

## 第4章 練習問題解答例

### 1.

- (i) 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ゆきます。
- (ii) この場合は従属変数がゼロ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ゆきます。
- (iii) 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斷分布モデルを用いて分析を行ってゆきます。

### 2.

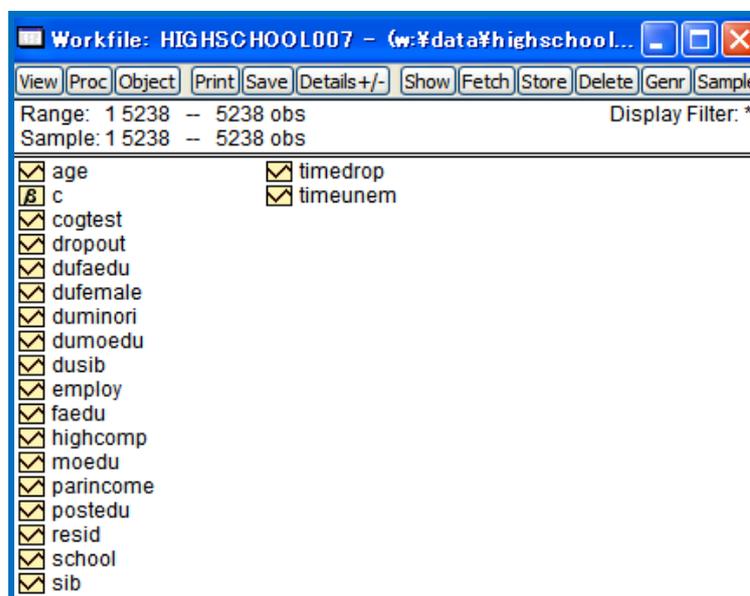
- (i) この場合は、従属変数に関しての上限が存在するというケースになります。つまり合格点以上の得点をとった受験者に関しては合格という事実が分かるのみで実際の得点は分らないという想定です。ただし、合格者についても属性データは存在するという想定ですので、従属変数に関して打ち切りが存在するが、説明変数に関する情報は利用可能なので、閲覧モデルを用いて分析します。ただし、この場合は今までの例とは異なり、モデルの推定プロセスにおいて **upper truncation point** を指定することになります。
- (ii) 大学教官グループに利用可能でない個人的属性が存在し、それが合格できる水準ま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. まず、使用するデータセットが含まれている `graduationml-data.zip` ファイルと説明ファイルである `graduationreadme.ml.txt` をダウンロードします。zip ファイルを解凍しますと、`jae-4486-5-data-2-txt` ファイルと `jae-4486-5-data-3-txt` の二つのテキストファイルが出来ます。ここでは `jae-4486-5-data-2-txt` ファイルを用いて分析してみましよう。

このテキストファイルを `excel` に読み込ませると（「スペースによって右または左に揃えられた固定長フィールドのデータ」として読み込ませないといけないので注意して下さい）、17 変数、5238 個のオブザベーションが含まれています。

左端から 2 列目、4 列目、5 列目、6 列目、7 列目がそれぞれ家族所得、父親教育年数、その関連ダミー、母親教育年数、その関連ダミーとなっています。16 列目が高校卒業のデータですが、卒業した場合が 0 で、その他卒業できなかった場合が 1,2,3 の値となっています。このままでは二値変数ではありませんので、卒業した場合は 1、それ以外をゼロとする新たな変数をエクセルで作成しましょう。（if 文を使って例えば `=IF(Q3<0.8,"1","0")` のように打ち込みコピーして下さい。この新たな変数はそのままでは `evIEWS` では数値として認識されませんので、一度「形式を指定して貼り付け」で、数値として貼り付けてください）出来上がったエクセルシートを `evIEWS` に読み込ませた状態が図 4.27 です。

図 4.27 データ読み込み



後は、1.5 と同様にプロビット・モデルの推定を行います。推定するモデルは

$$P_i = p_r(\text{highcomp} = 1) = \beta_0 + \beta_1 \text{Parincome} + \beta_2 \text{father's education} + \beta_3 \text{dummy for father's education} \\ + \beta_4 \text{mother's education} + \beta_5 \text{dummy for mother's education}$$

です。推定結果は図 4.28 に示されています。

図 4.28 推定結果

	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.066297	0.183810	-5.801068	0.0000
PARINCOME	0.073604	0.025850	2.847325	0.0044
FAEDU	0.076183	0.012652	6.021404	0.0000
DUFAEDU	0.778212	0.177656	4.380455	0.0000
MOEDU	0.105233	0.015447	6.812375	0.0000
DUMOEDU	1.041488	0.214107	4.864341	0.0000
McFadden R-squared	0.074273	Mean dependent var	0.905307	
S.D. dependent var	0.292818	S.E. of regression	0.285371	
Akaike info criterion	0.582282	Sum squared resid	426.0774	
Schwarz criterion	0.589800	Log likelihood	-1518.996	
Hannan-Quinn criter.	0.584911	Restr. log likelihood	-1640.869	
LR statistic	243.7451	Avg. log likelihood	-0.289995	
Prob(LR statistic)	0.000000			
Obs with Dep=0	496	Total obs	5238	
Obs with Dep=1	4742			

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。