

自由度調整済み分散説明率

通常分散説明率は以下のように表現できる。

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = 1 - \frac{\frac{1}{n}SS_{residual}}{\frac{1}{n}SS_{total}} = \frac{\frac{1}{n}SS_{total} - \frac{1}{n}SS_{residual}}{\frac{1}{n}SS_{total}}$$

残差と従属変数の(標本における)分散比を1から引いたものに等しく、モデルによって従属変数の分散がどの程度減少したかを、元の従属変数の分散に対する割合で表した誤差減少率(Proportional Reduction in Error; PRE)である。

しかしこれらの分散は標本における統計量であり、母集団に於けるそれらの不偏推定量ではない。つまり母集団に於ける誤差減少率の推定量にはならない。

母集団に於ける誤差減少率の推定量とするには、従属変数の分散を不偏分散に、残差分散を残差の平均平方に書き換えなければならない。そうして得られた統計量が「自由度調整済み分散説明率」である。

$$adjusted\ R^2 = 1 - \frac{\frac{1}{n-p-1}SS_{residual}}{\frac{1}{n-1}SS_{total}}$$

自由度調整済み分散説明率と無調整の分散説明率の関係は、

$$\frac{SS_{residual}}{SS_{total}} = \frac{n-p-1}{n-1}(1 - adjusted\ R^2) = 1 - R^2$$

より、

$$adjusted\ R^2 = 1 - \frac{n-1}{n-p-1}(1 - R^2)$$

である。

自由度調整済み分散説明率は、定義上、無調整の分散説明率よりも値が小さくなる(自分で考えてみよう)。また、無調整の分散説明率は決して負の値を取らないが、自由度調整済み分散説明率は、よほど当てはまりが悪くて独立変数の数が多い重回帰分析では、負の値を取る事もある。例えば、 $n=200$, $p=10$ の時、無調整の決定係数が.05 以下だと自由度調整済み分散説明率はマイナスになる。