Prof. Dr. Roland Jeske

Aufgabensammlung Statistik 2021

1 Univariate Analyse

Aufgabe 1.1
In einer Erhebung wurden folgende Merkmale erhoben:

- Alter in Jahren
- Nationalität
- Familienstand
- Geschlecht
- Raucher (ja/nein)
- Anzahl der täglich konsumierten Zigaretten
- Telefonnummer
- Distanz zwischen Wohn- und Arbeitsort
- Position im Unternehmen nach: Sachbearbeitung/untere Führungsposition/ mittleres Management/oberes Management
- monatliches Nettoeinkommen (in €)
- Typ des ersten PKWs (Keinen/Kleinwagen/Mittelklasse/Luxusklasse)
- Postleitzahl des Wohnortes
- Englischkenntnisse (A1, A2, B1, B2, C1, C2)

Beurteilen Sie die Skalierung der Merkmale und beschreiben Sie die Merkmale nach den Kriterien stetig/diskret und qualitativ/quantitativ.

Aufgabe 1.2
Betrachten Sie folgende Datensätze:

i) Die Landtagswahl im Freistaat Bayern ergab am 14.10.2018 folgendes Ergebnis:

Partei	Wähleranteil (in %)	Sitze
CSU	37,2	85
Grüne	17,5	38
Freie Wähler	11,6	27
AfD	10,2	22
SPD	9,7	22
FDP	5,1	11
sonstige	8,7	0

ii) Die Punkteverteilung einer Mathematikklausur (maximal waren 100 Punkte erreichbar) ergab folgende Punktezahlen:

$$\begin{gathered} 29 , 83 , 100 , 7 , 52 , 78 , 53 , 20 , 95 , 28 , 72 \\ 50 , 50 , 65 , 81 , 42 , 58 , 100 , 43 , 65 , 18 , 90 \\ 39 , 80 , 12 , 100 , 27 , 50 , 100 , 50 , 50 , 100 , 38 . \end{gathered}$$

Klassieren Sie die Ergebnisse zu Noten, wobei Sie folgende Regeln beachten sollten:

unter 50 Punkten: Note 5
ab 50 Punkten: Note 4
ab 60 Punkten: Note 3
ab 75 Punkten: Note 2
ab 90 Punkten: Note 1

iii) Die Notenverteilung des Anfängerjahrgangs 2018/9 der Fakultät BW im Fach Mathematik ist wie folgt gegeben:

Note	Häufigkeit
1	10
2	12
3	22
4	22
5	108

a) Stellen Sie die jeweiligen Daten grafisch dar. Welche Abbildungsform erscheint Ihnen geeignet?

b) Ermitteln Sie geeignete Lagemaße für die Daten.

Aufgabe 1.3
Betrachten Sie folgende Datensätze:

a) 1, 7, 62, 42, 47, 48, 12, 34, 51, 2, 53, 17, 34, 23, 36, 17, 8, 26, 36, 28.

b) 12, 78, 56, 75, 112, 45, 37, 26, 63.

c) 32, 17, 22, 18, 67, 88, 54, 6.

Berechnen Sie jeweils folgende Quantile:
Median, Quartile, Quintile, 5%-Quantil, 10%-Quantil, 90%- Quantil und 95%-Quantil.

Aufgabe 1.4
Betrachten Sie folgenden Datensatz:

$$10 , 2 , 9 , 1 , 8 , 10 , 1 , 1 , 7 , 10 , 3 , 2 , 4 , 9 , 10 , 3 .$$

Berechnen Sie arithmetisches Mittel, Median und Modus.
Welches der Lagemaße erscheint Ihnen geeignet, um das „Zentrum“ des Datensatzes zu beschreiben?

Aufgabe 1.5
Die Erhebung der monatlichen Konsumausgaben für Genussmittel von 200 zufällig befragten Haushalten ergab folgende Werte:

Ausgaben (€) von …	bis unter …	Häufigkeit
0	25	15
25	50	45
50	100	60
100	150	50
150	200	15
200	300	10
300	500	5

a) Berechnen Sie das arithmetische Mittel, Median und Modus.

b) Berechnen Sie zudem die Standardabweichung und den Quartilsabstand.

c) Zeichnen Sie ein Histogramm.

d) Welche Aussagen können Sie über die Schiefe machen?

Aufgabe 1.6
10 Studierende wurden nach der Häufigkeit befragt, mit der sie im letzten Monat die Mensa besucht haben:

$$5, \quad 7, \quad 6, \quad 4, \quad 5, \quad 2, \quad 0, \quad 12, \quad 0, \quad 5.$$

a) Berechnen Sie arithmetisches Mittel, Median und Modus.

b) Berechnen Sie Standardabweichung, Spannweite und Quartilsabstand.

Aufgabe 1.7
In einer Gruppe von Studierenden wurde folgende Altersverteilung erhoben:

Alter	Häufigkeit
17	1
18	5
19	11
20	7
21	1
23	1
28	1

a) Berechnen Sie arithmetisches Mittel, Median und Modus.

b) Beurteilen Sie die Schiefe der Daten.

Aufgabe 1.8
Eine Pflegeeinrichtung konnte in den letzten fünf Jahren folgende Umsatzsteigerungen (jeweils gegenüber dem Vorjahresergebnis) erzielen:

$$+10\%, +2\%, -8\%, +5\%, +6\%.$$

Wie hoch ist die durchschnittliche jährliche Umsatzsteigerung innerhalb der fünf Jahre?

Aufgabe 1.9
Im Zuge einer Lebensmitteluntersuchung wurde bei Frischfleischproben von Hackfleisch die Anzahl der Keime je mg ermittelt. Die Stichprobe ergab folgende Werte (Keime je mg):

$$180, 175, 190, 180, 185, 250, 175, 190, 175, 195, 195$$

Drei Lebensmittelchemiker ziehen daraus folgende Schlüsse:

- Matt Wurst meint, dass die meisten Proben 180 Keime je mg besaßen.
- G. Hacktes kommt zu dem Schluss, dass durchschnittlich 190 Keime je mg zu verzeichnen waren.
- H. Schee hingegen behauptet, dass in gut der Hälfte der Proben mindestens 185 Keime je mg zu beobachten waren.

Wer von den Lebensmittelchemikern hat Recht, wer hat Unrecht? Begründen Sie Ihre Beurteilung kurz.

Aufgabe 1.10
Die Liegedauer von Patient/innen nach einer speziellen Operation ist nach Geschlechtern differenziert wie folgt erhoben worden:

- Weibliche Patientinnen: 6,10,5,3,4,6,7,3
- Männliche Patienten: 6,8,12,5,3,5,7,6,8

a) Erstellen Sie getrennt nach Geschlecht Boxplots für die erhobenen Daten.

b) Welche Schlüsse können Sie hinsichtlich des Vergleichs beider Gruppen ziehen?

Aufgabe 1.11
Eine Großstadt und ihre umgebenden Landkreise bilden gemeinsam eine Metropolregion. Für die einzelnen Gebiete wurden folgende Indikatoren ermittelt:

- Arztdichte in Anzahl niedergelassener Ärzte je 1.000 Einwohner
- Unfalldichte in Anzahl von Verkehrsunfällen je 1.000 zugelassenen Kraftfahrzeugen

Für die Teilregionen wurden folgende Kenngrößen berechnet:

Gebiet	Arztdichte	Unfalldichte
Großstadt	6,4	25,7
Umgebende Landkreise	3,6	38,4

a) Berechnen Sie die Arztdichte der gesamten Metropolregion, wenn bekannt ist, dass in der Großstadt 5000 Ärzt/innen und in den umgebenden Landkreisen 2400 Ärzt/innen niedergelassen sind.

b) Berechnen Sie die Unfalldichte der gesamten Metropolregion, wenn bekannt ist, dass in der Metropole 420.000 und in den Landkreisen 240.000 Kraftfahrzeuge zugelassen sind.

2 Bivariate Analyse

Aufgabe 2.1
In einer Stichprobe wurden bei 8 Unternehmen Umsatz und Werbebudget erhoben:

Unternehmen	Umsatz (in Mio. €)	Werbebudget (in Mio. €)
1	998	5,1
2	124	0,8
3	256	2,5
4	786	4,6
5	261	1,9
6	308	2,0
7	359	2,9
8	612	4,2

a) Ermitteln Sie ein geeignetes Maß für den Zusammenhang der beiden Merkmale.

b) Passen Sie eine lineare Regressionsgerade an die Daten an. Begründen Sie die Wahl der abhängigen und der unabhängigen Variablen.

c) Beurteilen Sie die Anpassungsgüte, indem Sie das Bestimmtheitsmaß berechnen.

Aufgabe 2.2
Die Verwaltungsgemeinschaft der beiden Örtchen Oberschlag und Niederschlag am Regen betreibt ein gemeinsames Verkehrsbüro. Das aktuelle Beherbergungsverzeichnis weist folgende Übernachtungsmöglichkeiten mit der Zertifizierung des Reiseanbieters PFUI sowie den Übernachtungspreisen für zwei Personen aus:

Beherbergungsstätte	Zertifizierung (Sterne)	Doppelzimmerpreis bzw. Übernachtungspreis für 2 Personen (in €)
Hotel zur schlechten Aussicht	3	79
Gasthof zur schalen Schenke	3+	75
Hotel zur harten Matratze	3	82
Gasthof zum letzten Tropfen	3	85
Hotel am Stolperstein	5	129
Pension Gebald-Wieder	3	57
Hotel Zwergblick	5	179
Pension Schlafnich-Guth	3	69
Stundenhotel Bleibtreu	2	50
Almhotel Lawinentobel	5	198
Hotel Höllenblick	4	99
Hotel zur letzten Absteige	2	45
Hotel Landeplatz	3	59
Gasthof zum blauen Karpfen	4+	88
Gasthof zum abgegebenen Löffel	3+	78
Campingplatz Hochwasser	2	32

Berechnen Sie ein geeignetes Zusammenhangsmaß zwischen Zertifizierung und Übernachtungspreis und interpretieren Sie den Wert.

Aufgabe 2.3

Student Thorsten arbeitet als Aushilfskellner in einem amerikanischen Steak-Restaurant. Die Trinkgelder[1] der Beschäftigten werden in eigens beschrifteten Sparschweinen mit den Initialen der Bedienungen gesammelt. Das „T-TIP“-Schwein wird von Thorsten nicht nur aufmerksam beachtet, sondern auch statistisch analysiert, indem er folgende Daten erhebt:

Rechnungssumme (€)	128,60	86,20	61,30	195,70	58,90	107,80	134,80	76,30
Trinkgeld (€)	6,40	3,80	1,70	4,30	3,10	5,20	5,20	3,70

a) Erklären Sie die Höhe von Thorstens Trinkgeldern durch die Höhe des Rechnungsbetrages, indem Sie die lineare Regressionsgerade bestimmen.

[1] engl.: tip

b) Beurteilen Sie Anpassung des Regressionsmodells.

Aufgabe 2.4
Das Studentenwerk einer Hochschule möchte die Zufriedenheit seiner Gäste mit der Qualität des Mensaessens erheben. Dazu wurden 400 Personen befragt, von denen 75% Studierende waren, der Rest bestand aus Hochschulbediensteten. Unter den Studierenden waren 65% mit der Qualität zufrieden, unter den Bediensteten waren es 56%.
Berechnen Sie den

a) Yule'schen Assoziationskoeffizienten.

b) korrigierten Kontingenzkoeffizienten nach Pearson

um den Zusammenhang zwischen Konsumentengruppe und Zufriedenheit zu beschreiben.

Aufgabe 2.5
Recht spontan beschließen sechs Personen auf einer Familienfeier (80. Geburtstag), einen Wettlauf über 1.000m zu machen. In Ermangelung einer Stoppuhr wurden statt der Laufzeiten lediglich die Platzierungen sowie das Alter festgehalten:

Startnummer	Person	Platz beim 1.000 m Lauf	Alter
1	Ludwig Lungenschwund	6	18
2	Ferdinand Fettpolster	4	37
3	Erna Ehrgeiz	2	20
4	Arthur Athletikus	1	31
5	Gerda Gefräßig	5	45
6	Frieda Fitness	3	80

Berechnen Sie ein geeignetes Maß für den Zusammenhang zwischen Alter und Platzierung.

Aufgabe 2.6
Die nachfolgende Tabelle gibt die Verteilung von geretteten und ertrunkenen Passagieren der MS „Titanic“ an:

	Verbleib	
Passagierklasse	gerettet	ertrunken
1. Klasse	202	123
2. Klasse	118	167
3. Klasse	178	528

Berechnen Sie den korrigierten Kontingenzkoeffizienten nach Pearson als Maß für den Zusammenhang zwischen Klasse und Verbleib der Passagiere.

3 Wahrscheinlichkeitsrechnung

Aufgabe 3.1
Sie planen eine Studentenparty und wissen, dass 80% der Teilnehmer Beefburger und 50% der Teilnehmer Veggieburger mögen. 45% der Teilnehmer mögen beide Burgersorten.

a) Wie groß ist die Wahrscheinlichkeit, dass ein Teilnehmer gar keine Burger mag?

b) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Partybesucher auch Liebhaber von Beefburgern ist, falls bekannt ist, dass er Veggieburger mag?

Aufgabe 3.2
Studentin Vera Vergesslich hat ihre sechsstellige Matrikelnummer vergessen, die sie für den Zugang zum Intranet ihrer Hochschule benötigt.
Wie viele Möglichkeiten gibt es, wenn

a) sie gar keine Erinnerung an ihre Matrikelnummer hat und alle Kombinationen durchprobiert?

b) ihr einfällt, dass der Geburtstag ihres Freundes, der 23., vorkommt?

c) ihr einfällt, dass der Geburtstag ihres Freundes, der 23., vorkommt, dieser aber weder am Anfang noch am Ende der Matrikelnummer steht?

d) ihr einfällt, dass der Geburtstag ihres Freundes, der 23., vorkommt, dieser aber weder am Anfang noch am Ende der Matrikelnummer steht und dass weiterhin keine der Ziffern doppelt vorkommt?

Aufgabe 3.3
Ein verliebter Student besucht regelmäßig seine Freundin mit der Bahn. In 70% der Fälle fährt er mit der Privatbahn Departa, in den anderen Fällen mit der deutschen Bummelbahn (kurz: DBB). Erfahrungsgemäß sind 80% der DBB-Züge und 30% der Züge von Departa verspätet.

a) Wie groß ist die Wahrscheinlichkeit dass der Student zu spät zu seiner Freundin kommt?

b) Wieder einmal kommt der arme Student verspätet zu seiner Freundin. Wie groß ist die Wahrscheinlichkeit, dass er mit Departa gefahren ist?

Aufgabe 3.4

Zum Nachweis einer HIV-Erkrankung wird seit mehr als als zwei Jahrzehnten der sog. ELISA Test (**E**nzyme-**L**inked **I**mmuno**s**orbent **A**ssay), der heute als sogenannter „**Such**test“ durchgeführt wird. Der ELISA-Test besitzt eine Sensitivität von 99,7% sowie eine Spezifität von 98,5%. Die Prävalenz für eine/n in Deutschland an HIV Erkrankte/n beträgt etwa 0,1%.

a) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Patient tatsächlich an HIV erkrankt ist, falls der ELISA-Test für ihn ein positives Testergebnis zeigt?

Im Fall eines positiven Testergebnisses wird anschließend als **Bestätigungs**test ein sog. Western-Blot-Test durchgeführt, der unabhängig vom ELISA-Test arbeitet und eine Sensitivität von 80% sowie eine Spezifität von 99,99% besitzt.

b) Wie groß ist nun die Wahrscheinlichkeit, dass ein zufällig ausgewählter Patient tatsächlich an HIV erkrankt ist, falls beide hintereinander geschaltete Tests positive Testergebnisse lieferten?

c) Zum Nachweis einer HIV-Erkrankung wird momentan das zuvor beschriebene Verfahren von zwei hintereinander geschalteten Tests in genau dieser **Reihenfolge** durchgeführt.
Was können die Gründe für die Durchführung in dieser Reihenfolge sein? Kann die umgekehrte Reihenfolge zu einem unsichereren Ergebnis führen?

Aufgabe 3.5

Aktie A besitzt eine erwartetete jährliche Rendite von 8% und ein Risiko in Form der Standardabweichung von 3 Prozentpunkten, während Aktie B eine Rendite von 12% bei einer Standardabweichung von 5 Prozentpunkten erwarten lässt.
Ein Portfolio bestehe zu 30% aus Aktien vom Typ A, der Rest wurde in Aktien vom Typ B angelegt.

a) Wie lautet die erwartete Portfolio-Rendite?

b) Berechnen Sie das Risiko des Portfolios, indem Sie die Standardabweichung des Portfolios ermitteln. Unterstellen Sie eine Korrelation von 0,4 zwischen Aktie A und B.

Aufgabe 3.6
Die Zufallsvariablen X und Y seien unkorreliert. Ferner ist bekannt, dass die Varianz von X größer als die von Y ist. Betrachten Sie nun die Zufallsvariable $Z = X - Y$. Ist die Varianz von Z größer, kleiner oder gleich groß wie die Varianz von X bzw. Y?

Aufgabe 3.7
Sie werfen 12 Mal mit einem fairen Würfel, der die Augenzahlen 1, 2, ...,6 zeigt. Berechnen Sie die Wahrscheinlichkeit, dass Sie

a) mindestens zweimal eine „6“

b) höchstens 4 gerade Augenzahlen

werfen.

Aufgabe 3.8
Textilgroßhändler Karl-Otto Rupt vertreibt unter anderem Baumwollhosen aus der Kollektion von Stefan Rangler, die hochwertigen „St.-Rangler-Jeans“. Er bezieht jedoch nicht nur 20 Original-St.-Rangler-Jeans, sondern kauft zusätzlich noch 80 Billig-Imitate, die er zufällig unter die Originalware mischt. Auf den ersten Blick können Originaljeans und Billig-Imitate nicht unterschieden werden. Lediglich nach dem ersten Waschgang zeigen sich deutliche Mängel bei den Imitaten.
Boutique-Besitzerin Ann Ungslos erwirbt bei K.-O. Rupt insgesamt zehn Jeans, die zufällig dem Lager entnommen werden.

a) Wie groß ist die Wahrscheinlichkeit, dass Ann Ungslos genau zwei Originaljeans erwirbt?

b) Wie groß ist die Wahrscheinlichkeit, dass Ann mindestens zwei Originaljeans erhält?

c) Wie groß ist die Wahrscheinlichkeit, dass die Boutique-Besitzerin mindestens sieben Imitate bekommt?

d) Mit welcher Wahrscheinlichkeit sind mindestens 2 aber höchstens 4 Originaljeans in der Lieferung an Ann Ungslos?

Aufgabe 3.9
Die Lebenszeitprävalenz von Personen über 70 Jahren mit Diabetes mellitus beträgt laut Robert Koch Institut etwa 20%.

a) Eine Pflegeeinrichtung betreut auf einer Station 20 Patienten, die älter als 70 Jahre sind. Wie groß ist die Wahrscheinlichkeit, dass mehr als 5 Patienten mit Diabetes mellitus zu betreuen sind?

b) Eine Krankenkasse hat 20.000 Mitglieder, die älter als 70 Jahre sind. Wie groß ist die Wahrscheinlichkeit, dass die Krankenkasse für mehr als 5.000 Mitglieder die Behandlungskosten von Diabetes mellitus zahlen muss?

c) Woran liegt es, dass in a) und b) nicht die gleiche Wahrscheinlichkeit erzielt wird?

Aufgabe 3.10
Eine medizinische Untersuchungstätigkeit sei (approximativ) normalverteilt mit einem Erwartungswert von 4,5 Minuten und einer Standardabweichung von 0,9 Minuten.

a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Untersuchung höchstens 6 Minuten beträgt?

b) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Untersuchung mindestens 4,2 Minuten beträgt?

c) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Untersuchung zwischen 3 und 6 Minuten dauert?

d) Welche Dauer wird von 95% aller Untersuchungen überschritten?

Aufgabe 3.11
Eine Studierendenvertretung richtet ein Sommerfest aus, zu der sie Speisen und Getränke anbietet.

a) Der Getränkelieferant sieht den Großauftrag als willkommenen Anlass, unter die Lieferung auf zufälliger Basis 10% der Bierflaschen durch Flaschen zu ersetzen, deren Mindesthaltbarkeitsdatum bereits abgelaufen ist.
Sie erwerben für sich und Ihre Freunde einen Kasten mit 20 Flaschen. Wie groß ist die Wahrscheinlichkeit, dass Sie mindestens zwei Flaschen erwerben, deren Mindesthaltbarkeitsdatum bereits abgelaufen ist?

b) Die Studierendenvertretung wird von erbosten Besuchern über diesen Mangel informiert und beschließt, sämtliche noch vorhandenen Bierflaschen auf das Mindesthaltbarkeitsdatum zu überprüfen. Es gilt dabei insgesamt 50 Kästen mit jeweils 20 Flaschen zu kontrollieren.
Wie groß ist die Wahrscheinlichkeit, dass die Prüfer mindestens 100 Flaschen mit minderwertigem Bier vorfinden?

c) Wie groß ist die Wahrscheinlichkeit, dass die Prüfer in b) mindestens 15 Flaschen überprüfen müssen, bevor sie zum ersten Mal eine Flasche mit Mangel vorfinden?

d) Am Grill herrscht Hochbetrieb. Bedauerlicherweise haben 5 der 100 eingekauften Steaks auf dem Weg zum Grill eine Zwischenstation auf dem zugegebenermaßen nicht ganz reinen Boden gemacht. Der Chefgriller legt sie dennoch auf den Grill und verkauft sie anschließend.
Sie kaufen für sich und ihre 5 Freunde jeweils ein Steak. Wie groß ist die Wahrscheinlichkeit, dass Sie mindestens eines der Steaks vom Boden erhalten haben?

Aufgabe 3.12
Die Lebensdauer einer Batterie (in Stunden) sei (approximativ) normalverteilt mit Erwartungswert 6.000 und Standardabweichung 120.

a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig der Produktion entnommene Batterie eine Lebensdauer von höchstens 6.100 Stunden aufweist?

b) Berechnen Sie die Wahrscheinlichkeit, dass eine zufällig ausgewählte Batterie eine Lebensdauer von mindestens 6.300 Stunden besitzt.

c) Welche Lebensdauer wird von 90% der Produktion nicht überschritten?

Aufgabe 3.13

Speziell für bulimische und geriatrische Patienten hat ein Hersteller das hochkalorische Nahrungsergänzungsmittel „Pummel-Figur“ als Drink kreiert. Produktionsbedingt unterliegt die Abfüllung leichten Schwankungen, die (approximativ) als normalverteilt mit Erwartungswert 100 ml und Standardabweichung 3 ml angesehen werden kann.

a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Flasche höchstens 105 ml enthält?

b) Welches Füllgewicht wird von 90% aller produzierten Einheiten überschritten?

Aufgabe 3.14

Berechnen Sie $F_X(2)$, d.h. die Verteilungsfunktion an der Stelle 2, wenn bekannt ist, dass

a) $X \sim Bin(20; 0,15)$,

b) $X \sim Hyp(5; 1; 3)$,

c) $X \sim G(0,6)$,

d) $X \sim N(0; 1)$,

e) $X \sim N(1; 4)$,

f) $X \sim R(0; 10)$,

g) $X \sim R(5; 10)$,

h) $X \sim Exp(0,1)$,

i) $X \sim Par(0,8;0,5)$,

Aufgabe 3.15

Tanja Tastisch ist begeisterte Anhängerin des Fußballvereins FC Bayern München. Als fachkundiger Fan weiß sie, dass die Wahrscheinlichkeit, dass ein Profispieler eines Spitzenclubs (FC Bayern München oder Borsussia Dortmund) einen Elfmeter mit Wahrscheinlichkeit von 0,8 verwandelt.

Am 28.04.2015 muss Fan Tastisch im DFB–Halbfinalspiel BVB Dortmund gegen Bayern München bei einem Spielstand von 1:1 in der Verlängerung einen Elfmeter-Krimi erleben, der sich wie folgt darstellt:

Schütze	Name	Verein	Ergebnis
1	Lahm	Bayern München	rutscht aus und verschießt
2	Gündogan	Borussia Dortmund	verwandelt Elfmeter zum 2:1
3	Alonso	Bayern München	rutscht aus und verschießt
4	Kehl	Borussia Dortmund	verwandelt Elfmeter zum 3:1
5	Götze	Bayern München	Langerak hält den Elfmeter, es bleibt beim 3:1
6	Hummels	Borussia Dortmund	Neuer hält den Elfmeter, es bleibt beim 3:1
7	Neuer	Bayern München	verschießt Ball an die Latte, das Spiel ist beendet

Wie groß ist die Wahrscheinlichkeit, dass

a) der FC Bayern München vier von vier Elfmetern verschießt?

b) der BVB Dortmund zwei von drei Elfmetern verwandelt?

c) Bayern München vier Elfmeter verschießt und Dortmund zwei von drei Elfmetern verwandelt?

Betrachten Sie nun allgemein das Aufeinandertreffen zweier Spitzenteams.

d) Welche Mindestanzahl von Elfmeterschüssen ist notwendig, bevor das Spiel zu Ende ist? Berechnen Sie die Wahrscheinlichkeit für dieses Ereignis.

e) Wie groß ist die Wahrscheinlichkeit, dass ein Elfmeterschießen nach sieben Elfmetern beendet ist?

4 Statistische Inferenz

Aufgabe 4.1
Mei Nüssli füllt Nussknabbereien in 100 g Packungen ab. Bedingt durch das schwankende Trockengewicht kommt es zu leichten Abweichungen vom Zielgewicht.
Eine von der Qualitätskontrolle entnommene Stichprobe, die exakt nachgewogen wurde, ergab folgende Werte:

100,1	99,9	100,2	100,0	100,4	100,1
100,0	100,3	100,2	99,8	100,1	

a) Berechnen Sie ein 95%-Konfidenzintervall für den unbekannten mittleren Füllwert, wenn die Grundgesamtkeit als normalverteilt angesehen werden kann.

b) Können Sie zum Niveau 5% nachweisen, dass das mittlere Füllgewicht signifikant mehr als 100 g beträgt?

Aufgabe 4.2
Wilhemine Rahl – seit Kindertagen kurz Minna gerufen – betreibt einen Naturbrunnen und füllt das so gewonnene Wasser, das berühmte „Minna-Rahl-Wasser“, in 500ml-Flaschen ab. Bedingt durch die im Wasser enthaltene Kohlensäure ist der Abfüllprozess leichten Schwankungen unterworfen. Die Füllmenge gilt als (approximativ) normalverteilt.

a) Formulieren Sie das Testproblem, das die hauseigene Qualitätskontrolle anwendet, um Abweichungen vom Zielwert zu entdecken.

b) Formulieren Sie das Testproblem, das das Eichamt anwendet, um dem Hersteller signifikante Unterschreitungen der Füllmenge von 500ml nachzuweisen.

c) Testen Sie zum Niveau 5%

$$H_0: \ \mu = 500 \qquad \text{vs.} \qquad H_0: \ \mu \neq 500,$$

wenn eine Zufallsstichprobe die exakt nachgemessenen Füllmengen (in ml) von

$494,9$	$500,3$	$501,5$	$496,9$	$499,7$
$502,1$	$496,5$	$498,8$	$501,7$	$500,6$

ergab.

d) Betrachten Sie nun die Stichprobe aus c) und die darin ermittelten Ergebnisse für Stichprobenmittel und Stichprobenvarianz als Pretest. In einer zukünftigen Stichprobe soll das Füllgewicht ermittelt werden, das mit mindestens 99%iger Wahrscheinlichkeit höchstens eine (absolute) Toleranz von 1 ml aufweisen soll. Welcher Stichprobenumfang ist dazu nötig?

Aufgabe 4.3

Betrachten Sie folgende Testprobleme:

a) In einer Erhebung von 50 Unternehmen zum beruflichen Gesundheitsmanagement (BGM) ergab die Korrelation zwischen den Merkmalen Umsatz und Ausgaben für das BGM einen Wert von 0,3. Testen Sie zum Niveau 5%, ob dieser Zusammenhang signifikant ist.

b) In der gleichen Erhebung von 50 Unternehmen wurde der Zusammenhang zwischen zwei nominal skalierten Merkmalen gemessen. Das eine Merkmal hatte drei, das andere Merkmal 2 Merkmalsausprägungen. Dabei ergab sich ein χ^2-Wert von 4,95 (das entspricht einem unkorrigierten Kontingenzkoeffizienten von 0,3). Testen Sie ebenfalls zum Niveau 5%, ob dieser Zusammenhang signifikant ist.

Aufgabe 4.4

Ein Arzneihersteller möchte die Wirkungsdauer (in Stunden) der beiden Medikamente Ratiofirm[2] und Inspirin[3] gegen Demenz untersuchen. Dazu werden jeweils 100 zufällig ausgewählte Demenzerkrankte mit einem der beiden Medikamente therapiert.

Die durchschnittlichen Wirkungsdauern (in Stunden) und deren Standardabweichungen sind der nachfolgenden Tabelle zu entnehmen:

[2] ratio (lat.): die Vernunft; firmare (lat.) stärken, bestärken

[3] inspirare (lat.): einhauchen

Präparat	Stichprobenmittel	Stichprobenstandardabweichung
Inspirin	3,8	1,6
Ratiofirm	3,6	1,8

a) Formulieren Sie das Testproblem, wenn Sie Unterschiede in den Wirksamkeiten aufzeigen möchten.

b) Formulieren Sie das Testproblem und führen Sie anschließend den Test durch, wenn Sie zum Niveau 5% zeigen möchten, dass Inspirin eine längere Wirkungsdauer hat als Ratiofirm.

Aufgabe 4.5
Harald Boh – von seinen Freunden kurz „Harry“ gerufen – ist leidenschaftlicher Liebhaber von Gummibärchen. Seine zweite Leidenschaft gilt der induktiven Statistik, so dass er beim Öffnen der nächsten Tüte gleich testen möchte, ob die verschiedenen Geschmacksrichtungen gleich verteilt sind. Dazu erhebt Harry Boh folgende Werte:

Farbe (Geschmack)	Häufigkeit
rot (Himbeer)	5
violett (Johannisbeer)	10
grün (Apfel)	7
gelb (Limette)	8
orange (Pfirsich)	12
weiß (Birne)	9

Testen Sie zum Niveau 5%, ob die Farben/Geschmacksrichtungen gleichmäßig verteilt sind.

Aufgabe 4.6
Betrachten Sie die Daten aus Beispiel 2.6.
Ein unkundiger Empiriker, der den korrigierten Kontingenzkoeffizienten nach Pearson berechnet dabei einen sehr geringen Wert erhalten hat (vgl. Lösung 2.6), schließt daraus, dass kein nennenswerter Zusammenhang existiert.

a) Testen Sie zum Niveau 1%, ob ein signifikanter Zusammenhang zwischen Klasse und Rettungsstatus besteht.

b) Halten Sie den Zusammenhang auch für marginal gering? Beschreiben Sie den Zusammenhang.

Aufgabe 4.7

Eine Wirtschaftsbranche besitzt 3.000 Unternehmen. Sie möchten mit jeweils 95%–iger Wahrscheinlichkeit folgende Größen mit entsprechender Genauigkeit schätzen:

a) den Anteil der Unternehmen, die am Programm „Seniorenfreundliches Unternehmen“ teilnehmen, auf 2 Prozentpunkte genau, wenn Sie

   i) wissen, dass maximal 25% der Unternehmen an dem Programm teilnehmen,

   ii) keinerlei Kenntnisse über den unbekannten Anteil haben.

b) den durchschnittlichen Umsatz der Branche mit einer maximalen Toleranz von 1%. Ein Pretest ergab eine relative Streuung (Variationskoeffizient) von 13%.

Wie viele Unternehmen müssen Sie jeweils bei reiner Zufallsauswahl (Ziehen ohne Zurücklegen) befragen?