データ分析（基礎）（オンライン対応版）[*1]

石田 淳
aishida@kwansei.ac.jp

目次

第 1 章

オンライン対応版イントロダクション

1.1　パンデミック対応の講義ノート

2020 年度秋学期は，COVID-19 感染防止のため，講義科目は基本的には対面の授業を行わずオンライン授業への移行となる．

本講義ノートは，オンライン対応版として作成したものである．さらに，各章に対応する講義スライドの PDF，R での分析のための補足資料も同時に公開する．

受講者は，毎回対応する講義ノート部分を読んで課題に応えていくことが求められる．ただし，数学的に詳細に記述している部分については，最初からすべてを完全に理解する必要はない．一番最初は，分からないところはいさぎよく飛ばして全体の流れを把握することにつとめよう．さらに，勉強意欲のある人は一つ一つの式展開を紙と鉛筆でフォローしてみたり，R での分析を自分で試してみたりしよう．

授業内容についての質疑応答は，LUNA 掲示板，メール，Zoom による web ミーティングで対応する予定である．

1.2　シラバス

1.2.1　授業目的

本講義では，社会調査で得られたサンプリング・データをまとめたり分析したりするために必要な，基礎的な推測統計学の知識・技能の修得を目的とする．そのために，各種の概念や技法を正確に理解するとともに，R を用いた具体的な分析過程を示すことで，受講生それぞれが実践的な分析を体験できる機会を設ける．

1.2.2　到達目標

サンプリング・データを自ら分析したり，推測統計学の手法を用いた計量社会学の論文を読むための基本的知識を体得する．

1.2.3　授業時間外の学習（準備学習等について）

データ分析（入門）をあらかじめ受講するか，同等の知識があることが望ましい．数学的知識としては，簡単な確率論の知識を必要とするが，講義内で基礎から説明するために，事前の数学的知識はデータ分析（入門）レベルで問題ない．事前に授業ファイルをアップするので，しっかりと予習・復習するとさらに効果的な学習となる．

1.2.4　授業計画【必ず確認すること】

	内容	配信日	課題〆切日
(1)	1 章 イントロダクション	9/24	10/7
(2)	2 章 記述統計学のおさらい	10/1	10/7
(3)	3 章 推測統計学の考え方	10/8	10/14
(4)	4 章 確率論の基礎	10/15	10/21
(5)	5 章 確率変数と確率分布	10/22	10/28
(6)	6 章 大数の法則と中心極限定理	10/29	11/4
(7)	7 章 標本と標本分布	11/5	11/11
(8)	8 章 統計的推定 (1)：点推定	11/12	11/18
(9)	9 章 統計的推定 (2)：区間推定	11/19	11/25
(10)	10 章 仮説検定 (1)：検定の考え方	11/26	12/2
(11)	11 章 仮説検定 (2)：母平均・母比率の検定	12/3	12/9
(12)	12 章 仮説検定 (3)：母平均の差・母比率の差の検定	12/10	12/16
(13)	13 章 回帰分析における推定と検定	12/17	12/23
(14)	14 章 まとめと補題	1/7	1/13

1.2.5 授業方法

- 本講義ノートと参考資料によるオンデマンド型オンライン講義として実施する．
- 各章の内容はすべて事前に OneDrive にアップロードしておく．授業開始前から読み進めて課題をこなしていってよい．配信日に対応したお知らせメールを LUNA 一斉メールで送る．
- 講義ノートに対応したパワーポイントファイルも参考資料として同時配信する．ただし，パワーポイントファイルはあくまで参考であって，講義ノートと微妙に対応していないところがあるかもしれない．あくまで本講義ノートの内容が正式な講義内容である．
- 統計的手法を扱うので授業では数学を用いる．ただし基本的には，中学高校レベルの基本的な四則演算と不等式の知識があればフォローできるものと期待する．一部で微分積分の知識が必要となるが，最低限の前提知識はその都度導入する．
- 定理の証明は基本的に読み飛ばしてよい．興味ある受講者は自分でフォローしてみよう．
- また，数学的にやや高度な箇所はすべて巻末の付録にまとめているので，興味ある人は確認してみよう．
- 実際の計算過程を体験してもらうために `R` での計算例を参考資料において紹介するので，`R` (`RStudio`) をインストールした学習環境の構築が望ましい．
- 各回ごとに対応した「課題」もしくは「小テスト」を LUNA 上で課す．各章の最後に課題・テストの指示がある．
- 課題や小テストでは，計算やグラフの作成を求めることがある．手計算・手書きでもできるが，`R` や `Excel` を用いることを推奨する．なお，受講生によってランダムに問題が変わることがある．
- 課題・小テスト〆切は配信日の一週間後に設定されているが，基本的には配信日に課題を完成させ提出することを推奨する．
- 課題ファイルはコピペチェッカーにかける．異なる受講者から同一ファイルが提出された場合は，どのような事情であろうと提出した受講者全員を不可と

する．

- 成績評価方法：毎回の授業に対応した小テストもしくは小課題の累積により評価する．
- 質疑応答は LUNA 掲示板もしくはメールにて (aishida@kwansei.ac.jp)．ただし，掲示板は毎日チェックするわけではないので，返答が遅れる可能性がある．
- Zoom を用いた参加任意の質問セッションを授業時間（木 1）に実施する．質問がある人もない人も自由に参加してよい．

1.3　参考文献

■入門　初歩の初歩から推測統計学のロジックをざっくりと理解したい人は，

小島寛之，2006，『完全独習　統計学入門』ダイヤモンド社

がよい．「使うのは中学数学だけ！」と銘打たれている．さらに，記述統計レベルからきちんと復習したい場合は，初学者向けのテキストとして

神林博史・三輪哲，2011，『社会調査のための統計学』技術評論社

がある．中高レベルの統計の復習として以下も参考になる．

結城浩，2016『数学ガールの秘密ノート　やさしい統計学』SB クリエイティブ

■本格入門　もう少しきちんと勉強する場合は，とっかかりとして

森棟公夫ほか，2008，『統計学』有斐閣
倉田博史・星野崇宏，2009，『入門統計解析』新世社

などがよいだろう．

■確率統計　さらに，数学的に確率統計を勉強する場合は

薩摩順吉，1989，『確率・統計』岩波書店
蓑谷千凰彦，1994，『統計学入門』東京図書

小針あき宏，1973，『確率・統計入門』岩波書店
渡辺澄夫・村田昇，2005，『確率と統計—情報学への架橋』コロナ社

などがある．薩摩（1989）は基本的なことがまんべんなく載っており，蓑谷（1994）は推定・検定に詳しい．小針（1973）は数学的展開が詳しい．渡辺・村田（2005）は，情報量の概念の説明が丁寧になされている．本講義の数学パートで用いられる数学は，初歩的集合論，初歩的微分・積分程度である．復習が必要であれば，

矢野健太郎・田代嘉宏，1993，『社会科学者のための基礎数学（改訂版）』裳華房
永野裕之 (著)・岡田謙介 (監修)，2016，『この 1 冊で腑に落ちる　統計学のための数学教室』ダイヤモンド社

などがよいだろう．

■多変量解析・ベイズ統計　重回帰分析を含めた多変量解析の教科書としては，

永田靖・棟近雅彦，2001，『多変量解析法入門』サイエンス社

が数学的に丁寧でわかりやすい．さらに，ベイズ統計モデリングを本格的に勉強する場合は以下も参照のこと．

浜田宏・石田淳・清水裕士，2019，『社会科学のためのベイズ統計モデリング』朝倉書店

■読み物　推測統計学の歴史的展開については，以下が参考になる．

蓑谷千凰彦，1997，『推測統計のはなし』東京図書
デイヴィット・サルツブルグ，2006，『統計学を拓いた異才たち』日本経済新聞社

1.4　R と RStudio の導入

R は，データ分析のためのオープンソース・フリーソフトウェアのプログラミング言語環境である（`https://www.r-project.org/`）．R 単体での利用も可能であるが，近年は R のための統合開発環境（IDE）である RStudio からの利用が一般的になっている（`https://rstudio.com/`）．

とりあえず, R や RStudio を触ってみたい場合には, RStudio と同じ機能を web 上で利用できるサービスである RStudio Cloud の利用を勧める（`https://rstudio.cloud/`）．ただし現時点で開発途上のベータ版であることに留意．アカウントを作成して（あるいは Google か GitHub アカウントで）ログインすると使える．ファイルをインポートしたりエクスポートすることも，パッケージもインストールすることもできる．

がっつり使いたい場合は, 自分のマシンに R と RStudio をインストールする．基本的に, OS (Windows, macOS, Linux) の違いに応じてそれぞれのインストラクションに沿ってインストールしていけば問題ないだろう．Windows については, ユーザー名を全角文字で設定している場合, R や RStudio がうまくインストールできないという問題がある．この問題への対処を含めて，現時点で日本語で一番まとまっている情報は, 矢内勇生氏のウェブサイト（`https://yukiyanai.github.io/jp/resources/`）である．

1.5　授業を始めるにあたって

これから推測統計学の基礎を学んでいくにあたって，いくつかの注意点を述べておく．

まず，本講義のゴールをどこに設定するかであるが，とりあえず「母平均の推定・検定」のロジックをしっかり理解することを目標としたい．これは実は，統計分析の実践という点からいうと入り口の入り口である．本講義では，安易に実践に向けた「マニュアル」を教えるのではなく，その基礎となるロジックの理解に注力したいと考えている．実践におけるマニュアルなんてものは実践においてその都度確認すればよろしい．

そのため，オンデマンド型オンライン講義の講義形態としては，講師のしょーもないべしゃり動画などでお茶を濁す軟派な形態ではなく，講義ノートをメインとする硬派な形態にした．詳細に書き下されたロジックを一つ一つ自分のペースでたどっていくことが，応用数学である統計学の学習には何よりも効果的であると信じるからである．リアルなロジックを深いレベルで理解するのは，結局のところ独習でしか成し得ないのである．

とはいえ，はっきりいって推測統計学のロジックは難しい．なので，初見で「わけわかめ」でも「そういうもの」だと思ってほしい．だって難しいんだもの．あと，こちらの説明が足りていないところもあると思う．そういうところは，何度か考えてみてそれでも分からなければぜひ質問に来て欲しい．遠慮は不要である．

推測統計学は，不確実性の中で限られた情報からより一般的な結論を導くために構築されたロジックの体系である．多くの人が統計を使うプロにはならない中で，大学において統計を学ぶ意味は何かというと（他の講義にも言えることだが），このようなロジックをより原理的なレベルで理解し，考え方の幅を身につけ，さまざまな問題に対応するための応用可能性を高めること，そのために自らの認知的強度を高める訓練をすることであると思われる．ボタンを押して結果を食べるくらいならハトでもできる．われわれは主体的な人間であるために，統計をそして数学を学ぶのである．知らんけど．

ということで，扉はすでに開かれている．Welcome to hell !!

課題 1.1. 本講義を受けるにあたっての意気込みを熱く語りなさい．*Word* ファイルでリポートを作成し，*10/7*（水）*23:59* までに *LUNA* 上で提出せよ．

第 2 章

記述統計学のおさらい

2.1　記述統計学とは

記述統計学 (descriptive statistics) とは，調査や観察，実験によって得られたデータを整理し，特性を記述するための方法論である．ここで関心の対象は**データの対象である集団そのものの特性**であり，それ以上でもそれ以下でもない[*1]．

一方，**推測統計学** (inferential statistics) とは記述統計学の知識の上に，確率論を駆使することによって作り上げられた方法論であって，関心の対象とする集団（**母集団**）の一部をランダムに取り出してできる部分集団（**標本**）の特性を調べることで，母集団の特性を推測する技法である．

われわれの最終目的は，推測統計学の基本ロジックを会得することである．そこで，まずは記述統計学をざっくりと確認して，そのあと確率論を勉強する．そしてその知識をもとに標本から母集団の特性を推測する方法を学習する．

データには大きく分けて，意味のある定量的な値で与えられる**量的データ**と，あるカテゴリーに属しているとか属していないといった情報をもつ**質的データ（カテゴリカル・データ）**がある．量的データは，例えば，身長，体重，年齢，所得，テストの点

[*1] 本講義ノートでは，重要な概念は太字で記し英語での表記を付け加えている．これには 2 つの重要な意味がある．まず，数学や統計学は言語集団に限定されない普遍的な科学言語であり，世界的に流通している．そして基本的に，どのような教科書でも同じ知識が教授されていることが期待される（もちろん，著者によって説明の仕方も違うし専門家の間で論争になっているような最先端の話題もある）．日本語でも，相当のことが学べるが，英語の世界の方が（科学に関していえば）圧倒的に広い．なので，今後英語の教科書で勉強してもよい．そのときのためというのが一点．もう一点は，統計を実践する際の R などのプログラミング言語の命令は，英語準拠で定められているため，対応する英単語を知っておくといろいろとはかどる．

数などである．質的データは，性別，学歴，ある意見への賛否などである．質的データもダミー変数化という工夫によって数量化することができる．本講義では基本的に，量的データを前提にして議論を進めていく．

2.2　平均，分散，標準偏差

2.2.1　定義

データの代表値としてもっともよく用いられる**平均** (mean)，ばらつきの指標である**分散** (variance)，そして分散の単位をもとの観測値の単位に戻した**標準偏差** (standard deviation) を導入する．n 個の観測値 $x_1, x_2, \cdots, x_n$ について，平均，分散，標準偏差はそれぞれ次のように定義される．

■平均

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{2.1}$$

■分散

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} \tag{2.2}$$

■標準偏差

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.3}$$

ここで導入した平均は，他にもある平均の定義と区別して，相加平均，算術平均ともいわれる．

分散の定義について注意しておく．他の教科書では分散もしくは標本分散として，n で割るのではなく，$n-1$ で割る定義を採用しているものがある．本講義でも，標本理論を導入した後でこの $n-1$ で割る分散を，**不偏分散** (unbiased variance) S^2 として導入することになる．そのわけは導入の際に詳しく説明することになる．いまのところは記述統計の範囲で分散といえば n で割る $\hat{\sigma}^2$ と考えておけばよい[*2]．

[*2] ただし，R の分散関数 `var()`，標準偏差関数 `sd()` は $n-1$ で割る定義を採用している．

また，分散と標準偏差の表記について，とくに変数 x の分散・標準偏差であることを強調する必要があるときには，$\hat{\sigma}_x^2, \hat{\sigma}_x$ と記すことにする．

■例（身長データ）

データ: 152.8, 150.1, 182.0, 163.2, 167.3, 160.2, 164.9, 161.4, 179.9, 172.2

$$
\begin{aligned}
\bar{x} =& \frac{1}{10}(152.8 + 150.1 + 182.0 + 163.2 + 167.3 + 160.2 + 164.9 + 161.4 + 179.9 \\
&+ 172.2) \\
=& 165.4
\end{aligned}
$$

$$
\begin{aligned}
\hat{\sigma}^2 =& \frac{1}{10}\{(152.8 - 165.4)^2 + (150.1 - 165.4)^2 + (182.0 - 165.4)^2 + (163.2 - 165.4)^2 \\
&+ (167.3 - 165.4)^2 + (160.2 - 165.4)^2 + (164.9 - 165.4)^2 + (161.4 - 165.4)^2 \\
&+ (179.9 - 165.4)^2 + (172.2 - 165.4)^2\} \\
=& \frac{1}{10}(158.76 + 234.09 + 275.56 + 4.84 + 3.61 + 27.04 + 0.25 + 16 + 210.25 + 46.24) \\
=& 97.66
\end{aligned}
$$

$$
\hat{\sigma} = \sqrt{97.66} = 9.88
$$

2.2.2　度数分布表からの計算

ある観測値の**度数** (frequency) をカウントし，表にしたものを**度数分布表** (frequency table) といい，度数の多さを面積で対応させたグラフで表したものを**ヒストグラム** (histogram) という[*3]．k 個の種類の観測値が観測されたとき，おのおのの観測値を v_i $(i = 1, \cdots, k)$，その観測値の度数を f_i $(i = 1, \cdots, k)$ とする．このとき $\sum_{i=1}^{k} f_i = n$ である．そして**相対度数** (relative frequency)，つまり観測度数のデータ総数 n に占める割合を $\hat{p}_i$ $(i = 1, \cdots, k)$ と表す．

[*3] 観測値が連続変量の場合，階級ごとに区切って度数分布表・ヒストグラムを作成することになるが，ここではその説明はしない．

■平均

$$\bar{x} = \sum_{i=1}^{k} \frac{f_i}{n} v_i = \sum_{i=1}^{k} \hat{p}_i v_i \tag{2.4}$$

■分散

$$\hat{\sigma}^2 = \sum_{i=1}^{k} \frac{f_i}{n} (v_i - \bar{x})^2 = \sum_{i=1}^{k} \hat{p}_i (v_i - \bar{x})^2 \tag{2.5}$$

■標準偏差

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\sum_{i=1}^{k} \hat{p}_i (v_i - \bar{x})^2} \tag{2.6}$$

■例（とある講義に出席する学生の年齢）　度数分布表は表 2.1，ヒストグラムは図 2.1 である．

$$\begin{aligned}\bar{x} =& 0.215 \times 18 + 0.275 \times 19 + 0.340 \times 20 + 0.105 \times 21 + 0.045 \times 22 + 0.020 \times 23 \\ =& 19.55\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2 =& 0.215(18 - 19.55)^2 + 0.275(19 - 19.55)^2 + 0.340(20 - 19.55)^2 \\ &+ 0.105(21 - 19.55)^2 + 0.045(22 - 19.55)^2 + 0.020(23 - 19.55)^2 \\ =& 1.398\end{aligned}$$

$$\hat{\sigma} = \sqrt{1.398} = 1.182$$

また，度数分布表から計算する方法を応用することで，ダミー変数化した 2 値カテゴリカルデータの平均，分散，標準偏差を計算することができる．2 つのカテゴリーが存在して，それぞれのカテゴリーの度数がカウントされているとき，カテゴリー 1 を “1”，カテゴリー 2 を “0” と数値化する．カテゴリー i の度数と相対度数をそれぞれ $f_i, \hat{p}_i$ $(i = 1, 2)$ とする．

■平均

$$\bar{x} = \frac{f_1 \times 1 + f_2 \times 0}{n} = \frac{f_1}{n} = \hat{p}_1 \tag{2.7}$$

表 2.1　とある講義に出席する学生の年齢

観測値 v	度数 f	相対度数 $\hat{p}$	累積相対度数
18	43	0.215	0.215
19	55	0.275	0.490
20	68	0.340	0.830
21	21	0.105	0.935
22	9	0.045	0.980
23	4	0.020	1.000
合計	200	1.000	

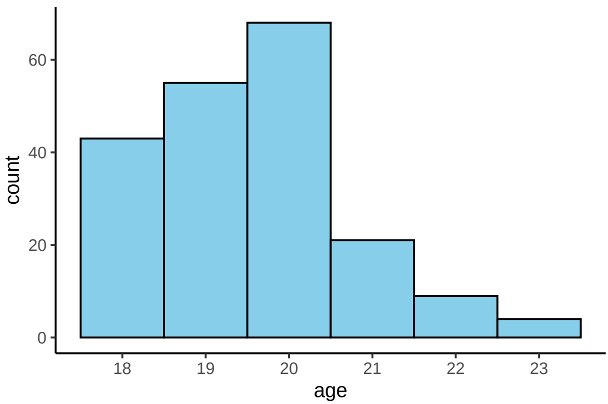

図 2.1　学生の年齢ヒストグラム

■分散

$$\hat{\sigma}^2 = \frac{f_1(1-\hat{p}_1)^2 + f_2(0-\hat{p}_1)^2}{n} = \hat{p}_1\hat{p}_2^2 + \hat{p}_2\hat{p}_1^2 = \hat{p}_1\hat{p}_2 \tag{2.8}$$

■標準偏差

$$\hat{\sigma} = \sqrt{\hat{p}_1\hat{p}_2} \tag{2.9}$$

つまり，平均はカテゴリー 1 の全体に占める割合である．

■例（とある講義に出席する学生の性別）　男性をカテゴリー 1，女性をカテゴリー 2 とする．男性が 64 人，女性が 36 であった．

$$\bar{x} = \frac{64}{100} = 0.64$$

$$\hat{\sigma}^2 = 0.64 \times 0.36 = 0.230$$

$$\hat{\sigma} = \sqrt{0.230} = 0.480$$

2.3　標準得点

データ $x_1, x_2, \cdots, x_n$ をそれぞれ a 倍して，b を足すことを考える．つまり，

$$z_i = ax_i + b, \qquad i = 1, \cdots, n \tag{2.10}$$

このようなデータ変換を**一次変換** (linear transformation) という．このとき，変換後の平均・分散・標準偏差は以下のようになる．

■平均

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = a\bar{x} + b \tag{2.11}$$

■分散

$$\hat{\sigma}_z^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{n}a^2(x_i - \bar{x})^2 = a^2\hat{\sigma}_x^2 \tag{2.12}$$

■標準偏差

$$\hat{\sigma}_z = \sqrt{a^2\hat{\sigma}_x^2} = |a|\hat{\sigma}_x \tag{2.13}$$

一次変換について，とくに

$$z_i = \frac{x_i - \bar{x}}{\hat{\sigma}_x} \tag{2.14}$$

を採用した場合，つまり，

$$a = \frac{1}{\hat{\sigma}_x}, \qquad b = -\frac{\bar{x}}{\hat{\sigma}_x}$$

とすると，$\bar{z} = 0$，$\hat{\sigma}_z^2 = 1$，$\hat{\sigma}_z = 1$ となる．このような変換を**標準化** (standardization) といい，z_i を**標準得点** (standard score) という．標準得点は z 得点ともいわれる．

ちなみに，受験の時におなじみのいわゆる「偏差値得点」は，標準得点をさらに $10z_i + 50$ と変換したものであり，平均は 50，標準偏差は 10 となるようにしたものである．

2.4　共分散と相関係数

2 変数のデータを考える．例えば，それぞれの個人に身長と体重を尋ねたとすると，それぞれの個人について観測値ベクトル (身長$_i$, 体重$_i$) が求まる．これを対象者分集めたものが 2 変数データである．

2 つの変数の結びつきを調べる上でもっとも基本となるのが，**共分散** (covariance) である．変数 x と y の共分散を c_{xy} で表すと，

$$\begin{aligned} c_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad (2.15) \\ &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{n} \end{aligned}$$

と定義される．1 変数の分散は非負の値しか取らないが，共分散はマイナスにもなることに注意．2 つの変数の関連が強いほど共分散はプラスもしくはマイナス方向に大きくなり，関連がなけれなないほど 0 に近づく．

x と y の標準得点について共分散をとったものを**相関係数** (correlation coefficient) という．

$$z_i = \frac{x_i - \bar{x}}{\hat{\sigma}_x}, \qquad w_i = \frac{y_i - \bar{y}}{\hat{\sigma}_y}$$

とすると，相関係数 r_{xy} は

$$r_{xy} = \frac{1}{n}\sum_{i=1}^{n} z_i w_i \tag{2.16}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\hat{\sigma}_x}\right)\left(\frac{y_i-\bar{y}}{\hat{\sigma}_y}\right) = \frac{c_{xy}}{\hat{\sigma}_x\hat{\sigma}_y} \tag{2.17}$$

と定義される．

相関係数は -1 から 1 の間の値をとり，$|r_{xy}| = 1$ のときそのときに限り，$y_i = ax_i+b$ という線形の関係がすべての i について成り立っている．そして，$|r_{xy}|$ が 1 に近づけば近づくほど，線形の関係に近づく．

■特性の証明

定理 2.1.

$$-1 \le r_{xy} \le 1$$

証明. 常に非負となる式 $(1/n)\sum_{i=1}^{n}(z_i \pm w_i)^2$ を展開する．

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(z_i \pm w_i)^2 &= \frac{1}{n}\sum_{i=1}^{n}(z_i^2 \pm 2z_iw_i + w_i^2)\\ &= \frac{1}{n}\sum_{i=1}^{n} z_i^2 \pm \frac{2}{n}\sum_{i=1}^{n} z_iw_i + \frac{1}{n}\sum_{i=1}^{n} w_i^2\\ &= 1 \pm 2r_{xy} + 1 = 2(1 \pm r_{xy}) \ge 0\end{aligned}$$

$\pm r_{xy} \ge -1$ より，$-1 \le r_{xy} \le 1$ である．　□

定理 2.2.

$$\forall i,\ y_i = ax_i + b \Longleftrightarrow |r_{xy}| = 1$$

証明. $(\Longrightarrow)$ $\forall i,\ y_i = ax_i + b$ とすると，式 (2.13) より，$\hat{\sigma}_y = |a|\hat{\sigma}_x$，また

$$\begin{aligned}c_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\\ &= \frac{a}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x}) = a\hat{\sigma}_x^2\end{aligned}$$

である．ゆえに，

$$r_{xy} = \frac{a\hat{\sigma}_x^2}{\hat{\sigma}_x |a| \hat{\sigma}_x} = \frac{a}{|a|}.$$

つまり，$a > 0$ のとき相関係数は 1 で，$a < 0$ のとき -1 となる．

$(\Longleftarrow)$ $r_{xy} = \pm 1$ とする．このとき，

$$c_{xy} = \pm \hat{\sigma}_x \hat{\sigma}_y = \pm \frac{\hat{\sigma}_y}{\hat{\sigma}_x} \hat{\sigma}_x^2$$

が成り立つ．$a = \pm \hat{\sigma}_y / \hat{\sigma}_x$ とすると，上式より

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = a \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})$$

が成り立つ．ゆえに，

$$\forall i, (y_i - \bar{y}) = a(x_i - \bar{x})$$

でなければならない．$b = \bar{y} - a\bar{x}$ とすると $\forall i$, $y_i = ax_i + b$ を得る． □

■例（身長と体重）　x を身長，y を体重とする．データは表 2.2，散布図（観測値ベクトルをプロットした図）は図 2.2 である．

表 2.2　身長と体重データ

個人 i	1	2	3	4	5	6	7	8	9	10
身長 x	152.8	150.1	182	163.2	167.3	160.2	164.9	161.4	179.9	172.2
体重 y	56.3	52.1	85.6	66.8	74.2	58.1	61.9	55.1	70.5	64.1

$$c_{xy} = 81.313, \qquad \hat{\sigma}_x = 9.882, \qquad \hat{\sigma}_y = 9.684$$
$$r_{xy} = \frac{81.313}{9.882 \times 9.684} = 0.850$$

身長と体重の関連はかなり強いと言える．

課題 2.1. *LUNA* のテストを *10/7*（水）*23:59* までに受けなさい．

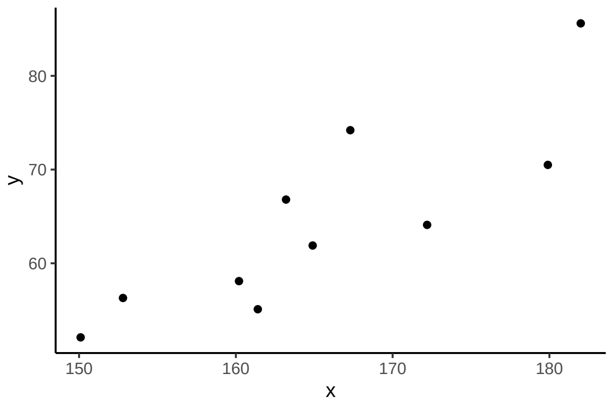

図 2.2　身長と体重の散布図

第 3 章

推測統計学の考え方

本章では，推測統計学の考え方を概説し，次章以降の導入とする．

3.1　データ分析の諸レベル

本学部において設定されているデータ分析科目はおおよそ表 3.1 のようにレベル分けされる．

表 3.1　データ分析の諸レベル

level	タイトル	やること	使う数学	対応する授業
1	記述統計学	データの特性を記述的に捉える	四則演算	入門
2	推測統計学	サンプルデータから母集団の特性を推測する	＋ 確率論	基礎
3	多変量解析	複数の変数からデータならびに母集団の特性を推測する	＋ 線形代数	応用

本講義が扱う「推測統計学」は，データ分析（入門）における記述統計学の知識を土台として，主に確率論を数学的道具として導入することで，サンプルデータから母集団の特性を推測することを目的としている．

3.2　母集団と標本

記述統計学と推測統計学の考え方の違いをさらに明確に理解するために，母集団と標本という用語についてもう少し丁寧に説明する．

母集団 (population) とは，調べたい対象の集団のことである．たとえば，関学の学生の生活状況や受講態度を知りたければ，母集団は「関学の学生全員」となる．西宮市の市民意識を調べたいのであれば，母集団は「西宮市民」，現在の日本の人口動態を知りたければ，母集団は「日本在住者全員」となるだろう．

一方，**標本** (sample) とは，母集団から何らかの方法で取り出された調査対象の集団のことである．母集団が関学生のときの標本の一つの例として「データ分析（基礎）の受講者」がある．あるいは，西宮市民を母集団として，正門前をほっつき歩いている人を 100 人つかまえるという標本もありうる．さらには，日本在住者全員を母集団として，標本として全国から無作為抽出した 2000 人を設定することもある．

図 3.1 は，母集団と標本の関係を示した模式図である．

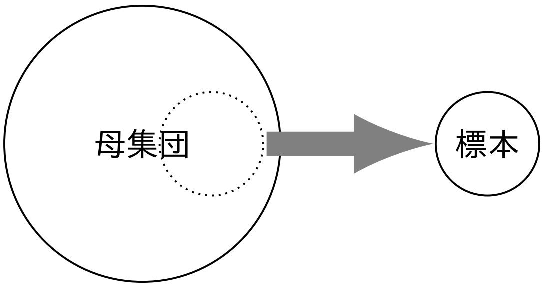

図 3.1　母集団と標本の関係

3.3　標本抽出

次に，関連して調査法についてごく簡単に確認しよう．

母集団全員を対象にした調査を**全数（悉皆）調査** (complete survey) という[*1]．別の言い方をすると，全数調査とは母集団全部を標本とする調査であるといえる．例えば，大学において卒業予定者を対象にした進路調査や，小学 6 年生，中学 3 年生を対象に行われる全国学力・学習状況調査，日本在住者全員が対象の国勢調査が全数調査の代表例である[*2]．

[*1] 「悉皆」は「しっかい」と読む．

[*2] 今年は国勢調査の実施年です．皆さん協力しましょう．というか国勢調査は国の統計法で「基幹統

一方，母集団の一部分である標本を対象にした調査のことを，**標本調査** (sample survey) という．視聴率調査，世論調査，出口調査，品質検査，面接・郵送アンケート調査など，実際に行われる大部分の調査が標本調査であると言ってよいだろう．

では，標本をどのように選び出すのだろうか．母集団から標本を選び出すことを一般的に**標本抽出** (sampling) という．標本抽出法としては，大まかに有意抽出法と無作為抽出法に分けられる．

有意抽出法 (purposive sampling) とは，調査者が恣意的に対象者を決める抽出法である．例えば，知り合いに頼む，街を歩いている人に頼む，あるいはある授業の受講者を対象者とするような抽出法である．このような手法は，頼みやすい人に頼めるという意味ではお手軽であるが，母集団特性を推測する際に抽出法に依存する系統誤差（バイアス）が発生する可能性がある．

一方，**無作為抽出法** (random sampling) とは，簡単に言うと，くじ引きで対象者を選ぶような方法である．より正確に言えば，母集団から選ばれる確率が等しくなるように標本を選ぶ方法である[*3]．無作為抽出法では，系統誤差（バイアス）は制御できるが，確率的試行において偶然に生じる偶然誤差（エラー）は依然として残る．

3.4　偶然誤差込みの推測

系統誤差と偶然誤差の説明には，よく味噌汁のたとえが用いられる．味噌汁を作って最後味噌を溶かしたあと，よくかき混ぜないで味見をすると（有意抽出），味の薄いところか濃いところにあたってしまって，鍋の全体の塩分濃度の判断を誤る危険性がある（系統誤差）．一方，よくかき混ぜて味見をした場合は（無作為抽出），鍋全体の塩加減をだいたいのところつかむことができるだろう．しかし，その場合でも，たまたま味の濃いところ，薄いところを掬ってしまうこともあるだろう（偶然誤差）．

なんとなくわかるような気もするがよくわからないような気もする．個人的に無作為抽出のイメージとしては，味噌汁をかき混ぜるイメージより，東京ドームを振り回すイメージの方が合っているように思うが，このイメージの詳細を書き下すといろい

計調査」とされており，調査を拒否したり虚偽の回答をするものは 50 万円以下の罰金の刑に処される（実際に処された人はいないが）．

[*3] 有意抽出法，無作為抽出法とも細かくは様々な手法やテクニックがあるが，本講義の射程を超えるのでここでは講じない．

ろと差し障りがあるので，機会があれば直接会ったときに教えます（図 3.2）.

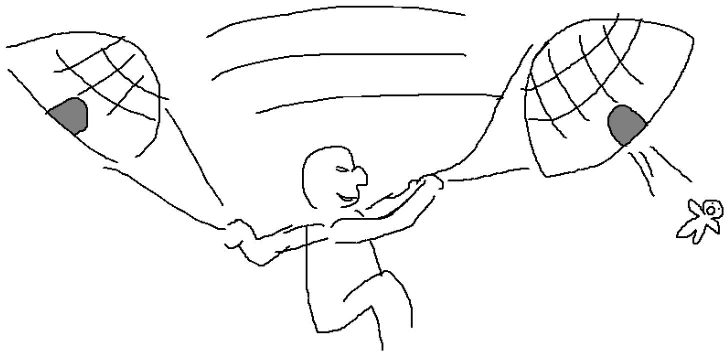

図 3.2　東京ドームを振り回すイメージ

もう一度まとめると，無作為抽出（ランダム・サンプリング）によって系統誤差（バイアス）は制御できる．それでも，偶然誤差（エラー）は残る．つまり，よくよく味噌汁をかき混ぜたけど，たまたま塩分濃度の濃いところをすくってしまうことがある．しかしながら，適切に無作為抽出法の調査設計がなされていれば，確率論と統計学の知識によって，どれくらいの偶然誤差が生じるかを見積もることができる．これこそが大きなコストをかけて無作為抽出調査を行う理由である．

無作為抽出と確率・統計の知識の組み合わせによって，調査結果より，たとえば「±5% くらいの誤差はあるが，母集団における内閣支持率は 60% である」とか，「±5cm くらいの誤差はあるが，母集団における平均身長は 170cm である」というような母集団の特性に関する推測を行うことが可能となる．

さらに，標本の大きさ（サンプルサイズ）[*4]を大きくすることで偶然誤差を減らすことができる（大数の法則）し，どの程度の精度を求めるかによって，どの程度のサ

[*4] 標本（サンプル）は一つの塊であるので，「サンプル数」というと抽出された塊としての標本がいくつかあるという表現になってしまう．そのため，標本のなかにどれくらいの対象者が含まれるかを述べる場合は，「サンプル数」ではなく「サンプルサイズ」と表現しなければならないと主張する流派があり，その警察組織として誤った表現を取り締まる「サンプル数警察」が暗躍しているので注意が必要である．

ンプルサイズが必要かを見積もることもできる．ゴイスーですね．

3.5　記述統計学と推測統計学

記述統計学においては得られたデータの特性のみが関心の対象であり，データを超えた推測は行わない．このとき，母集団と標本は概念上等しいものとなっている（図 3.3）．一方，推測統計学においては，無作為抽出された標本をデータとして，そこからより大きな対象である母集団の特性を確率・統計の知識を用いて推測するのである（図 3.4）．

図 3.3　記述統計学における母集団と標本の関係

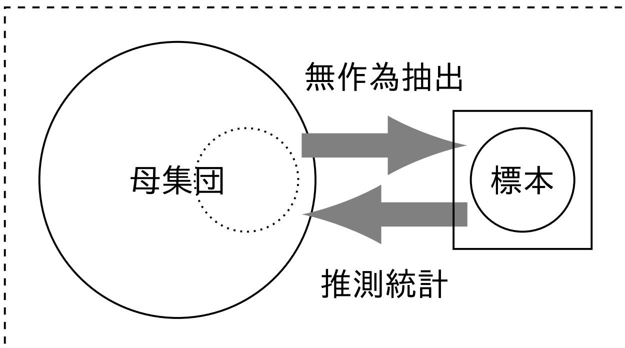

図 3.4　推測統計学における母集団と標本の関係（四角内が認識対象で点線の四角は推測範囲）

課題 3.1. 記述統計学と推測統計学の違いを自分自身の言葉でまとめよ．*Word* ファイルでリポートを作成し，*10/14*（水）*23:59* までに *LUNA* 上で提出せよ．

第 4 章

確率論の基礎

ここから，いよいよ推測統計学の基本的な考え方を学んでいく．そのために必要となる数学的道具が確率論である．つづく，4–6 章において確率論を導入し，確率分布を導入し，独立同分布，大数の法則，中心極限定理という理論的コアが導入される．これらの章が一番抽象度が高いところである．しっかりと息を止めて深くまで潜行していく必要がある．

本章ではまず，確率論の基礎を学ぶ．推測統計学の基本コンセプトは，ランダムに抽出された標本から母集団の特性を推測することにあるが，この際標本抽出のランダムネスによる推測の不確実性がどうしても生じてしまう．不確実性の数学モデルである確率論が必要とされるゆえんである．

4.1　標本空間と事象

4.1.1　標本空間

確率 (probability) とは，ある出来事が起こる「確からしさ」を数量的に定めたものである．確率論の基礎になるのが集合を使った起こりうる出来事の表現である．

コインやサイコロを投げる，あるいは壷の中にいくつか入っている小玉の一つを取り出す，あるいは東京ドームをぶん回して中の人 1 人を取り出すといった操作のことを確率論では**試行** (trial) という．ある試行を行ったときに生じるすべての可能な結果の集合のことを**標本空間** (sample space) あるいは全事象といい，Ω で表す．また，標本空間の任意の要素を**標本点** (sample point) といって ω で表す．定義から $\omega \in \Omega$

である[*1]．結果がある区間の実数をとりうる場合，標本空間はその区間のすべての点の集合になる．図 4.1 は標本空間の模式図である．

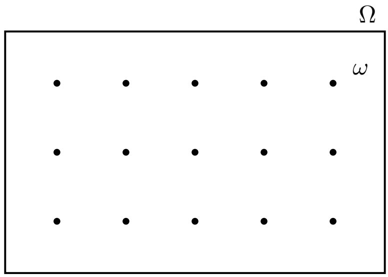

図 4.1　標本空間

■例（コイン投げ試行の標本空間）　コインを投げたときの裏表の標本空間は

$$\Omega = \{\text{ 裏}, \text{表 }\}$$

である．コインを 2 枚同時に投げたときの標本空間は

$$\Omega = \{(\text{裏}, \text{裏}), (\text{裏}, \text{表}), (\text{表}, \text{裏}), (\text{表}, \text{表})\}$$

である．

■例（サイコロ投げ試行の標本空間）　サイコロを投げたときの目の標本空間は

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

である．

4.1.2　事象

標本空間の部分集合のことを**事象** (event) という．事象は確率的に起こりうる出来事を意味する．標本点をただ 1 つ含む事象のことを**根元事象** (elementary event) とよび，2 つ以上の標本点を含む事象のことを複合事象という．標本点を含まない事象

[*1] 一般に，$a \in A$ は要素 a が集合 A に属することを示す．

のことを**空事象** (empty event) といい，$\emptyset$ で表す．ある事象を A で表すと $A \subset \Omega$ である[*2].

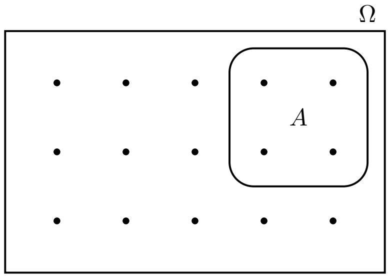

図 4.2　事象の例

■例（コイン投げ試行の事象）　コイン投げ試行において考えうる事象は

$$\{\,裏, 表\,\}, \qquad \{\,裏\,\}, \qquad \{\,表\,\}, \qquad \emptyset$$

の 4 つであり，それぞれ「裏か表が出る」，「裏が出る」，「表が出る」，「裏も表も出ない」である．このうち $\{\,裏\,\}, \{\,表\,\}$ は根元事象であり，$\{\,裏, 表\,\}$ は複合事象，$\emptyset$ は空事象である．

■例（サイコロ投げ試行の事象）　サイコロ投げ試行において考えうる事象は全部で $2^6 = 64$ 通りある．例えば，事象 $\{1, 2, 3\}$ は「3 以下の目が出る」という事象であり，$\{2, 4, 6\}$ は「偶数の目が出る」という事象である．

4.1.3　事象の演算

ここで，サイコロ投げ試行の事象として

$$A = \{1, 2, 3\}, \quad B = \{2, 4, 6\}$$

を例とする．「A と B のうち，少なくとも一方が起こる事象」のことを A と B の**和事象** (union of events) とよび $A \cup B$ で表す．演算としては集合演算の和集合と同じ

[*2] $A \subset B$ は，A は B の部分集合であることを示す．

であり，$A \cup B$ の場合

$$A \cup B = \{1, 2, 3, 4, 6\}$$

となる．「A と B が同時に起こる事象」を A と B の**積事象** (intersection of events) とよび $A \cap B$ で表す．演算としては集合演算の積集合と同じであり，$A \cap B$ の場合

$$A \cap B = \{2\}$$

となる．$A \cap B = \emptyset$ のとき A と B は**互いに排反** (mutually exclusive) であるという．最後に，「A が起こらないという事象」を A の**補事象** (complementary event) とよび，A^c で表す．集合演算における補集合と同じであり，例の場合

$$A^c = \{4, 5, 6\}, \quad B^c = \{1, 3, 5\}$$

である．

それぞれの事象演算を**ベン図** (Venn diagram) で表せば図 4.3–4.5 のようになる．灰色部分が該当する演算の結果の集合を示す．

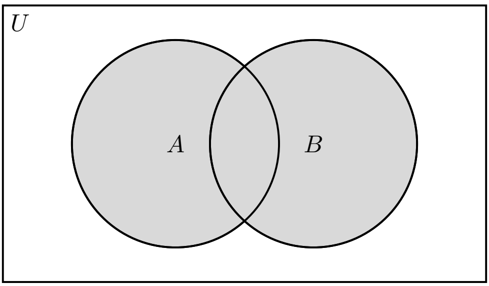

図 4.3　和事象 $A \cup B$ のベン図

さらに，集合演算 $\cup, \cap, (\ )^c$ に関して，付録 A.1 のような基本定理が成り立つ．以下の証明で用いているので，数学的に関心のある人は確認して欲しい．

4.2　確率の定義

確率 (probability) とは，事象の起こりやすさを定量的に示す測度 (measurement) であり，事象 A の確率を $P(A)$ と表す．つまり，確率とは集合を引数とする集合関

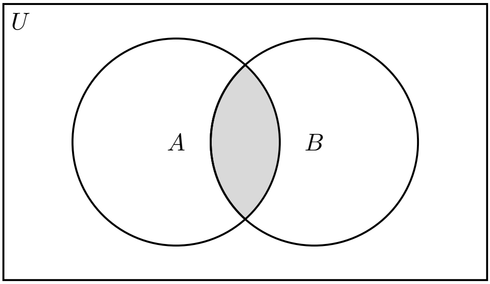

図 4.4　積事象 $A \cap B$ のベン図

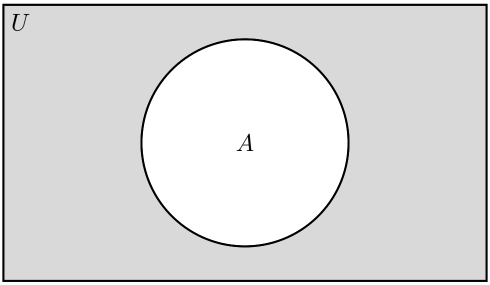

図 4.5　補事象 A^c のベン図

数の一種である．

確率を具体的にどのように定義するかということは，これまで一つの論争点であった．ここでは，ラプラスによる古典的定義，頻度による定義，そしてコルモゴロフによる公理主義的定義を紹介する．本講義では，直感的には古典的定義と頻度による定義，数学的には公理主義的定義に立って議論を行う．

定義 4.1 (**確率の古典的定義**)**.** ある試行について，標本空間の要素数が n であり，それぞれの根元事象は同程度に確からしく起こるとする．ある事象 A についてその要素数が r のとき，A の確率 $P(A)$ を

$$P(A) = \frac{r}{n} \tag{4.1}$$

と定義する．

■例（サイコロ投げの確率）　ゆがみのないサイコロのそれぞれの目（根元事象）は，同程度に確からしく起こると考えてよい．このとき，確率の古典的定義によれば，偶数の目が出る確率は $3/6 = 0.5$ である．

この定義には直感的には分かりやすいが，根元事象が同程度に確からしい場合にしか定義できないという問題がある．

定義 4.2 (**確率の頻度による定義**)**.** ある試行を n 回行ったとき，ある事象 A が r 回起こったとする．n を限りなく大きくしていくとき，r/n が一定の値 p に近づくならば，A の確率を

$$P(A) = p = \lim_{n\to\infty} \frac{r}{n} \tag{4.2}$$

と定義する[*3]．

■例（コイン投げの確率）　ゆがみのないコインを投げて表が出るかどうかをチェックする．試行回数を n，表が出る回数を r とすると，試行回数が多くなればなるほど r/n は $1/2$ に近づくことが確認できる．

n の極限は数学的には定義しうるものの，経験的には観察不可能であるという問題がある．

コルモゴロフによる確率の公理主義的定義は，確率の経験的内容に関する議論を留保したままで，数学的取り扱いについての形式的側面を公理化したものであり，この定義の登場によってより体系的に確率を数学的に取り扱うことが可能になった．

公理 4.1 (**確率の公理**)**.** 標本空間 Ω における各事象 A について，次の *3* つの条件を満たす実数 $P(A)$ が存在するとき，$P(A)$ を事象 A が起こる確率という．

(1) すべての事象 A について，$0 \leq P(A) \leq 1$.

[*3] $\lim_{n\to\infty} f(n)$ は極限を取るという操作を表す．数学的には厳密な定義が必要であるが，ここでは，n を限りなく大きくしていったときに，$f(n)$ が限りなく近づいていく値，とラフに理解すればよい．

(2) $P(\Omega) = 1$.

(3) $A_1, A_2, A_3, \cdots$ が互いに排反のとき

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots.$$

4.3　確率の性質

4.3.1　確率の基本定理

公理から直ちに導かれる確率の基本定理を確認する．

定理 4.1 (確率の基本定理)**.**

(1) $P(\emptyset) = 0$.

(2) $P(A^c) = 1 - P(A)$.

(3) Ω が n 個の根元事象 $E_i(i = 1, 2, \cdots, n)$ から成り立っているならば，

$$\sum_{i=1}^{n} P(E_i) = 1.$$

(4) $A \subset B \Longrightarrow P(A) \leq P(B)$.

証明. (1) 恒等律もしくは支配律より $\emptyset \cap \Omega = \emptyset$, $\emptyset \cup \Omega = \Omega$ である．ゆえに，

$$P(\emptyset \cup \Omega) = P(\Omega) = P(\emptyset) + P(\Omega)$$

である．$P(\Omega) = 1$ なので，$P(\emptyset) = 0$ である．

(2) 矛盾律より $A \cap A^c = \emptyset$，また排中律より $A \cup A^c = \Omega$ なので

$$P(A \cup A^c) = P(\Omega) = P(A) + P(A^c)$$

より，$P(A^c) = 1 - P(A)$ を得る．

(3) 根元事象は定義より互いに排反であるので

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + P(E_2) + \cdots + P(E_n)$$

である．前提条件より $\Omega = E_1 \cup E_2 \cup \cdots \cup E_n$ なので，定理が成立する．

(4) 集合的に A が B に含まれる，つまり A が B の部分集合であることを $A \subset B$ と

表す．$A \subset B$ であれば，$B = A \cup (A^c \cap B)$ かつ $A \cap (A^c \cap B) = \emptyset$ となる（ベン図を描いて確認せよ）．ゆえに，

$$P(B) = P(A) + P(A^c \cap B)$$

となる．確率の公理 (1) より，$P(A^c \cap B) \geq 0$ なので，定理が成立する．　$\square$

次の定理は**加法定理**として知られている．

定理 4.2 (**加法定理**)**.** 任意の事象 A, B について

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{4.3}$$

証明. ベン図を検討するか，もしくは集合演算基本公式を適用することで

$$\begin{aligned} A &= (A \cap B^c) \cup (A \cap B) \\ B &= (A^c \cap B) \cup (A \cap B) \\ A \cup B &= (A \cap B^c) \cup (A^c \cap B) \cup (A \cap B) \end{aligned}$$

かつ，$(A \cap B^c), (A^c \cap B), (A \cap B)$ が互いに排反であることが分かる．よって，

$$\begin{aligned} P(A) &= P(A \cap B^c) + P(A \cap B) \\ P(B) &= P(A^c \cap B) + P(A \cap B) \\ P(A \cup B) &= P(A \cap B^c) + P(A^c \cap B) + P(A \cap B) \end{aligned}$$

である．ここから，

$$\begin{aligned} P(A \cup B) &= P(A \cap B^c) + P(A^c \cap B) + P(A \cap B) \\ &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) + P(A \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

を得る．　$\square$

A, B が互いに排反のときは $A \cap B = \emptyset$ なので

$$P(A \cup B) = P(A) + P(B)$$

となる．**事象が相互排反のときは，事象の和の確率は確率の和！！**

4.3.2　古典的定義の導出

ある試行について，標本空間の要素数が n であり，それぞれの根元事象 E_i は同程度に確からしく起こるとする．

このとき，$P(E_1) = P(E_2) = \cdots = P(E_n) = \alpha$ であり，定理 4.1 より，

$$\sum_{i=1}^{n} P(E_i) = n\alpha \Longleftrightarrow \alpha = \frac{1}{n}$$

となる．つまり，$P(E_i) = 1/n$ である．

さらに，根元事象の和で構成される複合事象 A について，その要素数が r のとき，A の確率は

$$P(A) = r\alpha = \frac{r}{n}$$

である．これで確率の古典的定義を公理論的に導出することができた．

■例（コイン投げの確率）　コイン投げにおける表と裏の出現が同様に確からしいとき，公理より以下のように確率を与えることができる．

$$P(\{\text{裏}, \text{表}\}) = 1,\quad P(\{\text{裏}\}) = P(\{\text{表}\}) = \frac{1}{2},\quad P(\emptyset) = 0.$$

■例（サイコロ投げの確率）　サイコロの目の出現が同様に確からしいとき，公理よりあたられるサイコロ投げの事象の確率の一部である．

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6},\quad P(\{1,2,3\}) = P(\{4,5,6\}) = \frac{1}{2}.$$

4.4　条件付き確率と独立性

4.4.1　条件付き確率

ある事象 A が起こったときに，さらに事象 B が起こる確率を知りたいとする．このような確率を A が起きたという条件のもとでの B の**条件付き確率** (conditional probability) といい，$P(B|A)$ で表す．

定義 4.3 (**条件付き確率**)**.** 事象 A が起きたという条件のもとでの事象 B の条件付き確率は

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{4.4}$$

で定義される．ただし，事象 A は $P(A) > 0$ であるとする．

この定義は，もともと標本空間 Ω 上で割り当てられた確率 $P(A)$ と $P(A \cap B)$ の比を維持したままで，A を縮小された標本空間と見なして事象 B の確率を再割り当てするという操作を意味している[*4]．

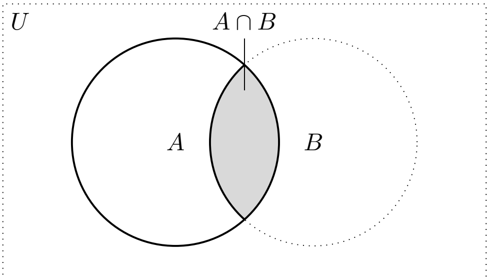

図 4.6　条件付き確率 $P(B|A)$ についてのベン図 (A を標本空間として $A \cap B$ に割り当てた確率測度)

■例（トランプの絵札）　トランプをよく切って 1 枚を引いたときそれがハートであったということがわかっている．このときさらにそのカードが絵札（11, 12, 13）である確率が知りたいとする．「ハート」という事象を A,「絵札」という事象を B として条件付き確率を求める．確率の古典的定義より

$$P(A) = 13/53, \quad P(B) = 12/53, \quad P(A \cap B) = 3/53$$

である．ゆえに条件付き確率 $P(B|A)$ は

$$P(B|A) = \frac{3/53}{13/53} = \frac{3}{13}$$

[*4] 詳しくは，P・G・ホエール，1978,『入門数理統計学』培風社，pp. 14-15 を参照のこと．

となるが，これは A を標本空間として絵札が出る古典的確率を直接求めた値に等しい．

条件 B のもとでの A の条件付き確率も同様に求められ，

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4.5}$$

となる．式 (4.4)，(4.5) の分母を払うと，

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B) \tag{4.6}$$

を得る．この式をとくに**乗法定理**とよぶことがある．

4.4.2　独立

条件付き確率において，事象 A が起こったことが，B が起きる確率に何の影響も与えないことがある．このとき，

$$P(B|A) = P(B) \tag{4.7}$$

となる．これを式 (4.4) に代入すると

$$P(A \cap B) = P(A)P(B) \tag{4.8}$$

を得る．式 (4.7)，(4.8) が成り立っているとき，事象 A と B は**独立** (independent) であるという．

独立の概念を n 個の事象に敷衍することができる．事象 $A_1, A_2, \cdots, A_n$ について，

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n) \tag{4.9}$$

が成り立つとき，事象 $A_1, A_2, \cdots, A_n$ は独立であるという．

■例（トランプの絵札）　さきほどのトランプの絵札 A の条件付き確率の例では，

$$P(B|A) = 3/13 \neq 12/53 = P(B)$$

であるので，統計的に独立ではない．これは「ジョーカー」の 1 枚があるためであって，もとの B の確率よりも，ハート B を条件としたときの条件付き確率 $P(B|A)$ の

ほうが，ジョーカーを排除したために絵札である確率が若干上昇する．ジョーカーを除いた 52 枚のトランプで同様の試行を行う場合，

$$P(A) = 13/52, \quad P(B) = 12/52, \quad P(A \cap B) = 3/52$$

であり，$P(A \cap B) = P(A)P(B)$ が成り立つので，この場合「絵札が出ること」と「ハートが出ること」は独立の事象である．

■例（コインを繰り返し投げる）　ゆがみのないコインを 10 回投げたときに，10 回連続で表が出る確率はどれくらいだろうか．t 回目に表が出るという事象を A_t と表す．確率を求めたい事象は $A_1 \cap A_2 \cap \cdots \cap A_{10}$ である．コインを 10 回投げたときの表裏の組み合わせのパターンは $2^{10} = 1024$ パターンある．どの組み合わせのパターンになるかは「同様に確からしい」と考えられる場合，確率の古典的定義から，

$$P(A_1 \cap A_2 \cap \cdots \cap A_{10}) = \frac{1}{2^{10}}$$

である．一方，事象 A_t の確率 $P(A_t)$ はそれぞれ 1/2 なので，

$$P(A_1)P(A_2)\cdots P(A_{10}) = \left(\frac{1}{2}\right)^{10}$$

なので，

$$P(A_1 \cap A_2 \cap \cdots \cap A_{10}) = P(A_1)P(A_2)\cdots P(A_{10})$$

であり，それぞれのコイン投げの結果は独立である．1 回 1 回の結果が独立となる試行のことを，独立試行ともいう．

4.5　最低限知っておくべきこと

以上の確率論の基礎知識は，今後推測統計を学んでいく上で最低限知っておくべき知識である．そのなかでも，推測統計に関して頻出する知識は以下の 3 点である．受講生諸君はこれを部屋の壁に貼り出し昼夜を問わず 100 万回唱えて血肉化すべきである．

1. 排反な事象のいずれかが起こる確率は，それぞれの**確率の和** (確率の公理 3)

2. 独立な事象が同時に起こる確率は，それぞれの**確率の積** (独立の定義)
3. 復元無作為抽出は独立試行

課題 4.1. *LUNA* のテストを *10/21*（水）*23:59* までに受けなさい.

第 5 章

確率変数と確率分布

引き続き，推測統計学に必須の確率論の知識を導入する．

5.1 確率変数

標本空間のそれぞれの根元事象に適当な数値を対応させた変数 X を考える．このとき，この変数 X がある特定の値 x とる確率が対応する根元事象の確率として定まる．このような変数を**確率変数** (random variable) という．X は変数であり試行を行う前は値は定まっていない．一方，x は試行後に定まった値を表すのがこの分野の慣例である[*1]．

確率変数 X がとびとびの値 $x_1, x_2, \cdots, x_k, \cdots$ しかとらないとき，X を**離散確率変数** (descrete random variable) という．以下，とりうる値は有限個であると仮定して話を進めよう．値が x_i をとるときの確率を

$$P(X = x_i) = p_i \qquad (i = 1, 2, \cdots, k)$$

というように表す．

[*1] 確率変数 X は，試行を行う前，あるいは試行結果が明らかになる前の不確実な変数であり，値が定まっていないものである．なかなかイメージしづらいかもしれないが，スロットやルーレットが回っている状態を思い浮かべてもよいだろう．個人的には X は，プルプルと小刻みに動いているイメージである．これを具現化するために，いちいち $(((X)))$ と表記してもよいが，めちゃめちゃ読みにくくなると思う．諸君も確率論を極めた暁には，文中の確率変数が独りでにプルプルし始めるであろう．

■例（サイコロの目を値としてとる確率変数）　サイコロの目をとる確率変数 X を考えると，それぞれの確率は

$$P(X = i) = \frac{1}{6} \qquad (i = 1, 2, 3, 4, 5, 6)$$

である．

5.2　離散型確率分布

確率 p_i は，値 x_i の関数であるから，これを明示化して

$$P(X = x_i) = f(x_i) \qquad (i = 1, 2, \cdots, k) \tag{5.1}$$

とも表す．このとき $f(x_i)$ は**確率関数** (probability function) といわれる．確率の性質から確率関数は

$$f(x_i) \geq 0 \qquad (i = 1, 2, \cdots, k) \tag{5.2}$$

$$\sum_{i=1}^{k} f(x_i) = 1 \tag{5.3}$$

を満たす．この確率関数によって定まる x と $f(x)$ との関係を**離散型確率分布** (probability distribution of discrete type) とよぶ．

■例（サイコロの目の確率分布）　表 5.1 と図 5.1 はサイコロの目の確率分布，表 5.2 と図 5.2 は 2 個のサイコロの目の和の確率分布を示している．

表 5.1　サイコロの目の確率分布

x	1	2	3	4	5	6
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6

5.3　連続型確率分布

確率変数 X が連続の値をとるとき，X を**連続確率変数** (continuous random variable) という．連続確率変数の場合，1 つ 1 つの値ごとではなく，値のとる範囲に

表 5.2　2 個のサイコロの目の和の確率分布

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

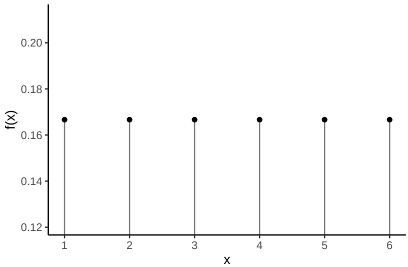

図 5.1　サイコロの目の確率分布

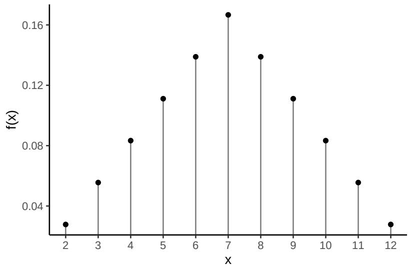

図 5.2　2 個のサイコロの目の和の確率分布

応じて確率が決まると考える．確率変数 X がある範囲の値をとる確率が，関数 $f(x)$ によって

$$P(a \leq X \leq b) = \int_a^b f(x)dx \tag{5.4}$$

と表され，かつ

$$\text{すべての } x \text{ について}, f(x) \geq 0 \tag{5.5}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{5.6}$$

を満たすとき，X は**連続型確率分布** (probability distribution of continuous type) をもつという．このとき関数 $f(x)$ を，**確率密度関数** (probability density function; PDF)，あるいは単に密度関数という．

定義より，確率変数がただ 1 つの値 a をとる確率は

$$P(a \leq X \leq a) = \int_a^a f(x)dx = 0$$

である．連続型の場合は，値の範囲にのみ意味があることに注意すること．

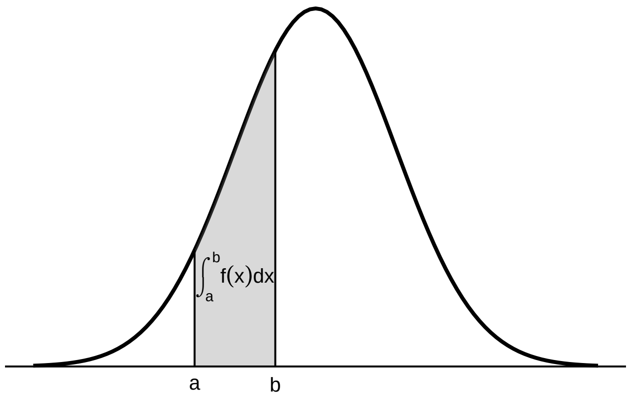

図 5.3　確率密度関数の例

さて，ここでいきなり積分記号 $\int$ が出てきて，戸惑った人もいるかもしれない．本当は（定）積分の導入には，厳密な議論が必要なのだが，ここでは最低限，定積分 $\int_a^b f(x)dx$ は，離散型確率分布のある範囲の確率を和 $\sum_i f(x_i)$ で表すことに対応しており，図形的には「連続関数 $f(x)$ と x 軸（$y=0$ の水平線）との間のにおける a から b までの範囲の面積」を示すことを理解しておけばよい（図 5.3 参照）．

5.4　期待値

5.4.1　期待値の定義

記述統計のところでデータに対して求めた平均や分散を確率変数について求める一般的な操作を**期待値** (expectation) という．

定義 5.1. 確率変数 X の任意の関数を $\varphi(X)$ とするとき，$\varphi(X)$ の**期待値**を離散確率変数の場合

$$\begin{aligned} E[\varphi(X)] &= \varphi(x_1)f(x_1) + \varphi(x_2)f(x_2) + \cdots + \varphi(x_k)f(x_k) \\ &= \sum_{i=1}^{k} \varphi(x_i)f(x_i) \end{aligned} \tag{5.7}$$

連続確率変数の場合

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f(x) dx \tag{5.8}$$

と定義する．

5.4.2　平均と分散

$\varphi(X) = X$ とするとき，確率変数 X の**平均** (mean) が定義される．離散確率変数の場合

$$\mu = E(X) = \sum_{i=1}^{k} x_i f(x_i), \tag{5.9}$$

連続確率変数の場合，

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx \tag{5.10}$$

となる．この確率変数の平均を「期待値」と呼ぶケースが多い．

$\varphi(X) = (X-\mu)^2$ とするとき，確率変数 X の**分散** (variance) が定義される．離散確率変数については

$$\begin{aligned} \sigma^2 &= V(X) = E[(X-\mu)^2] \\ &= (x_1-\mu)^2 f(x_1) + (x_2-\mu)^2 f(x_2) + \cdots + (x_k-\mu)^2 f(x_k) \\ &= \sum_{i=1}^{k} (x_i-\mu)^2 f(x_i) \end{aligned} \tag{5.11}$$

また，連続確率変数については

$$\begin{aligned} \sigma^2 &= V(X) = E[(X-\mu)^2] \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \end{aligned} \tag{5.12}$$

と定義される．分散のルートをとったものを確率変数 X の**標準偏差** (standard deviation) といい，σ で表す．すなわち，$\sigma = \sqrt{\sigma^2}$ である．

5.4.3 一次変換と標準化

平均と分散についての一次変換の結果を確認し，確率変数についての標準化の概念を導入する．

定理 5.1.

$$E(aX+b) = a\mu + b \tag{5.13}$$

証明. 離散的な場合について証明する．連続的な場合も同様に成り立つ．

$$\begin{aligned} E(aX+b) &= \sum_{i=1}^{k}(ax_i+b)f(x_i) \\ &= a\sum_{i=1}^{k}x_i f(x_i) + b\sum_{i=1}^{k}f(x_i) \\ &= a\mu + b \end{aligned}$$

□

定理 5.2.

$$V(aX+b) = a^2V(X) \tag{5.14}$$

証明.

$$\begin{aligned} V(aX+b) &= E[(aX+b-E(aX+b))^2] \\ &= E[(aX+b-a\mu-b)^2] \\ &= a^2E[(X-\mu)^2] = a^2V(X) \end{aligned}$$

□

また，分散は

$$\begin{aligned} V(X) &= E[(X-\mu)^2] \\ &= E(X^2-2X\mu+\mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \end{aligned}$$

と展開することで結局

$$V(X) = E(X^2) - \mu^2 \tag{5.15}$$

と分解することができる．

ここで確率変数の**標準化** (standardization) を導入する．確率変数 X が平均 μ，分散 σ^2 をもつとき，確率変数 Z を

$$Z = \frac{X - \mu}{\sigma}$$

と定義すれば，Z について

$$E(Z) = 0, \quad V(Z) = 1$$

が成立する．Z を標準化確率変数という．

5.5　二項分布

確率分布には様々なタイプのものがあるが，以下では具体例として，もっとも基本的な離散型確率分布である二項分布を導入する．

5.5.1　ベルヌーイ試行とベルヌーイ分布

結果が 2 種類しかない試行を**ベルヌーイ試行** (Bernoulli trials) という．多くの場合，2 つの結果は「成功（S)」と「失敗（F)」と名付けられる．ゆえにベルヌーイ試行の標本空間は

$$\Omega = \{S, F\}$$

である．「コイン投げ試行」などは典型的なベルヌーイ試行である．ベルヌーイ確率変数 X を，事象 $\{S\}$ に対して 1，$\{F\}$ に対して 0 と定める．このとき，ベルヌーイ試行における確率関数は，

$$P(X = x) = f(x) = p^x(1 - p)^{1-x}, \qquad x = 0, 1 \tag{5.16}$$

と定義される．定義より

$$f(x) = \begin{cases} p & (x=1) \\ 1-p & (x=0) \end{cases}$$

である．p はベルヌーイ試行の成功確率であり，$1-p$ は失敗確率である．このように定義される確率分布を**ベルヌーイ分布**という．

■ベルヌーイ分布の平均

$$E(X) = p \times 1 + (1-p) \times 0 = p$$

■ベルヌーイ分布の分散

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

5.5.2　順列の数と組合せの数

二項分布の導入の前に，順列の数と組合せの数について定義しておく．

5 人が参加するコンテストにおいて，3 位までの順位を付けることを考える．このとき，可能な順位付けは，まず，5 人のうち 1 人が 1 位になるのは 5 通り，残り 4 人のうち 1 人が 2 位になるのは 4 通り，残り 3 人のうち 1 人が 3 位になるのは 3 通り，なので全部で

$$5 \times 4 \times 3 = 60 \text{ 通り}$$

になる．

ここで自然数 n に対してその**階乗** (factorial) を

$$n! = \prod_{k=1}^{n} k = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1 \tag{5.17}$$

と定義しておく[*2]．$n=0$ のとき，$0! = 1$ と定義する．

[*2] 「エヌ！」と読む．

さて，一般に，n 個のものから r 個を取り出して並べるときの**順列の数** (permutation) は，

$$
\begin{aligned}
{}_nP_r &= n \times (n-1) \times (n-2) \times \cdots \times (n-(r-1)) \\
&= \frac{n!}{(n-r)!}
\end{aligned}
$$

となる．

順位を付けずに組合せだけを考える場合，例えば 5 人のうち 3 人だけを合格とするような場合，3 人の人が選ばれて順位が付けられたときの順列の数は $3 \times 2 \times 1 = 6$ なので，これを先の順列の数から割れば組合せの数がわかる．つまり

$$\frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10 \text{ 通り}$$

になる．

一般に，n 個のものから r 個を取り出すときの**組合せの数** (combination) は，

$$
\begin{aligned}
{}_nC_r = \begin{pmatrix} n \\ r \end{pmatrix} &= \frac{{}_nP_r}{r!} \\
&= \frac{n!}{r!(n-r)!}
\end{aligned}
$$

となる．

5.5.3　二項分布

同じ成功確率 p をもち，かつ独立なベルヌーイ試行 (つまり，各試行の成功・失敗事象が独立) を n 回繰り返す．各回のベルヌーイ試行の確率変数を $X_i \ (i = 1, 2, \cdots, n)$ と表記する．先ほどと同じく X_i は 0,1 の値をとる．このとき，

$$X = X_1 + X_2 + \cdots + X_n$$

とすると，X もまた確率変数であり，値は 0 から n までの整数値をとり，「n 回の試行のうち成功した回数」を表す．

例えば，1 回の試行の成功確率が 0.3 のとき，5 回試行を行って 3 回成功する確率を求めよう．成功失敗の生起順序は関係ないのであるから，5 回試行を行って 3 回成

功する場合の数は組合せの数として求まり，

$$ {}_5C_3 = \frac{5!}{3!2!} = 10 $$

である．5 回試行を行って 3 回成功する確率は，独立であることに注意すると，

$$ 0.3 \times 0.3 \times 0.3 \times 0.7 \times 0.7 = 0.013 $$

である．この確率を場合の数だけ足し合わせると $10 \times 0.013 = 0.13$ という確率を得る．

一般に，成功確率が p であるベルヌーイ試行を n 回行うとき，x 回成功する確率は

$$ P(X = x) = f(x) = {}_nC_x p^x (1-p)^{n-x}, \qquad x = 0, 1, \cdots, n \tag{5.18} $$

で与えられる．この確率分布を**二項分布** (binomial distribution) という．この二項分布は試行回数 n と成功確率 p を定めれば，形状が一意に決定される．このように，分布の形を定める定数のことを**分布のパラメータ** (parameters)，もしくは母数という．パラメータ n, p をもつ二項分布をとくに

$$ Bi(n, p) $$

と表すことにする．また，ある確率変数 X の分布が二項分布 $Bi(n, p)$ になるとき，「X は $Bi(n, p)$ に従う」と言い方をして，

$$ X \sim Bi(n, p) \tag{5.19} $$

と示すことがある．$Bi(1, p)$ という特殊ケースが先のベルヌーイ分布である．

ところで，「二項定理」を用いると，

$$ \sum_{x=0}^{n} {}_nC_x p^x (1-p)^{n-x} = (p + (1-p))^n = 1 \tag{5.20} $$

となり，確かに確率分布であることが確かめられる．

二項確率変数は，独立なベルヌーイ確率変数の和であるから，確率変数の和についての定理 6.2, 定理 6.3（次章にて導入）より以下が成立する．

■二項分布の平均

$$ E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np $$

■二項分布の分散

$$V(X) = V(X_1) + V(X_2) + \cdots + V(X_n) = np(1-p)$$

ところで，二項分布に従う確率変数 X が独立したベルヌーイ確率変数の和であるという事実から，以下のような定理もほぼ自明に成立する．

定理 5.3. 確率変数 X が $Bi(n,p)$ に従い，確率変数 Y が $Bi(m,p)$ に従っており，かつ，X と Y が独立であるなら，

$$X + Y \sim Bi(n+m, p).$$

このように元の分布のそれらの和の分布が同じ種類の分布であるとき，そうした種類の分布は**再生的** (reproductive) であるという．

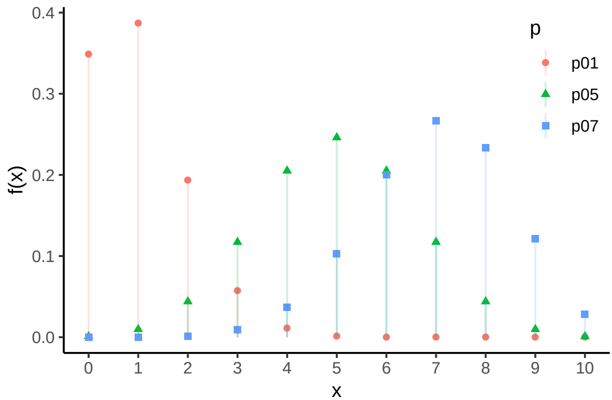

図 5.4　二項分布の例（$n = 10, p = 0.1, 0.5, 0.7$）

課題 5.1. $n = 5, 10, 20$ それぞれの場合の，$p = 0.3, 0.5, 0.7$ の二項分布の棒グラフを作成してレポートにまとめよ．`R` や `Excel` を使ってもよいし，手書きでもよい． *Word* か *pdf* で作成して，*10/28*（水）*23:59* までに *LUNA* 上で提出せよ．

第 6 章

大数の法則と中心極限定理

6.1　本章の結論

本章は本講義の数学的な山場である．この段階で，この山を越えられず少なくない人が討ち死にするかもしれない．御用心御用心.

本章の結論はズバリ以下の通りである.

> 独立かつ同一分布（iid）から得られるデータの相加平均（値の総和をデータの個数で割ったもの）は，サンプルサイズ n が十分に大きいとき，データ分布の平均 μ を中心に正規分布し，n が大きくなるにつれ分布の分散は小さくなる

このことはつまり，ランダムサンプリングによって十分に大きいサンプルが得られた場合，もとのデータ分布の形がわからなくても，データを要約したりデータの構造を示す値の精度や誤差については理論的に予測がつくことを保証しており，これをもとに統計的推定（の評価）や検定が可能になるのである.

これは相当にヤバい結論なのだが，本章ではできるだけ数学的正確さを犠牲にせず，このヤバさを実感していきたいと思う．ただし，詳細はどうしても割愛せざるを得ない．詳細は巻末の付録にまとめているので，興味ある向きはフォローして欲しい.

6.2　iid!! iid!! iid!!

本章で取り上げる大数の法則や中心極限定理の前提となる仮定が，**独立かつ同一の分布に従う** n 個の確率変数

$$X_1, X_2, \cdots, X_n$$

である．英語では **independent and identical distribution** であり，**iid** と略される．

たとえば，同じコインを n 回（n 個同時に）投げたときの裏表は，iid な n 個のベルヌーイ確率変数として表せる．

あるいは，同じサイコロを n 回（n 個同時に）投げたときの出目は，iid な n 個の同一の離散確率変数として表せる．

われわれにとって重要なのは次の例である．集団から n 人を（重複を許して）抽選で選び出したとき（ランダム・サンプリング）の選び出された人の特徴は，iid な n 個の同一の確率変数として表すことができる．

iid はめちゃめちゃめちゃめちゃ[*1]重要なので，普段の授業であれば毎回の授業で受講生に 100 万回ほど復唱してもらう．諸君も自宅にて毎日 100 万回復唱して欲しい．

6.3　同時確率分布

6.3.1　同時確率関数

サイコロ投げとコイン投げを同時に行う．このときサイコロの目を X で表し，コインの表（$Y=1$），裏（$Y=0$）を確率変数 Y で表したとき，それぞれの確率変数の値が**同時**に出る確率を求めることができる（表 6.1 参照）．例えば，サイコロの目が 3 でコインは表である確率は，「3」という事象を A,「表」という事象を B とすると，

$$P(A \cap B) = P(X=3, Y=1) = \frac{1}{12}$$

と表される．

[*1] めちゃめちゃめちゃめちゃめ

表 6.1　コイン投げとサイコロ投げの同時確率分布

	x						
y	1	2	3	4	5	6	$h(y)$
1	1/12	1/12	1/12	1/12	1/12	1/12	1/2
0	1/12	1/12	1/12	1/12	1/12	1/12	1/2
$g(x)$	1/6	1/6	1/6	1/6	1/6	1/6	1

一般に，X と Y が離散確率変数であるとき，$X = x_i$ であり同時に $Y = y_j$ である確率を

$$P(X = x_i, Y = y_j) = f(x_i, y_j) \qquad (i = 1, 2, \cdots, k; j = 1, 2, \cdots, l) \tag{6.1}$$

と表す．このとき，$f(x, y)$ は**同時確率関数** (joint probability function) という．また，x, y と $f(x, y)$ の関係を**同時確率分布** (joint probability distribution) という．

同時確率分布は，多変量解析において欠かすことのできない知識であるが，本講義の目的はあくまで推測統計学のエッセンスを学ぶことにあるので，詳細に検討することはしない．連続確率変数についての同時確率分布についても本来は導入すべき所だが，複雑さを避けるためここでは離散型同時確率分布のみに焦点を当てて最低限の知識を習得する．離散型同時確率分布における基本的性質は，同様に連続型同時確率分布にも適用できる．

確率の特性より，同時確率関数は

$$f(x_i, y_j) \geq 0 \qquad (i = 1, 2, \cdots, k; j = 1, 2, \cdots, l) \tag{6.2}$$

$$\sum_{i=1}^{k} \sum_{j=1}^{l} f(x_i, y_j) = 1 \tag{6.3}$$

を満たさなければならない．

X 単独の確率関数を $g(x)$，Y 単独の確率関数を $h(y)$ とすると，

$$g(x) = \sum_{j=1}^{l} f(x, y_j), \qquad h(y) = \sum_{i=1}^{k} f(x_i, y) \tag{6.4}$$

によって求めることができる．それぞれ X と Y の**周辺確率関数** (marginal probability function) とよび，その分布を**周辺確率分布** (marginal probability distribution) という．

■例（サイコロ投げとコイン投げの同時確率分布における周辺確率分布）　表 6.1 から，サイコロ投げの周辺確率関数は，

$$g(x) = \sum_{j=1}^{2} f(x, y_j) = f(x,1) + f(x,0) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6} \tag{6.5}$$

コイン投げの周辺確率関数は，

$$h(y) = \sum_{i=1}^{6} f(x_i, y) = f(1,y) + f(2,y) + \cdots + f(6,y) = 6\frac{1}{12} = \frac{1}{2} \tag{6.6}$$

である．

式 (4.8) の独立性の条件を確率変数に適用すると，確率変数 X と Y について

$$f(x,y) = g(x)h(y) \tag{6.7}$$

が常に成り立つとき，X と Y は独立であるという．

6.3.2　共分散

確率変数 X と Y の**共分散** (covariance) を

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \tag{6.8}$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{l}(x_i-\mu_X)(y_j-\mu_Y)f(x_i,y_j) \tag{6.9}$$

と定義する．ここで，μ_X, μ_Y はそれぞれ確率変数 X, Y の平均である．独立性と共分散の以下の関係は重要である．

定理 6.1. 確率変数 X と Y が独立であれば，共分散はゼロである．すなわち，

$$Cov(X,Y) = 0.$$

証明. 式 (6.9) より，展開すると，

$$
\begin{aligned}
Cov(X,Y) =& \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j f(x_i,y_j) - \mu_Y \sum_{i=1}^{k} x_i \sum_{j=1}^{l} f(x_i,y_j) \\
& - \mu_X \sum_{j=1}^{l} y_j \sum_{i=1}^{k} f(x_i,y_j) + \mu_X\mu_Y \sum_{j=1}^{l}\sum_{i=1}^{k} f(x_i,y_j) \\
=& \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j f(x_i,y_j) - \mu_Y \sum_{i=1}^{k} x_i g(x_i) \\
& - \mu_X \sum_{j=1}^{l} y_j h(y_j) + \mu_X\mu_Y \\
=& E(XY) - \mu_X\mu_Y
\end{aligned}
$$

という展開式を得る．ここで，

$$
E(XY) = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j f(x_i,y_j) \tag{6.10}
$$

である．独立性の仮定より，すべての i と j について，$f(x_i,y_j) = g(x_i)h(y_j)$ が成り立つ．これを式 (6.10) に代入すると，

$$
\begin{aligned}
E(XY) &= \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j g(x_i)h(y_j) \\
&= \sum_{i=1}^{k} x_i g(x_i) \sum_{j=1}^{l} y_j h(y_j) \\
&= \mu_X\mu_Y
\end{aligned}
$$

となるので，$E(XY) - \mu_X\mu_Y = 0$ である． □

共分散と標準偏差を用いて，確率変数 X と Y の**相関係数** (correlation) を次のように定義できる．

$$
\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y} \tag{6.11}
$$

6.4　確率変数の和

確率変数の和についての性質を確認する．前節に引き続き，離散確率変数のみ検討するが，ここでの結果は連続確率変数の場合でもまったく同じである．

定理 6.2. *2* つの離散確率変数 X, Y の和 $X+Y$ の期待値は

$$E(X+Y) = E(X) + E(Y) \tag{6.12}$$

である．

証明.

$$\begin{aligned}
E(X+Y) &= \sum_{i=1}^{k}\sum_{j=1}^{l}(x_i + y_j) f(x_i, y_j) \\
&= \sum_{i=1}^{k} x_i \sum_{j=1}^{l} f(x_i, y_j) + \sum_{j=1}^{l} y_j \sum_{i=1}^{k} f(x_i, y_j) \\
&= \sum_{i=1}^{k} x_i g(x_i) + \sum_{j=1}^{l} y_j h(y_j) \\
&= E(X) + E(Y)
\end{aligned}$$

□

一般に，n 個の確率変数について

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) \tag{6.13}$$

が成立する．

定理 6.3. *2* つの離散確率変数 X, Y の和 $X+Y$ の分散は

$$V(X+Y) = V(X) + V(Y) + 2Cov(X, Y) \tag{6.14}$$

である．X と Y が独立の時

$$V(X+Y) = V(X) + V(Y) \tag{6.15}$$

である．

証明.

$$
\begin{aligned}
V(X+Y) &= E[(X+Y-E(X+Y))^2] \\
&= E[((X-\mu_X)+(Y-\mu_Y))^2] \\
&= E[(X-\mu_X)^2]+E[(Y-\mu_Y)^2]+2E[(X-\mu_X)(Y-\mu_Y)] \\
&= V(X)+V(Y)+2Cov(X,Y)
\end{aligned}
$$

□

一般に，n 個の確率変数が独立のとき，

$$
V(X_1+X_2+\cdots+X_n)=V(X_1)+V(X_2)+\cdots+V(X_n) \tag{6.16}
$$

が成立する．

定理 6.4. $X_1, X_2, \cdots X_n$ が独立で，かつ平均 μ，分散 σ^2 の同一の確率分布に従うとする ***(iid)***. このとき，X_i の相加平均（データの平均）

$$
\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{X_1+X_2+\cdots+X_n}{n}
$$

の期待値と分散は

$$
E(\bar{X})=\mu, \qquad V(\bar{X})=\frac{\sigma^2}{n}
$$

となる．

証明.

$$
E(\bar{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)=\frac{n\mu}{n}=\mu \tag{6.17}
$$

$$
V(\bar{X})=V\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}V\left(X_i\right)=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n} \tag{6.18}
$$

□

定理 6.4 の結果は，標本理論において大変重要になる．

■例（サイコロの出目の平均・分散）　サイコロ試行の平均（期待値）は $\mu = 3.5$ であり，分散は $\sigma^2 = 2.92$ である．互いに影響が及ばないようにサイコロを n 個投げる，あるいは n 回サイコロ投げを繰り返したときの出目の相加平均の平均（期待値）は $E(\bar{X}) = 3.5$，相加平均の分散は $V(\bar{X}) = 2.92/n$ である．分散は n が大きくなるほど急速に小さくなる（図 6.1）．

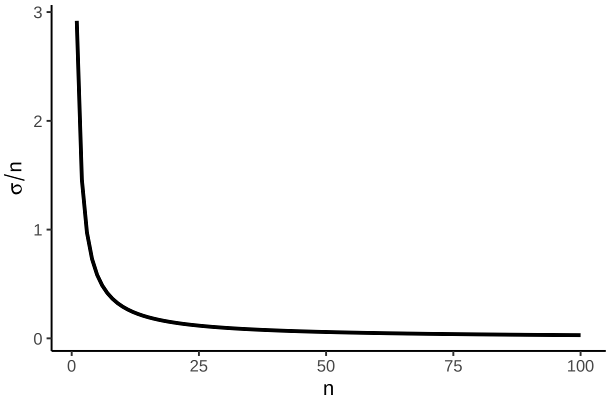

図 6.1　サイコロの出目の相加平均の分散

6.5　大数の法則

6.5.1　成功割合の分布

成功確率が p でかつ独立のベルヌーイ試行の成功割合を $\hat{p}$ とすると，

$$\hat{p} = \frac{X}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

である．ここで，$X = X_1 + X_2 + \cdots + X_n$ は試行回数 n，成功確率 p の二項確率変数であり，その一次変換の $\hat{p}$ もやはり確率変数である．

確率変数の一次変換の定理 5.1, 定理 5.2 より，確率変数 $\hat{p}$ の平均は

$$E(\hat{p}) = E(X/n) = (1/n)E(X) = p,$$

分散は

$$V(\hat{p}) = (1/n)^2 V(X) = \frac{p(1-p)}{n}$$

となる．つまり，n が大きくなるにつれて，平均は変わらないものの，分散はどんどん小さくなることが分かる．

$\hat{p}$ についての確率関数は式 (5.18) より，$n\hat{p} = X$ を代入すると，

$$P(\hat{p} = x/n) = f(x) = {}_nC_x p^x (1-p)^{n-x}, \qquad x = 0, 1, \cdots, n \tag{6.19}$$

となる．つまり，成功割合の確率関数としては二項分布の確率関数と同じである．成功割合の分布は図 6.2 のようになる[*2]．

[*2] ここでは，連続確率分布における確率密度関数のような，実現値の範囲の確率にグラフの面積が対応する形式で表現している．というのも，確率関数のまま分布をグラフにすると，n が大きくなればなるほど確率関数の定義域である $[0, 1]$ においてとられる点は多くなるので，当然ながら 1 点の確率は小さくなり傾向の変化は見にくくなるためである．

以下はそのようなグラフの構成法である（マニアックなので興味のある人だけが見ればよい）．

確率変数 $\hat{p}$ の実現値をここでは $\pi = x/n$ と表すことにする．グラフを構成するとき，x が 1 単位動いたときの範囲の確率と，それに対応して π が 1 単位動いたときの範囲の確率が同じになる必要がある．そうしないと全範囲で 1 にならないからである．一般的に，x から $x + \Delta x$ までの範囲の成功回数をとる確率は，近似的に

$$P(x \leq X < x + \Delta x) = f(x)\Delta x$$

と計算することができる．x が動く単位を Δx，π が動く単位を $\Delta \pi$ とすると，

$$\Delta x = 1, \qquad \Delta \pi = \frac{1}{n}$$

であるので，

$$P(x \leq X < x + 1) = f(x) = \varphi(\pi)\frac{1}{n} = P\left(\pi \leq \hat{p} < \pi + \frac{1}{n}\pi\right)$$

が成立する必要がある．ゆえに，π の範囲の確率にグラフの面積を対応させる関数は

$$\varphi(\pi) = nf(n\pi) = nf(nx)$$

となる．これが，連続確率分布における確率密度関数に対応する．

ここで行った操作は，より一般的には確率密度関数の**変数変換**という操作であり，確率変数 X を何らかの変換 $Y = h(X)$ によって Y へと変換するとき，Y についての確率密度関数 $g(y)$ は，X についての密度関数 $f(x)$ を用いて

$$g(y) = f(x)\left|\frac{dx}{dy}\right| \tag{6.20}$$

と変換できる．

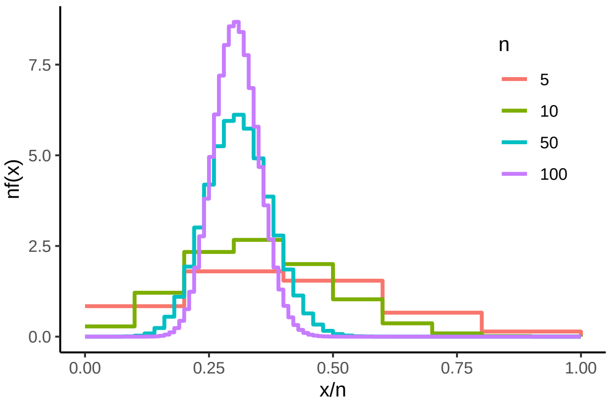

図 6.2　成功割合分布の例（$p = 0.3, n = 5, 10, 50, 100$）

6.5.2　大数の法則

成功割合分布の分散式 $p(1-p)/n$ や分布のグラフ図 6.2 から予想されるように，試行回数 n が多くなければなるほど，成功割合の分布は成功確率 p に集中していくように見える．この直感は「大数の法則」によって理論的に確かめられる．以下では大数の法則の結論のみを導入するが，証明の詳細に関心のある向きは，付録 A.2 節を参照されたい．

さて，チェビチェフの不等式 (付録 A.2 節参照) を，n 回独立ベルヌーイ試行の成功割合 $\hat{p}$ について適用すると，

$$\lim_{n\to\infty} P(|\hat{p} - p| \geq \epsilon) = 0 \tag{6.21}$$

が言える．このことは成功割合 $\hat{p}$ が区間 $[p-\epsilon, p+\epsilon]$ の外に出る確率が，ϵ をどんなに小さくとっても，n が大きくなるにつれて 0 に近づくと言うことを意味している．逆に言えば，成功割合 $\hat{p}$ が区間 $[p-\epsilon, p+\epsilon]$ に入っている確率は 1 に近づく．このとき，$\hat{p}$ は p に**確率収束** (converge in probability) するといって，

$$\operatorname*{plim}_{n\to\infty} \hat{p} = p \tag{6.22}$$

と表記する．これが**大数の法則** (law of large numbers) の特殊ケースであり，これによって確率の頻度による定義に理論的な妥当性が与えられる．

より一般に，以下が成立する．

定理 6.5 (大数の法則)**.** $X_1, X_2, \cdots X_n$ が独立で，かつ平均 μ，分散 σ^2 の同一の確率分布に従うとする *(iid)*．このとき，X_i の相加平均

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

は確率変数であり，$\bar{X}$ は μ に確率収束する．すなわち，

$$\operatorname*{plim}_{n\to\infty} \bar{X} = \mu. \tag{6.23}$$

証明. 付録 A.2 節参照. □

大数の法則から，ベルヌーイ試行の場合，成功割合は試行を多くすると成功確率そのものに近づくということが理論的に言える．また，一般にある 1 つの分布から n 個の値をランダムに重複を許して抜き出して相加平均をとったとき，抜き出す数が多ければ多いほど分布の平均に近づくということが理論的に言える．これらは直感的には明らかではあるが，理論的に妥当であることがきちんと証明できるという点が重要である．

6.6　中心極限定理

成功割合分布のグラフ図 6.2 からさらに予想されることは，n が大きくなればなるほど，グラフは釣り鐘型のなめらかな形をするようになるということである．実際に，成功割合分布は，n が大きくなるとき，**正規分布** (normal distribution) に近づくということが知られている．これこそが，確率統計学の大定理である**中心極限定理** (central limit theorem) の特殊ケースである．

6.6.1　二項分布についての中心極限定理

二項分布について以下のことが成り立つ．詳細は付録 A.3 節を参照のこと．

定理 6.6 (二項分布についての中心極限定理)**.** 二項分布 $Bi(n, p)$ に従う確率変数 X を標準化して

$$Z = \frac{X - \mu}{\sigma}$$

とするとき，$n \to \infty$ であれば，確率変数 Z は，確率密度関数

$$g(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \tag{6.24}$$

をもつ連続確率分布に従う[*3].

証明. 付録 A.3 節参照. □

確率密度関数 (6.24) をもつ連続分布を**標準正規分布** (standard normal distribution) という．これは平均が 0，分散が 1 の特殊な正規分布である．より一般的に，連続変数 y に対して

$$h(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \tag{6.25}$$

の形の分布を，平均 μ 分散 σ^2 の**正規分布** (normal distribution) といって，

$$N(\mu, \sigma^2)$$

で表す．標準正規分布の場合 $N(0, 1)$ である．

ここまでの知見をまとめると，以下のようになる．

> 二項分布 $Bi(n, p)$ に従う X を標準化した Z の分布は，n が大きくなると標準正規分布 $N(0, 1)$ に近づく．

6.6.2　相加平均についての中心極限定理

中心極限定理は，二項分布だけではなくより一般的な分布について成立する．中心極限定理の 1 つの形式として，次の相加平均の中心極限定理は大変重要である．証明は，この講義のレベルを超えるので，ここでは証明なしで導入する．

[*3] $\exp\{x\}$ は指数関数 e^x を表す.

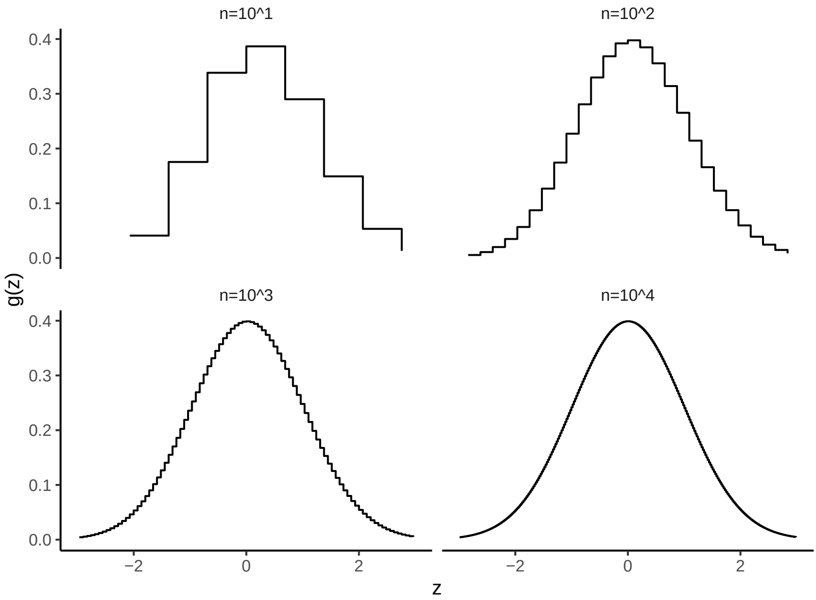

図 6.3　$g(z)$ の分布（$p = 0.3, n = 10^1, 10^2, 10^3, 10^4$）

定理 6.7 (平均についての中心極限定理)**.** $X_1, X_2, \cdots, X_n$ が独立で，平均 μ 分散 σ^2 をもつ同一の分布に従っているとする ***(iid)*.** $X_1, X_2, \cdots, X_n$ の相加平均

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

に対して，標準化した確率変数を

$$Z_n = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

とすると，$n \to \infty$ のとき，Z_n の分布は標準正規分布 $N(0, 1)$ に従う．

また，$n \to \infty$ のとき，Z_n の分布は標準正規分布 $N(0, 1)$ に従うことから，変数変換によって，$\bar{X}$ 自身は正規分布 $N(\mu, \sigma^2/n)$ に近似的に従うことがいえる[*4]．

[*4] 厳密に言えば，中心極限定理によって言えるのは，$n \to \infty$ のとき Z_n の分布が標準正規分布にど

ここで示されていることは，標本理論に引きつけて言えば，ある 1 つの分布から n 個の値をランダムに重複を許して抜き出して相加平均をとったとき，抜き出す数が多ければ多いほど**相加平均の分布は正規分布に近づく**ということである．それがどんな分布であろうと $\cdots\cdots$．これはかなり恐ろしいことである．このような分布の特性を**漸近正規性** (asymptotic normality) という．また，この話から正規分布（とくに標準正規分布）というものが，統計学にとっていかに大事なものであるかが分かるだろう．

課題 6.1. 大数の法則と中心極限定理について自分自身の言葉でまとめよ．*Word* ファイルでリポートを作成し，*11/4*（水）*23:59* までに *LUNA* 上で提出せよ．

こまでも近づく（分布収束）ことであり，$\bar{X}$ の分布についてはあくまで変数変換による近似である．

第 7 章

標本と標本分布

いよいよここから，これまでの確率論の知識をもとに，調査によって得られた標本を分析する方法を習得する．まず，本章では，これまでの確率論の枠組みを母集団と標本との関係に適用していく．それゆえ，前章まででの概念に，母集団と標本に関連した具体的な名前が割り当てられて出てくる．とくに，キーとなる概念が**標本分布** (sampling distribution) である．この概念の意味が分からず，さらに多くの人がここで脱落するかもしれない．注意深くついてきて欲しい．

7.1　正規分布の性質

最初に前章のつづきとして，前章に登場した，統計理論にとってたいへん重要な連続確率分布である正規分布の一般的な性質を確認しよう．連続確率変数 Y が，

$$h(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \tag{7.1}$$

の形の分布をもつとき，Y は平均 μ 分散 σ^2 の正規分布 $N(\mu, \sigma^2)$ というのであった．正規分布は平均を頂点（モード）とする左右対称の分布をする．このとき，平均 μ と分散 σ^2 が分布を決定するパラメータであり，μ は分布の頂点を σ^2 は分布の幅を決める[*1]．

確率変数 Y を標準化すると，変数変換公式 (付録 A.4 節参照) より，

$$g(z) = \sigma h(\sigma z + \mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-z^2}{2}\right\} \tag{7.2}$$

[*1] R の正規分布の密度関数は `dnorm(x, mean, sd)` であるが，この関数では分散ではなく標準偏差を指定することに注意．

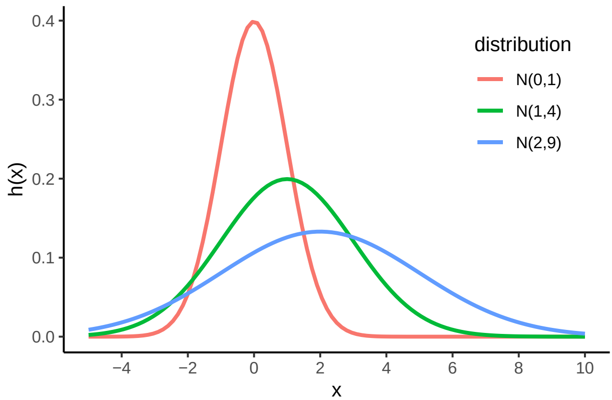

図 7.1　正規分布の例

となり，Z は標準正規分布 $N(0,1)$ に従うことが分かる．つまり，どのような正規分布であろうと，標準化することで標準正規分布に帰着させることができる．

7.1.1　確率計算

正規分布に従う確率変数が，ある範囲の値をとる確率を知るためには，正規分布の確率密度関数 (7.1) の積分演算で求めればよい．普通には解けないので，コンピュータプログラムに頼ったり，たいていの統計の教科書の巻末に載っている分布表を利用する．しかし，最低限には以下のことを知っていれば十分である．

標準正規分布 $N(0,1)$ に従う Z について，以下が成立する．

$$P(-1 \leq Z \leq 1) = \int_{-1}^{1} g(z)dz = 0.6827, \quad P(Z \leq -1, 1 \leq Z) = 1 - 0.6827 = 0.3173$$
$$P(-2 \leq Z \leq 2) = \int_{-2}^{2} g(z)dz = 0.9545, \quad P(Z \leq -2, 2 \leq Z) = 1 - 0.9545 = 0.0455$$
$$P(-3 \leq Z \leq 3) = \int_{-3}^{3} g(z)dz = 0.9973, \quad P(Z \leq -3, 3 \leq Z) = 1 - 0.9973 = 0.0027$$

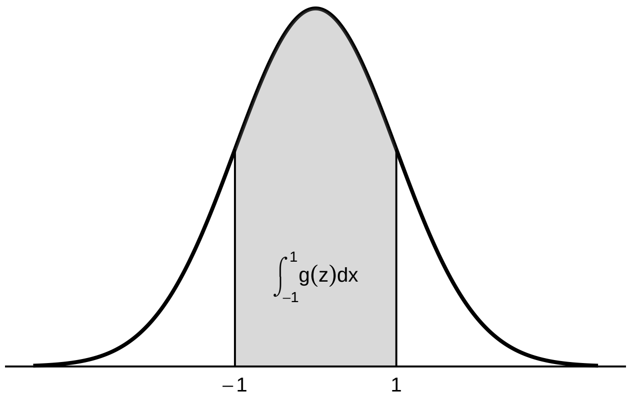

図 7.2　標準正規分布の積分計算

また，後の統計的検定に際しては，次の確率計算が頻繁に用いられる．

$$P(-1.96 \leq Z \leq 1.96) = 0.95, \qquad P(Z \leq -1.96, 1.96 \leq Z) = 0.05$$
$$P(-2.576 \leq Z \leq 2.576) = 0.99, \qquad P(Z \leq -2.576, 2.576 \leq Z) = 0.01$$

つまり，Z が $N(0,1)$ に従うとき，$|Z| \geq 1.96$ となる確率は高々 5% であり，$|Z| \geq 2.576$ となる確率は高々 1% であることが分かる．こうした確率の値は，仮説検定において「有意水準」という名前で重要な役割を与えられている．

標準正規分布から任意の正規分布に変換すると，$Y = \sigma Z + \mu$ なので，正規分布 $N(\mu, \sigma^2)$ に従う Y について，以下が成立する．

$$P(\mu - \sigma \leq Y \leq \mu + \sigma) = 0.6827, \qquad P(Y \leq \mu - \sigma, \mu + \sigma \leq Y) = 0.3173$$
$$P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) = 0.9545, \qquad P(Y \leq \mu - 2\sigma, \mu + 2\sigma \leq Y) = 0.0455$$
$$P(\mu - 3\sigma \leq Y \leq \mu + 3\sigma) = 0.9973, \qquad P(Y \leq \mu - 3\sigma, \mu + 3\sigma \leq Y) = 0.0027$$

$$P(\mu - 1.96\sigma \leq Y \leq \mu + 1.96\sigma) = 0.95, \qquad P(Y \leq \mu - 1.96\sigma, \mu + 1.96\sigma \leq Y) = 0.05$$
$$P(\mu - 2.576\sigma \leq Y \leq \mu + 2.576\sigma) = 0.99, \qquad P(Y \leq \mu - 2.576\sigma, \mu + 2.576\sigma \leq Y) = 0.01$$

7.1.2　正規分布の再生性

最後に正規分布の再生性の定理を証明なしで導入しておく．

定理 7.1. $X_1, X_2, \cdots, X_n$ が独立で，すべて正規分布 $N(\mu, \sigma^2)$ に従うとき，

(1) $X_1 + X_2 + \cdots + X_n$ は $N(n\mu, n\sigma^2)$ に従い，
(2) $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$ は $N(\mu, \sigma^2/n)$ に従う．

また，$\bar{X}$ を標準化して

$$Z_n = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

とすると，Z_n は $N(0,1)$ に従う．

中心極限定理は任意の分布について，$\boldsymbol{n}$ **を大きくすると**，相加平均を標準化した Z_n の分布が標準正規分布に近づくことを述べているが，正規分布の場合は n の大小に関係なく，再生性より標準正規分布に従うことが保証される．

7.2　母集団と標本

われわれが調査対象としたい集団全体を**母集団** (population) という．母集団を全部調査することができれば（このような調査を悉皆調査といった），2 章のような記述統計学だけで事足りる．一方，母集団の一部を取り出し分析することで，母集団の特性を推測する場合には，確率論を基礎にする**推測統計学** (inference statistics) の知識が必要不可欠である．

母集団から調査対象として選び出された対象を**標本** (sample) という．標本を選び出すことを**標本抽出** (sampling) という．特に，標本を等確率で選び出すことを**無作為抽出** (random sampling) という[*2]．実際に調査を実施する場合，抽出法の実践的

[*2] 抽出方法にも，抽出した対象を 1 回 1 回元に戻す「復元抽出」と元に戻さない「非復元抽出」がある．理論的には，標本間の独立性を仮定するためには復元抽出が必要であるが，母集団が十分に大きい場合は抽出方法としてほとんど差がなくなるため，実際にはあまり気にしなくてよい．

なテクニックについての知識も習得する必要が出てくるが，この授業では触れないことにする．

母集団の何らかの特性について，その分布を**母集団分布** (population distribution) という．母集団から無作為抽出を行うとき，母集団分布は何らかの確率分布 $f(x)$ であると仮定することができる．つまり，母集団における特性の分布によってランダムに抽選を行った場合，ある特性の人が選ばれる確率が変わる[*3]．これを確率分布 $f(x)$ によって表現するのである．これによって，ある特性についての標本のそれぞれの値 $X_1, X_2, \cdots, X_n$ は $f(x)$ に従う確率変数であると見なすことができる[*4]．特に，無作為抽出標本の場合

> 標本 $X_1, X_2, \cdots, X_n$ は，**同一の**母集団分布 $f(x)$ に従う n 個の**独立な**確率変数である

と見なすことができる．iid !! iid !! iid !! このことが以下の議論において決定的に重要な前提となる．**これ以降標本といったときには，特別の断りがない限りは，無作為抽出標本を指すものとする．**

7.3　母数と統計量

7.3.1　母平均と母分散

ある確率分布を特定する値をパラメータあるいは母数というのであった．特に，母集団分布の平均 μ を**母平均** (population mean)，分散 σ^2 を**母分散** (population variance) という．母平均と母分散（あるいは母標準偏差）は，記述的な意味で重要であるだけでなく，推定や検定においても重要な役割を演じる．

[*3] たとえば，社会学部を母集団とするとランダムサンプリングで男女が同じくらい選ばれやすいが，理工学部であれば圧倒的に男性が選ばれやすい．つまり，性別という特性についての母集団分布が異なる．

[*4] 確率変数である標本の実現値は普通小文字で $x_1, x_2, \cdots, x_n$ と表す．

7.3.2　統計量

標本 $X_1, X_2, \cdots, X_n$ を引数とする任意の関数

$$f(X_1, X_2, \cdots, X_n)$$

を**統計量** (statistic) と呼ぶ．標本 $X_1, X_2, \cdots, X_n$ はそれぞれ母集団分布に従う確率変数であるから，統計量もまた確率的に分布する．この統計量の分布を**標本分布** (sampling distribution) という[*5]．

統計量としては，最大値，最小値など比較的単純なものから，不平等度の指数であるジニ係数など複雑なものまで様々な種類がある．そのなかでも，もっとも基本となるのが標本平均と標本分散である．

7.3.3　標本平均

標本 $X_1, X_2, \cdots, X_n$ から

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

によって与えられる統計量を**標本平均** (sample mean) という．

母集団分布が母平均 μ，母分散 σ^2 をもつとき，標本平均の期待値（期待値）と分散は，6 章の定理 6.4（式 (6.17), (6.18)）がそのまま適用できるので，

$$E(\bar{X}) = \mu, \qquad V(\bar{X}) = \frac{\sigma^2}{n}$$

となる．また，

$$\sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

は標本平均分布についての標準偏差を示すもので，これを特に母平均の点推定の**標準誤差** (standard error) と呼ぶことがある．

[*5] スライムが確率変数とすると，それを何らかの形で組み合わせたキングスライムは統計量であり確率的に分布する．

7.3.4　標本分散

標本 $X_1, X_2, \cdots, X_n$ から

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

によって与えられる統計量を**標本分散** (sample variance) という．

母集団分布が母平均 μ，母分散 σ^2 をもつとき，標本分散 $\hat{\sigma}^2$ の期待値は，

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\{(X_i - \mu) - (\bar{X} - \mu)\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\{(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2\} \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - 2\left\{\frac{1}{n}\sum_{i=1}^{n}X_i - \mu\right\}(\bar{X} - \mu) + (\bar{X} - \mu)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\bar{X} - \mu)^2
\end{aligned}$$

より，

$$\begin{aligned}
E(\hat{\sigma}^2) &= \frac{1}{n}\sum_{i=1}^{n}E[(X_i - \mu)^2] - E[(\bar{X} - \mu)^2] \\
&= \frac{1}{n}(n\sigma^2) - \frac{\sigma^2}{n}
\end{aligned}$$

ゆえに，

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \tag{7.3}$$

となる．ここで，$E[(X_i - \mu)^2] = \sigma^2$ は分散の定義そのものから，$E[(\bar{X} - \mu)^2] = \sigma^2/n$ は定理 6.4 から得られる．

このことから分かるとおり，標本分散の期待値は母分散と等しくなく，$-\sigma^2/n$ だけズレがある．そしてこのズレは標本サイズが小さければ小さいほど大きくなる．統計量の期待値が母数と等しくない場合，母数を統計量から推定する場合偏りがあるこ

とを意味しているため，その統計量を母数の推定量として用いるのは都合が悪い（詳しくは次章）．そこで以下，標本分散としては，次の**不偏標本分散** (unbiased sample variance) S^2 を用いることにする．

■不偏標本分散

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{7.4}$$

平方和を n ではなく $n-1$ で割ることだけが（不偏でない）標本分散と異なる．不偏分散の期待値は

$$E(S^2) = \sigma^2 \tag{7.5}$$

となる．

■標本標準偏差

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{7.6}$$

7.4　χ^2 分布

ここで，不偏標本分散の標本分布に関連して，χ^2 分布を導入しよう．

定義 7.1. $Z_1, Z_2, \cdots, Z_k$ を独立な，標準正規分布 $N(0,1)$ に従う確率変数とすると，

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

は自由度 k の $\boldsymbol{\chi^2}$ **分布** *(χ^2 distribution)* に従う．自由度 k の χ^2 分布を $\chi^2(k)$ と書く．「カイジジョウブンプ」と読む．

ここで，**自由度** (degree of freedom) という概念が登場する．自由度は統計において非常に重要な概念なのだがわかりにくい．本講義のレベルでは，自由度とは「制約なく自由に動くことのできる確率変数の数」であり，自由度により分布の形が変わる分布があり，χ^2 分布はその一つであると理解しておけばよい．

正規母集団からの標本に関して，以下のことが成立することは自明である．

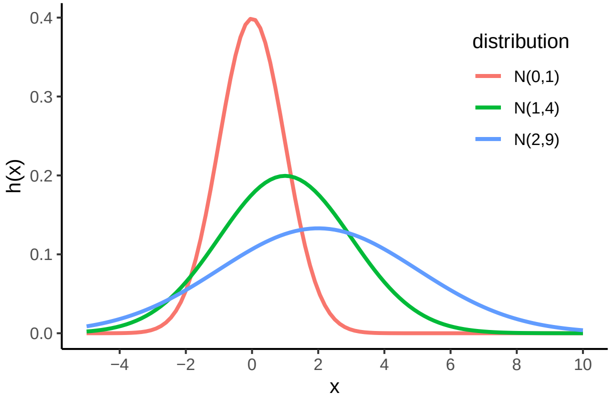

図 7.3　χ^2 分布の例

定理 7.2. 正規分布 $N(\mu, \sigma^2)$ に従う母集団から無作為に抽出された，大きさ n の標本 $X_1, X_2, \cdots, X_n$ について，

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \left(\frac{X_1-\mu}{\sigma}\right)^2 + \left(\frac{X_2-\mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n-\mu}{\sigma}\right)^2$$

は自由度 n の χ^2 分布 $\chi^2(n)$ に従う．

定理 7.2 における母平均 μ を標本平均 $\bar{X}$ に置き換えた場合，次のような定理が成立する．この定理は証明を与えないが，

$$\sum_{i=1}^{n}(X_i-\bar{X}) = 0$$

という制約によって，添字が 1 から $n-1$ 番目までの変数の値を決めれば，最後の n 番目の変数の値は自動的に決まることから，χ^2 分布の自由度は 1 つ減って $n-1$ になることに注意.

定理 7.3. 正規分布 $N(\mu, \sigma^2)$ に従う母集団から無作為に抽出された，大きさ n の標

本 $X_1, X_2, \cdots, X_n$ について，

$$U = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)S^2}{\sigma^2}$$

は自由度 $n-1$ の χ^2 分布 $\chi^2(n-1)$ に従う．

このU という統計量の標本分布から，不偏標本分散を用いた母分散の推定ができる（詳しくは付録 A.7 節参照）．なお，非正規母集団分布（一部例外を除く）を仮定する場合でも，U は漸近的に $\chi^2(n-1)$ に従うことが知られている．

7.5　標本平均の標本分布

ここでの最終的な目的は，標本平均の標本分布について理解することである．標本分布がどのような分布になるかは，**母集団分布が正規分布か否か，母分散が既知か否か，標本サイズが十分に大きいか否かにかかっている**．これらのことに注意しよう．

7.5.1　母集団分布が正規分布で母分散が既知の場合

正規分布の再生性の定理 7.1 より，$X_1, X_2, \cdots, X_n$ が独立で，すべて同一の正規分布 $N(\mu, \sigma^2)$ に従うとき，標本平均の標準化得点を

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

とすると，**$\boldsymbol{Z_n}$ は標準正規分布 $\boldsymbol{N(0,1)}$ に従う．このことは，標本サイズの大小にかかわらず成立する．**

7.5.2　母集団分布が正規分布以外で母分散が既知の場合

6 章で導入した相加平均についての中心極限定理（定理 6.7）より，標本 $X_1, X_2, \cdots, X_n$ が独立で，平均 μ 分散 σ^2 をもつ同一の分布に従っているとき，**$\boldsymbol{n}$ が十分に大きいとき**[*6]**，標本平均の標準化得点 $\boldsymbol{Z_n}$ は近似的に標準正規分布 $\boldsymbol{N(0,1)}$ に従う．**このような性質を「漸近正規性」ともいう．

[*6] 蓑谷（1994: 279）は「経験則として 30 以上あれば大標本と見なすことができる」としている．

n が小さく中心極限定理が適用できない場合，Z_n の標本分布は不明である．

7.5.3　母集団分布が正規分布で母分散が未知の場合

実際の調査データを分析する場合，母分散が既知であることはむしろ希である．母分散が未知の場合，標本平均の標準化得点を計算することができない．そこで，母分散を不偏標本分散 S^2 で置き換えた統計量

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \tag{7.7}$$

を定義する．この統計量を **t 統計量** (t statistic) と呼ぶ．

t 統計量は以下に定義する t 分布に従うことが知られている．

定義 7.2. Z を標準正規分布 $N(0,1)$ に従う確率変数，Y を自由度 k の χ^2 分布 $\chi^2(k)$ に従う確率変数であるとする．また，Z と Y は独立であるとする．このとき確率変数

$$t = \frac{Z}{\sqrt{Y/k}}$$

が従う確率分布を，自由度 k の **t 分布** *(t distribution)* といって，$t(k)$ と表す．

t 分布表はたいていの統計の教科書に載っているので，必要に応じて参照すればよい．

定理 7.4. 正規分布 $N(\mu, \sigma^2)$ に従う母集団から無作為に抽出された，大きさ n の標本 $X_1, X_2, \cdots, X_n$ について，t 統計量は自由度 $n-1$ の t 分布に従う．

証明. 標本平均の標準化得点

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

は標準正規分布に従う．また，定理 7.3 より，

$$U = \frac{(n-1)S^2}{\sigma^2}$$

は自由度 $n-1$ の χ^2 分布に従う．ゆえに，

$$\frac{Z_n}{\sqrt{U/(n-1)}}$$

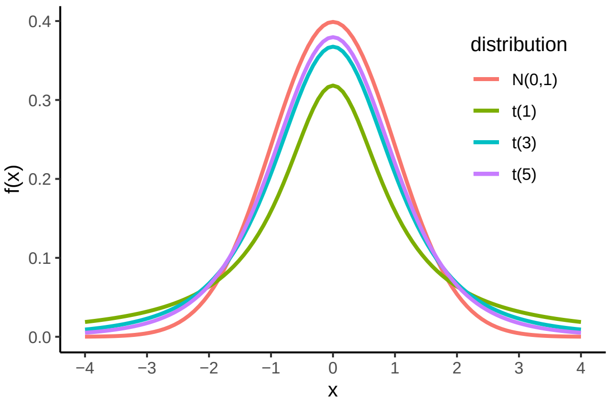

図 7.4　t 分布の例

は自由度 $n-1$ の t 分布に従う．これを変形すると，

$$\frac{Z_n}{\sqrt{U/(n-1)}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ (n-1)} = \frac{\sqrt{n}(\bar{X}-\mu)}{S}$$

となり，式 (7.7) を得る．Z_n と U の独立性は証明なしで認めることとする．　□

t 分布 $t(k)$ は自由度 k が大きくなればなるほど，標準正規分布 $N(0,1)$ に近づき，$k \to \infty$ のとき $N(0,1)$ に一致することが知られている．つまり，$\boldsymbol{n}$ が十分に大きいとき，$\boldsymbol{t}$ 統計量は近似的に標準正規分布 $\boldsymbol{N(0,1)}$ に従うと見なしてよい．一方，$\boldsymbol{n}$ が小さいときは，$\boldsymbol{t}$ 統計量は $\boldsymbol{t}$ 分布に従う[*7]．

7.5.4　母集団分布が正規分布以外で母分散が未知の場合

結論だけ述べると，大標本の場合中心極限定理が援用できるので，$\boldsymbol{n}$ が十分に大きいとき，$\boldsymbol{t}$ 統計量は近似的に標準正規分布 $\boldsymbol{N(0,1)}$ に従うと見なしてよい．しかしながら，$\boldsymbol{n}$ が小さいときは，$\boldsymbol{t}$ 統計量が従う分布は不明である．

[*7] 母分散が未知のとき，標準正規分布に比べて t 分布を用いると，一般的により保守的な推定・検定結果が導かれるため，大標本の場合でも t 分布に従うとして分析してもよい．

ここまでの考察結果を表にまとめると以下のようになる（蓑谷 1994: 278–9）．ただし，$\sim$ は確率変数がある確率分布に従うこと，$\approx$ は大標本の元で近似的に従うことを意味する．

表 7.1　標本平均の標本分布

母分散 σ^2	n	母集団分布	
		正規分布	非正規分布
既知	大	$Z_n \sim N(0,1)$	$Z_n \approx N(0,1)$
	小	$Z_n \sim N(0,1)$	不明
未知	大	$t \approx N(0,1)$	$t \approx N(0,1)$
	小	$t \sim t(n-1)$	不明

課題 7.1. *LUNA* のテストを *11/11*（水）*23:59* までに受けなさい．

第 8 章

統計的推定 (1): 点推定

8.1　推定量

前章では，母集団分布が分かっているとき，またあるいは，母集団分布が不明であっても，標本サイズ n が十分大なるときに母数は既知であることを前提として，標本分布を求めた．例えば，標本平均の標本分布の場合，常に母平均 μ は既知であることが前提とされていたわけである．このとき標本分布は，経験的に言うと，ある母数を前提としたときに，何度も何度も標本抽出を行い統計量を計算して分布を作ると，どういう形になるかということを示している．

しかしながら，調査結果を分析するときは普通，標本はただ 1 回だけ抽出される．ただ 1 回だけ抽出された標本を元に，母集団の特性，特に母数を推測することが求められる[*1]．そのための枠組みの一つが**統計的推定** (statistical estimation) である．

母数を推定するために標本から求められた統計量を**推定量** (estimator) という．母数を一般的に θ で表すとき，この推定量を $\hat{\theta}$ で表す．推定量は母数の推定に用いられる統計量であるので，一般の統計量と同じく，標本 $X_1, X_2, \cdots, X_n$ の関数

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$$

である．標本の各々は確率変数であるので，推定量 $\hat{\theta}$ もまた確率変数である[*2]．標本の各々が，ある実現値 $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ をとるとき，これを推定量に代入して得られる値のことを**推定値** (estimate) という．

推定には，推定値によって，母数を推定する**点推定** (point estimation) と，真の母

[*1] これを筆者は個人的に「統計的推測の根本問題」と呼んでいる．
[*2] プルプルしている．

数が入っている確からしい区間を求める**区間推定** (interval estimation) の 2 種類がある．本章では，まず点推定に焦点を当てる．

8.2　母平均・母分散の点推定

ここまでの議論から，母平均・母分散の点推定は以下のようにするのが一般的である．

8.2.1　母平均の推定

母平均 μ は標本平均 $\bar{X}$ によって推定する．

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

8.2.2　母分散の推定

母分散 σ^2 は，不偏標本分散 S^2 によって推定する．

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

8.2.3　不偏推定量

標本平均の平均（期待値）は，

$$E(\bar{X}) = \mu,$$

不偏標本分散の平均（期待値）は，

$$E(S^2) = \sigma^2$$

であった．このことは，標本平均・不偏標本分散は，それぞれ推定したい母数を平均的に偏りなく言い当てていることを示している．推定値の平均が母数と一致するとき，この推定量を**不偏推定量** (unbiased estimator) という．

8.2.4　標準誤差

点推定値が，どの程度正確かを示すために，推定値の分布（標本分布）のばらつきを情報として示すことがある．その際，点推定値の標準偏差が一般的に用いられる．これを**標準誤差** (santard error) といって s.e. と略す．

母平均の点推定の場合，母分散が既知の場合は

$$\sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

が標準誤差である．母集団が未知の場合は σ を S で置き換え，$S/\sqrt{n}$ を標準誤差の近似値として用いる．

一般的に分布のばらつきの指標である**標準偏差** (standard deviation: s.d.) と，標本分布のばらつきの指標である標準誤差 (s.e.) は，**ぜんぜん別物である**．標準偏差はたんに分布のばらつきを示したいときに使い，標準誤差は点推定の精度を示したいときに使う．間違って使うと，地球がその矛盾に耐えきれずに大爆発を起こすので，心して使って欲しい．

8.3　最尤法

一般的には，より複雑な統計モデルにおける母数を求める場合や，母集団分布が特殊な分布の時にも，それらの母数をきちんと推定量を求められる，一般的な手続きがあれば便利である．そのような方法としてはモーメント法と最小二乗法，最尤法がある．

モーメント法 (moment estimation) は，母平均を相加平均で推定する方法を一般化した方法である．付録 A.5 節にて解説しているので関心のある人はフォローして欲しい．

最小二乗法 (least square estimation) は，各データとの距離の二乗和が最小になるように推定値を選択する方法であり，回帰分析の回帰係数の推定の際に用いられる．

最尤法 (maximum likelihood estimation) は，確率的に最も尤もらしい推定値を選択する方法であり，多変量解析における基本的な推定法になっている．

以下では，最尤法のさわりを導入して，その雰囲気を味わってもらおう[*3].

8.3.1　尤度関数

ここで母集団分布を決める母数を一般に複数ある場合も含めて，$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ とする．母集団からの標本抽出が行われた段階では，これらの母数は確率的に決まっているのではなく，所与の値として与えられている．つまり，母数は確率変数である標本 X_i の前提条件となっており，標本として任意の値が抽出される確率 $P(X_i = x_i)$ は厳密には，$\boldsymbol{\theta}$ が起こったときの条件付き確率

$$P(X_i = x_i | \boldsymbol{\theta})$$

なのである．そこで，標本 X_i の従う確率密度関数を条件付き確率密度関数 $f(x_i|\boldsymbol{\theta})$ として正確に書くことにする．

このとき，標本 $X_1, X_2, \cdots, X_n$ は，（条件付き）同時確率密度関数

$$f(x_1, x_2, \cdots, x_n|\boldsymbol{\theta}) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n|\boldsymbol{\theta})$$

に従うことになるが，標本は同一の（条件付き）確率密度関数に従い，標本間は相互に独立であるから，

$$f(x_1, x_2, \cdots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) = f(x_1|\boldsymbol{\theta}) \times f(x_2|\boldsymbol{\theta}) \times \cdots \times f(x_n|\boldsymbol{\theta}) \qquad (8.1)$$

が成立する[*4].

調査データの分析段階では，われわれは標本の実際の値（データ）$x_1, x_2, \cdots, x_n$ を得ており，そこから未知の母数 $\boldsymbol{\theta}$ を推定量によって推定しようとしている．そこで，データ $x_1, x_2, \cdots, x_n$ を所与として，母数 $\boldsymbol{\theta}$ をいろいろと動かしたとき，どのような母数が**尤もらしい** (likely) かを調べるような関数を考える．このような関数を**尤**

[*3] 最尤法の知識は，社会調査士の D 科目としては必須ではないが，より実践的な統計解析には欠かせない知識なので，ここではあえて紹介している．数学的にややレベルが高いので，一読してよく理解できなくても気にしなくてよい．今後，もし高度な統計と関わることがあれば思いだしてもらう程度でよい．

[*4] 記号 $\prod$ は $\sum$ 記号のかけ算版でプロダクトと読む．$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times \cdots \times x_n$ である．

度関数 (likelihood function) といい，

$$L(\boldsymbol{\theta}|x_1, x_2, \cdots, x_n) = f(x_1, x_2, \cdots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \tag{8.2}$$

と定義する．尤度関数 $L(\boldsymbol{\theta}|x_1, x_2, \cdots, x_n)$ は，結局 $\boldsymbol{\theta}$ を条件としたときの，$x_1, x_2, \cdots, x_n$ の実現確率と等しくなるが，L それ自体は $\boldsymbol{\theta}$ の実現確率を示しているわけではないことに注意せよ．$\boldsymbol{\theta}$ はすでに決まっていることなので，決して確率的には動かない．ただデータからどのような値が尤もらしいのかを推測しているのである．

そして，尤度関数の値を最も高くするような $\boldsymbol{\theta}$ の値を，「最も尤もらしい推定値」なので，**最尤推定値** (maximum likelihood estimate) と呼ぶ．また，実現値を確率変数 $X_1, X_2, \cdots, X_n$ に置き換えて，確率変数の関数の形にしたものを**最尤推定量** (maximum likelihood estimator) という．推定量 $\hat{\theta}$ が最尤推定量であることを特に強調する場合には，$\hat{\theta}_{\mathrm{ML}}$ などと表記する．

最尤推定量の計算に当たっては，計算を簡単にするために，単調増加関数である対数関数によって尤度関数を変換することが多い[*5]．これを対数尤度関数

$$\log L(\boldsymbol{\theta}) = \log\left(L(\boldsymbol{\theta}|x_1, x_2, \cdots, x_n)\right) = \sum_{i=1}^{n} \log f(x_i|\boldsymbol{\theta}) \tag{8.3}$$

という．この関数の $\boldsymbol{\theta}$ についての最大化問題を解くことによって最尤推定値，そして最尤推定量を得る．

8.3.2　ベルヌーイ分布の母数 p の最尤推定量

ベルヌーイ分布 $Bi(1, p)$ に従い，1 か 0 の値をとる標本 $X_1, X_2, \cdots, X_n$ について，その条件付き確率密度関数は

$$f(x_1, x_2, \cdots, x_n|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

[*5] 人間でもコンピュータでもかけ算より足し算の方が簡単である．対数は値の大小関係を保ったままかけ算を足し算に変えるため，数値計算においてしばしば重要な役割を果たす．

である．成功回数の確率変数を $X = X_1 + X_2 + \cdots + X_n$ とし，その実現値を $x = x_1 + x_2 + \cdots + x_n$ とすると，

$$f(x_1, x_2, \cdots, x_n | p) = p^x (1-p)^{n-x}$$

となる．ゆえに，母数 p の尤度関数は

$$L(p | x_1, x_2, \cdots, x_n) = p^x (1-p)^{n-x}$$

であり，対数尤度関数は

$$\log L(p | x_1, x_2, \cdots, x_n) = x \log p + (n-x) \log(1-p)$$

となる．このとき p について関数を最大化する 1 階の必要条件は，p で微分して 0 と等しくなればよいので，

$$\frac{\partial \log L}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = 0$$

より，

$$\frac{x}{p} = \frac{n-x}{1-p} \Longleftrightarrow p = \frac{x}{n}$$

である[*6]．実現値を確率変数に戻して推定量の形にすると，結局，p の最尤推定量 $\hat{p}_{\mathrm{ML}}$ は

$$\hat{p}_{\mathrm{ML}} = \hat{p} = \frac{X}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

となる．成功回数 X が与えられたときの二項分布 $Bi(n, p)$ の母数 p の最尤推定量もこれと同一になる．

また，中心極限定理より，$\hat{p}_{\mathrm{ML}}$ の標本分布は平均 p，分散 $p(1-p)/n$ の正規分布に近似する[*7]．$\hat{p}_{\mathrm{ML}}$ の標準誤差は $\sqrt{\hat{p}_{\mathrm{ML}}(1-\hat{p}_{\mathrm{ML}})/n}$ に近似する．

---

*6 細かくは，2 階の十分条件，つまり 2 階微分して極点を代入したときに負となるという条件も確認しなければならない．

$$\left. \frac{\partial^2 \log L}{\partial p^2} \right|_{p=\frac{x}{n}} = -\frac{n^3}{(n-x)x} < 0$$

なので，$0 < p < 1$ のとき，$p = x/n$ は極大値である．また，$x = 0$ のとき，$p = 0, L(0) = 1$．このとき，$\partial \log L / \partial p = -n/(1-p) < 0$ であるので $p = 0$ が最大点．また $x = n$ のとき，$p = 1, L(1) = 1$，$\partial \log L / \partial p = n/p > 0$ であるので，$p = 1$ が最大点になる．

*7 ベルヌーイ分布の平均は p，分散は $p(1-p)$ であったことに注意．

■コイン投げの確率の最尤推定　表 1，裏 0 としてコインを 10 回投げたとき，データ（実現値）として

$$(0, 1, 1, 1, 1, 0, 1, 1, 0, 1)$$

が得られたとする．表の出現確率である母数 p の尤度関数は

$$L(p|0, 1, 1, 1, 1, 0, 1, 1, 0, 1) = p^7(1-p)^3$$

であり（図 8.1），対数尤度関数は

$$\log L(p|0, 1, 1, 1, 1, 0, 1, 1, 0, 1) = 7\log p + 3\log(1-p)$$

である（図 8.2）．p の最大化問題を解くと，

$$\hat{p}_{\mathrm{ML}} = \frac{7}{10} = 0.7$$

を得る．

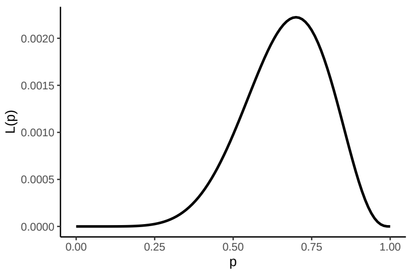

図 8.1　尤度関数 $L(p)$

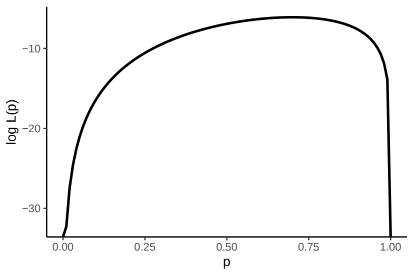

図 8.2　対数尤度関数 $\log L(p)$

正規分布の母数 μ, σ^2 の最尤推定量については付録 A.6 節を参照のこと．

8.4　点推定の基準

ここまで，いくつかの推定法のうち最尤法について見てきた．では，数ある推定量の候補の中で，どのような推定量がより望ましいと判断されるのであろうか．こうした推定量の望ましさの基準についても，統計理論の発展の中でさまざまに検討されている．こうした基準について簡単に触れて，最後に最尤法の利点について確認する．

8.4.1　不偏性

母数 θ の推定量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ は確率変数であるので確率分布を持つ．これを標本分布というのであった．さて，この $\hat{\theta}$ について，

$$E(\hat{\theta}) = \theta \tag{8.4}$$

が成立するとき，この推定量は**不偏** (unbiased) であるといい，このような推定量を**不偏推定量** (unbiased estimator) という．言葉で言えば，推定量の平均（期待値）が母数に一致するということである．また，$E(\hat{\theta}) - \theta$ を推定量 $\hat{\theta}$ の偏り (bias) と呼ぶが，不偏性を満たさないとき，$\hat{\theta}$ はプラスもしくはマイナスの偏りを持ち，平均的に θ を過大推定もしくは過小推定する．

8.2 節ですでに検討したように，標本平均は母平均の，不偏標本分散は母分散の，それぞれ不偏推定量である．

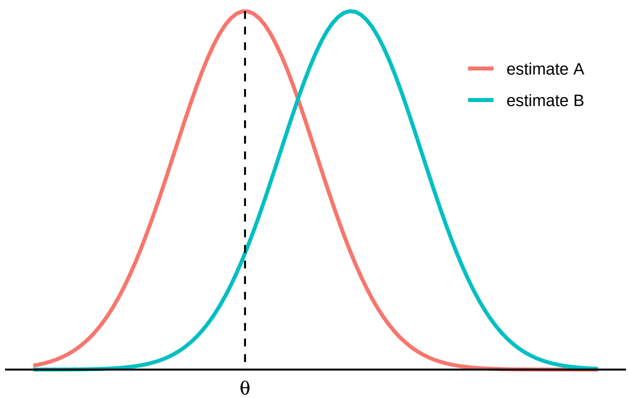

図 8.3　θ の不偏推定量 A の分布と不偏でない推定量 B の分布

8.4.2　有効性

母数 θ の不偏推定量がいくつかの種類存在するとき，そのなかでも分散が小さくなる推定量を選んだ方が推定の精度は高くなる．これが有効性もしくは効率性 (efficiency) の基準である．とくに，母数 θ の不偏推定量のなかで，もっとも分散の小さくなる推定量を**有効推定量** (efficient estimator)，あるいは**最小分散不偏推定量** (minimum variance unbiased estimator) と呼ぶ．有効推定量かどうかの判定には，やや高度な数学を用いなければならないので，ここでは詳細は省く．例えば，母集団分布が正規分布 $N(\mu, \sigma^2)$ である場合，標本平均は母平均の有効推定量であることが知られている．

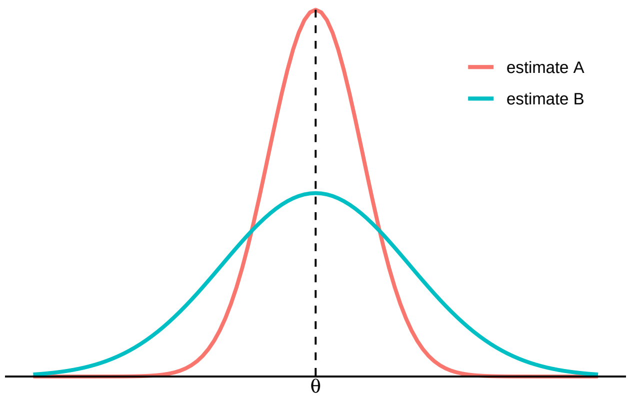

図 8.4　θ の有効推定量 A の分布と有効でない推定量 B の分布

8.4.3　一致性

以下で導入する一致性，漸近正規性，漸近有効性は大標本における特性である．

標本サイズ n が大きくなるに従い，推定量 $\hat{\theta}_n$ が母数 θ に近づくとき，つまり，$\hat{\theta}_n$

が θ に確率収束し，

$$\operatorname*{plim}_{n\to\infty} \hat{\theta}_n = \theta \tag{8.5}$$

が成り立つとき，推定量 $\hat{\theta}_n$ を**一致推定量** (consistent estimator) という．大数の法則（定理 6.5）より，標本平均は母平均の一致推定量である．また，証明は省略するが，不偏標本分散もまた母分散の一致推定量であることが知られている．特殊なケースを除いて一般的に，モーメント法・最尤法で求めた推定量は一致性をもつことが知られている．

8.4.4　漸近正規性

$n \to \infty$ のとき，標本分布が正規分布に近似する特性を**漸近正規性** (asymptotic normality) というのであった．この性質を満たす推定量を**漸近正規推定量** (asymptotically normal estimator) と呼ぶ．

平均についての中心極限定理（定理 6.7）より，標本平均は母平均の漸近正規推定量である．特殊なケースを除いて一般的に，モーメント法・最尤法で求めた推定量は漸近正規性をもつことが知られている．

8.4.5　漸近有効性

標本分布が漸近的に正規分布になる推定量のうち，その漸近正規分布の分散（漸近分散という）が最小となる推定量を**漸近有効推定量** (asymptotically efficient estimator) という．モーメント法による推定量は漸近有効性を持たないが，特殊なケースを除いて一般的に，最尤法で求めた推定量は漸近有効性をもつことが知られている．

8.4.6　最尤法の利点

最尤推定量が他の種類の推定量と比べて優れているのは，有効性を持つ場合が多いこと，さらには漸近的には（特殊なケースを除いて）必ず漸近有効性を持つことである[*8]．

[*8] 最尤法の特性について，さらに詳しくはホエール（1978: 198）や蓑谷（1994: 268-70）を参照のこと．

しかしながら，最尤法を用いる場合，推定には必ず母集団分布の分布型を指定する必要がある．しかしながら，仮定するモデルが複雑になると最尤推定量を解析的に解くことができなくなる．この場合，統計ソフトが提供する数値計算アルゴリズムを利用して推定値を求める．一方，モーメント法では分布型を仮定する必要がないため，推定に際して扱いやすいという側面もあるため，モーメント法をさらに一般化した手法 (GMM; generalized methods of moments) が用いられることもある．

課題 8.1. *LUNA* のテストを *11/18*（水）*23:59* までに受けなさい．

第 9 章

統計的推定 (2): 区間推定

前章では，母数を一点で推定してきたが，標本分布を活かして「母数が入っている可能性の高い区間」を推定することができる．このような方法を**区間推定** (interval estimation) という．

9.1 信頼区間

より具体的に区間推定とは，真の母数の値 θ が，ある区間 $[L, U]$ に入る確率を $1-\alpha \quad (0 \leq \alpha \leq 1)$ となるように保証する方法であり，

$$P(L \leq \theta \leq U) = 1 - \alpha \tag{9.1}$$

となる確率変数 L, U を求めるものである．$1-\alpha$ は**信頼係数** (confindence coefficient) と呼ばれ，このとき，区間 $[L, U]$ を $\mathbf{100(1-\alpha)\%}$ **信頼区間** (confidence interval) と呼ぶ[*1]．

調査データの分析段階で θ はすでに決まった定数であるので，θ が確率的に動くことは想定されていない[*2]．母数 θ ではなく，区間 $[L, U]$ が確率変数として標本抽出において動くと考える．

ゆえに信頼区間の経験的な解釈としては，「同じ母集団から繰り返し異なった標本をとって，その都度信頼区間を計算すると，繰り返し試行が十分に多ければ，θ を区間内に含むものの割合が $1-\alpha$ となる」というものである（図 9.1）．

[*1] 一般に，連続数直線上の点 $a, b (a < b)$ について，それら（端点）を含んだ区間を閉区間といって $[a, b]$ と表す．端点を含まない区間は開区間といって (a, b) と表す．開区間は順序対（ベクトル）表記と同じであるが，文脈上問題なければ特に断りなく両者の意味で使い分ける．

[*2] 母数を確率変数と仮定する統計の枠組みがベイズ統計学である．

しかしながら，ふつう標本抽出はただの 1 回限りである．たとえば，$\alpha = 0.05$ として，だいたい 100 回中 95 回が信頼区間に母数を含むとしても，今回のただ 1 回のケースがそれに該当するかどうかは分からない*3．たまたま，外れの 5 回のうちの 1 回が実現したのかもしれない．そのことは 1 回の標本抽出だけではどこまでいっても分からないが，「まあ，外れるといっても 100 回中高々 5 回なので大丈夫でしょう (外れてるかも知らんけど)」と誤りの余地を残して結論を導くのである．これが区間推定である．

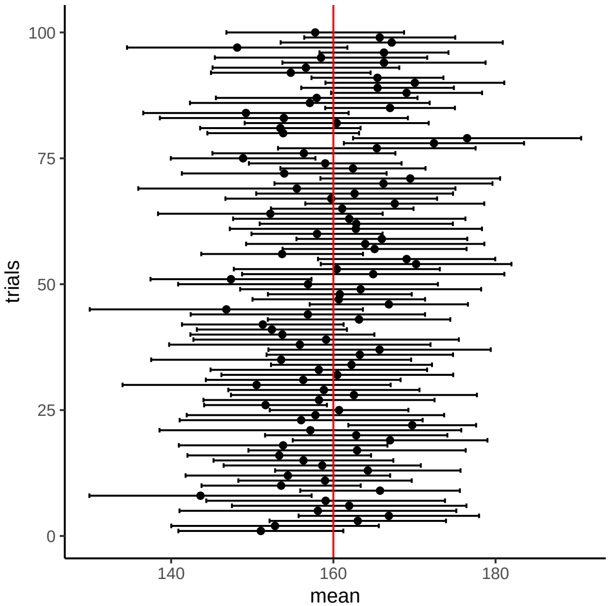

図 9.1　同一母集団からとった異なる標本の信頼区間（95% 信頼区間であれば，θ を含む信頼区間の割合は 0.95 に近くなる．）

*3 God only knows っちゅうやつですわ．

9.2　母平均の信頼区間

以下，母平均の信頼区間を計算する．母平均については，表 9.1 のように既に状況ごとに従う標本平均の標本分布が分かっている．母平均の信頼区間の計算に関して言えば，大きくは以下の 3 パターンに分けられる．典型的なのは (2) のケースである．

(1) 母分散 σ^2 が既知で，標本平均の標準化得点 Z_n が（漸近的に）標準正規分布 $N(0,1)$ に従う場合
(2) **母分散 σ^2 が未知で，標本平均の t 統計量が漸近的に標準正規分布 $N(0,1)$ に従う場合**
(3) 母分散 σ^2 が未知で，標本平均の t 統計量が t 分布 $t(n-1)$ に従う場合

このそれぞれについて具体的に見ていこう．

表 9.1　標本平均の標本分布（再掲）

母分散 σ^2	n	母集団分布：正規分布	母集団分布：非正規分布
既知	大	$Z_n \sim N(0,1)$	$Z_n \approx N(0,1)$
	小	$Z_n \sim N(0,1)$	不明
未知	大	$t \approx N(0,1)$	$t \approx N(0,1)$
	小	$t \sim t(n-1)$	不明

9.2.1　母分散 σ^2 が既知で，標本平均の標準化得点 Z_n が（漸近的に）標準正規分布 $N(0,1)$ に従う場合

標準正規分布 $N(0,1)$ において，それ以上の値が実現する確率が $\alpha/2$ となるような値を $z_{\alpha/2}$ とおく．すなわち，

$$P(z_{\alpha/2} \leq Z) = \alpha/2.$$

このとき，$N(0,1)$ に従う確率変数 Z について

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha \tag{9.2}$$

が成り立つ（図 9.2 参照）．$z_{\alpha/2}$ を「両側 $100\alpha\%$ 点」というように呼ぶことがある．

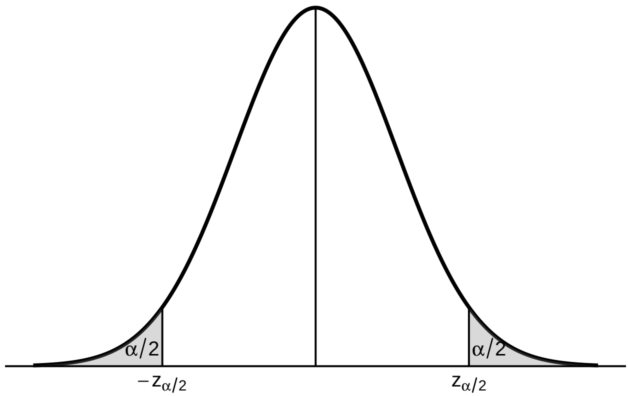

図 9.2　標準正規分布

7.1.1 節での知識を使うと，例えば，$\alpha = 0.05$ とすると，

$$P(-z_{0.025} = -1.96 \leq Z \leq z_{0.025} = 1.96) = 0.95$$

となる．あるいは，$\alpha = 0.01$ とすると，

$$P(-z_{0.005} = -2.576 \leq Z \leq z_{0.005} = 2.576) = 0.99$$

となる．

いま，母分散 σ^2 が既知で，標本平均の標準化得点

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

が（漸近的に）標準正規分布 $N(0,1)$ に従うことが分かっているとする．Z_n を式 (9.2) に代入すると，

$$P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq z_{\alpha/2}\right) = 1 - \alpha. \tag{9.3}$$

カッコ内の不等式を μ について解くと，

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \tag{9.4}$$

となる．ゆえに，95% 信頼区間は

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\quad \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right], \tag{9.5}$$

99% 信頼区間は

$$\left[\bar{X} - 2.576\frac{\sigma}{\sqrt{n}},\quad \bar{X} + 2.576\frac{\sigma}{\sqrt{n}}\right] \tag{9.6}$$

によって計算できる．

■例（平均身長の区間推定 1）　母集団分布が正規分布に従い，母分散が 100 センチであることが経験的に分かっているとする．このとき，標本サイズ 16 のデータから得られた標本平均が 165 センチであった．このときの，母平均の 95% 信頼区間は

$$\left[165 - 1.96\frac{10}{4},\quad 165 + 1.96\frac{10}{4}\right],$$

まとめると，$[160.1, 169.9]$ となる．95% の確からしさで，母平均 μ は区間 $[160.1, 169.9]$ に含まれると結論づけられる．

9.2.2　母分散 σ^2 が未知で，標本平均の t 統計量が漸近的に標準正規分布 $N(0,1)$ に従う場合

この場合，t 統計量

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$$

は漸近的に標準正規分布 $N(0,1)$ に従うので，式 (9.2) の Z に t を代入して，

$$P\left(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha \tag{9.7}$$

となる．ゆえに，95% 信頼区間は

$$\left[\bar{X} - 1.96\frac{S}{\sqrt{n}},\quad \bar{X} + 1.96\frac{S}{\sqrt{n}}\right], \tag{9.8}$$

99% 信頼区間は

$$\left[\bar{X} - 2.576\frac{S}{\sqrt{n}},\ \ \bar{X} + 2.576\frac{S}{\sqrt{n}}\right] \tag{9.9}$$

によって計算できる．

■例（平均身長の区間推定 2）　母分散は未知であるが，標本サイズは 900 と十分に大きく t 統計量は漸近的に標準正規分布に従っていると見なしてよいとする．このとき，標本平均が 165 センチ，不偏標本分散が 100 センチであることが分析結果から分かった．このときの，母平均の 95% 信頼区間は

$$\left[165 - 1.96\frac{10}{30},\ \ 165 + 1.96\frac{10}{30}\right],$$

まとめると，$[164.35, 165.65]$ となる．95% の確からしさで，母平均 μ は区間 $[164.35, 165.65]$ に含まれると結論づけられる．

9.2.3　母分散 σ^2 が未知で，標本平均の t 統計量が t 分布 $t(n-1)$ に従う場合

以下のケースは，社会心理学実験など小標本における推定として重要である．

自由度 $n-1$ の t 分布 $t(n-1)$ において，それ以上の値が実現する確率が $\alpha/2$ となるような値を $t_{\alpha/2}(n-1)$ とおく．このとき，$t(n-1)$ に従う確率変数 t について

$$P(-t_{\alpha/2}(n-1) \leq t \leq t_{\alpha/2}(n-1)) = 1 - \alpha \tag{9.10}$$

が成り立つ．この式の左辺カッコの中に

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$$

を代入してまとめると，

$$P\left(\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right) = 1 - \alpha \tag{9.11}$$

となる．ゆえに，$100(1-\alpha)$% 信頼区間は

$$\left[\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}},\ \ \bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right] \tag{9.12}$$

によって計算できる．

■例（平均身長の区間推定 3）　母集団分布は正規分布であるが母分散は未知であり，標本サイズは 16 と小さく中心極限定理は適用できない．このとき，標本平均が 165 センチ，不偏標本分散が 100 センチであることが分析結果から分かった．

$$t_{0.025}(15) = 2.131$$

である．このときの，母平均の 95% 信頼区間は

$$\left[165 - 2.131\frac{10}{4},\ \ 165 + 2.131\frac{10}{4}\right],$$

まとめると，$[159.67, 170.33]$ となる．95% の確からしさで，母平均 μ は区間 $[159.67, 170.33]$ に含まれると結論づけられる．

9.3　ベルヌーイ分布の母数 p についての区間推定

確率 p のベルヌーイ分布 $Bi(1,p)$ に従う独立な n 個の確率変数 $X_1, X_2, \cdots, X_n$ について考える．$X = \sum_{i=1}^{n} X_i$ として，

$$\bar{X} = \frac{X}{n} = \hat{p}, \qquad E(\bar{X}) = p, \qquad V(\bar{X}) = \frac{p(1-p)}{n}$$

である．試行回数 n が十分に大きい場合，中心極限定理より，$\hat{p} = \bar{X}$ を標準化した

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \tag{9.13}$$

は，漸近的に標準正規分布 $N(0,1)$ に従う[*4]．式 (9.2) に Z を代入してまとめると，

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha \tag{9.14}$$

[*4] X は二項分布 $Bi(n,p)$ に従う確率変数でもあるので，6 章でフォローした二項分布についての中心極限定理（定理 6.6）を直接適用して，

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

は，漸近的に標準正規分布 $N(0,1)$ に従うことから，右辺の分母分子を n で割って，

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

としても同様である．

となる．しかしながら，このままでは信頼区間の両端の式に母数 p が含まれており信頼区間が計算できない．そこで p を $\hat{p}$ で置き換えて近似計算すると，

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha \tag{9.15}$$

を得る．このときの，母数 p の 95% 信頼区間は

$$\left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \ \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right], \tag{9.16}$$

99% 信頼区間は

$$\left[\hat{p} - 2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \ \hat{p} + 2.576\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \tag{9.17}$$

によって計算できる．

■例（支持率の区間推定）　無作為抽出調査を実施し対象者に現内閣への支持・不支持を尋ねた．支持を 1，不支持を 0 とする確率変数 X_i は，真の支持率 p を母数とするベルヌーイ分布 $Bi(1,p)$ に従う[*5]．100 人の標本において支持者が 60 人であった．このとき，真の支持率 p の 95% 信頼区間は

$$\left[0.6 - 1.96\sqrt{\frac{0.6 \times 0.4}{100}},\ \ 0.6 + 1.96\sqrt{\frac{0.6 \times 0.4}{100}}\right],$$

まとめると，$[0.504, 0.696]$ となる．標本支持率が変わらず 0.6 で標本サイズが 1000 の場合の 95% 信頼区間は，区間 $[0.570, 0.630]$ になり，より予測精度が向上することが確認できる．

そのほか，母分散 σ^2 の区間推定については，付録 A.7 節を参照のこと．

課題 9.1. *LUNA* のテストを *11/25*（水）*23:59* までに受けなさい．

[*5] あるいは $X = \sum_{i=1}^{n} X_i$ は二項分布 $Bi(n,p)$ に従う．

第 10 章

仮説検定 (1): 仮説検定の考え方

本章から本講義の最終目的である仮説検定に入る．まず，本章では仮説検定の考え方を具体例に則して導入する．

10.1　仮説

統計学で言う**仮説検定** (hypothesis testing) とは，母集団について仮定された命題を，標本を調べることで検証する方法である．以下説明のための例として，

> 母分散が既知である正規母集団分布 $N(\mu, \sigma^2)$ から，大きさ n の標本を得たときに，母平均 μ の特性について検定する

ことを考える．

数値例

母分散が 100 である身長の正規母集団分布 $N(\mu, 100)$ から，大きさ 100 の標本を得たときに，母平均 μ の特性について検定する．

検証の対象となる仮説を**帰無仮説** (null hypothesis) といい，記号 H_0 で表す．多くの場合，仮説検定でターゲットになるのは母数がどのような値をとるか，ということである．例えば，母平均 μ がある値 μ_0 であるという仮説を帰無仮説とするとき，

$$H_0: \ \mu = \mu_0$$

などと表す．帰無仮説と対立する仮説を**対立仮説** (alternative hypothesis) といっ

て，H_1 で表す．例えば，先の H_0 に対する可能な対立仮説の 1 つは，

$$H_1: \ \mu \neq \mu_0$$

である．このとき，関心の対象は母平均が値 μ_0 であるか，否か，ということである．このような形式の対立仮説を**両側対立仮説** (two-sided alternative hypothesis) といい，このときの仮説検定を特に**両側検定** (two-sided test) という．

一方，母平均がある値 μ_0 であるか，それを上回る（下回る）かということに興味がある場合，対立仮説を

$$H_1': \ \mu > \mu_0$$

とすることがある．このような形式の対立仮説を**片側対立仮説** (one-sided alternative hypothesis) といい，このときの仮説検定を**片側検定** (one-sided test) という．また，特に上述のような「上回る」という対立仮説を「右」片側対立仮説という．逆に $H_1'': \ \mu < \mu_0$ のように，「下回る」という対立仮説を「左」片側対立仮説という．

母平均に関する仮説検定など，多くの場合仮説検定は，**帰無仮説が正しいと仮定したときに得られる標本の特性と，実際の標本の特性がほとんど矛盾することを導きだし，その結果より帰無仮説を棄却して対立仮説を採用する**という論証プロセスをとる[*1]．この論法は，ある命題が真であることを主張するために，その否定命題を検証し矛盾を導き出すことで，元の命題が真であることを主張する**背理法**の統計版である．

数値例

$$H_0: \ \mu = 170, \qquad H_1: \ \mu \neq 170$$

つまり，母平均は（何らかの基準である）170cm であると帰無仮説を立てて，それを棄却して対立仮説の 170cm ではない，と言いたい．

[*1] 逆に，「帰無仮説が棄却されない」ということを主張するために，仮説検定を用いるケースもあるが，本章で検討するケースはすべて対立仮説の主張が真の目的である．

10.2　検定統計量

次に，調査で得られた標本をもとに，母数に関連する統計量を計算する．この統計量を特に**検定統計量** (test statistic) という．例えば，母平均 μ についての検定統計量としては，標本平均 $\bar{X}$ あるいはそれを変形した Z_n 統計量や t 統計量が用いられることが一般的である．そして，**帰無仮説が正しいと仮定したときの**検定統計量の標本分布を計算する．例の場合，標本平均の標本分布は帰無仮説が正しければ，正規分布 $N(\mu_0, \sigma^2/n)$ に従う．つまり，

$$\bar{X} \sim N(\mu_0, \sigma^2/n), \quad Z_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1).$$

また，実際の標本から得られた統計量の値を計算する．

数値例

$$\bar{X} \sim N(170, 1), \quad Z_n = \bar{X} - 170 \sim N(0, 1)$$

実際の標本から得られた標本平均は 173cm であったとする．つまり，Z_n の実現値は $z_n = 173 - 170 = 3$ である．

10.3　有意水準と棄却域

帰無仮説の下で得られる検定統計量の標本分布において，実際の標本統計量の値が得られる可能性がかなり低ければ，帰無仮説が誤りであった可能性が高いと言えるであろう．このとき帰無仮説を**棄却** (reject) し，対立仮説を**採択** (accept) する．帰無仮説が棄却された場合，検定は**有意** (significant) であるという．

帰無仮説を棄却するための基準をあらかじめ設定しておく．帰無仮説を棄却すべき統計量の値の集合を**棄却域** (rejection region) という．そして，棄却域内に値が入る確率を**有意水準** (significant level) という．帰無仮説となる母数の値から遠いところを棄却域に定めるが，両側検定の場合両側に，片側検定の場合その片側に棄却域を定

める．例えば，両側対立仮説の場合，

$$|Z_n| > z_{\alpha/2} \iff \left(\bar{X} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \vee \left(\mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}\right) \tag{10.1}$$

が有意水準 $100 \times \alpha\%$ の棄却域になる[*2]．つまり，

$$P(|Z_n| > z_{\alpha/2}|\mu = \mu_0) = \alpha \tag{10.2}$$

である[*3]．このときの $z_{\alpha/2}$ を棄却域の**境界値** (critical value) という．片側の場合，例えば H_1： $\mu > \mu_0$ とすると，有意水準 $100 \times \alpha\%$ の棄却域は

$$Z_n > z_\alpha \tag{10.3}$$

である．つまり，

$$P(Z_n > z_\alpha|\mu = \mu_0) = \alpha. \tag{10.4}$$

多くの場合，有意水準として 5%，もしくは 1% が採用される．両側検定で 5% の有意水準の場合，棄却域は

$$|Z_n| > 1.96 \iff \left(\bar{X} < \mu_0 - 1.96\frac{\sigma}{\sqrt{n}}\right) \vee \left(\mu_0 + 1.96\frac{\sigma}{\sqrt{n}} < \bar{X}\right)$$

となる．一方，片側検定で 5% の有意水準の場合，棄却域は

$$Z_n > 1.645 \iff \bar{X} > \mu_0 + 1.645\frac{\sigma}{\sqrt{n}}$$

となる（図 10.1, 図 10.2）．

実際の標本統計量の値が棄却域に「落ちる」とき，帰無仮説を棄却し帰無仮説を採択する．一方，標本統計量の値が棄却域に「落ちない」とき，帰無仮説は棄却されない．だからといって，帰無仮説の正しさが論証されたわけではないことに注意．「帰無仮説が間違っているということはない」という消極的な言明を得たにすぎないのである．

[*2] ここで記号 $\vee$ は，「または (or)」を表す．また，より正確には，棄却域は式 (10.1) を満たすような区間であるが，ここでは区間の条件と区間を同一視する．

[*3] この式の意味は「$\mu = \mu_0$ を前提条件として，Z_n の絶対値が境界値 $z_{\alpha/2}$ よりも大きくなる確率が α である」．

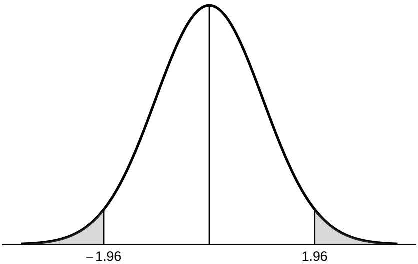

図 10.1　両側検定における棄却域の例（$Z_n \sim N(0,1)$ のときの有意水準 5% 両側検定）

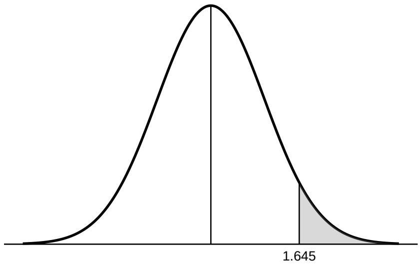

図 10.2　片側検定における棄却域の例（$Z_n \sim N(0,1)$ のときの有意水準 5% 右片側検定）

数値例

$z_n = 3 > 1.96$，または $173 > 170 + 1.96 \times 1$ であるので，標本統計量の値は両側 5% 水準の棄却域に落ちる．帰無仮説を棄却し対立仮説を採用する．つまり，(高々 5% の誤りの可能性はあるが) 母平均は 170cm ではないといえる．

ところで，式 (10.1) より，

$$\begin{aligned}|Z_n| > z_{\alpha/2} &\Longleftrightarrow \left(\bar{X} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \vee \left(\mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}\right) \\ &\Longleftrightarrow \left(\mu_0 > \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \vee \left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > \mu_0\right) \qquad (10.5)\end{aligned}$$

と変形できるので，$H_0 : \mu = \mu_0$ を帰無仮説とする有意水準 $100 \times \alpha$% の両側検定は，$100 \times (1-\alpha)$% 信頼区間の外側に μ_0 が位置するかどうかによっても判断することができる．つまり，

$$\mu_0 \notin \left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \qquad (10.6)$$

のとき[*4]，帰無仮説 H_0 は $100 \times \alpha$% 水準で棄却される．同様にして，$H_0 : \mu = \mu_0$

[*4] $a \notin A$ は，要素 a が集合（区間）A に含まれないことを意味する．

を帰無仮説とする有意水準 $100 \times \alpha\%$ の右片側検定の場合

$$\begin{aligned} Z_n > z_\alpha &\iff \bar{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \\ &\iff \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}} > \mu_0 \end{aligned} \tag{10.7}$$

となるので，$100 \times (1-2\alpha)\%$ 信頼区間の下限より下に μ_0 が位置するかどうかによって帰無仮説を棄却できるかどうかを判断することができる．左片側検定の場合は，$100 \times (1-2\alpha)\%$ 信頼区間の上限より上に μ_0 が位置するかどうかによって判断する．

数値例

母平均の 95% 信頼区間は $[173 - 1.96 \times 1, 173 + 1.96 \times 1] = [171.04, 174.96]$ である．帰無仮説 $\mu_0 = 170$ は 95% 信頼区間の外側にあるので，帰無仮説は 5% 水準で棄却できる．

10.4　p 値

有意水準は理論的には，検定に先立って定めておくものであるが，多くの統計のコンピュータ・パッケージは $\boldsymbol{p}$ **値** (p value) を出力する．p 値とは，検定統計量 T の値が t_r だったときに，t_r が境界値になるような有意水準 p の値のことである．具体的に，帰無仮説が $\theta = \theta_0$ のとき，対立仮説を $\theta \neq \theta_0$ とする両側検定においては，

$$P(|T| > t_r | \theta = \theta_0) = p$$

が p 値となる．$p < 0.05$ であれば有意水準 5% で，$p < 0.01$ であれば有意水準 1% で帰無仮説は棄却できることが分かる．

数値例

標準正規分布の性質から $P(|Z_n| > 3 | \theta = 170) = 0.0027$ である．$p < 0.05$ であるので，有意水準 5% で帰無仮説は棄却できる（有意水準 1% でも棄却できる）．

10.5　第 1 種の過誤と第 2 種の過誤

帰無仮説の棄却・棄却しないの決定については表 10.1 のように 4 つのパターンがある．

表 10.1　帰無仮説の棄却・棄却しないの決定

	H_0 を棄却	H_0 を棄却しない
H_0 が真	第 1 種の過誤	正しい判定
H_0 が偽	正しい判定	第 2 種の過誤

帰無仮説の棄却・棄却しないの決定に際しては，確率論的に不可避に 2 種類の誤りの可能性が発生する．1 つは，H_0 が真であるにもかかわらず，帰無仮説を棄却する誤りで，これを**第 1 種の過誤** (error of the first kind) という．第 1 種の過誤の確率は α そのものであり，有意水準を調整することによって第 1 種の過誤の確率をコントロールすることができる．一方，H_0 が偽である（その反対に H_1 が真である）にもかかわらず，H_0 を棄却しない（その結果 H_1 を採択しない）という誤りを，**第 2 種の過誤** (erro of the second kind) という．第 2 種の過誤の確率は α と n を所与とすると，コントロールすることはできない．

10.6　仮説検定のプロセス

まとめると，仮説検定は以下のプロセスをとる．

(1) 母集団の特性に関して，帰無仮説を設定する．同時に，帰無仮説に対立する対立仮説を設定する．
(2) 有意水準を決定する．
(3) 帰無仮説が正しいと仮定したときの，検定統計量の標本分布を導き出し，有意水準に対応する棄却域を求める．
(4) 実際の標本から統計量の値を計算する．
(5) その値が棄却域に落ちるときは，帰無仮説を棄却し対立仮説を採用する．一方，

棄却域外に落ちるときは，帰無仮説は棄却されない.

課題 10.1. *LUNA* のテストを *12/2*（水）*23:59* までに受けなさい.

第 11 章

仮説検定 (2): 母平均・母比率の検定

前章で仮説検定の概念的導入が終わったので，本章からは仮説検定の実践例をいくつか紹介する．

11.1　母平均についての仮説検定のパターン

母平均については信頼区間の計算と同様に，標本平均が従う標本分布（表 11.1）に応じて以下の 3 パターンに分けられる：

(1) 母分散 σ^2 が既知で，標本平均の標準化得点 Z_n が（漸近的に）標準正規分布 $N(0,1)$ に従う場合
(2) 母分散 σ^2 が未知で，標本平均の t 統計量が漸近的に標準正規分布 $N(0,1)$ に従う場合
(3) 母分散 σ^2 が未知で，標本平均の t 統計量が t 分布 $t(n-1)$ に従う場合

このそれぞれについて，検定法を具体的に見ていこう．このとき，帰無仮説を

$$H_0:\ \mu = \mu_0$$

とする．

11.1.1　母分散 σ^2 が既知で，標本平均の標準化得点 Z_n が（漸近的に）標準正規分布 $N(0,1)$ に従う場合

このケースの場合，

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

表 11.1　標本平均の標本分布（再々掲）

母分散 σ^2	n	母集団分布	
		正規分布	非正規分布
既知	大	$Z_n \sim N(0,1)$	$Z_n \approx N(0,1)$
	小	$Z_n \sim N(0,1)$	不明
未知	大	$t \approx N(0,1)$	$t \approx N(0,1)$
	小	$t \sim t(n-1)$	不明

が（漸近的に）標準正規分布 $N(0,1)$ に従う．検定に際しては，この Z_n を検定統計量とする．

■有意水準 5% の両側検定　対立仮説を

$$H_1:\ \mu \neq \mu_0$$

とする．

$$P(|Z_n| > 1.96) = 0.05 \tag{11.1}$$

であるので，

$$|Z_n| > 1.96 \Longleftrightarrow \left(\bar{X} < \mu_0 - 1.96\frac{\sigma}{\sqrt{n}}\right) \vee \left(\mu_0 + 1.96\frac{\sigma}{\sqrt{n}} < \bar{X}\right) \tag{11.2}$$

が棄却域であり，$\bar{X}$ が棄却域に落ちるとき，帰無仮説 H_0 を棄却し H_1 を採択する．$\bar{X}$ が棄却域に落ちないとき，帰無仮説 H_0 は棄却しない．

有意水準が 1% のときは，境界値として ± 1.96 の代わりに ± 2.576 を用いる．

■有意水準 5% の右片側検定　対立仮説を

$$H_1:\ \mu > \mu_0$$

とする．

$$P(Z_n > 1.645) = 0.05 \tag{11.3}$$

であるので，

$$Z_n > 1.645 \iff \bar{X} > \mu_0 + 1.645\frac{\sigma}{\sqrt{n}} \tag{11.4}$$

が棄却域であり，$\bar{X}$ が棄却域に落ちるとき，帰無仮説 H_0 を棄却し H_1 を採択する．$\bar{X}$ が棄却域に落ちないとき，帰無仮説 H_0 は棄却しない．

左片側検定については，棄却域を $Z_n < -1.645$ とすればよい．有意水準が 1% のときは，境界値として 1.645 の代わりに 2.326 を用いる．

■例（平均身長の検定 1）　母集団分布が正規分布に従い，母分散が 100 センチであることが経験的に分かっているとする．このとき，標本サイズ 16 のデータから得られた標本平均が 165 センチであった．このとき

$$H_0: \ \mu = 160, \quad H_1: \ \mu \neq 160$$

として 5% 水準，1% 水準それぞれで検定を行う．

$$Z_n = \frac{4(165-160)}{10} = 2$$

である．5% 水準の両側検定の場合，棄却域は

$$|Z_n| > 1.96 \iff \left(\bar{X} < 160 - 1.96\frac{10}{4} = 155.1\right) \vee \left(160 + 1.96\frac{10}{4} = 164.9 < \bar{X}\right)$$

であるので，5% 水準では帰無仮説は棄却され，検定は有意となる．一方，1% 水準の両側検定の場合，棄却域は

$$|Z_n| > 2.576 \iff \left(\bar{X} < 160 - 2.576\frac{10}{4} = 153.56\right) \vee \left(160 + 2.576\frac{10}{4} = 166.44 < \bar{X}\right)$$

であるので，1% 水準では帰無仮説は棄却されず，検定は有意ではない．

次に，

$$H_0: \ \mu = 160, \quad H_1': \ \mu > 160$$

として 5% 水準，1% 水準それぞれで検定を行う．

5% 水準の右片側検定の場合，棄却域は

$$Z_n > 1.645 \iff \bar{X} > 160 + 1.645\frac{10}{4} = 164.11$$

であるので，5% 水準では帰無仮説は棄却され，検定は有意となる．一方，1% 水準の右片側検定の場合，棄却域は

$$Z_n > 2.326 \Longleftrightarrow \bar{X} > 160 + 2.326\frac{10}{4} = 165.82$$

であるので，1% 水準では帰無仮説は棄却されず，検定は有意ではない．

11.1.2　母分散 σ^2 が未知で，標本平均の t 統計量が漸近的に標準正規分布 $N(0,1)$ に従う場合

この場合，t 統計量

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$$

は漸近的に標準正規分布 $N(0,1)$ に従う．t 統計量を検定統計量とする．

■有意水準 5% の両側検定　対立仮説を

$$H_1 : \ \mu \neq \mu_0$$

とする．

$$P(|t| > 1.96) = 0.05 \tag{11.5}$$

であるので，

$$|t| > 1.96 \Longleftrightarrow \left(\bar{X} < \mu_0 - 1.96\frac{S}{\sqrt{n}}\right) \vee \left(\mu_0 + 1.96\frac{S}{\sqrt{n}} < \bar{X}\right) \tag{11.6}$$

が棄却域である．有意水準が 1% のときは，境界値として ± 1.96 の代わりに ± 2.576 を用いる．

■有意水準 5% の右片側検定　対立仮説を

$$H_1 : \ \mu > \mu_0$$

とする．

$$P(t > 1.645) = 0.05 \tag{11.7}$$

であるので，

$$t > 1.645 \Longleftrightarrow \bar{X} > \mu_0 + 1.645\frac{S}{\sqrt{n}} \tag{11.8}$$

が棄却域である．

左片側検定については，棄却域を $t < -1.645$ とすればよい．有意水準が 1% のときは，境界値として 1.645 の代わりに 2.326 を用いる．

■例（平均身長の検定 2）　母分散は未知であるが，標本サイズは 900 と十分に大きく t 統計量は漸近的に標準正規分布に従っていると見なしてよいとする．このとき，標本平均が 165 センチ，不偏標本分散が 100 センチであることが分析結果から分かった．このとき H_0：$\mu = 160$,　H_1：$\mu \neq 160$ として 5% 水準，1% 水準それぞれで検定を行う．t 統計量は，

$$t = \frac{30(165-160)}{10} = 15$$

である．

5% 水準の両側検定の場合，棄却域は

$$|t| > 1.96 \Longleftrightarrow \left(\bar{X} < 160 - 1.96\frac{10}{30} = 159.35\right) \vee \left(160 + 1.96\frac{10}{30} = 160.65 < \bar{X}\right)$$

であるので，5% 水準では帰無仮説は棄却され，検定は有意となる．一方，1% 水準の両側検定の場合，棄却域は

$$|t| > 2.576 \Longleftrightarrow \left(\bar{X} < 160 - 2.576\frac{10}{30} = 159.14\right) \vee \left(160 + 2.576\frac{10}{30} = 160.86 < \bar{X}\right)$$

であるので，1% 水準でも帰無仮説は棄却され，検定は有意となる．

次に，H_0：$\mu = 160$,　H_1'：$\mu > 160$ として 5% 水準，1% 水準それぞれで検定を行う．

5% 水準の右片側検定の場合，棄却域は

$$|t| > 1.645 \Longleftrightarrow \bar{X} > 160 + 1.645\frac{10}{30} = 160.55$$

であるので，5% 水準では帰無仮説は棄却され，検定は有意となる．一方，1% 水準の両側検定の場合，棄却域は

$$|t| > 2.326 \Longleftrightarrow \bar{X} > 160 + 2.326\frac{10}{30} = 160.78$$

であるので，1% 水準でも帰無仮説は棄却され，検定は有意となる．

11.1.3　母分散 σ^2 が未知で，標本平均の t 統計量が t 分布 $t(n-1)$ に従う場合

この場合，t 統計量

$$t = \frac{\sqrt{n}(\bar{X}-\mu)}{S}$$

は自由度 $n-1$ の t 分布 $t(n-1)$ に従う．t 統計量を検定統計量とする．

■有意水準 5% の両側検定　対立仮説を

$$H_1:\ \mu \neq \mu_0$$

とする．

$$P(|t| > t_{0.025}(n-1)) = 0.05 \tag{11.9}$$

であるので，

$$|t| > t_{0.025}(n-1) \iff \left(\bar{X} < \mu_0 - t_{0.025}(n-1)\frac{S}{\sqrt{n}}\right) \vee \left(\mu_0 + t_{0.025}(n-1)\frac{S}{\sqrt{n}} < \bar{X}\right) \tag{11.10}$$

が棄却域である．

有意水準が 1% のときは，境界値として $\pm t_{0.025}(n-1)$ の代わりに $\pm t_{0.005}(n-1)$ を用いる．

■有意水準 5% の右片側検定　対立仮説を

$$H_1:\ \mu > \mu_0$$

とする．

$$P(t > t_{0.05}(n-1)) = 0.05 \tag{11.11}$$

であるので，

$$t > t_{0.05}(n-1) \iff \bar{X} > \mu_0 + t_{0.05}(n-1)\frac{S}{\sqrt{n}} \tag{11.12}$$

が棄却域である．

左片側検定については，棄却域を $t < -t_{0.05}(n-1)$ とすればよい．有意水準が 1% のときは，境界値として $t_{0.05}(n-1)$ の代わりに $t_{0.01}(n-1)$ を用いる．

■例（平均身長の検定 3）　母集団分布は正規分布であるが母分散は未知であり，標本サイズは 16 と小さく中心極限定理は適用できない．このとき，標本平均が 165 センチ，不偏標本分散が 100 センチであることが分析結果から分かった．t 統計量は

$$t = \frac{4(165-160)}{10} = 2$$

である．また，自由度 15 の t 分布の境界値として以下が与えられている．

$$t_{0.05}(15) = 1.753, \quad t_{0.025}(15) = 2.131, \quad t_{0.01}(15) = 2.602, \quad t_{0.005}(15) = 2.947$$

このとき $H_0:\ \mu = 160, \quad H_1:\ \mu \neq 160$ として 5% 水準，1% 水準それぞれで検定を行う．

5% 水準の両側検定の場合，棄却域は

$$|t| > 2.131 \Longleftrightarrow \left(\bar{X} < 160 - 2.131\frac{10}{4} = 154.67\right) \vee \left(160 + 2.131\frac{10}{4} = 165.328 < \bar{X}\right)$$

であるので，5% 水準では帰無仮説は棄却されない．したがって，1% 水準の両側検定の場合でも帰無仮説は棄却されない．一応確認しておくと，棄却域は

$$|t| > 2.947 \Longleftrightarrow \left(\bar{X} < 160 - 2.947\frac{10}{30} = 152.63\right) \vee \left(160 + 2.947\frac{10}{30} = 167.37 < \bar{X}\right)$$

である．

次に，$H_0:\ \mu = 160, \quad H_1':\ \mu > 160$ として 5% 水準，1% 水準それぞれで検定を行う．

5% 水準の右片側検定の場合，棄却域は

$$|t| > 1.753 \Longleftrightarrow \bar{X} > 160 + 1.753\frac{10}{4} = 164.38$$

であるので，5% 水準では帰無仮説は棄却され，検定は有意となる．一方，1% 水準の両側検定の場合，棄却域は

$$|t| > 2.326 \Longleftrightarrow \bar{X} > 160 + 2.326\frac{10}{30} = 166.51$$

であるので，1% 水準では帰無仮説は棄却されず，検定は有意ではない．

11.2　ベルヌーイ分布の母数 p についての検定

帰無仮説を

$$H_0:\ p = p_0$$

とする．$X_1, X_2, \cdots, X_n$ を独立にベルヌーイ分布 $Bi(1,p)$ に従う確率変数とする．$\hat{p} = \sum_{i=1}^{n} /n$ としたとき，

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \fallingdotseq \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

は，漸近的に標準正規分布 $N(0,1)$ に従う．Z を検定統計量とする．

■有意水準 5% の両側検定　対立仮説を

$$H_1:\ p \neq p_0$$

とする．

$$P(|Z| > 1.96) = 0.05$$

であるので，ここでは Z をバラさず表記すると，

$$(Z < -1.96) \vee (1.96 < Z)$$

が棄却域であり，$\bar{X}$ が棄却域に落ちるとき，帰無仮説 H_0 を棄却し H_1 を採択する．$\bar{X}$ が棄却域に落ちないとき，帰無仮説 H_0 は棄却しない．

有意水準が 1% のときは，境界値として ± 1.96 の代わりに ± 2.576 を用いる．

■有意水準 5% の右片側検定　対立仮説を

$$H_1:\ p > p_0$$

とすると，棄却域は $Z > 1.645$ である．

左片側検定については，棄却域を $Z < -1.645$ とすればよい．有意水準が 1% のときは，境界値として 1.645 の代わりに 2.326 を用いる．

■**例（支持率の検定）**　無作為抽出調査を実施し対象者に現内閣への支持・不支持を尋ねた．支持を 1，不支持を 0 とする確率変数 X_i は，真の支持率 p を母数とするベルヌーイ分布 $Bi(1,p)$ に従う[*1]．100 人の標本において支持者が 60 人であった．つまり標本支持率は $\hat{p}=0.6$ である．前回の真の支持率が 0.55 であったので，今回の真の支持率が有意に上昇しているか知りたい．ゆえに

$$H_0:\ p=0.55,\quad H_1:\ p>0.55$$

を検定する．

$$Z=\frac{0.6-0.55}{\sqrt{\frac{0.55(1-0.55)}{100}}}=1.005$$

なので，片側 1% 水準の境界値 2.326 はもちろん，片側 5% 水準の境界値 1.645 をも下回る．ゆえに，帰無仮説は棄却されない．前回調査時の支持率よりも今回の支持率が有意に上昇しているとは言えない．

一方，標本支持率が変わらず 0.6 で標本サイズが 1000 の場合，$Z=3.178$ なので，1% 水準で帰無仮説は棄却され，対立仮説が採用される．この場合，1% の有意水準で，前回調査時の支持率よりも今回の支持率が有意に上昇していると言える．

課題 11.1. *LUNA* のテストを *12/9*（水）*23:59* までに受けなさい．

[*1] あるいは $X=\sum_{i=1}^{n}X_i$ は二項分布 $Bi(n,p)$ に従う．

第 12 章

仮説検定 (3): 母平均の差・母比率の差の検定

前章にて最低限の目標である「母平均についての推定・検定を導入する」ことを達成した．ここまでで，基本的な推測統計学のロジックはある程度理解できたものと思う．

本章では，より実践的な検定法として母平均の差の検定，母比率の差の検定を導入し，さらに最後にいくつかの検定法について言及する．

12.1　母平均の差の検定

母平均の差の検定は，t 統計量と t 分布を用いることから，**t 検定** (t test) とも呼ばれる．

平均 μ_x, 分散 σ_x^2 の分布 $f(x)$ をもつ母集団から，大きさ m の標本 $X_1, X_2, \cdots, X_m$ を抽出する．同時に，平均 μ_y, 分散 σ_y^2 の分布 $f(y)$ をもつ母集団から，大きさ n の標本 $Y_1, Y_2, \cdots, Y_n$ を抽出する．このとき，2 つの母集団分布は独立であると仮定する．例えば，無作為抽出された標本のうち，男性標本を男性母集団からの抽出，女性標本を女性母集団からの抽出と考える（図 12.1）．

ここで帰無仮説を

$$H_0 :\ \mu_x = \mu_y$$

あるいは，同じことだが表現を変えて，

$$H_0 :\ d_0 = \mu_x - \mu_y = 0$$

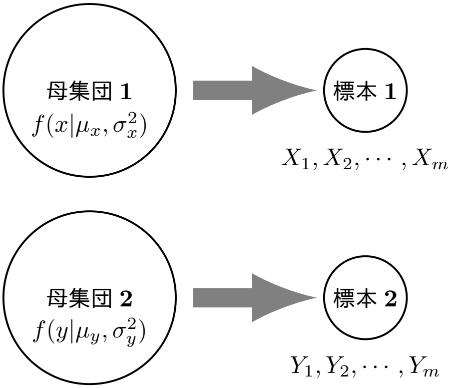

図 12.1　2 母集団と標本の関係

とする．対立仮説は両側検定の場合，

$$H_1: \ \mu_x \neq \mu_y \quad (d_0 \neq 0)$$

であり，片側検定の場合

$$H_1': \ \mu_x > \mu_y \quad (d_0 > 0), \quad \text{もしくは}, \quad H_1'': \ \mu_x < \mu_y \quad (d_0 < 0)$$

である．ここでは，両側検定だけを説明するが，これまでと同様にして片側検定も行うことができる．

このとき，検定統計量としては，それぞれの標本平均を

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

として，その差をとった統計量

$$D = \bar{X} - \bar{Y}$$

を元にしたものを用いる．標本平均の期待値（平均）の特性から，

$$E(D) = \mu_x - \mu_y$$

であり，帰無仮説の下で $E(D) = 0$ となる．

以下の検定では，大きく分けて 3 つのパターンに分けて考える必要がある：

(1) 母分散が既知の場合
(2) 母分散は未知ではあるが，母分散は等しいと仮定してよい場合（分散均一性の仮定）
(3) 母分散は未知であり，母分散は等しいと仮定できない場合

以下では簡単のため，(1)，(2) のケースのみ紹介する[*1]．

12.1.1　母分散が既知の場合

2 つの母集団の母集団分布を正規分布，$N(\mu_x, \sigma_x^2), N(\mu_y, \sigma_y^2)$ であると仮定する．このとき，D の分布は，「正規分布の再生性」，「正規母集団の標本平均の標本分布」の特性より，正規分布に従うことが分かっている．より具体的には，

$$D \sim N(\mu_x - \mu_y, (\sigma_x^2/m) + (\sigma_y^2/n)).$$

ここで，D を標準化すると，帰無仮説の下で $\mu_x - \mu_y = 0$ なので，

$$Z = \frac{D - E(D)}{\sqrt{V(D)}} = \frac{\bar{X} - \bar{Y}}{\sqrt{(\sigma_x^2/m) + (\sigma_y^2/n)}} \tag{12.1}$$

となる．$Z \sim N(0,1)$ なので，

$$|Z| > z_{\alpha/2} \tag{12.2}$$

を棄却域として検定を実施する．両側 5% 検定の場合棄却域は $|Z| > 1.96$，両側 1% 検定の場合棄却域は $|Z| > 2.576$ である．

なお，2 つの母集団の母分布が正規分布でない場合も，標本サイズ m, n が十分に大きいならば，中心極限定理を用いて，Z が漸近的に標準正規分布 $N(0,1)$ に従うことが分かっているので，この場合上と同様の手続きによって検定できる．

[*1] (3) のケースでも，それぞれの母集団分布が正規分布で標本サイズが少ない場合，そして母集団分布が未知でも標本サイズが大きい場合には検定を実施することができる．たいていの統計ソフトには「等分散性を仮定しない母平均の差の検定（Welch の検定）」が実装されている．

12.1.2　母分散は未知ではあるが，母分散は等しいと仮定してよい場合

2 つの母集団の母集団分布を正規分布，$N(\mu_x, \sigma_x^2), N(\mu_y, \sigma_y^2)$ であると仮定する．共通する分散 $\sigma_x^2 = \sigma_y^2$ を

$$\begin{aligned} S^2 &= \frac{1}{m+n-2}\left[\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] \\ &= \frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2} \end{aligned}$$

によって推定することにする．これを式 (12.1) の σ_x^2, σ_y^2 に代入すると，

$$t = \frac{\bar{X} - \bar{Y}}{S\sqrt{(1/m) + (1/n)}} \tag{12.3}$$

という統計量を得る．これを検定統計量とする．この統計量は自由度が $m+n-2$ の t 分布 $t(m+n-2)$ に従うことが分かっている[*2]．ゆえに，両側 $100 \times \alpha\%$ 検定の場合棄却域は

$$|t| > t_{\alpha/2}(m+n-2) \tag{12.4}$$

である．

なお，2 つの母集団の母分布が正規分布でない場合は，標本サイズ m, n が十分に大きいならば，中心極限定理を用いて，t が漸近的に標準正規分布 $N(0,1)$ に従うことが分かっている．この場合，$|t| > z_{\alpha/2}$ を棄却域として検定を実施すればよい．

■例（テストの平均点の男女差）　男性 100 人，女性 100 人を標本として何らかのテストを実施した．過去の知見から，それぞれの母集団分布は正規分布で，母分散は等しいと仮定してよい．男性の平均得点は 63 で不偏分散は 50，女性の平均得点は 65 で不偏分散は 48 であった．このとき「母平均が等しい」という帰無仮説に対して，5% 水準，1% 水準で両側検定を実施する．

$$t = \frac{65 - 63}{\sqrt{\frac{99*50+99*48}{198}}\sqrt{(1/100) + (1/100)}} = 2.020$$

[*2] t の計算で用いる確率変数は $m+n$ 個だが，S^2 の計算の際に $\bar{X}, \bar{Y}$ についての制約で自由度を 2 失っているので，自由度は $m+n-2$ である．

である．$t_{0.025}(198) = 1.972, t_{0.005}(198) = 2.601$ であるので，5% 水準の場合 t は棄却域に落ちるが，1% 水準の場合棄却域には落ちない．ゆえに，有意水準 5% 水準では両者の平均点に差があると言えるが，1% 水準では差があるとは言えない，という結果になった．

12.2　母比率の差の仮説検定

確率 p_1 のベルヌーイ分布 $Bi(1, p_1)$ に従う母集団から，大きさ m の標本 $X_1, X_2, \cdots, X_m$ を抽出する．同時に，確率 p_2 のベルヌーイ分布 $Bi(1, p_2)$ に従う母集団から，大きさ n の標本 $Y_1, Y_2, \cdots, Y_n$ を抽出する．このとき，2 つの母集団分布は独立であると仮定する．

帰無仮説と対立仮説を

$$H_0:\ p_1 - p_2 = 0, \qquad H_1:\ p_1 - p_2 \neq 0$$

とする．それぞれの標本平均（割合）を

$$\hat{p}_1 = \frac{1}{m}\sum_{i=1}^{m} X_i, \qquad \hat{p}_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

としたとき，検定統計量

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{(1/m) + (1/n)}} \tag{12.5}$$

は，n が十分に大きいとき，漸近的に標準正規分布にしたがう．ただし，

$$\hat{p} = \frac{m\hat{p}_1 + n\hat{p}_2}{m+n}$$

である．$Z \sim N(0,1)$ なので，

$$|Z| > z_{\alpha/2} \tag{12.6}$$

を棄却域として検定を実施する．両側 5% 検定の場合棄却域は $|Z| > 1.96$，両側 1% 検定の場合棄却域は $|Z| > 2.576$ である

■例（支持率の差の検定）　男性と女性 2 つの母集団に対して無作為抽出調査を実施し対象者に現内閣への支持・不支持を尋ねた．それぞれの標本において，支持を 1，不支持を 0 とする確率変数は，真の支持率 p_1, p_2 を母数とするベルヌーイ分布に従う．男性標本では 100 人の標本において支持者が 60 人であった．つまり標本支持率は $\hat{p}_1 = 0.6$ である．女性標本では 100 人の標本において支持者が 40 人であった（標本支持率は $\hat{p}_2 = 0.4$）．2 つの母集団で真の支持率に有意差があるかを知りたい．ゆえに

$$H_0:\ p_1 - p_2 = 0, \quad H_1:\ p1 - p_2 \neq 0$$

を検定する．

$$Z = \frac{0.6 - 0.4}{\sqrt{0.5(1-0.5)}\sqrt{(1/100) + (1/100)}} = 2.828$$

なので，両側 5% 水準の境界値 1.96 はもちろん，両側 1% 水準の境界値 2.576 を上回る．ゆえに，どちらの水準を設定しても帰無仮説は棄却され，男性と女性 2 つの母集団において支持率は異なる（有意差がある）という結論を得る．

12.3　そのほかの検定

そのほか，知りたい母数に応じてさまざまな検定がある．今後さらに異なる検定手法にであった場合でも，数理統計学的にきちんと理解しようと思うと，どれも結構難しかったりするが，検定の基本的なロジックは同じなので，基本的なところの理解は早いと思う．その際には，(1) 帰無仮説は何か，(2) どのような検定統計量が使われているか，(3) どのような分布が使われているか，に注意すればよいだろう．

入門・基礎レベルでよく取り扱われるものとして，たとえば，相関係数の検定がある．これは，(1) 母集団における相関係数が 0 であるという帰無仮説のもと，(2) 標本相関係数を変形した検定統計量を用い，(3) それが t 分布に従うことを利用して検定や区間推定を行う．

あるいは，クロス集計表の独立性の検定も社会学では頻出する．これは，(1) 母集団において，クロス表を構成する 2 つの分布（周辺分布）が独立であるという帰無仮説のもと，(2) ピアソンの χ^2 統計量を検定統計量として，(3) それが χ^2 分布に従う

ことを利用した検定になっている[*3].

課題 12.1. *LUNA* のテストを *12/16*（水）*23:59* までに受けなさい.

[*3] 森棟ほか (2008: 10 章)，倉田・星野 (2009: 8 章) を参照．独立性の検定については，参考資料「2020 秋データ分析基礎 (12)_参考 (独立性の検定).pdf」も参照のこと．

第 13 章

回帰分析における推定と検定

本講義の最後に，回帰分析における推定と検定について紹介する．本章では，単回帰分析における回帰係数 β の推定・検定のみに焦点を当てて説明するが，同様のロジックで重回帰分析の偏回帰係数の推定・検定も理解可能である．記述的な単回帰分析については，すでに導入済みとして話を進めるが，最初におさらいとして記述的な単回帰分析について簡単に確認する．

もしかしたら，受講者によっては回帰分析をまったく習っていない人もいるかもしれない．その場合，本章の内容はかなり難しく感じるかもしれない．その場合は細部はともかく，すくなくとも回帰係数の推定・検定について，ざっくりと理解できれば問題ない．

13.1　記述的な単回帰分析

記述的な意味での**回帰分析** (regressoin analysis)，とくに独立変数（説明変数）が 1 つの単回帰分析は，散布図上に散布図のちらばりを要約する直線を引くことに他ならない（図 13.1）．

x を独立変数（制約なく動く変数），y を従属変数（x の変化に伴って変化する変数）とすると，直線

$$y = a + bx \tag{13.1}$$

の係数 a, b として，散布図を表現する最適のものを選ぶ問題となる．b をとくに，**回帰係数** (regression coefficient) という．

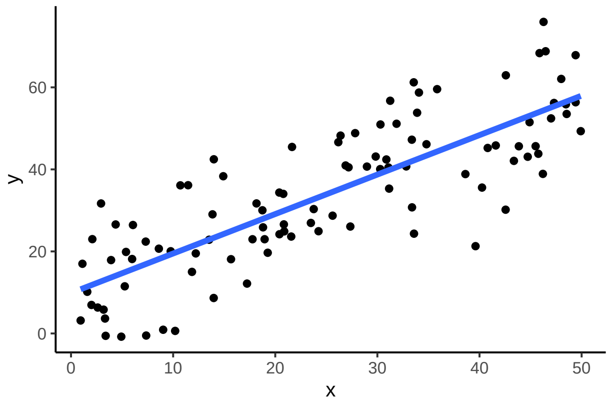

図 13.1　x–y 平面上の散布図に直線を引く

これを解くために，散布図上の点 (x_i, y_i) のそれぞれについて

$$\forall i \in \{1, 2, \cdots, n\}, y_i = a + bx_i + e_i \tag{13.2}$$

という関係を仮定する[*1]．ここで，e_i は点 (x_i, y_i) の直線 $y = a + bx$ からの y 方向の残差である．この残差の二乗和 S_e

$$S_e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i) \tag{13.3}$$

を最少にする a, b の値を求めることを**最小二乗法** (least square method) という．

途中の計算を飛ばして結論だけを言うと，最少二乗法でもとめた a, b の推定値 $\hat{a}, \hat{b}$ は

$$\hat{a} = \bar{y} - \hat{b}\bar{x}, \qquad \hat{b} = \frac{c_{xy}}{\hat{\sigma}_x^2}$$

で求められる．つまり，回帰係数は最少二乗推定値は，x と y の共分散を x の分散で割ることで得られる．

[*1] $\forall a \in A$ は，「集合 A に属するすべての（任意の）i について」を意味する．

実際の分析においては，手計算で推定値を出すことはほとんどない．統計ソフトには回帰分析の機能が必ず備わっている[*2]．たとえば R を使って図 13.1 のデータを回帰分析させると，以下のようになる．

```
> summary(lm(y ~ x, dat))

Call:
lm(formula = y ~ x, data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-26.6956  -7.0253  -0.8862   6.0086  21.5698

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.86720    2.02106   4.882 4.08e-06 ***
x            0.96227    0.06952  13.842  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.48 on 98 degrees of freedom
Multiple R-squared:  0.6616,Adjusted R-squared:  0.6582
F-statistic: 191.6 on 1 and 98 DF,  p-value: < 2.2e-16
```

ここから，

$$y = 9.86720 + 0.96227x$$

という直線による散布図の要約を得る．つまり，x と y の関係は，$x = 0$ のとき $y = 9.87$ で，x が 1 単位大きくなるごとに，y もおよそ 0.96 ずつ大きくなる傾向であると要約できる．

しかし，記述のレベルではこれはどこまで行っても標本データの要約である．対象のデータが母集団からの適切なランダムサンプリングで得られたとすると，ここから母集団上の x と y の関係についてどのようなことが言えるだろうか．そこで以下では確率を入れ込んだ回帰モデルを想定する．

[*2] Excel でもできる．

13.2　標準線形回帰モデル

母集団上で確率変数 Y が（確率変数ではない）変数 x と係数 α, β によって，平均が規定される正規分布に従っていると考える．つまり，

$$Y \sim N(\alpha + \beta x, \sigma^2) \tag{13.4}$$

と仮定する[*3]．

母集団からのランダムサンプリングによって，n 個の独立な確率変数 $Y_1, Y_2, \ldots, Y_n$ が得られる[*4]．確率変数 $Y_1, Y_2, \ldots, Y_n$ は，

$$Y_i = \alpha + \beta x_i, \qquad Y_i \sim N(\alpha + \beta x_i, \sigma^2) \tag{13.5}$$

もしくは，誤差確率変数 ε_i を導入して別の表現で書けば

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \text{ iid} \tag{13.6}$$

と表すことができる．この仮定によって，ε について

1. $\forall i, E(\varepsilon_i) = 0$
2. $\forall i, V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$
3. $\forall i \neq j, Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0$

が成り立つ．

以上の仮定を図解すると図 13.2 のようになる．母集団分布は赤色の正規分布であり，x のレベルで平均パラメータが動く[*5]．ただし，分散は x のレベルで変わらない（分散均一性の仮定）．そして，実際の散布図データは，この母集団分布からランダムに抽出されたものである[*6]．

[*3] 回帰分析の古典的な理解では独立変数 x は確率変数ではない．しかし，最近の考え方は X も確率変数と考える．ここでは，古典的な理解に準じる．

[*4] 各 Y_i には x の値として x_i が付随する．

[*5] 本当は，分布の赤い線は x のレベルに応じてぎっしり詰まっている．カマボコを上から俯瞰することを想像せよ．

[*6] カマボコをぶるんぶるんしたら，黒いボツボツがふぁらーと落ちてくるみたいなイメージである．

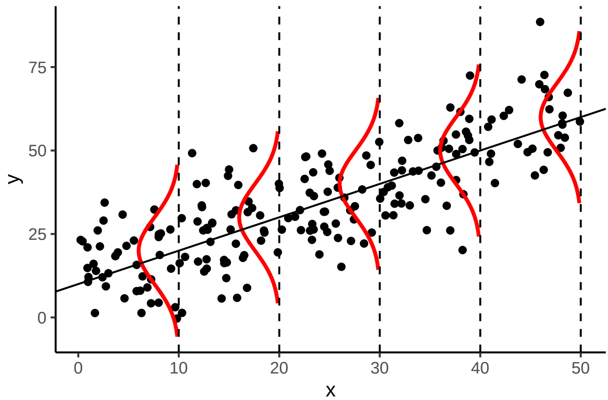

図 13.2　母集団分布と標本データ

13.3　最小二乗推定量

この節以降の数式部分は適宜読み飛ばして結論部分のみ理解すればよい．

各 i についての残差 $e_i = Y_i - \alpha - \beta x_i$ の平方和（残差平方和）

$$S_e = \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2$$

を最小化する α, β の推定量 $\hat{\alpha}, \hat{\beta}$ を**最小二乗推定量** (least squares estimator) という．

$\hat{\alpha}, \hat{\beta}$ を求めるのは，$S_e = f(\alpha, \beta)$ の最小化問題を解けばよい．その一階の条件は，

$$\frac{\partial S_e}{\partial \alpha} = -2 \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i) = 0 \tag{13.7}$$

$$\frac{\partial S_e}{\partial \beta} = -2 \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i) x_i = 0 \tag{13.8}$$

である[*7]．これより，連立方程式

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} x_i \tag{13.9}$$

$$\sum_{i=1}^{n} x_i Y_i = \hat{\alpha} \sum_{i=1}^{n} x_i + \hat{\beta} \sum_{i=1}^{n} x_i^2 \tag{13.10}$$

を得る．これらを**正規方程式** (normal equation) という．式 (13.9) より，

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{13.11}$$

を得る．これを式 (13.10) に代入して

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{13.12}$$

を得る．ここで，x と Y の偏差積和と x の平方和

$$S_{xY} = \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

を用いれば，

$$\hat{\beta} = \frac{S_{xY}}{S_{xx}} \tag{13.13}$$

を得る[*8]．

ここで何より重要なのは，Y_i が確率変数であり，サンプリングによって確率的に値が変化するものであり，それに応じて $\hat{\beta}$ も確率変数として，サンプリングによって確率的に値が変化する，ということである[*9]．したがって，$\hat{\beta}$ は標本分布をもつ．ではそれはどのような分布か．

[*7] 本当は，2 階の条件まで確認する必要がある．詳しくは，森棟ほか（2008: 369–371）

[*8] このモデルの仮定の下では，最少二乗推定量と最尤推定量は等しくなる．詳細は，付録 A.8 節を参照のこと．

[*9] つまりプルプルしている．

13.4　最小二乗推定量 $\hat{\beta}$ の特徴

最小二乗推定量 $\hat{\beta}$ はどのような特徴をもつだろうか.

$$S_{xY} = \sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y}) = \sum_{i=1}^{n}(x_i - \bar{x})Y_i \tag{13.14}$$

に注意すると，式 (13.13) の $\hat{\beta}$ は

$$\hat{\beta} = \sum_{i=1}^{n} w_i Y_i \tag{13.15}$$

と表せる．ただし，

$$w_i = \frac{(x_i - \bar{x})}{S_{xx}} \tag{13.16}$$

である．この式は，最小二乗推定量 $\hat{\beta}$ が確率変数 Y_i の線形結合からなる**線形推定量**であることを示している．さらに式 (13.15) に $Y_i = \alpha + \beta x_i + \varepsilon_i$ をぶち込むと

$$\begin{aligned}
\hat{\beta} &= \sum_{i=1}^{n} w_i(\alpha + \beta x_i + \varepsilon_i) \\
&= \alpha \sum_{i=1}^{n} w_i + \beta \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i \\
&= \beta + \sum_{i=1}^{n} w_i \varepsilon_i
\end{aligned} \tag{13.17}$$

となる．ここから両辺の期待値をとって，

$$E(\hat{\beta}) = \beta \tag{13.18}$$

を得る．これは，最小二乗推定量 $\hat{\beta}$ が**不偏推定量**であることを示している．さらに $\hat{\beta}$ の分散は，式 (13.17) を使って

$$\begin{aligned} V(\hat{\beta}) = E[(\hat{\beta}-\beta)^2] &= E\left[\left(\sum_{i=1}^{n} w_i\varepsilon_i\right)^2\right] \\ &= E\left[\sum_{i=1}^{n} w_i^2\varepsilon_i^2 + \sum_{i\neq j} w_i w_j \varepsilon_i\varepsilon_j\right] \\ &= \sum_{i=1}^{n} w_i^2 E(\varepsilon_i^2) + \sum_{i\neq j} w_i w_j E(\varepsilon_i\varepsilon_j) \\ &= \sigma^2\sum_{i=1}^{n} w_i^2 = \frac{\sigma^2}{S_{xx}} \end{aligned} \tag{13.19}$$

を得る．$V(\hat{\beta})$ は他の任意の線形不偏推定量の分散と比べて最も小さくなることが知られている[*10]．その意味で最小二乗推定量 $\hat{\beta}$ は**最良線形不偏推定量** (best linear unbiased estimator: BLUE) と呼ばれる[*11]．

13.5　最小二乗推定量 $\hat{\beta}$ の標本分布

7 章の正規分布の再生性の定理 7.1 をより一般化したものとして以下の定理が成り立つ（森棟ほか 2008: 235–236）．

定理 13.1. 独立な確率変数 X_1 と X_2 の分布が各々 $N(\mu_1,\sigma_1^2), N(\mu_2,\sigma_2^2)$ であるとすると，aX_1+bX_2 *(*a, b は定数*)* の分布は $N(a\mu_1+b\mu_2, a^2\sigma_1^2+b^2\sigma_2^2)$ になる．

ここから，$\varepsilon_i \sim N(0,\sigma^2)$ なので

$$\sum_{i=1}^{n} w_i\varepsilon_i \sim N\left(0, \sigma^2\sum_{i=1}^{n} w_i^2\right), \tag{13.20}$$

さらに，正規確率変数 $X \sim N(\mu,\sigma^2)$ の線形変換に対して $aX+b \sim N(a\mu+b, a^2\sigma^2)$

---

*10 **ガウス＝マルコフの定理** (Gauss-Markov theorem) という．

*11 ここまでの話では，誤差の正規性の仮定はいらないことに注意．

となるので，式 (13.17) より

$$\hat{\beta} \sim N(\beta, \sigma^2/S_{xx}) \tag{13.21}$$

である[*12].

つまり，$\hat{\beta}$ の標本分布は，真の値 β を中心とした正規分布である.

13.6　誤差分散の推定量

13.6.1　予測値・残差

回帰式による Y_i の予測値 (回帰値) を $\hat{Y}_i$ とすると，

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i \tag{13.22}$$

であり，回帰モデルの残差 (residual) は

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} x_i) \tag{13.23}$$

と定義できる．予測値の総和は，式 (13.11) を使って，

$$\sum_{i=1}^{n} \hat{Y}_i = n(\hat{\alpha} + \hat{\beta}\bar{x}) = n\bar{Y} \tag{13.24}$$

である．また，これより，残差の総和は

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{Y}_i = 0 \tag{13.25}$$

である．また，

$$\begin{aligned} \sum_{i=1}^{n} e_i x_i &= \sum_{i=1}^{n} [(Y_i - \bar{Y}) - \hat{\beta}(x_i - \bar{x})] x_i \\ &= S_{xY} - \hat{\beta} S_{xx} = 0 \end{aligned} \tag{13.26}$$

[*12] ここでは，正規確率変数の再生性から $\hat{\beta}$ の分布を導き出した．しかしながら，誤差分布を正規分布と仮定しなくても，ある条件の下では iid であって十分に n が大きければ，中心極限定理から $\hat{\beta}$ の分布は漸近的に正規分布に従うことが知られている．また，同一分布でなくても独立性が確保されていれば，ある条件の下では $\hat{\beta}$ の分布は漸近的に正規分布に従う（佐和 1979: 68–71）.

も成り立つ．これは，残差ベクトルと x ベクトルの内積が 0，つまり 2 つのベクトルが直交することを示す．以上により，残差の自由度は $n-2$ であることが分かる．

残差の平方和 (residual sum of squares; RSS)S_e は

$$S_e = \sum_{i=1}^{n} e_i^2 \tag{13.27}$$

と定義される．

13.6.2　誤差分散の推定量

誤差の分散 σ^2 については

$$\hat{\sigma}^2 = \frac{S_e}{n-2} \tag{13.28}$$

を推定量として用いることが一般的である．このとき，$\hat{\sigma}^2$ について，

1. $(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi(n-2)$
2. $E(\hat{\sigma}^2) = \sigma^2$
3. $\hat{\sigma}^2$ と $\hat{\alpha}$，$\hat{\sigma}^2$ と $\hat{\beta}$ は独立

が成り立つ（倉田・星野 2009: 287–288）．

13.7　信頼区間と仮説検定

式 (13.21) より，$\hat{\beta}$ を標準化して

$$Z_{\hat{\beta}} = \frac{\hat{\beta}-\beta}{\sqrt{\sigma^2/S_{xx}}} \sim N(0,1) \tag{13.29}$$

を得る．しかし，通常母数 σ^2 は未知であるので，このままでは信頼区間の導出，仮説検定の検定統計量としては使えない．

ところで，$\hat{\sigma}^2$ の特性より $X = (n-2)\hat{\sigma}^2/\sigma^2$ は自由度 $n-2$ の χ^2 分布に従い $\hat{\beta}$ とは独立である．ゆえに，$Z_{\hat{\beta}}/\sqrt{X/(n-2)}$ は自由度 $n-2$ の t 分布に従う．すなわち，

$$t_{\hat{\beta}} = \frac{Z_{\hat{\beta}}}{\sqrt{X/(n-2)}} = \frac{\hat{\beta}-\beta}{\sqrt{\hat{\sigma}^2/S_{xx}}} \sim t(n-2). \tag{13.30}$$

つまり，$\hat{\beta}$ を標準化する際に $\hat{\beta}$ の分散をデータから推定して得られた t 統計量は自由度 $n-2$ の t 分布に従う．

13.7.1　信頼区間

自由度 $n-2$ の t 分布 $t(n-2)$ において，それ以上の値が実現する確率が $\gamma/2$ となるような値を $t_{\gamma/2}(n-2)$ とおく[*13]．このとき，$t(n-2)$ に従う確率変数 $t_{\hat{\beta}}$ について

$$P(-t_{\gamma/2}(n-2) \leq t_{\hat{\beta}} \leq t_{\gamma/2}(n-2)) = 1-\alpha \tag{13.31}$$

が成り立つ．この式の左辺カッコの中に

$$t_{\hat{\beta}} = \frac{\hat{\beta}-\beta}{\sqrt{\hat{\sigma}^2/S_{xx}}}$$

を代入してまとめると，

$$P\left(\hat{\beta} - t_{\gamma/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \leq \beta \leq \hat{\beta} + t_{\gamma/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right) = 1-\gamma \tag{13.32}$$

となる．ゆえに，β の $100(1-\gamma)\%$ 信頼区間は

$$\left[\hat{\beta} - t_{\gamma/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}},\ \ \hat{\beta} + t_{\gamma/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right] \tag{13.33}$$

によって計算できる．

13.7.2　検定統計量

帰無仮説，対立仮説をそれぞれ

$$H_0 : \beta = 0, \qquad H_1 : \beta \neq 0 \tag{13.34}$$

とする．このとき，帰無仮説が正しいという仮定の下で検定統計量は自由度 $n-2$ の t 分布に従う．つまり，

$$t^*_{\hat{\beta}} = \frac{\hat{\beta}}{\sqrt{\hat{\sigma}^2/S_{xx}}} \sim t(n-2). \tag{13.35}$$

[*13] 切片 α と表記が被るので，ここでは γ を用いる．

$|t^*_{\hat{\beta}}| \geq t_{\alpha/2}(n-2)$ のとき，両側 $100\alpha\%$ の有意確率で帰無仮説を棄却し対立仮説を採用する．$|t^*_{\hat{\beta}}| < t_{\alpha/2}(n-2)$ の場合帰無仮説は棄却できない．

13.7.3　検定の実際

図 13.1 のデータの分析結果から，検定の仮定を再現する．R の結果出力における x の Estimate の値が $\hat{\beta}$ に対応し，Std. Error の値が $\sqrt{\hat{\sigma}^2/S_{xx}}$ に対応する．ゆえに，t 値は

$$t^*_{\hat{\beta}} = \frac{0.96227}{0.06952} = 13.842.$$

となる．データサイズは $n = 100$ なので，自由度 98 の t 分布より，

$$P(|T| > 13.842) = 2 * 10^{-16}$$

が p 値として得られる．

13.8　さらに学ぶために

ここでは，最低限回帰分析の結果が得られたとき，各係数の横についている p 値やアスタリスクが何を意味しているのかが分かればよい．まとめると，

(1) $H_0 : \beta = 0, H_1 : \beta \neq 0$ と設定し，
(2) 帰無仮説のもとでの t 値を算出し，
(3) $t(n-2)$ 分布から棄却域に落ちるかどうかで有意性を判断（あるいは p 値で判断）

という手順を経る．帰無仮説が棄却されることが何を意味するのかというと，$\beta \neq 0$ であるので，少なくとも変数 x の変化によって Y が変化しないことはない，つまり，「独立変数は従属変数の変動に何らかの効果がある」といえる．

回帰モデルの評価においては，決定係数と適合度検定の理解が不可欠である．さらに学ぶためには，付録 A.10 節も参照．

課題 13.1. *LUNA* のテストを *12/23*（水）*23:59* までに受けなさい．

第 14 章

まとめと補題

ここまで，推測統計学の基礎を順を追って学んできた．

推測統計学はロジックそのものが難しいこともあって，わかりにくいところも多かったかもしれない．でもそれはある意味仕方のないことだと思う．だって難しいんだもの．

より十全な理解を求めるのであれば，ここで展開されるロジックを繰り返し繰り返しフォローして，自らの納得のレベルに落とし込んでいく必要がある．すべての人にとって推測統計学の十全な知識が必要だとは思わないが，向学心のある人はぜひ学びを続けていって欲しい．Welcome to deeper hell !!

課題 14.1. 本講義で何を学んだのか．振り返ってまとめのリポートを書こう．さらに，やる気のある人はエクストラの課題に取り組もう．これまでのテストでできなかった箇所に再挑戦したり，実践的なデータ分析（仮想データでもよい）をしてみてもよい．エクストラ課題にはプラスアルファで加点する．*Word* ファイルでリポートを作成し，*1/13*（水）*23:59* までに *LUNA* 上で提出せよ．

付録 A

A.1　集合演算の基本公式

集合演算 $\cup, \cap, (\ )^c$ に関して，以下のような基本定理が成り立つ．

(1) 交換律

$$A \cup B = B \cup A \tag{A.1}$$
$$A \cap B = B \cap A \tag{A.2}$$

(2) 分配律

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{A.3}$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{A.4}$$

(3) 同一律

$$A \cap U = A \tag{A.5}$$
$$A \cup \emptyset = A \tag{A.6}$$

(4) 補元律

$$A \cup A^c = U \tag{A.7}$$
$$A \cap A^c = \emptyset \tag{A.8}$$

(5) ベキ等律

$$A \cup A = A \tag{A.9}$$
$$A \cap A = A \tag{A.10}$$

(6) 有界律

$$A \cup U = U \tag{A.11}$$

$$A \cap \emptyset = \emptyset \tag{A.12}$$

(7) 吸収律

$$A \cup (A \cap B) = A \tag{A.13}$$

$$A \cap (A \cup B) = A \tag{A.14}$$

(8) 結合律

$$A \cup (B \cup C) = (A \cup B) \cup C \tag{A.15}$$

$$A \cap (B \cap C) = (A \cap B) \cap C \tag{A.16}$$

(9) 対合律

$$(A^c)^c = A \tag{A.17}$$

(10) ド・モルガン律

$$(A \cup B)^c = A^c \cap B^c \tag{A.18}$$

$$(A \cap B)^c = A^c \cup B^c \tag{A.19}$$

A.2　大数の法則

まず，チェビチェフの不等式を導入する．

定理 A.1 (チェビチェフの不等式)**.** 確率変数 X の平均を μ，標準偏差を σ とする．このとき，k を任意の正の数とすると，

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{A.20}$$

が成立する．

証明. 離散の場合について証明する．$\sum$ を $\int$ に変えることによって連続の場合も同様に証明できる．X のとりうる値の集合を R とする．R は次の 2 つの集合に分割で

きる．

$$R_1 = \{x||x-\mu| \geq k\sigma\}$$
$$R_2 = \{x||x-\mu| < k\sigma\}$$

ここで，$R = R_1 \cup R_2$, $R_1 \cap R_2 = \emptyset$ である．このとき，分散 σ^2 は

$$\sigma^2 = \sum_{x\in R}(x-\mu)^2 f(x) = \sum_{x\in R_1}(x-\mu)^2 f(x) + \sum_{x\in R_2}(x-\mu)^2 f(x)$$

と分解できる．一般に，$(x-\mu)^2 f(x) \geq 0$ であるので，

$$\sigma^2 \geq \sum_{x\in R_1}(x-\mu)^2 f(x). \tag{A.21}$$

このとき，R_1 に属する x については，

$$|x-\mu| \geq k\sigma \Longrightarrow (x-\mu)^2 \geq k^2\sigma^2$$

が成り立つので，

$$\sum_{x\in R_1}(x-\mu)^2 f(x) \geq \sum_{x\in R_1} k^2\sigma^2 f(x). \tag{A.22}$$

式 (A.21), (A.22) より，

$$\frac{1}{k^2} \geq \sum_{x\in R_1} f(x)$$

が成り立つ．ところで，

$$\sum_{x\in R_1} f(x) = P(|x-\mu| \geq k\sigma)$$

であるので，定理の不等式が成立する．　□

チェビチェフの不等式 (A.20) は，$k\sigma = \epsilon\ (> 0)$ とおくと

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \tag{A.23}$$

と表すこともできる．

チェビチェフの不等式は，直感的には平均 μ からプラス方向マイナス方向に $\pm k\sigma$ 以上離れているような x が実現する確率が $1/k^2$ 以下になることを言っている．

次に，大数の法則を証明する．

定理 A.2 (大数の法則)**.** $X_1, X_2, \cdots X_n$ が独立で，かつ平均 μ，分散 σ^2 の同一の確率分布に従うとする．このとき，X_i の相加平均

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

は確率変数であり，$\bar{X}$ は μ に確率収束する．すなわち，

$$\operatorname*{plim}_{n \to \infty} \bar{X} = \mu. \tag{A.24}$$

証明. 式 (6.17) より $E(\bar{X}) = \mu$, また式 (6.18) より $V(\bar{X}) = \sigma^2/n$ であるので，チェビチェフの不等式は

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

となる．ここで，右辺は $n \to \infty$ のとき 0 に近づき，左辺は確率なので非負である．ゆえに，

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0 \tag{A.25}$$

が成立する． □

A.3 二項分布についての中心極限定理

以下，薩摩（1989: 84-6）を参考にしながら，二項分布についての中心極限定理の証明の道筋をできるだけ丁寧にフォローする．

定理 A.3 (二項分布についての中心極限定理)**.** 二項分布 $Bi(n, p)$ に従う確率変数 X を標準化して

$$Z = \frac{X - \mu}{\sigma}$$

とするとき，$n \to \infty$ であれば，確率変数 Z は，確率密度関数

$$g(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \tag{A.26}$$

をもつ連続確率分布に従う．

証明. 確率変数 Z の値 z は具体的には，

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{np(1-p)}} \tag{A.27}$$

である．X の確率関数

$$f(x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \tag{A.28}$$

を使って，Z の確率密度関数をうまく定義したい．x が動く 1 単位は $\Delta z = 1$ である．このときそれにともなって z が動く 1 単位は

$$\Delta z = \frac{1}{\sqrt{np(1-p)}} \tag{A.29}$$

である．

$$\lim_{n\to\infty} \Delta z = 0$$

であるので，Δz は n が大きくなるにつれてどんどん小さくなり 0 に近づく．極限においては z は連続変数であると見なすことができる．

さて，$x \leq X < x + \Delta x$ に対応する確率と $z \leq Z < \Delta z$ に対応する確率を一致させるためには，

$$P(x \leq X < x + \Delta x) = f(x) = g(z)\Delta z = P(z \leq Z < \Delta z)$$

が成立しなければならない．ここから，

$$g(z) = \sqrt{np(1-p)} f(x). \tag{A.30}$$

また，同様にして，

$$g(z + \Delta z) = \sqrt{np(1-p)} f(x+1) \tag{A.31}$$

を得る．ここで，

$$f(x+1) = {}_nC_{x+1} p^{x+1} (1-p)^{n-x-1} = \frac{n!}{(x+1)!(n-x-1)!} p^{x+1} (1-p)^{n-x-1} \tag{A.32}$$

である．式 (A.31) を式 (A.30) で割って整理すると，

$$\frac{g(z+\Delta z)}{g(z)} = \frac{f(x+1)}{f(x)} = \frac{(n-x)p}{(x+1)(1-p)} \tag{A.33}$$

となる．ところで，式 (A.29) を使って，式 (A.27) を x についての式に直すと，

$$x = np + \frac{z}{\Delta z} \tag{A.34}$$

であるので，これを式 (A.33) に代入すると，

$$\frac{g(z+\Delta z)}{g(z)} = \frac{np(1-p) - pz/\Delta z}{np(1-p) + (1-p)(z/\Delta z + 1)} \tag{A.35}$$

を得る．さらに，式 (A.29) より $np(1-p) = 1/(\Delta z)^2$ を式 (A.35) 代入して整理すると，

$$\frac{g(z+\Delta z)}{g(z)} = \frac{1 - pz\Delta z}{1 + (1-p)(z\Delta z + (\Delta z)^2)} \tag{A.36}$$

となる．ここから，

$$\begin{aligned}\frac{g(z+\Delta z) - g(z)}{\Delta z} &= \left(\frac{g(z+\Delta z)}{g(z)} - 1\right)\frac{g(z)}{\Delta z} \\ &= \frac{-z-(1-p)\Delta z}{1+(1-p)(z\Delta z + (\Delta z)^2)} g(z)\end{aligned}$$

を得る．

$$\frac{g(z+\Delta z) - g(z)}{\Delta z} = \frac{-z-(1-p)\Delta z}{1+(1-p)(z\Delta z + (\Delta z)^2)} g(z)$$

の両辺で $n \to \infty$，つまり $\Delta z \to 0$ とすると，

$$\frac{dg(z)}{dz} = -zg(z)$$

を得る．

$$\begin{aligned}\frac{dg(z)}{dz} = -zg(z) &\iff \frac{1}{g(z)}\frac{dg(z)}{dz} = -z \\ &\iff \frac{d}{dz}\log g(z) = -z\end{aligned}$$

さらに両辺を z について積分して，

$$\frac{d}{dz}\log g(z) = -z \iff \log g(z) = -\frac{1}{2}z^2 + C' \quad (C'\text{は積分定数})$$
$$\iff g(z) = C\exp\left\{-\frac{1}{2}z^2\right\} \quad (C\text{ は積分定数})$$

となる．確率密度関数の全積分は 1 なので，

$$1 = \int_{-\infty}^{\infty} g(z)dz = C\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}z^2\right\}dz = C\sqrt{2\pi}$$

となるので*1，

$$C = \frac{1}{\sqrt{2\pi}}$$

を得る．結局，$g(z)$ は $n \to \infty$ のとき

$$g(z) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2}\right\} \tag{A.37}$$

となる． □

A.4　変数変換

定理 A.4. 連続な導関数をもつ単調関数 ϕ によって，確率変数 X が $Y = \phi(X)$ に変換されるとき，Y についての確率密度関数 $g(y)$ は，X についての密度関数 $f(x)$ を

---

*1 積分 $\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}z^2\right\}dz$ を解くには，いくつかの積分法のアドバンストな知識が必要になる．

$$I = \int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}x^2\right\}dx$$

として，I^2 を極座標変換 $x = r\cos\theta, y = r\sin\theta$ を施して積分する．すなわち，

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}x^2 - \frac{1}{2}y^2\right\}dxdy \\ &= \int_0^{2\pi}\int_0^{\infty} r\exp\left\{-\frac{1}{2}r^2\right\}drd\theta \\ &= \int_0^{2\pi}\left[-\exp\left\{-\frac{1}{2}r^2\right\}\right]_0^{\infty}d\theta \\ &= \int_0^{2\pi}d\theta = 2\pi \end{aligned}$$

ゆえに，$I = \sqrt{2\pi}$ となる．

用いて

$$g(y) = \left|\frac{dx}{dy}\right| f(x) = \left|\frac{d\phi^{-1}(y)}{dy}\right| f(\phi^{-1}(y)) \tag{A.38}$$

と表される．

証明． 単調関数 ϕ が増加関数の場合を考える．X が x と $x+\Delta x$ の間にある確率と，Y が y と $y+\Delta y$ の間にある確率とは等しくなければならない．ゆえに，

$$\begin{aligned} & P(x \leq X \leq x+\Delta x) = P(y \leq Y \leq y+\Delta y) \\ \Longleftrightarrow & \int_x^{x+\Delta x} f(s)ds = \int_y^{y+\Delta y} g(t)dt \end{aligned}$$

このとき，$\Delta x, \Delta y$ が十分小なるとき，近似的に

$$f(x)\Delta x = g(y)\Delta y \Longleftrightarrow g(y) = \frac{\Delta x}{\Delta y} f(x)$$

となるので，$\Delta x \to \infty, \Delta y \to 0$ とすると $g(y) = (dx/dy)f(x)$，定理を得る．単調関数 ϕ が減少関数の場合は，$\int_x^{x+\Delta x} f(s)ds = -\int_y^{y+\Delta y} g(t)dt$ であり，dx/dy は負となるので定理を得る． $\square$

標準化の場合は，$X = \sigma Z + \mu$ なので，

$$g(z) = \sigma f(\sigma z + \mu) \tag{A.39}$$

となる．逆に，$Z = (X-\mu)/\sigma$ なので

$$f(x) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right) \tag{A.40}$$

である．

A.5 モーメント法

何らかの確率分布に従う確率変数 X を k 乗したものの期待値を，確率変数 X の **（原点まわりの） $\boldsymbol{k}$ 次のモーメント**という．つまり，

$$E(X^k).$$

母集団分布のモーメントを，標本モーメント，すなわち標本 $X_1, X_2, \cdots, X_n$ から計算される $X_1^k, X_2^k, \cdots, X_n^k$ の相加平均

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k$$

によって推定することによって, 母数の推定量を求める方法を**モーメント法** (methods of moment) という.

A.5.1　モーメント法による母平均の推定

母平均は 1 次のモーメントなので，母平均は標本平均

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

によって推定される.

A.5.2　モーメント法による母分散の推定

母分散は式 (5.15) より，

$$V(X) = E(X^2) - E(X)^2 \tag{A.41}$$

と 2 次のモーメントと 1 次のモーメントに分解できるのであった．ゆえに，母分散の推定量は

$$\begin{aligned} E(X^2) - E(X)^2 &= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - n2\bar{X}^2 + n\bar{X}^2\right) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i\bar{X} + n\bar{X}^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \hat{\sigma}^2 \end{aligned}$$

であるので，母分散は（不偏でない）標本分散によって推定される.

A.5.3 モーメント法によるベルヌーイ分布の母数 p の推定

1 か 0 の値をとるベルヌーイ確率変数 X_i の期待値は，p を成功確率とすると

$$E(X_i) = p$$

である．ベルヌーイ分布 $Bi(1,p)$ に従う母集団からの標本 $X_1, X_2, \cdots, X_n$ は，それぞれ 1 か 0 の値をとるベルヌーイ確率変数であるので $X = X_1 + X_2 + \cdots + X_n$ とすると，確率変数 X は成功回数を示す．そして，母数 p の推定量は

$$\hat{p} = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

で与えられる．$\hat{p}$ は成功割合を示す確率変数で，X の従う二項分布 $Bi(n,p)$ の母数 p の推定量にもなっていることに注意.

K・ピアソンによって提唱されたモーメント法は，直感的に分かりやすく，また母平均や母分散の推定においては母集団分布を仮定する必要がないという利点がある．しかしながら，推定量のもつ望ましさの基準から見ると，必ずしもよい推定量を導出しないという批判が R・A・フィッシャーによってなされた．そして，これに代わってフィッシャーが提唱したのが**最尤法** (maximum likelihood method) である．

A.6 正規分布の母数 μ, σ^2 の最尤推定量

標本 $Y_1, Y_2, \cdots, Y_n$ がそれぞれ独立に正規分布 $N(\mu, \sigma^2)$ に従うとする．Y_i の実現値 y_i について，確率密度関数は

$$h(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}$$

である．そして $Y_1, Y_2, \cdots, Y_n$ についての同時確率密度関数は

$$h(y_1, y_2, \cdots, y_n|\mu, \sigma^2) = \prod_{i=1}^{n} h(y_i|\mu, \sigma^2)$$

である．ところで,

$$\log(h(y_i|\mu, \sigma^2)) = -\frac{\log(2\pi\sigma^2)}{2} - \frac{(y_i-\mu)^2}{2\sigma^2}$$

であるので，μ, σ^2 についての対数尤度関数は

$$\log L(\mu, \sigma^2 | y_1, y_2, \cdots, y_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

となる．この対数尤度関数にもとづき，母数のペア μ, σ^2 の最大化問題を解く．1 階の条件は，

$$\frac{\partial \log L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) = 0$$
$$\frac{\partial \log L(\mu, \sigma^2)}{\partial \sigma^2} = \frac{1}{2\sigma^2}\left(-n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right) = 0$$

である．この連立方程式を解けば，

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i = \mu^*, \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu^*)^2 = (\sigma^2)^*$$

を得る[*2]．実現値を確率変数に戻して推定量の形にすると，

$$\hat{\mu}_{ML} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad \hat{\sigma}^2_{ML} = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

となるのであって，結局，母平均 μ の最尤推定量は標本平均であり，母分散 σ^2 の最尤推定量は（不偏でない）標本分散に他ならない．

[*2]

$$f_{xx}(a, b) = \left.\frac{\partial^2 f(x, y)}{\partial x^2}\right|_{(x,y)=(a,b)}$$

のように書くことにすると，一般に (a, b) が関数 $f(x, y)$ の極大値であるための 2 階の条件は

$$f_{xx}(a, b) < 0, \qquad f_{yy}(a, b) < 0$$

かつ

$$\begin{vmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{vmatrix} = f_{xx}(a, b) f_{yy}(a, b) - (f_{xy}(a, b))^2 > 0$$

である．これを検証すると

$$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{n}{\sigma^2} < 0, \qquad \left.\frac{\partial^2 \log L}{\partial (\sigma^2)^2}\right|_{\sigma^2=(\sigma^2)^*} = -\frac{n}{2((\sigma^2)^*)^2} < 0, \qquad \left.\frac{\partial^2 \log L}{\partial \mu \partial \sigma^2}\right|_{\mu=\mu^*} = 0$$

より，2 階の条件も満たすことが分かる．

A.7　母分散 σ^2 の区間推定・検定

A.7.1　母分散 σ^2 の区間推定

母平均の区間推定ほど頻繁に利用するわけではないが，定理 7.3 を用いて母分散の区間推定もできる．$X_i \sim N(\mu, \sigma^2)$ のとき

$$\frac{(n-1)S^2}{\sigma^2}$$

は自由度 $n-1$ の χ^2 分布 $\chi^2(n-1)$ に従うので，$\chi^2(n-1)$ の右側 $\alpha\%$ 点（その値を超える確率が α になる点）を $\chi^2_\alpha(n-1)$ とすると，

$$P\left(\chi^2_{1-\alpha/2}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2}(n-1)\right) = 1-\alpha. \tag{A.42}$$

カッコ内の不等式を σ^2 について解くと，

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right) = 1-\alpha \tag{A.43}$$

となる．ここから，母分散 σ^2 の信頼区間を計算することができる．

A.7.2　母分散 σ^2 の検定

先の「母分散の区間推定」における議論を応用すると，母分散についての仮説検定もできる．

母分散について，

$$H_0:\ \sigma^2 = \sigma_0^2$$

を帰無仮説とする．このとき，

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

という特性を用いる．結論だけ述べておくと，両側検定の場合，

$$\left(S^2 < \frac{\sigma_0^2\chi^2_{1-\alpha/2}(n-1)}{(n-1)}\right) \vee \left(\frac{\sigma_0^2\chi^2_{\alpha/2}(n-1)}{(n-1)} < S^2\right)$$

が棄却域となり，片側検定の場合，

$$\frac{\sigma_0^2\chi^2_{\alpha}(n-1)}{(n-1)} < S^2$$

が棄却域となる．

A.8　回帰係数の最尤推定量

確率変数 $Y_1, Y_2, \cdots, Y_n$ がそれぞれ独立に正規分布 $N(\alpha+\beta x_i, \sigma^2)$ に従っているという仮定から，Y_i の実現値 y_i について，確率密度関数は

$$h(y_i|\alpha+\beta x_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma^2}\right\} \tag{A.44}$$

である．そして $Y_1, Y_2, \cdots, Y_n$ についての同時確率密度関数は

$$h(y_1, y_2, \cdots, y_n|\alpha+\beta x_i, \sigma^2) = \prod_{i=1}^{n} h(y_i|\alpha+\beta x_i, \sigma^2) \tag{A.45}$$

である．α, β についての対数尤度関数は

$$\begin{aligned}\log L(\alpha, \beta|y_1, y_2, \cdots, y_n) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{S_e}{2\sigma^2}\end{aligned} \tag{A.46}$$

となる．σ^2 を定数とすると，第一項は α, β に対して定数であるので，結局 $\log L(\alpha, \beta)$ の最大化問題と S_e の最小化問題は同じことである．

ゆえに，式 (13.6) の仮定のもとでは，最小二乗推定量と最尤推定量は等しい．

A.9　回帰分析の最小二乗推定量 $\hat{\alpha}$ の特徴

回帰分析の切片の最小二乗推定量として得られる $\hat{\alpha}$ の特徴を確認する．

A.9.1 期待値・分散

式 (13.11) より，$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$ である．また式 (13.6) の $Y_i = \alpha + \beta x_i + \varepsilon_i$ より，$\bar{Y} = \alpha + \beta\bar{x} + n^{-1}\sum_{i=1}^{n}\varepsilon_i$ なので，これを α の式に代入すると，

$$\hat{\alpha} = \alpha + \beta\hat{x} + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i - \hat{\beta}\bar{x} \tag{A.47}$$

を得る．$\hat{\alpha}$ の期待値は

$$E(\hat{\alpha}) = \alpha + \beta\hat{x} + E\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right) - \bar{x}E(\hat{\beta}) = \alpha \tag{A.48}$$

である．$\hat{\alpha}$ の分散は

$$\begin{aligned}
V(\hat{\alpha}) &= E[(\hat{\alpha} - \alpha)^2] \\
&= E\left[\bar{x}(\beta - \hat{\beta}) + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right]^2 \\
&= E[\bar{x}(\beta - \hat{\beta})]^2 + 2E\left[\frac{\bar{x}}{n}(\beta - \hat{\beta})\left(\sum_{i=1}^{n}\varepsilon_i\right)\right] + E\left(\sum_{i=1}^{n}\varepsilon_i\right)^2 \\
&= \bar{x}^2 V(\hat{\beta}) + \frac{1}{n^2}V\left(\sum_{i=1}^{n}\varepsilon_i\right) \\
&= \sigma^2\left(\frac{\bar{x}^2}{S_{xx}} + \frac{1}{n}\right)
\end{aligned} \tag{A.49}$$

である．ただし，

$$\begin{aligned}
E\left[\frac{\bar{x}}{n}(\beta - \hat{\beta})\left(\sum_{i=1}^{n}\varepsilon_i\right)\right] &= E\left[-\frac{\bar{x}}{n}\left(\sum_{i=1}^{n}w_i\varepsilon_i\right)\left(\sum_{i=1}^{n}\varepsilon_i\right)\right] \\
&= -\frac{\bar{x}}{n}\sum_{i=1}^{n}w_i E(\varepsilon_i^2) - \frac{\bar{x}}{n}\sum_{i\neq j}w_i E(\varepsilon_i\varepsilon_j) \\
&= -\frac{\bar{x}\sigma^2}{n}\sum_{i=1}^{n}w_i = 0
\end{aligned} \tag{A.50}$$

に注意すること．

A.9.2　α の信頼区間と仮説検定

$\hat{\beta}$ と同様にして $\hat{\alpha}$ の標本分布は

$$\hat{\alpha} \sim N\left(\alpha, \sigma^2\left(\frac{\bar{x}^2}{S_{xx}} + \frac{1}{n}\right)\right) \tag{A.51}$$

である．σ^2 の推定量として $\hat{\sigma}^2$ を用いて，$100(1-\gamma)\%$ 信頼区間は

$$\left[\hat{\alpha} - t_{\gamma/2}(n-2)\sqrt{\hat{\sigma}^2\left(\frac{\bar{x}^2}{S_{xx}} + \frac{1}{n}\right)}, \hat{\alpha} + t_{\gamma/2}(n-2)\sqrt{\hat{\sigma}^2\left(\frac{\bar{x}^2}{S_{xx}} + \frac{1}{n}\right)}\right]$$

によって計算できる．また，帰無仮説，対立仮説をそれぞれ

$$H_0 : \beta = 0, \qquad H_1 : \beta \neq 0 \tag{A.52}$$

としたときの検定統計量は

$$t^*_{\hat{\alpha}} = \frac{\hat{\alpha}}{\sqrt{\hat{\sigma}^2\left(\frac{\bar{x}^2}{S_{xx}} + \frac{1}{n}\right)}} \sim t(n-2). \tag{A.53}$$

である．

A.10　回帰分析の決定係数とモデルの適合度検定

A.10.1　決定係数

式 (13.23) より，

$$Y_i - \bar{Y} = e_i + (\hat{Y}_i - \bar{Y}) \tag{A.54}$$

である．この両辺を 2 乗して総和をとると，

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} e_i^2 + 2\sum_{i=1}^{n} e_i(\hat{Y}_i - \bar{Y}) \\ &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} e_i^2 \end{aligned} \tag{A.55}$$

となる．ここで，

$$\sum_{i=1}^{n} e_i(\hat{Y}_i - \bar{Y}) = \hat{\alpha}\sum_{i=1}^{n} e_i + \hat{\beta}\sum_{i=1}^{n} e_i x_i - \bar{Y}\sum_{i=1}^{n} e_i = 0 \tag{A.56}$$

を用いた．式 (A.55) は標本データ Y の分散の分解と見なすことができる．つまり，全変動 (total sum of squares; TSS) を回帰値の変動 (explained sum of squares; ESS) と残差変動 (residual sum of squares; RSS) に分解している．

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} e_i^2 \tag{A.57}$$

$$\text{全変動 (TSS)} = \text{回帰値の変動 (ESS)} + \text{残差変動 (RSS)} \tag{A.58}$$

回帰モデルの適合度指標としての**決定係数** (coefficient of determination)R^2 は，次のように定義される，すなわち，

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{S_e}{S_{YY}}. \tag{A.59}$$

R^2 の特性として，$0 \leq R^2 \leq 1$ がいえる．また，式 (A.56) を用いると，

$$\begin{aligned} \text{ESS} = S_{\hat{Y}\hat{Y}} &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \\ &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})(e_i + \hat{Y}_i - \bar{Y}) \\ &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})(Y_i - \bar{Y}) = S_{Y\hat{Y}} \end{aligned} \tag{A.60}$$

である．つまり，ESS は Y と回帰値 $\hat{Y}$ の共分散でもある．ここから，

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{(S_{Y\hat{Y}})^2}{S_{\hat{Y}\hat{Y}} S_{YY}} = (r_{Y\hat{Y}})^2 \tag{A.61}$$

がいえる．つまり，決定係数は Y と回帰値 $\hat{Y}$ の相関係数の 2 乗である．だから R^2

という．単回帰の場合とくに，

$$\begin{aligned}\mathrm{ESS} = S_{\hat{Y}\hat{Y}} &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \\ &= \sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}x_i - \bar{Y})^2 \\ &= \hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 = \hat{\beta}^2 S_{xx} \end{aligned} \tag{A.62}$$

より，$R^2 = (r_{xY})^2$ でもある．

独立変数が p 個ある重回帰分析では，S_e の自由度 $n-p-1$ と S_{YY} の自由度を考慮した**自由度調整済み決定係数** (adjusted R^2) を用いる．自由度調整済み決定係数を $\bar{R}^2$ とすると，

$$\bar{R}^2 = 1 - \frac{S_e/(n-p-1)}{S_{YY}/(n-1)} \tag{A.63}$$

である．

A.10.2　モデルの適合度検定

p 個の独立変数を持つ重回帰モデル

$$Y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \tag{A.64}$$

について，そのうちの k 個の変数 $x_{p-k+1}, x_{p-k+2}, \cdots, x_p$ が Y の変動に影響を与えないという帰無仮説を設定する．すなわち，

$$H_0 : \beta_{p-k+1} = \beta_{p-k+2} = \cdots = \beta_p = 0, \tag{A.65}$$

$$H_1 : \exists h \in \{p-k+1, p-k+2, \cdots, p\}, \beta_h \neq 0. \tag{A.66}$$

帰無仮説における残差平方和を RSS_0，対立仮説におけるそれを RSS_1 とすると，帰無仮説が正しい場合それらの差は 0 に近いはずである．そこで次のような検定統計量を定義する．

$$F = \frac{(\mathrm{RSS}_0 - \mathrm{RSS}_1)/k}{\mathrm{RSS}_1/(n-p-1)} = \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}\frac{n-p-1}{k} \tag{A.67}$$

この統計量は F 統計量 (F statistics) もしくは分散比 (variance ratio) と呼ばれ，帰無仮説の下で自由度 $(k, n-p-1)$ の F 分布に従う．そこで有意水準を α とおいたとき，F の値が $F_\alpha(k, n-p-1)$ を上回るとき帰無仮説を棄却し，下回るとき帰無仮説を棄却しない，という手順で検定を行う．

帰無仮説として，すべての独立変数が Y の変動に影響を与えない，つまり

$$H_0 : \forall h \in \{1, 2, \cdots, p\}, \beta_h = 0 \tag{A.68}$$

とおく．帰無仮説のモデルをヌル・モデル (null model) という．このとき，$\hat{Y}_i = \hat{\alpha} = \bar{Y}, e_i = Y_i - \bar{Y}$ なので $\mathrm{RSS}_0 = \mathrm{TSS}$ であり，$\mathrm{RSS}_1 = \mathrm{RSS}$ である．ゆえに，F 統計量は

$$F = \frac{(\mathrm{TSS} - \mathrm{RSS})/p}{\mathrm{RSS}/(n-p-1)} = \frac{R^2/p}{(1-R^2)/(n-p-1)} = \frac{R^2}{1-R^2}\frac{n-p-1}{p} \tag{A.69}$$

となる．とくに，$p = 1$ の単回帰モデルの場合，

$$F = \frac{\mathrm{ESS}}{\mathrm{RSS}/(n-2)} = \frac{\hat{\beta}^2 S_{xx}}{S_e/(n-2)} = \frac{\hat{\beta}^2}{\hat{\sigma}^2/S_{xx}} = \left(t^*_{\hat{\beta}}\right)^2 \tag{A.70}$$

となり，$H_0 : \beta = 0$ のときの t 統計量の 2 乗に等しくなる．一般的に，t 分布と F 分布の関係から，確率変数 t が自由度 $n-2$ の t 分布に従うとき，t^2 は自由度 $(1, n-2)$ の F 分布に従う．つまり，$p = 1$ の単回帰モデルの場合，β の有意性検定とモデルの適合度検定は本質的に同じである[*3]．

[*3] どちらの帰無仮説・対立仮説も同じであり，用いる検定統計量と分布が異なるが，p 値は同じになる．