存活分析期末報告
Bladder cancer

M122040017
吳俞憲

目錄

內　　容

PART 01

資料介紹

1. 研究動機
2. 變數介紹

資 料 介 紹

資料來源: MSK, Clin Cancer Res 2023

癌症介紹:
膀胱癌是常見的泌尿系統惡性腫瘤，最常見的症狀包括血尿、頻尿和排尿困難。它通常發生在60歲以上的男性身上，且具有高復發率。但只要通過定期覆診，治癒的機會也相當高。

研究動機:
膀胱是我們日常生活中很重要的器官，然而膀胱癌確切的成因尚未了解，所以想要透過簡單的分析來更了解有什麼因素會影響膀胱癌。

變 數 介 紹

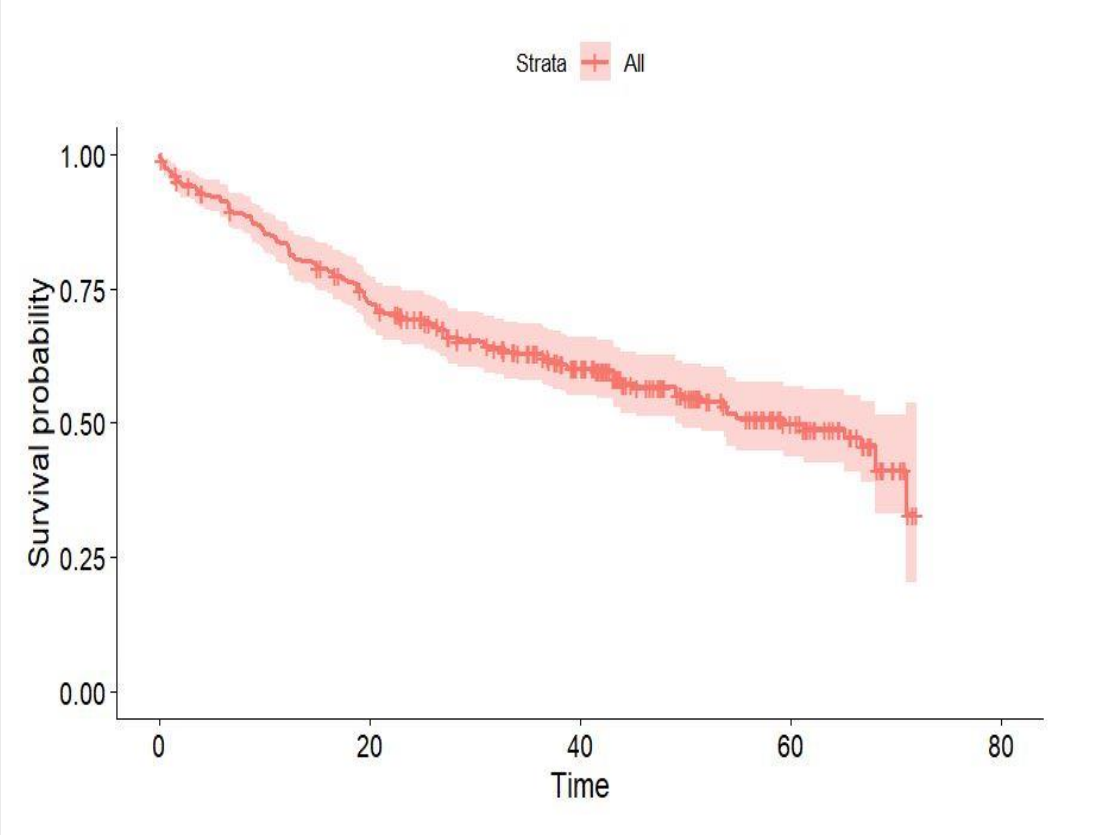
變數名稱	變數解釋
Cancer.Type.Detailed	Bladder Urothelial Carcinoma(膀胱尿路上皮癌) Upper Tract Urothelial Carcinoma(上泌尿道尿路上皮癌) Urethral Urothelial Carcinoma(尿道尿路上皮癌)
Age	診斷年紀
Fraction.Genome.Altered(FGA)	拷貝數受影響的百分比
MSI.Score	微星體不穩定性的分數
Mutation.Count	基因突變數量
TMB	腫瘤突變負荷
Sex	性別
Overall.Survival..Months	存活時間(月)
Overall.Survival.Status	1:事件發生，0:censoring

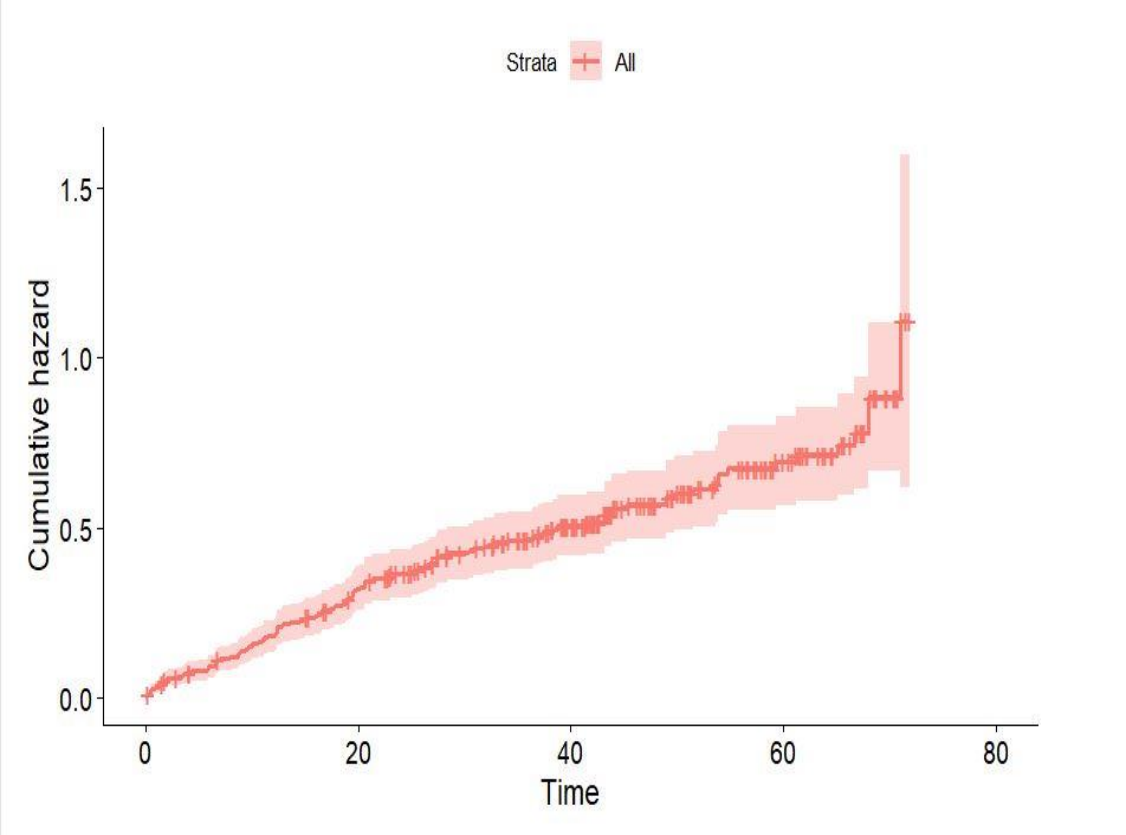
PART 02

模型應用及解釋

1. Km’s estimator and NA’s estimator
2. Log-rank test
3. Cox PH model
4. Local test

K-M estimator for S(t) and N-A estimator for H(t)

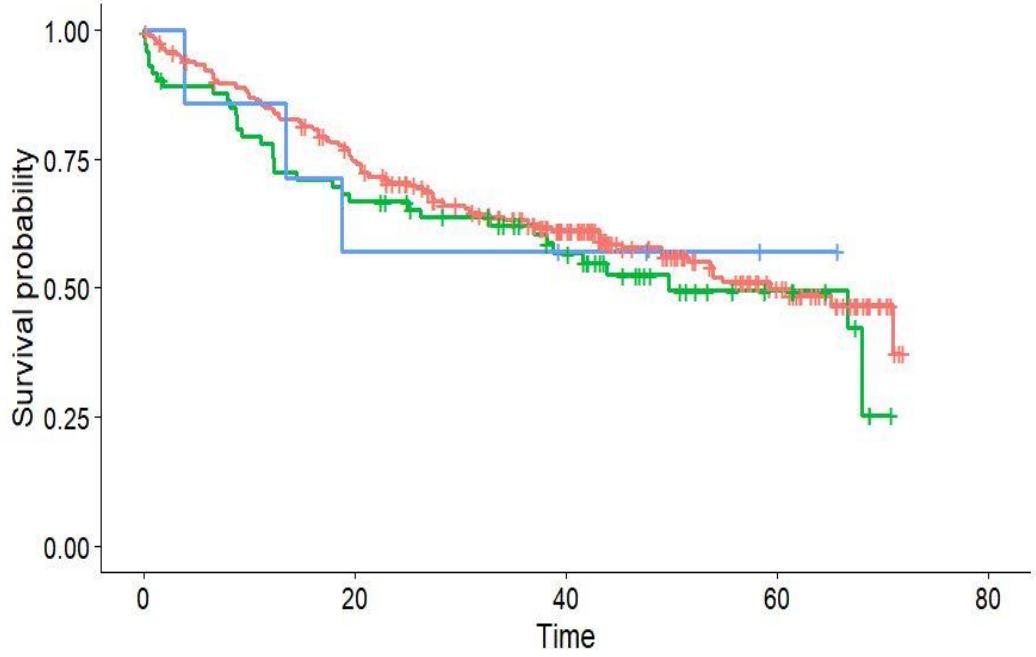




Log-rank test

- Bladder Urothelial Carcinoma
- Upper Tract Urothelial Carcinoma
- Urethral Urothelial Carcinoma

$$H_0 : h_1(t) = h_2(t) = h_3(t)$$




```
Call:
survdiff(formula = Surv(event, time_new) ~ factor(cancer), data = all)

                                                N Observed Expected (O-E)^2/E (O-E)^2/V
factor(cancer)=Bladder Urothelial Carcinoma   258      110   115.47   0.25938   1.15910
factor(cancer)=Upper Tract Urothelial Carcinoma  73       36    30.46   1.00588   1.27142
factor(cancer)=Urethral Urothelial Carcinoma      7        3     3.06   0.00129   0.00132

 Chisq= 1.3  on 2 degrees of freedom, p= 0.5
```

Cox PH model

```
Call:
coxph(formula = Surv(event, time_new) ~ factor(cancer) + factor(sex) +
    age + Genome + MSI + Mutation + tmb, data = all)

  n= 338, number of events= 149

                                              coef exp(coef)  se(coef)      z Pr(>|z|)
factor(cancer)Upper Tract Urothelial Carcinoma 0.305145  1.356822 0.197696  1.544 0.122708
factor(cancer)Urethral Urothelial Carcinoma    0.373815  1.453268 0.595900  0.627 0.530456
factor(sex)Male                                0.279171  1.322033 0.183597  1.521 0.128370
age                                            0.034195  1.034786 0.008518  4.015 5.96e-05 ***
Genome                                         2.025266  7.578124 0.562070  3.603 0.000314 ***
MSI                                           -0.128048  0.879811 0.053231 -2.406 0.016150 *
Mutation                                      -0.289414  0.748702 0.075134 -3.852 0.000117 ***
tmb                                            0.293592  1.341236 0.077716  3.778 0.000158 ***
```

```
Concordance= 0.662  (se = 0.022 )
Likelihood ratio test= 63.47  on 8 df,   p=1e-10
Wald test            = 43.57  on 8 df,   p=7e-07
Score (logrank) test = 46.74  on 8 df,   p=2e-07
```

可以看出三個test 的p value都很小，代表這個模型比什麼都不放的模型擬合的要好

大多數變數都是顯著的，可能需要去查看有沒有共線性問題

查看模型是否有共線性問題

利用廣義方差膨胀因子(GVIF)

	GVIF	Df	GVIF^(1/(2*Df))
factor(cancer)	1.092636	2	1.022395
factor(sex)	1.043605	1	1.021570
age	1.008580	1	1.004281
Genome	1.158694	1	1.076427
MSI	1.450463	1	1.204352
Mutation	241.757186	1	15.548543
tmb	247.401589	1	15.729005

刪除掉tmb

	GVIF	Df	GVIF^(1/(2*Df))
factor(cancer)	1.069990	2	1.017056
factor(sex)	1.033671	1	1.016696
age	1.012368	1	1.006165
Genome	1.062969	1	1.031004
MSI	1.224292	1	1.106477
Mutation	1.180215	1	1.086377

Cox PH model

```
Call:
coxph(formula = Surv(event, time_new) ~ factor(cancer) + factor(sex) +
    age + Genome + MSI + Mutation, data = all)

  n= 338, number of events= 149

                                              coef exp(coef)  se(coef)      z Pr(>|z|)
factor(cancer)Upper Tract Urothelial Carcinoma  0.360678  1.434302  0.196445  1.836  0.06635 .
factor(cancer)Urethral Urothelial Carcinoma     0.157291  1.170336  0.593617  0.265  0.79103
factor(sex)Male                                 0.227266  1.255164  0.182692  1.244  0.21350
age                                             0.035824  1.036473  0.008699  4.118 3.82e-05 ***
Genome                                          1.497989  4.472684  0.532815  2.811  0.00493 **
MSI                                            -0.070067  0.932331  0.052785 -1.327  0.18437
Mutation                                       -0.020311  0.979894  0.010026 -2.026  0.04278 *
```

```
Concordance= 0.653  (se = 0.023 )
Likelihood ratio test= 49.62  on 7 df,   p=2e-08
Wald test            = 36.76  on 7 df,   p=5e-06
Score (logrank) test = 38  on 7 df,   p=3e-06
```

可以看出三個test 的p value都很小，代表這個模型比什麼都不放的模型擬合的要好

Local test

查看膀胱癌類別是否對存活有影響

```
Likelihood ratio test

Model 1: Surv(event, time_new) ~ factor(sex) + age + Genome + MSI + Mutation
Model 2: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + MSI + Mutation
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -771.37
2   7 -769.77  2 3.1964     0.2023
```

```
Wald test

Model 1: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + MSI + Mutation
Model 2: Surv(event, time_new) ~ factor(sex) + age + Genome + MSI + Mutation
  Res.Df Df  Chisq Pr(>Chisq)
1    142
2    144 -2 3.3831     0.1842
```

兩種test的p value>0.05，所以膀胱癌類別沒有影響

Local test

查看性別是否對存活有影響

```
Likelihood ratio test

Model 1: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + MSI + Mutation
Model 2: Surv(event, time_new) ~ factor(cancer) + age + Genome + MSI +
    Mutation
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   7 -769.77
2   6 -770.56 -1 1.5918     0.2071
```

```
Wald test

Model 1: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + MSI + Mutation
Model 2: Surv(event, time_new) ~ factor(cancer) + age + Genome + MSI +
    Mutation
  Res.Df Df  Chisq Pr(>Chisq)
1    142
2    143 -1 1.5475     0.2135
```

兩種test的p value>0.05，所以性別沒有影響

Local test

查看MSI是否對存活有影響

```
Likelihood ratio test

Model 1: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + MSI + Mutation
Model 2: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + Mutation
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   7 -769.77
2   6 -771.01 -1 2.4751     0.1157
Wald test

Model 1: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + MSI + Mutation
Model 2: Surv(event, time_new) ~ factor(cancer) + factor(sex) + age +
    Genome + Mutation
  Res.Df Df Chisq Pr(>Chisq)
1    142
2    143 -1 1.762     0.1844
```

```
Call:
coxph(formula = Surv(event, time_new) ~ MSI, data = all)

  n= 338, number of events= 149

       coef exp(coef) se(coef)     z Pr(>|z|)
MSI -0.08655   0.91709  0.03797 -2.28   0.0226 *
```

導致的原因可能是MSI這個變數的效應已經被其他變數解釋完了

目前的模型

```
Call:
coxph(formula = Surv(event, time_new) ~ age + Genome + Mutation,
    data = all)

  n= 338, number of events= 149

              coef exp(coef)  se(coef)     z Pr(>|z|)
age       0.037501  1.038213  0.008493  4.416 1.01e-05 ***
Genome    1.393367  4.028389  0.518259  2.689  0.00718 **
Mutation -0.027374  0.972997  0.009022 -3.034  0.00241 **
```

```
Concordance= 0.638  (se = 0.023 )
Likelihood ratio test= 43.74  on 3 df,   p=2e-09
Wald test            = 34.29  on 3 df,   p=2e-07
Score (logrank) test = 30.28  on 3 df,   p=1e-06
```

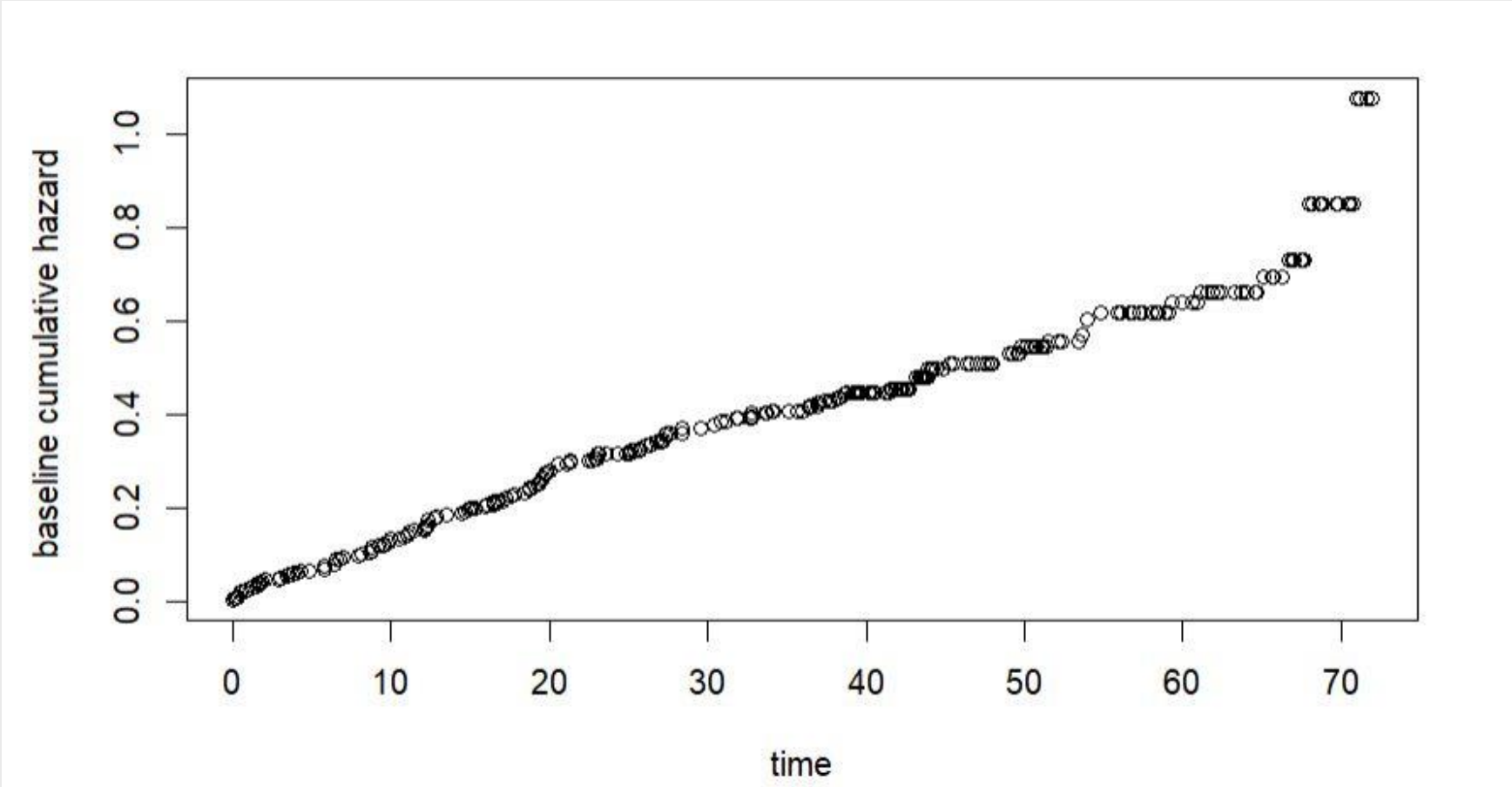
與原本的模型的Concordance相差不大

PART 03

模型基本假設

1. Breslow‘s baseline hazard function
2. Schoenfeld residual
3. Cox - Snell residual

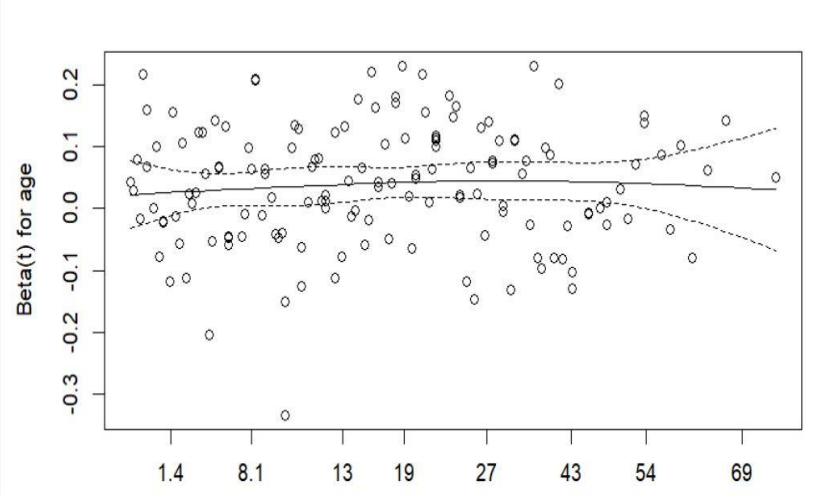
Breslow‘s baseline hazard function



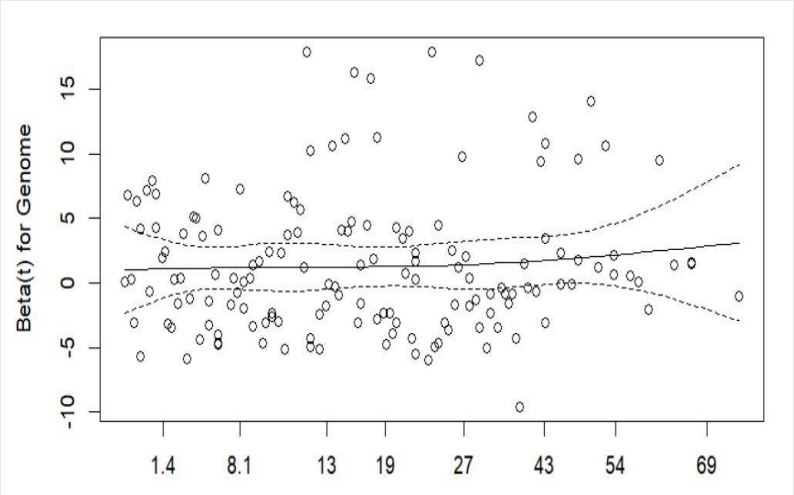
圖片呈現出斜直線，代表每個時間的風險都差不多

Schonfeld residuals

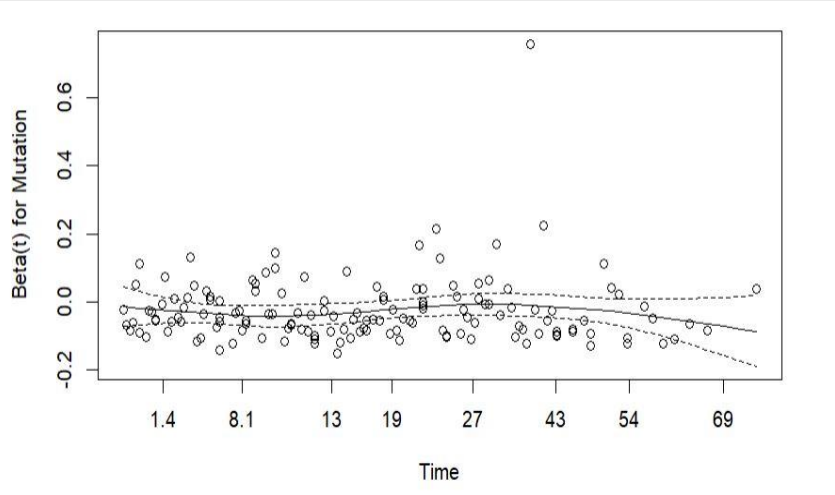
Age



FGA



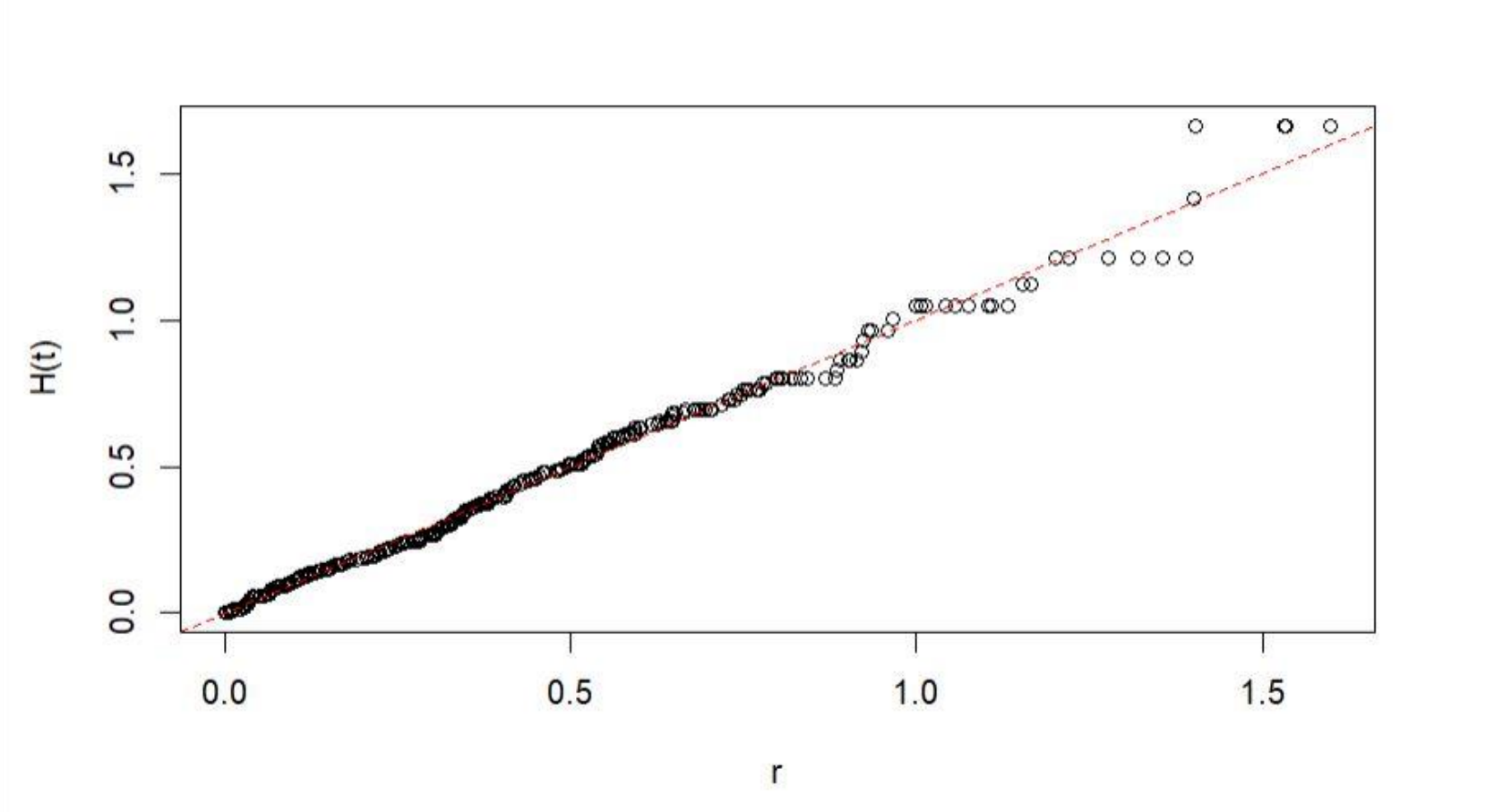
Mutation



	chisq	df	p
age	0.4226	1	0.52
Genome	0.5711	1	0.45
Mutation	0.0774	1	0.78
GLOBAL	1.0511	3	0.79

每個變數的p value都非常大，代表變數有滿足模型假設

Cox-snell residuals



大部分的點都落在虛線上，所以模型有符合PH model的假設

PART 04

結論

結論

最終模型:

$$h(t|z)=h_0(t)e^{0.0375age+1.3934Genome+(-0.0273)Mutation}$$

1.此模型有滿足Cox PH model 的假設

2.Age增加風險也是增加的，這跟我們認知的是一樣的

3.FGA越大，風險是增加的

4.突變數量越多，風險是降低的

引用

FGA and Mutation:

https://www.thehyve.nl/articles/fraction-of-genome-altered-total-mutations-cbioportal

Copy number:

https://baike.baidu.hk/item/%E6%8B%B7%E8%B2%9D%E6%95%B8/10042577

Tmb:

https://geneonline.news/tmb-in-prostate-cancer/

膀胱癌:

https://heal-oncology.com/cancer/%E8%86%80%E8%83%B1%E7%99%8C/

https://bladdercancercanada.org/en/guidebook-translations/%E4%B8%8A%E6%B3%8C%E5%B0%BF%E9%81%93%E5%B0%BF%E8%B7%AF%E4%B8%8A%E7%9A%AE%E7%99%8C-utuc-%E6%82%A3%E8%80%85%E6%8C%87%E5%8D%97/

報告結束感謝聆聽