第 6 回　重回帰モデル：推定①

[1] 重回帰モデルによる母集団の表現

変数 $x_1,\cdots,x_k$ が y にどのように影響するかを統計的に推測するとき，「母集団」は何か？

母集団（知りたい対象の全体）： y と $x_1,\cdots,x_k$ が発生する仕組み

↓

この仕組みを知れば，実際に y と $x_1,\cdots,x_k$ がどう発生するかをすべて説明できる。

重回帰モデル： y と $x_1,\cdots,x_k$ が発生する仕組みは，次の関係で表されると考える。

$$y=\beta_0+\beta_1x_1+\cdots+\beta_kx_k+u \quad \cdots①$$

$x_1,\cdots,x_k$ ： **説明変数**　（非確率変数）

u　　： **誤 差 項**　（確率変数（$x_1,\cdots,x_k$ で説明できない要因））

$\beta_0,\cdots,\beta_k$ ： **パラメータ**（y の分布の特徴を表す定数）

y　　： **従属変数**（①式に従って発生する確率変数）

⇒ 重回帰モデルで表される母集団の特徴：$\beta_0,\beta_1,\cdots,\beta_k$ と u の発生方法

⇒ これらがわかれば，y と $x_1,\cdots,x_k$ の発生する仕組み（y と $x_1,\cdots,x_k$ の関係）がわかる。

・なぜ単回帰モデルから重回帰モデルへの拡張が必要なのか？

(1) 一般に，ある変数 y に影響する要因はたくさんある。このような状況では，ある要因 x が変数 y に及ぼす影響だけを知りたいとしても，「条件一定化」が必要になる。

(2) 重要な要因を見落とすと，最小二乗推定量の望ましさがなくなる（除外変数の問題）。

(3) 重要な要因は x だけであっても，x は単回帰モデルでは表現できない形で y に影響するかもしれない（関数形の問題）。

重回帰モデルの必要性(1)～(3)について，具体例を使って調べてみる。

(1) 条件一定化の必要性

教育年数 educ の時給 wage への効果を知りたいとき，次の式を測ればよいのか？

$$wage = \beta_0 + \beta_1 educ + u \quad \cdots \quad ②$$

実際には exper（労働市場への参加年数）も wage に影響するとすれば，単回帰モデル②を使うと，教育年数の時給への影響は正しく測れない。

・偶然に「条件一定化」が成立する場合（偶然に正しい関係が測れる場合）

	wage	educ	exper	
A さん	1600 円	16 年	4 年	⇒ 教育年数が 1 年増えると
B さん	1200 円	12 年	4 年	時給は 100 円上がる。

単回帰モデル②の関係　　無視される部分

・「条件一定化」が成立しない場合（正しい関係が測れない場合）

	wage	educ	exper	
C さん	1600 円	16 年	4 年	⇒ 教育年数が 1 年増えても
D さん	1600 円	12 年	8 年	時給は上がらない？

単回帰モデル②の関係　　無視される部分

D さんは参加年数が長いため，1600 円のうちいくらかは exper の効果である。
400 円が exper の効果ならば，教育の効果はやはり 1 年当たり 100 円である。

この場合，②より適切なのは次の重回帰モデルである。

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + v \quad \cdots \quad ③$$

(2) 除外変数の問題

③式が正しいにもかかわらず②式を推定すると,②式の誤差項 u は次のように表せる。

$$u = \beta_2 \exp er + v$$

この場合，②式の誤差項 u は説明変数 educ と次のような関係をもつ。

$$Cov(educ, u) = \beta_2 Cov(educ, \exp er) + Cov(educ, v)$$

⇒ educ と v が無相関であっても，educ と exper が相関関係をもてば，$Cov(educ, u) \neq 0$

⇒ 単回帰モデルの仮定(S2)が成り立たない（**除外変数**の問題）

（統計学の復習）

確率変数 Y と Z の線形関数 $V = aY + bZ$ （a, b：定数）に関係する分散と共分散

$$V - E(V) = a\{Y - E(Y)\} + b\{Z - E(Z)\}$$

$$Var(V) = E[\{V - E(V)\}^2] = a^2 Var(Y) + b^2 Var(Z) + 2abCov(Y,Z)$$

$$Cov(X,V) = E[\{X - E(X)\}\{V - E(V)\}] = aCov(X,Y) + bCov(X,Z)$$

(3) 関数形の問題

年齢 age が時給 wage に及ぼす影響を知りたいとき，線形の単回帰モデルを考えると，

$$wage = \beta_0 + \beta_1 age + u \quad \cdots \quad ④$$

つまり，wage は age とともに一定割合で増えると想定する。しかし，実際には，
wage は age とともに逓減的に増えることが多い。つまり，$\beta_2 < 0$ として

$$wage = \beta_0 + \beta_1 age + \beta_2 (age)^2 + v \quad \cdots \quad ⑤$$

⑤式が正しいときに④式を推定すれば，除外変数と同様の問題が生じる。