LOS Data Analysis

7109018032謝綺珊
7109018037李蕙君

01

Descriptive Statistics

Variables

```
> head(LOS)
```

	ID	AGE	FUNC	PT	OT	ST	FIM_F	FIM_S	BAR_F	BAR_S	sex	ht	dm	in_date	out_date
1	1	21	神外	1	1	0	48	50	10	15	0	0	0	2000/11/1	2000/12/19
2	2	21	神外	1	1	0	67	81	30	65	1	0	0	2000/10/25	2000/11/30
3	3	21	骨疫	1	0	0	125	125	95	100	1	0	0	2002/5/3	2002/5/6
4	4	28	骨疫	1	1	0	83	89	35	50	1	0	0	2001/8/13	2001/9/20
5	5	29	神外	1	0	0	18	18	0	0	0	0	0	2000/11/15	2001/1/9
6	6	30	內疫	1	1	0	87	99	20	60	0	0	0	2003/9/17	2003/10/25

共544個樣本

ID : 患者編號

AGE : 年齡

FUNC : 經由該門診辦理住院

PT : 物理治療

OT : 職能治療

ST : 語言治療

FIM_F : 入院時功能獨立量表分數

FIM_S : 出院時功能獨立量表分數

BAR_F : 入院時巴氏量表分數 (35以下可申請外籍看護)

BAR_S : 出院時巴氏量表分數

sex : 性別

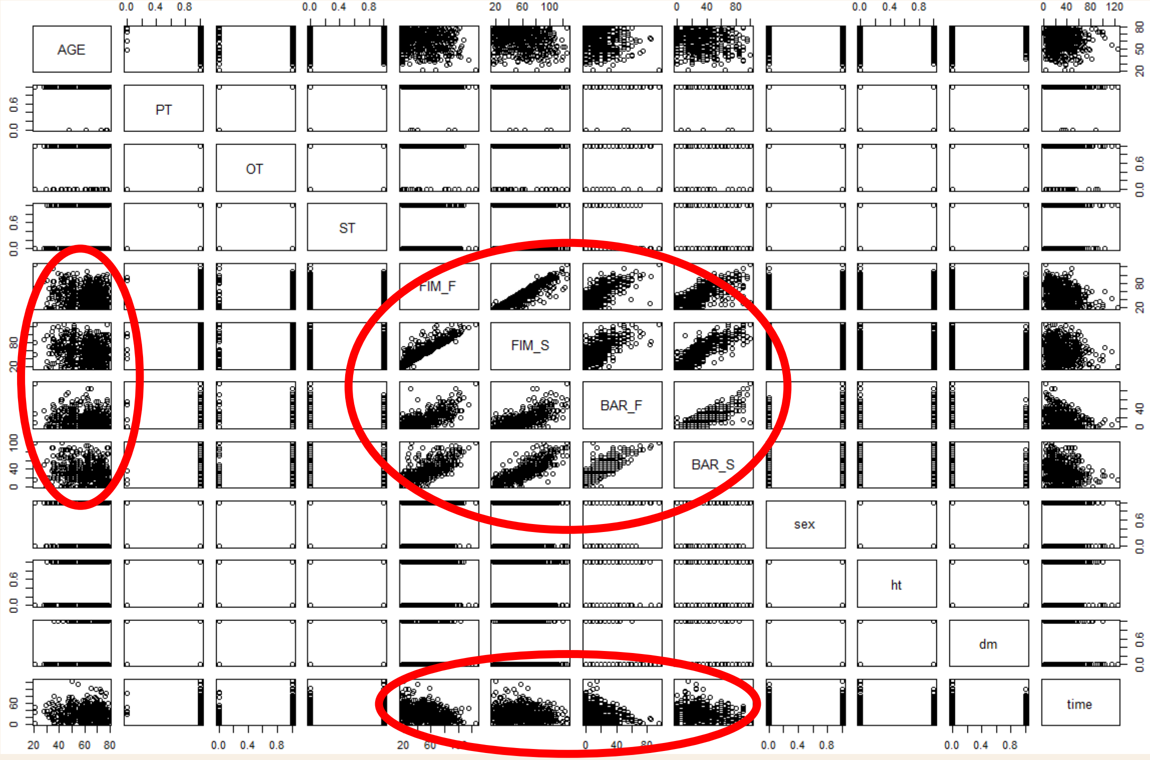
ht : 高血壓

dm : 糖尿病

in_date : 入院日期

out_date : 出院日期

Correlation



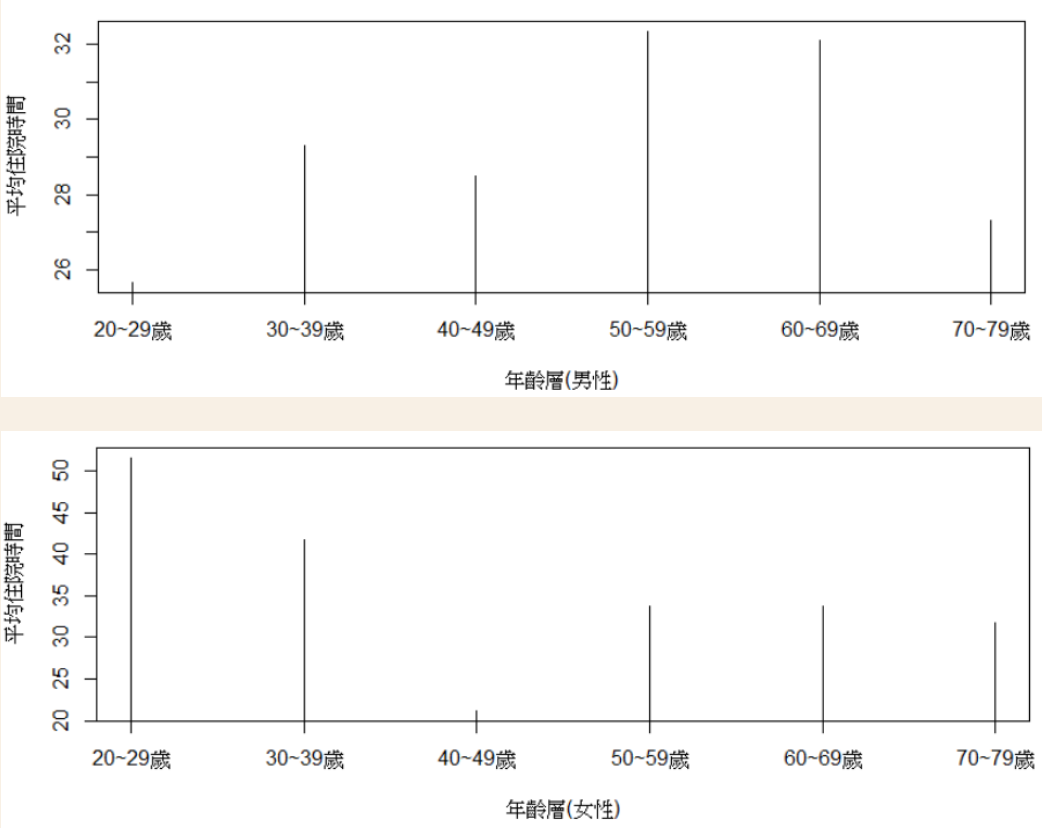

```
> cor(LOS[,7:10])
          FIM_F     FIM_S     BAR_F     BAR_S
FIM_F 1.0000000 0.9131617 0.7252305 0.7280111
FIM_S 0.9131617 1.0000000 0.7053760 0.8140139
BAR_F 0.7252305 0.7053760 1.0000000 0.8174083
BAR_S 0.7280111 0.8140139 0.8174083 1.0000000
```

```
> cor(LOS$AGE,LOS$FIM_F)
[1] -0.1647077
> cor(LOS$AGE,LOS$BAR_F)
[1] -0.1348637
> cor(LOS$AGE,LOS$FIM_S)
[1] -0.1756479
> cor(LOS$AGE,LOS$BAR_S)
[1] -0.1651406
```

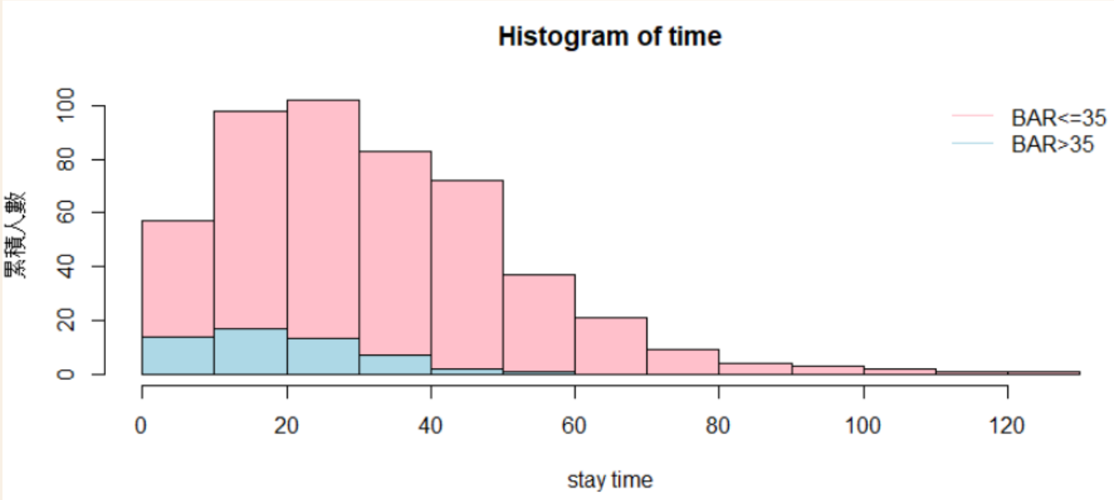
```
> cor(LOS$AGE,as.numeric(time))
[1] -0.0004152204
> cor(LOS$AGE,LOS$FIM_S-LOS$FIM_F)
[1] -0.06202067
> cor(as.numeric(time),LOS$FIM_F)
[1] -0.2496124
> cor(as.numeric(time),LOS$BAR_F)
[1] -0.2706034
```

按年齡做分組

20~29歲
30~39歲
40~49歲
50~59歲
60~69歲
70~79歲



FIM分數若為35分以下，則可以申請外籍看護
有490人在入院時分數已達可申請看護標準



02

Model and Results

Survival Analysis

研究或分析樣本所觀察到的某一段時間長度之分配，一段時間長度通常是從一特定事件起始之時間原點(起始時間點)直到某一特定事件發生的時間點

- 時間長度
  時間原點到事件發生時間點

- 事件：死亡、患病、復發、提早退出試驗

- 資料 { 完整
設限

在此次LOS中是分析影響住院時間長短的變數

Survival Analysis

- survival function
  某特定時間點下，個案可以活過此特定時間點的機率

$$S(t) = P(T > t) = \int_t^\infty f(u)\ du$$

  CDF : $F(t) = P(T \leq t) = 1 - S(t)$

- hazard function
  下一個瞬間事件發生的機率

$$h(t) = \lim_{\delta \to 0} \frac{1}{\delta} P(t \leq T < t + \delta | T \geq t)$$

  累積風險：$H(t) = \int_0^t h(u)\ du$

- $h(t) = \frac{-\frac{d[S(t)]}{dt}}{S(t)} = \frac{f(t)}{S(t)} = -\frac{d}{dt}\log[S(t)]$
  $S(t) = \exp(-\int_0^t h(u)\ du)$

Cox Regression

假設解釋變數對風險的作用是成比例，則hazard function和基線風險度(baseline hazard)的關係可表示為

$$h(t|x) = h_0(t)e^{\beta^T x}$$

比例風險假設：解釋變量對與風險的作用所帶來的風險比是不隨時間改變

$$h(t) = \lambda$$

$$S(t) = \exp\left(-\int_0^t h(u)\ du\right) = \exp(-\lambda t) = e^{-\lambda t}$$

$$f(t) = h(t) * S(t) = \lambda * e^{-\lambda t}$$

因 λ 為一定值，故可以推出 Cox Regression 為

$$\log\big(HR(x)\big) = log\frac{h(t|x)}{h_0(t)} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

Cox Regression

Cox 比例風險模型下的概似函數為

$$L = \prod_{i=1}^{n} \{h_0(t)e^{\beta^T x_i} \exp\left(-\int_0^{t_i} h_0(u)e^{\beta^T x_i}\, du\right)\}^{\delta_i}\{\exp\left(-\int_0^{t_i} h_0(u)e^{\beta^T x_i}\, du\right)\}^{1-\delta_i}$$

無法估計，所以使用partial likelihood，它漸進一致(asymptotically the same)完整的likelihood

時間點發生事件的條件機率為

$$\frac{h_0(t_j)\exp(\beta^T x_{i_j})}{\sum_{k\in R_j} h_0(t_j)\exp(\beta^T x_k)} = \frac{\exp(\beta^T x_{i_j})}{\sum_{k\in R_j}\exp(\beta^T x_k)}$$

其中 i_j 為實驗對象， x_{i_j}為解釋變數， t_j為生存時間

用這個性質忽略$h_0(t)$ ，使用MLE去估計β
把所有發生事件時的條件機率相乘，得到偏概似函數

$$L_p = \prod_j \frac{\exp(\beta^T x_{i_j})}{\sum_{k\in R_j}\exp(\beta^T x_k)}$$

$$l_p = \sum_j \beta^T x_{i_j} - \sum_j \log(\sum_{k\in R_j}\exp(\beta^T x_k))$$

透過 $\frac{dl_p}{d\beta} = \sum_j x_{i_j} - \sum_j \frac{\sum_{k\in R_j} x_k \exp(\beta x_k)}{\sum_{k\in R_j}\exp(\beta x_k)} = 0$ ，求解 $\hat{\beta}_{MLE}$

Cox Regression

使用不同變數建立模型

```
> model1 = coxph(time~FIM_F+FIM_S+BAR_F,data = LOS.train)
> summary(model1)
Call:
coxph(formula = time ~ FIM_F + FIM_S + BAR_F, data = LOS.train)

  n= 400, number of events= 400

           coef exp(coef)  se(coef)      z Pr(>|z|)    
FIM_F  0.023328  1.023602  0.006639  3.514 0.000441 ***
FIM_S -0.019380  0.980806  0.005724 -3.386 0.000710 ***
BAR_F  0.020419  1.020629  0.005177  3.944 8.02e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

      exp(coef) exp(-coef) lower .95 upper .95
FIM_F    1.0236     0.9769    1.0104    1.0370
FIM_S    0.9808     1.0196    0.9699    0.9919
BAR_F    1.0206     0.9798    1.0103    1.0310

Concordance= 0.62  (se = 0.016 )
Likelihood ratio test= 56.6  on 3 df,   p=3e-12
Wald test            = 61.45  on 3 df,   p=3e-13
Score (logrank) test = 61.19  on 3 df,   p=3e-13
```

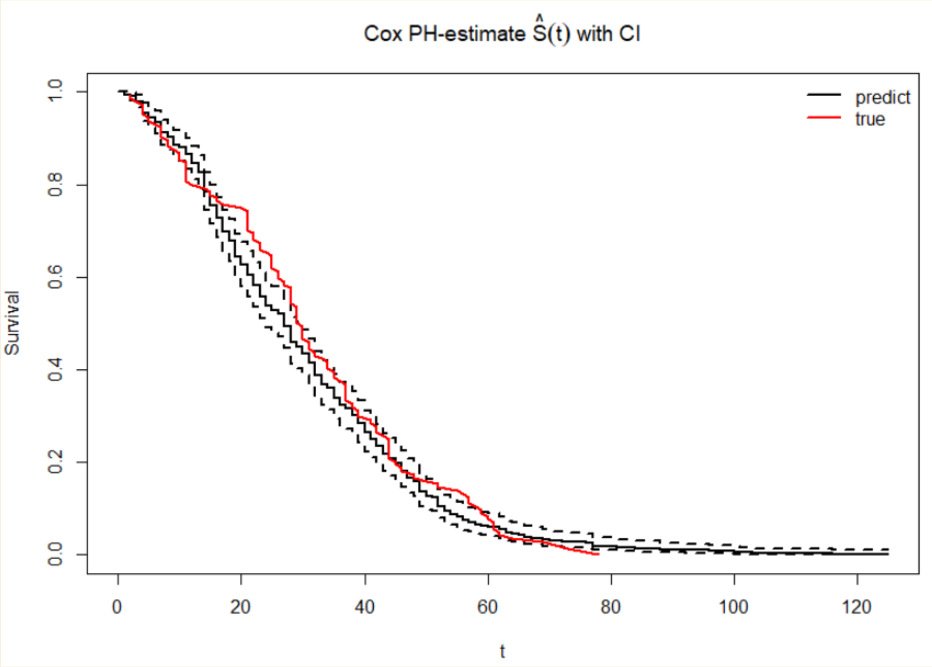
```
> model = coxph(time~AGE+FIM_F+FIM_S+BAR_F+BAR_S+PT+OT+ST+sex+ht+dm,data =
 LOS.train)
> summary(model)
Call:
coxph(formula = time ~ AGE + FIM_F + FIM_S + BAR_F + BAR_S + 
    PT + OT + ST + sex + ht + dm, data = LOS.train)

  n= 400, number of events= 400

           coef exp(coef)  se(coef)      z Pr(>|z|)    
AGE    0.005324  1.005339  0.004577  1.163  0.24472    
FIM_F  0.021578  1.021812  0.006697  3.222  0.00127 ** 
FIM_S -0.015450  0.984669  0.006644 -2.326  0.02004 *  
BAR_F  0.026119  1.026463  0.006590  3.963 7.39e-05 ***
BAR_S -0.007466  0.992562  0.005637 -1.324  0.18536    
PT     0.480933  1.617583  0.755881  0.636  0.52461    
OT    -0.064365  0.937663  0.252897 -0.255  0.79910    
ST    -0.198766  0.819742  0.111114 -1.789  0.07364 .  
sex    0.141240  1.151701  0.111742  1.264  0.20624    
ht     0.205265  1.227851  0.105399  1.948  0.05147 .  
dm    -0.041958  0.958910  0.142005 -0.295  0.76763    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
      exp(coef) exp(-coef) lower .95 upper .95
AGE      1.0053     0.9947    0.9964    1.0144
FIM_F    1.0218     0.9787    1.0085    1.0353
FIM_S    0.9847     1.0156    0.9719    0.9976
BAR_F    1.0265     0.9742    1.0133    1.0398
BAR_S    0.9926     1.0075    0.9817    1.0036
PT       1.6176     0.6182    0.3677    7.1166
OT       0.9377     1.0665    0.5712    1.5393
ST       0.8197     1.2199    0.6593    1.0192
sex      1.1517     0.8683    0.9252    1.4337
ht       1.2279     0.8144    0.9987    1.5096
dm       0.9589     1.0429    0.7259    1.2666

Concordance= 0.621  (se = 0.016 )
Likelihood ratio test= 66.27  on 11 df,   p=6e-10
Wald test            = 71.01  on 11 df,   p=8e-11
Score (logrank) test = 71.28  on 11 df,   p=7e-11
```

Cox Regression

模型比較

```
> anova(model1,model)
Analysis of Deviance Table
 Cox model: response is  time
 Model 1: ~ FIM_F + FIM_S + BAR_F
 Model 2: ~ AGE + FIM_F + FIM_S + BAR_F + BAR_S + PT + OT + ST + sex + ht + 
 dm
   loglik  Chisq Df P(>|Chi|)
1 -1972.2
2 -1967.4 9.6724  8    0.2888
```

Estimated Survival Function



Thank you for your attention