

**WIRTSCHAFTSWISSENSCHAFTLICHES
PRÜFUNGSSEKRETARIAT**

FAKULTÄT FÜR EMPIRISCHE HUMANWISSENSCHAFTEN UND WIRTSCHAFTSWISSENSCHAFT
DER UNIVERSITÄT DES SAARLANDES

Von der/dem Studierenden auszufüllen (Bitte leserlich und in Blockschrift):

Name der Prüfung: Deskriptive Statistik und Wahrscheinlichkeitsrechnung
Semester, dem die Prüfung zugeordnet ist: SS 2021 (z. B. WS 2019/2020, SS 2020)
(Prüfungen im Februar/April = WS; Prüfungen im August/Oktobre = SS)
Nachname, Vorname der/des Studierenden: _____
Matrikelnummer der/des Studierenden: _____

Erklärung

Hiermit erkläre ich, dass mir die für diese Prüfung relevanten Zulassungsvoraussetzungen aus der für mich geltenden Prüfungsordnung bekannt sind.

Mir ist damit bewusst, dass diese Prüfungsleistung als nicht abgelegt zählt, wenn die hierfür vorgesehenen Zulassungsvoraussetzungen nicht erfüllt sind.

Mir ist bekannt, dass die Teilnahme an der Prüfung zudem die ordnungsgemäße Anmeldung zur Prüfung voraussetzt. Die Teilnahme bei versäumter Anmeldung hat die Ungültigkeit der Prüfung zur Folge.

Zudem ist mir bekannt, dass eine nicht bestandene Prüfung zweimal wiederholt werden kann. Die Wiederholung einer bestandenen Prüfung ist nicht zulässig.

Datum: _____ **Unterschrift der/des Studierenden:** _____

Von der Prüferin/Vom Prüfer auszufüllen:

Aufgabe	Punkte	Max. Punkte	Bemerkungen
1		16	
2		12	
3		11	
4		19	
5		8	
6		5	
7		23	
8		17	
9		9	
<i>Summe</i>		120	

bestanden Note: _____

nicht bestanden Unterschrift der Prüferin/des Prüfers: _____

KLAUSURHEFT ZUR
 BACHELOR-PRÜFUNG
 DESKRIPTIVE STATISTIK UND WAHRSCHEINLICHKEITSRECHNUNG
 SOMMERSEMESTER 2021

PD Dr. Martin Becker

Hinweise für die Klausurteilnehmer

- Die Klausur besteht aus insgesamt 9 Aufgaben. Prüfen Sie die Vollständigkeit Ihres Exemplares nach; spätere Reklamationen können nicht berücksichtigt werden.
- Es sind insgesamt 120 Punkte (= 16 + 12 + 11 + 19 + 8 + 5 + 23 + 17 + 9) erreichbar.
- Als Hilfsmittel sind zugelassen: Taschenrechner (auch mit Grafikfähigkeit), 2 selbstgestellte DIN-A4 Blätter bzw. 4 selbstgestellte (einseitige) DIN-A4 Seiten. Benötigte Tabellen finden Sie am Ende dieses Klausurheftes.
- Bei der Korrektur werden **nur** die Lösungen auf diesen Blättern berücksichtigt.
- Bei mehreren Lösungsvorschlägen muss die gültige Lösung eindeutig gekennzeichnet sein.
- Mit Ausnahme der Multiple-Choice-Aufgaben muss der Lösungsweg klar ersichtlich sein.
- Alle Punkte innerhalb von Zahlen sind stets Dezimalpunkte (und keine Tausenderpunkte).

Bewertungsteil — Bitte nicht beschreiben							
Aufgabe	(a)	(b)	(c)	(d)	(e)	(f)	Σ
1		■	■	■	■	■	
2		■	■	■	■	■	
3					■	■	
4						■	
5				■	■	■	
6				■	■	■	
7							
8					■	■	
9				■	■	■	
Σ							

Aufgabe 1 (16 Punkte)

Markieren Sie jeweils mit einem Kreuz pro Aussage im betreffenden Kästchen, ob die unten stehenden Aussagen wahr oder falsch sind.

Richtige Antworten geben 2 Punkte, falsche Antworten und nicht bearbeitete Aussagen 0 Punkte (Aussagen mit zwei Kreuzen zählen als nicht bearbeitet!).

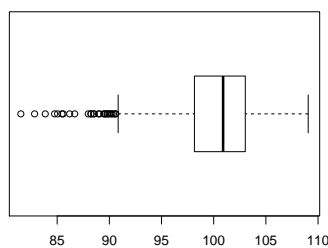
- | | wahr | falsch |
|--|--------------------------|--------------------------|
| 1. Jeder Median eines Merkmals X kommt mindestens einmal in der Urliste zu X vor. | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Ist (a_i, b_j) eine Ausprägung des zweidimensionalen Merkmals (X, Y) , dann ist die gemeinsame relative Häufigkeit $r(a_i, b_j)$ dieser Ausprägung niemals größer als die zugehörige relative Randhäufigkeit $r(a_i)$ der Ausprägung a_i des Merkmals X . | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. Es seien A und B zwei Ereignisse in einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Dann gilt stets $P(A) \geq P(A \cap B)$. | <input type="checkbox"/> | <input type="checkbox"/> |
| 4. Ist (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum und sind $A, B \in \mathcal{F}$ mit $P(A) > 0$ und $P(B) > 0$, so gilt stets:
$P(A B) + P(B A) = 1$ | <input type="checkbox"/> | <input type="checkbox"/> |
| 5. Die Anzahl der Fehlversuche, bevor man zum ersten Mal mit einem 6-seitigen fairen Würfel eine 5 oder 6 würfelt, ist $\text{Geom}(\frac{1}{3})$ -verteilt. | <input type="checkbox"/> | <input type="checkbox"/> |
| 6. Die Summe von drei unabhängigen $\text{Pois}(4)$ -verteilten Zufallsvariablen ist $\text{Pois}(12)$ -verteilt. | <input type="checkbox"/> | <input type="checkbox"/> |
| 7. Der Korrelationskoeffizient $\text{Korr}(X, Y)$ zweier Zufallsvariablen X und Y ist betragsmäßig stets höchstens so groß wie die zugehörige Kovarianz $\text{Cov}(X, Y)$ der beiden Zufallsvariablen. | <input type="checkbox"/> | <input type="checkbox"/> |
| 8. Die Zufallsvariablen X_1 und X_2 seien unkorreliert und für die Standardabweichungen gelte $\text{Sd}(X_1) = 3$ sowie $\text{Sd}(X_2) = 4$. Dann gilt für die Standardabweichung der Summe $X_1 + X_2$ $\text{Sd}(X_1 + X_2) = 7$. | <input type="checkbox"/> | <input type="checkbox"/> |

Aufgabe 2 (12 Punkte)

Markieren Sie jeweils die korrekte Antwort mit einem Kreuz im betreffenden Kästchen. Es ist jeweils genau ein Kreuz korrekt.

Richtige Antworten geben 3 Punkte, falsche Antworten und nicht bearbeitete Aufgabenteile 0 Punkte (Aufgabenteile mit mehr als einem Kreuz zählen als nicht bearbeitet!).

1. Kreuzen Sie an, auf welche Merkmalseigenschaften der folgende Box-Plot *am ehesten* hindeutet:



- (a) leptokurtisch und rechtssteil
(b) leptokurtisch und linkssteil
(c) platykurtisch und rechtssteil
(d) platykurtisch und linkssteil

2. Die Ränge $\text{rg}(X)_1, \dots, \text{rg}(X)_8$ zur (der Einfachheit halber sortierten) Urliste

4, 4, 4, 8, 9, 9, 10, 12

des ordinalskalierten Merkmals X lauten:

- (a) 2, 2, 2, 4, 5.5, 5.5, 7, 8
(b) 1, 1, 1, 2, 3, 3, 4, 5
(c) 1, 2, 3, 4, 5, 6, 7, 8
(d) 1.5, 1.5, 1.5, 4, 5.5, 5.5, 7, 8

3. Wenn Sie 3-mal mit Zurücklegen aus einer Urne mit 6 unterscheidbaren Kugeln ziehen, dann beträgt die Wahrscheinlichkeit dafür, dass alle gezogenen Kugeln unterschiedlich sind:

- (a) $\frac{3!}{6^3}$
(b) $\frac{(6)_3}{6^3}$
(c) $\frac{6!}{3^6}$
(d) $\frac{3!}{3^6}$

4. Von den 200 Studierenden, die in einem bestimmten Semester sowohl an der Mathematik- als auch an der Statistikklausur teilgenommen haben, haben 160 Studierende die Statistikklausur bestanden, 140 Studierende haben die Mathematikklausur bestanden und 130 Studierende haben sowohl die Statistik- als auch die Mathematik- klausur bestanden. Mit welcher Wahrscheinlichkeit hat ein(e) zufällig ausgewählte(r) Studierende(r) wenigstens eine der beiden Klausuren bestanden?

(a) 75%

(b) 80%

(c) 85%

(d) 90%

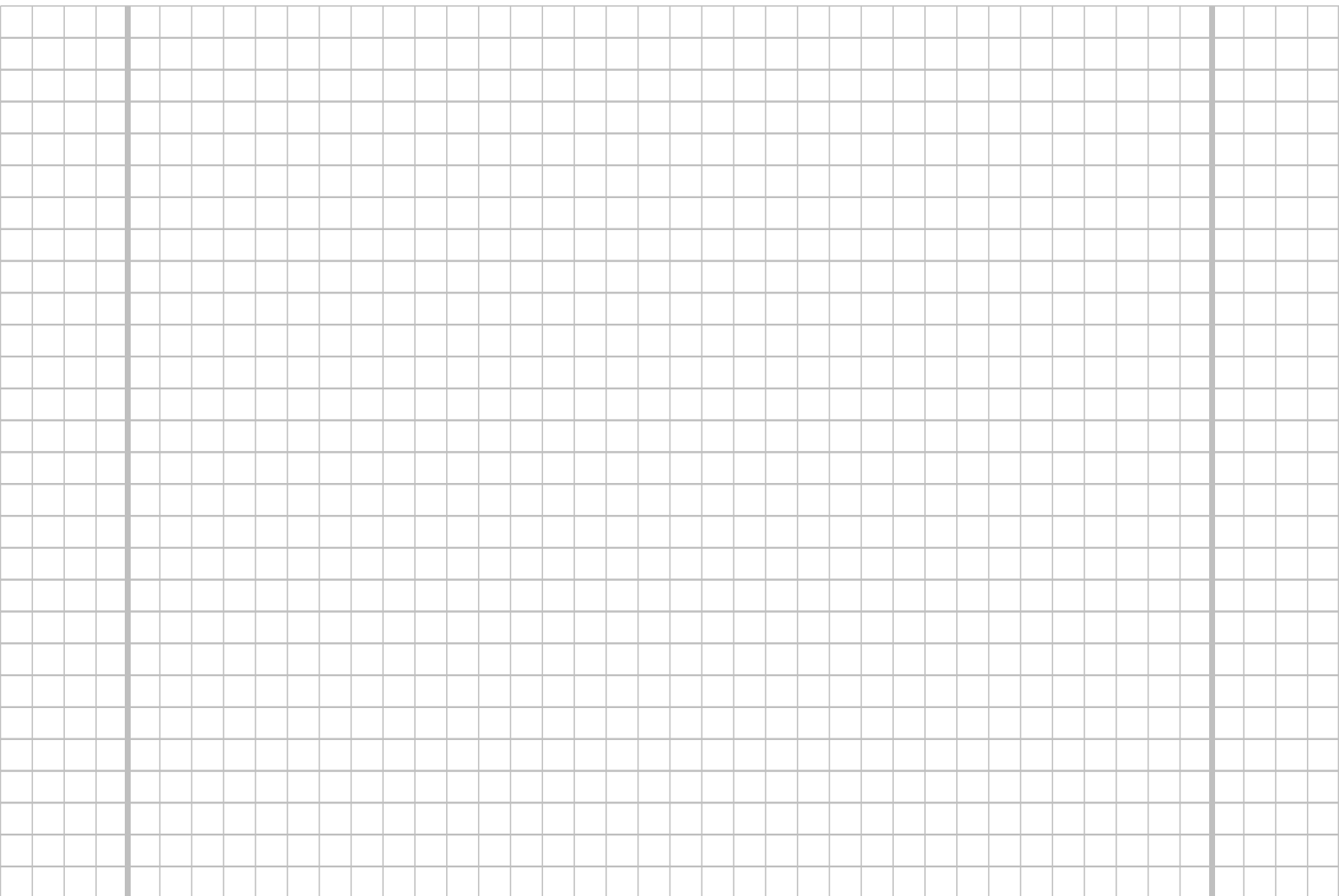
Aufgabe 3 (4 + 1 + 5 + 1 = 11 Punkte)

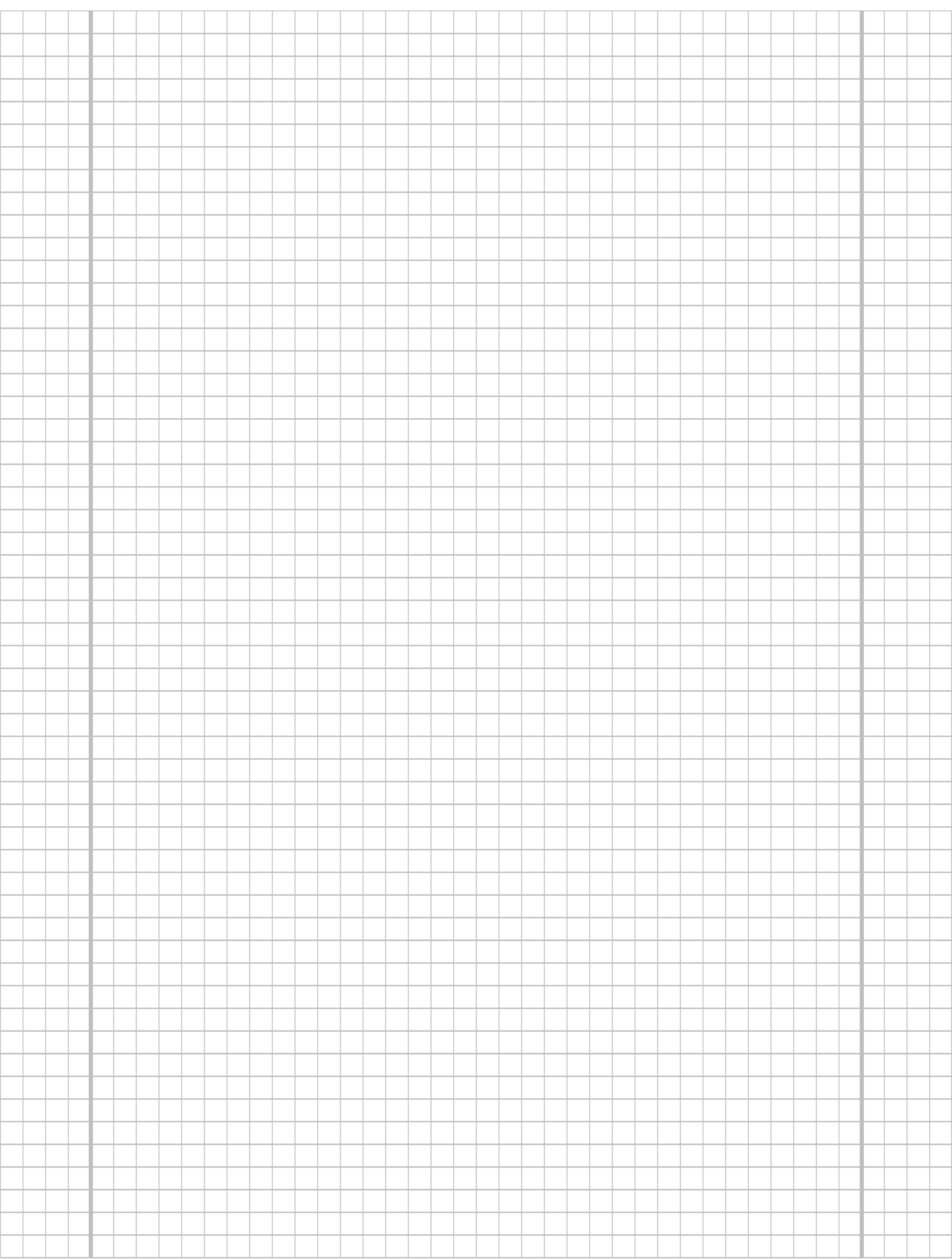
Zu einem erhobenen Merkmal X sei die folgende empirische Verteilungsfunktion gegeben:

$$F(x) = \begin{cases} 0.00 & \text{für } x < 2 \\ 0.02 & \text{für } 2 \leq x < 4 \\ 0.36 & \text{für } 4 \leq x < 6 \\ 0.64 & \text{für } 6 \leq x < 8 \\ 0.84 & \text{für } 8 \leq x < 10 \\ 0.94 & \text{für } 10 \leq x < 12 \\ 1.00 & \text{für } x \geq 12 \end{cases}$$

Außerdem sei die Länge der Urliste $n = 50$ bekannt.

- (a) Erstellen Sie eine Tabelle der absoluten und relativen Häufigkeiten.
- (b) Wie groß ist der Anteil der Urlisteneinträge, die Werte von weniger als 5 annehmen?
- (c) Berechnen Sie den arithmetischen Mittelwert und die empirische Varianz des Merkmals X .
- (d) Bestimmen Sie einen Median des Merkmals X .





Aufgabe 4 (6 + 4 + 3 + 3 + 3 = 19 Punkte)

Zu einem kardinalskalierten Merkmal sei die folgende (zur einfacheren Bearbeitung der Aufgabe bereits sortierte) Urliste der Länge $n = 30$ gegeben:

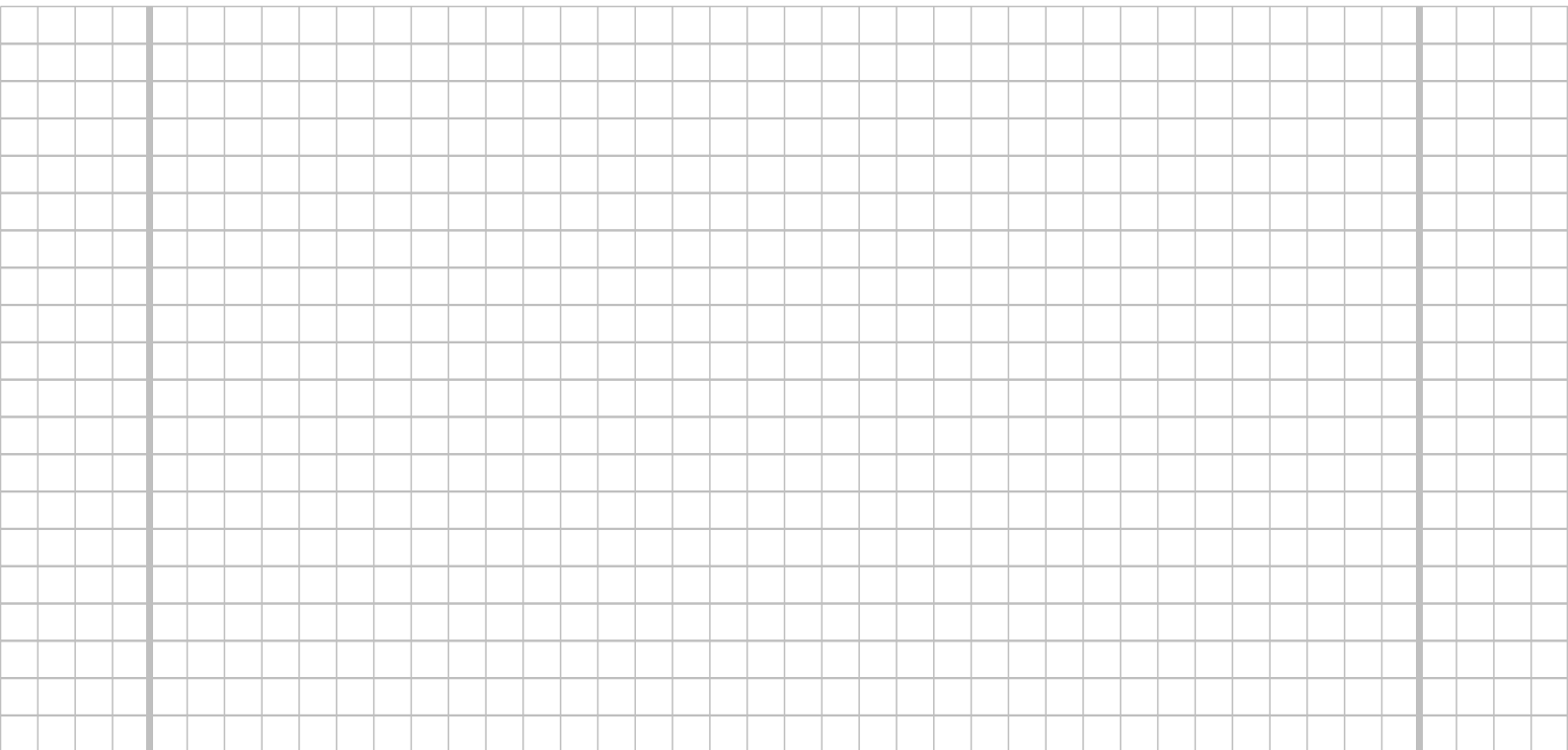
11.70, 12.50, 13.00, 14.93, 16.42, 17.03, 17.36, 17.68, 17.73, 18.62, 18.75, 19.06, 19.15, 19.21, 19.24, 19.37, 19.71, 19.91, 20.30, 20.49, 20.66, 20.81, 20.84, 20.94, 21.27, 21.40, 21.52, 21.61, 21.81, 21.99

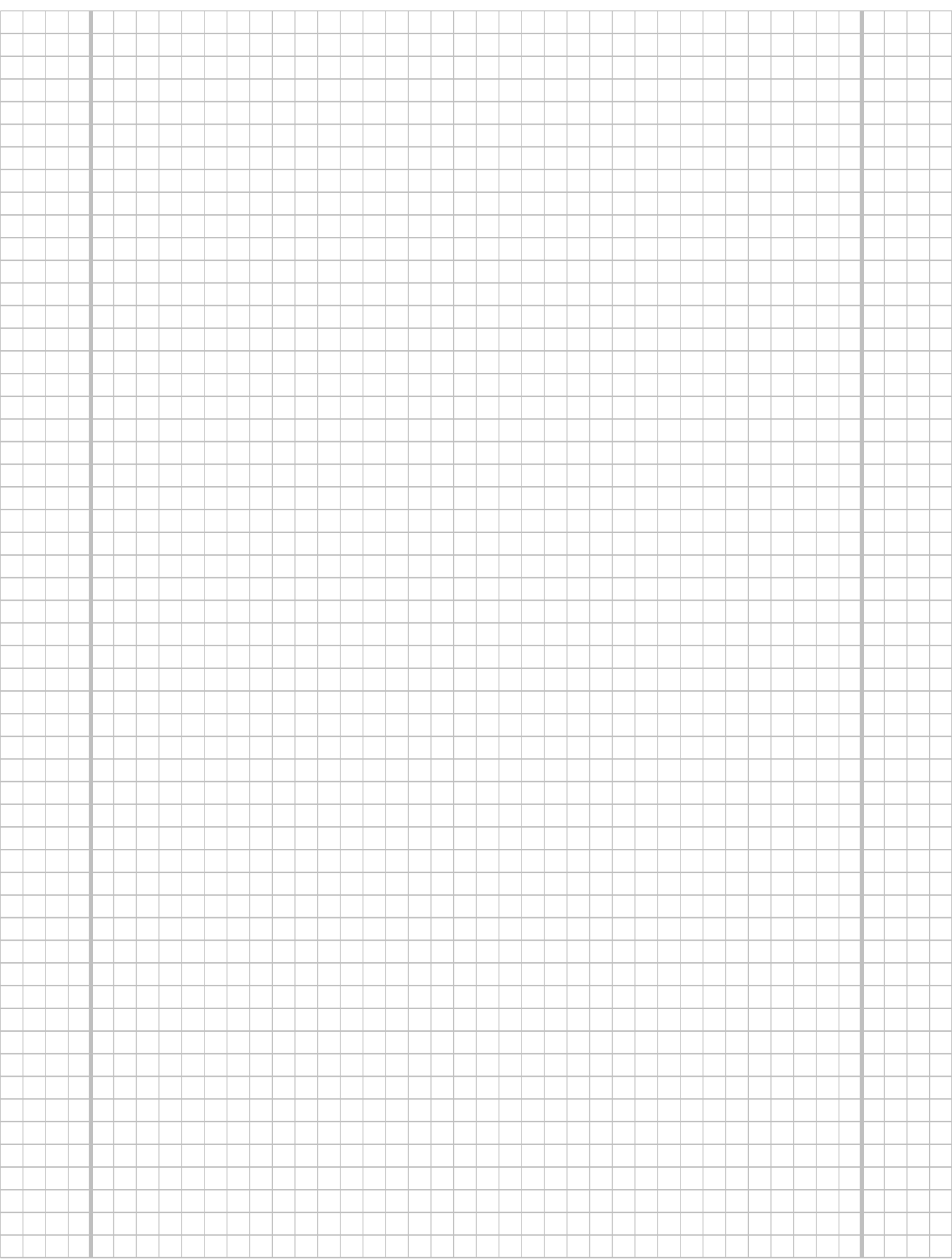
- (a) Führen Sie eine Klassierung der erhobenen Daten auf Grundlage der Klassen

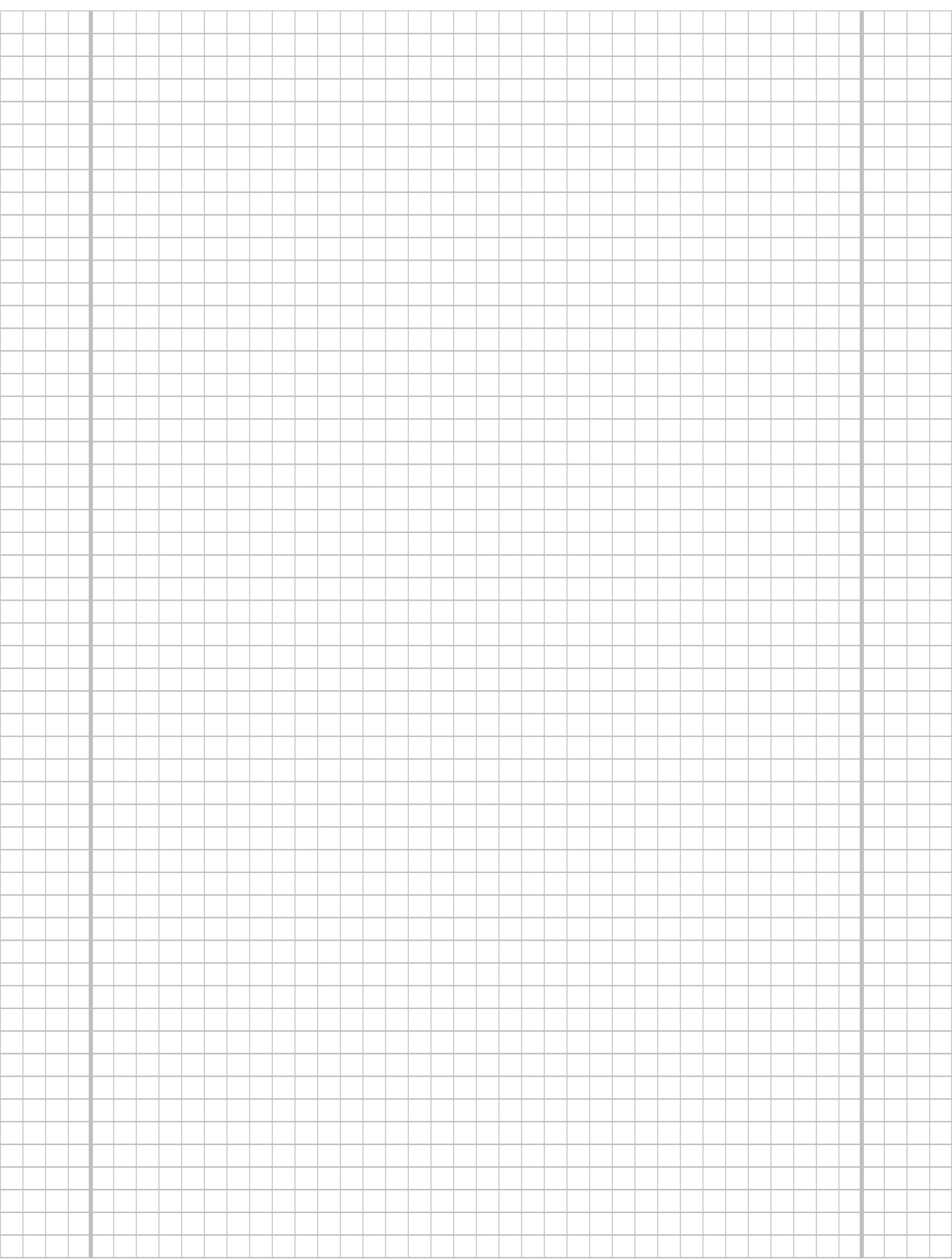
$$K_1 = (10, 14], K_2 = (14, 18], K_3 = (18, 20], K_4 = (20, 22]$$

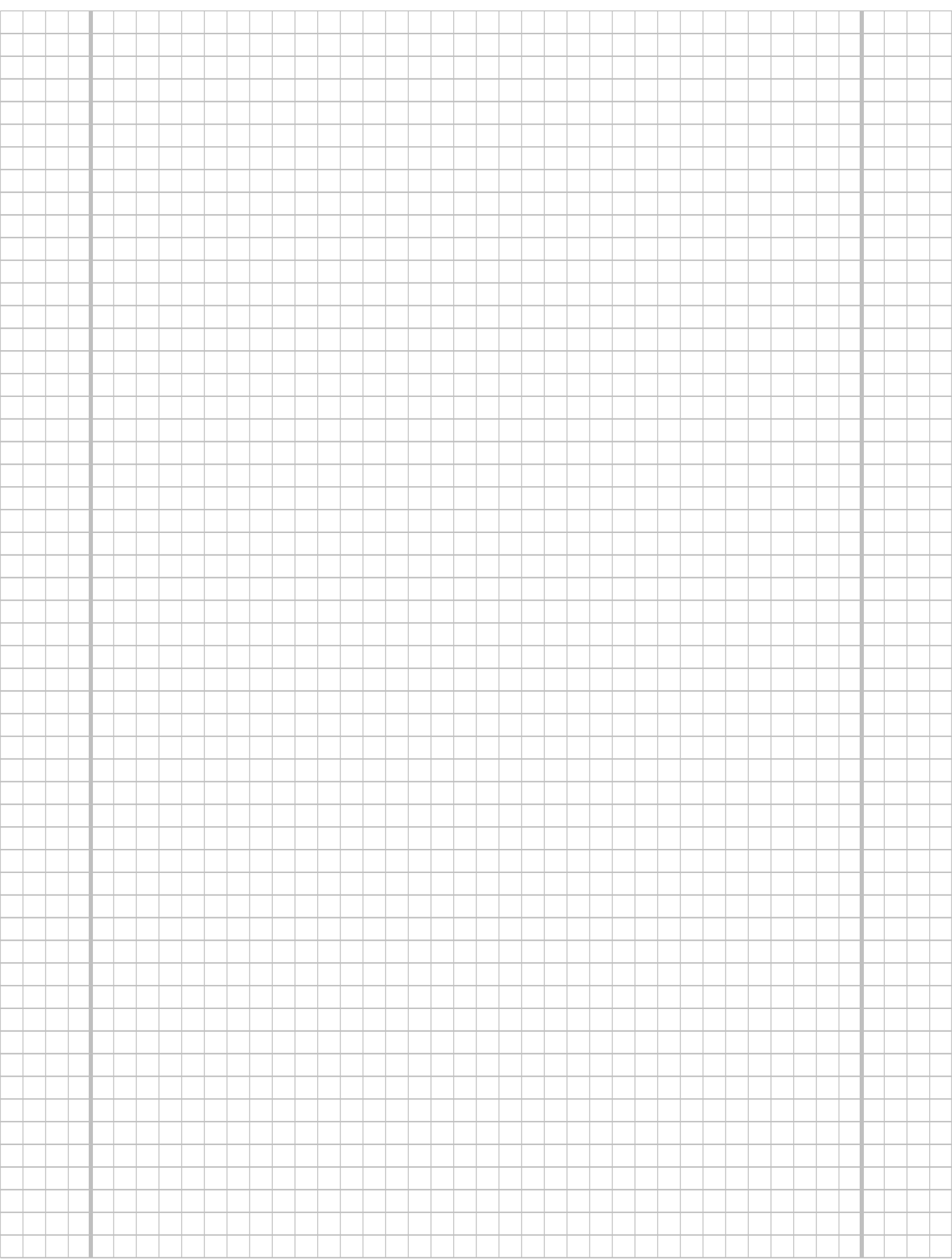
durch. Geben Sie insbesondere die jeweiligen Klassenbreiten, Klassenmitten, absoluten und relativen Klassenhäufigkeiten, Häufigkeitsdichten sowie die Werte der empirischen Verteilungsfunktion an den Klassengrenzen an.

- (b) Stellen Sie die (approximative) empirische Verteilungsfunktion des Merkmals aus der Klassierung der Daten auf.
- (c) Berechnen Sie aus den klassierten Daten den (approximativen) arithmetischen Mittelwert der Daten. Wie groß ist die relative Abweichung vom tatsächlichen (aus der Urliste bestimmten) Mittelwert von 18.834?
- (d) Bestimmen Sie (aus der Urliste) die Anzahl von Merkmalswerten zwischen 12 und 18. Welche Näherung für diese Anzahl können Sie aus der in Teil (b) aufgestellten empirischen Verteilungsfunktion berechnen?
- (e) Bestimmen Sie (unter Verwendung der bekannten Konvention zur eindeutigen Festlegung von Quantilen bei kardinalskalierten Merkmalen) *das* obere Quartil sowohl exakt aus der Urliste als auch approximativ mit Hilfe der Verteilungsfunktion für die klassierten Daten.





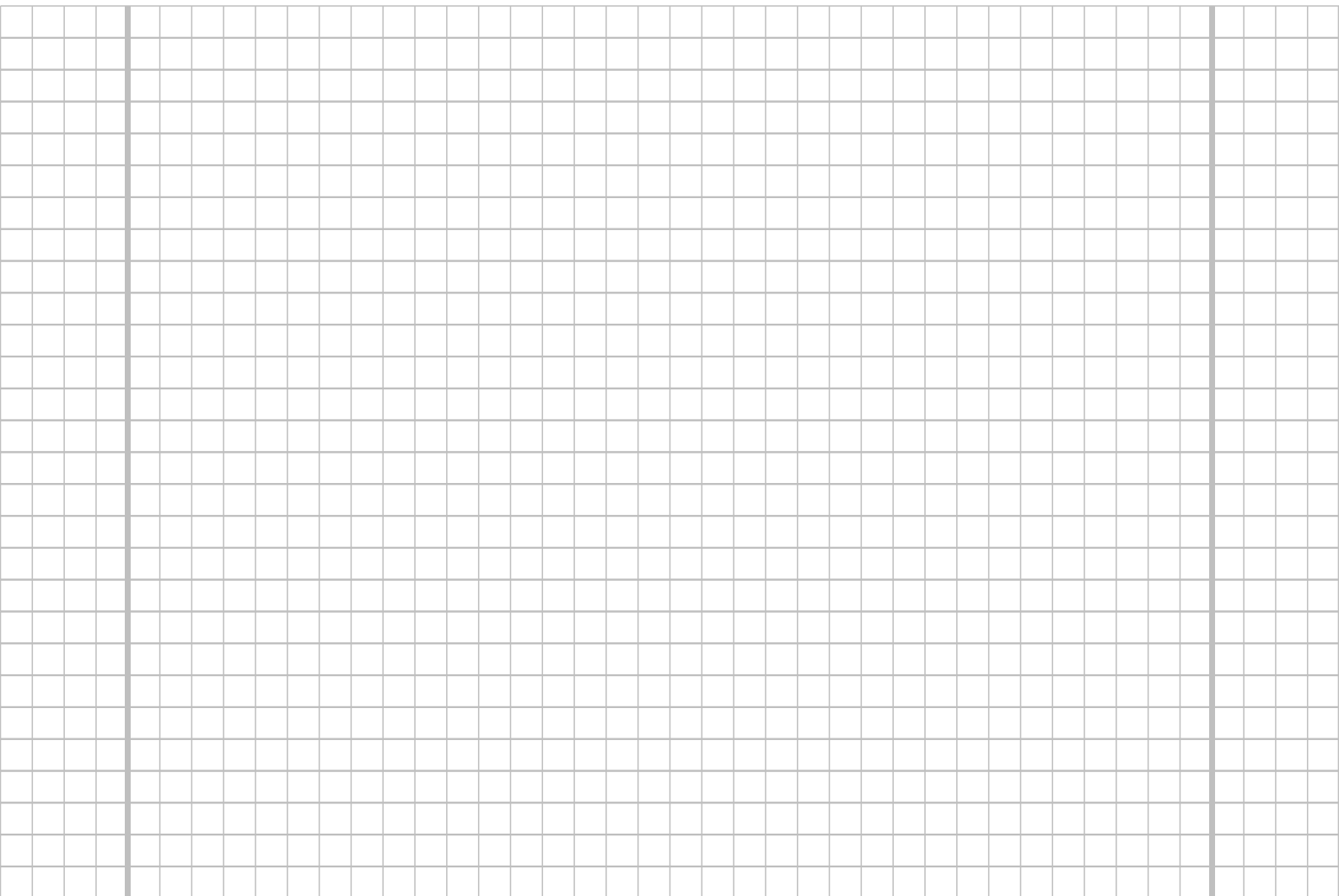


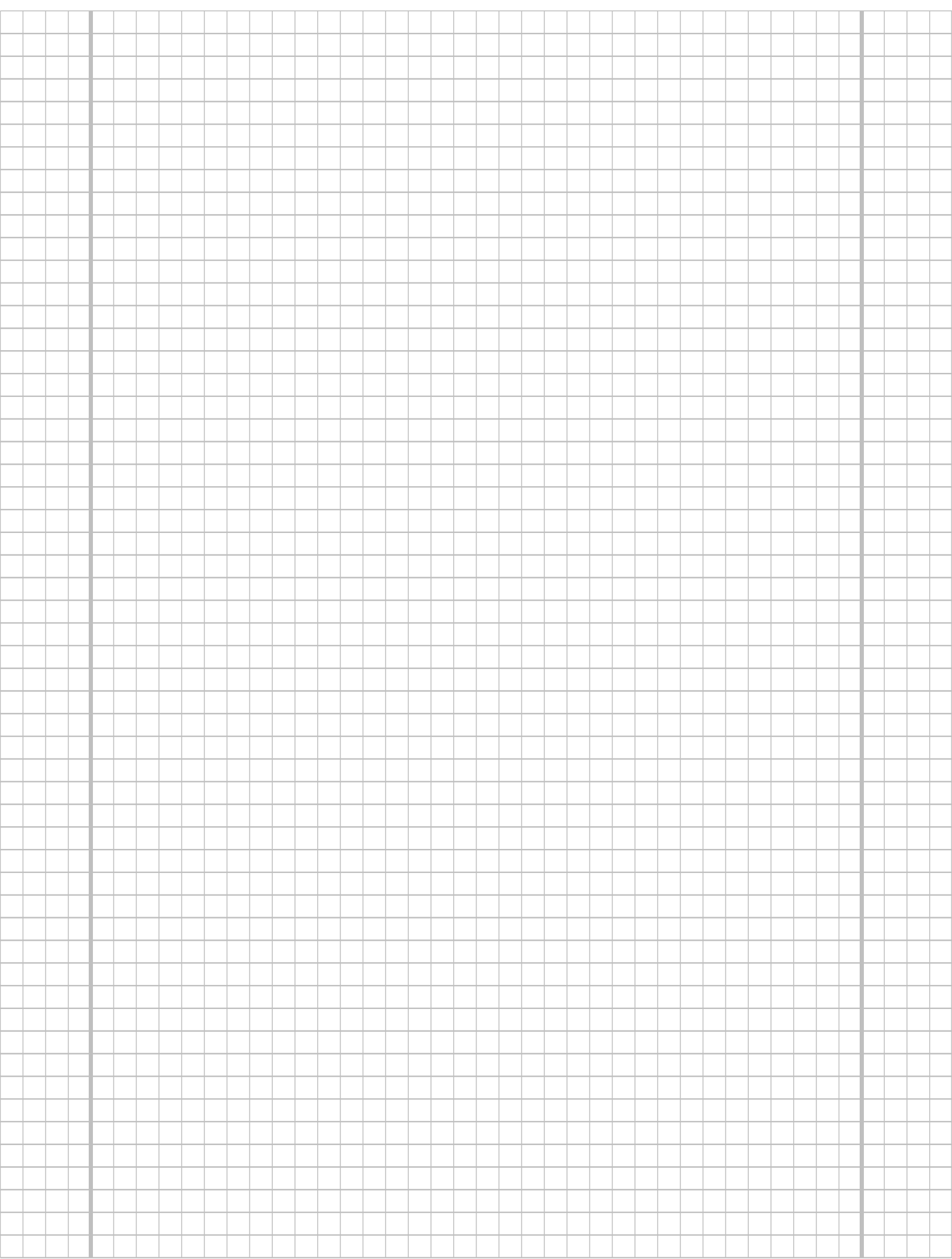


Aufgabe 5 (4 + 2 + 2 = 8 Punkte)

Zur Unterstützung der Bekämpfung einer Pandemie werden auch Schnelltests eingesetzt, die im Vergleich zu aufwändigeren Testverfahren eine geringere Zuverlässigkeit aufweisen. Gehen Sie im Folgenden davon aus, dass diese Tests eine tatsächlich vorhandene Infektion (bei asymptomatischen Personen) nur mit einer Wahrscheinlichkeit von 80% erkennen, während bei nicht infizierten (asymptomatischen) Personen fälschlicherweise mit einer Wahrscheinlichkeit von 0.15% ein positives Testergebnis entsteht. Gehen Sie außerdem davon aus, dass in einer zum betreffenden Zeitpunkt vergleichsweise günstigen Infektionslage die Wahrscheinlichkeit des tatsächlichen Vorliegens einer Infektion (bei asymptomatischen Personen) nur 0.05% beträgt.

- (a) Mit welcher Wahrscheinlichkeit wird ein Schnelltest bei einer (zufällig ausgewählten und asymptomatischen) untersuchten Person ein positives Ergebnis liefern?
- (b) Mit welcher Wahrscheinlichkeit ist eine (mit einem Schnelltest) positiv getestete (asymptomatische) Person tatsächlich gar nicht infiziert?
- (c) Vor dem Hintergrund des Ergebnisses aus Teil (b) wird der Nutzen der Schnelltests bei der aktuellen Infektionslage in Frage gestellt, indem pauschalisiert ein „hoher Anteil falscher Testergebnisse“ beklagt wird. Entkräftigen Sie diese pauschalisierte Aussage durch die Berechnung der Wahrscheinlichkeit dafür, dass eine (asymptomatische) Person in der beschriebenen Infektionslage ein *korrektes* Testergebnis erhält.

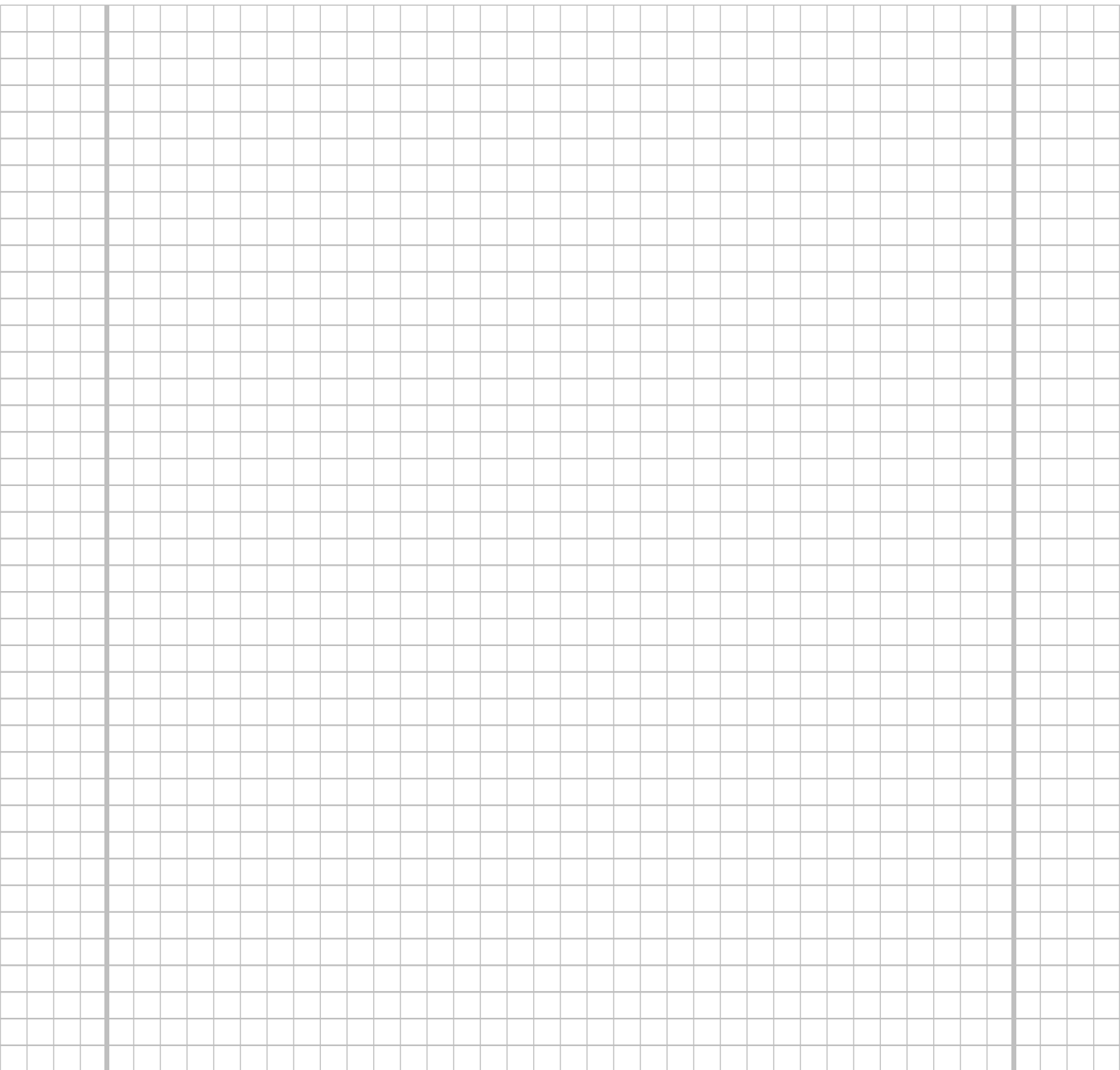


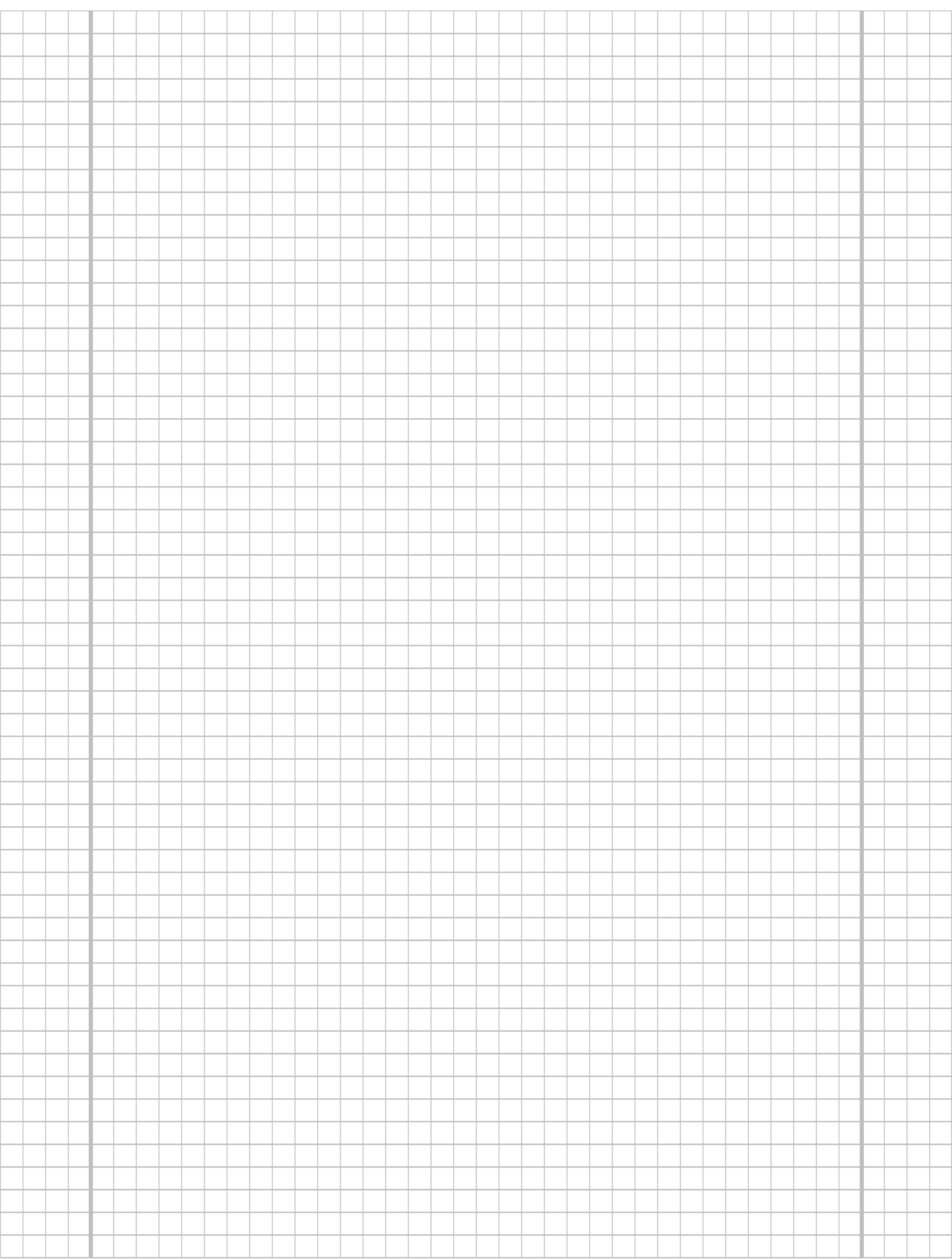


Aufgabe 6 (2 + 1 + 2 = 5 Punkte)

Die Lebensdauer bis zum Ausfall einer Server-Festplatte lasse sich als eine exponentialverteilte Zufallsvariable auffassen. Im Mittel vergehen bis zum Ausfall 8 Jahre.

- (a) Welche Standardabweichung hat die Lebensdauer bis zum Ausfall?
- (b) Mit welcher Wahrscheinlichkeit beträgt die Lebensdauer bis zum Ausfall mehr als 10 Jahre?
- (c) Berechnen Sie das 0.80-Quantil der Lebensdauer bis zum Ausfall.





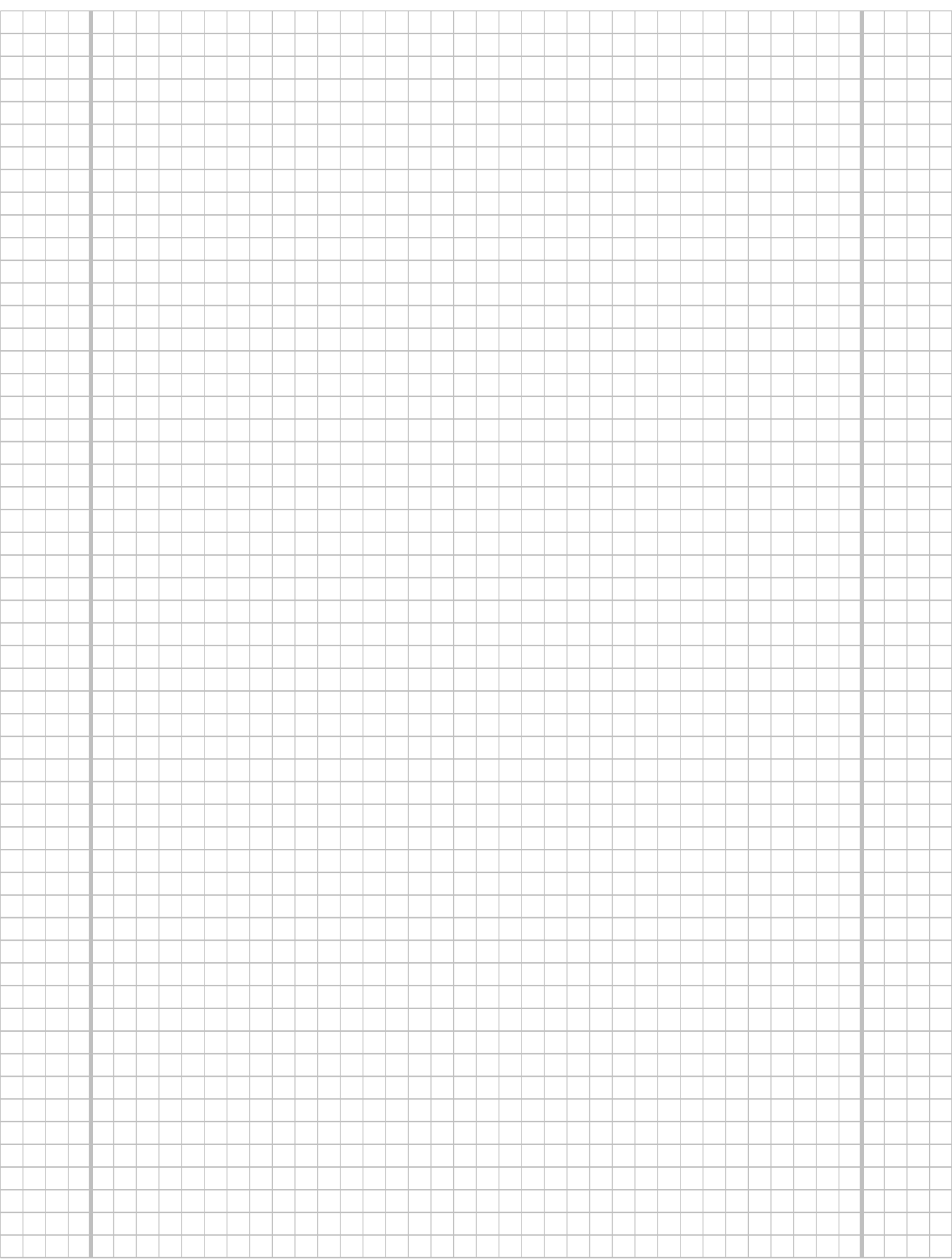
Aufgabe 7 (5 + 2 + 5 + 6 + 3 + 2 = 23 Punkte)

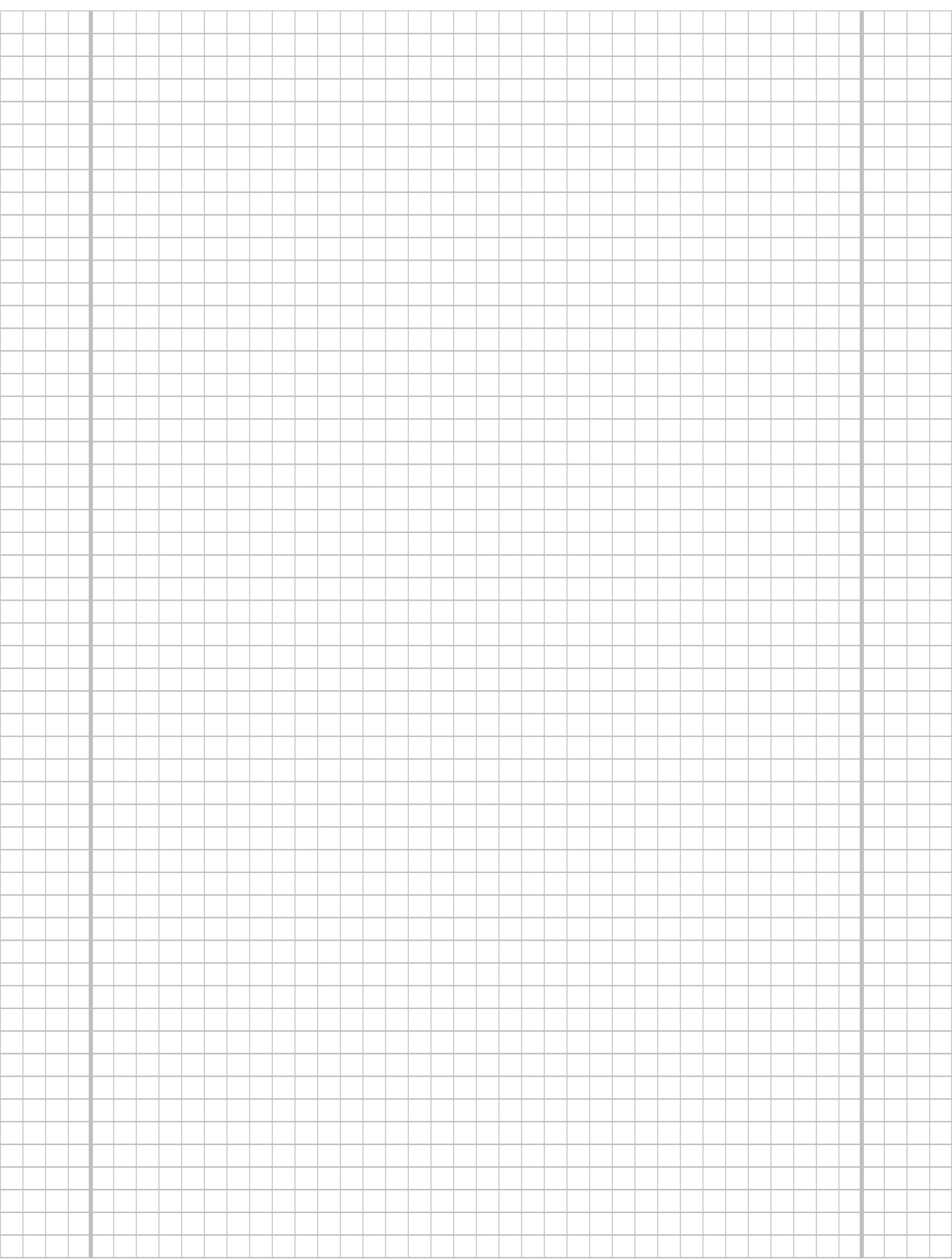
Die Verteilung einer stetigen Zufallsvariablen X sei durch die folgende Dichtefunktion gegeben:

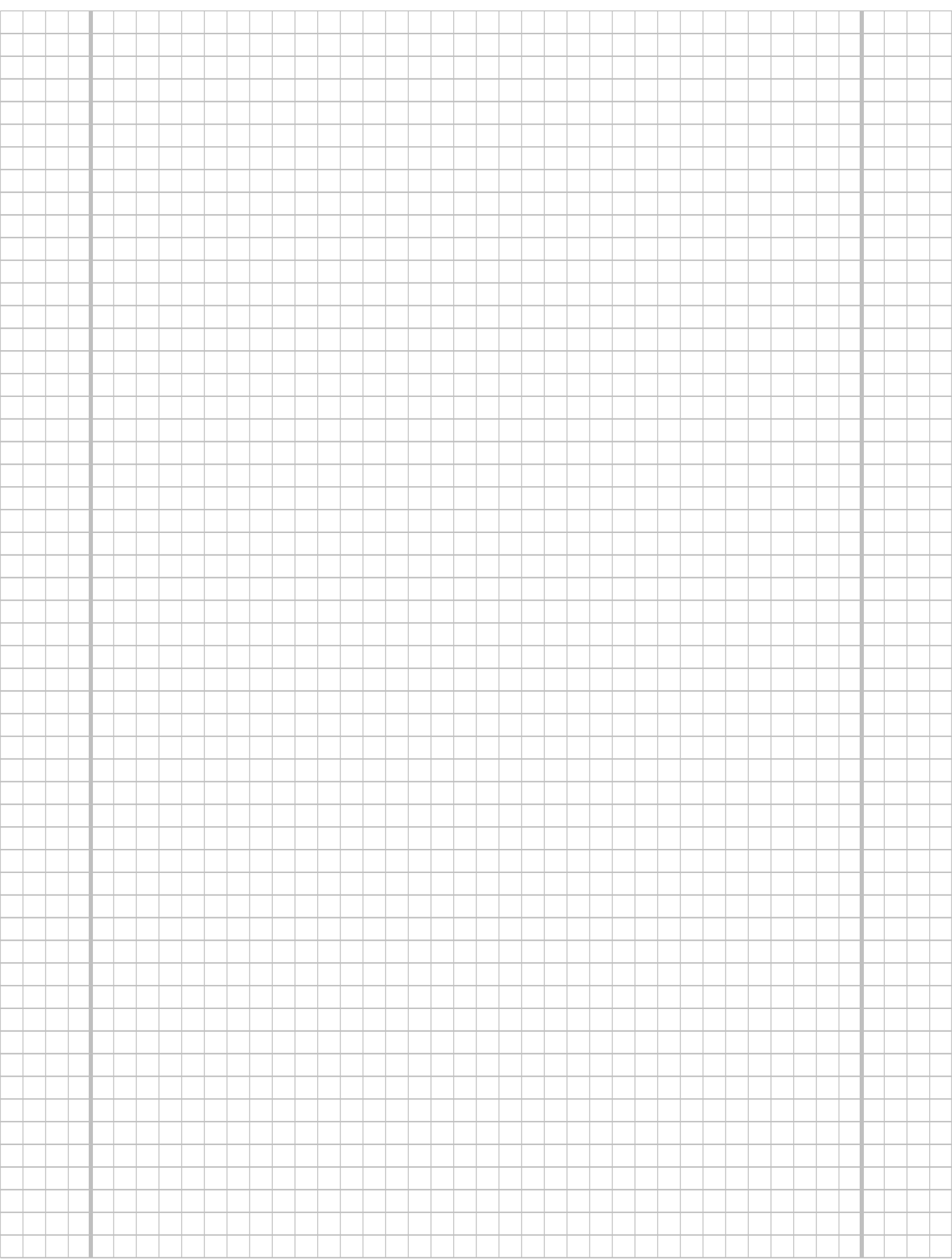
$$f_X(x) = \begin{cases} \frac{1}{6}x + \frac{1}{3} & \text{für } -2 \leq x < 1 \\ -\frac{1}{2}x + 1 & \text{für } 1 \leq x \leq 2 \\ 0 & \text{sonst} \end{cases}$$

- (a) Berechnen Sie die Verteilungsfunktion F_X von X .
- (b) Berechnen Sie $P(\{X > \frac{3}{2}\})$ und $P(\{-1 \leq X \leq \frac{3}{2}\})$.
- (c) Bestimmen Sie den Median und das obere Quartil von X .
- (d) Bestimmen Sie den Erwartungswert $E(X)$.
- (e) Bestimmen Sie eine Dichtefunktion von $Y := 2 \cdot X + 4$.
- (f) Bestimmen Sie Erwartungswert und Varianz von $Y := 2 \cdot X + 4$.

Hinweis: $\text{Var}(X) = 0.7\bar{2}$







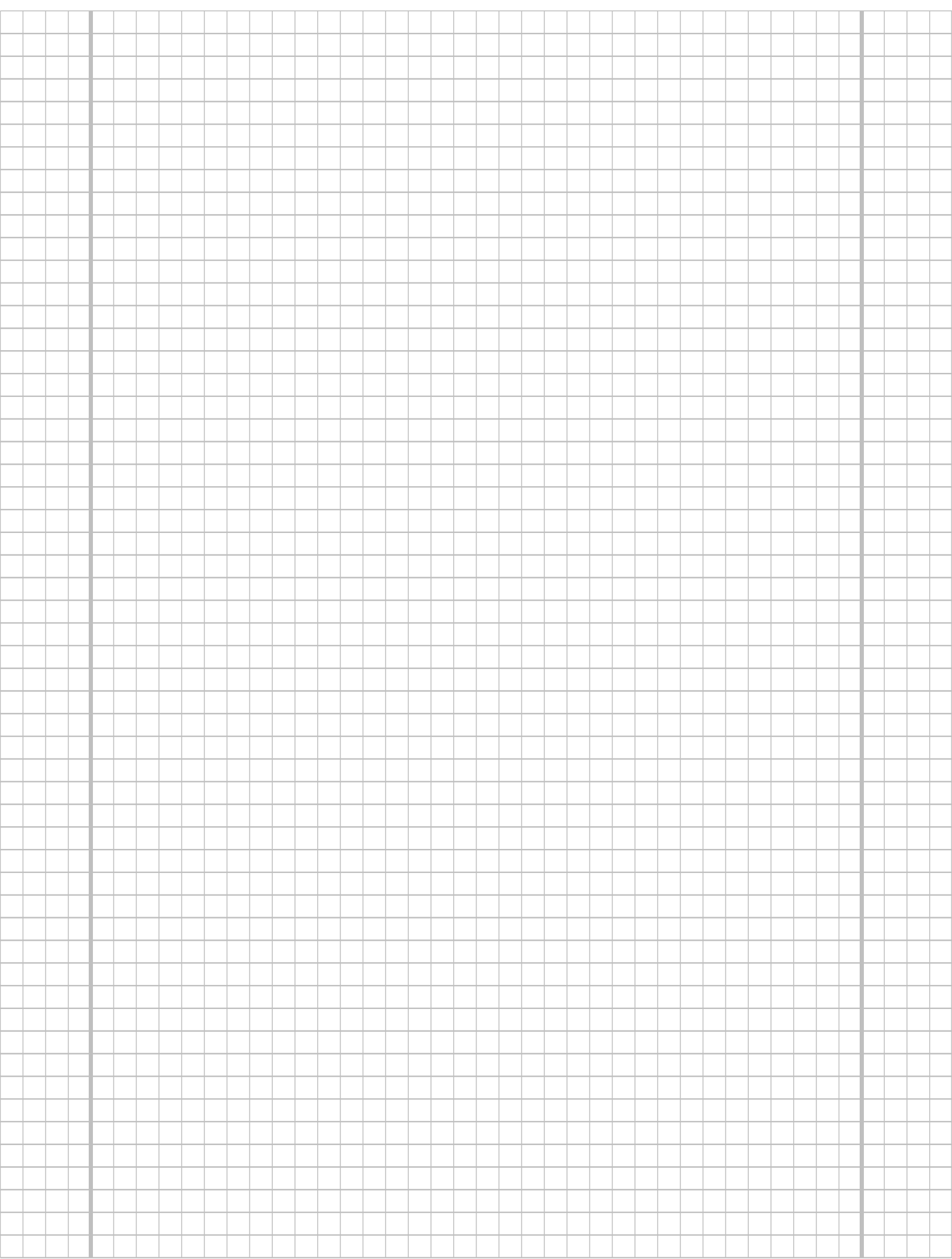
Aufgabe 8 (2 + 3 + 9 + 3 = 17 Punkte)

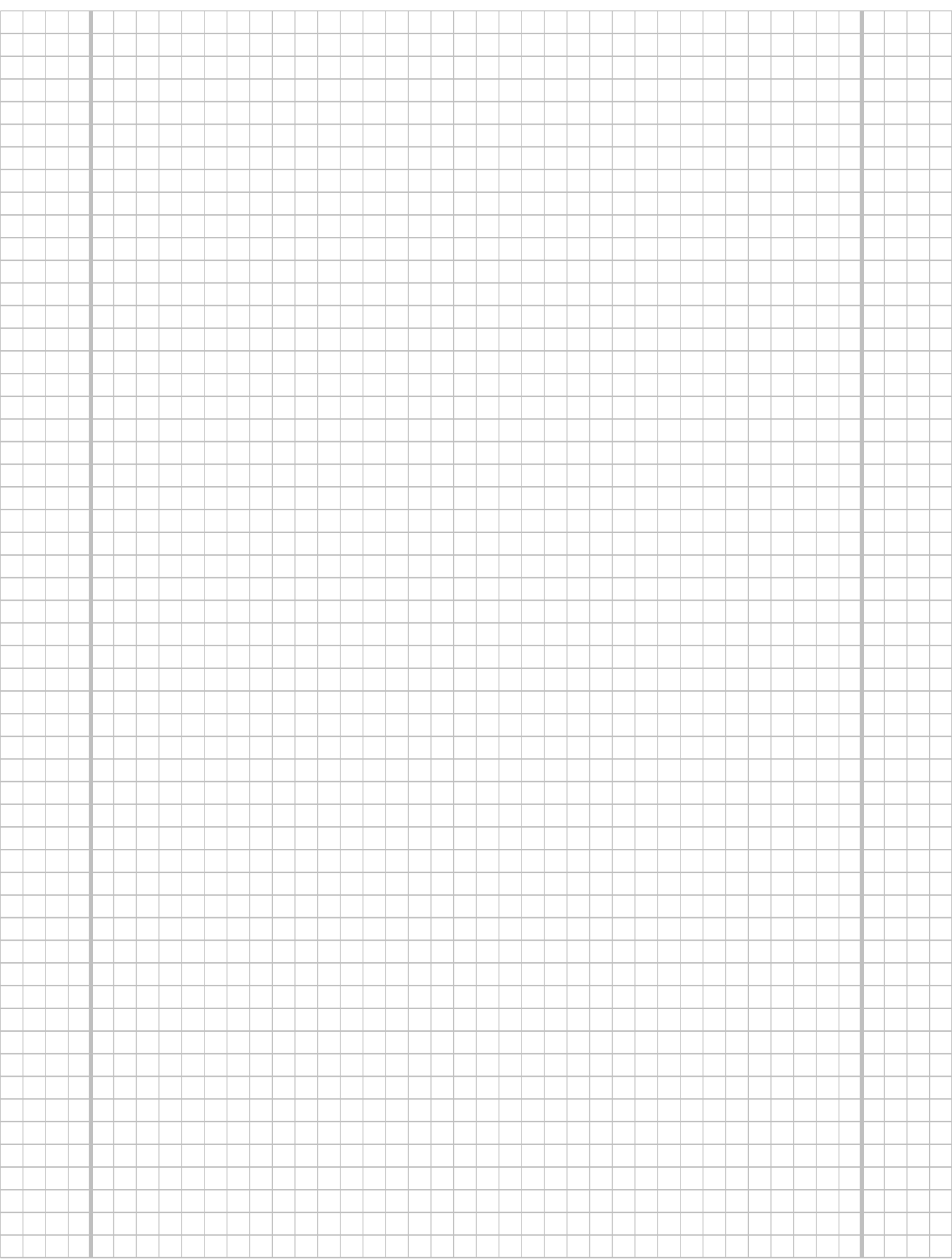
Gegeben sei die folgende Tabelle der gemeinsamen Wahrscheinlichkeitsverteilung zu einem zweidimensionalen diskreten Zufallsvektor (X, Y) :

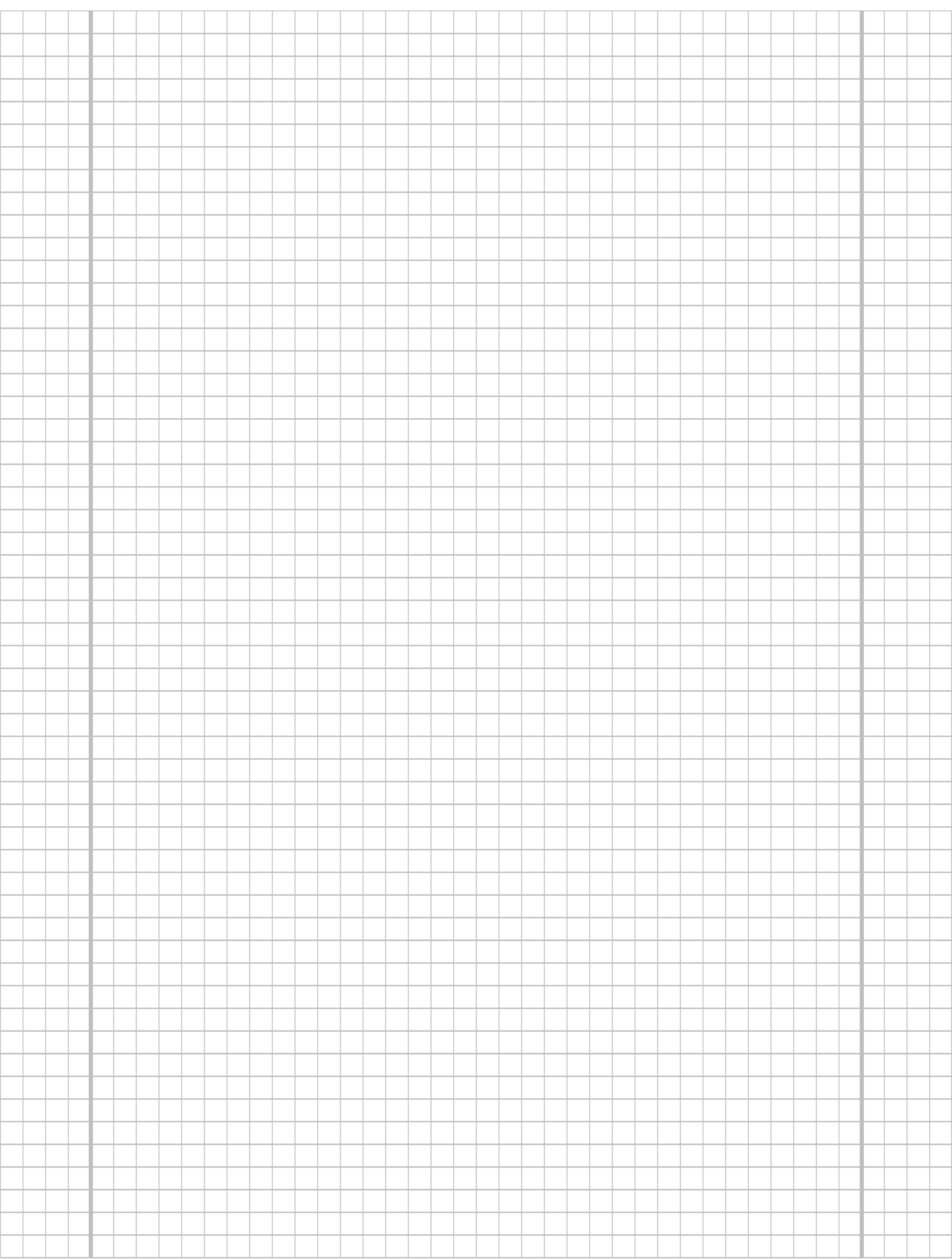
$X \setminus Y$	-2	0	4	$p_{i\cdot}$
1	0.06	0.06	0.18	
2	0.14	0.28	0.08	
4	0.1	0.06	0.04	
$p_{\cdot j}$				

- (a) Ergänzen Sie die obige Tabelle (in den vorgesehenen Feldern) um ihre Randverteilungen.
- (b) Geben Sie die bedingten Wahrscheinlichkeitsverteilungen von Y unter der Bedingung $X = x_i$ für alle $x_i \in T(X)$ über die zugehörigen (bedingten) Wahrscheinlichkeitsfunktionen an.
- (c) Berechnen Sie $E(X)$, $E(Y)$, $\text{Var}(X)$, $\text{Var}(Y)$, $\text{Cov}(X, Y)$ sowie $\text{Korr}(X, Y)$.
- (d) Berechnen Sie $E(-3X + 2Y)$ sowie $\text{Var}(-3X + 2Y)$.







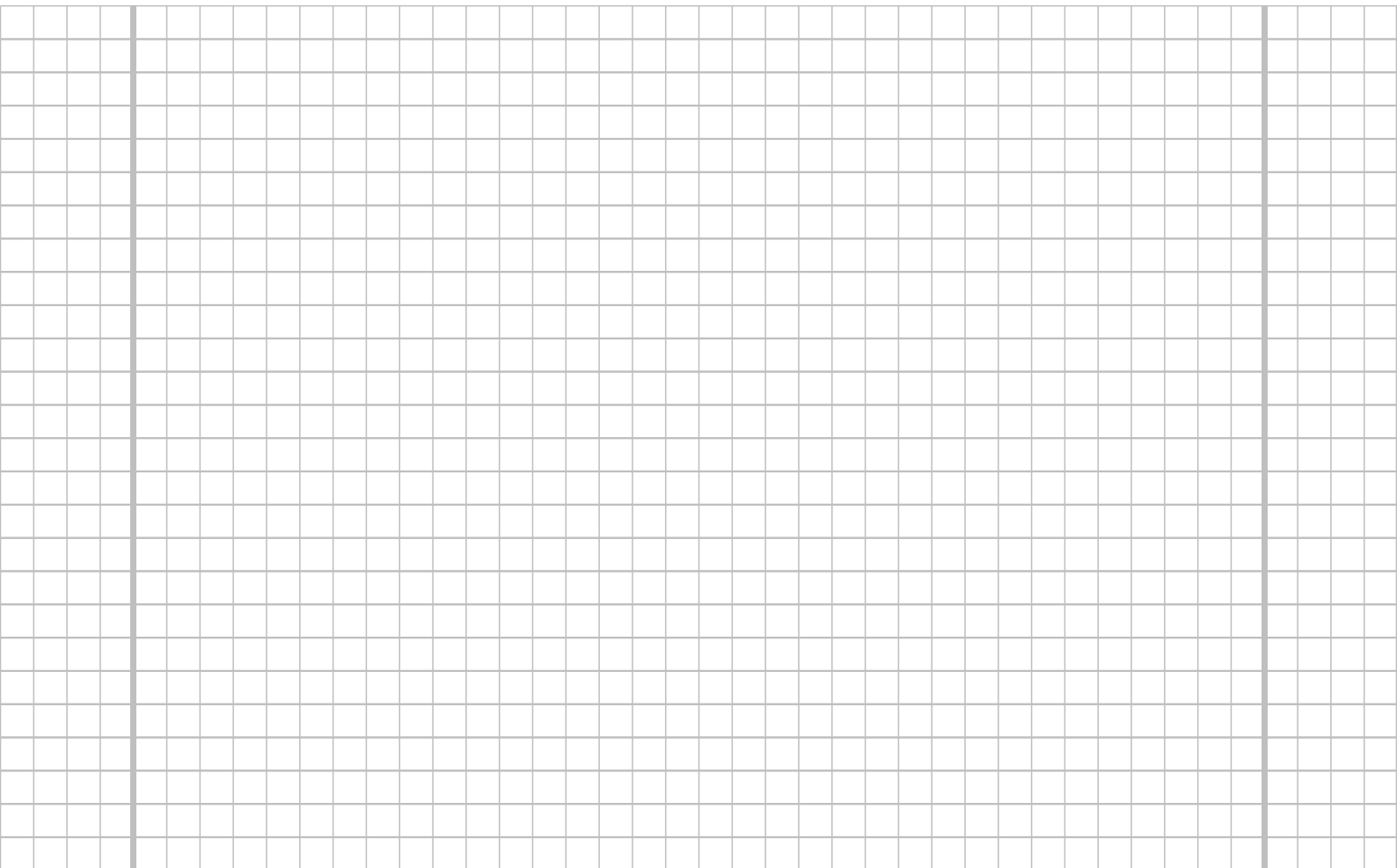


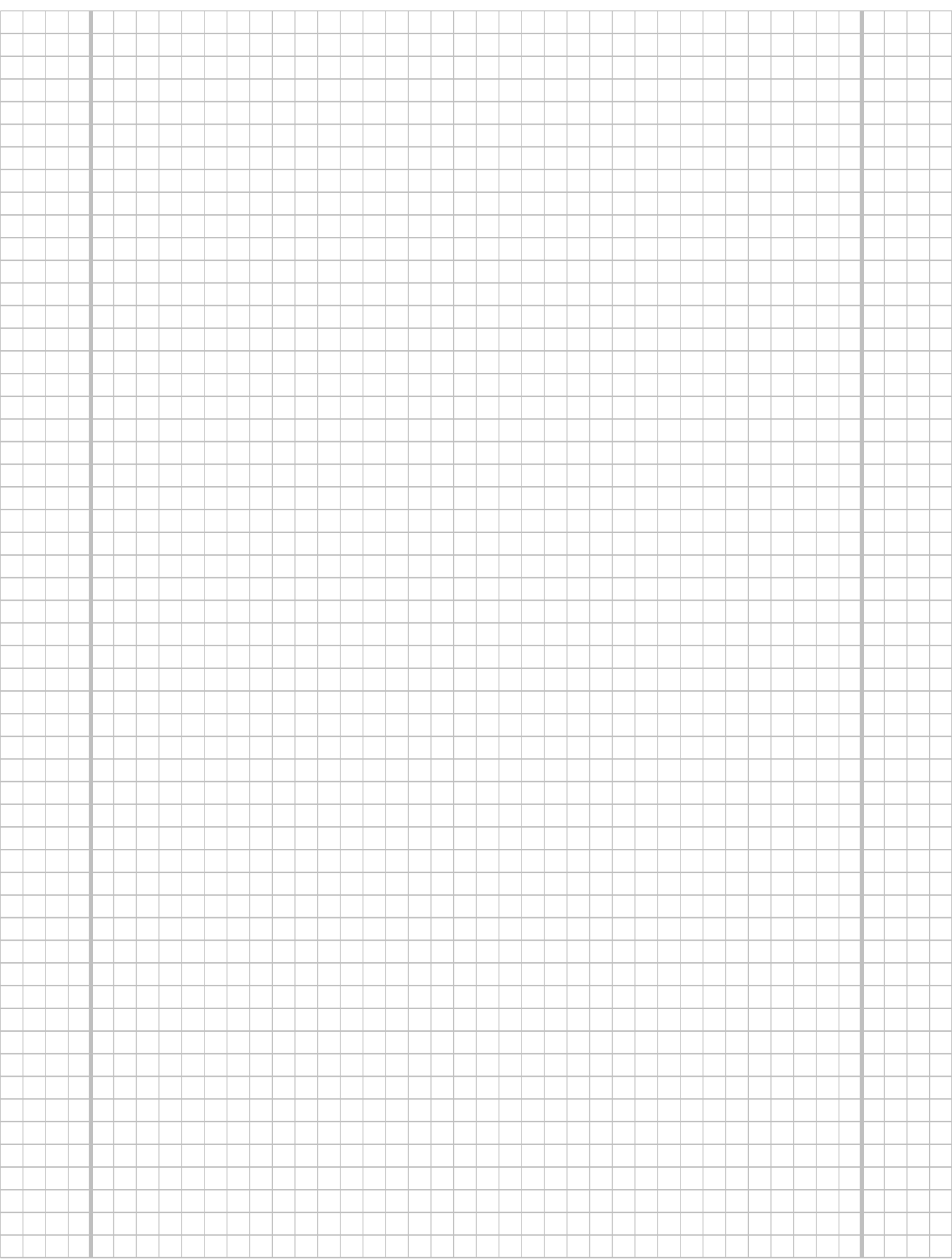
Aufgabe 9 (3 + 2 + 4 = 9 Punkte)

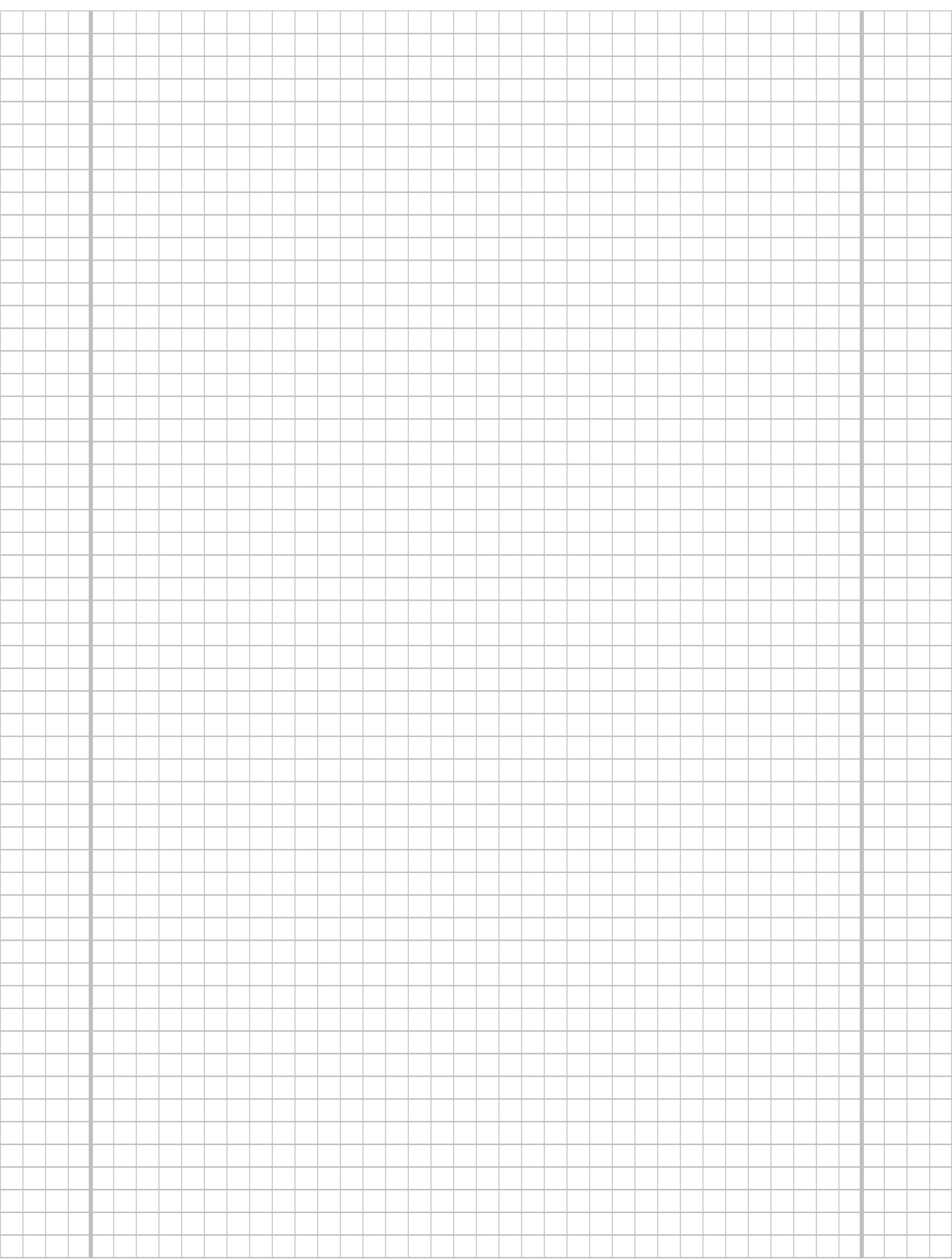
In einem (fiktiven) Impfzentrum stehen täglich 360 Impfdosen zur Verfügung. Auf Basis der Erfahrungen aus der jüngeren Vergangenheit soll davon ausgegangen werden, dass vergebene Impftermine mit einer Wahrscheinlichkeit von 10% (ohne rechtzeitige Absage) nicht in Anspruch genommen werden. Für die individuelle Inanspruchnahme der Impftermine soll dabei Unabhängigkeit unterstellt werden. Um die Verschwendung ungenutzter Impfdosen zu reduzieren, plant das Impfzentrum, täglich zusätzliche Feierabendtermine zu vergeben, zu denen eine Impfung in Aussicht gestellt, aber nicht garantiert wird.

- (a) Wie ist die Anzahl der (täglich) nach Ablauf der regulären Impftermine *übrig gebliebenen* Impfdosen Y exakt verteilt? Geben Sie auch den Erwartungswert $E(Y)$ sowie die Varianz $\text{Var}(Y)$ der Anzahl der übrig gebliebenen Impfdosen an.
- (b) Das Impfzentrum entschließt sich dazu, täglich 30 zusätzliche Feierabendtermine zu vergeben. Mit welcher (näherungsweise mit einem geeigneten Grenzwertsatz zu bestimmenden) Wahrscheinlichkeit reichen die übrig gebliebenen Impfdosen aus, um alle 30 Impfwillingen der Feierabendtermine auch tatsächlich impfen zu können?
- (c) Verwenden Sie einen geeigneten Grenzwertsatz, um näherungsweise einen um den zugehörigen Erwartungswert symmetrischen Bereich zu bestimmen, in dem sich die Anzahl der (täglich) übrig gebliebenen Impfdosen mit einer Wahrscheinlichkeit von 0.95 realisiert. (*Die Grenzen des Bereichs müssen nicht gerundet werden!*)

Hinweis: Verwenden Sie zur Bearbeitung der Aufgabe die Tabelle zur Standardnormalverteilung auf Seite 26!







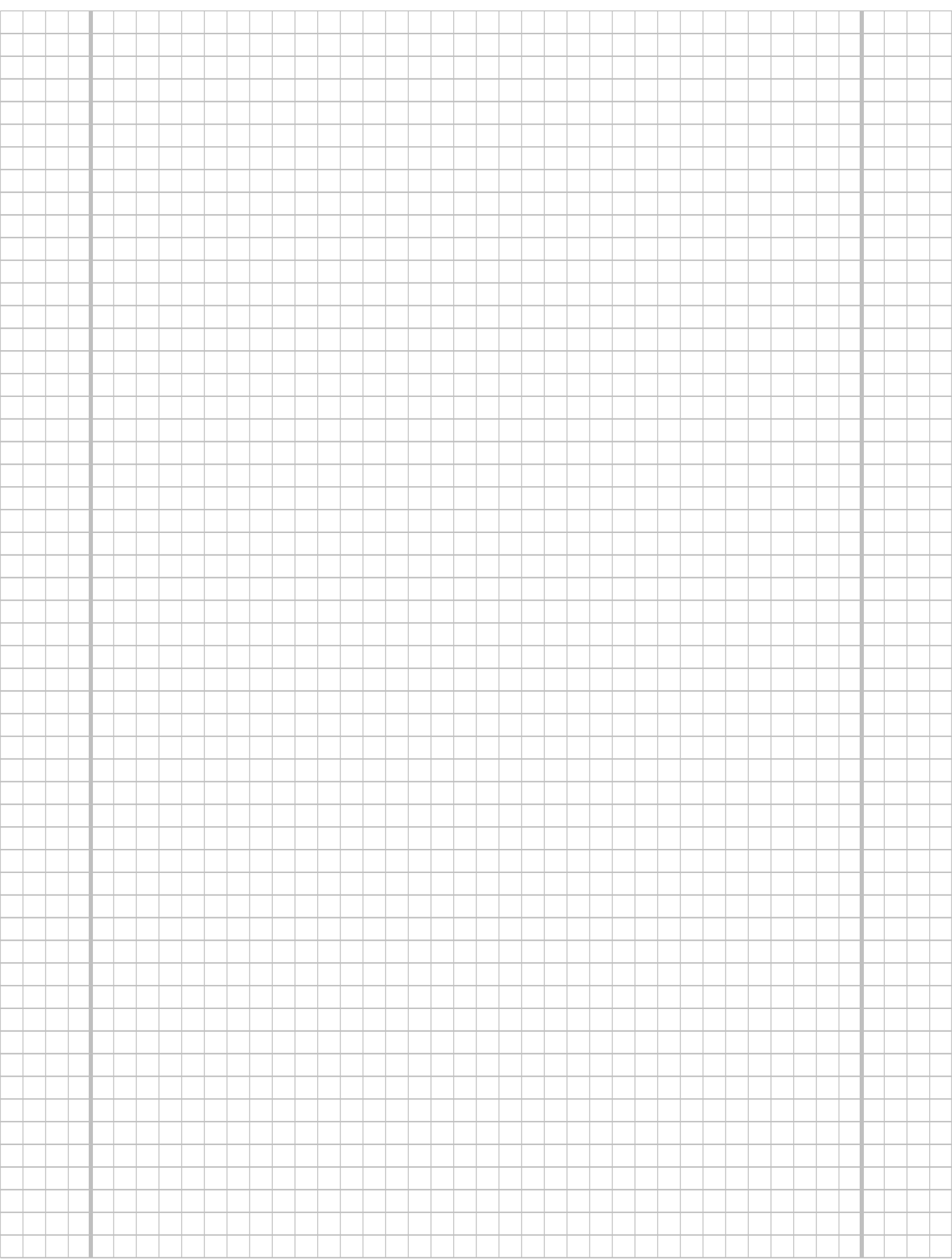


Tabelle zur Standardnormalverteilung

$$F_{N(0,1)}(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$$

$$\Phi(x) = 1 - \Phi(-x)$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998