Analysis of Covariance (ANCOVA)

Covariance analysis is a kind of analysis of variance in which the data are adjusted (or controlled) for the presence of one or more other variables. Covariance is used when the response variable y, in addition to being affected by the treatments, is also linearly related to another variable x.

A simple analysis of covariance model looks like this:

$$y_{ij} = \alpha_j + \beta x_i + e_{ij}$$

y_{ij} … i[th] observation at level j of the factor
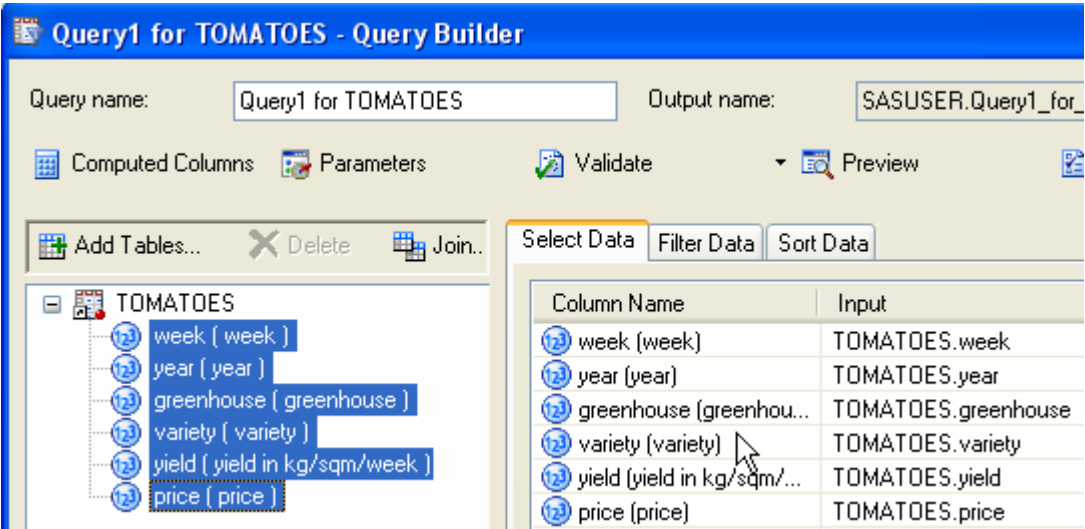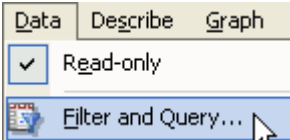α_j … j[th] effect of factor a
β … regression coefficient
x_{ij} … independent variable with the ith observation at level j of the factor
e_{ij} … random effekt with the ith observation at level j of the factor

As an example we use dataset tomatoes.
As there are several time points with no yield we filter these data and use only those observations where yield>0.
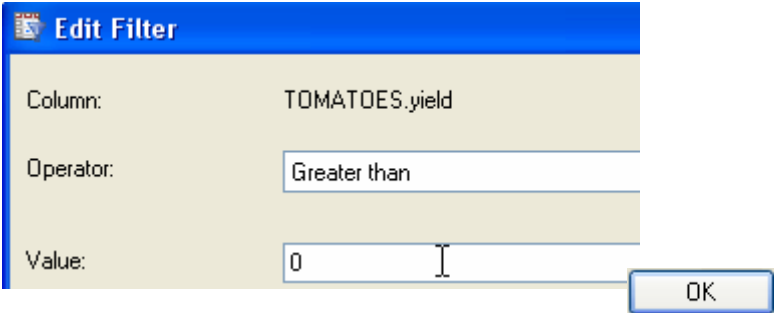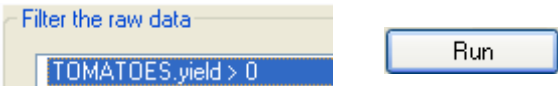

To create a Filter we have to choose
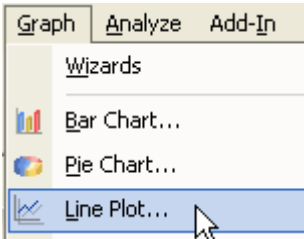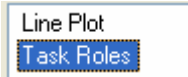

and in the follow

As a condition we select

This results in the following condition .

Now we create some plots to get an idea about the distribution of our data and create a line plot. We select
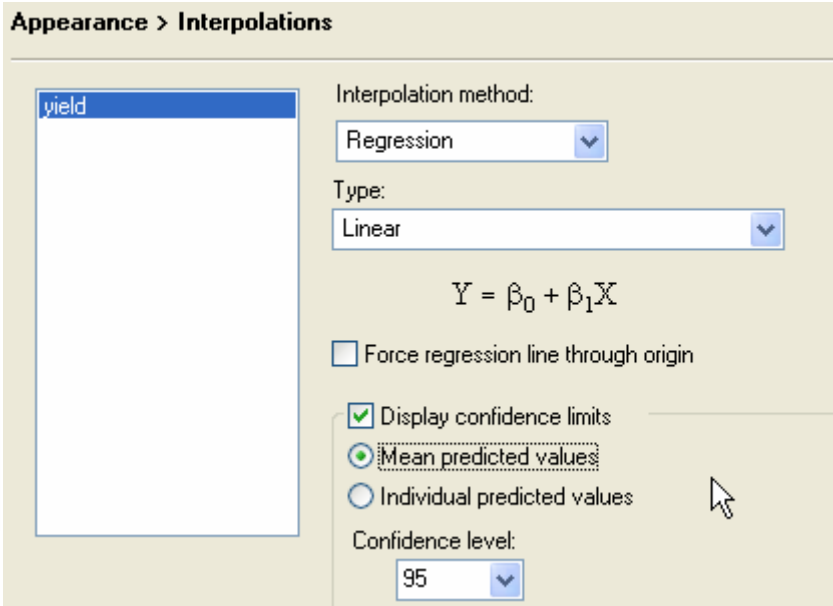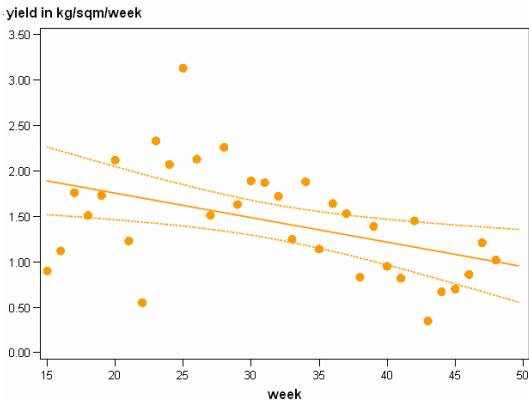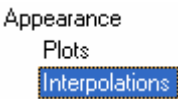

In the left window we choose

We want to know if yield is influenced by variety, greenhouse and year (grouping variables). Obviously yield depends on week so it forms the horizontal axis. Yield is our dependent variable, so we put it on the vertical axis.
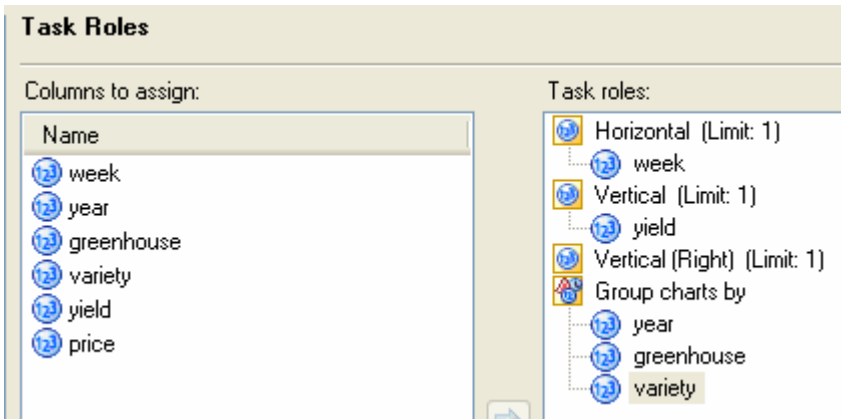
As an interpolation mode we try

And in a second step we select Quadratic.
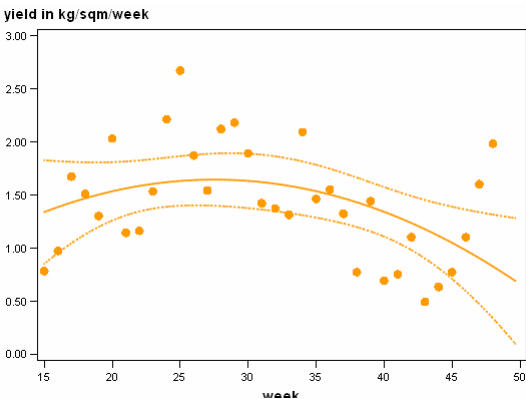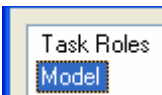
Maybe polynomials of degree 2 fit better than a linear regression, as relation between yield and week is hardly linear The created plots show a high variance but a similar evolution of yield.

$$Y = \beta_0 + \beta_1 X$$

Based on these results we start with an analysis of covariance with factors variety, greenhouse and year. That kind of analysis is handled with "Lineare Models".
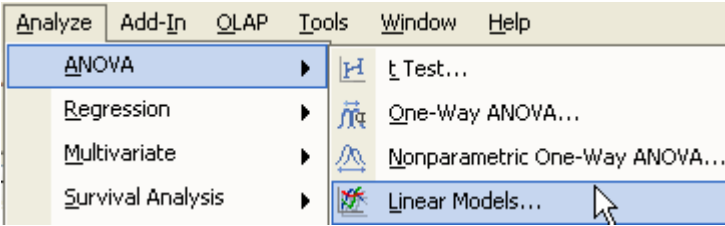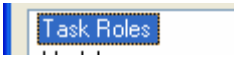

With we select corresponding to our plots "yield" as dependent, "week" as quantitative and "greenhouse, variety, year" as classification variables.
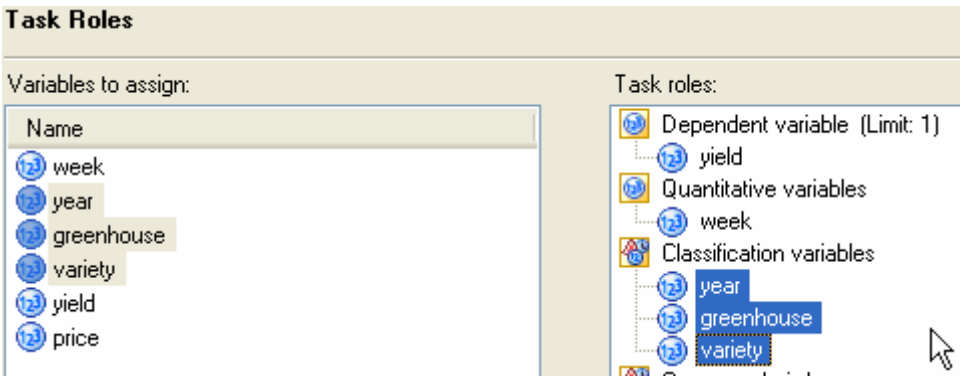
 Now we build our model
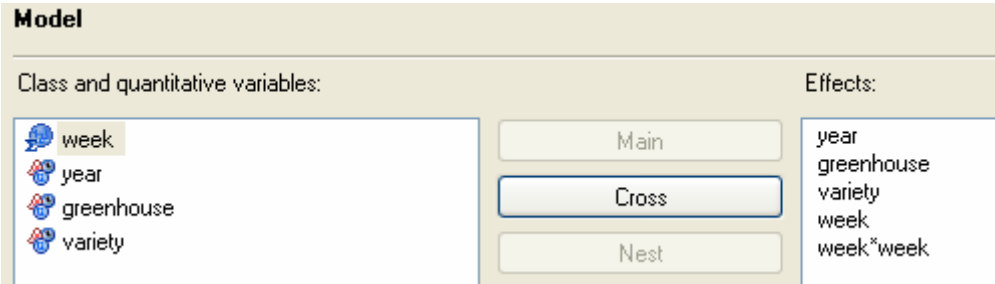
To get a polynomial of degree 2 we select week in the left window and press

To get an idea about the quality of our model we plot residuals vs. predicted yield.

Results

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	63.6140439	7.9517555	27.40	<.0001
Error	659	191.2219262	0.2901698		
Corrected Total	667	254.8359701			

From "Pr>F" we recognise that there are some important influences.

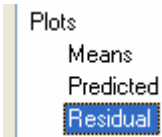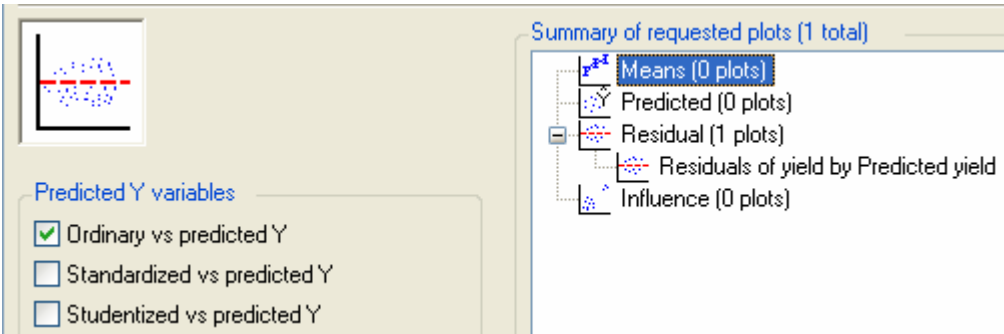| R-Square	Coeff Var	Root MSE	yield Mean
0.249627	38.01656	0.538674	1.416946

"R-Square" is relatively low.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
greenhouse	1	0.05872498	0.05872498	0.20	0.6530
variety	4	0.16613405	0.04153351	0.14	0.9660
year	1	0.34567911	0.34567911	1.19	0.2755
week	1	12.80359098	12.80359098	44.12	<.0001
week*week	1	20.66163611	20.66163611	71.21	<.0001

The only significant results are week and week*week

Residual of yield
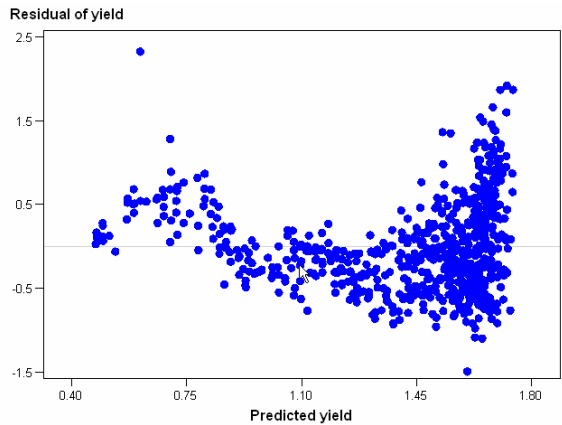
The residual plot shows some trend in the distribution of our data, which means, that we should try to enhance our model.

We can do this in several ways.

We apply a polynomial of degree 3.
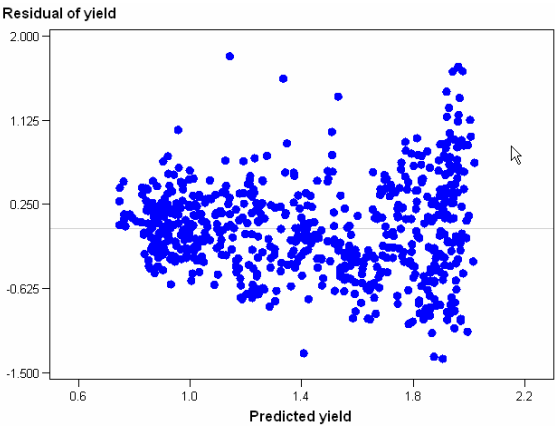
The distribution of residuals now looks quite good.

The "R-Square" value raise to ~ 40%.

The value that we are really interested in is variety. So we try to further enhance our model. We can do this by nesting week, week*week, week*week*week within greenhouse, variety and year.

We select "greenhouse, variety, year" in the left window and "week, week*week, week*week*week" in the left window and press

This leads to a further enhancement of the "R-Square" value to ~ 41%.

At last we try to use interactions between factors.

Muss noch geändert werden

Comparing two regression curves
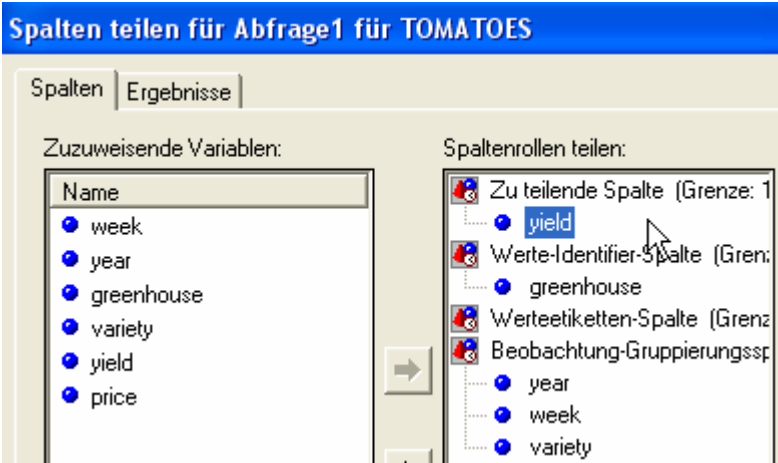
If we want to compare development between regression curves in respect to "greenhouse", we have to split "yield" corresponding to "greenhouse".

We use a prefix for the columns which are to be created "grh".

In a second step we create a new variable "diff" which is the difference between the "ghr1" and "grh2".

This new variable will become significant if there is a different development within "greenhouse".

Lineare Modelle für Geteilte Abfrage1 für TOMATOES

Spalten | Modell erstellen | Modelloptionen | Post-Hoc-Tests | Plots |

Zuzuweisende Variablen:

Name
year
Diff
week
variety
grh1
grh2

Lineare Modellrollen:
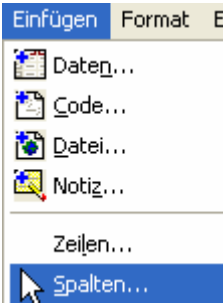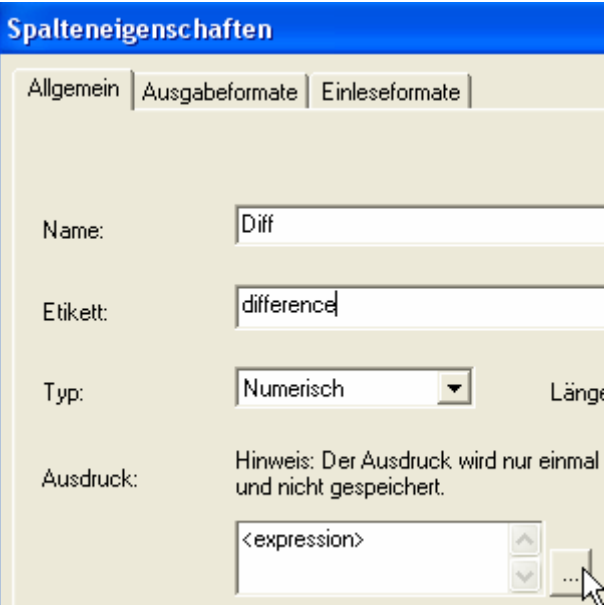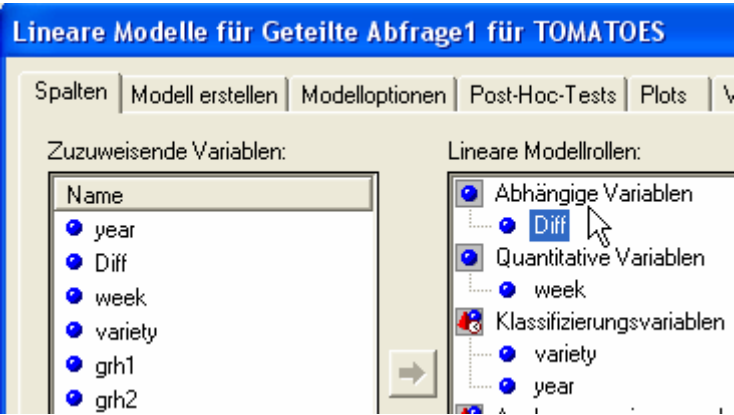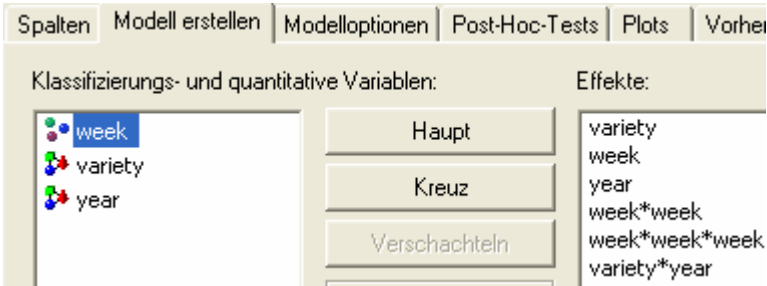
- Abhängige Variablen
  - Diff
- Quantitative Variablen
  - week
- Klassifizierungsvariablen
  - variety
  - year

 Our model for "diff" is similar to that of "yield".

Spalten | Modell erstellen | Modelloptionen | Post-Hoc-Tests | Plots | Vorher

Klassifizierungs- und quantitative Variablen:

- week
- variety
- year

[Haupt]
[Kreuz]
[Verschachteln]

Effekte:

variety
week
year
week*week
week*week*week
variety*year

Results

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	1.87702379	0.15641865	0.94	0.5089
Error	320	53.35701945	0.16674069		
Corrected Total	332	55.23404324			

R-Square	Coeff Var	Root MSE	Diff Mean
0.033983	−2168.690	0.408339	−0.018829

Source	DF	Type III SS	Mean Square	F Value	Pr > F
variety	4	0.62267870	0.15566968	0.93	0.4446
week	1	0.26067375	0.26067375	1.56	0.2121
year	1	0.00731703	0.00731703	0.04	0.8342
week*week	1	0.20101000	0.20101000	1.21	0.2730
week*week*week	1	0.15549209	0.15549209	0.93	0.3349
variety*year	4	0.61773224	0.15443306	0.93	0.4489

As there are no significant results and the "R-Square" values is very low, we believe in a similar development of tomatoes within "greenhouse".