

# 前書き

このテキストは、2011年から開講されてきた、「初級統計学講座」のテキストに加筆・修正を加えまとめたものである。

株式会社すうがくぶんかは、社会人向けの数学塾として2010年に創業した。「初級統計学講座」は創業当時からの講座の一つであり、すうがくぶんかを支える人気講座でもある。2018年、通算3回目の大改定となる今回は以下の項目を改定した。

- 1) 初級統計学講座の運営で培われた、社会人に統計を教えた経験を講座に反映させた。
- 2) 一部数学的なコラムを追加して内容を増補した。
- 3) R言語の script を本文中に挿入した。

本書で学ぶカリキュラムは、2018年前期までの初級統計学からかなりのボリュームアップとなったが、この量は必要と判断し大幅に内容を増やした。これは2011年より始まった統計検定試験(この講座は2級に対応)に対応する必要からというところが大きい。最近の統計検定試験は以前よりも本質的な統計学および数学の理解が問われている。

今回のカリキュラム改訂には、これまでに初級統計学、またその他の弊社講座を受講してくださった、たくさんの受講生から得た経験が各所に生かされている。多くの社会人の皆さんとの交流の中で生まれた本書は、従来の教科書とはひと味違ったオリジナルなものに仕上がったと思う。重ねて、受講生の方々にはお礼を申し上げさせていただきたい。

2018年秋 株式会社すうがくぶんか 代表取締役 瀬下 大輔



# 目次

第 1 章	序論	1
1.1	統計学の現状	1
1.2	統計的考え方	2
1.3	統計学の分類とその対象	5
1.3.1	記述統計学と推測統計学	5
1.3.2	母集団と標本	6
1.4	統計学の歴史-3つの統計思想-	7
1.4.1	国勢学	7
1.4.2	近代統計学の起こり	8
1.4.3	確率論と統計学	9
1.5	いくつかの注目すべき事例	11
第 2 章	データの分類と整理	13
2.1	量的変数と質的変数	13
2.2	尺度	14
2.3	離散変数と連続変数	14
2.4	時間による分類	15
2.5	データの次元	16
2.6	tidy データ	16
第 3 章	記述統計の方法：可視化	19
3.1	度数分布表とヒストグラム	20
3.1.1	度数分布表	20
3.1.2	ヒストグラム	21
第 4 章	記述統計の方法：数値要約	23
4.1	平均値・中央値・最頻値	23
4.2	四分位点と箱ひげ図	25
4.3	分散・標準偏差	27
4.4	データの標準化と偏差値	30
4.4.1	標準化	30
4.4.2	偏差値	31

第 5 章	2 変量の関係	33
5.1	相関	33
5.2	相関の図示	34
5.2.1	散布図	34
5.2.2	分割表	34
5.3	相関係数	35
5.3.1	単相関係数	35
5.4	質的変数を含む場合の指標	38
5.4.1	相関比	38
5.4.2	クラメールの連関係数	39
第 6 章	確率分布	41
6.1	確率分布とは	41
6.2	確率分布はなぜ重要か	44
6.3	ベルヌーイ試行と二項分布	45
6.4	ポアソン分布*	46
6.5	正規分布	48
6.5.1	正規分布の性質と確率	48
6.5.2	標準正規分布と標準正規分布表	50
6.6	標本分布	51
第 7 章	推定	55
7.1	推定とは	55
7.1.1	推定の基礎知識	55
7.1.2	点推定と区間推定	55
7.1.3	【点推定】 不偏性と一致性*	56
7.1.4	パラメトリックとノンパラメトリック*	58
7.1.5	モーメント法と最尤法*	58
7.2	中心極限定理	59
7.2.1	中心極限定理	59
7.3	母平均に関する推定	61
7.3.1	【区間推定】 母平均	61
7.3.2	【区間推定】 2つの母平均の差	65
7.3.3	【区間推定】 対応ある標本の場合	68
7.4	その他の推定	70
7.4.1	【区間推定】 母分散 母標準偏差	70
7.4.2	【区間推定】 母集団比率	71
第 8 章	検定	73
8.1	検定とは	73
8.1.1	検定の考え方	73

8.1.2	仮説検定 . . . . .	73
8.1.3	帰無仮説と対立仮説 . . . . .	75
8.1.4	真実と判断 . . . . .	75
8.1.5	検定の種類と目的 . . . . .	76
8.1.6	検定の手順 . . . . .	76
8.2	母平均の比較値との差の検定 . . . . .	77
8.3	母平均の差の検定 . . . . .	78
8.4	等分散検定 . . . . .	81
8.5	独立性の検定： $\chi^2$ 検定 . . . . .	81
8.6	分散分析（1 要因） . . . . .	83
8.7	分散分析の概要（2 要因） . . . . .	84
8.7.1	2 要因の分散分析 計算 . . . . .	86
第 9 章	回帰分析 . . . . .	89
9.1	回帰分析とは . . . . .	89
9.2	回帰分析の手順 . . . . .	90
9.2.1	回帰分析にまつわるその他のこと . . . . .	102
第 10 章	重回帰分析 . . . . .	105
10.1	重回帰分析の手順 . . . . .	105
10.2	回帰分析の注意点 . . . . .	117
10.2.1	より良い回帰式を得るには . . . . .	117
10.2.2	母集団について . . . . .	118
10.2.3	標準化残差 . . . . .	119
10.2.4	多重共線性 . . . . .	119
10.2.5	重回帰分析で変数の影響度を調べる。 . . . .	120
付録 A	統計学に必要な数学 . . . . .	123
.1	関数 . . . . .	123
.2	積分記号とその意味 . . . . .	124
.3	合成関数 . . . . .	125
.4	微分 . . . . .	126
.4.1	微分係数と導関数 . . . . .	126
.4.2	微分の応用：最大最小問題 . . . . .	129
.4.3	合成関数の微分法 . . . . .	130
.5	行列 . . . . .	131
.5.1	数学の時間：行列 . . . . .	131
.5.2	単位行列 . . . . .	134
付録 B	R-Tutorial . . . . .	139
.6	R とは . . . . .	139
.7	R を動かしてみよう . . . . .	139

---

.8	オブジェクトを作ろう . . . . .	140
.9	ファイルを読み込んでみよう . . . . .	142
.10	基本的な統計量を求めてみよう . . . . .	144
.11	分布を調べるために、グラフを書こう . . . . .	145
.12	相関を見てみよう . . . . .	146
.13	回帰分析を試してみる . . . . .	147
付録 C	時系列データ	149
.14	時系列データの取り扱い . . . . .	149
.14.1	変化率・指数化 . . . . .	149
.14.2	幾何平均 . . . . .	150
.14.3	移動平均 . . . . .	151
付録 D	サンプリング論	153
.15	標本抽出 . . . . .	153
.16	調査方法 . . . . .	155
.17	標本の適切な大きさ . . . . .	156
.18	量的調査と質的調査 . . . . .	157
.19	探索型調査と検証型調査 . . . . .	158
付録 E	アンケート作成の方法	159
.20	アンケートの構成 . . . . .	159
.21	回答の方法 . . . . .	160
.22	よくない質問 . . . . .	161
付録 F	復習問題	165
索引		173

# 第 1 章

## 序論

### 1.1 統計学の現状

**問題 1** 統計学 (statistics) はどんなところに使われているか。できるだけ書いてみなさい。

上の問題でみたように、医学、社会学、教育、理学、人文学、金融、経済、マーケティングなどなど (挙げるときりが無い)、統計学は様々な分野で活用されていて各学問ごとに、その分野専門の統計学が作られている。計量経済学とか、計量心理学とか、学問の名前の前に、「計量」とか「数理」○○とついているものは、だいたいその分野に統計学を導入した学問であると思って良い。統計学は現在あまりにも多くの分野で使われ、収集の付かない事態に陥っているといつてよい。裏返せば、どの分野でも使いたくなるほど魅力的で汎用性の高い学問であるということだろう。

統計学の手法はこのように、現在ではあまりにも広範囲に開発され過ぎている。そのため、その全体像は誰にも掴むことができない状態になっているのである。このような事情から、この講座では広範に発達した各分野の統計学に共通する主要な概念にしばって説明を行うことにした。

様々な分野の知識が入り乱れるこのような状況の中で、確かな頼りになるのは、やはり数式に基づいた統計学である。その理由の一例を挙げよう。統計学には同じ概念なのに、分野が違うせいで全く違う呼び名となっている対象が数多くある。そのようなときに、いくら文章の説明を読んでも細かな点ばかりが気になり、結局同じ概念なのか、似ているだけで違う概念なのか判断がつかない。しかし、数式を見れば一目瞭然、すぐに判断できるのである。

私は統計学とは何か？と  
言う質問は、スマートフォンとは何かという質問に似  
ていると思っています。ス  
マートフォンとは何かと尋  
ねられたら、あなたはど  
う答えますか？「電話、メ  
ール、ゲーム・・・うーん  
なんでもできるね！」と  
答えてスマートフォンを  
知らない相手は納得する  
でしょうか？

ひとつ例をあげてみま  
しょう。最も重要な統計  
量のひとつである標準  
偏差ですが、これは金  
融の文脈ではボラティ  
リティと呼ばれます。近  
接分野であるはずの  
経済学ではリスクと呼  
ばれます。これらはど  
れもばらつきを表す指  
標でありその式は同じ  
なのです。

このような考え方のもとに、本講座では出来る限り数式を丁寧に記述し、それと格闘することで本当に使える統計学を目指している。慣れない数式を追う事は、最初は辛いと感じる事もあるだろうが、おそらくこの講座を学び終える頃にはいくらか親しみを感じられるようになっていくはずである。

前提とするのは中学校の数学までで、高校数学については必要となった時点で、学習しながら進めていきます。数学についてはあまり心配せずとも大丈夫です。ただ中学範囲の式の展開や因数分解は前提なしに用いる場合があります。その辺りに不安がある方は、軽くおさらいしておいてください。

カイザー・ファンク『ヤバイ統計学』は題名はすこしくだけた感じですが、統計学の本質をやばいくらい面白く解説してくれます。似たような柔らかな統計本の中でも特に好きな一冊です。

## 1.2 統計的考え方

さて、統計学を学び始める前に、少し統計学の周辺について知識を広げておこう。こういう機会は統計学を本格的に学び始めてしまうと中々ないので、いくらか統計学と距離がある今のうちに、それを済ませてしまおうというわけだ。

まずここでは、統計学全体に通底するいくつかの思考法について見ていく事にしよう。アメリカの統計学者であり、ラジオ会社との専属契約を結び、人々に批判的なデータ分析を啓蒙するカイザー・ファンクは自著の中で、統計学者に共通する思考法のパターンについて以下の5つを挙げている。

- 1) 統計学者は統計上の平均という一般的な概念をあまり気にしない。むしろ平均からのずれ、偏差にこだわる。
- 2) あらゆることに合理的な説明を求めるのは私たちの本能的な欲求であるが、ばらつきについての筋の通った説明ができなくてもかまわない。統計学者は相関関係のパターンを見ているだけで幸せになれる。
- 3) 統計学者は、微妙な違いを見逃していないかと常に目を光らせる。
- 4) 統計に基づく判断は、2種類の間違いのあいだで折り合いをつけるように基準を調整できる。
- 5) 統計学者は「統計的検定」という特殊な方法を使って、例えば証拠と犯罪のつじつまが合うかどうか判断する。統計学者は奇跡を信じない。

[p.12 ヤバイ統計学]

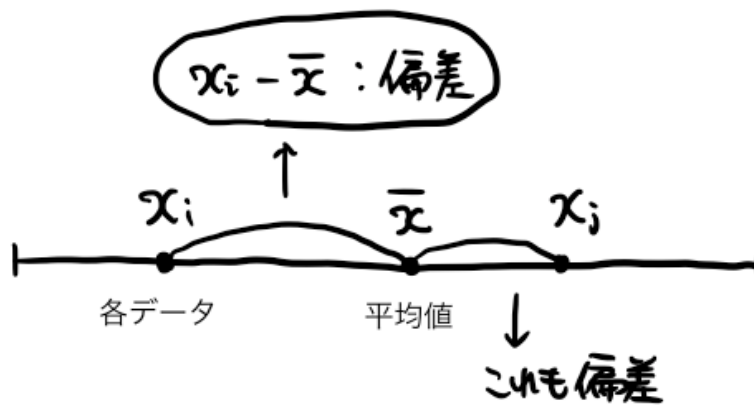
統計を学んだものなら、なるほどニヤリとさせられるものばかりである。ここに挙げられた統計学的思考は全て、統計学の本質をつく鋭い描察であることは疑いない（詳しくは本書を参照）。しかし、この中で特に私が強調しておきたいと思うのは1)、2)、5)である。その3点について少し解説をしてみよう。

ここで出てきた偏差という概念は統計学を理解する上でとても重要な概念です。偏差は各データが平均値からどれだけ離れているかを表す概念であり、数式では  $x_i - \bar{x}$  と表されます。 $x_i$  は各データであり、 $\bar{x}$  は平均値を表します。

まず1)のずれにこだわるというところだが、こだわりが生まれる理由は、ずれこそは理想世界の住人である数学者にとっての、最大の敵ともいえるべき存在だからに他ならない。数学者は調和を好み、拡散した状況を好まない。このずれの原因を突き止めてやろうと躍起になるのである。

ずれはばらつきとも言い表せる。平均からのずれを偏差 (deviation) と呼ぶが、我々の現実に照らして考えてみても、問題を引き起こすのはいつも平均 (予測できる物事) ではなく、そこから外れた (予想外の) 出来事であるはずだ。





日本人の多くがあれほど混んでいる（殺人的な乗車率の）電車に毎日乗り込んで通勤するのは、なぜか考えてみると良い。すぐに平均よりも、そこからのずれ「偏差」の方が注目に値するのだということが納得いただけるだろう。

**問題2** 平均からのずれが引き起こす問題について、何か一つ事例を挙げて説明してみなさい。

続いて、2) について説明する。これは統計学という学問と通常の科学と呼ばれるものを比較して考えてみると分かり易い。例えば、「ある薬がある病に効く」ということをどう証明するか考えてみよう。

通常の科学、この場合は薬理学では、おそらく以下のような説明になるだろう。

「この薬には、〇〇という成分が、××グラム含まれていて、服用後□□分後に、ドコドコという器官にホニャララ% 吸収され、血中濃度が☆☆  $\mu\text{g}/\text{l}$  になり、病原体に対して、△△という反応を起こしますので、この薬はこの病に効きます。」

一方の統計学はこうである。

「100人に服用させたら、95人治ったので、この薬はこの病に効きます。」

私は薬理学に詳しいわけではないので、想像で書いています。全然違う(笑)という場合は授業後に教えてください。

**問題3** 上の例を参考にしながら、通常の科学の方法と統計学の方法の利点と欠点を指摘してみなさい。

最後に5)について考えよう。統計学者は奇跡を信じないとある。念のために言っておくが、これは別に特定の宗教批判を展開しているとか、そういうわけではない(統計学者には信仰心に篤い人だって、多分たくさんいる)。次の問題を考える事で、この統計学者の真意を問いただしてみよう。

数学者、統計学者、会計士の登場するアメリカンジョークにこんなものがあるそうです。

【数学者と統計学者と会計士が、同じ仕事に応募した。

面接官は数学者を呼び入れて、問うた。「 $2+2$ は何になりますか?」数学者は「4です」と答えた。

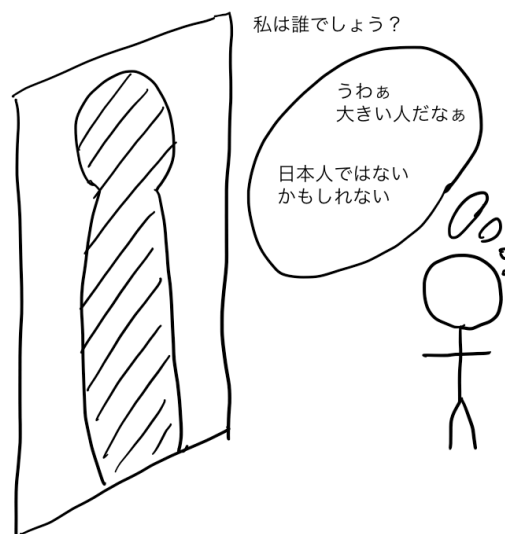
面接官は念をおした。「ぴったり4ですか?」数学者は疑わしいような姿で面接官を報告、答えた。「はい、ぴったり4です」

次に面接官は統計学者を呼んで入れて同じ質問をした。「 $2+2$ は何になりますか?」統計学者は「平均して、4、誤差10パーセントですが、だいたい4です」

終わりに面接官は会計士を呼び入れて、同じ質問をした。「 $2$ たす $2$ は何になりますか?」会計士はドアに鍵をかけて窓のシェードを下げて、面接官の横に座って問った。「 $2+2$ をいくつにしたいんですか?」

**問題4** 命題「身長2m以上の人は、日本人でない」は正しいか。

この命題の真偽に象徴されるように、統計学者はあらゆる問題を確率的に判断している。つまり、「非常に稀な出来事が起こる事はない。」と考えるのである。しかし不思議な事に、それと矛盾するような考え方。「数回であれば例外も起こりうる。」も認める。この一見ダブルスタンダードに見える考え方こそ、統計学を学んだものの考え方である。



問題 5\* 「偶然」ということについて、議論して見なさい。

竹内啓, 『偶然とは何か-その積極的意味-』は様々な話題に触れながら、統計の考え方に触れられる名著です。数学の本のわりに新書で安いのもオススメできるポイントですね。

## 1.3 統計学の分類とその対象

### 1.3.1 記述統計学と推測統計学

次に統計学を大きく分類する事を考えよう。統計学は大きく分けると、標題の通り 2 つに別れる。これは各統計学の目的の違いを手がかりにした分類である。

記述統計学・推測統計学

記述統計学 (descriptive statistics)・・・集団としての特徴を記述するため、観測対象となった各個体を観測し、得られたデータを整理・要約などして、データを正しく効率的に把握するための統計学。

推測統計学 (inferential statistics)・・・一部を観察して、そこから論理性のある推測で全体の法則性の発見にいたることを目的とした統計学

問題 6 以下のニュースや日々の出来事には、どちらの統計学の範疇にある手法が用いられているか、考えてみよ。

- ・生活保護受給者ついに 215 万人を超える。 (厚生労働省)
- ・半沢直樹！ 最終回瞬間視聴率 40 % 超え！ (TBS)
- ・Male finger lengths related to attractiveness(The Daily Telegraph/英)

The Daily Telegraph は 1855 年創刊の英国の大衆紙です。記事の内容は、「人差し指よりも、薬指が長い男性ほど、異性を惹きつける」というものでした。

問題7 記述統計学的な事例と推測統計学的な事例をそれぞれ一つずつ挙げて、それがなぜその分類となるか説明しなさい。

### 1.3.2 母集団と標本

次に我々は、統計学的手法をどのような対象に対して用いているかということについてよくよく考えてみなければならない。

統計的な手法を用いて何かを分析するとき、結論として何を示したいかについて考えない人はいないと思いますが、その結論がどのような集団について（限定的に）言えるのかということには無頓着であることが多いように思います。

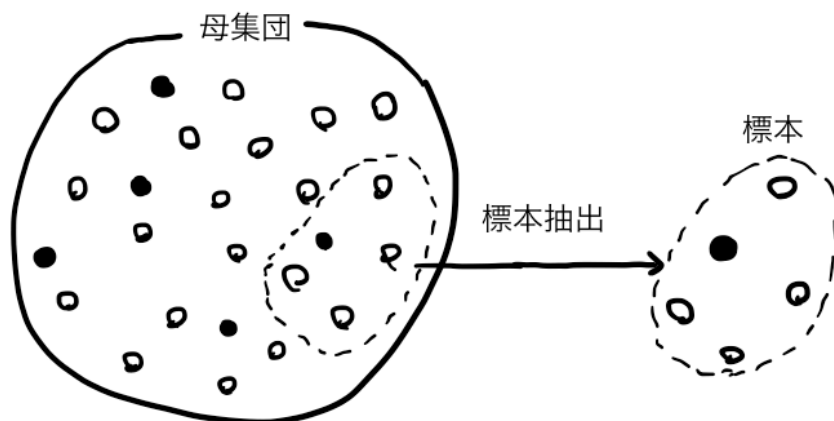
実は、この sampling という作業は大変でデリケートな作業で、統計分析に足るようなデータをきちんと集めるには、大変な苦労があるのです。Gabege In! Gabage Out! という評語が統計学にはありますが、これはどんなに高級な分析をしても、ゴミのようなデータを用いてはゴミのような結果しか出ないという戒めです。本講座ではあまりこの問題について面倒な議論をする事を避けています。この問題には付録にて簡単にですが触れてあります。しかしそこに書いてある内容は sampling 論の極一部であり、もっと奥深いのです。

母集団、または標本のサイズのことを母数と呼ぶ人がいますが、統計学ではサイズのことを母数とは決していいません。なぜなら母数という言葉には違う意味があるからです。でも日常の用法ではサイズのことを分母にくる数だから母数と呼ぶこともありますね。

#### 母集団・標本

我々がこれからの調査によって知りたいと思う集団全体のことを 母集団 (universe, population) という。また母集団から分析のために選び出された要素、またはその属性値を 標本 (sample) と呼び、標本を選び出すことを 標本抽出 (sampling) または、単に 抽出 と呼ぶ。母集団または標本に含まれるデータの数をその集合の サイズ (size) と呼ぶことがある。

母集団全体を調査する場合を 全数調査 (complete survey) もしくは 悉皆調査 と呼ぶ。また母集団から選ばれた標本を調査することを 標本調査 (sample survey) という。



例えば、日本人全員の意識調査を行おうとしたときの母集団は、日本人全員である。しかしながら、現実的には、予算、時間等の関係から、母集団全体を調べることが不可能である場合が多い。その場合には全体の傾向を代表すると思われる適切な標本を調査し、それを分析して母集団についての推測を行うのである。

問題8 以下の調査における母集団と、用いられていると思われる調査方法を言え。

- 1) 衆議院選挙 当選速報
- 2) ○○社 健康診断
- 3) 雑誌 読者アンケート

問題9\* 横の注釈にある、「母集団全てを調べるのが不可能あるいは無意味な場合」の実例を考えて見なさい。

予算、時間等の問題とは別に、調査の性質上、母集団全てを調べるのが不可能あるいは無意味な場合があります。それはどのような場合か考えてみてください。

## 1.4 統計学の歴史-3つの統計思想-

統計学は数学の中で比較的新しい分野だという評をもらう事が多い。しかし、それはどこまでを統計学と見なすかによる。統計学には以下の小節で述べるように3つの源流がある。それが一つに纏まって今の統計学を構成している。

この節で我々はその3つの源流を概観し、統計学の歴史を辿っていくことにしよう。

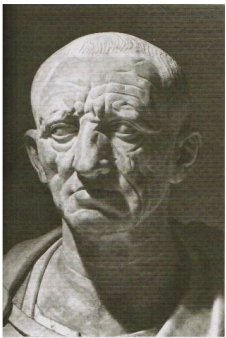
### 1.4.1 国勢学

3つの源流のうち最も古くからあるのは、国の実態を把握する為に用いられた統計学である。この流れを遡っていくと、古代文明にまで遡る。古来より為政者・政治家は自

この辺りの記述は総務省統計局の提供するホームページ『統計学習の指導のために』やデイヴィッド・サルツブルグ『統計学を拓いた異才たち』を参考にしています。

ら治める国について、なるべく正確な情報を得ようと努力してきた。徴兵や納税といった国力を左右する大事は、国の正確な情報を持っていなければ大変困難な作業になる。それゆえ、国家運営には国に関する情報をまとめ、もれなく管理する術が欠かせなかったのである。

古代ローマにはケンソルと呼ばれる職の役人がおり、ケンソルの行う調査を指して、センサスと呼んでいたようです。他の政務官職が毎年選出されるのに対し、ケンソルは5年に1度、18ヶ月の任期で選出される。定員は2名



Marcus Porcius Cato Censorius

今、我々が知る事が出来る情報の内、最古の記述は古代エジプト文明におけるピラミッド建設のための基礎調査であると言われる。その他にも紀元前 2300 年ごろの中国においても、大規模な人口調査があったとされている。

また、ローマ帝国では初代皇帝アウグストゥス (Gaius Julius Ceaser Octavianus Augustus B.C.63-A.C.14) の時代に大規模な国勢調査 (Census) がおこなわれたという記録が残っている。今も国勢調査のことをセンサス (Census)というが、これはこの時代の名残であるとされる。

15~17 世紀、ヨーロッパの大航海時代には、各国がその国力の増大を目指して競い合う時代に入った。国家の繁栄は人口や貿易に依るという考えから、産業や貿易に関する数量データの調査・研究が各地で行なわれるようになる。

特にドイツでは、学問的な体系をもった国勢学 (census studies)が発達し、人口のみでなく細部にわたっての調査が行われるようになった。これにより、情報の利を得たドイツはヨーロッパの中でも大国へと成長し、この流れはヨーロッパ全域に広がっていく。統計学を statistics(英) と呼ぶが、これはドイツ語の statistik(独) から来ているとされている (statistik(独) の語源はラテン語の status(国家・状態))。

こうして、現代でも行なわれている国勢調査に基づく国状の把握という目的を持った統計学が完成していくのである。

#### 1.4.2 近代統計学の起こり

現代まで残っているものの中で、近代統計学に関連する記述が見られる最初の書物は、ジョン・グラント (John Grant)(1620-1674/英) の「死亡表に関する自然のおよび政治的諸観察」(1662) である。ここには例えば以下のような記述がある。

6. かつて生を享けた者の全体の約 1/3 は 4 歳以下で、また 36/100 は 6 歳以下で死亡すること。

9. 1/100 は老年のために死ぬということ。

52. ロンドンにおいては 13 人の女子に対して 14 人の男子があり、地方においては 14 人の女子に対して 15 人の男子しかいこと。

[抜粋, 死亡表に関する自然のおよび政治的諸観察]

前の小節でみた国勢学が、全てのデータを収集し、記述し、整理することが目的であるとすれば、グラントの卓越性は丁寧選ばれた小集団 (すなわち標本) を観察し、そのデータから一見不秩序に見える物事の背後に潜む一般的な関係を記述しようとした点に

ある。

また、これと同時期にはウイリアム・ペティ (William Petty)(1632-1687/英) が「政治算術」という本を著している。この本の中でペティは戦争により疲弊したイギリスを他の欧州諸国と定量的に比較し、国政に役立ててくれるよう嘆願したのである。これは一見「国勢学」の流れに位置するように思えるが、その内容は単なる事実の整理・記述に留まっておらず、グラントがしたのと同じように、その背後にある因果関係の分析などにも触れており、やはり国勢学よりもさらに発展的であると考えるのがよいと思う。



William Petty/英

**問題 10** 国勢学と近代統計学の違いについてまとめてみよ。

### 1.4.3 確率論と統計学

上でみた流れとは別に数学者が偶然を扱う理論として考えだしたのが確率論 (probability theory)である。確率論は数学者ブлез・パスカル (Blaise Pascal) (1623-1662/仏) とピエール・ド・フェルマー (Pierre de fermat)(1600 初-1665/仏) のサイコロ賭博に関する往復書簡によってその基礎が生まれたとされている。

今まで見た流れは、現在の分類で言えば数学というよりはどちらかというところ、社会科学のような方法論に近い。ここでやっと数学の出番である (とはいってもグラントやペティとパスカル・フェルマーはほぼ同時代人ではあるが)。統計学に数学の視点が入る事によって、観察が主な手段であった統計学に、理論からの光が差し込む事になる。

**問題 11\*** 経験確率 (empirical probability)と、理論確率 (theoretical probability)という言葉がある。その語句の意味を調べてみなさい。また、確率論でよく使われるサイコロ投げの例でそれを説明して見なさい。

サイコロ賭博を途中でやめる際に、掛け金をどのように配分するのが公平かというような内容について話し合っていたのだといわれています。賭博よりもそちらの方に熱中してしまうあたりさすが偉大な数学者ですね。

フェルマーとパスカルが確率論を創始して後、統計学の発展は主に（完全に数学者だけではないが）数学者の手にゆだねられることになる。ベイズ (Thomas Bayes 1702-61/英)、ラプラス (Pierre-Simon Laplace 1749-1827/仏)、ガウス (Carl Friedrich Gauss 1777-1855/独) などの数学者が次々に生まれ、統計学を発展させた。驚くべき事に、この後わずか 200 年くらいでほぼ現在の形の統計学を作り上げてしまうのである。以下にこの後の統計学の発展を担った幾人かを挙げておく事にしよう。

#### 近代統計学を築いた人々

1) ラプラス (Pierre-Simon Laplace 1749-1827/仏)

古典的確率論を完成させた。統計学の基礎を作った。

2) ガウス (Carl Friedrich Gauss 1777-1855/独)

正規分布の発見、最小二乗法等。

3) ゴルトン (Francis Galton 1822-1911/英)

遺伝学の数学理論、「回帰」の導入。

4) ピアソン (Karl Pearson 1851-1936/英)

相関係数、 $\chi^2$  統計量など。近代数理統計学の数理的基礎。

5) ゴセット (William Gosset 1876-1937/英)

$t$  分布、小標本についての理論。

6) フィッシャー (Ronald Fisher 1890-1962/英)

母集団と標本に関する理論。統計的推測理論、実験計画法、 $F$  分布

7) ワルド (Abraham Wald 1902-1950/トランシルバニア)

統計的決定理論、検定の理論。

以上で見たように、歴史的には 3 つの源流が統計学を作り上げてきたと言ってよいだろう。そして現在では、この 3 つの大きな流れは統合され一つの巨大な体系を形作っているのである。

他分野で活躍した多くの数学者がそうであるように、統計学を作り上げた数学者たちも、面白いエピソードには事欠かない人生を送っています。中でも表中 6) のゴセットはすべての論文を「student」という名前で投稿しました。今後学習が進むと student の  $t$  検定という検定手法が登場しますが、これはゴセットの手によるものです。彼がなぜそんなことをしたのか興味がある方は、講師にお尋ねください。



## 1.5 いくつかの注目すべき事例

この章の最後に復習を兼ねて、いくつか統計学の応用事例について考えてみることにする。

**問題 12** 東京ディズニーランドにはファストパスという制度があるが、あの制度の中心的なアイデアは何か。説明しなさい。ちなみにあの制度は統計学者が考案した制度である。

**問題 13** あるスポーツ系雑誌によるところでは、皇居ランナーを続けている年数が長いほど、その人の年取は高いという。相関係数は  $r = 0.72$ 。「したがって、年取を上げたいなら今すぐ皇居ランナーになるべきである」という主張がなされていたが、これは正しいだろうか。

まだ相関関係ということに関して詳しく話していませんが、ここではどちらかが多いと、もう片方も多いという単純な関係のことだと考えてください。ちなみに問題の  $r = 0.72$  は一般的に強い相関があると考えられます。