

Excel[®]2010 を用いた統計学実習

林 譲

まえがき

現代は、コンピュータによる情報の処理と伝達がますます高速になり、人間による情報の処理も迅速性を要求されています。それに伴い、統計学を現場で使う人は計算を迅速に行える技量を持っている必要があります。Excel[®]は多くの統計計算を提供していますので、統計学を現場で実践するスキルを獲得するには最良のソフトウェアの一つでしょう。本書は Excel[®]を使って学習を進めますが、多くの統計的な計算法を解説するのではなく、 t -分布と χ^2 乗分布だけに焦点を当て、これらを理解し現場で使うためのコンピュータ実習を提供します。読者は高等学校の数学の基礎的な知識と、Excel[®]に関する初歩的な知識を持っていることを前提としています。

本書におけるコンピュータ実習はモンテカルロシミュレーションです。これは、同じ条件でのくり返し実験（たとえば、さいころを何度も振るなど）を仮想体験する方法です。実験から平均や標準偏差を推定するとき、可能な限り多くのくり返し実験を行う理由は、最も信頼できる推定値を得るためです。しかし、現実にくり返せる実験の数は限られていますから、実験から推定する値が真の値と一致することはありません。そこで、統計学の検定は、信頼区間という概念で、この不一致の度合いに対処しています。本書の目標は、多数のくり返し実験が可能なモンテカルロシミュレーション実習を通して、この推定値のあいまいさを体験し、信頼区間の概念と検定方法の基礎を理解することです。

本書で扱う話題は平均値と標準偏差の検定だけです。この守備範囲の狭さは、現代の動向には反しているように思えるかもしれません。しかし、著者は、知識を効率よく利用するために最も重要なことは、知識を使う技術ではなく、知識の理解にあると信じています。

コンピュータを使って統計学を勉強するのは能率が良いと認めたとしても、コンピュータの使い方を学ぶのを躊躇する人は多いと思います。筆者もそのうちの一人です。理由は幾つかありますが、最も重大なことは、ノウハウ本における専門用語の解説にあります。残念ながら、対象となる読者の知識水準の設定が不定であり、用語の解説が不十分なため、同業者だけに通じる職業用語の氾濫である本が多いとの印象があります。コンピュータ用語は日々進化していますので、スラングとして廃れる用語が多いという事実を考慮しても、用語解説の不十分さは許容しがたいものがあります。本書は科学の本ですから、コンピュータ用語も科学用語も同様に、なるべく明確に記述するように努めました。

数学や統計学を勉強するとき、体を静止させて頭だけを働かせるのは効率の良い勉強法とは限りません。「習うより慣れよ」という諺の通り、目、耳などあらゆる感覚器官を通しての体験が理解の助けになることもあります。しかし、統計学は、体験しにくい学問です。なぜならば、無限大の状況を想像する必要がある場合が多いからです。

統計学の体験的な勉強法の例として、シュハートのノーマルチップスを紹介します[1]。横 3cm、縦 1.5cm のボール紙（チップ）を 998 枚用意し、1~60 までの整数をそれぞれのチップに書きます。平均が 30、標準偏差が 10 の正規分布となるように整数を書くとすると、整数 30 と書くチップの数はもっとも多く 40 枚であり、30 から離れるに従ってチップの数

は減少します。このように書いた 998 枚のチップを箱の中に入れ、良くかき混ぜてから、ある決まった数のチップを取り出せば、平均 30, 標準偏差 10 の正規分布の母集団からランダムに試料を抜き出した時の実験結果を模倣できます。この方法によって、平均と標準偏差の推定値のばらつき、統計的検定の曖昧さなどを体験できます。

文献[1]によれば、この 998 枚のチップを作るのに、二人がかりで約 4 時間かかったそうです。これと同じことを、本書では、Excel[®]の乱数発生機能を使って行います。人にもよりますが、数分で準備ができます。このように、本書はコンピュータの利点を生かした勉強法を提供します。たとえば、現実には決して得られない分布 (t-分布と χ^2 乗分布) を自分で作ることもできますので、理解の助けになると思います。

筆者のような門外漢の著作により、読者が統計学の基礎を理解し現場で使える知識として構築できれば、望外の喜びです。

林 譲

平成 23 年 11 月 3 日

千葉県佐倉市

本書の読み方

本書は、統計学の基本的な解説と Excel[®]の使い方を示してあります。後者は各省の付録に示してありますので、Excel[®]に精通している読者は、付録の項を省略すれば、効率よく本書を読み進められます。

脚注には、薬学、化学、統計学、確率論などの専門用語と、Excel[®]の使用法に関する簡単な説明を載せてあります。しかし、基本的な Excel[®]の操作で、かつ科学でよく使うものは本文中で説明してあります。

Excel[®]のセルにタイプ入力する文字・数字は「文字・数字」と示します。リボンにあるタブ、グループ、コマンドの名称は[名称]と示します。また、ボタン、ダイアログボックス名など、Excel[®]のコンテンツとしての表示も[表示]として表します。

著作権

本書が含む文書、画像などの権利は、著者（林譲）および／もしくはこれらの使用を許可しているベンダーが保有しています。従って、本書の内容全体または内容の一部をコピー、改変、配布、表示、ライセンスおよび販売することはできません。

本書に含まれる Excel[®]の内容は、Microsoft Corporation のガイドラインに従って掲載しています。

目次

第 1 章 一様分布の統計量の推定	6
1.1 一様乱数の発生.....	6
1.2 データのグラフ表示.....	9
1.3 平均と標準偏差の計算.....	9
1.4 度数の計算とヒストグラムの作成.....	10
1.5 さいころ実験の体験.....	11
1.6 平均の推定値のばらつき.....	12
1.7 標準偏差の推定値のばらつき.....	14
1.8 真の値と推定値.....	15
1.9 付録.....	15
グラフ体裁の変更 (1).....	15
グラフ体裁の変更 (2).....	17
配列数式.....	19
第 2 章 正規乱数と正規分布	22
2.1 正規乱数の発生と表示.....	22
2.2 データの独立性.....	25
2.3 正規乱数のヒストグラム.....	25
2.4 正規分布の標準偏差と確率.....	27
2.5 正規分布の標準偏差 σ の性質.....	29
2.6 付録.....	30
分析ツールの読み込み.....	30
一様乱数と正規乱数のツールの違い.....	31
グラフ体裁の変更 (1).....	31
グラフ体裁の変更 (2).....	31
参照する配列を固定する方法 (F4 の利用).....	32
第 3 章 平均の区間推定	33
3.1 正規分布における平均値の推定.....	33
3.2 t-分布における平均値の推定.....	34
3.3 t-分布のヒストグラム.....	35
3.4 t-分布の確率密度.....	37
3.5 t-分布の性質.....	39
3.6 平均値の区間推定.....	40
3.7 信頼区間と宝くじ.....	42
3.8 系統誤差と有意差検定.....	43

3.9 帰無仮説と有意差検定	44
3.10 対になったデータの t 検定	45
3.11 付録	46
信頼区間のグラフ表示	46
第 4 章 標準偏差の区間推定	48
4.1 χ^2 分布のヒストグラム	48
4.2 χ^2 分布の確率密度	50
4.3 分散の区間推定	50
4.4 信頼区間のくり返し数 n に対する依存性	51
4.5 くり返し数と統計量の信頼性	53
参考文献	54