

Assignment 1.

Answer the following questions with the reading attached.

- Q1. What is the data type of survival data?
- Q2. Find the example of survival data from the internet.
- Q3. What are the main characteristics of survival analysis?
- Q4. Before learning the survival analysis, what statistical tools you have learned could be applied to analyze the survival data?
- Q5. What do you feel about the survival analysis?

Survival Analysis 在近代應用醫學統計的重要角色

- 臨床醫學研究者，經常在臨床計劃治療 (Clinical Trials) 實驗中，欲證明甲治療方法比乙治療方法更能使病人的壽命延長，或者更能延緩病人的某些症狀之出現。
- 因此 time to event 成爲此類研究計劃中的 primary interested endpoint，換言之，在應用醫學統計分析中，time to event 是其統計 model 的主要應變數 (primary outcome variable)。
- 因爲 Survival Analysis 的資料有其獨特的性質，所以它常常被單獨地提出討論。以下將略爲說明其資料特性：
 - ◆ 一、Survival time data 的分佈，一般不會呈現常態分佈 (normal distribution)，在最常用的應用醫學統計方法中，如 t-test, ANOVA, Regression Analysis 等等，其最基本的假設條件就是資料呈現常態分佈，但對 survival time data 而言，通常這是不存在的。
 - ◆ 二、Survival time data 經常會有 censoring 的情況出現，甚麼是 censoring？簡單地說，當 event (如死亡) 沒有發生時，就是 censoring！筆者舉以上表的例子說明：
 - 病人 A 在 study opening date 當日就接受治療，而在 3 月 31 日死亡，在 Survival time data 中，病人 A 的資料是完整的；因爲此病人的資料顯示其 survival status 是死亡的(即 event 發生了)，而且在死亡前，被追蹤了 3 個月，就病人 A 的完整資料情況而言，並不屬於 censoring。
 - 病人 B 同樣地在 study opening date 當日就接受治療，而且一直存活至 study closing date (6/30/ 1998)，在 Survival Analysis 中，病人 B 的資料是不十分完整的，因爲 event (死亡) 沒有發生，僅知道病人 B “至少” 存活 6 個月！由於 study 已經結束了，所以病人 B 的資料不顯示“何時死亡發生”，也許病人 B 可以繼續存活 10 年，20 年或者病人 B 在 7 月 1 日就死亡。這種情況出現於 survival time data 時，就稱爲 censoring，或是稱爲 right-censoring，又或稱爲 type II censoring。在此僅就被應用於醫學統計的 type II censoring 提出討論。
 - 病人 C，其自 2 月 1 日起接受治療，但在 3 月 31 日做完門診複查後，便杳無音訊 (loss of follow-up)，病人 C 的資料在 Survival Analysis 中，亦屬不完整的！因爲僅知道病人 C “至少” 存活兩個月，但無從得知其 4 月 1 日至 6 月 30 日間之實際情形，此種情況也被稱爲 type II censoring！
 - 病人 D 的資料和病人 B 的資料相類似，均爲 type II censoring，唯一不同點在於病人 D 從 6 月 1 日才開始接受治療。因此，

其資料顯示為“至少”存活一個月。

- ◆ 三、臨床醫學論文中之 Survival Analysis 通常強調的是 survival time data 的中位數 (median) 而非平均數 (mean) ，
 - 理由：除非在 survival time data 中沒有 censoring 的情況發生；否則 The estimates of the mean will always be biased 。

敘述完 survival time data 之特性後，就幾種經常應用於醫學統計 Survival Analysis 的方法及常見的錯誤，做簡明扼要的介紹。

- 對臨床醫學研究者或病患而言，最 fundamental 的問題是：經過某種治療後，病人存活 XX 年的機會有多少？
 - 此問題的答案，則有仰賴於 estimation of the survival curve 求得。
 - 在 estimate survival curve 時，最常用的方法為 Kaplan-Meier estimator ，又稱為 Product Limit estimator ，大約超過 95% 的臨床醫學學術論文在 estimate survival curve 時會使用此一方法。
 - Kaplan-Meier estimator 其優點為簡單易懂，而且對 censoring 的 cases 做了某種程度的調整 (adjustment) ，其實此方法的數學理論十分簡易。
 - 但應注意的是，在某些情況下，Kaplan-Meier estimator 無法求得 median survival time ，例如：
 - ◆ study period 太短，
 - ◆ events (例如死亡) 太少，或者是
 - ◆ censoring 的 cases 太多。
- 如何比較 survival curves ? 從臨床醫學研究者的角度說，就是究竟甲治療方式的 survival time distribution 與乙治療方式的 survival time distribution 在統計上有無顯著的差異 (statistically significant difference) ? 至少有數百篇的統計博士論文在探討此一問題，可分類為 nonparametric method, parametric method 及 semi-parametric method 來討論。
 - nonparametric method 基本上不需要 the distribution of survival time data 的假設條件 (assumptions) ，最常見的此類方法為
 - Log-Rank Test ，其實 Log-Rank Test 就是 Cochran-Mantel-Haenszel Test 的應用，而其最基本及重要的假設條件必須是 the ratio of event rates 為一常數 (constant) with respect to time 。舉例而言，甲治療方式的死亡率是乙治療方式的死亡率之 1.5 倍，這個 1.5 倍的差異存在於治療後的任一時間，亦即治療後一個月，甲方式的死亡率是乙方式的 1.5 倍，而治療後三年，甲方式的死亡率仍是乙方式的 1.5 倍。而此一基本假設條件和 Cox Proportional Hazards Model (見下文) 的基本假設條件是完全相同的!!! 更一步說，Log-Rank Test 是

Cox Proportional Hazards Model 的特例 (special case)。

醫學統計 Survival Analysis 常見的錯誤

● 錯誤一：因為 the ratio of event rates 不為一 constant，所以在論文寫作時，捨棄 semi-parametric method 之 Cox Proportional Model，而選擇 nonparametric method 的 Log-Rank Test，此錯誤的發生在於不瞭解 Log-Rank Test 的 assumption。

- 另一種常在 nonparametric method 中見到的方法為 Generalized Wilcoxon Test，當應用 Generalized Wilcoxon Test 時，常會面臨選擇 Gehan's Generalized Wilcoxon Test，或者 Peto's Generalized Wilcoxon Test，又或者 other types of Generalized Wilcoxon Test，雖然有關這幾種 tests 的細節，不在本文討論範圍內，但在此提醒讀者，當使用統計 package 時，務必先瞭解該 package 採用何種 Generalized Wilcoxon Test。不論是何種 Generalized Wilcoxon Test，當沒有 censoring 出現的情況下，Generalized Wilcoxon Test 就是 Wilcoxon Rank Sum Test。

在臨床醫學論文中，究竟應該用 Generalized Wilcoxon Test 或者 Log-Rank Test 呢？兩者之差別在那裏？

其實這兩種 tests 的差異在於 “how to weigh the event” ！

- Log-Rank Test 對不論何時發生的 event，皆給予 “equal weight”，
- Generalized Wilcoxon Test 則給早期的 event 較重的 weight，而給晚期的 event 較輕的 weight。When the event rate is greater early in the study than toward the end, the Generalized Wilcoxon Test is the more appropriate test since it gives greater weight to the earlier differences。換言之，The differences in the early survival experience of two populations are more likely to be detected by the Generalized Wilcoxon Test than by the Log-Rank Test。
- 但因為有各種不同型態的 Generalized Wilcoxon Test 存在，所以在臨床醫學期刊中，若 Generalized Wilcoxon Test 及 Log-Rank Test 這兩種方式的結論相同時，絕大多數會選擇 Log-Rank Test，僅有當此兩種方式之結論不同時，並且在臨床

醫學上早期的差異有醫學上的意義時，才會選擇 Generalized Wilcoxon Test 的結果做論。

以 parametric method 處理 survival time data 時，首先其假設條件為 data 是哪一種分布 (distribution)? 在此常見到的分布是

- exponential distribution,
- log-normal distribution,
- Weibull distribution,
- gamma distribution 和
- log-logistic distribution 等。

確認 data 的 distribution 以後，最常選用 Likelihood Ratio Test 作統計分析，Likelihood Ratio Test 有時也稱為 Log Likelihood Ratio Test。一般而言，parametric method 的 power 較之於 nonparametric method 的 power 則會大些。所以若 assumptions 正確的話，Likelihood Ratio Test 會比 Log-Rank Test 或 Generalized Wilcoxon Test 容易 reject Null Hypothesis。在應用 Likelihood Ratio Test 時，應該注意其 data 必須為特殊的 distribution 外，同時其 event probabilities are constant over time。易言之，不論治療後一個月或者一年，其死亡率是固定不變的。

- 錯誤二：在醫學論文寫作中，選取了 Likelihood Ratio Test，但卻不知其 survival time data 為何種 distribution！

在 semi-parametric method 中最常見的方法是 Cox Proportional Hazards Model，以下簡稱為 Cox Model。Cox Model 在臨床醫學期刊中，被十分廣泛地使用，主要在於其 assumptions 不如 parametric method 的 assumptions 嚴格，而又有類似 Regression Analysis 的方便，也就是說 Cox Model 不需要假設 any particular distribution of survival time data，而且可以將多個 risk factors (例如病人的治療方式，吸煙與否，嚼檳榔與否，年紀等等) 同時考慮於此 model 中，來探討 the association between those risk factors and survival time data。最重要的是，Cox Model 可以求取 adjusted risk ratio for the treatment effect。換句話說，在 Cox Model 中，經調整過 (adjusted) 其它 risk factors 後，甲治療方法比乙治療方法減少 XX% 的死亡機會。雖然 Cox Model 是很好的統計方法，但請注意其假設條件和 Log-Rank Test 的假設條件是一樣的，亦就是 risk ratio (或者 the ratio of event rates) 為一常數，不會變動的。因此不論是治療後一個月，還是治療後一年，此一 ratio 乃 fixed。

綜合上述，介紹了一些常用的 Survival Analysis methods and models，每一個

方式都有其優缺點及假設條件，瞭解了這些 methods 的使用條件，才能求取正確的 p-value，做出正確的論文結論。

References:

- Armitage, P., and Colton, T. (1998). Encyclopedia of Biostatistics, John Wiley & Sons, New York.
- Cox, D.R., and Oakes, D. (1984). Analysis of Survival Data, Chapman & Hall, New York.
- Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples, *Biometrika*, 52, 203-223.
- Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *J Am Stat Assoc*, 53, 457-481.
- Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data, John Wiley & Sons, New York.
- Matthews, D., and Farewell, V.T. (1996). Using and Understanding Medical Statistics, Karger, New York.
- Peto, R., and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion), *J R Stat Soc A*, 135, 185-206.