

生存時間データに対する Kaplan-Meier 図と比例ハザードモデルの基本。

がん治療における患者の生存率の推移など、一般に「時間-イベント型データ（ここでは生存時間データとよぶ）」について回帰分析などのモデル解析を行う場合には『(Cox の) 比例ハザードモデル解析』が用いられる。生存時間データは「ある時点」での「あるできごと（イベント）の有無」を集計するので、その図式化や検定方法なども他の連続的な数値のデータ解析手法とは異なり、推移図は『Kaplan-Meier プロット』と呼ばれる手法で描画し、また群間比較検定には『ログ・ランク検定』という方法がよく用いられる。回帰分析の際に用いるハザードモデルとは「ハザード（危険）」の度合いが各種の影響因子によりどのように影響を受けるかを評価するモデルをいう。

回帰式を模式的な数式で表すと次のようになる。 β はそれぞれの因子の影響の程度を表す係数であり、データにこのモデルをあてはめることで算出する。

$$\text{対数ハザード比} = \text{切片} + \beta_1 \cdot \text{影響因子 1} + \beta_2 \cdot \text{影響因子 2} + \dots + \text{誤差項}$$

ここでいう対数ハザード比とは、基準ハザードとよばれるすべての因子の影響がないとした場合のハザードと、どれか一つ以上の要因の影響がある場合のハザードとの比、と定義される。この比が常に一定で影響要因とハザード比 (Hazard Ratio, HR) とが比例すると仮定したものが比例ハザードモデルで、多くの生存時間データの解析に用いられる。実際にはこの比例性の仮定が成り立たない場合もあり確認が必要となるが、ここでは仮定が正しいとして話を進める。

生存時間をグラフ化した生存曲線ではある時間でのイベント発生の有無を考慮するので、横軸に時間、縦軸にイベント未発生率（例えばイベントを死亡と考えると縦軸は生存率）をとると階段状の右下がりのグラフとなる。簡単な例として次の 5 例の被験者の生存曲線を描いてみると図（左）のようになる。

例：5 名の被験者の生存時間が、それぞれ 2、3、5、8、10 か月

この場合は、イベントが起きた時間ごとにその人数割合を減算したら図が描ける。しかし実際のデータでは「打ち切り(censored)」とよばれるイベントが起きた時間を特定できない場合もある。例えば、試験の途中でその試験から辞退した被験者や試験終了時までイベントが起これなかった被験者が相当する。特に試験途中で欠損となった場合をドロップアウト (dropout) 例ということもある。次の 5 例のデータを考えてみる。例：5 名の被験者の生存時間が、それぞれ 2、3、5*、8、10 か月（ただし*印は打ち切り例として扱う）とする。具体的な計算式を示す。結果は図（右）のようになる。

0 年目：5 例ともに生存なので生存率は 1.0

（1 年目：イベントは起こっていないので生存率はそのまま 1.0）

2 年目：1 例でイベントが起こったので、生存率は $1.0 \times (1 - 1/5) = 0.8$

3 年目：1 例でイベントが起こったので、生存率は $0.8 \times (1 - 1/4) = 0.6$

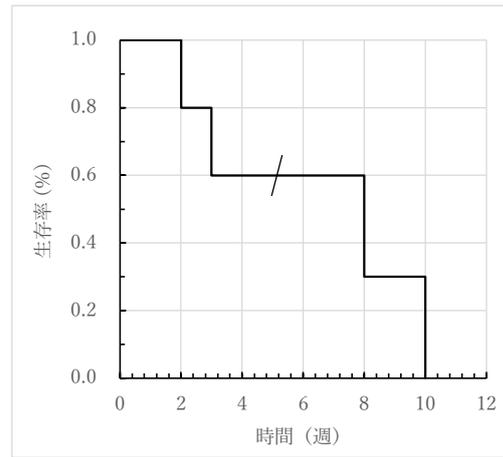
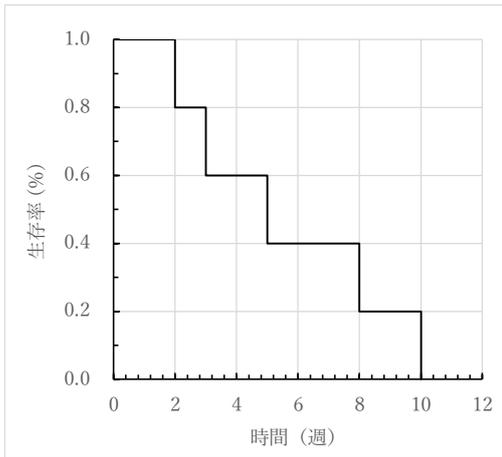
5 年目：1 例が欠損となったので生存率は変化しないがこのデータは今後除外される。

8 年目：1 例でイベントが起こったので、生存率は $0.6 \times (1 - 1/2) = 0.3$

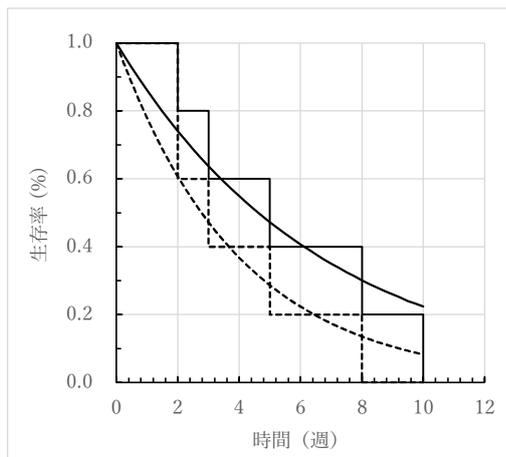
（ここで末尾の項の分母が 3 ではなく 5 年目の被験者を除いた生存人数 2 を用いている）

10 年目：1 例でイベントが起こったので、生存率は $0.3 \times (1 - 1/1) = 0.0$

欠損があることをわかりやすくするためにその時刻に斜線 (/) などを記すことがある。



比例ハザードモデル解析では生存時間データを説明するモデルを決める。そのイメージ図を示した(下図曲線)。図は実線(例えば抗がん剤 A 使用群)と点線(例えば抗がん剤 B 使用群)の 2 群でデータが得られたときの Kaplan-Meier 図とハザードモデルによる曲線(イメージ)を示す。この例では A 群よりも B 群のほうが生存率の減少が早く B 群のほうが抗がん剤としての効果が低い(有意性を評価する場合はきちんと統計学的な解析を行う、このシミュレーションでは B に対する A のハザード比を 0.6 とした)。ハザード(危険な度合い)が小さいということは治療効果があったと理解できる。



先に触れたように比例ハザードの仮定は必ずしも成立しないので、厳密には比例ハザード性の検証など十分な統計的評価が必要である。普段の解析ではそこまで考えることが難しく、単にソフトウェアにデータを代入して得られた結果をそのまま使用してしまいがちであるが、まずは Kaplan-Meier 図によりデータのプロットをよく見てその特徴を知る習慣が必要である。

ハザードモデルの模式的な数式を用いた解析はソフトウェアでは「Cox の比例ハザード回帰」などのメニューを選択することで実施可能で、回帰分析の結果としていくつかの数値が得られるが、重要な値は係数 β および β の推定誤差であり、これらの結果から $HR = \exp(\beta)$ (ある影響要因の値が「1 単位」変化したときの HR) が得られる。また推定誤差を用いて HR の 95% 信頼区間が得られ、その範囲が 1 を含まない場合に HR は有意、すなわち有意な影響因子であると判断できる。同時に結果には有意性を判定した p 値が示される。下表にあるソフトウェアからの出力例を示した。回帰分析の p 値を確認し、推定誤差から信頼区間を求める等の手法は回帰分析に共通の手法である。

重要な値は係数 β および β の推定誤差であり、これらの結果から $HR = \exp(\beta)$ (ある影響要因の値が「1 単位」変化したときの HR) が得られる。また推定誤差を用いて HR の 95% 信頼区間が得られ、その範囲が 1 を含まない場合に HR は有意、すなわち有意な影響因子であると判断できる。同時に結果には有意性を判定した p 値が示される。下表にあるソフトウェアからの出力例を示した。回帰分析の p 値を確認し、推定誤差から信頼区間を求める等の手法は回帰分析に共通の手法である。

独立変数	係数 (推定値)	標準誤差	P 値	ハザード比	ハザード比の 95% 信頼区間	
				(推定値の指数)	下限	上限
変数 A	-1.15	0.456	0.0XXX	0.317	0.13	0.774
変数 B	0.495	0.195	0.0XXX	1.64	1.119	2.404
変数 C	-0.0567	0.00751	0.0XXX	0.945	0.931	0.959