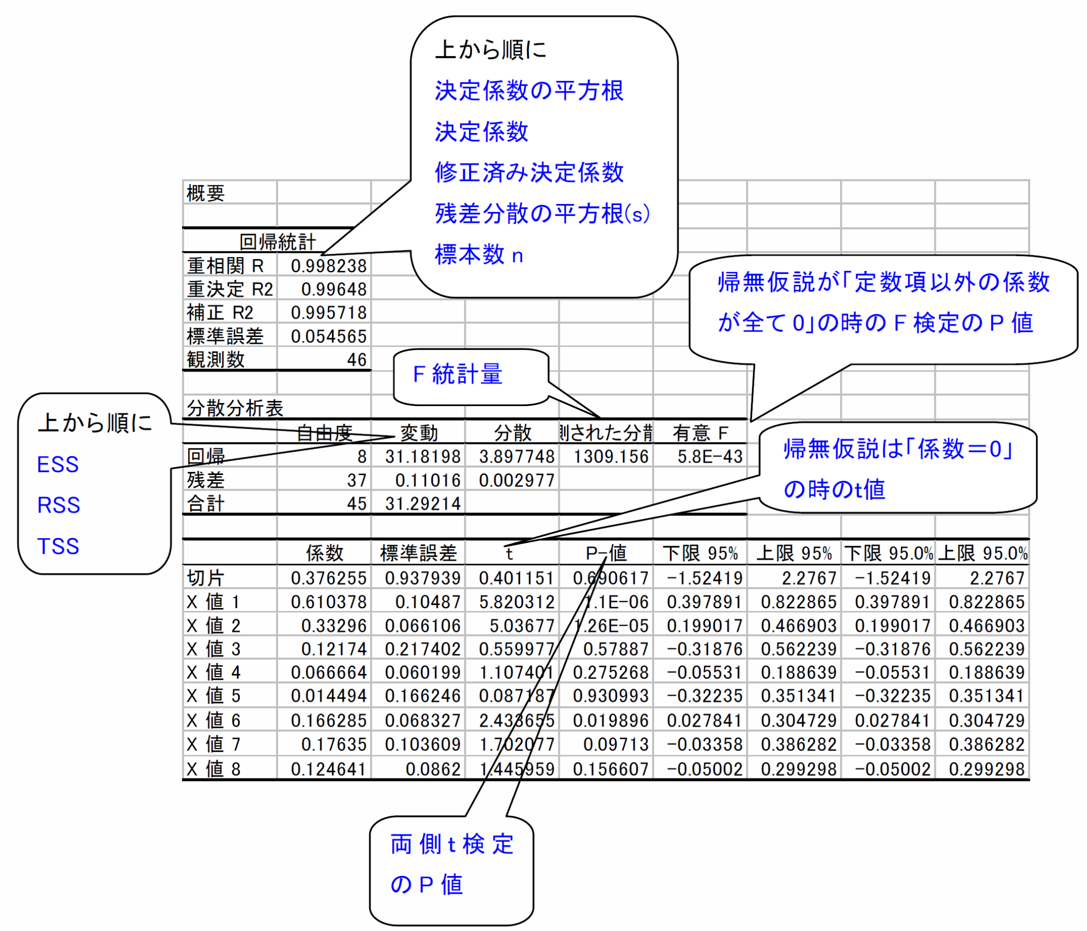
計量経済学練習問題

劉　慶豊*

平成 25 年 7 月 15 日

1　Excel による回帰分析

1.1　出力結果の読み方



概要

回帰統計	
重相関 R	0.998238
重決定 R2	0.99648
補正 R2	0.995718
標準誤差	0.054565
観測数	46

分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	8	31.18198	3.897748	1309.156	5.8E-43
残差	37	0.11016	0.002977		
合計	45	31.29214			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	0.376255	0.937939	0.401151	0.690617	-1.52419	2.2767	-1.52419	2.2767
X 値 1	0.610378	0.10487	5.820312	1.1E-06	0.397891	0.822865	0.397891	0.822865
X 値 2	0.33296	0.066106	5.03677	1.26E-05	0.199017	0.466903	0.199017	0.466903
X 値 3	0.12174	0.217402	0.559977	0.57887	-0.31876	0.562239	-0.31876	0.562239
X 値 4	0.066664	0.060199	1.107401	0.275268	-0.05531	0.188639	-0.05531	0.188639
X 値 5	0.014494	0.166246	0.087187	0.930993	-0.32235	0.351341	-0.32235	0.351341
X 値 6	0.166285	0.068327	2.433655	0.019896	0.027841	0.304729	0.027841	0.304729
X 値 7	0.17635	0.103609	1.702077	0.09713	-0.03358	0.386282	-0.03358	0.386282
X 値 8	0.124641	0.0862	1.445959	0.156607	-0.05002	0.299298	-0.05002	0.299298

*E-mail:qliu@res.otaru-uc.ac.jp。

1.2　練習問題

問題一 表 1. は Excel による

$$\begin{aligned}\log Y &= \beta_1 + \beta_2 \log L + \beta_3 \log K \\ &\quad + \beta_4 \log H1 + \beta_5 \log H2 + \beta_6 \log H3 \\ &\quad + \beta_7 \log H4 + \beta_8 \log H5 + \beta_9 \log H6\end{aligned} \tag{1}$$

の推定結果である。

表 1.

概要								
回帰統計								
重相関 R	0.998238							
重決定 R2	0.99648							
補正 R2	0.995718							
標準誤差	0.054565							
観測数	46							
分散分析表								
	自由度	変動	分散	測された分散	有意 F			
回帰	8	31.18198	3.897748	1309.156	5.8E-43			
残差	37	0.11016	0.002977					
合計	45	31.29214						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	0.376255	0.937939	0.401151	0.690617	-1.52419	2.2767	-1.52419	2.2767
X 値 1	0.610378	0.10487	5.820312	1.1E-06	0.397891	0.822865	0.397891	0.822865
X 値 2	0.33296	0.066106	5.03677	1.26E-05	0.199017	0.466903	0.199017	0.466903
X 値 3	0.12174	0.217402	0.559977	0.57887	-0.31876	0.562239	-0.31876	0.562239
X 値 4	0.066664	0.060199	1.107401	0.275268	-0.05531	0.188639	-0.05531	0.188639
X 値 5	0.014494	0.166246	0.087187	0.930993	-0.32235	0.351341	-0.32235	0.351341
X 値 6	0.166285	0.068327	2.433655	0.019896	0.027841	0.304729	0.027841	0.304729
X 値 7	0.17635	0.103609	1.702077	0.09713	-0.03358	0.386282	-0.03358	0.386282
X 値 8	0.124641	0.0862	1.445959	0.156607	-0.05002	0.299298	-0.05002	0.299298

1. 推定された回帰式を書いてください。
2. $\beta_8 = 0$ かどうかの両側検定を行ってください。有意水準は 5% とする。
3. 定数項以外の係数がすべて 0 であるという帰無仮説に関して検定を行ってください。

解答

1. 係数の推定結果を代入して得られた回帰式は

$$\begin{aligned}\log Y &= 0.38 + 0.61\log L + 0.33\log K \\ &\quad + 0.12\log H1 + 0.07\log H2 + 0.01\log H3 \\ &\quad + 0.17\log H4 + 0.18\log H5 + 0.12\log H6\end{aligned} \tag{2}$$

である。

2. 帰無仮説を $H_0 : \beta_8 = 0$, 対立仮説を $H_A : \beta_8 \neq 0$ として t 検定を行う。結論は同じだが検定を行う方法は以下の 2 種類がある。

   (a) 表の結果により β_8 の両側検定の P 値が $0.097 = 9.7\%$ であり、有意水準 5% より大きいであるため、帰無仮説を棄却できない。$\beta_8 = 0$ と考えられる。

   (b)
   $$t = \frac{\hat{\beta}_8 - 0}{s_{\beta_8}} = \frac{0.18 - 0}{0.10} = 1.8.$$
   この t 値が自由度が $n - K = 46 - 9 = 37$ の t 分布の 2.5%（両側検定であるため、元の有意水準 5% を 2 で割る）有意水準点 1.96 より小さいため、帰無仮説は棄却できない。$\beta_8 = 0$ と考えられる。

3. F 検定になる。有意水準を 5% とする。帰無仮説は $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$. 対立仮説は H_1 : 帰無仮説ではない。表から検定統計量を読み取る。$f = 1309.16$、対応する P 値は $5.8E - 43 = 5.8 \times 10^{-43}$ 殆どゼロとなっており、有意水準 5% より小さいため、帰無仮説は棄却される。

問題二 表 2. は Excel による

$$\log Y = \beta_1 + \beta_2\log L + \beta_3\log K + \beta_7\log H4 + \beta_8\log H5 \tag{3}$$

の推定結果である。

表 2.

概要								
回帰統計								
重相関 R	0.998006							
重決定 R2	0.996015							
補正 R2	0.995626							
標準誤差	0.055148							
観測数	46							
分散分析表								
	自由度	変動	分散	測された分散	有意 F			
回帰	4	31.16745	7.791862	2562.028	1.38E-48			
残差	41	0.124693	0.003041					
合計	45	31.29214						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	0.948916	0.619966	1.530593	0.13355	-0.30313	2.200964	-0.30313	2.200964
X 値 1	0.707843	0.060668	11.66756	1.32E-14	0.585322	0.830363	0.585322	0.830363
X 値 2	0.336529	0.057582	5.844375	7.2E-07	0.22024	0.452817	0.22024	0.452817
X 値 3	0.235699	0.041541	5.673919	1.26E-06	0.151805	0.319592	0.151805	0.319592
X 値 4	0.190143	0.097784	1.94453	0.058714	-0.00733	0.387621	-0.00733	0.387621

1. 推定された回帰式を書いてください。
2. 修正済みの決定係数で判断する場合、問題一と問題二の回帰式のどっちを選択するかに関して答えなさい。
3. $\beta_8 = 0$ かどうかの両側検定を行ってください。有意水準は 5% とする。
4. 問題一にある表 1. の結果とあわせて、

$$\begin{aligned}\log Y &= \beta_1 + \beta_2 \log L + \beta_3 \log K \\ &\quad + \beta_4 \log H1 + \beta_5 \log H2 + \beta_6 \log H3 \\ &\quad + \beta_7 \log H4 + \beta_8 \log H5 + \beta_9 \log H6\end{aligned} \tag{4}$$

に関して $H_0 : \beta_4 = \beta_5 = \beta_6 = \beta_9 = 0$ に関して検定しなさい。（ヒント：F 検定になる。）

解答

1. 推定された回帰式は

$$\log Y = 0.95 + 0.71 \log L + 0.34 \log K + 0.24 \log H4 + 0.19 \log H5$$

となる。

2. 表 2 の修正済みの決定係数は 0.9956 で表 1 の修正済みの決定係数 0.9957 より小さいため、問題一の回帰式を選ぶ。

3. 帰無仮説を $H_0 : \beta_8 = 0,$ 対立仮説を $H_A : \beta_8 \neq 0$ として t 検定を行う。表 2 の結果により β_8 の両側検定の P 値が $0.059 = 5.9\%$ であり、有意水準 5% より大きいであるため、帰無仮説を棄却できない。$\beta_8 = 0$ と考えられる。

4. 有意水準を 5% として、$H_0 : \beta_4 = \beta_5 = \beta_6 = \beta_9 = 0$ に関して F 検定を行う。帰無仮説の元ではモデルが

$$\log Y = \beta_1 + \beta_2 \log L + \beta_3 \log K + \beta_7 \log H4 + \beta_8 \log H5 \tag{5}$$

となるため、表 2 が帰無仮説に対応している、表 2 から $RSS\,(H_0) = 0.1247$ を読み取る。そして、表 1 は対立仮説に対応しているので、表 2 から $RSS\,(H_A) = 0.1102$ を読み取る。f 値の公式に代入して

$$\begin{aligned} f &= \frac{n-K}{m}\frac{RSS\,(H_0) - RSS\,(H_A)}{RSS\,(H_A)} \\ &= \frac{46-9}{4}\frac{0.1247 - 0.1102}{0.1102} \\ &= 1.2171 \end{aligned}$$

となる。自由度が $(m, n-K) = (4, 37)$ の F 分布の 5 ％有意水準点が F 分布表から読み取り、2.63 である（自由度が $(4, 37)$ の f 値がテキストの付表にないため、自由度が $(4, 37)$ と $(4, 50)$ の f 値を読み取って、その平均値を利用した）。$f = 1.2171 < 2.63$ であるので、帰無仮説は棄却できない。

問題三 講義用ホームページにある「性別ダミー」のデータを利用して身長（X）と体重（Y）の関係を表す線形モデル（$Y_i = \alpha + \beta X_i + u_i$）を推定し推定式を書きなさい。データの前半 25 人を女性、後半 25 人を男性として、性別ダミーのデータを作成して、ダミー変数のデータを取り入れたモデルを再推定し推定式を書きなさい。ダミー変数の係数に関して t 検定を行い、さらに $\bar{R}^2$ を利用してダミーを入れたと入れていないモデルのどちらを選択するかを判断しなさい。さらに推定式を男性式と女性式に分けて書きなさい。その上、ダミー変数の係数が表す男女差について説明しなさい。

解答

表 3.

回帰統計	
重相関 R	0.715644
重決定 R2	0.512146
補正 R2	0.501983
標準誤差	8.77458
観測数	50

分散分析表

	自由度	変動	分散	された分	有意 F
回帰	1	3879.704	3879.704	50.39019	5.19E-09
残差	48	3695.676	76.99325		
合計	49	7575.38			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	−56.2212	16.30417	−3.44827	0.001184	−89.0029	−23.4394	−89.0029	−23.4394
身長(cm)	0.698723	0.098431	7.098605	5.19E-09	0.500815	0.896632	0.500815	0.896632

推定結果：

$$\hat{Y}_i = -56.22 + 0.6987X_i$$

となった。

表 4.

回帰統計	
重相関 R	0.759558
重決定 R2	0.576928
補正 R2	0.558925
標準誤差	8.257725
観測数	50

分散分析表

	自由度	変動	分散	された分	有意 F
回帰	2	4370.449	2185.225	32.04611	1.66E-09
残差	47	3204.931	68.19002		
合計	49	7575.38			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	−58.5731	15.36882	−3.81117	0.000401	−89.4912	−27.655	−89.4912	−27.655
男女ダミー	6.266884	2.336061	2.682672	0.010049	1.567336	10.96643	1.567336	10.96643
身長(cm)	0.693992	0.09265	7.490472	1.48E-09	0.507604	0.880379	0.507604	0.880379

ダミー変数 M_i を女性なら 1、男性なら 0 として定義する。ダミー変数を取り入れた推定結果：

$$\hat{Y}_i = -58.57 + 6.27M_i + 0.694X_i$$

となった。

M_i の t 値が 2.68 となり、対応している P 値は $0.01 < 5\%$ であるので、両側検定 5% で M_i の係数が 0 である帰無仮説が棄却される。すなわち、男女の性別による身長と体重の間の関係に差が存在する。

ダミー変数を取り入れた式の $\bar{R}^2 = 0.558925$ となり、ダミー変数を取り入れていない式の $\bar{R}^2 = 0.501983$ より大きなので、ミー変数を取り入れたモデルを選択する。

推定結果を男性式と女性式に分けて書く。

$$
\begin{aligned}
\text{男性式：}\hat{Y}_i &= -58.57 + 0.694X_i \\
\text{女性式：}\hat{Y}_i &= (-58.57 + 6.27) + 0.694X_i \\
&= -52.3 + 0.694X_i
\end{aligned}
$$

二つの式の定数項が異なる。女性式の定数項が男性式より 6.27 大きくなっている。すなわち、同じ身長であっても女性の体重が男性より重くなりがちである。

問題四 偏相関係数の計算法を述べた上で、その役割を説明しなさい。

解答 第三章の資料の 4、5、6 ページを参照しなさい。

問題五 Y は生産額、K は資本額、L は労働の投入量とする。コブ · ダグラス生産関数は $Y = \alpha K^\beta L^\gamma$ となり、以下のように変形できる。

$$\log Y = \alpha^* + \beta \log K + \gamma \log L$$

収穫の不変性の意味合いを説明し、係数 β と γ がどのような条件を満たすとき収穫の不変性を持つか説明しなさい。

解答 第三章資料 35 ページとテキスト 100 ページを参照しなさい。

問題六 自由度修正済み $\bar{R}^2$ の公式を書きなさい。その上、何故モデルを選択するとき R^2 ではなく $\bar{R}^2$ を利用するのかを説明しなさい。

解答 第四章の資料の 11 と 12 ページおよびテキスト 116 ページを参照しなさい。

問題七 テキスト 149 ページ練習問題 3 を答えなさい。追加に、トランスログ式の DW 値を利用して正の自己相関があるかどうかを検定しなさい。

解答 a.　$DW \approx 2 - 2\hat{\phi}$ の公式を利用すれば、自己相関係数の推定値 $\hat{\phi} \approx (2 - DW)/2 = 0.1$。

b.　F 検定を行って二つの式から一つを選択するため、$(\log L)^2, (\log K)^2$ と $(\log L \log K)$ の係数を β_4, β_5 と β_6 と表記して、以下の仮説を検定すれば良い。有意水準を 5% とする。

$$
\begin{aligned}
H_0 &: \beta_4 = 0, \beta_5 = 0, \beta_6 = 0. \\
H_A &: \beta_4 \neq 0, \beta_5 \neq 0, \beta_6 \neq 0.
\end{aligned}
$$

帰無仮説のモデルは短い式で、対立仮説のモデルはトランスログ式であるので、$RSS\,(H_0) = 0.88,\ RSS\,(H_a) = 0.66$。$F$ 値の公式より、

$$\begin{aligned} f &= \frac{n-K}{m}\frac{RSS\,(H_0) - RSS\,(H_A)}{RSS\,(H_A)} \\ &= \frac{27-6}{3}\frac{0.88-0.66}{0.66} = 2.33. \end{aligned}$$

F 値 2.33 が自由度 $(3, 21)$ の F 分布の 5% 臨界値 3.1 より小さいため、帰無仮説を採択する。短いモデルを選択する。

追加の問題に関しては、仮説を以下のようにする。

$$\begin{aligned} &H_0: \text{自己相関がない、}\phi = 0 \\ &H_1: \text{正の自己相関がある}\phi > 0 \end{aligned}$$

観測個数 $n = 27,\ K = 6,\ K - 1 = 5,\ DW = 1.8$ なので、5% の検定では上限と下限がそれぞれ $U = 1.86$ と $L = 1.0$。$L < DW < U$ なので系列相関があるかどうか結論できない（不決定)。

問題八 1980 年から 2008 年のデータで消費（C_t）と国内総生産（GDP_t）との関係を分析する。1992 年と 1993 年の間に構造変化があったかどうかを F 検定で検定してください。ただし、推定の結果で帰無仮説のモデルの RSS= 3.566、対立仮説の RSS= 1.368 とし、自由度 $(2, 25)$ の F 分布の 5% 臨界値 $F_{0.05}\,(2, 25) = 3.39$ である。

解答 ダミー変数 D_t が 1980 − 1992 年の期間中で 1 と、1993 年以後は 0 と定仮説を立てる。

$$\begin{aligned} &H_0: \lambda = \gamma = 0 \\ &H_1: \lambda\text{、}\gamma\text{のいずれか又はすべて 0 ではない} \end{aligned}$$

帰無仮説と対立仮説のモデルを構築する。

$$\begin{aligned} &\text{帰無仮説のモデル：} \quad C_t = \alpha + \beta GDP_t + u_t \\ &\text{対立仮説のモデル：} \quad C_t = \alpha + \lambda D_t + \beta GDP_t + \gamma D_t \times GDP_t + u_t \end{aligned}$$

F 値を計算する

$$\begin{aligned} f &= \frac{n-K}{m}\frac{RSS\,(H_0) - RSS\,(H_A)}{RSS\,(H_A)} \\ &= \frac{29-5}{2}\frac{3.66-1.368}{1.368} = 20.105 \end{aligned}$$

$f >$ 臨界値 $= 3.39$ なので帰無仮説を棄却する。構造変化があったと認める。

問題九 多重共線性がある場合のデータ回帰分析の結果の特徴を述べなさい。対処法を説明しなさい。

解答 多重共線性がある場合、変数間の相関が高く、１つの変数が相関の高い他の変数の効果を吸収してしまい、係数の推定結果が本来の符号を示さない。t 値も有意になりにくい。互いに相関が高い変数の一方をモデルから削除する方法で解決できる。

問題十 モデル $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ の誤差項 u_i に不均一分散があるとする。分散の構造は $V\left(\frac{u_i}{x_{2i}}\right) = \sigma^2$ の場合、どのようにデータを変換して推定すれば分散不均一性を解決できるかについて述べなさい。その理由も説明しなさい。

解答 定数項を $\frac{1}{x_{2i}}$、y_i, x_{2i} と x_{3i} をそれぞれ $\frac{y_i}{x_{2i}}, 1, \frac{x_{3i}}{x_{2i}}$ と変換してモデル

$$\frac{y_i}{x_{2i}} = \beta_1 \frac{1}{x_{2i}} + \beta_2 + \beta_3 \frac{x_{3i}}{x_{2i}} + \frac{u_i}{x_{2i}}$$

を最小 2 乗法で推定すれば良い。それは新しいモデルの誤差項 $\frac{u_i}{x_{2i}}$ の分散が σ^2 となり、i に依存しなく、分散均一になったためである。