医療職のための統計シリーズ

医療職のための学び直し－研究デザインから論文報告までの生物統計学の道標－ 第14回　回帰モデリング

篠崎（シノザキ）　智大（トモヒロ）＊

Ⅰ　はじめに

前３回の連載で，回帰モデル一般の考え方（第11回）と，２値変数（第12回）または生存時間変数（第13回）が結果変数であるようなイベント発症に対する回帰モデルの解説を行ってきた。ここまでは回帰（regression）とそのモデル（regression models）そのものの理解を目的とし，研究論文を読む場面で，あるいは自身で解析を行う前段階での知識整理を主に意図していた。今回は，自らが報告のために統計解析を行う場面で，これから推定すべき回帰モデルを決めるためのモデリング（modeling：モデル化やモデル特定model specificationともいう）の考え方を紹介する[1]。

Ⅱ　回帰モデルでの説明変数

回帰とは「サブグループ平均値」であり，回帰における説明変数とはそのサブグループを決めるための変数（例えば年齢）あるいは変数の組み合わせ（例えば年齢と性別）であった。ここで考えるのは，回帰モデルにおける説明変数の使い方である。

（１）　回帰モデルの形状

図１は連載第11回と第12回で示したデータである。さらにこの図には相異なる回帰モデルが重ね描きしてある。図１（A）には，年齢５歳ごとに得られた20人ずつの収縮期血圧データと，それらの平均値がプロットされている。直線はこのデータに回帰モデル

$E[\text{収縮期血圧} \mid \text{年齢}=x\text{歳}]=a+bx$

を当てはめたものである（推定値は$a=24.06$，$b=1.56$となる）。同じ図の曲線は同じデータに

$E[\text{収縮期血圧} \mid \text{年齢}=x\text{歳}]=a+bx+cx^2$

という回帰モデルを当てはめた結果である（$a=139.11$，$b=-2.56$，$c=0.034$）。

図１（B）には，年齢５歳ごとに20人ずつから記録した動脈硬化発症データ（ありなら１，なしなら０）と，発症ありの年齢別割合データ

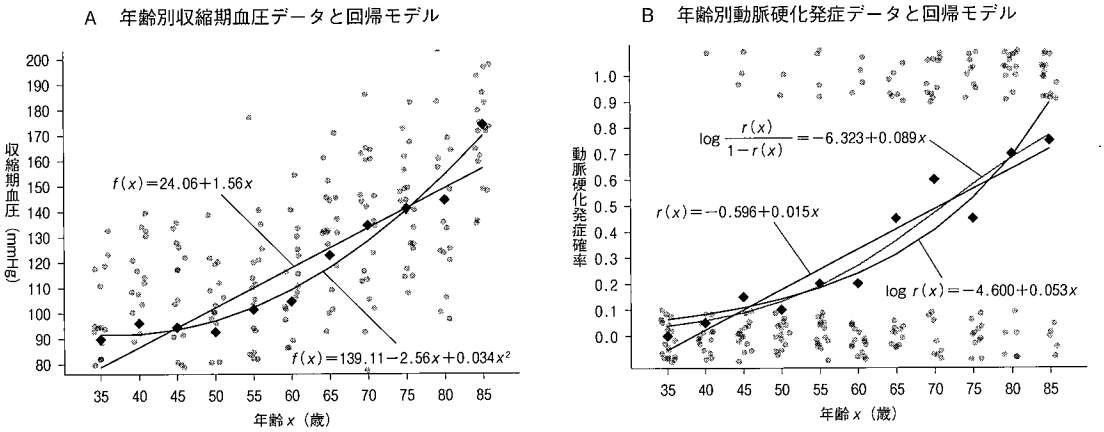

図１　年齢５歳ごとに測定されたデータと当てはめた回帰モデル

＊東京理科大学工学部情報工学科講師