存活分析NOTE

Data : 01\_Length of Hospital Stay (LOS)

544名患者住院時間的資料，feature : 物治、語治、職治、功能獨立評定(FIM)(評估比較患者殘疾嚴重程度)、巴氏量表(BAR)(日常生活功能評估，測量病患的治療效果及退化情形(35下可申請外籍看護))、性別、高血壓、糖尿病

FIM



Q :

1. 三種治療在相同程度患者，是否有效減少住院時間
2. 變量和時間長短之關係
3. 預測

存活分析 :

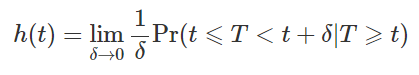
分析影響死亡率的變數(影響存活率的重要因子)，在LOS終究是分析影響住院時間長短的變數。

Survival function : S( t ) = P( T > t )

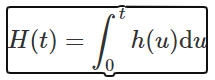
Cdf:



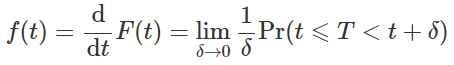
Hazard function :下一個瞬間事件發生的機率

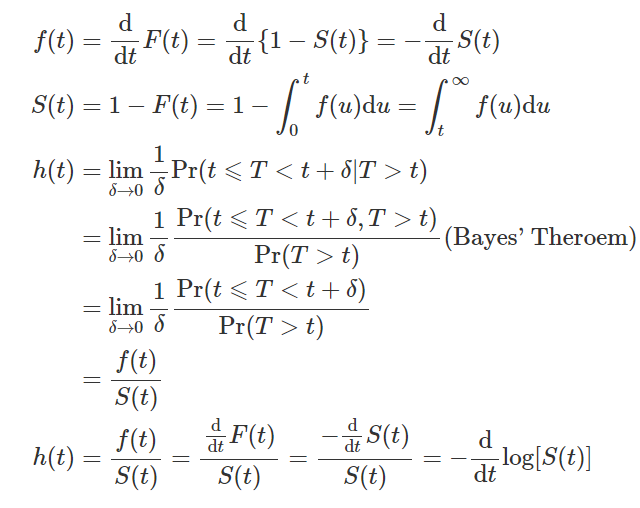


累積風險:

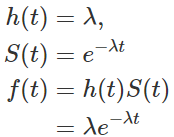


生存時間pdf:

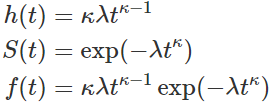


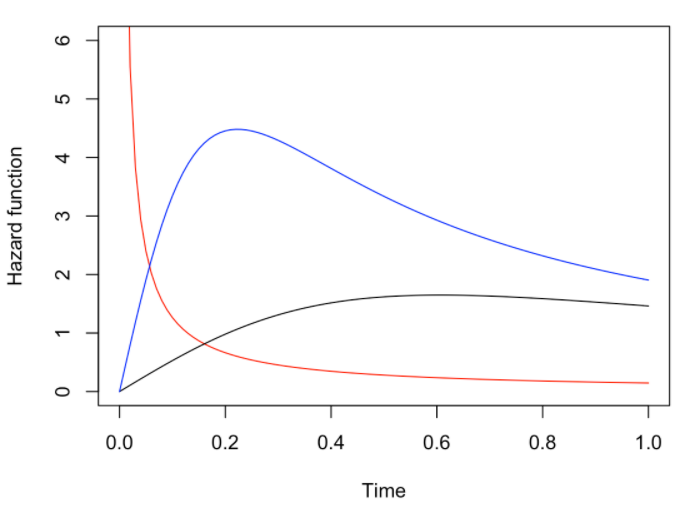


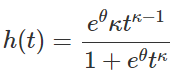
假設每個時間點發生事件的機率相同，也就是



假設服從weibull，則 (此假設較上面假設更靈活)

 k決定風險度曲線的形狀，k<1風險度隨時間下降，反之，上升。K = 1，則為指數分布，k=2，為線性關係。

假設服從Log-logistic，則



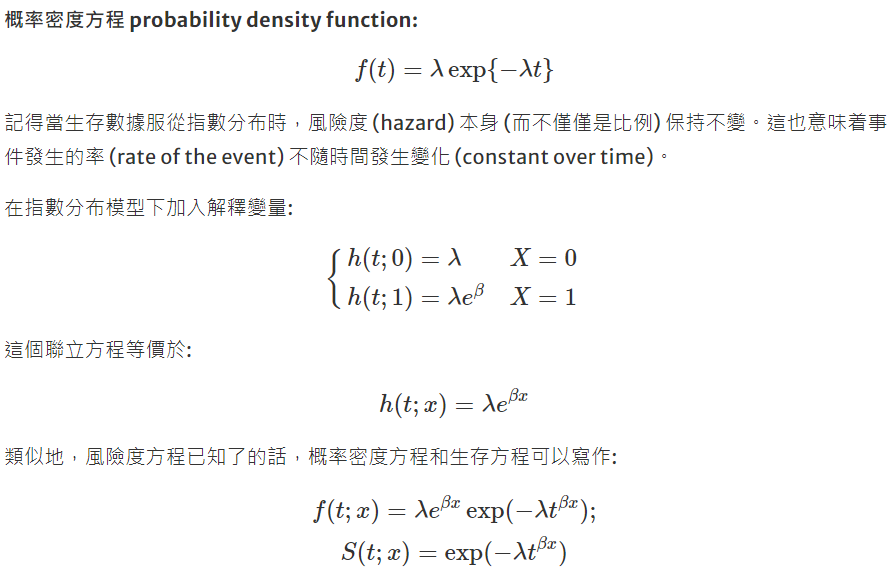
在weibull分布下，風險度指隨時間單調遞增或遞減。在此假設下，風險度可以隨時間有增有減。

回歸模型:

Likelihood function:

(這組資料無缺失值，所以L = ㄇf(ti) )

指數分布:



Weibull分布:

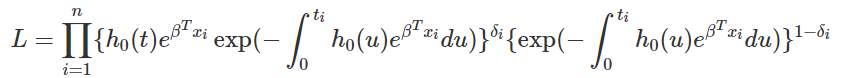
COX比例風險模型

假設解釋變量對風險度得作用是成比例，則風險度function和基線風險度(baseline hazard )的關係可表示為



h0(t)為被比較的基線組成員的風險度函數。Cox在1972年提出，可以忽略基線風險在模型中的存在，這個模型對預測變量對於風險的的效果用beta進行參數估計，因此又稱半參數化模型。

Likelihood:



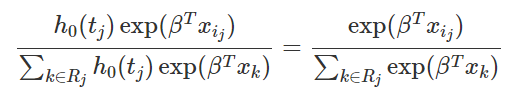
無法估計(?)

要使用partial likelihood

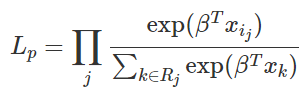
Ij > 實驗對象，tj > 生存時間，xij > 解釋變量，

對於潛在對象集合(在無線接近生存時間之前都沒有發生事件的觀測對象)

則ij在tj發生事件的條件機率為



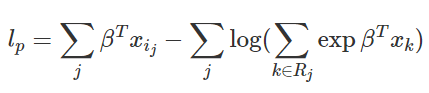
則partial likelihood Lp =



稱為partial likelihood 因為他只取用了生存過程中的一部份(意思是潛在對象?)

但他漸進一致(asymptotically the same)完整的likelihood。Cox利用這個性質忽略h0(t)，使用MLE去估計beta，且透過Fisher information matrix的反矩陣來得到Beta.hat的變異數。

Log-likelihood:

(估計Beta)

