

Department of Natural Sciences

<http://www.dina.kvl.dk/~torbenm/DINA/survival>

Log-rank test and Cox-regression part I

Torben Martinussen

torbenm@dina.kvl.dk

Outline

- The Logrank Test.
- Cox-regression
- Worm example
- Logrank Tests and Cox-regression

The Logrank test

Comparison of survival in two groups.

Idea: at each observed death time we make a 2×2 table

time t_i	death	survived	at risk
group 1	f_{i1}	$r_{i1} - f_{i1}$	r_{i1}
group 2	f_{i2}	$r_{i2} - f_{i2}$	r_{i2}
total	f_i	$r_i - f_i$	r_i

Usually $f_i = 1$, so either $f_{i1} = 0, f_{i2} = 1$ or $f_{i1} = 1, f_{i2} = 0$.

With the same mortality in the groups (H_0) we expect

$$E_{i1} = r_{i1} \frac{f_i}{r_i}, \text{ and } V_i = \frac{r_{i1} r_{i2} f_i (r_i - f_i)}{r_i^2 (r_i - 1)}$$

Why?

The Logrank test

Now the expected number of deaths E_1 is: $E_1 = \sum_i E_{i1}$, and $V = \sum_i V_i$
and the **Logrank test** is

$$LR = \frac{(O_1 - E_1)^2}{V} \sim \chi_1^2 \quad \text{under } H_0$$

where $O_1 = \sum_i f_{i1}$ is no. of deaths in Gr. 1. What makes us reject the null hypothesis?

Note: $O_1 + O_2 = E_1 + E_2 \Leftrightarrow O_1 - E_1 = -(O_2 - E_2) \Leftrightarrow \frac{(O_1 - E_1)^2}{V} = \frac{(O_2 - E_2)^2}{V}$

For the worm data we have for :

```
> survdiff(Surv(time,status==1)~treatF)
```

Call:

```
survdiff(formula = Surv(time, status == 1) ~ treatF)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
treatF=0	600	526	510	0.522	1.51
treatF=1	294	265	281	0.947	1.51

```
Chisq= 1.5 on 1 degrees of freedom, p= 0.22
```

p-value of 0.22. What does that mean?

Stratified Logrank Tests

For each strata we calculate the observed, $O_{s,k}$, and expected number of deaths, $E_{s,k}$.

	Age < 10	Age \geq 10
Treat K, I	O_{00}, E_{00}, V_0	O_{01}, E_{01}, V_1
Treat F	O_{10}, E_{10}, V_0	O_{11}, E_{11}, V_1

$$\frac{(O_{00} - E_{00})^2}{V_0}$$

$$\frac{(O_{01} - E_{01})^2}{V_1}$$

That is two separate χ_1^2 -tests for effect of Fenbendazole, one in each stratum defined by age (< 10/ \geq 10)

One test only (the stratified logrank):

$$\frac{[(O_{00} + O_{01}) - (E_{00} + E_{01})]^2}{V_0 + V_1} \sim \chi_1^2$$

Stratified Logrank Tests

Doing it in R:

```
> survdiff(Surv(time,status==1)~treatF+strata(age.gr1))
```

Call:

```
survdiff(formula = Surv(time, status == 1) ~ treatF +  
          strata(age.gr1))
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
treatF=0	600	526	500	1.32	3.74
treatF=1	294	265	291	2.27	3.74

Chisq= 3.7 on 1 degrees of freedom, p= 0.0532

Cox-Regression

- Cox-Regression is the regression technique for survival analysis ;
- Regression techniques are very useful for dealing with many covariates;
- Can be used to learn about treatment effect while correcting for other covariates.

Hazard function

$$\lambda_i(t)dt = P(T_i \in [t, t + dt] \mid \text{alive at time } t)$$

Cox model:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 X_{i1} + \dots + \beta_p X_{ip}),$$

where $\lambda_0(t)$ is baseline hazard for a subject with covariates 0.

Note: $\lambda_0(t)$ is not further specified!

The Cox model

The intensity is assumed to be

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 X_{i1} + \dots + \beta_p X_{ip}),$$

The regression coefficients β_1, \dots, β_p represent the effects of the covariates. β_1 is the effect of X_{i1} when we have corrected for the other covariates.

β_1 may be interpreted in terms of the relative risk when the covariate X_{i1} is increased 1:

$$\frac{\lambda_0(t) \exp(\beta_1 (X_{i1} + 1) + \dots + \beta_p X_{ip})}{\lambda_0(t) \exp(\beta_1 X_{i1} + \dots + \beta_p X_{ip})} = \exp(\beta_1)$$

If $\beta_1 > 0$ the risk of dying increases as X_{i1} increases, and if $\beta_1 < 0$ the risk of dying decreases as X_{i1} increases.

The quantity

$$\hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_p X_{ip}$$

is called the prognostic index for the i th subject.

Two-Sample Cox Model

Consider a simple 2-sample situation where we wish to study effect of Ivermectin to hatching times.

Defining a covariate for the i th cocoon as

$$X_i = \begin{cases} 1 & \text{Fenbendazole} \\ 0 & \text{Not Fenbendazole} \end{cases}$$

Then the hazard can be written as

$$\lambda_i(t) = \lambda_0(t) \exp(\beta X_i)$$

giving the contributions $\lambda_0(t) \exp(\beta)$ when $X_i = 1$ and $\lambda_0(t)$ when $X_i = 0$.

Note again that $\lambda_0(t)$ is unspecified meaning that which distribution is unspecified?

Two-Sample Cox Model

The Cox model is fitted to the data as follows

```
> fit1<-coxph(Surv(time,status==1)~treatF)
> fit1
Call:
coxph(formula = Surv(time, status == 1) ~ treatF)
      coef exp(coef) se(coef)      z      p
treatF -0.0937      0.91  0.0764 -1.23 0.22

Likelihood ratio test=1.52  on 1 df, p=0.217  n= 894
> exp(0.0937)
[1] 1.098230
> summary(fit1)
      coef exp(coef) se(coef)      z      p
treatF -0.0937      0.91  0.0764 -1.23 0.22
      exp(coef) exp(-coef) lower .95 upper .95
treatF      0.91      1.10      0.784      1.06
Likelihood ratio test= 1.52  on 1 df,  p=0.217
Wald test              = 1.51  on 1 df,  p=0.220
Score (logrank) test = 1.51  on 1 df,  p=0.220
```

Three different tests for no effects: the **Likelihood-Ratio-test**, the **Score** and the **Wald test**.

The Wald test is given as $\left(\hat{\beta}/SE(\hat{\beta})\right)^2$ which is χ_1^2 .

Example: Melanoma data

Consider the Cox-model for the melanoma with the explanatory variables sex and log(thickness) (lt):

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 \cdot \text{sex}_i + \beta_2 \cdot \text{lt}_i)$$

We get:

	coef	exp(coef)	p
sex	0.458	1.58	8.8e-02
log(thickness)	0.781	2.18	6.9e-07

What does the relative risks of 1.58 and 2.18 mean?

Cox Model and K-Sample

Worm-data. Consider all three treatments. Define covariates:

$$X_{i1} = \begin{cases} 1 & \text{Fenbendazole} \\ 0 & \text{Otherwise} \end{cases} \quad X_{i2} = \begin{cases} 1 & \text{Ivermectin} \\ 0 & \text{Otherwise} \end{cases}$$

We can investigate the effect of the treatments through a Cox model with

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2})$$

- Intensity for controls is proportional to the intensity for the other groups with different proportionality factors.
- What is the intensity for the control group?

Cox Model and K-Sample

R-analysis

```
> summary(fit1)
Call:
coxph(formula = Surv(time, status == 1) ~ treatF + treatI)
      coef exp(coef) se(coef)      z      p
treatF 0.166      1.18  0.0885 1.87 6.1e-02
treatI 0.598      1.82  0.0883 6.78 1.2e-11

      exp(coef) exp(-coef) lower .95 upper .95
treatF      1.18      0.847   0.992   1.40
treatI      1.82      0.550   1.530   2.16
Rsquare= 0.051  (max possible= 1 )
Likelihood ratio test= 47  on 2 df,  p=6.17e-11
Wald test           = 49.1  on 2 df,  p=2.22e-11
Score (logrank) test = 50.2  on 2 df,  p=1.27e-11
```

The score test is equiv. to the the logrank test!

Cox Model and K-Sample

How about comparing Ivermectin and Fenbendazole? We can get it by re-parametrizing the problem. Define

$$X_{i3} = \begin{cases} 1 & \text{Control} \\ 0 & \text{Otherwise} \end{cases}$$

and consider the Cox model

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_2 X_{i2} + \beta_3 z_{i3})$$

This is exactly the same model as before both with reparametrized effects. Using R:

```
> fit1<-coxph(Surv(time,status==1)~treatI+treatK)
> summary(fit1)
Call:
coxph(formula = Surv(time, status == 1) ~ treatI + treatK)
      coef exp(coef) se(coef)      z      p
treatI  0.432     1.541  0.0881  4.91 9.3e-07
treatK -0.166     0.847  0.0885 -1.87 6.1e-02
      exp(coef) exp(-coef) lower .95 upper .95
treatI    1.541     0.649   1.296    1.83
treatK    0.847     1.180   0.712    1.01
Rsquare= 0.051  (max possible= 1 )
Likelihood ratio test= 47  on 2 df,  p=6.17e-11
Wald test           = 49.1  on 2 df,  p=2.22e-11
Score (logrank) test = 50.2  on 2 df,  p=1.27e-11
```

The Stratified Cox model

The stratified Cox model contains different baselines for different strata :

$$\lambda_{ik}(t) = \lambda_k(t) \exp(\beta_1 X_{i1} + \dots + \beta_p X_{ip}) \quad k = 1, \dots, K$$

$\lambda_k(t)$ is the baseline hazard for a subject in strata k .

- The regression coefficients β_1, \dots, β_p represent the effects of the covariates as in the simple Cox regression model.
- Worm-data: we might stratify according to different age groups.
- The baselines gives the mortality of the age groups when all other covariates are 0.

Stratified Log-rank

Define age-strata as below.

Now

```
> age.gr1<-0*age+1*(age<10);  
> age.gr2<-0*age+1*( (age>=10)&(age<15) );  
> age.gr<-2-(2*age.gr1+1*age.gr2)  
> fit1<-coxph(Surv(time,status==1)~treatF+treatI+strata(age.gr))  
> fit1
```

	coef	exp(coef)	se(coef)	z	p
treatF	0.111	1.12	0.0891	1.25	2.1e-01
treatI	0.585	1.79	0.0889	6.58	4.8e-11

```
Likelihood ratio test=46.2 on 2 df, p=9.23e-11 n= 894
```

```
> summary(fit1)
```

```
Likelihood ratio test= 46.2 on 2 df, p=9.23e-11  
Wald test = 48.6 on 2 df, p=2.84e-11  
Score (logrank) test = 49.7 on 2 df, p=1.64e-11
```

The Score test is equivalent to the stratified log-rank test.

The fitted stratified Cox model looks like this :

$$\lambda_{ik}(t) = \lambda_k(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2}) \quad k = 1, 2, 3$$

Where we have different baselines for the three different age-strata.

Exercise using R

- (1) Make a log-rank test comparing the groups defined by `age.gr`. Redo the test but now stratified for treatment
- (2) Make a Cox-analysis using `age.gr` as explanatory variable. Make conclusions in terms of relevant relative risks.
- (3) Redo the analysis in (2) but now also include `treatment`. Give relevant relative risks estimates.
- (4) Make a stratified Cox-analysis with `treatment` as explanatory variable and stratified by `age.gr`. Give relevant relative risks estimates.