

Cox proportional hazards model with XLSTAT-Life

[DemoCox.xls](#)

Dataset to run a Cox proportional hazards model

An Excel sheet with both the data and results can be downloaded by clicking [here](#).

The data have been obtained in Kalbfleisch and Prentice (The Statistical Analysis of Failure Time Data, Wiley, 2002, p. 119) and represent a clinical trial investigating the effect of covariates on time to death of patients with lung cancer. Our goal is to determine which covariate influences the survival time.

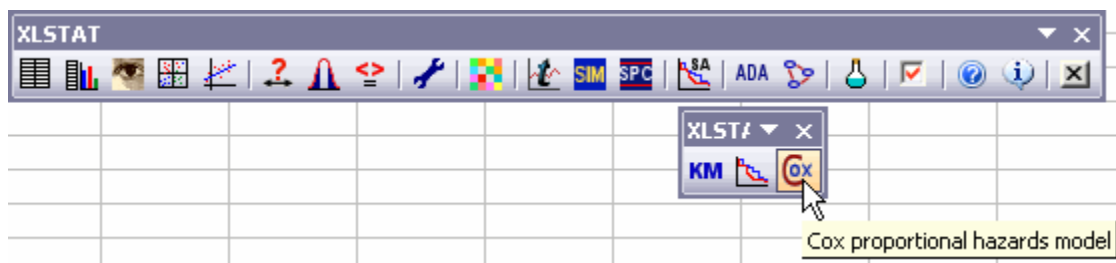
Cox proportional hazards model

The proportional hazard model has been introduced by Cox (1972) and it is based on a classical regression scheme. The estimation of the model is performed with a specific type of maximum likelihood estimation called partial likelihood. An estimation of the coefficients of the model is obtained supposing the proportional hazard hypothesis holds.

In the dataset, the daysurv variable is the time data; the censoring variable is the status variable (1 for death, 0 for censored). The covariates are the performance status of the patient at the beginning of the study (perfstatus), the age of the patient at the beginning of the study (age), the number of month since lung cancer diagnostic at the beginning of the study (month) and the presence of an earlier treatment.

Setting up a Cox proportional hazards model

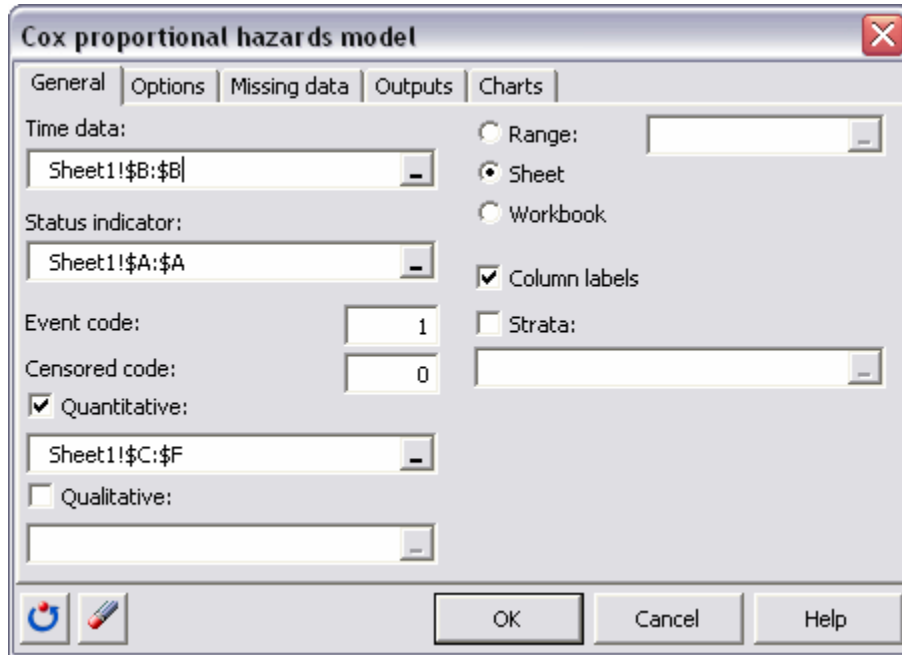
After opening XLSTAT, select the **XLSTAT / XLSTAT-Life / Cox Proportional hazards model** command, or click on the corresponding button of the **XLSTAT-Life** toolbar (see below).



Once you've clicked on the button, the Cox proportional hazards model box will appear. Select the data on the Excel sheet. The **Time data** corresponds to the durations when the patients either

died or were censored. The "Status indicator" describes whether a patient died (event code=1) or was censored (censored code = 0) at a given time.

The covariates are all quantitative and can be selected in the **quantitative** box.



Other options can be selected on the other tabs of the dialog box like stratification of the model, individual residuals computation, ties handling method selection and so on.

The computations begin once you have clicked on **OK**. The results will then be displayed on a new Excel sheet.

Interpreting the results of a Cox proportional hazards model

The first table displays a summary of the data. We can see that the number of observed times (time steps) is different than the number of observations. We then have to use a tie handling technique (Breslow's method is the default one, but Efron's method is also available).

Summary statistics:			
Total observed	Total failed	Total censored	Time steps
137	128	9	101

Descriptive statistics are then displayed for each variable (in our example, all variable are quantitative, but if qualitative variables are selected a different table will be displayed). For the treatment of qualitative variables, please refer to the help of XLSTAT for more details.

Descriptive statistics:							
Variable	Observation	with missing	without missing	Minimum	Maximum	Mean	std. deviation
perfstatus	137	0	137	10,000	99,000	58,569	20,040
month	137	0	137	1,000	87,000	8,774	10,612
age	137	0	137	34,000	81,000	58,307	10,542
priorthrp	137	0	137	0,000	10,000	2,920	4,563

The next table gives several indicators of the quality of the model (or goodness of fit). These results are equivalent to the R2 and to the analysis of variance table in linear regression and ANOVA. The most important value to look at is the probability of Chi-square test on the log ratio. This is equivalent to the Fisher's F test: we try to evaluate if the variables bring significant information by comparing the model as it is defined with a simpler model with no impact of the covariates. In this case, as the probability is lower than 0.0001, we can conclude that significant information is brought by the variables.

Goodness of fit statistics:		
Statistic	Independent	Full
Observations	128	128
DF	0	4
-2 Log(Likelihood)	1011,768	969,987
AIC	1011,768	977,987
SBC	1011,768	989,395
Iterations	1	5

Test of the null hypothesis H0: beta=0			
Statistic	DF	Chi-square	Pr > Chi²
-2 Log(Likelihood)	4	41,781	< 0,0001
Score	4	45,262	< 0,0001
Wald	4	43,072	< 0,0001

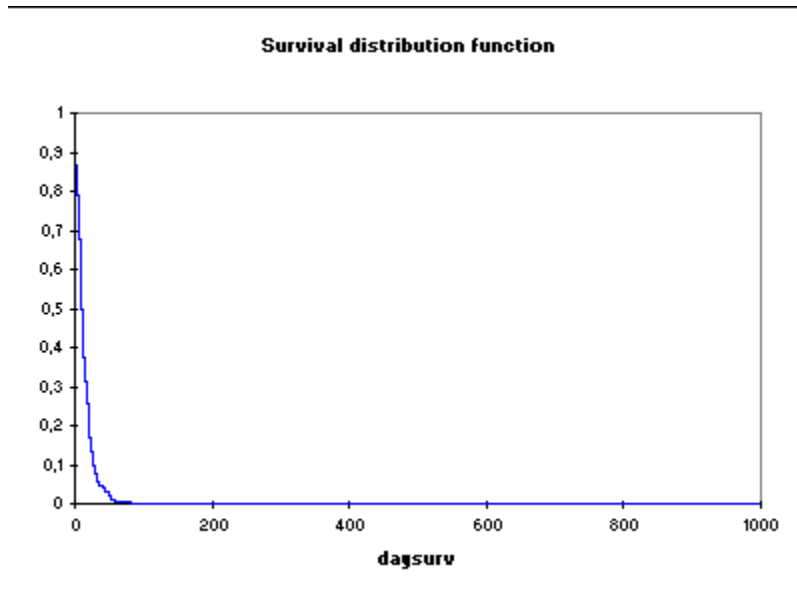
The following table gives details on the model. This table is helpful in understanding the effect of the various variables.

Regression coefficients:							
Variable	Value	Standard error	Wald Chi-Square	Pr > Chi²	Hazard ratio	Lower bound	Upper bound (95%)
perfstatus	-0,033	0,005	39,052	< 0,0001	0,967	0,957	0,977
month	0,002	0,009	0,036	0,849	1,002	0,984	1,020
age	-0,002	0,009	0,060	0,807	0,998	0,980	1,016
priorthrp	-0,006	0,022	0,083	0,773	0,994	0,952	1,038

On this table we can see from looking at the probability of the Chi-squares that the variable most influencing survival time is perfstatus. This shows that the performance status of the patient at

beginning of the study has a significant effect on survival time. The hazard ratio is obtained as the exponential of the parameter estimate.

Finally, the cumulative hazard function is displayed:



This study has shown that the only covariate with a significant impact is the performance status. The coefficient being negative shows that when a patient has a low performance status his survival time is greater. The other covariates do not have a significant effect on the survival time.