

❖ 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6回 二つの変数の関係と要約

カワハラ タクヤ サカマキ ケンタロウ
川原 拓也*1 坂巻 顕太郎*2

I はじめに

医学系研究では、各対象者の年齢と収縮期血圧のような、二つの特性（変数）の関係に着目することがある。介入効果の評価が目的の研究では、治療などの介入変数と治療などの結果変数（アウトカム）の二つの変数の関係に興味がある。ただし、医学系研究で検討される変数の関係は様々であり、例えば、喫煙の有無とがんの発生（曝露とアウトカムの関係）、QOL（Quality of life）質問票の構成（質問項目間の関係）、介入前と介入後24週時点の重症度スコア（介入前後の変化）、心疾患と血圧値（真に測定したいアウトカム（true outcome）とそれを代替するアウトカム（surrogate outcome）の関係）、などがある。今回は、介入（もしくは曝露）とアウトカムの二変数の関係に主に着目する。正確には「介入」と「曝露」は異なるものだが、可読性を考慮し、今回は「介入」を用いて説明する（一部を除く）。

二変数の関係を考える際、相関（correlation）と因果（causation）の違いに注意が必要である。（統計学における）相関とは、二つの変数の間に直線関係に近い傾向が見られるという意味¹⁾で、例えば、一方の変数の値が大きいともう一方の変数の値も大きいという関係である。年収が高いと収縮期血圧も高いという関係は「相関」である。変数の相関は、一方の変数がもう一方の変数に影響するという因果関係を必ずしも意味しない¹⁾。先述の例では、年収の上昇が収縮期血圧の上昇に影響する、またはその反対の関係を意味するものではなく、二変数の相関は因果関係ではない。実際には、年収と収縮期血圧の両方に影響する年齢を介して二変数の相関が生じているという見かけ上の相関（擬似相

関)¹⁾であることにも注意が必要である。介入効果の評価が目的の研究においても同様な問題（介入とアウトカムの因果関係が歪められる問題）が生じる可能性がある。その問題を交絡（バイアス）という。本解説では、交絡等により二変数の関係が歪められていない状況を想定する。

以下、介入効果を評価する研究における介入とアウトカムの関係（関連）の検討方法を概説する。連載第5回²⁾でも述べたように、変数の型によって変数の要約方法が異なるため、II節で離散変数（離散アウトカム）、III節で連続変数（連続アウトカム）の説明をする。介入が離散変数（アスピリン服用・非服用など）の場合、群ごと（服用グループ・非服用グループなど）でアウトカムの要約（一変数の要約）²⁾を行い、これらの群間比較により介入とアウトカムの関連を検討する。介入（説明変数）とアウトカムがともに連続変数の場合はIV節で説明する。

II 離散アウトカムの要約と群間比較

(1) 分割表（クロス集計表）による要約

群間での離散アウトカムの要約には分割表（contingency table）を用いることができる。分割表はクロス集計表（cross-tabulation table）ともよばれる。分割表に記載する内容（頻度と割合）²⁾は研究デザイン（データの集め方）と関連するため、以下では、研究デザインごとに説明する。各研究デザインについては以前の連載³⁾⁴⁾を参照してほしい。分割表において、介入を行と列のどちらに置くか、介入・非介入や改善の有無などの順序をどうするかは任意である（介入を行におく、順序はアルファベット順などがある）が、本解説で紹介するような形式を目にすることが多い。表に記載する合計や

* 1 東京大学医学部附属病院臨床研究推進センター助教

* 2 横浜市立大学データサイエンス推進センター特任准教授