

Aufgabe 1 (Hierarchisches Clustern)

(14 + 6 = 20 Punkte)

Sechs Objekte werden mit dem average-linkage-Verfahren geclustert. Dabei ist die folgende Distanzmatrix gegeben.

	{1}	{2}	{3}	{4}	{5}	{6}
{1}	0	25	8	1	10	10
{2}	25	0	5	32	5	25
{3}	8	5	0	13	2	10
{4}	1	32	13	0	13	17
{5}	10	5	2	13	0	20
{6}	10	25	10	17	20	0

- Clustern Sie die Objekte mit dem Average-Linkage-Verfahren. Kommentieren Sie dabei die einzelnen Schritte z.B. durch Hervorhebungen in den einzelnen Matrizen.
- Zeichnen Sie das Dendrogramm und benennen Sie eine plausible Partition.

Aufgabe 2 (Parameterschätzung, Zweistichprobentests) (10 + 6 + 4 + 12 = 32 Punkte)

In der Weihnachtsmanufaktur am Nordpol werden LED-Lichterketten mit Zeitschalter hergestellt. Dieser schaltet die Kette nach Einbruch der Dunkelheit zunächst ein und nach einer zufälligen Zeit $X \in [0; 1]$ wieder aus (Angabe in Stunden). Der Statistikwichtel nimmt an, dass die Wartezeit X eine Dreieck-Verteilung mit der folgenden Dichte $f_a : [0; 1] \rightarrow \mathbb{R}$ mit $0 < a < 1$ besitzt:

$$f_a(x) = 2 \min\left(\frac{x}{a}, \frac{1-x}{1-a}\right) = \begin{cases} 2x/a & \text{falls } x < a \\ 2(1-x)/(1-a) & \text{falls } x \geq a \end{cases}$$

Im Testbetrieb werden 20 Schaltvorgänge beobachtet und die Zeiten X_1, \dots, X_{20} bis zum Ausschalten gemessen. Der Statistikwichtel sucht eine Schätzung für a auf Grundlage der beobachteten Werte x_1, \dots, x_{20} .

Hinweis: In den folgenden Aufgaben ist stets der vollständige Rechenweg anzugeben, die bloße Angabe von Endergebnissen reicht nicht aus.

- Berechnen Sie einen Momentenmethode-Schätzer \hat{a}_{MM} für a . Leiten Sie zunächst $E(X_1) = \frac{a+1}{3}$ her.
- Prüfen Sie, ob der von Ihnen in a) berechnete Schätzer \hat{a}_{MM} erwartungstreu ist und berechnen Sie den MSE, d.h. $E((\hat{a}_{MM} - a)^2)$. Dabei dürfen Sie ohne weitere Rechnung ausnutzen, dass $\text{var}(X_1) = \frac{1}{18}(a^2 - a + 1)$.
- Aus der Messung wird der Wert $\bar{x} = 0.32$ berechnet. Ermitteln Sie zunächst die zugehörige MM-Schätzung. Erläutern Sie, weshalb der MM-Schätzer für sich genommen hier problematisch ist. Welchen Nutzen könnte er dennoch haben?
- Die in der Stichprobe gefundene durchschnittliche Wartezeit von 0.32 Stunden wird vom Oberwichtel als zu kurz angesehen. Der Elektrowichtel entwirft daraufhin eine andere Schaltung. Es soll geprüft werden, ob diese Schaltung eine höhere erwartete Laufzeit bis zum Abschalten hat. Mit den beiden Versionen der Lichterkette (Kette 1 mit ursprünglicher Schaltung, Kette 2 mit neuer Schaltung) werden nun jeweils 12 Messungen durchgeführt und es ergeben sich folgende Ergebnisse:

Kette 1	0.276	0.277	0.287	0.288	0.304	0.330	0.338	0.341	0.347	0.361	0.386	0.413
Kette 2	0.315	0.391	0.397	0.398	0.399	0.404	0.434	0.435	0.445	0.448	0.454	0.455

- Erläutern Sie, welche Modellannahmen hier erfüllt sein müssen, damit der Wilcoxon-Test zur Anwendung kommen kann.
- Führen Sie unter diesen Annahmen den Wilcoxon-Test zum Signifikanzniveau $\alpha = 1/100$ aus (d.h. gehen Sie auf Hypothesen, Teststatistik, Test inklusive Schwellenwert sowie Testentscheidung ein). Beachten Sie auch die Quantiltabelle im Anschluss an die Aufgabe.

Quantile $w_{0,01}(n_1, n_2)$ der Wilcoxon-Verteilung

$n_2 \backslash n_1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5
3	6	6	6	6	6	7	7	8	8	8	9	9	9	10	10	11	11	11	12	12	12	13	13	14
4	10	10	10	11	12	12	13	14	14	15	16	16	17	18	18	19	20	20	21	22	22	23	24	24
5	15	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
6	21	21	23	24	25	26	28	29	30	31	33	34	35	37	38	40	41	42	44	45	46	48	49	51
7	28	29	30	32	33	35	36	38	40	41	43	45	46	48	50	52	53	55	57	59	60	62	64	65
8	36	37	39	41	43	44	46	48	50	52	54	57	59	61	63	65	67	69	71	73	75	77	79	82
9	45	47	49	51	53	55	57	60	62	64	67	69	72	74	77	79	82	84	86	89	91	94	96	99
10	55	57	59	62	64	67	69	72	75	78	80	83	86	89	92	94	97	100	103	106	109	111	114	117
11	66	68	71	74	76	79	82	85	89	92	95	98	101	104	108	111	114	117	120	124	127	130	133	137
12	78	81	84	87	90	93	96	100	103	107	110	114	117	121	125	128	132	135	139	143	146	150	154	157
13	92	94	97	101	104	108	112	115	119	123	127	131	135	139	143	147	151	155	159	163	167	171	175	179
14	106	108	112	116	119	123	128	132	136	140	144	149	153	157	162	166	171	175	179	184	188	193	197	201
15	121	124	128	132	136	140	145	149	154	158	163	168	172	177	182	187	191	196	201	206	211	215	220	225
16	137	140	144	149	153	158	163	168	173	178	183	188	193	198	203	208	213	219	224	229	234	239	245	250
17	154	158	162	167	172	177	182	187	192	198	203	209	214	220	225	231	236	242	247	253	259	264	270	276
18	172	176	181	186	191	196	202	208	213	219	225	231	237	242	248	254	260	266	272	278	284	290	296	302
19	192	195	200	206	211	217	223	229	235	241	247	254	260	266	273	279	285	292	298	304	311	317	324	330
20	212	216	221	227	233	239	245	251	258	264	271	278	284	291	298	304	311	318	325	332	338	345	352	359
21	233	237	243	249	255	262	268	275	282	289	296	303	310	317	324	331	338	345	353	360	367	374	382	389
22	255	259	265	272	278	285	292	299	307	314	321	329	336	344	351	359	366	374	381	389	397	404	412	420
23	278	283	289	296	303	310	317	325	332	340	348	356	364	371	379	387	395	403	411	419	427	435	444	452
24	302	307	314	321	328	336	343	351	359	367	376	384	392	400	409	417	425	434	442	451	459	468	476	485
25	327	333	339	347	355	362	371	379	387	396	404	413	421	430	439	448	456	465	474	483	492	501	510	518

Für die 0,99-Quantile gilt $w_{0,99}(n_1, n_2) = n_1(n_1 + n_2 + 1) - w_{0,01}(n_1, n_2)$

Aufgabe 3 (Regression)

(20 Punkte)

(20 Punkte)

Gegeben sei die folgenden 10 Datensätze von Objekten mit den drei metrischen Merkmalen x, y, z :

Objekt	1	2	3	4	5	6	7	8	9	10
x	2	2	3	3	4	5	6	7	7	10
y	1	2	-1	3	-2	-3	4	-4	5	-2
z	4	4	3	5	2	0	5	2	7	6

Der Datensatz wird in einen Data Frame `alldata` eingelesen und mit R wird eine lineare Regression durchgeführt. Als Ergebnis erhalten Sie die folgende Konsolenausgabe:

```

1 Call:
2   lm(formula = z ~ x + y, data = alldata)
3
4 Residuals:
5     Min       1Q   Median       3Q      Max
6 -1.99666 -0.21003  0.08498  0.58498  1.55698
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept)  ?.?????    0.8484   2.100  0.07384 .
11 x            ?.?????    ?.?????  ?.????  0.04380 *
12 y            ?.?????    ?.?????  ?.????  0.00331 **
13 ---
14 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
15
16 Residual standard error: ?.???? on 7 degrees of freedom
17 Multiple R-squared:  0.7554, Adjusted R-squared:  ?.?????
18 F-statistic: ???.?? on 2 and 7 DF,  p-value: 0.007235

```

Berechnen Sie die zehn Einträge in den Zeilen 10,11,12,16,17,18 der Konsolenausgabe, die mit `?.????` bzw. `?.??? bzw. ???.??` „maskiert“ sind. Geben Sie dabei auch den vollständigen Rechenweg an. Hinweise:

- Sie dürfen dabei auf die Datentabelle und die sichtbaren Einträge der Konsolenausgabe zurückgreifen.
- Sie dürfen ohne besonderen Nachweis verwenden, dass für die Modellmatrix X gilt:

$$(X^T X)^{-1} = \frac{1}{51492} \begin{pmatrix} 26789 & -4361 & -903 \\ -4361 & 881 & 147 \\ -903 & 147 & 609 \end{pmatrix}$$

- Die Angabe der gesuchten Einträge ohne einen wenigstens ansatzweisen Rechenweg ergibt keine Punkte.

Aufgabe 4 (Multiple Choice) (1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 12 Punkte)

Die nachfolgenden Aussagen sind entweder wahr (Antwort „ja“) oder falsch (Antwort „nein“). Geben Sie jeweils die richtige Antwort an:

Vorlesung Datenanalyse

a) Hat die Objektdistanzmatrix nur ganzzahlige Einträge, so gilt dies bei single linkage auch für jede Cluster-Distanzmatrix.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
b) Durch Umordnen der Objekte im Dendrogramm lässt sich bei Centroid- und Ward-Verfahren stets ein überschneidungsfreies Dendrogramm erzeugen	<input type="checkbox"/> ja	<input type="checkbox"/> nein
c) Erwartungstreue Schätzer haben stets minimale Schätzervarianz	<input type="checkbox"/> ja	<input type="checkbox"/> nein
d) Bei einem erwartungstreuen Schätzer ist der Bias Null.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
e) Der chi-Quadrat-Test hat eine besonders niedrige Fehlerwahrscheinlichkeit 2. Art.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
f) Für die Anwendung des Kolmogoroff-Smirnoff-Tests ist es erforderlich, dass die Merkmale stetig verteilt sind.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
g) Die Teststatistiken des Gesamtmodell-Tests in der Regression lassen sich auf das Bestimmtheitsmaß zurückführen.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
h) Die KQ-Schätzer der linearen Regression sind unter Normalverteilungsannahme auch ML-Schätzer	<input type="checkbox"/> ja	<input type="checkbox"/> nein

Vorlesung Simulation

i) Die Warteschlangentheorie fordert bei der Analyse von $M M 1$ Wartesystemen, dass gilt $\mu < \lambda$.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
j) Das Gesetz von Little gilt nicht für Wartesysteme mit seriellen Schaltern.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
k) Beim Wartesystem $M M 1$ mit ungeduldligen Kunden ist die Ankunftsrate unabhängig vom Zustand des Systems.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
l) Beim Wartesystem $M M s$ stehen s serielle Schalter zur Abfertigung bereit.	<input type="checkbox"/> ja	<input type="checkbox"/> nein

Aufgabe 5 (Simulationssoftware) (3 + 6 + 1 = 10 Punkte)

- Nennen Sie die drei grundlegenden Ansätze zur Umsetzung eines Computermodells einer Simulation.
- Nennen Sie zu jedem der drei grundlegenden Ansätze einen Vorteil, einen Nachteil sowie ein konkretes Beispielwerkzeug.
- Erläutern Sie kurz den hauptsächlichen Vorteil einer hierarchischen Simulationsumgebung im Vergleich zu einer nicht-hierarchischen Simulationsumgebung.

Aufgabe 6 (Zufallszahlen) (2 + 3 + 6 = 11 Punkte)

- Nehmen Sie an, die Zufallsvariable X sei Bernoulli-verteilt. Geben Sie ein Beispiel für die Verwendung einer solchen Zufallsvariable in einem Simulationsmodell. Erläutern Sie Ihr Beispiel kurz.
- Formulieren Sie einen Algorithmus zur Generierung einer Realisierung von X auf Basis einer Realisierung einer im Intervall $[0,1]$ gleichverteilten Zufallsvariable U .
- Schlagen Sie ein Verfahren zur rechnerischen Generierung von Realisierungen von U vor. Benennen Sie das Verfahren, formulieren Sie eine parametrisierte Rechenvorschrift für die Generierung mit dem Verfahren, benennen Sie die Parameter und erläutern Sie kurz, was bei der Wahl der Parameter in der Praxis zu beachten ist.

Aufgabe 7 (ARENA)

(2 + 13 = 15 Punkte)

- a) Erläutern Sie kurz den Unterschied zwischen den Attributen und Variablen in ARENA.
- b) Frau M. Zug plant eine neue Touristenattraktion für die Stadt zu schaffen. Die neue Attraktion soll ein Museum sein, in dem man eine Modelleisenbahnanlage, die die Stadt in Miniaturformat darstellt, bewundern kann. Während der Planung, bittet sie Sie darum, Ihr bei der Simulation der Besucherströme im neuen Touristenmagnet zu unterstützen. Sie hofft, durch die Simulation einen besseren Eindruck über Auslastung zu erhalten und so mögliche Ressourcen-Engpässe frühzeitig identifizieren zu können. Sie beschreibt die erwarteten Abläufe wie folgt:
- Etwa alle 7 Minuten trifft ein neuer Modelleisenbahnbesucher ein.
 - Der Besucher hat die Wahl entweder eine geführte Tour mit einem unserer Modelleisenbahnspezialisten zu machen, oder auf eigene Faust das Museum zu erkunden. Etwa 20 % der Besucher wählen eine geführte Tour.
 - Die Teilnehmerzahl für die Führungen ist limitiert auf 15 Personen. Wir versuchen diese Teilnehmerzahl zu erreichen. Damit die Besucher allerdings nicht zu lange warten müssen, ist die maximale Wartezeit auf 30 Minuten festgelegt und die Führung wird notfalls mit geringerer Teilnehmerzahl durchgeführt.
 - Die Führungen werden jeweils von einem unserer 6 Experten durchgeführt und dauern ca. 30 Minuten.
 - Die Besucher, die sich gegen eine geführte Tour entscheiden, können sich ein Headset mit Audioguide ausleihen. Die Besucher mit einem Headset bestaunen die Modelleisenbahn wahrscheinlich 45 Minuten, minimal 20 Minuten und maximal 90 Minuten.
 - Sie können davon ausgehen, dass alle Besucher, die keine geführte Tour mitmachen, ein Headset haben wollen. Leider ist die Anzahl der Headsets beschränkt. Wir haben momentan nur 30 Headsets zur Verfügung.
 - Falls alle Headsets entliehen sind, wartet der Besucher nicht darauf, dass wieder eines frei wird, sondern geht ohne Headset durch das Museum. Ein Museumsbesuch ohne Headset dauert zwischen 20 und 40 Minuten.
 - Nach der geführten Tour oder dem Besuch, mit oder ohne Headset, verlässt ungefähr die Hälfte der Besucher unmittelbar das Museum. Die andere Hälfte geht noch in unseren Andenkenshop. Dort bleiben sie im Schnitt 15 Minuten und verlassen anschließend das Museum. Den Bezahlvorgang im Shop müssen Sie nicht modellieren.

Skizzieren Sie ein entsprechendes ARENA-Modell zur Simulation der neuen Touristenattraktion. Benennen Sie alle Module aussagekräftig und notieren Sie zu treffende Einstellungen in den Flowchart- und Data-Modulen.

– Flowchart Module

- * Create: Zwischenankunftszeit | Anzahl Entitäten pro Ankunft
- * Dispose
- * Process: Seize-Delay-Release Typ | (Bearbeitungs-/ Delay-Zeit) | (Ressource(n))
- * Decide: by-condition + Bedingung/ by-chance + Wahrscheinlichkeiten
- * (Adjustable) Batch: Anzahl / (maximale Wartezeit)
- * Separate: Duplicate Original / Split Batch
- * Assign: Art | Name | Wert
- * Record: Count / Time Interval

– Data Module

- * Ressource: Ressource | Anzahl

Bitte beachten Sie, dass Sie **nicht** alle der hier aufgeführten Flowchart-Module in Ihrem Modell verwenden müssen und hier nicht aufgeführte Module nicht beschrieben werden müssen. Sie müssen desweiteren keine Angaben zum Run-Setup machen und können davon ausgehen, dass als Standard-Zeiteinheit Minuten eingestellt ist. Dokumentieren Sie (falls nötig) getroffene Annahmen.