Seminar notes

Medical Biostatistics 2

Georg Heinze

Section of Clinical Biometrics
Core Unit for Medical Statistics and Informatics
Medical University of Vienna
Spitalgasse 23, A-1090 Vienna, Austria

e-mail: georg.heinze@meduniwien.ac.at

Version 2008-04

Part 1: Statistical models

Class 1: General issues on statistical modeling

Statistical tests and statistical models

In the basic course on Medical Biostatistics, several statistical tests were introduced. The course closed by presenting a statistical model, the linear regression model. Here, we start with a review of statistical tests and show how they can be represented as statistical models. Then we extend the idea of statistical models and discuss application, presentation of results, and other issues related to statistical modeling.

What is a statistical test?

In its simplest setting, a statistical test compares the values of a variable between two groups. Often we want to infer whether two groups of patients actually belong to the same population. We specify a null hypothesis and reject it if the observed data does not give evidence that the hypothesis holds. For simplification we restrict the hypothesis to the comparison of means, as the mean is the most important and most obvious feature of any distribution. If our patient groups belong to the same population, they should exhibit the same mean. Thus, our null hypothesis states “the means in the two groups are equal”.

To perform the statistical test, we need two pieces of information for each patient: his/her group membership, and his/her value of the variable to be compared. (And so far, it is of no importance whether the variable we want to compare is a scale or a nominal variable.)

In short, a statistical test verifies hypotheses about the study population.

As an example, consider the rat diet example of the basic lecture. We tested the equality of weight gains between the groups of high protein diet and low protein diet.

What is a statistical model?

A statistical model establishes a relationship between variables, e. g., a rule how to predict a patient’s cholesterol level from his age and body mass index. Estimating the model parameters, we can quantify this relationship and (hopefully) predict cholesterol levels:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	153.115	8.745		17.509	.000
	Body-mass-index	1.179	.326	.283	3.620	.001
	Age	.756	.091	.648	8.293	.000

a. Dependent Variable: Cholesterol level

In this model, we would estimate a patient's cholesterol level from age and body-mass-index as

Cholesterol = 153.1+1.179*BMI + 0.756*Age

The regression coefficients (parameters) are:

1.179 for BMI
0.756 for age.

They have the following interpretation:

Comparing two patients of the same age which differ in their BMI by 1 kg/m2, the heavier person's cholesterol level is on average 1.179 units higher than that of the slimmer person.

and

Comparing two patients with the same BMI which differ in their age by one year, the older person will on average have a cholesterol level 0.756 units higher than the younger person.

The column labeled "Sig." informs us whether these coefficients can be assumed to be 0, the p-values in that column refer to testing that the corresponding regression coefficients are zero. If they were actually zero, then these variables had no effect on cholesterol, as can be demonstrated easily:

Cholesterol = 180 + 0*BMI + 0*Age

In the above equation, the cholesterol level is completely independent from BMI and age. No matter which values we insert for BMI or Age, the cholesterol level will not change from 180.

Summarizing, we can get out more of a statistical model than we can get out of a statistical test: not only do we test the hypothesis of 'no relationship', we also obtain an estimate of the magnitude of the relationship, and even a prediction rule for cholesterol.

Response or outcome variable

Statistical models, in their simplest form, and statistical tests are related to each other. We can express any statistical test as a statistical model, in which the P-value obtained by statistical testing is delivered as a ‘by-product’.

In our example of a statistical model, the cholesterol level is our outcome or response variable. Generally, any variable we want to compare between groups is an outcome or response variable.

In the rat diet example, the response variable is the weight gain.

Independent variable

The statistical model provides an equation to estimate values of the response variable by one or several independent variables. The denotation ‘independent’ points at their role in the model: their part is an active one; namely to explain differences in response and not to be explained themselves. In our example, these independent variables were BMI and age.

In the rat diet example, we consider the diet group (high or low protein) as independent variable.

The interpretability of estimated regression coefficients is of special importance. Since the interpretation of coefficients is not clear in some models, in the field of medicine such models are seldom used. Models which allow a clear interpretation of their results are generally preferred.

Representing a statistical test by a statistical model

Recall the rat diet example. We can represent the t-test which was applied to the data as linear regression of weight gain on diet group:

Weight gain = b0 + b1*D

where D=1 for the high protein group, and D=0 for the low protein group.

Now the regression coefficients b0 and b1 have a clear interpretation:

b0 is the mean weight gain in the low protein group (because for D=0, we have Weight gain = b0 + b1*0).

b1 is the excess average weight gain in the high protein group, compared to the low protein group, or, put another way, the difference in mean weight gain between the two groups.

Clearly, if b1 is significantly different from zero, then the type of diet influences weight gain. Let’s proof by applying linear regression to the rat diet data:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	139,000	14,575		9,537	,000
	Dietary group	-19,000	10,045	-,417	-1,891	,076

a. Dependent Variable: Weight gain (day 28 to 84)

For comparison, consider the results of the t-test:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Weight gain (g)	Equal variances assumed	,015	,905	1,891	17	,076	19,0000	10,0453	-2,1937	40,1937
	Equal variances not assumed			1,911	13,082	,078	19,0000	9,9440	-2,4691	40,4691

For interpreting the coefficient corresponding to ‘Dietary group’, we must know how this variable was coded. Actually, 1 was the code for the high protein group, and 2 for the low protein group. Inserting the codes into the regression model we obtain

Weight gain = 139 – 19 = 120 for the high protein group and
Weight gain = 139 – 19*2 = 101 for the low protein group,

which exactly reproduces the means of weight gain in the two groups. The p-value associated with Dietary group exactly resembles that of a two-sample t-test.

Other relationships exist for other statistical tests, e. g., the chi-square test has its analogue in logistic regression, or the log-rank test for comparing survival data can be expressed as a simple Cox regression model. Both will be demonstrated in later sessions.

Uncertainty of a model

Since a model is estimated from a sample of limited size, we cannot be sure that the estimated values resemble exactly those of the underlying population. Therefore, it is important that when reporting results we also state how precise our estimates are. This is usually done by supplying confidence intervals in addition to point estimates.

Even in the hypothetical case where we actually *know* the population values of regression coefficients, the structure of the equation may be insufficient to predict a patient's outcome with 100% certainty. Therefore, we should give an estimate of the predictive accuracy of a model. In linear regression, such a measure is routinely computed by any statistical software, it's called R-squared. This measure (sometimes called the coefficient of determination) describes the proportion of variance of the outcome variable that can be explained by variation in the independent variables. Usually, we don't know or we won't consider all the causes of variation of the outcome variable. Therefore, R-squared seldom approaches 100%.

In logistic or Cox regression models, there is no unique definition of R-squared. However, some suggestions have been made and some of them are implemented in SPSS. In these kinds of models, R-squared is typically lower than in linear regression models. For logistic regression models, this is a consequence of the discreteness of the outcome variable. Usually we can only estimate the *percentage* of patients that will experience the event of interest. This means, that we know how many patients on average will have the event, but we cannot predict exactly who of them will or won't. In survival (Cox) models, it's the longitudinal nature of the outcome which prohibits its precise prediction.

Summarizing, there are two sources of uncertainty related to statistical models: one source is due to limited sample sizes, and the other source due to limited ability of a model's structure to predict the outcome.

Types of responses – types of models

The type of response defines the type of model to use. For scale variables as responses, we will most often use the linear regression model. For binary (nominal) outcomes, the logistic regression model is the model of choice. (There are other models for binary data, but with less appealing interpretability of results.) For survival outcomes (time to event data), the Cox regression model is useful. For repeated measurements on scale outcomes, the analysis of variance for repeated measurements can be applied.

Univariate and multivariable models

A univariate model is the translation of a simple statistical test into a statistical model: there is one independent variable and one response variable. The independent variable may be nominal, ordinal or scale.

A multivariable model uses more than one independent variable to explain the outcome variable. Multivariable models can be used for various purpose, some of them are listed in the next subsection but one.

Often, univariate (*crude*) and multivariable (*adjusted*) models are contrasted in one table, as the following example (from a Cox regression analysis) shows [1]:

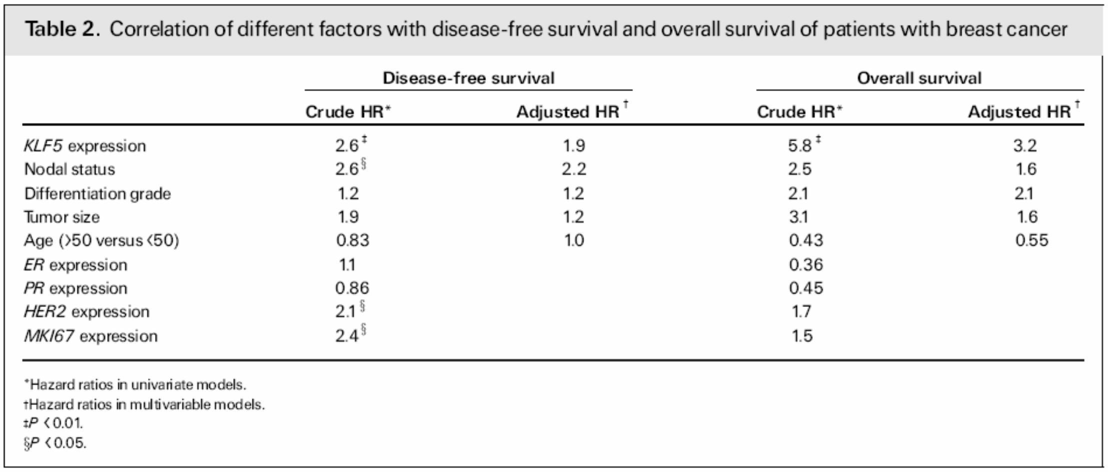


Table 2. Correlation of different factors with disease-free survival and overall survival of patients with breast cancer

	Disease-free survival		Overall survival	
	Crude HR*	Adjusted HR†	Crude HR*	Adjusted HR†
KLF5 expression	2.6‡	1.9	5.8‡	3.2
Nodal status	2.6§	2.2	2.5	1.6
Differentiation grade	1.2	1.2	2.1	2.1
Tumor size	1.9	1.2	3.1	1.6
Age (>50 versus <50)	0.83	1.0	0.43	0.55
ER expression	1.1		0.36	
PR expression	0.86		0.45	
HER2 expression	2.1§		1.7	
MKI67 expression	2.4§		1.5	

*Hazard ratios in univariate models.
†Hazard ratios in multivariable models.
‡$P < 0.01$.
§$P < 0.05$.

Univariate and multivariable models may yield different results. These differences are caused by correlation between the independent variables: some of the variation in variable X1 may be reflected by variation in X2.

In the above table, wee see substantial differences in the estimated effects for KLF5 expression, nodal status and tumor size, but not for differentiation grade. It was shown that KLF5 expression is correlated with nodal status and tumor size, but not with differentiation grade. Therefore, the univariate effect of differentiation grade does not change at all by including KLF5 expression into the model. On the other hand, the effect of KLF5 is reduced by about 40%, caused by the simultaneous consideration of nodal status and tumor size.

In other examples, the reverse may occur; an effect may be insignificant in a univariate model and only be confirmable statistically if another effect is considered simultaneously:

As an example, consider the relationship of sex and cholesterol level:

			Mean	Std Deviation
Sex	male	Cholesterol level	212,50	15,62
	female	Cholesterol level	215,30	15,85

As outlined earlier, the ‘effect’ of sex (2=female, 1=male) on cholesterol level could also be demonstrated by applying a univariate linear regression model:

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	209,698	5,525		37,952	,000
	Sex	2,802	3,457	,090	,811	,420

a. Dependent Variable: Cholesterol level

Both analyses (comparison of means and linear regression) yield the same result: mean cholesterol level in females is about 2.8 units higher than mean cholesterol level in males. The difference is not significant, as revealed by a t-test (or a univariate regression model) with a p-value of 0.42.

If adjusted by body weight (which is on average higher in males), we obtain the following regression model:

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	175,729	12,749		13,784	,000
	Weight (kg)	,378	,129	,339	2,928	,004
	Sex	7,132	3,622	,228	1,969	,052

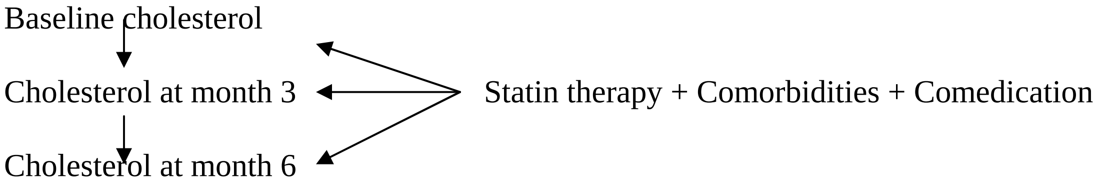
a. Dependent Variable: Cholesterol level

Now, the effect of sex on cholesterol is much more pronounced (comparing equal-weighted males and females, the difference is 7.132) and marginally significant (P=0.052).

Sometimes multivariable models are falsely denoted as ‘multivariate’ models. However, one should be careful not to confuse these two concepts.

Multivariate models

A multivariate model is a model with several outcome variables explained by the same set of independent variables. As an example, consider a study in which two different statin products are compared in their ability to decrease cholesterol levels. Patient’s cholesterol levels are repeatedly assessed, beginning with a baseline examination before start of treatment, and examinations after three and six months of statin therapy. A simultaneous evaluation of all these cholesterol measurements makes sense because the repeated cholesterol levels will be correlated within a patient, and this correlation should be taken into account.



Purposes of multivariable models

The two main purposes of multivariable models are

- Defining a prediction rule of the outcome
- Adjusting effects for confounders

The typical situation for the first purpose is a set of candidate variables, from which some will enter the final (best explaining) model. There are several strategies to identify such a subset of variables:

- Option 1: variable selection based on significance in univariate models: all variables that show a significant effect in univariate models are included. Usually the significance level is set to 0.15-0.25.
    - Pros: evaluates whether significant unadjusted associations with the outcome remain if adjusted for other important effects
    - Cons: sometimes variables are only significant if adjusted for other effects (example: see above), such relationships would be missed

- Option 2: variable selection based on significance in multivariable model: starting with a multivariable model including all candidate variables, one eliminates non-significant effects one-by-one until all effects in the model are significant. Variants of this method allow re-entering of variables at later steps or start with an empty model and subsequently include variables one-by-one (backward/stepwise/forward selection)
    - Pros: automated procedure, can be independently reproduced
    - Cons: the results obtained can be very unstable and must be assumed to be biased. Careful validation should follow such an analysis, resampling techniques (the bootstrap or permutation) can shed some light on the inherent but obscured variability; these validation algorithms are – unfortunately – not readily available in standard software

- Option 3: the ‘PurposefulSelection’ algorithm (Z. Bursac, C. H. Gauss, D. K. Williams, and D. Hosmer: *A Purposeful Selection of Variables Macro for Logistic*

Regression. SAS Global Forum 2007, Paper 173-2007, http://www2.sas.com/proceedings/forum2007/TOC.html). This variable selection procedure selects variables not only based on their significance in a multivariable model, but also if their omission from the multivariable model would cause the regression coefficients of other variables in the model to change by more than, say, 20%. The algorithm needs several cycles until it converges to a final model, where all variables that are contained satisfy both conditions. The algorithm could be executed by hand, but with many variables a computer program is needed.
  - Pro: automated procedure, can be independently reproduced
  - Pro: very useful if the purpose of the model is to adjust the effect of some variable for potential confounders; one can be sure that the algorithm does not miss any important confounders (among those which are presented to the algorithm)
  - Cons: needs more ‘tuning’ parameters than options 1 or 2 (but ‘default’ settings perform quite satisfactory)
  - Cons: although the results are assumed to be less biased than those of options 1 and 2, it is not yet sure whether there is residual bias

- Option 4: variable selection based on substance matter knowledge: this is the best way to select variables, as it is not data-driven and it is therefore considered as yielding unbiased results.
  - Pros: no bias
  - Cons: not automated, needs some thinking, hard to justify that selection was really made without looking at the data

The optimal choice of variable selection method has always been a matter of debate. The first option should be avoided if possible. The second option should only be used in conjunction with careful validation using resampling techniques. Among all ‘automatic’ selection procedures, the third one is currently state-of-the-art and should be applied. It however needs specialized software (there is one implementation in SAS but not in SPSS) . The fourth option is generally preferred by statisticians (passing the buck to their clinical partners).

A worked example

Consider cholesterol as outcome variable. The candidate predictors are: sex, age, BMI, WHR (waist-hip-ratio), and sports (although this variable is ordinal, we treat it as a scale variable here for simplicity).

Variable	Univariate B (P)	Model 1*: B (P)	Model 2** B (P)	Model 3*** B (P)
Sex	2.802 (0.420)			4.201 (0.148)
Age	0.766 (<0.001)	0.736 (<0.001)	0.756 (<0.001)	0.721 (<0.001)
BMI	1.257 (0.006)	0.962 (0.032)	1.179 (0.001)	0.952 (0.015)
WHR	31.87 (0.007)	4.487 (0.656)		14.13 (0.226)
Sports	-7.017 (0.002)	-1.125 (0.596)		
(Constant)		156.1 (<0.001)	153.1 (<0.001)	
R-squared		51.5%	51.1%	52.6%

* Selection based on univariate P<0.25
** Selection based on multivariable P<0.05
*** Selection based on multivariable P<0.1 and change in B of other variables of more than 20%

While model 2 can be easily calculated by SPSS, model 1 needs hand selection after all univariate models have been estimated and model 3 needs many side calculations.

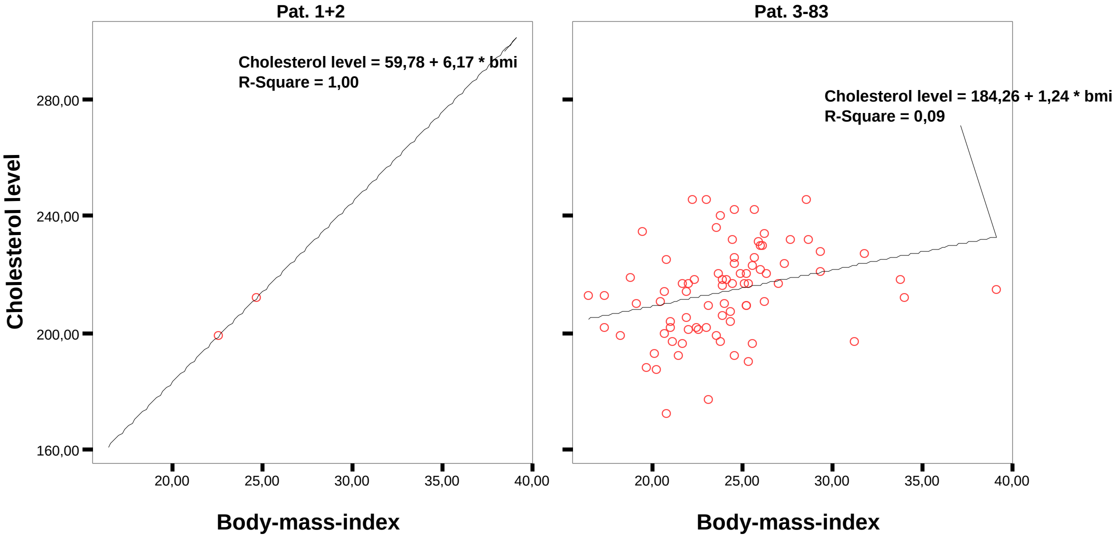
Model 3 selected Sex, Age, BMI and WHR as predictors of cholesterol. Age and BMI were selected based on their significance (P<0.1) in the multivariable model. On the other hand, Sex was selected because dropping it from the model would cause the B of WHR to change by -63%. Similarly, dropping WHR from the model would imply a change in B of Sex by -44%. Therefore, both variables were left in the model. Dropping sports from the model including all 5 variables will cause a change in B of BMI of +17%, and has less impact on the other variables. Since sports was not significant (P=0.54) and the maximum change in B was 17% (less than the pre-specified 20%), it was eliminated.

There are some typical situations (among others) in which multivariable modeling is used to adjust an effect for confounders:

- if a new candidate variable (e. g., a new biomarker) should be established as predictor of the outcome (e. g., survival after diagnosis of cancer), independent of known predictors (e. g., tumor stage, nodal status etc.)
- if in an observational study one wants to separate the effects of two variables which are correlated (e. g., type of medication and comorbidities)
- to asses the treatment effect in a randomized trial

How many independent variables can be included in a multivariable model? There are some guidelines addressing this issue. First of all it should be discussed why it is important to restrict the number of candidate variables. In the extreme case, the number of variables equals the number of subjects. In this situation the results cannot be generalized, as they only reflect the sample at hand.

As an example, consider a regression line which is based on two observations, compared to a regression line based on all other patients:



The red line is a linear regression line based on data from the first two patients only. Although the fit for these two patients is perfect, as confirmed by an R-Square of 1 (=100%), it is not transferable to the other patients. A regression line computed from patients 3-83 yields substantially different results, with an R-Square of only 9%. Typically, the results based on a small sample show a more extreme relationship than would be obtained in a larger sample. Such results are termed ‘overfit’.

In general, using too many variables with too few independent subjects *tends* to over-estimate relationships (as shown in the example above), the results are unstable (i. e., they change greatly by leaving out one subject or one variable from the model). As a rule of thumb, there should be at least 10 subjects for each variable in the model (or for each candidate variable when automated variable selection is applied). In logistic regression models, this rule is further tightened: if there are n events and m non-events, then the number of subjects should exceed 10min(n, m). In Cox regression models for survival data, the 10-subjects-rule applies to the number of deaths.

Confounding

Univariate models describe the crude relationship between a variable (let’s call it the *exposure* for the time being; it could also be the *treatment* in a randomized trial) and an outcome. Often the crude relationship may not only reflect the effect of the exposure, but may also reflect the effect of an extraneous factor, a confounder, which is associated with the exposure. A confounder is an extraneous factor that is:

- associated with the exposure in the source population
- a determinant of the outcome, independent of the exposure, and
- not part of the causal pathway from the exposure to the outcome

This implies, that the crude measure of effect reflects a mixture of the effect of the exposure and the effect of confounding factors. When confounding exists, analytical methods must be used to separate the effect of the exposure from the effects of the confounding factor(s). Multivariable modeling is one way to control confounding (another way would be stratification, which is not considered here).

Confounding is not much of an issue in randomized trials, as the randomization procedure automatically makes the treatment group allocation independent from any other factor that may be related to the outcome. However, it has been proposed to include important factors into multivariable modeling to reduce the variability of the outcome.

However, in observational studies addressing the issue of confounding is a *must*. As an example, consider the relationship between type of hypertension medication (e. g., betablockers vs. angiotensin converting enzyme inhibitors) and the outcome after kidney transplantation in an observational study. If patients had not been randomized to receive either betablocker or ACEI, it is not possible to conclude which of the two types of treatment is better without considering confounders (e. g., heart or vascular diseases), because patients with more favorable baseline characteristics may have been more likely to receive one of the two medications than to receive the other.

Effect modification

Effect modification means that the size of the effect of a variable depends on the level of another variable. Presence of effect modification can be assessed by adding interaction terms to a model:

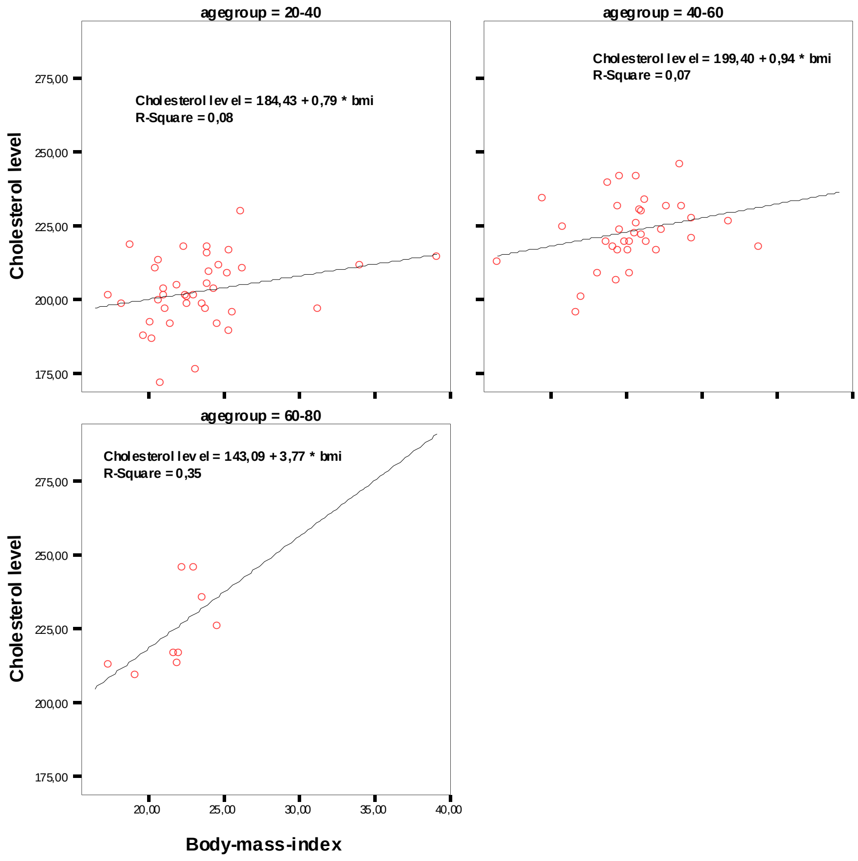
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	218.519	27.718		7.884	.000
	Age	-.805	.636	-.691	-1.266	.209
	Body-mass-index	-1.712	1.208	-.411	-1.417	.160
	bmiage	.069	.028	1.538	2.478	.015

a. Dependent Variable: Cholesterol level

Here, Bmiage = Age*BMI. The effect of body mass index on cholesterol is modified by age; we have different effects of BMI on cholesterol at different ages.

Significant and relevant effect modification indicates the use of subgroup analyses (separate models for patients divided into groups defined by the effect modifier). In our example, we would divide the patients into young, middle-aged and old subjects and present separate (univariate) regression models explaining cholesterol by BMI.



Usually, we retain the assumption of no effect modification unless we proof the opposite. Here, a significant effect modification is present, as indicated by a p-value of 0.015.

Assumptions of various models

Various assumptions underlie statistical models. Some of them are common to all models, some are specific to linear or Cox regression.

Common assumptions of all models

- Effects of independent variables sum up (additivity):
  All models that will be used in our course are of a linear structure. That is, the kernel of the model equation is always a linear combination of regression coefficients and independent variables, e. g. Cholesterol=b0+b1*age+b2*BMI. This structure implies that the effect of age and BMI sum up, but do not multiply. The additivity principle can be relaxed by including interaction terms into the model equation, or by taking the log of the outcome variable: recall that additivity on the log scale is equivalent to multiplicity on the original scale.

- No interactions (no effect modification):
  The assumption of no effect modification is usually retained unless the opposite can be proven; there is no use in establishing a complex model if a simpler model fits the data equally well.

Common assumptions of models involving scale variables

- Linearity:
  Consider the regression equation Cholesterol=b0+b1*age+b2*BMI. Both independent variables age and BMI have by default a linear effect on cholesterol: comparing two patients of age 30 and 31 leads to the same difference in cholesterol as a comparison of two patients aged 60 and 61. The linearity assumption can be relaxed by including quadratic and cubic terms for scale variables, as was demonstrated in the basic course.

Assumptions of linear models

Model-specific assumptions concern the distribution of the residuals, i. e. the distance between the predicted and the observed values of the outcome variable. These assumptions are:

- Residuals are normally distributed
  This can easily be checked by a histogram of residuals.

- Residuals have a constant variance. A plot of residuals against predicted values should not show any increase or decrease of the spread of the residuals.
- Residuals are uncorrelated to each other
  This assumption could be violated if subjects were not sampled independently, but were recruited in clusters. If the assumption of independence is violated, we must account for the clustering by including so-called random effects into the model. A random effect (as the opposite of a fixed effect) is not of interest per se, it rather serves to adjust for the dependency of observations within a cluster.
- Residuals are uncorrelated to independent variables
  If a scatter plot of residuals versus an independent variable shows some systematic dependency, it could be a consequence of a violation of the linearity assumption, or it might also indicate a misspecification, e. g., the constant has been omitted.

Assumptions of Cox regression models

- Proportional hazards assumption:
  As will be demonstrated later, Cox regression assumes that although the risk to die may vary over time, the risk ratio between two groups of patients is constant over the whole range of follow-up. This is a meaningful assumption which holds in the majority of data sets. Including interactions of covariates with follow-up time, thus generating a time-dependent effect, is one way to relax the proportional hazards assumption.

As the validity of the models results is crucially depending on the validity of model assumptions, estimation of statistical models should always be followed by a careful investigation of the model assumptions.

Part 2: Analysis of binary outcomes

Class 2: Diagnostic studies

Assessment of diagnostic tests

Example: Mine workers and pneumoconiosis (Campbell and Machin [1]).
Consider a sample of mine workers, whose forced expiratory volume 1 (FEV-1) values and pneumoconiosis status (present/absent) were measured. FEV-1 values are given as percent of reference values. Pneumoconiosis was diagnosed by clinical evaluation.

	Pneumoconiosis		
	Present	Absent	Total
FEV1<80%	22	6	28
FEV1>80%	5	7	12
Total	27	13	40

	Disease		
Test result	Present	Absent	Total
positive	True positives (TP)	False positives (FP)	TP+FP
negative	False negatives (FN)	True negatives (TN)	FN+TN
Total	TP+FN	FP+TN	TP+FP+FN+TN

Assume that FEV-1 should be used to assess pneumoconiosis status. In order to quantify its ability to detect pneumoconiosis, the following diagnostic measures are useful:

- Sensitivity (Se): the probability of a positive test given the disease is present. Se=TP/(TP+FN)
- Specificity (Sp): the probability of a negative test given the diseases is absent. Sp=TN/(FP+TN)
- Accuracy (Ac): the probability of a correct test result. Ac=(TP+TN)/(TP+FP+FN+TN)

In our example, these values calculate as follows:

Sensitivity:	81.5%
Specificity:	53.8%

Accuracy: 72.5%

An ideal test exhibits a sensitivity and a specificity both close to 100%.

From our sample of mine workers, we estimate a pretest probability of the disease as 27/40=67.5%. Now assume that a mine worker's FEV-1 is measured, and it falls below 80% of the reference value. How does this test result affect our pretest probability? We can quantify the posttest probability (positive predictive value) as 78.6%. Generally, it is defined as

- Posttest probability of presence of the disease (Positive predictive value, PPV): Probability of the disease given the test result is positive. PPV=TP/(TP+FP)

The ability of a positive test result to change our prior (pretest) assessment is quantified by the positive likelihood ratio (PLR). It is defined as the ratio of posttest odds and pretest odds. Odds are another way to express probabilities. Generally, the odds of an event are given by

- Odds = Probability of event/(1 – Probability of event)

Therefore, pretest odds are calculated as

- Pretest odds: Pretest probability/(1 - Pretest probability)

Similarly, posttest odds are given by

- Posttest odds: Posttest probability/(1 – Posttest probability) = PPV / (1 – PPV)

The positive likelihood ratio (PLR) can then be calculated as

- PLR = Posttest odds / Pretest odds

Some simple calculus results in

- PLR = Se / (1 – Sp)

In our example, the positive likelihood ratio is thus 0.815/(1 – 0.538) = 1.764. This means that a positive test result increases the odds for presence of disease by the 1.764-fold.

What's the advantage of PLR?

Since Se and Sp are conditional probabilities, conditioning on presence or absence of disease, these numbers are independent of the prevalence of the disease in a given population. By contrast, the positive and negative predictive values are conditional probabilities, conditioning on positive or negative test results, respectively. We would obtain different values for PPV or NPV in populations that exhibit different pretest disease probabilities, as can be exemplified:

Assume, we investigate FEV-1 in workers of a different mine, and obtain the following sample:

	Pneumoconiosis		
	Present	Absent	Total
FEV1<80%	22	60	82
FEV1>80%	5	70	75
Total	27	130	157

The key measures characterizing the performance of the diagnostic test calculate as follows:

Sensitivity:	22/27 = 81.5%	*unchanged*
Specificity:	70/130 = 53.8%	*unchanged*
Pretest probability:	27/157 = 17.2%	*changed (much lower)*
Posttest probability:	22/82 = 26.8%	*changed (lower)*
PLR:	0.815/(1 – 0.538) = 1.764	*unchanged!*

We see that the positive likelihood ratio is unchanged. It is independent of the pretest probability (or prevalence). In other words, a positive test result still increases the probability of presence of disease by the 1.764-fold. However, since we start at a pretest probability of 17.2%, this increase results in a lower value for the posttest probability than before.

Similarly, we have

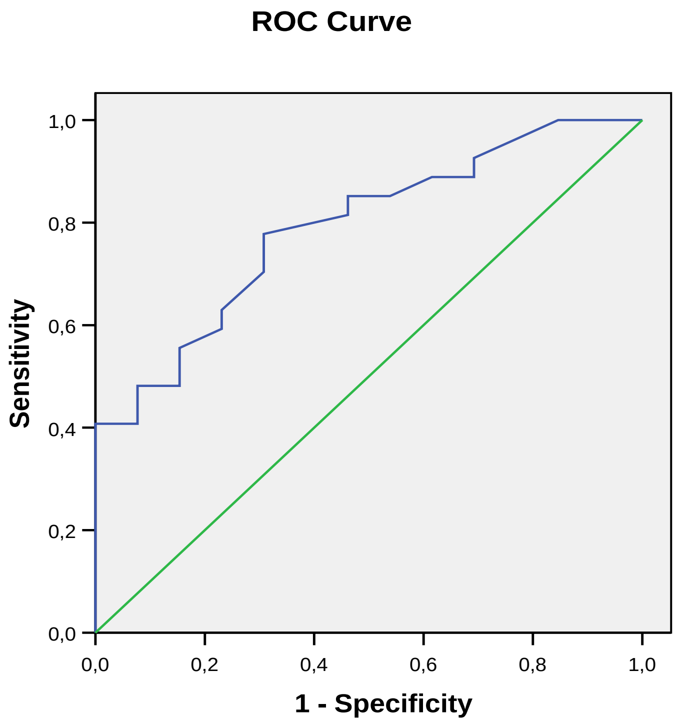
- Posttest probability of absence of the disease (Negative predictive value, NPV): Probability of absence of disease given the test result is negative. NPV=TN/(TN+FN)
- Negative likelihood ratio: Sp / (1 – Se)

expressing the increase of the probability of *absence* of disease caused by a negative test result.

Receiver-operating-characteristic (ROC) curves

In the example given above, we chose a cut-off value of 80% of reference as defining a positive or negative test result. Selecting different cut-off values would change sensitivity and specificity of the diagnostic test. Sensitivity and specificity resulting from various cut-off values can be plotted in a so-called receiver operating characteristic (ROC) curve.

We see that generally, there is a trade-off between sensitivity and specificity: the higher the cut-off value, the higher the sensitivity (TP rate), but the lower the specificity (TN rate), as more healthy workers are classified as diseased.



Diagonal segments are produced by ties.

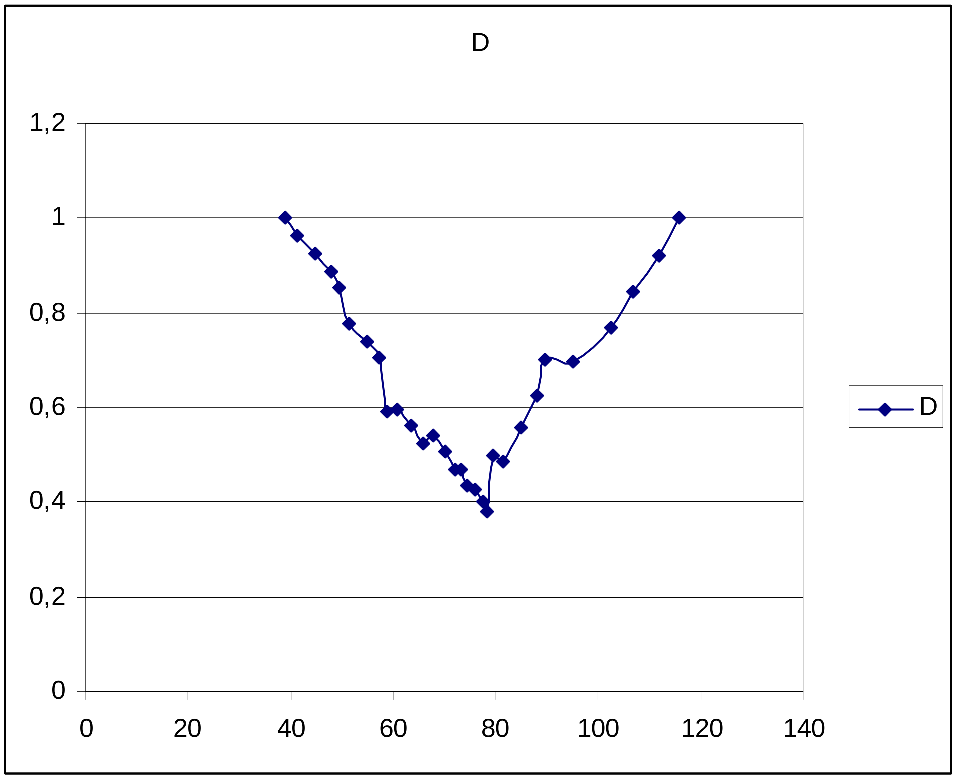
Note that on the x-axis, by convention 1-Specificity is plotted. A global criterion for a test is the area under the ROC curve, often denoted as the c-index. Generally, this value falls into the range 0 to 1. It can be interpreted as the probability that a randomly chosen diseased worker has a lower FEV-1 value than a randomly chosen healthy worker. Clearly, if the c-index is 0.5, it means that the healthy or the diseased worker may have a higher FEV-1 value, or, put another way, that the test is meaningless. This is expressed by the diagonal line in the ROC curve: the area under this line is exactly 0.5, and if the ROC curve of a test more or less follows the diagonal, such a test would be meaningless in detecting the disease. A common threshold value for the c-index to denote a test as "useful" is 0.8. Our FEV-1 test has a c-index of 0.789, which is marginally below the threshold value of 0.8. Because it is based on a very small sample, it is useful to state a 95% confidence interval for the index, which is given by [0.647, 0.931]. Since 0.5 is outside this interval, we can prove some correlation of the test with presence of disease. However, our data is compatible with c-indices ranging from 0.65 on, meaning that we cannot really prove the usefulness of the test to detect pneumoconiosis in mine workers.

How should a cut-off value be chosen for a quantitative test?

A simple approach would take that value that maximizes the sum of Se and Sp. A more elaborated way to obtain a cut-off value is to take that value that minimizes the distance between the ROC-curve and the upper left corner of the panel (the point where Se and Sp assume 100%). This (Euclidian) distance can be calculated as

$D = \text{sqrt}((1\text{-Se})^2 + (1\text{-Sp})^2)$

A graph plotting D against various cut-off values can be used to validate the identified cut off level:



Here we see that the “best” cut-off value is indeed 80. The inverse peak at a cut-off value of 80 underlines the uniqueness of that value.

Both approaches outlined above put the same weight on a high sensitivity and a high specificity. However, sometimes it is more useful to attain a certain minimum level of sensitivity, because it may be more harmful or costly to overlook presence of disease than to falsely diagnose the disease in a healthy person. In such cases, one would consider only such values as cut-points where the sensitivity is at least 95% (or 99%), and select that value that maximizes the specificity.

ROC curves can also be used to compare diagnostic markers. A test A is preferable over a test B, it the ROC curve of A is always above the ROC curve of B.

Class 3: Risk measures

Absolute measures to compare the risk in two groups

The following example [18] is a prospective study, which compares the incidences of dyskinesia after ropinirole (ROP) or levodopa (LD) in patients with early Parkinson's disease. The results show that 17 of 179 patients who took ropinirole and 23 of 89 who took levodopa developed dyskinesia. The data are summarized in the following table:

	Presence of dyskinesia		
	Yes	No	Total
Group			
Levodopa	23	66	89
Ropinirole	17	162	179
Totals	40	228	268

The risk of having dyskinesia among patients who took LD is 23/89 = 0.258, whereas the risk of developing dyskinesia among patients who took ROP is 17/179 = 0.095. Therefore, the absolute risk reduction is ARR=0.258-0.095=0.163. Since ARR is a point estimate, it is desirable to have an interval estimate as well which reflects the uncertainty in the point estimate due to limited sample size. A 95% confidence interval can be obtained by a simple normal approximation by first computing the variance of ARR. The standard error of ARR is then simply the square root of the variance. Adding +/-1.96 times the standard error to the ARR point estimate yields a 95% confidence interval. To compute the variance of the ARR, let's first consider variances for the risk estimates in both groups. These calculate as risk(1-risk)/N.
Summarizing, we have

Risk of dyskinesia in LD:	23/89 = 0.258
Standard error (square root of variance):	sqrt(0.258(1-0.258)/89) = 0.04638
Risk of dyskinesia in ROP:	17/179 = 0.095
Standard error:	sqrt(0.095(1-0.095)/179) = 0.02192

Absolute risk reduction (ARR):	0.258 – 0.095 = 0.163
Standard error of ARR:	Sqrt(0.258(1-0.258)/89 + 0.095(1-0.095)/179) = 0.0513
95% Confidence interval for ARR:	0.163 +/- 1.96 * 0.0513 = [0.062, 0.264]

A number related to the ARR is the number needed to treat (NNT). It is defined as the reciprocal of ARR, thus we have

Number needed to treat: NNT = 1 / ARR

The NNT is interpreted as the number of patients who must be treated in order to expect one healed patient. The larger the NNT, the more useless is the treatment.

A 95% confidence interval for NNT can be obtained by taking the reciprocal of the confidence interval of ARR. In our example, we have

NNT = 1/ 0.163 = 6.13
95% Confidence interval: [1/0.264, 1/0.062] = [3.8, 15.9]

Note: *if ARR is close to 0, the confidence interval for NNT such obtained may not include the point estimate. This is due to the singularity of NNT in case of ARR=0: in this situation NNT is actually infinite. For illustration, consider an example where ARR (95% C.I.) is 0.1 (-0.05, 0.25). The NNT (95% C.I.) would be calculated as 10 (-20, 4). The confidence interval does not contain the point estimate. However, this confidence interval is not correctly calculated. In case that the confidence interval of ARR covers the value 0, the confidence interval of NNT must be redefined as (-20 to -∞, 4 to ∞). Thus it contains all values between -20 and -∞, and at the same time all values between 4 and infinity.*

This can be proven empirically by computing the NNT for some ARR values inside the confidence interval, say for -0.03, -0.01, +0.05 and +0.15; we would end up in NNT values of -33, -10, +20 and +6.7, which are all inside the redefined interval but not in the original interval.

Considering the NNT at an ARR of 0, we would have to treat an infinite number of patients in order to observe one successfully treated patient.

ARR is an absolute measure to compare the risk between two groups. Thus it reflects the underlying risk without treatment (or with standard treatment) and has a clear interpretation for the practitioner.

Relative measures to compare the risk between two groups

The next two popular measures are the relative risk (RR) and the relative risk reduction (RRR). The relative risk is the ratio of risks of the treated group and the control group, and also called the risk ratio. The relative risk reduction is derived from the relative risk by subtracting it from one, which is the same as the ratio between the ARR and the risk in the control group. A 95% confidence intervals for RR can be obtained by first calculating the standard error of the log of RR, then computing a confidence interval for log(RR), and then taking the antilog to obtain a confidence interval of RR. In our example, the RR and the RRR calculate as follows:

Relative risk: RR = 0.095 / 0.258 = 0.368
Relative risk reduction: RRR = 1 – 0.368 = 0.632

These numbers are interpreted as follows: the risk of developing dyskinesia after treatment by ROP is only 0.368 times the risk of developing dyskinesia after treatment by LD. This means, the risk of developing dyskinesia is reduced by 63.2% if treatment ROP is applied.

One disadvantage of RR is that its value can be the same for very different clinical situations. For example, a RR of 0.167 would be the outcome for both of the following clinical situations: 1) when the risks for the treated and control groups are 0.3 and 0.05, respectively; and for 2) a risk of 0.84 for the treated group and of 0.14 for the control group. RR is clear on a proportional scale, but has no real meaning on an absolute scale. Therefore, it is generally more meaningful to use relative effect measures for summarizing the evidence and absolute measures for application to a concrete clinical or public health situation [2].

The odds ratio (OR) is a commonly used measure of the size of an effect and may be reported in case control studies, cohort studies, or clinical trials. It can also be used in retrospective studies and cross-sectional studies, where the goal is to look at associations rather than differences.

The odds can be interpreted as the number of events relative to the number of nonevents. The odds ratio is the ratio between the odds of the treated group and the odds of the control group.

Both odds and odds ratios are dimensionless. An odds ratio less than 1 means that the odds have decreased, and similarly, an OR greater than 1 means that the odds have increased.

It should be noted that ORs are hard to comprehend [3] and are frequently interpreted as an approximate relative risk. Although the odds ratio is close to the relative risk when the outcome is relatively uncommon [2] as assumed in case-control studies, there is a recognized problem that odds ratios do not give a good approximation of the relative risk when the control group risk is “high”. Furthermore, an odds ratio will always exaggerate the size of the effect compared to a relative risk. When the OR is less than 1, it is smaller than the RR, and when it is greater than 1, the OR exceeds the RR. However, the interpretation will not, generally, be influenced by this discrepancy, because the discrepancy is large only for large positive or negative effect size, in which case the qualitative conclusion will remain unchanged. The odds ratio is the only valid measure of association regardless of whether the study design is follow-up, case-control, or cross sectional. Risks or relative risks can be estimated only in follow-up designs.

The great advantage of odds ratios is that they are the result of logistic regression, which allows adjusting effects for imbalances in important covariates. As an example, assume that patients in the LD groups were on average older than those in the ROP group. In such a case it would be difficult to judge from the crude (unadjusted) relative risk estimate whether the advantage of ROP is just due to the age imbalance or really an effect of treatment. Therefore, even if the underlying risk is not low, the OR is used to describe an effect size which is adjusted for imbalance in other covariates.

Summary: calculation of risk measures and 95% confidence intervals

Consider the general case where we have a table of the following structure:

	Disease		
	Present	absent	Total
Group			
Control	A	B	A+B
Treated	C	D	C+D
Totals	A+C	B+D	N=A+B+C+D

The following describes the calculation of the measures and the associated 95% confidence intervals:

Measure	Use in	Type of estimate	Calculation
Risk in control group	Cohort study	Point estimate	RC=A/(A+B)

		Standard error (SE)	VC=RC*(1-RC)/(A+B)
			SE=sqrt(VC)

		95% confidence interval	RC +/- 1.96*SE
Risk in treated group	Cohort study	Point estimate	RT=C/(C+D)

		Standard error (SE)	VT=RT*(1-RT)/(C+D)
			SE=sqrt(VT)

		95% confidence interval	R +/- 1.96*SE
Absolute risk reduction (*useful only if RT<RC!*)	Cohort study	Point estimate	ARR=A/(A+B)-C/(C+D)
		Standard error (SE)	SE=Sqrt(VC + VT)
		95% confidence interval	ARR +/- 1.96*SE
Number needed to treat (*if RT<RC*)	Cohort study	Point estimate	1/ARR
		95% confidence	1/(ARR+1.96*SE), 1/(ARR-

		interval	1.96*SE) If point estimate is not contained in the interval, the interval is redefined as -∞ to 1/(ARR+1.96*SE) and 1/(ARR-1.96*SE) to ∞

Relative risk	Cohort study	Point estimate
		95% confidence interval
			RR = RT/RC
			logRR = log(RR)
			V=1/A-1/(A+B)+1/C-1/(C+D)
			SE=sqrt(V)
			logL=logRR-1.96*SE
			logU=logRR+1.96*SE
			
			95% Confidence interval:
			[exp(logL), exp(logU)]

Relative risk reduction (*makes only sense if RR<1!*)	Cohort study	Point estimate	RRR = 1-RR
		95% confidence interval	[1-exp(logU), 1-exp(logL)]
Odds ratio	Case-control study	Point estimate	OR = RT*(1-RC)/(1-RT)/RC

		95% confidence interval	logOR = log(OR)
			V=1/A + 1/B + 1/C + 1/D
			SE=sqrt(V)
			logL=logOR-1.96*SE
			logU=logOR+1.96*SE
			
			95% Confidence interval:
			[exp(logL), exp(logU)]

Estimation of all the risk measures presented in this section and computation of 95% confidence intervals is facilitated by the Excel application “RiskEstimates.xls” which is available at the author’s homepage

http://www.meduniwien.ac.at/user/georg.heinze/RiskEstimates.xls

Class 4: Logistic regression

Simple logistic regression

A possibility to extend the analysis of studies of a binary outcome to more than one explaining variable, is analysis by logistic regression. This method is an analogue of linear regression for binary outcomes. The regression equation estimated by logistic regression is given by:

Pr(Y=1) = 1 / (exp(-b0 – b1X))

where X and Y denote the independent and binary dependent variables, respectively. This equation describes the association of X and the probability that Y assumes the value 1.

The regression equation may be transformed into:

Log(Pr(Y=1)/Pr(Y=0)) = b0 + b1X

which is linear on the right hand side, and has the so-called logistic function on the left hand side (hence the name "logistic regression").

The expression

Pr(Y=1)/Pr(Y=0)

is equal to the odds of Y=1, such that we are actually modeling the log odds by a linear model. Thus, the regression coefficient b1 has the meaning:

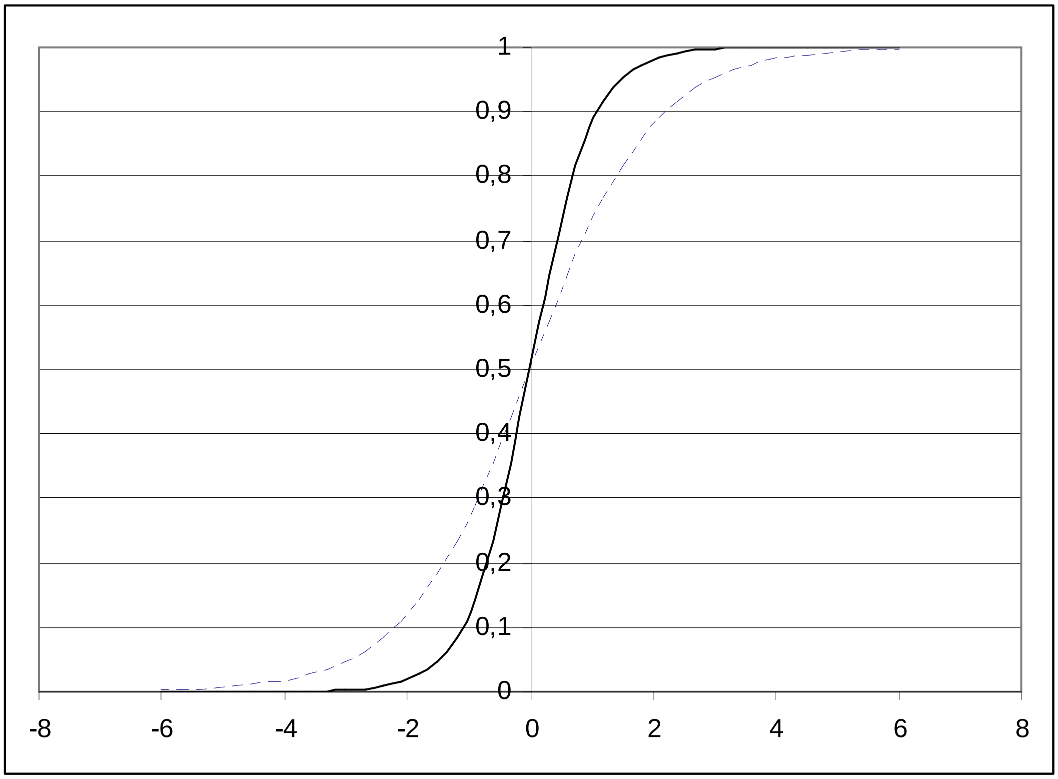
If X changes by 1, then the log odds change by b1.

This is equivalent to:

If X changes by 1, then the odds change by exp(b1).

Since a change in odds is called an odds ratio, we can directly compute odds ratios from the regression coefficients which are given in the output of any statistical software package for logistic regression. These odds ratio refer to a comparison of two subjects differing in X by one unit.

For b0=0 and b1=1 (dashed line) or b1=2 (solid line), the logistic equation yields:



The higher the value of b1, the steeper is the slope of the curve. In the extreme case of b1=0, the curve will be a flat line. Values of b0 different from 0 will shift the curve to the left (for positive b0) or to the right (for negative b0). Negative values of b1 will mirror the curve, it will fall from the upper left corner to the lower right corner of the panel.

By estimating the curve parameters b0 and b1, we can quantify the association of a an independent variable X with a binary outcome variable Y. The regression parameter b1 has a very intuitive meaning: it is simple the log of the odds ratio associated with a one-unit increase of X. Put another way, exp(b1) is the factor by which the odds for an event (Y=1) change if X is increased by 1.

Now assume that X is not a scale variable, but a dichotomous factor itself. It could be an indicator of treatment, for instance X=1 defines the new treatment, and X=0 the standard therapy. Of course, the curve will now reduce to two points, i. e. the probability of an event in group X=1 and the probability of an event in group X=0. Estimating these two probabilities by means of logistic regression will exactly yield the relative frequencies of events in these two points. So, logistic regression can be used for analysis of a two-by-two table, yielding relative frequencies and an odds ratio, but it can also be extended to scale variables, and one can even mix both in one model.

Examples

The following two examples are based on the same study, where the aim was to identify risk factors for low birth weight (lower than 2500 grams) [5]. 189 newborn babies were included in the study, 69 of them had low birth weight.

Simple logistic regression for a scale variable

Let's first consider age of the mother as independent variable, a scale variable. For convenience, the age of the mother is expressed as decade, such that odds ratio estimates refer to a 10-year change in age instead of a 1-year change.

The results of logistic regression analysis using SPSS is given by the following table:

Variables in the Equation

								95,0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	age_decade	-,512	,315	2,635	1	,105	,600	,323	1,112
	Constant	,385	,732	,276	1	,599	1,469		

a. Variable(s) entered on step 1: age_decade.

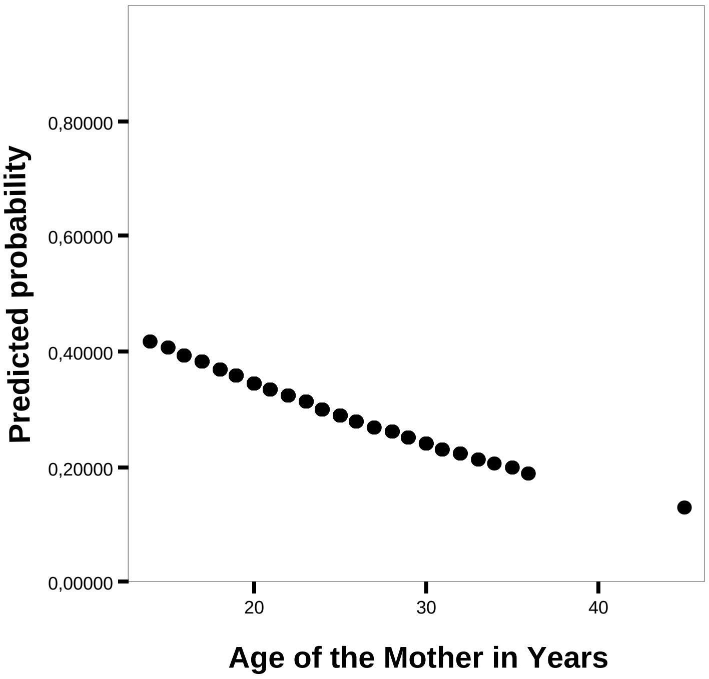
There are two 'variables' in the model: age_decade and Constant. The column labeled B contains the regression coefficient estimates. Thus, the regression equation reads as

Log (Odds of low birth weight) = b0 + b1 Age_Decade = 0.385 – 0.512 Age_Decade

We cannot learn much from this equation unless we take a look at the column Exp(B), which contains the odds ratio estimate for Age_Decade: 0.6, with a 95% confidence interval of [0.32, 1.11]. This means that the risk of low birth weight decreases to the 0.6fold with every decade of mother's age. Put another way, each decade reduces the risk for low birth weight by 40% (1-0.6, corresponding to the formula for relative risk reduction).

However, we see that the confidence interval contains the value 1 which would mean that mother's age has absolutely no influence on low birth weight. With our data, we cannot rule out that situation. A 95% confidence interval containing the null hypothesis value is always accompanied by an insignificant p-value; here it is 0.105, which is clearly above the commonly accepted significance level of 0.05.

Despite the non-significant result, let's have a look at the estimated regression curve:



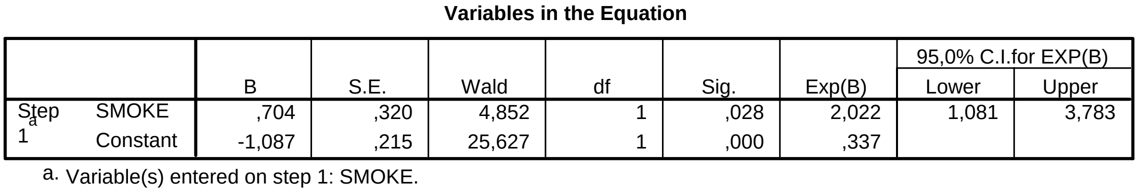
Simple logistic regression for a nominal (binary) variable

Now let's consider smoking as independent variable. A cross-tabulation of smoking and birth weight yields:

Smoking Status During Pregnancy (1/0) * Low Birth Weight (<2500g) Crosstabulation

			Low Birth Weight (<2500g)		
			0	1	Total
Smoking Status During Pregnancy (1/0)	0	Count	86	29	115
		% within Low Birth Weight (<2500g)	66,2%	49,2%	60,8%
	1	Count	44	30	74
		% within Low Birth Weight (<2500g)	33,8%	50,8%	39,2%
Total		Count	130	59	189
		% within Low Birth Weight (<2500g)	100,0%	100,0%	100,0%

We see that half of the mothers of low weight babies were smoking, but only one third of the mothers of normal weight babies. Analysis by logistic regression yields:



Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	SMOKE	,704	,320	4,852	1	,028	2,022	1,081	3,783
	Constant	-1,087	,215	25,627	1	,000	,337		

a. Variable(s) entered on step 1: SMOKE.

The odds ratio corresponding to smoking (95% confidence interval) is 2 (1.1, 3.8). Thus, smoking during pregnancy is a risk factor for low birth weight. (The very same result is obtained if the data of the contingency table given above was entered into RiskEstimates.xls)

Multiple logistic regression

Using multiple logistic regression, it is now possible to obtain not only crude effects of variables, but also adjusted effects. The following covariables are available:

Variable	Description	Range of values
AGE	Age of the mother	14-45
LWT	Weight in pounds at the last menstrual period	80-250
PTL	History of premature labor	0, 1, 2, 3
HT	History of hypertension	0, 1
SMOKE	Smoking status during pregnancy	0, 1

Let’s fit a multivariable logistic regression model. The analysis is done in four steps:

1. Check the number of events/nonevents and compare with number of variables
2. Fit the model
3. Interpret the model results
4. Check model assumptions

Ad 1: We have 59 events (cases of low birth weight), and 130 nonevents (cases of normal birth weight). The number of covariates is 5. Since 5<59/10, we are allowed to fit the model.

Ad 2: Fitting the model with SPSS, we obtain the following output:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	AGE	-,046	,034	1,754	1	,185	,955	,893	1,022
	LWT	-,015	,007	5,159	1	,023	,985	,972	,998
	SMOKE	,559	,340	2,692	1	,101	1,748	,897	3,408
	PTL	,690	,339	4,129	1	,042	1,993	1,025	3,876
	HT	1,771	,688	6,629	1	,010	5,877	1,526	22,632
	Constant	1,674	1,067	2,462	1	,117	5,333		

a. Variable(s) entered on step 1: AGE, LWT, SMOKE, PTL, HT.

Ad 3: This table contains the following data:

Column label	Contents
B	Estimated regression coefficients
S.E.	Their standard errors
Wald	Wald Chi-Squared, computed as $(B/SE)^2$
Df	Degrees of freedom
Sig.	Two-sided P-value for testing the hypothesis B=0
Exp(B)	Estimated odds ratio referring to a unit increase in the variable, computed as Exp(B)
Lower	Lower 95% confidence limit for odds ratio, computed as Exp(B-1.96S.E.)
Upper	Upper 95% confidence limit for odds ratio, computed as Exp(B+1.96S.E.)

Exercise: Try to figure out of that table which variables affect the outcome (low birth weight), and in which way they do!

The last line contains the estimate for the constant, which was denoted as b0 in the outline of simple logistic regression. The most important columns are the odds ratio estimates, the confidence limits and the P-value. We learn that last weight, history of premature labor and hypertension are independent risk factors for low birth weight.

SPSS outputs some other tables which are useful to interpret results:

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	23,344	5	,000
	Block	23,344	5	,000
	Model	23,344	5	,000

This table contains a test for the hypothesis that all regression coefficients related to covariates are zero (equivalent to: all odds ratios are one). SPSS performs the estimation in Steps and Blocks, which are only of relevance, if automated variable selection is applied (which is not the case here). The result of the test is “P<0.001” which means that the null hypothesis of no effect at all is implausible.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	211,328[a]	,116	,163

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

The model summary provides two Pseudo-R-Square measures, which yield quite the same result: about 11-16% of the variation in birth weight (depending on the way of calculation) can be explained by our five predictors.

Ad 4: checking assumptions of the model

First, let’s look at the regression equation, which can be extracted from the regression coefficients of the first output table:

Log odds(low birth weight) = 1.67 - 0.46 AGE - 0.015 LWT + 0.559 SMOKE + 0.69 PTL + 1.771 HT

The equation can be re-written as:

Pr(low birth weight) = 1 / (1 + exp(-1.67 + 0.46 AGE + 0.015 LWT - 0.559 SMOKE - 0.69 PTL - 1.771 HT))

Thus, we can predict the probability of low birth weight for each individual in the sample. These predictions can be used to assess the model fit, which is done by the Hosmer and Lemeshow Test:

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,862	8	,552

This test mainly tests the hypothesis that important predictors are still missing from the regression equation. In our case it is not significant, indicating adequacy of the model. How’s it done? The subjects of the sample are categorized in deciles corresponding to their predicted probabilities, and then the number of observed events (cases of low birth weight) in each decile is compared to the number expected from the predicted probabilities (i. e., the sum of predicted probabilities).

Contingency Table for Hosmer and Lemeshow Test

		Low Birth Weight (<2500g) = 0		Low Birth Weight (<2500g) = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	18	17,191	1	1,809	19
	2	15	16,047	4	2,953	19
	3	16	15,284	3	3,716	19
	4	16	14,637	3	4,363	19
	5	16	14,646	4	5,354	20
	6	10	13,169	9	5,831	19
	7	14	12,321	5	6,679	19
	8	10	11,457	9	7,543	19
	9	8	9,734	11	9,266	19
	10	7	5,512	10	11,488	17

If the expected and observed numbers differ by more than what can be expected from random variation, 'lack of fit' is still present, meaning that important predictors are missing from the model. Such effects could be

- Other variables explaining the outcome
- Nonlinear effects of continuous variables (e. g., of AGE or LWT)
- Interactions of variables (e. g., smoking in combination with hypertension could be worse than just the sum of the main effects of smoking and hypertension)

Another assessment of model fit is given by the classification table.

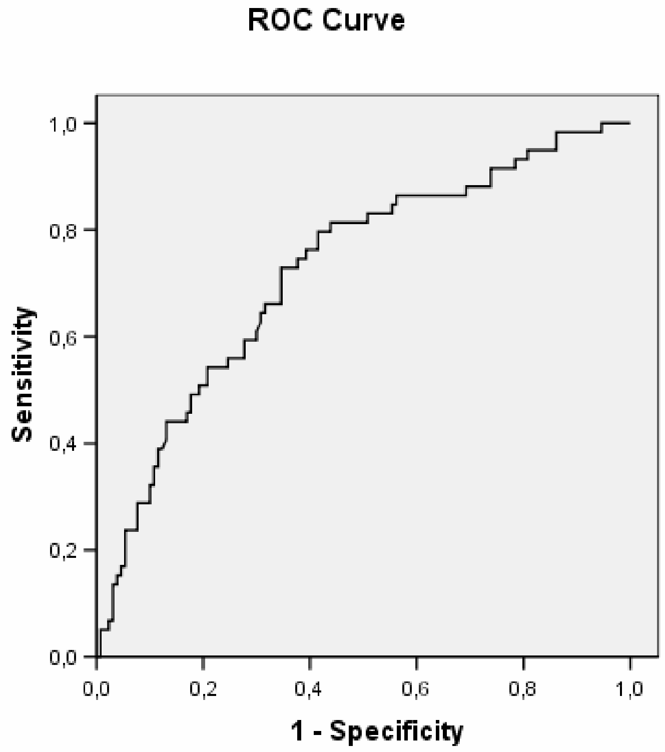
Classification Table[a]

			Predicted		
			Low Birth Weight (<2500g)		Percentage Correct
	Observed		0	1	
Step 1	Low Birth Weight (<2500g)	0	121	9	93,1
		1	45	14	23,7
	Overall Percentage				71,4

a. The cut value is ,500

Here, subjects are classified according to their predicted probability for low birth weight, with predicted probabilities above 0.5 defining the 'high risk group', for which we would predict low birth weight. We see that overall 71.4% can be classified correctly.

Another way to assess model fit is to use not only one cut value, but all possible values, constructing a ROC curve (with the predicted probabilities as 'test' variable, and the outcome as 'state' variable):



Diagonal segments are produced by ties.

From this ROC cure, the area under the curve (the so-called c-index) can be computed. In our example, it is 0.723. This number can again be interpreted as the probability of proversion (or probability of concordance):

Comparing two randomly chosen subjects with different outcome, then our model assigns a higher risk score (predicted probability) to the subject with unfavorable outcome with 72.3% probability.

Clearly, if the c-index is 0.5, the model cannot predict anything. By contrast, a c-index close to 1.0 indicates a perfect model fit.

We must carefully distinguish goodness-of-fit, which is expressed by the c-index, from proportion of explained variation, which is expressed by R-Square.

In case-control studies, the variation of the outcome is set by the design of the study; it is simply the proportion of cases among all subjects.

In cohort studies, the variation of the outcome reflects its prevalence in the study population.

Therefore, measures taking into regard the outcome variation, like R-square, are not suitable for case-control studies.

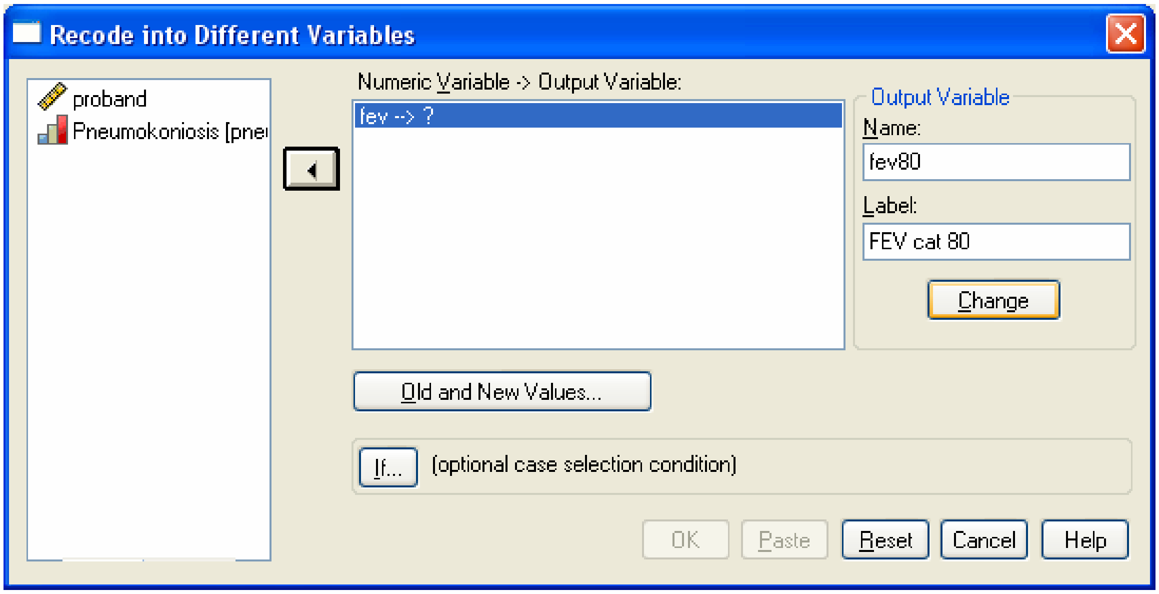
SPSS Lab 1: Analysis of binary outcomes

Define a cut-point

Open the data set pneumo.sav.

We want to define a cut-point of 80 for the variable FEV-1. Choose

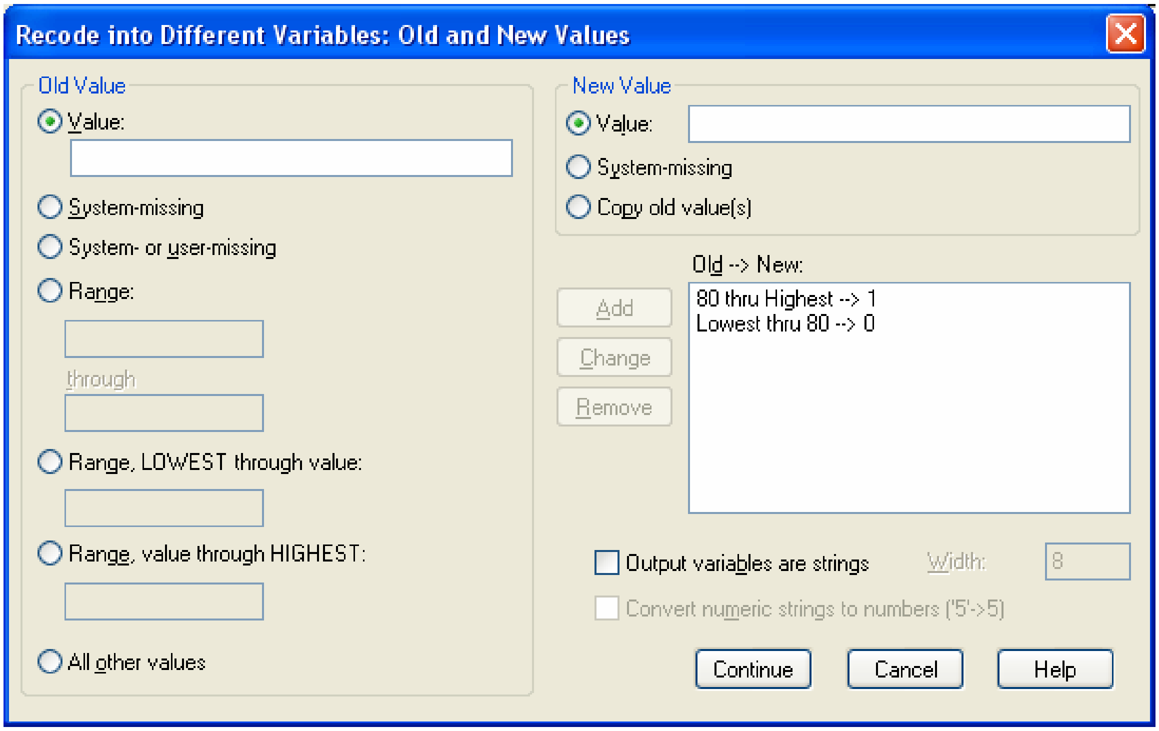
Transform-Recode-Into different variables…



Numeric Variable -> Output Variable:
fev --> ?

Output Variable
Name: fev80
Label: FEV cat 80

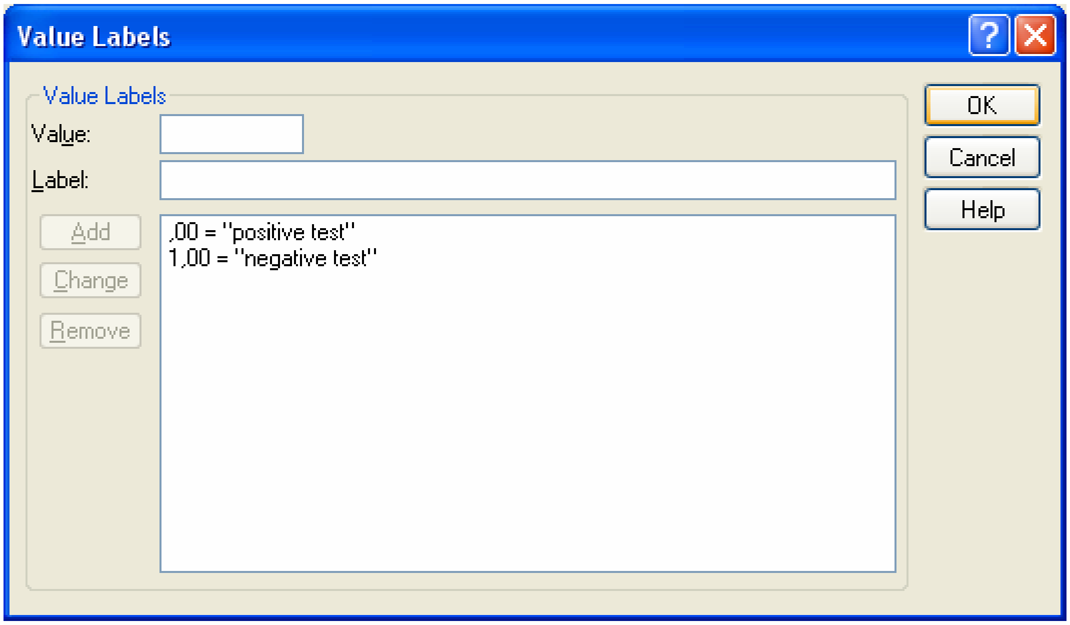
Choose fev as input variable. Define fev80 as output variable, labelled ‘FEV cat 80’ (or something similar). Press ‘Change’ to accept the new name. Then press ‘Old and New Values…’:


Recode into Different Variables: Old and New Values
Old Value
● Value:
○ System-missing
○ System- or user-missing
○ Range:
through
○ Range, LOWEST through value:
○ Range, value through HIGHEST:
○ All other values
New Value
● Value:
○ System-missing
○ Copy old value(s)
Old --> New:
Add
Change
Remove
□ Output variables are strings
Width:
8
□ Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help

80 thru Highest --> 1
Lowest thru 80 --> 0

Fill in the value '80' (without quotation marks) in the field 'Range, value through HIGHEST', and define 1 in the field 'New Value'. Press 'Add' to accept this choice. In the field 'Range, LOWEST through value:', fill in '80', and in the field 'New Value', define 0. Again, press 'Add' to confirm. Press 'Continue'. Back at the first dialogue, press 'OK'.
A new variable, 'FEV80' has been added to the data sheet. We learn that the value 80 was categorized as 1. This is controlled by the sequence we use to define recoding instructions. In our example '80 thru Highest' precedes 'Lowest thru 80'. Thus, the program first applies the first instruction to all subjects. As soon as a subject is categorized, it will not be recoded again by a subsequent instruction.

In the variable view, we can now assign labels to the values 0 and 1:



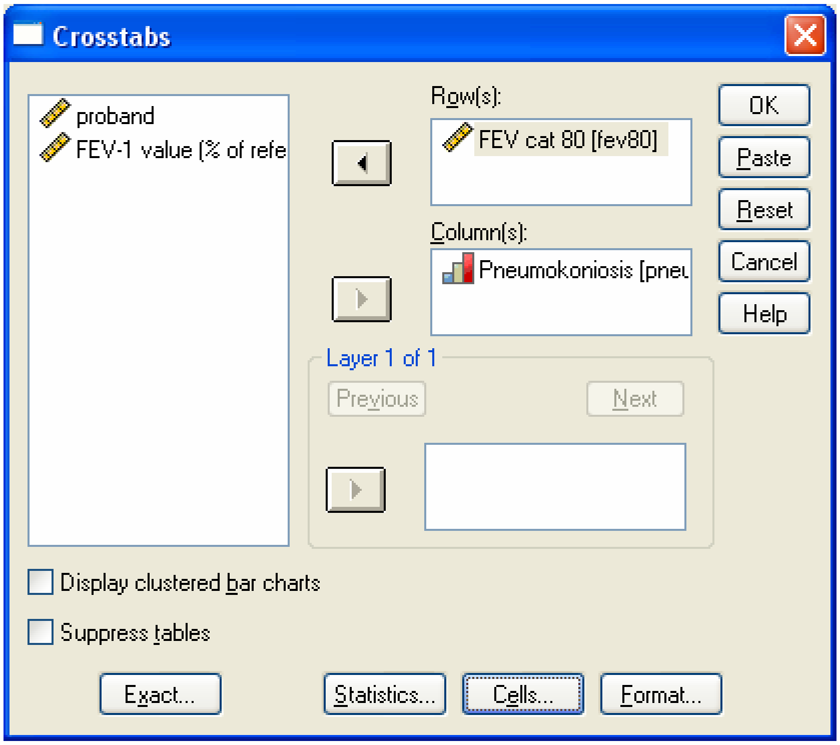
,00 = "positive test"
1,00 = "negative test"

2x2 tables: computing row and column percentages

We can now use SPSS to compute a cross table of the diagnostic test and the disease status. Choose

Analyze-Descriptive Statistics-Crosstabs…

Move pneumo (the status variable) into the field labeled ‘Column(s)’ and FEV80 (the test variable) into the ‘Row(s)’ field:



Crosstabs

proband
FEV-1 value (% of refe

Row(s):
FEV cat 80 [fev80]

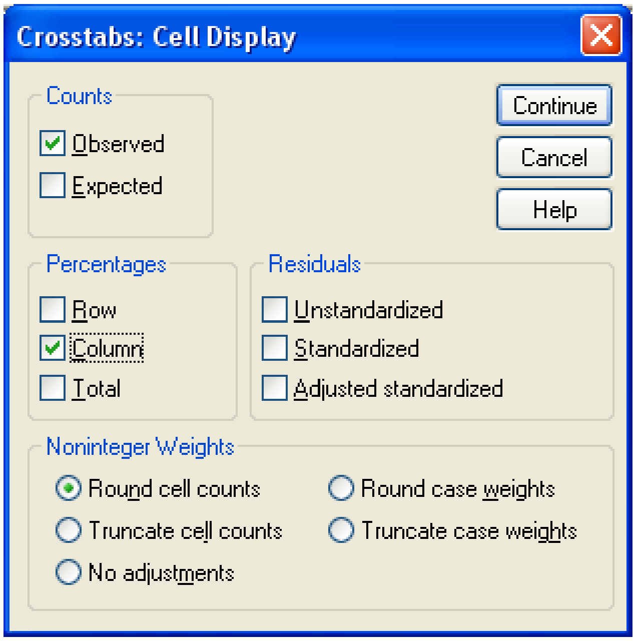
Column(s):
Pneumokoniosis [pneu

Layer 1 of 1

□ Display clustered bar charts

□ Suppress tables

Press ‘Cells…’ and choose ‘Column’ percentages to obtain the sensitivity and specificity of the test.



Crosstabs: Cell Display

Counts
☑ Observed
□ Expected

Percentages
□ Row
☑ Column
□ Total

Residuals
□ Unstandardized
□ Standardized
□ Adjusted standardized

Noninteger Weights
● Round cell counts
○ Round case weights
○ Truncate cell counts
○ Truncate case weights
○ No adjustments

FEV cat 80 * Pneumokoniosis Crosstabulation

			Pneumokoniosis		Total
			absent	present	
FEV cat 80	positive test	Count	6	22	28
		% within Pneumokoniosis	46,2%	81,5%	70,0%
	negative test	Count	7	5	12
		% within Pneumokoniosis	53,8%	18,5%	30,0%
Total		Count	13	27	40
		% within Pneumokoniosis	100,0%	100,0%	100,0%

The sensitivity is defined as the true positive rate (among the subgroup of diseased), which can be read as 81.5%. Similarly, the specificity is defined as true negative rate and computes to 53.8%.

To obtain positive and negative predictive value, we repeat the analysis, requesting ‘Row percentages’ instead of column percentages:

FEV cat 80 * Pneumokoniosis Crosstabulation

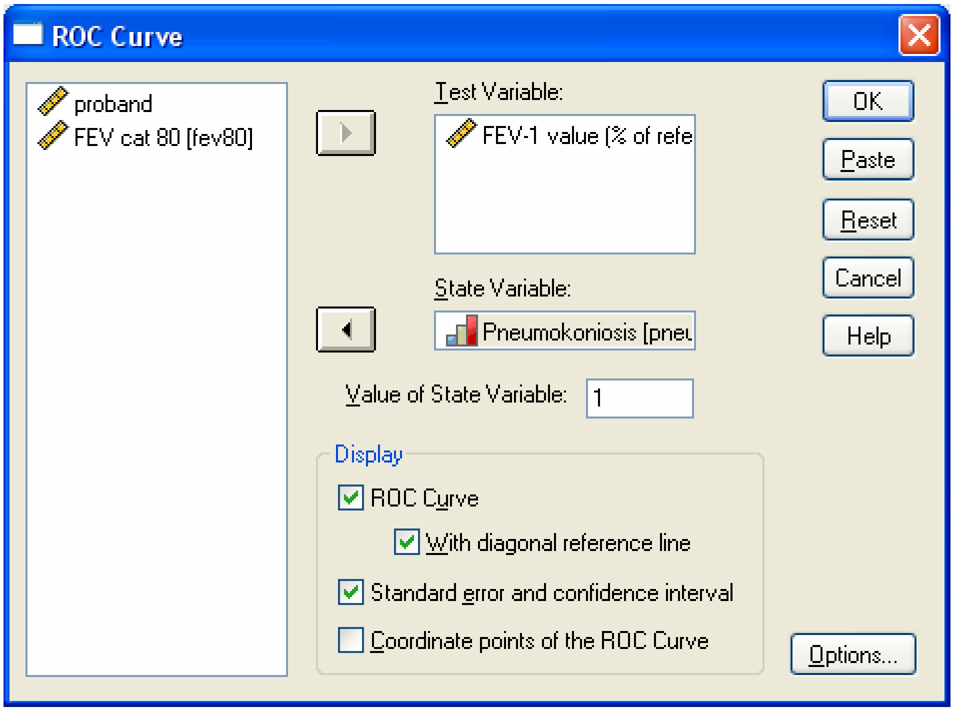
			Pneumokoniosis		Total
			absent	present	
FEV cat 80	positive test	Count	6	22	28
		% within FEV cat 80	21,4%	78,6%	100,0%
	negative test	Count	7	5	12
		% within FEV cat 80	58,3%	41,7%	100,0%
Total		Count	13	27	40
		% within FEV cat 80	32,5%	67,5%	100,0%

Looking at the same cells as before, we read a positive predictive value of 78.6% and a negative predictive value of 58.3%.

ROC curves

From the menu, choose

Graphs-ROC Curve…



ROC Curve

proband
FEV cat 80 [fev80]

Test Variable: FEV-1 value (% of refe

State Variable: Pneumokoniosis [pneu

Value of State Variable: 1

Display

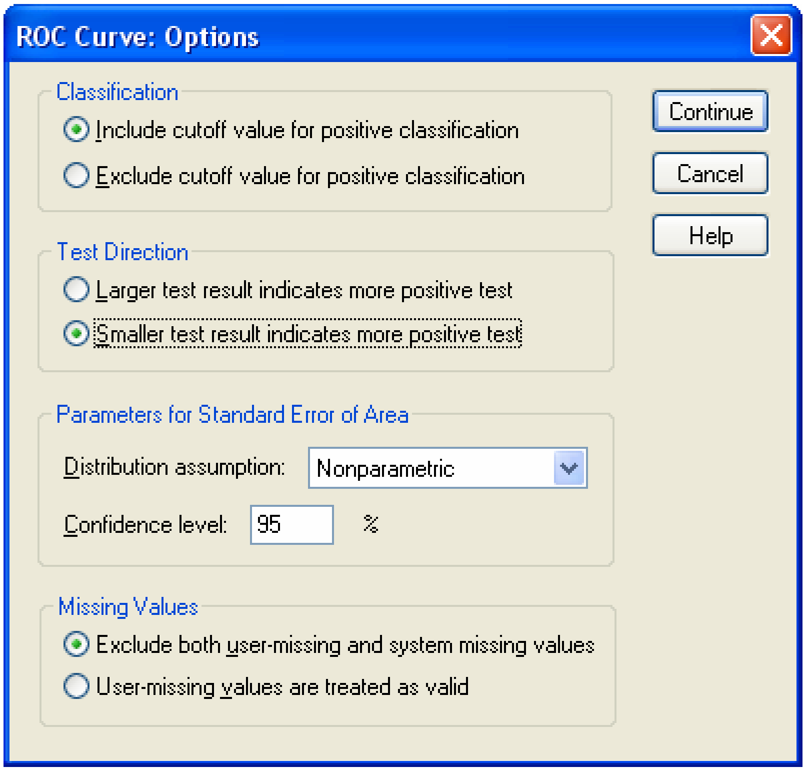
☑ ROC Curve

☑ With diagonal reference line

☑ Standard error and confidence interval

☐ Coordinate points of the ROC Curve

Define ‘FEV-1 value’ as test variable, and ‘Pneumokoniosis’ as state variable. Insert 1 as the value of the state variable that defines presence of disease. Check ‘ROC Curve’, ‘With diagonal reference line’ and ‘Standard error and confidence interval’. Press ‘Options…’:



ROC Curve: Options

Classification

● Include cutoff value for positive classification

○ Exclude cutoff value for positive classification

Test Direction

○ Larger test result indicates more positive test

● Smaller test result indicates more positive test

Parameters for Standard Error of Area

Distribution assumption: Nonparametric

Confidence level: 95 %

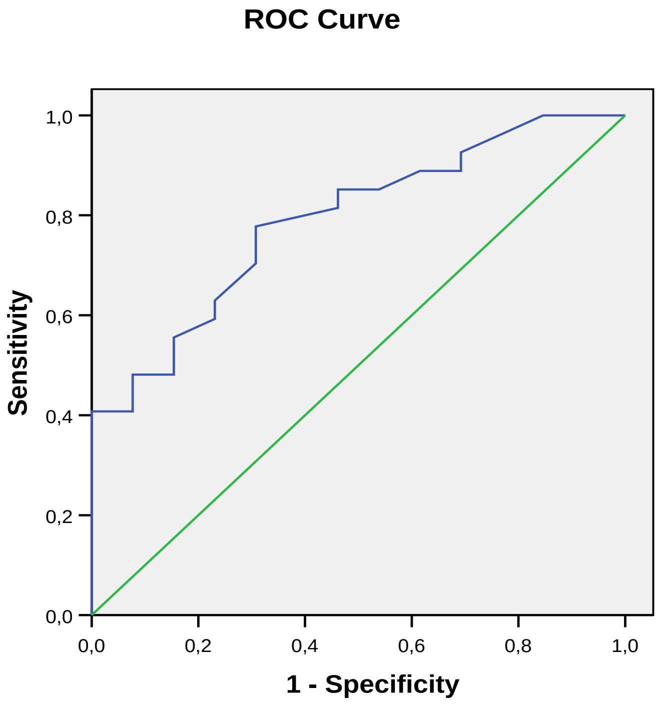
Missing Values

● Exclude both user-missing and system missing values

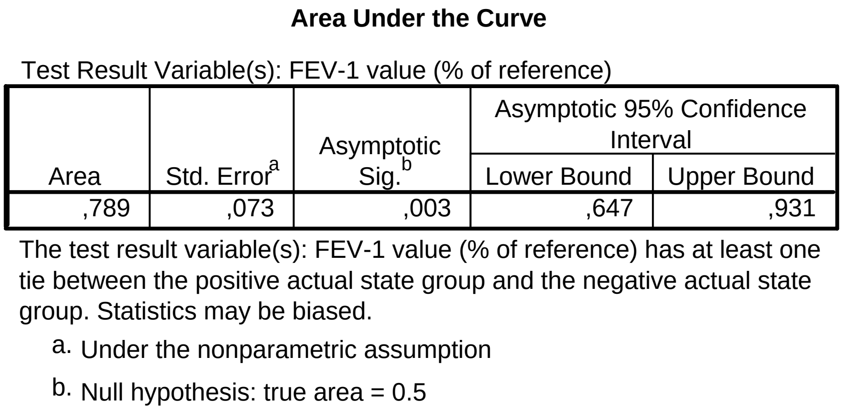
○ User-missing values are treated as valid

It’s crucial to check ‘Smaller test result indicates more positive test’, as small values of FEV-1 indicate a poor lung function. Confirm by pressing ‘Continue’ and ‘OK’.

We obtain the ROC curve and the summary about the area under the curve:



Diagonal segments are produced by ties.



Area Under the Curve

Test Result Variable(s): FEV-1 value (% of reference)

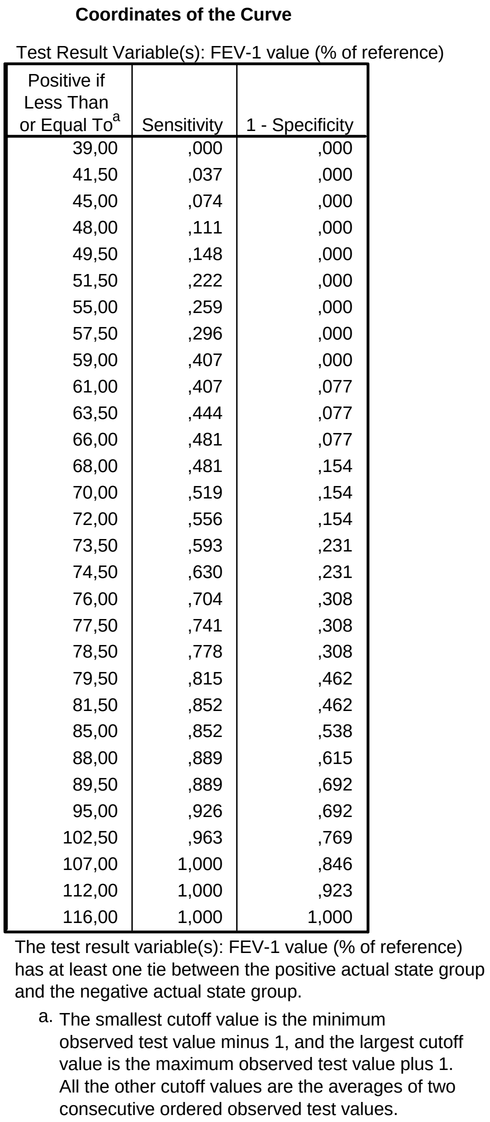
Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,789	,073	,003	,647	,931

The test result variable(s): FEV-1 value (% of reference) has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

[a] Under the nonparametric assumption

[b] Null hypothesis: true area = 0.5

Sometimes it’s useful to further examine the coordinates of the ROC curve (e. g., for defining an optimal cut value). The coordinates are output if the corresponding box ‘Coordinate points of the ROC curve’ is checked:



Coordinates of the Curve

Test Result Variable(s): FEV-1 value (% of reference)

Positive if Less Than or Equal To[a]	Sensitivity	1 - Specificity
39,00	,000	,000
41,50	,037	,000
45,00	,074	,000
48,00	,111	,000
49,50	,148	,000
51,50	,222	,000
55,00	,259	,000
57,50	,296	,000
59,00	,407	,000
61,00	,407	,077
63,50	,444	,077
66,00	,481	,077
68,00	,481	,154
70,00	,519	,154
72,00	,556	,154
73,50	,593	,231
74,50	,630	,231
76,00	,704	,308
77,50	,741	,308
78,50	,778	,308
79,50	,815	,462
81,50	,852	,462
85,00	,852	,538
88,00	,889	,615
89,50	,889	,692
95,00	,926	,692
102,50	,963	,769
107,00	1,000	,846
112,00	1,000	,923
116,00	1,000	1,000

The test result variable(s): FEV-1 value (% of reference) has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

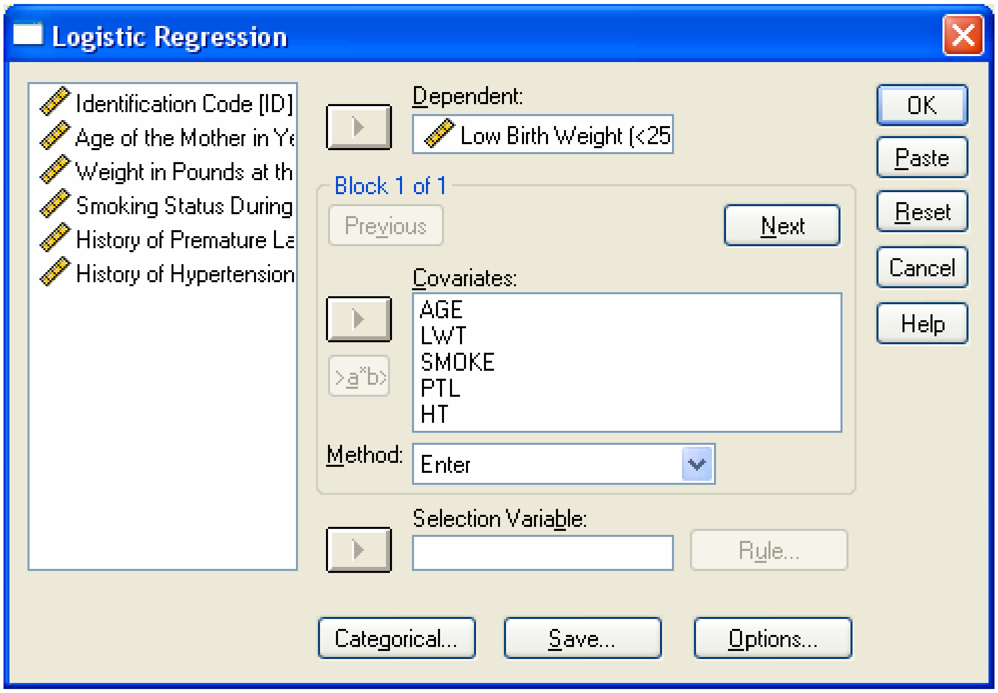
This table can be copied into Excel, say, for further analyses.

Logistic regression

We start with the low birth weight data set (lowbwt.sav). Logistic regression is called by choosing from the menu

Analyze-Regression-Binary logistic…

Define LOW as dependent variable, and the risk factors as covariates:



Identification Code [ID]
Age of the Mother in Ye
Weight in Pounds at th
Smoking Status During
History of Premature La
History of Hypertension

Dependent: Low Birth Weight (<25

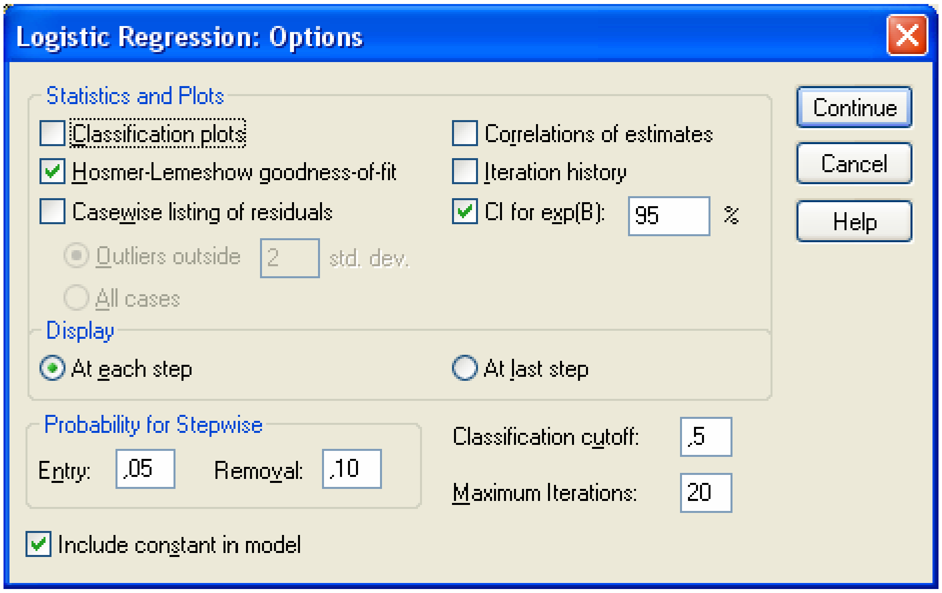
Block 1 of 1

Covariates:
AGE
LWT
SMOKE
PTL
HT

Method: Enter

Selection Variable:

Press ‘Options…’ and check ‘Hosmer-Lemeshow goodness-of-fit’ and ‘CI for exp(B): 95%’:



Statistics and Plots

☐ Classification plots
☑ Hosmer-Lemeshow goodness-of-fit
☐ Casewise listing of residuals
● Outliers outside 2 std. dev.
○ All cases
☐ Correlations of estimates
☐ Iteration history
☑ CI for exp(B): 95 %

Display

● At each step ○ At last step

Probability for Stepwise

Entry: ,05 Removal: ,10

Classification cutoff: ,5

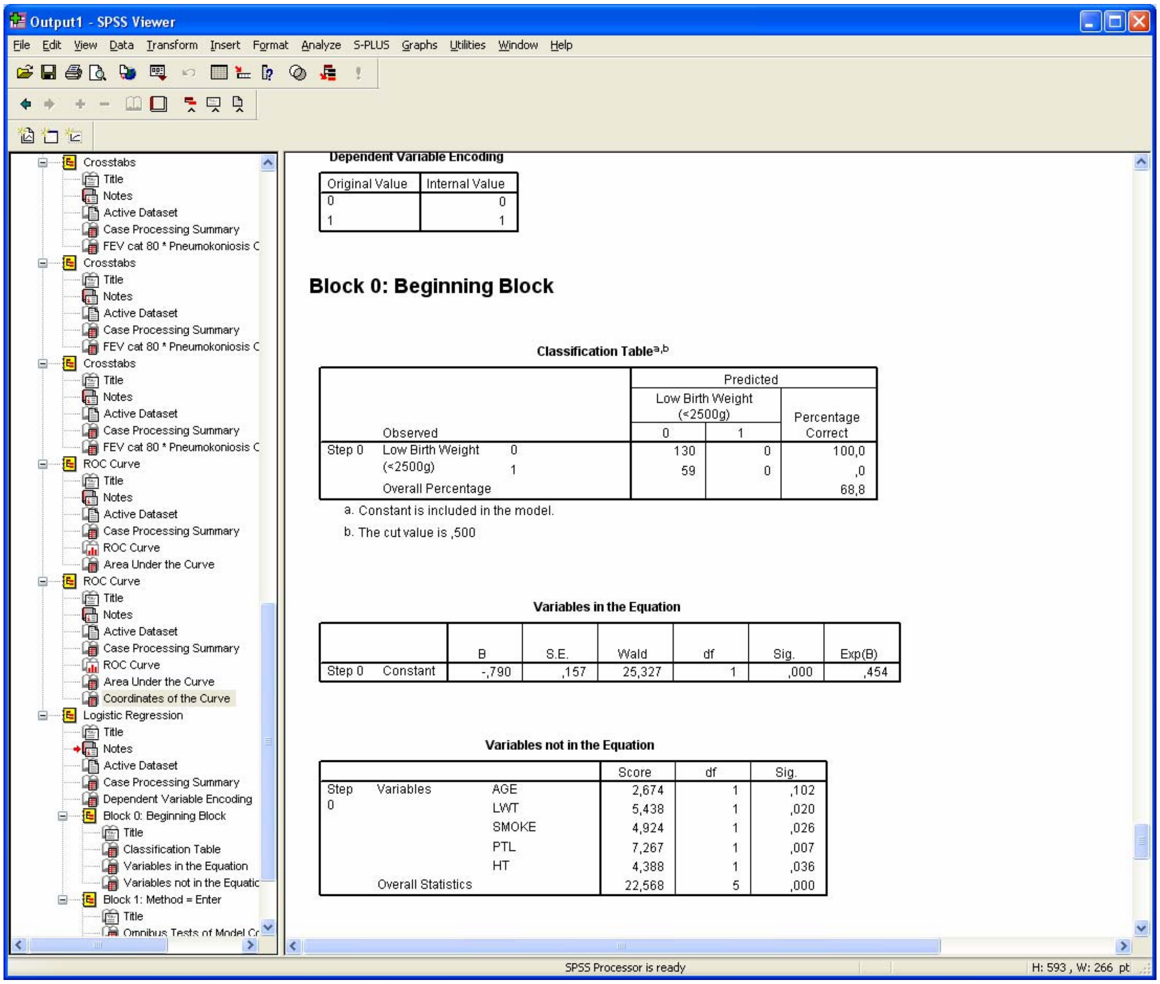
Maximum Iterations: 20

☑ Include constant in model

It’s very important to include the constant in the model!

Confirm by pressing ‘Continue’ and ‘OK’.

The program not only outputs the results of the procedure, but also some steps in between. Results labeled as ‘Block 0’ refer to pre-fit results, i.e. ‘what happens if we include certain variables in the model’:



Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			Low Birth Weight (<2500g)		Percentage Correct
	Observed		0	1	
Step 0	Low Birth Weight (<2500g)	0	130	0	100,0
		1	59	0	,0
	Overall Percentage				68,8

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

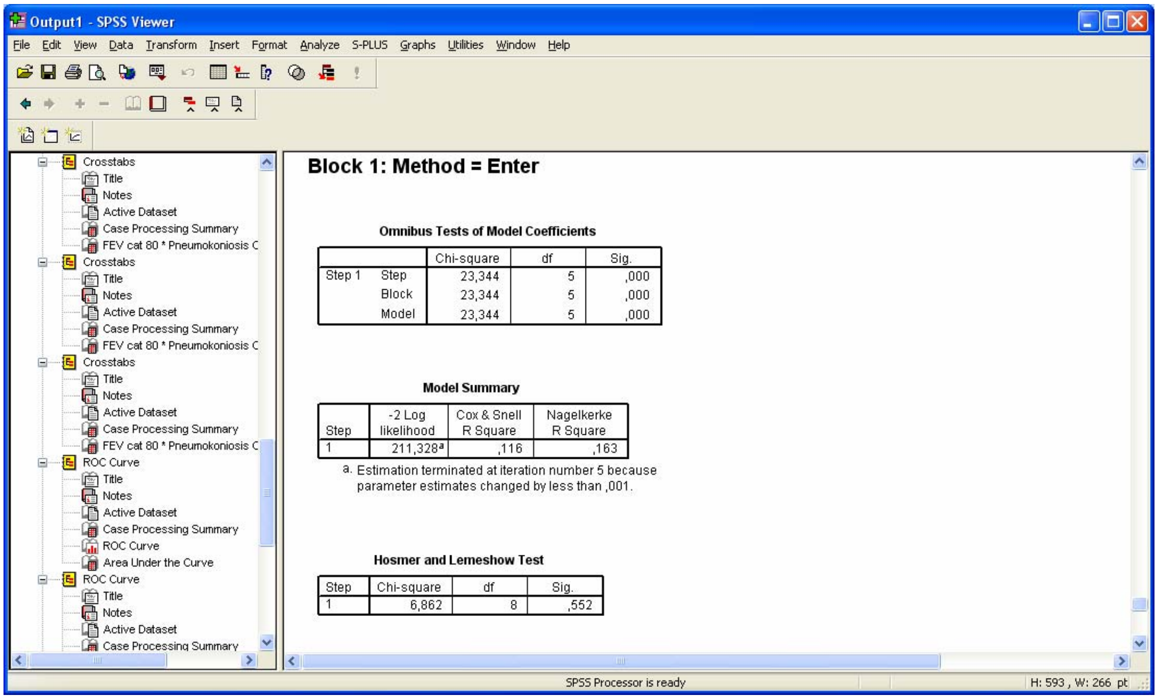
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-,790	,157	25,327	1	,000	,454

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	AGE	2,674	1	,102
		LWT	5,438	1	,020
		SMOKE	4,924	1	,026
		PTL	7,267	1	,007
		HT	4,388	1	,036
	Overall Statistics		22,568	5	,000

We see, at step 0, only the constant is in the model, and all other variables are not. SPSS performs some tests evaluating whether inclusion of these variables would significantly improve this null model. We see that with the exception of AGE, inclusion of any other variable would improve the model significantly.

Now let's have a look at Block 1. We requested to enter all specified covariates simultaneously (as defined by 'Method: Enter' in the logistic regression dialogue). If the 'Enter'-method was requested, SPSS performs one single step: entering all variables. We will later choose other methods, which may produce more than one step. The results contain several tables, which have been explained previously:



Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	23,344	5	,000
	Block	23,344	5	,000
	Model	23,344	5	,000

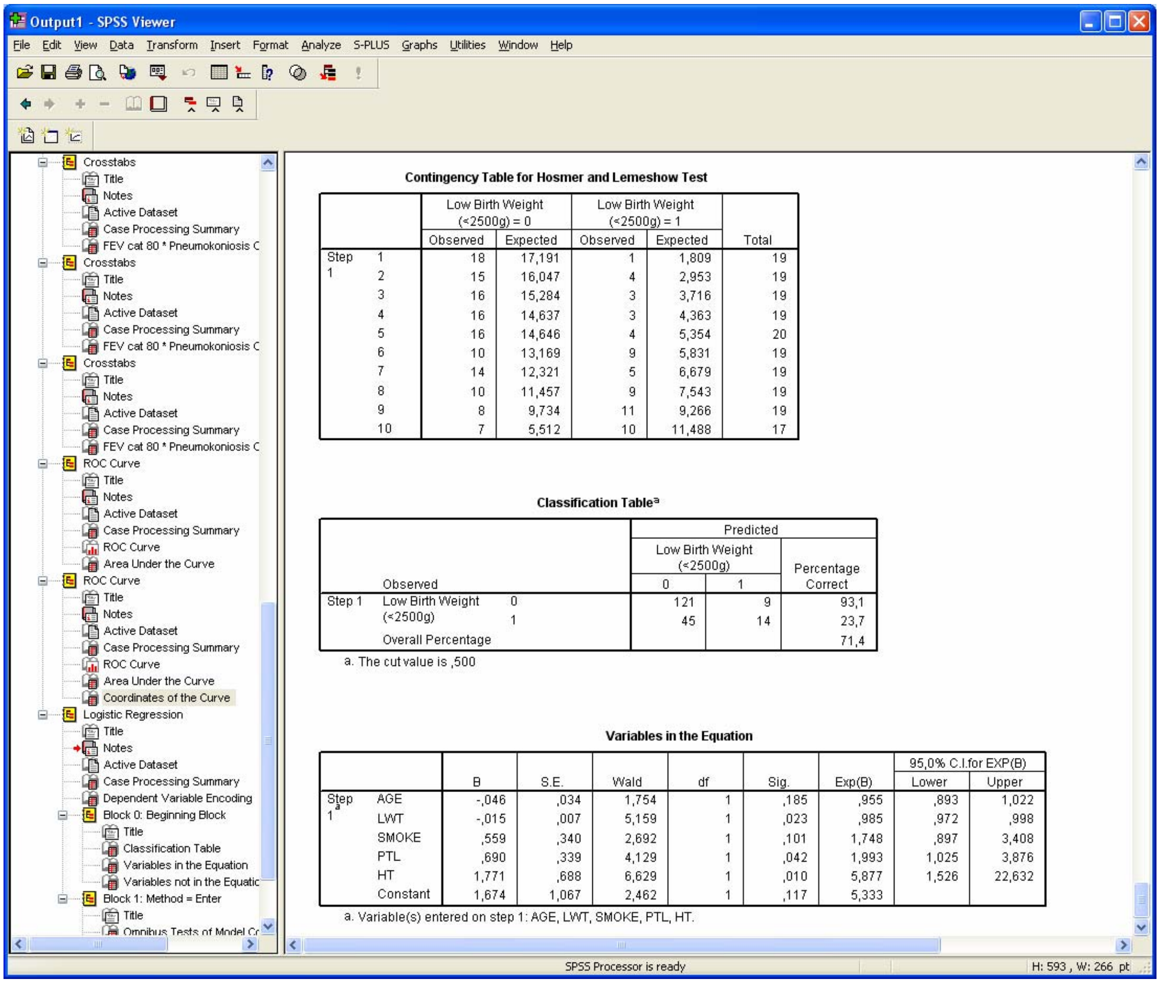
Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	211,328[a]	,116	,163

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,862	8	,552



Contingency Table for Hosmer and Lemeshow Test

		Low Birth Weight (<2500g) = 0		Low Birth Weight (<2500g) = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	18	17,191	1	1,809	19
	2	15	16,047	4	2,953	19
	3	16	15,284	3	3,716	19
	4	16	14,637	3	4,363	19
	5	16	14,646	4	5,354	20
	6	10	13,169	9	5,831	19
	7	14	12,321	5	6,679	19
	8	10	11,457	9	7,543	19
	9	8	9,734	11	9,266	19
	10	7	5,512	10	11,488	17

Classification Table[a]

			Predicted		
			Low Birth Weight (<2500g)		Percentage Correct
Observed			0	1	
Step 1	Low Birth Weight (<2500g)	0	121	9	93,1
		1	45	14	23,7
	Overall Percentage				71,4

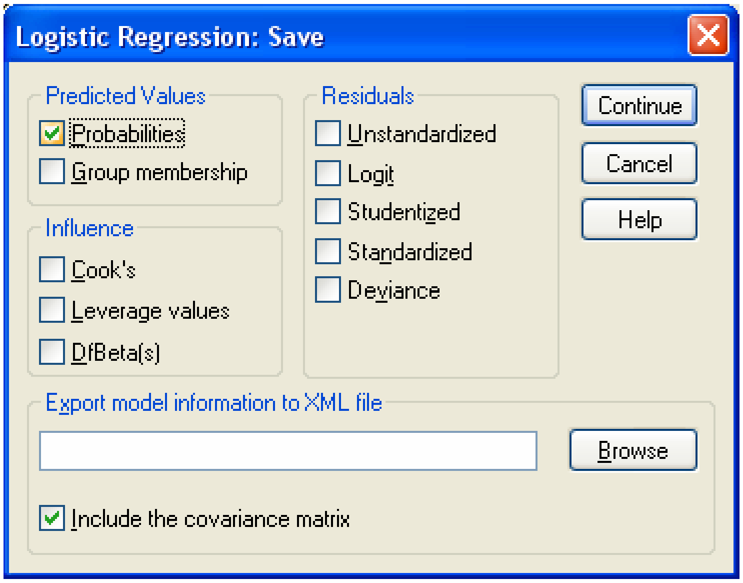
a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	AGE	-,046	,034	1,754	1	,185	,955	,893	1,022
	LWT	-,015	,007	5,159	1	,023	,985	,972	,998
	SMOKE	,559	,340	2,692	1	,101	1,748	,897	3,408
	PTL	,690	,339	4,129	1	,042	1,993	1,025	3,876
	HT	1,771	,688	6,629	1	,010	5,877	1,526	22,632
	Constant	1,674	1,067	2,462	1	,117	5,333		

a. Variable(s) entered on step 1: AGE, LWT, SMOKE, PTL, HT.

Now suppose we want to compute, for each subject, the probability of low weight birth. At the logistic regression dialogue, press ‘Save’ and choose ‘Predicted values: Probabilities’ as shown below:



Predicted Values
☑ Probabilities
□ Group membership

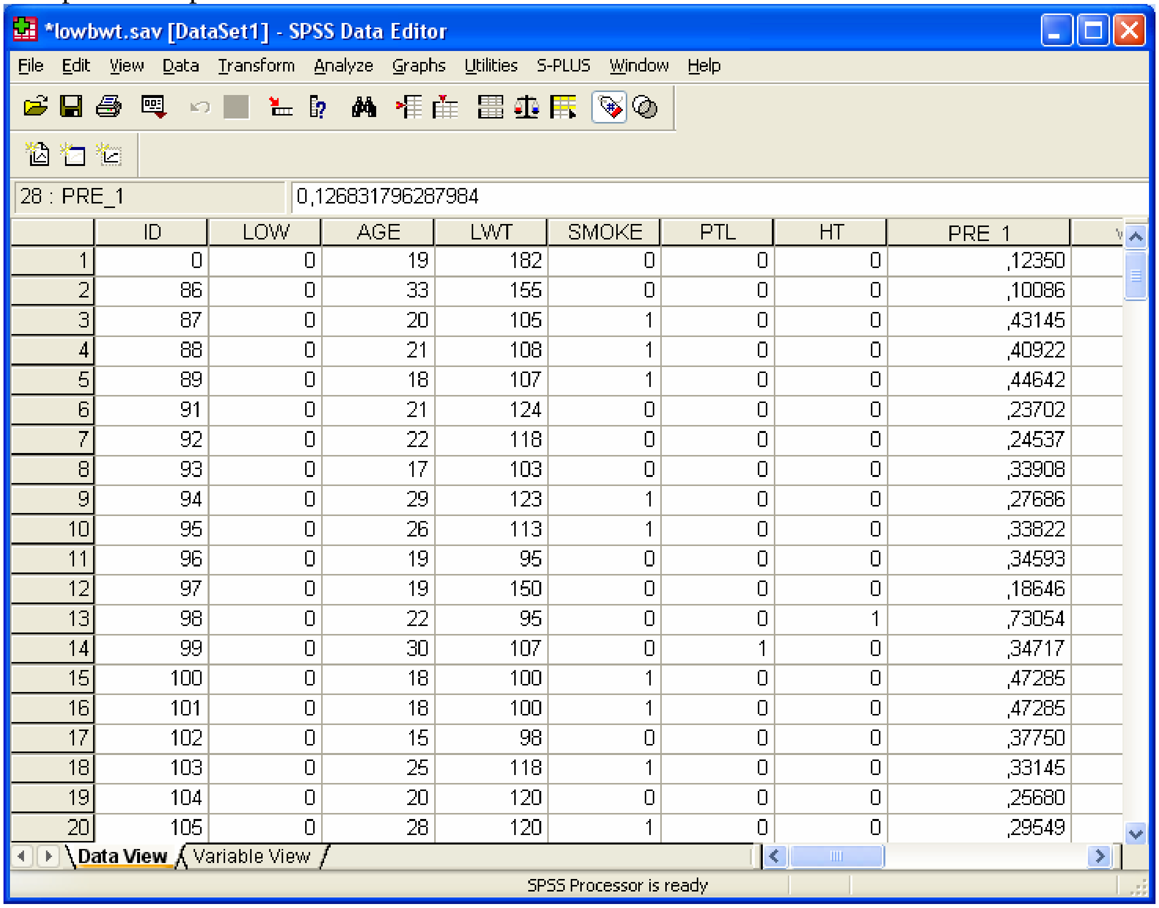
Influence
□ Cook's
□ Leverage values
□ DfBeta(s)

Residuals
□ Unstandardized
□ Logit
□ Studentized
□ Standardized
□ Deviance

Export model information to XML file

☑ Include the covariance matrix

The predicted probabilities are now contained in a new column of the data editor:


*lowbwt.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View
SPSS Processor is ready

28 : PRE_1 0,126831796287984

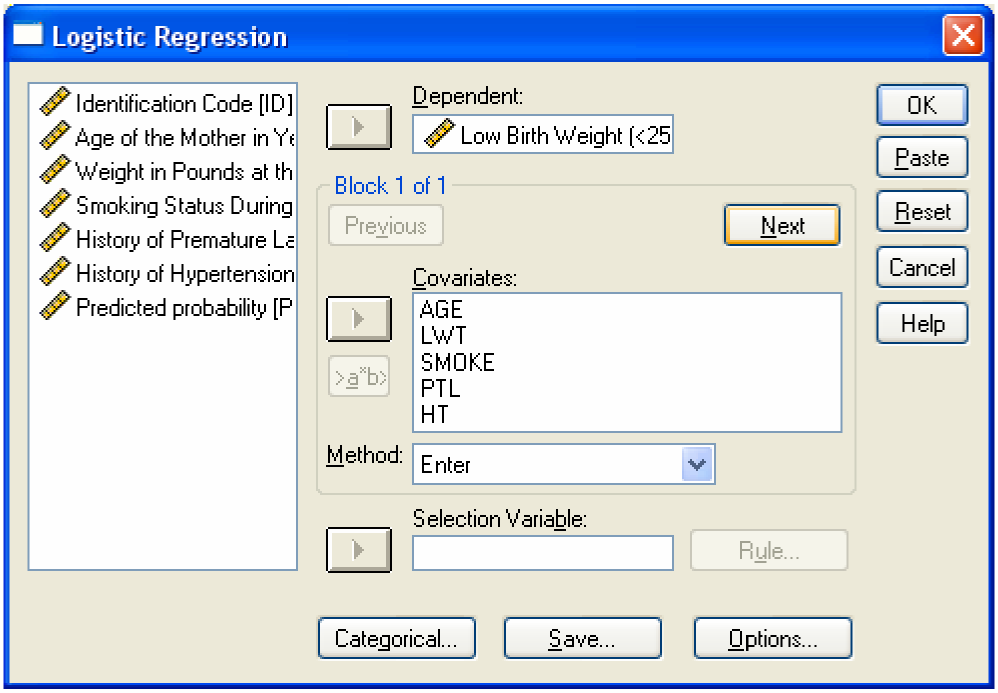
	ID	LOW	AGE	LWT	SMOKE	PTL	HT	PRE_1
1	0	0	19	182	0	0	0	,12350
2	86	0	33	155	0	0	0	,10086
3	87	0	20	105	1	0	0	,43145
4	88	0	21	108	1	0	0	,40922
5	89	0	18	107	1	0	0	,44642
6	91	0	21	124	0	0	0	,23702
7	92	0	22	118	0	0	0	,24537
8	93	0	17	103	0	0	0	,33908
9	94	0	29	123	1	0	0	,27686
10	95	0	26	113	1	0	0	,33822
11	96	0	19	95	0	0	0	,34593
12	97	0	19	150	0	0	0	,18646
13	98	0	22	95	0	0	1	,73054
14	99	0	30	107	0	1	0	,34717
15	100	0	18	100	1	0	0	,47285
16	101	0	18	100	1	0	0	,47285
17	102	0	15	98	0	0	0	,37750
18	103	0	25	118	1	0	0	,33145
19	104	0	20	120	0	0	0	,25680
20	105	0	28	120	1	0	0	,29549

We can now obtain a ROC curve to assess the model fit using PRE_1 as test variable and LOW as state variable (and selecting 'Higher value indicates more positive test result').

Checking for interactions

It is often desirable to check for interactions (effect modifications) that may exist between a model's covariates. For this purpose, forward selection could be applied, starting with a model containing all variables in question.

The forward selection of significant interactions can now be performed as follows: first, open the logistic regression dialogue:

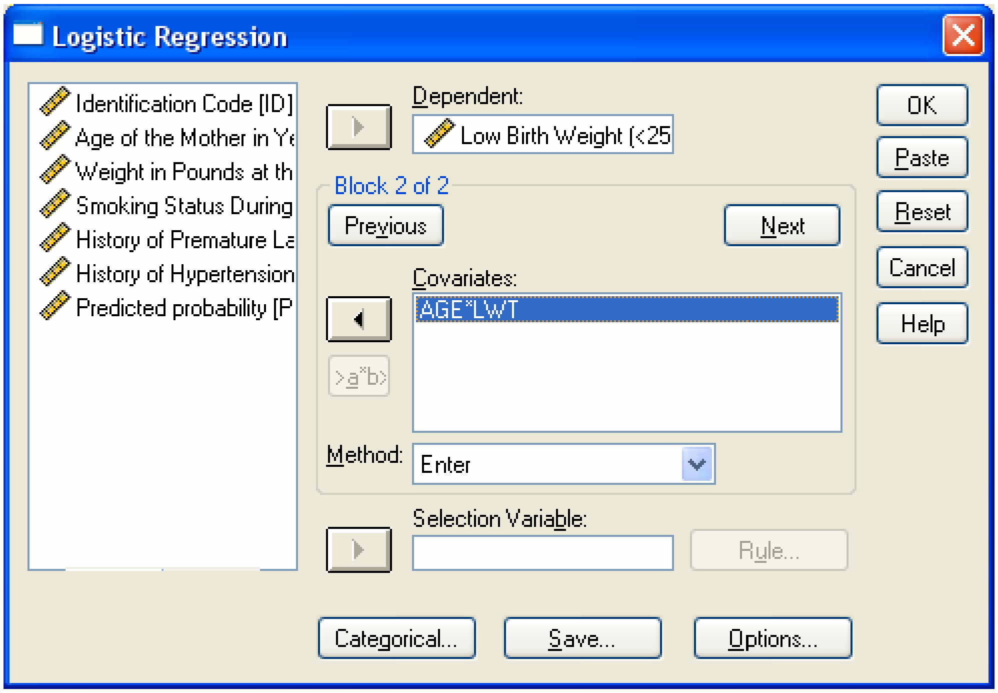


Dependent: Low Birth Weight [<25

Covariates:
AGE
LWT
SMOKE
PTL
HT

Method: Enter

Click on 'Next' to define another 'Block':



Logistic Regression

Identification Code [ID]
Age of the Mother in Ye
Weight in Pounds at th
Smoking Status During
History of Premature La
History of Hypertension
Predicted probability [P

Dependent: Low Birth Weight (<25

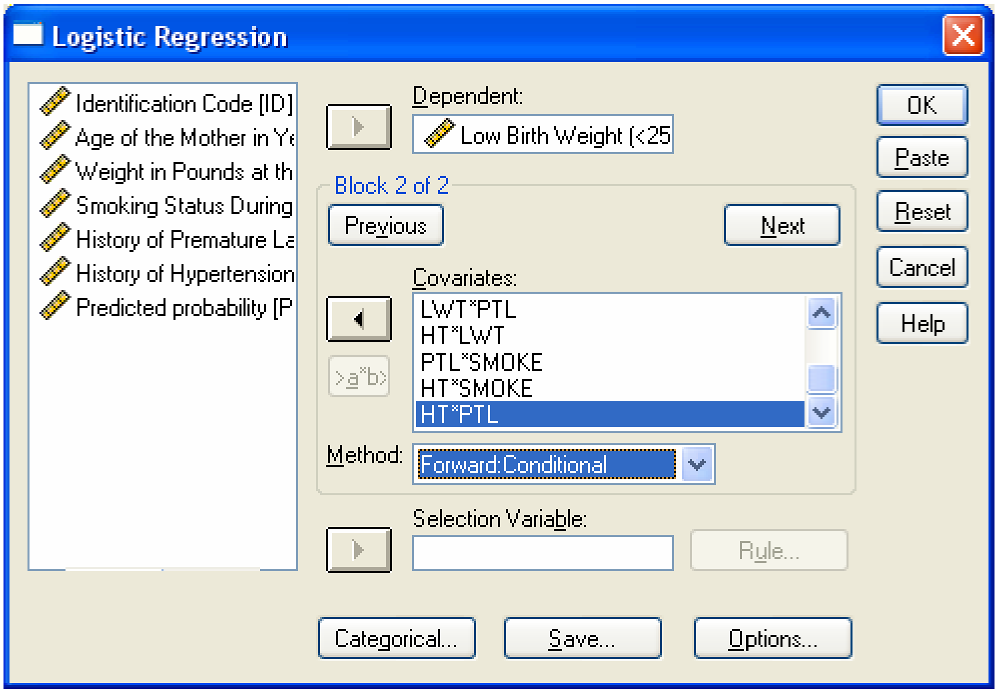
Block 2 of 2

Covariates:
AGE*LWT

Method: Enter

Selection Variable:

Here, we select all possible pairs of covariates (click on the first covariate, then hold the Strg key and click on the second covariate), and press the button '>a*b>' to include their interaction as a candidate covariate in the model. After having defined all possible pairs, we change the 'Method:' to 'Forward: Conditional':



Logistic Regression

Identification Code [ID]
Age of the Mother in Ye
Weight in Pounds at th
Smoking Status During
History of Premature La
History of Hypertension
Predicted probability [P

Dependent: Low Birth Weight (<25

Block 2 of 2

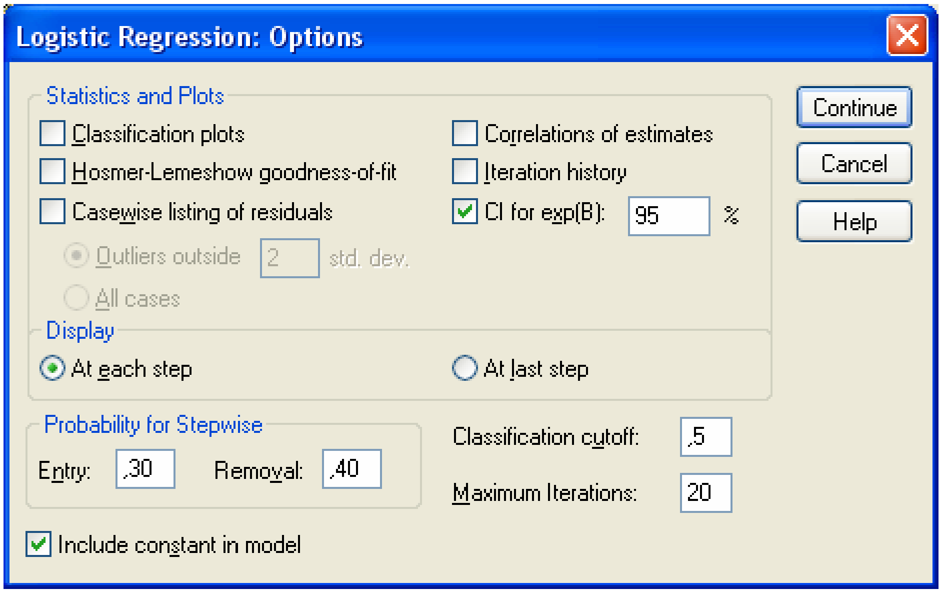
Covariates:
LWT*PTL
HT*LWT
PTL*SMOKE
HT*SMOKE
HT*PTL

Method: Forward:Conditional

Selection Variable:

Press 'OK'. In this example, none of the interaction terms is significant (significance margins are defined at the 'Options…' dialogue, so there is no output in Block 2.

To exemplify the stepwise selection of interactions, we temporarily change the significance margin for entry to 0.3 (instead of the usual .05) and for removal to 0.40:



Logistic Regression: Options

Statistics and Plots

□ Classification plots
□ Hosmer-Lemeshow goodness-of-fit
□ Casewise listing of residuals
○ Outliers outside 2 std. dev.
○ All cases
□ Correlations of estimates
□ Iteration history
☑ CI for exp(B): 95 %

Display

● At each step
○ At last step

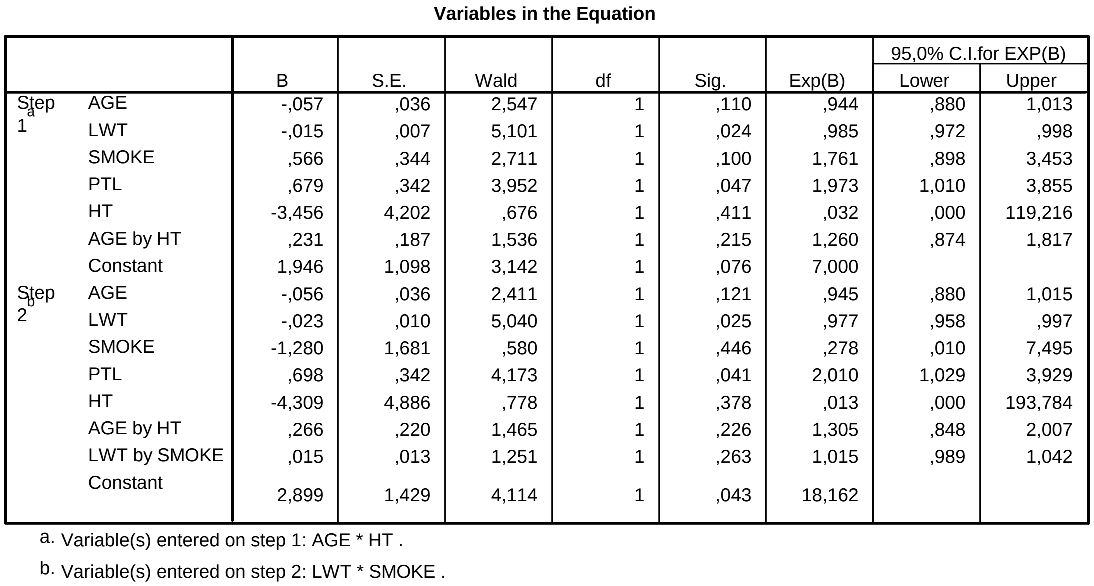
Probability for Stepwise

Entry: ,30
Removal: ,40

Classification cutoff: ,5
Maximum Iterations: 20

☑ Include constant in model

The most important table is the one labeled ‘Variables in the Equation’, and from this table only the results for ‘Step 2’ are interesting (as these are the final results):



Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	AGE	-,057	,036	2,547	1	,110	,944	,880	1,013
	LWT	-,015	,007	5,101	1	,024	,985	,972	,998
	SMOKE	,566	,344	2,711	1	,100	1,761	,898	3,453
	PTL	,679	,342	3,952	1	,047	1,973	1,010	3,855
	HT	-3,456	4,202	,676	1	,411	,032	,000	119,216
	AGE by HT	,231	,187	1,536	1	,215	1,260	,874	1,817
	Constant	1,946	1,098	3,142	1	,076	7,000		
Step 2[b]	AGE	-,056	,036	2,411	1	,121	,945	,880	1,015
	LWT	-,023	,010	5,040	1	,025	,977	,958	,997
	SMOKE	-1,280	1,681	,580	1	,446	,278	,010	7,495
	PTL	,698	,342	4,173	1	,041	2,010	1,029	3,929
	HT	-4,309	4,886	,778	1	,378	,013	,000	193,784
	AGE by HT	,266	,220	1,465	1	,226	1,305	,848	2,007
	LWT by SMOKE	,015	,013	1,251	1	,263	1,015	,989	1,042
	Constant	2,899	1,429	4,114	1	,043	18,162		

a. Variable(s) entered on step 1: AGE * HT .

b. Variable(s) entered on step 2: LWT * SMOKE .

Two interactions have been included in the model: Age by HT and LWT by SMOKE. This means that the effect of age depends on history of hypertension (if a significance level of 0.3 was postulated!), and that the effect of mother’s weight depends on her smoking status.

For non-hypertensive mothers, the effect of age is -0.056, or expressed as an odds ratio, 0.945 which means that with each year, the risk of delivering a low weight baby decreases by 5.5%.

For hypertensive mothers, the effect of age is -0.056+0.266=0.21 corresponding to an odds ratio of 1.23, which means that with each year of age, the risk for a low weight baby increases by 23%.

References

[1] Tong, D et al. Expression of KLF5 is a prognostic factor for disease-free survival and overall survival in patients with breast cancer. Clin Cancer Res. 2006; 12(8):2442-8.
[2] Egger M. Meta-analysis. principles and procedures. BMJ 1997;315:1533–7.
[3] Davies HTO, Crombie IK, Tavakoli M. When can odds ratios mislead? BMJ 1998; 316:989–91.
[4] Rascol O, Brooks D, Korczyn AD, et al. A fiveyear study of the incidence of dyskinesia in patients with early Parkinson’s disease who were treated with ropinirole or levodopa. N Engl J Med 2000; 342:1484–91.
[5] Hosmer, D, Lemeshow, St. Applied logistic regression. New York: Wiley, 2000.
[6] Campbell, M, Machin, D. Medical statistics. New York: Wiley, 1995.
[7] Reisinger, J et al. Prospective Comparison of Flecainide Versus Sotalol for Immediate Cardioversion of Atrial Fibrillation. American Journal of Cardiology 1998; 81:1450-1454.

Part 3: Analysis of survival outcomes

Class 5: Survival outcomes

Definition of survival data

Examples of survival data:

- Survival after an operation (e. g., resection of a tumor)
- Functional graft survival (e. g., after kidney transplant)
- Time to rejection (e. g., after transplant)
- Functional survival of a bypass
- Time to recurrence of disease (e. g., tumor)
- Time to remission

All these examples have in common, that we are interested in the time that elapses from a well-defined starting point (e. g., operation, onset of therapy, diagnosis), until occurrence of a particular event (e. g, death, time of progression, rejection, etc.). This time is generally called survival time.

However, it is very unlikely, that at time of analysis the event of interest has occurred in all patients of a study.

Example: dogs. Consider the following experiment: 11 dogs are treated with an experimental drug, and for each dog we take records of its survival time. The data looks as follows:

```
Name               Entry     Last date      min                                                max
                                             01JAN2004                                     01NOV2005
                                            *------------------------------------------------------*

Matt           01JAN2004     01DEC2004      |E--------------------------d                          |
Andy           01FEB2004     01APR2005      |  E----------------------------------d                |
Jack           01MAR2004     01JAN2005      |     E------------------------d                       |
Sim            01APR2004     01JUN2005      |       E----------------------------------d           |
Jimmy          01MAY2004     01JUN2005      |          E-------------------------------d           |
Phil           01JUN2004     01NOV2005      |            E----------------------------------------d|
Bart           01JUL2004     01MAY2005      |              E------------------------a              |
Tommy          01JUL2004     01OCT2005      |              E-------------------------------------a |
Teddy          01SEP2004     01OCT2005      |                   E--------------------------------d |
Jody           01OCT2004     01NOV2005      |                      E------------------------------a|
Dolly          01OCT2004     01NOV2005      |                      E------------------------------d|
                                            *------------------------------------------------------*
```

The letters ‘d’ and ‘a’ indicate whether a dog was dead or alive at the ‘last contact’ date.

Summary:

Median survival = 14 months

or
One-year survival = 81% +/- 12%

Note that we are not able to measure the complete survival time for some dogs in our experiment. Dogs with status ‘alive’ have incomplete or censored information on survival times. We know the exact survival time for Matt, Andy, Jack, Sim, Jimmy, Phil, Teddy and Dolly. Such observations are called complete. For Jody all we know is that he survived more than 13 months, we don’t know if his survival time is 14 months or 26 months.

What can cause censoring to occur?

Compare Jody to Tommy (who ran away) and Bart (who was eaten by Dolly).

An important assumption is that a subject’s survival time is independent of any mechanism which causes that individual’s survival time to be censored at a particular time. This assumption is called the ‘non-informative censoring assumption’. Standard methods for analysis of survival data (as presented throughout here) require this assumption to hold.

Censoring is inappropriate if the censoring mechanism is in any way related to the probability of the event of interest. We call such mechanisms ‘informative censoring’. For example, if a subject is removed from study because he is moribund, treating the survival time as censored would introduce bias to the estimate of survival.

Tommy is lost to follow-up because he ran away.

- Does running away indicate he was seeking a place to die? (increased risk)
- Or was he seeking a postman to harass? (vitality associated with decreased risk)

Bart was eaten by Dolly.

Could this be related to treatment (a toxic reaction that weakened or killed him)?

In a clinical trial of a new therapy for diabetes, would you consider an observation to be censored if the patient:

- Committed suicide?
- Was killed in a car accident?

Competing risks:

Consider a study of an intervention to prevent death from stroke. The population is at risk of death from many other causes. Should patients who die from other causes (e. g., cancer, myocardial infarction, etc.) be censored?

- Pros: the conclusion of the study is more focused if only study-related causes of death are taken into account.
- Cons: it is often difficult to identify the precise cause of death with any certainty.

To be on the safe side, it might be wiser to consider ‘death from all causes’ as the event.

Henceforth, we will (have to) assume that Bart actually left town and refused further cooperation with the investigators.

What can we do about censoring?

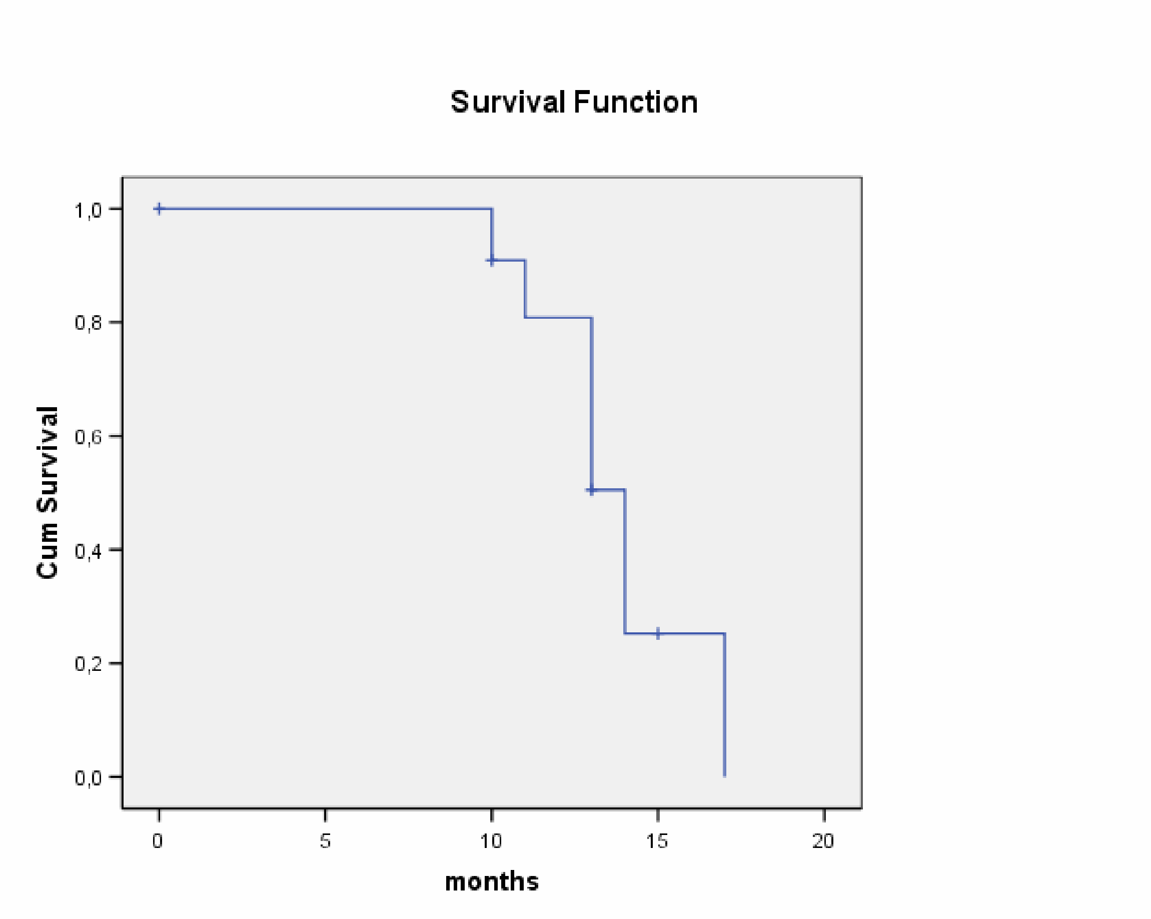
- Wait until all individuals experience the event
    - Impossible if subjects are lost to follow-up
    - May take too long

- Ignore censored cases
    - Introduces bias to estimates of survival
    - Example: For some disease, 60 patients are treated by A and 60 by B, and then followed up for one year. In group A, two patients die in the first month, the rest survive. In group B, each month three patients die. Ignoring censoring gives average survival of 6 months in group B and 1 month for group A.

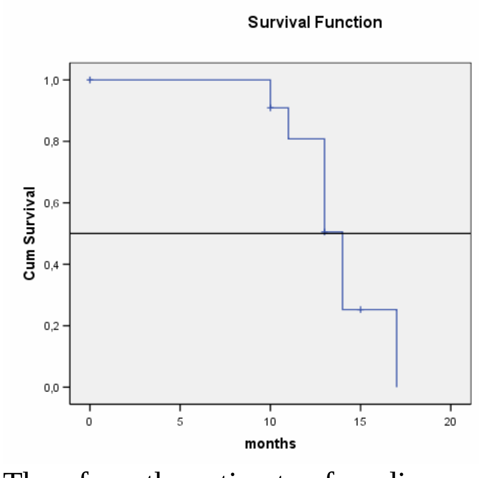
Potentially censored survival data can be properly dealt with using the methods presented in the following.

Kaplan-Meier estimates of survival functions

The Kaplan-Meier method [1] can use the information that lies in censored observations efficiently. The result of a Kaplan-Meier analysis is a survival curve like the following:



This plot shows time on the horizontal axis, and the probability to survive on the vertical axis. The curve crosses 50% survival at the time of 14 months:



Therefore, the estimate of median survival is 14 months. Similarly, we can obtain estimates of the 25th or the 75th percentile of survival time.

How are the Kaplan-Meier estimates computed?

First, all survival episodes are synchronized to an (artifical) common starting point:

```
Name           Start     months      min                                                        max
                                     0                                                           17
                                    *-------------------------------------------------------------*
Matt               0         11     |S---------------------------------------d                    |
Andy               0         14     |S-------------------------------------------------d          |
Jack               0         10     |S-----------------------------------d                        |
Sim                0         14     |S-------------------------------------------------d          |
Jimmy              0         13     |S----------------------------------------------d             |
Phil               0         17     |S-----------------------------------------------------------d|
Bart               0         10     |S-----------------------------------a                        |
Tommy              0         15     |S-----------------------------------------------------a      |
Teddy              0         13     |S----------------------------------------------d             |
Jody               0         13     |S----------------------------------------------a             |
Dolly              0         13     |S----------------------------------------------d             |
                                    *-------------------------------------------------------------*
```

I. e., the survival time is computed by subtracting the entry date from the last observation date.

Next, at each time point at which an event occurred, an estimate of the conditional survival probability is obtained:

Survival Table

	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	10,000	dead	,909	,087	1	10
2	10,000	alive	.	.	1	9
3	11,000	dead	,808	,122	2	8
4	13,000	dead	.	.	3	7
5	13,000	dead	.	.	4	6
6	13,000	dead	,505	,158	5	5
7	13,000	alive	.	.	5	4
8	14,000	dead	.	.	6	3
9	14,000	dead	,253	,149	7	2
10	15,000	alive	.	.	7	1
11	17,000	dead	,000	,000	8	0

The shortest survival time was 10 months. Eleven dogs lived up to 10 months. One dog died at 10 months. Ten dogs survived at least a little bit longer than 10 months (we assume that censoring at 10 months means: surviving a little bit longer than 10 months). Thus, the survival probability at time 10 is 10/11=90.9%.

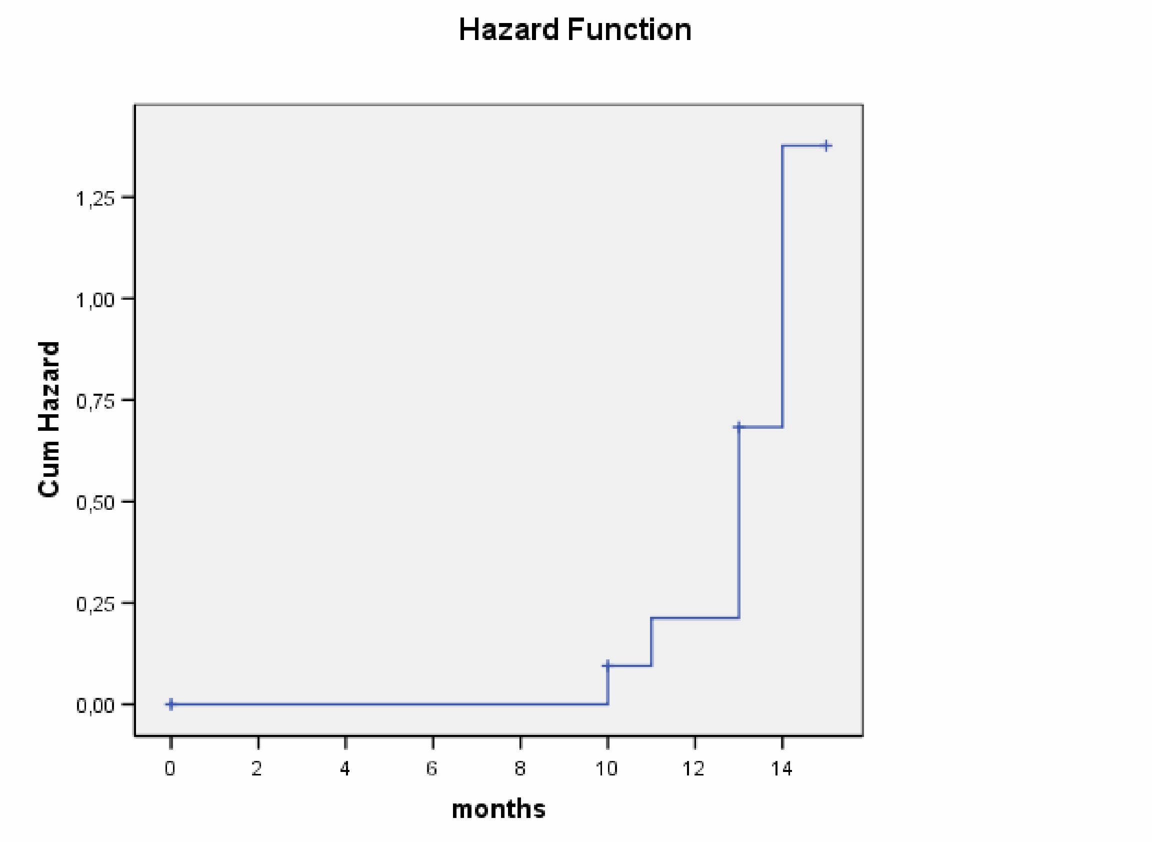
Immediately after the 10th month, one dog was lost due to censoring. Now, only 9 are still under observation.

After 11 months, another dog dies. 9 dogs were under observation up to 11 months, 8 dogs survived that time point. The survival probability at 11 months is thus 8/9=88.9%, but this probability is conditional that a dog lived up 11 months. To obtain a cumulative survival probability, we must consider survival at all time points before 11. Thus we have a cumulative survival probability of 90.9% x 88.9% = 80.8%.

The computation is carried on until the last dog has died or disappeared.

Please note that the Kaplan-Meier curve is a step function: the curve stays at a certain level until the next event occurs. At each event time, the curve drops.

The step heights provide an estimate of the risk of death at certain time points. Relating the step heights of the Kaplan-Meier curve to the total height of the curve at the times the steps occur, and cumulating them generates a so-called ‘cumulative hazard plot’. We see that the relative step heights of the Kaplan-Meier curve and so the risk to die increase by time.



Kaplan-Meier curves can also be used to compare groups, as will be demonstrated below.

Simple tests

Basically there are two nonparametric tests available to compare survival curves between groups: the log rank test and the generalized Wilcoxon test. A third test (the Tarone-Ware test) is a kind of mixture of these two.

The null hypothesis of these tests is:

There is no difference in the distribution of survival times.

The alternative hypothesis states:

The groups differ in their survival time distributions.

The log rank test is obtained by constructing a set of 2 by 2 tables, one at each distinct event time. In each table, the death rates are compared between the two groups, conditional on the number of subjects at risk in the groups. Observed death rates are compared to those expected under the null hypothesis. The information from each table is then combined into a single test statistic.

The table at time t(j) has the following structure:

Group	Died at t(j)	Did not die at tj	Total
1	2	38	40
2	8	52	60
Total	10	90	100

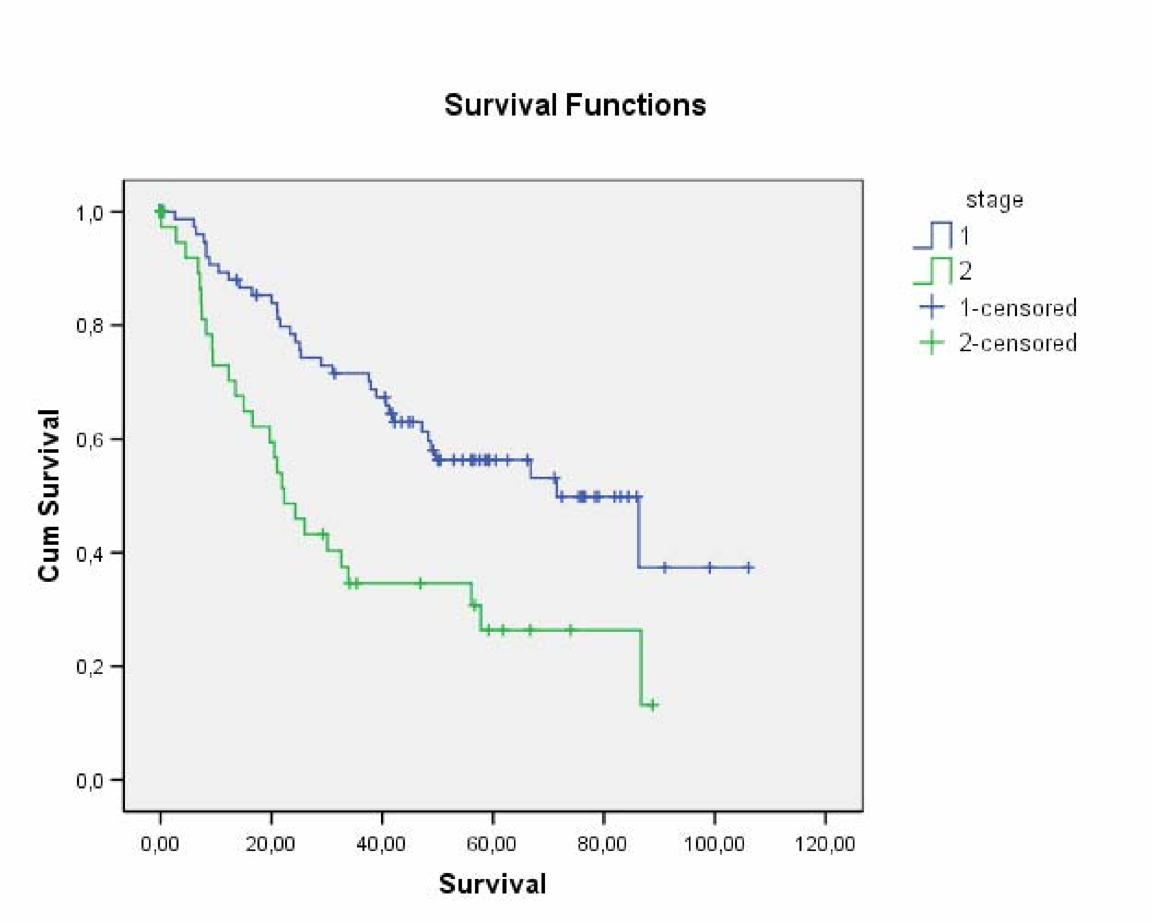
In the example given above, we observe two deaths at time t(j) in group 1, and 8 deaths in group 2. In group 1, we would expect 40% of all deaths that occurred at this time (because the proportion of subjects in group 1 is 40 out of 100 at this time). In total, 10 deaths occurred, thus we would expect 4 in group 1. However, there were only 2 deaths in group 1. The contribution to the log rank statistic of this table is therefore 2 – 4 = -2.

The contributions are summed up over all tables and related to their variance. The resulting test statistic follows a chi-squared distribution with one degree of freedom. This means, we can look up a p-value by using a table of quantiles of a chi-squared distribution with one degree of freedom.

The generalized Wilcoxon test constructs its test statistic in the very same way, but it weights the contributions of each table by the number of subjects that are at risk just before t(j). Thus, early time points obtain more weight compared to late time points. This is often desirable if the effect of an experimental condition vanishes with ongoing time. By contrast, the log rank test has optimal power to detect differences in event rates if those differences are manifest during all the follow-up period.

Example: lung cancer study [1]

Battarcharjee et al [2] studied the association of gene expression on survival of lung cancer patients. 125 individuals were followed-up from diagnosis until death or last contact date. The most important predictor of survival is tumor stage, a variable combining tumor grading, nodal status and presence or absence of metastases into a ‘high risk’ and a ‘low risk’ classification. The Kaplan-Meier estimates of these two groups are as follows:



Percentiles

stage	25,0%		50,0%		75,0%	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
1	.	.	71,500	13,233	25,300	6,878
2	86,700	17,626	22,300	3,344	9,400	3,296
Overall	.	.	48,800	10,925	20,500	3,044

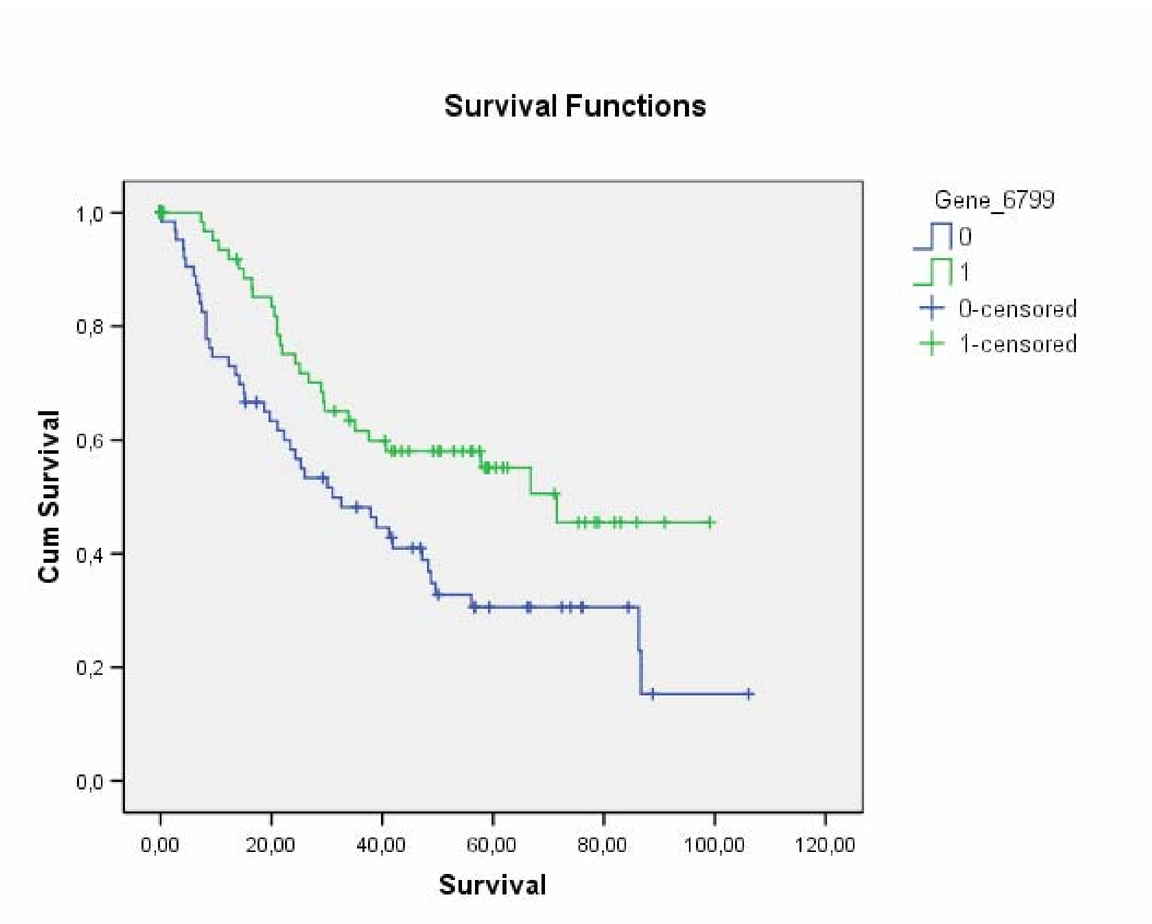
Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	11,229	1	,001
Breslow (Generalized Wilcoxon)	12,389	1	,000

Test of equality of survival distributions for the different levels of stage.

We see that survival in the high risk group (stage 2) is much worse than in the low risk group (stage 1). The median survival time is 22.3 months in stage 2, compared to 71.5 in stage 1. Both tests confirm the importance of stage for survival after diagnosis.

The following plots show Kaplan-Meier curves for groups defined by the expression of particular genes. The groups are defined by expression below or above the median expression of that gene in the sample.

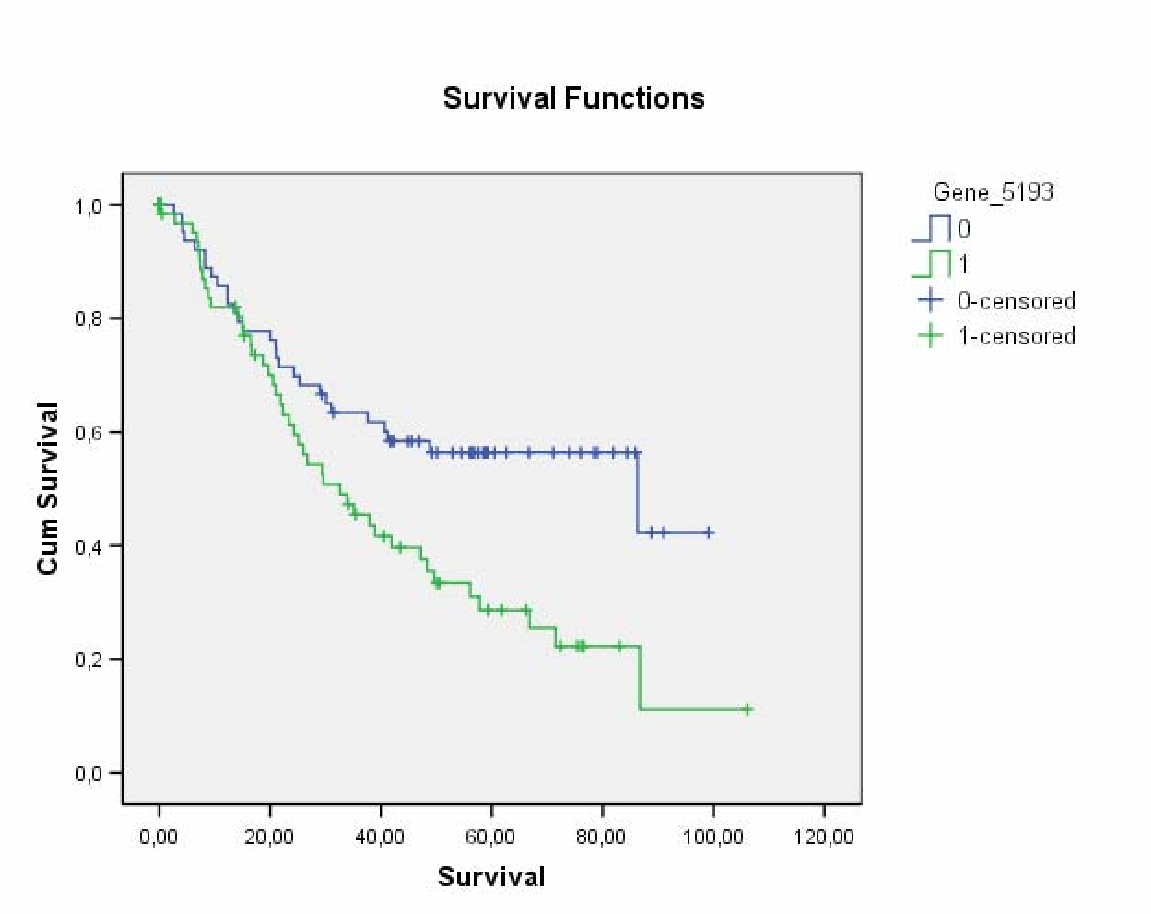


Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	6,889	1	,009
Breslow (Generalized Wilcoxon)	7,628	1	,006

Test of equality of survival distributions for the different levels of Gene_6799.

For this gene, both tests are significant. We see a constant gap between the curves, i. e. the patients with gene expression below the median (group 0) have constantly higher risk to die than patients with high gene expression.

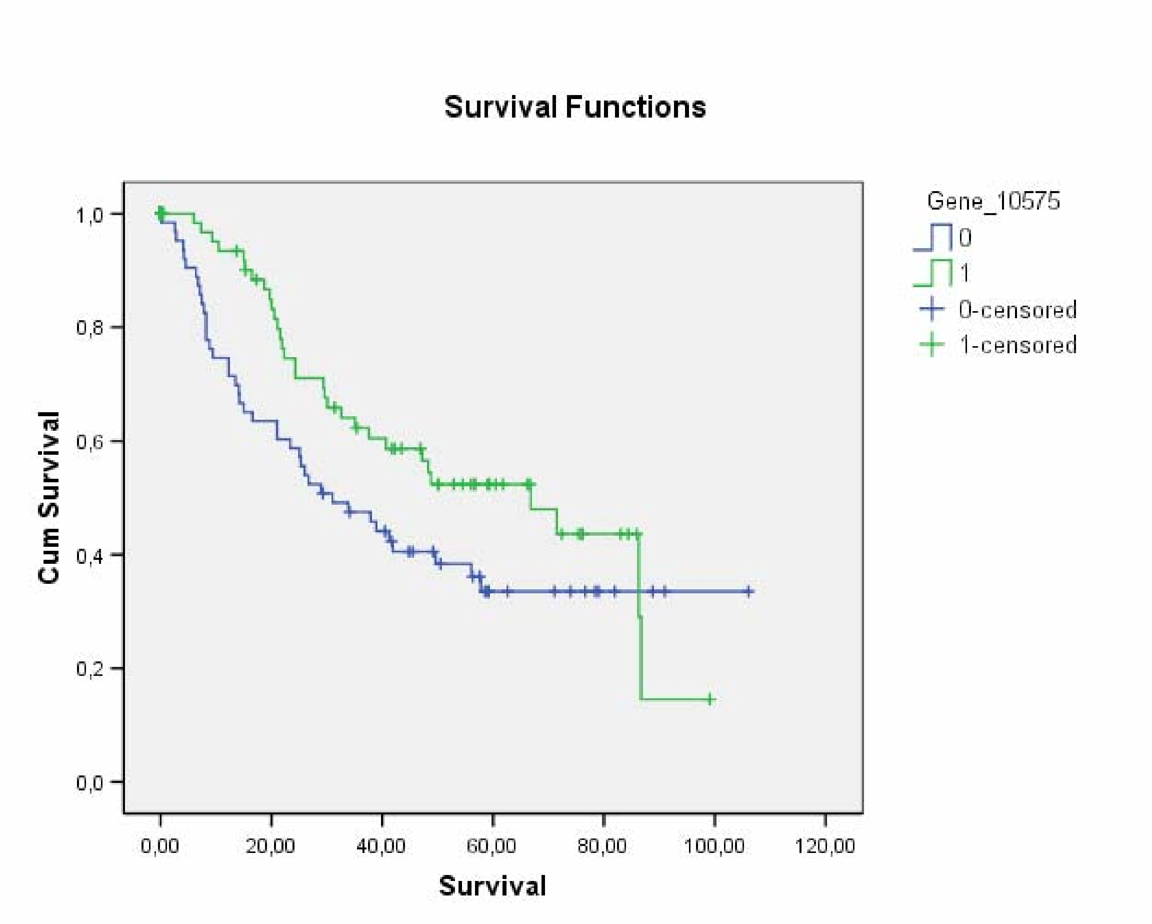


Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	7,323	1	,007
Breslow (Generalized Wilcoxon)	3,684	1	,055

Test of equality of survival distributions for the different levels of Gene_5193.

Here we see that the two groups have equal survival during the first year, but then patients with expression values above the median do worse than the other group. This is reflected by an insignificant Wilcoxon test (which emphasizes differences at early times), but a significant log rank test, which assigns equal weight to all death times.



Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	3,241	1	,072
Breslow (Generalized Wilcoxon)	6,699	1	,010

Test of equality of survival distributions for the different levels of Gene_10575.

For this gene, we see marked differences during the first 18 months, then the curves run in parallel until surviving patients of the low expression group make up the differences from 60 months on. The Wilcoxon test, which puts more weight on early times, yields a significant p-value, while the log rank test does not.

Since discrepancies or agreement of the test results may provide further insight in survival mechanisms, both tests should be carried out and reported. As a matter of course, not only p-values should be reported; they should rather accompany the description of the survival time distribution, which is best done by a table of median survival times and their 25th and 75th percentiles. Alternatively, one may also present survival rates at particular time points. For example, the 3-year survival rates (standard errors) are

71.5% (5.3%) for stage 1 patients and
34.6% (7.9%) for stage 2 patients.

Class 6: Cox regression

Basics

So far, we have tested differences in survival curves by using log rank or generalized Wilcoxon tests. These tests (in particular, the log rank test) can also be extended by using a regression model instead, the so-called Cox proportional hazards regression model [3]. The Cox regression model is a semi-parametric model, i. e., it imposes no assumptions about the distribution of survival times (e. g., exponential or normal distribution). On the other hand, it cannot be used to predict survival times for individuals.

Consider the contingency table at t(j) we have already been working with:

Group	Died at t(j)	Did not die at tj	Total
1	2	38	40
2	8	52	60
Total	10	90	100

Cox regression extends the group comparison which is performed by the log rank test to estimating a group difference. The group difference is quantified by an estimate of relative risk.

Considering the table shown above, the relative risk of group 2 vs group1 is 8/60 / 2/40 = 2.67.

Cox regression combines the information of all tables that correspond to all distinct event times. The final relative risk estimate is that one that best explains the observed group differences in the data (maximum likelihood principle).

Technically, the following equation is used to model the logarithm of a relative risk (please note that big X denotes a variable, and little x a particular value of that variable):

$$\mathrm{Log}(R(X=x) / R(X=0)) \quad = \quad B * x$$

Thus, Cox regression can easily be extended to a scale variable X.

If we compare two subjects which differ in X (say, systolic blood pressure) by 5 units, we have

Log(RR) = b * 5

or

RR = exp(b * 5)

If we compare two subjects coming from two different groups (coded as 1 and 0), we have

Log(RR) = b * 1

or

RR = exp(b)

As an example, consider the variable stage of the lung cancer data set. Requesting a Cox regression analysis, we obtain the following table:

Variables in the Equation

							95,0% CI for Exp(B)	
	B	SE	Wald	Df	Sig.	Exp(B)	Lower	Upper
stage	,847	,260	10,584	1	,001	2,333	1,400	3,885

The program automatically supplies the value for Exp(B), which corresponds to the relative risk referring to a comparison of high risk to low risk patients. We learn that the relative risk for high stage patients is 2.3 compared to low stage patients. The program also supplies a 95% confidence interval for this estimate.

The relative risk estimates may assume only positive values. A relative risk estimate of 1 means that the groups to be compared don't differ in terms of survival. A relative risk estimate <1 means that a subject with a higher value for X has lower risk than a subject with a lower value for X. The reverse applies if the relative risk estimate is >1.

While the relative risk estimate is the most intuitive result of a Cox regression analysis, the regression coefficient B, which equals the logarithm of the relative risk, is used for testing and for confidence interval estimation, because its distribution is symmetric and approximately normal (it may assume values between minus and plus infinity).

The program first computes a standard error (SE) for the regression coefficient. Assuming a normal distribution (which is reasonable in moderate-sized samples), we obtain a 95% confidence interval for the regression coefficient as

B – 1.96 SE, B + 1.96 SE

Transforming this confidence interval to the exponential scale, we obtain a 95% confidence interval for the relative risk:

exp(B – 1.96 SE), exp(B + 1.96 SE)

This confidence interval is given in the output of the program.

The so-called Wald statistics (named after a famous statistician) is computed as

Wald = $(B/SE)^2$

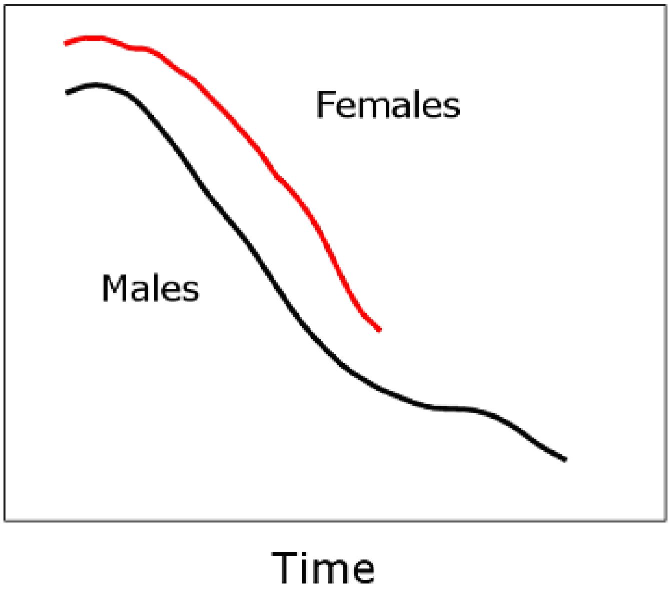
It is approximately chi-squared distributed with one degree of freedom, if the null hypothesis B=0 applies. If not, then high values for the Wald statistic are implausible, as indicated by small P-values (Sig.). In our example, we observe a P-value of 0.001 which means that the data is not plausible given that null hypothesis. As a consequence, we conclude that the stage is indeed an important predictor of survival.

Previous models that have been dealt with (linear regression, logistic regression) can be used to predict the outcome of future patients.

From the results of Cox regression, it is not straightforward to predict the death times of patients. Cox regression places no assumptions about the distribution of death time. In linear regression, we assume that at least the residuals should be normally distributed. Since we are not predicting death times here, we also do not compute residuals (at least not in the usual sense as observed minus expected outcome).

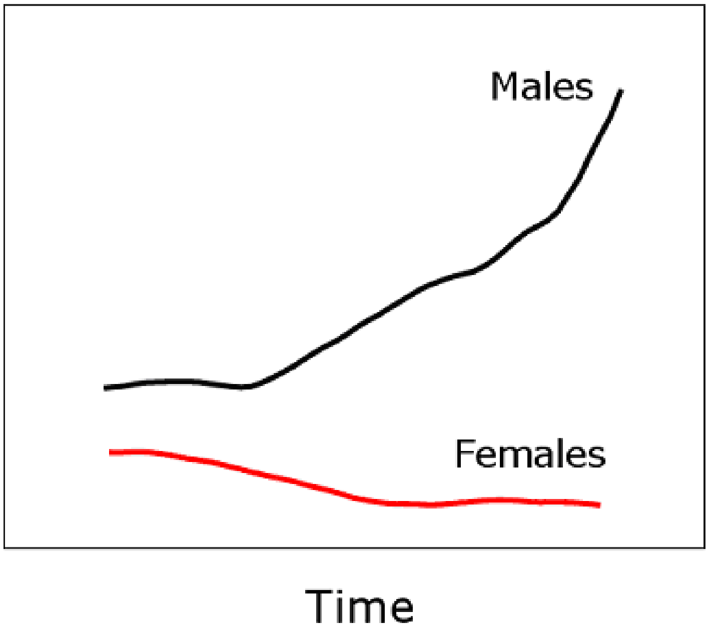
Assumptions

The most important assumption of the Cox model is the so-called proportional hazards assumption. Hazard means the instantaneous risk to die, i.e. the probability to die within the next minute. (Therefore, the relative risk estimated by Cox's model is often denoted by hazard ratio, although both terms mean the same.) 'Proportional hazards' means, that the ratio between the hazards of two patient groups remains constant over the complete follow-up period. Consider the following example, depicting follow-up time on the X-axis and hazard on the Y-axis:



In this example we see that although the hazard (the probability to die within the next minute) decreases with time, the distance between the groups remains constant. Thus, the proportional hazards assumption is fulfilled here.

Another example:



In this example, the hazard for females remains constant, but the hazard for males increases with time. Thus, the proportional hazards assumption is not justified in this example.

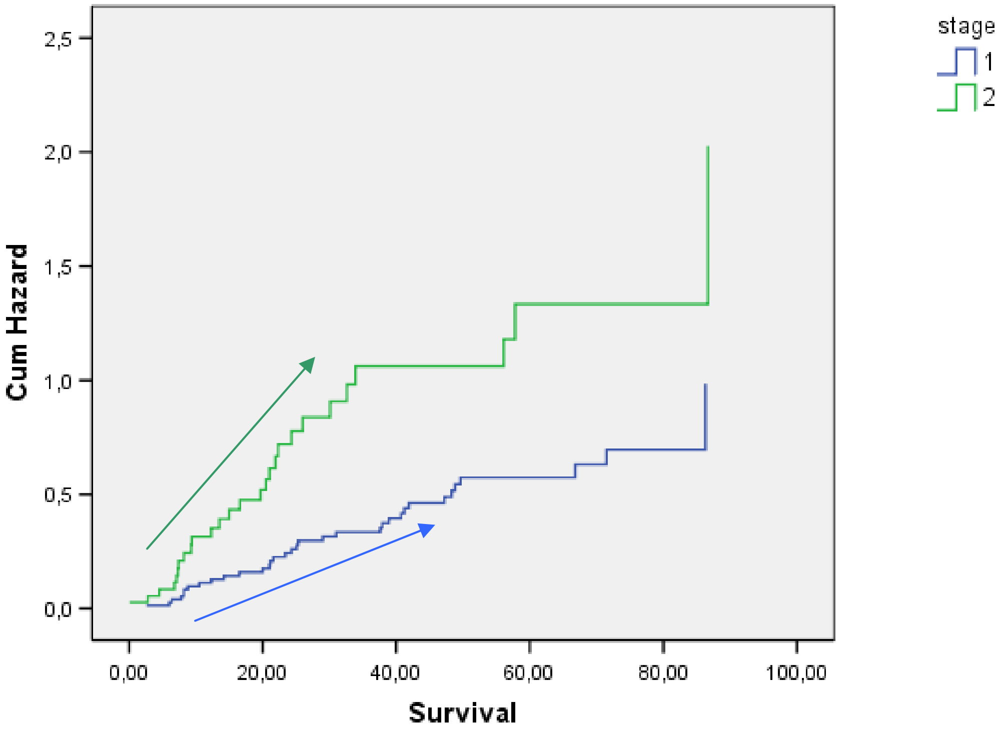
Why is the proportional hazards assumption so important?

Since in Cox regression we use the information from all tables evaluated at all distinct death times, and compute one relative risk (from now on, let’s use the more correct denomination ‘hazard ratio’) estimate which should apply to all death times, we must make sure that this simplification is really justified. It is of course only justified, if the group difference remains constant over the whole range of follow-up time we are dealing with.

How can we verify the proportional hazards assumption?

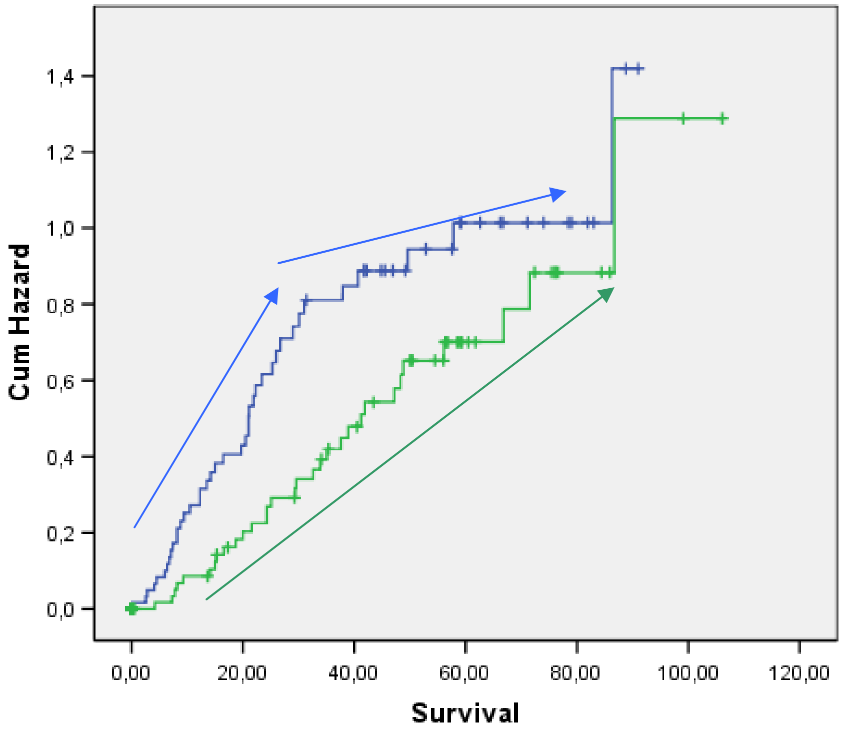
In small to moderate-sized samples, it is difficult to compute a graph showing the hazard in two groups as a function of time as shown above (because deaths occur rather rarely). It is, however, possible to plot the *cumulative* hazard in two groups.

Such a plot of the cumulative hazard is shown below:



Proportional hazards hold if we observe a picture similarly to the one above: The cumulative hazard increases in both groups (it actually cannot decrease). The rate of increase is sometimes higher, sometimes lower. However, comparing the rates between the groups, we see that they are more or less proportional.

A plot of non-proportional cumulative hazards could resemble the following:



Here we see that first the group of patients with low gene expression (group 0, blue, upper line) has higher risk, after twenty months suddenly the cumulative hazard in this group grows at much lower rate, while the cumulative hazard in the high gene expression (group 1, green, lower line) increases constantly.

Computing one relative risk estimate for all the range would over-simplify the situation:

Variables in the Equation

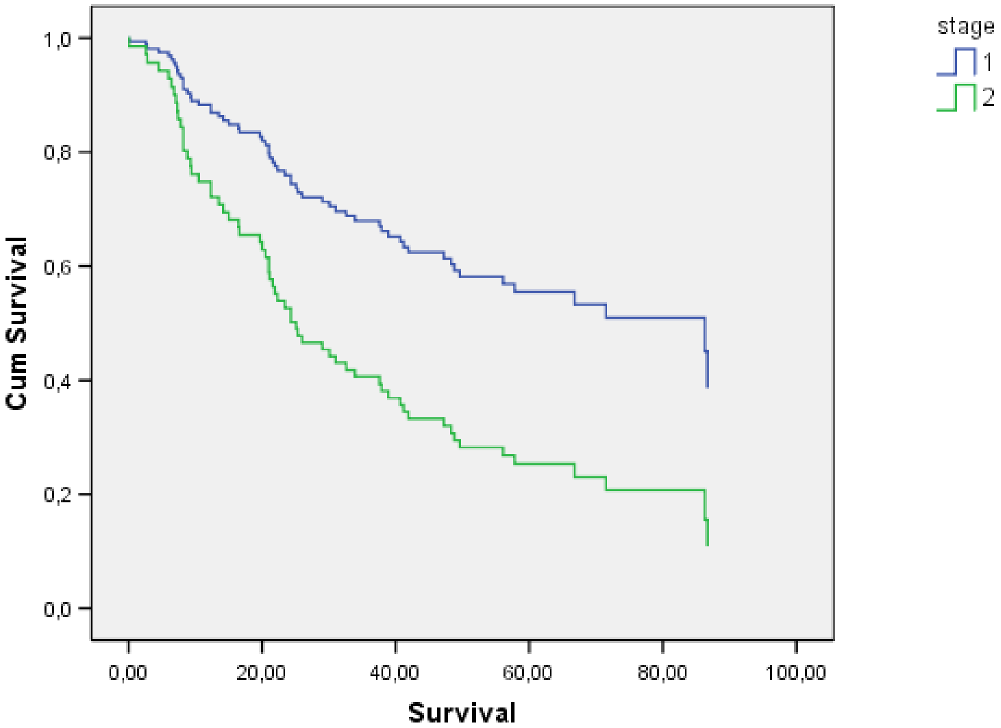
	B	SE	Wald	df	Sig.	Exp(B)
Gene_1791	-,454	,240	3,586	1	,058	,635

Although in total, we see a positive effect of high gene expression, from the cumulative hazards plot we must assume that the benefit of the high expression group only last until about 24 months. We must assume a much larger effect during the first 24 months, and almost an inversion of effect after that time point.

The proportional hazards assumption will be further dealt with in Lec 8.

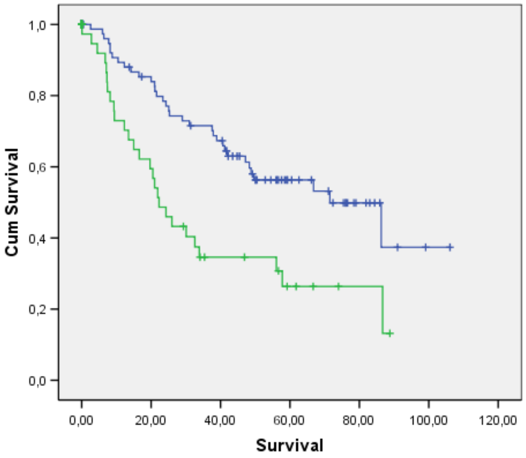
Estimates derived from the model

As has been already mentioned, the Cox model does not provide estimates of individual survival times. However, one can obtain survival curves (similar to the Kaplan-Meier method) which are based on the group difference estimated by the model. In order to compute these survival curves, the programs makes use of the baseline survival function, which is basically the Kaplan-Meier estimate computed from the total sample, and separates the survival curves for two groups following the estimated relative risk between those groups:



Comparing these model-based survival curves to Kaplan-Meier estimates, we see that the model-based curves assume perfect proportional hazards. The constant risk ratio between the groups leads to constantly increasing differences between the survival curves, which cumulate the death hazard over time.

By contrast, the Kaplan-Meier method estimates survival curves separately for each group, without imposing any assumption on proportionality of hazards. Thus, the distance between the curves may decrease and extend.



Relationship Cox regression – log rank test

The Cox regression model is closely related to the log rank test comparing two groups. The p-value obtained by the log rank test is approximately equal to the p-value corresponding to a binary independent variable in Cox regression, as the following comparison shows:

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	11,229	1	,001
Breslow (Generalized Wilcoxon)	12,389	1	,000

Test of equality of survival distributions for the different levels of stage.

Variables in the Equation

							95,0% CI for Exp(B)	
	B	SE	Wald	df	Sig.	Exp(B)	Lower	Upper
stage	,847	,260	10,584	1	,001	2,333	1,400	3,885

This close relationship is the reason why the log-rank test is also labeled “Mantel-Cox” test.

Class 7: Multivariable Cox regression

The equation for Cox regression can easily be extended to multiple covariates:

Log(R(X=x, Z=z) / R(X=0, Z=0)) = b1 * x + b2 * z

The regression coefficients have the same interpretation as in univariable Cox regression. Using multivariable Cox regression, we are now able to adjust effects for others:

Variables in the Equation

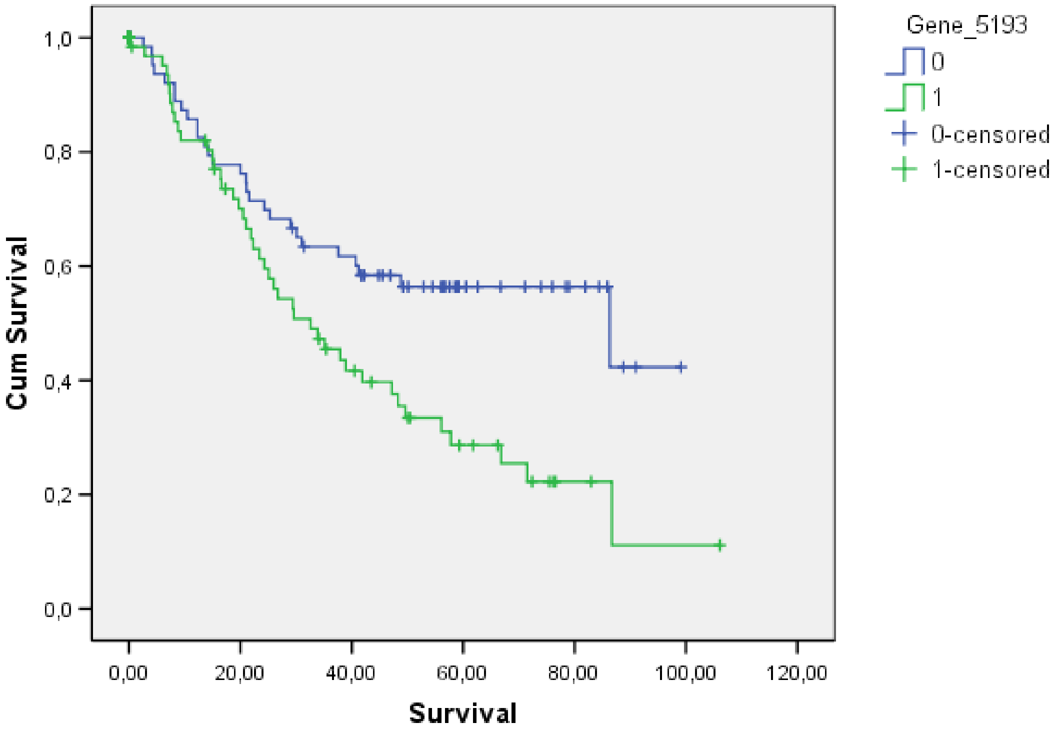
	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
stage	,927	,262	12,478	1	,000	2,526	1,510	4,223
nikotin	,162	,311	,273	1	,601	1,176	,639	2,164
gender	,343	,269	1,622	1	,203	1,409	,831	2,387
Age	,014	,014	,960	1	,327	1,014	,986	1,042

The numbers shown in column ‘Exp(B)’ are now adjusted relative risk (hazard ratio) estimates; each effect is adjusted for any other effect in the model.

The inclusion of nikotin, gender and age doesn’t change our conclusions about the importance of the factor stage.

The same model building strategies that have already been discussed also apply to multivariable Cox regression. As a rule of thumb, we must keep in mind that the number of candidate variables for model building must not exceed one tenth of the number of events (deaths).

Suppose a candidate marker ‘Gene 5193’ should be evaluated in its ability to predict survival of lung cancer patients. For simplicity, the expression values of this marker again have been categorized (1: above median, 0: below median). The marker proves satisfactory in univariable analyses:



Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	7,323	1	,007
Breslow (Generalized Wilcoxon)	3,684	1	,055

Test of equality of survival distributions for the different levels of Gene_5193.

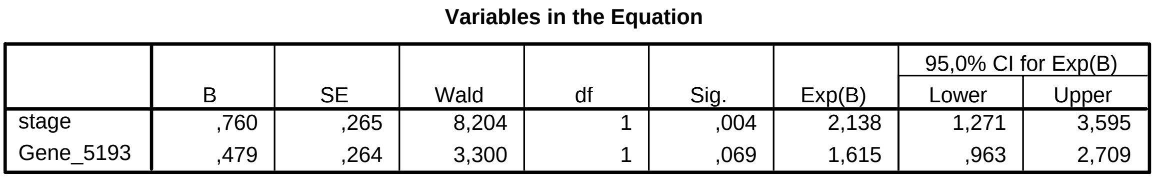
Results of univariable Cox regression:

Variables in the Equation

							95,0% CI for Exp(B)	
	B	SE	Wald	df	Sig.	Exp(B)	Lower	Upper
Gene_5193	,648	,244	7,063	1	,008	1,911	1,185	3,081

These results suggest that Gene 5193 is a risk factor: patients with higher gene expression have a 1.9fold higher risk than patients with lower gene expression.

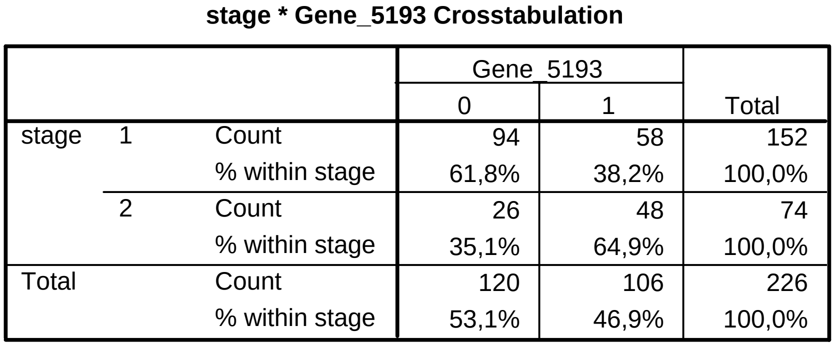
However, a marker is only important if it adds information to the survival prediction based on clinical (histopathological) evaluation of a patient. Therefore, we should consider adjustment at least for the most important factor: stage.



Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
stage	,760	,265	8,204	1	,004	2,138	1,271	3,595
Gene_5193	,479	,264	3,300	1	,069	1,615	,963	2,709

Adjusting for stage, we learn that the hazard ratio drops from 1.9 to 1.6, and the p-value now exceeds the significance level of 5%. Therefore, we cannot prove that Gene 5193 is an independent predictor of patient survival. The drop in hazard ratio comparing univariable to multivariable analyses is a result of the correlation between stage and Gene 5193:



stage * Gene_5193 Crosstabulation

			Gene_5193		Total
			0	1	
stage	1	Count	94	58	152
		% within stage	61,8%	38,2%	100,0%
	2	Count	26	48	74
		% within stage	35,1%	64,9%	100,0%
Total		Count	120	106	226
		% within stage	53,1%	46,9%	100,0%

While only 38.2% of the stage 1 patients have gene expression above the median, the corresponding number for stage 2 patients is 64.9%. The chi-square test yields a p-value < 0.001.

Similarly as in other regression models (linear or logistic), positive correlation between variables leads to a diminished effect in multivariable regression. The reverse happens if variables are negatively correlated.

With scale independent variables, we must verify the linearity assumption by considering nonlinear effects. The additivity assumption can be checked by screening interactions. These assessments of model assumptions can be carried out in the same way as has been demonstrated for linear or logistic regression, and is therefore not considered again in this text.

Of special importance in Cox regression is the assessment of the proportional hazards assumption. It will be dealt with in detail in Lec 8.

Stratification: another way to address confounding

Cox regression provides an alternative way to adjust for confounding variables, namely stratification. Stratification does not mean that one reports results from subgroups.

Rather, the data are grouped into strata that are defined by different levels of stratification variables, and then an estimate of the hazard ratio is computed that best fits the data in all strata. Note that for stratification factors, no hazard ratios are estimated.

- Stratification can be used if for a variable non-proportional hazards are detected. Note, however, that the effect of such variables cannot be quantified if they are used as stratification variables.
- Only nominal or ordinal variables can be used as stratification variables, as stratification by a scale variable would assign each individual to a separate subgroup, making any estimation impossible. Scale variables can only be used for stratification if they are grouped into few categories prior to estimation of a Cox model.

As an example, compare the following analyses of Gene_5193, adjusted for stage:

Adjusting for stage in multivariable estimation:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
Gene_5193	,479	,264	3,300	1	,069	1,615
stage	,760	,265	8,204	1	,004	2,138

Adjusting for stage by stratification:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
Gene_5193	,443	,266	2,781	1	,095	1,558

There will always be differences between the results obtained by stratification and those obtained by multivariable estimation. These differences result from the relaxed assumption of proportional hazards for variable stage in analysis by stratification. The differences will be negligible if the proportional hazards assumption holds well for the stratification variable, and will be larger if hazards are non-proportional between the different levels of the stratification variable Therefore, stratification can be used if we want to adjust for a variable that

- exhibits non-proportional hazards
- and is not of interest by itself

Class 8: Assessing model assumptions

Proportional hazards assumption

We have already defined ‘proportional hazards’ as the situation where the ratio between the hazards of two patient groups remains constant over the complete follow-up period.

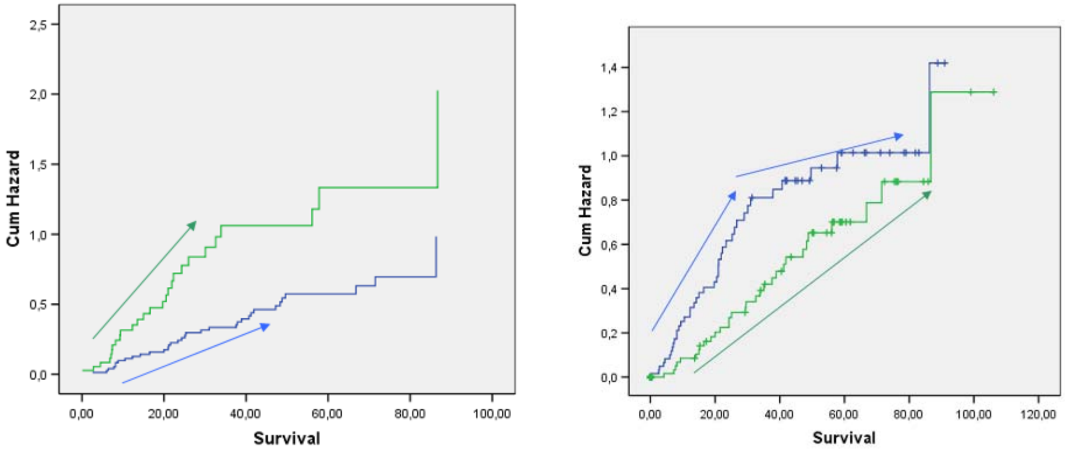
Since Cox hazard ratio estimates crucially depend on the validity of the proportional hazards assumption, it is necessary that this assumption is verified in a Cox model. As with assumption checking in other regression models, there are two options:

- Graphical checks
- Statistical testing of violations of the proportional hazards assumption

Graphical checks of the PH assumption

Cumulative hazards plots

Comparing the cumulative hazard between two groups has already been outlined in Lec 6. These plots show the increasing cumulative risk as function of follow-up time. Therefore, assuming proportional hazards, we should observe two lines which diverge with increasing follow-up time. Proportional hazards approximately hold for variable stage (as shown in the left panel). If the proportional hazards assumption does *not* hold, there is no constant divergence; periods of divergence may be followed by periods of convergence, and the cumulative hazards may even cross. Such a situation is shown in the right panel of the graph below.

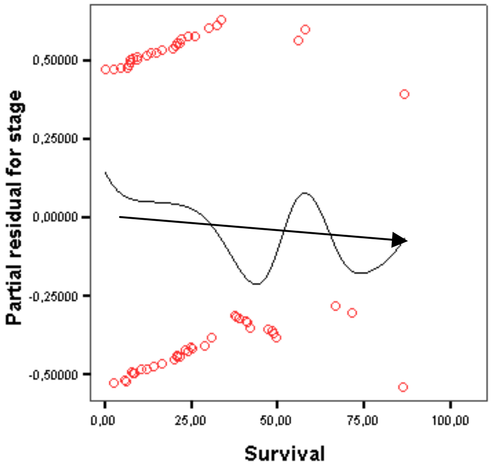


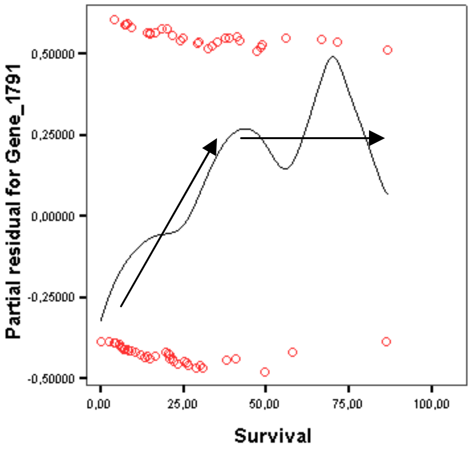
While we can use these plots to judge upon the proportional hazards assumption in univariable models, it is not straightforward to apply these plots to multivariable models, as the cumulative hazard may be confounded by other variables.

Schoenfeld (partial) residuals

So-called partial residuals [4] can be used to assess the proportional hazards assumption in multivariable or univariable models. These residuals are frequently denoted by the name of their inventor, D. Schoenfeld. We can compute partial residuals for each variable in the model, hence the name ‘partial’.

In brief, one such residual is computed at each death time, and it is basically defined as the difference between the observed covariate value of the subject that failed (experienced the event) at that time, and the covariate value of the average individual that would be expected to fail from the estimated Cox model. If the partial residuals are on average positive at early times and negative at later times, we can conclude that the hazard ratio decreases with time, meaning that hazards are non-proportional. Sometimes, it is useful to insert a line of moving averages (a so-called *smoother*) into the partial residuals plots, which allows an easier interpretation. The two plots below show the partial residuals corresponding to the cumulative hazards plots shown above:





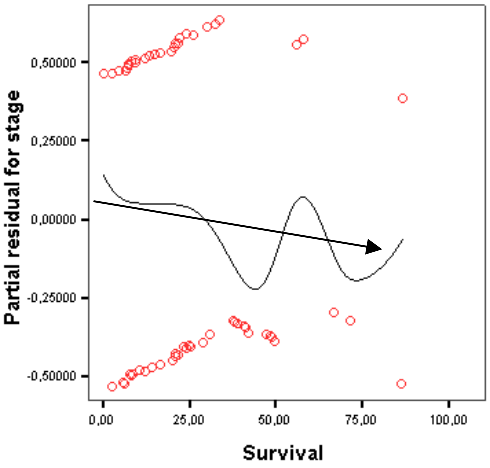
While we see an almost constant smoother line for variable stage, there is a sharp increase during the first 24 months in the residuals for Gene_1791, followed by a constant period. (Occasional fluctuations that result from the small number of patients at risk after 60 months, say, should not be over-interpreted.)

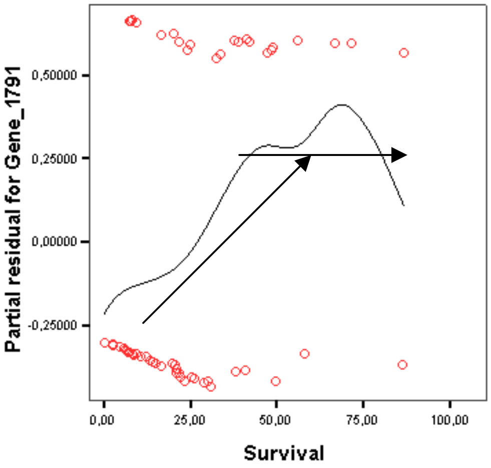
The most important advantage of partial residuals is their use in multivariable modeling. The non-proportionality of hazards observed in Gene_1791 could disappear if we adjust for other variables. In other situations, non-proportional hazards appear only in multivariable modeling, but are not present in a univariate model. If a Cox model is fit including the variables stage and Gene_1791, we obtain the following table of estimated and hazard ratios:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
stage	,788	,262	9,071	1	,003	2,199	1,317	3,673
Gene_1791	-,539	,267	4,087	1	,043	,583	,346	,984

Plotting the partial residuals, we obtain the following graphs:





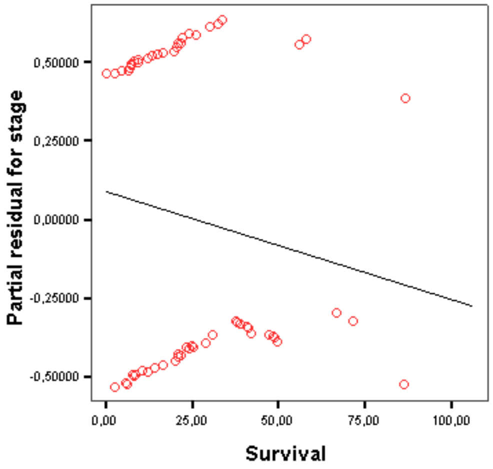
Now we observe, as before, a slight decrease in the residuals of stage, and the same, even more unique, picture for Gene_1791.

Testing violations of the PH assumption

There are two ways to formally test the proportional hazards assumption, leading to approximately the same results:

- Testing the slope of partial residuals
- Testing an interaction of covariate with time

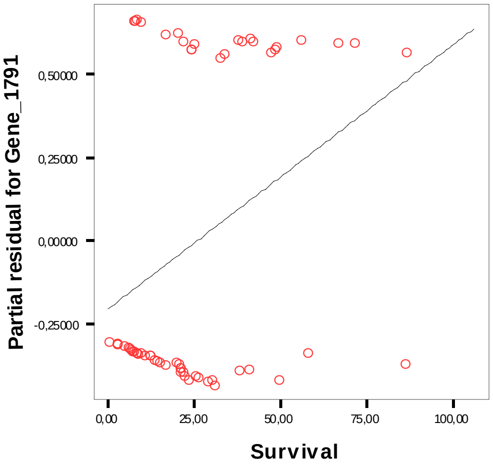
If a linear regression model is computed, using the partial residuals as dependent variable and time (or log of time) as independent variable, then we are able to test whether the partial residuals change (linearly) with time. This test can be used to assess the proportional hazards assumption:



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,089	,100		,886	,379
	Survival	-,003	,003	-,144	-1,121	,267

a. Dependent Variable: Partial residual for stage



Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-,205	,094		-2,178	,033
	Survival	,008	,003	,338	2,754	,008

a. Dependent Variable: Partial residual for Gene_1791

Alternatively, one may fit a Cox model with time-dependent effects. In this model we relax the assumption that a hazard ratio remains constant over the whole follow-up period, and we estimate a possible time-dependency of the hazard ratio. Technically, we include an interaction of the variable of interest with time in the model. If this interaction is significant, then we may conclude that the effect of that variable depends on follow-up time. The simplest setting is obtained if the hazard ratio is allowed to change linearly with time, as in the following example:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
stage	1,207	,429	7,924	1	,005	3,345
T_COV_*stage	-,015	,014	1,101	1	,294	,986

In the table shown above, we assume now the following relationship between the effect of stage and time: B = 1.207 – 0.015*time. This means, that there is a small and not significant decline in the effect of stage. At beginning of follow-up, the effect of stage is 1.207 (corresponding to a hazard ratio of 3.345). With every month, it declines by 0.015 (or in terms of hazard ratios, the hazard ratio reduces by 1.4%).

If we assume a linear relationship of the effect of Gene 1791 and time, we obtain the following Cox model:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
Gene_1791	-1,258	,440	8,183	1	,004	,284
Gene_1791*T_COV_	,032	,015	4,786	1	,029	1,033

At beginning, high expression of Gene 1791 has a protective effect. The B estimate is -1.258, corresponding to a hazard ratio of 0.28. With every month, the effect decreases by 0.032 (in terms of hazard ratios: the initial hazard ratio 0.28 multiplies by 1.033 every month). After roughly 40 months, we already see a hazard ratio of 1. Afterwards, high expression has a negative effect on the death hazard.

In summary, both tests reveal a significant time-dependency of the hazard ratio of Gene 1791 (p=0.008 and 0.029 for the slope-of-partial-residuals test and the interaction-with-time test). On the other hand, based on these tests we would not conclude that the proportional hazards assumption is violated for stage (p=0.267 and p=0.294, respectively).

What to do if PH assumption is violated?

There are some options how to proceed if the proportional hazards assumption is violated in a model.

Ignoring the non-proportionality

If non-proportionality of hazards is ignored in a model, then we obtain an average estimate of the hazard ratio which summarizes the time-dependency into one number. During some periods, this number will overestimate the true effect, for others it will underestimate the true effect. This option will be chosen if for a variable a time-dependent effect has been detected, but the direction of the effect remains the same throughout the follow-up period (either positive or negative). In case of Gene 1791, we learned that before 40 months, high expression has a protective effect, and afterwards, high expression is associated with higher risk. Averaging the effect over the time axis still yields a significant average hazard ratio of 0.58, because more deaths occur before 40 months than afterwards.

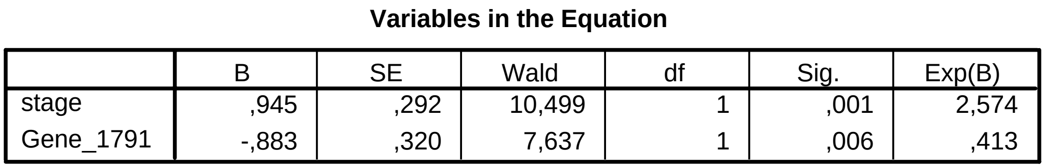
Dividing the time axis

Another option is to divide the time axis into two parts, and estimate separate hazard ratios estimates for each part. Suppose we want to divide the time axis at 40 months. For a proper subgroup analysis, we must follow the principle:

- *All subjects must enter the first subgroup*. The *censoring indicator has to be set to 'censored' for all subjects which lived longer than 40 months*, and the survival time is set to 40 months for these subjects.
- Only the subjects *who lived longer than 40 months* enter the second subgroup.

The results for our Cox model are as follows:

First subgroup (up to 40 months, N=113, 48 events):



Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
stage	,945	,292	10,499	1	,001	2,574
Gene_1791	-,883	,320	7,637	1	,006	,413

Second subgroup (longer than 40 months, N=57, 13 events):

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
stage	,169	,677	,063	1	,802	1,185
Gene_1791	,578	,608	,903	1	,342	1,783

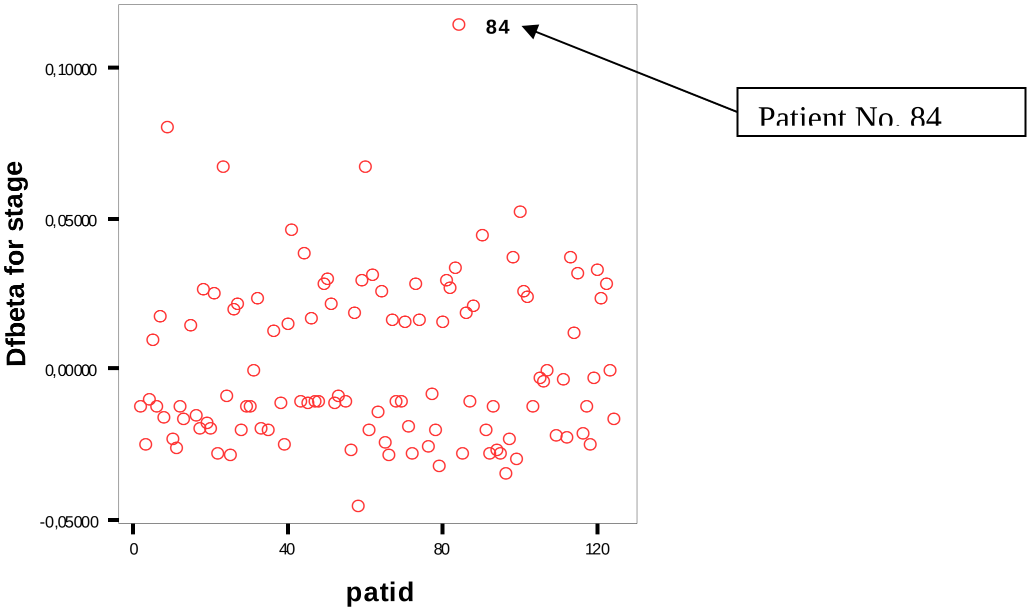
We see that the effect of stage is higher in the first subgroup than in the second one (2.57 vs. 1.19), and that the hazard ratio of Gene 1791 is 0.4 in the first subgroup, but 1.8 in the second one. Because of the small number of events in the second subgroup, we cannot statistically proof any effect of these variables after 40 months of follow-up. However, we already proved that the effect of Gene 1791 changes significantly with time. In other words, the hazard ratio of 0.413 computed for the first time period is significantly different from the hazard ratio of 1.783 corresponding to the second time period.

A third option to deal with non-proportional hazards is to model the time-dependency. This has already been done in the previous subchapter, assuming a linear relationship between the effect and time. It is also possible to assume nonlinear relationships of an effect with time, but this goes beyond the scope of this text.

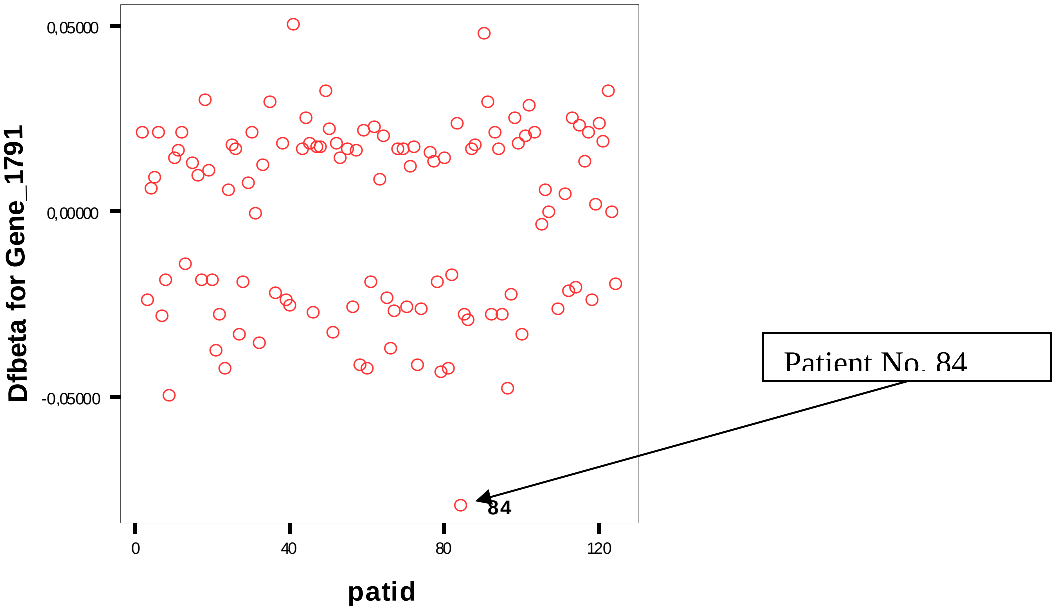
Influential observations

Influential observations can be detected by investigating the so-called DfBeta diagnostics, that can computed for each subject and covariate. This statistic measures the change in regression coefficient of the corresponding covariate, if a subject is left out from model estimation. If this value is high (positive or negative), the subject has more influence on the results than other subjects. Investigation of the characteristics of such subjects could lead to the detection of new risk factors.

Plotting the DfBeta values for stage against patient ID, we learn that patient 84 has a noticeably high value:



Interestingly, the same patient is identified as influential point when plotting the DfBeta of Gene 1791:



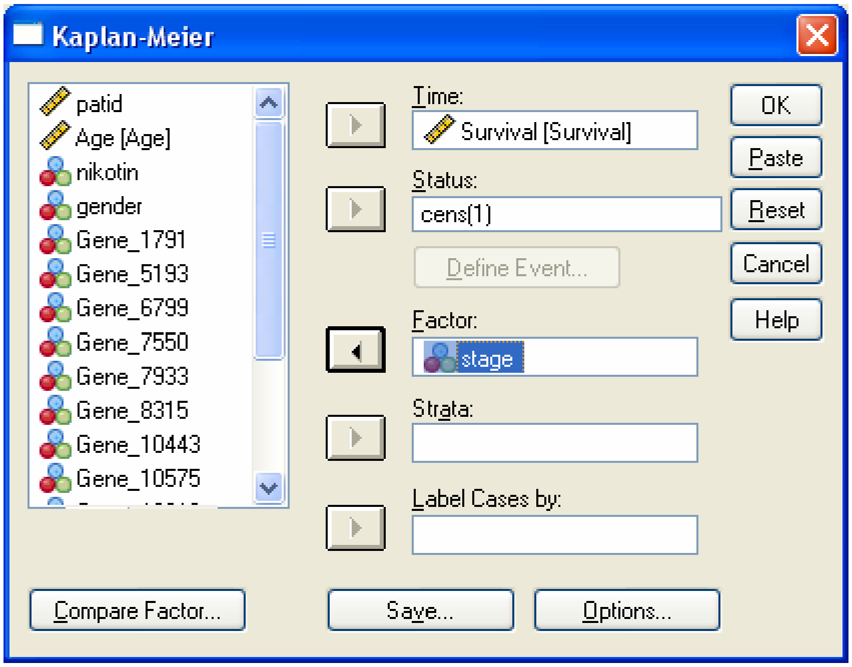
SPSS Lab 2: Analysis of survival outcomes

All methods of survival analysis are demonstrated using the data set lungcancer.sav. In this data set, we have the following variables:

Patid …	A patient identifier
Survival …	survival time in months after diagnosis of lung cancer
Age …	age of the patient at diagnosis of lung cancer
Cens …	censoring indicator (1=dead, 0=alive)
Stage …	a composite variable from TNM-stage: 1=low risk, 2=high risk
Nikotin …	smoking status at time of diagnosis (1=yes, 0=no)
Gender …	sex, 1=male, 0=female
Gene_1791, Gene_5193, Gene_6799, Gene_7550, Gene_7933, Gene_8315, Gene_10443, Gene_10575, Gene_10912, Gene_11304, Gene_11686, Gene_12519…	Expression of 12 genes, categorized into 1, expression above median, and 0, expression below median

Kaplan-Meier analysis

First we compute Kaplan-Meier curves for groups defined by stage. Call the dialogue Analyze-Survival-Kaplan-Meier… and define:



Time: Survival [Survival]

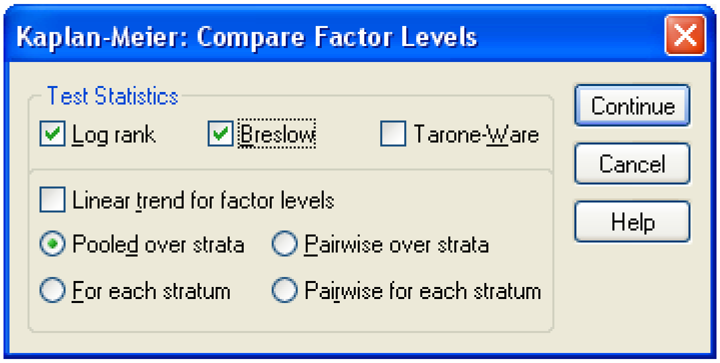
Status: cens(1)

Factor: stage

Strata:

Label Cases by:

The code ‘1’ has to be defined as the event (death). The button ‘Compare Factor…’ lets us choose tests to compare the groups defined by stage:



Test Statistics

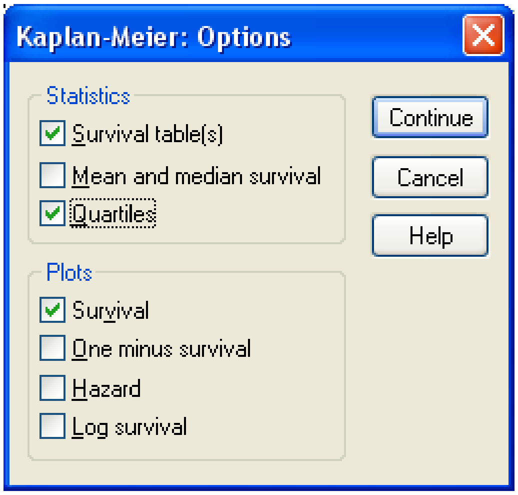
☑ Log rank ☑ Breslow □ Tarone-Ware

□ Linear trend for factor levels

● Pooled over strata ○ Pairwise over strata

○ For each stratum ○ Pairwise for each stratum

We select ‘Log rank’ and ‘Breslow’ and confirm by clicking on ‘Continue’. At the submenu ‘Options’, we request survival tables, quartiles and survival plots:



Statistics

☑ Survival table(s)

□ Mean and median survival

☑ Quartiles

Plots

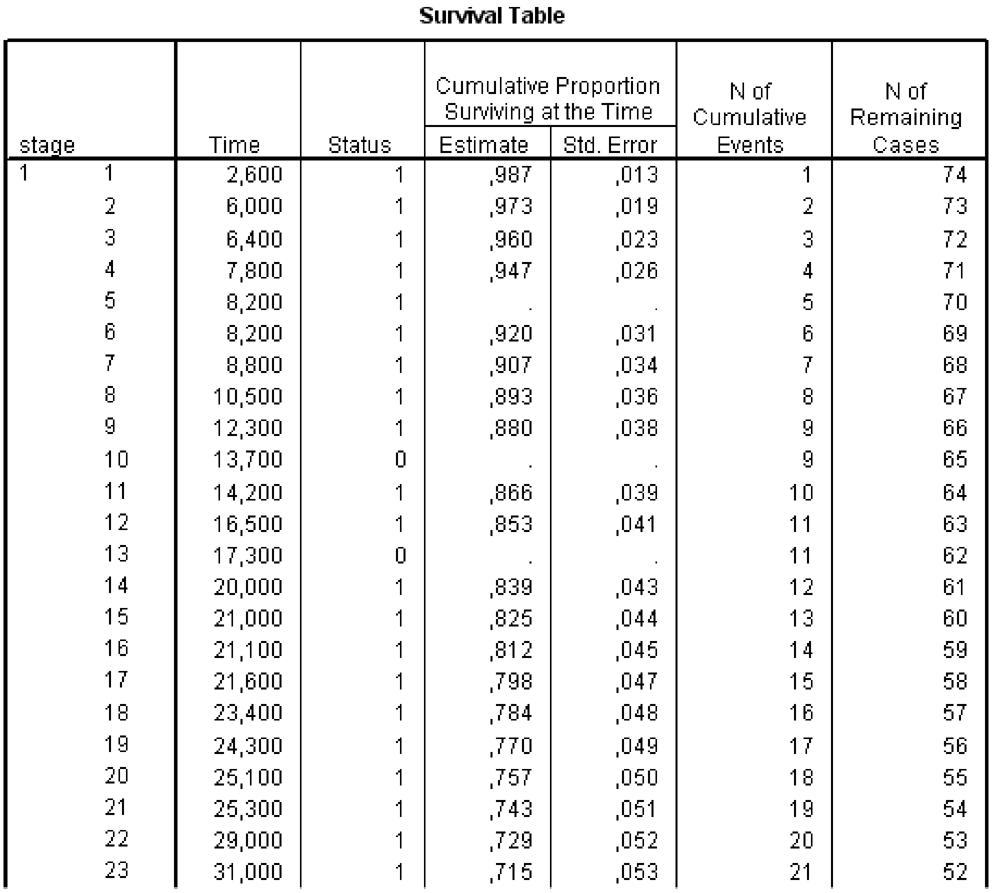
☑ Survival

□ One minus survival

□ Hazard

□ Log survival

Finally, we obtain a table with survival function estimates for each group:



Survival Table

stage		Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
				Estimate	Std. Error		
1	1	2,600	1	,987	,013	1	74
	2	6,000	1	,973	,019	2	73
	3	6,400	1	,960	,023	3	72
	4	7,800	1	,947	,026	4	71
	5	8,200	1	.	.	5	70
	6	8,200	1	,920	,031	6	69
	7	8,800	1	,907	,034	7	68
	8	10,500	1	,893	,036	8	67
	9	12,300	1	,880	,038	9	66
	10	13,700	0	.	.	9	65
	11	14,200	1	,866	,039	10	64
	12	16,500	1	,853	,041	11	63
	13	17,300	0	.	.	11	62
	14	20,000	1	,839	,043	12	61
	15	21,000	1	,825	,044	13	60
	16	21,100	1	,812	,045	14	59
	17	21,600	1	,798	,047	15	58
	18	23,400	1	,784	,048	16	57
	19	24,300	1	,770	,049	17	56
	20	25,100	1	,757	,050	18	55
	21	25,300	1	,743	,051	19	54
	22	29,000	1	,729	,052	20	53
	23	31,000	1	,715	,053	21	52

…

The subsequent table shows the median survival, and 25th and 75th percentiles in both groups and overall:

Percentiles

stage	25,0%		50,0%		75,0%	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
1	.	.	71,500	13,233	25,300	6,878
2	86,700	17,626	22,300	3,344	9,400	3,296
Overall	.	.	48,800	10,925	20,500	3,044

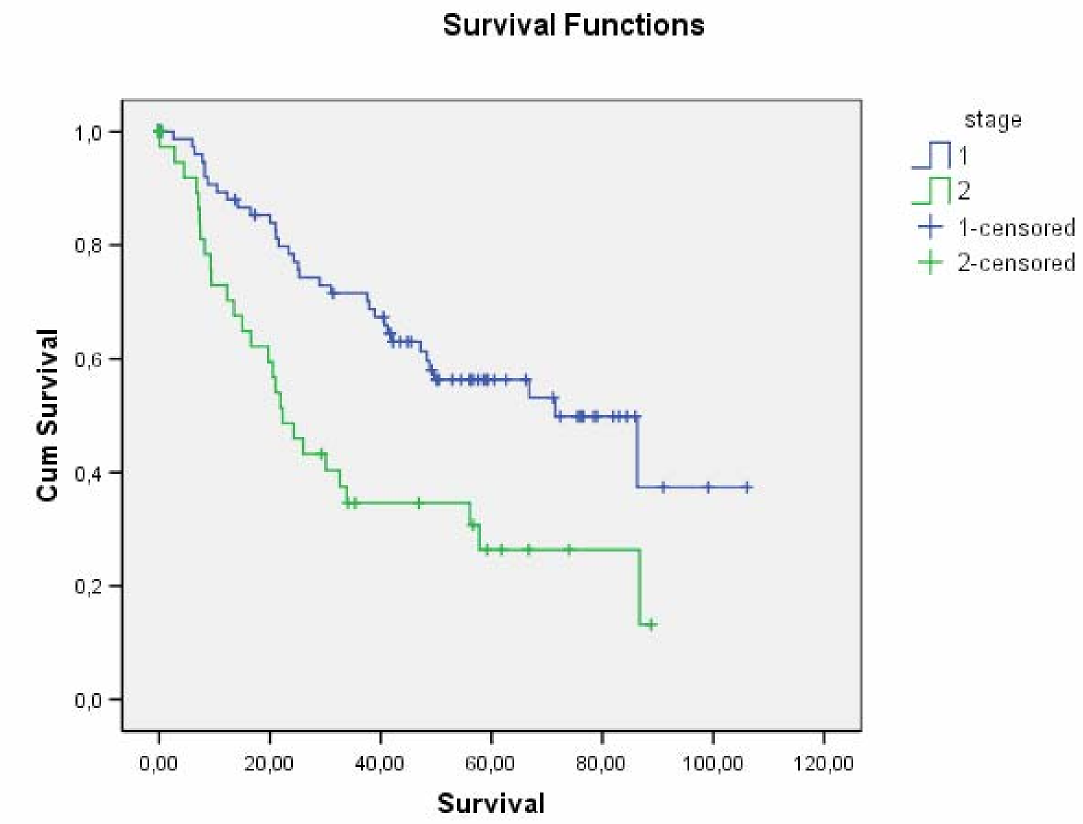
Afterwards, we obtain a table with the results of the log rank and the generalized Wilcoxon tests:

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	11,229	1	,001
Breslow (Generalized Wilcoxon)	12,389	1	,000

Test of equality of survival distributions for the different levels of stage.

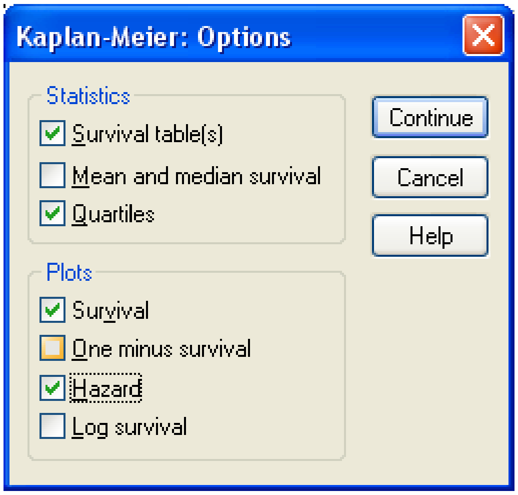
Finally, the Kaplan-Meier plot is produced:



Please note that these curves start at the shortest death time. In our analysis, the shortest death time is close enough to 0 such that the plot seems to start at 0. In some analyses however, it may be necessary to add, for each group, a 'ghost observation' with survival time and censoring indicator both set to 0 to have the Kaplan-Meier curves starting at 0.

Cumulative hazards plots

Cumulative hazards plots can be obtained by checking 'Hazard' in the options submenu:



Kaplan-Meier: Options

Statistics

☑ Survival table(s)

☐ Mean and median survival

☑ Quartiles

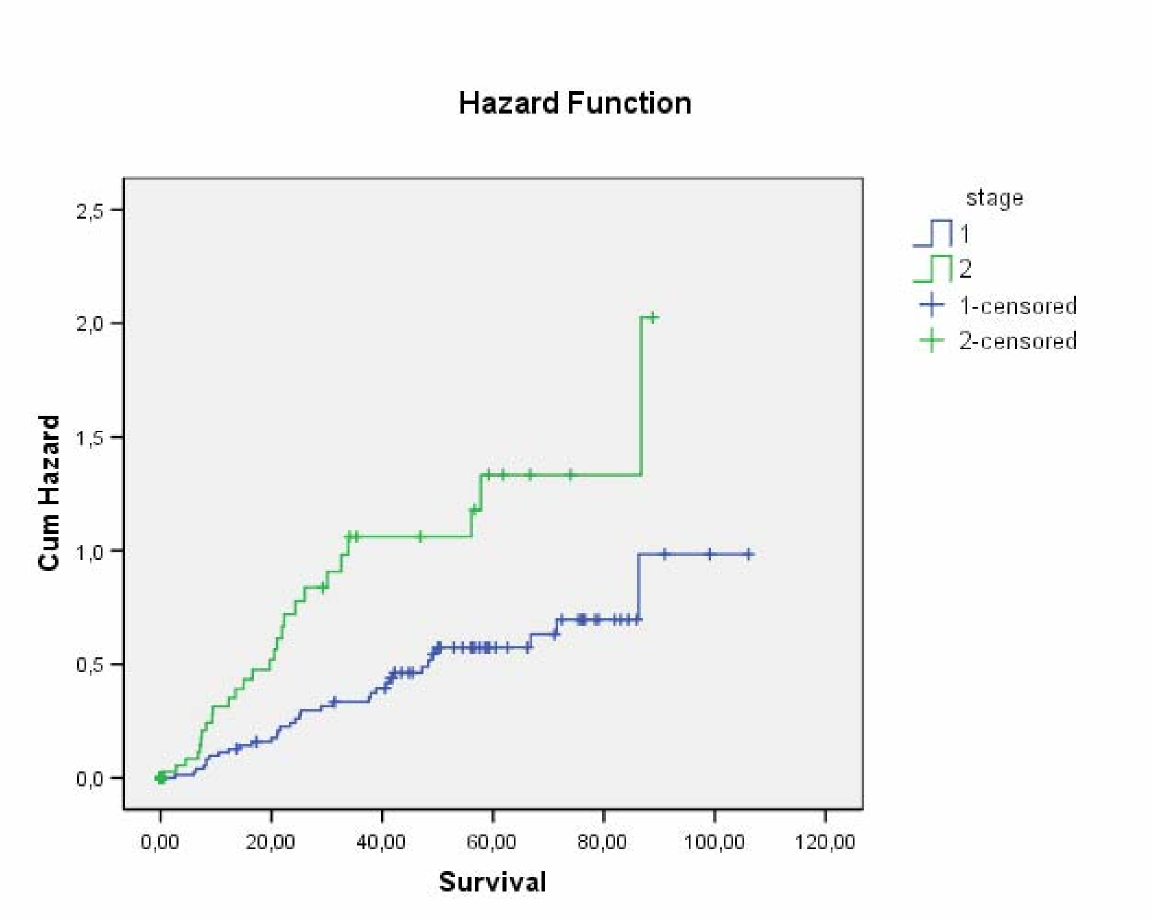
Plots

☑ Survival

☐ One minus survival

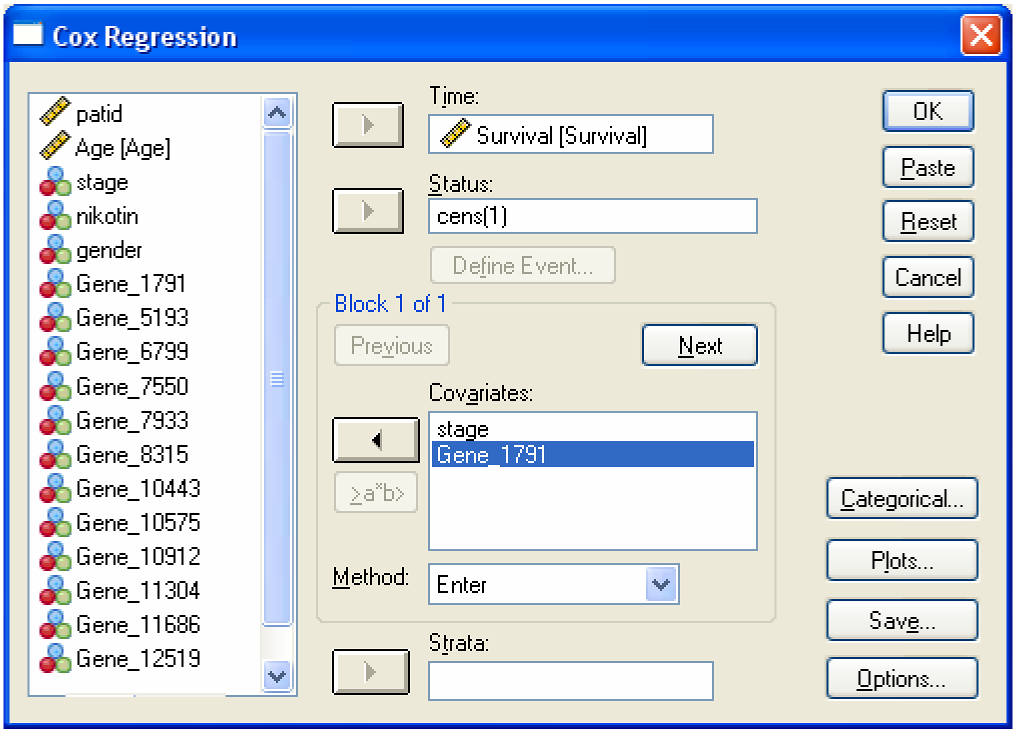
☑ Hazard

☐ Log survival



Cox regression

Cox regression is called by choosing Analyze-Survival-Cox regression…:



Time: Survival [Survival]

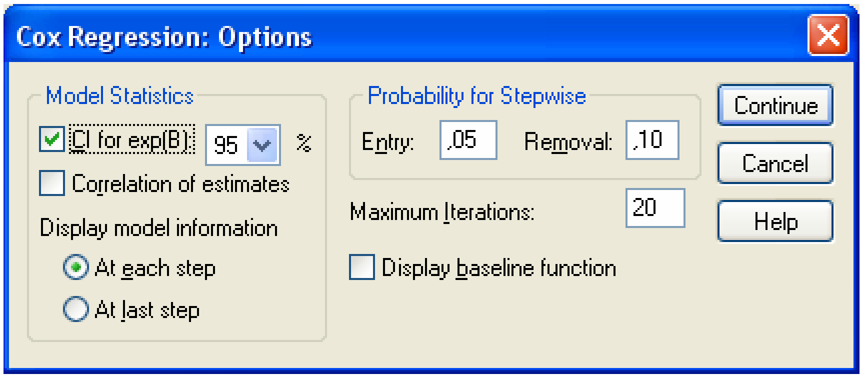
Status: cens(1)

Covariates: stage, Gene_1791

Method: Enter

Strata:

The status variable has to be defined the same way as before. In the Options subdialogue, we request 95% confidence limits for Exp(B):



☑ CI for exp(B): 95 %

□ Correlation of estimates

● At each step

○ At last step

Entry: ,05 Removal: ,10

Maximum Iterations: 20

□ Display baseline function

The output starts with a global (‘omnibus’) test of model coefficients, meaning that it tests the hypothesis that all effects are zero:

Omnibus Tests of Model Coefficients[a,b]

-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.
500,588	15,383	2	,000	14,219	2	,001	14,219	2	,001

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 514,807

b. Beginning Block Number 1. Method = Enter

This hypothesis is clearly rejected. At least one model effect is not zero in our model. The next table is the most important one:

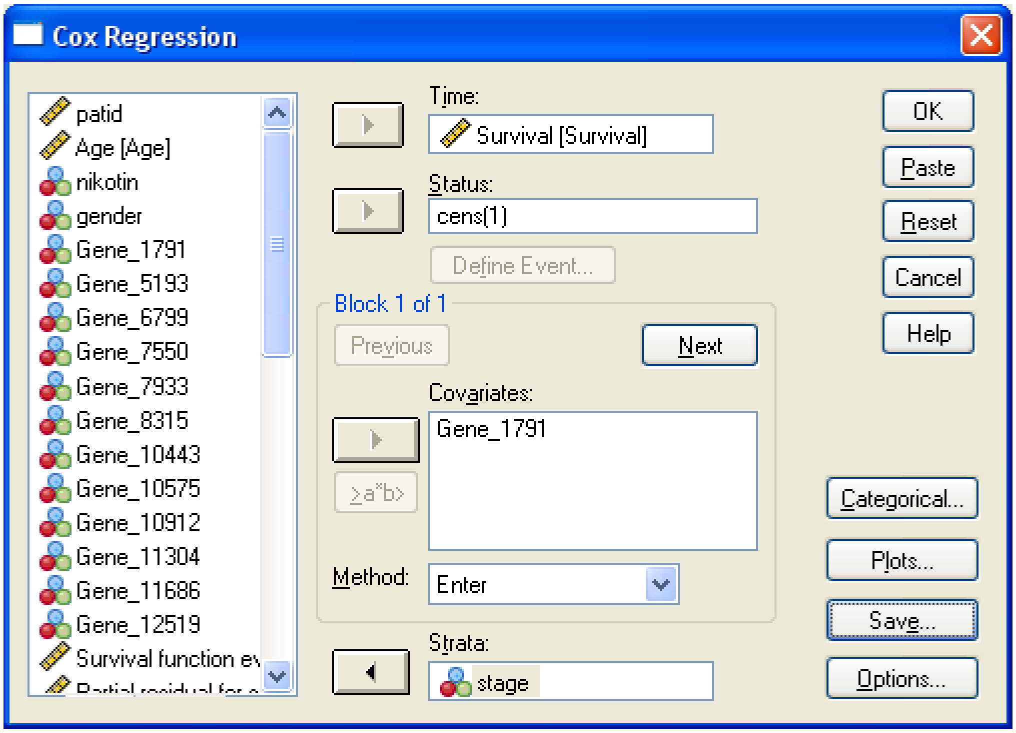
Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
stage	,788	,262	9,071	1	,003	2,199	1,317	3,673
Gene_1791	-,539	,267	4,087	1	,043	,583	,346	,984

This table contains regression coefficients (B), their standard errors (SE), the Wald statistics, the p-values (Sig.), and the estimated hazard ratios (Exp(B)) and associated 95% confidence intervals.

Stratified Cox model

A stratified Cox model can be obtained by defining a stratification factor:



Time: Survival [Survival]

Status: cens(1)

Covariates: Gene_1791

Method: Enter

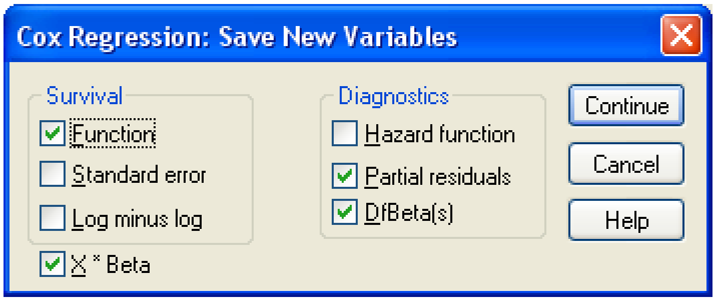
Strata: stage

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
Gene_1791	-,548	,267	4,209	1	,040	,578	,342	,976

Partial residuals and DfBeta plots

In the Save subdialogue, we may request to save certain statistics into the data matrix. The following could be of interest:



Survival

☑ Function

☐ Standard error

☐ Log minus log

☑ X * Beta

Diagnostics

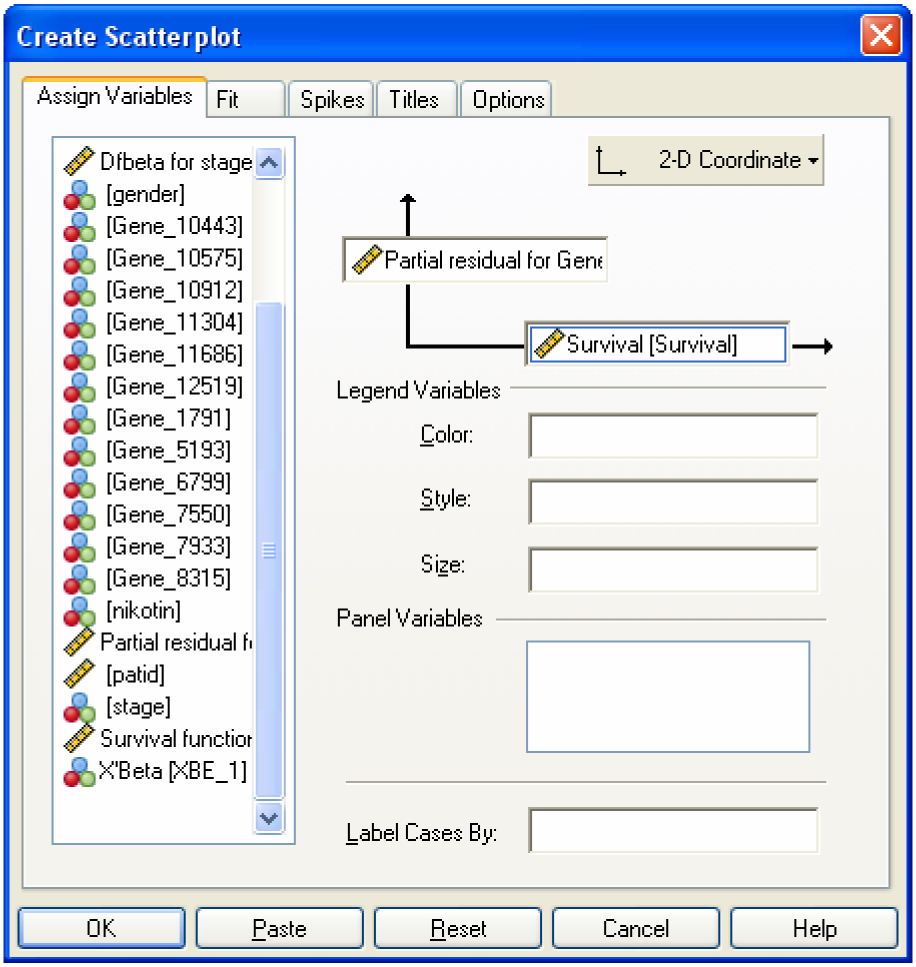
☐ Hazard function

☑ Partial residuals

☑ DfBeta(s)

Survival function saves the survival function estimates for each subject at the time of failure or at the last observation time. X*Beta saves the risk score for each function; it's the covariates times the coefficients (the linear part of the regression equation). Under 'Diagnostics', we may request partial residuals, which are of interest to detect violations of the proportional hazards assumption, or DfBeta(s), which can be used to identify observations that influence the results more than others. All these statistics are computed for each subject (with the exception of partial residuals; these are computed for each subject that had an event).

Partial residuals can be plotted against survival time by calling Graphs-Interactive-Scatterplot…, and defining the partial residuals for Gene 1791 as vertical axis variable, and survival (the survival time) as horizontal axis variable:

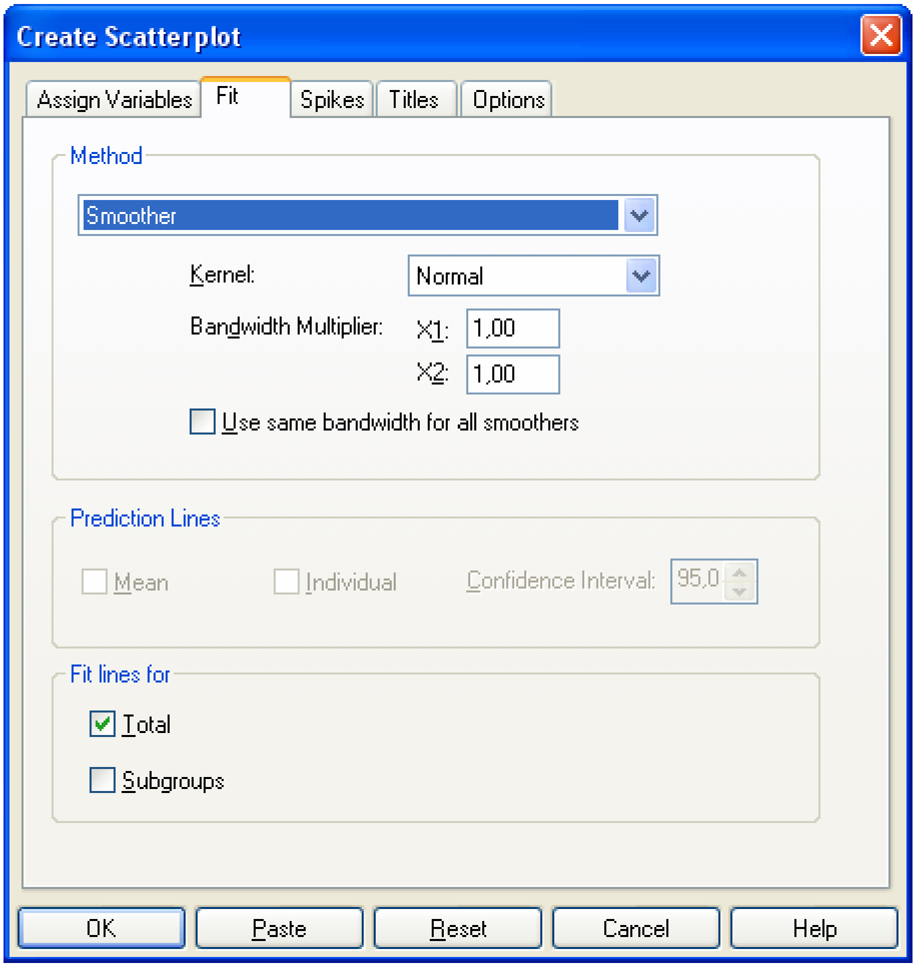

Create Scatterplot

2-D Coordinate

Partial residual for Gene

Survival [Survival]

On the ‘Fit’ view, we first request the smoother:



Method

Smoother

Kernel: Normal

Bandwidth Multiplier: X1: 1,00

X2: 1,00

□ Use same bandwidth for all smoothers

Prediction Lines

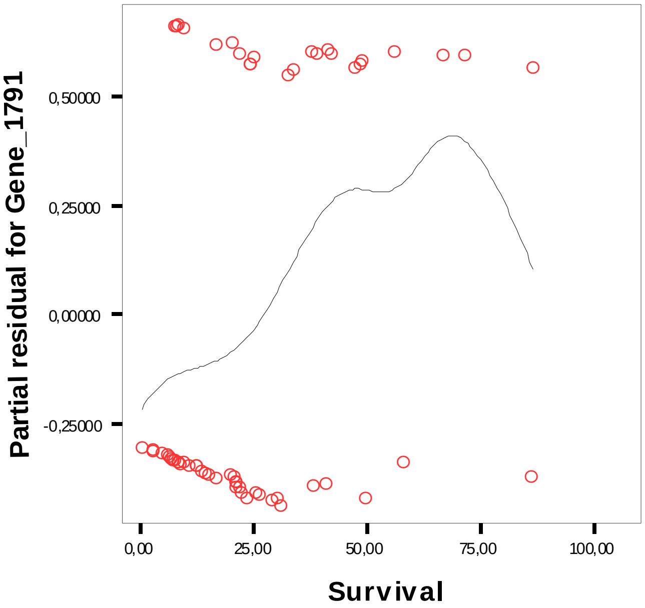
□ Mean □ Individual Confidence Interval: 95,0

Fit lines for

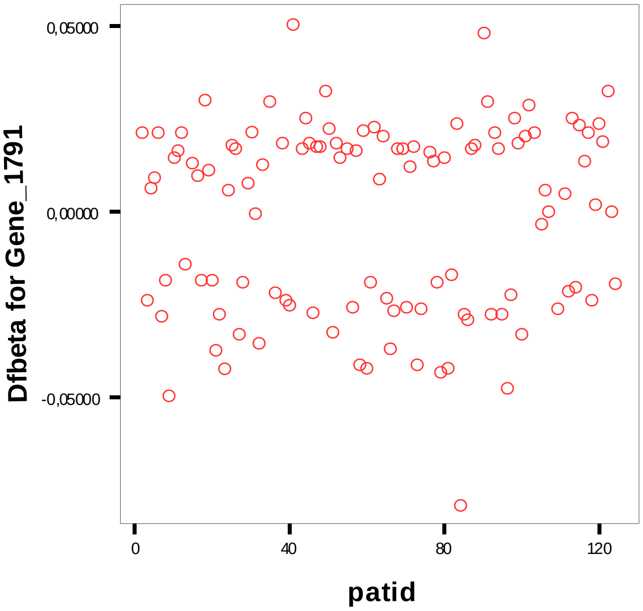
☑ Total

□ Subgroups

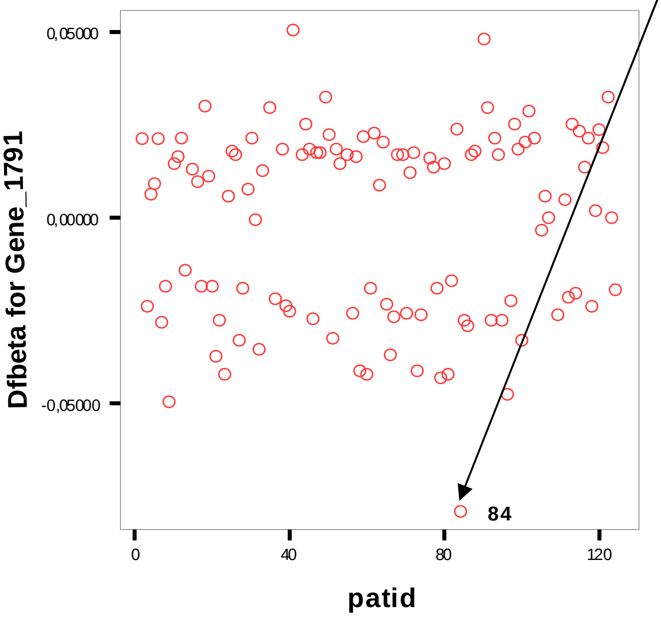
The resulting plot shows the flamboyant time-dependence of the effect of Gene 1791:



A plot of influential diagnostics can be obtained in a similar way, producing a scatter plot of DfBeta statistics against patient ID (remember to deselect the smoother, it is not useful here!):

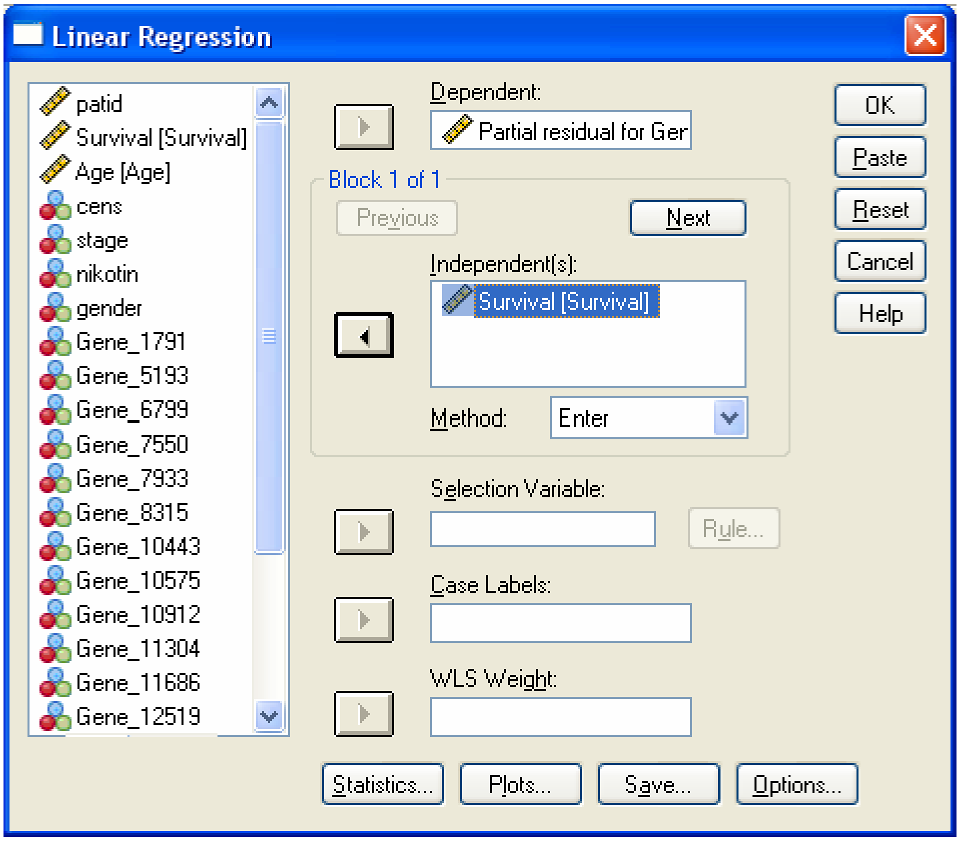


By double-clicking the graph, and choosing the symbol ⊞, we can use the cursor to identify the most extreme points:



Testing the slope of partial residuals

Using linear regression, we can test whether partial residuals increase or decrease with time. Choose Analyze-Regression-Linear… and define Partial residual for Gene 1791 as dependent variable, and survival (the survival time) as independent variable as shown below:



Dependent: Partial residual for Ger

Independent(s): Survival [Survival]

Method: Enter

Coefficients[a]

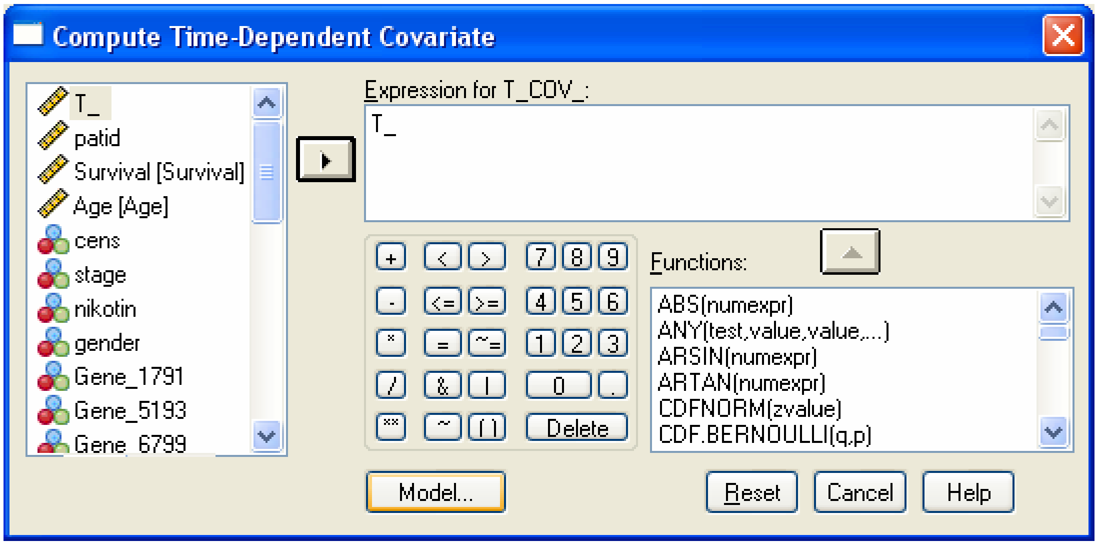
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,205	,094		-2,178	,033
	Survival	,008	,003	,338	2,754	,008

a. Dependent Variable: Partial residual for Gene_1791

We see that survival time has indeed as significant effect on the residuals, thus we conclude that the effect of Gene 1791 changes with time.

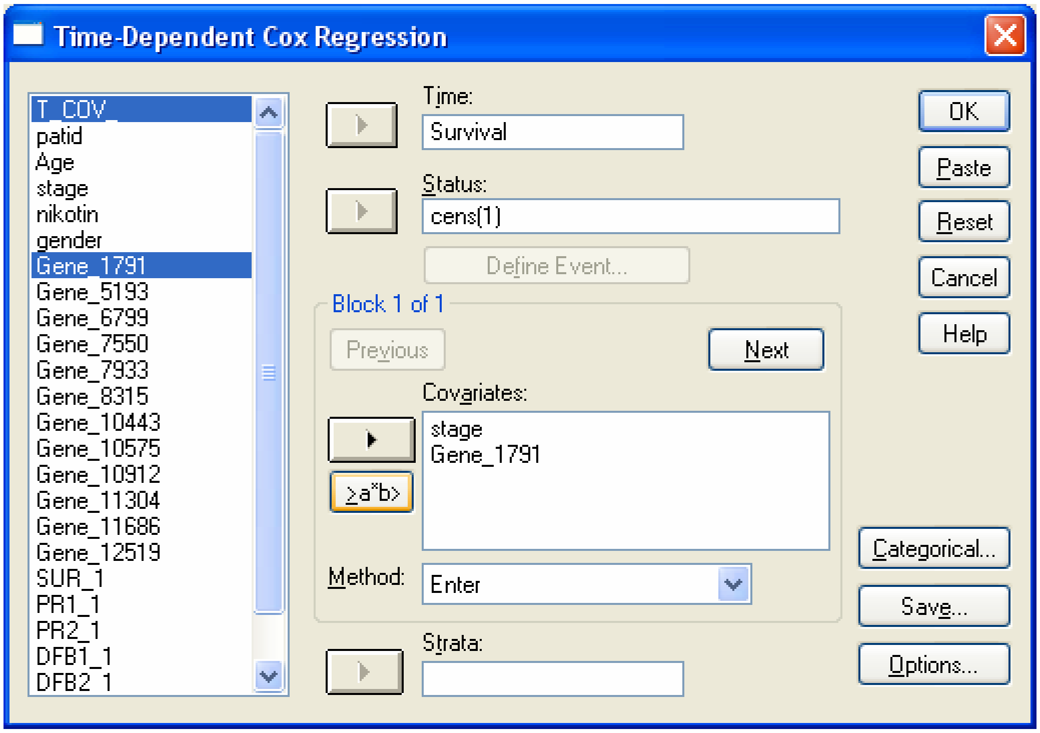
Defining time-dependent effects

Interactions of covariates with time (time-dependent effects) can be specified using the menu Analyze-Survival-Cox w/ Time-Dep Cov…. First we have to define how the interaction term should involve time (either linearly, or as logarithm etc.). We first choose a linear interaction with time, by moving the system variable T_ (standing for 'time') into the field 'Expression for T_COV_'. This variable denotes the survival time. Please note that the variable 'survival' may not be used instead of T_ to define interactions with time!



Expression for T_COV_: T_

Next, we proceed by clicking on 'Model…'. Time, Status and the covariates stage and Gene_1791 can be defined as before. However, before we click on OK, we have to select T_COV_ and Gene_1791 at the same time (select one of these variables and hold the STRG key, then select the other one). Having selected both of them, click on the '>a*b>'-button:

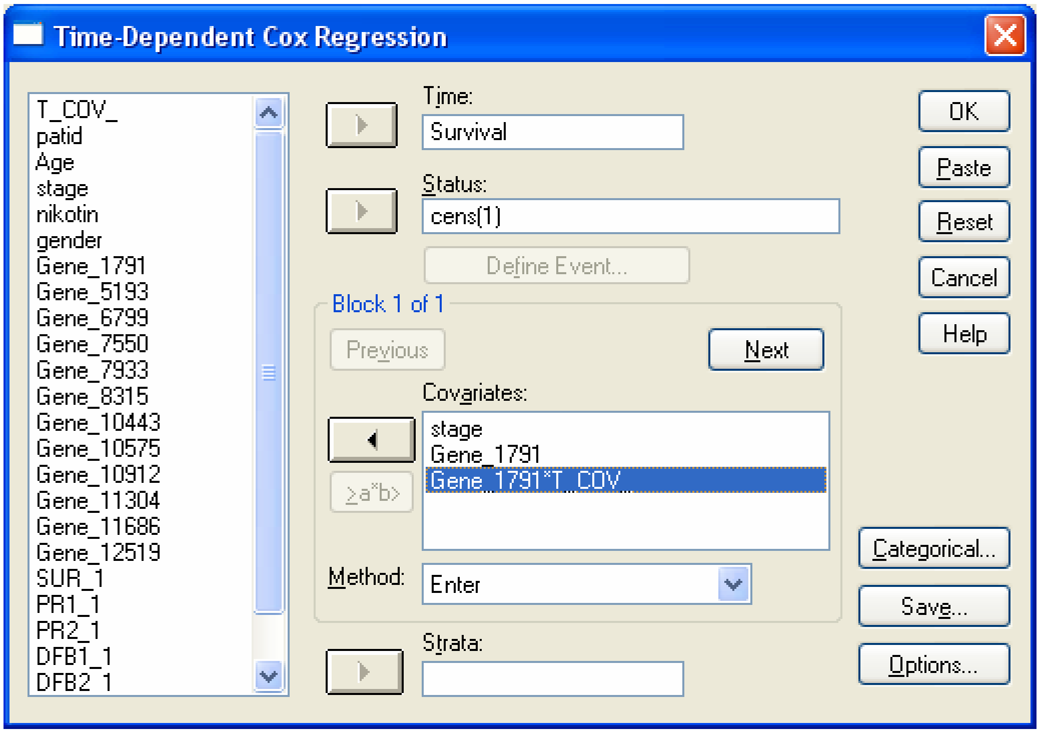


Time: Survival

Status: cens(1)

Covariates:
stage
Gene_1791

Method: Enter



Time: Survival

Status: cens(1)

Covariates:
stage
Gene_1791
Gene_1791*T_COV_

Method: Enter

Strata:

Now we can click on OK:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
stage	,798	,263	9,225	1	,002	2,222
Gene_1791	-1,554	,515	9,108	1	,003	,211
Gene_1791*T_COV_	,037	,016	5,630	1	,018	1,038

We see that the effect of Gene_1791 indeed depends on time. At the beginning of the follow-up time (when time=0), we observe an adjuste hazard ratio of 0.211. This hazard ratio increases by 3.8% (or: 'multiplies by 1.038') with every month.

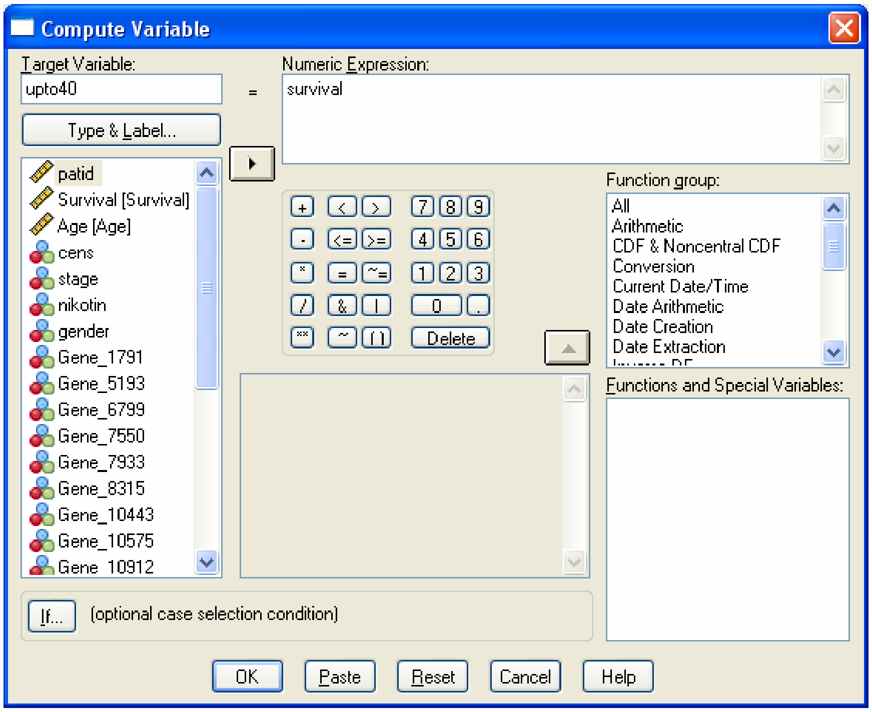
Dividing the time axis

When dividing the time axis and fitting separate models for the two subgroups, we must follow the two principles that have already been outlined above (please note that the cutpoint of 40 months is completely arbitrary, as alternative we may also choose that time at which half of the total number of events have already occurred):

- *All subjects must enter the first subgroup*. The *censoring indicator has to be set to 'censored' for all subjects which lived longer than 40 months*, and the survival time is set to 40 months for these subjects.
- Only the subjects *who lived longer than 40 months* enter the second subgroup.

The first subgroup analysis needs a redefinition of survival time. All subjects that lived longer than 40 months have to be censored at 40 months. This is done in several steps. For the first subgroup analysis, we create a new variable 'upto40' containing the survival times redefined by the principle given above:

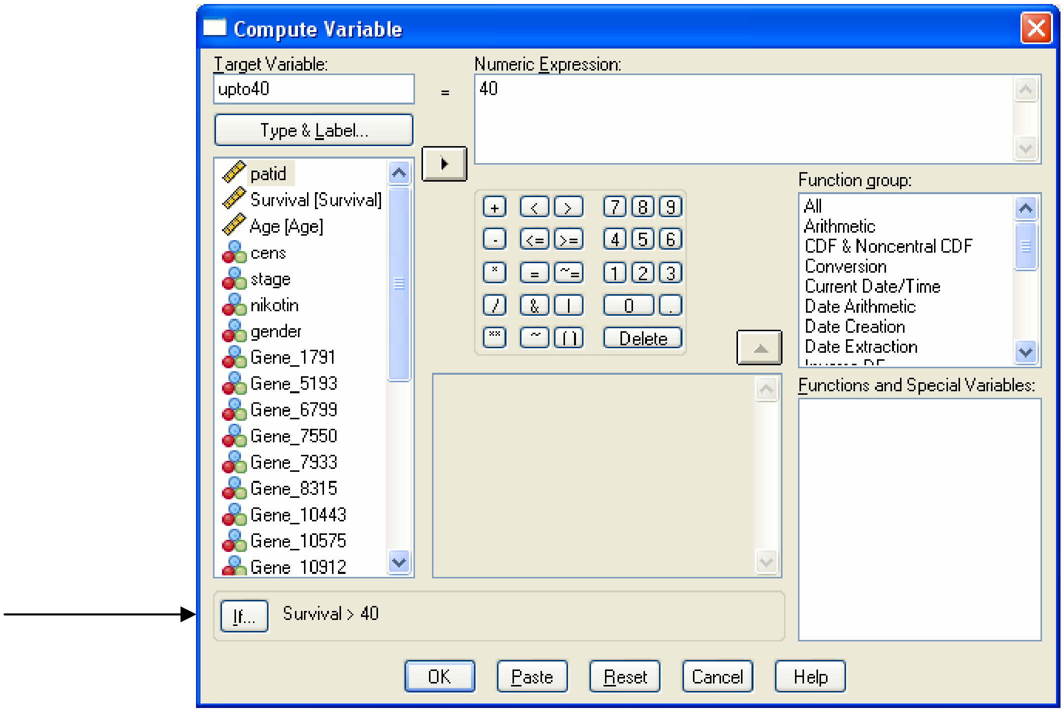
First, we set upto40 equal to survival for all subjects.



Target Variable: upto40

= Numeric Expression: survival

Next, for subjects living longer than 40 months, we restrict the survival time to 40 months:



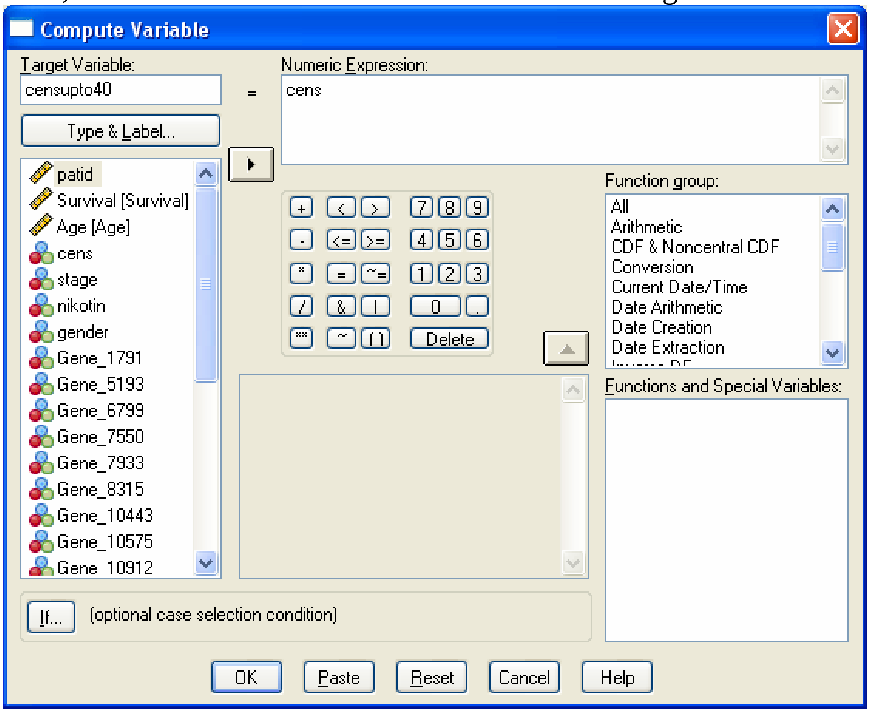
Target Variable: upto40

Numeric Expression: 40

If... Survival > 40

Then, we create a new censoring indicator. It should contain the true dead/alive status for subjects having been followed-up shorter than 40 months. For all other subjects, it should indicate 'alive', as these subjects lived longer than 40 months.

First, we define the new status indicator as assuming the same values as the original one:

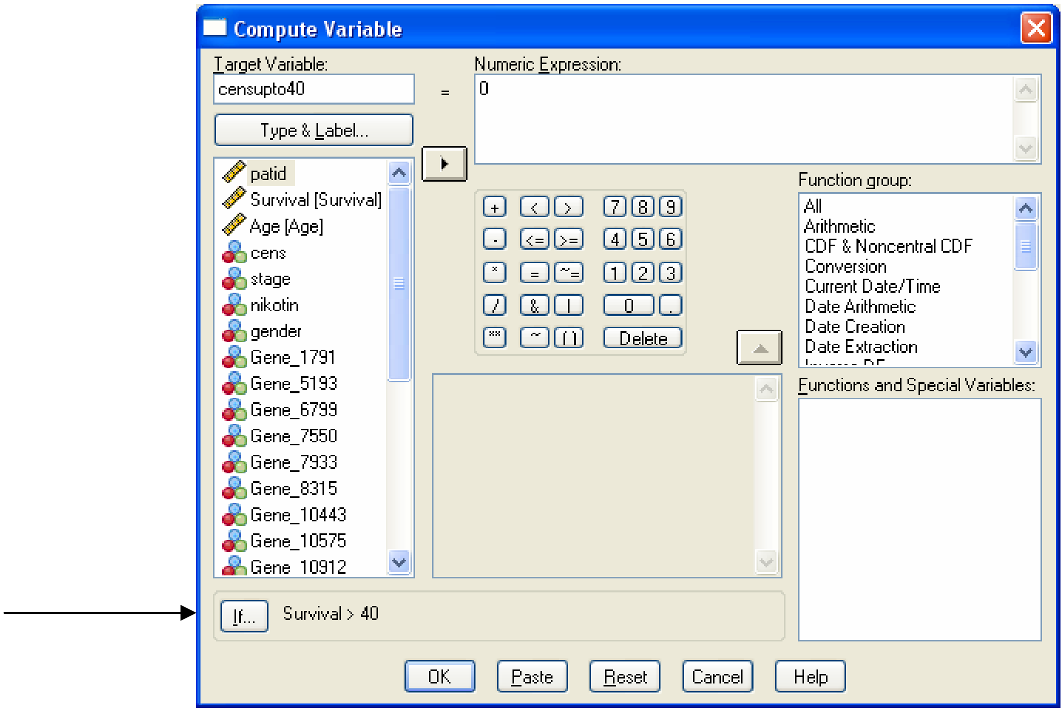


Target Variable: censupto40

Numeric Expression: cens

If... (optional case selection condition)

Then we change it to 0 for subjects which lived longer than 40 months:

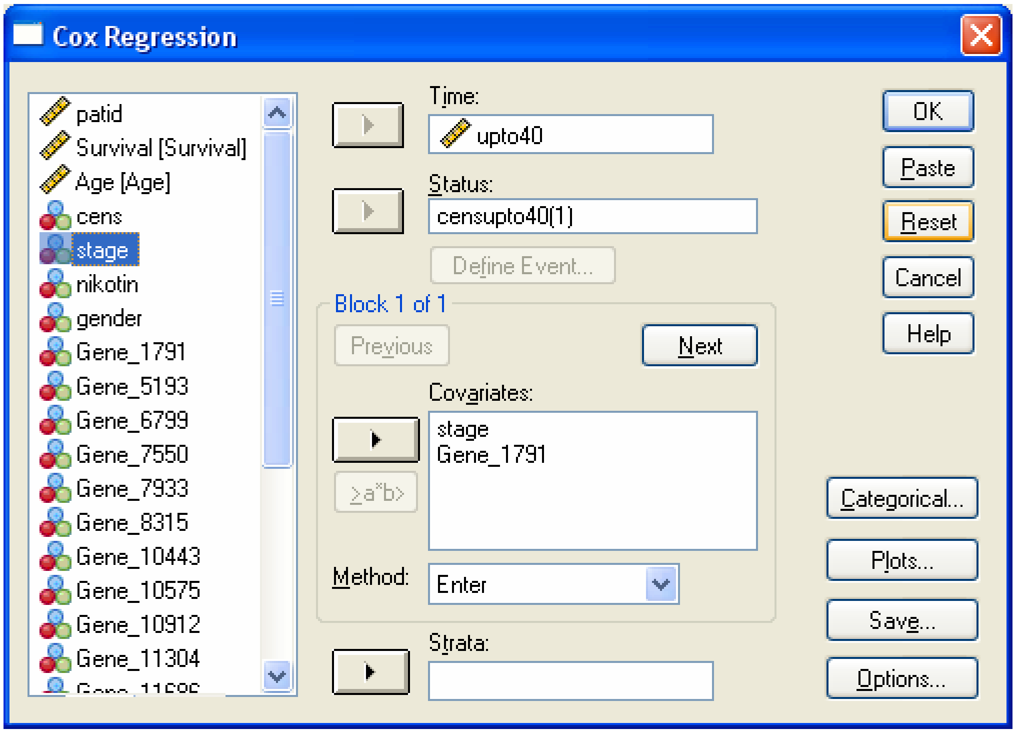


Target Variable: censupto40

Numeric Expression: 0

Survival > 40

Now we call Cox regression, using the redefined survival time and status:



Time: upto40

Status: censupto40(1)

Block 1 of 1

Covariates:
stage
Gene_1791

Method: Enter

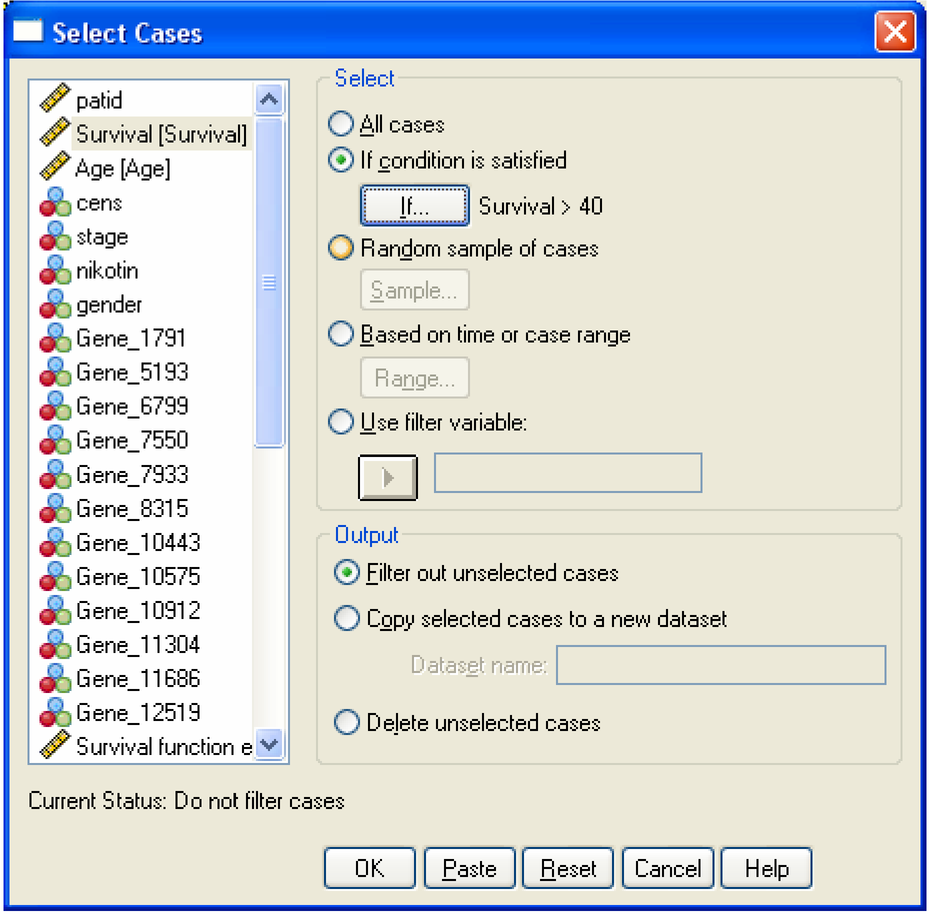
Strata:

We obtain the following table of hazard ratio estimates:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
stage	,945	,292	10,499	1	,001	2,574	1,453	4,560
Gene_1791	-,883	,320	7,637	1	,006	,413	,221	,773

The second subgroup analysis can be obtained by simply requesting a Cox regression analysis with the original survival time and censoring indicator, but restricting the analysis to all subjects with a survival time longer than 40 months. Choose Data-Select Cases and request 'If Survival > 40':



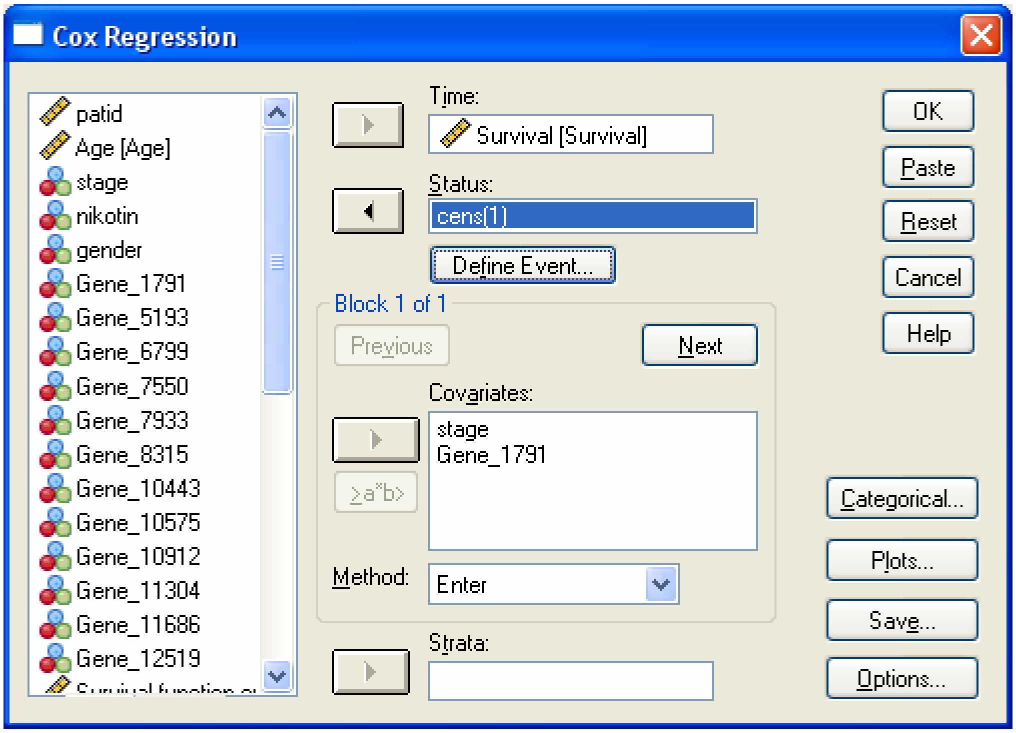
Select: ● If condition is satisfied

If... Survival > 40

Output: ● Filter out unselected cases

Current Status: Do not filter cases

After this selection, we call Cox regression using the original survival time and status variables:



Time: Survival [Survival]

Status: cens(1)

Covariates: stage, Gene_1791

Method: Enter

Strata:

The resulting table refers to all subjects that lived longer than 40 months.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
stage	,169	,677	,063	1	,802	1,185	,314	4,465
Gene_1791	,578	,608	,903	1	,342	1,783	,541	5,876

We see that the hazard ratio estimates are quite different from the estimates we computed for the first 40 months. (Although they are not significantly different from 0, there is, for Gene_1791, a significant difference if we compare 1.783 to 0.413.)

References

[1] Kaplan EL, Meier P. (1958) Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53: 457-481
[2] Bhattacharjee, A., et al (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. PNAS 98, 13790-13795.
[3] Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society B 34: 187-220.
[4] Schoenfeld, D. (1982). Partial residuals for the Cox proportional hazards model. Biometrika 69, 239-241.

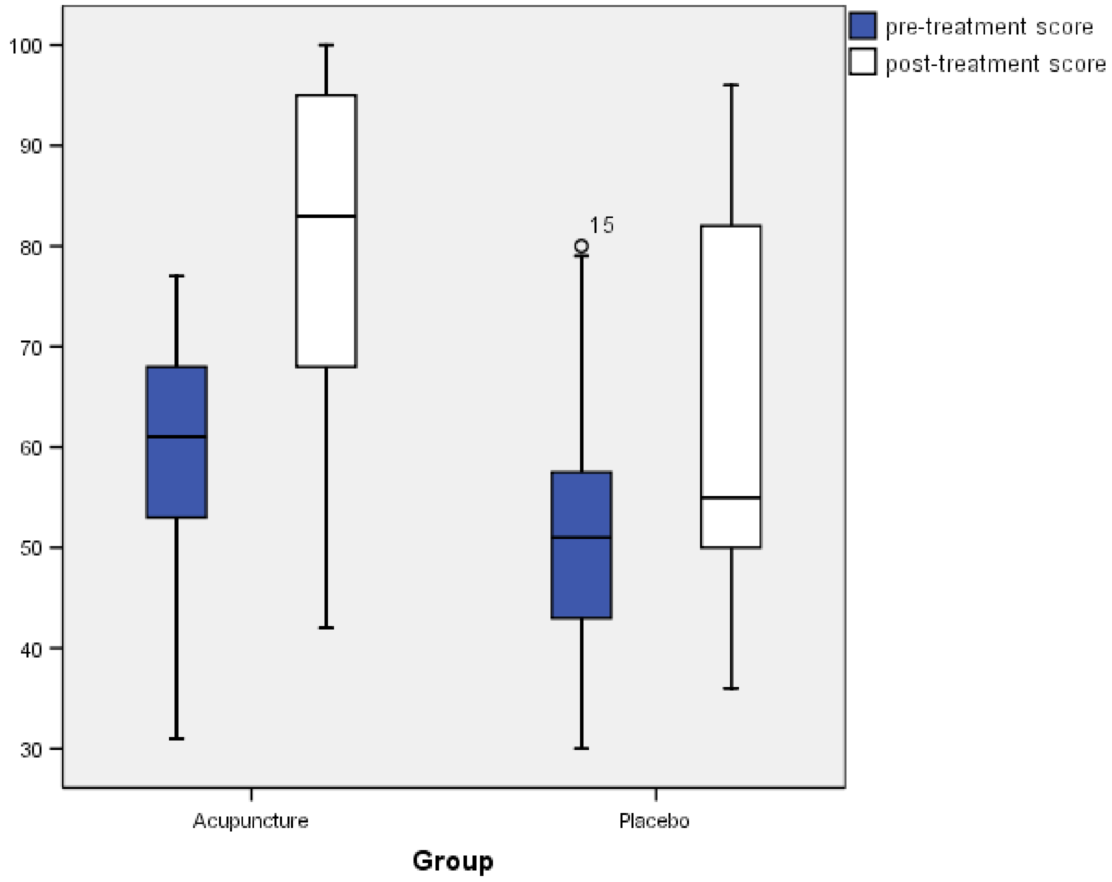
Part 4: Analysis of repeated measurements

Class 9: Pretest-posttest data

Pretest-posttest data

Kleinhenz et al [1] randomised 52 patients with shoulder pain to either true or sham acupuncture (‘placebo’). Patients were assessed before and after treatment using a 100 point rating scale of pain and function, with lower scores indicating poorer outcome.

A box plot illustrates the distribution of pre- and post-treatment scores, stratified by treatment:



We see that the post-treatment functional scores in the acupuncture group are clearly higher than those in the placebo group, but there is already an imbalance between the groups at baseline, with better scores in the acupuncture group.

Although we notice an improvement in both groups, we cannot see the (variation of) individual improvements. The plot may even obscure occasional decreases in functional scores in some patients.

A proper analysis should therefore

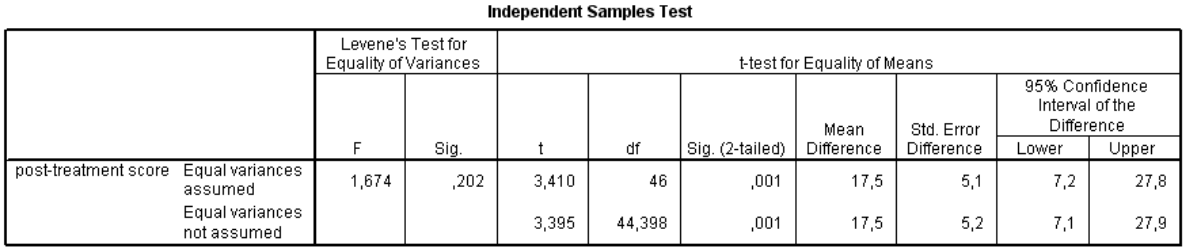
- Adjust for the baseline imbalance
- take into account individual changes in functional scores

Change scores

Let's first ignore the baseline imbalance and compare the post-treatment scores using a t-test:

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
post-treatment score	Acupuncture	25	80,3200	16,83182	3,36636
	Placebo	23	62,8261	18,70744	3,90077



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
post-treatment score	Equal variances assumed	1,674	,202	3,410	46	,001	17,5	5,1	7,2	27,8
	Equal variances not assumed			3,395	44,398	,001	17,5	5,2	7,1	27,9

We see that there is a highly significant difference in post-treatment functional scores between patients treated by acupuncture and patients treated by placebo. However, so far we are not sure whether this significant effect is just a product of the baseline imbalance between the groups.

The easiest way to accomplish both requirements is to use change scores. These are simply defined as the difference between post-treatment and pre-treatment score, computed for each patient separately. Change scores can be computed as *raw change scores* or as *percent change scores*:

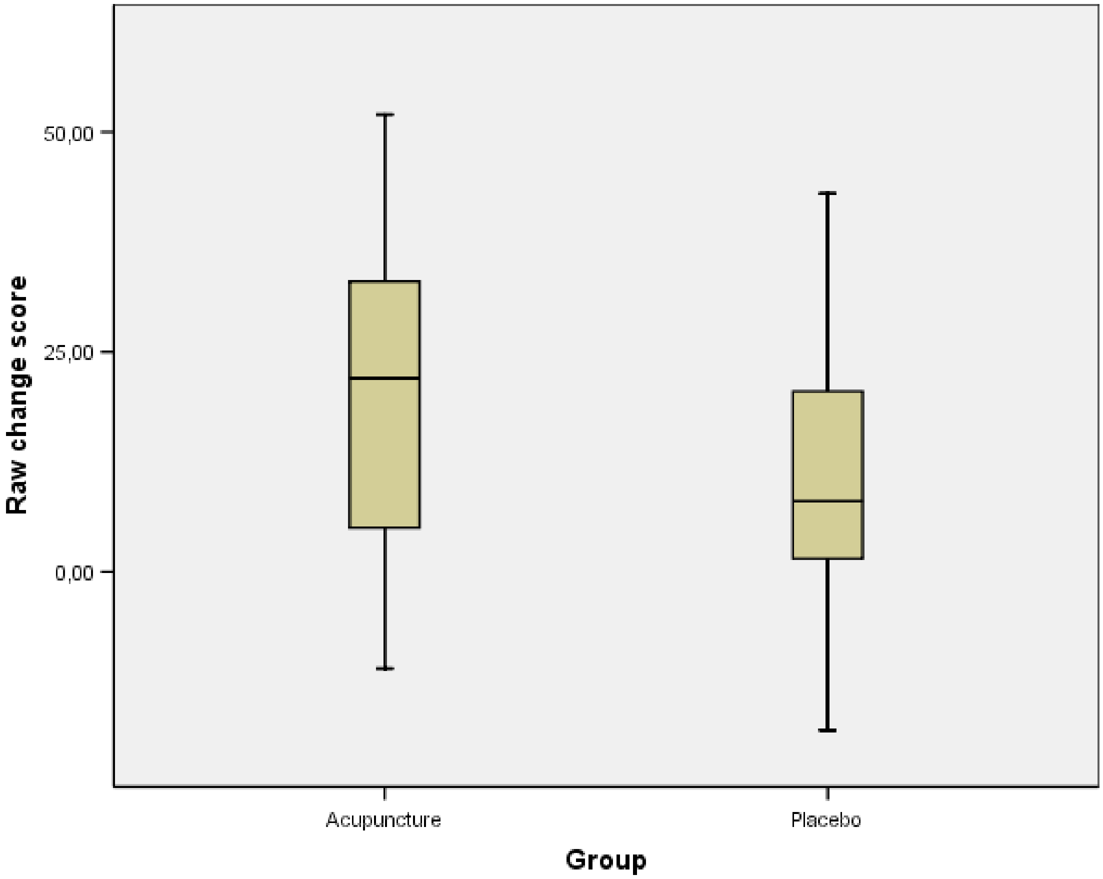
Raw change scores: $RCS = Y - X$

Percent change scores: $PCS = (Y - X)/X \cdot 100\%$

We will use raw change scores, if a change of 5 units has the same interpretation at all levels of pre-treatment scores. Sometimes, this may not be useful, and then percent change scores are more appropriate. Using percent change scores, we assign the same importance to a change from, e. g., 20 to 15 (-25%) as to a change from 40 to 30 (-25%).

Having computed RCS or PCS, we can compute measures of location (mean or median) and scale (standard deviation or quartiles) and compare them between the groups. If the change scores have a symmetric distribution, we can use the t-test for comparison, otherwise a t-test after log transformation or the Mann-Whitney-U (Wilcoxon rank-sum) test can be applied.

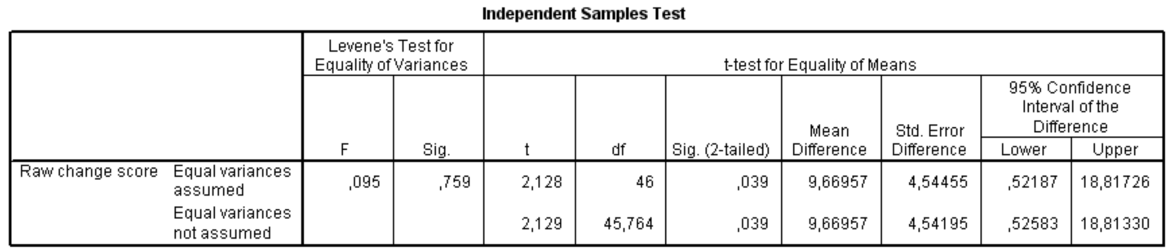
In our example, the change scores have the following distributions:



The boxplot shows approximate normal distributions of the raw change sores, such that use of the t-test is indicated:

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Raw change score	Acupuncture	25	20,8000	15,82982	3,16596
	Placebo	23	11,1304	15,61847	3,25668

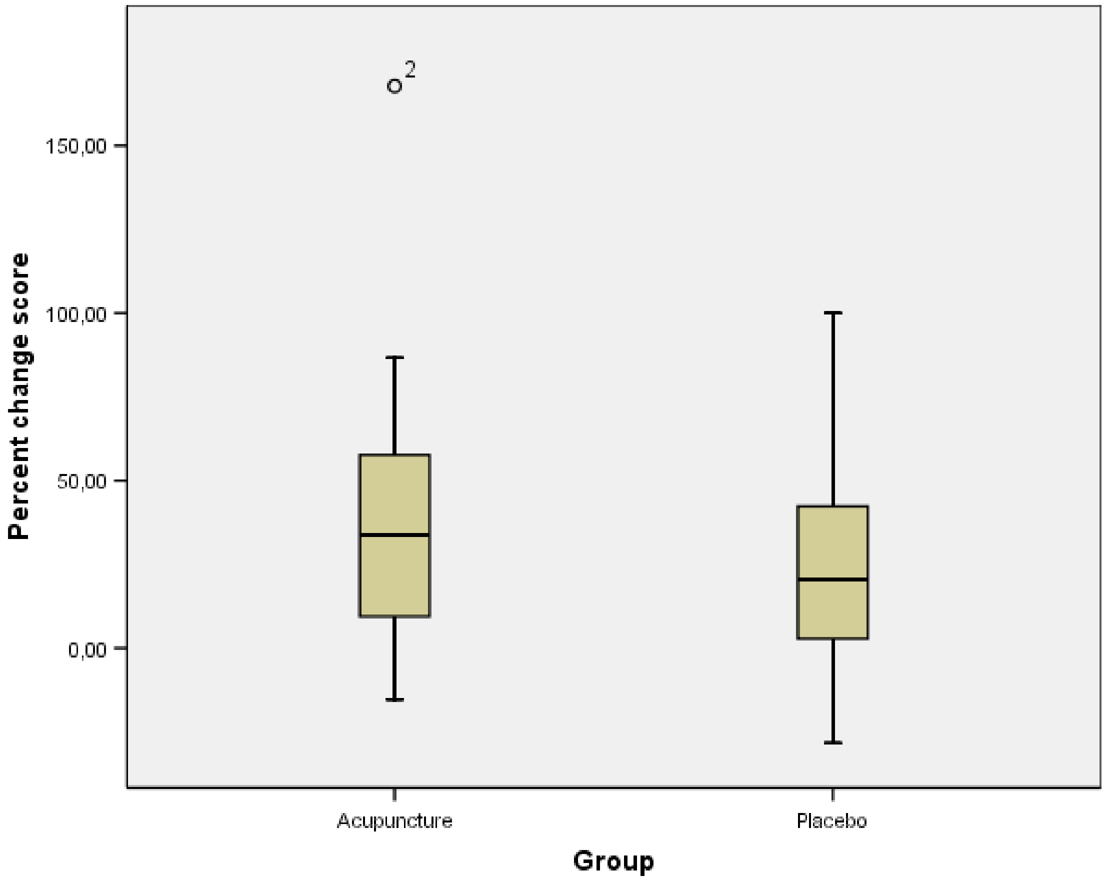


Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw change score	Equal variances assumed	,095	,759	2,128	46	,039	9,66957	4,54455	,52187	18,81726
	Equal variances not assumed			2,129	45,764	,039	9,66957	4,54195	,52583	18,81330

The t-test shows a significant difference between the two groups (p=0.039).

Using percent change scores, we arrive at a similar conclusion:



Because of the outlier observed in the acupuncture group, we might prefer the Mann-Whitney-U test for comparison:

			Count	Median	Percentile 25	Percentile 75
Group	Acupuncture	Percent change score	25	33,78	8,17	58,18
	Placebo	Percent change score	23	20,59	2,78	43,33

Test Statistics(a)

	Percent change score
Mann-Whitney U	220,000
Wilcoxon W	496,000
Z	-1,393
Asymp. Sig. (2-tailed)	,164

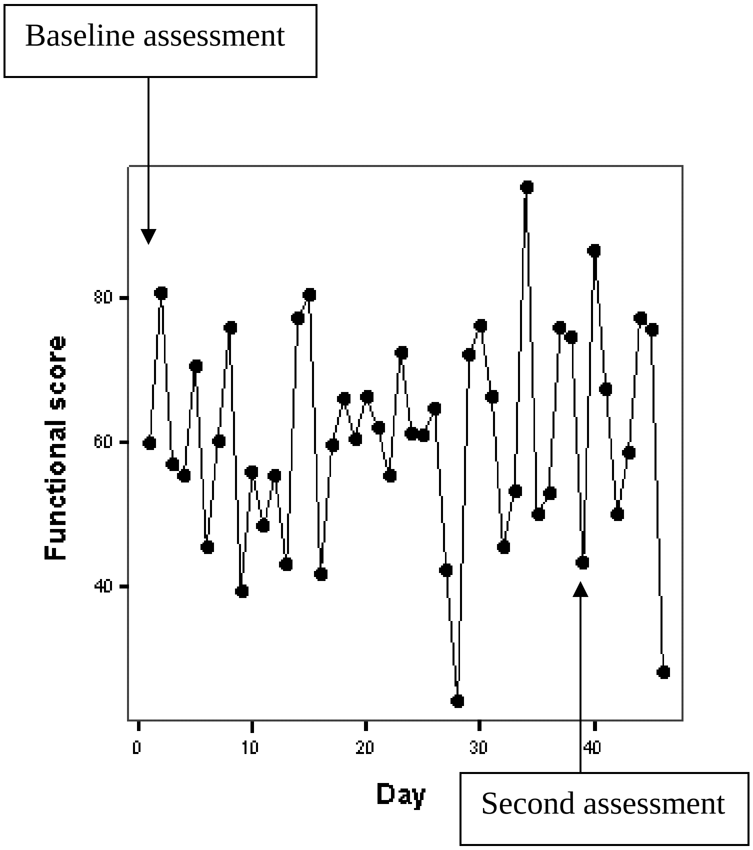
a Grouping Variable: Group

The difference in relative (percent) changes is not so strong as in raw changes, such that significance is not reached.

Comparing change scores we have to take into account the possibility of the so-called *regression to the mean* effect. This effect is responsible for the general observation that patients with low functional scores before start of treatment tend to have greater change scores than patients with better functional scores prior to treatment. If baseline functional scores are very heterogeneous, this may affect the change scores leading to a biased analysis. Using alternative options of analysis we can circumvent the regression to the mean effect and obtain an unbiased group comparison. One such option is to use analysis of covariance, and will be presented later.

The regression to the mean effect

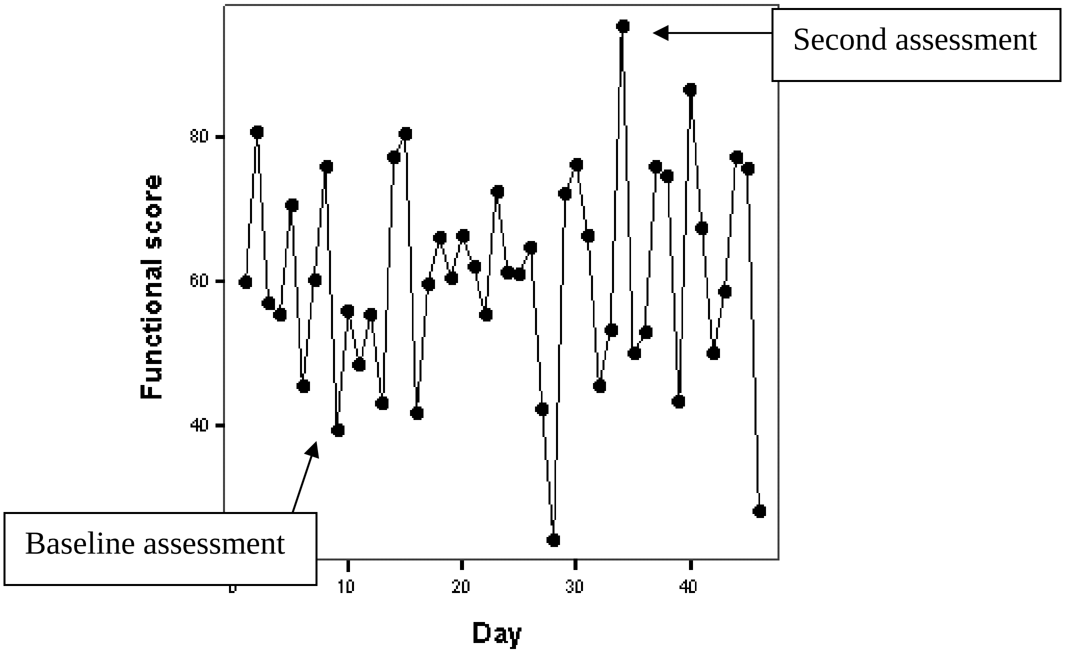
Assume an ineffective treatment of patients with shoulder pain, leading to fluctuations in repeatedly assessed functional scores that are purely due to the randomly changing constitution of the patients at subsequent days.



Baseline assessment

Second assessment

If a patient is – by chance – in a good constitution at the day at which the baseline measurement is taken, it is very likely that he/she will be in a worse constitution at the next assessment, leading to a decrease in functional score.



Second assessment

Baseline assessment

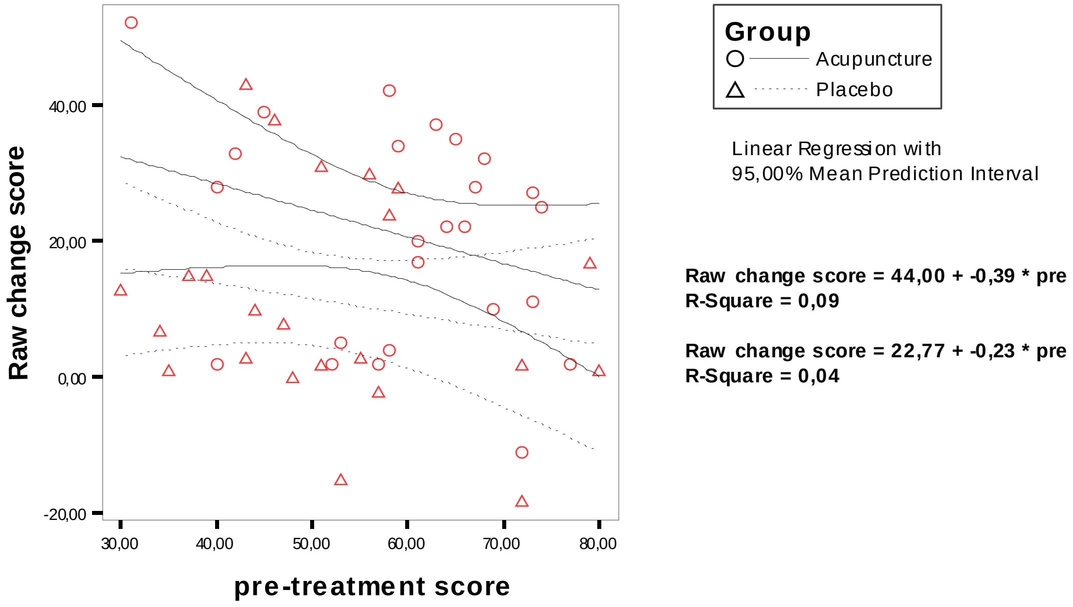
If, on the other hand, a patient's baseline measurement was taken on a day at which he/she was in bad constitution, then it is more likely that the next assessment will yield a better score, leading to an improvement in functional score.

Thus, low baseline functional scores are – in our hypothetical example of no treatment effect – associated with positive change scores, and high baseline functional scores correlate with negative change scores.

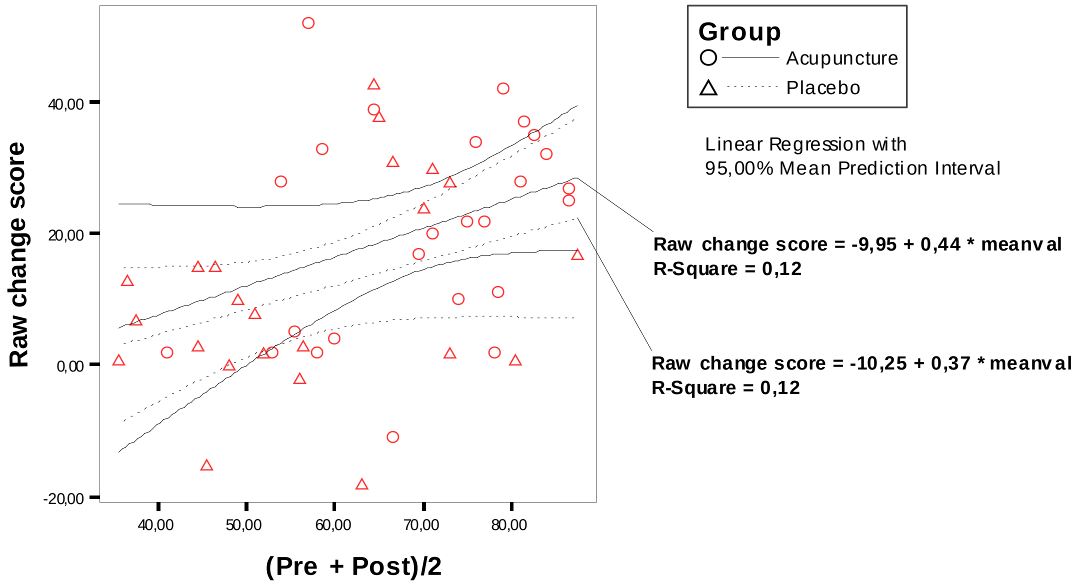
If a treatment effect is present, we must still assume that the change scores will tend to be negatively correlated with the baseline scores.

Assuming that there is no effect of treatment, we would assume the mean of the two assessments as the best guess of a patient's long-time functional score.

The regression to the mean effect becomes most evident, if we correlate change scores with baseline values:



We see that in both groups, the change scores are negatively correlated with pre-treatment scores. This obvious negative correlation is a logical result from comparing the expressions Y - X and X. Therefore, for determining whether the magnitude of change scores correlate with baseline values, we have to correct for this automatic correlation by replacing X by (X+Y)/2 (the mean) at the x-axis. The mean again serves as the best estimate of the long-time average of a patient's functional score.



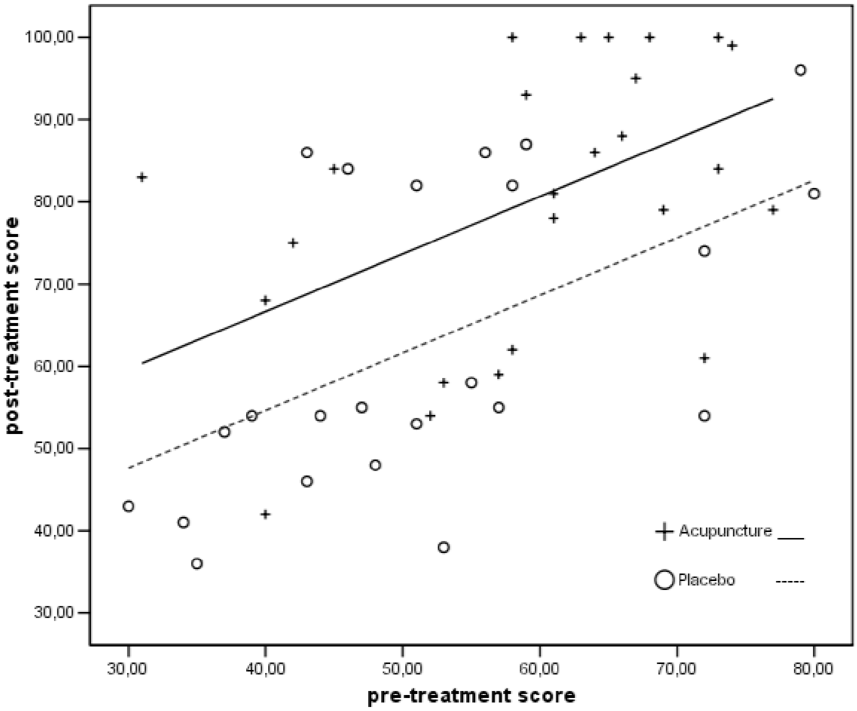
After correction for the regression-to-the-mean effect, we actually notice a positive correlation between the magnitude of functional score and the change in the scores, leading to the conclusion that patients with higher scores benefit more from treatment (and placebo!) than patients with lower scores.

Analysis of covariance

In order to correct for the regression-to-the-mean effect when comparing the treatment effect between acupuncture and placebo groups, the simple comparison of change scores should be replaced by an analysis of covariance (ANCOVA).

ANCOVA is the same as ANOVA (or, in a two-group situation, a t-test), but it allows for taking into account covariates which can be used for assessing *adjusted treatment effects*. Technically, we fit a linear regression model with two variables – one defining the treatment groups and one representing the baseline assessment. The change score or the post-treatment scores may be used as outcome, leading to the same conclusion.

From ANCOVA we obtain an estimate of the treatment effect (the difference in change score), which is adjusted for baseline imbalance using multivariable estimation. In other words, we obtain the average difference in change score (or post-treatment score) between a patient of the acupuncture group and a patient of the placebo group, assuming both had the same baseline values.



The above plot shows the way ANCOVA adjusts for pre-treatment (baseline) scores. There are more patients with low pre-treatment scores from the placebo group than from the acupuncture group. ANCOVA fits two regression lines with the same slope into the data cloud. The slopes of the regression lines are equal, because we assume the same relationship between pre-treatment and post-treatment scores in both patient groups. The vertical distance between the regression lines represents the difference in post-treatment scores between the two groups, assuming equal pre-treatment scores. Since we allow the post-treatment scores to depend on baseline values, this distance is independent from the baseline values and we can interpret it as the baseline-adjusted treatment effect.

ANCOVA analysis yields the following table of regression coefficients:

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	50,693	13,642		3,716	,001
	pre-treatment score	,700	,174	,480	4,014	,000
	Group	-12,019	4,655	-,309	-2,582	,013

a Dependent Variable: post-treatment score

We see that after optimal adjustment for baseline imbalance, we end up with a difference in post-treatment scores of -12, referring to the difference Placebo (code 2) – Acupuncture (code 1). Thus, patients treated by acupuncture have post-treatment scores which are on average 12 units higher than those of patients in the placebo group with equal pre-treatment scores. This effect is significant (p=0.013).

Summarizing the three types of analysis, we have:

Type of analysis	Treatment effect	P-value	Remark
t-test on post-treatment scores	+17.5	0.001	Assumed to be too large because of the baseline imbalance that has not been accounted for
t-test on raw change scores	+9.7	0.039	Must be assumed as too small because of the regression-to-the-mean effect (the lower baseline values in the placebo group are associated with higher change scores)
ANCOVA	+12.0	0.013	Assumed to supply the most unbiased results

Class 10: Visualizing repeated measurements

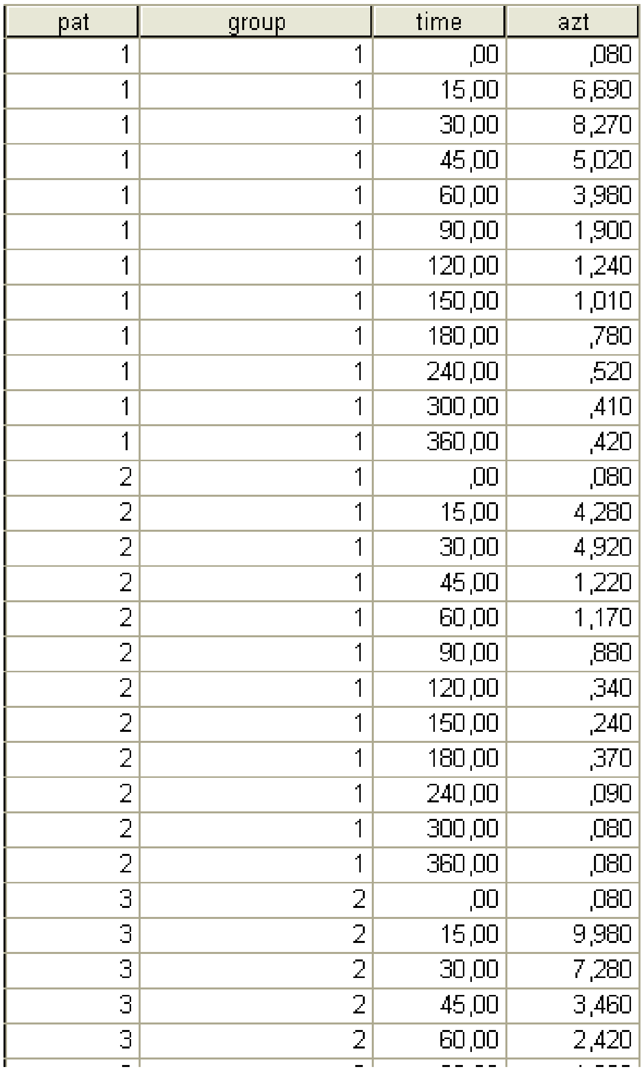
Introduction

In some medical studies the coarse of a parameter is followed over a pre-defined period of time. Such studies may have several purposes:

- To describe the course of a parameter after an intervention
- To identify typical shapes of courses
- To compare the course of a parameter between groups of patients defined by an experimental condition

Example: zidovudine [2]

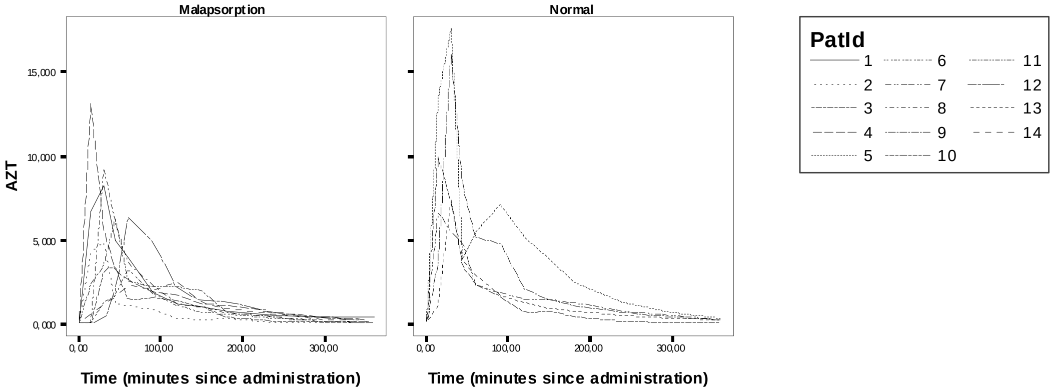
The following table (AZT.sav) shows the serum zidovudine levels (AZT) of AIDS patients at various time points after administration of AZT. Some patients are known to possess abnormal fat absorption (malapsorption):



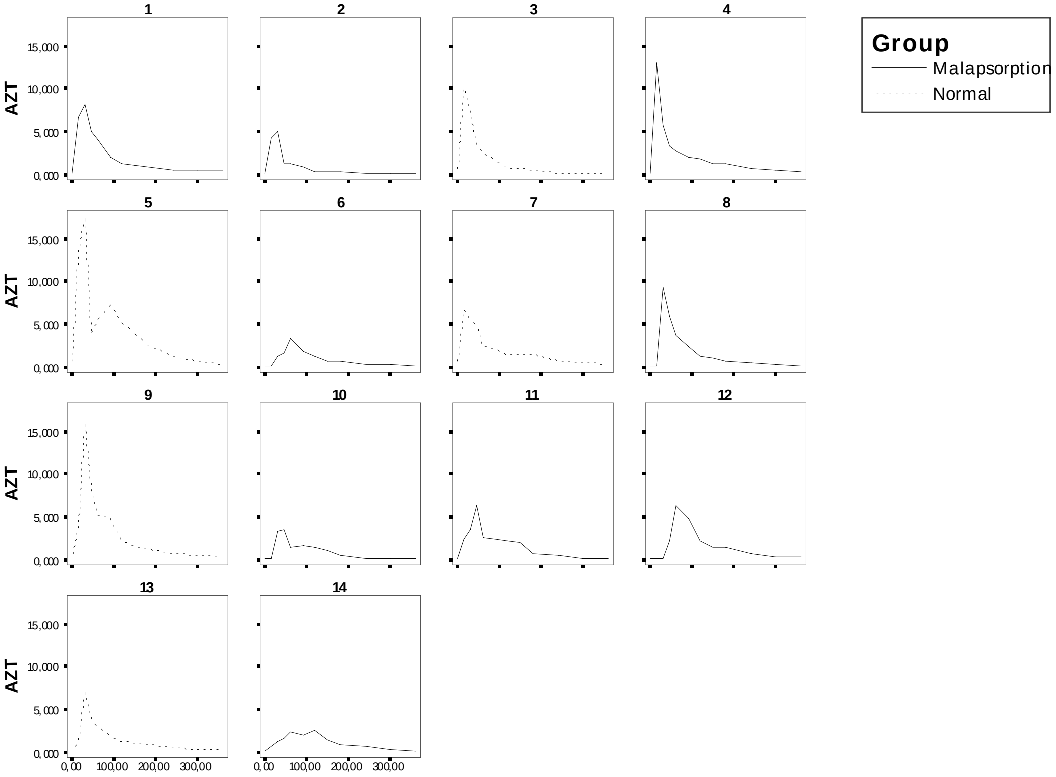
pat	group	time	azt
1	1	,00	,080
1	1	15,00	6,690
1	1	30,00	8,270
1	1	45,00	5,020
1	1	60,00	3,980
1	1	90,00	1,900
1	1	120,00	1,240
1	1	150,00	1,010
1	1	180,00	,780
1	1	240,00	,520
1	1	300,00	,410
1	1	360,00	,420
2	1	,00	,080
2	1	15,00	4,280
2	1	30,00	4,920
2	1	45,00	1,220
2	1	60,00	1,170
2	1	90,00	,880
2	1	120,00	,340
2	1	150,00	,240
2	1	180,00	,370
2	1	240,00	,090
2	1	300,00	,080
2	1	360,00	,080
3	2	,00	,080
3	2	15,00	9,980
3	2	30,00	7,280
3	2	45,00	3,460
3	2	60,00	2,420

Individual curves

The first step in the analysis of repeated measurements should always consist of a graphical display of individual curves (separately for each patient). With a small number of patients, one may draw all curves into one diagram. In our data set, there are nine patients with malapsorption and five patients with normal fat absorption. Therefore, we can draw two diagrams, one for each group:



Alternatively, separate diagrams for each patient may provide a clearer view:



Why do we depict individual curves before further analyses? There are several reasons:

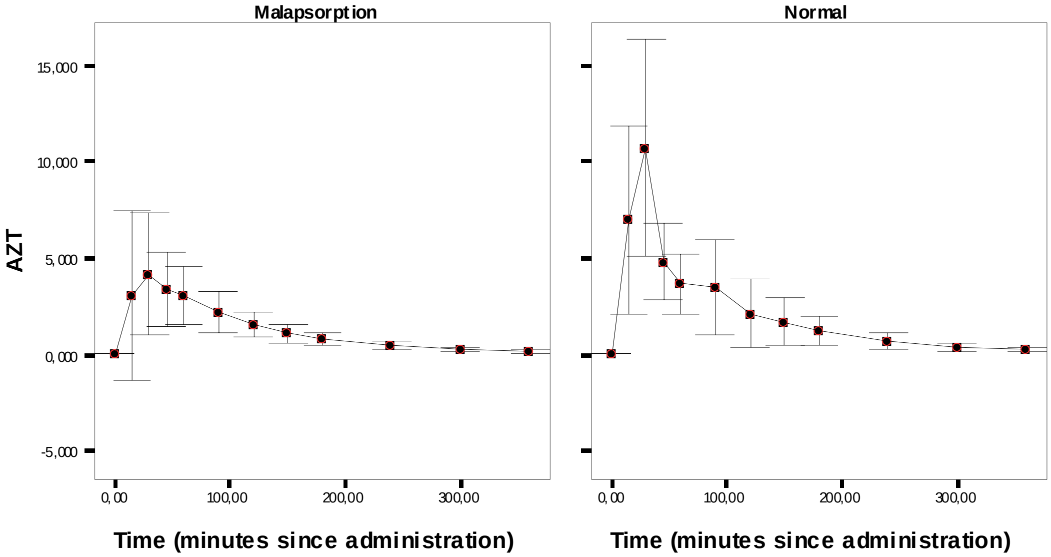
- First, explorative data analysis provides insight into the distribution of variables, and some types of distributions demand special statistical methods (e. g., for skewed distributions we should not use methods based on the assumption of normally distributed data). For this purpose, all individual curves should be equally scaled, i. e., the diagrams should all have the same y-axis and x-axis definitions.

- Second, we may identify outliers or data errors by inspecting individual curves.
- Third, different shapes of time courses may be identified by inspecting individual curves. In the example shown above, two types are identified: we see curves with a marked peak at the beginning (e.g., patients 1, 3, 4, 5, 8, 9, 13), and some with a slower ascent and a smoother peak occurring later (6, 14).

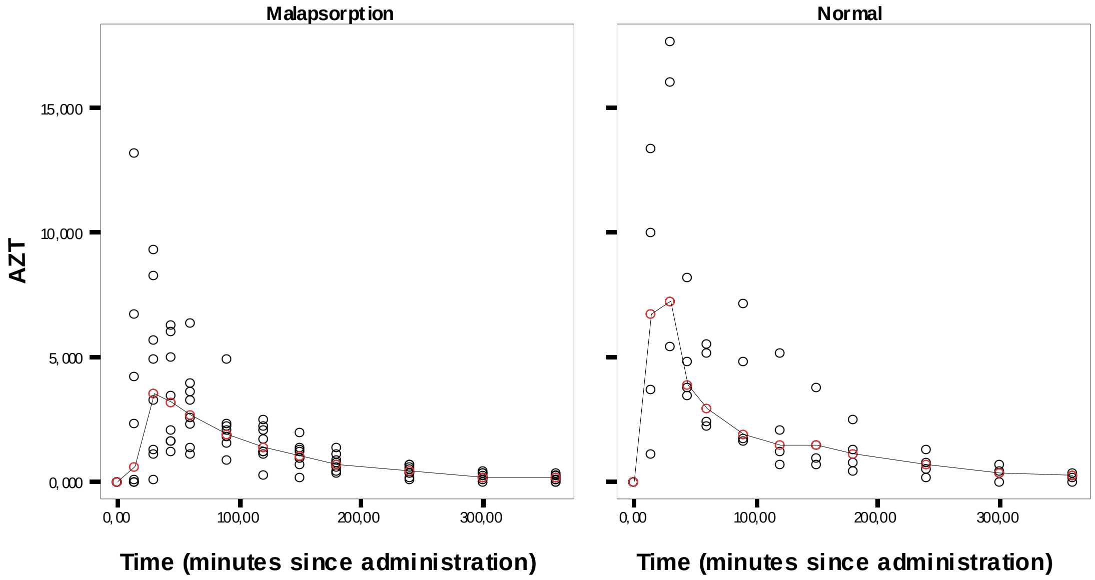
Grouped curves

Only if the shapes of the curves are similar, grouping the curves by showing the course of means or medians over time is reaonable. In any case, also the variation of the values should be properly shown, either by displaying standard deviation as error bars or by adding curves referring to 25th and 75th percentiles (or minimum and maximum).

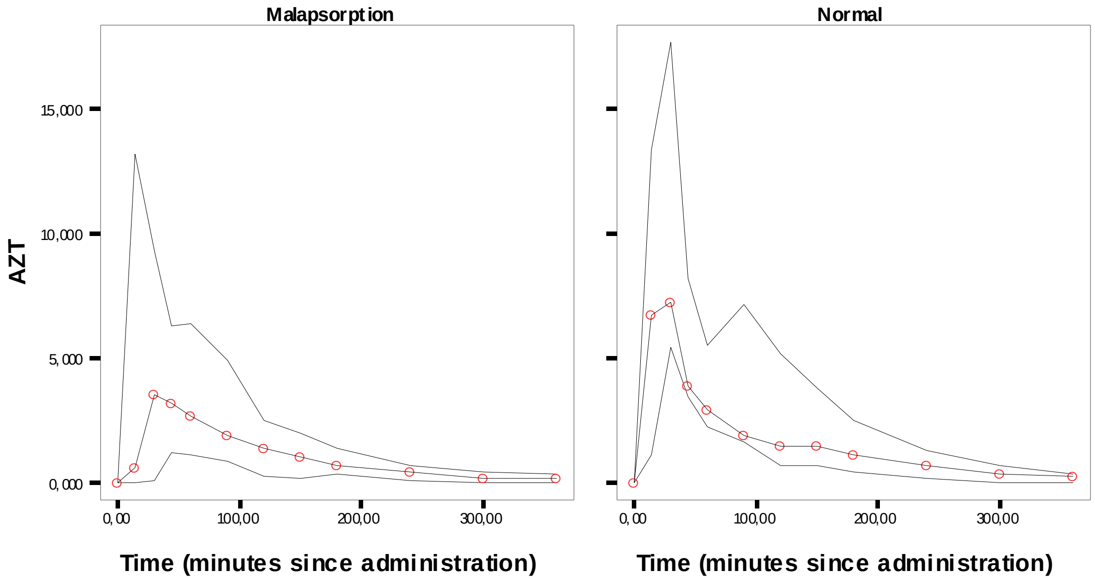
For the example shown above, a plot of the mean course with standard deviation, grouped by type of fat absorption, yields the following diagram:



In this example we notice that in the malabsorption group, the error bar representing the standard deviation reaches negative values, which are not possible for serum levels. This indicates that the distribution of serum AZT levels is not symmetric. To be sure, we could depict its distribution by separate dot plots (because of the small N!) for each group and time point:



In the above plot, the lines represent the medians at each time point. Clearly, the variation of data points above the median is greater than that of data points below the median. Therefore, we should not use parametric descriptive statistics, but rather the median and, because of the small number of subjects, minimum and maximum (instead of 25th and 75th percentiles):



The shape of individual curves can be quite different from that of mean (median) curves. Several different patterns of curves may even average to a mean (median) curve, which is by itself not typical for the individual curves. Therefore, individual curves should be compared to the aggregated curves in order to detect such effects.

Drop-outs

Special attention should be paid to patients for whom measurements after a particular time point are unavailable. Reasons for such drop-outs are:

- Patients who die
- Patients who are lost-to-follow-up
- Patients unwilling to further participate in the study
- Patients for whom measurements cannot be taken
- …

If the frequency of drop-outs is imbalanced between the groups, or if the missing-at-random assumption is violated (patients of group A die more often than patients in group B, and the remaining patients of group A show better values than those of group B) than we should carefully describe these numbers. An additional analysis of survival may be required, if the number of drop-outs is considerable, or if differences in the number of drop-outs between groups must be assumed. Recent methodological research focused on the joint modeling of time-to-drop out and longitudinal measurements of a scale variable (cf. [3]).

In any case, drop-outs should be documented and their frequencies verbally described or presented in a flow-chart (see Medical Biostatistics 1).

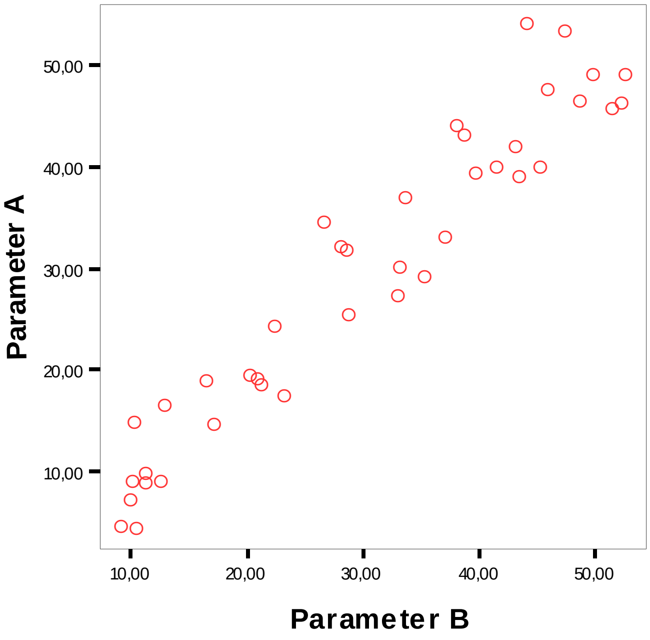
Correlation of two repeatedly measured variables

A commonly asked question in the context of repeated measurements is

“Is there a correlation in changes in parameter A with changes in parameter B?”

The quick and simple solution would involve computing a correlation coefficient between A and B. The problem with that approach is that we have repeated measurements for those parameters for each patient. We cannot assume that values taken from the same patient are independent, which, however, is a crucial assumption underlying the computation of correlation coefficients. Therefore, this violated assumption has to be accounted for. One easy way to account for it is to compute *partial correlation coefficients*. Partial correlation coefficients adjust the correlation between A and B for the different average levels of the patients.

Consider the following (hypothetical) example. Let there be two parameters A and B, each measured in 8 repeated assessments of 5 patients. Ignoring the ‘patient effect’, we obtain the following scatter plot and correlation coefficient:

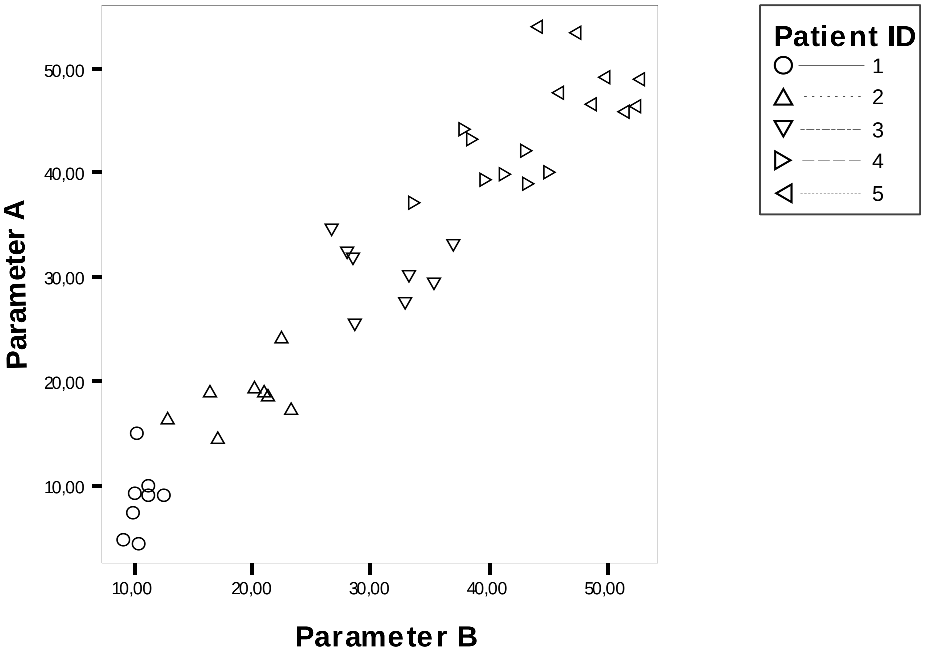


Correlations

		Parameter A	Parameter B
Parameter A	Pearson Correlation	1	,960(**)
	Sig. (2-tailed)		,000
	N	40	40
Parameter B	Pearson Correlation	,960(**)	1
	Sig. (2-tailed)	,000	
	N	40	40

** Correlation is significant at the 0.01 level (2-tailed).

A proper statistical analysis has to adjust for the ‘patient effect’. If we mark the data points in the scatter plot grouping the measurements on each different patient, we clearly see that the parameters are only correlated because if in a patient parameter A is high, then also B is high, but if in the same patient the parameter A changes, then parameter B may increase or decrease. There is absolutely no systematic correlation in the intra-individual changes of parameters A and B:



The partial correlation coefficient is -0.013 (P=0.942), so there is no correlation of *intra-individual changes* of A and B:

Correlations

Control Variables			Parameter A	Parameter B
dum1 & dum2 & dum3 & dum4	Parameter A	Correlation	1,000	-,013
		Significance (2-tailed)	.	,942
		df	0	34
	Parameter B	Correlation	-,013	1,000
		Significance (2-tailed)	,942	.
		df	34	0

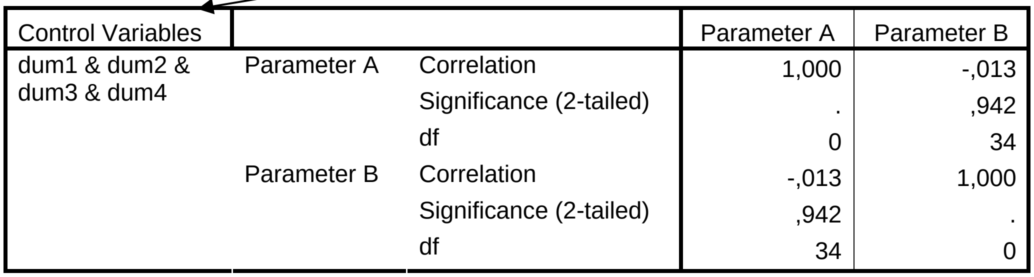
Technically, it is computed by requesting a partial correlation. The menu dialogue asks us for specifying the adjusting variables. These can be binary or scale variables. As ‘patient’ is a nominal variable (neither binary nor scale), we must transform it into several binary variables, which are called ‘dummy variables’. There are several options to define the values of these dummy variables. The most common option is the ‘reference category’ coding:

- Define a reference category.
- With N levels, create N – 1 dummy variables referring to all other categories.
- The dummy variables are all 0 for the reference category.
- For any other category, the corresponding dummy variable is 1.

In our example, the variable ‘PatID’ has 5 levels: 1, 2, 3, 4, 5. Let’s define 5 as the reference level (without loss of generality). The dummy variables are therefore defined as follows:

PatID	Dummy1	Dummy2	Dummy3	Dummy4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	0	0	0	0

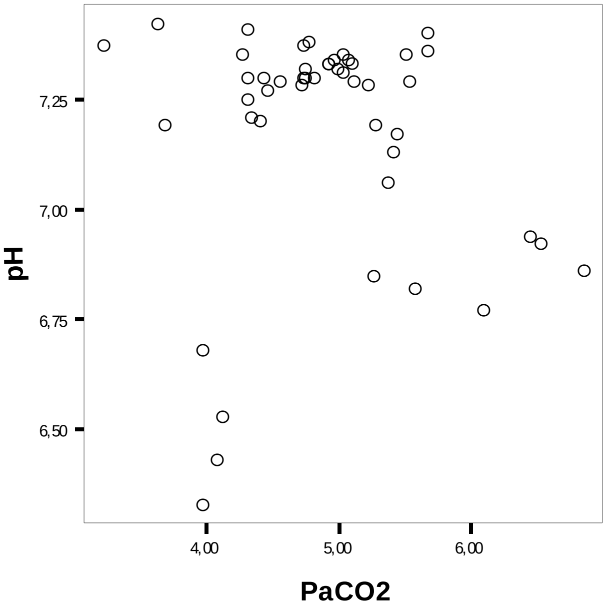
Now the set of dummy variables can be used to adjust for the patient effect in a partial correlation analysis (labeled by ‘control variables’ in the SPSS output).



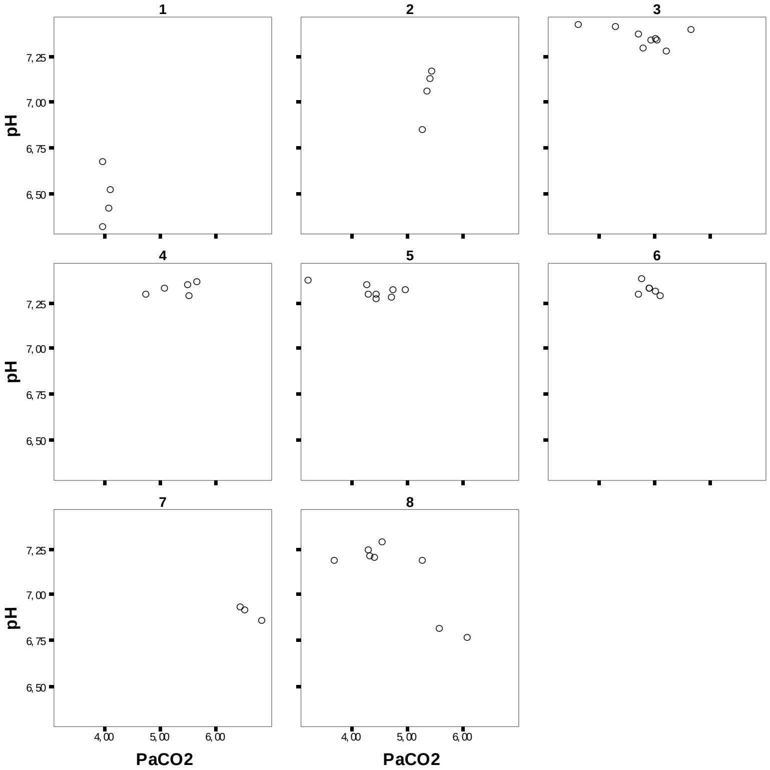
Control Variables			Parameter A	Parameter B
dum1 & dum2 & dum3 & dum4	Parameter A	Correlation	1,000	-,013
		Significance (2-tailed)	.	,942
		df	0	34
	Parameter B	Correlation	-,013	1,000
		Significance (2-tailed)	,942	.
		df	34	0

Another example: correlation of changes in serum pH and serum $PACO_2$ [4]. In eight patients, repeated measurements on serum pH and serum PACO2 have been taken, with a varying number of measurements per patient. The scientific question was to evaluate a correlation of changes in pH and PACO2 (does PACO2 change if pH changes?).

In a marginal scatter plot (ignoring the patient factor), we see no correlation between pH and PaCO2 (a correlation coefficient is computed as -0.065, but please note that this value is wrong because of the ignored dependence of values that are measured on the same patient)



A panel scatterplot shows the data cloud for each of the eight patients:



The partial correlation in this example computes to -0.507 (p=0.001):

Correlations

Control Variables			pH	PaCO2
dum1 & dum2 & dum3 & dum4 & dum5 & dum6 & dum7	pH	Correlation	1,000	-,507
		Significance (2-tailed)	.	,001
		df	0	38
	PaCO2	Correlation	-,507	1,000
		Significance (2-tailed)	,001	.
		df	38	0

In summary, correlation of original data is wrong here, and should be replaced by partial correlation.

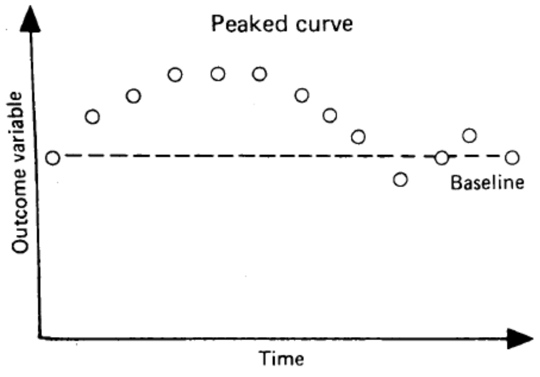
Class 11: Summary measures

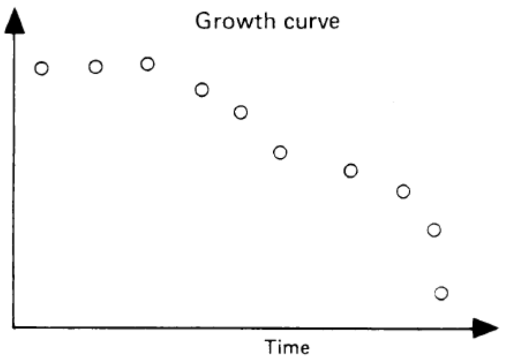
Before we move to the discussion of repeated measurements-ANOVA, which models all repeated measurements, an approach is presented which circumvents the formulation of an explicit model, and proves useful in a variety of situations.

The key idea of this approach is to compute a summary measure out of each individual curve, which summarizes all the important information into one number per patient. This measure can then be compared between groups, either in a crude comparison or in a baseline-adjusted analysis.

Matthews et al [5] identified two different typical shapes of curves:

- *growth curves* have their maximum at the end or at the beginning of the time course, and
- *peaked curves*, which have a peak somewhere in the middle of the time course.





The following summary measures may be used for growth curves:

- Last value
- Difference between last and first value, either computed from raw values or from a regression analysis
- Slope of the growth curve

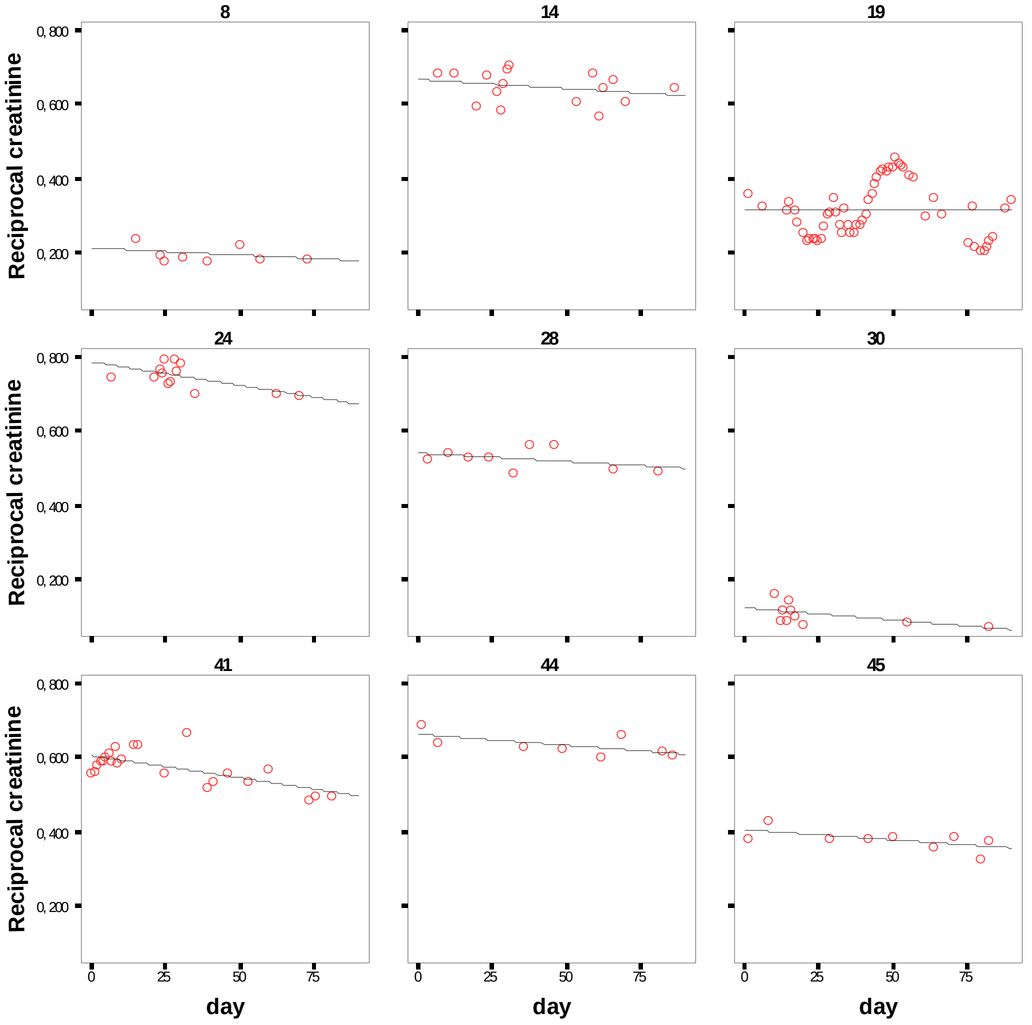
Sometimes measurements are taken repeatedly to describe the course of the concentration of a drug in the body. To describe such peaked curves the following summary measures are useful:

- Maximum concentration (C_{max})
- Time to maximum concentration (T_{max})
- Minimum concentration (C_{min}) between two subsequent applications
- Time above a threshold value
- Area under the curve, which is used as a measure of drug absorption
- Half-time: time to half of maximum concentration ($T_{1/2)}$

Example: slope of reciprocal creatinine values

A decline in kidney function after a kidney transplantation is indicated by a rise in creatinine measurements. Usually, such a rise is detected by plotting reciprocal creatinine values (1/Creatinine) against time. In terms of the taxonomy presented by Matthews et al, such reciprocal creatinine courses are growth curves. We can therefore use linear regression on the individual lines to obtain two pieces of information:

- The slope of the regression lines can be used to describe the kidney function
- The values predicted from the regression line at particular days (say, day90 after transplantation) may serve as a daily-fluctuation corrected value to be used in further analysis, e. g., for predicting the long-term outcome of the kidney function.



Using SPSS, we can compute for each patient the slope of the regression line and an estimate of the (reciprocal and original-scale) creatinine value at day 90 after transplantation. The table below shows these values. The slopes are computed per 90 days, i. e they describe the decline in reciprocal creatinine from day 0 to day 90:

Pat ID	Predicted creatinine value at day 90	Slope per 90 days
8	5.63	-0.0339
14	1.60	-0.0436
19	3.18	-0.0005
24	1.49	-0.1140
28	2.00	-0.0405
30	15.80	-0.0581
41	2.01	-0.1092
44	1.64	-0.0534
45	2.81	-0.0494

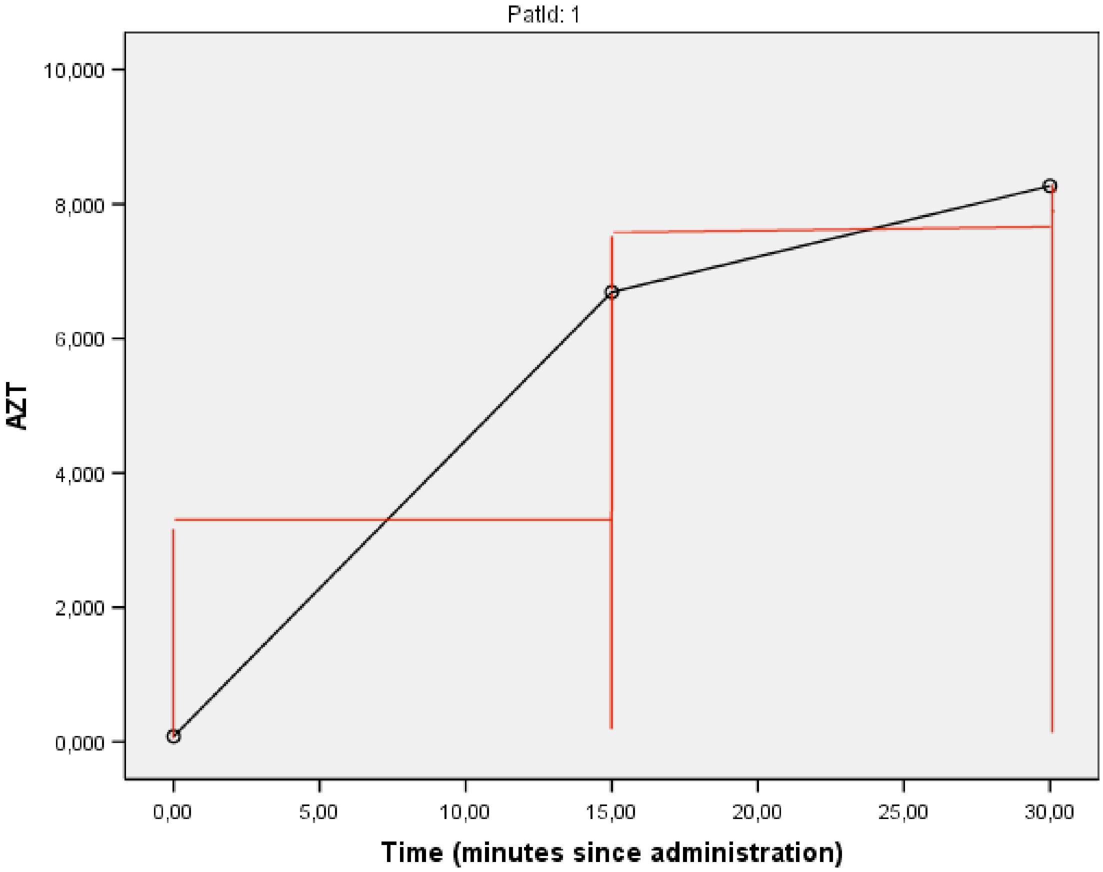
Example: area under the curve

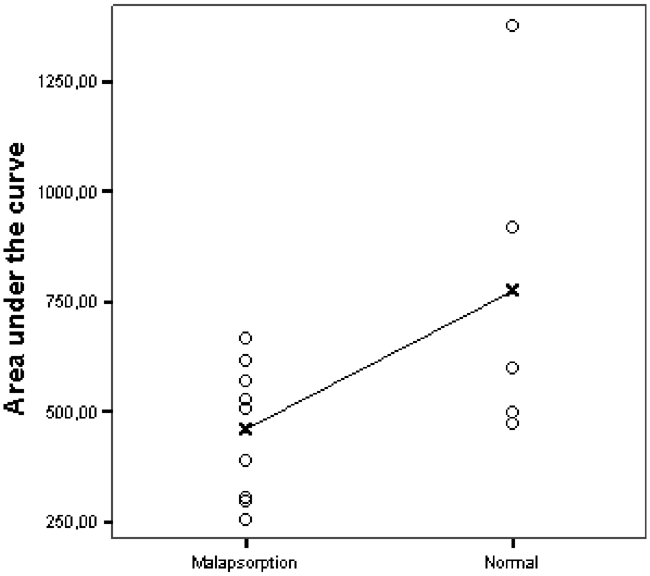
Consider the AZT data of Lec 10. The area under the curve can be computed using the trapezoidal rule. The area under the curve between two subsequent time points T1 and T2 is based on an assumed linear course between these two time points, and thus estimated by:

A12 = (X2 + X1) (T2 – T1) / 2

Where X1 and X2 denote the concentrations at times T1 and T2, respectively. The area under the curve (AUC) is obtained by summing up all partial areas between subsequent time points.

The diagram below shows the way the area is computed using the trapezoidal rule on the first three measurements on patient 1. By replacing the trapezoidal sections (under the black line) by rectangulars (under the red line) of average height, the area can easily be computed using the formula length*height.





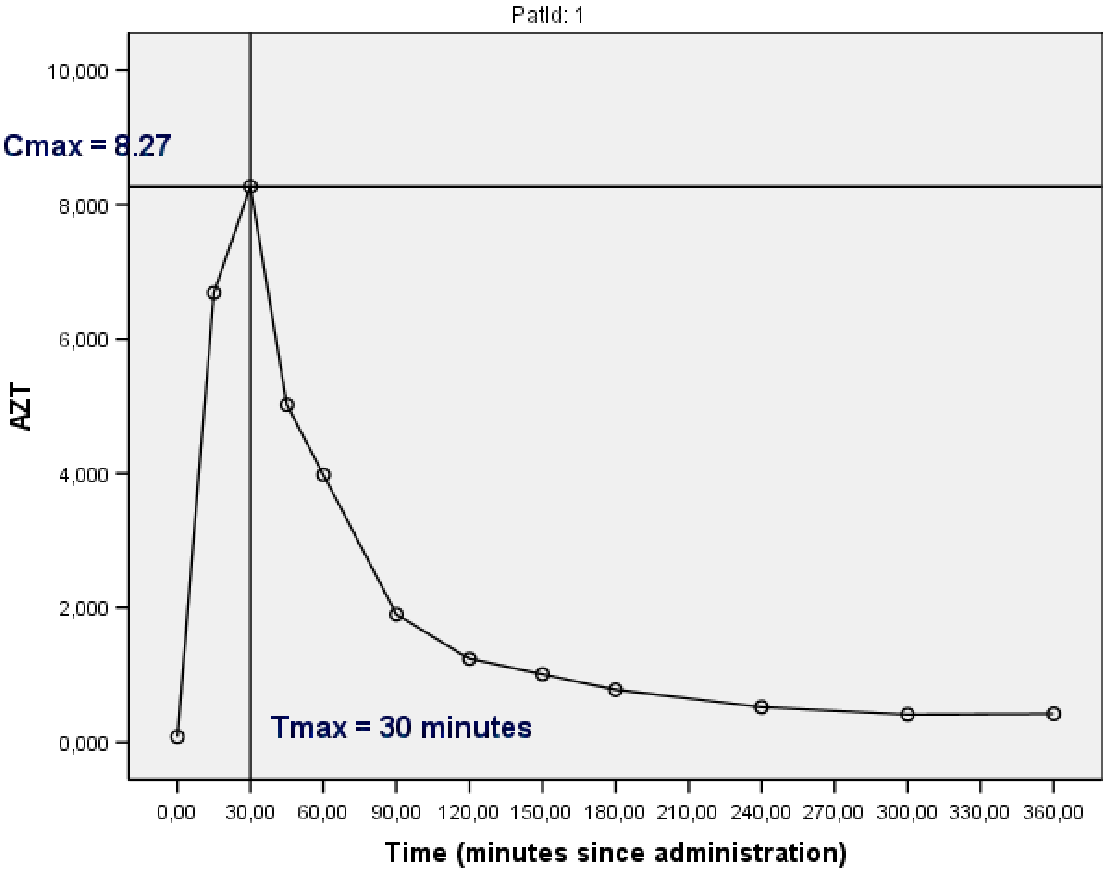
Comparing the area under curve between the patient groups, we notice higher values in patient with normal fat absorption. A t-test on log-transformed area under the curve values reveals a p-value of 0.045, and the ratio of geometric means is 1.62 (95% confidence interval 1.01 – 2.61), meaning that the AUC is 1.6 times higher in patients with normal fat absorption than in patients with malabsorption.

The area under the curve can be computed in SPSS using the instructions given in the SPSS lab section.

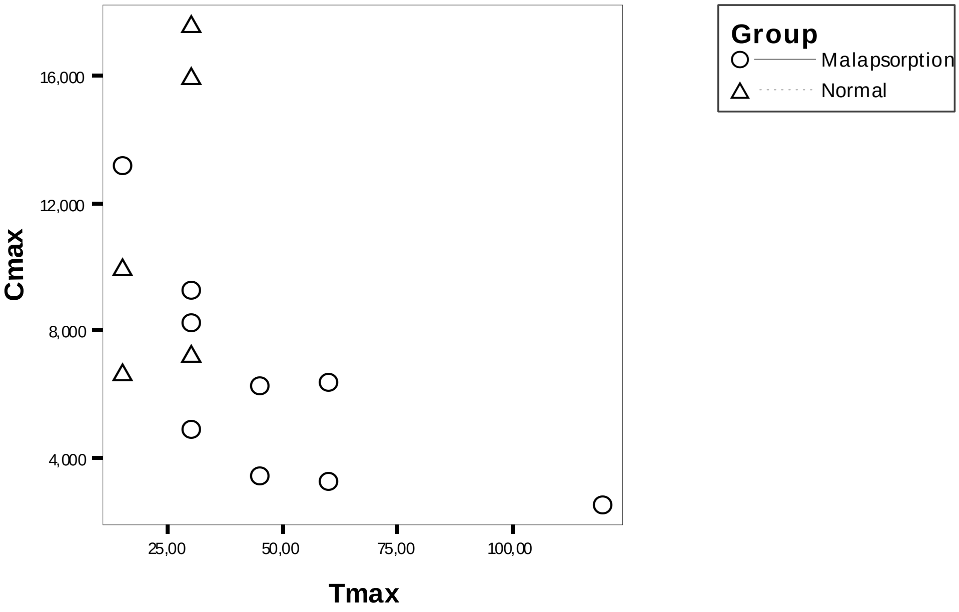
Example: Cmax vs. Tmax

From the individual AZT curves, we can distill the maximum concentration Cmax and the time to maximum concentration Tmax in order to find associations between the rise time and the maximum absorption of the drug.

For patient 1, we observe a Cmax of 8.27 occuring at a Tmax of 30 minutes:



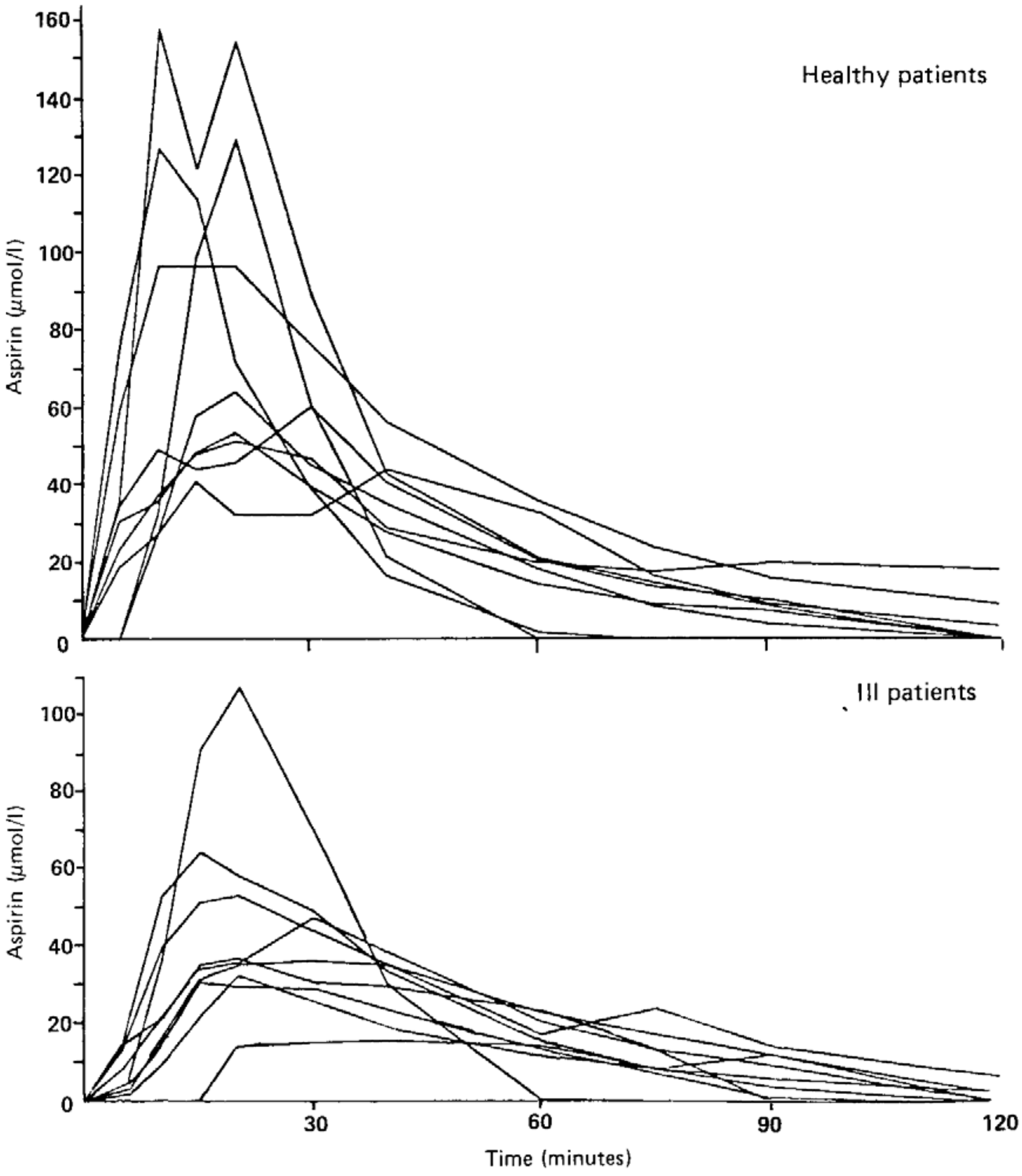
Computed from all patients, we notice a negative correlation between Tmax and Cmax (high maximum concentrations occur earlier than low maximum concentrations):



Example: aspirin absorption

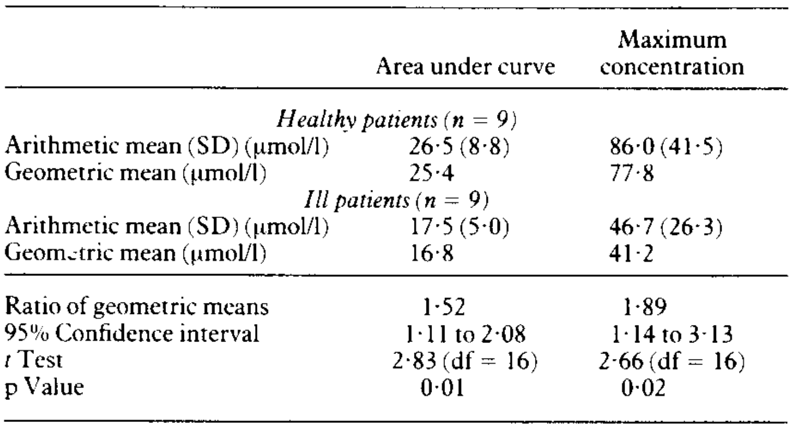
Another example, reproduced from [5], deals with aspirin absorption in healthy and ill persons. The basic research question was: Do ill persons have reduced aspirin absorption?

The individual aspirin concentrations are depicted in the following plot:



We could answer the research question by using two summary measures: the area under the curve (AUC) and the maximum value (Cmax). Both are meaningful and familiar to pharmacologists. The table below gives statistics of these two summary measures for comparing the groups. Since the distributions of both summary measures are skewed, a data transformation or use of a non-parametric method to compare groups is indicated:

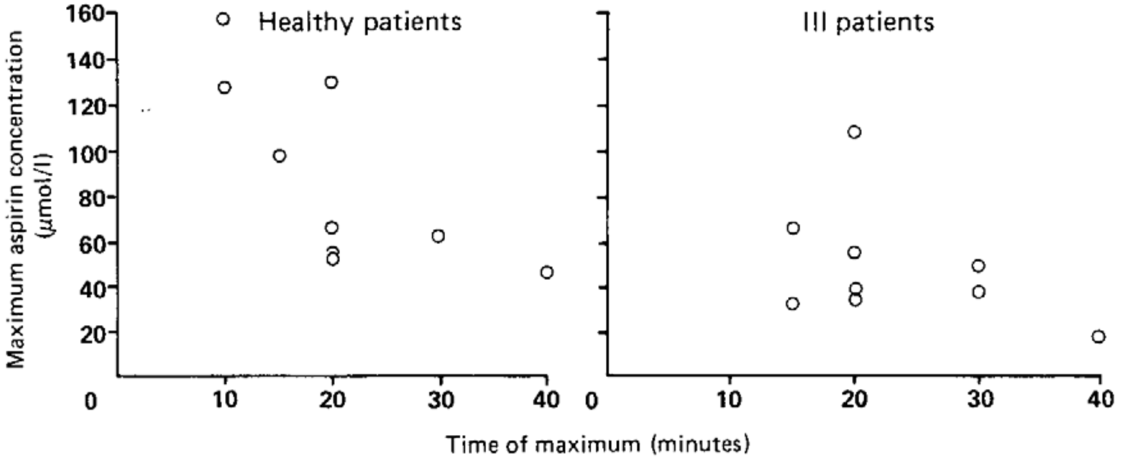
TABLE I—*Analysis of data from aspirin study*



	Area under curve	Maximum concentration
Healthy patients (n = 9)		
Arithmetic mean (SD) (μmol/l)	26·5 (8·8)	86·0 (41·5)
Geometric mean (μmol/l)	25·4	77·8
Ill patients (n = 9)		
Arithmetic mean (SD) (μmol/l)	17·5 (5·0)	46·7 (26·3)
Geometric mean (μmol/l)	16·8	41·2
Ratio of geometric means	1·52	1·89
95% Confidence interval	1·11 to 2·08	1·14 to 3·13
t Test	2·83 (df = 16)	2·66 (df = 16)
p Value	0·01	0·02

As can be seen, the standard deviation increases with increasing mean (8.8 vs. 5.0), therefore the summary measures should be log transformed and then group difference expressed as the ratio of the geometric means with 95% confidence interval. There is strong evidence that higher maxima and higher overall values occur in healthy persons than in ill persons.

The next figure shows the Cmax values plotted against the Tmax values. This plot clearly shows that the maximum concentrations tend to be lower and occur later in the ill patients. Additionally, the Cmax appears to be negatively correlated with Tmax, i. e., higher peaks occur earlier than lower peaks.



Class 12: ANOVA for repeated measurements

Extension of one-way ANOVA

ANOVA compares the means of a scale variable between levels of a factor. In its simplest setting, there are two levels, which means that ANOVA is actually a two-group comparison and could be replaced by a t-test.

A basic assumption of ANOVA is that all observations are mutually independent. This assumption is violated, if several measurements are taken from each subject. We cannot treat these repeated measurements as being independent. ANOVA for repeated measurements provides a toolbox to account for the dependence of the observations taken on the same individuals.

Between-subject and within-subject effects

So far, we have only considered effects which relate to experimental conditions (e. g., a comparison of treatment A with treatment B) that were varied between individuals. Some individuals were randomized to receive A, others to receive B. However, there was no subject receiving both. Thus, we call ‘treatment’ in this simple example a ‘between-subject effect’.

What is a within-subject effect?

By contrast, a within-subject effect is an effect that varies within subjects.

Example: consider an ophthalmologic trial, where two types of lenses are implanted in the same subjects, and the outcome is a visus assessment one week after the operation. Since each subject receives both lenses, the factor ‘lens’ is a within-subject effect. A proper statistical analysis must account for the within-subject nature of this effect, if no covariates are present, a paired t-test could be done, or alternatively, type of lens could be specified as within-subject effect in an ANOVA.

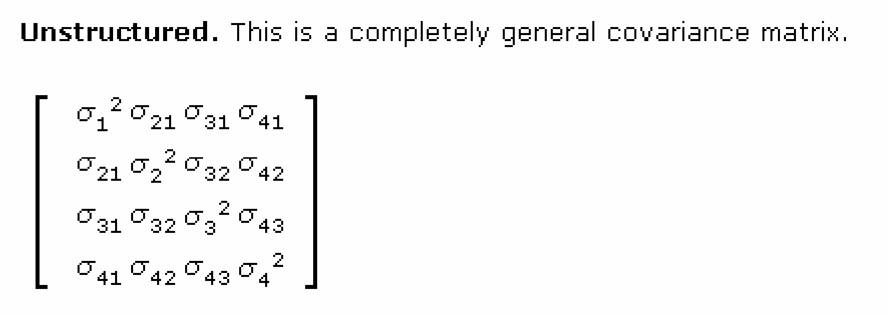
If we take several serial measurements on the same individual, then these measurements differ by the time at which they were taken. Therefore, the factor ‘time’ constitutes a within-subject effect: it is varied within subjects.

Consider a trial where serial measurements of pain (on a visual analogue scale, VAS) are taken on the same patients after start of a pain treatment: patients are randomized either to electro-stimulated acupuncture or to standard acupuncture. In this example, we have two types of effects: time is a within-subject effect, type of acupuncture a between-subject effect.

Specifying a within-subject covariance structure

It was already outlined that a proper statistical analysis must account for the inherent dependence of observations taken on the same subjects. This can be done by specifying a particular structure that is assumed to describe the correlation of observations on the same individual. The program then estimates the parameters of this structure. We call such a structure the covariance structure. Several covariance structures are available, among these four are introduced here:

Unstructured covariance



Unstructured. This is a completely general covariance matrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

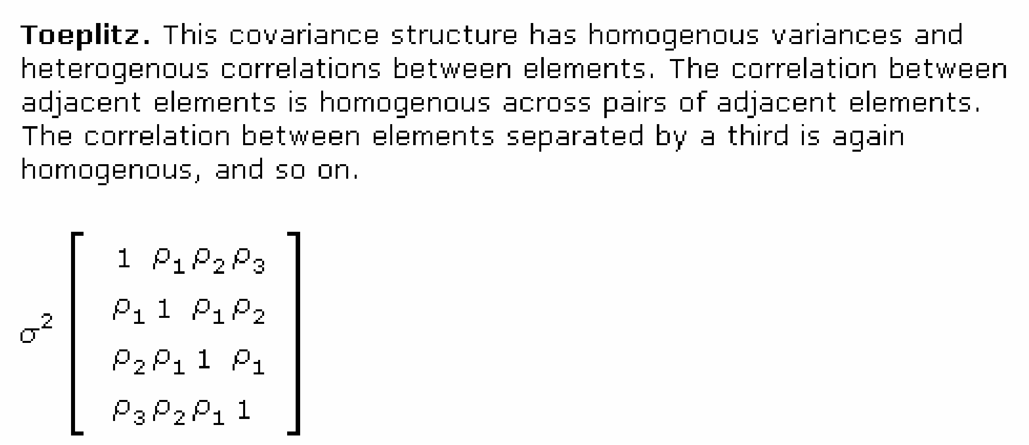
The rows and columns in the matrix shown above correspond to the time points at which measurements are taken. Between each pair of time points there exists a correlation (or covariance), and using an unstructured covariance matrix, we estimate each covariance separately. With N time points, we end up with N(N+1)/2 covariance parameters that have to be estimated. Of course, this structure offers highest flexibility. Variances (standard deviations) are allowed to differ from time point to time point. However, since many parameters have to estimated, it can only be applied if the number of measurements is small and the number of subjects comparably high, otherwise the ‘1:10’ rule of thumb would be violated and estimates highly unstable.

Since these conditions do not always hold in real life, several covariance structures have been proposed that simplify the unstructured covariance, but at the cost of lower flexibility.

Toeplitz covariance structure

The Toeplitz structure imposes two restrictions on the covariance matrix:

- It assumes the same variance for each time point,
- it assumes that the correlations between subsequent measurements are the same,
- it assumes that the correlations between measurements separated by a third are the same, etc.



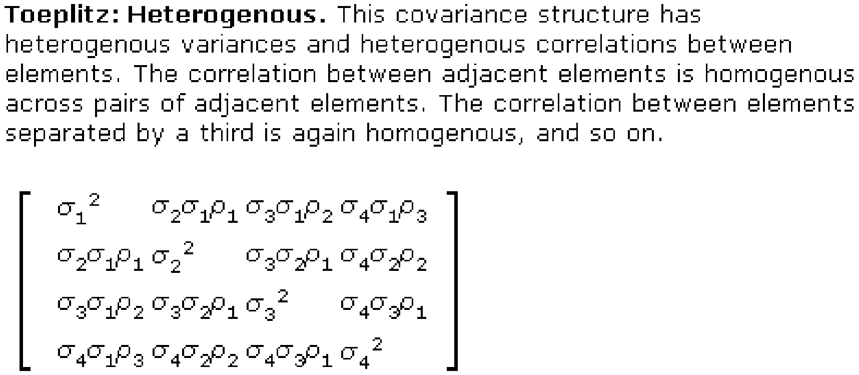
Toeplitz. This covariance structure has homogenous variances and heterogenous correlations between elements. The correlation between adjacent elements is homogenous across pairs of adjacent elements. The correlation between elements separated by a third is again homogenous, and so on.

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

This covariance structure estimates N parameters, and is thus much more parsimonious than the unstructured covariance. However, the assumptions that are imposed may not always hold.

Toeplitz-Heterogenous

To relax the assumption of equal variances at each time point, one may choose Toeplitz-heterogenous structure:

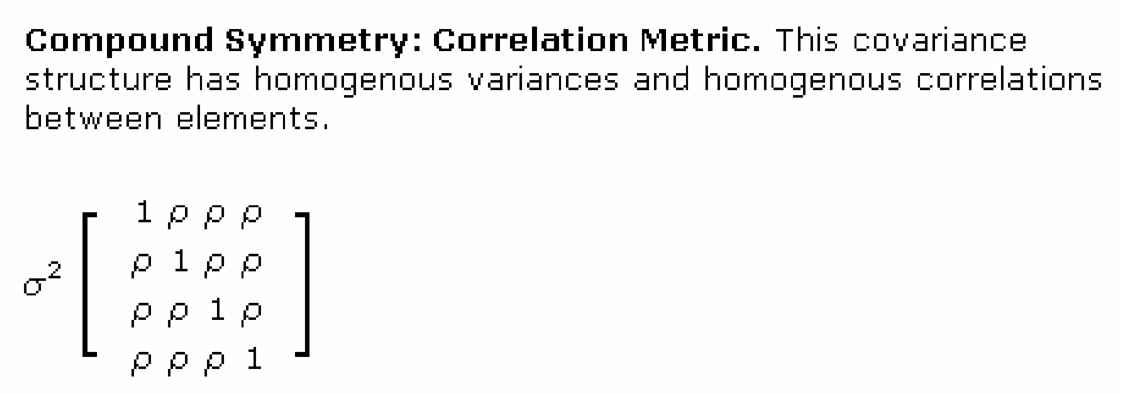


Toeplitz: Heterogenous. This covariance structure has heterogenous variances and heterogenous correlations between elements. The correlation between adjacent elements is homogenous across pairs of adjacent elements. The correlation between elements separated by a third is again homogenous, and so on.

$$\begin{bmatrix} \sigma_1{}^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_2 & \sigma_4\sigma_1\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2{}^2 & \sigma_3\sigma_2\rho_1 & \sigma_4\sigma_2\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3{}^2 & \sigma_4\sigma_3\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4{}^2 \end{bmatrix}$$

This covariance structure needs 2N-1 covariance parameters, thus it is about twice as ‘costly’ as the standard Toeplitz structure, yet it provides a much more flexible way to define the covariance.

Compound symmetry structure



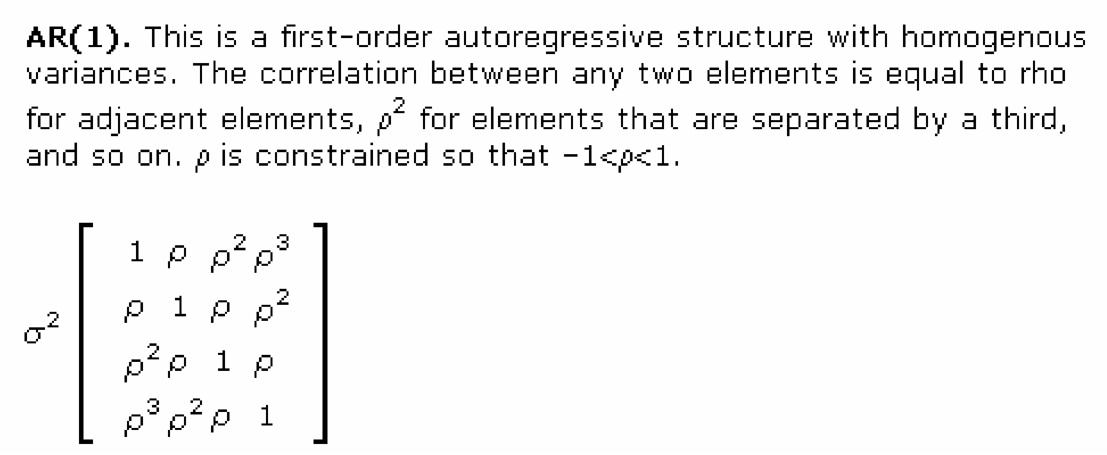
Compound Symmetry: Correlation Metric. This covariance structure has homogenous variances and homogenous correlations between elements.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Compound symmetry means that we are assuming the same variance (standard deviation) of measurements at each time point, and that we are assuming the same correlation between measurements taken at different times. This covariance structure may be most adequate if we have a stable-over-time process, e. g. with blood pressure measurements, or if there is a relatively long time elapsing between two subsequent measurements such that we can assume that a precedent measurement does no longer take influence on a subsequent one. Only two parameters have to estimated: the variance of the measurements and the within-subject correlation.

Autoregressive covariance structure

The autoregressive (AR(1)) covariance structure can be used, if high autocorrelation between measurements must be assumed, which fades out with time:



AR(1). This is a first-order autoregressive structure with homogenous variances. The correlation between any two elements is equal to rho for adjacent elements, ρ^2 for elements that are separated by a third, and so on. ρ is constrained so that $-1<\rho<1$.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

We see that since rho is a number between -1 and 1, the correlation drops off with increasing time between two measurements. However, the magnitude of this drop-off may not be adequately modeled with this structure. On the other hand, similarly to the compound symmetry structure, only two parameters are estimated.

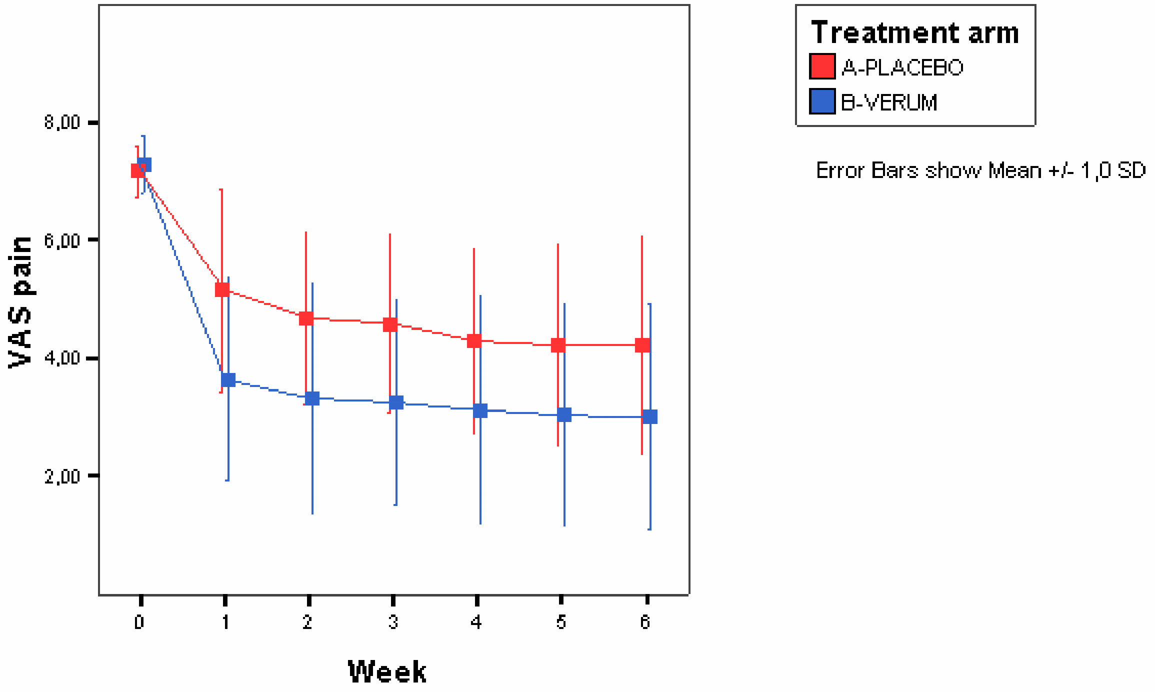
Specification of a RM-ANOVA

The specification of the covariance structure is the most crucial part of an RM-ANOVA analysis. All other specifications are very similar to simple ANOVA or regression analysis.

Model formulation

The most simplest setting assumes a between-subject factor like treatment arm, and time as within-subject facture (time). Additionally, we may consider covariates such as baseline measurements or demographic variables (age, sex) that the treatment effect should be adjusted for.

As an example, consider an acupuncture trial where two acupuncture techniques (electro-stimulated, standard) are compared in their effectiveness to remove cervical pain [6]. Patients were randomized to either electrical or standard acupuncture. Pain was measured by a visual analogue scale (VAS): before start of treatment, and one, two, …, six weeks after start of treatment:



Conducting a repeated measures ANOVA, we have the following effects:

Time	Within-subject
Treatment	Between-subject

Additionally, we should consider a possible **Time by treatment interaction**, which can be dropped from the model if it is not statistically significant. A significant time-by-treatment interaction would indicate that at each time point, the difference of the VAS scores between the two treatment arms is different. If there is not such interaction, we can assume the same difference between treatment arms throughout the follow-up period (which simplifies the interpretation of results a lot!).

If we observe a significant effect of time, we may conduct post-hoc comparisons of measurements taken at different times. Of course, such post-hoc comparisons must be properly corrected for the inherent multiple testing.

Now the question arises which covariance structure is the most adequate. While the unstructured covariance matrix provides the most flexible fit to the observed data, it requires estimation of many additional parameters (the elements of the covariance matrix), which we are not really interested in. Such parameters are called nuisance parameters. Generally, estimation of many parameters with few independent subjects should be avoided, because results can be unstable. By contrast, the Toeplitz or the AR(1) structures are less flexible, but more parsimonious, as they require less parameters to be estimated. The so-called ‘Akaike information criterion’ (AIC) can be used to decide whether to use a more flexible or a more restrictive covariance matrix. It penalizes the likelihood of the model (i. e., the probability of the observed data given the estimated model parameters) by the number of estimated parameters. AIC is supplied in the program output and by running the analysis with different covariance structures, one may compare AIC und choose that covariance structure that yields the smallest AIC (by convention, AIC is scaled such that small numbers indicate better fit).

In our example, we have:

Covariance structure	-2 log Likelihood	Model parameters	AIC	
Unstructured	167.835	34	209.835	
Toeplitz	200.511	19	212.511	
AR(1)	203.466	15	207.466	← minimum AIC!

We see that the unstructured covariance requires the estimation of additional 19 parameters (compared to AR(1)), which initially yields a better likelihood value. After penalization for the number of parameters, AR(1) has the optimal AIC value. Therefore, interpretation of results should be based on the AR(1) covariance structure.

Therefore, we proceed using the AR(1) structure. We adjust for the **baseline VAS** assessment (denoted by vas0_mean in the SPSS output below) by including baseline VAS as covariate (also considered a between-subject effect).

With the AR(1) covariance we obtain the following table of global tests:

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	17,648	,200	,660
group	1	18,262	4,131	,057
Week	5	85,740	2,248	,057
group * Week	5	85,740	,373	,866
vas0_mean	1	17,655	,024	,879

a Dependent Variable: VAS.

For the interaction test, the procedure calculates a p-value of 0.866, which means that the interaction term should be dropped from the model:

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	17,679	,202	,659
group	1	17,841	3,913	,064
Week	5	90,740	2,286	,053
vas0_mean	1	17,685	,024	,879

a Dependent Variable: VAS.

We see that the effect of the baseline measurements is not significant, nevertheless this effect is retained in the analysis (it is not of importance whether this effect is significant or not, we only want to adjust for it).

Although both treatment effect and time effect (row labeled ‘Week’) are not significant at the usual 5% level, we take a look at the parameter estimates:

Estimates of Fixed Effects[b]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	1,993854	6,780319	17,685	,294	,772	-12,269273	16,256981
[group=A-PLACEBO]	1,593299	,805477	17,841	1,978	,064	-,100024	3,286623
[group=B-VERUM]	0[a]	0	.	.	.	.	.
[Week=1]	,575499	,213562	99,717	2,695	,008	,151783	,999215
[Week=2]	,348734	,193197	97,853	1,805	,074	-,034666	,732134
[Week=3]	,312273	,168507	95,624	1,853	,067	-,022227	,646774
[Week=4]	,100772	,138565	93,290	,727	,469	-,174379	,375923
[Week=5]	,012948	,098677	90,874	,131	,896	-,183065	,208960
[Week=6]	0[a]	0	.	.	.	.	.
vas0_mean	,142963	,929271	17,685	,154	,879	-1,811858	2,097784

a. This parameter is set to zero because it is redundant.

b. Dependent Variable: VAS.

The VAS measurements in the PLACEBO group are about 1.6 cm higher than in the VERUM group. Since there is no interaction with time, we can assume that the effect is constant over time (as can also be seen from the error bar plot above). Regarding the time effect, the program automatically assumes the last week as the reference category, and all other time points are compared to week 6. In week 1, the measurements are on average 0.56 cm higher, and decline thereafter. After adjusting for multiple testing (multiplying the p-values by 5), the only significant difference is between week 1 and week 6. Thus, we conclude the final effect of treatment is not yet attained after the first week. However, significances of comparisons of time points should not be overinterpreted. With a larger sample size, we may have obtained a significance in comparing weeks 3 and 6 or weeks 4 and 6. Emphasis should be put on the magnitude of the effect, which clearly shows that VAS measurements decline with ongoing acupuncture treatment.

With some data sets, the unstructured covariance specification may lead to a convergence failure, i. e., the program does not supply reliable results. Such a failure is often indicated by a warning like:

Warnings

```
Iteration was terminated but convergence has not been achieved. The MIXED
procedure continues despite this warning. Subsequent results produced are based on
the last iteration. Validity of the model fit is uncertain.
```

More restrictive covariance specifications may remove this convergence failure, but at the cost of less flexibility.

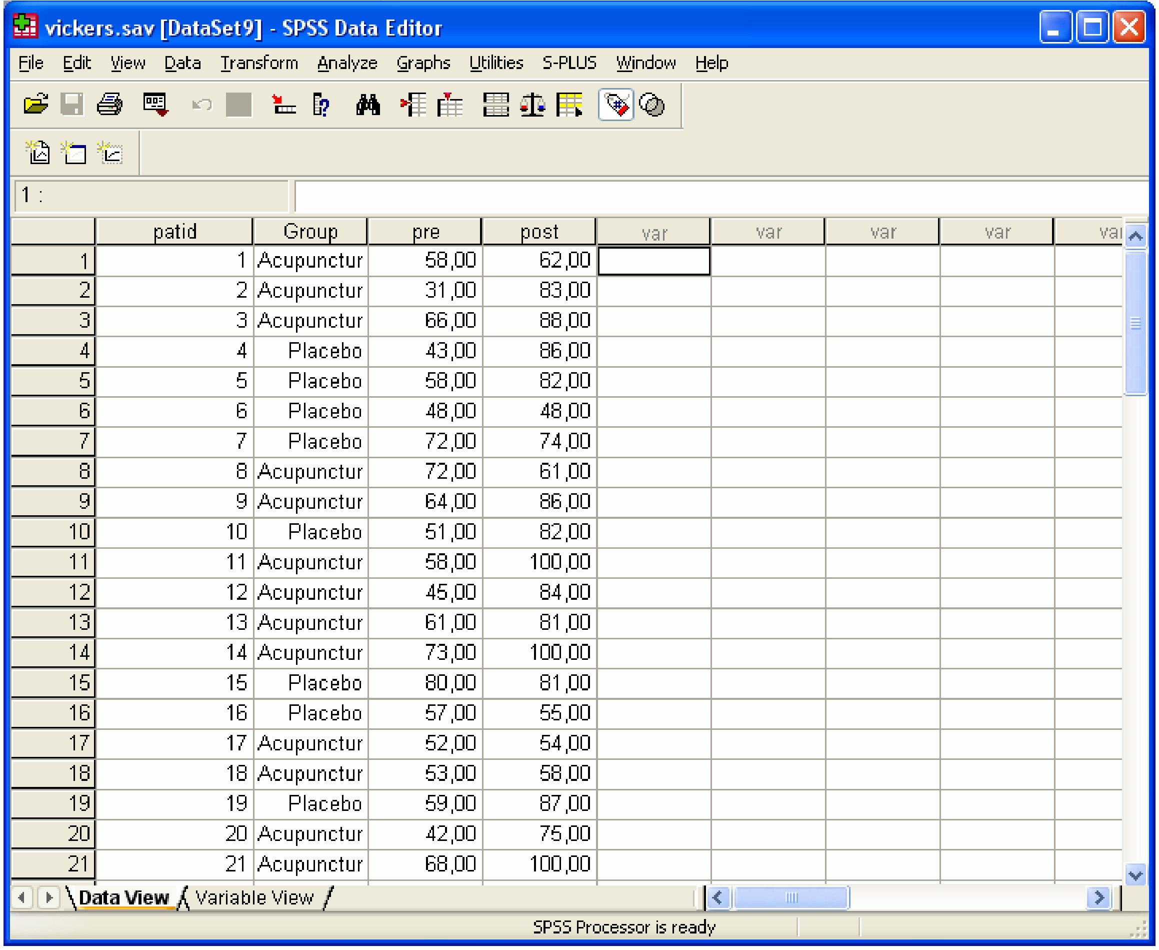
Summarizing, a repeated measures ANOVA can be conducted in the following steps:

1. Check normal distribution in all subgroups and at all time points
2. Specify model, include treatment by time interaction
3. Check AIC using different assumptions of covariance structure, use that structure that yields the smallest AIC
4. Interpret the results; drop interaction from model if not significant

SPSS Lab 3: Analysis of repeated measurements

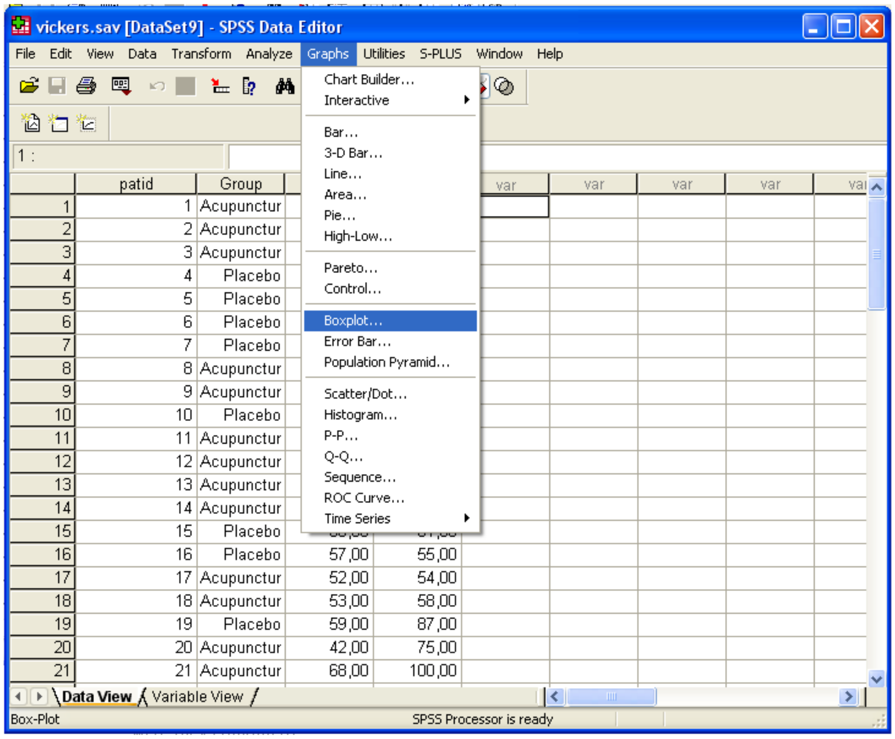
Pretest-posttest data

The pretest-posttest data analysis is exemplified on the data set vickers.sav from ref. [1]:


vickers.sav [DataSet9] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 :
Data View
Variable View
SPSS Processor is ready

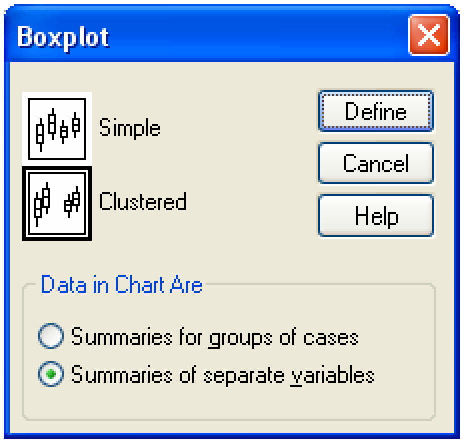
	patid	Group	pre	post	var	var	var	var	vai
1	1	Acupunctur	58,00	62,00					
2	2	Acupunctur	31,00	83,00					
3	3	Acupunctur	66,00	88,00					
4	4	Placebo	43,00	86,00					
5	5	Placebo	58,00	82,00					
6	6	Placebo	48,00	48,00					
7	7	Placebo	72,00	74,00					
8	8	Acupunctur	72,00	61,00					
9	9	Acupunctur	64,00	86,00					
10	10	Placebo	51,00	82,00					
11	11	Acupunctur	58,00	100,00					
12	12	Acupunctur	45,00	84,00					
13	13	Acupunctur	61,00	81,00					
14	14	Acupunctur	73,00	100,00					
15	15	Placebo	80,00	81,00					
16	16	Placebo	57,00	55,00					
17	17	Acupunctur	52,00	54,00					
18	18	Acupunctur	53,00	58,00					
19	19	Placebo	59,00	87,00					
20	20	Acupunctur	42,00	75,00					
21	21	Acupunctur	68,00	100,00					

A grouped boxplot comparing pretest and posttest scores in both acupuncture and placebo groups can be obtained by first selecting Graphs-Boxplot…:



	patid	Group		
1	1	Acupunctur		
2	2	Acupunctur		
3	3	Acupunctur		
4	4	Placebo		
5	5	Placebo		
6	6	Placebo		
7	7	Placebo		
8	8	Acupunctur		
9	9	Acupunctur		
10	10	Placebo		
11	11	Acupunctur		
12	12	Acupunctur		
13	13	Acupunctur		
14	14	Acupunctur		
15	15	Placebo		
16	16	Placebo	57,00	55,00
17	17	Acupunctur	52,00	54,00
18	18	Acupunctur	53,00	58,00
19	19	Placebo	59,00	87,00
20	20	Acupunctur	42,00	75,00
21	21	Acupunctur	68,00	100,00

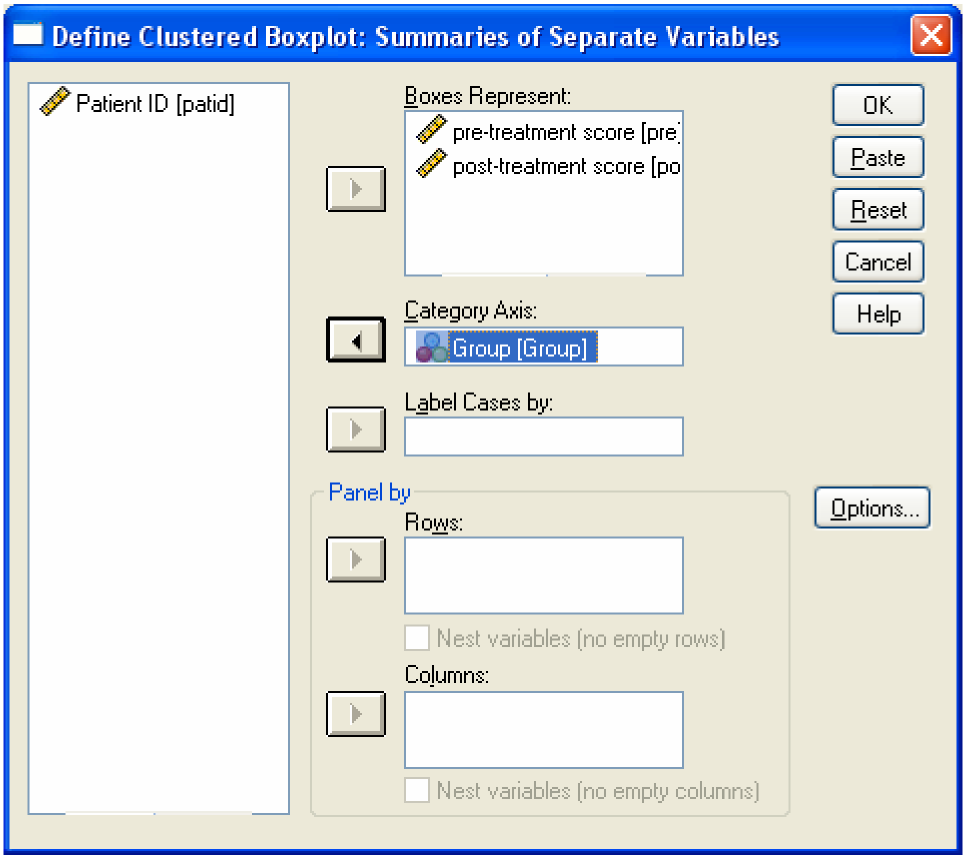
Select ‘Clustered’ and ‘Summaries of separate variables’:



Data in Chart Are

○ Summaries for groups of cases

● Summaries of separate variables



Define Clustered Boxplot: Summaries of Separate Variables

Patient ID [patid]

Boxes Represent:

pre-treatment score [pre]

post-treatment score [po

Category Axis:

Group [Group]

Label Cases by:

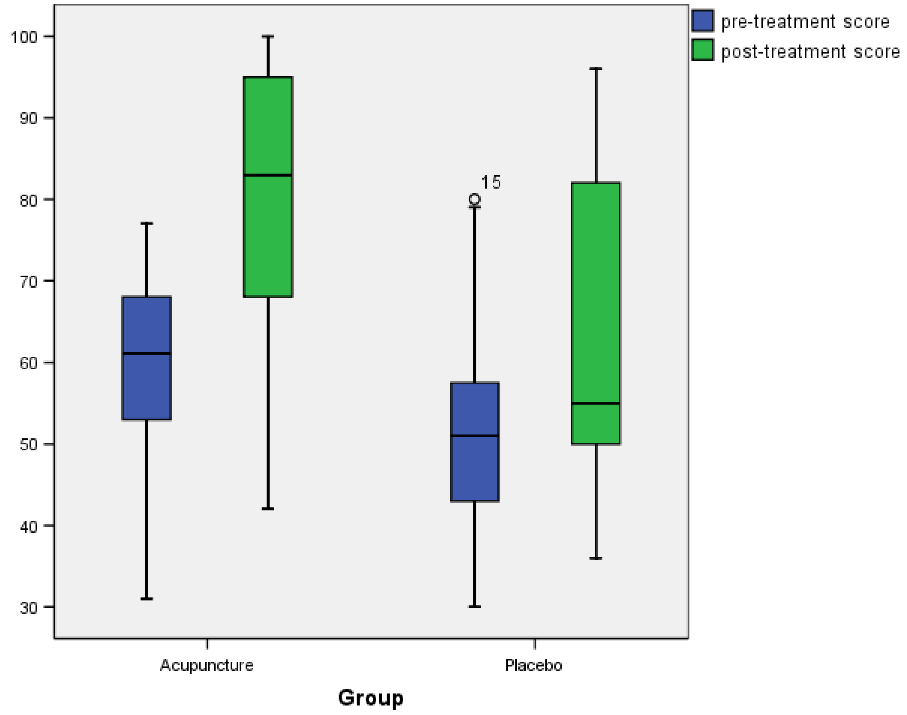
Panel by

Rows:

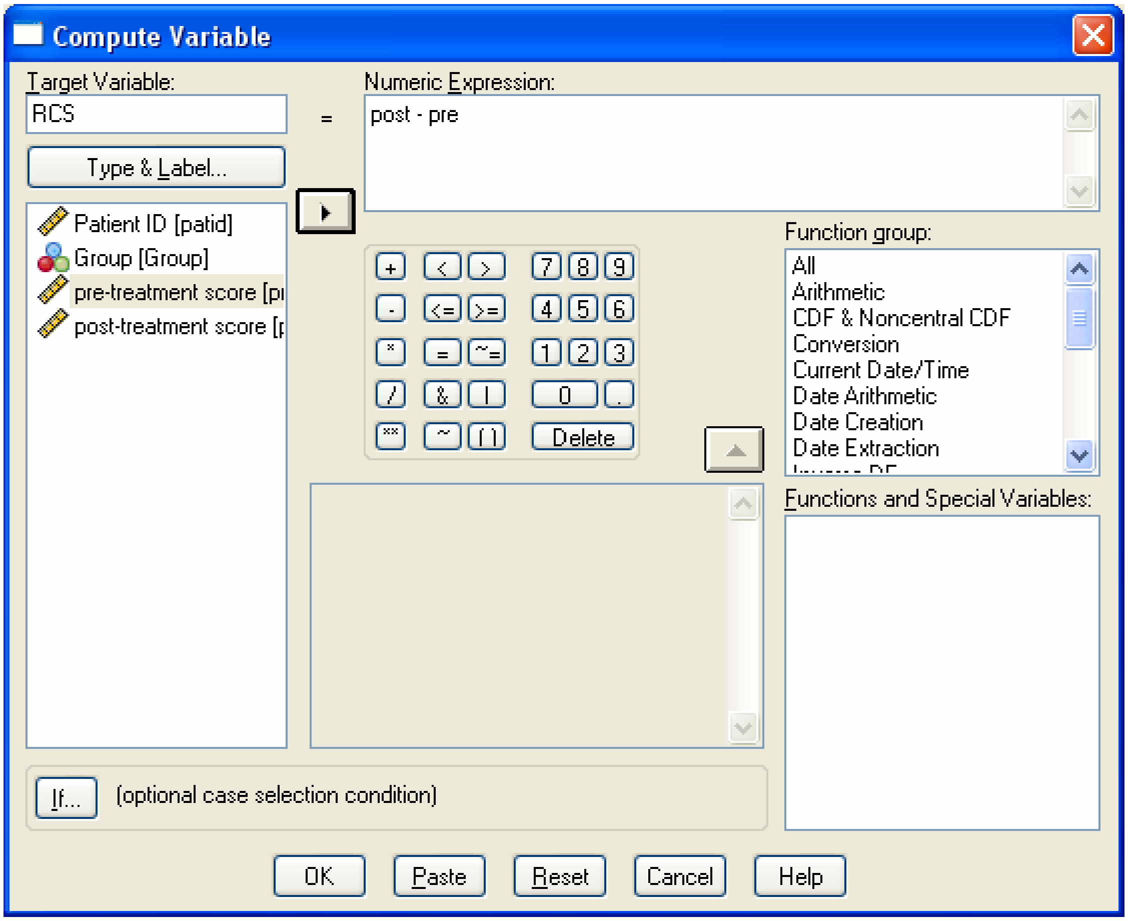
□ Nest variables (no empty rows)

Columns:

□ Nest variables (no empty columns)



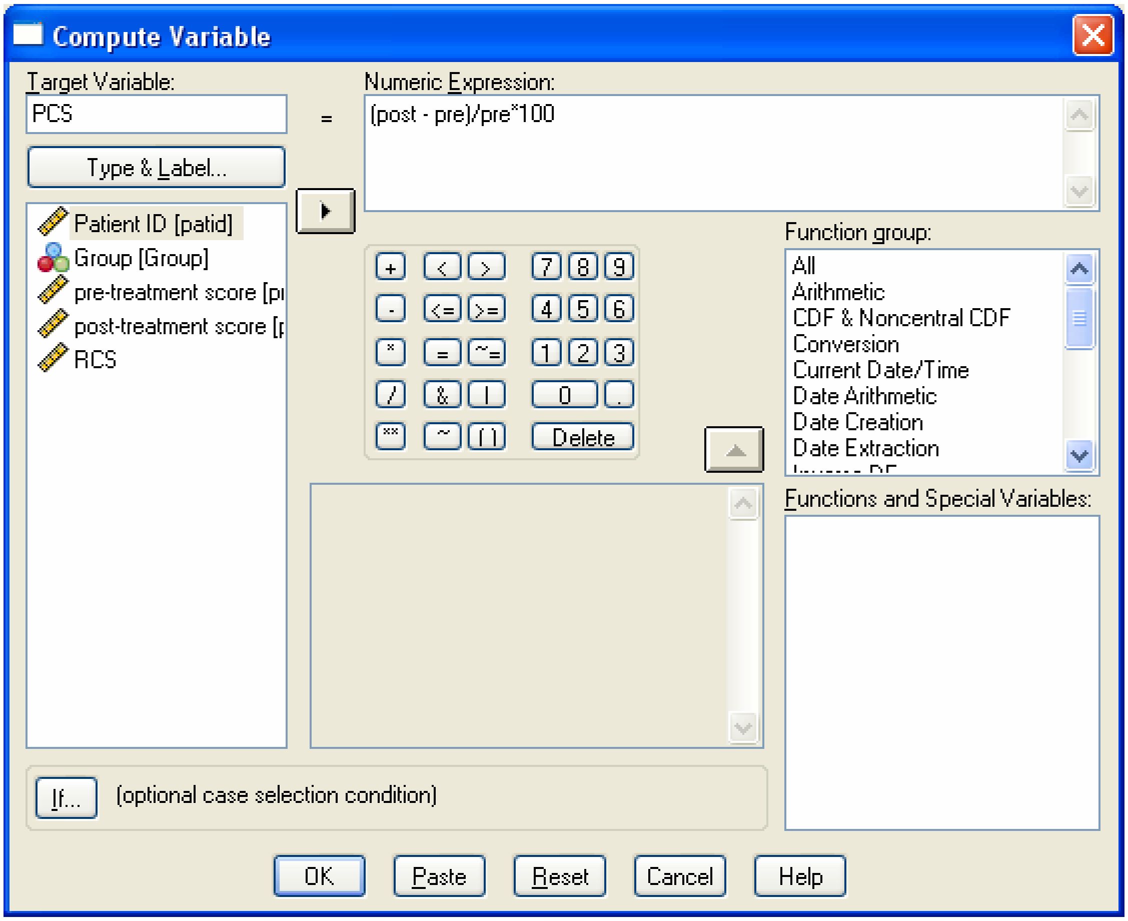
Computation of change scores can be done using the Transform-Compute menu dialogue. For raw change scores, specify:



Target Variable: RCS

Numeric Expression: post - pre

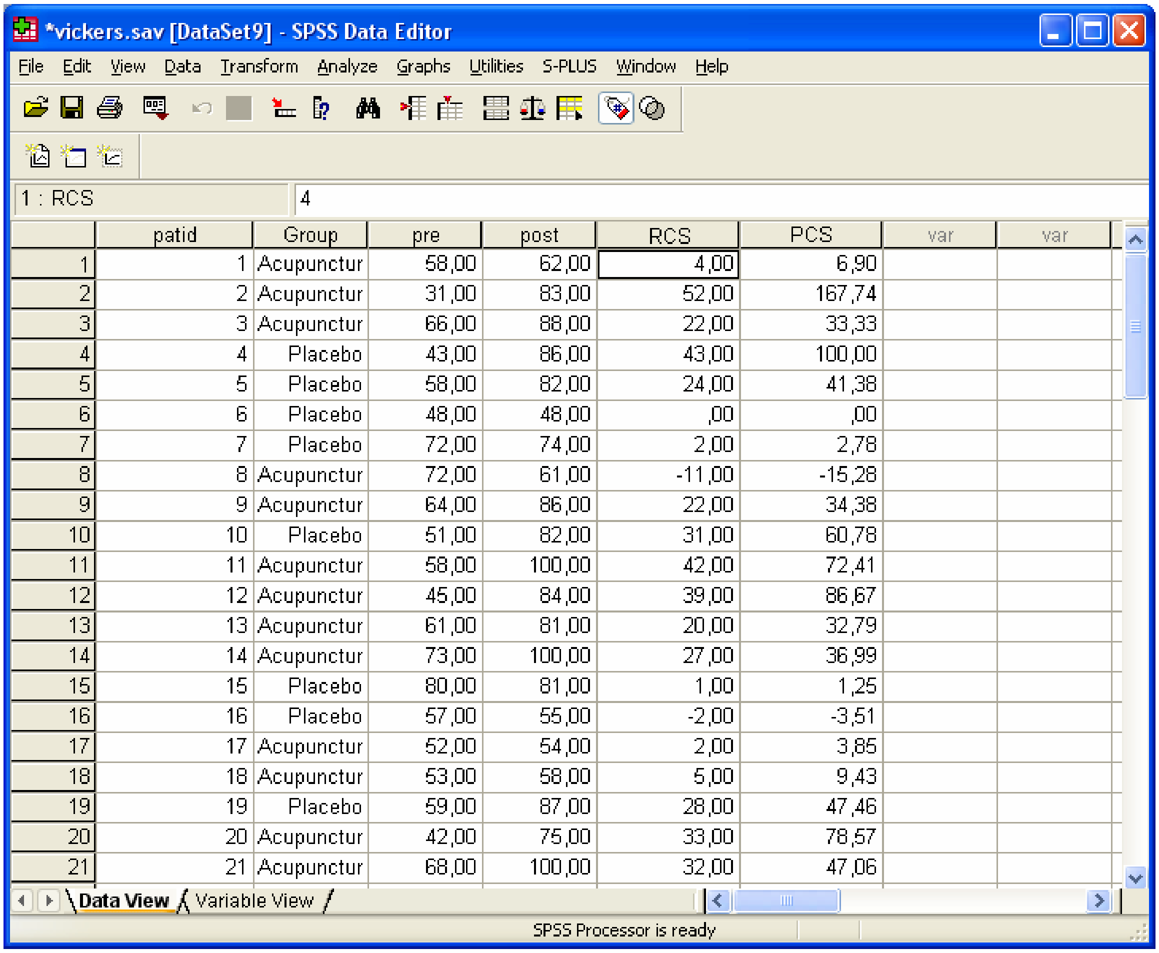
For percent change scores, compute:



Target Variable: PCS

Numeric Expression: (post - pre)/pre*100

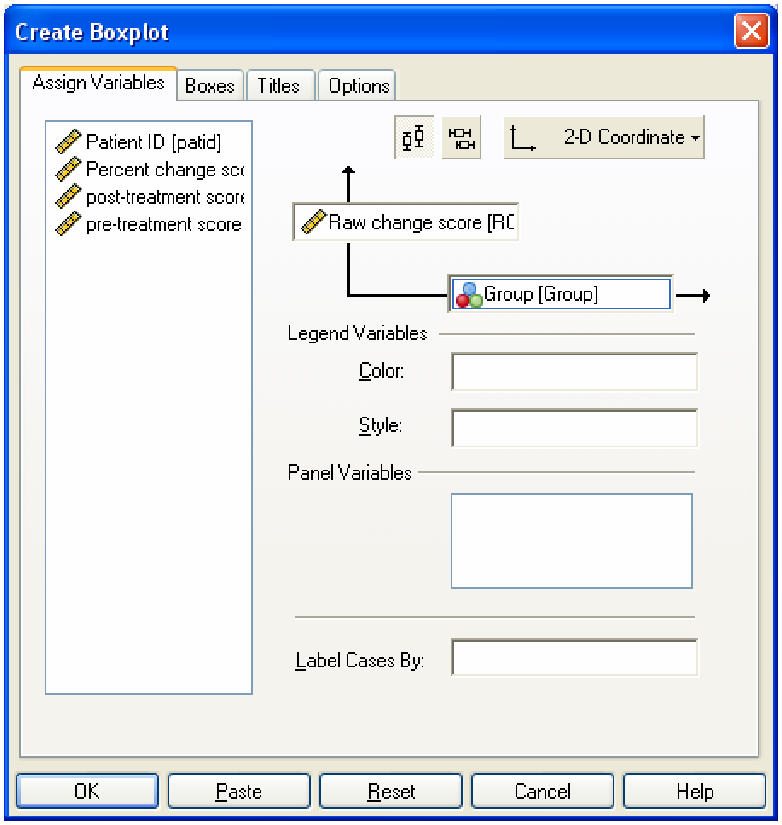
The variables are added to the data editor window:


*vickers.sav [DataSet9] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View
SPSS Processor is ready

1 : RCS 4

	patid	Group	pre	post	RCS	PCS	var	var
1	1	Acupunctur	58,00	62,00	4,00	6,90		
2	2	Acupunctur	31,00	83,00	52,00	167,74		
3	3	Acupunctur	66,00	88,00	22,00	33,33		
4	4	Placebo	43,00	86,00	43,00	100,00		
5	5	Placebo	58,00	82,00	24,00	41,38		
6	6	Placebo	48,00	48,00	,00	,00		
7	7	Placebo	72,00	74,00	2,00	2,78		
8	8	Acupunctur	72,00	61,00	-11,00	-15,28		
9	9	Acupunctur	64,00	86,00	22,00	34,38		
10	10	Placebo	51,00	82,00	31,00	60,78		
11	11	Acupunctur	58,00	100,00	42,00	72,41		
12	12	Acupunctur	45,00	84,00	39,00	86,67		
13	13	Acupunctur	61,00	81,00	20,00	32,79		
14	14	Acupunctur	73,00	100,00	27,00	36,99		
15	15	Placebo	80,00	81,00	1,00	1,25		
16	16	Placebo	57,00	55,00	-2,00	-3,51		
17	17	Acupunctur	52,00	54,00	2,00	3,85		
18	18	Acupunctur	53,00	58,00	5,00	9,43		
19	19	Placebo	59,00	87,00	28,00	47,46		
20	20	Acupunctur	42,00	75,00	33,00	78,57		
21	21	Acupunctur	68,00	100,00	32,00	47,06		

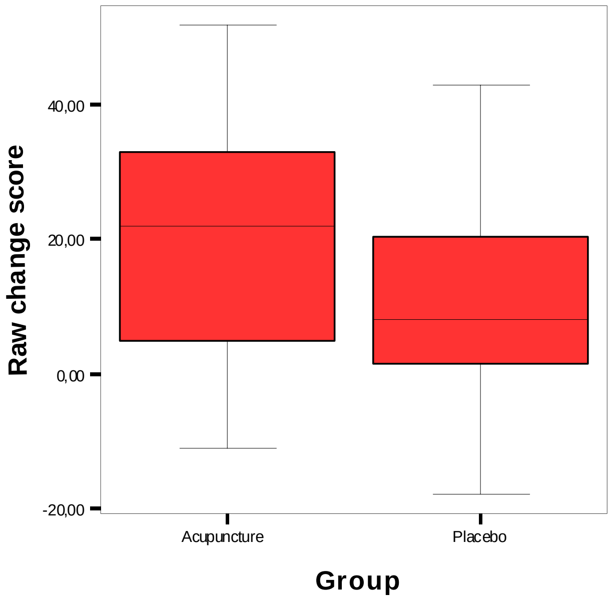
You may assign labels to raw and percent change scores in the variable view. Now groups can easily compared, calling Graphs-Interactive-Boxplot:


Create Boxplot

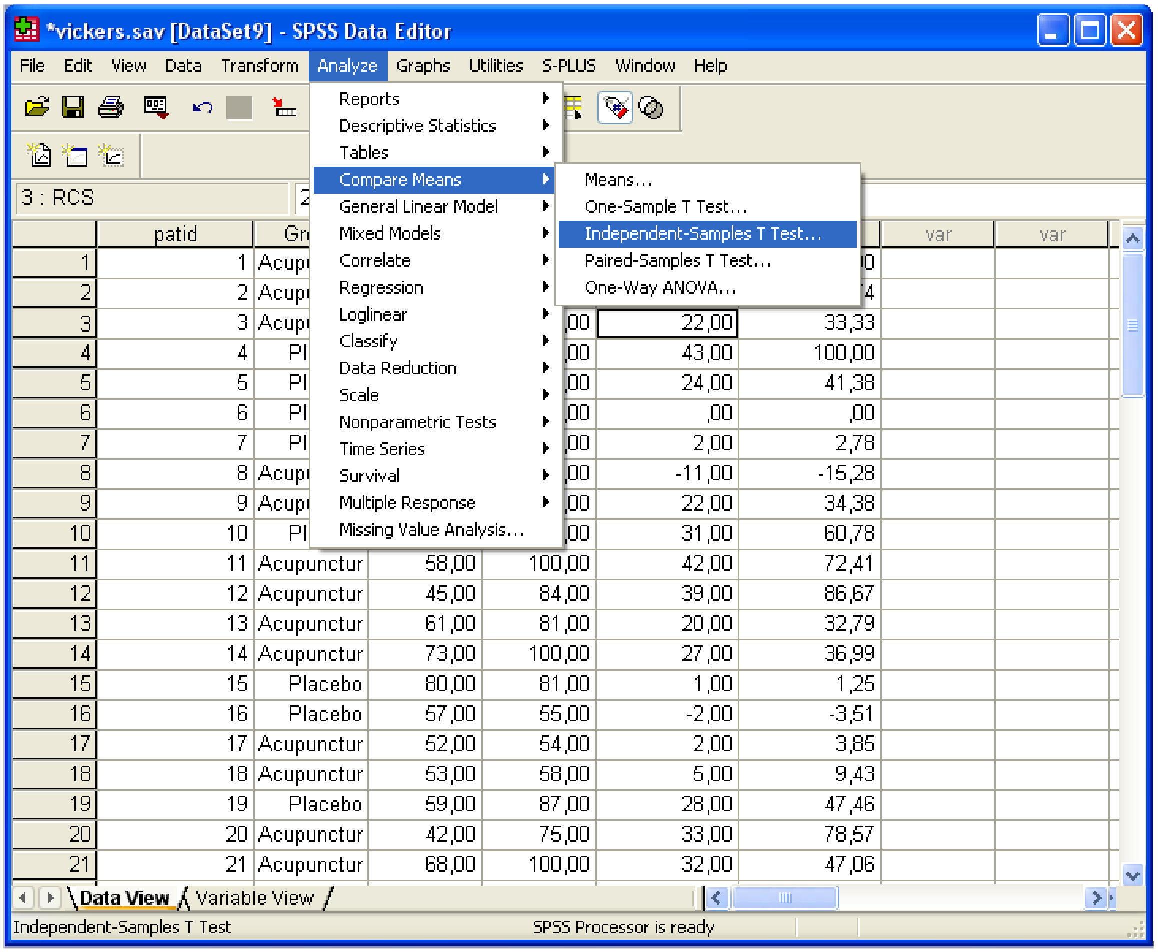
Patient ID [patid]
Percent change sc
post-treatment score
pre-treatment score

Raw change score [RC
Group [Group]

Legend Variables
Color:
Style:
Panel Variables
Label Cases By:



Raw change scores can be compared by means of a table of means and standard deviations and by a t-test for independent samples:

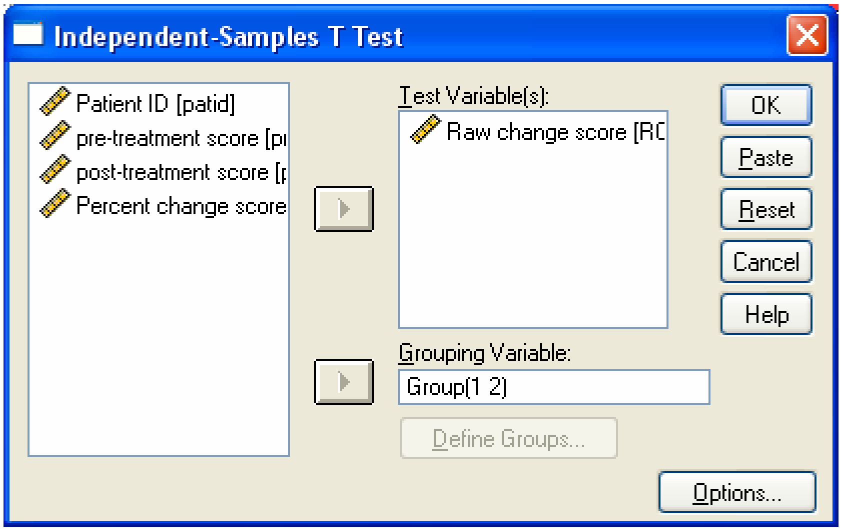


3 : RCS

	patid	Gr					var	var
1	1	Acup				0		
2	2	Acup				4		
3	3	Acup		,00	22,00	33,33		
4	4	Pl		,00	43,00	100,00		
5	5	Pl		,00	24,00	41,38		
6	6	Pl		,00	,00	,00		
7	7	Pl		,00	2,00	2,78		
8	8	Acup		,00	-11,00	-15,28		
9	9	Acup		,00	22,00	34,38		
10	10	Pl		,00	31,00	60,78		
11	11	Acupunctur	58,00	100,00	42,00	72,41		
12	12	Acupunctur	45,00	84,00	39,00	86,67		
13	13	Acupunctur	61,00	81,00	20,00	32,79		
14	14	Acupunctur	73,00	100,00	27,00	36,99		
15	15	Placebo	80,00	81,00	1,00	1,25		
16	16	Placebo	57,00	55,00	-2,00	-3,51		
17	17	Acupunctur	52,00	54,00	2,00	3,85		
18	18	Acupunctur	53,00	58,00	5,00	9,43		
19	19	Placebo	59,00	87,00	28,00	47,46		
20	20	Acupunctur	42,00	75,00	33,00	78,57		
21	21	Acupunctur	68,00	100,00	32,00	47,06		

Independent-Samples T Test   SPSS Processor is ready

Remember to define the group codes. In our data set, the groups are coded as 1 and 2.



Test Variable(s): Raw change score [RC

Grouping Variable: Group(1 2)

Group Statistics

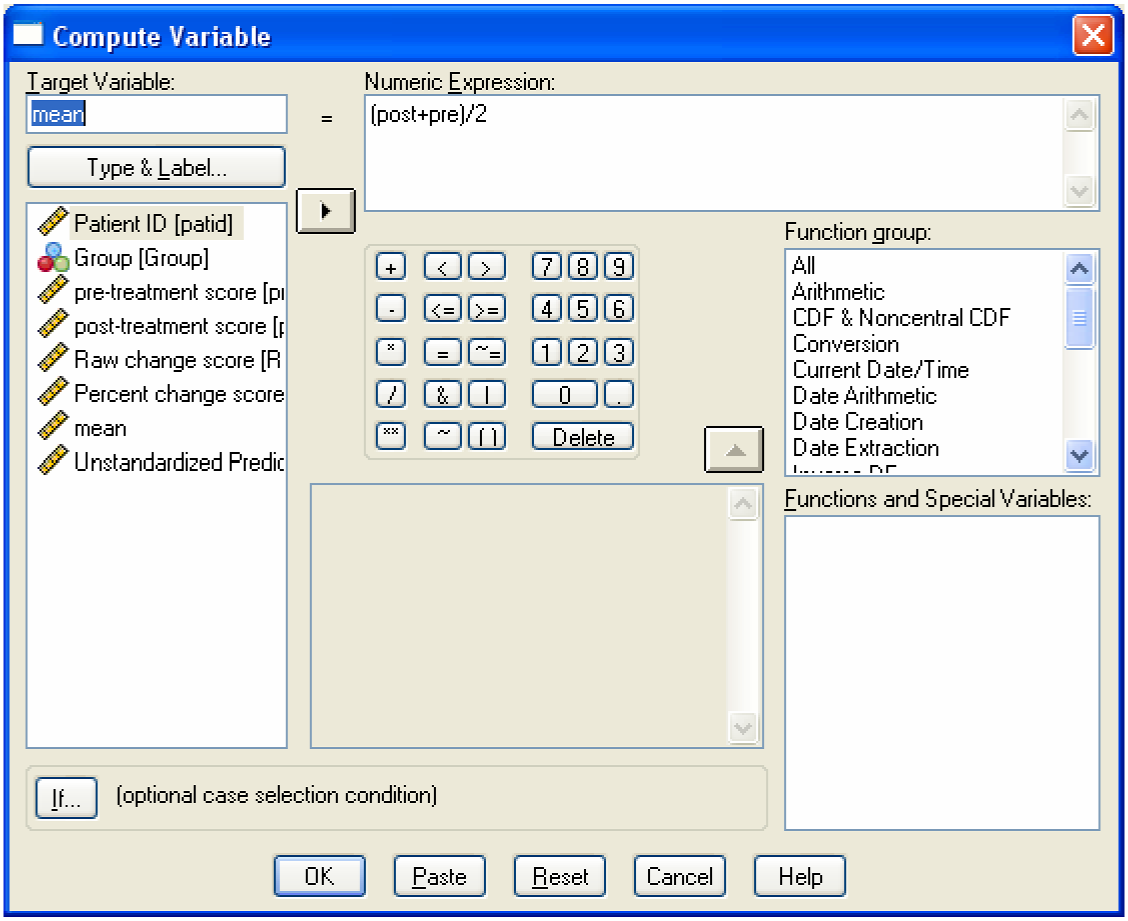
	Group	N	Mean	Std. Deviation	Std. Error Mean
Raw change score	Acupuncture	25	20,8000	15,82982	3,16596
	Placebo	23	11,1304	15,61847	3,25668

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Raw change score	Equal variances assumed	,095	,759	2,128	46	,039	9,66957	4,54455	,52187	18,81726
	Equal variances not assumed			2,129	45,764	,039	9,66957	4,54195	,52583	18,81330

Correlation of change score with baseline measurement

To compute the regression-to-the-mean corrected correlation of change score with baseline measurement, we have to compute the mean of pretest and posttest scores first (Transform-Compute):

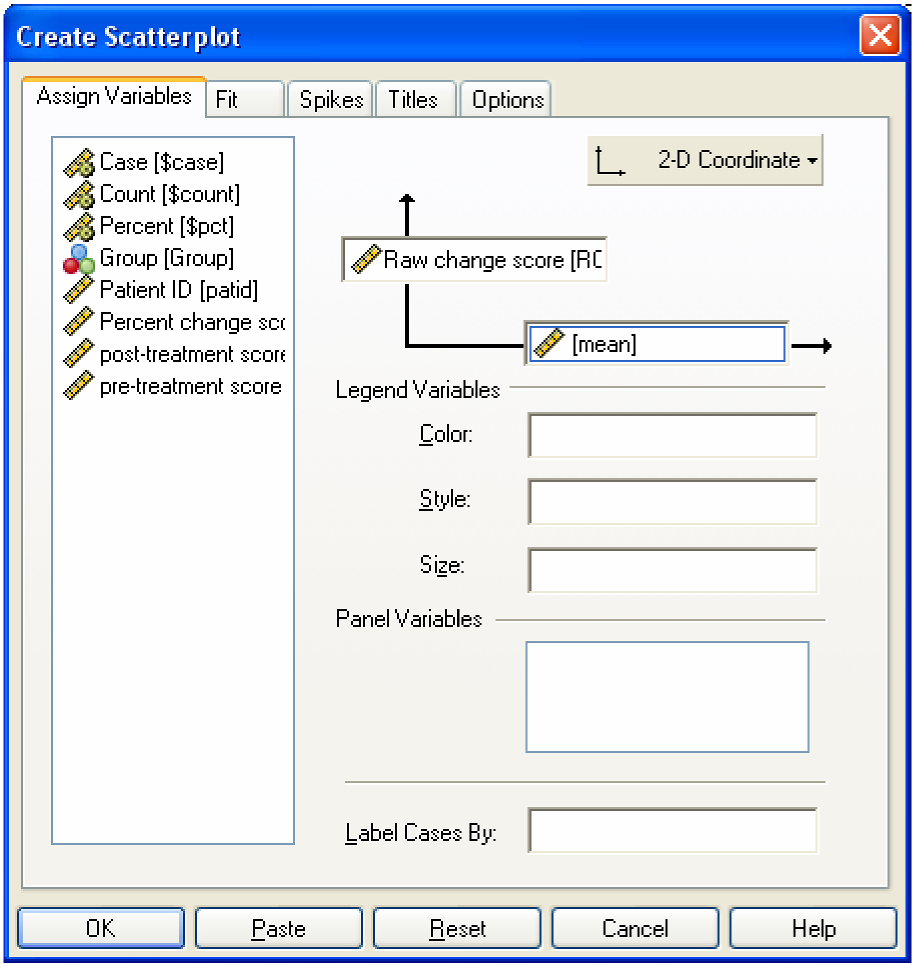


Target Variable: mean

=

Numeric Expression: (post+pre)/2

Then we can produce a scatterplot of raw change scores vs. mean of pre- and post-treatment score (Graphs-Interactive-Scatterplot):

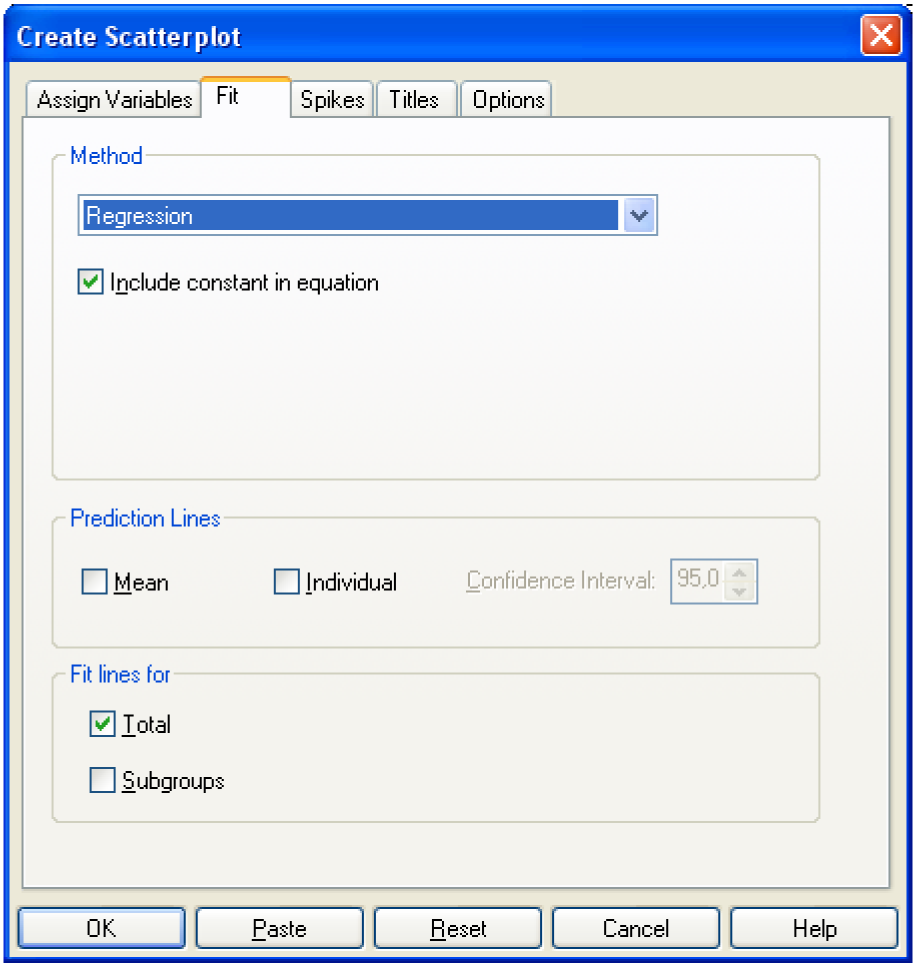


2-D Coordinate

Raw change score [RC

[mean]

At the “Fit” view, we may request a regression line:



Method

Regression

☑ Include constant in equation

Prediction Lines

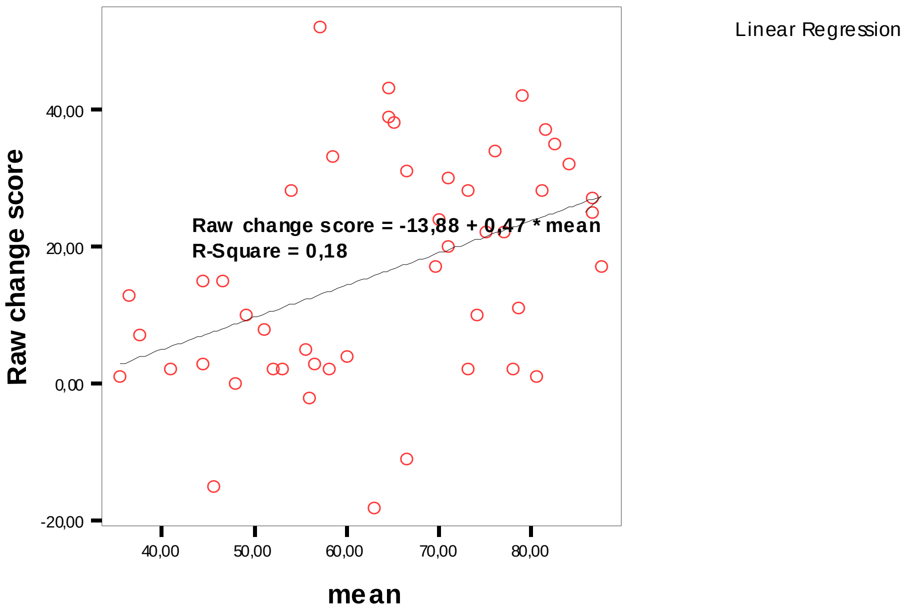
□ Mean □ Individual Confidence Interval: 95,0

Fit lines for

☑ Total

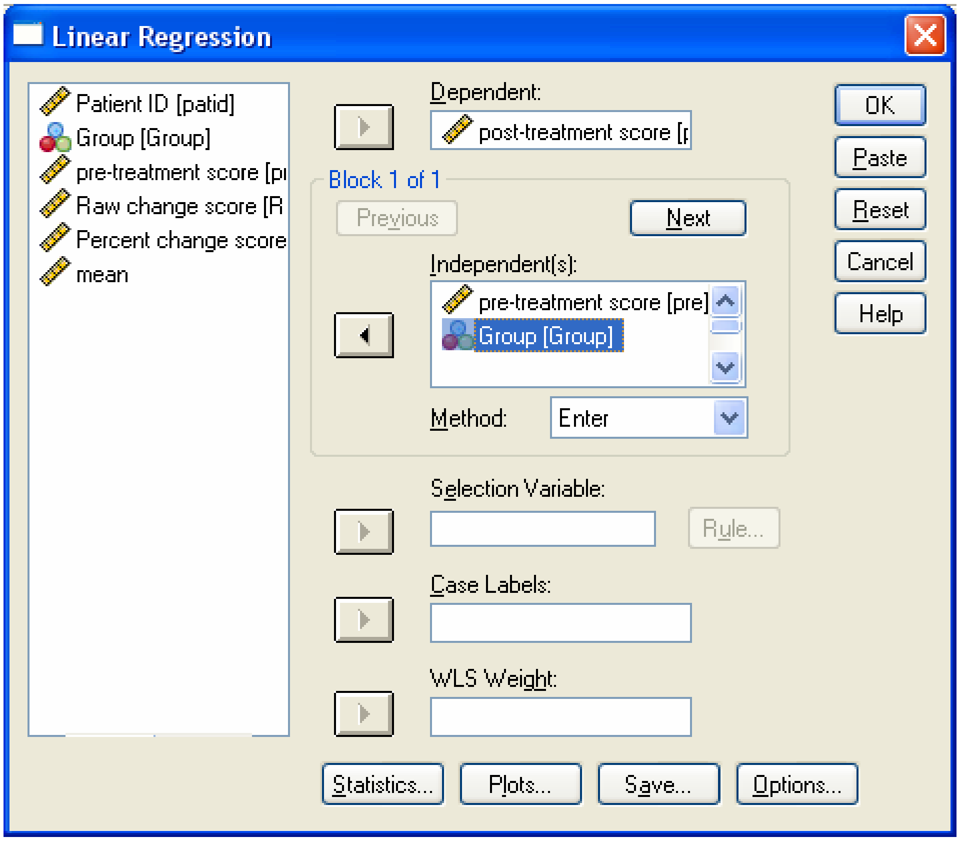
□ Subgroups

The program automatically outputs the regression equation, but one may hide this additional description (by double-clicking the graph, selecting the description, and choosing ‘Hide label’ from the context menu).



Analysis of covariance

The baseline-adjusted comparison of posttest scores between groups (ANCOVA) can be obtained by selecting Analyze-Regression-Linear from the menu:



Dependent: post-treatment score [ṛ

Independent(s): pre-treatment score [pre], Group [Group]

Method: Enter

The program outputs a table of regression coefficients:

Coefficients(a)

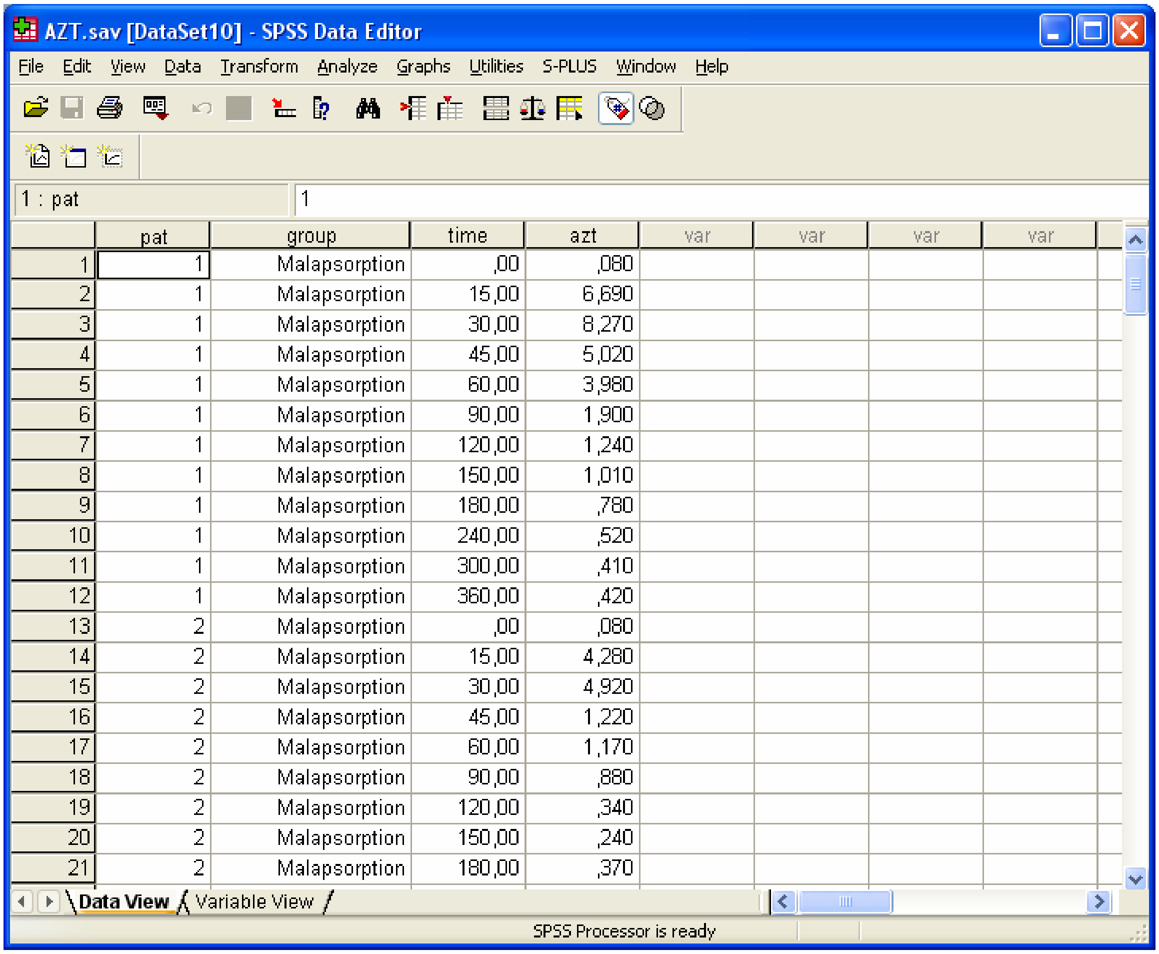
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	50,693	13,642		3,716	,001
	pre-treatment score	,700	,174	,480	4,014	,000
	Group	-12,019	4,655	-,309	-2,582	,013

a Dependent Variable: post-treatment score

The coefficient B for Group is -12.019. Technically, a change in Group by 1 unit means a reduction of post-treatment score by 12.019. Practically, group 2 is placebo and group 1 is acupuncture. Thus, the placebo group has a baseline-adjusted average post-treatment score which is by 12.019 units lower than that of the acupuncture group.

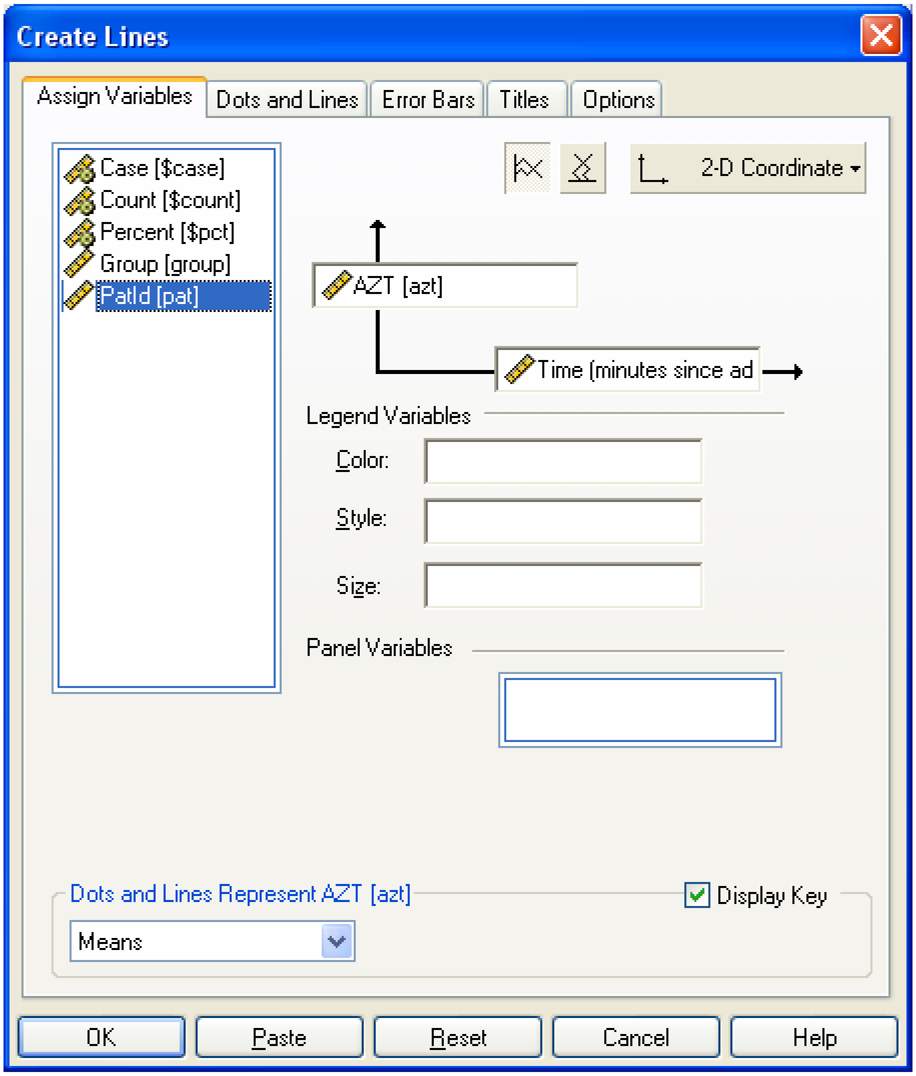
Individual curves

Open the data set AZT.sav:

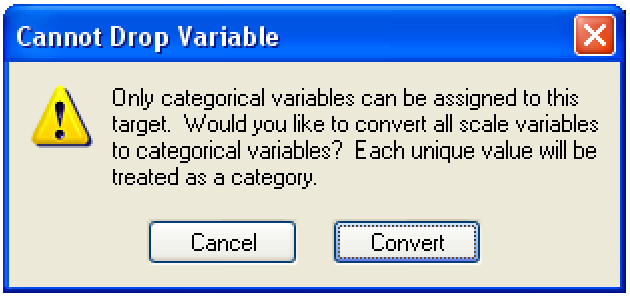

AZT.sav [DataSet10] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 : pat
1
Data View
Variable View
SPSS Processor is ready

	pat	group	time	azt	var	var	var	var
1	1	Malapsorption	,00	,080				
2	1	Malapsorption	15,00	6,690				
3	1	Malapsorption	30,00	8,270				
4	1	Malapsorption	45,00	5,020				
5	1	Malapsorption	60,00	3,980				
6	1	Malapsorption	90,00	1,900				
7	1	Malapsorption	120,00	1,240				
8	1	Malapsorption	150,00	1,010				
9	1	Malapsorption	180,00	,780				
10	1	Malapsorption	240,00	,520				
11	1	Malapsorption	300,00	,410				
12	1	Malapsorption	360,00	,420				
13	2	Malapsorption	,00	,080				
14	2	Malapsorption	15,00	4,280				
15	2	Malapsorption	30,00	4,920				
16	2	Malapsorption	45,00	1,220				
17	2	Malapsorption	60,00	1,170				
18	2	Malapsorption	90,00	,880				
19	2	Malapsorption	120,00	,340				
20	2	Malapsorption	150,00	,240				
21	2	Malapsorption	180,00	,370				

Individual curves, plotted into several panels, can be obtained by choosing Graphs-Interactive-Line:

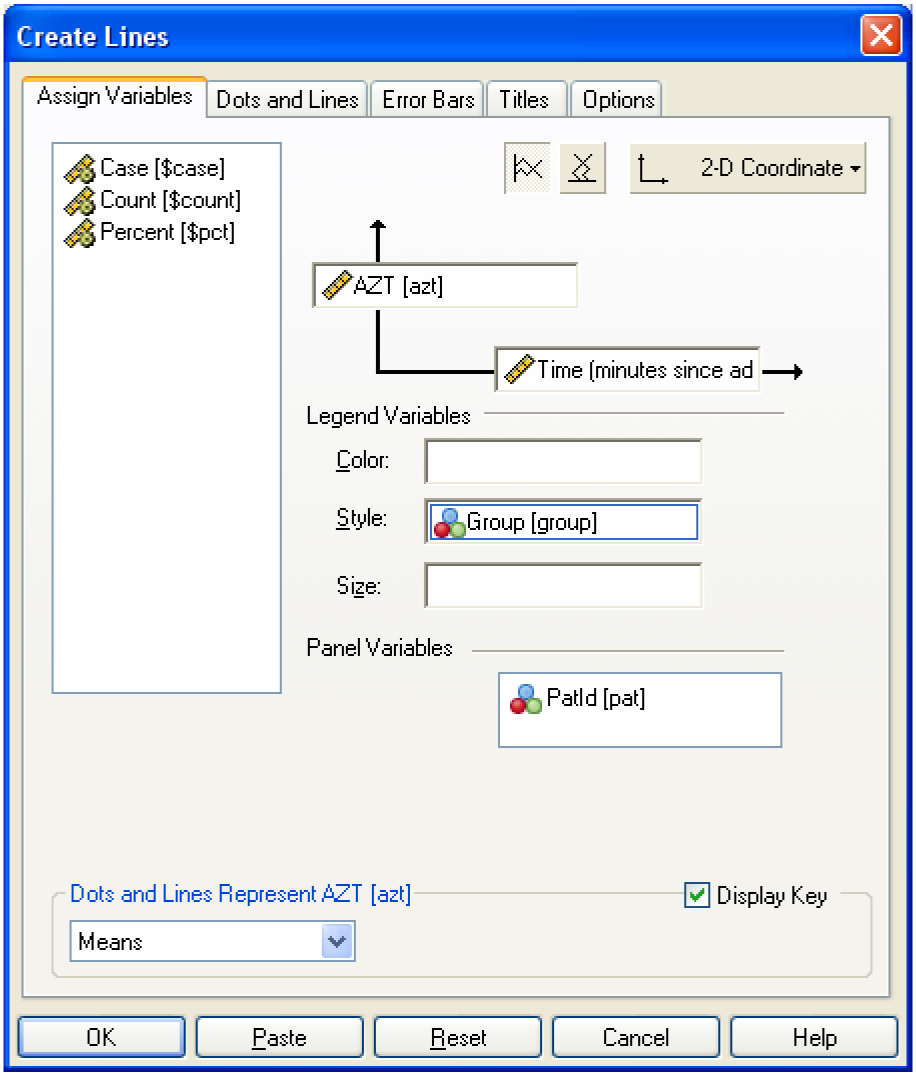


Put AZT and Time into the fields corresponding to the y- and x-axis, respectively. The move PatID into the field 'Panel Variables'. A window pops up:



Only categorical variables can be assigned to this target. Would you like to convert all scale variables to categorical variables? Each unique value will be treated as a category.

This message reminds us that the PatID variable should be of nominal type but it is (erranously) of scale type. We can immediately change the type of PatID by selecting 'Convert'. The same warning pops up if we move Group into the field 'style' (to have different line styles for either group):


Create Lines

Assign Variables | Dots and Lines | Error Bars | Titles | Options

Case [$case]
Count [$count]
Percent [$pct]

2-D Coordinate

AZT [azt]

Time (minutes since ad

Legend Variables

Color:

Style: Group [group]

Size:

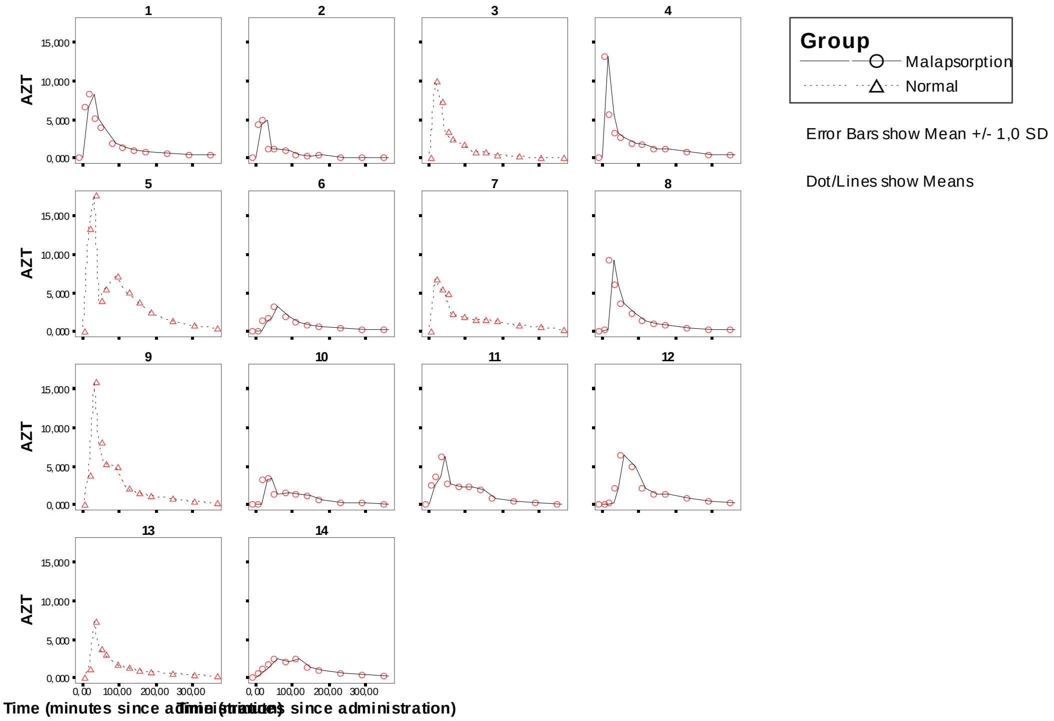
Panel Variables

PatId [pat]

Dots and Lines Represent AZT [azt]

☑ Display Key

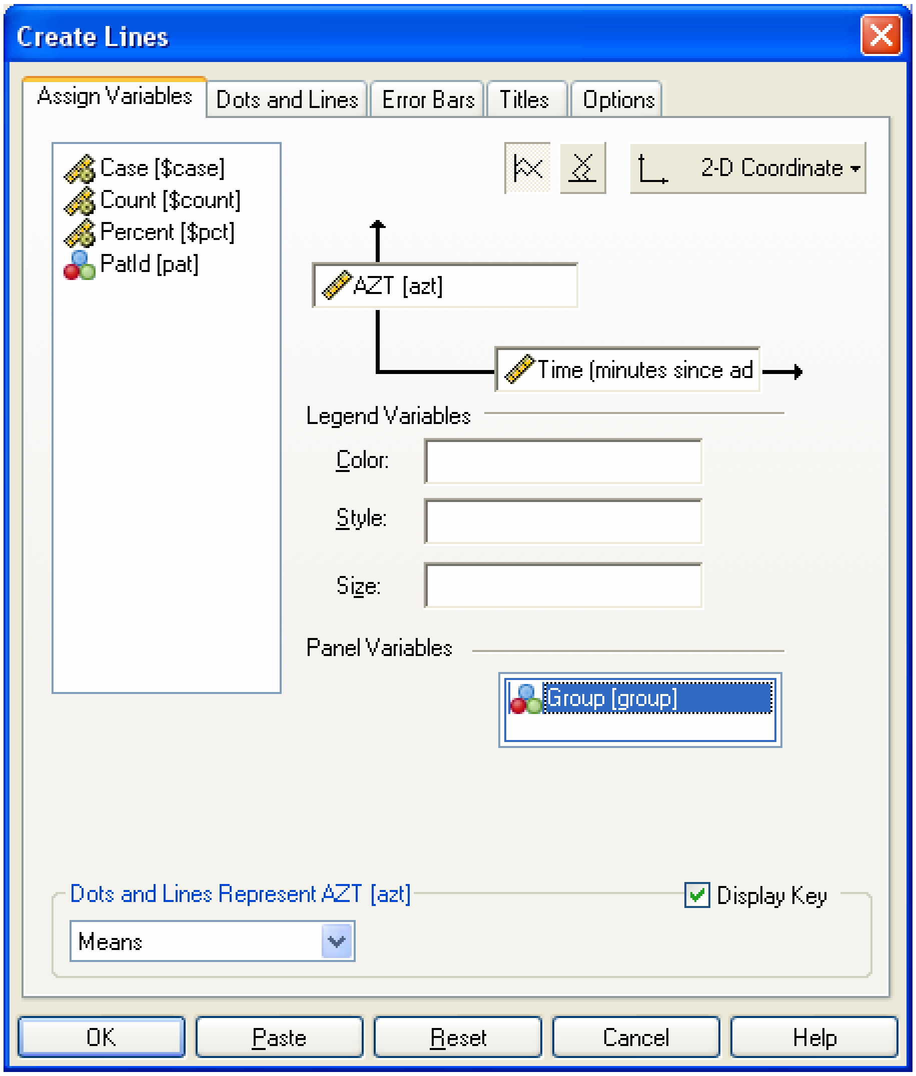
Means



Actually, SPSS computes mean curves, but the means are computed per patient and time point, such that the plot actually depicts the raw observations. The legends on the right hand side can be selected (after double-clicking the graph) and deleted.

Grouped curves

A plot of mean curves with standard deviations represented as error bars can be obtained in a similar way, just omitting to specify PatID to define panels on the ‘Assign Variables’ view, and moving Group into the ‘Panel variables’ field:



AZT [azt]

Time (minutes since ad

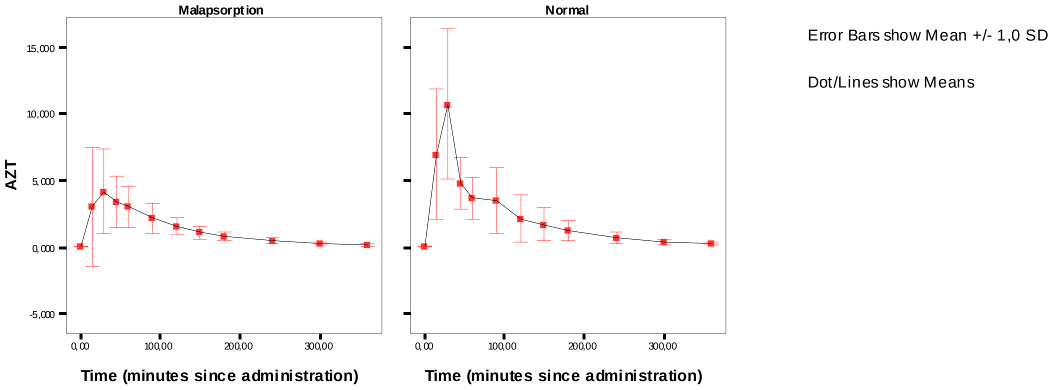
Group [group]

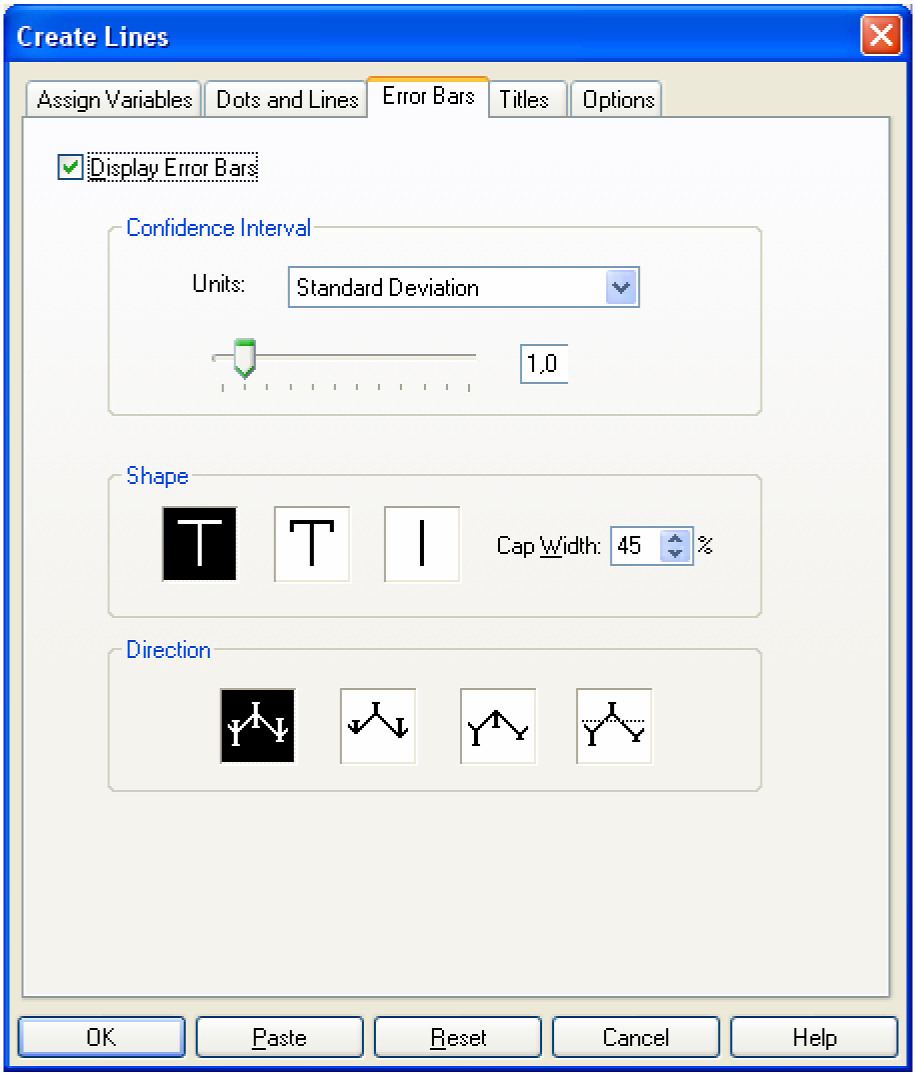
Dots and Lines Represent AZT [azt]

Means

☑ Display Key

On the ‘Error Bars’ view, select Units: Standard deviation and move the slider to 1,0:





☑ Display Error Bars

Units: Standard Deviation

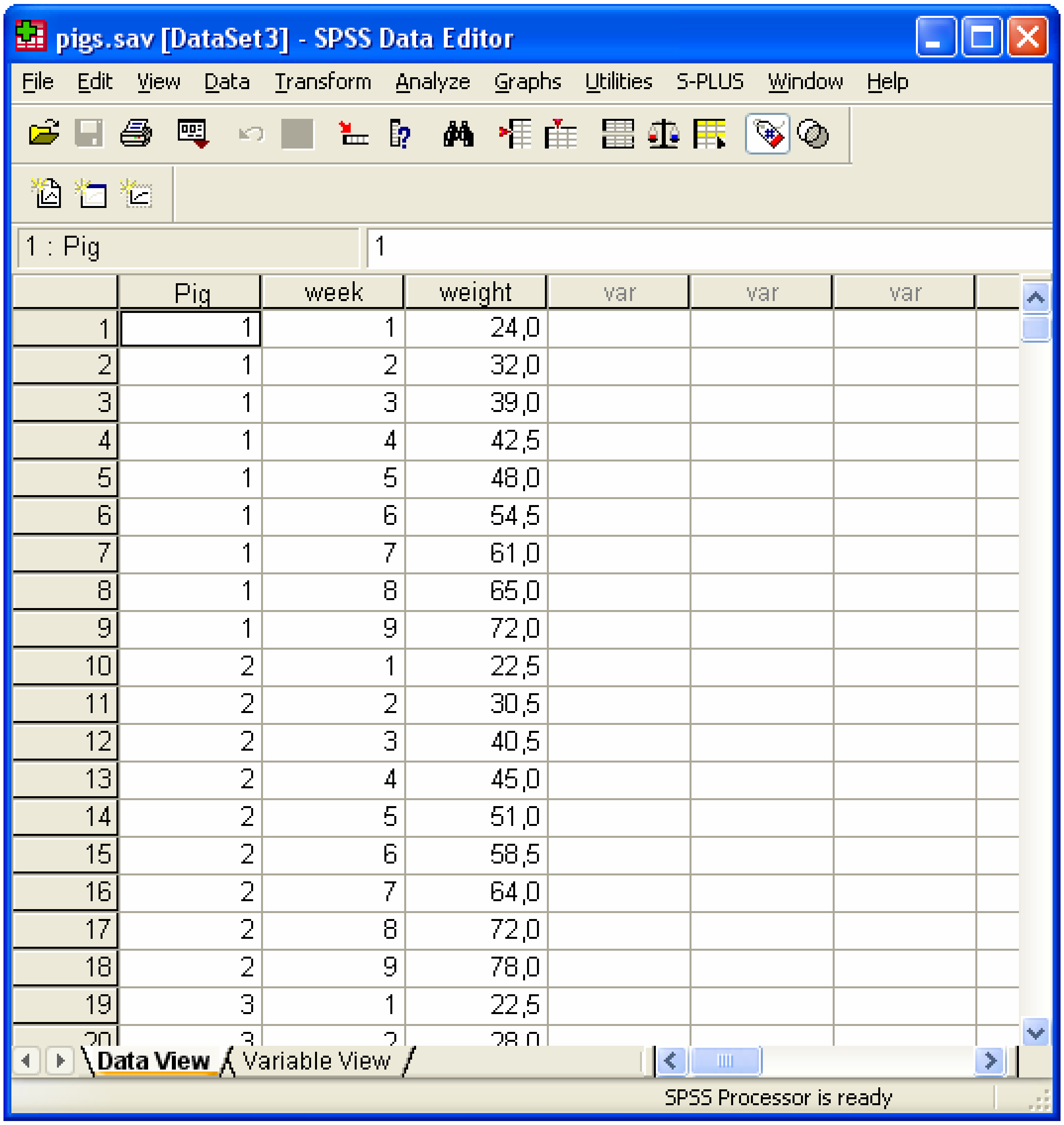
1,0

Cap Width: 45 %

Similarly, one can obtain a plot of median and p25/p75 curves. We start by requesting the median curves, and then we add lines representing p25 and p75.

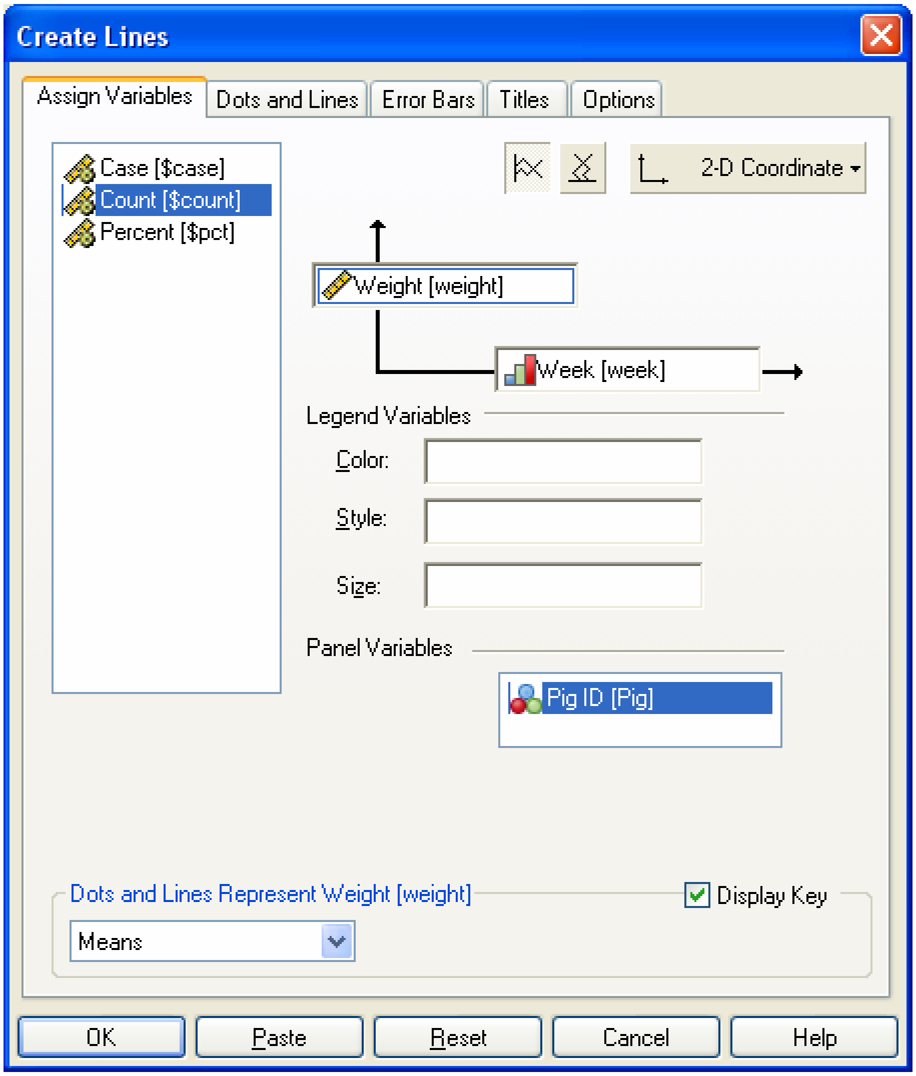
Computing summary measures

Summary measures suitable for growth curves are exemplified on the data set pigs.sav which contains serial weight measurements on 48 pigs. Clearly, the data is in long format:


pigs.sav [DataSet3] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 : Pig
1
Data View
Variable View
SPSS Processor is ready

	Pig	week	weight	var	var	var
1	1	1	24,0			
2	1	2	32,0			
3	1	3	39,0			
4	1	4	42,5			
5	1	5	48,0			
6	1	6	54,5			
7	1	7	61,0			
8	1	8	65,0			
9	1	9	72,0			
10	2	1	22,5			
11	2	2	30,5			
12	2	3	40,5			
13	2	4	45,0			
14	2	5	51,0			
15	2	6	58,5			
16	2	7	64,0			
17	2	8	72,0			
18	2	9	78,0			
19	3	1	22,5			
20	3	2	28,0			

First, we draw individual curves to check the data values (choose graphs-interactive-Line):



Create Lines

Case [$case]
Count [$count]
Percent [$pct]

Weight [weight]

Week [week]

Legend Variables

Color:

Style:

Size:

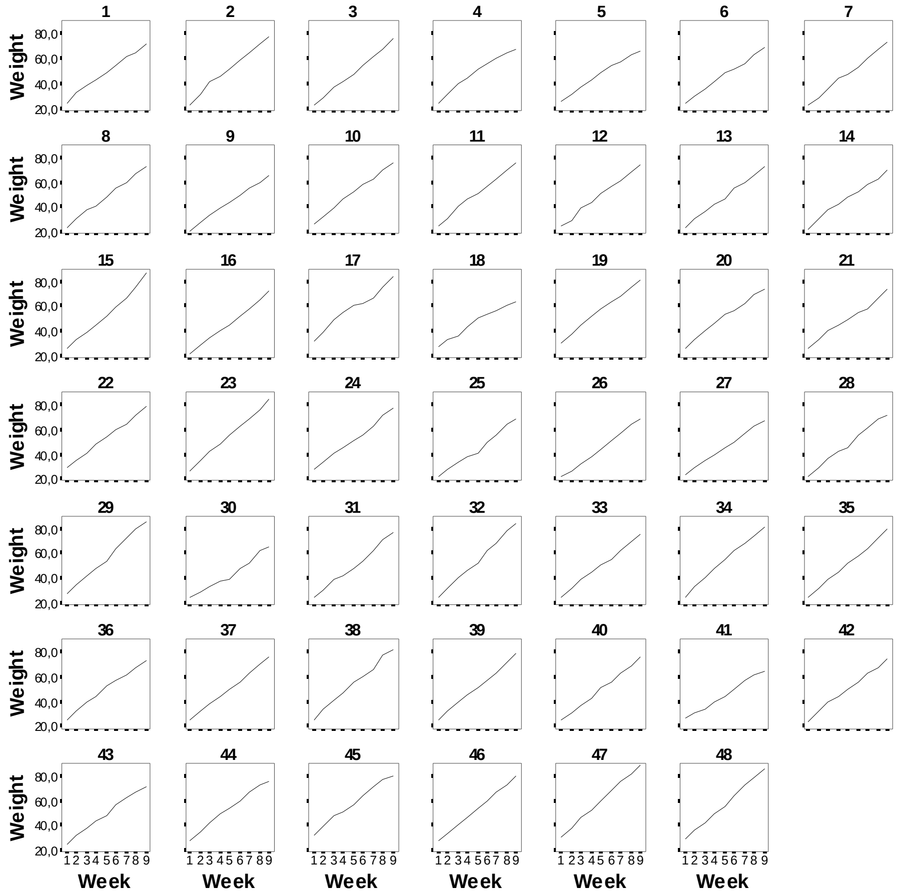
Panel Variables

Pig ID [Pig]

Dots and Lines Represent Weight [weight]

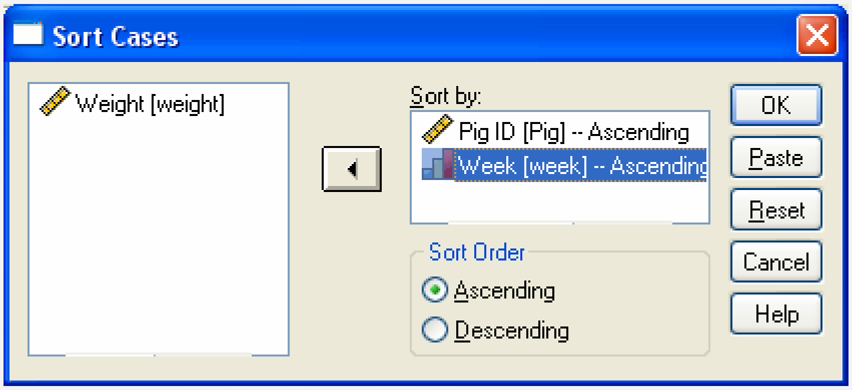
Means

☑ Display Key



Extracting raw first and last values of each subject

To compute the weight change during 9 weeks of diet, we have to extract the first and last weight measurements and compute their difference. First, sort the data set by pig ID and week (Data-Sort Cases):



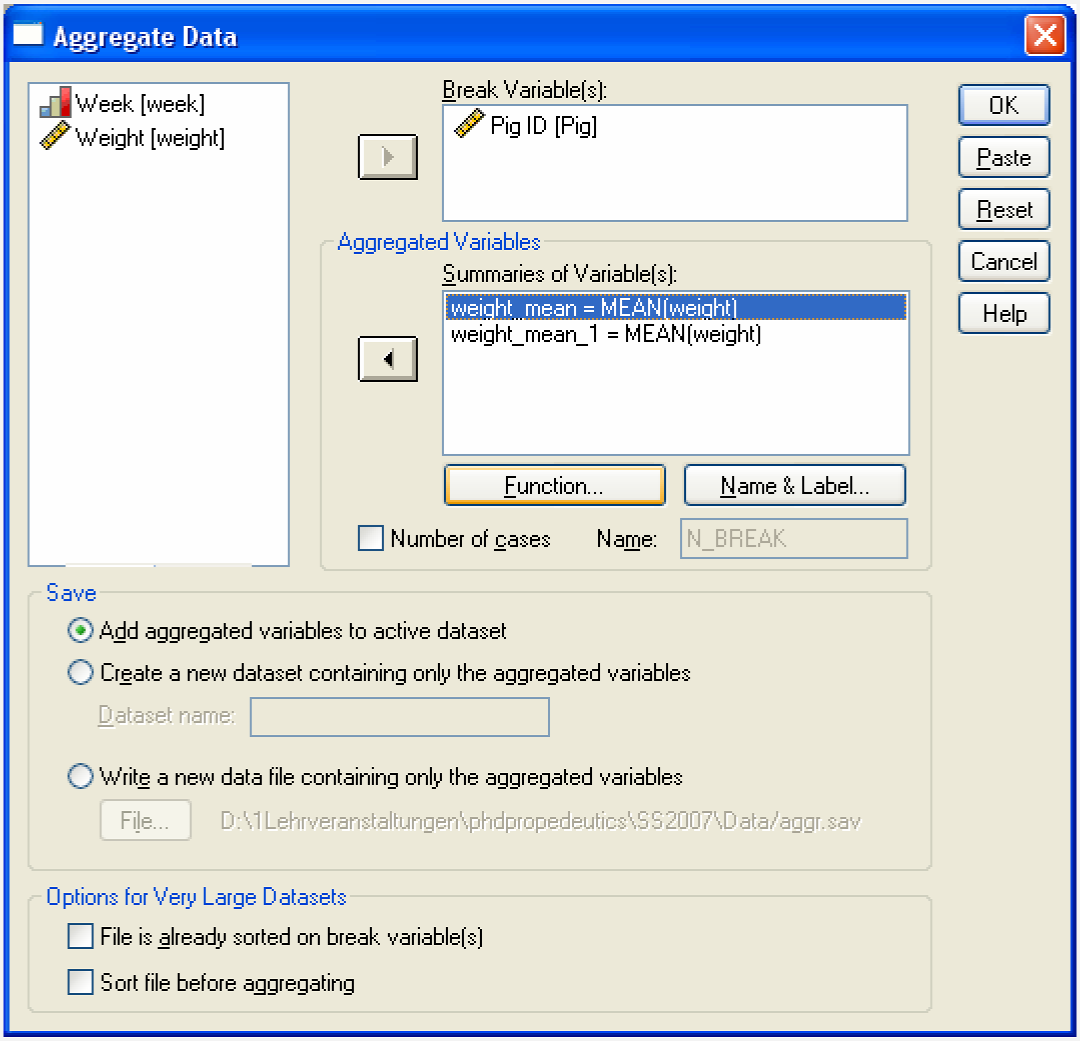
Sort Cases

Weight [weight]

Sort by:
Pig ID [Pig] -- Ascending
Week [week] -- Ascending

Sort Order
● Ascending
○ Descending

Next, choose Data-Aggregate. Move Pig ID to the field ‘Break Variable(s)’ and move ‘weight’ *twice* into the field ‘Summaries of Variable(s)’. Please note that if in other data sets there any other baseline characteristics variables, you must also move them into the field ‘Break Variable(s)’.



Week [week]
Weight [weight]

Break Variable(s):
Pig ID [Pig]

Aggregated Variables

Summaries of Variable(s):
weight_mean = MEAN(weight)
weight_mean_1 = MEAN(weight)

□ Number of cases Name: N_BREAK

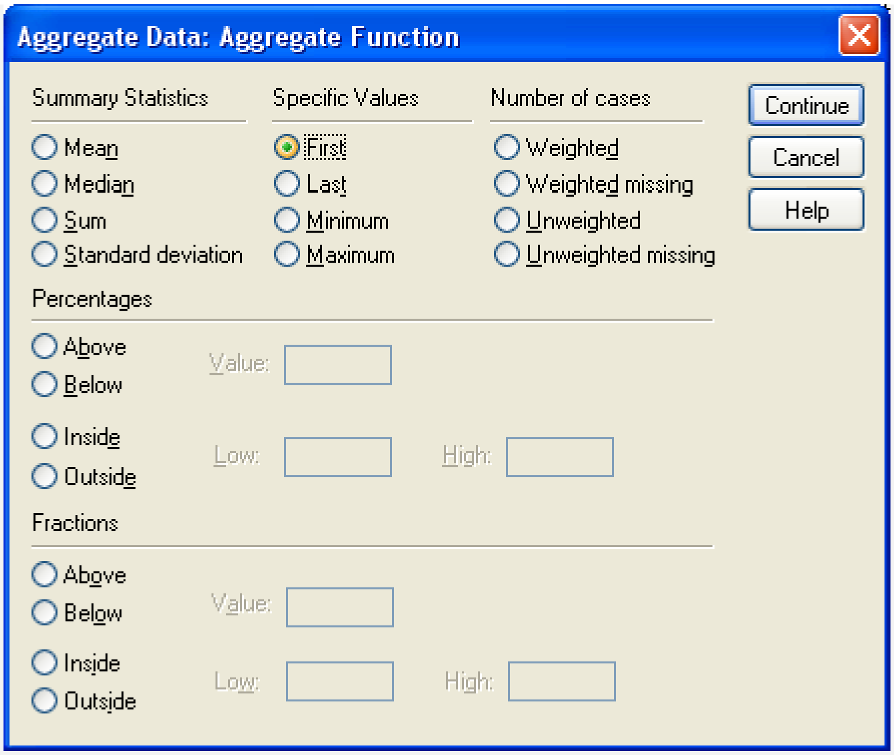
Save

● Add aggregated variables to active dataset
○ Create a new dataset containing only the aggregated variables
Dataset name:
○ Write a new data file containing only the aggregated variables
D:\1Lehrveranstaltungen\phdpropedeutics\SS2007\Data/aggr.sav

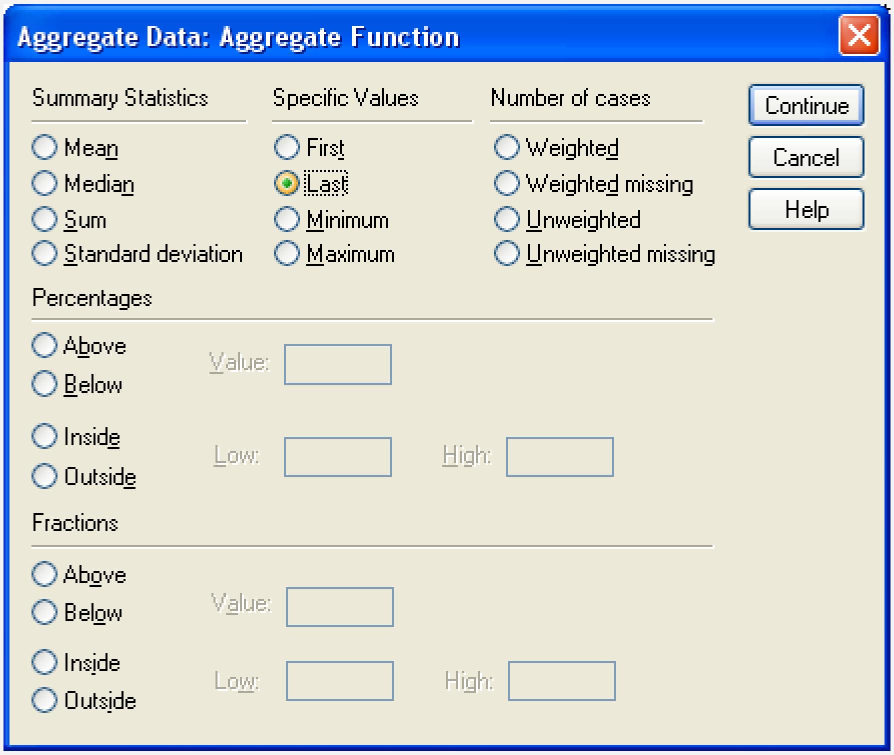
Options for Very Large Datasets

□ File is already sorted on break variable(s)
□ Sort file before aggregating

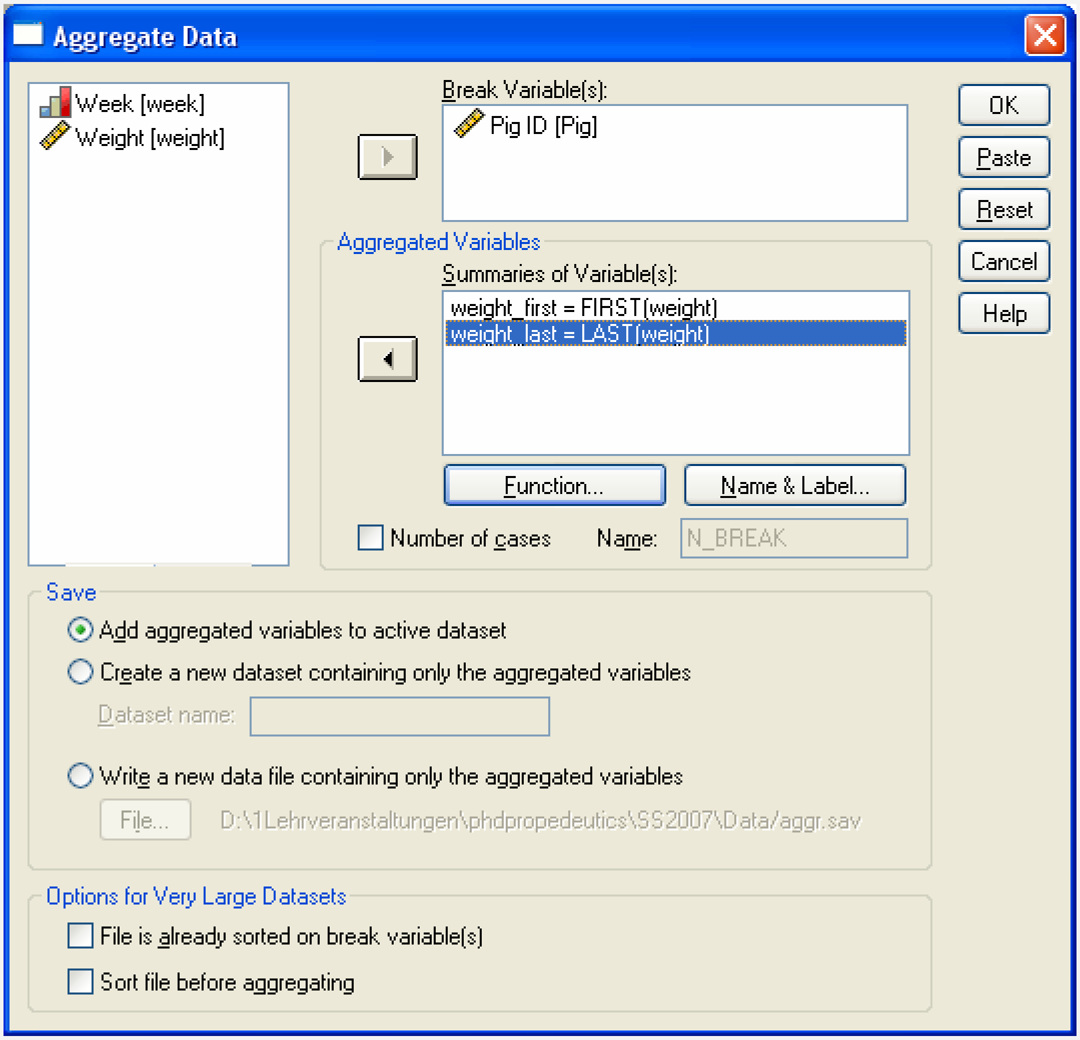
This would compute a new data set with one line per pig, containing two variables corresponding to the mean weight. We select the first of the two variable summaries and click on ‘Function…’:



Here, we select ‘First’, which can be found below ‘Specific Values’. Confirm by pressing ‘Continue’. Next, select the second summary variable and again click on ‘Function…’. Now we select ‘Last’:



The main dialogue should read as:



Week [week]
Weight [weight]

Break Variable(s):
Pig ID [Pig]

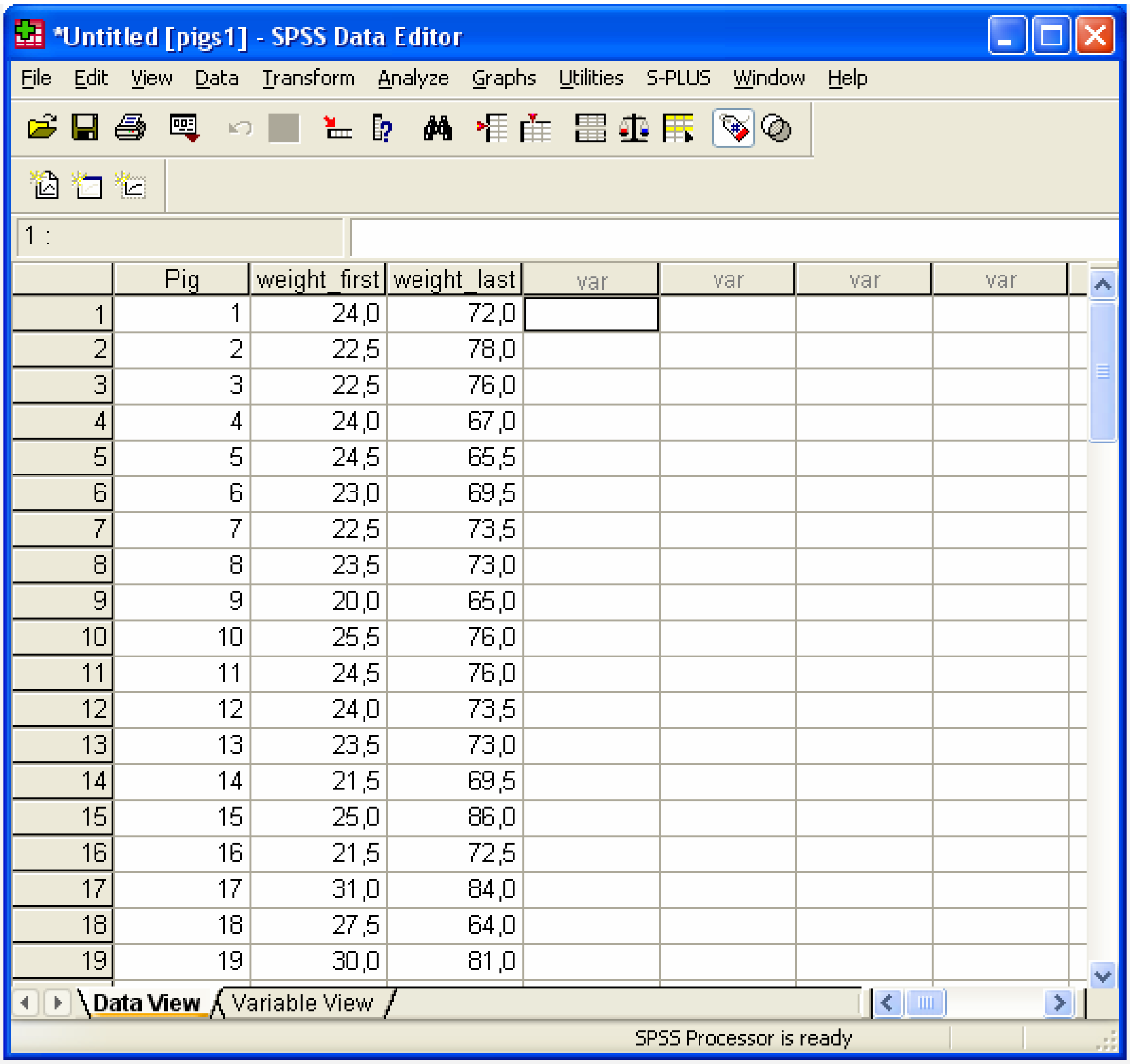
Summaries of Variable(s):
weight_first = FIRST(weight)
weight_last = LAST(weight)

□ Number of cases Name: N_BREAK

● Add aggregated variables to active dataset
○ Create a new dataset containing only the aggregated variables
Dataset name:
○ Write a new data file containing only the aggregated variables
File... D:\1Lehrveranstaltungen\phdpropedeutics\SS2007\Data/aggr.sav

□ File is already sorted on break variable(s)
□ Sort file before aggregating

In the ‘Save’ field, select ‘Create a new dataset containing only the aggregated variables’, and name the new data set ‘pigs1’. Now, you may press OK. The new, aggregated, data set is displayed in a new SPSS data editor window:

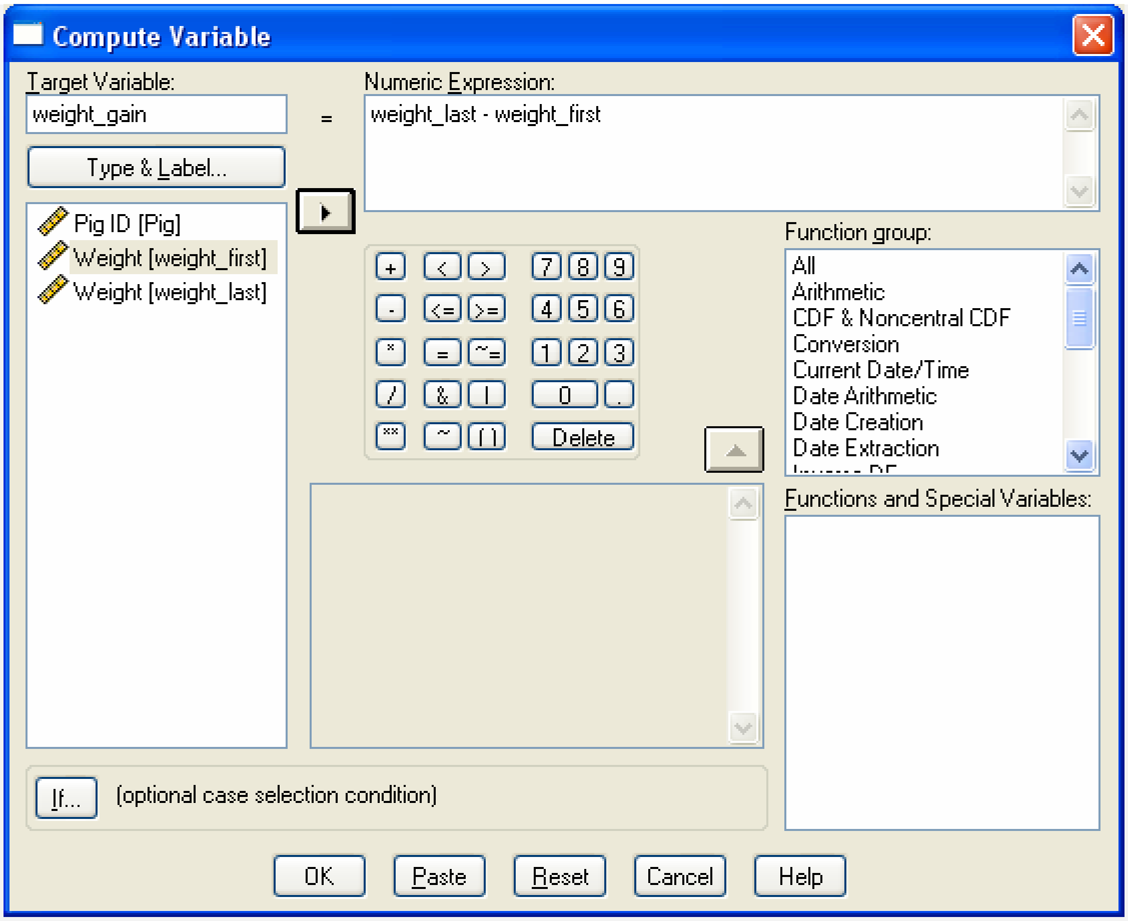

*Untitled [pigs1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 :
Data View
Variable View

	Pig	weight_first	weight_last	var	var	var	var
1	1	24,0	72,0				
2	2	22,5	78,0				
3	3	22,5	76,0				
4	4	24,0	67,0				
5	5	24,5	65,5				
6	6	23,0	69,5				
7	7	22,5	73,5				
8	8	23,5	73,0				
9	9	20,0	65,0				
10	10	25,5	76,0				
11	11	24,5	76,0				
12	12	24,0	73,5				
13	13	23,5	73,0				
14	14	21,5	69,5				
15	15	25,0	86,0				
16	16	21,5	72,5				
17	17	31,0	84,0				
18	18	27,5	64,0				
19	19	30,0	81,0				

SPSS Processor is ready

Now, choose ‘Transform-Compute’ to compute the difference:

A new variable is created containing the weight gain for each of the pigs as a summary measure in the new data set. The distribution of thesese values can be depicted using a histogram (Graphs-Interactive-Histogram).



Compute Variable

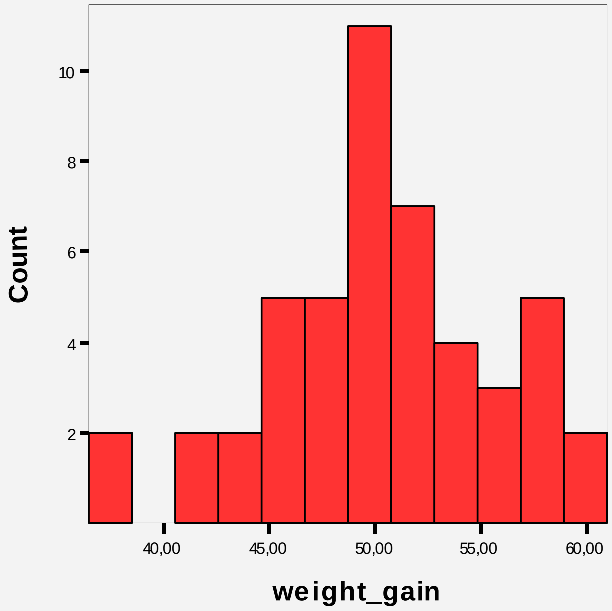
Target Variable: weight_gain = Numeric Expression: weight_last - weight_first

Pig ID [Pig]
Weight [weight_first]
Weight [weight_last]

Function group:
All
Arithmetic
CDF & Noncentral CDF
Conversion
Current Date/Time
Date Arithmetic
Date Creation
Date Extraction

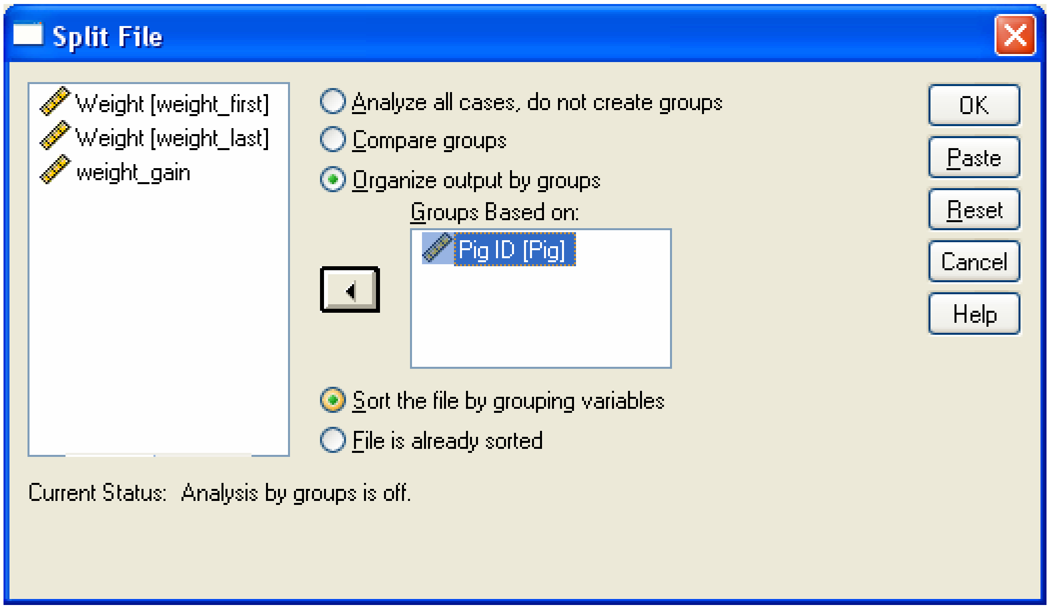
Functions and Special Variables:

If... (optional case selection condition)



Computing regression-stabilized first and last values of each subject

Regression-stabilized differences can be obtained by performing linear regression on the weight measurements of each pig. First, revert to the data set pigs.sav. Then, split the file by pig ID (Data-Split file):



Split File

Weight [weight_first]
Weight [weight_last]
weight_gain

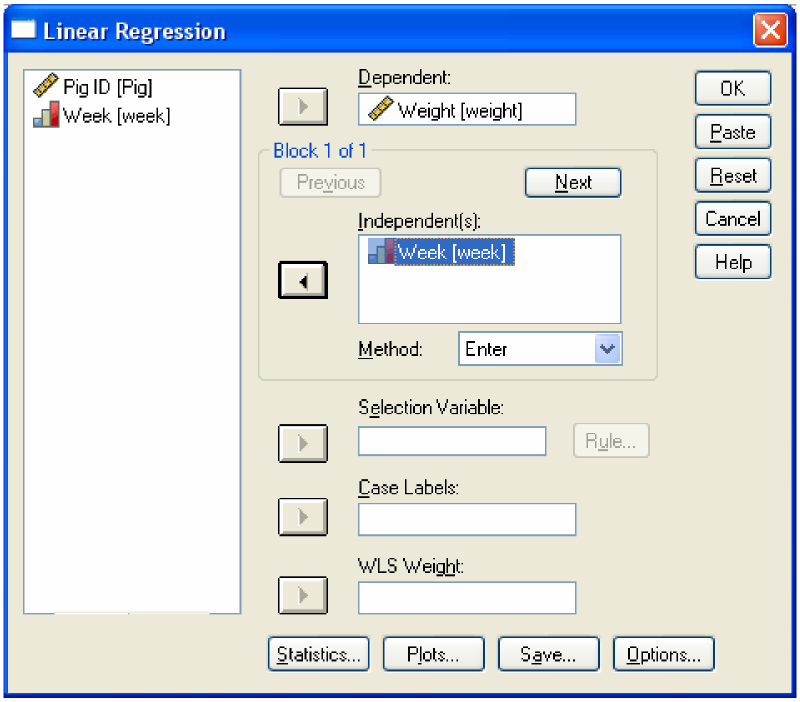
○ Analyze all cases, do not create groups
○ Compare groups
● Organize output by groups

Groups Based on: Pig ID [Pig]

● Sort the file by grouping variables
○ File is already sorted

Current Status: Analysis by groups is off.

Next, call linear regression (Analyze-Regression-linear). Choose ‘Weight’ as the dependent and ‘Week’ as the independent variables:



Linear Regression

Pig ID [Pig]
Week [week]

Dependent: Weight [weight]

Block 1 of 1

Independent(s): Week [week]

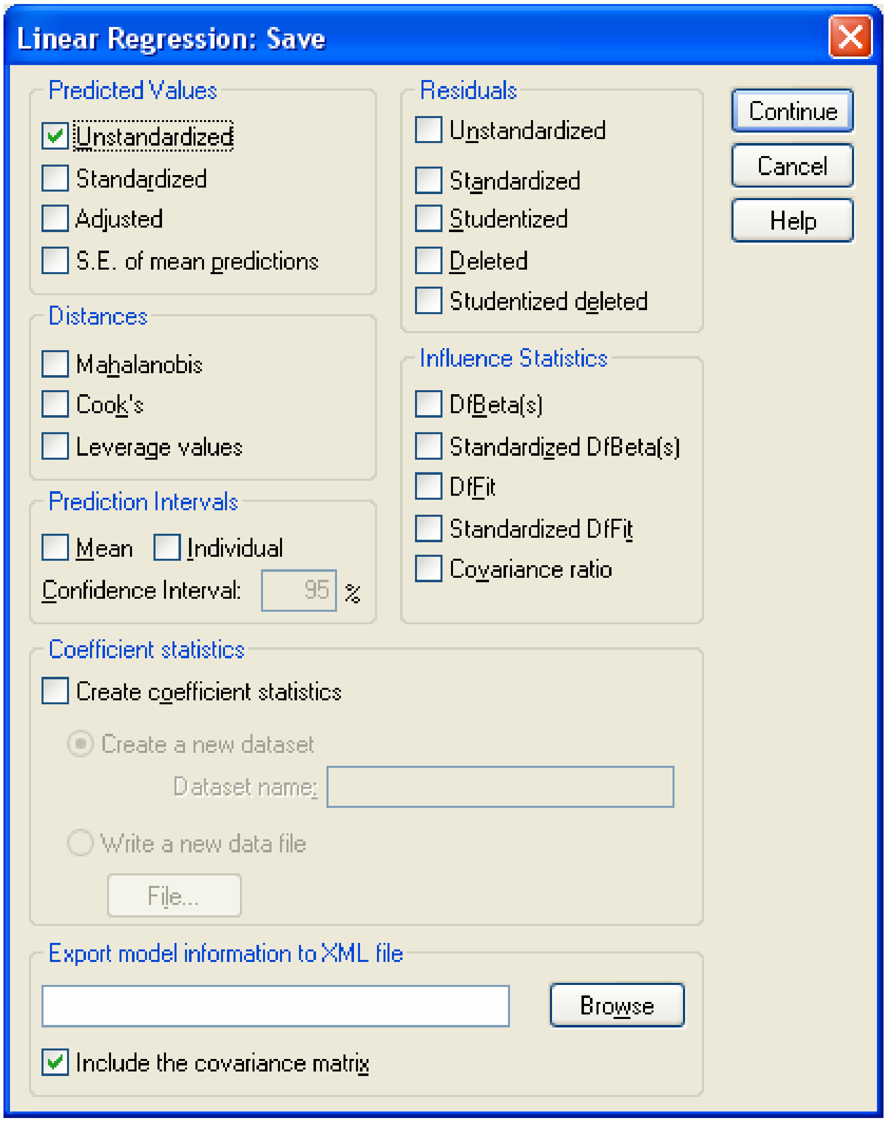
Method: Enter

Selection Variable:

Case Labels:

WLS Weight:

In the ‘Save…’ submenu, choose ‘Unstandardized predicted values’:



Linear Regression: Save

Predicted Values

☑ Unstandardized
□ Standardized
□ Adjusted
□ S.E. of mean predictions

Distances

□ Mahalanobis
□ Cook's
□ Leverage values

Prediction Intervals

□ Mean □ Individual

Confidence Interval: 95 %

Residuals

□ Unstandardized
□ Standardized
□ Studentized
□ Deleted
□ Studentized deleted

Influence Statistics

□ DfBeta(s)
□ Standardized DfBeta(s)
□ DfFit
□ Standardized DfFit
□ Covariance ratio

Coefficient statistics

□ Create coefficient statistics

● Create a new dataset

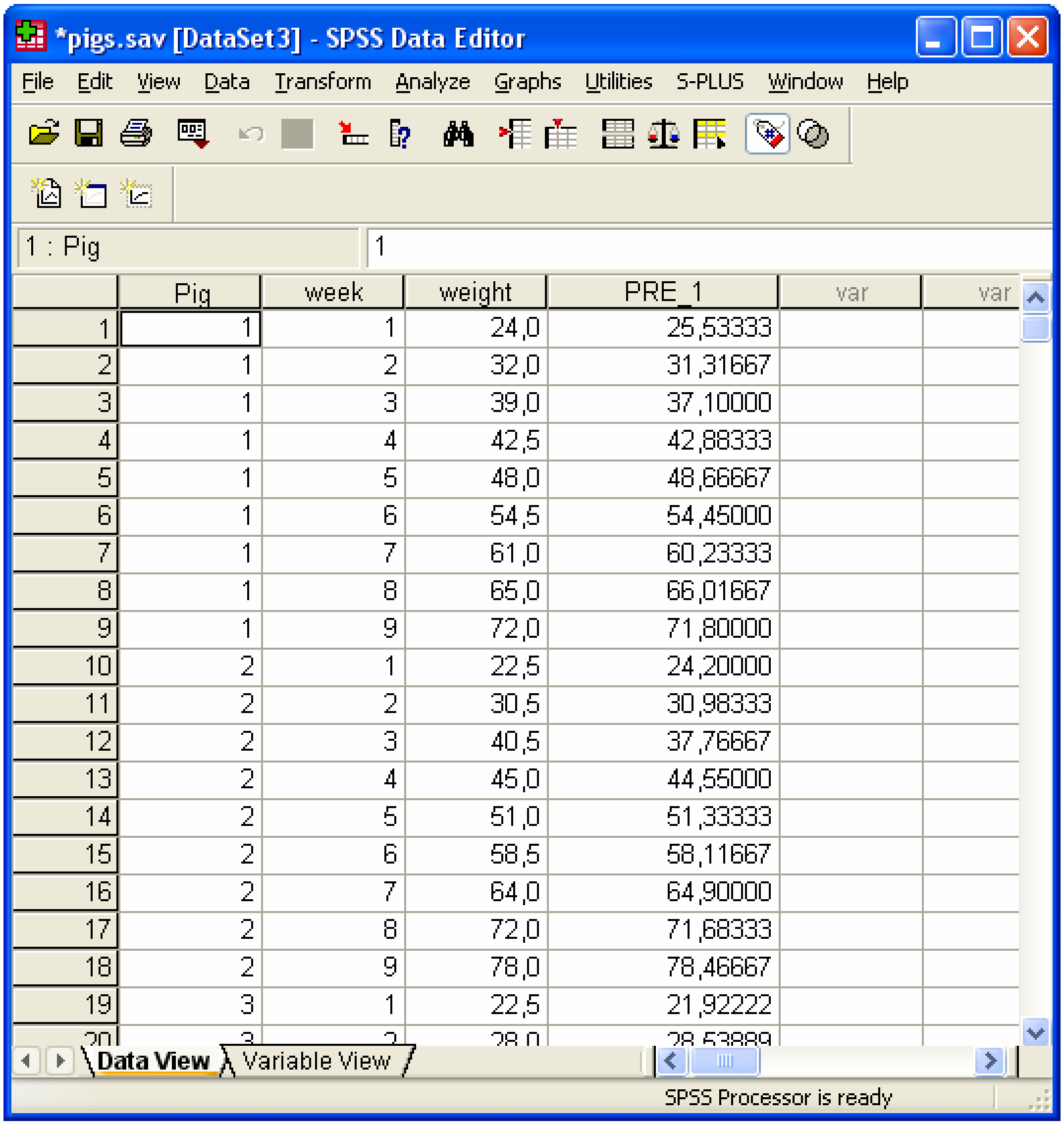
Dataset name:

○ Write a new data file

Export model information to XML file

☑ Include the covariance matrix

Then click OK. The individual regression equations are not of interest per se. We are only interested in the predicted values, which have been added to the data set in the data editor:


*pigs.sav [DataSet3] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View

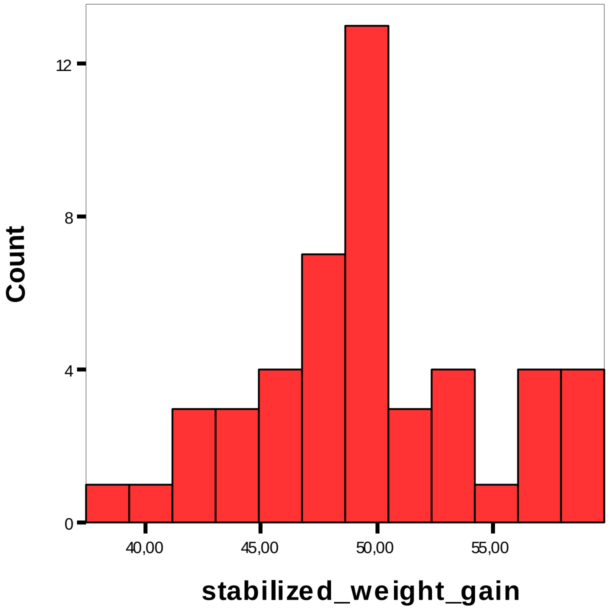
1 : Pig 1

	Pig	week	weight	PRE_1	var	var
1	1	1	24,0	25,53333		
2	1	2	32,0	31,31667		
3	1	3	39,0	37,10000		
4	1	4	42,5	42,88333		
5	1	5	48,0	48,66667		
6	1	6	54,5	54,45000		
7	1	7	61,0	60,23333		
8	1	8	65,0	66,01667		
9	1	9	72,0	71,80000		
10	2	1	22,5	24,20000		
11	2	2	30,5	30,98333		
12	2	3	40,5	37,76667		
13	2	4	45,0	44,55000		
14	2	5	51,0	51,33333		
15	2	6	58,5	58,11667		
16	2	7	64,0	64,90000		
17	2	8	72,0	71,68333		
18	2	9	78,0	78,46667		
19	3	1	22,5	21,92222		
20	3	2	28,0	28,53889		

SPSS Processor is ready

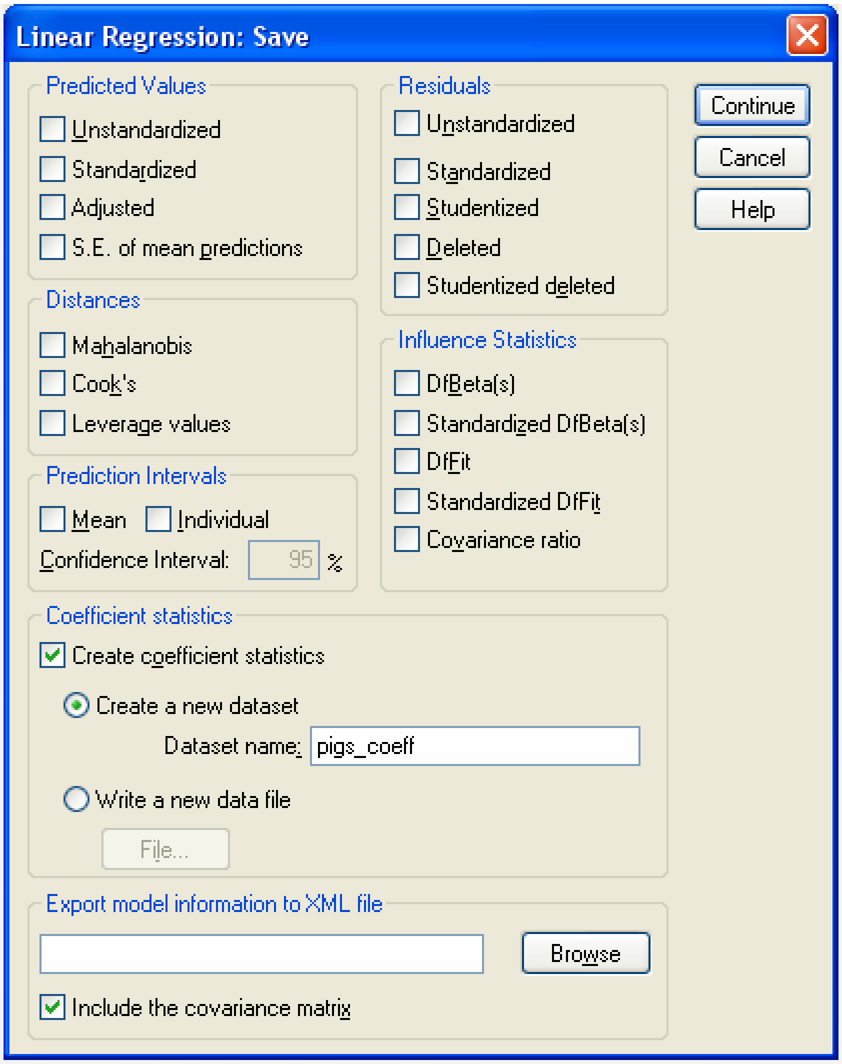
Now, we can proceed as in the preceding section on extracting the first and last value, to obtain a regression-stabilized estimate of the weight gain between weeks 1 and 9.

Again, the stabilized weight gain can be depicted in a histogram:



Computing the slope for each subject

We start with the data set pigs.sav. Use file splitting by pig ID (Data-Split file…) as in the preceding section. Choose Analyze-Regression-Linear and select weight and week as dependent and independent variables, respectively. Now press 'Save…':



Predicted Values

☐ Unstandardized
☐ Standardized
☐ Adjusted
☐ S.E. of mean predictions

Distances

☐ Mahalanobis
☐ Cook's
☐ Leverage values

Prediction Intervals

☐ Mean ☐ Individual
Confidence Interval: 95 %

Residuals

☐ Unstandardized
☐ Standardized
☐ Studentized
☐ Deleted
☐ Studentized deleted

Influence Statistics

☐ DfBeta(s)
☐ Standardized DfBeta(s)
☐ DfFit
☐ Standardized DfFit
☐ Covariance ratio

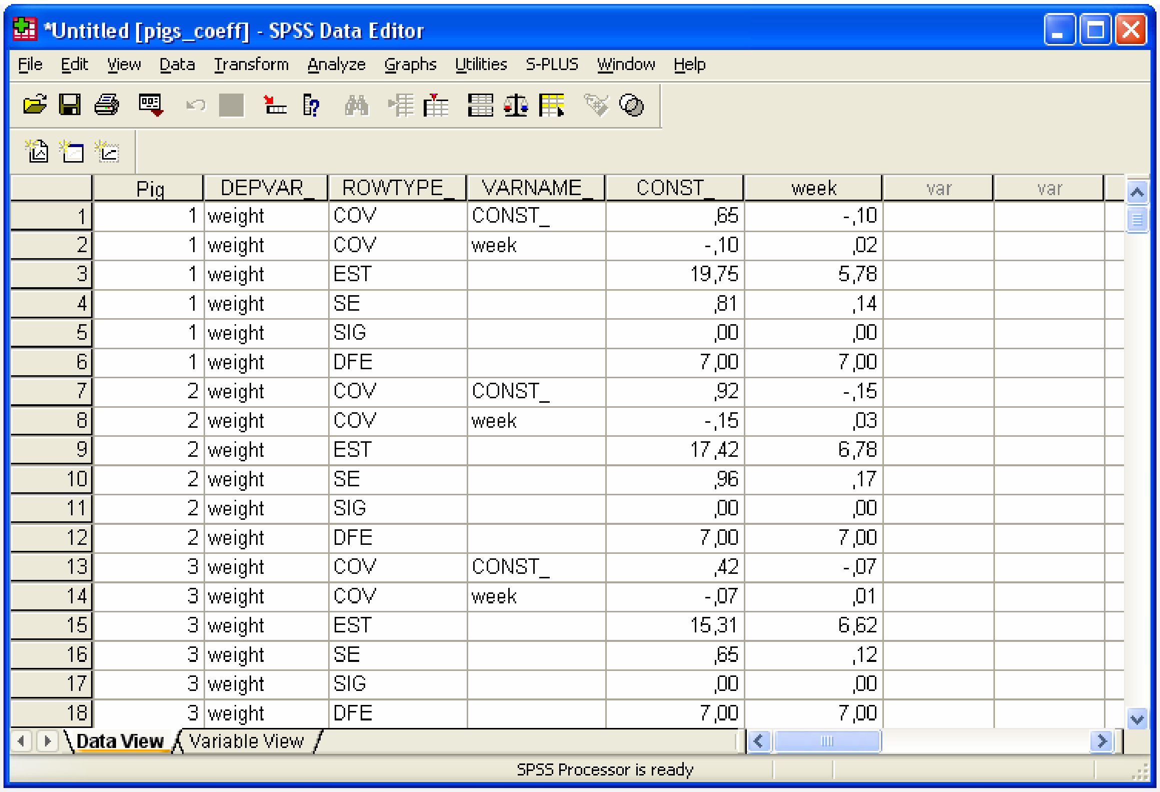
Coefficient statistics

☑ Create coefficient statistics
● Create a new dataset
Dataset name: pigs_coeff
○ Write a new data file
File...

Export model information to XML file

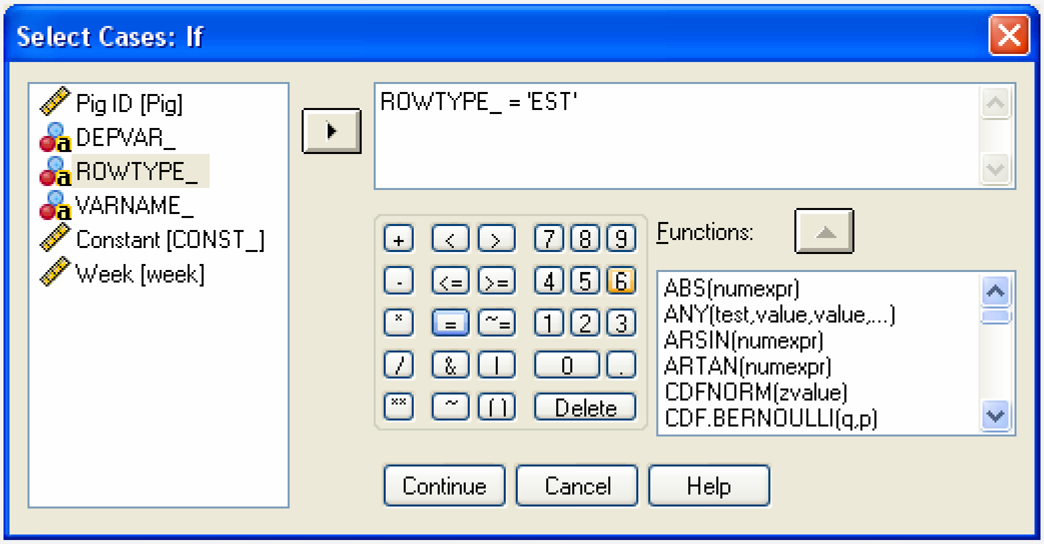
Browse
☑ Include the covariance matrix

Uncheck ‘Unstandardized’ predicted values, but check ‘Create coefficient statistics’. Name the new data set ‘pigs_coeff’. Confirm by ‘Continue’ and ‘OK’. Then, change to the new data set:

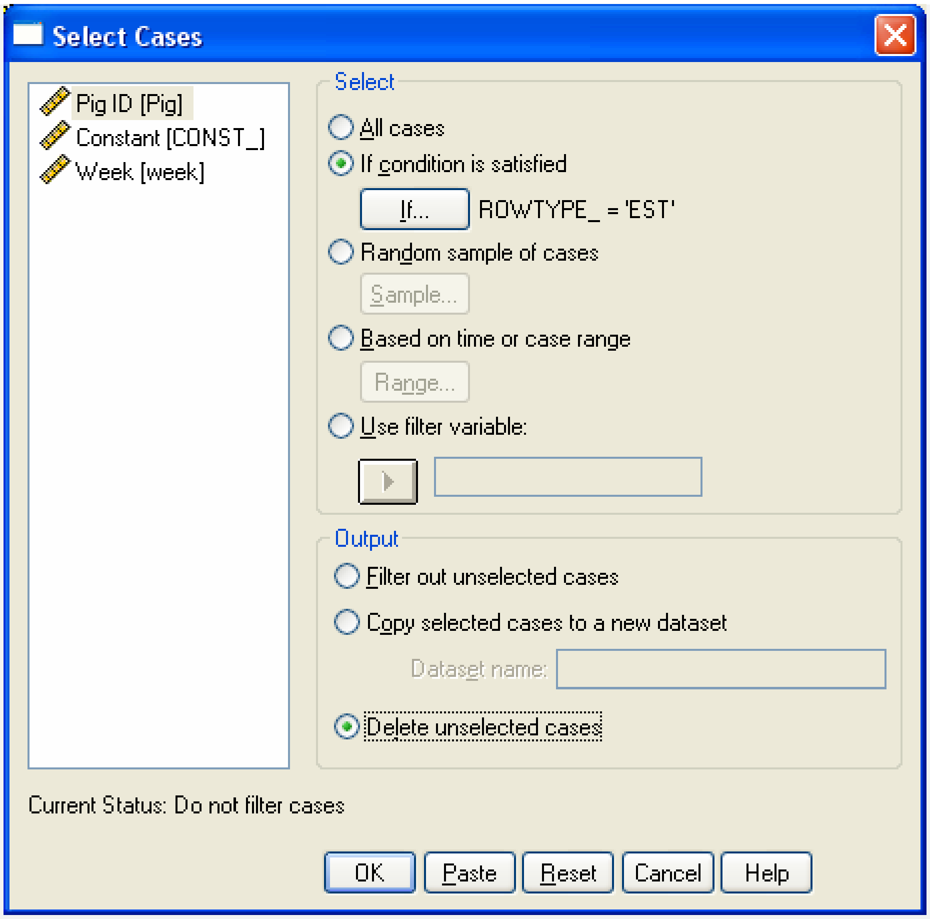

*Untitled [pigs_coeff] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View
SPSS Processor is ready

	Pig	DEPVAR_	ROWTYPE_	VARNAME_	CONST_	week	var	var
1	1	weight	COV	CONST_	,65	-,10		
2	1	weight	COV	week	-,10	,02		
3	1	weight	EST		19,75	5,78		
4	1	weight	SE		,81	,14		
5	1	weight	SIG		,00	,00		
6	1	weight	DFE		7,00	7,00		
7	2	weight	COV	CONST_	,92	-,15		
8	2	weight	COV	week	-,15	,03		
9	2	weight	EST		17,42	6,78		
10	2	weight	SE		,96	,17		
11	2	weight	SIG		,00	,00		
12	2	weight	DFE		7,00	7,00		
13	3	weight	COV	CONST_	,42	-,07		
14	3	weight	COV	week	-,07	,01		
15	3	weight	EST		15,31	6,62		
16	3	weight	SE		,65	,12		
17	3	weight	SIG		,00	,00		
18	3	weight	DFE		7,00	7,00		

Next, choose ‘Data-Select Cases’. Select ‘If condition is satisfied’ and click on ‘If’. Fill in as follows:

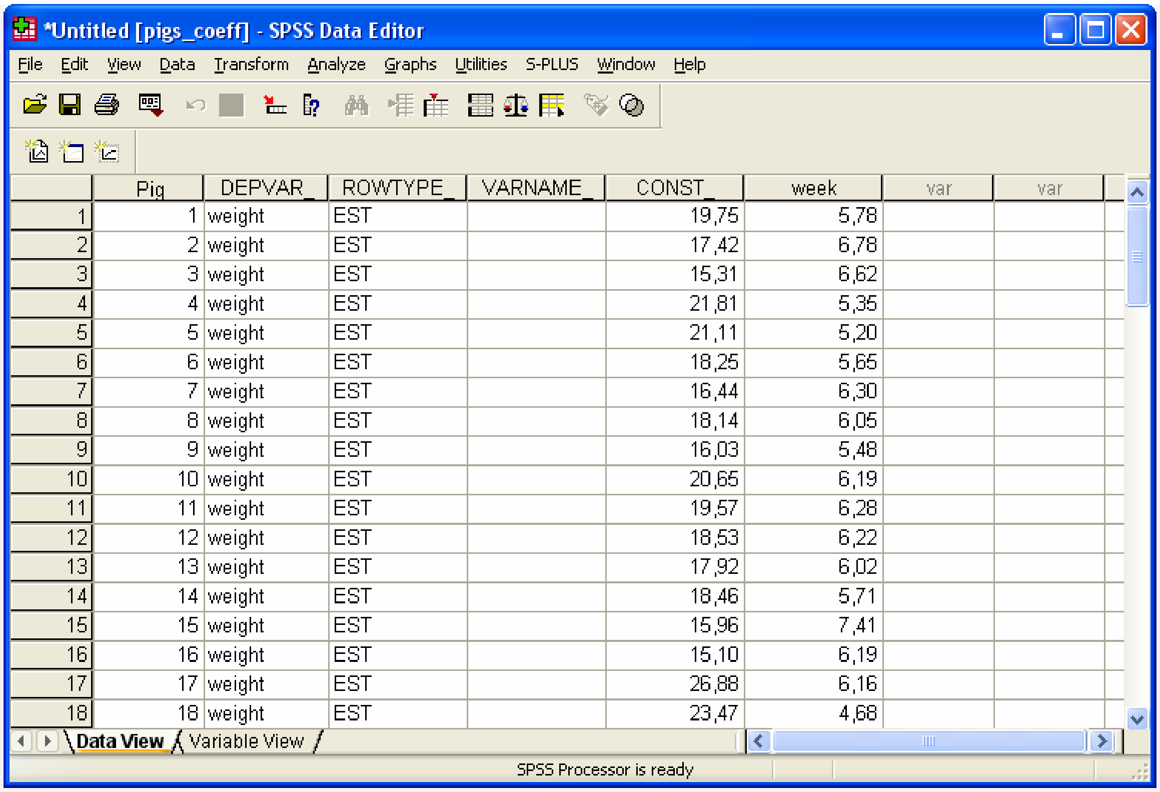


ROWTYPE_ = 'EST'



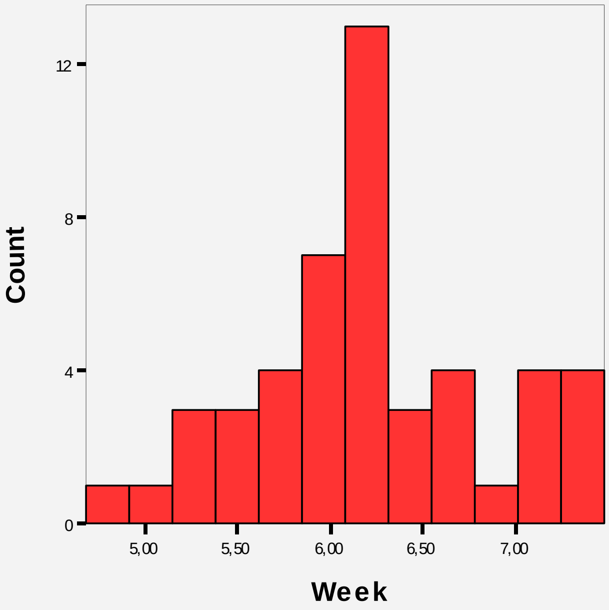
Select ‘Delete unselected cases’. After pressing OK, we arrive at the following data set (see next page).

Now, there is only one row per pig, containing each individual’s regression equation. For pig no. 1, the regression equation is weight = 19.75 + 5.78 * week. For pig no. 2, it is 17.42 + 6.78 * week, etc. Clearly, the slopes are contained in the column named ‘week’ (as these numbers are the regression coefficients corresponding to the independent variable ‘week’). The slopes can be interpreted as the average weight gain per week and are distributed as shown by the histogram below.

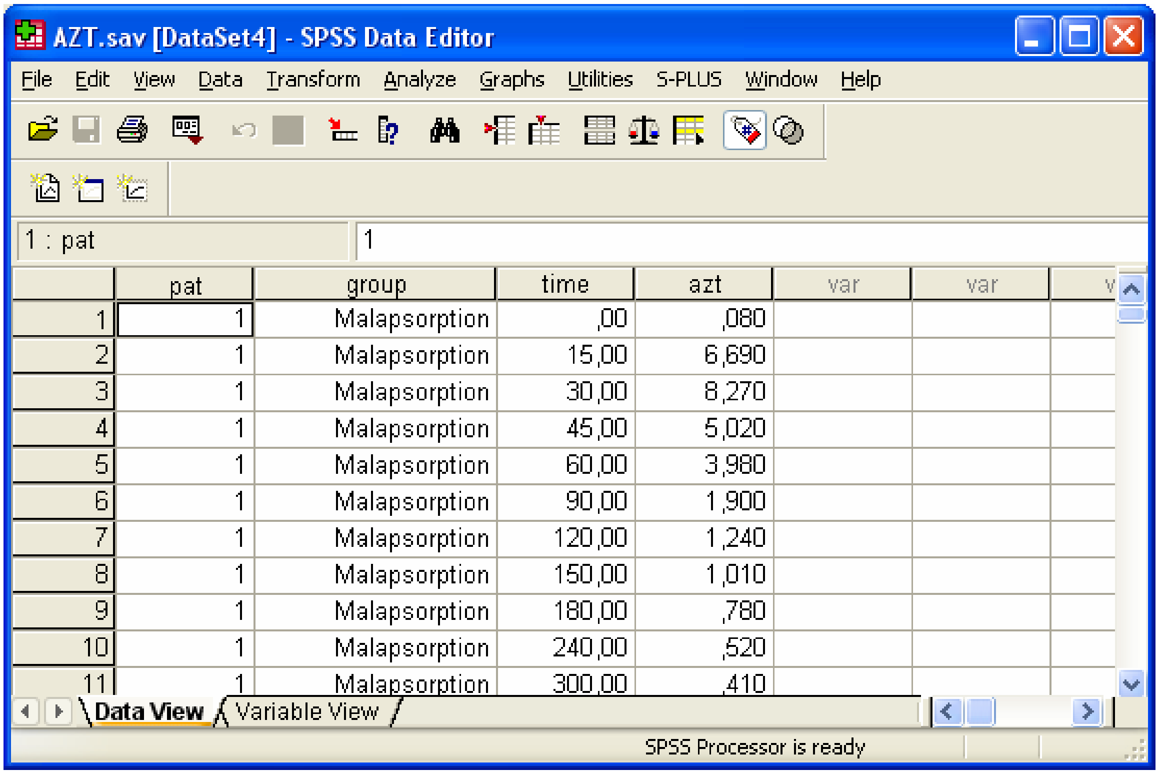

*Untitled [pigs_coeff] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View

	Pig	DEPVAR_	ROWTYPE_	VARNAME_	CONST_	week	var	var
1	1	weight	EST		19,75	5,78		
2	2	weight	EST		17,42	6,78		
3	3	weight	EST		15,31	6,62		
4	4	weight	EST		21,81	5,35		
5	5	weight	EST		21,11	5,20		
6	6	weight	EST		18,25	5,65		
7	7	weight	EST		16,44	6,30		
8	8	weight	EST		18,14	6,05		
9	9	weight	EST		16,03	5,48		
10	10	weight	EST		20,65	6,19		
11	11	weight	EST		19,57	6,28		
12	12	weight	EST		18,53	6,22		
13	13	weight	EST		17,92	6,02		
14	14	weight	EST		18,46	5,71		
15	15	weight	EST		15,96	7,41		
16	16	weight	EST		15,10	6,19		
17	17	weight	EST		26,88	6,16		
18	18	weight	EST		23,47	4,68		

SPSS Processor is ready



Summary measures for peaked curves are exemplified on the data set AZT.sav:

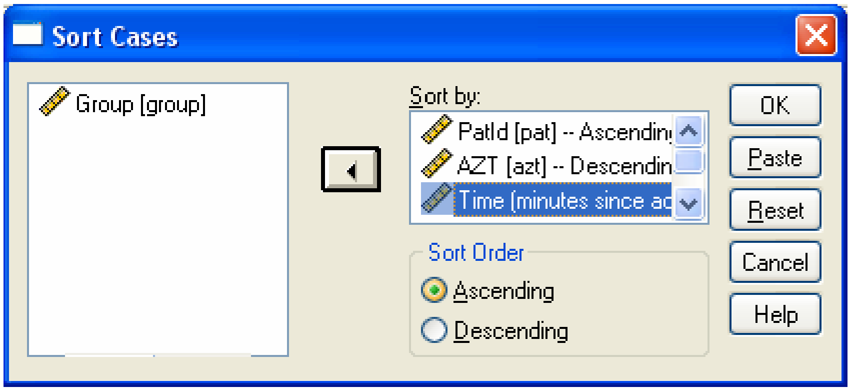

AZT.sav [DataSet4] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View
SPSS Processor is ready

1 : pat 1

	pat	group	time	azt	var	var	v
1	1	Malapsorption	,00	,080			
2	1	Malapsorption	15,00	6,690			
3	1	Malapsorption	30,00	8,270			
4	1	Malapsorption	45,00	5,020			
5	1	Malapsorption	60,00	3,980			
6	1	Malapsorption	90,00	1,900			
7	1	Malapsorption	120,00	1,240			
8	1	Malapsorption	150,00	1,010			
9	1	Malapsorption	180,00	,780			
10	1	Malapsorption	240,00	,520			
11	1	Malapsorption	300,00	,410			

Extracting C_{max} and T_{max}

First, sort the data set by patient ID, descending AZT value and ascending time (Data-Sort Cases):

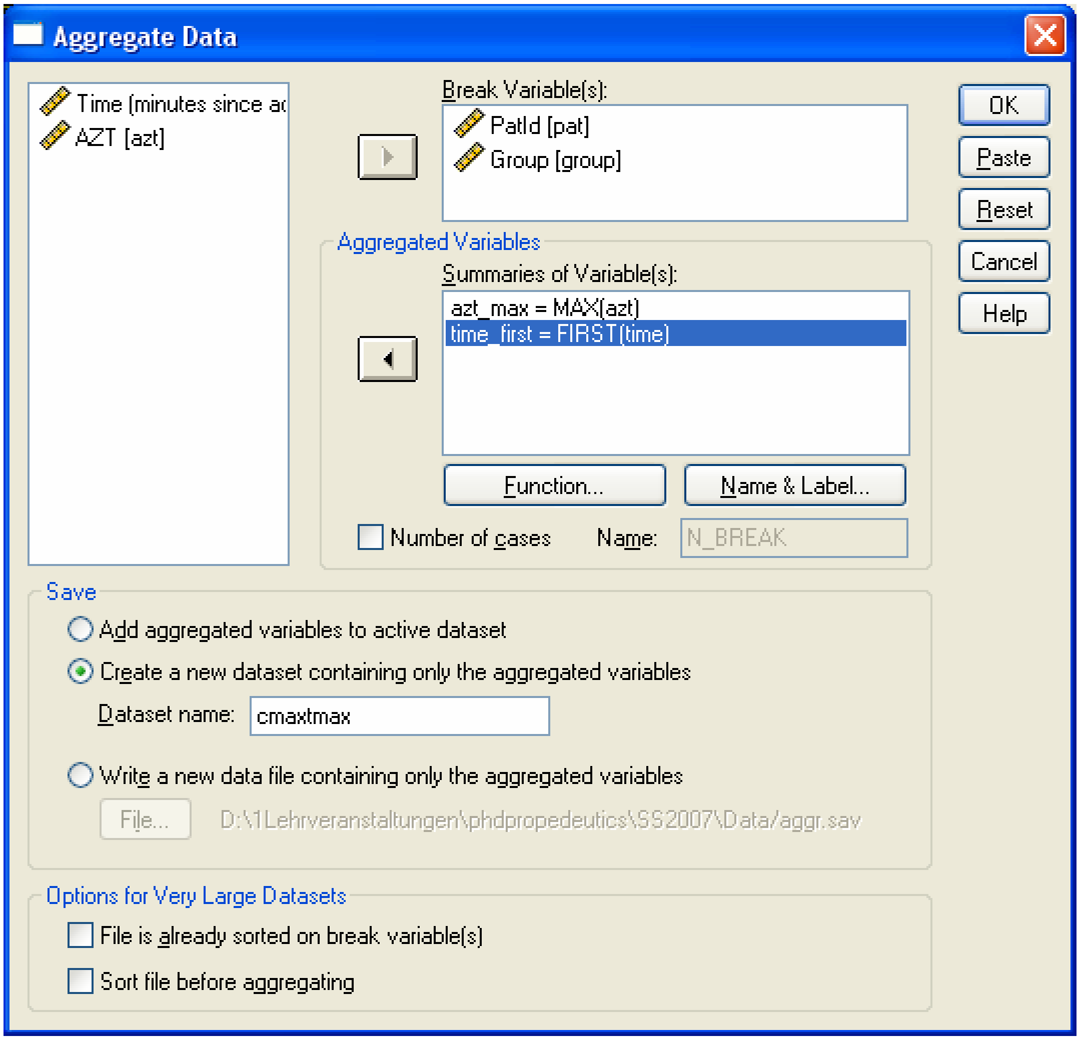


Group [group]

Sort by:
PatId [pat] -- Ascendin
AZT [azt] -- Descendin
Time (minutes since ac

Sort Order
● Ascending
○ Descending

Next, choose 'Data-Aggregate'. Define 'PatID' and 'Group' as Break Variable(s) and move AZT and time into the field 'Summaries of Variable(s)'. For azt, change the summary function to 'Maximum' and for time, change to 'First value'. Select 'Create a new dataset containing only the aggregated variables' and give it a name (e. g., CmaxTmax):



Break Variable(s):
PatId [pat]
Group [group]

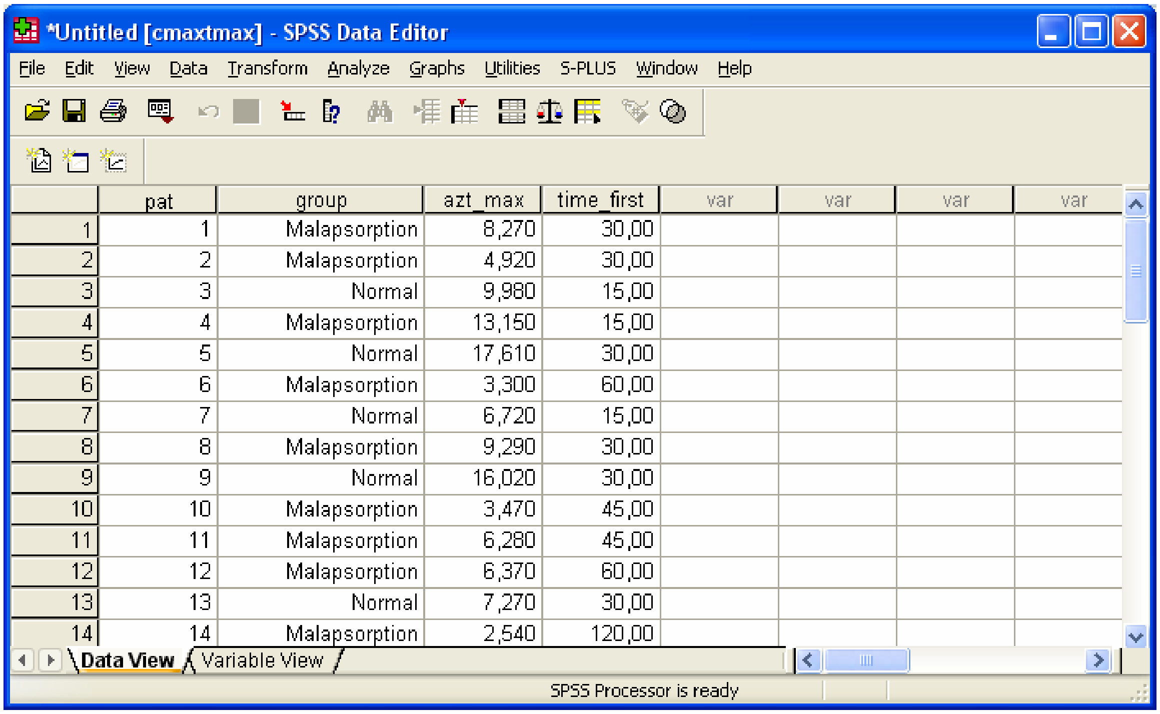
Aggregated Variables
Summaries of Variable(s):
azt_max = MAX(azt)
time_first = FIRST(time)

□ Number of cases Name: N_BREAK

Save
○ Add aggregated variables to active dataset
● Create a new dataset containing only the aggregated variables
Dataset name: cmaxtmax
○ Write a new data file containing only the aggregated variables
D:\1Lehrveranstaltungen\phdpropedeutics\SS2007\Data/aggr.sav

Options for Very Large Datasets
□ File is already sorted on break variable(s)
□ Sort file before aggregating

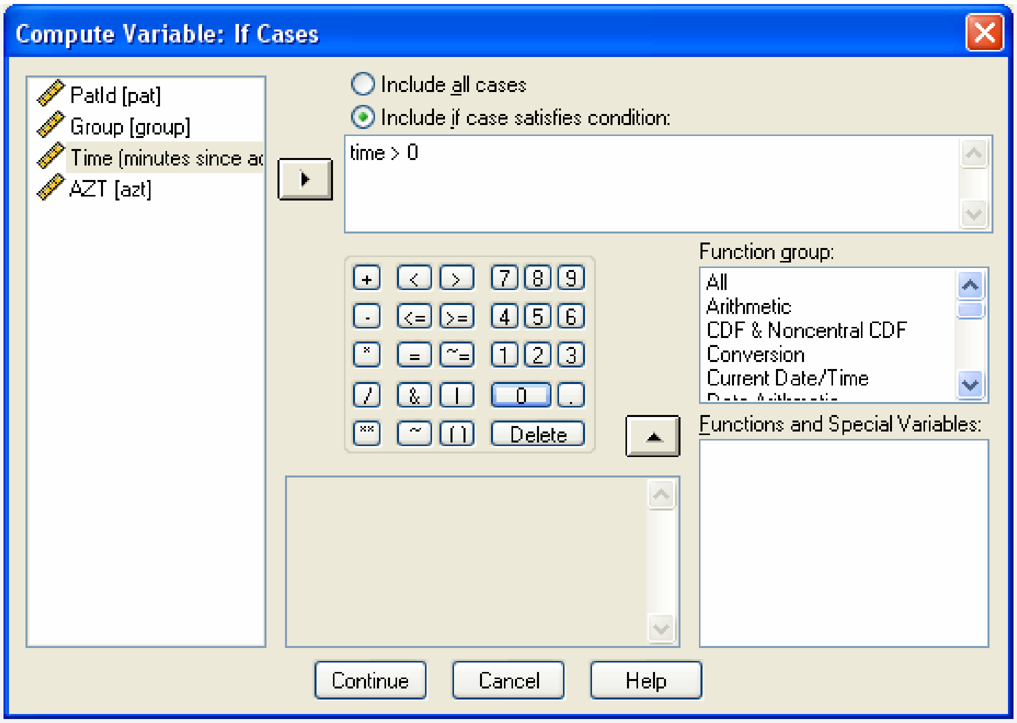
Change to the new data set, it contains the Cmax (azt_max) and Tmax (time_first) values:


*Untitled [cmaxtmax] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View
SPSS Processor is ready

	pat	group	azt_max	time_first	var	var	var	var
1	1	Malapsorption	8,270	30,00				
2	2	Malapsorption	4,920	30,00				
3	3	Normal	9,980	15,00				
4	4	Malapsorption	13,150	15,00				
5	5	Normal	17,610	30,00				
6	6	Malapsorption	3,300	60,00				
7	7	Normal	6,720	15,00				
8	8	Malapsorption	9,290	30,00				
9	9	Normal	16,020	30,00				
10	10	Malapsorption	3,470	45,00				
11	11	Malapsorption	6,280	45,00				
12	12	Malapsorption	6,370	60,00				
13	13	Normal	7,270	30,00				
14	14	Malapsorption	2,540	120,00				

Computing the area under the curve

The area under the curve can be computed using the trapezoidal rule. First, revert to the original AZT.sav data set. Next, sort by PatID and ascending time (Data-Sort Cases). Then, call ‘Transform-Compute’. The height of the partial rectangles is called ‘height’. The numeric expression should read (azt + lag(azt)) / 2. Remember that the lag function moves the preceding azt value into the respective subsequent line. We do not want this computation for the first line of each patient (since there is no rectangle to compute). Therefore, click on ‘If…’ and request ‘time > 0’:

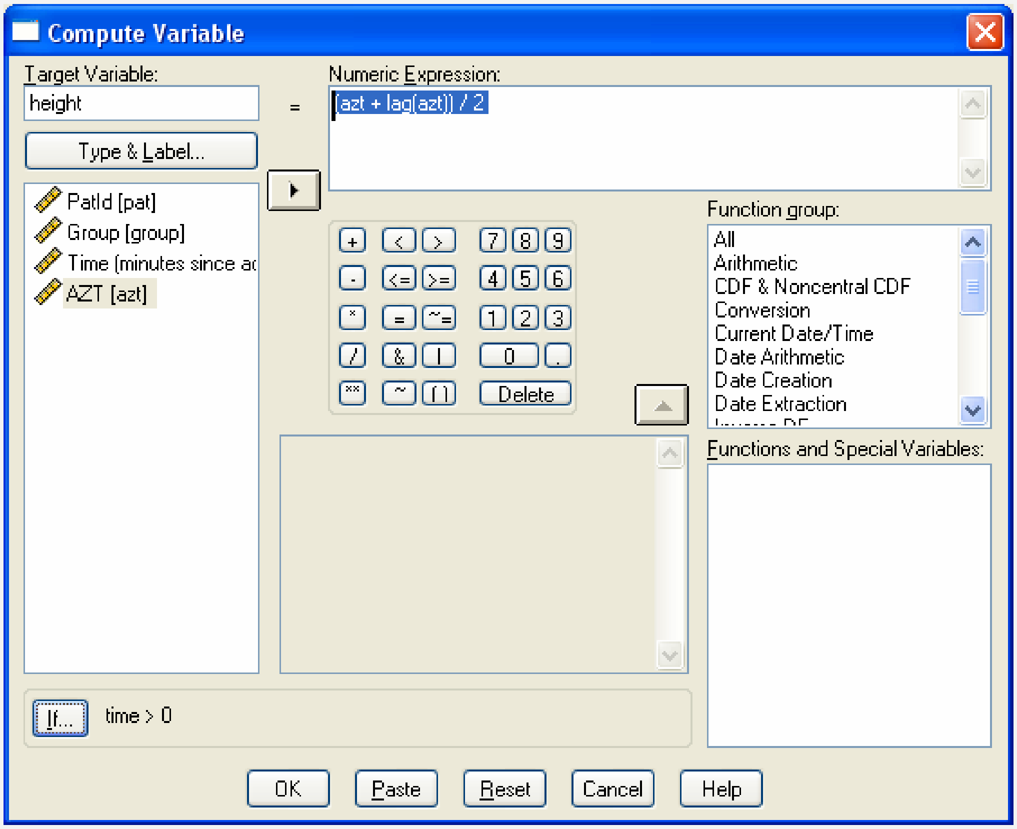


Compute Variable: If Cases

○ Include all cases

● Include if case satisfies condition:

```
time > 0
```



Compute Variable

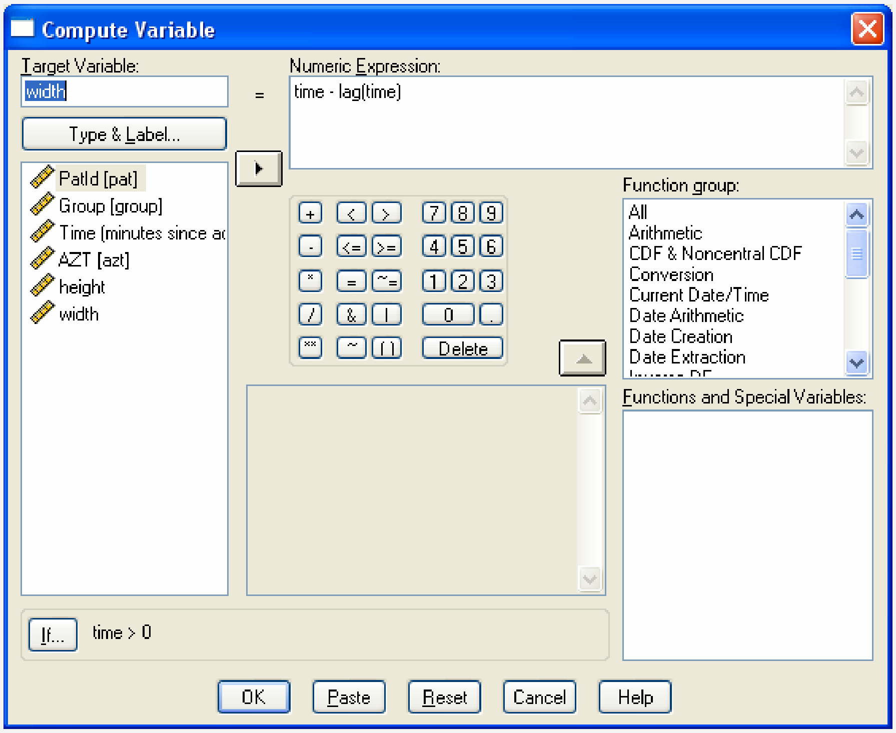
Target Variable: height

Numeric Expression:

```
(azt + lag(azt)) / 2
```

If... time > 0

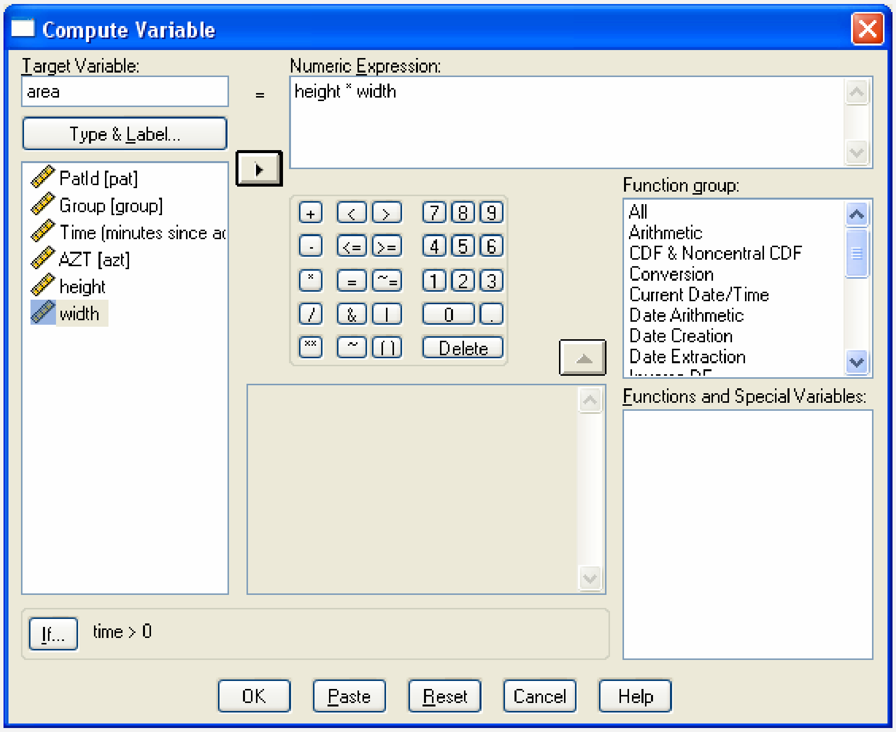
The width of the time intervals is computed similarly, by requesting width = time – lag(time). Finally, the partial areas are computed by area = height * time. [You may also compute the areas directly by area = (time – lag(time))*(azt + lag(azt))/2.]


Compute Variable

Target Variable: width

Numeric Expression: time - lag(time)

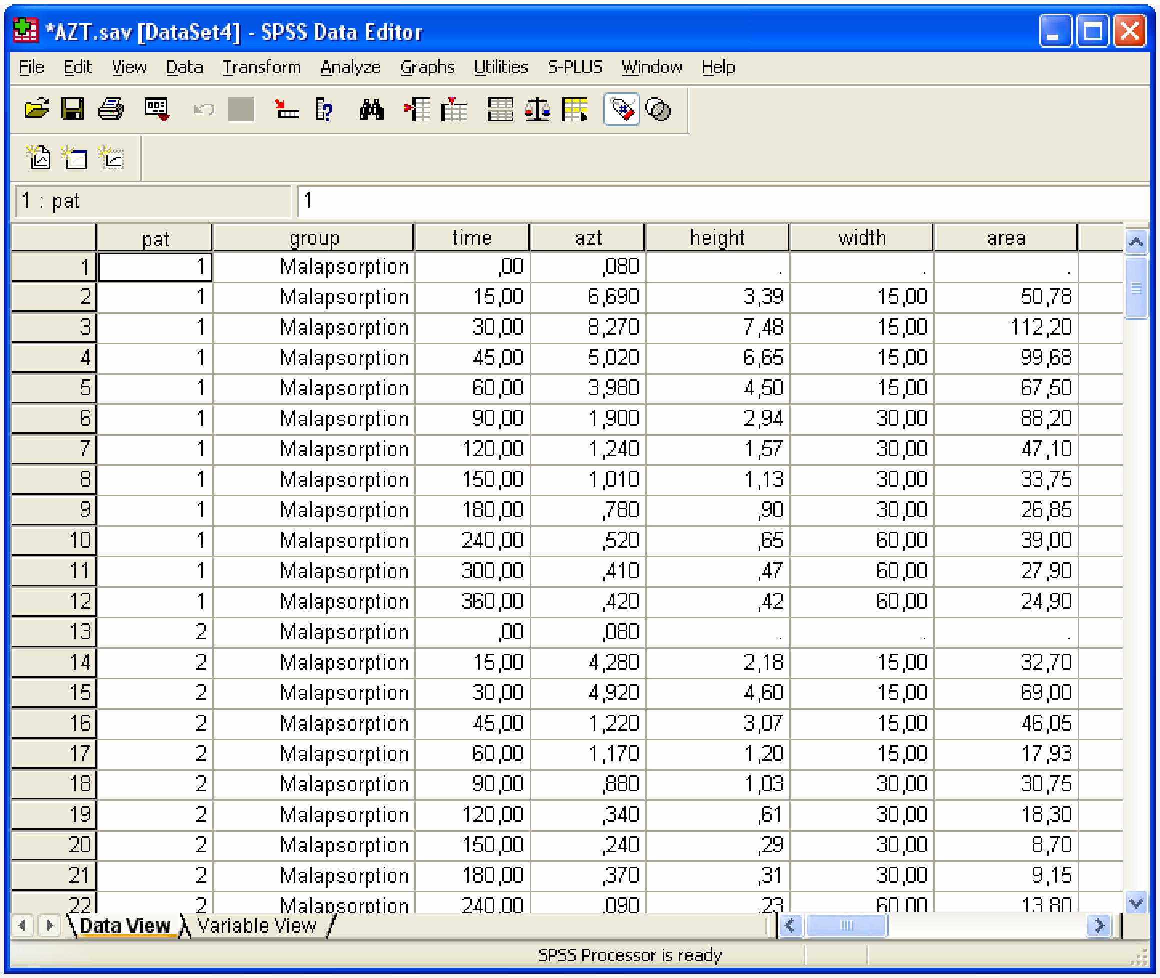
If... time > 0


Compute Variable

Target Variable: area

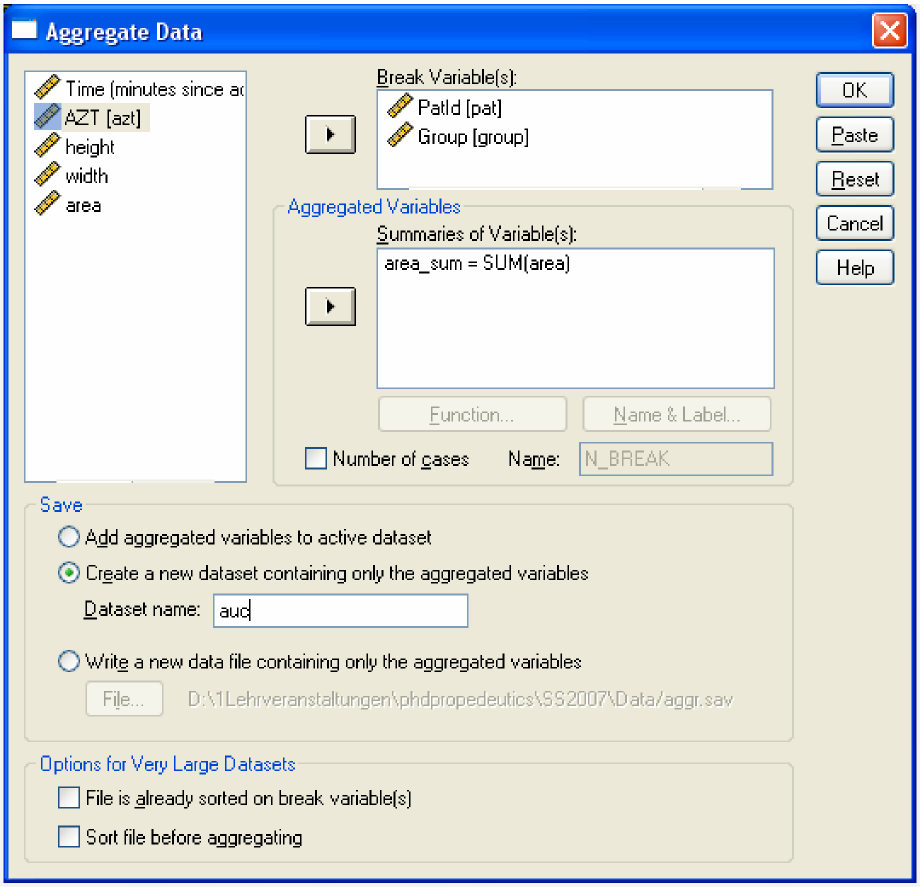
Numeric Expression: height * width

If... time > 0


*AZT.sav [DataSet4] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 : pat
1
Data View
Variable View
SPSS Processor is ready

	pat	group	time	azt	height	width	area
1	1	Malapsorption	,00	,080	.	.	.
2	1	Malapsorption	15,00	6,690	3,39	15,00	50,78
3	1	Malapsorption	30,00	8,270	7,48	15,00	112,20
4	1	Malapsorption	45,00	5,020	6,65	15,00	99,68
5	1	Malapsorption	60,00	3,980	4,50	15,00	67,50
6	1	Malapsorption	90,00	1,900	2,94	30,00	88,20
7	1	Malapsorption	120,00	1,240	1,57	30,00	47,10
8	1	Malapsorption	150,00	1,010	1,13	30,00	33,75
9	1	Malapsorption	180,00	,780	,90	30,00	26,85
10	1	Malapsorption	240,00	,520	,65	60,00	39,00
11	1	Malapsorption	300,00	,410	,47	60,00	27,90
12	1	Malapsorption	360,00	,420	,42	60,00	24,90
13	2	Malapsorption	,00	,080	.	.	.
14	2	Malapsorption	15,00	4,280	2,18	15,00	32,70
15	2	Malapsorption	30,00	4,920	4,60	15,00	69,00
16	2	Malapsorption	45,00	1,220	3,07	15,00	46,05
17	2	Malapsorption	60,00	1,170	1,20	15,00	17,93
18	2	Malapsorption	90,00	,880	1,03	30,00	30,75
19	2	Malapsorption	120,00	,340	,61	30,00	18,30
20	2	Malapsorption	150,00	,240	,29	30,00	8,70
21	2	Malapsorption	180,00	,370	,31	30,00	9,15
22	2	Malapsorption	240,00	,090	,23	60,00	13,80

Next, we sum up the partial areas across all measurements on the same subject. Call ‘Data-Aggregate’ and fill in the fields as follows. Change the function corresponding to the summary of ‘area’ to ‘Sum’:



Aggregate Data

Time (minutes since ac
AZT [azt]
height
width
area

Break Variable(s):
PatId [pat]
Group [group]

Aggregated Variables

Summaries of Variable(s):
area_sum = SUM(area)

□ Number of cases Name: N_BREAK

Save

○ Add aggregated variables to active dataset

● Create a new dataset containing only the aggregated variables

Dataset name: auc

○ Write a new data file containing only the aggregated variables

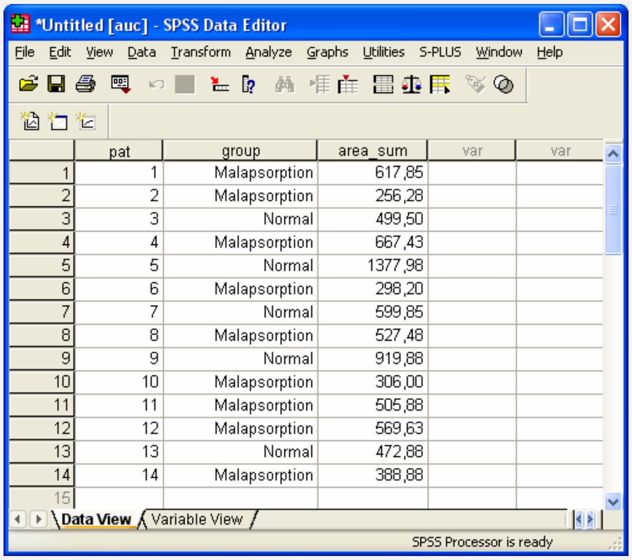
File... D:\1Lehrveranstaltungen\phdpropedeutics\SS2007\Data/aggr.sav

Options for Very Large Datasets

□ File is already sorted on break variable(s)

□ Sort file before aggregating

The final data set is as follows:



*Untitled [auc] - SPSS Data Editor

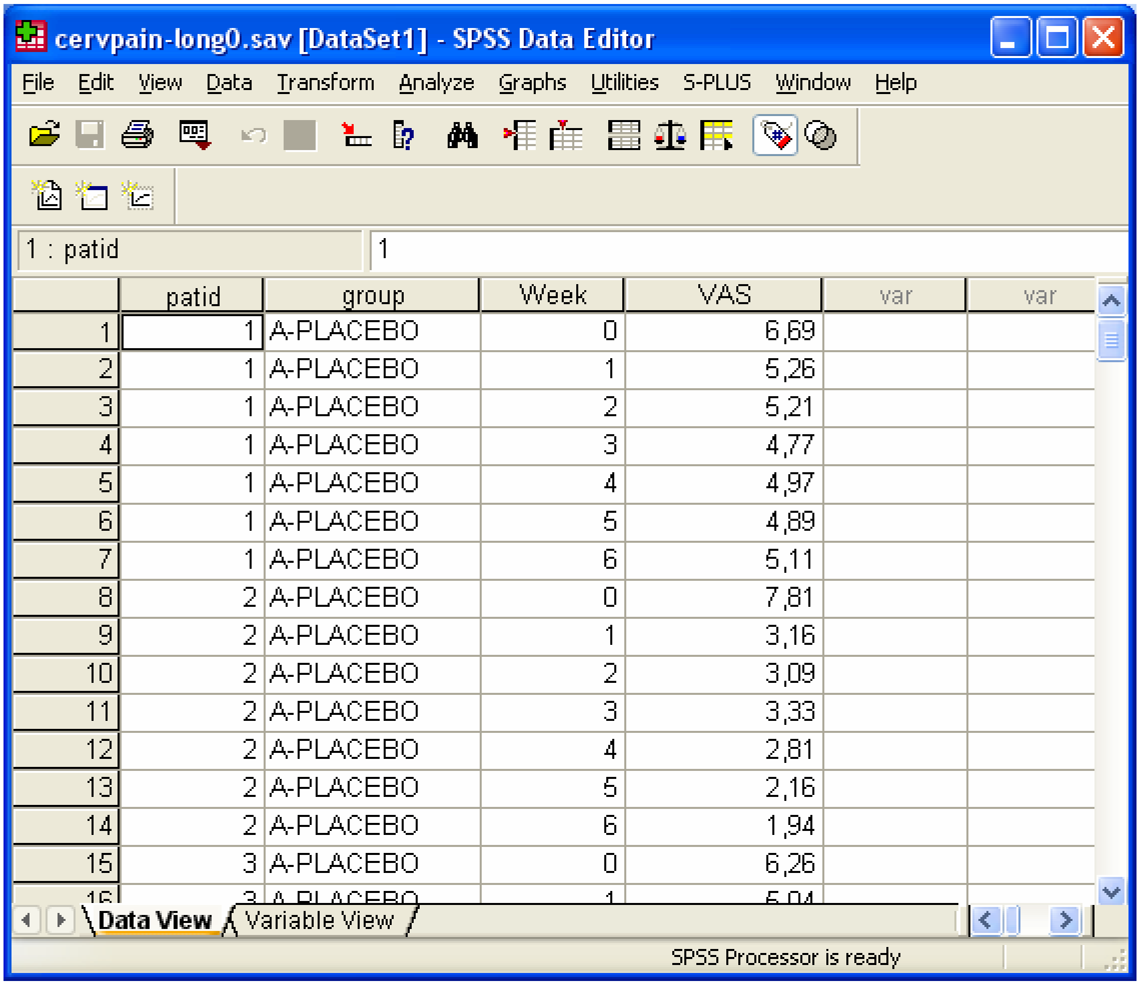
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
Data View
Variable View
SPSS Processor is ready

	pat	group	area_sum	var	var
1	1	Malapsorption	617,85		
2	2	Malapsorption	256,28		
3	3	Normal	499,50		
4	4	Malapsorption	667,43		
5	5	Normal	1377,98		
6	6	Malapsorption	298,20		
7	7	Normal	599,85		
8	8	Malapsorption	527,48		
9	9	Normal	919,88		
10	10	Malapsorption	306,00		
11	11	Malapsorption	505,88		
12	12	Malapsorption	569,63		
13	13	Normal	472,88		
14	14	Malapsorption	388,88		
15					

Please note that the computation of AUC, Cmax and Tmax could also be done in one task, but care must be taken when sorting the data (you need to resort the data before you compute the partial areas).

ANOVA for repeated measurements

Open the data set cervpain-long0sav. This data set contains the data in long format, but with an additional line (‘week’=0) for the baseline measurement:

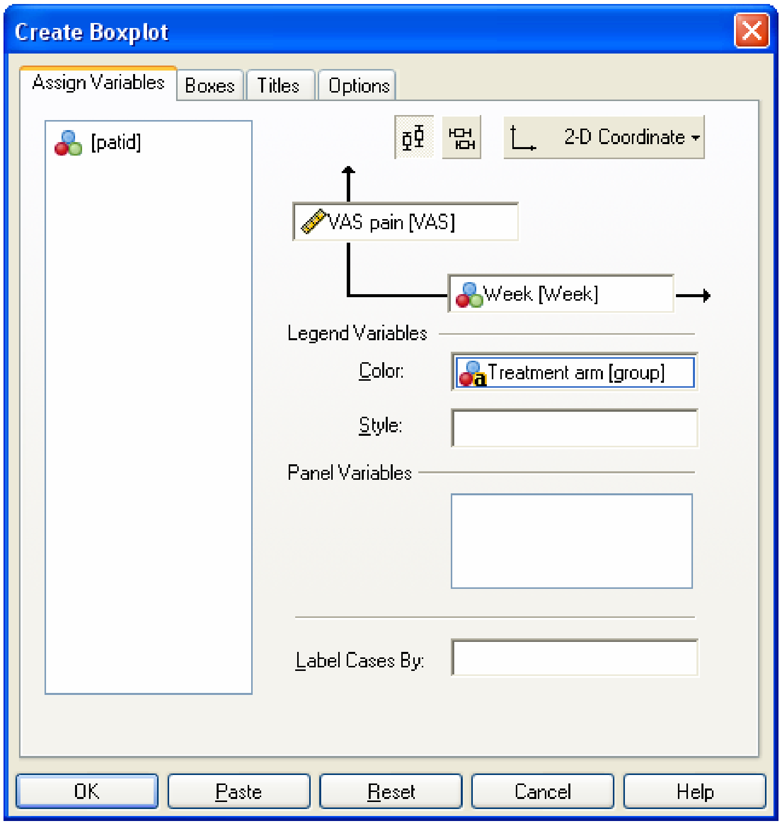


1 : patid | 1

	patid	group	Week	VAS	var	var
1	1	A-PLACEBO	0	6,69		
2	1	A-PLACEBO	1	5,26		
3	1	A-PLACEBO	2	5,21		
4	1	A-PLACEBO	3	4,77		
5	1	A-PLACEBO	4	4,97		
6	1	A-PLACEBO	5	4,89		
7	1	A-PLACEBO	6	5,11		
8	2	A-PLACEBO	0	7,81		
9	2	A-PLACEBO	1	3,16		
10	2	A-PLACEBO	2	3,09		
11	2	A-PLACEBO	3	3,33		
12	2	A-PLACEBO	4	2,81		
13	2	A-PLACEBO	5	2,16		
14	2	A-PLACEBO	6	1,94		
15	3	A-PLACEBO	0	6,26		
16	3	A-PLACEBO	1	6,04		

SPSS Processor is ready

For an exploratory descriptive analysis, we first create a box plot (Graphs-Interactive-Boxplot):



Create Boxplot

Assign Variables | Boxes | Titles | Options

[patid]

2-D Coordinate

VAS pain [VAS]

Week [Week]

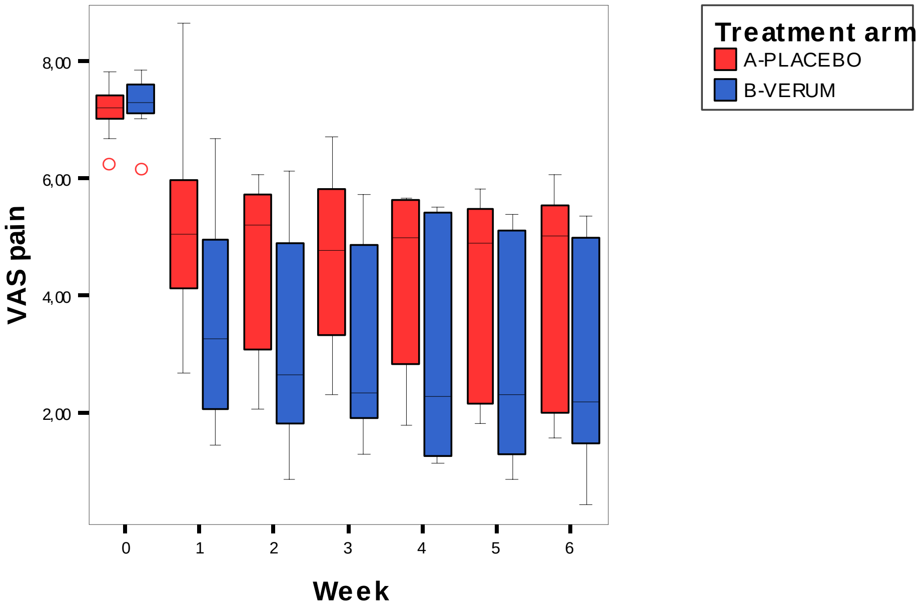
Legend Variables

Color: Treatment arm [group]

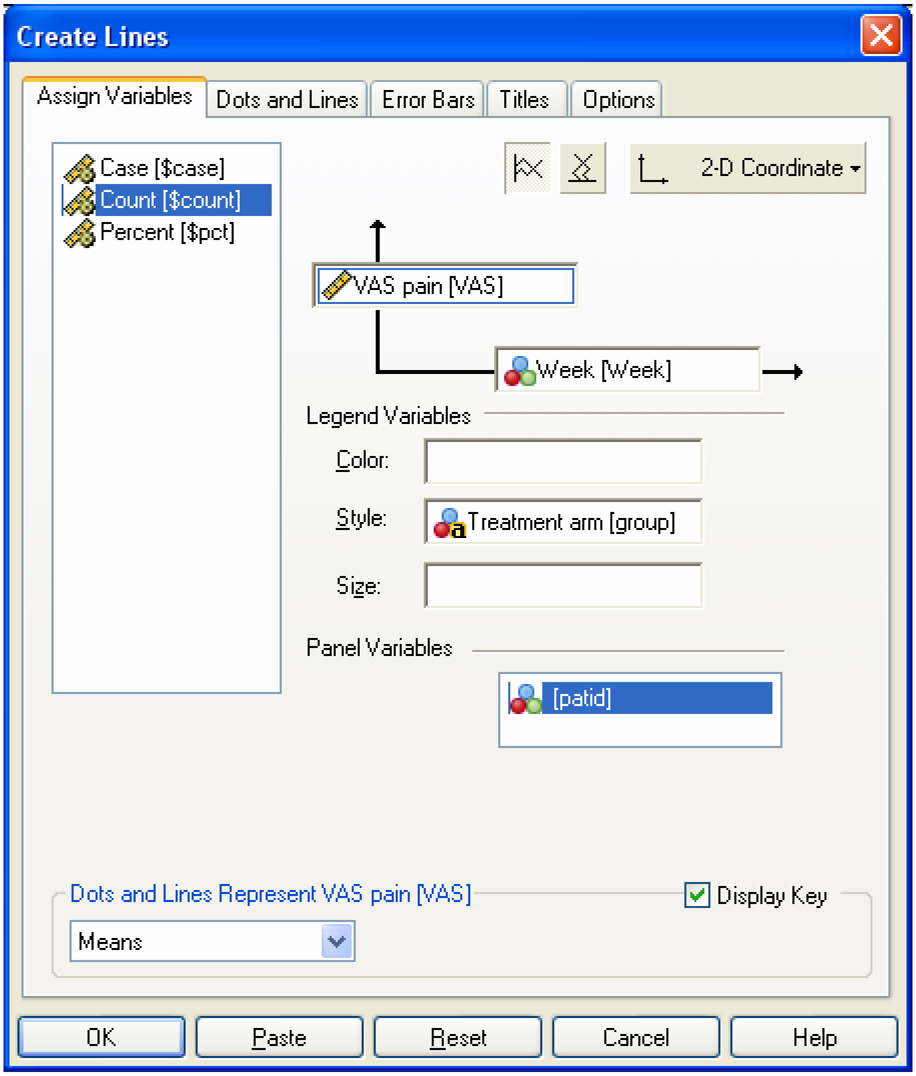
Style:

Panel Variables

Label Cases By:



Next, we create a plot of individual lines. Select Graphs-Interactive-Line:



VAS pain [VAS]

Week [Week]

Color:

Style: Treatment arm [group]

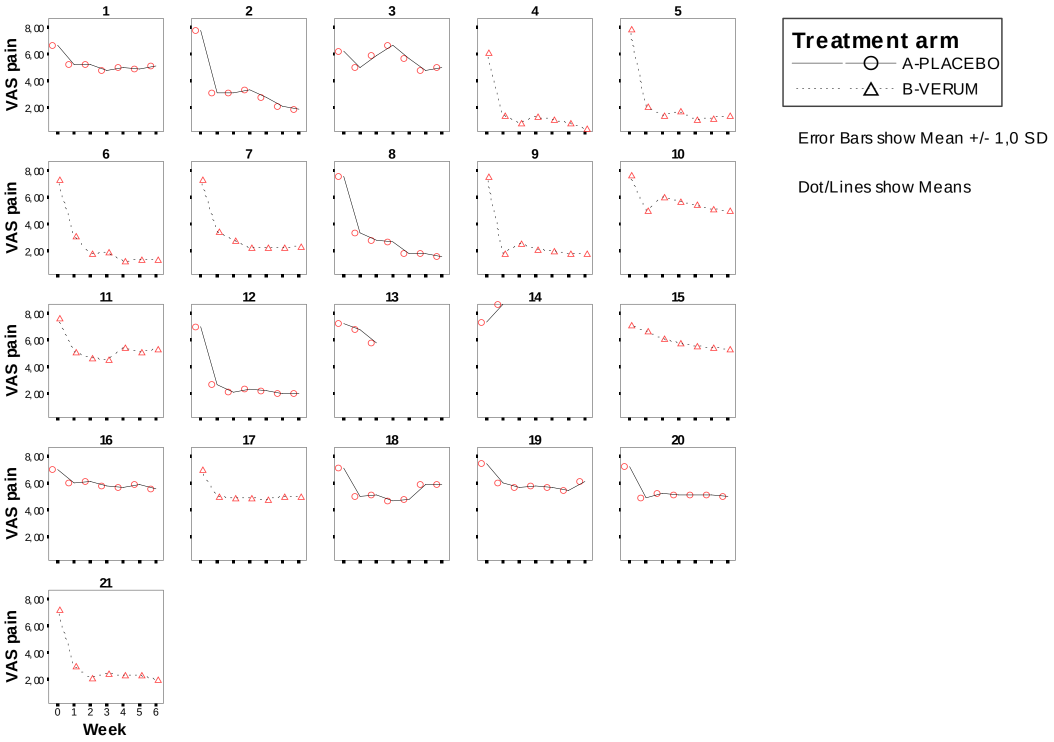
Size:

[patid]

Dots and Lines Represent VAS pain [VAS]

☑ Display Key

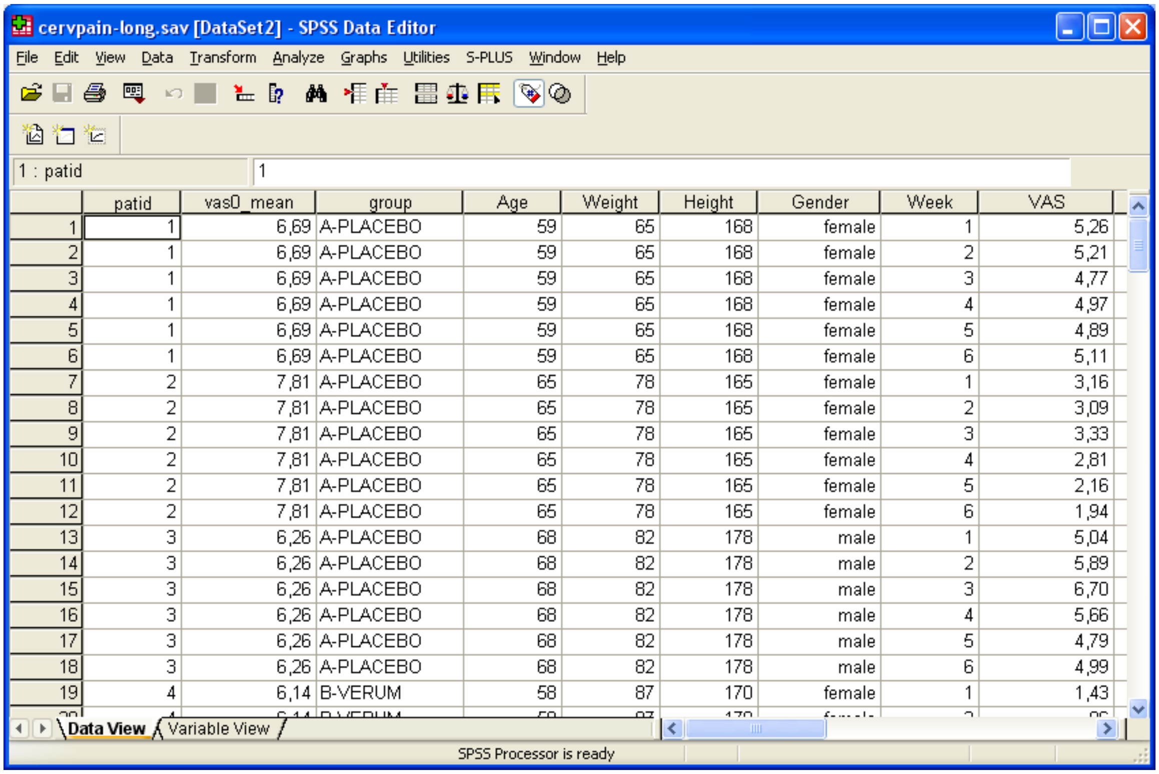
Means



Error Bars show Mean +/- 1,0 SD

Dot/Lines show Means

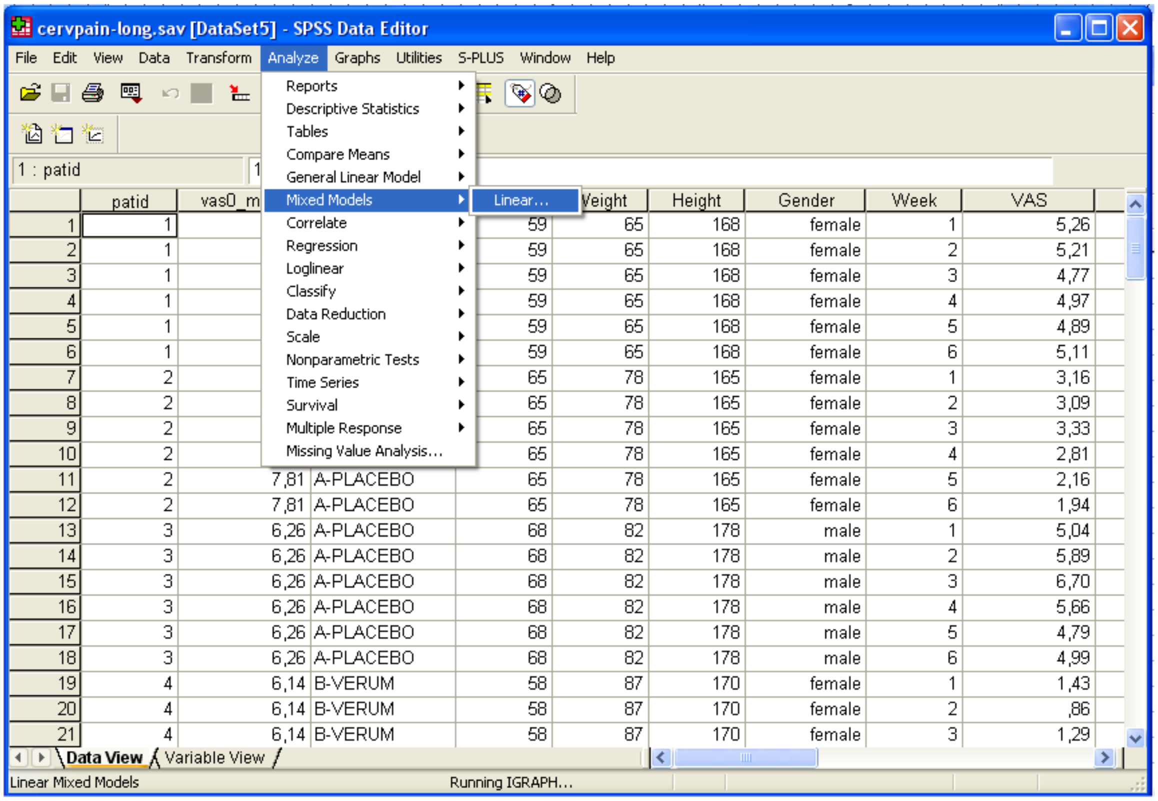
For the repeated measures-ANOVA, we need the data set in the usual long format, with the baseline measurement as a covariate (not a separate line). Open cervpain-long.sav (without the '0'):



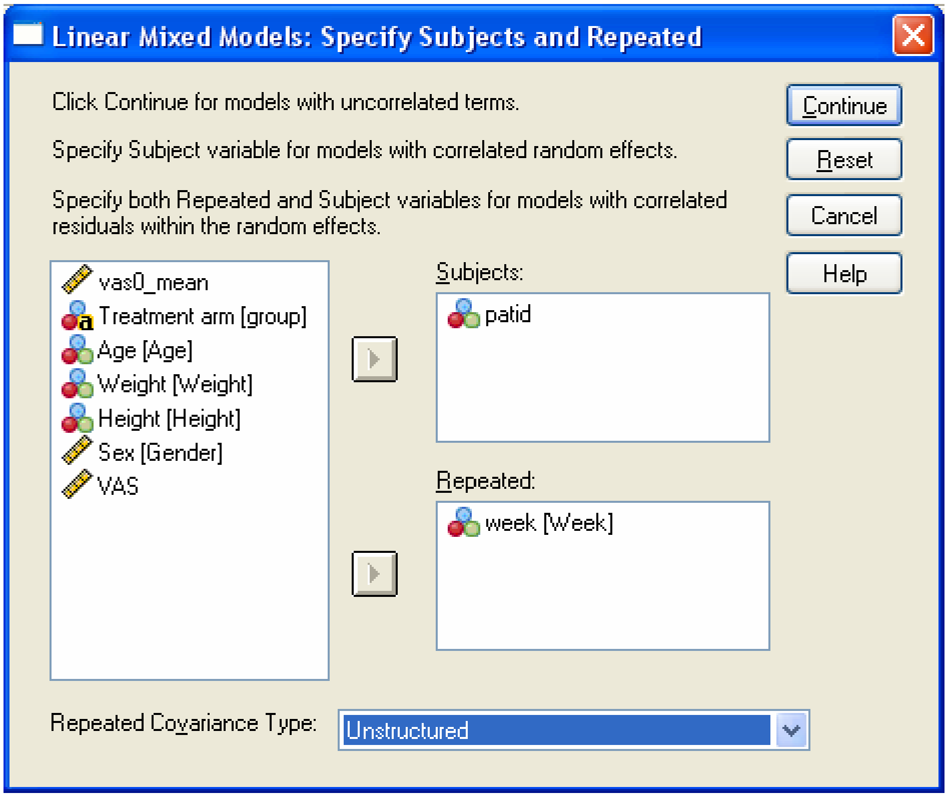
cervpain-long.sav [DataSet2] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 : patid
1
Data View
Variable View
SPSS Processor is ready

	patid	vas0_mean	group	Age	Weight	Height	Gender	Week	VAS
1	1	6,69	A-PLACEBO	59	65	168	female	1	5,26
2	1	6,69	A-PLACEBO	59	65	168	female	2	5,21
3	1	6,69	A-PLACEBO	59	65	168	female	3	4,77
4	1	6,69	A-PLACEBO	59	65	168	female	4	4,97
5	1	6,69	A-PLACEBO	59	65	168	female	5	4,89
6	1	6,69	A-PLACEBO	59	65	168	female	6	5,11
7	2	7,81	A-PLACEBO	65	78	165	female	1	3,16
8	2	7,81	A-PLACEBO	65	78	165	female	2	3,09
9	2	7,81	A-PLACEBO	65	78	165	female	3	3,33
10	2	7,81	A-PLACEBO	65	78	165	female	4	2,81
11	2	7,81	A-PLACEBO	65	78	165	female	5	2,16
12	2	7,81	A-PLACEBO	65	78	165	female	6	1,94
13	3	6,26	A-PLACEBO	68	82	178	male	1	5,04
14	3	6,26	A-PLACEBO	68	82	178	male	2	5,89
15	3	6,26	A-PLACEBO	68	82	178	male	3	6,70
16	3	6,26	A-PLACEBO	68	82	178	male	4	5,66
17	3	6,26	A-PLACEBO	68	82	178	male	5	4,79
18	3	6,26	A-PLACEBO	68	82	178	male	6	4,99
19	4	6,14	B-VERUM	58	87	170	female	1	1,43

RM-ANOVA is called by choosing the menu ‘Analyze-Mixed Models-Linear’:



	patid	vas0_m			Veight	Height	Gender	Week	VAS
1	1			59	65	168	female	1	5,26
2	1			59	65	168	female	2	5,21
3	1			59	65	168	female	3	4,77
4	1			59	65	168	female	4	4,97
5	1			59	65	168	female	5	4,89
6	1			59	65	168	female	6	5,11
7	2			65	78	165	female	1	3,16
8	2			65	78	165	female	2	3,09
9	2			65	78	165	female	3	3,33
10	2			65	78	165	female	4	2,81
11	2	7,81	A-PLACEBO	65	78	165	female	5	2,16
12	2	7,81	A-PLACEBO	65	78	165	female	6	1,94
13	3	6,26	A-PLACEBO	68	82	178	male	1	5,04
14	3	6,26	A-PLACEBO	68	82	178	male	2	5,89
15	3	6,26	A-PLACEBO	68	82	178	male	3	6,70
16	3	6,26	A-PLACEBO	68	82	178	male	4	5,66
17	3	6,26	A-PLACEBO	68	82	178	male	5	4,79
18	3	6,26	A-PLACEBO	68	82	178	male	6	4,99
19	4	6,14	B-VERUM	58	87	170	female	1	1,43
20	4	6,14	B-VERUM	58	87	170	female	2	,86
21	4	6,14	B-VERUM	58	87	170	female	3	1,29


Linear Mixed Models: Specify Subjects and Repeated

Click Continue for models with uncorrelated terms.

Specify Subject variable for models with correlated random effects.

Specify both Repeated and Subject variables for models with correlated residuals within the random effects.

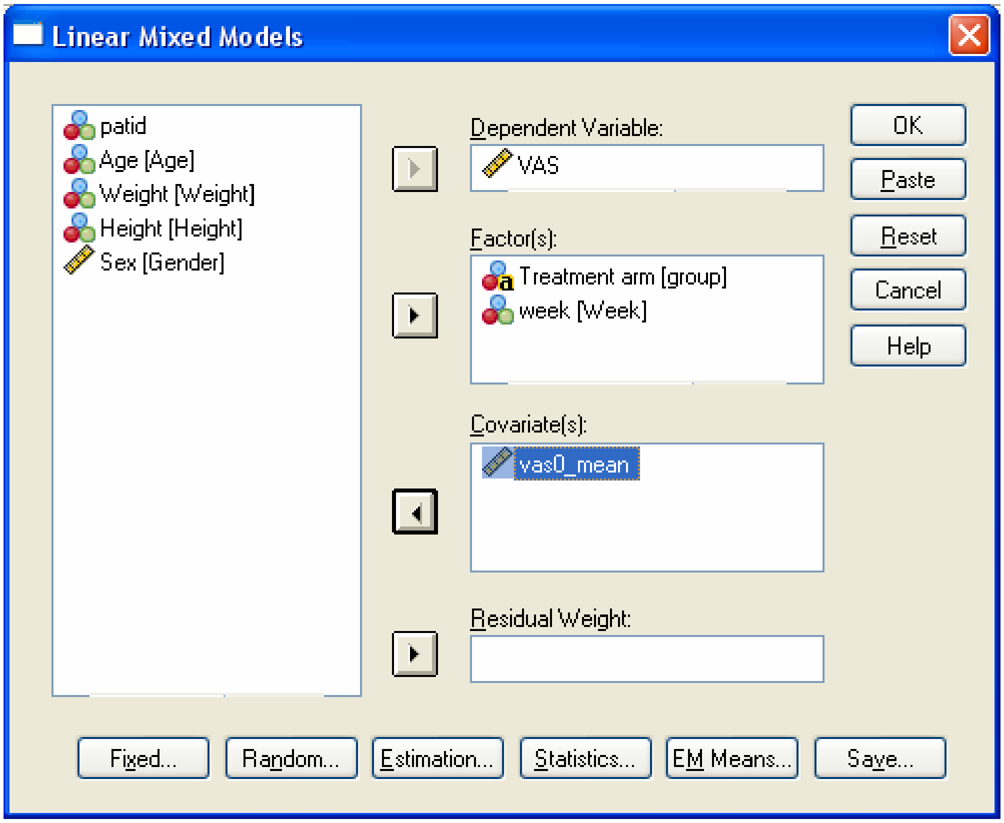
vas0_mean
Treatment arm [group]
Age [Age]
Weight [Weight]
Height [Height]
Sex [Gender]
VAS

Subjects: patid

Repeated: week [Week]

Repeated Covariance Type: Unstructured

We move patid into the field ‘Subjects’ and ‘week’ into the field ‘Repeated’. As a covariance structure, we choose ‘Unstructured’. At the next menu, we select VAS as the dependent variable, week and treatment as factors, and the baseline VAS measurement as the covariate:



patid
Age [Age]
Weight [Weight]
Height [Height]
Sex [Gender]

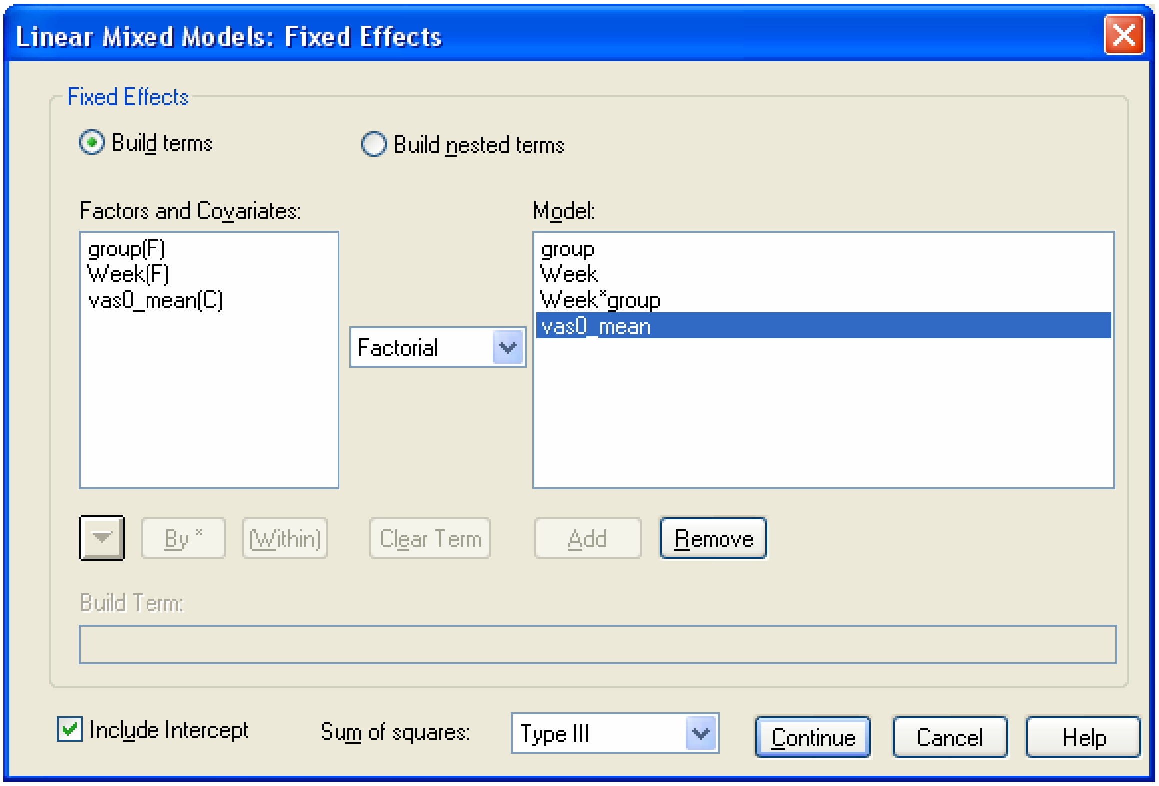
Dependent Variable: VAS

Factor(s):
Treatment arm [group]
week [Week]

Covariate(s):
vas0_mean

Residual Weight:

Next, press 'Fixed…' and select the following:



● Build terms ○ Build nested terms

Factors and Covariates:
group(F)
Week(F)
vas0_mean(C)

Factorial

Model:
group
Week
Week*group
vas0_mean

Build Term:

☑ Include Intercept

Sum of squares: Type III

We want to estimate the following effects:

Effect	Choose…	Name in our data set
Treatment	Factorial	group
Time	Factorial	week
Treatment by time	Interaction	week*group
Baseline VAS	Factorial	vas0_mean

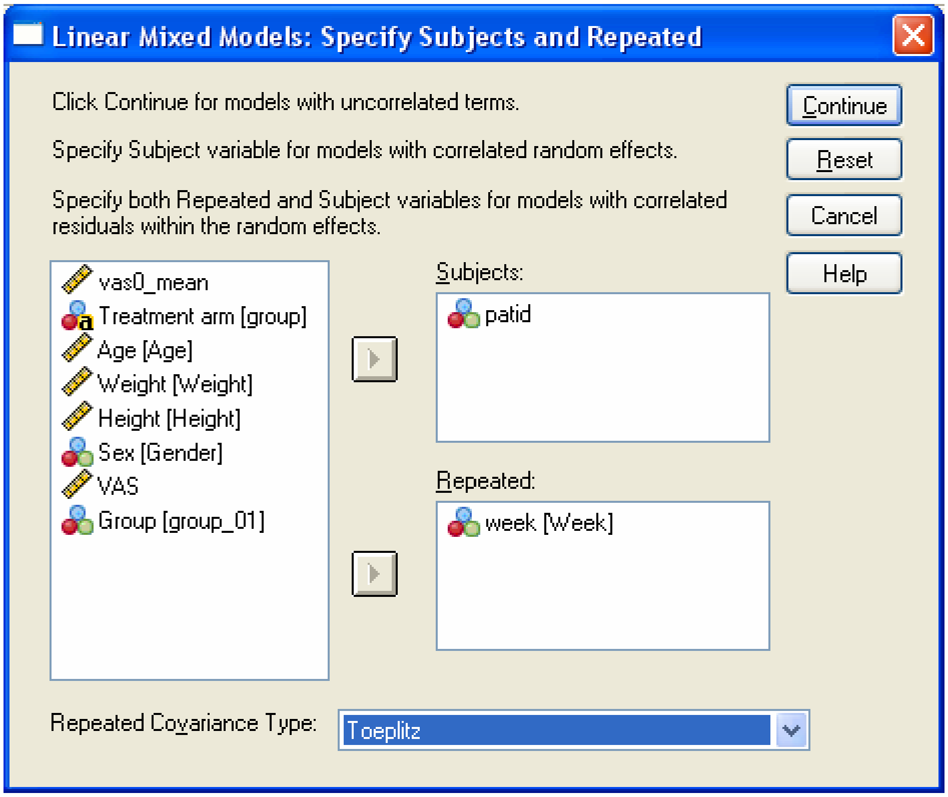
After having defined these model terms, we click on 'Continue' and OK. For the time being, we are only interested in the Akaike information criterion of the model, to judge the adequacy of the assumed covariance structure:

Information Criteria(a)

-2 Restricted Log Likelihood	167,835
Akaike's Information Criterion (AIC)	209,835
Hurvich and Tsai's Criterion (AICC)	221,103
Bozdogan's Criterion (CAIC)	286,367
Schwarz's Bayesian Criterion (BIC)	265,367

The information criteria are displayed in smaller-is-better forms.
a Dependent Variable: VAS.

The AIC is 209.835 for the unstructured covariance. We repeat the analysis, this time specifying the Toeplitz structure at the very beginning:



Click Continue for models with uncorrelated terms.

Specify Subject variable for models with correlated random effects.

Specify both Repeated and Subject variables for models with correlated residuals within the random effects.

- vas0_mean
- Treatment arm [group]
- Age [Age]
- Weight [Weight]
- Height [Height]
- Sex [Gender]
- VAS
- Group [group_01]

Subjects: patid

Repeated: week [Week]

Repeated Covariance Type: Toeplitz

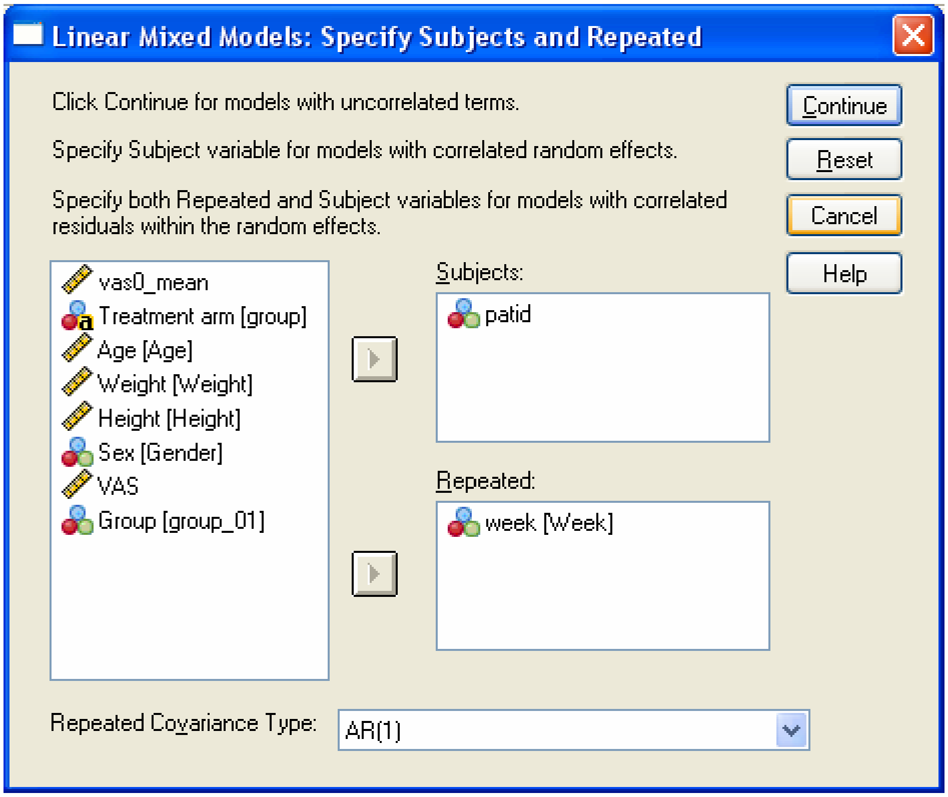
We obtain the following table of information criteria:

Information Criteria(a)

-2 Restricted Log Likelihood	200,511
Akaike's Information Criterion (AIC)	212,511
Hurvich and Tsai's Criterion (AICC)	213,377
Bozdogan's Criterion (CAIC)	234,377
Schwarz's Bayesian Criterion (BIC)	228,377

The information criteria are displayed in smaller-is-better forms.
a Dependent Variable: VAS.

Another turn, using the AR(1) structure:



Click Continue for models with uncorrelated terms.

Specify Subject variable for models with correlated random effects.

Specify both Repeated and Subject variables for models with correlated residuals within the random effects.

Subjects: patid

Repeated: week [Week]

Repeated Covariance Type: AR(1)

Information Criteria(a)

-2 Restricted Log Likelihood	203,466
Akaike's Information Criterion (AIC)	207,466
Hurvich and Tsai's Criterion (AICC)	207,585
Bozdogan's Criterion (CAIC)	214,755
Schwarz's Bayesian Criterion (BIC)	212,755

The information criteria are displayed in smaller-is-better forms.
a Dependent Variable: VAS.

This structure yields the smallest AIC. Therefore, we continue with the AR(1) structure.

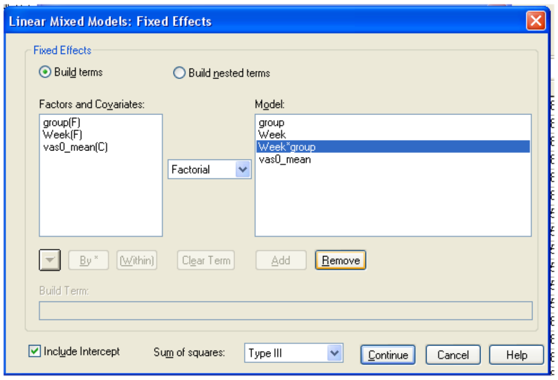
Next, we decide whether to keep or to drop the interaction of treatment and time (group * week in the table below).

Type III Tests of Fixed Effects(a)

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	17,648	,200	,660
group	1	18,262	4,131	,057
Week	5	85,740	2,248	,057
vas0_mean	1	17,655	,024	,879
group * Week	5	85,740	,373	,866

a Dependent Variable: VAS.

We learn that the interaction is not significant and it can therefore be dropped from the model. We recall the menu ‘Analyze-Mixed models-Linear’ with the same specifications, but remove the interaction in the ‘Fixed effects’ submenu (select the interaction and press ‘Remove’):


Linear Mixed Models: Fixed Effects

Fixed Effects

● Build terms ○ Build nested terms

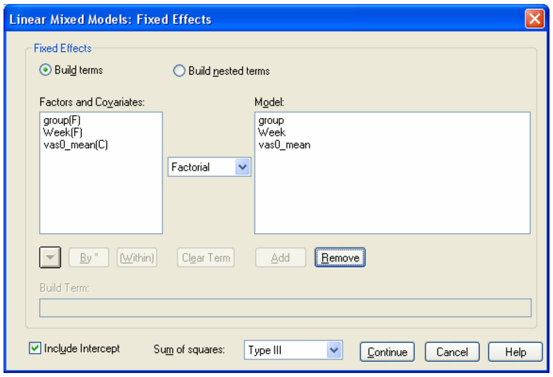
Factors and Covariates: group(F) Week(F) vas0_mean(C)

Model: group Week Week*group vas0_mean

Factorial

☑ Include Intercept

Sum of squares: Type III


Linear Mixed Models: Fixed Effects

Fixed Effects

● Build terms ○ Build nested terms

Factors and Covariates: group(F) Week(F) vas0_mean(C)

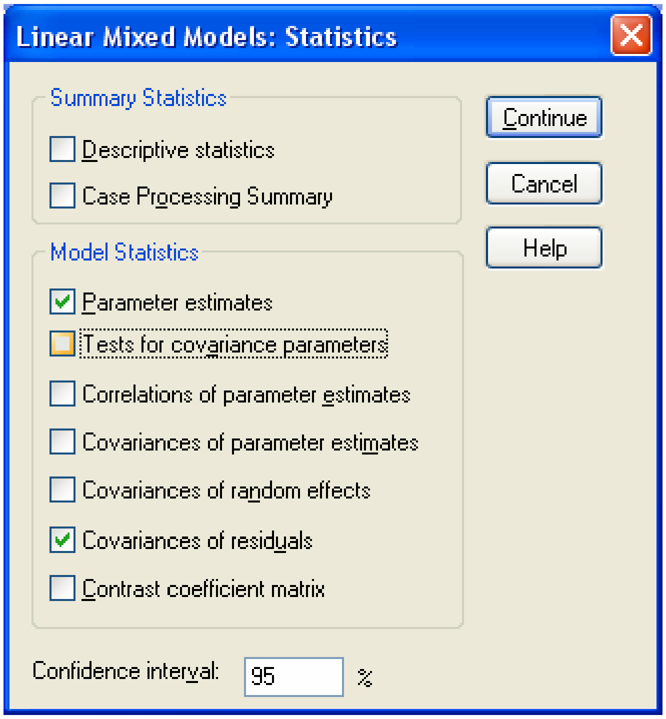
Model: group Week vas0_mean

Factorial

☑ Include Intercept

Sum of squares: Type III

Now we proceed by clicking ‘Continue’. Next, we call the ‘Estimates’ submenu and select the following:



Linear Mixed Models: Statistics

Summary Statistics

- ☐ Descriptive statistics
- ☐ Case Processing Summary

Model Statistics

- ☑ Parameter estimates
- ☐ Tests for covariance parameters
- ☐ Correlations of parameter estimates
- ☐ Covariances of parameter estimates
- ☐ Covariances of random effects
- ☑ Covariances of residuals
- ☐ Contrast coefficient matrix

Confidence interval: 95 %

In the resulting output, the most important table is now the one containing the parameter estimates:

Estimates of Fixed Effects[b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1,993854	6,780319	17,685	,294	,772	-12,269273	16,256981
[group=A-PLACEBO]	1,593299	,805477	17,841	1,978	,064	-,100024	3,286623
[group=B-VERUM]	0[a]	0	.	.	.	.	.
[Week=1]	,575499	,213562	99,717	2,695	,008	,151783	,999215
[Week=2]	,348734	,193197	97,853	1,805	,074	-,034666	,732134
[Week=3]	,312273	,168507	95,624	1,853	,067	-,022227	,646774
[Week=4]	,100772	,138565	93,290	,727	,469	-,174379	,375923
[Week=5]	,012948	,098677	90,874	,131	,896	-,183065	,208960
[Week=6]	0[a]	0	.	.	.	.	.
vas0_mean	,142963	,929271	17,685	,154	,879	-1,811858	2,097784

a. This parameter is set to zero because it is redundant.

b. Dependent Variable: VAS.

We see that the treatment effect is 1.58 (p=0.064). This means that on average, the placebo treated group has VAS scores which are about 1.59 units higher than those of the electro-stimulated acupuncture group. Since we could not find a significant interaction of treatment effect and time, we may assume that the treatment effect is constant over the whole range of follow-up (6 weeks).

From the estimates referring to the time effect, we see that both treatments lead to a reduction of VAS values between week 1 and the subsequent weeks. The parametrization of the time effect is such that week 6 is the reference category. All preceding weeks are compared to this week. Since the estimate is 0.575 for the first week, the VAS scores are 0.575 cm higher in the first week compared to week 6 (in both groups, since there is no time-by-treatment interaction). In week 2, the scores are 0.348 cm higher. After week 3, the VAS scores do not change much anymore as can be seen from the very low estimates (0.10, 0.01 for the difference between weeks 4 and 5 both compared to week 6).

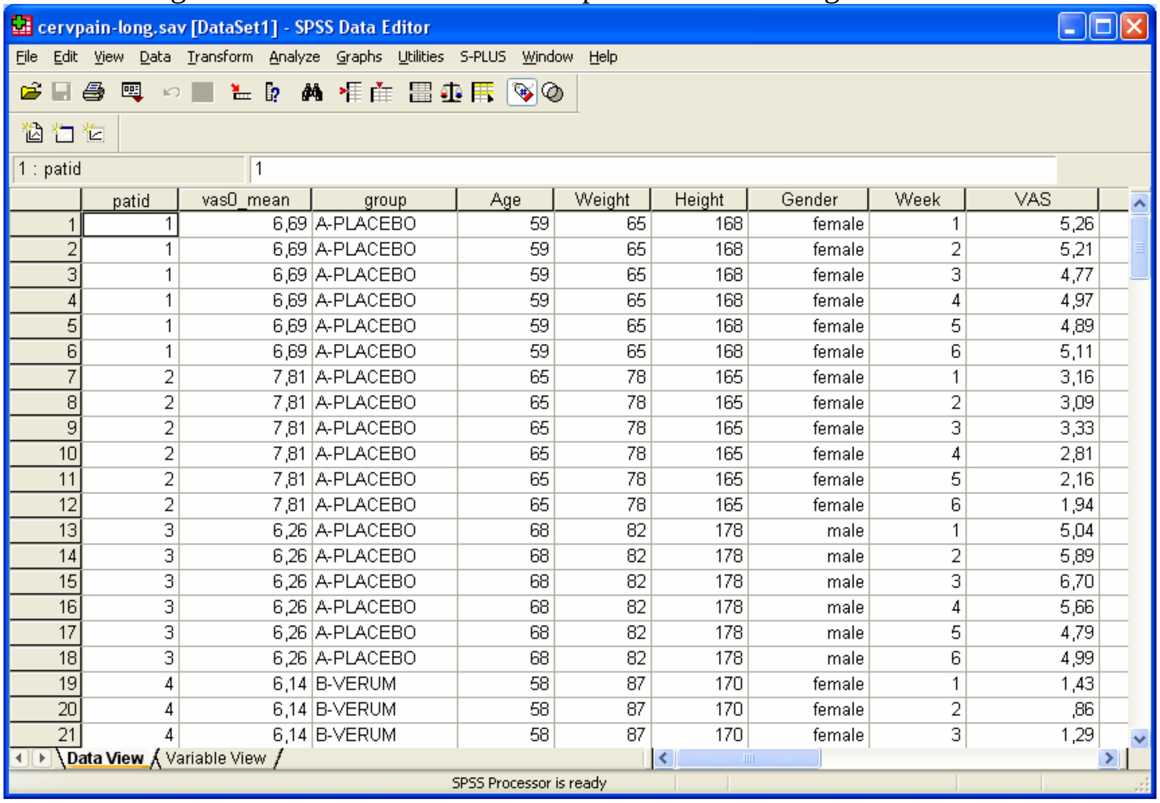
The baseline VAS has no effect on later VAS measurements (p=0.879). This could be a results from the very small range of baseline VAS measurements (there was not much difference in the baseline VAS measurements between the patients).

Restructuring a longitdudinal data set

A longitudinal data set, i. e., a data set involving repeated measurements on the same subjects, can be represented in two formats:

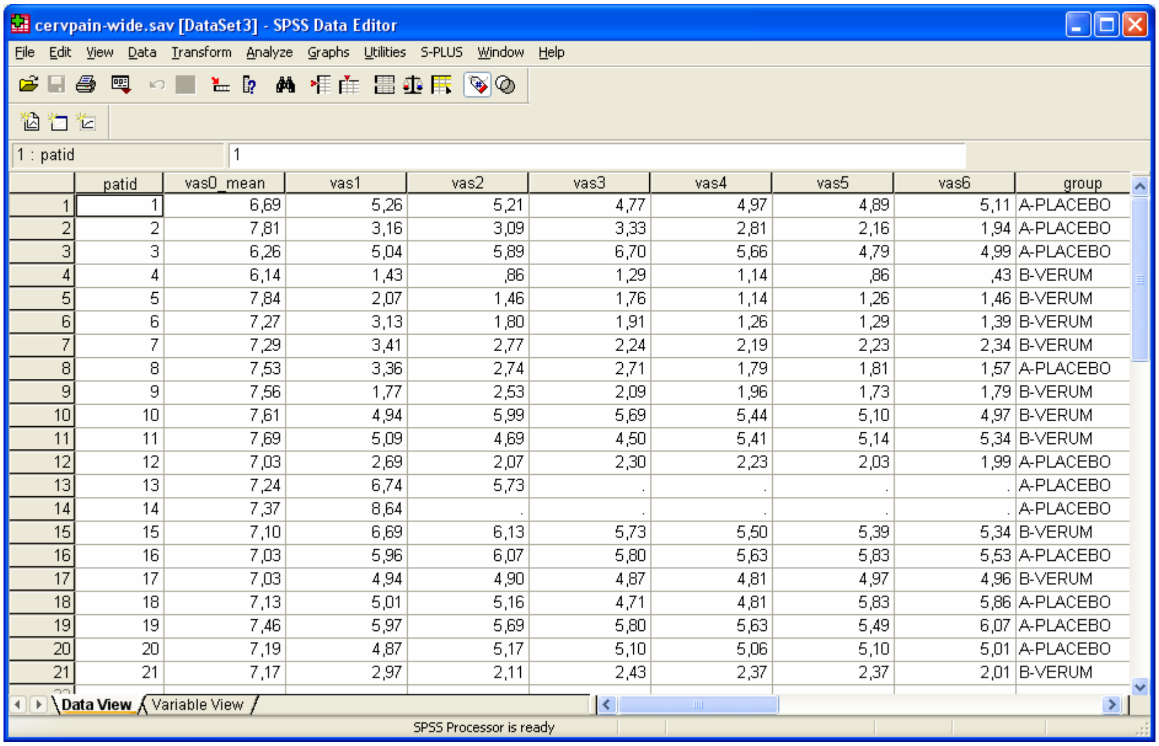
- The ‘long’ format: each row of data corresponds to one time point at which measurements are taken. Each subject is represented by multiple rows.
- The ‘wide’ format: each row of data corresponds to one subject. Each of several serial measurements is represented by multiple columns.

The following screenshots show the cervical pain data set in long …


cervpain-long.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 : patid
1
Data View
Variable View
SPSS Processor is ready

	patid	vas0_mean	group	Age	Weight	Height	Gender	Week	VAS
1	1	6,69	A-PLACEBO	59	65	168	female	1	5,26
2	1	6,69	A-PLACEBO	59	65	168	female	2	5,21
3	1	6,69	A-PLACEBO	59	65	168	female	3	4,77
4	1	6,69	A-PLACEBO	59	65	168	female	4	4,97
5	1	6,69	A-PLACEBO	59	65	168	female	5	4,89
6	1	6,69	A-PLACEBO	59	65	168	female	6	5,11
7	2	7,81	A-PLACEBO	65	78	165	female	1	3,16
8	2	7,81	A-PLACEBO	65	78	165	female	2	3,09
9	2	7,81	A-PLACEBO	65	78	165	female	3	3,33
10	2	7,81	A-PLACEBO	65	78	165	female	4	2,81
11	2	7,81	A-PLACEBO	65	78	165	female	5	2,16
12	2	7,81	A-PLACEBO	65	78	165	female	6	1,94
13	3	6,26	A-PLACEBO	68	82	178	male	1	5,04
14	3	6,26	A-PLACEBO	68	82	178	male	2	5,89
15	3	6,26	A-PLACEBO	68	82	178	male	3	6,70
16	3	6,26	A-PLACEBO	68	82	178	male	4	5,66
17	3	6,26	A-PLACEBO	68	82	178	male	5	4,79
18	3	6,26	A-PLACEBO	68	82	178	male	6	4,99
19	4	6,14	B-VERUM	58	87	170	female	1	1,43
20	4	6,14	B-VERUM	58	87	170	female	2	,86
21	4	6,14	B-VERUM	58	87	170	female	3	1,29

… and wide format:

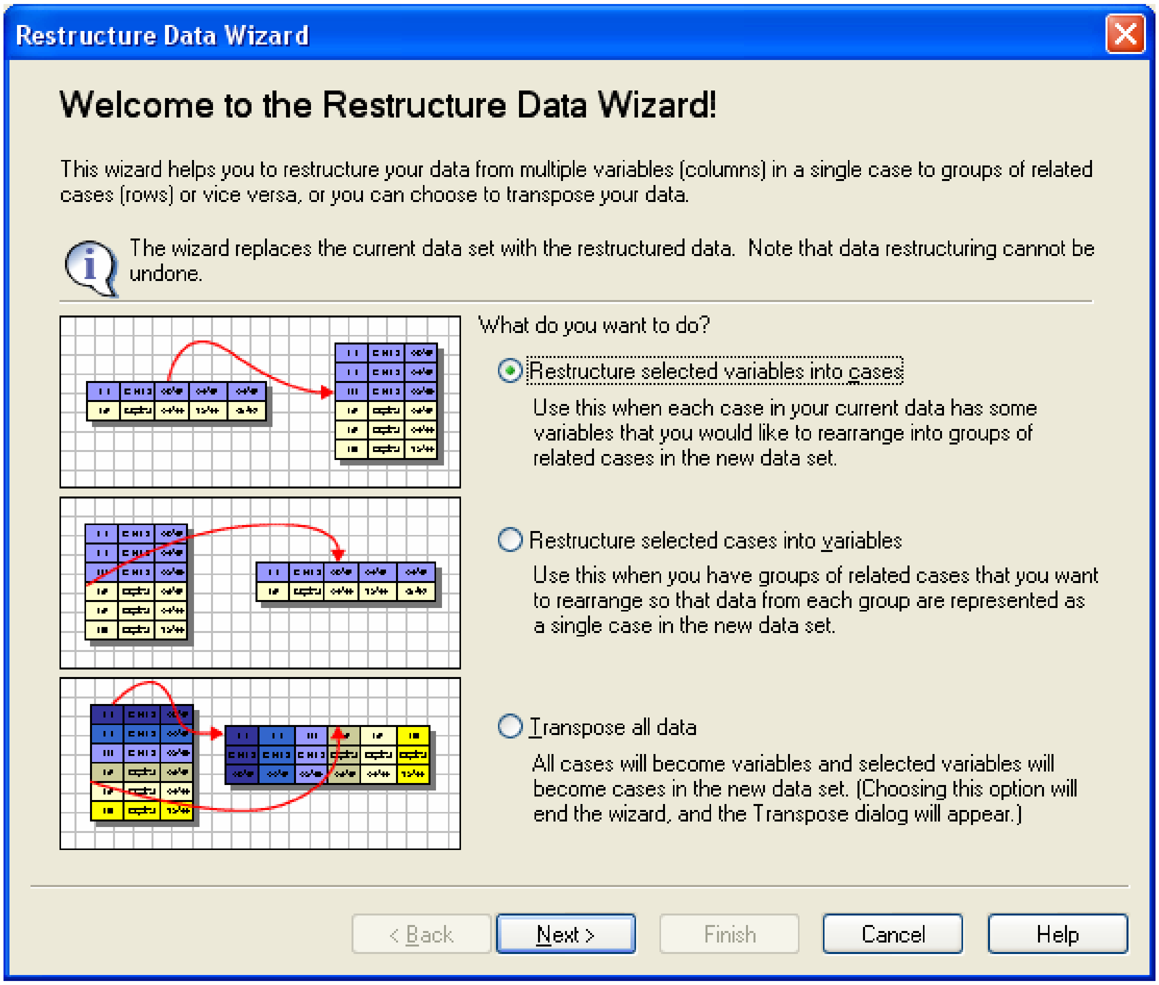

cervpain-wide.sav [DataSet3] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities S-PLUS Window Help
1 : patid
1
Data View
Variable View
SPSS Processor is ready

	patid	vas0_mean	vas1	vas2	vas3	vas4	vas5	vas6	group
1	1	6,69	5,26	5,21	4,77	4,97	4,89	5,11	A-PLACEBO
2	2	7,81	3,16	3,09	3,33	2,81	2,16	1,94	A-PLACEBO
3	3	6,26	5,04	5,89	6,70	5,66	4,79	4,99	A-PLACEBO
4	4	6,14	1,43	,86	1,29	1,14	,86	,43	B-VERUM
5	5	7,84	2,07	1,46	1,76	1,14	1,26	1,46	B-VERUM
6	6	7,27	3,13	1,80	1,91	1,26	1,29	1,39	B-VERUM
7	7	7,29	3,41	2,77	2,24	2,19	2,23	2,34	B-VERUM
8	8	7,53	3,36	2,74	2,71	1,79	1,81	1,57	A-PLACEBO
9	9	7,56	1,77	2,53	2,09	1,96	1,73	1,79	B-VERUM
10	10	7,61	4,94	5,99	5,69	5,44	5,10	4,97	B-VERUM
11	11	7,69	5,09	4,69	4,50	5,41	5,14	5,34	B-VERUM
12	12	7,03	2,69	2,07	2,30	2,23	2,03	1,99	A-PLACEBO
13	13	7,24	6,74	5,73	.	.	.	.	A-PLACEBO
14	14	7,37	8,64	.	.	.	.	.	A-PLACEBO
15	15	7,10	6,69	6,13	5,73	5,50	5,39	5,34	B-VERUM
16	16	7,03	5,96	6,07	5,80	5,63	5,83	5,53	A-PLACEBO
17	17	7,03	4,94	4,90	4,87	4,81	4,97	4,96	B-VERUM
18	18	7,13	5,01	5,16	4,71	4,81	5,83	5,86	A-PLACEBO
19	19	7,46	5,97	5,69	5,80	5,63	5,49	6,07	A-PLACEBO
20	20	7,19	4,87	5,17	5,10	5,06	5,10	5,01	A-PLACEBO
21	21	7,17	2,97	2,11	2,43	2,37	2,37	2,01	B-VERUM

With SPSS, SAS and other statistics programs, it is possible to switch between these two formats. We exemplify the format switching on the cervical pain data set.

Switching from wide to long

We start with the data set cervpain-wide.sav as depicted above. From the menu, select Data-Restructure:



Restructure Data Wizard

Welcome to the Restructure Data Wizard!

This wizard helps you to restructure your data from multiple variables (columns) in a single case to groups of related cases (rows) or vice versa, or you can choose to transpose your data.

The wizard replaces the current data set with the restructured data. Note that data restructuring cannot be undone.

What do you want to do?

● Restructure selected variables into cases

Use this when each case in your current data has some variables that you would like to rearrange into groups of related cases in the new data set.

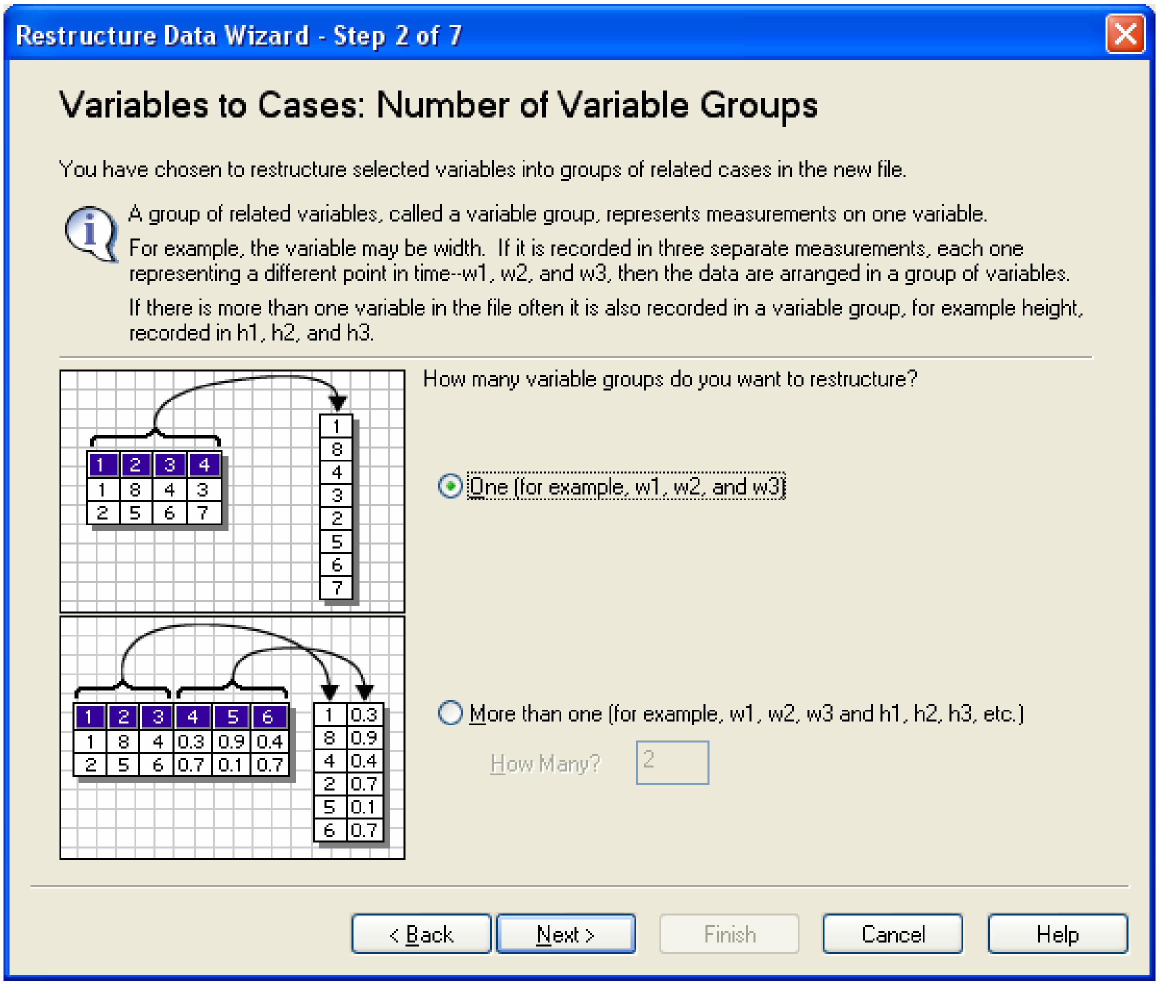
○ Restructure selected cases into variables

Use this when you have groups of related cases that you want to rearrange so that data from each group are represented as a single case in the new data set.

○ Transpose all data

All cases will become variables and selected variables will become cases in the new data set. (Choosing this option will end the wizard, and the Transpose dialog will appear.)

Choose ‘Restructure selected variables into cases’, press ‘Next >’:



Variables to Cases: Number of Variable Groups

You have chosen to restructure selected variables into groups of related cases in the new file.

A group of related variables, called a variable group, represents measurements on one variable.

For example, the variable may be width. If it is recorded in three separate measurements, each one representing a different point in time--w1, w2, and w3, then the data are arranged in a group of variables.

If there is more than one variable in the file often it is also recorded in a variable group, for example height, recorded in h1, h2, and h3.

How many variable groups do you want to restructure?

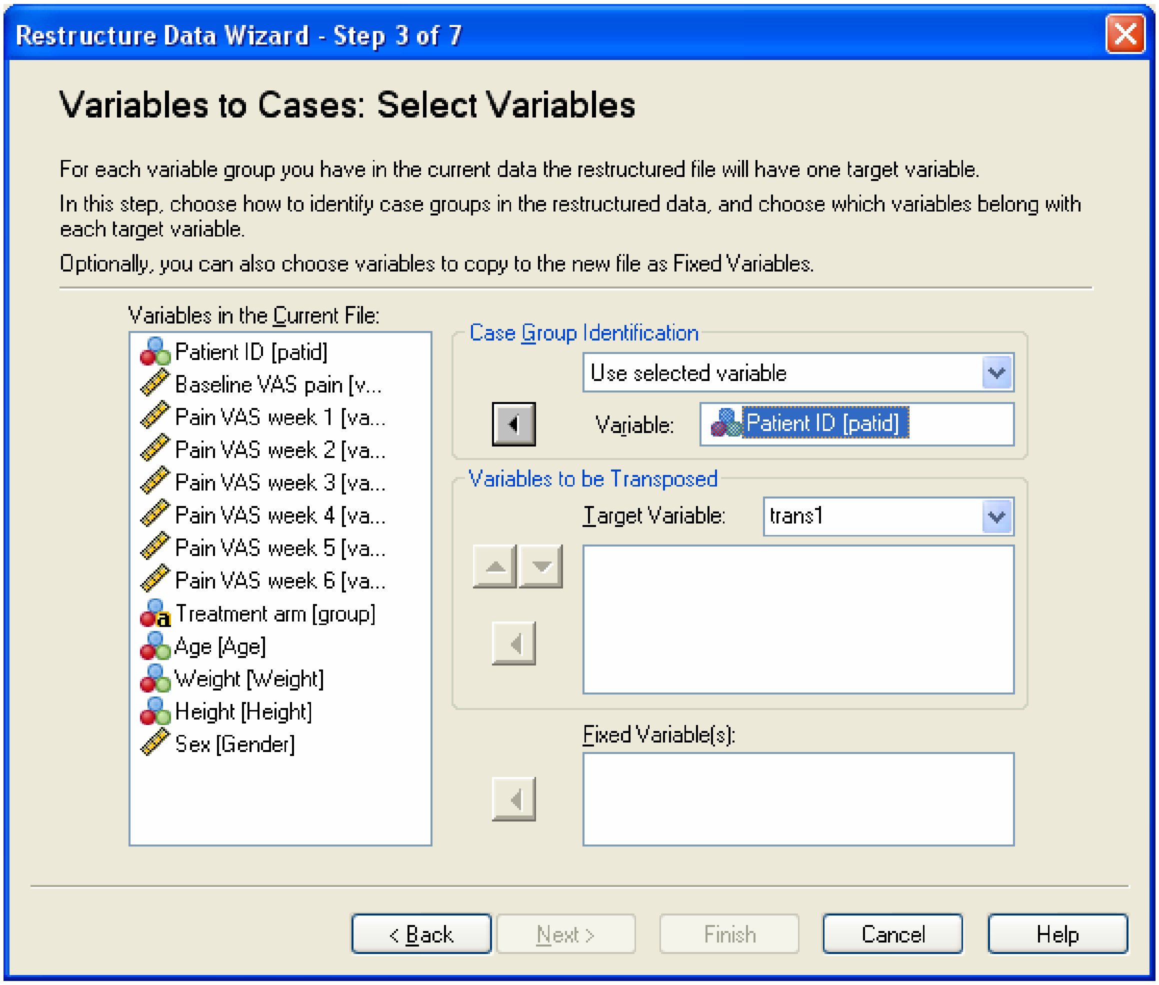
● One (for example, w1, w2, and w3)

○ More than one (for example, w1, w2, w3 and h1, h2, h3, etc.)

How Many? 2

Now the dialogue asks us whether we want to restructure one or several variable groups. In our case, we have only one variable with repeated measurements, but in general, one will have more than one such variable.

Now we have to define the subject identifier. This is done by changing ‘Case group identification’ to ‘Use selected variable’, and moving ‘Patient ID’ into the field ‘Variable’:



Restructure Data Wizard - Step 3 of 7

Variables to Cases: Select Variables

For each variable group you have in the current data the restructured file will have one target variable.

In this step, choose how to identify case groups in the restructured data, and choose which variables belong with each target variable.

Optionally, you can also choose variables to copy to the new file as Fixed Variables.

Variables in the Current File:

- Patient ID [patid]
- Baseline VAS pain [v...
- Pain VAS week 1 [va...
- Pain VAS week 2 [va...
- Pain VAS week 3 [va...
- Pain VAS week 4 [va...
- Pain VAS week 5 [va...
- Pain VAS week 6 [va...
- Treatment arm [group]
- Age [Age]
- Weight [Weight]
- Height [Height]
- Sex [Gender]

Case Group Identification

Use selected variable

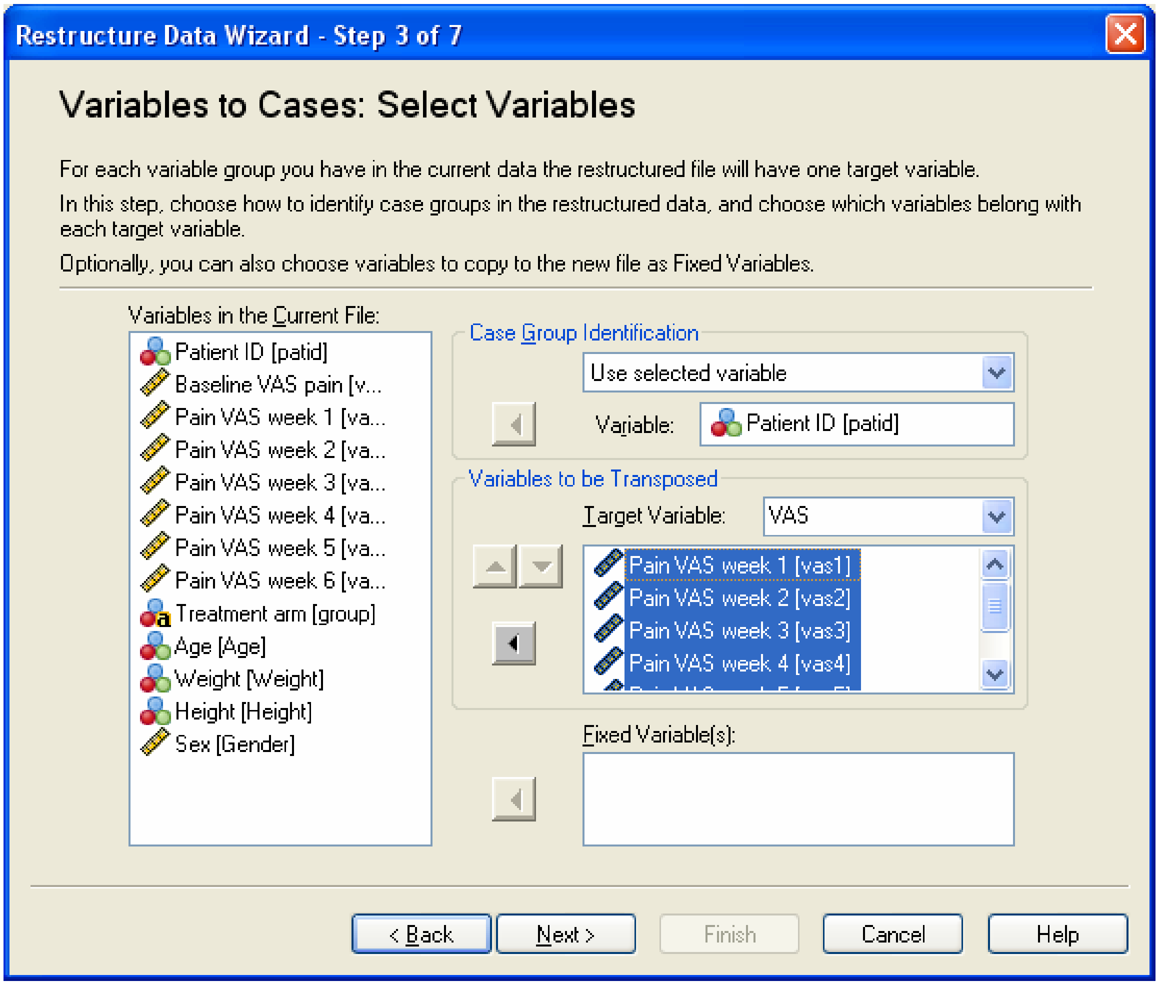
Variable: Patient ID [patid]

Variables to be Transposed

Target Variable: trans1

Fixed Variable(s):

Then we have to define the columns which correspond to the repeatedly measured variable. We change ‘Target Variable’ to ‘VAS’ (directly writing into the field), and move the 6 variables labeled ‘Pain VAS week 1’ to ‘Pain VAS week 6’ into the field ‘Variables to be Transposed’:



Restructure Data Wizard - Step 3 of 7

Variables to Cases: Select Variables

For each variable group you have in the current data the restructured file will have one target variable.

In this step, choose how to identify case groups in the restructured data, and choose which variables belong with each target variable.

Optionally, you can also choose variables to copy to the new file as Fixed Variables.

Variables in the Current File:

- Patient ID [patid]
- Baseline VAS pain [v...
- Pain VAS week 1 [va...
- Pain VAS week 2 [va...
- Pain VAS week 3 [va...
- Pain VAS week 4 [va...
- Pain VAS week 5 [va...
- Pain VAS week 6 [va...
- Treatment arm [group]
- Age [Age]
- Weight [Weight]
- Height [Height]
- Sex [Gender]

Case Group Identification

Use selected variable

Variable: Patient ID [patid]

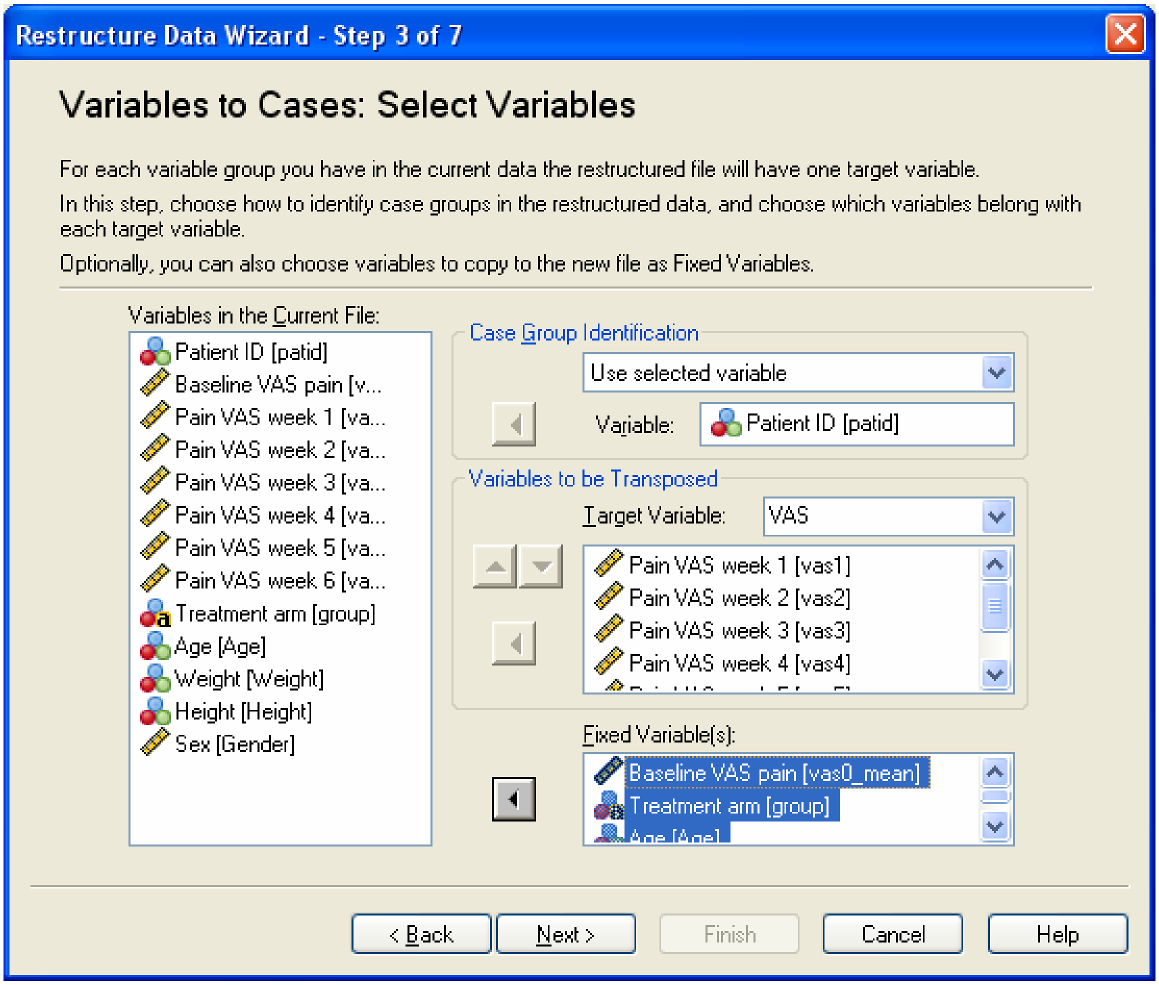
Variables to be Transposed

Target Variable: VAS

- Pain VAS week 1 [vas1]
- Pain VAS week 2 [vas2]
- Pain VAS week 3 [vas3]
- Pain VAS week 4 [vas4]

Fixed Variable(s):

Please take care of the correct sequence of these variables. All other variables, which constitute the baseline characteristics, are moved to the field ‘Fixed Variables’:



Restructure Data Wizard - Step 3 of 7

Variables to Cases: Select Variables

For each variable group you have in the current data the restructured file will have one target variable.

In this step, choose how to identify case groups in the restructured data, and choose which variables belong with each target variable.

Optionally, you can also choose variables to copy to the new file as Fixed Variables.

Variables in the Current File:

- Patient ID [patid]
- Baseline VAS pain [v...
- Pain VAS week 1 [va...
- Pain VAS week 2 [va...
- Pain VAS week 3 [va...
- Pain VAS week 4 [va...
- Pain VAS week 5 [va...
- Pain VAS week 6 [va...
- Treatment arm [group]
- Age [Age]
- Weight [Weight]
- Height [Height]
- Sex [Gender]

Case Group Identification

Use selected variable

Variable: Patient ID [patid]

Variables to be Transposed

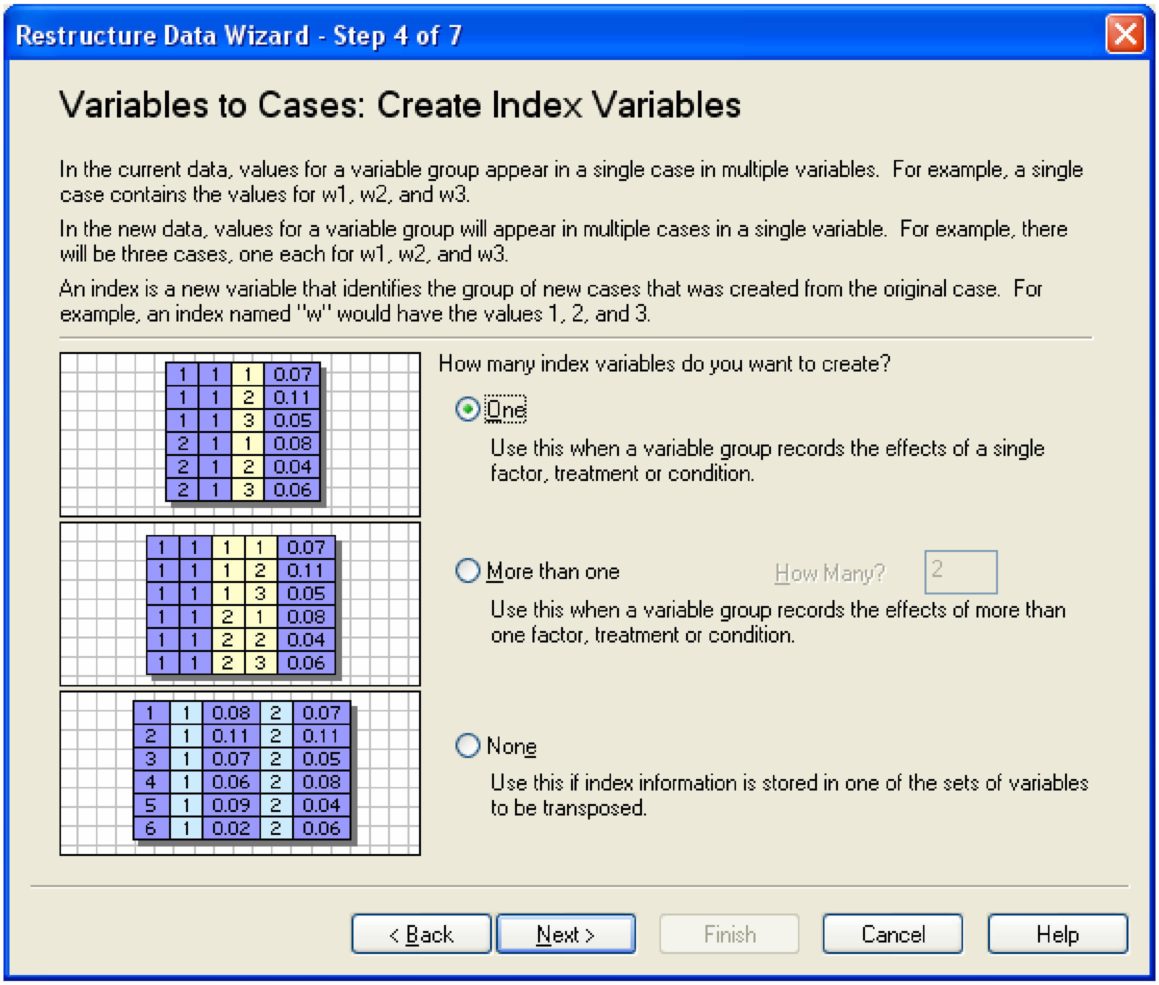
Target Variable: VAS

- Pain VAS week 1 [vas1]
- Pain VAS week 2 [vas2]
- Pain VAS week 3 [vas3]
- Pain VAS week 4 [vas4]

Fixed Variable(s):

- Baseline VAS pain [vas0_mean]
- Treatment arm [group]
- Age [Age]

Then press ‘Next >’. The program now asks us to define an index variable. This variable is later used to define the time points of the serial measurements. Therefore, we could name it ‘week’. We only need one index variable:



Variables to Cases: Create Index Variables

In the current data, values for a variable group appear in a single case in multiple variables. For example, a single case contains the values for w1, w2, and w3.

In the new data, values for a variable group will appear in multiple cases in a single variable. For example, there will be three cases, one each for w1, w2, and w3.

An index is a new variable that identifies the group of new cases that was created from the original case. For example, an index named "w" would have the values 1, 2, and 3.

1	1	1	0.07
1	1	2	0.11
1	1	3	0.05
2	1	1	0.08
2	1	2	0.04
2	1	3	0.06

1	1	1	1	0.07
1	1	1	2	0.11
1	1	1	3	0.05
1	1	2	1	0.08
1	1	2	2	0.04
1	1	2	3	0.06

1	1	0.08	2	0.07
2	1	0.11	2	0.11
3	1	0.07	2	0.05
4	1	0.06	2	0.08
5	1	0.09	2	0.04
6	1	0.02	2	0.06

How many index variables do you want to create?

● One

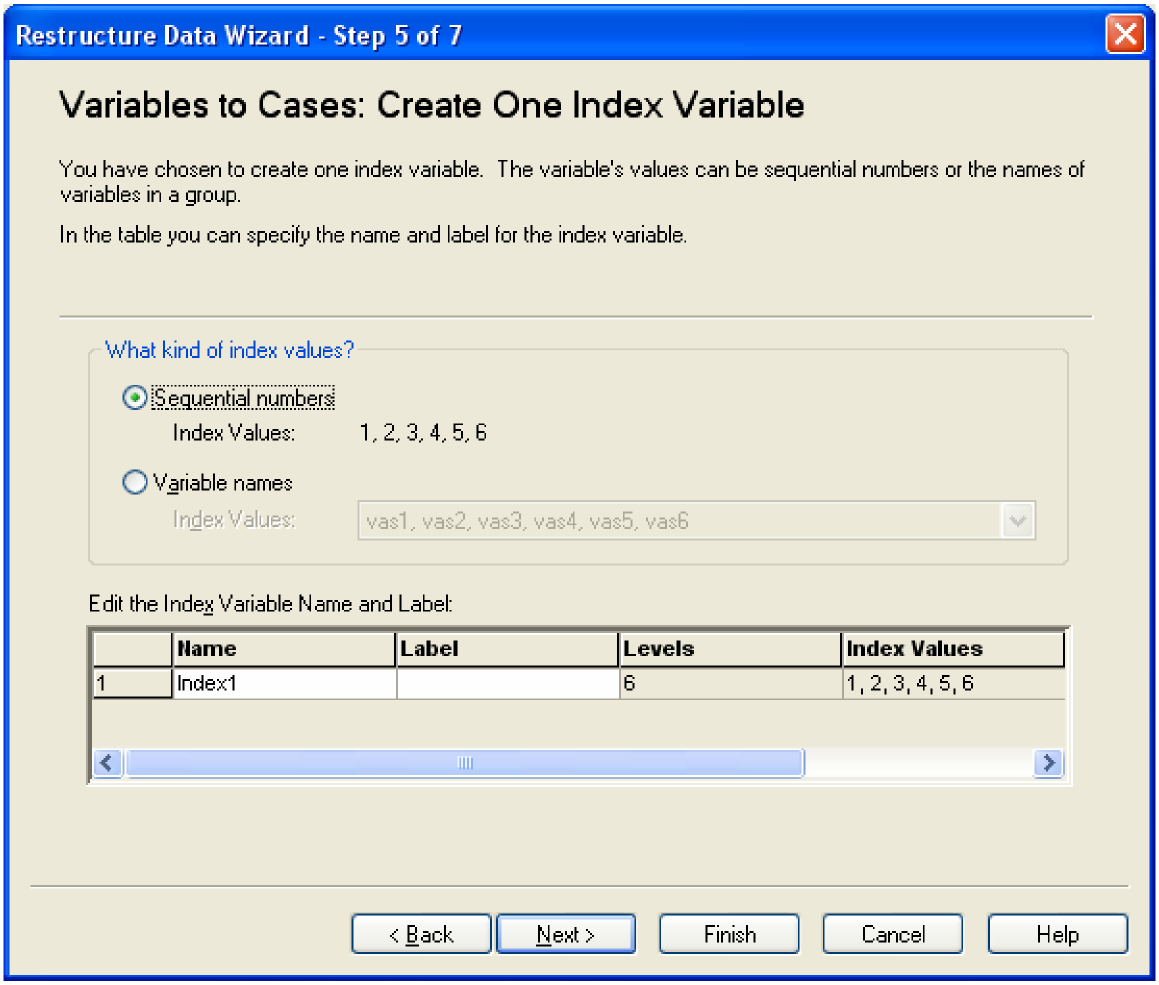
Use this when a variable group records the effects of a single factor, treatment or condition.

○ More than one   How Many? 2

Use this when a variable group records the effects of more than one factor, treatment or condition.

○ None

Use this if index information is stored in one of the sets of variables to be transposed.



Restructure Data Wizard - Step 5 of 7

Variables to Cases: Create One Index Variable

You have chosen to create one index variable. The variable's values can be sequential numbers or the names of variables in a group.

In the table you can specify the name and label for the index variable.

What kind of index values?

● Sequential numbers

Index Values: 1, 2, 3, 4, 5, 6

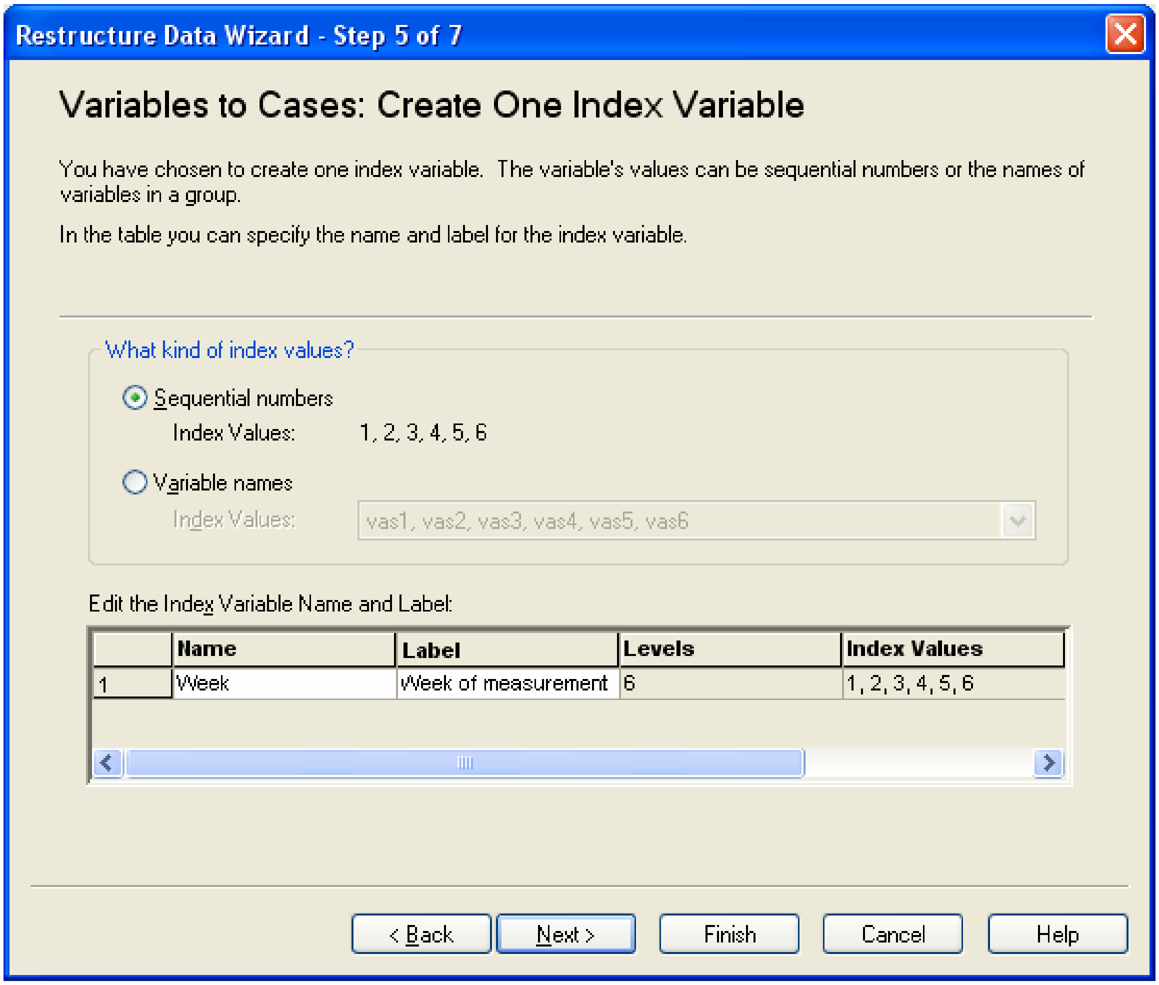
○ Variable names

Index Values: vas1, vas2, vas3, vas4, vas5, vas6

Edit the Index Variable Name and Label:

	Name	Label	Levels	Index Values
1	Index1		6	1, 2, 3, 4, 5, 6

We request sequential numbers, and change the name to ‘week’:



Restructure Data Wizard - Step 5 of 7

Variables to Cases: Create One Index Variable

You have chosen to create one index variable. The variable's values can be sequential numbers or the names of variables in a group.

In the table you can specify the name and label for the index variable.

What kind of index values?

● Sequential numbers

Index Values: 1, 2, 3, 4, 5, 6

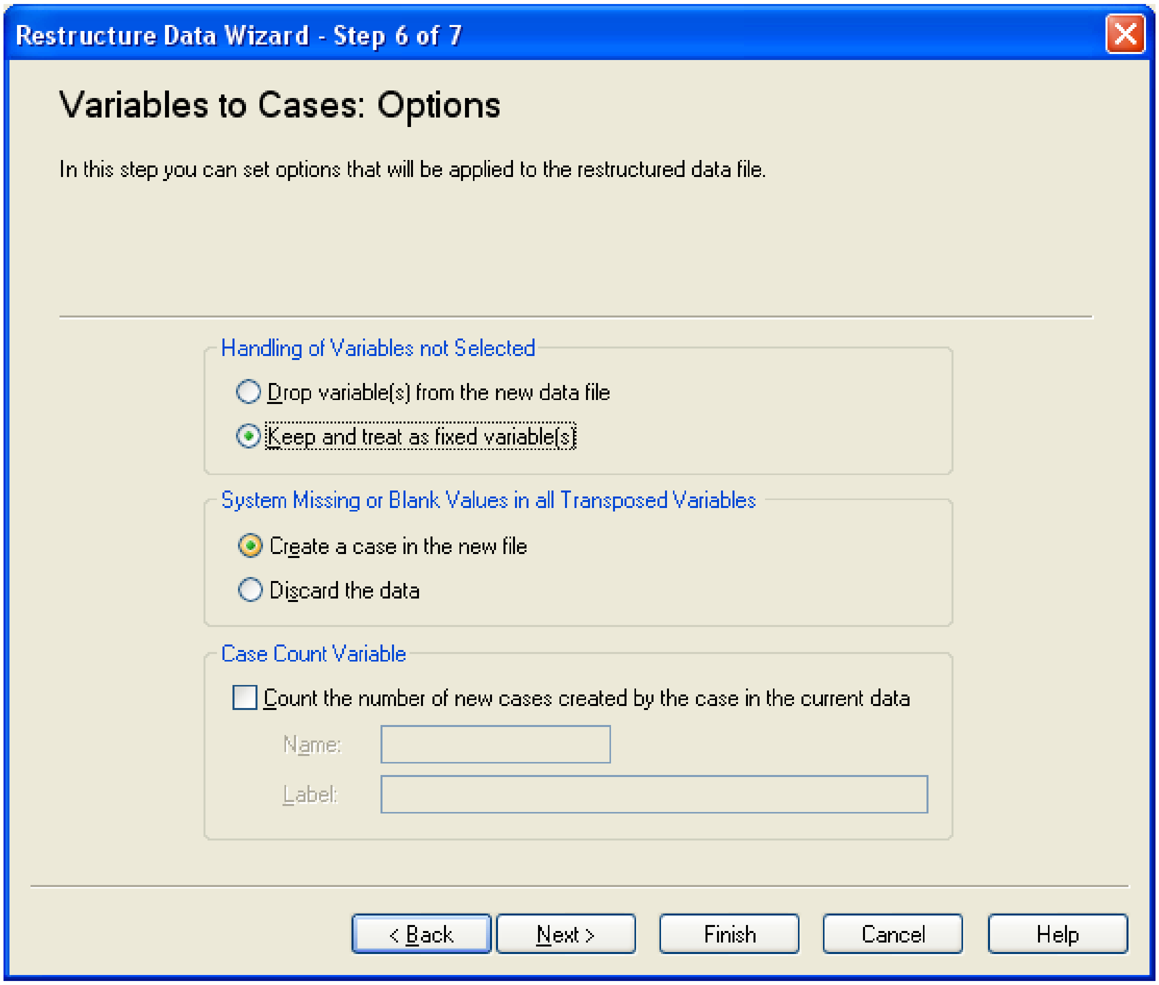
○ Variable names

Index Values: vas1, vas2, vas3, vas4, vas5, vas6

Edit the Index Variable Name and Label:

	Name	Label	Levels	Index Values
1	Week	Week of measurement	6	1, 2, 3, 4, 5, 6

In the options dialogue that appears subsequently, we request to keep any variables that were not selected before as ‘fixed’ variables, and also to keep rows with missing entries.



Restructure Data Wizard - Step 6 of 7

Variables to Cases: Options

In this step you can set options that will be applied to the restructured data file.

Handling of Variables not Selected

○ Drop variable(s) from the new data file

● Keep and treat as fixed variable(s)

System Missing or Blank Values in all Transposed Variables

● Create a case in the new file

○ Discard the data

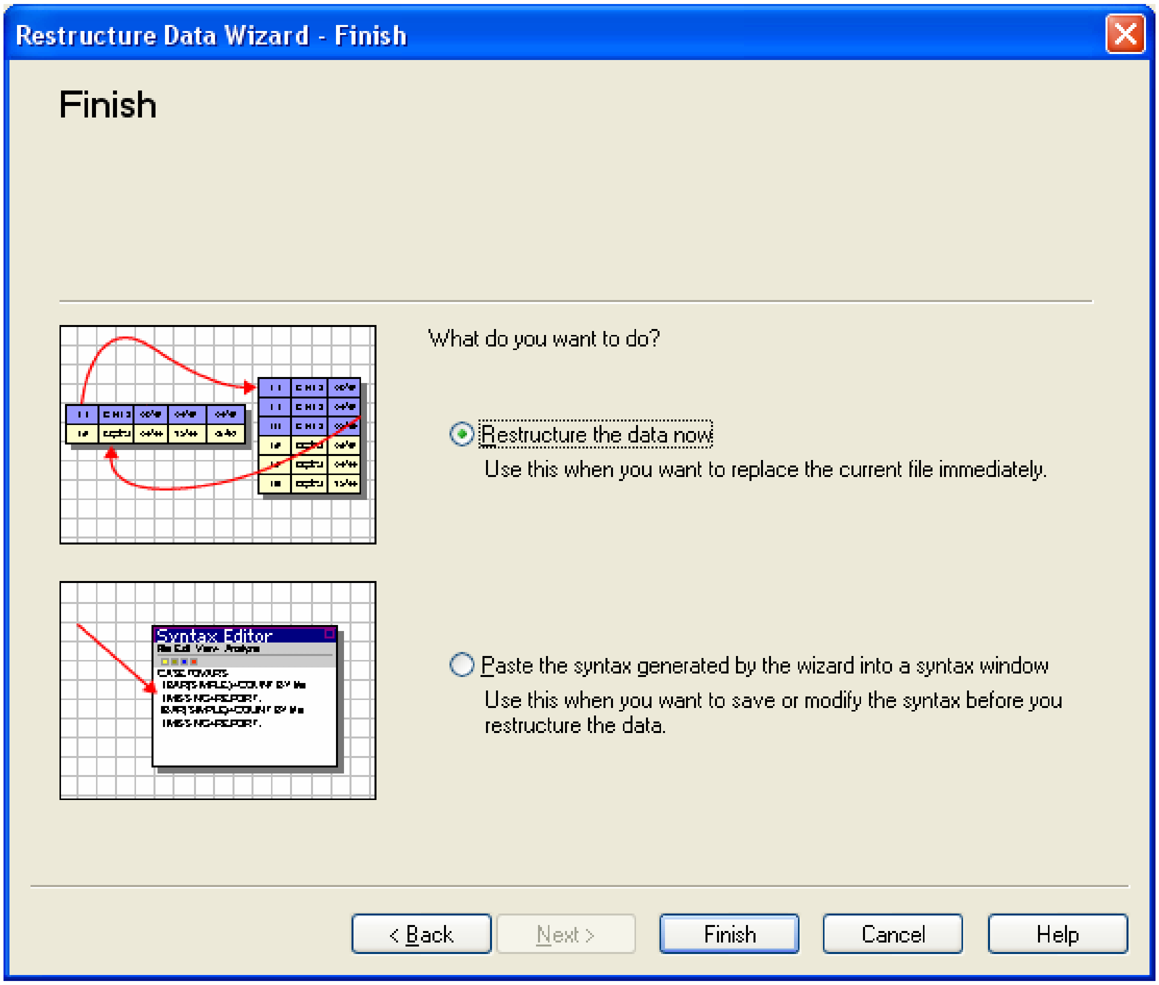
Case Count Variable

□ Count the number of new cases created by the case in the current data

Name:

Label:

In the next dialogue, we are asked if we want to create SPSS syntax which does the restructuring of the data for later reference (SPSS syntax can be used to perform ‘automatic’ analyses, or to keep track of what we have done):



Restructure Data Wizard - Finish

Finish

What do you want to do?

● Restructure the data now

Use this when you want to replace the current file immediately.

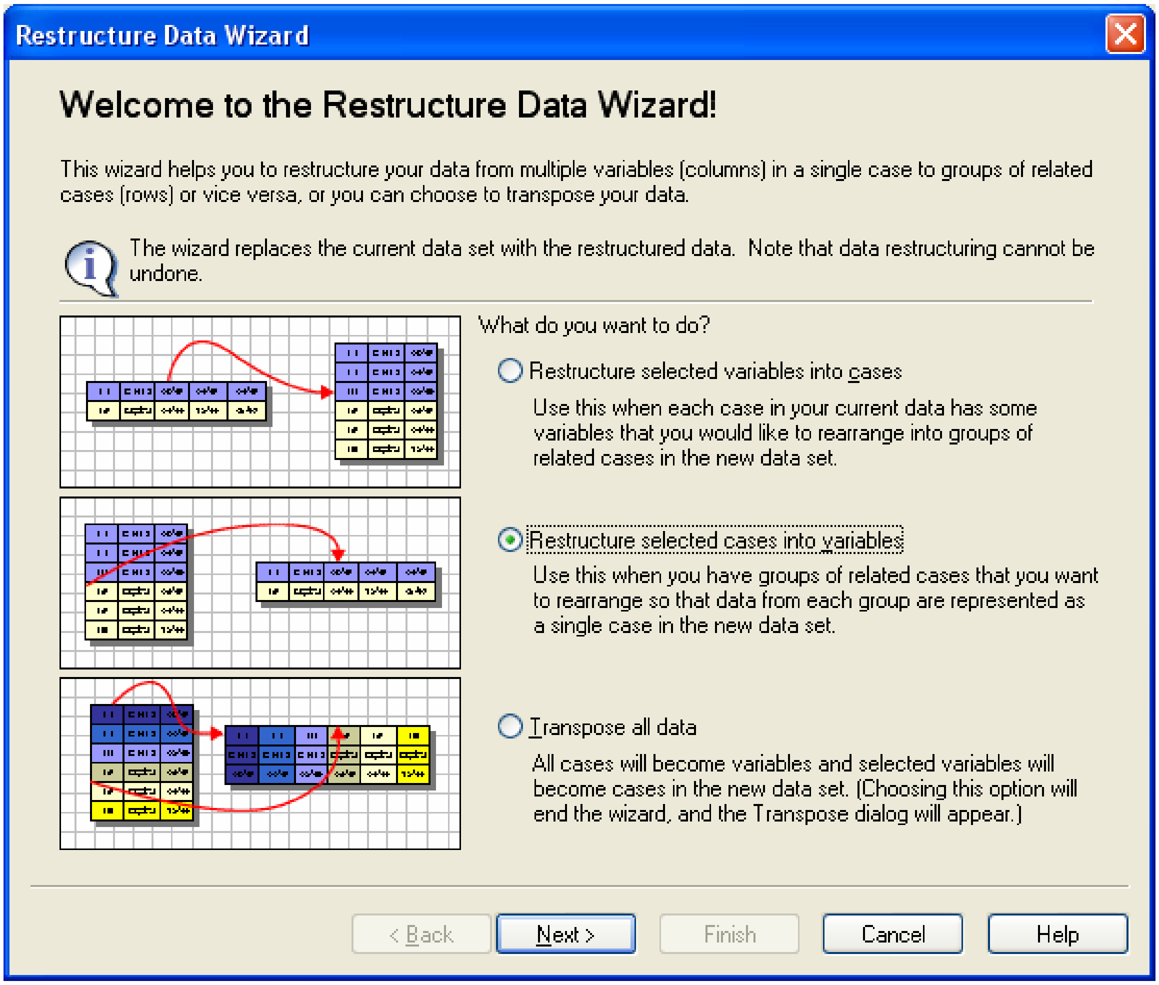
○ Paste the syntax generated by the wizard into a syntax window

Use this when you want to save or modify the syntax before you restructure the data.

After pressing ‘Finish’, the data set is immediately restructured into long format. You should save the data set now using a different name.

Switching from long to wide

We start with the data set in long format (cervpain-long.sav). Select Data-Restructure from the menu, and choose ‘Restructure selected cases into variables’:



Welcome to the Restructure Data Wizard!

This wizard helps you to restructure your data from multiple variables (columns) in a single case to groups of related cases (rows) or vice versa, or you can choose to transpose your data.

The wizard replaces the current data set with the restructured data. Note that data restructuring cannot be undone.

What do you want to do?

○ Restructure selected variables into cases

Use this when each case in your current data has some variables that you would like to rearrange into groups of related cases in the new data set.

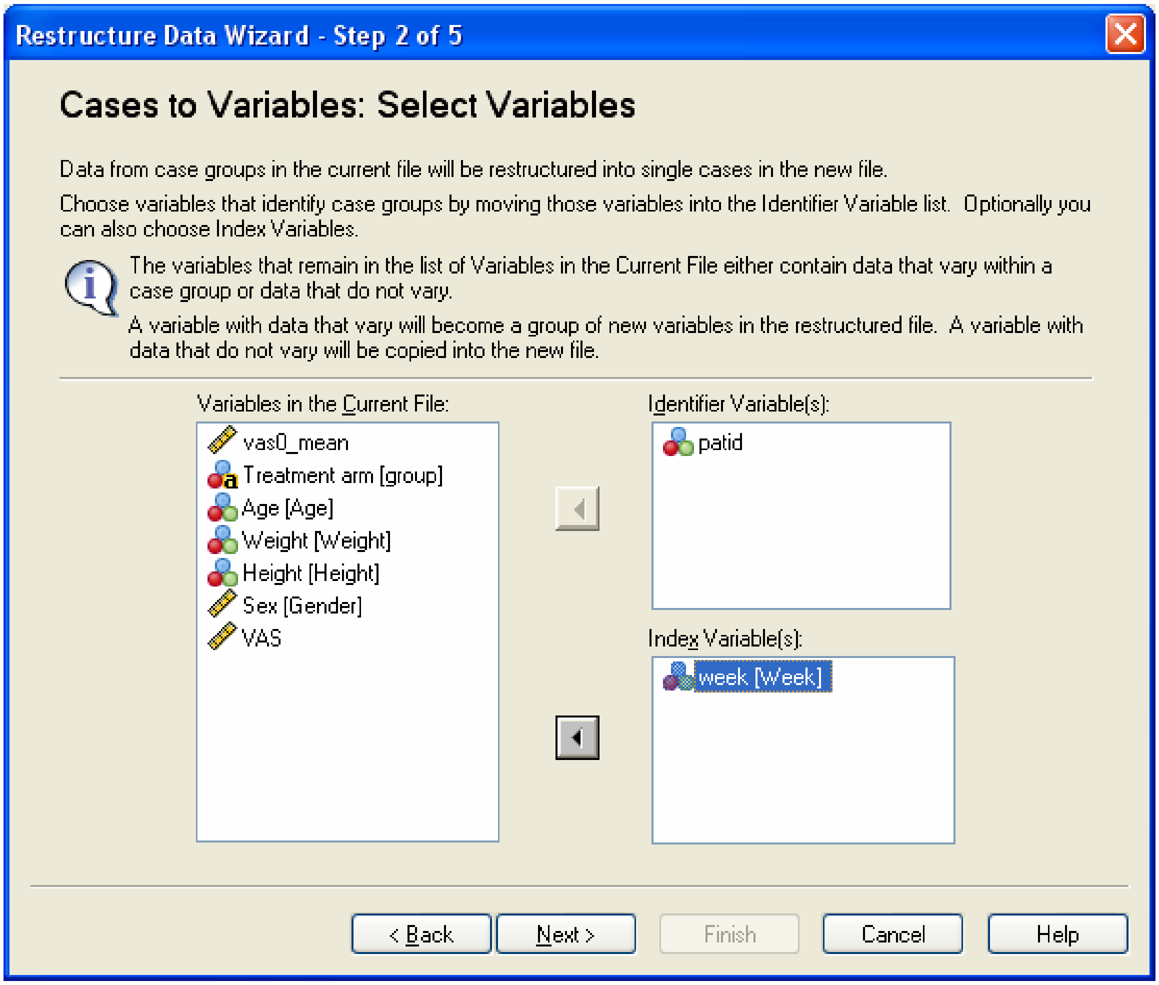
● Restructure selected cases into variables

Use this when you have groups of related cases that you want to rearrange so that data from each group are represented as a single case in the new data set.

○ Transpose all data

All cases will become variables and selected variables will become cases in the new data set. (Choosing this option will end the wizard, and the Transpose dialog will appear.)

Define ‘Patient ID’ as ‘Identifier Variable’ and ‘Week’ as ‘Index variable’:



Cases to Variables: Select Variables

Data from case groups in the current file will be restructured into single cases in the new file.

Choose variables that identify case groups by moving those variables into the Identifier Variable list. Optionally you can also choose Index Variables.

The variables that remain in the list of Variables in the Current File either contain data that vary within a case group or data that do not vary.

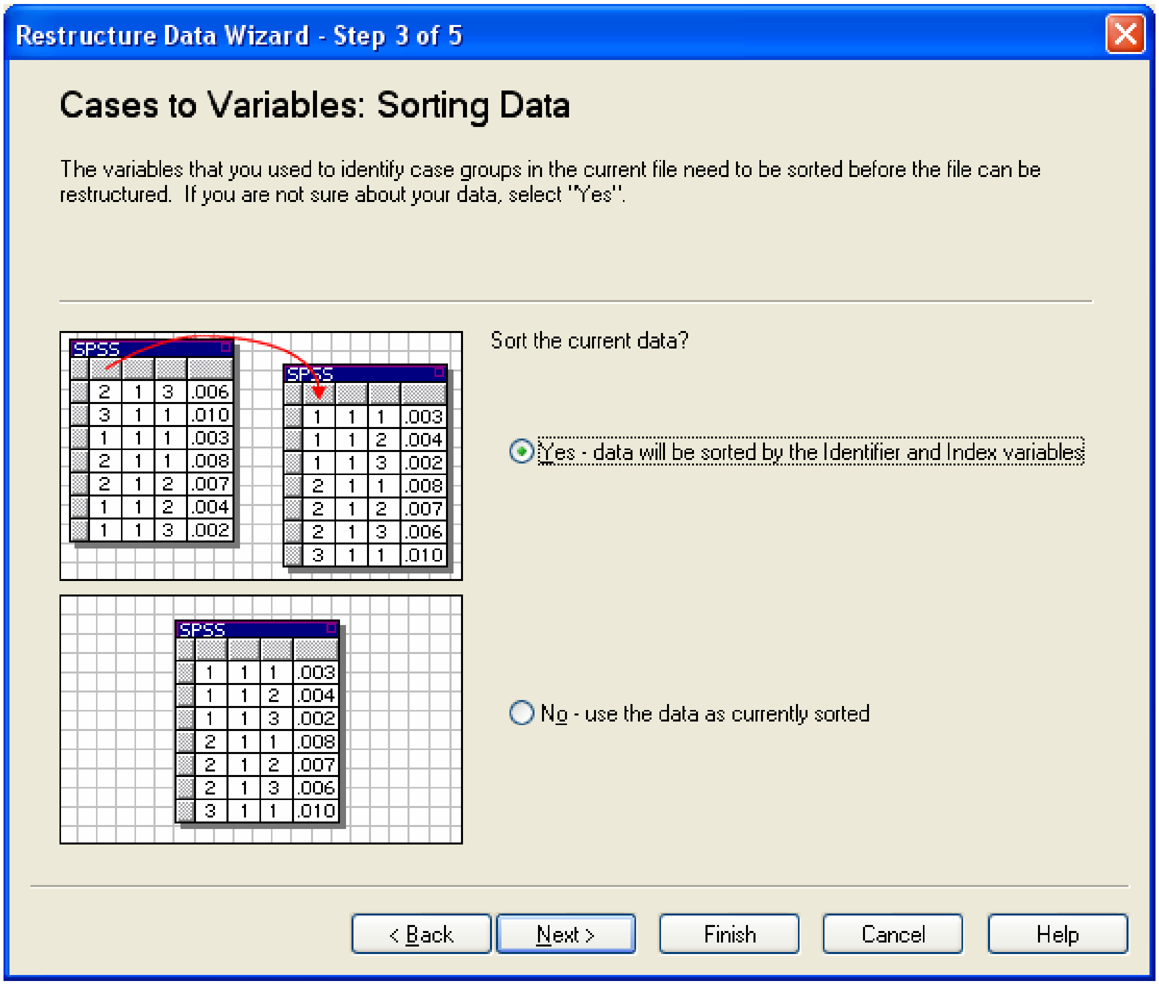
A variable with data that vary will become a group of new variables in the restructured file. A variable with data that do not vary will be copied into the new file.

Variables in the Current File:
- vas0_mean
- Treatment arm [group]
- Age [Age]
- Weight [Weight]
- Height [Height]
- Sex [Gender]
- VAS

Identifier Variable(s):
- patid

Index Variable(s):
- week [Week]

Next, we are asked if the data should be sorted (always select ‘yes’):



Restructure Data Wizard - Step 3 of 5

Cases to Variables: Sorting Data

The variables that you used to identify case groups in the current file need to be sorted before the file can be restructured. If you are not sure about your data, select "Yes".

SPSS

2	1	3	.006
3	1	1	.010
1	1	1	.003
2	1	1	.008
2	1	2	.007
1	1	2	.004
1	1	3	.002

SPSS

1	1	1	.003
1	1	2	.004
1	1	3	.002
2	1	1	.008
2	1	2	.007
2	1	3	.006
3	1	1	.010

SPSS

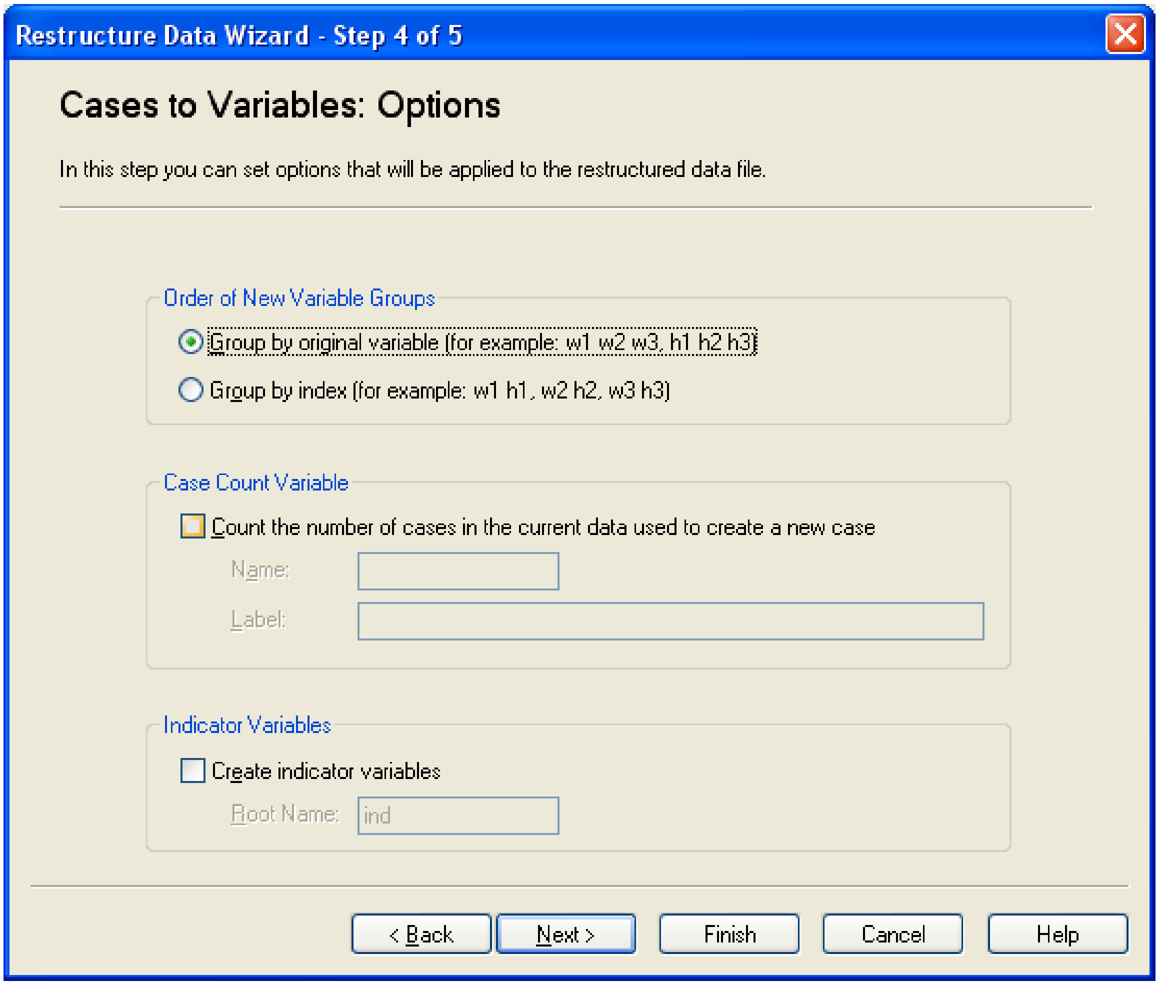
1	1	1	.003
1	1	2	.004
1	1	3	.002
2	1	1	.008
2	1	2	.007
2	1	3	.006
3	1	1	.010

Sort the current data?

● Yes - data will be sorted by the Identifier and Index variables

○ No - use the data as currently sorted

The order of the new variable groups is only relevant, if more than one variable is serially measured. In our case, we have only the VAS scores as repeated variable. Optionally, one may also create a column which counts the number of observations that were combined into one row for each subject.



Restructure Data Wizard - Step 4 of 5

Cases to Variables: Options

In this step you can set options that will be applied to the restructured data file.

Order of New Variable Groups

● Group by original variable (for example: w1 w2 w3, h1 h2 h3)

○ Group by index (for example: w1 h1, w2 h2, w3 h3)

Case Count Variable

□ Count the number of cases in the current data used to create a new case

Name:

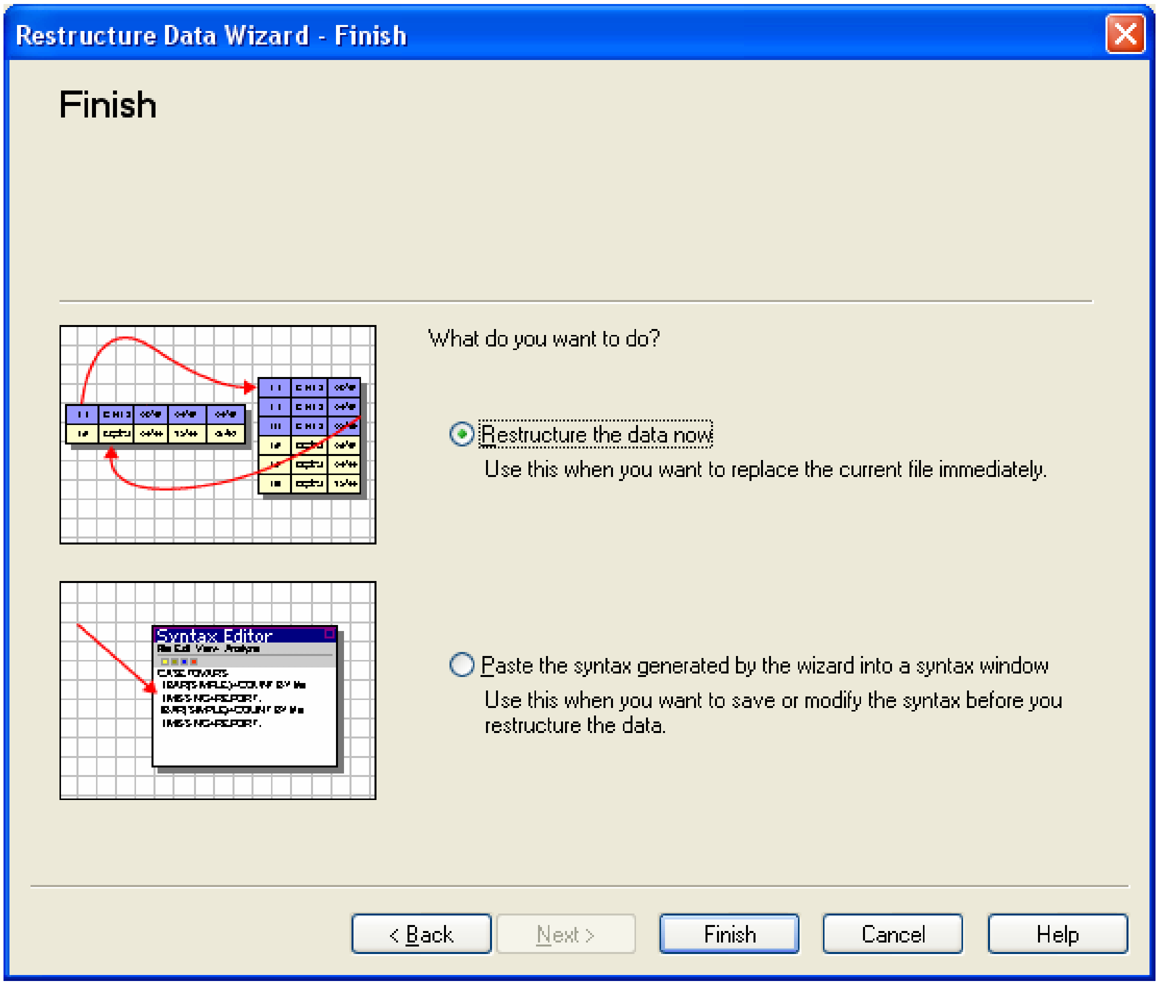
Label:

Indicator Variables

□ Create indicator variables

Root Name: ind

Pressing ‘finish’, we obtain the data set in wide format. The VAS scores are automatically named ‘VAS.1’ to ‘VAS.6’.



Restructure Data Wizard - Finish

Finish

What do you want to do?

● Restructure the data now

Use this when you want to replace the current file immediately.

○ Paste the syntax generated by the wizard into a syntax window

Use this when you want to save or modify the syntax before you restructure the data.

References

[1] J. Kleinhenz, K. Streitberger, J. Windeler, A. Gussbacher, G. Mavridis, and E. Martin. Randomised clinical trial comparing the effects of acupuncture and a newly designed placebo needle in rotator cuff tendonitis. Pain, 83:235-241, 1999.

[2] M. Bland. An Introduction to Medical Statistics, 3rd ed. Oxford University Press, Oxford, 1995.

[3] X. Guo and B. P. Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. The American Statistician 58: 16-24, 2004.

[4] M. Bland and D. Altman. Calculating correlation coefficients with repeated observations: Part 1: correlation within subjects. British Medical Journal, 310:446, 1995.

[5] J. N. S. Matthews, D. Altman, M. Campbell and P. Royston. Analysis of serial measurements in medical research. British Medical Journal, 300:230-235, 1990.

[6] S. Sator-Katzenschlager, et al. Electrical stimulation of auricular acupuncture points is more effective than conventional manual auricular acupuncture in chronic cervical pain: A pilot study. Anesth Analg, 97:1469-73, 2003.

Concluding remarks

I appreciate to receive any comments on the present lecture notes (suggestions, corrections, etc.) by e-mail: georg.heinze @ meduniwien.ac.at.

The data files that are referred to in these notes can be downloaded from
http://www.meduniwien.ac.at/msi/biometrie/lehre
(Click on the link Medical Biostatistics 2)