

経済統計学

竹内 清 著

丸善株式会社

經濟統計学

竹内 清 著

丸善株式会社

著者の略歴

現職 東北大学経済学部教授
昭和 27 年 東京商科大学卒業

経 済 統 計 学

¥ 2,500

昭和 54 年 11 月 30 日 発 行

©1979

著 作 者 ^{たけ}竹 ^{うち}内 ^{きよし}清

発 行 者 飯 泉 新 吾

発 行 所 丸 善 株 式 会 社

郵便番号 103 東京都中央区日本橋二丁目 3 番 10 号

印刷 日東紙工株式会社・製本 切石製本株式会社

3034—2415—7924

序 文

本書は、経済統計分析の基礎についてまとめたものである。従来、経済統計学の書物としては、主として経済統計データを対象としたデータ論の立場からまとめたものと、素材たる経済統計データを分析するための統計的方法に主眼をおいてまとめたものに大別することができるであろう。経済統計学の講義を通じて得た経験からすると、データ論の立場に傾斜した場合、学生の側からするとある程度平板的な感じをもち、退屈さを感じることが少なくないようである。それを補うには、適宜、統計的分析法を組み入れることが必要である。第一の視点に重点をおいた経済統計学の教科書あるいは講義は、大学院学生あるいは実務家などのようにある程度の子備知識をもった者に適している面をもつように思われる。このような意味で、本書は、第一の視点をふまえつつ、主として第二の立場にたってまとめたものである。

主として経済統計データを素材として統計的に分析する場合、それらのデータが、どの機関により、どのような目的で、いつどこで、どのような調査方法を用いて収集、整理されたものなのか、対象たるデータの諸性質についての的確な検討をふまえることなしに、形式的に統計的方法を適用することは危険である。本書は、このことを念頭におきつつ、統計データを作る立場を含め、主としてこれを利用、分析する立場にたってまとめられている。

本書の対象読者層としては、経済学を含む社会科学系の大学の専門課程に在学中の学生諸君ならびに会社や官庁などに勤めている実務家の人々を主として念頭においているが、大学などにおける一般の統計学の教科書としても利用で

きるように、内容、配列を考えている。経済統計学を講義してきた経験からすると、最初の段階で統計学一般の基礎知識を整理しておいた方が、後段の講義にとって好都合である。このような観点から、また本書を一応自己完結的な形のものにするために、第1章と第2章では、経済観測データのまとめと関連して、記述統計の基礎的部分である度数分布および度数分布特性値の問題がそれぞれまとめられる。第1章では累積分布の応用として、所得分布などのように社会、経済の分野における集中現象の分析と関連してローレンツ曲線の問題が考察される。

第3章では指数の問題を取りあげ、主として現在わが国や諸外国で利用されている代表的な経済指数である物価指数や数量指数などを対象として、その構成法や利用法を考察している。

第4章では確率の問題を取り扱うが、まず確率の種類、性質、演算規則などを考察し、さらに確率変数、分布関数、確率密度関数、確率変数の分布特性値などについてまとめる。なお実際の応用に当たり重要な役割を果たす大数の法則についてふれるが、この章は後章の基礎的部分を構成する。第5章では既知のいくつかの確率分布について考察するが、まず離散型の確率分布として二項分布、超幾何分布、多項分布、ポアソン分布などを取りあげ、それらの関係を明らかにしている。ついで連続型の確率分布として、統計学で最も基本的な役割を演ずる分布である正規分布のほか、一様分布、対数正規分布、パレート分布などが考察される。

第6章では、標本の関数として表わされる確率変数である統計量のしたがう確率分布である標本分布を考察する。まず大標本の場合に重要な働きをする中心極限定理を、実際の応用の観点から考察し、ついで統計分析において基本的な役割を果たしている χ^2 分布、 t 分布、 F 分布の性質を要約した上で、平均値、平均値の差および比率についての標本分布を導出する。この章は、以下の章における推定、検定、分散分析や回帰分析などの基礎を構成する。第7章では点推定と区間推定を中心として、推定の問題を考察する。ここではまず、点推定量としてもつべき望ましい性質をまとめ、それらの推定量を見つけ出す方

法として、最尤法と最小二乗法を中心に考察する。ついで区間推定の問題と関連して、平均値、平均値の差、比率および比率の差についての信頼区間の設定を導いている。

第8章では検定の問題を取り扱うが、まず統計的検定の考え方の基本構造について、平均値の検定の問題を例にとりながら、やや詳しく考察する。大標本の場合と小標本の場合に分けながら、平均値および平均値の差の検定の問題を考えるが、小標本の場合の後者の問題と関連して分散の比の検定についても言及する。ついで比率および比率の差の検定の問題を考えた後、適合度の検定について、対数正規分布と関連した問題を取りあげ、若干の側面からこれを考察する。これとの関連で、正規確率紙あるいは対数正規確率紙を用いた分析も例示する。第9章では分散分析法の問題を、母数模型にもとづいて一元配置法および二元配置法の基本について考察する。

第10章では、線形回帰モデルを中心として、2変数の場合の単純回帰分析および3変数の場合の多元回帰分析の問題を考察する。ここでは、最小二乗法による回帰パラメーターの推定および推定量の信頼度を中心に議論が展開される。第11章では標本調査の問題を取りあげる。経済統計データは標本調査を通じて得られたものが多いが、その最も基本的なものとして単純無作為抽出法をまず考察する。これと関連して平均や総計、あるいは比率の推定の問題のほか、標本の大きさの決定の問題が取り上げられる。ついで補助情報の活用による推定や検定の精度をあげるための層別抽出法を考え、その中で最も簡単でしかも実際上有効な比例抽出法を考察する。さらにネイマンの最適割当法を取りあげ、これらを通じて層別の効果の測定方法などを分析する。

本書では、紙幅の関係もあり、経済統計学の中で重要な問題領域である時系列の問題は割愛せざるを得なかった。最後の参考文献にあげた拙著（『需要予測入門』、丸善、昭和46年）を参照して頂ければ幸いである。またここ数十年間急速に発展し、統計学の体系の中で一つの独立の学派、分野を形成するに至っているベイズ統計学の立場からの接近法についても、その詳細は割愛せざるを得なかった。特に意思決定の観点からは、ベイズの接近法は極めて

有用であるが、これについての展開は別の機会にゆずりたい。なお本書の性格からして、計量経済学的接近法についての展開は省略するとともに、数学的な厳密さはある程度犠牲にして平易さを主に考えていることをつけ加えておきたい。

本書が成るに当たっては、丸善出版部の多くの方々の暖かい励ましと御協力の負うところが極めて大きい。ここにあらためて厚く御礼を申し上げる次第である。

昭和54年9月

竹 内 清

目 次

1. 経済観測データのまとめ方

—度数分布—

1・1 はじめに	1
1・2 度数分布	3
1・2・1 度数分布表	4
1・2・2 度数分布図	8
1・2・3 度数分布の形	11
1・3 ローレンツ曲線	13

2. 経済観測データのまとめ方

—度数分布特性値—

2・1 平均値	17
2・1・1 算術平均	18
2・1・2 幾何平均	20
2・1・3 メディアン（中位数）	21
2・1・4 四分位数	24
2・2 ち ら ば り	26
2・2・1 絶対的なちらばりの測度	27

2・2・2 相対的なちらばりの測度	28
2・3 ひ ず み	29
2・4 と が り	30
2・5 チェビシェフの不等式	31

3. 指 数

3・1 指 数 の 概 念	35
3・2 指 数 の 構 成 法	35
3・2・1 比率の単純平均	36
3・2・2 比率の加重平均	38
3・3 物価指数の経済学的意味	44
3・4 地域間の物価水準および数量水準の比較	45
3・5 連 鎖 指 数	48
3・6 指数の構成，利用に際しての留意点	49

4. 確 率

4・1 確 率 の 定 義	51
4・1・1 先 験 的 確 率	51
4・1・2 経 験 的 確 率	52
4・1・3 幾何学的確率	53
4・2 確 率 の 性 質	54
4・3 確 率 変 数	60

4・3・1 確 率 変 数	60
4・3・2 分 布 関 数	61
4・3・3 確率密度関数	62
4・4 確率変数の分布特性値	64
4・4・1 数学的期待値	64
4・4・2 分 散	67
4・4・3 積率 (モーメント)	70
4・5 大 数 の 法 則	72
4・6 2次元の分布関数	74
4・6・1 2次元確率変数と分布関数	74
4・6・2 2次元確率密度関数	75
4・6・3 条件付き期待値	76
4・6・4 相 関 積 率	77

5. 確 率 分 布

5・1 離散型の確率分布	79
5・1・1 二 項 分 布	79
5・1・2 超 幾 何 分 布	84
5・1・3 多 項 分 布	87
5・1・4 ポアソン分布	88
5・2 連続型の確率分布	92
5・2・1 正 規 分 布	92
5・2・2 一 様 分 布	96
5・2・3 対数正規分布	97
5・2・4 パレート分布	100

6. 標 本 分 布

6・1 標本分布の性質	103
6・2 中心極限定理	104
6・3 χ^2 分 布	106
6・4 t 分 布	109
6・5 F 分 布	112
6・6 平均値の標本分布	113
6・6・1 標本平均の期待値	113
6・6・2 標本平均の標準誤差	114
6・7 平均値の差の標本分布	116
6・7・1 平均値の差の期待値	116
6・7・2 平均値の差の標準誤差	116
6・8 比率の標本分布	117

7. 推 定

7・1 統計的推定の考え方	119
7・2 点 推 定	120
7・2・1 推定量としてもつべき望ましい性質	120
7・2・2 推定量の見つけ方	125
7・3 区 間 推 定	131
7・4 平均値の信頼区間	132

7・4・1 大標本の場合	132
7・4・2 小標本の場合	134
7・5 平均値の差の信頼区間	135
7・5・1 大標本の場合	135
7・5・2 小標本の場合	136
7・6 比率の信頼区間	137
7・7 比率の差の信頼区間	138

8. 検 定

8・1 統計的検定の考え方	141
8・2 平均値の検定	143
8・2・1 大標本の場合	143
8・2・2 小標本の場合	146
8・3 平均値の差の検定	147
8・3・1 大標本の場合	147
8・3・2 小標本の場合	149
8・4 分散の比の検定	149
8・5 比 率 の 検 定	152
8・6 比率の差の検定	153
8・7 適合度の検定	155
8・7・1 検 定	155
8・7・2 コルモゴロフの適合度基準	159
8・7・3 正規確率紙による検定	160

9. 分散分析法

9・1 分散分析法の考え方	163
9・2 一元配置法による分散分析	164
9・3 二元配置法による分散分析	171

10. 回 帰 分 析

10・1 問題の定式化	179
10・2 単純線形回帰モデル	181
10・2・1 最小二乗法によるパラメーターの推定	181
10・2・2 相 関 係 数	183
10・2・3 最小二乗推定量の信頼度	184
10・2・4 数 値 例	188
10・3 対数線形回帰モデル	189
10・4 多元回帰分析	191
10・4・1 最小二乗推定量の不偏性	192
10・4・2 偏相関係数と重相関係数	194
10・4・3 偏回帰係数の信頼度	195
10・4・4 数 値 例	195

11. 標 本 調 査

11・1 単純無作為抽出法	199
11・1・1 単純無作為抽出法の概念	199
11・1・2 乱数表の使い方	201

11・2 単純無作為抽出における推定	204
11・2・1 平均の推定	204
11・2・2 総計の推定	206
11・2・3 比率の推定	208
11・3 標本の大きさの決定	211
11・4 層別抽出法	213
11・4・1 層別抽出法の内容	213
11・5 比例抽出法	216
11・5・1 標本の大きさの決定	218
11・5・2 ネイマンの最適割当法	219
11・5・3 層別の効果	220
11・6 多段抽出法	222
付 表	225
付表 1 標準正規分布表	225
付表 2 t 分布表	226
付表 3 χ^2 分布表	227
付表 4 F 分布表 (1) ($\alpha=0.05$)	228
付表 5 F 分布表 (2) ($\alpha=0.025$)	229
付表 6 F 分布表 (3) ($\alpha=0.01$)	230
参 考 文 献	231
索 引	235

1

経済観測データのまとめ方 ——度数分布——

1・1 はじめに

本章では、経済データを直接的に作成する立場だけでなく、既成の経済データを利用する立場を含めて、経済観測データをまとめる場合の要点を考察する。官庁統計、その他二次的な既成の経済データを利用する立場にたっても、それらがどのような目的で、どこの機関により、何時、どのような調査方法を適用して収集されたのか、またどのような仕方でどのようにデータがまとめられているのか、等々についてその基本を的確に理解することなしに、経済データの十全な利用は考えられないであろう。ここで記述する方法は、たんに経済データだけでなく、一般のデータの場合にも当然適用可能であることはいうまでもない。分析、決定の基礎は、データのまとめ方にある。データのまとめ方の当否は、分析、決定の当否につながるといえよう。

統計的な経済データを実際に分析し、意思決定に役立てるためには、第一段階の手順として、まず調査目的ないし利用目的にしたがって、分析対象となっている統計集団の特性を表わす標識——単数または複数個——に関してデータを要約、整理することが必要である。ところで集団の特性を表わす標識としては、質的標識と量的標識——変量ともよばれる——に分類できる。たとえば、ある都市における小売店を対象の集団と考えた場合、これを都市部とか郊外といった地域別や売上高によって分類することができるであろうが、前者は質的

標識による分類の例であり、後者は量的標識による分類の例である。実際の分析に当たっては、調査目的や利用目的あるいは問題の性質に応じて、質的標識との関連において量的標識——変量——を分析することが必要である。変量のまとめには度数分布 (frequency distribution) が有用である。

統計的分析では、調査対象全体の集まりと調査対象のもつ一定の特性に関する観測値全体の集まりを、概念上区別しておくことと混乱の危険を避けることができるであろう。前者を祖集団 (universe) といい、後者を母集団 (population) という。たとえば、ある都市における世帯全体の集まりは対象そのものの全体集団であり、祖集団を構成し、それぞれの世帯の一定期間の消費支出とか世帯人員などの量的標識についての観測値の全体の集まりは母集団を構成する。しかし前後の関係などから混同の恐れのないときは、母集団によってどちらをも表わすことにする。

統計的推測の観点からは、母集団と標本 (sample) の概念を明確に区別することが基本的に重要である。フィッシャー (R.A. Fisher) は、すでに 1922 年の論文において、当時の統計学界の混乱の原因の一つとして、この点を強調している¹⁾。標本は、母集団を構成する要素の部分的な集まりであるが、通常、標本というときは、確率的な抽出過程を通じて構成されるものを指す。一般には、標本というときは、母集団を構成する要素を確率的に無作為に (ランダムに) 抽出して構成される部分集団を指すが、時として、個々の要素または個々の観測値そのものをさすこともあるので、どちらの意味で使われているかを、前後関係から判断することが必要である。無作為抽出された標本は、無作為標本 (random sample) または確率標本 (probability sample) とよばれる。

母集団全体を調査する方式は全数調査またはセンサス (census) とよばれる。官庁統計では、たとえば、総理府統計局の国勢調査や事業所統計調査、通産省の工業統計調査や商業統計調査、農林水産省の農業センサス、その他は全数調査の典型例である。全数調査は一般に時間と費用がかかるものであり、特別の

1) R.A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society*, A, 222, 1922, pp. 310-311.

目的以外には、官庁統計などでも、標本を用いて調査する方式である標本調査 (sample survey) を採用することが多い。たとえば、官庁統計では、総理府統計局の家計調査、農林水産省の農家経済調査や漁家経済調査、経済企画庁の消費と貯蓄の動向調査、その他は標本調査の代表例である。後の章で標本抽出や標本調査についてより詳しく考察する。

母集団の分布の特性値、たとえば、平均値 (mean, average)、比率 (proportion)、分散 (variance)、標準偏差 (standard deviation)、相関係数 (coefficient of correlation) などは、パラメーターまたは母数 (parameter) とよばれる。標本についてのそれらの特性値は統計量 (statistic) とよばれる。

いわゆるネイマン＝ピアソン (J. Neyman - E.S. Pearson) 流の統計的推理においては、パラメーターは、特定の実際の問題では、通常未知ではあっても、一定と考えられる (しかしベイズ統計学では、これは一定の事前分布にしたがう確率的な変量と想定して、議論を進める)。統計量は一般に標本ごとに異なった値をとり、標本抽出に伴う変動をもっている。

統計的推論では、母集団と標本を明確に区別することを背景として、一般にパラメーターにはギリシャ文字を、また統計量にはローマ字の小文字を用いることが多い。

記述統計学的接近では、母集団と標本を概念上区別しないで、統計データを処理、分析するが、そこでの基本的な手法は、母集団と標本の概念を明確に規定する推測統計学においても、基本的な働きをすることはいうまでもない。第1章および第2章では記述統計学におけるデータのまとめ方を踏まえつつ、主として変量のまとめ方に関連して、1変量の場合の度数分布を中心とした問題を取り扱うことにする。

1・2 度数分布

調査、観測を通じて集団に関する情報、データが得られた場合、集団の量的構造を記述するためには、通常、度数分布表 (table of frequency distribution)