

一般線形モデルから一般化線形モデル（GLM）への拡張【第122回生物統計学】

1 概要

データ解析において統計学的な検定を行いますが、一般的な統計モデルは正規分布を仮定したもので、t検定や分散分析などは一般線形モデルと呼ばれています。一般線形モデルでは、目的変数 y は正規分布に従い、説明変数 x によらず、 Y の分布の標準偏差が一定であることを仮定したのになります。しかし、臨床試験において正規分布に従わないデータも存在し、この場合一般線形モデルでは対応が困難となってしまいます。正規分布に従わないデータを組み込むための統計モデルに一般化線形モデル（Generalized Linear Model: GLM）というものがあります。今回は線形モデルから GLM への拡張について紹介します。

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + e$$

Y : 目的変数 x : 説明変数 β : 切片 β : 傾き ε : 誤差

2 一般化線形モデル

一般化線形モデルは、正規分布に従わないデータに対して採用されるモデルで、誤差構造（error structure）、リンク関数（link function）、および線形予測子（linear predictor）の3つの要素によって構成されています。正規分布以外の確率分布を扱うことができ、GLM は一般線形モデルの拡張したモデルとなっています。

2.1 誤差構造、リンク関数と線形予測子

誤差構造は任意の指数分布族型（ポアソン分布や二項分布など）の確率分布を仮定してデータを変換せずに扱うため、直接パラメータを推定すること可能となります。また、従属変数が正規分布に従わない $y_i = \beta_1 + \beta_2 x_i$ の式をモデル化すると、正確さが失われてしまいます。そこで、従属変数を対数変換してから、モデル化することで、モデルの正確さが向上することが知られています。すなわち、 $\log(y_i) = \beta_0 + \beta_1 X_1$ となり、この左辺をリンク関数と呼びます。また、右辺を線形予測子と呼びます。

2.2 ポアソン回帰モデルの算出

カウントデータなどに有効なポアソン回帰モデルは GLM の1つで、誤差構造をポアソン分布に従うと仮定した場合、期待値と分散は以下のようになります。

$$P(y_i | \lambda_i) = \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!}$$

$$E[X] = V[X] = \lambda$$

ポアソン分布に従うデータを回帰分析する場合、 $y_i = \log(\beta_0 + \beta_1 X_1)$ のような式を組み、観測された y_i のデータは平均 λ_i に従います。ここで、先ほどの従属変数を対数変換すると、 $\log(y_i) = \beta_0 + \beta_1 X_1$ となります。こ

の式に任意の数値を代入することで確率を求めることができます。

3 まとめ

ヒト臨床試験においても、正規分布に従わないデータが存在しており、その場合には GLM を用いて回帰分析を実施する際に、高い精度で予測することが可能となります。GLM の応用としてロジスティック回帰や一般化線形混合モデルなどがあり、より現実的な統計モデルとなり介入効果を高めることができます。統計モデルを正しく理解し活用することでより良い結果を得ることが出来るようになります。

4 参考文献

- 久保 拓弥, データ解析のための統計モデリング入門 一般線形化モデル・階層ベイズモデル・MCMC, 岩波書店 2012

ヒト臨床試験（ヒト試験）で得られる結果は、様々な誤差を含んでいます。この誤差を小さくすることで介入効果を増大させることができます。オルトメディコは、多分野の専門家を有するため、様々なアプローチにより誤差を最小化する試験運営が可能です。引き続き、皆様にご満足いただけるような高品質なヒト試験を提供させていただきますので、今後ともどうぞ宜しくお願い申し上げます。