

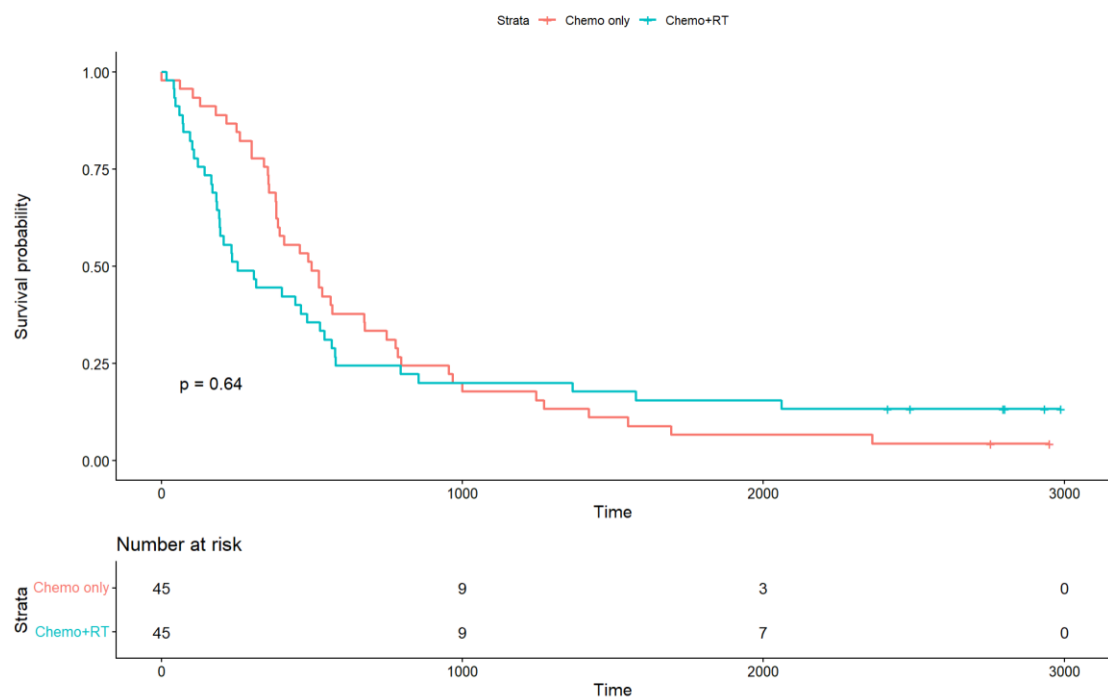
HW 4: Diagnose the validation of Cox PH model for the gastric data

Gastric cancer data are shown in “gastric cancer data.csv”, where Treatment= 0 if chemotherapy only; 1 if chemotherapy with radiotherapy.

Please answer the following questions and summarize and explain your results.

Computer outputs without any interpretations are not acceptable.

1. Please draw Kaplan-Meier curves stratified by treatments and interpret the result.



Log-rank test:

H_0 : The two survival curves are identical, i.e. $S_1(t) = S_2(t)$ for all t .

H_1 : The two survival curves are not identical, i.e. $S_1(t) \neq S_2(t)$ for some t .

Treatment=0 (化療) 與 Treatment=1 (化療+放療) 的生存曲線，在 1000 天左右的位置生存曲線交叉，整體上差異不大。根據 Log-rank 檢定，p 值= 0.64 (> 0.05)，兩組之間的生存差異不顯著。因此，我們沒有足夠的證據認為放射治療能顯著改變病人的生存時間。

2. Please fit the data with Cox PH model and interpret the result.

$$h(t \mid \text{Treatment}) = h_0(t) \exp(\beta \cdot \text{Treatment})$$

$$H_0 : \beta_{\text{Treatment}} = 0 \quad (HR = 1)$$

$$H_1 : \beta_{\text{Treatment}} \neq 0 \quad (HR \neq 1)$$

```
> summary(cox1)
Call:
coxph(formula = surv_obj ~ Treatment, data = df)

    n= 90, number of events= 82

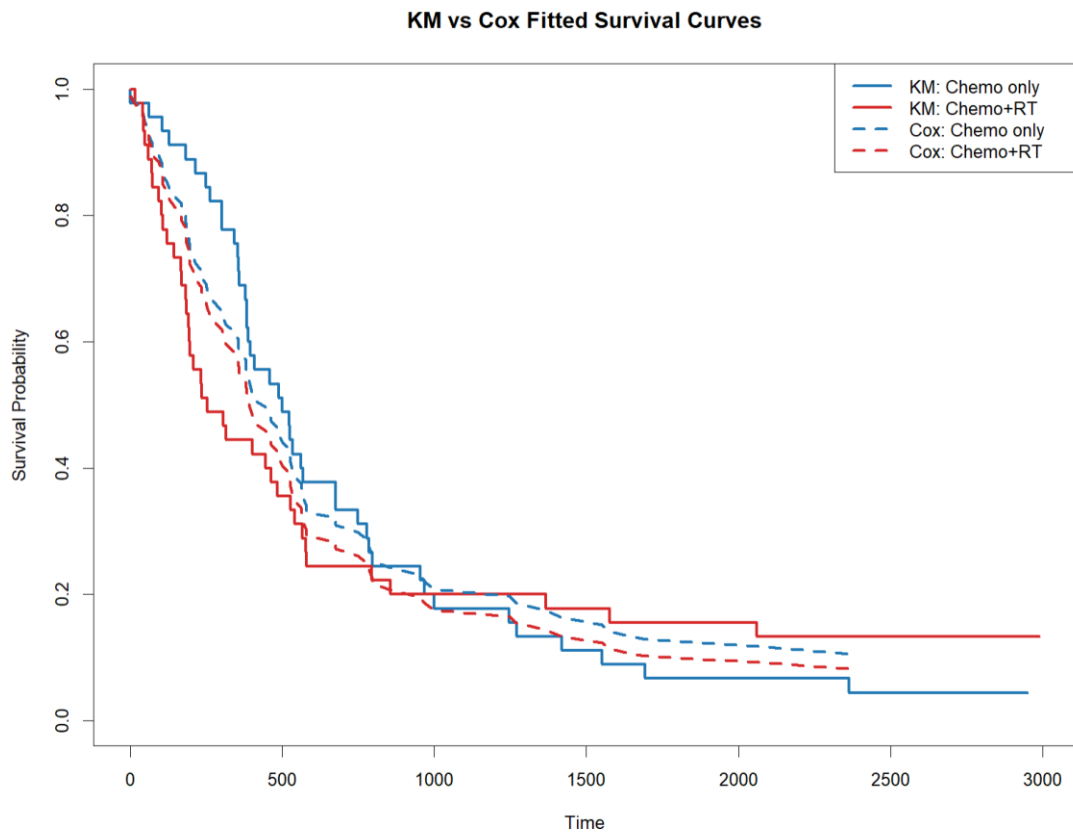
              coef exp(coef) se(coef)      z Pr(>|z|)
Treatment 0.1051    1.1109  0.2233 0.471  0.638

              exp(coef) exp(-coef) lower .95 upper .95
Treatment      1.111      0.9002  0.7171      1.721

Concordance= 0.562 (se = 0.031 )
Likelihood ratio test= 0.22 on 1 df,  p=0.6
Wald test               = 0.22 on 1 df,  p=0.6
Score (logrank) test = 0.22 on 1 df,  p=0.6
```

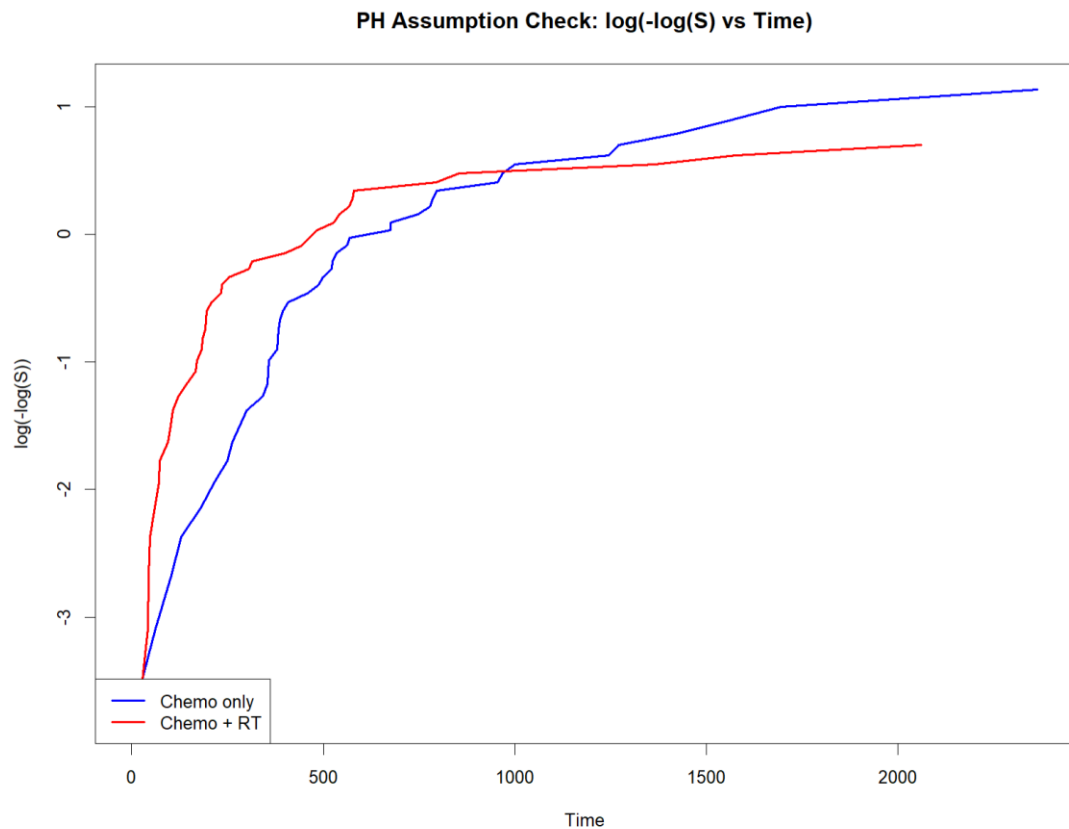
根據 Cox 比例風險模型的結果：治療方式（化放療 vs. 化療）對病人存活時間的影響不顯著。化放療組的估計風險比 (HR = 1.11) 稍高於化療組，但差異並不顯著 (p = 0.638)。表示根據資料，加上放射治療並沒有明顯降低或提高死亡風險。因此，無法拒絕虛無假設，治療方式沒有顯著影響生存風險。

3. Following Question 2, draw the fitted survival curve on the same figure of Question 1 and compare with Kaplan-Meier curves. Use the result to conduct model checking.



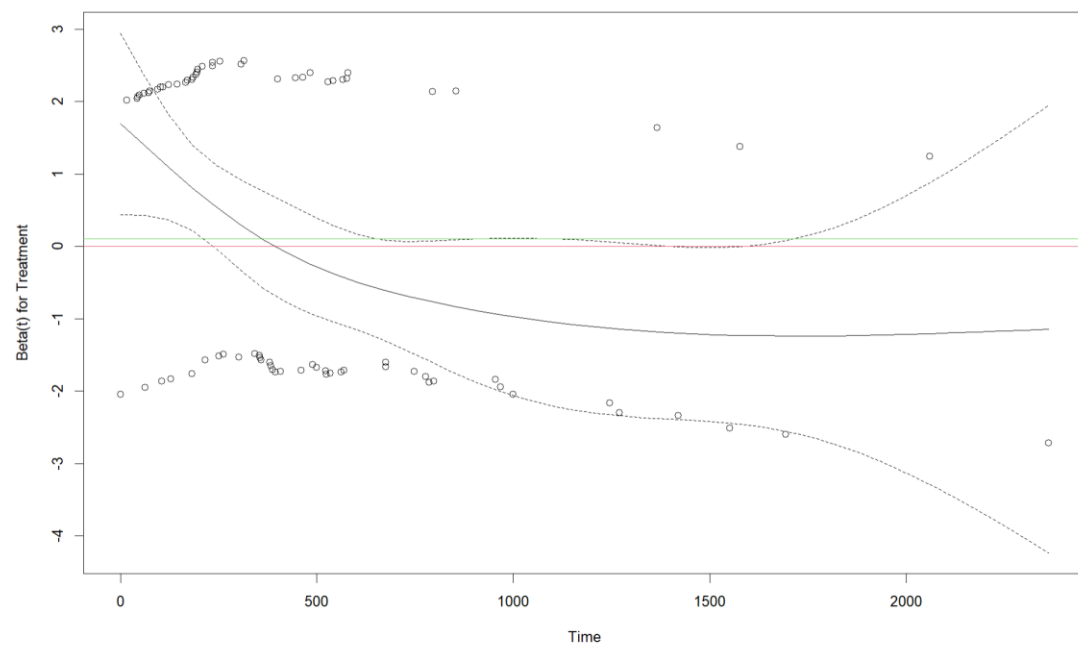
Cox 擬合曲線與 KM 曲線有明顯差距，有可能是因為 Cox hazard ratio 隨時間變化，比例風險假設可能不成立，所以 Cox 曲線擬合得不好。

4. Use the log-log plots to diagnose Cox model.



從 $\log\text{-}\log$ plot 觀察到兩組曲線不太平行且在後期有交叉，加放療組曲線下降至低於單純化療組，顯示比例風險假設不成立。這與 Cox 擬合 KM 曲線距離較遠的現象一致，表示 Cox 模型在此資料中對生存函數的估計不好，Cox 模型用固定危險比描述 Treatment 效果，會錯估後期加放療組的事件風險（生存機會）。

5. Use Schoenfeld residuals for testing time-varying effect, and draw time-varying effect $\beta(t)$.



```
> zph.cox
      chisq df      p
Treatment   11  1 0.00093
GLOBAL      11  1 0.00093
```

Treatment:

$H_0 : \beta(t) = \beta$ (The hazard ratio for Treatment is constant over time,proportional hazards holds)

$H_1 : \beta(t) \neq \beta$ (The hazard ratio for Treatment changes over time,proportional hazards does not hold)

GLOBAL:

H_0 : Proportional hazards assumption holds for all covariates

H_1 : Proportional hazards assumption is violated for at least one covariate

Schoenfeld 殘差檢定顯示 Treatment 的比例風險假設不成立 ($p < 0.001$)，表示其危險比隨時間變化；同時也違反整體模型 PH 的假設。

程式碼網址:

https://github.com/Lai-jun-yan/Survival_data_analysis/tree/master/%E4%BD%9C%E6%A5%AD3