

Prof. Dr. Ivo Mossig  
Institut für Geographie  
Tel: 0421 / 218 67410  
ivo.mossig@uni-bremen.de

Dipl.-Geogr. Lars Schieber  
Institut für Geographie  
Tel: 0421 / 218 67113  
lars.schieber@uni-bremen.de

# Deskriptive Statistik

WS2009/10

## II Bivariate deskriptive Statistik

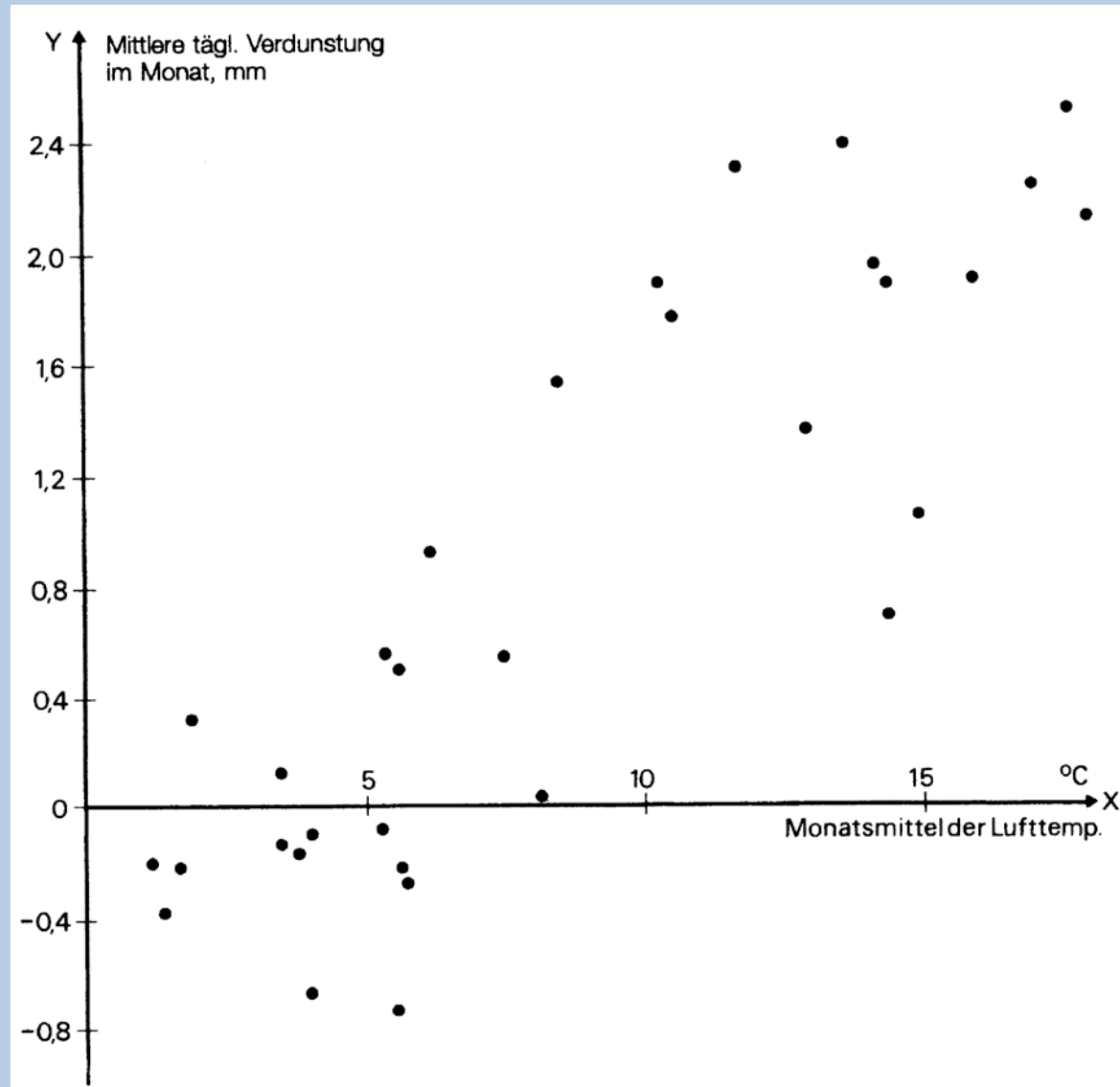
6 Einführung in die Korrelations- und  
Regressionsanalyse

# 6.1 Regressionsanalyse

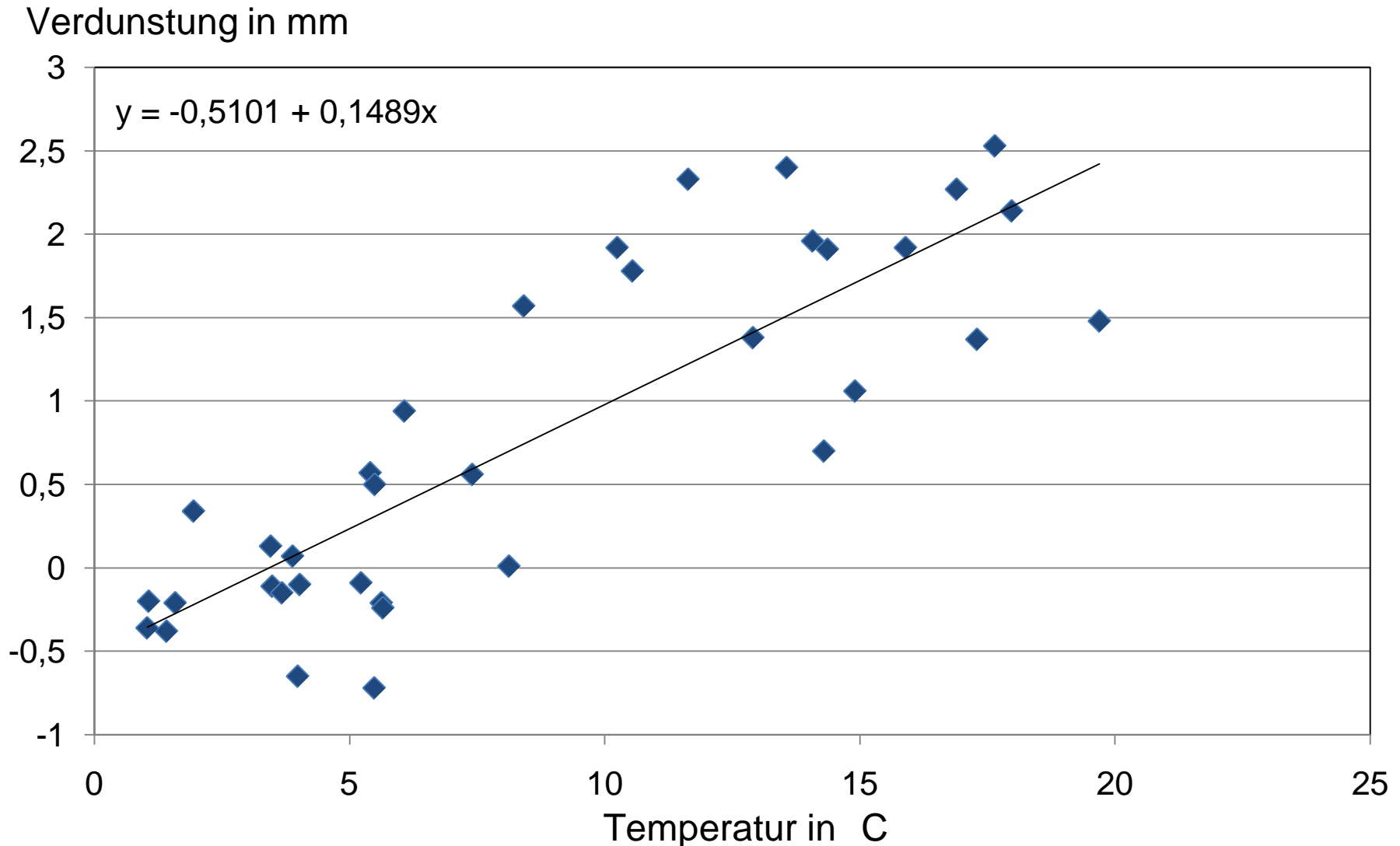
## Mittlere Tagessummen der Verdunstung im Monat und mittlere monatliche Lufttemperatur der Landverdunstungsmeßanlage Senne 1973 - 1975

Jahr/ Monat		Mittlere tägl. Verdunstung im Monat (mm)	Mittlere monatl. Lufttemperatur (°C)	Jahr/ Monat		Mittlere tägl. Verdunstung im Monat (mm)	Mittlere monatl. Lufttemperatur (°C)	Jahr/ Monat		Mittlere tägl. Verdunstung im Monat (mm)	Mittlere monatl. Lufttemperatur (°C)
	$i$	$y_i$	$x_i$		$i$	$y_i$	$x_i$		$i$	$y_i$	$x_i$
1973				1974				1975			
Januar	1	-0,20	1,06	Januar	13	-0,65	3,98	Januar	25	-0,09	5,22
Februar	2	-0,38	1,41	Februar	14	-0,11	3,48	Februar	26	0,34	1,94
März	3	0,07	3,88	März	15	0,50	5,49	März	27	0,13	3,45
April	4	0,57	5,40	April	16	1,57	8,41	April	28	0,94	6,07
Mai	5	2,33	11,63	Mai	17	1,78	10,54	Mai	29	1,92	10,24
Juni	6	1,92	15,89	Juni	18	2,40	13,56	Juni	30	1,91	14,36
Juli	7	1,37	17,29	Juli	19	1,96	14,07	Juli	31	2,53	17,64
August	8	2,14	17,97	August	20	2,27	16,89	August	32	1,48	19,69
September	9	0,70	14,29	September	21	1,38	12,90	September	33	1,06	14,90
Oktober	10	0,01	8,12	Oktober	22	-0,21	5,62	Oktober	34	0,56	7,40
November	11	-0,10	4,02	November	23	-0,72	5,48	November	35	-0,15	3,67
Dezember	12	-0,36	1,03	Dezember	24	-0,24	5,65	Dezember	36	-0,21	1,58

# Streuungsdiagramm der Werte zur Bestimmung des Zusammenhangs zwischen der Verdunstung in Abhängigkeit von der Lufttemperatur



# Zusammenhang zwischen der Verdunstung in Abhängigkeit von der Lufttemperatur



## Regressionskonstante a:

$$a = \frac{\sum_{i=1}^n x_i^2 * \sum_{i=1}^n y_i - \sum_{i=1}^n x_i * \sum_{i=1}^n x_i y_i}{n * \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$

## Regressionskoeffizient b:

$$b = \frac{n * \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i * \sum_{i=1}^n y_i}{n * \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$

## Rechenschema zur Bestimmung der Regressionskonstanten a und des Regressionskoeffizienten b

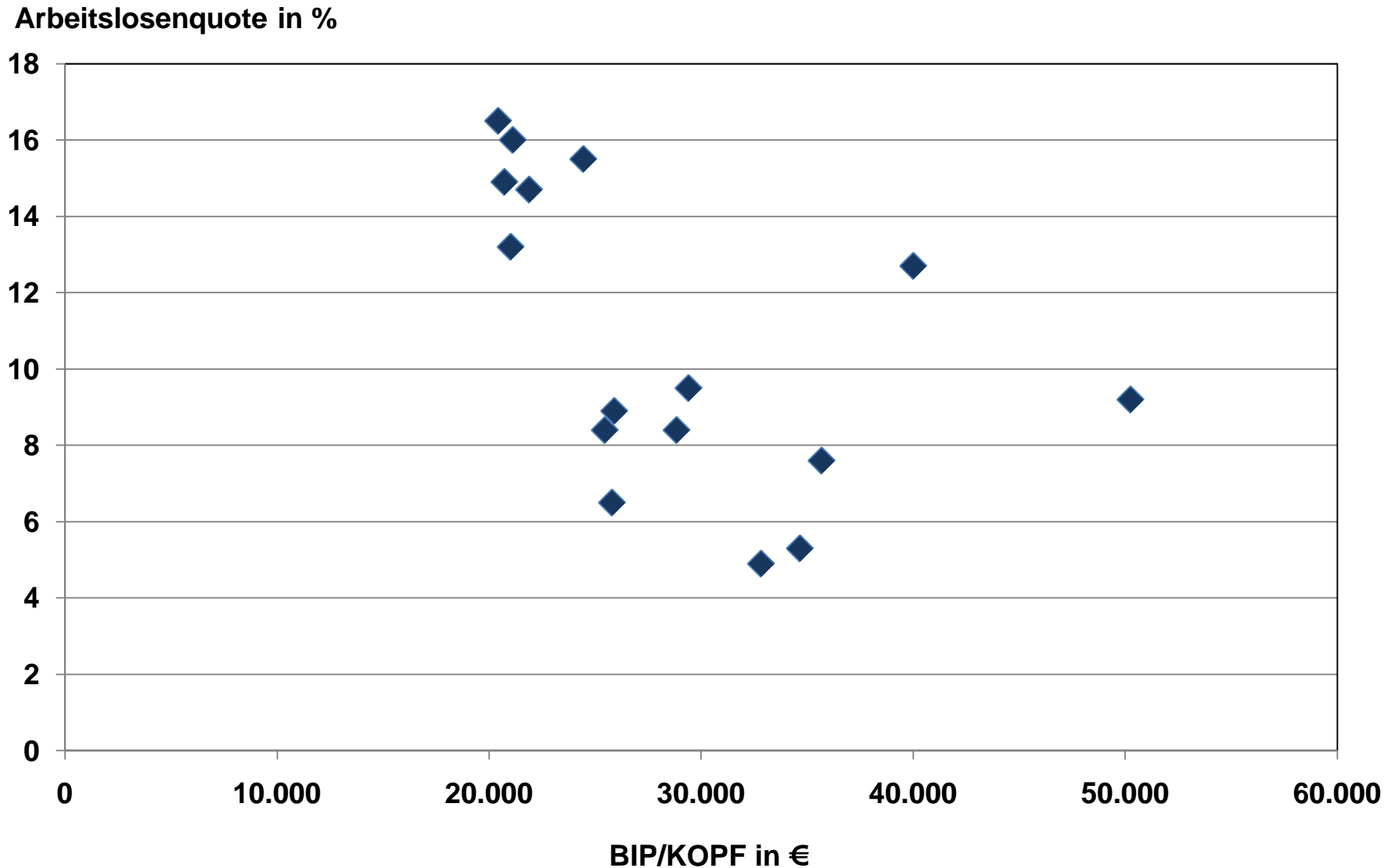
	Verdunstung	Lufttemperatur		
$i$	$y_i$	$x_i$	$x_i y_i$	$x_i^2$
1	-0,2	1,06	-0,212	1,1236
2	-0,38	1,41	-0,5358	1,9881
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
36	-0,21	1,58	-0,3318	2,4964
	$\sum_{i=1}^{36} y_i = ?$	$\sum_{i=1}^{36} x_i = ?$	$\sum_{i=1}^{36} x_i y_i = ?$	$\sum_{i=1}^{36} x_i^2 = ?$

# Datengrundlage zur Überprüfung der Fragestellung, ob und wie die Arbeitslosenquote (ALQ) eines Bundeslandes von der regionalen Wirtschaftskraft (gemessen am BIP/Kopf) abhängt

Bundesland	Arbeitslosen- quote 2007 in %	Bruttoinlands- produkt/Kopf (BIP/Kopf) 2007 in €
Baden-Württemberg	4,9	32.833
Bayern	5,3	34.666
Berlin	15,5	24.458
Brandenburg	14,9	20.734
Bremen	12,7	40.015
Hamburg	9,2	50.254
Hessen	7,6	35.689
Mecklenburg-Vorp.	16,5	20.440
Niedersachsen	8,9	25.913
Nordrhein-Westfalen	9,5	29.416
Rheinland-Pfalz	6,5	25.808
Saarland	8,4	28.852
Sachsen	14,7	21.900
Sachsen-Anhalt	16	21.131
Schleswig-Holstein	8,4	25.467
Thüringen	13,2	21.031

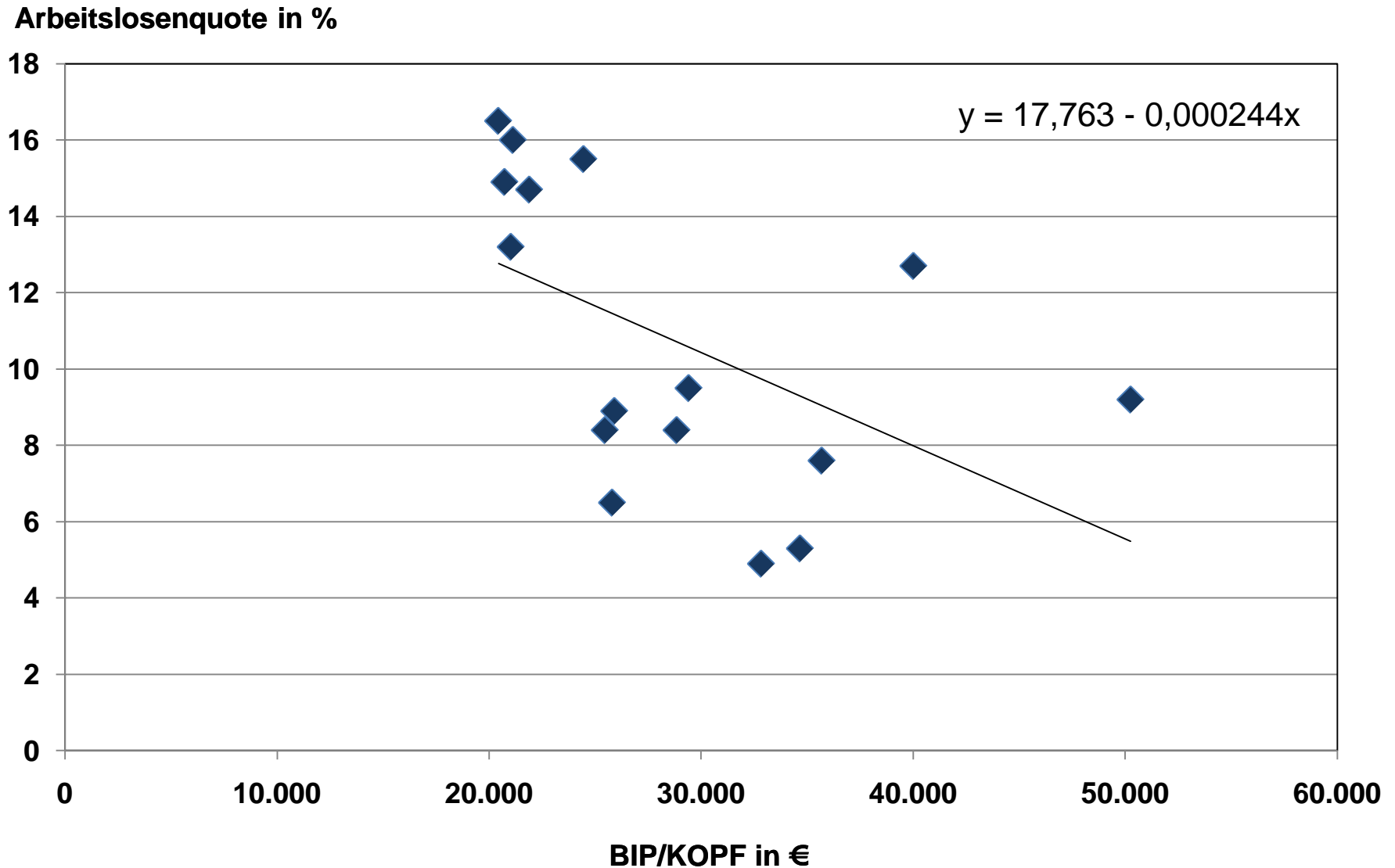
Quelle: Statistisches Bundesamt

# **Streuungsdiagramm bezüglich der Arbeitslosenquote 2007 in Abhängigkeit vom BIP/Kopf 2007 auf der Basis der Bundesländer**





# **Streuungsdiagramm bezüglich der Arbeitslosenquote 2007 in Abhängigkeit vom BIP/Kopf 2007 auf der Basis der Bundesländer**



# Bestimmtheitsmaß B

$$B = \frac{b^2 \cdot s_x^2}{s_y^2}$$

$$0 \leq B \leq 1 \quad \text{bzw.} \quad 0\% \leq B \leq 100\%$$

Das Bestimmtheitsmaß B gibt an, wieviel Prozent der Varianz der Y-Werte sich auf die Varianz der X-Werte zurückführen läßt.

## Produktmoment Korrelationskoeffizient r (nach Pearson)

$$r^2 = B; \quad \text{bzw.} \quad r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{KOVARIANZ}{s_x \cdot s_y}$$



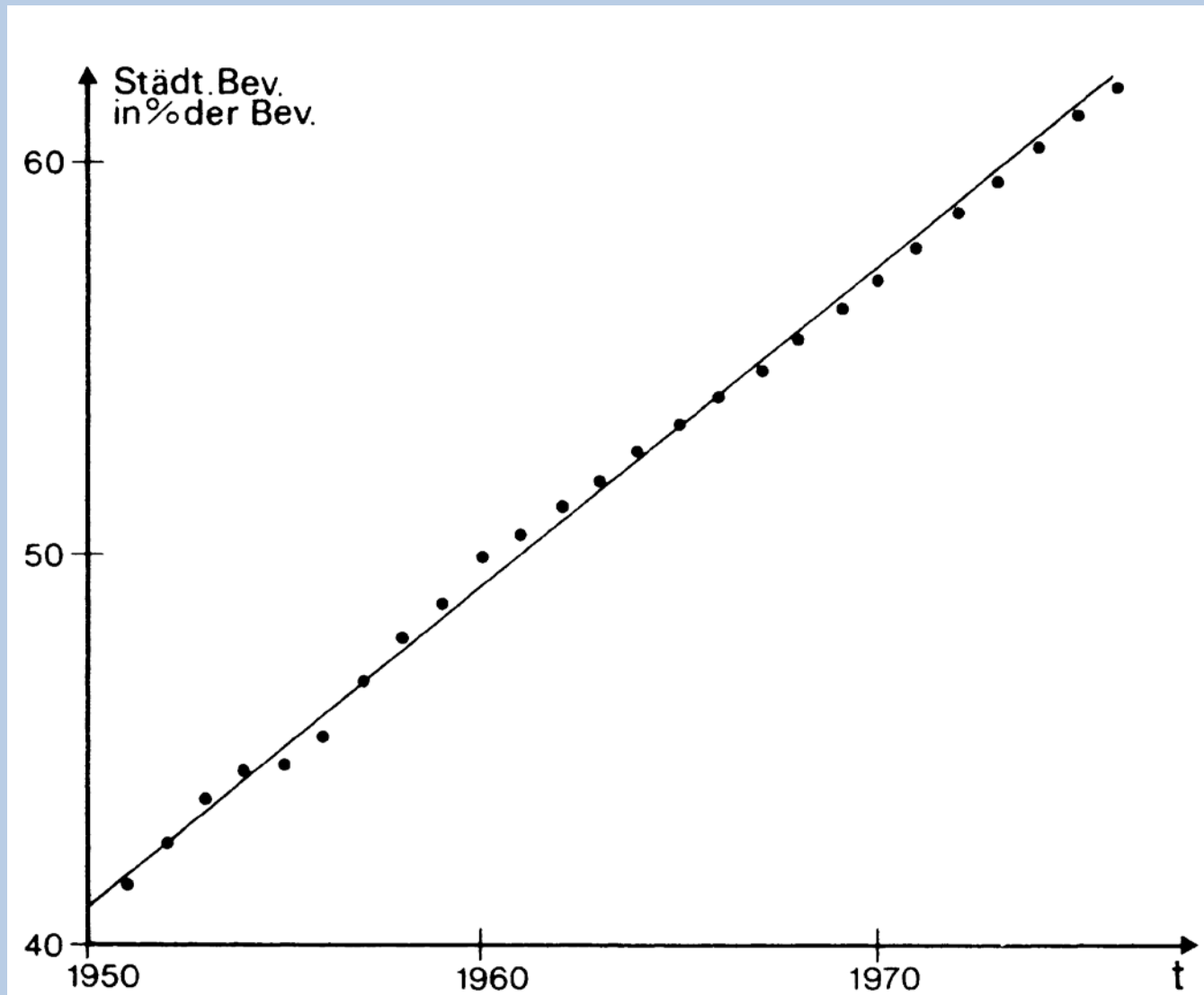
# Entwicklung der Natalität und Verstädterung in der Sowjetunion 1950 - 1976

Jahr	Natalität	Städt. Bev. in % d. Bev.
1950	26,7	40,2
1951	27,0	41,6
1952	26,5	42,7
1953	25,1	43,8
1954	26,6	44,4
1955	25,7	44,6
1956	25,2	45,4
1957	25,4	46,7
1958	25,3	47,9
1959	25,0	48,8
1960	24,9	49,9
1961	23,8	50,5
1962	22,4	51,2
1963	21,2	51,9
1964	19,6	52,6
1965	18,4	53,3
1966	18,2	54,0
1967	17,4	54,7
1968	17,3	55,5
1969	17,0	56,3
1970	17,4	57,0
1971	17,8	57,9
1972	17,8	58,8
1973	17,6	59,6
1974	18,0	60,4
1975	18,1	61,3
1976	18,4	61,9

Quelle: Bahrenberg et al. (1999, S. 145)

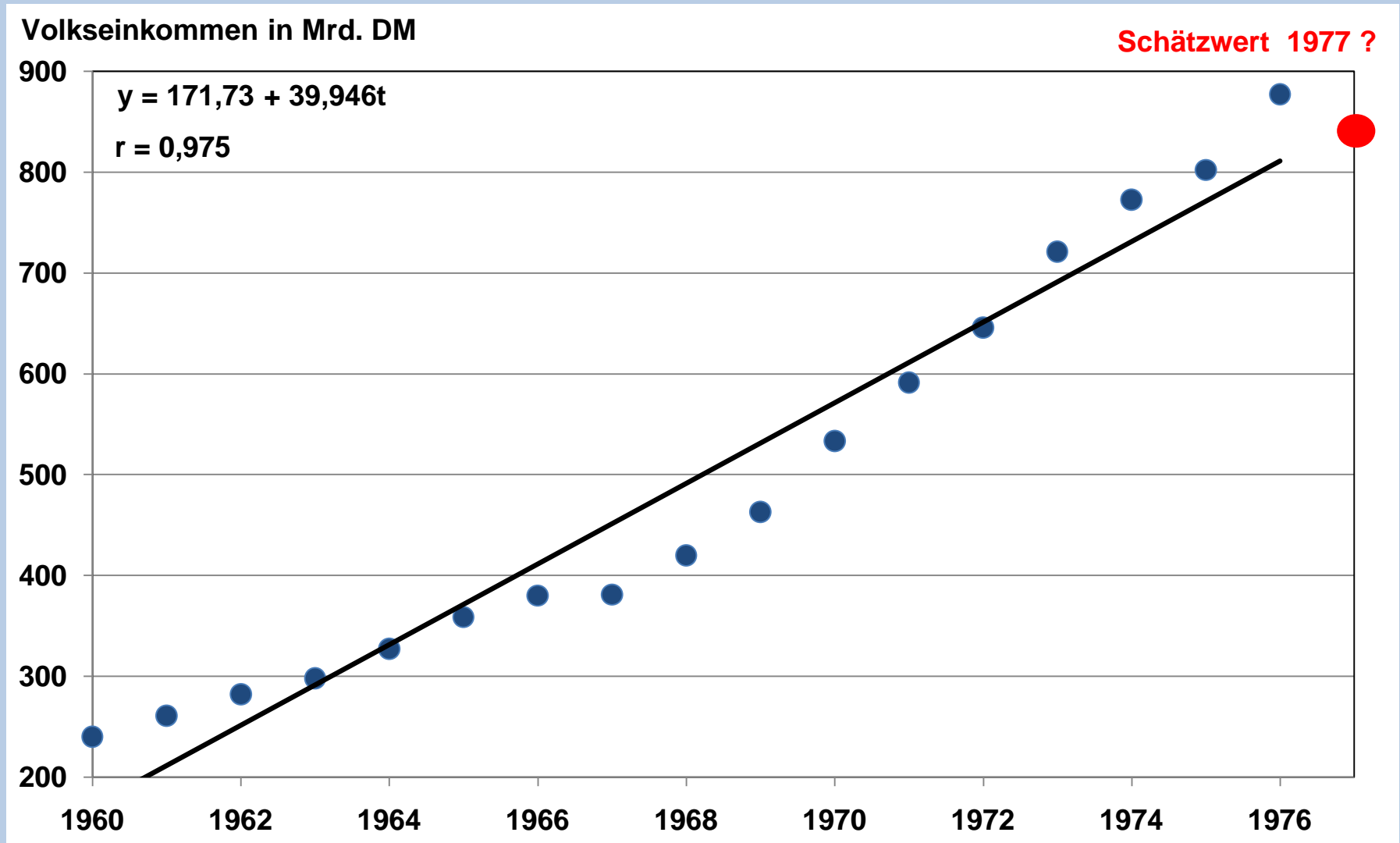
# Entwicklung der Verstädterung in der Sowjetunion 1950-1976

$$y = -1538,6 + 0,8101x \quad \text{bzw. mit Zeitachsentransformation } y = 41,058 + 0,81t$$



Quelle: Bahrenberg et al. (1999, S. 146)

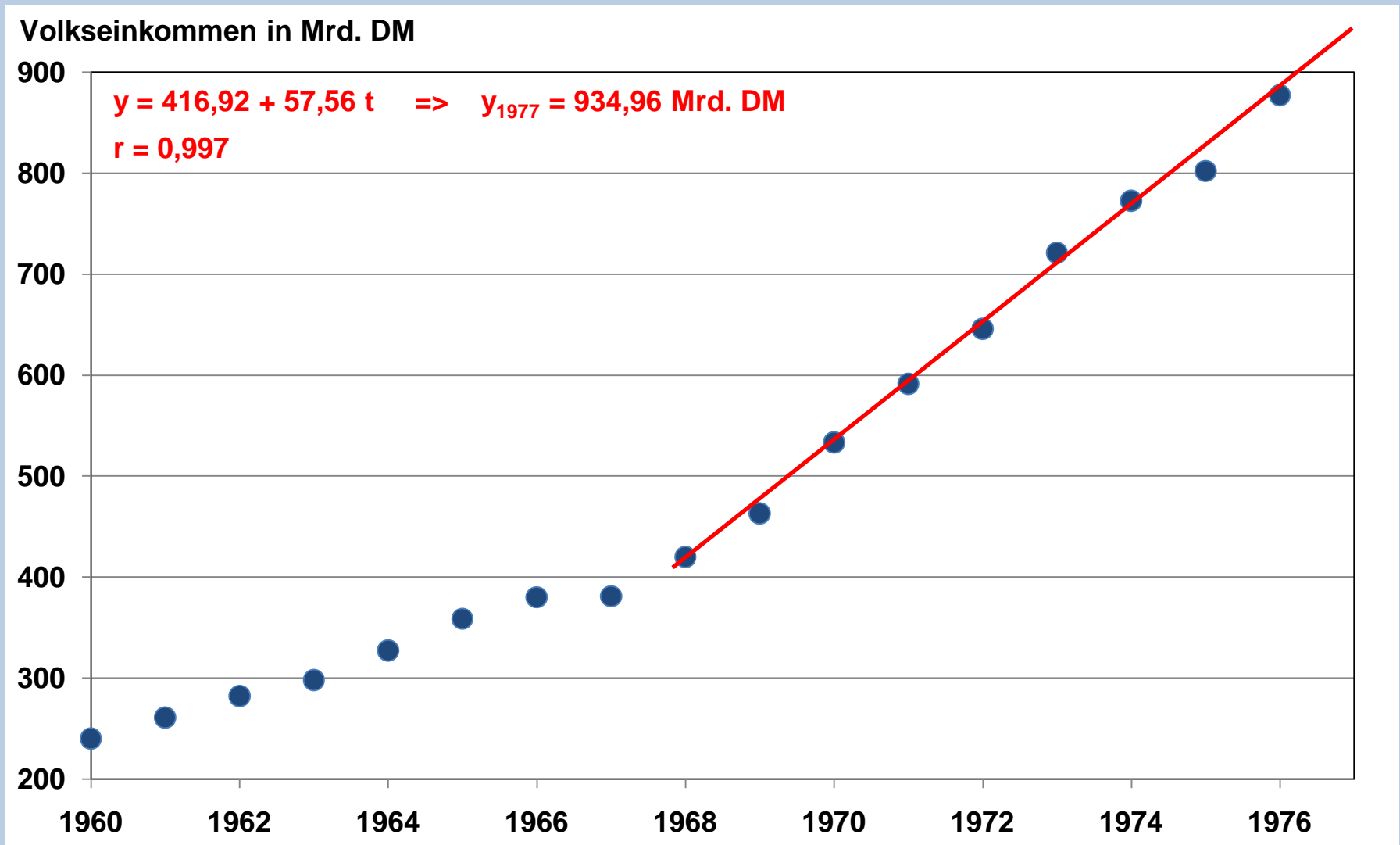
# Die Entwicklung des Volkseinkommens in der Bundesrepublik Deutschland 1960 - 1977



Quelle: Eigene Darstellung nach Bahrenberg et al. 1999, S. 168

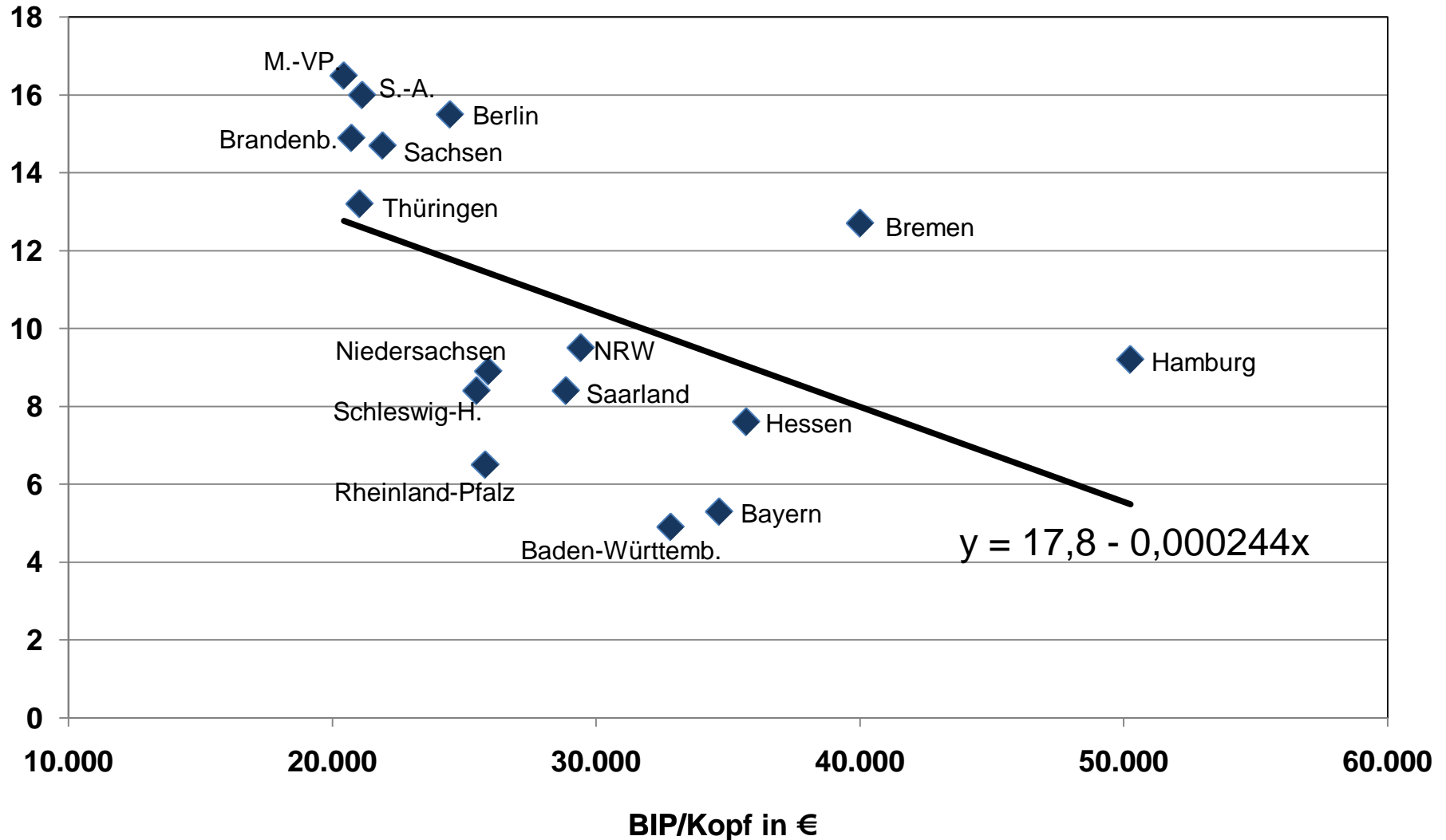


## Zerlegung der Regression für den Zeitraum 1968 - 1976



# **Streuungsdiagramm bezüglich der Arbeitslosenquote 2007 in Abhängigkeit vom BIP/Kopf 2007 auf der Basis der Bundesländer**

**Arbeitslosenquote in %**





## Residuen der Regression bezüglich der Arbeitslosenquote 2007 in Abhängigkeit von der regionalen Wirtschaftskraft 2007

Bundesland	Arbeitslosen- quote 2007 in %	Bruttoinlands- produkt/Kopf 2007 in €	Residuum $e_i$
Rheinland-Pfalz	6,5	25.808	-5,0
Baden-Württemberg	4,9	32.833	-4,9
Bayern	5,3	34.666	-4,0
Schleswig-Holstein	8,4	25.467	-3,1
Niedersachsen	8,9	25.913	-2,5
Saarland	8,4	28.852	-2,3
Hessen	7,6	35.689	-1,5
Nordrhein-Westfalen	9,5	29.416	-1,1
Thüringen	13,2	21.031	0,6
Brandenburg	14,9	20.734	2,2
Sachsen	14,7	21.900	2,3
Sachsen-Anhalt	16	21.131	3,4
Hamburg	9,2	50.254	3,7
Berlin	15,5	24.458	3,7
Mecklenburg-Vorp.	16,5	20.440	3,7
Bremen	12,7	40.015	4,7

Quelle: Eigene Berechnungen nach Stat. Bundesamt.

# Rangkorrelationskoeffizient nach Spearman

$$r = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad ; \text{ mit } d_i = |\text{Rang } x_i - \text{Rang } y_i|$$

**Fiktives Beispiel: Zusammenhang zwischen der Semesterzahl und der Zufriedenheit mit der Wahl des Studienfaches**  
**(1 = sehr zufrieden - 6 = sehr unzufrieden)**

Semesterzahl	Zufriedenheit	Rg. Sem.	Rg. Zufr.	di	di <sup>2</sup>
4	1				
5	3				
3	5				
1	2				
4	6				
1	5				
2	3				
4	2				
5	4				
6	4				
1	2				

## Achtung:

**Korrekturfaktor beachten, wenn „Bindungen“ auftreten**

$$r = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1) - (T_x + T_y)}$$

$$\text{mit } T_x = \frac{1}{2} \sum_{j=1}^k (t_{xj}^3 - t_{xj})$$

$$\text{und } T_y = \frac{1}{2} \sum_{j=1}^l (t_{yj}^3 - t_{yj})$$

wobei: k: Die Anzahl der Bindungen bei X

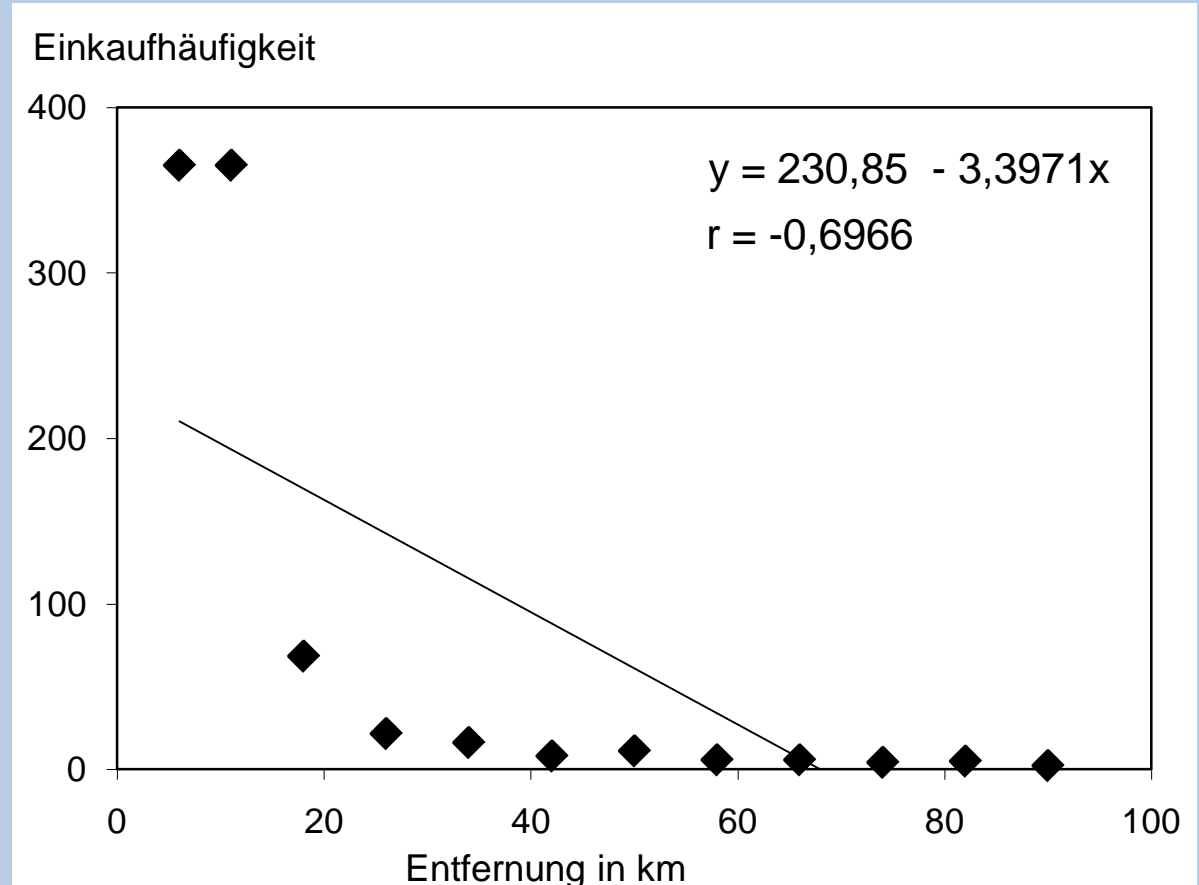
l: Die Anzahl der Bindungen bei Y

$t_{xj}$ : Anzahl der Stichprobenelemente in der j-ten Bindung von X

$t_{yj}$ : Anzahl der Stichprobenelemente in der j-ten Bindung von Y

# Häufigkeit des Einkaufs von Familien aus dem Umland von Turku in Abhängigkeit von der Entfernung der Wohnstandorte von Turku

Entfernung ( $x_i$ )	Einkäufe ( $y_i$ )
6	365
11	365
18	68,5
26	21,7
34	16,1
42	8,0
50	11,1
58	5,7
66	5,5
74	4,2
82	5,0
90	2,2



# Die Exponentialfunktion:

Variablentransformation:

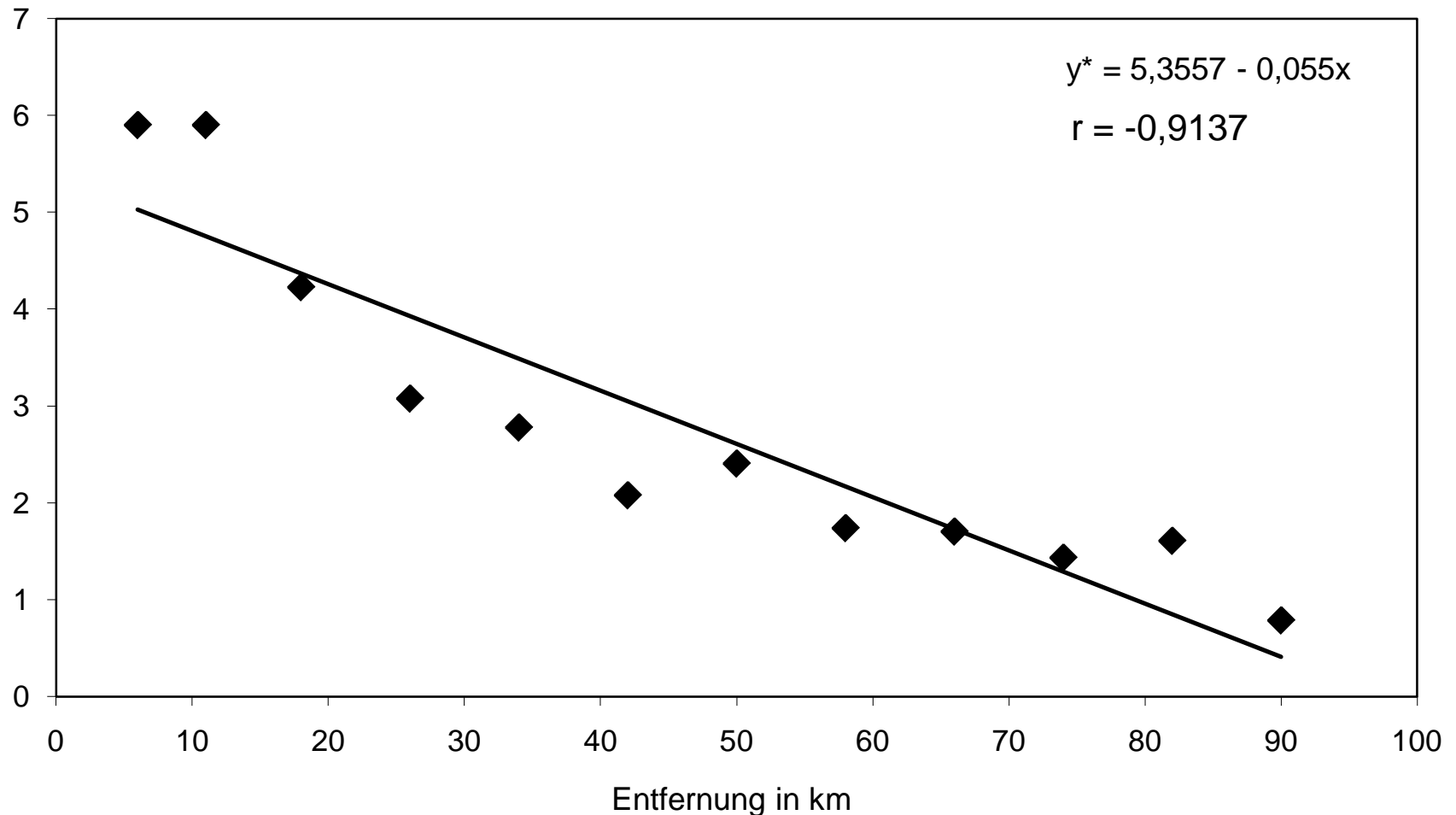
$$Y^* = \ln Y$$

$$X^* = X$$

Entfernung ( $x_i$ )	Einkäufe ( $y_i$ )	$X^*=X_i$	$Y^*=\ln(Y_i)$
6	365	6	5,8999
11	365	11	5,8999
18	68,5	18	4,2268
26	21,7	26	3,0773
34	16,1	34	2,7788
42	8,0	42	2,0794
50	11,1	50	2,4069
58	5,7	58	1,7405
66	5,5	66	1,7047
74	4,2	74	1,4351
82	5,0	82	1,6094
90	2,2	90	0,7885

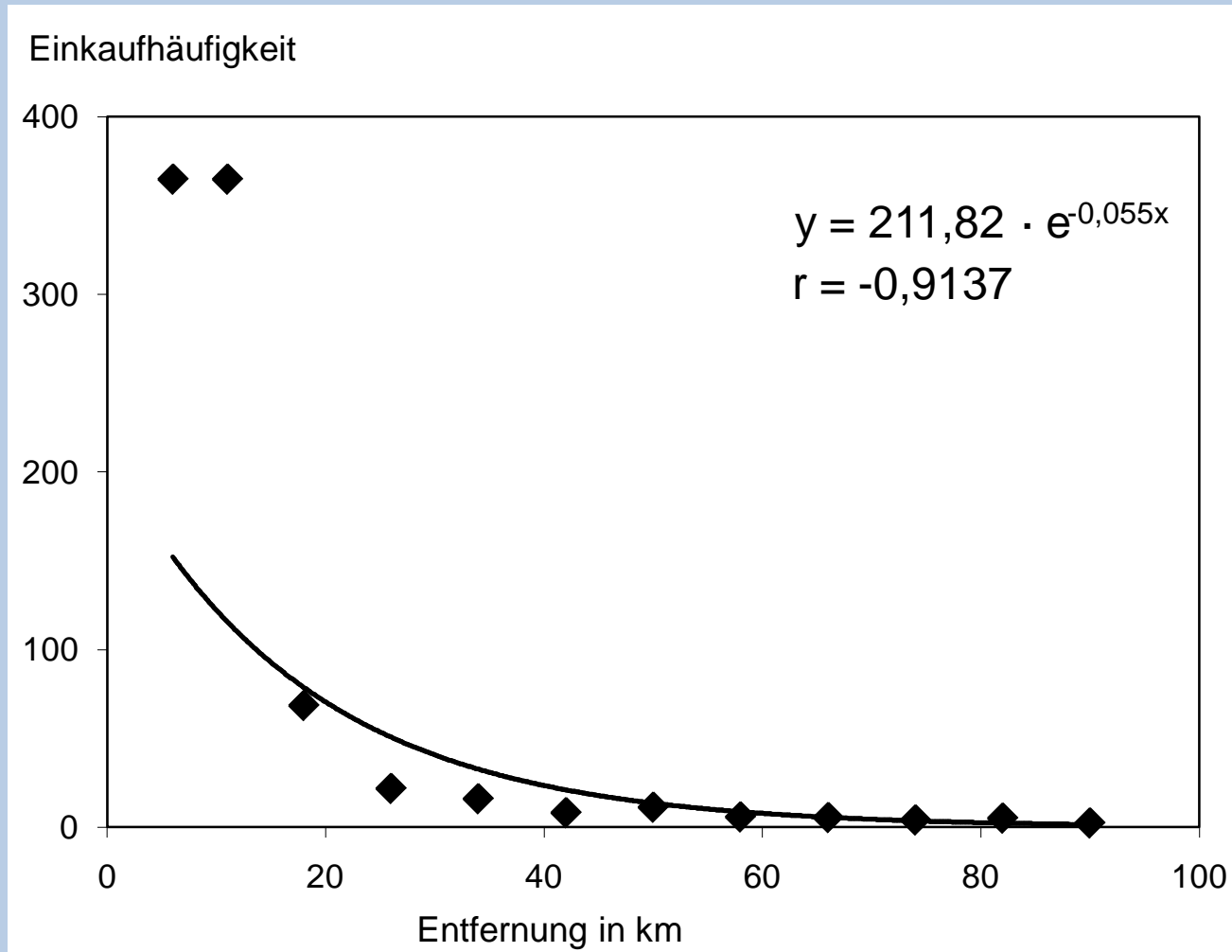
# Häufigkeit des Einkaufs von Familien aus dem Umland von Turku in Abhängigkeit von der Entfernung der Wohnstandorte von Turku

ln Y (Einkaufshäufigkeit, logarithmierte Werte)



Quelle: Eigene Darstellung nach Bahrenberg et al. 1999, S. 193

## Exponentialfunktion bezüglich der Häufigkeit des Einkaufs von Familien aus dem Umland von Turku in Abhängigkeit von der Entfernung der Wohnstandorte von Turku



# Potenzfunktion bezüglich der Häufigkeit des Einkaufs von Familien aus dem Umland von Turku in Abhängigkeit von der Entfernung der Wohnstandorte von Turku

