

110年衛生福利資料科學中心

R軟體推廣課程《進階篇》

存活分析

講師：劉品崧 統計分析師

花蓮慈濟醫院

課程大綱

- 存活分析的應用情境與資料結構
- 無母數分析模型
- KM plot繪圖技巧
- Cox proportional hazard model
- 衍伸議題討論

存活分析的應用情境與資料結構

- Time, event, or time-to-event
- 範例資料：pneumonia.fst

Time, event, or time-to-event

- 到底用要什麼觀察觀點？

治療組	受試者編號	存活年數	觀察結果
現行藥物	A1	4	0
	A2	6	1
	A3	3	0
	A4	7	1
	A5	6	1
新型藥物	B1	3	0
	B2	5	1
	B3	3	1
	B4	6	0
	B5	6	0

觀察結果：0 = 存活；1 = 死亡

比較平均存活時間

現行藥物

$$= \frac{4 + 6 + 3 + 7 + 6}{5} = 5.2 \text{年}$$

新型藥物

$$= \frac{3 + 5 + 3 + 6 + 6}{5} = 4.6 \text{年}$$

比較死亡比例

現行藥物

$$= \left(\frac{0 + 1 + 0 + 1 + 1}{5} \right) * 100\% = 60\%$$

新型藥物

$$= \left(\frac{0 + 1 + 1 + 0 + 0}{5} \right) * 100\% = 40\%$$

Time, event, or time-to-event

- 到底用要什麼觀察觀點？

治療組	受試者編號	存活年數	觀察結果
現行藥物	A1	4	0
	A2	6	1
	A3	3	0
	A4	7	1
	A5	6	1
新型藥物	B1	3	0
	B2	5	1
	B3	3	1
	B4	6	0
	B5	6	0

觀察結果：0 = 存活；1 = 死亡

比較死亡事件發生率

現行藥物

$$= \left(\frac{0 + 1 + 0 + 1 + 1}{4 + 6 + 3 + 7 + 6} \right) * 100\% = 11.5\%$$

新型藥物

$$= \left(\frac{0 + 1 + 1 + 0 + 0}{3 + 5 + 3 + 6 + 6} \right) * 100\% = 8.7\%$$

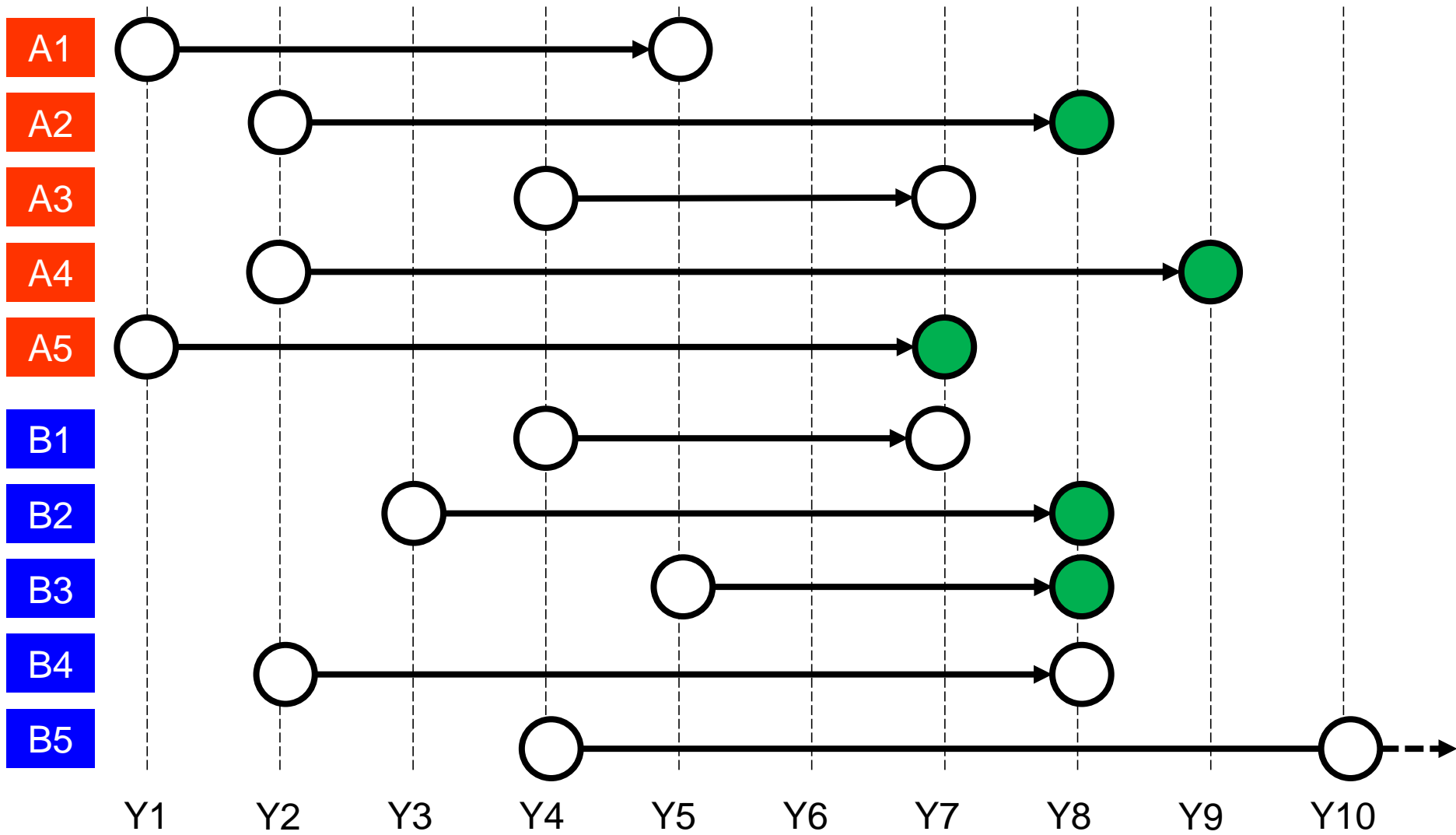
存活分析的應用情境與特性

- 存活分析方法是用来研究或分析樣本所觀察到的某一段時間查度之分配。一段時間長度通常是從一特定事件起始之時間原點(time zero, index date)直到某一特定事件發生的時間點。並非所有研究對象，其被追蹤(follow-up)的時間都足夠長，這樣的資料通常被稱為設限資料(censored data)。
- 重點：觀察多久 & 觀察結果是什麼

林建甫(2020)存活分析

研究樣本的追蹤狀況

● 死亡
○ 設限



觀察年數	觀察結果
4	0
6	1
3	0
7	1
6	1
3	0
5	1
3	1
6	0
6	0

肺炎住院病人的存活分析

- Scenario

- 2014年6月住院且具有肺炎診斷(480.xx ~ 486.xx)的病人
- 出院日為index date

- Primary outcome

- 出院後再次因為肺炎而入住院治療
- 若遇到死亡，或是2014.12.31仍然存活，則進行設限(censored)

- Secondary outcome

- 出院後死亡，若到2014.12.31仍然存活，則進行設限(censored)

範例資料探索：pneumonia.fst

讀取範例資料

```
pneumonia <- read_fst("pneumonia.fst", as.data.table = T)
```

```
head(pneumonia, 20) # 前20筆資料
```

```
summary(pneumonia) # 整體摘要性統計
```

```
summary(pneumonia$age) # 數值型
```

```
table(pneumonia$agegp) # 類別型，0為65歲以下，1為65歲(含)以上
```

無母數分析模型

- 生命表的計算
- Kaplan-Meier estimator
- Nelson-Aalen estimator
- 比較存活函數的差異

生命表的計算

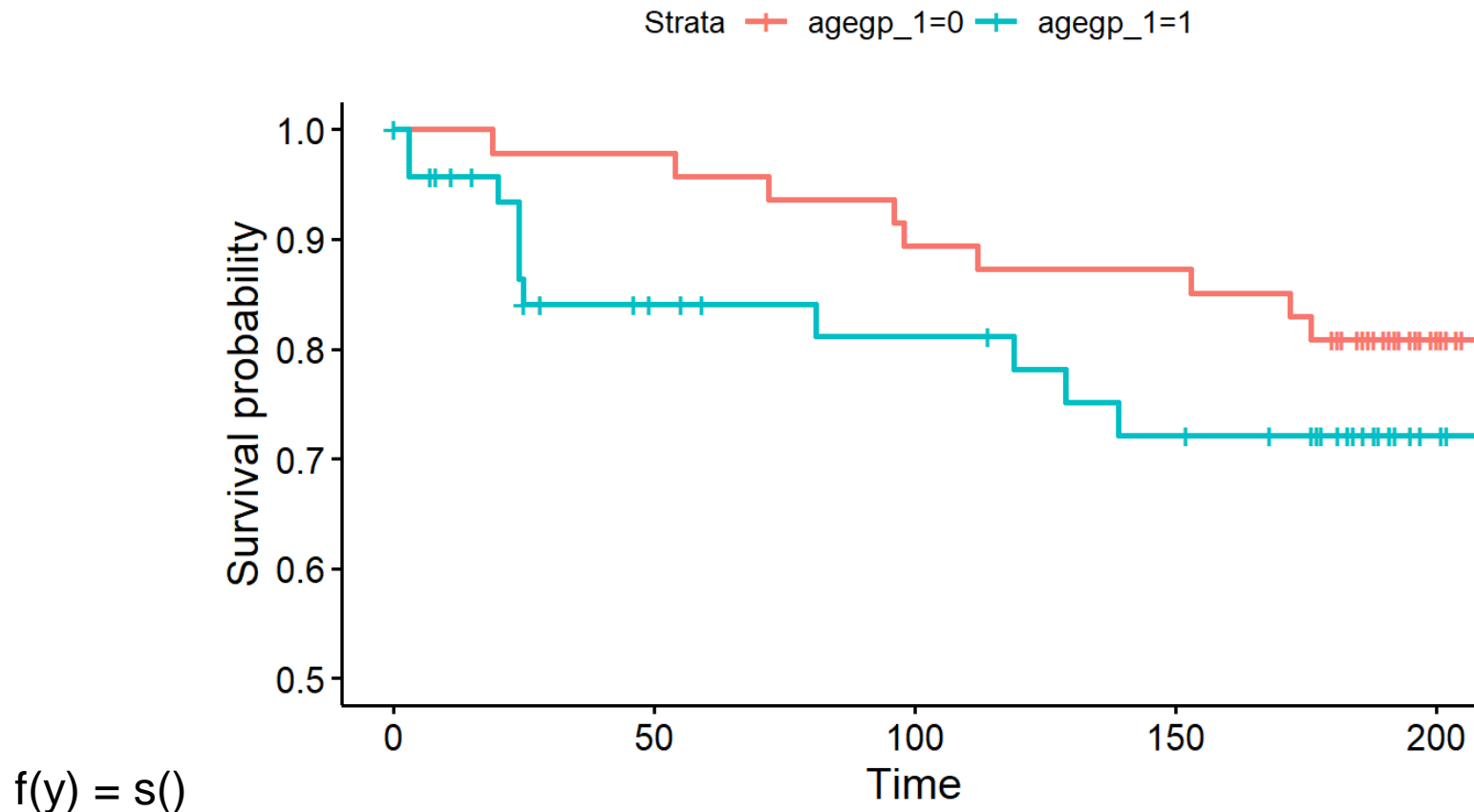
- 使用仍存活的樣本(risk set)計算條件存活機率

```
> summary(survfit(Surv(re_adm_ft, re_adm) ~ 1, data = pneumonia))  
Call: survfit(formula = Surv(re_adm_ft, re_adm) ~ 1, data = pneumonia)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
3	94	2	0.979	0.0149	0.950	1.000
19	88	1	0.968	0.0184	0.932	1.000
20	87	1	0.956	0.0213	0.916	0.999
24	86	3	0.923	0.0279	0.870	0.980
25	83	1	0.912	0.0297	0.856	0.972
54	77	1	0.900	0.0316	0.840	0.964
72	74	1	0.888	0.0334	0.825	0.956
81	73	1	0.876	0.0351	0.810	0.947
96	72	1	0.864	0.0367	0.795	0.939

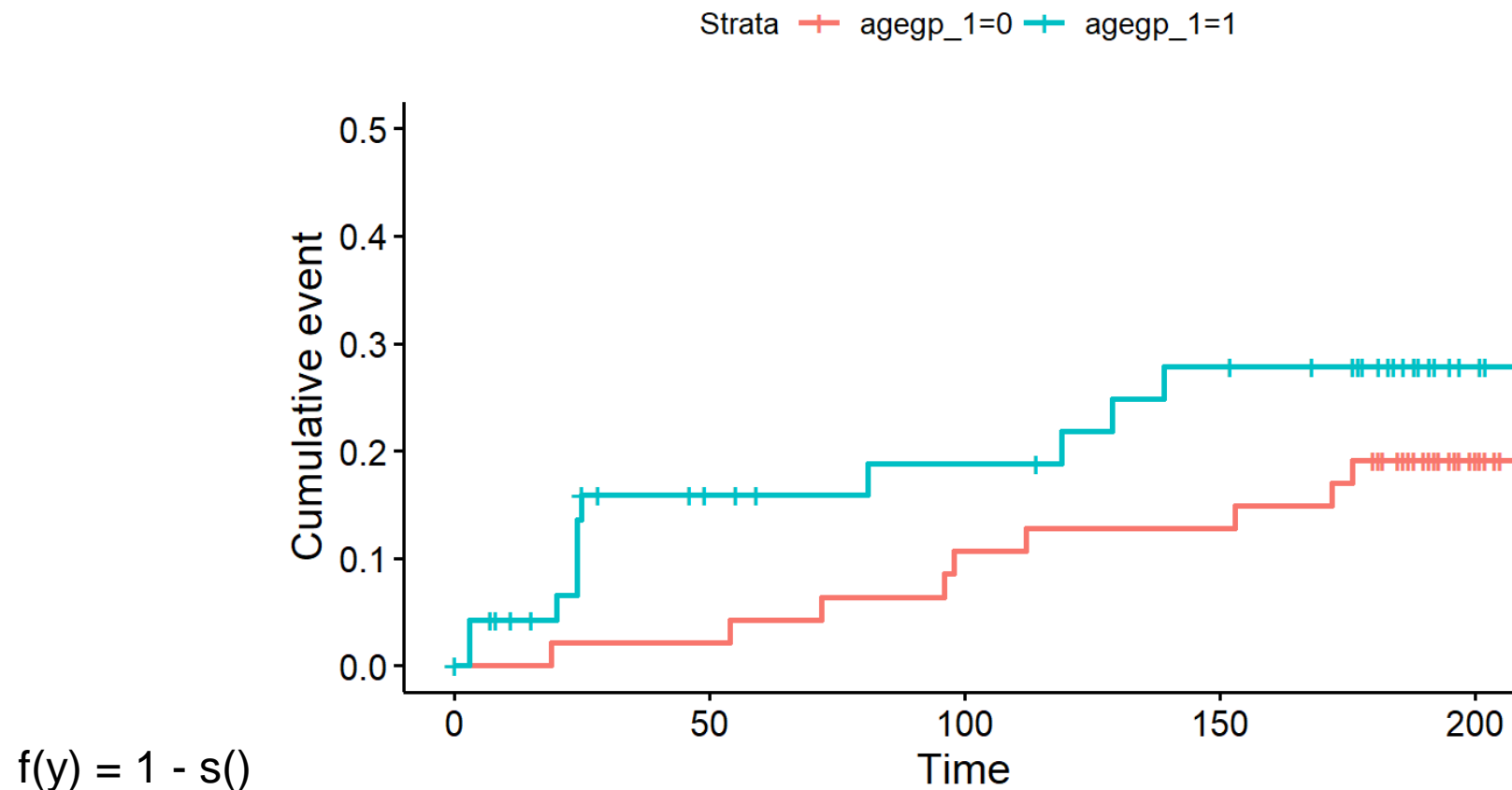
Kaplan-Meier estimator, survival probability

- `ggsurvplot(fit_1a, ylim = c(0.5, 1))`



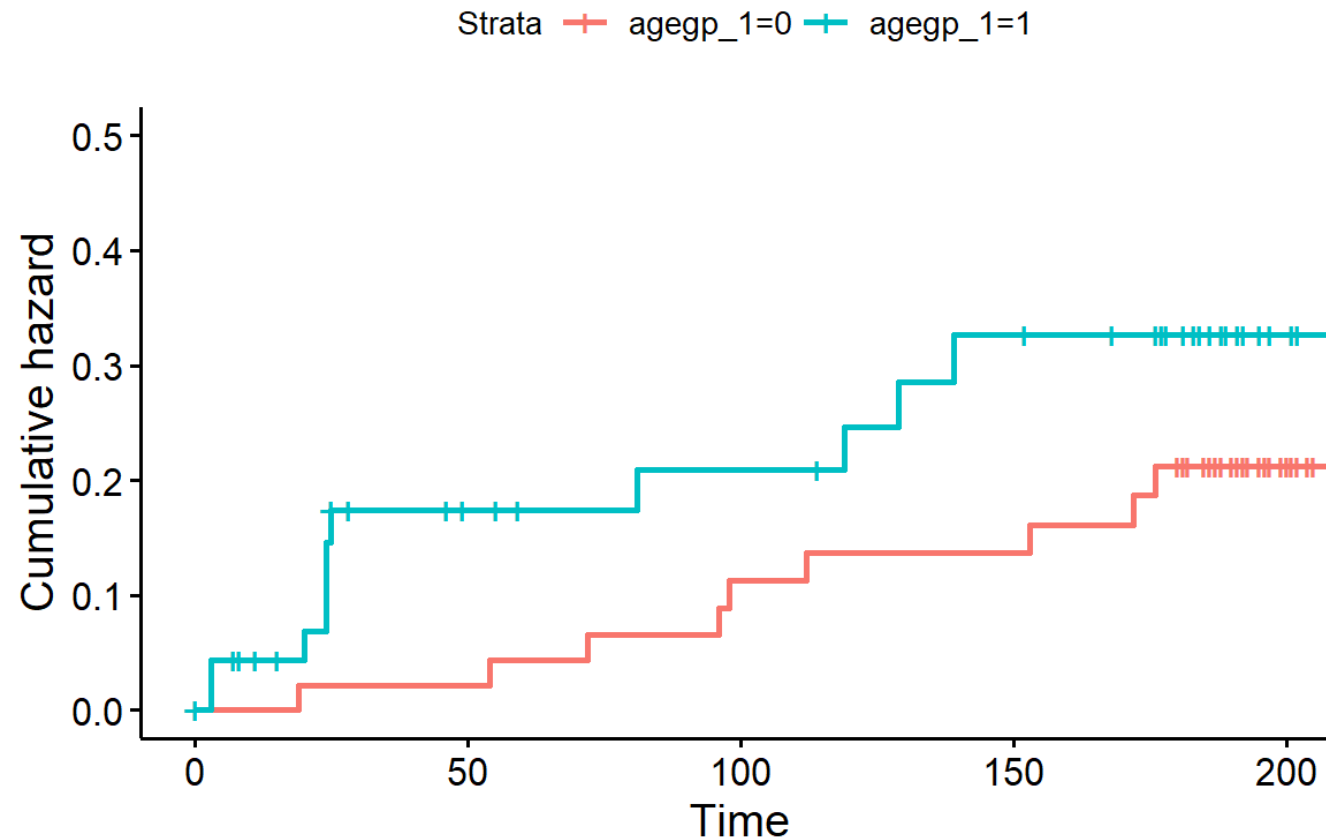
Kaplan-Meier estimator, cumulative event

- `ggsurvplot(fit_1a, fun = "event", ylim = c(0, 0.5))`



Nelson-Aalen estimator

- `ggsurvplot(fit_1a, fun = "cumhaz", ylim = c(0, 0.5))`



$$f(y) = -\log(s())$$

比較存活函數的差異

- 比較各個時間點上的實際事件數目與期望事件數目
- 檢定不同組別的事件發生機率(存活函數)是否有差異

```
> # log rank test  
> survdiff(Surv(death_ft, death) ~ agegp_1, data = pneumonia)  
Call:  
survdiff(formula = Surv(death_ft, death) ~ agegp_1, data = pneumonia)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
agegp_1=0	47	2	10.34	6.73	14.8
agegp_1=1	48	17	8.66	8.04	14.8

Chisq= 14.8 on 1 degrees of freedom, p= 1e-04

Cox proportional hazard model

- 估計模型
- 模型基本假設檢定
- 模型殘差診斷

估計模型：單變項Cox PH model

```
> # 以再次肺炎為結果變項，自變項放入pneumonia_pre，1代表過去半年曾因為肺炎就醫  
> fit_2a <- coxph(Surv(re_adm_ft, re_adm) ~ pneumonia_pre, data = pneumonia)  
>
```

模型定義 & 配適

```
> # 檢視模型摘要  
> summary(fit_2a)
```

Call:

```
coxph(formula = Surv(re_adm_ft, re_adm) ~ pneumonia_pre, data = pneumonia)
```

```
n= 95, number of events= 20
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
pneumonia_pre	0.8767	2.4029	0.4529	1.936	0.0529 .

因子顯著性檢定

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

	exp(coef)	exp(-coef)	lower .95	upper .95
pneumonia_pre	2.403	0.4162	0.989	5.838

因子Hazard ratio估計值及信賴區間

```
Concordance= 0.603 (se = 0.055 )
```

```
Likelihood ratio test= 3.54 on 1 df, p=0.06
```

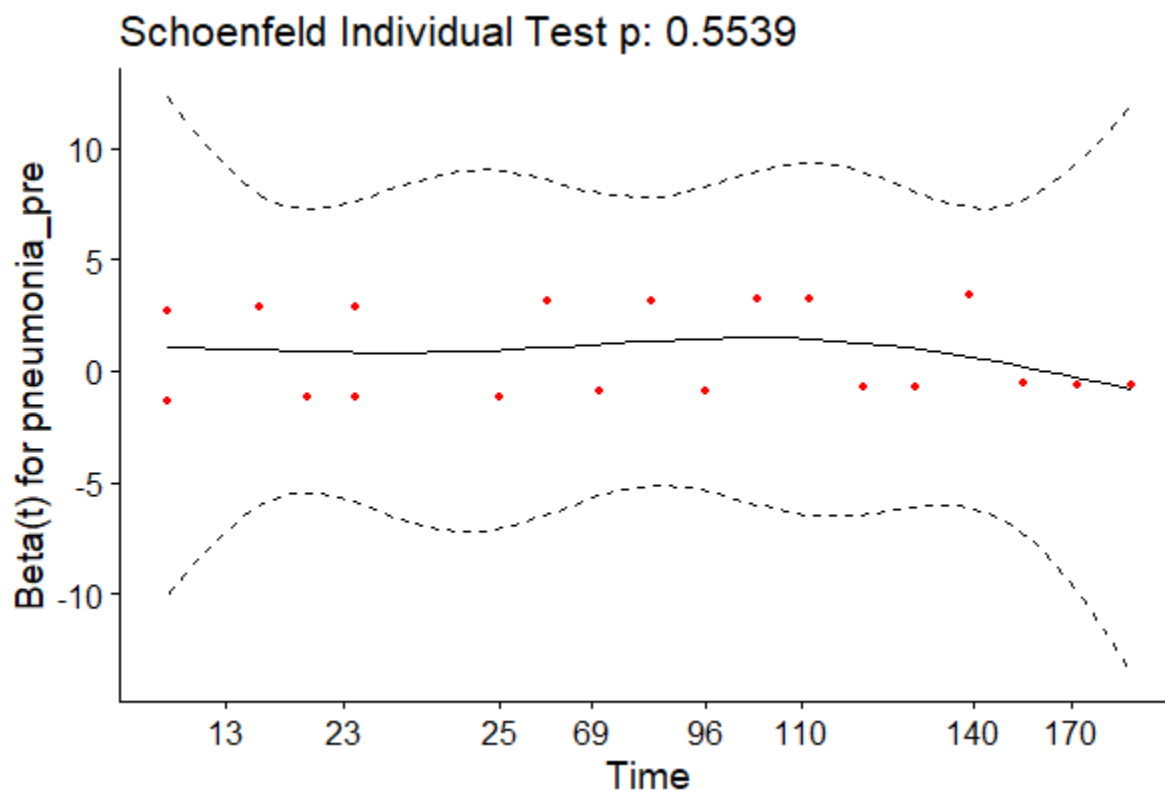
```
Wald test = 3.75 on 1 df, p=0.05
```

```
Score (logrank) test = 3.98 on 1 df, p=0.05
```

模型基本假設(**proportional hazard**)檢定

檢定hazard是否在不同時間保持等比例關係(不可以跟時間有關)

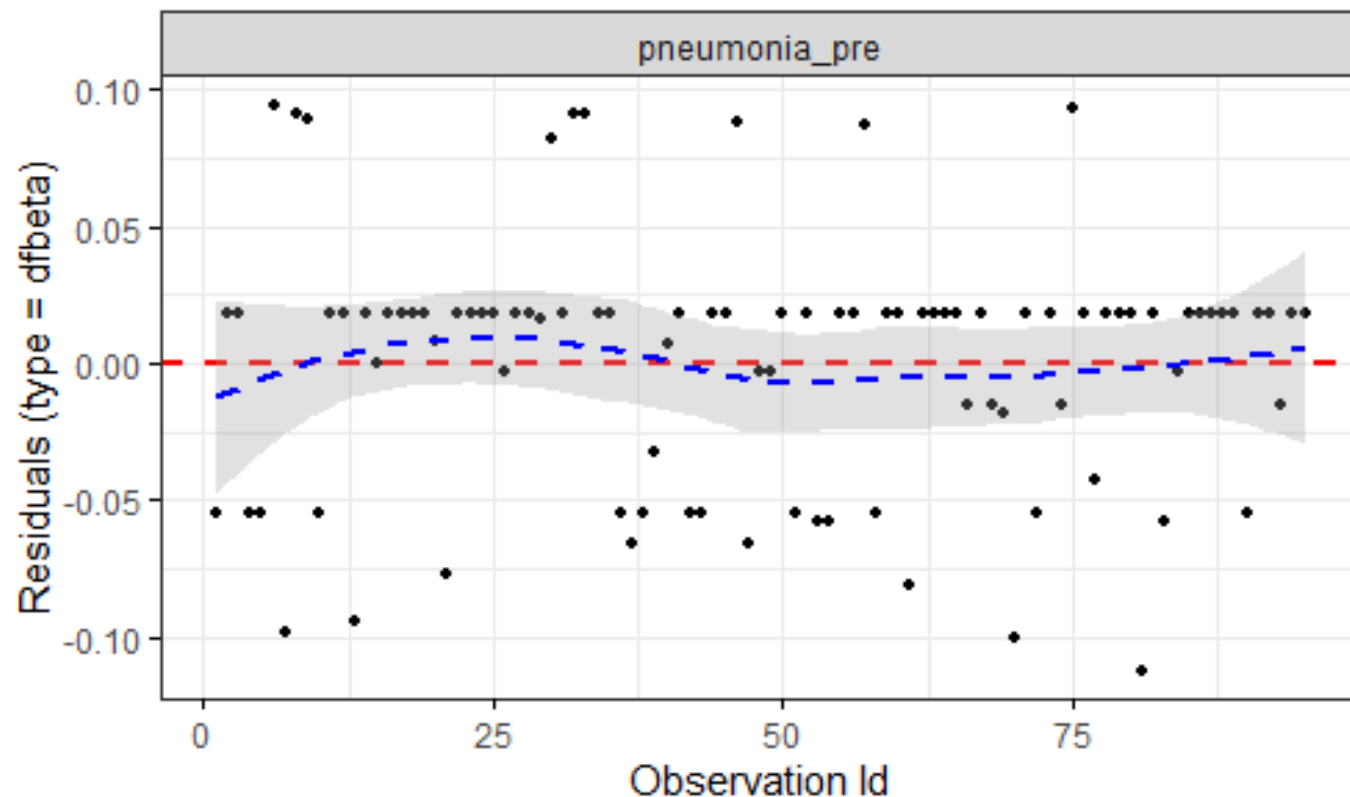
```
ggcoxzph(cox.zph(fit_2a)) # survminer::ggcoxzph
```



模型殘差診斷

模擬每抽掉一個觀察值，對迴歸係數的影響

```
ggcoxdiagnostics(fit_2a, type = "dfbeta")
```



估計模型：多變項Cox PH model

```
coxph(Surv(re_adm_ft, re_adm) ~ agegp_1 + sex_1 + incgp_1 +  
incgp_2 + med_center_1 + bed_day + cci + pneumonia_pre + htn  
+ dm + hyperlipidemia, data = pneumonia)
```

	exp(coef)	lower .95	upper .95	Pr(> z)	
agegp_1	0.6871	0.19827	2.381	0.5539	
sex_1	2.8074	0.88344	8.921	0.0801	.
incgp_1	0.6879	0.23586	2.006	0.4933	
incgp_2	0.2823	0.06457	1.234	0.0929	.
med_center_1	1.2169	0.40342	3.671	0.7275	
bed_day	1.1314	1.06476	1.202	6.74e-05	***
cci	1.0862	0.82543	1.429	0.5551	
pneumonia_pre	2.5677	0.90847	7.257	0.0753	.
htn	0.9130	0.30422	2.740	0.8710	
dm	0.1991	0.03347	1.184	0.0760	.
hyperlipidemia	2.3927	0.22614	25.317	0.4685	

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

衍伸模型

- 分層Cox模型
- 競爭風險模型

衍伸模型：分層Cox PH model

```
coxph(Surv(re_adm_ft, re_adm) ~ agegp_1 + sex_1 + incgp_1 +  
incgp_2 + strata(med_center_1) + bed_day + cci +  
pneumonia_pre + htn + dm + hyperlipidemia, data = pneumonia)
```

	exp(coef)	lower .95	upper .95	Pr(> z)	
agegp_1	0.6492	0.18532	2.274	0.499321	
sex_1	2.8131	0.88734	8.918	0.078930	.
incgp_1	0.7280	0.25132	2.109	0.558457	
incgp_2	0.2825	0.06319	1.263	0.098019	.
bed_day	1.1268	1.06080	1.197	0.000106	***
cci	1.0781	0.81641	1.424	0.596139	
pneumonia_pre	2.7606	0.96084	7.932	0.059326	.
htn	0.9225	0.30602	2.781	0.886128	
dm	0.2065	0.03513	1.214	0.080894	.
hyperlipidemia	2.5569	0.23645	27.650	0.439612	

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

衍伸模型：競爭風險模型

```
crr(  
  ftime = pneumonia$re_adm_ft,  
  fstatus = pneumonia$crr_re_adm,  
  cov1 = pneumonia[, .(agegp_1, sex_1, incgp_1,  
    incgp_2, med_center_1, bed_day, cci,  
    pneumonia_pre, htn, dm, hyperlipidemia)]  
)
```

課程討論 & Final remark

- Summary

- 存活分析的應用時機
- 使用R語言指令完成存活分析
 - 圖型繪製
 - 模型配適
 - 模型估計
 - 模型診斷
 - 衍伸模型

- See more

- Official website & Cheat sheet
 - survival
 - survminer

- Stack Overflow



- STHDA

- Build by A. Kassambara
creator of survminer

