

Mo, 18.3.2013

Frühjahrstagung des VDSt in Wolfsburg

Ökologische Korrelation und Regression mit Excel und PSPP



Wolfsburg

Hartmut Bömermann Regional- und Kommunalstatistik

Themen

- » Ökologische Variablen
- » Korrelation und Regression mit Excel
- » Korrelation und Regression mit PSPP



Daten

- » Ausgangspunkt ist die Datenmatrix mit Anteilswerten bzw. Quotienten
- » Es gibt **keine absoluten Häufigkeiten**. Könnte das ein Problem sein?
- » Keine der Variablen bezieht sich auf die **Wahlberechtigten** (das Elektorat)
- » Die meisten Variablen basieren auf Individualeigenschaften
- » Variable *pAusl* ist eine reine Kontextvariable
- » Variable *mHHGr* ist keine Individualeigenschaft
- » Ziel:
 Berechnung **ökologischer Korrelationen** und
ökologischer Regressionen mit Excel

SBNR	SBNNAME	pWbt	pU18	p18U30	p30U45	p45U65	p65+	pAlleinErz	mHHGr	pRK	pEV	pX	pAlo18U65	pAusl	pEingeb	pAussied	pMH	pOMH	pBG_HH	pSGB2Quote
111	Volkshochschule F	66,9	8,9	21,6	28,3	25,0	16,2	2,6	172,8	44,1	15,6	40,2	4,7	17,2	5,2	10,3	32,7	67,3	3,6	3,3
121	Christoph-Scheine	71,6	9,3	27,8	28,3	20,7	13,8	2,8	157,0	42,9	14,8	42,3	4,7	19,8	5,4	5,8	31,0	69,0	6,1	5,3
122	Christoph-Scheine	69,0	7,0	34,1	29,0	19,2	10,7	2,0	148,7	49,0	11,6	39,4	2,1	22,1	5,9	4,7	32,7	67,3	3,2	2,6
131	Gnadenthal-Schule	78,8	8,0	22,3	22,4	20,3	27,0	3,6	213,7	48,5	17,6	33,9	3,5	18,6	4,0	11,3	33,9	66,1	9,9	6,3
132	Gnadenthal-Schule	52,7	7,2	27,2	21,9	15,8	28,0	4,2	228,3	50,1	17,3	32,6	4,6	20,2	5,8	11,3	37,3	62,7	9,9	6,5
141	Volksschule Auf de	80,2	15,1	10,8	18,5	32,3	23,2	2,0	211,1	59,0	14,8	26,2	2,1	2,8	3,7	4,7	11,2	88,8	2,3	1,3
142	Volksschule Auf de	77,6	14,5	13,5	24,0	23,0	25,0	4,0	193,9	56,8	15,4	27,8	2,5	4,5	4,5	7,5	16,5	83,5	1,7	1,4
143	Volksschule Auf de	78,0	18,1	15,1	20,2	30,2	16,4	7,7	230,1	57,6	18,9	23,5	2,0	5,1	3,9	8,6	17,6	82,4	1,2	1,1
144	Volksschule Auf de	62,5	9,6	17,2	20,4	25,1	27,6	1,9	208,9	51,0	19,7	29,3	1,8	9,4	5,9	8,7	24,0	76,0	3,4	1,8
145	Volksschule Auf de	71,7	15,7	18,3	24,4	28,1	13,5	5,2	199,2	49,2	17,8	33,0	3,1	7,8	5,1	11,3	24,1	75,9	3,9	3,5

Deskriptive Maße

pWbt				
A	B	C	pL	
1	SBNR	SBNAME		
2	111	Volkshochschule F	66,9	
3	121	Christoph-Scheine	71,6	
4	122	Christoph-Scheine	69,0	
5	131	Gnadenthal-Schule	78,8	
6	132	Gnadenthal-Schule	52,7	
7	141	Volksschule Auf de	80,2	1
8	142	Volksschule Auf de	77,6	1
9	143	Volksschule Auf de	78,0	1
10	144	Volksschule Auf de	62,5	

In Excel:
benannte Bereiche
(„Namen definieren“)

In Formel benannte Bereiche wie Variablen ansprechen:

=Min(pWbt)

=Max(pWbt)

=Mittelwert(pWbt)

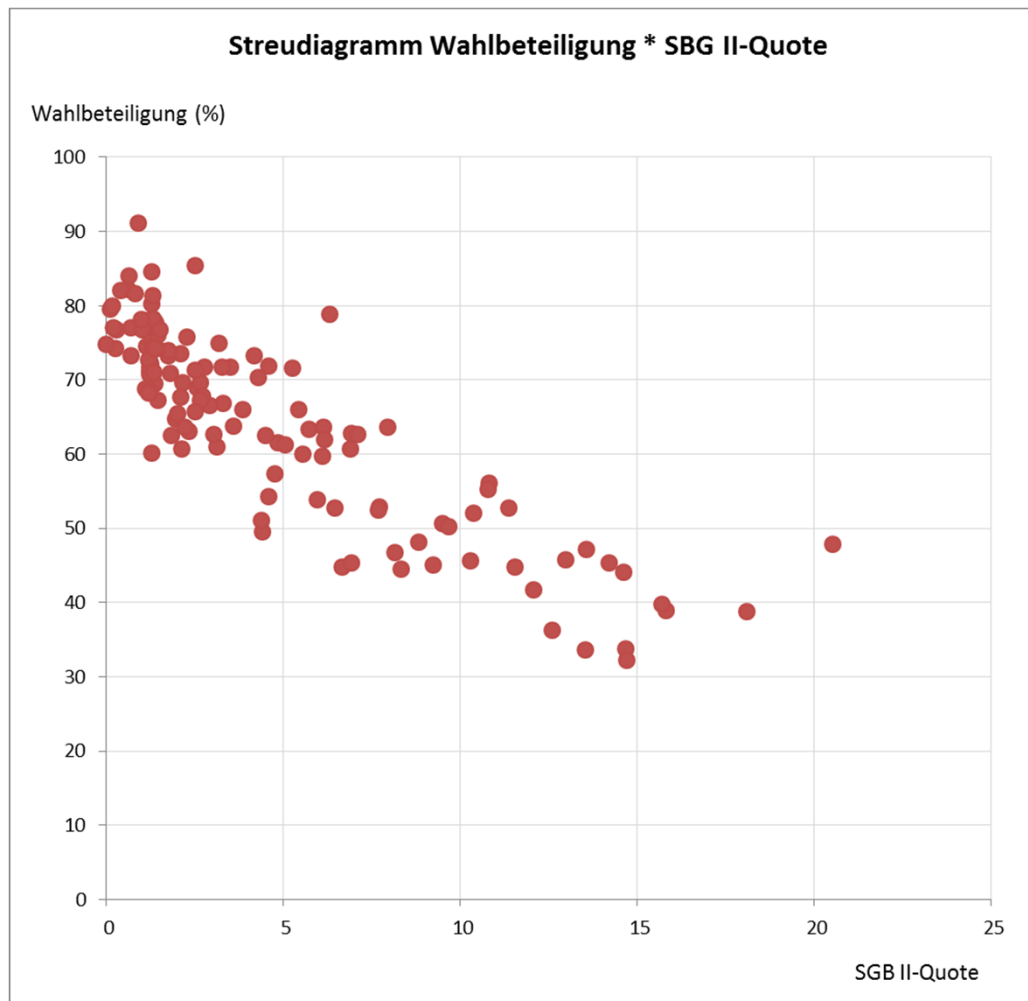
=StAbw.S(pWbt)

Variable	Min	Max	Mittelw	Standardabw
pWbt	32,20	110,10	64,89	13,57
pU18	0,00	0,00	17,07	4,07
p18U30	9,06	34,11	15,68	3,79
p30U45	15,64	32,78	22,21	3,21
p45U65	15,76	38,68	26,22	3,34
p65+	6,73	31,80	18,82	5,25
pAlleinErz	0,93	10,34	4,62	1,67
mHHGr	1,49	2,85	2,20	0,28
pRK	22,32	81,27	52,06	11,09
pEV	7,09	31,76	18,27	4,23
pX	11,65	62,03	29,68	9,76
pAlo18U65	0,31	10,89	3,84	2,48
pAusl	0,87	42,04	11,97	8,85
pEingeb	1,56	13,58	4,95	2,02
pAussied	1,74	37,85	15,69	8,23
pMH	5,43	80,61	32,61	17,32
pOMH	19,39	94,57	67,39	17,32
pBG_HH	0,00	19,52	5,18	4,61
pSGB2Quote	0,00	20,53	4,68	4,50

Problem:

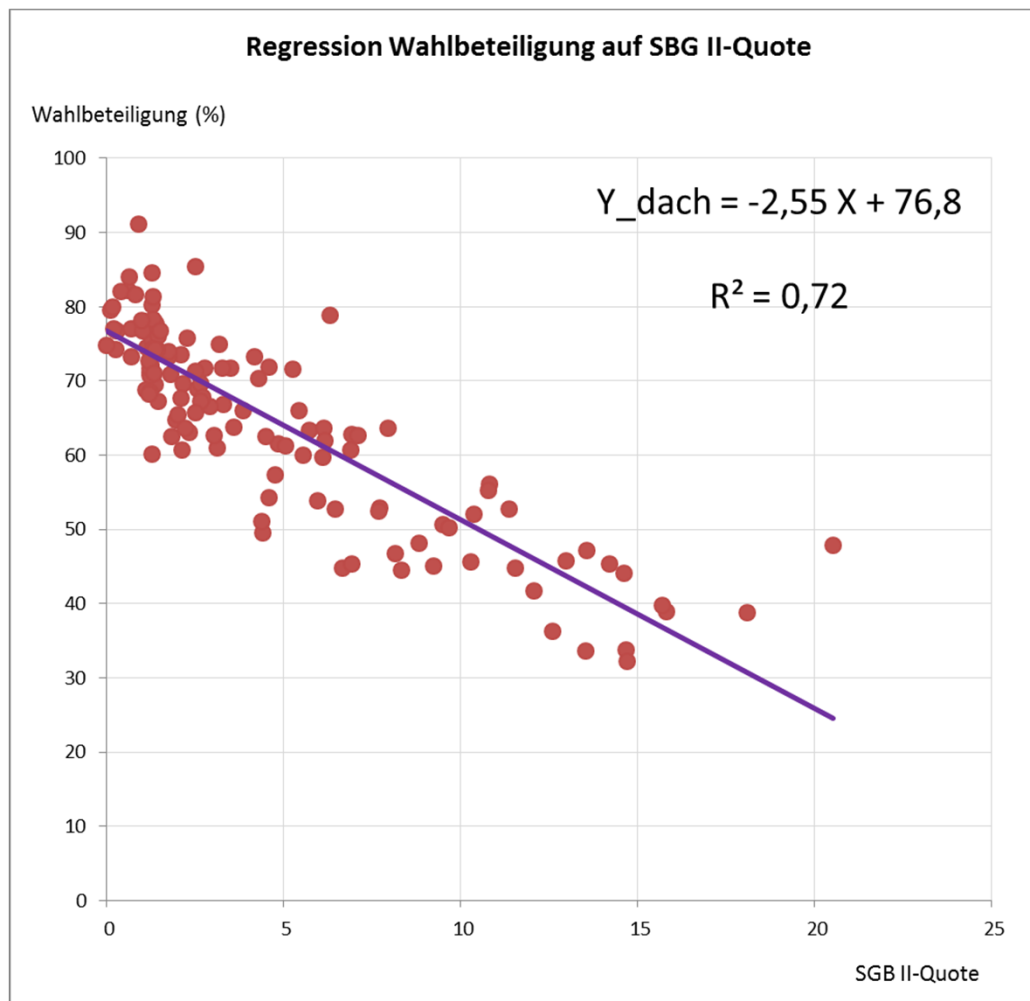
- » Kann aus Anteilswerten der Aggregate der richtige Gesamtwert errechnet werden?
- » Alle Aggregate gehen mit dem **Gewicht 1** in die Berechnung ein.
- » Anteil muss/müsste mit **Population** gewichtet werden.

Streudiagramm



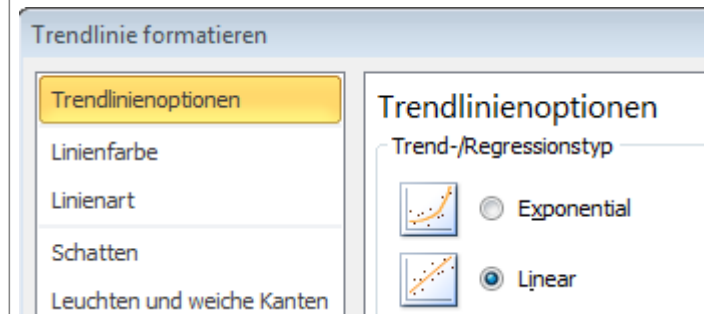
- » Streudiagramm der Anteilswerte bzw. Quote
- » Punkte sind gleich groß, da Gewicht = 1
- » Alternativ: Blasendiagramm, Blasenfläche abhängig von Population

Streudiagramm mit Regressionsgerade

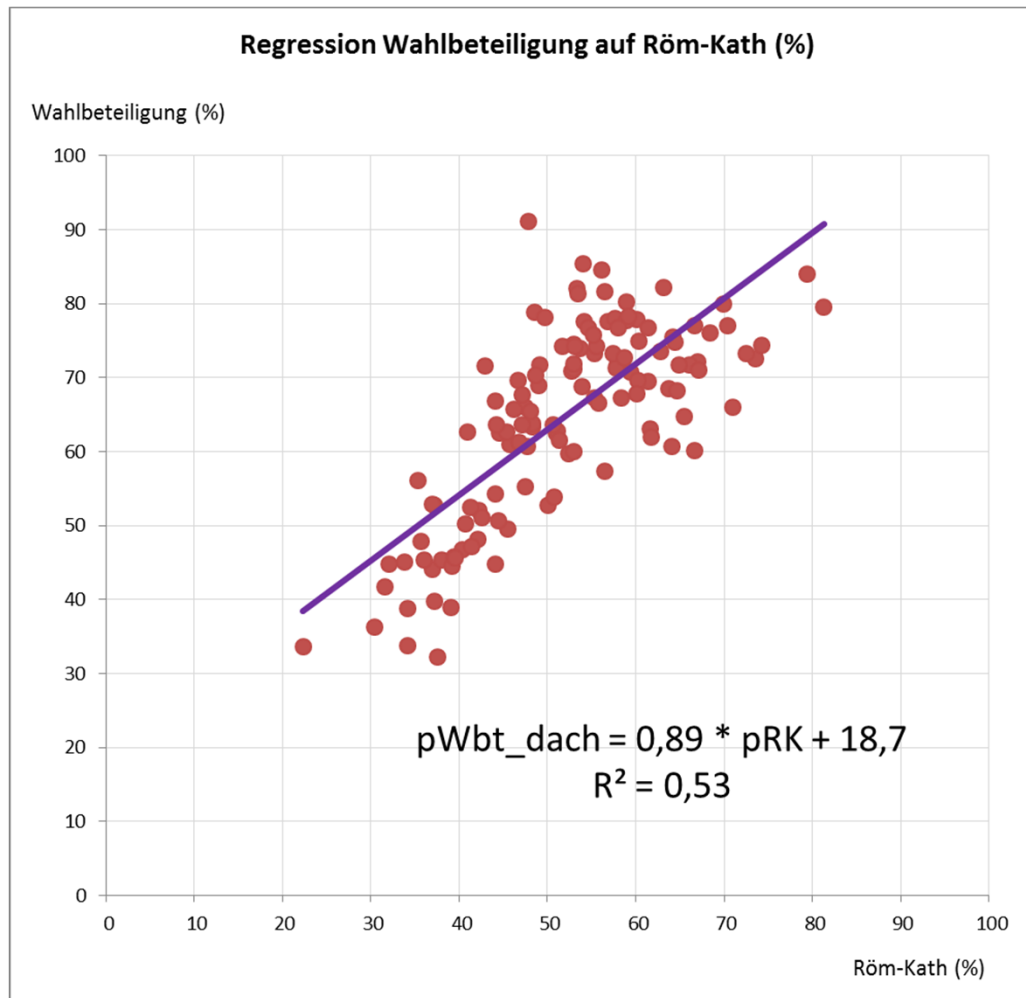


- » Streudiagramm mit Regressionsgerade
- » Varianzerklärung 72%
- » Je 2,6 %-Punkte SGBII-Anteil sinkt die Wahlbeteiligung um 1-%-Punkt.

Kontextmenü „Trendlinie hinzufügen“:



Streudiagramm mit Regressionsgerade



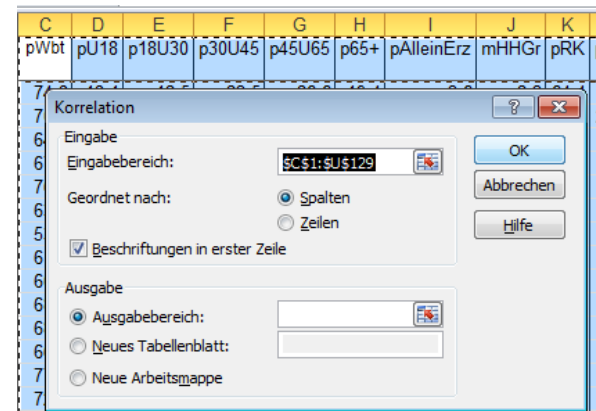
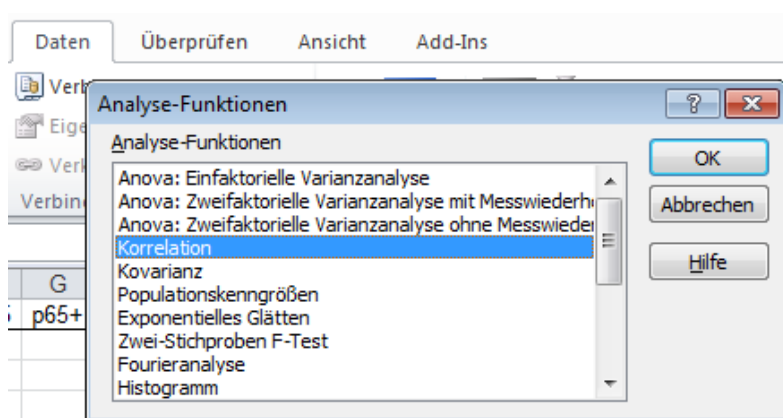
- » Streudiagramm mit Regressionsgerade
- » Varianzerklärung 53 %
- » Je 1 %-Punkte RK-Anteil steigt die Wahlbeteiligung um 0,89-%-Punkte.

Korrelationsmatrix

- » Berechnung der Korrelationsmatrix mit dem Add-In „Analyse-Funktionen“ (ungewichtet)

	pWbt	pU18	p18U30	p30U45	p45U65	p65+	pAlleinErz	mHHGr	pRK	pEV	pX	pAlo18U65	pAusl	pEingeb	pAussied	pMH	pOMH	pBG_HH
pWbt	1,00																	
pU18	-0,04	1,00																
p18U30	-0,36	-0,43	1,00															
p30U45	0,15	0,08	0,34	1,00														
p45U65	0,39	-0,02	-0,50	-0,42	1,00													
p65+	-0,05	-0,50	-0,29	-0,65	-0,01	1,00												
pAlleinErz	-0,41	0,35	0,13	-0,11	-0,27	-0,12	1,00											
mHHGr	0,16	0,78	-0,46	-0,10	0,15	-0,31	0,11	1,00										
pRK	0,73	0,07	-0,48	0,03	0,40	0,02	-0,37	0,36	1,00									
pEV	-0,12	-0,03	0,01	-0,09	0,09	0,01	0,12	-0,09	-0,49	1,00								
pX	-0,77	-0,07	0,54	0,01	-0,49	-0,03	0,36	-0,36	-0,93	0,12	1,00							
pAlo18U65	-0,83	0,01	0,31	-0,23	-0,35	0,14	0,49	-0,25	-0,78	0,14	0,82	1,00						
pAusl	-0,83	-0,01	0,50	-0,07	-0,49	0,00	0,35	-0,26	-0,85	0,07	0,94	0,86	1,00					
pEingeb	-0,74	0,09	0,32	-0,18	-0,43	0,08	0,41	-0,16	-0,80	0,06	0,88	0,79	0,85	1,00				
pAussied	-0,82	0,06	0,19	-0,22	-0,18	0,07	0,33	-0,10	-0,72	0,40	0,65	0,70	0,68	0,59	1,00			
pMH	-0,90	0,03	0,38	-0,16	-0,38	0,04	0,38	-0,20	-0,87	0,23	0,89	0,87	0,93	0,83	0,89	1,00		
pOMH	0,90	-0,03	-0,38	0,16	0,38	-0,04	-0,38	0,20	0,87	-0,23	-0,89	-0,87	-0,93	-0,83	-0,89	-1,00	1,00	
pBG_HH	-0,86	0,08	0,38	-0,17	-0,43	0,05	0,53	-0,14	-0,77	0,14	0,82	0,94	0,87	0,78	0,73	0,88	-0,88	1,00
pSGB2Quote	-0,85	0,12	0,36	-0,14	-0,44	0,02	0,56	-0,15	-0,77	0,14	0,82	0,94	0,87	0,77	0,72	0,88	-0,88	0,98

Korrelationskoeffizienten ≥ 0.7 fett wegen $r * r = R^2$. $0,7^2 \approx 50\%$ Varianzerklärung



Multiple Regression

- » Berechnung der Regression mit dem Add-In „Analyse-Funktionen“

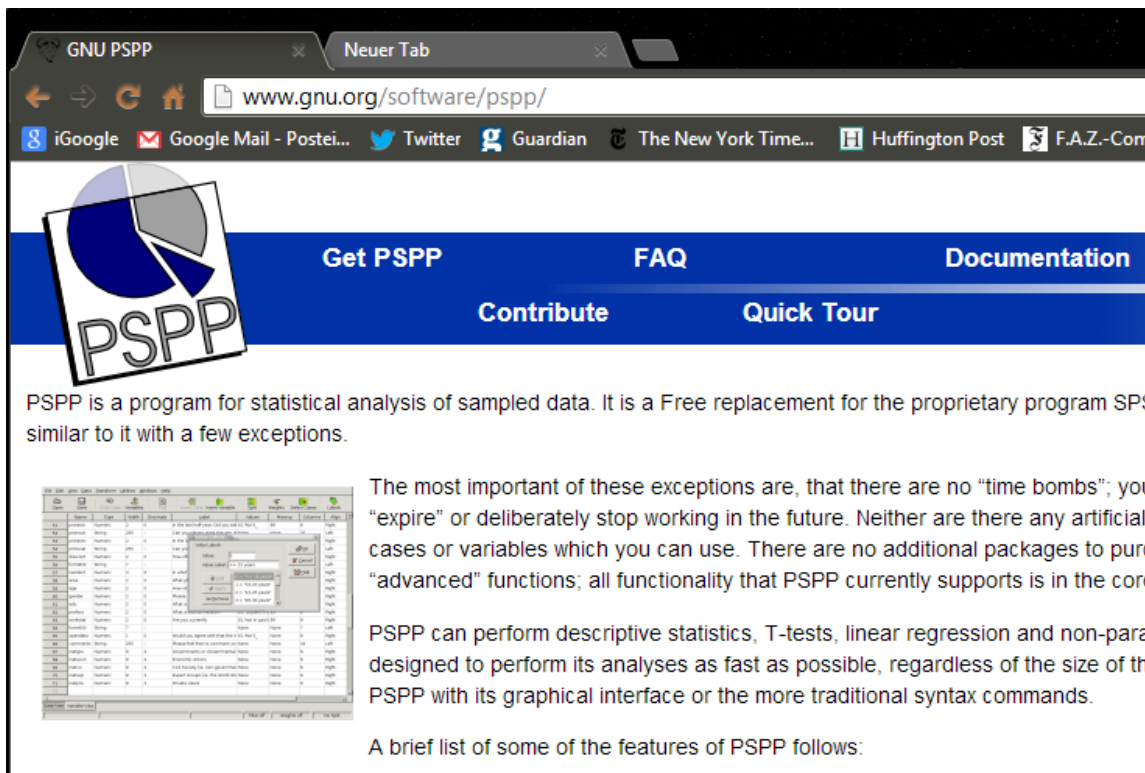
	A	B	C	D	E	F	G
1	AUSGABE: ZUSAMMENFASSUNG						
2							
3	<i>Regressions-Statistik</i>						
4	Multipler Korrelationskoeffizient	0,910					
5	Bestimmtheitsmaß	0,827					
6	Adjustiertes Bestimmtheitsmaß	0,825					
7	Standardfehler	5,686					
8	Beobachtungen	128					
9							
10	ANOVA						
11		<i>Freiheitsgrade (df)</i>	<i>Quadratsummen (SS)</i>	<i>Mittlere Quadratsumme (MS)</i>	<i>Prüfgröße (F)</i>	<i>F krit</i>	
12	Regression	2	19356,535	9678,268	299,326	0,000	
13	Residue	125	4041,697	32,334			
14	Gesamt	127	23398,232				
15							
16		<i>Koeffizienten</i>	<i>Standardfehler</i>	<i>t-Statistik</i>	<i>P-Wert</i>	<i>Untere 95%</i>	<i>Obere 95%</i>
17	Schnittpunkt	86,018	1,255	68,548	0,000	83,534	88,501
18	pSGB2Quote	-0,703	0,234	-3,004	0,003	-1,166	-0,240
19	pMH	-0,547	0,061	-8,988	0,000	-0,667	-0,427
20							
21							
22	AUSGABE: RESIDUENPLOT						
23							
24	<i>Beobachtung</i>	<i>Schätzung pWbt</i>	<i>Residuen</i>				
25	1	65,807	1,093				
26	2	65,386	6,214				
27	3	66,303	2,697				
28	4	63,026	15,774				
29	5	64,044	9,244				

Analyse mit dem SPSS-Clone PSPP



PSPP

- » PSPP ist ein freier Clone des Programms SPSS
- » www.gnu.org/software/pspp/



PSPP is a program for statistical analysis of sampled data. It is a Free replacement for the proprietary program SPSS similar to it with a few exceptions.

The most important of these exceptions are, that there are no “time bombs”; you “expire” or deliberately stop working in the future. Neither are there any artificial cases or variables which you can use. There are no additional packages to purchase “advanced” functions; all functionality that PSPP currently supports is in the core.

PSPP can perform descriptive statistics, T-tests, linear regression and non-parametric tests. It is designed to perform its analyses as fast as possible, regardless of the size of the data set. PSPP with its graphical interface or the more traditional syntax commands.

A brief list of some of the features of PSPP follows:

PSPP

```
*Einlesen — PSPPIRE Syntax Editor
Datei Bearbeiten Ausführen Fenster Hilfe
38 |
39 CORRELATION
40 /VARIABLES = pWbt pU18 p18U30 p30U45 p4
41 /PRINT = TWOTAIL SIG.
42
43
```

Output — PSPPIRE Output Viewer
Datei Bearbeiten Fenster Hilfe

☞ CORRELATIONS CORRELATIONS
CORRELATION
/VARIABLES = pWbt pU18 p18U30 p30U45 p45U65 p65 pAlleinErz mHHGr
pSGB2Quote p65_A
/PRINT = TWOTAIL SIG.

Correlations

		pWbt	pU18	p18U30	p30U45	p45U65	p65	pAlleinErz	r
pWbt	Pearson Correlation	1,00	-,04	-,37	,15	,39	-,05	-,40	
	Sig. (2-tailed)		,64	,00	,08	,00	,57	,00	
	N	128	128	128	128	128	128	128	
pU18	Pearson Correlation	-,04	1,00	-,43	,08	-,02	-,50	,35	
	Sig. (2-tailed)	,64		,00	,39	,82	,00	,00	
	N	128	128	128	128	128	128	128	

- » PSPP verwendet den gleichen Prozeduraufruf wie SPSS
- » Die Berechnungen mit PSPP sind auch gewichtet möglich.

PSPP

REGRESSION

```

/VARIABLES= pSGB2Quote pMH
/DEPENDENT= pWbt
/STATISTICS= COEFF R ANOVA.

```

Model Summary

R-Quadrat	R Square	Adjusted R Square	Std. Error of the Estimate
,91	,83	,83	5,68

ANOVA

	Sum of Squares	df	Mean Square	F	Significance
Regression	19360,19	2	9680,10	299,65	,00
Residual	4038,04	125	32,30		
Total	23398,23	127			

Coefficients

	B	Std. Error	Beta	t	Significance
(Constant)	86,01	1,25	,00	68,60	,00
pSGB2Quote	-,70	,23	-,23	-3,02	,00
pMH	-,55	,06	-,70	-9,00	,00

- » PSPP verwendet den gleichen Prozeduraufruf wie SPSS
- » Berechnet werden Determinationskoeffizient, F-Statistik der Modellanpassung, unstandardisierte und standardisierte Regressionskoeffizienten, t-Werte, p-Werte

Kontakt

Hartmut Bömermann

030 – 9021 3585

hartmut.boemermann@statistik-bbb.de

