

Introduction  
oo

K-M plot  
oooo

Local test  
oooooooooooo

Schoenfeld residual  
ooo

Breslow Baseline  
o

Cox-Snell residual  
o

Conclusion  
o

# Breast Cancer Dataset

Reporter:Wei-Bo Huang

May 29 2023

Introduction  
oo

K-M plot  
oooo

Local test  
oooooooooooo

Schoenfeld residual  
ooo

Breslow Baseline  
o

Cox-Snell residual  
o

Conclusion  
o

## 1 Introduction

## 2 K-M plot

## 3 Local test

## 4 Schoenfeld residual

## 5 Breslow Baseline

## 6 Cox-Snell residual

## 7 Conclusion

# About Dataset

The data set contains patient records from a 1984-1989 trial conducted by the German Breast Cancer Study Group (GBSG) of 720 patients with node positive breast cancer; it retains the 686 patients with complete data for the prognostic variables.

Columns	Description
pid	patient identifier
age	age,year
meno	menopausal status (0= premenopausal, 1= postmenopausal)
size	tumor size, mm
grade	tumor grade
nodes	number of positive lymph nodes
pgr	progesterone receptors (fmol/l)
er	estrogen receptors (fmol/l) hormon hormonal therapy 0= no, 1= yes
rfstime	recurrence free survival time; days to first of recurrence, death or last follow-up
status	0= alive without recurrence, 1= recurrence or death

Introduction  
oo

K-M plot  
●ooo

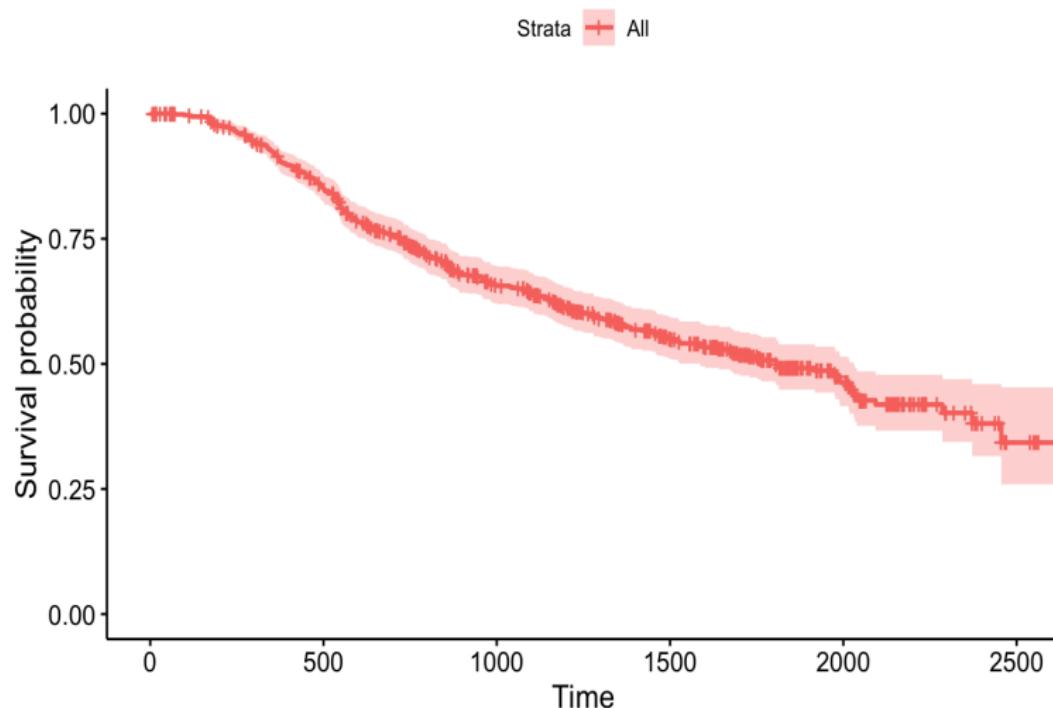
Local test  
oooooooooooo

Schoenfeld residual  
ooo

Breslow Baseline  
o

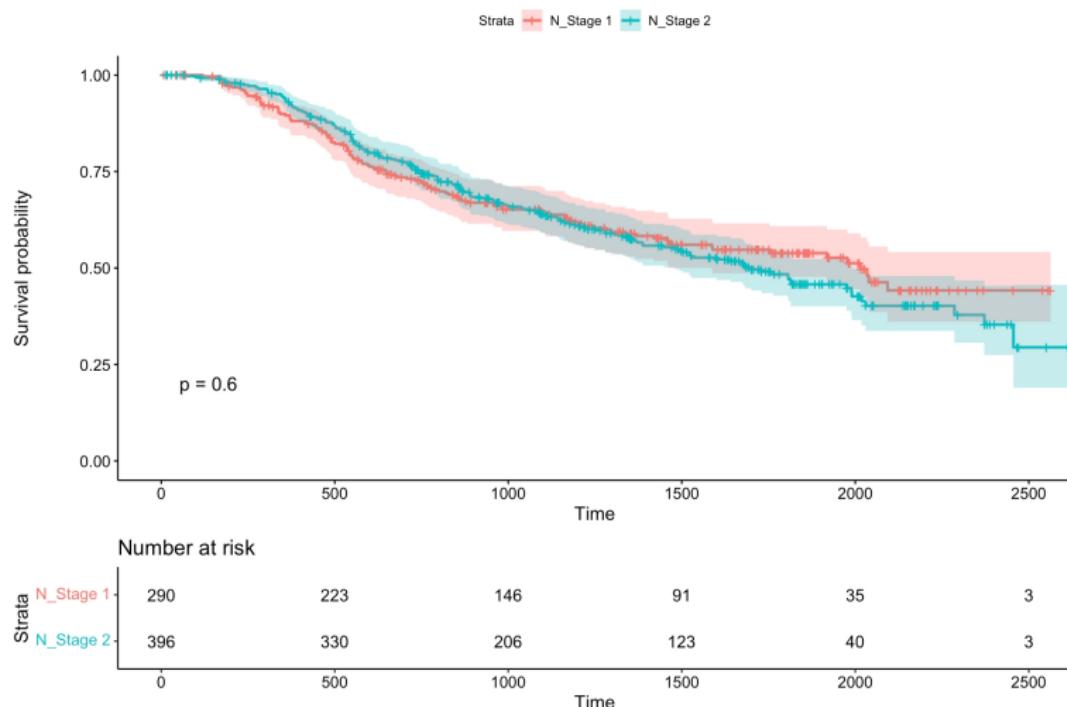
Cox-Snell residual  
o

Conclusion  
o



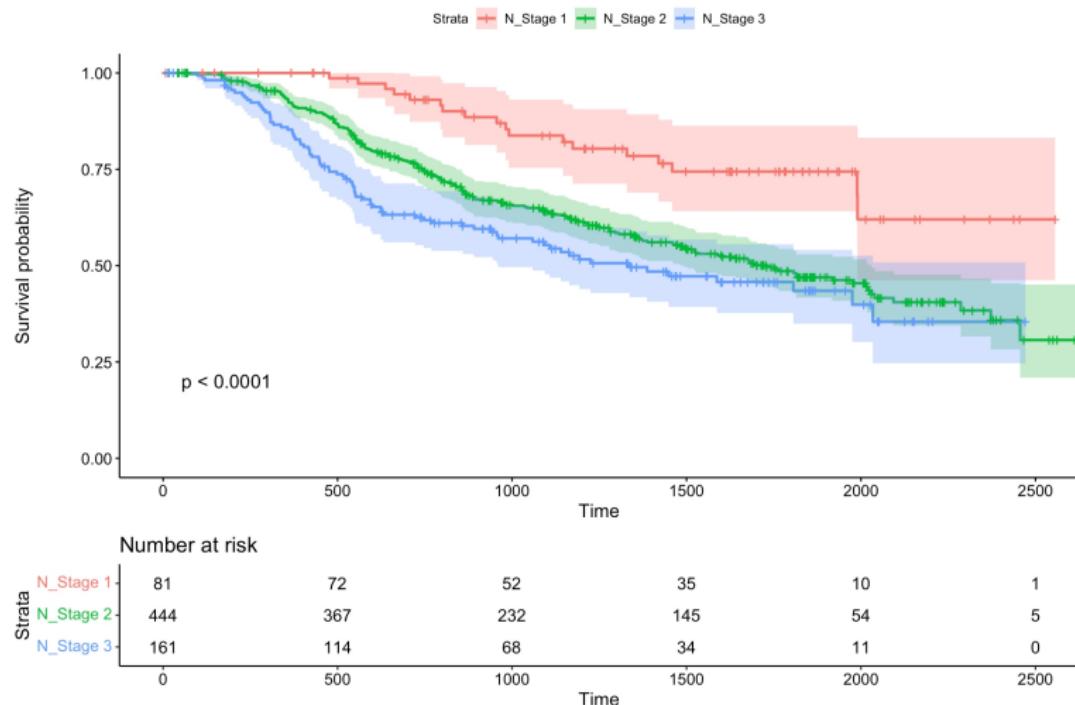
# Meno

Kaplan-Meier curves



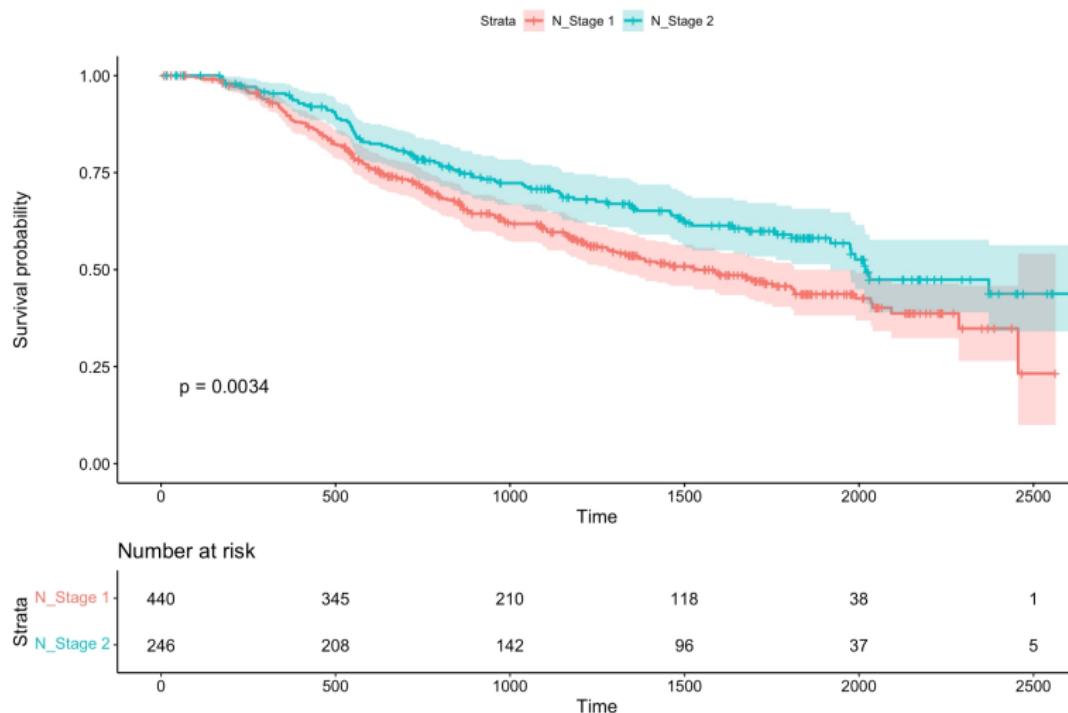
# Grade

Kaplan-Meier curves



# Hormon

Kaplan-Meier curves



# Cox PH model

n= 686, number of events= 299

	coef	exp(coef)	se(coef)	z	Pr(> z )	
age	-0.0093924	0.9906516	0.0092733	-1.013	0.311136	
meno	0.2672772	1.3064025	0.1833366	1.458	0.144882	
size	0.0077164	1.0077463	0.0039497	1.954	0.050739	.
grade	0.2802894	1.3235128	0.1060553	2.643	0.008221	**
nodes	0.0498939	1.0511596	0.0074094	6.734	1.65e-11	***
pgr	-0.0022378	0.9977647	0.0005758	-3.887	0.000102	***
er	0.0001674	1.0001674	0.0004477	0.374	0.708431	
hormon	-0.3372029	0.7137640	0.1289618	-2.615	0.008929	**

# Likelihood ratio test Wald test Score test(age)

```
# Likelihood ratio test
````{r}
Fit_reduced = coxph(Surv(rfstime, status)~ age, data = cgd)
chiLR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chiLR ##卡方值
1-pchisq(chiLR, 1) ##p-value
````
```

```
[1] 101.2803
[1] 0
```

```
# Wald test
````{r}
Cov <- c(1) ##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%%solve(var11)%%(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
````
```

```
[,1]
[1,] 1.025846
      [,1]
[1,] 0.3111364
```

```
````{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(horman), data = cgd,init=c(Fitw$coefficients[1:2],0,Fitw$coefficients[4:10]),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[3])%%Fit_age$var[3,3]%%score.vector[3]

1-pchisq(chiSC_age,1) #p value
````
```

```
[,1]
[1,] 0
```

# Likelihood ratio test Wald test Score test(meno)

```
```{r}
Fit_reduced = coxph(Surv(rfstime, status)~ meno, data = cgd)
chILR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chILR ##卡方值
1-pchisq(chILR, 1) ##p-value
```

```

```
[1] 101.5786
[1] 0
```

```
Cov <- c(2) ##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%%solve(var11)%%(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
```

```

```
[,1]
[1,] 2.125325
[,1]
[1,] 0.1448821
```

```
```{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(hormon), data = cgd,init=c(Fitw$coefficients[1:3],0,Fitw$coefficients[5:10]),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[4])%%Fit_age$var[4,4]%%score.vector[4]

1-pchisq(chiSC_age,1) #p value
```

```

```
[,1]
[1,] 0.0003997243
```

# Likelihood ratio test Wald test Score test(size)

```
```{r}
Fit_reduced = coxph(Surv(rfstime, status)~ size, data = cgd)
chiLR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chiLR ##卡方值
1-pchisq(chiLR, 1) ##p-value
```

```

```
[1] 86.1764
[1] 0
```

```
Cov <- c(3) ##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%%solve(var11)%%(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
```

```

```
[,1]
[1,] 3.816883
[,1]
[1,] 0.05073855
```

```
```{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(hormon), data = cgd,init=c(Fitw$coefficients[1:4],0,Fitw$coefficients[6:10]),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[5])%%Fit_age$var[5,5]%%score.vector[5]

1-pchisq(chiSC_age,1) #p value
```

```

```
[,1]
[1,] 0
```

# Likelihood ratio test Wald test Score test(grade)

```
```{r}
Fit_reduced = coxph(Surv(rfstime, status)~ grade, data = cgd)
chiLR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chiLR ##卡方值
1-pchisq(chiLR, 1) ##p-value
```

```

```
[1] 81.8749
[1] 0
```

```
Cov <- c(4) ##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%%solve(var11)%%(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
```

```

```
[,1]
[1,] 6.984712
[1,] 
[1,] 0.008220888
```

```
```{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(hormon), data = cgd,init=c(Fitw$coefficients[1:5],0,Fitw$coefficients[7:10]),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[6])%%Fit_age$var[6,6]%%score.vector[6]

1-pchisq(chiSC_age,1) #p value
```

```

```
[,1]
[1,] 0
```

# Likelihood ratio test Wald test Score test(nodes)

```
```{r}
Fit_reduced = coxph(Surv(rfstime, status)~ nodes, data = cgd)
chiLR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chiLR ##卡方值
1-pchisq(chiLR, 1) ##p-value
```

```

```
[1] 51.82479
[1] 6.068479e-13
```

```
Cov <- c(5) ##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%%solve(var11)%%(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
```

```

```
[,1]
[1,] 45.3453
[,1]
[1,] 1.651834e-11
```

```
```{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(hormon), data = cgd,init=c(Fitw$coefficients[1:6],0,Fitw$coefficients[8:10]),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[7])%%Fit_age$var[7,7]%%score.vector[7]

1-pchisq(chiSC_age,1) #p value
```

```

```
[,1]
[1,] 0
```

# Likelihood ratio test Wald test Score test(pgr)

```
```{r}
Fit_reduced = coxph(Surv(rfstime, status)~ pgr, data = cgd)
chiLR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chiLR ##卡方值
1-pchisq(chiLR, 1) ##p-value
```

```

```
[1] 67.81484
[1] 2.220446e-16
```

```
Cov <- c(6) ##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%*%solve(var11)%*%(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
```

```

```
[,1]
[1,] 15.1065
[1,] 0.0001016129
```

```
```{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(hormon), data = cgd,init=c(Fitw$coefficients[1:7],0,Fitw$coefficients[9:10]),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[8])%*%Fit_age$var[8,8]%*%score.vector[8]

1-pchisq(chiSC_age,1) #p value
```

```

```
[,1]
[1,] 0
```

# Likelihood ratio test Wald test Score test(er)

```
```{r}
Fit_reduced = coxph(Surv(rfstime, status)~ er, data = cgd)
chiLR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chiLR ##卡方值
1-pchisq(chiLR, 1) ##p-value
```

```

```
[1] 97.15835
[1] 0
```

```
```{r}
Cov <- c(7) ##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%%solve(var11)%%*(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
```

```

```
[,1]
[1,] 0.1398491
[1,]
[1,] 0.7084311
```

```
```{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(hormon), data = cgd,init=c(Fitw$coefficients[1:8],0,Fitw$coefficients[10]),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[9])%%Fit_age$var[9,9]%%score.vector[9]

1-pchisq(chiSC_age,1) #p value
```

```

```
[,1]
[1,] 0
```

# Likelihood ratio test Wald test Score test(hormon)

```
```{r}
Fit_reduced = coxph(Surv(rfstime, status)~ hormon, data = cgd)
chiLR = 2*(Fitw$loglik[2]- Fit_reduced$loglik[2])
chiLR ##卡方值
1-pchisq(chiLR, 1) ##p-value
````
```

```
[1] 93.0389
[1] 0
```

```
Cov <- c(8)##收集變數
beta1hat = Fitw$coefficients[Cov]
beta10 = rep(0, 1)
var11 = Fitw$var[Cov,Cov]
chiWald = t(beta1hat - beta10)%%solve(var11)%%(beta1hat- beta10)
chiWald ##卡方值
1-pchisq(chiWald,1) ##p-value
````
```

```
[,1]
[1,] 6.836924
[1,] 
[1,] 0.008929245
```

```
```{r}
Fit_age=coxph(Surv(rfstime, status)~ pid + age + factor(meno) + size + factor(grade) + nodes + pgr + er +
factor(hormon), data = cgd,init=c(Fitw$coefficients[1:9],0),iter=0)
score.vector=colSums(coxph.detail(Fit_age)$score)

chiSC_age=t(score.vector[10])%%Fit_age$var[10,10]%%score.vector[10]

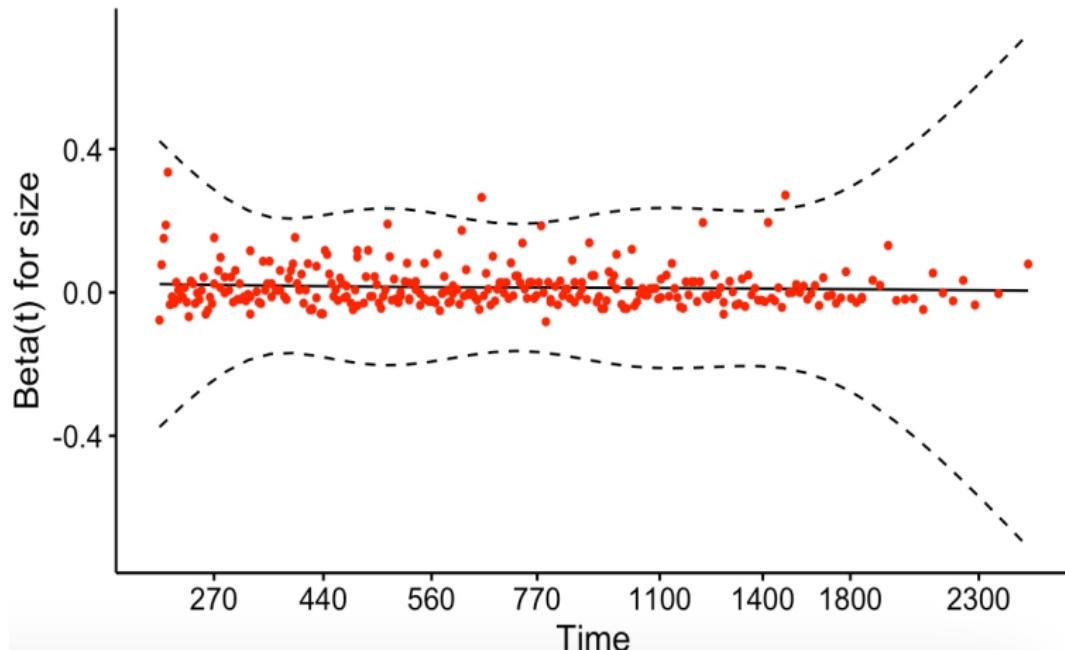
1-pchisq(chiSC_age,1) #p value
````
```

```
[1,]
[1,] 0
```

|        | Cox PH model | Likelihood ratio | Wald         | Score        |
|--------|--------------|------------------|--------------|--------------|
| age    | 0.311136     | 0                | 0.3111364    | 0            |
| meno   | 0.144882     | 0                | 0.1448821    | 0.0003997243 |
| size   | 0.050739     | 0                | 0.05073855   | 0            |
| grade  | 0.008221     | 0                | 0.008220888  | 0            |
| nodes  | 1.65e-11     | 1.110223e-16     | 1.651834e-11 | 0            |
| pgr    | 0.000102     | 0                | 0.0001016129 | 0            |
| er     | 0.70483      | 0                | 0.7084311    | 0            |
| hormon | 0.008929     | 0                | 0.008929245  | 0            |

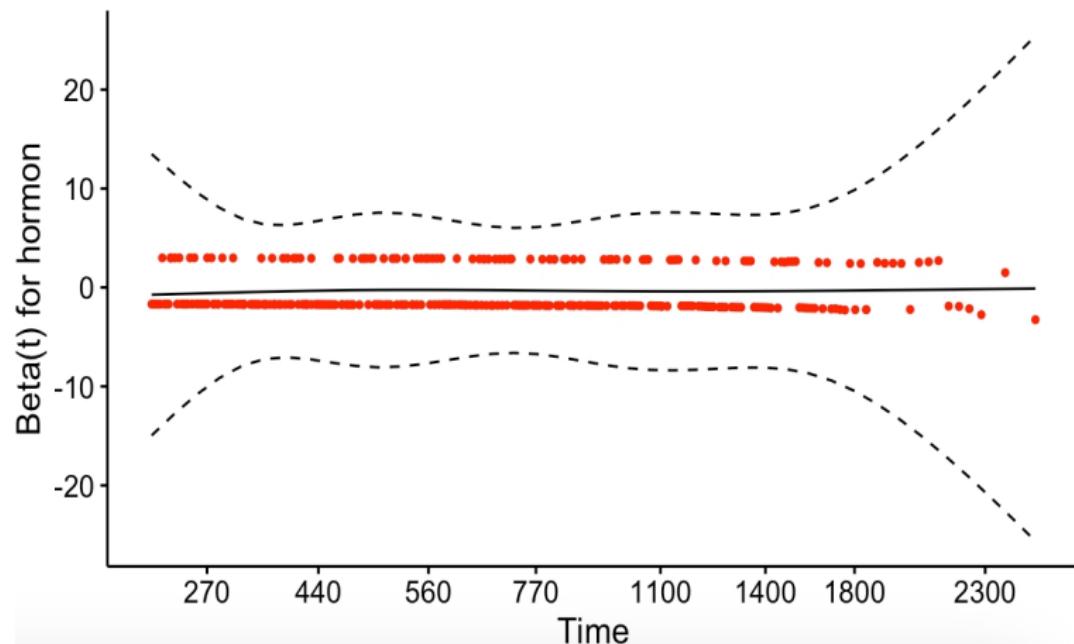
size

Schoenfeld Individual Test p: 0.2783



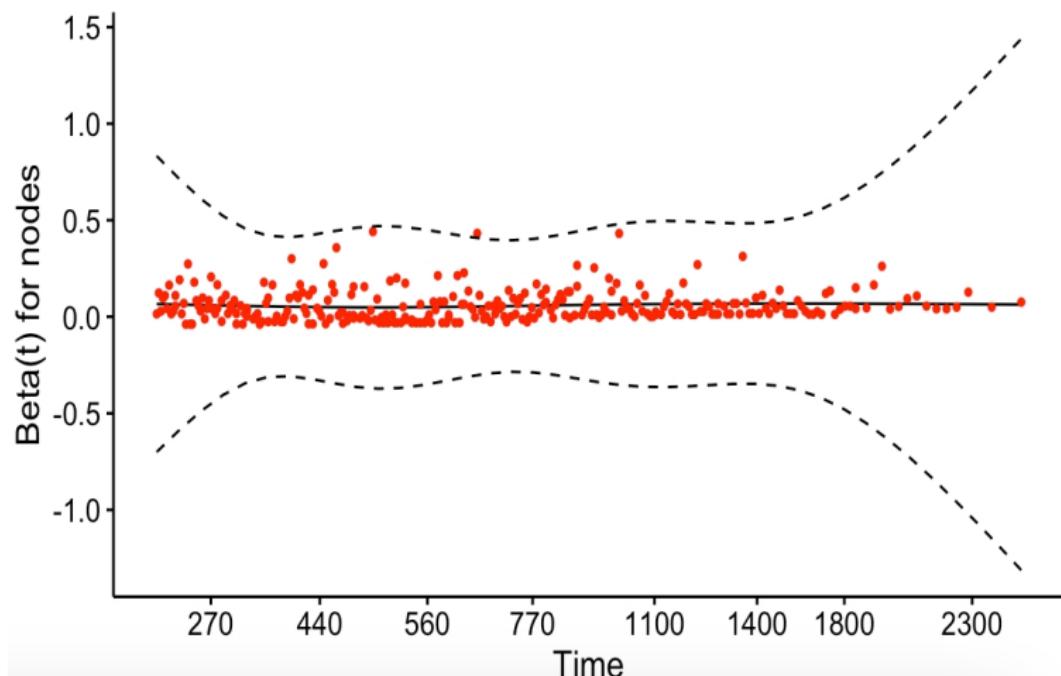
## nodes

Schoenfeld Individual Test p: 0.6334



# hormon

Schoenfeld Individual Test p: 0.3656



Introduction  
oo

K-M plot  
oooo

Local test  
oooooooooooo

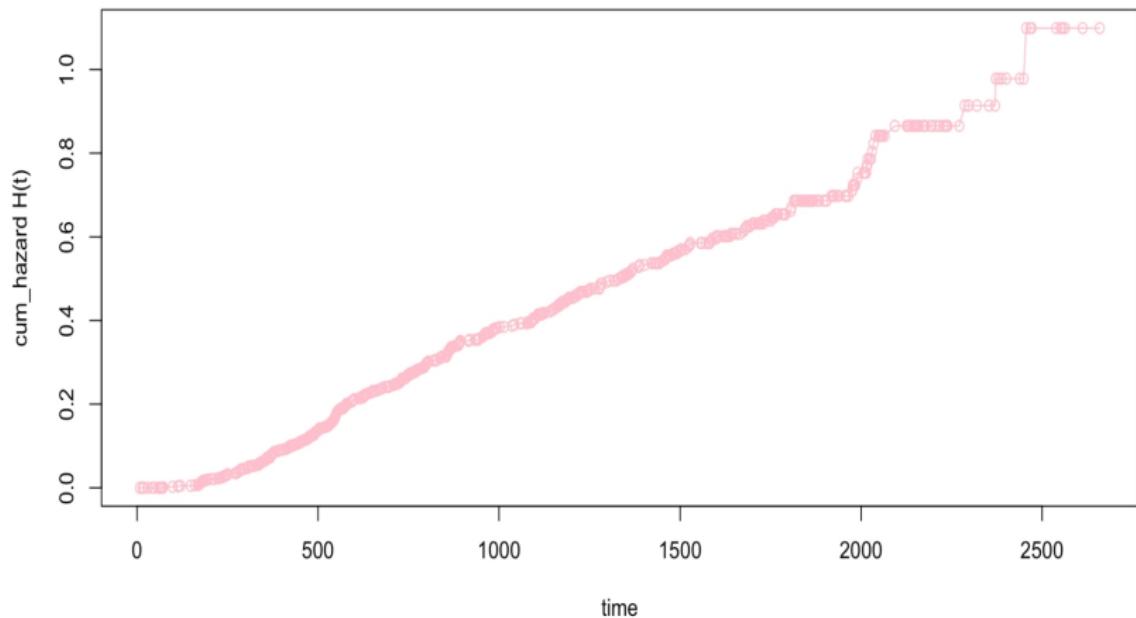
Schoenfeld residual  
ooo

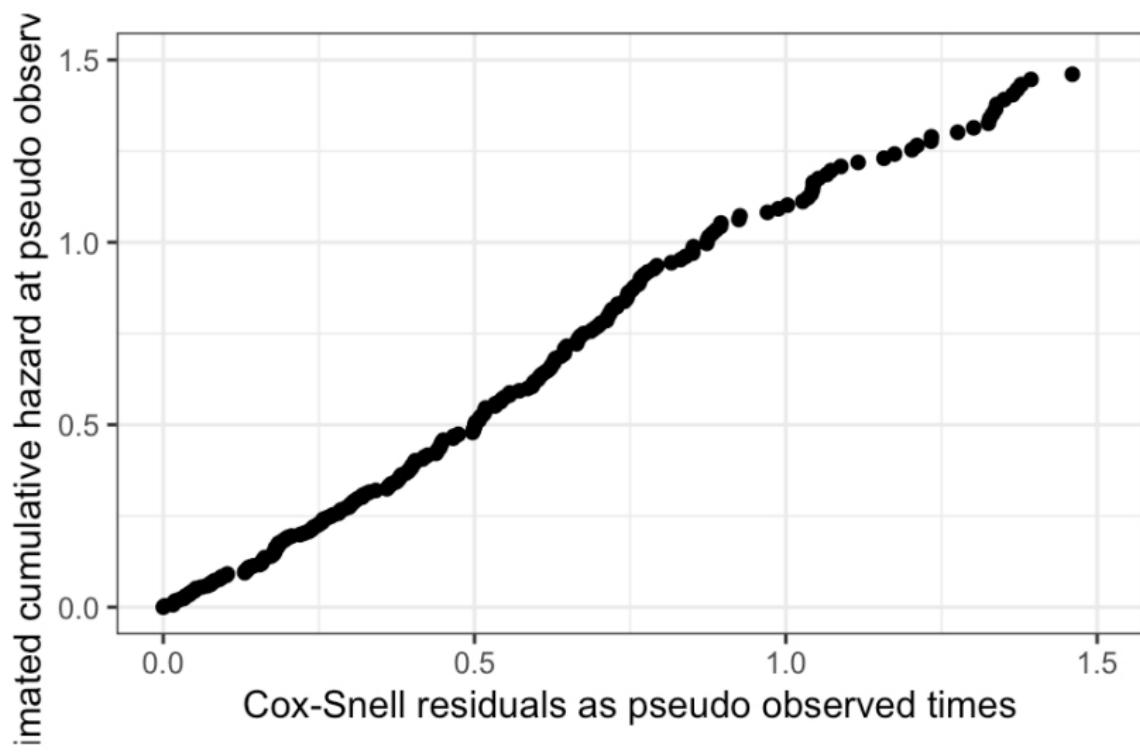
Breslow Baseline  
●

Cox-Snell residual  
o

Conclusion  
o

### Breslow's cumulative baseline hazard function





1. The risk increases by 1.306 times for each tumor grade increase.
2. For every 1 year increase in age, the risk increases by 0.9906516 times.
3. In the table, the wald test is very close to the Cox PH model, four are significant and four are not significant, but in terms of likelihood test and Score test, eight are significant. This may be something I have to figure out how to deal with later.

Source:

<https://www.kaggle.com/datasets/utkarshx27/breast-cancer-dataset-used-royston-and-altman>