

統計演習 (15)

生存分析

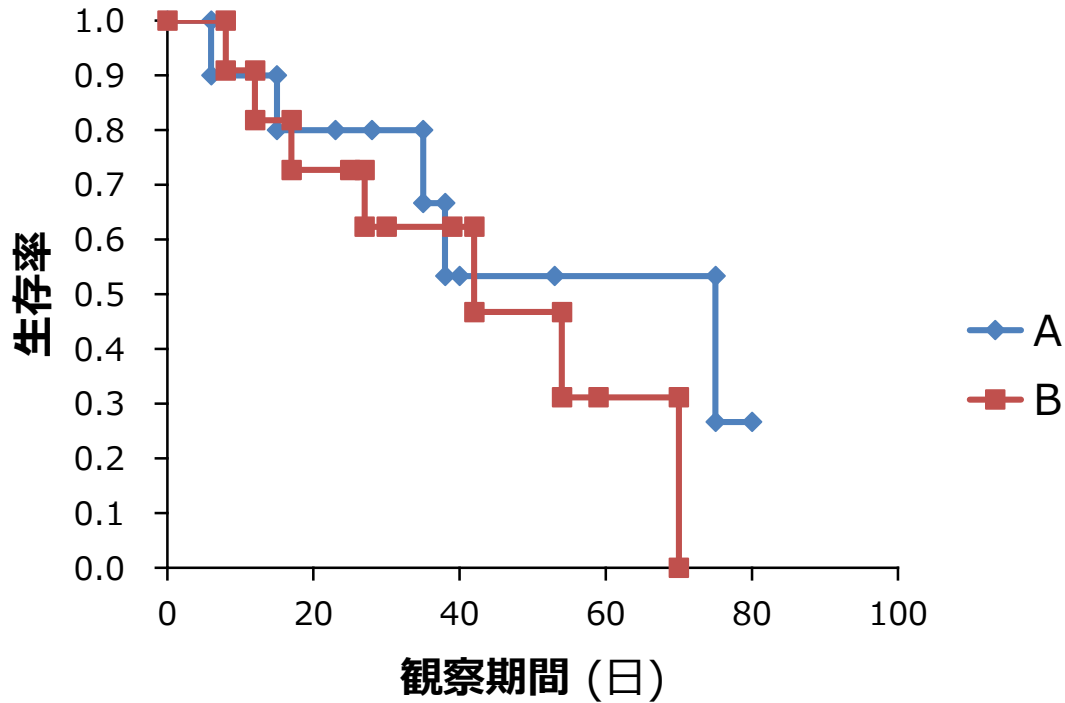
高谷 智英

Jan 22, 2020





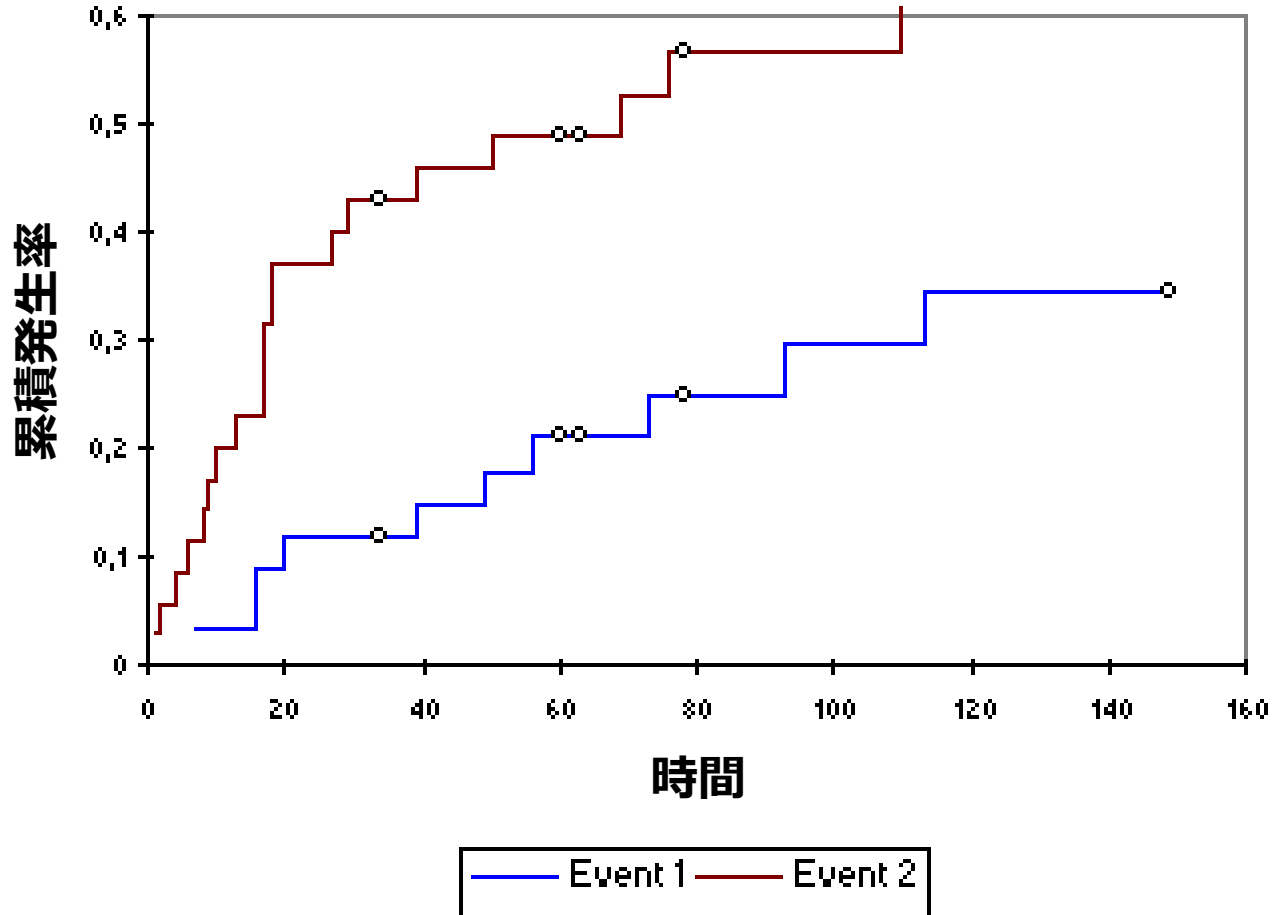
生存曲線



生存曲線とは、ある明確に定義された**イベント (出来事, エンドポイント)** の発生を時間に対してプロットした曲線である。イベントは必ずしも死亡である必要はなく、一度だけ生じるものであれば良い。



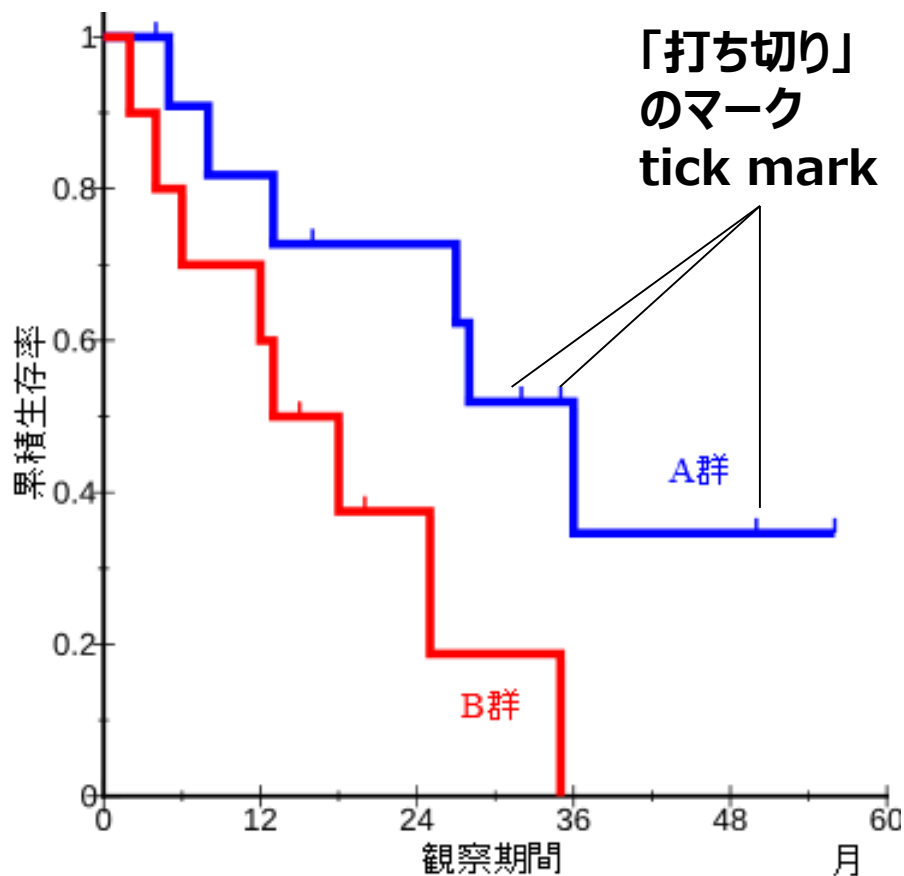
イベントの累積発生率を描いた生存曲線は右肩上がりになる



イベントの累積発生率を時間に対してプロットした生存曲線は右肩上がりになる。



生存分析は打ち切り (censor) を扱うことができる



例) 生死を扱う場合、解析時点で生存中の患者や、ある時期まで生存したことは知られているが、その後の情報がない患者などは「**打ち切り**」として扱う。



生存曲線の作成

生命保険数理法：

一定の期間（月や年）ごとの生存率を計算する。
サンプル数が大きいても比較的計算が容易である。

Kaplan-Meier 法：

イベントが発生するごとに生存率が再計算される。
サンプル数が大きくなると計算が複雑に...
→ コンピュータにより解決



Kaplan-Meier 法による生存率の計算

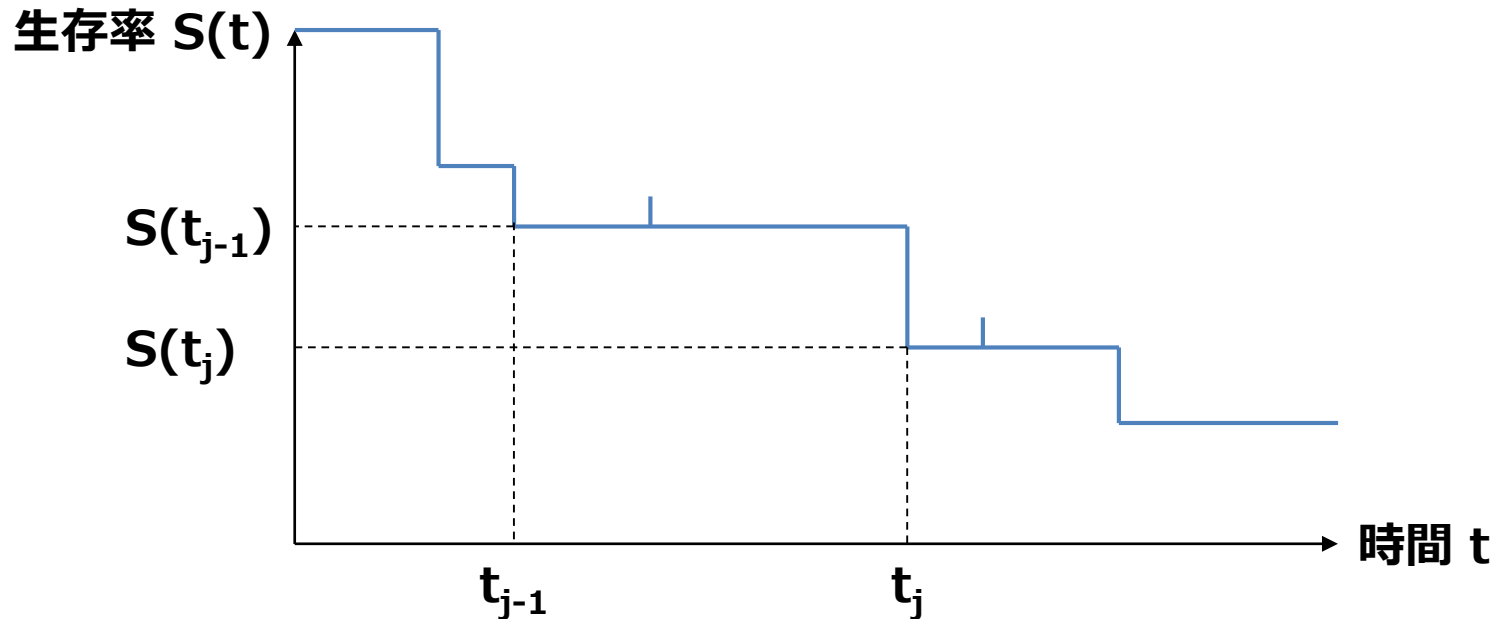
時間 t における生存率を $S(t)$ とすると、

$$S(t_j) = S(t_{j-1}) \times (1 - q_j)$$

q_j (ハザード) : t_j で死亡する確率

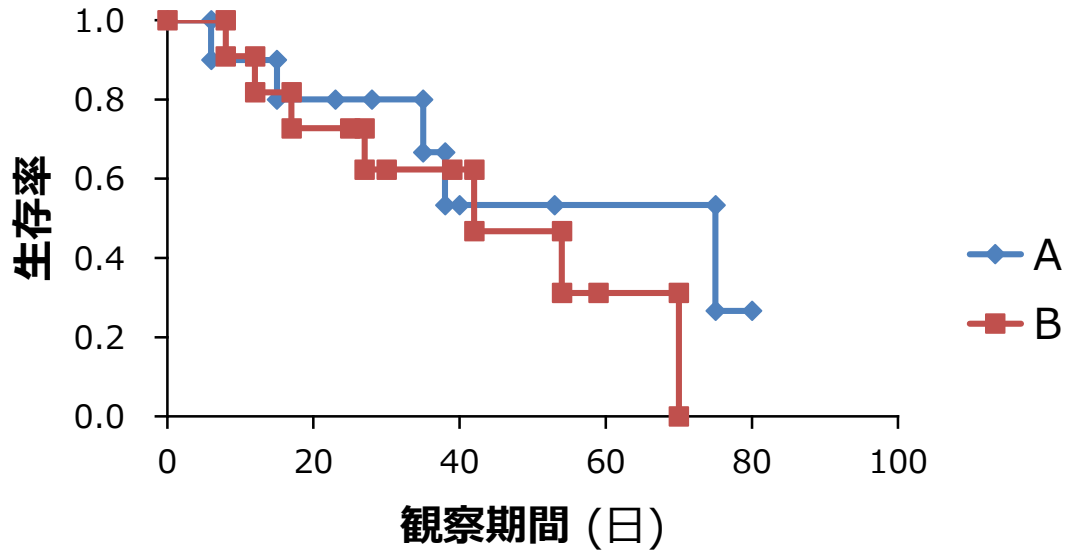
t_j の時点で死亡した数を、 t_j まで生存した数で割った値。

イベント発生前に打ち切りとなったサンプルを計算から除外する。





ログランク検定



Kaplan-Meier 法により作成した複数の生存曲線の差を検定する。
イベント発生時点における分割表の χ^2 検定を利用する。

	A群	B群
イベント発生あり	12%	27%
イベント発生なし	88%	73%