

Statistik Aufgabensammlung

Sommersemester 2017

Prof. Dr. Stefan Etschberger – Hochschule Augsburg

Anmerkungen zu den Übungsaufgaben:

- ▶ Nach der Vorlesung finden Sie jeweils in der Aufgabensammlung die für die jeweilige Woche zu bearbeitenden Aufgaben; besprochen und gelöst werden die Aufgaben in der darauf folgenden Woche in den Übungsgruppen.
- ▶ Grundlagen in R sind ein wichtiger, obligatorisch zu erlernender Bestandteil des Kurses; alle in den Übungsaufgaben behandelten Lösungen in R sind prüfungsrelevant und müssen auch bei veränderter Aufgabenstellung (ohne Rechner) gelöst werden können.
- ▶ Es gibt für die Klausur keine Einschränkung auf nur eine Aufgabe mit R. Klausuraufgaben mit R könnten in der Prüfung bei verschiedenen Themen als Teilaufgabe oder als separate Aufgabe eingebaut sein. R-Teile in der Klausur können, müssen aber nicht als single choice formuliert sein.
- ▶ Es gibt *kein* vorgefertigtes „cheat-sheet“ mit den wichtigsten Funktionen in R für die Klausur; bitte schreiben Sie sich die wichtigsten Funktionen inkl. Parametern auf Ihre selbsterstellte Formelsammlung. Vorausgesetzt werden für die Klausur alle in den Lösungshinweisen der Übungsaufgaben verwendeten Funktionen.
- ▶ (Teil)aufgaben, deren Lösungen (auch) in R erarbeitet werden sollen, sind am Rand mit dem Symbol  gekennzeichnet.

Inhalt

Aufgaben zur mehrdimensionalen Differentialrechnung

Aufgabe 1: Maximum gesucht	4
Aufgabe 2: Extremwerte	5
Aufgabe 3: Kompakte Definitionsmengen . .	6
Aufgabe 4: Kritische Punkte	7
Aufgabe 5: Polynom	8

Aufgaben zu R Grundlagen

Aufgabe 6: RStudio und erste Versuche . . .	10
Aufgabe 7: Zuweisungen und Variablen . . .	12
Aufgabe 8: Vektoren	14
Aufgabe 9: Funktionen einer reellen Variablen	16
Aufgabe 10: Funktionsplots	19
Aufgabe 11: Mehrere Merkmale: Data Frames	21
Aufgabe 12: Skalenniveaus und Data Frames	23
Aufgabe 13: Datenimport aus Textdateien . .	24
Aufgabe 14: R-Skripten als Logbuch	25
Aufgabe 15: Deskriptives mit R	26
Aufgabe 16: Einfache Grafiken in R	28
Aufgabe 17: Emp. Vtlgs.f. Quantil Boxplot .	29

Aufgaben zur deskriptiven Statistik

Aufgabe 18: Häufigkeit1b	30
Aufgabe 19: Lageparameter	31
Aufgabe 20: Lageparameter	32
Aufgabe 21: Lage Streuung	33
Aufgabe 22: Lage Streuung Vtgl.fkt.	34
Aufgabe 23: Lageparameter Konzentration . .	35
Aufgabe 24: Lageparameter Konzentration . .	36
Aufgabe 25: Konzentration	37
Aufgabe 26: Konzentration	38
Aufgabe 27: Lage Konzentration	39
Aufgabe 28: Preisindex	40
Aufgabe 29: Preisindex	41
Aufgabe 30: Rangkorrelation	42
Aufgabe 31: Lage Korrelation	43
Aufgabe 32: Kontingenzkoeffizient	44
Aufgabe 33: Kontingenzkoeffizient	45
Aufgabe 34: Korrelation Regression	46
Aufgabe 35: Korrelation Regression	47
Aufgabe 36: Korrelation Regression	48
Aufgabe 37: Korrelation Regression	49
Aufgabe 38: Korrelation Regression	50
Aufgabe 39: Regression	51
Aufgabe 40: Regression	52

Aufgaben zur Kombinatorik

Aufgabe 41: Kombinationen	53
Aufgabe 42: Kombinationen	54
Aufgabe 43: Kombinationen	55

Aufgabe 44: Zählprinzip	56
Aufgabe 45: Kombinationen Zählprinzip . . .	57
Aufgabe 46: Zählprinzip	58

Aufgaben zur Wahrscheinlichkeitstheorie

Aufgabe 47: Laplace-Wahrscheinlichkeit . . .	59
Aufgabe 48: Wahrscheinlichkeiten	60
Aufgabe 49: Wahrscheinlichkeit	61
Aufgabe 50: bedingte Wahrscheinlichkeit . .	62
Aufgabe 51: bedingte Wahrscheinlichkeit . .	63
Aufgabe 52: bedingte Wahrscheinlichkeit . .	64
Aufgabe 53: bedingte Wahrscheinlichkeit . .	65
Aufgabe 54: bedingte Wahrscheinlichkeit . .	66
Aufgabe 55: bedingte Wahrscheinlichkeit . .	67
Aufgabe 56: bedingte Wahrscheinlichkeit . .	68
Aufgabe 57: Verteilungen	69
Aufgabe 58: Verteilungen	70
Aufgabe 59: Verteilungen	71
Aufgabe 60: Verteilungen	72
Aufgabe 61: Verteilungen	73
Aufgabe 62: Verteilungen	74
Aufgabe 63: Verteilungen	75
Aufgabe 64: Verteilungen	76
Aufgabe 65: Verteilungen	77
Aufgabe 66: Verteilungen	78
Aufgabe 67: Verteilungen	79
Aufgabe 68: Verteilungen	80
Aufgabe 69: Verteilungen	81
Aufgabe 70: Verteilungen	82
Aufgabe 71: Verteilungen	83
Aufgabe 72: Verteilungen	84
Aufgabe 73: Erwartungswert Varianz	85
Aufgabe 74: Erwartungswert Varianz	86
Aufgabe 75: Erwartungswert Varianz	87
Aufgabe 76: Erwartungswert Varianz	88
Aufgabe 77: Erwartungswert Varianz	89
Aufgabe 78: Erwartungswert Varianz	90
Aufgabe 79: stetige Zufallsvariablen	91
Aufgabe 80: stetige Zufallsvariablen	92
Aufgabe 81: Erwartungswert Varianz	93
Aufgabe 82: Erwartungswert Varianz	94
Aufgabe 83: Kovarianz	95
Aufgabe 84: Kovarianz	96

Aufgaben zur induktiven Statistik

Aufgabe 85: Punktschätzer	97
Aufgabe 86: Punktschätzer	98
Aufgabe 87: Intervallschätzer	99
Aufgabe 88: Intervallschätzer	100
Aufgabe 89: Intervallschätzer	101
Aufgabe 90: Tests Fehler 1. Art	102
Aufgabe 91: Tests Erwartungswert	103
Aufgabe 92: Tests Erwartungswert	104

Aufgabe 93: Intervallschätzer	105
Aufgabe 94: Intervallschätzer	106
Aufgabe 95: Intervallschätzer Tests	107
Aufgabe 96: Intervallschätzer	108
Aufgabe 97: Konfidenzintervall Anteil	109
Aufgabe 98: Tests Anteil	110
Aufgabe 99: Tests Fehler	111
Aufgabe 100: Tests Kontingenz	112
Aufgabe 101: Tests Kontingenz	113
Aufgabe 102: Tests Kontingenz	114
Aufgabe 103: Tests Kontingenz	115

Aufgaben zur mehrdimensionalen Differentialrechnung

Aufgabe 1

Mehrdimensionale Differentialrechnung: Maximum gesucht (MULTDIFF1.4)

Bestimmen Sie zur Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

$$f(x,y) = x^3 - x^2 \cdot \ln(y^2 + 1) - 3x$$

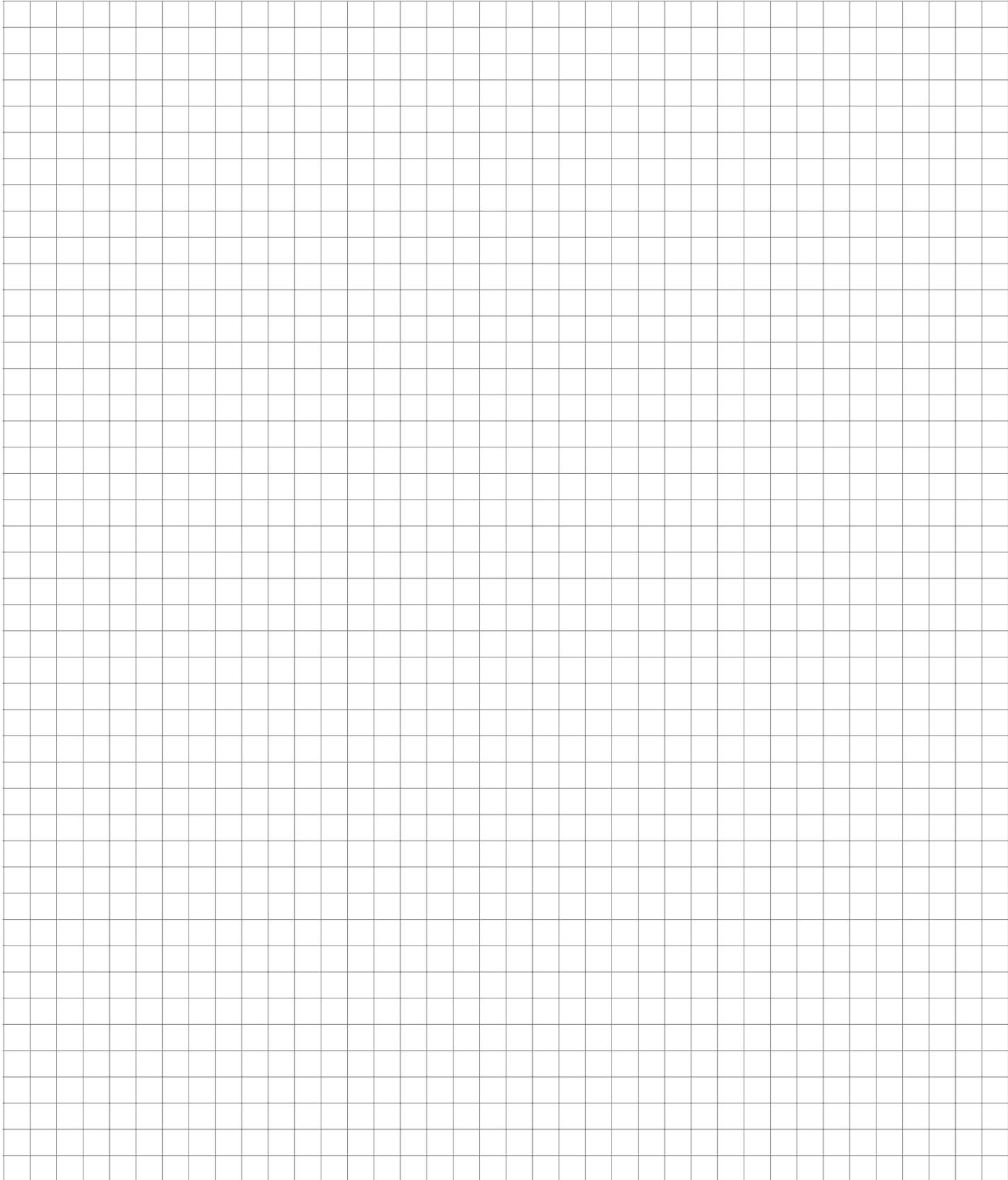
- a) den Funktionswert $f(-1,0)$,
- b) den Gradienten,
- c) die Hesse-Matrix,
- d) alle Nullstellen des Gradienten,
- e) die Definitheitseigenschaft der Hesse-Matrix in jeder Nullstelle des Gradienten,
- f) eine lokale Maximalstelle.



Aufgabe 2

Berechnen Sie alle Extremwerte der Funktion $f : (0; \infty)^2 \rightarrow \mathbb{R}$ mit

$$f(x,y) = xy + 2x - \ln(x^2y)$$



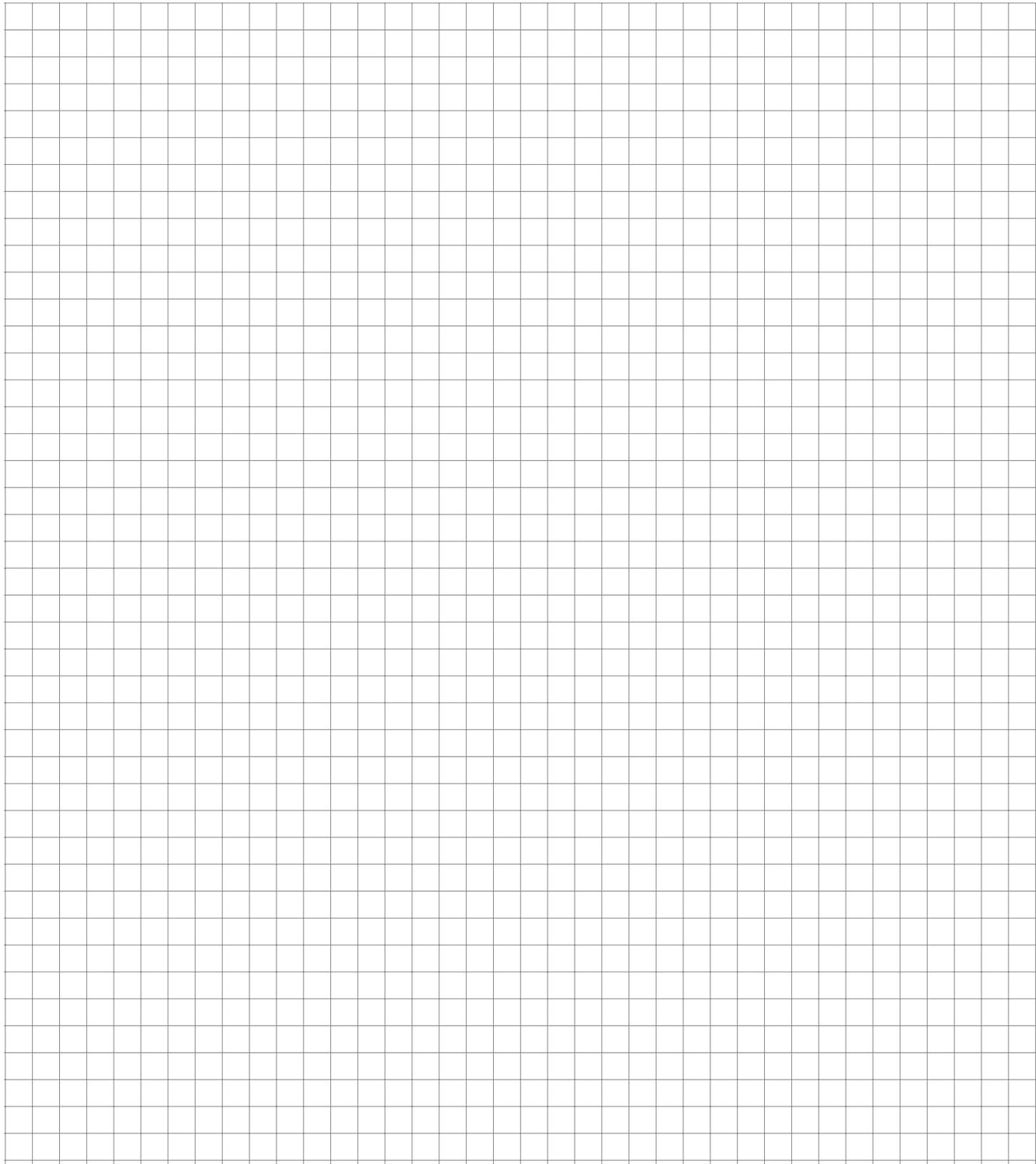
Aufgabe 3 Mehrdimensionale Differentialrechnung: Kompakte Definitionsmengen (MULTDIFF1.8)

Bestimmen Sie die absoluten Maxima und Minima der folgenden Funktionen:

a) $f : [0,5] \times [-3,3] \rightarrow \mathbb{R}$ mit $f(x,y) = x^2 + xy + y^2 - 6x$

b) $g : D \rightarrow \mathbb{R}$ mit $D = \{x \in \mathbb{R}_+^2 : y + 2x \leq 2\}$ und $g(x,y) = x^2 + y^2$

c) $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $h(x,y) = x^4 + \frac{1}{2}y^2 + 2xy$



Aufgabe 4

Gegeben sei folgende Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

$$f(s, t) = (s^2 + t^2) \cdot e^{-t}.$$

- Berechnen Sie den Funktionswert $f(2, 3)$.
- Berechnen Sie den Gradienten ∇f .
- Berechnen Sie alle kritischen Stellen, also die Nullstellen von ∇f .
(es muss keine Hessematrix von f berechnet werden)

Im Folgenden werde die Funktion

$$g(a, b) = a^3 + a^2 \cdot b - b^2 - 4 \cdot b$$

betrachtet. Zu g wurden bereits die kritischen Stellen $(0, -2)$, $(1, -\frac{3}{2})$ sowie $(-4, 6)$ und die folgende Hessematrix H_g berechnet:

$$H_g = \begin{pmatrix} 6a + 2b & 2a \\ 2a & -2 \end{pmatrix}$$

- Untersuchen Sie die Funktion $g(a, b)$ in den angegebenen kritischen Stellen auf lokale Extrema und Sattelpunkte.

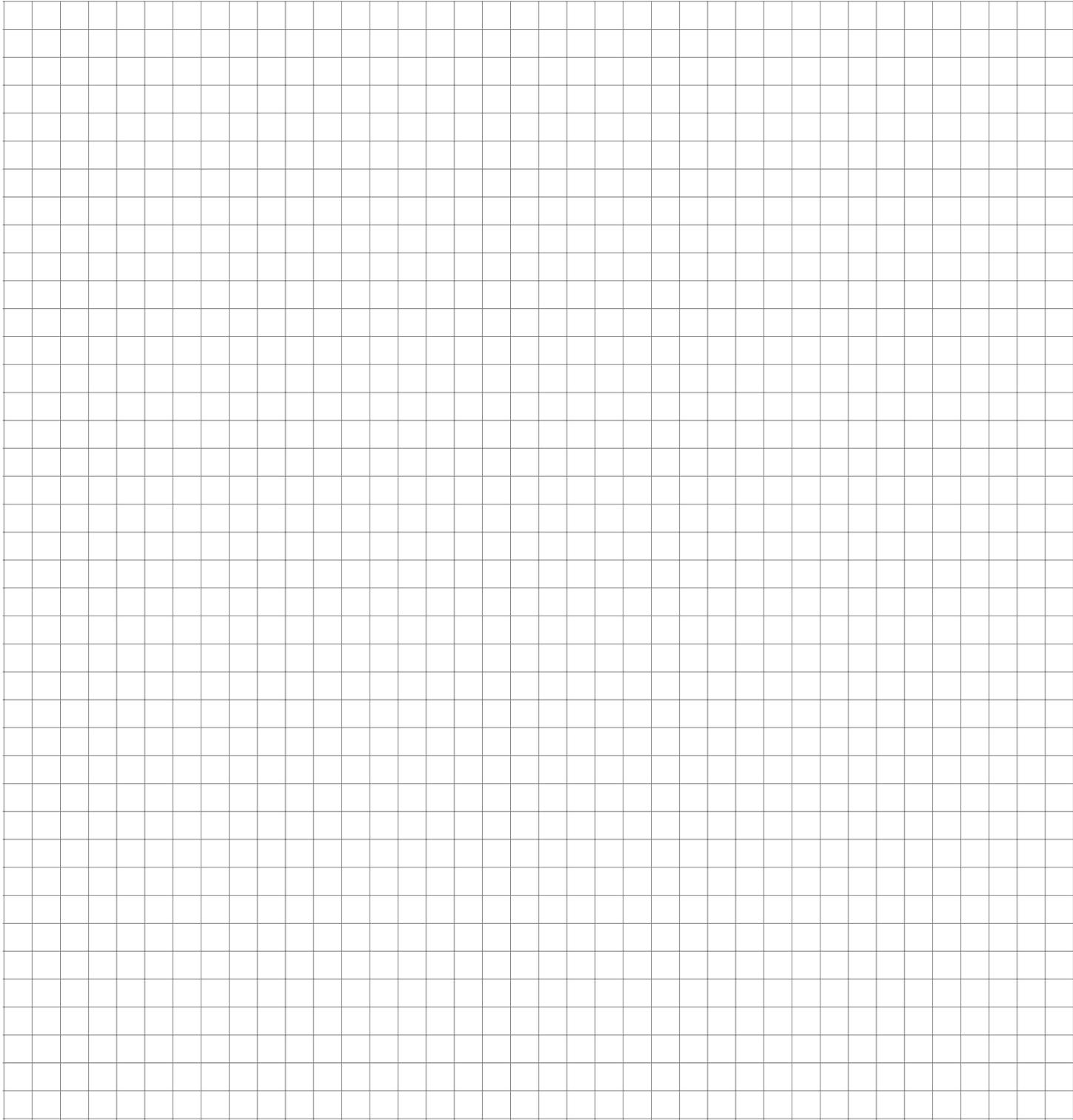


Aufgabe 5

Gegeben ist die reelle Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

$$f(x,y) = x^4 - 2x^2 + (2x^2 + 1)y^2.$$

- Bestimmen Sie den Gradienten von f .
- Bestimmen Sie alle Punkte, an denen der Gradient von f gleich 0 ist.
- Bestimmen Sie die Hessematrix $H_f(x,y)$.
- Bestimmen Sie alle Maximal- und Minimalstellen sowie die zugehörigen Werte von f .



Aufgaben zu R Grundlagen

Aufgabe 6

R: RStudio und erste Versuche (1)

Installation und Kennenlernen von R und RStudio

R

(Sofern Sie über keinen eigenen Rechner verfügen, können Sie im Rechnerraum im W-Gebäude arbeiten; dort sind R und Rstudio installiert)

- a) Installieren Sie R von <http://goo.gl/ALaUXu> (für Windows) bzw. von <http://cran.r-project.org/bin/> für andere Plattformen.

R ist das Programm, das in der Vorlesung verwendet wird, um Daten zu verarbeiten und Ergebnisse als Text oder Grafik auszugeben; es ist in der Rohfassung nicht sehr komfortabel zu bedienen. Deswegen arbeiten wir in diesem Kurs mit RStudio, einer sehr komfortablen und mächtigen integrierten Entwicklungsumgebung.

- b) Installieren Sie RStudio von <http://goo.gl/RX11dj>.
c) Öffnen Sie RStudio. Klicken Sie in den linken unteren Bereich des Fensters („Console“), tippen Sie

```
1 + 2
```

und schließen Sie die Eingabe mit Enter ab.

In der Kommandozeile der Konsole werden alle Anweisungen eingegeben und Textrückmeldungen des Programms ausgegeben; dazu gehören Ergebnisse, aber auch Hinweise, Warnungen und Fehlermeldungen, falls etwas nicht geklappt hat. Die Kommandozeile eignet sich auch prima als Taschenrechner. Kennt man die Bedeutung einer Funktion nicht, kann man ein Fragezeichen voranstellen und bekommt eine Erklärung (rechts im Hilfebereich).

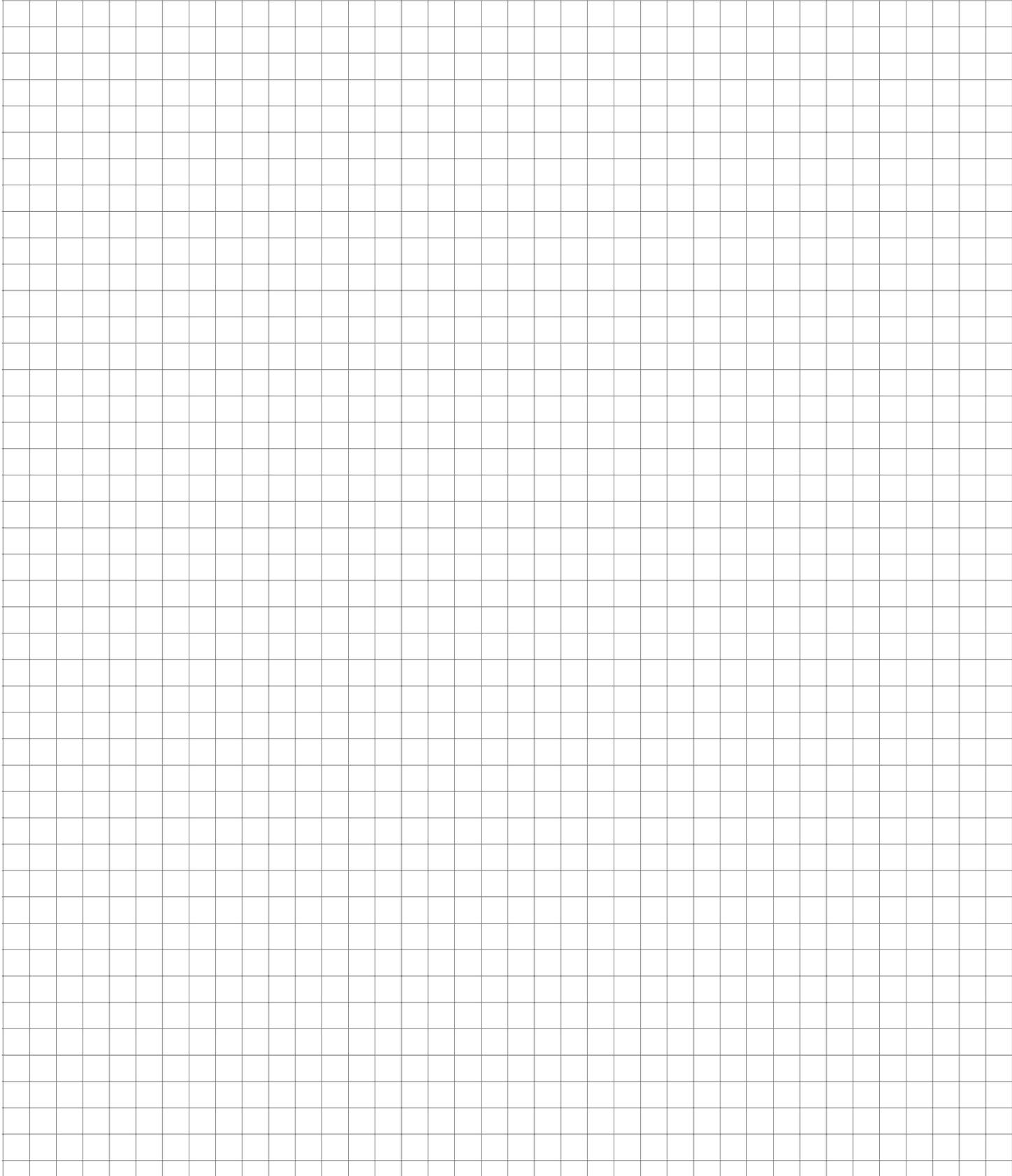
Bei Rechenoperationen gelten die Vorrangregeln der Mathematik (Potenz vor Punkt vor Strich). Der Dezimaltrenner ist ein Punkt (kein Komma). Exponential-, Logarithmus- bzw. Quadratwurzeln berechnet man über Funktionsaufrufe, das Argument steht in runden Klammern. Groß- und Kleinschreibung macht einen Unterschied. Stellt man einer Zeile ein #-Zeichen voran, wird die Zeile von R nicht ausgeführt.

- d) Geben Sie folgende Ausdrücke ein und erklären Sie jeweils das Ergebnis

```
2 + 3 * 4
(2 + 3) * 4
0.2 * 3 - 1.1
0,2 * 3
2^3^2
(2^3)^2
exp(1)
?exp
```

```
log(exp(1))  
sqrt(16)  
16^(1/2)  
Sqrt(16)  
# Das ist ein Kommentar.
```

e) Suchen Sie die Hilfsfunktionen zu den verwendeten Operationen und lesen Sie diese.



Aufgabe 7

R: Zuweisungen und Variablen (2)

Variablen, Zuweisungen und Funktionen

R

Zahlen (und andere Objekte) können in R in Variablen gespeichert werden. Dazu kann der Zuweisungsoperator = oder alternativ <- beispielsweise folgendermaßen verwendet werden:

```
x = 3.5
x2 <- 1.5 # funktioniert genauso mit x2 = 1.5
```

Mit diesen Variablen kann dann weitergerechnet werden. In Variablennamen dürfen Buchstaben, Ziffern (nicht als erstes Zeichen), Punkte und Unterstriche (_) vorkommen. Diese Bezeichner dürfen keine Leerzeichen enthalten. Auch hier ist Groß- und Kleinschreibung zu beachten.

- a) Weisen Sie der Variablen x den Wert 4 zu. Weisen Sie dann der Variablen x.2 den folgenden Wert zu:

$$\sqrt{3x^2 + \ln\left(\frac{1}{e^x}\right) + 5}.$$

Funktionsaufrufe schreibt man in R mit einem Funktionsbezeichner, auf den direkt (keine Leerstelle!) ein Paar runder Klammern folgt. Innerhalb der runden Klammern können ein oder mehrere Argumente oder Parameter der Funktion stehen. Funktionen kann man auch verschachtelt aufrufen.

Die Funktion ls() gibt die in der aktuellen Sitzung definierten Objekte aus. Mit rm(<Var>) kann man eine Variable löschen, wenn man ihren Bezeichner anstatt <Var> in die runden Klammern schreibt.

- b) Überlegen Sie was folgende Zeilen ausgeben und führen Sie diese dann in R aus, um Ihr Ergebnis zu überprüfen.

```
x
x.2
X
x + x.2
x.Produkt = x * x.2
x.Produkt
x.Produkt = x.Produkt * x
ls()
rm(x)
x
ls()
```

Außer Zahlen kann R auch mit Zeichenketten umgehen. Diese können in Objekten gespeichert werden, indem man die Zeichenkette in Anführungsstriche setzt. Zeichenketten, die Zahlen beinhalten werden nicht als Zahlen interpretiert. Man kann mit ihnen also nicht rechnen.

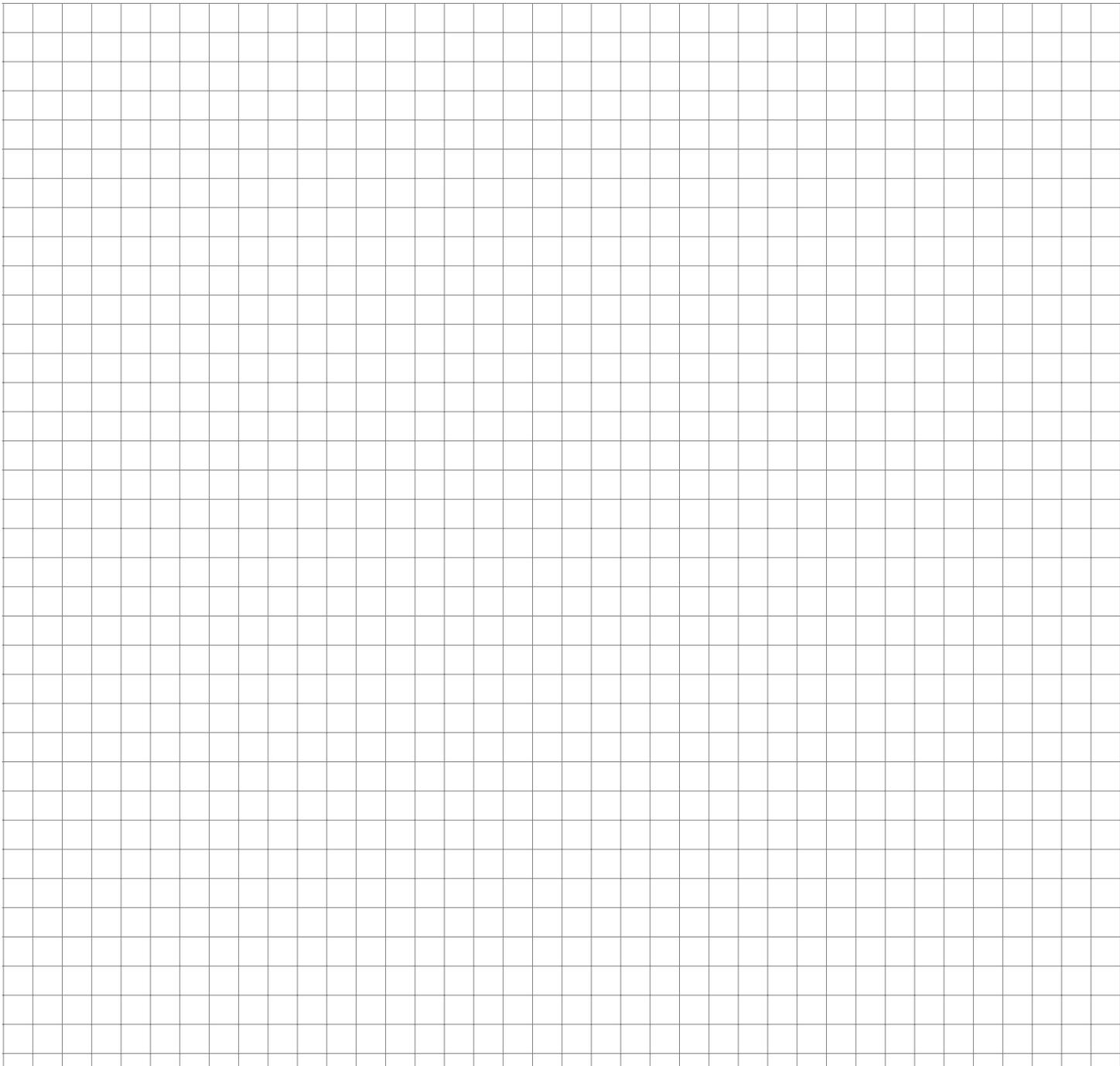
- c) Welche Ausgabe bewirken folgende Zeilen? Überlegen Sie, bevor Sie die Eingabe in R ausprobieren.

```
tubby.1 = "Tinky-Winky"
tubby.2 = "Dipsy"
Zahl = 10
keine.Zahl = "10"
Zahl + 1
keine.Zahl + 1
```

Tricks zur Ein- und Ausgabe:

- ▶ *Ist eine Eingabe in einer Zeile nicht vollständig, kann R das mit einem „+“-Zeichen anzeigen; die Eingabe kann dann vervollständigt werden.*
- ▶ *Sofortige Hilfe bei der Eingabe einer Funktion erhält man, wenn man nach Eingabe der ersten Buchstaben des Funktionsbezeichners die Tabulator-Taste betätigt. Die möglichen Funktionen werden dann zur Auswahl aufgelistet und können dann ausgewählt werden.*
- ▶ *Mit der ↑-Taste auf der Tastatur kann der letzte (oder bei zweimaligem Drücken der vorletzte usw.) Befehl wieder sichtbar gemacht und dann nochmals ausgeführt oder verändert werden.*
- ▶ *Im RStudio-Fenster finden Sie (meistens rechts oben) einen Reiter History. Auch dort werden alle eingegebenen Befehle abgespeichert.*
- ▶ *Im Reiter Environment werden alle Objekte der aktuellen Sitzung aufgelistet.*

d) Probieren Sie die angesprochenen Tricks zur Ein- und Ausgabe aus.



Aufgabe 8

R: Vektoren (3)

Daten: Vektoren

R

Eine Urliste von Daten eines Merkmals wird in R durch einen Vektor repräsentiert. Zur Erzeugung eines Vektors dient die Funktion `c()`. die Einträge der Urliste werden dann zum Beispiel als Argumente von `c()` durch Kommata getrennt angegeben. Als Ausprägungen sind Zahlen oder Zeichenketten möglich. R versucht dann durch die Art der Argumente automatisch zu entscheiden, ob es sich um ein nominales oder ein metrisches Merkmal handelt.

- a) Legen Sie eine Urliste für das Merkmal `x` an, das die Werte 1, 4, 2, 1.5 enthält. Geben Sie `x` aus. Legen Sie ein weiteres Merkmal `Geschlecht` mit den Werten Mann, Frau, Frau, Frau an. Geben Sie auch `Geschlecht` aus. Das dritte Merkmal `z` soll die Werte 1, 2, 1, "1" enthalten. Ist `z` für R nominal oder metrisch? Überprüfen Sie Ihre Entscheidung.

Vektoren aufeinanderfolgender ganzer Zahlen werden mit dem Doppelpunkt-Operator gebildet. `2:5` steht zum Beispiel für den Vektor mit den Zahlen 2, 3, 4, 5. Mit der Funktion `seq()` kann man genauer Vektoren als Folgen von Zahlen erzeugen. `seq(from=2, to=3, by=0.2)` erzeugt zum Beispiel den Vektor (2, 2.2, 2.4, 2.6, 2.8, 3). Mit `rep()` werden Werte oder ganze Vektoren vervielfacht als Vektor ausgegeben. Zum Beispiel ergibt `rep(c(1,2), 3)` den Vektor (1,2,1,2,1,2). Die Hilfe-Seiten (Aufruf über `?seq` bzw. `?rep`) erklären die Details.

- b) Erzeugen Sie folgende Vektoren in R:

```
## [1] 5 6 7 8 9
## [1] 10 9 8 7 6 5 4 3 2 1
## [1] -0,10 -0,08 -0,06 -0,04 -0,02 0,00
## [1] 10000 12500 15000 17500 20000
## [1] -3 -2 -1 0 1 2 -3 -2 -1 0 1 2 -3 -2 -1 0 1
## [18] 2
## [1] 5,0 6,0 7,0 8,0 9,0 10,0 10,1 10,2 10,3 10,4
## [11] 10,5
```

Rechenoperationen können zwischen (numerischen) Vektoren elementweise ausgeführt werden. Hat ein Vektor weniger Elemente als ein anderer, werden die Elemente vom Beginn des kürzeren Vektors einfach solange wiederholt, bis die Länge der beiden Vektoren gleich ist. Die Länge eines Vektors kann mir der Funktion `length()` ausgelesen werden. Die Summe aller Elemente eines Vektors wird mit `sum()` errechnet. Beispielsweise ergibt mit `x=1:5` und `y = c(10.1, 10.5)` die Summe `x+y` den Vektor (11.1, 12.5, 13.1, 14.5, 15.1). Analog funktioniert `-`, `*`, `/`.

- c) Gegeben sind die Vektoren

```
x = 4:2
y = seq(from = 0.1, to = 0.5, by = 0.1)
```

Erklären Sie, was folgende Ausdrücke ergeben und überprüfen Sie Ihr Ergebnis in R:

```
x + y
x * y
x^3 + 1
2 * x - 3 * y
n = length(x + y)
sum(x + y)/n
```

Teile oder einzelne Elemente eines Vektors können mit der Angabe der Indexwerte in eckigen Klammern ausgegeben werden. Auch Bedingungen mit Vergleichsoperatoren (z.B. < für kleiner als oder == für ist gleich) sind möglich in eckigen Klammern. Verknüpfungen zwischen Vergleichen sind mit logisch UND (&) beziehungsweise ODER (|) möglich.

d) Gegeben sind die Vektoren

```
x = seq(from = 0, to = 100, by = 2)
y = 100:1
```

Schreiben Sie die Ergebnisse folgender Ausdrücke auf und überprüfen Sie anschließend Ihr Ergebnis in R:

```
x[3]
y[c(1, 3, 10)]
x[1:4]
x[x > 91]
x[x > 20 & x <= 30]
y[y == 5 | y > 95 | y < 3]
```

Anmerkung: Die Ausgabe von Relationen wie $x < y$ auf Vektoren in R sind Vektoren mit den Ausprägungen TRUE beziehungsweise FALSE. Diese sogenannten logischen Vektoren können zur Indizierung von Vektoren verwendet werden; Elemente mit einem Index von TRUE werden ausgewählt, die mit Wert FALSE werden übergangen.

e) Was ergeben folgende Zeilen in R:

```
x = seq(from = 0.2, to = 2, by = 0.3)
y = -3:3
x < y
x^2 < x
Index = x^2 < x
x[Index]
y[Index]
```



Aufgabe 9

Funktionen einer reellen Variablen

R

In R können Funktionen einer reellen Variable mit dem Aufruf `function(Variable) {Funktionsterm}` definiert werden. Diese Funktionen können beliebigen Bezeichnern zugewiesen werden. Soll beispielsweise die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R} \quad \text{mit} \quad f(x) = \frac{x}{2} + 1$$

definiert werden, kann man schreiben:

```
f = function(x) { x/2 + 1 }
```

Die Funktion kann dann mit Konstanten sowie Variablen Skalaren bzw. Vektoren aufgerufen werden. Der Funktionswert wird zurückgegeben bzw. am Bildschirm ausgegeben:

```
f(2)
## [1] 2
x = 3.2
y = f(x) # Funktionswert von f(3.2) wird in Variable y abgespeichert.
y
## [1] 2,6
x = -2:2 # Mehr als ein x-Wert
x
## [1] -2 -1 0 1 2
f(x)
## [1] 0,0 0,5 1,0 1,5 2,0
```

Wertetabellen können über die Funktion `data.frame()` generiert bzw. ausgegeben werden.

```
Tabelle = data.frame(x, f(x)) # f(x) wird zu f.x., Spaltenname damit Bezeichner
Tabelle # Standardausgabe, durchnummerierte Zeilen
##   x f.x.
## 1 -2 0,0
## 2 -1 0,5
## 3  0 1,0
## 4  1 1,5
## 5  2 2,0
```

Mit `print()` als Wrapper können beispielsweise die Zeilennummern unterdrückt werden.

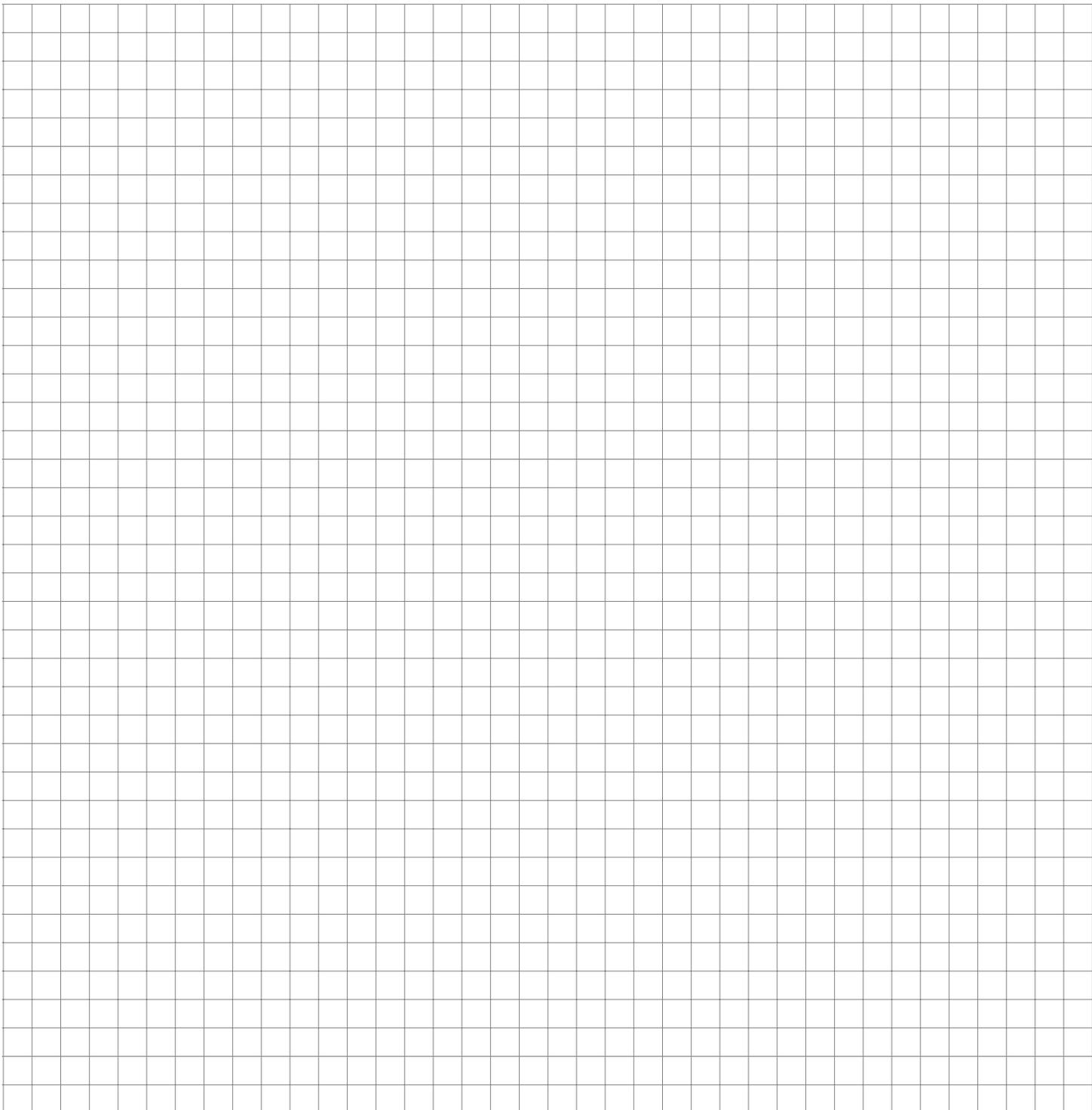
```
print(Tabelle, row.names=FALSE) # Ausgabe ohne Zeilennummern
##   x f.x.
## -2 0,0
## -1 0,5
##  0 1,0
##  1 1,5
##  2 2,0
```

Führen Sie folgende Aufgaben in \mathbb{R} aus:

- a) Definieren Sie die Funktion $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ mit

$$g(t) = \frac{100}{1 + 20 \cdot e^{-4t}}$$

- b) Berechnen Sie die Funktionswerte für Werte von $t \in \{0, 0.1, 1, 10\}$.
c) Generieren Sie einen Vektor t , der die Zahlen von 0 bis 2 in Schrittweiten von 0.2 enthält.
d) Geben Sie eine Wertetabelle der Werte von t und $g(t)$ ohne Angabe der Zeilennummer aus.



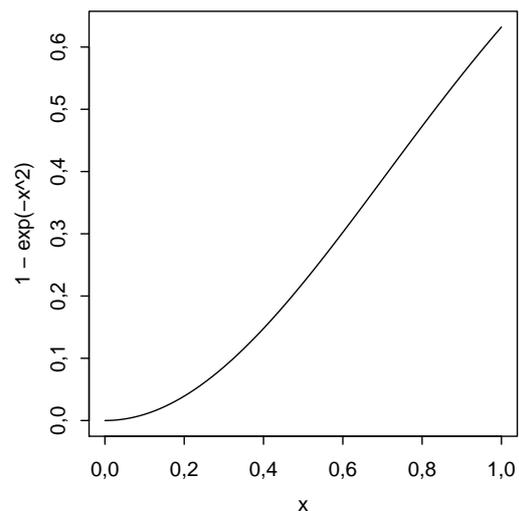
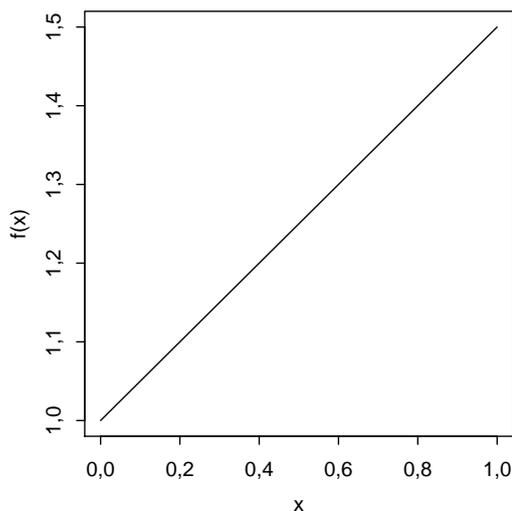
Aufgabe 10

Graphen von Funktionen einer reellen Variablen

R

Mit der Funktion `curve()` kann in R sehr komfortabel der Graph einer Funktion einer reellen Variablen gezeichnet werden. Als Argument wird eine selbstdefinierte Funktion oder ein Text eines Funktionsterms (in Abhängigkeit von x) akzeptiert.

```
f = function(x) { x/2 + 1 }  
curve(f)  
curve(1 - exp(-x^2))
```

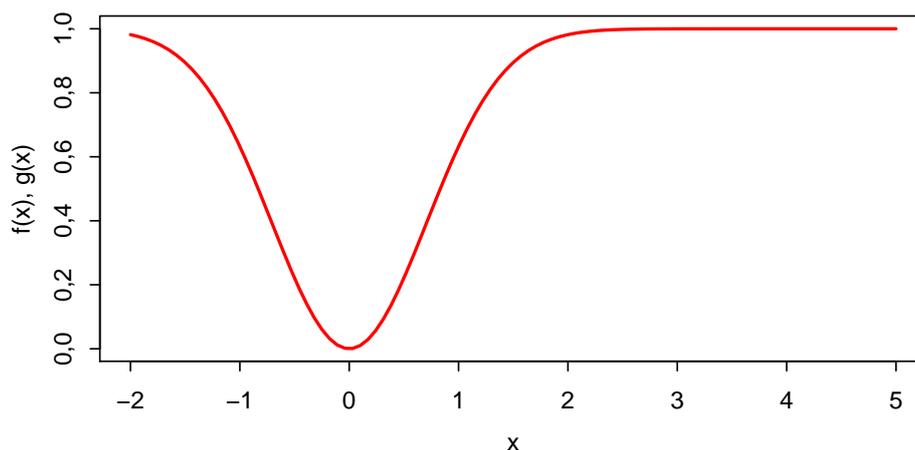


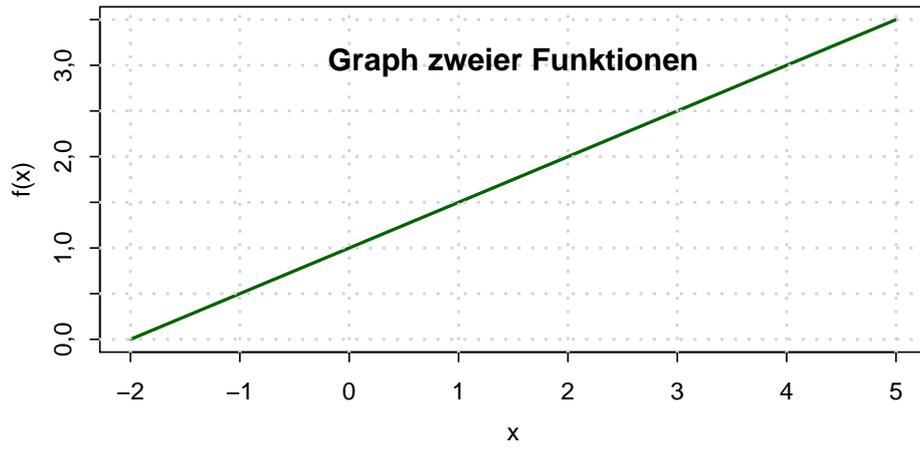
`curve()` hat einige nützliche Parameter:

- ▶ `from`, `to` für den Definitionsbereich,
- ▶ `add` für die Überlagerung mehrerer Kurven,
- ▶ `col` für die Farbe der Kurve,
- ▶ `xlab`, `ylab`, `main` für die Beschriftung der Abszisse, Ordinate, der Überschrift
- ▶ `lwd` für die Strichdicke der Kurve

Mit `grid()` kann man ein Gitter in den Graphen einzeichnen.

```
curve(1 - exp(-x^2), from = -2, to = 5, lwd = 2, col = "red", xlab = "x",  
      ylab = "f(x), g(x)", main = "Graph zweier Funktionen")  
curve(f, add = TRUE, col = "green", lwd = 2)  
grid(lwd = 2)
```



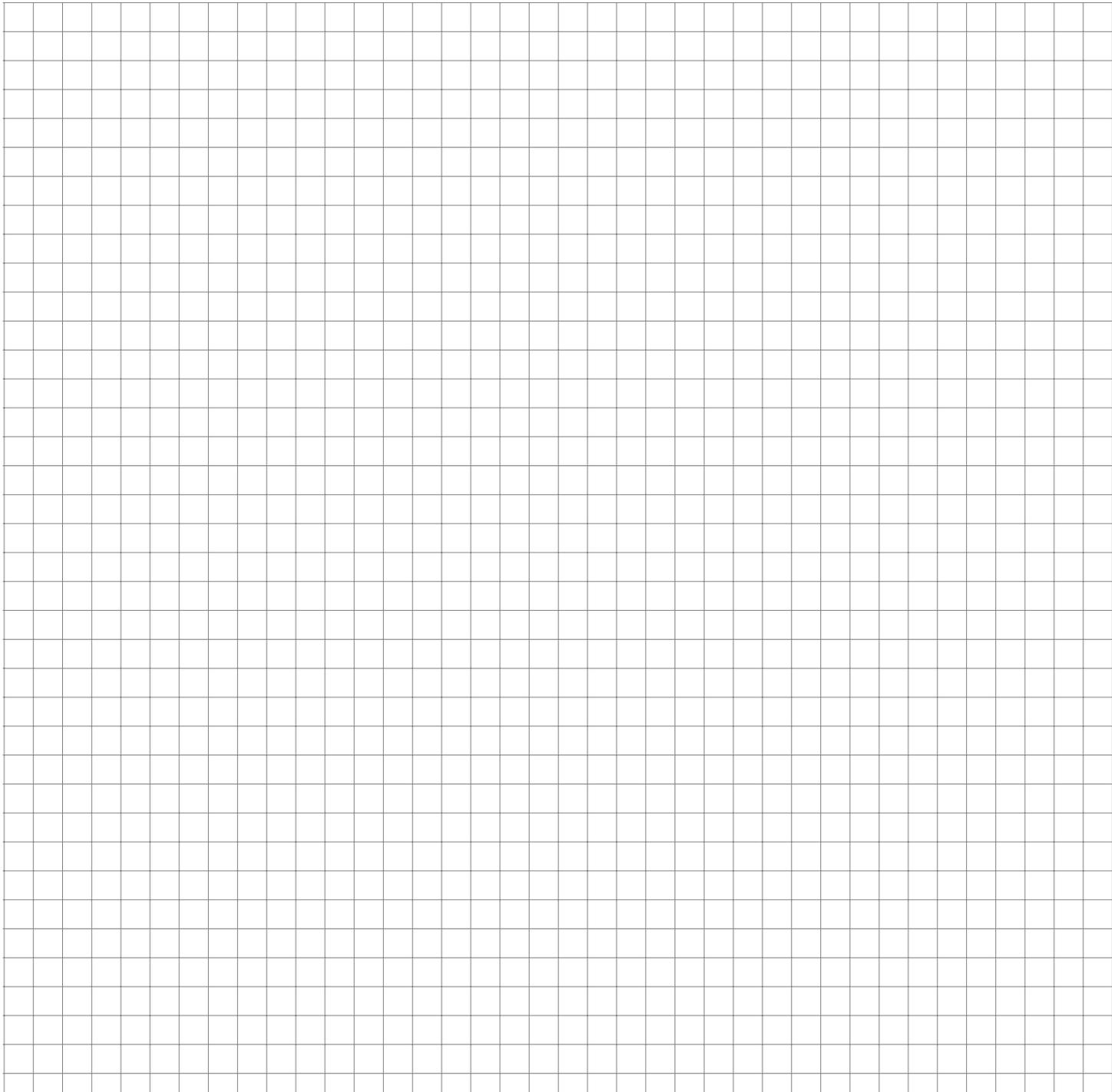


Führen Sie folgende Aufgaben in \mathbb{R} aus:

- a) Definieren Sie die Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$g(t) = \frac{100}{1 + 20 \cdot e^{-4t}}$$

- b) Zeichnen Sie den Graphen von g mit einem sinnvollen Definitionsbereich $D_g \subset \mathbb{R}$, mit Beschriftung der Achsen sowie einer geeigneten Überschrift.
- c) Definieren Sie eine Funktion g_2 , die identisch zu g ist, nur um eine Einheit gemäß $g_2(t - 1) = g(t)$ verschoben.
- d) Zeichnen Sie in den Graphen aus Teilaufgabe a) zusätzlich den Graphen von g_2 ein.
- e) Zeichnen Sie ein Gitter in das Koordinatensystem.



Daten mit mehreren Merkmalen als Data Frames

Urlisten mit mehr als einem Merkmal kann man sinnvoll in einer Tabelle abbilden. Die Zeilen entsprechen den Objekten, die Spalten den Merkmalen. In R heißen solche Datenmatrizen *Data Frames*; diese können mit der Funktion `data.frame()` gebildet werden. Ein Beispiel: Es soll ein Data Frame erstellt werden, der aus 6 Studenten, deren Alter, den Matrikelnummern, einer Klausurnote und der Information, ob diese Klausur bestanden wurde besteht:

```
namen <- c("Arno", "Bert", "Carl", "Doro", "Edda", "Fred")
alter <- c(19, 21, 20, 22, 20, 27)
matrnR <- c(101010, 101007, 200000, 123456, 654321, 111111)
note <- c(1, 5, 2.3, 2.7, 1.3, 4)
bestanden <- (note < 4.3)
Studenten <- data.frame(Name = namen, Alter = alter, Matrikelnummer = matrnR,
  Note = note, Bestanden = bestanden)
Studenten
```

##	Name	Alter	Matrikelnummer	Note	Bestanden
## 1	Arno	19	101010	1,0	TRUE
## 2	Bert	21	101007	5,0	FALSE
## 3	Carl	20	200000	2,3	TRUE
## 4	Doro	22	123456	2,7	TRUE
## 5	Edda	20	654321	1,3	TRUE
## 6	Fred	27	111111	4,0	TRUE

Um auf einzelne Merkmale bzw. Objekte eines Data Frames zuzugreifen, kann der Zeilen- und Spaltenindex durch ein Komma getrennt in eckigen Klammern angegeben werden. Dabei können auch wie bei Vektoren mehrere Zeilen- bzw. Spaltenindexwerte angegeben werden, bzw. über Auswahloperatoren gebildet werden. Um auf ein Merkmal zuzugreifen, kann der Bezeichner des Merkmals mit einem `$`-Zeichen an den Bezeichner des dataframes angehängt werden.

- a) Was gibt R jeweils nach folgenden Zeilen aus? Überprüfen Sie Ihre Antwort in R.

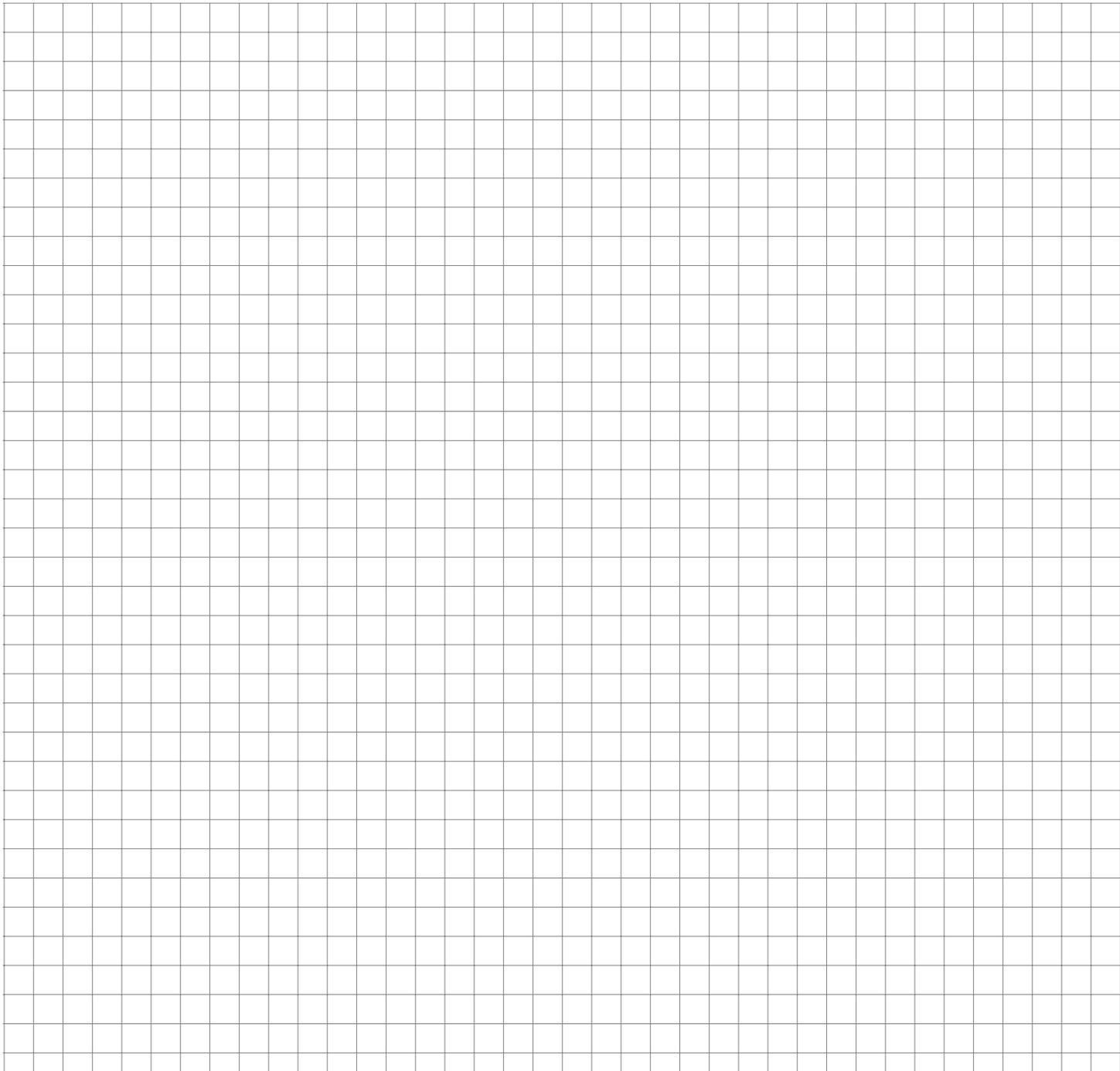
```
Studenten[1, 3]
Studenten[1:3, c(1, 4)]
Studenten[2, ]
Studenten[, 4]
Studenten$Note
Studenten[, 3:5]
Studenten[note < 4, ]
```

b) Erzeugen Sie einen Data Frame der folgenden 7 Kinder gemäß der Datentabelle

Alter	Geschlecht	Taschengeld (in Euro)	besitzt Fahrrad
6	männlich	12	ja
7	männlich	18	ja
6	weiblich	14	nein
7	weiblich	20	ja
8	männlich	26	ja
7	weiblich	20	ja
8	weiblich	20	nein

c) Generieren Sie in R einen Data Frame des Alters und des Taschengeldes aller Kinder mit höchstens 7 Jahren.

d) Berechnen Sie in R die Summe des Taschengeldes aller Fahrradbesitzer.



Datentypen und Umgang mit Data Frames

R kann zwischen metrischen, ordinalen und nominalen Merkmalen unterscheiden. Das Skalenniveau eines Merkmals kann über die Funktion `str()` abgefragt werden. Um explizit ein bestimmtes Skalenniveau abzufragen, können die Funktionen `is.numeric()` für metrisches, `is.ordered()` für ordinales sowie `is.factor()` für nominale Merkmale verwendet werden.

Metrische Merkmale werden typischerweise automatisch von R erkannt, wenn gültige Zahlen eingelesen werden. Falls metrische Ausprägungen als Zeichenketten eingegeben wurden, können sie mittels `as.numeric()` umgewandelt werden.

Nominale Merkmale können mit der Funktion `factor()` angelegt, bzw. mit `as.factor()` umgewandelt werden.

Ordinale Merkmale sind in R auch factors, haben aber zusätzlich eine Rangfolge der Ausprägungen hinterlegt. Zum Anlegen eines solchen Merkmals wird `ordered()`, zum umwandeln `as.ordered()` verwendet.

- a) Lesen Sie die Hilfe-Seiten zu `factor()` und und entscheiden Sie dann, was die folgenden Zeilen in R bewirken. Überprüfen Sie Ihre Lösung mit R.

```

Merkmal1 = c(1, 2, 3, 2, 1, 2, 3)
Merkmal2 = c("2.1", 1, 3, "-3e2", 2000, "-0.2", "-.3")
Merkmal3 = c("gut", "gut", "katastrophal", "mittel", "katastrophal",
             "gut", "mittel")
is.numeric(Merkmal1)
is.numeric(Merkmal2)
Merkmal2 = as.numeric(Merkmal2)
Merkmal2
is.numeric(Merkmal2)
is.factor(Merkmal3)
Merkmal3 = factor(Merkmal3)
is.factor(Merkmal3)
Merkmal3
Merkmal3 = ordered(Merkmal3)
Merkmal3
Merkmal3 = ordered(Merkmal3, levels = c("katastrophal",
                                         "mittel", "gut"))
Merkmal3

```

- b) Gegeben ist der folgende Data Frame „Studenten“:

```

namen <- c("Arno", "Bert", "Carl", "Doro", "Edda", "Franz")
geschlecht <- c(0, 0, 0, 1, 1, 0)
alter <- c(19, 21, 20, 22, 20, 27)
note <- c("sehr gut", "durchgefallen", "gut", "gut", "sehr gut",
         "gut")

Studenten <- data.frame(Name = namen, Alter = alter, Geschlecht = geschlecht,
                       Note = note)

```

Kodieren Sie in R den Datentyp der jeweiligen Merkmale richtig (metrisch, nominal,

Einlesen von Daten aus Textdateien

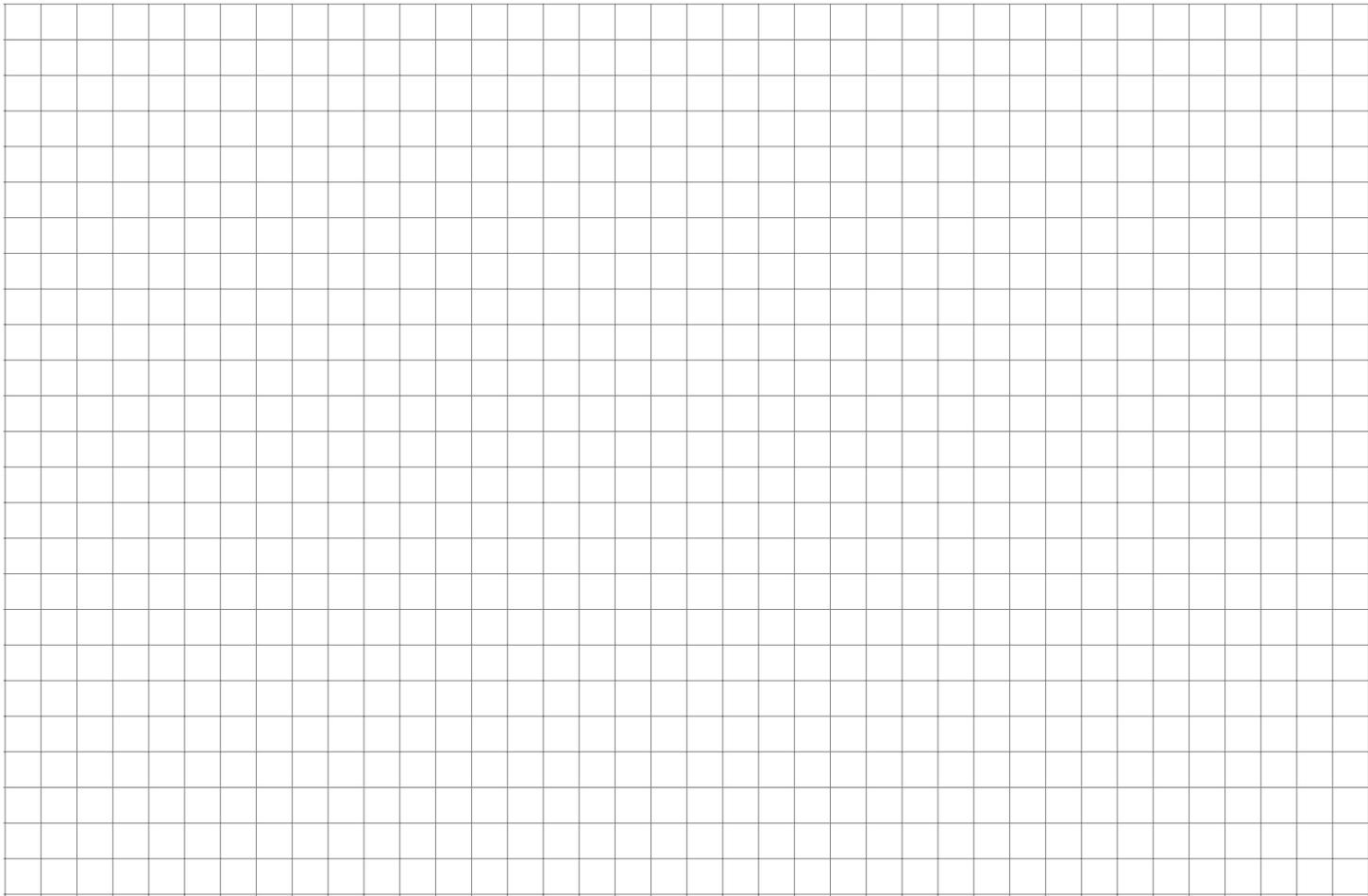


Wenn Daten als Tabelle in einer Textdatei vorliegen können sie leicht in R eingelesen werden; dazu ist es gut, wenn in der Tabelle die Zeilen den Objekten und die Spalten den Merkmalen entsprechen. Über die graphische Benutzeroberfläche in RStudio funktioniert das Importieren einfach über den Knopf „Import Dataset“ im Reiter Environment. Dabei kann eine lokale Datei ausgewählt werden oder eine Datei, die über das Netzwerk via URL erreichbar ist. Klickt man auf Import werden Die Daten als Data Frame unter dem angegebenen Bezeichner zum Beispiel mit der Funktion `read.csv()` eingelesen.

- Importieren Sie mit der graphischen Benutzeroberfläche von RStudio die Umfragedaten aus der Vorlesung von der Adresse <http://goo.gl/Mg6kmj> und speichern Sie den Data Frame unter der Bezeichnung „Umfrage“.
- Importieren Sie die Daten nochmals, diesmal aber indem Sie eine lokale Kopie der csv-Datei auf Ihrer Festplatte anlegen und das Einlesen über die Funktion `read.csv()` bewerkstelligen.

Die Funktion `head()` zeigt die ersten Objekte eines Data Frames an, mit `summary()` bekommt man einen Überblick über die Verteilung der Ausprägungen in den einzelnen Merkmalen.

- Verschaffen Sie sich einen Überblick über die Daten, indem Sie sich die Struktur des Data Frames mit `str()`, die ersten Objekte mit `head()` und die Verteilung der Ausprägungen in den einzelnen Merkmalen mit `summary()` ansehen.



R-Skripten: Führen eines Logbuches der eigenen Analysen

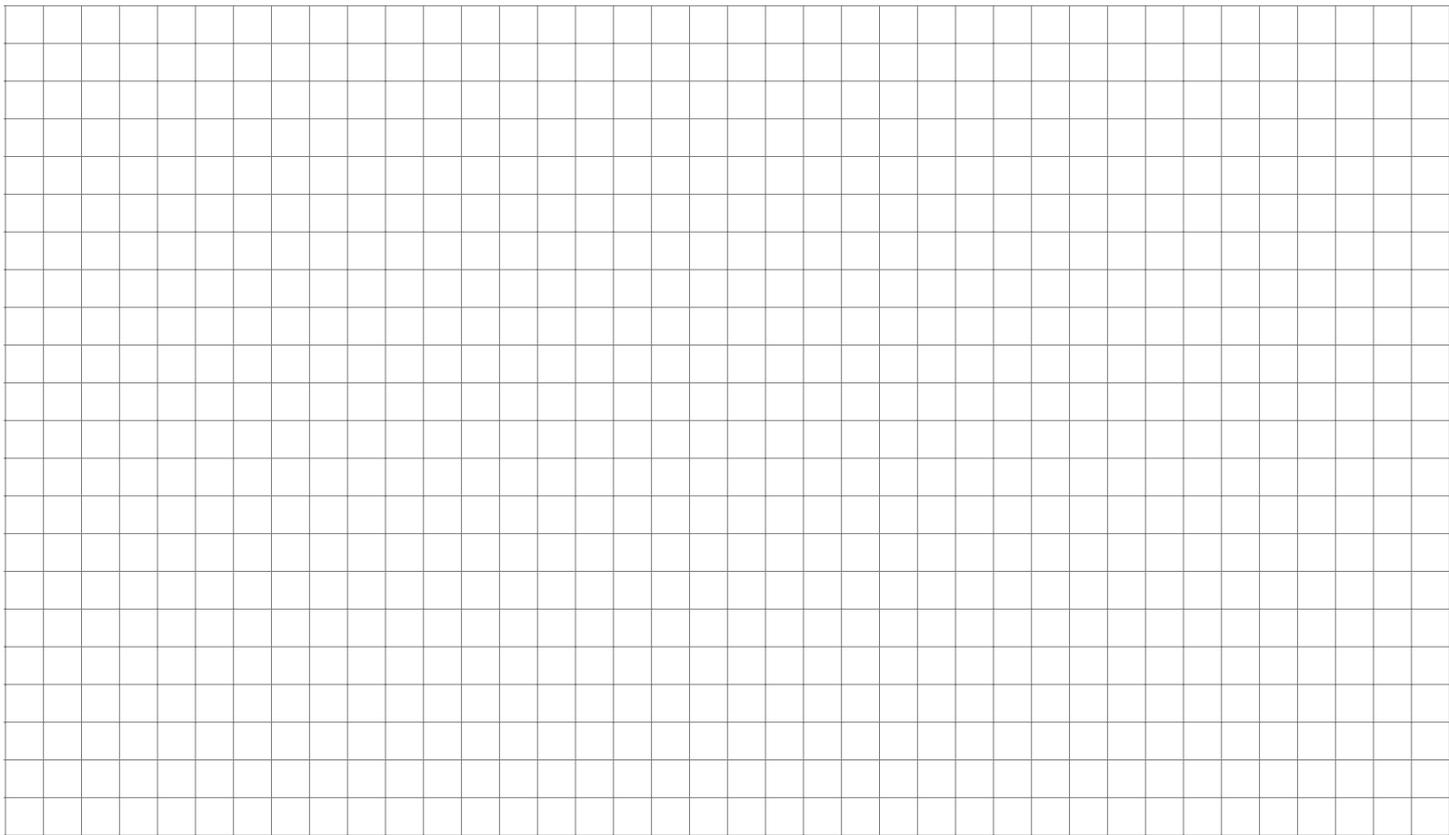
R

In vielen Fällen besteht eine statistische Untersuchung aus mehr als einem Schritt. Meistens werden Daten eingelesen, bereinigt, aufbereitet, verdichtet, graphisch dargestellt usw. Um diesen Ablauf zu dokumentieren kann man eine Textdatei mit der Endung .R, ein sogenanntes R-Skript erstellen und alle Kommandos dort ablegen, mit Kommentaren dokumentieren und für spätere Wiederverwendung abspeichern.

- Legen Sie eine .R-Datei mit dem Bezeichner Statistik-Uebung.R an (in RStudio über File -> New -> R-Script) und schreiben Sie in diese Datei in die ersten Zeilen als Kommentar (#-Zeichen voranstellen) Ihren Namen, das Datum sowie eine Anmerkung, dass diese Datei alle R-Lösungen der Übungsaufgaben enthält.
- Fügen Sie für jede bis hierher bearbeitete Aufgabe nach einem entsprechenden Kommentar den jeweiligen R-Code in diese Datei ein und schreiben Sie zu möglichst vielen Zeilen einen Kommentar, in dem Sie eine Anmerkung schreiben was die Zeile bewirkt.

Um eine Zeile aus einem R-Skript in R auszuführen, kann der Cursor in die entsprechende Zeile platziert werden; durch die Tastenkombination Strg-Enter (auf englischsprachigen Tastaturen Ctrl-Enter) wird die Zeile in die Console kopiert und ausgeführt; danach springt der Cursor in die nächste Zeile des Skripts. Wiederholt man das mehrmals, werden der Reihe nach alle Zeilen ausgeführt (Kommentarzeilen werden übergangen). Möchte man mehr als eine Zeile ausführen, kann man den entsprechenden Teil des Skripts mit der Maus markieren und mit Strg-Enter ausführen.

- Führen Sie die Befehle der ersten R-Aufgabe zunächst zeilenweise aus und beobachten Sie die Ein- und Ausgaben in der Console, danach markieren Sie die komplette Aufgabe und wiederholen die Ausführung.



Aufgabe 15

Häufigkeiten in R und Umgang mit fehlenden Werten

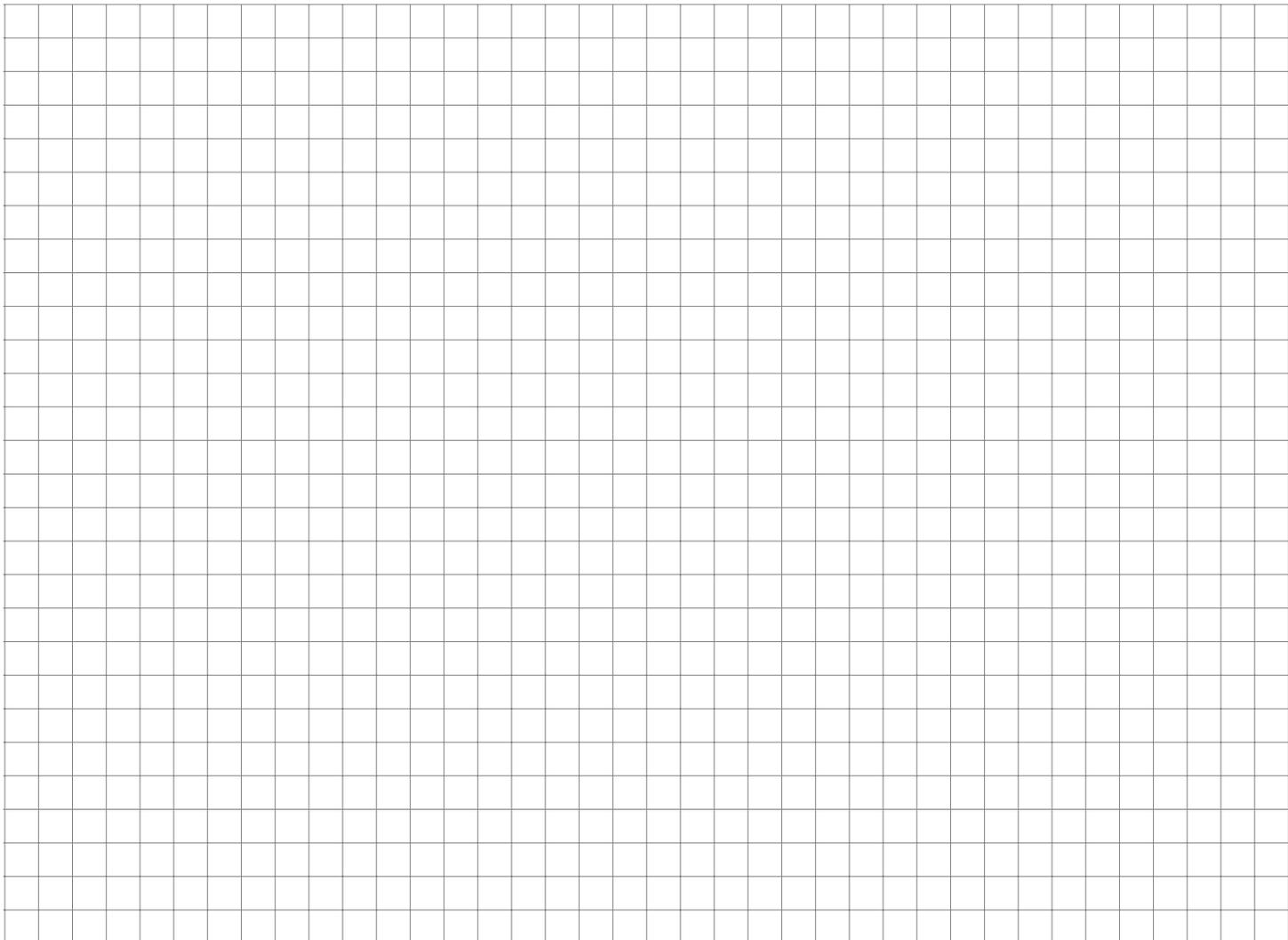


Häufigkeitsauszählungen können in R mit `table()` erstellt werden. `cumsum()` bildet kumulierte Summen. `mean()` berechnet das arithmetische Mittel, `median()` den Median

- Lesen Sie von <http://goo.gl/dZkICg> die Daten der Vorlesungsumfrage ein und bilden Sie jeweils eine Tabelle der absoluten und relativen sowie der absoluten kumulierten und relativen kumulierten Häufigkeiten des Merkmals Alter.
- Wandeln Sie das Merkmal `MatheZufr` (Antwort auf „Waren Sie zufrieden mit Ihrer Leistung in der Matheklausur“) in ein ordinale Merkmal mit sinnvoller Reihenfolge um.
- Berechnen Sie den Median und das arithmetische Mittel aller metrischen Merkmale der eingelesenen Daten.

Ausprägungen fehlender Werte werden in R mit `NA` (Not Available) dargestellt, Objekte mit fehlenden Werten können mittels `na.omit()` gelöscht werden. Die Funktion `sort()` gibt einen metrisch oder ordinal skalierten Vektor in aufsteigender Reihenfolge zurück.

- Für ordinale Merkmale ist der Median zwar definiert, mit der eingebauten Funktion in R erhält man aber eine Fehlermeldung. Lösen Sie das Problem und berechnen Sie den Median aller (vorhandenen) Ausprägungen des Merkmals `MatheZufr`.



Grafiken



Grafiken kann man in R mit `plot()` erzeugen. `plot` versucht anhand der übergebenen Merkmalstypen aus dem Kontext zu entscheiden, welche Art von Grafik erzeugt werden soll:

a) Probieren Sie folgende plots in R aus:

```
# Umfragedaten einlesen
Daten <- read.csv("http://goo.gl/yMeyJp", sep = ";", dec = ",")

# plot mit einem nominalen Merkmal als Argument zeigt
# ein Balkendiagramm
plot(Daten$Farbe)

# Farben können als optionales Argument angegeben werden
# Ausrichtung der Beschriftung über las
plot(Daten$Farbe, col = c("blue", "yellow", "red", "black",
  "grey", "white"), las = 2)

# zwei nominale Merkmale werden als 'spineplot'
# gegenübergestellt
plot(Daten$Farbe, Daten$Geschlecht)

# ein metrisches und ein nominales Merkmal wird als
# Liste von Boxplots gezeichnet
plot(Daten$Geschlecht, Daten$Alter)

# zwei metrische Merkmale verarbeitet plot() als
# Streuplot
plot(Daten$AlterV, Daten$AlterM)

# mehr als zwei metrische Merkmale führen zu einer
# Streuplotmatrix:
plot(Daten[, c("Alter", "AlterM", "AlterV")])
```

`plot()` versteht je nach Kontext viele Parameter zur Anpassung der Grafiken. Die Liste aller Grafikeinstellungen von `plot()` und deren möglichen Werte findet man unter `?par`. Einige wichtige Parameter sind:

- ▶ *main*: Überschrift der Grafik
- ▶ *xlab*, *ylab*: Beschriftung der Abszisse bzw. Ordinate
- ▶ *pch*, *col*, *cex*: ein Wert oder ein Vektor, der das Symbol, das gezeichnet wird, beeinflusst; dabei steht *pch* für die Form des Symbols, *col* für die Farbe und *cex* für die Größe des Symbols.

Mit der Funktion `rgb()` kann man beliebige Farben mischen und deren Transparenz festlegen.

- b) Probieren Sie folgende Ausdrücke in R aus und experimentieren Sie mit Änderungen der Parameter

```
# Speichere zunächst 2 Farben
# rgb(): "r"rot, "g"rün, "b"lau;
# Argumente 1 bis 3: Anteile an rot, grün, blau
# 4. Argument: Transparenz (1 ist undurchsichtig, 0 ist komplett transparent)
Farben = c(rgb(1, 0, 0, 0.2), rgb(0, 0, 1, 0.2))
Symbole = c(20,18) # Kreis und Raute

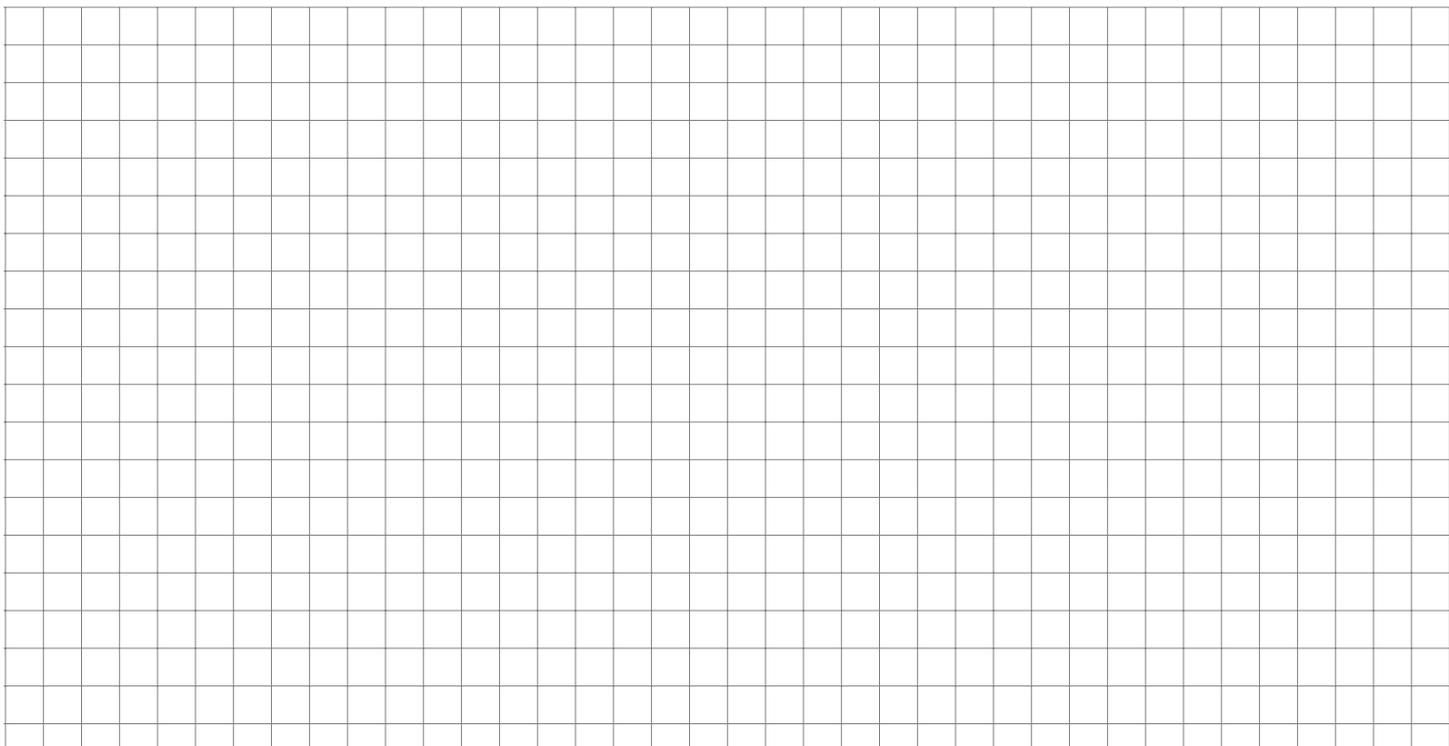
plot(Daten$AnzSchuhe, Daten$AusgSchuhe,
     col=Farben[Daten$Geschlecht], # farbliche Markierung über Geschlecht
     pch=Symbole[Daten$Geschlecht], # unterschiedliche Symbole je nach Geschlecht
     xlab="Wieviele Paar Schuhe besitzen Sie?", # Beschriftung der Abszisse
     ylab="Ausgaben in den letzten 12 Monaten für Schuhe") # Ordinate

plot(Daten[,c("Alter", "AlterV", "AlterM")],
     col=Farben[Daten$Geschlecht], # farbliche Markierung über Geschlecht
     pch=Symbole[Daten$Geschlecht]) # unterschiedliche Symbole je nach Geschlecht

# Alternative zu spineplot: Mosaikplot
mosaicplot(Daten$Farbe ~ Daten$Geschlecht,
           shade=TRUE, # farbige Markierung
           las=2, # Ausrichtung der Achsenbeschriftung
           xlab="Geschlecht",
           ylab="Wunschfarbe", main="")

# Boxplot, horizontale Ausrichtung, eingefärbt
plot(Daten$Geschlecht, Daten$Alter, col = rgb(0, 0, 1, 0.2),
     horizontal = TRUE, las = 2, lwd = 1.5, pch = 20)
```

- c) Generieren Sie einen Streuplot, in dem die Ausgaben für Schuhe und die Anzahl der Schuhe gegenübergestellt werden. Trennen Sie die Farben nach dem Merkmal „Farbe“ und färben Sie mit der jeweils in der Umfrage genannten Lieblingsfarbe ein, jeweils mit 30% Transparenz. Zeichnen Sie für die Frauen ausgefüllte Dreiecke und für die Männer ausgefüllte Quadrate.



Aufgabe 17

R: Emp. Vtlgs.f. Quantil Boxplot (11_Quantile_Boxplot)

Empirische Verteilungsfunktion, Quantile, Boxplot

R

mit der Funktion $ecdf()$ wird eine empirische Verteilungsfunktion erstellt. Wird das Ergebnis von $ecdf()$ in die Funktion $plot()$ eingesetzt erhält man eine grafische Darstellung dieser Verteilungsfunktion. Die Funktion $quantile()$ liefert empirische Quantile einer Urliste. Dabei gibt es verschiedene Varianten, wie diese Quantile berechnet werden können; mit dem Parameter $type$ kann man diese unterschiedlichen Definitionen ansprechen. In der Vorlesung wurde $type = 2$ definiert.

Gegeben ist die folgende Urliste:

```
x = c(1, 2, 1, 2, 10, 10, 20, 1, 2, 1)
```

Lösen Sie folgende Aufgaben jeweils zuerst auf Papier und dann mit R:

- Geben Sie die empirische Verteilungsfunktion $F(x)$ an und zeichnen Sie den Graph von F .
- Berechnen Sie die folgenden empirischen Quantile \tilde{x}_{p_i} mit

```
p = c(0.2, 0.25, 0.4, 0.5, 0.6, 0.75, 0.99)
```

- Zeichnen Sie (von Hand) einen Boxplot zum Merkmal X und überprüfen Sie mit der Assgabe von R das Ergebnis.



Aufgaben zur deskriptiven Statistik

Aufgabe 18

Deskriptiv: Häufigkeit1b (1)

Ein Einzelhändler registriert für einen Exklusivartikel im Verlauf von 30 Verkaufstagen folgende Verkaufszahlen:

Tag	1	2	3	4	5	6	7	8	9	10
Anzahl	5	2	3	0	0	1	3	6	0	2
Tag	11	12	13	14	15	16	17	18	19	20
Anzahl	1	0	1	0	2	3	5	1	0	0
Tag	21	22	23	24	25	26	27	28	29	30
Anzahl	3	5	3	1	0	0	0	6	3	1

- Berechnen Sie die absoluten und relativen Häufigkeiten der Ausprägungen sowie die absolute kumulierte Häufigkeit für $x = 4$.
- Erstellen Sie das zugehörige Balkendiagramm und das Kreisdiagramm mithilfe der absoluten Häufigkeiten.

R

R



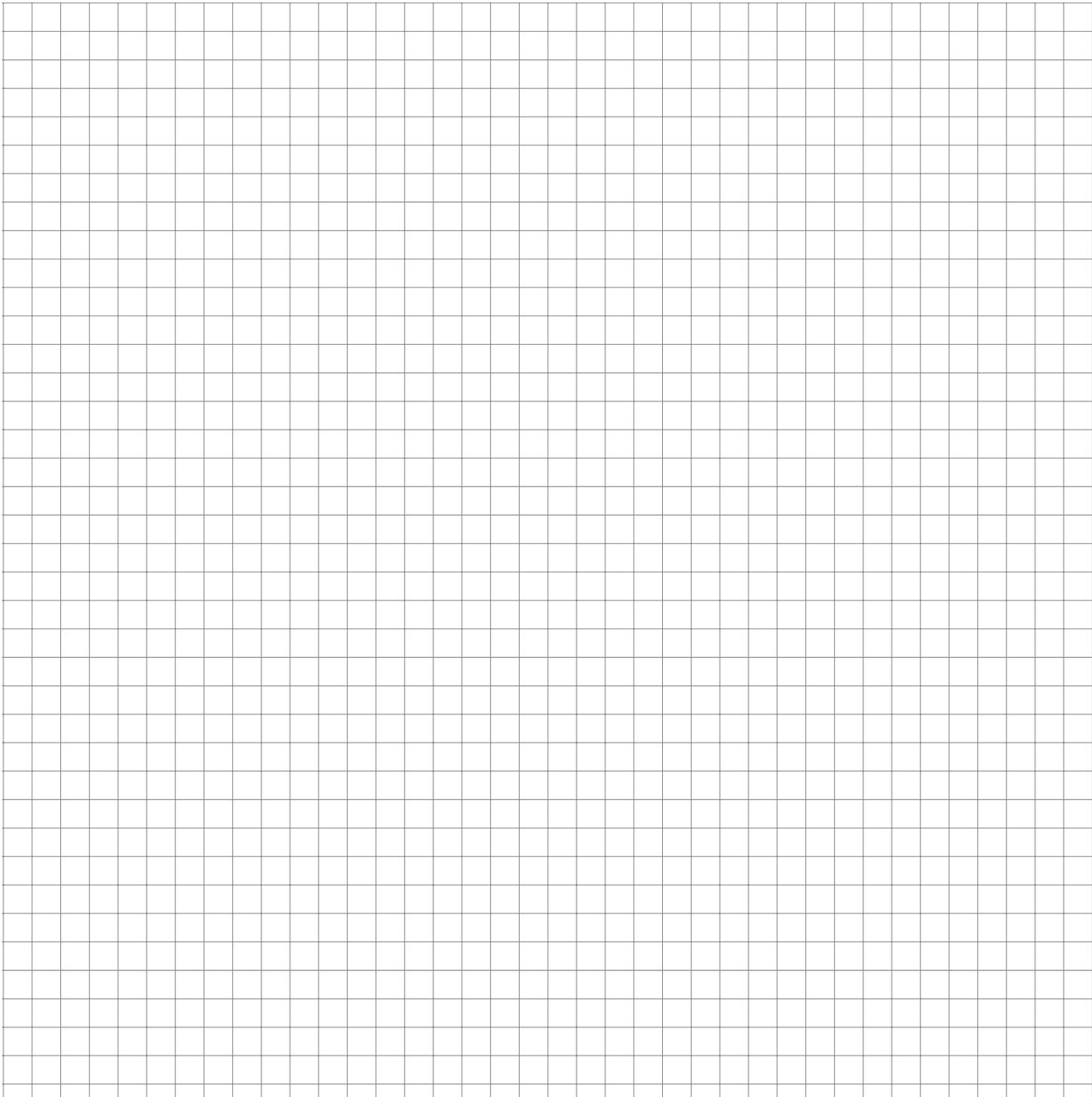
Aufgabe 19

Deskriptiv: Lageparameter (10)

Das Ergebnis der Untersuchung eines kardinalskalierten Merkmals X sei in folgender Tabelle wiedergegeben:

Ausprägung	1	2	3	4	7
Anzahl	4	4	6	4	2

- Bestimmen Sie das arithmetische Mittel, den Modus und den Median.
- Berechnen Sie die mittlere quadratische Abweichung.
- Obige Daten werden nun mittels der Intervalle $[0; 3)$, $[3; 4)$ und $[4; 7]$ klassiert. Bestimmen Sie die Rechteckhöhen des Histogramms.



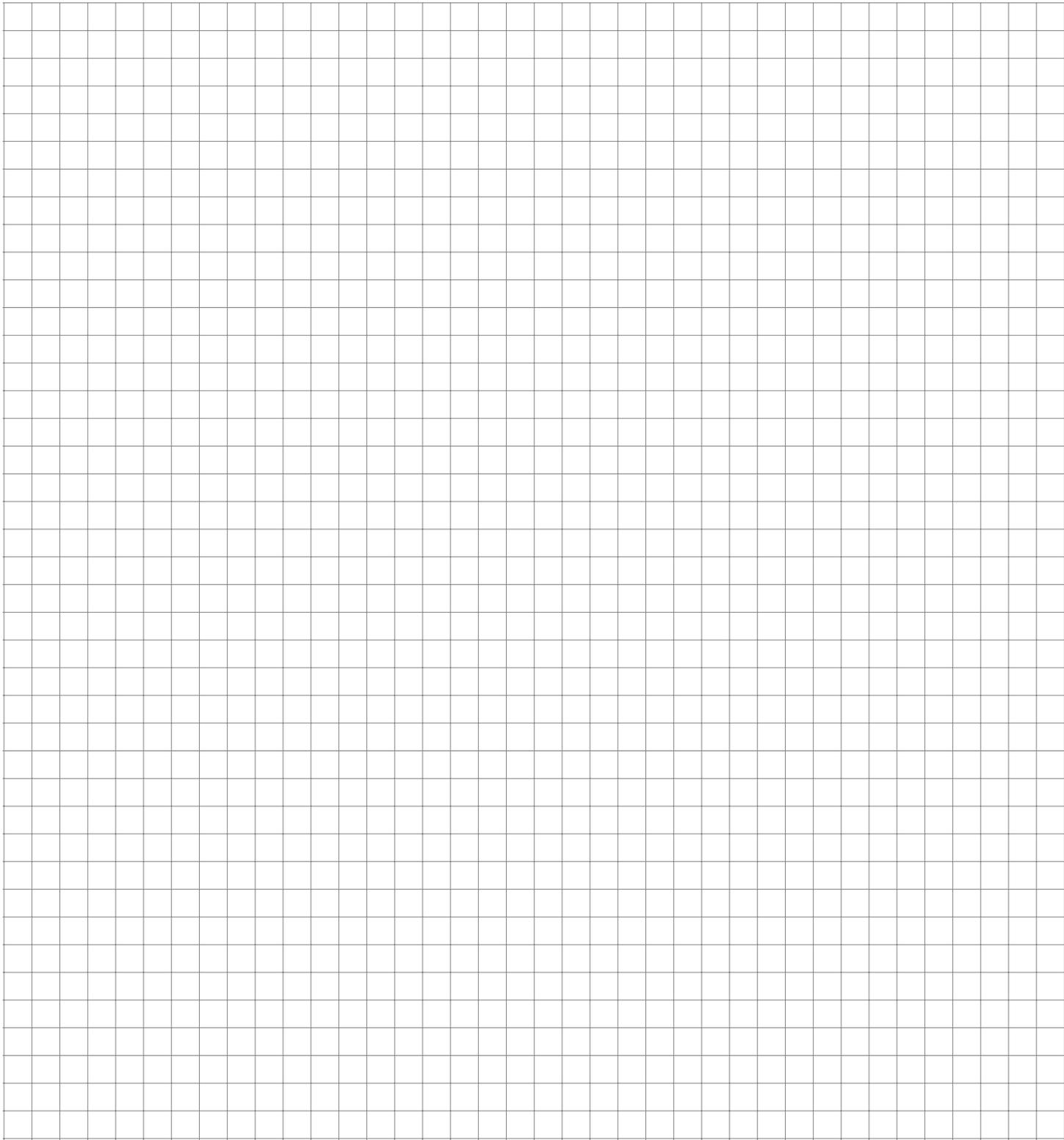
Aufgabe 20

Deskriptiv: Lageparameter (2)

Eine Umfrage über den Bierkonsum Münchner Bürger ergibt bei 10 Personen folgende Zahlenreihe (Liter pro Woche):

3 10 1 2 3 0 2 1 0 3

Berechnen Sie den Modalwert, den Median, das arithmetische Mittel, die Spannweite, die mittlere quadratische Abweichung, die Standardabweichung und den Variationskoeffizienten.



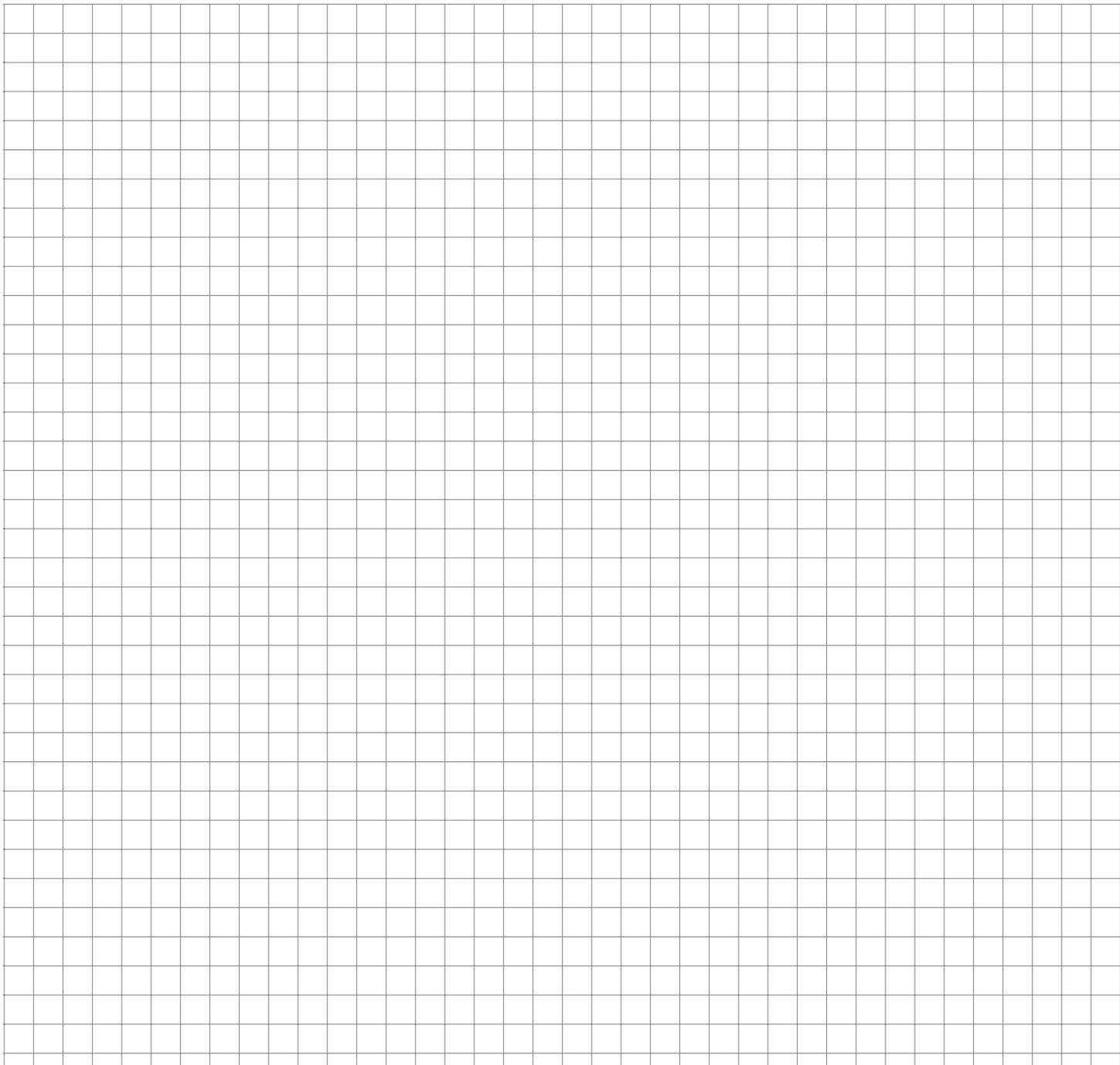
Aufgabe 21

Deskriptiv: Lage Streuung (Koerpergroesse_FussballerWM2010)

In der deutschen Fußballnationalmannschaft wird die Körpergröße (in cm) einiger (aktiver wie ehemaliger) Spieler erhoben. Es ergeben sich folgende Daten:

Körpergröße	[150, 160)	[160, 170)	[170, 175)	[175, 180)	[180, 190)	[190, 210)
Anzahl	5	35	47	38	101	27

- Zeichnen Sie ein Histogramm der Daten.
- Schätzen Sie das arithmetische Mittel und den Median der Körpergröße.
- Schätzen Sie die Standardabweichung der Körpergröße, indem Sie davon ausgehen, dass alle Nennungen pro Klasse jeweils in der Klassenmitte liegen. Ist dieser Wert größer oder kleiner als die tatsächliche Standardabweichung in der Grundgesamtheit?



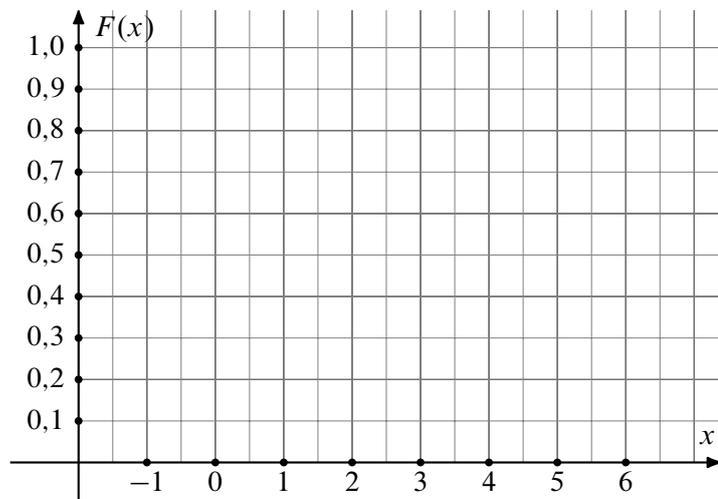
Aufgabe 22

Deskriptiv: Lage Streuung Vtgl.fkt. (Verteilung_2014_01)

10 Personen werden befragt, wieviel Sie in den Weihnachtsferien zugenommen haben. Für die empirische Verteilungsfunktion des abgefragten Merkmals $X \hat{=}$ „Gewichtszunahme vom 23. Dezember bis zum darauf folgenden 7. Januar“ ergibt sich:

$$F(x) = \begin{cases} 0,0 & \text{für } x < -1,0 \\ 0,1 & \text{für } -1,0 \leq x < 1,0 \\ 0,3 & \text{für } 1,0 \leq x < 1,5 \\ 0,4 & \text{für } 1,5 \leq x < 2,0 \\ 0,9 & \text{für } 2,0 \leq x < 6,0 \\ 1,0 & \text{für } x \geq 6,0 \end{cases}$$

- Zeichnen Sie $F(x)$ in nebenstehendes Koordinatensystem ein.
- Schreiben Sie die ursprünglichen Daten als Urliste von X auf.
- Berechnen Sie das arithmetische Mittel von X .
- Bestimmen Sie den Median von X .
- Berechnen Sie die mittlere quadratische Abweichung sowie die Standardabweichung von X .



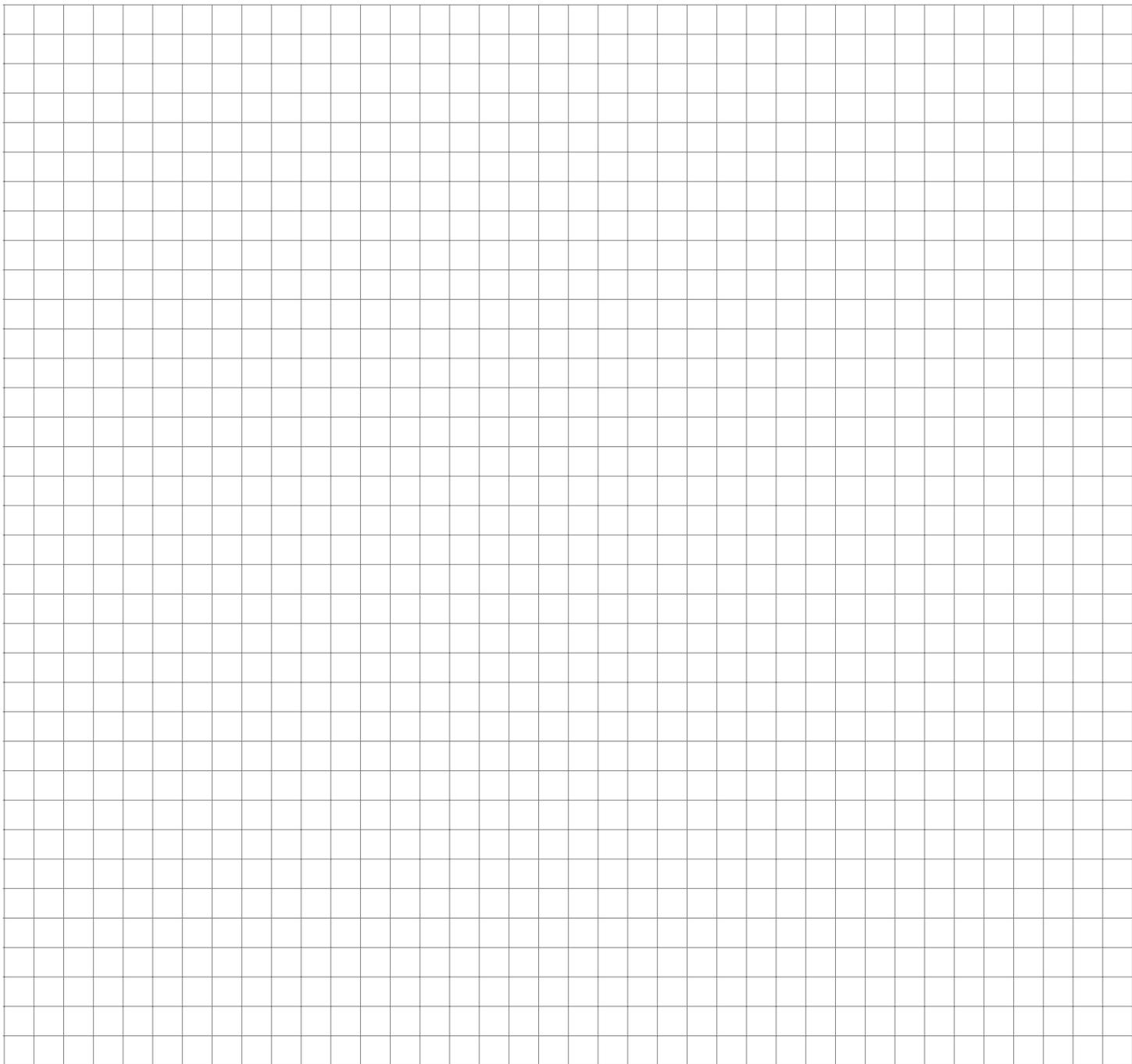
Aufgabe 23

Deskriptiv: Lageparameter Konzentration (3)

Ein bestimmtes Gut wird von genau 7 Firmen produziert. Folgende Tabelle gibt an, wie viele tausend Stück jede Firma herstellt:

Firma:	A	B	C	D	E	F	G
prod. Stückzahl:	3	2	3	5	6	15	6 (tausend Stück)

- Bestimmen Sie den Medianwert der produzierten Stückzahlen.
- Skizzieren Sie für x -Werte aus dem Intervall $[0;20]$ den Verlauf der Funktion $F(x) =$ Anteil der Firmen, die höchstens $1000 \cdot x$ Stück produzieren.
- Errechnen Sie die Knickpunkte der zugehörigen Lorenzkurve.
- Errechnen Sie den normierten Gini-Koeffizienten.
- Bestimmen Sie den Konzentrationskoeffizienten CR_2 .

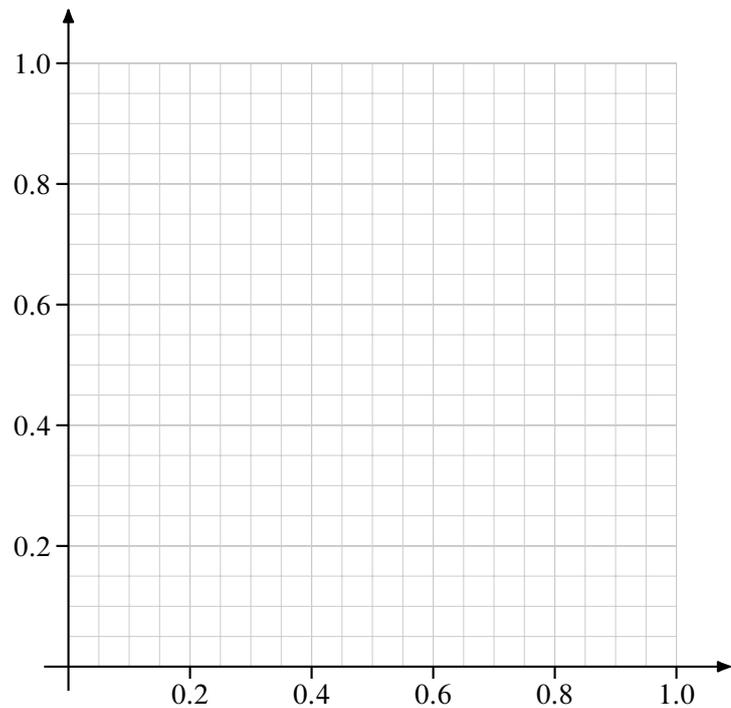


Aufgabe 24

Bei einer Umfrage unter den Kindern einer Schulklasse nach der Anzahl ihrer Geschwister ergaben sich die folgenden Daten:

Anzahl Geschwister (a_i)	0	1	2	3	5	6
Häufigkeit (h_i)	6	4	5	2	2	1

- Erstellen Sie eine Tabelle, die die Koordinaten der Lorenzkurve enthält. Zeichnen Sie die Lorenzkurve in nebenstehendes Diagramm.
- Bestimmen Sie den Mittelwert, den Median, die Standardabweichung, $f(4)$ und $F(4)$ der Verteilung.
- Bestimmen Sie das 25 %- sowie das 75 %-Quantil der Daten.
- Zeichnen Sie einen Boxplot der Daten.



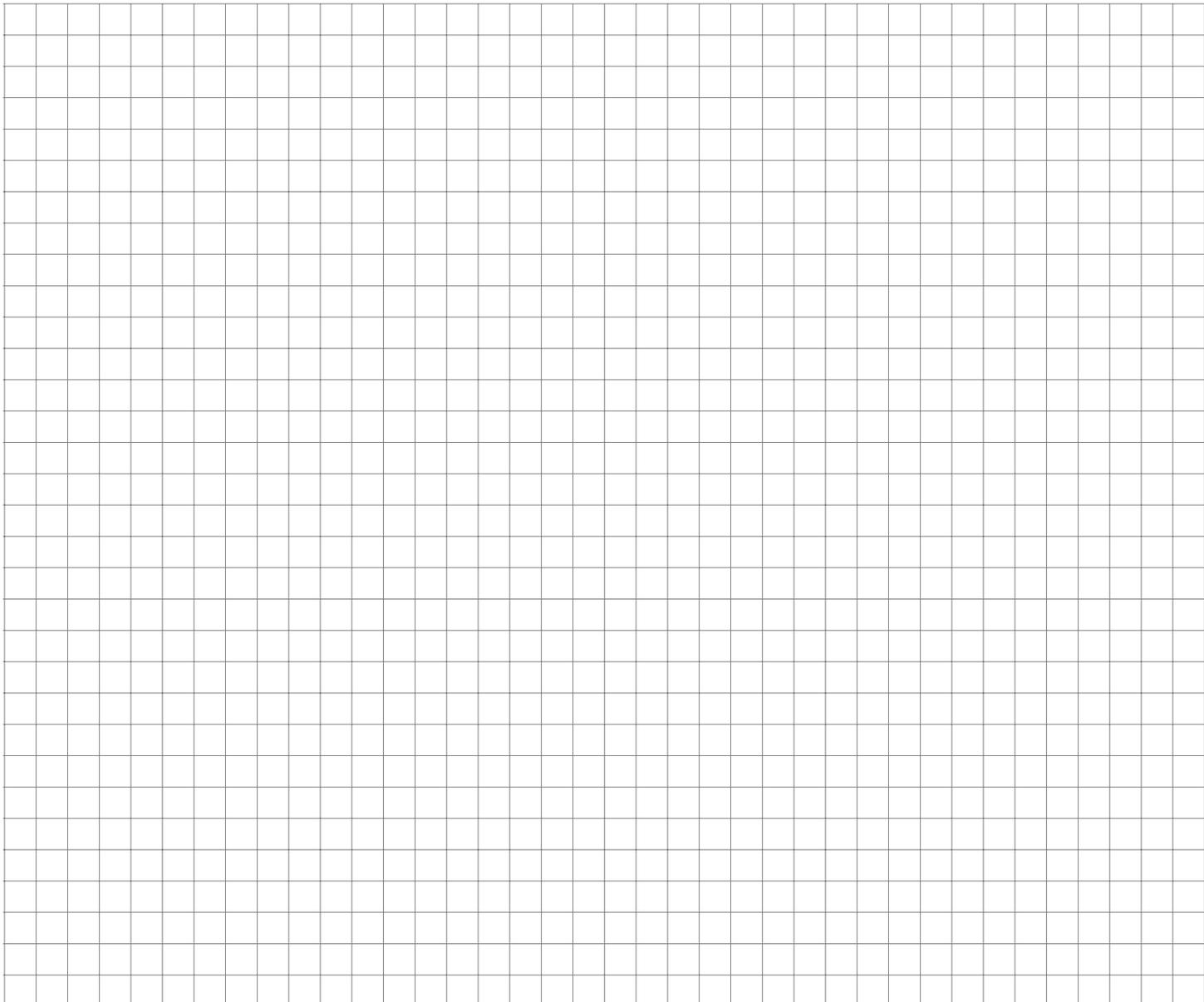
Aufgabe 25

Deskriptiv: Konzentration (2014_10_19)

Die Firma CelebWedCake liefert zu einem Festpreis von 200.000 € eine exklusive Premium-Hochzeitstorte an Prominente. In den letzten 5 Jahren wurden insgesamt 20 von diesen Torten verkauft. Pro Kunde ist die Anzahl der verkauften Torten in dieser Zeitspanne mittels der verschiedenen Ausprägungen a_i und den zugehörigen absoluten Häufigkeiten h_i erfasst:

i	1	2	3	4
a_i	1	2	4	5
h_i	5	3	1	h_4

- Bestimmen Sie h_4 .
- Zeichnen Sie die Lorenzkurve,
- berechnen Sie den normierten Gini-Koeffizienten sowie
- den Herfindahl- und
- den Exponentialindex der Anzahl der verkauften Torten pro Kunde.



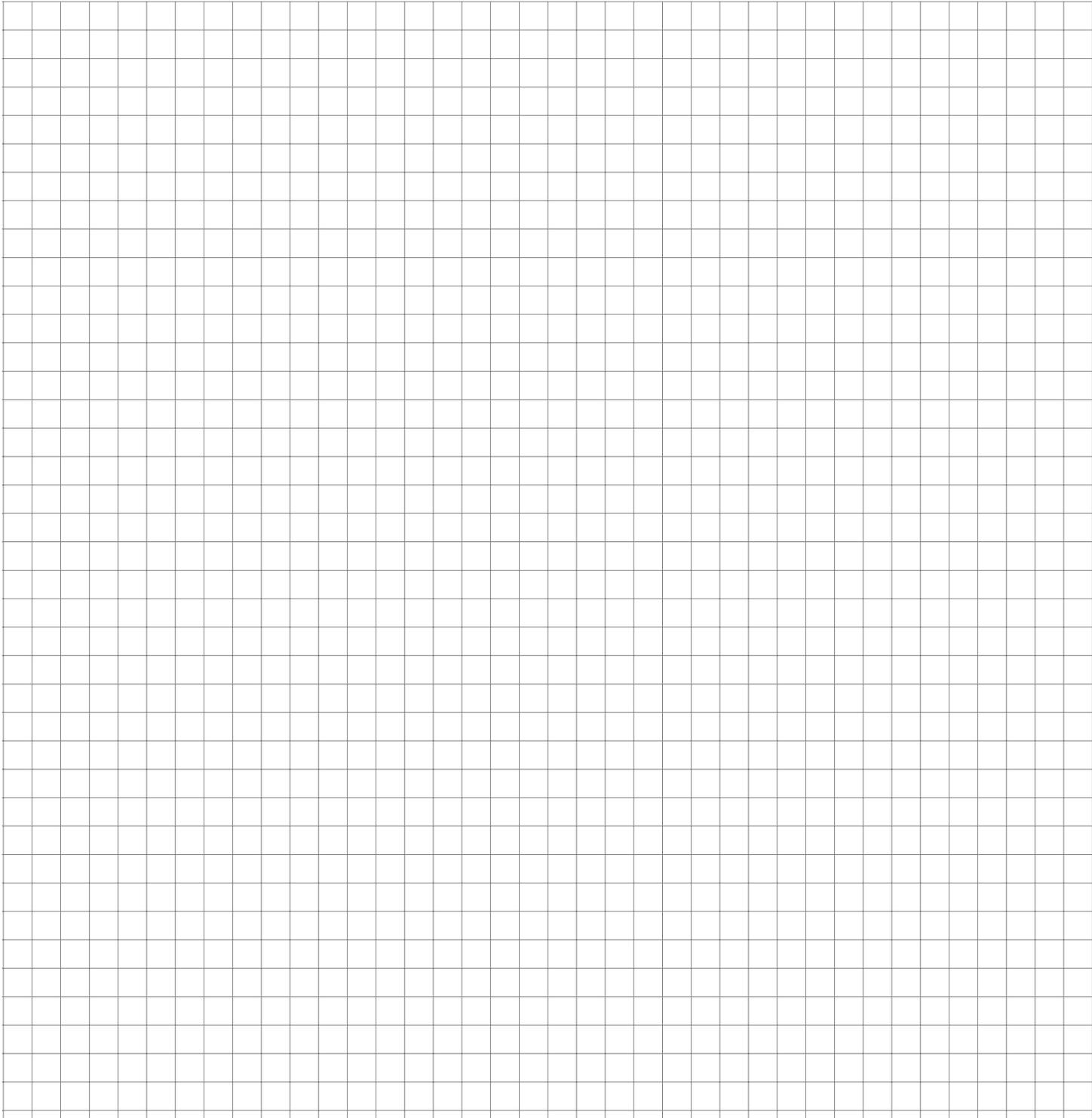
Aufgabe 26

Deskriptiv: Konzentration (18)

Die folgende Tabelle gibt jeweils den jährlichen Umsatz der weltweit 10 umsatzstärksten Softwareunternehmen an:

Nr. des Unternehmens:	1	2	3	4	5	6	7	8	9	10
Umsatz [in Mrd. US-Dollar]:	4	49	4	22	5	18	12	6	7	6

- Zeichnen Sie die Lorenzkurve des Umsatzes.
- Berechnen Sie den Gini-Koeffizienten und den normierten Gini-Koeffizienten.
- Berechnen Sie den Herfindahl- sowie den Exponentialindex.



Aufgabe 27

Deskriptiv: Lage Konzentration (Konzentration_Prosecco)

Pia lädt 11 Freundinnen zu einem Damenabend ein. Es gibt 8 Flaschen Prosecco. Pro Flasche kann Sie 5 Gläser ausschenken.

- ▶ 3 Freundinnen müssen fahren und trinken nichts vom Prosecco,
 - ▶ 3 Freundinnen trinken jeweils 3 Gläser,
 - ▶ 4 Freundinnen trinken jeweils 5 Gläser,
 - ▶ 1 Freundin trinkt 8 Gläser und
 - ▶ Pia übernimmt den Rest.
- a) Welchen Anteil am Prosecco muss Pia trinken?
- b) Geben Sie die Häufigkeitsverteilung der Gläser pro Dame an.
- c) Bestimmen Sie den Median, das arithmetische Mittel und die Standardabweichung der Gläser pro Dame.
- d) Welche Werte nimmt die zur Anzahl der Gläser x pro Dame gebildete empirische Verteilungsfunktion $F(x)$ bei $x = 2$ und bei $x = 5$ an?
- e) Zeichnen Sie die Lorenzkurve und
- f) berechnen Sie den normierten Gini-Koeffizienten der Gläseranzahl pro Dame.



Aufgabe 28

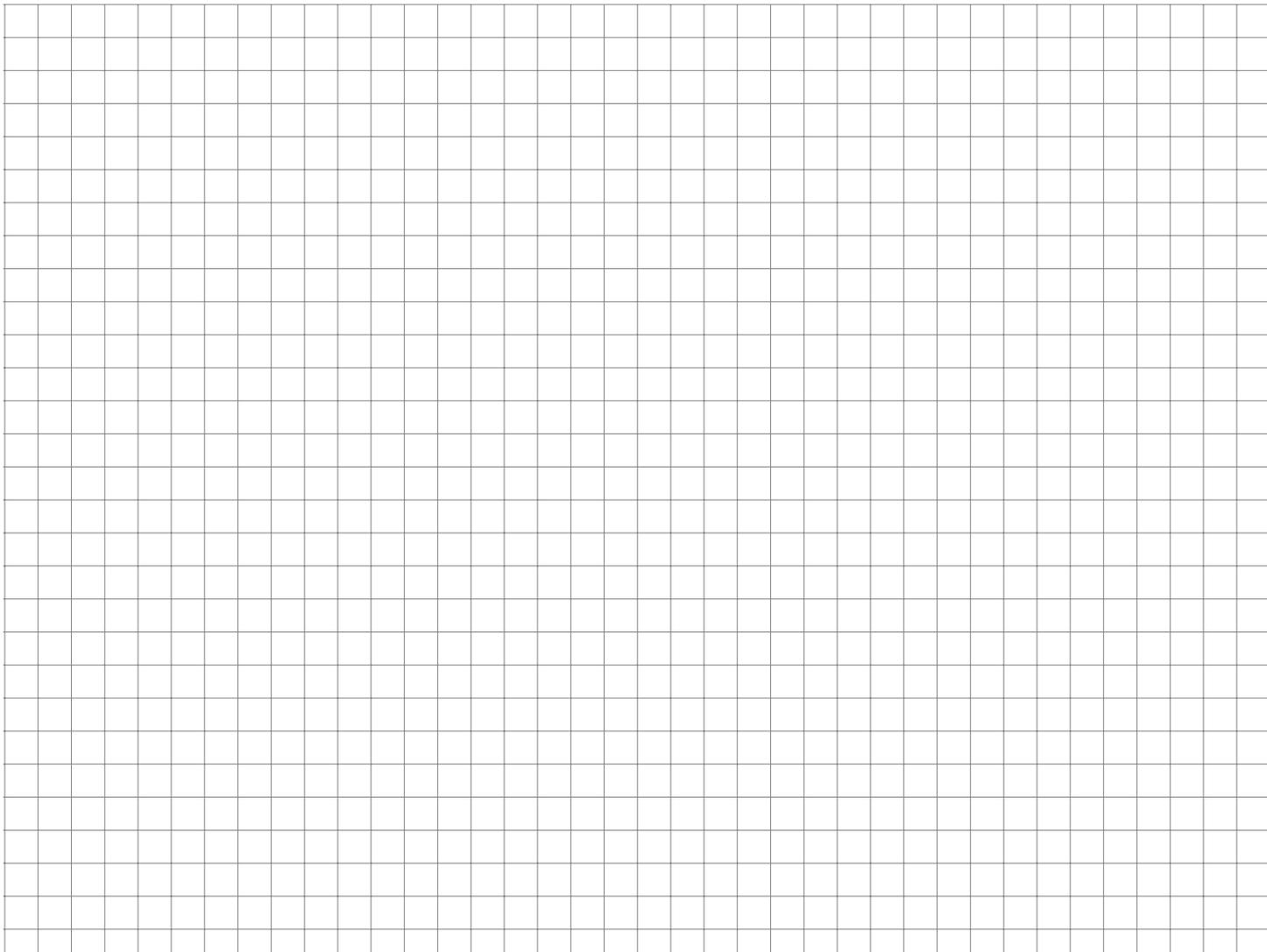
Deskriptiv: Preisindex (15)

An 5 aufeinanderfolgenden Zeitpunkten wurden Preise p und Mengen q zweier Güter G_1 und G_2 festgestellt:

Gut	Jahr	1		2		3		4		5	
		p_1	q_1	p_2	q_2	p_3	q_3	p_4	q_4	p_5	q_5
G_1		2,5	5	3	4	2	6	3,5	3	4	2
G_2		3	6	2	10	4	5	4	4	3	10

Betrachtet werden im Folgenden die Berichtszeit 5 und die Basiszeit 1.

- Berechnen Sie jeweils den Preisindex von Laspeyres, Paasche, Fisher sowie den von Marshall-Edgeworth.
- Bestimmen Sie jeweils zu Gut 1 bzw. Gut 2 die durchschnittlichen Mengen \bar{q}^{G_1} , \bar{q}^{G_2} der konsumierten Mengen über alle 5 Jahre. Berechnen Sie damit den Preisindex von Lowe.
- Angenommen die Mengen zu den Zeitpunkten 1 bzw. 5 sind nicht bekannt. Wie hoch ist der maximale bzw. minimale Wert für den Preisindex von Laspeyres?



Aufgabe 29

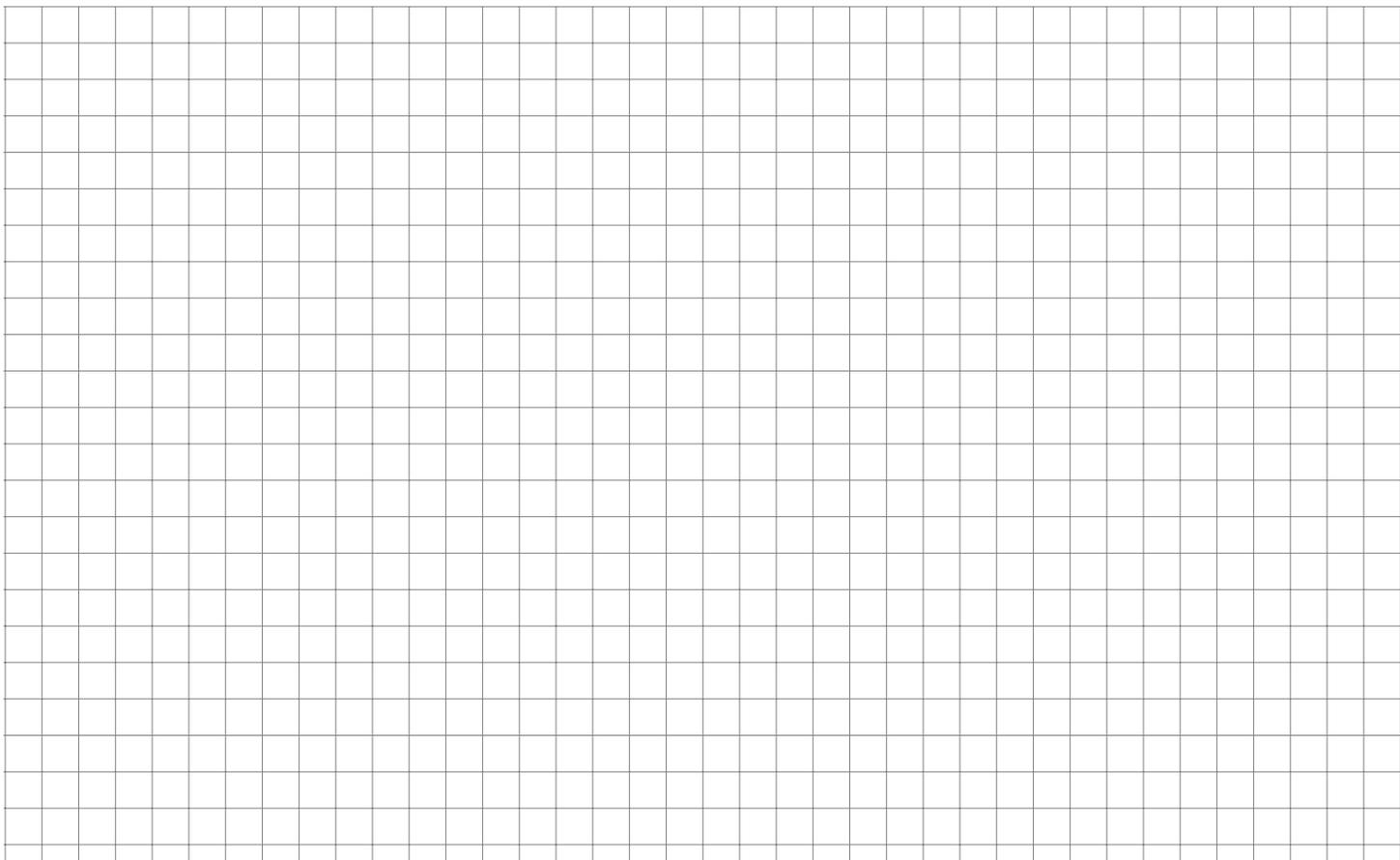
Deskriptiv: Preisindex (Preisindex2)

Für zwei Güter G_1, G_2 sind in den Jahren $t = 1, \dots, 5$ folgende Preise x_t (zu G_1) bzw. y_t (zu G_2) pro Mengeneinheit (= ME) registriert worden.

t	1	2	3	4	5	
x_t	2	2	3	3	5	(€)
y_t	4	5	8	7	8	(€)

Im Jahr 1 wurden 9 ME von G_1 und 8 ME von G_2 verbraucht, im Jahr 4 waren dies 12 ME von G_1 und 9 ME von G_2 .

- ▶ Bei den im folgenden betrachteten Indexzahlen besteht der Warenkorb nur aus G_1 und G_2 ; als Basisperiode ist stets das Jahr 1 angesetzt.
 - a) Offensichtlich sind die Verbrauchsmengen nicht für jedes Jahr bekannt. Für welche $t \in \{2,3,4,5\}$ läßt sich dennoch ein Laspeyres-Preisindex P_{1t}^L ermitteln? (Kurze Begründung, aber keine Berechnung erforderlich!)
 - b) Für welche $t \in \{2,3,4,5\}$ läßt sich ein Paasche-Preisindex P_{1t}^P ermitteln? Berechnen Sie *einen* solchen Indexwert P_{1t}^P .
- ▶ Nun beschreibt $y = \hat{a} + \hat{b}x$ die Regressionsgerade mit dem Preis von G_1 als unabhängiger und dem Preis von G_2 als abhängiger Variablen.
 - c) Berechnen Sie \hat{a} und \hat{b} .
 - d) Prognostizieren Sie auf Basis dieser Regression den Preis pro ME des Gutes G_2 , wenn G_1 pro ME 6 € kostet.



Aufgabe 30

Deskriptiv: Rangkorrelation (6)

Zwei Personen sollen fünf verschiedene Produkte A bis E durch Angabe einer Reihenfolge beurteilen. Die Befragung ergab folgende Ergebnisse:

Produkt	Person I	Person II
A	5	3
B	2	1
C	3	4
D	4	2
E	1	5

Berechnen Sie den Korrelationskoeffizienten von Spearman.



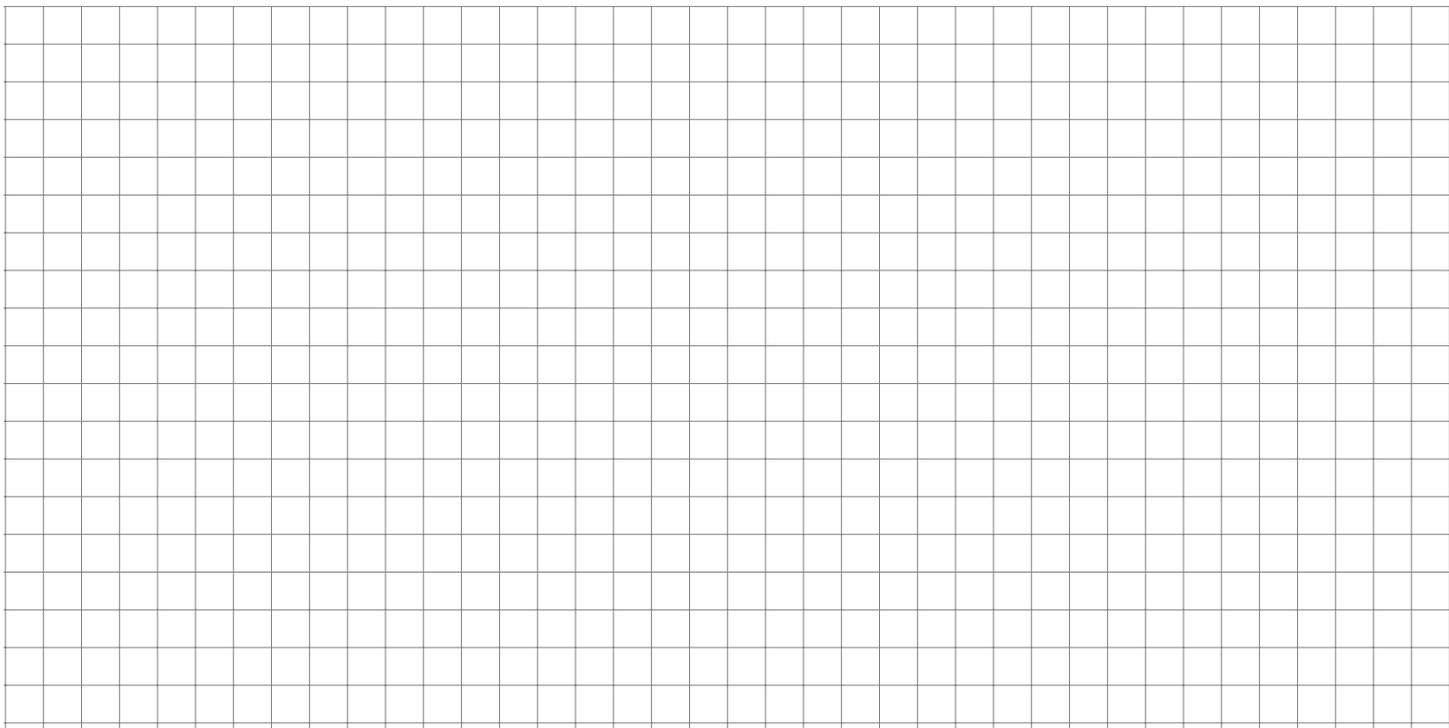
Aufgabe 31

Deskriptiv: Lage Korrelation (11)

Ein Betrieb hat im Kalenderjahr 2004 zwölf neue Mitarbeiter eingestellt. Von diesen sind unter anderem folgende Daten bekannt:

Mitarbeiter Nr.	Geschlecht	Ausbildungsdauer (in Jahren)	Abschlussnote
1	männlich	9	4
2	weiblich	10	2
3	weiblich	10	4
4	männlich	11	4
5	weiblich	12	2
6	weiblich	13	2
7	weiblich	14	1
8	männlich	15	3
9	männlich	16	2
10	männlich	17	3
11	weiblich	19	3
12	männlich	22	2

- a) Geben Sie die Skalierung der drei Merkmale Geschlecht, Ausbildungsdauer und Abschlussnote an.
- b) Ermitteln Sie für jedes der drei Merkmale die folgenden Größen, soweit diese aufgrund des jeweiligen Skalenniveaus sinnvollerweise berechnet werden können:
- (1) Modus
 - (2) Median
 - (3) Arithmetisches Mittel
 - (4) Mittlere quadratische Abweichung
 - (5) Variationskoeffizient
- c) Geben Sie für jedes der zwei Merkmalspaare
- (1) Geschlecht – Abschlussnote
 - (2) Ausbildungsdauer – Abschlussnote
- einen statistisch sinnvollen Korrelationskoeffizienten an.
(Die Korrelationskoeffizienten müssen nicht berechnet werden.)



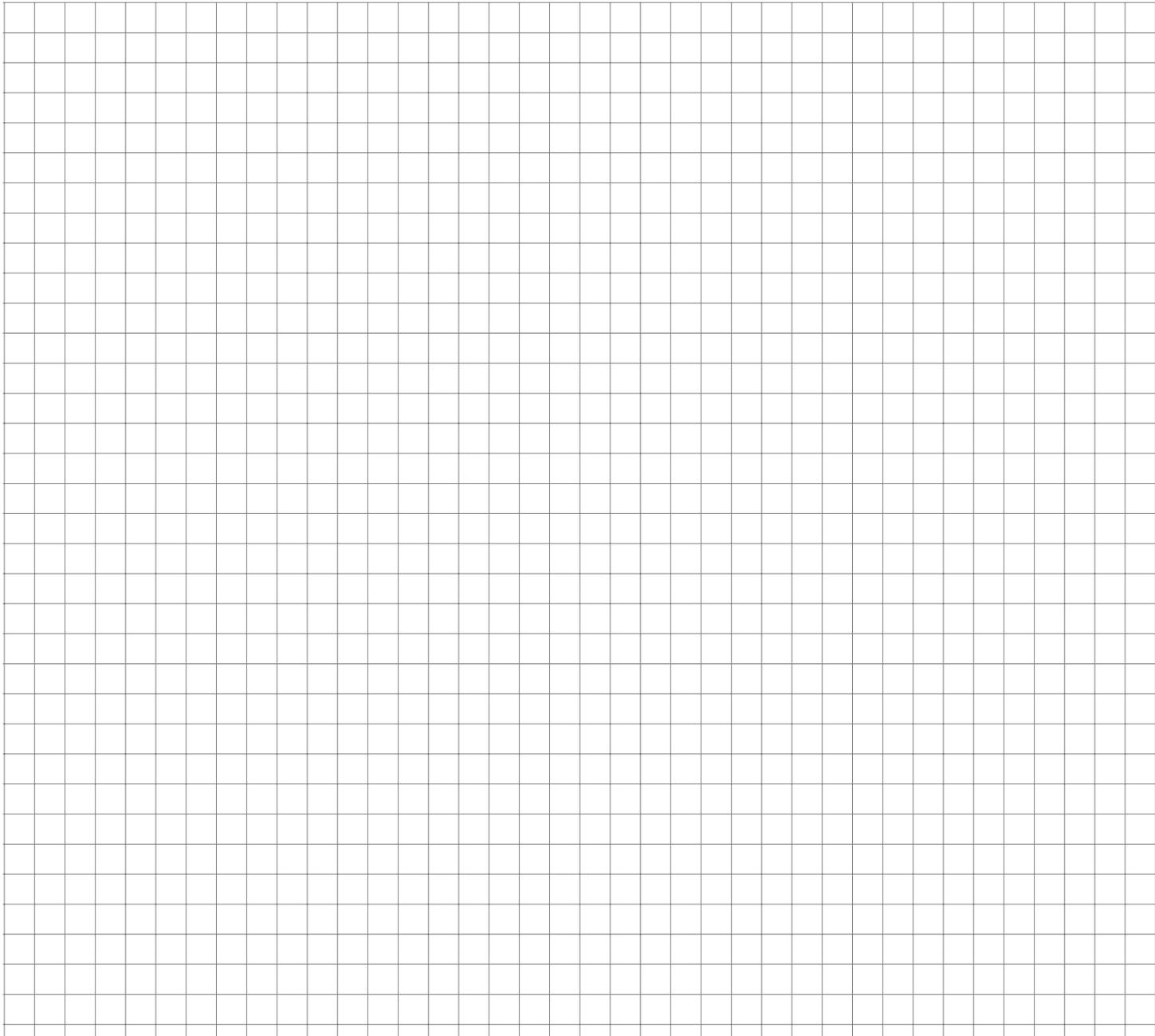
Aufgabe 32

Deskriptiv: Kontingenzkoeffizient (4)

Bei einer Befragung von Passanten in einer Fußgängerzone bezüglich ihres Bierkonsums in Litern pro Woche und ihrer Selbsteinschätzung als Fußballfan ergaben sich folgende Daten:

20 Fußballfans und 120 Nichtfußballfans gaben einen Bierkonsum von höchstens 1 Liter pro Woche an. Zwischen 1 und 3 Liter pro Woche trinken 210 Fußballfans und 200 Nichtfußballfans. 150 Fußballfans und 90 Nichtfußballfans gaben einen Bierkonsum von mindestens 7 Liter an. 145 Fußballfans und 65 Nichtfußballfans lagen in der verbleibenden Zwischen-
gruppe.

- Stellen Sie die zugehörige Kontingenztabelle auf.
- Errechnen Sie die Randhäufigkeiten.
- Berechnen Sie die bedingte Verteilung des Bierkonsums für Fußballfans.
- Sind Bierkonsum und die Fußballaffinität unabhängig?
- Berechnen Sie ein geeignetes Zusammenhangsmaß für die beiden Merkmale.



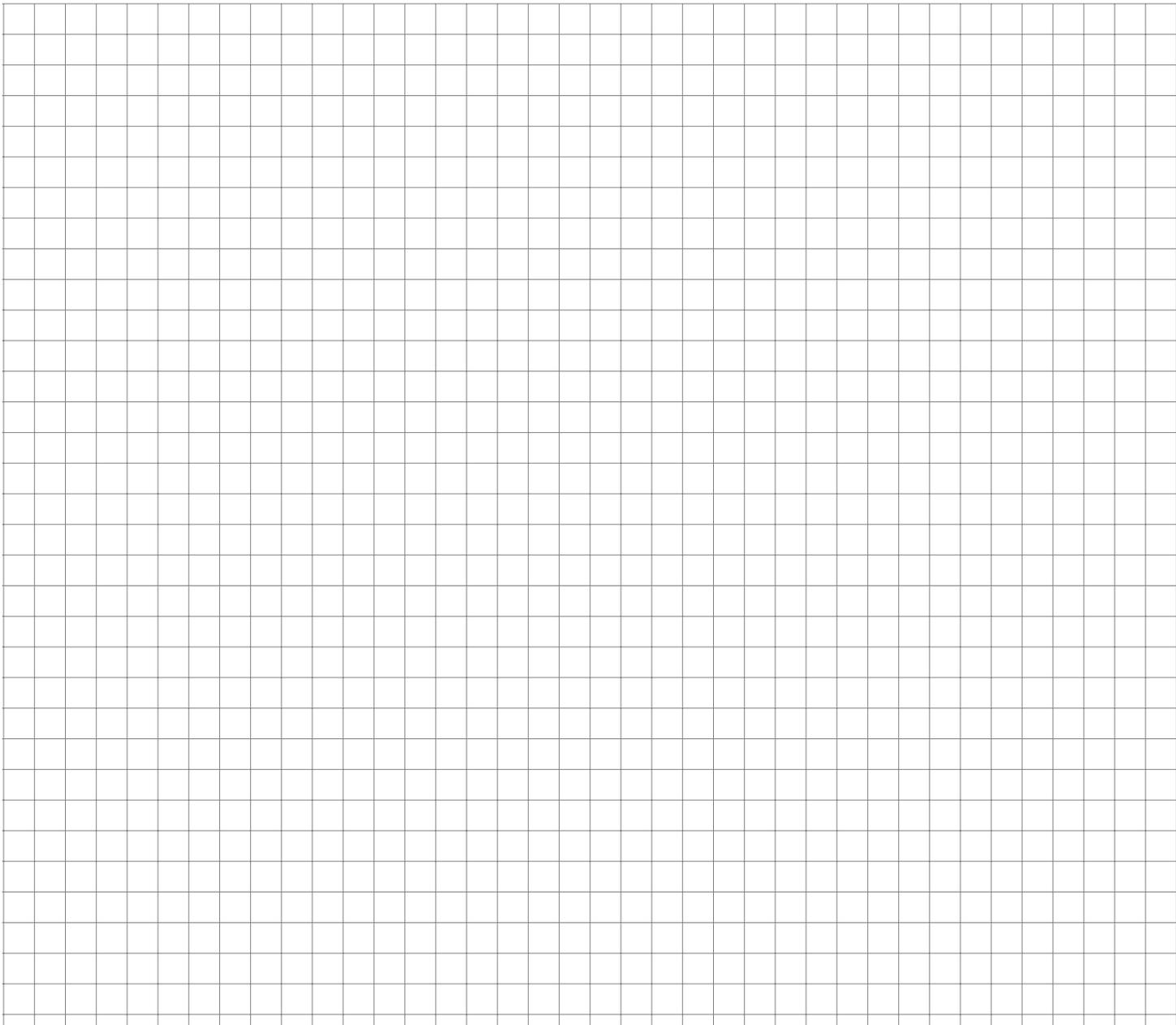
Aufgabe 33

Deskriptiv: Kontingenzkoeffizient (Mensa_Kontingenzkoeffizient)

An einer Hochschule sollen die Studierenden ihre Mensa bezüglich der Qualität des Essens beurteilen. In einer Voruntersuchung haben 50 Studenten aus vier Studienjahren befragt ein bestimmtes Gericht bezüglich des Geschmacks als schlecht, mittel bzw. gut bewertet. Folgende Häufigkeitstabelle fasst die Ergebnisse zusammen:

Bewertung	Studienjahr			
	1	2	3	4
schlecht	0	6	3	1
mittel	5	9	1	0
gut	5	5	6	9

Berechnen Sie den normierten Kontingenzkoeffizient zwischen der Zugehörigkeit zum Studienjahr und der vergebenen Bewertung.



Aufgabe 34

Deskriptiv: Korrelation Regression (Regression_2014_12_29)

Bei 8 zufällig ausgewählten Arbeitnehmerinnen und Arbeitnehmern wird die Schuhgröße S und das jährliche Einkommen E erfasst. Es ergeben sich folgende Werte:

i	1	2	3	4	5	6	7	8
S_i	36	44	40	49	38	34	34	45
E_i	41910	53860	37360	73450	46720	39560	20470	69040

- Berechnen Sie einen geeigneten Korrelationskoeffizienten zwischen Schuhgröße und Einkommen. Interpretieren Sie das Ergebnis.
- Bestimmen Sie die Regressionsparameter eines linearen Modells, in dem die Höhe des Einkommens in Abhängigkeit von der Schuhgröße beschrieben wird.
- Wieviel Einkommen würden Sie gemäß diesem Modell bei einer Person wie zum Beispiel Dirk Nowitzki mit Schuhgröße 54 und bei jemanden mit Größe 35 (z.B. Kylie Minogue) erwarten?
- Zeichnen Sie die Werte zusammen mit der Regressionsgerade in ein geeignetes Koordinatensystem ein.
- Wie hoch ist der Determinationskoeffizient dieses Modells? Was sagt diese Größe aus?
- Bewerten Sie das Modell bezüglich Kausalität versus Korrelation und geben Sie eine potentielle latente Variable an.

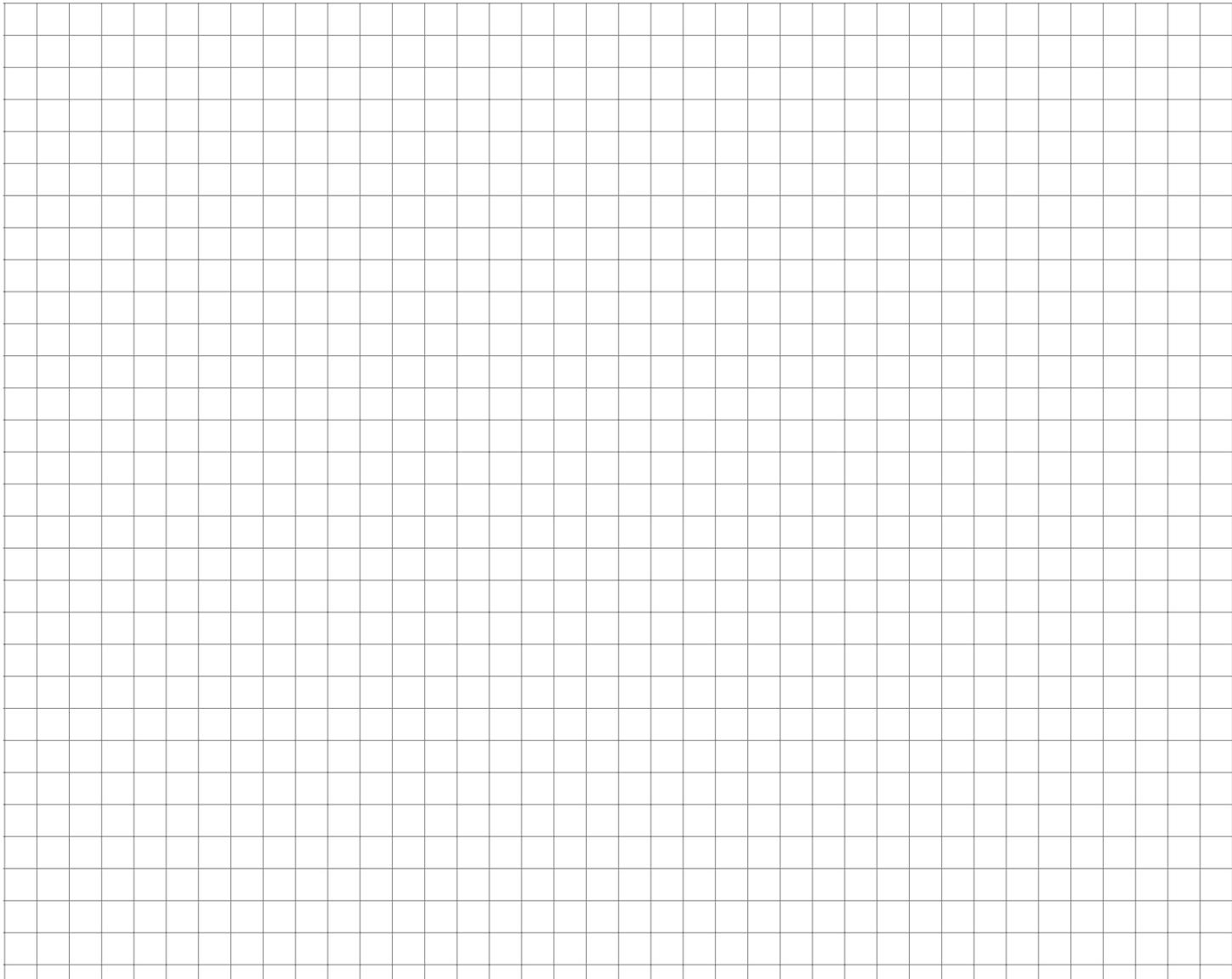


Aufgabe 35

In einem Unternehmen fragt man sich, ob zwischen Umsatz und Marketingkosten ein Zusammenhang besteht. Folgende betrieblichen Daten (in 1000 €) liegen vor:

Marketingkosten/Kunde	Umsatz/Kunde
1,4	210
1,8	220
1,9	240
2,4	241
2,8	320
3,2	400
3,6	410
4,0	480

- Erstellen Sie ein Streuungsdiagramm ($y = \text{Umsatz}$, $x = \text{Marketingkosten}$) und berechnen Sie den Bravais-Pearson- und den Rangkorrelationskoeffizienten.
- Stellen Sie die Regressionsgerade $\hat{y} = \hat{a} + \hat{b}x$ auf und berechnen Sie den Determinationskoeffizienten.



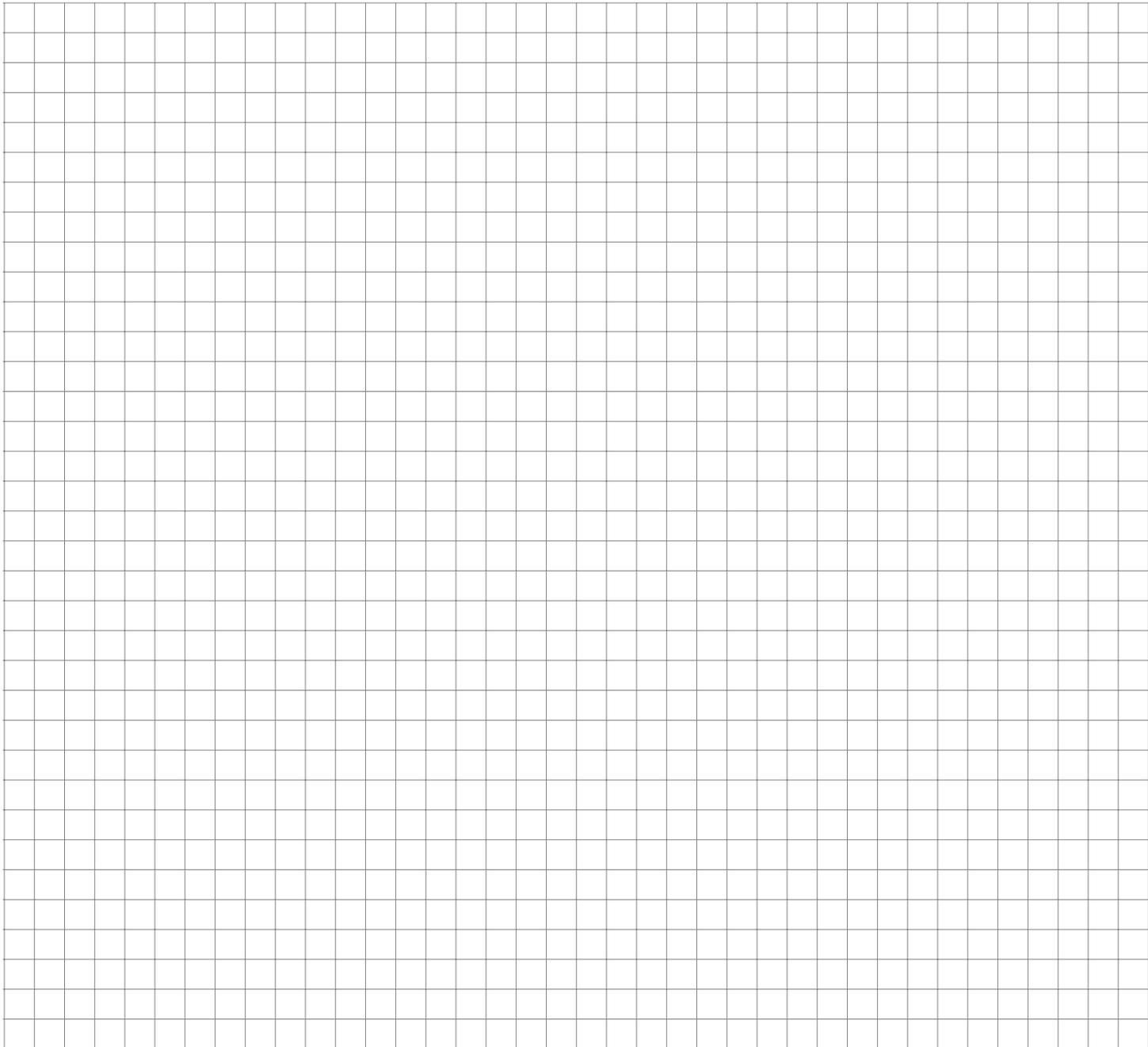
Aufgabe 36

Deskriptiv: Korrelation Regression (12)

Von einer Firma sind über mehrere Jahre hinweg die Umsätze und die Beschäftigtenzahlen bekannt:

Jahr t :	1	2	3	4	5	6
Umsatz x_t (in Millionen €):	60	55	57	61	65	62
Anzahl y_t der Beschäftigten:	1000	1100	960	840	800	700

- Berechnen Sie den Variationskoeffizienten des Umsatzes.
- Berechnen Sie den Rangkorrelationskoeffizienten von Spearman zwischen den beiden Merkmalen Umsatz und Beschäftigtenzahl.
- Berechnen Sie die Regressionsgerade $\hat{y} = \hat{a} + \hat{b}t$ der Beschäftigtenzahl in Abhängigkeit von der Zeit. Mit welcher Anzahl der Beschäftigten ist im Jahr 8 zu rechnen?



Aufgabe 37

An 5 aufeinander folgenden Zeitpunkten wurden Preise p und Mengen q eines Gutes festgestellt:

Zeitpunkt	1	2	3	4	5
p	2,5	3	2	3,5	4
q	5	4	6	3	2

- Berechnen Sie den Rangkorrelationskoeffizienten von Spearman. Welche Vermutung wird durch das Ergebnis nahe gelegt?
- Bestimmen Sie die Regressionsgerade $\hat{q} = \hat{a} + \hat{b} p$.
- Wie groß ist der Korrelationskoeffizient von Bravais-Pearson? (Beachten Sie Ihr Ergebnis aus Teil b!).

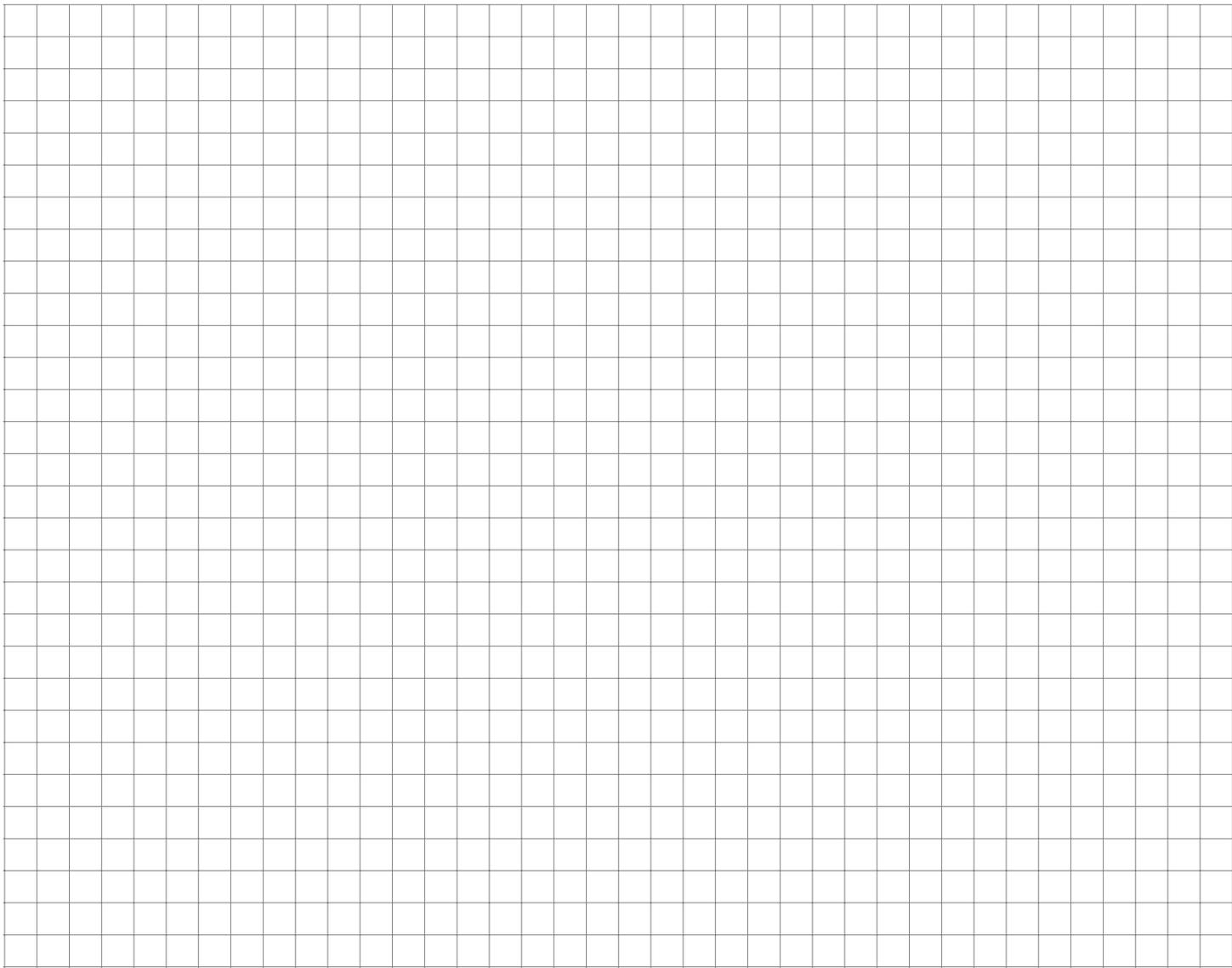


Aufgabe 38

Zwischen der Anzahl der Besucher eines Freibades und der Tageshöchsttemperatur wird ein Zusammenhang vermutet. Es wurden folgende Daten erhoben:

Tag	Besucheranzahl	Höchsttemperatur (in Grad Celsius)
1	340	35
2	150	25
3	250	28
4	300	32
5	240	26
6	220	28

- Berechnen Sie ein geeignetes Zusammenhangsmaß zwischen der Besucheranzahl und der Höchsttemperatur.
- Berechnen Sie die Regressionskoeffizienten der linearen Regression, wenn die Höchsttemperatur als einzige Einflussgröße für die Besucheranzahl erachtet wird.
- Mit welcher Besucheranzahl ist bei einer Höchsttemperatur von 30° zu rechnen?



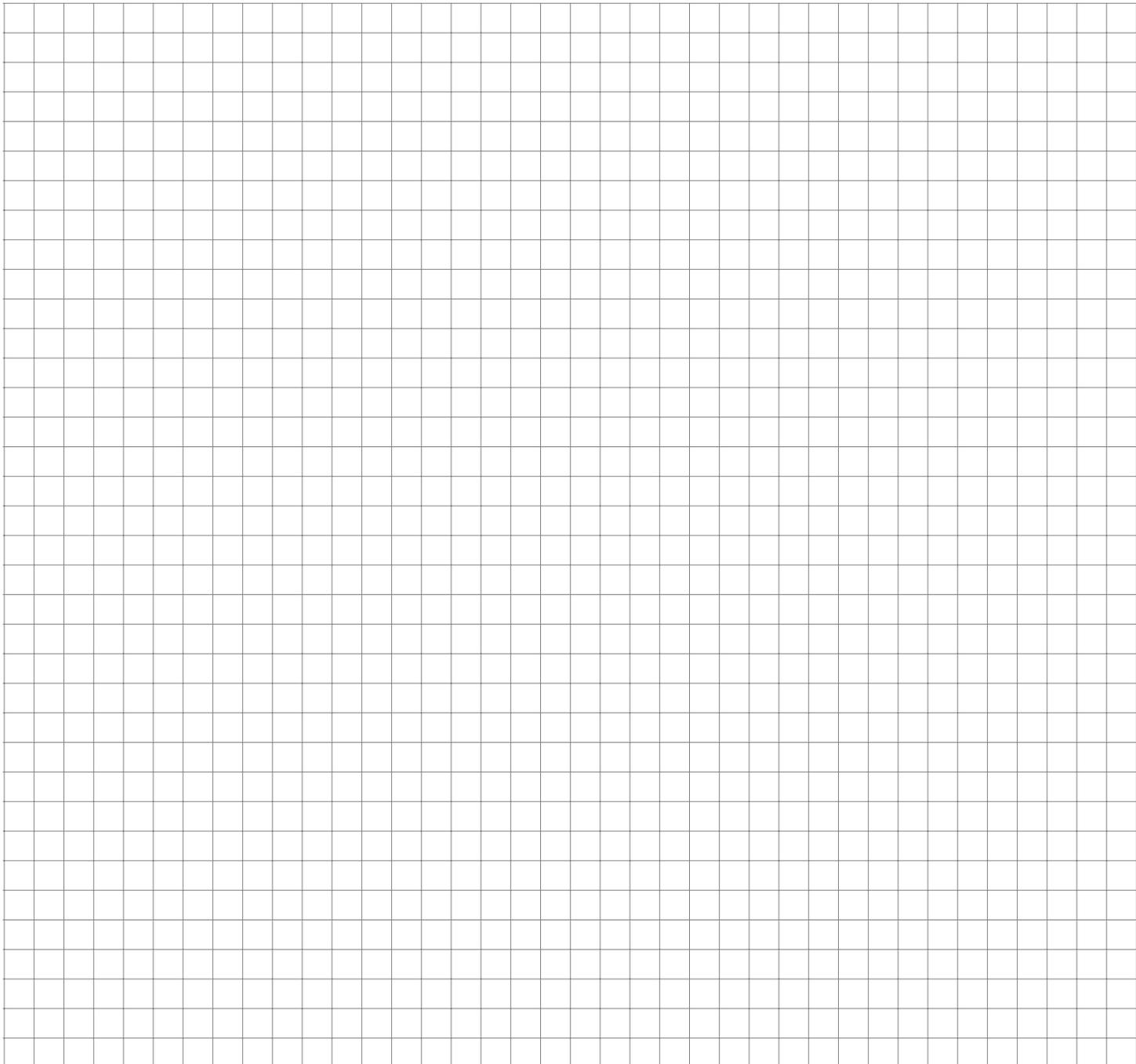
Aufgabe 39

Deskriptiv: Regression (Gehalt_AnzahlToreWM2010)

Das Jahreseinkommen einiger Fußballnationalspieler ist zusammen mit der Anzahl der Tore, die sie in Länderspielen für Deutschland erzielen konnten in folgender Tabelle dargestellt:

Spieler Nummer	Jahreseinkommen [in Mio. €]	Anzahl Tore
1	1,1	0
2	0,8	1
3	2,3	3
4	4,2	3
5	1,7	1
6	0,9	0
7	3,7	4
8	0,7	1
9	2,8	2

- Stellen Sie ein lineares Regressionmodell der Toranzahl in Abhängigkeit vom Spielereinkommen auf.
- Geben Sie den Determinationskoeffizienten an und interpretieren Sie ihn.
- Wieviele Tore würden Sie nach diesem Modell bei einem Einkommen von 10 Mio. erwarten?
- Wieviele Tore müsste nach diesem Modell ein Spieler mehr schießen, wenn er 1 Mio. € mehr verdient?



Aufgabe 40

Deskriptiv: Regression (14)

Zu verschiedenen Zeitpunkten wird der Wasserstand x_t der Isar gemessen.

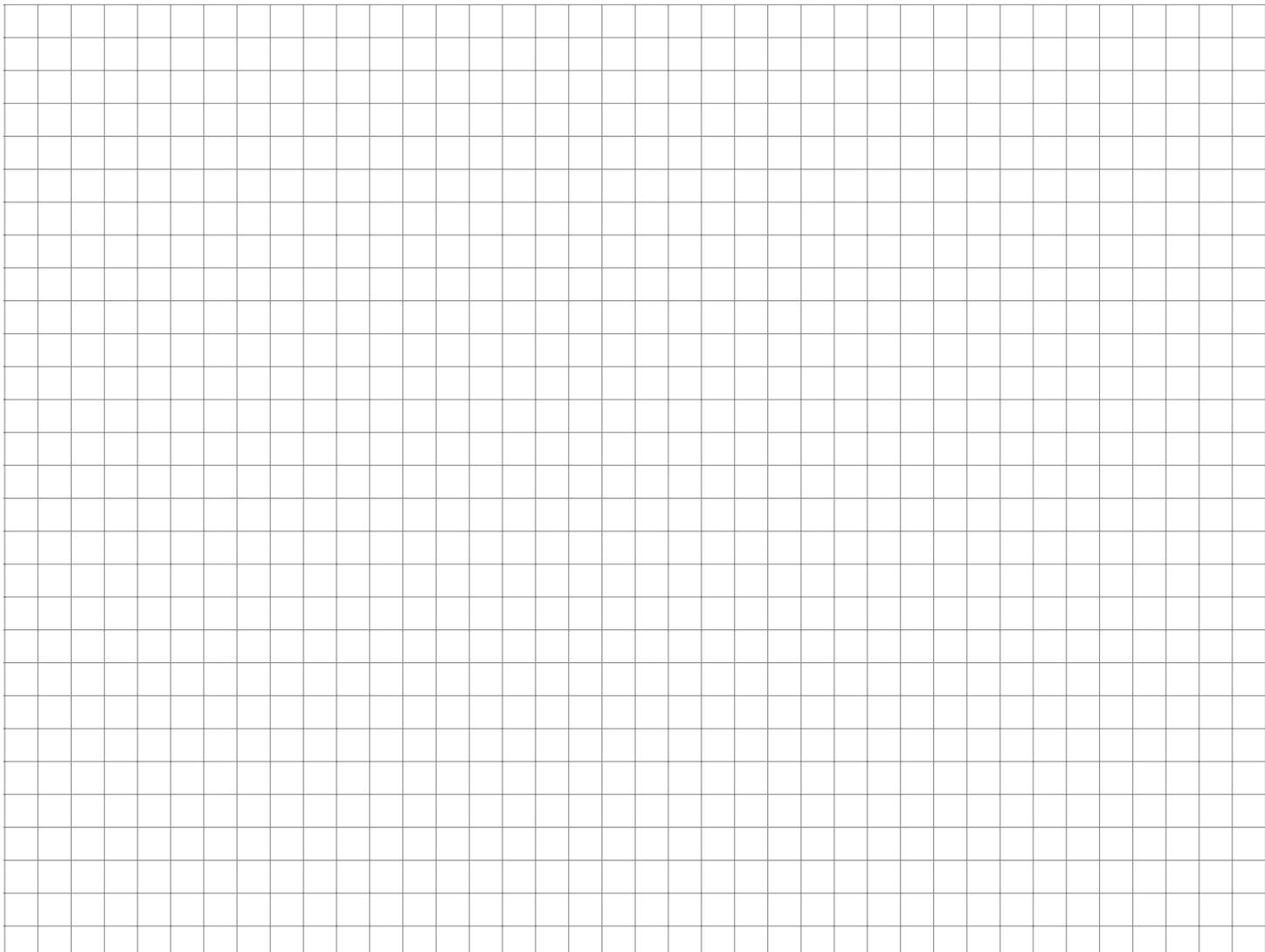
t	1	2	3	4	5	6
Wasserstand in cm	100	110	?	?	125	120

Die Messwerte zum Zeitpunkt 3 und 4 sind leider verloren gegangen. Es ist jedoch Folgendes bekannt:

$$\sum_{t=1}^6 x_t = 705 \quad \text{und} \quad \sum_{t=1}^6 x_t^2 = 83425$$

Außerdem ist bekannt, dass der Messwert zum Zeitpunkt 4 größer ist als der Messwert zum Zeitpunkt 3.

- Ermitteln Sie die fehlenden Messwerte zum Zeitpunkt 3 und 4.
- Prognostizieren Sie den Wasserstand zum Zeitpunkt $t = 7$ mittels einer linearen Regression.
- Ermitteln Sie den Determinationskoeffizienten der Regression.
- Begründen Sie kurz, ob Sie das Vorgehen aus Teilaufgabe b) für sinnvoll erachten.

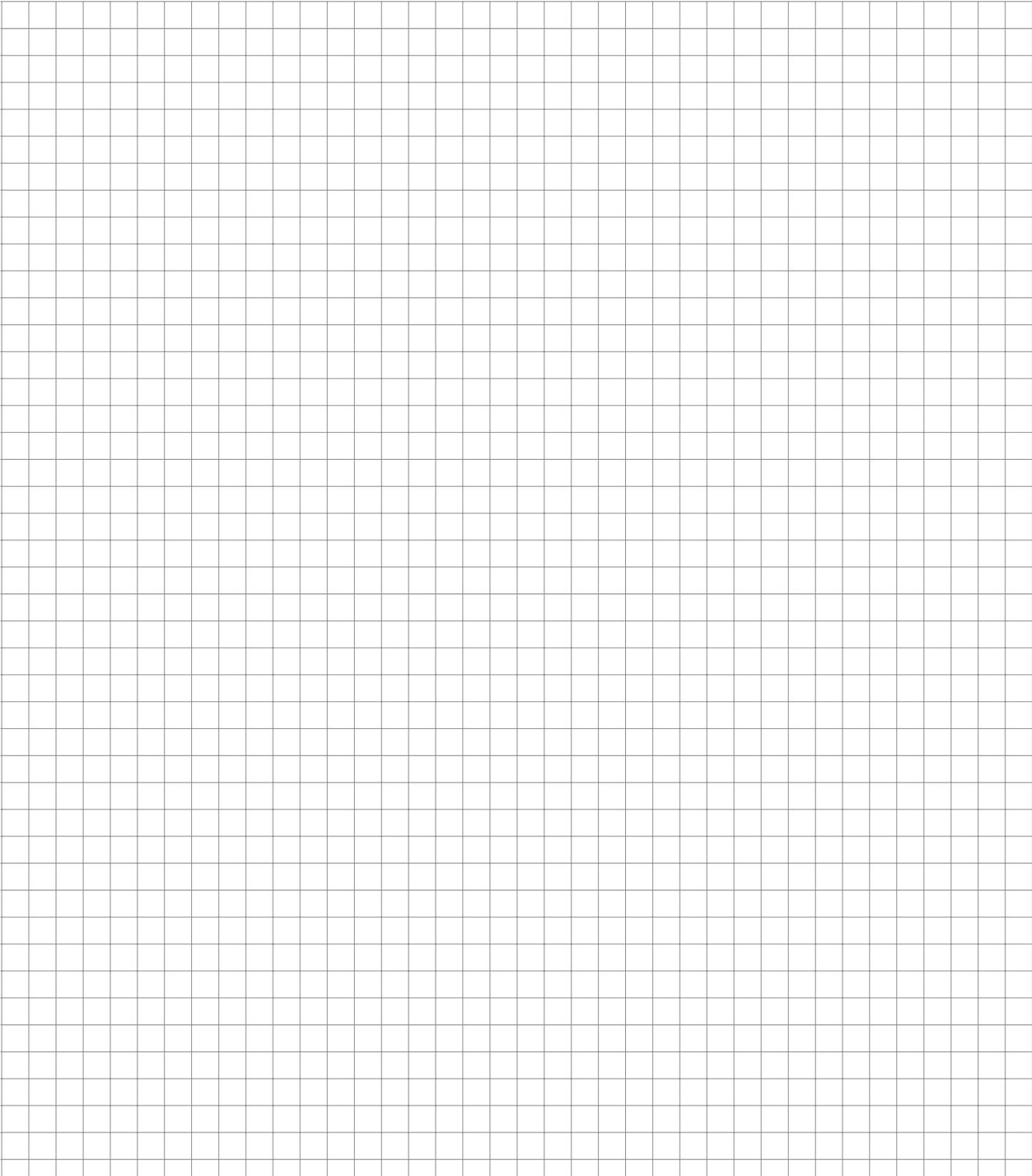


Aufgaben zur Kombinatorik

Aufgabe 41

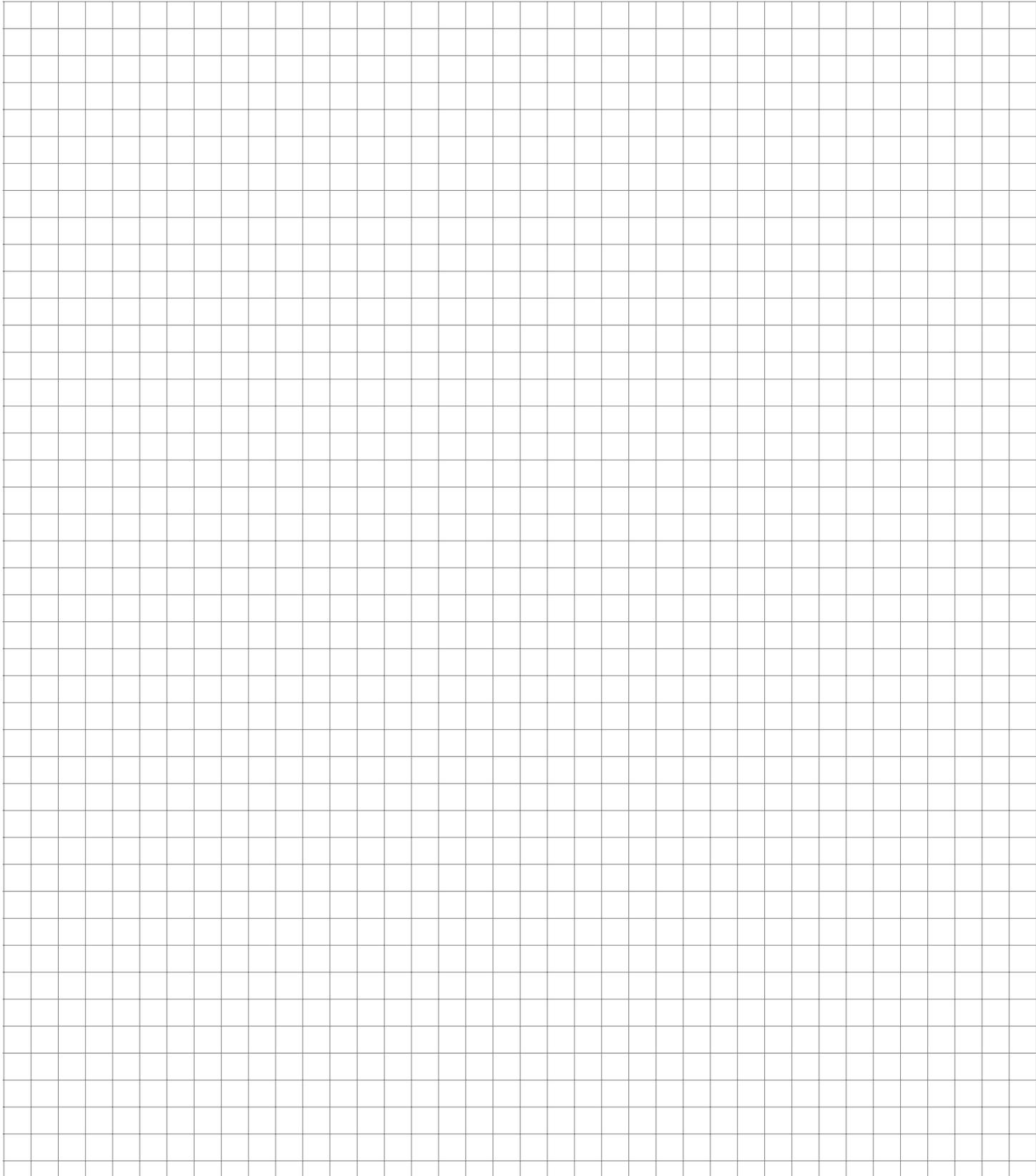
Kombinatorik: Kombinationen (1)

Wie viele verschiedene Möglichkeiten gibt es, 5 Buchstaben aus den 26 Buchstaben des Alphabets nebeneinander zu legen, wenn Wiederholungen zulässig bzw. nicht zulässig sind?



Aufgabe 42

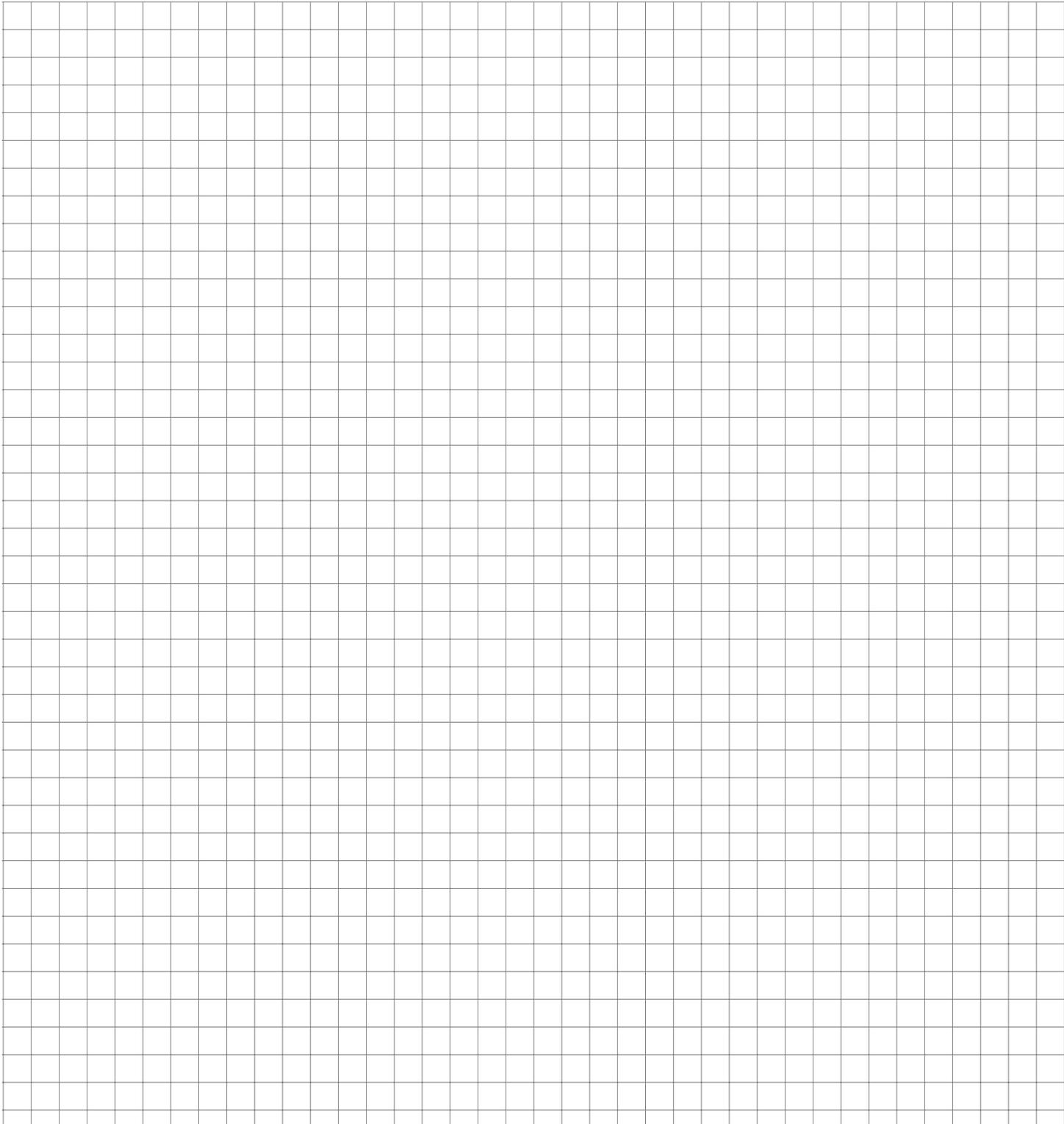
Bei der Beurteilung der Klangqualität von 10 Lautsprecher-Boxen ist in der Weise zu verfahren, dass die Tester jeweils zwei Boxen durch aufeinander folgendes Anhören miteinander vergleichen. Um die Objektivität der Tester zu überprüfen, soll auch jede Box mit sich selbst in der angegebenen Weise verglichen werden. Wie viele Hörvergleiche sind durchzuführen, wenn es auf die Reihenfolge, in der zwei Boxen angehört werden, nicht ankommt?



Aufgabe 43

Ein Kartenspiel mit 32 verschiedenen Karten soll so unter 4 Spieler aufgeteilt werden, dass jeder genau 8 Karten erhält.

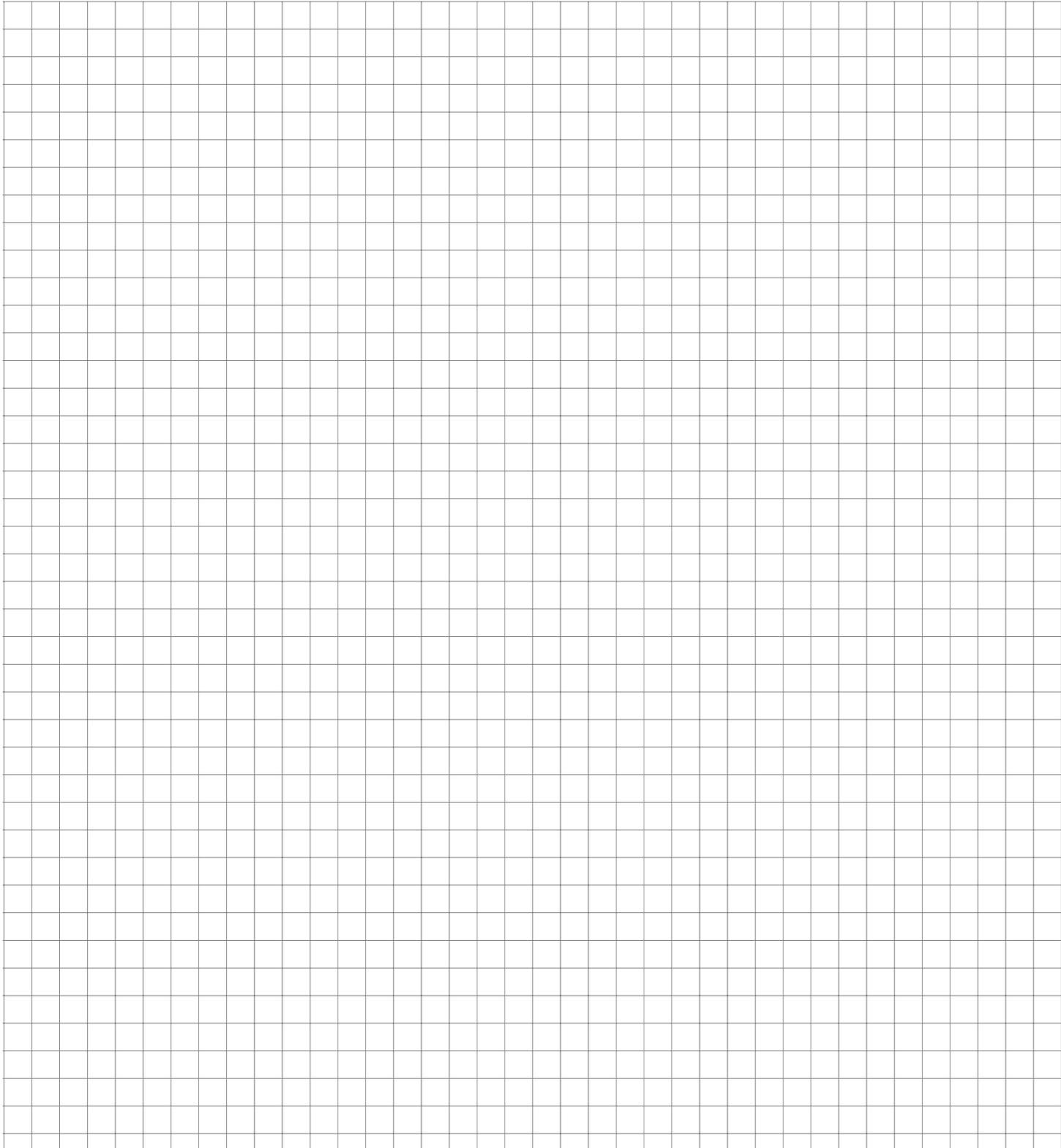
- Wie viele unterschiedliche Möglichkeiten gibt es?
- Wie viele Möglichkeiten gibt es, dass ein bestimmter Spieler alle vier Asse erhält?
- Wie viele Möglichkeiten gibt es, dass einer der Spieler alle vier Asse erhält?
- Bilden Sie den Quotienten des Ergebnisses von c) und a) und interpretieren Sie den erhaltenen Wert.



Aufgabe 44

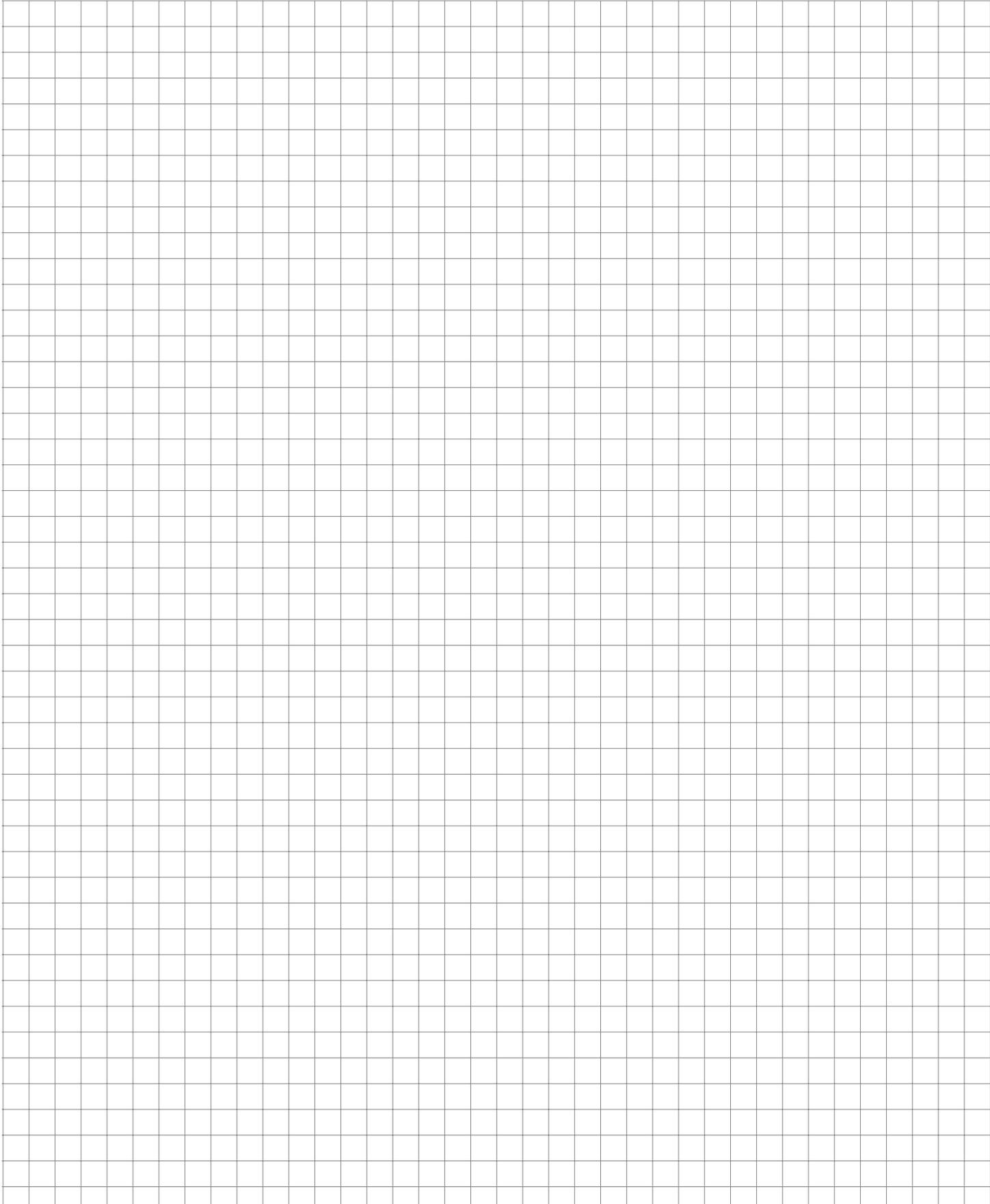
Gegeben seien die Ziffern 1, 2, 3, 4, 5, 6, 7, 8, 9.

- Wie viele dreistellige Zahlen können daraus gebildet werden, wenn jede Ziffer höchstens einmal vorkommen darf?
- Wie viele der so gebildeten Zahlen sind gerade, wie viele ungerade?
- Wie viele dieser Zahlen sind durch 5 teilbar?
- Wie viele dieser Zahlen sind kleiner als 200 bzw. größer als 500?



Aufgabe 45

Wie viele Möglichkeiten gibt es, im Zahlenlotto „6 aus 49“ genau 3, 4, 5, beziehungsweise 6 richtige Zahlen anzukreuzen?



Aufgabe 46

Kombinatorik: Zählprinzip (Aufgabenreihenfolge_2014_01)

Eine Statistik-Klausur bestehe aus insgesamt 10 Aufgaben mit den (absteigend sortierten) Punktzahlen

22, 20, 16, 12, 12, 10, 8, 8, 8, 6.

Die Bearbeitung der einzelnen Aufgaben sei in beliebiger Reihenfolge zulässig.

- Wie viele unterschiedliche Anordnungen (unterschiedliche Bearbeitungsreihenfolgen) gibt es, wenn alle Aufgaben bearbeitet werden?
- Wie viele unterschiedliche Anordnungen (unterschiedliche Auswahlen der Aufgaben sowie unterschiedliche Bearbeitungsreihenfolgen) gibt es, wenn nur 5 Aufgaben bearbeitet werden?
- Eine Studentin verfolgt die Strategie, die Aufgaben in absteigender Reihenfolge der erreichbaren Punktzahlen zu bearbeiten. Haben mehrere Aufgaben eine übereinstimmende Punktzahl, wählt Sie irgendeine Anordnung dieser Aufgaben. Wie viele unterschiedliche Bearbeitungsreihenfolgen zur Bearbeitung aller Aufgaben bleiben bei dieser Strategie möglich?



Aufgaben zur Wahrscheinlichkeitstheorie

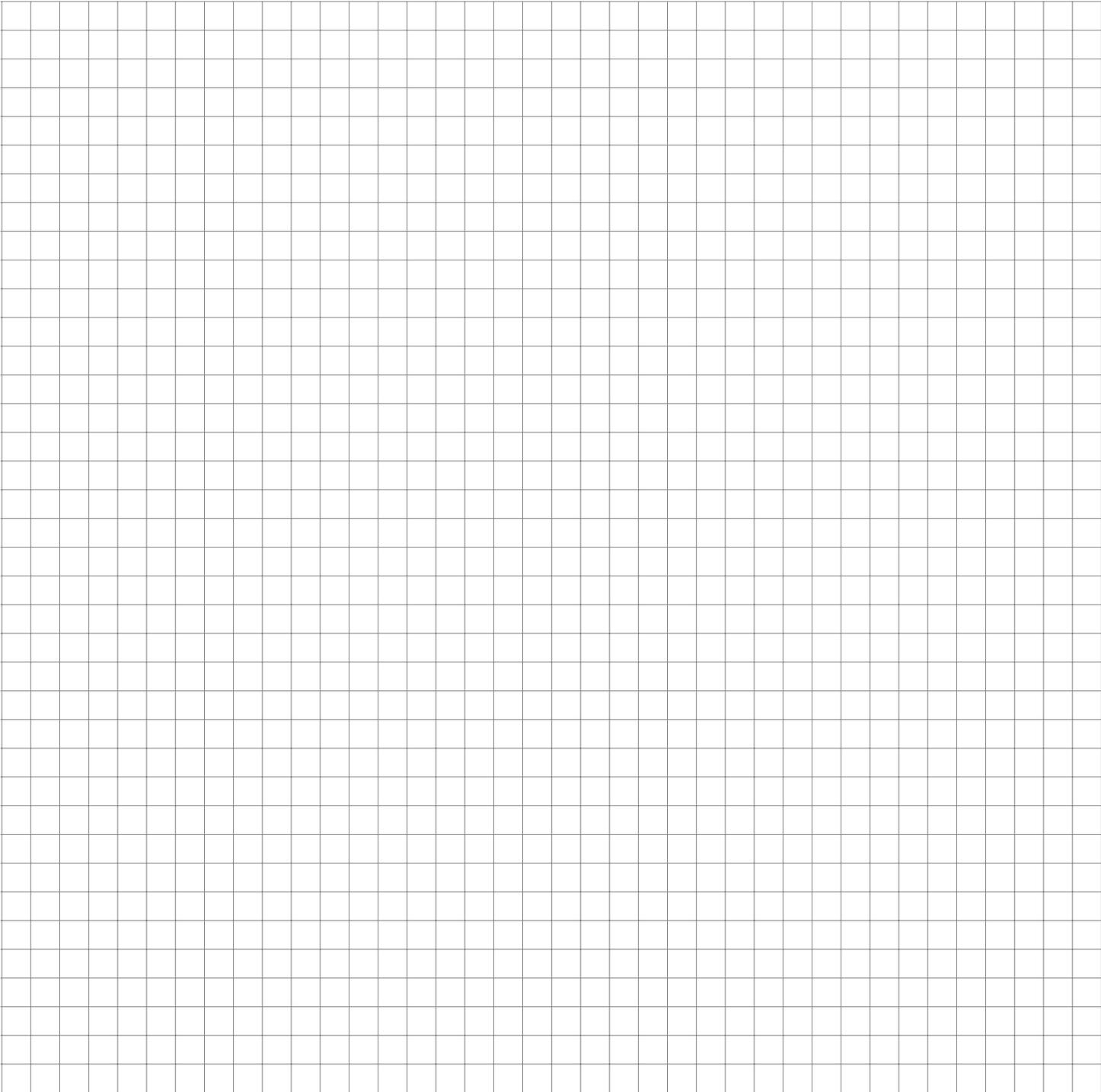
Aufgabe 47

WTheorie: Laplace-Wahrscheinlichkeit (1)

Wie groß ist die Wahrscheinlichkeit, mit viermaligem Werfen eines Würfels

- a) viermal 6
- b) keine 6
- c) mindestens eine 6
- d) der Reihe nach 6,6,6,5
- e) dreimal 6 und einmal 5
- f) genau die Augensumme 7
- g) mindestens zweimal die gleiche Zahl

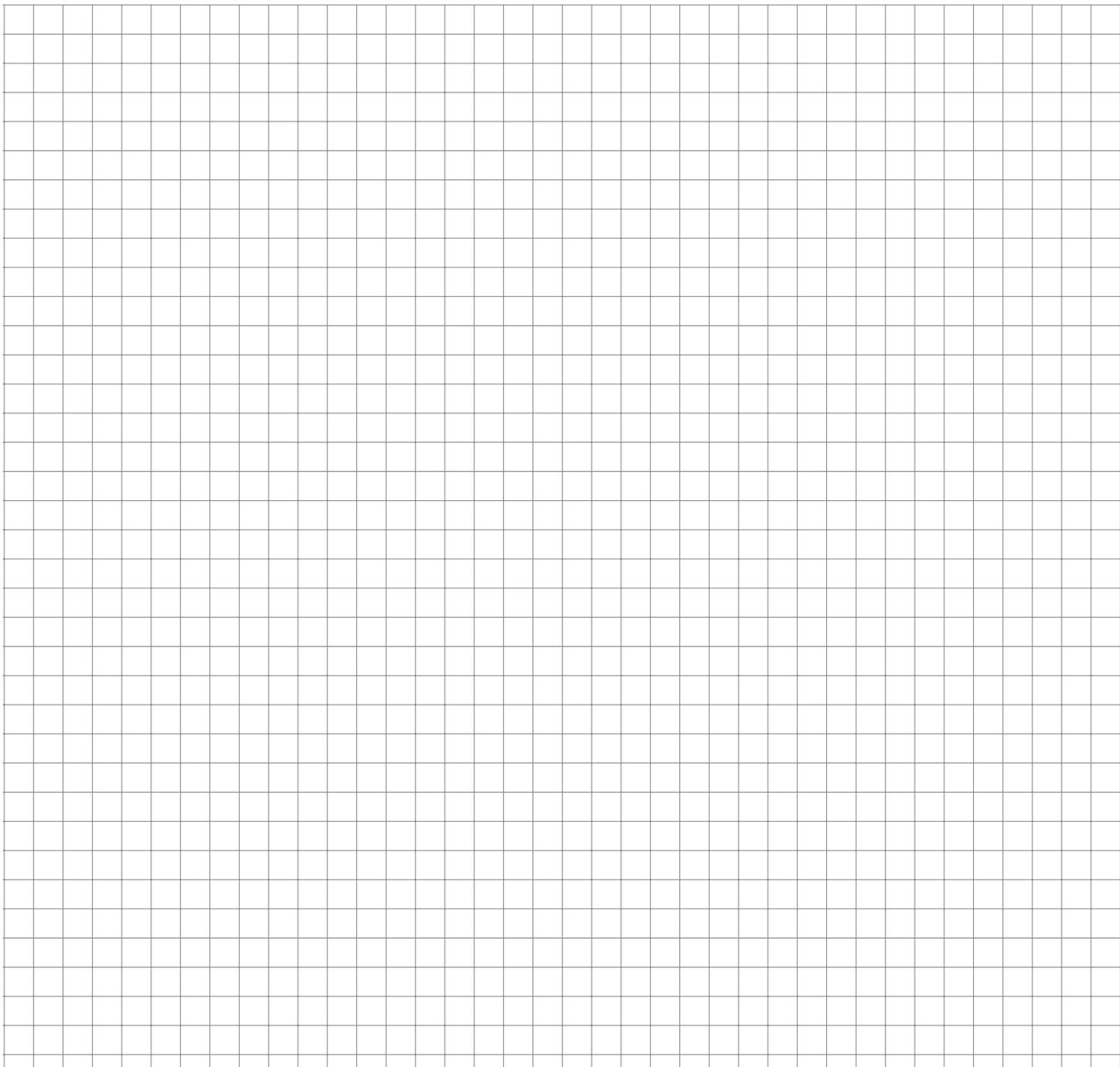
zu erhalten?



Aufgabe 48

In einem Raum befinden sich n Personen, von denen niemand am 29. Februar Geburtstag hat. Nehmen Sie weiterhin an, dass Sie selbst auch nicht am 29. Februar Geburtstag haben.

- Sei $n = 3$. Wie groß ist die Wahrscheinlichkeit, dass mindestens zwei der drei Personen am gleichen Tag (Tag und Monat) Geburtstag haben?
- Wie viele Leute müssen sich im Raum befinden, so dass die Wahrscheinlichkeit mindestens 50 % beträgt, dass mindestens zwei Personen am gleichen Tag Geburtstag haben?
- Sei $n = 100$. Wie groß ist die Wahrscheinlichkeit, dass mindestens noch eine Person am selben Tag Geburtstag hat wie Sie selbst?
- Wie viele Personen müssen sich im Raum befinden, so dass die Wahrscheinlichkeit mindestens 50 % beträgt, dass mindestens noch eine Personen am gleichen Tag Geburtstag hat wie Sie selbst?

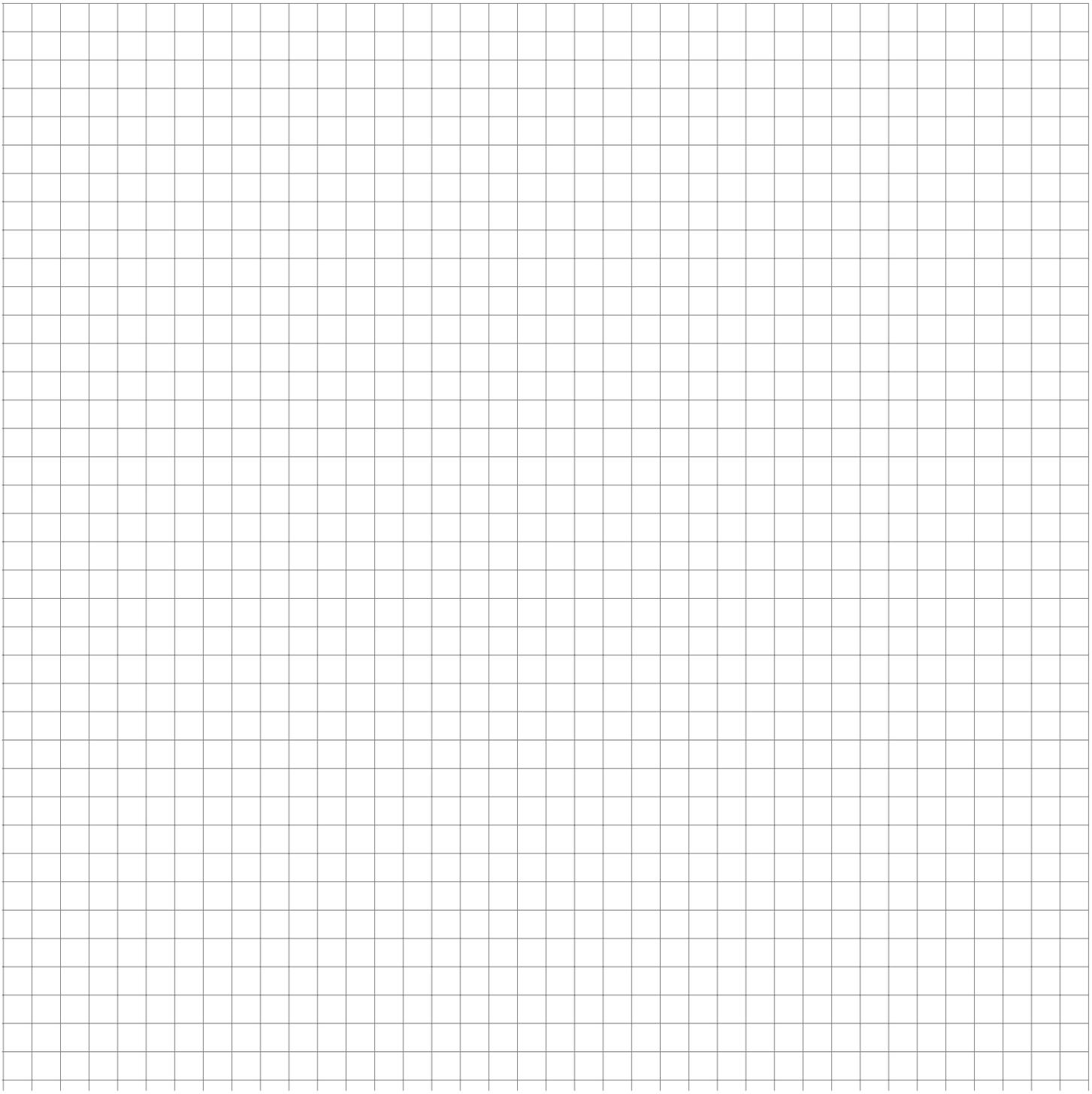


Aufgabe 49

Ein dreimotoriges Flugzeug stürzt ab, wenn der Hauptmotor in der Mitte ausfällt oder beide Seitenmotoren ausfallen. Es wird angenommen, dass jeder der Flugzeugmotoren mit der Wahrscheinlichkeit p auf einem bestimmten Flug ausfällt. Ferner wird angenommen, dass der Ausfall eines Motors unabhängig vom Verhalten der anderen Motoren erfolgt.

A bezeichne das Ereignis, dass ein Flugzeug dieses Typs infolge von Motorversagen abstürzt.

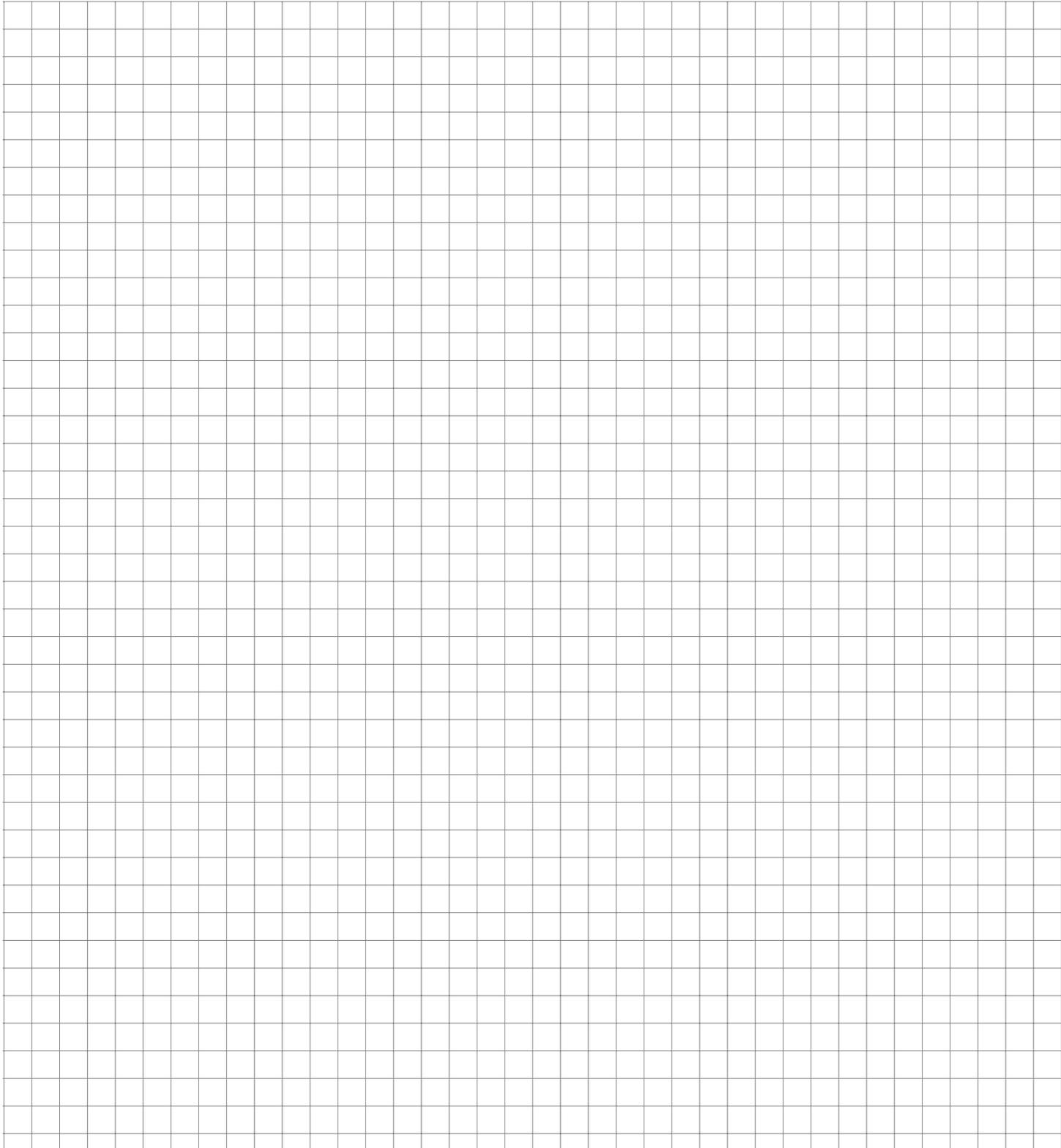
- Ist die Wahrscheinlichkeit $P(A)$ größer oder kleiner als p ? Bitte begründen Sie Ihre Antwort.
- Bestimmen Sie $P(A)$ für $p = 0,01$.



Aufgabe 50

Ein Kraftfahrzeughändler weiß aus langjähriger Erfahrung, dass bei den in Zahlung genommenen Wagen 50% Mängel am Motor, 70% an der Karosserie und 30% an Motor und Karosserie aufweisen. Wie groß ist die Wahrscheinlichkeit dafür, dass ein in Zahlung genommener Wagen

- a) ohne Mängel an Motor und Karosserie ist,
- b) auch einen Mangel am Motor besitzt, wenn bekannt ist, dass die Karosserie schadhaft ist?

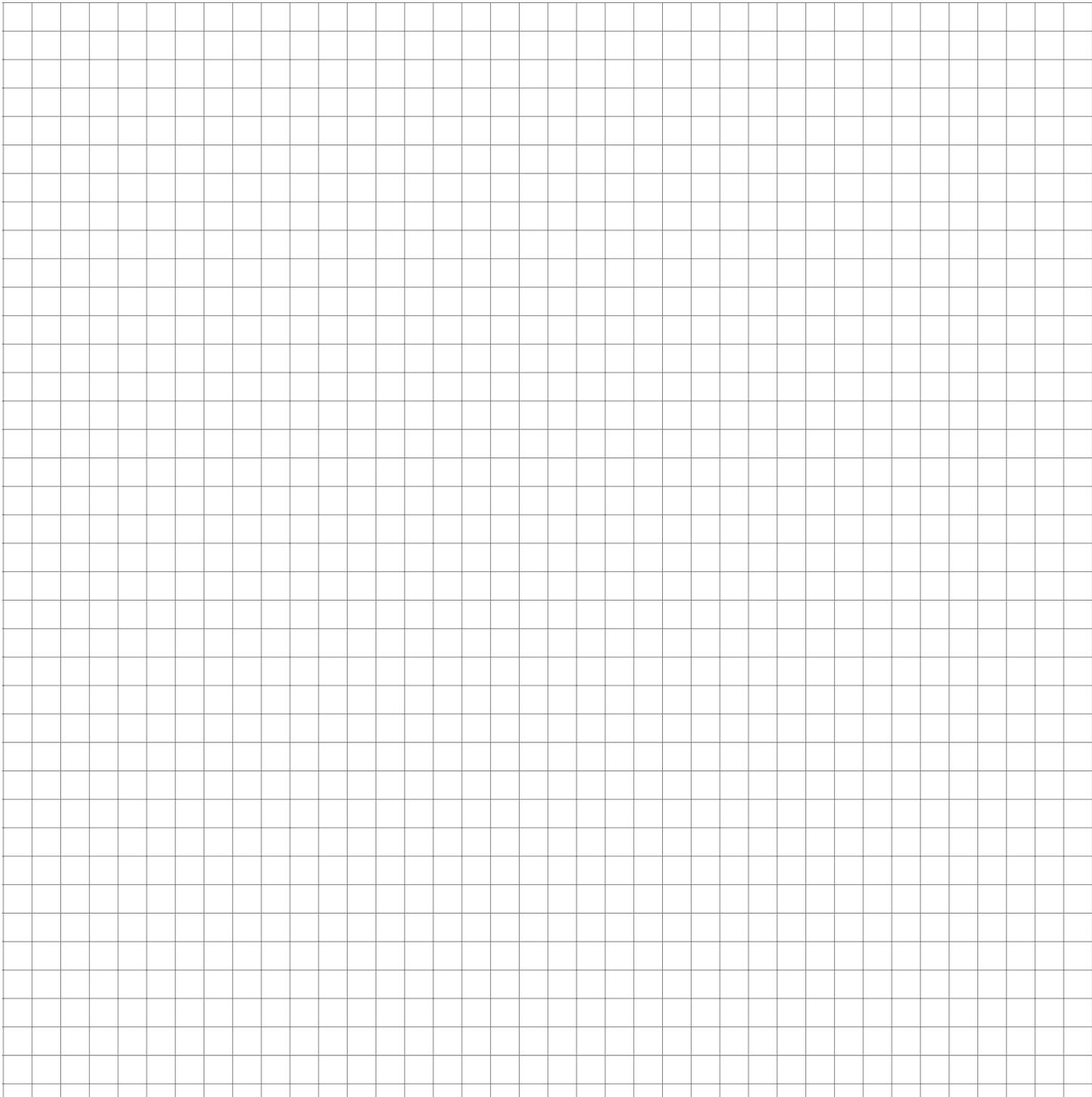


Aufgabe 51

Der Bauer Bertram hat 3 Hühner (Erna, Lisa und Moni). Erna ist seine Lieblingshenne, denn sie liefert durchschnittlich 40 % aller pro Jahr gelegten Eier, während Lisa und Moni nur jeweils 30 % schaffen. Da die Eier ein Mindestgewicht haben müssen, gibt es einen gewissen Ausschuß (A). Bei Erna und Lisa beträgt er jeweils 3 % und bei Moni 5 %.

Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Ei

- a) von Lisa stammt
- b) zu klein ist
- c) von Lisa stammt, wenn bekannt ist, dass es zu klein ist?

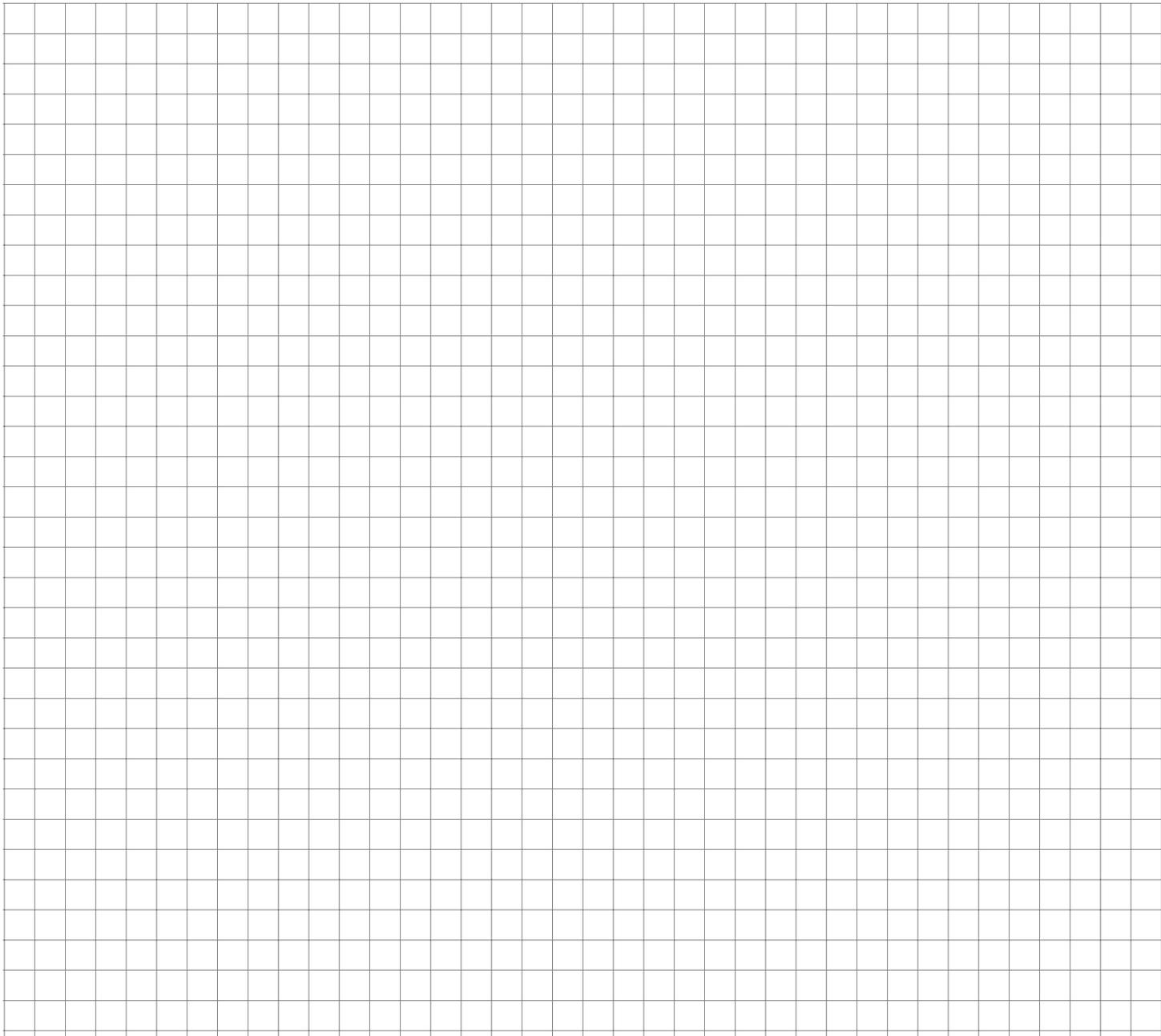


Aufgabe 52

WTheorie: bedingte Wahrscheinlichkeit (3Koenige_2014_01)

Alexandra, Bernhard und Claudio sind als heilige drei Könige verkleidet von Haus zu Haus unterwegs. Bei jedem Haus lassen Sie den Zufall entscheiden, wer von den dreien ein Gedicht aufsagen darf. Dazu würfeln sie jeweils vorher einmal mit einem fairen Würfel. Alexandra sagt das Gedicht, wenn eine 1 fällt, Bernhard bei einer 2 oder 3 und Claudio darf bei 4, 5 oder 6 rezitieren. Alexandra sagt das Gedicht mit einer Wahrscheinlichkeit von 80 % perfekt (mit einer Wahrscheinlichkeit von 20 % ist mindestens ein kleiner Fehler dabei), Bernhard sagt das Gedicht mit einer Wahrscheinlichkeit von 90 % perfekt auf, Claudio mit einer Wahrscheinlichkeit von 95 %.

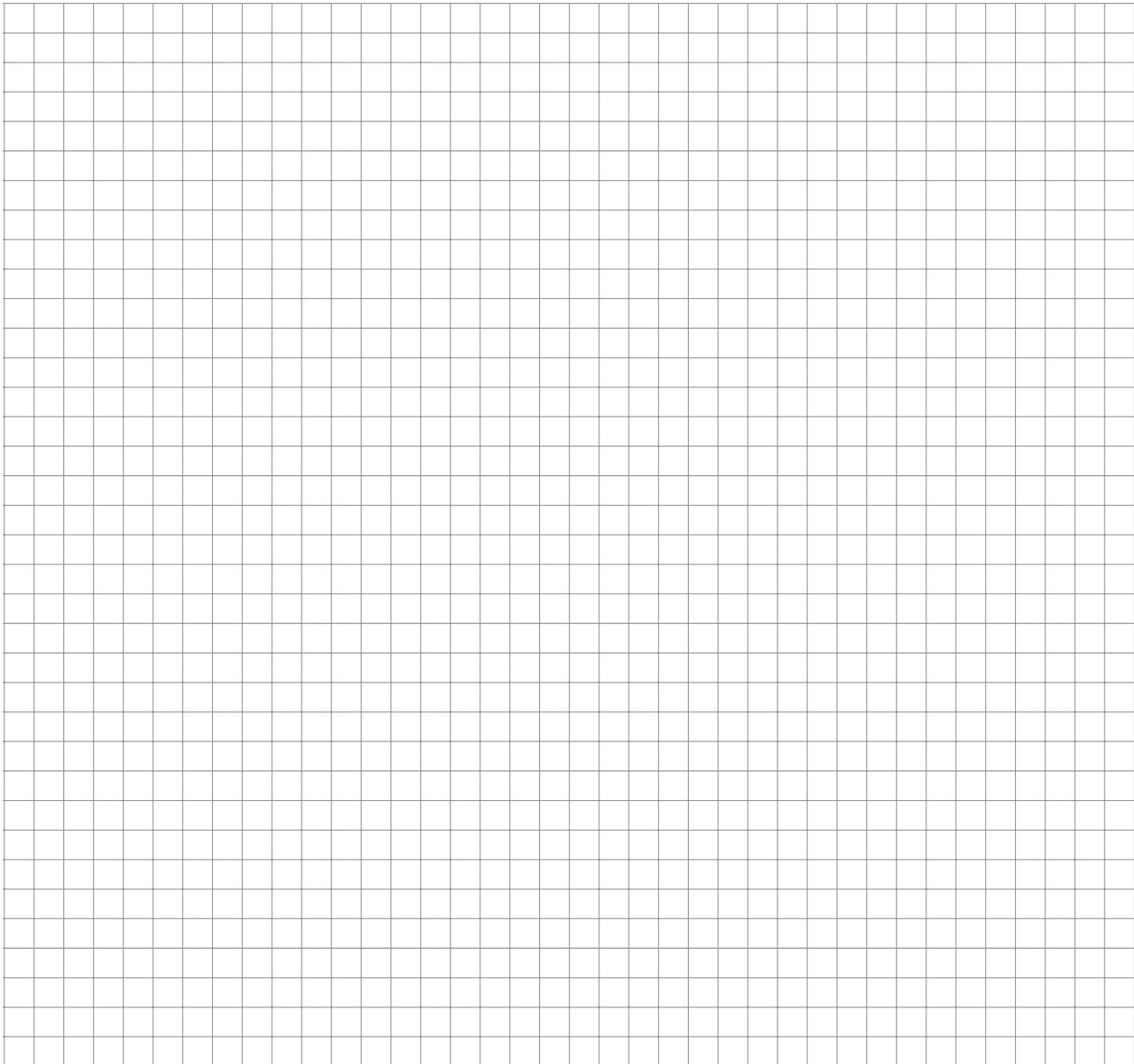
- a) Mit welcher Wahrscheinlichkeit wird das Gedicht bei einem beliebigen Haus perfekt zum Vortrag gebracht?
- b) Frau Maier erzählt am Tag nach dem Besuch der drei Ihrer Nachbarin, dass sich bei Ihr ein Sternsinger beim Gedicht ganz schön verhaspelt hätte. Mit welcher Wahrscheinlichkeit hat Bernhard das Gedicht bei Frau Maier aufgesagt?



Aufgabe 53

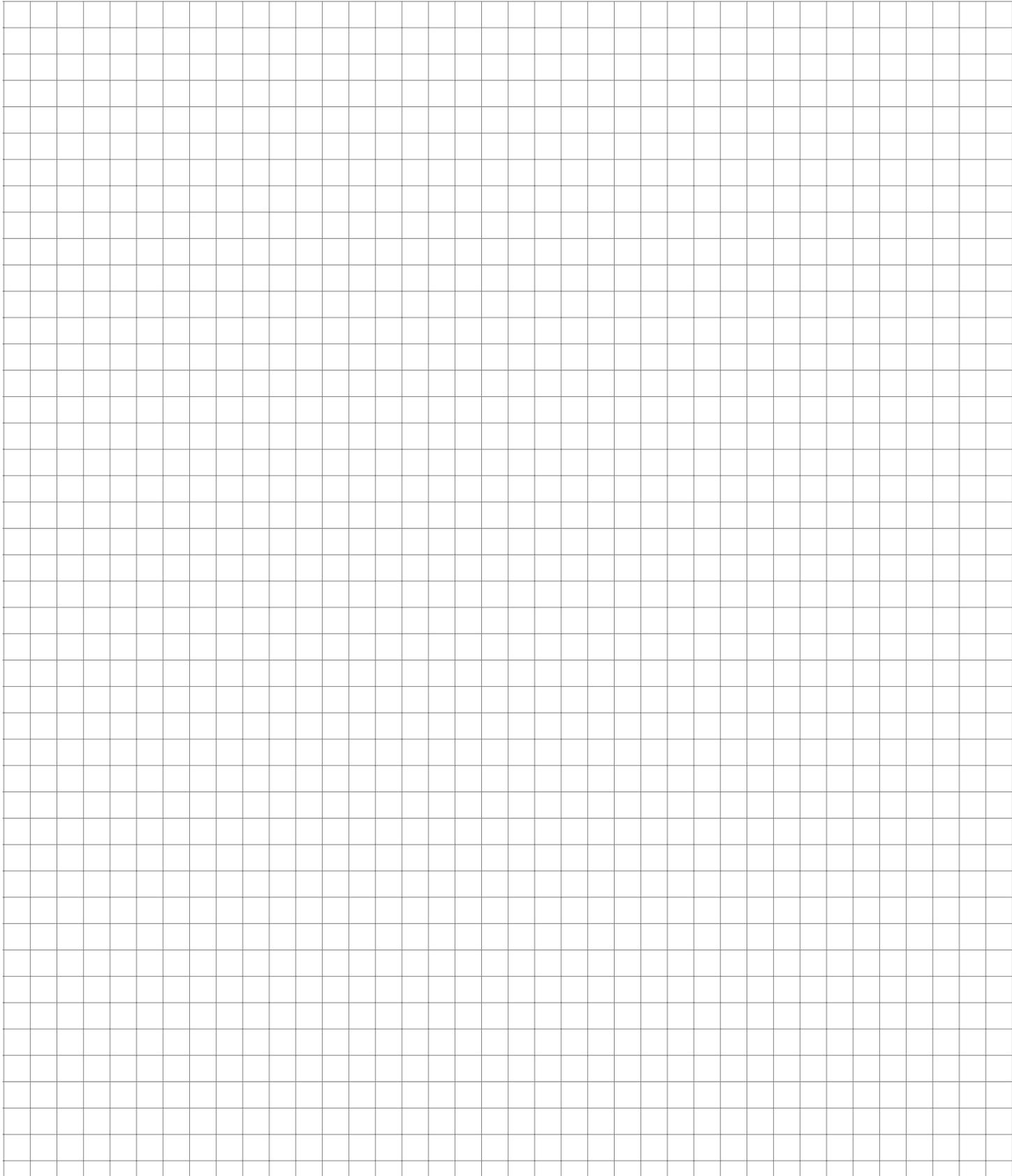
In einer Großbank kommen 80 % der männlichen Kreditkunden ihren Kreditverpflichtungen pünktlich nach, 15 % schleppend nach, und bei 5 % muss die Bank den Kredit abschreiben. Bei den weiblichen Kreditkunden sind die entsprechenden Zahlen 85 %, 10 % und 5 %. Von den Kreditkunden der Bank sind 70 % männlich.

- Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person ihren Kreditverpflichtungen pünktlich nachkommt?
- Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus dem Kreis aller Kreditkunden ausgewählte Person weiblich ist, falls die Person ihren Kreditverpflichtungen nur schleppend nachkommt?
- Sind die Ereignisse „Kunde ist männlich“ und „Kunde zahlt pünktlich“ stochastisch unabhängig?



Aufgabe 54

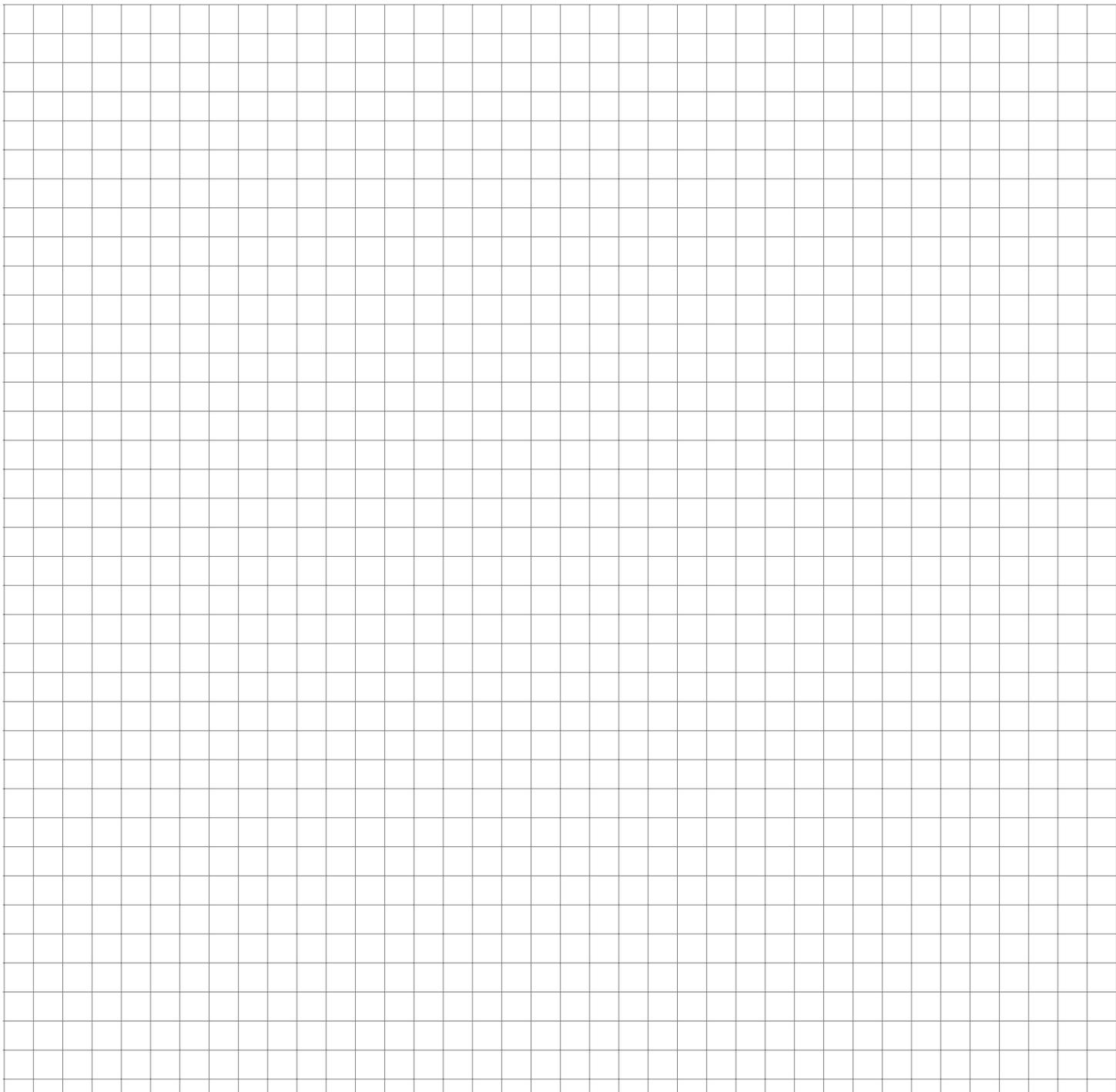
In der Stadt D wird im Mittel zu 10 % schwarz gefahren. 70 % der Schwarzfahrer haben keine Fahrkarte, während die anderen 30 % gefälschte oder illegal besorgte Karten besitzen. Von den ehrlichen Fahrgästen haben im Mittel 5 % ihre Fahrkarte vergessen. Mit welcher Wahrscheinlichkeit ist ein kontrollierter Fahrgast, der keine Karte vorzeigen kann, ein Schwarzfahrer?



Aufgabe 55

10.000 Flugreisende, die aus einem südlichen Land nach Deutschland einreisen werden auf eine ansteckende tropische Krankheit getestet. Ein positiver Test deutet auf eine Erkrankung hin, allerdings nicht sicher. 9 Leute, bei denen der Test positiv ausgefallen ist sind tatsächlich krank. 9899 Leute mit negativem Testergebnissen sind nicht krank. Insgesamt war der Test bei 9900 Untersuchten negativ. Wie hoch ist die Wahrscheinlichkeit, dass bei einer beliebig unter diesen 10.000 Flugreisenden ausgewählten Person

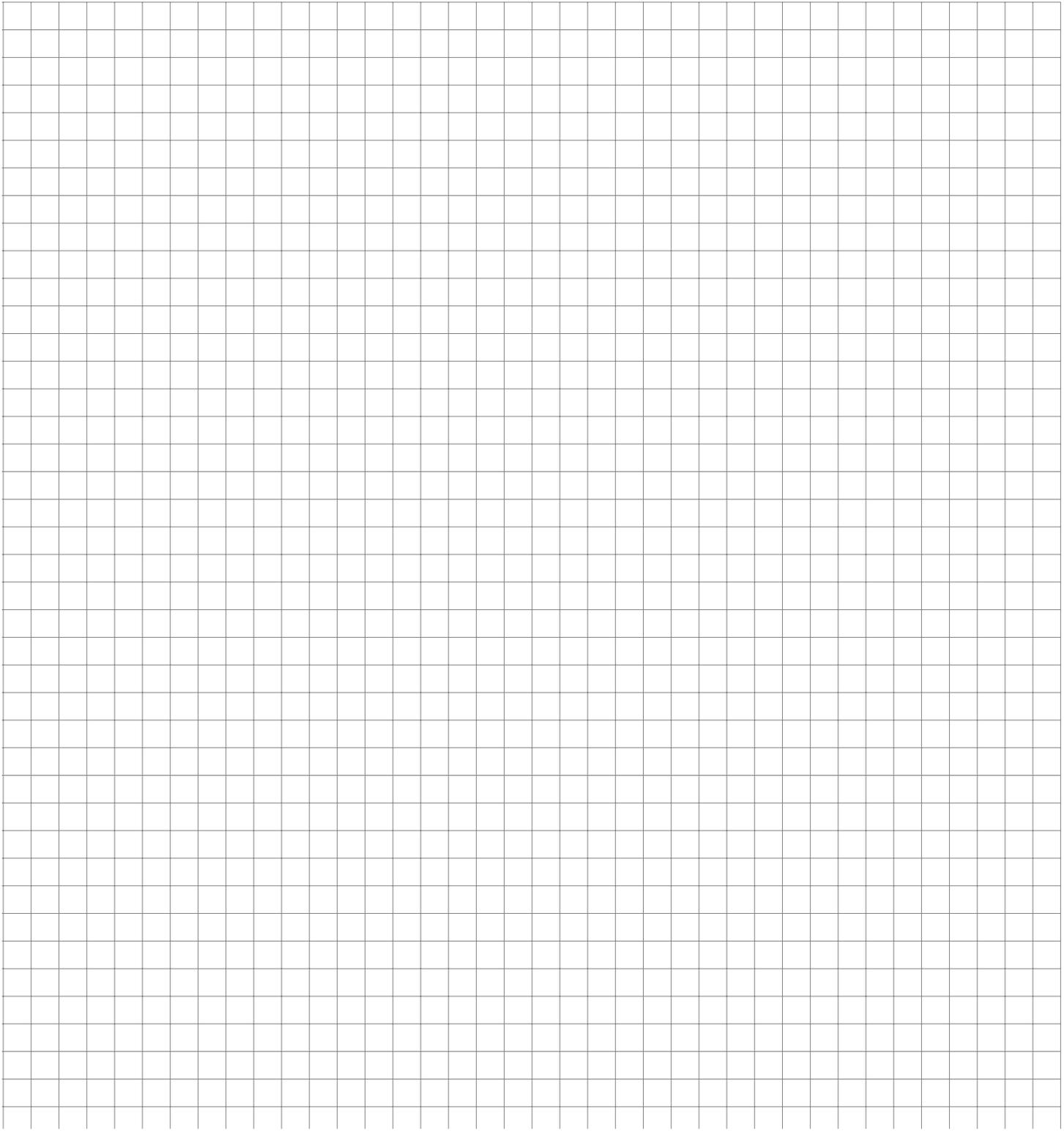
- a) Der Test positiv ausfällt,
- b) die Person krank ist,
- c) die Person gesund ist, obwohl der Test positiv ausgefallen ist,
- d) der Test positiv ausfällt, wenn bekannt ist, dass die Person gesund ist.



Aufgabe 56

Geben Sie zu den Ereignissen A, B die bedingte Wahrscheinlichkeit $P(A|B)$ an, wenn

- a) $A \subset B$,
- b) $B \subset A$,
- c) $A = \Omega$,
- d) $B = \Omega$,
- e) $A \cap B = \{\}$.

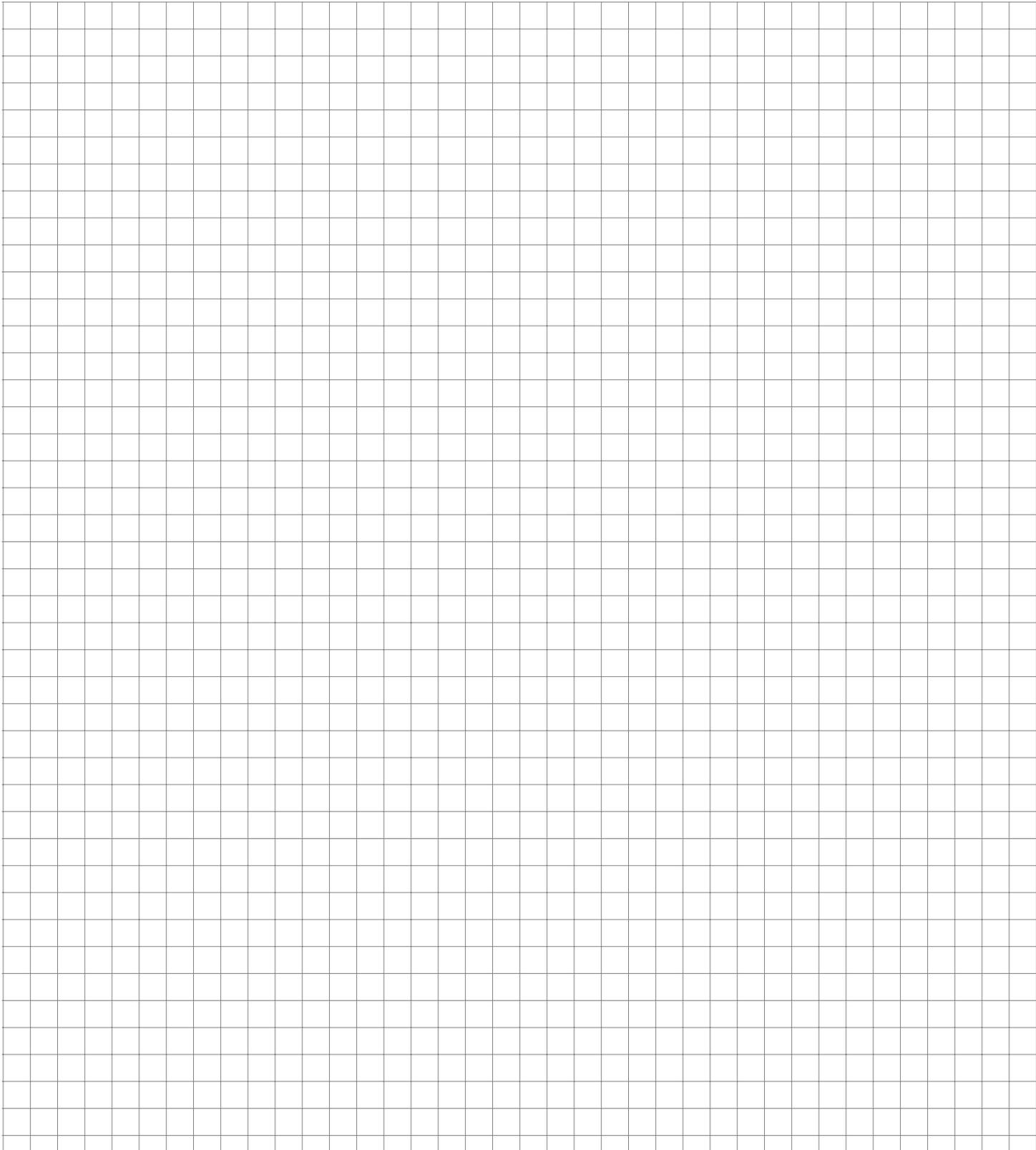


Aufgabe 57

Ein Schießbudenbesitzer hat festgestellt, dass die Trefferwahrscheinlichkeit in den späten Abendstunden 0,1 pro Schuss beträgt.

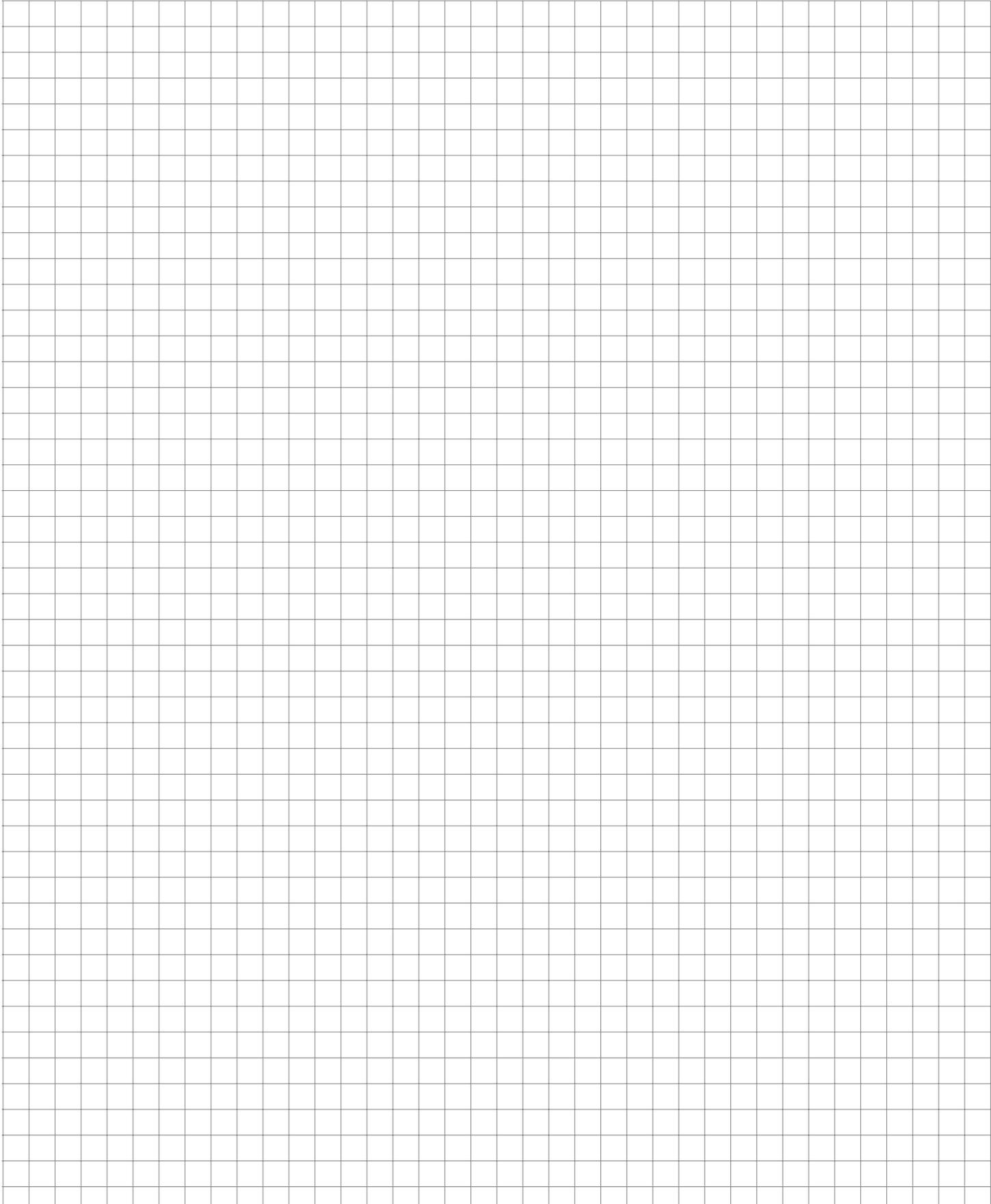
- a) Wie hoch ist die Wahrscheinlichkeit, bei 5 Schüssen mindestens 2 Treffer zu erzielen?
- b) Wie viele Schüsse sind notwendig, um mit einer Wahrscheinlichkeit von mindestens 0,9 mindestens einen Treffer zu erzielen?

R



Aufgabe 58

Eine binomialverteilte Zufallsvariable X habe einen Erwartungswert von 2 und eine Varianz von $4/3$. Wie groß ist die Wahrscheinlichkeit für $x=2$?



Aufgabe 59

In der Klausur zur Statistik werden 25 Multiple-Choice-Fragen gestellt mit jeweils vier Antwortmöglichkeiten, von denen genau eine richtige anzukreuzen ist. Wie wahrscheinlich ist es, mindestens 12 Punkte zu erhalten, wenn man nur rät?

R

A large grid of graph paper, consisting of 25 columns and 30 rows of small squares, intended for the student to perform calculations or show their work for the problem.

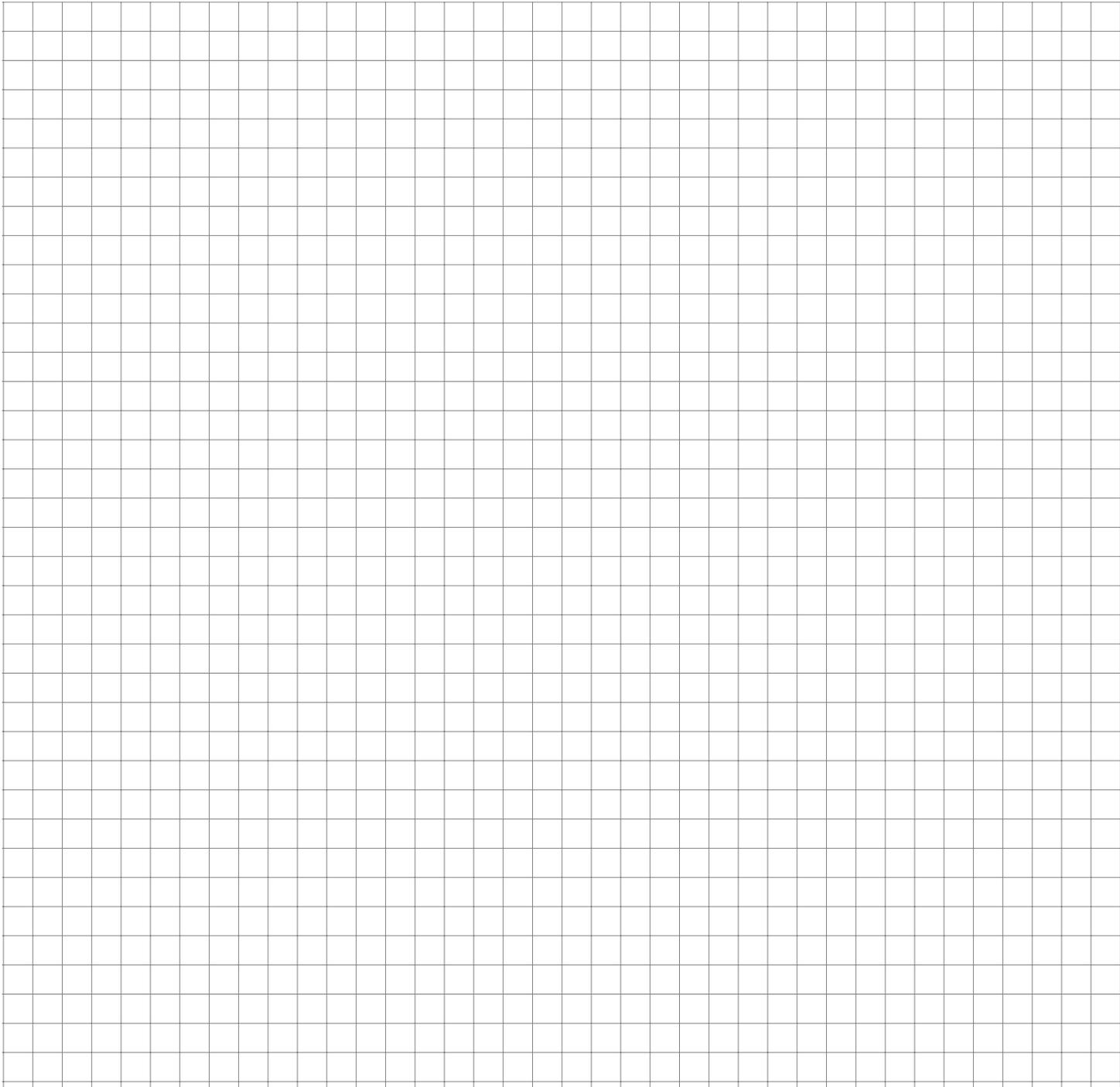
Aufgabe 60

Im Laufe eines Jahres werden von 52 aufeinanderfolgenden Ausgaben einer wöchentlich erscheinenden Zeitschrift 11 beliebige Ausgaben mit einer bestimmten Annonce versehen. Wie groß ist die Wahrscheinlichkeit, dass ein Leser von 20 beliebigen (aber verschiedenen) Ausgaben

R

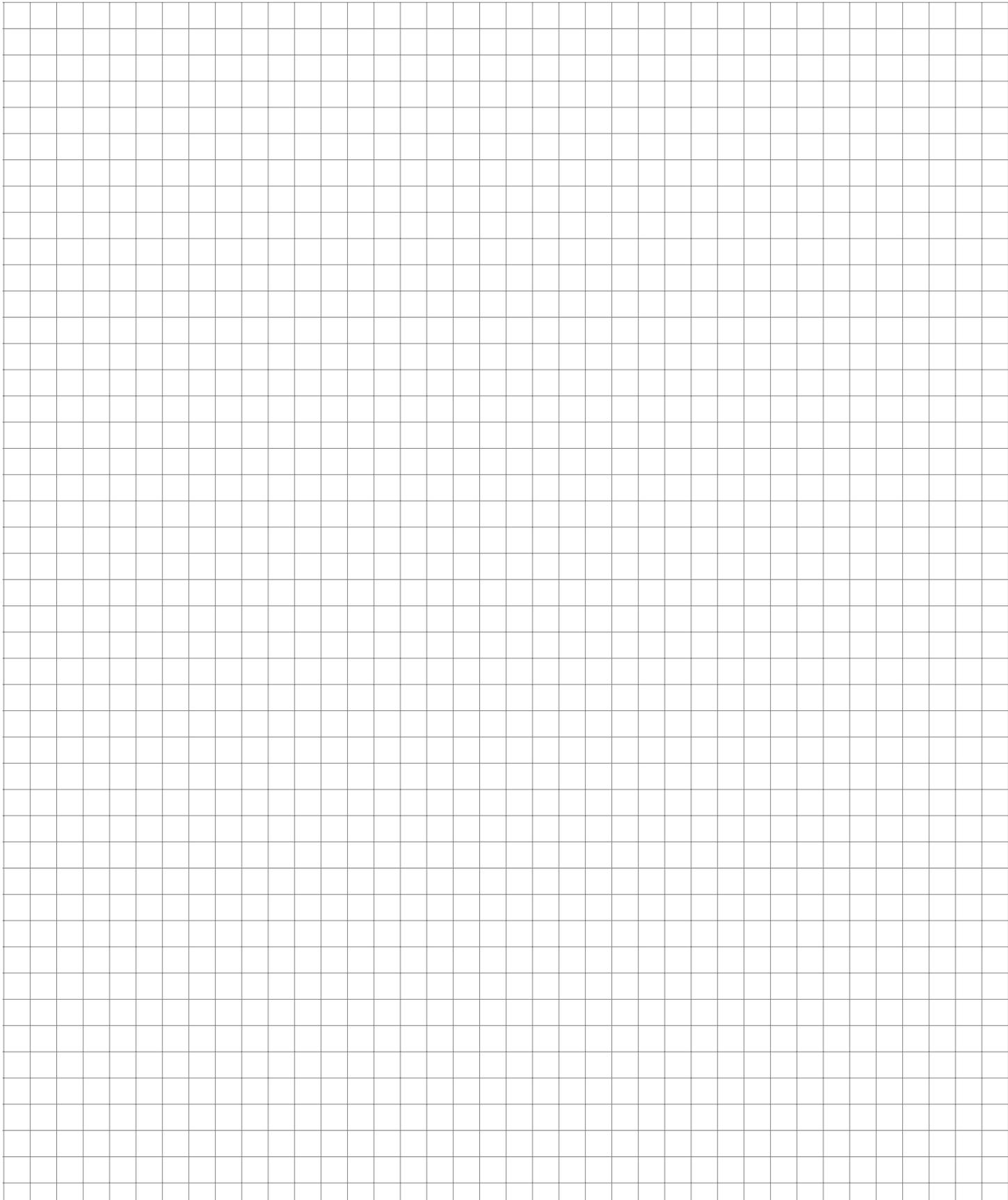
- a) zwei Ausgaben
- b) keine Ausgabe
- c) 20 Ausgaben
- d) sämtliche 11 Ausgaben
- e) mindestens eine Ausgabe

mit einer Annonce erhält?



Aufgabe 61

Unter den 20 Passagieren eines Charterfluges befinden sich zwei Bewaffnete, die das Flugzeug entführen wollen. Zehn Passagiere werden zufällig ausgewählt und genau untersucht. Wie groß ist die Wahrscheinlichkeit dafür, dass die beiden Bewaffneten unentdeckt bleiben?



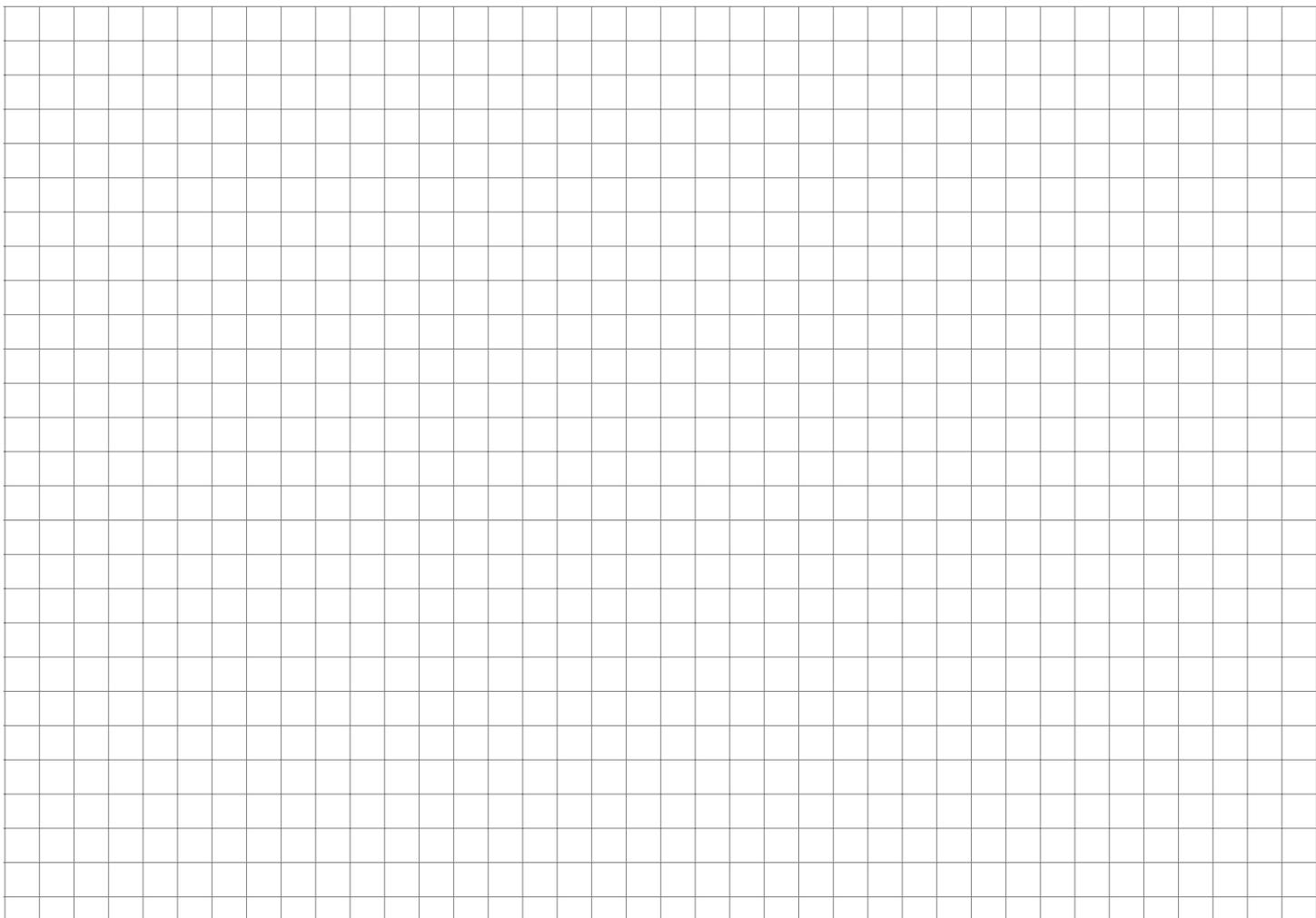
Aufgabe 62

Den drei Studentinnen Anna, Julia und Laura steht eine Statistikklausur bevor. Leider hatten die drei keine Zeit, die Vorlesung zu verfolgen. So verlassen Sie sich auf die Aussage des Dozenten, dass alle 5 Klausuraufgaben zufällig aus einer Liste von 50 veröffentlichten Aufgaben ausgewählt werden.

Die Klausur ist bestanden, wenn mindestens vier Aufgaben richtig gelöst werden. Gehen Sie im Folgenden davon aus, dass eine nicht vorbereitete Aufgabe sicher falsch und eine vorbereitete Aufgabe sicher richtig gelöst wird.

Alle drei wählen eine bestimmte Anzahl von Aufgaben zufällig und unabhängig voneinander aus und bereiten sich auf diese Fragen intensiv vor. Anna bereitet sich auf die Hälfte der Aufgaben vor. Julia geht davon aus, dass es reicht, sich auf vierzig der fünfzig Aufgaben vorzubereiten. Laura möchte nichts dem Zufall überlassen, schafft es aber wegen einer Krankheit nur sich auf 45 Aufgaben vorzubereiten.

- Mit welcher Wahrscheinlichkeit bestehen die drei Kandidatinnen jeweils die Prüfung?
- Wie hoch ist die Wahrscheinlichkeit, dass alle drei bestehen?
- Wie hoch ist die Wahrscheinlichkeit, dass mindestens eine von den dreien besteht?
- Wie hoch ist die Wahrscheinlichkeit, dass Anna in drei Klausurversuchen (mit jeweils den gleichen Konditionen und den gleichen vorbereiteten Aufgaben wie beim ersten Versuch) durchfällt? Wie hoch ist diese Wahrscheinlichkeit bei Laura?

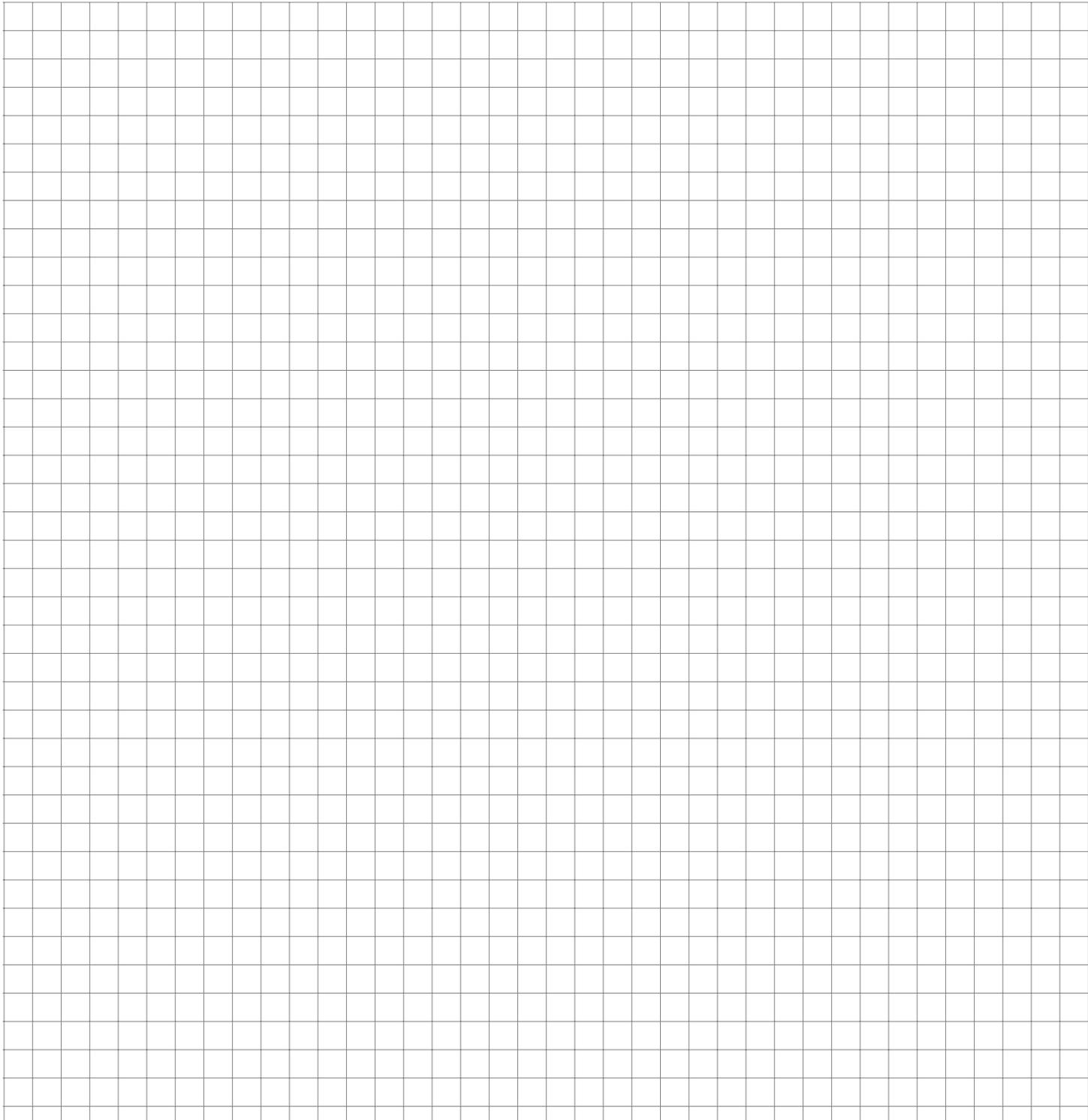


Aufgabe 63

Das Rechenzentrum der Hochschule habe festgestellt, dass während einer Betriebszeit von einem Tag mit der Wahrscheinlichkeit 0,905 kein Ausfall des Systems zu verzeichnen ist. Die Anzahl der Systemausfälle sei Poisson-verteilt.

- Bestimmen Sie den Parameter λ der Poisson-Verteilung.
- Wie groß ist die Wahrscheinlichkeit, dass an einem Tag genau zwei Systemausfälle zu verzeichnen sind?
- Berechnen Sie die Wahrscheinlichkeit, dass es bei 5 gleichartigen Systemen, die unabhängig voneinander laufen, zu mindestens einem Ausfall am Tage kommt.

R



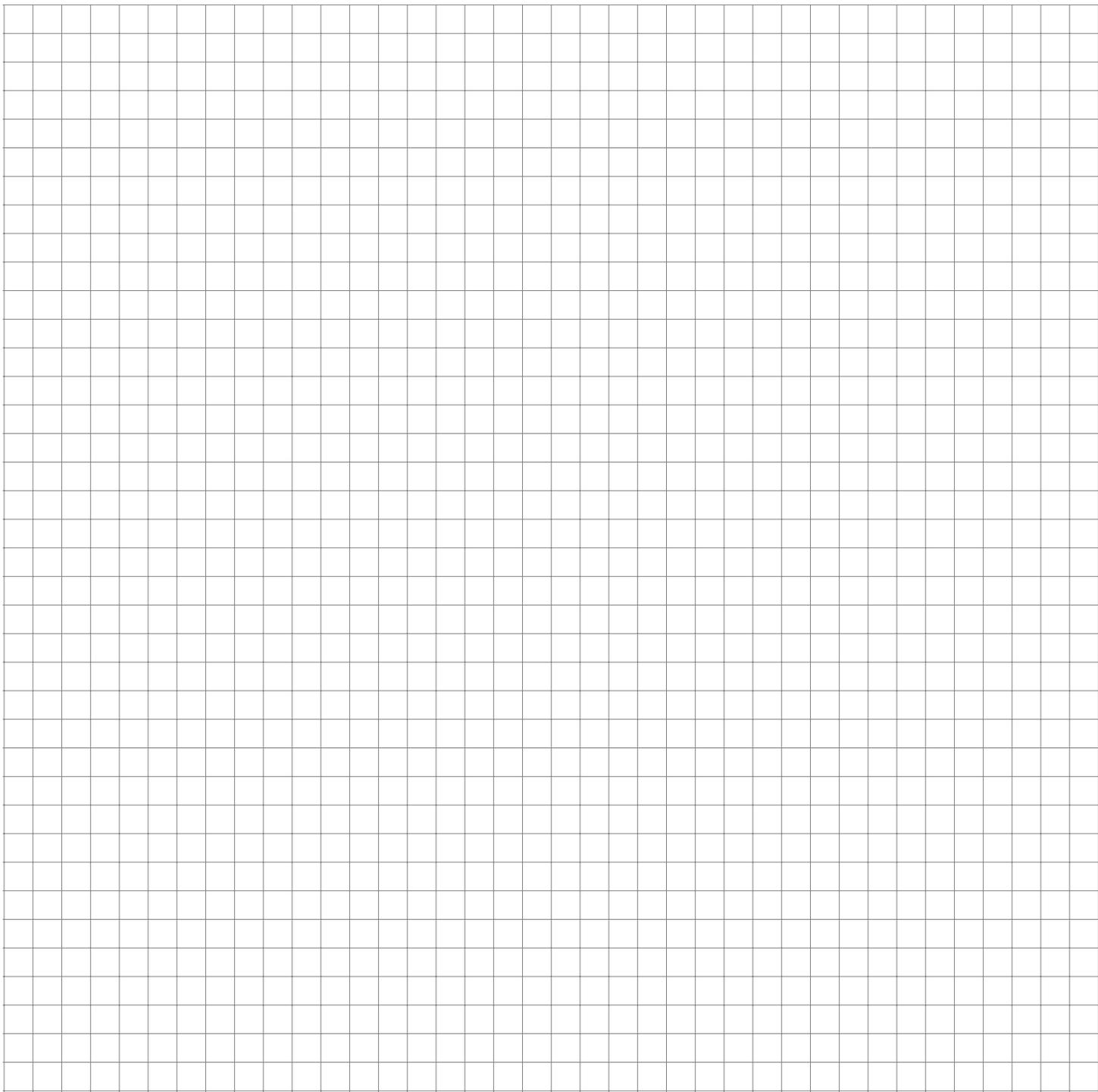
Aufgabe 64

In einer Online-Redaktion weiß man, dass ein Webredakteur gemessen am output sehr wenige sprachliche Fehler produziert. Im Durchschnitt werden drei Fehler pro Monat festgestellt. Die Anzahl der Fehler pro Monat kann als Poisson-verteilt angenommen werden und ist jeweils unabhängig von den anderen Monaten.

Wie groß ist die Wahrscheinlichkeit,

- a) dass der Redakteur mehr als 9 Fehler pro Monat begeht,
- b) für mehr als 3 Fehler, wenn man weiß, dass er schon 2 Fehler gemacht hat,
- c) dass er während eines Jahres in mindestens 3 Monaten keinen Fehler produziert?

R



Aufgabe 65

Die portugiesische Fußballnationalmannschaft schießt pro Spiel durchschnittlich 1 Tor. Gehen Sie davon aus, dass die Anzahl der Tore pro Spiel poissonverteilt ist. Wie groß ist die Wahrscheinlichkeit, dass die Mannschaft

- a) in einem Spiel höchstens 1 Tor erzielt,
- b) in einem Spiel genau 7 Tore erzielt,
- c) während der Gruppenphase einer Fußball-WM (3 Spiele) mindestens einmal 7 Tore schießt?

A large grid of graph paper for calculations, consisting of 20 columns and 30 rows of small squares.

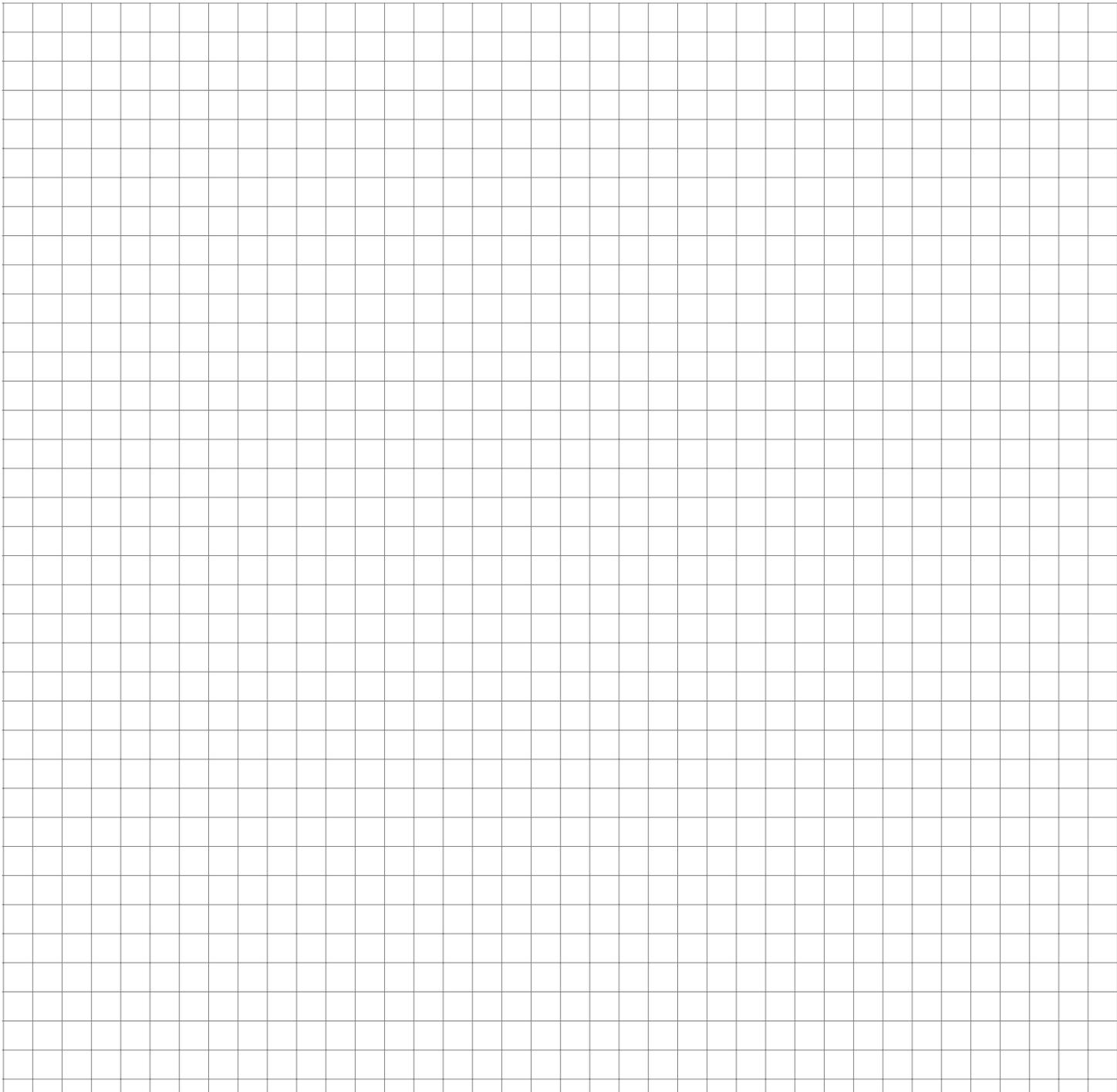
Aufgabe 66

Johanna und Benedikt haben zu Beginn Ihres Studiums geheiratet, sich aber schon vor dem Abschluss wieder scheiden lassen. Im Verlauf der folgenden $n = 40$ Jahre begegnen sich die beiden jedoch häufiger wieder, wobei allerdings die Wahrscheinlichkeit evtl. wieder zu heiraten in jedem Jahr nur $1/200$ beträgt. Wie groß ist dann (bei Unabhängigkeit) die Wahrscheinlichkeit, dass die beiden

- a) nicht wieder heiraten ($x=0$)
- b) noch einmal heiraten ($x=1$)?

Berechnen Sie die beiden Ergebnisse jeweils mit der Binomial- sowie der Poissonverteilung und beurteilen Sie die Abweichungen.

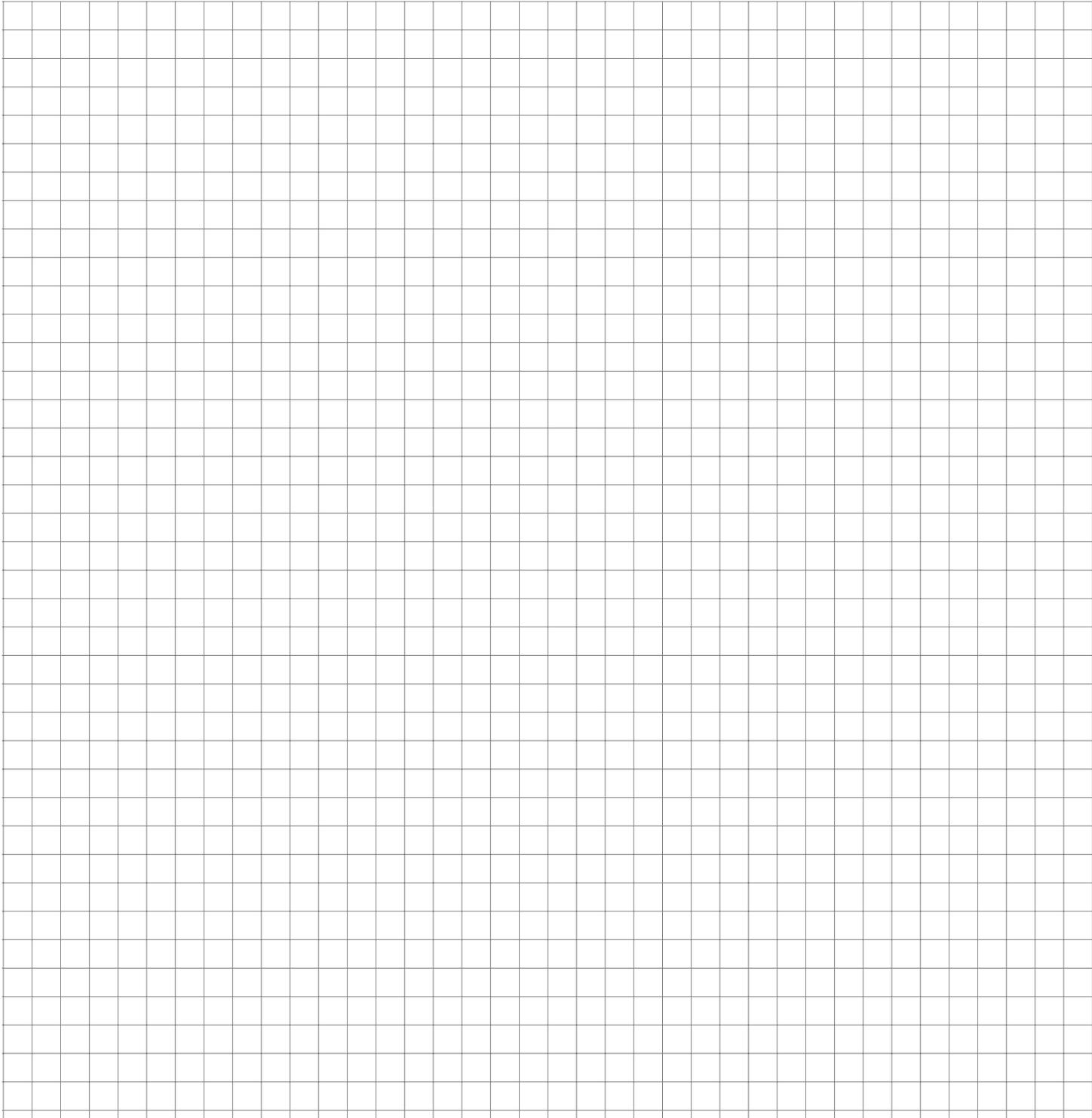
R



Aufgabe 67

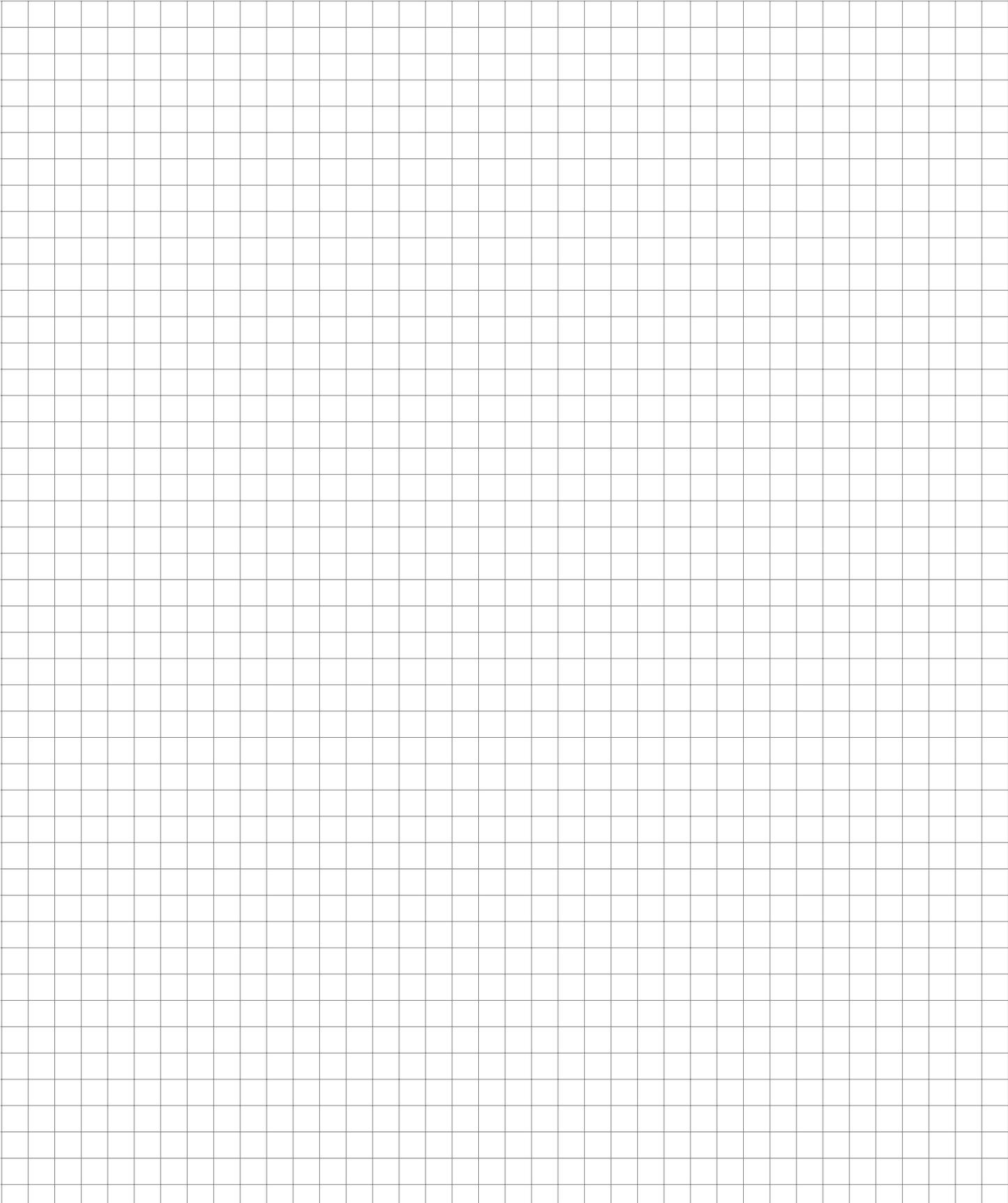
Ein Computerhersteller will eine neue Bestückungsmaschine für Platinen kaufen. Die Ausschussrate soll höchstens 5 % sein. Zur Kontrolle wird ein Probelauf mit 20 Platinen durchgeführt. Sind mehr als k Platinen fehlerhaft bestückt, so muss die Produktion gestoppt und kostenfrei nachgebessert werden.

- Mit welcher Wahrscheinlichkeit erhält man bei einer tatsächlichen Ausschussrate von 5 % höchstens 3 fehlerhafte Platinen?
- Wie muss die Zahl k gewählt werden, damit die Wahrscheinlichkeit für einen Produktionsstopp trotz ausreichender Ausschussrate kleiner als 10 % ist?



Aufgabe 68

Die Gesamtdauer X eines Projektes wird als normalverteilt mit dem Parameter $\mu = 10$ (Wochen) angenommen. Ferner wird für die Wahrscheinlichkeit $P(8 \leq X \leq 12)$ der Wert 0,8 geschätzt. Man bestimme den Parameter σ .

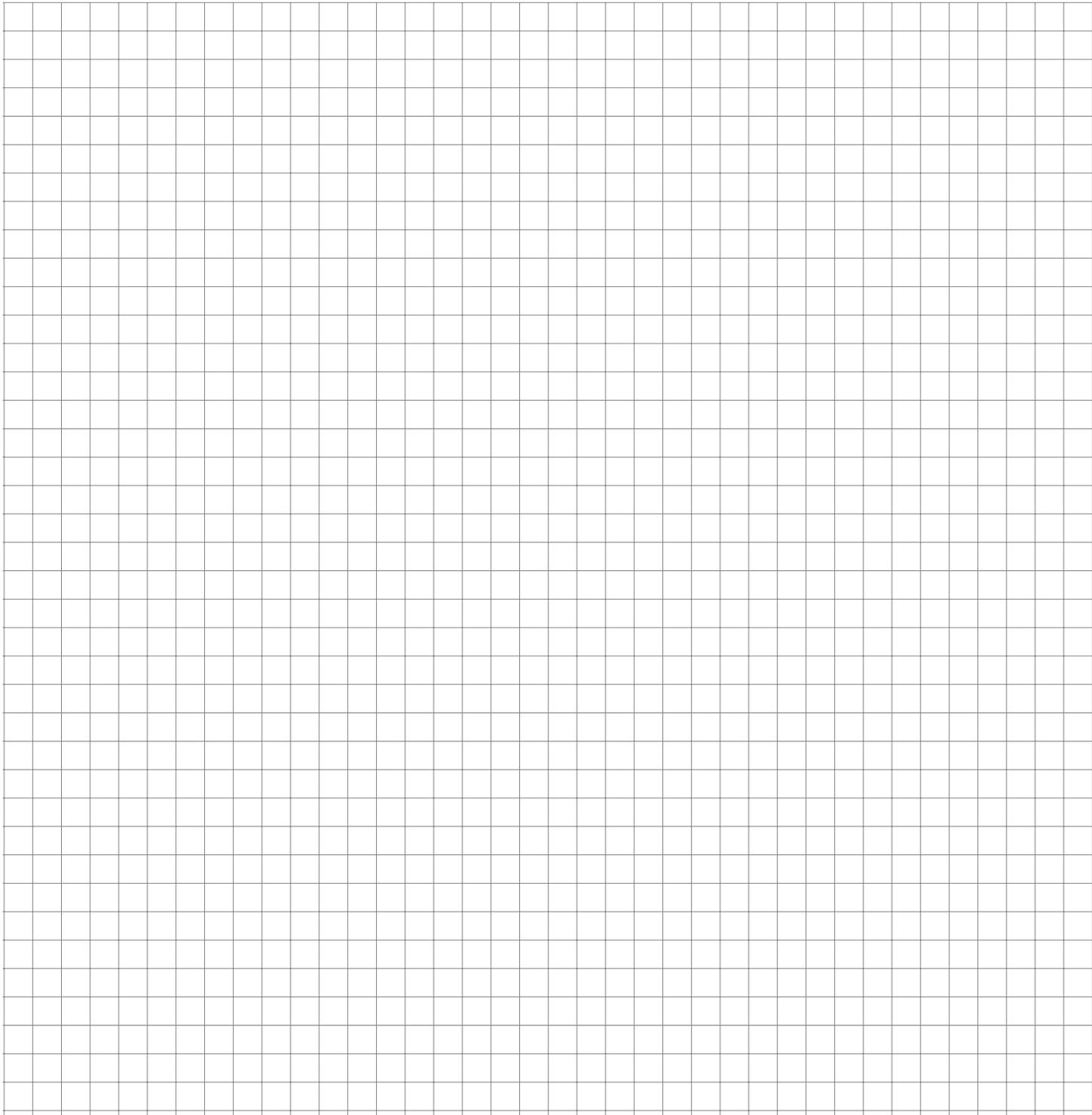


Aufgabe 69

Das Körpergewicht X (in kg) zufällig ausgewählter Personen aus einer Grundgesamtheit sei normalverteilt mit den Parametern μ und σ . Es gilt:

$$P(X \leq 80) = \frac{1}{2} \quad \text{und} \quad P(X \leq 70) = \frac{1}{4}.$$

- Geben Sie μ und σ an.
- Berechnen Sie $P(X \geq 100)$.
- Wieviel Prozent der Personen der Grundgesamtheit, die mindestens 100 kg wiegen, wiegen über 110 kg?



Aufgabe 70

Die Zufallsvariable X beschreibt den täglichen Umsatz in einer Eisdiele. Es wird angenommen, dass $X \sim N(\mu, \sigma)$ gilt. Außerdem sei bekannt, dass der Umsatz an 30,854 % der Tage mindestens 1500 € und an 30,854 % der Tage weniger als 900 € beträgt.

- Bestimmen Sie μ und σ .
- Wie groß ist die Wahrscheinlichkeit, dass der Umsatz an einem Tag mehr als 3000 € beträgt?
- Wie hoch müsste der Umsatz mindestens gemäß der Verteilung an den 5 % besten Tagen sein?
- Wie groß ist die Wahrscheinlichkeit, dass der Umsatz negativ ist? Was bedeutet das?



Aufgabe 71

Eine normalverteilte Zufallsvariable X soll untersucht werden. Zwei Tatsachen sind von X bekannt:

- ▶ $P(X > 20) = 20\%$
- ▶ $P(X < 1) = 1\%$

Bestimmen Sie damit:

- a) $\text{Sta}[X] = \sigma$ sowie $E[X] = \mu$
- b) $P(X = 20)$
- c) $P(X \leq 20)$
- d) $P(X \geq 25)$
- e) $P(X \geq 25 | X \geq 20)$
- f) $P(X \geq 20 | X \geq 25)$



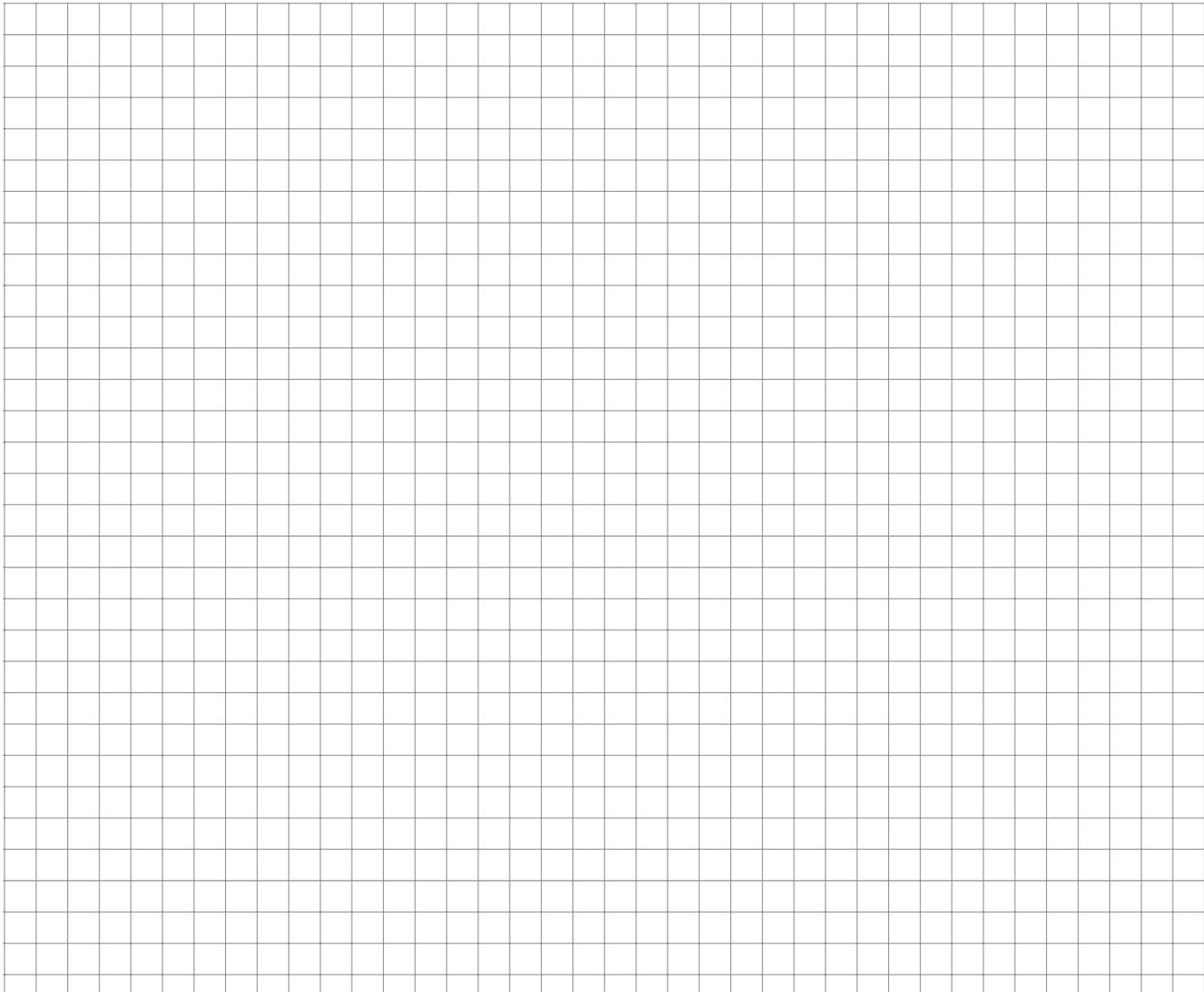
Aufgabe 72

Schokoladennikoläuse mit einem Sollgewicht von 200g sollen bzgl. ihres Gewichts kontrolliert werden. Es stellt sich heraus, dass

- ▶ das Gewicht X der Nikoläuse normalverteilt ist,
- ▶ die Wahrscheinlichkeit, dass ein Nikolaus mindestens 200g wiegt bei 30 % liegt und
- ▶ ein Nikolaus mit einer Wahrscheinlichkeit von 99 % höchstens 210g wiegt.

Berechnen Sie bzw. geben Sie ohne Rechnung aber mit Begründung an:

- Wie groß ist die Standardabweichung σ sowie der Erwartungswert μ von X ?
- Wie groß ist die Wahrscheinlichkeit, zufällig einen Nikolaus mit einem Gewicht von exakt 200g (± 0 g) auszuwählen?
- Wie groß ist die Wahrscheinlichkeit, dass ein Nikolaus weniger als 190g wiegt?
- Nikoläuse mit weniger als 195g werden aussortiert. Wie groß ist die Wahrscheinlichkeit, dass ein Nikolaus aus diesem Ausschuss zwischen 190g und 195g wiegt?

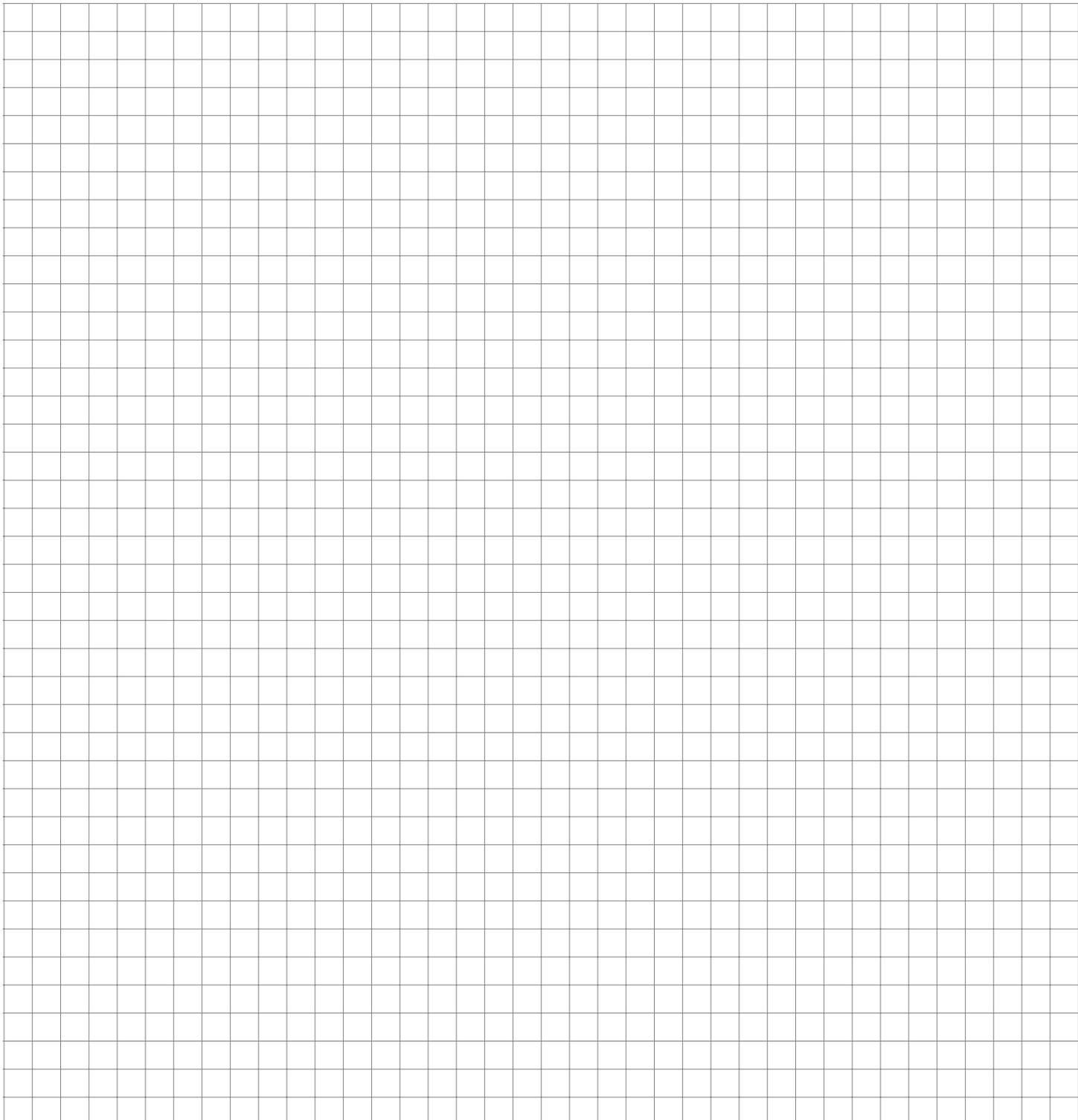


Aufgabe 73

Die zufallsabhängige Nachfrage X nach einem Gut in einer Zeitperiode ist gemäß der folgenden Wahrscheinlichkeitsfunktion verteilt:

$$P(x = n) = \frac{n}{40} \quad \text{für } n = 1, \dots, 8 \quad \text{und} \quad P(x = 9) = \frac{1}{10}$$

- Skizzieren Sie den Verlauf der Wahrscheinlichkeitsfunktion.
- Mit welcher Wahrscheinlichkeit werden mindestens 6 Stück des Gutes nachgefragt?
- Berechnen Sie den Erwartungswert und die Standardabweichung der Nachfrage.

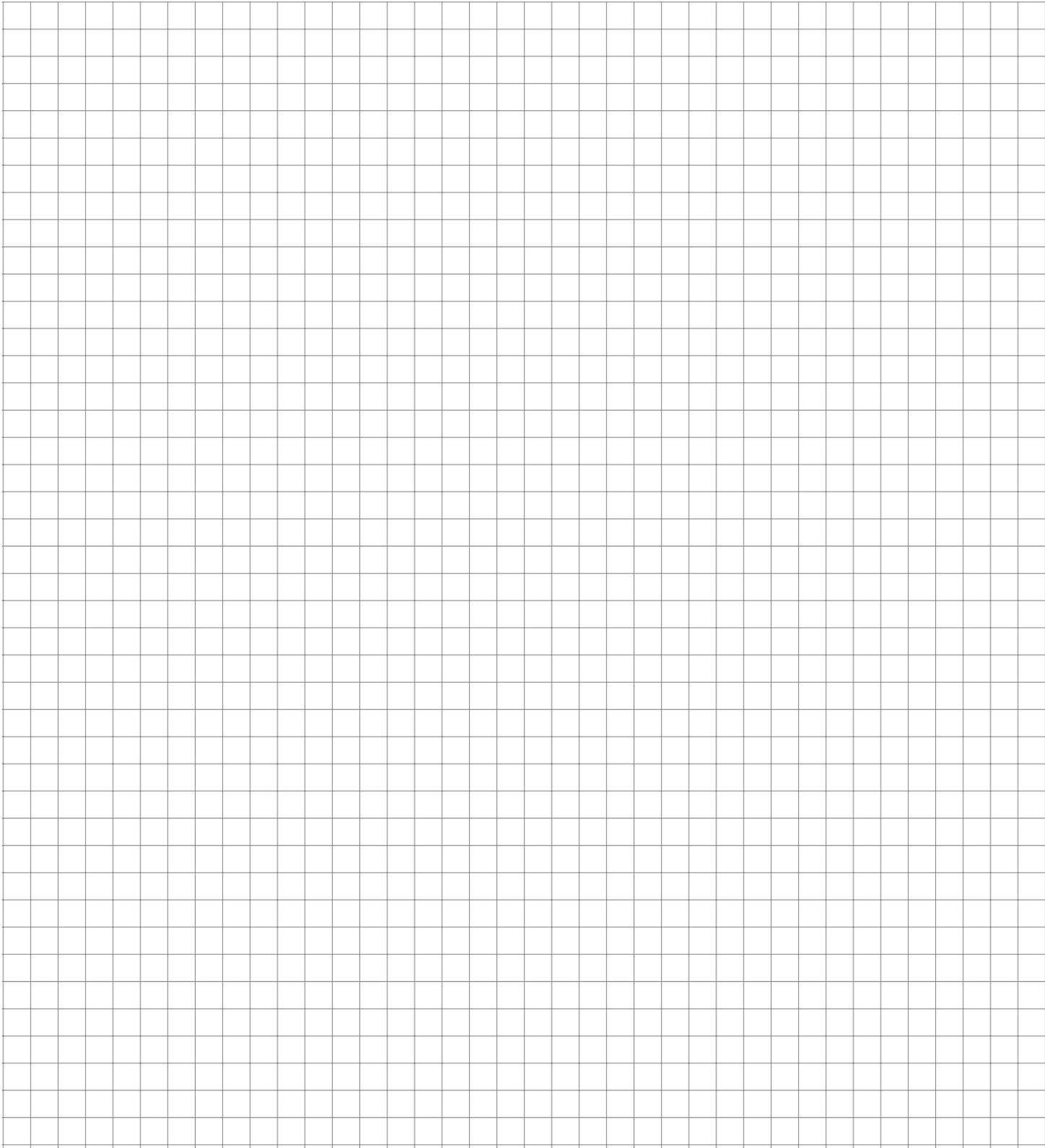


Aufgabe 74

WTheorie: Erwartungswert Varianz (20)

Die Lebensdauer einer Maschine sei eine über dem Zeitintervall $[0,65]$ gleichverteilte Zufallsvariable. Berechnen Sie

- den Erwartungswert der Lebensdauer
- die Varianz der Lebensdauer
- die Wahrscheinlichkeit, dass die Lebensdauer zwischen 13 und 39 liegt.

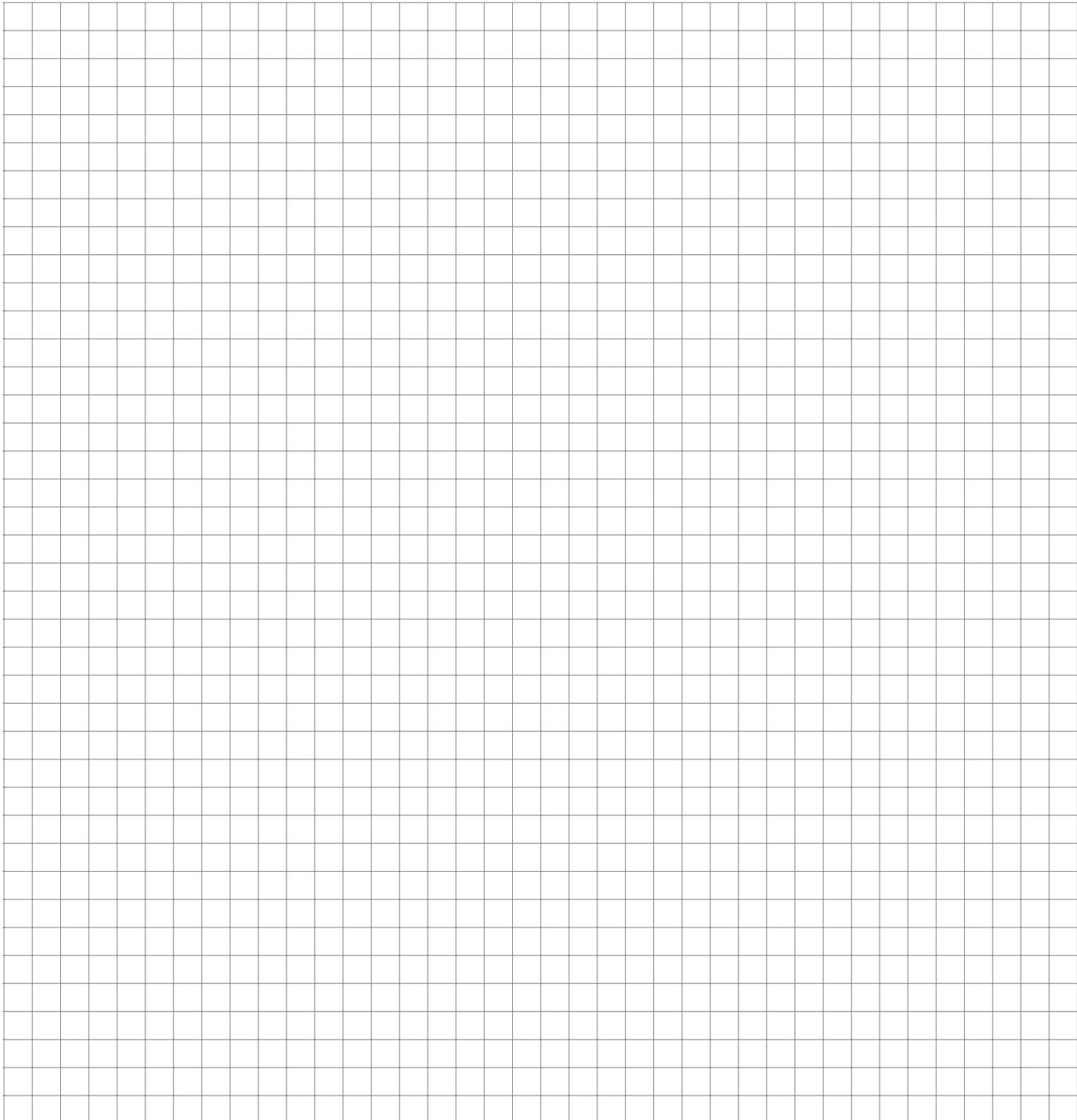


Aufgabe 75

Eine Unternehmung sieht sich auf dem Absatzmarkt zufällig schwankender Nachfrage gegenübergestellt. Die Höhe der Nachfrage X sei folgendermaßen verteilt:

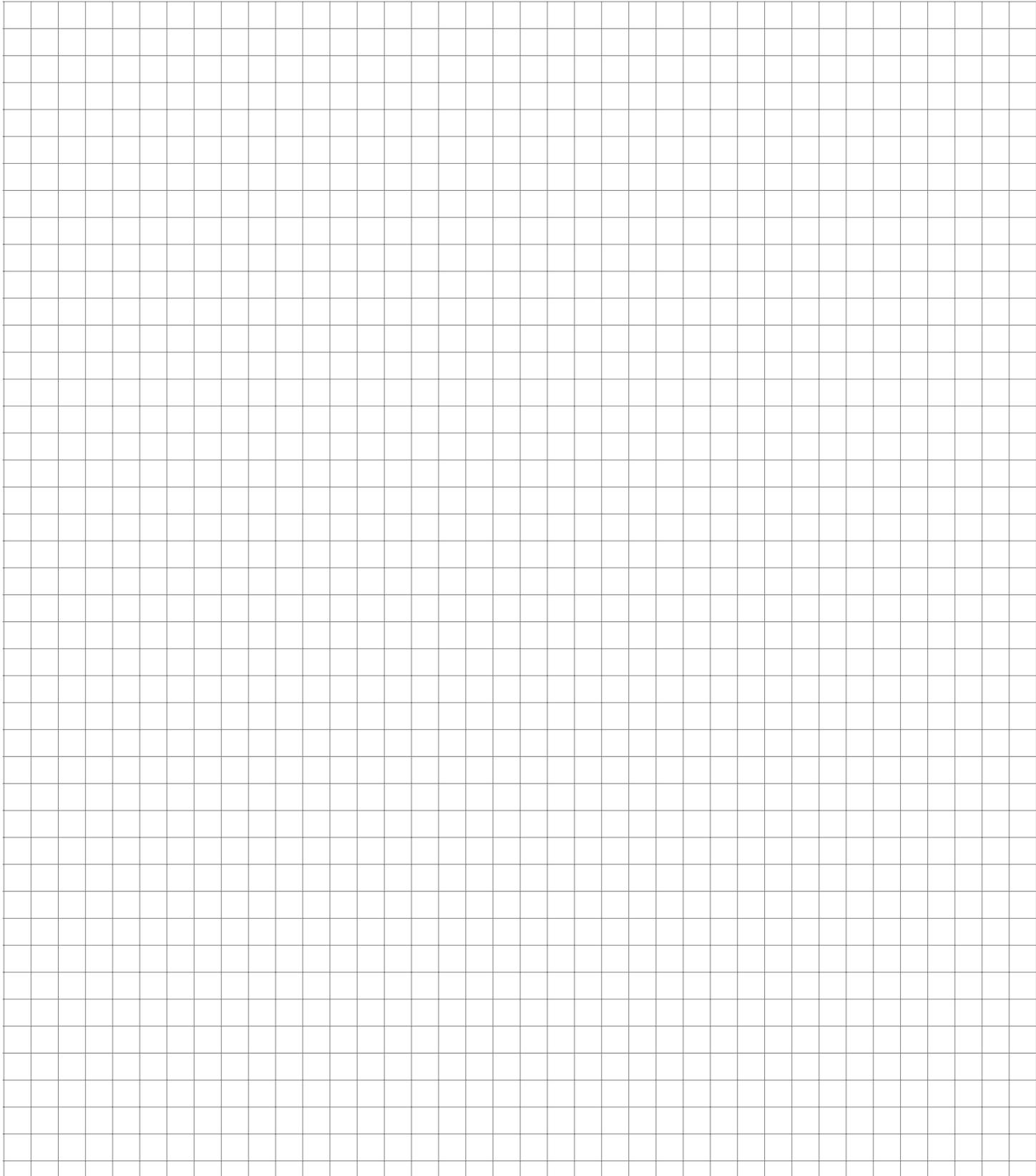
$$f(x) = \begin{cases} \frac{1}{12}, & \text{für } 0 \leq x \leq 12 \\ 0, & \text{sonst} \end{cases}$$

Die Produktion der Unternehmung wird unmittelbar abgesetzt, d.h. es existieren keine Absatzlager. Die Kostenfunktion der Unternehmung lautet $Y = 2X + 10$. Man gebe den Erwartungswert und die Varianz der Kosten an.



Aufgabe 76

Ein Röhrenwerk produziert Stahlröhren, deren Durchmesser produktionsbedingten Schwankungen unterliegen. Für den Innendurchmesser X_1 hat man $E(X_1) = 800$ und $\text{Var}(X_1) = 0,01$ und für den Außendurchmesser X_2 hat man $E(X_2) = 810$ und $\text{Var}(X_2) = 0,02$. Bestimmen Sie Erwartungswert und Varianz für die Wandstärke der Röhren, wenn angenommen werden kann, dass Innen- und Außendurchmesser voneinander unabhängig schwanken.

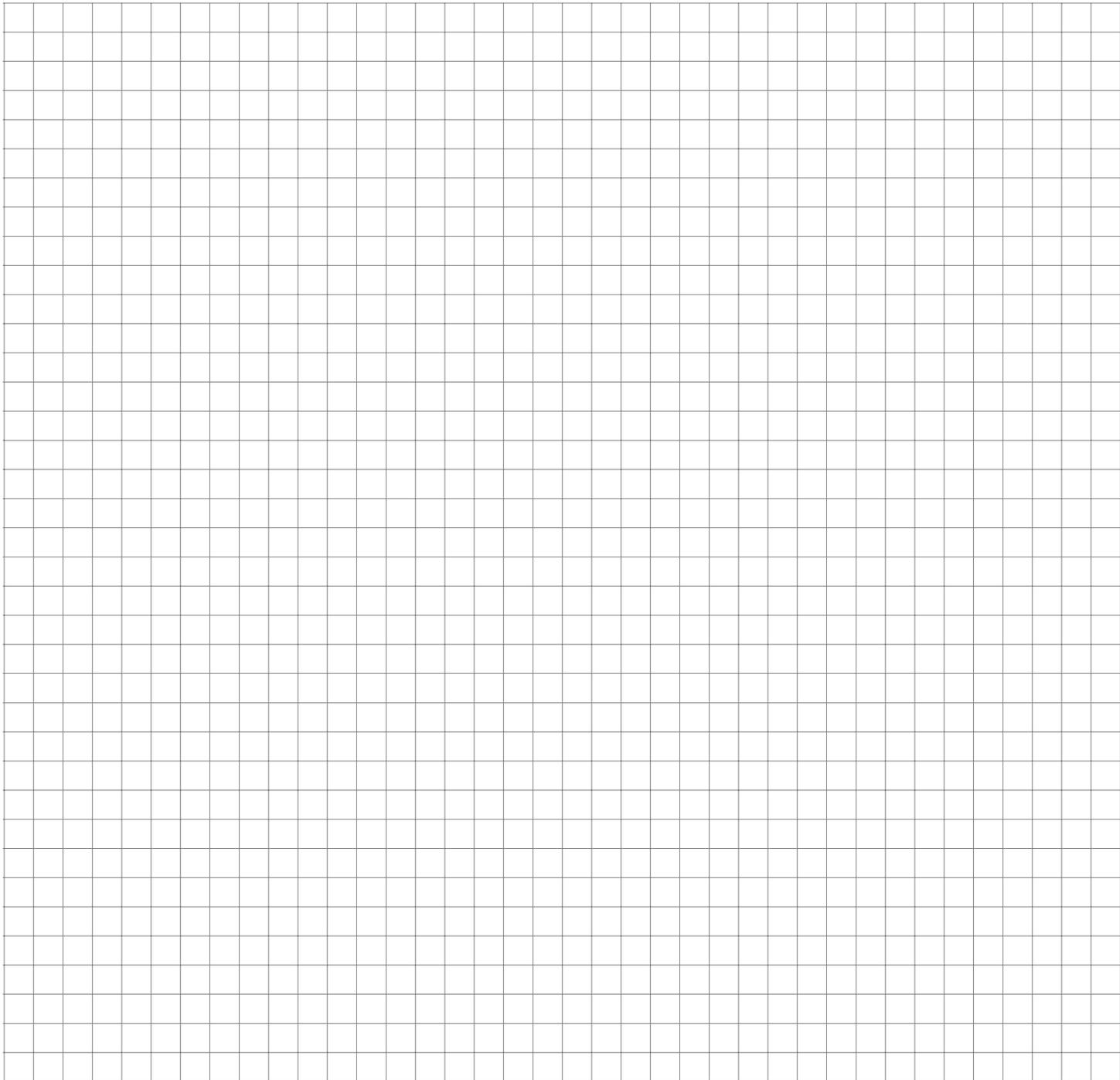


Aufgabe 77

Der Besitzer eines Zeitschriftenladens hat für einen längeren Zeitraum in der Vergangenheit folgende tägliche Nachfrageverteilung nach einer bestimmten Tageszeitung beobachtet:

pro Tag nachgefragte Exemplare	0	1	2	3	4	> 4
Nachfragewahrscheinlichkeit	0,20	0,30	0,20	0,20	0,10	0

Er rechnet für die Zukunft mit keiner Änderung der Nachfrageverteilung. Der Einkaufspreis eines Exemplars beträgt 0,50 €, der Verkaufspreis 1,50 €. Unverkaufte Exemplare können nicht zurückgegeben werden. Für einen längeren Zeitraum muss er eine feste Zahl von Zeitungen pro Tag bestellen. Wie viele Zeitungen pro Tag sollte er bestellen, um seinen erwarteten Gewinn zu maximieren?



Aufgabe 78

Die Versicherungsgesellschaft *Ollionz* verlangt als Prämie das 1,3-fache des Erwartungswertes ihrer Zahlungen an den Versicherungsnehmer. Es werden einjährige Lebensversicherungen des folgenden Typs betrachtet:

Der Betrag von 100.000 € ist zu zahlen, wenn der Versicherungsnehmer innerhalb eines Jahres nach Abschluss des Vertrages stirbt. Im Erlebensfall ist keine Zahlung zu leisten.

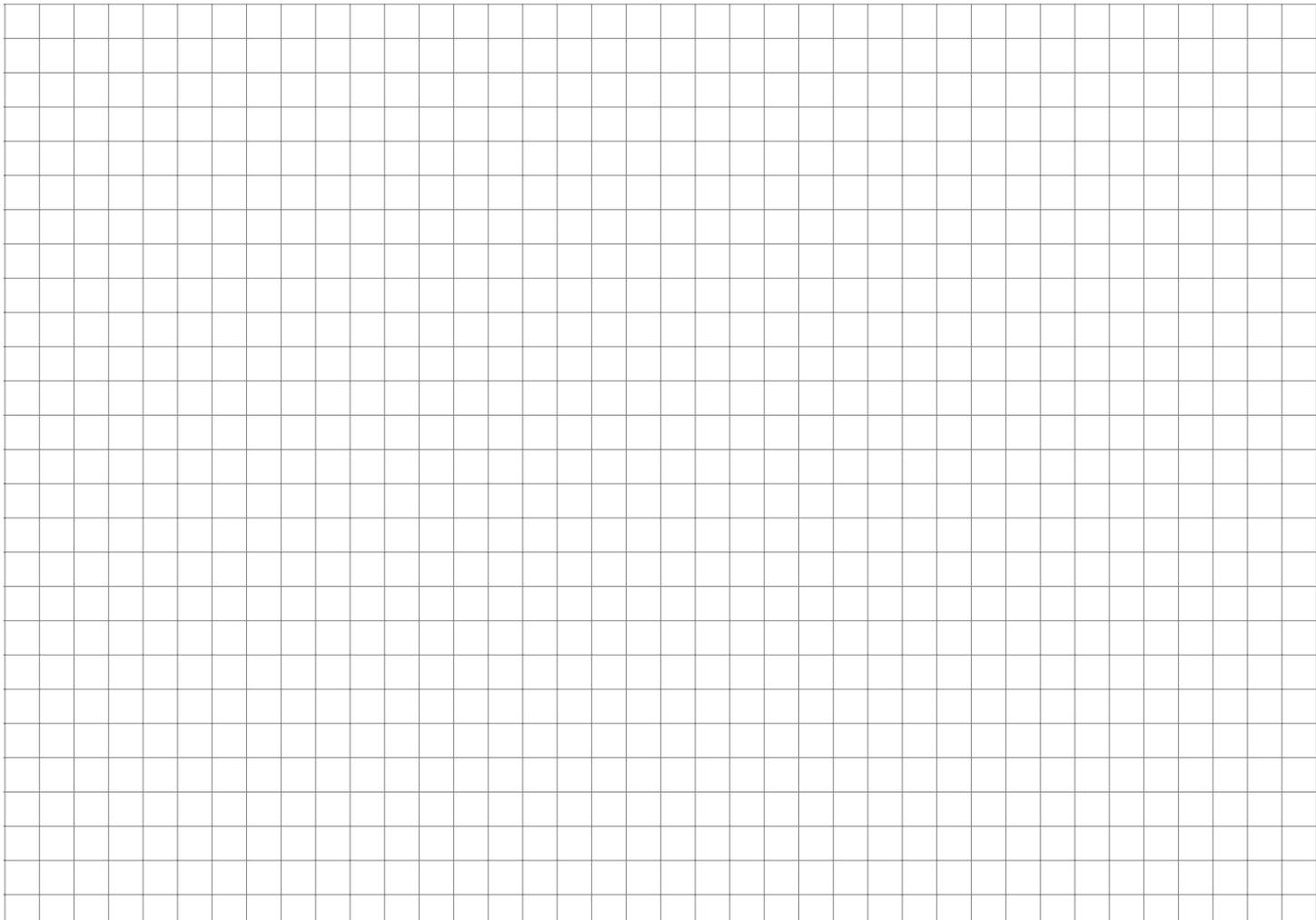
Die einjährige Sterbewahrscheinlichkeit betrage 0,006.

- Wie hoch ist die Prämie für einen derartigen Vertrag?
- Betrachten Sie den aus einem solchen Vertrag resultierenden Gewinn G . Wie hoch sind der Erwartungswert $E[G]$ und die Varianz $\text{Var}[G]$ des Gewinns?
- Berechnen Sie die Standardabweichung σ sowie das Intervall $[\mu - 3\sigma; \mu + 3\sigma]$ (den Drei-Sigma-Bereich). Interpretieren Sie dieses Intervall.
- Die Versicherungsgesellschaft schließt 50.000 derartige Verträge ab. Bestimmen Sie den Drei-Sigma-Bereich des durchschnittlichen Gewinns pro Vertrag

$$\bar{G} = \frac{1}{50.000} [G_1 + G_2 + \dots + G_{50.000}] ,$$

der aus den Verträgen resultiert und interpretieren Sie diese Zahl.

Hinweis: Setzen Sie die Unabhängigkeit der Einzelgewinne G_i voraus.

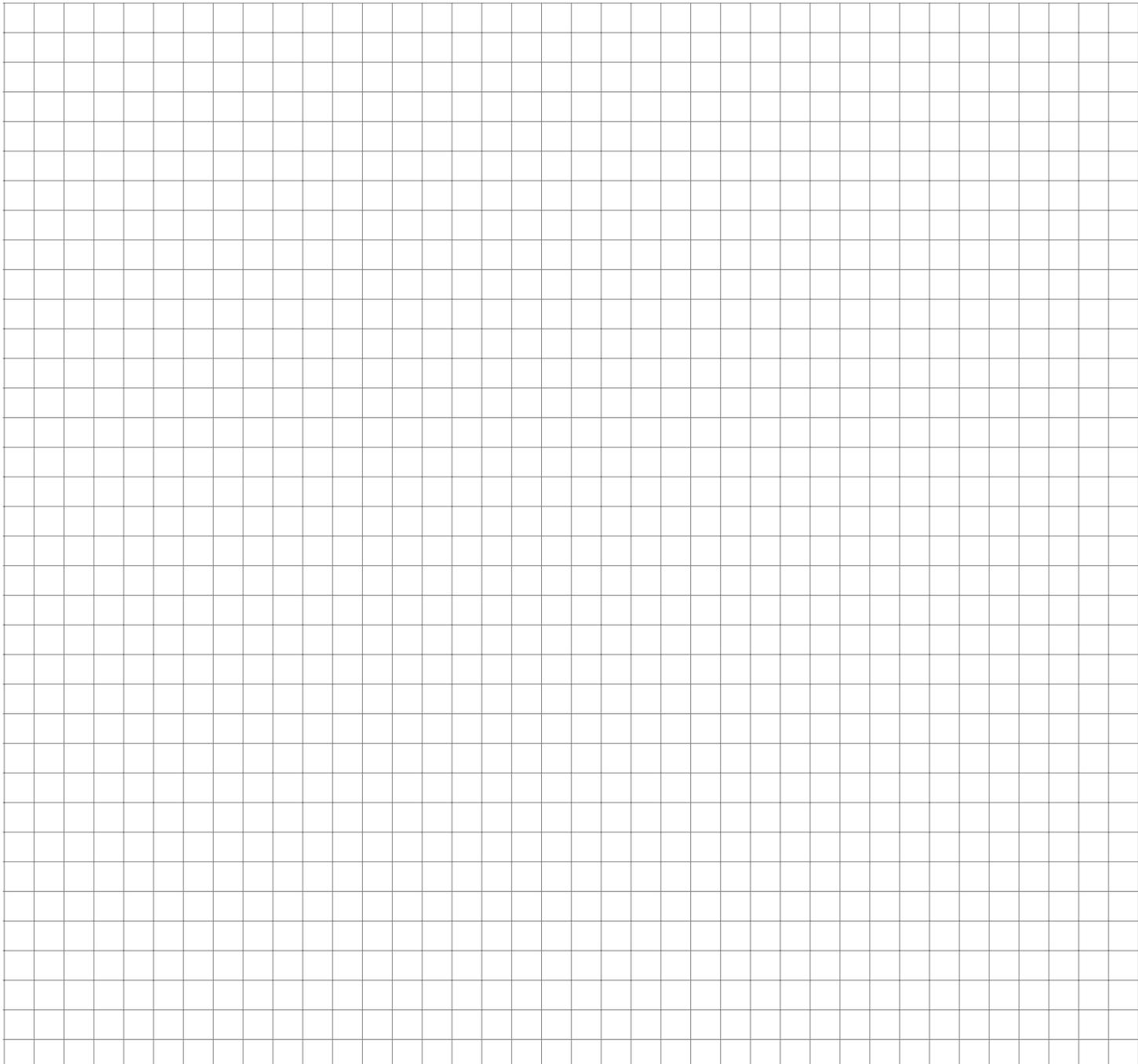


Aufgabe 79

Gegeben ist zur Zufallsvariablen X die Dichtefunktion f gemäß

$$f(x) = \begin{cases} \frac{1}{x \cdot \ln 2} & \text{für } 1 \leq x \leq 2 \\ 0 & \text{sonst} \end{cases}$$

- Skizzieren Sie den Graph von f .
- Bestimmen Sie zu X die Verteilungsfunktion $F(x) = P(X \leq x)$ und skizzieren Sie auch deren Graph.
- Berechnen Sie $P(1.2 \leq X \leq 1.8)$.
- Berechnen Sie $E[X]$ sowie
- $\text{Var}[X]$.

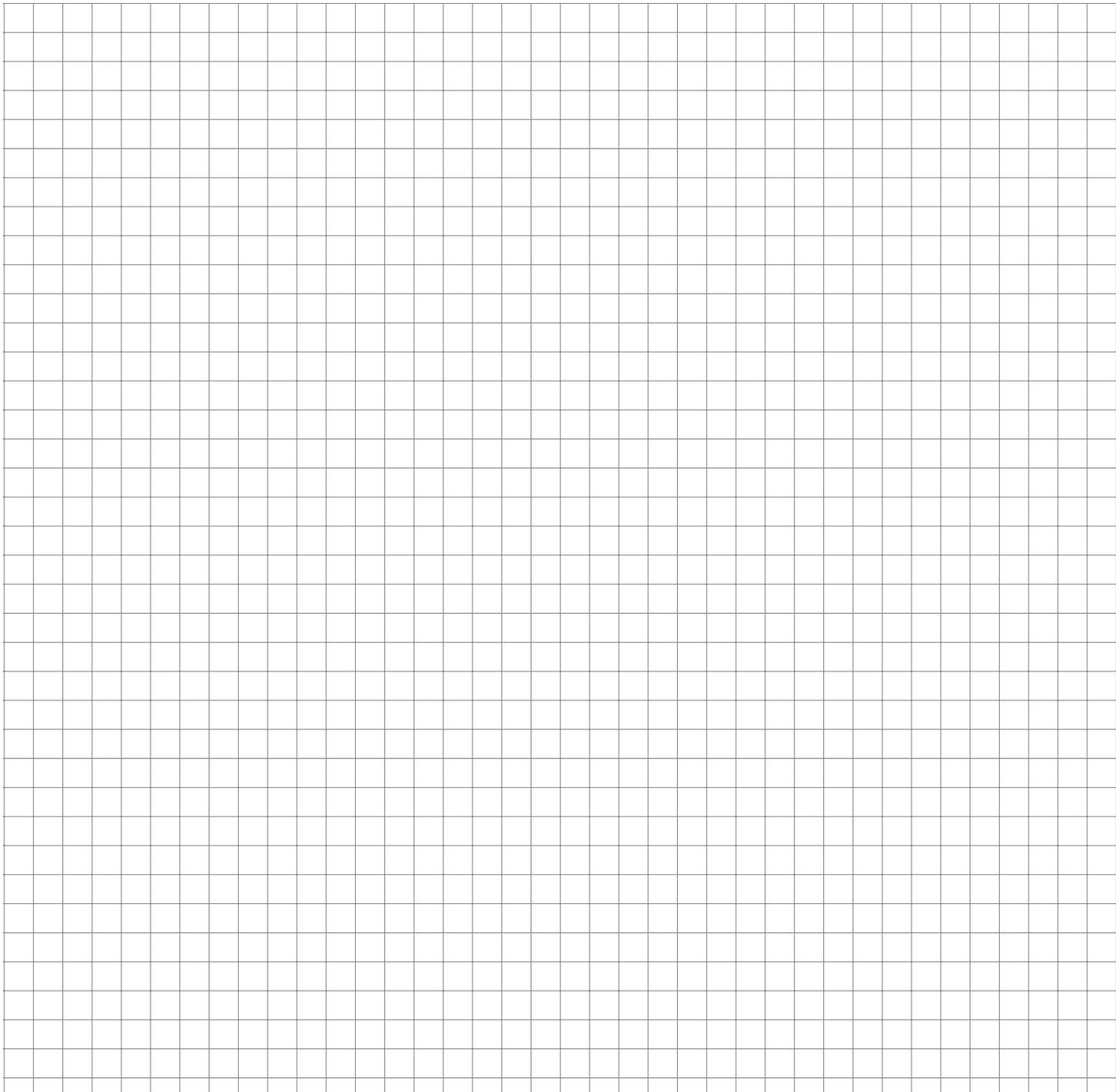


Aufgabe 80

Gegeben ist eine stetige Zufallsvariable X mit der zugehörigen Dichtefunktion $f : \mathbb{R} \rightarrow \mathbb{R}$:

$$f(x) = \begin{cases} a \cdot \frac{1}{10}x & \text{für } 0 \leq x \leq 10 \\ a \left(2 - \frac{1}{10}x\right) & \text{für } 10 < x \leq 20 \\ 0 & \text{sonst} \end{cases}$$

- Welchen Wert muss a haben, so dass f eine Dichtefunktion ist.
- Berechnen Sie $P(X \leq 12)$.
- Bestimmen Sie zu X die Verteilungsfunktion $F(x)$.
- Benutzen Sie F , um $P(3 \leq X \leq 25)$ zu ermitteln.



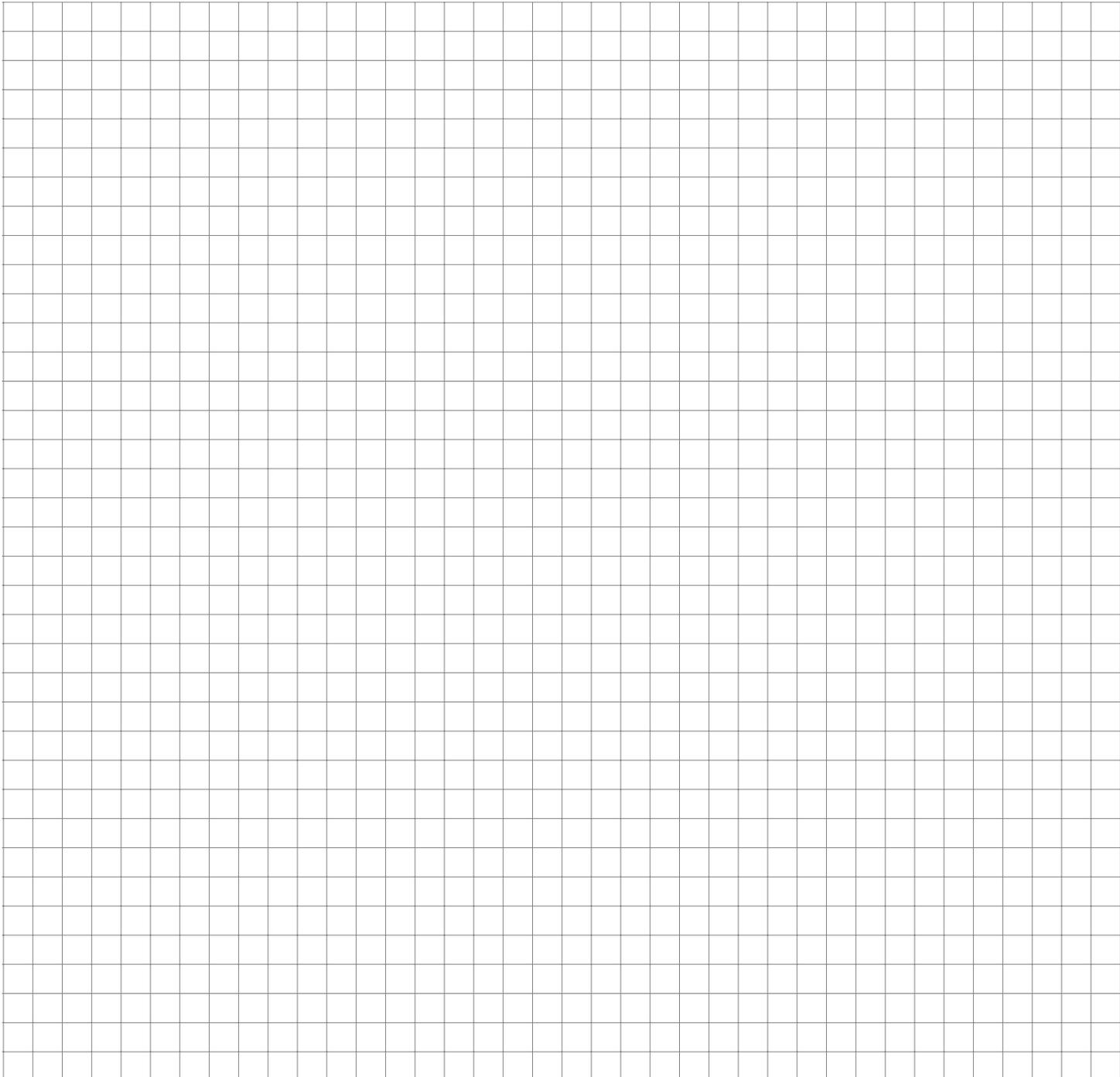
Aufgabe 81

Für eine Zufallsvariable X sei die Wahrscheinlichkeit $P(|X - 10| \geq 2,5)$ mit p^* abgekürzt.

Berechnen Sie p^* für die Fälle, daß X

- im Intervall $[5; 15]$ gleichverteilt ist
- einer $N(\mu; \sigma)$ -Verteilung mit $\mu = 10$ und $\sigma^2 = 4$ genügt
- die Werte 3, 6, 9, 12, 15 jeweils mit der Wahrscheinlichkeit 0,2 annimmt
- Poisson-verteilt ist mit dem Parameter $\lambda = 10$.

Empfehlung: Veranschaulichen Sie den Bereich $\{x \in \mathbb{R} : |x - 10| \geq 2,5\}$ auf einer Zahlengeraden.



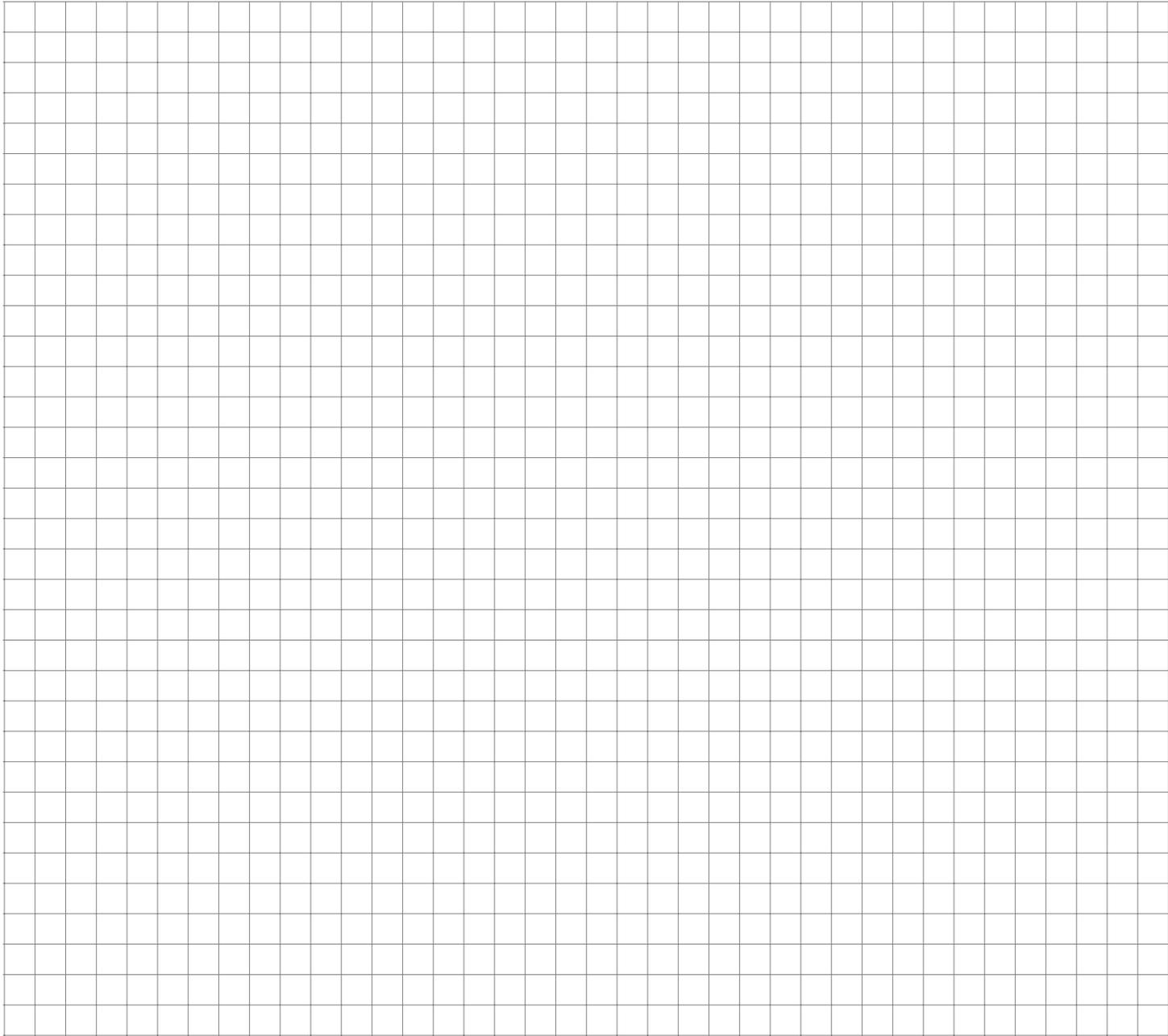
Aufgabe 82

Die Zufallsvariable X beschreibt die Anzahl der möglichen Tore der deutschen Nationalmannschaft bei Länderspielen. Es kommen nur die folgenden Ergebnisse mit den angegebenen Wahrscheinlichkeiten vor:

x	0	1	2	3	4
$P(X = x)$	0,4	0,5	0,07	0,025	0,005

Wie groß ist der

- Erwartungswert von X ,
- die Varianz und die Standardabweichung von X und
- $P(|X - E[X]| \leq \text{Sta}[X])$?



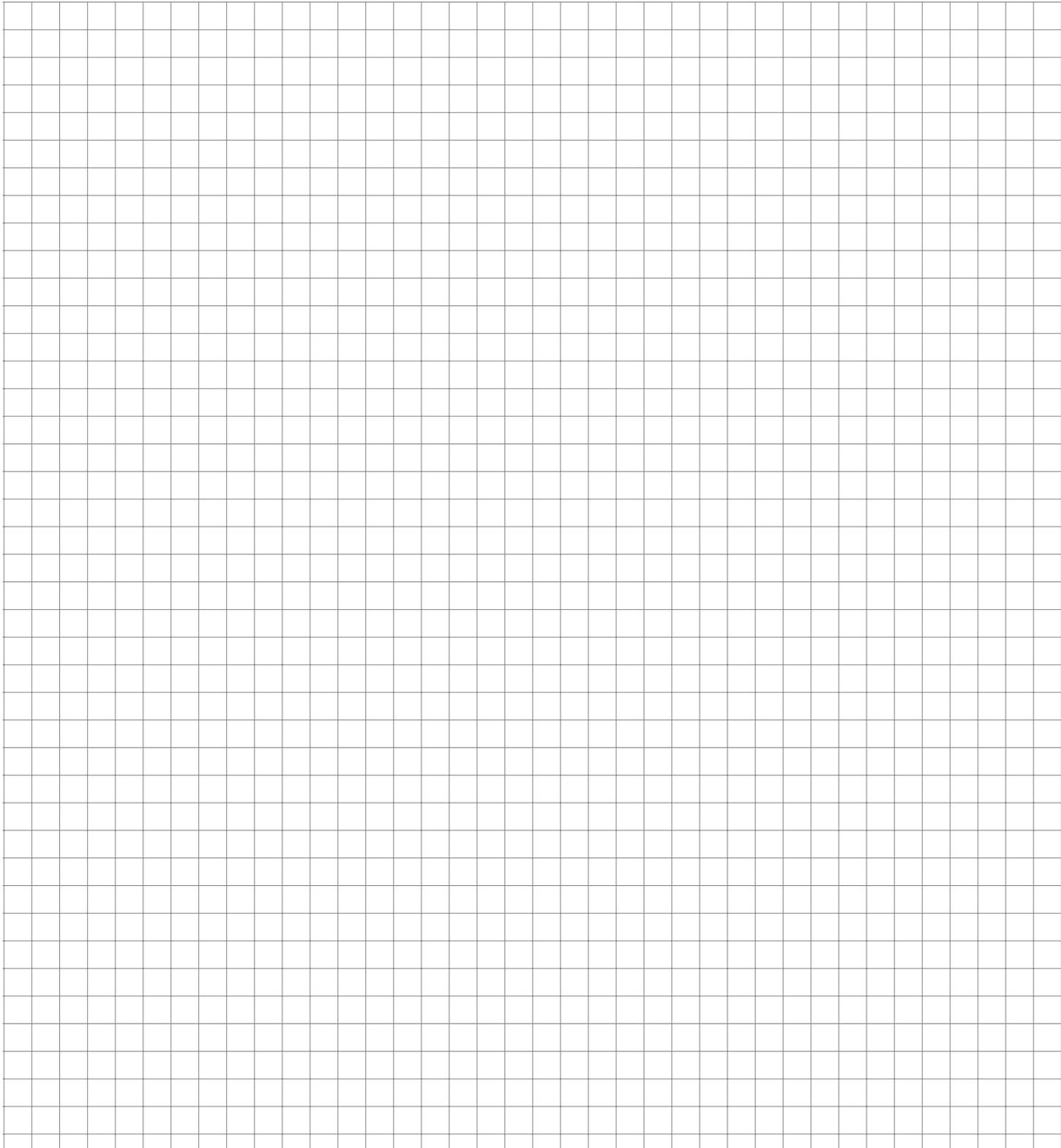
Aufgabe 83

WTheorie: Kovarianz (24)

Die Zufallsvariablen X und Y haben die nebenstehende gemeinsame Wahrscheinlichkeitsfunktion. Man berechne

	y	0	1	2
x				
1		0,1	0,2	0,3
2		0,2	0	0,2

- die Randverteilungen, Erwartungswerte und Varianzen für X und Y ,
- die Kovarianz und Korrelation zwischen X und Y .



Aufgabe 84

Man betrachte die Klausurergebnisse in Mathematik (Zufallsvariable X) und Statistik (Zufallsvariable Y). Die gemeinsame Wahrscheinlichkeitsfunktion der zweidimensionalen Zufallsvariablen (X, Y) sei durch die folgende Tabelle gegeben:

$x \downarrow y \rightarrow$	1	2	3	4	5
1	0,04	0,03	0,02	0,01	0,00
2	0,04	0,10	0,03	0,02	0,01
3	0,02	0,08	0,20	0,08	0,02
4	0,01	0,02	0,04	0,10	0,03
5	0,00	0,01	0,03	0,03	0,03

- a) Bestimmen Sie die Wahrscheinlichkeiten
 - (i) in Mathematik zu bestehen und in Statistik nicht zu bestehen
 - (ii) in beiden Klausuren zu bestehen
 - (iii) in beiden Klausuren nicht zu bestehen
 - (iv) in beiden Klausuren besser als 3 zu erhalten
 - (v) in beiden Klausuren zwischen 2 und 4 zu erreichen.
- b) Geben Sie die Randwahrscheinlichkeits- und Randverteilungsfunktionen an.
- c) Sind die beiden Zufallsvariablen unabhängig?
- d) In Mathematik erzielen Beate, Peter, Helga und Bernd die Noten 1, 2, 3 und 4. Wie sehen für diese vier Kandidaten die Notenchancen bei der Statistiklausur aus?



Aufgaben zur induktiven Statistik

Aufgabe 85

Induktiv: Punktschätzer (1)

Der Erwartungswert μ in der Grundgesamtheit soll durch die Stichprobenfunktion

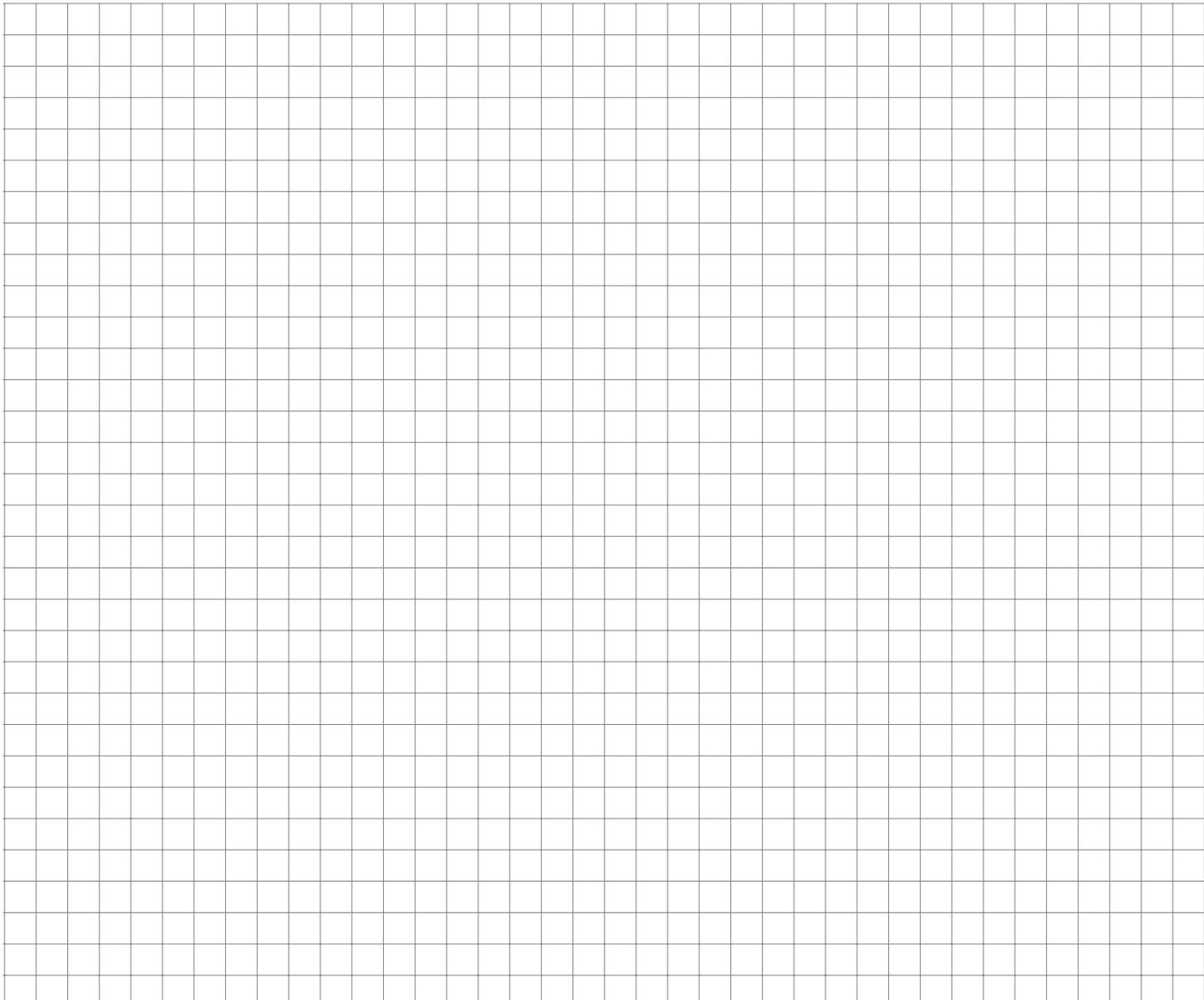
$$\hat{\Theta} = \frac{1}{n} (X_1 + \dots + X_n)$$

$$\hat{\Theta}' = \frac{2}{n^2} \sum_{i=1}^n i X_i$$

$$\hat{\Theta}'' = \frac{2}{n(n+1)} (X_1 + 2X_2 + \dots + nX_n)$$

geschätzt werden. Prüfen Sie die Erwartungstreue der Schätzfunktionen und ermitteln Sie die wirksamste unter diesen.

Hinweise: Es gilt $\sum_{i=1}^n i = \frac{1}{2} n(n+1)$ und $\sum_{i=1}^n i^2 = \frac{1}{3} n(n+1)(n + \frac{1}{2})$



Aufgabe 86

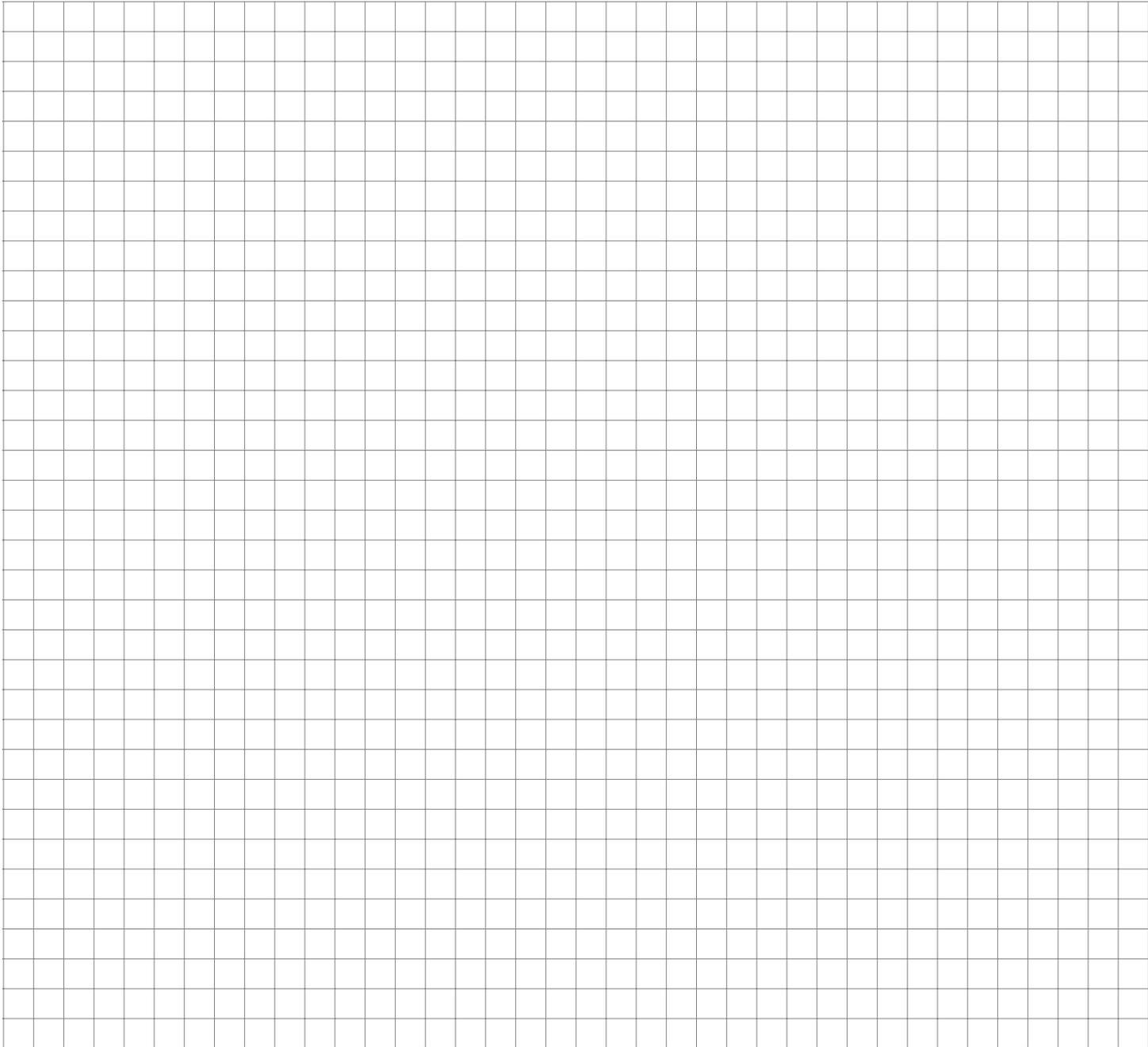
Induktiv: Punktschätzer (Punktschaetzer)

Gegeben sei eine einfache Stichprobe X_1, \dots, X_n aus einer Grundgesamtheit mit unbekanntem Erwartungswert μ . Zur Schätzung von μ wird der Einsatz von linearen Stichprobenfunktionen der folgenden Gestalt erwogen:

$$\hat{\theta}_1 = \frac{1}{n-1} \sum_{i=1}^{n-1} X_i + \alpha \cdot X_n \quad \text{mit} \quad -1 \leq \alpha \leq 1$$

- Bestimmen Sie α so, dass $\hat{\theta}_1$ erwartungstreu für μ ist.
- Welche Schätzfunktion ist wirksamer: Die erwartungstreue Variante von $\hat{\theta}_1$ von Teilaufgabe a) oder eine neue Funktion $\hat{\theta}_2$ mit

$$\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n X_i ?$$



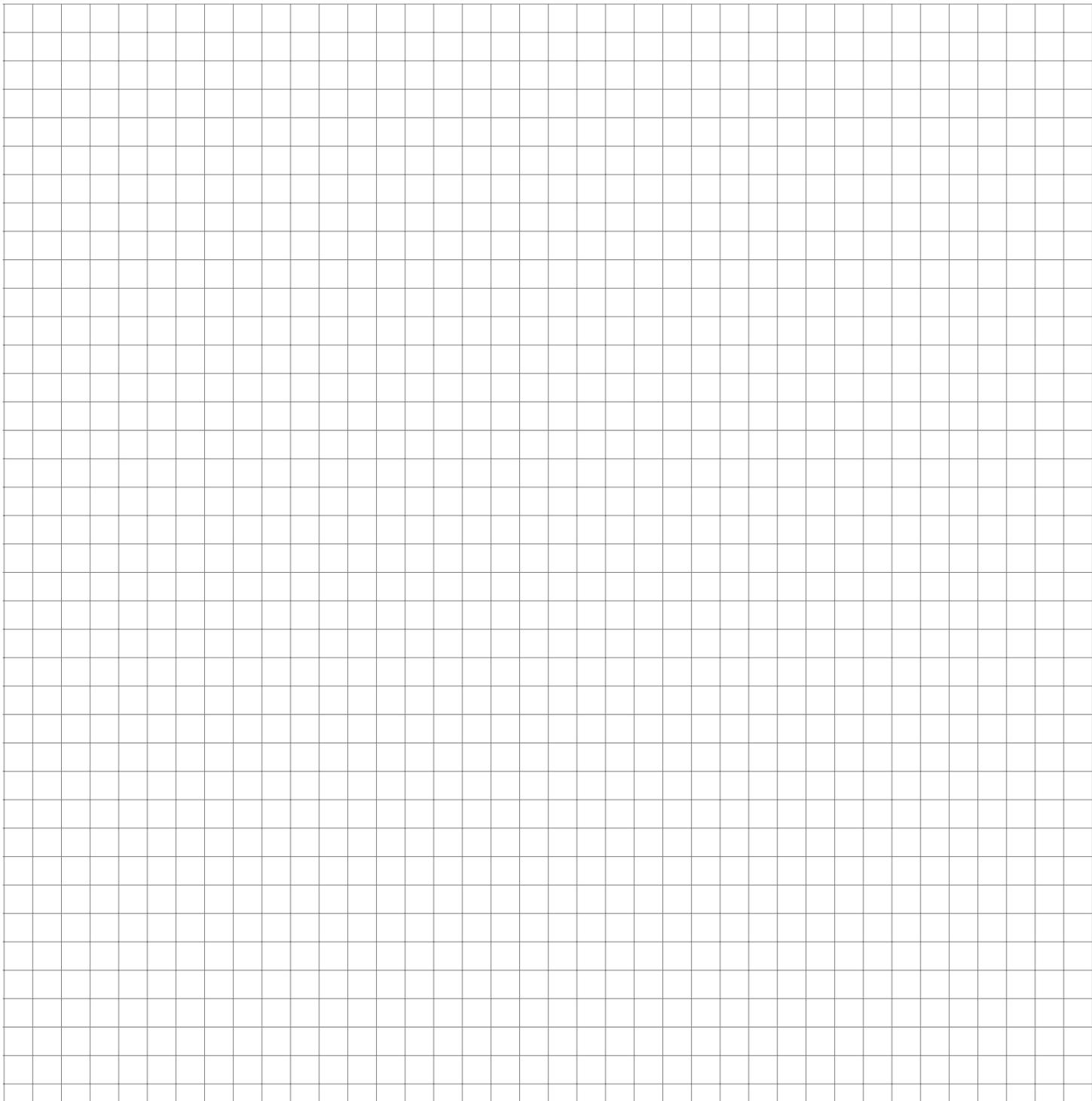
Aufgabe 87

Bei der Prüfung der Füllmenge von Fruchtsaftflaschen ergaben sich folgende Werte:

ccm	197	198	199	200	201	202	203	204	205	206	207
Anzahl	2	1	3	1	3	1	2	1	1	0	1

Nach Angaben des Abfüllers ist die Füllmenge normalverteilt mit einer Varianz von $\sigma^2 = 2,25$.

- Man gebe ein Schätzintervall für den Erwartungswert μ zum Niveau $1 - \alpha = 0,94$.
- Welcher Stichprobenumfang n garantiert eine Länge von 1 für das Schätzintervall?

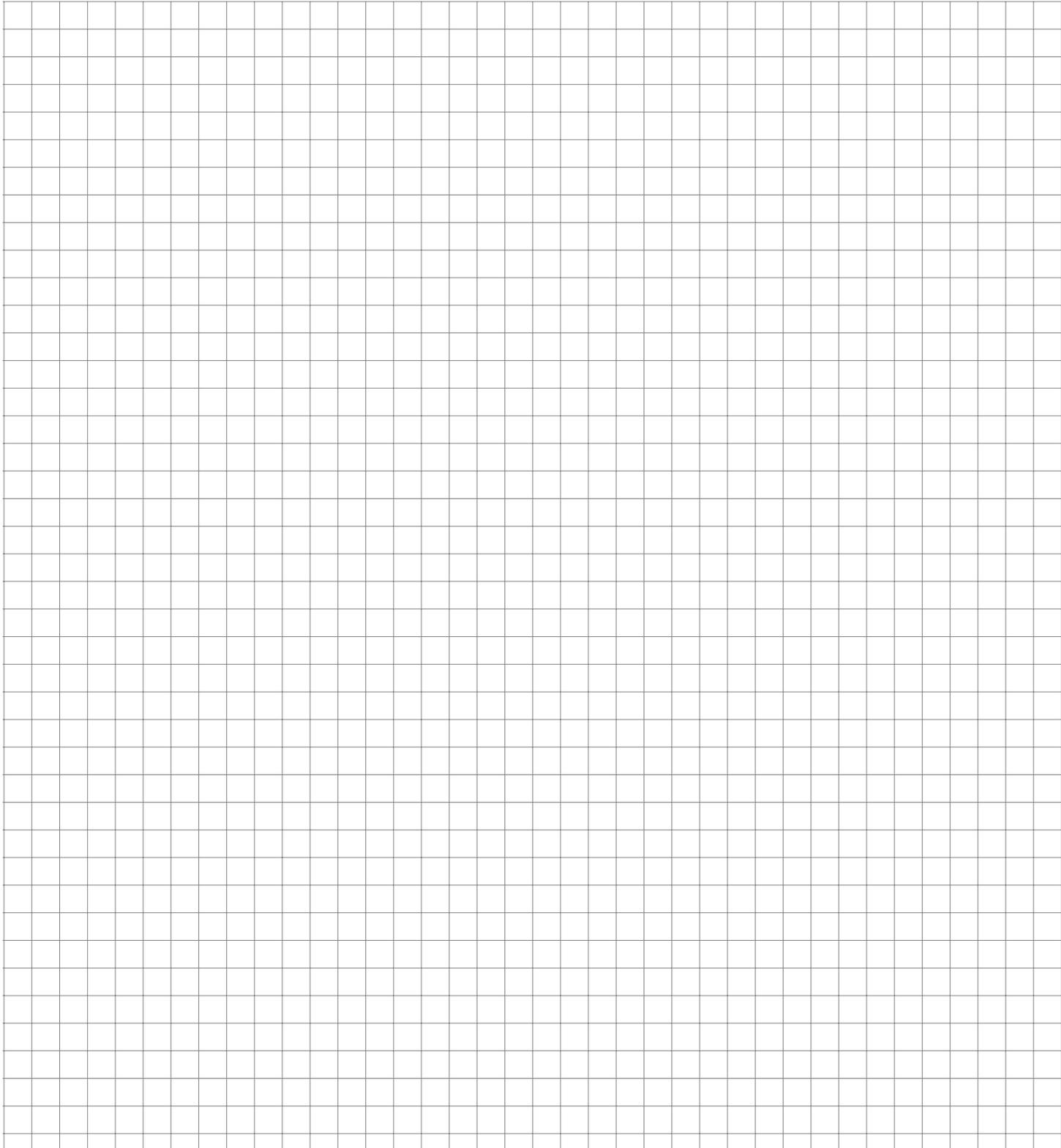


Aufgabe 88

Induktiv: Intervallschätzer (3)

X_1, \dots, X_{31} beschreibe eine einfache Stichprobe aus einer beliebig verteilten Grundgesamtheit. Aus den Ergebnissen wurden $\bar{x} = 9$ und $s^2 = 31/4$ errechnet. Zur Irrtumswahrscheinlichkeit $\alpha = 0,05$ bestimme man

- ein Schätzintervall für den Erwartungswert μ ,
- unter der Annahme, dass die Grundgesamtheit normalverteilt ist, ein Schätzintervall für die Varianz σ^2 .



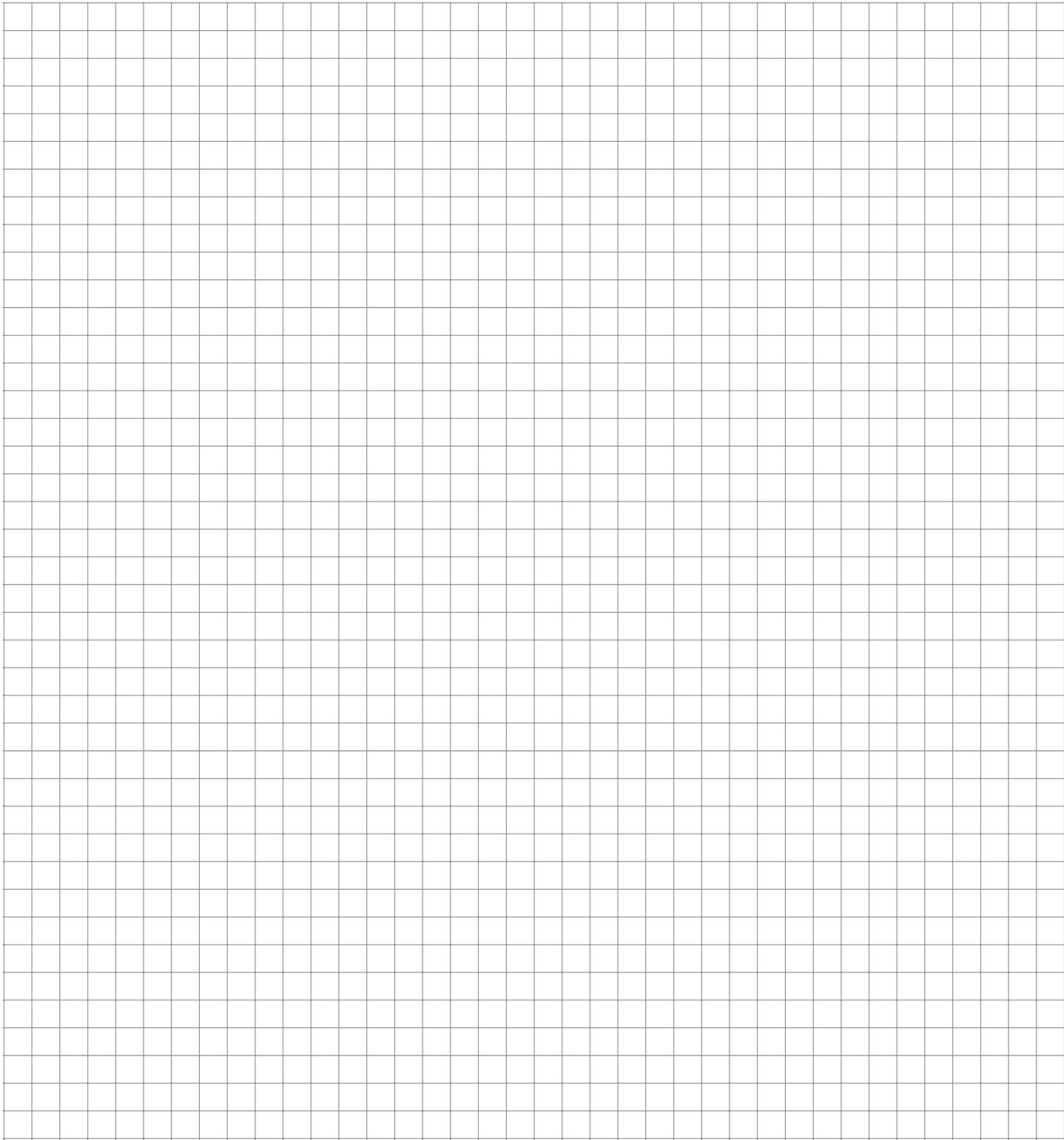
Aufgabe 89

Induktiv: Intervallschätzer (12)

Ein Barista dosiert in einer Espresso-Bar die Menge Kaffeepulver in Gramm bei 10 zufällig ausgewählten Espresso-Bezügen folgendermaßen:

7.3	8.2	7.0	9.2	8.1	6.9	7.1	8.5	9.0	8.5
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

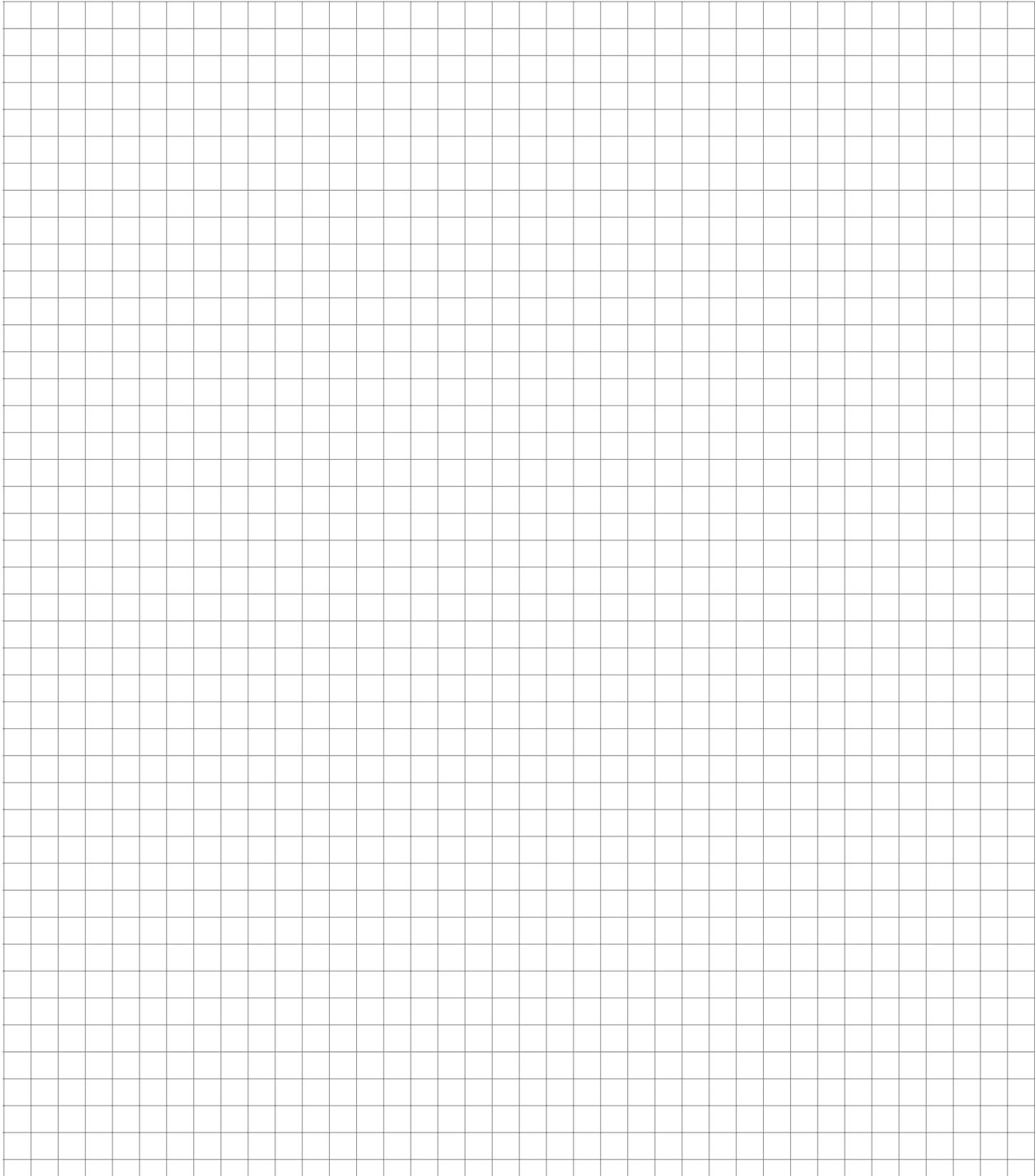
Berechnen Sie für diesen Barista ein Konfidenzintervall für die Varianz der Kaffeemenge pro Espresso zum Konfidenzniveau 0,95.



Aufgabe 90

Induktiv: Tests Fehler 1. Art (4)

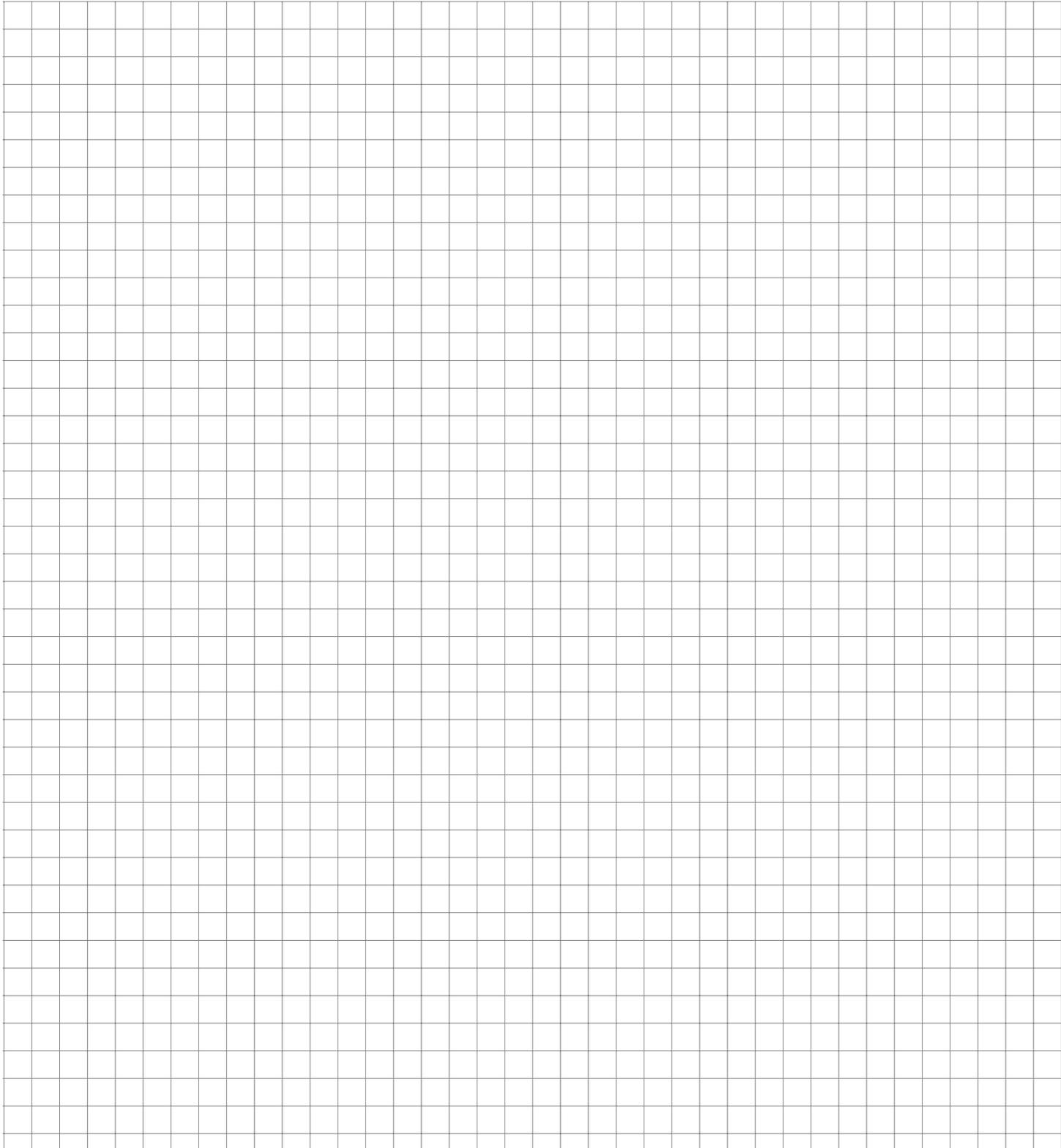
In einem Spielkasino werden Zweifel geäußert, dass ein bestimmter Würfel fair ist, d.h. alle Zahlen gleich häufig auftreten. Der Spielleiter fordert einen Zweifler auf, ein Signifikanzniveau α zwischen 0,01 und 0,40 anzugeben, zu dem die Hypothese H_0 , dass der Würfel fair ist, getestet werden soll. Welches α wird der Zweifler wählen, wenn er möchte, dass der Würfel aus dem Spiel genommen wird?



Aufgabe 91

Induktiv: Tests Erwartungswert (5)

Ein Arbeiter braucht für die Bearbeitung eines Werkstücks im Durchschnitt 7 Minuten (420 sec. = μ_0). Ein Fachmann schlägt, um eine Zeitersparnis zu erreichen ($\mu < \mu_0$), eine andere Bearbeitungsart vor und will die Effektivität seines Vorschlags mithilfe einer Stichprobe vom Umfang $n = 16$ testen. Führen Sie diesen Test (Hypothese $H_0 : \mu = \mu_0$ gegen $H_1 : \mu < \mu_0$) zum Signifikanzniveau $\alpha = 0,05$ bzw. $0,01$ durch. Dabei sei ferner vorausgesetzt, dass die Grundgesamtheit normalverteilt ist. Die Stichprobe ergab folgende Werte: $\bar{x} = 408$ und $s = 25,7$.

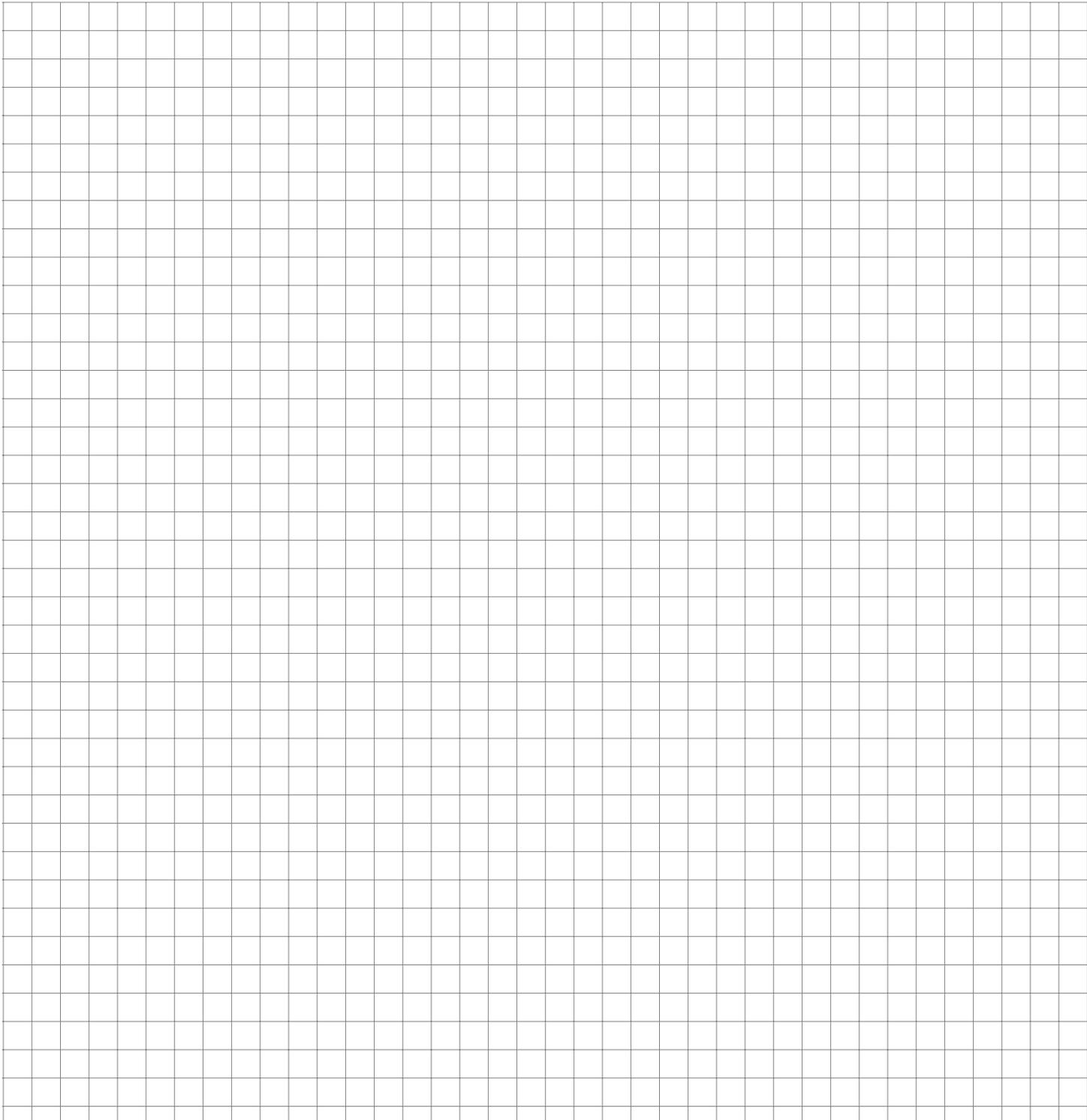


Aufgabe 92

Induktiv: Tests Erwartungswert (6)

Die Personalabteilung eines Großunternehmens hat den Verdacht, dass die Mitarbeiter die Mittagspause (maximal 1 Stunde) im Durchschnitt überziehen. Mit einer einfachen Stichprobe der Pausenlänge von 20 Mitarbeitern soll getestet werden, ob die Zeiten eingehalten werden oder ob im Mittel überzogen wird. Es ergibt sich für die Pausendauer ein Stichprobenmittel von 70 Minuten und eine Stichprobenstandardabweichung von 15 Minuten. Die Pausendauer eines Mitarbeiters kann als normalverteilte Zufallsvariable angenommen werden.

Formulieren Sie Nullhypothese und Gegenhypothese und testen Sie zum Signifikanzniveau von 5 %.

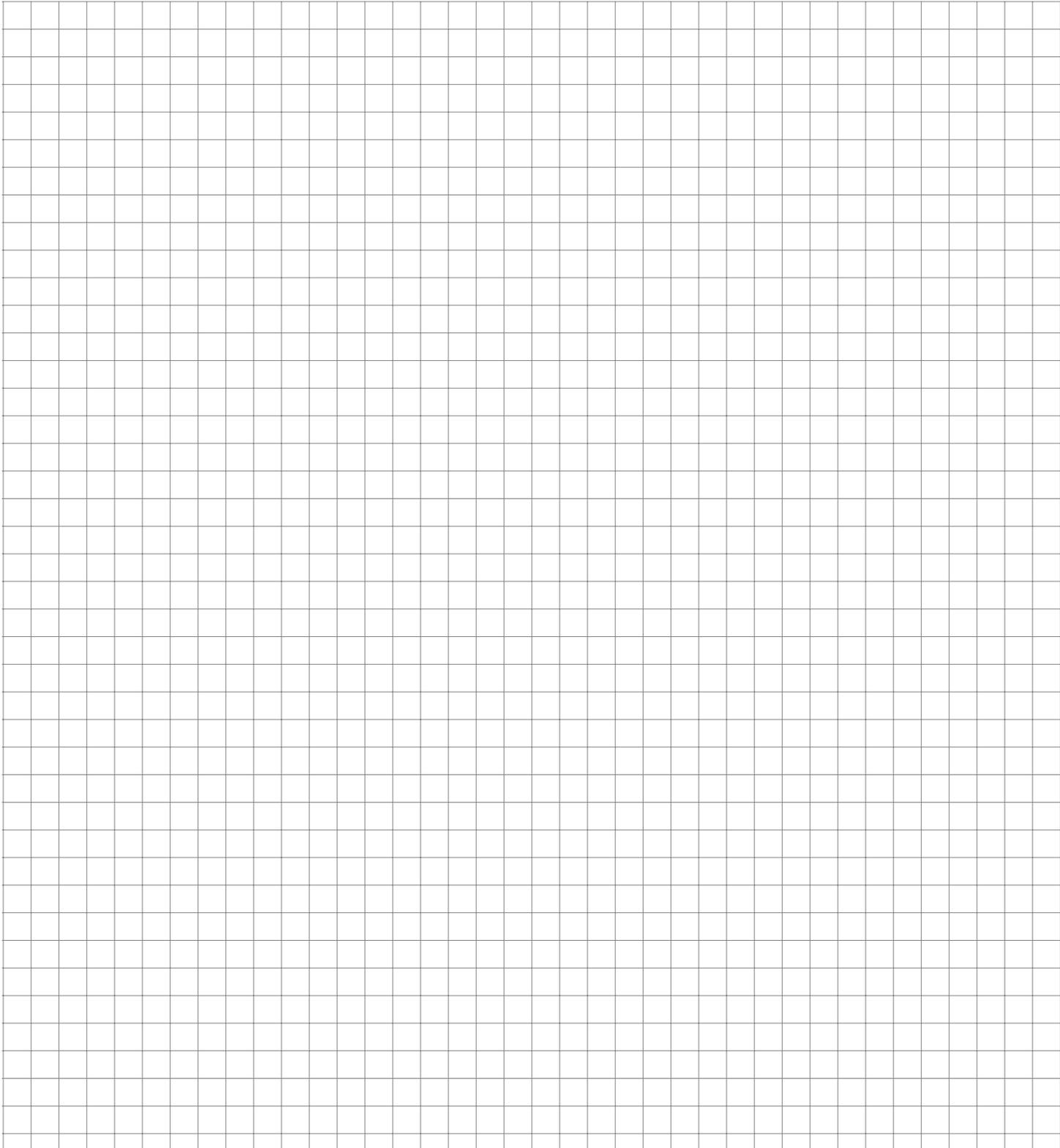


Aufgabe 93

Induktiv: Intervallschätzer (8)

Angeblich sollen Studierende sich in der Nacht vor einer Klausur kürzer in der Tiefschlafphase befinden also im Durchschnitt aller Nächte. Eine einfache Stichprobe von 61 Studenten wird diesbezüglich untersucht. Im Durchschnitt wurde in dieser Stichprobe 48 Minuten Tiefschlaf in den letzten 24h vor der Klausur gemessen, mit einer Stichprobenvarianz von 196.

Bestimmen Sie ein 95 %-Konfidenzintervall für die Tiefschlaflänge aller Studierenden am Tag vor der Prüfung.

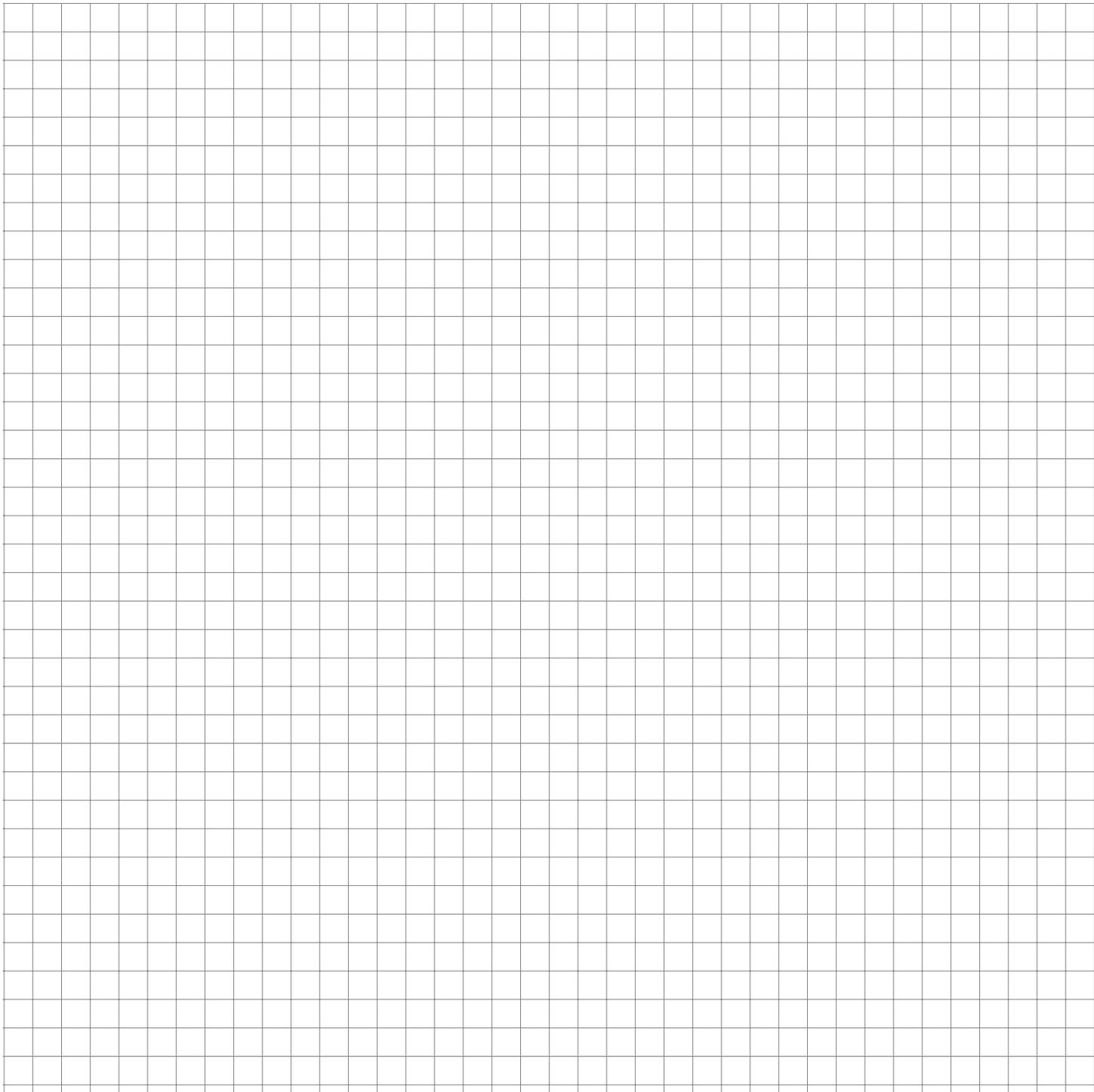


Aufgabe 94

Induktiv: Intervallschätzer (9)

Die Hochschule X möchte wissen, wie gut Ihre Studenten über internationale aktuelle Nachrichten aus der Politik informiert sind. 40 zufällig ausgewählten Studierenden werden Fragen zu 100 Nachrichten der letzten beiden Monate gestellt. Im Durchschnitt können die Befragten 58 Fragen richtig beantworten bei einer Stichprobenstandardabweichung von 3,2.

- Berechnen Sie ein 99 %-Konfidenzintervall für die durchschnittlich von allen Studenten der Hochschule richtig beantwortete Anzahl der Fragen.
- Im Landesdurchschnitt aller Studenten aller Hochschulen ergeben sich 65 Punkte. Testen Sie zum Signifikanzniveau von 5 %, ob der der Durchschnitt der Punktzahl an der Hochschule X niedriger ist als im Landesdurchschnitt.



Aufgabe 95

Induktiv: Intervallschätzer Tests (KI_KlausurGewicht)

Studierende beschwerten sich, dass die Klausuren in Statistik zu schwer seien. Der Dozent möchte das natürlich im Sinne der Studierenden verbessern und lässt durch das Prüfungsamt die Metallklammern der Klausuren durch eine leichtere Variante aus Kunststoff ersetzen. Die Studierenden glauben aber nicht, dass dadurch die Klausuren leichter werden. Der Dozent möchte das beweisen und zieht eine einfache Stichprobe von 10 Klausuren, bei der er folgende Gewichte misst:

65.28 65.84 64.47 63.82 66.64 62.55 65.74 64.90 65.87 65.29

Gehen Sie im folgenden davon aus, dass es sich beim Gewicht der Klausur um eine normaverteilte Zufallsvariable handelt.

- Bestimmen Sie zum Konfidenzniveau 0,95 ein symmetrisches Schätzintervall für den Erwartungswert des Gewichts einer Klausur.
- Mit Metallklammer hatten die Klausuren früher ein durchschnittliches Gewicht von 65,86 g. Testen Sie zum Signifikanzniveau von 5 % auf Basis der Stichprobe, ob die Klausuren jetzt leichter sind.
- Was bedeutet bei dem Test von b) der Fehler 2. Art?



Aufgabe 96

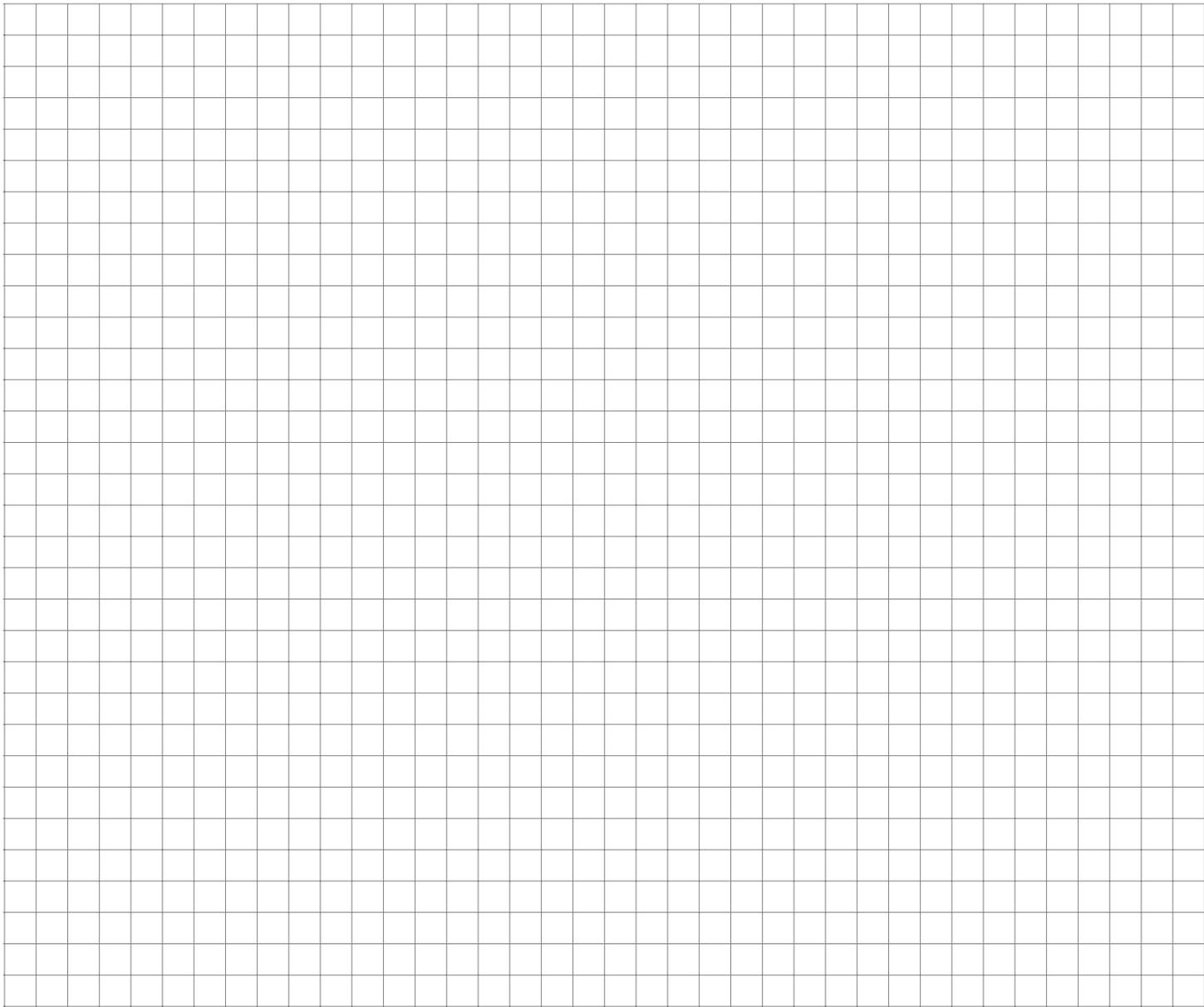
Induktiv: Intervallschätzer (KI_ToreWM2010)

Die deutsche Nationalmannschaft hat in 50 zufällig ausgewählten Länderspielen die folgenden Toranzahlen geschossen:

Anzahl der Tore pro Spiel	0	1	2	3	4
Anzahl der Spiele mit diesem Ergebnis	18	22	5	3	2

Gehen Sie im folgenden davon aus, dass die erhobenen Toranzahlen aus einer einfachen Stichprobe der Grundgesamtheit aller Länderspiele der deutschen Nationalmannschaft stammen.

- Bestimmen Sie das 95 %-Konfidenzintervall für den Erwartungswert μ der Toranzahl in allen Länderspielen.
- Die Standardabweichung der Tore pro Spiel in der Grundgesamtheit aller Länderspiele sei jetzt mit $\sigma = 1,0$ gegeben. Wie groß muss der Umfang einer Stichprobe sein, damit das 95 %-Konfidenzintervall für μ nicht breiter als 0,5 Tore ist?

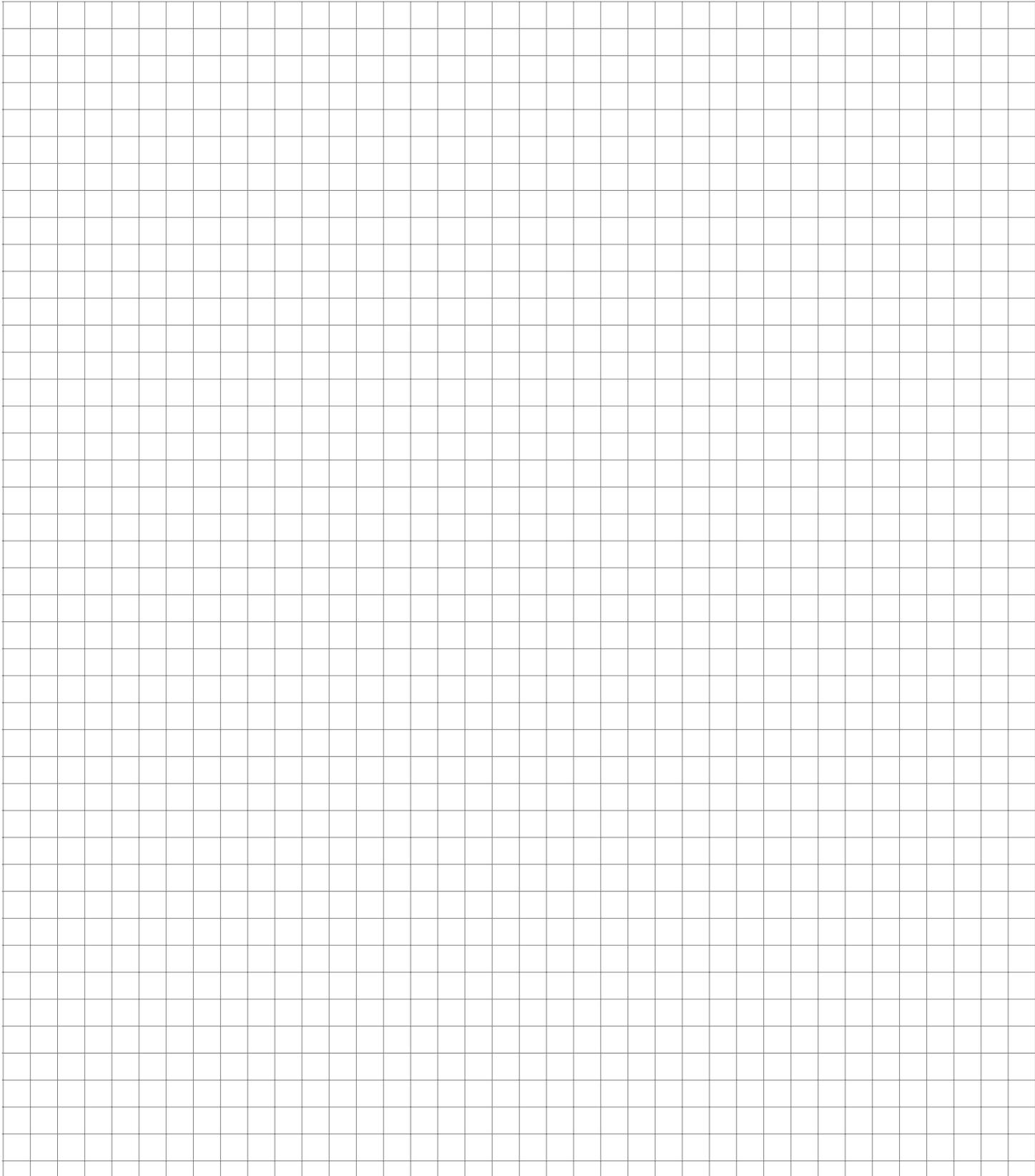


Aufgabe 97

Induktiv: Konfidenzintervall Anteil (10)

Am Tag der Bundestagswahl werden kurz nach dem Schließen der Wahllokale 300 zufällig ausgewählte Wahlberechtigte gefragt, ob sie gewählt hatten. 250 der Befragten bejahten das.

Schätzen Sie mit der Approximation der Normalverteilung zum Konfidenzniveau 99 % ein Konfidenzintervall für die Wahlbeteiligung aller Wahlberechtigten.

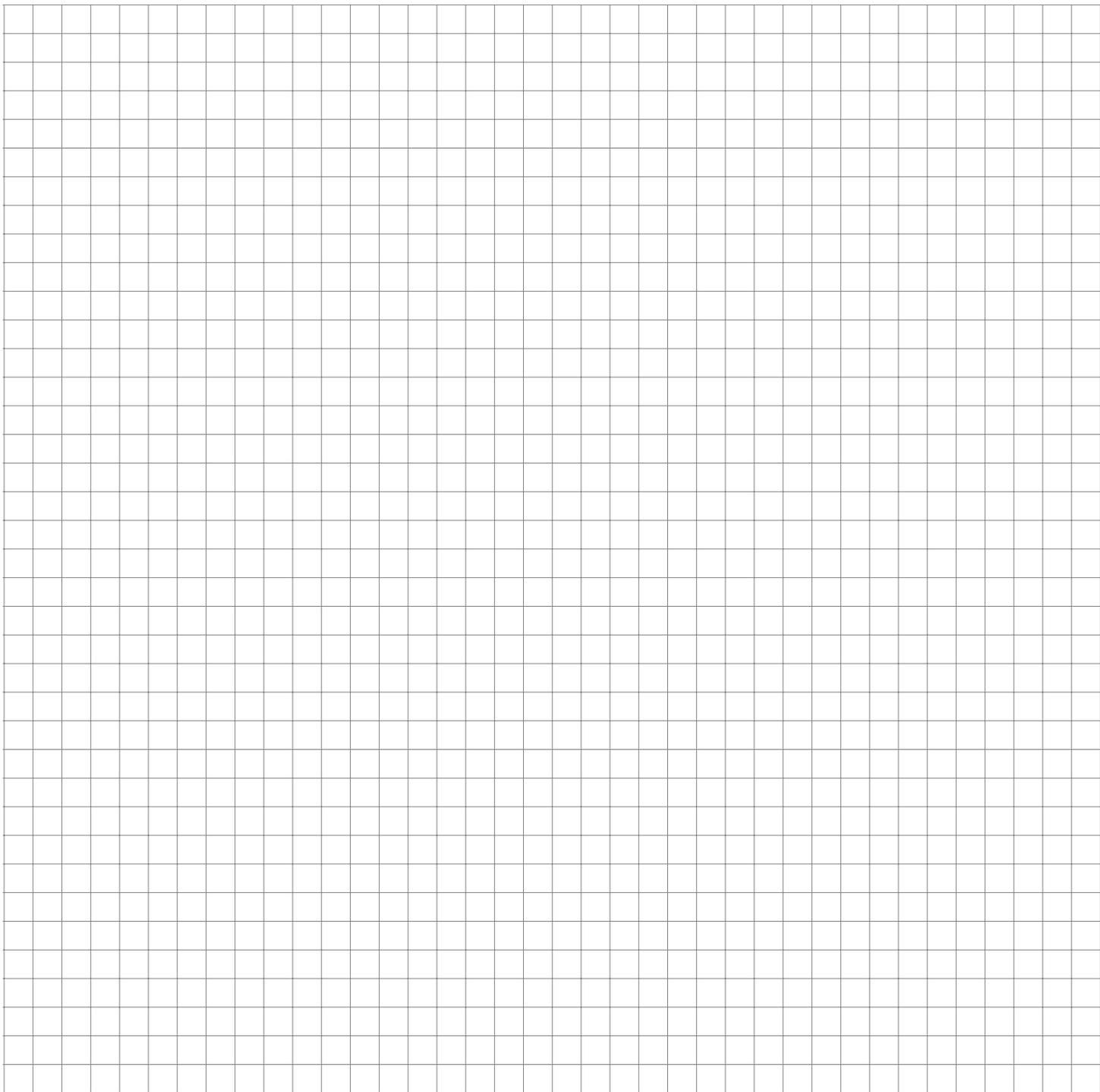


Aufgabe 98

Induktiv: Tests Anteil (11)

Herr Meyer betreibt einen Schnellimbiss für Vegetarier. Eines Tages wird er von einem Kunden wegen Betrugs und Etikettenschwindels verklagt. Der Kunde kann per Fotos nachweisen, dass 8 von 25 von ihm bestellten Gemüsesuppen gar nicht vegetarisch waren, da sich eine Fliege darin befand. Das Gericht verlangt auf Basis dieser als einfach akzeptierten Stichprobe einen Hypothesentest, mit dem der Anteil μ aller Gemüsesuppen mit Fliegen als über den gesetzlich zugelassenen 10 % nachgewiesen wird.

- Ist die Approximation durch die Normalverteilung in diesem Fall gerechtfertigt?
- Formulieren Sie H_0 sowie H_1 .
- Führen Sie den Hypothesentest zu einem Signifikanzniveau von $\alpha = 2,28\%$ durch.



Aufgabe 99

Induktiv: Tests Fehler (13)

Herr Untermann möchte auf der Karriereleiter in seiner Firma, einem Telekommunikationsunternehmen, etwas vorankommen und schlägt deshalb folgende Maßnahme vor: Der Datendurchsatz der bisher üblichen Flatrates beim Internetzugang von Kunden soll ab einem Volumen von 1GB downloads pro Monat extrem gedrosselt werden. Die alten Konditionen können die Kunden dann optional zukaufen. Bisher hat Herr Untermanns Firma einen Marktanteil von 45 % bei dieser Art flatrates. Eine Stichprobe unter allen potentiellen Kunden vom Umfang $n = 2500$ ergab, dass immerhin noch 43% der potentiellen Kunden diese Leistung mit den verschlechterten Konditionen nachfragen wollen. Herr Untermann schließt daraus, dass sich die Kunden von der Verschlechterung der Vertragsbedingungen nicht abschrecken lassen und freut sich auf die Mehreinnahmen durch seinen Plan.

- Formulieren Sie in diesem Szenario die Nullhypothese H_0 und die Alternative H_1 .
- Worin besteht in diesem Beispiel das Risiko, den Fehler 1. Art zu begehen?
- Was bedeutet hier der Fehler 2. Art?
- Würden Sie an der Stelle von Herrn Untermann ein möglichst kleines Signifikanzniveau α zugrunde legen und dadurch einen größeren Fehler 2. Art (β) in Kauf nehmen oder umgekehrt β klein halten und dabei ein größeres α zulassen?



Aufgabe 100

100 zufällig ausgewählten Personen einer bestimmten Bevölkerungsschicht werden zwei Fragen vorgelegt, nämlich ob sie (mindestens) ein Smartphone besitzen bzw. ob sie soziale Netze im Internet an mehr als 3 Stunden am Tag nutzen. Es ergeben sich folgende Antworten:

>3h soziale Netze	Smartphone	
	ja	nein
ja	12	24
nein	38	26

Testen Sie zum Signifikanzniveau $\alpha = 5\%$, ob in dieser Bevölkerungsschicht die beiden Merkmale „Smartphonebesitz“ bzw. „starke Nutzung sozialer Netzwerke“ voneinander unabhängig sind.



Aufgabe 101

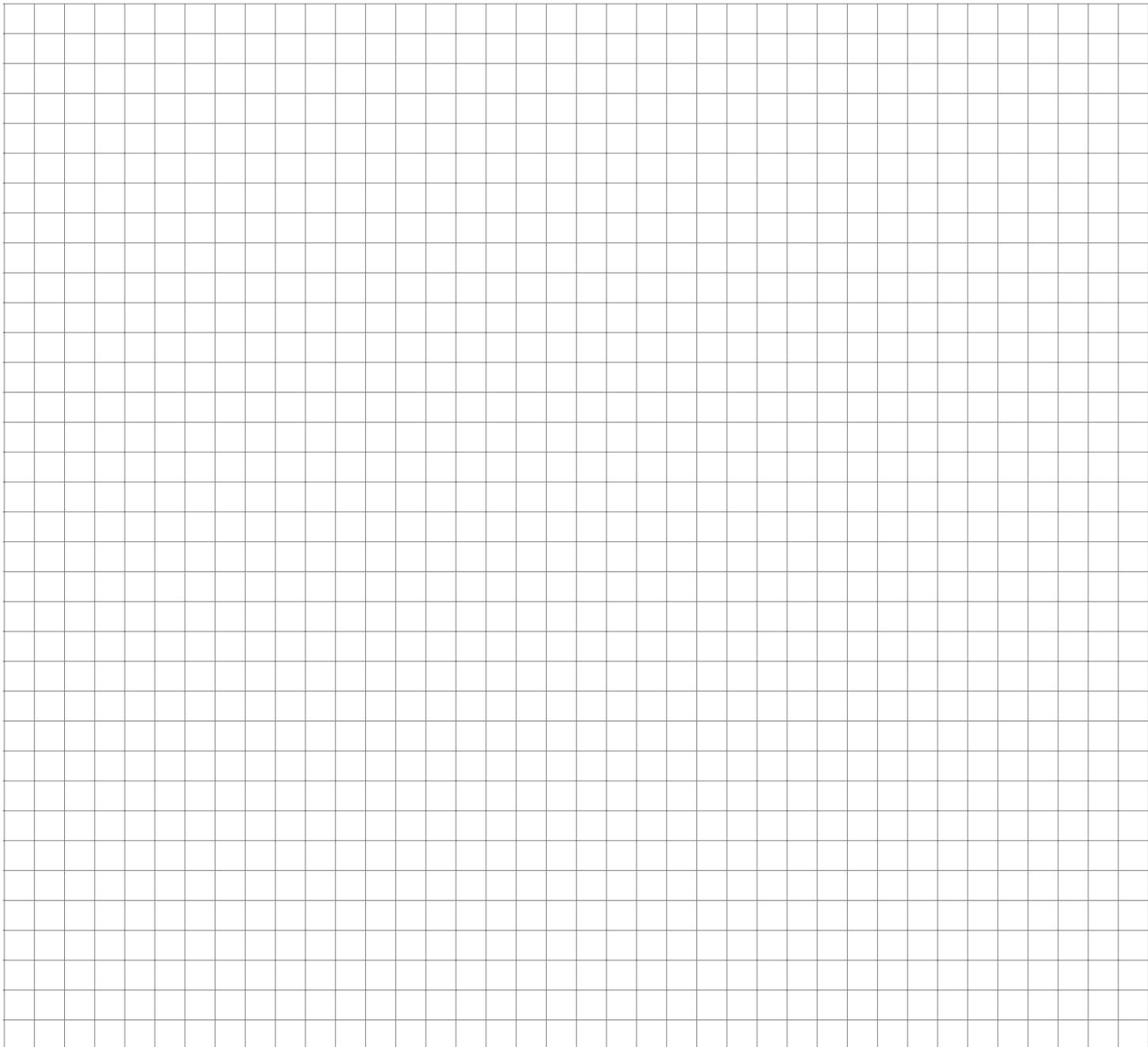
Induktiv: Tests Kontingenz (15)

Wie hängen der Bierkonsum in Litern pro Woche und die Selbsteinschätzung als Fußballfan zusammen? Personen einer einfachen Stichprobe wurden diesbezüglich befragt:

20 Fußballfans und 120 Nichtfußballfans gaben einen Bierkonsum von höchstens 1 l pro Woche an. Zwischen 1 und 3 l pro Woche trinken 210 Fußballfans und 200 Nichtfußballfans. 150 Fußballfans und 90 Nichtfußballfans gaben einen Bierkonsum von mindestens 7 l an. 145 Fußballfans und 65 Nichtfußballfans lagen in der verbleibenden Zwischengruppe.

Es soll nun auf Basis der Stichprobe die Unabhängigkeit der Fußballaffinität vom Bierkonsum in der Gesamtbevölkerung getestet werden.

- Formulieren Sie H_0 bzw. H_1 .
- Was bedeutet hier der Fehler 1. bzw. der Fehler 2. Art?
- Führen Sie den Test zum Signifikanzniveau von 10 % durch.



Aufgabe 102

500 Vollzeitbeschäftigte eines Krankenhauses wurden für eine Studie zur Abhängigkeit von Überstunden und dem Geschlecht zufällig ausgewählt. Die Probanden wurden nach Ihrer durchschnittlichen Wochenarbeitszeit im letzten Jahr befragt. Dabei wurde die Wochenarbeitszeit in die drei Klassen „unter 40“, „von 40 bis unter 45“ und „45 und mehr“-Wochenarbeitsstunden eingeteilt; außerdem wurden die Mitarbeiter nach Geschlecht unterschieden. Die Ergebnisse der Stichprobe sind der folgenden Tabelle zu entnehmen:

	durchschn. Wochenarbeitszeit [h]		
	unter 40	40 bis unter 45	45 und mehr
Frauen	165	55	25
Männer	175	45	35

Testen Sie bitte zur Irrtumswahrscheinlichkeit von 5 %, ob in diesem Krankenhaus die wöchentliche Arbeitszeit unabhängig vom Geschlecht ist.



Aufgabe 103

Ist die Fahrtzeit mit dem PKW für eine vorgegebene Strecke vom Geschlecht des Fahrers bzw. der Fahrerin abhängig? Diese Frage soll mit einer einfachen Stichprobe untersucht werden. Es ergibt sich:

	Fahrtzeit		
	kurz	mittel	lang
Mann	524	455	221
Frau	413	263	124

- Formulieren Sie H_0 bzw. H_1 .
- Was bedeutet der Fehler 1. Art hier?
- Testen Sie zum Signifikanzniveau von 5 %, ob die Fahrtzeit unabhängig vom Geschlecht ist.

