
LOGIT (CASE=変数, COND, NCHOICE=n, NREC=変数, SUFFIX=リスト, 非線形オプション)
従属変数 独立変数 ;
あるいは 従属変数 条件付変数 | 多項(選択肢) 変数 ;

機能 :

LOGIT は条件付きロジット や 多項ロジットモデルの推定に使われます. モデル中の説明変数は, 観測対象毎に選ばれる選択肢にともなって変化してもよいし, あるいは観測対象毎の特性を表している場合でもよく, さらに両方の性質を持っていても使えます. 選択肢の数については制限がありません.

使用法:

ロジットモデルには3つの基本的な形があります. 第1はすべての選択肢について, 各観測対象の持つ独立変数の値が同じ場合です. 観測対象が変われば独立変数の値もかわります. この場合の独立変数を, 多項変数と呼ぶこともあります. 第2は, 独立変数が選択肢の特性を表すもので, 第3は両方の性質の独立変数を持つ混合モデルです. 第1のタイプは多項ロジットと呼ばれ, 独立変数にともなう係数が, 1つを除いた全ての選択肢について異なっています. 第2のタイプは条件付きロジットと呼ばれ, 独立変数は選択肢にともなって値を変えます. もちろん観測対象が変われば独立変数の値も変わります.

1. 2項, あるいは多項ロジット モデル: OLSQ や PROBIT と同様です.

```
LOGIT   従属変数   多項変数  (選択肢の特性);
LOGIT  Y  C  X1  X2 ... XK ;
```

Y は0/1, あるいは1/2, あるいは任意の2個の整数値を採ることができます. Y が2個以上の整数値を採る場合を多項ロジットモデルといいます. 係数名は, 各独立変数名に各選択肢に対する Y の値を付け加えたものとします. 係数は最小の選択に対する係数をゼロとして基準化されます. つまりYが0/1ならば, 係数 C0,X10,X20,...,XK0 はゼロに基準化されて, C1,X11,X21,...,XK1 が推定されます. Yが 1/2/3 の値を取る場合は, 選択肢1に対応する係数C1,X11,X21,...,XK1がゼロに基準化され, C2,X12,X22,...,XK2; C3,X13,X23,...,XK3 が推定されます.

```
LOGIT (NCHOICE=2) Y C X1 X2 ... XK ;
```

とすると, TSP は Y の範囲を2つの選択肢だけであると確認します. 推定されるモデルには, K+1 個の係数と K+1 個の変数があります.

2. 条件付きロジット モデル:

```
LOGIT (COND ,NCHOICE=n)  従属変数    条件変数  (選択肢特性) ;
```

たとえば

```
LOGIT (COND, NCHOICE=2) Y X Z ;
```

では2つの選択肢に対応する変数 X1,X2,Z1,Z2 を探します. 係数 X と Z が推定されます. 定数項Cは条件変数としては使えません (Cは選択肢によって変化しないので, 識別できないためです). ダミー変数の特定の選択肢に対して, Cを混合ロジットモデルで多項変数として使うことができます. この場合は, 上の例は,

```
LOGIT (COND, NCHOICE=2) Y X Z | C ;
```

となります．C により, 選択肢にかかわらず固定された係数と, 選択肢毎に変化する定数ダミーがモデルに入ります．CASE オプションは, 観測対象当たり 1 つのケースでなく, 観測対象当たり 1 つの選択肢を持つデータセットを作れます．例えば,

```
LOGIT (CASE=V) Y X Z ;
```

とすれば, 変数 V はケースを区別する ID 変数で, 同じケースに関する隣接する観測値は同じ値を持たないといけません．この場合は, 変数名は本来の名前が使えます．つまり X1,X2,Z1,Z2 ではなく X,Z でよいのです．ケース毎に同じ数の観測値がある必要はありません．各ケースで最初の Y だけを有効な選択肢の数として調べます.

ケース毎のデータセットで複数の観測値がある場合, NREC を用いることができます.

```
LOGIT (NREC=W) Y X Z ;
```

W はケース毎のレコード数を指定し, この場合 CASE オプションを付けた ID 変数を与える必要はありません.

3．混合ロジットモデル:

このコマンドの一般的な形は

```
LOGIT (COND,NCHOICE=N)  従属変数  条件付き変数 | 独立変数(多項変数) ;
```

となります．例えば

```
LOGIT (COND,NCHOICE=3) Y ZA ZB | XA XB XC ;
```

となります．Y は 1,2,3 の値を取ります．TSP は 3 つの選択肢に対応する条件変数, ZA1,ZA2,ZA3,ZB1,ZB2,ZB3 を探します．XA,XB,XC はケース毎に変わる多項変数です．係数 ZA,ZB,XA2,XB2,XC2,XA3,XB3,XC3 が, XA1,XB1,XC1 をゼロに基準化して推定されます.

オプション:

非線形推定に関する標準のオプションが使えます．このマニュアル中の, NONLINEAR のセクションを参照して下さい．何も指定しない場合は, HITER=N と HCOV=N がヘッセ行列の近似と標準誤差の計算に使われます．さらにロジットモデルには, 次のオプションがあります.

CASE= ケース変数 これはケース毎に複数の観測値があるとしてデータのインプットをする場合に使われます．この変数は, 同じケースに属する観測値は同じ値を持ちます．また同一のケースに属する, あるいはケース変数が同じ値を持つ観測値は, 隣接しないといけません．ケース変数としてどの様な変数が使われてもかまいません.

COND/**NOCOND** 条件付モデルと混合モデル対純粋な多項モデルの推定に対して指定します (使用法の部分を参照して下さい)．もし CASE=あるいは NREC=を用いた場合, COND が仮定されるので, 指定する必要はありません.

NCHOICE= 選択肢の数 全ての観測値が同じ数の選択肢を持つ場合に使われます．プログラムは, データがこの条件を満たすかどうか検討します．これは CASE=あるいは NREC=と一緒には用いません, この場合には常に最初の選択をとります.

NREC= 選択肢を数える変数 ケース毎に多数の選択肢がある場合に使われます．この変数は, 各ケースに含まれる観測値の数を指定します．通常, データ中のNREC変数値は観測値毎に繰り返されます．しかし, 各ケースにおける最初の観測値の値だけが計算に使われます．NREC=3はできませんが, NCHOICE=3はできます．

SUFFIX=リスト は, 選択の部分集合を選ぶ別の方法です．たとえば, NCHOICE=2はSUFFIX=(1,2)と同じです．SELECTコマンドもこのために使えます．このオプションは十分にテストされていません．

例:

```
NOSUPRES @DPDX;
LOGIT Y C X;
```

このプロシジャーの使用法についての他の例は, 使用法のセクションを参照して下さい．

アウトプット:

ロジットプロシジャーのプリントアウトは, TSPの他の非線形推定法のプリントアウトと似ています．まずタイトルがきます．次に各選択肢の選択状況を示す度数分布表が与えられます．次に初期値, そして反復毎の結果が印刷されます．印刷量は, PRINT,SILENT のオプションによって調整されます．以下, 反復計算の収束状況に関する情報, 尤度関数の値, そして係数推定値, その漸近的な標準誤差, そして t-統計量の表が続きます．SUPRES されていなければ, 係数の分散共分散行列もプリントされます．@DPDXあるいは@DPDZ行列が, SUPRESされていなければプリントされます．データ領域には, 次の結果が保存されます．

名前	タイプ	長さ	変数の説明
@RNMS	リスト	パラメータ数	係数名
@LOGL	スカラー	1	対数尤度関数
@IFCONV	スカラー	1	収束状況 (1は成功)
@KLRSQ	スカラー	1	多項ロジットでのKullback-Leiblerの R^2
@RSQ	スカラー	1	バイナリーロジットのYと@FITの2乗した相関係数
@SSR	スカラー	1	バイナリーロジットの残差 (Y-@FIT) 平方和
@GRAD	ベクトル	パラメータ数	極大値における尤度関数の勾配
@COEF	ベクトル	パラメータ数	係数推定値
@SES	ベクトル	パラメータ数	係数推定値の標準誤差
@T	ベクトル	パラメータ数	t 統計値
%T	ベクトル	パラメータ数	t 統計量の p 値
@VCOV	ベクトル	パラメータ数**2	係数推定値の共分散行列
@DPDX	行　列	変数の数×選択肢数	多項変数に関する平均確率導関数の行列 この行列は, (係数セット間の差のように) ゼロに基準化される係数セットに対して不変である
@DPDZ	行　列	変数の数×選択肢数	条件変数の平均確率導関数の行列 (変数の数=選択肢数×条件変数) この行列は, 次元が (条件数×選択肢数) の小行列が, 縦に選択肢数の数だけ並ぶ またブロック対称で, 選択肢数が同一でないと計算されない
@FIT	行　列 又は系列	観測値数×選択肢数 又は観測値数	選択肢数が2より大で各ケースに1個の観測値しかない場合, 推定された確率が計算される．選択肢数が2のときやケース ごとに複数の観測値があるようなデータセットの場合は, 観測値数×1の系列変数になる．選択肢が2なら, @FITは 最も頻度の高い選択に対する確率の結果だけを含む.

方法:

もし C_t が t 番目の観測対象に対する選択のセットで, また t 番目の観測対象が C_t から i 番目の選択肢を撰んだならば, 選択確率の表現は次のようになります．

$$P(i\text{が }C_t|Z\text{より選択される}) = \exp(Z_{ti}b)/\sum_j \exp(Z_{tj}b)$$

そして尤度関数は

$$\log L = -\sum_t \log[\sum_j \exp((Z_{tj} - Z_{ti})b)]$$

となります. 推定される係数ベクトルは b です. Z 変数のいくつかが選択肢に対応して変化しなければ, Z 行列は,固定型の Z 変数行列と,次元が選択肢数に等しい恒等行列のクロネッカー積になっていると考えればよいでしょう. しかし実際の計算では, クロネッカー積を使わず, 条件付き変数と, 多項ロジットにおける独立変数を, 記憶領域を節約するように別々の方法で扱います.

パラメータ b に関する尤度関数の最大化は, ニュートン法によってなされます. 尤度関数が大域で凹型ですから, 計算は簡単です. 指定のないときは初期値は零ですが, @START によって初期値を与えることもできます. このマニュアル中の NONLINEAR のセクションを参照して下さい.

尤度関数に含まれる指数関数 (EXP()) と導関数は, 数値のオーバーフローと零で除することを避けるため評価されます. この様な状況が起きれば, 適切な上限がおかれます. この方法は, 多少不正確な結果を生じるかも知れませんが, 計算を中止するよりはましでしょう. この様な問題が生じる観測対象は, @FIT ベクトル中で 1 か 0 の値によって判別することができます.

推定前に LOGIT はデータとこの条件のフラッグとの単変量完全分離と準完全分離をチェックします (モデルはこの場合識別されません). このチェックがないと, 1つ以上の RHS 変数がいくつかの観測値に対して Y を完全に予測し, それらの係数は + 無限大あるいは − 無限大をゆっくりと繰り返します.

基準化された R^2 は定数項だけのモデルに関する適合度の尺度です：Estrella(1998) 参照.

参考文献:

Albert, A., and J.A. Anderson, “On the Exsistence of Maximum likelihood Estimates in Logistic Regression Models,” in **Biomerika,** 71, 1984.

Amemiya, Takeshi, **Advanced Econometrics,** Harvard University Press, 1985,Chapter 11.

Cameron, A. Colin, and Frank A. G. Windmeijer, “Am R-squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models,” **Journal of Econometrics,** 77, 1997, pp.329-342.

Maddala, G. S., **Limited Dependent Variables and Qualitative Variables in Econometrics,** Cambridge University Press, 1983, Chapter 2 and 3.

McFadden, Daniel S., “Conditional Logit Analysis of Qualitative Choice Behavior,” in Zarembka, P., ed., **Frontiers in Econometrics,** Academic Press, 1973.

McFadden, Daniel S., “Quantal Choice Analysis: A Survey,” **Annals of Economic and Social Mearsurment,** Vol.5, 1976, pp.363-390.

Nerlove, Marc and S. James Press, **Univariate and Multivariate Loglinear and Logistic Models,** Rand Report No. R-1306-EDA/NIH, 1973.

Train, Kenneth, **Quantitative Choice Analysis,** The MIT Press, Cambridge, MA, 1986.