

# Kappa係数の意外な落とし穴とその対処法の紹介

○江口 凌平、菊池 宏和

(シミック株式会社)

## 背景：

カテゴリカルデータの一致度を評価する方法の一つにカッパ係数( $\kappa$ 係数)がある。例えば、二人の観察者間の評価といった質的データ(カテゴリカルデータ)の一致度を、偶然による一致度を考慮した上で算出することが可能である。しかし、一致セル・不一致セルの割合に偏りが存在する場合には、一致率が同じでも $\kappa$ 係数が異なる場合があり、 $\kappa$ 係数の解釈が困難になる可能性がある。

## 検証：

SASのシミュレーションにより、 $2 \times 2$ クロス表データの値の偏りを変動させた場合の $\kappa$ 係数の動作特性を報告する。また、偶然による一致度を調整し、値の偏りに頑健な指標であるPABAK、AC1統計量を紹介し、 $\kappa$ 係数同様に $2 \times 2$ クロス表データの値の偏りを変動させた場合の動作特性を報告する。

## 結果・まとめ：

$\kappa$ 係数のシミュレーションの結果、一致セル・不一致セルの割合の不均衡が大きい場合、 $\kappa$ 係数はその影響を受け、値が変動しやすい傾向にあることが分かった。したがって、 $\kappa$ 係数の値を解釈する際には、一致セル・不一致セルの割合の偏り具合を確認する必要がある。また、PABAK、AC1統計量のシミュレーションの結果、一致セル・不一致セルの割合の不均衡が大きい場合でも、 $\kappa$ 係数の場合と比較して値が変動しにくい傾向にあることが分かった。したがって、 $\kappa$ 係数だけではなく、より偏りに頑健な統計指標であるPABAK、AC1統計量も併せて評価する必要があると考えられる。