＼ー／ ＼ー／ ＼ー／　やさしいパソコン統計入門　＼ー／ ＼ー／ ＼ー／

2014 年度　面接授業資料　安田正實（千葉学習センター）

[統計学テキスト／ pasokon 統計入門.tex]

1　統計の概要

これから学ぼうとする統計学は、学校教育の科目では「数学」の中に取り扱われています。数学の知識は統計を理解し、活用するためには必要な事柄です。しかし、数学とはちょっと異なった側面をもっていることに、みなさんは気が付かれているかも知れません。数学が公理的で明確な論理をもっていますが、一方、統計の目的には真の状態を把握しようと、調査、データの解析をしますが、結論としての正解が曖昧になる点があります。基礎的な部分には数学を必要としますし、多くの数式でいろいろな統計モデルが展開されていきます。しかし統計は決して、正確で間違いのない真理を与えてはくれません。これから、統計学における重要な事項を学びますが、とくにここで強調したい点は、統計を理解し、それをさまざまな状況で活用できるためには、問題として提起された事柄についての貴重な経験、いろいろな関連分野での知識をもつことが重要であり、心理学、経済学、経営学などをはじめとし、社会科学の分野や工学、医学において役立て活用できるよう、明晰な思考、批判的な解釈の基本的な技能を学んでいこうと考えています。

1.1　メディアからの統計

《統計改革ーもっと生かして使うには》朝日新聞「社説」２０１０年１２月１４日（火）からの引用です：

政府の公的統計を内外に広く公開することで、新たな活用の道を開くことができないだろうか。社会保障や公共事業などの政策をつくる際、判断材料とするために政府はさまざまな公的統計を定期的に集めている。多くは現状を把握するためのデータで、政府も利用の拡大に取り組んではいる。従来は行政担当者だけが使う想定のもとに、さまざまな利用制限が設けられていたが、２００９年４月施行の法改正で、社会全体が活用できる共有財産と位置づけられた。背景には、「個票」と呼ばれる統計の個別データを学術研究や経済活動に二次的に利用できるようにして、その成果を政策に反映させようとする世界的な動向がある。政策に一貫性を持たせたり、政策の変更を国民に納得してもらったりするためには、統計に基づく実証的根拠が以前にも増して大切だ。だが、政府だけでは限界がある。統計データを公開し、多くの研究者や専門家に、官僚では思いつかない視点から分析・検証してもらうことが重要になってきた。変化が早まり、社会が複雑化する中では、政策の場をオープンにしたほうが質も効率も高められる、との判断がそこにはある。だが、施行から１年半たった今も、二次利用はあまり進んでいない。窓口の独立行政法人統計センターによると昨年度の利用件数は２７、今年度は１２月１３日時点で４２にとどまる。学者や研究者の間から問題点としてあがっているのは、二次利用が可能な形で提供されている統計数の少なさだ。そうした統計は国勢調査や全国消費実態調査、就業構造基本調査など、まだ２０に満たない。今年度中にもう少し増える見込みだが、公的統計のうち特に公共性が高いとされる基幹統計だけでも５０余りあることを考えれば、公開のペースは遅い。個人情報保護の観点からデータを加工処理する作業などに時間がかかるとはいえ、政府全体で取り組もうという姿勢は見えない。日本の公的統計は縦割り行政の弊害が大きく、同じ省内でも個々の統計の関連づけといった体系的な整備がなされていないものが少なくない。データの取り方や設問が異なり、複数の統計を組み合わせる高度な分析が困難な例も目立つ。今後は利用にも配慮した統計の総合管理が必要だ。所得格差の拡大や少子高齢化といった問題を抱えている日本が、積極的な統計整備を進めれば、課題解決に向けて世界の英知を集めることも夢ではないだろ

う。官僚に頼らない政治主導の政策づくりにも生かせる。さまざまな効用を視野に、「統計大国」への道を歩みたい。

1.2　統計の制度、法律

http://law.e-gov.go.jp/htmldata/H19/H19HO053.html 統計法（平成 19 年法律第 53 号)

【目的】　第一条　この法律は、公的統計が国民にとって合理的な意思決定を行うための基盤となる重要な情報であることにかんがみ、公的統計の作成及び提供に関し基本となる事項を定めることにより、公的統計の体系的かつ効率的な整備及びその有用性の確保を図り、もって国民経済の健全な発展及び国民生活の向上に寄与することを目的とする。

【基本計画】　政府は、公的統計の整備に関する施策の総合的かつ計画的な推進を図るため、公的統計の整備に関する基本的な計画を定めなければならない。総務大臣は、統計委員会の意見を聴いて、基本計画の案を作成し、閣議の決定を求めなければならない。政府は、統計をめぐる社会経済情勢の変化を勘案し、及び公的統計の整備に関する施策の効果に関する評価を踏まえ、おおむね五年ごとに、基本計画を変更するものとする。

【基幹統計】　（国勢統計）総務大臣は、本邦に居住している者として政令で定める者について、人及び世帯に関する全数調査を行い、これに基づく統計を作成しなければならない。

（国民経済計算）内閣総理大臣は、国際連合の定める国民経済計算の体系に関する基準に準拠し、国民経済計算の作成基準を定め、これに基づき、毎年少なくとも一回、国民経済計算を作成しなければならない。

【調査票情報等の保護】　（調査票情報等の適正な管理）（調査票情報等の利用制限）（守秘義務）（調査票情報等の提供を受けた者による適正な管理）（調査票情報の提供を受けた者の守秘義務等）

1.3　統計の窓口

「政府統計の総合窓口」では、任意時点で検索・出力が可能な形式でデータの提供をしています。おもな統計調査は国勢調査，経済センサス，人口推計，労働力調査，家計調査，消費者物価指数など統計局が実施され、政府の施策、予算の策定など、国家の維持には欠かせない情報です。

検索には「e-stat」いれてみましょう。

このホームページには、いわゆる官庁統計とよばれる重要なデータが毎日発表され、年次データ、月別の集計等により更新されています。ttps://www.e-stat.go.jp/estat/guide/basic/index.tml では活用術として説明されています。

上の切り抜き部分には、右上に統計について勉強しよう、「統計を知る・学ぶ」というリンク個所があります。これをクリックすると、楽しく統計を勉強できるとおもいます。

下のホームページは ttp://www.pref.ciba.lg.jp/toukei/toukeidata/hiroba/ 千葉県の統計をまとめた総合案内です。一度見てみると面白いでしょう。

2　統計調査のいろいろ

統計調査あるいはアンケートも同じですが、実行をする際には、実施の規模が違っていても、まず注意しておくこととして、つぎが挙げられます。

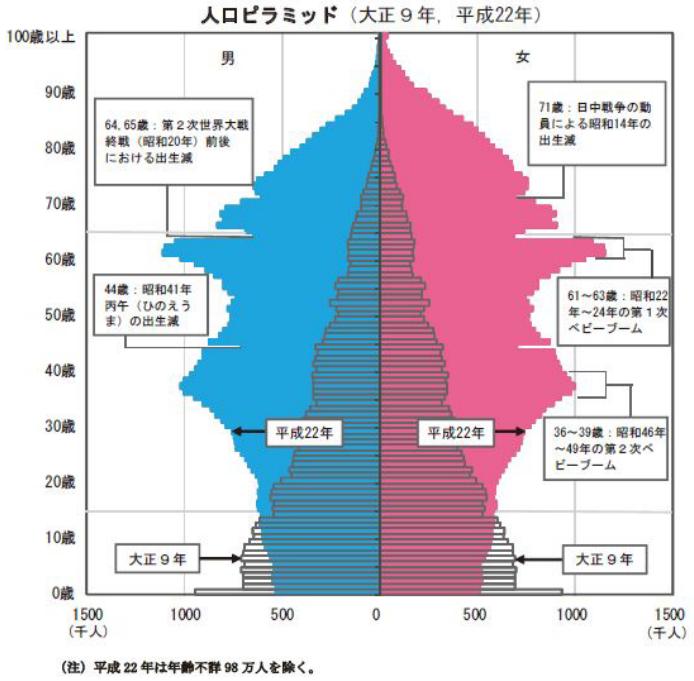
- 調査の目的、主体者、法律
- 調査事項、調査内容
- 方法と時期、集計方法
- 調査の対象や地域
- 調査表の形式、項目
- 集計結果の公表、報告書、結果の概要

それでは、まず最も大規模であり、重要な日本の代表的な統計調査からみてきましょう。

2.1　国勢調査

国勢調査は，我が国に住んでいるすべての人と世帯を対象とする国の最も重要な統計調査です。　調査の結果は，国や地方公共団体の行政施策のほか，民間企業等でも様々な場面で利用されています。調査対象は日本に住居するすべての人ですから、センサス (Population Census) とよばれます。一部分のみを抽出して調べる標本調査ではなく、正確なデータが求められます。5 年に一度調査員が各家庭に回って調査票を回収しています。

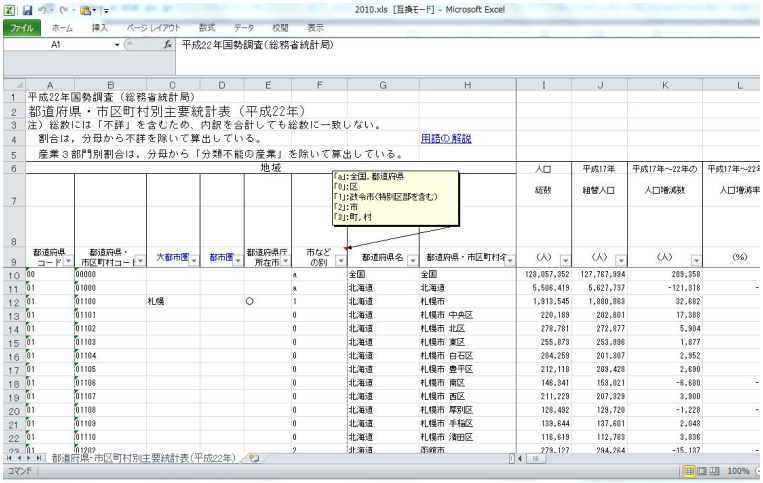
平成 22 年 10 月 1 日現在の我が国の人口は 1 億 2805 万 7352 人，世帯数は 5195 万 504 世帯で、これを集計することもコンピュータのおかげで迅速に行われるようになりました。集計結果から、よく知られているよ



（注）平成 22 年は年齢不詳 98 万人を除く。

う、人口ミラミッドをつくることができます。富士山型（大正９年）からひょうたん型（平成 22 年）へ大きく変化していることも統計局の e-stata には説明がされていて、少子化対策の重要なテーマで、単に集計だけではなく、現況を取り巻く社会環境の経済的、文化的な要因を考察する難しい課題となっています。

公表されている Excel file をダウンロードできますから、これからグラフを作成することできます。



「e-stat」に登録しておくと、更新の情報が送られてきます。

更新情報

平成 25 年 10 月 29 日　抽出詳細集計（全国及び北海道など 12 都道府県）

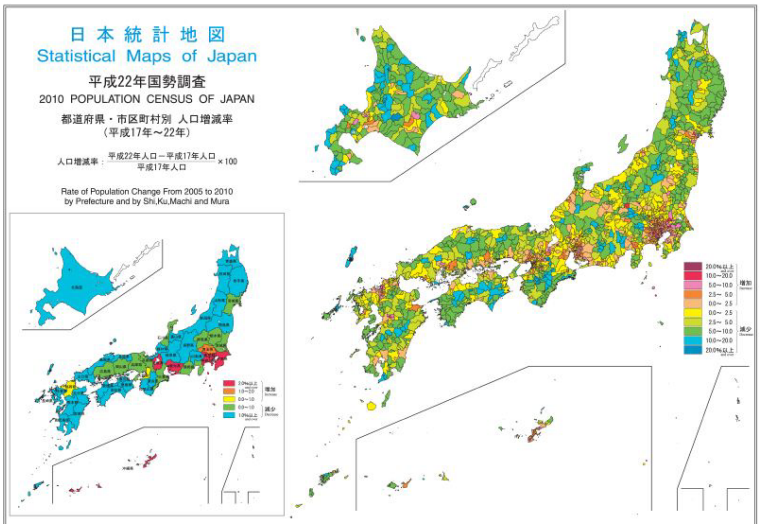
平成 25 年 10 月 29 日　従業地・通学地による抽出詳細集計

平成 25 年 10 月 29 日　キロ圏・距離帯及び大都市圏・都市圏

平成 25 年 7 月 30 日　抽出詳細集計（群馬県など 13 府県分）

平成 25 年 3 月 26 日　移動人口の職業等集計

数値データをグラフや統計地図として処理をしている。都道府県別の人口増減率についてが発表されています。分析データを簡単に把握し、表現する方法です。



2.2　経済センサス

経済センサスは、事業所及び企業の経済活動の状態を明らかにし、我が国における包括的な産業構造を明らかにするとともに、事業所・企業を対象とする各種統計調査の実施のための母集団情報を整備することを目的としています。経済センサスは、事業所・企業の基本的構造を明らかにする「経済センサス-基礎調査」と事業所・企業の経済活動の状況を明らかにする「経済センサス-活動調査」の二つから成り立っています。

地方消費税は最終的に消費が行われた都道府県の税収となるよう、各都道府県の「消費に相当する額」に応じてあん分されています。この「消費に相当する額」は、地方税法施行令及び同法施行規則に定められた「消費に関連する指標」に基づいて計算され、その一つとして経済センサスの都道府県別従業者数が利用される。清算の後に都道府県の収入となった地方消費税の 1/2 は、安定的な財政基盤確立のため、市町村へあん分して交付されます。あん分は経済センサスの市町村別従業者数等に基づいて行われる。

主要な指標から　平成 24 年 2 月 1 日現在の我が国の企業等の数は 412 万 8215 企業、事業所数は 576 万 8489 事業所、従業者数は 5583 万 7 千人。平成 23 年 1 年間の売上（収入）金額は 1335 兆 6410 億円、付加価値額は 244 兆 7620 億円。$e-stat$ のトップページからホーム ＞ 統計データ ＞ 経済センサス ＞ 経済センサス総合ガイド ＞ 基礎調査 ＞ 基礎調査結果の利用 ＞ と進めばこんなデータがわかります。もちろん経済施策の分析、効果を調べ、新たな策定のためには重要な役割を果たします。

- 我が国の総事業所数は 635 万 6 千事業所
  - うち事業内容等が不詳の事業所を除いた事業所数は 604 万 3 千事業所
  - うち事業内容等が不詳の事業所を除いた従業者数は 6286 万 1 千人（平成 21 年 7 月 1 日現在（確報値））
- 産業分類別（大分類 19、中分類 97、小分類 591）に事業所数及び従業者数がわかります
- 男女別に従業者数を把握することで男女共同参画の実態などが明らかになります

3　統計を知る学ぶために

学習指導要領改訂により内容の充実が図られた統計教育学習に対応すべく、高校生用統計学習サイト「How to 統計（平成 16 年開設）」の全面的な見直しを行い、「なるほど統計学園高等部」として、平成 25 年 4 月 5 日から公開しています。

[主なコンテンツ] 統計の作成・分析 （統計の企画・計画からデータ収集・分析・結論までを学ぶことができます。）主要統計データ （統計局のデータを中心に、日本国内や世界の様々な統計データの探し方を案内します。）統計分析事例 （統計分析から導き出される興味深い様々なトピックスを分かりやすく解説します。）豆知識 （よく統計が使われる業界やその場面の紹介のほか、統計の発展に貢献した人々のエピソードや統計の歴史等を解説しています。`http://www.stat.go.jp/index.htm`

4　統計が最強はうそ！？

統計にダマされないためには、だます方法を知るべきということで、統計結果を悪い使い方、あるいは結果として欺くことに陥る危険を解説しています。

書籍名：**統計でウソをつく法: 数式を使わない統計学入門** 著者：ダレル・ハフ、翻訳：高木秀玄、出版社：講談社, 1968. ISBN：4061177206, 9784061177208.

内容の一部紹介：だまされないためには、だます方法を知ることだ！かの有名な英国の政治家ディズレーリは言った——ウソには 3 種類ある。ウソ、みえすいたウソ、そして統計だ——と。確かに私たちが見たり聞いたり読んだりするものに統計が氾濫しているし、「平均」とか「相関関係」とか「トレンド」とか言って数字を見せられ、グラフを示されると、怪しい話も信じたくなる。しかし、統計数字やグラフは、必ずしも示されている通りのものではない。目に見える以上の意味がある場合もあるし、見かけより内容がないかもしれないのである。私たちにとって、統計が読み書きの能力と同じぐらい必要になっている現在、「統計でだまされない」ためには、まず「統計でだます方法」を本書によって知ることが必要なのである！

この本の目次:

§1 かたよりはサンプルにつき物　§2 “平均”でだます法　§3 小さい数字はないも同然　§4 大山鳴動ネズミ 1 匹　§5 びっくりグラフ　§6 絵グラフの効用　§7 こじつけた数字　§8 因果はめぐる　§9 統計操縦法　§10 統計のウソを見破る 5 つのカギ

■サンプリングをしたときに，それが元のデータ（全体）のサンプルとして偏りがないかを確認しなければ，

そのサンプリングによるデータは，意味がない。(＊母集団と標本抽出) ■アンケートをした場合，回答者が本当のことを言うとは限らない。たとえば，「主に読んでいる本」を調査しようと思ったら，「何を読んでいますか」と質問などするより「訪問して，古雑誌を買いたい」と言った方が，ずっと多くのことがわかるということ。ただし，これでも本当のことはわからない。これでわかるのは，読んでいる本ではなく買った本であるからだ。（＊質問項目の立て方、作為的な誘導質問と意図とする結論）■サンプリング調査の結果が，もとになるサンプリングより正しくないことも事実なのであるが，データが何回も統計的操作で濾過され，小数点のついた平均値に姿を変える頃には，その結果はもとのデータとは似ても似つかないような確信の臭気を身につけ始めるのである。（＊データからあるべき計算結果の有効桁数、無意味で冗長的な計算）■サンプルの基礎は「ランダム」という性質がなければならない。つまり，サンプルは「母集団」からまったく偶然に選ばなければならない。ランダム・サンプルであるかどうかの判定は次のようになされる。「母集団の中のすべての人あるいはものは，等しくサンプルに選ばれるチャンスがあるか？」■さて，少人数のグループを使うことが意味があるのはこうだからだ。つまり，大きなグループを使ったのでは，偶然による差がどうしても小さくなってしまうし，それでは，大見出しが使えるような結果はえられないからである。■試行回数が十分に多い場合に限って，平均の法則は説明や予測の役に立つのだ。■何も知らないことの方が，不正確なことを知っているより健全である場合が多いのであって，少しばかり勉強するというのは，かえって危険なことなのかもしれない。■常識というものは，もっともらしく正確で，厳然としている 3.6 などという数字の前では，どうも弱いのであって，この場合は，調査から誰にもわかるようなこと，すなわち，多くの家庭は小家族で，大家族はほんのわずかであるという事実も常識にはかなわなかったのである。■誤りが起こるのは，結果が研究者から煽動的あるいは情報不足の記者を通して読者に届くまでの情報の濾過過程にある。そして読者というものは，その過程で姿を消してしまった数字には気がつかないのである。誤解の多くは，「標準」や平均に分布幅についての注があれば避けることができる。■プラス・マイナスの誤差ということは，いつも心にとめておかなければならないことで，誤差が書いてないときでも，あるいは書いてなければなおさらのこと，注意しなければいけない。ときには，数学的には実在し，しかも証明できても，現実には意味がないほど小さい差をめぐって，大騒ぎされることがある。－要するに「目くそ鼻くそを笑う」である。■グラフの目盛りを拡大してみせるものに対して曰く－これは「国民所得 10 %の増加」というのを「…10 %の飛躍的伸び」とする編集方法と同じようなものである。しかし，それより断然効果的である。なぜならグラフには形容詞や副詞がないので，客観性という幻影がこわされることがないからである。ここには誰からも，突っつかれるようなものは何もないのである。

このほかにも、NHK の番組「数字のからくり、データの真実」(http://www.nhk.or.jp/gendai/kiroku/detail02_3375_all.html) があります。統計データを挙げて説明する主張には、それなりの明確さ、正確性などをもつ反面、一方には意図的な作為や誤りが含まれているときがあることを注視すべきで必要があります。数字には強くなるべき！！(http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1066282195)

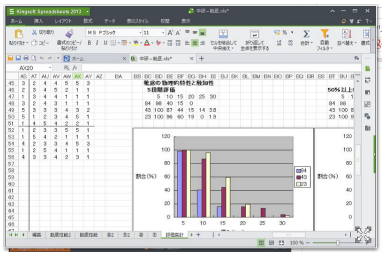
5　F. ナイチンゲールによる政府への「主張する統計」

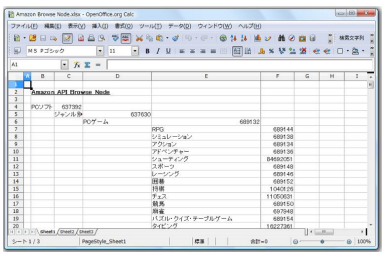
有名な白衣の天使とよばれるナイチンゲールは、当時のヨーロッパ転換期となったクリミア戦争（1850 年ごろ）で、戦地の死傷者に関する統計データをもちいて、当時のイギリス政府に働きかけに貢献した、「戦う天使」であった。有名な統計図「蝙蝠のはね」というヒストグラムをもちいた。戦争での負傷より、負傷者救済のための病院管理が整備されていないという主張をしたといわれています。http://www.stat.go.jp/teacher/c2epi3.htm, http://www7.ocn.ne.jp/~ooguro/nightingale.htm

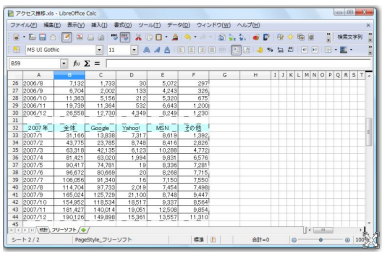
多尾清子 [統計学者としてのナイチンゲール (2001)] によれば、ナイチンゲールの統計学は「記述（数理）統計学と社会統計学のそれぞれの定義に基づいた内容を併せもつもの」で、記述統計学は数学の実際的応用に志向し、社会統計学は社会生活の合法則性の究明に志向している。社会統計学の対象は集団であり、その性質を数量的に研究する。多尾はナイチンゲールの統計学を「記述社会統計学」と名づけている。ナイチンゲールの 7 つの素顔 : http://www.nightingale-a.com/page-fn7-7.html 決してやさしい天使とよばれる看護師などではなく、無知な政府、不理解者へと戦った戦士なのです。

6　パソコンと表計算ソフト

今日、パソコンで利用される統計学ソフトウェアには、よく知られた (1) 表計算ソフトと統計解析を本格的に処理して、(2) プログラミングできるソフトウェアがあります。代表的なものとしては、(1) では、マイクロソフトの excel(有料) がよく知られていますが、下図で左側から KINGSOFT Office（有料）、Apache OpenOffice(無料)、LibreOffice(無料) など、互換性を保つソフトなどがあります。







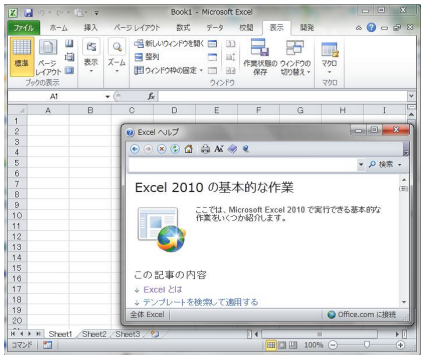
6.1　専門家もつかうソフトは

「もうバカ高い統計ソフトを医学研究に使う時代ではない」と指摘したブログ http://www.kamakuraheart.org/wordpress/?p=1115 では、これまで、まともに統計処理をしようとすると、ば

か高い (20 万円から 50 万円) ものお金を払って、ソフトを購入せねばなりませんでした。世の中にはそのようなソフトとして有名なのは、SAS, STATA, SPSS あたりがあります。それぞれソフト一本の基本セットのみで、20 万円ぐらいしますし、色々なパッケージを揃えると簡単に 50 万円ぐらいになります。といっています。また `https://sites.google.com/site/statinfra/comparison` では、統計ソフトの比較をしています。このブログの結論でも”医学部学生・大学院生・レジデントで初めて統計学を学ぶ際には，統計ソフト R が費用・入手しやすさ・学習環境・将来性の 4 点から最も勧めることができる．ただし，すでに使い慣れた統計ソフトがあり，今後も継続して使い続けられる見込みがあれば，わざわざソフトを他に変えるまでもない．”としていますが、これは日常の医学研究で積み重ねたデータを将来のテーマでの解析に使うことを重要な主眼としています。もし本格的に統計データの利用をめざすならば、このソフトウェア R が (1) 無料，(2) 機能が豊富，(3) 学習資料も無料で入手できるメリットでお勧めです。もちろんこれは報酬を目的としてグループの活動ではなく、統計をもっと多くの人々に利用、応用して素晴らしい社会に作り上げる研究者たちの努力によって支えられています。

6.2　エクセルで何ができる？

よく知られた表計算ソフトの代表としては Microsoft 社 エクセル EXCEL があります。この講義でもこれを用いてみましょう。



エクセルを起動して Help 命令で「エクセルとは」という説明には”Excel は Microsoft Office system に含まれる表計算プログラムです。Excel では、データを分析し、十分な情報に基づいて意思決定を行うために、ブック (スプレッドシートのコレクション) を作成して書式を設定できます。特に、Excel を使用すると、データのトラッキング、データを分析するためのモデルの構築、そのデータについて計算を実行するための数式の作成、さまざまな方法でのデータのピボット、本格的な外観のグラフを使用したデータの表示を実行できます。”と書かれています。

主な機能として、取り上げている項目には

● テンプレートを検索して適用する ● 新しいブックを作成する ● ブックを保存する ● ワークシートにデータを入力する ● 数値の表示形式を設定する ● セルの罫線を適用する ● Excel のテーブルを作成する ● セルの網かけを適用する ● データをフィルターにかける ● データを並べ替える ● 数式を作成する ● データをグラフにする ● ワークシートを印刷する ● アドインをアクティブにして使用する

があります。さらに統計のいろいろな分析を手っ取り早く行うには

ファイル > オプション > アドイン > と進んで、**分析ツール** を起動項目に追加すると、最初のリボンメニューに**分析>データ分析** が表示されます。

データ分析ウィザード（選択メニュー）には、基本統計量 (データセルをアクティブにして、統計量にチェックすれば、平均、標準誤差、中央値、最頻値、標準偏差、分散など) や、相関、共分散、ヒストグラム、回帰分析などを実行できます。これを使わなくても、デフォルトのままで同様のことをデータ処理できます。ただし、一般に使われる統計の用語とは異なった、“くせ”のある単語が用いられてもいますから、注意が必要です。

また Vidual Basic for Application(VBA) という専門的なプログラムを起動、実行するには、**ツール（T）→マクロ（M）→ Visual Basic Editor(V)** をクリックします。これはプログラミング言語ですが、コンパイル不要で、エクセルでのマクロ（処理を記録し、繰り返しをおこなう）機能をより発展させることができます。メニューの新たに「開発」が追加され、コードリボンに「ＶＢＡ，マクロ」、アドインリボンに「アドイン」が表示されます。その昔、パソコンの黎明期によくもちいたＢａｓｉｃ言語の発展形です。

組み込み関数の例 : 二項分布の密度関数は=BINORMDIST(A2,A3,A4,FALSE) で、3 個のセルには、成功回数セル A2 = 6, 繰り返し回数 A3 = 10, 成功確率 A4 = 0.5, 確率密度 FALSE(0)=密度関数、TRUE(1)=分布関数 (累積密度) に対する値の計算は、0.205078 と結果が得られますが、これは二項係数の式 COMBIN(n,x)=$\binom{n}{x}$ と確率の計算 $p^x(1-p)^{n-x}$ から $\binom{10}{6}(0.5)^6(1-0.5)^{10-6} = \frac{105}{512} = 0.205078$ を計算。

配列数式というセルの組（かたまり）に対する組み込み関数では、統計学では、データ分析のヒストグラム集計、数学のベクトル、行列の計算ができます。たとえば、行列 $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ に対して、行列入力は 2 × 2 セルをアクティブにして、={1,2;3,4}（ただし 2 と 3 の間にはセミコロン）をアクティブセルの一つに入れて、Ctrl+Shift+Enter とします。単に Enter だけではありません。同様に 2 次行列 B を入力して、行列の演算 $A+B$、$A-B$、$A*B$ の結果を書くべきセルをアクティブにして、配列数式の入力 Ctrl+Shift+Enter を行います。積 (multiply, 何倍かにする) は MMULTI, 逆行列 (inverse、逆の、あべこべ) は MINVERSE でもできます。しかし数値のみに限られ、あまり大きな行列では不適です。

この他の組み込み関数では、乱数生成に=RAND()（引数なし）からコインやさいころ投げをシミュレートできます。その集計の度数分布表やヒストグラムの作成には=FREQUENCY(データ配列, 区間配列) など多くの種類があり、有効に使えばたいへん便利です。

実際パソコンを使いながら、実習をしていきましょう。