

データの入手から計量分析まで

北村 友宏

2020年4月30日改訂

1 はじめに

実際のデータを入手して分析し、その結果から何かを明らかにすることを実証分析といいます。これは学部
の卒業論文、大学院の修士論文・博士論文、その他学術論文などの執筆のための研究において有用なアプ
ローチの1つです。価格、需要、生産、消費、所得など、経済データの実証分析をするための計量経済学という学
問分野があり、その手法を用いた実証分析を計量分析、あるいは計量経済分析といいます。

計量分析は、データを入手・整理し、それをパソコンの統計解析ソフトに読み込んで実行します。本稿で
は、計量分析を実行したことのない読者を想定し、データの入手から計量分析までの具体的な方法を、簡単な
例を用いて解説します。本稿で示した手順に従って作業をすれば、一連の計量分析を練習できるようになっ
ています。なお、本稿の内容は加藤（2012, 2019）および鹿野（2015）に基づいています。

2 分析例

経済理論の1つに、「所得の一定割合が消費に振り向けられる。ただし、所得がないときもある程度の消費
が必要である」という考え方があります。 c を消費額、 y を所得額として、この考え方を数式で表すと、

$$c = \beta_0 + \beta_1 y \quad (1)$$

のように、 c を y の1次関数として表すことができます。この関数をケインズ型消費関数といいます。 β_0 は
基礎消費と呼ばれ、たとえ所得がゼロであっても生活のために（借金しながら）どの程度消費する必要がある
かを表します。また、 β_1 は限界消費性向と呼ばれ、所得額が1単位増えると消費額が何単位変化するかを表
します。

このケインズ型消費関数が、現実世界にも当てはまるかどうかを検証するのが、計量分析の1つです。本稿
では、2014年の日本の47都道府県における消費額 c と所得額 y の数値データを用い、ケインズ型消費関数
の基礎消費 β_0 と限界消費性向 β_1 を求めます。どちらもゼロでないことが分かれば、ケインズ型消費関数が
2014年の日本にも当てはまると考えて差し支えないでしょう。

しかしながら、横軸に y 、縦軸に c をとった平面上に、現実の c と y の値の組み合わせの点が(1)式に従っ
て一直線上に並ぶわけではありません。ほぼ必ず、一次関数の直線からのズレが発生します。そこで、(1)式
の左辺の変数 c の値の、1次関数の直線からのズレ（乖離）を表す誤差項（攪乱項） u を(1)式の右辺に追加
して、

$$c = \beta_0 + \beta_1 y + u \quad (2)$$

のように考えます。この (2) 式に通常の最小二乗法 (Ordinary Least Squares, OLS) を適用し、 β_0 と β_1 を推定します。

3 統計解析ソフトの準備

前節の計量分析を実行するために、パソコンで統計解析ソフトを使用できる状態にしましょう。本稿では、無料で利用でき、日本語にも対応している gretl を紹介します。同ソフトは、gretl の公式 HP (<http://gretl.sourceforge.net/>) からダウンロードできます。Windows 版と Mac 版が用意されていますので、お使いの OS に対応したものをダウンロード・インストールしてください。手順は、どちらの OS でもほとんど同じです。本稿では詳細な説明は省略しますが、詳細は、加藤 (2012, 2019) や鹿野 (2015) をご参照ください。ただし、加藤 (2012) は、当時の日本語非対応版での解説となっています。

以下では、Windows 用の gretl 2020b (2020 年 4 月 30 日時点での最新版) の使用を前提として解説します。

4 データの入手

2014 年の日本の 47 都道府県における所得と消費のデータは『平成 26 年全国消費実態調査』に収録されており、政府統計の総合窓口 e-Stat の HP から探してダウンロードすることができます。ここでは、2014 年の各都道府県における、勤労者世帯の可処分所得額と消費支出額 (ともに単位は円) の 1 世帯当たり月間平均データを用います。それでは、以下の作業を順に実行してみましよう (2020 年 4 月 30 日時点での探し方です)。

1. デスクトップに新しいフォルダを作成し、「消費関数推定」という名前を付ける (全角日本語使用可)。
2. 政府統計の総合窓口 e-Stat の HP (<https://www.e-stat.go.jp/>) にアクセス。
3. 「統計データを探す」→「分野」→「企業・家計・経済」→「全国消費実態調査」→「平成26年全国消費実態調査の「ファイル | 件数 更新日」で、「2,408 件 2017-12-27」と記載されているセルをクリック→家計収支に関する結果 [121 件] の「総世帯 [10 件]」と辿る。
4. 地域編、表番号 13 の地域別 1 世帯当たり 1 か月間の収入と支出の勤労者世帯の「EXCEL」をクリックし、「消費関数推定」フォルダに保存。ファイル名が ra01302.xls となっていることを確認。

5 データの整理

入手したデータは、そのままでは統計解析ソフトを用いた分析に使えません。そこで、データの整理が必要となります。イメージとしては、表 1 のような Excel ファイルのデータセットを作成します。

表 1 において、最も左の列の 1,2,3, ..., 48 は Excel ファイルの行番号、最も上の行の A,B,C,D は Excel ファイルの列記号だと考えてください。続いて表 1 の、各セルの内容をご覧ください。Excel ファイルの 1 行目には変数名が入っています。2 行目には 1 番目の個体 (ここでは北海道) の各変数の数値、3 行目には 2 番目の個体 (ここでは青森県)、4 行目には 3 番目の個体、...といった規則となっています。上記のような形に整理すればよいのです。ただし、変数名や個体名 (都道府県名) を含め、全て半角英数字で入力しましょう。というのは、現行バージョンの gretl は日本語対応ですが、セルに全角日本語が入力された Excel ファイルを読み込むと文字化けするからです。以下の作業を順に最後まで実行すれば、gretl に読み込める形への整理が完了します。

表 1 整理後の Excel ファイルの形

	A	B	C	D
1	id	prefecture	expenditure	income
2	1	Hokkaido	262,135	322,712
3	2	Aomori	239,223	306,329
⋮	⋮	⋮	⋮	⋮
48	47	Okinawa	211,968	271,543

1. Excel を起動し、所得と消費.xlsx という名前で新規ファイルを保存（全角日本語使用可）。
2. セル A1 に id, セル B1 に prefecture, セル C1 に expenditure, セル D1 に income と入力。
3. ra01302.xls を開き、セル AJ16 “Hokkaido” をコピー。
4. 所得と消費.xlsx のセル B2 を選択して貼り付け。
5. ra01302.xls の「セル AK15 からセル CD15 まで」（北海道以外の都府県名）の範囲をコピー。
6. 所得と消費.xlsx のセル B3 を選択し、行列を入れ替えて貼り付け。
7. 「セル A2 からセル A48 まで」のセル 47 個にそれぞれ 1, 2, …, 47 の通し番号を入力（オートフィル機能を使うと楽）。
8. ra01302.xls の「セル AJ89 からセル CD89 まで」（各都道府県の消費支出）の範囲をコピー。
9. 所得と消費.xlsx のセル C2 を選択し、行列を入れ替えて貼り付け。
10. ra01302.xls の「セル AJ86 からセル CD86 まで」（各都道府県の可処分所得）の範囲をコピー。
11. 所得と消費.xlsx のセル D2 を選択し、行列を入れ替えて貼り付け。
12. 所得と消費.xlsx を上書き保存して閉じる。

6 統計解析ソフトへの読み込み

ここでいよいよ、統計解析ソフト gretl の出番です。gretl では、表 1 の形に整理された Excel ファイル (xls, xlsx 両方) を読み込むことができます。以下の作業を順に実行してみましょう。

1. gretl を起動。
2. 所得と消費.xlsx を、gretl の画面にドラッグ・アンド・ドロップ。
3. 出てきたダイアログボックスの、インポートを開始する場所: の列: と行: がともに 1 になっていることを確認し、「OK」をクリック。
4. 「インポート可能なシートを 1 個見つけました」で始まるメッセージが表示されるので、「閉じる」をクリック。
5. 「インポートされたデータは…（中略）…解釈し直しますか？」というメッセージが表示されるので、「いいえ」をクリックすると、データが読み込まれる。
6. 「id」から「income」までの 4 つをドラッグして選択し、その上で右クリック→「データ（値）を表示」と操作すると、全変数の観測値リストが新規ウィンドウにて表示される。確認したら閉じる。
7. メニューバーから「ファイル」→「データに名前を付けて保存」と操作し、所得と消費.gdt という名前で「消費関数推定」フォルダに保存（全角日本語使用可）。

図1 表示された推定結果

gretl: モデル

ファイル 編集(E) 検定(D) 保存(S) グラフ(G) 分析(A) LaTeX

モデル 1

モデル 1: 最小二乗法(OLS), 観測: 1-47
 従属変数: expenditure
 不均一分散頑健標準誤差, バリエーション HC1

	係数	標準誤差	t値	p値	
const	51588.4	17189.0	3.001	0.0044	***
income	0.624469	0.0482975	12.93	9.31e-017	***
Mean dependent var	272055.0	S.D. dependent var	20787.01		
Sum squared resid	4.74e+09	S.E. of regression	10285.10		
R-squared	0.781440	Adjusted R-squared	0.756139		
F(1, 45)	167.1754	P-value(F)	9.31e-17		
Log-likelihood	-499.7839	Akaike criterion	1003.568		
Schwarz criterion	1007.268	Hannan-Quinn	1004.960		

7 計量分析

いよいよ (2) 式のケインズ型消費関数に通常最小二乗法を適用し、基礎消費と限界消費性向を求めます。以下の作業を順に実行しましょう。

1. メニューバーから「モデル」→「通常最小二乗法」と操作。
2. 出てきたウィンドウ左側の変数リストにある expenditure をクリックし、3つの矢印のうち上の青い右向き矢印をクリック。これで、推定式の左辺の変数（被説明変数，従属変数）が expenditure（消費支出）となる。
3. ウィンドウ左側の変数リストにある income をクリックし、3つの矢印のうち真ん中の緑の右向き矢印をクリック。これで、推定式の右辺の変数（説明変数，独立変数）が income（可処分所得）になる。最初から説明変数リストに入っている const は推定式の切片（定数項）のことで、ここでは基礎消費を指す。
4. 「頑健標準誤差を使用する」にチェック。これで、(2) 式の誤差項 u のバラつき（分散）に関する仮定が誤っていても、より厳密な分析ができるようになる。
5. 「OK」をクリックすると、結果が図1のように、新しいウィンドウに表示される。
6. 表示された「gretl: モデル 1」のウィンドウのメニューバーから「ファイル」→「名前を付けて保存」と操作。
7. 「標準テキスト」を選び、「OK」をクリック。
8. 消費関数推定結果.txt という名前で「消費関数推定」フォルダに保存。すると、表示された推定結果をそのままテキストファイルで保存できる。

これで、ケインズ型消費関数の推定結果を出力・保存できました。図1の画面を見ながら、結果を読み取りましょう。まずは income の行をご覧ください。係数が 0.624469 となっています。income の係数は (2) 式の

y の係数 β_1 で、前述のとおり β_1 は限界消費性向を表すので、限界消費性向が約 0.62 と推定されました。所得・消費両変数の単位は円ですから、所得額が 1 円高くなると、消費額が約 0.62 円増加する傾向がある、という解釈ができます。

ただし、この 0.62 という値は、あくまで推定値ですので、真の限界消費性向も 0.62 であるとは限りません。今回用いた 2014 年の都道府県別データにおける可処分所得と消費支出の数値は、データとして観測される前はどのような値になるかが分からず、さまざまな値をとり得る中でたまたま実現した数値です。よって、そのデータを用いて推定した限界消費性向も、「ゼロを中心にバラつき、たまたま 0.62 になった」という可能性があります。

そこで仮説検定を行い、限界消費性向の真の値がゼロかどうかを検証します。図 1 の画面の、income の行の右のほうを見ると、 t 値が 12.93、 p 値が $9.31e-017$ （これは 9.31×10^{-17} という意味）となっています。これは、仮に「income の係数（限界消費性向）がゼロ」だとすると、12.93 という t 値は 9.31×10^{-17} 、つまりほぼ 0% の確率でしか出てこないことを意味します。正確には、「 t 値の絶対値が 12.93 以上になる確率」が 9.31×10^{-17} ということです。よって「income の係数（限界消費性向）がゼロ」とは考えにくく、限界消費性向の真の値はゼロから離れていると判断してよいでしょう。このことを「限界消費性向は統計的に有意にゼロと異なる」などと表現します。もし p 値が 0.1 を超えていれば、すなわち 10% を超える確率でその t 値が出てくるなら、その確率は小さすぎないので「真の係数がゼロ」もあり得ると判断することが（計量分析では）多いです。本稿では結果の読み方の説明にとどめますので、仮説検定の詳細は、例えば鹿野（2015）や加藤（2019）を参照してください。

続いて、図 1 の画面の const の行をご覧ください。係数が 51588.4 と表示されています。const は推定式の切片（定数項）ですから、(2) の β_0 に相当し、前述のとおり β_0 は基礎消費を表すので、基礎消費が 51,588.4 円と推定されました。たとえ所得が 0 円であっても、生活のために毎月約 51,588 円の消費が必要である、という解釈ができます。先ほどの限界消費性向と同様に、基礎消費の真の値がゼロかどうかを仮説検定により検証しましょう。図 1 の画面を見ると、const の t 値は 3.001、 p 値は 0.0044 です。このことから、仮に「const の係数（基礎消費）がゼロ」だとすると、3.001 という t 値が 0.0044、つまり 0.44% のわずかな確率でしか出てこないことになり、「const の係数（基礎消費）がゼロ」とは考えにくく、基礎消費の真の値はゼロから離れていると判断できます。よって、基礎消費も統計的に有意にゼロと異なります。

以上より、2014 年の日本の 47 都道府県における所得と消費のデータを用いた計量分析の結果、ケインズ型消費関数の限界消費性向が約 0.62、基礎消費が 51,588.4 円と推定され、どちらもゼロでないこととみなせることが分かりました。よって、「所得の一定割合が消費に振り向けられる。ただし、所得がないときもある程度の消費が必要である」というケインズ型消費関数の理論が、2014 年の日本にも当てはまることが明らかになりました。

8 おわりに

本稿では、経済理論における関数の 1 つであるケインズ型消費関数の、現実データを用いた推定を例に、データの入手方法から計量分析の結果の読み方までを解説しました。これから計量分析を初めて実行するという方は、まずは本稿の分析を再現してみてください。それができたら、次は自分の興味のあるデータでの分析や、より高度な手法での分析にも挑戦してみましょう。

参考文献

加藤久和（2012）『gretl で計量経済分析』日本評論社

加藤久和（2019）『やさしい計量経済学—プログラミングなしで身につける実証分析—』オーム社

鹿野繁樹（2015）『新しい計量経済学—データで因果関係に迫る』日本評論社