The stratified Cox Procedure

Lisa Borsi, Marc Lickes & Lovro Soldo

28. March 2011

Recall:

1. Cox PH-Model

2. PH-Assumption

Recall:

1. Cox PH-Model
   - Model from Week 3

$$h(t, X) = h_0(t) e^{\sum_i \beta_i X_i}$$

2. PH-Assumption

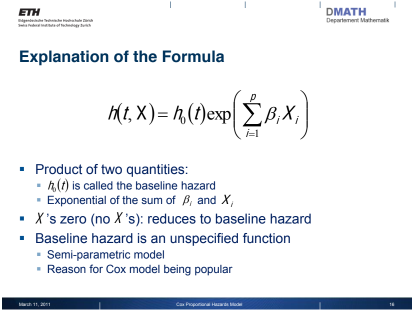


Figure: Definition from Presentation 3

Recall:

1. Cox PH-Model
   - Model from Week 3

$$h(t, X) = h_0(t)e^{\sum_i \beta_i X_i}$$

2. PH-Assumption
   - Hazard ration const. over time

$$HR := \frac{\hat{h}(t, X^*)}{\hat{h}(t, X)} = const$$

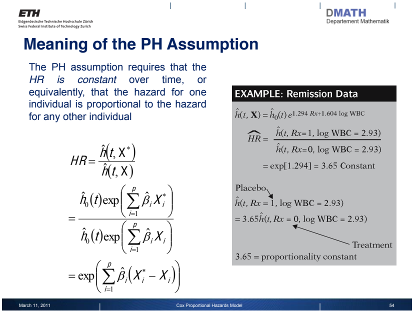


Figure: PH-Assumption from Presentation 3

Preview - bit more in detail

1. The stratified general Cox (SC) model
   - What to do if some variables do not satisfy the PH assumption?
   - Put them into stratas
2. No interaction assumption
   - Same β_i for each strata.
   - What if the β_i's are different for each strata?
3. Comparing: Likelihood, Wald & score test
   - What are LR, Wald & score test?
   - How are they related ?
4. Examples
5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?
6. Summary

The stratified general Cox (SC) model

The stratified general Cox (SC) model

The objects we consider,...

The Variables

- Denote by $X_1, ..., X_p$ for $p \in \mathbb{N}$ the variables satisfying the PH assumption.

The objects we consider,...

The Variables

- Denote by $X_1, ..., X_p$ for $p \in \mathbb{N}$ the variables satisfying the PH assumption.
- Denote by $Z_1, ..., Z_k$ for $k \in \mathbb{N}$ the variables not satisfying the PH assumption.

The objects we consider,...

The Variables

- Denote by $X_1, \ldots, X_p$ for $p \in \mathbb{N}$ the variables satisfying the PH assumption.
- Denote by $Z_1, \ldots, Z_k$ for $k \in \mathbb{N}$ the variables not satisfying the PH assumption.
- Define a single new variable Z^*
  1. Categorize each Z_i
  2. Form combinations of categories (strata)
  3. The strata are the categories of Z^*

The objects we consider,...

The Variables

- Denote by $X_1, ..., X_p$ for $p \in \mathbb{N}$ the variables satisfying the PH assumption.
- Denote by $Z_1, ..., Z_k$ for $k \in \mathbb{N}$ the variables not satisfying the PH assumption.
- Define a single new variable Z^*
  1. Categorize each Z_i
  2. Form combinations of categories (strata)
  3. The strata are the categories of Z^*
- k^* denotes the number of combinations, i.e. Z^* has k^* categories.

Illustrate this immediately with an example.

Illustrate this immediately with an example.

EXAMPLE

		Age		
		Young	Middle	Old
Treatment status	Placebo	1	2	3
	Treatment	4	5	6

Z^* = new variable with six categories
Stratify on Z^*

The model

The general SC model

The general Cox model is given by

$$h_g(t, X) = h_{0g} \exp[\beta_1 X_1 + \ldots + \beta_p X_p]$$

with $g = 1, \ldots, k^*$ strata defined from Z^*

A few observations

- The same coefficients: for each stratum $\beta_1, \ldots, \beta_p$.
- BUT: the baseline hazard functions $h_{0g}(t)$ may be different for each stratum.
- $X_1, \ldots, X_p$ directly included in the model, but Z^* appears only through the different baseline hazard functions.

The model

The general SC model

The general Cox model is given by

$$h_g(t, X) = h_{0g} \exp[\beta_1 X_1 + ... + \beta_p X_p]$$

with $g = 1, ..., k^*$ strata defined from Z^*

A few observations

- The same coefficients: for each stratum $\beta_1, ..., \beta_p$.
- BUT: the baseline hazard functions $h_{0g}(t)$ may be different for each stratum.
- $X_1, ..., X_p$ directly included in the model, but Z^* appears only through the different baseline hazard functions.

The model

The general SC model

The general Cox model is given by

$$h_g(t, X) = h_{0g} \exp[\beta_1 X_1 + \ldots + \beta_p X_p]$$

with $g = 1, \ldots, k^*$ strata defined from Z^*

A few observations

- The same coefficients: for each stratum $\beta_1, \ldots, \beta_p$.
- BUT: the baseline hazard functions $h_{0g}(t)$ may be different for each stratum.
- $X_1, \ldots, X_p$ directly included in the model, but Z^* appears only through the different baseline hazard functions.

The model

The general SC model

The general Cox model is given by

$$h_g(t, X) = h_{0g} \exp[\beta_1 X_1 + \ldots + \beta_p X_p]$$

with $g = 1, \ldots, k^*$ strata defined from Z^*

A few observations

- The same coefficients: for each stratum $\beta_1, \ldots, \beta_p$.
- BUT: the baseline hazard functions $h_{0g}(t)$ may be different for each stratum.
- $X_1, \ldots, X_p$ directly included in the model, but Z^* appears only through the different baseline hazard functions.

How to obtain the regression coefficients $\beta_1, ...\beta_p$

The estimation procedure

1. We calculate the likelihood functions for each strata

Strata:	$1, \ldots, k^*$
Likelihood:	$L_1, \ldots, L_{k^*}$
Hazard:	$h_1, \ldots, h_{k^*}$

2. We multiply all the likelihood function we have obtained this way.

$$L = L_1 \times \ldots \times L_{k^*}$$

3. Estimate $\hat{\beta}$ by max likelihood of $L(\beta)$.

Does the model really stand this ways?

Question:

- One fundamental assumption is missing! Which one?
- A hint: What do you know about the strata ?

Does the model really stand this ways?

Question:

- One fundamental assumption is missing! Which one?
- A hint: What do you know about the strata ?

Answer:

No interaction assumption between the strata and $X_1, ..., X_p$ is missing. Otherwise the model would look differently. More to this later.

Example: Leukemia data

Testing the PH assumption in R

Input:

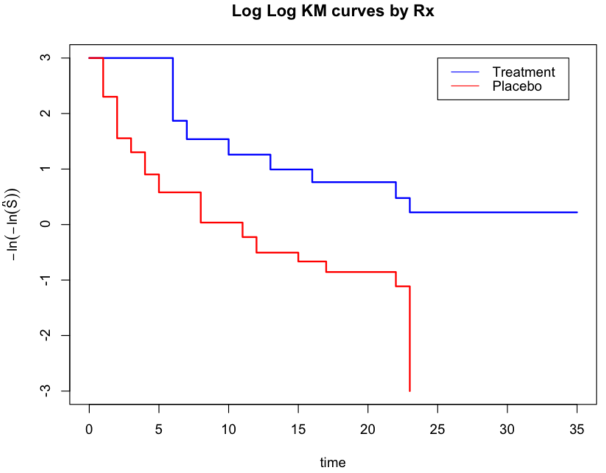
```
> cox2 <- coxph(Surv(t,status) ~ logWBC + Rx + sex,
 data=leuk, method="breslow")
> cox.zph(cox2)
```

Output:

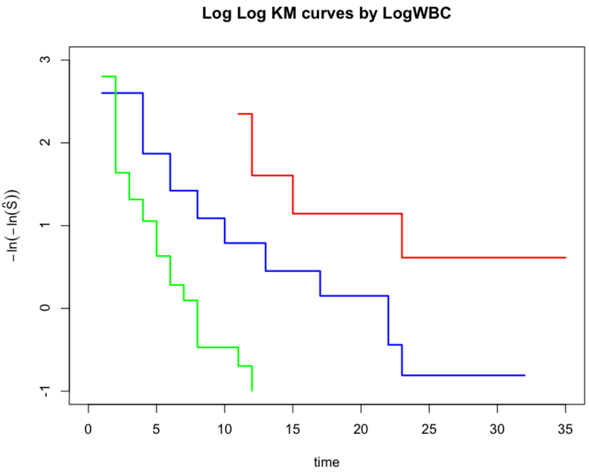
```
           rho chisq      p
logWBC  0.0657 0.189 0.6638
Rx     -0.1150 0.411 0.5214
sex    -0.3656 3.839 0.0501
GLOBAL      NA 3.969 0.2648
```

$0.0501 \approx 0.05$
$\Rightarrow$ PH assumption is violated for SEX

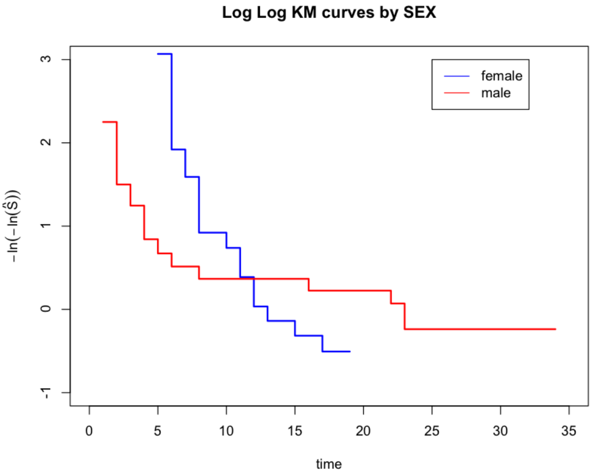
Example: Leukemia data



Example: Leukemia data



Example: Leukemia data



PH assumption not satisfied for SEX

Example: Leukemia data

PH assumption not satisfied for SEX

- use stratified Cox model
- control for SEX by stratification
- include logWBC and Rx in the model
- hazard function for stratified Cox model:
  $h_g(t, \mathbf{X}) = h_{0g}(t)\exp[\beta_1 \text{ logWBC} + \beta_2 \text{Rx}]$,
  $g = 1$ for females, $g = 2$ for males.

Example: Leukemia data

Input

```
> cox3 <- coxph(Surv(t,status) ~ Rx + logWBC + strata(sex), data=leuk,
method="breslow")
```

Output

```
          coef  exp(coef) se(coef)      z  Pr(>|z|)
logWBC 1.3902     4.0158   0.3376 4.118 3.83e-05 ***
Rx     0.9310     2.5370   0.4718 1.973   0.0485 *

       exp(coef) exp(-coef) lower .95 upper .95
logWBC     4.016     0.2490     2.072     7.783
Rx         2.537     0.3942     1.006     6.396
```

- Hazard ratio for the effect of Rx: $e^{\hat{\beta}_2} = e^{0.931} = 2.537$
- 95% confidence interval for Rx: $(1.006, 6.396)$
  formula: $exp(\hat{\beta}_2 \pm 1.96 \times \text{stand.error}(\hat{\beta}_2))$

Test significance of Rx adjusted for logWBC and SEX

R-output

```
> cox3$loglik[2]
[1] -57.55975
> cox4$loglik[2]
[1] -59.64846
> #Likelihood ratio statistic (significance of Rx)
> pchisq(-2*cox4$loglik[2]-(-2*cox3$loglik[2]), 1, lower.tail = FALSE)
[1] 0.040966
> cox3$coefficient[2]
0.9309646
> sqrt(diag(cox3$var))[2]
[1] 0.4717663
> pchisq((cox3$coefficient[2]/sqrt(diag(cox3$var))[2])^2, 1,
lower.tail = FALSE)
0.04845462
```

- LR and Wald tests lead to same conclusions

Let's have a closer look at the hazard functions

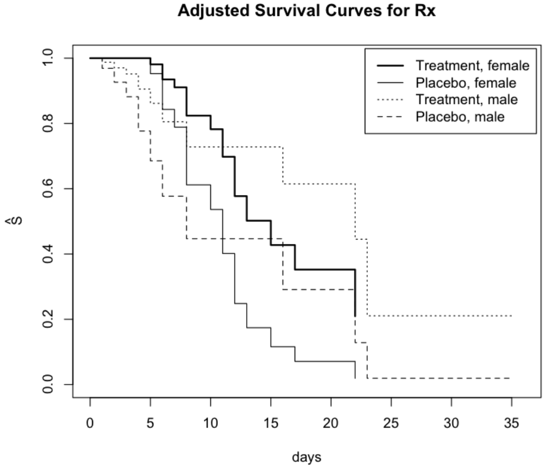
Stratified Cox model for females and males:

- Females (g = 1): $h_1(t, \mathbf{X}) = h_{01}(t)\exp[\beta_1 \text{logWBC} + \beta_2 \text{Rx}]$
- Males (g = 2): $h_2(t, \mathbf{X}) = h_{02}(t)\exp[\beta_1 \text{logWBC} + \beta_2 \text{Rx}]$

Note:

- h_{01} and h_{02} are different $\Rightarrow$ different survival curves for females and males
- β_1 and β_2 are the same $\Rightarrow$ same $\hat{\text{HR}}$'s for females and males $(e^{\hat{\beta}_1}, e^{\hat{\beta}_2})$
- in this model: no interaction between variables in the model and stratified variables

Example: Leukemia data



No interaction assumption

No interaction assumption

The interaction model

The interaction model

$$h_g(t) = h_{0g}(t)\exp[\beta_{1g}X_1 + \ldots + \beta_{pg}X_p] \quad (1)$$

The difference to the no-interaction model is that the coefficients β depend on the strata.

The interaction model

The interaction model

$$h_g(t) = h_{0g}(t)\exp[\beta_{1g}X_1 + ... + \beta_{pg}X_p] \tag{1}$$

The difference to the no-interaction model is that the coefficients β depend on the strata.

An alternative interaction model

$$h_g(t) = h_{0g}(t)[\beta_1^* X_1 + ... + \beta_p^* X_p + \sum_{j=1}^{k^*-1} \sum_{i=1}^{p} \beta_{ij}^* X_i Z_j^*] \tag{2}$$

where the β^* do not involve g.

The interaction models

Claim

The formulations (1) and (2) are equivalent.

Proof.

See blackboard, we derive the equivalence for the previous example. □

Likelihood ratio test

- The test statistic is given by

$$LR = -2\ln L_R - (-2\ln L_F)$$

  where L_R denotes the no interaction model, and L_F the interaction model.

- under H_0 : no-interaction, we have that

$$LR \sim \chi^2_{p(k^*-1)}$$

Comparing: Likelihood, Wald & score test

Comparing: Likelihood, Wald & score test

Comparing Likelihood, Wald & Score Test

In this Section we will compare the following tests, which are all based on Likelihood Theory.

1. Likelihood
2. Wald
3. Score

Assumptions and Notations:

As usual:

- X denotes the data
- $\theta = (\theta_1, \ldots, \theta_p)$ are the parameters we want to estimate.
- $L(\theta, X)$ will denote the likelihood or partial likelihood function.
- We maximize the likelihood (or log-likelihood) with $\hat{\theta}(X) = \hat{\theta}$.
- H_0 is $\theta = \theta_0$.

The definition of the score vector

Fisher's score vector $U(\theta)$.

The first derivative of the log-likelihood function is called *(Fisher's) score vector*,

$$U(\theta) = \frac{\partial}{\partial\theta} \ln L(\theta, X)$$

The definition of the score vector

Fisher's score vector $U(\theta)$.

The first derivative of the log-likelihood function is called *(Fisher's) score vector*,

$$U(\theta) = \frac{\partial}{\partial \theta} \ln L(\theta, X)$$

Remember: the derivative w.r.t. a vector is a vector, with entries:

$$U_i(\theta) = \tfrac{\partial}{\partial \theta_i} \ln(L(\theta, X)).$$

- For U we have

$$E_\theta[U(\theta)] = 0$$

- At a maximum $\hat{\theta}$ we have:

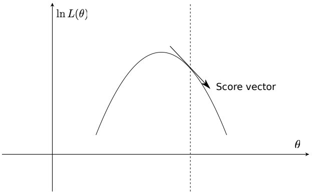
$$U(\hat{\theta}) = (0, \dots, 0)$$

To illustrate the meaning of the score vector

If we just have one parameter the score vector is the direction of the tangent at θ.

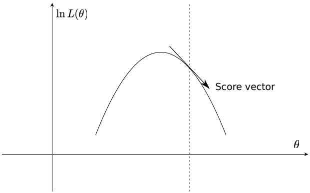
To illustrate the meaning of the score vector

If we just have one parameter the score vector is the direction of the tangent at θ.



To illustrate the meaning of the score vector

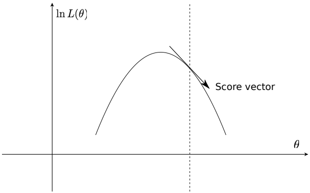
If we just have one parameter the score vector is the direction of the tangent at θ.



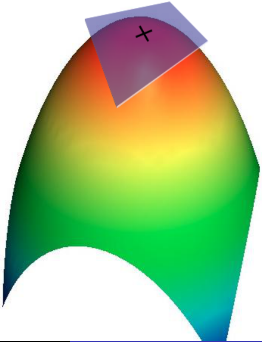
For two parameters we can think of the vectors as spanning the tangential plane.

To illustrate the meaning of the score vector

If we just have one parameter the score vector is the direction of the tangent at θ.

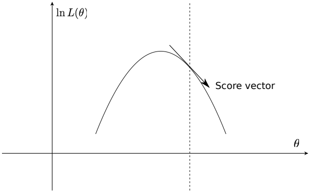


For two parameters we can think of the vectors as spanning the tangential plane.

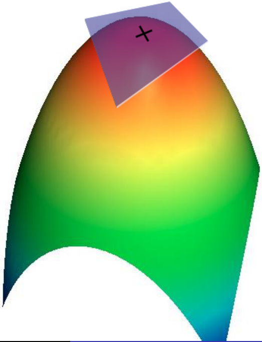


To illustrate the meaning of the score vector

If we just have one parameter the score vector is the direction of the tangent at θ.



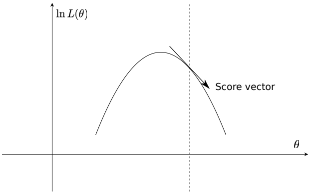
For two parameters we can think of the vectors as spanning the tangential plane.



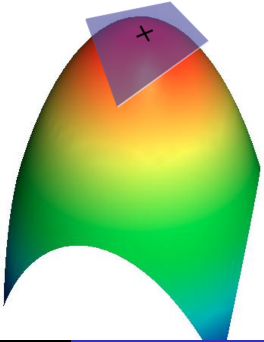
For more it's a tangential hyperplane.

To illustrate the meaning of the score vector

If we just have one parameter the score vector is the direction of the tangent at θ.



For two parameters we can think of the vectors as spanning the tangential plane.



For more it's a tangential hyperplane.

Not so easy to draw, but I guess you can imagine it! :)

Covariance matrix

We can also look at the second momentum, this is called:

The Fisher information matrix $i(\theta)$.

The Fisher information matrix is a $p \times p$ matrix defined by:

$$i(\theta) = E_\theta \left[U(\theta)^T U(\theta) \right] = \left\{ -E_\theta \left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln L(\theta, X) \right] \right\}_{\substack{i=1,\dots,p \\ j=1,\dots,p}}$$

Covariance matrix

We can also look at the second momentum, this is called:

The Fisher information matrix $i(\theta)$.

The Fisher information matrix is a $p \times p$ matrix defined by:

$$i(\theta) = E_\theta\left[U(\theta)^T U(\theta)\right] = \left\{-E_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln L(\theta, X)\right]\right\}_{\substack{i=1,\dots,p\\ j=1,\dots,p}}$$

Mostly this is to hard to compute, thus we estimate i with:

$$\hat{i}(\theta) =: I_{i,j}(\theta) := -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln L(\theta, X), \quad i = 1, \dots, p, \ j = 1, \dots, p$$

Who was Fisher?

Figure: R. A. Fisher (1890-1962)

- One of the most important theoretical biologists, geneticists, evolutionary theorists and statistician of the 20th Century.
- Founder of maximum likelihood principle and variance statistics
- Anders Hald about Fisher:

  a genius who almost single-handedly created the foundations for modern statistical science

Who was Fisher?

Figure: R. A. Fisher (1890-1962)

- One of the most important theoretical biologists, geneticists, evolutionary theorists and statistician of the 20th Century.
- Founder of maximum likelihood principle and variance statistics
- Anders Hald about Fisher:

  a genius who almost single-handedly created the foundations for modern statistical science

Who is Cox?

- b.1924 Birmingham, England
- 1988: member of the Department of Statistics at Oxford University
- 1990: won the Kettering Prize and Gold Medal for Cancer Research for "the development of the Proportional Hazard Regression Model."
- 1994 retired

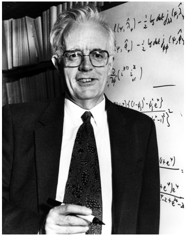


Figure: D. R. Cox (1924 -)

We are ready to compare the tests

We already know the *likelihood ratio test* statistic:

$$LR := -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right)$$

We are ready to compare the tests

We already know the *likelihood ratio test* statistic:

$$LR := -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right)$$

The Wald test statistic is defined as:

$$W := (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0)$$

We are ready to compare the tests

We already know the *likelihood ratio test* statistic:

$$LR := -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right)$$

The Wald test statistic is defined as:

$$W := (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0)$$

And the score test uses the Fisher score vector:

$$S := U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0)$$

We are ready to compare the tests

We already know the *likelihood ratio test* statistic:

$$LR := -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right) \rightarrow \chi^2_p$$

The Wald test statistic is defined as:

$$W := (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \rightarrow \chi^2_p$$

And the score test uses the Fisher score vector:

$$S := U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0) \rightarrow \chi^2_p$$

Note: They all converge to a χ^2_p distribution.
The advantage of score test: it does not depend on $\hat{\theta}$.

Let's look at an example

Given:

- Censored sample of size n of exponentially dying population with hazard rate λ.
- observed survival time of individual i is T_i with variable for failure δ_i.

Test for null hypothesis: $\lambda_0 = 1$.

Let's look at an example

Given:

- Censored sample of size n of exponentially dying population with hazard rate λ.
- observed survival time of individual i is T_i with variable for failure δ_i.

Test for null hypothesis: $\lambda_0 = 1$.

The Likelihood function is given by:

$$L(\lambda, (T_i, \delta_i)) = \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda T_i} = \lambda^D e^{-\lambda T}$$

where $T = \sum_{i=1}^{n} T_i$, is the total time
and $D = \sum_{i=1}^{n} \delta_i$ is the observed number of deaths.

Calculate the important functions

First we calculate:

$$\begin{aligned}
\text{Log Likelihood:} \quad \ln L(\lambda) &= D\ln(\lambda) - \lambda T \\
\text{Score vector:} \quad U(\lambda) &= \tfrac{d}{d\lambda}\ln L(\lambda) = \tfrac{D}{\lambda} - T \\
\text{Fisher vector:} \quad I(\lambda) &= -\tfrac{d^2}{d\lambda^2}\ln L(\lambda) = \tfrac{D}{\lambda^2}
\end{aligned} \tag{3}$$

Calculate the important functions

First we calculate:

$$\begin{array}{rrcl} \text{Log Likelihood:} & \ln L(\lambda) & = & D\ln(\lambda) - \lambda T \\ \text{Score vector:} & U(\lambda) & = & \frac{D}{\lambda} - T \\ \text{Fisher vector:} & I(\lambda) & = & \frac{D}{\lambda^2} \end{array} \tag{3}$$

We compute the maximum likelihood estimator, from (3).

$$\frac{D}{\lambda} - T = 0 \rightsquigarrow \hat{\lambda} = \frac{D}{T}$$

Comparing the tests:

Likelihood ratio test:

$$
\begin{aligned}
LR &= -2\left(\ln L(\lambda_0, X) - \ln L(\hat{\lambda}, X)\right) \\
&= -2\left((D\ln(1) - 1\cdot T) - (D\ln(\tfrac{D}{T} - \tfrac{D}{T}\cdot S)\right) \\
&= 2\left(T - D + D\ln(\tfrac{D}{T})\right)
\end{aligned}
$$

Comparing the tests:

Likelihood ratio test:

$$\begin{aligned} LR &= -2\left(\ln L(\lambda_0, X) - \ln L(\hat{\lambda}, X)\right) \\ &= -2\left((D\ln(1) - 1\cdot T) - \left(D\ln(\tfrac{D}{T} - \tfrac{D}{T}\cdot S\right)\right) \\ &= 2\left(T - D + D\ln(\tfrac{D}{T})\right) \end{aligned}$$

Wald test:

$$\begin{aligned} W &= (\hat{\lambda} - \lambda_0)^T I(\hat{\lambda})(\hat{\lambda} - \lambda_0) \\ &= \left(\tfrac{D}{T} - 1\right) \tfrac{D}{(D/S)^2} \left(\tfrac{D}{T} - 1\right) = \tfrac{(D-T)^2}{D} \end{aligned}$$

Comparing the tests:

Likelihood ratio test:

$$\begin{aligned} LR &= -2\left(\ln L(\lambda_0, X) - \ln L(\hat{\lambda}, X)\right) \\ &= -2\left((D\ln(1) - 1\cdot T) - (D\ln(\tfrac{D}{T} - \tfrac{D}{T}\cdot S)\right) \\ &= 2\left(T - D + D\ln(\tfrac{D}{T})\right) \end{aligned}$$

Wald test:

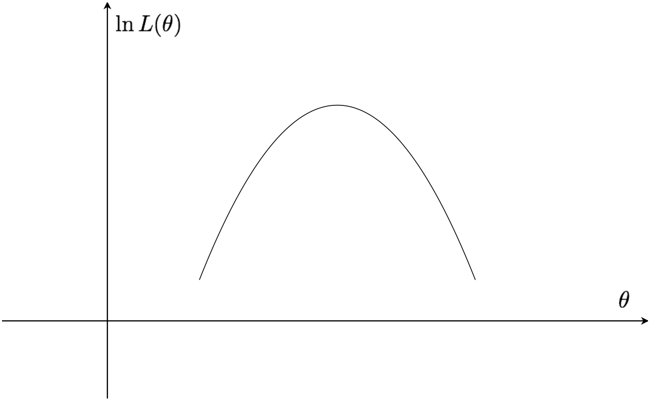
$$\begin{aligned} W &= (\hat{\lambda} - \lambda_0)^T I(\hat{\lambda})(\hat{\lambda} - \lambda_0) \\ &= \left(\tfrac{D}{T} - 1\right)\tfrac{D}{(D/S)^2}\left(\tfrac{D}{T} - 1\right) = \tfrac{(D-T)^2}{D} \end{aligned}$$

Score test:

$$\begin{aligned} S &= U(\lambda_0)^T I^{-1}(\lambda_0) U(\lambda_0) \\ &= \left(\tfrac{D}{T} - 1\right)\tfrac{1}{D^2}\left(\tfrac{D}{T} - 1\right) = \tfrac{(D-T)^2}{D} \end{aligned}$$

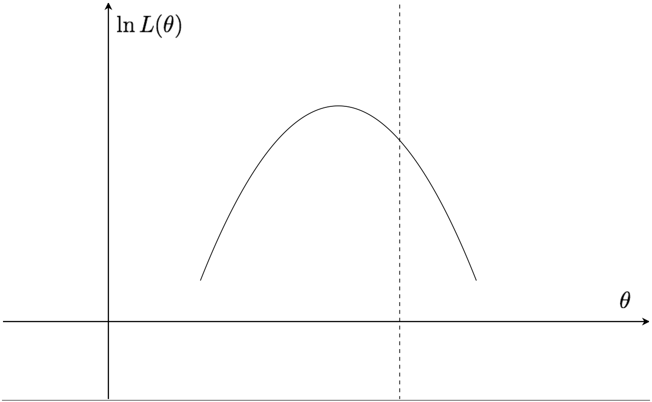
Comparing the tests "intuitively".

Assume the plot of $\ln L(\theta)$ against θ looks like this:



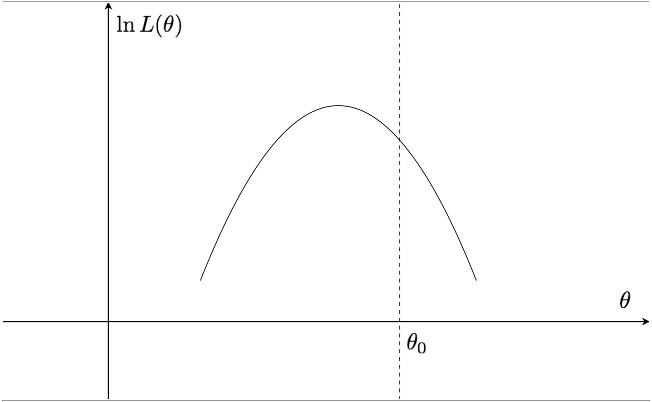
Comparing the tests "intuitively".

What would be this value?



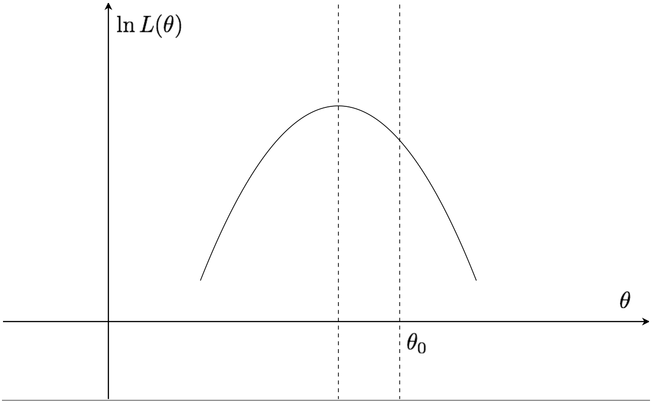
Comparing the tests "intuitively".

Correct. It's θ_0.



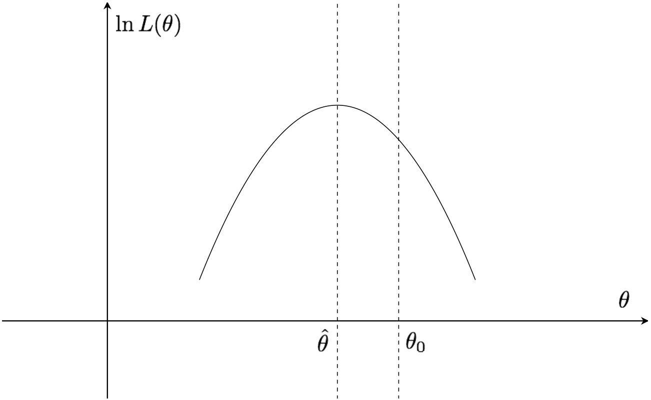
Comparing the tests "intuitively".

And what about this one?



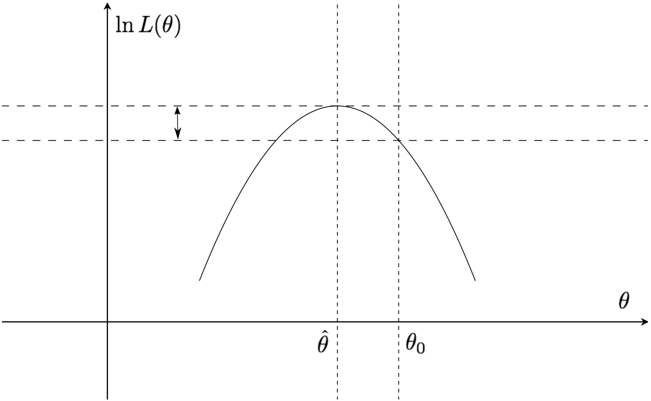
Comparing the tests "intuitively".

Correct. It's $\hat{\theta}$. Maximizing $\ln L(\theta)$.



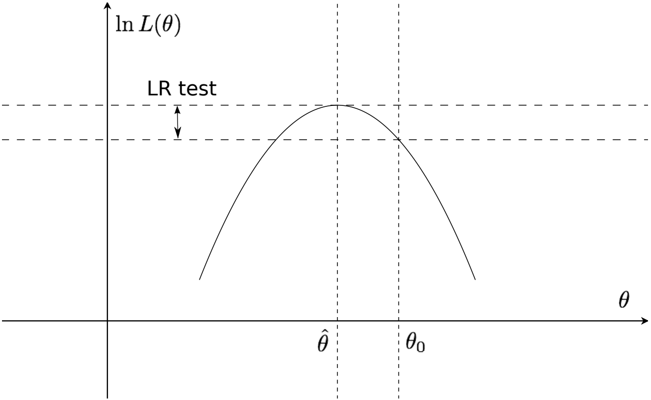
Comparing the tests "intuitively".

What would be this quantity?



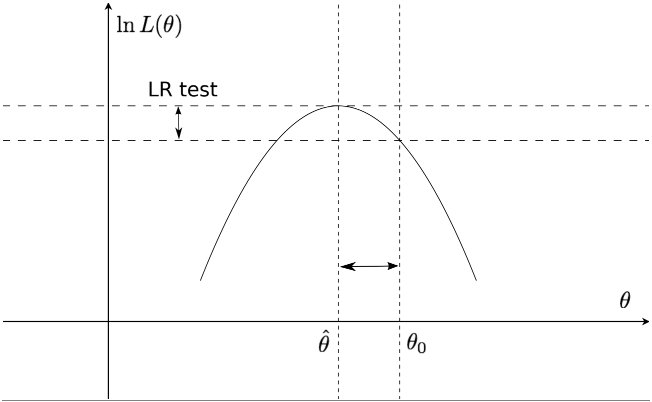
Comparing the tests "intuitively".

That's the Likelihood ratio test. Comparing $\ln L(\theta_0)$ with $\ln L(\hat{\theta})$.



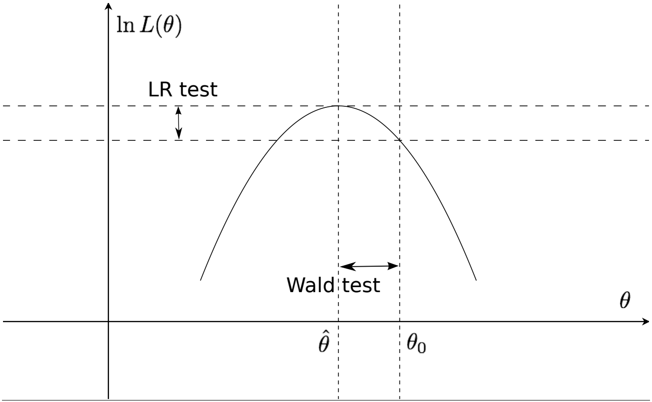
Comparing the tests “intuitively”.

And this one?



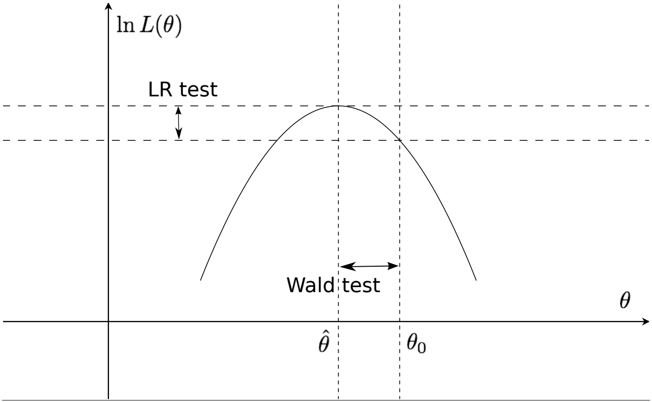
Comparing the tests "intuitively".

That's it. The Wald test.



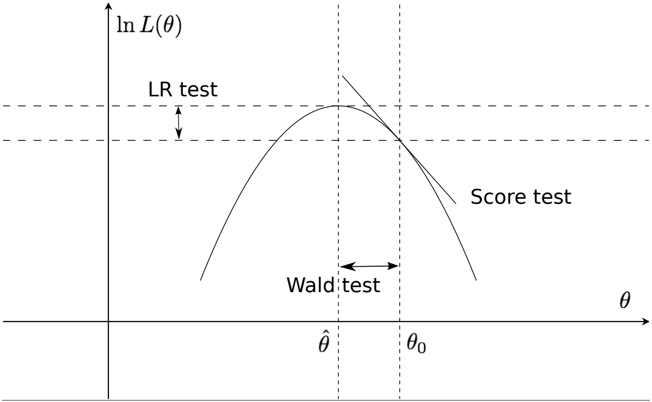
Comparing the tests “intuitively”.

Where would be the score test?



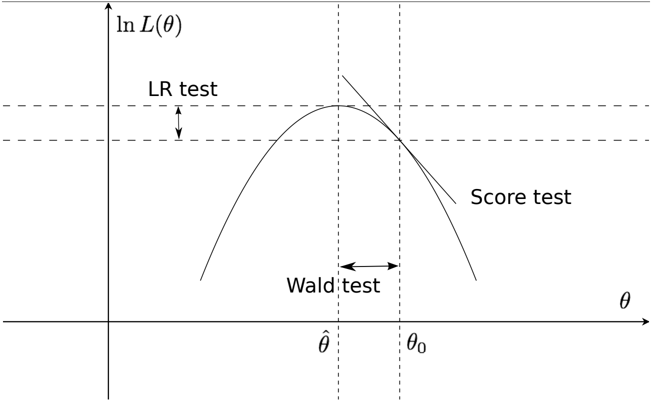
Comparing the tests "intuitively".

Great.



Comparing the tests "intuitively".

Great. We see how the tests are "related" to the likelihood function.



Example

Example

Example involving several stratification variables

Data:

- survival times in days for 137 lung cancer patients
- exposure variable of interest: treatment status (standard=1, test=2)
- control variables: cell type (4 types), performance status, disease duration, age, prior therapy status

How to proceed:

- check PH assumption
- stratify on the variables not satisfying PH assumption
- check no-interaction assumption: fit interaction model and test for significance

Example involving several stratification variables

	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf. Interval]		P(PH)
Treatment	0.290	0.207	0.162	1.336	0.890	2.006	0.628
Large cell	0.400	0.283	0.157	1.491	0.857	2.594	0.033
Adeno cell	1.188	0.301	0.000	3.281	1.820	5.915	0.081
Small cell	0.856	0.275	0.002	2.355	1.374	4.037	0.078
Perf. Stat	–0.033	0.006	0.000	0.968	0.958	0.978	0.000
Dis. Durat.	0.000	0.009	0.992	1.000	0.982	1.018	0.919
Age	–0.009	0.009	0.358	0.991	0.974	1.010	0.198
Pr. Therapy	0.007	0.023	0.755	1.007	0.962	1.054	0.145

No. of subjects = 137 Log likelihood = –475.180

- cell type and performance status do not verify PH assumption
- use stratified Cox model: stratify on cell type and perf_stat

Example involving several stratification variables

stratified Cox analysis:

- stratify on 2 variables: cell type and performance status
- new categorical variable Z^*
- categories of Z^* = all possible combinations of categories of cell type and performance status
- cell type: 4 categories
- performance status $\in (0, 100)$, but need categorical variable

$$\Rightarrow \text{ new variable PSbin } = \begin{cases} 1 & \text{if performance } \geq 60, \\ 0 & \text{otherwise.} \end{cases}$$

- Z^* has $k^* = 4 \times 2 = 8$ categories

Example involving several stratification variables

Stratified Cox analysis in R:

Input

```
cox_3 <- coxph(Surv(time,status) ~ treatment + age +
strata(large_cell,small_cell,adeno_cell,PSbin),
data=vets, method="breslow")
summary(cox_3)
```

Output

```
               coef  exp(coef)  se(coef)      z  Pr(>|z|)
treatment  0.125398  1.133599  0.208471  0.602    0.547
age       -0.001307  0.998694  0.010103 -0.129    0.897

          exp(coef) exp(-coef) lower .95 upper .95
treatment    1.1336     0.8821    0.7534     1.706
age          0.9987     1.0013    0.9791     1.019
```

Example involving several stratification variables

No-interaction model

- $h_g(t, \mathbf{X}) = h_{0g}(t)\exp[\beta_1\text{Treatment}+\beta_2\text{Age}]$, $g = 1, 2, \ldots, 8$

Is an interaction model appropriate?

- define an interaction model
- test $\text{H}_0=$ no-interaction model is acceptable by LR test

Example: interaction model

Hazard function for interaction model:

$$\begin{aligned} h_g(t,\mathbf{X}) &= h_{0g}(t)\exp[\beta_1 \text{Treatment} + \beta_2 \text{Age} \\ &+\beta_{11}(Z_1^* \times \text{Treatment}) + \cdots + \beta_{17}(Z_7^* \times \text{Treatment}) \\ &+\beta_{21}(Z_1^* \times \text{Age}) + \cdots + \beta_{27}(Z_7^* \times \text{Age})] \end{aligned}$$

where $Z_1^* =$ large cell, $Z_2^* =$ small cell, $Z_3^* =$ adeno cell, $Z_4^* =$ PSbin, $Z_5^* = Z_1^* \times Z_4^*$, $Z_6^* = Z_2^* \times Z_4^*$ and $Z_7^* = Z_3^* \times Z_4^*$

Example: interaction model

Input

```
cox_4 <- coxph(Surv(time,status) ~
        treatment*strata(large_cell,small_cell,adeno_cell,PSbin)+
        age*strata(large_cell,small_cell,adeno_cell,PSbin),
        data=vets, method="breslow")
```

Test for significance of the interaction model

Likelihood ratio (LR) test

- H_0: no-interaction model acceptable,
  i.e. β's of interaction terms are all zero
- $\text{LR} = -2(\text{LR}_R - \text{LR}_F)$
  $R =$ no-interaction model, $F =$ interaction model
- $\text{LR} \sim \chi^2_{14}$ (14 coefficients of interaction terms)
- if H_0 rejected then interaction model is appropriate

Test for significance of the interaction model

Likelihood ratio test in R:

```
> red$loglik[2]
[1] -262.0195
> full$loglik[2]
[1] -249.972
> pchisq(-2*(red$loglik[2]-full$loglik[2]), 14, lower.tail=FALSE)
[1] 0.04462534
```

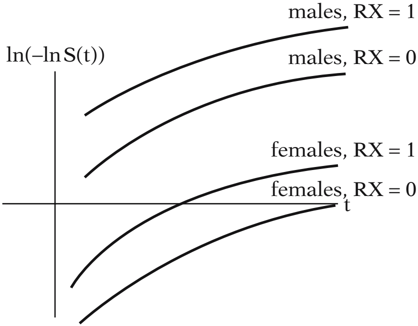
$\Rightarrow$ no-interaction assumption is rejected at 5% level
so interaction model is preferred

Graphical Comparison of the models so far

Graphical Comparison of the models so far

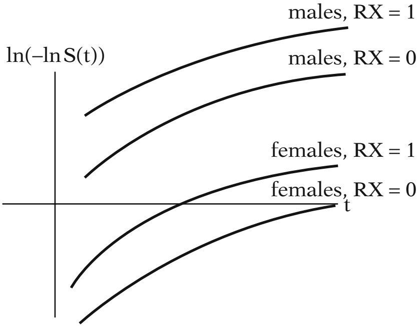
Graphical View - Example 1

- PH assumption? For which variable?
- Interaction?



Graphical View - Example 1

- PH assumption? For which variable?
  - PH for both RX and SEX, we see that since the curves are parallel.
- Interaction?

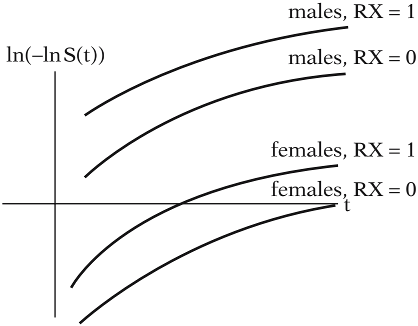


$$h_0(t)e^{\beta_1 RX+\beta_2 SEX}$$

Graphical View - Example 1

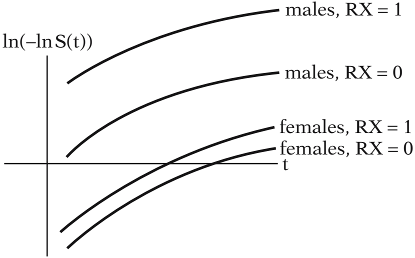
- PH assumption? For which variable?

- Interaction?
  - No interaction, because they have the same distance.



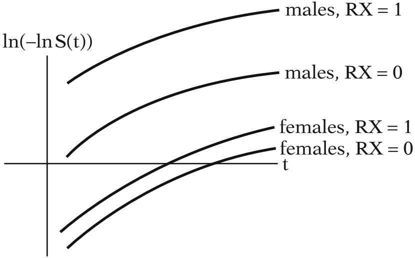
Graphical View - Example 2

- PH assumption? For which variable?
- Interaction?



Graphical View - Example 2

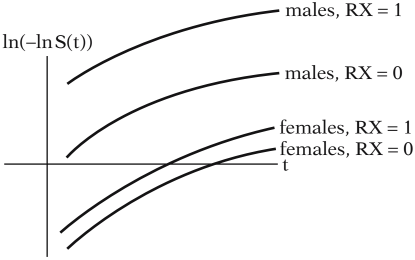
- PH assumption? For which variable?
- PH for both RX and SEX (parallel)
- Interaction?



$$h_0(t)e^{\beta_1 RX+\beta_2 SEX+\beta_3 RX\times SEX}$$

Graphical View - Example 2

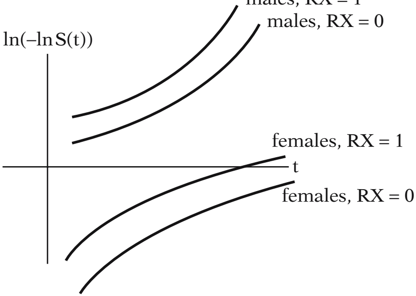
- PH assumption? For which variable?
  - PH for both RX and SEX (parallel)
- Interaction?
  - Between treatment and SEX
  - Effect for the treatment is larger for males then females. Bigger distance between RX=1 to RX=0.



$$h_0(t)e^{\beta_1 RX + \beta_2 SEX + \beta_3 RX \times SEX}$$

Graphical View - Example 3

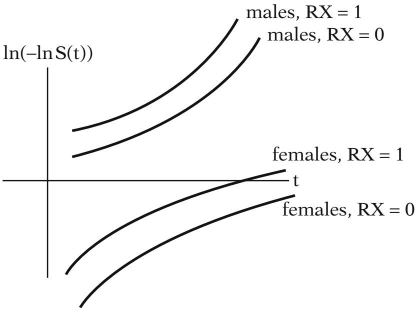
- PH assumption?
- Interaction?



Graphical View - Example 3

- PH assumption?

- PH assumption for RX. Stratified, for SEX.



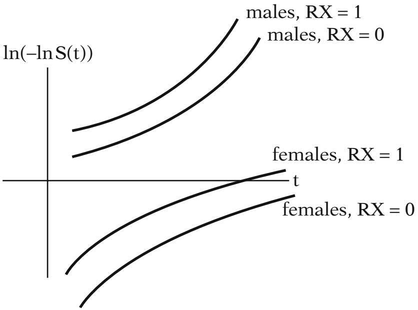
- Interaction?

$$h_{00}(t)e^{\beta_1 RX},$$

$$h_{01}(t)e^{\beta_1 RX}$$

Graphical View - Example 3

- PH assumption?
  - PH assumption for RX. Stratified, for SEX.
  - Notice the strata are no longer parallel.
- Interaction?

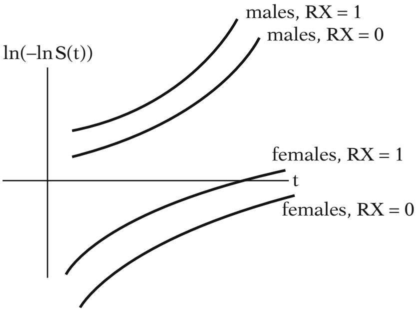


$$h_{00}(t)e^{\beta_1 RX},$$

$$h_{01}(t)e^{\beta_1 RX}$$

Graphical View - Example 3

- PH assumption?
  - PH assumption for RX. Stratified, for SEX.
  - Notice the strata are no longer parallel.
- Interaction?
  - Distance between RX=1 and RX=0 is the same indicating no interaction between RX and SEX.

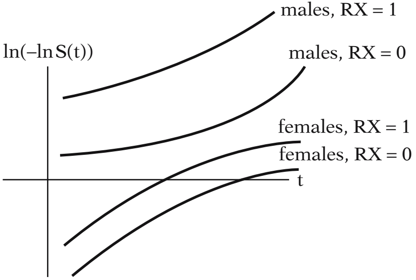


$$h_{00}(t)e^{\beta_1 RX},$$

$$h_{01}(t)e^{\beta_1 RX}$$

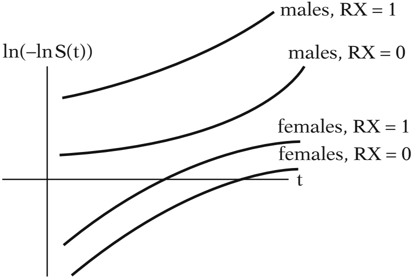
Graphical View - Example 4

- PH assumption?
- Interaction?



Graphical View - Example 4

- PH assumption?
- PH assumption for RX. Stratified, for SEX.
- Interaction?

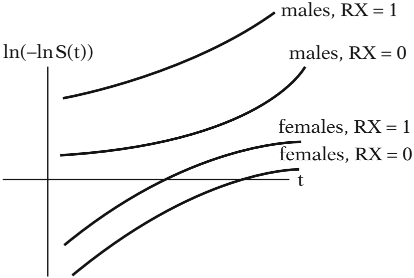


$$h_{00}(t)e^{\beta_1 RX+\beta_2 RX\times SEX},$$

$$h_{01}(t)e^{\beta_1 RX+\beta_2 RX\times SEX}$$

Graphical View - Example 4

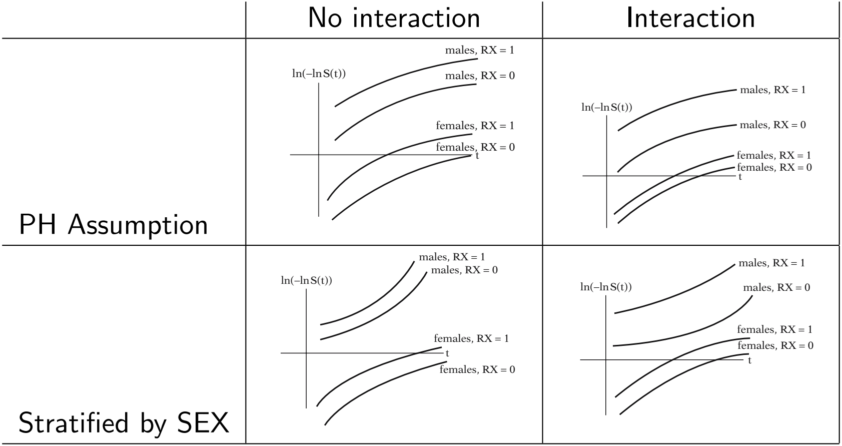
- PH assumption?
  - PH assumption for RX. Stratified, for SEX.
- Interaction?
  - Interaction between RX and SEX, bigger distance for RX=1 and RX=0.



$$h_{00}(t)e^{\beta_1 RX+\beta_2 RX\times SEX},$$

$$h_{01}(t)e^{\beta_1 RX+\beta_2 RX\times SEX}$$

Overview of the examples



Summary

Summary

Summary

1. The stratified general Cox (SC) model
   - What to do if some variables do not satisfy the PH assumption?

   ⇒ Put them into stratas - general SC model:

   $$h_g(t,X) = h_{0g}\exp[\beta_1 X_1 + ... + \beta_p X_p]$$

2. No interaction assumption
   - Same β_i for each strata.
   - What if the β_i's are different for each strata?

   ⇒ Test the no-interaction model against the full (interaction) model:

   $$h_g(t) = h_{g0}(t)[\beta_1^* X_1 + ... + \beta_p^* X_p + \sum_{j=1}^{k*-1}\sum_{i=1}^{p}\beta_{ij}^* X_i Z_j^*]$$

   (using log-likelihood, Wald or score test).

Summary

1. The stratified general Cox (SC) model
   - What to do if some variables do not satisfy the PH assumption?

   $\Rightarrow$ Put them into stratas - general SC model:

$$h_g(t, X) = h_{0g} \exp[\beta_1 X_1 + ... + \beta_p X_p]$$

2. No interaction assumption
   - Same β_i for each strata.
   - What if the β_i's are different for each strata?

   $\Rightarrow$ Test the no-interaction model against the full (interaction) model:

$$h_g(t) = h_{g0}(t)[\beta_1^* X_1 + ... + \beta_p^* X_p + \sum_{j=1}^{k*-1} \sum_{i=1}^{p} \beta_{ij}^* X_i Z_j^*]$$

(using log-likelihood, Wald or score test).

Summary

1. The stratified general Cox (SC) model
   - What to do if some variables do not satisfy the PH assumption?

   ⇒ Put them into stratas - general SC model:

$$h_g(t,X) = h_{0g} \exp[\beta_1 X_1 + ... + \beta_p X_p]$$

2. No interaction assumption
   - Same β_i for each strata.
   - What if the β_i's are different for each strata?

   ⇒ Test the no-interaction model against the full (interaction) model:

$$h_g(t) = h_{g0}(t)[\beta_1^* X_1 + ... + \beta_p^* X_p + \sum_{j=1}^{k*-1} \sum_{i=1}^{p} \beta_{ij}^* X_i Z_j^*]$$

(using log-likelihood, Wald or score test).

Summary

1. The stratified general Cox (SC) model
   - What to do if some variables do not satisfy the PH assumption?

   $\Rightarrow$ Put them into stratas - general SC model:

$$h_g(t,X) = h_{0g}\exp[\beta_1 X_1 + \ldots + \beta_p X_p]$$

2. No interaction assumption
   - Same β_i for each strata.
   - What if the β_i's are different for each strata?

   $\Rightarrow$ Test the no-interaction model against the full (interaction) model:

$$h_g(t) = h_{g0}(t)[\beta_1^* X_1 + \ldots + \beta_p^* X_p + \sum_{j=1}^{k*-1}\sum_{i=1}^{p}\beta_{ij}^* X_i Z_j^*]$$

   (using log-likelihood, Wald or score test).

Summary

3. Comparing: Likelihood, Wald & score test
   - What are LR, Wald & score test?

$\Rightarrow$ Definitions:

$$
\begin{aligned}
LR &:= -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right) \to \chi_p^2 \\
W &:= (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \to \chi_p^2 \\
S &:= U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0) \to \chi_p^2
\end{aligned}
$$

   - How are they "related" ?

Summary

3. Comparing: Likelihood, Wald & score test
   - What are LR, Wald & score test?

   ⇒ Definitions:

$$
\begin{aligned}
LR &:= -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right) \rightarrow \chi_p^2 \\
W &:= (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \rightarrow \chi_p^2 \\
S &:= U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0) \rightarrow \chi_p^2
\end{aligned}
$$

   - How are they "related" ?

Summary

3. Comparing: Likelihood, Wald & score test
   - What are LR, Wald & score test?

   ⇒ Definitions:

$$
\begin{aligned}
LR &:= -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right) \rightarrow \chi_p^2 \\
W &:= (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \rightarrow \chi_p^2 \\
S &:= U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0) \rightarrow \chi_p^2
\end{aligned}
$$

   - How are they "related" ?

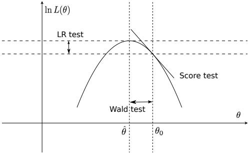
Summary

3. Comparing: Likelihood, Wald & score test
   - What are LR, Wald & score test?

   ⇒ Definitions:

$$
\begin{aligned}
LR &:= -2\left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X)\right) \rightarrow \chi^2_p \\
W &:= (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \rightarrow \chi^2_p \\
S &:= U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0) \rightarrow \chi^2_p
\end{aligned}
$$

   - How are they "related" ?



Summary

4. Example
   - Sum up of theory: survival times of 137 lung cancer patients

   $\Rightarrow$ check PH assumption

   $\Rightarrow$ stratify on the variables not satisfying PH assumption

   $\Rightarrow$ check no-interaction assumption
5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?

   $\Rightarrow$ If the curves within one stratus are parallel, the model assumed the PH assumption.

   $\Rightarrow$ If the distance between different strata is the same, the model assumed no interaction.

Summary

4. Example
   - Sum up of theory: survival times of 137 lung cancer patients

   ⇒ check PH assumption

   ⇒ stratify on the variables not satisfying PH assumption

   ⇒ check no-interaction assumption

5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?

   ⇒ If the curves within one stratus are parallel, the model assumed the PH assumption.

   ⇒ If the distance between different strata is the same, the model assumed no interaction.

Summary

4. Example
   - Sum up of theory: survival times of 137 lung cancer patients

   ⇒ check PH assumption

   ⇒ stratify on the variables not satisfying PH assumption

   ⇒ check no-interaction assumption

5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?

   ⇒ If the curves within one stratus are parallel, the model assumed the PH assumption.

   ⇒ If the distance between different strata is the same, the model assumed no interaction.

Summary

4. Example
   - Sum up of theory: survival times of 137 lung cancer patients
   - $\Rightarrow$ check PH assumption
   - $\Rightarrow$ stratify on the variables not satisfying PH assumption
   - $\Rightarrow$ check no-interaction assumption
5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?
   - $\Rightarrow$ If the curves within one stratus are parallel, the model assumed the PH assumption.
   - $\Rightarrow$ If the distance between different strata is the same, the model assumed no interaction.

Summary

4. Example
   - Sum up of theory: survival times of 137 lung cancer patients

   ⇒ check PH assumption

   ⇒ stratify on the variables not satisfying PH assumption

   ⇒ check no-interaction assumption

5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?

   ⇒ If the curves within one stratus are parallel, the model assumed the PH assumption.

   ⇒ If the distance between different strata is the same, the model assumed no interaction.

Summary

4. Example
   - Sum up of theory: survival times of 137 lung cancer patients
   - ⇒ check PH assumption
   - ⇒ stratify on the variables not satisfying PH assumption
   - ⇒ check no-interaction assumption
5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?
   - ⇒ If the curves within one stratus are parallel, the model assumed the PH assumption.
   - ⇒ If the distance between different strata is the same, the model assumed no interaction.

Summary

4. Example
   - Sum up of theory: survival times of 137 lung cancer patients

   $\Rightarrow$ check PH assumption

   $\Rightarrow$ stratify on the variables not satisfying PH assumption

   $\Rightarrow$ check no-interaction assumption
5. Graphical comparison of the tests so far
   - Could we recognize the different models in their $\ln(-\ln S)$ plot?

   $\Rightarrow$ If the curves within one stratus are parallel, the model assumed the PH assumption.

   $\Rightarrow$ If the distance between different strata is the same, the model assumed no interaction.

Thank you for your attention.

Questions?

References - further reading

- D. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text (Statistics for Biology and Health)*. Springer, 2nd ed., 2005.
- J. Fox, *Applied Regression Analysis, Linear Models, and Related Methods*. SAGE Publications, 1997.
- G. Rodriguez, "Review of likelihood theory."
- J. Klein and M. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, second ed., 2003.