

生存時間解析における3次スプラインを用いた事例の紹介

○室永 遼太郎、浦狩 保則

(シミック株式会社)

自明の記法を用いて、比例ハザードモデルを対数で表現すると

$$\log(h(t)) = \log(h_0(t)) + \beta X + \alpha Z$$

である。 X と Z はともに共変量であるが、本発表では X について議論する。

比例ハザードモデルの解析では、年齢や白血球数などの連続量の説明変数(X とする)は通常、それらの線形成分がモデルに含まれる。しかし、このような説明変数の効果を線形成分のみで解釈することが好ましくないことがある。この問題への一つの対処法として、 X の非線形関数 $g(X)$ を用いて以下のようにモデル化し解析する。

$$\log(h(t)) = \log(h_0(t)) + \beta g(X) + \alpha Z$$

また、比例ハザードモデルの解析における重要な仮定は、ハザードの比例性である。これはハザード関数におけるある共変量の効果が時間上で一定であることを意味する。しかし実際には、共変量の効果が時間上で変動する場合がある。このように比例ハザード性を満たさない共変量(X とする)が見出された場合、以下のよう X の回帰係数を時間の関数とするアプローチが主流となっている。

$$\log(h(t)) = \log(h_0(t)) + f(t)X + \alpha Z$$

$f(t)$: 時間の関数で表した回帰係数

最も単純な方法は、 $f(t)$ を時間の線形関数とすることであるが、そのようなモデル当てはめの柔軟性を高めるために $f(t)$ として時間の非線形関数を用いる方法が推奨されている。

本発表では、上記の $g(X)$ と $f(t)$ のそれぞれに3次スプラインを用いた場合を取り上げ、実際のデータを用いてSASでの実装方法を紹介する。