

食道癌

存活分析期末報告 B082040005 高念慈

2023/05/29

目錄

◆ 動機目的

◆ 模型

◆ 變數、資料介紹

◆ 分析結果、解釋

◆ 變數挑選

◆ 結論

◆ 參考資料



動機目的

■ 動機

- 食道雖然不像肝臟幾乎沒感覺，但也是個大家平常不會特別注意到的一個器官，能找出影響食道癌的特徵，有助於防患未然

■ 目的

- 從108個變數中找出影響食道癌的變數，並做出解釋

 今健康

 今健康

食道癌初期症狀不明顯， 患者發現時多已中晚期。

 吞嚥卡痛	 體重減輕
 聲音沙啞	 摸到硬塊

資訊來源：臺北榮民總醫院胸腔外科主治醫師 徐博奎



變數、資料介紹



Esophageal Carcinoma (TCGA, Nature 2017)

Whole exome sequencing of 559 esophageal and gastric carcinoma tumors with matched normals. [PubMed](#)

[Summary](#)
[Clinical Data](#)
[CN Segments](#)

Selected: 559 patients | 559 samples


[Custom Selection](#)
[Columns](#)
[Groups](#)

[Study Page Help](#)

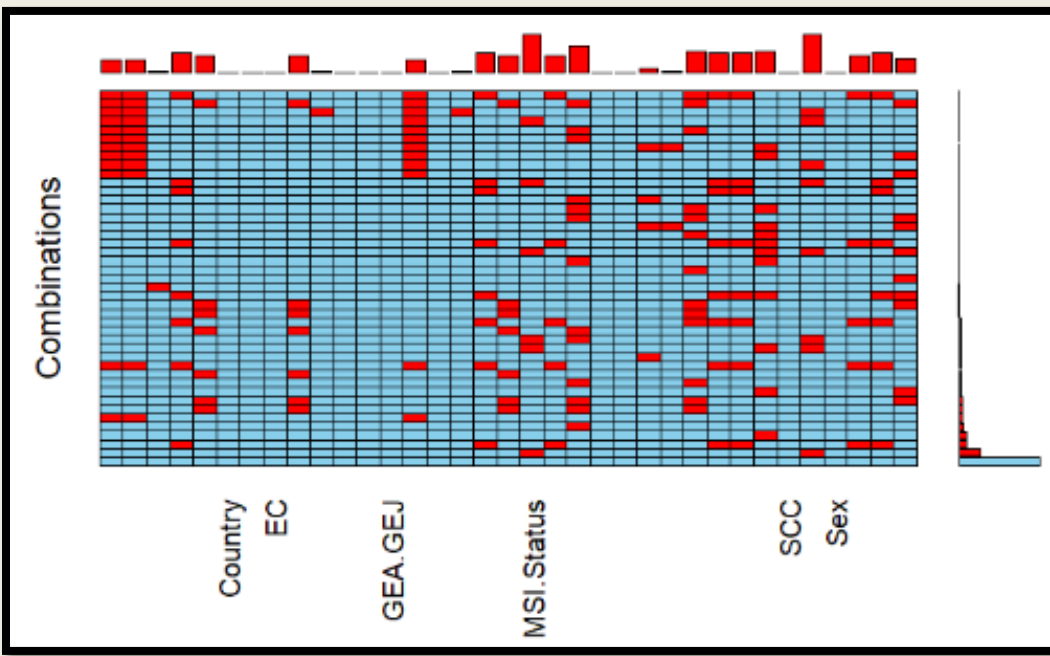
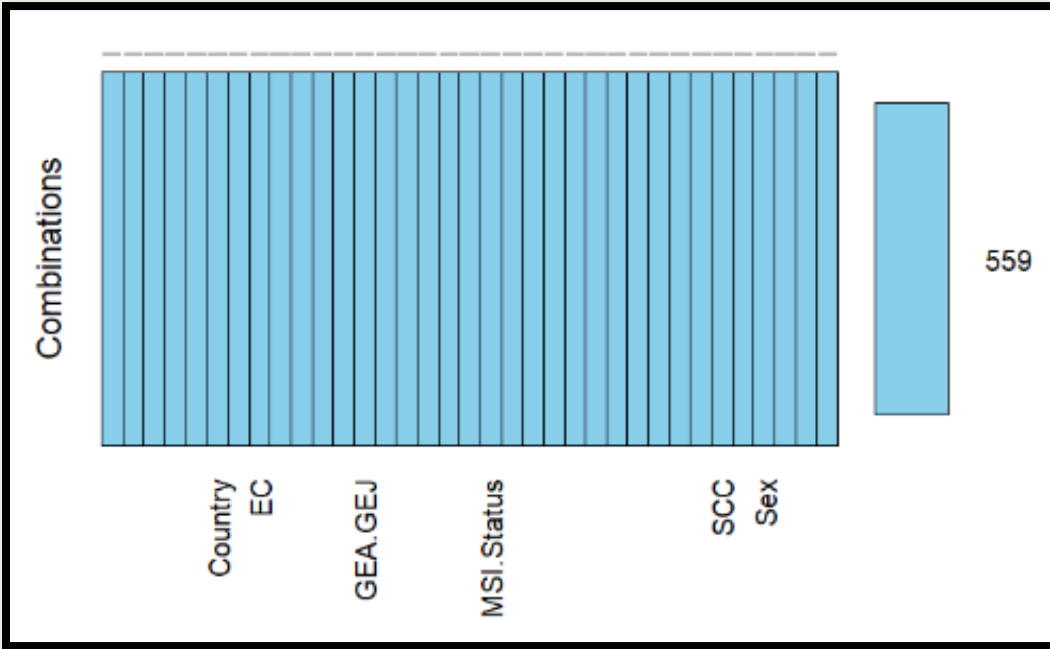
Patient ID	Sample ID	Cancer Study	Cancer Type	Cancer Type Detailed	Number of Samples Per Patient	Mutation Count	Fraction Genome Altered	Diagnosis Age	Sample Type	Sex	Absolute cancer dna fraction	Absolute Extract Ploidy	Absolute Extract Purity	American Joint Committee on Cancer Metastasis Stage Code	American Joint Committee on Cancer Tumor Stage Code	ARHGAP26-ARHGAP6-CLDN18 Rearrangem
TCGA-2H-A9GF	TCGA-2H-A9GF-01	stes_tcga_pub	Esophagogastric Cancer	Esophageal Adenocarcinoma	1	344	0.5418	67.04	Primary	Male	0.75	2.84	0.68	M0	T3	
TCGA-2H-A9GH	TCGA-2H-A9GH-01	stes_tcga_pub	Esophagogastric Cancer	Esophageal Adenocarcinoma	1	215	0.3645	44.35	Primary	Male	0.71	2.93	0.62	M0	T1	
TCGA-2H-A9GI	TCGA-2H-A9GI-01	stes_tcga_pub	Esophagogastric Cancer	Esophageal Adenocarcinoma	1	285	0.4343	68.71	Primary	Male	0.67	2.66	0.6	M0	T3	
TCGA-2H-A9GK	TCGA-2H-A9GK-	stes_tcga_pub	Esophagogastric Cancer	Esophageal Adenocarcinoma	1	319	0.2518	43.99	Primary	Male	0.47	2.07	0.47	M0	T3	

- 資料來源 : https://www.cbioportal.org/study/clinicalData?id=stes_tcga_pub

前處理

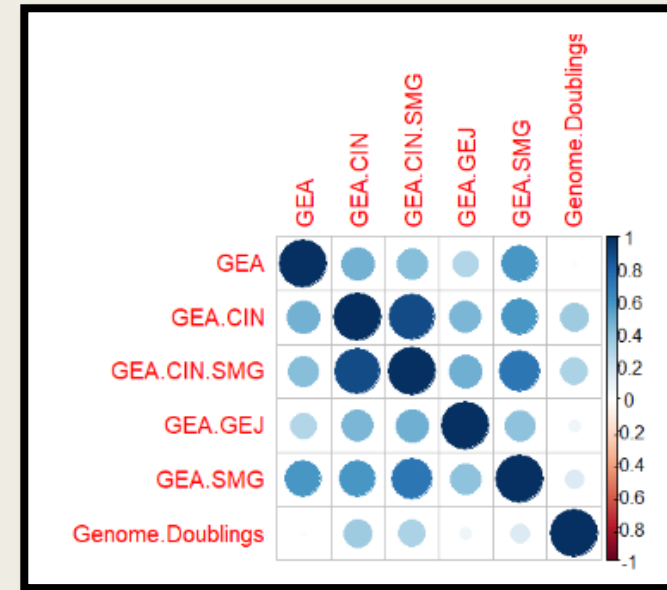
原資料大小 : 559 obs. of 108 variables

- $559 * 0.8 = 447.2$
 - ✓ 移除缺失值超過 $559 - 448 = 111$ 的資料 → 剩 56 variables
- 去掉沒用 & 重複/相關性高 & 變異性小的變數 → 剩 35 variables
- 補缺失值
 - ✓ 分出存活和死亡，分別填補各列中位數，類別變數則取眾數



➤ 補缺失值前

➤ 相關性



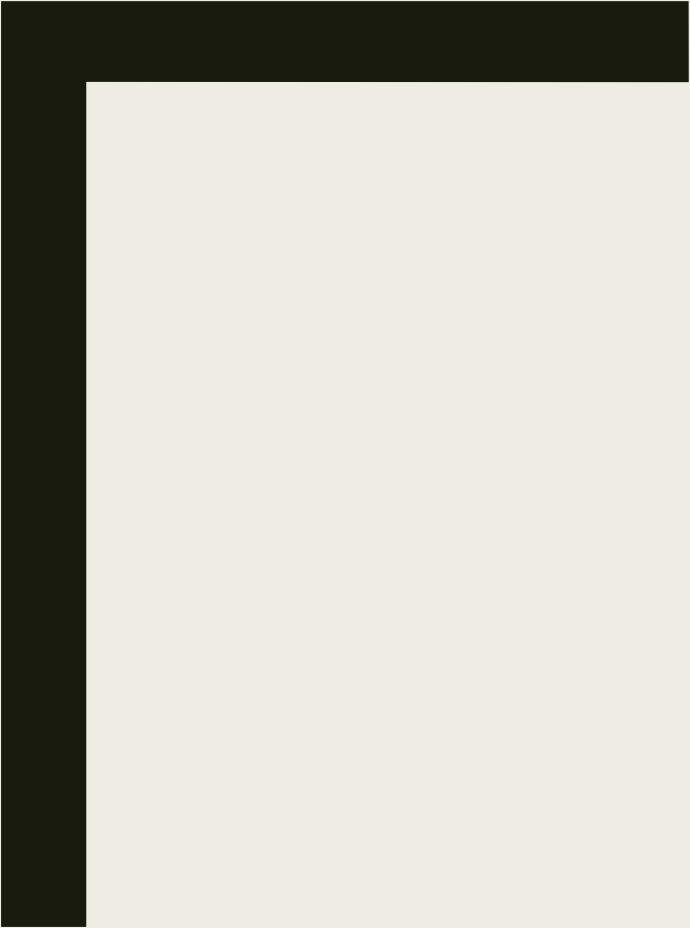
➤ 補缺失值後

- "Absolute.Extract.Ploidy" : 絕對.提取物.倍性 , (num.)
倍性水平是指雙倍體體細胞 (2n) 或單倍體配子體細胞 (1n) 中染色體組的數量。
- "Absolute.Extract.Purity" : 絕對提取物純度 , (num.)
- "Diagnosis.Age" : 診斷年齡 , (num.) , 首次診斷出病症或疾病的年齡
- "ARID1A.Protein.Coding" : ARID1A 蛋白質編碼 , (int.0:NO,1:YES)
- "CDKN2A.Methylation" : CDKN2A 甲基化 , (logi.FALSE,TRUE)
CDKN2A 甲基化可能會導致基因的表達受到抑制或關閉 , 從而影響細胞週期的調控
- "Country" : 國家 , (chr.)
- "EBV.Positive" : 伊普斯坦-巴爾病毒陽性 , (int.) , EBV 是一種 DNA 病毒 , 它可以感染人類的 B 淋巴細胞 , 並在體內引起多種疾病。(喉嚨痛、疲勞、淋巴結腫大相關)
- "Overall.Survival..Months." : 總體.生存..月 , (num.) , 從首次診斷出疾病到結束 , 目標!
- "Overall.Survival.Status" : 總體.生存.狀態 , (chr.1:DECEASED , 0:LIVING) , 目標!
- "Percent.Tumor.Cells" : 癌細胞百分比 , (num.)
- "Percent.Tumor.Nuclei" : 癌細胞核百分比 , (num.)
- "Person.Neoplasm.Cancer.Status" : 人體腫瘤癌症狀態 , (chr.TUMOR_FREE/WITH_TUMOR) , 描述腫瘤的有無、腫瘤的種類、腫瘤的大小和擴散程度 , 以及其他與腫瘤狀態相關的信息

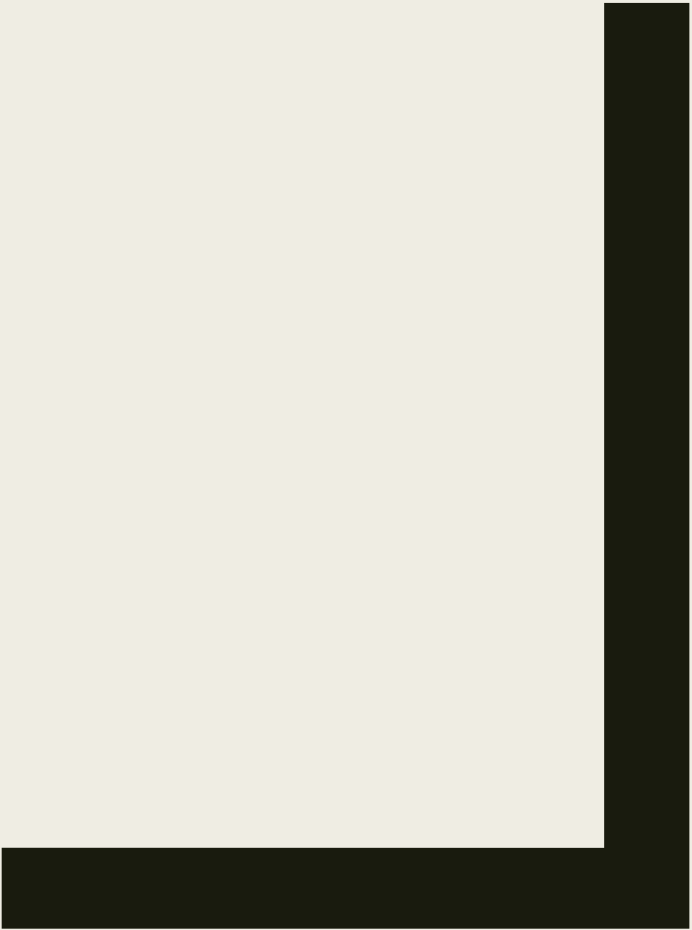
- "EC" , (int.0:NO,1:YES) , EC 可能代表食道攝影檢查、食道黏膜刷取檢查、食道鏡檢查
- "EC.SMG" , (int.0:NO,1:YES) , EC.SMG 可能代表食道超聲內視鏡檢查，也可能指食道超聲引導下淋巴結活檢
- "Estimated.Leukocyte.Percentage" : 估計的白血球百分比 , (num.) , 提供關於白血球組成的信息，幫助醫生判斷患者的免疫狀態和健康狀況
- "Fraction.Genome.Altered" : 基因組改變比例 , (num.) , 腫瘤細胞中發生基因組改變的比例，如果腫瘤細胞中有多個區域發生了基因組改變，值就會較高。
- "Gastric.Classification" : 胃.分類 , (chr.)
- "GEA.CIN.SMG" , (int.0:NO,1:YES) , 可能表示在食道癌檢查中同時進行食道細胞異常檢查和腹部超聲檢查
- "GEA.GEJ" , (int.0:NO,1:YES) , 可能表示在食道癌檢查中同時進行上消化道鏡檢查以及專注於胃食道連接處的評估
- "GEA.SMG" , (int.0:NO,1:YES) , 可能表示在食道癌檢查中同時進行腹部超聲和上消化道鏡檢查
- "Genome.Doublings" : 基因體倍增 , (int.) , 細胞的染色體數目增加為正常倍數的兩倍，它會影響細胞的功能和癌症的發展。
- "PIK3CA.Protein.Coding" : PIK3CA 蛋白質編碼 , (int.0:NO,1:YES) PIK3CA 基因突變或過度表達與多種腫瘤的發生和進展有關，包括乳腺癌、卵巢癌、大腸癌、胃癌等

- **"Neoplasm.Histologic.Grade" : 腫瘤.組織學.分級 , (chr.)** , 低分級通常表示腫瘤細胞較接近正常組織 , 分化程度較好 , 高分級則表示腫瘤細胞較不正常 , 分化程度較差
- **"HER2.Amplification.Status" : 人類表皮生長因子受體2基因放大的狀態 , (int.0:正常,1:放大)** , HER2 放大可能促進癌症細胞的生長和擴散 , 在乳腺癌中相對常見
- **"KRAS.Protein.Coding" : KRAS蛋白編碼 , (int.0:NO,1:YES)** KRAS 基因是一個重要的腫瘤抑制基因 , 它在調節細胞生長和分化中起著關鍵的作用 , 特別是大腸癌和肺癌等
- **"MLH1.Methylation" : MLH1甲基化 , (logi.FALSE,TRUE)** MLH1甲基化導致基因組中的MLH1基因失去表達 , 這意味著蛋白質無法正常產生。缺乏 MLH1蛋白質會干擾 DNA 配對修復機制 , 使得 DNA 錯誤無法被及時修復 , 增加細胞的突變風險
- **"MSI.Status" : 微衛星不穩定性的狀態 , (chr.)** , 微衛星穩定 (MSS) : 腫瘤組織中微衛星序列保持穩定 , 沒有明顯的不穩定性 , 微衛星不穩定 (MSI) : 腫瘤組織中出現微衛星不穩定性 , 表明微衛星序列的錯誤或缺陷
- **"Mutation.Rate" : 突變率 , (num.)** , 在一個特定基因或基因群中突變事件發生的頻率
- **"New.Neoplasm.Event.Post.Initial.Therapy.Indicator" : 新的腫瘤。事件.後.初始.治療.指標 , (chr,YES/NO)** , 在初次治療後是否出現新的腫瘤事件。Positive/Yes : 可能意味著癌症復發、轉移或出現了新的原發腫瘤 , Negative/No : 意味著患者在初次治療後沒有出現新的腫瘤或腫瘤復發的跡象

- "RHOA.Protein.Coding" : RHOA 蛋白質編碼，(int.0:NO,1:YES) RHOA 基因的突變或異常活化可能導致細胞的遷移、侵襲和轉移能力增強，從而促進腫瘤的進展和轉移
- "Surgical.Margin.Resection.Status" : 手術邊緣切除狀態，(chr.)，指在手術過程中腫瘤或病變組織周圍的切緣是否被完全切除的狀態。正常組織 (Negative Margin，切緣未檢測到腫瘤或病變組織) 不確定 (Close Margin，切緣附近可能存在腫瘤或病變組織) 腫瘤或病變組織殘留 (Positive Margin，切緣檢測到腫瘤或病變組織)
- "SCC" : 鱗狀細胞癌，(int.,0:NO,1:YES)
- "SCC.SMG"，(int.,0,1)，(int.,0:NO,1:YES)，"SMG" 可以代表 SLNM、Mediastinoscopy、Groove Node Dissection等檢查項目
- "SCNA.Cluster.GEA" : 基因組拷貝數異常在胃食道腺癌中的聚類分析，(chr.)，描述基因組拷貝數異常的模式，使用聚類分析方法將相似的拷貝數異常模式的腫瘤分到同一個聚類中
- "Sex" : 性別
- "TMB..nonsynonymous." : 腫瘤組織中非同義突變的數量，(num.)，高 TMB.nonsynonymous 暗示著腫瘤組織中存在大量的基因突變，可能對免疫系統的識別和攻擊產生影響
- "TP53.Protein.Coding" : TP53 蛋白質編碼，(int.0:NO,1:YES) TP53 是人類體內最為重要的腫瘤抑制基因之一，TP53 的突變可能導致 p53 蛋白質的功能喪失或減弱，從而降低細胞對 DNA 損傷的應答和防止腫瘤發展的能力。
- "Tumor.Stage" : 腫瘤階段，(chr.)



變數挑選

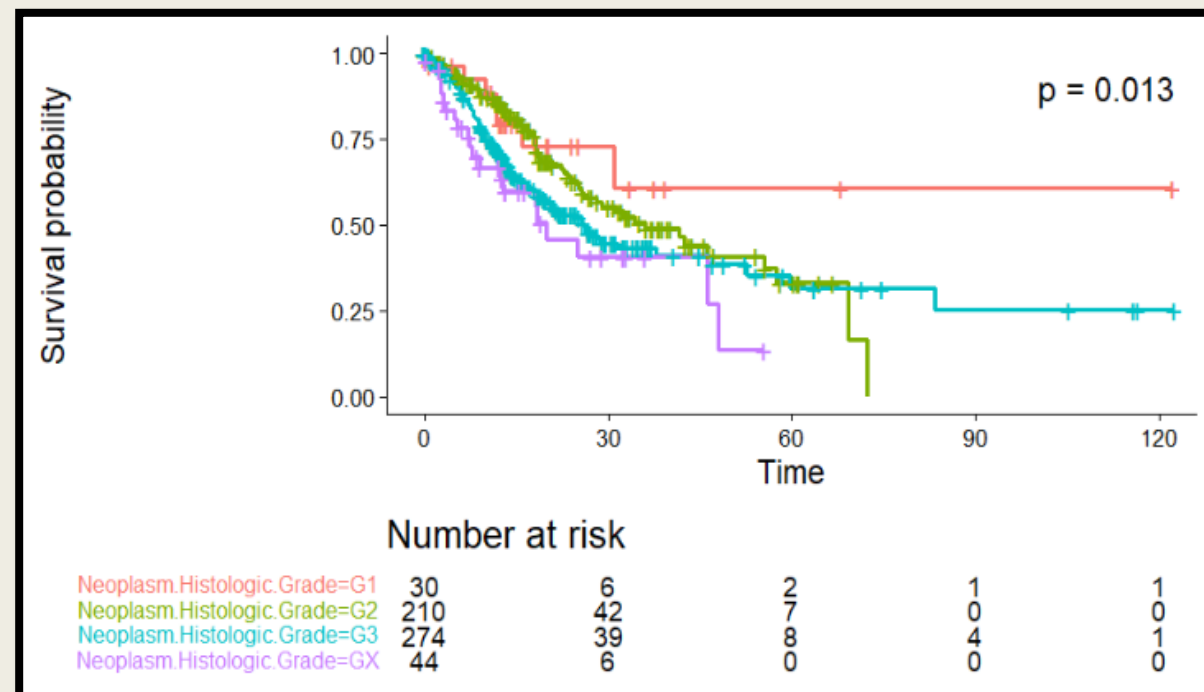
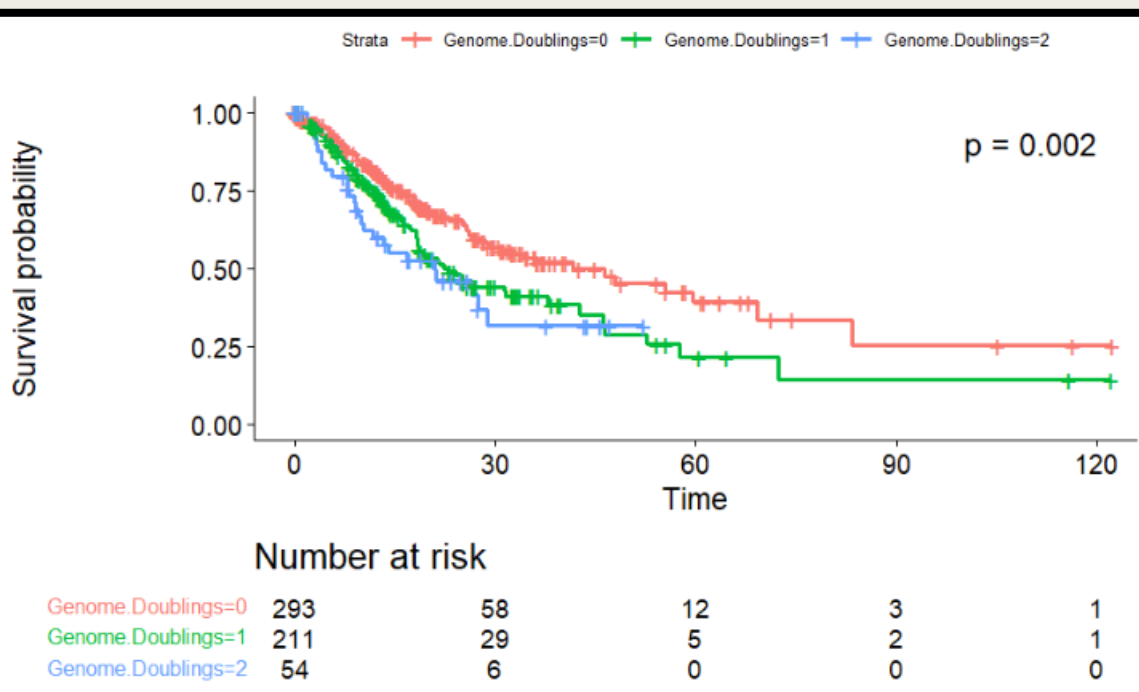
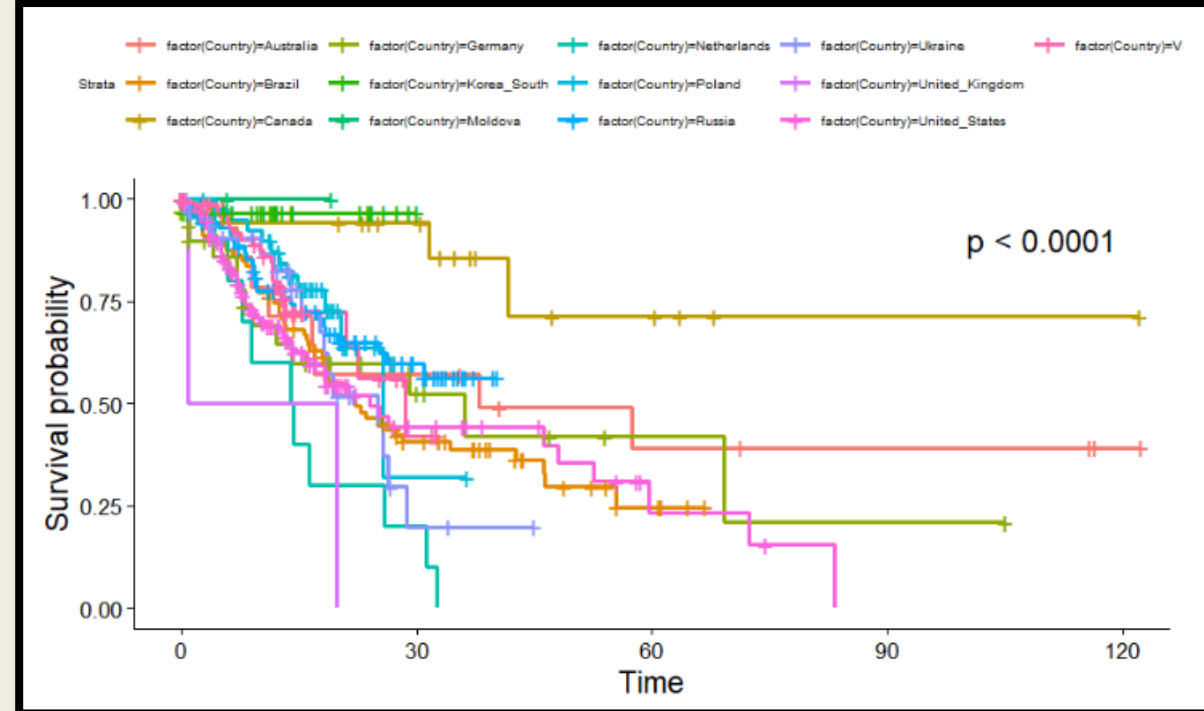
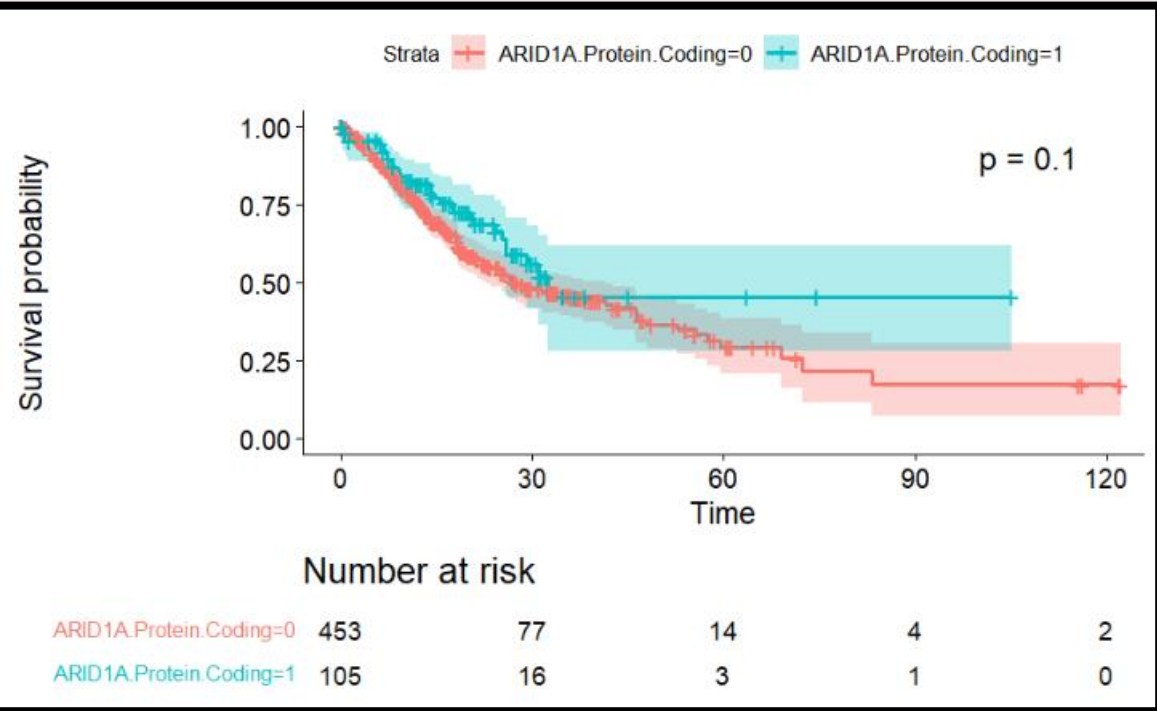


Log rank test

- test 整個風險函數有無差異
- 不考慮其他變數
- 權重 : $p = q = 0$
- $H_0: h_1(t) = h_2(t)$

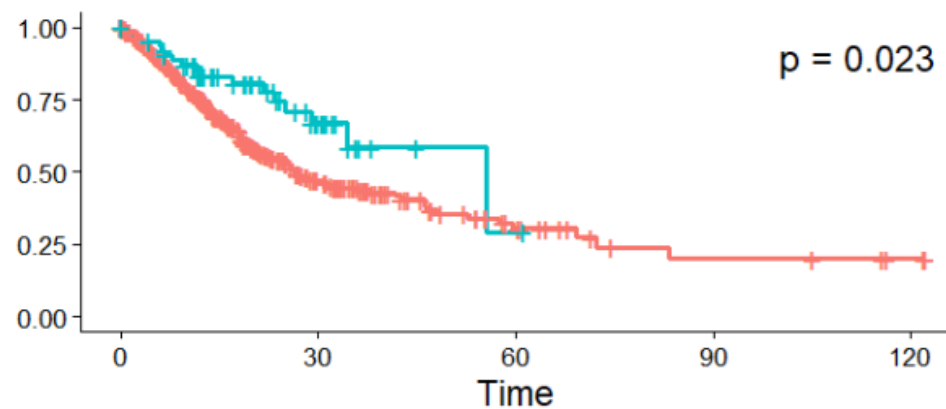
Group(類別變數):

1. ARID1A.Protein.Coding : ARID1A 蛋白質編碼, (int.0:NO,1:YES)
2. CDKN2A.Methylation : CDKN2A 甲基化, (logi.FALSE,TRUE)
3. Country : 國家, (chr.)
4. EBV.Positive : 伊普斯坦-巴爾病毒陽性, (int.)
5. EC, (int.0:NO,1:YES)
6. Gastric.Classification : 胃.分類, (chr.)
7. GEA.CIN.SMG, (int.0:NO,1:YES)
8. GEA.GEJ, (int.0:NO,1:YES)
9. Genome.Doublings : 基因體倍增, (int.)
10. Neoplasm.Histologic.Grade : 腫瘤.組織學.分級, (chr.)
11. HER2.Amplification.Status" : 人類表皮生長因子受體2基因放大的狀態, (int.0:正常,1:放大)
12. KRAS.Protein.Coding : KRAS蛋白質編碼, (int.0:NO,1:YES)
13. MLH1.Methylation : MLH1甲基化, (logi.FALSE,TRUE)
14. MSI.Status : 微衛星不穩定性的狀態, (chr.)
15. New.Neoplasm.Event.Post.Initial.Therapy.Indicator" : 新的腫瘤.事件.後.初始.治療.指標, (chr,YES/NO)
16. Person.Neoplasm.Cancer.Status : 人體腫瘤癌症狀態, (chr.TUMOR_FREE/WITH_TUMOR)
17. PIK3CA.Protein.Coding : PIK3CA 蛋白質編碼, (int.0:NO,1:YES)
18. RHOA.Protein.Coding : RHOA 蛋白質編碼, (int.0:NO,1:YES)
19. Surgical.Margin.Resection.Status : 手術邊緣切除狀態, (chr.)
20. SCC : 鱗狀細胞癌, (int.,0:NO,1:YES)
21. SCNA.Cluster.GEA : 基因組拷貝數異常在胃食道腺癌中的聚類分析, (chr.)
22. Sex : 性別
23. TP53.Protein.Coding : TP53 蛋白質編碼, (int.0:NO,1:YES)
24. Tumor.Stage : 腫瘤階段, (chr.)



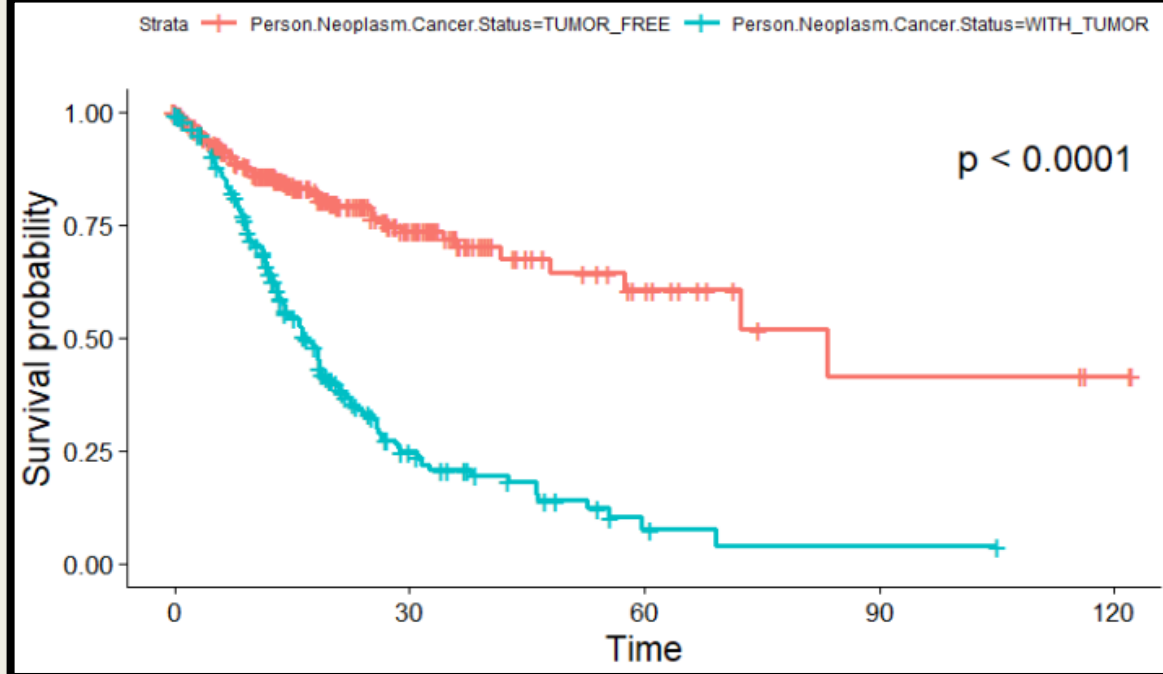
Survival probability

Strata + MLH1.Methylation=0 + MLH1.Methylation=1



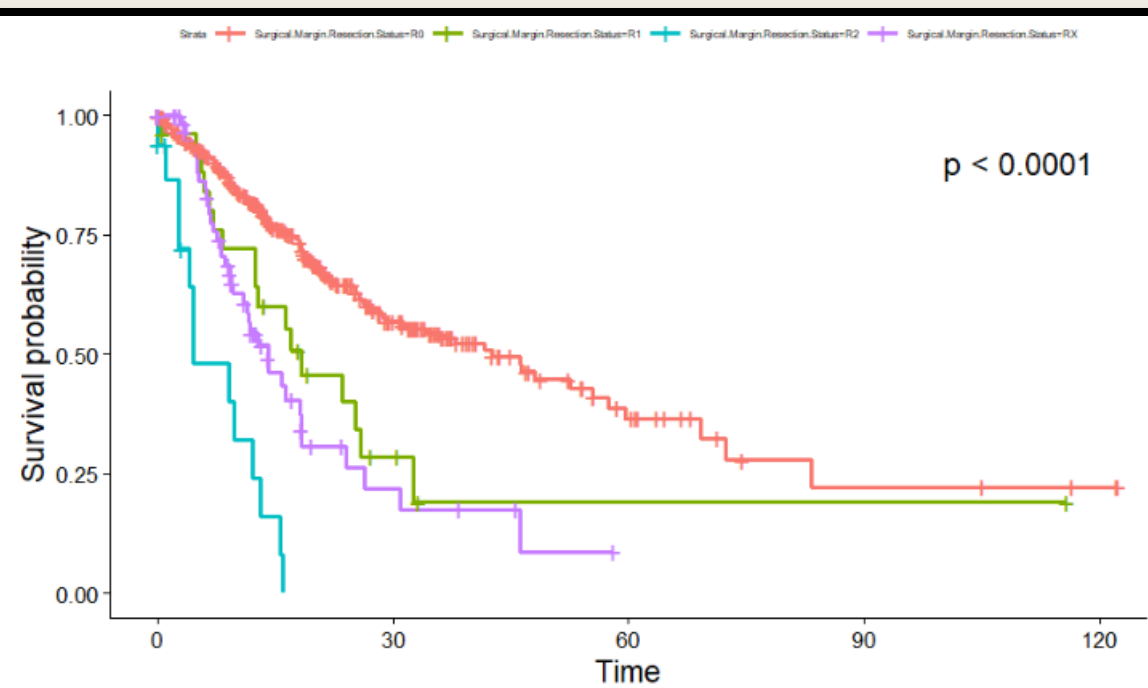
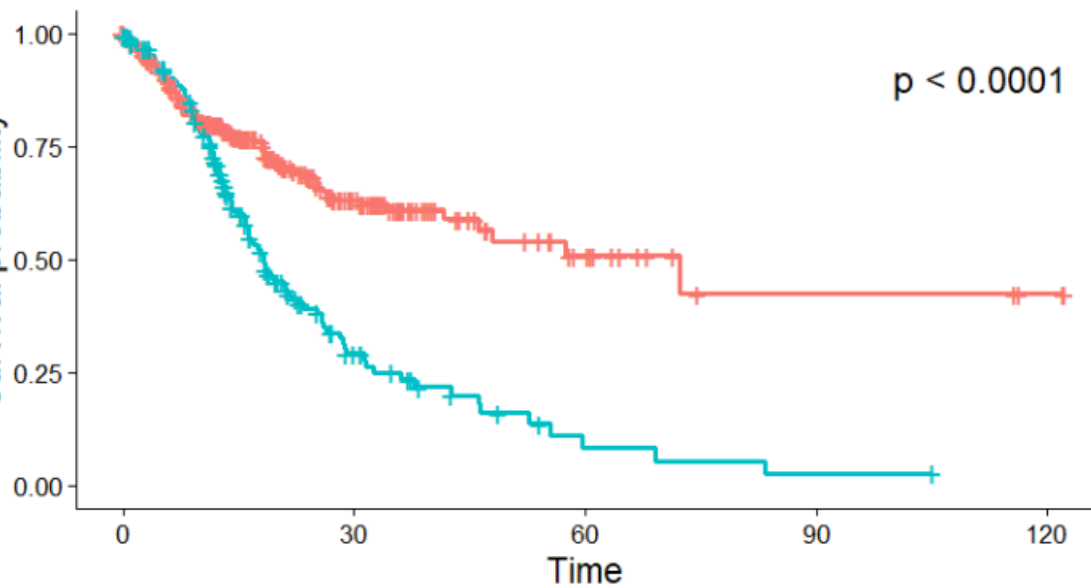
Number at risk

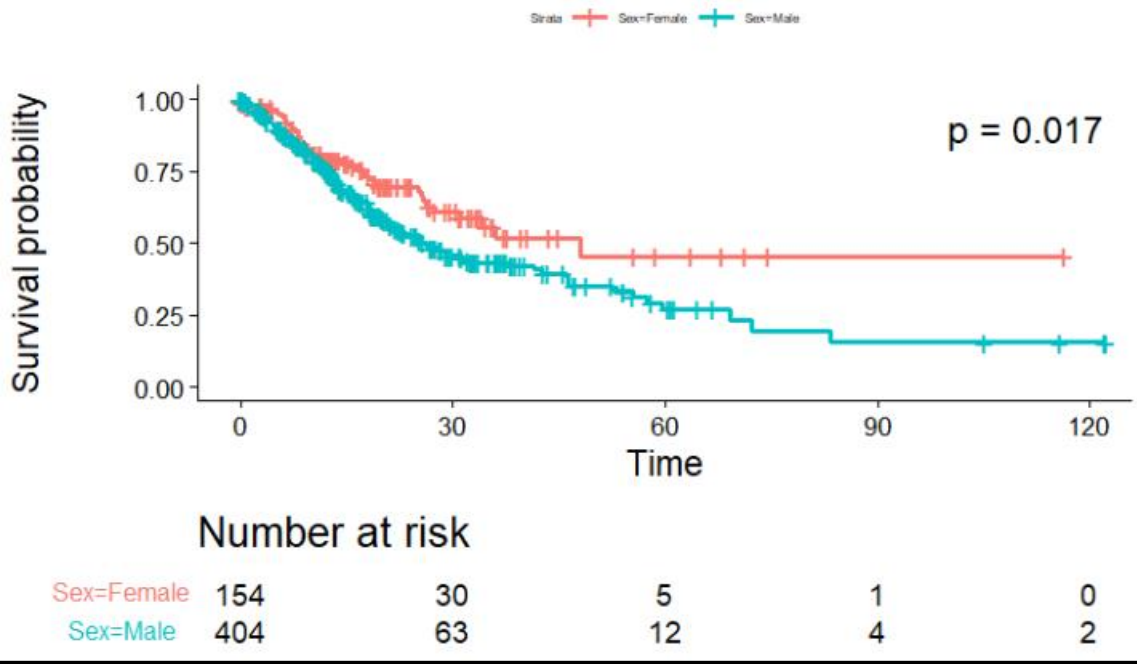
MLH1.Methylation=0	494	78	16	5	2
MLH1.Methylation=1	64	15	1	0	0



Survival probability

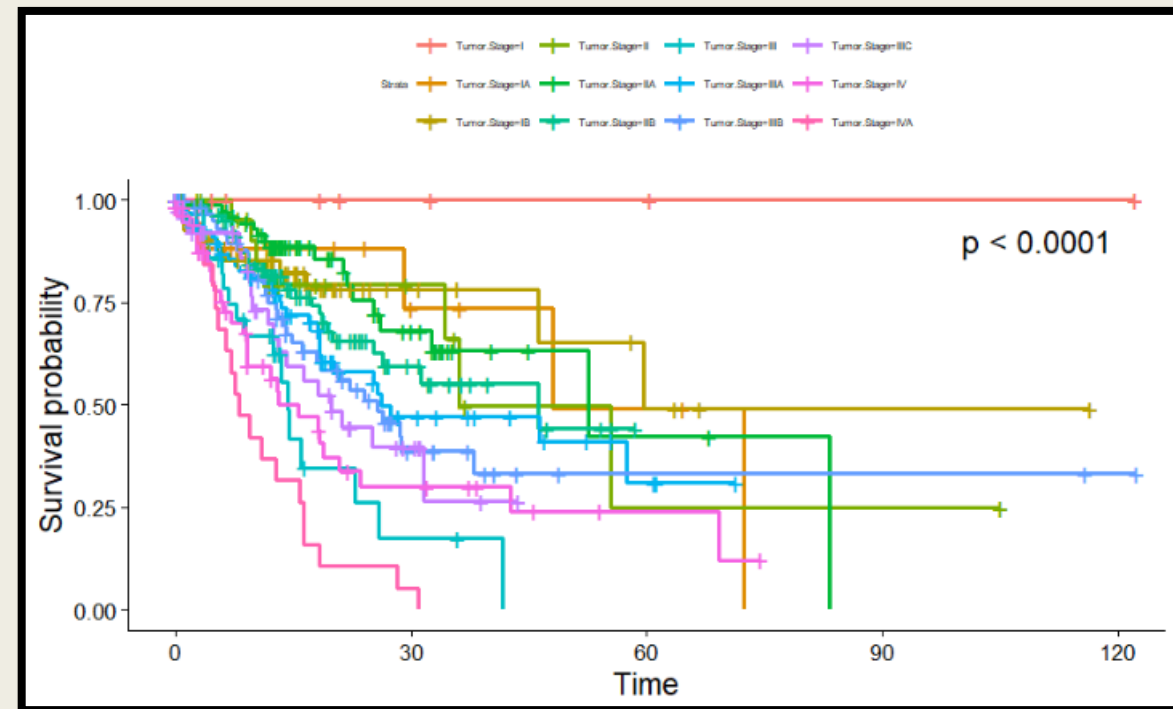
Strata + New.Neoplasm.Event.Post.Initial.Therapy.Indicator=NO + New.Neoplasm.Event.Post.Initial.Therapy.Indicator=Y





- 24個類別變數中，9個變數分不分組對危險函數有顯著差別
- Country : 國家 , (chr.)
- Genome.Doublings : 基因體倍增 , (int.)
- Neoplasm.Histologic.Grade : 腫瘤.組織學.分級 , (chr.)
- MLH1.Methylation : MLH1甲基化 , (logi.FALSE,TRUE)

- New.Neoplasm.Event.Post.Initial.Therapy.Indicator" : 新的腫瘤。事件.後.初始.治療.指標 , (chr,YES/NO)
- Person.Neoplasm.Cancer.Status : 人體腫瘤癌症狀態 , (chr.TUMOR_FREE/WITH_TUMOR)
- Surgical.Margin.Resection.Status : 手術邊緣切除狀態 , (chr.)
- Sex : 性別
- Tumor.Stage : 腫瘤階段 , (chr.)



Local test

看連續變數是否對食道癌生存有顯著影響

9個連續變數中，只有3個對食道癌的生存有顯著影響

Wald test :

- pvalue:0.4105631 > 0.05 , 不拒絕H0 , 係數顯著為0 , Absolute.Extract.Ploidy對食道癌生存沒有顯著影響
- pvalue:0.1194808 > 0.05 , 不拒絕H0 , 係數顯著為0 , Absolute.Extract.Purity對食道癌生存沒有顯著影響
- pvalue:0.01779964 < 0.05 , 拒絕H0 , 係數不顯著為0 , 年齡對食道癌生存有顯著影響

Likelihood ratio test :

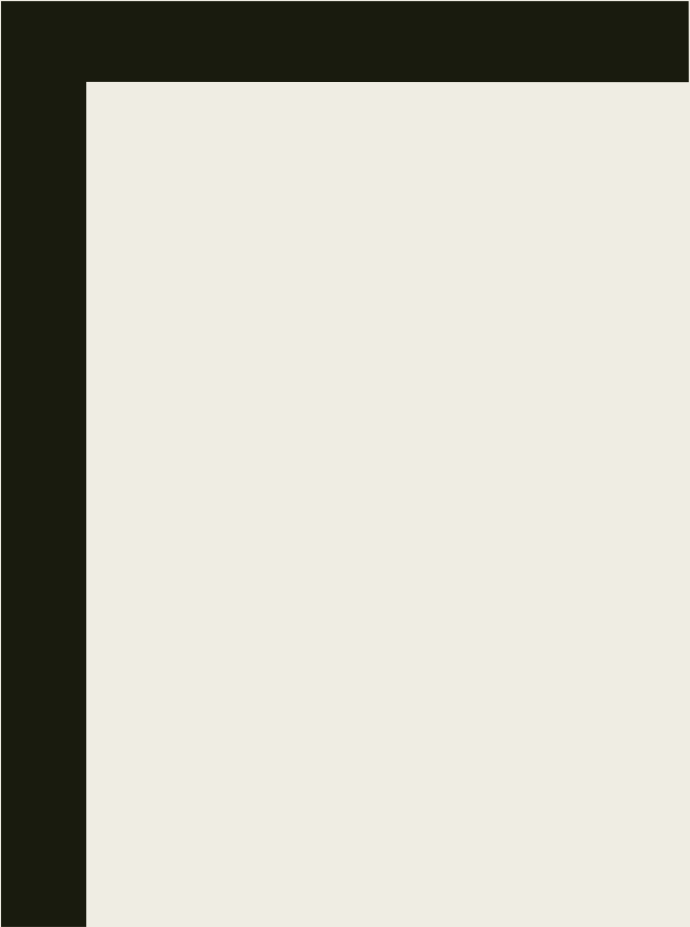
- p-value:0.0496991 < 0.05 , 拒絕H0 , 係數不顯著為0 , 白血球百分比對食道癌生存有顯著影響
- p-value:0.0483102 < 0.05 , 拒絕H0 , 係數不顯著為0 , 基因組改變比例對食道癌生存有顯著影響
- p-value:0.4504393 > 0.05 , 不拒絕H0 , 係數顯著為0 , 突變率對食道癌生存沒有顯著影響

Score test :

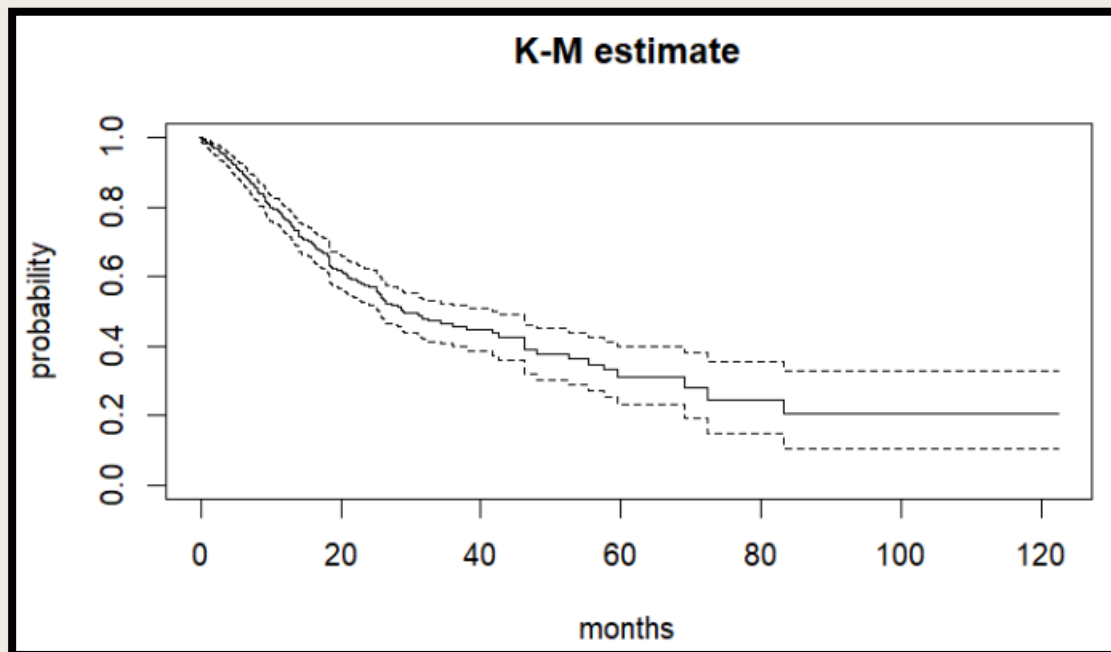
- p-value:0.7248046 > 0.05 , 不拒絕H0 , 係數顯著為0 , 癌細胞百分比對食道癌生存沒有顯著影響
- p-value:0.7152439 > 0.05 , 不拒絕H0 , 係數顯著為0 , 癌細胞核百分比對食道癌生存沒有顯著影響
- p-value:0.7928437 > 0.05 , 不拒絕H0 , 係數顯著為0 , 腫瘤組織中非同義突變的數量對食道癌生存沒有顯著影響

- 更新最後資料變數：DATA6，559 obs. of 14 variables

```
> str(data6)
'data.frame': 559 obs. of 14 variables:
 $ Diagnosis.Age : num 70 51.2 62.4 52.1 74.1 ...
 $ Country : chr "United_States" "Russia" "Russia" "Russia" ...
 $ Estimated.Leukocyte.Percentage : num 0.173 0.428 0.224 0.248 0.138 0.502 0.14 0.254 0.254 0.254 ...
 $ Fraction.Genome.Altered : num 0.523 0.0876 0.206 0.2297 0.4391 ...
 $ Genome.Doublings : num 0 1 0 0 1 0 1 0 1 0 ...
 $ Neoplasm.Histologic.Grade : chr "G2" "G3" "G3" "G3" ...
 $ MLH1.Methylation : num 0 1 0 0 1 0 0 0 0 0 ...
 $ New.Neoplasm.Event.Post.Initial.Therapy.Indicator: num 0 0 1 0 0 0 0 1 1 1 ...
 $ Overall.Survival..Months. : num 58.1 26.7 11.7 19.6 11 ...
 $ Overall.Survival.Status : num 0 0 0 0 0 0 0 0 0 0 ...
 $ Person.Neoplasm.Cancer.Status : num 0 0 0 0 0 0 0 1 1 1 ...
 $ Surgical.Margin.Resection.Status : chr "RX" "R0" "R0" "RX" ...
 $ Sex : chr "Male" "Female" "Male" "Male" ...
 $ Tumor.Stage : chr "I" "II" "I" "III" ...
```

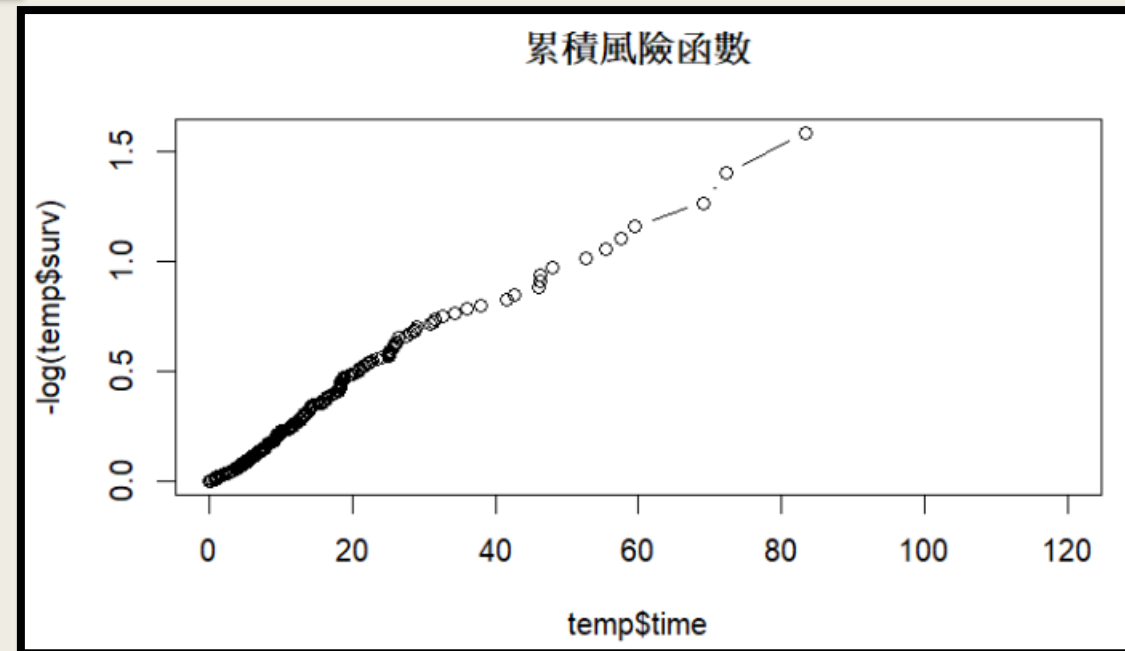


模型



Kaplan-Meier estimate & 累積風險函數

不考慮任何變數

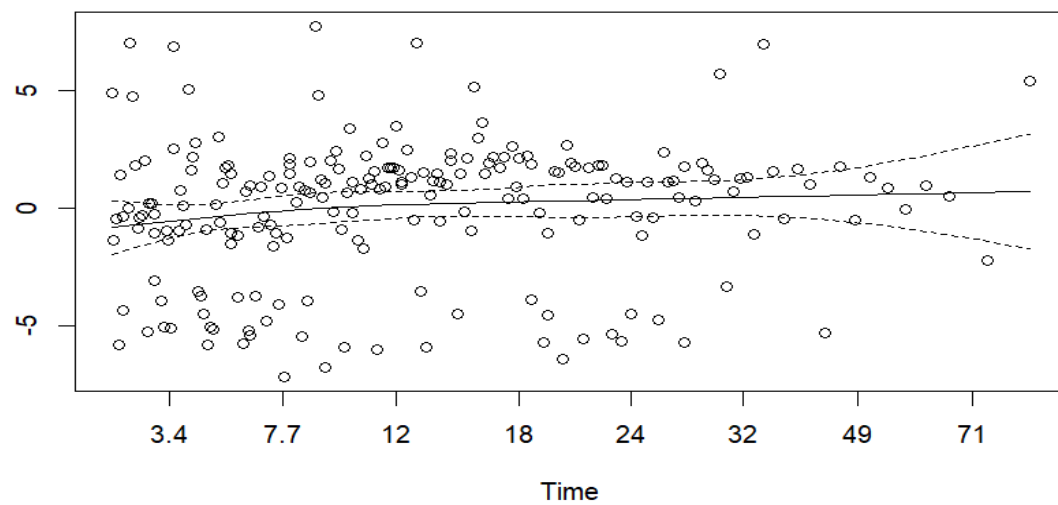


Cox Proportional Hazard (PH) Model

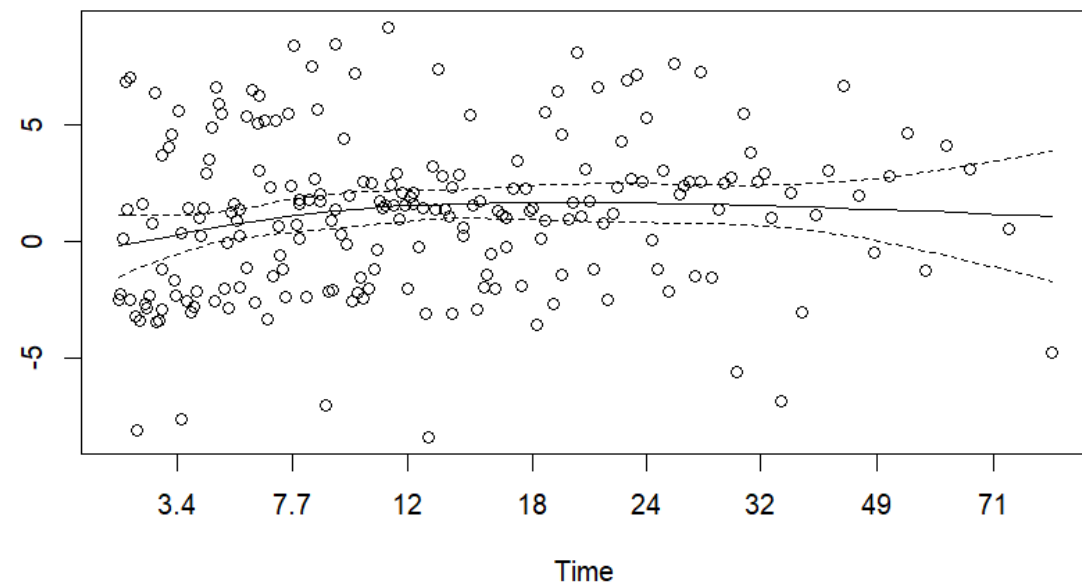
- Time varying cox model (Schoenfeld residuals test)
- H_0 : 係數隨時間變化
- New.Neoplasm.Event.Post.Initial.Therapy.Indicator , $p = 5.5e-05 < 0.05$, 變數係數隨時間變化
- Person.Neoplasm.Cancer.Status , $p = 9.3e-05 < 0.05$, 拒絕 H_0 , 變數係數隨時間變化

	chisq	df	p
Diagnosis.Age	2.10e-01	1	0.64713
Country	1.72e+01	12	0.14233
Estimated.Leukocyte.Percentage	2.89e-02	1	0.86506
Fraction.Genome.Altered	4.36e-01	1	0.50909
Genome.Doublings	3.61e-02	1	0.84925
Neoplasm.Histologic.Grade	4.22e+00	3	0.23846
MLH1.Methylation	8.34e-04	1	0.97697
New.Neoplasm.Event.Post.Initial.Therapy.Indicator	1.63e+01	1	5.5e-05
Person.Neoplasm.Cancer.Status	1.53e+01	1	9.3e-05
Surgical.Margin.Resection.Status	2.07e+00	3	0.55795
Sex	3.06e-03	1	0.95588
Tumor.Stage	2.73e+00	3	0.43528
GLOBAL	6.20e+01	29	0.00035

Beta(t) for New.Neoplasm.Event.Post.Initial.Therapy.Indicat



Beta(t) for Person.Neoplasm.Cancer.Status



最後模型 : Time Dependent Covariates + AIC 自動挑選變數

```
Call:
coxph(formula = Surv(Overall.Survival..Months., Overall.Survival.Status) ~
      Diagnosis.Age + Country + Genome.Doublings + Neoplasm.Histologic.Grade +
      New.Neoplasm.Event.Post.Initial.Therapy.Indicator + Surgical.Margin.Resection.Status +
      Tumor.Stage + tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator) +
      tt(Person.Neoplasm.Cancer.Status), data = data6, tt = function(x,
t, ...) x * log(t + 1))
```

	coef	exp(coef)	se(coef)	z	p
Diagnosis.Age	1.711e-02	1.017e+00	7.145e-03	2.394	0.0167
CountryBrazil	1.502e-01	1.162e+00	4.142e-01	0.363	0.7169
CountryCanada	-9.570e-01	3.841e-01	6.917e-01	-1.384	0.1665
CountryGermany	-4.507e-01	6.372e-01	4.902e-01	-0.919	0.3579
CountryKorea_South	-1.567e+00	2.087e-01	1.072e+00	-1.462	0.1439
CountryMoldova	-1.309e+01	2.071e-06	1.315e+03	-0.010	0.9921
CountryNetherlands	8.635e-01	2.371e+00	5.117e-01	1.688	0.0915
CountryPoland	-4.931e-02	9.519e-01	4.866e-01	-0.101	0.9193
CountryRussia	4.547e-01	1.576e+00	4.418e-01	1.029	0.3033
CountryUkraine	2.821e-01	1.326e+00	4.627e-01	0.610	0.5420
CountryUnited_Kingdom	7.925e-01	2.209e+00	8.070e-01	0.982	0.3261
CountryUnited_States	1.207e-01	1.128e+00	4.230e-01	0.285	0.7754
CountryVietnam	6.005e-01	1.823e+00	4.572e-01	1.313	0.1890
Genome.Doublings	2.701e-01	1.310e+00	1.104e-01	2.445	0.0145
Neoplasm.Histologic.GradeG2	2.352e-01	1.265e+00	4.155e-01	0.566	0.5713
Neoplasm.Histologic.GradeG3	6.148e-01	1.849e+00	4.115e-01	1.494	0.1352
Neoplasm.Histologic.GradeGX	1.248e+00	3.482e+00	4.944e-01	2.523	0.0116
New.Neoplasm.Event.Post.Initial.Therapy.Indicator	-1.133e+00	3.221e-01	4.760e-01	-2.380	0.0173
Surgical.Margin.Resection.StatusR1	3.442e-02	1.035e+00	3.122e-01	0.110	0.9122
Surgical.Margin.Resection.StatusR2	1.478e+00	4.382e+00	3.503e-01	4.217	2.47e-05
Surgical.Margin.Resection.StatusRX	6.301e-01	1.878e+00	2.544e-01	2.477	0.0133
Tumor.StageII	-6.835e-02	9.339e-01	3.104e-01	-0.220	0.8257
Tumor.StageIII	5.291e-01	1.697e+00	3.003e-01	1.762	0.0781
Tumor.StageIV	7.562e-01	2.130e+00	3.298e-01	2.293	0.0218
tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator)	4.635e-01	1.590e+00	1.950e-01	2.377	0.0175
tt(Person.Neoplasm.Cancer.Status)	4.638e-01	1.590e+00	9.057e-02	5.122	3.03e-07

Likelihood ratio test=207.4 on 26 df, p=< 2.2e-16
n= 559, number of events= 211



分析結果、解釋



$$h(t|z) = h_0(t)e^{0.017 \times \text{Diagnosis.Age} + 0.1501 \times \text{CountryBrazil} - 0.957 \times \text{CountryCanada} - 0.451 \times \text{CountryGermany} - 1.567 \times \text{CountryKoreasouth} - 13.087 \times \text{CountryMoldova} + 0.864 \times \text{CountryNetherlands} + \dots + 0.4635 \times \text{tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator)}}$$

主要挑有顯著的變數做解釋

	coef	exp(coef)	se(coef)	z	p-value
:---	:---:	:---:	:---:	:---:	:---:
Diagnosis.Age	1.711e-02	1.017e+00	7.145e-03	2.394	0.0167
Genome.Doublings	2.701e-01	1.310e+00	1.104e-01	2.445	0.0145
Neoplasm.Histologic.GradeGX	1.248e+00	3.482e+00	4.944e-01	2.523	0.0116
New.Neoplasm.Event.Post.Initial.Therapy.Indicator	-1.133e+00	3.221e-01	4.760e-01	-2.380	0.0173
Surgical.Margin.Resection.StatusR2	1.478e+00	4.382e+00	3.503e-01	4.217	2.47e-05
Surgical.Margin.Resection.StatusRX	6.301e-01	1.878e+00	2.544e-01	2.477	0.0133
Tumor.StageIV	7.562e-01	2.130e+00	3.298e-01	2.293	0.0218
tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator)	4.635e-01	1.590e+00	1.950e-01	2.377	0.0175

+ 診斷年齡每增加一歲，食道癌死亡的風險增加\$1.017\$倍

+ 基因體倍增(細胞的染色體數目增加為正常倍數的兩倍)1，食道癌死亡的風險增加\$1.310\$倍

+ 基因體倍增(細胞的染色體數目增加為正常倍數的四倍)2，食道癌死亡的風險增加\$1.310^2\$倍

+ 腫瘤分級GX(無法判斷)比分級G1(輕微)，食道癌死亡的風險增加\$3.482\$倍

+ **初次治療後出現新的腫瘤事件比沒事的，食道癌死亡的風險增加\$0.3221 \times t^{1.590}\$倍**

+ 當時間越久，t越大，出現事件比沒事的食道癌死亡的風險，\$0.3221 \times t^{1.590}\$，越來越大

+ R0：原發腫瘤部位顯微鏡下未見癌細胞

+ R1：癌細胞在顯微鏡下出現在原發腫瘤部位

+ R2：區域淋巴結處有殘留腫瘤

+ 手術邊緣切除狀態 R2(區域淋巴結處有殘留腫瘤)比 R0，食道癌死亡的風險增加\$4.382\$倍

+ 手術邊緣切除狀態 RX(無法評估)比 R0，食道癌死亡的風險增加\$1.878\$倍

+ 腫瘤階段IV比階段I(輕微)，食道癌死亡的風險增加\$2.130\$倍

結論

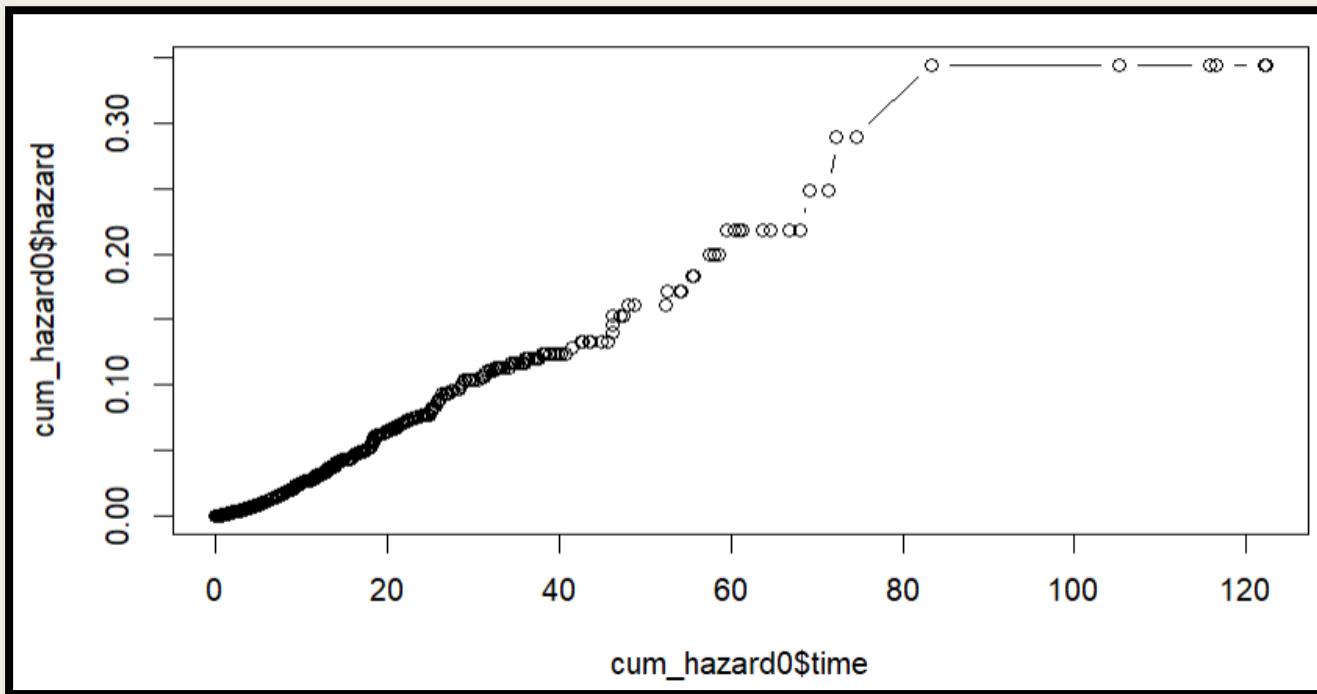
資料中有很多基因相關變數，
但各基因主要影響的器官不同，
雖然基因突變容易造成癌症，
但能直接影響食道癌的還是需要再進一步分析才能發現

Model Check

Check的模型為：Data6，559 obs. of 14 variables

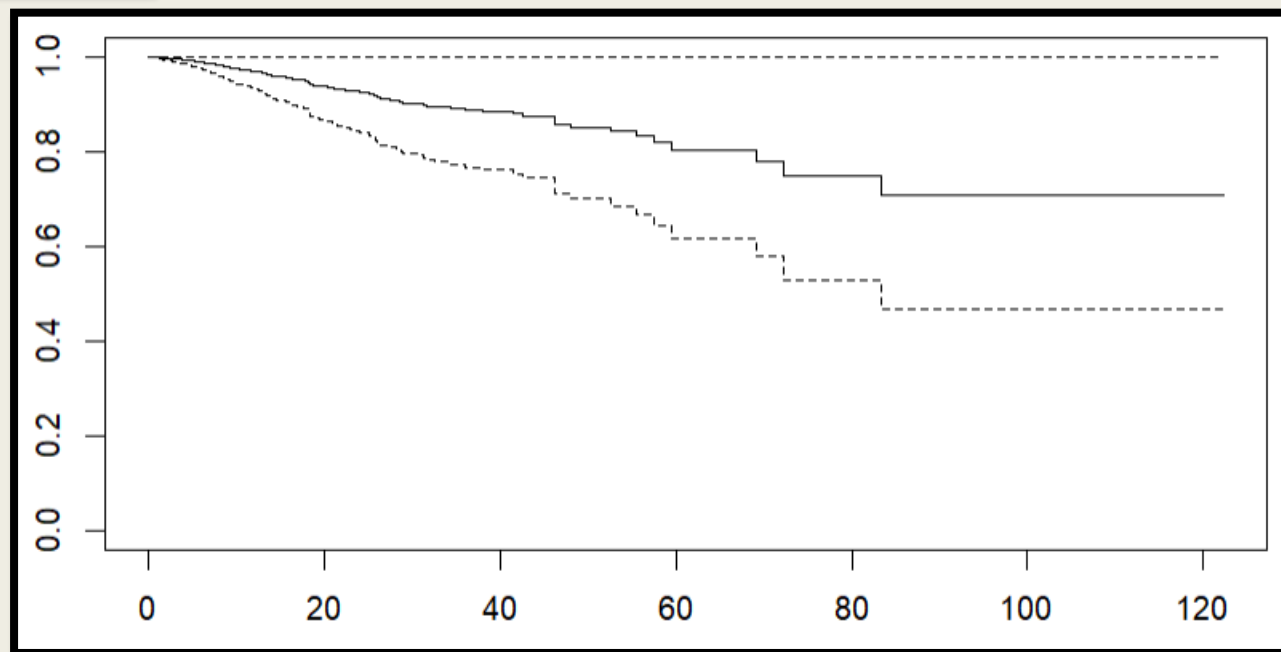
	coef	exp(coef)	se(coef)	z	Pr(> z)	
Diagnosis.Age	1.808e-02	1.018e+00	7.351e-03	2.460	0.01391	*
CountryBrazil	2.398e-01	1.271e+00	4.316e-01	0.556	0.57848	.
CountryCanada	-8.436e-01	4.301e-01	7.014e-01	-1.203	0.22908	.
CountryGermany	-4.149e-01	6.604e-01	5.016e-01	-0.827	0.40805	.
CountryKorea_South	-1.500e+00	2.232e-01	1.086e+00	-1.381	0.16730	.
CountryMoldova	-1.314e+01	1.967e-06	1.317e+03	-0.010	0.99204	.
CountryNetherlands	9.758e-01	2.653e+00	5.348e-01	1.825	0.06804	.
CountryPoland	-3.476e-03	9.965e-01	4.968e-01	-0.007	0.99442	.
CountryRussia	4.549e-01	1.576e+00	4.485e-01	1.014	0.31037	.
CountryUkraine	2.777e-01	1.320e+00	4.661e-01	0.596	0.55140	.
CountryUnited_Kingdom	8.566e-01	2.355e+00	8.133e-01	1.053	0.29220	.
CountryUnited_States	1.592e-01	1.173e+00	4.313e-01	0.369	0.71210	.
CountryVietnam	6.382e-01	1.893e+00	4.656e-01	1.371	0.17050	.
Estimated.Leukocyte.Percentage	-6.688e-01	5.123e-01	7.369e-01	-0.908	0.36409	.
Fraction.Genome.Altered	-8.142e-01	4.430e-01	5.306e-01	-1.534	0.12493	.
Genome.Doublings	3.340e-01	1.397e+00	1.222e-01	2.734	0.00626	**
Neoplasm.Histologic.GradeG2	2.263e-01	1.254e+00	4.199e-01	0.539	0.58986	.
Neoplasm.Histologic.GradeG3	6.107e-01	1.842e+00	4.182e-01	1.460	0.14423	.
Neoplasm.Histologic.GradeGX	1.224e+00	3.402e+00	5.042e-01	2.429	0.01515	*
MLH1.Methylation	-1.240e-01	8.834e-01	2.853e-01	-0.434	0.66398	.
New.Neoplasm.Event.Post.Initial.Therapy.Indicator	-1.004e+00	3.662e-01	5.790e-01	-1.735	0.08278	.
Person.Neoplasm.Cancer.Status	-1.701e-01	8.435e-01	5.718e-01	-0.298	0.76604	.
Surgical.Margin.Resection.StatusR1	1.955e-02	1.020e+00	3.132e-01	0.062	0.95023	.
Surgical.Margin.Resection.StatusR2	1.495e+00	4.459e+00	3.635e-01	4.113	3.91e-05	***
Surgical.Margin.Resection.StatusRX	6.348e-01	1.887e+00	2.598e-01	2.443	0.01455	*
SexMale	-4.893e-02	9.523e-01	1.887e-01	-0.259	0.79546	.
Tumor.StageII	-6.411e-02	9.379e-01	3.108e-01	-0.206	0.83657	.
Tumor.StageIII	5.305e-01	1.700e+00	3.012e-01	1.761	0.07824	.
Tumor.StageIV	7.852e-01	2.193e+00	3.321e-01	2.364	0.01807	*
tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator)	4.227e-01	1.526e+00	2.314e-01	1.827	0.06770	.
tt(Person.Neoplasm.Cancer.Status)	5.201e-01	1.682e+00	2.286e-01	2.275	0.02290	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

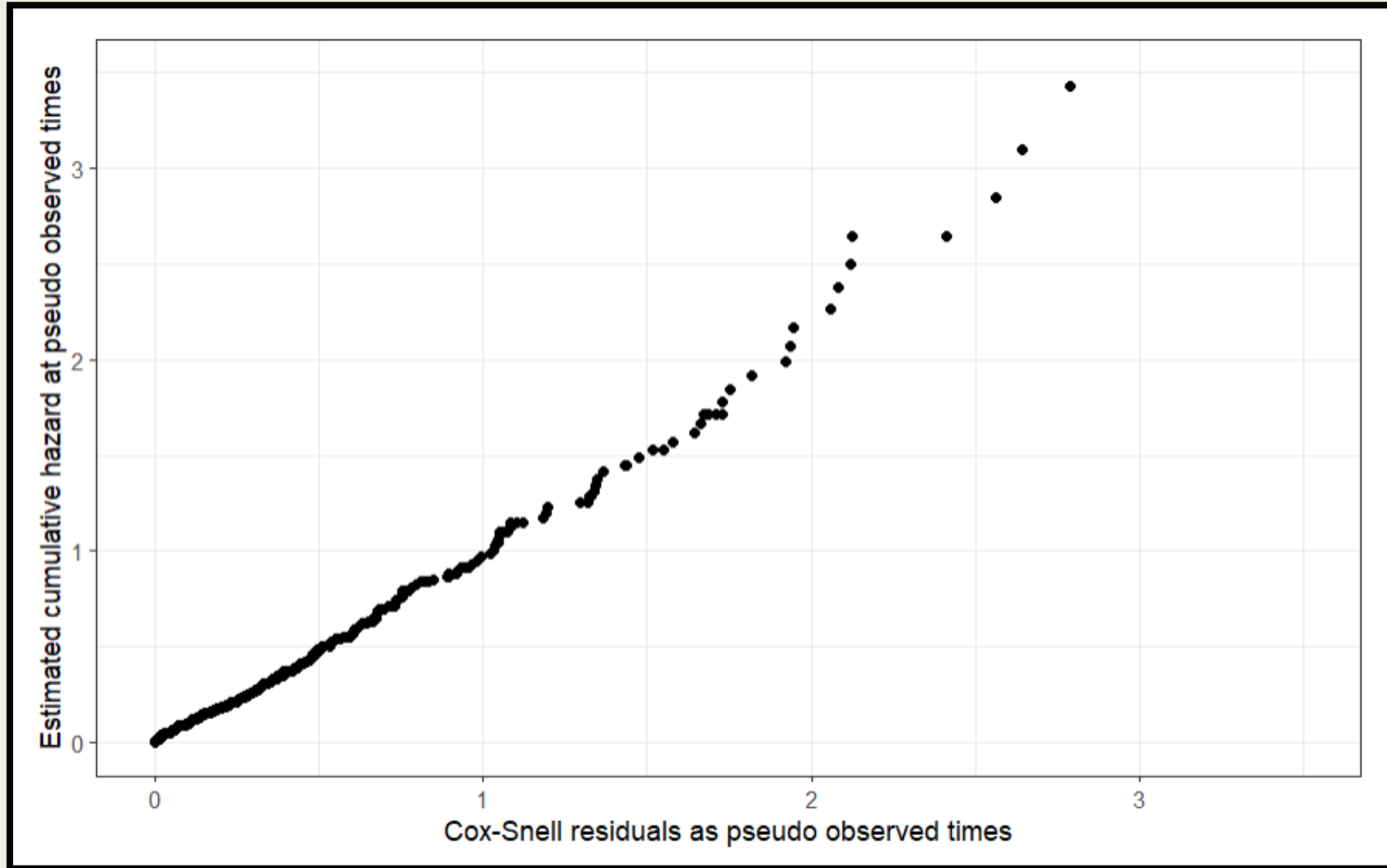


Cox baseline cumulative hazard function
&
Kaplan-Meier estimate

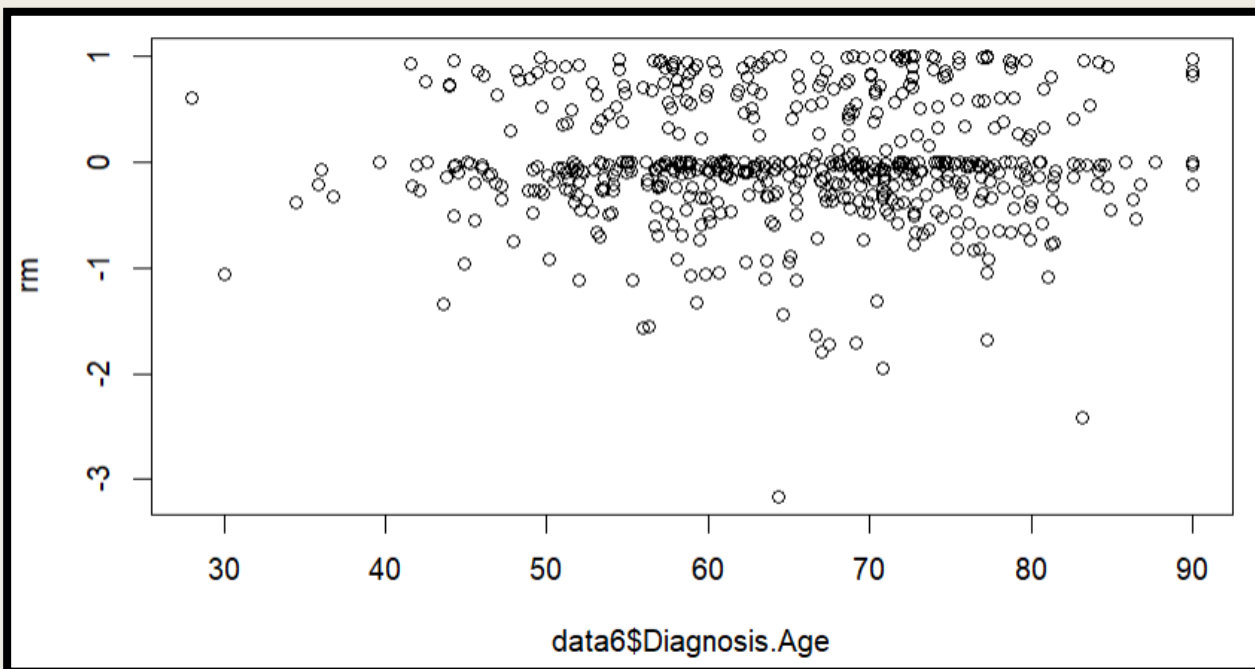
考慮所有變數



Cox snell's residual

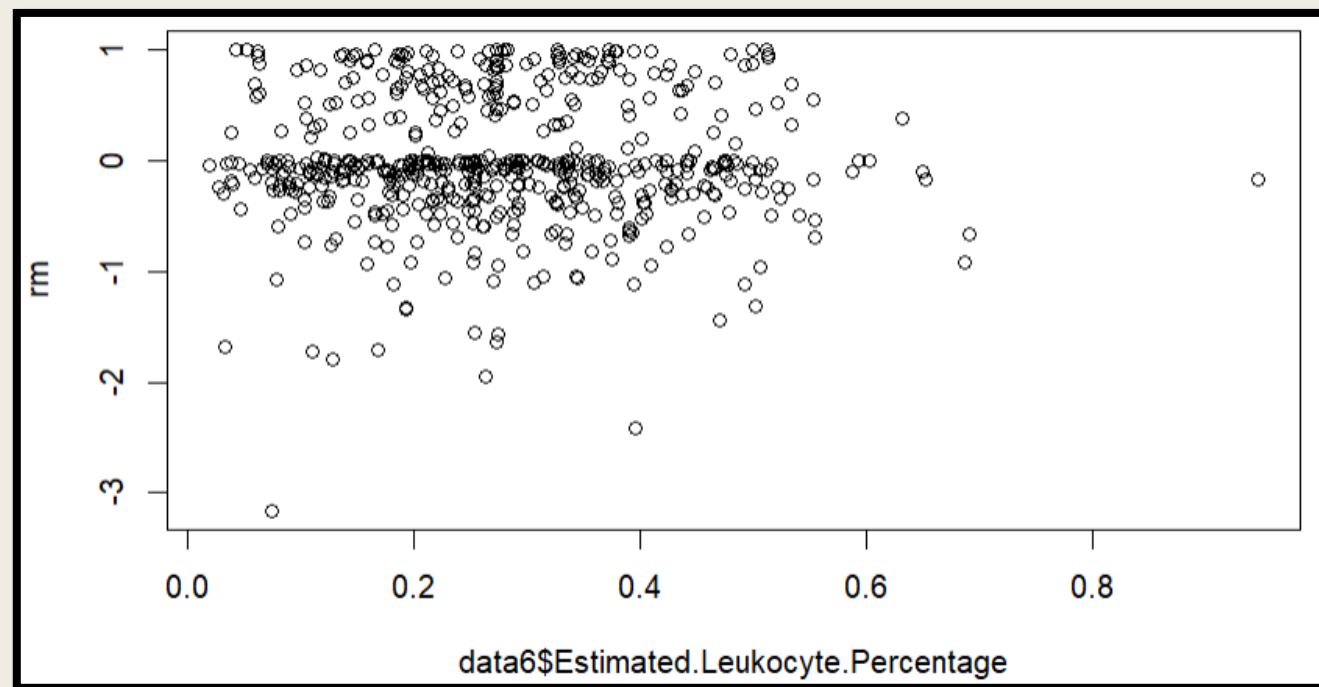


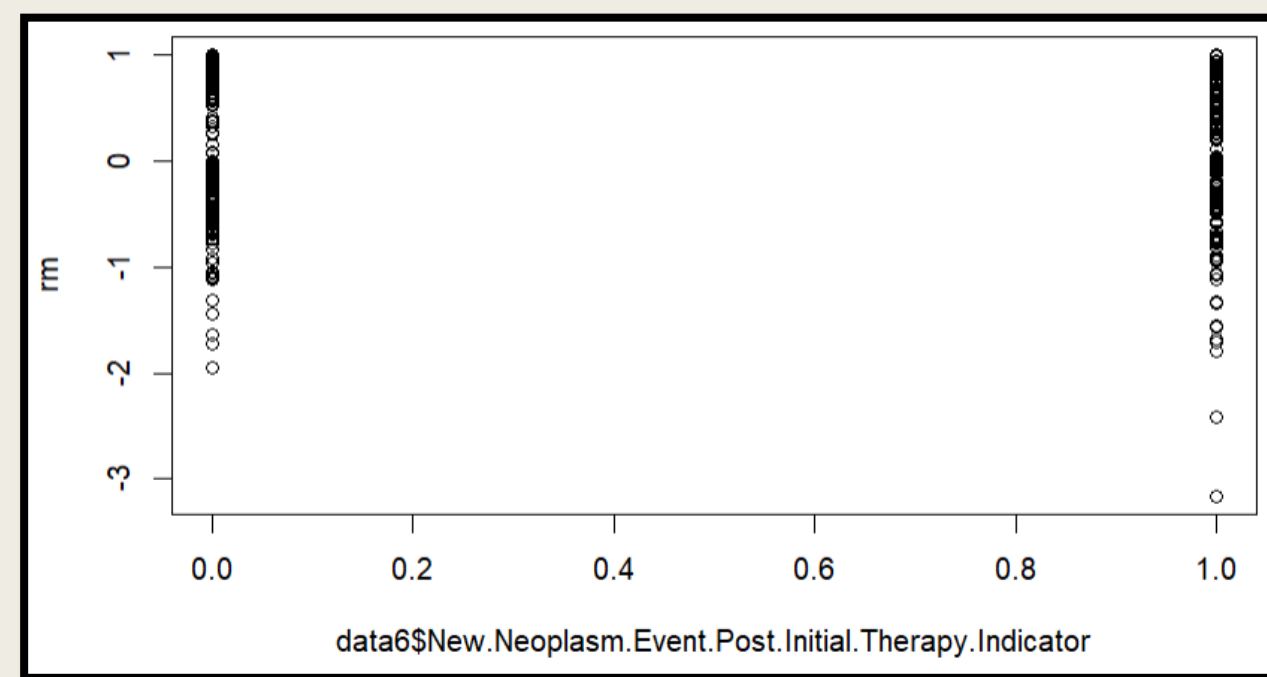
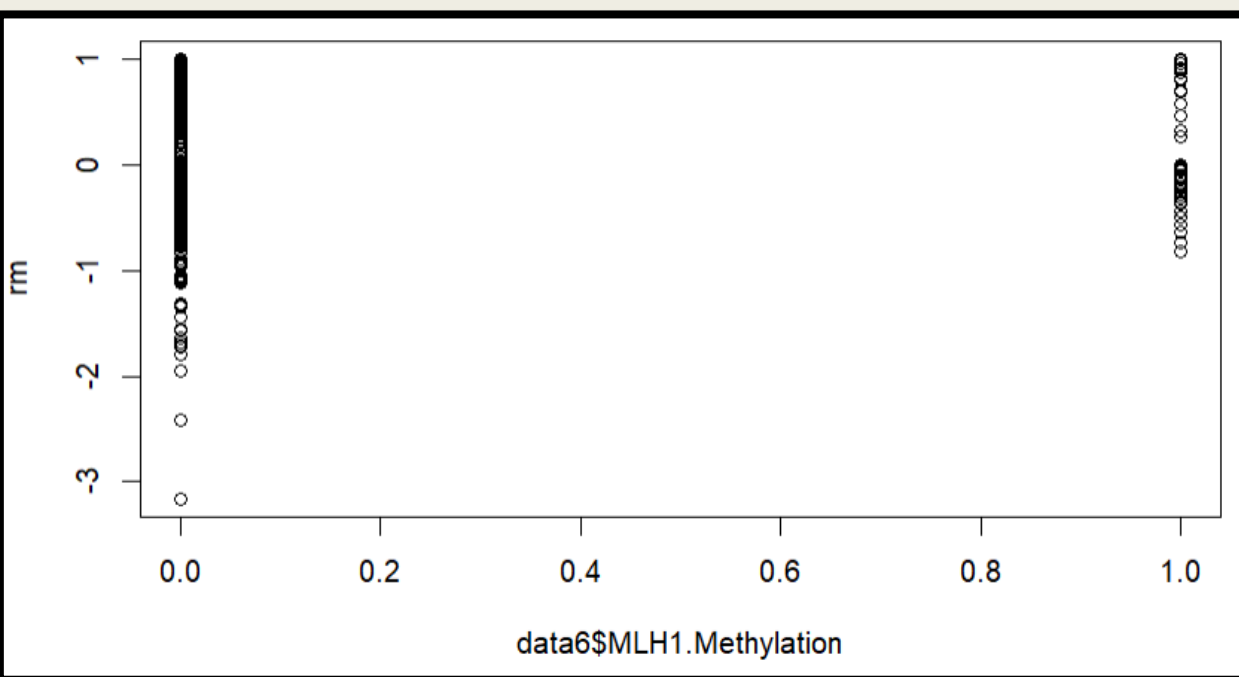
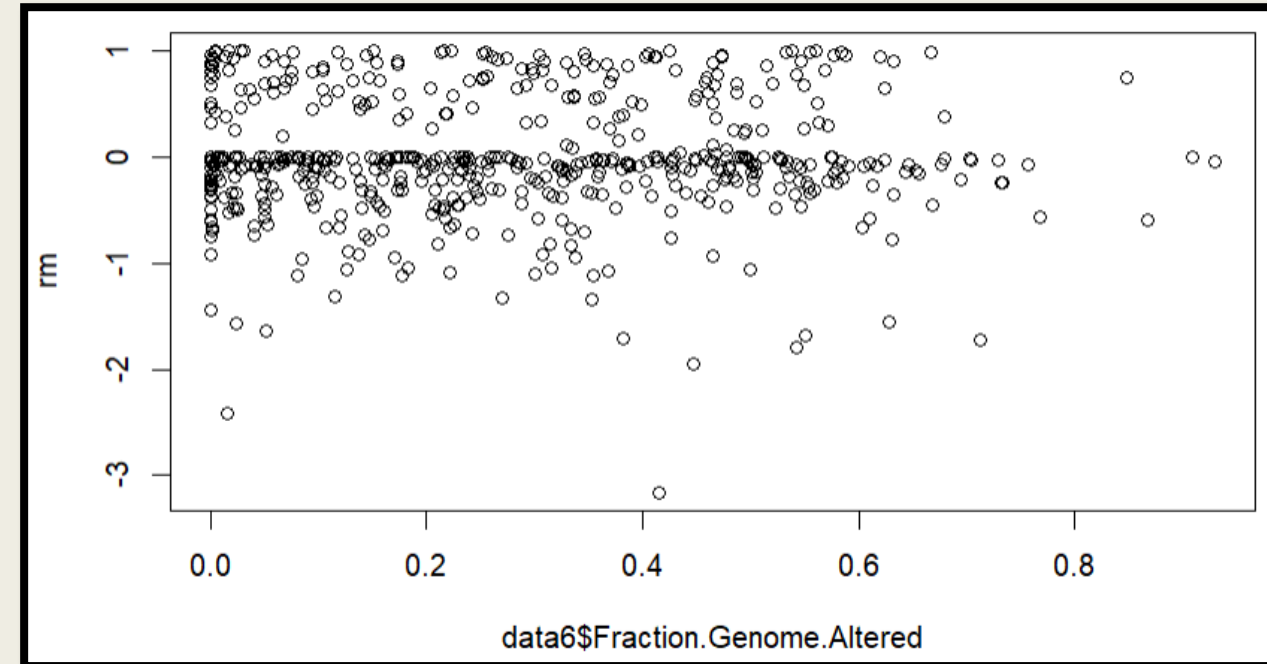
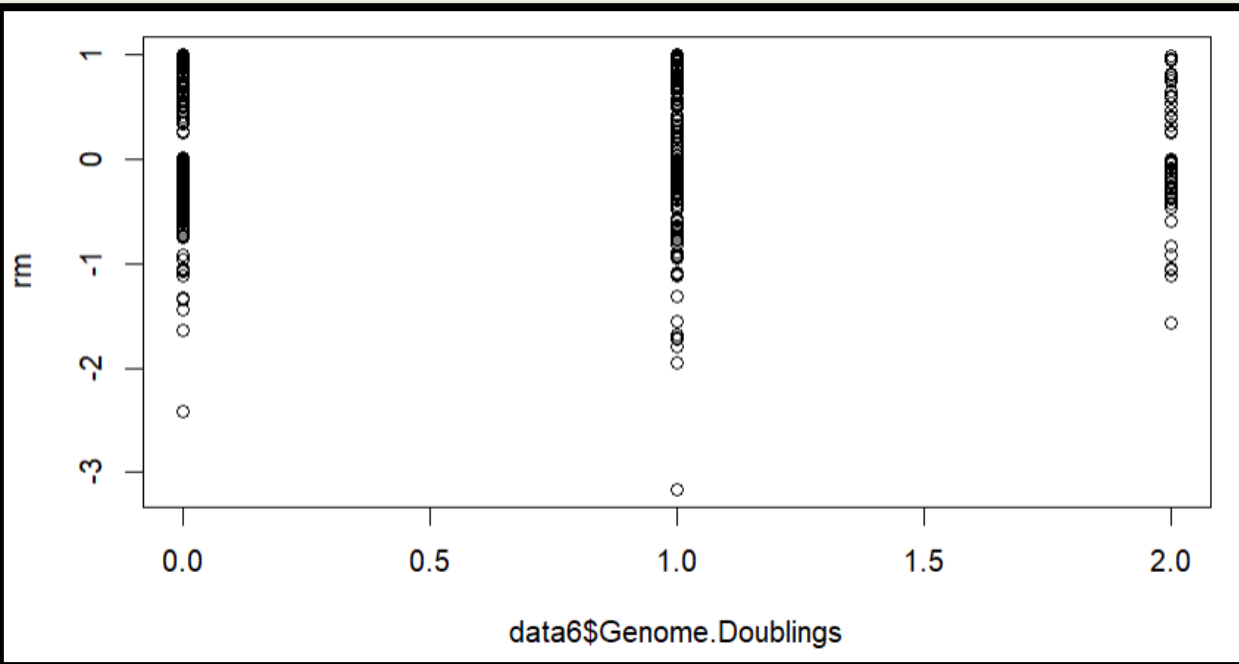
用(snell殘差, event)估計累積風險函數，接近45度線，模型應該足夠



Martingale residual

對連續變數做殘插圖，
看起來皆沒有特別的趨勢，模型可能足夠





參考資料

2023/05/29

- <https://www.rdocumentation.org/packages/VIM/versions/6.2.2/topics/aggr>
- <https://blog.yjtseng.info/post/2020-05-13-survival-curve/>
- https://rpubs.com/kaz_yos/resid_cox
- [step function – Rdocumentation](#)
- 資料來源：
- https://www.cbioportal.org/study/clinicalData?id=stes_tcga_pub