集合の中からいくつかの要素を取り出して一列に並べることです。順列の公式は高校でも学びますが、樹形図を作成して具体的なイメージを持つことが肝心です。式と樹形図を連動させて考えられるようになりましょう。

(1) 直順列

例 2.2

「A」「C」「E」「R」と書かれたカードが 1 枚ずつあります。これを並べて、英配列を作ります。

この程度の数であれば、全ての配列を一覧にする**樹形図**が簡明です。悩む前に作ってみましょう。描けば、規則性に気づくこともあります。そのまま英語として読める配列が多いですね。

$$4 \times 3 \times 2 \times 1 \left(= \frac{4!}{0!} = {}_4P_4 \right) = 24 \text{ [通り]}$$

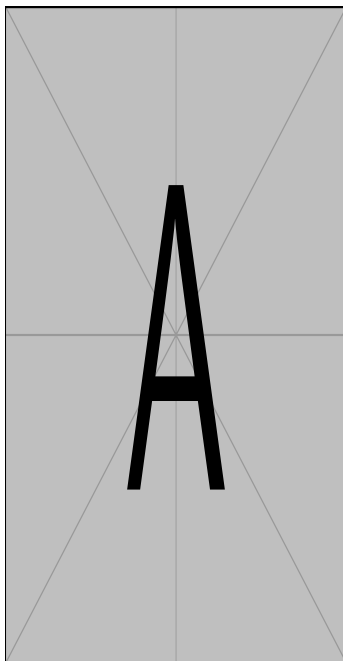


図 2.3 例 2.2 の樹形図

例 2.3

もし、この 4 枚中 3 枚だけを使うのであれば、樹形図の左の 3 列だけを考えればよいことになります。

$$4 \times 3 \times 2 \left(= \frac{4!}{1!} = {}_4P_3 \right) = 24 \text{ [通り]}$$

3 枚使うということは、1 枚残すということです。この 1 枚を 4 枚目として並べても、残しても、同じことです。

一般的に、全て異なる n から k を配列させるとき、式 (2.5) で計算される配列が作れます。これを ${}_nP_k$ と書きます*³。

$${}_nP_k = \frac{n!}{(n-k)!} \quad (2.5)$$

類題 2.2

カード「A」「C」「E」「R」「T」から3枚を並べます。何通りできますか。

答え

$${}_5P_3 = \frac{5!}{(5-3)!} = 5 \times 4 \times 3 = 60 \text{ [通り]}$$

2.3 組合せ

2.4 確 率

章 末 問 題

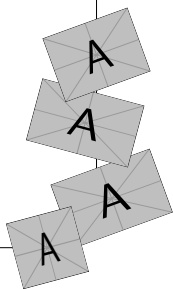
2.1

* 3 順列 (permutation) の頭文字をとっています。

第 3 章

第 3 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



3.1 节 3.1 标题

山路を登りながら、こう考えた。

3.2 节 3.2 标题

山路を登りながら、こう考えた。

3.3 节 3.3 标题

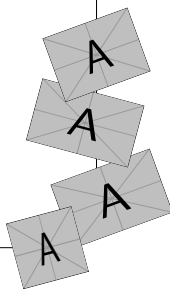
山路を登りながら、こう考えた。

章 末 問 題

第 4 章

第 4 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



4.1 节 4.1 标题

山路を登りながら、こう考えた。

4.2 节 4.2 标题

山路を登りながら、こう考えた。

4.3 节 4.3 标题

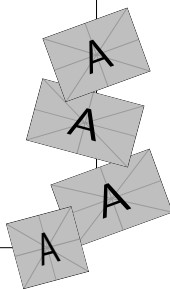
山路を登りながら、こう考えた。

章 末 問 題

第 5 章

第 5 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



5.1 节 5.1 标题

山路を登りながら、こう考えた。

5.2 节 5.2 标题

山路を登りながら、こう考えた。

5.3 节 5.3 标题

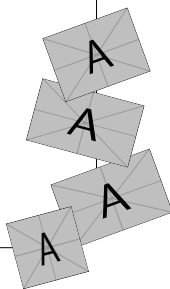
山路を登りながら、こう考えた。

章 末 問 題

第 6 章

第 6 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



6.1 节 6.1 标题

山路を登りながら、こう考えた。

6.2 节 6.2 标题

山路を登りながら、こう考えた。

6.3 节 6.3 标题

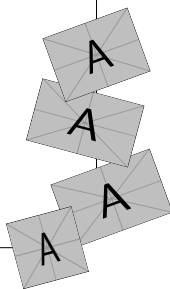
山路を登りながら、こう考えた。

章 末 問 題

第 7 章

第 7 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



7.1 节 7.1 标题

山路を登りながら、こう考えた。

7.2 节 7.2 标题

山路を登りながら、こう考えた。

7.3 节 7.3 标题

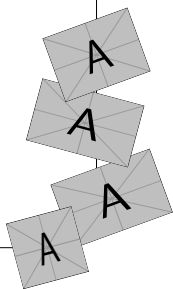
山路を登りながら、こう考えた。

章 末 問 題

第 8 章

第 8 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



8.1 节 8.1 标题

山路を登りながら、こう考えた。

8.2 节 8.2 标题

山路を登りながら、こう考えた。

8.3 节 8.3 标题

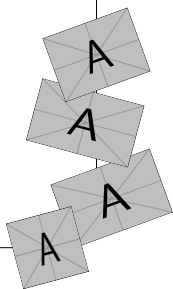
山路を登りながら、こう考えた。

章 末 問 題

第 9 章

第 9 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



9.1 节 9.1 标题

山路を登りながら、こう考えた。

9.2 节 9.2 标题

山路を登りながら、こう考えた。

9.3 节 9.3 标题

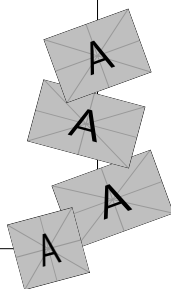
山路を登りながら、こう考えた。

章 末 問 題

第 10 章

第 10 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきっと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閤僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



10.1 节 10.1 标题

山路を登りながら、こう考えた。

10.2 节 10.2 标题

山路を登りながら、こう考えた。

10.3 节 10.3 标题

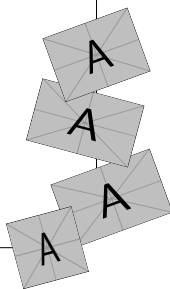
山路を登りながら、こう考えた。

章 末 問 題

第 11 章

第 11 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきつと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閤僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



11.1 节 11.1 标题

山路を登りながら、こう考えた。

11.2 节 11.2 标题

山路を登りながら、こう考えた。

11.3 节 11.3 标题

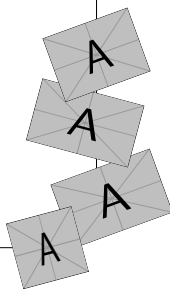
山路を登りながら、こう考えた。

章 末 問 題

第 12 章

第 12 章标题

世に住むこと二十年にして、住むに甲斐ある世と知った。二十五年にして明暗は表裏のごとく、日のあたる所にはきっと影がさすと悟った。三十の今日はどう思うている。—— 喜びの深きとき憂いよいよ深く、楽みの大いなるほど苦しみも大きい。これを切り放そうとすると身が持てぬ。片づけようとすれば世が立たぬ。金は大事だ、大事なものが殖えれば寝る間も心配だろう。恋はうれしい、嬉しい恋が積もれば、恋をせぬ昔がかえって恋しかろ。閨僚の肩は数百万人の足を支えている。背中には重い天下がおぶさっている。うまい物も食わねば惜しい。少し食えば飽き足らぬ。存分食えばあとが不愉快だ。……



12.1 节 12.1 标题

山路を登りながら、こう考えた。

12.2 节 12.2 标题

山路を登りながら、こう考えた。

12.3 节 12.3 标题

山路を登りながら、こう考えた。

章 末 問 題

