

```

id sex;
by age;
run;

data _null_;
file "sample5_2_1_2.txt";
put '年齢', ' ', '男', ' ', ' ', '女';
run;

data _null_;
set out2;
file "sample5_2_1_2.txt" mod;
put age ' ', 'x1 ', 'x2';
run;

```

このプログラムにより作成される SAMPLES_2_1_2.TXT (男女別年齢各歳別人口のファイル) は以下のように書き出される。

```

年齢, 男, 女
00 ,2002 ,1752
01 ,2173 ,1827
02 ,1941 ,2110
03 ,2113 ,2398
04 ,2246 ,2133
05 ,2243 ,1809
.
.
81 ,152 ,465
82 ,368 ,375
83 ,177 ,528
84 ,171 ,459
85 ,644 ,1678

```

このファイルを表計算ソフトウェアで読み込みグラフ機能を用いて人口ピラミッドを作成する(図 6.2.1 参照)。

ここまで多変量解析のように複雑な統計処理を扱ってこなかったが、SASでは複雑な多変量解析についてもかなり簡易に記述できるようになっている。それらについては多くの参考書が既に出版されているのでそれらを参照されたい。

6.2.2 STATA を用いた統計分析

ここでは stata を用いた実証分析の事例を紹介する。⁸ここで取り上げる分析の目的は、総務庁「家計調査」と「貯蓄動向調査」の世帯マッチを行なったデー

⁸この分析は総務庁からの委託を受け、永瀬も参加した「統計的マッチングにより発生する誤差の要因等の検証に関する調査研究会」において行った推計作業の一部である。永瀬(1999)参照

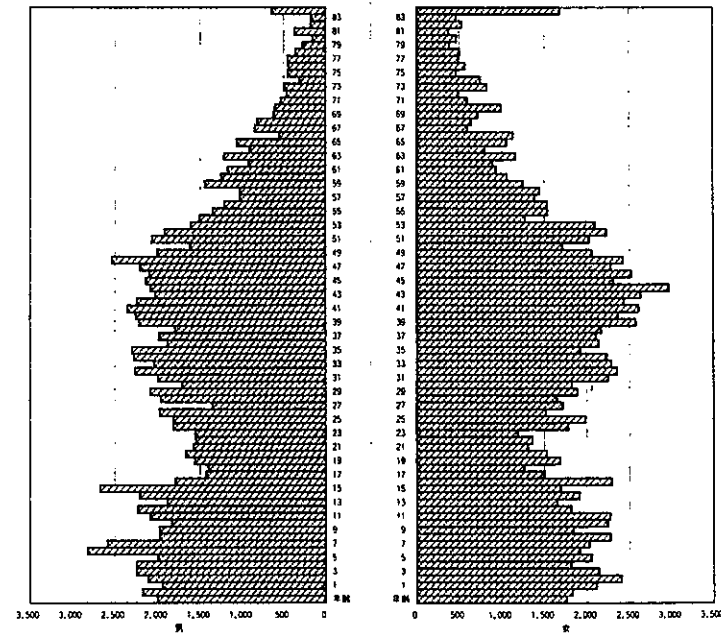


図 6.2.1: 1996年のニューヨークの人口ピラミッド

タを用いた子どもコストの推計である。教育費の高さや負担が出生率を下げる
としばしば指摘されるが、子どもコストがどの程度かかっているのか、個票データ
を用いた推計はほとんど行われていない。子どもコストは、教育費など直接
のコストだけではなく、財や時間の消費の変化を含めたものとして、さらには
ライフサイクルを見通しての長期的なものとしてもとらえるべきであり、また
社会的負担と給付、私的負担と便益の変化を勘案して、子ども需要の推計を行
うことも必要だろう。ここでは、紙面の都合上、上記の研究成果の一部である
エンゲル関数を用いた短期的な子どもコストの推計を紹介する。

データ読み込み、外挿データとの照合

大容量のデータは、ファイルスペースの節約のために、連続した数字の羅列
として保存されている場合が多く、この研究会で利用した『家計調査』や『貯蓄
動向調査』も例外ではない。一世帯について、例えば 1015398950293……と
いったような数字が、そのデータ量に応じて最高 2300 程度続き、こうしたデー
タが約 8000 世帯について羅列されているのが家計調査の月間データである。こ
の数列に意味づけを与えるのがデータの読み込み作業である。データにはカラム
票がついており、一例を挙げれば 3 カラム目から 10 カラムは世帯の識別番号、
次の 2 カラムは世帯の住む県番号、次の 3 カラムは市町村番号が、その次の 1
カラムは世帯主の性別が記されている、といった情報が添付されている。デー
タ分析は、この数字の羅列を情報に従い、変数として意味づけてソフトに認識
させるところからはじまるだろう。

はじめの 10 カラムを世帯番号として `ident` と名付け、次の 2 カラムは県番と
して `ken` と名付け、次の 3 カラムは `city` と名付けることにしよう。その例が以
下の辞書ファイル (`infile.dct` と名付ける) である。 `long`、`byte` などではデータ
の大きさをあらかじめ指定し、スペースを節約するためのコマンドである。また
" " は変数につけるラベルである。変数名は 8 桁以下の必要があるが、作表の
際にはラベルを表側や表頭に出力出来る (なお以下のプログラムはデータの秘
匿のために、データの実際の並び等とは異なるものとなっている)。%2 は 2 カ
ラムをその変数として読み込むことの指示である。

```
dictionary using c:data{
  _column(3)    long   ident   %10f  "identifier"
  _column(14)   byte   ken      %2f
  _column(16)   byte   city     %3f
  _column(19)   byte   sex      %1f
  _column(20)   byte   hshoyu   %1f  "1 own 2,3 rental"
  _column(23)   byte   hsize    %3f  "nobe menseki"
}
```

stata は対話方式でも行うことが出来るが、あらかじめ作成したプログラムを
流す方式は、プログラムミスを発見修正しやすいので、よりお勧めする。stata
のプログラムには

```
"ファイル名.do"
"ファイル名.log"
"ファイル名.dct"
```

がある。

データの読み込み、作表、作図などを指示するプログラムを作成した場合は、
拡張子を `do` とする。 `dct` ファイルは上で説明した辞書ファイルの拡張子である。
計算結果は、バッチ方式であれば、 `log` ファイルに自動的に出力されるが、対話
方式のもとでプログラムを流す場合は、 `log` ファイル名を指定しないと出力さ
れないから、以下のコマンドをプログラムのはじめにつけておくと便利である。

```
log using ファイル名.log, replace
```

`replace` は同じファイル名で内容を置換したい場合のみつける。

以下では辞書ファイル (`infile.dct`) を用いての読み込む例を示す。まずは
`summarize` で平均値、分散、最大値、最小値を出力し、正しい読みこみがされ
たかどうかをチェックする。全世帯が読み込めたか、性別は男性 1 女性 2 であ
るはずだが、 `sex` の最大値は 2、最小値は 1 と正しく出ているかなど、読み込み
のチェックと確認に時間をかけたい。

データの読み込みが正しく行われているようであれば、データを保存する。
`compress` によってファイルスペースを節約できる。 `infile.dct` 別に作った上
で、読み込みおよび作成された stata データの保存プログラムは以下のよう
なる。

```
infile using c:\infile.dct
summarize
compress
save c:\kakei.dta
```

「家計調査」においては、世帯や世帯員メンバーに関する情報に続いてその世帯が当該月に使用した用途分類の金額がカラム票の指定する位置に並んでいる（「家計調査-用途別分類」の場合）。例えば食費であれば、穀物費、内訳の米類、パン、麺類、穀粉その他、次いで魚介類、内訳の生鮮魚介...と行儀良く並んでいるので、カラム票に沿って注意深く辞書ファイルを作り上げ、読み込みをチェックすれば読み込みは終わりである。

ところが、「貯蓄動向調査」については、変数の位置がカラム票では特定できない。ある特定の数字（の羅列）が、郵便貯金なのか、銀行預金なのか、それとも社内預金か、株式か、投資信託か等を区別する方法は、続く変数が、どの資産を表すかにかかっている。この場合には、@を用いたプログラムが大変便利となる。例えば、X1_1, X1_2, X1_3 から X2_1, X2_2, X2_3 ... と続く3変数のセットのうちその後続く v1, v2, v3, ... いずれかの値が1であれば、その直前の変数は、郵便貯金の今期残高、財形残高、前期残高を意味すると定義づけられているとしよう。これをそれぞれ sav1, sav2, sav3 と名付けるとして、まず郵便貯金今期残高 sav1=0 を作成し、その上で、replace sav1=X1_1 if v2==1, と順繰りに ... if v23==1, といくつも先の変数まで書き出して変数を作成するのは大変な手間である。これは@の利用で簡略化できる。

```
gen sav1=0
for 1-23/1,1(n):replace sav1=X@1_1 if v@1==1
gen sav2=0
for 1-23/1,1(n):replace sav2=X@1_2 if v@1==1
gen sav3=0
for 1-23/1,1(n):replace sav3=X@1_3 if v@1==1
```

なお分析に必要な土地価格のデータがないため、ここでは市町村別土地価格として公示価格データを利用し、市町村別コードとともに入力したファイルを作成した。そのための辞書ファイルが tochi.dct であり、その結果、tochi.dta を stata データセットとして保存する。この土地価格データを同じような方法で、「家計調査」と「貯蓄動向調査」とを完全照合したデータ merge.dta に付加

することにする。照合のためには、照合キー（ここでは city）でソートをかけておく必要がある。以下がそのプログラムである。照合の結果、_merge という変数が出る。両データで無事に city が照合できた場合は、_merge は3で示されるが、どちらか一方しかない場合、1か2となる。「家計調査」はすべての市町村を同時にカバーしないので、_merge はすべて3ではないが、_merge によって、マッチマージがうまくいったかどうか、確認できる。

```
infile using c:\tochi.dct
sort city
save c:\tochi.dta
use c:\merge.dta
sort city
merge city using c:\tochi
summarize city age1 tochi95
tabulate _merge
drop if _merge==1
drop if _merge==2
```

分析対象の限定

マッチできた世帯のうち、収入内訳が詳しくわかるのは勤労者世帯のみである。そこで分析対象を勤労者世帯 (jiei==0) に限ろう。なお jie1 という変数の値が0 場合、と特定する場合には、== と等号を2つつける必要がある。

```
drop if jie1==1
```

非勤労者世帯を落とすと世帯数は、1725 世帯となる。この1725 サンプルで分析を初めて良いものだろうか。非勤労者世帯には、月間収入データがない。このため収入項目のわかる勤労者世帯に分析を限定することを決め、上のコマンドで安易に分析をはじめがちである。コンピュータは大変優秀で、コマンド一つでさりと非勤労者世帯を落としてくれる。

しかしデータ分析を行うにあたって、単純な表作成によるチェックは欠かさない。「非勤労者世帯」を除くのは、収入がわからないことその他に「非勤労者世帯」の収入・消費構造は、サラリーマン世帯とかなり異なるだろうという想定からである。ところが、非勤労者世帯を除く前に以下のようなコマンドでクロス表を作っておけば簡単にわかることであるが、

tabulate 年齢階層変数, summarize (jiei)

年齢階層が上がるほど非勤労者世帯は、いわゆる自営世帯だけではなく、サラリーマンを引退した世帯が加わっていくことが示される。例えば55歳未満では1から2割を占めるに過ぎなかった非勤労者世帯は、55-59歳層では3割近く、60-64歳層では6割超に拡大し、70歳代前半では8割を超えてしまう。

つまり高齢勤労者世帯は、若い勤労者世帯の集計とは性格を異にしている。つまり自発的に「労働市場からの引退を遅らせている世帯」であること、おそらく年金などの非勤労所得の水準が低いか、あるいは、良い仕事機会に恵まれているか、特定の性格を持つ限られた世帯を中心とした集計になってしまうことが想像される。また同居の子がいる世帯主を考えるため分析は世帯主が54歳以下の1343世帯とする。

カテゴリーデータの作成：用語の定義との照合

子ども変数 「家計調査」には、直接に家族構成の変数が入っていない。世帯主に関する情報（性別、年齢、就業状況等）に続いて、他の世帯員に関する情報が、世帯メンバー1、世帯メンバー2、世帯メンバー3...と並ぶ。そこで世帯員員の「(世帯主との)続柄」という変数から、世帯主がその親と同居している世帯なのかどうか、また世帯主の「子」や「孫」がいる世帯かどうか、未就学児童数は何人いる世帯か、特定年齢以下の子ども数は何人いるか、などといった家族類型の変数を自分で作成して世帯の状況をとらえる必要がある。

子どもコストという点では、大学生を持つ世代の分析も重要となるが、「家計調査」では同居人の情報はとれるが、別居中の家族の情報は含まれない。子どもの年齢が上がるほど、独立した、あるいは、親元を離れて下宿している子どもがいる世帯が増えると考えられるが、作成できる「子ども数」の変数では、独立した子ども数はとらえられない。これは作成した変数の限界であり、「子ども数」という自分自身の命名に惑わされないよう、解釈に注意を必要とする。実際にこの形で定義した「子ども数」という変数がどのように世帯主の年齢階層別に変化しているか、いくつかの指標をとって比較してみる。

世帯主の「子」もしくは「孫」として定義した子ども数は、6歳以下児数、7-13

表 6.2.1: 世帯の子ども数と妻の有業率

世帯主年齢階級	6歳以下児数	7-13歳以下児数	14-22歳以下児数	18歳以下児数	妻有取入率	妻有業率	妻本格就業率
24	0.563	0.000	0.000	0.563	12.5%	12.5%	62.5%
29	0.825	0.063	0.000	0.888	25.0%	32.5%	21.3%
34	1.179	0.154	0.000	1.333	19.4%	25.9%	15.9%
39	1.012	0.748	0.035	1.787	22.4%	33.5%	13.4%
44	0.374	1.128	0.494	1.958	32.8%	40.0%	20.8%
49	0.079	0.565	1.062	1.404	39.7%	51.4%	25.3%
54	0.017	0.077	0.809	0.434	36.2%	50.6%	24.3%
計	0.516	0.525	0.476	1.360	30.0%	40.0%	20.1%

妻本格就業率とは妻の勤労月収が8万円より多い世帯

歳階級児数、14-22歳階級児数、0-18歳階級児数それぞれについて表6.2.1の通りである。それぞれのピークは世帯主30-34歳階級の6歳以下児平均1.18人、40-44歳階級の7-13歳児平均1.13人、45-49歳階級の14-22歳児平均1.06人である。また18歳以下児数、22歳以下児数は40-44歳階級でそれぞれ平均1.96人、2.00人であり、世帯主年齢階級がこれより上がると、その人数は低下するから、世帯主が45歳以上の子ども数0の意味には独立という形での無子が含まれるようになってくることが示されている。表6.2.1はtabulateコマンド等を用いて作成したものである。

妻の就業状態 妻の就業状況も子どもコストの分析に重要な変数となる。「家計調査」から妻の就業についてわかるのは、当該月の妻の勤労収入と妻のふだんの職種である。ただしここで「妻」は、次の問題をはらんでいる。三世帯同居世帯で、親世代が世帯主である場合、「妻」は、「おばあちゃん」の就業状況や収入を指し示すことになり、子どもの母親の就業状況に関する変数を作成することにはならなくなる。そこでこのような誤りが起きる可能性を確認するため、世帯類型を続柄から作成すると(表6.2.2)、世帯主が若夫婦でない、というケースは1995年の世帯主が54歳以下の勤労者世帯には見られないことが確認された。なお表6.2.2からは、世帯主が35-39歳層では1割に過ぎない老親との同居世帯が、45-49歳層では2割近くにまで増えることがわかる。興味深いことに、

表 6.2.2: 世帯類型

世帯主年齢階級	～29	～34	～39	～44	～49	～54	計
核家族	89%	89%	85%	79%	75%	74%	81%
祖父母と同居	4%	6%	10%	14%	18%	16%	13%
うち祖父母の一方のみと同居	2%	4%	6%	9%	15%	13%	9%
一人親家庭その他	7%	4%	5%	6%	7%	9%	6%
標本数	96	201	254	264	292	236	1,343

増加分は片親との同居が中心である。夫婦の一方が亡くなった場合に、子世帯と同一家計になるというのが現代の多世代同居の主流であることが推察できる。なお、「家計調査」では、「住居」および「家計」が別であることが同一世帯の定義である。物理的に同一住宅に住んでいるとしても、家計が別であれば、別世帯という定義であることに加えて、勤労者世帯に限っていることも、多世代同居率が低く出る要因となっているだろう。なお世帯主が29歳以下のサンプルが少ないのは、「家計調査」が、独立生計でかつ二人以上の世帯に限ったデータであるためである。

このような確認作業をしつつ、妻の就業状態について、次の三種の変数を作成した。「同居世帯員に世帯主の妻があり、職業が無職でない（以下「妻有職」）」、「（家計調査11月において）妻に勤労収入がある（以下「妻有収入」）」と、「（家計調査11月における妻の勤労収入が8万円以上（以下「妻本格就業」）」。

これまでの研究から、既婚女性の労働時間の長短やその自由度は、正社員であるか、非正規社員であるかで大きく異なることが知られている。また既婚女性の時間あたり賃金率も、正社員と非正規社員とでは、3.4割程度の差異があることが知られている。そこで、正社員か、パートかといった情報が知りたいが、「家計調査」には、労働時間や従業上の地位についての調査項目はない。職種の内訳がわかるが、その区分けは常用労務者、臨時日雇い労務者、民間職員、官公職員、商人・職人、自由業、その他であり、いわゆるパートと正社員との差異はこの職種分けからは推測できない。そこで年収が非課税となる境目程度の勤労収入月額（8万円）を就業がどの程度本格的であるかを見るための代理指標とした。

ところがさらに検討すると、「妻有収入」と「妻有職」とでは、サンプルの30%と40%と、かなりの差異が出ることがわかった。どうしたことだろう。勤労収入がない有職女性が多いからだろうか。妻が「商人・職人、自由業、その他」の場合には、定義から勤労収入はないはずである。しかし職種と収入との関係をクロス表に作成すると、勤労収入があるはずの「民間職員」でも妻の収入が11月に記載されていない世帯が17%あった。記入もれもありうるが、隣接月の家計簿に11月分の妻の勤労収入を記載した可能性もある。この点は10月、11月、12月の家計調査のデータを世帯番号で時系列でつなげば改善できるだろう。また有職であっても11月には仕事をしない妻が多かった可能性も考えられる。民間企業で事務に従事しているパート労働者は「民間職員」の分類に入るだろう。年収を一定以内に抑えるという就業調整目的を行っているパートは労働省「パートタイム労働者総合実態調査」によればパートの約4割程度に達しているが、12月の繁忙月を避けて11月中に仕事を休み調整をした者が多いのかもしれない。

このように度々確認をとるのは、作成したデータが何を意味するものなのか、大量データを分析する場合ほど、注意深いチェックが欠かせないからである。データ数が多いほど、読み込みのミスやあるいは重要な点を見落として思いこみで変数を作成してしまうこともありうるからである。

データの把握

データ内容を把握するために様々なクロス集計を行う。この作業は丁寧に行う。便利なコマンドをいくつかを紹介する。

連続した世帯主の年齢(age1)という変数から、世帯主の年齢階級データを作成する便利な方法はrecode()である。2変数のtabulateによって、世帯主年齢階級と子ども数とのクロスをとる。

```
gen agecat=recode(age1, 29,34,39,44,49,54)
tabulate agecat child
```

tabulateに、summarizeを組み合わせると階級別平均を出すことができる。ここでは世帯主年齢別の世帯主勤労所得平均を出す例を示す。また子どもの人数(child)、世帯主年齢階級(agecat)別に、()内の平均値を出すこともできる。

ここでwifeincは妻の勤労収入であり、世帯主年齢階級別に、かつ、子ども数別に妻の勤労収入が出てくることになる。

```
tabulate agecat, summ (income)
tabulate child agecat, summ(wifeinc)
```

このたくいのような作表および作図によってデータ間の関係を見る。

340ページでは短期の子どもコストの推計に、消費支出に占める食料費の割合を用いた。そこでモデルの検討と前後するが、計量分析をする前のデータのチェックについてもこのセクションで触れる。

外れ値は、計量分析の係数をかなり左右する。例えば変数Aと変数Bとに強い相関がある場合でも、いくつかの外れ値によって、相関係数は弱くなりうる。計量分析をする前に、データ分布の形状を見ておく必要がある。

```
summarize 変数名, detail
```

は当該変数について、最大値4つと最小値4つ、および、上下から10%、5%、1%ポイントの値が表示されるから、異常値の有無のあたりをつけることに役立つだろう。またデータの分布の偏りを示すskewnessや分布の尖度を示すkurtosisなども表示される。以下は食費シェア(share)を取り上げた。meanと中央値が近く、またkurtosisも正規分布の3に近いので、通常の最小2乗法で扱いやすいデータ分布となっている。

```
summarize share, detail
```

Percentiles			Smallest		
1%	.0618443	.016668			
5%	.1048141	.0305948			
10%	.1337908	.0346536	Obs	1348	
25%	.1862042	.0352135	Sum of Wgt.	1348	
			Mean	.2674191	
50%	.2566126		Std. Dev.	.1129203	
			Largest		
75%	.3316313	.7102163			
90%	.421131	.7462845	Variance	.012751	
95%	.4680865	.797015	Skewness	.7836428	
99%	.5968968	.8360072	Kurtosis	4.253493	

2変数をプロットすることで、変数間の関係と散らばりを目で見ることが出来る。

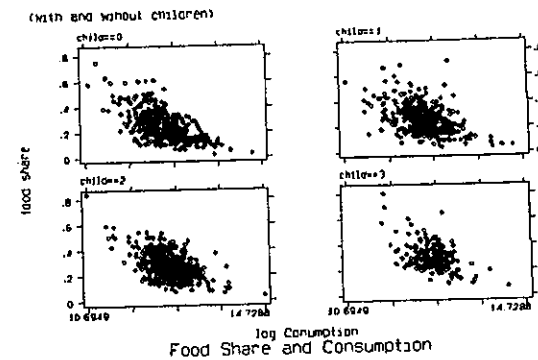


図 6.2.2: 食費シェア(消費支出に占める食費割合)と消費支出(対数)

作成した食費シェア(share)と消費支出との関係を子ども数別に、世帯主45歳未満でプロットしたものが図6.2.2である。ここでは子ども数3人以上は人数が少なくなるため、まとめて3として示した。グラフは消費支出額と食費シェアの負の相関を示している。シェアという形で率にしていること、消費支出の対数をとっていることによって外れ値の問題は回避されているようである。ただし世帯の食費シェアに子ども数がどの程度の異なる影響を与えているか、という点に関しては、プロット図では不十分である。

```
sort child
graph shFF lnC, ylab(0, 0.2, 0.4, 0.6, 0.8) by (child)
  rlab(0, 0.2, 0.4, 0.6, 0.8)
  t1("(with and without children)")
  title ("Food Share and Consumption")
```

これに対して分母を経常収入とした食費割合はややばらつきが増す。その月の食費が経常収入を上回る世帯が多少ある。listによって、外れ値のチェックをすると、食費シェアが1を上回る世帯が2世帯あった。実際にその世帯の月間経常収入、勤労収入等と年間の世帯収入とを対比させるなど、関連のありそうな変数についての金額を具体的にチェックすると、年間収入に比べて今月の経常収入がかなり低く、貯蓄取り崩しが高く、この月に限って何らかの特別な理由で経常収入が低かった世帯であろうと推察される。

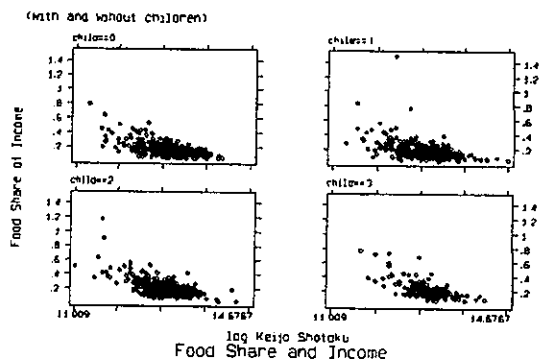


図 6.2.3: 食費シェア（経常収入に占める食費割合）と経常収入（対数）

```
label variable lnK "log Keijo Shotoku"
label variable shF "Food Share of Income"
graph shF lnK, ylab( 0.2, 0.4, 0.6,0.8,1.0,1.2,1.4) by (child)
  rlab( 0.2, 0.4, 0.6,0.8,1.0,1.2,1.4) t1("(with and without children)")
  title ("Food Share and Income")
*find outliers
list income1-income5 x1 x5 x22 if shF>1 & shF<=.
list income1-income5 x1 x5 x22 if shF>0.8 & shF<=.
by child: corr shF lnK if shF<1
by child: corr shF lnK if shF
```

こうした外れ値を除くと、例えば上の図 6.2.3 で示した 2 変数の相関は、子ども数 1 において 1 サンプルが除かれるだけだが、相関係数が -0.4568 から -0.5196 (349 サンプル) に変化する。なお同様に 1 サンプル除かれる子ども数 2 においては -0.5711 が -0.5701 に変化する程度である。なお基準を 0.9 以上とすると、子ども数 2 においても 1 サンプル除かれ、相関係数は -0.5603 となる。

経常収入でなく、消費額を用いた食費シェア (図 6.2.2) の振れが少ないのは、消費額は多少の収入の変動があっても大きく振れにくいからである。エンゲル関数の推計には消費額に占める食費割合が通常用いられる。

経済モデルと分析

短期的な子どもコストの推計として、研究蓄積があるのは、等価尺度によるコストの推計である。子どものいない世帯と、子どもが 1 人、2 人、3 人等の世帯があるとしよう。すべての世帯が同一の収入実額を受け取っているとしても、生活費が異なるから、子どもの人数や年齢等の応じて、世帯員が享受する生活水準は異なるだろう。消費における規模の経済があるため、単純に消費を人数で割ったものでは不十分である。また教育費や子ども衣料等、子どものための支出を単純集計するのみでも不十分である。というのは、子どもを持つことによって大人向けの財支出についても、節約が行われ、財支出構造全体が変わると考えられるからである。

そこである世帯を基準として、家族構成が異なる世帯が、基準世帯と同じ「効用」水準を達成するために必要な支出額は何倍かを計算するのが等価尺度 (equivalence scales) である。その簡便な方法である Engel モデルを推計してみよう。

Engel モデルは、貧しい世帯ほど食費シェアが高いという経験事実を用いて、エンゲル係数 (消費支出に占める食料費額) を生活水準の簡便な指標と見なす。つまり食費シェアが同一の世帯は、子ども数等が異なっても同一の生活水準を享受していることを仮定している。同一の食費シェアを達成するために必要とされる支出増が、子どもコストの等価尺度となる。

Tsakoglou(1991) の定式化を用いて、

$$\text{食費シェア} = b_0 + b_1 \cdot \ln X + b_2 \cdot (\text{Children}) + b_3 \cdot (\text{other factors})$$

とする。なお $\ln X$ は消費支出の対数、食費シェアは消費支出に占める食費の割合である。ここで 2 種類の食費シェアを考慮した。食費シェア 1 は食費全体であり、食費シェア 2 は食費から一般外食費を除いて推計したものである。⁹

エンゲル型等価尺度は、 $\exp(-b_2/b_1)$ によって求めることができる。通常の最小自乗法のコマンドは regress 変数名 である。

⁹この方法は武蔵 (1992) にならった。

推計結果

他の説明変数としては、年齢階層、妻の就業状態、持ち家や社宅、公営住宅住まい等を入れた。住宅関連の変数は、貸家世帯よりも、帰属家賃および社宅や公営住宅の家賃補助の分、実際の生活水準に対して、シェア関数の分母となる消費額が下がるからである。また妻の就業状態も同様に、分母の消費額に含まれないが実消費に含まれる家事労働の差異を考慮するためである。様々な推計作業を行ったが、18歳以下の子どもが一人増えた場合、子どもなしの夫婦と同じ効用水準に夫婦の効用水準に戻す場合、外食費を入れた食費シェアを用いると22%の、外食費も入れた食費シェア1では17%程度の消費支出増が必要であると計測された。

妻が就業者の場合、外食費を含まない食費シェアには有意な変化はないが、外食費を含めた食費シェアは上がる。妻の就業によって「外食」が家事労働の一部代替するためであり、就業コストとともいえよう。「妻有収入」によるコストの増加は、食費シェア1で4-8%である。「妻本格的就業」（月収8万円以上）を説明変数すると、係数の有意性がやや高まり、また係数が上がり、コスト増加は13%強と推計されている。その例は図6.2.4に示す通りである。

発展課題：生涯を見通した子ども需要と子どもコストの推計

上記のEngelモデルは確立された推計手法の一つではあるが、現代の子どもコストの推計という点では、非常に重要な2点を捨象している。今日子ども数は長期的には避妊等によって選択できる側面があること、また世帯収入は外生的に与えられているものではなく、妻の就業時間によって増減出来るという点である。Engelモデルは子ども数も世帯所得も外生変数として扱うから、推計する子どもコストは短期に過ぎない。従来、このモデルは生活水準の確保のために、所得税の賦課や家族手当等に関して、世帯間のバランスのあり方を検討する上での指標として利用されてきたが、今日の少子化の問題には、長期の視点にたつてははじめから子どもを持たない（あるいは少数持つ）ことを選択することの影響も大きいと考えられる。この場合、子どもを持つことが今期の生活水準に与える影響だけでなく、中年期、および老年期の資産形成や生活水準に

Source	SS	df	MS			
Model	7.40359864	12	.616966553	Number of obs = 1343		
Residual	9.64545946	1330	.007252225	F(12, 1330) = 85.07		
				Prob > F = 0.0000		
				R-squared = 0.4343		
				Adj R-squared = 0.4291		
				Root MSE = 0.0851		
Total	17.0490581	1342	.012704216			
share	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnC	-.1455653	.0050801	-28.654	0.000	-.1555311	-.1355994
chdno	.0233696	.0026493	8.821	0.000	.0181723	.0285668
agect3	.0110067	.0107027	1.028	0.304	-.0099893	.0320027
agect4	.0362854	.0106159	3.418	0.001	.0154596	.0571112
agect5	.0689779	.0108042	6.384	0.000	.0477827	.0901731
agect6	.0802301	.0106864	7.508	0.000	.0592661	.1011194
agect7	.0753634	.0113641	6.632	0.000	.0530699	.097657
wlph	.0172604	.0059732	2.890	0.004	.0055424	.0289784
lfnoth	.0121938	.0043979	2.773	0.006	.0035663	.0208213
mochiie	.0359834	.0062961	5.715	0.000	.023632	.0483348
shataku	.0182186	.0079205	2.300	0.022	.0026805	.0337566
kosha	.0274161	.0099676	2.751	0.006	.0078622	.04697
_cons	1.975381	.0627297	31.490	0.000	1.852321	2.098441

記号の説明

share	食料費シェア	agect7	世帯主年齢階級 50-54 歳
lnC	消費支出対数	wlph	妻本格就業
chdno	18歳以下子ども数	lfnoth	他の有業人員数
agect3	世帯主年齢階級 30-34 歳	mochiie	持ち家世帯
agect4	世帯主年齢階級 35-39 歳	shataku	社宅住まい
agect5	世帯主年齢階級 40-44 歳	kosha	公営住宅住まい
agect6	世帯主年齢階級 45-49 歳	_cons	定数項

図 6.2.4: エンゲル関数の推計

表 6.2.3: 世帯主年収階層別 18 歳以下の子ども数

世帯主年 齢階級	全体				持ち家世帯			
	第1四分 位以下	第2四分 位以下	第3四分 位以下	第3四分 位より上	第1四分 位以下	第2四分 位以下	第3四分 位以下	第3四分 位より上
~29	1.04	0.72	0.84	0.76	1.75	1.33	1.25	1.00
~34	1.33	1.53	1.45	1.10	1.20	1.67	1.13	1.53
~39	1.87	1.86	1.83	1.58	2.16	2.00	2.04	1.59
~44	1.84	2.13	1.91	1.94	2.06	2.08	2.00	1.98
~49	1.03	1.59	1.47	1.48	1.09	1.56	1.47	1.45
~54	0.53	0.47	0.35	0.36	0.60	0.53	0.38	0.36

どのような（正のあるいは負の）影響を与えているかを含めて計測し、また税制・社会保障等の再分配も含めて子どもコストの推計をしていく必要がある。

「家計調査」と「貯蓄動向調査」を完全照合したのは、資産のデータとフローのデータ双方を利用するためである。子どもの有無と資産蓄積の関連までも見ることができる。筆者の委託研究では、長期の子どもコストの推計という、ここでは示さない部分にも取り組んだ。しかし与えられた紙面もつきたのでここでこの事例は終わりとする。ただし資産面も含めた作表例を示す。表 6.2.3 は所得を四分位に分けて、子ども数を見た例である。

```
egen incp1=pctile(incs), p(25) by(agecat)
egen incp2=pctile(incs), p(50) by(agecat)
egen incp3=pctile(incs), p(75) by(agecat)
sort agecat
by agecat: summarize incp1-incp3
gen p1=1 if incs<=incp1
replace p1=2 if incs<=incp2 & incs>incp1
replace p1=3 if incs<=incp3 & incs>incp2
replace p1=4 if incs>incp3 & incs!="."
tabulate agecat p1, summarize (child)
```

個票利用の留意点

この項では、入門者に対して stata という統計パッケージを用いて、短期の子どもコストの推計を中心に実証事例を紹介した。統計解析パッケージの利用による個票データの再集計によって、新たな調査コストをかけずに、どれだけ

豊かで興味深い情報を引き出せるか、その一端を示せばと願う。

ただし、入門者に対して、どうしても強調したいのは、コンピュータが、膨大なデータ分析を文字通り命令通り、まさに一瞬にしてやっつけてのけるだけに、どれだけマイクロデータ分析には注意が必要かという点である。クロス表が単純な二変数の関係を示すのに対して、高度な統計手法は、その因果関係をも指し示してくれる。しかし高度な統計手法はさまざまな「仮定」を置いてなりたつ場合が多い。単純なクロス表を作ることや、データをプロットして見ること、さらに異常値等の内容まで見ること（自分の読み込みによる誤解であって異常値ではないのかもしれない）は、現実社会の意味ある分析を行うために、どうしても必要なことである。この手間を厭うと、一見もつもらしいが意味のない分析を行う危険性がある。労を惜まず、さまざまなクロス表やグラフを作成してデータ構造をつかむ中で、また多くのヒントとアイデアを得られる。

6.2.3 S-Plus

Sとは？ 事前に提示した「統計モデル」をデータに当てはめる仮説検証的データ解析(Confirmatory Data Analysis)に対峙する概念として、Tukey (1977)は、データの構造自体を発見的に分析する「探索的データ解析(EDA, Exploratory Data Analysis)」を提唱した。ある意味で、「データ・マイニング(Data Mining)」の先駆と位置付けられる考え方である。S言語は、Tukeyの思想を実現するUNIX上の計算環境として、1980年代にAT&Tベル研究所が開発したプログラミング言語である。このため開発当初から、箱ひげ図や散布図行列などの統計グラフを用いてデータを眺める関数群が充実していた。今では、SAS-JMPなど多くのデータ解析ソフトウェアで実現されている、ダイナミック・グラフィックス、すなわちグラフ上にプロットされた点とデータベース上の観測データの関係性を視覚的に明示する方法なども、S言語の特長とされていた。Sは、「言語」と銘打たれているように、SAS-IMLと類似した統計ツール開発のための行列型プログラミング言語として、明解なシンタックスと必要十分な関数群を揃えている。オブジェクト指向の考え方もいち早く導入されており、データ解析に限らず、行列や多重配列の演算、乱数実験を伴うアプリケーション開発のプロト