Kapitel III

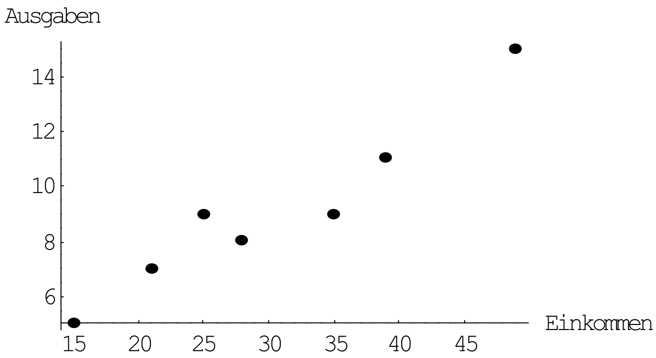
Regressionsanalyse

Lösungen

A. Rechenaufgaben

1.

1.



2.

x_i	y_i	x_i^2	$x_i \cdot y_i$
35	9	1225	315
49	15	2401	735
21	7	441	147
39	11	1521	429
15	5	225	75
28	8	784	224
25	9	625	225
212	64	7222	2150

$$\begin{cases} 7a_0 + 212a_1 = 64 \\ 212a_0 + 7222a_1 = 2150 \end{cases} \Rightarrow \quad a_0 = 1.1414, \quad a_1 = 0.2642.$$

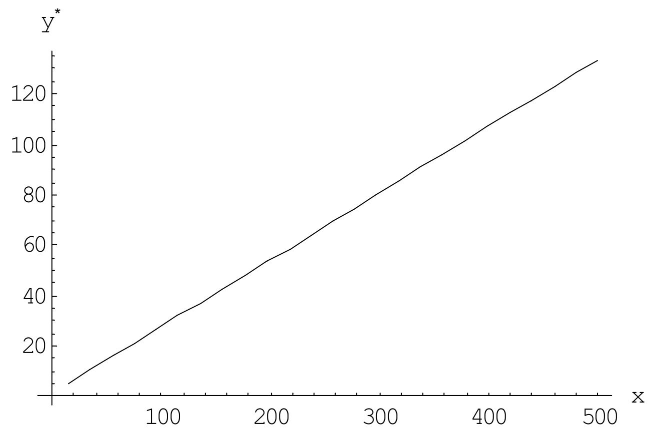
bzw.

$\bar{x} = \frac{212}{7} \approx 30.2857, \qquad \bar{y} = \frac{64}{7} \approx 9.1429$

x_i	y_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
35	9	22.22448966	-0.67346937
49	15	350.2244893	109.612245
21	7	86.22449006	19.8979592
39	11	75.93877526	16.1836735
15	5	233.6530617	63.3265306
28	8	5.224489861	2.61224491
25	9	27.93877566	0.75510203
212	64	801.4285714	211.714286

$a_1 = \frac{s_{xy}}{s_2^x} = \frac{211.7142}{801.4286} \approx 0.2642\,. \qquad a_0 = \bar{y} - a_1 \bar{x} = 9.1429 - 0.2642 \cdot 30.2857 \approx 1.1414$

$$y^* = 1.1414 + 0.2642x$$



3.

x_i	y_i	y_i^*	u_i^*	u_i^{*2}
35	9	10.3884	-1.3884	1.92765456
49	15	14.0872	0.9128	0.83320384
21	7	6.6896	0.3104	0.09634816
39	11	11.4452	-0.4452	0.19820304
15	5	5.1044	-0.1044	0.01089936
28	8	8.5390	-0.5390	0.29052100
25	9	7.7464	1.2536	1.57151296
212	64	64.0002	-0.0002[1]	4.92834292

[1] Die Summe ist nur wegen Rundungsfehler nicht gleich Null.

4.
$a_0 = 1.1414$ besagt:

Bei einem Einkommen von 0 wären die Ausgaben für Lebensmittel gleich $114.14! Das Ergebnis ist jedoch nicht glaubhaft, da Null außerhalb des Definitionsbereichs der Funktion liegt.

$a_1 = 0.2642$ besagt:

Eine Erhöhung des Einkommens um eine Einheit würde zu einer Erhöhung von Ausgaben für Lebensmittel um 0.2642 Einheiten führen.

5.
V1:
Für vorgegebene Werte der Variablen $x_i,\ i = 1,2,...,7,$ ist im Mittel zu erwarten. dass die zufällige Störvariable u Nulle ist:

$$E(u_i) = 0 \text{ bzw. } E(u) = 0 .$$

Mit anderen Worten: Unter den Haushalten mit gleichem Einkommen manche geben mehr und manche weniger für Lebensmittel aus als vorausgesagt. Die Summe der Residuen und damit auch das arithmetische Mittel der Fehler sind also gleich Null.

V2:
Die Werte der Zufallsvariablen sind nicht miteinander korreliert. Dies bedeutet. dass die Haushalte unabhängig voneinander entscheiden. wie viel sie für Lebensmittel ausgeben.

V3:
Die Zufallsvariable u ist normalverteilt. Mit anderen Worten: Die Verteilung der Ausgaben für Lebensmittel bei allen Familien mit gleichen Einkommen ist normal.

V4:
Die Varianz der Zufallsvariablen u ist für alle Werte $u_i,\ i = 1,2,...,7,$ gleich und konstant:

$$E(u_i^2) = \sigma_u^2 .$$

(Diese Eigenschaft der Störvariablen u wird al *Homoskedastizität* bezeichnet.)

Damit ist die Streuung der Punkte um die Regressionsgerade für alle x_i ähnlich.

6.
Siehe 3.

$$s_u^2 = \frac{\sum_{i=1}^{n} u_i^{*2}}{n-(m+1)} = \frac{4.9283}{7-(1+1)} = 0.9857 .$$

$$s_u = 0.9928 .$$

7.

$$B_{yx} = \frac{\sum_{i=1}^{n}(y_i^* - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

y_i	y_i^*	$(y_i - \bar{y})^2$	$(y_i^* - \bar{y})^2$
9	10.3884	0.02040816	1.55130585
15	14.0872	34.3061225	24.4462438
7	6.6896	4.59183672	6.01861077
11	11.4452	3.4489796	5.30065109
5	5.1044	17.1632653	16.3093668
8	8.539	1.30612244	0.36467795
9	7.7464	0.02040816	1.95017234
64	64.0002	60.8571429	55.9410286

$$B_{yx} = \frac{55.9410286}{60.8571429} \approx 0.92.$$

Das Einkommen bestimmt zu etwa 92% die Ausgaben für Lebensmittel. Es handelt sich um eine gute Approximation.

8.

$$X = \begin{pmatrix} 1 & 35 \\ 1 & 49 \\ 1 & 21 \\ 1 & 39 \\ 1 & 15 \\ 1 & 28 \\ 1 & 25 \end{pmatrix}.$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 35 & 49 & 21 & 39 & 15 & 28 & 25 \end{pmatrix} \cdot \begin{pmatrix} 1 & 35 \\ 1 & 49 \\ 1 & 21 \\ 1 & 39 \\ 1 & 15 \\ 1 & 28 \\ 1 & 25 \end{pmatrix} = \begin{pmatrix} 7 & 212 \\ 212 & 7222 \end{pmatrix}.$$

$$(X^T X)^{-1} = \begin{pmatrix} 7 & 212 \\ 212 & 7222 \end{pmatrix}^{-1} = \frac{1}{5610}\begin{pmatrix} 7222 & -212 \\ -212 & 7 \end{pmatrix} \approx \begin{pmatrix} 1.2873 & -0.0378 \\ -0.0378 & 0.0012 \end{pmatrix}$$

$$s_{a_i} = s_u \sqrt{x^{(ii)}}, \quad i = 0,1. \qquad\qquad s'_{a_i} = \frac{s_{a_i}}{a_i}, \quad i = 0,1.$$

$$s_{a_0} = 0.9928 \cdot \sqrt{1.2873} \approx 1.1264, \qquad s'_{a_0} = \frac{s_{a_0}}{a_0} = \frac{1.1264}{1.1414} \approx 0.9869. \quad \text{annehmbar.}$$

$$s_{a_1} = 0.9928 \cdot \sqrt{0.0012} \approx 0.0344, \qquad s'_{a_1} = \frac{s_{a_1}}{a_1} = \frac{0.0344}{0.2642} \approx 0.1302. \quad \text{annehmbar.}$$

9.

$$\left[a_i - t_{n-m-1} \cdot s_{a_i},\ a_i + t_{n-m-1} \cdot s_{a_i}\right],\ i = 0,1.$$

$$A_0 \in \left[1.1414 - 2.571 \cdot 1.1264,\ 1.1414 + 2.571 \cdot 1.1264\right] \approx \left[-1.7546,\ 4.0374\right].$$

$$A_1 \in \left[0.2642 - 2.571 \cdot 0.0344,\ 0.2642 + 2.571 \cdot 0.0344\right] \approx \left[0.1758,\ 0.3526\right].$$

10.

$$n = 7, \quad a_1 = 0.2644, \qquad \alpha = 0.01, \qquad s_{a_1} = 0.0344$$

S1:

$$H_0 : A_1 = 0, \qquad H_1 : A_1 > 0.$$

S2:

Da σ_u nicht bekannt ist. wird die $t-$Verteilung benutzt.

S3:

Der Test ist wegen $H_1 : H_1 > 0$ rechtsseitig. $df = 7 - 2 = 5$. Damit liegt die Ablehnungsregion rechts von 3.365.

S4:

$$t_{stat} = \frac{a_1 - A_1}{s_{a_1}} = \frac{0.2642}{0.0344} \approx 7.6802$$

S5:

Wegen $7.6802 > 3.365$ wird die Nullhypothese abgelehnt. Dies bedeutet. dass mit zunehmendem Einkommen die Ausgaben für Lebensmittel wachsen.

2.

Arbeitstabelle 1

y_i	x_{i1}	x_{i2}	x_{i1}^2	x_{i2}^2	$x_{i1} \cdot y_i$	$x_{i2} \cdot y_i$	$x_{i1} \cdot x_{i2}$
148	5	2	25	4	740	296	10
76	14	0	196	0	1064	0	0
100	6	1	36	1	600	100	6
126	10	3	100	9	1260	378	30
194	4	6	16	36	776	1164	24
110	8	2	64	4	880	220	16
114	11	3	121	9	1254	342	33
86	16	1	256	1	1376	86	16
198	3	5	9	25	594	990	15
92	9	1	81	1	828	92	9
70	19	0	361	0	1330	0	0
120	13	3	169	9	1560	360	39
1434	118	27	1434	99	12262	4028	198

1.

$$\begin{cases} 12a_0 + 118a_1 + 27a_2 = 1434 \\ 118a_0 + 1434a_1 + 198a_2 = 12262 \\ 27a_0 + 198a_1 + 99a_2 = 4028 \end{cases}$$

x_1	x_2	x_3	x_0
12	118	27	1434
118	1434	198	12262
27	198	99	4028
1	9.8333	2.2500	119.5000
0	273.6667	-67.5000	-1839.0000
0	-67.5000	38.2500	801.5000
1	0	4.6754	185.5786
0	1	-0.2467	-6.7199
0	0	1.0000	16.1061
1	0	0	110.2761
0	1	0	-2.7473
0	0	1	16.1061

$a_0 = 110.2761 \quad a_1 = -2.7473 \quad a_2 = 16.1061$

$$y^* = 110.2761 - 2.7473x_1 + 16.1061x_2$$

Der Koeffizient $a_1 = -2.7473$ besagt: Jedes zusätzliches Fahrerfahrungsjahr reduziert bei gleichbleibender Anzahl der Verkehrsdelikte den monatlichen Versicherungsbeitrag um etwa $2.7473.

Der Koeffizient $a_2 = 16.1061$ besagt: Jeder weitere Verkehrsdelikt erhöht bei gleichbleibender Anzahl der Fahrerfahrungsjahre den monatlichen Versicherungsbeitrag um etwa $16.1061.

Arbeitstabelle 2

y_i	x_{i1}	x_{i2}	y_i^*	$(y_i^* - \bar{y})^2$	$(y_i - \bar{y})^2$	u_i^2
148	5	2	128.75189	85.5974686	812.25	370.489739
76	14	0	71.81422	2273.93361	1892.25	17.5207542
100	6	1	109.8985	92.1888023	380.25	97.9803023
126	10	3	131.12166	135.062981	42.25	26.2314012
194	4	6	195.92364	5840.57275	5550.25	3.70039085
110	8	2	120.51008	1.02026161	90.25	110.461782
114	11	3	128.37439	78.7547979	30.25	206.623088
86	16	1	82.4258	1374.49631	1122.25	12.7749056
198	3	5	182.56479	3977.16774	6162.25	238.245708
92	9	1	101.65669	318.383712	756.25	93.2516618
70	19	0	58.07787	3772.67805	2450.25	142.137184
120	13	3	122.87985	11.423386	0.25	8.29353602
1434	118	27	1433.99938	17961.2799	19289	1327.71045

2.

$$\bar{y} = \frac{1438}{12} = 119.5$$

$$r^2 = \frac{17961.2799}{19289.0000} = 0.93116698 \approx 0.93.$$

Der monatliche Versicherungsbeitrag wird zu etwa 93% durch die Fahrerfahrung und die Anzahl der Verkehrsdelikte bestimmt.
Die Approximation ist sehr gut.

3.

$$y^* = 110.2761 - 2.7473 \cdot 12 + 16.1061 \cdot 4 \approx 141.7329$$

4.

$$s_u^2 = \frac{1327.71045}{12-(2+1)} = 147.5233833 \approx 147.5234; \qquad s_u \approx 12.1459.$$

$$\left(X^T X\right)\left(X^T X\right)^{-1} = \begin{pmatrix} 12 & 118 & 27 \\ 118 & 1434 & 198 \\ 27 & 198 & 99 \end{pmatrix} \begin{pmatrix} 1.449 & -0.089 & -0.216 \\ -0.089 & b & 0.011 \\ -0.216 & 0.011 & c \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$118 \cdot (-0.089) + 1434b + 198 \cdot 0.011 = 1 \quad \Rightarrow \quad b \approx 0.0065$$
$$27 \cdot (-0.216) + 198 \cdot 0.011 + 99 \cdot c = 1 \quad \Rightarrow \quad c \approx 0.0470$$

$$s_{a_1} = 12.1459 \cdot \sqrt{0.0065} \approx 0.9792,$$

$$s_{a_2} = 12.1459 \cdot \sqrt{0.0470} \approx 2.6332,$$

$$A_1 \in [-2.7473 - 2.262 \cdot 0.9792, -2.7473 + 2.262 \cdot 0.9792\ ] \approx [-4.9622,\ -0.5323]$$

$$A_2 \in [16.1061 - 2.262 \cdot 2.6332,\ 16.1061 + 2.262 \cdot 2.6332\ ] \approx [10.1498,\ 22.0624].$$

5.

S1:

$$H_0:\ A_1 = A_2 = 0, \qquad H_1:\ A_1 \text{ und/oder } A_2 \neq 0$$

Step 2:

$$MSR = \frac{\sum_{i=1}^{n}(y_i^* - \bar{y})^2}{m} = \frac{17961.2799}{2} = 8980.63995 \approx 8980.645$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - y_i^*)^2}{n-m-1} = \frac{SST - SSR}{n-m-1} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i^* - \bar{y})^2}{n-m-1} = \frac{19289 - 17961.2799}{12-2-1} \approx 147.52$$

$$F = \frac{MSR}{MSE} = \frac{8989.64}{147.52} \approx 60.94$$

Step 3:

$$F = 60.94 \geq 4.26 = F_{2,9;0.05} \quad \Rightarrow \text{ reject } H_0$$

6.

A_1 :

S1:

$$H_0:\ A_1 = 0, \quad H_1:\ A_1 < 0.$$

S2:

Da σ_u nicht bekannt ist, wird die $t-$Verteilung benutzt.

S3:

Der Test ist wegen $H_1:\ A_1 < 0$ linksseitig. $df = 12 - 3 = 9$. Damit liegt die Ablehnungsregion links von -1.833.

S4:

$$t_{stat} = \frac{a_1 - A_1}{s_{a_1}} = -\frac{2.7473}{0.9890} \approx -2.778\,.$$

S5:
Wegen $-2.778 < -1.833$ wird die Nullhypothese abgelehnt.

A_2 :

S1:

$$H_0 : A_2 = 0, \quad H_1 : A_2 > 0.$$

S2:
Da σ_u nicht bekannt ist, wird die $t-$Verteilung benutzt.

S3:
Der Test ist wegen $H_1 : A_2 > 0$ rechtsseitig. $df = 12 - 3 = 9$. Damit liegt die Ablehnungsregion rechts von 2.262.

S4:

$$t_{stat} = \frac{a_2 - A_1}{s_{a_2}} = \frac{16.10612}{2.6362} \approx 6.1096 .$$

S5:
Wegen $6.1096 > 1.833$ wird die Nullhypothese abgelehnt.

3.

1.

Arbeitstabelle

y_i	x_{i1}	x_{i2}	x_{i1}^2	x_{i2}^2	$x_{i1}y_i$	$x_{i2}y_i$	$x_{i1}.x_{i2}$
2.1	15	2	225	4	31,5	4.2	30
3.0	28	4	784	16	84.0	12	112
1.6	20	4	400	16	32.0	6.4	80
2.1	22	3	484	9	46.2	6.3	66
1.2	10	2	100	4	12.0	2.4	20
10	95	15	1993	49	205.7	31.3	308

$$\begin{cases} 5a_0 + 95a_1 + 15a_2 = 10 \\ 95a_0 + 1993a_1 + 308a_2 = 205.7 \\ 15a_0 + 308a_1 + 49a_2 = 31.3 \end{cases}$$

x_1	x_2	x_3	x_0
5	95	15	10.0000
95	1993	308	205.7000
15	308	49	31.3000
1	19	3.0000	2.0000
0	1	0.1223	0.0835
0	23	4.0000	1.3000
1	0	0.6755	0.4133
0	1	0.1223	0.0835
0	0	1.1862	-0.6207
1	0	0	0.7668
0	1	0	0.1475
0	0	1	-0.5233

$a_0 = 0.7668,\ a_1 = 0.1475,\ a_2 = -0.5233$

$$y^* = 0.7668 + 0.1475x_1 - 0.5233x_2$$

2.

$$\bar{y} = \frac{10}{5} = 2$$

Arbeitstabelle 2

y_i^*	$(y_i^* - \bar{y})^2$	$(y_i - \bar{y})^2$	u_i^{*2}
1.9327	0.00452929	0.01	0.02798929
2.8036	0.64577296	1.00	0.03857296
1.6236	0.14167696	0.16	0.00055696
2.4419	0.19527561	0.01	0.11689561
1.1952	0.64770304	0.64	2.304E-05
9.997	1.63495786	1.82	0.18403786

$$r^2 = \frac{1.63495786}{1.82} = 0.898328494 \approx 0.90.$$

Die Ersparnisse sind damit zu etwa 90% durch das Einkommen und die Anzahl der Kinder erklärt (eine sehr gute Approximation).

3.

$$s_u^2 = \frac{0.18403786}{5-(2+1)} = 0.09201893 \approx 0.0920, \quad s_u \approx 0.3033$$

$$95 \cdot (-0.03139) + 1993b - 0.10313 \cdot 308 = 1 \;\Rightarrow\; b \approx 0.0180$$

$$15 \cdot (-0.56951) + 308 \cdot (-0.10314) + 49c = 1 \;\Rightarrow\; c \approx 0.8431$$

(

$$(X^T X)^{-1} = \begin{pmatrix} 5 & 95 & 15 \\ 95 & 1993 & 308 \\ 15 & 308 & 49 \end{pmatrix}^{-1} = \begin{pmatrix} 2.504932735 & -0.031390135 & -0.569506726 \\ -0.031390135 & 0.017937220 & -0.103139013 \\ -0.569506726 & -0.103139013 & 0.843049327 \end{pmatrix}$$

)

$$s_{a_1} = 0.3033 \cdot \sqrt{0.0180} \approx 0.0407, \qquad s'_{a_1} = \frac{0.0407}{0.1475} \approx 0.2759 \quad \Rightarrow \text{ annehmbar.}$$

$$s_{a_2} = 0.3033 \cdot \sqrt{0.8431} \approx 0.2785 \qquad s'_{a_2} = \left|\frac{0.2785}{-0.5233}\right| \approx 0.5322 \quad \Rightarrow \text{ annehmbar.}$$

$$A_1 \in \left[0.1475 - 4.303 \cdot 0.0407, \quad 0.1475 + 4.303 \cdot 0.0407\right] \approx \left[-0.0276,\ 0.3226\right]$$

$$A_2 \in [-0.5233 - 4.303 \cdot 0.2785, \quad -0.5233 + 4.303 \cdot 0.2785] \approx [-1.7217,\ 0.6751]$$

4.

A_1 :

S1:

$$H_0:\ A_1 = 0, \quad H_1:\ A_1 > 0.$$

S2:

Da σ_u nicht bekannt ist, wird die t – Verteilung benutzt.

S3:

Der Test ist wegen $H_1:\ A_1 > 0$ rechtsseitig. $df = 5 - 3 = 2$. Damit liegt die Ablehnungsregion rechts von 2.920.

S4:

$$t_{stat} = \frac{a_1 - A_1}{s_{a_1}} = \frac{0.1475}{0.0407} \approx 3.62.$$

S5:

Wegen $3.62 > 2.920$ wird die Nullhypothese abgelehnt.

A_2 :

S1:

$$H_0:\ A_2 = 0, \quad H_1:\ A_2 < 0.$$

S2:

Da σ_u nicht bekannt ist, wird die t – Verteilung benutzt.

S3:

Der Test ist wegen $H_1:\ A_2 < 0$ linksseitig. $df = 5 - 3 = 2$. Damit liegt die Ablehnungsregion links von -2.920.

S4:

$$t_{stat} = \frac{a_2 - A_2}{s_{a_2}} = -\frac{0.5233}{0.2785} \approx -1.8790.$$

S5:

Wegen $-1.8790 > -2.920$ wird die Nullhypothese nicht abgelehnt.

4.

Arbeitstabelle

y_i	x_{i1}	x_{i2}	x_{i1}^2	x_{i2}^2	$x_{i1}y_i$	$x_{i2}y_i$	$x_{i1}.x_{i2}$
2.1	15	2	225	4	31,5	4.2	30
3.0	28	4	784	16	84.0	12	112
1.6	20	4	400	16	32.0	6.4	80
2.1	22	3	484	9	46.2	6.3	66
1.2	10	2	100	4	12.0	2.4	20
10	95	15	1993	49	205.7	31.3	308

$$\begin{cases} 5a_0 + 95a_1 + 15a_2 = 10 \\ 95a_0 + 1993a_1 + 308a_2 = 205.7 \\ 15a_0 + 308a_1 + 49a_2 = 31.3 \end{cases}$$

x_1	x_2	x_3	x_0
5	95	15	10
95	1993	308	205.7
15	308	49	31.3
1	19	3	2
0	188	23	15.7
0	23	4	1.3
1	0	0.6755	0.4133
0	1	0.1223	0.0835
0	0	0.1861	-0.6207
1	0	0	0.7668
0	1	0	0.1475
0	0	1	-0.5233

$$a_0 = \frac{171}{223} \approx 0.7668, \quad a_1 = \frac{329}{2230} \approx 0.1475, \quad a_2 = -\frac{1167}{2230} \approx -0.5233$$

$$y^* = 0.7668 + 0.1475 x_1 - 0.5233 x_2$$

2.

$$\bar{y} = \frac{10}{5} = 2$$

Arbeitstabelle 2

y_i^*	$(y_i^* - \bar{y})^2$	$(y_i - \bar{y})^2$	u_i^{*2}
1.9327	0.00452929	0.01	0.02798929
2.8036	0.64577296	1.00	0.03857296
1.6236	0.14167696	0.16	0.00055696
2.4419	0.19527561	0.01	0.11689561
1.1952	0.64770304	0.64	2.304E-05
9.997	1.63495786	1.82	0.18403786

$$r^2 = \frac{1.63495786}{1.82} = 0.898328494 \approx 0.90.$$

Die Ersparnisse sind damit zu etwa 90% durch das Einkommen und die Anzahl der Kinder erklärt (eine sehr gute Approximation).

3.

$$s_u^2 = \frac{0.18403786}{5-(2+1)} = 0.09201893 \approx 0.0920, \quad s_u \approx 0.3033$$

$$95 \cdot (-0.03139) + 1993b - 0.10313 \cdot 308 = 1 \;\Rightarrow\; b \approx 0.0180$$

$$15 \cdot (-0.56951) + 308 \cdot (-0.10314) + 49c = 1 \;\Rightarrow\; c \approx 0.8431$$

(

$$(X^T X)^{-1} = \begin{pmatrix} 5 & 95 & 15 \\ 95 & 1993 & 308 \\ 15 & 308 & 49 \end{pmatrix}^{-1} = \begin{pmatrix} 2.504932735 & -0.031390135 & -0.569506726 \\ -0.031390135 & 0.017937220 & -0.103139013 \\ -0.569506726 & -0.103139013 & 0.843049327 \end{pmatrix}$$

)

$s_{a_1} = 0.3033 \cdot \sqrt{0.0180} \approx 0.0407$, $\quad s'_{a_1} = \dfrac{0.0407}{0.1475} \approx 0.2759 \quad \Rightarrow$ annehmbar.

$s_{a_2} = 0.3033 \cdot \sqrt{0.8431} \approx 0.2785$ $\quad s'_{a_2} = \left|\dfrac{0.2785}{-0.5233}\right| \approx 0.5322 \quad \Rightarrow$ annehmbar.

$$A_1 \in [0.1475 - 4.303 \cdot 0.0407, \quad 0.1475 + 4.303 \cdot 0.0407] \approx [-0.0276,\ 0.3226]$$

$$A_2 \in [-0.5233 - 4.303 \cdot 0.2785, \quad -0.5233 + 4.303 \cdot 0.2785] \approx [-1.7217,\ 0.6751]$$

4.

A_1:

S1:

$H_0: A_1 = 0, \quad H_1: A_1 > 0.$

S2:

Da σ_u nicht bekannt ist, wird die $t-$Verteilung benutzt.

S3:

Der Test ist wegen $H_1: A_1 > 0$ rechtsseitig. $df = 5 - 3 = 2$. Damit liegt die Ablehnungsregion rechts von 2.920.

S4:

$$t_{stat} = \frac{a_1 - A_1}{s_{a_1}} = \frac{0.1475}{0.0407} \approx 3.62.$$

S5:

Wegen $3.62 > 2.920$ wird die Nullhypothese abgelehnt.

A_2:

S1:

$H_0: A_2 = 0, \quad H_1: A_2 < 0.$

S2:

Da σ_u nicht bekannt ist, wird die $t-$Verteilung benutzt.

S3:

Der Test ist wegen $H_1: A_2 < 0$ linksseitig. $df = 5 - 3 = 2$. Damit liegt die Ablehnungsregion links von -2.920.

S4:

$$t_{stat} = \frac{a_2 - A_2}{s_{a_2}} = -\frac{0.5233}{0.2785} \approx -1.8790\,.$$

S5:

Wegen $-1.8790 > -2.920$ wird die Nullhypothese nicht abgelehnt.

5.

y_i	x_{i1}	x_{i2}	x_{i1}^2	x_{i2}^2	$x_{i1}y_i$	$x_{i2}y_i$	$x_{i1}x_{i2}$
745	36	66	1296	4356	26820	49170	2376
895	37	68	1369	4624	33115	60860	2516
442	47	64	2209	4096	20774	28288	3008
440	32	53	1024	2809	14080	23320	1696
1598	1	101	1	10201	1598	161398	101
4120	153	352	5899	26086	96387	323036	9697

$$\begin{cases} 5a_0 + 153a_1 + 352a_2 = 4120 \\ 153a_0 + 5899a_1 + 9697a_2 = 96387 \\ 352a_0 + 9697a_1 + 26086a_2 = 323036 \end{cases}$$

x_1	x_2	x_3	x_0
5	153	352	4120
153	5899	9697	96378
352	9697	26086	323036
1	30.6000	70.4000	824
0	1217.2000	- 1074.2000	-29694
0	-1074.2000	1305.2000	32988
1	0	97.4050	1570.2710
0	1	-0.8825	-24.3879
0	0	357.2000	6790.4754
1	0	0	-281.4270
0	1	0	-7.6110
0	0	1	19.0103

$a_0 = -281.4270$

$a_1 = -7.6110$

$a_2 = 19.0103$

$$y^* = -281.4270 - 7.6110x_1 + 19.0103x_2$$

2.

$$\bar{y} = \frac{4120}{5} = 824$$

Arbeitstabelle 2

y_i^*	$(y_i^* - \bar{y})^2$	$(y_i - \bar{y})^2$	u_i^{*2}
699.2568	15560.8659	6241	2092.44035
729.6664	8898.82809	5041	27335.1993
577.5152	60754.7566	145924	18364.3694
482.5669	116576.562	147456	1811.94098
1631.0023	651252.712	599076	1089.15181
	853043.725	903738	50693.1018

$$r^2 = \frac{853043.725}{903738} = 0.943906 \approx 0.94.$$

Der Preis ist damit zu etwa 94% durch das Alter und die Fläche der Ferienwohnungen erklärt (eine sehr gute Approximation).
3.

$$s_u^2 = \frac{50693.1018}{5-(2+1)} = 25346.5509, \qquad s_u = 159.2060015 \approx 159.2060$$

$$153 \cdot (-0.26579) + 5899b + 9697 \cdot 0.00247 = 1 \quad \Rightarrow \quad b \approx 0.0030$$

$$352 \cdot (-0.27269) + 9697 \cdot 0.00247 + 26086c = 1 \quad \Rightarrow \quad c \approx 0.0028$$

(

$$(X^T X)^{-1} = \begin{pmatrix} 5 & 153 & 352 \\ 153 & 5899 & 9697 \\ 352 & 9697 & 26086 \end{pmatrix}^{-1} = \begin{pmatrix} 27.530696866 & -0.265793712 & 0.272690473 \\ -0.265793712 & 0.003001952 & 0.002470653 \\ 0.272690473 & 0.002470653 & 0.002799552 \end{pmatrix}$$

)

$$s_{a_1} = 159.2060 \cdot \sqrt{0.0030} \approx 8.7200, \qquad s'_{a_1} = \left|\frac{8.7200}{-7.611}\right| \approx 1.1457 \quad \Rightarrow \text{ nicht annehmbar.}$$

$$s_{a_2} = 159.2060 \cdot \sqrt{0.0028} \approx 8.4244 \qquad s'_{a_2} = \frac{8.4244}{19.0103} \approx 0.4431 \quad \Rightarrow \text{ annehmbar.}$$

$$A_1 \in [-7.6110 - 4.303 \cdot 8.7200, \quad -7.6110 + 4.303 \cdot 8.7200] \approx [-45.1332, \ 29.9111]$$

$$A_2 \in [19.0103 - 4.303 \cdot 8.4244, \quad 19.0103 + 4.303 \cdot 8.4244] \approx [-17.2399, \ 55.2605]$$

4.

A_1 :

S1:

$H_0: A_1 = 0, \quad H_1: A_1 < 0.$

S2:

Da σ_u nicht bekannt ist, wird die $t-$Verteilung benutzt.

S3:

Der Test ist wegen $H_1: A_1 < 0$ linksseitig. $df = 5-3 = 2$. Damit liegt die Ablehnungsregion links von -2.920.

S4:

$$t_{stat} = \frac{a_1 - A_1}{s_{a_1}} = -\frac{7.6110}{8.7200} \approx -0.873 .$$

S5:

Wegen $-0.873 > -2.920$ wird die Nullhypothese nicht abgelehnt.

A_2 :

S1:

$H_0: A_2 = 0, \quad H_1: A_2 > 0.$

S2:

Da σ_u nicht bekannt ist, wird die $t-$Verteilung benutzt.

S3:

Der Test ist wegen $H_1: A_2 > 0$ rechtsseitig. $df = 5-3 = 2$. Damit liegt die Ablehnungsregion rechts von 2.920.

S4:

$$t_{stat} = \frac{a_2 - A_1}{s_{a_2}} = \frac{19.0103}{8.4244} \approx 2.2666 .$$

S5:

Wegen $2.266 < 2.920$ wird die Nullhypothese nicht abgelehnt.

B. SPSS-Aufgaben

1.

- ***Analysieren -> Regression -> Linear…***

- Übertragen Sie die Variable *cpr* in das Eingabefeld ***Abhängige Variable*** und die Variablen *yverf* und *zins* für ***Block 1 von 1*** (für die erste Schätzgleichung) in das Eingabefeld ***Unabhängige***.

  Zur gleichzeitigen Berechnung der zweiten Schätzgleichung wählen Sie:

  Nächste

- Übertragen Sie die zusätzliche unabhängige Variable *lq*

  (Aus dem Auswahlfeld ***Methode*** können andere Verfahren zum Einschluss der unabhängigen Variablen in die Regressionsgleichung gewählt werden. Hier für beide Blöcke ***Einschluss***. Die Methode ***Einschluss*** ist die Standardeinstellung und bedeutet, dass die gewählten unabhängigen Variablen in einem Schritt in die Regressionsgleichung eingeschlossen werden.)

  Statistiken

- (Im Dialogfeld ***Lineare Regression: Statistiken*** sind ***Schätzungen*** und ***Anpassungsgüte des Modells*** voreingestellt.)
  Wählen Sie noch:

  Konfidenzintervalle, Deskriptive Statistik, Teil- und partielle Korrelationen und ***Kollinearitätsdiagnose***.

  Weiter

- ***Diagramme***

  Übertragen Sie die Variable *ZRESID in das Feld ***Y*** und die Variable *ZPRED in das Feld ***X*** und wählen Sie die Optionen ***Histogramm***, ***Alle partiellen Diagramme erzeugen*** und ***Normalverteilungsdiagramm***.

  Weiter
- ***Speichern***
  Wählen Sie **folgende** Optionen:

Bei

Vorhersagewerten: Nicht standardisiert

Residuen: Nicht standardisiert, Studentisiert

Vorhersageintervallen: Mittelwert, Individuell

Weiter

- ***OK***

Ausgabe und Interpretation

Koeffizienten

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	Kollinearitätsstatistik	
		Regressionskoeffizient B	Standardfehler	Beta			Toleranz	VIF
1	(Konstante)	51.765	10.561		4.902	0.000		
	yverf	0.862	0.007	1.007	127.646	0.000	0.928	1.078
	zins	-5.313	1.355	-0.31	-3.920	0.001	0.928	1.078
2	(Konstante)	-18.077	45.006		-0.402	0.691		
	yverf	0.844	0.013	0.986	64.897	0.000	0.237	4.219
	zins	-7.031	1.704	-0.041	4.127	0.000	0.557	1.797
	lq	1.424	0.893	0.028	1.594	0.123	0.176	5.669

1.
In der obigen Tabelle werden die Regressionskoeffizienten der beiden Regressionsmodelle sowie Angaben zu Signifikanzprüfungen der Koeffizienten aufgeführt.

Modell 1:

$$y^* = 51.767 + 0.862x_1 - 5.313x_2$$

Modell 1:

$$y^* = -18.077 + 0.844x_1 - 7.031x_2 + 1.424x_3$$

2.

- Eine Erhöhung von <u>nur</u> Einkommen um eine Einheit hat zur Folge, dass sich der Privatkonsum um 0.862 Einheiten erhöht.
- Eine Erhöhung von <u>nur</u> Zins um eine Einheit hat zur Folge, dass sich der Privatkonsum um 7.031 Einheiten reduziert.
- Eine Erhöhung von <u>nur</u> Lohnquote um eine Einheit hat zur Folge, dass sich der Privatkonsum um 1.424 Einheiten erhöht.

3.
Modell 1:

$A_0: \quad p-value = 0.000 < 0.05:$ signifikant
$A_1: \quad p-value = 0.000 < 0.05:$ signifikant

Modell 2:

$A_0: \quad p-value = 0.691 \geq 0.05:$ nicht signifikant
$A_1: \quad p-value = 0.000 < 0.05:$ signifikant
$A_2: \quad p-value = 0.000 < 0.05:$ signifikant
$A_3: \quad p-value = 0.123 \geq 0.05:$ nicht signifikant

4.

(In der Spalte ***Beta*** werden die Beta-Koeffizienten (es sind standardisierte Regressionskoeffizienten) für die Erklärungsvariablen aufgeführt. Beta-Koeffizienten würden sich als Regressionskoeffizienten ergeben, wenn vor der Anwendung der linearen Regression alle Variablen standardisierte (in z-Werte transformiert):

$$z = \frac{x - \bar{x}}{s}.$$

Mit der Standardisierung werden die Abweichungen der Messwerte der Variablen von ihrem Mittelwert in Standardabweichungen ausgedrückt. Sie sind dann dimensionslos.
Im Unterschied zu den Regressionskoeffizienten sind die Beta-Koeffizienten deshalb von der Dimension der erklärenden Variablen unabhängig und daher miteinander vergleichbar. Es zeigt sich, dass der absolute Beta-Koeffizient für das verfügbare Einkommen den für den Zinssatz bei weitem übersteigt. Damit wird sichtbar, dass das verfügbare Einkommen als bedeutsamste Variable den weitaus größten Erklärungsbeitrag liefert. Aus den unstandardisierten Regressionskoeffizienten für die beiden Variablen ist dieses nicht erkennbar. Aufgrund der Größenverhältnisse dieser Regressionskoeffizienten könnte man

das Gegenteil vermuten. Allerdings darf bei dieser vergleichenden Beurteilung der relativen Bedeutung der Variablen zur statistischen Erklärung nicht übersehen werden, dass auch die Beta-Koeffizienten durch Multikollinearität nicht unabhängig voneinander und insofern in ihrer Aussagekraft eingeschränkt sind.)

5.

Modellzusammenfassung[c]

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,999[a]	,9984	,9983	9,5180
2	,999[b]	,9985	,9984	9,2664

a. Einflußvariablen : (Konstante), Zinssatz (%), verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)

b. Einflußvariablen : (Konstante), Zinssatz (%), verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM), Lohnquote (%)

c. Abhängige Variable: Privater Konsum, in Preisen 1985 (Mrd. DM)

Mit den Bestimmtheitsmaßen von 0.9984 bzw. 0.9995 ist der Werte für beide Modelle nahezu 1, so dass fast die gesamte Variation von *cpr* im Modell 1 durch die Variation von *yverf* und *zins* und im Modell 2 durch *yverf* , *zins* und *lq* erklärt wird. Das korrigierte Bestimmtheitsmaß ist ein Bestimmtheitsmaß, das die Anzahl der erklärenden Variablen sowie die Anzahl der Beobachtungen berücksichtigt.

6.

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	1566437,632	2	783218,816	8645,589	,000[b]
	Nicht standardisierte Residuen	2536,568	28	90,592		
	Gesamt	1568974,200	30			
2	Regression	1566655,812	3	522218,604	6081,767	,000[c]
	Nicht standardisierte Residuen	2318,389	27	85,866		
	Gesamt	1568974,200	30			

a. Abhängige Variable: Privater Konsum, in Preisen 1985 (Mrd. DM)

b. Einflußvariablen : (Konstante), Zinssatz (%), verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)

c. Einflußvariablen : (Konstante), Zinssatz (%), verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM), Lohnquote (%)

Modell 1:

$$F_{stat} = \frac{\frac{1566437.632}{2}}{\frac{2536.568}{31-(2+1)}} = 8645.589$$

$$F_{krit(\alpha=0.05;df_1=2;df_2=28)} = 3.34$$

Wegen $F_{stat} > F_{krit}$ wird die Nullhypothese $A_1 = A_2 = 0$ abgelehnt. Dies wird auch durch das Ergebnis $Sig = p - value = 0 < 0.05 = \alpha$ bestätigt.

Modell 2:

$$F_{stat} = \frac{\frac{1566655.812}{3}}{\frac{2318.389}{31-(3+1)}} = 6081.767$$

$$F_{krit(\alpha=0.05;df_1=3;df_2=27)} = 2.96$$

Wegen $F_{stat} > F_{krit}$ wird die Nullhypothese $A_1 = A_2 = 0$ abgelehnt. Dies wird auch durch das Ergebnis $Sig = p - value = 0 < 0.05 = \alpha$ bestätigt.

7.
Ausschnitte aus der Tabelle „**Koeffizienten**“

Modell 1:

	95% Konfidenzintervall	
Modell 1	Untergrenze	Obergrenze
(Konstante)	30.133	73.401
yverf	0.848	0.876
zins	-8.089	-2.536

Modell 2:

	95% Konfidenzintervall	
Modell 2	Untergrenze	Obergrenze
(Konstante)	-110.423	74.269
yverf	0.817	0.871
zins	-10.526	-3.535
	-0.409	3.257

8.
Durch Kovarianz. Dazu gibt es folgende Tabelle:

Korrelation der Koeffizienten[a]

Modell			Zinssatz (%)	verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)	Lohnquote (%)
1	Korrelationen	Zinssatz (%)	1,000	-,269	
		verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)	-,269	1,000	
	Kovarianzen	Zinssatz (%)	1,837	-,002	
		verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)	-,002	4,560E-5	
2	Korrelationen	Zinssatz (%)	1,000	,440	-,632
		verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)	,440	1,000	-,863
		Lohnquote (%)	-,632	-,863	1,000
	Kovarianzen	Zinssatz (%)	2,903	,010	-,963
		verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)	,010	,000	-,010
		Lohnquote (%)	-,963	-,010	,798

a. Abhängige Variable: Privater Konsum, in Preisen 1985 (Mrd. DM)

- Wenn beide Variablen gleichzeitig steigen oder fallen, ist der Koeffizient positiv. Z.B. im Modell 2 die Beziehung zwischen Zinssatz und Einkommen).
- Wenn die eine Variable steigt und die andere fällt, ist der Koeffizient negativ. Z.B, im Modell 2 die Beziehung zwischen Zinssatz und Lohnquote).

9.

Kollinearitätsdiagnose[a]

Modell	Dimension	Eigenwert	Konditionsindex	Varianzanteile (Konstante)	verfügbares Einkommen, in Kaufkraft 1985 (Mrd. DM)	Zinssatz (%)	Lohnquote (%)
1	1	2,941	1,000	,00	,01	,00	
	2	,044	8,174	,08	,99	,12	
	3	,015	13,981	,91	,01	,88	
2	1	3,938	1,000	,00	,00	,00	,00
	2	,045	9,347	,00	,27	,05	,00
	3	,017	15,253	,02	,00	,60	,00
	4	,000	92,861	,97	,73	,35	1,00

a. Abhängige Variable: Privater Konsum, in Preisen 1985 (Mrd. DM)

Nur im Modell 2 liegt in diesem Falle eine sehr starke Kollinearität vor, woanders sehr schwache.

2.

Koeffizienten

Modell		Nicht standardisierte Koeffizienten RegressionskoeffizientB	Standardfehler	Standardisierte Koeffizienten Beta	T	Sig.
1	(Konstante)	28,828	3,773		7,641	,000
	Stadtbevölkerung (%)	,097	,033	,240	2,943	,004
	Alphabetisierungsrate (%)	,217	,036	,494	6,090	,000
	Bruttoinlandsprodukt / Kopf	-6,516E-6	,000	-,005	-,054	,957
	Tägliche Kalorienaufnahme	,005	,002	,269	2,808	,006

In der obigen Tabelle werden die Regressionskoeffizienten der beiden Regressionsmodelle sowie Angaben zu Signifikanzprüfungen der Koeffizienten aufgeführt.

$$y^* = 28.828 + 0.097x_1 + 0.217x_2 - 6.516\mathrm{E} - 6x_3 + 0.005x_4$$

Wir sehen, dass der Koeffizient von x_3 praktisch gleich Null ist.

2.

$$A_0: \quad p - value = 0.000 < 0.05: \quad \text{signifikant}$$
$$A_1: \quad p - value = 0.004 < 0.05: \quad \text{signifikant}$$
$$A_2: \quad p - value = 0.000 < 0.05: \quad \text{signifikant}$$
$$A_3: \quad p - value = 0.957 \geq 0.05: \quad \text{nicht signifikant}$$
$$A_4: \quad p - value = 0.006 < 0.05: \quad \text{signifikant}$$

3.

Modellzusammenfassung[b]

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,893[a]	,798	,786	4,695

a. Einflußvariablen : (Konstante), Tägliche Kalorienaufnahme, Alphabetisierungsrate (%), Stadtbevölkerung (%), Bruttoinlandsprodukt / Kopf

b. Abhängige Variable: Durchschnittliche Lebenserwartung für Männer

Mit dem Bestimmtheitsmaß von 0.798 ist der Wert gut, so dass die 4 Faktoren (Anteil der Bevölkerung, die in den Städten leben, Anteil der Bevölkerung, die lesen kann, Bruttosozialprodukt je Einwohner und die durchschnittliche Kalorienaufnahme je Person und Tag) zu etwa 79.8% die Variation der Lebenserwartung eines Mannes bestimmen. Das korrigierte Bestimmtheitsmaß 0.786 berücksichtigt die Anzahl der erklärenden Variablen sowie die Anzahl der Beobachtungen berücksichtigt.

4.

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	6016,202	4	1504,051	68,224	,000[b]
	Nicht standardisierte Residuen	1521,149	69	22,046		
	Gesamt	7537,351	73			

a. Abhängige Variable: Durchschnittliche Lebenserwartung für Männer

b. Einflußvariablen : (Konstante), Tägliche Kalorienaufnahme, Alphabetisierungsrate (%), Stadtbevölkerung (%), Bruttoinlandsprodukt / Kopf

$$F_{stat} = \frac{\frac{6016.202}{4}}{\frac{1521.149}{69}} = 68.224$$

$$F_{krit(\alpha=0.05;df_1=4;df_2=69)} = 2.5046$$

Wegen $F_{stat} > F_{krit}$ wird die Nullhypothese $A_1 = A_2 = A_3 = A_4 = 0$ abgelehnt. Dies wird auch durch das Ergebnis $Sig = p - value = 0 < 0.05 = \alpha$ bestätigt.

5.

Ausschnitt aus der Tabelle „**Koeffizienten**“

Modell		Sig.	95,0% Konfidenzintervalle für B	
			Untergrenze	Obergrenze
1	(Konstante)	,000	21,302	36,354
	Stadtbevölkerung (%)	,004	,031	,163
	Alphabetisierungsrate (%)	,000	,146	,288
	Bruttoinlandsprodukt / Kopf	,957	,000	,000
	Tägliche Kalorienaufnahme	,006	,001	,008

6.
Durch Kovarianz. Dazu gibt es folgende Tabelle:

Korrelation der Koeffizienten[a]

Modell			Tägliche Kalorienaufnahme	Alphabetisierungsrate (%)	Stadtbevölkerung (%)	Bruttoinlandsprodukt / Kopf
1	Korrelationen	Tägliche Kalorienaufnahme	1,000	-,250	-,311	-,517
		Alphabetisierungsrate (%)	-,250	1,000	-,353	-,176
		Stadtbevölkerung (%)	-,311	-,353	1,000	-,109
		Bruttoinlandsprodukt / Kopf	-,517	-,176	-,109	1,000
	Kovarianzen	Tägliche Kalorienaufnahme	2,991E-6	-1,542E-5	-1,778E-5	-1,085E-7
		Alphabetisierungsrate (%)	-1,542E-5	,001	,000	-7,615E-7
		Stadtbevölkerung (%)	-1,778E-5	,000	,001	-4,359E-7
		Bruttoinlandsprodukt / Kopf	-1,085E-7	-7,615E-7	-4,359E-7	1,474E-8

a. Abhängige Variable: Durchschnittliche Lebenserwartung für Männer

- Wenn beide Variablen gleichzeitig steigen oder fallen, ist der Koeffizient positiv.
- Wenn die eine Variable steigt und die andere fällt, ist der Koeffizient negativ.

7.

Kollinearitätsdiagnose[a]

Modell	Dimension	Eigenwert	Konditionsindex	Varianzanteile (Konstante)	Stadtbevölkerung (%)	Alphabetisierungsrate (%)	Bruttoinlandsprodukt / Kopf	Tägliche Kalorienaufnahme
1	1	4,465	1,000	,00	,00	,00	,01	,00
	2	,427	3,234	,01	,00	,00	,44	,00
	3	,071	7,927	,05	,77	,00	,17	,01
	4	,029	12,411	,09	,14	,99	,05	,03
	5	,008	23,074	,85	,09	,01	,34	,96

a. Abhängige Variable: Durchschnittliche Lebenserwartung für Männer

Alle Werte in dieser Spalte sind kleiner als 30. Es gibt keine starke Kollinearität.

(Letzte Aktualisierung 08.12.23)