

Vorname: .....

Nachname: .....

Matrikel-Nr.: .....

Studiengang: .....

Versuch Nr.: .....

## Klausur Angewandte Informatik

Prüfer	Etschberger, Jansen, Wesp
Prüfungsdatum	23. Januar 2019
Prüfungsort	Augsburg
Studiengang	BW, IM

Bearbeitungszeit:	90 Minuten
Punkte:	90

Die Klausur umfasst	7 Aufgaben auf 22 Seiten
---------------------	--------------------------

Zugelassene Hilfsmittel	Schreibzeug, Taschenrechner, der nicht 70! berechnen kann, ein mit dem Namen versehenes Din-A4 Blatt mit handgeschriebenen Notizen (keine Kopien oder Ausdrucke)
-------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Weitere Regularien:**

- ▶ Bitte überprüfen Sie *vor* Bearbeitungsbeginn die Vollständigkeit der Klausurangabe.
- ▶ Tragen Sie Ihren Namen und Ihre Matrikelnummer auf dem Deckblatt ein.
- ▶ Die Heftung der Klausur darf nicht verändert werden.
- ▶ Bitte tragen Sie die Lösung zu den jeweiligen Aufgaben *nur* direkt im Anschluss an die jeweilige Angabe ein. Sollte der Platz dort nicht ausreichen, verwenden Sie die Ersatzblätter am Ende der Klausurangabe.
- ▶ Ergebnisse (auch Zwischenergebnisse) müssen mit mind. 4 gültigen Ziffern angegeben werden.
- ▶ Der Lösungsweg muss klar dokumentiert werden.
- ▶ Die Klausur ist in ordentlich lesbarer Form zu bearbeiten. Schwer lesbare Teile der Klausur werden als ungültig ersatzlos gestrichen.
- ▶ Die Klausur unterliegt der für Sie zur Zeit gültigen Prüfungsordnung.
- ▶ Bitte verwenden Sie *keine rote Farbe* zur Bearbeitung der Klausur.

<b>Aufgabe</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>Punkte</b>	<input style="width: 40px; height: 40px;" type="text"/>	<input style="width: 40px; height: 40px;" type="text"/>	<input style="width: 40px; height: 40px;" type="text"/>	<input style="width: 40px; height: 40px;" type="text"/>	<input style="width: 40px; height: 40px;" type="text"/>	<input style="width: 40px; height: 40px;" type="text"/>	<input style="width: 40px; height: 40px;" type="text"/>
<i>maximal</i>	15	15	5	6	14	26	9

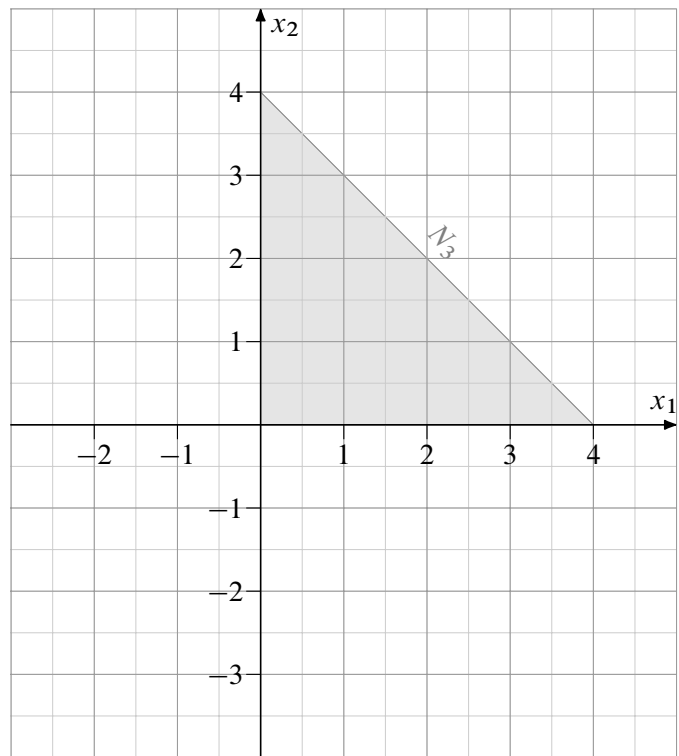
# Aufgabe 1

15 Punkte

Gegeben ist das folgende lineare Optimierungsproblem mit

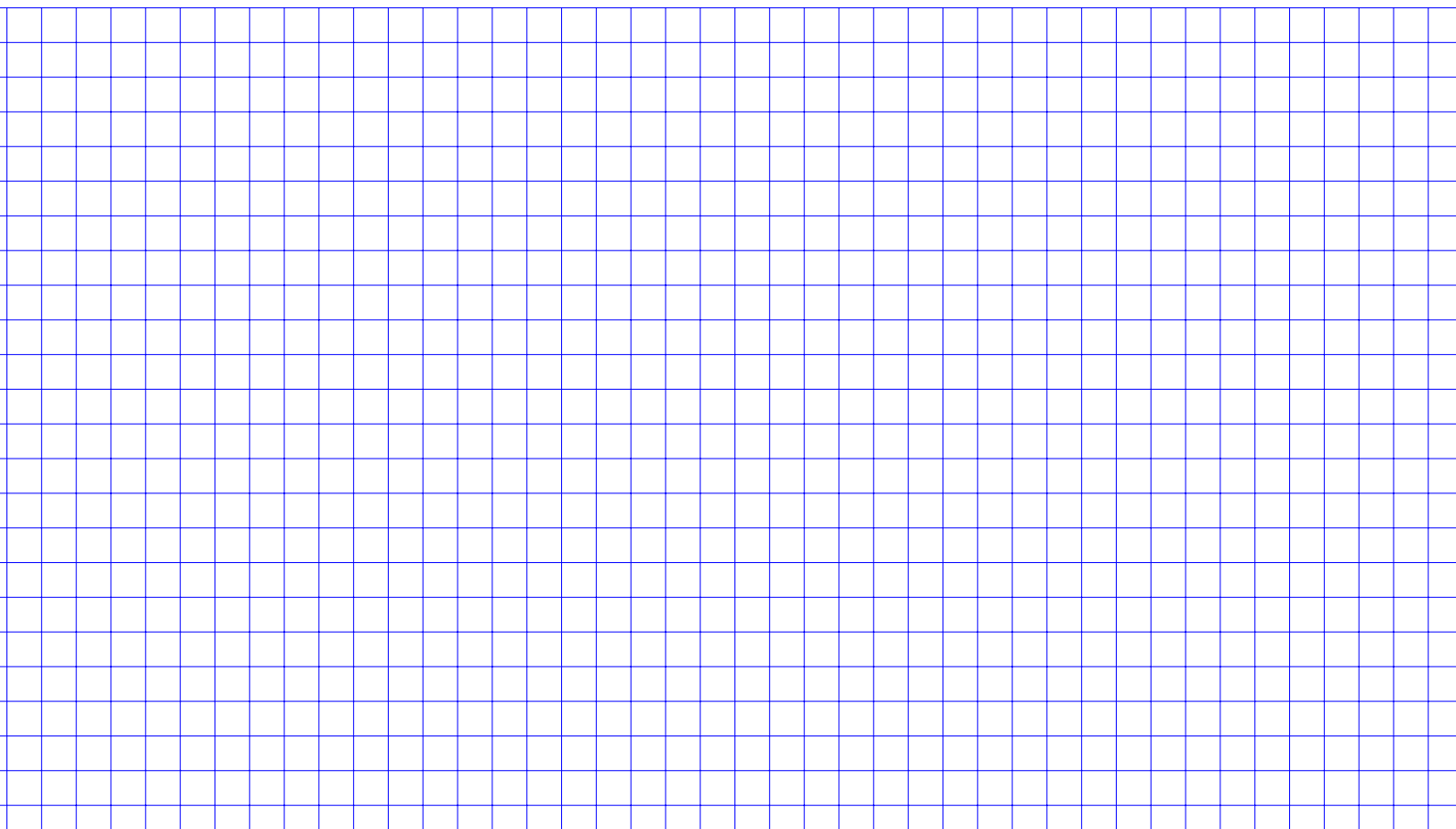
- ▶ den Strukturvariablen  $x_1, x_2 \in \mathbb{R}_+$ ,
- ▶ der Zielfunktion  $f: \mathbb{R}_+^2 \rightarrow \mathbb{R}$  und
- ▶ den Nebenbedingungen  $N_1, \dots, N_4$  (dabei ist  $N_3$  nur teilweise gegeben).

$f$	$x_1$	+	$3x_2$	$\rightarrow$	max
$N_1$	$-x_1$	+	$x_2$	$\leq$	2
$N_2$	$x_1$	-	$x_2$	$\leq$	2
$N_3$	<input type="text"/>	+	<input type="text"/>	$\leq$	<input type="text"/>
$N_4$	$x_1$	-	$x_2$	$\leq$	3



a) Die graphische Repräsentation von  $N_3$  ist im Koordinatensystem rechts eingezeichnet. Füllen Sie in der Tabelle oben die fehlenden Felder () von  $N_3$  aus.

- b) Zeichnen Sie  $N_1, N_2, N_4$  sowie den Zulässigkeitsbereich des Problems in das Koordinatensystem ein.
- c) Begründen Sie, ob und ggf. welche Nebenbedingungen zu vernachlässigen sind.
- d) Bestimmen Sie graphisch (Berechnung der Schnittpunkte ist nötig) die optimale Lösung des linearen Optimierungsproblems.
- e) Geben Sie den Funktionsterm  $g(x_1, x_2)$  einer alternativen Zielfunktion an, so dass alle Punkte des zulässigen Bereichs optimal sind, welche auf dem Rand von  $N_1$  liegen.



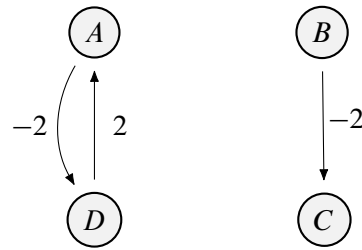


## Aufgabe 2

15 Punkte

Es sollen kürzeste Wege durch einen Graph nach dem Algorithmus von *Floyd-Warshall* gefunden werden.

- a) Gegeben ist dazu die unvollständige Zeichnung eines Graphen und die dazugehörige ebenfalls jeweils unvollständige Kosten- bzw. Vorgängermatrix  $C_0$  bzw.  $P_0$ .



$$C_0 = \begin{pmatrix} 0 & 6 & \infty & \square \\ \infty & 0 & \square & \infty \\ \infty & \square & 0 & -1 \\ 2 & 4 & 2 & \square \end{pmatrix},$$

$$P_0 = \begin{pmatrix} \square & A & 0 & A \\ 0 & B & \square & 0 \\ 0 & 0 & C & \square \\ D & \square & D & D \end{pmatrix}$$

Ergänzen Sie die fehlenden Kanten des Graphen mit deren Kantengewichten, sowie die fehlenden Einträge in der Kosten- und der Vorgängermatrix.

- b) Der vorletzte Iterationsschritt wurde bereits durchgeführt und Sie haben  $C_3$  vorliegen. Führen Sie nun den letzten Iterationsschritt durch, indem Sie die Matrix  $C_4$  an den fehlenden Stellen ergänzen.

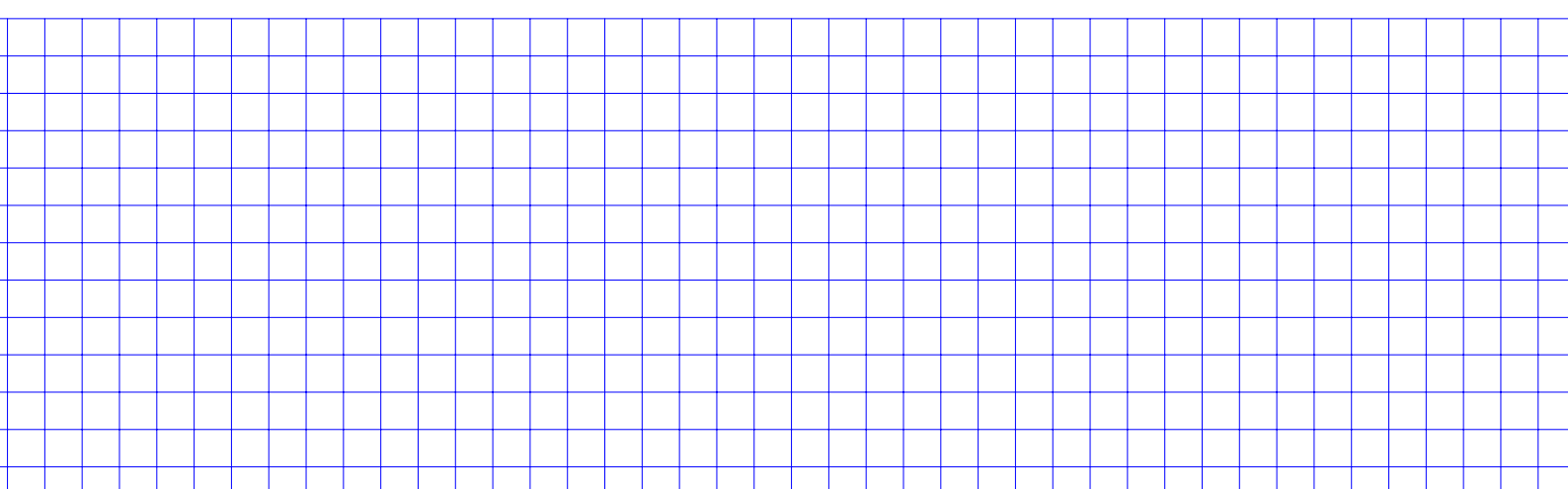
$$C_3 = \begin{pmatrix} 0 & 6 & 4 & -2 \\ \infty & 0 & -2 & -3 \\ \infty & \infty & 0 & -1 \\ 2 & 4 & 2 & 0 \end{pmatrix}$$

$$C_4 = \begin{pmatrix} \square & \square & \square & -2 \\ \square & \square & \square & -3 \\ \square & \square & \square & -1 \\ 2 & 4 & 2 & 0 \end{pmatrix}$$

- c) Gegeben ist jetzt zu einem *neuen* Graph mit 5 Knoten die Vorgängermatrix  $P_5$  nach der letzten Iteration des Algorithmus mit

$$P_5 = \begin{pmatrix} A & B & D & C & A \\ B & A & C & E & D \\ C & D & C & E & D \\ B & E & D & D & C \\ E & B & A & C & E \end{pmatrix}.$$

Dabei ist bekannt, dass die Zeilen zu Beginn alphabetisch aufsteigend mit den Bezeichnungen der Knoten von A bis E angeordnet waren. Geben Sie dazu den kürzesten Weg vom Knoten D nach A an.





### Aufgabe 3

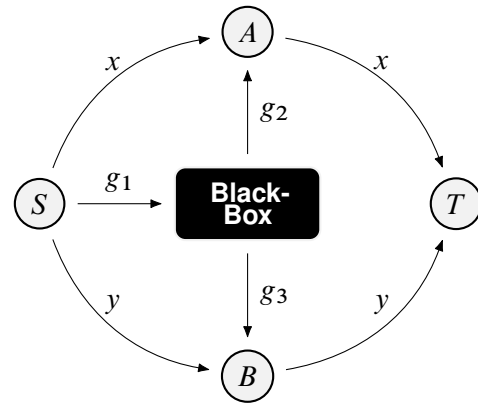
5 Punkte

Auf den nebenstehenden Graph wurde der Algorithmus von Dijkstra angewendet, wobei ein kürzester Weg von  $S$  nach  $T$  gesucht wird.

Alle Kanten des Graphen sind gerichtet. Die Gewichte  $g_1$ ,  $g_2$  und  $g_3$  sind unbekannt. Die Black-Box (BB) beschreibt einen Teil des Graphen, der nicht näher bekannt ist.

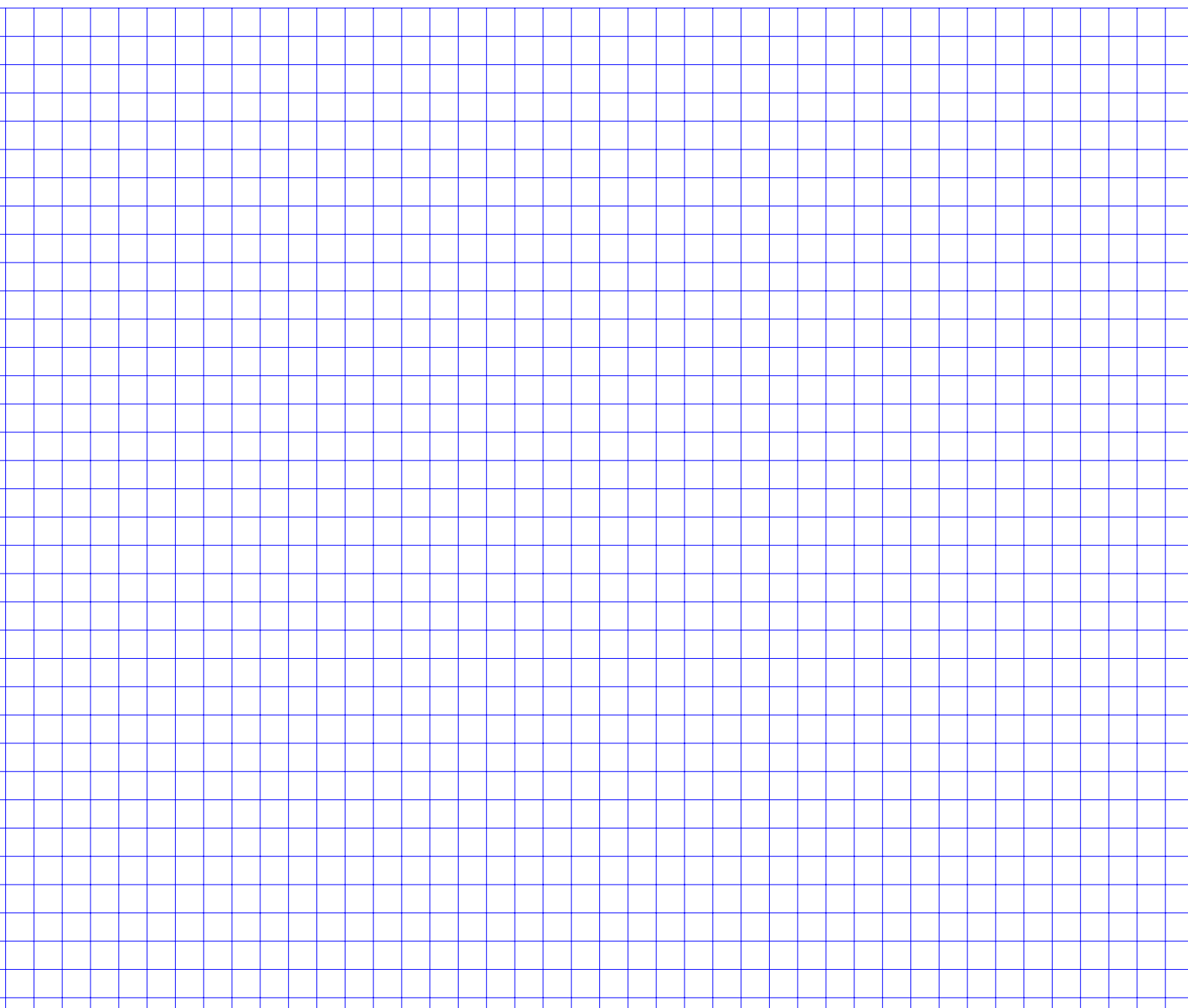
Außerdem ist bekannt, dass der kürzeste Weg von  $S$

- ▶ nach  $A$  bzw.  $B$  stets über BB führt,
- ▶ nach  $A$  ein Drittel von  $y$  entspricht,
- ▶ nach  $B$  die Hälfte von  $x$  beträgt,
- ▶ nach  $T$  sowohl über  $A$  wie über  $B$  gleich ist.



Nach dem letzten Iterationsschritt steht nur noch  $(A, a)$  und  $(B, b)$ , mit  $a, b \in \mathbb{N}$  in der Warteschlange.

Entscheiden Sie, ob Sie für  $y = 9$  sagen können, welcher Knoten als nächstes laut Algorithmus abgearbeitet wird, oder begründen Sie, warum dies nicht möglich ist.





## Aufgabe 4

6 Punkte

Eine Marketingberatung untersucht die Aufenthaltszeit von potentiellen Kunden auf Webshop-Seiten. Dazu wird die Verweildauer in Sekunden (Merkmal  $X$ ) auf den Webshop-Seiten eines Unternehmens bei 600 Besuchern der Webseite analysiert und aufgezeichnet. Außerdem ist noch das Geschlecht dieser Leute bekannt (Merkmal  $Y$ ). Die Beobachtungen  $(x_i, y_i)$  können als Ergebnis einer einfachen Stichprobe angesehen werden. Getrennt nach dem Geschlecht der Surfer ergibt sich folgende Häufigkeitstabelle für die Verweildauern (in Sekunden) und die Anzahl der Leute:

	Verweildauer auf Seite des Web-Shops									
	40	60	80	100	120	140	160	180	200	220
Frau	1	4	31	42	49	46	16	6	4	1
Mann	2	17	46	83	119	88	41	3	1	0

Getrennt nach Geschlecht bleiben die Untersuchten  $\bar{x}_{\text{Frau}} = 119.7$  bzw.  $\bar{x}_{\text{Mann}} = 117.45$  Sekunden auf der Webshop-Seite (muss nicht nachgerechnet werden).

Es soll mit einem  $\chi^2$ -Test zu einem Signifikanzniveau von 5 % getestet werden, ob das Geschlecht in der Grundgesamtheit abhängig von der Verweildauer ist. Dazu werden alle Beobachtungen von  $X$  in 5 Intervalle eingruppiert. Es ergibt sich folgende Kontingenztabelle:

Geschlecht	[40,76]	(76,112]	(112,148]	(148,184]	(184,220]
Frau	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Mann	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

- Füllen Sie die Kontingenztabelle aus.
- Berechnen Sie den Summand der Testgröße  $v$  für die Frauen und die Dauerklasse [40,76].
- Für die Testgröße ergibt sich  $v \approx 9.63$  (muss nicht nachgerechnet werden). Führen Sie den Test durch und leiten Sie daraus ab, ob die beiden Merkmale in der Grundgesamtheit unabhängig sind.
- Warum wäre eine Abhängigkeit der beiden Parameter kein Widerspruch zum fast identischen Wert von  $\bar{x}_{\text{Frau}}$  und  $\bar{x}_{\text{Mann}}$ ?





## Aufgabe 5

14 Punkte

Gegeben ist die  $(1\,000\,000 \times 2)$ -Datenmatrix  $A$  mit zwei metrischen Merkmalen  $X_1, X_2$  (Graphische Darstellung siehe Streuplot rechts).

Die Varianz-Kovarianzmatrix  $S$  von  $A$  ergibt sich mit

$$S = \begin{pmatrix} 2.0824 & -1.4464 \\ -1.4464 & 2.9292 \end{pmatrix}.$$

Es soll eine Hauptkomponentenanalyse zu  $A$  durchgeführt werden.

- a) Berechnen Sie näherungsweise die Eigenwerte  $\lambda_1, \lambda_2$  zu  $S$ .

(Hinweis: Falls Sie Teilaufgabe a) nicht lösen konnten, rechnen Sie bitte mit der (alternativen) Varianz-Kovarianz-matrix

$$S = \begin{pmatrix} 1.4704 & -1.6128 \\ -1.6128 & 6.5296 \end{pmatrix}$$

und den zugehörigen Eigenwerten  $\lambda_1 = 7, \lambda_2 = 1$  weiter.)

- b) Wieviel Prozent der Information können in der ersten Hauptkomponente konzentriert werden?  
c) Berechnen Sie die Eigenvektoren von  $S$ .  
d) Zeichnen Sie mit den Eigenvektoren die Hauptachsen der Hauptkomponentenanalyse in die Grafik ein. Benutzen Sie dabei den Schwerpunkt der Daten  $(x_1, x_2) = (5, 8)$  als Ursprung.

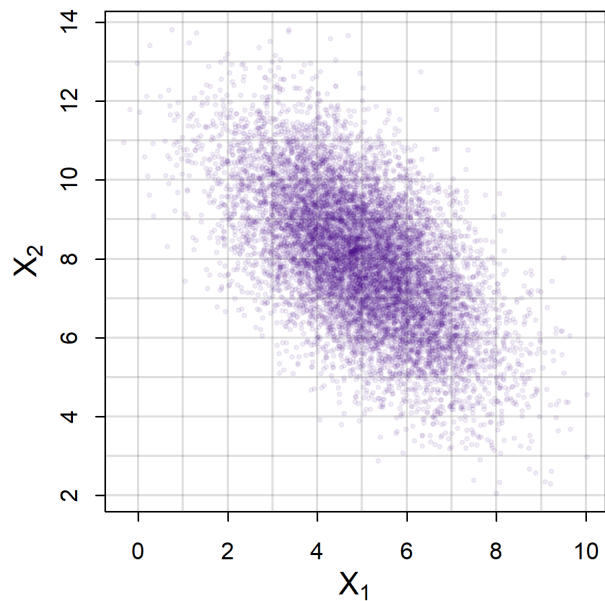
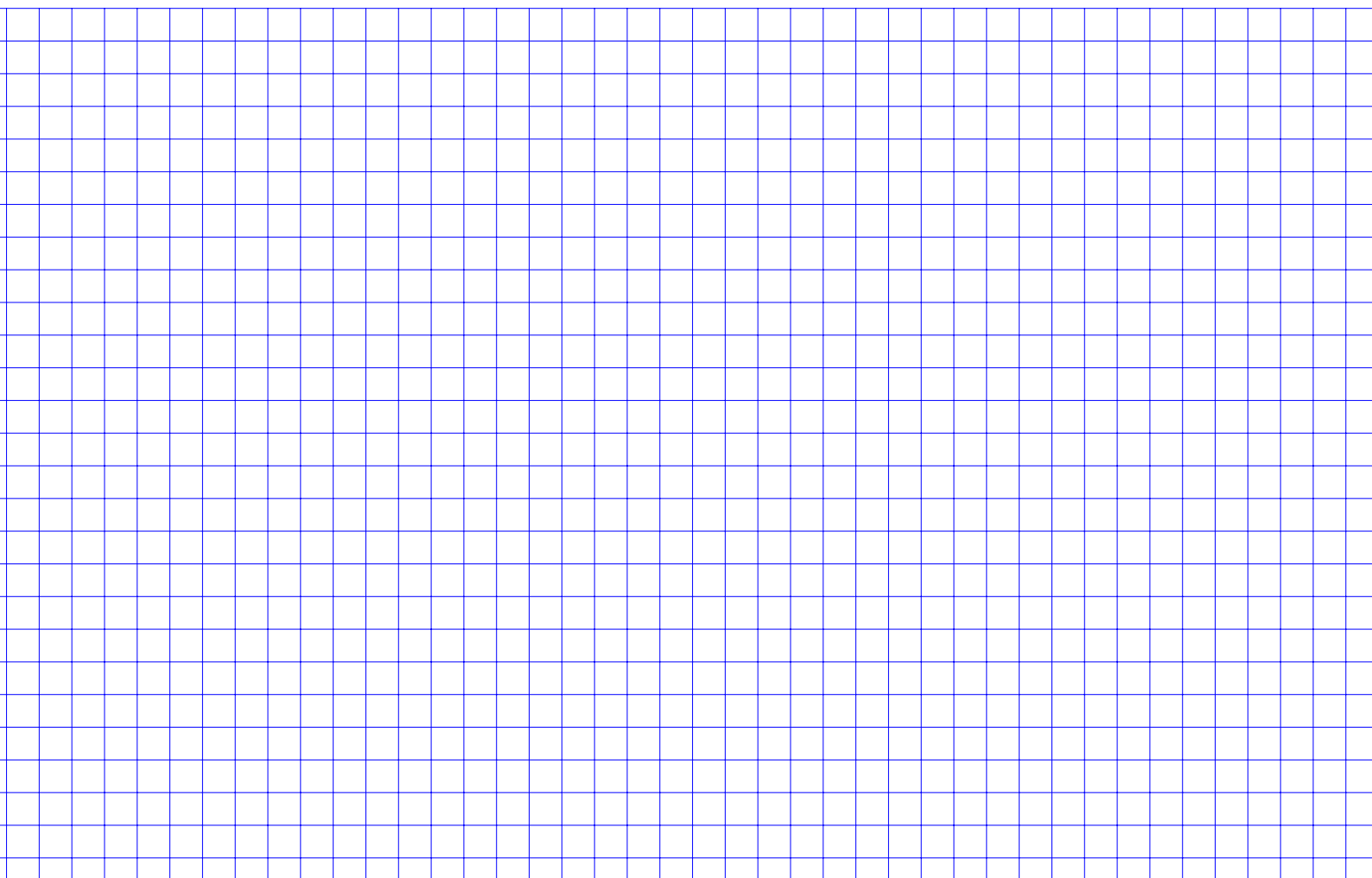


Abbildung 1: Streuplot von  $A$





## Aufgabe 6

26 Punkte

R

a) Gegeben ist ein Dataframe (`daten`) mit drei Spalten (`nachname`, `vorname` und `geburtsjahr`). Geben sie jeweils die R-Befehle an mit denen der dataframe wie jeweils beschrieben geändert wird:

1. Sie wollen eine weitere Zeile mit Ihren persönlichen Daten hinzufügen.

2. Sie wollen eine Spalte mit einem logischen Vektor hinzufügen, der genau dann TRUE ist falls die Person nach 1989 geboren ist. Die neue Spalte soll den Namen `u30` haben.

3. Die 42. Zeile sei nun die mit Ihrem persönlichen Eintrag. Geben Sie zwei Möglichkeiten an, wie Sie Ihren Nachnamen ändern können.

4. Wie könnte man die Daten im Dataframe (erste Spalte, alphabetisch aufsteigend) sortieren?

5. Nachdem Sie alle obigen Änderungen gemacht haben, stellen Sie fest, dass Ihnen die Spaltenüberschriften nicht gefallen. Wie Ändern Sie diese in `NAM`, `VOR`, `GEB` und `U30` um?

b) Sie erzeugen einen numerischen Vektor mittels `x <- seq(1, 2, by = 0.1)`. Welche der folgenden Befehle erzeugt den gleichen Vektor?

	wahr	falsch
<code>c(1, 0.1, times = 10)</code>	<input type="checkbox"/>	<input type="checkbox"/>
<code>seq(1, 2, length = 10)</code>	<input type="checkbox"/>	<input type="checkbox"/>
<code>seq(1, 2, length.out = 11)</code>	<input type="checkbox"/>	<input type="checkbox"/>
<code>seq(to=2, from=1, 0.1)</code>	<input type="checkbox"/>	<input type="checkbox"/>
<code>rep(1, 0.1, times = 11)</code>	<input type="checkbox"/>	<input type="checkbox"/>
<code>10:20/10</code>	<input type="checkbox"/>	<input type="checkbox"/>

c) Seien a und b zwei gleich lange numerische Vektoren der Länge n. Sie führen den Befehl `a!=b` aus. Was passiert?

	wahr	falsch
das Resultat ist ein numerischer Vektor	<input type="checkbox"/>	<input type="checkbox"/>
das Resultat ist ein logischer Vektor	<input type="checkbox"/>	<input type="checkbox"/>
das Resultat ist ein Vektor der Länge 1	<input type="checkbox"/>	<input type="checkbox"/>
das Resultat ist ein Vektor der Länge n	<input type="checkbox"/>	<input type="checkbox"/>
es wird ein Fehler ausgegeben	<input type="checkbox"/>	<input type="checkbox"/>
es wird eine Warnung ausgegeben	<input type="checkbox"/>	<input type="checkbox"/>
dem Vektor b wird die Fakultät von a zugewiesen	<input type="checkbox"/>	<input type="checkbox"/>
das Ergebnis ist TRUE falls jeweils die Fakultät von a gleich b ist, sonst FALSE	<input type="checkbox"/>	<input type="checkbox"/>

d) Welche Aussagen bezüglich Faktoren in R sind korrekt?

	wahr	falsch
Faktoren benötigt man für ordinale Merkmale	<input type="checkbox"/>	<input type="checkbox"/>
Faktoren benötigt man für kardinale Merkmale	<input type="checkbox"/>	<input type="checkbox"/>
Faktoren benötigt man für nominale Merkmale	<input type="checkbox"/>	<input type="checkbox"/>
in Data Frames sind Zeichenketten standardmäßig Faktoren	<input type="checkbox"/>	<input type="checkbox"/>
<code>is.factor()</code> wandelt numerische Werte in Faktoren um	<input type="checkbox"/>	<input type="checkbox"/>
es gibt keine Faktoren in R	<input type="checkbox"/>	<input type="checkbox"/>

e) Geben Sie an, was die Ausführung der folgenden Zeilen R-Code ausgibt (drei Antworten!).

```
x <- c(3,1,4,1,5,9,pi)
y <- 2 < x

y
all(y)
any(y)
```

- f) Aus der Statistikvorlesung kennen Sie das *Geburtstagsparadoxon*, in dem es darum geht, wie groß die Wahrscheinlichkeit ist, dass unter  $n$  Leuten mindestens 2 Leute am gleichen Tag Geburtstag haben. In R gibt es zum Berechnen dafür die Funktion `pbirthday()` (Hilfe: siehe Abb. 2). Wie berechnen Sie mit diesem Vorwissen und der Hilfe von R die Wahrscheinlichkeit dafür, dass in einer Gruppe von 250 Leuten genau 3 Leute die gleiche vierstellige EC-Karten PIN haben?

birthday (stats)

R Documentation

## Probability of coincidences

### Description

Computes answers to a generalised *birthday paradox* problem. `pbirthday` computes the probability of a coincidence and `qbirthday` computes the smallest number of observations needed to have at least a specified probability of coincidence.

### Usage

```
qbirthday(prob = 0.5, classes = 365, coincident = 2)
pbirthday(n, classes = 365, coincident = 2)
```

### Arguments

<code>classes</code>	How many distinct categories the people could fall into
<code>prob</code>	The desired probability of coincidence
<code>n</code>	The number of people
<code>coincident</code>	The number of people to fall in the same category

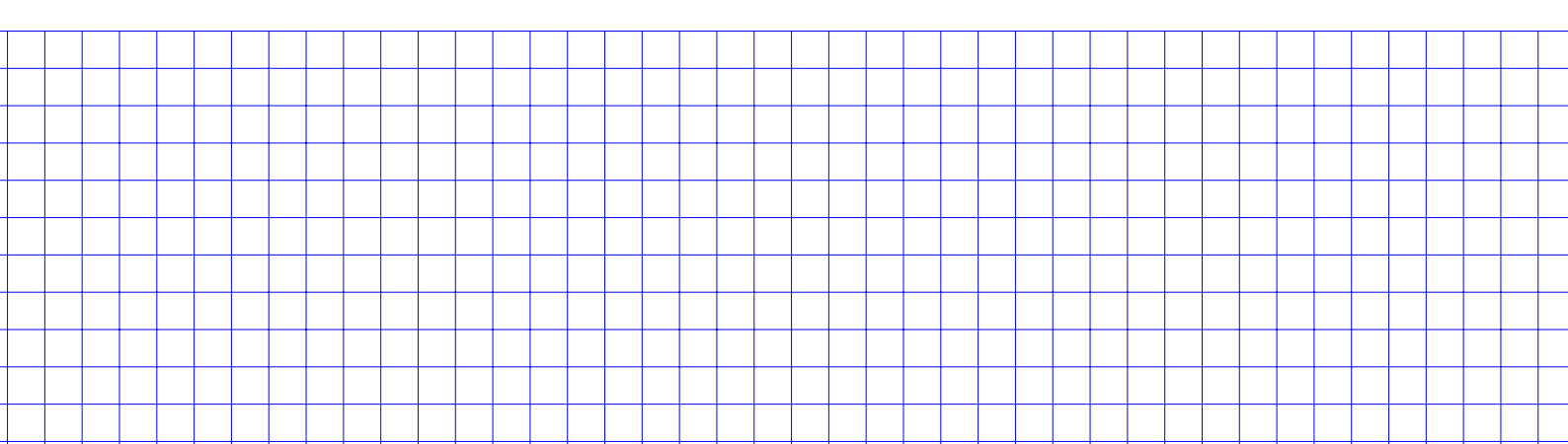
Abbildung 2: Hilfe zur R-Funktion `pbirthday()`



- g) Was versteht man unter der *Monte-Carlo-Methode*? Geben Sie *kein* Beispiel aus der Vorlesung an, sondern erläutern Sie die zugrunde liegende Idee.



- h) Wozu dient das R-Paket `knitr`?



















# Tabellen

## $\alpha$ -Fraktile der $\chi^2$ -Verteilung mit $n$ Freiheitsgraden

$\downarrow \alpha \setminus n \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.005	0.00	0.01	0.07	0.21	0.41	0.68	0.99	1.34	1.73	2.16	2.60	3.07	3.56	4.07	4.60
0.01	0.00	0.02	0.11	0.30	0.55	0.87	1.24	1.65	2.09	2.56	3.05	3.57	4.11	4.66	5.23
0.025	0.00	0.05	0.22	0.48	0.83	1.24	1.69	2.18	2.70	3.25	3.82	4.40	5.01	5.63	6.26
0.05	0.00	0.10	0.35	0.71	1.15	1.64	2.17	2.73	3.33	3.94	4.57	5.23	5.89	6.57	7.26
0.1	0.02	0.21	0.58	1.06	1.61	2.20	2.83	3.49	4.17	4.87	5.58	6.30	7.04	7.79	8.55
0.2	0.06	0.45	1.01	1.65	2.34	3.07	3.82	4.59	5.38	6.18	6.99	7.81	8.63	9.47	10.31
0.25	0.10	0.58	1.21	1.92	2.67	3.45	4.25	5.07	5.90	6.74	7.58	8.44	9.30	10.17	11.04
0.4	0.28	1.02	1.87	2.75	3.66	4.57	5.49	6.42	7.36	8.30	9.24	10.18	11.13	12.08	13.03
0.5	0.45	1.39	2.37	3.36	4.35	5.35	6.35	7.34	8.34	9.34	10.34	11.34	12.34	13.34	14.34
0.6	0.71	1.83	2.95	4.04	5.13	6.21	7.28	8.35	9.41	10.47	11.53	12.58	13.64	14.69	15.73
0.75	1.32	2.77	4.11	5.39	6.63	7.84	9.04	10.22	11.39	12.55	13.70	14.85	15.98	17.12	18.25
0.8	1.64	3.22	4.64	5.99	7.29	8.56	9.80	11.03	12.24	13.44	14.63	15.81	16.98	18.15	19.31
0.9	2.71	4.61	6.25	7.78	9.24	10.64	12.02	13.36	14.68	15.99	17.27	18.55	19.81	21.06	22.31
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68	21.03	22.36	23.68	25.00
0.975	5.02	7.38	9.35	11.14	12.83	14.45	16.01	17.53	19.02	20.48	21.92	23.34	24.74	26.12	27.49
0.99	6.63	9.21	11.34	13.28	15.09	16.81	18.48	20.09	21.67	23.21	24.73	26.22	27.69	29.14	30.58
0.995	7.88	10.60	12.84	14.86	16.75	18.55	20.28	21.95	23.59	25.19	26.76	28.30	29.82	31.32	32.80

$\downarrow \alpha \setminus n \rightarrow$	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
0.005	5.14	5.70	6.26	6.84	7.43	8.03	8.64	9.26	9.89	10.52	11.16	11.81	12.46	13.12	13.79
0.01	5.81	6.41	7.01	7.63	8.26	8.90	9.54	10.20	10.86	11.52	12.20	12.88	13.56	14.26	14.95
0.025	6.91	7.56	8.23	8.91	9.59	10.28	10.98	11.69	12.40	13.12	13.84	14.57	15.31	16.05	16.79
0.05	7.96	8.67	9.39	10.12	10.85	11.59	12.34	13.09	13.85	14.61	15.38	16.15	16.93	17.71	18.49
0.1	9.31	10.09	10.86	11.65	12.44	13.24	14.04	14.85	15.66	16.47	17.29	18.11	18.94	19.77	20.60
0.2	11.15	12.00	12.86	13.72	14.58	15.44	16.31	17.19	18.06	18.94	19.82	20.70	21.59	22.48	23.36
0.25	11.91	12.79	13.68	14.56	15.45	16.34	17.24	18.14	19.04	19.94	20.84	21.75	22.66	23.57	24.48
0.4	13.98	14.94	15.89	16.85	17.81	18.77	19.73	20.69	21.65	22.62	23.58	24.54	25.51	26.48	27.44
0.5	15.34	16.34	17.34	18.34	19.34	20.34	21.34	22.34	23.34	24.34	25.34	26.34	27.34	28.34	29.34
0.6	16.78	17.82	18.87	19.91	20.95	21.99	23.03	24.07	25.11	26.14	27.18	28.21	29.25	30.28	31.32
0.75	19.37	20.49	21.60	22.72	23.83	24.93	26.04	27.14	28.24	29.34	30.43	31.53	32.62	33.71	34.80
0.8	20.47	21.61	22.76	23.90	25.04	26.17	27.30	28.43	29.55	30.68	31.79	32.91	34.03	35.14	36.25
0.9	23.54	24.77	25.99	27.20	28.41	29.62	30.81	32.01	33.20	34.38	35.56	36.74	37.92	39.09	40.26
0.95	26.30	27.59	28.87	30.14	31.41	32.67	33.92	35.17	36.41	37.65	38.89	40.11	41.34	42.56	43.77
0.975	28.85	30.19	31.53	32.85	34.17	35.48	36.78	38.08	39.36	40.65	41.92	43.19	44.46	45.72	46.98
0.99	32.00	33.41	34.81	36.19	37.57	38.93	40.29	41.64	42.98	44.31	45.64	46.96	48.28	49.59	50.89
0.995	34.27	35.72	37.16	38.58	40.00	41.40	42.80	44.18	45.56	46.93	48.29	49.64	50.99	52.34	53.67