

## 回帰分析におけるモデルの適合度の評価【第40回生物統計学】

### 1 はじめに

回帰分析とは、独立変数を原因となる変数、従属変数（アウトカム）を結果となる変数ととらえて、変数どうしの因果関係( $Y=aX+b$ の関係)を分析するために開発された統計学的手法です。回帰分析は、従属変数が連続変数か 2 値変数 かによって手法が異なり、前者は線形回帰分析、後者は 2 項ロジスティック回帰分析が用いられます。これらの解析からは、2 つの主な情報が得られます。1 つ目は冒頭でも述べたように、個々の独立変数と従属変数の関係に関する情報です。2 つ目は、解析で得られたモデルがどれほどデータと適合しているかです。今回は、統計ソフトで線形回帰分析、2 項ロジスティック回帰分析を行った場合に得られるモデルの適合度について解説します。

### 2 モデルの適合度の評価

モデルの適合度とは、モデルつまり回帰式が、従属変数をどれほどうまく説明できているかを表したものです。つまり、適合度によってモデルの質を評価することができます。以下、個別に説明します。

#### ● 線形回帰分析

線形回帰分析の適合性を評価する方法には、F 検定と  $R^2$ (決定係数、寄与率)があります。

回帰分析における F 検定とは、対象者全員の従属変数とその平均値であると仮定した場合よりも、独立変数がどれほど従属変数をよりよく予測できるかを検定するものです。独立変数による予測が、偶然によって期待される場合よりもいい場合には、F 値は大きくなります。F 値が大きいほど P 値は小さくなり、独立変数と従属変数の間に関連がないという帰無仮説は棄却されます。

F 検定の主な限界は、それが「モデル全体」の適合度を評価するもので、どの独立変数が有意であるかなどを知ることはできないことです。例えば、モデルが 3 つの独立変数を含んでいて、そのうち 2 つが従属変数と有意の関連がなくとも、モデル全体では関連が有意になるということがあります。もう 1 つの限界は、モデルに投入された独立変数群が全体として従属変数と統計学的に有意に関連していたとしても、個々の独立変数がどれほど従属変数を量的に説明しているかを知ることができないことです。では、こうした限界を持つ F 検定がどのように役立つのでしょうか。それは、F 検定が有意でないときには、モデル自体に問題があることがわかるということです。個々には、従属変数と有意に関連している独立変数があるのに、モデル全体として有意でない場合には、従属変数と有意な関連がない独立変数を削除することにより解決することがあります。

$R^2$  は、独立変数がどれほどよく従属変数を説明するかを定量的に示してくれるという意味で、一般的には F 検定よりも有用です。 $R^2$  は 0 から 1 の間の値をとり、1 に近づくほど、そのモデルで従属変数の変動を説明できることを意味します。 $R^2$  は F 値よりも多くの情報を提供してくれますが、独立変数の数が増えれば、

それらがあまり意味のない変数であっても、 $R^2$  値が大きくなってしまいう問題があります。この欠点を補うために、調整  $R^2$  という統計量が開発されています。調整  $R^2$  では、変数が増えても、それが従属変数の説明に役立たない場合には、調整  $R^2$  が有利にならないように計算が工夫されています。

### ● 2 項ロジスティック回帰分析

2 項ロジスティック回帰分析の適合度の検定には、尤度比検定が用いられます。これは F 検定に似たものですが、分布はカイニ乗分布をとります。独立変数による予測が、偶然によって期待される場合よりもいい場合には、カイニ乗値は大きくなります。尤度比検定のカイニ乗値が大きいほど、P 値は小さくなり、独立変数は従属変数と関連があるということになります。カイニ乗分布による P 値を用いる場合には、サンプルサイズが大きいことが前提となっており、サンプルサイズが 100~80 の場合によく近似されることが知られています。

この検定よりも、優れた適合度検定として、ホスマー-レムショー検定があります。この検定では、従属変数の推定確立に基づいて、対象者をサンプルサイズが等しい 10 のグループに分割し、それぞれのサブグループにおける従属変数の期待値と観測値を比較します。適合度の高いモデルは、それぞれのサブグループにおいて、従属変数の尤度が観測された尤度に近くなるため、カイニ乗値が小さくなり、P 値は有意でない値を取ります。

### 3 まとめ

今回は、回帰分析における適合度についてまとめました。回帰分析は、原因となる変数から従属変数を予測することが可能となるので非常に便利ですが、モデルが適切ではないと何も意味がありません。解析結果をよく検討し、適切な回帰モデルを選択するようにしましょう。

#### 参考文献

医学研究のための多変量解析 第 2 版