Übung 1: Mathematische und statistische Grundlagen

Aufgabe 1

Berechnen Sie die Schiefe und Kurtosis folgender Zufallsvariablen:

a) X besitzt die Dichte

$$f_X(x) = \begin{cases} 2x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases} .$$

b) $X \sim \mathrm{Bin}(4; 0{,}3)$ und besitzt damit die Wahrscheinlichkeitsfunktion

$$f_X(x) = \begin{cases} \binom{4}{x} 0{,}3^x 0{,}7^{4-x} & \text{für } x \in \{0, 1, 2, 3, 4\} \\ 0 & \text{sonst} \end{cases} .$$

Aufgabe 2

Eine faire Münze wird 1000 mal geworfen. Approximieren Sie mit Hilfe des zentralen Grenzwertsatzes die Wahrscheinlichkeit, dass häufiger als 530 mal Kopf fällt.

Aufgabe 3

Der Zufallsvektor $\mathbf{X} = (X_1, X_2, X_3)^T$ sei dreidimensional $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-verteilt mit

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} \qquad \text{und} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix} .$$

a) Geben Sie die Verteilung von X_1 an.

b) Ermitteln Sie die Verteilung des zweidimensionalen Zufallsvektors $(X_1, X_2)^T$, bedingt gegeben $X_3 = 3$.

Übung 2: Klassisches lineares Modell I

Aufgabe 1

Ein Mitarbeiter in der Marketingabteilung eines Finanzdienstleisters vermutet, dass es einen positiven Zusammenhang zwischen den Ausgaben (in Millionen Euro) für Werbung und dem Absatz (in Millionen Euro) des beworbenen Finanzproduktes gibt. In den letzten drei Jahren wurden folgende Zahlen beobachtet:

Absatz y	Ausgaben x
5	2
8	3
9	4

Es wird im Folgenden davon ausgegangen, dass ein linearer Zusammenhang zwischen dem Absatz und den Ausgaben vorliegt.

a) Bestimmen Sie die Designmatrix $\mathbf{X}$ und den KQ-Schätzer $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_0, \widehat{\beta}_1\right)^T$.

b) Bestimmen Sie eine Schätzung für die Varianz-Kovarianzmatrix des KQ-Schätzers $\widehat{\boldsymbol{\beta}}$.

c) Ermitteln Sie den Wert des Bestimmtheitsmaßes R^2 und des adjustierten Bestimmtheitsmaßes R_a^2.

d) Testen Sie die Signifikanz der erklärenden Variablen x bei einem Signifikanzniveau von $\alpha = 0{,}05$ und $\alpha = 0{,}1$.

e) Ermitteln Sie für den Regressionskoeffizienten β_1 das 95%-Konfidenzintervall.

Aufgabe 2

Es soll der Zusammenhang zwischen einer abhängigen Variablen y und zwei erklärenden Variablen x_1 und x_2 untersucht werden. Bei x_1 handelt es sich dabei um ein intervallskaliertes Merkmal und bei x_2 um eine kategoriale Variable, die drei Ausprägungen annehmen kann. Für die Variablen y, x_1 und x_2 liegen folgende Beobachtungen vor:

y	x_1	x_2
5	4	1
3	3	1
9	5	2
10	6	2
10	2	3
15	5	3

a) Stellen Sie ein geeignetes Regressionsmodell auf und wählen Sie dabei $x_2 = 1$ als Referenzklasse der kategorialen Variablen.

b) Berechnen Sie den KQ-Schätzer für das in a) aufgestellte Regressionsmodell und verwenden Sie dafür:

$$\left(X^T X\right)^{-1} = \frac{1}{22} \begin{pmatrix} 60 & -14 & 17 & -11 \\ -14 & 4 & -8 & 0 \\ 17 & -8 & 38 & 11 \\ -11 & 0 & 11 & 22 \end{pmatrix}$$

c) Testen Sie die Signifikanz der erklärenden Variablen x_2 bei einem Signifikanzniveau von $\alpha = 0{,}01$.

Übung 3: Klassisches lineares Modell II

Aufgabe 1

Es wird ein lineares Modell mit sechs zu schätzenden Parametern $\beta_0, \ldots, \beta_5$ betrachtet.

a) Ermitteln Sie den Wert der F-Statistik für die Testsituation

$$H_0 : (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T = \mathbf{0} \quad \text{gegen} \quad H_1 : (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T \neq \mathbf{0}$$

und den Fall, dass $n = 40$ Beobachtungen vorliegen und das Bestimmtheitsmaß $R^2 = 0{,}2$ beträgt. Beurteilen Sie anschließend, ob die Nullhypothese H_0 bei einem Signifikanzniveau von $\alpha = 0{,}05$ zu verwerfen ist.

b) Es liegen nun $n = 400$ Beobachtungen vor. Beurteilen Sie, ob die Nullhypothese H_0 jetzt zu verwerfen ist.

c) Der KQ-Schätzer für $\boldsymbol{\beta}$ sei nun durch $\widehat{\boldsymbol{\beta}} = (2, 2, 3, 3, 4, 1)^T$ gegeben. Berechnen Sie das 95%-Konfidenzintervall für die Linearkombination $\mathbf{c}^T\widehat{\boldsymbol{\beta}}$ mit $\mathbf{c}^T = (1, 2, 1, 4, 5, 3)$ und $\widehat{\sigma}\left(\mathbf{c}^T\widehat{\boldsymbol{\beta}}\right) = 0{,}9$ für den Fall, dass $n = 40$ Beobachtungen vorliegen.

d) Der KQ-Schätzer für $\boldsymbol{\beta}$ sei wieder durch $\widehat{\boldsymbol{\beta}} = (2, 2, 3, 3, 4, 1)^T$ gegeben und $n = 40$. Ferner gelte für den Vektor mit den Werten der erklärenden Variablen $\mathbf{x}_* = (1, 1, 3, 2, 2, 1)^T$ und $\widehat{\sigma}\left(\mathbf{x}_*^T\widehat{\boldsymbol{\beta}}\right) = 0{,}3$. Als Schätzung für den Varianzparameter (mittlerer quadratischer Fehler) σ^2 liegt der Wert $\widehat{\sigma}^2 = 4$ vor. Berechnen Sie mittels diesen Angaben das $(1-\alpha)$-Prognoseintervall für y zum Signifikanzniveau $\alpha = 0{,}05$.

Aufgabe 2

Für ein lineares Modell der Form

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

wurde folgende Stichprobe erhoben:

y	x_1	x_2
120	3	10
108	5	7
92	-2	3
61	1	-12
198	-5	21
21	-2	-29

a) Prüfen Sie bei einem Signifikanzniveau von $\alpha = 0{,}025$, ob mindestens eine der beiden abhängigen Variablen einen signifikanten Einfluss auf die Zielvariable hat.

b) Führen Sie einen beidseitigen Test für die Nullhypothese

$$H_0 : \beta_1 = -1 \quad \wedge \quad \beta_2 = 1 \qquad \text{gegen} \qquad \beta_1 \neq -1 \quad \vee \quad \beta_2 \neq 1$$

durch. Kann die Nullhypothese bei einem Signifikanz von $\alpha = 0{,}025$ abgelehnt werden?

Übung 4: Allgemeines lineares Modell

Aufgabe 1

Betrachtet wird das lineare Modell

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

und die Stichprobe:

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

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

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

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

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

Verwenden Sie dafür:

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

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

Aufgabe 2

Die folgende Tabelle enthält für die zwölf Risikoklassen (i, j) eines Autohaftpflichtportfolios die Gesamtschadenhöhen $S_{i,j}$:

Fahrzeuggewicht i	Alter j			
	21-30 ($j=1$)	31-40 ($j=2$)	41-50 ($j=3$)	51-60 ($j=4$)
leicht ($i=1$)	1859	1872	1430	2028
mittel ($i=2$)	2376	1872	1716	2340
schwer ($i=3$)	2772	2184	2002	2730

a) Stellen Sie ein allgemeines lineares Modell für die logarithmierten Gesamtschadenhöhen $\ln(S_{ij})$ auf und wählen Sie dabei die Risikoklasse $(1, 1)$ als Referenzklasse.

b) Bestimmen Sie die Schätzungen für die erwarteten Gesamtschadenhöhen der zwölf Risikoklassen und berücksichtigen Sie nur Koeffizienten, die auf einem Signifikanzniveau von $\alpha = 0{,}05$ signifikant sind. Verwenden Sie hier bei die folgenden Ergebnisse:

	Estimate	Std. Error	t value	p value
(Intercept)	7.59625	0.04248	178.834	$2.06e^{-12}$ ***
Fahrzeuggewicht.mittel	0.14270	0.04248	3.360	0.015239 *
Fahrzeuggewicht.schwer	0.29685	0.04248	6.989	0.000427 ***
Alter .x2	−0.15662	0.04905	−3.193	0.018761 *
Alter .x3	−0.30440	0.04905	−6.206	0.000807 ***
Alter .x4	0.01883	0.04905	0.384	0.714344

Übung 5: Modellprüfung und Modellwahl

Aufgabe 1

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 Variablen erhoben. Man erhielt damit:

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 Modell aufstellen soll. Dieser passt ein lineares Modell mit Berücksichtigung aller vier Variablen an die Daten an sowie alle möglichen Submodelle. Die resultierenden Ergebnisse sind in der dazugehörigen Tabelle angegeben.

a) Welche der Variablen würden sich für eine einfache lineare Regression eignen?

b) Welches Modell würde der Experte dem Vorstand empfehlen, wenn er als Auswahlkriterium das adjustierte Bestimmtheitsmaß verwendet? Würde sich auch das normale Bestimmtheitsmaß für die Modellauswahl eignen?

c) Wie würde sich die Entscheidung aus Aufgabenteil b) ändern, wenn statt dem Bestimmtheitsmaß eines der beiden Informationskriterien für die Modellauswahl verwendet wird?

d) Führen Sie mithilfe der drei Modellauswahlkriterien jeweils eine Vorwärts-Selektion durch und starten Sie dabei beim Nullmodell. Für welches Modell würden Sie sich dann jeweils entscheiden? Stimmen diese Modelle mit den globalen Lösungen aus den Aufgabenteilen b) und c) überein?

d) Führen Sie mithilfe der drei Modellauswahlkriterien jeweils eine Rückwärts-Selektion durch und starten Sie dabei mit Modell 16. Für welches Modell würden Sie sich dann jeweils entscheiden? Stimmen diese Modelle mit den Globalen Lösungen aus den Aufgabenteilen b) und c) überein?

f) Vergleichen Sie die P-Werte der Regressionskoeffizienten in den verschiedenen Modellen. Was fällt an ihnen auf und woran können diese Unregelmäßigkeiten liegen?

Aufgabe 2

Es wird wieder die Situation aus Aufgabe 1 von Übungsblatt 2 betrachtet.

a) Ermitteln Sie eine Schätzung für die Varianz-Kovarianzmatrix der Residuen $\widehat{\varepsilon}$.

b) Berechnen Sie die standardisierten Residuen r_i für $i = 1, 2, 3$ und beurteilen Sie damit und unter Verwendung einer gängigen Daumenregel, ob es sich bei den Beobachtungen y_1, y_2, y_3 um Ausreißer handelt.

c) Beurteilen Sie anhand einer gängigen Daumenregel für die Hebelwerte, ob die Beobachtungen y_1, y_2, y_3 selbst-schätzend sind. Erläutern Sie ferner, ob die Anwendung dieser Daumenregel in diesem Fall sinnvoll ist.

Modell	Intercept		x_1		x_2		x_3		x_4		R^2	R^2_a	AIC	BIC
	$\widehat{\beta}_0$	P-Wert	$\widehat{\beta}_1$	p-Wert	$\widehat{\beta}_2$	p-Wert	$\widehat{\beta}_3$	p-Wert	$\widehat{\beta}_4$	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			

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			

			Schritte des Auswahlverfahrens				
		Start-Modell	1	2	3	4	End-Modell
Im Modell vorhandene Variablen	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			

			Schritte des Auswahlverfahrens				
		Start-Modell	1	2	3	4	End-Modell
Im Modell vorhandene Variablen	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					

Übung 6: Modellüberprüfung und Modellwahl II

Aufgabe 1

Ein Unternehmen möchte mithilfe eines Klassischen linearen Modell untersuchen, ob die Marketingkosten des Unternehmens einen signifikanten Einfluss auf den Umsatz des Unternehmens hat. Für die Daten aus den letzten 6 Jahren ergaben sich folgende Werte:

Umsatz y	Marketingkosten x
10	5
13	7
13	6
5	3
30	12
9	4

a) Bestimmen Sie den KQ-Schätzer $\widehat{\boldsymbol{\beta}}$.

b) Das Unternehmen vermutet, dass sich innerhalb der Daten Ausreißer befinden, die das Ergebnis verzerren könnten. Untersuchen Sie, ob man mithilfe der standardisierten Residuen eine der Beobachtungen als Ausreißer einstufen könnte.

c) Untersuchen Sie die Hebelwerte der einzelnen Beobachtungen. Prüfen Sie mithilfe der Cook-Distanz und einer bekannten Daumenregel, ob es sich bei der Beobachtung mit dem größten Hebelwert, um eine Einflussreiche Beobachtung handelt?

d) Berechnen Sie für die im Aufgabenteil c) untersuchte Beobachtung das studentisierte Residuum und vergleichen Sie das Ergebnis mit dem im Aufgabenteil b) berechneten standardisierten Residuum.

Übung 7: Verallgemeinerte lineare Modelle I

Aufgabe 1

Weisen Sie nach, dass die Geometrische Verteilung zur Exponential-Dispersions-Familie gehört und berechnen Sie den Erwartungswert und die Varianz der Verteilung.

Aufgabe 2

Ein Bankunternehmen sucht ein geeignetes Modell für die Einschätzung der Kreditfähigkeit von potenziellen Kreditnehmern. Dabei werden die Merkmale "Alter"(x_1), "Monatsgehalt"(x_2), "Laufzeit"(x_3) und "Höhe des Kredites"(x_4) erhoben. Mithilfe einer Stichprobe von $n = 300$ Beobachtungen wurden ein klassisches lineares Modell, ein Logit- und ein Probit-Modell für die Prognose der Kreditfähigkeit aufgestellt. Man erhielt folgende Ergebnisse:

	Lineares Modell		Logit-Modell		Probit-Modell	
	Estimate	p value	Estimate	p value	Estimate	p value
Intercept	$-1{,}105$	$1{,}81 \cdot 10^{-4}$	$-8{,}437$	$3{,}34 \cdot 10^{-7}$	$-4{,}627$	$9{,}67 \cdot 10^{-7}$
x_1	$1{,}328 \cdot 10^{-2}$	$2{,}15 \cdot 10^{-3}$	$6{,}936 \cdot 10^{-2}$	$2{,}40 \cdot 10^{-3}$	$3{,}820 \cdot 10^{-2}$	$4{,}27 \cdot 10^{-3}$
x_2	$2{,}031 \cdot 10^{-4}$	$1{,}19 \cdot 10^{-4}$	$1{,}068 \cdot 10^{-3}$	$1{,}46 \cdot 10^{-4}$	$5{,}683 \cdot 10^{-4}$	$5{,}24 \cdot 10^{-4}$
x_3	$2{,}790 \cdot 10^{-2}$	$3{,}49 \cdot 10^{-5}$	$1{,}456 \cdot 10^{-1}$	$7{,}09 \cdot 10^{-5}$	$8{,}455 \cdot 10^{-2}$	$6{,}68 \cdot 10^{-5}$
x_4	$2{,}303 \cdot 10^{-6}$	$2{,}23 \cdot 10^{-2}$	$1{,}159 \cdot 10^{-5}$	$2{,}46 \cdot 10^{-2}$	$6{,}859 \cdot 10^{-6}$	$2{,}25 \cdot 10^{-2}$

a) Prognostizieren Sie für einen Neukunden, der 52 Jahre alt ist, ein Monatsgehalt von 4413 € erhält, eine Kredithöhe von 103.361 möchte und eine Lauflänge von 17 Jahren anstrebt, die Kreditfähigkeit mithilfe der drei angegebenen Modelle. Interpretieren Sie die Ergebnisse und beurteilen Sie, welches der Modelle am wenigsten geeignet ist.

b) Welchen von den folgenden 10 möglichen Kreditnehmern sollte die Bank einen Kredit gewähren, wenn alle möglichen Kosten einer Fehleinschätzung gleich gewichtet werden? Geben Sie den Anteil an richtig klassifizierten Kreditnehmern an.

i	x_1	x_2	x_3	x_4	y_i
1	52	4413	17	103361	1
2	45	3649	17	61044	1
3	37	3548	14	85822	1
4	52	4040	20	84167	1
5	38	5134	11	116721	1
6	40	2609	13	65051	0
7	44	4571	13	67189	0
8	23	2333	6	57439	0
9	32	4041	7	74440	0
10	31	4013	10	62341	0

c) Wie würden sich die Anteile der richtig klassifizierten Kreditnehmer ändern, wenn die Kosten für einen nicht zurückgezahlten Kredit im Durchschnitt 6 mal so hoch sind, wie die Kosten eines nicht vergebenen Kredites, den der Kreditnehmer zurückgezahlt hätte? Wie verändert sich der Anteil der vergebenen Kredite?

Aufgabe 3

In einem verallgemeinerten linearen Modell ist die Nullhypothese

$$H_0 : \beta_j = 0 \qquad \text{gegen} \qquad H_1 : \beta_j \neq 0$$

mit Hilfe der Wald-Statistik zu testen. Für die ML-Schätzung von β_j und deren Varianz gilt

$$\widehat{\beta}_j = 2 \quad \mathrm{Var}(\widehat{\beta}_j) = 1{,}5.$$

Berechnen Sie den p-Wert und geben Sie an, ob die Nullhypothese zum Signifikanzniveau von $\alpha = 0{,}05$ zu verwerfen ist.

Übung 8: Verallgemeinerte lineare Modelle II

Aufgabe 1

Gegeben sei ein verallgemeinertes lineares Modell und zu testen sei die allgemeine lineare Hypothese

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d} \qquad \text{gegen} \qquad H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{d}$$

mit $\text{rang}(\mathbf{C}) = 2$. Für die Log-Likelihoodfunktion resultieren die Werte

$$\ln\left(\mathrm{L}\left(\widehat{\boldsymbol{\beta}}_{H_0;y_1,\ldots,y_n}\right)\right) = -1{,}3 \qquad \text{und} \qquad \ln\left(\mathrm{L}\left(\widehat{\boldsymbol{\beta}}_{H_1;y_1,\ldots,y_n}\right)\right) = 2{,}3.$$

Prüfen Sie, ob die Nullhypothese H_0 zum Signifikanzniveau $\alpha = 0{,}05$ bzw. $\alpha = 0{,}01$ zu verwerfen ist.

Aufgabe 2

Bisher wurde in einem Unternehmen ein verallgemeinertes lineares Modell M_1 mit $k = 3$ erklärenden Variablen eingesetzt. Das Unternehmen steht vor der Entscheidung, ob zukünftig ein erweitertes Modell M_2 verwendet werden sollte, bei dem zu den bisherigen drei erklärenden Variablen zwei weitere hinzugefügt wurden. D.h zu testen ist

$$H_0 : M_1 \qquad \text{gegen} \qquad H_1 : M_2$$

Bei Vorliegen von $n = 20$ Beobachtungen betragen die skalierten Devianzen

$$\Delta^*_{M_1} = 7{,}2 \qquad \text{und} \qquad \Delta^*_{M_2} = 1{,}1$$

a) Entscheiden Sie anhand dieser Informationen, ob die Nullhypothese $H_0 : M_1$ zum Signifikanzniveau $\alpha = 0{,}05$ abzulehnen ist.

b) Für die Log-Likelihood des Modells M_2 hat man $\ln\left(L\left(\widehat{\beta}_{M_2;y_1,\ldots,y_n}\right)\right) = 5{,}4$ erhalten. Berechnen Sie die Log-Likelihood von M_1 und vom saturierten Modell.

c) Wir würde die Entscheidung ausfallen, wenn M_1 kein Submodell von M_2 wäre?

d) Wie groß müsste der Stichprobenumfang mindestens sein, damit man sich im Aufgabenteil c) für das kleinere Modell entscheidet? Nehmen Sie dafür an, dass sich die Log-Likelihood Werte der beiden Modelle nicht verändern.

Übung 9: Grundlagen der Zeitreihenanalyse

Aufgabe 1

Betrachtet wird die Zeitreihe $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = U_1 \sin(2\pi\delta t) + U_2 \cos(2\pi\delta t) \qquad \text{für } t \in \mathbb{Z},$$

wobei U_1 und U_2 stochastisch unabhängig sind mit

$$\mathbb{E}[U_1] = \mathbb{E}[U_2] = 0 \qquad \text{und} \qquad \mathrm{Var}(U_1) = \mathrm{Var}(U_2) = \sigma^2.$$

Zeigen Sie, dass die Zeitreihe schwach stationär ist.
(Hinweis: Verwenden Sie $\sin(\alpha)\sin(\beta) + \cos(\alpha)\cos(\beta) = \cos(\alpha - \beta)$)

Aufgabe 2

Betrachtet wird der stochastische Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = \beta_1 + \beta_2 t + \varepsilon_t \qquad \text{für } t \in \mathbb{Z}$$

sowie $\beta_1, \beta_2 \in \mathbb{R}$ und $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

a) Untersuchen Sie, ob $(X_t)_{t\in\mathbb{Z}}$ schwach stationär ist.

b) Zeigen Sie, dass $(\nabla X_t)_{t\in\mathbb{Z}}$ mit $\nabla X_t = X_t - X_{t-1}$ schwach stationär ist.

Aufgabe 3

Betrachtet wird der stochastische Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = \varepsilon_t \varepsilon_{t-1} \qquad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \mathrm{IID}(0, \sigma^2)$. Bestimmen Sie die Mittelwert- und Autokovarianzfunktion von $(X_t)_{t\in\mathbb{Z}}$ und geben Sie an, ob $(X_t)_{t\in\mathbb{Z}}$ schwach stationär ist.

Aufgabe 4

Es seien $(X_t)_{t\in\mathbb{Z}}$ und $(Y_t)_{t\in\mathbb{Z}}$ zwei schwach stationäre und unkorrelierte Prozesse. Zeigen Sie, dass dann auch $(Z_t)_{t\in\mathbb{Z}}$ mit $Z_t := X_t + Y_t$ für $t \in \mathbb{Z}$ schwach stationär ist.

Übung 10: ARMA- und ARIMA-Prozesse I

Aufgabe 1

Im Folgenden gelte $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$. Benennen Sie die untenstehenden stochastischen Prozesse:

a) $X_t = \frac{4}{5}X_{t-1} - \frac{1}{2}X_{t-2} + \varepsilon_t$

b) $X_t = \varepsilon_t + \frac{1}{4}\varepsilon_{t-1} + \frac{1}{2}\varepsilon_{t-2}$

c) $X_t = \frac{3}{2}X_{t-1} - \frac{1}{2}X_{t-2} + \varepsilon_t + \frac{2}{5}\varepsilon_{t-1} + \frac{1}{5}\varepsilon_{t-2}$

d) $X_t = \frac{1}{2}X_{t-1} + \frac{1}{2}X_{t-2} + \varepsilon_t - \frac{3}{4}\varepsilon_{t-1} - \frac{1}{4}\varepsilon_{t-2}$

Untersuchen Sie die Prozesse auf Parameterredundanz, schwache Stationarität, Kausalität und Invertierbarkeit.

Aufgabe 2

Betrachtet wird der stochastische Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = \varepsilon_t + 0{,}7\varepsilon_{t-1} - 0{,}3\varepsilon_{t-2} + 0{,}2\varepsilon_{t-3},$$

wobei $\varepsilon \sim \mathrm{WN}(0, 50)$ gilt. Geben Sie die Autokovarianzfunktion $\gamma(h)$ des stochastischen Prozesses $(X_t)_{t\in\mathbb{Z}}$ an.

Übung 11: ARMA- und ARIMA-Prozesse II

Aufgabe 1

Bestimmen Sie für den AR(2)-Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = \frac{6}{13}X_{t-1} - \frac{10}{13}X_{t-2} + \varepsilon_t \qquad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$ das zugehörige Autoregressive Polynom $\Phi_2(z)$ sowie dessen Nullstellen in Polardarstellung. Stellen Sie zusätzlich die Autokorrelationsfunktion $\rho(h)$ als Funktion dieser Nullstellen dar.

Aufgabe 2

Ermitteln Sie mithilfe der Methode der unbestimmten Koeffizienten die kausale Darstellung des stochastischen Prozesses

$$X_t = 1{,}3X_{t-1} - 0{,}4X_{t-2} + \varepsilon_t - 0{,}2\varepsilon_{t-1}$$

mit $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$.
(Hinweis: Verwenden Sie, dass die Differenzengleichung 2. Ordnung

$$0{,}4\psi_{j-1} - 1{,}3\psi_j + \psi_{j+1} = 0$$

die Lösung

$$\psi_j = c_1\left(\frac{1}{z_1}\right)^j + c_2\left(\frac{1}{z_2}\right)^j$$

besitzt, wobei z_1 und z_2 die Nullstellen des Polynoms $\Phi_2(z) = 1 - 1{,}3z + 0{,}4z^2$ sind und $c_1 + c_2 = 1$ gilt.)

Lehrstuhl für BWL, insb. Mathematik und Statistik
Prof. Dr. Michael Merz
Regressionsmodelle mit Anwendungen in der Versicherungs- und Finanzwirtschaft
Wintersemester 2017/18

Übung 12: Schätzung und Prognose

Aufgabe 1

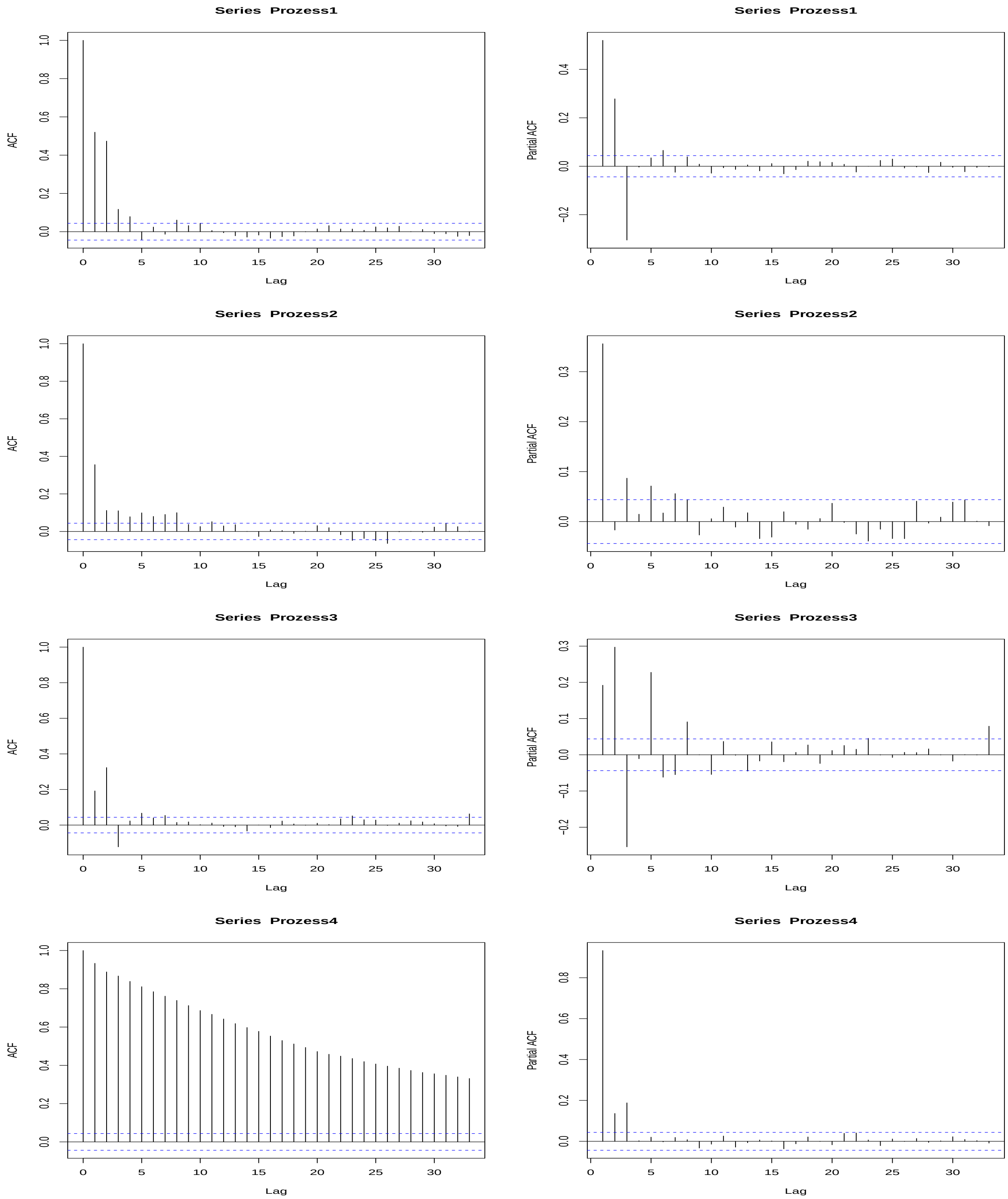
Benennen Sie die folgenden vier stochastischen Prozesse:

a) $X_t = \frac{3}{4}X_{t-1} + \frac{1}{5}X_{t-3} + \varepsilon_t$

b) $X_t = \varepsilon_t + \frac{1}{4}\varepsilon_{t-1} + \frac{1}{2}\varepsilon_{t-2} - \frac{1}{5}\varepsilon_{t-3}$

c) $X_t = \frac{9}{20}X_{t-1} + \frac{2}{5}X_{t-2} - \frac{3}{10}X_{t-3} + \varepsilon_t$

d) $X_t = \frac{4}{5}X_{t-1} + \frac{1}{20}X_{t-2} + \varepsilon_t - \frac{9}{20}\varepsilon_{t-1} - \frac{1}{4}\varepsilon_{t-2}$



Aufgabe 2

Für einen stochastischen Prozess $(X_t)_{t\in\mathbb{Z}}$ wurde die Zeitreihe

t	X_t
1	4
2	14
3	10
4	16
5	6
6	14
7	10
8	14

realisiert. Bestimmen Sie die die empirische Autokorrelationsfunktion für die gegebene Zeitreihe.

Aufgabe 3

Für den schwach stationären AR(2)-Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim$WN$(0,\sigma^2)$ wurden anhand von $T = 100$ Beobachtungen die Schätzungen $\widehat{\rho}(1) = 0{,}8$ und $\widehat{\rho}(2) = 0{,}5$ sowie

$$\overline{X}_T = \frac{1}{T}\sum_{t=1}^{100} X_t = 2$$

ermittelt. Bestimmen Sie hiermit die Momentenschätzungen für die Parameter ϕ_1, ϕ_2 und μ.

Aufgabe 4

Betrachtet wird der stochastische Prozess $(X_t)_{t\in\mathbb{Z}}$ mit

$$X_t = \varepsilon_t + 0{,}7\varepsilon_{t-1} - 0{,}3\varepsilon_{t-2} + 0{,}2\varepsilon_{t-3},$$

wobei $\varepsilon \sim$ WN$(0, 50)$ gilt. Bestimmen Sie den besten affin-linearen Prädiktor

$$\widehat{X}^{opt}_{T+h|T} = a_0 + \sum_{t=1}^{T} a_{T+1-t} X_t$$

für $T = 2$ und $h = 2$.