

## Übung 3: Allgemeines lineares Modell & Modellwahl und Variablenselektion

### Aufgabe 1

Betrachtet wird das lineare Modell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

für die Beobachtungen:

$y$	$x_1$	$x_2$
120	4	12
150	6	18
180	5	15
160	8	25

- a) Berechnen Sie den KQ-Schätzer für  $\beta = (\beta_0, \beta_1, \beta_2)^T$  und die Residuen  $\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \hat{\varepsilon}_3, \hat{\varepsilon}_4)^T$ . Verwenden Sie dafür:

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} \frac{77}{6} & -24 & \frac{43}{6} \\ -24 & 62 & -19 \\ \frac{43}{6} & -19 & \frac{35}{6} \end{pmatrix}$$

- b) Berechnen Sie den Aitken-Schätzer für  $\beta = (\beta_0, \beta_1, \beta_2)^T$ , die Residuen  $\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \hat{\varepsilon}_3, \hat{\varepsilon}_4)^T$  und  $\hat{\sigma}^2$  unter der Voraussetzung, dass

$$\varepsilon \sim N \left( 0, \sigma^2 \begin{pmatrix} 0,5 & 0 & 0 & 0 \\ 0 & 0,25 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \right)$$

gilt. Verwenden Sie dafür:

$$(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} = \begin{pmatrix} \frac{201}{38} & -\frac{161}{19} & 2,5 \\ -\frac{161}{19} & \frac{1001}{19} & -17 \\ 2,5 & -17 & 5,5 \end{pmatrix}$$

- c) Vergleichen Sie die berechneten Residuen aus den Aufgabenteilen a) und b) miteinander.

## Aufgabe 2

Ein Unternehmen möchte ein klassisches lineares Modell für den Umsatz  $y$  aufstellen. Dafür wurden der Umsatz der letzten 20 Jahre sowie die Ausprägungen von vier weiteren erklärenden Variablen erhoben, die in der folgenden Tabelle zusammengefasst sind:

$i$	$y_i$	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$i$	$y_i$	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$i$	$y_i$	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$
1	190	5	17	24	66	8	423	52	46	19	29	15	510	52	54	7	9
2	96	5	8	-13	45	9	405	47	42	57	23	16	525	62	59	63	48
3	255	37	33	5	20	10	185	23	20	42	36	17	380	48	46	9	17
4	382	42	43	17	19	11	70	11	8	43	15	18	30	6	3	16	11
5	270	29	26	4	32	12	123	16	18	19	55	19	210	14	17	-53	37
6	303	42	40	37	41	13	120	7	10	-17	42	20	116	12	13	-35	33
7	303	31	29	-19	13	14	600	41	46	-25	18						

Der Vorstand des Unternehmens zieht einen externen Berater hinzu, der mithilfe der Daten ein geeignetes lineares Regressionsmodell aufstellen soll. Dieser passt ein lineares Regressionsmodell mit Berücksichtigung aller vier erklärenden Variablen an die Daten an sowie alle möglichen Submodelle. Die resultierenden Ergebnisse sind in der nachfolgenden Tabelle angegeben.

- Welche der erklärenden Variablen würden sich für eine einfache lineare Regression eignen?
- Welches Modell würde der Experte dem Vorstand empfehlen, wenn er als Auswahlkriterium das adjustierte Bestimmtheitsmaß verwendet? Würde sich auch das gewöhnliche Bestimmtheitsmaß für die Modellauswahl eignen?
- Wie würde sich die Entscheidung aus Aufgabenteil b) ändern, wenn statt dem adjustierten Bestimmtheitsmaß eines der beiden Informationskriterien AIC und BIC für die Modellauswahl verwendet wird?
- Führen Sie mit Hilfe der drei Modellauswahlkriterien  $R_a^2$ , AIC und BIC jeweils eine Vorwärts-Selektion durch und starten Sie dabei jeweils beim Nullmodell. Für welches Modell würden Sie sich jeweils entscheiden? Stimmen diese Modelle mit den globalen Lösungen aus den Aufgabenteilen b) und c) überein?
- Führen Sie mit Hilfe der drei Modellauswahlkriterien  $R_a^2$ , AIC und BIC jeweils eine Rückwärts-Selektion durch und starten Sie dabei mit dem Modell 16 (volles Modell). Für welches Modell würden Sie sich jeweils entscheiden? Stimmen diese Modelle mit den globalen Lösungen aus den Aufgabenteilen b) und c) überein?
- Vergleichen Sie die  $p$ -Werte der Schätzungen  $\hat{\beta}_1$  und  $\hat{\beta}_2$  in den verschiedenen Modellen. Was fällt auf und wodurch werden diese Unregelmäßigkeiten verursacht?

Modell	Intercept		$x_1$		$x_2$		$x_3$		$x_4$		$R^2$	$R_a^2$	AIC	BIC
	$\hat{\beta}_0$	P-Wert	$\hat{\beta}_1$	p-Wert	$\hat{\beta}_2$	p-Wert	$\hat{\beta}_3$	p-Wert	$\hat{\beta}_4$	p-Wert				
1	274.80	$4.47 \cdot 10^{-07}$	-	-	-	-	-	-	-	-	0	0	263.7907	265.7822
2	44.2433	0.175	7.9229	$7.63 \cdot 10^{-08}$ ***	-	-	-	-	-	-	0.8069	0.7961	232.903	235.8902
3	10.9296	0.643 *	-	-	9.1305	$1.18 \cdot 10^{-10}$ ***	-	-	-	-	0.9054	0.9001	218.6332	221.6204
4	266.2035	$2.41 \cdot 10^{-06}$ ***	-	-	-	-	0.8597	0.501	-	-	0.0255	-0.02864	265.274	268.2612
5	358.883	0.000312 ***	-	-	-	-	-	-	-2.761	0.258456	0.07033	0.01869	264.3321	267.3193
6	2.318	0.9166	-5.478	0.0571 .	14.944	$8.55 \cdot 10^{-05}$ ***	-	-	-	-	0.924	0.9151	216.2488	220.2317
7	34.8241	0.2486	8.6423	$3.19 \cdot 10^{-08}$ ***	-	-	-1.1514	0.0532 .	-	-	0.846	0.8278	230.3785	234.3615
8	30.1890	0.594	8.0247	$3.18 \cdot 10^{-07}$ ***	-	-	-	-	0.3643	0.760	0.808	0.7854	234.7897	238.7727
9	5.7046	0.7900	-	-	9.6012	$5.23 \cdot 10^{-11}$ ***	-0.8379	0.0374 *	-	-	0.9272	0.9186	215.3902	219.3731
10	26.5287	0.49	-	-	9.0373	$6.37 \cdot 10^{-10}$ ***	-	-	-0.4239	0.60	0.9069	0.896	220.3	224.2829
11	354.809	0.000438 ***	-	-	-	-	1.015	0.424676	-2.961	0.234225	0.1055	0.0003089	265.56	269.5429
12	0.8369	0.968360	-3.8630	0.176062	13.5925	0.000241 ***	-0.6446	0.114080	-	-	0.9353	0.9232	215.0302	220.0088
13	49.3604	0.1531	-7.8339	0.0137 *	17.1417	$3.09 \cdot 10^{-05}$ ***	-	-	-1.3790	0.0867 .	0.9371	0.9253	214.4652	219.4439
14	-4.9162	0.926	9.0080	$1.09 \cdot 10^{-07}$ ***	-	-	-1.2892	0.040 *	1.0008	0.373	0.8536	0.8262	231.3565	236.3352
15	10.3728	0.7713	-	-	9.5667	$3.87 \cdot 10^{-10}$ ***	-0.8252	0.0501 .	-0.1247	0.8676	0.9273	0.9137	217.3544	222.333
16	36.0955	0.323572	-6.1495	0.079963 .	15.6730	0.000272 ***	-0.4268	0.325150	-1.0189	0.240419	0.9412	0.9255	215.1309	221.1053

Modell	Variablen	$R_a^2$	AIC	BIC
1	-			

Modell	Variablen	$R_a^2$	AIC	BIC
2	$x_1$			
3	$x_2$			

Modell	Variablen	$R_a^2$	AIC	BIC
4	$x_3$			
5	$x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
6	$x_1, x_2$			
7	$x_1, x_3$			

Modell	Variablen	$R_a^2$	AIC	BIC
8	$x_1, x_4$			
9	$x_2, x_3$			

Modell	Variablen	$R_a^2$	AIC	BIC
10	$x_2, x_4$			
11	$x_3, x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
12	$x_1, x_2, x_3$			
13	$x_1, x_2, x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
14	$x_1, x_3, x_4$			
15	$x_2, x_3, x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
16	$x_1, x_2, x_3, x_4$			

Im Modell vorhandene Variablen	Schritte des Auswahlverfahrens					
	Start-Modell	1	2	3	4	End-Modell
	$R_a^2$	-				
	AIC	-				
BIC	-					

Modell	Variablen	$R_a^2$	AIC	BIC
1	-			

Modell	Variablen	$R_a^2$	AIC	BIC
2	$x_1$			
Modell	Variablen	$R_a^2$	AIC	BIC
3	$x_2$			

Modell	Variablen	$R_a^2$	AIC	BIC
4	$x_3$			
Modell	Variablen	$R_a^2$	AIC	BIC
5	$x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
6	$x_1, x_2$			
Modell	Variablen	$R_a^2$	AIC	BIC
7	$x_1, x_3$			

Modell	Variablen	$R_a^2$	AIC	BIC
8	$x_1, x_4$			
Modell	Variablen	$R_a^2$	AIC	BIC
9	$x_2, x_3$			

Modell	Variablen	$R_a^2$	AIC	BIC
10	$x_2, x_4$			
Modell	Variablen	$R_a^2$	AIC	BIC
11	$x_3, x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
12	$x_1, x_2, x_3$			
Modell	Variablen	$R_a^2$	AIC	BIC
13	$x_1, x_2, x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
14	$x_1, x_3, x_4$			
Modell	Variablen	$R_a^2$	AIC	BIC
15	$x_2, x_3, x_4$			

Modell	Variablen	$R_a^2$	AIC	BIC
16	$x_1, x_2, x_3, x_4$			

Im Modell vorhandene Variablen	Schritte des Auswahlverfahrens					
	Start-Modell	1	2	3	4	End-Modell
	$R_a^2$	$x_1, x_2, x_3, x_4$				
	AIC	$x_1, x_2, x_3, x_4$				
BIC	$x_1, x_2, x_3, x_4$					