Cox Regression

File: COX

THE DATA

The dataset is from a childcare study. A cohort of women who returned to work after the birth of their child was studied for a period of 24 months. The study focuses on the second spell of childcare. We make the assumption that the first spell (or period) of childcare is undertaken by the women herself or her partner. We are interested in what happens when the women returns to work.

There are 341 mothers (and babies) in the study.

THE VARIABLES

id unique mother identification number

start start of childcare spell number 2 (months)

end end of childcare spell number 2 (months)

inc family income
0 = £30K + gross
1 = Up to £30K gross

gender baby's gender
0 = boy
1 = girl

mumage mother's age (years) at birth

childm type of care childminder
0 = no
1 = yes

nursery type of care nursery
0 = no
1 = yes

type type of care
1 = relative
2 = childminder
3 = nursery

GET TO KNOW THE DATA

1. What is the mean time (months) for spell 2 to begin?

2. What is the mean time (months) for spell 2 to end?

3. What proportion of families earn less than £30K net?

4. How many boys and how many girls are in the study?

5. How old is the youngest mother?

6. How old is the oldest mother?

7. What is the mean and standard deviation of mothers age (years)?

GETTING READY TO MODEL THE DATA

Construct a time (or duration) variable called **time** using **start** and **end.**

8. Compute the mean and the standard deviation of this variable.

Display the frequencies of the new variable **time**.

9. How many and what proportion of cases are censored?

Compute a variable **status** that identifies censored cases.

> *If end is equal to 25 then the case is censored and status = 0.*
>
> *If end is less than 25 then the case is not censored (i.e. the event took place) and status =1.*

Check that the correct number of censored cases identified by your new variable **status**.

THEORISING OUR DATA

10. What might be the effect of family income on the (survival) duration of the second childcare spell?

11. What might be the effect of the type of childcare on the (survival) duration of the second childcare spell?

12. What might be the effect of the child's gender on the (survival) duration of the second childcare spell?

13. What might be the effect of mother’s age on the (survival) duration of the second childcare spell?

EXPLORING THE DATA

Explore the relationships between mean durations (using the variable **time**) and the explanatory variables.

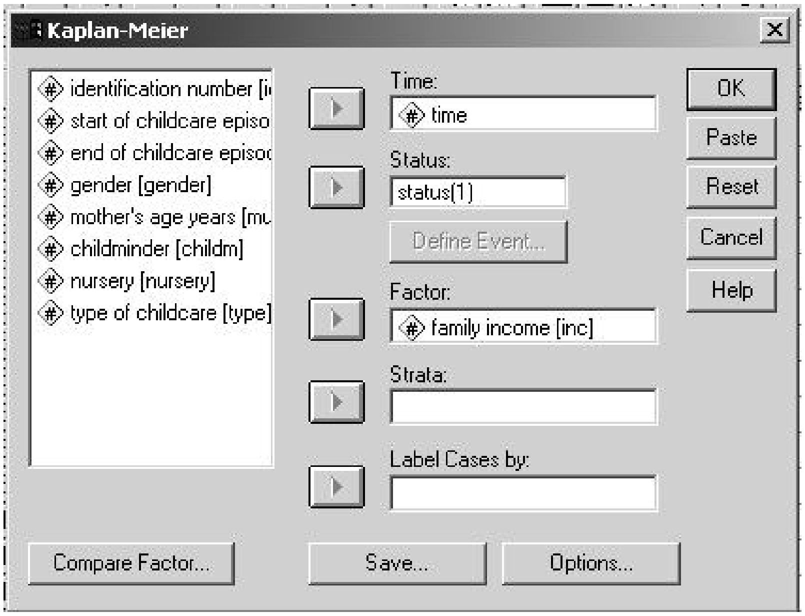
Compare the means for **inc, gender, childm, nursery**
(Hint: Use an Anova Table might help you)

14. What do these results suggest?

Either scatter plot the relationship between mother’s age (**mumage**) and duration (**time**), or compute a correlation.

15. What does this suggest?

KAPLAN-MEIER PLOT


Kaplan-Meier

identification number [i
start of childcare episo
end of childcare episo
gender [gender]
mother's age years [mu
childminder [childm]
nursery [nursery]
type of childcare [type]

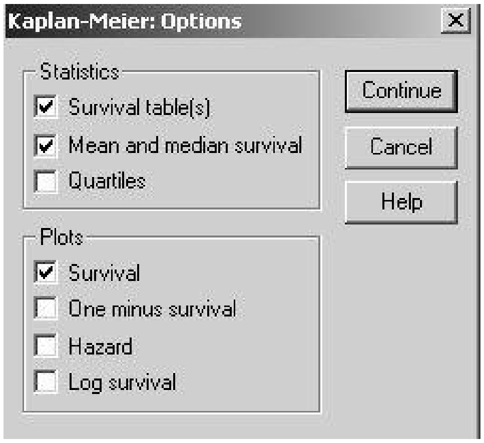
Time: time

Status: status(1)

Factor: family income [inc]

Strata:

Label Cases by:


Kaplan-Meier: Options

Statistics
- ☑ Survival table(s)
- ☑ Mean and median survival
- □ Quartiles

Plots
- ☑ Survival
- □ One minus survival
- □ Hazard
- □ Log survival

Plot a Kaplan-Meier curve for income.

16. What does this suggest?

17. Now compare plots for *Survival* with *One minus survival*, *Hazard* and *Log survival*. What do these plots tell us?

Plot a Kaplan-Meier (survival) curve for gender.

18. What does this suggest?

FITTING THE COX MODEL

Now fit the Cox regression model. Use the menus

Analyze
    Survival
        Cox regression.

Time is the duration (**time**).
Status is our censored indicator variable (**status**).
Don’t forget to define the event (in our case it is 1).
Enter the covariates **inc, gender, mumage, childm, nursery**.

19. Which covariates are significant?

Now fit the Cox regression model with only the significant covariates.

20. What are the effects of family income and type of childcare?

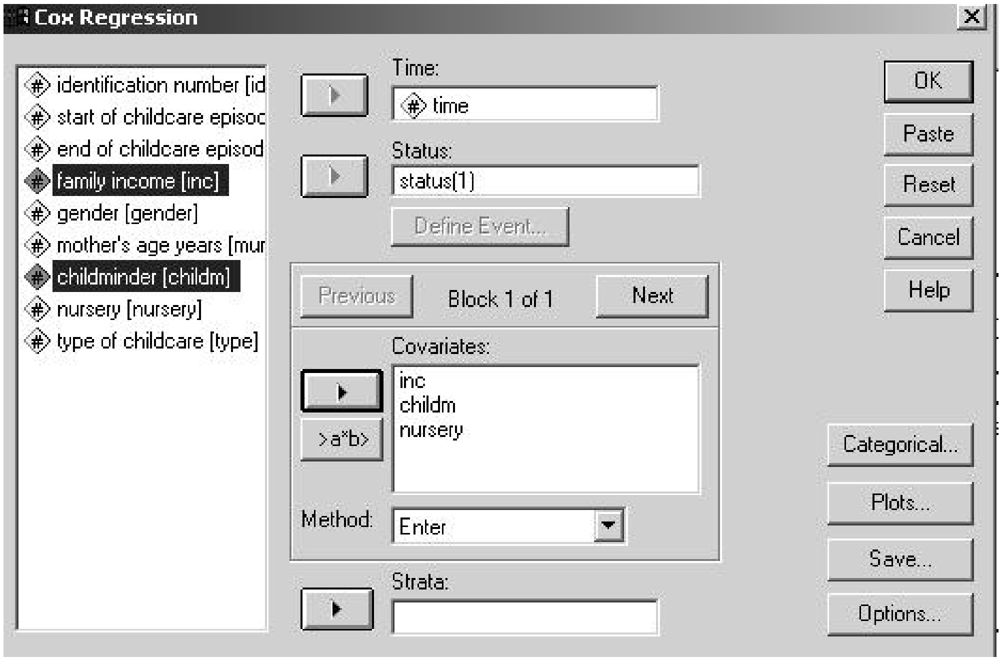
21. Could type of childcare have a different effect depending on family income?

Fit an interaction between income **(inc)** and childminder care (**childm**) and income (**inc**) and nursery care (**nursery**).

To do this click on **inc**; hold down the Ctrl key; then click on **childm**.

Both variables should be highlighted in blue (see below).

Add the interaction to the model using the **>a*b>** button.



identification number [id
start of childcare episoc
end of childcare episod
family income [inc]
gender [gender]
mother's age years [mur
childminder [childm]
nursery [nursery]
type of childcare [type]

Time: time

Status: status(1)

Covariates:
inc
childm
nursery

Method: Enter

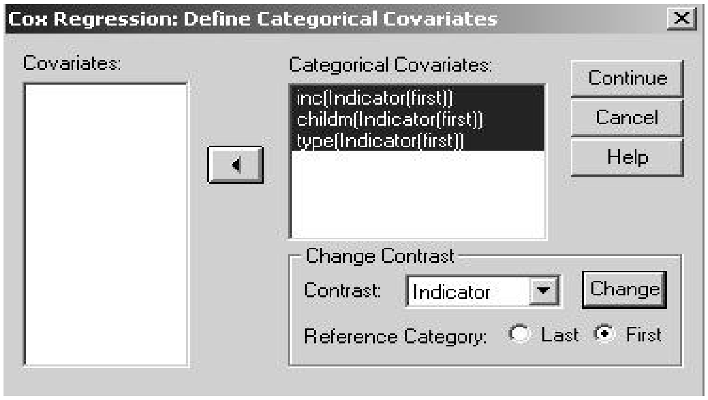
Strata:

22. Which variables are significant?

23. Might the cost of nurseries be implicated in these results?

Now fit the model and include **inc** and a new variable called **type**. This variable records the type of care in spell number 2 in a single variable.

Because **type** is a categorical (three category) variable rather than a dummy variable we must specify this. Click on the Categorical... button.



Cox Regression: Define Categorical Covariates

Categorical Covariates:
inc(Indicator(first))
childm(Indicator(first))
type(Indicator(first))

Contrast: Indicator

Reference Category: ○ Last ● First

In SPSS you must specify that a variable is categorical and it is a good idea to change the contrast to first (like you might do in a logistic regression).

24. Which variables are significant?

THE ACCELERATED LIFE MODEL

Compute a variable for the $\log_e$ of time (**ltime**).

25. Examine the distribution of this new variable and the compute its mean and standard deviation.

26. Plot a scatter plot of **time** and **ltime**.

Fit a linear regression model with **ltime** as the dependent variable and **inc, gender, mumage, childm, nursery** as independent variables.

27. Which variables are significant?

28. What do you notice about the signs of the estimates (B) compared with the Cox model?

29. Why might this be?

30. Why is this model inappropriate?

Well Done!