sas club

Der Business Analytics Club für SAS User

MEHR ALS LINEAR ODER LOGISTISCH?

QUANTILS REGRESSIONEN UND ADAPTIVE SPLINES IN SAS

MIHAI PAUNESCU

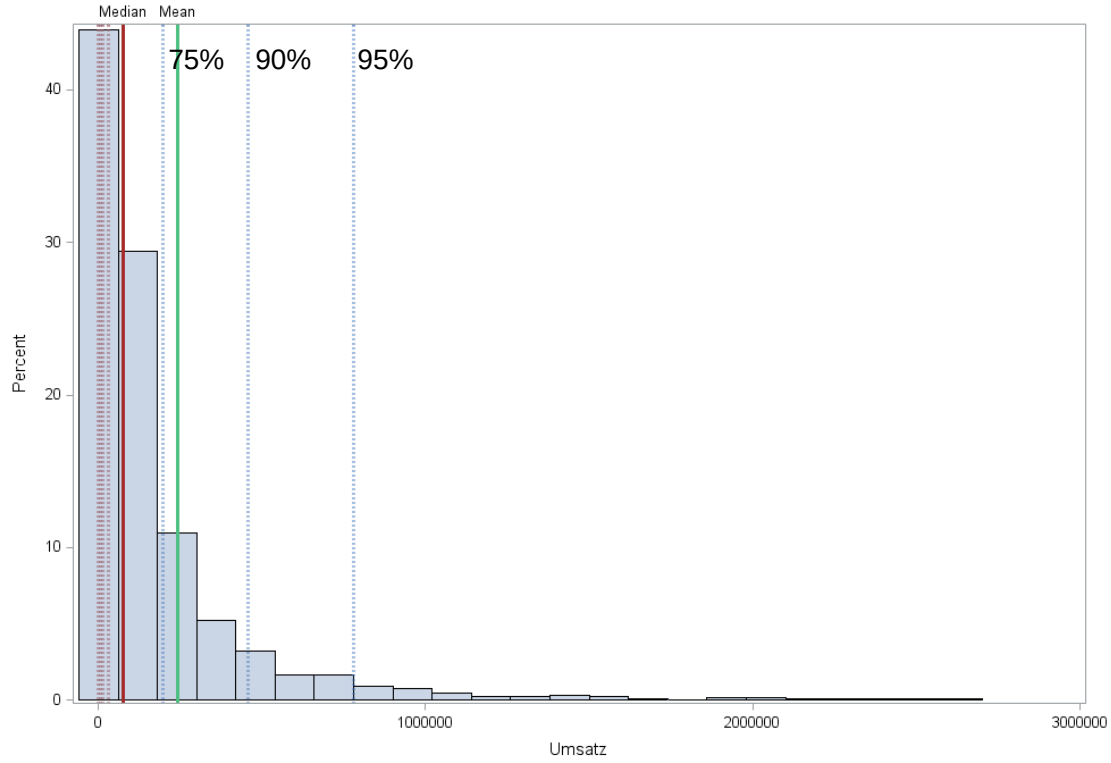
QUANTILE


```
proc univariate data=dat;
 ods select moments quantiles;
 var sales;
run;
```

Basic Statistical Measures			
Location		**Variability**	
Mean	241825.8	**Std Deviation**	823581
Median	**75875.1**	**Variance**	6,78E+16
Mode	.	**Range**	18605198
		Interquartile Range	168416

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	18605230.40
99%	2699730.30
95%	779877.70
90%	458327.90
75% Q3	198564.00
50% Median	**75875.10**
25% Q1	30147.60
10%	13607.10
5%	5598.60
1%	500.75
0% Min	32.55

QUANTILE IM HISTOGRAMM



Quantiles (Definition 5)	
Level	**Quantile**
100% Max	18605230.40
99%	2699730.30
95%	779877.70
90%	458327.90
75% Q3	198564.00
50% Median	75875.10
25% Q1	30147.60
10%	13607.10
5%	5598.60
1%	500.75
0% Min	32.55

```
ods graphics on / width=1000px height=700px;
proc sgplot data=dat;
 where sales < 2699730.30;
 histogram sales;
 refline 779877.70 458327.90 198564.00 / axis=x lineattrs=(color=bigb pattern=dot thickness=3);
 refline 30147.60 13607.10 5598.60 500.75 /axis=x lineattrs=(color=dapk pattern=dot thickness=3);
 refline 75875.10 /axis=x label='Median' lineattrs=(color=brown thickness=3);
 refline 241825.8 /axis=x label='Mean' lineattrs=(color=big thickness=3);
run;
```

PROC QUANTREG UND MEDIAN


```
proc quantreg data=dat;
  model Sales= / quantile=(0.5);
run;
```

Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > |t|
Intercept	1	75875.10	2.654.220	70.670.691	81.079.509	28.59	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	18605230.40
99%	2699730.30
95%	779877.70
90%	458327.90
75% Q3	198564.00
50% Median	75875.10
25% Q1	30147.60
10%	13607.10
5%	5598.60
1%	500.75
0% Min	32.55

KLASSIFIKATIONSVARIABLEN


```
proc quantreg data=dat ci=resampling ;
 class cust_grp;
 model Sales=cust_grp / quantile=(0.5) seed=12345;
run;
```

Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > |t|
Intercept		1	47335.60	3.071.371	41.313.237	53.357.963	15.41	<.0001
Cust_grp	I	1	88078.60	6.903.957	74.541.278	101615.92	12.76	<.0001
Cust_grp	O	1	21870.35	5.324.410	11.430.214	32.310.486	4.11	<.0001
Cust_grp	N	0	0.0000	0.0000	0.0000	0.0000	.	.

```
proc means data=dat P1 median P90;
 var sales;
 class Cust_grp;
run;
```

Analysis Variable : Sales

Cust_grp	N Obs	Median
I	793	135414.2
O	1024	69062.08
N	1004	47323.3

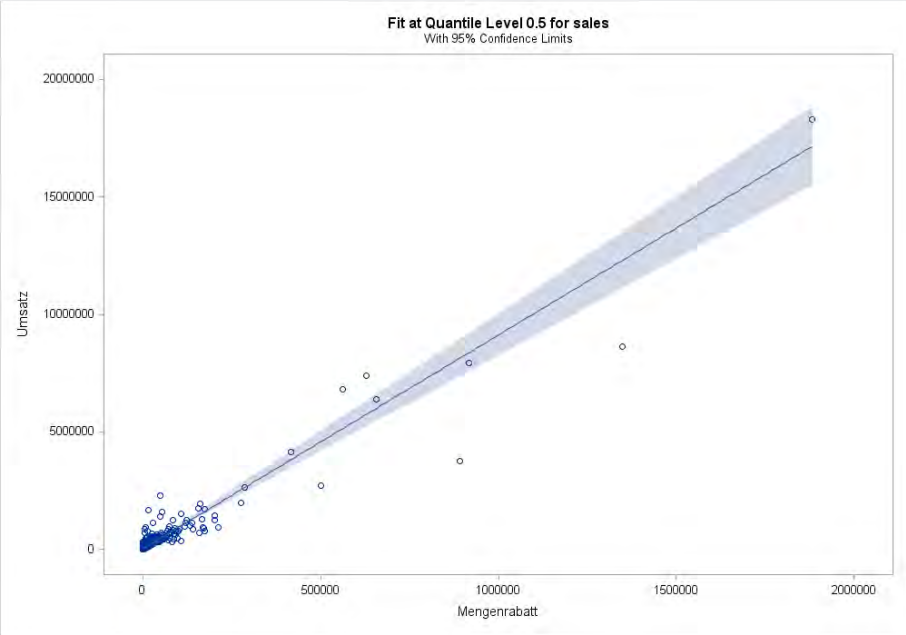
47335.6 + 88078.6 = 135414.2

MEDIAN SCHÄTZUNG FÜR INTERVALVARIABLE

```
proc quantreg data=dat ci=resampling plots=(fitplot);
 where cust_grp='n';
 model Sales=Rabatt_Menge / quantile=(0.5) seed=12345;
run;
```

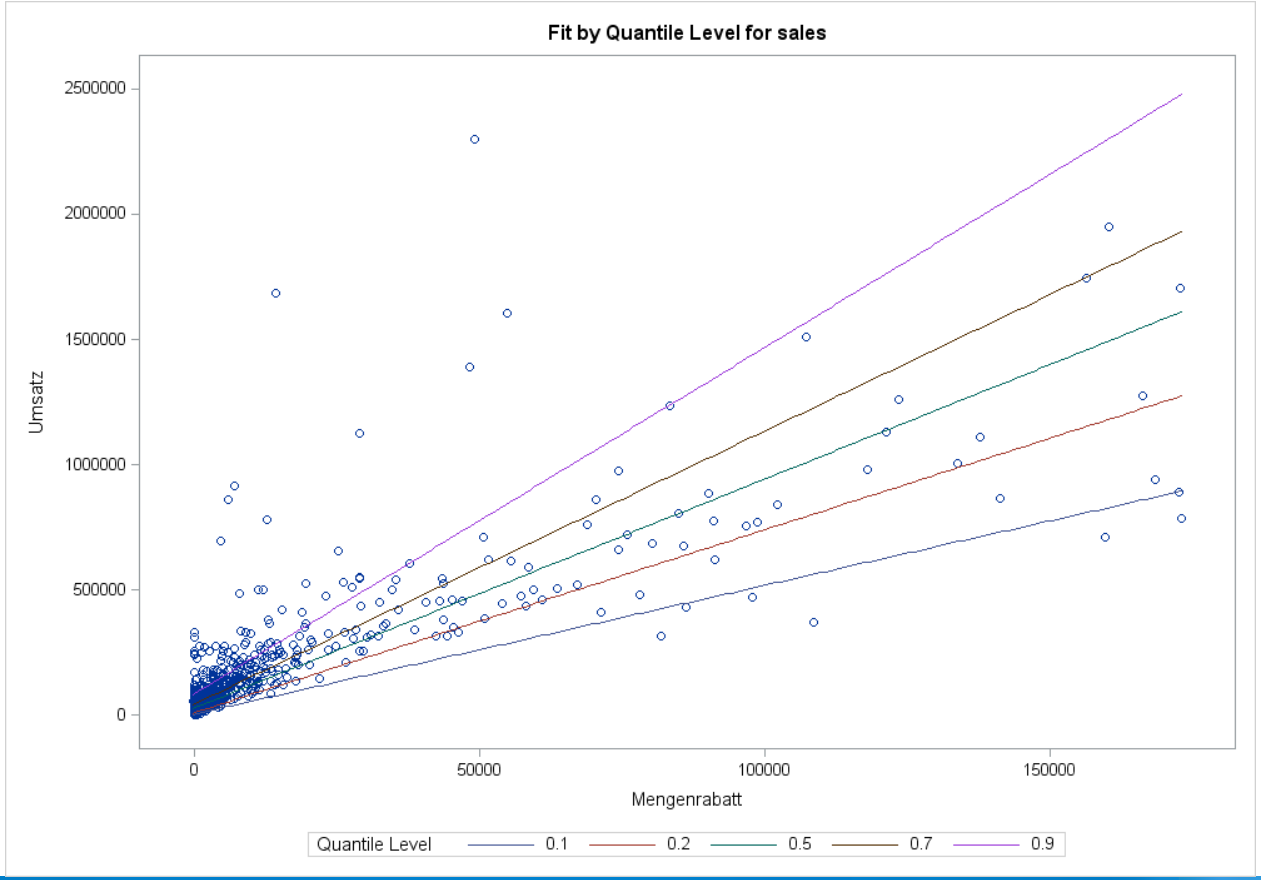
Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > |t|
Intercept	1	28552.1	5873.371	17026.581	40077.615	4.86	<.0001
Rabatt_Menge	1	9.1	0.4522	8.2082	9.9829	20.12	<.0001



QUANTILSCHÄTZUNG FÜR INTERVALVARIABLE

```
proc quantreg data=dat ci=resampling plots=(fitplot);
 ods select fitplot;
 where cust_grp='n' and Rabatt_Menge < 200000;
 model Sales=Rabatt_Menge / quantile=(0.1 0.2 0.5 0.7 0.9)  seed=12345;
run;
```



13.83

10.9

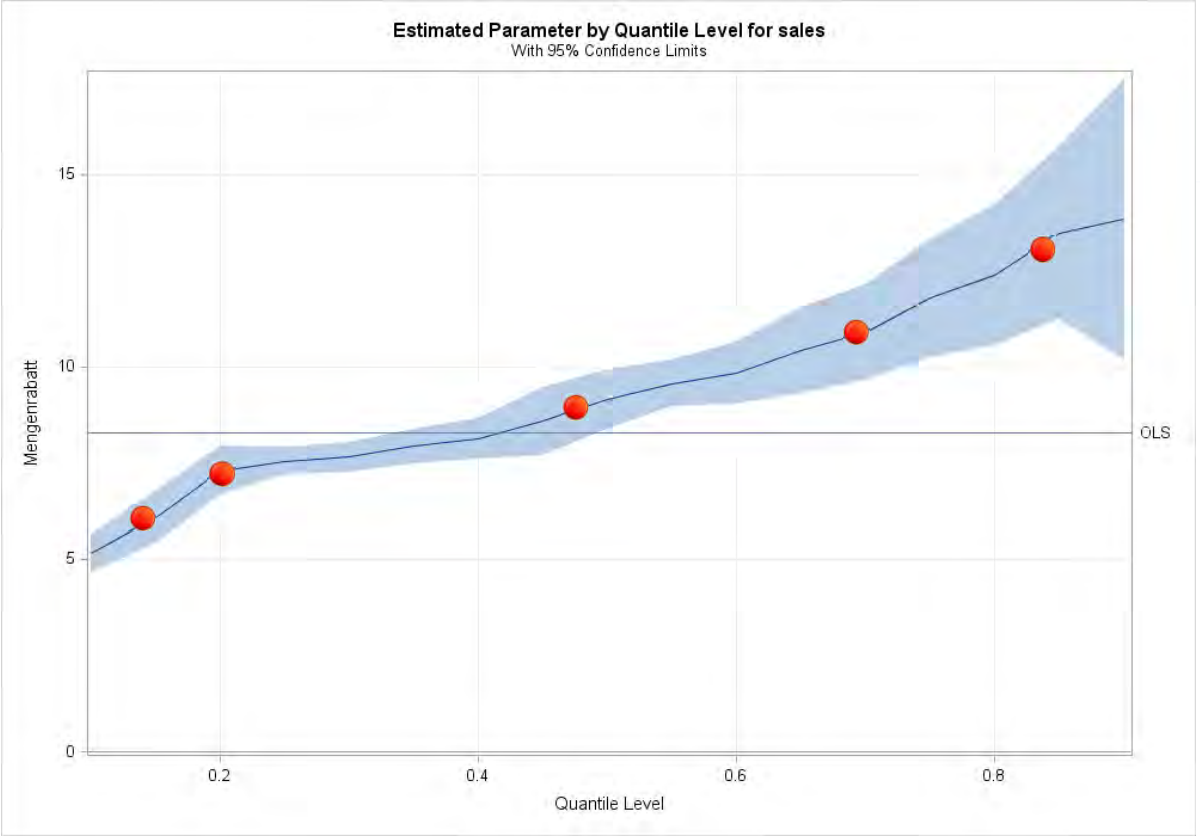
9.14

7.31

5.15

QUANATILE PROZESS PLOT

```
proc quantreg data=dat ci=resampling;
ods select quantplot;
where cust_grp='n' and Rabatt_Menge < 200000;
model Sales=Rabatt_Menge
/quantile=(0.1 to 0.9 by 0.05)  plot=(quantplot /unpack ols) seed=1268 ;
run;
```



13.83

10.9

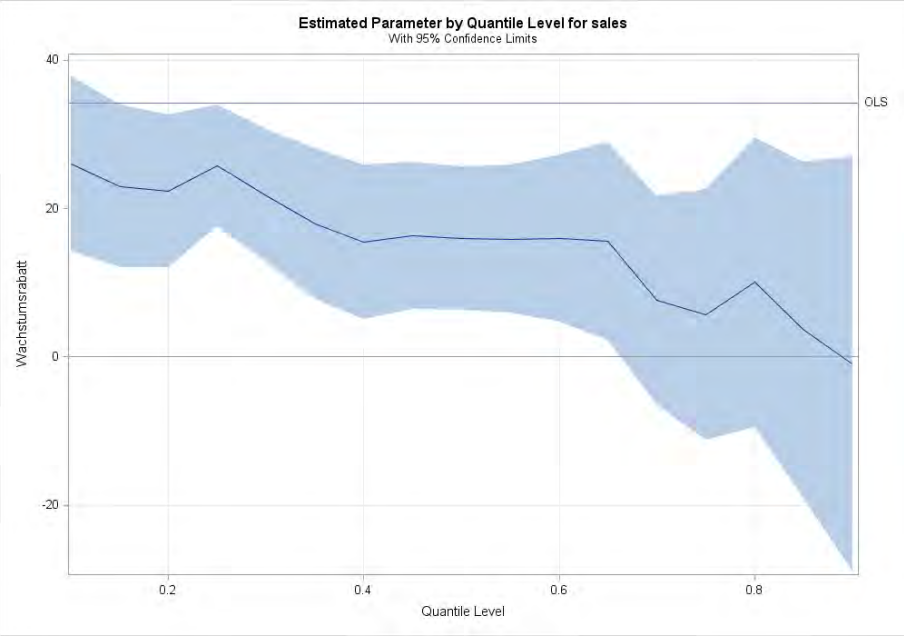
9.14

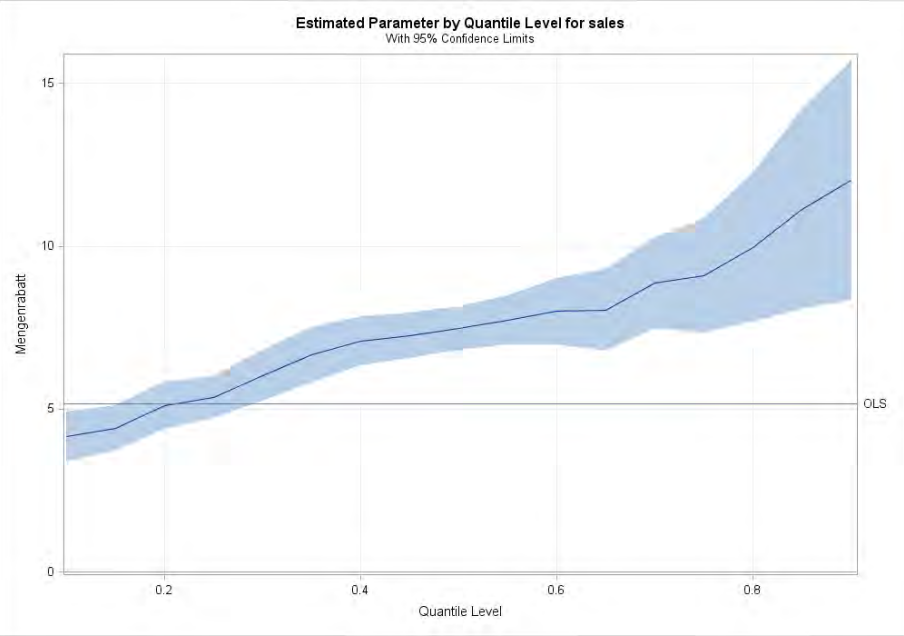
7.31

5.15

MULTIVARIATE QUANTILSREGRESSION


```
proc quantreg data=dat ci=resampling;
where cust_grp='n';
 model Sales=Rabatt_Wachstum Rabatt_Menge  /
 quantile=(0.1 to 0.9 by 0.05)  plot=(quantplot /unpack ols) seed=1268 ;
run;
```





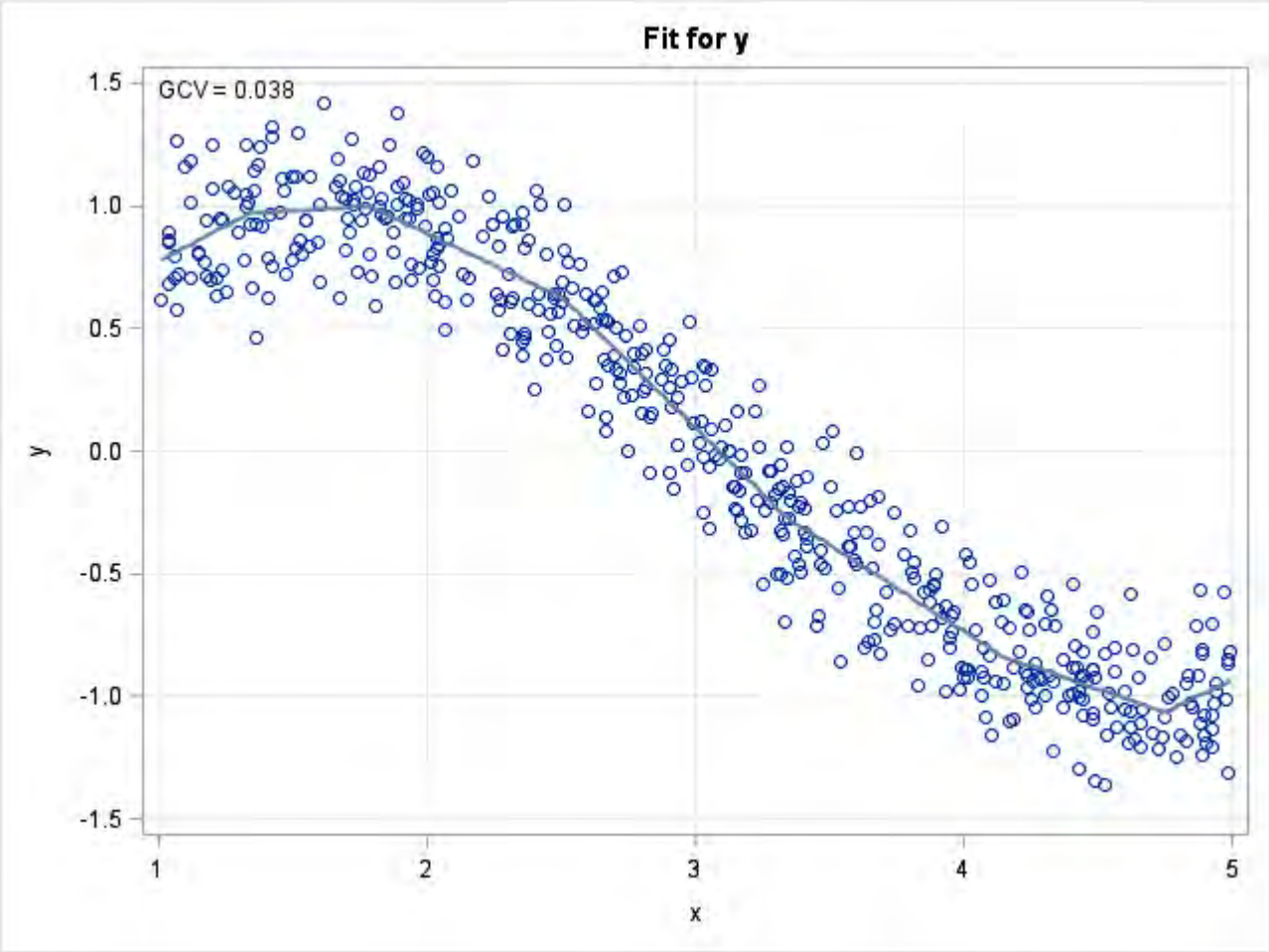
FAZIT

- Verwendung:
  - Zusammenhänge entdecken für extreme Bereiche der Zielvariable
  - Robuste Medianschätzung gegenüber Ausreißer, ohne Verteilungsannahmen

- Hinweise:
  - Ist nicht äquivalent zu linearen Regressionen für Segmente von Beobachtungen
  - Schneller: HPQUANTSELECT ab SAS/STAT 13.2

ADAPTIVEREG

MULTIVARIATE ADAPTIVE REGRESSION SPLINES



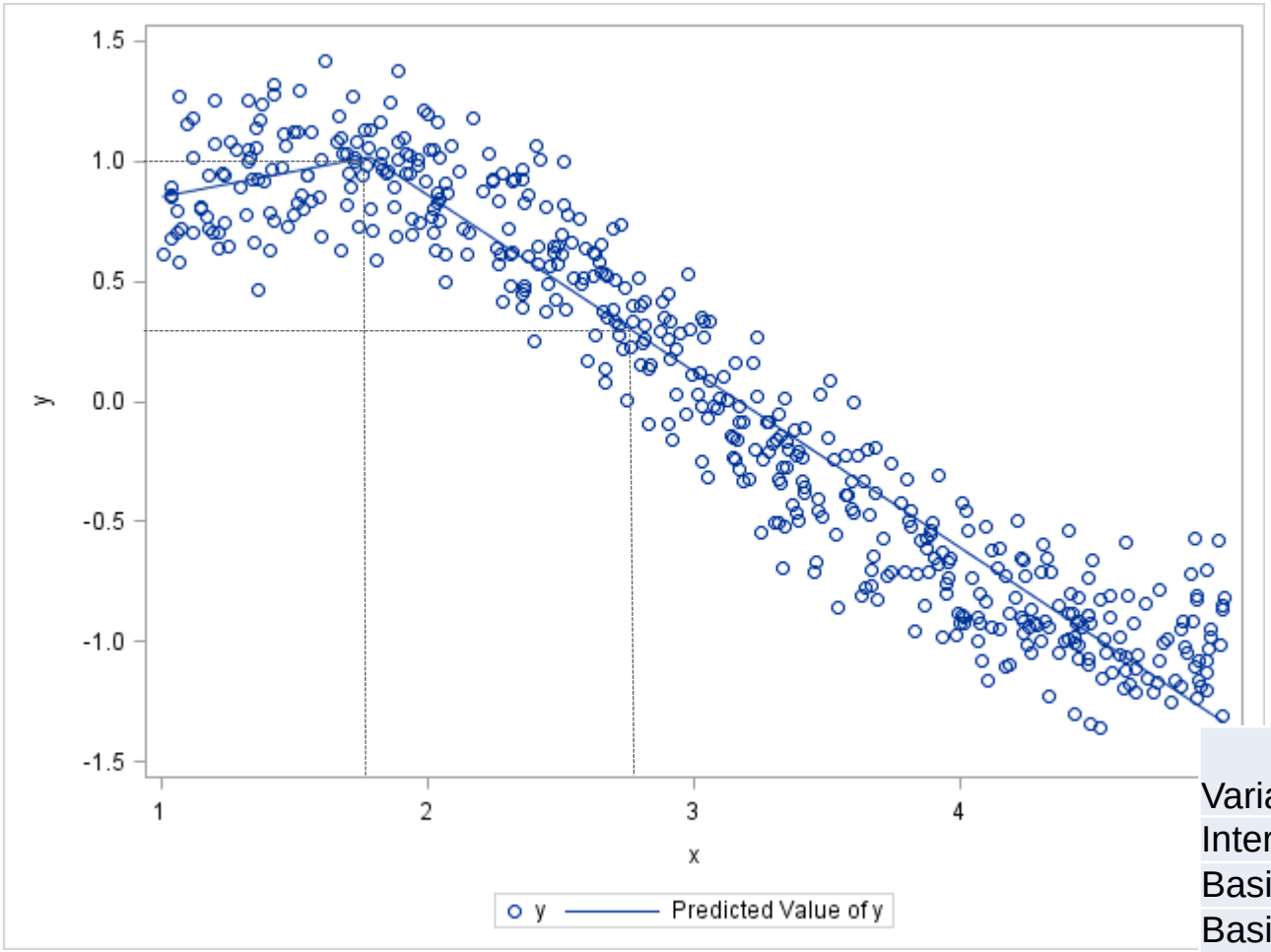

```
proc adaptivereg
plots=all
details=bases;

model y = x;

run;
```

ADAPTIVEREG

SELEKTION VON GEEIGNETEN KNOTEN WERTEN



Basis1	MAX(x - 1.8,0)
Basis2	MAX(1.8 - x,0)

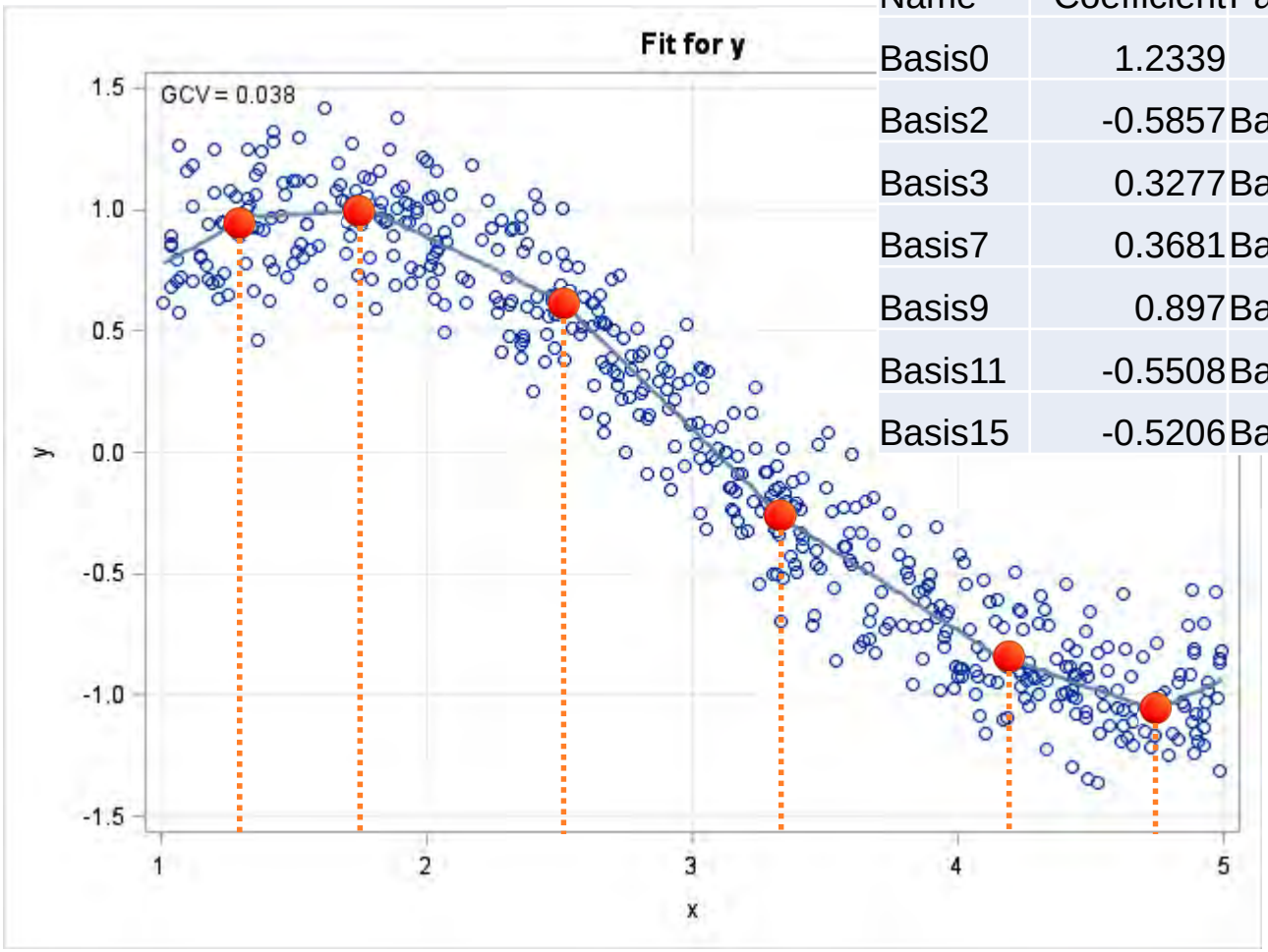
```
data ds2;
set ds;
Basis1 = max(x-1.8, 0);
Basis2 = max(1.8 - x, 0);
run;

proc reg data=ds2;
 model y = basis1 basis2;
run;
```

Variable	Parameter Estimate	t Value	Pr > |t|
Intercept	1.02	52.04	<.0001
Basis1	-0.73	-67.82	<.0001
Basis2	-0.21	-3.31	0.001

ADAPTIVEREG

REDUKTION DES MODELS DURCH SELEKTIONSMETHODEN



Regression Spline Model after Backward Selection				
Name	Coefficient	Parent	Variable	Knot
Basis0	1.2339		Intercept	
Basis2	-0.5857	Basis0	x	1.7865
Basis3	0.3277	Basis0	x	4.1447
Basis7	0.3681	Basis0	x	3.3424
Basis9	0.897	Basis0	x	4.7489
Basis11	-0.5508	Basis0	x	2.5061
Basis15	-0.5206	Basis0	x	1.3345

ADAPTIVEREG

GEEIGNET FÜR KLASSIFIKATIONS-VARIABLEN UND MISSING VALUES

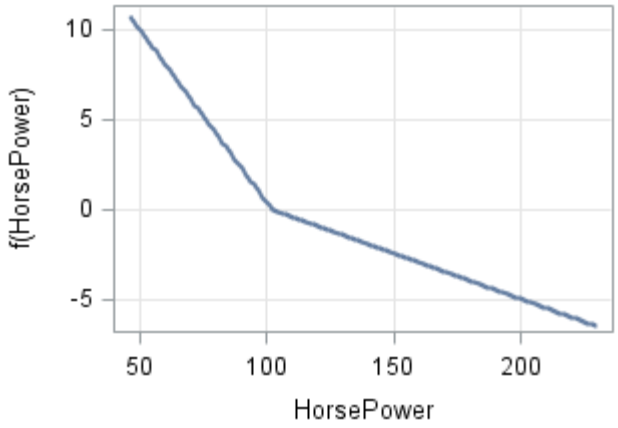
```
proc adaptivereg data=autompg plots=all details=bases;
class cylinders year origin;
model mpg = cylinders displacement horsepower
weight acceleration year origin / additive;
run;
```

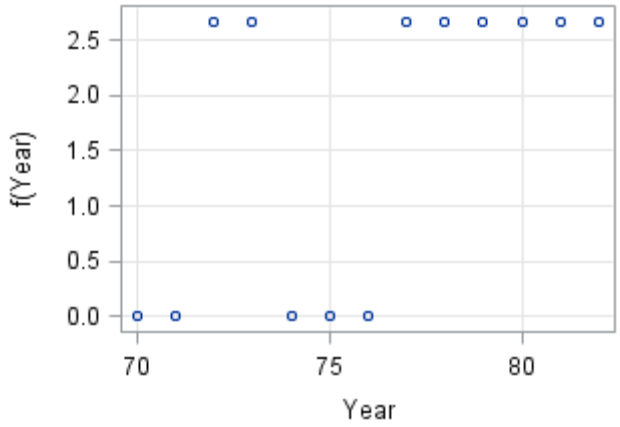
Basis Information

Name	Transformation
Basis0	1
Basis1	Basis0*MAX(Weight - 3139,0)
Basis2	Basis0*MAX(3139 - Weight,0)
Basis3	**Basis0*NOT(MISSING(HorsePower))**
Basis4	**Basis0*MISSING(HorsePower)**
Basis5	**Basis3*MAX(HorsePower - 102,0)**
Basis6	**Basis3*MAX(102 - HorsePower,0)**
Basis7	Basis0*(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)
Basis8	Basis0*NOT(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)
Basis9	Basis0*MAX(Displacement - 85,0)
Basis10	Basis0*MAX(85 - Displacement,0)
Basis11	Basis0*MAX(Displacement - 97,0)
Basis12	Basis0*MAX(97 - Displacement,0)
Basis13	Basis0*MAX(Acceleration - 21,0)
Basis14	Basis0*MAX(21 - Acceleration,0)
Basis15	Basis3*MAX(Displacement - 105,0)
Basis16	Basis3*MAX(105 - Displacement,0)

VISUALISIERUNG DER MODEL-KOMPONENTEN

Basis Information		
Name	Coefficient	Transformation
Basis3	-4.03	Basis0*NOT(MISSING(HorsePower))
Basis4	0	Basis0*MISSING(HorsePower)
Basis5	-0.05	Basis3*MAX(HorsePower - 102,0)
Basis6	0.19	Basis3*MAX(102 - HorsePower,0)
Basis7	2.67	Basis0*(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)
Basis8	0	Basis0*NOT(Year = 80 OR Year = 82 OR Year = 81 OR Year = 79 OR Year = 78 OR Year = 77 OR Year = 73 OR Year = 72)





PROC Adaptivereg
http://support.sas.com/resources/papers/proceedings13/457-2013.pdf

Weitere Nicht-Parametrische Regressionen: PROC GAM, PROC LOESS