

Survival Analysis Using SPSS

By Hui Bian

Office for Faculty Excellence

Survival analysis

- ❖ What is survival analysis
 - Event history analysis
 - Time series analysis
- ❖ When use survival analysis
 - Research interest is about time-to-event and event is discrete occurrence.
- ❖ Examples of survival analysis
 - Duration to the hazard of death
 - Adoption of an innovation in diffusion research
 - Marriage duration
- ❖ Characteristics of survival analysis
 - At any time point, events may occur
 - Factors influence events include two types: time-constant and time-dependent (age).

Survival analysis

- Survival analysis focuses on hazard function
 - Hazard: the event of interest occurring
 - Hazard might be death, engine breakdown, adoption of innovation, etc.
 - Hazard rate: is the instantaneous probability of the given event occurring at any point in time. It can be plotted against time on the X axis, forming a graph of the hazard rate over time.
 - Hazard function: the equation that describe this plotted line is the hazard function.
 - Hazard ratio: also called relative risk: $\text{Exp}(B)$ in SPSS.

Survival analysis



Type of survival analysis

- Nonparametric: no assumption about the shape of hazard function. Hazard function is estimated based on empirical data, showing change over time, for example, Kaplan-Meier survival analysis.
- Semi-parametric: no assumption about the shape of hazard function, but make assumption about how covariates affect the hazard function, for example: Cox regression
- Parametric: specify the shape of baseline hazard function and covariates effects on hazard function in advance.
 - Maximum likelihood method
 - Used when time is itself considered a meaningful independent variable.
 - Used for predictive modeling
 - Software: Stata

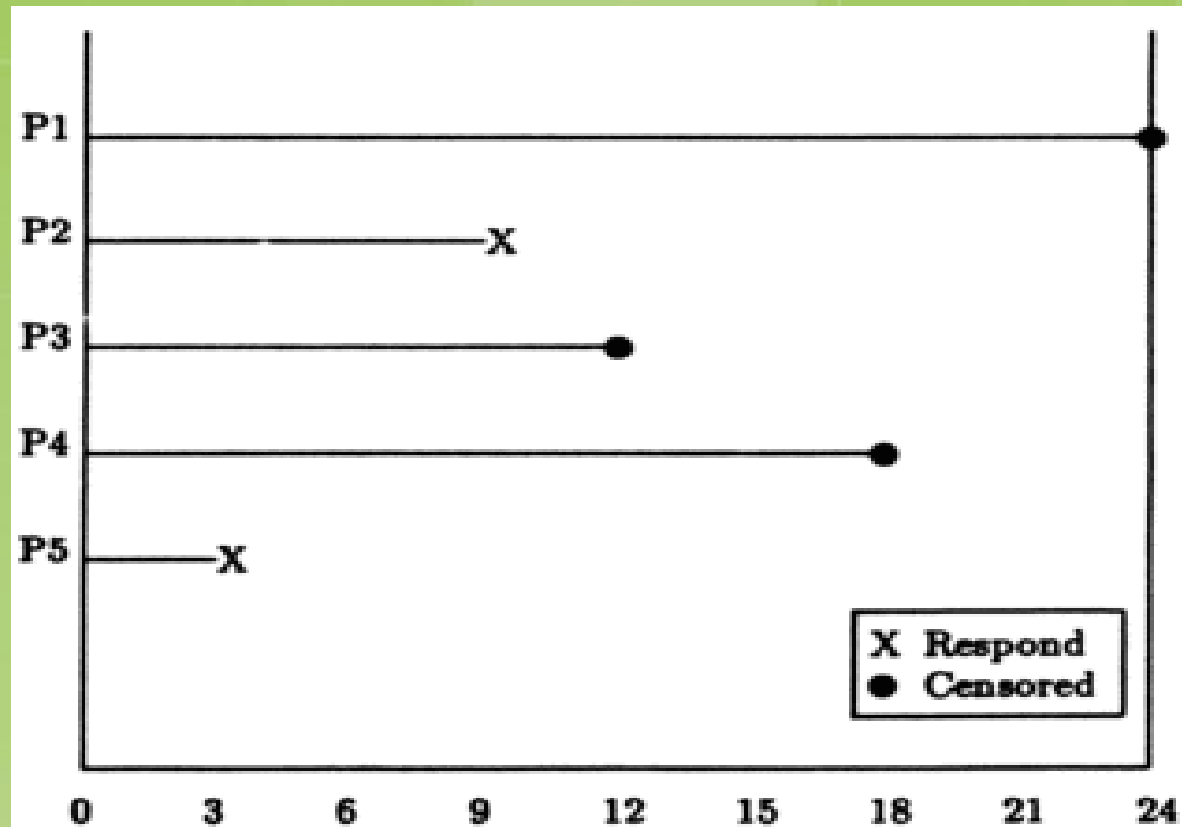
Survival analysis

Terms

- Events: what terminates an episode (such as death, adoption of an innovation), it is the change which causes the subject to transition from one state to another.
- Durations: the number of time units an individual spends in a given state.
- Dependent: probability of an event.
- Survival function, $s(t)$: is the cumulative frequency of the proportion of the sample **Not experiencing** the event by time t . In another word, it is the probability of event will NOT occur until time t .
- Censored cases: data are censored if events start before (left-censored) or ended after (right-censored) the period of observation.

Survival analysis

◇ Censored cases



Survival analysis

- ❖ Censored cases: unique characteristics of survival analysis.
 - For some cases, the event simply doesn't occur before the end of study.
 - For some cases, they drop out from the study for reasons unrelated to the study.
 - For some cases, we lost track of their status sometime before the end of the study.

Survival analysis



Outline of topics

- Life tables
- Kaplan-Meier
- Cox regression
- Cox regression with a time-dependent covariate

Life tables

- ❖ Life Tables is a descriptive procedure for examining the distribution of time-to-event variables. We also can compare the distribution by levels of a factor variable.
- ❖ The basic idea of life tables is to subdivide the period of observation into smaller time intervals. Then the probability from each of the intervals are estimated.

Life tables

Variables

- Time variable (duration variable): must be a continuous variable.
- Status variable: binary or categorical variable, represents the event of interest.
- Factor variable: categorical variable.

Assumption

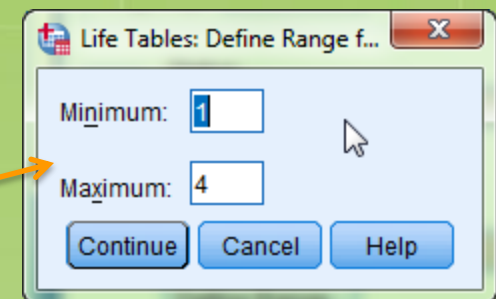
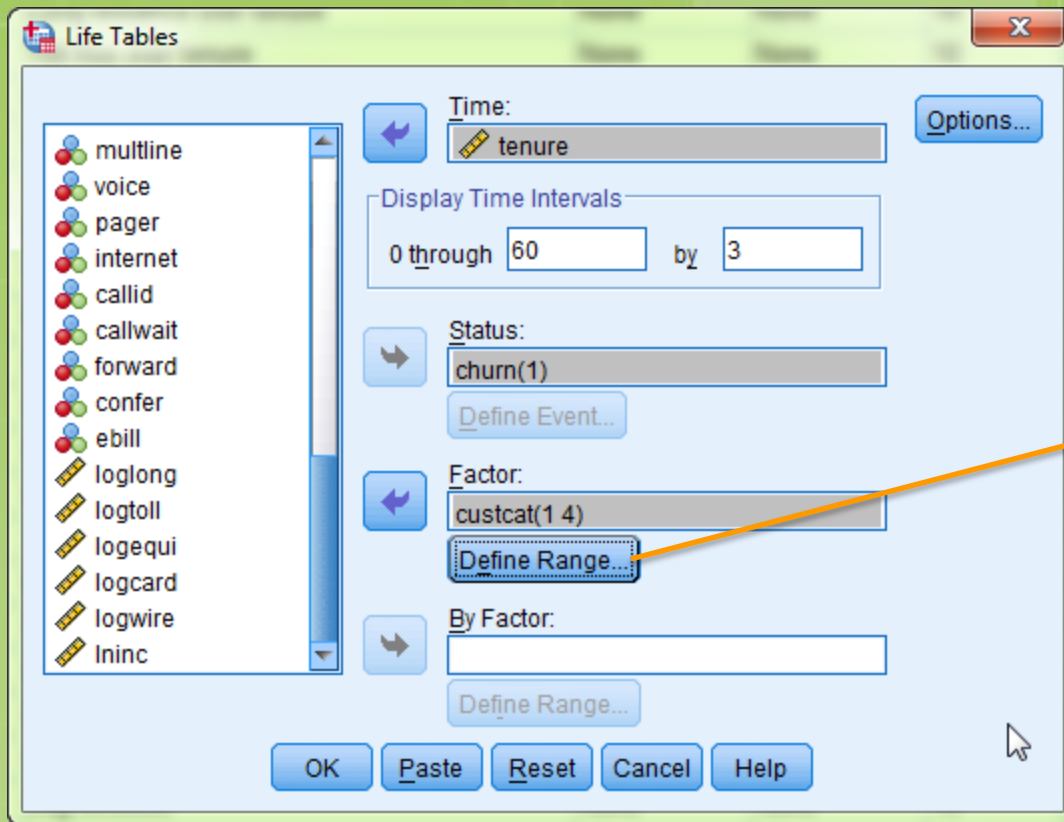
- The probability for the event of interest should depend **only on time**. Cases that enter the study at different times should behave similarly.
- No systematic differences between censored and uncensored cases

Life tables

- ◆ **Example** (from IBM SPSS 20.0): data file name: telco
 - Examine distribution of customer time to churn by customer category.
 - Time variable: tenure (in month)
 - Status variable: churn (binary: 1 = Churn, 0 = Not churn)
 - Factor: custcat (four categories)
- ◆ Go to Analyze > Survival > Life Tables

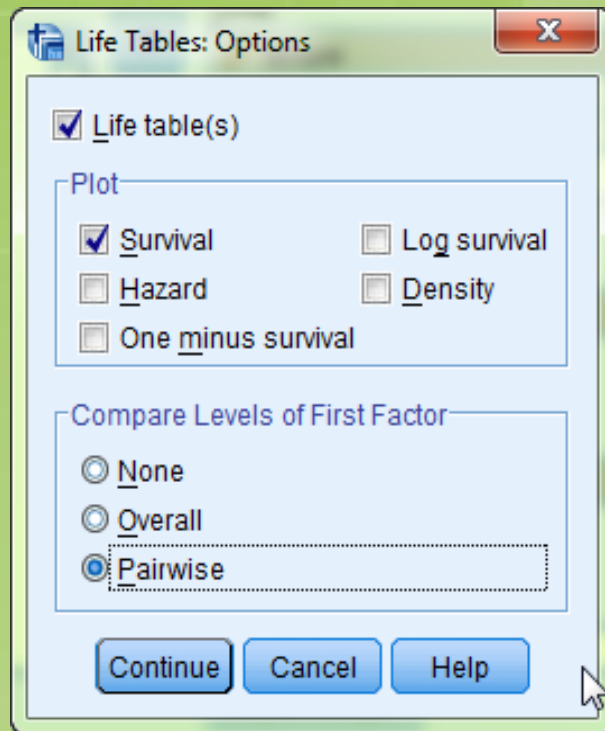
Life tables

Run analysis



Life tables

Click *Options*



1. Survival: display the cumulative survival function on a linear scale
2. Hazard: display the cumulative hazard function on a linear scale.

Life tables

SPSS Outputs: life table

Life Table								
			Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving
First-order Controls		Interval Start Time						
custcat Customer category	1 Basic service	0	266	3	264.500	10	.04	.96
		3	253	10	248.000	17	.07	.93
		6	226	12	220.000	10	.05	.95
		9	204	11	198.500	10	.05	.95
		12	183	13	176.500	6	.03	.97
		15	164	10	159.000	5	.03	.97
		18	149	15	141.500	1	.01	.99
		21	133	6	130.000	4	.03	.97
		24	123	10	118.000	4	.03	.97
		27	109	4	107.000	2	.02	.98
		30	103	9	98.500	4	.04	.96
		33	90	17	81.500	3	.04	.96
		36	70	7	66.500	2	.03	.97
		39	61	5	58.500	3	.05	.95
		42	53	6	50.000	1	.02	.98
		45	46	7	42.500	1	.02	.98
		48	38	7	34.500	0	.00	1.00
		51	31	5	28.500	0	.00	1.00
		54	26	8	22.000	0	.00	1.00
		57	18	6	15.000	0	.00	1.00
		60	12	12	6.000	0	.00	1.00
	2 E-service	0	217	0	217.000	1	.00	1.00
		3	216	0	216.000	3	.01	.99
		6	213	2	212.000	2	.01	.99
		9	209	2	208.000	4	.02	.98
		12	203	2	202.000	5	.02	.98
		15	196	2	195.000	6	.03	.97

Life tables

- ◆ SPSS outputs: life table
 - *Interval Start Time*. The beginning value for each interval. Each interval extends from its start time up to the start time of the next interval.
 - *Number Withdrawing during Interval*: the number of **censored** cases in this interval. These are still active customers, but so far they have not been customers longer than the time period indicated by this interval.
 - *Number Exposed to Risk*. The number of surviving cases minus one half the censored cases. This is intended to account for the effect of the censored cases.

Life tables

- SPSS outputs: life table
 - *Number of Terminal Events*. The number of cases that experience the terminal event in this interval. These are customers with churn = 1.
 - *Proportion Terminating*. The ratio of terminal events to the number exposed to risk ($10/264.5=0.0378$).
 - *Proportion Surviving*. One minus the proportion terminating.

Life tables

SPSS Outputs: life table

Cumulative Proportion Surviving at End of Interval	Std. Error of Cumulative Proportion Surviving at End of Interval	Probability Density	Std. Error of Probability Density	Hazard Rate	Std. Error of Hazard Rate
.96	.01	.013	.004	.01	.00
.90	.02	.022	.005	.02	.01
.86	.02	.014	.004	.02	.00
.81	.02	.014	.004	.02	.01
.78	.03	.009	.004	.01	.00
.76	.03	.008	.004	.01	.00
.75	.03	.002	.002	.00	.00
.73	.03	.008	.004	.01	.01
.71	.03	.008	.004	.01	.01
.69	.03	.004	.003	.01	.00
.67	.03	.009	.005	.01	.01
.64	.04	.008	.005	.01	.01
.62	.04	.006	.004	.01	.01
.59	.04	.011	.006	.02	.01
.58	.04	.004	.004	.01	.01
.56	.04	.005	.004	.01	.01
.56	.04	.000	.000	.00	.00
.56	.04	.000	.000	.00	.00
.56	.04	.000	.000	.00	.00
.56	.04	.000	.000	.00	.00
.56	.04	.000	.000	.00	.00

Life tables

- ◆ SPSS Outputs: life table
 - *Cumulative Proportion Surviving at End of Interval*. The proportion of cases surviving from the start of the table to the end of the interval $(266-10-17)/266=0.8984$ (second row).
 - *Probability Density*. An estimate of the probability of experiencing the terminal event during the interval.
 - *Hazard Rate*. An estimate of experiencing the terminal event during the interval, conditional upon surviving to the start of the interval.

Life tables



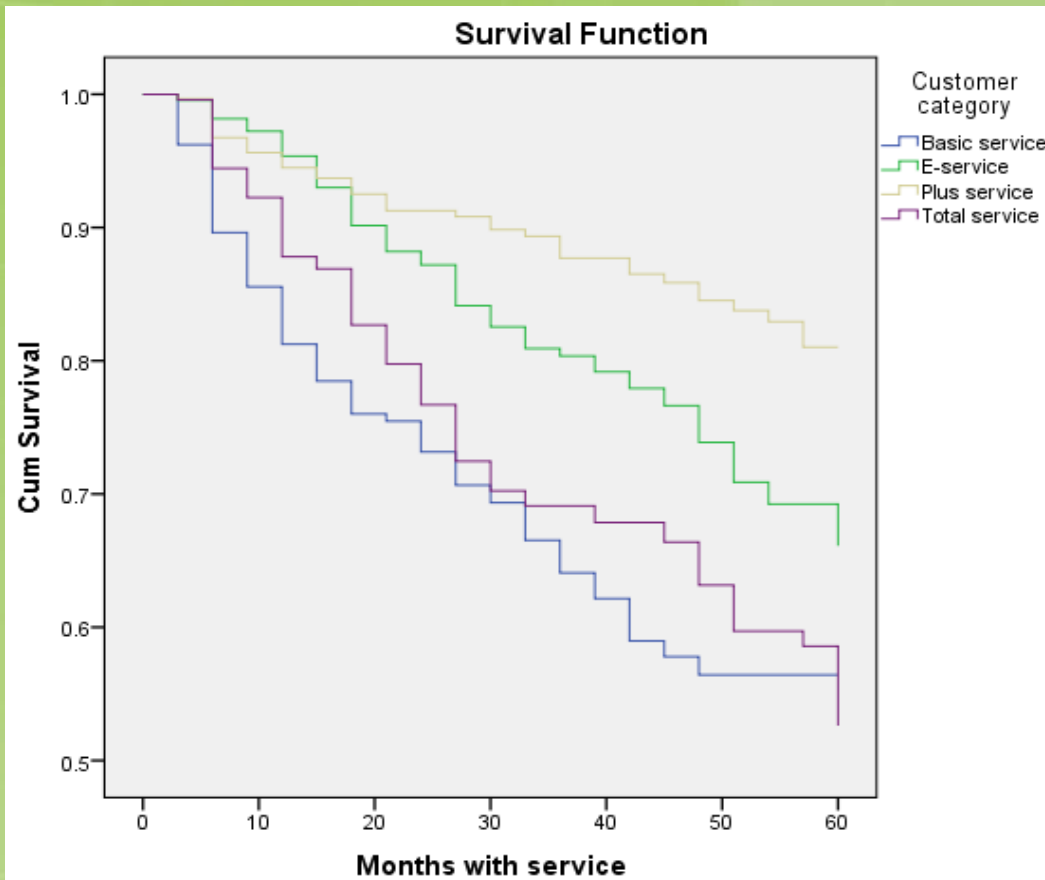
SPSS Outputs: life table

- The greatest number and proportion of terminal events occur within the first year, which suggests that customers should be monitored more closely during their first year to be sure of their satisfaction with the company's service.

Life Table								
		Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving	
First-order Controls custcat Customer category	1 Basic service	Interval Start Time 0	266	3	264.500	10	.04	.96
		3	253	10	248.000	17	.07	.93
		6	226	12	220.000	10	.05	.95
		9	204	11	198.500	10	.05	.95
		12	183	13	176.500	6	.03	.97
		15	164	10	159.000	5	.03	.97
		18	149	15	141.500	1	.01	.99
		21	133	6	130.000	4	.03	.97
		24	123	10	118.000	4	.03	.97
		27	109	4	107.000	2	.02	.98
		30	103	9	98.500	4	.04	.96
		33	90	17	81.500	3	.04	.96
		36	70	7	66.500	2	.03	.97
		39	61	5	58.500	3	.05	.95
		42	53	6	50.000	1	.02	.98
		45	46	7	42.500	1	.02	.98
		48	38	7	34.500	0	.00	1.00
		51	31	5	28.500	0	.00	1.00
		54	26	8	22.000	0	.00	1.00
		57	18	6	15.000	0	.00	1.00
		60	12	12	6.000	0	.00	1.00
	2 E-service	0	217	0	217.000	1	.00	1.00
		3	216	0	216.000	3	.01	.99
		6	213	2	212.000	2	.01	.99
		9	209	2	208.000	4	.02	.98
		12	203	2	202.000	5	.02	.98
		15	196	2	195.000	6	.03	.97

Life tables

SPSS Outputs: survival function



1. The horizontal axis shows the time to event. The vertical axis shows the probability of survival.
2. Any point on the survival curve shows the probability that a customer of a given service category will **remain a** customer past that time.
3. Total service and Basic service customers have the lowest survival curves, and E-service customers have lower curves than Plus service customers.

Life tables

SPSS Outputs

Comparisons for Control Variable: custcat

Overall Comparisons^a

Wilcoxon (Gehan) Statistic	df	Sig.
49.179	3	.000

a. Comparisons are exact.

Pairwise Comparisons^a

(I) custcat	(J) custcat	Wilcoxon (Gehan) Statistic	df	Sig.
1	2	18.640	1	.000
	3	37.154	1	.000
	4	2.949	1	.086
2	1	18.640	1	.000
	3	5.515	1	.019
	4	9.222	1	.002
3	1	37.154	1	.000
	2	5.515	1	.019
	4	27.229	1	.000
4	1	2.949	1	.086
	2	9.222	1	.002
	3	27.229	1	.000

a. Comparisons are exact.

1. Wilcoxon test is used to compare survival distribution among groups, with the test statistic based on differences in group mean scores.
2. Since the significance value of the test is less than 0.05, we conclude that the survival curves are different across the group.
3. Pairwise comparisons show which two groups are significantly different in survival curves.

Kaplan-Meier procedure

- ◆ The Kaplan-Meier procedure is a method of estimating time-to-event models in the presence of censored cases.
- ◆ A descriptive procedure for examining the distribution of time-to-event variables. We also can compare the distribution by levels of a factor variable or produce separate analyses by levels of a stratification variable.
- ◆ Censored cases (right-censored cases) are those for which the event of interest has not yet happened.

Kaplan-Meier procedure



Assumptions

- Probabilities for the event of interest should depend **only on time** after the initial event without covariates effects.
- Cases that enter the study at different times (for example, patients who begin treatment at different times) should behave similarly.
- Censored and uncensored cases behave the same. If, for example, many of the censored cases are patients with more serious conditions, your results may be biased.

Kaplan-Meier procedure

- ◆ *Example* (from IBM SPSS 20.0) : data file: pain_medication
 - A pharmaceutical company is developing an anti-inflammatory medication for treating chronic arthritic pain. The research interest is the time it takes for the drug to take effect and how it compares to an existing drug. Shorter times to effect are considered better.
- ◆ Event: drug takes effect

Kaplan-Meier procedure



Variables

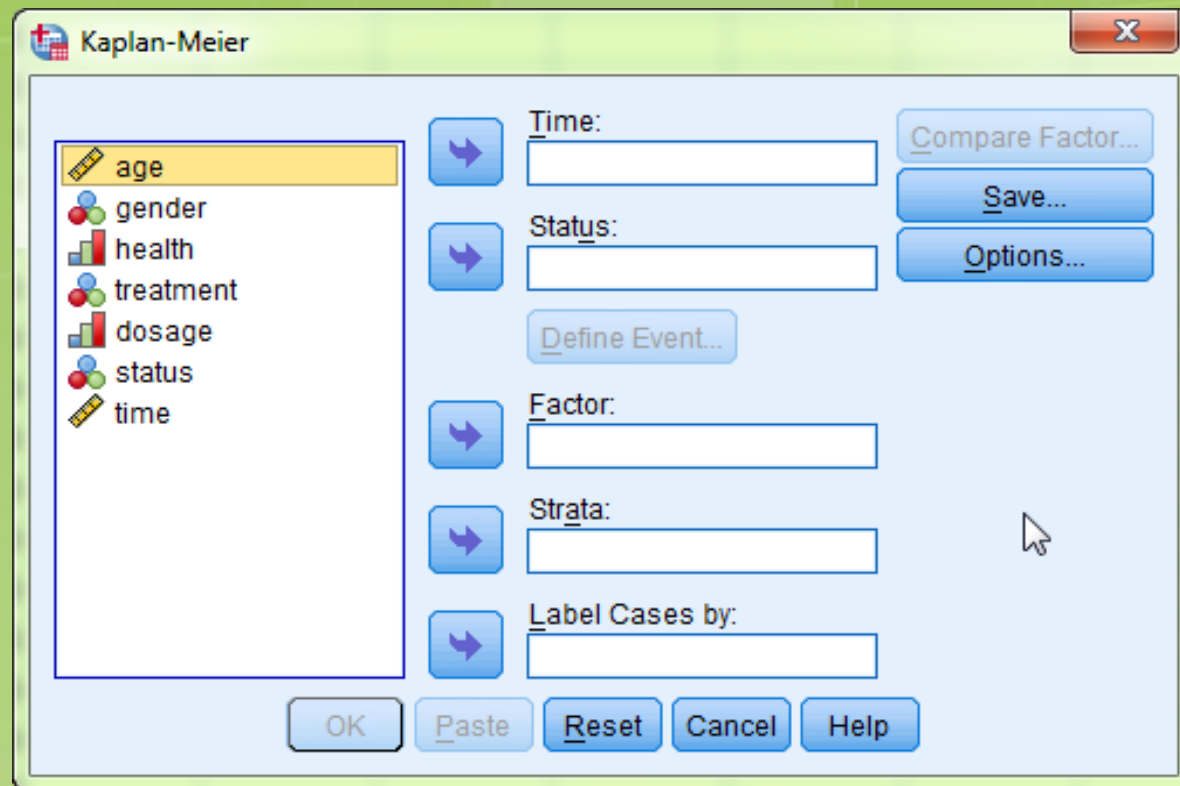
- Time variable (duration variable): must be a continuous variable
- Status variable: categorical or continuous variable, represents the event of interest (drug has effect or not).
- Factor variable: categorical variable, represents a causal effect (type of treatment for example).
- Stratification variable: categorical variable.

Kaplan-Meier procedure

- ◆ We have Factor variable: Treatment (0 = New drug, 1 = old drug), Status variable: status (0 = censored, 1 = take effect), Time variable: time
- ◆ We want to compare the effect of two different drugs.
- ◆ Null hypothesis: whether survival function is the same between different levels of the factor variable (between old and new drug) .

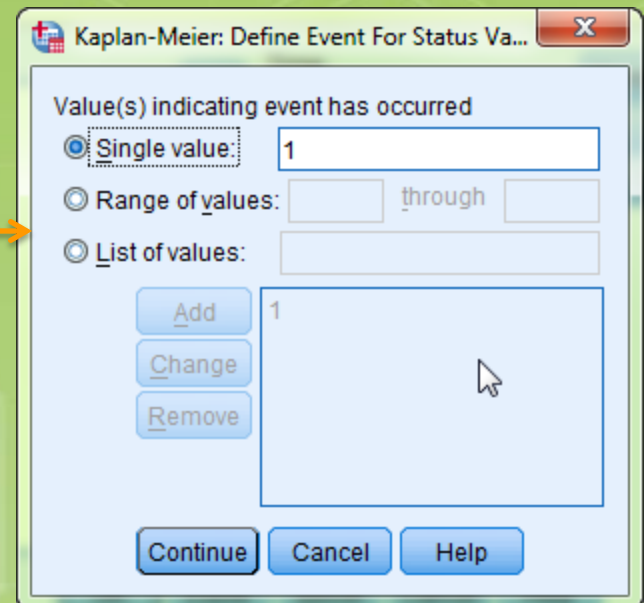
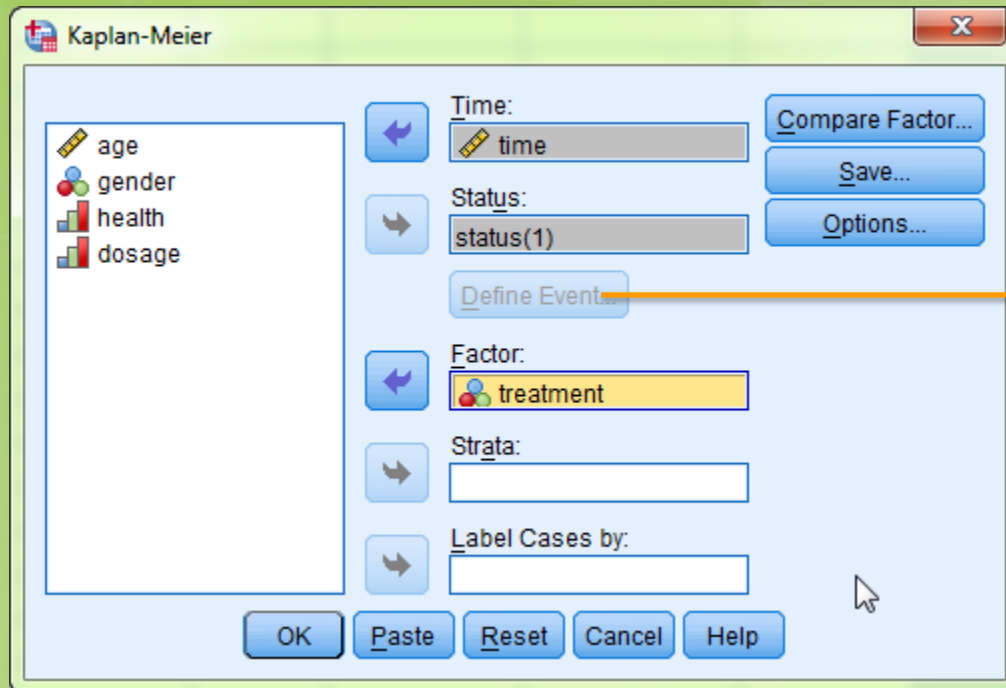
Kaplan-Meier procedure

- Go to Analyze > Survival > Kaplan-Meier



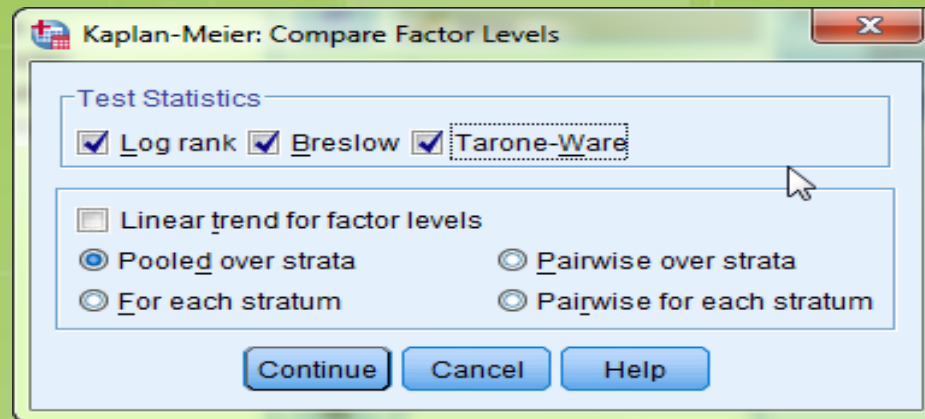
Kaplan-Meier procedure

◆ Analyze data



Kaplan-Meier procedure

Click *Compare Factor*



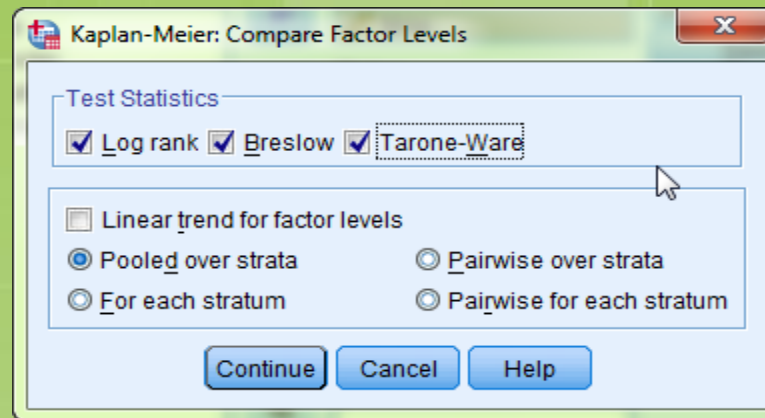
Log rank: Tests equality of survival functions by weighting all time points the same.

Breslow: Tests equality of survival functions by weighting all time points by the number of cases at risk at each time point.

Tarone-Ware: Tests equality of survival functions by weighting all time points by the square root of the number of cases at risk at each time point.

Kaplan-Meier procedure

Compare factor



Pooled over strata: a single test is computed for all factor levels, testing for equality of survival function across all levels of the factor variable.

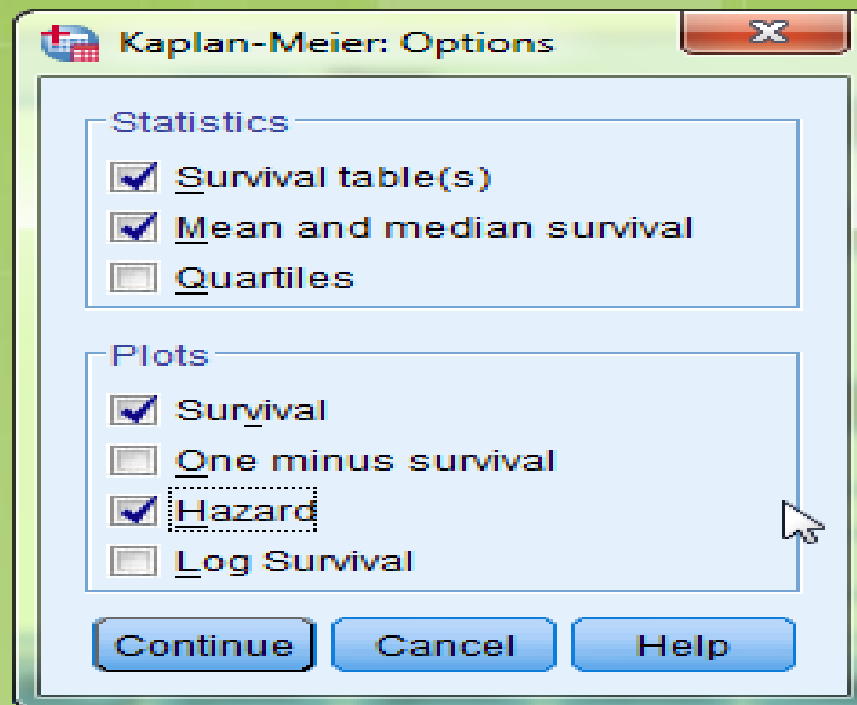
Pairwise over strata: a separate test is computed for each pair of factor levels when a pooled test shows non-equality of survival functions.

For each stratum: a separate test is computed for group formed by the stratification variable.

Pairwise for each stratum: a separate test is computed for each pair of factor variable, for each stratum of the stratification variable.

Kaplan-Meier procedure

Click *Options*



Kaplan-Meier procedure

SPSS outputs: survival table

Case Processing Summary				
treatment Treatment	Total N	N of Events	Censored	
			N	Percent
0 New drug	104	79	25	24.0%
1 Existing drug	96	74	22	22.9%
Overall	200	153	47	23.5%

The survival table is a descriptive table that details the time until the drug takes effect. The table is sectioned by each level of Treatment, and each observation occupies its own row in the table. The table is very large.

Survival Table								
treatment Treatment	ID	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases	
				Estimate	Std. Error			
0 New drug	1	1006	.600	1 Taken effect	.	.	1	103
	2	1075	.600	1 Taken effect	.981	.013	2	102
	3	1057	.700	1 Taken effect	.971	.016	3	101
	4	1159	.800	1 Taken effect	.962	.019	4	100
	5	1094	.900	1 Taken effect	.952	.021	5	99
	6	1009	1.100	1 Taken effect	.	.	6	98
	7	1135	1.100	1 Taken effect	.933	.025	7	97
	8	1001	1.200	1 Taken effect	.	.	8	96
	9	1171	1.200	1 Taken effect	.913	.028	9	95
	10	1032	1.300	1 Taken effect	.	.	10	94
	11	1048	1.300	1 Taken effect	.	.	11	93
	12	1108	1.300	1 Taken effect	.885	.031	12	92
	13	1105	1.400	1 Taken effect	.875	.032	13	91
	14	1049	1.500	1 Taken effect	.	.	14	90
	15	1133	1.500	1 Taken effect	.	.	15	89
	16	1197	1.500	1 Taken effect	.846	.035	16	88
	17	1024	1.600	1 Taken effect	.	.	17	87
	18	1051	1.600	1 Taken effect	.827	.037	18	86
	19	1037	1.700	1 Taken effect	.	.	19	85
	20	1086	1.700	1 Taken effect	.	.	20	84
	21	1122	1.700	1 Taken effect	.798	.039	21	83
	22	1119	1.800	1 Taken effect	.	.	22	82
	23	1179	1.800	1 Taken effect	.	.	23	81
	24	1185	1.800	1 Taken effect	.	.	24	80
	25	1186	1.800	1 Taken effect	.760	.042	25	79
	26	1007	1.900	1 Taken effect	.	.	26	78
	27	1169	1.900	1 Taken effect	.	.	27	77
	28	1181	1.900	1 Taken effect	.731	.043	28	76

Kaplan-Meier procedure



Survival table

- Time. The time at which the event or censoring occurred.
- Status. Indicates whether the case experienced the terminal event or was censored.
- Cumulative Proportion Surviving at the Time. The proportion of cases surviving from the start of the table until this time. When multiple cases experience the terminal event at the same time, these estimates are printed once for that time period and apply to all the cases whose drug took effect at that time.

Kaplan-Meier procedure



Survival table

- N of Cumulative Events. The number of cases that have experienced the terminal event from the start of the table until this time.
- N of Remaining Cases. The number of cases that, at this time, haven't yet to experience the terminal event or be censored.

Kaplan-Meier procedure

Means and Median for Survival time

Means and Medians for Survival Time								
treatment Treatment	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
0 New drug	4.867	.360	4.162	5.572	3.700	.292	3.128	4.272
1 Existing drug	5.185	.350	4.499	5.871	4.100	1.131	1.884	6.316
Overall	5.014	.252	4.520	5.507	3.900	.272	3.367	4.433

a. Estimation is limited to the largest survival time if it is censored.

Since there is a lot of overlap in the confidence intervals, it is unlikely that there is much difference in the "average" survival time. If confidence intervals do not overlap between levels, differences in effect on time to event can be inferred.

Kaplan-Meier procedure

Overall comparison

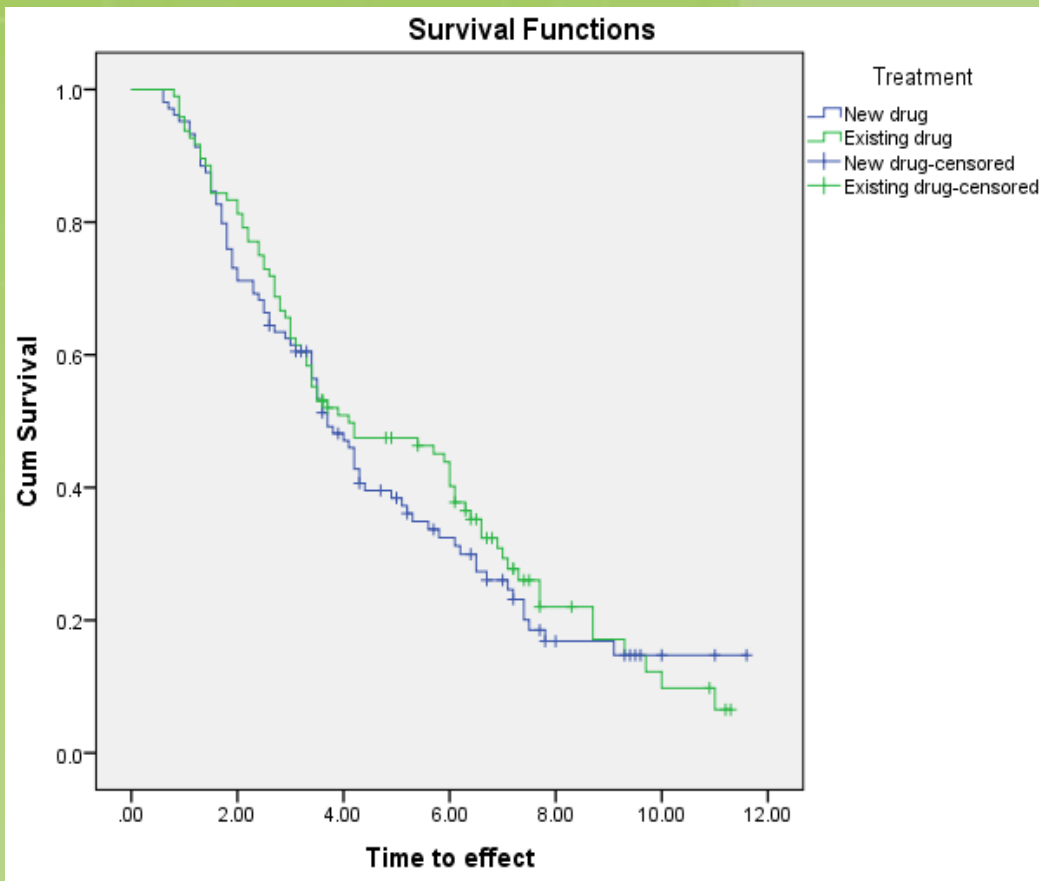
Overall Comparisons			
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.379	1	.538
Breslow (Generalized Wilcoxon)	.748	1	.387
Tarone-Ware	.705	1	.401

Test of equality of survival distributions for the different levels of treatment Treatment.

This table provides overall tests of the equality of survival times across groups. Since the significance values of the tests are all greater than 0.05, there is no statistically significant difference between two treatments in survival time.

Kaplan-Meier procedure

Survival function



1. The survival curves give a visual representation of the life tables.
2. The horizontal axis shows the time to event. In this plot, drops in the survival curve occur whenever the medication takes effect in a patient.
3. The vertical axis shows the probability of survival (probability of **not experience** the treatment effect).

Cox regression

- ❖ The Cox Regression procedure is useful for modeling the time to a specified event, based upon the values of given **covariates**.
- ❖ One or more covariates are used to predict a status (event).
- ❖ The central statistical output is the **hazard ratio**.
- ❖ Data contain censored and uncensored cases.
- ❖ Similar to logistic regression, but Cox regression assesses relationship between survival time and covariates .

Cox regression

Terms

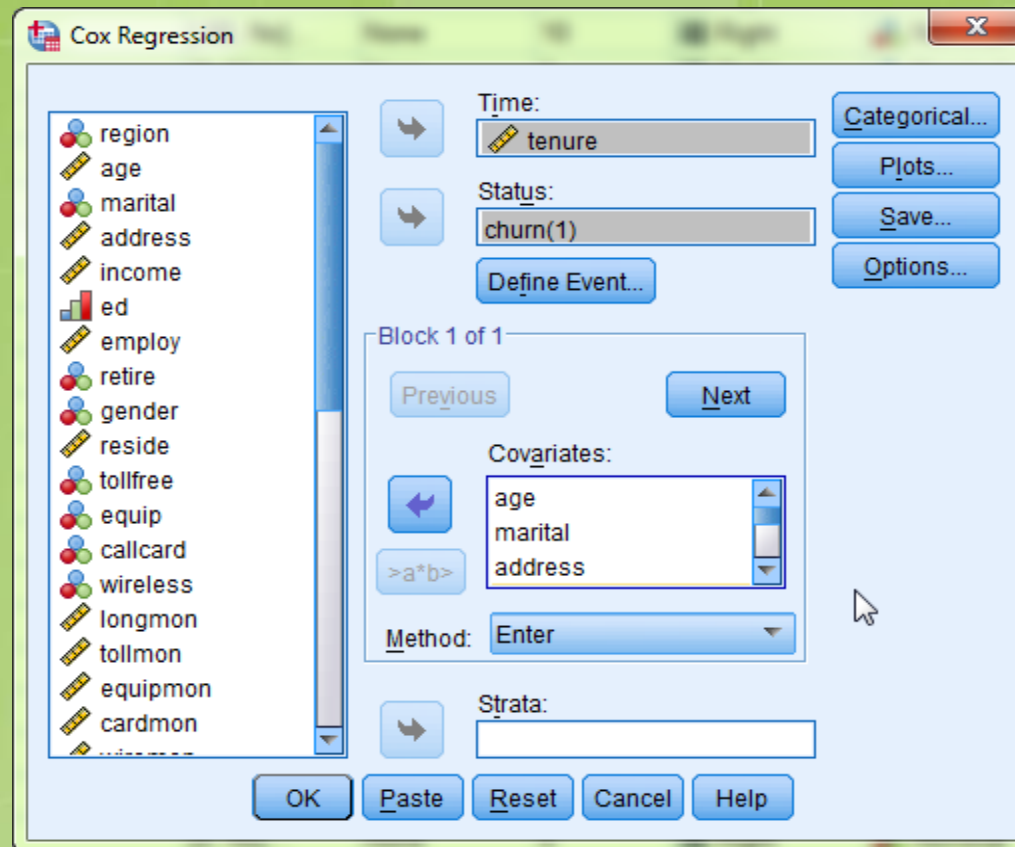
- Status variable: the dependent in Cox regression, should be binary variable.
- Time variable: measures duration to the event defined by the status variable (continuous or discrete).
- Covariates: independent/predictor variables. They can be categorical or continuous. They also can be time-fixed or time-dependent.
 - Interaction terms
 - Categorical covariates: SPSS automatically convert them into a set of dummy variables, omitting one category.

Cox regression

- ◆ **Example** (from SPSS): data file: telco
 - Use Cox Regression to determine which attributes are associated with shorter "time to churn".
 - Time variable: tenure (month with services)
 - Status variable: churn (0 = No, 1 = Yes).
 - Covariates: age, marital, address, employ, retire, gender, reside, and custcat
- ◆ Go to Analyze > Survival > Cox Regression

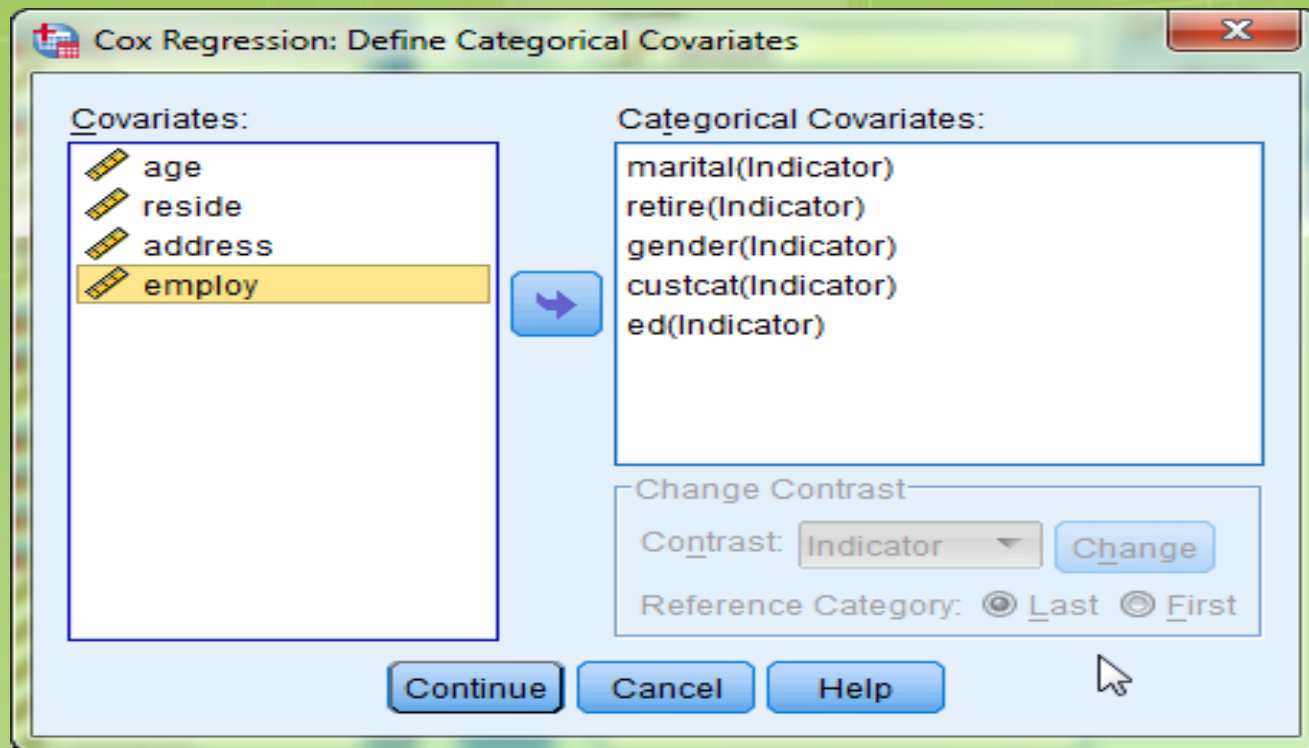
Cox regression

Run Cox regression



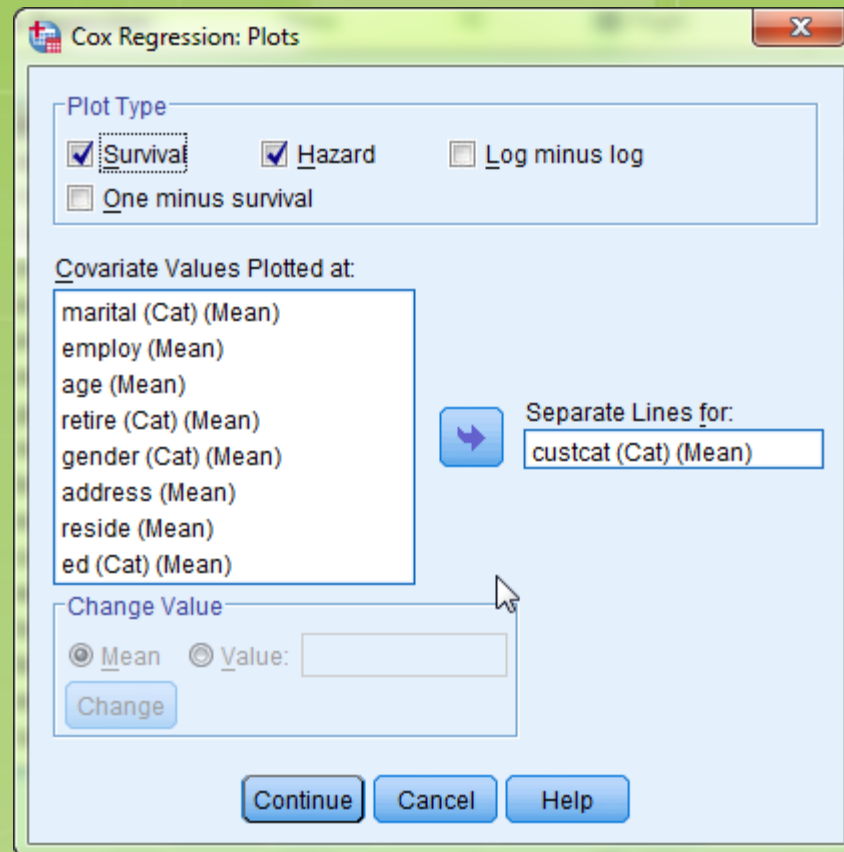
Cox regression

- Click Categorical



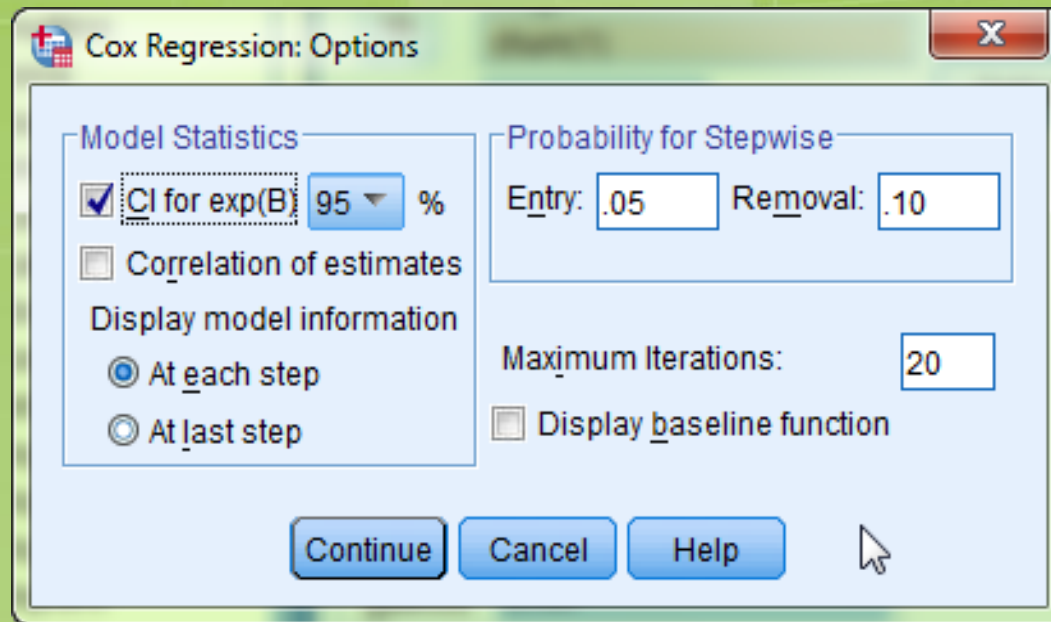
Cox regression

Click Plots



Cox regression

- Click Options



Cox regression

SPSS Outputs

Categorical Variable Codings^{a,d,e,f,g}

	Frequency	(1) ^c	(2)	(3)	(4)
marital ^b 0=Unmarried	505	1			
1=Married	495	0			
ed ^b 1=Did not complete high school	204	1	0	0	0
2=High school degree	287	0	1	0	0
3=Some college	209	0	0	1	0
4=College degree	234	0	0	0	1
5=Post-undergraduate degree	66	0	0	0	0
retire ^b .00=No	953	1			
1.00=Yes	47	0			
gender ^b 0=Male	483	1			
1=Female	517	0			
custcat ^b 1=Basic service	266	1	0	0	
2=E-service	217	0	1	0	
3=Plus service	281	0	0	1	
4=Total service	236	0	0	0	

a. Category variable: marital (Marital status)

b. Indicator Parameter Coding

c. The (0,1) variable has been recoded, so its coefficients will not be the same as for indicator (0,1) coding.

d. Category variable: ed (Level of education)

e. Category variable: retire (Retired)

f. Category variable: gender (Gender)

g. Category variable: custcat (Customer category)

Cox regression

SPSS Outputs

Block 1: Method = Enter

Omnibus Tests of Model Coefficients^a

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
3242.673	236.536	14	.000	283.691	14	.000	283.691	14	.000

a. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
age	-.005	.008	.303	1	.582	.995	.979	1.012
marital	.333	.162	4.243	1	.039	1.395	1.016	1.915
address	-.059	.010	33.348	1	.000	.943	.924	.962
ed			8.100	4	.088			
ed(1)	-.464	.278	2.780	1	.095	.629	.364	1.085
ed(2)	-.264	.238	1.221	1	.269	.768	.481	1.226
ed(3)	-.180	.237	.579	1	.447	.835	.525	1.329
ed(4)	.101	.218	.213	1	.644	1.106	.721	1.695
employ	-.073	.012	38.686	1	.000	.930	.909	.951
retire	.323	.611	.280	1	.597	1.381	.417	4.572
gender	-.037	.122	.092	1	.761	.964	.759	1.224
reside	-.059	.058	1.027	1	.311	.943	.842	1.056
custcat			29.583	3	.000			
custcat(1)	.285	.168	2.885	1	.089	1.330	.957	1.847
custcat(2)	-.539	.171	9.880	1	.002	.584	.417	.816
custcat(3)	-.502	.196	6.532	1	.011	.605	.412	.890

Cox regression

- ◆ SPSS Outputs: variables in the equation
 - $\text{Exp}(B)$, which can be interpreted as the predicted change in the hazard for a unit increase in the predictor.
 - For binary covariates, hazard ratio is the estimate of the ratio of the hazard rate in one group to the hazard rate in another group.
 - The value of $\text{Exp}(B)$ for marital means that the churn hazard for an unmarried customer is 1.395 times that of a married customer.
 - The value of $\text{Exp}(B)$ for address means that the churn hazard is reduced by $100\% - (100\% \times 0.943) = 5.7\%$ for each year a customer has lived at the same address.

Cox regression

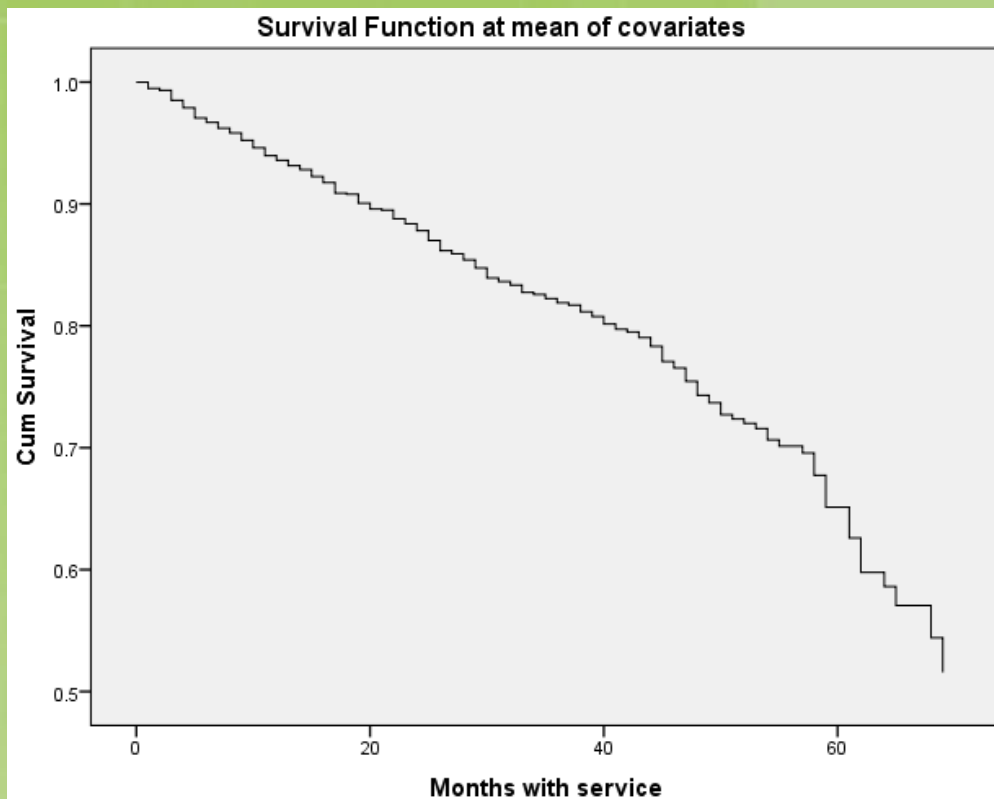
SPSS Outputs

Covariate Means and Pattern Values					
	Mean	Pattern			
		1	2	3	4
age	41.684	41.684	41.684	41.684	41.684
marital	.505	.505	.505	.505	.505
address	11.551	11.551	11.551	11.551	11.551
ed(1)	.204	.204	.204	.204	.204
ed(2)	.287	.287	.287	.287	.287
ed(3)	.209	.209	.209	.209	.209
ed(4)	.234	.234	.234	.234	.234
employ	10.987	10.987	10.987	10.987	10.987
retire	.953	.953	.953	.953	.953
gender	.483	.483	.483	.483	.483
reside	2.331	2.331	2.331	2.331	2.331
custcat(1)	.266	1.000	.000	.000	.000
custcat(2)	.217	.000	1.000	.000	.000
custcat(3)	.281	.000	.000	1.000	.000

This table displays the average value of each predictor variable, plus a pattern for each level of custcat. The four patterns correspond to each of the customer types, each with otherwise "average" predictor values.

Cox regression

SPSS Outputs: survival function



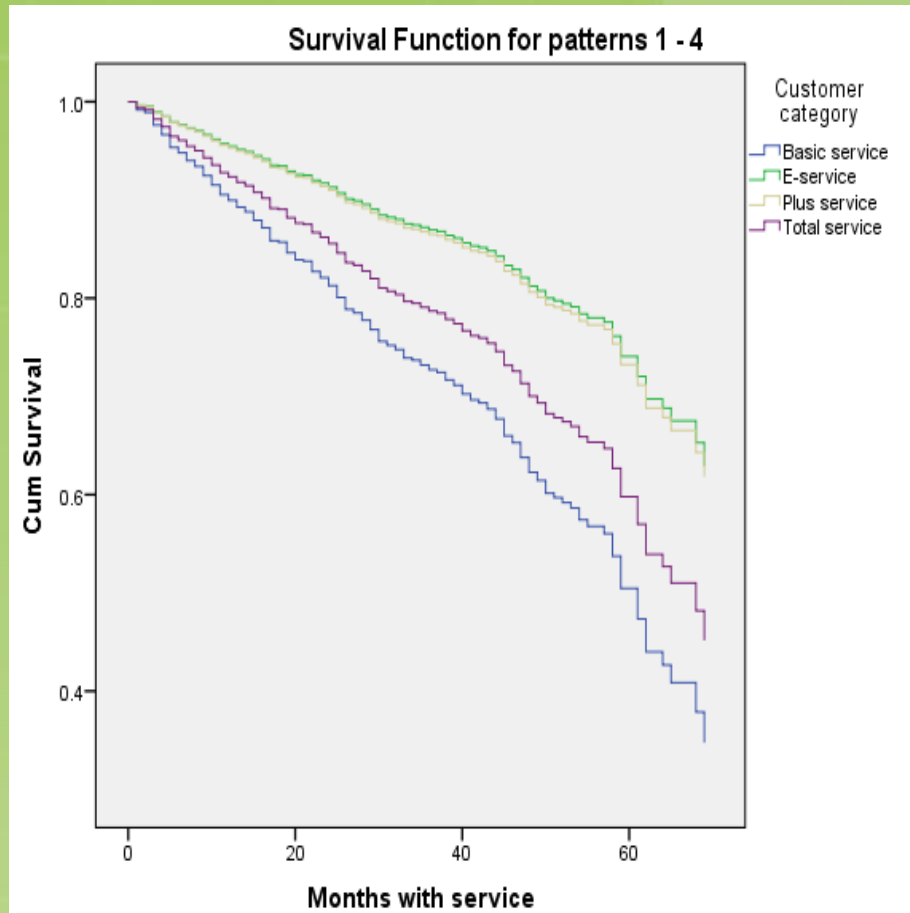
The basic survival curve is a visual display of the model-predicted time to churn for the "average" customer.

The horizontal axis shows the time to event.

The vertical axis shows the probability of survival.

Cox regression

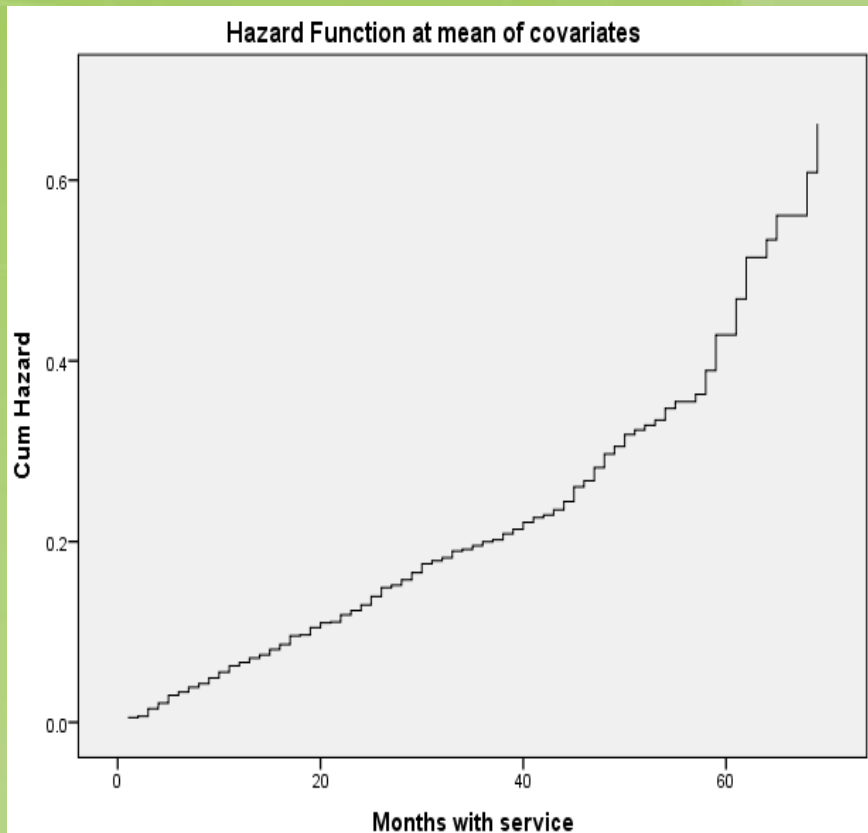
SPSS Outputs



1. The plots show the effect of customer category.
2. Total service and Basic service customers have lower survival curves because, they are more likely to have shorter times to churn.

Cox regression

SPSS Outputs



1. The horizontal axis shows the time to event.
2. The vertical axis shows the cumulative hazard, equal to the negative log of the survival probability.

Cox regression with a time-dependent covariate



Time-dependent covariates

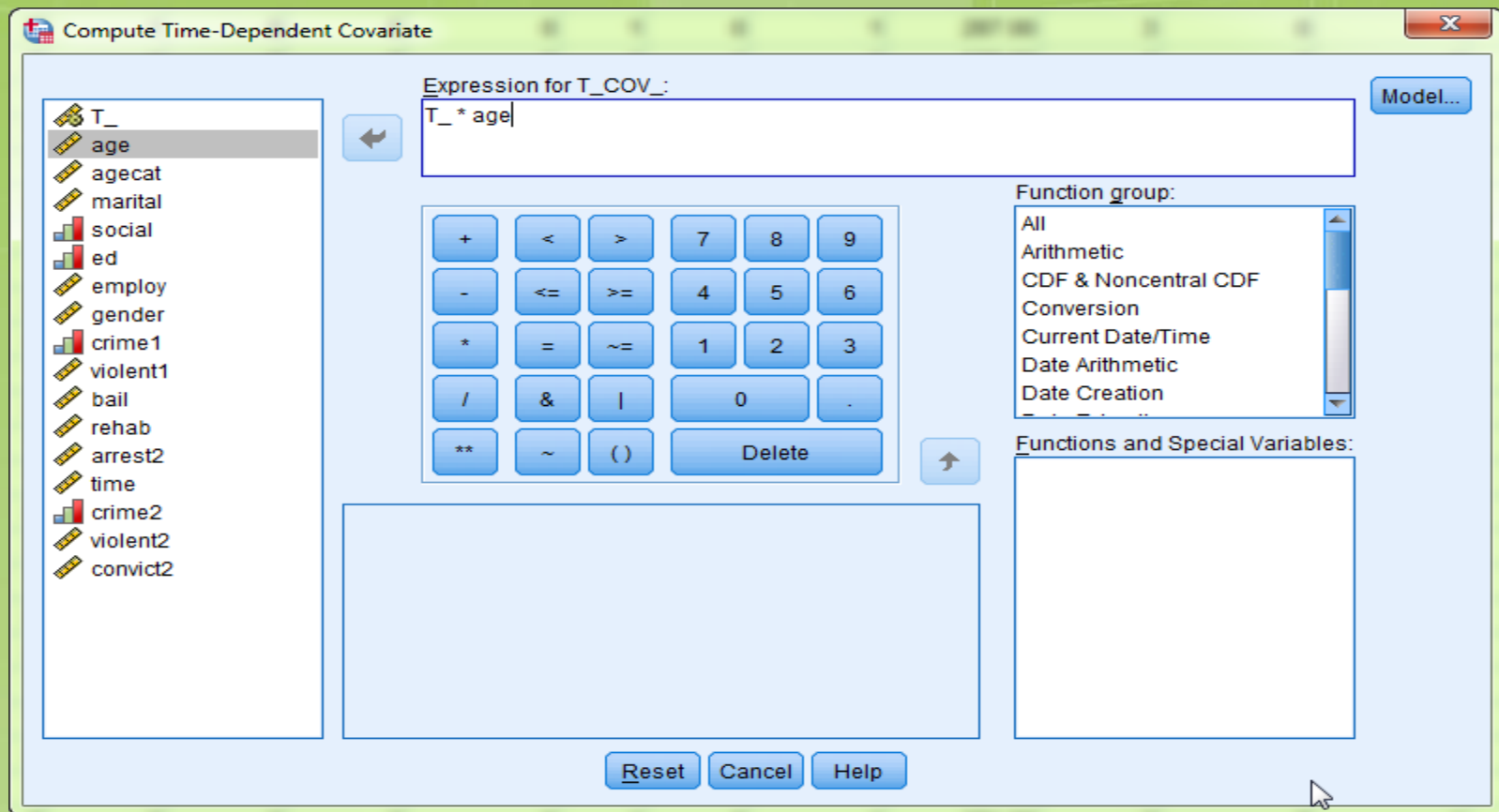
- Simple product of the time variable and the covariate.
- Some variables may have different values at different time periods but aren't systematically related to time. In such cases, you need to define a segmented time-dependent covariate, which can be done using logical expressions.

Cox regression with a time-dependent covariates

- ◆ **Example** (from IBM SPSS 20.0): data file: recidivism
 - A government law enforcement agency is concerned about recidivism rates in their area of jurisdiction. One of the measures of recidivism is the time until second arrest for offenders. The agency would like to model time to rearrest using Cox Regression, but are worried the proportional hazards assumption is invalid across age categories.
- ◆ Go to Analyze > Survival > Cox with time-dependent covariate

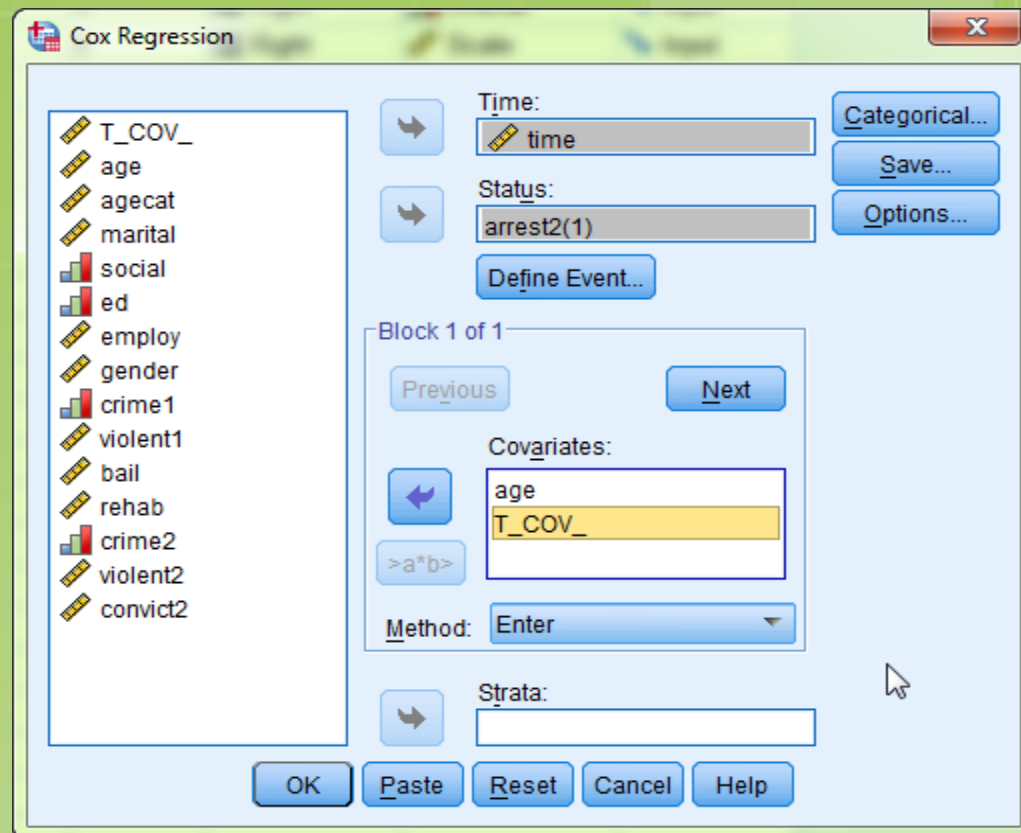
Cox regression with a time-dependent covariate

Run analysis: the product of variable T_ and variable age



Cox regression with a time-dependent covariate

- Click Model and back to Cox regression window



The image shows the 'Cox Regression' dialog box in SPSS. On the left is a list of variables: T_COV_, age, agecat, marital, social, ed, employ, gender, crime1, violent1, bail, rehab, crime2, violent2, and convict2. In the center, the 'Time' variable is set to 'time' and the 'Status' variable is set to 'arrest2(1)'. Below these, 'Block 1 of 1' is shown with 'Previous' and 'Next' buttons. The 'Covariates' section contains 'age' and 'T_COV_'. The 'Method' is set to 'Enter'. At the bottom, there is a 'Strata' section and buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. On the right side of the dialog, there are buttons for 'Categorical...', 'Save...', and 'Options...'.

Time: time

Status: arrest2(1)

Block 1 of 1

Previous Next

Covariates: age T_COV_

Method: Enter

Strata:

OK Paste Reset Cancel Help

Categorical... Save... Options...

Dependent variable:
arrest2

Time variable: time
Covariates: age and
new product
variable: T_cov

Cox regression with a time-dependent covariate

SPSS outputs

Variables in the Equation								
	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
age	-.026	.010	6.346	1	.012	.975	.955	.994
T_COV_	.000	.000	10.736	1	.001	1.000	1.000	1.000

Covariate Means	
	Mean
age	29.095
T_COV_	8068.214

The time-dependent covariate has a significance value less than 0.05, which means it contributes to the model, but the value of the coefficient is very small.

We found that the effect of age on recidivism is time-dependent, and added a term to the model that helps to account for that dependence.

The image features a green background with a subtle pattern of hexagons. A dark green horizontal bar is at the top. The text "THANK YOU" is centered in a bold, yellow, 3D-style font.

THANK YOU