

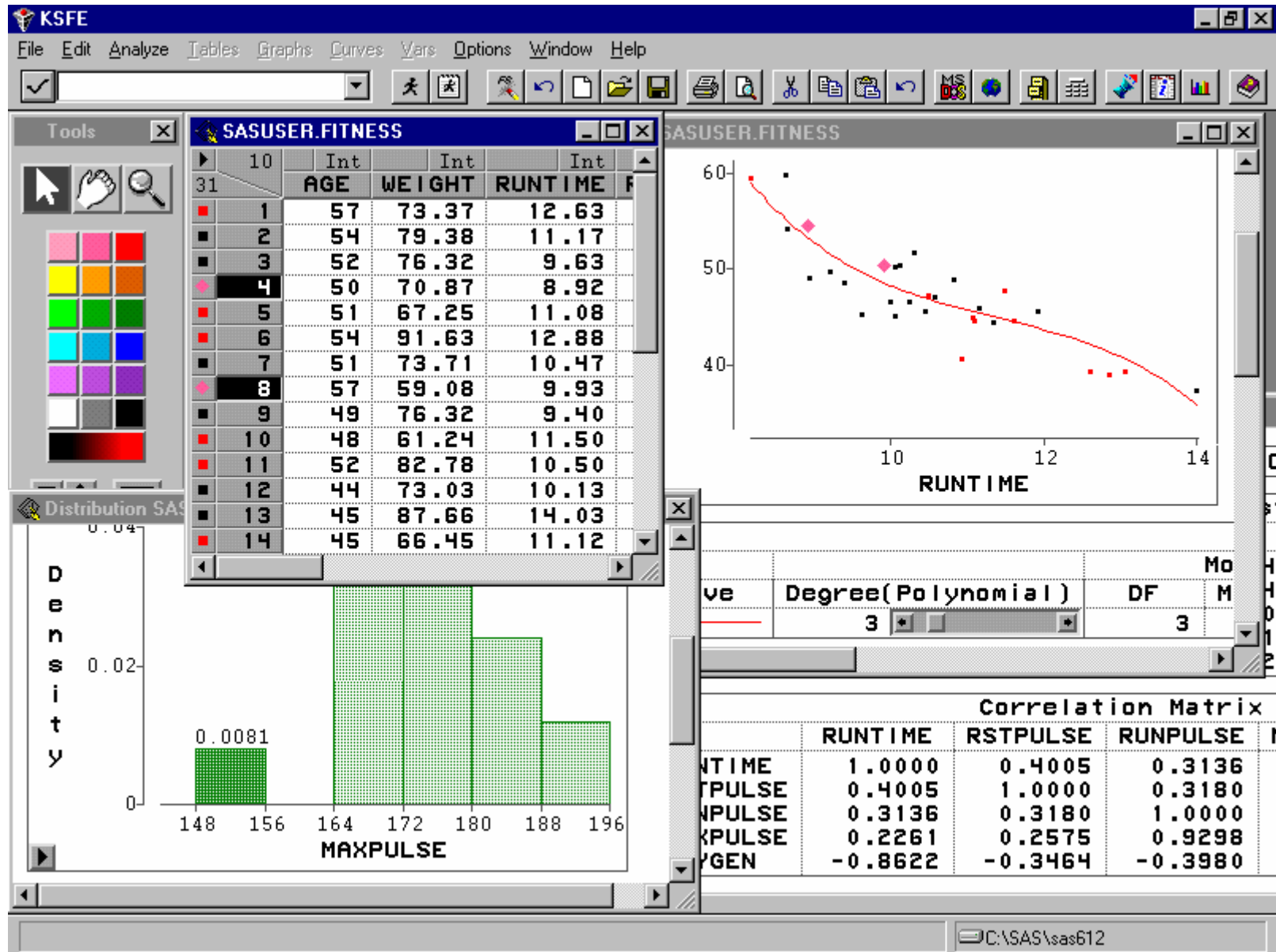
SAS/INSIGHT

interaktive Datenanalyse

Iris Krammer
SAS Institute GmbH
CC Datamining

Agenda:

- Einleitung
- Datendarstellung
- Datenvisualisierung
(grafikorientiert)
- Datenanalyse
(zahlenorientiert)
- Informationen



Insight ist...

- interaktiv
- menügesteuert
- für Experten und Nichtstatistiker

Möglichkeiten

- Kommunikation der Objekte untereinander
- Brushing
- Freeze
- visuelle Identifikation von
 - Ausreißern
 - auffälligen Beobachtungen
 - Zusammenhängendurch Farben, Symbole und Anklicken

Datendarstellung

- Spreadsheet
 - Anzahl Variablen
 - Anzahl Beobachtungen
- Möglichkeiten
 - editieren
 - sortieren
 - suchen
 - transformieren
 - definieren

KSFE

File Edit Analyze Tables Graphs Curves Vars Options Window Help

BEISPIEL.AIR

9	Label	Int	Int	Int	Int	Int	Int	Int	Int	Int
168	DATETIME	DAY	HOUR	CO	O3	SO2	NO	DUST	WIND	
1	13NOV89:00:00	Mon	0	0.63	0.98	1.073	1.1768	1.489	2.01	
2	13NOV89:01:00	Mon	1	0.63	0.98	0.894	0.5469	1.563	1.62	
3	13NOV89:02:00	Mon	2	0.47	0.73	0.894	0.2930	1.270	2.53	
4	13NOV89:03:00	Mon	3	0.63	0.85	0.858	0.3857	0.879	1.37	
5	13NOV89:04:00	Mon	4	0.31	1.34	0.787	0.1855	0.781	2.54	
6	13NOV89:05:00	Mon	5	0.40	1.10	0.894	0.2393	1.147	1.99	
7	13NOV89:06:00	Mon	6	0.63	1.10	1.037	0.6592	1.636	1.42	
8	13NOV89:07:00	Mon	7	2.22	2.20	2.110	2.4658	2.393	0.96	
9	13NOV89:08:00	Mon	8	5.11	3.42	4.972	5.0391	3.857	1.32	
10	13NOV89:09:00	Mon	9	1.76	1.83	3.290	1.8213	5.225	1.92	
11	13NOV89:10:00	Mon	10	0.82	2.69	2.110	0.8398	4.199	2.45	
12	13NOV89:11:00	Mon	11	0.57	6.47	1.431	0.3369	2.051	2.94	
13	13NOV89:12:00	Mon	12			3	0.3662	1.587	1.40	
14	13NOV89:13:00	Mon	13			0	0.3700	2.051	1.05	
15	13NOV89:14:00	Mon	14			6	0.4150	1.904	0.84	
16	13NOV89:15:00	Mon	15			8	0.5518	1.855	1.23	
17	13NOV89:16:00	Mon	16			3	1.6553	3.369	0.48	
18	13NOV89:17:00	Mon	17			9	3.7012	5.542	0.29	
19	13NOV89:18:00	Mon	18			7	4.8096	7.373	0.98	
20	13NOV89:19:00	Mon	19			7	5.5566	7.642	1.38	
21	13NOV89:20:00	Mon	20			7	4.7168	6.299	1.23	
22	13NOV89:21:00	Mon	21			6	4.6143	5.347	0.57	
23	13NOV89:22:00	Mon	22			2	5.0879	4.785	0.72	
24	13NOV89:23:00	Mon	23			7	4.0186	4.468	1.07	
25	14NOV89:00:00	Tue	0			2	1.9043	4.688	0.98	
26	14NOV89:01:00	Tue	1			3	0.9619	4.395	2.41	
27	14NOV89:02:00	Tue	2	1.28	1.10	0.930	1.4404	3.418	1.04	
28	14NOV89:03:00	Tue	3	1.21	1.34	0.858	1.1865	3.052	1.43	
29	14NOV89:04:00	Tue	4	1.17	1.46	1.073	1.1865	3.320	1.37	
30	14NOV89:05:00	Tue	5	1.44	1.59	1.216	1.0840	3.247	1.82	
31	14NOV89:06:00	Tue	6	2.18	2.32	1.753	2.7002	3.198	0.77	

Find Next
Move to First
Move to Last
Sort...
New Observations
New Variables
Define Variables...
Fill Values...
Extract
Data Options...

Set variable name, label, default role, ...

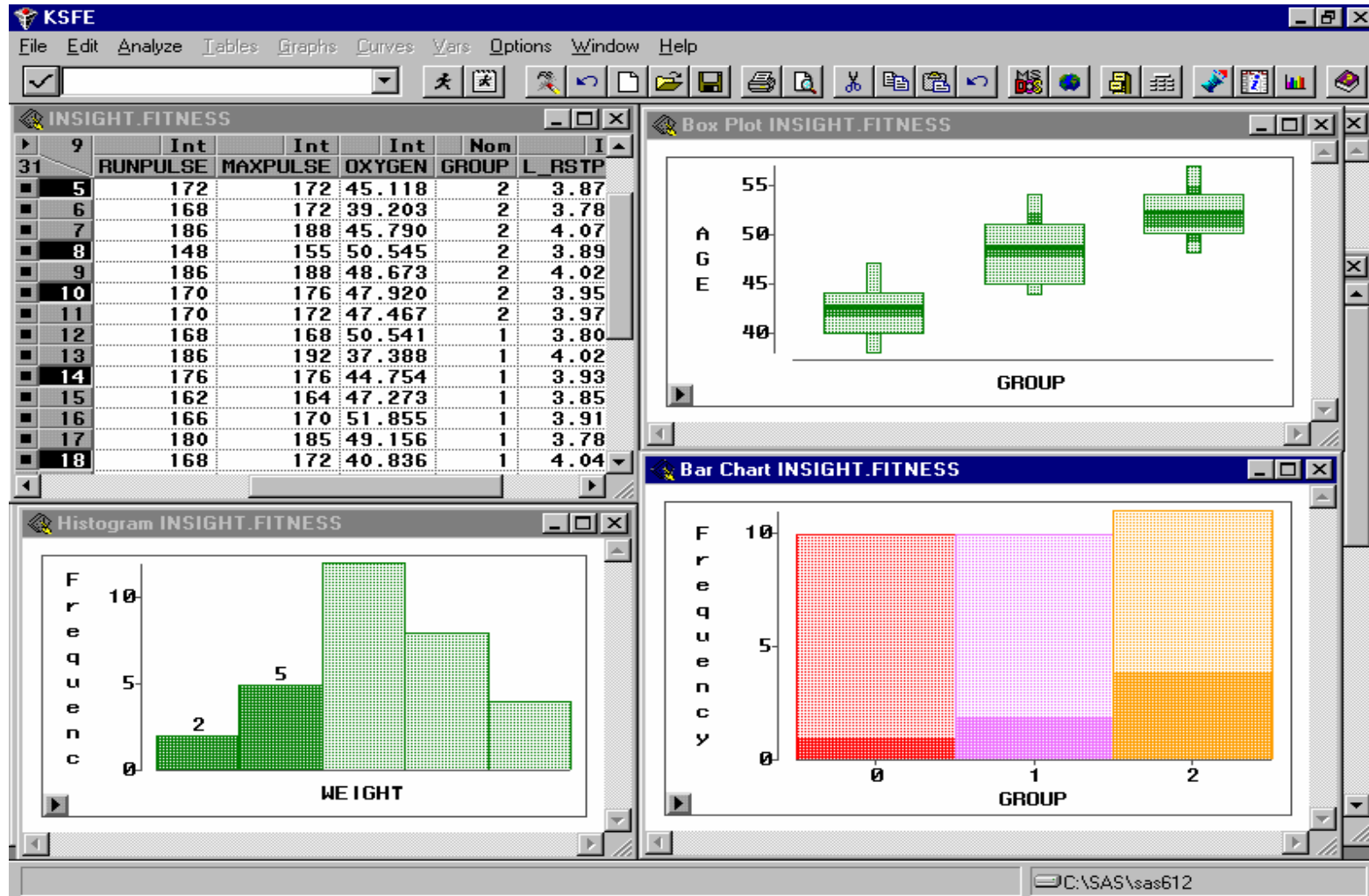
C:\SAS\sas612

grafische Datenvisualisierung

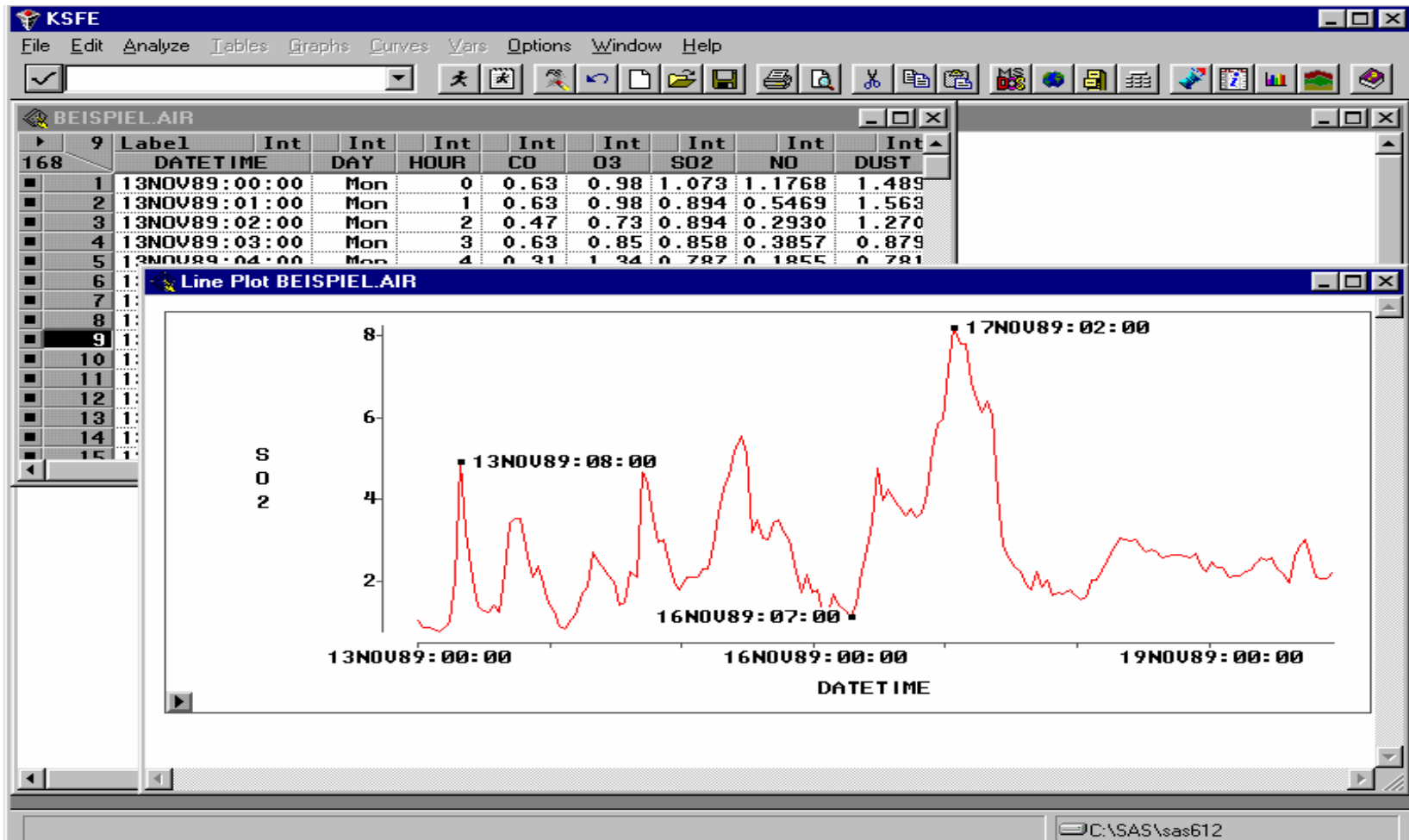
- Histogramme / Balkendiagramme
- Boxplots / Mosaikplots
- Liniendiagramme
- Streuungsdiagramme
- 3-dimensionale Rotating Plots

- Werkzeuge

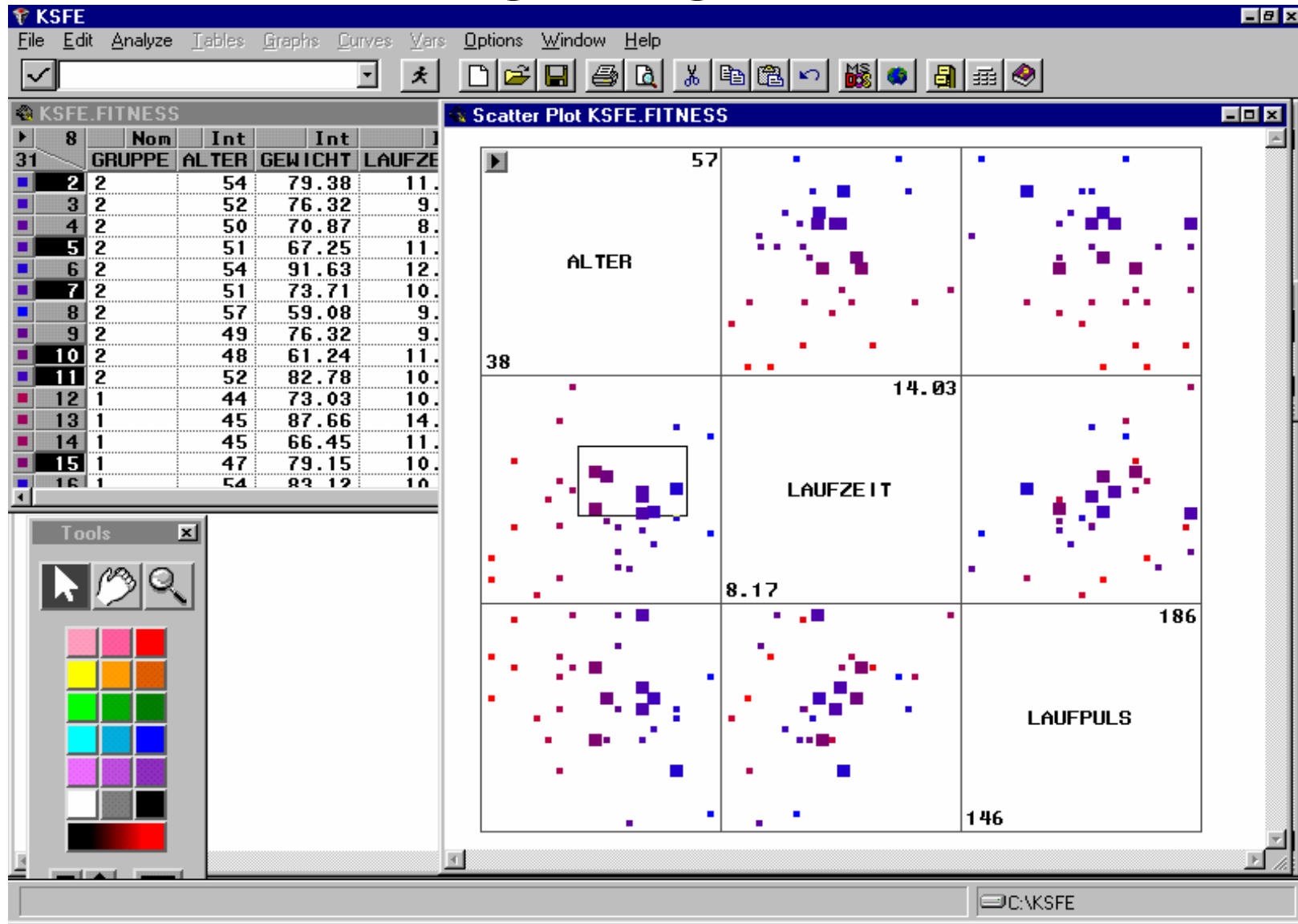
Datenvisualisierung



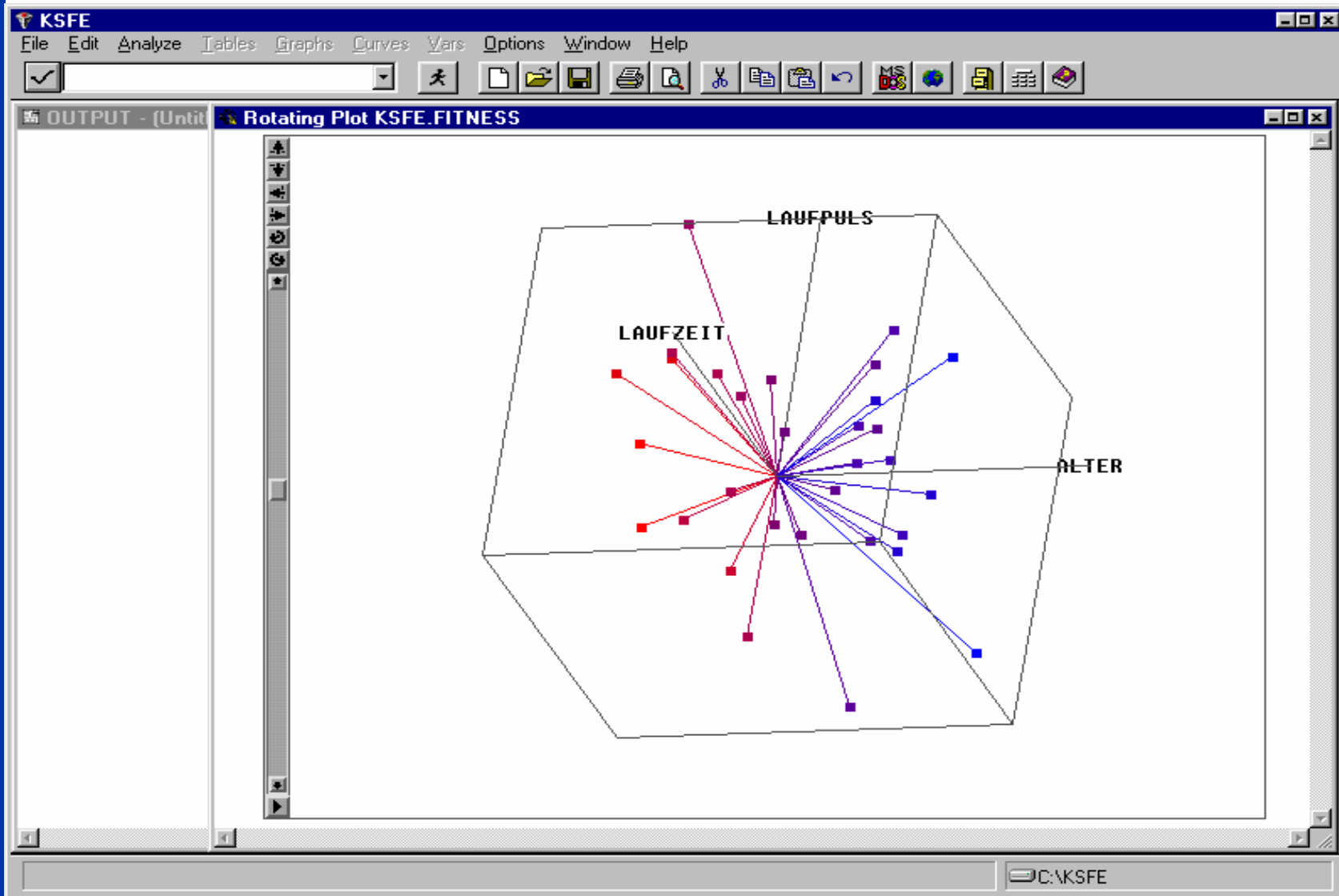
Linendiagramme



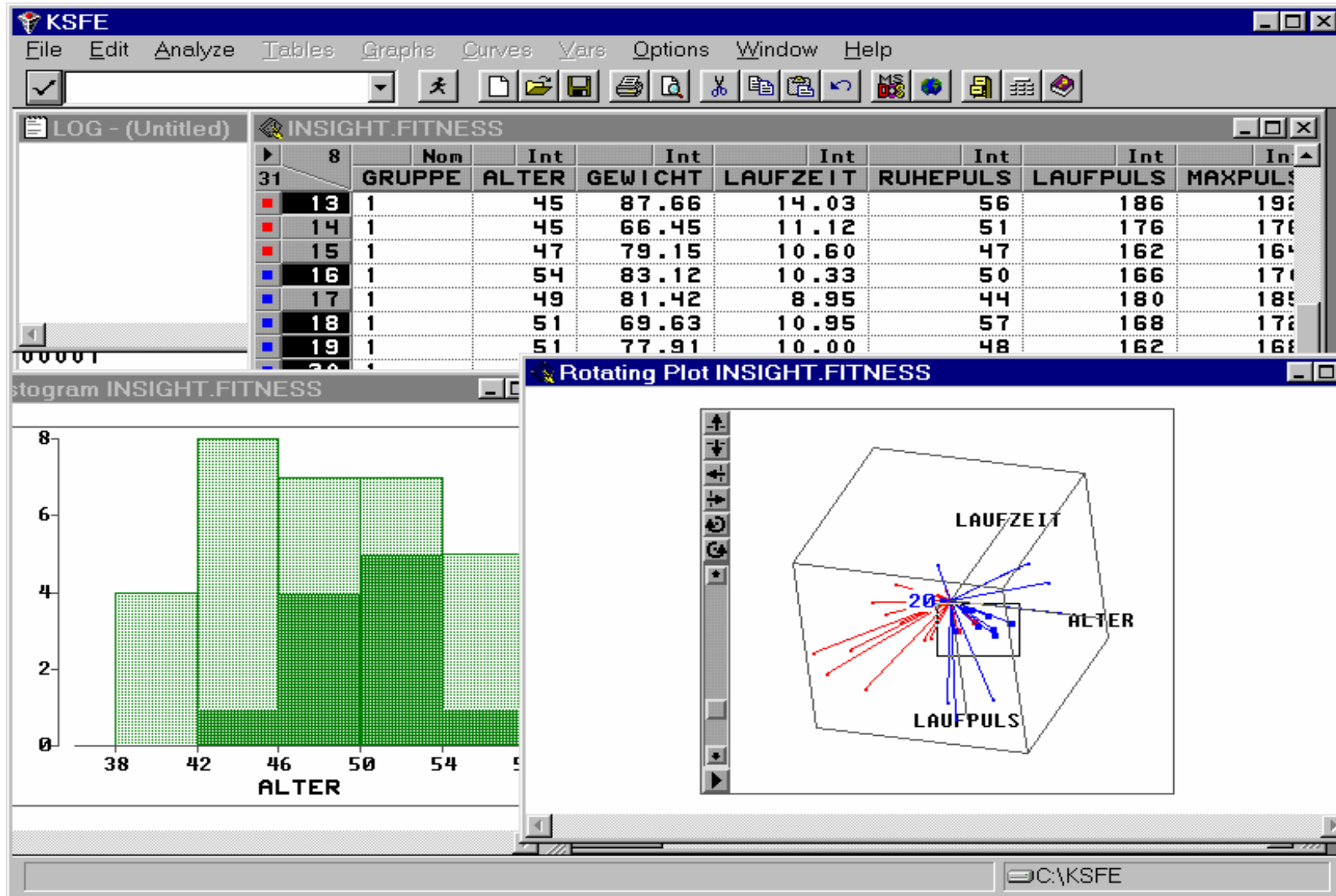
Streuungsdiagramme



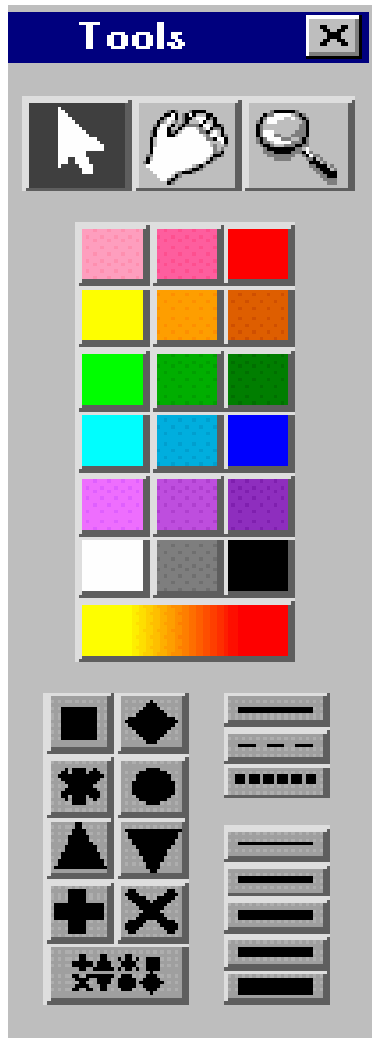
3-D Rotating Plot



Brushing



Das Tools-Fenster



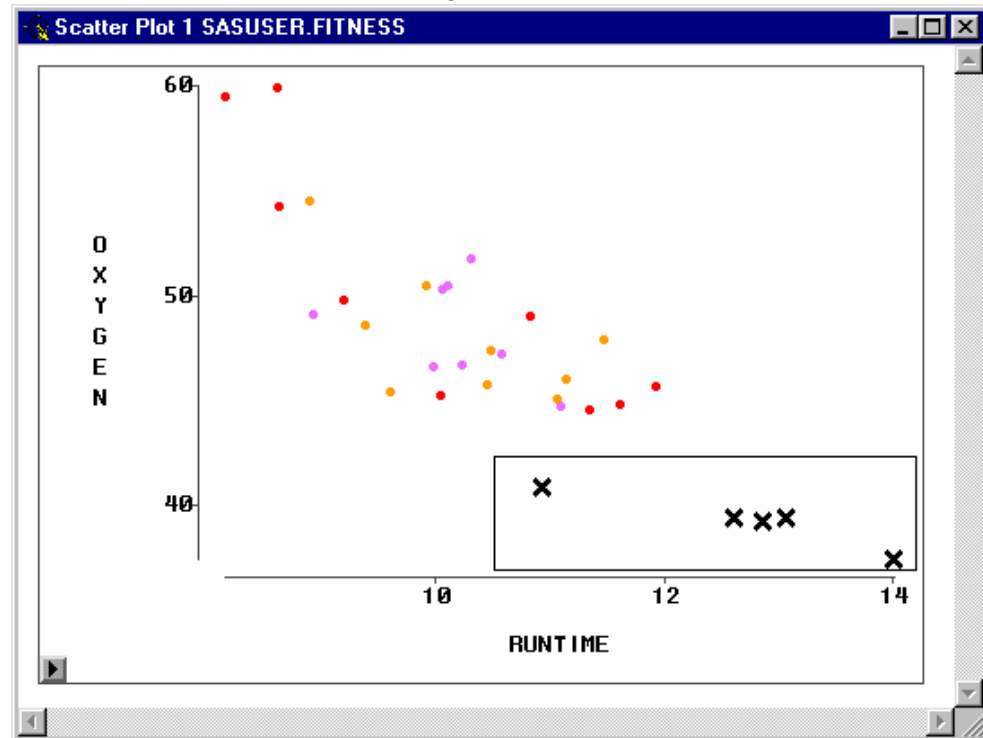
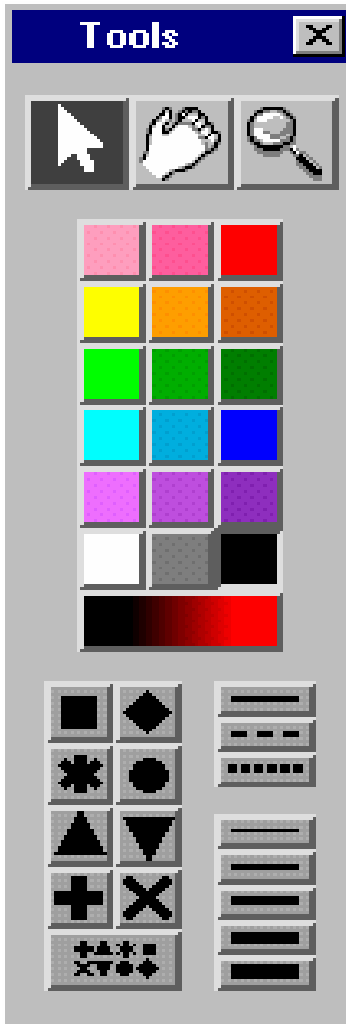
Möglichkeiten zum:

- drehen bestimmter Plots
 - verschieben und
 - vergrößern der Fenster,
- sowie
- anzeigen
 - einfärben
 - markieren

bestimmter Beobachtungen

Das Tools-Fenster

Zum Beispiel:
brush (markiertes Feld)
verschiedene Symbole/Farben



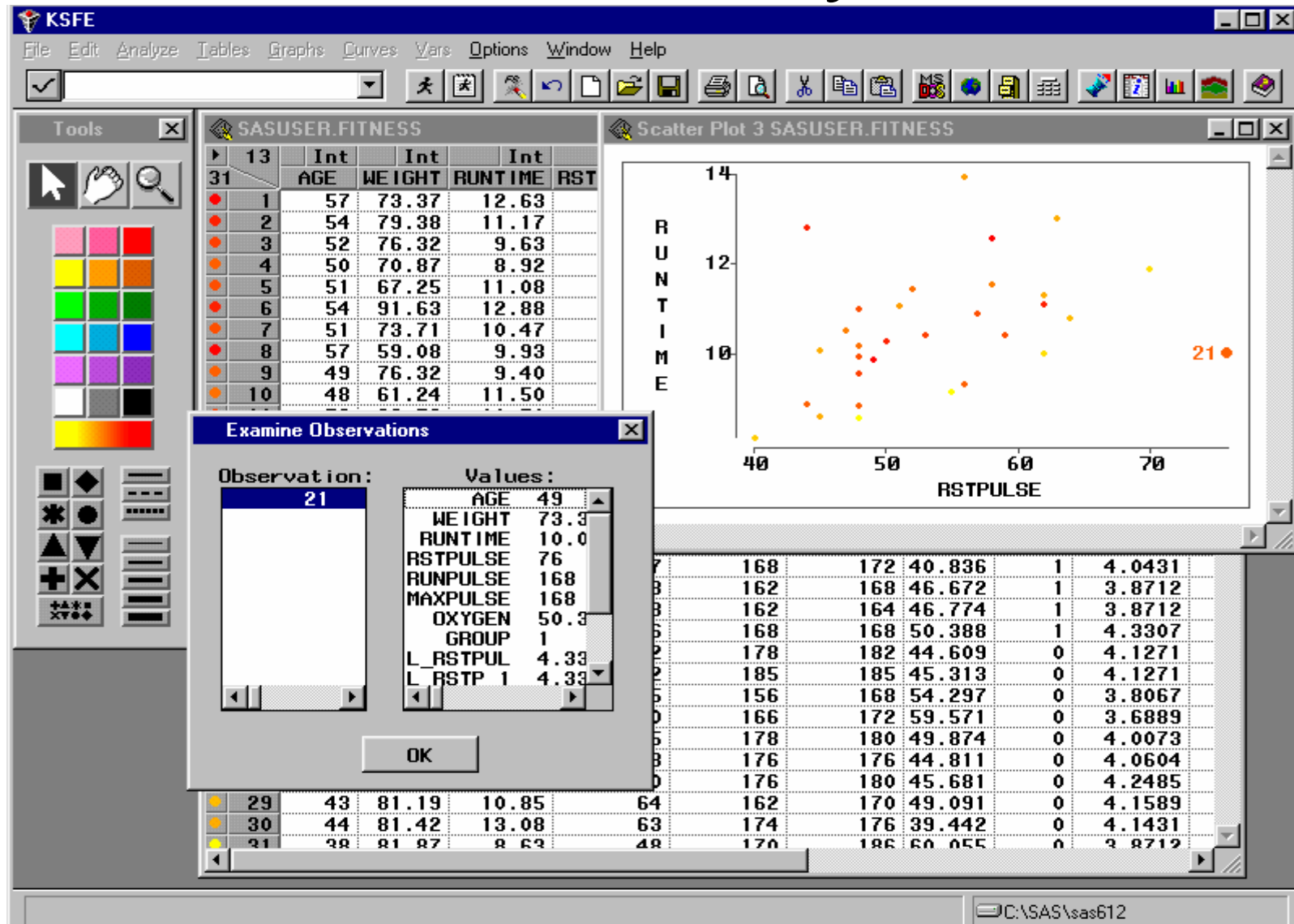
Das Tools-Fenster

The screenshot displays the SAS user interface with the following components:

- Tools Window:** Located on the left, it contains a palette of colored squares and various tool icons (arrow, hand, magnifying glass, etc.). A red arrow points from the Tools window to the data table, and a yellow arrow points from the Tools window to the second scatter plot.
- SASUSER.FITNESS Table:** A data table with columns AGE, WEIGHT, RUNTIME, and RST. The data is as follows:

	AGE	WEIGHT	RUNTIME	RST
1	57	73.37	12.63	
2	54	79.38	11.17	
3	52	76.32	9.63	
4	50	70.87	8.92	
5	51	67.25	11.08	
6	54	91.63	12.88	
7	51	73.71	10.47	
8	57	59.08	9.93	
9	49	76.32	9.40	
10	48	61.24	11.50	
11	52	82.78	10.50	
12	44	73.03	10.13	
13	45	87.66	14.03	
14	45	66.45	11.12	
15	47	79.13	10.60	
16	54	83.12	10.33	
17	49	61.42	8.95	
18	51	69.63	10.95	
19	51	77.91	10.63	
20	48	91.63	10.25	
21	49	73.37	10.08	
22	44	89.47	11.37	
23	40	75.07	10.07	
24	44	85.84	8.65	
25	42	68.15	8.17	
26	38	89.02	9.22	
27	47	77.45	11.63	
28	40	75.98	11.95	
29	43	81.19	10.85	
30	44	81.42	13.08	
31	38	81.87	8.63	
- Scatter Plot SASUSER.FITNESS:** A plot of RUNTIME (y-axis, 10-14) versus RUNPULSE (x-axis, 150-180).
- Scatter Plot 2 SASUSER.FITNESS:** A plot of OXYGEN (y-axis, 40-50) versus AGE (x-axis).

Ausreißeranalyse

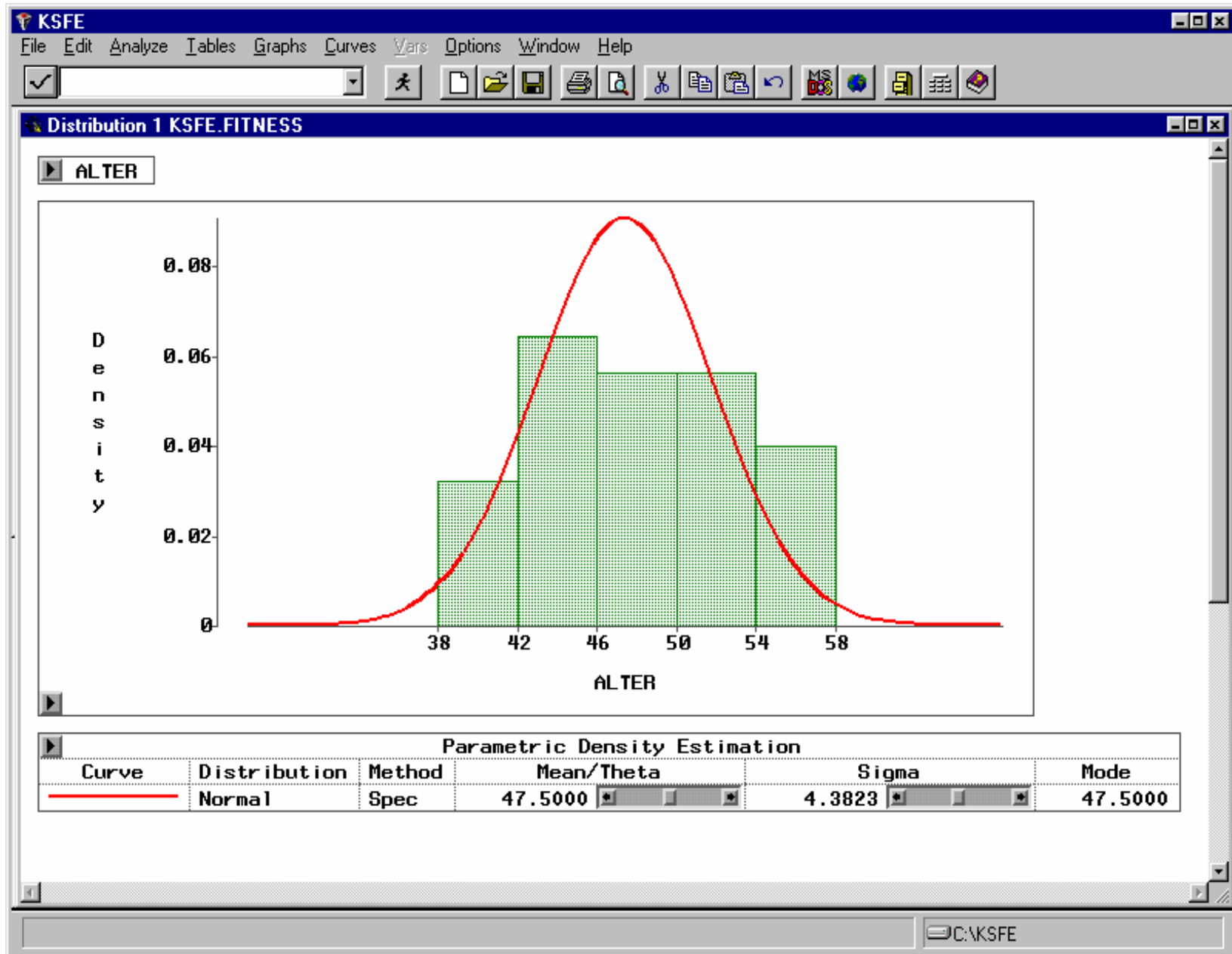


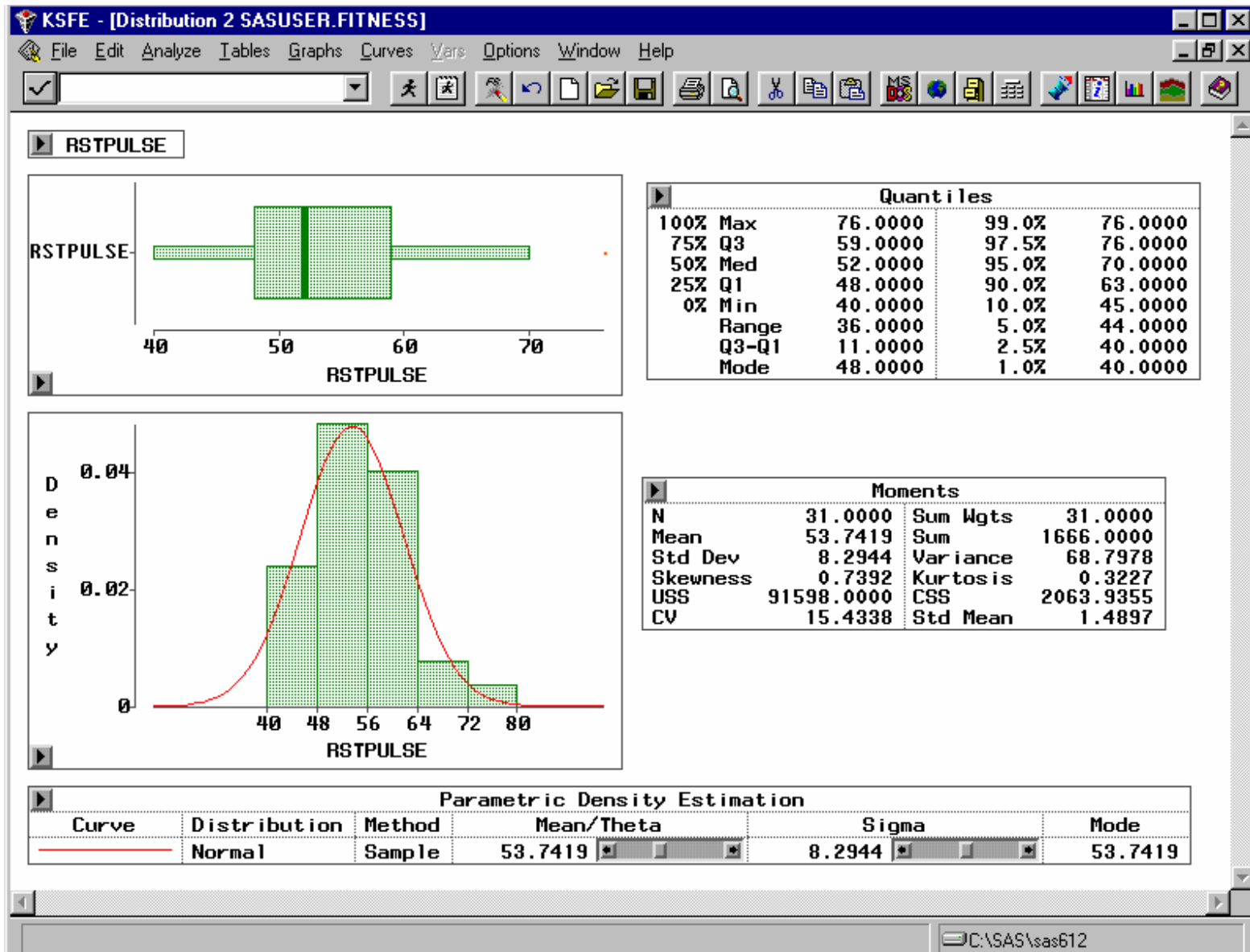
Datenanalyse

Zum Beispiel:

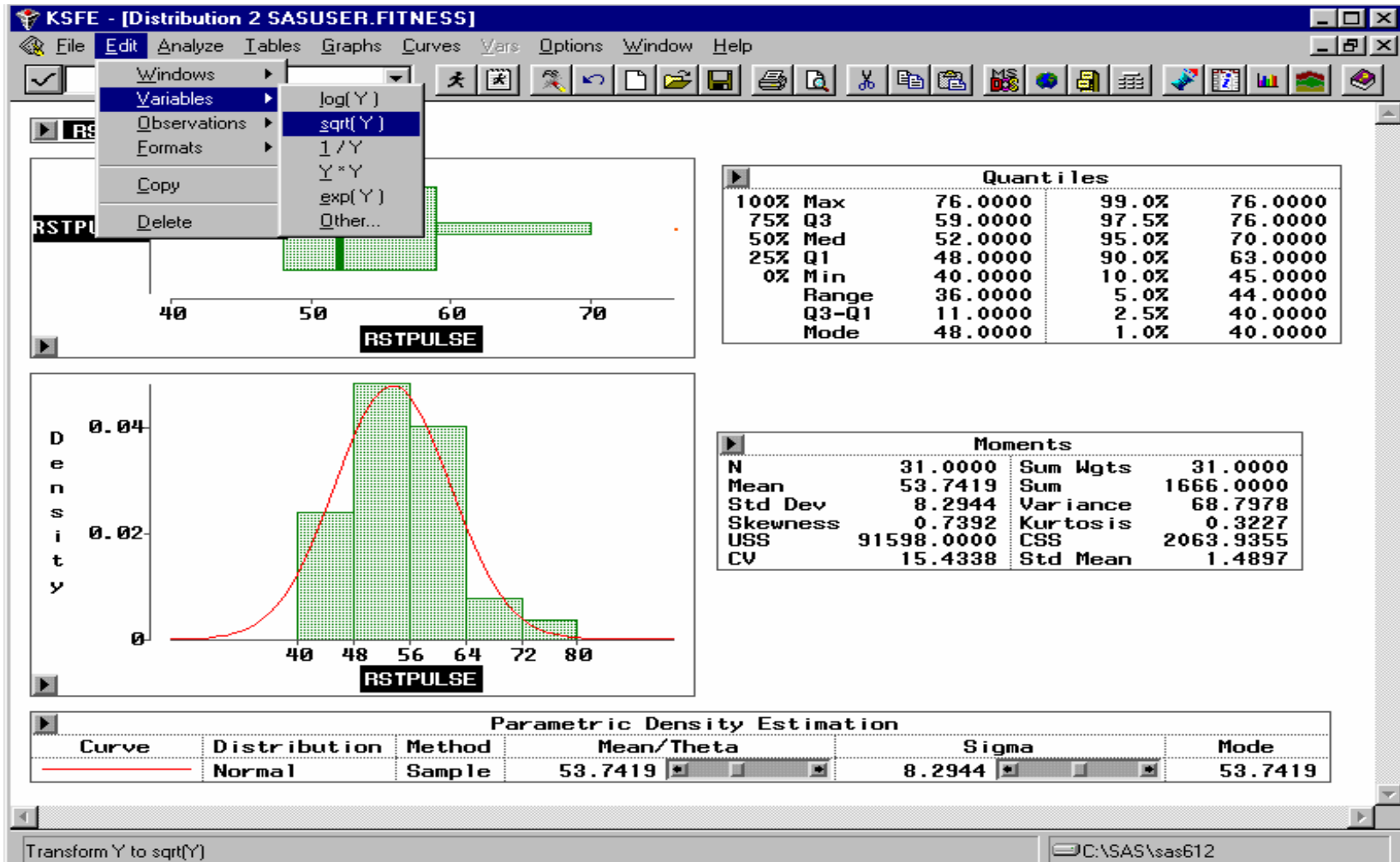
- Univariate Verteilungsanalyse
- Transformation
- Lineare Regressionsanalyse
- Logistische Regression
- Varianzanalyse
- Korrelationsanalyse
- Hauptkomponentenanalyse

Verteilungsanalysen

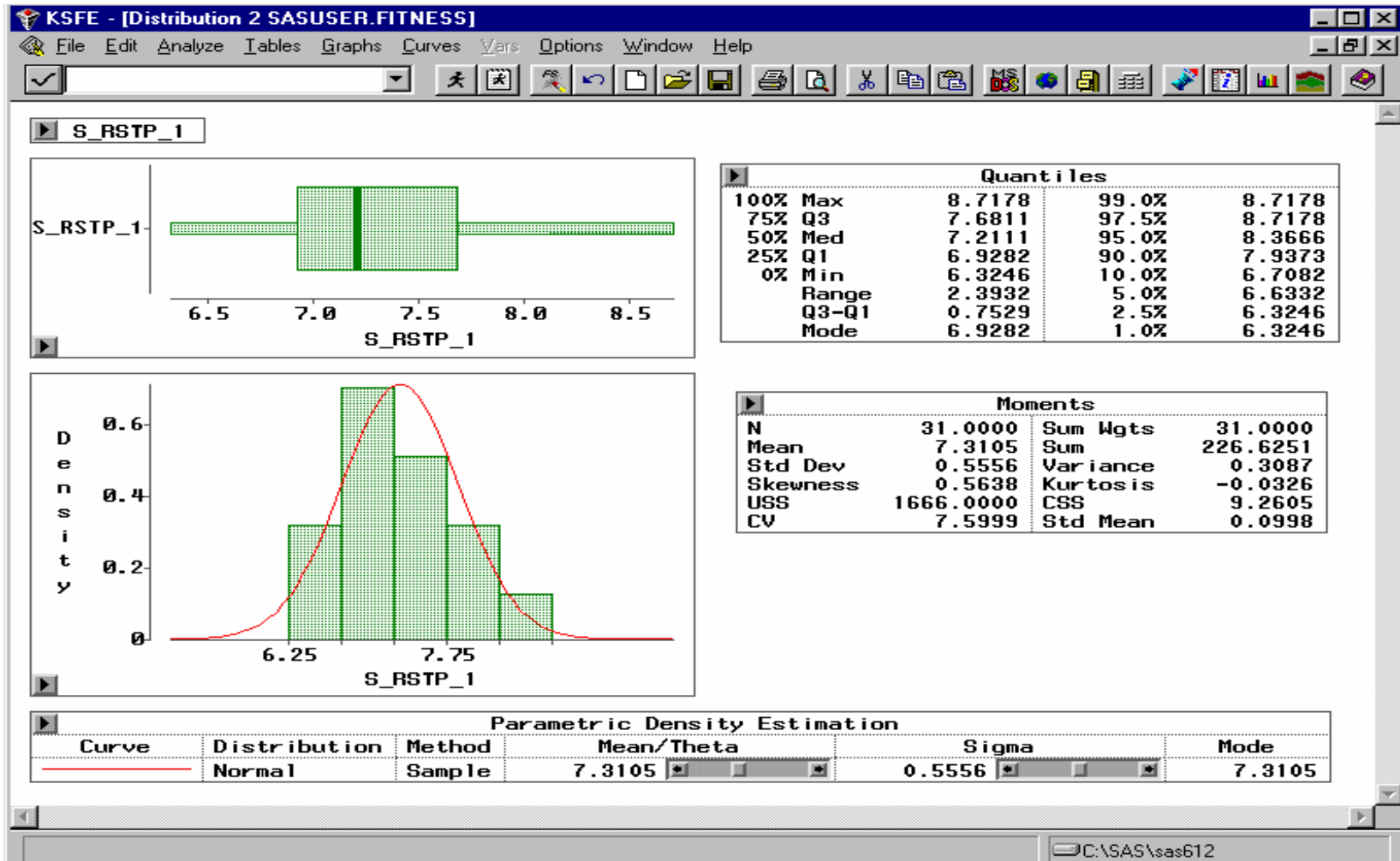




Transformation



Transformation

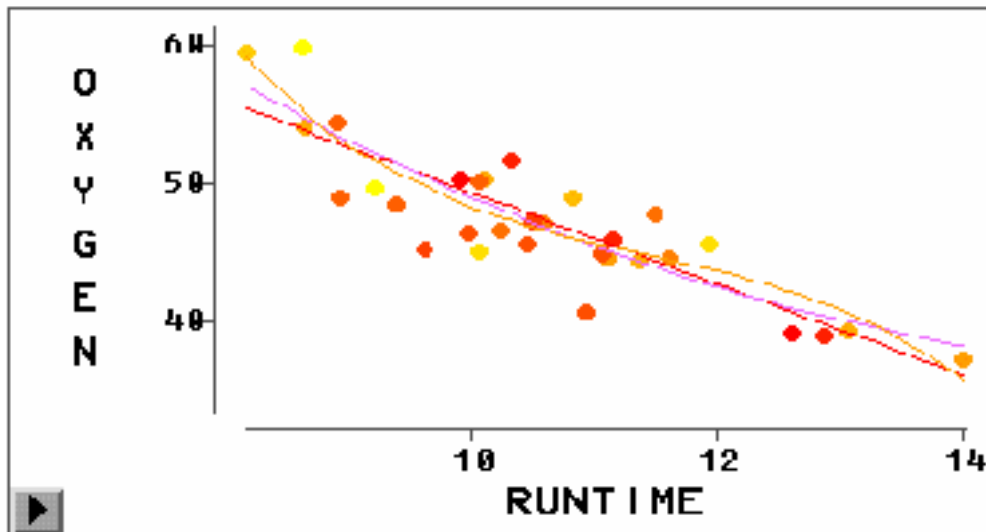





Fit(YX)

- **General-Linear-Model**
 - Regressionsanalysen
 - Varianzanalysen
 - Kovarianzanalysen
- **Generalized-Linear-Model**
 - logistische Regressionsanalysen
 - Poisson Regression
- ...

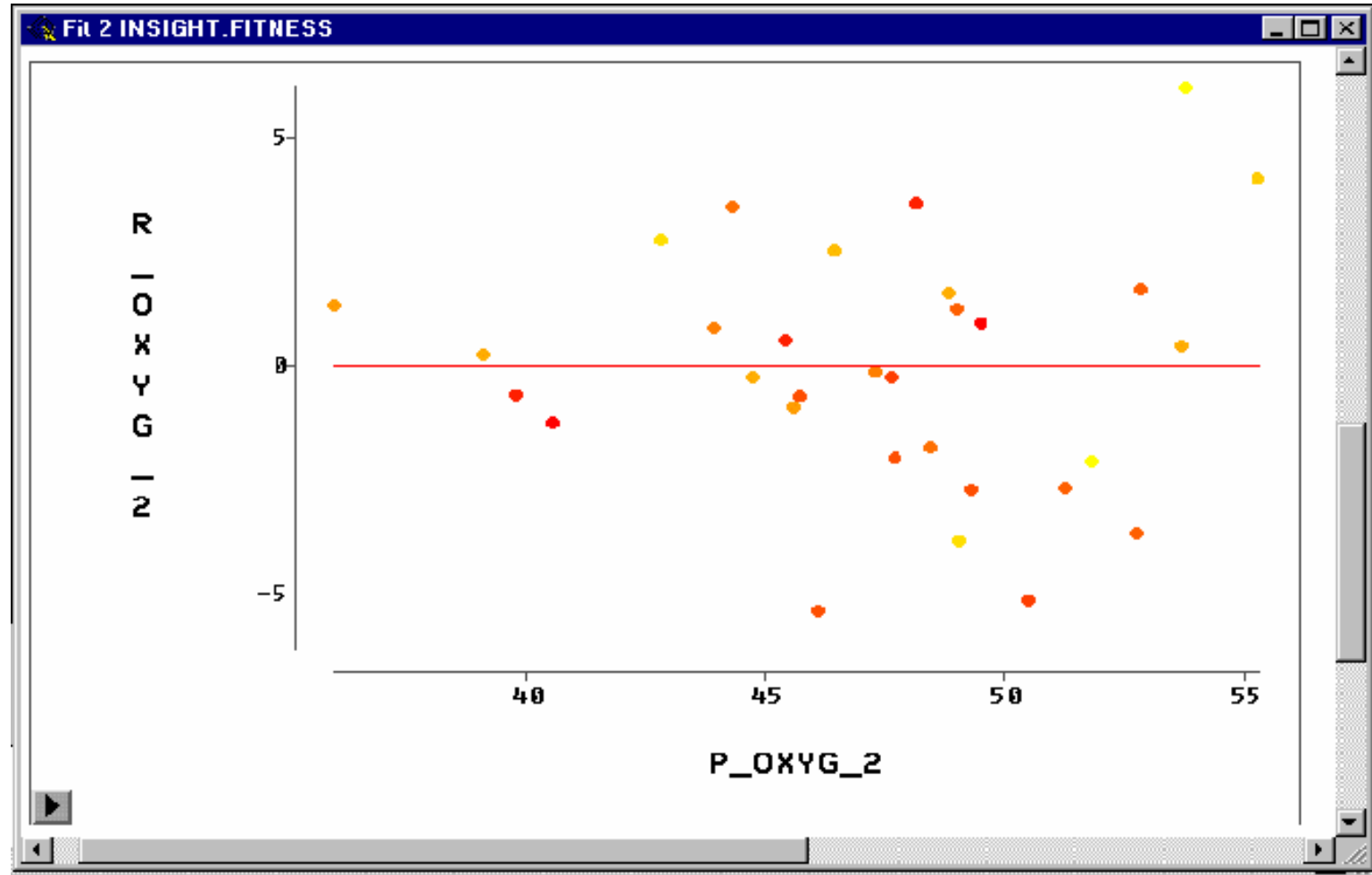
Optische Modellbeurteilung

▶ **Model Equation**
OXYGEN = 82.4218 - 3.3106 RUNTIME

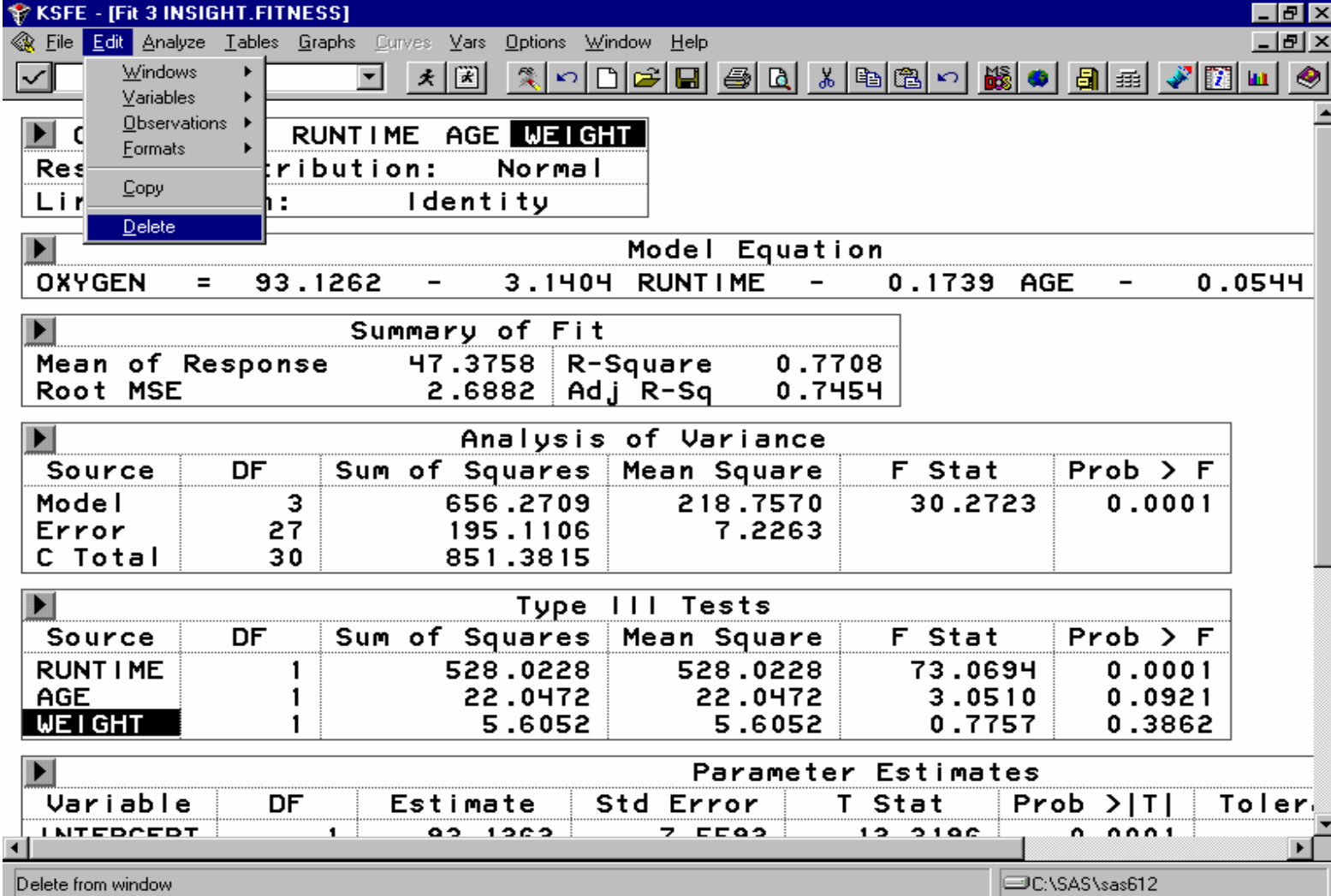


		Parametric Regression			
		Model		Error	
Curve	Degree(Polynomial)	DF	Mean Square	DF	Me
	1 <input type="text" value="1"/>	1	632.9001	29	
	2 <input type="text" value="2"/>	2	323.7232	28	
	3 <input type="text" value="3"/>	3	223.1787	27	

Residuenplots



Statistische Beurteilung



The screenshot displays the SAS software interface with a regression analysis window titled "KSFE - [Fit 3 INSIGHT.FITNESS]". The window shows the following results:

Model Equation

$$\text{OXYGEN} = 93.1262 - 3.1404 \text{ RUNTIME} - 0.1739 \text{ AGE} - 0.0544 \text{ WEIGHT}$$

Summary of Fit

Mean of Response	47.3758	R-Square	0.7708
Root MSE	2.6882	Adj R-Sq	0.7454

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	3	656.2709	218.7570	30.2723	0.0001
Error	27	195.1106	7.2263		
C Total	30	851.3815			

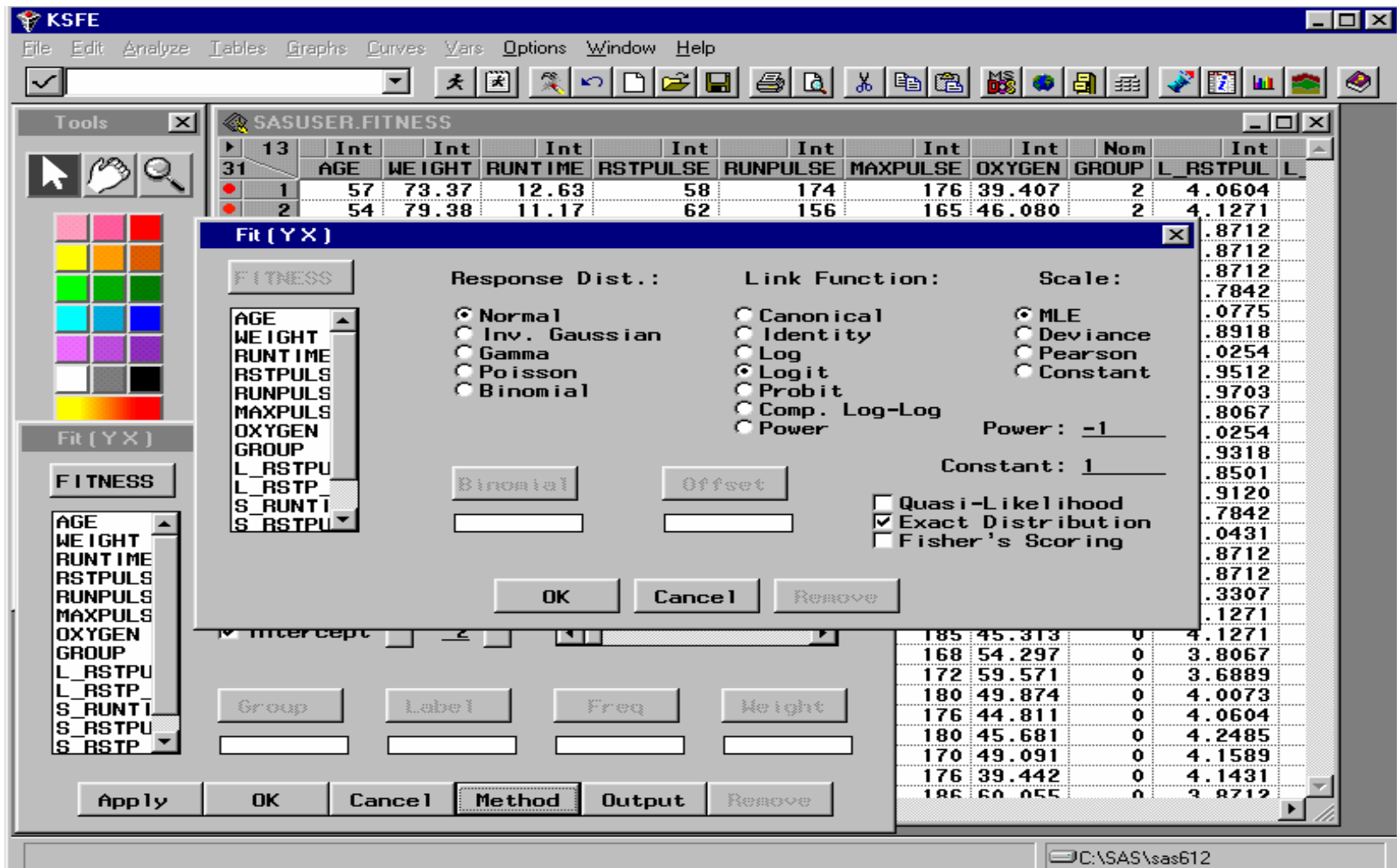
Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
RUNTIME	1	528.0228	528.0228	73.0694	0.0001
AGE	1	22.0472	22.0472	3.0510	0.0921
WEIGHT	1	5.6052	5.6052	0.7757	0.3862

Parameter Estimates

Variable	DF	Estimate	Std Error	T Stat	Prob > T	Toler.
INTERCEPT	1	93.1262	2.5592	36.3906	0.0001	

Große Auswahl an Optionen



The screenshot shows the SAS interface with a data table and a 'Fit (Y X)' dialog box. The data table contains the following information:

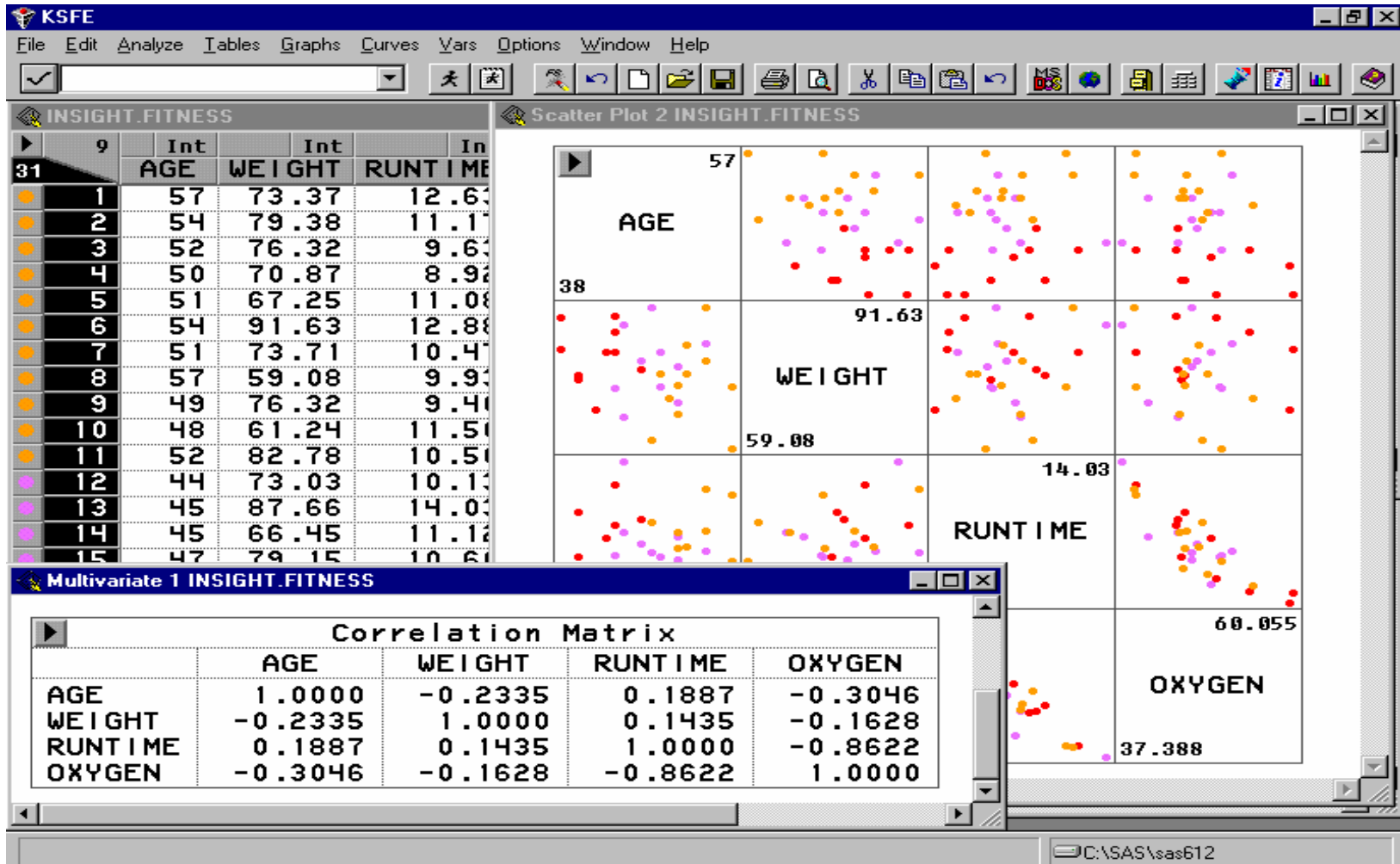
	Int	Int	Int	Int	Int	Int	Int	Int	Nom	Int
	AGE	WEIGHT	RUNTIME	RSTPULSE	RUNPULSE	MAXPULSE	OXYGEN	GROUP	L_RSTPUL	L
1	57	73.37	12.63	58	174	176	39.407	2	4.0604	
2	54	79.38	11.17	62	156	165	46.080	2	4.1271	

The 'Fit (Y X)' dialog box is open, showing the following options:

- Response Dist.:**
 - Normal
 - Inv. Gaussian
 - Gamma
 - Poisson
 - Binomial
- Link Function:**
 - Canonical
 - Identity
 - Log
 - Logit
 - Probit
 - Comp. Log-Log
 - Power
- Scale:**
 - MLE
 - Deviance
 - Pearson
 - Constant

Additional options in the dialog include 'Power: -1', 'Constant: 1', 'Quasi-Likelihood', 'Exact Distribution' (checked), and 'Fisher's Scoring'. Buttons for 'OK', 'Cancel', and 'Remove' are visible at the bottom of the dialog.

Korrelationsanalyse



Statements

```
proc insight data=sasuser.fitness;
```

```
SCATTER AGE WEIGHT RUNTIME RSTPULSE *  
        AGE WEIGHT RUNTIME RSTPULSE;
```

```
MULT RUNPULSE MAXPULSE OXYGEN;
```

```
BAR AGE WEIGHT MAXPULSE OXYGEN;
```

```
FIT AGE = WEIGHT RUNTIME RSTPULSE;
```

```
run;
```

Verfügbarkeit der SAS/INSIGHT Software

- verfügbar auf allen Plattformen
- Empfehlung:
 - Workstation oder schneller PC
 - **großer Bildschirm**

grafische Datenauswertung

- intuitiver Weg
- Alternative zur klassischen Datenanalyse
- kurz zusammengefaßt:
das SAS System bietet auch hier
alle Möglichkeiten

wichtige Informationen

- **SAS/Insight User's Guide**
Version 6 , 3rd Edition , 1996 , #55582
Anwendungs- und Nachschlagewerk
- **DISK '93 Proceedings**
Heinz-Bernd Spieker: Neuerungen in SAS/Insight Software
- **SEUGI Proceedings '94**
Gerhard Held: Dynamic Data Analysis with SAS/Insight
Software p 784
- **Online Help System**
- **16./17.4.98: Insight-Schulung bei SAS-Institute**

Fragen ?

