Kaplan-Meier

- Kaplan Meier can be used to compare two or more treatment groups on their survival times.
- Duration is measured from a well-defined time origin until the occurrence of some particular event of interest or end-point.
- Kaplan-Meier is the usual technique performed to analyse survival-time data.
- The Kaplan Meier technique is the univariate version of survival analysis.

Example

- 25 lung cancer patients :
- 1#, 5#, 6, 6, 9#, 10, 10, 10#, 12, 12, 12, 12, 12#, 13#, 15#, 16#, 20#, 24, 24#, 27#, 32, 34#, 36#, 36#, 44# months
- variable “Status” tells which case is censored (denoted by 0) and which case is an event (dying of lung-cancer, denoted by 1).

Sample result

Factor group = control

	Survival time	Standard error	95% confidence interval
Mean (Limited to 36)	21	5	(12, 30)
Median	12	2	(7, 17)

Factor group = active

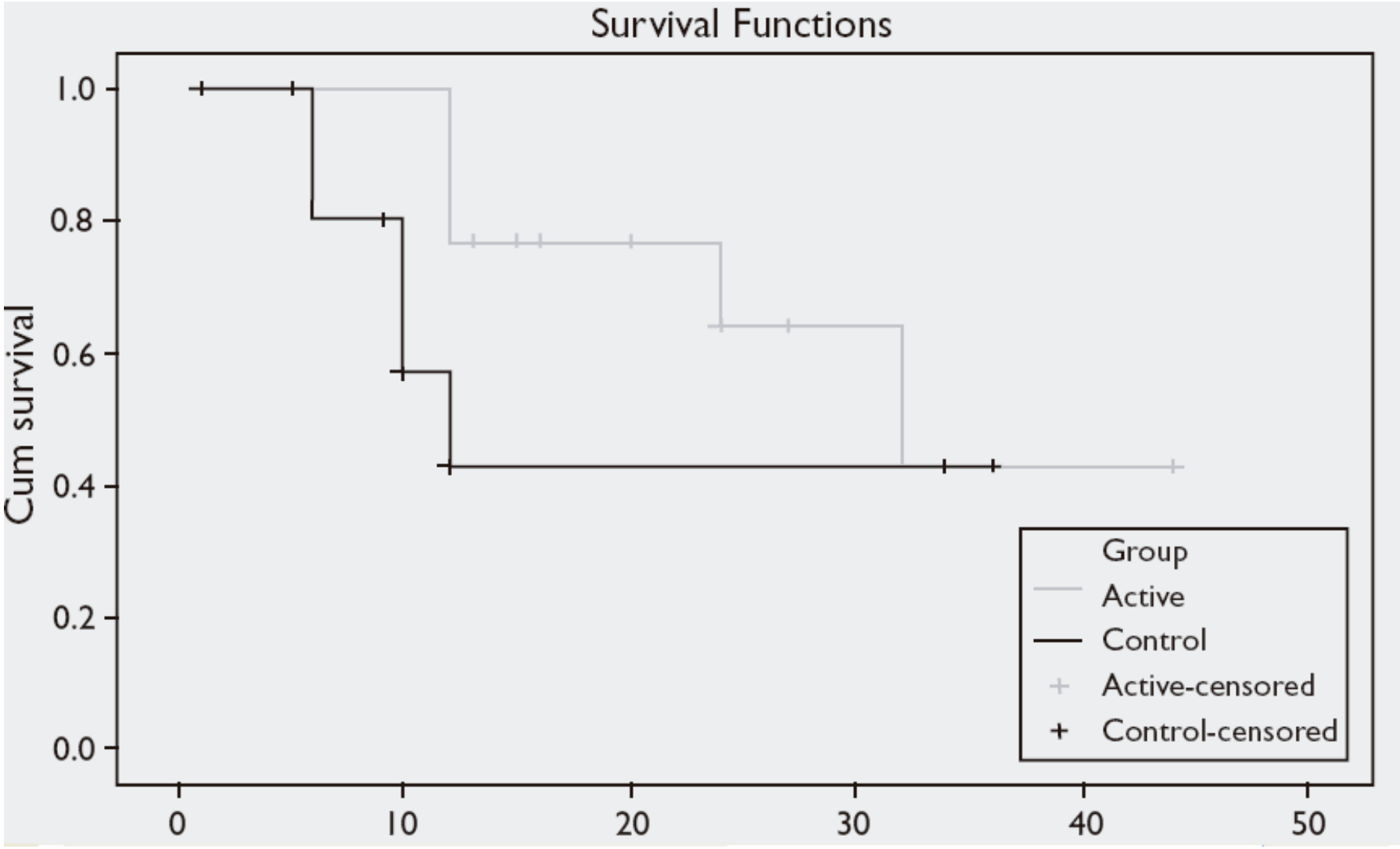
	Survival time	Standard error	95% confidence interval
Mean (Limited to 44)	31	4	(23, 39)
Median	32	8	(17, 47)

	Total	Number of events	Number censored	Percent censored
Group control	12	5	7	58.33
Group active	13	5	8	61.54
Overall	25	10	15	60.00

Test statistics for equality of survival distributions for group

	Statistic	df	Significance
Log rank	1.77	1	.1835

Sample Plot


Survival Functions
Cum survival
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Group
Active
Control
Active-censored
Control-censored

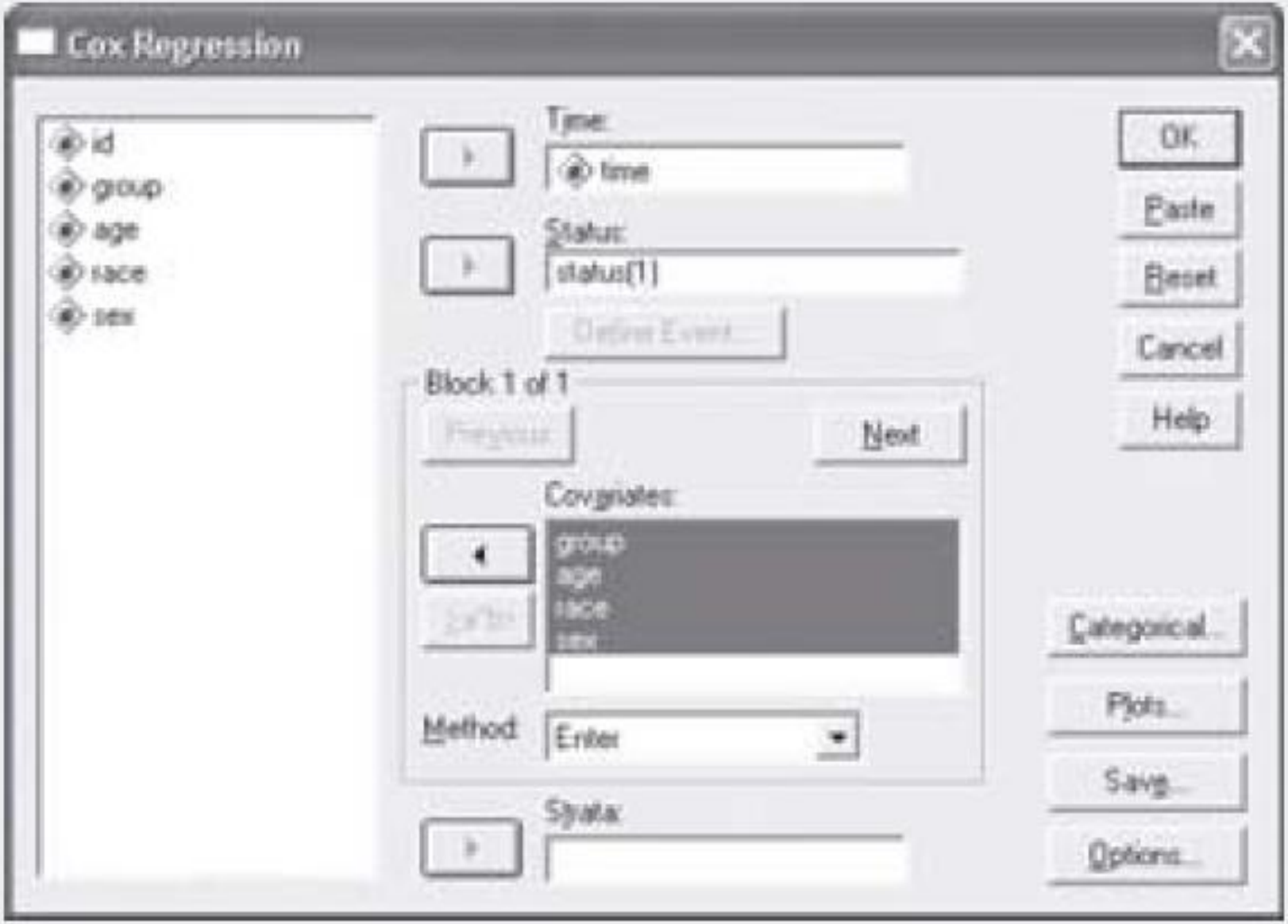
Cox Regression

YH Chan

Cox Regression

- to look at a confounder model to determine whether the groups differ after adjusting for confounding factors such as differences of demographics.
- To perform a Cox regression, go to
  - Analyse,
  - Survival,
  - Cox regression.

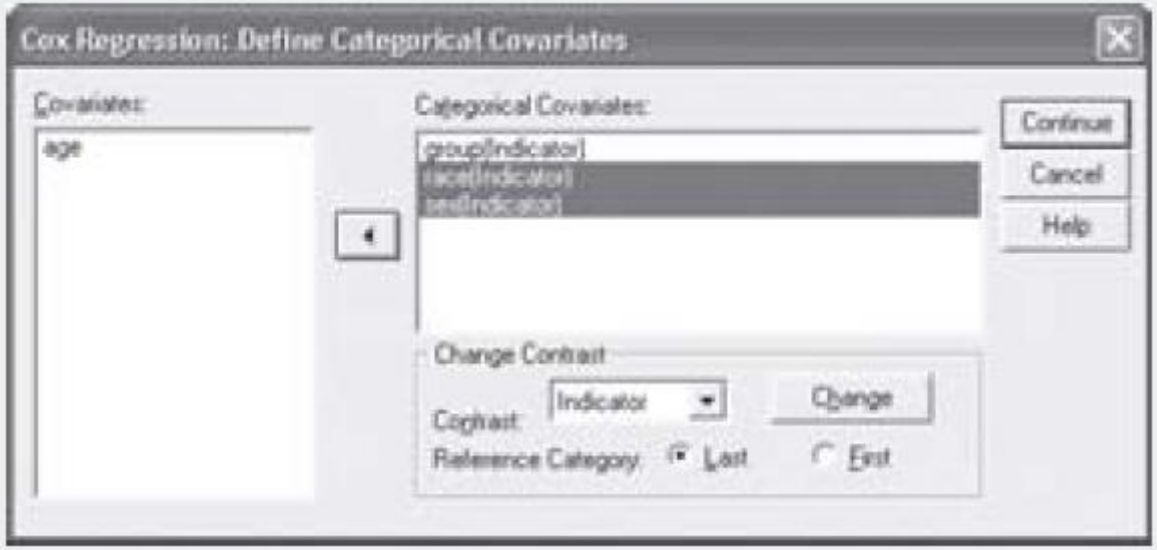
Template VI. Cox regression: lung cancer example.



Categorical Variables

- The declaration for the categorical variables is similar to logistic regression by clicking on the “Categorical” button. The group with the longer survival time would be the reference group.
- In this example, we put group, race and sex as the categorical covariates.

Template VII. Declaration of categorical variables.



hazard ratio

- Click on “Options” to invoke the 95% CI for the hazard ratio (HR), given by the expression exp(B)

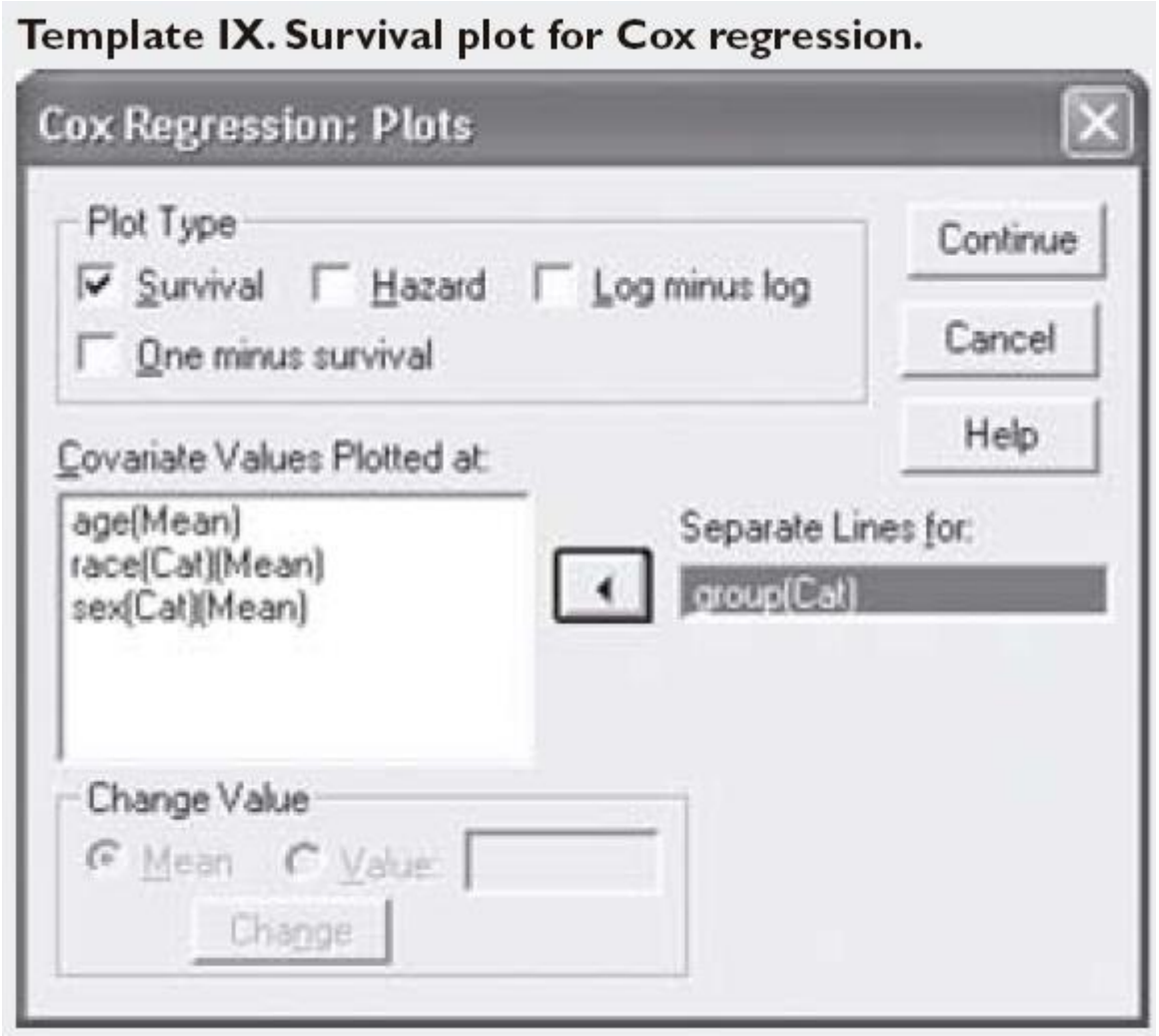
Template VIII. Invoking the 95% CI for the hazard ratio.

hazard ratio

- The interpretation for the hazard ratio is similar to that of the odds ratio.
  - A value of one means there is no differences between two groups in having a “shorter time to event”.
  - A HR >1 means that the group of interest comparing to the reference group (based on the categorical declaration) likely have a shorter time to event.
  - A HR <1 means that the group of interest less likely to have a shorter time to event.

Plots

- Click on “Plots” to get the following requester.
- Click on “Survival” and choose Separate Lines for “group”.



Results

- The reference category for group is active, race is “other race” and sex is female.

Table VIa. Categorical definition.

		Categorical variable codings			
		Frequency	(1)	(2)	(3)
Group	1.00=control	12	1		
	2.00=active	13	0		
Race	1=chinese	15	1	0	0
	2=indian	5	0	1	0
	3=malay	2	0	0	1
	4=other	3	0	0	0
Sex	1=male	17	1		
	2=female	8	0		

Results

- The table gives the p-values (Sig) and the hazard ratios (Exp(B)) of the variables.
- We have to check for multi-collinearity by observing whether the SE of all the variables are small

Table VIb. Estimates of variables in Cox regression.

Variables in the equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Group	1.841	.911	4.086	1	.043	6.302	1.058	37.550
Sex	3.670	1.435	6.542	1	.011	39.263	2.358	653.769
Age	.115	.043	7.137	1	.008	1.122	1.031	1.220
Race			2.066	3	.559			
Race(1)	-.307	1.181	.068	1	.795	.735	.073	7.448
Race(2)	.983	1.299	.573	1	.449	2.672	.210	34.060
Race(3)	.907	1.469	.381	1	.537	2.476	.139	44.085

Results

- Since adjusting for confounder model, our interest is only in the variable “group”.
- p-value for Cox-Regression is 0.043 (statistically significant) compared to the earlier Kaplan Meier analysis (log-rank p=0.1835).
- The HR is 6.302 (95% CI 1.058 - 37.55), comparing the control with the active, the control is likely to have a shorter time to event and in this example, the event is death.
- Why Cox is different with Kaplan-Meier?

Table VIb. Estimates of variables

	B	Sig.
Group	1.841	.043
Sex	3.670	.011
Age	.115	.008
Race		.559
Race(1)	-.307	.795
Race(2)	.983	.449
Race(3)	.907	.537

Why? Confounder?

- Statistical differences noted for gender and age.
- Men and older people were worse off.
- Cross-tab -> more men than women in control group (p = 0.673)
- Mean age is higher in the active group.

Table VIc. Cross-tabulation between group and gender.

The sex of the patient * group cross-tabulation

			Group		
			Control	Active	Total
Sex of patient	Male	Count	9	8	17
		% within group	75.0%	61.5%	68.0%
	Female	Count	3	5	8
		% within group	25.0%	38.5%	32.0%
Total		Count	12	13	25
		% within group	100.0%	100.0%	100.0%

Table VId. Age differences between group (p=0.737).

Group statistics

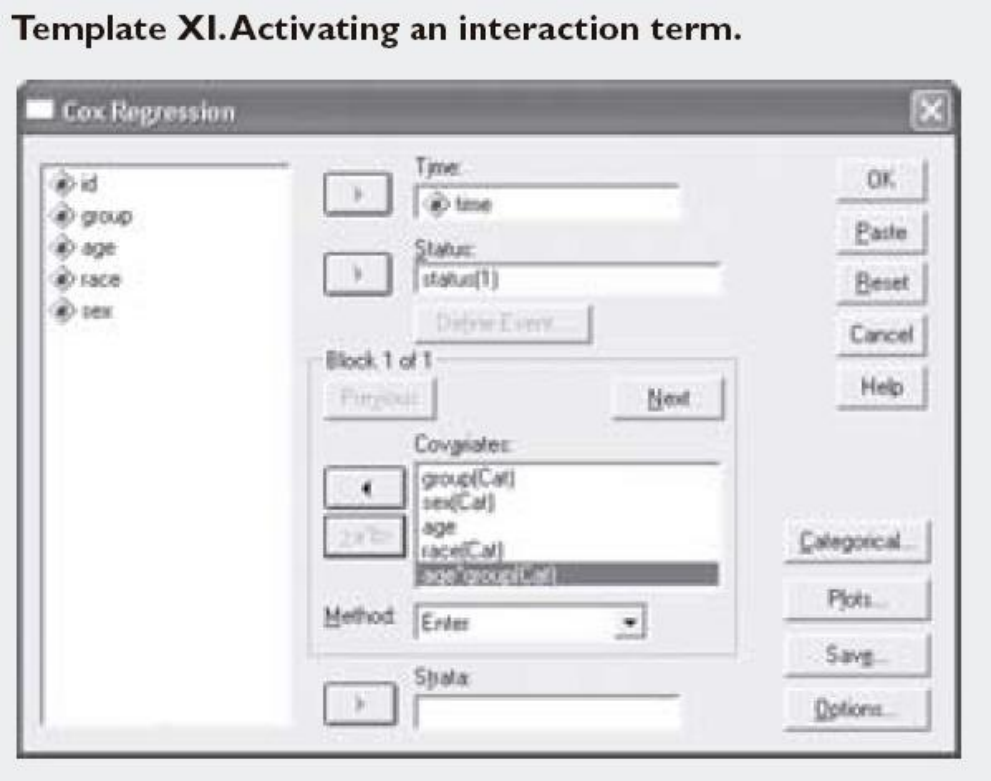
	Group	N	Mean	Std. deviation	Std. error mean
Age	active	13	31.6923	16.16263	4.48271
	control	12	29.5833	14.73683	4.25416

Therefore!

- Don’t stop at the univariate analysis but to always perform a multivariate analysis to identify the real situation!
- How about interaction between gender and “group”, or age and “group”? How to test for it?

How to test for interaction?

- Within Cox Regression, select both age and “group” at the same time by pressing the “Ctrl” key. The button “>a*b>” becomes “enabled”, so click on it. Repeat the same for gender and “group”.



	B	SE	Wald	df	Sig.
Group	-5.524	4.891	1.276	1	.259
Sex	1.687	1.716	.966	1	.326
Age	.082	.055	2.186	1	.139
Race			3.171	3	.366
Race(1)	-.869	1.341	.420	1	.517
Race(2)	1.112	1.261	.777	1	.378
Race(3)	1.018	1.570	.421	1	.517
Age*group	.121	.089	1.823	1	.177
Group*sex	5.584	3.261	2.933	1	.087

- none of the interaction terms are significant. This implies that regardless of age or gender, the active group is performing better.

Exercise

- Open breast cancer survival dataset from SPSS.
- Variables of interest;
  - age
  - categorical histology grade,
  - oestrogen receptor status,
  - progesterone receptor status,
  - pathological tumour size and
  - lymph node status.
- to determine the predictors for a shorter survival time to death.

Kaplan-Meier

- Do a Kaplan-Meier analysis for all these variables first;
  - age
  - categorical histology grade,
  - oestrogen receptor status,
  - progesterone receptor status,
  - pathological tumour size and
  - lymph node status.
- Identify which variables are significant and explain the biological plausibility.

Based on your earlier results, are these coding correct?

		Frequency	(1)	(2)
histgrad	1=1	56	0	0
	2=2	352	1	0
	3=3	252	0	1
cr	0=negative	262	0	
	1=positive	398	1	
pr	0=negative	299	0	
	1=positive	361	1	
pathscat	1=<=2cm	457	0	0
	2=2-5cm	196	1	0
	3=>5cm	7	0	1
ln_yesno	0=no	485	0	
	1=yes	175	1	

- Reference group;
  - histology grade is grade 1
  - for er, pr and lymph node is negative
  - and tumour size is ≤2cm.
- Do the Cox regression, first without controlling for interaction, then repeat but controlling for interaction.

Discuss this result

	B	SE	Wald	df	Sig.
Age	-.021	.014	2.200	1	.138
histgrad			.872	2	.647
histgrad(1)	.778	1.036	.564	1	.453
histgrad(2)	.942	1.056	.796	1	.972
cr	-.022	.432	.003	1	.959
pr	-.455	.422	1.159	1	.282
pathscat			6.005	2	.050
pathscat(1)	.638	.336	3.614	1	.057
pathscat(2)	1.484	.776	3.658	1	.056
ln_yesno	.724	.337	4.605	1	.032

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
age	-.021	.014	2.200	1	.138	.980	.953	1.007
pathscat			6.005	2	.050			
pathscat(1)	.638	.336	3.614	1	.057	1.893	.980	3.657
pathscat(2)	1.484	.776	3.658	1	.056	4.412	.964	20.200
ln_yesno	.724	.337	4.605	1	.032	2.063	1.065	3.997
histgrad			.872	2	.647			
histgrad(1)	.778	1.036	.564	1	.453	2.177	.286	16.587
histgrad(2)	.942	1.056	.796	1	.372	2.564	.324	20.300
er	.022	.432	.003	1	.959	1.022	.438	2.385
pr	.455	.422	1.159	1	.282	1.576	.689	3.605

Discussion

- Those with a positive lymph node more likely to have a shorter time to death (HR = 2.06, 95% CI 1.07 - 4.0, p = 0.032). Tumour size is “just off statistical significance”.
- Should we conclude that only women with a positive lymph node are at a higher risk?
- What happens if we include a lymph node * tumor size interaction?

lymph node * tumor size interaction

	B	SE	Wald	df	Sig.
Age	-.023	.014	2.845	1	.092
histgrad			1.165	2	.559
histgrad(1)	1.047	1.067	.962	1	.327
histgrad(2)	1.161	1.081	1.153	1	.283
cr	-.063	.424	.022	1	.881
pr	-.516	.413	1.556	1	.212
pathscat			8.520	2	.014
pathscat(1)	-.179	.501	.128	1	.721
pathscat(2)	3.100	1.102	7.904	1	.005
ln_yesno	.006	.505	.000	1	.990
ln_yesno*pathscat			8.564	2	.014
ln_yesno*pathscat(1)	1.670	.707	5.574	1	.018
ln_yesno*pathscat(2)	-1.847	1.547	1.425	1	.233

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
age	-.023	.014	2.845	1	.092	.977	.951	1.004
pathscat			8.520	2	.014			
pathscat(1)	-.179	.501	.128	1	.721	.836	.313	2.233
pathscat(2)	3.100	1.102	7.904	1	.005	22.189	2.557	192.566
ln_yesno	.006	.505	.000	1	.990	1.006	.374	2.706
histgrad			1.165	2	.559			
histgrad(1)	1.047	1.067	.962	1	.327	2.848	.352	23.068
histgrad(2)	1.161	1.081	1.153	1	.283	3.192	.384	26.563
er	.063	.424	.022	1	.881	1.065	.464	2.447
pr	.516	.413	1.556	1	.212	1.675	.745	3.766
ln_yesno*pathscat			8.564	2	.014			
ln_yesno*pathscat(1)	1.670	.707	5.574	1	.018	5.312	1.328	21.248
ln_yesno*pathscat(2)	-1.847	1.547	1.425	1	.233	.158	.008	3.274

lymph node * tumor size interaction

- lymph node status is no more statistically significant but tumour size and their interaction are.
- Regardless of the lymph node status, subjects with tumour size >5cm are at risk (HR=22.19, 95% CI 2.56 - 192.57, p=0.005) and for subjects with tumour size 2 - 5cm, they are at a higher risk if they have a positive lymph node (HR=5.31, 95% CI 1.33 - 21.25, p=0.018).