第５章　練習問題解答例

(1)
① この場合は分析対象、従属変数は例えば保有している場合 1、保有していない場合 0、という離散的な値をとることになる。したがってプロビット・モデルまたはロジット・モデルを用いて分析を行っていきます。
② この場合は従属変数が 0 という水準で打ち切られているが、説明変数に関する情報は利用可能なので、閲覧モデルを用いて分析を行っていきます。
③ この場合は従属変数・説明変数のいずれに関しても打ち切りが存在している場合なので切断分布モデルを用いて分析を行っていきます。

(2)
① この場合は、従属変数に関しての上限が存在するというケースになります。つまり合格点以上の得点をとった受験者に関しては合格という事実が分かるのみで、実際の得点は分らないという想定です。ただし、合格者に関しても属性データは存在するという想定ですので、従属変数に関して打ち切りが存在するが、説明変数に関する情報は利用可能なので、閲覧モデルを用いて分析します。ただし、この場合は今までの例とは異なり、モデルの推定プロセスにおいて upper truncation point を指定することになります。
② 大学教官グループに利用可能でない個人的属性が存在し、それが合格できる水準まで十分な学習を積むかどうかという意思決定プロセスに影響を与え、かつ実際の入学試験の得点にも影響を与えるということはそれほど無理のない想定でしょう（ただし、ここでは個人が十分な努力をすることによって入学試験合格水準に到達できるという仮定をしています。現実的ではないかもしれませんが、とりあえずこう仮定して分析をすすめてみましょう）。したがってこの場合にはサンプル・セレクション・モデルの適用が可能となります。選択メカニズム方程式は、従属変数が合格、不合格という離散的な変数であり、利用可能な個人的属性（全部あるいは一部）が説明変数であるプロビット・モデルによって定式化されます。入試得点決定方程式では、利用可能な個人的属性（全部あるいは一部）を説明変数として、不合格者の入試得点（合格者の場合は一定の値）を従属変数として定式化されます。

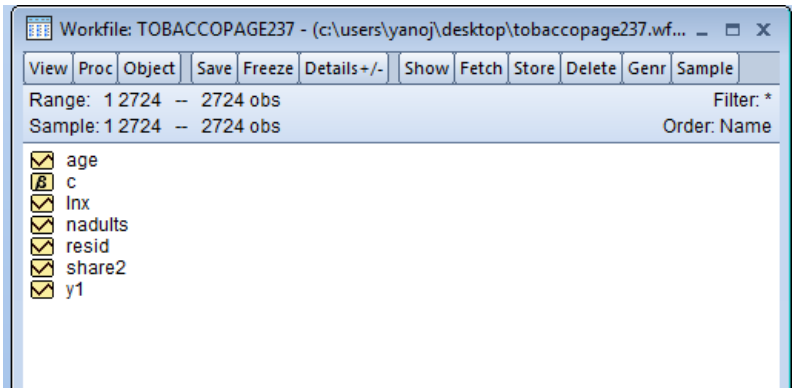
(3)
　Verbeek 教授のサイトにある tobacco.wf1 からここでの分析に必要なデータのみを含んでいるのが、tobacco_page237.wf1 のファイルです。このファイルを Eviews に読み込ませた状態が下図になります。変数はそれぞれ以下のようになっています。
　Share2：タバコ支出割合
　age：0 から 4 歳児の年齢

lnx：家計の総支出額

nadults：家計の成人者数

y1：タバコ支出割合が正ならば 1、そうでない場合は 0

データ読み込み



まずは、1.5 と同様にプロビット・モデルの推定を行います。以下のモデルは一つの例示にすぎません。Tobacco.wf1 には多くの変数が含まれていますので、その他の特定化にも是非チャレンジしてみて下さい。

推定するモデルは。

$$P_i = p_r(share2 > 0) = \beta_0 + \beta_1 \ln x + \beta_2 nadults + \beta_3 age$$

です。推定結果は次の図に示されています。

推定結果

EViews - [Equation: LNXNADULTSAGE Workfile: TOBACCO2PROBITNKIDS12NADULTS::Tobac

File Edit Object View Proc Quick Options Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y1
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 02/21/14 Time: 12:59
Sample: 1 2724
Included observations: 2724
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	3.205430	0.788608	4.064667	0.0000
LNX	-0.254771	0.059051	-4.314432	0.0000
NADULTS	0.156021	0.034653	4.502398	0.0000
AGE	-0.135910	0.018333	-7.413249	0.0000

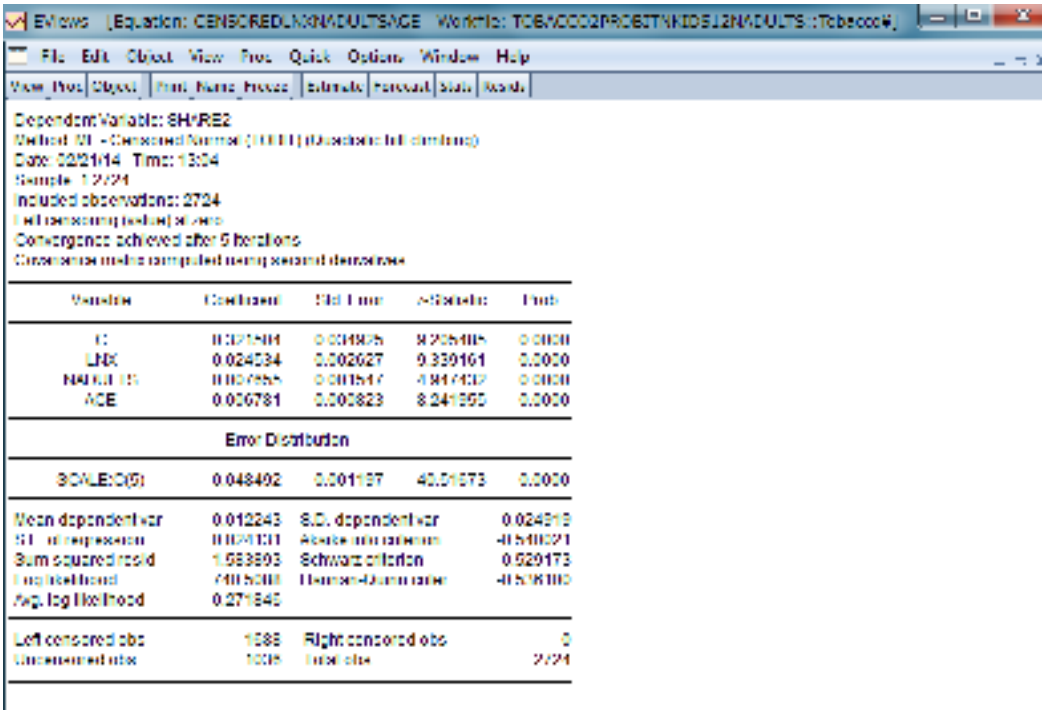
McFadden R-squared	0.020136	Mean dependent var	0.380323
S.D. dependent var	0.485555	S.E. of regression	0.479453
Akaike info criterion	1.304632	Sum squared resid	625.2618
Schwarz criterion	1.313310	Log likelihood	-1772.908
Hannan-Quinn criter.	1.307768	Deviance	3545.816
Restr. deviance	3618.682	Restr. log likelihood	-1809.341
LR statistic	72.86530	Avg. log likelihood	-0.650847
Prob(LR statistic)	0.000000		

Obs with Dep=0	1688	Total obs	2724
Obs with Dep=1	1036		

得られた結果は、家計の所得水準が高いほど（家計の支出額おおいほど）、幼児の年齢数が多いほど、家計の成人者数が少ないほど、タバコへの支出割合が少なくなるということを示しています。

次に閲覧回帰モデルと切断回帰モデルを推定してみましょう。2.2 と同様なプロセスにより、以下の推定結果が得られます。

閲覧回帰モデル



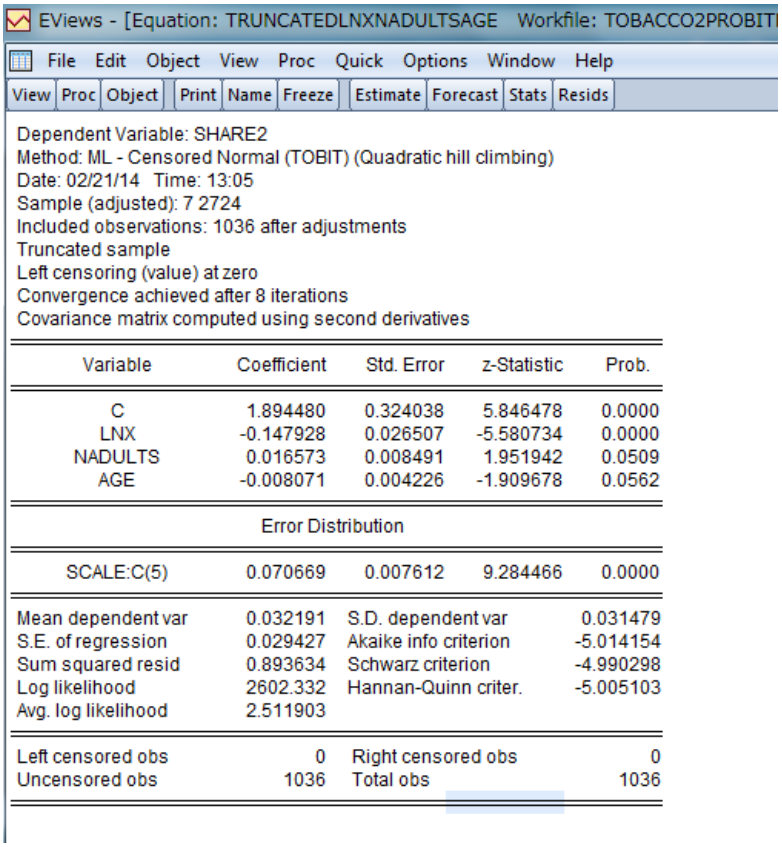
Dependent Variable: SHARE2
Date: 02/21/14 Time: 13:04
Included observations: 2724
Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	[illegible]	[illegible]	[illegible]	[illegible]
LNX	0.024234	0.002627	9.339161	0.0000
NADULTS	[illegible]	[illegible]	[illegible]	[illegible]
AGE	0.006781	0.002823	8.241595	0.0000
Error Distribution				
SCALE:C(5)	0.048402	0.001197	40.51573	0.0000

Mean dependent var	0.012243	S.D. dependent var	0.024519
S.E. of regression	[illegible]	Akaike info criterion	[illegible]
Sum squared resid	1.583893	Schwarz criterion	0.629173
Log likelihood	[illegible]	Hannan-Quinn criter.	[illegible]
Avg. log likelihood	0.271845		

Left censored obs	1588	Right censored obs	0
Uncensored obs	1036	Total obs	2724

切断回帰モデル



EViews - [Equation: TRUNCATEDLNXNADULTSAGE Workfile: TOBACCO2PROBIT

File Edit Object View Proc Quick Options Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: SHARE2
Method: ML - Censored Normal (TOBIT) (Quadratic hill climbing)
Date: 02/21/14 Time: 13:05
Sample (adjusted): 7 2724
Included observations: 1036 after adjustments
Truncated sample
Left censoring (value) at zero
Convergence achieved after 8 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.894480	0.324038	5.846478	0.0000
LNX	-0.147928	0.026507	-5.580734	0.0000
NADULTS	0.016573	0.008491	1.951942	0.0509
AGE	-0.008071	0.004226	-1.909678	0.0562
Error Distribution				
SCALE:C(5)	0.070669	0.007612	9.284466	0.0000

Mean dependent var	0.032191	S.D. dependent var	0.031479
S.E. of regression	0.029427	Akaike info criterion	-5.014154
Sum squared resid	0.893634	Schwarz criterion	-4.990298
Log likelihood	2602.332	Hannan-Quinn criter.	-5.005103
Avg. log likelihood	2.511903		

Left censored obs	0	Right censored obs	0
Uncensored obs	1036	Total obs	1036

基本的には probit モデルと同様の推定結果が得られています。

それでは、サンプル・セレクション・モデルを推定してみましょう。以下のようなモデルを考えましょう。

［選択モデル］ $P_i = p_r(share2 > 0) = \beta_0 + \beta_1 age$

［回帰モデル］ $share2 = \beta_0 + \beta_1 \ln x + \beta_2 nadults$

このモデルも例示にすぎませんが、タバコを吸うか吸わないかの意志決定には幼児の年齢が影響し、吸う場合の支出額の決定には、所得水準、成人者数が関係し、この二つの意志決定には我々が観察できない相関する要因があると考えているわけです。推定結果は以下になります。

推定結果

Equation: HECKIT01 Workfile: TOBACCO2WITHY1PROBIT02::Tobacco\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SHARE2
Method: Heckman Selection
Date: 02/21/14 Time: 14:04
Sample: 1 2724
Included observations: 2724
Selection Variable: Y1
Estimation method: Maximum likelihood (Quadratic Hill Climbing)
Covariance matrix: Default (Hessian - observed)
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Response Equation - SHARE2				
C	0.394602	0.028399	13.89513	0.0000
LNX	-0.026320	0.002127	-12.37480	0.0000
NADULTS	0.002176	0.001291	1.685962	0.0919
Selection Equation - Y1				
C	-0.006805	0.048777	-0.139523	0.8890
AGE	-0.125803	0.017938	-7.013240	0.0000
Interaction terms				
@LOG(SIGMA)	-3.522969	0.028004	-125.8014	0.0000
@ATAN(RHO)*2/PI	-0.335145	0.200731	-1.669617	0.0951
SIGMA	0.029512	0.000826	35.70891	0.0000
RHO	-0.205870	0.114885	-1.791959	0.0733

Mean dependent var	0.032191	S.D. dependent var	0.031479
S.E. of regression	0.018337	Akaike info criterion	-0.295853
Sum squared resid	0.913568	Schwarz criterion	-0.280666
Log likelihood	409.9521	Hannan-Quinn criter.	-0.290364

得られた結果は我々の想定通りであり、また sigma、rho 推定結果はサンプル・セレクションの存在を示唆しています。予想されたように、家族所得、両親の教育年数はいずれも

子供の高校卒業確率に有意な正の影響を持つことを示唆する結果となっています。