2010/10/26

「経済の実証分析の方法について」

神戸大学・経済学部　谷崎 久志

1　最小二乗法について

線型モデルの係数の値をデータから求める時に用いられる手法 $\Longrightarrow$ 最小二乗法

1.1　最小二乗法と回帰直線

$(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ のように n 組のデータがあり，X_i と Y_i との間に以下の線形関係を想定する。

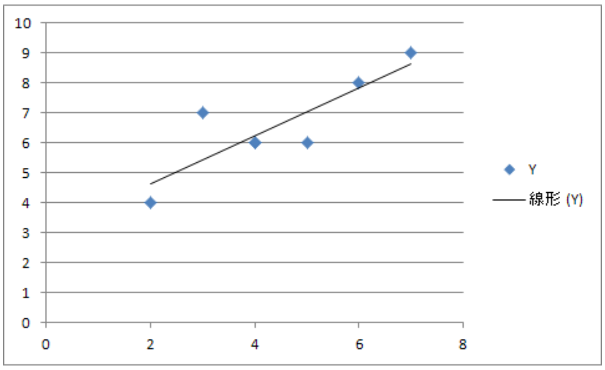
$$Y_i=\alpha+\beta X_i,$$

X_i は説明変数，Y_i は被説明変数，α,β はパラメータとそれぞれ呼ばれる。
上の式は回帰モデル (または，回帰式) と呼ばれる。目的は，切片 α と傾き β をデータ $\{(X_i,Y_i), i=1,2,\cdots,n\}$ から推定すること。

例：
表 1：X と Y のデータ ($n=7$)

X	Y
2	4
4	6
5	6
3	7
4	6
6	8
7	9

図 1：X (横軸) と Y (縦軸) の散布図 ($n=7$)



データから直線を求める。

以下のように対数線形を想定することもできる。

$$\log Y_i=\alpha'+\beta'\log X_i,$$

切片 α' と傾き β' をデータ $\{(\log X_i,\log Y_i),\ i=1,2,\cdots,n\}$ から推定する

β と β' の意味の違いに注意。

$$\beta=\frac{\mathrm{d}Y_i}{\mathrm{d}X_i}$$
$$\beta'=\frac{\mathrm{d}\log Y_i}{\mathrm{d}\log X_i}=\frac{100\mathrm{d}Y_i/Y_i}{100\mathrm{d}X_i/X_i}$$

i を時点とすると，$100\mathrm{d}X_i/X_i$ は i 時点の瞬間的な成長率を表す。

$\beta=X$ が 1 単位増えたときに Y がどれだけ増えるか (限界係数) —> X,Y の単位に注意

$\beta'=X$ が 1% 増えたときに Y が何 % 増えるか (Y の X 弾力性)

1.2　切片 α と傾き β の推定

次のような関数 $S(\alpha,\beta)$ を定義する。

$$S(\alpha,\beta)=\sum_{i=1}^n u_i^2=\sum_{i=1}^n(Y_i-\alpha-\beta X_i)^2$$

このとき，

$$\min_{\alpha,\beta} S(\alpha,\beta)$$

となるような α,β を求める (最小自乗法)。このときの解を $\hat{\alpha},\hat{\beta}$ とする。
最小化のためには，

$$\frac{\partial S(\alpha,\beta)}{\partial\alpha}=0,\qquad \frac{\partial S(\alpha,\beta)}{\partial\beta}=0$$

を満たす α,β が $\hat{\alpha},\hat{\beta}$ となる。すなわち，$\hat{\alpha},\hat{\beta}$ は，

$$\sum_{i=1}^n(Y_i-\hat{\alpha}-\hat{\beta}X_i)=0, \tag{1}$$
$$\sum_{i=1}^n X_i(Y_i-\hat{\alpha}-\hat{\beta}X_i)=0, \tag{2}$$

を満たす。
$\hat{\alpha},\hat{\beta}$ について解くと，

$$\hat{\beta}=\frac{\sum_{i=1}^n(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^n(X_i-\overline{X})^2}$$
$$\hat{\alpha}=\overline{Y}-\hat{\beta}\overline{X}$$

となる。ただし，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

とする。

$\Longrightarrow$ 最小二乗法を経済学に応用する。

2　需要関数 (一般的な話)

2 つの財を考える。

需要量： Q_1, Q_2　　価格： P_1, P_2　　所得： Y

● $Y \uparrow \Longrightarrow Q_1 \uparrow$

● $P_1 \uparrow \Longrightarrow Q_1 \downarrow$

● $P_2 \uparrow \Longrightarrow Q_2 \downarrow \Longrightarrow \begin{cases} Q_1 \uparrow & \text{(代替的)} \\ Q_1 \downarrow & \text{(補完的)} \end{cases}$

線形で表す。

$Q_1 = a + bY + cP_1 + dP_2$

——>　a, b, c, d は定数項と係数

符号条件：

$b > 0$,　$c < 0$

$d < 0$ (補完的)，または，$d > 0$ (代替的)

a, b, c をどのようにして求めるか？——> 最小二乗法

または，対数線形で表す。

$\log Q_1 = a' + b' \log Y + c' \log P_1 + d' \log P_2$

——>　a', b', c', d' は定数項と係数

符号条件は線形の場合と同じ。すなわち，

$b' > 0$,　$c' < 0$

$d' < 0$ (補完的)，または，$d' > 0$ (代替的)

しかし，値の意味は異なる。

b, c, d は限界係数

b', c', d' は弾力性

3　データ

１世帯当たり年間の品目別支出金額，購入数量及び平均価格（全世帯・勤労者世帯）

『家計調査年報 (平成 21 年)』(総務省統計局)

1982〜2009 年の年次データ

====> パンの需要関数

====> 米，麺類も考慮する必要あり

====> 利用データ

・１世帯当たり年間のパンの購入量 (単位：1g)

・１世帯当たり年間の消費支出 (単位：円)

　　$\longrightarrow$ 所得と考える

・パン 100g 当たりの価格 (単位：円)

・麺類 100g 当たりの価格 (単位：円)

・米 1kg 当たりの価格 (単位：円)

表 2：利用したデータ

	米		パン		麺		マーガリン		総支出額	物価指数
	数量	金額	数量	金額	数量	金額	数量	金額		
	1kg	円/1kg	1g	円/100g	1g	円/100g	1g	円/100g	円	2000年=100
1982	164.22	435.98	42161	54.04	36854	45.32	3001	64.79	3038024	82.9
1983	160.14	448.20	41745	55.88	36492	47.97	2897	63.06	3114247	84.4
1984	158.06	461.69	40890	57.62	36500	49.21	2870	63.12	3195829	86.3
1985	154.51	477.41	39545	59.42	36099	50.20	2666	63.38	3277373	88.1
1986	150.96	482.80	39532	60.86	35859	50.74	2692	61.93	3316493	88.6
1987	142.60	482.67	38710	61.53	34576	51.83	2587	55.61	3371326	88.7
1988	132.04	478.40	39218	61.75	33971	52.65	2524	56.98	3493468	89.3
1989	128.40	486.37	39927	63.99	33603	54.71	2430	57.87	3592205	91.3
1990	125.78	497.33	39157	66.71	32890	57.14	2243	60.03	3734084	94.1
1991	123.82	499.36	39659	69.57	32615	61.44	2194	62.17	3925358	97.3
1992	120.58	516.05	39697	70.75	33401	61.06	2112	61.26	4003931	98.9
1993	121.93	536.85	40209	70.51	35085	59.80	2037	60.69	4022955	100.2
1994	107.99	587.50	40458	71.08	35760	58.37	2000	59.37	4006086	100.8
1995	106.42	496.64	38766	71.97	35096	56.77	1931	58.01	3948741	100.7
1996	104.91	476.26	38436	72.74	34804	55.90	1815	57.56	3946187	100.8
1997	102.81	460.70	38333	74.39	35061	56.77	1802	57.44	3999759	102.7
1998	103.53	439.24	38287	74.10	34956	55.98	1763	57.54	3938235	103.3
1999	101.99	427.60	39246	73.00	34963	55.37	1751	56.43	3876091	103.0
2000	99.24	405.62	38433	71.58	34884	53.81	1707	56.43	3807937	102.2
2001	97.29	393.60	37492	70.30	34959	52.55	1645	54.15	3708649	101.5
2002	93.84	389.95	43507	61.48	36769	50.01	1530	55.16	3671438	100.6
2003	93.97	396.46	45104	60.28	36988	49.11	1528	55.33	3622095	100.3
2004	87.62	425.65	45905	60.15	37692	47.90	1518	53.83	3635703	100.3
2005	89.48	367.64	44122	59.50	36187	46.04	1443	54.08	3606377	100.0
2006	85.10	363.92	44497	59.69	35335	46.11	1383	54.37	3539316	100.3
2007	85.33	359.55	45238	59.90	35486	46.26	1381	54.72	3573382	100.3
2008	88.55	352.70	44445	63.49	35899	50.10	1433	60.70	3563187	101.7
2009	85.11	358.32	45599	63.52	36615	50.31	1387	69.36	3500848	100.3

物価指数 ＝ 平成 17 年基準・総合消費者物価指数

グラフにする。

図２： １世帯当たり年間のパンと麺の購入量 (単位：1g)

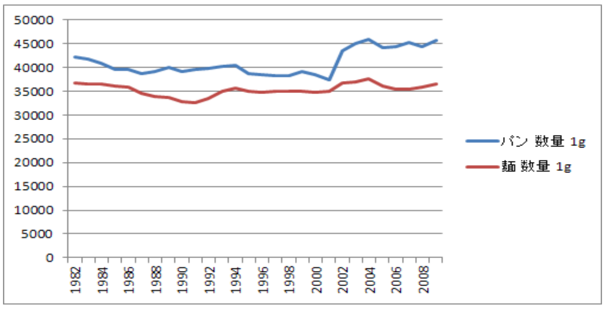


図３： １世帯当たり年間の米の購入量 (単位：1kg)

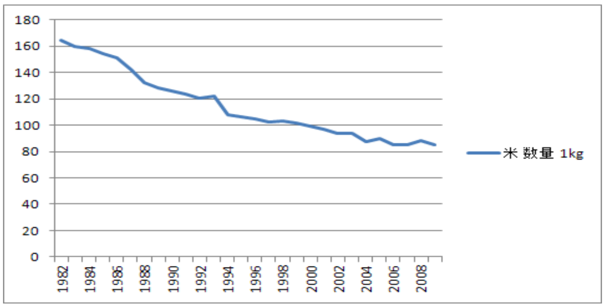


図４： １世帯当たり年間の消費支出 (単位：円)

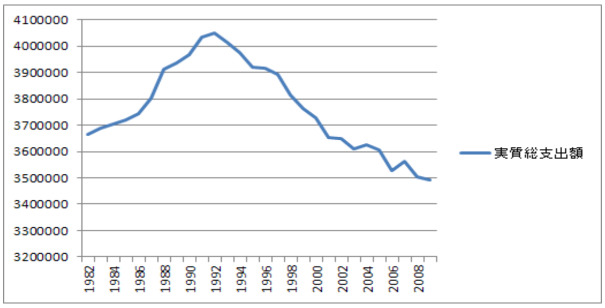


図５：パンと麺の実質価格 (単位：円/100g)

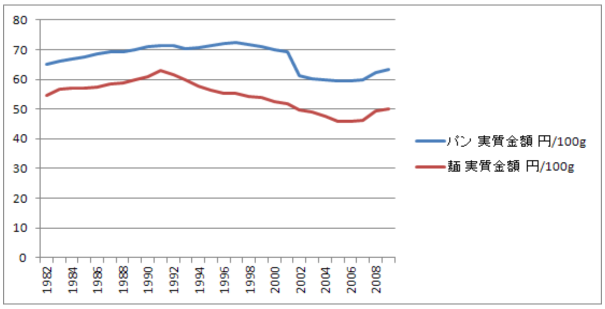
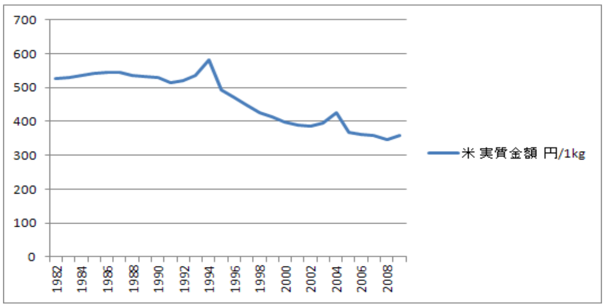


図６：米の実質価格 (単位：円/1kg)



4　パンの需要関数

4.1　線形

Excel を使って推定できる　「アドイン」で「分析ツール」を組み込む)。

Q_1 = １世帯当たり年間のパンの購入量 (単位：1g)
Y = １世帯当たり年間の消費支出 (単位：円)
P_1 = パン 100g 当たりの価格 (単位：円)
P_2 = 麺類 100g 当たりの価格 (単位：円)
P_3 = 米 1kg 当たりの価格 (単位：円)

$$Q_1 = \underset{(13.80)}{72671.5} + \underset{(1.016)}{0.00247}\ Y - \underset{(7.954)}{689.7}\ P_1 + \underset{(1.254)}{158.6}\ P_2 - \underset{(1.259)}{7.00}\ P_3$$

$$s = 982.0, \quad R^2 = 0.884, \quad \overline{R}^2 = 0.864, \quad DW = 0.996$$

() 内の数値は t 値を表す。

結果の解釈：

・符号条件は？ —> 理論モデルとの整合性
・推定式の当てはまりは？ —> R^2 の値
・係数の有意性は？ —> t 値の値

パンと米は補完的（？），パンと麺類は代替的

$0 \leq R^2 \leq 1$ が成り立ち，１に近ければ，データのプロットは直線に近くなる。
この場合，R^2 は 1 に近いので当てはまりはいい。

() 内の値が，絶対値で目処として 2 以下であれば，その説明変数は被説明変数に影響を与えない。
すなわち，Y, P_2, P_3 が Q_1 に影響を与えていない。
Y の係数は正であるはず。

Y (年間収入) が１円増えれば，Q_1 (パンの購入量) は 0.00247g 増える。
P_1 (パンの価格) が 100g 当たり１円増えれば，Q_1 (パンの購入量) は 689.7g 減る。
P_2 (麺類の価格) が 100g 当たり１円増えれば，Q_1 (パンの購入量) は 158.6g 増える。
P_3 (米の価格) が 1kg 当たり１円増えれば，Q_1 (パンの購入量) は 7g 減る。
係数の値の意味が分かりにくいので，対数変換して，推定しなおす。

4.2　対数線形

$$\log Q_1 = \underset{(4.373\,)}{12.20} + \underset{(0.7927)}{0.1686} \log Y - \underset{(7.834\,)}{1.123} \log P_1 + \underset{(1.654\,)}{0.2765} \log P_2 - \underset{(1.337\,)}{0.0839} \log P_3$$

$$s = 0.0234,\quad R^2 = 0.885,\quad \overline{R}^2 = 0.866,\quad DW = 1.000$$

() 内の数値は t 値を表す。

パンと米は補完的（？），パンと麺類は代替的
Y が 1% 増えれば，Q_1 が 0.1686% 増える。—> パンは必需品
P_1 が 1% 増えれば，Q_1 が 1.123% 減る。—> パンは他のもので代替可能

4.3　財の追加 & 対数線形

パンと補完的と考えられる財を追加する。—> マーガリン

P_4 = マーガリン 100g 当たりの価格 (単位：円)

図 7： 1 世帯当たり年間のマーガリンの購入量 (単位：g)

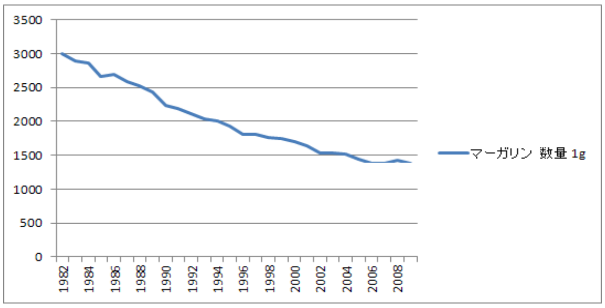
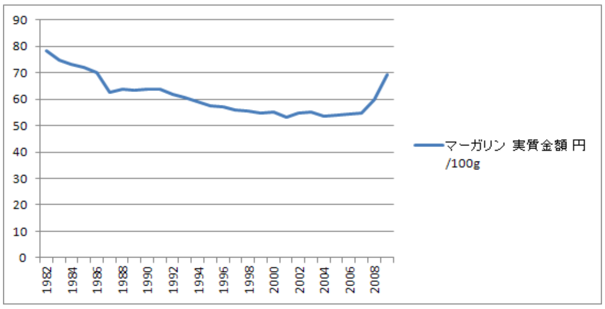


図 8：マーガリンの実質価格 (単位：円/100g)



$$\log Q_1 = \underset{(0.457)}{2.311} + \underset{(2.377)}{0.8627} \log Y - \underset{(7.397)}{1.026} \log P_1 - \underset{(0.4564)}{0.1038} \log P_2 - \underset{(2.335)}{0.1518} \log P_3 + \underset{(2.27113\,)}{0.220859} \log P_4$$

$$s = 0.0215,\quad R^2 = 0.907,\quad \overline{R}^2 = 0.886,\quad DW = 1.554$$

() 内の数値は t 値を表す。

t 値から判断して，P_2 を除き，すべての説明変数が被説明変数に影響を与えている。

パンと米は補完的，パンと麺類は補完的，パンとマーガリンは代替的—> 予想に反する

●麺類の価格を除いて推定しなおす。

$$\log Q_1 = \underset{(1.188)}{4.006} + \underset{(2.981)}{0.7445} \log Y - \underset{(9.845)}{1.065} \log P_1 - \underset{(2.454)}{0.1556} \log P_3 + \underset{(2.911)}{0.1882} \log P_4$$

$$s = 0.0212,\quad R^2 = 0.906,\quad \overline{R}^2 = 0.890,\quad DW = 1.478$$

() 内の数値は t 値を表す。

R^2 は 0.9 以上で当てはまりは良い。

$\log Y$，$\log P_1$ の係数の値は，正，負なので，理論的にも正しい。

t 値は絶対値で 2 以上なので，すべての説明変数が被説明変数に影響を与えている。

パンと米は補完的，パンとマーガリンは代替的—> 予想に反する

この講義ノートは，
`http://ht.econ.kobe-u.ac.jp/~tanizaki/class/2010`
からダウンロード可